

**„Nowoczesne metody, leki i terapie w ochronie zdrowia i gospodarce Europy XXI wieku – interdyscyplinarne kształcenie w obszarze nauk biomedycznych na studiach II i III stopnia”, POKL.04.03.00-00-060/12**

Zadanie nr 1: Przygotowanie, otwarcie i realizacja trzech nowych specjalności na studiach II stopnia na kierunku Biotechnologia

Materiały dydaktyczne do wykładu w ramach przedmiotu:

**Matematyczne metody w naukach biomedycznych**  
dla Specjalności:

Łączna liczba godzin dydaktycznych: 30

**Autorzy: Anna Gambin, Urszula Foryś, Jacek Miękiś, Bartosz Wilczyński**

Projekt realizowany ze środków Unii Europejskiej  
w ramach Europejskiego Funduszu Społecznego

# PLAN

wykład 1: wprowadzenie;

**wykłady 2-4:** matematyczny model ekspresji genów  
(**dr hab. Jacek Miękiś**);

**wykłady 5-7:** algorytmy detekcji i analizy strukturalnych  
zmian w genomie człowieka (**dr hab. Anna Gambin**);

**wykłady 8-10:** modele aktywności sekwencji  
regulatorowych (**dr Bartosz Wilczyński**);

**wykłady 11-13:** modele procesów nowotworowych  
(**dr hab. Urszula Foryś**).

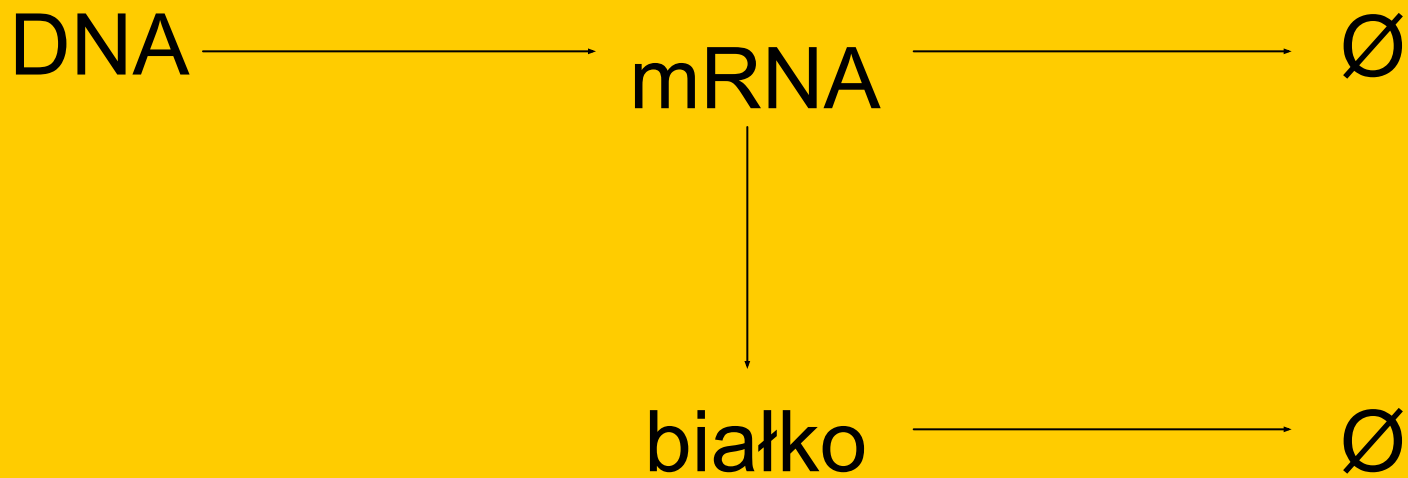
# Matematyczny model ekspresji genów

Jacek Miękiś  
Instytut Matematyki Stosowanej i Mechaniki  
Uniwersytet Warszawski

## Cele wykładu

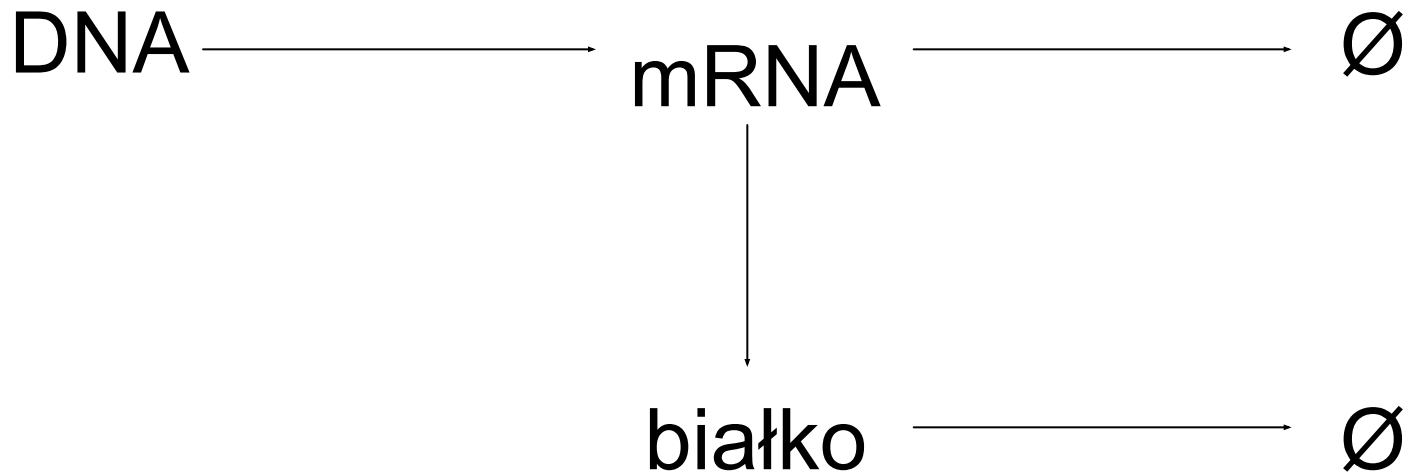
- 1 Wprowadzenie podstawowych koncepcji matematycznych niezbędnych w ilościowym opisie zachowania się komórki biologicznej- układy równań różniczkowych zwyczajnych (kinetyka chemiczna) oraz proste procesy stochastyczne typu urodzin i śmierci (ekspresja genu czyli produkcja białka).
2. Zastosowanie metod matematycznych do opisu procesów biochemicznych zachodzących w komórce – prosty model produkcji białka w komórce.
3. Dyskusja wyników matematycznych. Uzasadnienie celowości wyprowadzania wzorów czyli rozwiązań matematycznych.

# Komórka matematyczna

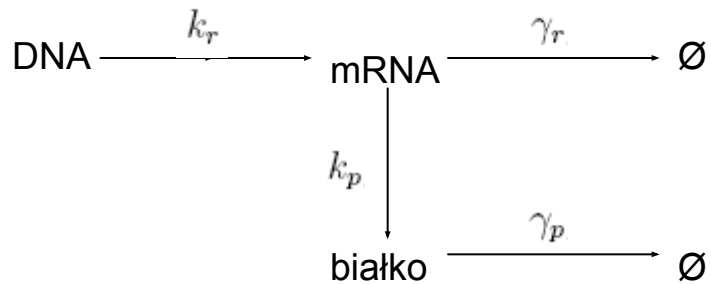


# Elementarne procesy biochemiczne

transkrypcja, translacja, degradacja mRNA i białka



# Poziom makroskopowy



$\rho_r$  - gęstość mRNA

$\rho_p$  - gęstość białka

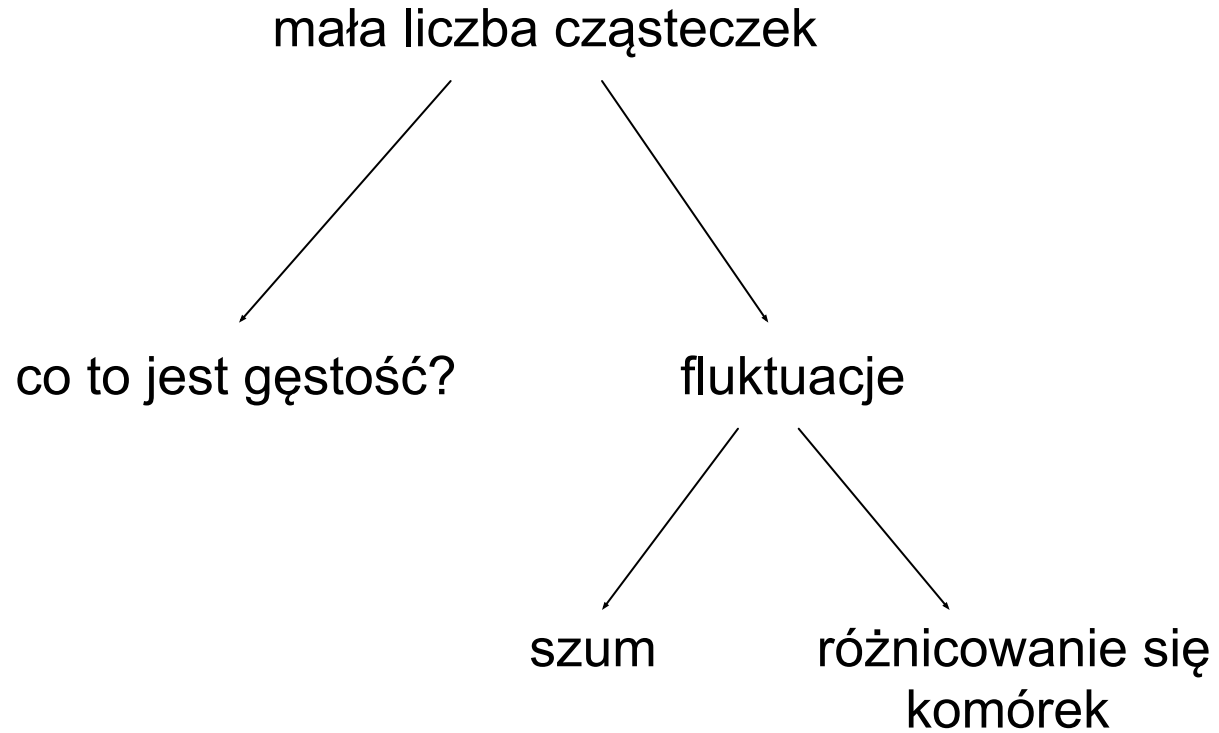
W stanie stacjonarnym mamy

$$\frac{d\rho_r}{dt} = k_r - \gamma_r \rho_r$$

$$\frac{d\rho_p}{dt} = k_p \rho_r - \gamma_p \rho_p$$

$$\rho_r = \frac{k_r}{\gamma_r} \quad \rho_p = \frac{k_p k_r}{\gamma_p \gamma_r}$$

# Problem





# Poziom mikroskopowy

$r$  - liczba cząsteczek mRNA

$p$  - liczba cząsteczek białka

prawdopodobieństwo zajścia w czasie  $(t, t+h)$

|                           |                         |                       |
|---------------------------|-------------------------|-----------------------|
| transkrypcji              | $(r \rightarrow r + 1)$ | $k_r h + o(h)$        |
| translacji                | $(p \rightarrow p + 1)$ | $r k_p h + o(h)$      |
| degradacji mRNA           | $(r \rightarrow r - 1)$ | $r \gamma_r h + o(h)$ |
| degradacji białka         | $(p \rightarrow p - 1)$ | $p \gamma_p h + o(h)$ |
| więcej niż jednej reakcji |                         | $o(h)$                |

proces urodzin i śmierci

Naszym celem jest obliczyć

wartość średnią liczby cząsteczek białka  $\langle p \rangle$

wariancję liczby cząsteczek białka

$$\text{Var}(p) = \langle (p - \langle p \rangle)^2 \rangle = \langle p^2 \rangle - \langle p \rangle^2$$

Ostatecznie dostajemy analityczne wyrażenie na wariancję RNA i białka

$$\text{var}(r) = \langle r \rangle$$

$$\frac{\text{var}(p)}{\langle p \rangle} = 1 + \frac{k_p}{\gamma_r + \gamma_p}$$

Thattai and Oudenaarden, PNAS 2001

# **Metody detekcji i analizy patogennych zmian strukturalnych w genomie człowieka**

Anna Gambin,  
Instytut Informatyki,  
Uniwersytet Warszawski

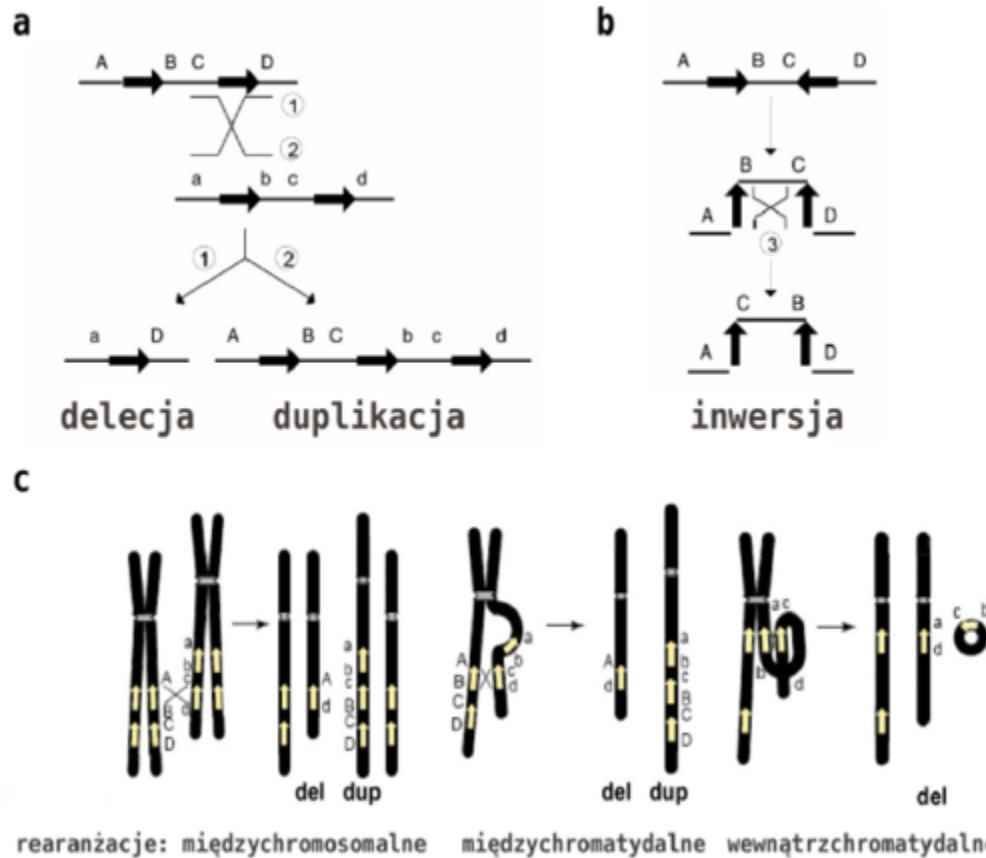
# PLAN WYKŁADU

- molekularne mechanizmy powstawania rearanżacji genomowych
- detekcja regionów narażonych na zmiany strukturalne
- analiza częstości zmian - korelacja z architekturą genomu: model regresji Poissona

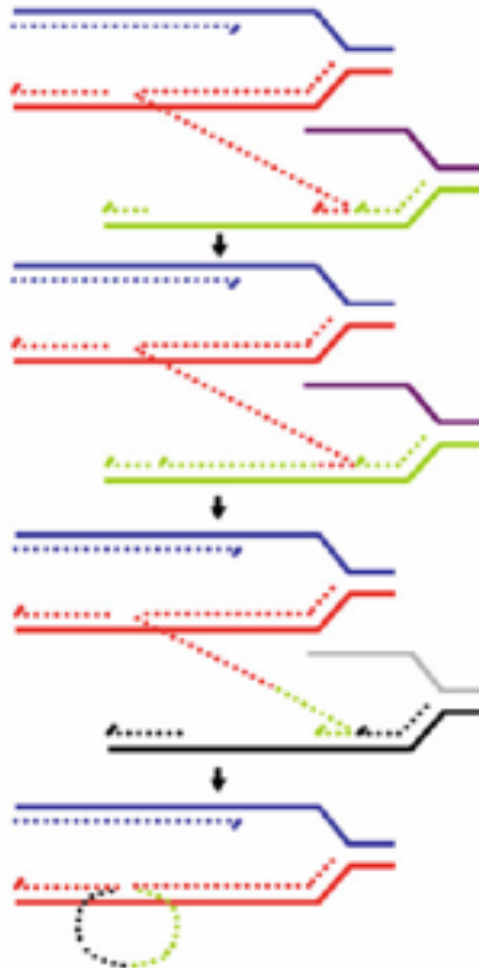
# CELE WYKŁADU

- poznanie podstawowych mechanizmów powstawania rearanżacji genomowych
- prezentacja algorytmów pozwalających na detekcję zmian strukturalnych i regionów narażonych na te zmiany
- modelowanie zależności między architekturą genomu, a częstością nawracających rearanżacji

# mechanizm NAHR: niealleliczna homologiczna rekombinacja

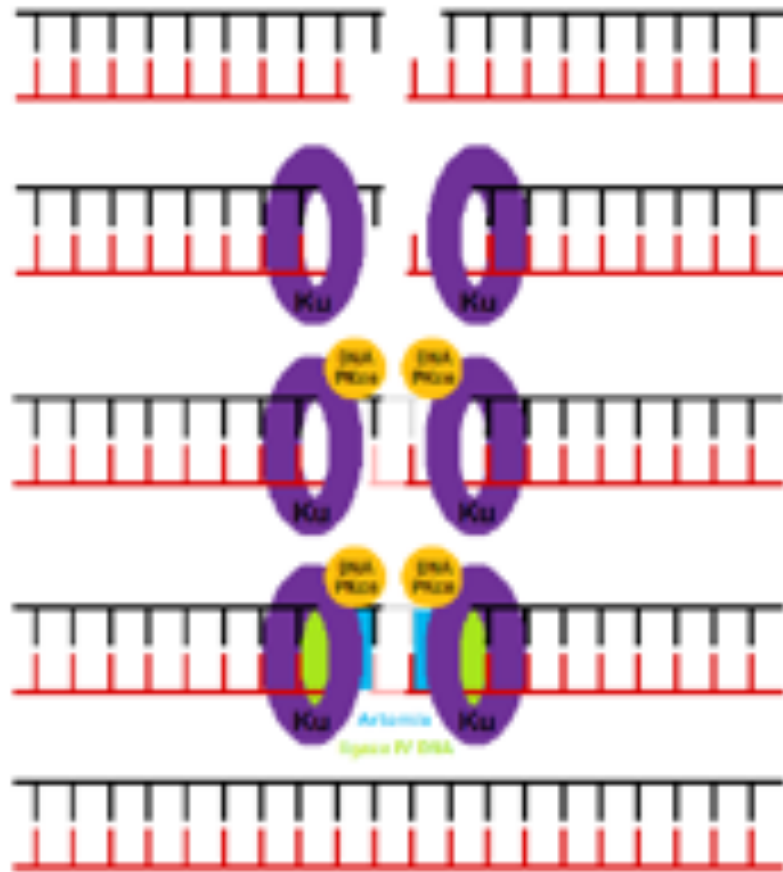


# bardziej złożone rearanżacje: mechanizm FoSTeS

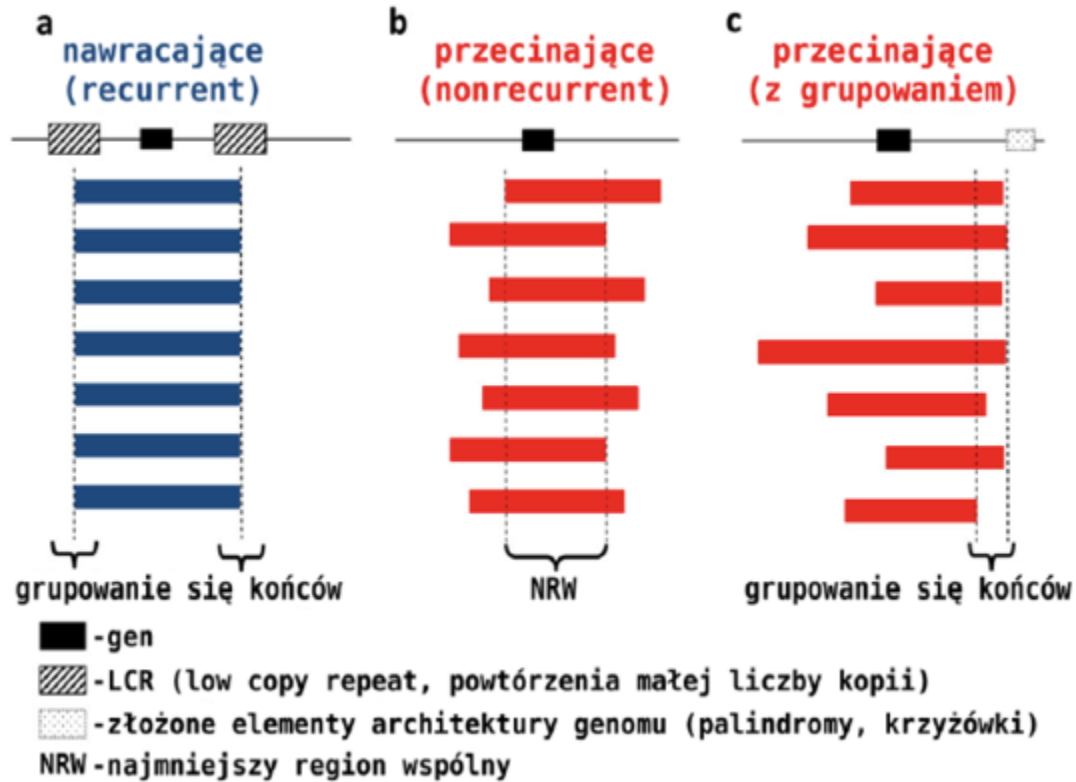




# mechanizm NHEJ



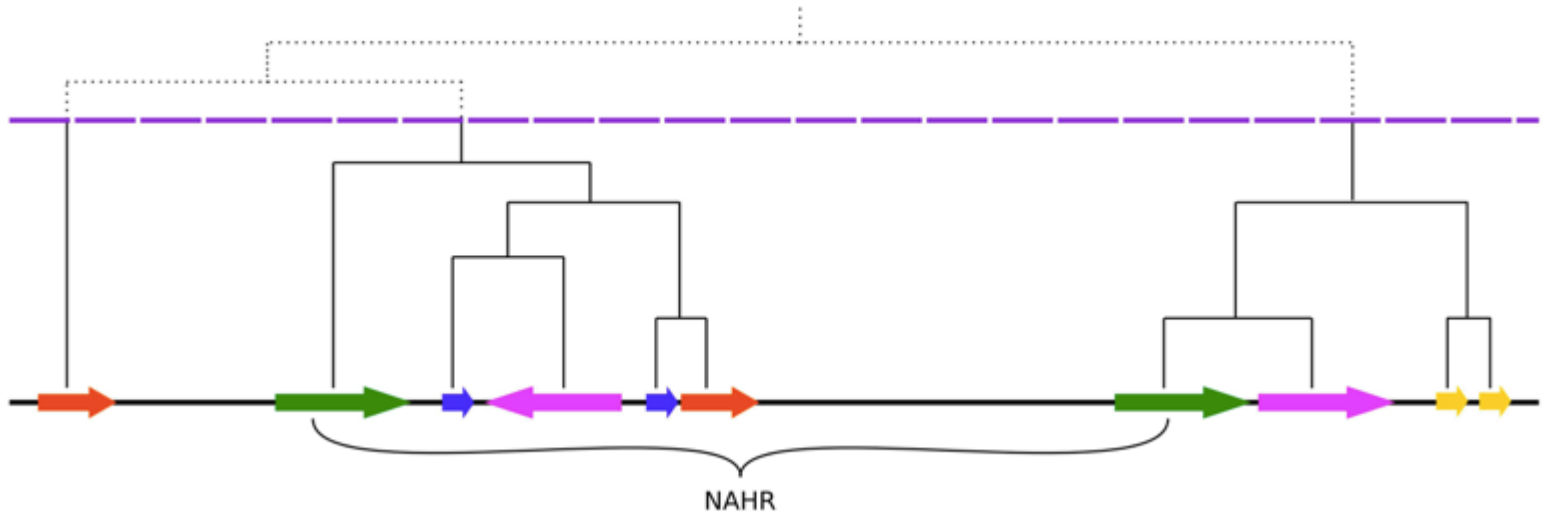
# zmiany patogenne



# detekcja niestabilnych rejonów

- pojęcie powtórzeń o niskiej liczbie kopii typu LCR oraz segmentalnych duplikacji.
- detekcja: algorytm MUMMER (tablice suffiksowe)
- identyfikacja regionów niestabilnych poprzez klastrowania elementów LCR

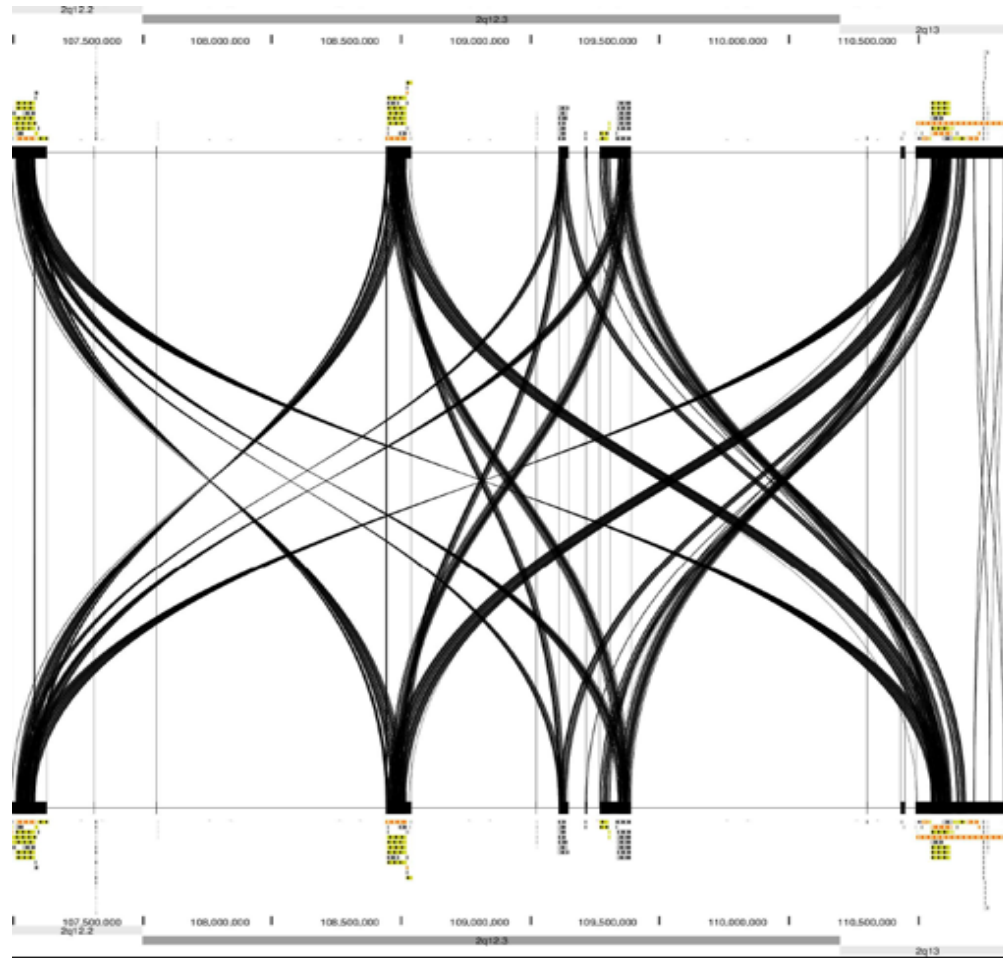
# klastrowanie regionów LCR (low copy repeat)



NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits, Dittwald et al.



# architektura genomu



**NAHR-mediated copy-number variants in a clinical population: mechanistic insights into both genomic disorders and Mendelizing traits, Dittwald et al.**

# korelacje z częstością zmian

- identyfikacja cech architektury genomu, które korelują z częstością występowania zmian strukturalnych:
  - zawartość GC
  - nasycenie specyficznymi motywami
  - występowanie sekwencji homologicznych
- próba budowy modelu statystycznego wyjaśniającego obserwowane korelacje

# Sekwencje niekodujące w regulacji genów

Bartek Wilczyński



# Motywy sekwencyjne

- Rola sekwencji regulatorowych w regulacji ekspresji
- Motywy sekwencyjne jako opis miejsc wiązania czynników transkrypcyjnych
- Zawartość informacyjna jako miara jakości motywu
- wyszukiwanie motywów optymalnych w sensie zawartości informacyjnej
- motywy a ewolucja

# Modyfikacje chromatyny

- Rola chromatyny w aktywności sekwencji regulatorowych
- Modyfikacje histonów jako "odczyt" stanu chromatyny
- Ukryte Modele Markowa jako narzędzie do opisu różnych funkcji fragmentów genomu
- Znajdowanie optymalnych Modeli Markowa opisujących zaobserwowane modyfikacje histonów

# Klasyfikacja obszarów regulatorowych

- Problem klasyfikacji ogólnie i w kontekście przewidywania aktywności sekwencji regulatorowych
- Metody rozwiązywania problemu klasyfikacji (na przykładzie lasów losowych i sieci Bayesowskich)
- Klasyfikacja obszarów regulatorowych na podstawie motywów i modyfikacji histonów

# Modele procesów nowotworowych

Urszula Foryś

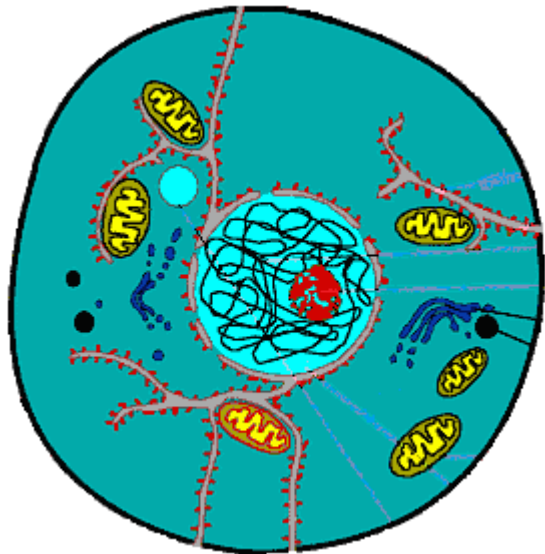
Instytut Matematyki Stosowanej i Mechaniki  
Uniwersytet Warszawski

# Cele wykładu

- poznanie podstawowych procesów towarzyszących progresji nowotworu
- poznanie podstawowych modeli opisujących wzrost guza litego
- modelowanie bardziej założeń zależności, takich jak angiogeneza
- metody modelowania terapii przeciwnowotworowych, takich jak terapia antyangiogenna

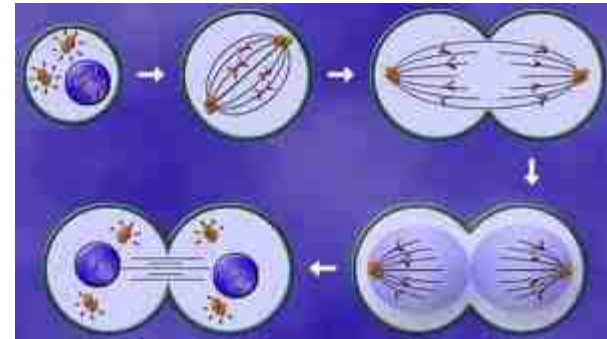
# Jakościowo, nie ilościowo

Chcemy opisać dynamikę guza łitego w początkowej fazie wzrostu.



schemat na podstawie: taenia-saginata.blog.onet.pl

Zauważmy, że już pojedyncza komórka jest obiektem skomplikowanym.

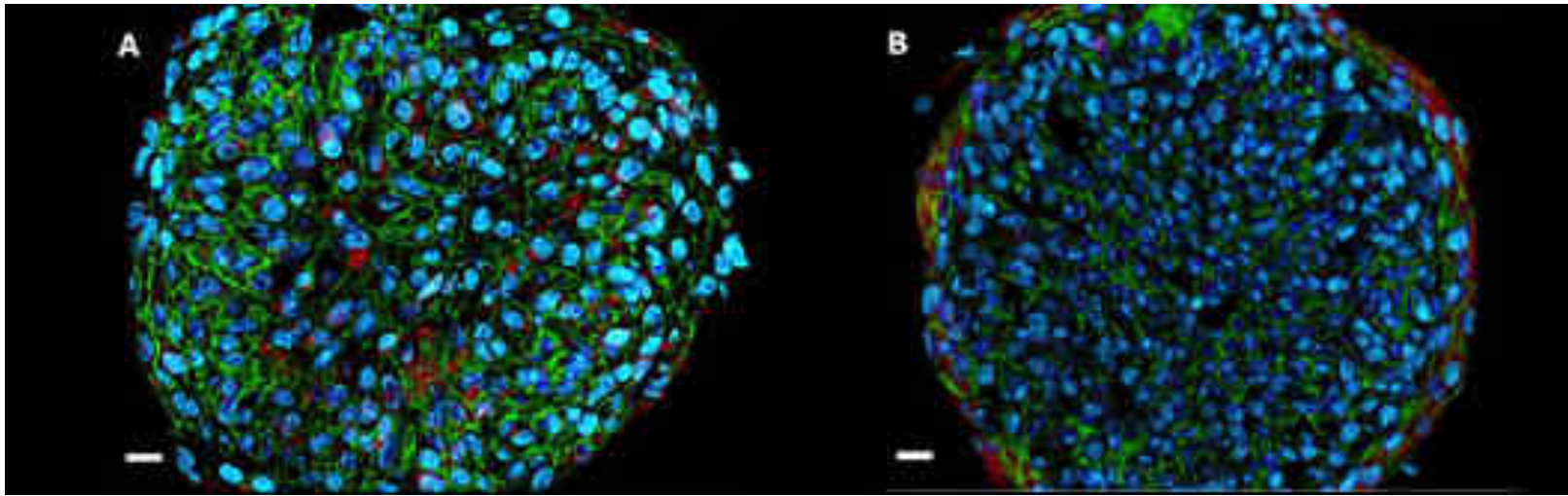


schemat na podstawie: nauka.katalogi.pl

Także przebieg cyklu komórkowego i podział komórki zależy od wielu różnych czynników.

# *Multicellular spheroid (MCS)*

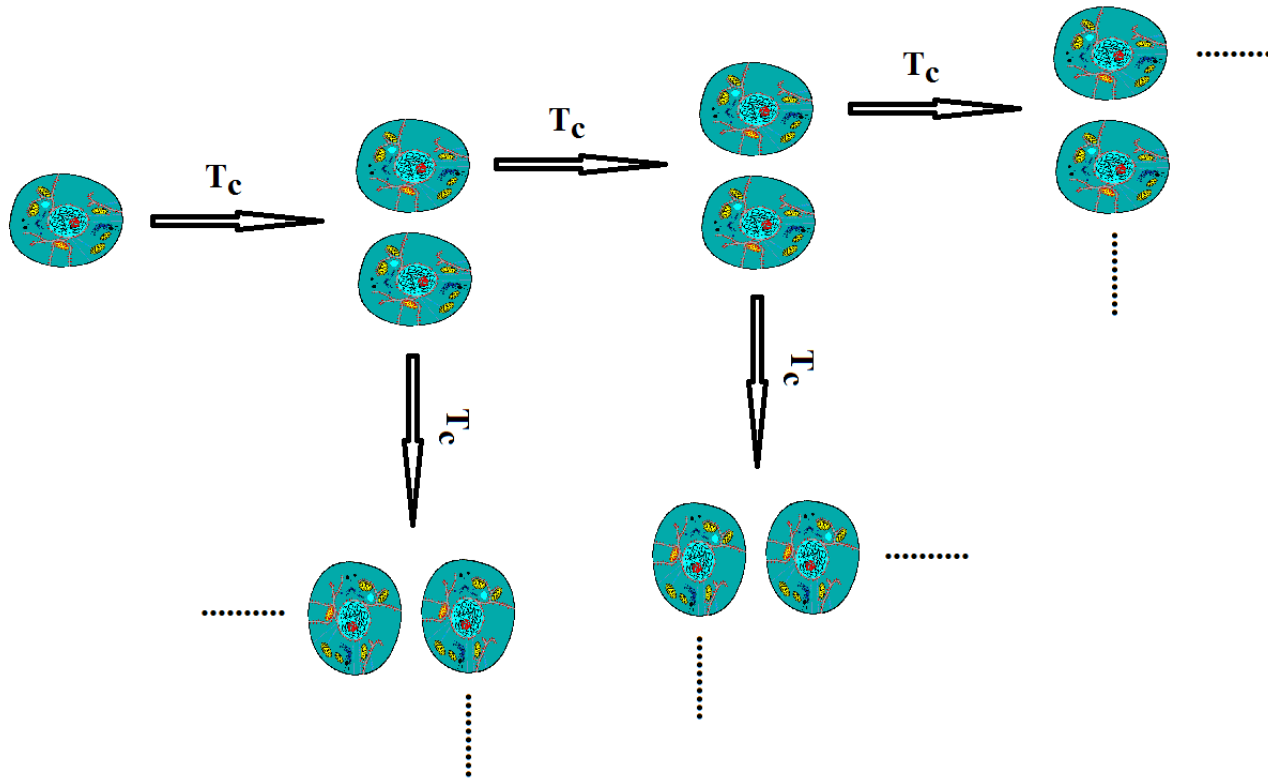
W początkowej fazie wzrostu guz lity tworzy dość jednorodny agregat komórkowy, który nazywamy **wielokomórkową sferoidą/multicellular spheroid (MCS)**.



Zdjęcie z eksperymentu na linii komórkowej MCF-10A komórek nabaonka piersi; Uniwersytet w Lund, [lu.se/cell-proliferation-group](http://lu.se/cell-proliferation-group)

Zakładamy więc, że komórki nowotworowe układają się równomiernie tworząc symetryczną strukturę, dlatego dynamikę procesu możemy opisać za pomocą **liczebności  $x(t)$  w chwili  $t$** .

# Najprostszy opis (podział synchroniczny)



Komórki dzielą się synchronicznie co ustalony czas  $T_c$ , czyli

$$x(T_c) = 2x(0) = 2x_0, x(2T_c) = 2x(T_c) = 2^2x_0, \dots, x(nT_c) = 2x((n-1)T_c) = 2^n x_0.$$

**Wniosek 1.** Otrzymujemy ciąg geometryczny o ilorazie  $q = 2$  i początkowym wyrazie  $x_0$ .



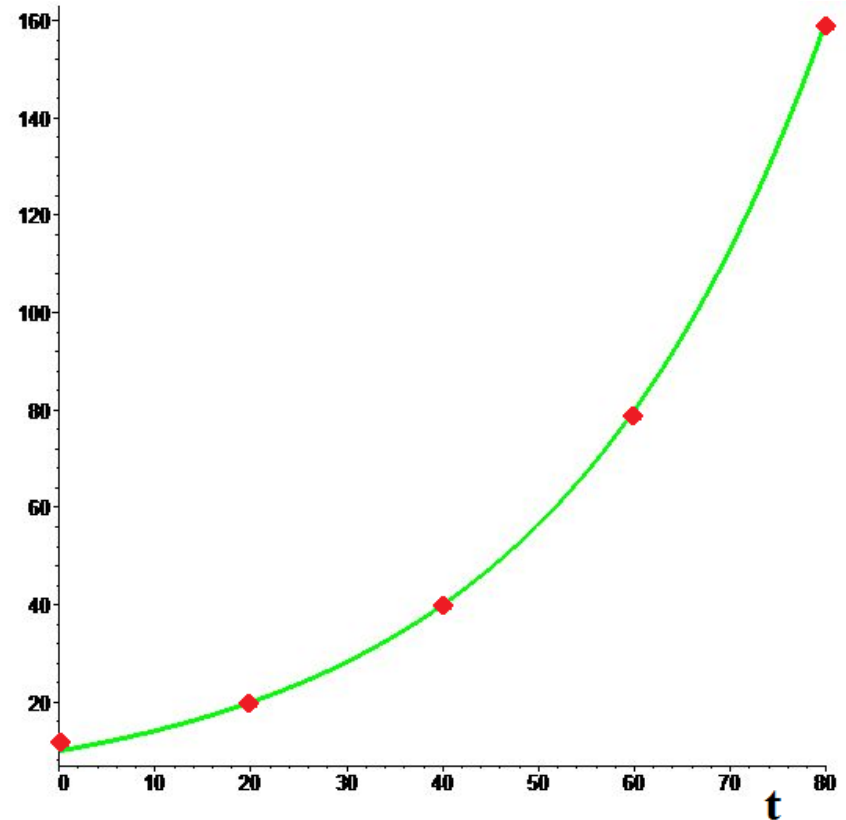
# Najprostszy opis (podział asynchroniczny)

W rzeczywistości mamy do czynienia z podziałami asynchronicznymi, otrzymujemy więc prostą zależność wykładniczą

$$x(t) = 2^{t/T_c} x_0,$$

w szczególności dla  $t = nT_c$  mamy

$$x(nT_c) = 2^n x_0.$$



# Bardziej realistyczne modele (model Gompertza)

- Zauważmy jednak, że w tak opisanym zagadnieniu czas podwojenia sferoidy  $T_c$  jest stały, nie zależy ani od czasu, ani od wielkości populacji komórkowej, czy innych charakterystyk tego procesu.
- Okazuje się, że w rzeczywistości **czas podwojenia wzrasta wraz z rosnącą wielkością populacji komórkowej** (czyli objętością sferoidy).
- Najczęściej stosowanym nieliniowym opisem wzrostu guza, który uwzględnia to zjawisko, jest **model Gompertza**:

$$x(t) = x_0 e^{\left(\frac{\alpha}{\beta} (1 - e^{-\beta t})\right)}$$

- W latach sześćdziesiątych XX w. Anna Laird, na podstawie porównań z danymi doszła do wniosku, że model ten bardzo dobrze przybliży wzrost nowotworu w początkowej fazie.

# Bardziej realistyczne modele (model Gompertza)

$$x(t) = x_0 e^{\left(\frac{\alpha}{\beta} (1 - e^{-\beta t})\right)}$$

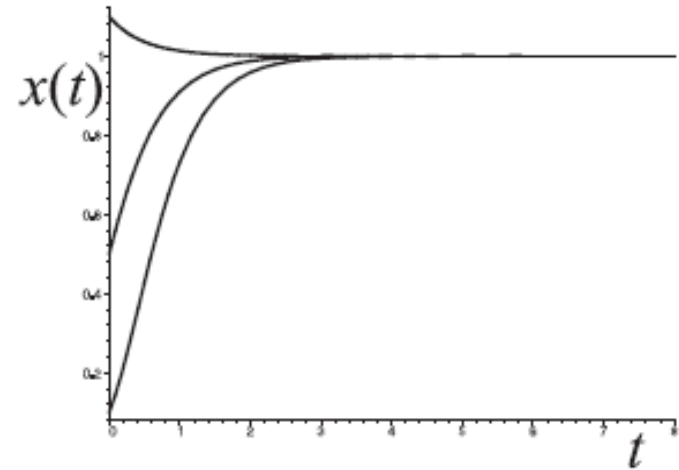
- $x(t)$  i  $x_0$  mają tę samą interpretację jak poprzednio (wielkość sferoidy)
- $\alpha$  odzwierciedla maksymalny współczynnik wzrostu nowotworu,
- $\beta$  opisuje stopień odchylenia krzywej Gompertza od krzywej wykładniczej,
- funkcja wykładnicza  $f(z) = e^z$  jest funkcją o podstawie  $e$ , gdzie  $e \approx 2,7$  jest liczbą Eulera,
- Parametry  $x_0$ ,  $\alpha$  oraz  $\beta$  spełniają zależność

$$\lim_{t \rightarrow \infty} x(t) = x_0 e^{\frac{\alpha}{\beta}}.$$

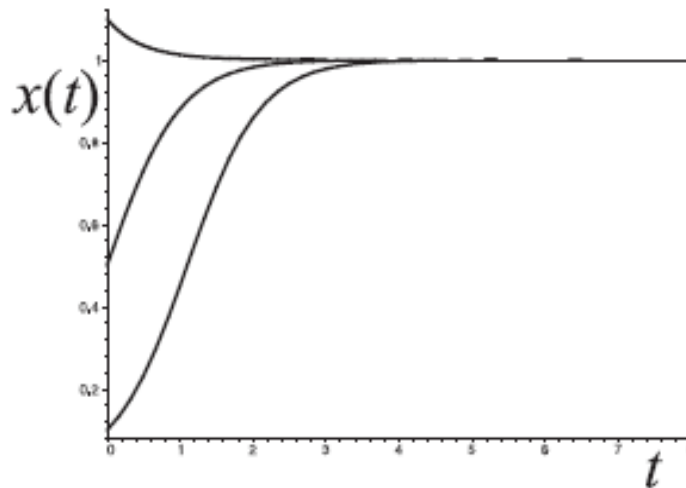
# Inne bardziej realistyczne modele

- Podobne jakościowe odwzorowanie procesu wzrostu guza litego otrzymujemy korzystając z innych modeli.

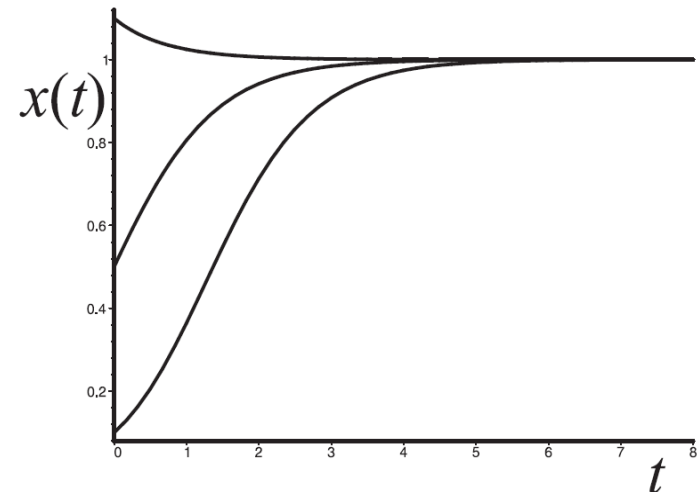
**Gompertz**



**Logistyczne**



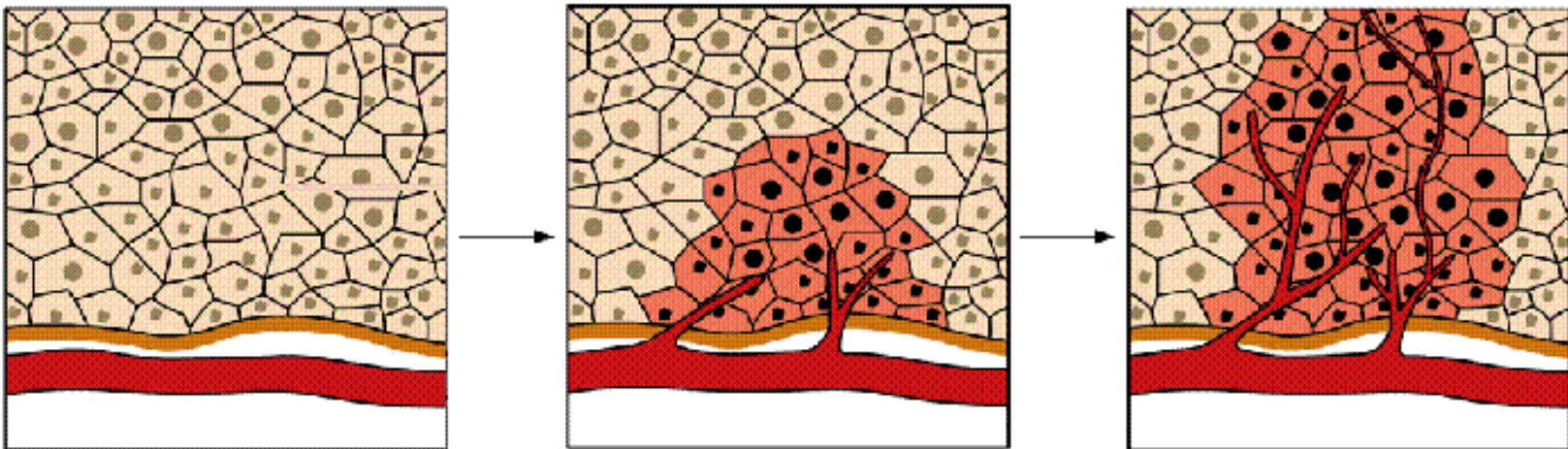
**Greenspan**



# Angiogeneza nowotworowa

- Jeśli guz przekracza pewną krytyczną wielkość (średnica ok. 1-2 mm), to składniki pokarmowe przestają docierać w odpowiedniej ilości do wszystkich komórek we wnętrzu guza.
- W celu dalszego rozwoju nowotwór zaczyna wydzielać odpowiednie substancje, które stymulują do wzrostu komórki śródbłonna naczyniowego i z istniejących naczyń krwionośnych powstają nowe, które penetrują wewnątrz nowotworu i dzięki temu nowotwór może dalej rosnąć.

Proces ten nazywamy **angiogenezą nowotworową**.



# Angiogeneza nowotworowa - modelowanie

## Model Hahnfeldta i in.

- $V(t)$  - objętość guza ( $\text{mm}^3$ )
- $K(t)$  - maksymalna objętość guza jaką bieżąca struktura naczyń jest w stanie odżywić

Model opisujący wzrost guza unaczynionego wyrażony jest za pomocą układu dwóch równań różniczkowych zwyczajnych:

$$\begin{cases} \dot{V} = -\lambda_1 V \ln \frac{V}{K}, \\ \dot{K} = -\lambda_2 K + bV - dKV^{2/3} - eKg. \end{cases}$$

# Angiogeneza nowotworowa - modelowanie

## Model Hahnfeldta i in.

model Gompertza

$$\begin{cases} \dot{V} = -\lambda_1 V \ln \frac{V}{K}, \\ \dot{K} = -\lambda_2 K + bV - dKV^{2/3} - eKg. \end{cases}$$

spontaniczna regresja naczyń  
krwionośnych

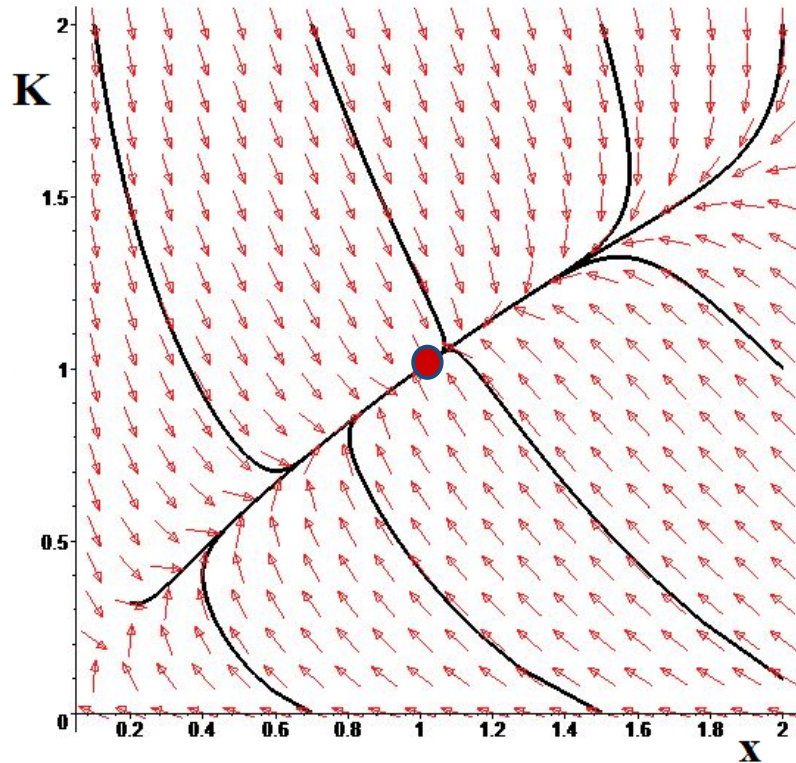
stymulacja wzrostu naczyń  
krwionośnych guza pod wpływem  
substancji wydzielanych przez  
komórki nowotworowe

inhibicja wzrostu naczyń poprzez  
różne czynniki wydzielane w  
obrębie guza

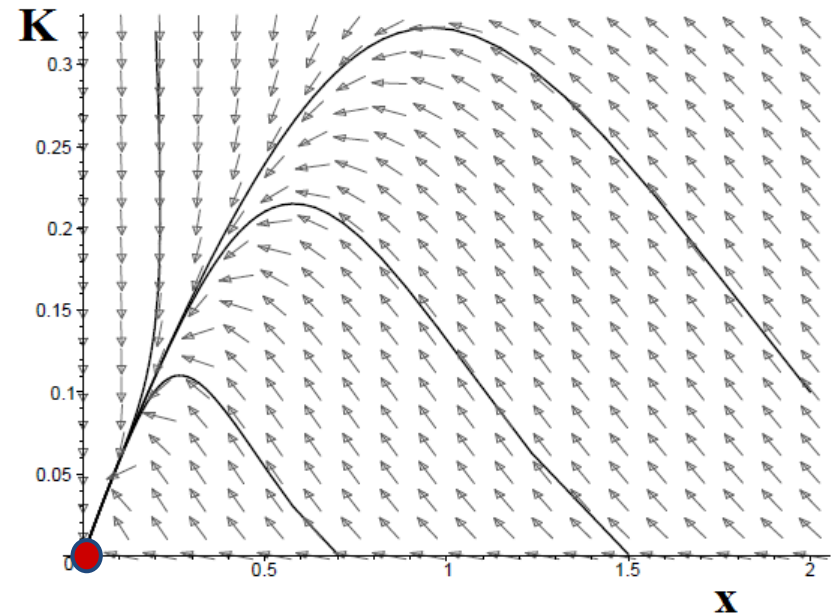
terapia antyangiogenna

# Angiogeneza nowotworowa - modelowanie

## Model Hahnfeldta i in.



Teoretyczna graniczna wielkość guza.



Możliwość wyleczenia przy odpowiednio wysokiej dawce.



**„Nowoczesne metody, leki i terapie w ochronie zdrowia i gospodarce Europy XXI wieku – interdyscyplinarne kształcenie w obszarze nauk biomedycznych na studiach II i III stopnia”, POKL.04.03.00-00-060/12**

Zadanie nr 1: Przygotowanie, otwarcie i realizacja trzech nowych specjalności na studiach II stopnia na kierunku Biotechnologia

Materiały dydaktyczne do ćwiczeń w ramach przedmiotu:

**Matematyczne metody w naukach biomedycznych**  
dla Specjalności:

Łączna liczba godzin dydaktycznych: 60

**Autorzy: Paulina Szymańska, Piotr Dittwald, Jan Poleszczuk**

Projekt realizowany ze środków Unii Europejskiej  
w ramach Europejskiego Funduszu Społecznego

# Matematyczny model ekspresji genów

## ĆWICZENIA

Paulina Szymańska MISDoMP  
Uniwersytet Warszawski

# Blok deterministyczny

1. Obliczanie pochodnych i całek prostych funkcji
2. Rozwiązywanie prostych równań różniczkowych zwyczajnych
3. Wykreślanie rozwiązań prostych równań, wizualizacja trajektorii czasowych
4. Badanie stabilności punktów krytycznych równań różniczkowych zwyczajnych
5. Analiza jakościowa wybranych układów równań różniczkowych zwyczajnych

# Blok stochastyczny

1. Znajdywanie wartości średniej i wariancji prostych rozkładów prawdopodobieństwa
2. Obliczanie prawdopodobieństw przy pomocy prawdopodobieństwa warunkowego i wzoru Bayesa
3. Konstrukcja prostych modeli stochastycznych przykładowych reakcji chemicznych – przejście od równań różniczkowych do odpowiadających im procesów urodzin i śmierci
4. Przykłady łańcuchów Markowa, znajdywanie stanów stacjonarnych

# **Metody detekcji i analizy patogennych zmian strukturalnych w genomie człowieka ćwiczenia**

Piotr Dittwald

MISDoMP, UW

# Ćwiczenia: mgr Piotr Dittwald

- blok biologiczny
- blok algorytmiczny
- blok statystyczny

UWAGA: powyższy podział ma charakter orientacyjny, a poruszana tematyka może na raz dotyczyć kilku bloków, np. zasoby biologiczne mogą być wykorzystane jako podstawa działania wybranych algorytmów i analizowane przy użyciu technik statystycznych

# BLOK BIOLOGICZNY

- podstawowe zasoby molekularne:
  - przeglądarka UCSC (genom oraz zbiory danych z nim związane),
  - bazy danych o zmianach strukturalnych: DGV, ISCA
- integracja zasobów ze środowiskiem programistycznym

# BLOK ALGORYTMICZNY

- programowanie w języku R, podstawowa analiza statystyczna przy użyciu R, wybrane biblioteki, funkcje, operacje na wektorach, obsługa plików, wizualizacja otrzymanych wyników
- operacje na przedziałach i ich wykorzystanie w analizie ludzkiego genomu
- uliniowanie globalne
- narzędzia do analizy genomu człowieka: BLAST, BLAT, CLUSTAL



# BLOK STATYSTYCZNY

- korelacja Pearsona/Spearmana,
- testy parametryczne i nieparametryczne, p-wartość
- wprowadzenie do regresji liniowej,
- porównywanie modeli (kryteria BIC, AIC)
- ANOVA

# Sekwencje niekodujące w regulacji genów - ćwiczenia

Piotr Dittwald

MISDoMP UW

# Motywy sekwencyjne

- Formaty zapisu motywów sekwencyjnych
- Obliczanie zawartości informacyjnej motywu
- Poprawka Laplace'a
- Wyszukiwanie instancji motywów w genomach i ustalanie progów istotności dla takich wystąpień
- Zastosowanie dostępnych metod optymalizacji motywów dla danych o wiązaniu czynników

# Modyfikacje chromatyny

- Dostępne publicznie dane o modyfikacjach histonów w projektach ENCODE i modENCODE
- Metody automatycznej segmentacji genomu przy pomocy
  - Ukrytych modeli Markowa (Chrom HMM)
  - Dynamicznych sieci Bayesowskich (Segway)
- Porównanie predykcji pochodzących z segmentacji genomu z faktycznymi miejscami regulatorowymi (np. baza enhancer.lbl.gov)

# Klasyfikacja obszarów regulatorowych

- Dostępne pakiety do klasyfikacji przy pomocy lasów losowych i sieci Bayesowskich
- Opis sekwencji genomowych w przestrzeni motywów i przestrzeni modyfikacji histonów
- konstrukcja zbioru uczącego
- Metody weryfikacji krzyżowej
- Analiza jakości predykcji przy pomocy krzywych ROC

# Modele procesów nowotworowych (ĆWICZENIA)

Jan Poleszczuk

MISDoMP, UW

# Podstawy teoretyczne (aspekt jakościowy)

## 1. Podstawy jakościowej teorii równań dyskretnych

- znajdowanie punktów krytycznych równania dyskretnego
- badanie lokalnej stabilności punktów krytycznych
- badanie globalnej stabilności punktów krytycznych dla pojedynczego równania dyskretnego (metoda pajęczynowa)
- analiza wybranych modeli dyskretnych

## 2. Podstawy jakościowej teorii równań różniczkowych

- znajdowanie punktów krytycznych równania różniczkowego
- badanie lokalnej stabilności punktów krytycznych
- badanie globalnej stabilności punktów krytycznych dla pojedynczego równania różniczkowego (metoda portretu fazowego)
- analiza wybranych modeli z czasem ciągłym

# Rozwiązania numeryczne (aspekt ilościowy)

- zapoznanie się z dostępnymi pakietami umożliwiającymi rozwiązywanie równań dyskretnych oraz równań różniczkowych zwyczajnych (np. MATLAB, Octave)
- implementacja wybranych modeli progresji nowotworu
- eksperymenty "numeryczne" (in silico) polegające np. na badaniu zmian w przebiegu rozwiązań pod wpływem zmian w wartościach parametrów
- elementy analizy wrażliwości modelu na zmiany w wartościach parametrów i/lub warunków początkowych



# Modelowanie różnych protokołów terapii

- matematyczny model opisujący zmiany stężenia leku w krwi w zależności od wielkości dawki i częstości podawania leku (farmakokinetyka)
- implementacja wybranego modelu progresji nowotworu z uwzględnieniem funkcji opisującej terapię
- badanie wpływu terapii na nowotwór (rozwiązania modelu) przy różnych dawkach i częstościach podawania leku
- modelowanie terapii łączonych (np. terapii antyangiogennej z chemioterapią)