MIKROMACIERZE

dr inż. Aleksandra Świercz
dr Agnieszka Żmieńko

Informacje ogólne

- Wykłady będą częściowo dostępne w formie elektronicznej

http://cs.put.poznan.pl/aswiercz

aswiercz@cs.put.poznan.pl

- Godziny konsultacji:

wtorek, godz. 9-11 w sali 2.6.5 w budynku BT

- Zaliczenie przedmiotu:

obecność + test końcowy = ocena końcowa

Obecność – maksymalnie **2,5** punkta

Test końcowy – maksymalnie **2,5** punkta

Zaliczenie od 3 punktów

Wykłady

3 pierwsze wykłady będą odbywały się co tydzień, nastepnie co 2 tygodnie

Pierwsze zajęcia laboratoryjne

Zapoznanie się laboratorium, w którym przeprowadzane są eksperymenty mikromacierzowe

zbiórka przy wejściu do ECBiG, II piętro przy laboratorium **2.7.6** (zamiast 2.6.21)

Proszę o punktualność!!!!

Plan wykładów

- Wprowadzenie do eksperymentów mikromacierzowych
- Rodzaje eksperymentów i typy mikromacierzy
- Odczytywanie obrazów z mikromacierzy
- Kontrola jakości i normalizacja
- Analiza statystyczna (wykrywanie genów z istotną statystycznie różnicą w ekspresji)
- Klastrowanie, klasyfikacja
- Dalsza analiza wybranych genów

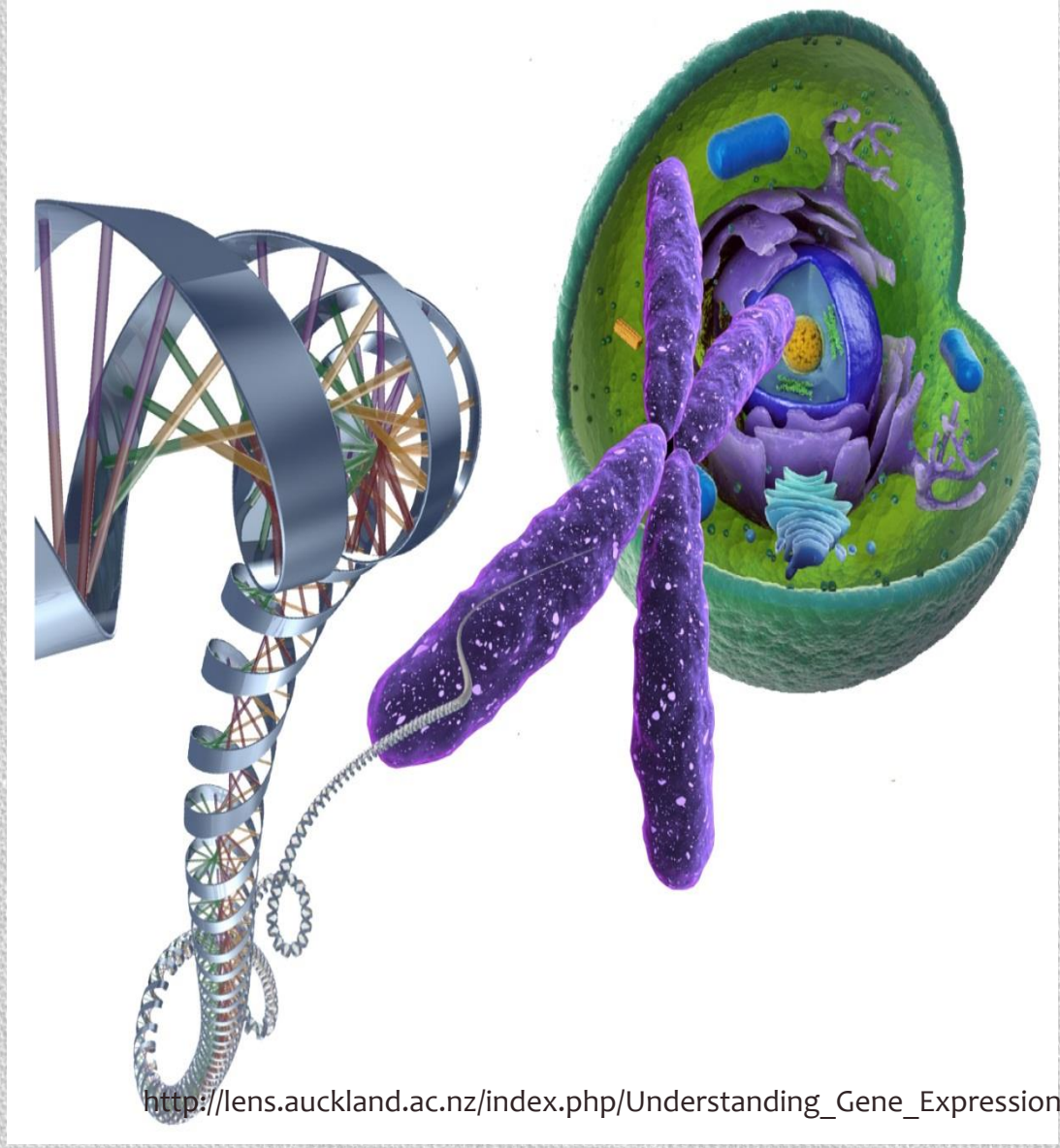
Jak działa komórka?

Kopia całego genomu przechowywana jest w każdej komórce.

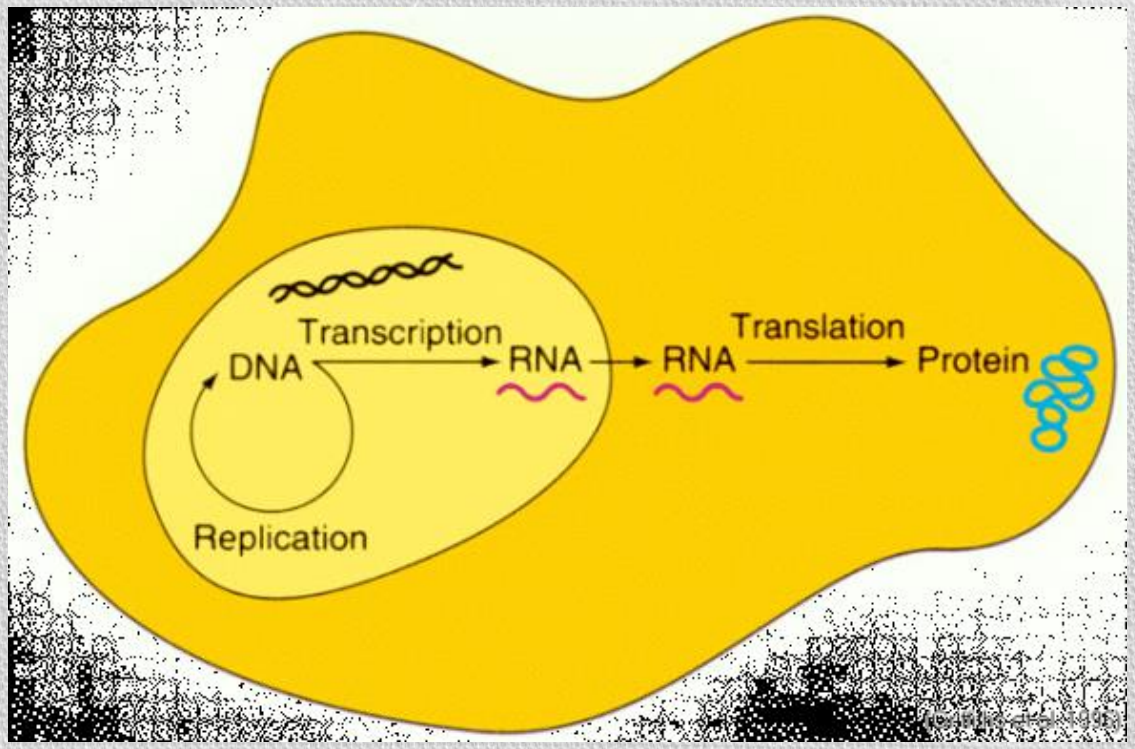
Ekspresja genu jest to proces, w którym informacja zakodowana w genie jest przepisywana na RNA.

Ekspresja genu zachodzi w komórce tylko wówczas, gdy zachodzi taka potrzeba.

W różnych komórkach inne geny ulegają ekspresji.



http://lens.auckland.ac.nz/index.php/Understanding_Gene_Expression



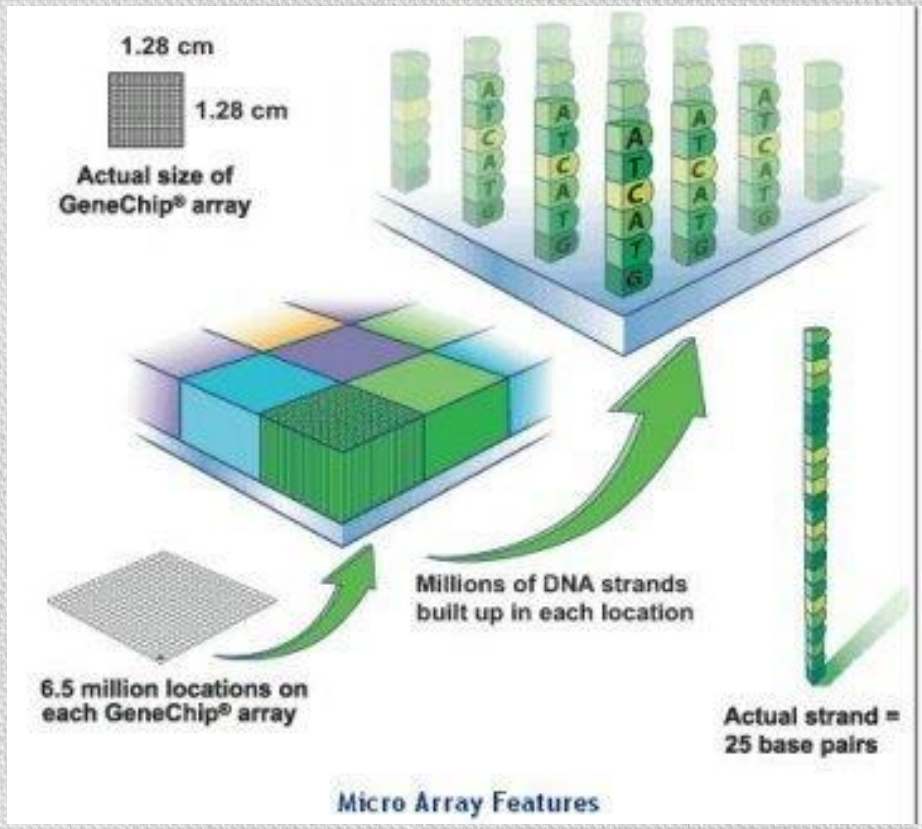
Eksperymenty, które badają ekspresję genu, mierzą ilość mRNA, aby sprawdzić, które geny uległy ekspresji w komórce.

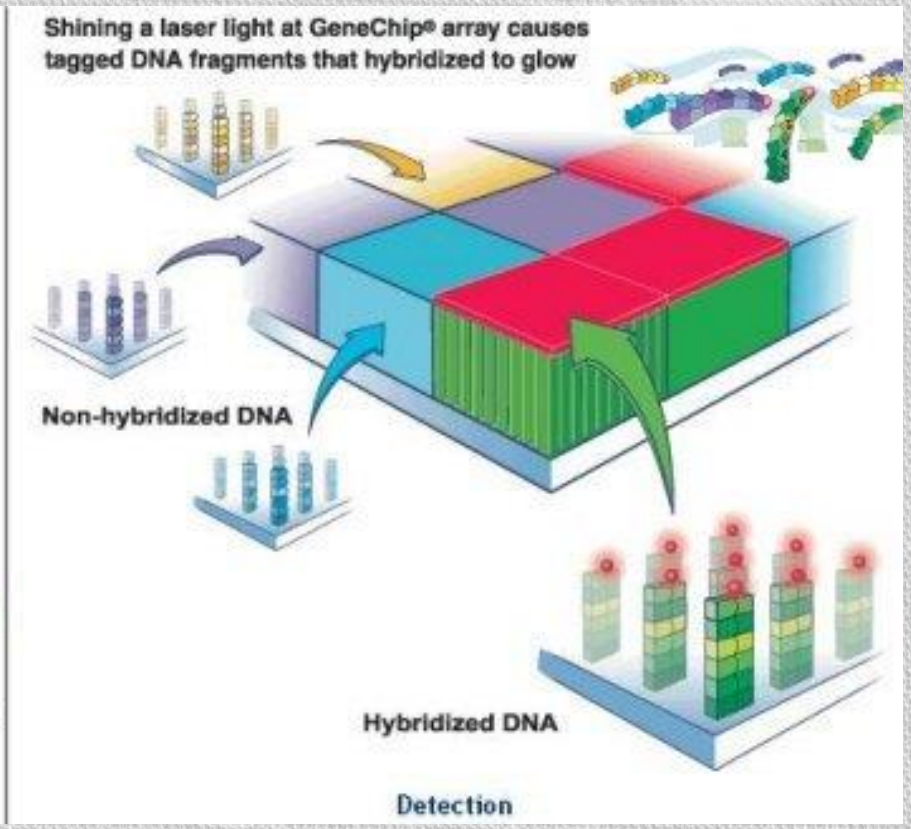
Cel eksperymentu mikromacierzowego

Celem eksperymentu jest zmierzenie obecności i liczności oznakowanych cząsteczek kwasu nukleinowego w danej próbce biologicznej, która zhybrydyzuje do mikromacierzy poprzez utworzenie dupleksów (Watson-Crick), a następnie będzie odczytana dzięki znakowaniu

„*Microarray Bioinformatics*" Dov Stekel

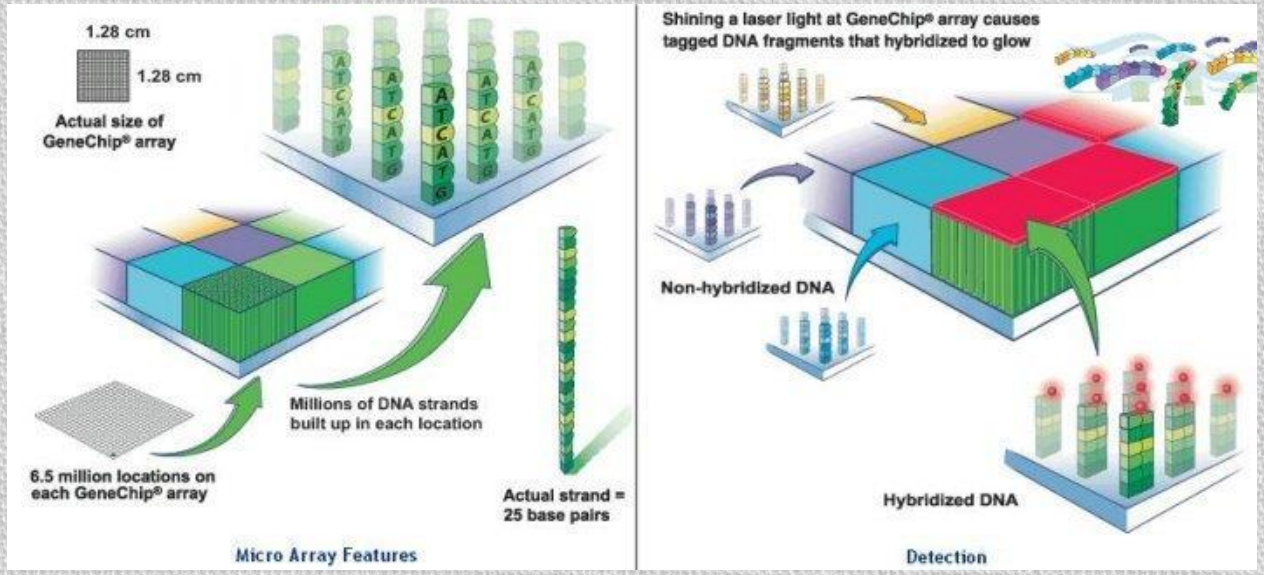
Co to jest mikromacierz?



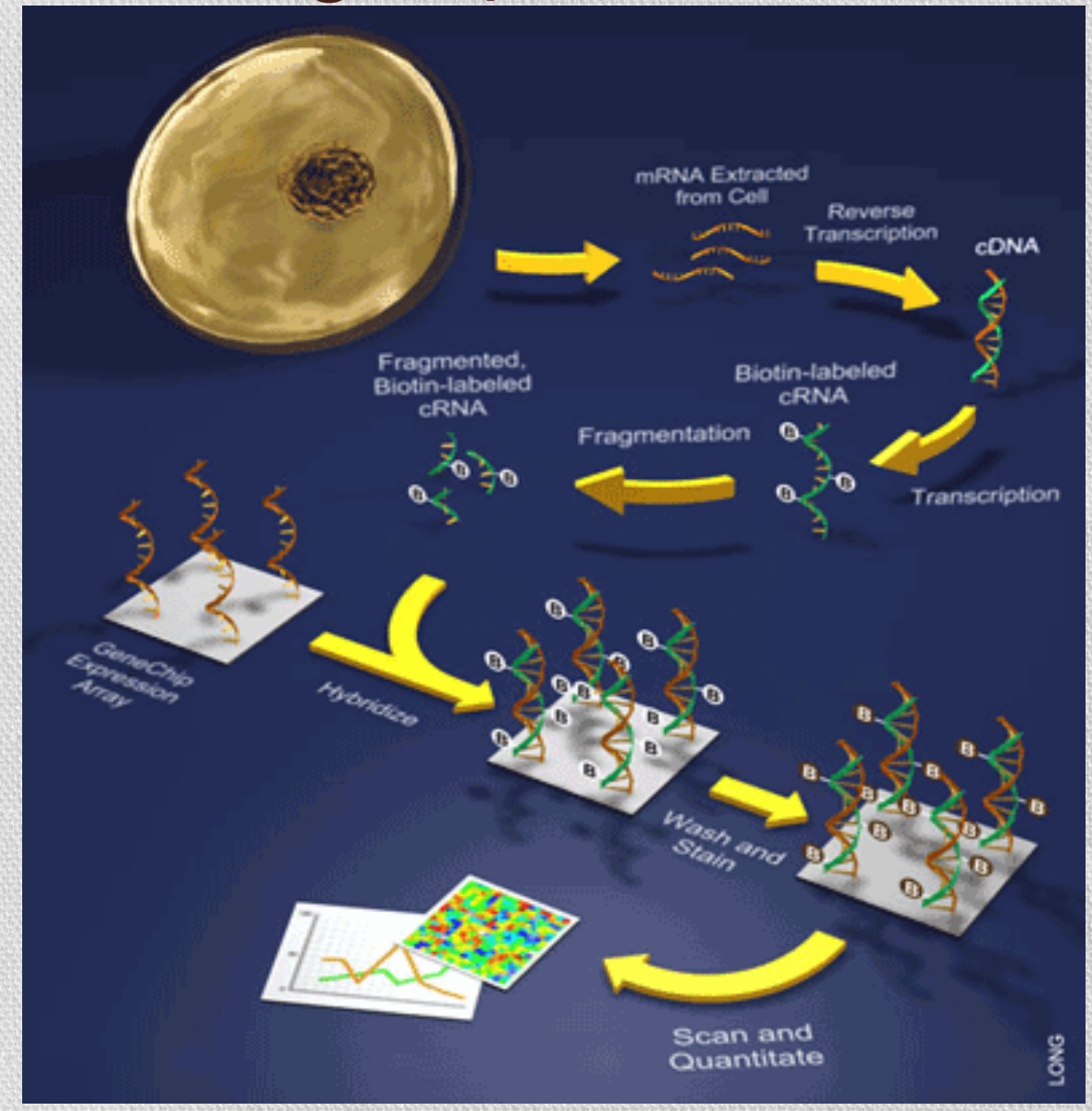


Co to jest mikromacierz?

- Mikromacierz DNA (określany także jako chip DNA) to zbiór, krótkich DNA przyczepionych do powierzchni szklanej płytki.
- Mikromacierzy można użyć do mierzenia poziomu ekspresji genów
- Każdy punkt na mikromacierzy zawiera specyficzną sekwencję DNA, która reprezentuje jeden z genów (sonda, ang. *probe*)



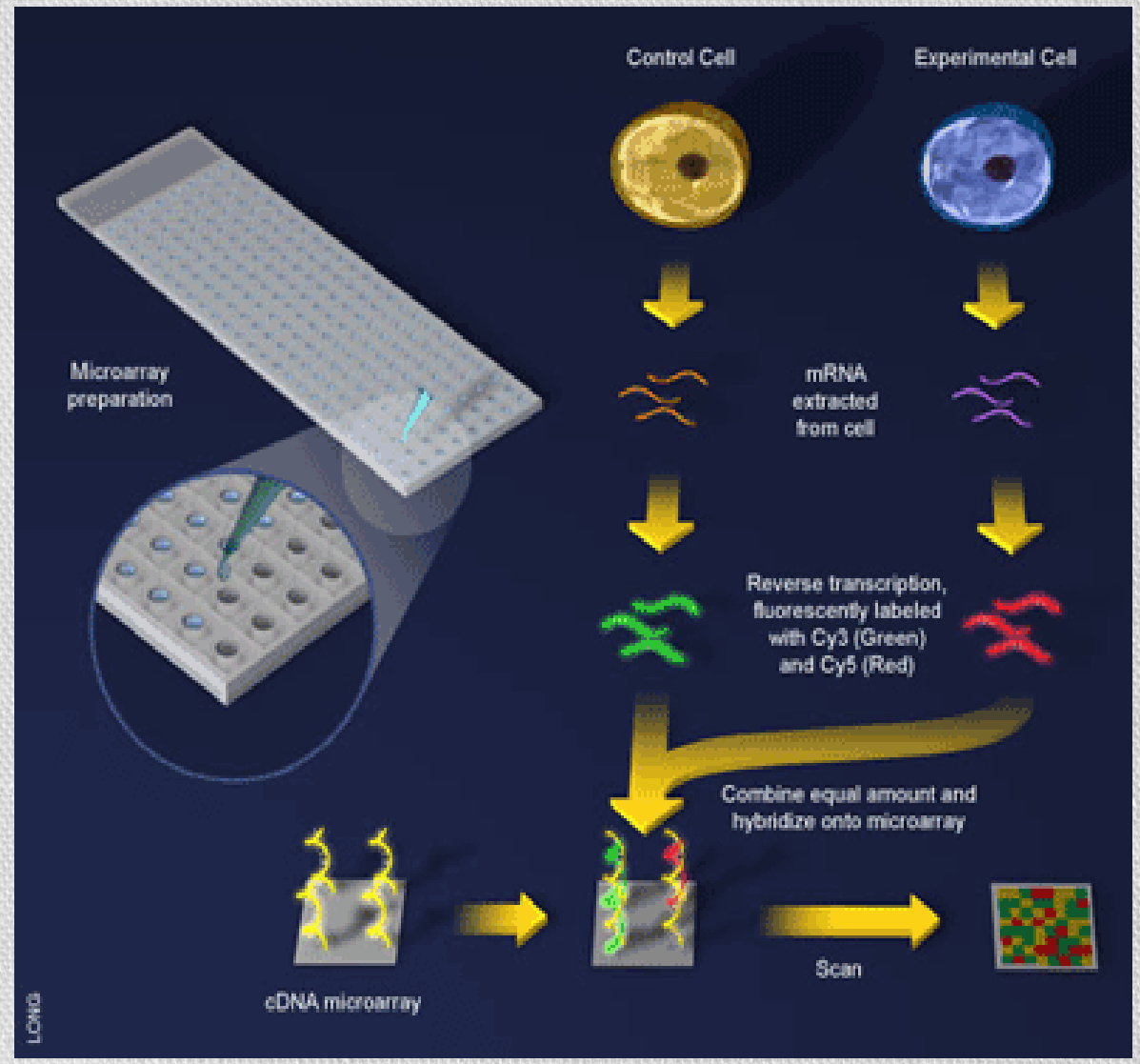
Przebieg eksperymentu



- ekstrakcja mRNA z komórki
- odwrotna transkrypcja -> cDNA
- transkrypcja + znakowanie biotyną -> RNA
- fragmentacja RNA
- hybrydyzacja RNA do oligonukleotydów na płytce
- zmywanie płytki
- skanowanie obrazu
- analiza

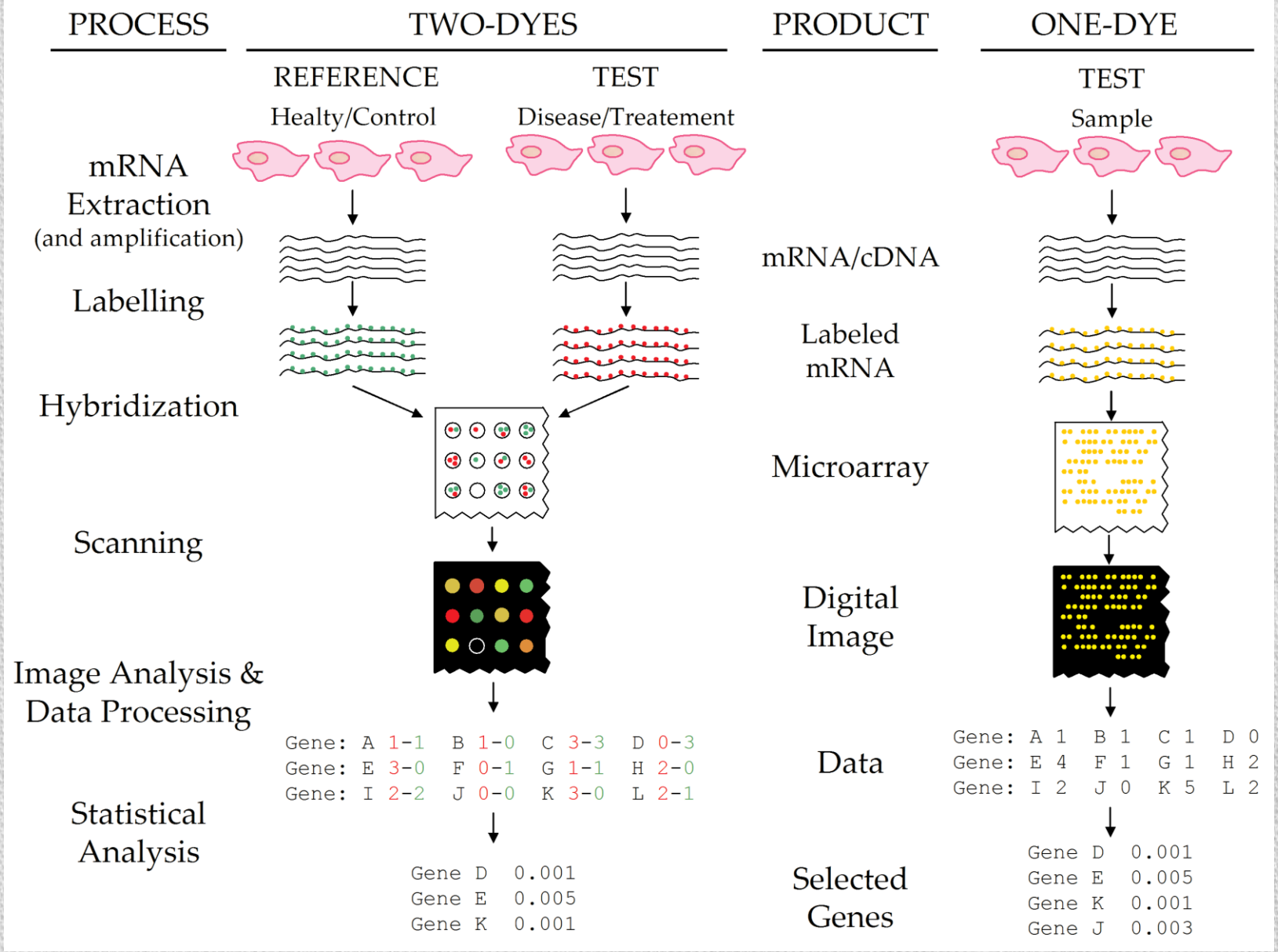
http://www.scq.ubc.ca/spot-your-genes-an-overview-of-the-microarray/

Eksperyment dwukolorowy



- Dwie oddzielne próbki (ang. *sample*)
- Każda z nich znakowana jest innym kolorem: pierwsza barwnikiem fluorescencyjnym zielonym **Cy3**, druga fluorescencyjnym czerwonym **Cy5**

Ogólny schemat przebiegu eksperymentu



Motywacja

- Porównanie ze sobą dwóch warunków (próbka kontrolna/próbka traktowana, chora/normalna) i znalezienie genów, które uległy zróżnicowanej ekspresji
- Porównanie więcej niż dwóch warunków (różne grupy chorobowe, wiele traktowań, np. szczepionki)
- Znalezienie grup, które jeszcze nie są zdefiniowane (np. nowe grupy chorobowe)
- Analiza serii czasowej (odpowiedź na szczepionkę, stadia rozwojowe, cykl komórkowy)
- Znalezienie wzorców, które powiedzą czy zastosowanie pewnej terapii przyniesie korzyści

Czy ważna jest technologia, w której przeprowadzamy eksperyment?

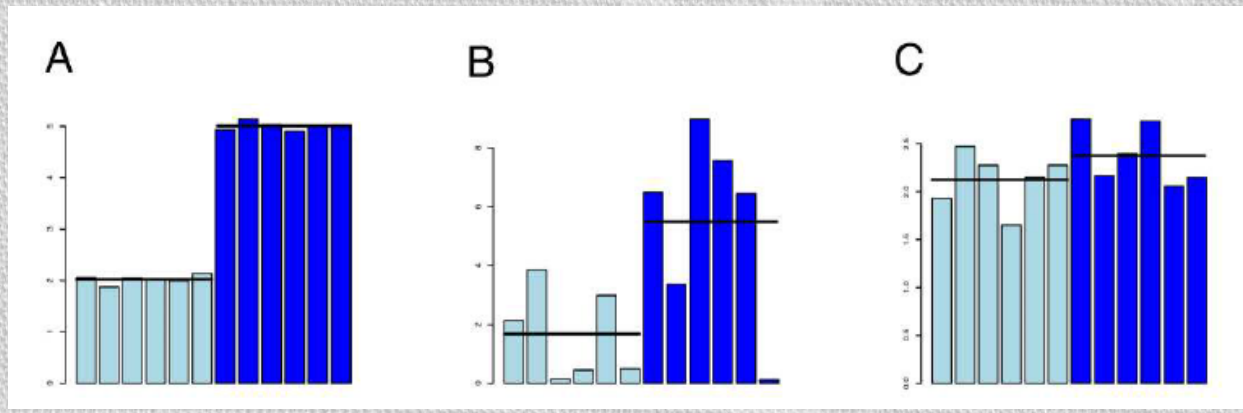
- Wybór technologii zależy od celu w jakim przeprowadzamy eksperyment i od środków jakie możemy na to przeznaczyć
- Od wybranej technologii zależy w jaki sposób eksperyment będzie przygotowywany i przeprowadzany
- Analiza niskiego poziomu (normalizacja, ocena jakości eksperymentu) jest inna dla różnych technologii

- Mikromacierze oligonukleotydowe vs cDNA (różnica w długości sond)
- Drukowane, syntetyzowane in situ (różna technika nanoszenia sond na płytkę)
- Jedno- i dwu-kanałowe (różna ilość próbek, które biorą udział w jednym eksperymencie)

Przykłady wykorzystania eksperymentów mikromacierzowych

Różna ekspresja genów

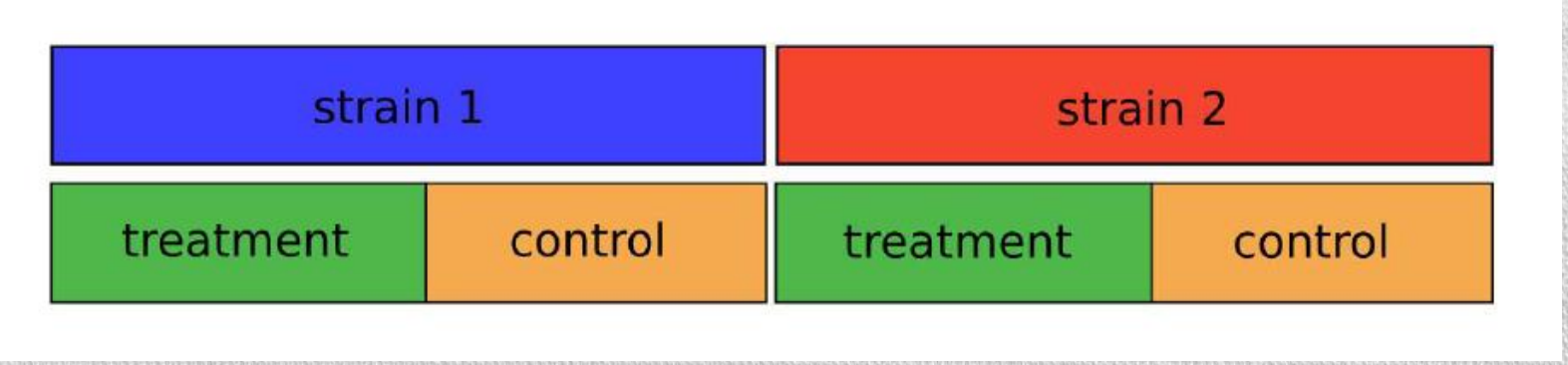
Najczęstsze pytanie w eksperymentach mikromacierzowych to, czy geny uległy mniejszej ekspresji (downregulated), czy podwyższonej (upregulated) dla różnych grup próbek (zdrowe/chore, dzikie/traktowane)



Wymagane są testy statystyczne, które będą potrafiły rozróżnić zmiany w ekspresji wynikające z innych warunków od zmian pomiędzy osobnikami tego samego gatunku (co najmniej 3 powtórzenia)

- Trudno wykryć geny, dla których różnica ekspresji jest mała, niewykrywalna przez metody statystyczne, lecz jest istotna
- Metody statystyczne to t-test, Wilcoxon

Różna ekspresja genów – kilka warunków



- Jeśli są więcej niż dwa badane warunki, lub są one zagnieżdżone, to odpowiednią metodą statystyczną jest ANOVA
- ***Wartości p-value muszą zostać skorygowane dla wielokrotnego testowania***

Analiza eksploracji danych

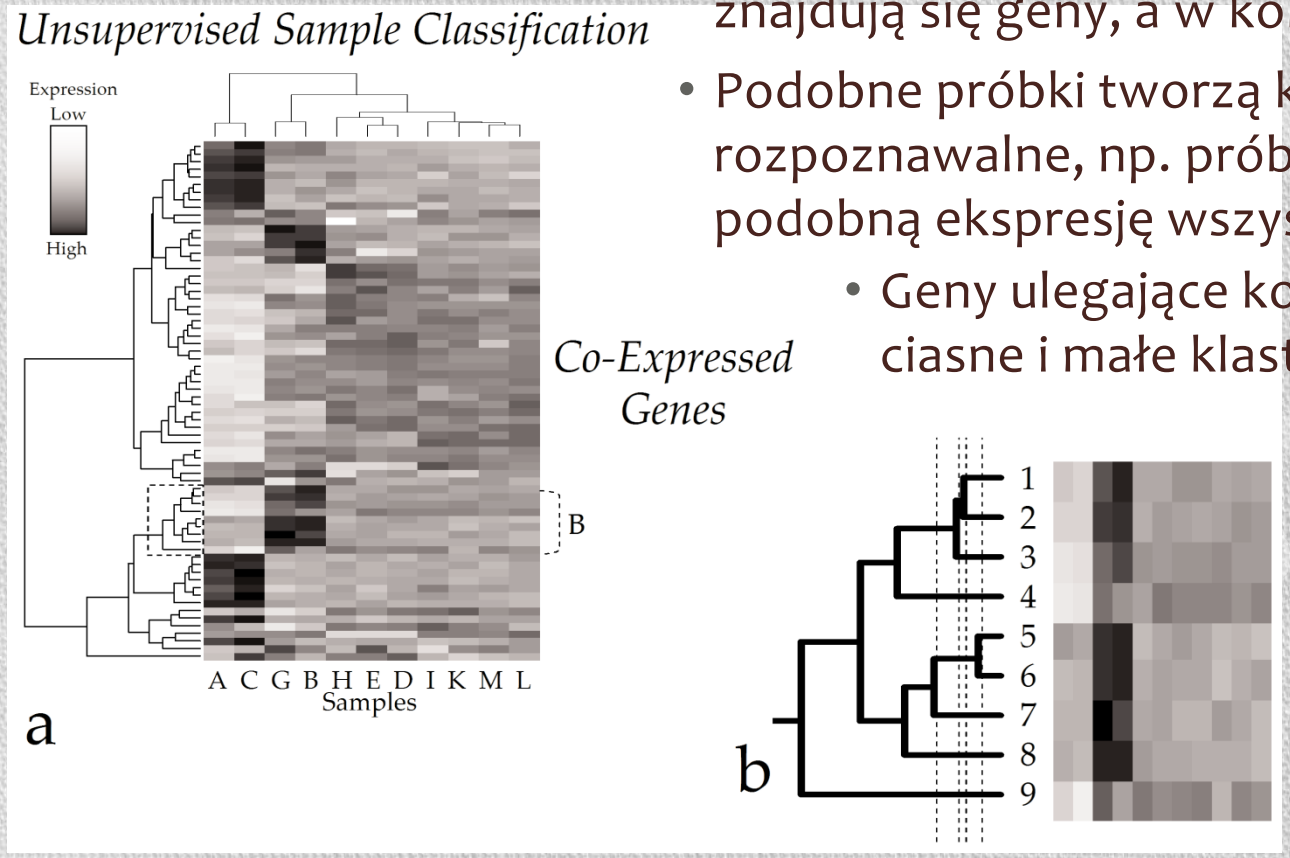
- Poszukiwanie genów o podobnym profilu ekspresji dla kilku próbek
- Metody eksploracji danych znajdą jakieś wzorce w danych, nie ważne, czy będą one znaczące dla nas czy też nie
- Metody obejmują klastrowanie (np. hierachiczne, podział, k-średnich) oraz projekcję (principal component analysis, skalowanie wielowymiarowe)
- Taki rodzaj analizy powinien być używany tylko w przypadku, gdy nie mamy żadnej wiedzy, którą moglibyśmy wykorzystać

Co możemy uzyskać dzięki metodom eksploracji danych?

Poznać funkcję nieznanych genów. Geny, które uległy podobnej ekspresji (co-expression) z dużym prawdopodobieństwem są regulowane przez te same czynniki lub też maja podobną funkcję

Analiza eksploracji danych - klastrowanie

- Podwójne hierarchczne klastrowanie wartości ekspresji genów (heatmap) – w wierszach znajdują się geny, a w kolumnach próbki.
- Podobne próbki tworzą klastry łatwo rozpoznawalne, np. próbka A i C maja bardzo podobną ekspresję wszystkich genów
  - Geny ulegające ko-ekspresji tworzą ciasne i małe klastry oznaczone na rysunku (a) jako B.
  - Na rysunku (b) hierarchiczne tworzenie klastrów dla wybranej grupy genów

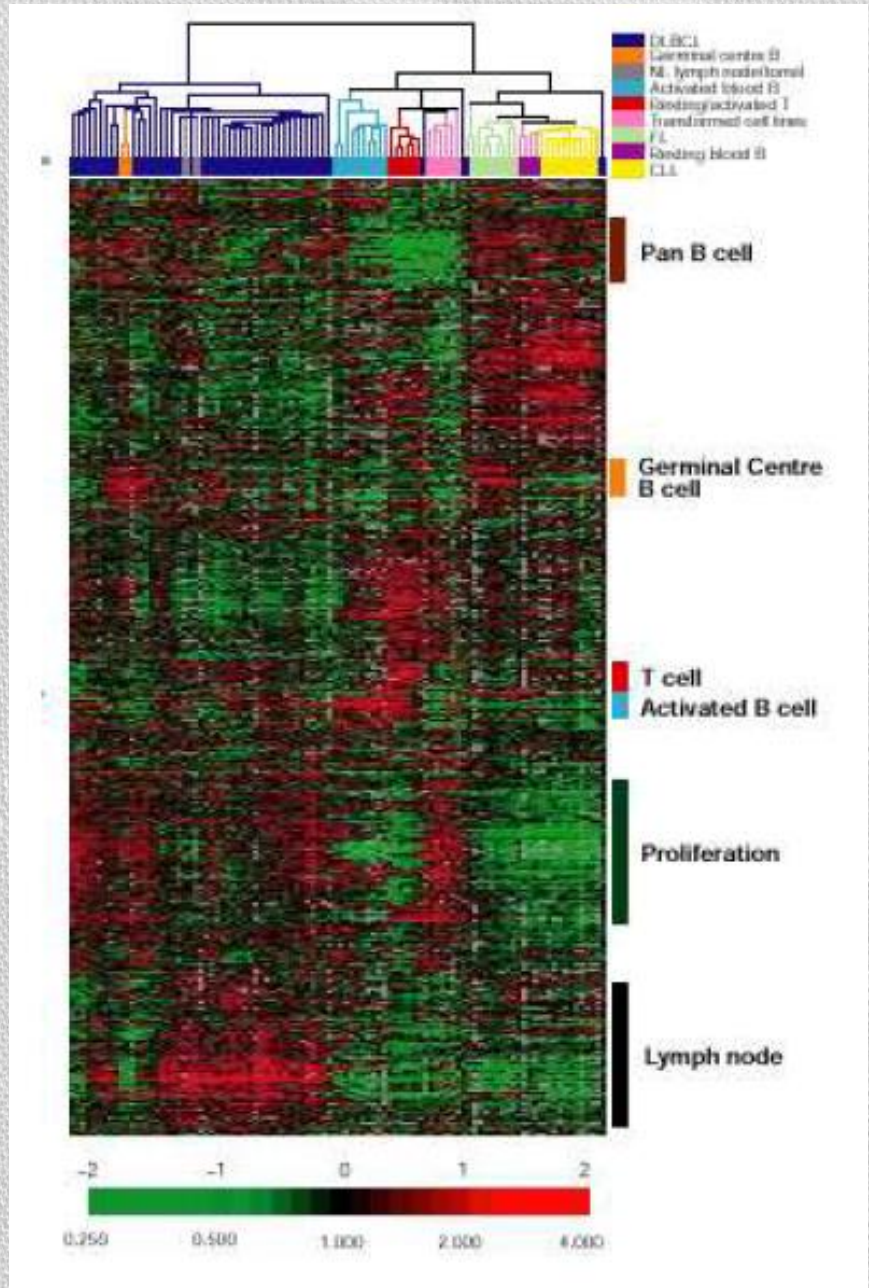


Alizadeh et al. Nature ,2000

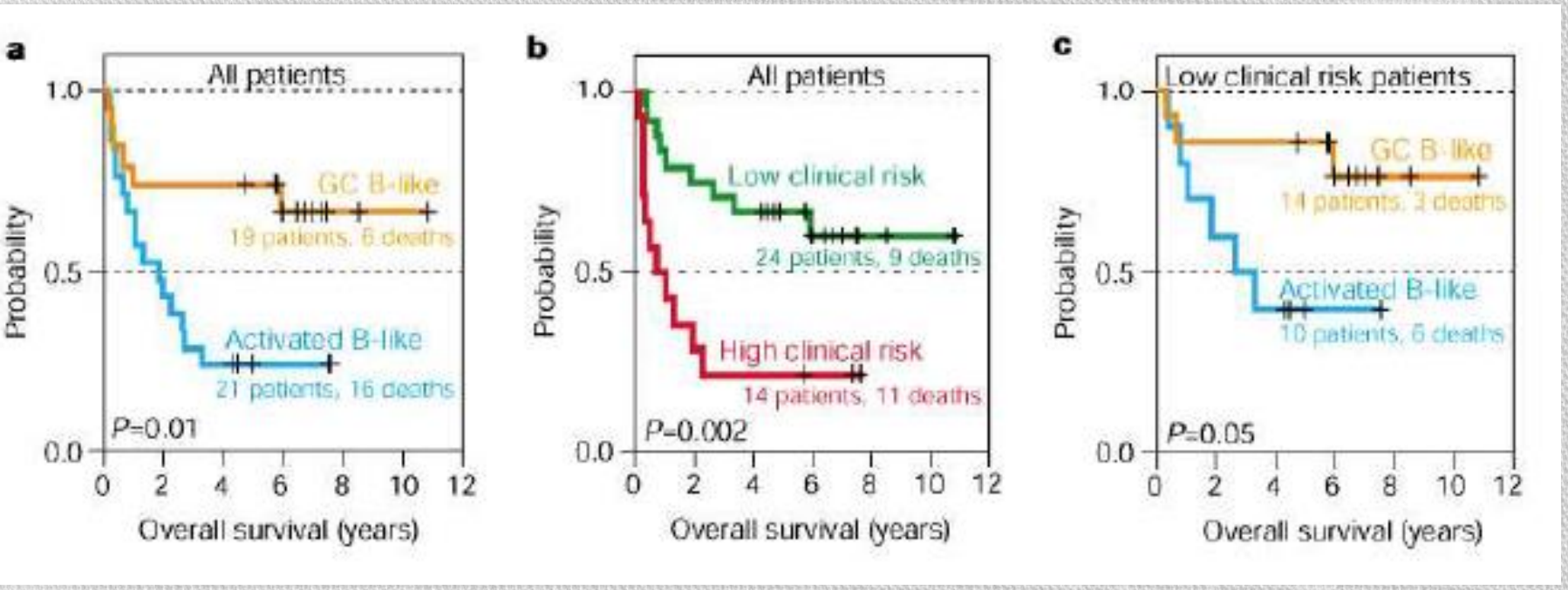
- Opublikowane w *Nature* **403**: 503-511 (2000)
- Profilowanie ekspresji genu dla DLBCL (Duffuse Large B-Cell Lymphoma)
- Lymphoma to rak krwi, w którym peryferyjne komórki krwi ulegają degeneracji i dzielą się bez kontroli
- DLBCL jest agresywną formą tej choroby, pochodzącej od B-limfocytów. 5 lat ma szansę przeżyć ok 40%

Alizadeh et al. *Nature,2000*

- Zaprojektowano specjalny chip cDNA, Lymphochip
- Drukowana macierz cDNA zawiera ok 17tys klonów związanych z białaczką
- Przeanalizowano 42 próbki DLBCL, normalne próbki z komórek B, oraz próbki z innych powiązanych chorób
- mRNA z tych próbek zostało zhybrydyzowane razem z kontrolnymi mRNA, z puli komórek białaczkowych
- Dane zostały zanalizowane za pomocą klastrowania



Alizadeh et al. *Nature*,2000



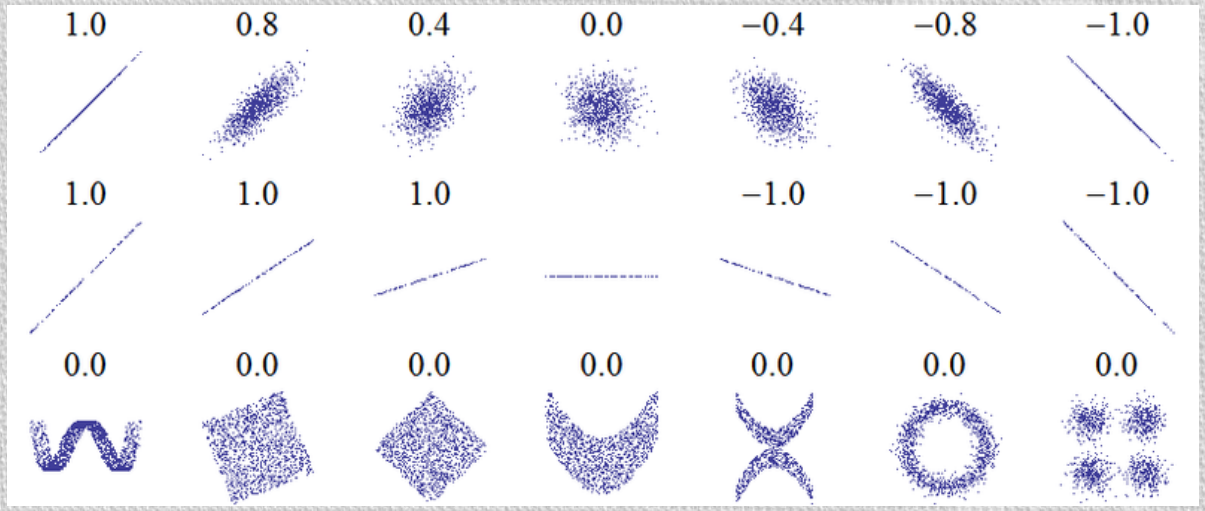
Analiza korelacji

- Jeśli poszukiwane są geny dla których profile ekspresji są podobne do genów podstawowych wystarczy policzyć korelację pomiędzy nimi i posortować po niej. Nie jest wymagane żadne klastrowanie.

- **Co to jest korelacja?**

Korelacja

- Współczynnik korelacji – liczba określająca w jakim stopniu zmienne są współzależne. Jest miarą korelacji dwu (lub więcej) zmiennych. Istnieje wiele różnych wzorów określanych jako współczynniki korelacji. Większość z nich jest normalizowana tak, żeby przybierała wartości od -1 (zupełna korelacja ujemna), przez 0 (brak korelacji) do +1 (zupełna korelacja dodatnia).



- Najczęściej stosowany jest współczynnik korelacji *r* Pearsona. W przypadku rozkładu dalekiego od dwuwymiarowego normalnego lub istnienia w próbie obserwacji odstających współczynnik korelacji Pearsona może fałszywie wskazywać na nieistniejącą korelację. Wady tej nie mają współczynniki rangowe, które z kolei mają mniejszą efektywność dla rozkładów bliskich normalnemu (np. rangi Spearmana)

Serie czasowe (time series)

- W analizie czasowej chcemy zazwyczaj znaleźć wzorce genów koekspresyjnych, czyli ze spójnym profilem ekspresji
- Znaczenie *time series* jest różne dla biologów (2-10 punktów czasowych) i dla statystyków (>200 punktów czasowych)
- Jako rozwiązanie (zazwyczaj nieoptymalne), używa się metody klastrowania aby znaleźć wspólne wzorce. Są one czasochłonne i nie ma znaczącej miary, która mogłaby być z nimi związana
- W odróżnieniu od metod eksploracji danych, bardziej popularnymi metodami są metoda k-średnich (k-means) lub też mapy samoorganizujące się

Klasyfikacja

- Klasyfikować (grupować) można różne próbki np.: traktowane/kontrolne, chore/normalne, różne stadia chorobowe, typ mutant/dziki, sukces/porażka terapii,…
- Klas może być więcej niż dwie
- *Celem klasyfikacji jest wyznaczenie zbioru genów, który będzie potrafił jednoznacznie rozróżnić, do której klasy należy badana próbka. Wybrany zbiór musi być reprezentatywny, tak aby nowego osobnika potrafił przydzielić do odpowiedniej klasy*
- W klasyfikacji można szukać charakterystycznych wzorców w **zbiorze treningowym**, a następnie oceniać przydział do klas w **zbiorze testowym**

Schemat klasyfikacji