

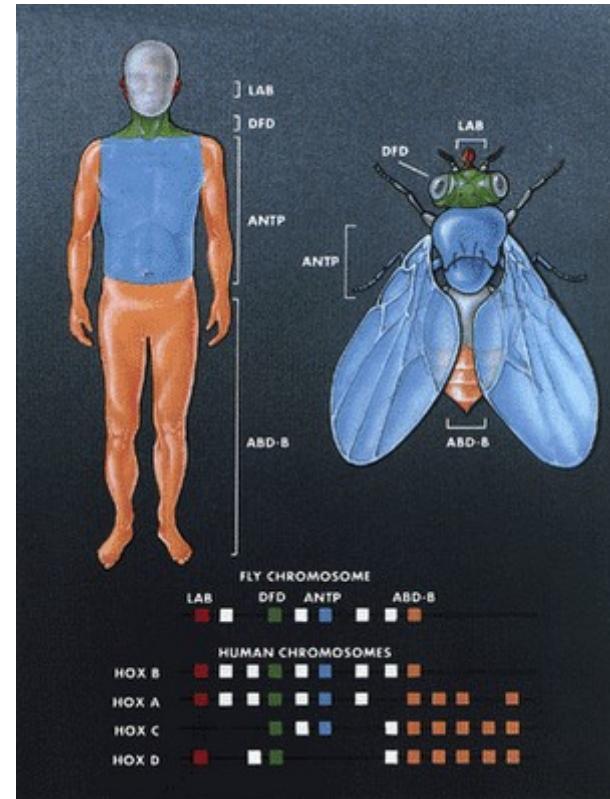
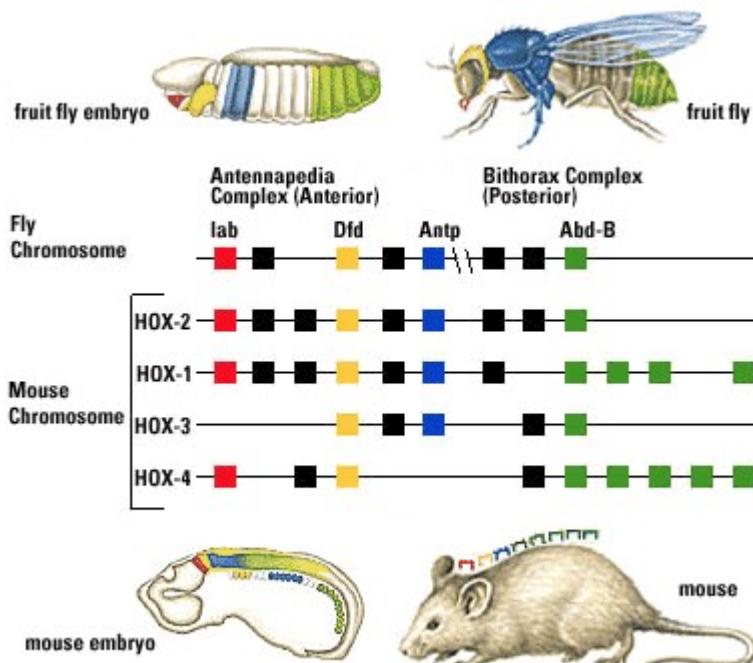
WBO 9

Rodziny genów
i określanie funkcji genów

30. kwietnia 2019

Bartek Wilczyński

Wiele genów zachowuje podobną sekwencję i funkcję poprzez miliony lat ewolucji



Jak powstały rodziny genów ?

- Nie znamy dokładnie początków życia, w szczególności mamy różne hipotezy nt. Życia przed powstaniem komórek
- Wiemy, że najprostsze organizmy komórkowe mają około tysiąca genów (*Mycoplasma genitalium*), a niektóre pasożyty mogą funkcjonować mając nawet około 500 genów
- Podejrzewa się, że pierwsze formy komórkowe prawdopodobnie stanowią pra-przodka całego żywego świata jaki znamy
- Jakie procesy stoją za powstaniem organizmów o dużo większej liczbie genów?

Duplikacja i specjacja

- W procesie replikacji chromosomów, duże fragmenty mogą zostać zduplikowane w tej samej komórce
- Następnie procesy mutacji następują niezależnie w różnych kopiach
- Jeśli duplikacja genu nie spowodowała negatywnych efektów ubocznych, presja selekcyjna na zduplikowane geny jest słabsza

Specjalizacja po duplikacji

- Jeśli wystarczy aby tylko jedna z kopii zduplikowanego genu zachowała funkcje przodka, drugi gen może nabywać nowych funkcji w procesie mutacji
- Jeśli mutacje nie doprowadzą do powstania użytecznej funkcji, często gen przestaje być używany i pozostaje **pseudogenem**
- Jeśli nowy gen zyskuje nową funkcję, może przyczynić się do procesu specjacji – powstania nowego gatunku z podpopulacji posiadającej zduplikowany gen

Zachowanie genów duplikowanych

- Selekcja osobników przebiega na podstawie fenotypu, a nie genotypu, zaś posiadanie dwóch kopii tego samego genu często nie wpływa bezpośrednio na fenotyp
- Chyba, że mamy do czynienia z genem, który musi mieć ścisłe regulowaną liczbę produktów (transkryptów, białek), bo zwiększenie liczby kopii DNA genu prowadzi często do zwiększonej ekspresji genu

Homologi i ich rodzaje

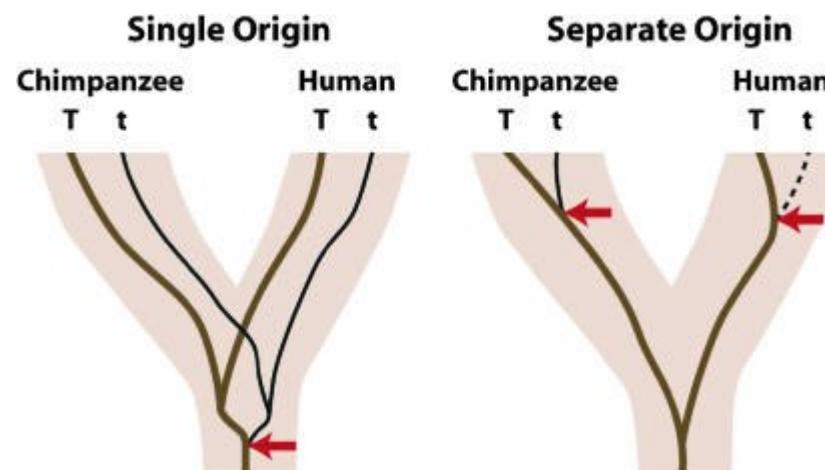
- **Geny pochodzące od wspólnego przodka nazywamy homologami.** Teoretycznie *jest to relacja równoważności*, choć nie mamy możliwości weryfikacji homologii ze 100% pewnością
- W praktyce, za homologi uznaje się geny o dużym (**arbitralnie**) podobieństwie sekwencyjnym, względem modeli losowych ewolucji sekwencji. *To nie jest relacja równoważności*, ze względu na brak przechodniości

Rodzaje homologów

- Pary homologów, pochodzące z duplikacji, nazywamy **paralogami**
- Pary homologów pochodzące ze specjacji, nazywamy **ortologami**
- Są jeszcze ksenologi – pochodzące z horyzontalnego transferu genów i ohnologi – pochodzące z duplikacji całych genomów, ale te rozważamy rzadziej.

Kolejność wydarzeń ewolucyjnych ma znaczenie

- Mamy zupełnie różne oczekiwania wobec paralogów jeśli chodzi o ich funkcje:
 - Ortologi powinny mieć niemal niezmienioną funkcję
 - Paralogi mogą dość istotnie odbiegać w swym działaniu

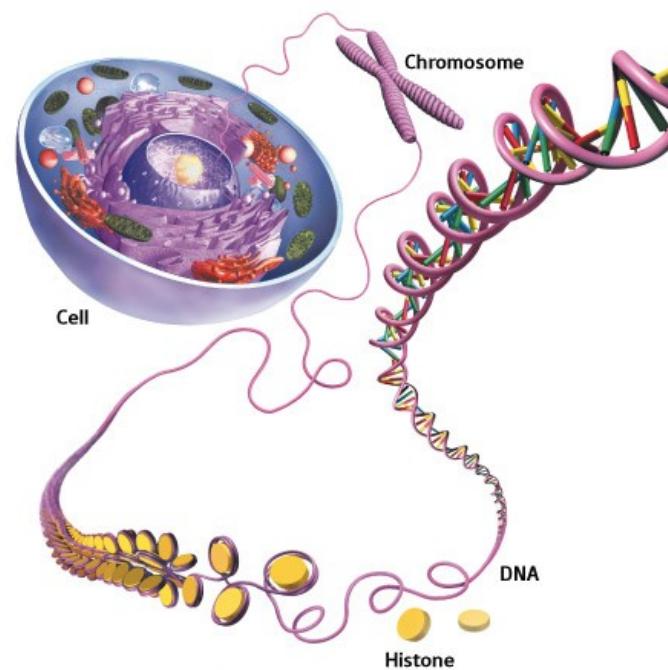


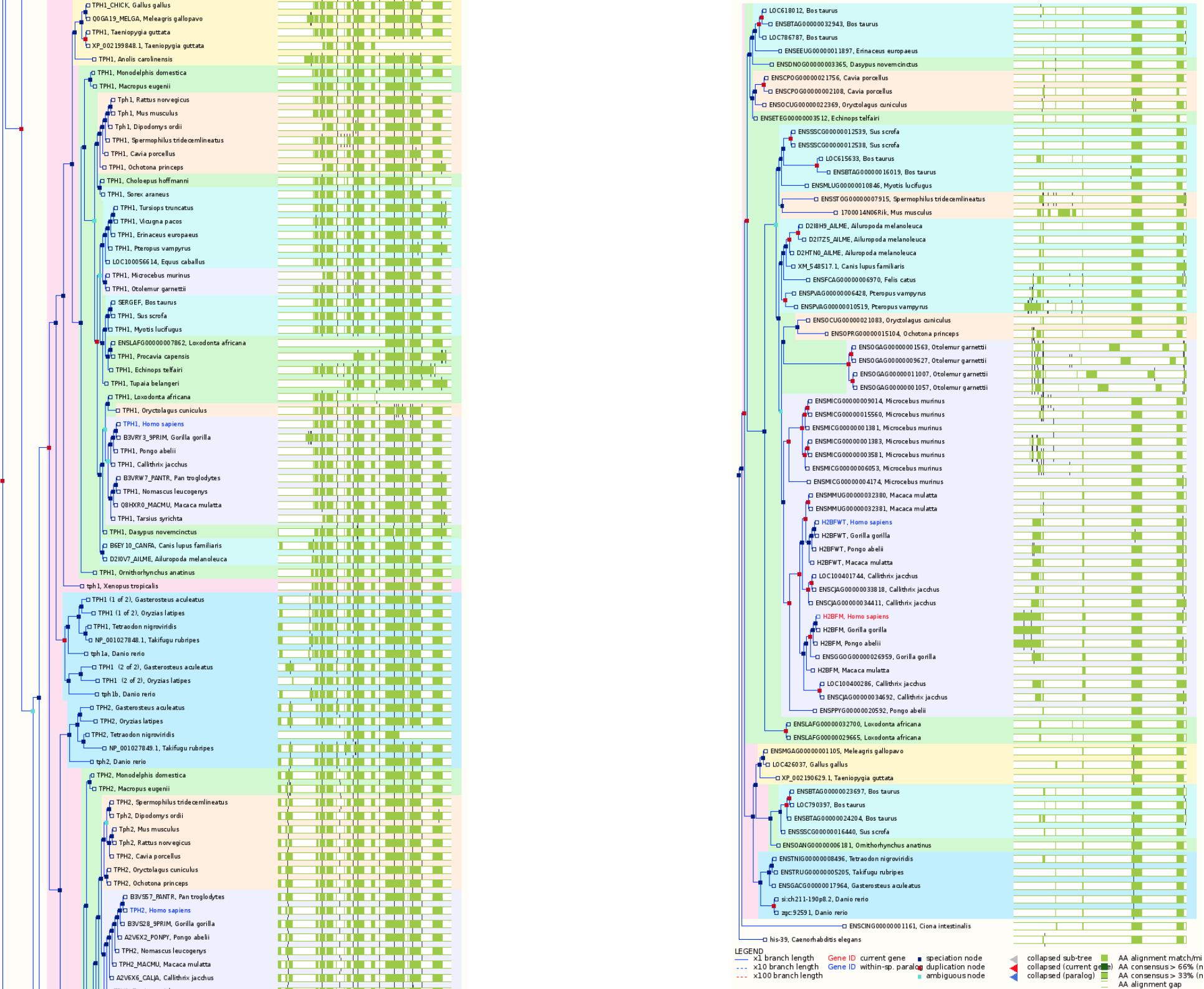
Przykład PAH i TPH

- PAH – hydroksylaza fenyloalaniny, to znany enzym, który występuje w pojedynczej kopii w niemal wszystkich zwierząt
- Mutacje w tym genie prowadzą do recesywnej choroby genetycznej – fenyloketonurii
- Paralog PAH – TPH – hydroksylaza tryptofanu, występuje w bardzo różnej liczbie kopii u różnych organizmów

Przykład 2. histony

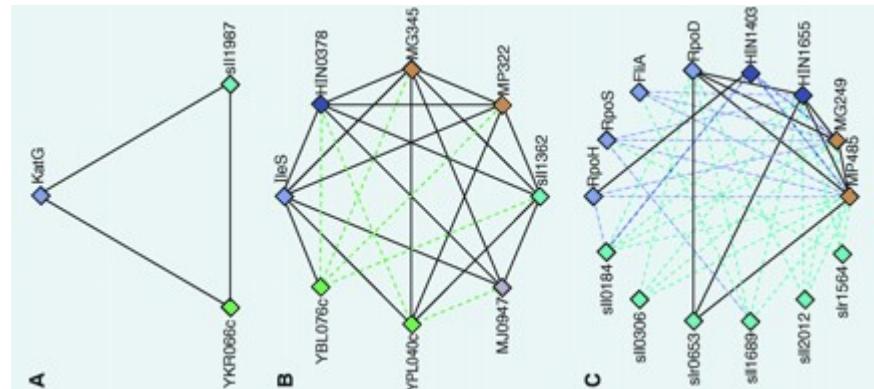
- Histony, które służą jako element strukturalny w komórkach eukariotypycznych, mają zupełnie inne wymagania – wiele niezależnych kopii w genomie jest potrzebnych dla produkcji milionów białek





Wykorzystanie BLASTa do znajdowania homologów

- BLAST Best hits – białka, które są dla siebie wzajemnie „trafieniami” BLASTa
- Możemy skonstruować graf, gdzie wierzchołkami są wszystkie geny, a krawędziami trafienia BLASTem
- Szukając „prawie-klik” w tym grafie możemy skonstruować COGs – klastry genów ortologicznych



Jak określić funkcje genów?

- Większość białek ma więcej niż jedną „funkcję”, i nie jest łatwe ich testowanie
- Zwykle jest to wykonywane eksperymentalnie, i bardzo czasochłonne
- Dla większości białek nadal nie wiemy jaka jest ich funkcja
- Od lat trwają wysiłki w systematyzacji tej wiedzy

Różne podejścia do opisu funkcji

- Gene Ontology – baza danych tworzona ręcznie przez wielu kuratorów, składająca się z kontrolowanego słownika funkcji w postaci DAGu (grafu skierowanego acyklicznego) i przypisań do niego
- Wikigenes – oparte na metodach text miningu automatycznie tworzone wiki na podstawie tekstów naukowych



the Gene Ontology

Search Browse BLAST Homolog Annotations Tools & Resources Help

Search GO

terms genes or proteins exact match

PAH

Phenylalanine-4-hydroxylase

protein from *Homo sapiens* (human)

[Term associations](#) ▾ [Gene product information](#) ▾ [Peptide Sequence](#) ▾ [Sequence information](#) ▾

Term Associations

Download all association information in:

▼ Filter associations displayed ?

Filter Associations

Ontology

All

biological process

cellular component

molecular function

Evidence Code

All

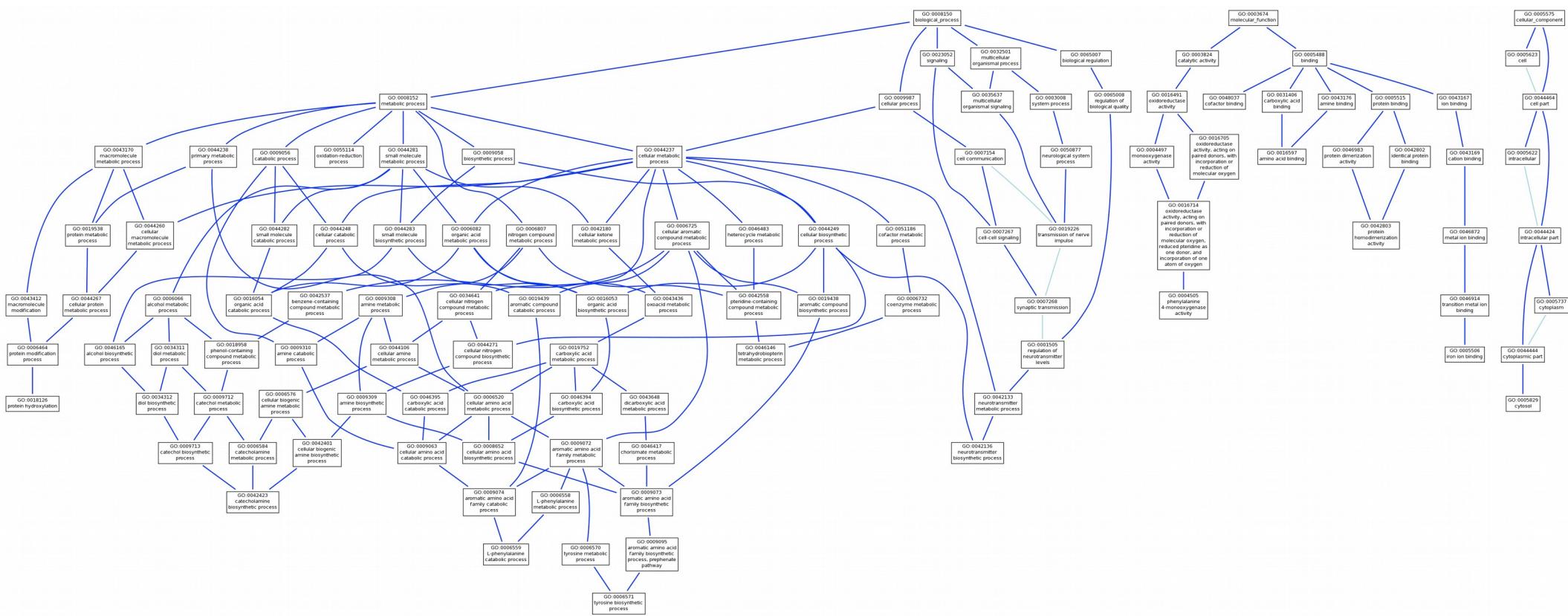
IC

IDA

IEA

Perform an action with this page's selected terms...

	Accession, Term		Ontology	Qualifier	Evidence
<input type="checkbox"/>	GO:0042423 : catecholamine biosynthetic process	71 gene products <input type="button" value="view in tree"/>	biological process		NAS
<input type="checkbox"/>	GO:0008652 : cellular amino acid biosynthetic process	6303 gene products <input type="button" value="view in tree"/>	biological process		TAS



[search][advanced]

Editor

[Edit this page](#)

[Discussion](#)

[History](#)

[Permanent link](#)

[Print](#)

[Export](#)^{new}

Share

[Send to a friend](#)

[Share](#)

Personal info

[Your homepage](#)

[Your articles](#)

[Create article](#)

View

[Font size](#)

[Color scheme](#)

Gene Review

[\[edit this page\]](#)

PAH - phenylalanine hydroxylase

Synonyms: PH, PKU, PKU1, Phe-4-monoxygenase, Phenylalanine-4-hydroxylase

Homo sapiens

[Green, E.K.](#), et al., [Chao, H.M.](#), et al., [Maass, A.](#), et al., [Teigen, K.](#), et al., [Ramus, S.J.](#), et al., et al.

Welcome! If you are familiar with the subject of this article, you can contribute to this open access knowledge base by deleting incorrect information, restructuring or completely rewriting any text.

Ideally this entry shall become one comprehensive and continuous article. Bulleted lists, for instance, were only used because it is

POWERED WITH
AUTHORSHIP TRACKING
TECHNOLOGY ■ ■ ■
Simply click in the text to find out who wrote what. Fair credit for authors. Always know your sources.

[mememoir.org](#)

Disease relevance of PAH

- We present the case of a girl affected by [classical phenylketonuria](#) who has been screened for mutations in the [PAH gene](#) [1].
- Benzo[a]pyrene-induced [DNA damage](#) and [p53](#) modulation in human [hepatoma](#) HepG2 cells for the identification of potential biomarkers for [PAH](#) monitoring and [risk assessment](#) [2].

society
in science
The Branco Weiss Fellowship

Testowanie istotności funkcji

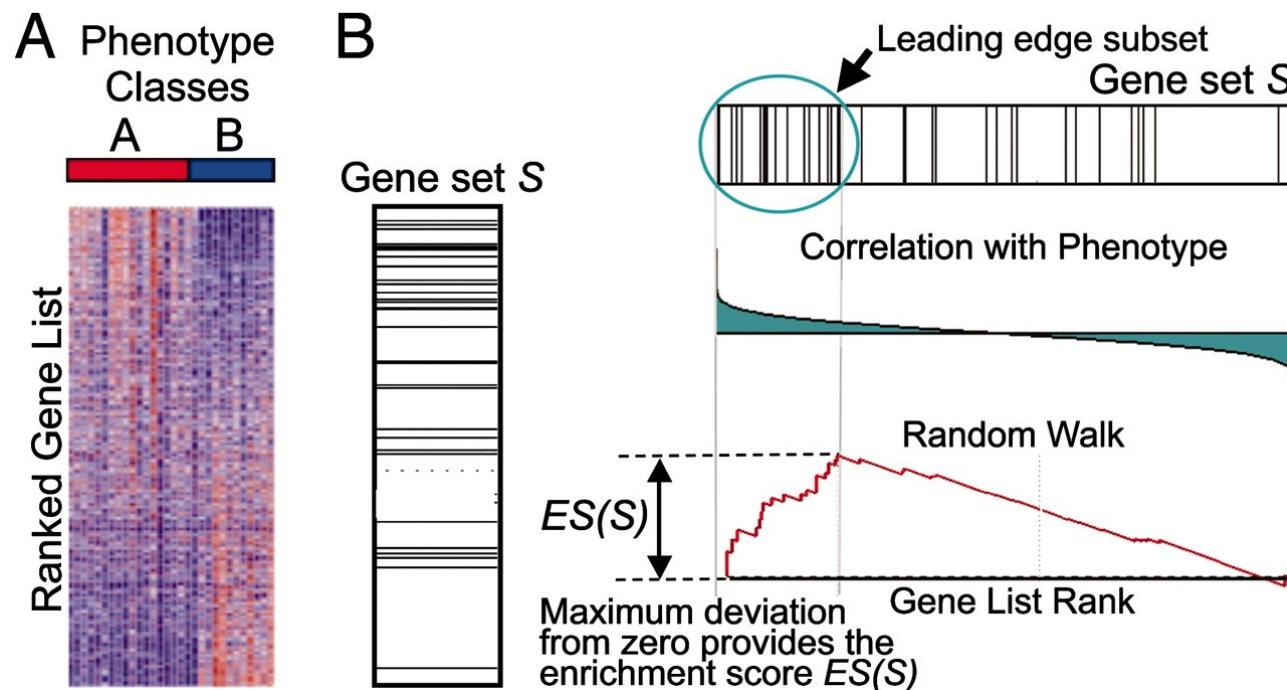
- Dla wielu zbiorów genów możemy interesować się ich funkcją (np. genów o zmienionej ekspresji, genów z mutacjami w określonej chorobie, genów wpływających na jakąś ilościową cechę)
- Często mamy dany zbiór genów, lub ranking genów (według wpływu na jakąś cechę lub ekspresji) i możemy zapytać, czy wśród genów w danym zbiorze nadreprezentowana jest jakąś funkcja. Jak odpowiedzieć na takie pytanie?

Istotne przecięcie zbiorów genów

- Najprościej jest zbadać, czy przecięcie zbioru genów posiadających daną funkcję z genami ze zbioru „zapytania” jest statystycznie istotne biorąc pod uwagę rozmiary tych zbiorów
- Odpowiada na to pytanie test Fishera, który korzysta z rozkładu hipergeometrycznego
- Standardowe pytanie: jaką jest szansa, że wyciągnę k lub więcej czarnych kul, jeśli losuję n kul spośród N kul w urnie, jeśli wiem, że kul czarnych jest M

Gene set enrichment analysis

- Jeśli mamy ranking genów to możemy zbadać, czy w górnej części rankingu mamy wzbogacenie którejś z interesujących nas funkcji wykorzystując miejsce największego wzbogacenia (podobnie jak w statystyce K-S) i empiryczne obliczanie p-wartości



Poprawki na testowanie wielu hipotez

- Ponieważ w przypadku zbiorów i rankingów genów testujemy zwykle tysiące hipotez związanych z różnymi funkcjami potencjalnie wzbogaonymi, musimy stosować poprawki na testowanie wielu hipotez
- Często stosuje się poprawkę Bonferroniego, czyli mnożenie p-wartości przez liczbę testowanych funkcji
- Lepszym sposobem jest wyliczanie false Discovery rate – FDR, który mierzy ile wyników fałszywie pozytywnych spodziewamy się w zadanej grupie predykcji