

# ゲノム研究におけるオントロジー の開発と利用

東大 情報理工 高井貴子

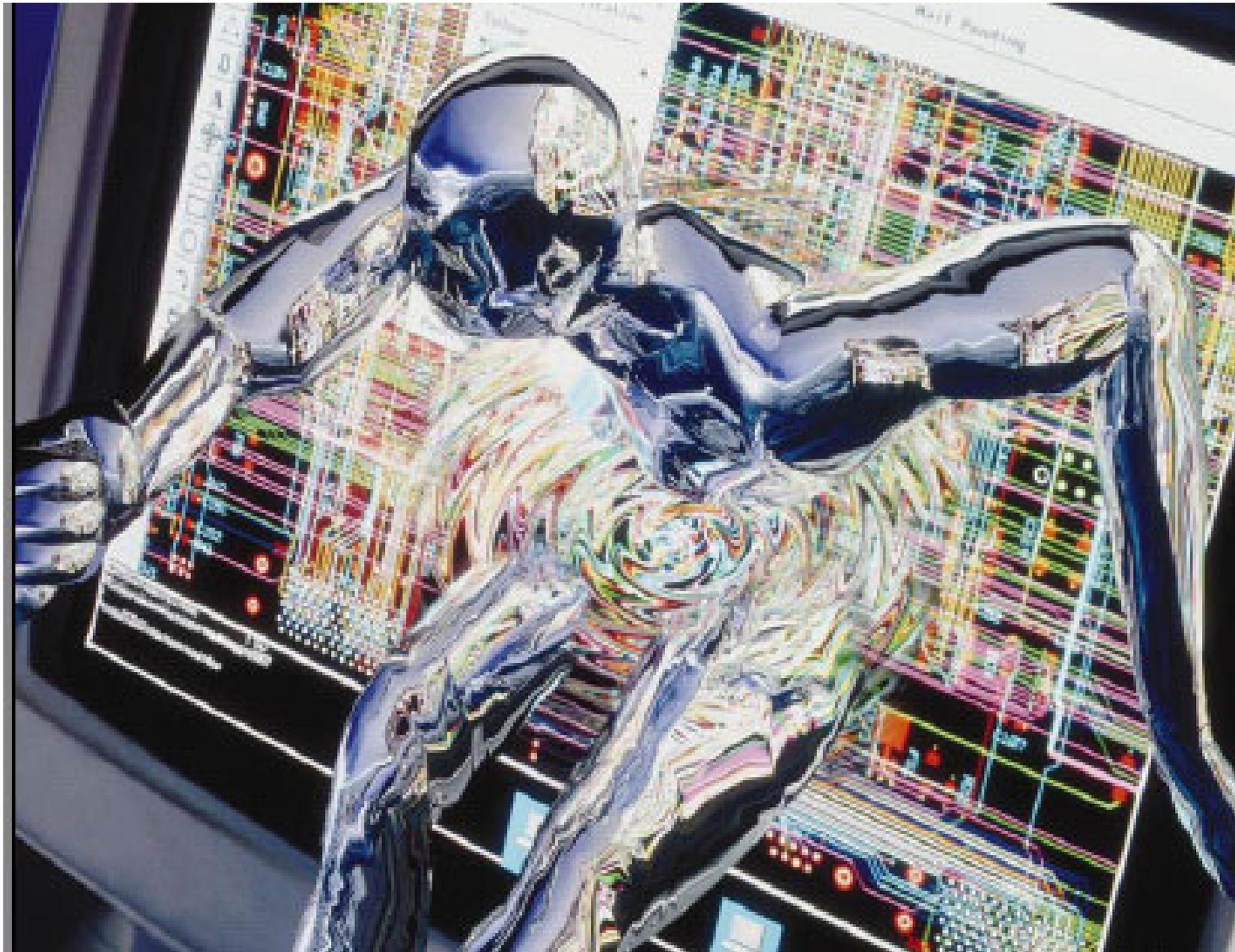
# ゲノム研究におけるオントロジーの開発と利用

- ❖ 生物学およびゲノム研究
- ❖ 計算機科学におけるオントロジー
- ❖ ゲノム研究とオントロジー

# ゲノム研究におけるオントロジーの開発と利用

- ❖ 生物学およびゲノム研究
- ❖ 計算機科学におけるオントロジー
- ❖ ゲノム研究とオントロジー

# Conceptual Biology



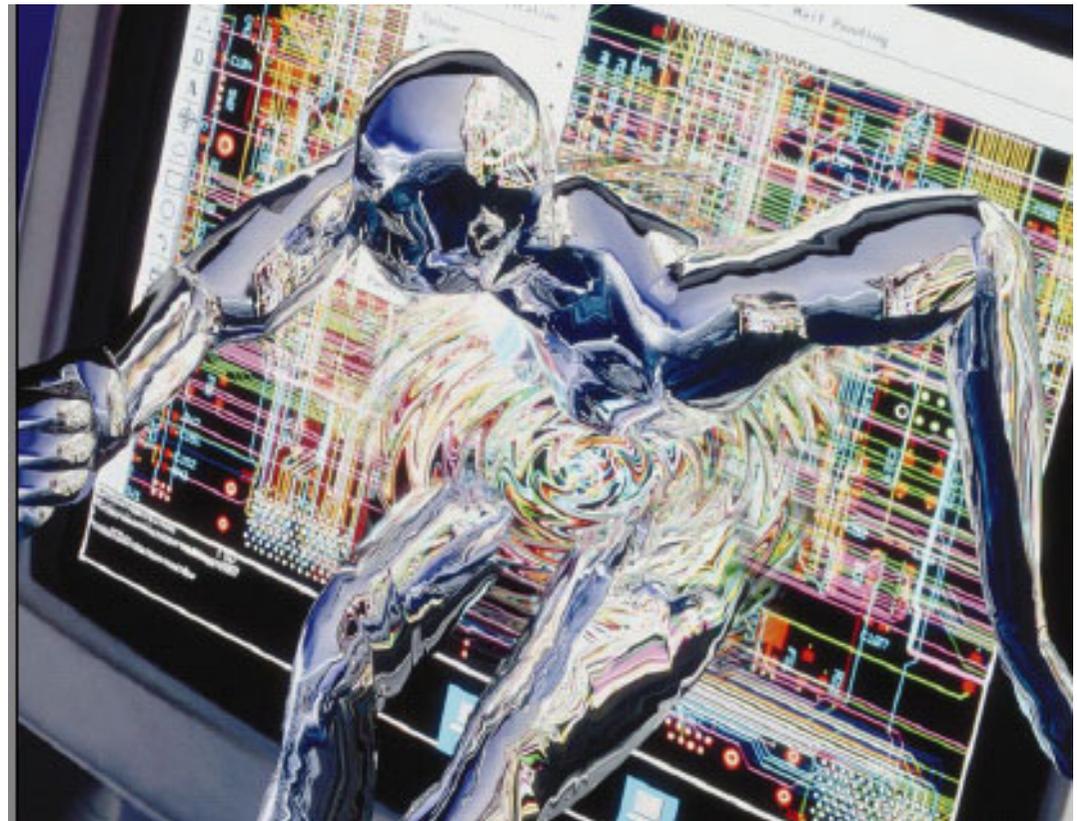
Nature, 416, 373 (2002)

# ゲノム研究における知識処理研究

## □ Conceptual Biology

- 概念の生成から概念の解析へ
- 既存の多量の概念を解析し、仮説を立て、それを実証できる能力が必要

Nature, 416, 373 (2002)



# 事実知識の獲得



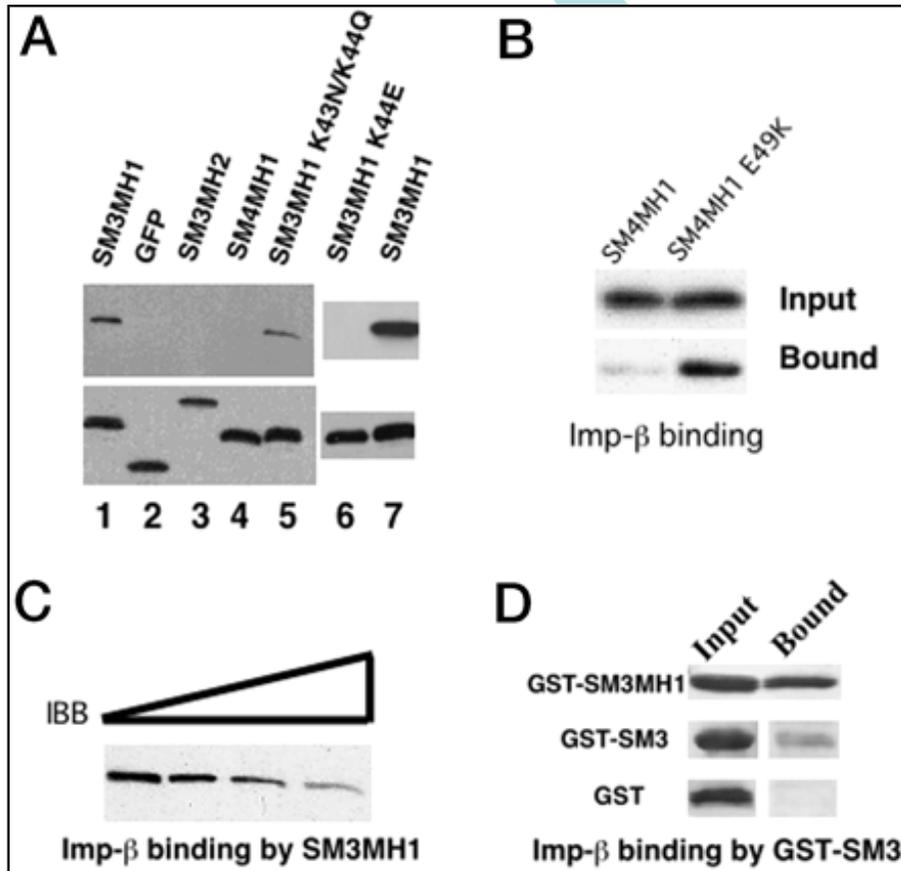
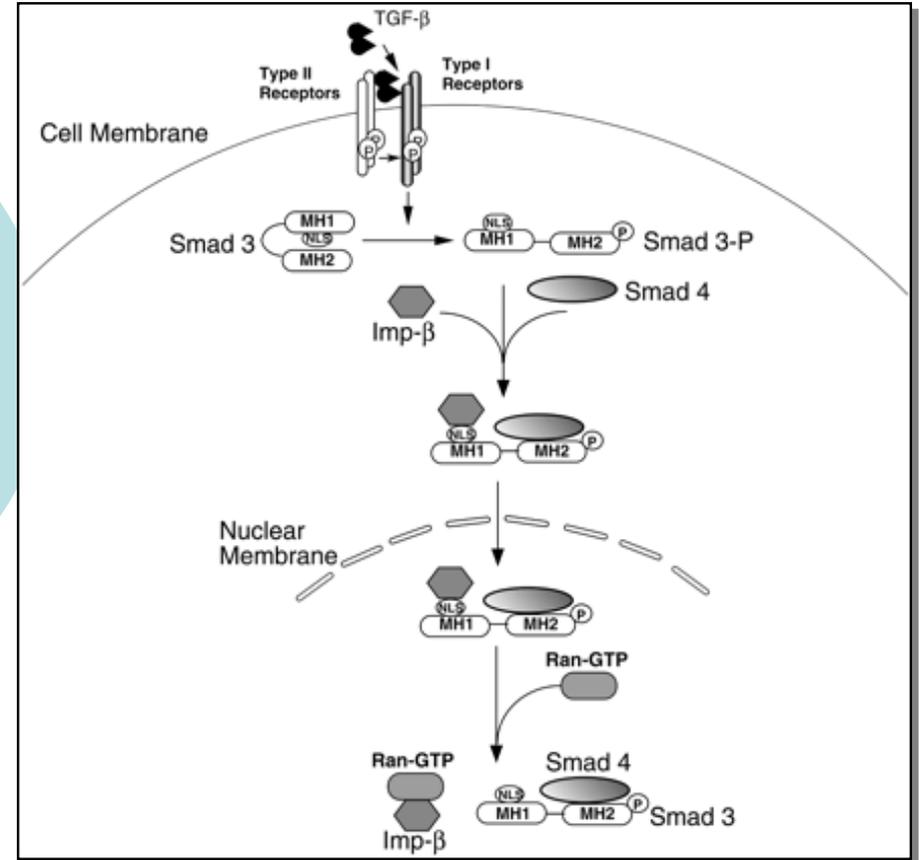
概念の獲得

核移行

実験データの解釈

生物学

実験の考案、実験手法の開発、  
実験の施行



解釈!



実験データ

# ゲノム研究がもたらしたパラダイムシフト

---

- データの解釈、概念の獲得、概念の利用は、専門家の頭の中で行われてきた
- ゲノム研究の登場により、概念が爆発的に増加し、人間の頭だけでは追いつかなくなりつつある
- 概念の処理にも、計算機の助けが必要

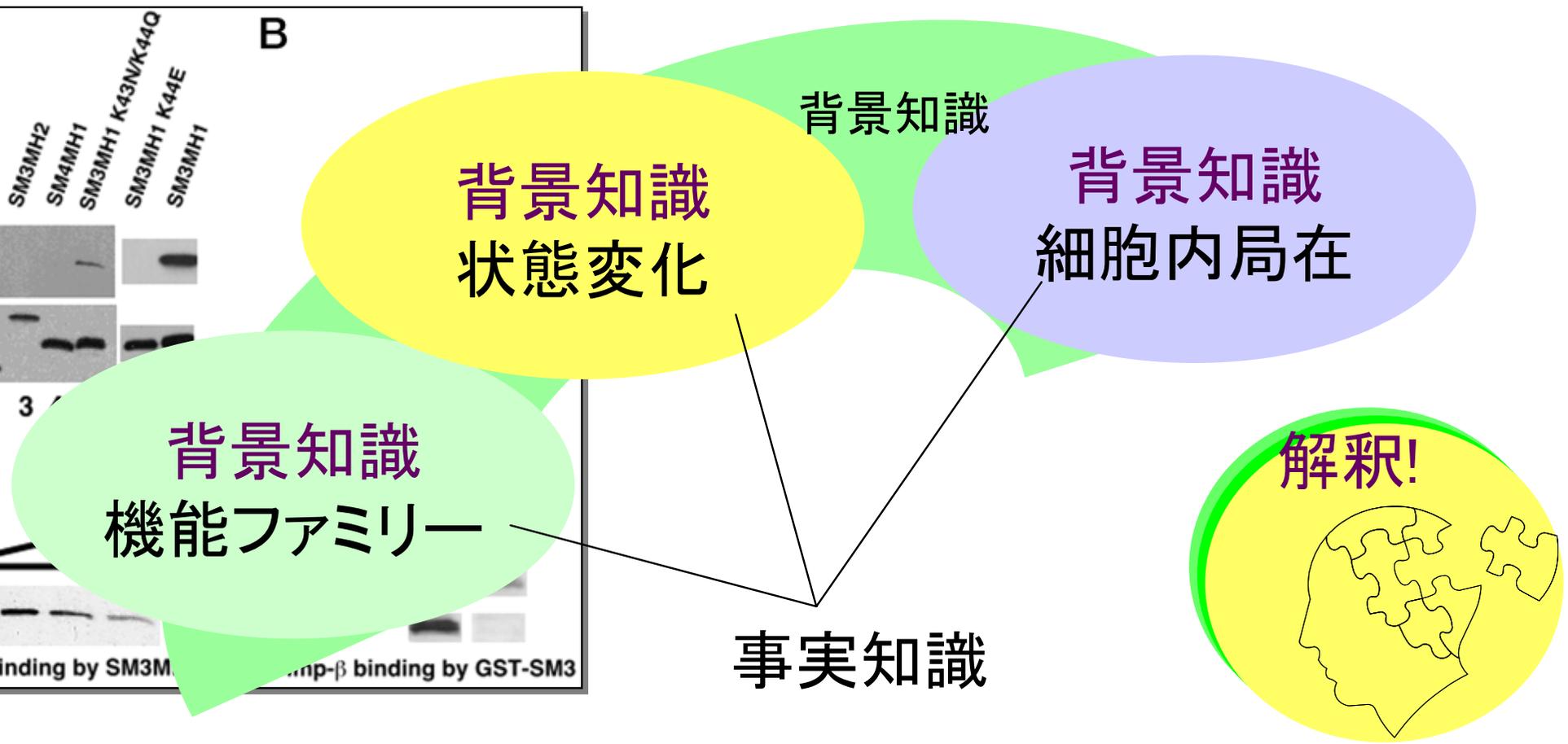
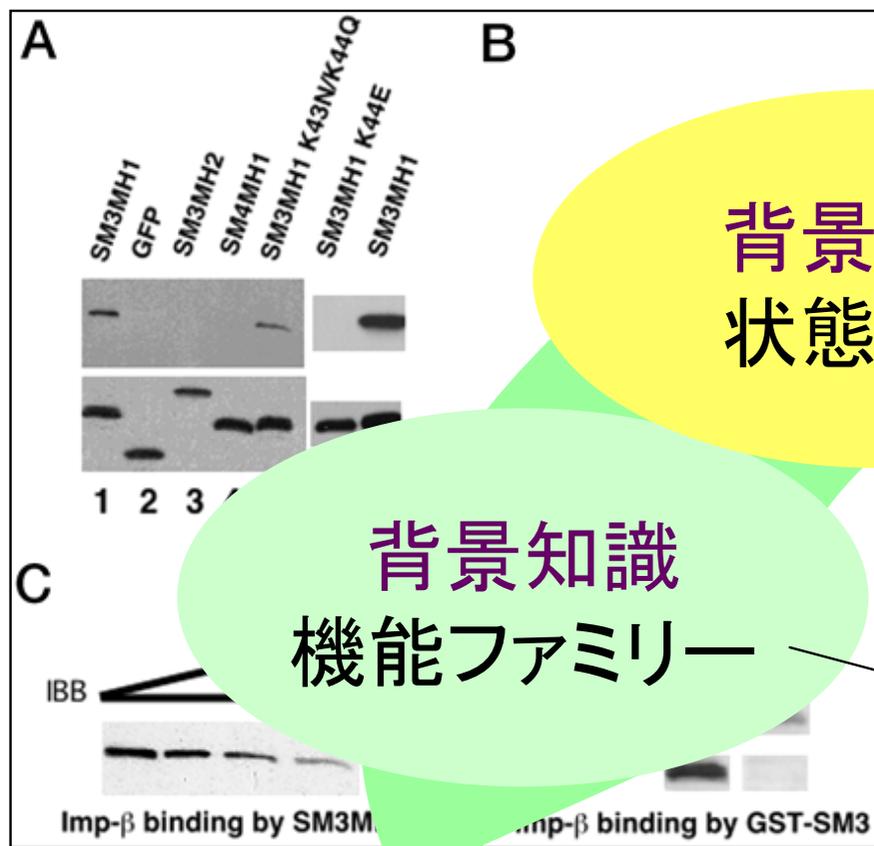
## ゲノム研究

- 遺伝子(遺伝子産物)の全セットから生命現象を理解
- 細胞内の全遺伝子を考察の対象とし、全分子の協調作用として生命現象を解析する

# データの解釈とは？

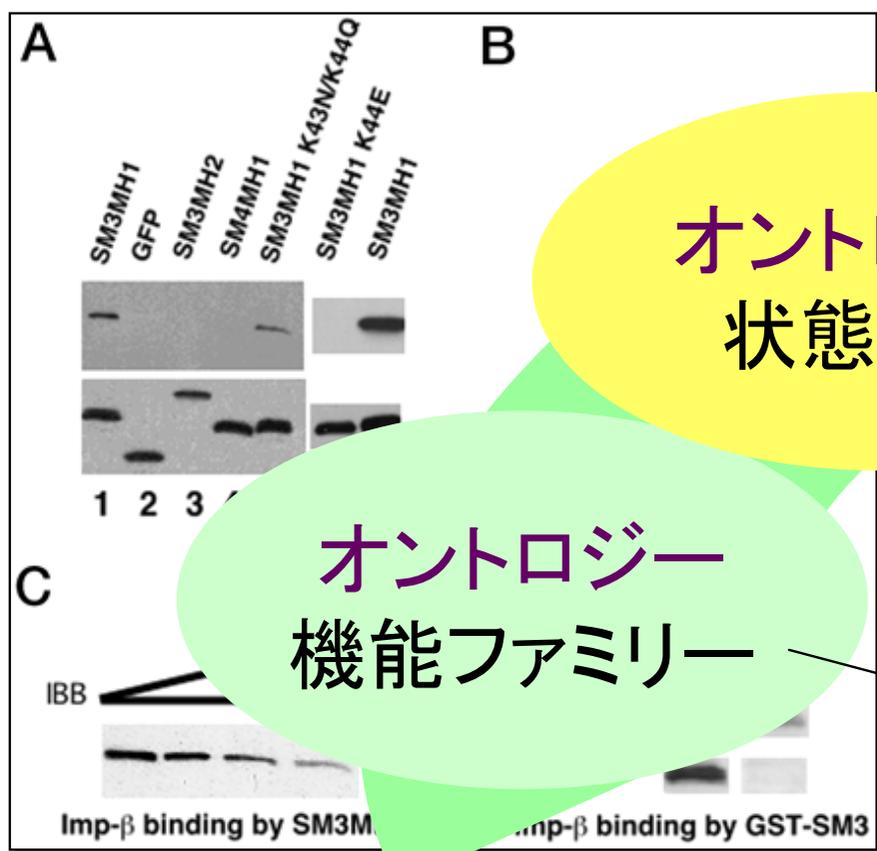
- 解釈: データを背景知識に照合して、事実知識を獲得する過程

Transportation of Smad3 to nucleus = Smad3 [機能ファミリー] + transportation [状態変化] + nucleus [細胞内局在]



# ゲノム研究は計算機が解釈を行うことを要求する

- **オントロジー**: 背景知識を計算機と人間が可読なように形式化したもの



# ゲノム研究におけるオントロジーの開発と利用

- ❖ 生物学およびゲノム研究
- ❖ 計算機科学におけるオントロジー
- ❖ ゲノム研究とオントロジー

# オントロジー Ontology

## ❖ 哲学の定義

- ❖ 存在に関する体系的な理論(存在論)
- ❖ アリストテレスが祖

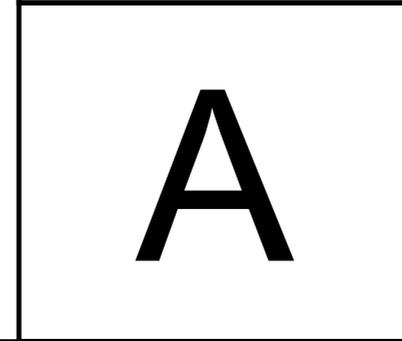
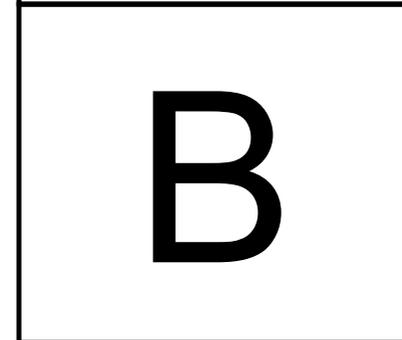
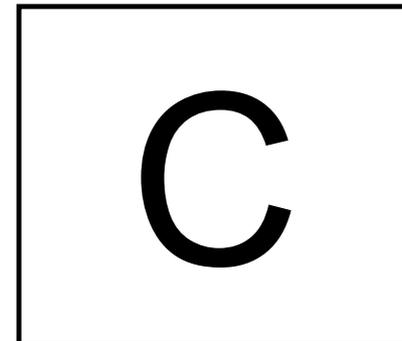
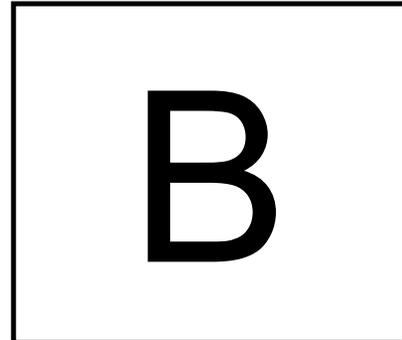
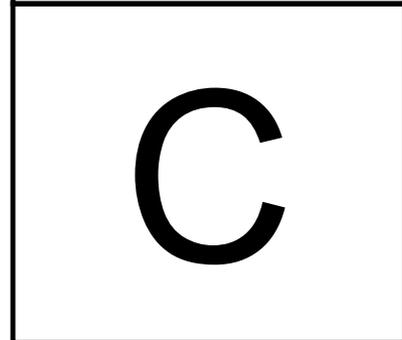
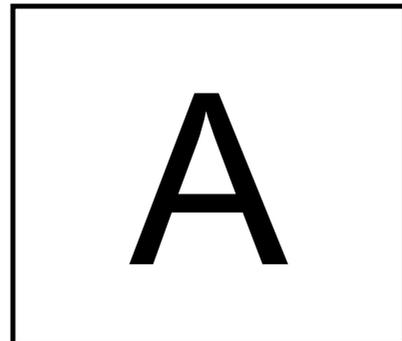
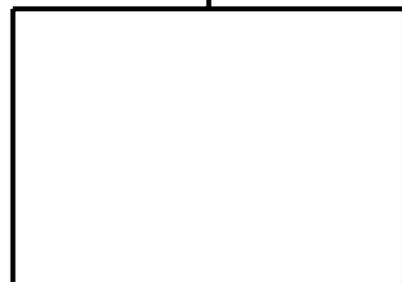
# オントロジー Ontology

## ❖ 人工知能の定義

- ❖ 概念化の明示的な記述 (Gruber, 92)
- ❖ 概念化: 対象とする世界に存在する概念とそれらの間に成立する関係を言う
- ❖ モデル構築者が対象とする世界をどのように「眺めたか」、言い換えるとその世界には「何が存在しているか」を明示化したもの
- ❖ 暗黙的な背景知識を人間と計算機に可読としたもの
- ❖ それを基に対象とする世界を計算機上に再構築できるもの

# ブロックワールド

ロボットハンドを使って初期状態の配置を  
ゴール状態の配置に積み替えること。そ  
の他の一切の作業も物体も存在しない。



初期状態

ゴール状態

## 概念化1

### オブジェクト 関係

block A	on(X,Y)
block B	above(X,Y)
block C	clear(X)
table A	holding(X)
hand A	handEmpty

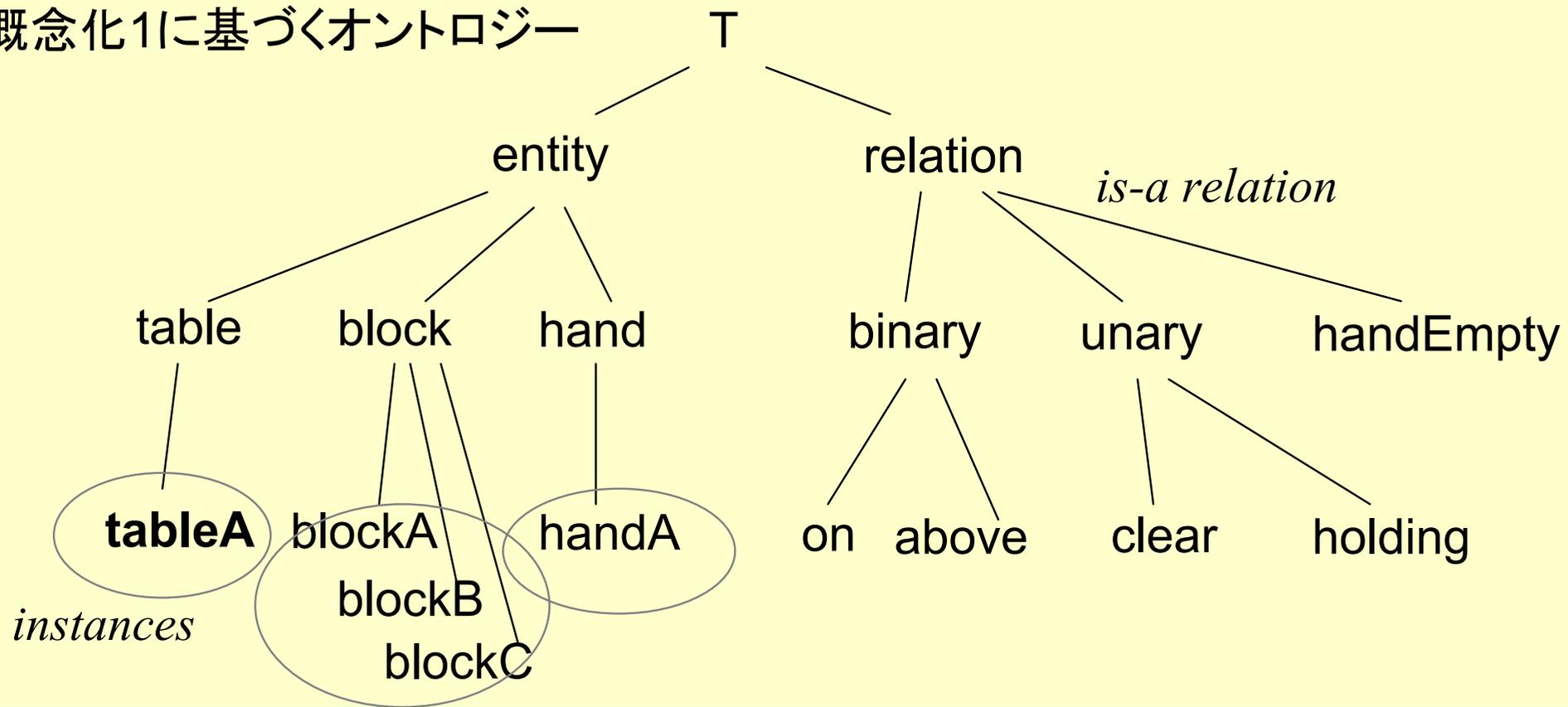
## 概念化2

### オブジェクト 関係

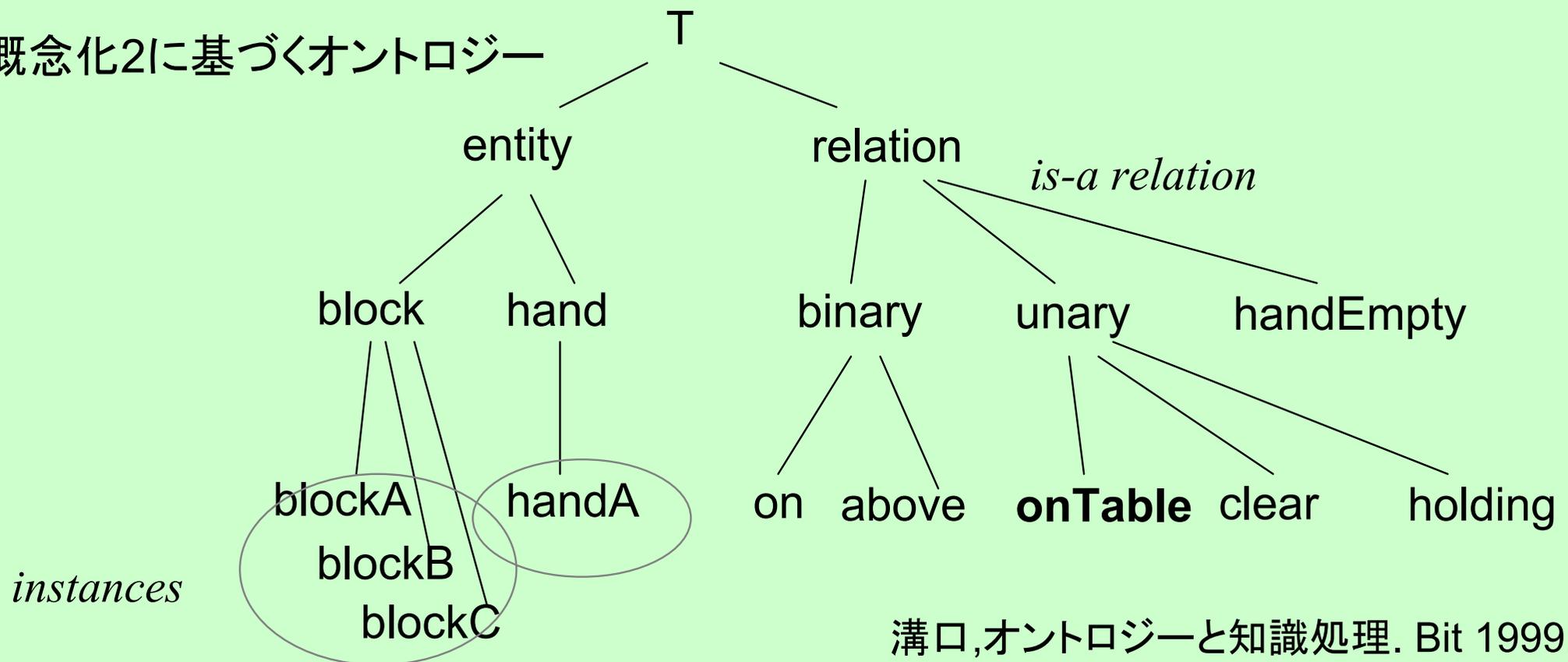
block A	on(X,Y)
block B	above(X,Y)
block C	onTable(X)
hand A	clear(X)
	holding(X)
	handEmpty

テーブルAの存在をオブジェクトとして認識するか、関係として認識するか？

# 概念化1に基づくオントロジー



# 概念化2に基づくオントロジー



# なぜオントロジーが必要か

- ❖ ブロックワールドを実行するプログラムを他人が理解できるか？
  - ❖ テーブルAをオブジェクトとして概念化せず、関係として概念化した
- ❖ モデル構築者が対象とする世界をどのように「眺めたか」、その世界には「何が存在しているか」と捉えたか
- ❖ これをオントロジーは明示化する

# オントロジーで使われる関係

## ❖ is-a関係

- ❖ 概念間の特殊汎化関係

- ❖ 例: 齧歯類 is-a ほ乳類, ほ乳類 is-a 脊椎動物

## ❖ part-of関係

- ❖ 概念間の部分全体関係

- ❖ 例: 爪 part-of 指, 指 part-of 手

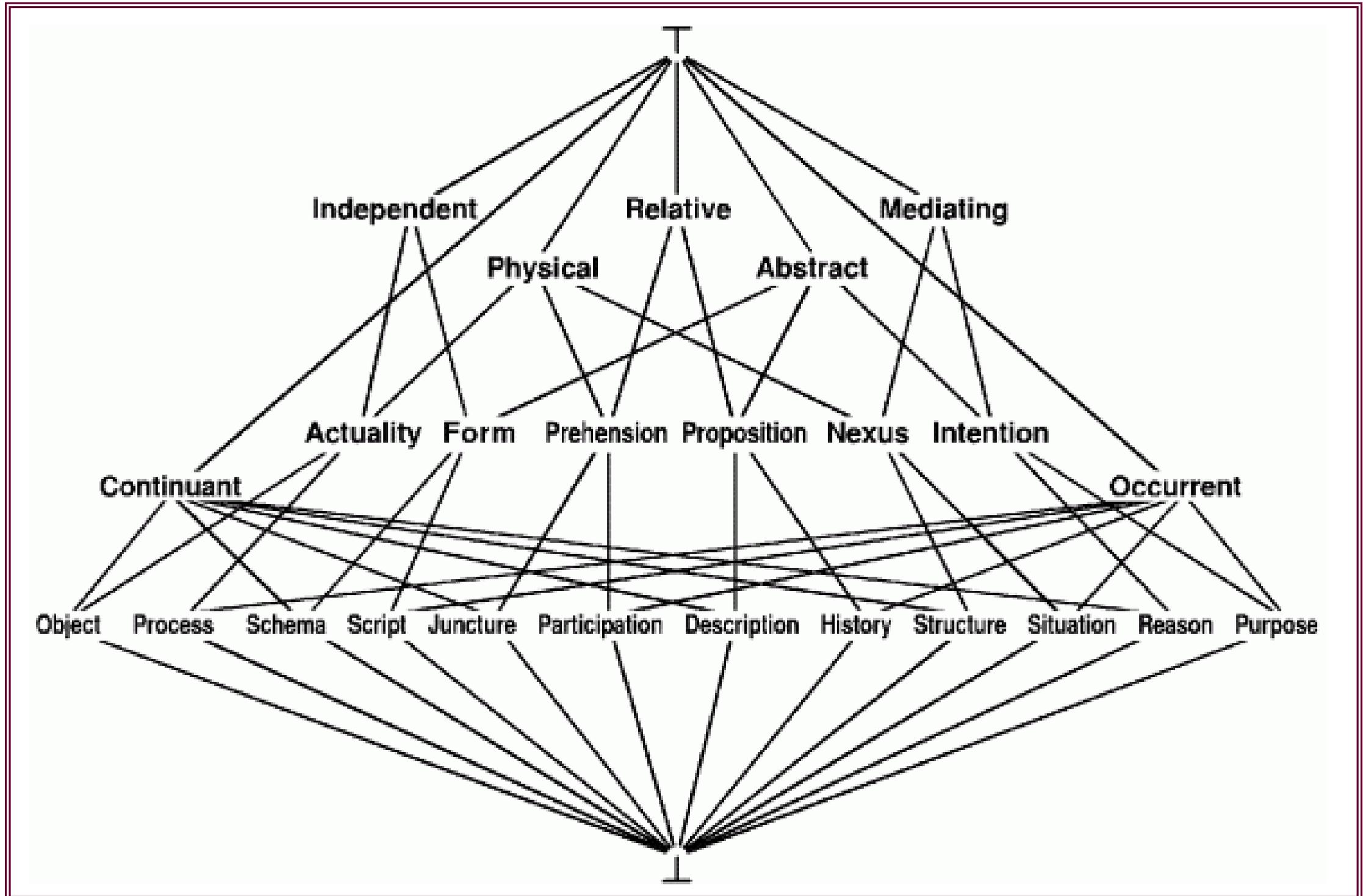
## ❖ instance-of関係

- ❖ 概念とインスタンスの間関係(インスタンス化)

# オントロジーの種類

- ❖ ドメインオントロジー
  - ❖ 対象とする世界がこの世の一部
- ❖ トップレベルオントロジー
  - ❖ 対象とする世界がこの世全体

# Sowaのトップレベルオントロジー



## 存在一般の体系(哲学オントロジー)

# ドメインオントロジーの例

- 乗り物
  - 陸上車
    - 車
      - 4輪車
        - 乗用車
        - トラック
      - オートバイ
    - 自動車
  - 船
  - 飛行機

- 乗り物(世界)
  - 種類
    - 陸上車
    - 船
    - 飛行機
  - 機能
    - 人を運ぶ
    - 荷物を運ぶ
  - 属性
    - 馬力
    - 大きさ
  - 構成物
    - エンジン
    - 車体
  - 交通システム

.....

# オントロジーに類似したもの

- ❖ 語彙集合（辞書等） Vocabulary
  - ❖ 処理の対象とする用語の集合
  - ❖ 語彙（用語）は概念を表現するための「ラベル」
- ❖ 概念階層 タキソノミー
  - ❖ 概念の「種類」という性質のみに注目して、階層的に概念を分類したものという
- ❖ すべてオントロジーに含めることも多い

- 乗り物
  - 陸上車
    - 車
      - 4輪車
        - 乗用車
        - トラック
      - オートバイ
    - 汽車
  - 船
  - 飛行機

## タキソノミー

- 乗り物(世界)
  - 種類
    - 陸上車
    - 船
    - 飛行機
  - 機能
    - 人を運ぶ
    - 荷物を運ぶ
  - 属性
    - 馬力
    - 大きさ
  - 構成物
    - エンジン
    - 車体
  - 交通システム
  - .....

## オントロジー



Search for  as  complete name  lock

Display  levels using filter:

Nucleotide  Protein  Structure  Genome  Popset  SNP  3D Domains  
 Domains  GEO Datasets  GEO Expressions  UniGene  UniSTS  PubMed Central  Gene  
 HomoloGene  MapView  LinkOut  BLAST  TRACE

**Lineage** (full): [root](#); [cellular organisms](#); [Eukaryota](#); [Fungi/Metazoa group](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Coelomata](#); [Deuterostomia](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Tetrapoda](#); [Amniota](#)

◦ **Mammalia** (mammals) *Click on organism name to get more information.*

◦ **Prototheria**

- **Monotremata** (egg-laying mammals)
  - **Ornithorhynchidae**
    - **Ornithorhynchus**
      - **Ornithorhynchus anatinus** (platypus)
  - **Tachyglossidae** (echidnas)
    - **Tachyglossus** (short-nosed echidnas)
      - **Tachyglossus aculeatus** (Australian echidna)
        - **Tachyglossus aculeatus aculeatus**
    - **Zaglossus**
      - **Zaglossus bruijnii** (long-beaked echidna)

◦ **Theria**

- **Eutheria** (placentals)
  - **Carnivora**
    - **Fissipedia**
      - **Canidae**
        - **Alopex**
          - **Alopex lagopus** (Arctic fox)
        - **Atelocynus**
          - **Atelocynus microtis** (small-eared dog)
        - **Canis**
          - **Canis adustus** (side-striped jackal)
          - **Canis aureus** (golden jackal)
          - **Canis familiaris** (dog)
          - **Canis latrans** (coyote)
          - **Canis lupus** (gray wolf)
            - **Canis lupus baileyi** (Mexican gray wolf)
            - **Canis lupus chanco** (Mongolian wolf)
            - **Canis lupus himalayensis** (Himalayan wolf)
            - **Canis lupus hodophilax** (Japanese wolf)
            - **Canis lupus indica**
            - **Canis lupus pallipes** (Indian wolf)

**NCBI Taxonomy**

- 細胞内の働き

- 代謝系

- Carbohydrate Metabolism
- TCA cycle
- Glycolysis
- Energy Metabolism

.....

- シグナル伝達系

- Receptor Tyrosine Pathway
- EGF Pathway
- PDGF Pathway
- GPCR Pathway

.....

- 細胞骨格系
- 輸送・分泌系
- 転写翻訳系
- タンパク質分解系

タキソノミー

- 細胞内の働き(世界)

- 種類

- 代謝系
- シグナル伝達系
- 細胞骨格系
- 輸送・分泌系
- 転写翻訳系
- タンパク質分解系

- 機能

- 変換する
- 化合物の構造
- 反応を起こす
- 移動する

- 属性

- 構造の種類
- 力の種類

- 構成物

- 対象物
- 装置
- 媒体

Why EGF Pathway is a Cell Signaling Pathway?

オントロジー

# タキソノミーとオントロジー

## ❖ 解釈を計算機可読に

### ❖ 事実として可読

タキソノミー(的)

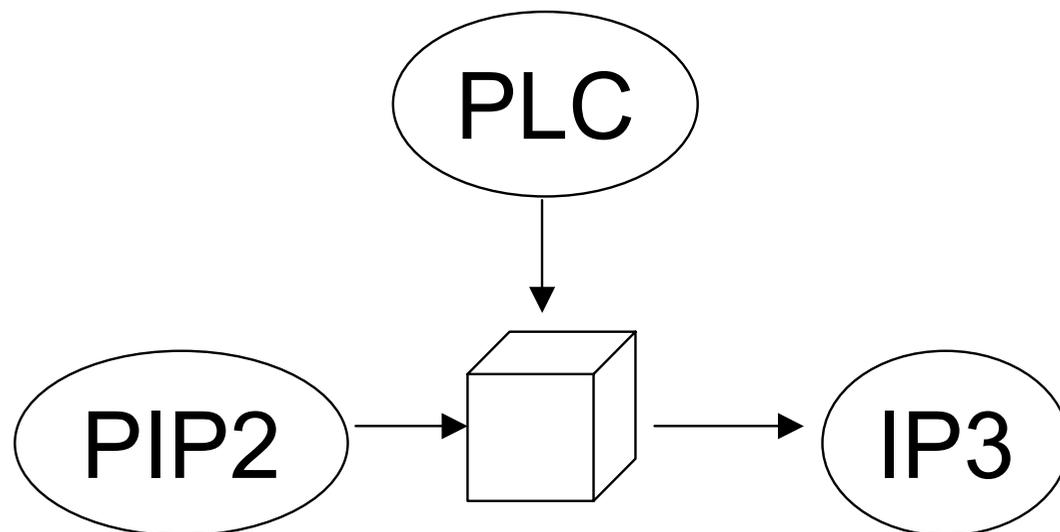
- ❖ 細胞内の働きには代謝系とシグナル伝達系がある

### ❖ ルールとして可読

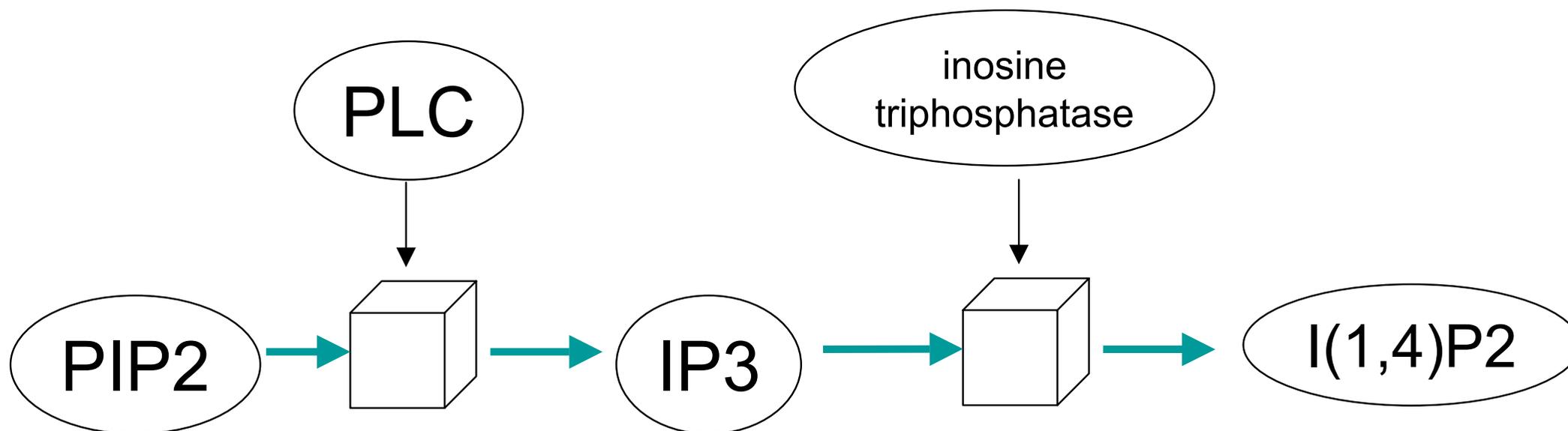
オントロジー(的)

- ❖ 代謝系：連鎖した反応の間に「化学構造の変化」が伝わる系
- ❖ シグナル伝達系：連鎖した反応の間に「反応を起こす活性」が伝わる系

## 単独の反応

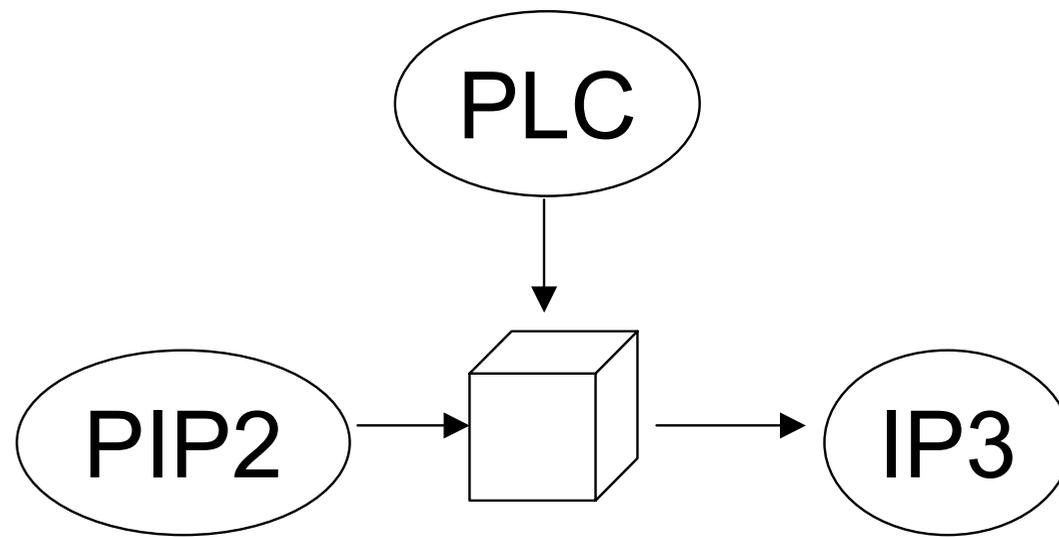


## 反応の連鎖(代謝系)

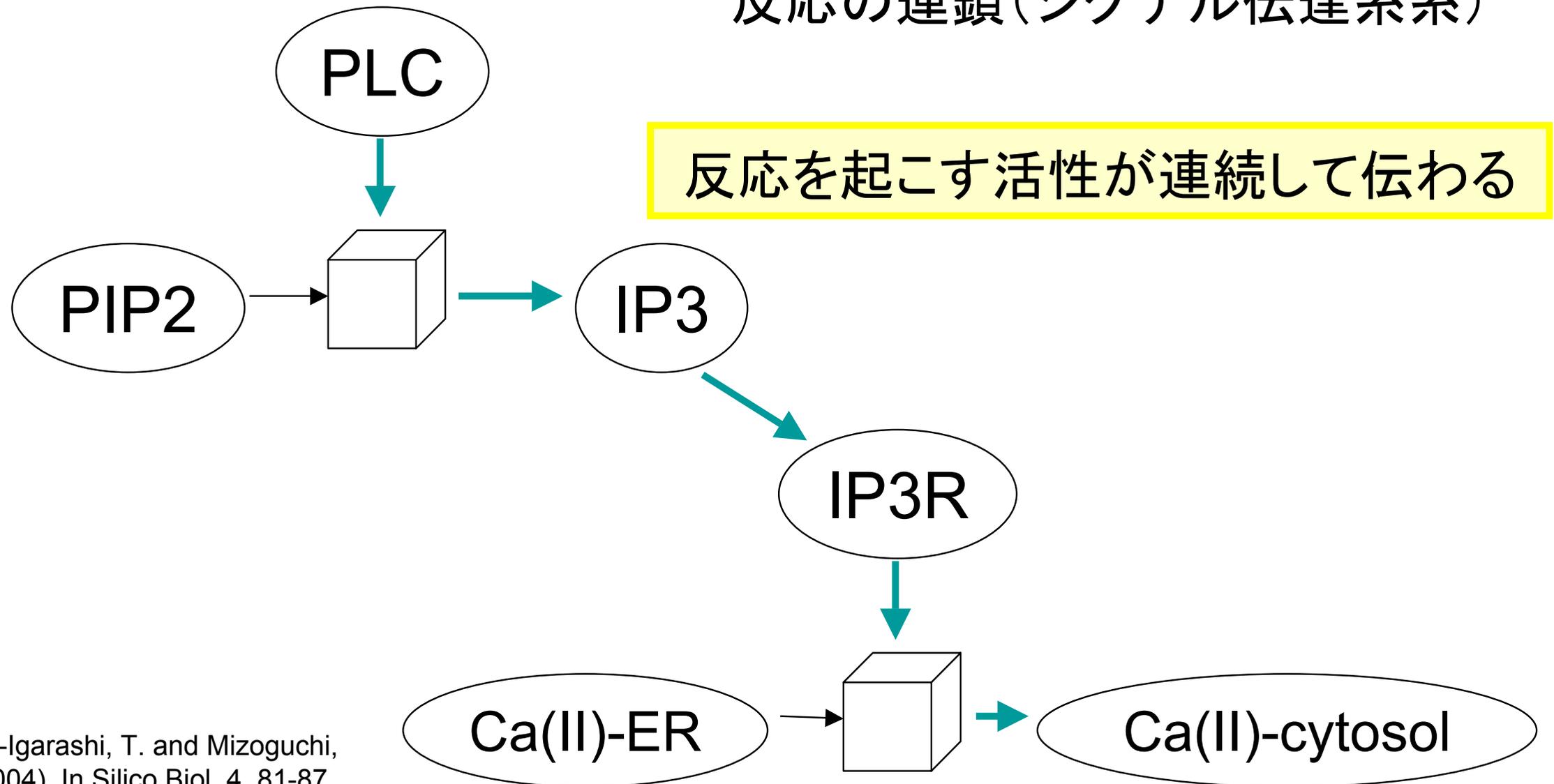


化合物の構造変化が連続して伝わる

# 単独の反応



# 反応の連鎖 (シグナル伝達系系)



- 細胞内の働き

- 代謝系

- Carbohydrate Metabolism
- TCA cycle
- Glycolysis
- Energy Metabolism

.....

- シグナル伝達系

- Receptor Tyrosine Pathway
- EGF Pathway
- PDGF Pathway
- GPCR Pathway

.....

- 細胞骨格系
- 輸送・分泌系
- 転写翻訳系
- タンパク質分解系

タキソノミー

- 細胞内の働き(世界)

- 種類

- 代謝系
- シグナル伝達系
- 細胞骨格系
- 輸送・分泌系
- 転写翻訳系
- タンパク質分解系

- 機能

- 変換する
- 化合物の構造
- 反応を起こす
- 移動する

- 属性

- 構造の種類
- 力の種類

- 構成物

- 対象物
- 装置
- 媒体

Why EGF Pathway is a Cell Signaling Pathway?

オントロジー

# 知識 Knowledge

## ✓ 事実知識 (ファクト Fact)

DNA配列、アミノ酸配列、糖脂質の修飾、細胞内局在、機能、発現量、相互作用分子、、、

## ✓ 推論的知識 (ルール Rule)

➤ 塩基はリン酸結合で連結されている

➤ ArgはTrpよりLysにより置換され易い

➤ 糖鎖が付くアミノ酸配列はAsn-X-Ser/Thr

➤ etc. ....

## タキソノミー: 事実知識の背景知識を構成する 概念を階層化

### ✓ 事実知識 (ファクト Fact)

DNA配列、アミノ酸配列、糖脂質の修飾、細胞内局在、機能、発現量、相互作用分子、、、

### ✓ 推論的知識 (ルール Rule)

- 塩基はリン酸結合で連結されている
- ArgはTrpよりLysにより置換され易い
- 糖鎖が付くアミノ酸配列はAsn-X-Ser/Thr

オントロジー: 推論的知識の背景知識を構成  
する概念とその構造を明示化

タキシノミー: 事実知識の背景知識を構成する  
概念を階層化

✓ 事実知識 (ファクト Fact)

DNA配列、**一致および類似性で判断する** 包  
内局在、機能、発現量、相互作用分子、、、

✓ 推論的知識 (ルール Rule)

人間は死ぬ

ソクラテスは人間である

ソクラテスは死ぬ

連結されている

は置換され易い

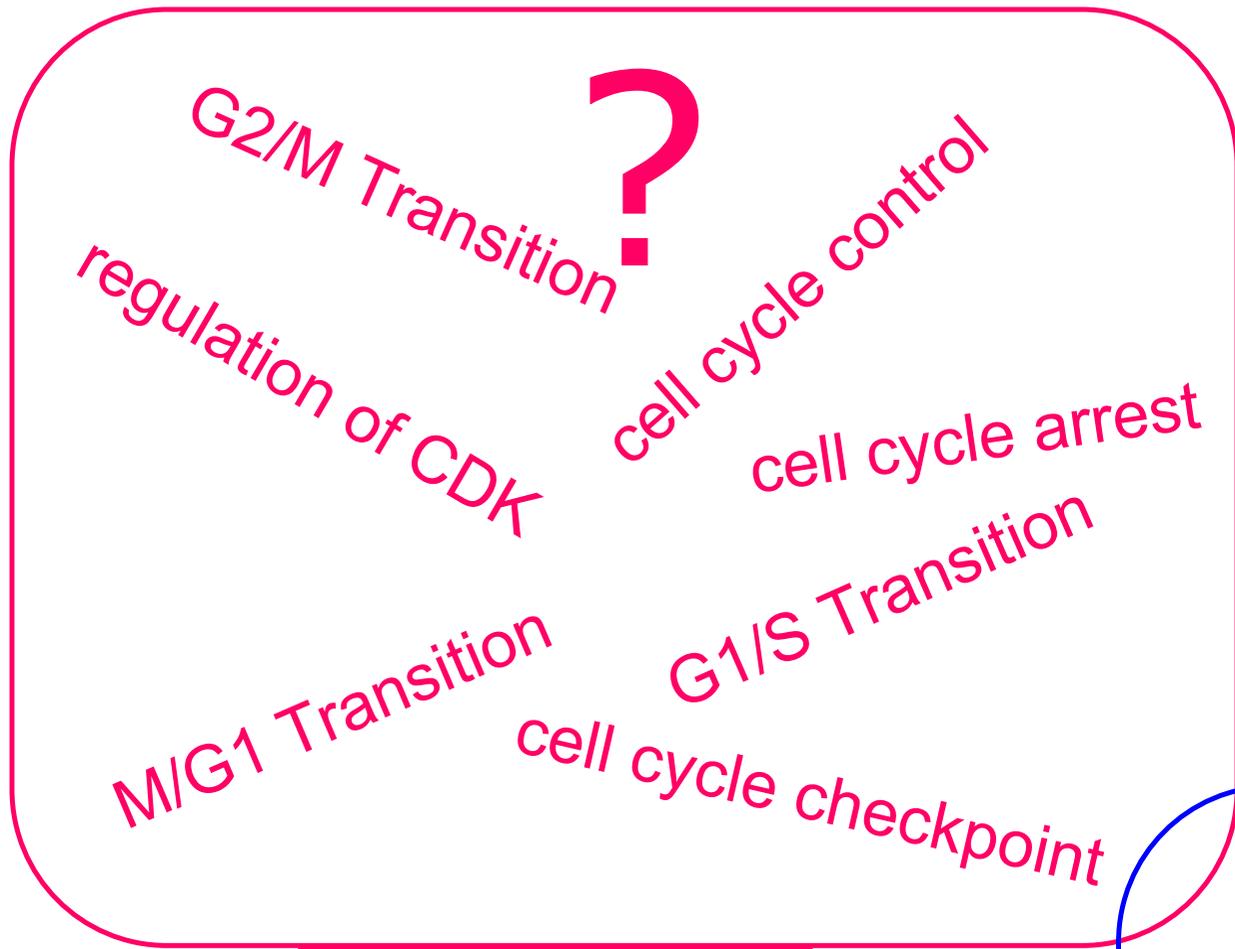
論理的に推論し判断する

オントロジー: 推論的知識の背景知識を構成  
する概念とその構造を明示化

# ゲノム研究におけるオントロジーの開発と利用

- ❖ 生物学およびゲノム研究
- ❖ 計算機科学におけるオントロジー
- ❖ ゲノム研究とオントロジー

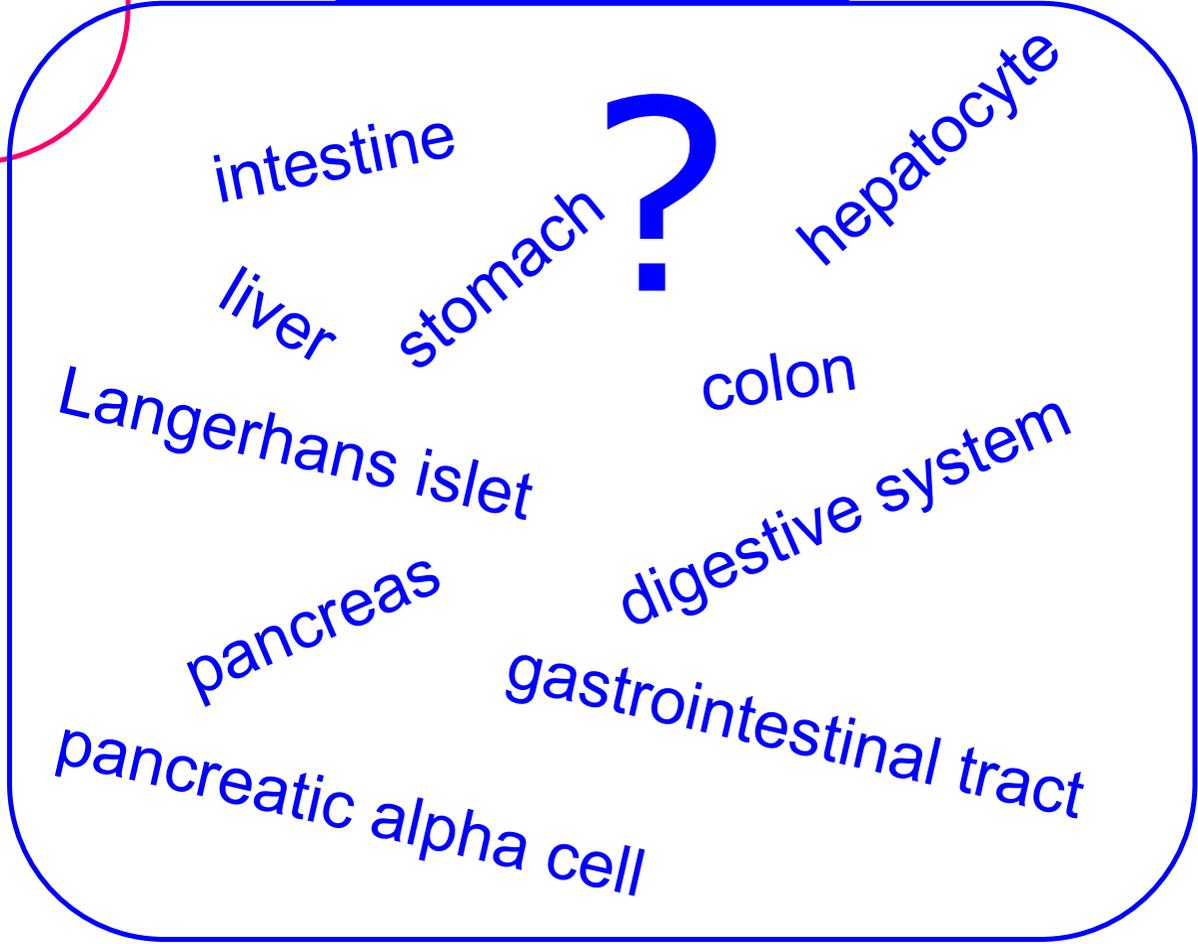
ある領域の概念の階層構造を  
つくりそれを利用する  
オントロジーの利用法



生物機能の概念

ある領域の概念を収集

組織/器官の概念



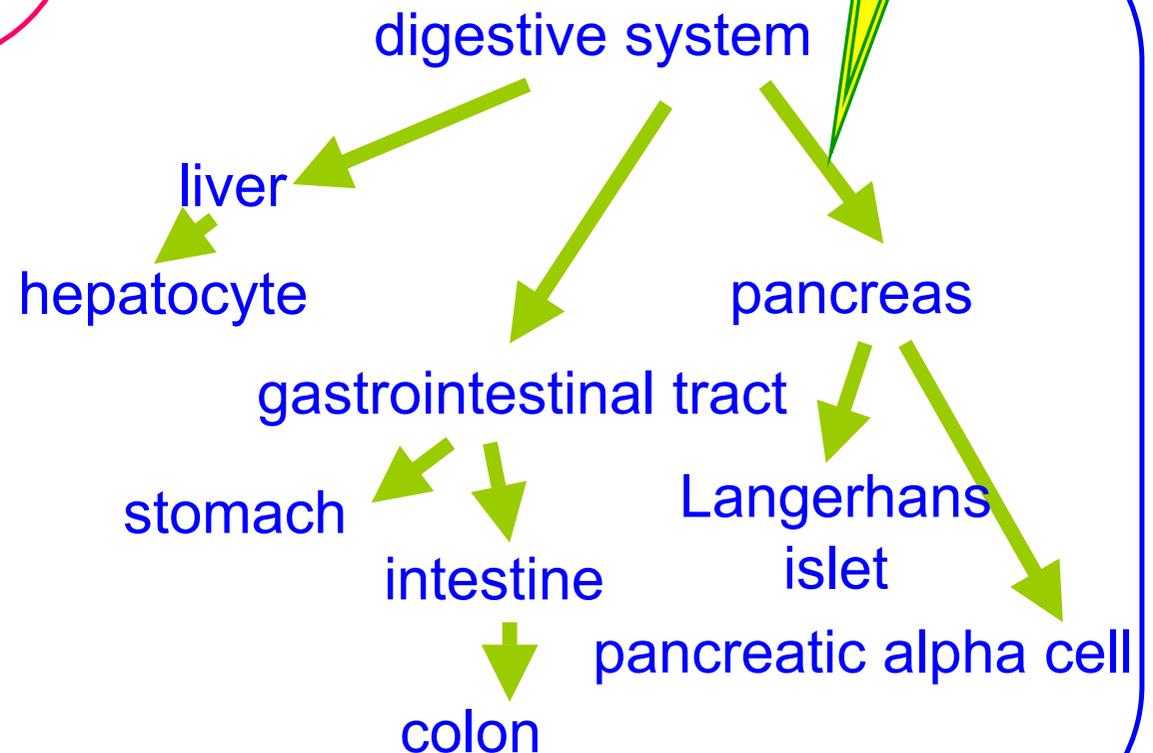
人間は概念を概念間の関係で理解している

## オントロジー 概念と概念の関係の定義

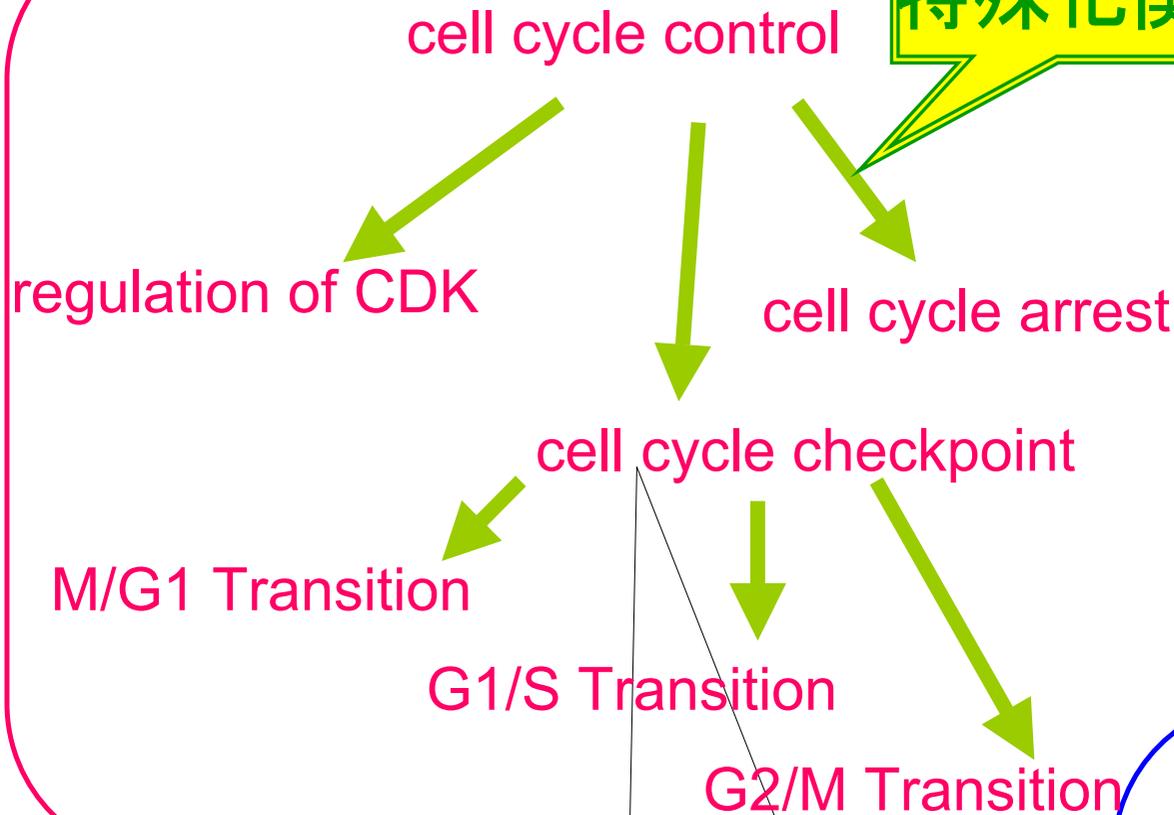
一貫した関係に基づく階層分類

## 部分関係

## 組織/器官の概念



## 特殊化関係

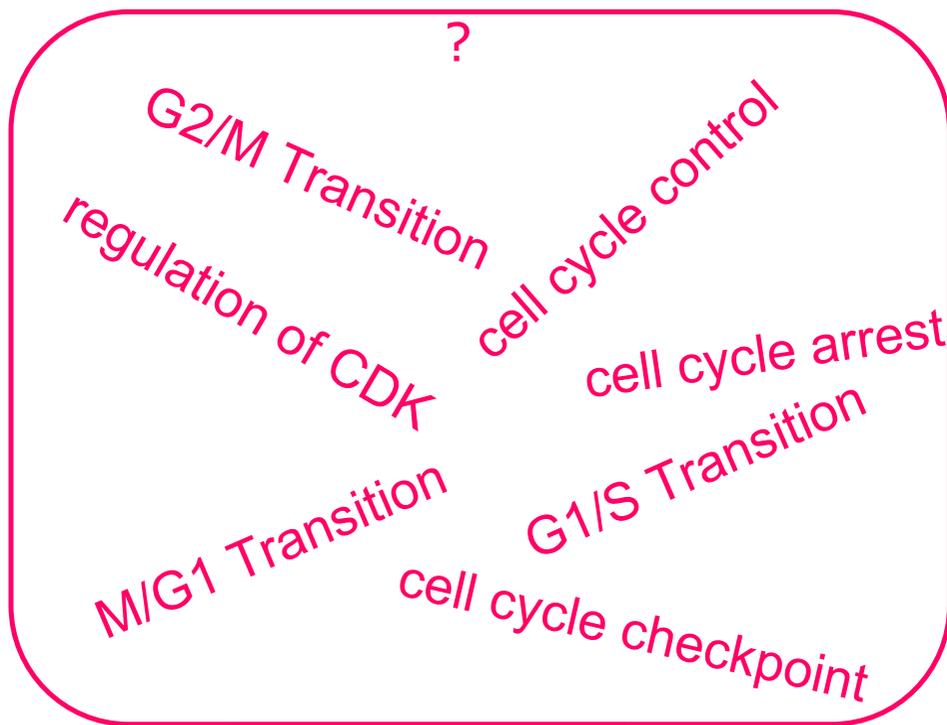


## 生物機能の概念

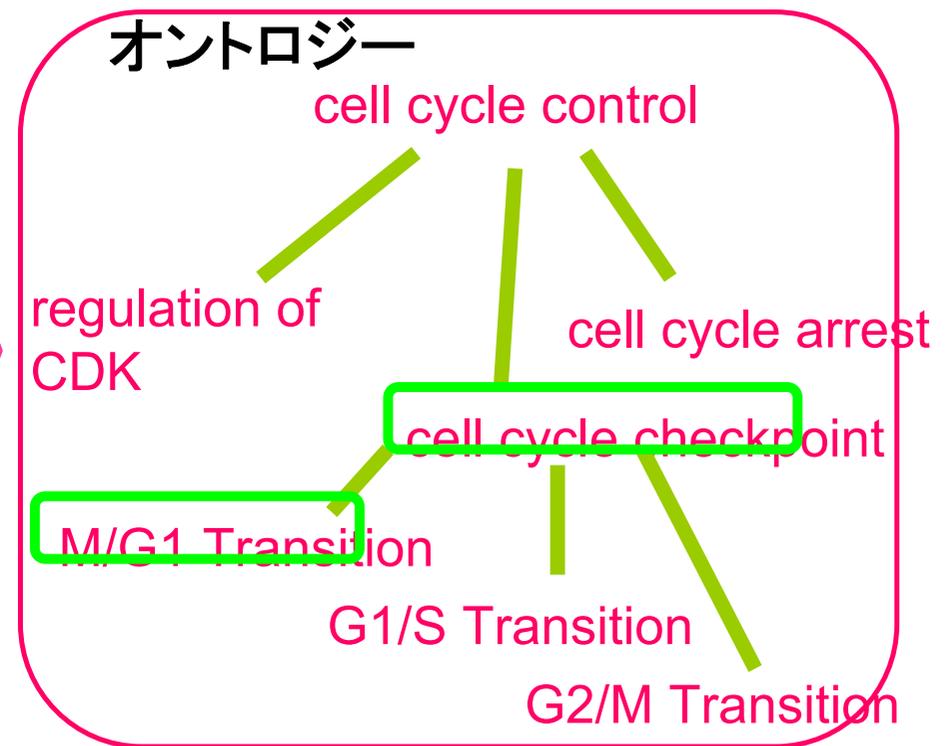
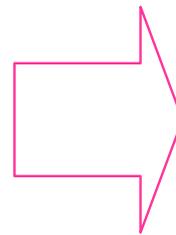
A point in eukaryotic cell cycle processes where the processes can be halted until all the required factors are fulfilled for entering to the next process.

## 概念の定義

語彙の統制(別名): starting point



生物機能の概念



生物機能の概念

## 概念を計算機で計算可能とする

計算する

$$A = B$$

$$A > B$$

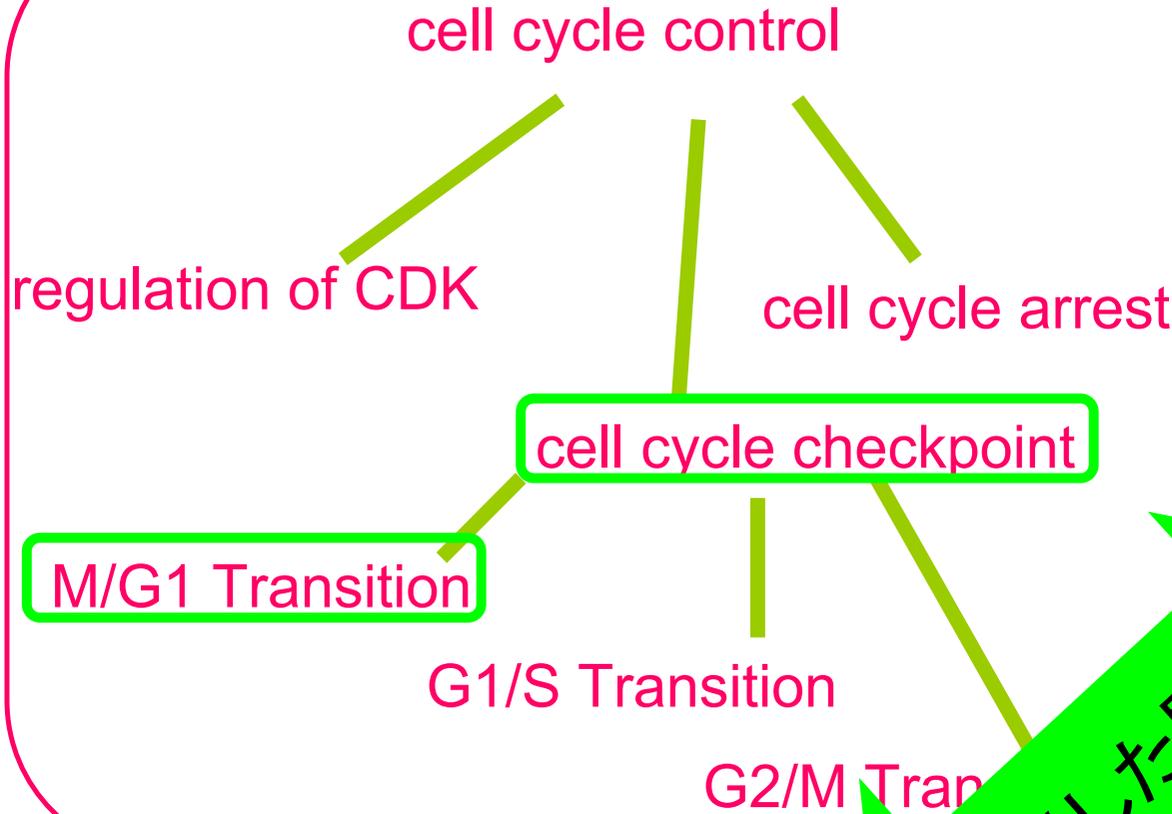
$$|A - B|$$

等しいかどうか？  
順番つけができる  
距離が定義できる

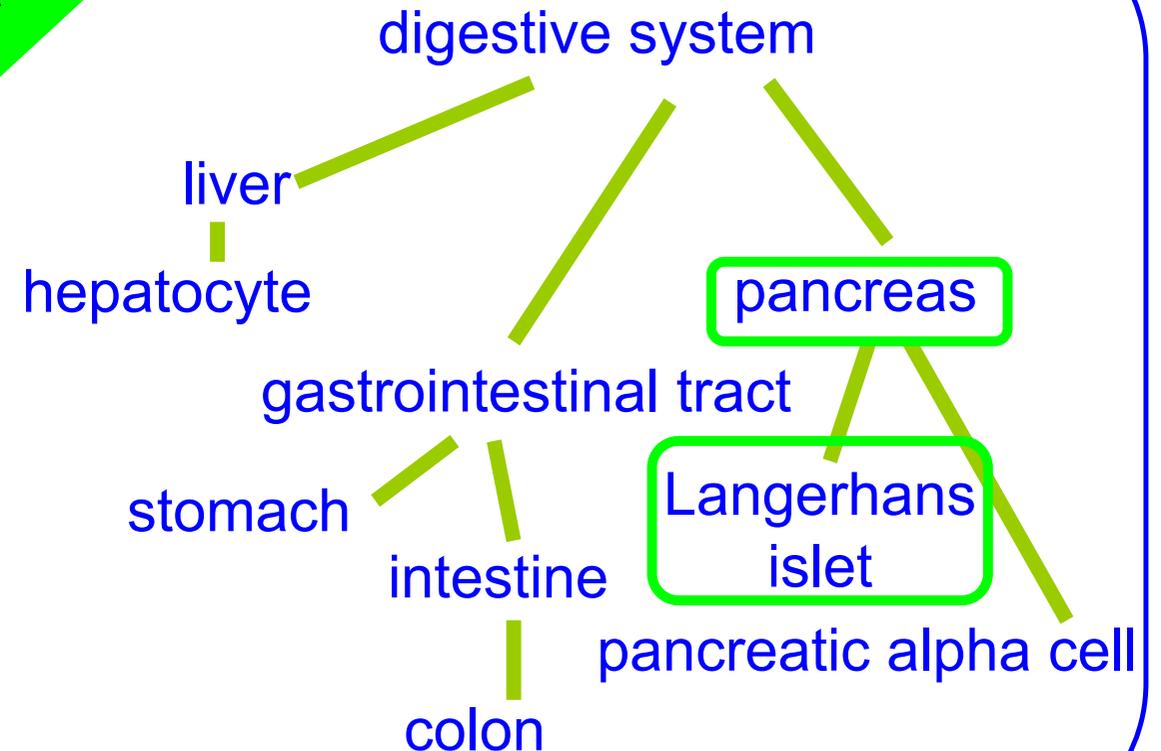
# オントロジーに基づくデータベース統合

## データ間のさまざまな関係づけ

データベースA  
Protein X  
AC:P02034 (配列)  
組織: pancreas  
生命現象: M/G1 Transition



## 組織/器官の概念



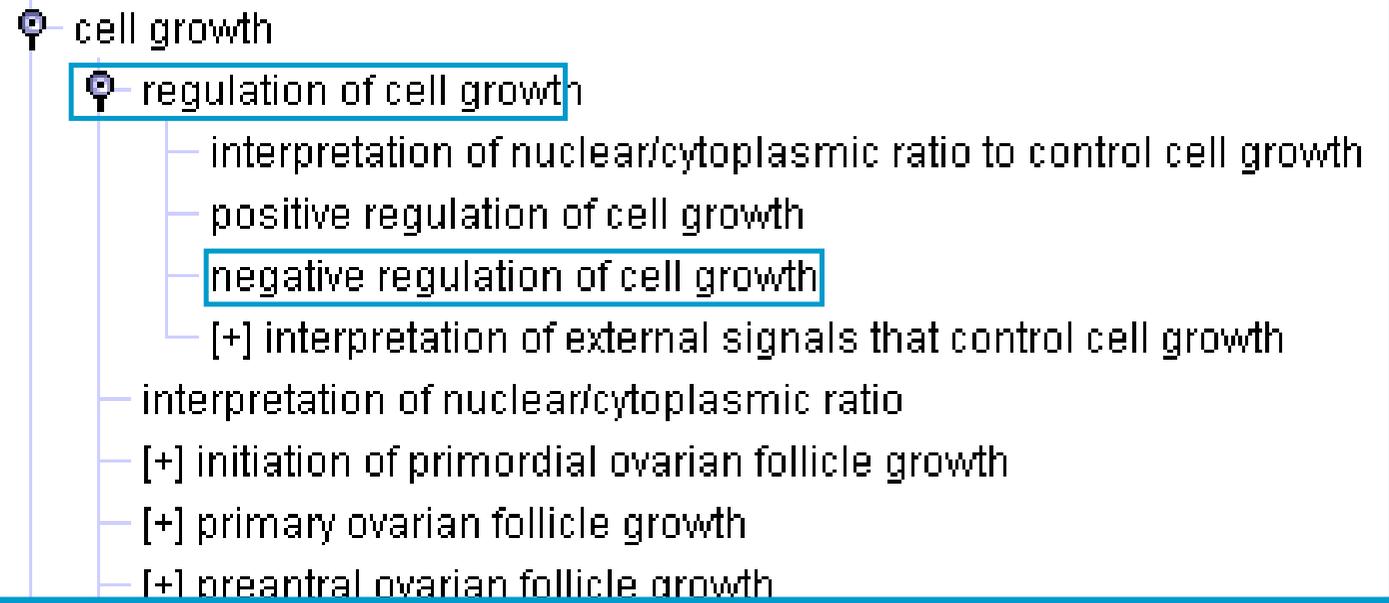
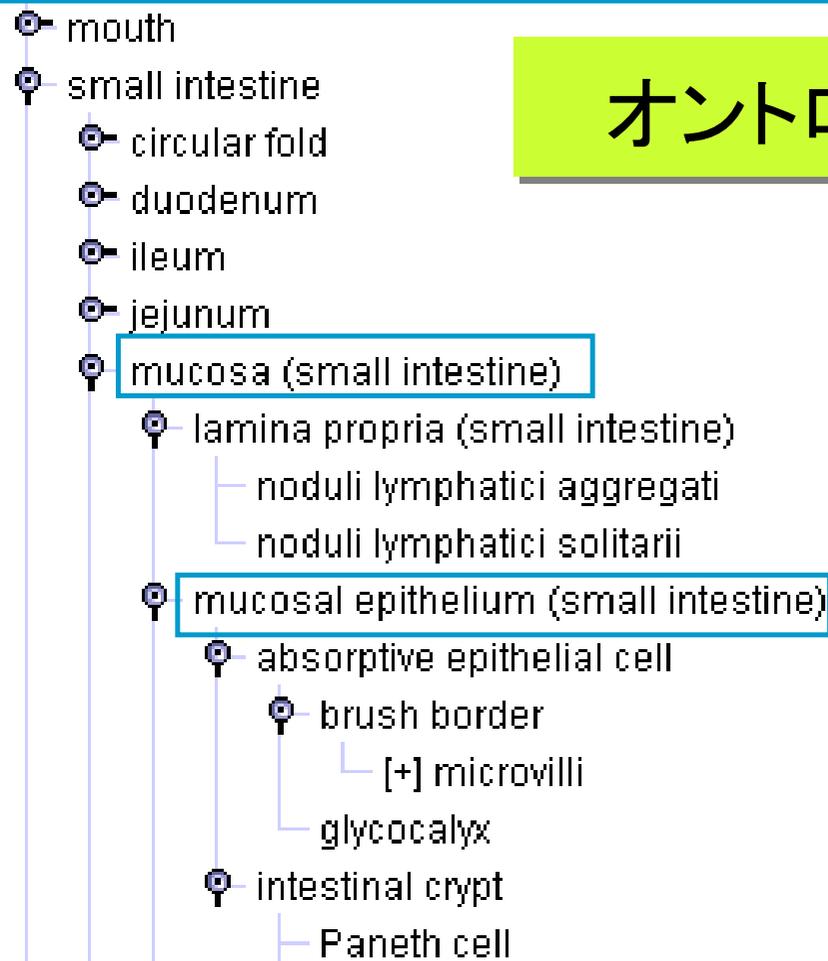
## 生物機能の概念

データベースB  
Protein Y  
AC:P00815 (配列)  
組織: Langerhans islet  
生命現象: cell cycle checkpoint

## オントロジーの利用

類似した関係

オントロジーを用いた検索条件拡張



A: 小腸粘膜 で 細胞増殖制御 に関与し Sos が含まれるパスウェイ

Q: 小腸粘膜上皮 で 細胞増殖阻害 に関与し Sos が含まれるパスウェイ

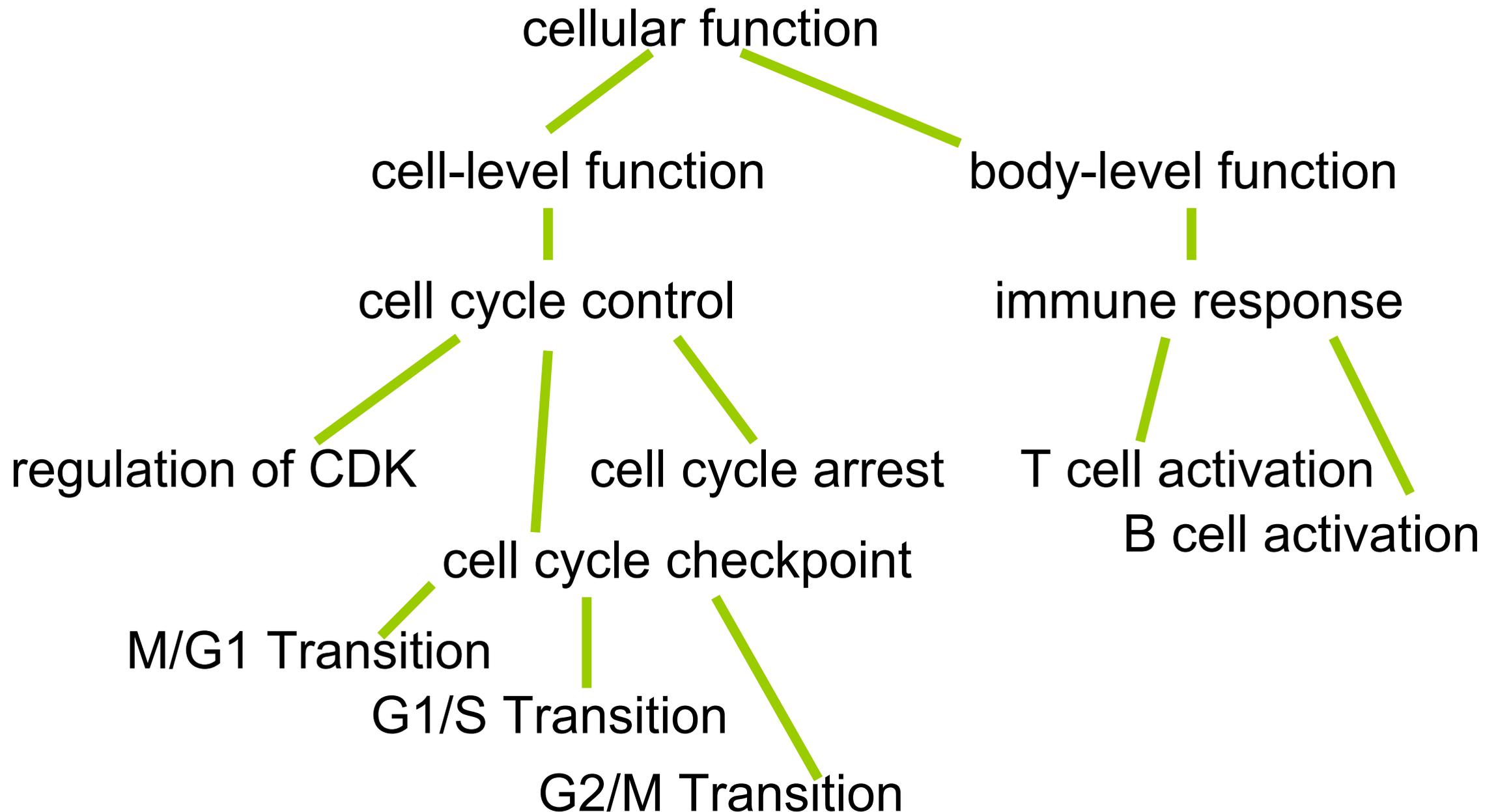
器官

生物機能

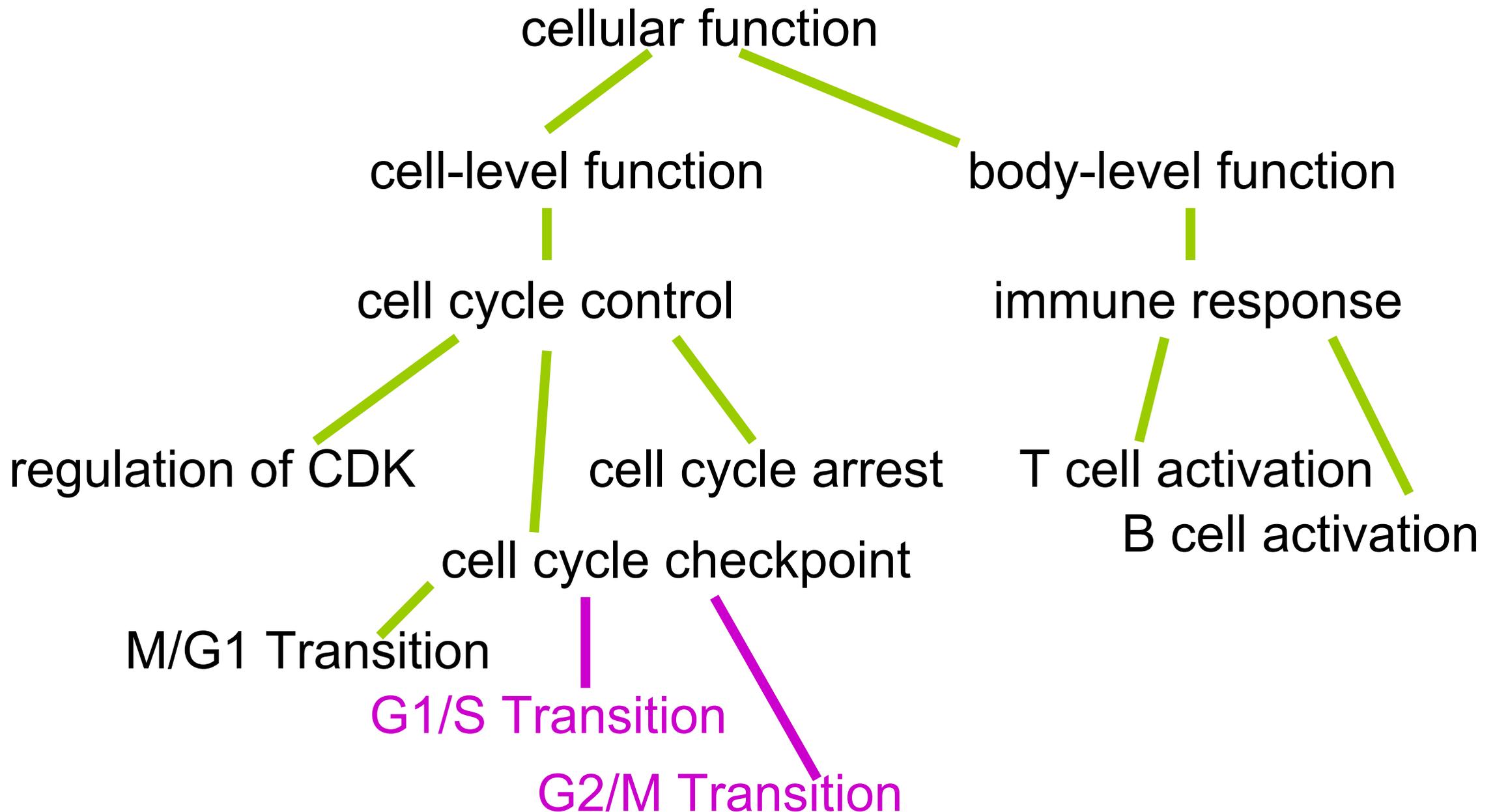
分子

# オントロジーの概念間の計算

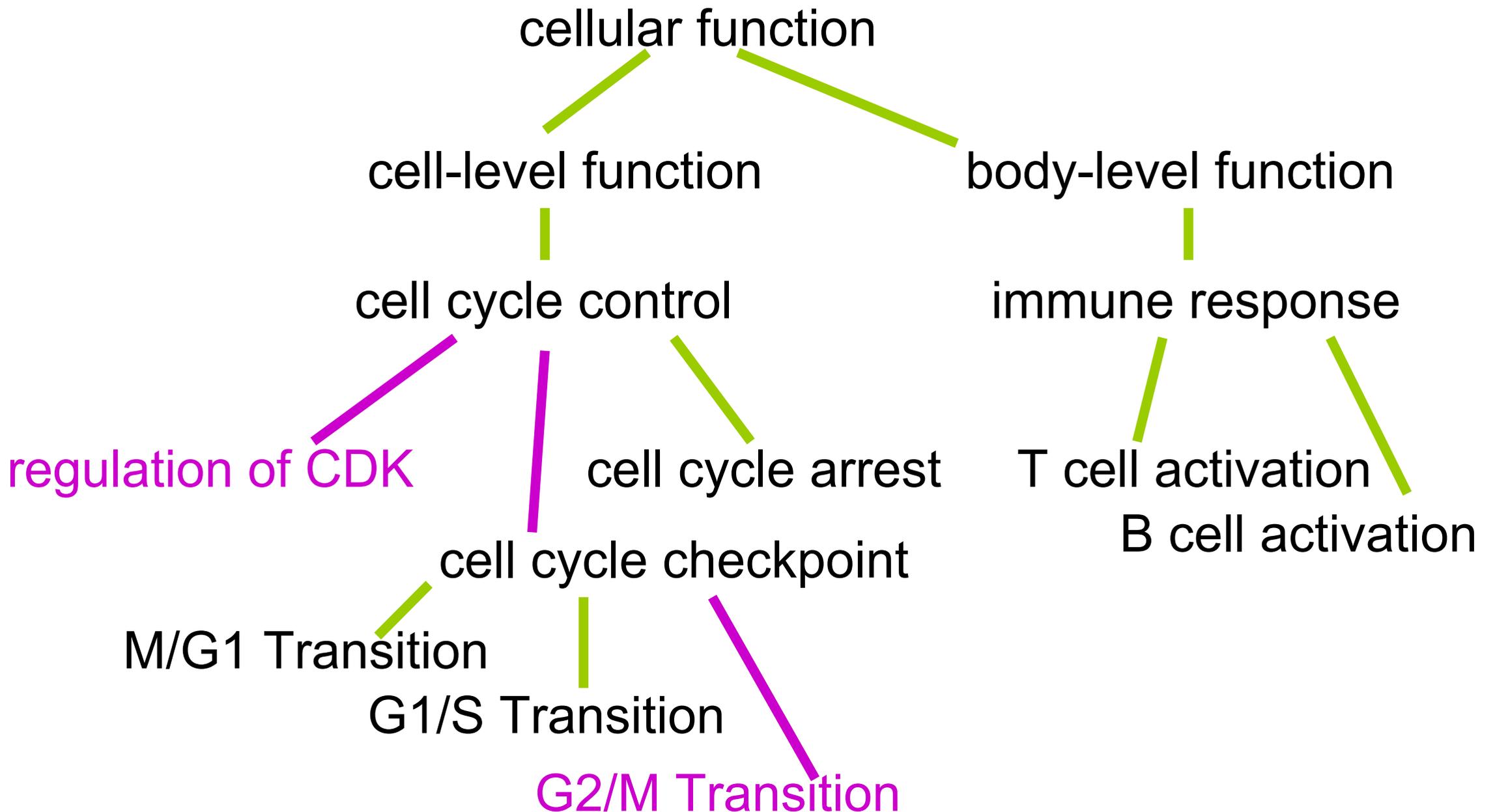
## 概念の汎化特殊化関係 is-a関係



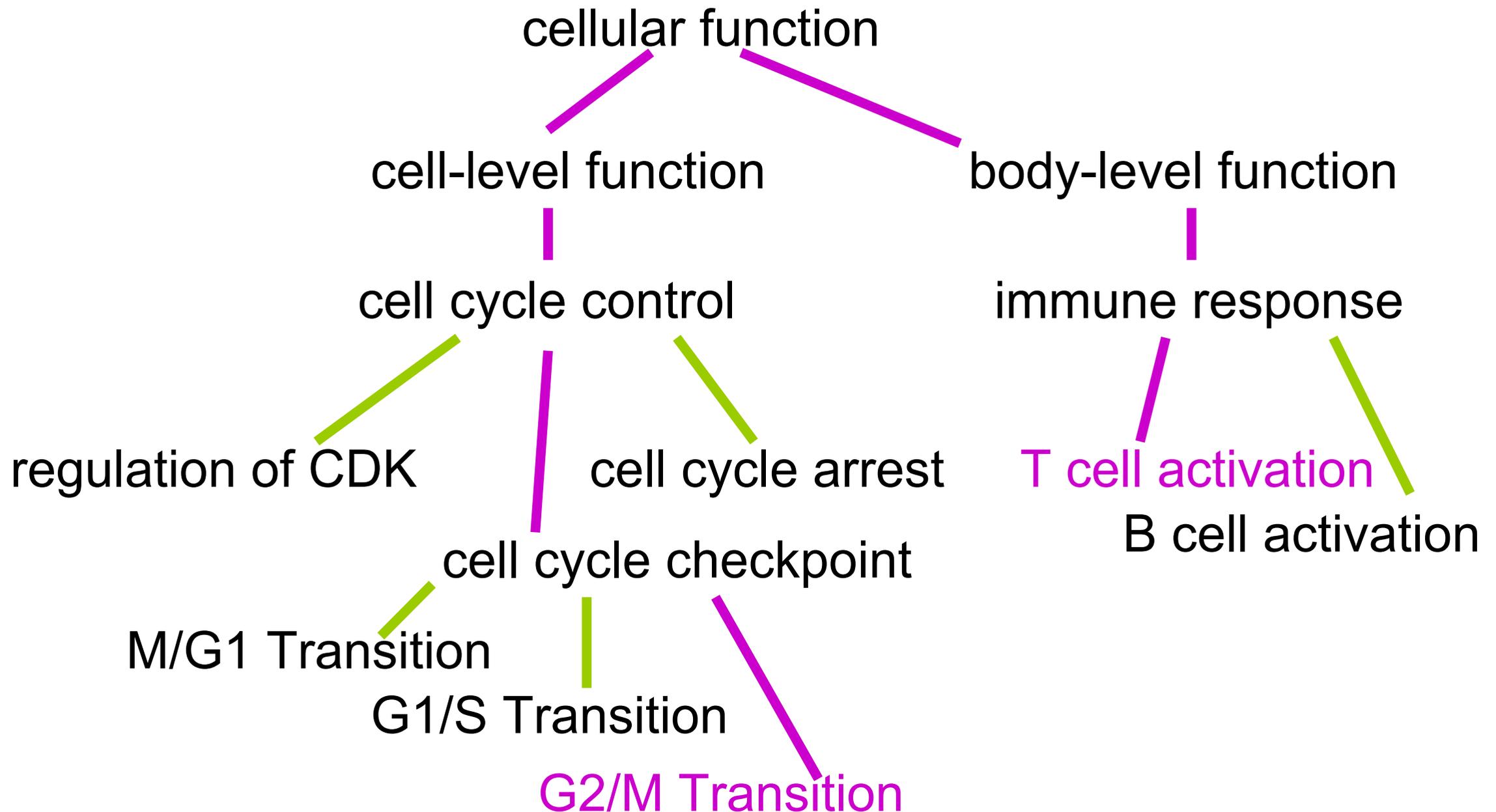
# オントロジーの概念間の計算



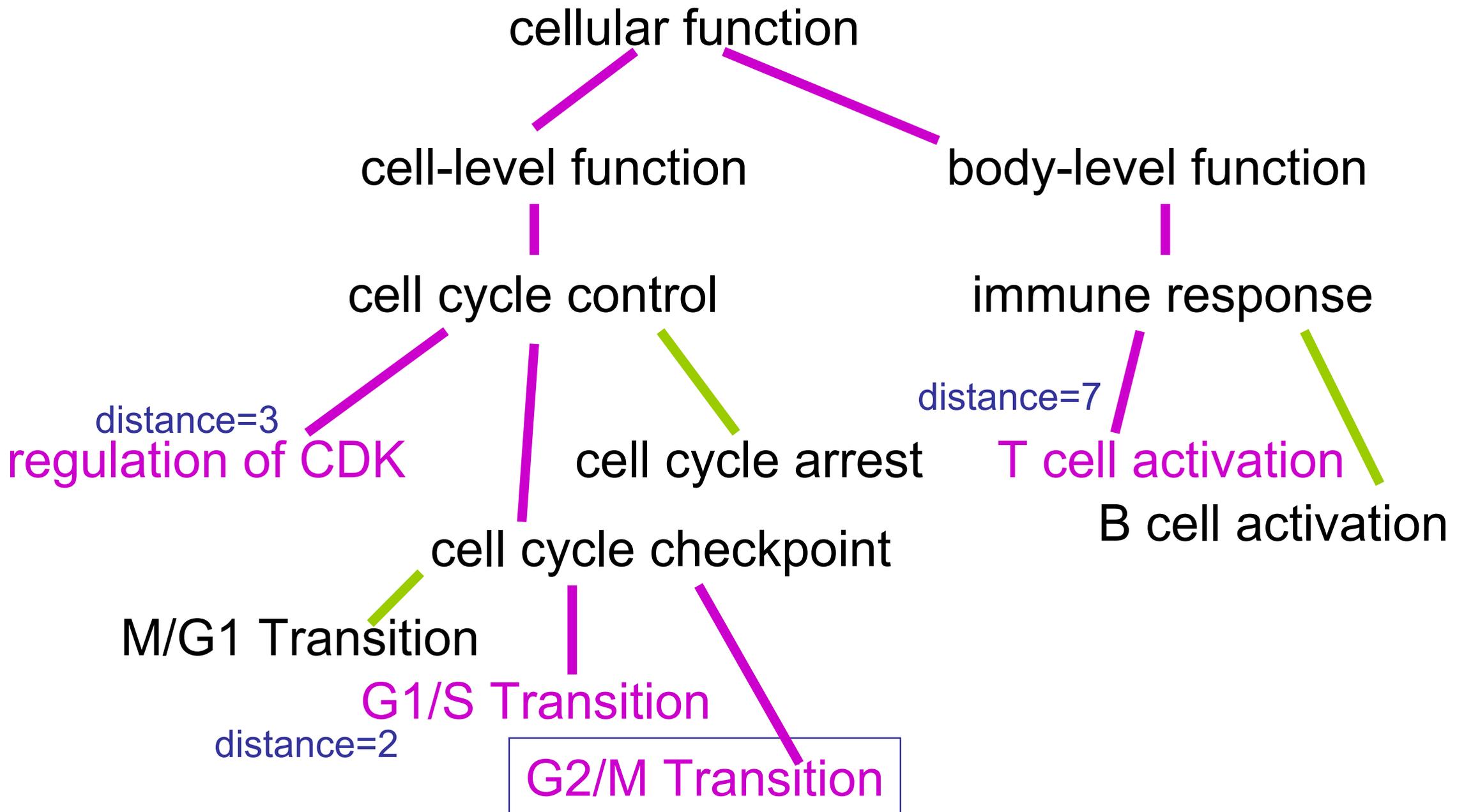
# オントロジーの概念間の計算



# オントロジーの概念間の計算



# オントロジーの概念間の計算



# 生物学を対象としたオントロジー

TAMBIS (Transparent Access to Multiple Biological Information Sources Ontology), UK  
MBO (Molecular Biology Ontology), Germany

- 生命学世界全体を対象とした、生物学のトップレベルオントロジー

- Entity
  - Generalized Structure
    - Abstract Structure
    - protocol
    - method
    - classification
    - reaction
    - pathway
    - Information source
    - Physical Structure
      - solid structure
      - biomolecular structure
      - component of biomolecular struct.
    - chemical
  - Generalised Substance
    - Body Substance
    - body fluid
    - tissue
    - Cellular Substance
  - Generalised Process
    - Biological Process
    - body process
    - biomolecular process

- Entity (continued)
  - Generalised Function
    - Biological Function
    - Molecular Modification
    - Binding
    - DNA Replication and Repair
    - Transport
    - Maintenance of Structure
    - Cellular Growth and Proliferation
    - Signal Transduction
    - Enzymic Function, Receptor, Hormone, Toxin, Inhibitor, Targeting, Stress
  - Modifier Concept
    - Aspect
    - Collection

# 生物学を対象としたオントロジー

ImMunoGeneTics (IMGIT), U Montpellier  
Gene Ontology (GO), USA and EU  
INTERACTIONS Ontology, SRI  
PharmGKB, NIGMS

- 対象領域を限定したドメインオントロジー

IMGIT, 免疫系の構造と機能

GO, 遺伝子産物の機能

INTERACTIONS, 代謝反応

PharmGKB, 医薬品の作用

# 生物学を対象としたオントロジー

Mouse Anatomical Dictionary, Jackson Lab.  
Digital Anatomical Foundation Model (FMA), U  
Washington  
CYTOMER, U Goettingen

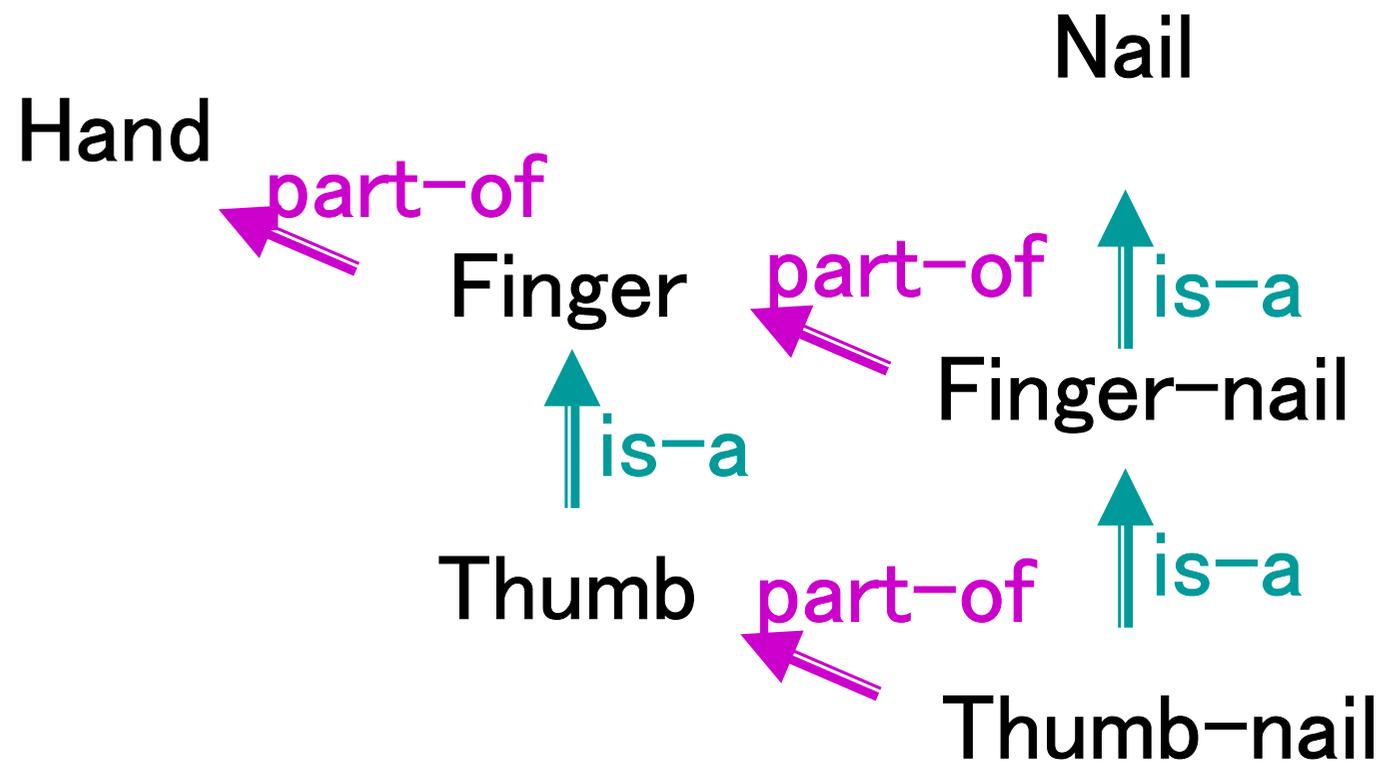
- 対象領域を限定したドメインオントロジー

Mouse Anatomical Dictionary, マウスの発生に  
おける組織と器官

FMA, ヒトの組織と器官

CYTOMER, ヒトの組織と器官

# is-a and part-whole relationships in tissue/organ-hierarchy



Ontology for Tissues / Organs

# GO, Gene Ontology

- 5ゲノムデータベースチームの共同プロジェクト  
ショウジョウバエ、酵母、マウス、シロイヌナズナ、線虫
- 真核モデル生物に共通な遺伝子産物の機能について定義
- 真核生物の遺伝子機能の比較のため(比較ゲノム学)
- Gene Ontologyが対象とする概念
  - Molecular Function: 単独分子の機能
  - Biological Process: 複数分子により制御される機能
  - Cellular Component: 分子が働く細胞内の場所

機能

構造

# 生物学を対象としたオントロジー

ImMunoGeneTics (IMG T), U Montpellier  
Gene Ontology (GO), USA and EU  
INTERACTIONS Ontology, SRI  
PharmGKB, NIGMS

- 対象領域を限定したドメインオントロジー

IMG T, 免疫系の構造と機能  
GO, 遺伝子産物の機能  
INTERACTIONS, 代謝反応  
PharmGKB, 医薬品の作用

機能

# 生物学を対象としたオントロジー

Mouse Anatomical Dictionary, Jackson Lab.  
Digital Anatomical Foundation Model (FMA), U  
Washington  
CYTOMER, U Goettingen

- 対象領域を限定したドメインオントロジー

Mouse Anatomical Dictionary, マウスの発生における組織と器官

FMA, ヒトの組織と器官

CYTOMER, ヒトの組織と器官

構造

# 機能概念はどの知識ドメインでも主張な対象

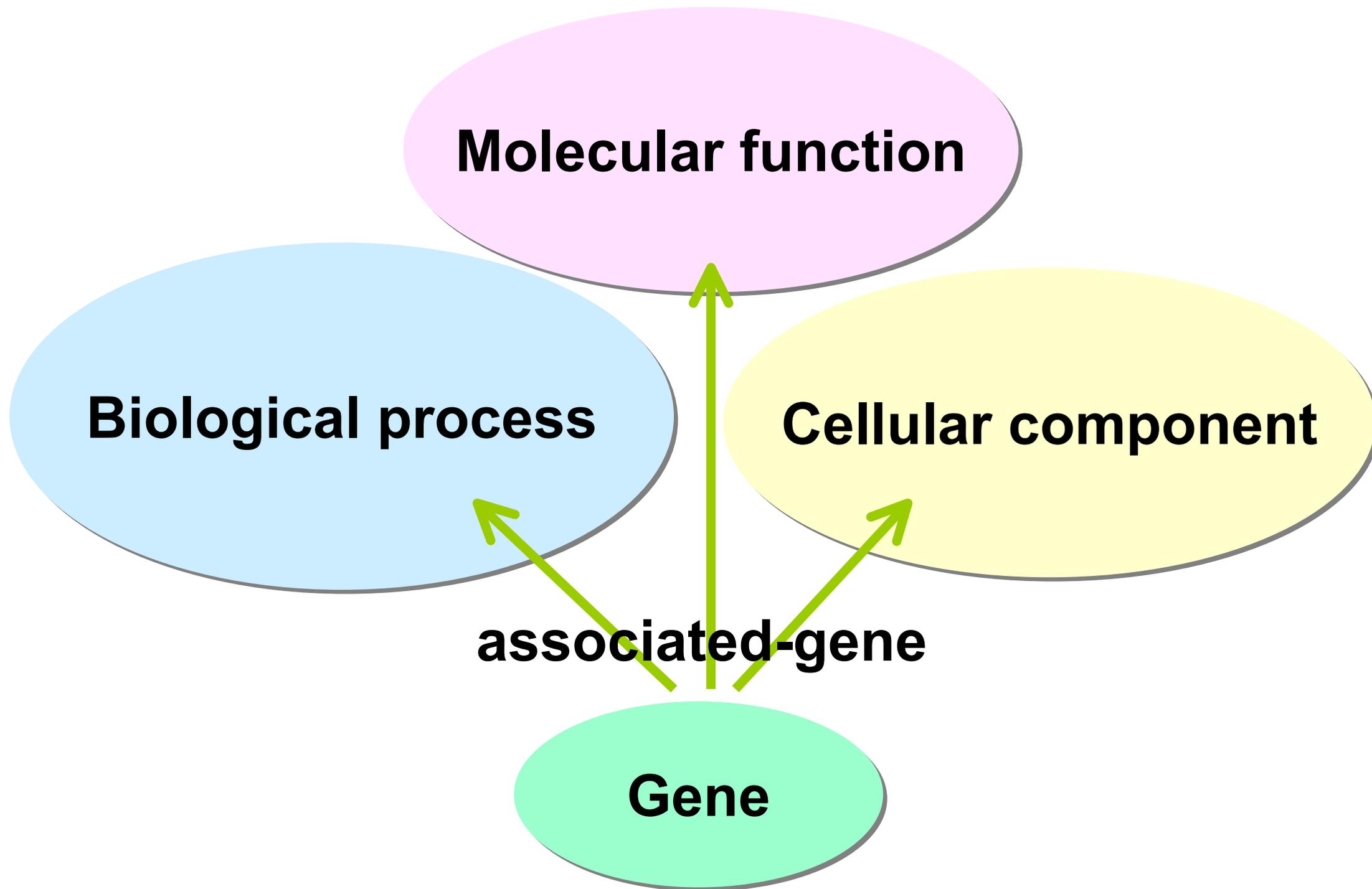
- 機能とは、対象物の属性の変化を、目的の下で解釈したもの
  - 椅子のリクライニング機能
    - 椅子の「背もたれの角度」の変化を、「人間が使用してくつろぐ」という目的の下で解釈した結果
  - NFkBの転写活性機能
    - NFkBの「染色体の特定の位置に結合」する変化を、「転写というプロセスを開始する」という目的の下で解釈した結果
- 機能の定義(解析)は目的を達成する方法を定義(解析)する鍵となる
  - 生物学の目的のひとつは、生物がもつ多種多様な機能の原理を知ること

# 機能概念はどの知識ドメインでも主張な対象

---

- 機能とは、対象物の属性の変化を、目的の下で解釈したもの
- 椅子のリクライニング機能
  - 椅子の「背もたれの角度」の変化を、「人間が使用してくつろぐ」という目的の下で解釈した結果
- NFkBの転写活性機能
  - NFkBの「染色体の特定の位置に結合」する変化を、「転写というプロセスを開始する」という目的の下で解釈した結果

構造は機能の説明に必要となる



**Gene Ontologyが考える遺伝子産物の機能の世界**

# AmiGO

## chemokine activity

**Accession:** GO:0008009

**Aspect:** molecular\_function

**Synonyms:** None

**Definition:**

The function of a family of chemotactic pro-inflammatory activation-inducible cytokines (GO:0005125) acting primarily upon hemopoietic cells in immunoregulatory processes; all chemokines possess a number of conserved cysteine residues involved in intramolecular disulfide bond formation.

### Term Lineage

GO:0003673 : Gene\_Ontology ( 149784 )

① GO:0003674 : molecular\_function ( 101079 )

① GO:0005488 : binding ( 27236 )

① GO:0005102 : receptor binding ( 1485 )

① GO:0005125 : cytokine activity ( 385 )

① **GO:0008009 : chemokine activity ( 79 )**

① GO:0001664 : G-protein-coupled receptor binding ( 107 )

① GO:0042379 : chemokine receptor binding ( 79 )

① **GO:0008009 : chemokine activity ( 79 )**

① GO:0004871 : signal transducer activity ( 6503 )

① GO:0005102 : receptor binding ( 1485 )

① GO:0005125 : cytokine activity ( 385 )

① **GO:0008009 : chemokine activity ( 79 )**

① GO:0001664 : G-protein-coupled receptor binding ( 107 )

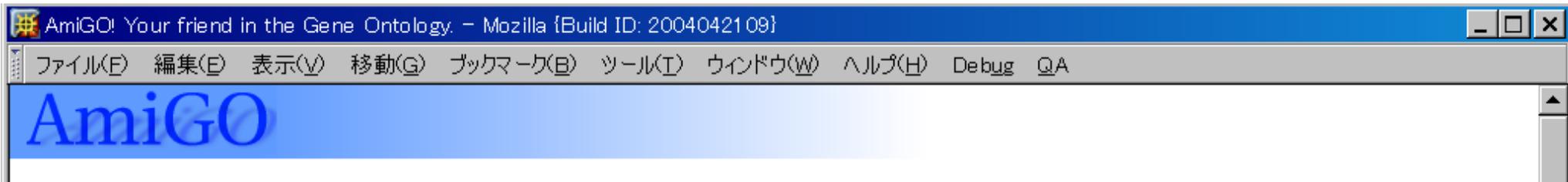
① GO:0042379 : chemokine receptor binding ( 79 )

① **GO:0008009 : chemokine activity ( 79 )**

[Graphical View](#)

### External References

① InterPro ( 5 )



**chemokine**

Accession  
Aspect: m  
Synonym  
Definition  
The i  
hem  
intra

Term Lineage  
GO:00036  
GO

External Refer  
InterPro

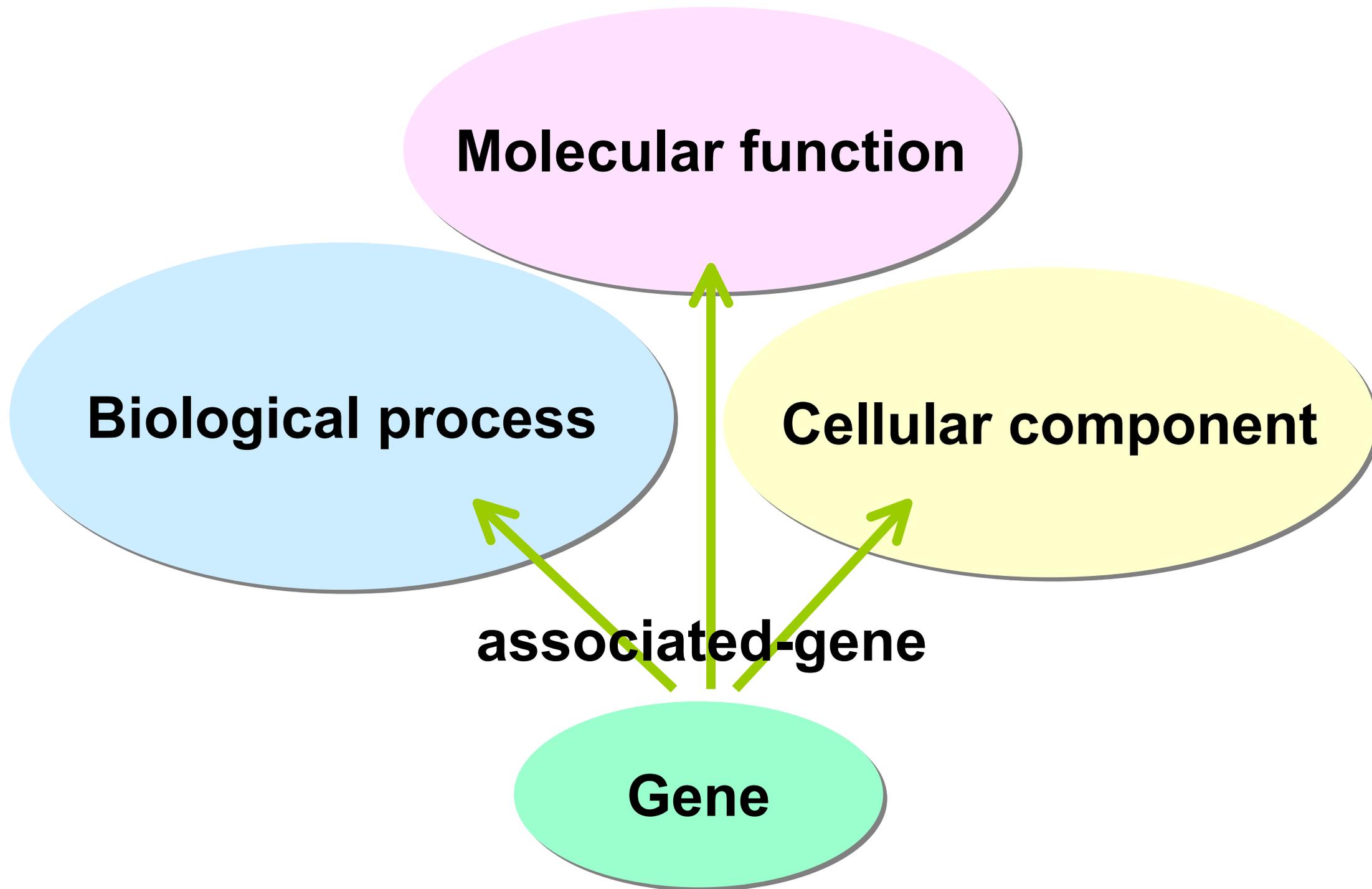
AmiGO! Your friend in the Gene Ontology. - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

GO:0008009 : chemokine

<a href="#">SY24_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A24 precursor
<a href="#">SY21_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A21 precursor
<a href="#">SY22_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A22 precursor
<a href="#">SZ11_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine B11 precursor
<a href="#">SY25_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A25 precursor
<a href="#">SY16_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A16 precursor
<a href="#">SZ13_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine B13 precursor
<a href="#">SZ14_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine B14 precursor
<a href="#">CO5_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Complement C5 precursor [Contains: C5a anaphyla
<a href="#">PLF4_HUMAN</a>	SWISS-PROT	NR - NOREFERENCE	Platelet factor 4 precursor
<a href="#">SZ10_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine B10 precursor
<a href="#">GRO_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Growth regulated protein precursor
<a href="#">IL8_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Interleukin-8 precursor
<a href="#">SY03_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A3 precursor
<a href="#">PLFV_HUMAN</a>	SWISS-PROT	NR - NOREFERENCE	Platelet factor 4 variant precursor
<a href="#">SY04_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A4 precursor
<a href="#">SY02_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Small inducible cytokine A2 precursor
<a href="#">SY05_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	
<a href="#">SY3L_HUMAN</a>	SWISS-PROT	NR - NOREFERENCE	
<a href="#">M12A_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	
<a href="#">M12B_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	
<a href="#">NAP4_HUMAN</a>	SWISS-PROT	<a href="#">TAS</a>	Neutrophil-activating protein 4

**遺伝子にアノテーションされる機能  
概念の標準化と計算機可読化**



**Gene Ontology**が考える遺伝子産物の機能の世界

**Total  
Gene  
Products  
Associated**

<b>SGD</b> <i>Saccharomyces cerevisiae</i> <a href="#">README</a>	6461
<b>FlyBase</b> <i>Drosophila melanogaster</i>	9159
<b>MGI</b> <i>Mus musculus</i>	14399
<b>TAIR</b> <i>Arabidopsis thaliana</i>	29762
<b>WormBase</b> <i>Caenorhabditis elegans</i>	11617
<b>RGD</b> <i>Rattus norvegicus</i>	3649

生物種	遺伝子数
酵母	6K
ハエ	14K
線虫	20K
ヒト	35K

# Gene Ontologyは厳密にはオントロジーと言えない

- ❖ 機能と構造概念のタキシノミーである
  - ❖ 概念の定義はプレーンテキストであり計算機可読でない
  - ❖ Biological Process, Molecular Function, Cellular Componentの概念間の関係が定義されていない
    - ❖ 各概念はassociated-gene関係で遺伝子に関連つけられているのみ
    - ❖ 機能と構造の本質的な関係を明示化することがオントロジーに求められる(暗黙的概念の明示化)
- ❖ part-of関係の定義があいまい

part-of in GO ::= made-of | belongs-to | physical-part-of |  
conceptual-part-of | subprocess-of | controls | causes |  
activates | inhibits | enclosed-by | binds-to

# AmiGO

## Gene Product: IL13\_HUMAN

**Full Name:** Interleukin-13 precursor

**Synonyms:** IPI00291841

**Datasource:** UniProt

### Associated to Terms:

Evidence	Term	Aspect
TAS	<input type="checkbox"/> <a href="#">antimicrobial humoral response (sensu Vertebrata)</a>	biological_process
TAS	<input type="checkbox"/> <a href="#">cell motility</a>	biological_process
NR	<input type="checkbox"/> <a href="#">cell proliferation</a>	biological_process
NR	<input type="checkbox"/> <a href="#">cell-cell signaling</a>	biological_process
NR	<input type="checkbox"/> <a href="#">chemokine activity</a>	molecular_function
TAS	<input type="checkbox"/> <a href="#">cytokine activity</a>	molecular_function
TAS	<input type="checkbox"/> <a href="#">inflammatory response</a>	biological_process
NR	<input type="checkbox"/> <a href="#">signal transducer activity</a>	molecular_function
NR	<input type="checkbox"/> <a href="#">signal transduction</a>	biological_process
TAS	<input type="checkbox"/> <a href="#">soluble fraction</a>	cellular_component

inflammatoryの  
機能達成のメカ  
ニズム(ルール)

### Peptide Sequences

>UN  
MAL  
SMV  
SLH

遺伝子IL\_13\_HUMAN (associated-gene) chemokine activity [MF]  
遺伝子IL\_13\_HUMAN (associated-gene) inflammatory response [BP]  
遺伝子IL\_13\_HUMAN (associated-gene) soluble fraction [CC]

Check/Uncheck All

# Gene Ontologyを用いた自動アノテーション研究

- ❖ 文献や遺伝子発現データ等のOMICSデータを用いたアノテーション支援への期待
- ❖ これまでの方法では1/3の遺伝子は機能アノテーションが付かない？

# ゲノムのアノテーション

## ❖ DNA配列に生物学の情報を追加する

❖ 遺伝子の位置 (エクソン、イントロン)

❖ 繰り返し配列、CpG、遺伝子多型

## ❖ 機能

❖ 遺伝子の機能

❖ 遺伝子産物の機能

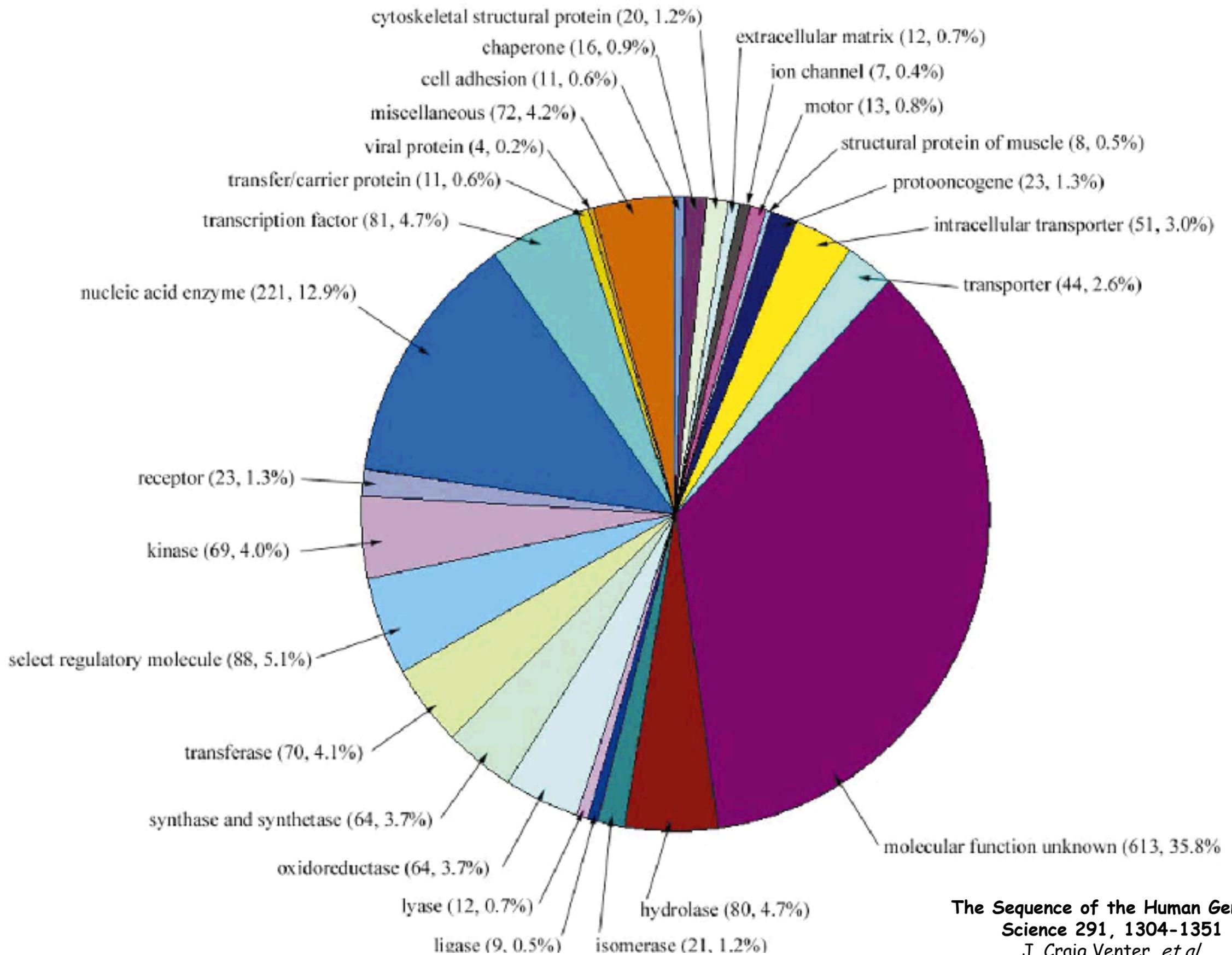
❖ 発現場所

❖ タンパク質の機能

❖ 二次代謝産物の機能

ゲノム配列データの  
解析から

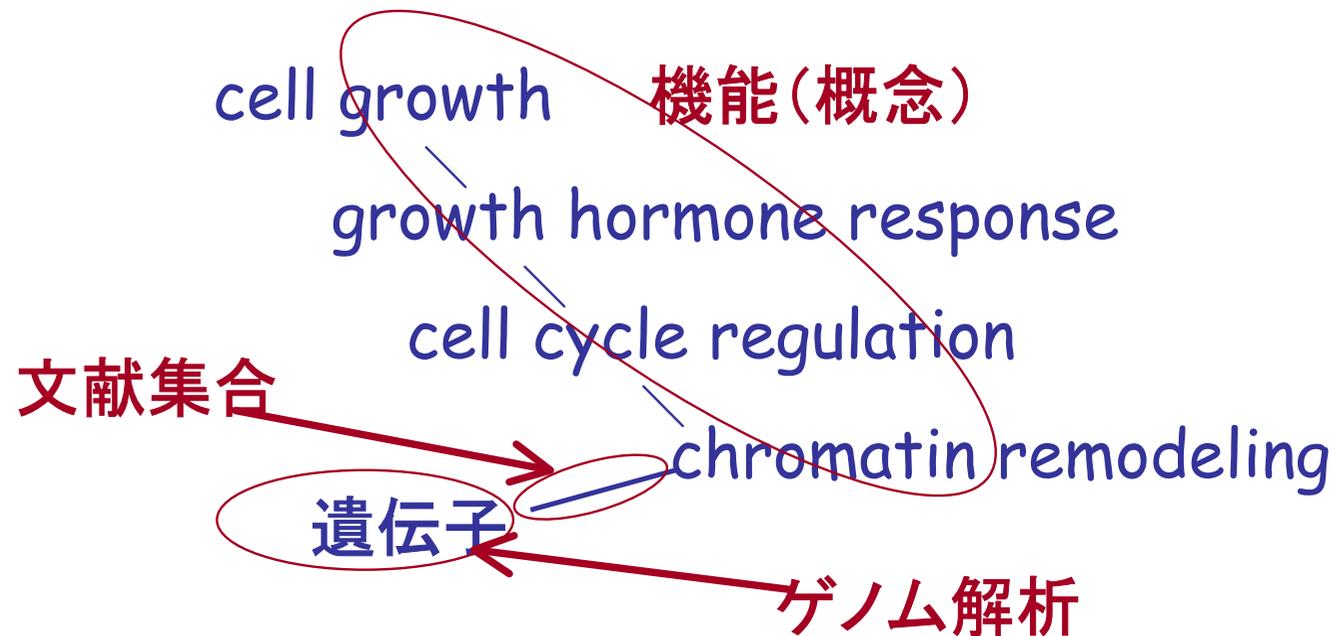
機能解析実験から



# Gene Ontologyを用いた自動アノテーション研究

- ❖ 文献や遺伝子発現データ等のOMICSデータを用いたアノテーション支援への期待
- ❖ これまでの方法では1/3の遺伝子は機能アノテーションが付かない？
- ❖ これまでの機能アノテーションは
  - ❖ 既存のデータベースに格納された情報の使い回し
- ❖ 文献や、網羅的実験データ(OMICSデータ)から機能情報を導出する計算機技術が必要

# GOを正確学習集合とする文献データのクラスタリング



Gene Ontology は 文献 - 機能(概念) - 遺伝子 を結んでいる!



Quick Search:

 Submit

[Site Map](#) | [Help](#) | [Full Search](#) | [Home](#)

[Community Info](#)   [Submit Data](#)   [BLAST](#)   [Primers](#)   [PatMatch](#)   [Gene/Seq Resources](#)   [Virtual Library](#)   [Contact SGD](#)

# Gene Ontology: Annotations

[Help](#)

[GO Tutorial](#)

## PGK1 GO ANNOTATIONS : [Function](#) | [Process](#) | [Component](#)

[PGK1 Locus Info](#)

### Function

Annotation(s)	Reference(s)	Evidence
<a href="#">phosphoglycerate kinase activity</a>	<p><b>Hitzeman RA, et al. (1980)</b> Isolation and characterization of the yeast 3-phosphoglycerokinase gene (PGK) by an immunological screening technique. <i>J Biol Chem</i> 255(24):12073-80</p> <p> </p>	<ul style="list-style-type: none"> <li>• IDA : <a href="#">Inferred from Direct Assay</a>Last updated on 2001-01-18</li> </ul>

### Process

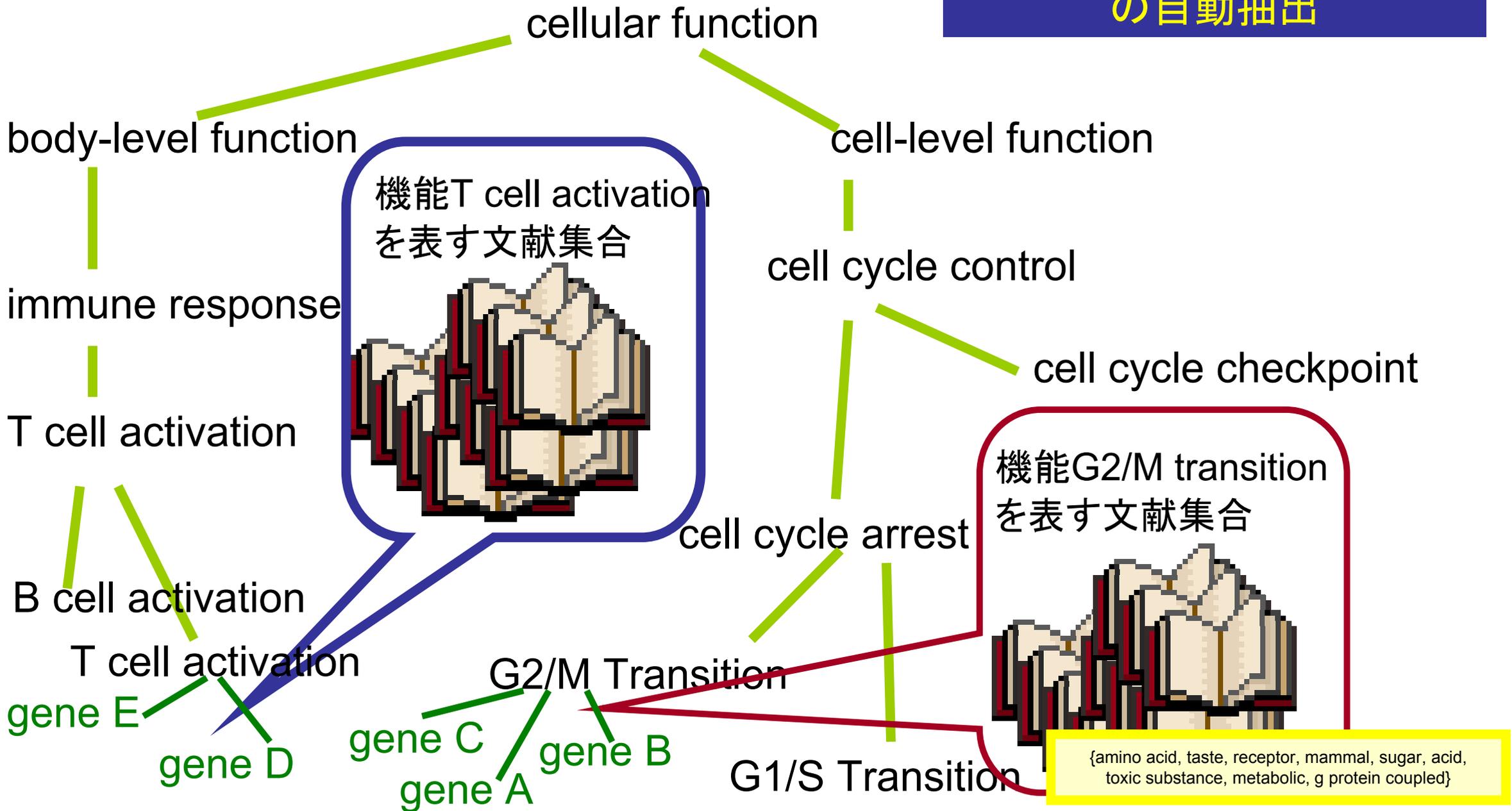
Annotation(s)	Reference(s)	Evidence
<a href="#">gluconeogenesis</a>	<p><b>Stryer L (1988)</b> Biochemistry (3rd ed.). New York: W. H. Freeman and Company</p> <p></p>	<ul style="list-style-type: none"> <li>• TAS : <a href="#">Traceable Author Statement</a>Last updated on 2001-01-19</li> </ul>
<a href="#">glycolysis</a>	<p><b>Stryer L (1988)</b> Biochemistry (3rd ed.). New York: W. H. Freeman and Company</p> <p></p>	<ul style="list-style-type: none"> <li>• TAS : <a href="#">Traceable Author Statement</a>Last updated on 2001-01-19</li> </ul>

### Component

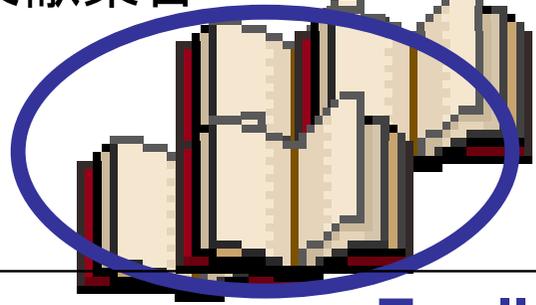
Annotation(s)	Reference(s)	Evidence
<a href="#">cytoplasm</a>	<p><b>Kumar A, et al. (2002)</b> Subcellular localization of the yeast proteome. <i>Genes Dev</i> 16(6):707-19</p> <p> </p>	<ul style="list-style-type: none"> <li>• IDA : <a href="#">Inferred from Direct Assay</a>Last updated on 2002-05-07</li> </ul>
<a href="#">cytosol</a>	<p><b>Alberts B, et al. (1994)</b> Molecular Biology of the Cell (3rd ed.). New York: Garland Publishing</p> <p></p>	<ul style="list-style-type: none"> <li>• TAS : <a href="#">Traceable Author Statement</a>Last updated on 2001-01-19</li> </ul>

# Gene Ontology は 文献 - 機能(概念) - 遺伝子 を結んでいる

GOと文献を用いた機能情報の自動抽出



機能未知の遺伝子Aを報告する文献集合



{amino acid, taste, receptor, mammal, sugar, acid, toxic substance, metabolic, g protein coupled}

G2/M transition

T cell activation

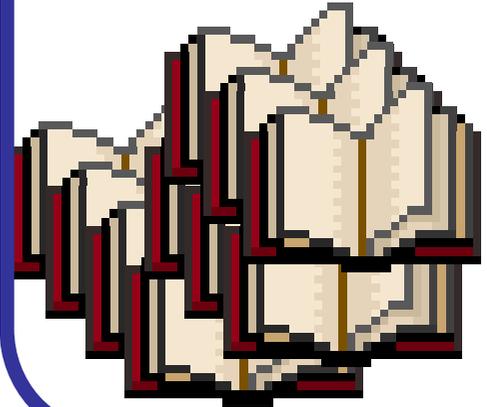
cellular function

GOと文献を用いた機能情報の自動抽出

body-level function

cell-level function

機能T cell activation  
を表す文献集合



cell cycle control

immune response

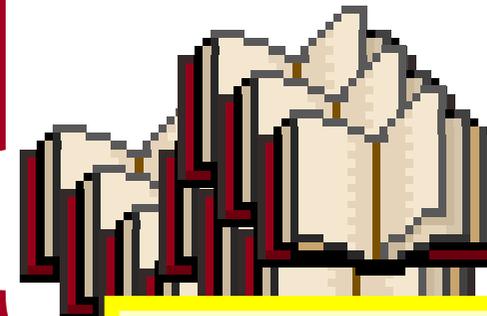
cell cycle checkpoint

T cell activation

cell cycle arrest

機能G2/M transition  
を表す文献集合

B cell activation



T cell activation

gene E

gene D

gene C

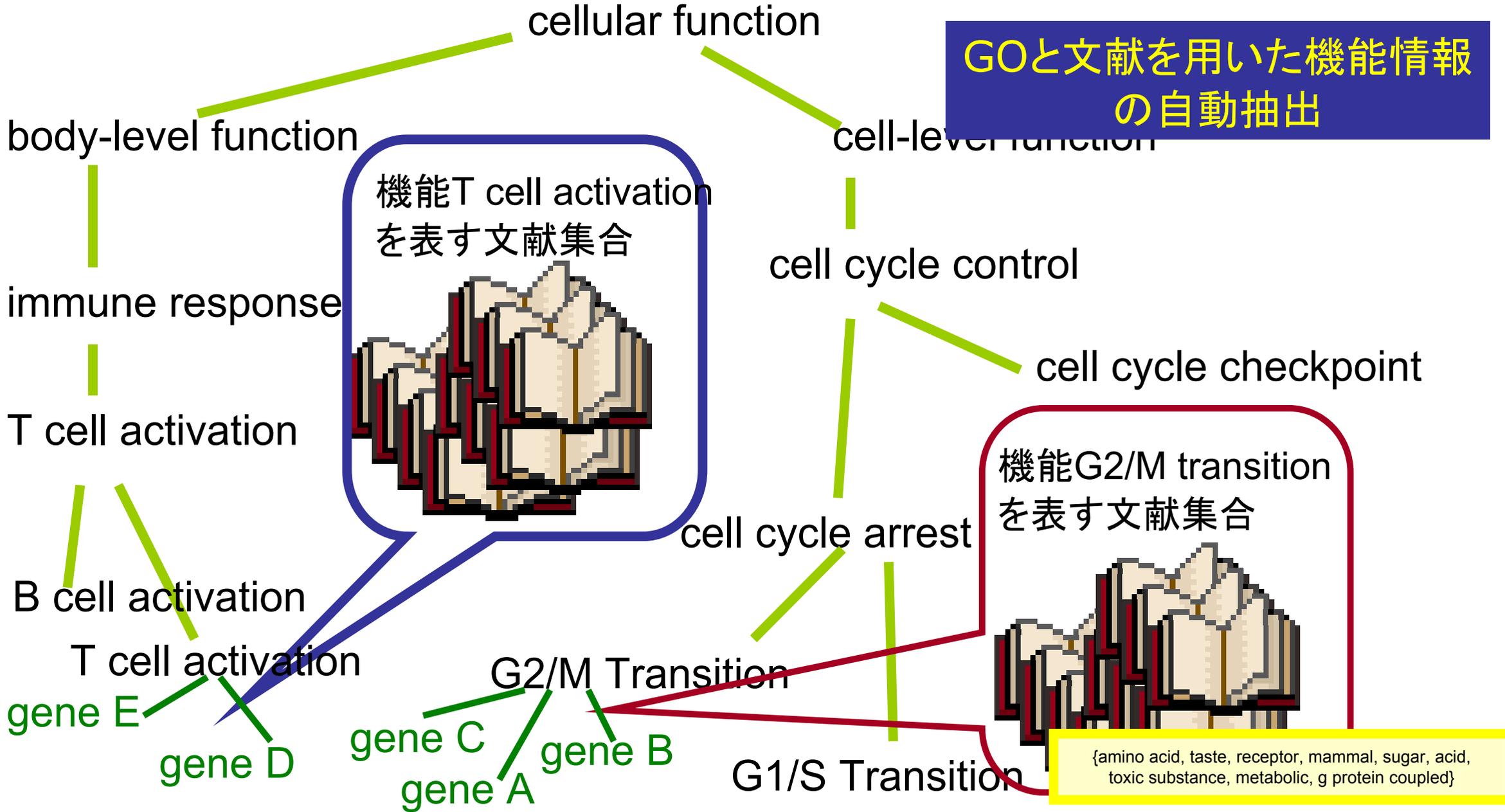
gene A

G2/M Transition

gene B

G1/S Transition

{amino acid, taste, receptor, mammal, sugar, acid, toxic substance, metabolic, g protein coupled}



# 文献の類似性を計算するには？

**Nature 2002 Mar 14;416(6877):199-202**

**An amino-acid taste receptor.**

The sense of **taste** provides animals with valuable information about the nature and quality of food. **Mammals** can recognize and respond to a diverse repertoire of chemical entities, including **sugars**, salts, **acids** and a wide range of **toxic substances**. This is noteworthy because L-**amino acids** function as the building blocks of proteins, as biosynthetic precursors of many biologically relevant small molecules, and as **metabolic** fuel. Thus, having a **taste** pathway dedicated to their detection probably had significant evolutionary implications. Here we identify and characterize a mammalian **amino-acid taste receptor**. This **receptor**, T1R1+3, is a heteromer of the **taste-specific** T1R1 and T1R3 **G-protein-coupled receptors**. We demonstrate that T1R1 and T1R3 combine to function as a broadly tuned L-**amino-acid** sensor responding to most of the 20 standard amino acids, but not to their D-enantiomers or other compounds. We also show that sequence differences in T1R receptors within and between species (human and mouse) can

**Cell. 2003 Oct 31;115(3):255-66.**

**The receptors for mammalian sweet and umami taste .**

The **taste** perception of monosodium glutamate (MSG) is termed 'umami'. Two putative **taste receptors** for **glutamate** have been identified, a truncated form of mGluR4 (taste-mGluR4) and the presumed heterodimer T1R1 + T1R3. Both **receptors** respond to **glutamate** when expressed in heterologous **cells**, but the **G protein** involved is not known. Galpha-Gustducin mediates the **transduction** of several **bitter** and **sweet** compounds; however, its role in umami has not been determined. We used standard two-bottle preference **tests** on alpha-gustducin knockout (KO) and wildtype (WT) mice to compare preferences for ascending concentrations of MSG and MSG + 5'-inosine monophosphate (IMP). A Latin Square was used to assign the order of tastants presented to each **mouse**. Statistical comparisons between KO and WT **mice** revealed that whereas WT mice preferred solutions of MSG and MSG + IMP over water, KO **mice** showed little preference for these **stimuli**. Denatonium and sucrose served as control **stimuli** and, as shown previously, WT mice preferred **sucrose** and avoided denatonium significantly more than did KO mice. Naive mice were also **tested**, and while prior exposure to **taste stimuli** influenced the magnitude of the preferences, experience did not change the overall pattern of intake. These data suggest that alpha-gustducin plays a role in **glutamate taste**.

# 文献の類似性を計算するには？

Nature 2002 Mar 14;416(6877):199-202

An **amino-acid taste receptor**.

The sense of **taste** provides animals with valuable information about the nature and quality of food.

{amino acid, taste, receptor, mammal, sugar, acid, toxic substance, metabolic, g protein coupled}

**receptor**, T1R1+3, is a heteromer of the **taste**-specific T1R1 and T1R3 **G-protein-coupled receptors**. We demonstrate that T1R1 and T1R3 combine to function as a broadly tuned L-**amino-acid** sensor responding to most of the 20 standard amino acids, but not to their D-enantiomers or other compounds. We also show that sequence differences in T1R receptors within and between species (human and mouse) can

Cell. 2

## 特徴語集合の類似性に置き換えて計算

The **taste** perception of monosodium glutamate (MSG) is termed 'umami'. Two putative **taste receptors** for **glutamate** have been identified, a truncated form of mGluR4 (taste-mGluR4) and the presumed heterodimer T1R1 + T1R3. Both **receptors** respond to **glutamate** when expressed in heterologous **cells**, but the **G protein** involved is not known. Galpha-Gustducin mediates the **transduction** of several **bitter** and **sweet** compounds; however, its role in umami has not been determined. We used standard two-bottle preference **tests** on alpha-gustducin knockout (KO) and wildtype (WT) mice to compare preferences for ascending concentrations of MSG and MSG + 5' inosine

{receptor, taste, glutamate, cell, g protein, sucrose, transduction, bitter, sweet, mouse, stimulus}

preferences, experience did not change the overall pattern of intake. These data suggest that alpha-gustducin plays a role in **glutamate taste**.

## 機能と発現量の相関解析

- ❖ GOアノテーションありの284遺伝子を機械学習し、機能と発現量プロフィールの相関ルールを取り出す
- ❖ ルールを用いてGOアノテーションなしの遺伝子を機能に関連つける
  
- ❖ 機械学習の精度を保証するために、各GO概念に少なくとも4遺伝子が付いていることが必要
- ❖ GOのis-a階層に基づき23の上位概念にまとめた

# 機能と発現量の相関解析

## Predicting Gene Ontology Biological Process From Temporal Gene Expression Patterns

Astrid Lægreid, Torgeir R. Hvidsten, Herman Midelfart, Jan  
Komorowski and Arne K. Sandvik

Genome Research, Vol 13, 965-979, 2003.

Ion homeostasis
Protein targeting
Blood coagulation
DNA metabolism
Intracellular signaling cascade
Energy pathways
Cell cycle
Oncogenesis
Circulation
Cell death
Developmental processes
Transcription
Defense (immune) response
Cell adhesion
Stress response
Protein metabolism and modification
Cell motility
Cell surface receptor linked signal transduction
Lipid metabolism
Transport
Cell organization and biogenesis
Cell proliferation
Amino acid and derivative metabolism

## Search GO

 Exact Match

 Terms

 Gene Products

## Search Filters

### Species

All	▲
A. aeolicus	▬
A. fulgidus	▼

### Datasource

All	▲
FlyBase	▬
SGD	▼

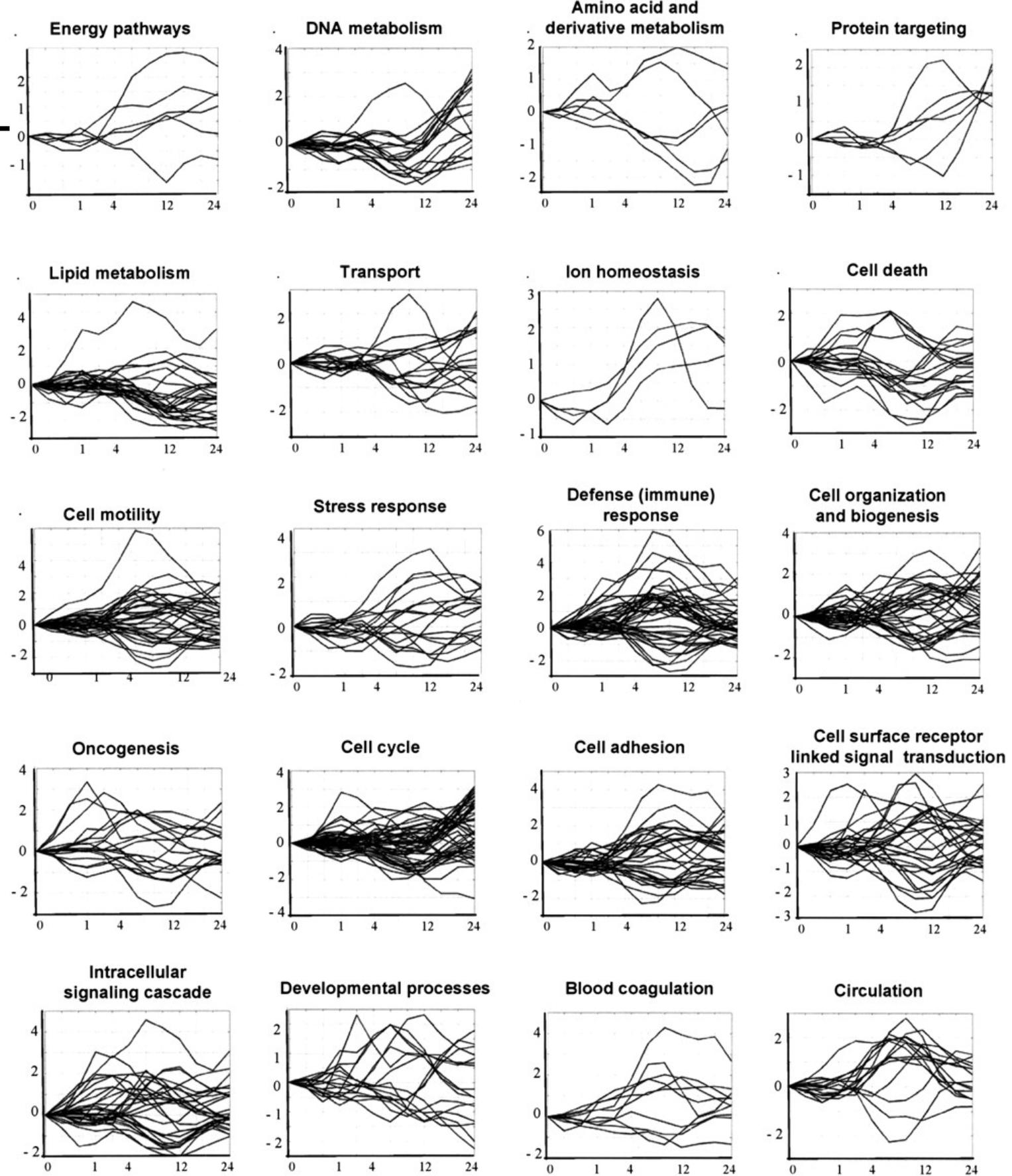
### Evidence Code

All Curator Approved	▲
IMP	▬
IGI	▼

[Advanced Query](#)
[XML](#)
[Flat File](#)
[Permalink](#)

- [-] **GO:0003673 : Gene\_Ontology ( 146200 )**
  - [-] **GO:0008150 : biological\_process ( 96312 )**
    - [+] **GO:0007610 : behavior ( 2293 )**
    - . **GO:0000004 : biological\_process unknown ( 26924 )**
    - [-] **GO:0009987 : cellular process ( 31905 )**
      - [-] **GO:0007154 : cell communication ( 8954 )**
        - [-] **GO:0007155 : cell adhesion ( 1135 )**
          - . **GO:0042710 : biofilm formation ( 0 )**
          - [-] **GO:0008037 : cell recognition ( 89 )**
            - . **GO:0009988 : cell-cell recognition ( 19 )**
            - [-] **GO:0009989 : cell-matrix recognition ( 0 )**
              - . **GO:0009990 : contact guidance ( 0 )**
            - [+] **GO:0008038 : neuronal cell recognition ( 47 )**
          - [+] **GO:0016337 : cell-cell adhesion ( 421 )**
          - [+] **GO:0007160 : cell-matrix adhesion ( 129 )**
          - [+] **GO:0000128 : flocculation ( 17 )**
          - [+] **GO:0030155 : regulation of cell adhesion ( 67 )**
            - . **GO:0030260 : cell invasion ( 25 )**
          - [+] **GO:0007267 : cell-cell signaling ( 1384 )**
          - [+] **GO:0030383 : host-pathogen interaction ( 269 )**
          - . **GO:0001759 : induction of an organ ( 4 )**
          - [+] **GO:0009875 : pollen-pistil interaction ( 7 )**
          - [+] **GO:0009991 : response to extracellular stimulus ( 141 )**
          - [+] **GO:0007165 : signal transduction ( 6780 )**
        - [+] **GO:0030154 : cell differentiation ( 2184 )**
        - [+] **GO:0050875 : cellular physiological process ( 25201 )**
        - [+] **GO:0006944 : membrane fusion ( 297 )**
        - [+] **GO:0050794 : regulation of cellular process ( 1955 )**
      - [+] **GO:0007275 : development ( 14496 )**
      - [+] **GO:0008371 : obsolete biological process ( 90 )**
      - [+] **GO:0007582 : physiological process ( 60310 )**
      - [+] **GO:0050789 : regulation of biological process ( 2533 )**

# 各機能ごとの 発現量プロフィール



## 機能と発現量の相関解析

- ❖ GOの23機能について相関ルールを抽出
- ❖ (変数を減らすために)プロフィールを3-4タイムポイント区間の"増加", "減少", "一定"の組み合わせで表現
- ❖ ラフ集合論を用いたルール抽出

## 相関ルール(association rule)

**A, B, C, ...** をあるデータベースの離散値属性とするとき

$$(A = a) \Rightarrow (C = c)$$

$$(A = a) \text{ AND } (B = b) \Rightarrow (C = c)$$

.....

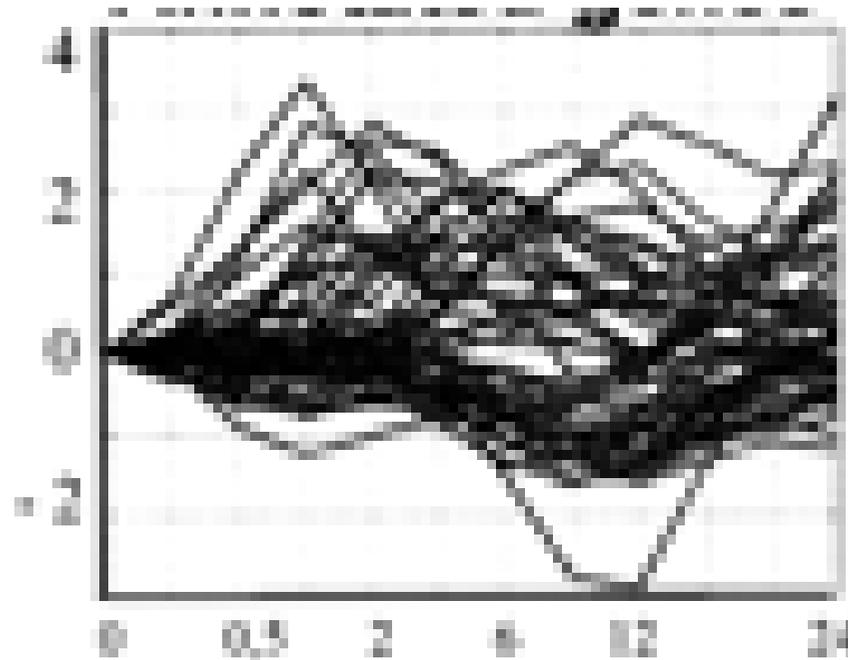
という形をした規則

# transcriptionの相関ルール

30MIN – 4H(Constant) AND 1H – 8H(*Decreasing*) AND **16H – 24H(Increasing)** → Process(transcription)

0MIN – 1H(Constant) AND **30MIN – 4H(Increasing)** AND 8H – 16H(*Decreasing*) AND 16H – 24H(*Decreasing*) → Process(transcription)

15MIN – 1H(*Decreasing*) AND 30MIN – 4H(Constant) AND **6H – 24H(Increasing)** → Process(transcription)



Process	Annotate genes	Reclassifications for known genes (correct)	Classifications for unknown genes
Ion homeostasis	4	4 (4)	4
Protein targeting	6	8 (6)	2
Blood coagulation	10	13 (10)	34
DNA metabolism	19	33 (19)	28
Intracellular signaling cascade	26	39 (25)	14
Energy pathways	6	6 (6)	0
Cell cycle	47	66 (44)	38
Oncogenesis	17	41 (17)	31
Circulation	15	17 (15)	3
Cell death	16	22 (16)	26
Developmental processes	15	23 (15)	21
Transcription	52	61 (47)	88
Defense (immune) response	37	44 (32)	20
Cell adhesion	30	32 (29)	19
Stress response	17	26 (17)	22
Protein metabolism and modification	33	45 (33)	35
Cell motility	32	40 (29)	21
Cell surface receptor linked signal transduction	27	38 (26)	34
Lipid metabolism	26	44 (25)	45
Transport	22	30 (22)	19
Cell organization and biogenesis	33	40 (30)	13
Cell proliferation	53	60 (46)	30
Amino acid and derivative metabolism	6	6 (6)	

Symbol (GenBank accession number)	Gene name	Molecular function	Comment	Reference (PMID)
<i>CCNG1</i> ( <a href="#">R45687</a> )	Cyclin G1	CDK kinase regulator	p53 target	11327114
<i>CDKN1C</i> ( <a href="#">R81336</a> )	Cyclin-dependent kinase inhibitor 1C	Cyclin-dependent protein kinase inhibitor	Tumor suppressor	7729684
<i>CAT</i> ( <a href="#">W89002</a> )	Catalase	Oxidoreductase	Tumor progression	8513880, 11597785
<i>ALDH3A2</i> ( <a href="#">H63779</a> )	Aldehyde dehydrogenase 10	Aldehyde dehydrogenase	Tumor progression	92393980
<i>ADD3</i> ( <a href="#">AA054129</a> )	Adducin 3 (gamma)	Membrane-cytoskeleton-associated protein	Tumor progression	9607561
<i>TFDP2</i> ( <a href="#">W46792</a> )	Transcription factor Dp-2 (E2F dimerization partner 2)	Transcription cofactor	Cell cycle regulation	7784053
<i>ATRX</i> ( <a href="#">N22858</a> )	Thalassemia/mental retardation syndrome	DNA helicase	Transcription and DNA repair	10362365, 10630641
<i>EPS15</i> ( <a href="#">N78949</a> )	Epidermal growth factor receptor pathway substrate 15	Kinase substrate	Growth regulation	93361014
<i>EGR1</i> ( <a href="#">H27557</a> )	Early growth response 1	Transcription factor	Tumor suppressor	9109500
<i>NR4A2</i> ( <a href="#">N22386</a> )	Nuclear receptor subfam 4, group A, m2 (Nurr1, Not)	Ligand-dependent nuclear receptor	Proto-oncogene	9592180
<i>NR4A3</i> ( <a href="#">W42606</a> )	Nuclear receptor subfam 4, group A, m 3 (Nor1)	Ligand-dependent nuclear receptor	Proto-oncogene	9592180

**False Positive for *oncogenesis*: Missing Annotation**

Gene symbol	Gene name	GenBank accession number	Predictions for the 23 broad cellular processes used for learning
<i>LOC80298</i>	Transcription termination factor-like protein	<a href="#">W95909</a>	Cell surface receptor-linked signal transduction, cell cycle
<i>EST</i> , unclustered		<a href="#">AA044605</a>	Protein metabolism and modification, developmental processes
	ESTs	<a href="#">AA059077</a>	Cell proliferation, protein metabolism and modification, cell motility, developmental processes
<i>EST</i> , unclustered		<a href="#">AA035657</a>	Protein metabolism and modification, developmental processes
<i>KIAA0455</i>	<i>KIAA0455</i> gene product	<a href="#">H19324</a>	Cell death, blood coagulation
	<i>Homo sapiens</i> cDNA FLJ13545 fis, clone PLACE1006867	<a href="#">W89018</a>	Lipid metabolism
<i>KIAA1391</i>	<i>KIAA1391</i> protein	<a href="#">H28360</a>	Cell surface receptor-linked signal transduction, transcription
	ESTs	<a href="#">H16592</a>	Cell motility, cell death
	ESTs	<a href="#">W78151</a>	Cell death, blood coagulation
<i>EST</i> , unclustered		<a href="#">H14500</a>	Cell death
	<i>Homo sapiens</i> clone 23645 mRNA sequence	<a href="#">N75026</a>	Cell death, blood coagulation
	<i>Homo sapiens</i> cDNA: FLJ21482 fis, clone COL05135	<a href="#">R87731</a>	Cell death, blood coagulation
<i>EST</i> , unclustered		<a href="#">H61274</a>	Cell death, blood coagulation
	<i>Homo sapiens</i> mRNA; cDNA DKFZp761K2024 (from clone DKFZp761K2024)	<a href="#">N63445</a>	Cell death, blood coagulation
	<i>Homo sapiens</i> mRNA; cDNA DKFZp564L0822 (from clone DKFZp564L0822)	<a href="#">W69445</a>	Cell surface receptor-linked signal transduction, cell death, blood coagulation
	<i>Homo sapiens</i> mRNA; cDNA DKFZp586I1823 (from clone DKFZp586I1823)	R60336, H15535	Lipid metabolism, transcription
<i>KIAA1628</i>	<i>KIAA1628</i> protein	<a href="#">N53427</a>	Lipid metabolism transcription
	ESTs	<a href="#">R60731</a>	Lipid metabolism transcription
<i>FLJ20643</i>	Hypothetical protein FLJ20643	<a href="#">AA018444</a>	Cell death, blood coagulation, oncogenesis
<i>KIAA0993</i>	<i>KIAA0993</i> protein	<a href="#">AA031778</a>	Cell surface receptor-linked signal transduction, transcription, oncogenesis

## Classification of Uncharacterized Genes

# 機能と発現量の相関解析

- ❖ 1/3は機能アノテーションできない問題を克服
- ❖ Gene Ontologyはタキソノミーとして利用
- ❖ 類似性

- ❖ 実験データの特徴を記号化
- ❖ 機械学習でルールを抽出

計算可能

機能を計算可能

- ◆ ベイズ推定(最大エントロピー法)
- ◆ ニューラルネットワーク
- ◆ 決定木
- ◆ Support Vector Machine
- ◆ .....

# 機能と発現量の相関解析

- ❖ 1/3は機能アノテーションできない問題を克服
- ❖ Gene Ontologyはタキソノミーとして利用
- ❖ 類似性

- ❖ 実験データの特徴を記号化
- ❖ 機械学習でルールを抽出

計算可能

機能を計算可能

- ◆ ベイズ推定(最大エントロピー法)
- ◆ ニューラルネットワーク
- ◆ 決定木
- ◆ Support Vector Machine

機能以外の生物概念のオントロジー(タキソノミー)開発

ルール知識の利用方法の開発へ

- 細胞内の働き
  - 代謝系
    - Carbohydrate Metabolism
    - TCA cycle
    - Glycolysis
  - Energy Metabolism
  - .....
  - シグナル伝達系
    - Receptor Tyrosine Pathway
    - EGF Pathway
    - PDGF Pathway
    - GPCR Pathway
  - .....
  - 細胞骨格系
  - 輸送・分泌系
  - 転写翻訳系
  - タンパク質分解系

## タキソノミー

ルール知識の利用方法の開発へ

- 細胞内の働き(世界)
  - 種類
    - 代謝系
    - シグナル伝達系
    - 細胞骨格系
    - 輸送・分泌系
    - 転写翻訳系
    - タンパク質分解系
  - 機能
    - 変換する
      - 化合物の種類
      - 反応を起こす
    - 移動する
  - 属性
    - 構造の種類
    - 力の種類
  - 構成物
    - 対象物
    - 装置
    - 媒体

Why EGF  
Pathway is a  
Cell Signaling  
Pathway?

オントロジー

# 人工知能研究 Artificial Intelligence

- ❖ 人工的な知能を作ることを目標とする
- ❖ ある問題に対して知的に振る舞うシステムを作るための理論と技術に関する研究
- ❖ 知能を計算可能とする研究
- ❖ 形式指向研究と内容指向研究
  - ❖ 形式: 知識をどのように記述し(計算)処理するか
  - ❖ 内容: そもそも知識とは何か?

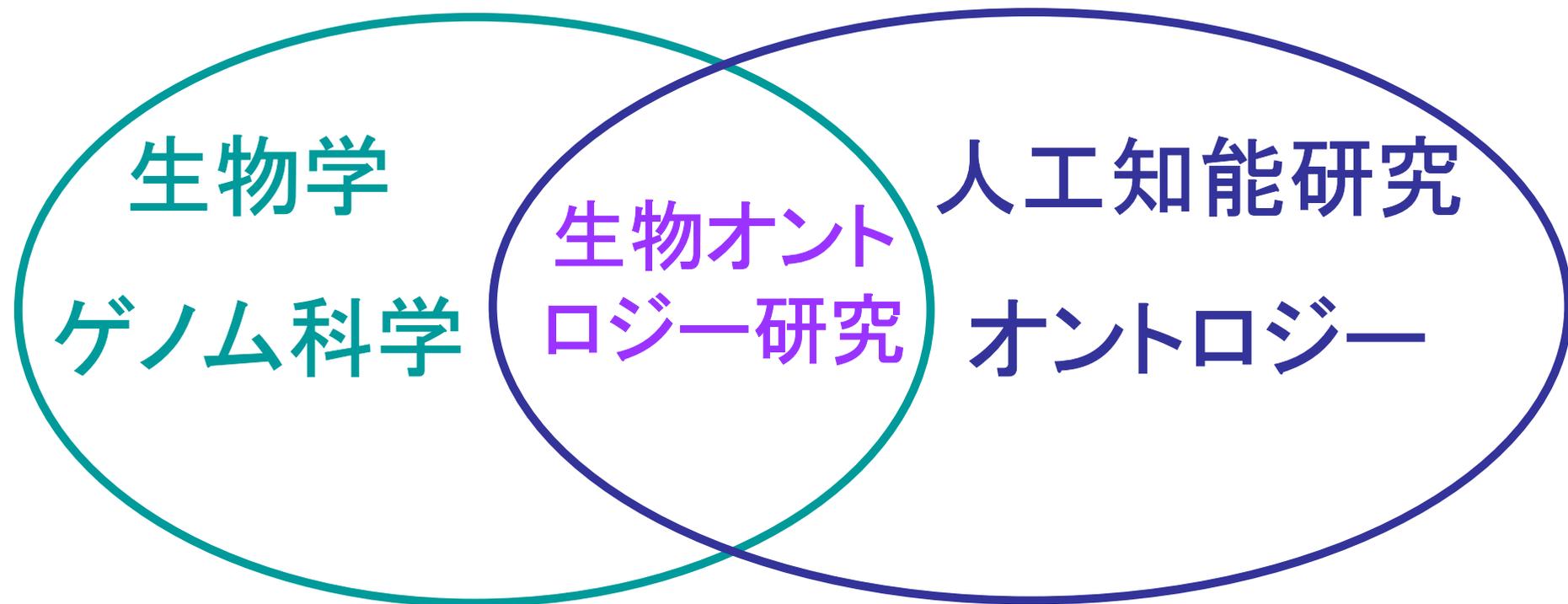
# 人工知能研究 Artificial Intelligence

## ❖ 内容指向研究

- ❖ 与えられた問題を解決するためには、どのような知識を用意すれば良いか？
- ❖ そもそも知識とは何か？
- ❖ 我々は認識をどのように構成し、知識の具体的内容はどのような構造と性質を持っているのか？

# オントロジーと生物学ドメイン

- ❖ 生物学ドメインの概念の特徴
  - ❖ 学問であるため、宣言性、共有性が高い
  - ❖ ゲノム解析の進展により、ドメインのエレメントが掌握されている





- ❖ 生物学ド
- ❖ 学問
- ❖ ゲノム
- トが重

イン

が高い

のエレメン

生物学

ゲノム科学

生物オント  
ロジー研究

人工知能研究

オントロジー

## 参考になる文献

- ❖ 溝口, オントロジーと知識処理, Bit, 1999.
- ❖ 溝口, 池田, 来村, オントロジー工学基礎論 –意味リンク, クラス, 関係, ロールのオントロジー的意味論–, 人工知能学会誌, Vol.14, pp.1019-1032, 1999
- ❖ 高井, 生物学の機能概念オントロジー, 人工知能学会誌, Vol.19, pp.137-143, 2003
- ❖ 武田, 上位オントロジー, 人工知能学会誌, Vol.19, pp.172-178, 2003
- ❖ 岡田 光弘. オントロジーの哲学的・論理学的背景. 人工知能学会誌 2002;17:224-231.
- ❖ 長尾 真. 知識と推論. 岩波書店. 1998.

## 参考になる文献

- ❖ S. Schulze-Kremer, Ontologies for Molecular Biology and Bioinformatics, In Silico Biology, 173-177, 2002
- ❖ P. Karp, Science, 2001, vol 293, pp.2040.
- ❖ MV Blagosklonny and AB Pardee, Unearthing the gems -Conceptual Biology-, Nature, 416, 373 (2002)
- ❖ The Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Res 2001;11(8):1425-33.
- ❖ The Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32: D258-D261.
- ❖ [www.geneontology.org](http://www.geneontology.org)