

# Genomika funkcjonalna

---

Wielkoskalowe analizy genetyczne

# Genomika funkcjonalna

---

- Kolejny po poznaniu sekwencji (struktury) genomu etap
- Poznanie funkcji wszystkich genów
- Zrozumienie interakcji pomiędzy genami - poziom systemowy

# *S. cerevisiae* jako organizm modelowy

---

- Łatwa hodowla i testy fenotypowe
- Fizjologia typowa dla mikroorganizmów
- Bezpieczne i nieszkodliwe dla środowiska
- Dobrze poznana genetyka klasyczna (analiza tetrad, mapy)
- Łatwe manipulacje genetyczne (“odwrotna genetyka”, plazmidy)
- Stabilna faza haploidalna oraz diploidalna
- Bardzo wysoka częstość rekombinacji homologicznej

# Genom *S. cerevisiae*

	Mendelian		Non-Mendelian							
Inheritance										
Nucleic acid	Double-stranded DNA			Double stranded RNA						
Location	Nucleus			Cytoplasm						
Genetic determinant	Chromosomes		2- $\mu$ m plasmid	Mitochondrial DNA		RNA Viruses				
Relative amount	85%		5%	10%		L-A	M	L-BC	T	W
Number of copies	2 sets of 16		60-100	~50 (8-130)		80%	10%	9%	0.5%	0.5%
Size (kb)	13,500 (200-2,200)		6.318	70-76		103	170	150	10	10
Deficiencies in mutants	All kinds		None	Cytochromes $\alpha$ - $\alpha_3$ & $\beta$		4,576	1.8	4.6	2.7	2.25
Wild-type	<i>YFG1</i> <sup>+</sup>		<i>cir</i> <sup>+</sup>	$\rho$ <sup>+</sup>		Killer toxin		None		
Mutant or variant	<i>yfg1-1</i>		<i>cir</i> <sup>o</sup>	$\rho$ <sup>-</sup>		<i>KIL-k<sub>1</sub></i>		<i>KIL-o</i>		

Pełna sekwencja znana od 1996 (pierwszy genom eukariotyczny)



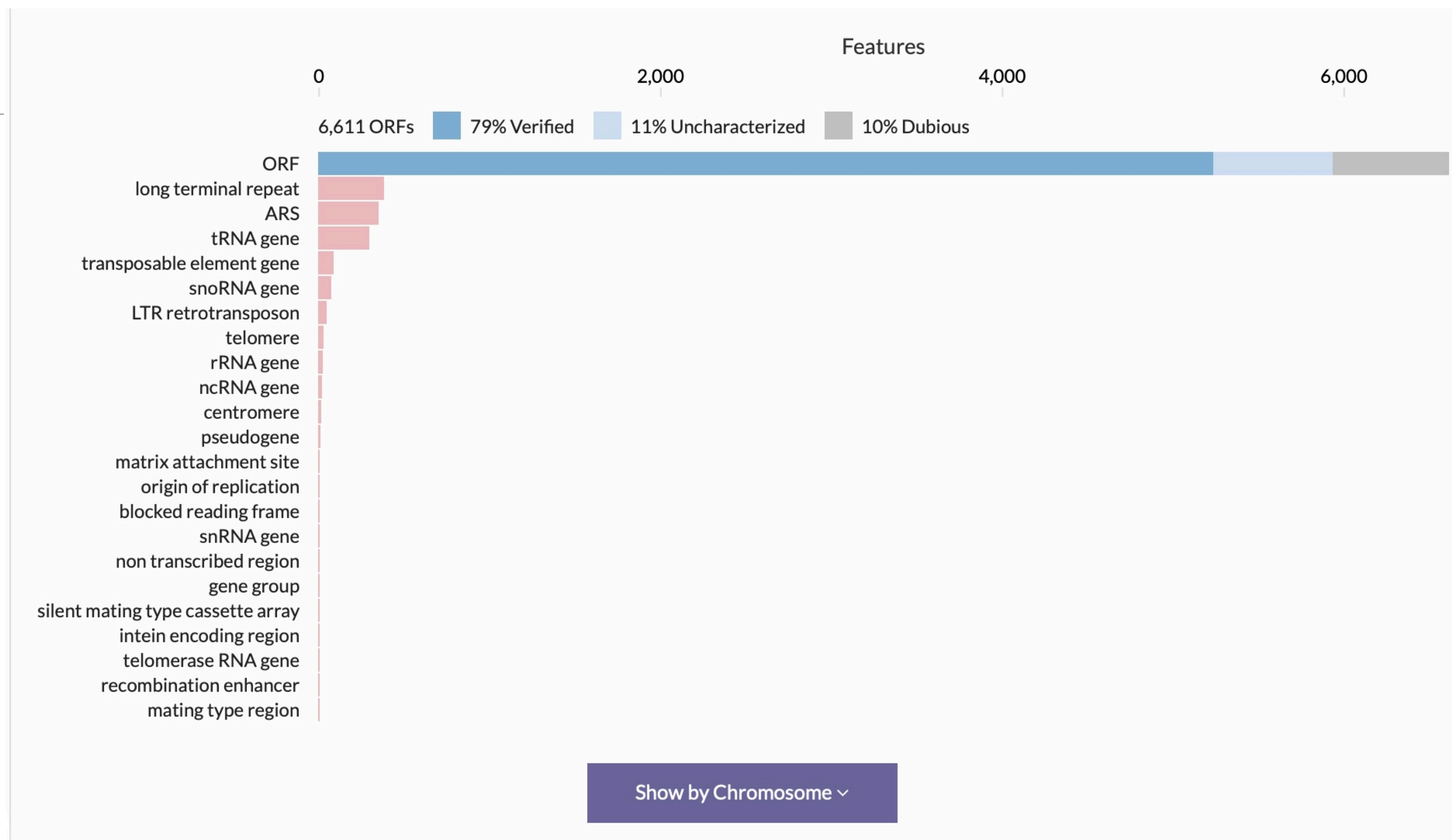
# Genom *S. cerevisiae*

---

- Genom zwarty, oszczędnie upakowany
  - Introny bardzo nieliczne (~200, 4% genów) i krótkie
  - Sekwencje regulatorowe dosyć krótkie
  - ~70% genomu to sekwencje kodujące białka
  - niewiele elementów repetytywnych
    - 52 kopie transpozonu TY, ~250 pojedynczych LTR

# Genom *S. cerevisiae*

- Około 12,5 Mb
  - *E. coli* – ~4,2 Mb; *Drosophila* ~160 Mb; człowiek ~3 000 Mb
- Około 6500 genów kodujących białka



Genome Inventory (as of 3/14/2022)

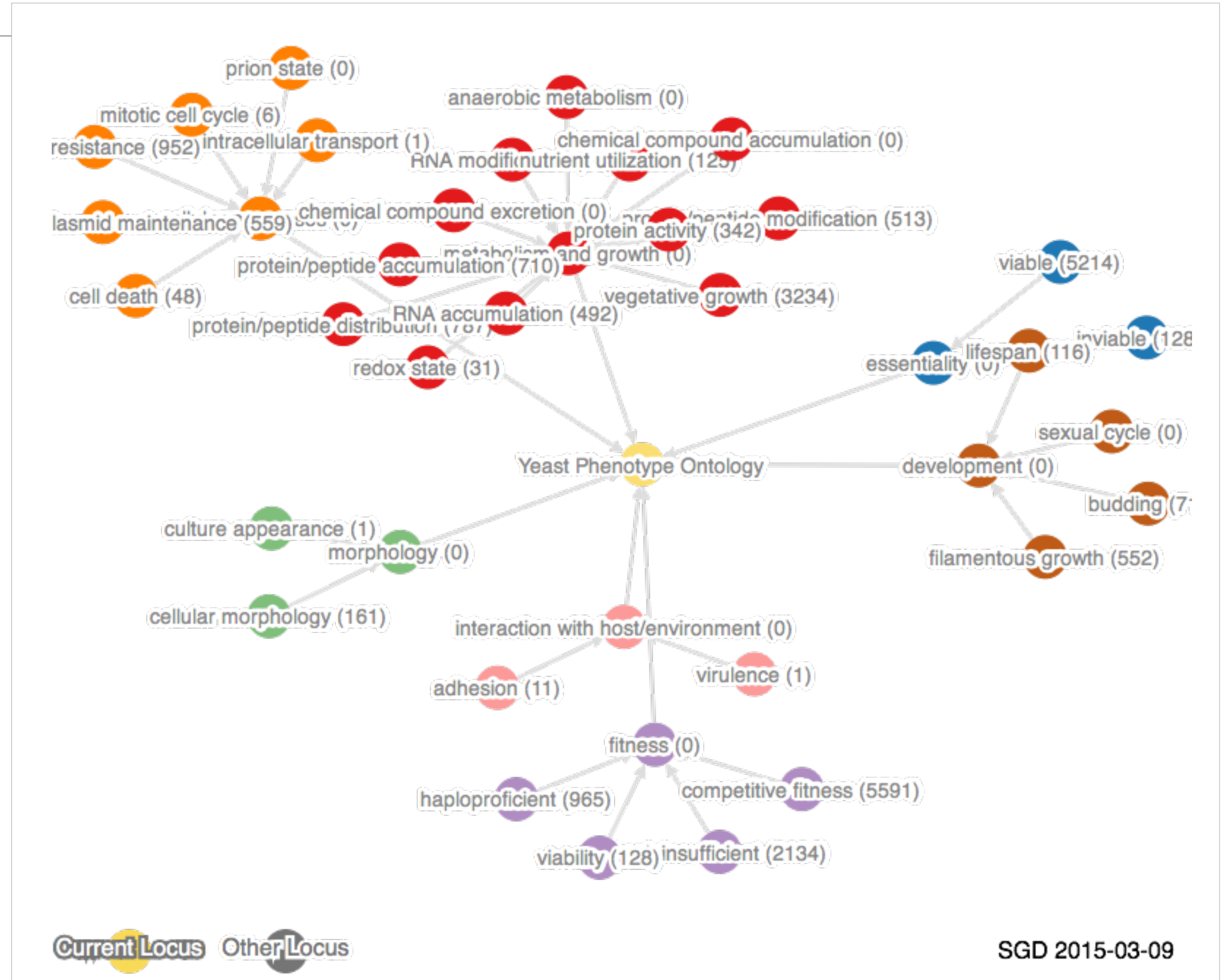
# Baza danych genomu drożdżowego

---

- SGD (Saccharomyces Genome Database) [www.yeastgenome.org](http://www.yeastgenome.org)
- Bardzo kompletna i starannie utrzymywana
- Personel SGD utrzymuje m. in. standaryzację nazw genów, referencyjną sekwencję genomu, annotację funkcji itp.
- GO - *Gene Ontology* - system hierarchicznych terminów opisujących funkcje, procesy biologiczne, lokalizację

# Standaryzacja

- Nomenklatura genów
  - gen: *ABC1*, mutant *abc1*, białko Abc1p
- Opis funkcji (GO - gene ontology)







# Gene ontology (GO)

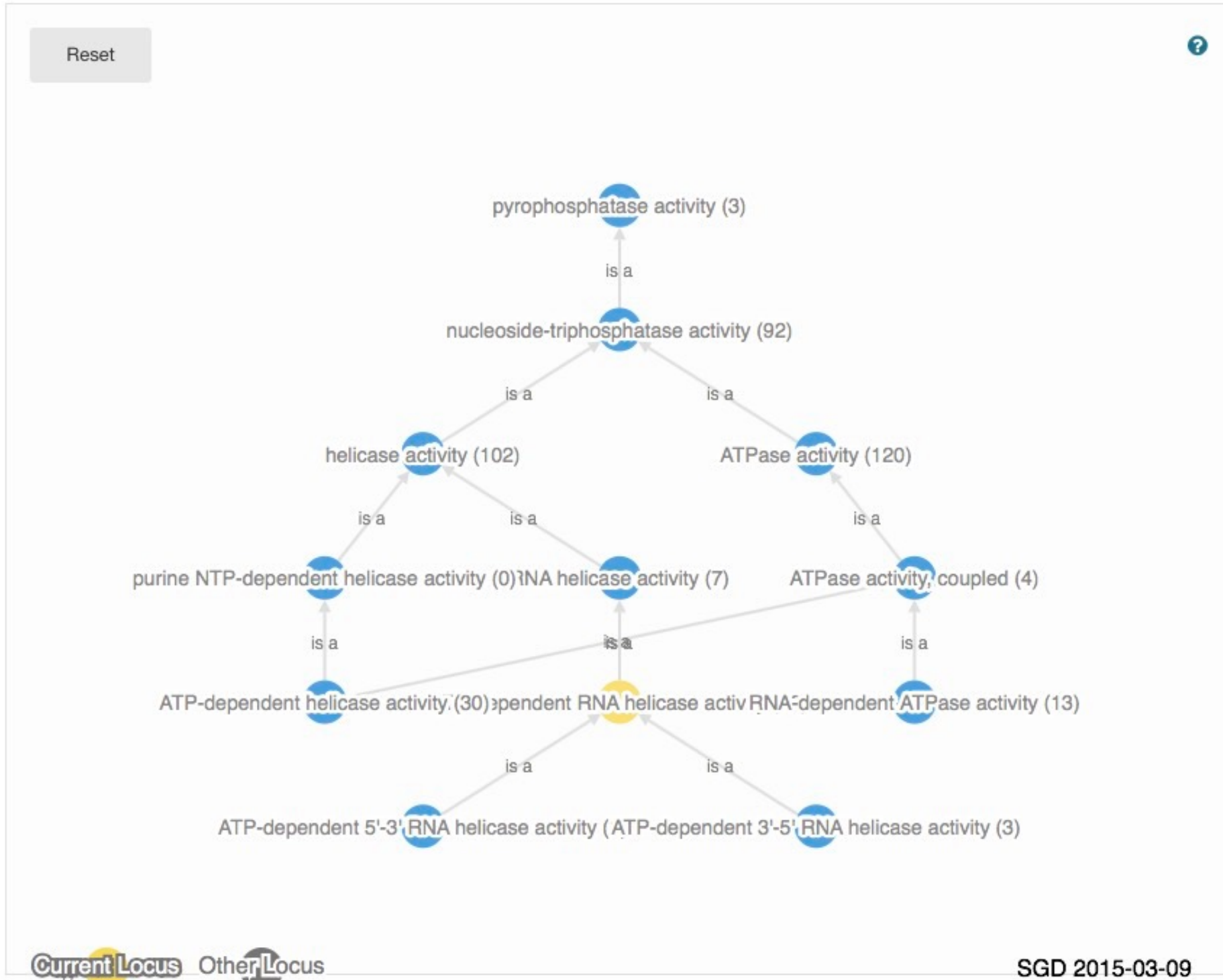
## Gene Ontology Term: **ATP-dependent RNA helicase activity**

### Overview

**GO ID:** [GO:0004004](#)  
**Aspect:** Molecular Function  
**Description:** Catalysis of the reaction: ATP + H<sub>2</sub>O = ADP + phosphate; this reaction drives the unwinding of an RNA helix.

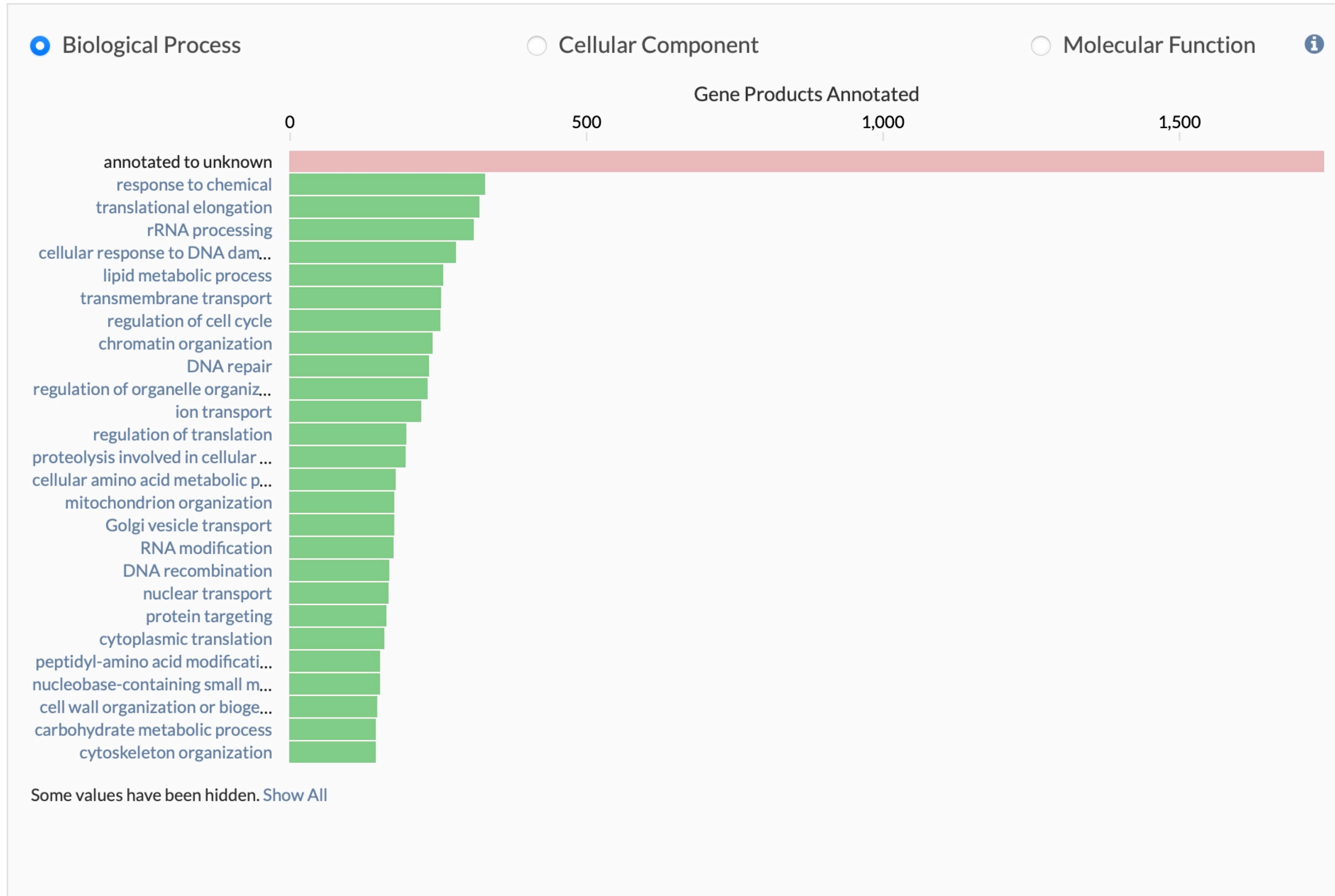
[View GO Annotations in other species in AmiGO](#)

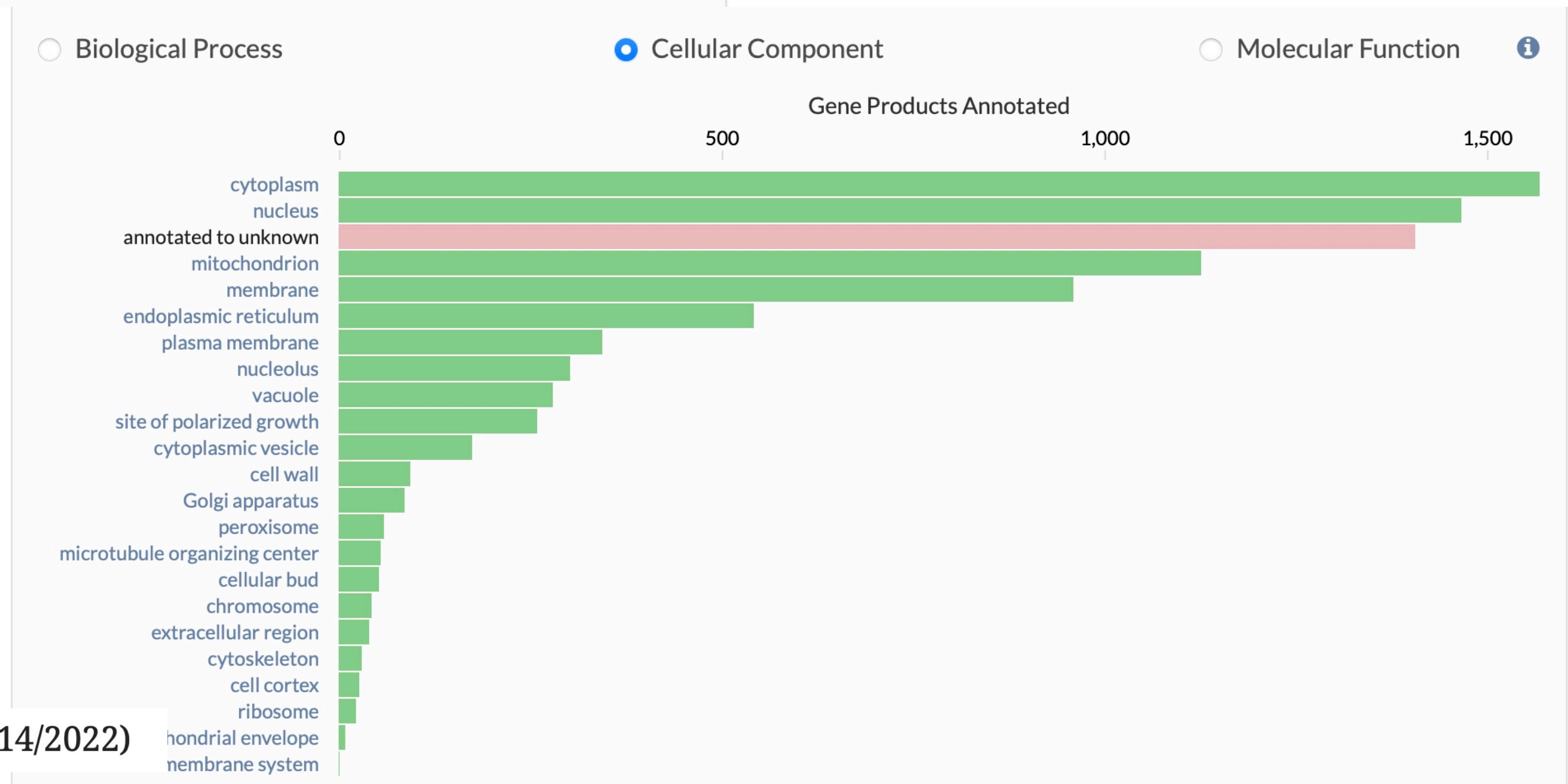
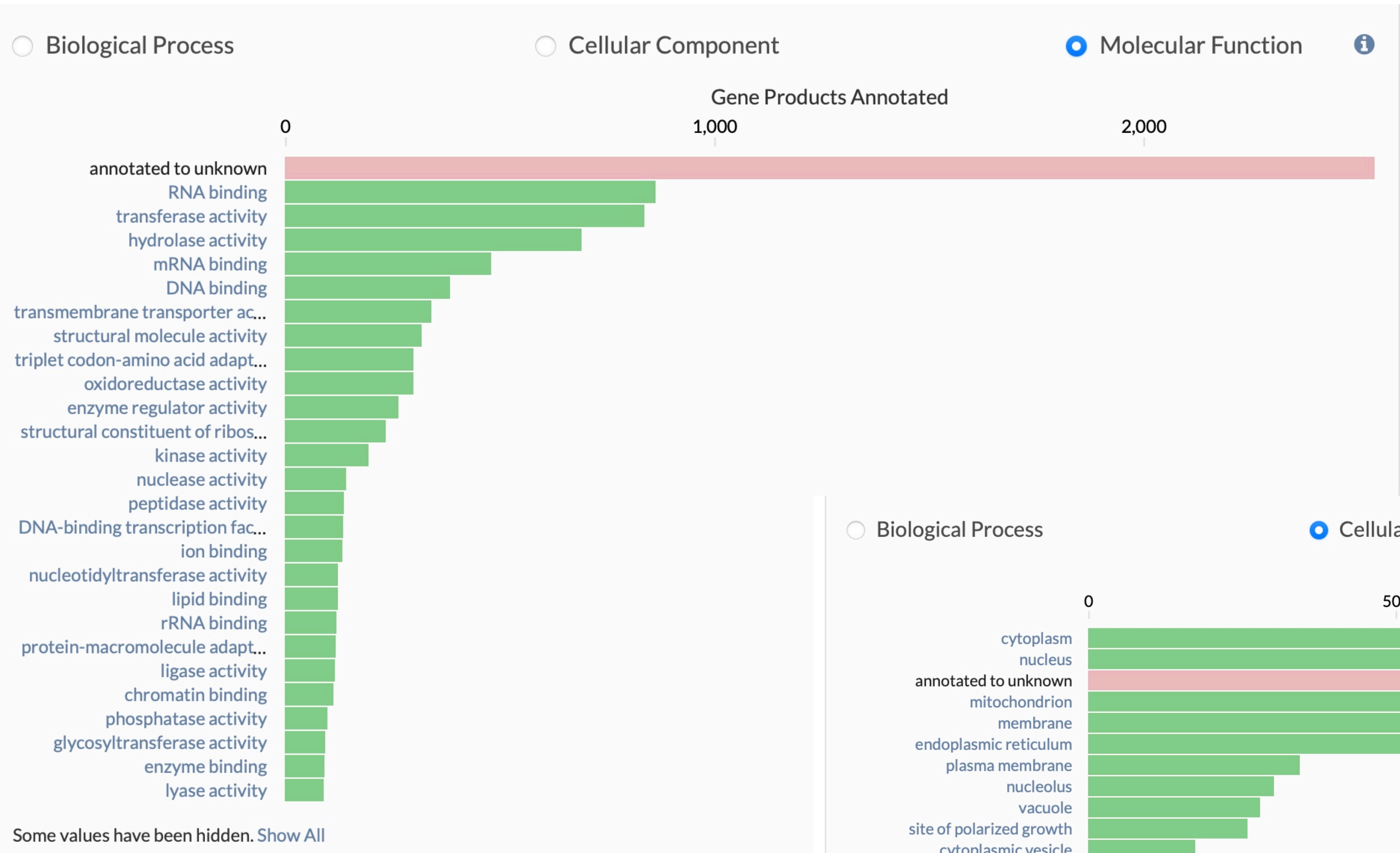
### Ontology Diagram





# Summary of Gene Ontology (GO) Annotations (as of 3/14/2022)





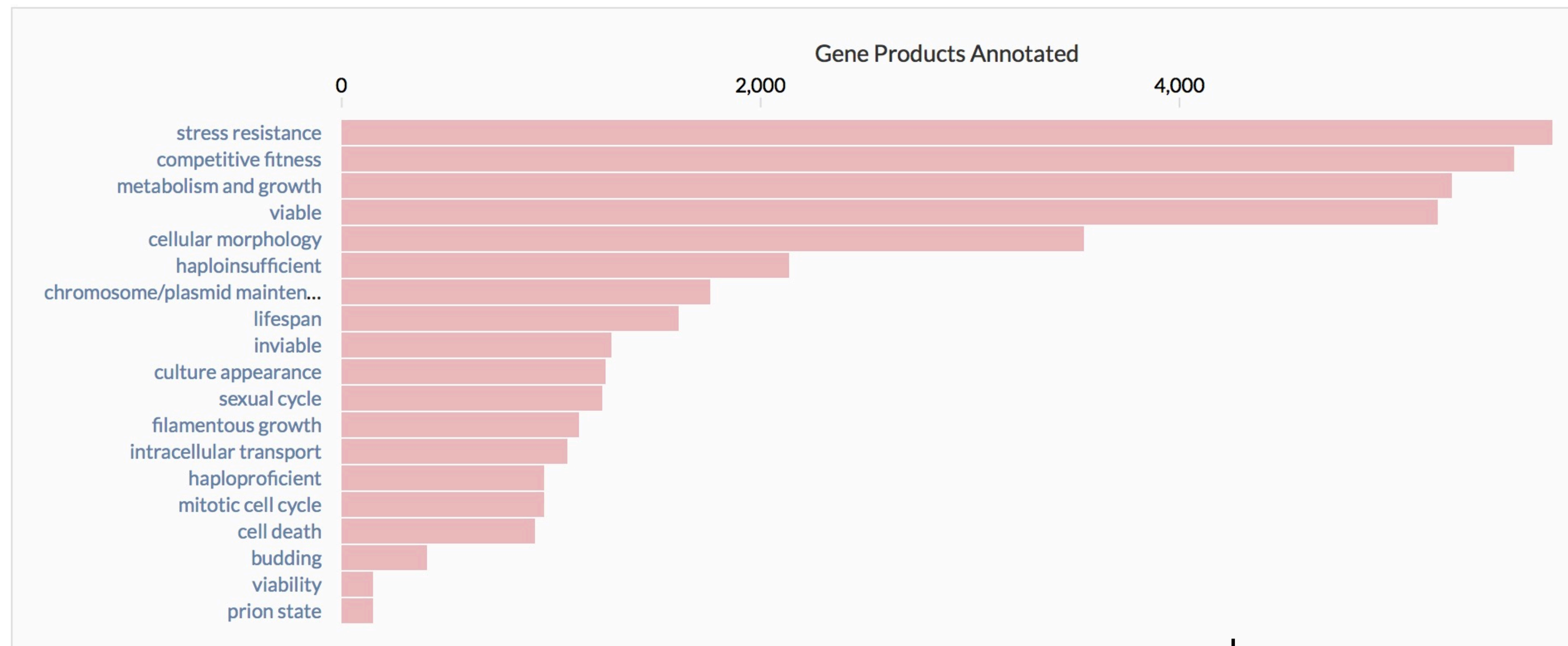
Summary of Gene Ontology (GO) Annotations (as of 3/14/2022)



# Najlepiej poznany genom eukariotyczny

## Summary of Phenotype Annotations (as of 3/6/2017)

This bar graph represents the state of phenotype annotations for the entire *S. cerevisiae* genome using a phenotype slim (i.e., a high-level subset of phenotype terms). This subset allows phenotypes to be grouped into broad categories. More information about phenotype annotation and the ontology can be found on SGD's [phenotype help page](#). All the phenotype data summarized on this graph can be found in our [phenotype\\_data.tab](#) file.

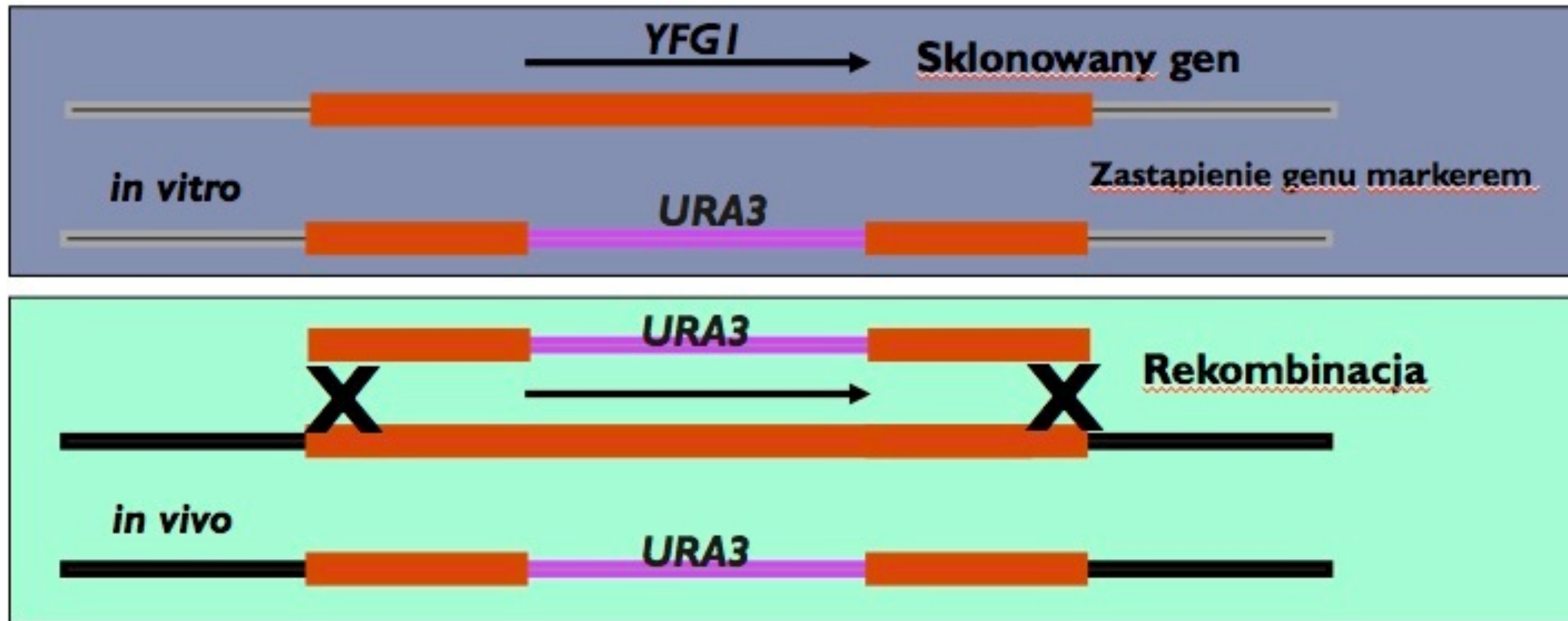


# Narzędzia genetyki molekularnej drożdży

---

- *S. cerevisiae* to jeden z najłatwiejszych w manipulowaniu organizmów
- Podstawowe właściwości
  - bardzo wysoka wydajność rekombinacji homologicznej
  - łatwość ukierunkowanej manipulacji genomem

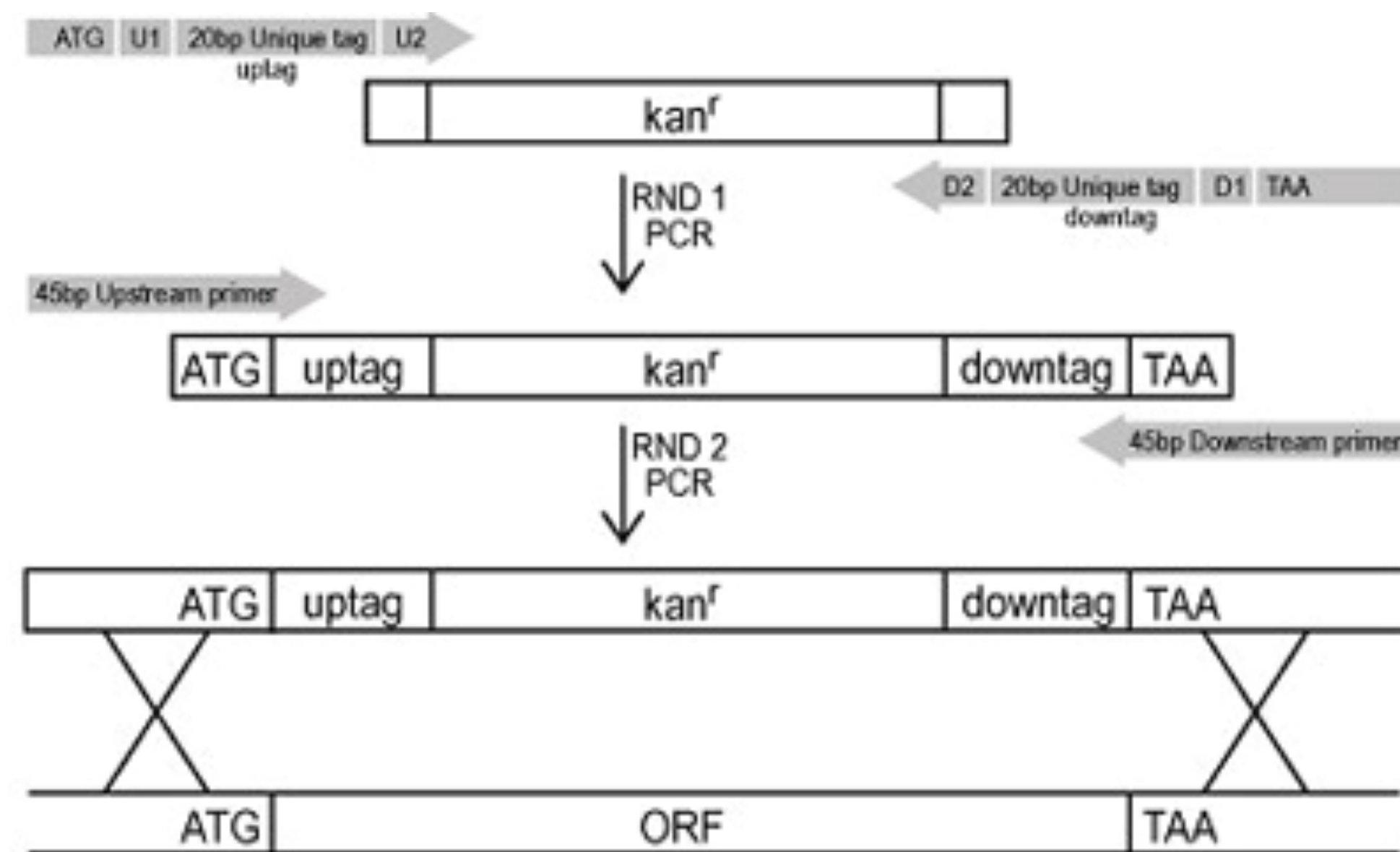
# Delecja genu przez rekombinację



Tzw. dysrupcja

# Dysrupcje – metoda PCR

- Homologiczne flanki nie muszą być długie, można je dodać przez PCR
- Kasetę *KanMX4*, nie pochodzący z drożdży, nie rekombinuje sama z genomem





# Kolekcje

---

- Dostępne kolekcje szczepów delecyjnych
- Diploidalne heterozygotyczne >6000
- Homozygotyczne ~4800
  - reszta – letalne u homozygot

# Kolekcje

---

- Inne kolekcje
  - fuzje ze znacznikami powinowactwa
  - ORF pod regulowanym promotorem
  - nadekspresja ORF ze znacznikiem powinowactwa
  - i wiele innych

# Projekty na skalę genomową

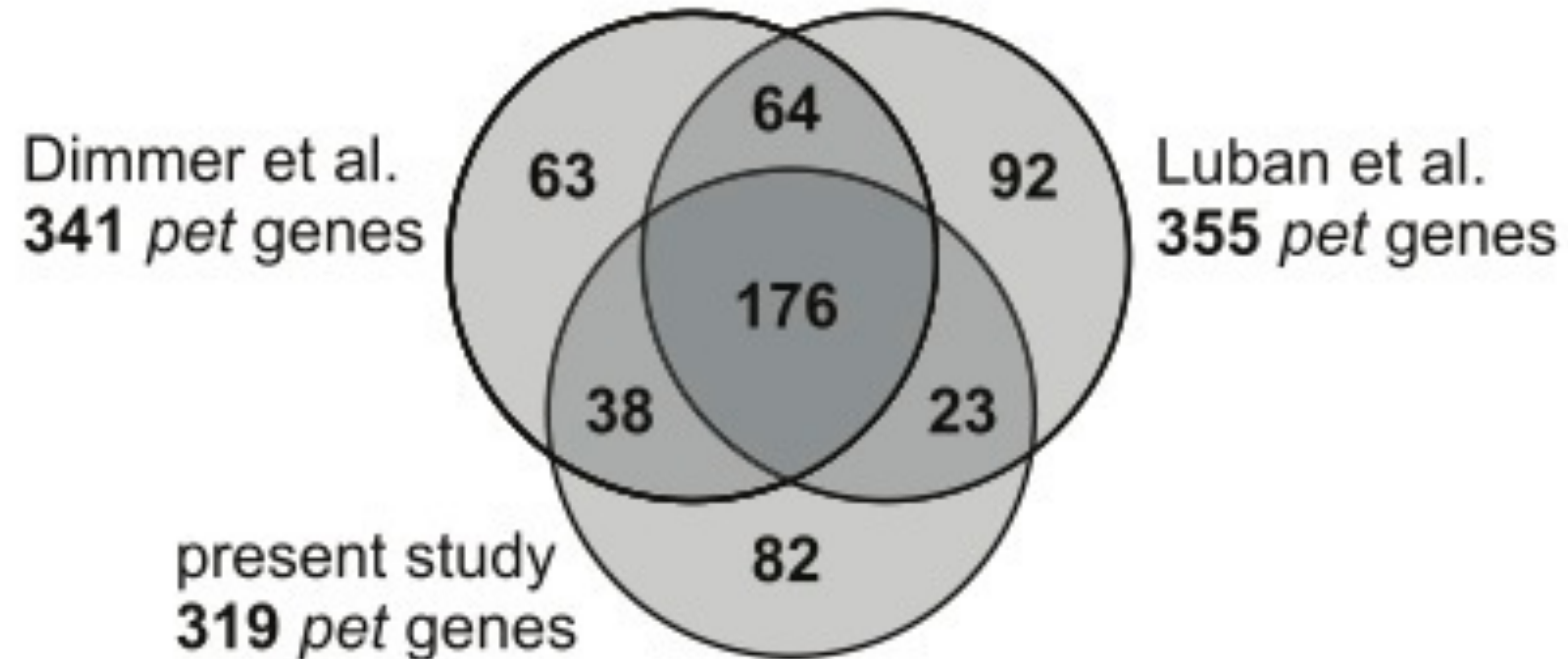
---

- Delecje (analiza fenotypowa)
- Nadekspresja białek (MORF)
- Zmiany ekspresji genów (mikromacierze, fuzje reporterowe)
- Lokalizacja białek w komórce (fuzje z GFP)
- Interakcje białek (system dwuhybrydowy)
- Interakcje genetyczne (np. syntetyczne letalne)
- Mapowanie QTL

# Problem analiz wysokoprzepustowych

---

**(a)**



Geny *pet* (niezbędne do oddychania)



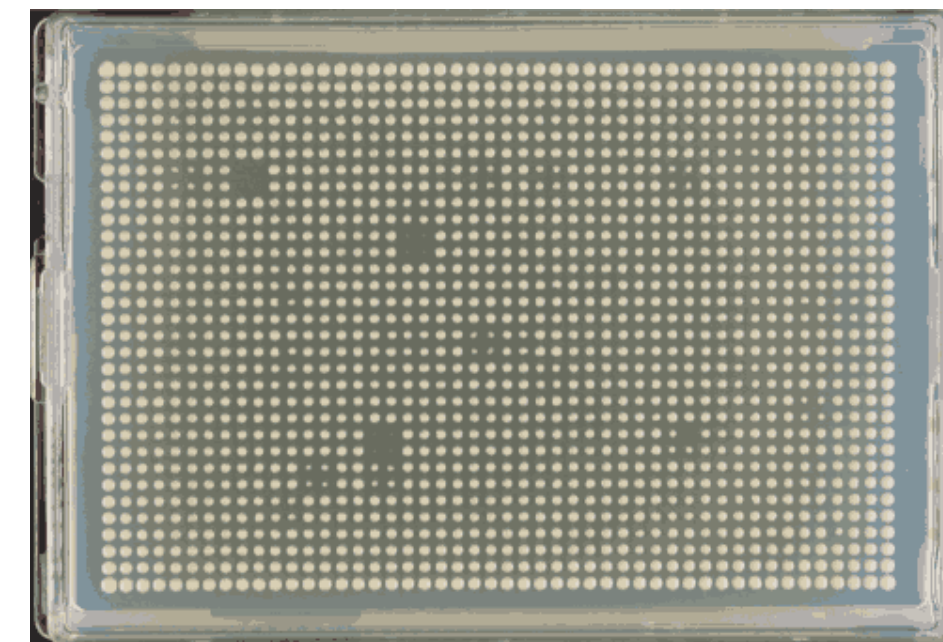
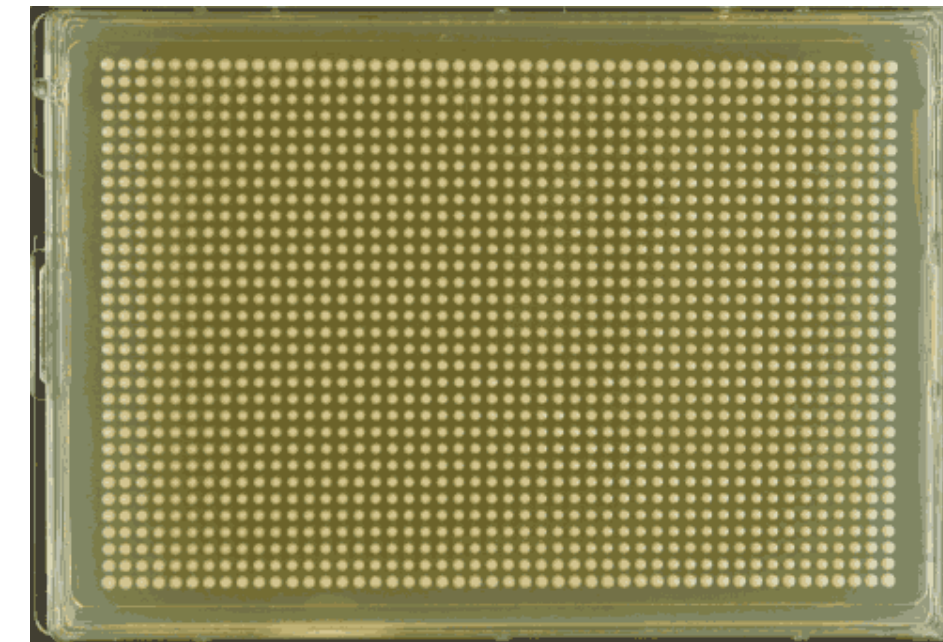
# Problem analiz wysokoprzepustowych

---

- Powtarzalność wyników w różnych badaniach jest niewielka
- Znaczny wpływ tła genetycznego i warunków doświadczalnych

# Analizy wysokoprzepustowe – roboty laboratoryjne

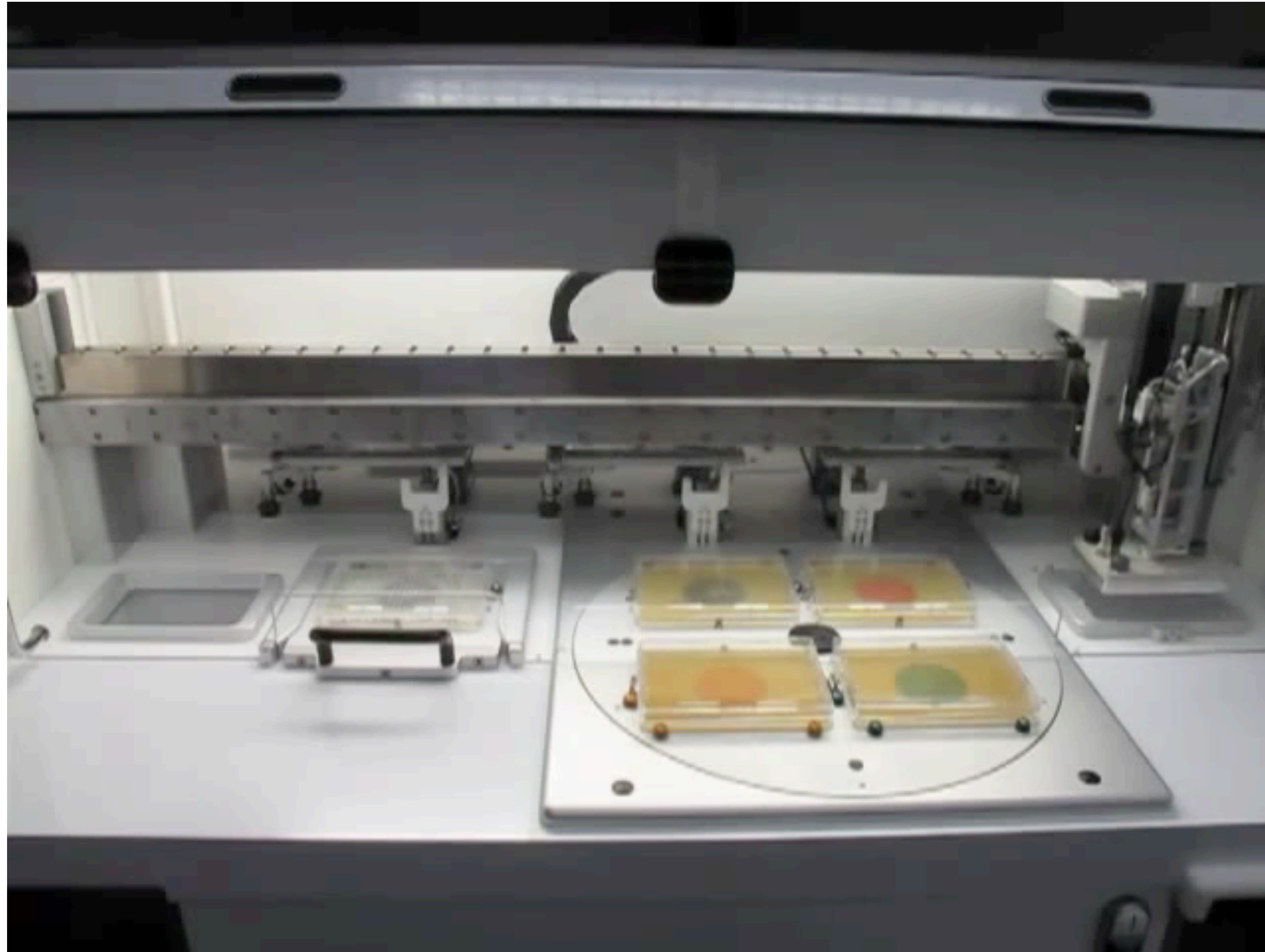
---



©Singer Instruments, UK

# Analizy wysokoprzepustowe – roboty laboratoryjne

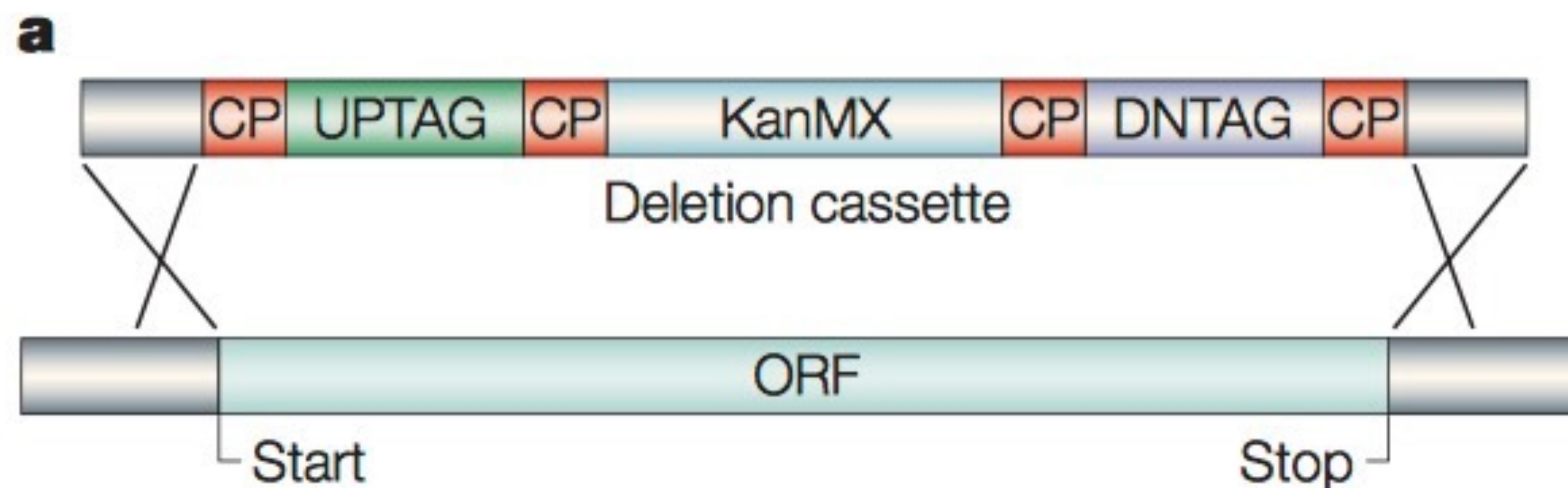
---





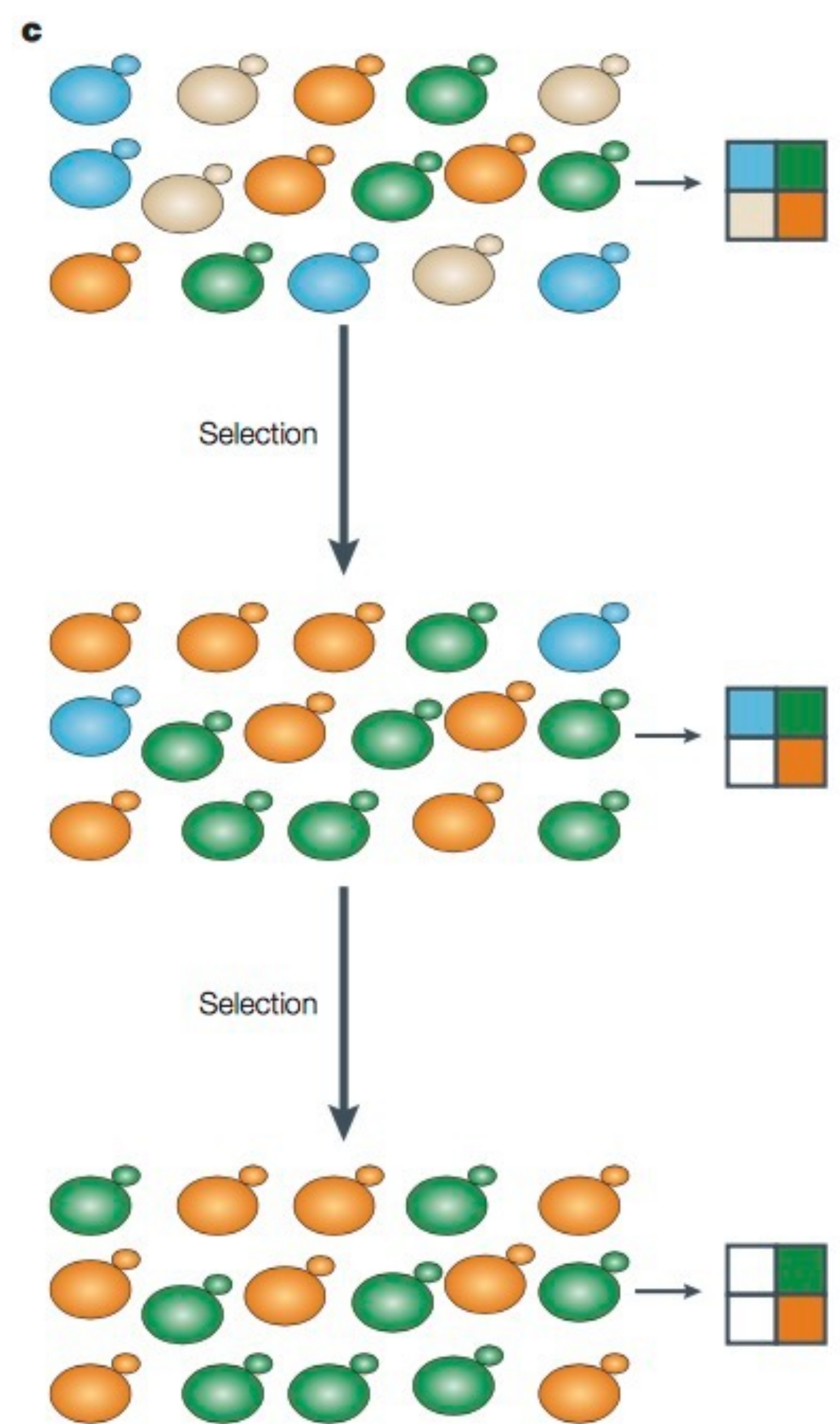
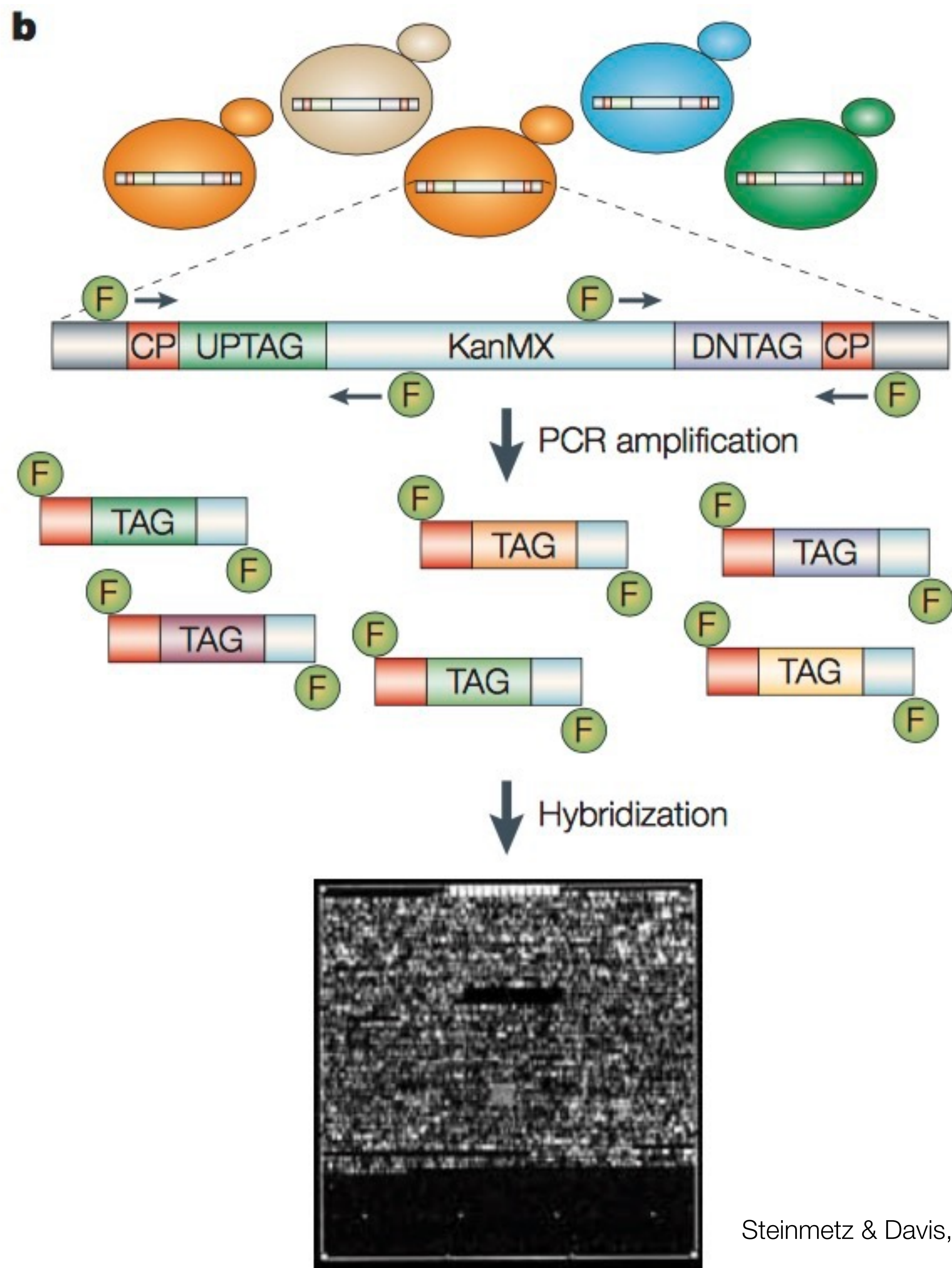
# Wykorzystanie kolekcji delecyjnych

---



CP - Common Primer - wspólny starter  
UPTAG, DNTAG - “kody kreskowe”, unikatowe sekwencje





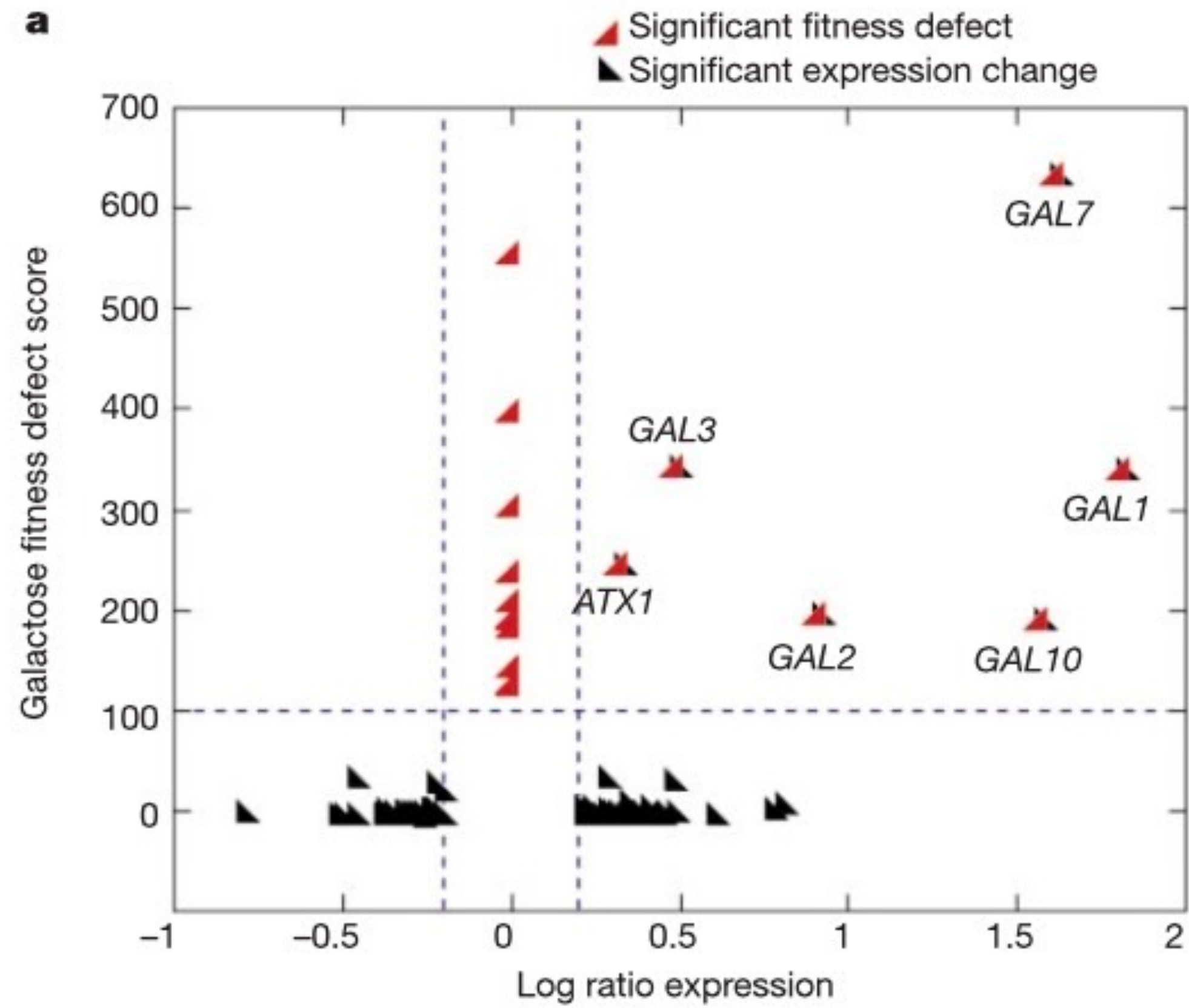
# Fenotyp a analiza ekspresji

---

- Dwa najczęstsze podejścia genomiki funkcjonalnej:
  - analiza ekspresji (transkryptomika) - zmiany poziomu mRNA w różnych warunkach
  - analiza fenotypowa - defekt wzrostowy (fitness) w określonych warunkach
- Czy wyniki się pokrywają?

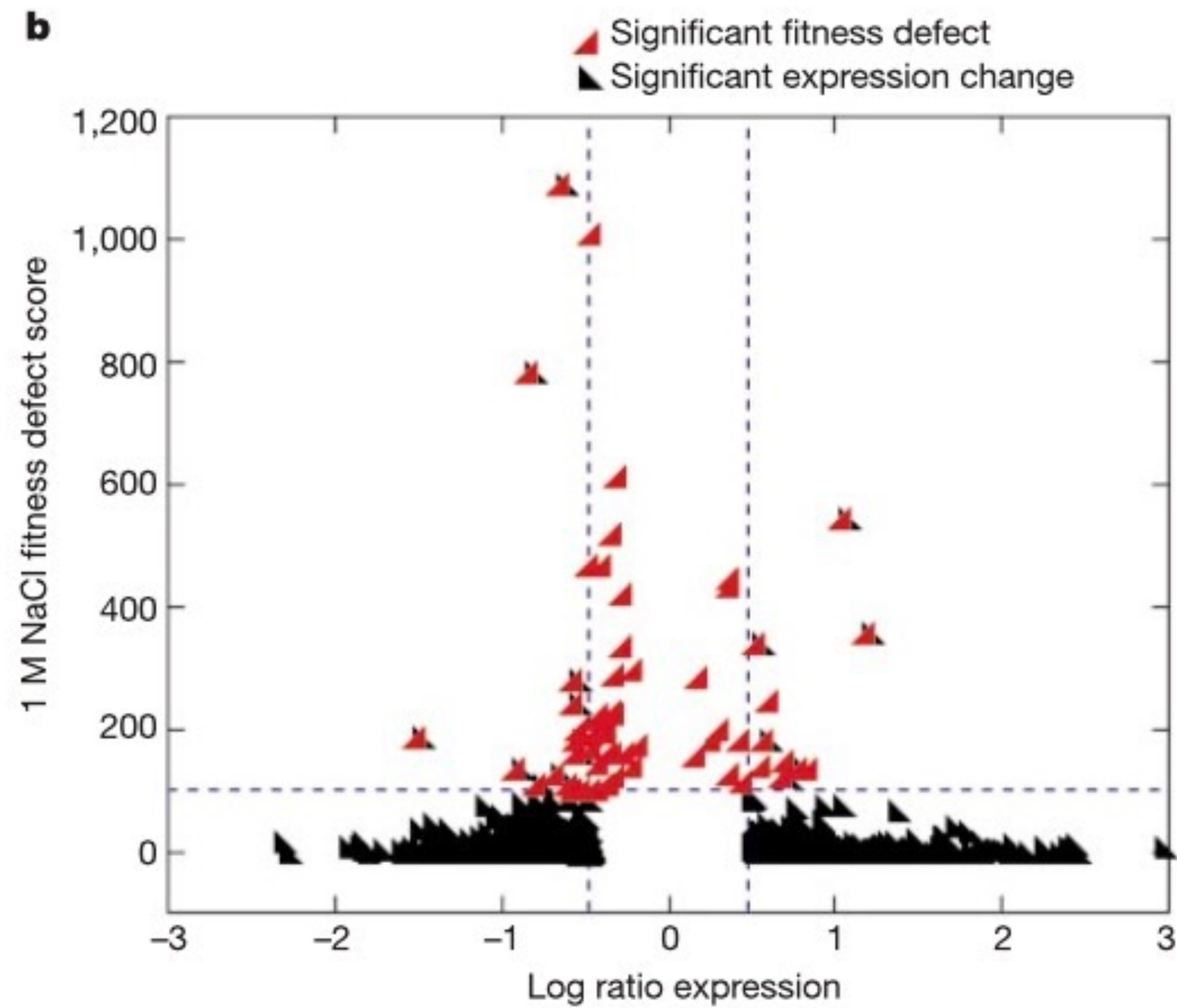


# Fenotyp a ekspresja



Wzrost na galaktozie

# Fenotyp a ekspresja



Tolerancja wysokiego stężenia soli

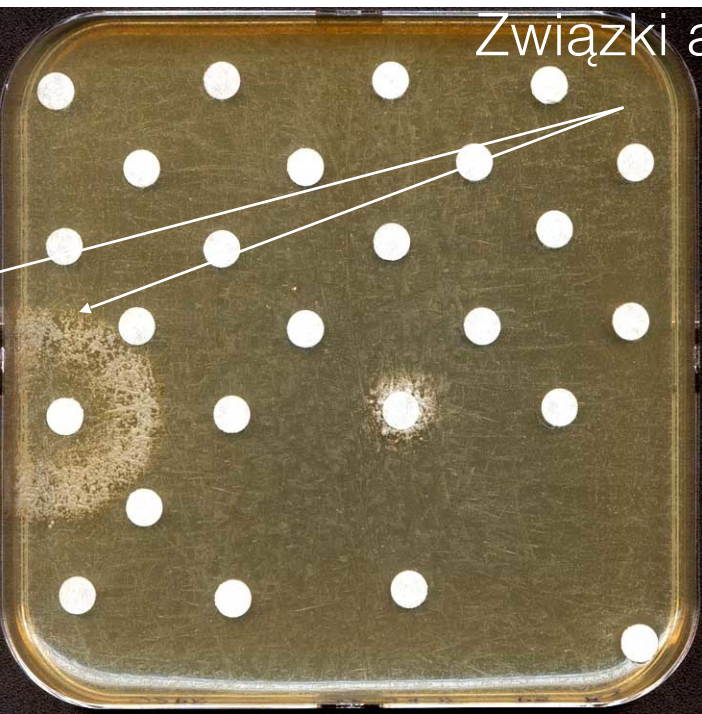
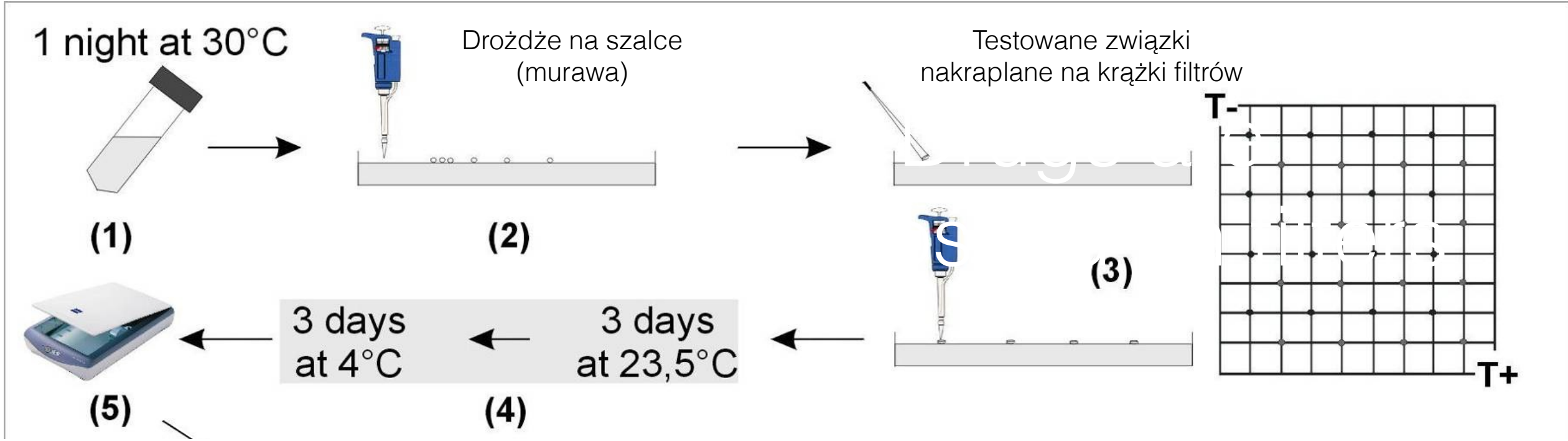


# Fenotyp a ekspresja

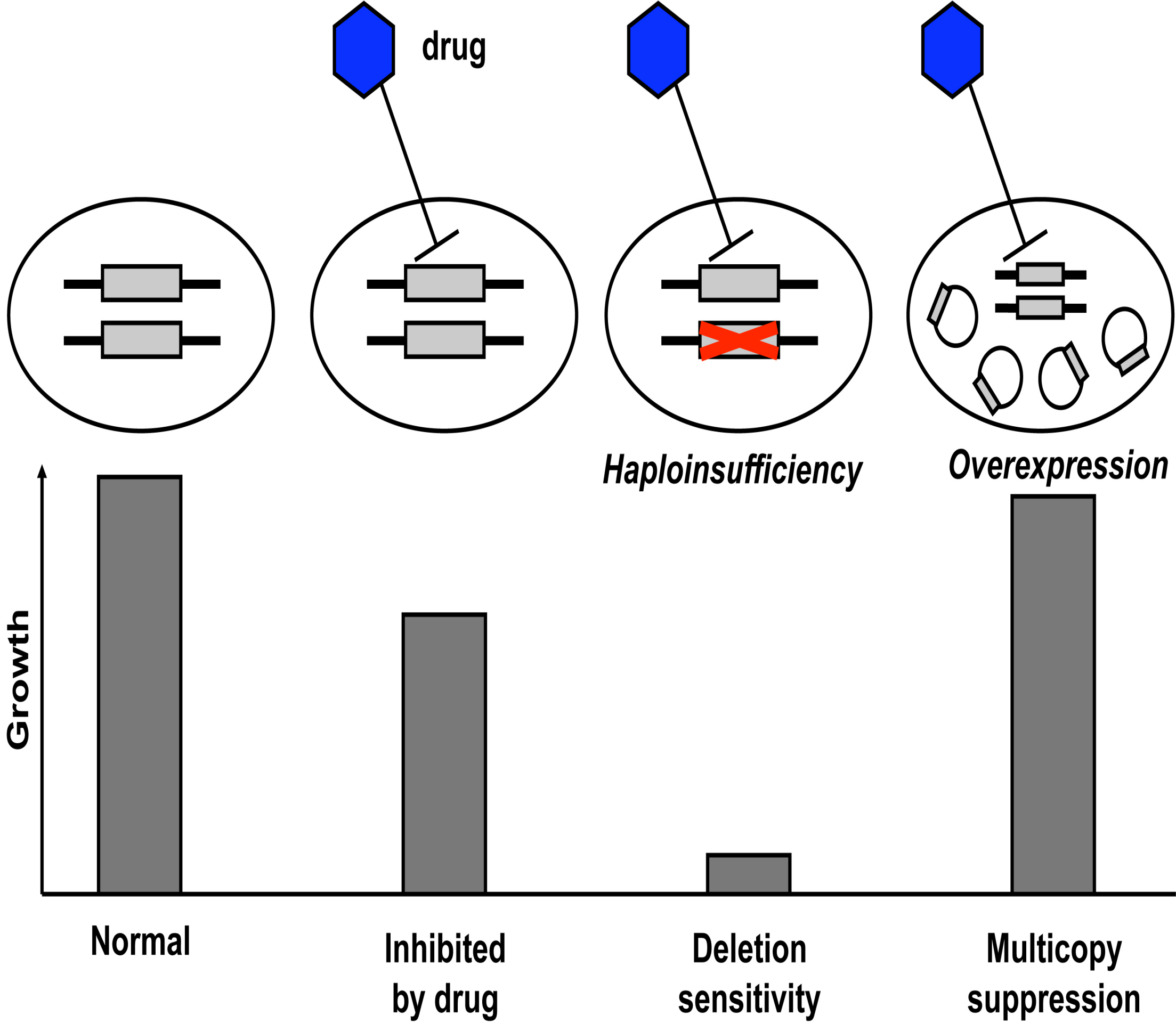
---

- Nakładanie się znaczącej zmiany ekspresji i defektu wzrostu
  - <7% dla wzrostu na galaktozie i 1M NaCl
  - ~7% dla wzrostu na niefermentowalnych źródłach węgla
  - ~16% dla sporulacji

# Identyfikacja substancji aktywnych



# Identyfikacja celów działania substancji aktywnych za pomocą genomiki drożdży

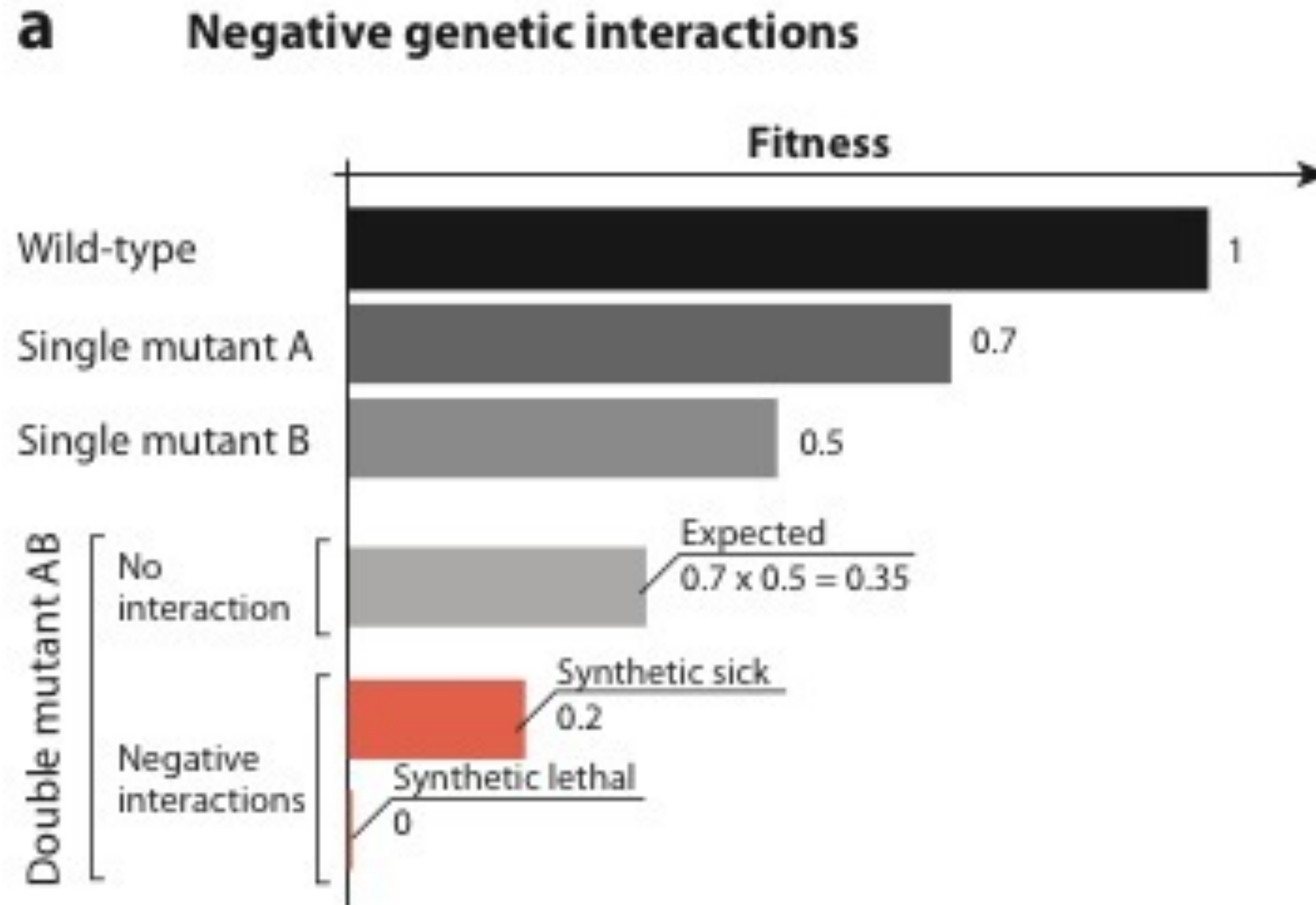


# Poszukiwanie interakcji genetycznych

---

- Oddziaływania łagodzące (np. supresja)
  - selekcja bezpośrednia
  
- Oddziaływania syntetyczne
  - syntetyczna letalność:
    - pojedyncze mutacje *gen1* i *gen2* nie są letalne, ale podwójny mutant *gen1, gen2* nie przeżywa
  - syntetyczne wzmocnienie
    - pojedyncze mutacje *gen1* i *gen2* słaby fenotyp, podwójny mutant *gen1, gen2* silny fenotyp (np. spowolnienie wzrostu)

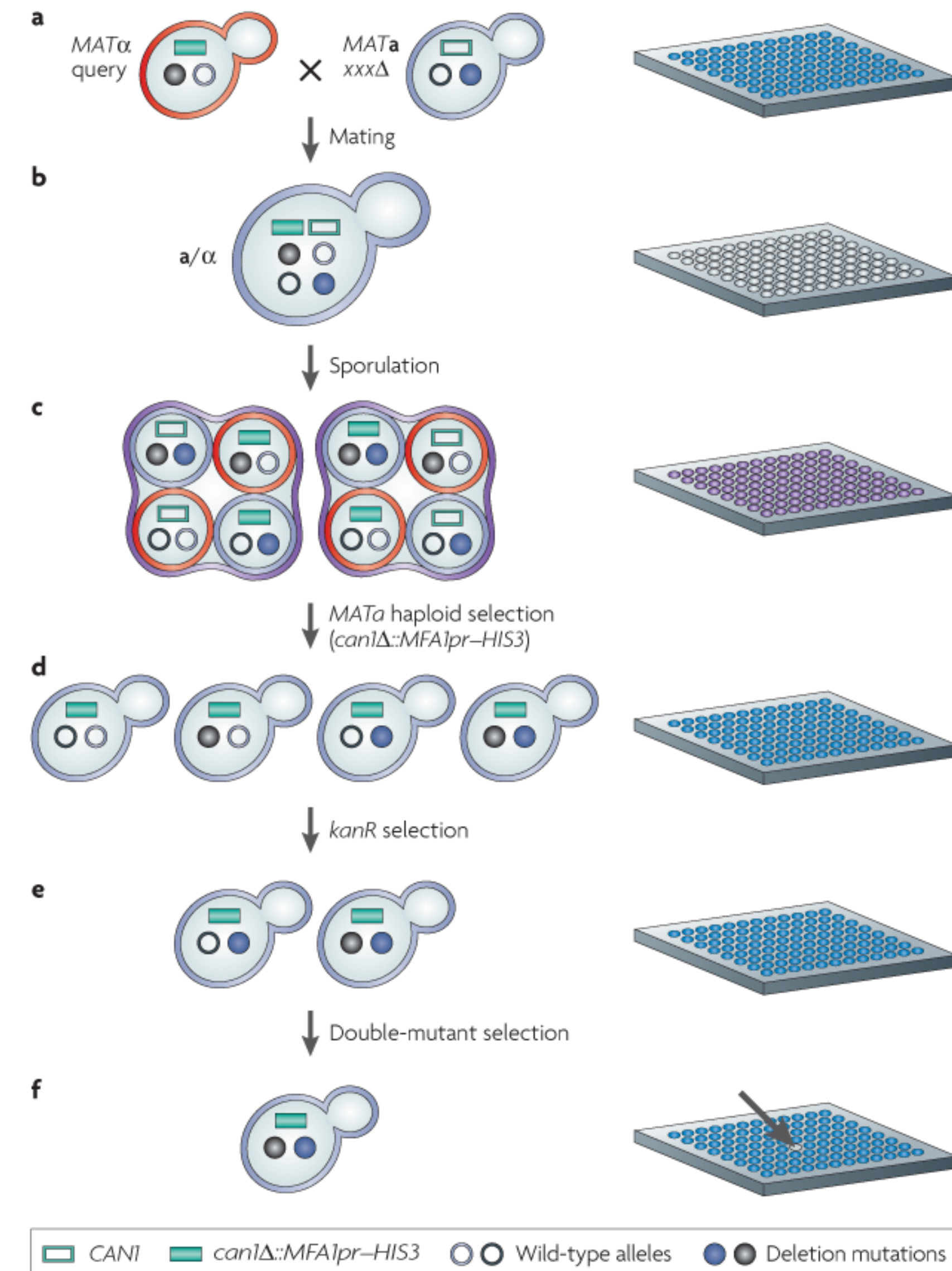
# Ujęcie ilościowe





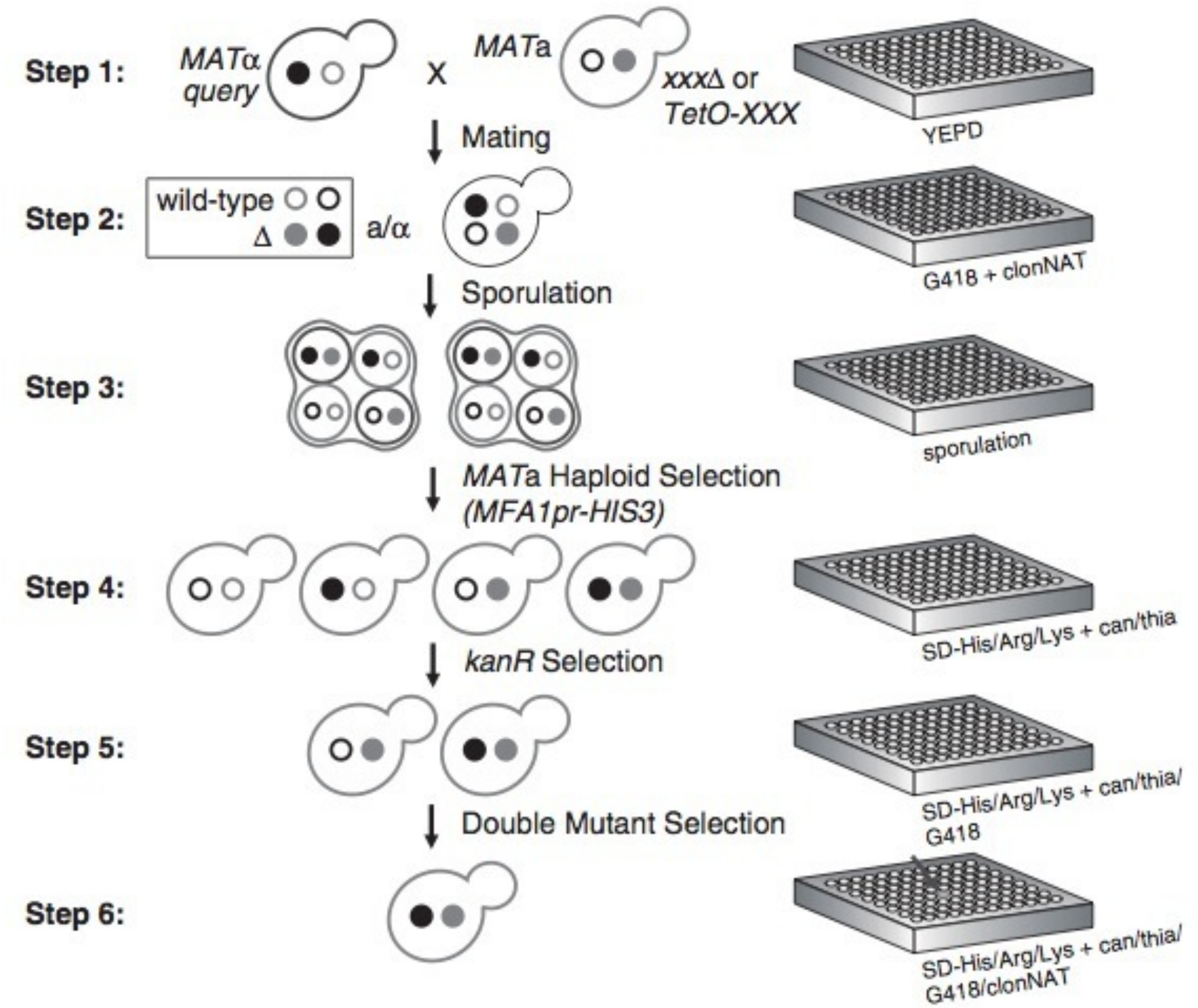
# SGA

- Synthetic Gene Array
- Kolekcja delecji, krzyżowana z badanym genem
- Sporulacja,
- Selekcja haploidów *MATa*
- Selekcja pojedynczych i podwójnych mutantów



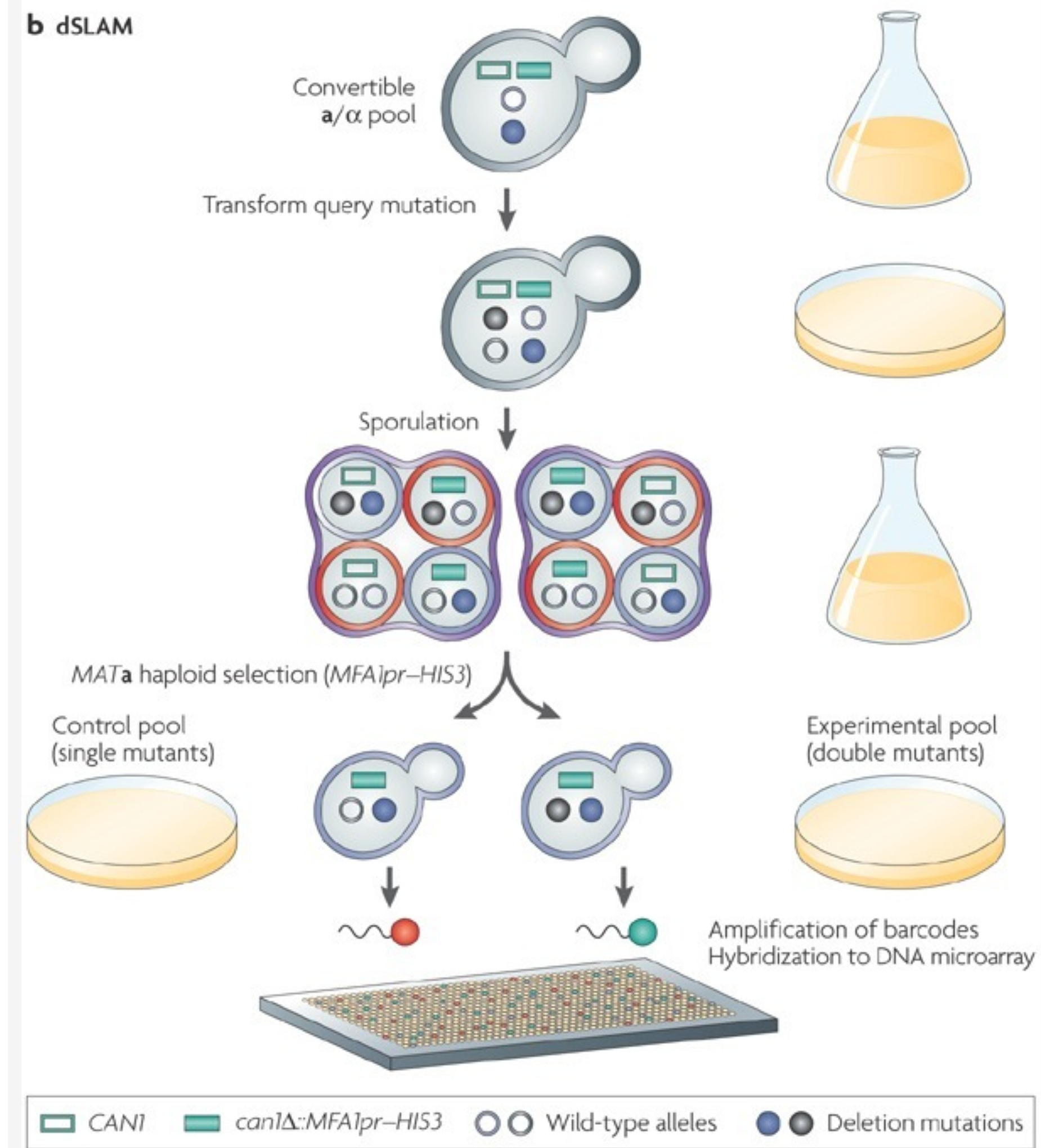


# SGA





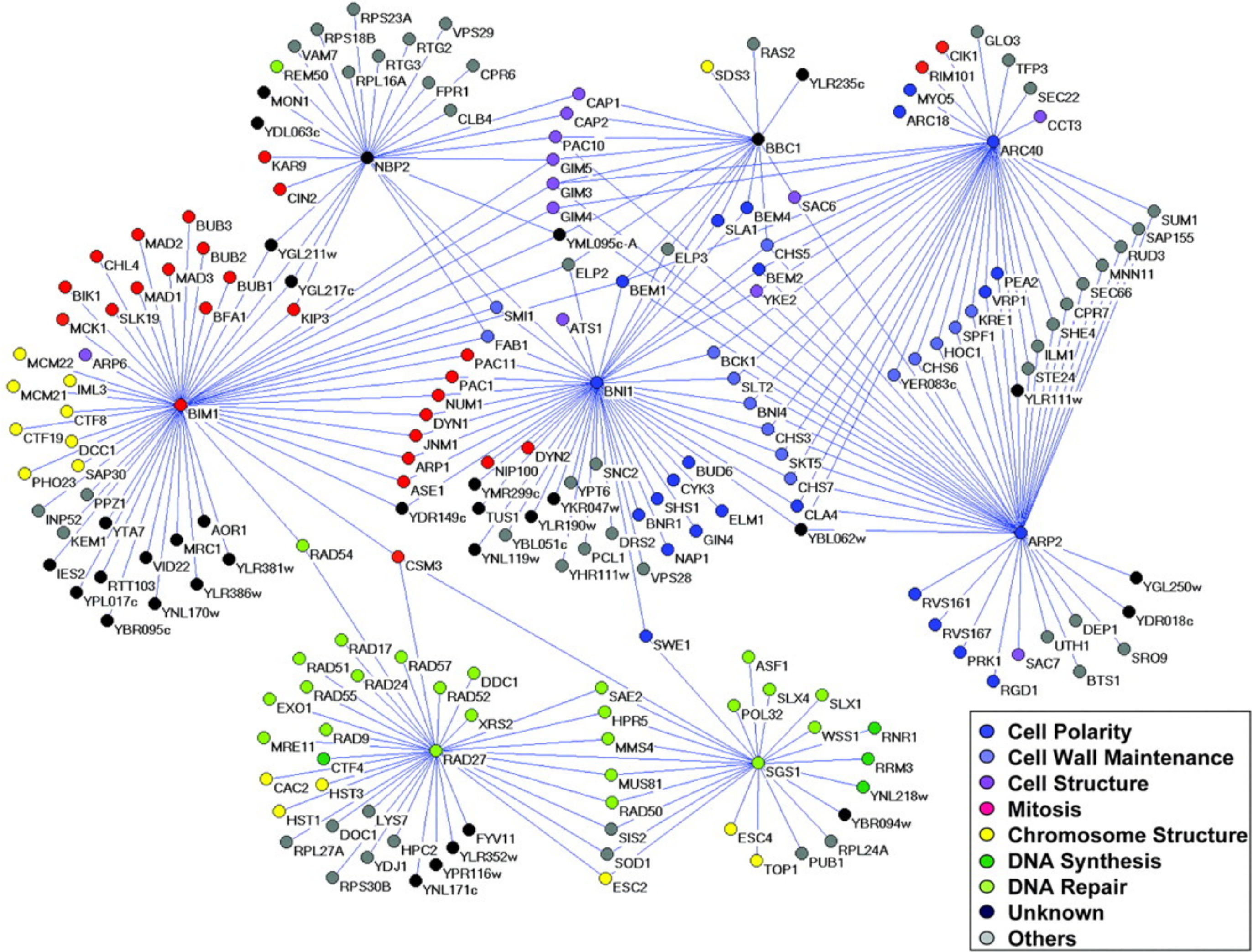
# dSLAM



Diploid-based synthetic lethality analysis with microarrays (dSLAM)



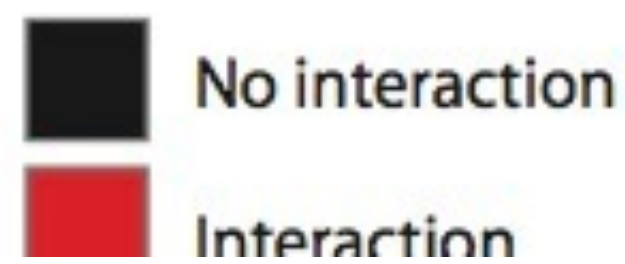
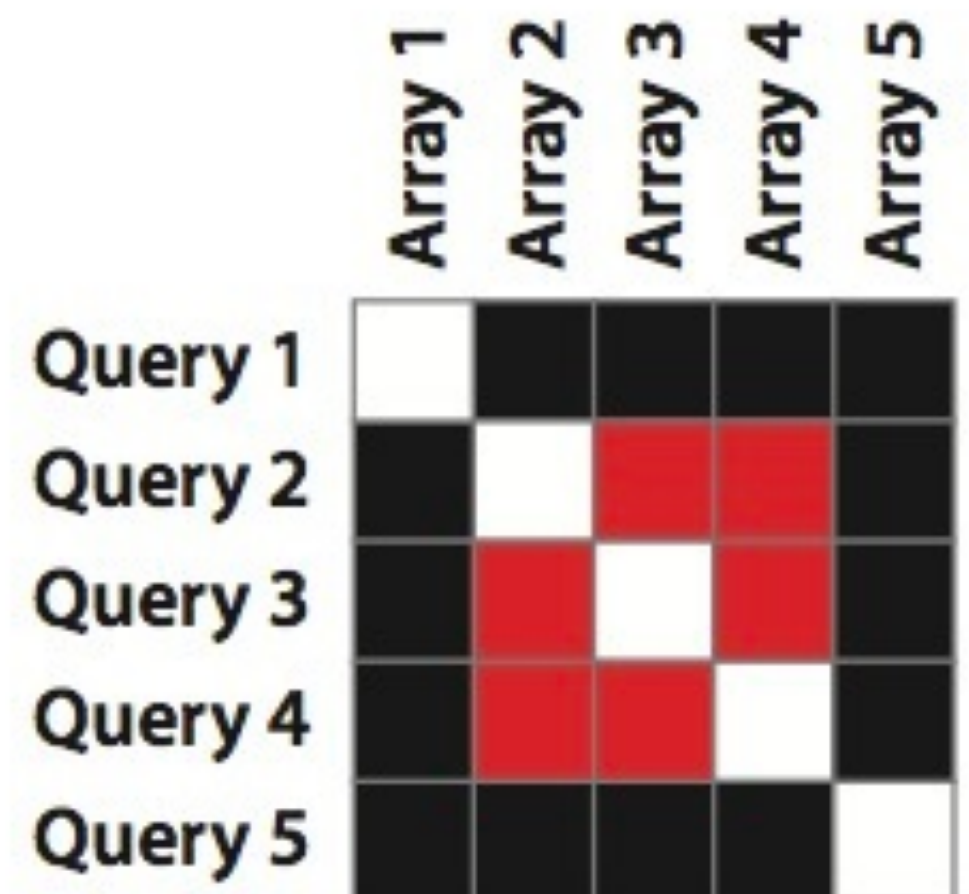
# Wyniki – sieci interakcji genetycznych



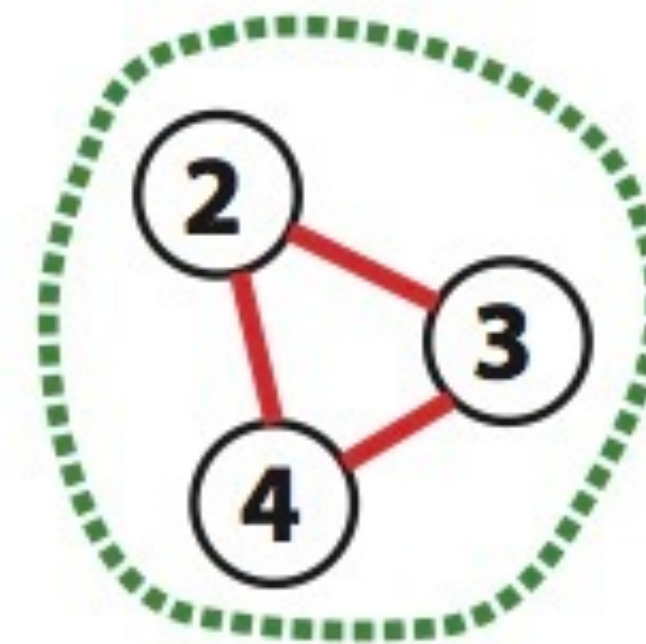


# Rekonstrukcja sieci interakcji

## Step 3: Build genetic interaction networks



Common biological process



**Explore function  
of gene cluster**

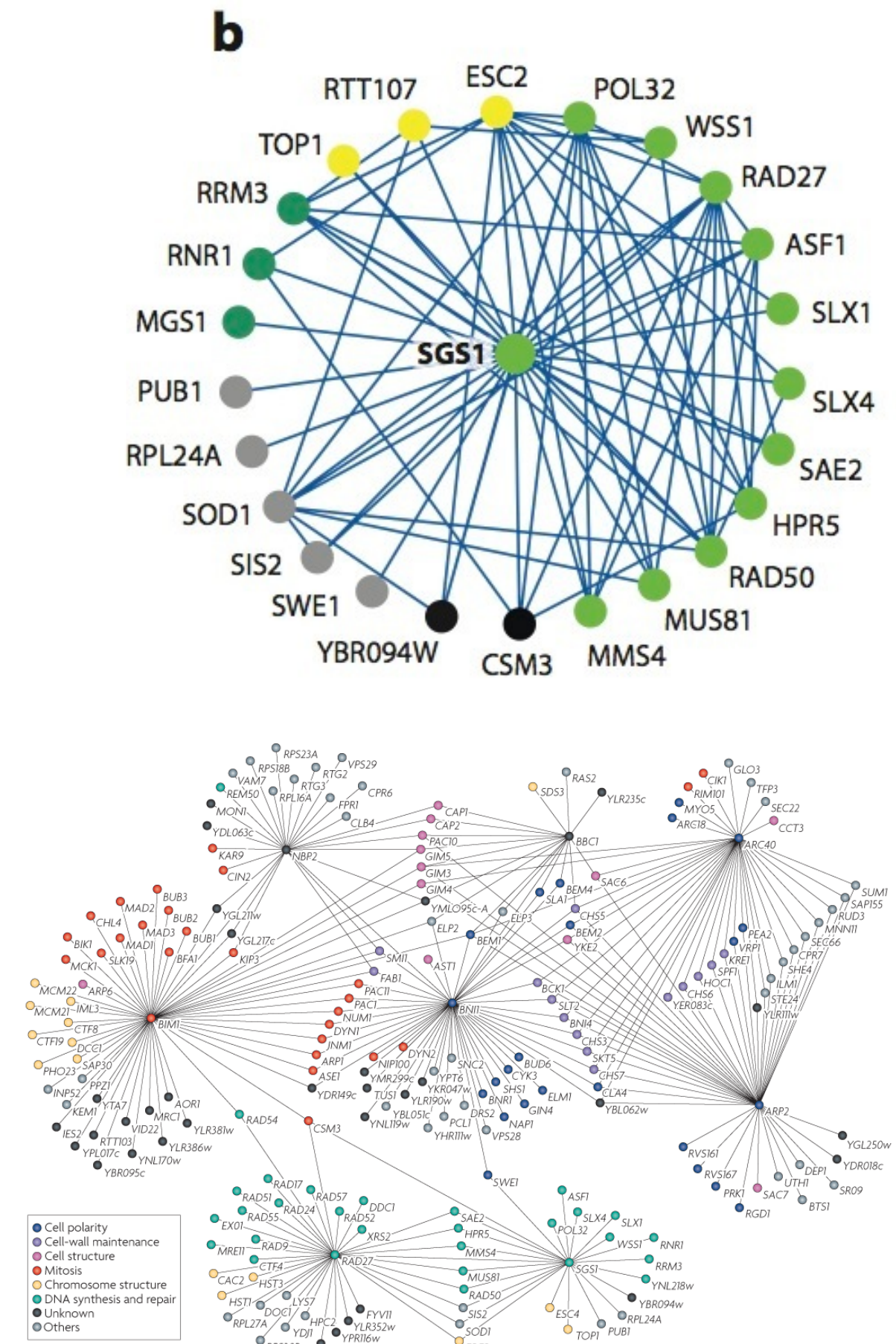
# Interakcje genetyczne – ujęcie systemowe

---

- Interakcje genetyczne wskazują na związki funkcji
  - Mogą wiązać elementy tego samego szlaku/kompleksu, ale też różnych szlaków, powiązanych funkcją
  - Zestaw interakcji (pozycja na mapie interaktomu genetycznego) może wskazywać na funkcję genu

# Sieci interakcji

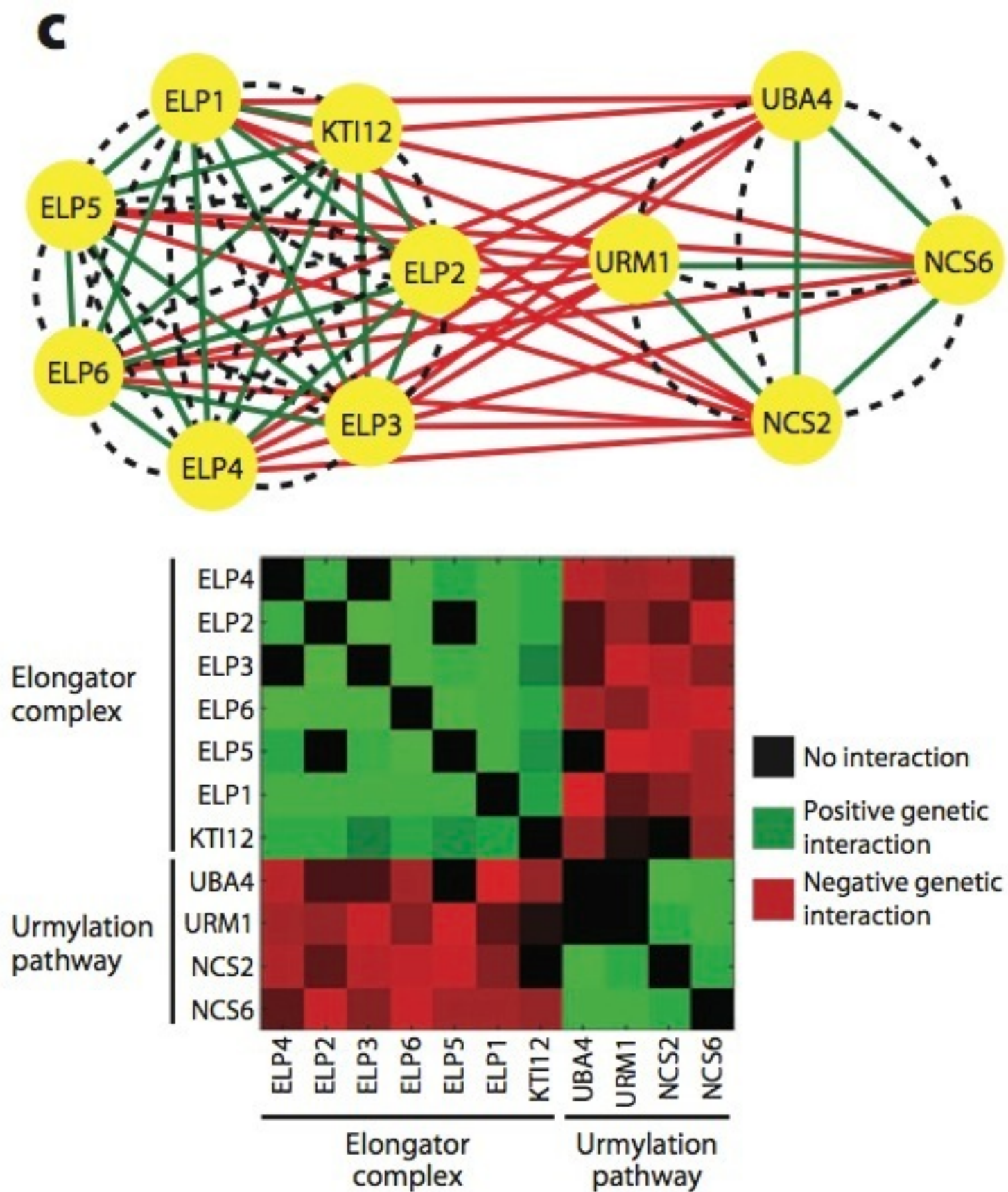
- Sieć interakcji syntetycznych letalnych jest rzadka – około 1%
- Interakcje syntetyczne są jednak częste pomiędzy genami o powiązanej funkcji (18%-25%)





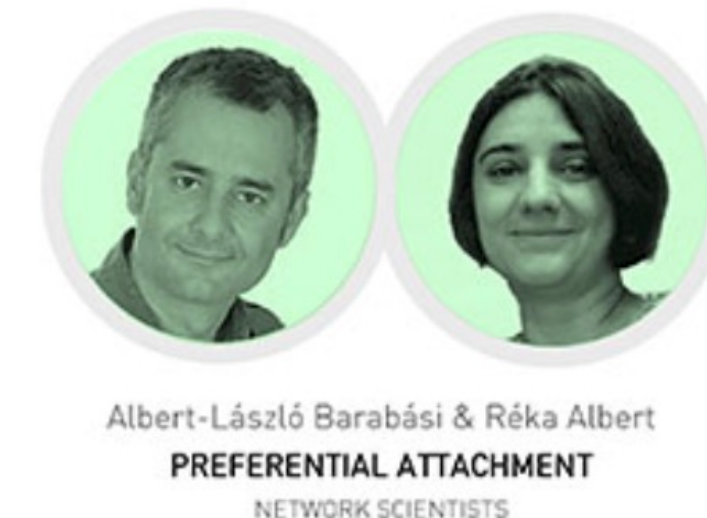
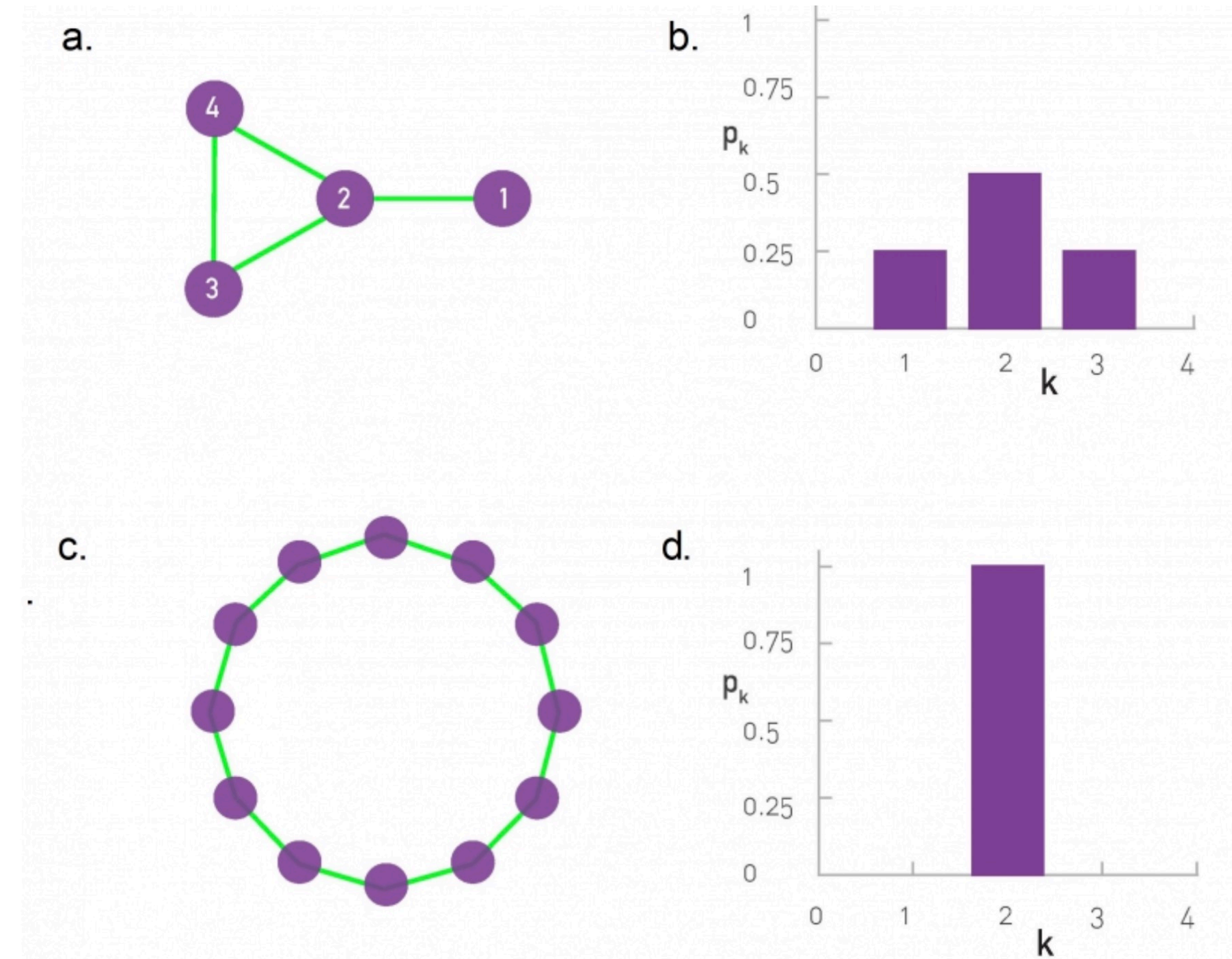
# Interakcje genetyczne a fizyczne

- Interakcje fizyczne i genetyczne rzadko się nakładają, choć częściej, niż przewidywano by dla pełnej losowości
- Nakładanie się interakcji genetycznych i fizycznych częste dla interakcji pozytywnych (epistaza)
- Interakcje negatywne z reguły pomiędzy różnymi kompleksami fizycznymi



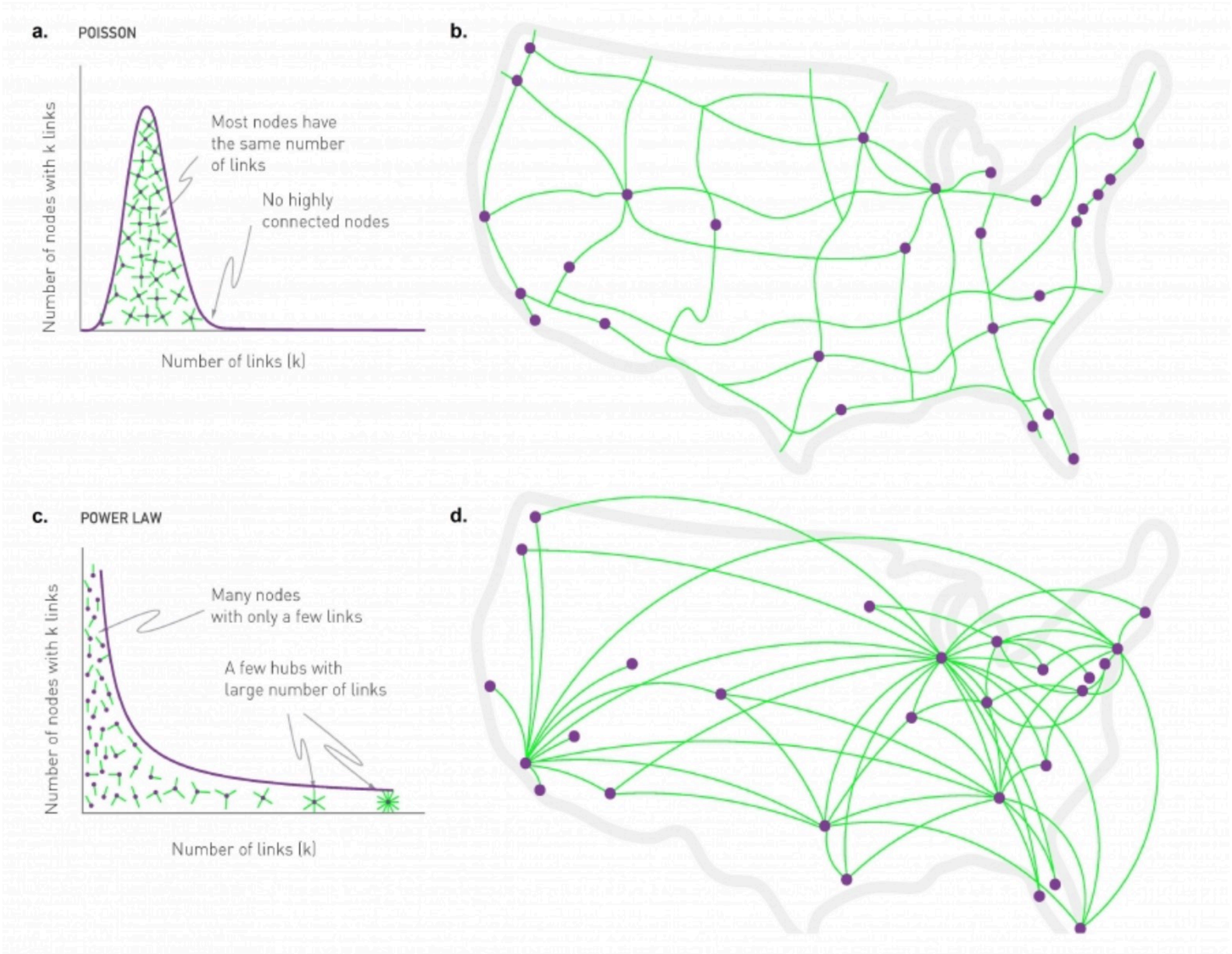
# Sieci biologiczne

- Zastosowanie pojęć teorii grafów i sieci
  - $N$  - liczba węzłów
  - $k$  - stopień węzła (liczba połączeń)
  - $L$  - całkowita liczba połączeń
  - $P(k)$  - rozkład prawdopodobieństwa znalezienia węzła o stopniu  $k$
- Najważniejsze odkrycie - opisanie sieci bezskalowych: Barabási & Albert.  
Emergence of scaling in random networks.  
Science, 286: 509, 1999.





# Sieci losowe i bezskalne

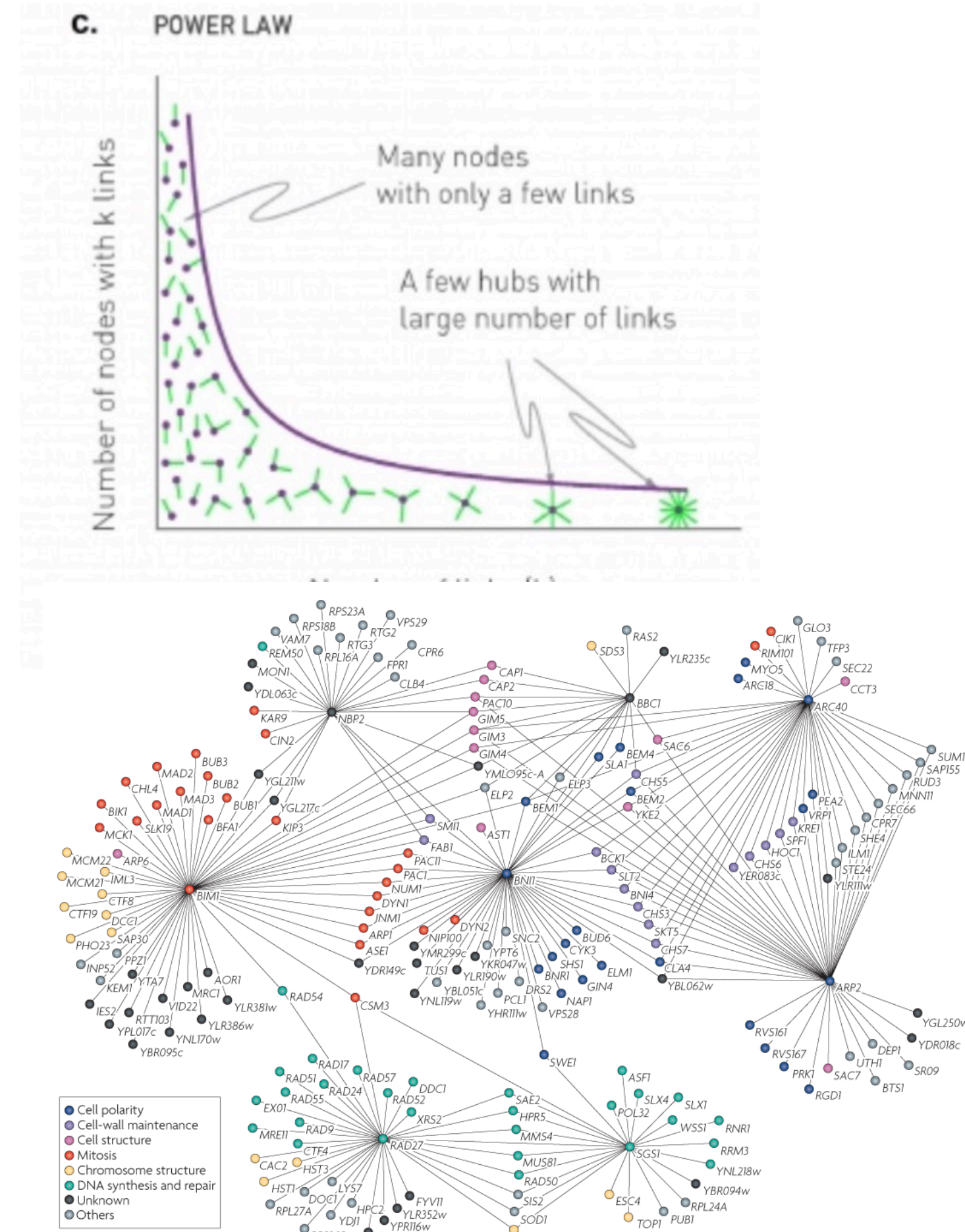


$$P(k) \sim \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(k) \sim k^{-\gamma}$$

# Sieci interakcji

- Sieci interakcji biologicznych mają charakter bezskalowy
- węzły centralne (hubs) z dużą liczbą połączeń
- węzły peryferyjne, z małą liczbą połączeń
- węzły centralne częściej odpowiadają genom niezbywalnym (których defekt jest letalny)



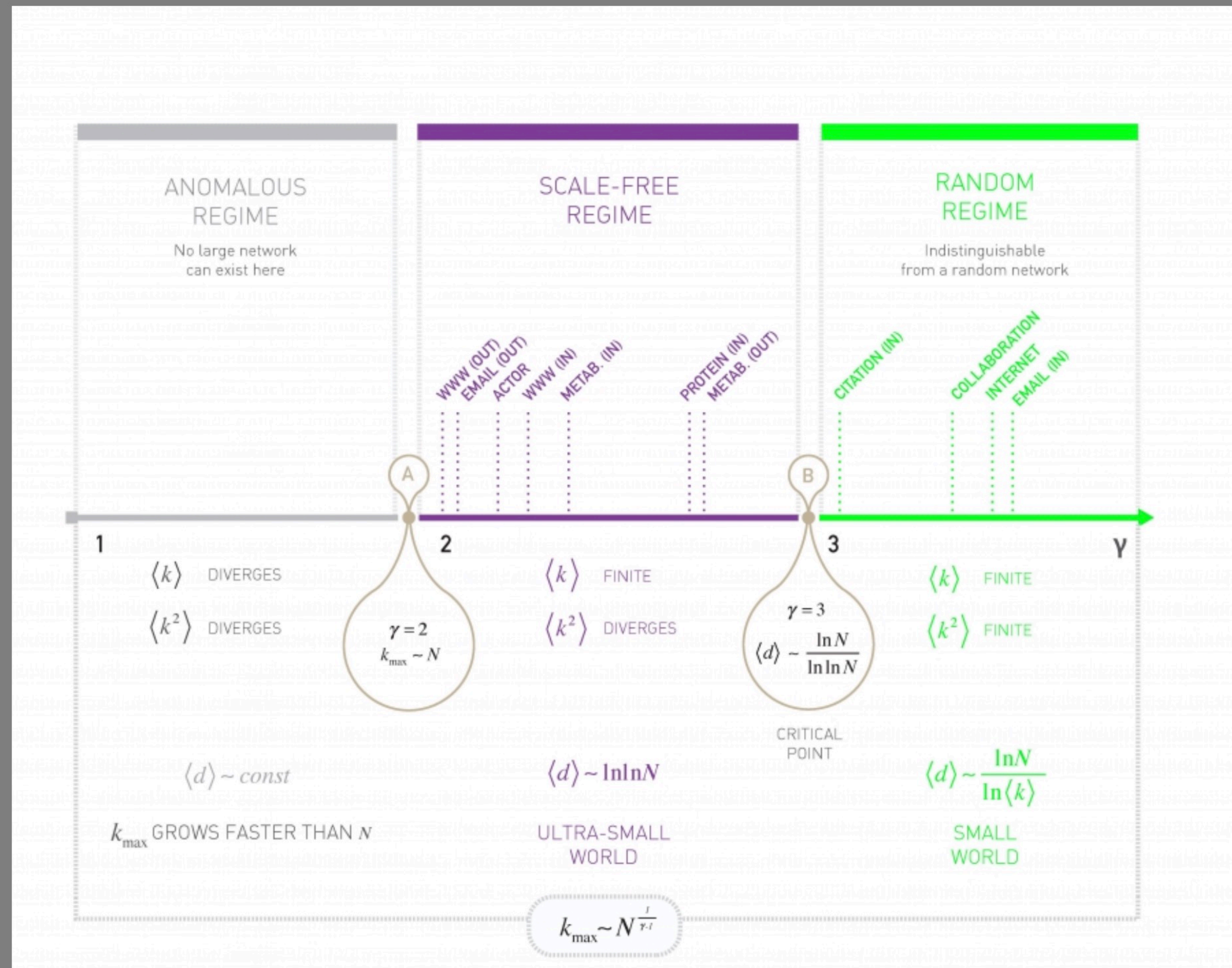


# Sieci biologiczne i inne

---

- Węzły centralne i peryferyjne
- “Mały świat” – długość najkrótszej ścieżki pomiędzy dwoma węzłami jest niewielka (~3,3 węzły u drożdży)
  - Niewielkie zwiększenie odległości przy zwiększaniu liczby węzłów (“ultra mały świat”)
- Podobne właściwości ma np. sieć połączeń lotniczych, Internet (WWW), sieci interakcji społecznych, liczba Erdősa wśród matematyków

# The $\gamma$ Dependent Properties of Scale-Free Networks

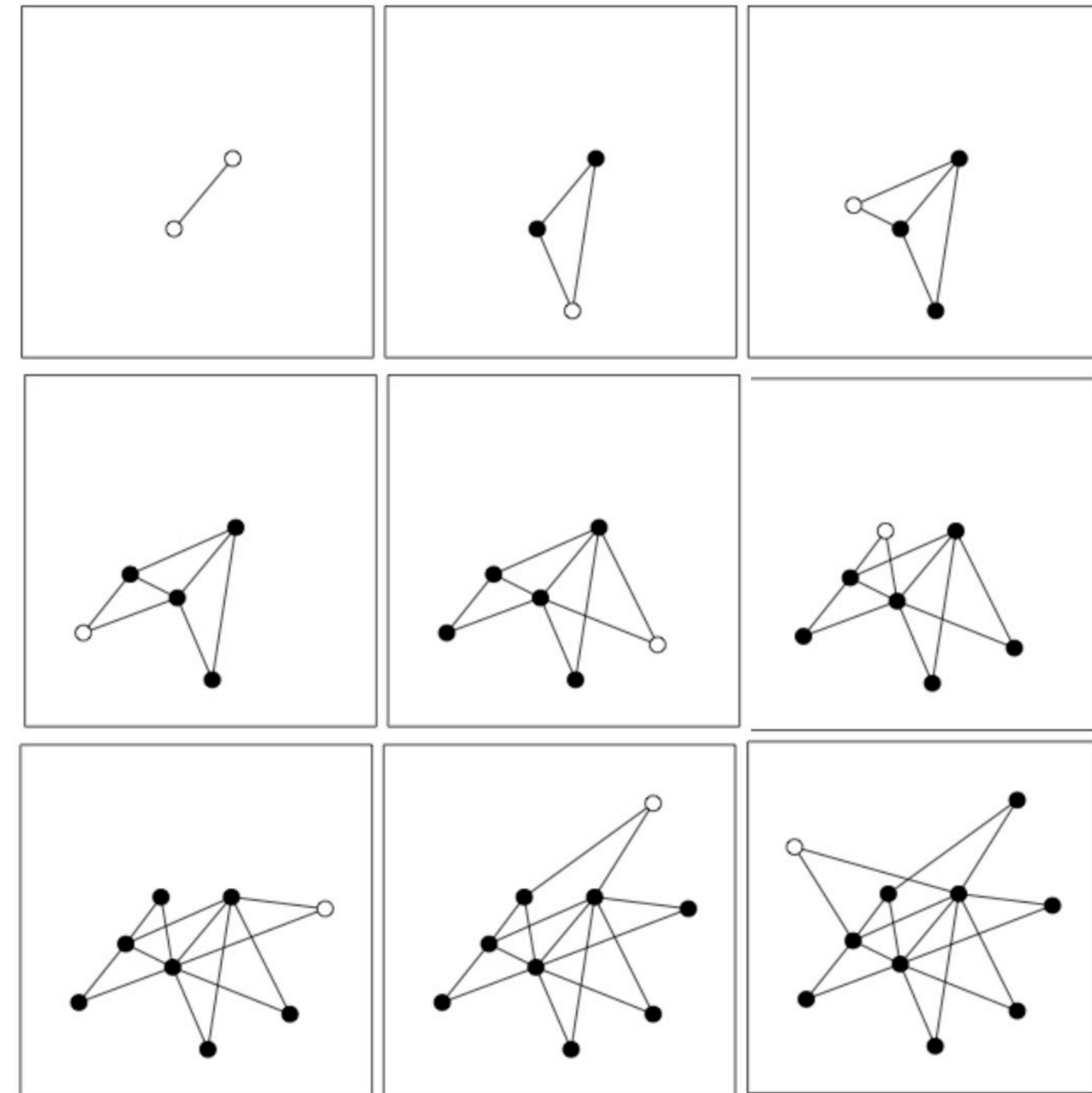


$$2 < \gamma < 3$$

# Ewolucja sieci bezskalowych

- Preferencyjne przyłączanie: model Barabásiego-Albert
- Nowy węzeł dołącza do istniejących z prawdopodobieństwem proporcjonalnym do stopnia węzła

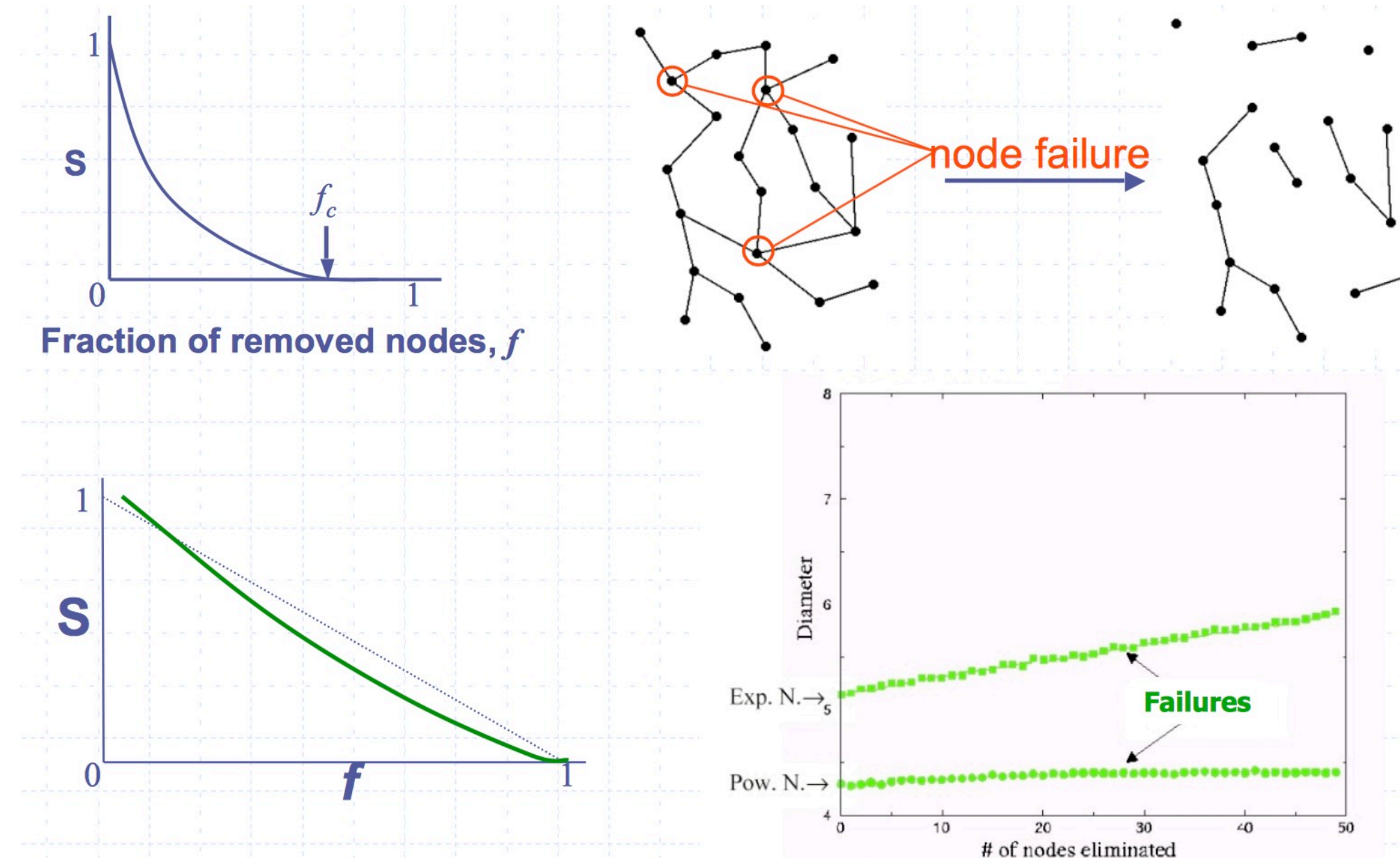
$$\Pi(k_i) = \frac{k_i}{\sum k_j}$$



# Krzepkość sieci

- Sieci bezskalowe są bardziej odporne na przypadkowe zaburzenia niż sieci losowe (wykładnicze)
- Są wrażliwe jeżeli atak skierowany jest na węzły centralne
- wykorzystanie znajomości sieci w projektowaniu leków itp.

**Error and attack tolerance of complex networks**  
Réka Albert, Hawoong Jeong and Albert-László Barabási  
*Nature* **406**, 378-382(27 July 2000)  
doi:10.1038/35019019





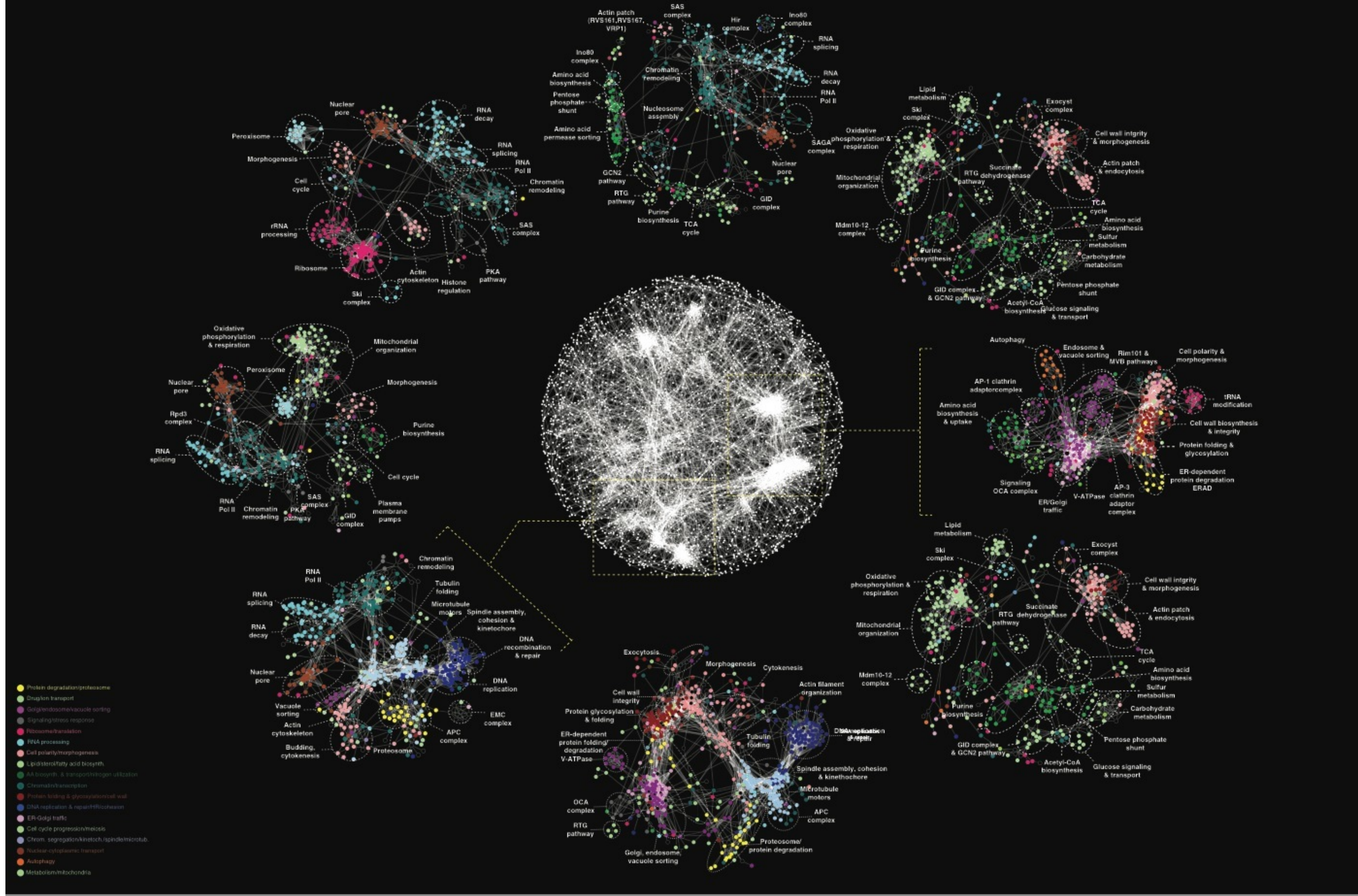
# Interakcje genetyczne a biologia systemów

---

- Badanie sieci interakcji funkcjonalnych na skalę całego organizmu to podstawa biologii systemów
- Interakcje genetyczne są ważnym elementem takiej sieci
  - Może nawet bardziej, niż interakcje fizyczne
  - Interakcje fizyczne identyfikują kompleksy, interakcje genetyczne mogą pokazać, w jakim kontekście te kompleksy funkcjonują
- Wszystkie dotychczasowe wyniki są bardzo niekompletne, nawet u drożdży
- Nie ma biologii systemów bez genetyki



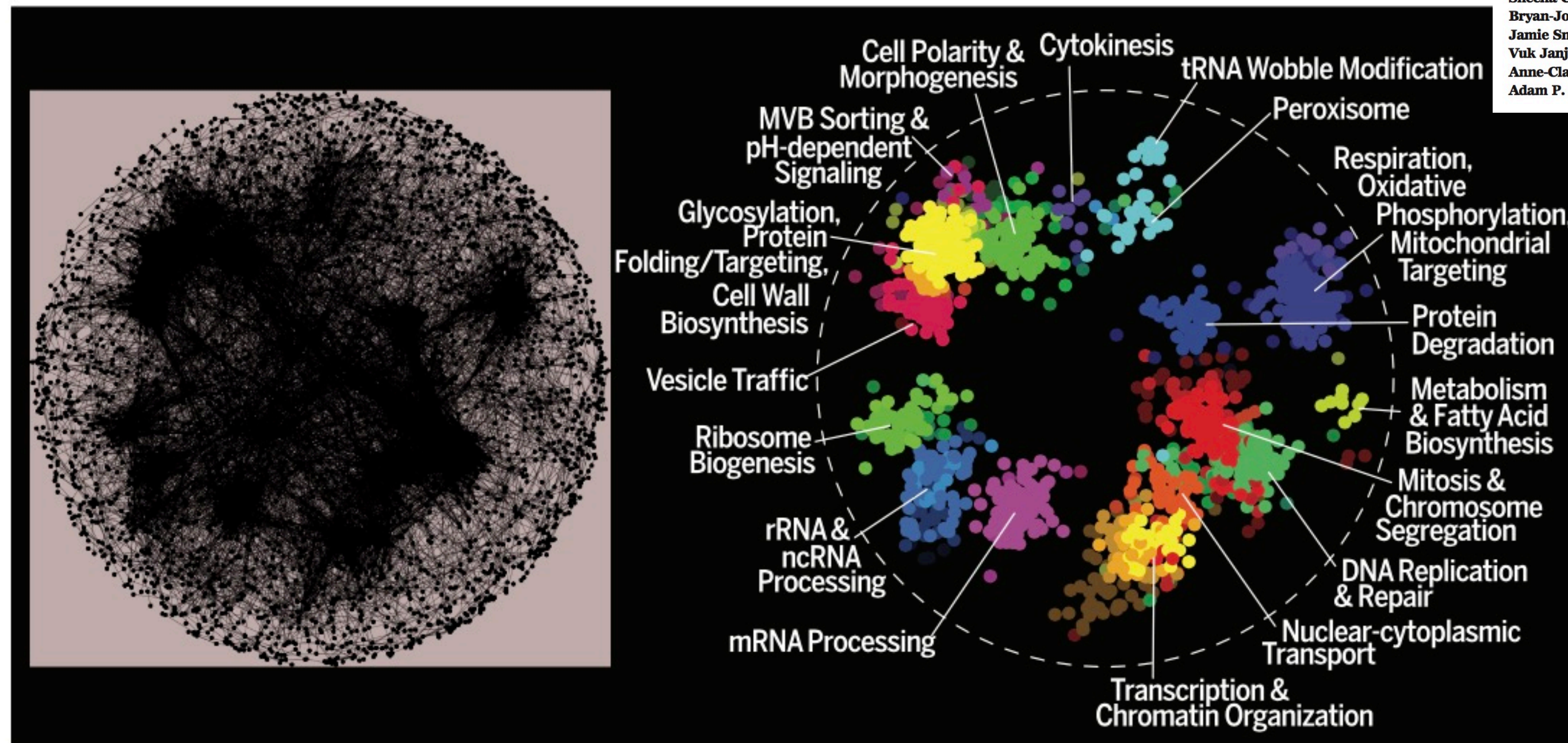
# The Genetic Landscape of a Cell





## A global genetic interaction network maps a wiring diagram of cellular function

Michael Costanzo,\* Benjamin VanderSluis,\* Elizabeth N. Koch,\* Anastasia Baryshnikova,\* Carles Pons,\* Guihong Tan,\* Wen Wang, Matej Usaj, Julia Hanchard, Susan D. Lee, Vicent Pelechano, Erin B. Styles, Maximilian Billmann, Jolanda van Leeuwen, Nydia van Dyk, Zhen-Yuan Lin, Elena Kuzmin, Justin Nelson, Jeff S. Piotrowski, Tharan Srikumar, Sondra Bahr, Yiqun Chen, Raamesh Deshpande, Christoph F. Kurat, Sheena C. Li, Zhijian Li, Mojca Mattiazzi Usaj, Hiroki Okada, Natasha Pascoe, Bryan-Joseph San Luis, Sara Sharifpoor, Emira Shuteriqi, Scott W. Simpkins, Jamie Snider, Harsha Garadi Suresh, Yizhao Tan, Hongwei Zhu, Noel Malod-Dognin, Vuk Janjic, Natasa Przulj, Olga G. Troyanskaya, Igor Stagljar, Tian Xia, Yoshikazu Ohya, Anne-Claude Gingras, Brian Raught, Michael Boutros, Lars M. Steinmetz, Claire L. Moore, Adam P. Rosebrock, Amy A. Caady, Chad L. Myers,† Brenda Andrews,† Charles Boone†

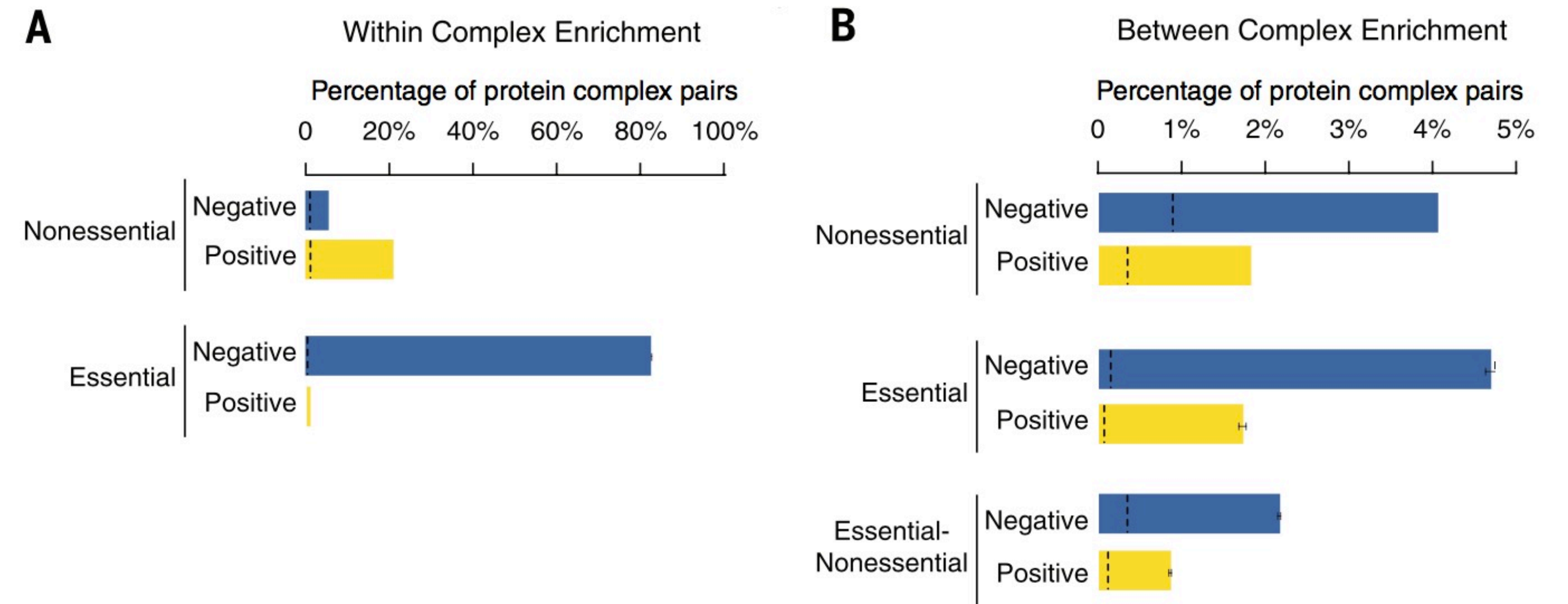


**A global network of genetic interaction profile similarities.** (Left) Genes with similar genetic interaction profiles are connected in a global network, such that genes exhibiting more similar profiles are located closer to each other, whereas genes with less similar profiles are positioned farther apart. (Right) Spatial analysis of functional enrichment was used to identify and color network regions enriched for similar Gene Ontology bioprocess terms.



# Interakcje genetyczne a fizyczne

- Produkty w różnych kompleksach - częstsze interakcje negatywne
- Produkty w tym samym kompleksie - pozytywne częstsze dla genów nie będących niezbywalnymi



**Fig. 8. Genetic interactions within and between protein complexes.** (A) The percentage of nonessential and essential complexes whose members were enriched for genetic interactions with each other and biased (i.e., coherent) for either mostly negative (blue) or mostly positive (yellow) interactions. (B) The percentage of nonessential-nonessential, essential-essential, or essential-nonessential complex-complex pairs found to be enriched for genetic interactions and biased (i.e., coherent) for either mostly negative (blue) or mostly positive (yellow) interactions. Black dashed lines indicate the background rate of coherent genetic interaction enrichment within individual complexes or between pairs of protein complexes. Error bars indicate the standard deviation across multiple samplings of different alleles for the same essential genes, where each gene is represented by a single, randomly selected allele in each sampling.

# Interakcje genetyczne a biologia systemów

---

- Badanie sieci interakcji funkcjonalnych na skalę całego organizmu to podstawa biologii systemów
- Interakcje genetyczne są ważnym elementem takiej sieci
  - Może nawet bardziej, niż interakcje fizyczne
  - Interakcje fizyczne identyfikują kompleksy, interakcje genetyczne mogą pokazać, w jakim kontekście te kompleksy funkcjonują
- Wszystkie dotychczasowe wyniki są bardzo niekompletne, nawet u drożdży

# Wyzwania

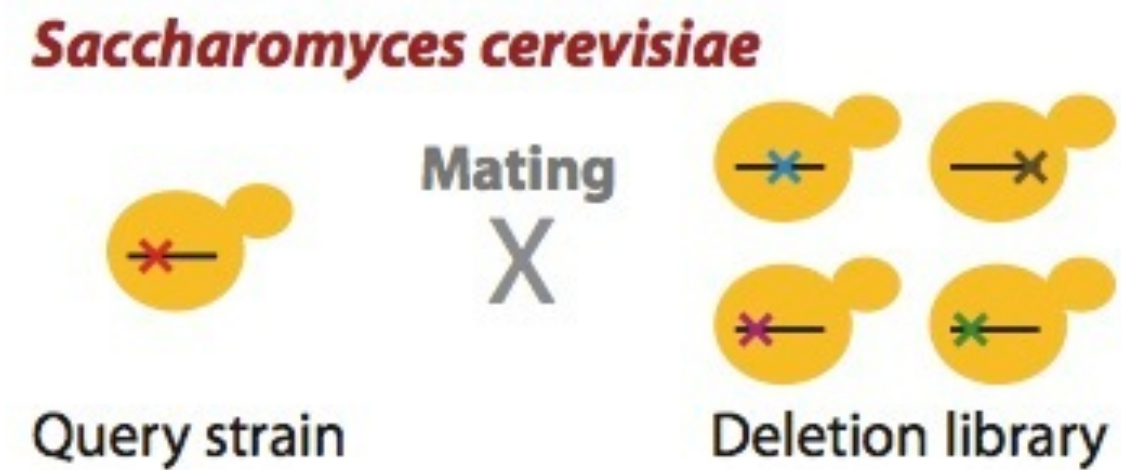
---

- Systematyczne badania interakcji genetycznych są obecnie w fazie początkowej
- Zagadnienia na przyszłość:
  - Oddziaływania wyższego rzędu niż podwójne (3 i więcej genów)
  - Wpływ środowiska i tła genetycznego
  - Allele inne, niż delecja (null) i nadekspresja – mniej ekstremalne formy zmienności genetycznej
  - Systematyczne analizy w innych, bardziej złożonych organizmach

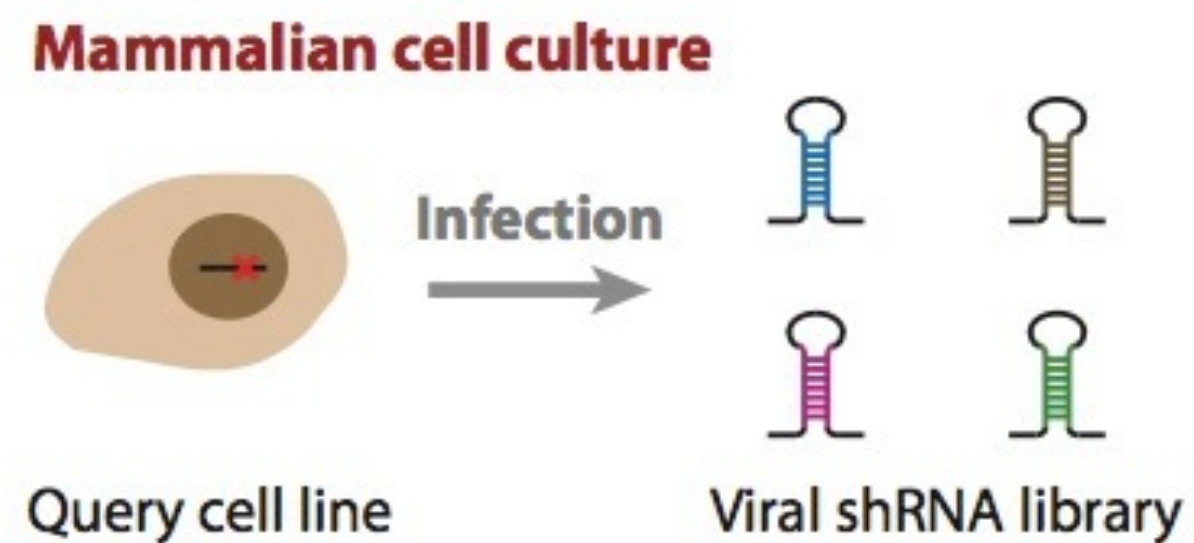
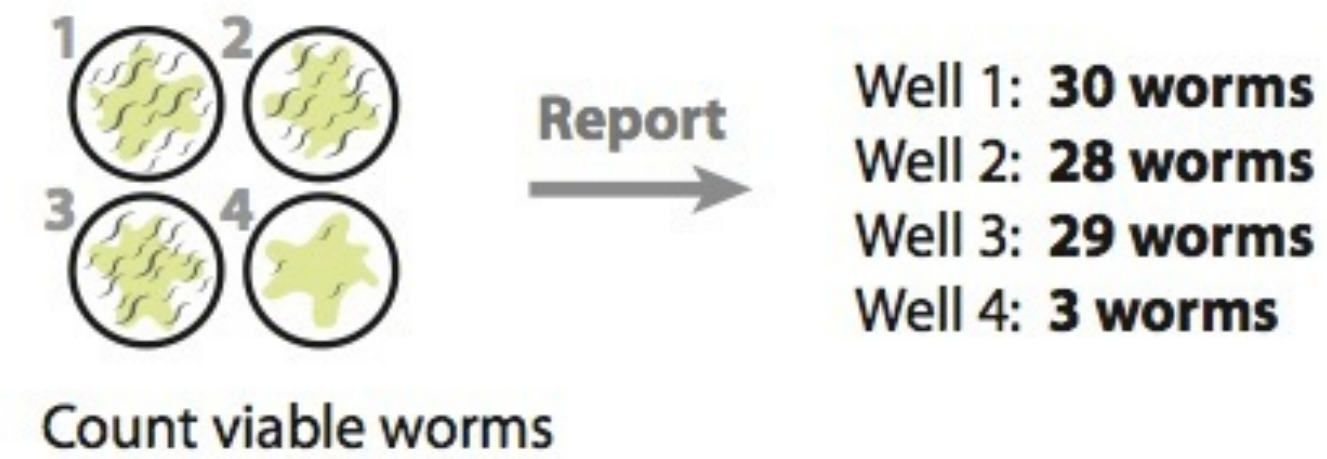
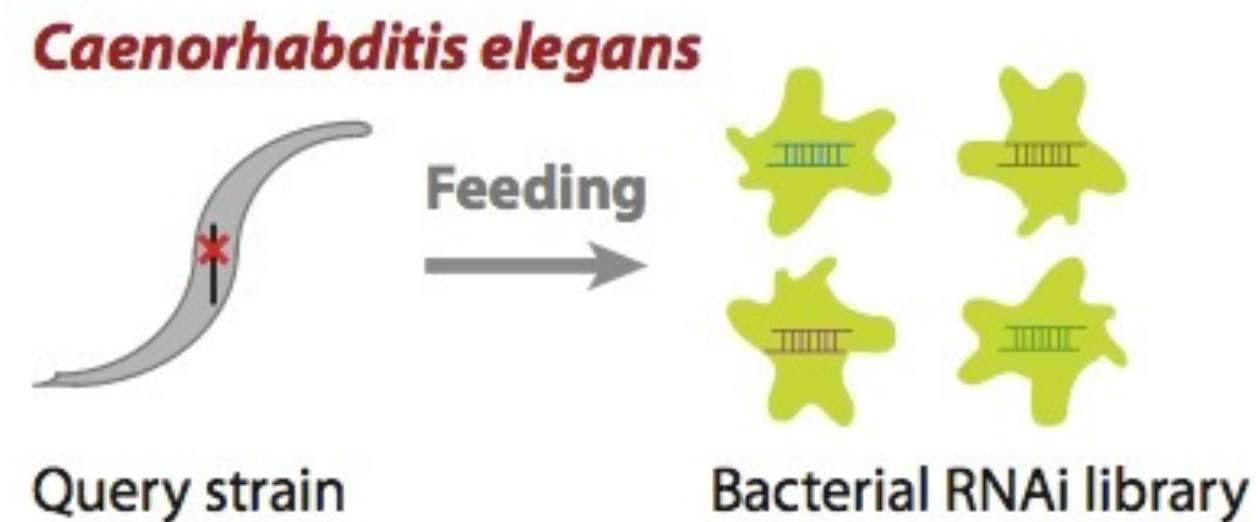
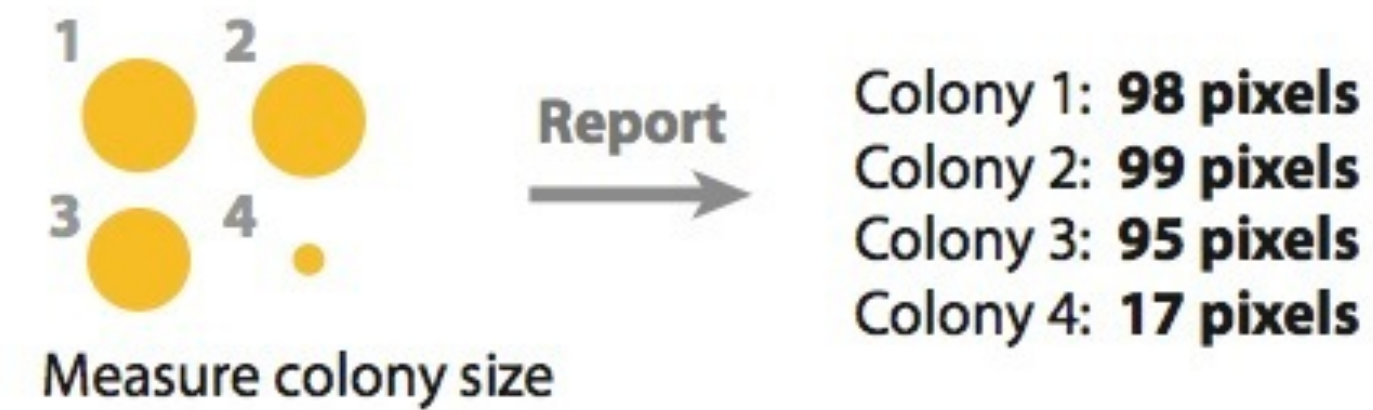


# Nie tylko drożdże

## Step 1: Generate double mutant



## Step 2: Score phenotype and identify interactions



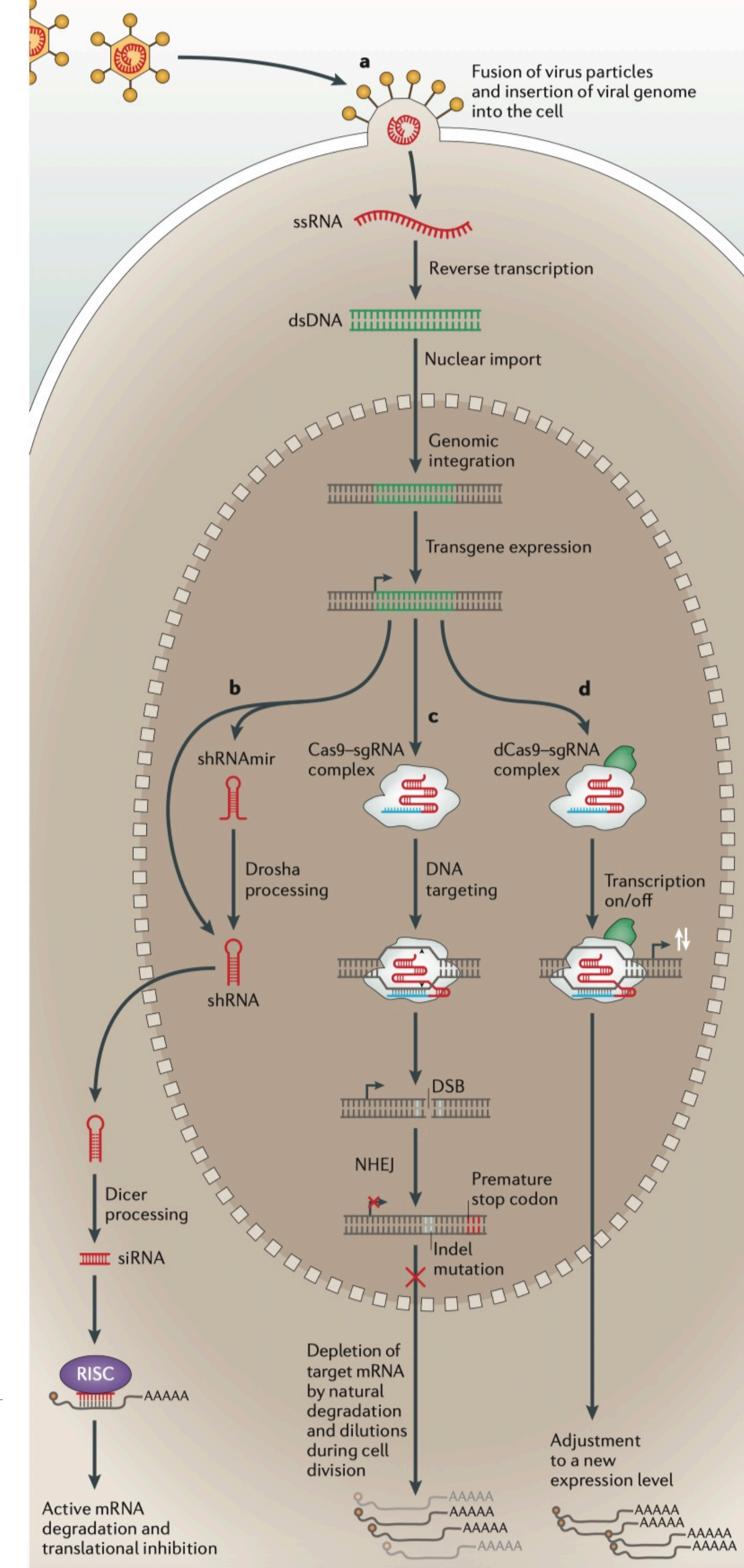


# Genomika funkcjonalna innych organizmów

- Wysokoprzepustowe wykorzystywanie metod odwrotnej genetyki
- interferencja RNA
- CRISPR-Cas9 i modyfikacje

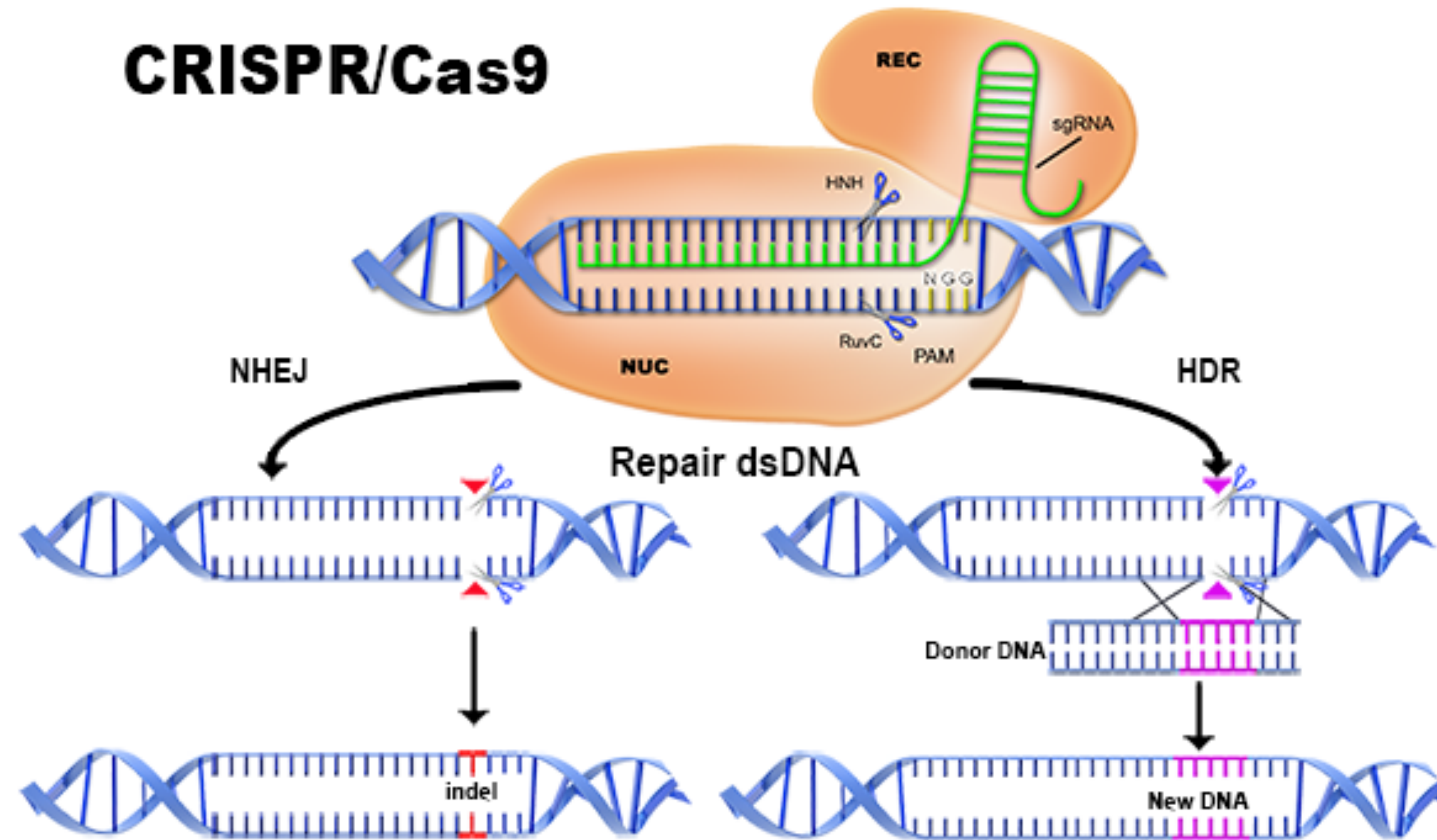
## High-throughput functional genomics using CRISPR-Cas9

Ophir Shalem, Neville E. Sanjana and Feng Zhang



# CRISPR-Cas9

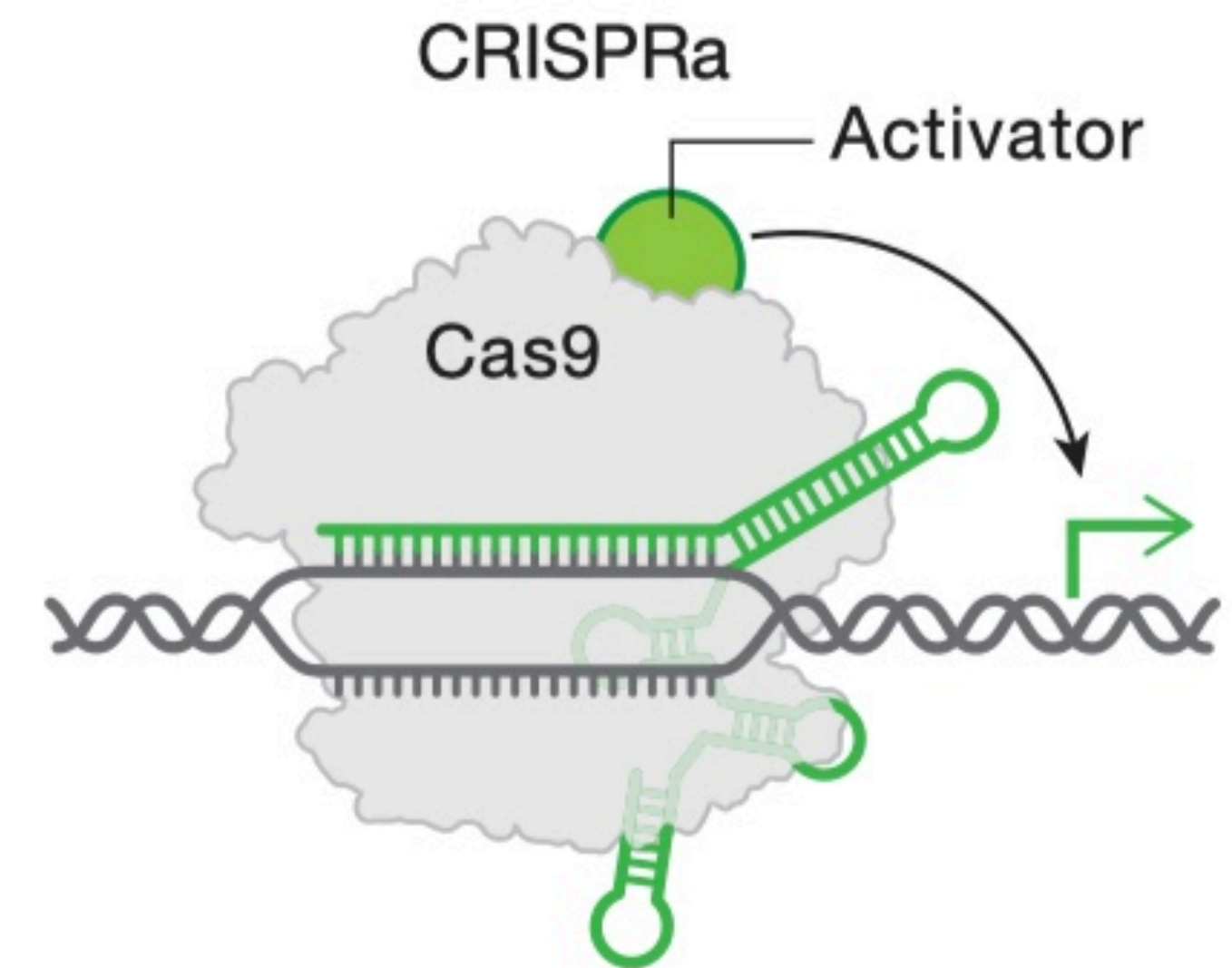
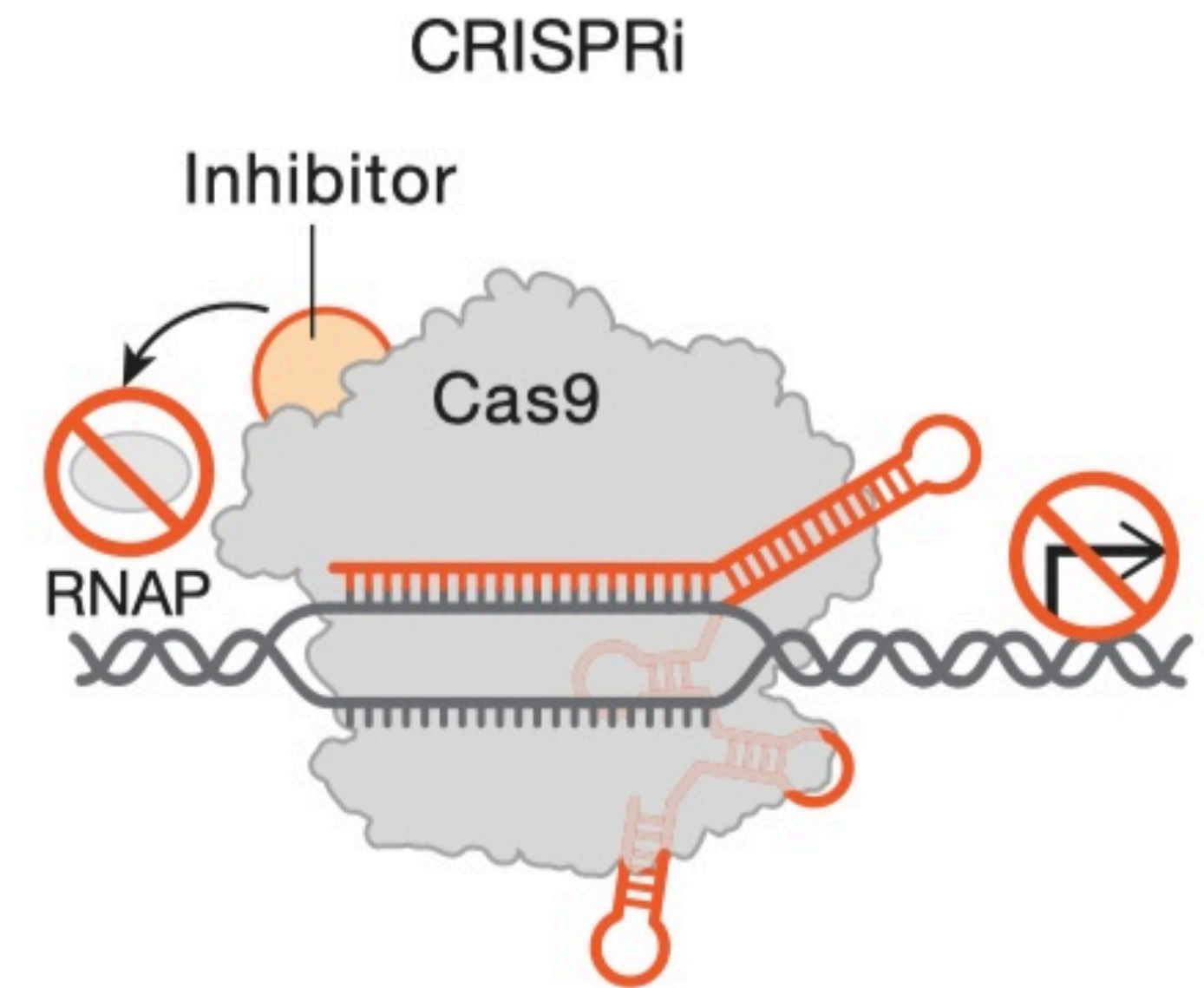
- Przecięty DNA - pęknięcie dwuniciowe (DSB- double strand break) - przez komórkę traktowane jako poważne uszkodzenie
- Indukowane systemy naprawcze
- Powstaje mała delecja
- W obecności donorowego DNA naprawa rekombinacyjna - wstawienie zmodyfikowanej sekwencji





# Modyfikacje - regulacja

- Cas9 nieaktywna - nie tnie DNA
- Związana domena aktywująca lub hamująca transkrypcje
- Związane domeny wprowadzające modyfikacje epigenetyczne (np. metylacja DNA) - dopiero rozwijane



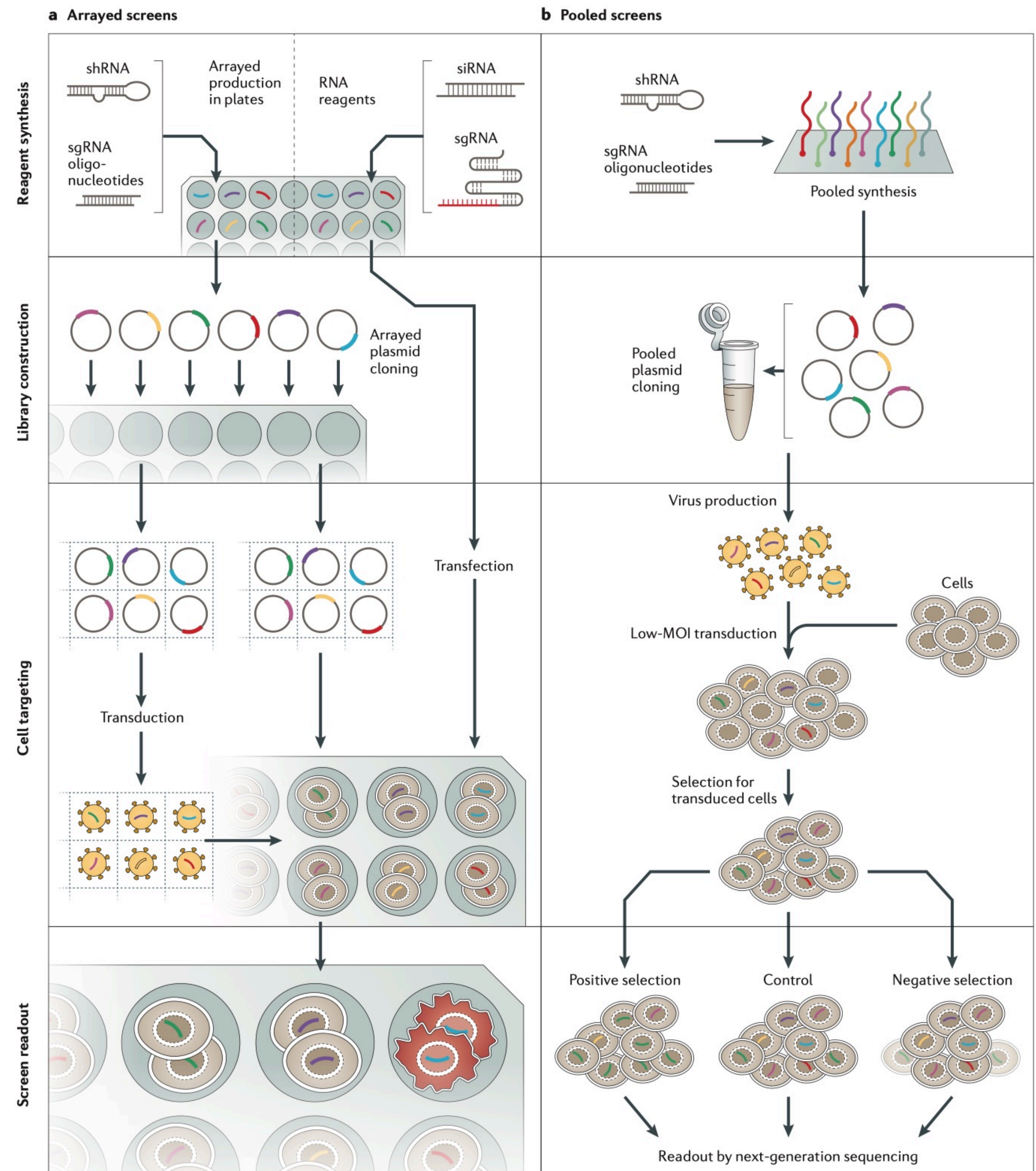


# Strategie screeningu

- screen macierzowy
- konstrukty przygotowywane i transfekowane osobno
- screen łączony
- konstrukty jako pula, po transfekcji klony identyfikowane przez NGS (rozpoznawany konkretny konstrukt)

## High-throughput functional genomics using CRISPR–Cas9

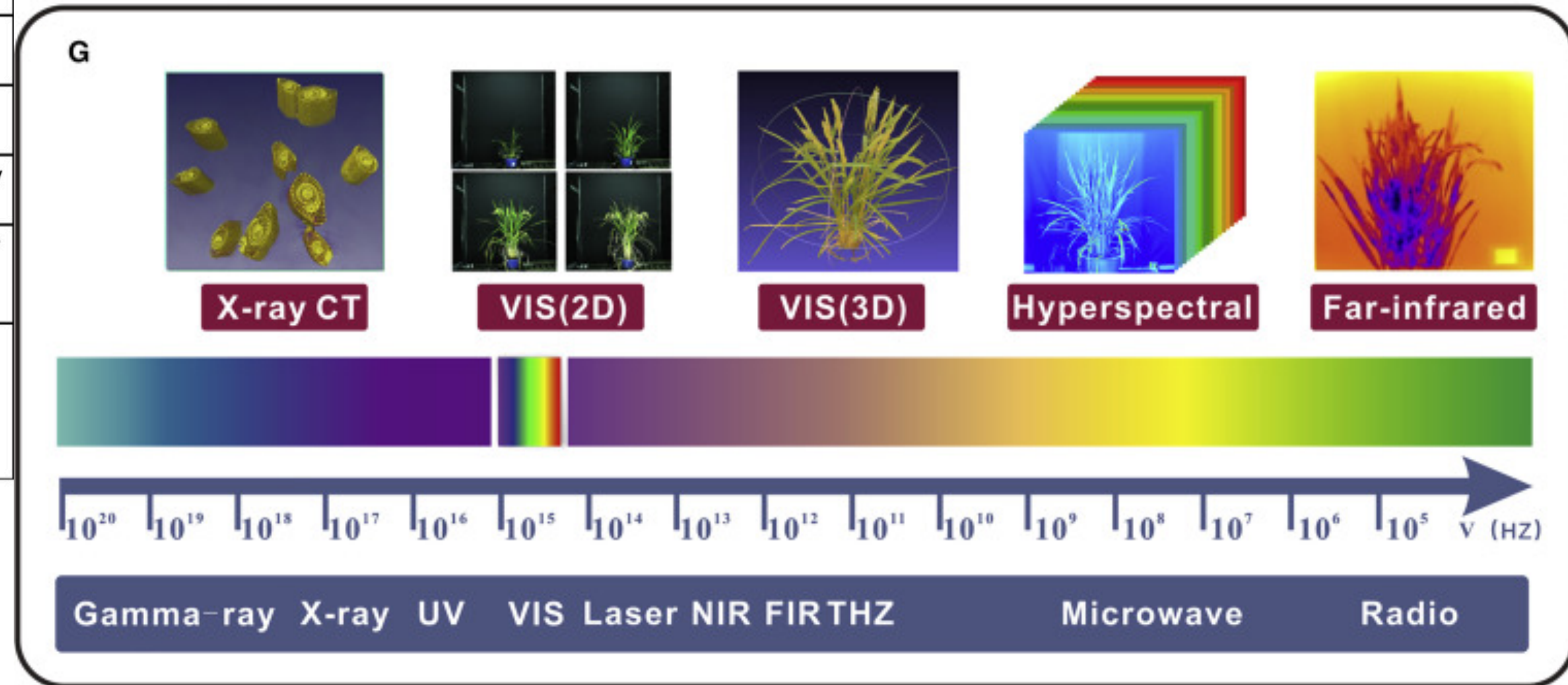
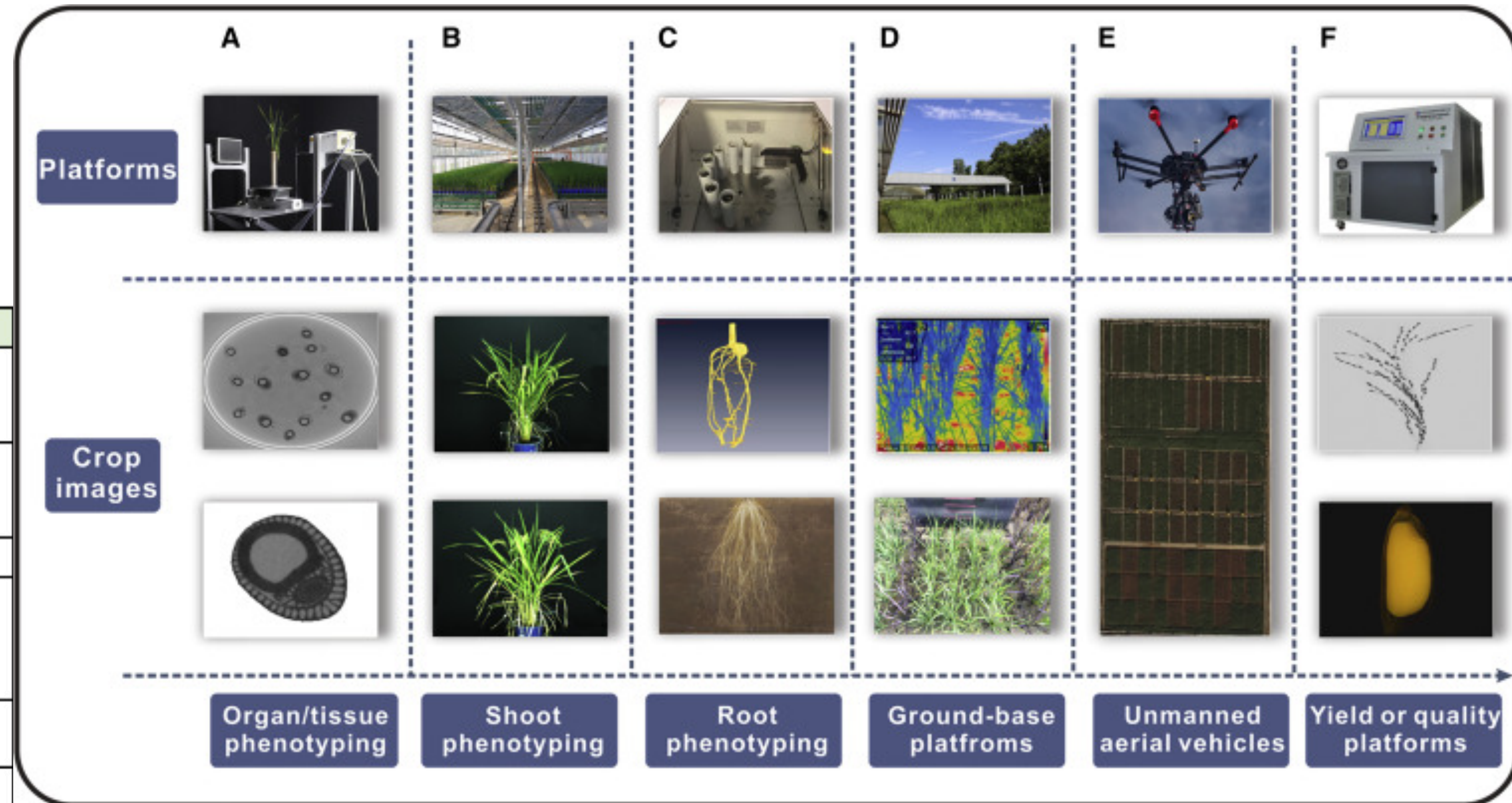
*Ophir Shalem, Neville E. Sanjana and Feng Zhang*





# Analizy fenotipu u rošlin

Applications	Platforms
Shoot phenotyping for <i>Arabidopsis</i> in the laboratory	PHENOPSIS (DB); GROWSCREEN (FLUORO); Phenoscope; Phenovator; PlantScreen
Shoot phenotyping for crops in the laboratory	TraitMill; Scanalyzer3D; PHENOARCH; HRPF
	PhenoBox
Root phenotyping in the laboratory	PlaRoM; Rhizoslides; Rhizoponics; RADIX; RhizoTubes
	GiARoots RootReader3D
	GROWSCREEN-Rhizo
	MRI-PET; PET-CT; MRI-CT
Ground-based phenotyping in the field	CPRS, a fixed phenotyping tower
	Field Scanalyzer, a rail-based gantry phenotyping system
	BreedVision, a self-propelled tractor equipped with multiple sensors mounted in a mobile dark chamber
Remote sensing in the field	Drones or UAVs equipped with multiple sensors





# Genomika cech wieloczynnikowych

---

- Dziedziczenie wieloczynnikowe - fenotyp zależy od interakcji alleli wielu genów oraz środowiska
- W odróżnieniu od fenotypów mendlowskich mają zwykle charakter zmienności ciągłej (ilościowej), a nie dyskretnej
- QTL - *Quantitative Trait Loci* - loci cech ilościowych - obszary genomu w istotny sposób wpływające na fenotyp cechy wieloczynnikowej

# Genomika cech wieloczynnikowych

---

- Ważne zagadnienie dla
  - genetyki człowieka
    - częste choroby, zmienność prawidłowa
  - genetyki roślin uprawnych
  - genetyki zwierząt hodowlanych
  - teorii ewolucji
- Głównie analizy statystyczne
- Czy można zastosować proste organizmy modelowe?



# QTL u drożdży

---

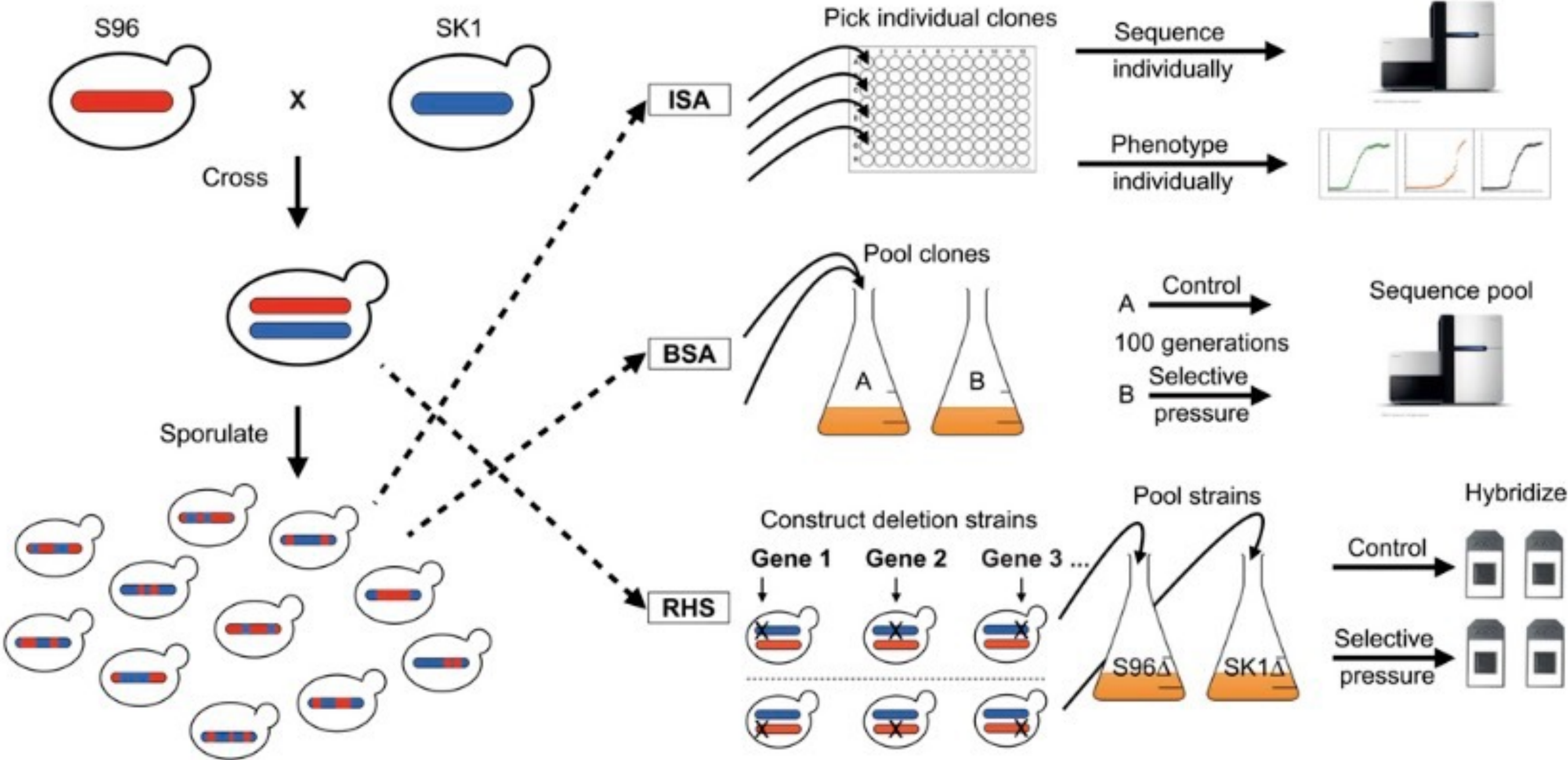
- QTL u drożdży można badać wykorzystując szczepy o różnym tle genetycznym
- *S. cerevisiae* - bardzo duża zmienność, nawet wśród szczepów laboratoryjnych

## QTL u drożdży - 3 strategie

---

- Bulk Segregant Analysis (BSA) - masowa analiza segregantów
- Individual Segregant Analysis (ISA) - analiza pojedynczych segregantów (równoległa)
- Reciprocal Hemizygosity Scanning (RHS) - analiza hemizygotycznych delecji

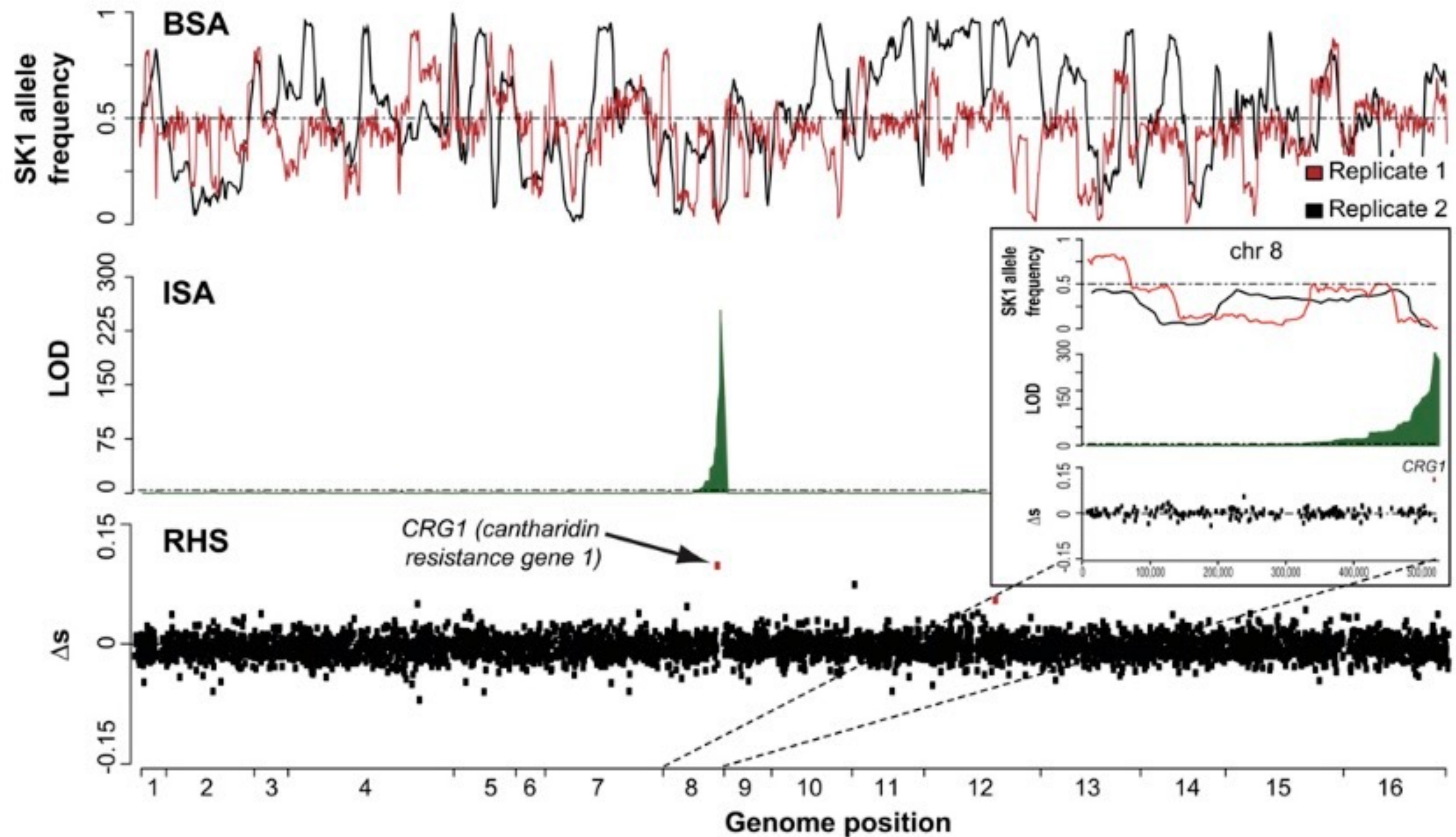
# QTL u drożdży



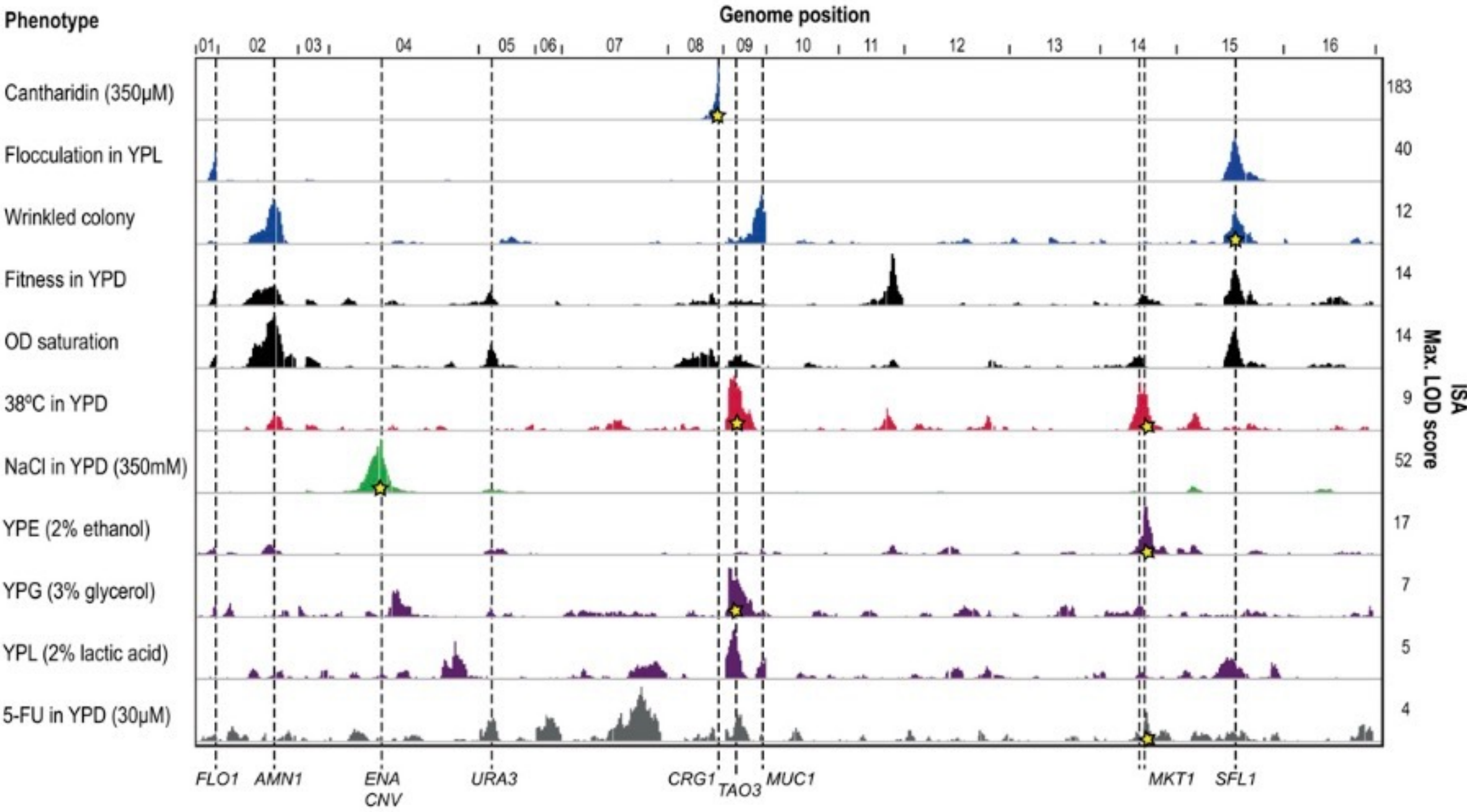
Wilkening et al. (2014), Genetics, 196:853-865



# QTL u drożdży



# QTL u drożdży - metoda ISA



# QTL u drożdży

---

- Metoda ISA pozwala na mapowanie cech, których nie można selekcjonować (np. kształt kolonii)
- Skuteczność zależy od liczby pojedynczych segregantów, które można przeanalizować
- Obecnie do ~1000