ゲノム解析で活躍するコンピューター

～ゲノム情報のためのデータベースと遺伝子機能の解析～

京都大学化学研究所
バイオインフォマティクスセンター
五斗 進

科学カフェ京都　第62回定例会　京都大学　2010/4/10

本日お話する内容

- ゲノムとゲノムプロジェクトについて
- ゲノムデータとデータベースについて
- ゲノムデータを使った解析について
  - 遺伝子の機能を調べる…

ゲノムとは

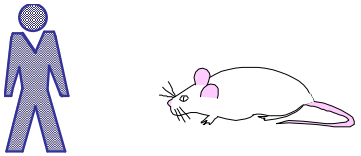
ゲノム（Genome）

・Gene（遺伝子）+ -ome（全体）：遺伝子の総体

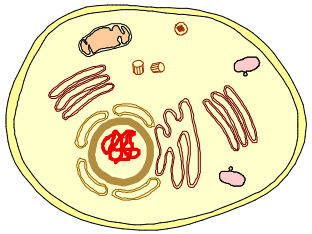
・Gene（遺伝子）+ Chromosome（染色体）

ゲノムとは

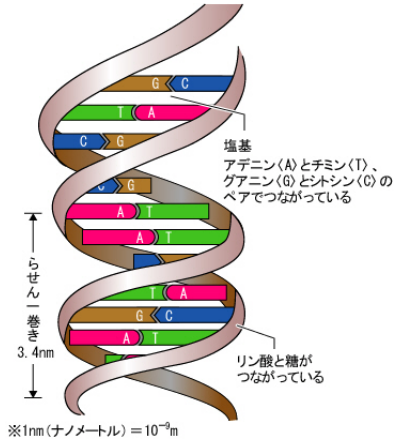
個体



細胞



染色体

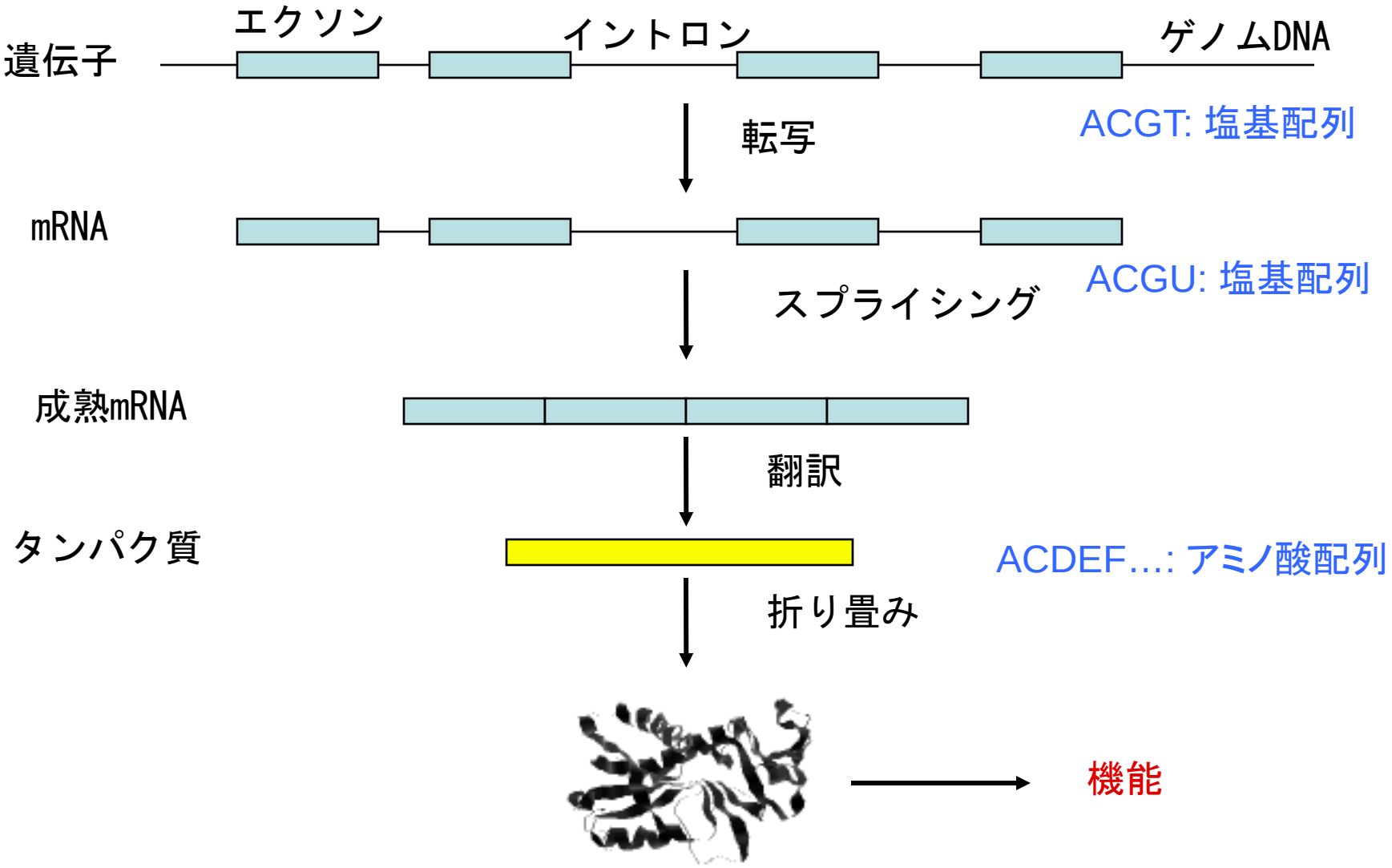


DNAの２重らせん構造

```
atggcgacccgcagccctggcgtcgtgattagtgatgatgaaccaggttatgaccttgat
ttattttgcatacctaatcattatgctgaggatttggaaagggtgtttattcctcatgga
ctaattatggacaggactgaacgtcttgctcgagatgtgatgaaggagatgggaggccat
cacattgtagccctctgtgt.....
```

ゲノムの全塩基配列

DNAとタンパク質（セントラルドグマ）



ゲノム研究の歴史

1900年前後	遺伝法則の発見（メンデル） ゲノム概念の提唱（ヴィンクラー）
1950年代	DNA２重らせん構造の発見 （ワトソン・クリック）
1960年代	遺伝暗号の解読（ニーレンバーグら）
1970年代	DNA塩基配列決定法の開発 ΦX174ファージゲノム配列決定（サンガーら）
1980年代	PCR法の開発（マリス） ヒトゲノム計画の提案（ダルベッコ）
1990年代	ヒトゲノム計画がスタート

ゲノム研究の歴史

1995年	ヘモフィルス菌ゲノムの解読（約200万塩基、2000遺伝子） （独立生活する生物のゲノム）
1996年	出芽酵母ゲノムの解読（約1200万塩基、6000遺伝子） （真核生物のゲノム）
1997年	枯草菌ゲノムの解読（約400万塩基、4000遺伝子） （日本を中心としたグループによる解読）
1998年	線虫ゲノムの解読（約9700万塩基、20000遺伝子） （多細胞生物のゲノム）
1999年	ヒト22番染色体ゲノムの解読
2000年	ショウジョウバエ、シロイヌナズナゲノムの解読
2001年	ヒトゲノムの概要配列発表（約30億塩基）

ポストゲノムプロジェクト

- トランスクリプトーム（Transcript + -ome）
  - 転写産物（RNAのこと）の総体
  - 細胞内で実際に mRNA として発現している遺伝子
- プロテオーム（Protein + -ome）
  - タンパク質の総体
  - 細胞内で実際にタンパク質として働いている遺伝子
- メタボローム（Metabolite + -ome）
  - 代謝産物（アミノ酸、グルコースなど）の総体
  - 代謝系で合成されている化合物

本日お話する内容

- ゲノムとゲノムプロジェクトについて
- ゲノムデータとデータベースについて
- ゲノムデータを使った解析について
  - 遺伝子の機能を調べる…

塩基配列データベース

- 1970年代の配列決定技術
  - サンガー法、マキサム・ギルバート法
- 自動DNAシーケンサーの開発
- 大量の塩基配列の産出
  - PIRタイプから著者によるサブミットへ
  - 三極体制によるデータの収集
    - 日本：DDBJ (DNA DataBank of Japan) @ 遺伝研
    - 米国：GenBank @ National Center for Biotechnology Information, National Institute of Health
    - 欧州：EMBL @ European Molecular Biology Laboratory
- ゲノムプロジェクト

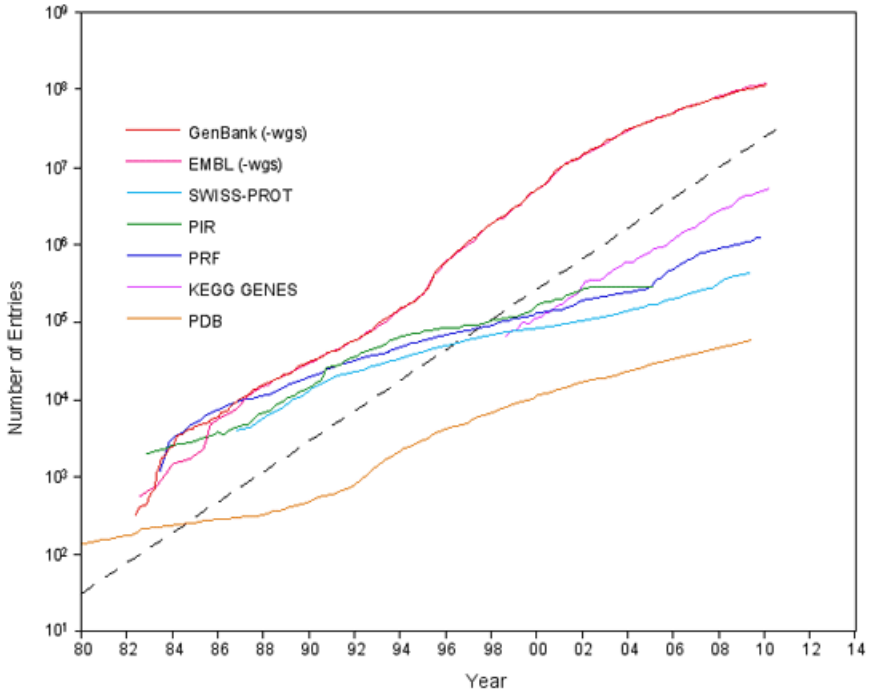
GenBank の例

```
LOCUS       X00617                1338 bp    DNA     linear   BCT 12-SEP-1993
DEFINITION  E.coli triose phosphate isomerase gene (TPI) (EC 5.3.1.1).
ACCESSION   X00617
VERSION     X00617.1  GI:43111
KEYWORDS    glycolysis gluconeogenesis; isomerase.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1338)
  AUTHORS   Pichersky,E., Gottlieb,L.D. and Hess,J.F.
  TITLE     Nucleotide sequence of the triose phosphate isomerase gene of
            Escherichia coli
  JOURNAL   Mol. Gen. Genet. 195 (1-2), 314-320 (1984)
   PUBMED   6092857
COMMENT     Data kindly reviewed (30-MAY-1985) by L.D. Gottlieb.
```

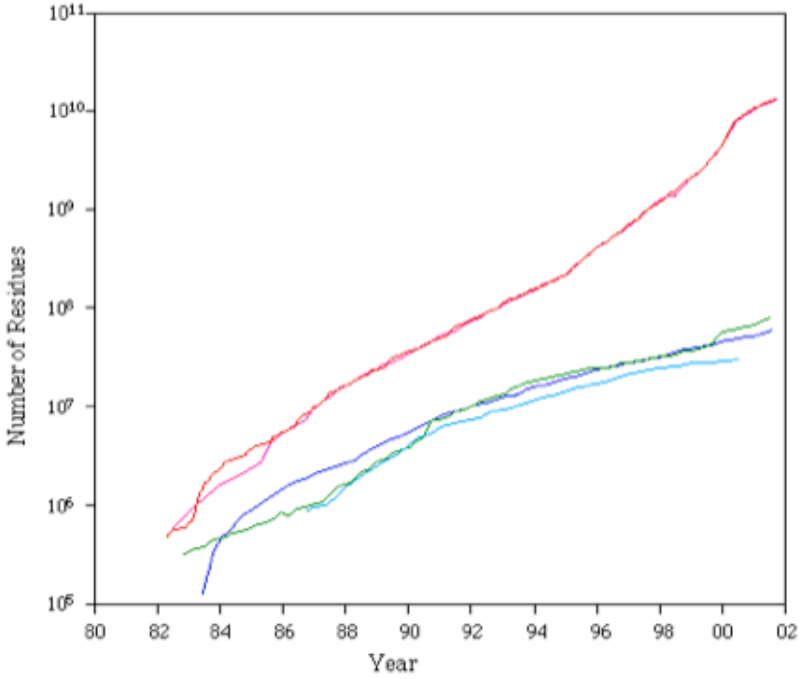
```
FEATURES             Location/Qualifiers
     source          1..1338
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     CDS             220..987
                     /note="unnamed protein product; isomerase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA25253.1"
                     /db_xref="GI:43112"
                     /db_xref="GOA:P0A858"
                     /db_xref="PDB:1TMH"
                     /db_xref="PDB:1TRE"
                     /db_xref="UniProtKB/Swiss-Prot:P0A858"
                     /translation="MRHPLVMGNWKLNGSRHMVHELVSNLRKELAGVAGCAVAIAPPE
                     MYIDMAKREAEGSHIMLGAQNVNLNLSGAFTGETSAAMLKDIGAQYIIIGHSERRTYH
                     KESDELIAKKFAVLKEQGLTPVLCIGETEAENEAGKTEEVCARQIDAVLKTQGAAAFE
                     GAVIAYEPVWAIGTGKSATPAQAQAVHKFIRDHIAKVDANIAEQVIIQYGGSVNASNA
                     AELFAQPDIDGALVGGASLKADAFAVIVKAAEAAKQA"
ORIGIN
        1 ctgcaggacg cctactaagg cggcggggaa aaacaaacgt tattacaccg agacagaagg
       61 tgcactgcgt tatgttgtcg cggacaacgg cgaaaagggg ctgaccttcg ctgttgaacc
      121 aattaagttg gcgctatctg antctcatac tgtttcacag acctgctgcc ctgcggcggc
      181 caatcttcct ttattcgctt ataagcgtgg agaattaaaa tgcgacatcc tttagtgatg
      241 ggtaactgga aactgaacgg cagccgccac atggttcacg agctggtttc taacctgcgt
      301 aaagagctgg caggtgttgc tggctgtgcg gttgcaatcg caccaccgga aatgtatatc
（中略）
     1021 ttactttcct taactcttcg ccttaacgca aaatctcaca ctgatgatcc tgaatttcct
     1081 cggctgaagc acggttaagc gtcagtagat ttcgttgtgt cgccagcaat acaaatgagt
     1141 tatcactctg ccgtaccatc gccagcccgt agcgtcccat atgttcccgc gcctcaggta
     1201 cttcttctgc cagcatcata aatgggctgc gttgtaccag ttcgctttcc gttacccgac
     1261 gcgcaggtat tcatgcccgc gcaaaccacc tggcagtggc aaccagcggc tgctgatgtt
     1321 cgccagattg ttatcgag
//
```

塩基配列データベースのサイズ

エントリーの数



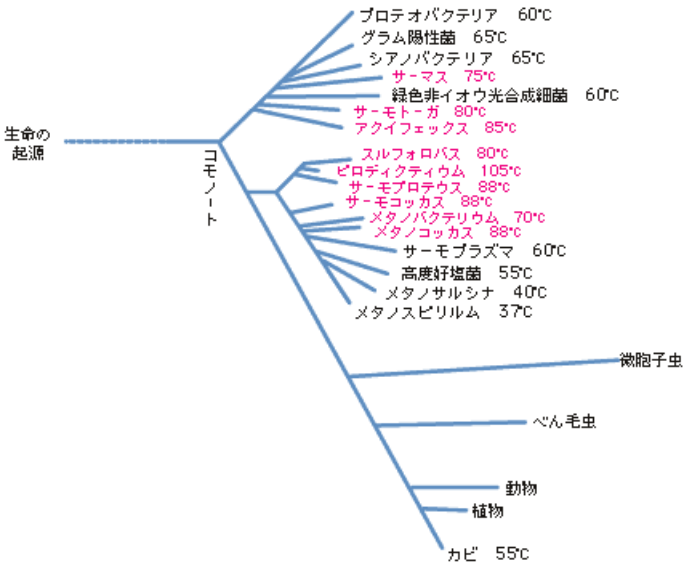
塩基の数

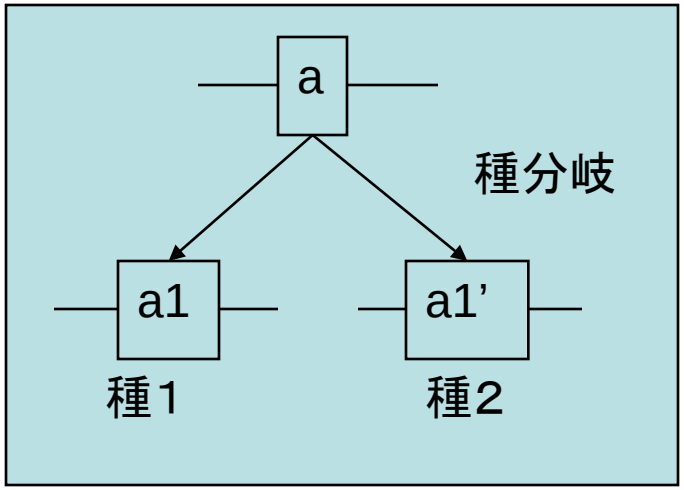


塩基配列の増加は約５年で１０倍のペース

配列データベースの使われ方

- 似たものを探す（ホモロジー検索）
  - ある（機能未知の）遺伝子の塩基配列を実験で決定
  - その配列をもとにして、データベース中に似た配列が登録されていないかを探す
- 前提：同じ機能を持つ遺伝子は似た配列を持つ





配列データベースの使われ方

- 大腸菌のある酵素と同じ機能を持つコレラ菌の酵素をアミノ酸配列で比較した結果

Top

```
>sp:TPIS_VIBCH [Q9KNR1] Triosephosphate isomerase (EC 5.3.1.1) (TIM)
           (Triose-phosphate isomerase).
          Length = 257

 Score =  315 bits (806), Expect = 1e-84
 Identities = 159/239 (66%), Positives = 186/239 (77%), Gaps = 2/239 (0%)
```

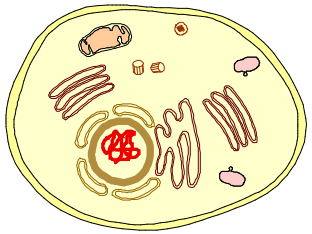
大腸菌
コレラ菌

```
Query: 1   MRHPLVMGNWKLNGSRHMVHELVSNLRKELAGVAGCAVAIAPPEMYIDMAKR--EAEGSH 58
           MR P+VMGNWKLNGS+ MV +L++ L  EL GV G  V +APP MY+D+A+R  +  G+
Sbjct: 1   MRRPVVMGNWKLNGSKAMVTDLLNGLNAELEGVEGVDVVVAPPAMYLDLAERLIKEGGNK 60

Query: 59  IMLGAQNVDLNLSGAFTGETSAAMLKDIGAQYIIIGHSERRTYHKESDELIAKKFAVLKE 118
           ++LGAQN D + SGA+TG+ S AMLKD GA +IIIGHSERR YHKESDE +AKKFA LKE
Sbjct: 61  LILGAQNTDTHNSGAYTGDMSPAMLKDFGASHIIIGHSERRDYHKESDEFVAKKFAFLKE 120

Query: 119 QGLTPVLCIGETEAENEAGKTEEVCARQIDAVLKTQGAAAFEGAVIAYEPVWAIGTGKSA 178
            GLTPV CIGETEA+NEAG+TE VCARQI+AV+   G  A  GA+IAYEP+WAIGTGK+A
Sbjct: 121 NGLTPVFCIGETEAQNEAGETEAVCARQINAVIDAYGVEALNGAIIAYEPIWAIGTGKAA 180

Query: 179 TPAQAQAVHKFIRDHIAKVDANIAEQVIIQYGGSVNASNAAELFAQPDIDGALVGGASL 237
           T   AQ +H  IR  IA  DA +AEQVIIQYGGSV   NAA  FAQPDIDGALVGGASL
Sbjct: 181 TADDAQRIHASIRALIAAKDAAVAEQVIIQYGGSVKPENAASYFAQPDIDGALVGGASL 239
```

ゲノムプロジェクトが出すデータ

個体

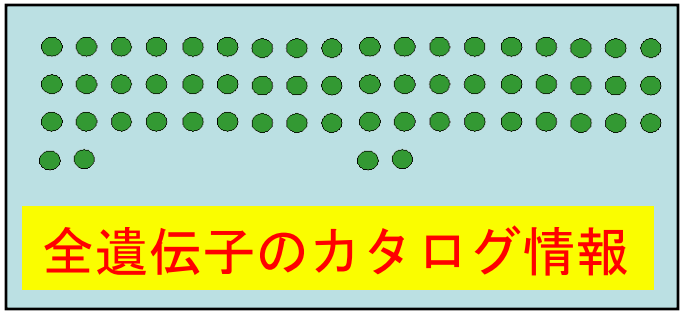
細胞



染色体

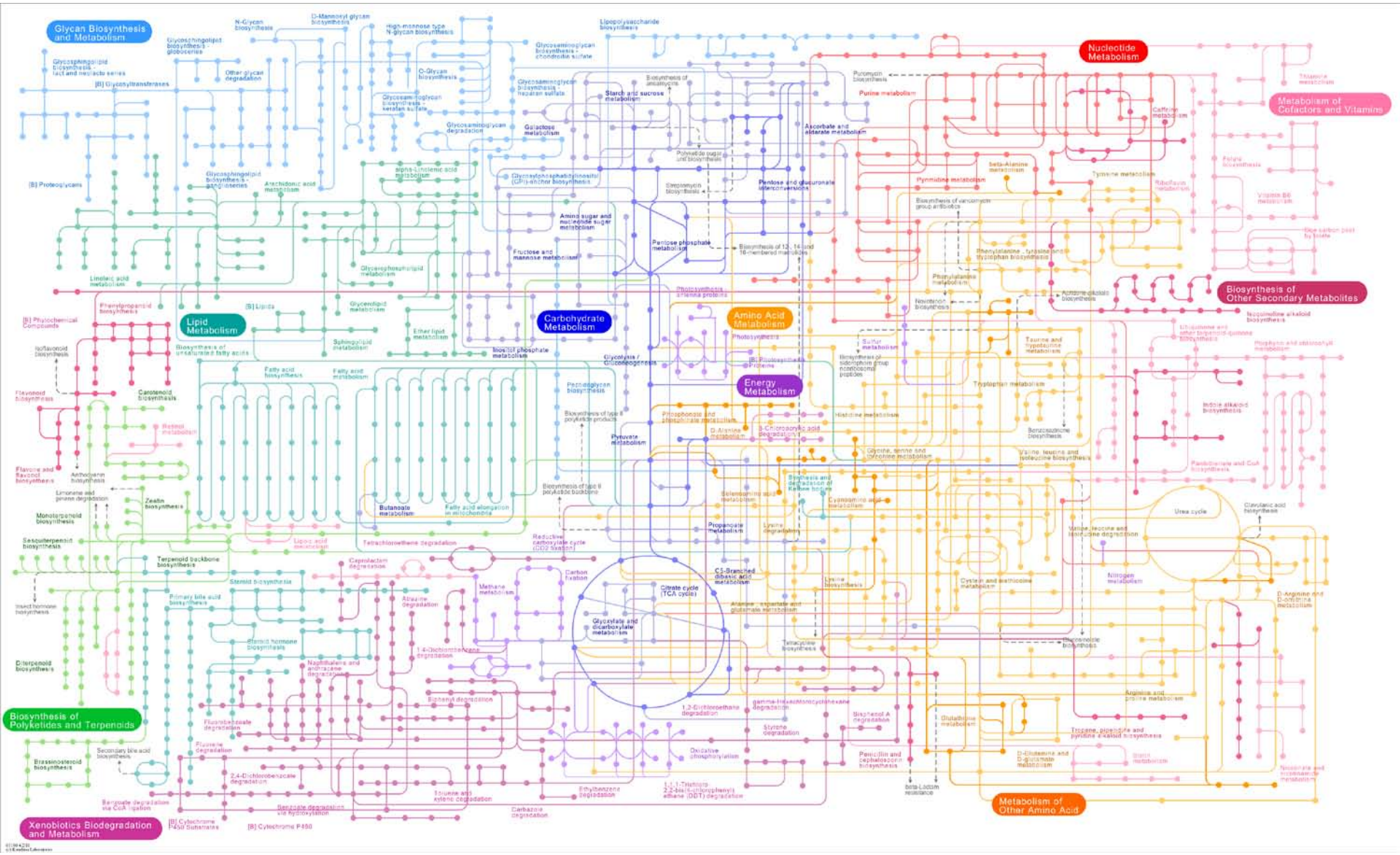
```
atggcgacccgcagccctggcgtcgtgattagtgatgatgaaccaggttatgaccttgat
ttattttgcatacctaatcattatgctgaggatttggaaagggtgtttattcctcatgga
ctaattatggacaggactgaacgtcttgctcgagatgtgatgaaggagatgggaggccat
cacattgtagccctctgtgt.....
```

ゲノムの全塩基配列

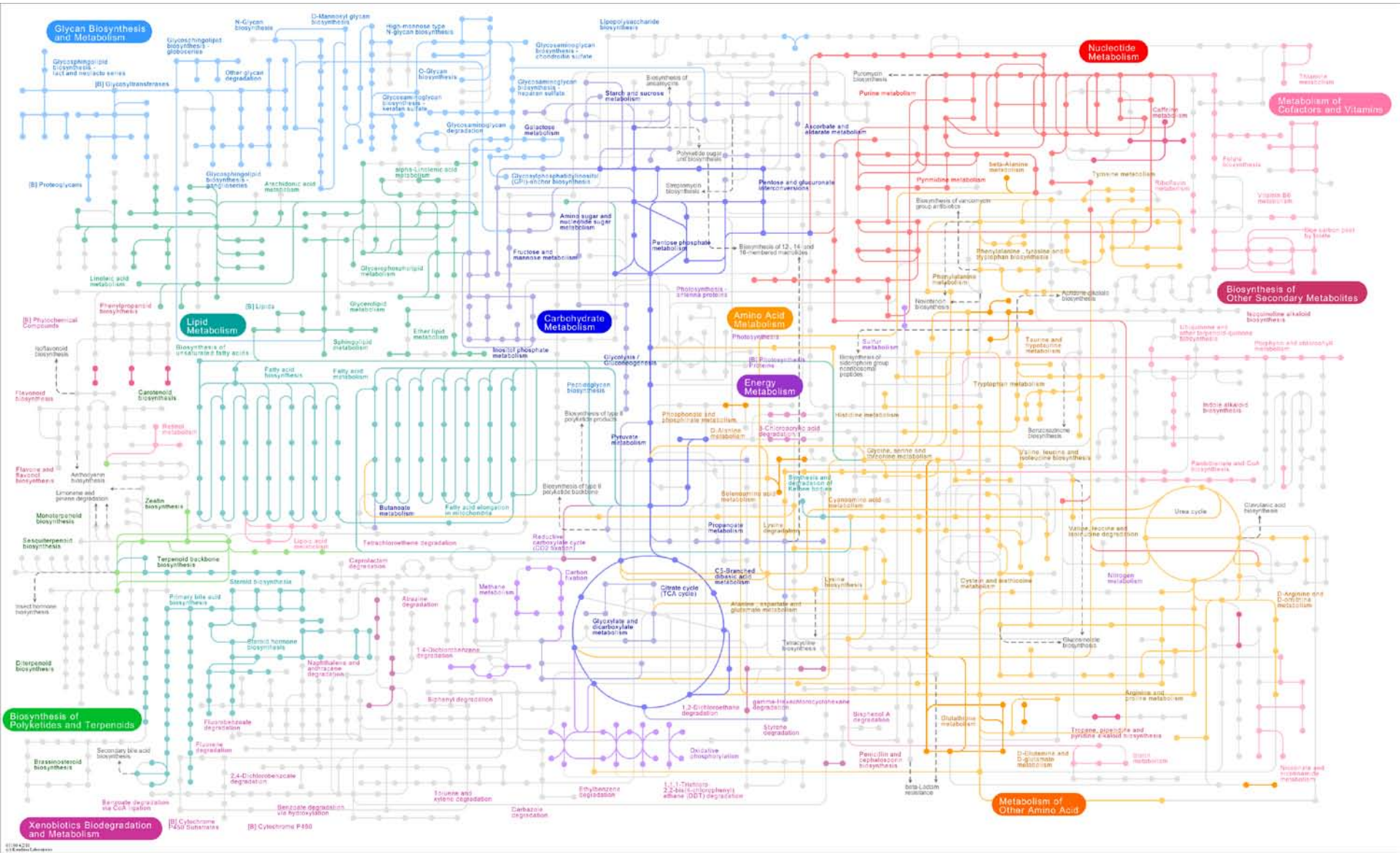


- 個々の遺伝子の機能についての情報
- ホモロジー検索だけで機能が分かる遺伝子は半分～2/3 程度

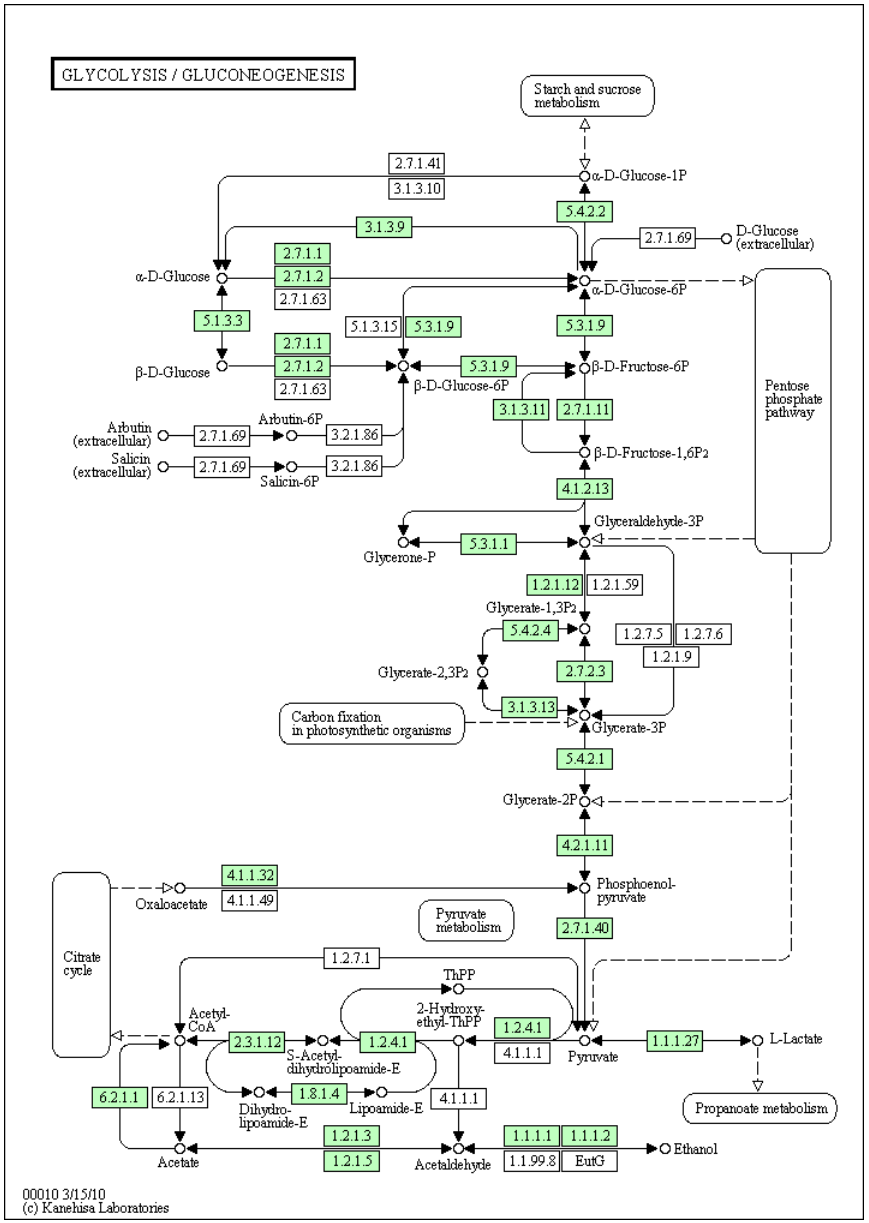
パスウェイの情報


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Biosynthesis of Polyketides and Terpenoids
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

パスウェイの情報


Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Amino Acid Metabolism
Energy Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Other Secondary Metabolites
Biosynthesis of Polyketides and Terpenoids
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid

パスウェイの情報



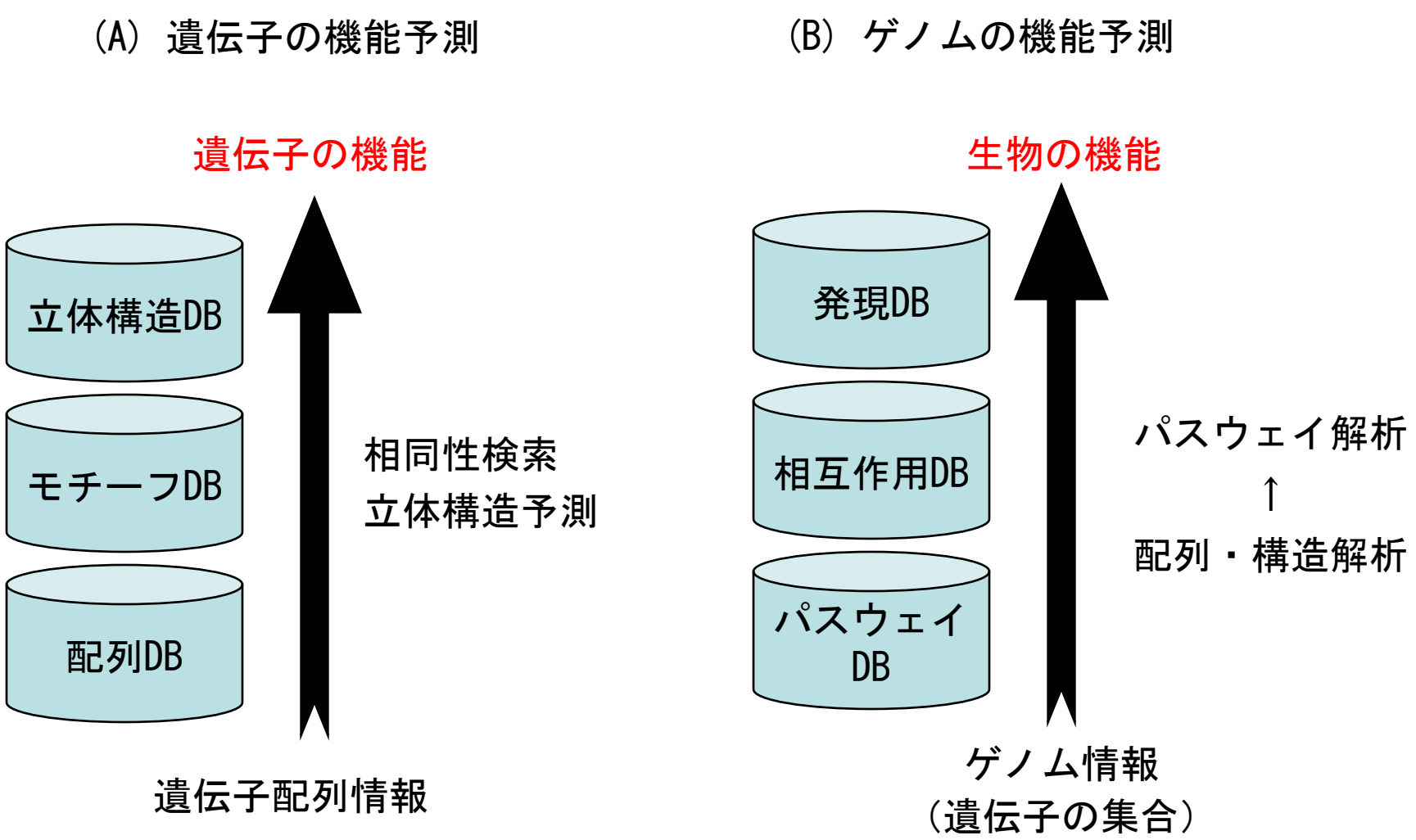
- 酵素／遺伝子と化合物のネットワーク

- ヒトの解糖系（体内に取り込んだ糖を分解して再利用する経路）
  - 緑：ヒトに対応する遺伝子がある酵素
  - 白：ヒトにない酵素

本日お話する内容

- ゲノムとゲノムプロジェクトについて
- ゲノムデータとデータベースについて
- ゲノムデータを使った解析について
  - 遺伝子の機能を調べる…

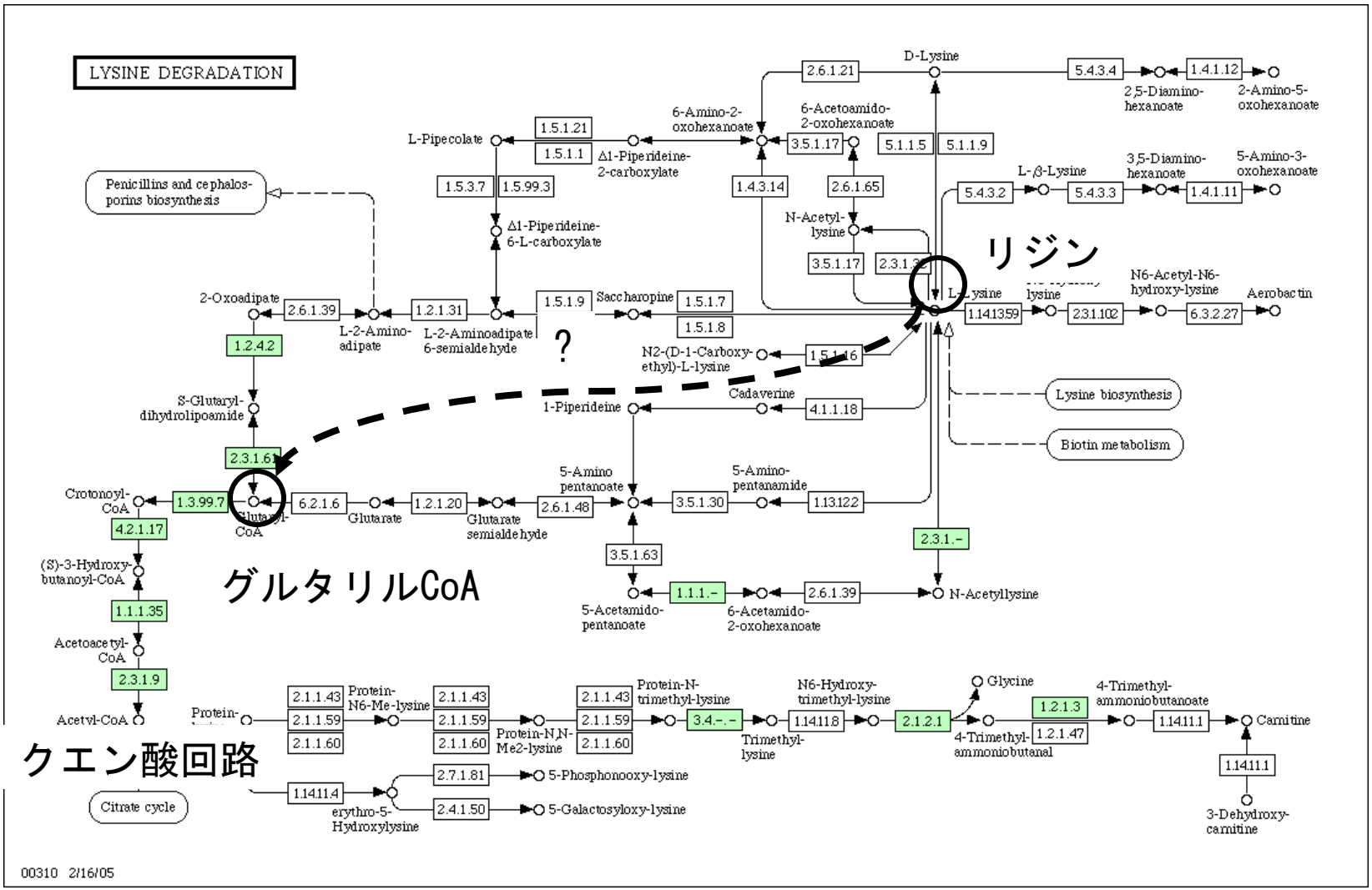
遺伝子の機能予測とゲノムの機能予測



ゲノムの機能予測をした後は

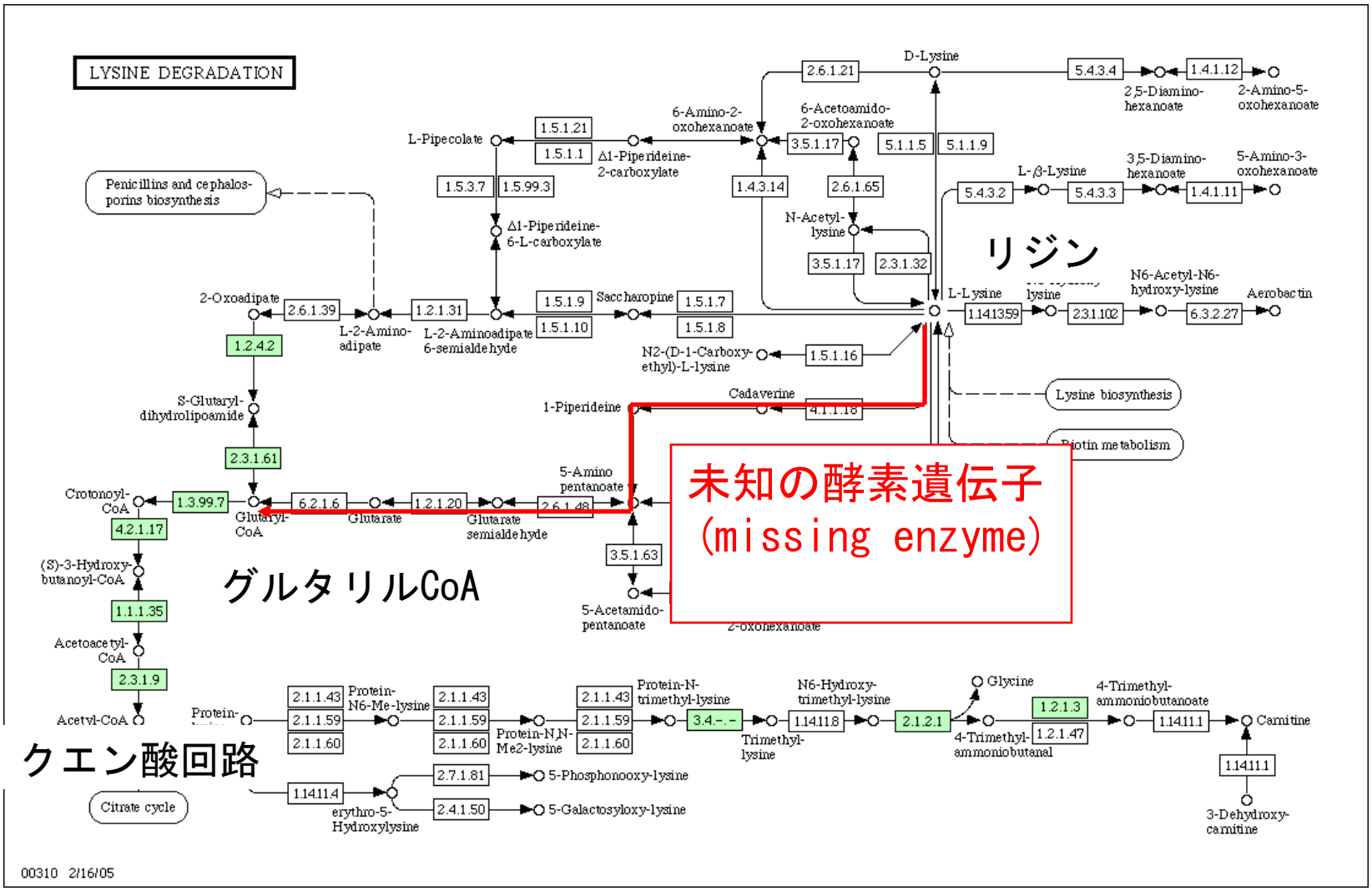
- 機能予測の抜けを探す
  - パスウェイ中で途切れているところ
  - 機能未知遺伝子との対応？
- 様々な情報を比較
  - 種に特徴的な機能は何か？
  - 機能未知遺伝子の機能予測
- パスウェイ（ネットワーク）のトポロジーを解析

機能予測の抜けの例



ゲノム情報から再構築された緑膿菌のリジン分解系

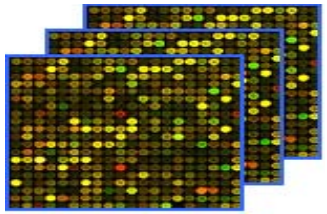
機能予測の抜けの例



生化学的な知識による緑膿菌のリジン分解系

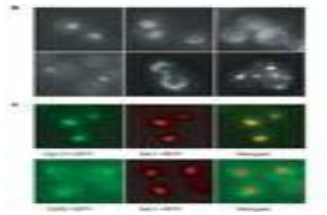
データ統合による知識抽出

マイクロアレイ
遺伝子発現

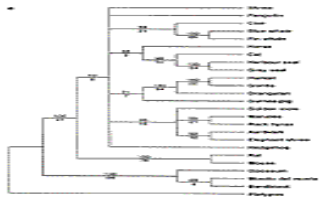


酵母２
ハイブリッド

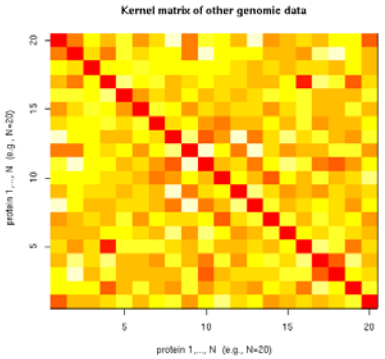
細胞内
局在情報



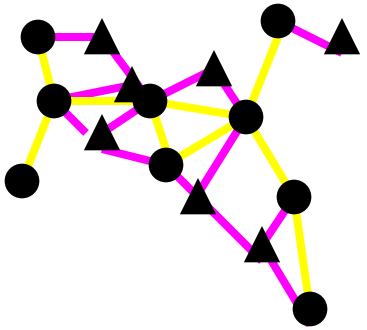
系統
プロファイル



類似度行列
（カーネル）



ネットワーク
推定



仮定：似たパターンを持つ遺伝子同士は機能的に関係している可能性が高い

カーネル

N 個のタンパク質 $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N$ を仮定

カーネル $K(\mathbf{x}, \mathbf{x}')$ は、タンパク質 $\mathbf{x}$ と $\mathbf{x}'$ の類似度

（数学的には、特徴ベクトルの内積）

タンパク質セットの類似度行列

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) \quad (i, j = 1,2,..., N)$$

カーネルの例

遺伝子 $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ の塩基構成比

$$\Phi(\mathbf{x}_1) = \begin{bmatrix} 0.1 \\ 0.4 \\ 0.2 \\ 0.3 \end{bmatrix}, \quad \Phi(\mathbf{x}_2) = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.3 \\ 0.2 \end{bmatrix}, \ldots$$

$$K(\mathbf{x}_1, \mathbf{x}_2) = \Phi(\mathbf{x}_1) \cdot \Phi(\mathbf{x}_2)$$

$$= 0.1 \times 0.2 + 0.4 \times 0.3 + 0.2 \times 0.3 + 0.3 \times 0.2 = 0.26$$

カーネルの例

カーネル行列:

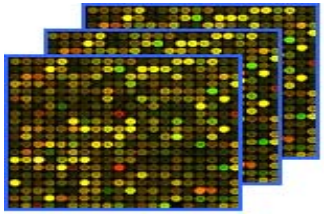
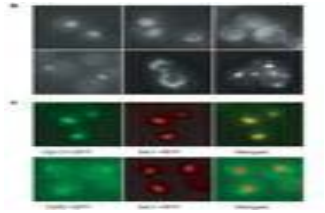
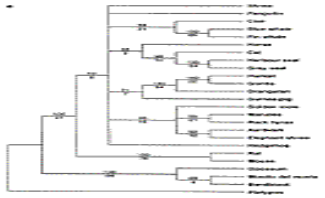
$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_2) \cdots \\ K(\mathbf{x}_2, \mathbf{x}_1) & K(\mathbf{x}_2, \mathbf{x}_2) \cdots \\ \vdots & \vdots \quad \ddots \end{bmatrix} = \begin{bmatrix} 0.3 & 0.26 \cdots \\ 0.26 & 0.26 \cdots \\ \vdots & \vdots \quad \ddots \end{bmatrix}$$

一種の類似度行列とみなせる

||

一種の遺伝子類似性ネットワーク

多様なデータと表現方法

	データ	表現方法
	マイクロアレイ遺伝子発現	数値ベクトル
	酵母2ハイブリッド（タンパク質間相互作用）	グラフ
	細胞内局在	ビットベクトル
	系統プロファイル	ビットベクトル

多様なデータとデータ型

数値ベクトル間の類似度を求める関数

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\| \mathbf{x}_i - \mathbf{x}_j \|^2 / \sigma^2)$$

グラフ上の頂点間の類似度を求める関数

$$K = \exp(-L)$$

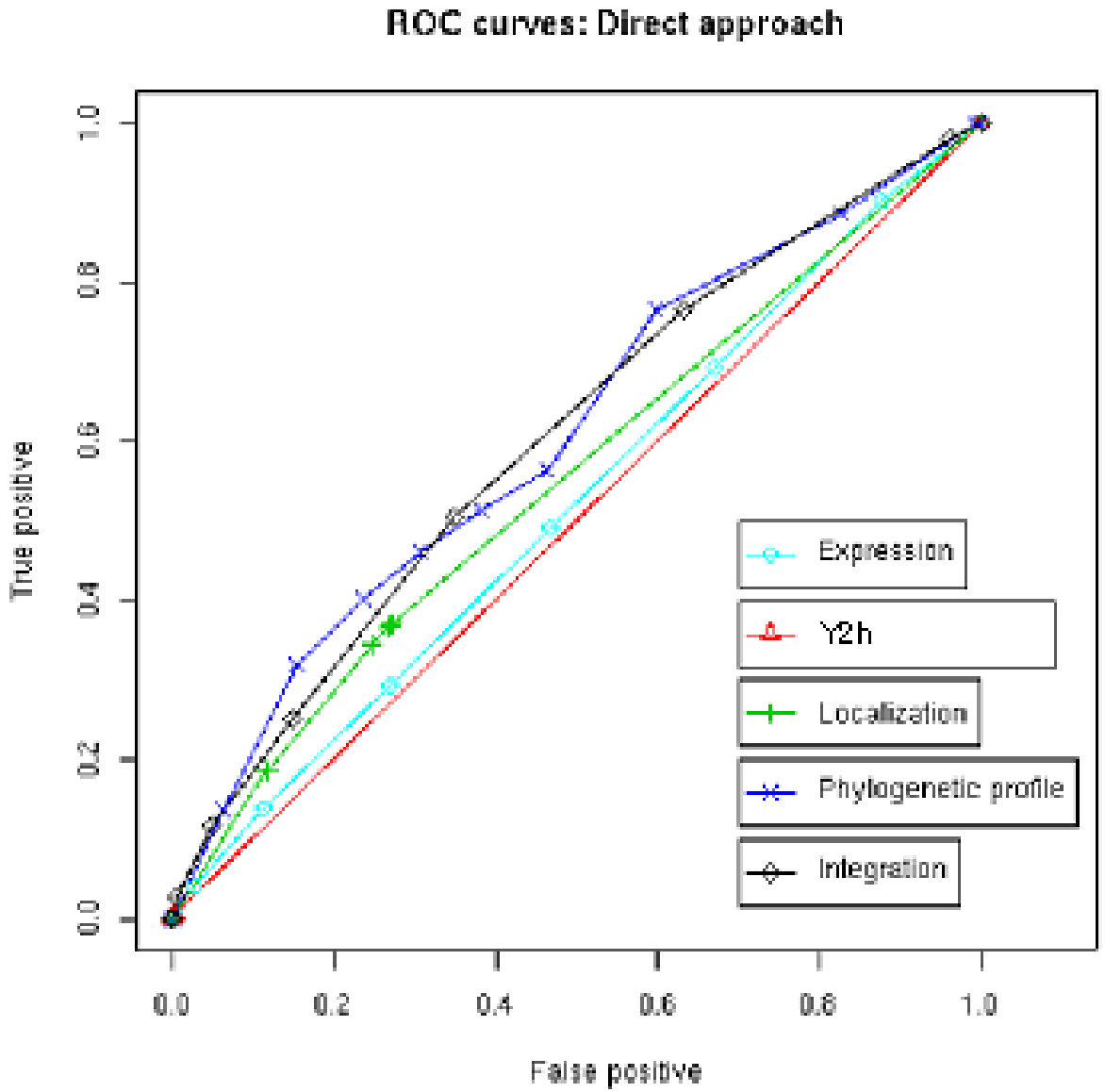
ここでLはグラフのラプラシアン

(Kondor, 2002)

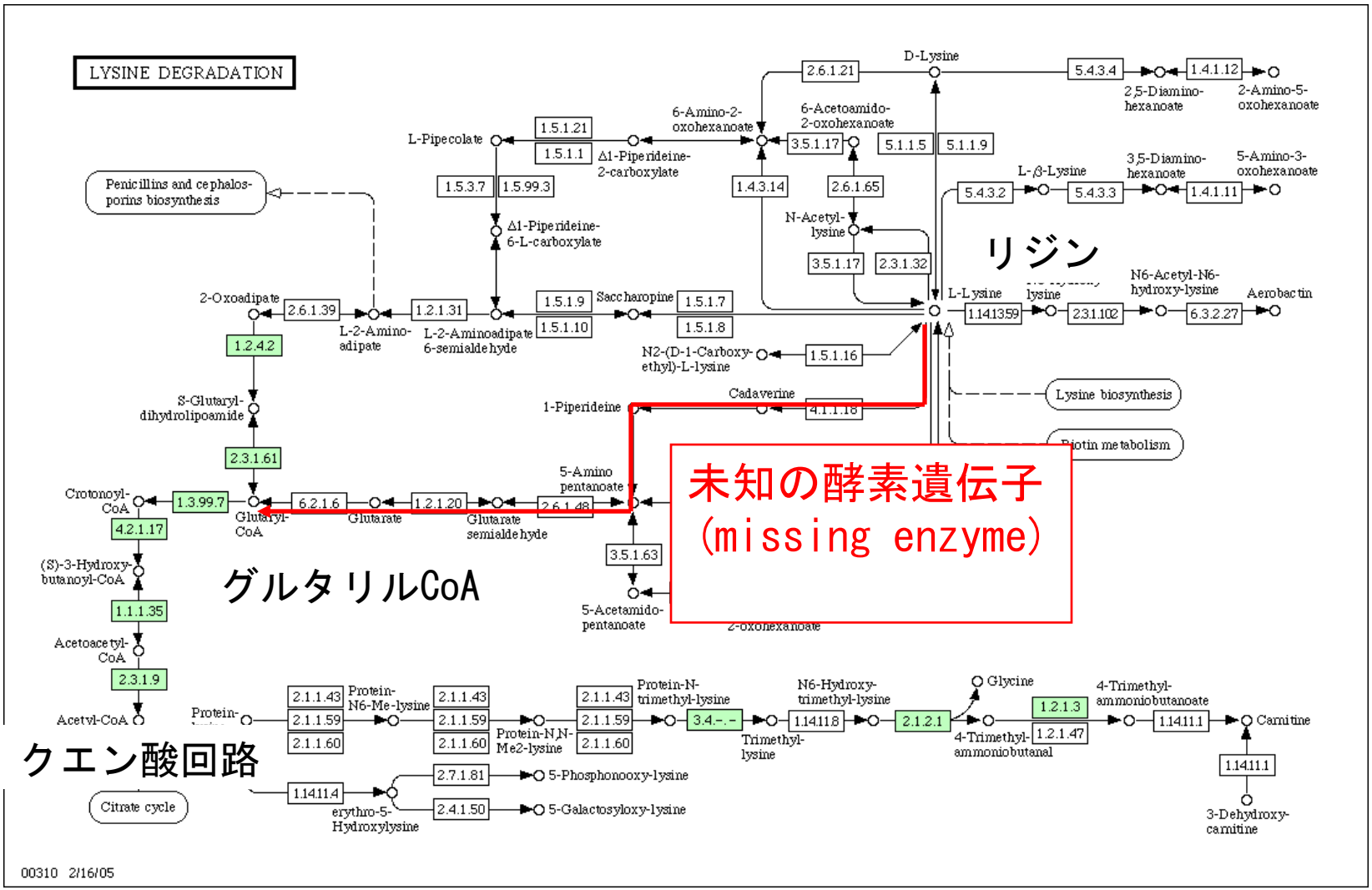
多様なデータとデータ型

- 各データにおける遺伝子間の距離をカーネル行列として定義
  - Kg: ゲノム上での遺伝子間の距離
  - Ke: 発現パターンの類似度
  - Kp: 系統プロファイルの類似度
- カーネルの和を取る
  - K = Kg + Ke + Kp
- 統合されたカーネル K を用いて遺伝子間の関係を変換
- 教師付き学習

教師なしの場合


ROC curves: Direct approach
True positive
False positive
0.0
0.2
0.4
0.6
0.8
1.0
Expression
Y2h
Localization
Phylogenetic profile
Integration

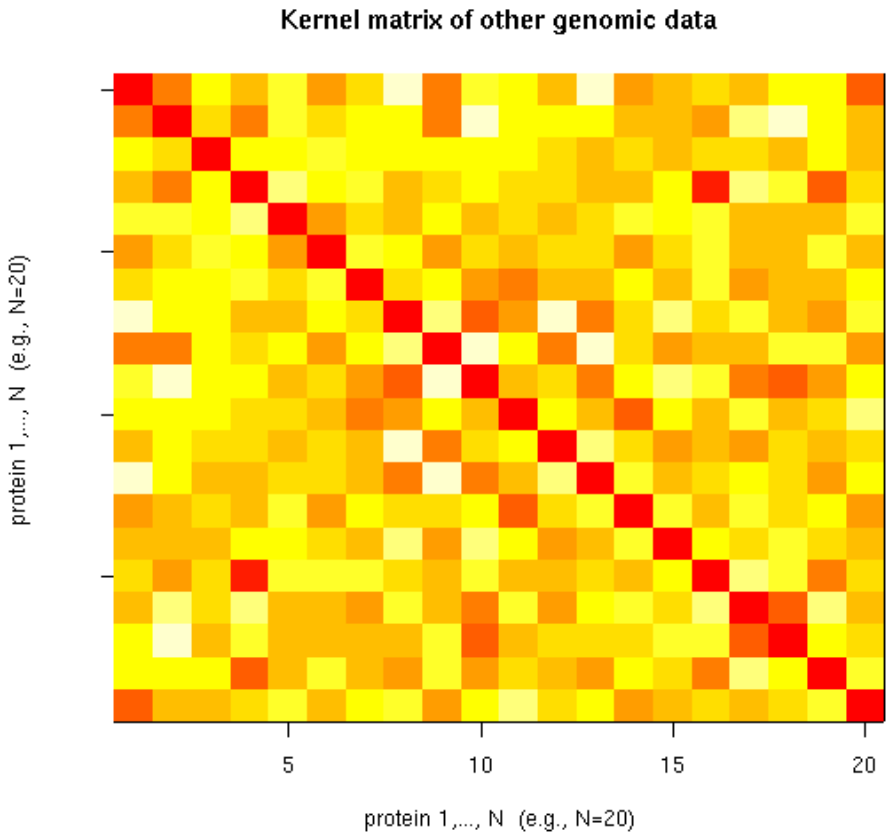
機能予測の抜けの例



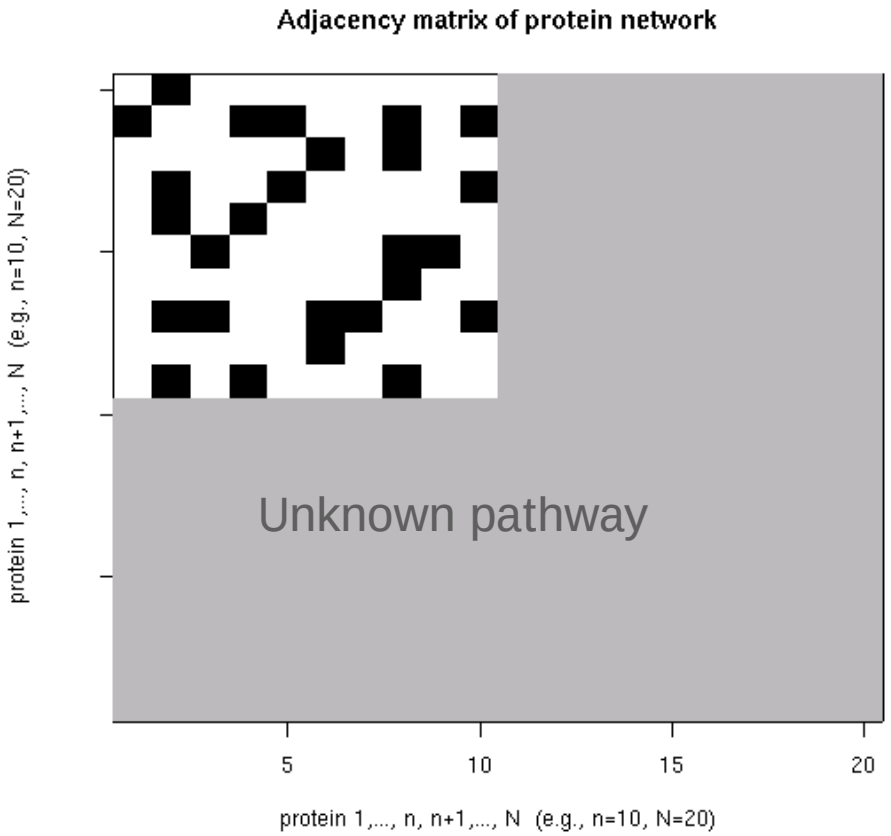
生化学的な知識による緑膿菌のリジン分解系

教師付き学習

発現データの類似度行列

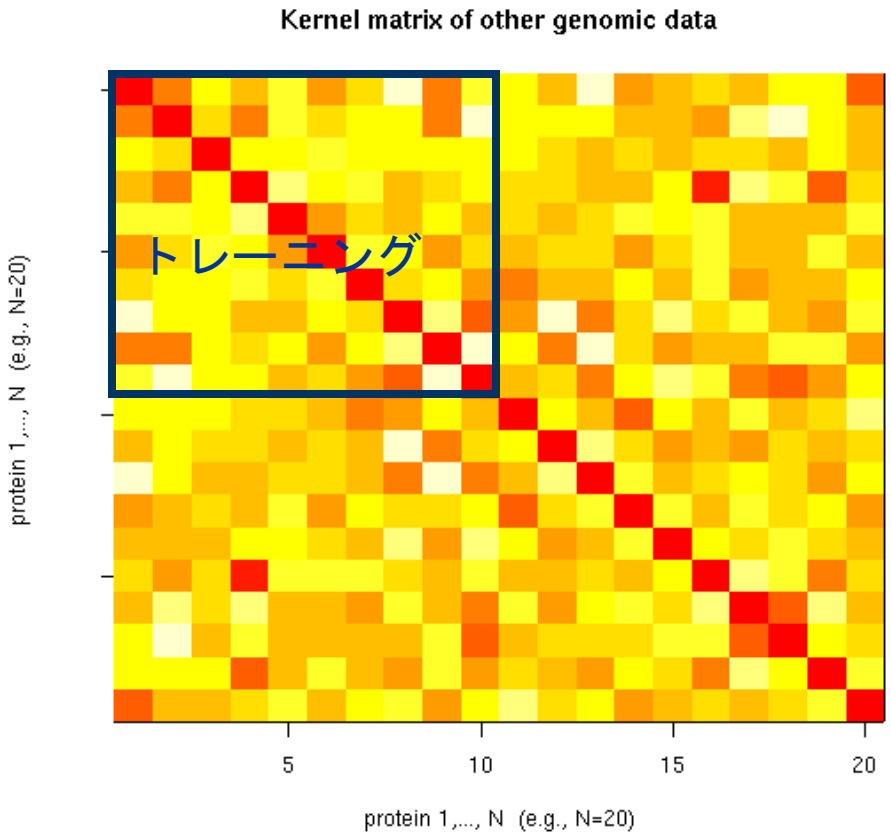


タンパク質ネットワーク

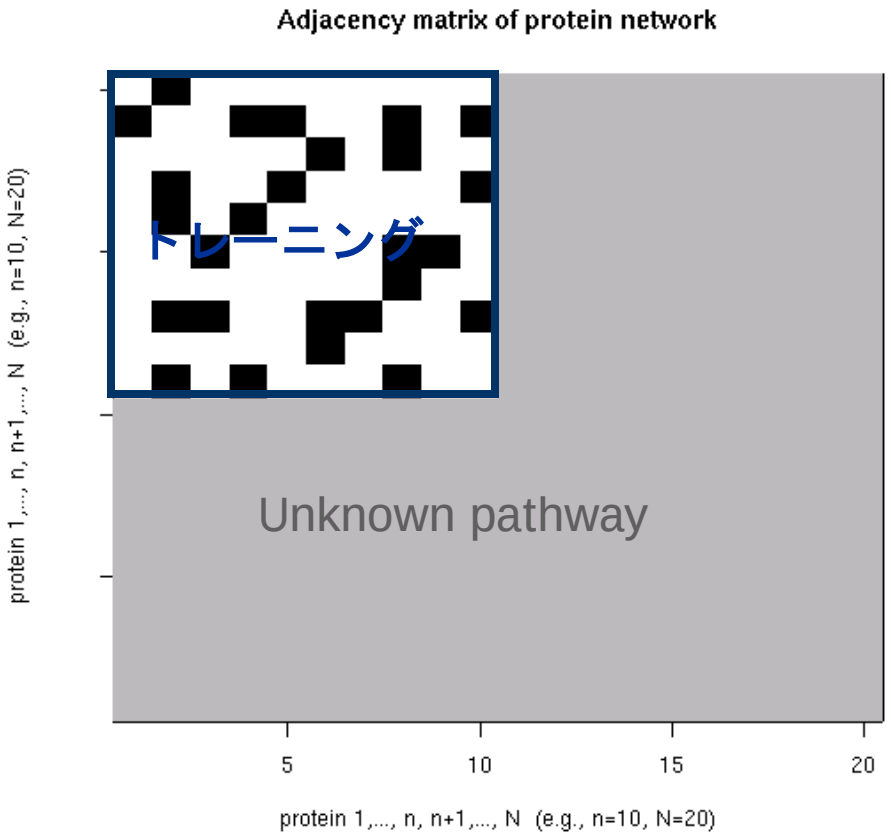


教師付き学習

発現データの類似度行列

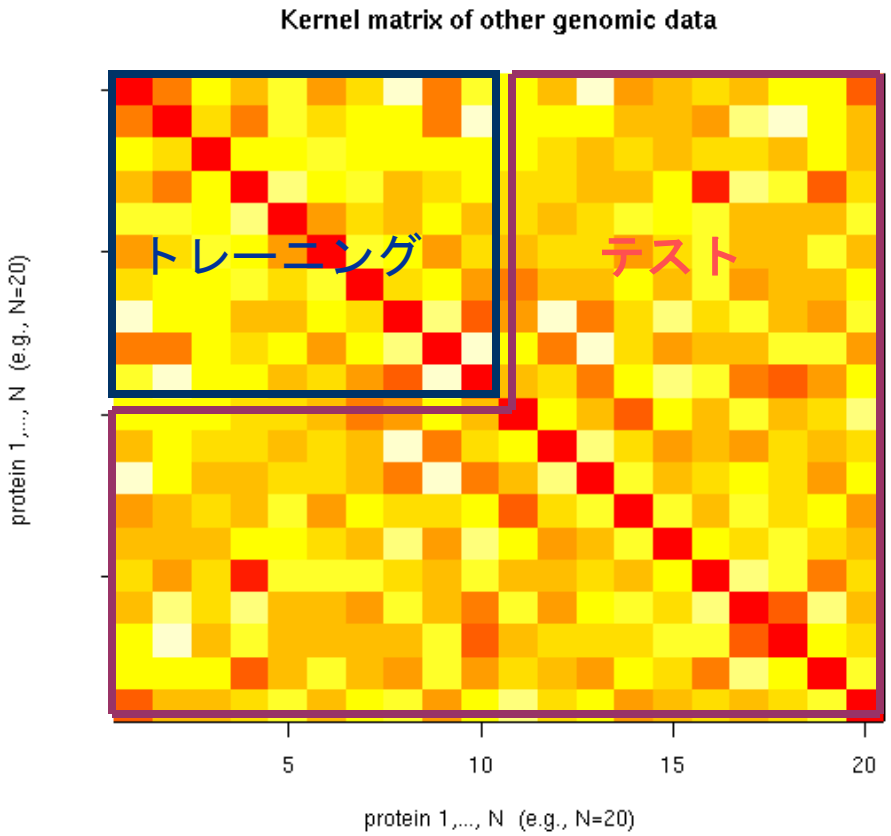


タンパク質ネットワーク

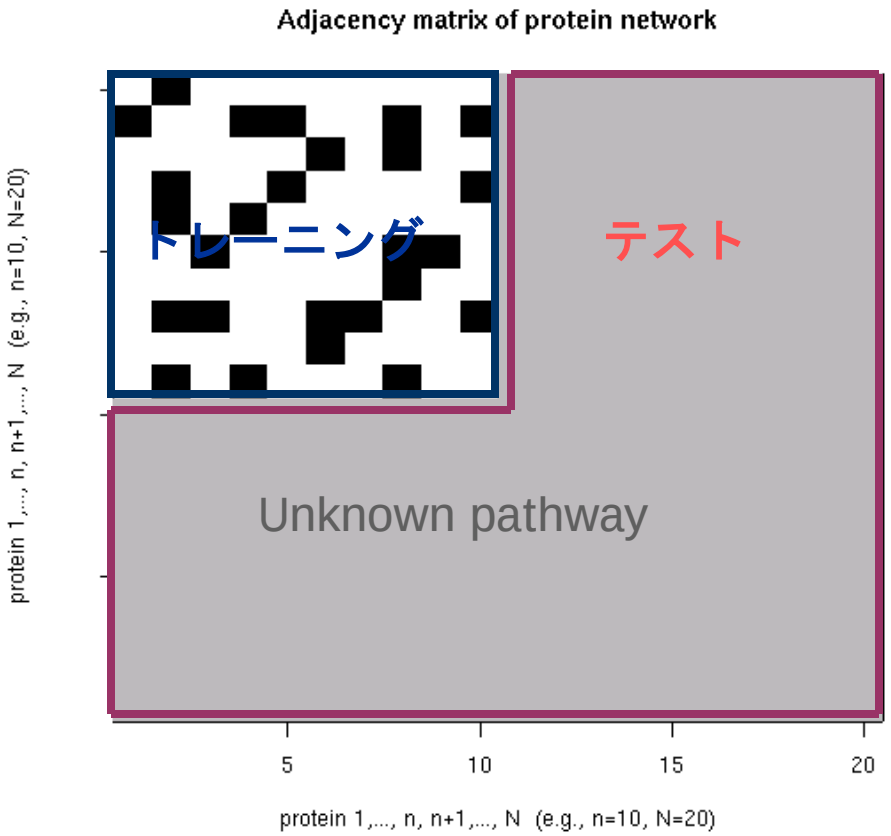


教師付き学習

発現データの類似度行列

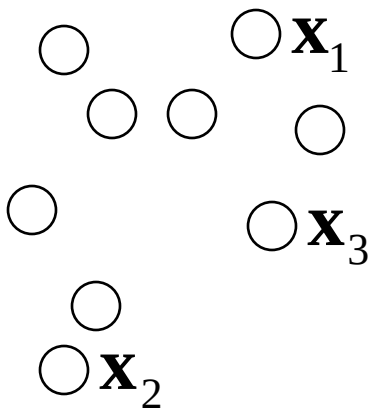


タンパク質ネットワーク



教師付き学習

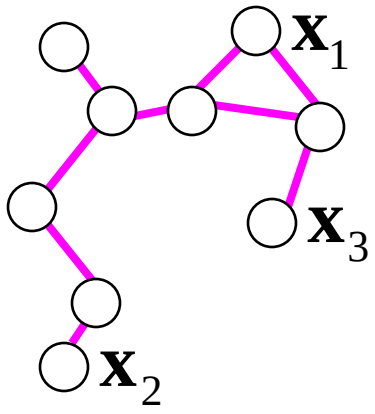
元の空間



○ ：トレーニングセット

教師付き学習

元の空間

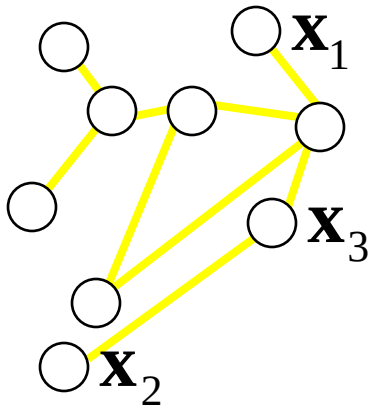


：教師なしで直接予測した結果

：トレーニングセット

教師付き学習

元の空間

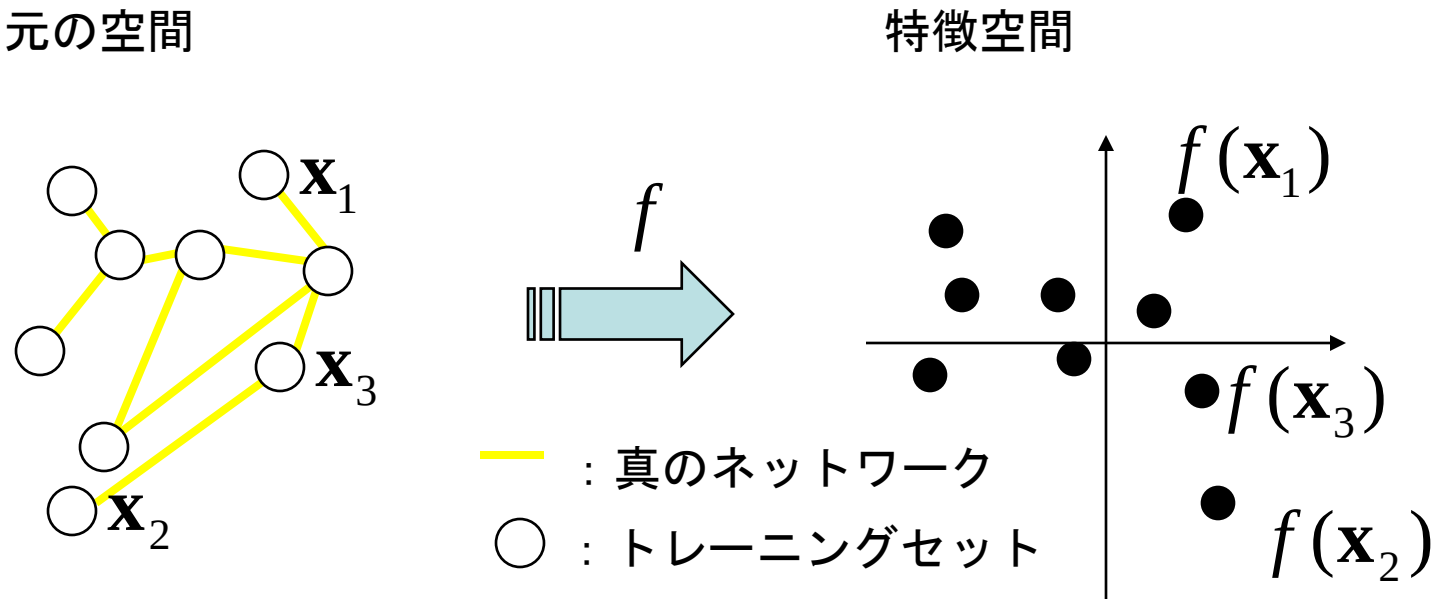


：真のネットワーク

：トレーニングセット

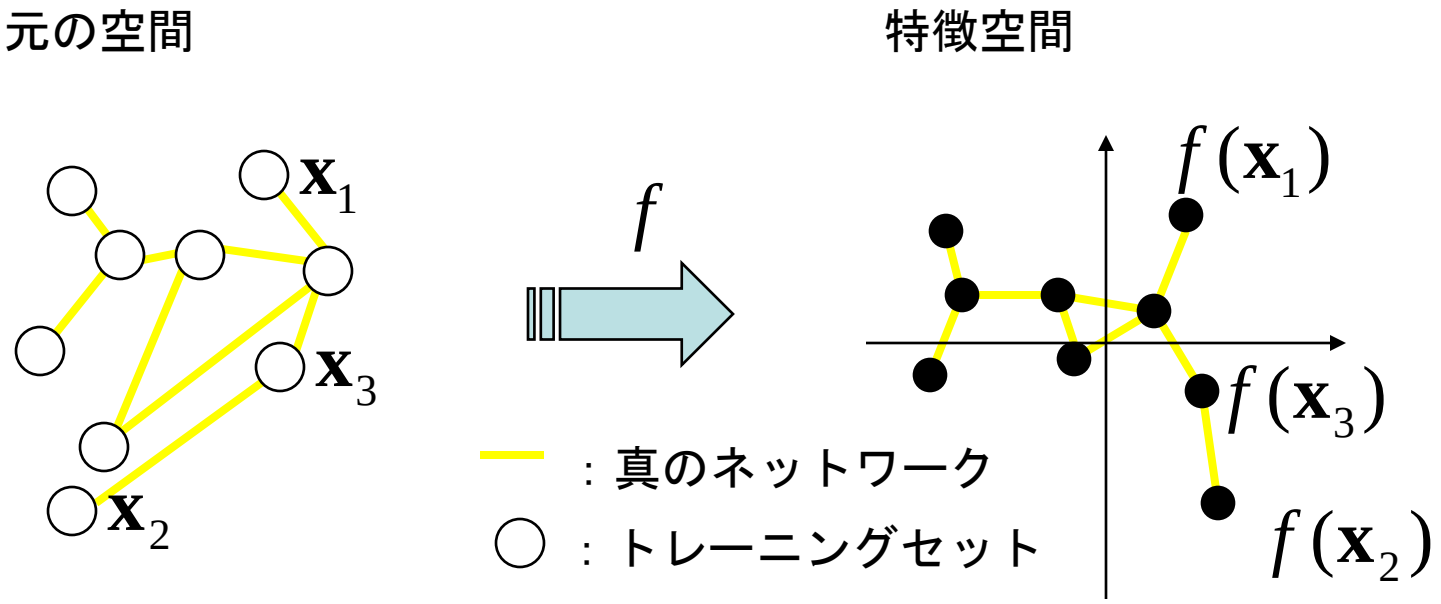
教師付き学習

ステップ１：相互作用するタンパク質ペアが近くにあるような特徴空間に射影

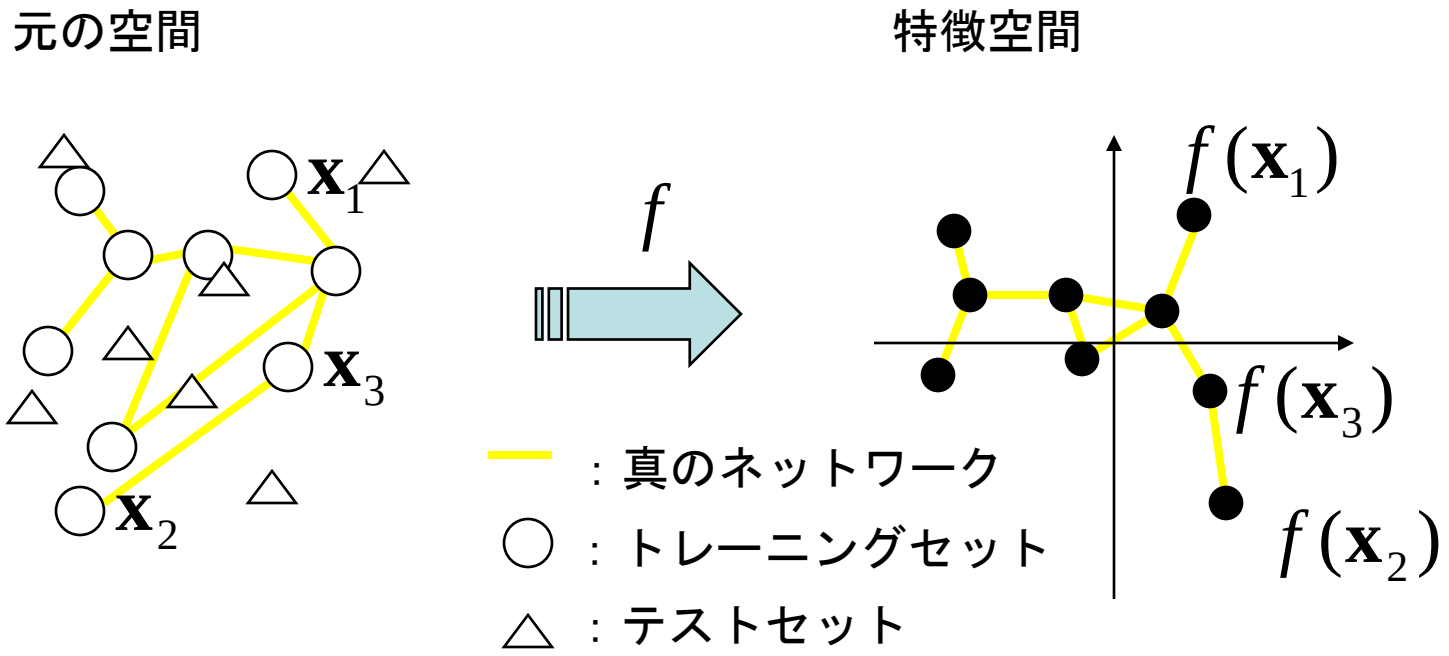


教師付き学習

ステップ１：相互作用するタンパク質ペアが近くにあるような特徴空間に射影

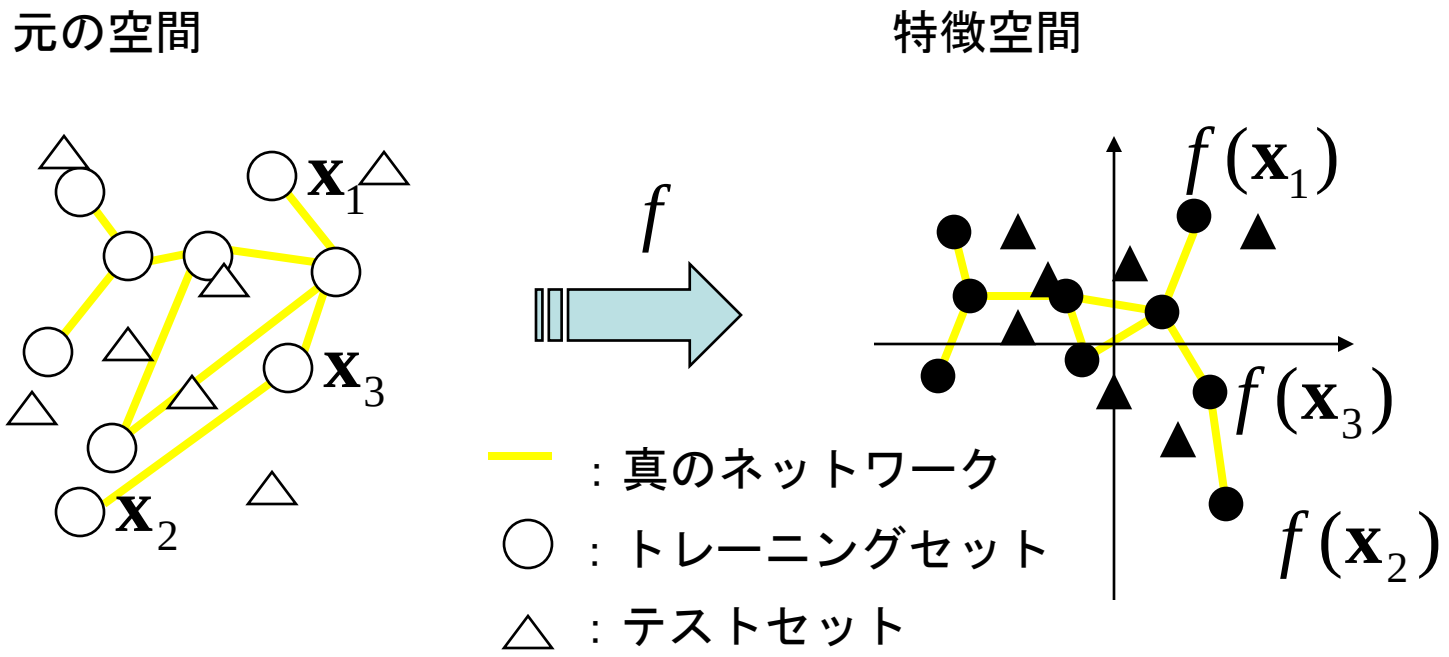


教師付き学習


元の空間
特徴空間
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
f
$f(\mathbf{x}_1)$
$f(\mathbf{x}_3)$
$f(\mathbf{x}_2)$
：真のネットワーク
：トレーニングセット
：テストセット

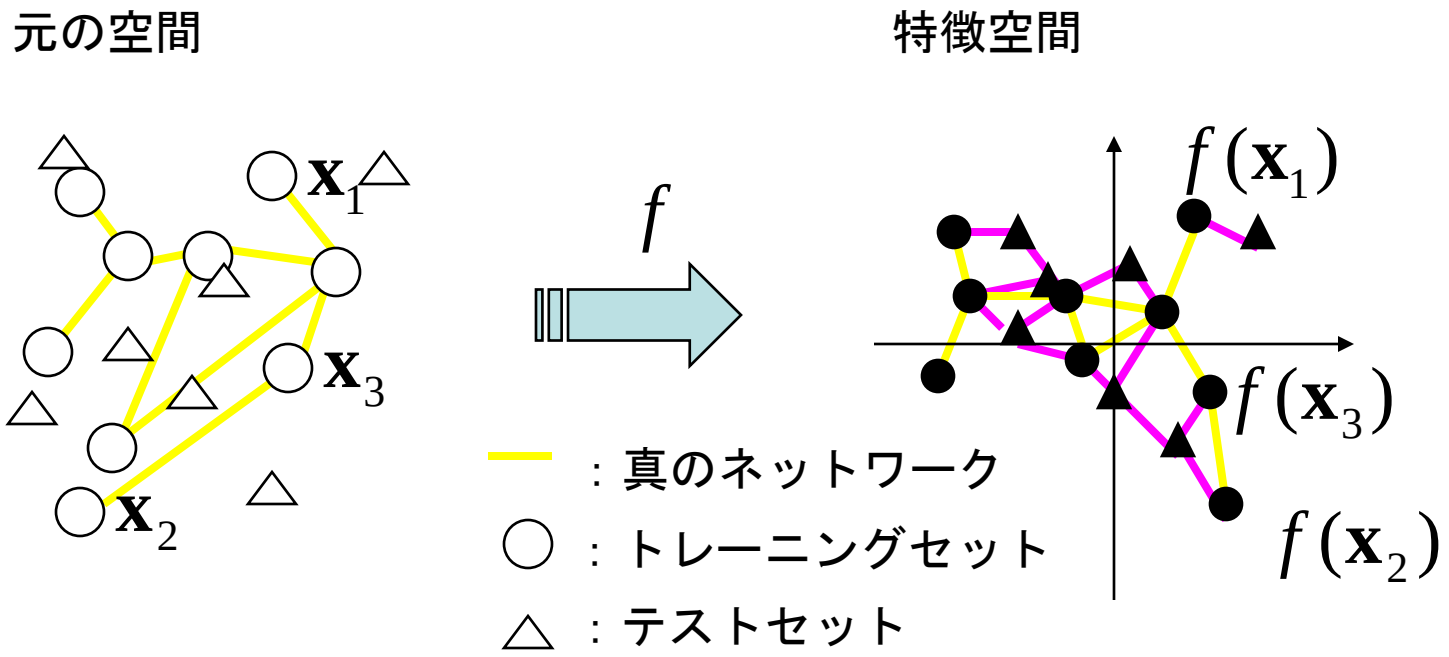
教師付き学習

ステップ２：テストセットに関与するタンパク質間相互作用を予測



教師付き学習

ステップ２：テストセットに関与するタンパク質間相互作用を予測



アルゴリズム

K_1 : 発現データの類似度行列

K_2 : ネットワークの類似度行列

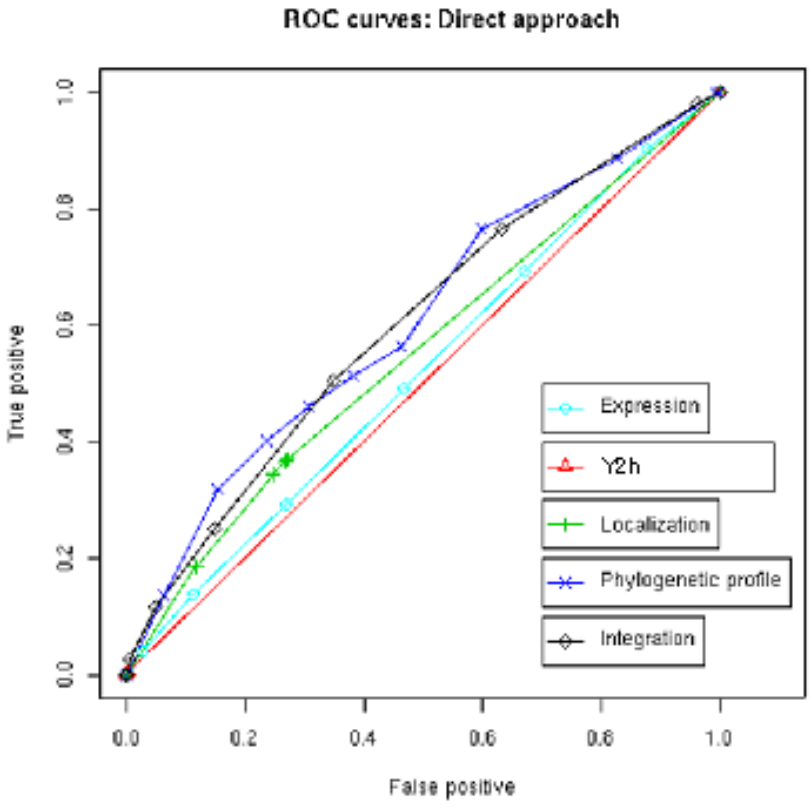
$$(\alpha_1, \alpha_2) = \arg\max \frac{\alpha_1^T K_1 K_2 \alpha_2}{\left(1+\lambda_1 \alpha_1^T K_1^2 \alpha_1\right)^{1/2} \left(1+\lambda_1 \alpha_2^T K_2^2 \alpha_2\right)^{1/2}}$$

$$f(\mathbf{x}) = \sum_{j=1}^{n} \alpha_{1j} K_1(\mathbf{x}_j, \mathbf{x})$$

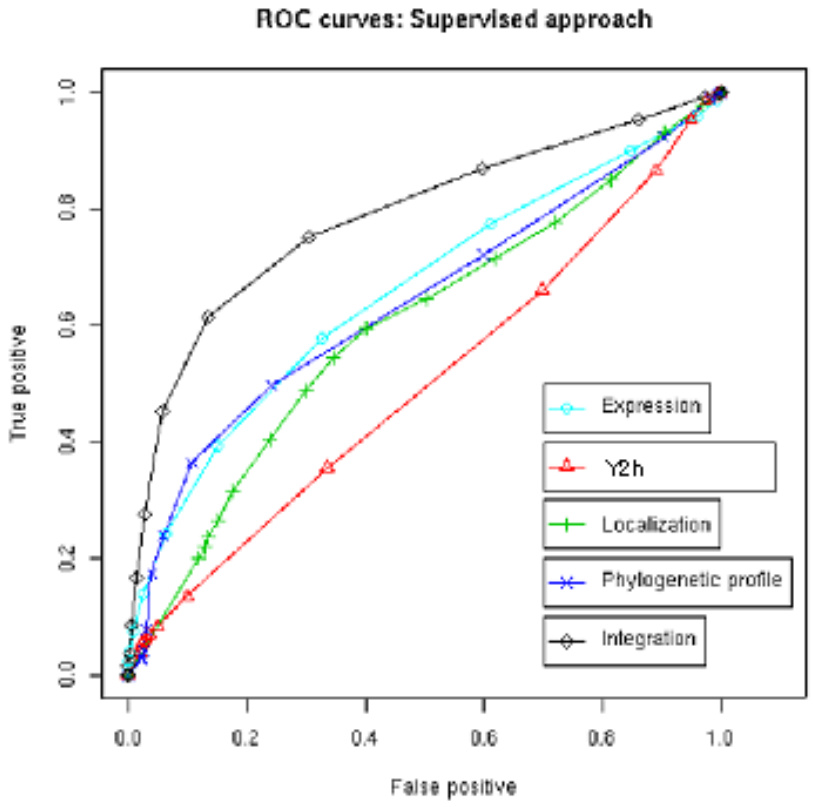
データの類似度行列が入力であることが特長

性能評価

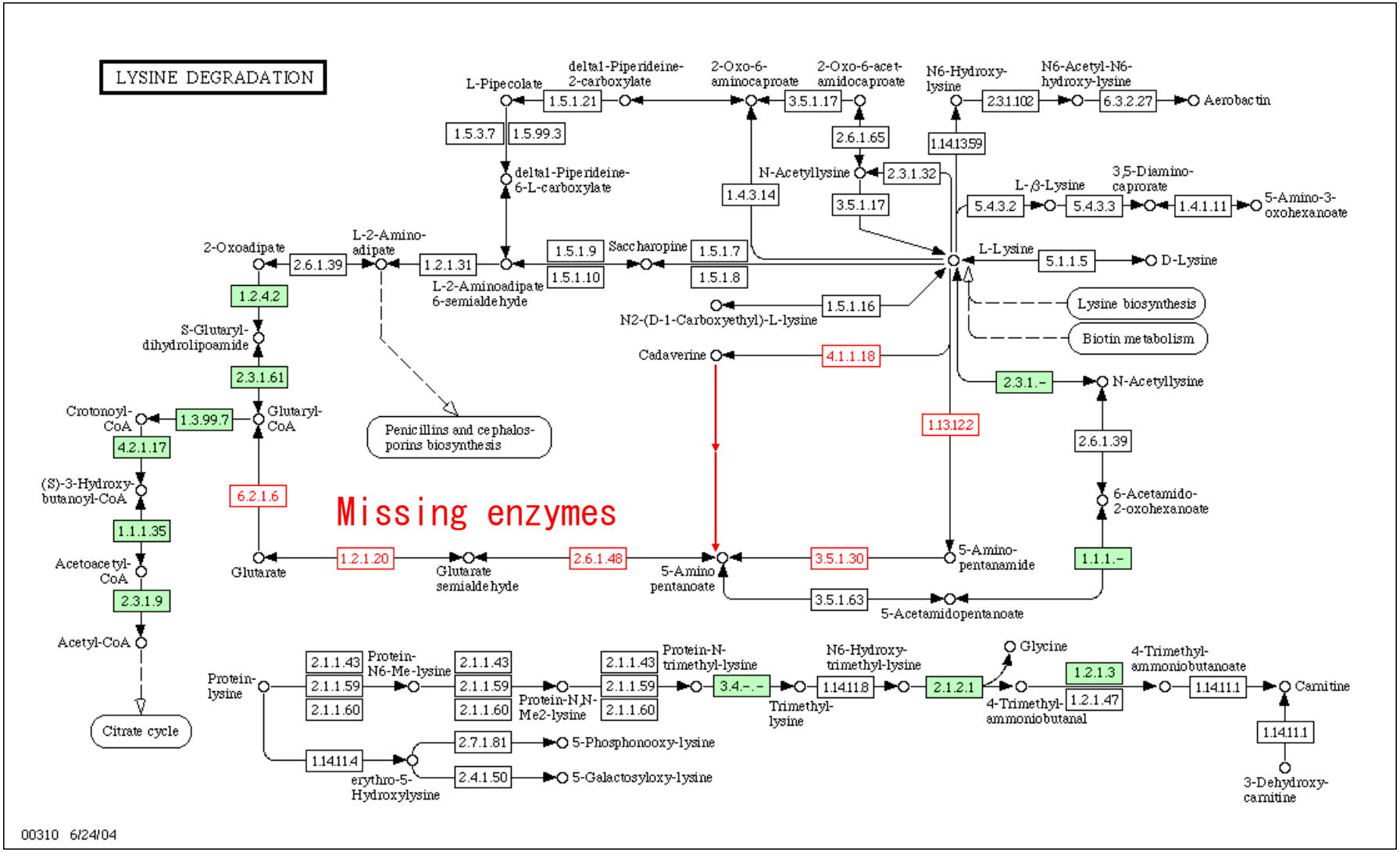
直接予測



教師付き学習



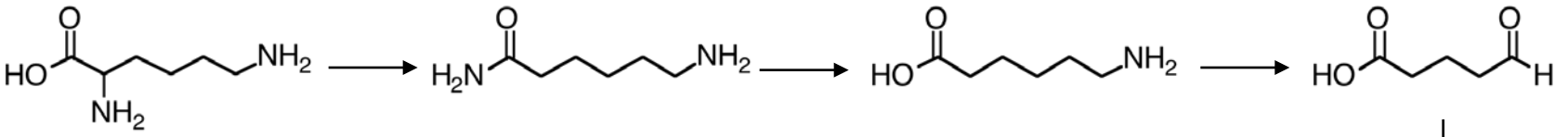
バクテリアの代謝系遺伝子の予測



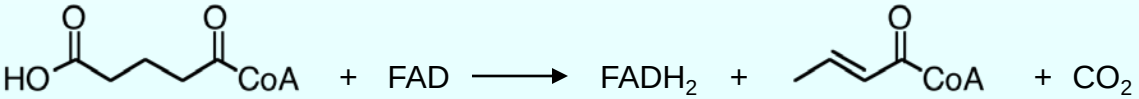
緑膿菌のリジン分解系

バクテリアの代謝系遺伝子の予測

Lysine degradation of *Pseudomonas aeruginosa*



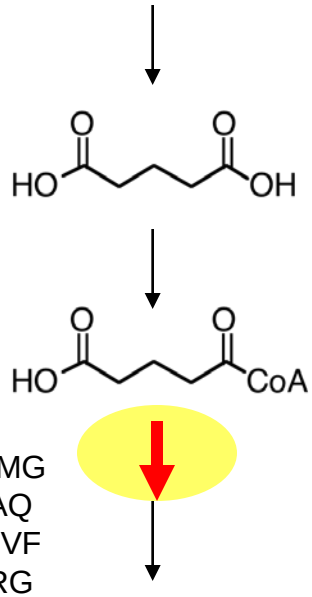
Glutaryl-CoA dehydrogenase



Glutaryl-CoA　　Crotonoyl-CoA

GcdH

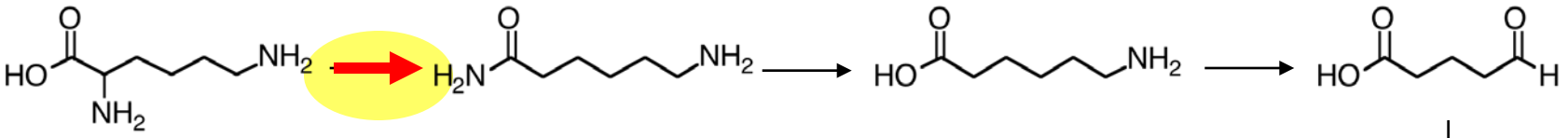
MVGKASFNWIDPLLLDQQLTEEERMVRDSAYQFAQDKLAPRVLEAFRHEQTDPAIFREMG
EVGLLGATIPEQYGGSGLNYVCYGLIAREVERIDSGYRSMMSVQSSLVMVPINEFGTEAQ
KQKYLPKLASGEWIGCFGLTEPNHGSDPGSMITRARKVDGGYRLTGSKMWITNSPIADVF
VVWAKDDAGDIRGFVLEKGWQGLSAPAIHGKVGLRASITGEIVMDNVFVPEENIFPDVRG
LKGPFTCLNSARYGISWGALGAAEACWHTARQYTLDRQQFGRPLAANQLIQKKLADMQTE
ITLALQGCLRLGRMKDEGTAAVEITSIMKRNSCGKALDIARMARDMLGGNGISDEFGVAR
HLVNLEVVNTYEGTHDVHALILGRAQTGIQAFY



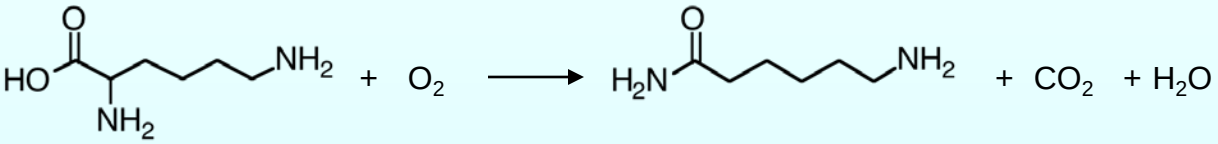
Citrate cycle

バクテリアの代謝系遺伝子の予測

Lysine degradation of *Pseudomonas aeruginosa*



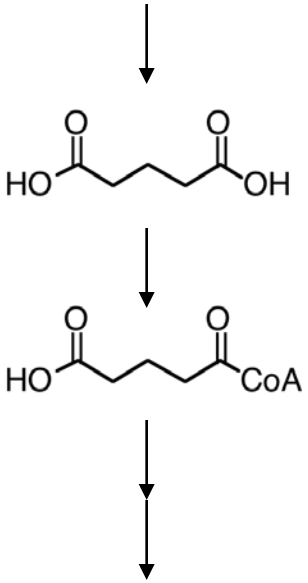
Lysine monooxygenase



L-lysine

5-aminopentanamide

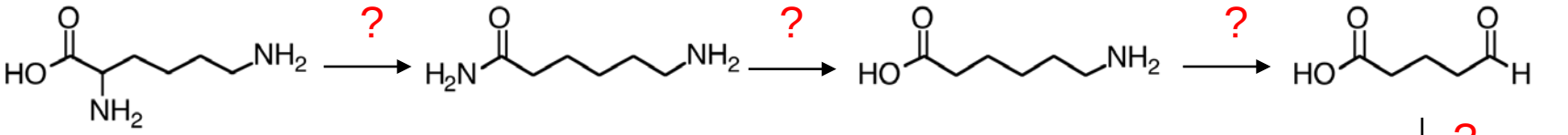
Gene: Unknown



Citrate cycle

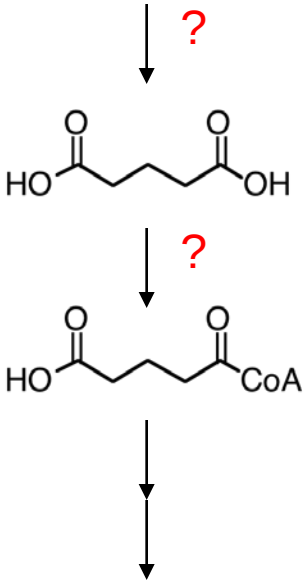
バクテリアの代謝系遺伝子の予測

Lysine degradation *Pseudomonas aeruginosa*



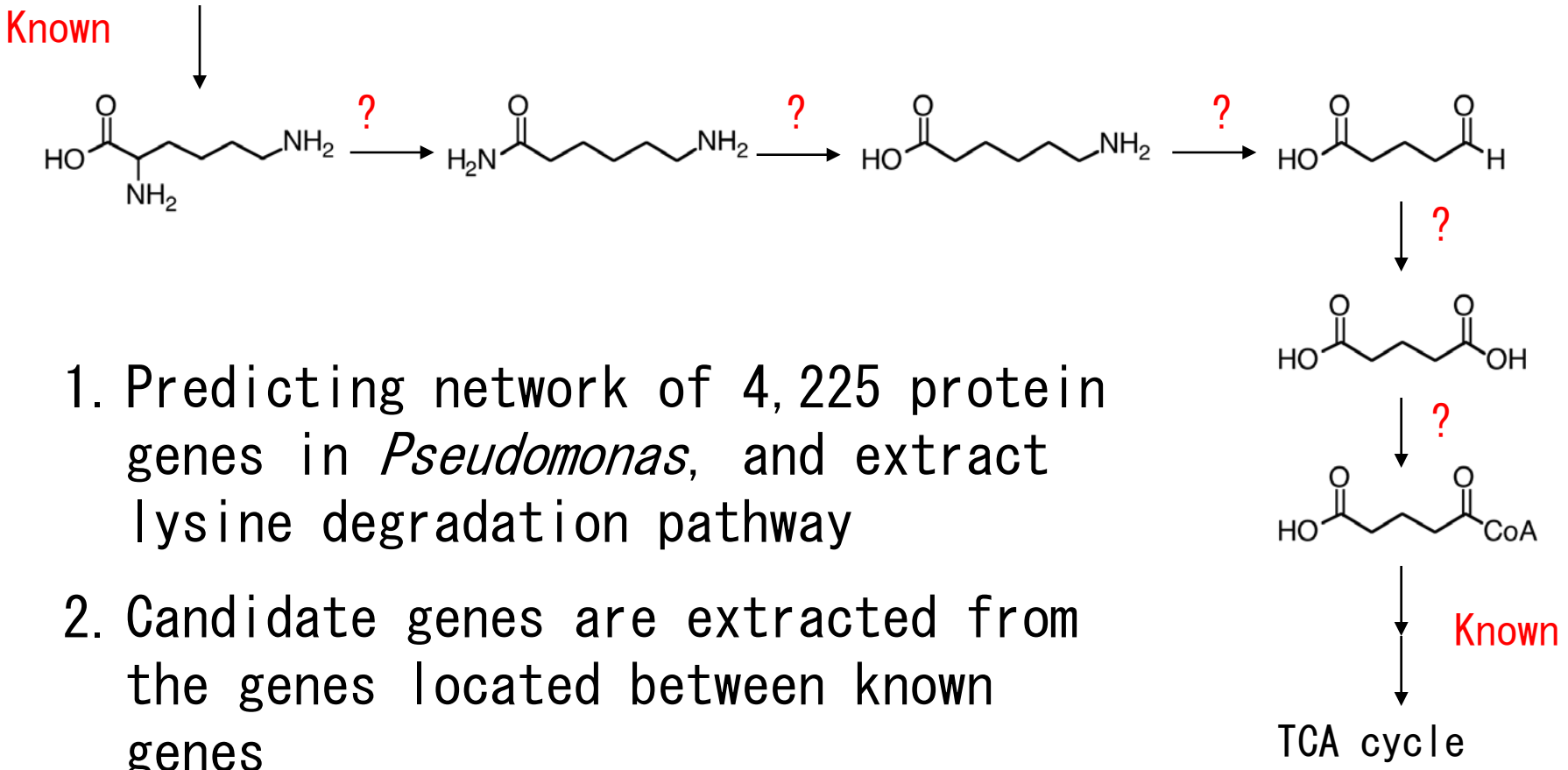
No sequence data for any other species

Gene: Unknown



Citrate cycle

予測方法



1. Predicting network of 4,225 protein genes in *Pseudomonas*, and extract lysine degradation pathway
2. Candidate genes are extracted from the genes located between known genes

予測に使える情報

- 機能的に関連のあるタンパク質の遺伝子は、ゲノム上で近い位置にある傾向
  (Bork, P. et al. , 1998)

- 機能的に関連のあるタンパク質は、同じような進化パターンを持つ傾向
  (Pazos, F., 2001; Pellegrini, M. et al, 1999)

系統プロファイル

- Pellegrini *et al.*
  - *Proc. Natl. Acad. Sci. USA*, 96:4285 (1999)
- オーソログ遺伝子のパターンを分類

E.coli	*S.cerevisiae*	*B.subtilis*	*H.influenzae*
遺伝子1	1	0	1
遺伝子2	1	1	0
遺伝子3	0	1	1
遺伝子4			0
遺伝子5	0	1	1
遺伝子6	1	1	0

同じパターンを持つ遺伝子は
進化的・機能的に関連がある

カーネル（類似度の表現）

・ ゲノム上での位置

$$K_{gen}(\mathbf{x}, \mathbf{x}') = \exp(-d / h)$$

ここで、d：遺伝子 $\mathbf{x}$ と $\mathbf{x}'$ 間の塩基数

・ 系統プロファイル

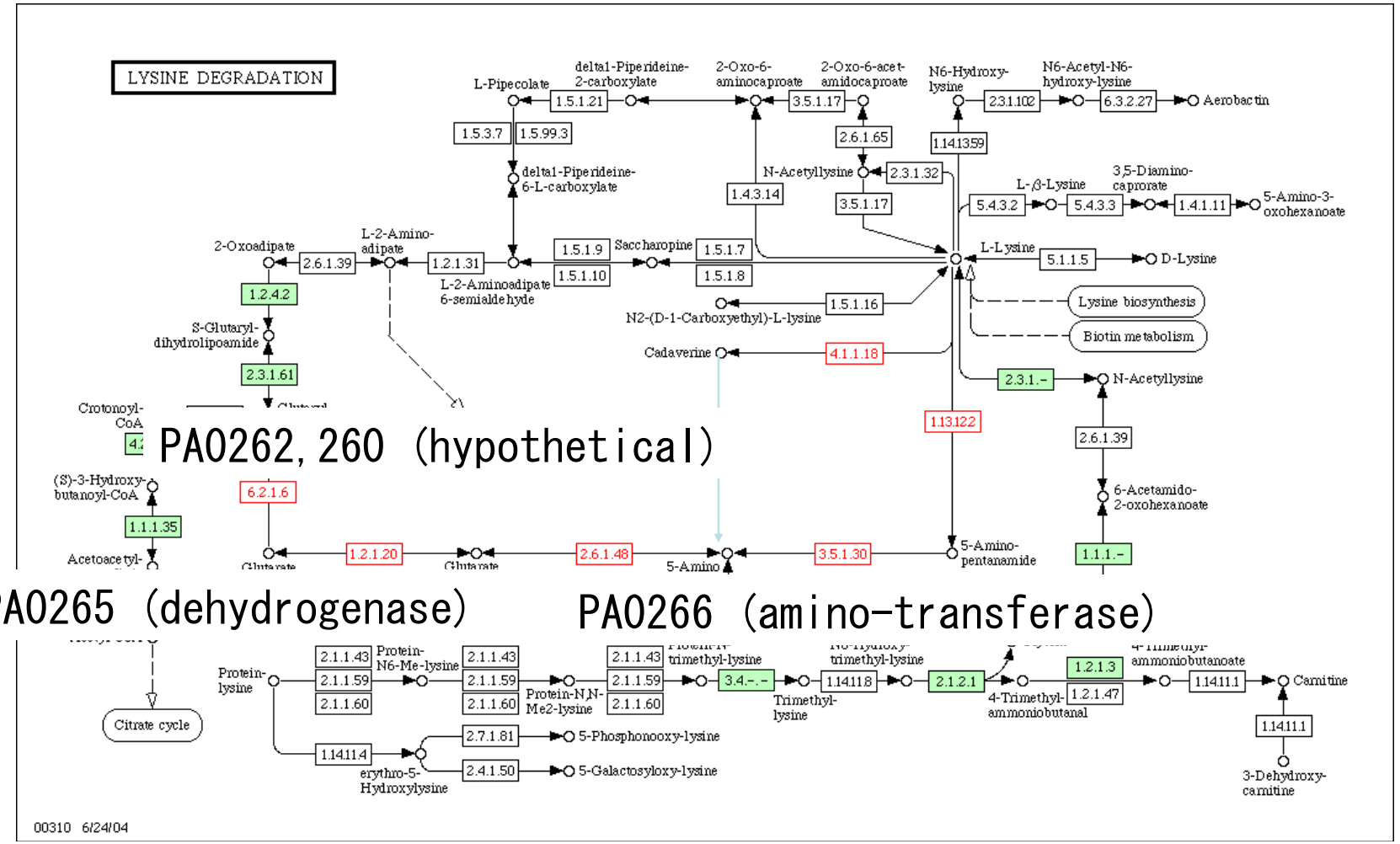
$$K_{phy}(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$

ここで、$\mathbf{x}$：系統プロファイル

・ 統合

$$K_{\text{int}} = K_{\text{gen}} + K_{\text{phy}}$$

予測結果


LYSINE DEGRADATION
L-Pipecolate
1.5.1.21
delta1-Piperideine-2-carboxylate
2-Oxo-6-aminocaproate
3.5.1.17
2-Oxo-6-acet-amidocaproate
N6-Hydroxy-lysine
2.3.1.102
N6-Acetyl-N6-hydroxy-lysine
6.3.2.27
Aerobactin
1.5.3.7
1.5.99.3
2.6.1.65
1.14.13.59
delta1-Piperideine-6-L-carboxylate
N-Acetyllysine
2.3.1.32
1.4.3.14
3.5.1.17
L-β-Lysine
5.4.3.2
3,5-Diamino-caprorate
5.4.3.3
1.4.1.11
5-Amino-3-oxohexanoate
2-Oxoadipate
2.6.1.39
L-2-Amino-adipate
1.2.1.31
L-2-Aminoadipate 6-semialdehyde
1.5.1.9
1.5.1.10
Saccharopine
1.5.1.7
1.5.1.8
L-Lysine
5.1.1.5
D-Lysine
1.2.4.2
S-Glutaryl-dihydrolipoamide
2.3.1.61
1.5.1.16
N2-(D-1-Carboxyethyl)-L-lysine
Lysine biosynthesis
Biotin metabolism
Cadaverine
4.1.1.18
2.3.1.-
N-Acetyllysine
Crotonoyl-CoA
1.13.12.2
2.6.1.39
6-Acetamido-2-oxohexanoate
(S)-3-Hydroxy-butanoyl-CoA
6.2.1.6
1.1.1.35
Acetoacetyl-
1.2.1.20
2.6.1.48
3.5.1.30
5-Amino-pentanamide
1.1.1.-
PA0262,260 (hypothetical)
PA0265 (dehydrogenase)
PA0266 (amino-transferase)
Citrate cycle
Protein-lysine
2.1.1.43
2.1.1.59
2.1.1.60
Protein-N6-Me-lysine
Protein-N,N-Me2-lysine
Protein-N,N,N-trimethyl-lysine
3.4.-.-
Trimethyl-lysine
1.14.11.8
2.1.2.1
4-Trimethyl-ammoniobutanal
1.2.1.47
1.2.1.3
4-Trimethyl-ammoniobutanoate
1.14.11.1
Carnitine
1.14.11.14
3-Dehydroxy-carnitine
1.14.11.4
erythro-5-Hydroxylysine
2.7.1.81
5-Phosphonooxy-lysine
2.4.1.50
5-Galactosyloxy-lysine
00310 6/24/04

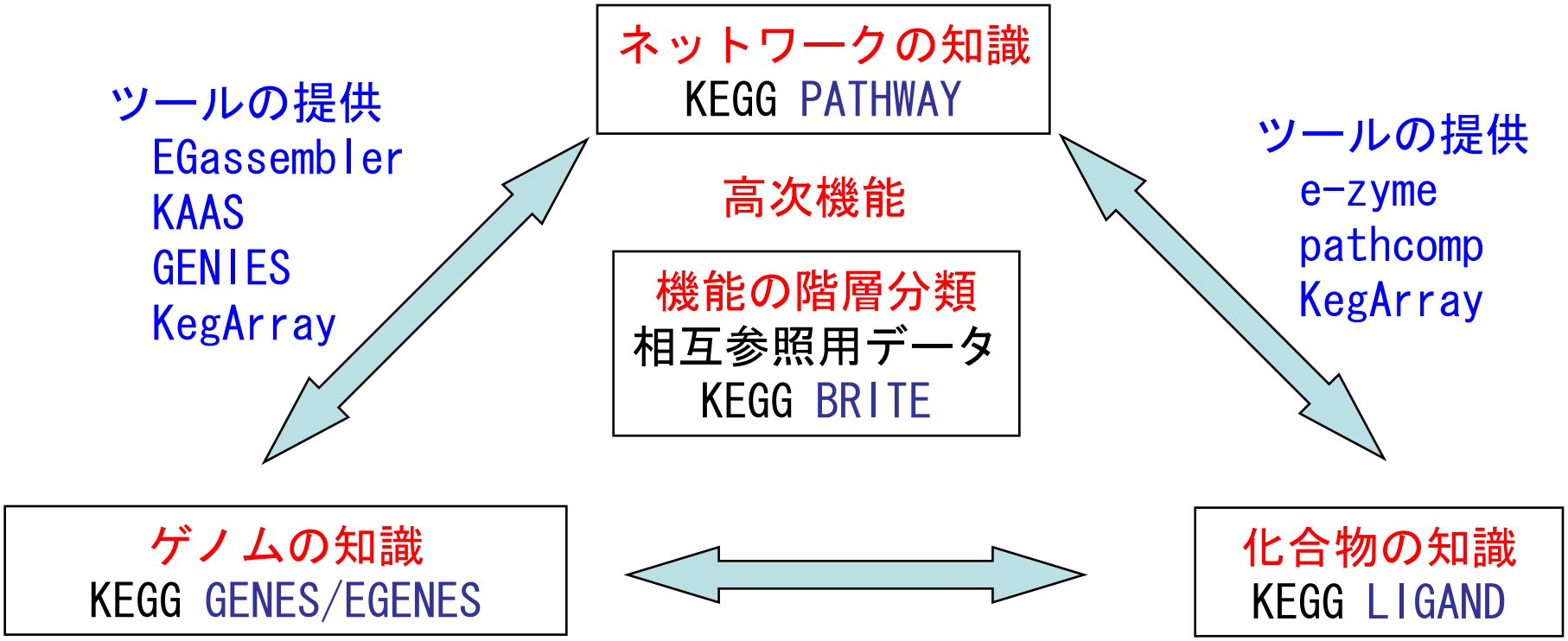
まとめ

- ゲノムをはじめとする多種多様なデータから生命システムを明らかにするにはデータを検索・解析技術が重要

- ハイスループットデータをデータベース化するだけでなく、既知の情報もデータベース化して、新しいデータと組み合わせて解析できるようにすることが重要

京都大学で構築中のデータベース

様々な種類のデータを「生命現象の総体」として立体的に再構築



研究者の知識をゲノムレベルのデータと結びつける

KEGG: Kyoto Encyclopedia of Gene and Genomes

http://www.genome.jp/kegg/

ご静聴ありがとうございました。