

wykład 10

Transkryptomika

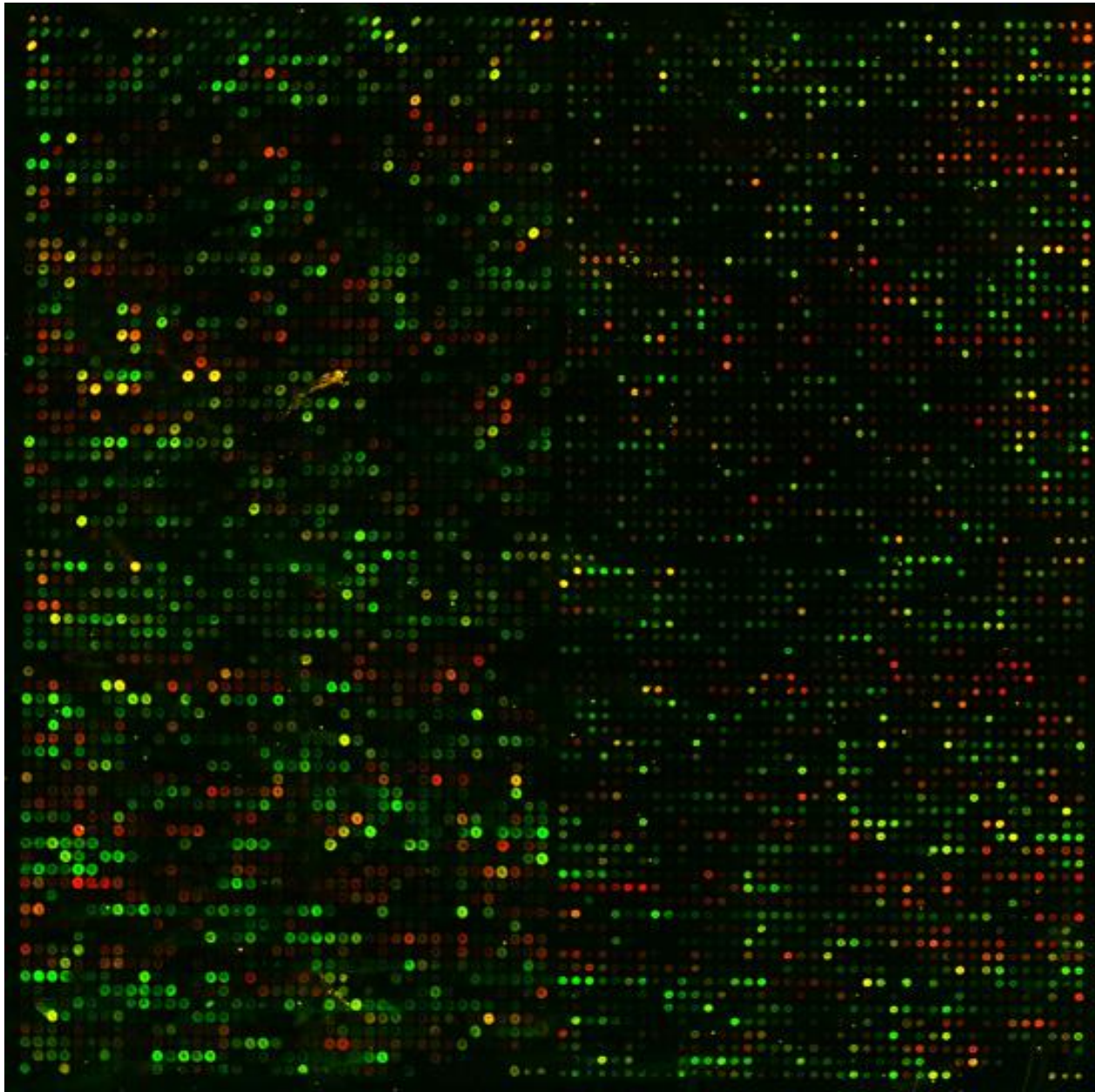
Mikromacierze

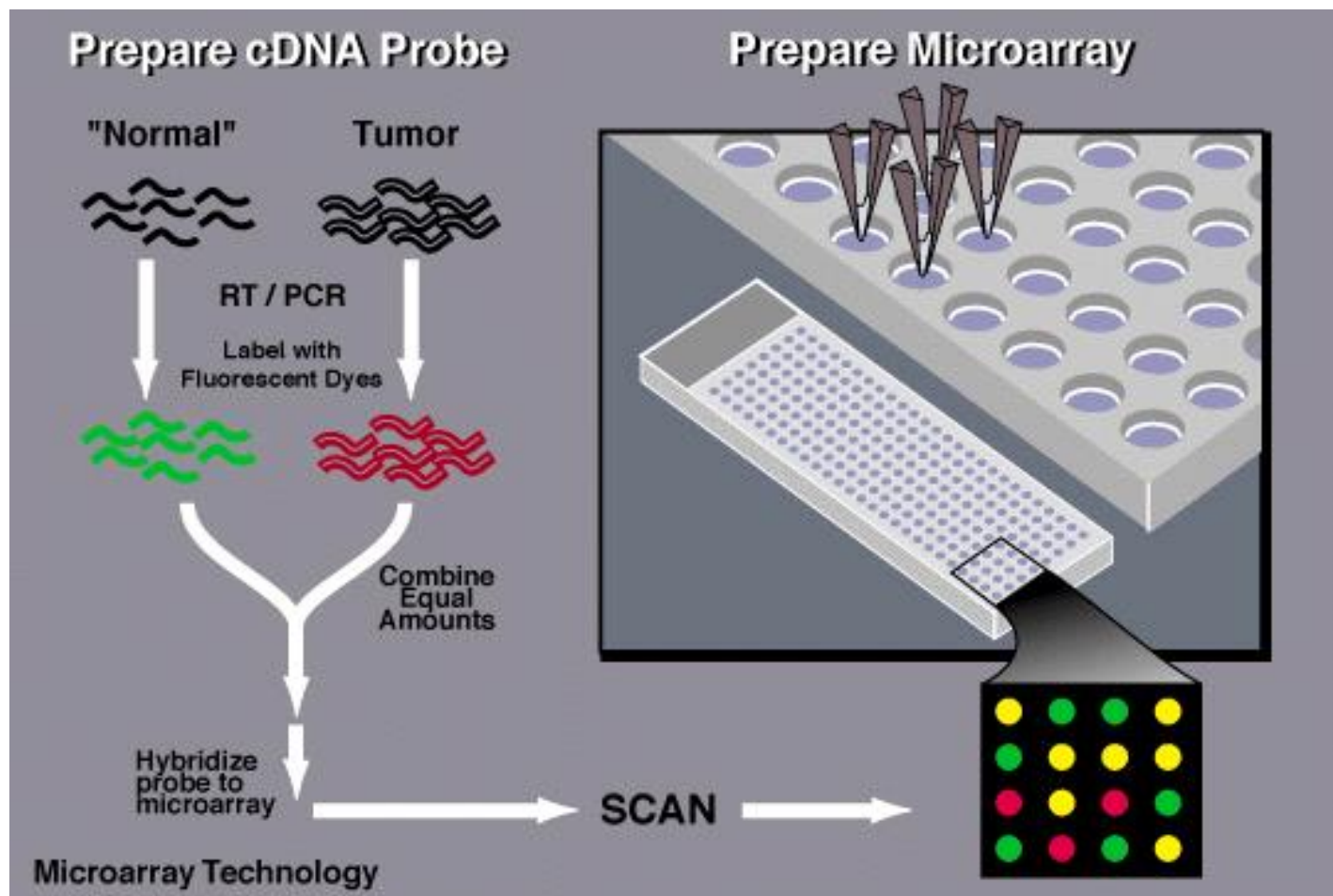
dr Jacek Śmietanski

jacek.smietanski@ii.uj.edu.pl

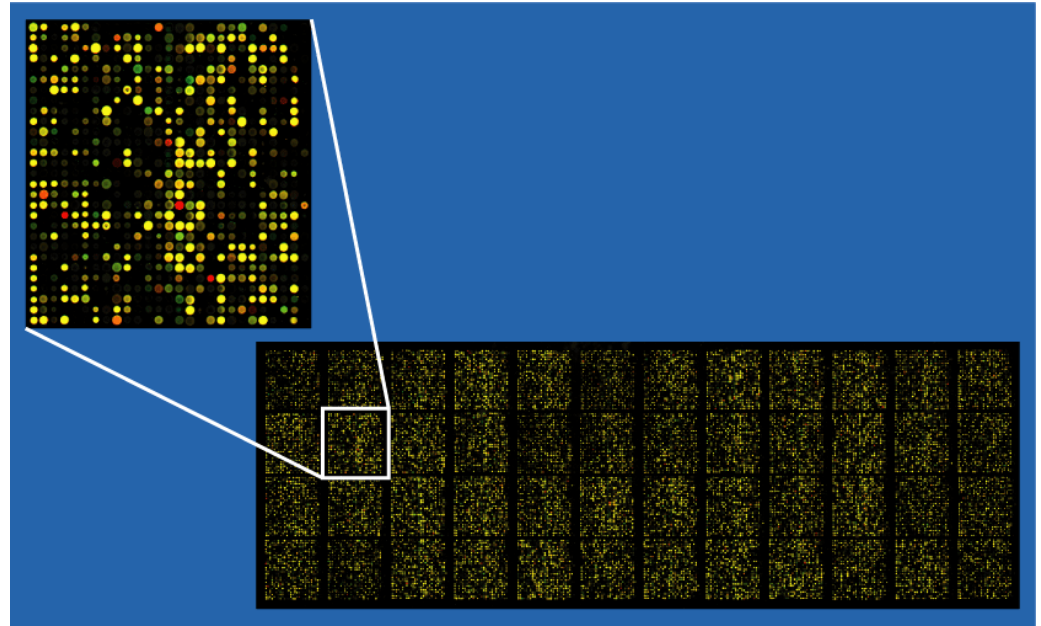
<http://jaceksmietanski.net>

Genom drożdży na mikromacierzy





Pozwalają na przeprowadzanie analiz genetycznych dla wielu (nawet tysiący) genów jednocześnie.
Badanie ekspresji genów.



Określenie zmian w poziomie ekspresji genów. W badaniu określa się obecność i ilość cząsteczek mRNA dla poszczególnych genów w danym momencie funkcjonowania komórki.

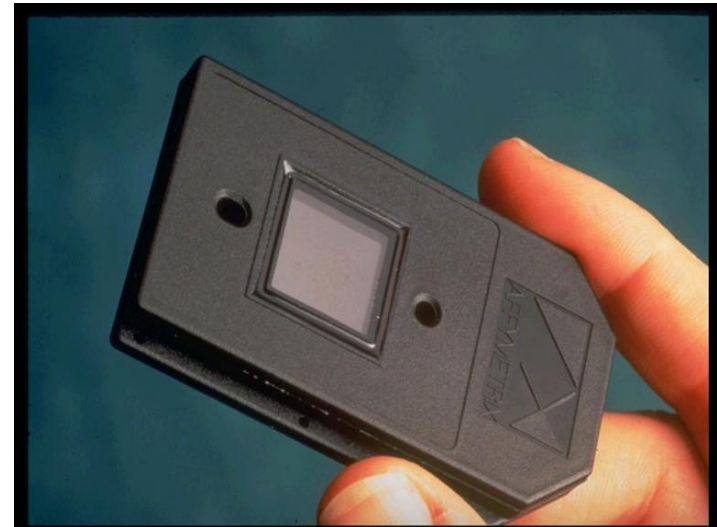
Np. jedna próbka pochodzi z komórki zdrowej, druga – z chorej.
Albo z komórek funkcjonujących w różnych warunkach środowiska (np. bakteria w warunkach tlenowych i beztlenowych).
Albo z komórek w różnych stadiach rozwoju (np. w różnych etapach mitozy).

Zwykle produkowana komercyjnie.

Każda płytka posiada tysiące ściśle określonych punktów, każdy zawiera nić pochodzącą z innego genu.

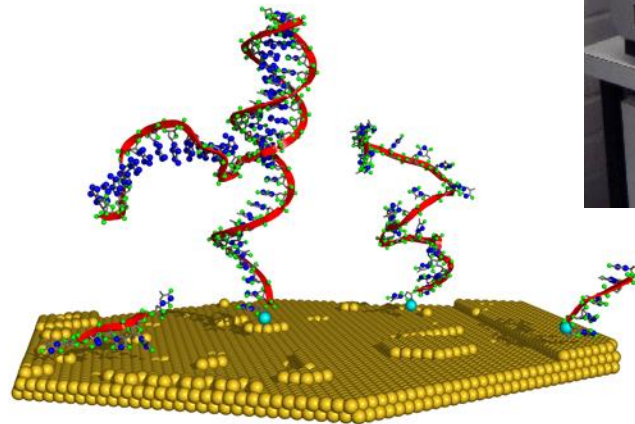
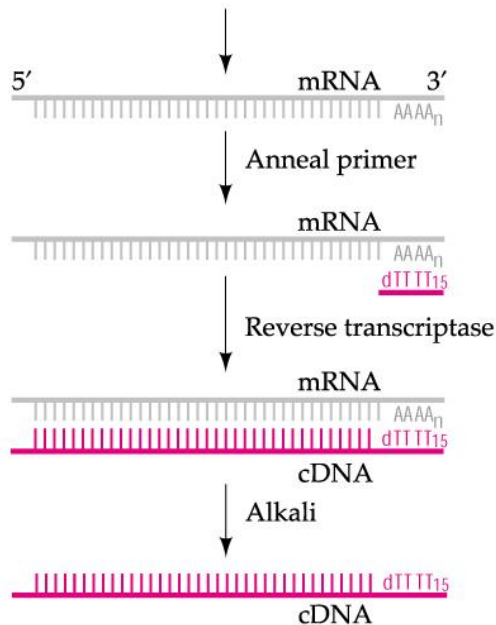
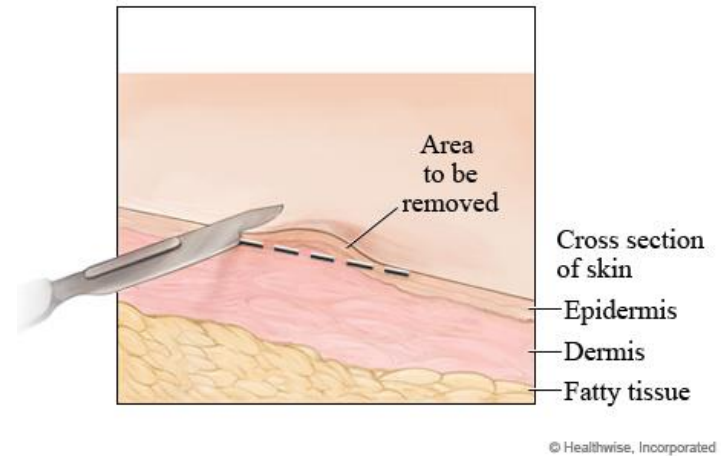
Nicią jest fragment nici cDNA lub syntetyczny oligonukleotyd.

Nici są dodawane automatycznie za pomocą igły aplikującej cDNA lub metodą podobną do tworzenia procesorów – z tego względu mikromacierze nazywane są też czipami genowymi (*Gene Chip*)

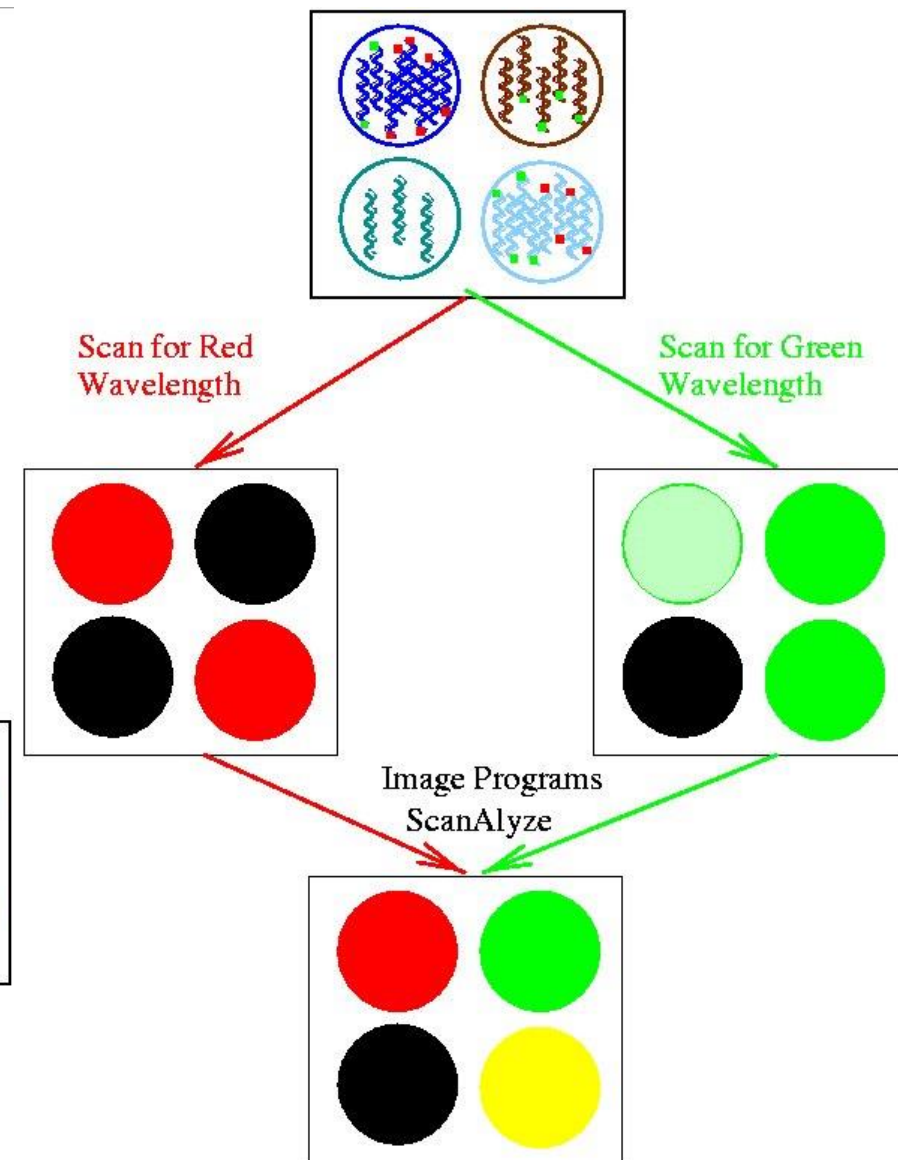
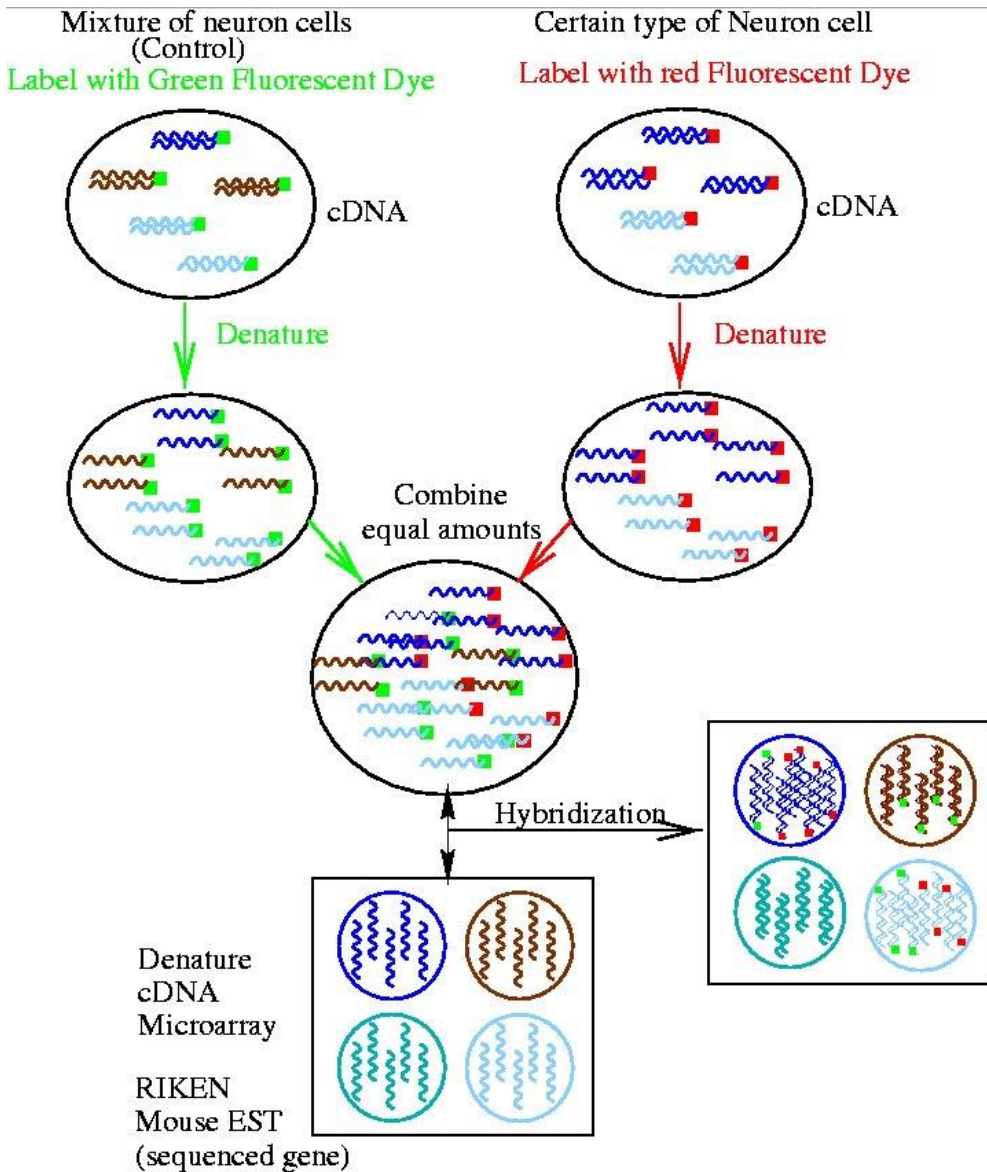


Przebieg eksperymentu

1. Pobranie próbek
2. Izolacja mRNA.
3. Odwrotna transkrypcja do cDNA.
4. Hybrydyzacja na płytce
5. Skanowanie mikromacierzy
6. Analiza danych



Eksperyment



ZIELONY - geny z próbki kontrolnej, które hybrydyzowały bardziej niż w badanej.

CZERWONY - geny z próbki badanej, które hybrydyzowały bardziej niż w kontrolnej.

ŻÓŁTY - geny z obu próbek hybrydyzowały w podobnym stopniu

CZARNY - obszary, w których żadna próbka nie hybrydyzowała do danej sekwencji DNA

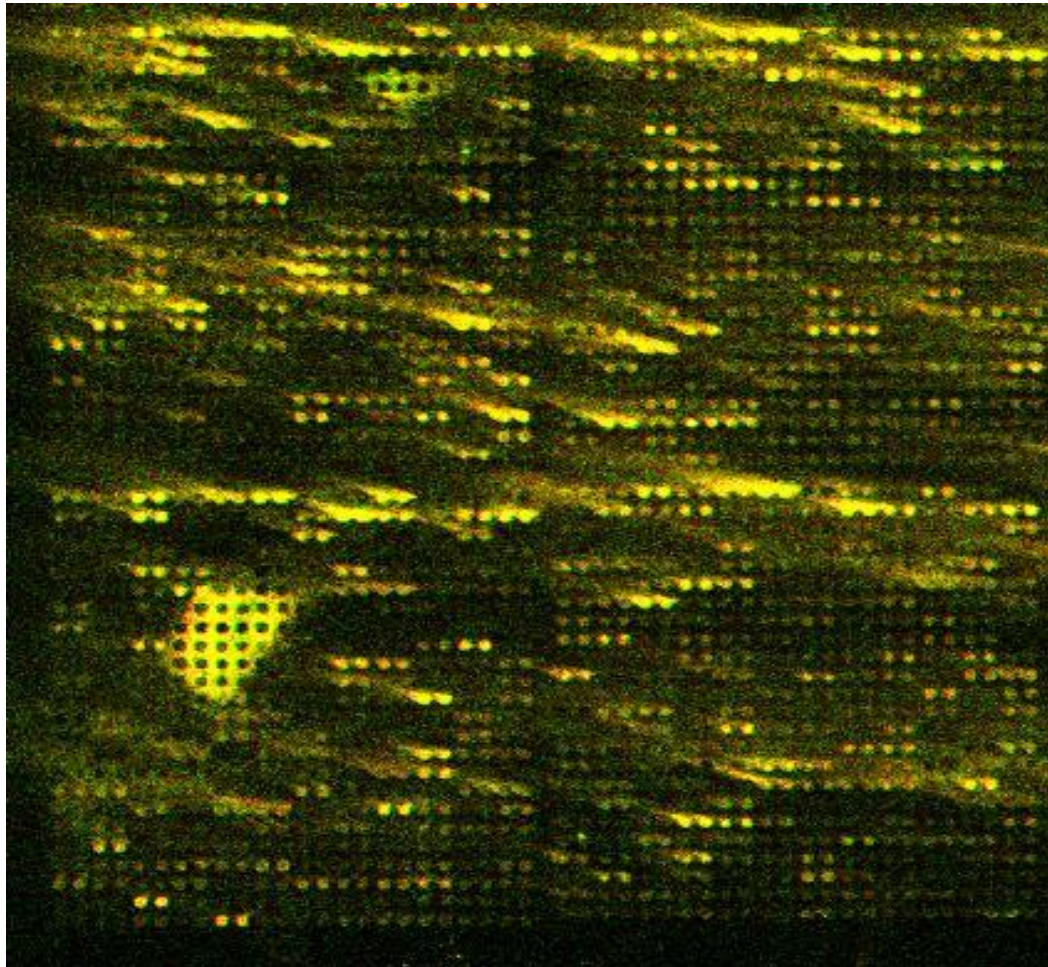
Analiza różnic w ekspresji genów pomiędzy dwiema próbkami pozwala nam zrozumieć rolę genów w poszczególnych stadiach życiowych czy w razie choroby komórki.



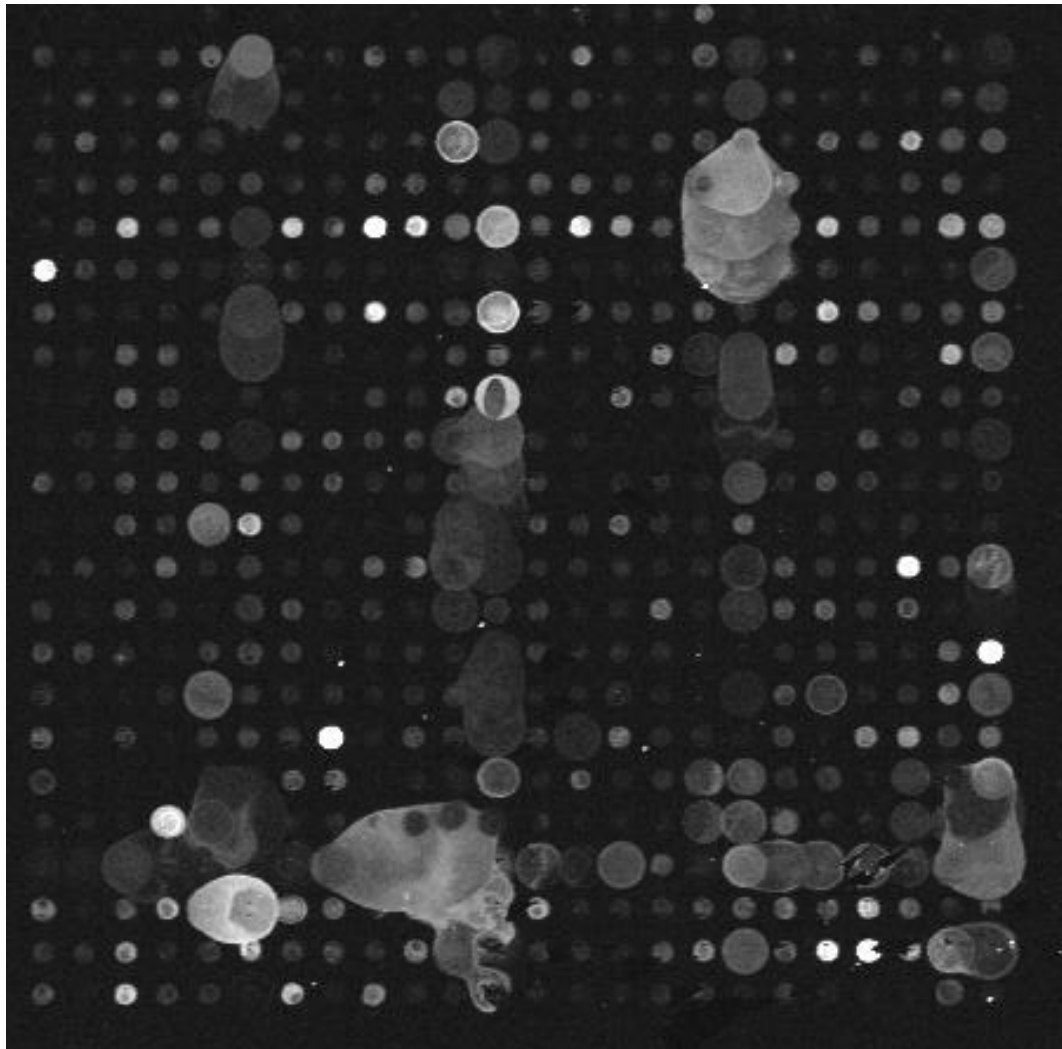
- 1) Identyfikacja punktów na obrazie odpowiadającym poszczególnym genom
- 2) Odczyt intensywności kolorów
- 3) Analizy statystyczne, itp.

Ogon komety (*comet tail*)

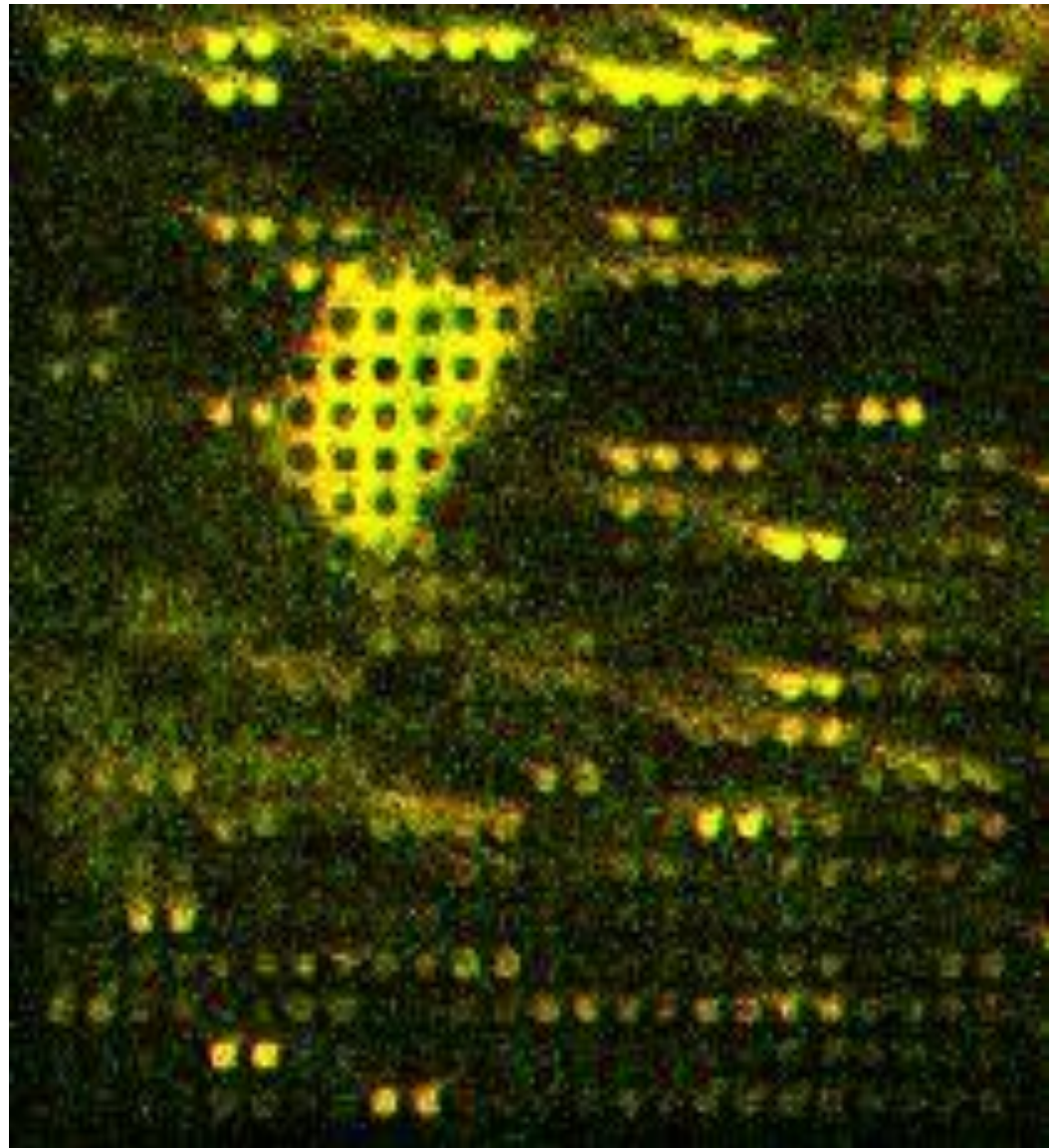
Najprawdopodobniej spowodowane niedostatecznie szybkim zanurzeniem próbki w roztworze.



Plamy, nierówne rozmiary, nachodzące na siebie



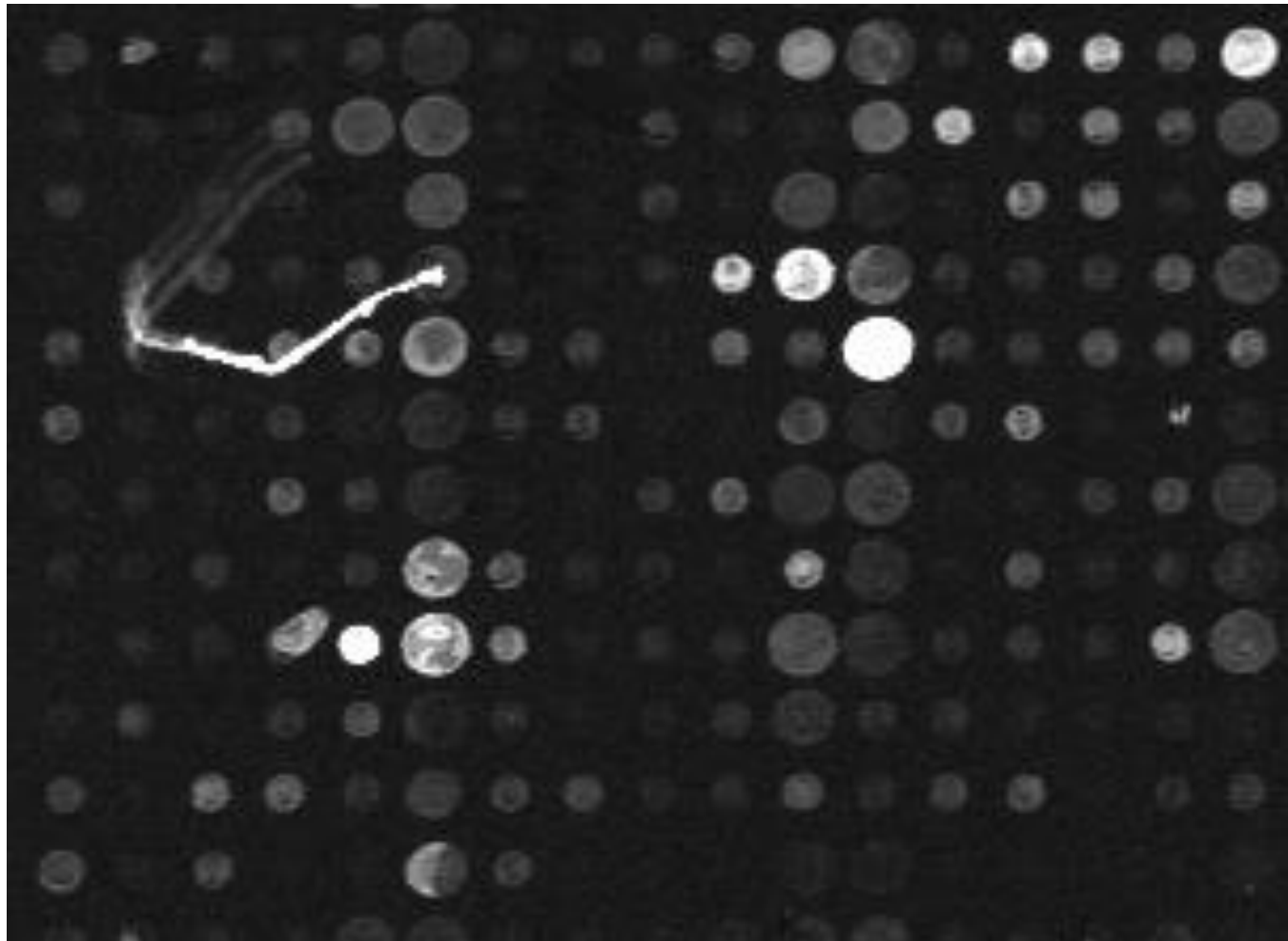
Jasne / nierówne tło.



Zachodzenie punktów.



Artefakty



Rozdzielczość: $10\text{ }\mu\text{m}$

Standardowy rozmiar punktu: $100\text{ }\mu\text{m}$

⇒ Średnica obiektu: 10 pixeli

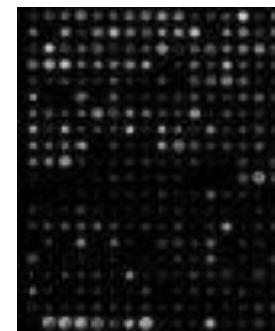
Format obrazu:

TIFF (tagged image file format) 16 bit (65,536 poziomów szarości)

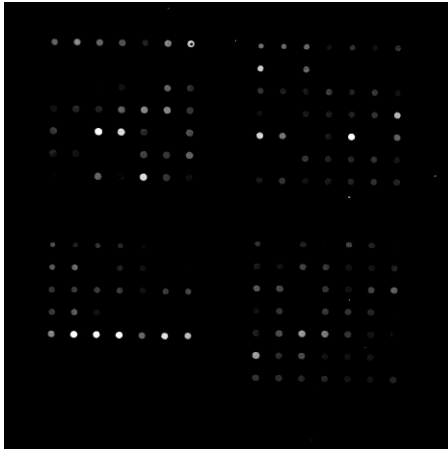
Obraz $1\text{ cm} \times 1\text{ cm}$ (16 bit) = 2Mb (bez kompresji)

Istnieją też inne formaty, np. SCN (Stanford)

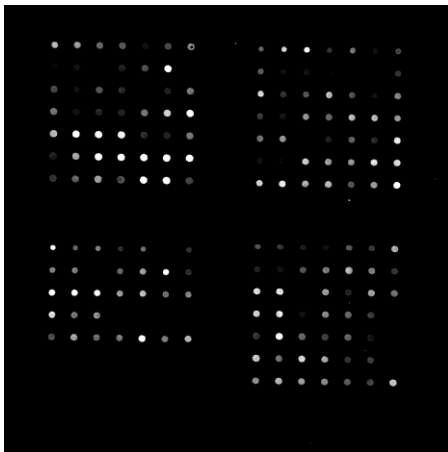
Oddzielne obrazy dla poszczególnych próbek.



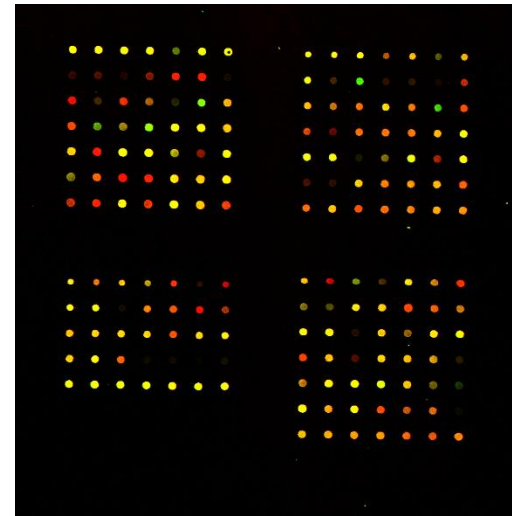
Cy3



Cy5



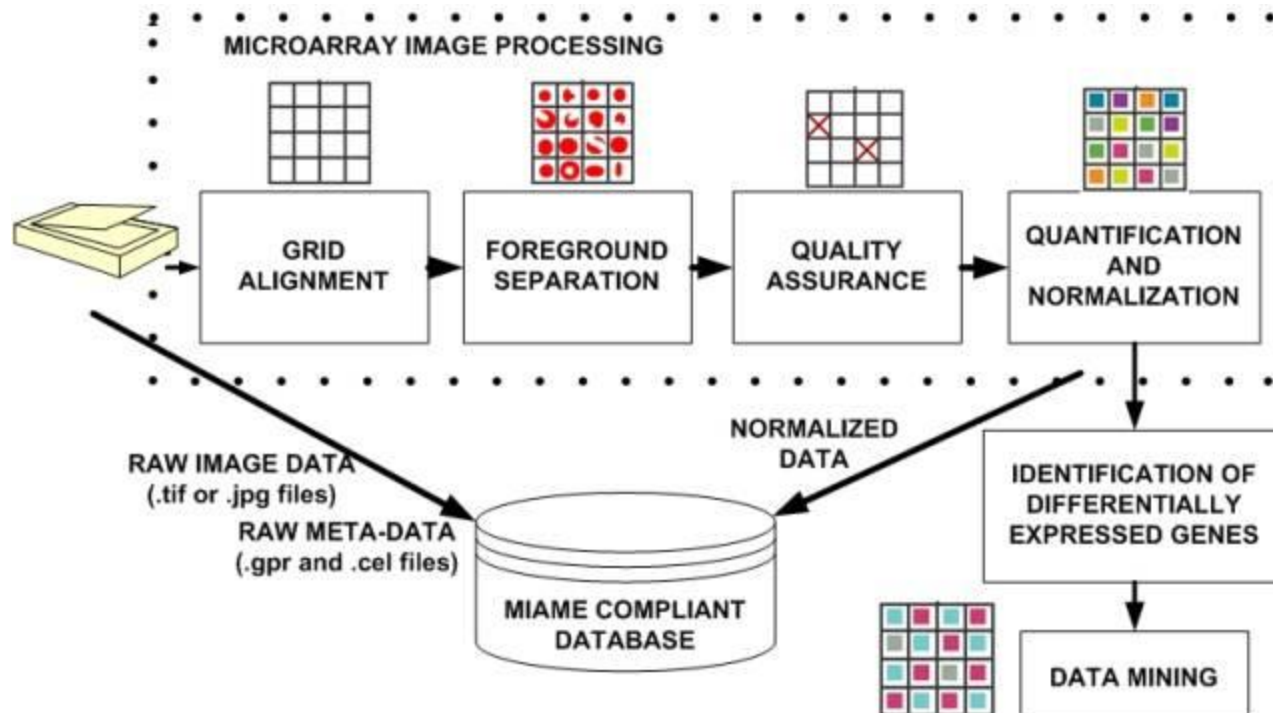
Pseudo-color overlay



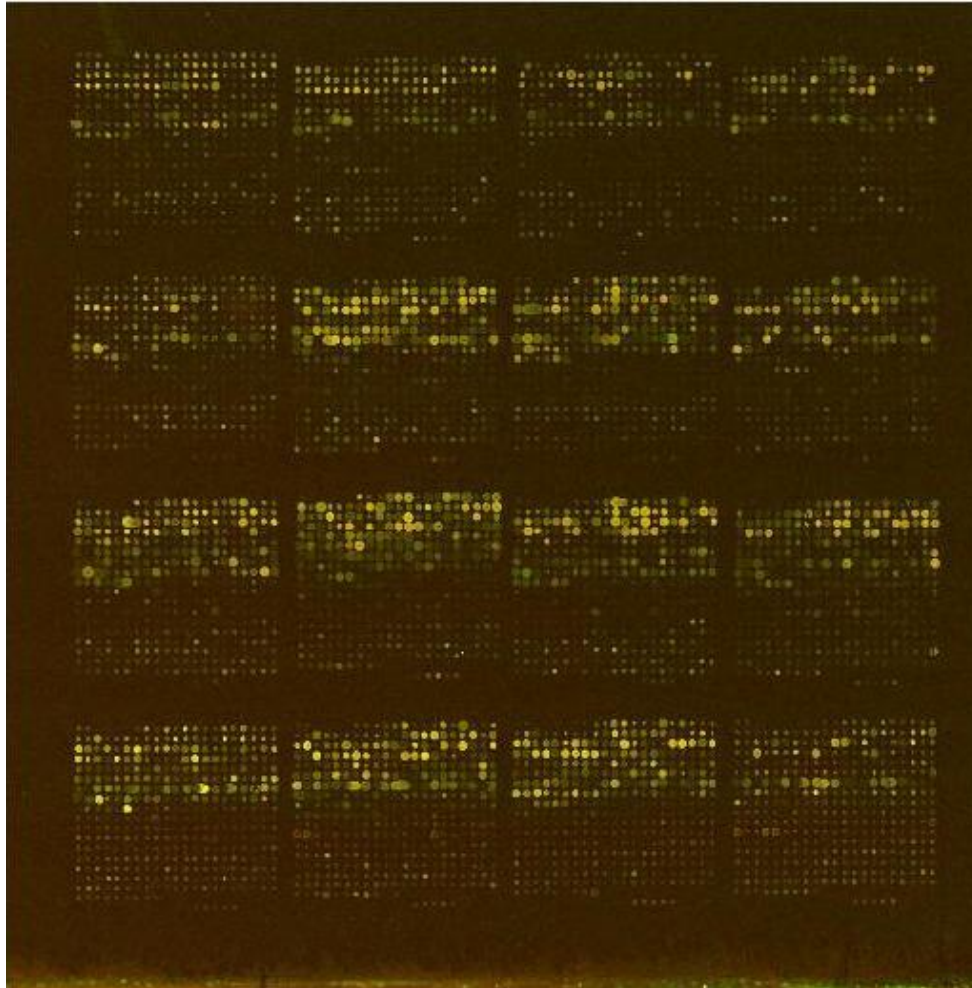
<i>Spot color</i>	<i>Signal strength</i>	<i>Gene expression</i>
yellow	Control = Treated	unchanged
red	Control < Treated	induced
green	Control > Treated	repressed

Etapy przetwarzania obrazów

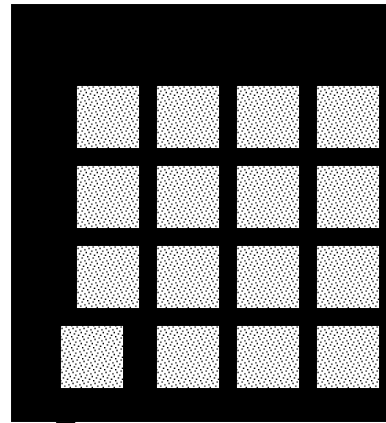
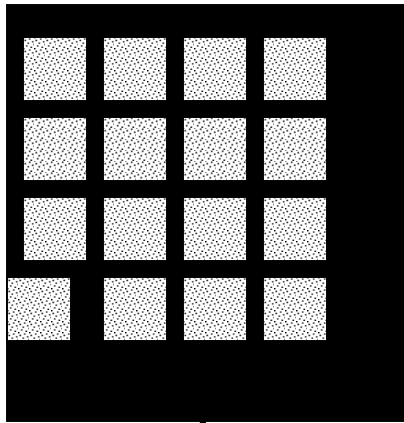
- Adresowanie (tworzenie siatki)
- Przypisanie współrzędnych do każdego punktu
- Segmentacja
- Oddzielenie sygnału od tła
- Analiza intensywności sygnału
- Analiza jakości pomiaru



This is the process of assigning coordinates to each of the spots.
Automating this part of the procedure permits high throughput analysis.



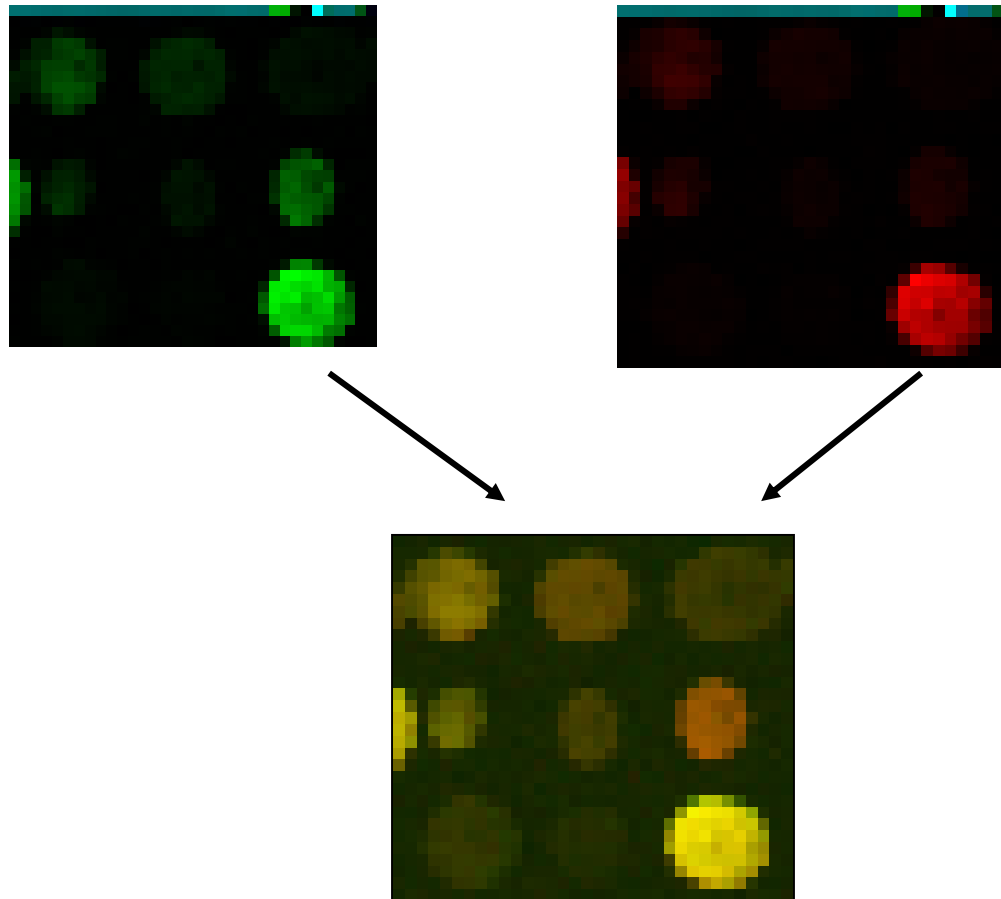
4 by 4 grids
19 by 21 spots per grid



Within the same batch of print runs; estimate translation of grids

4 by 4 grids

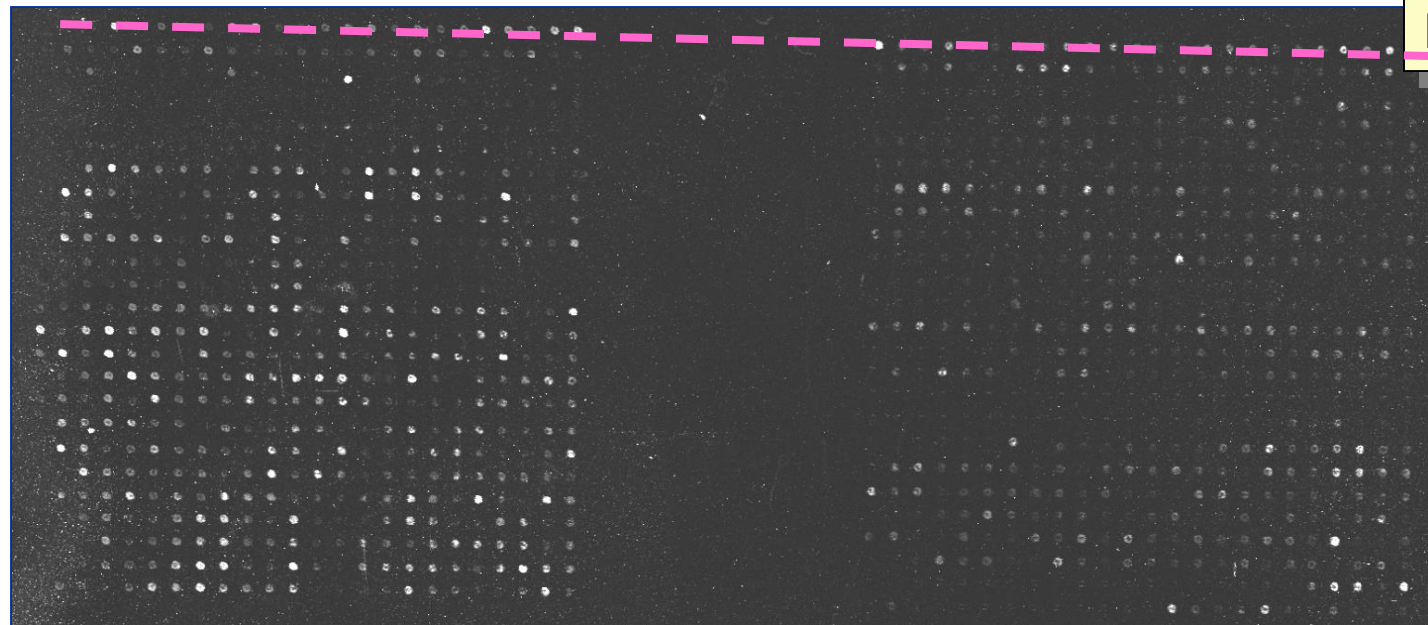
- Other problems:
- Misregistration
 - Rotation
 - Skew in the array



Misregistration of the red and green channels

Rotation of the array in the image

Skew in the array



Rotation

- Analiza podstawowej struktury obrazu (determinowanej przez urządzenie)
- Pozycjonowanie macierzy na obrazie
- Rozdział na wiersze i kolumny w grupach (*grids*)
- Identyfikacja względnego przesunięcia grup
- Rozdział na wiersze i kolumny wewnątrz każdej grupy
- Korekty (przesunięcia) pojedynczych punktów

New Grid

Grids To Add

☐ 1
☒ 4
☐ 16
☐ 32

Columns

Rows

Col Spacing (um)

Row Spacing (um)

XRes (um)

YRes (um)

Tip Spacing (um)

Spot Width (pixels)

Spot Height (pixels)

First Grid Position

Left

Top

Create Cancel

ScanAlyze

Klasyfikacja pikseli do tła lub sygnału.

Metody:

- *Fixed circles*
- *Adaptive circles*
- *Adaptive shape*
- *Edge detection*
- *Seeded Region Growing* (R. Adams and L. Bischof (1994): rozrost obszarów począwszy od „punktów zasiewu” stosownie do różnicy pomiędzy jasnościami pikseli i średniej jasności sąsiadujących obszarów
- Metody bazujące na histogramie



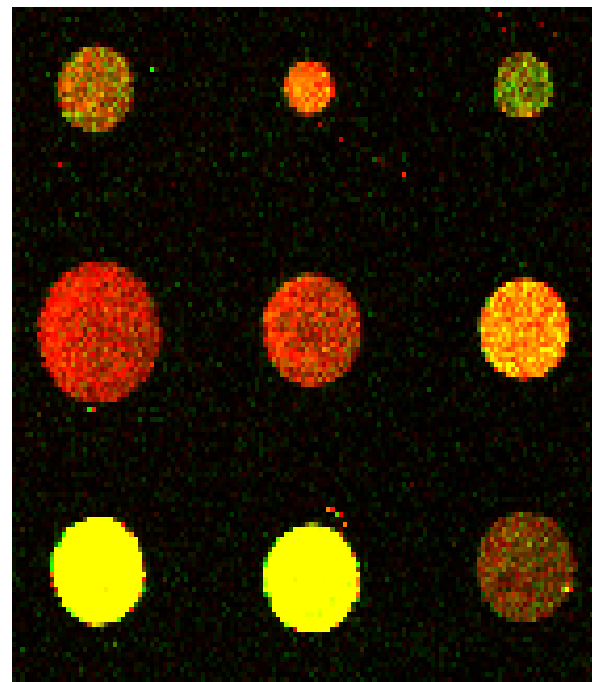
Fixed circle	ScanAlyze, GenePix, QuantArray
Adaptive circle	GenePix, Dapple, SignalViewer (uses ellipse)
Adaptive shape	Spot, region growing and watershed
Histogram	ImaGene, QuantArray, DeArray and adaptive thresholding

Dopasowuje okręgi o zadanej na sztywno średnicy do wszystkich obszarów na obrazie.

Łatwa w implementacji.

Obszary powinny mieć ten sam kształt i wielkość.

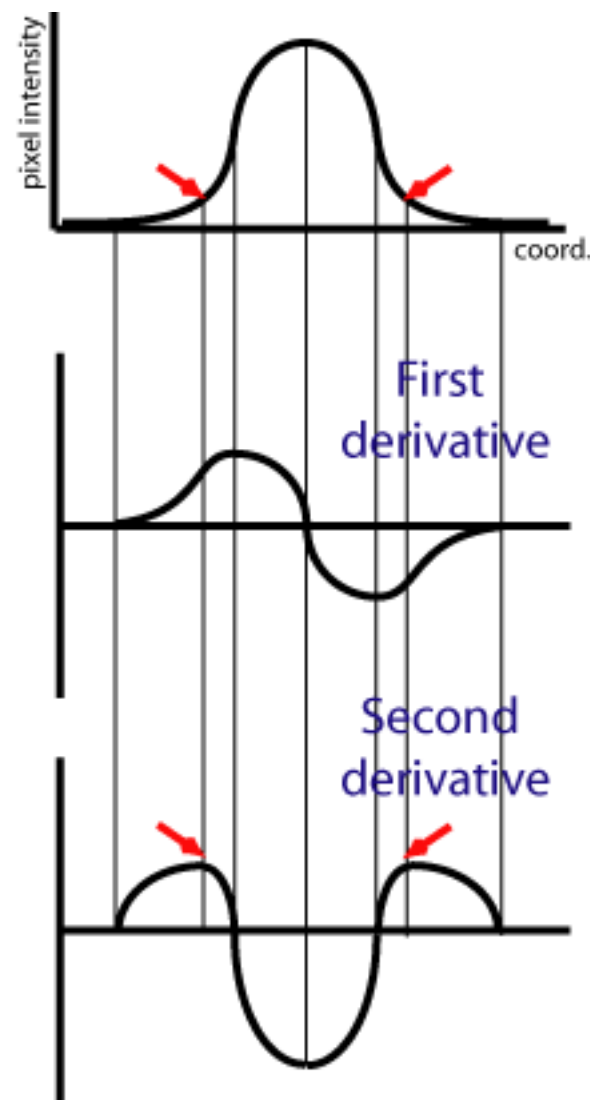
*May not be good
for this example*

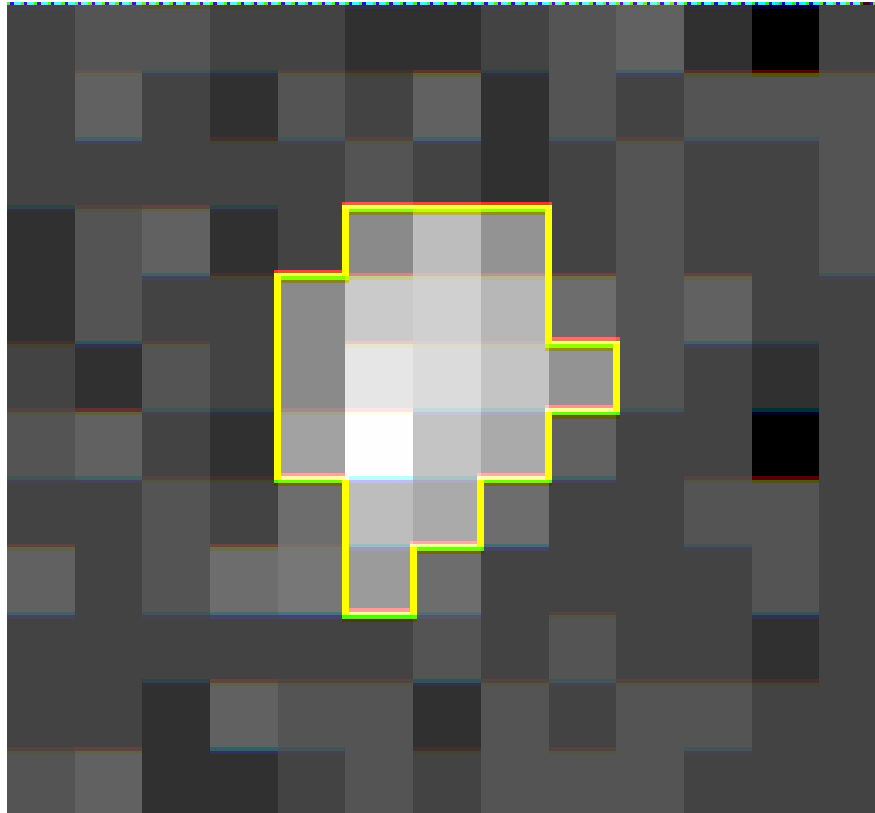


Średnica okręgu jest szacowana indywidualnie dla każdego obszaru.

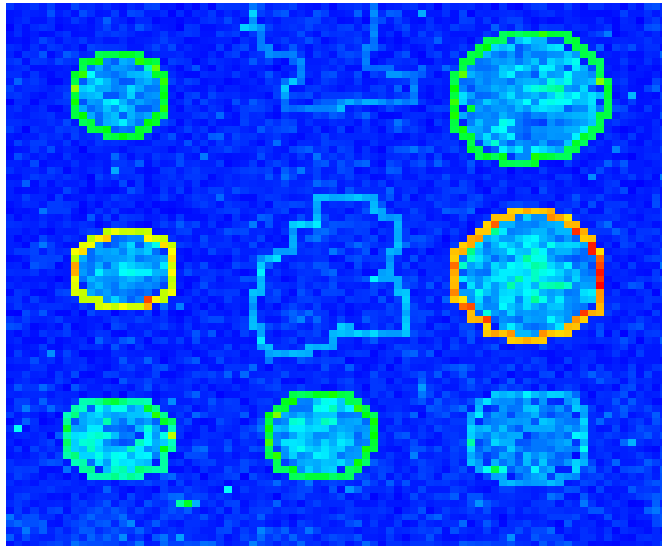
Wykorzystywane są tu metody gradientowe wykrywania krawędzi (*second derivative*)

Problematyczna dla nieokrągłych kształtów, np. owalnych.

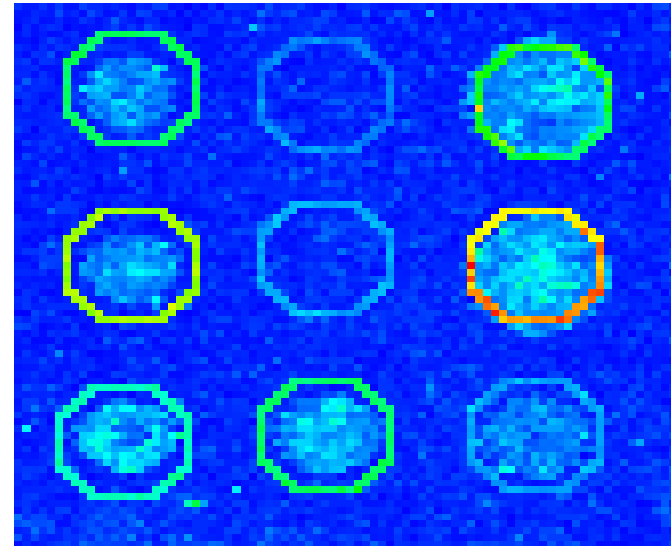




- Zbyt mały obszar
- Kształt inny niż okrąg



SRG

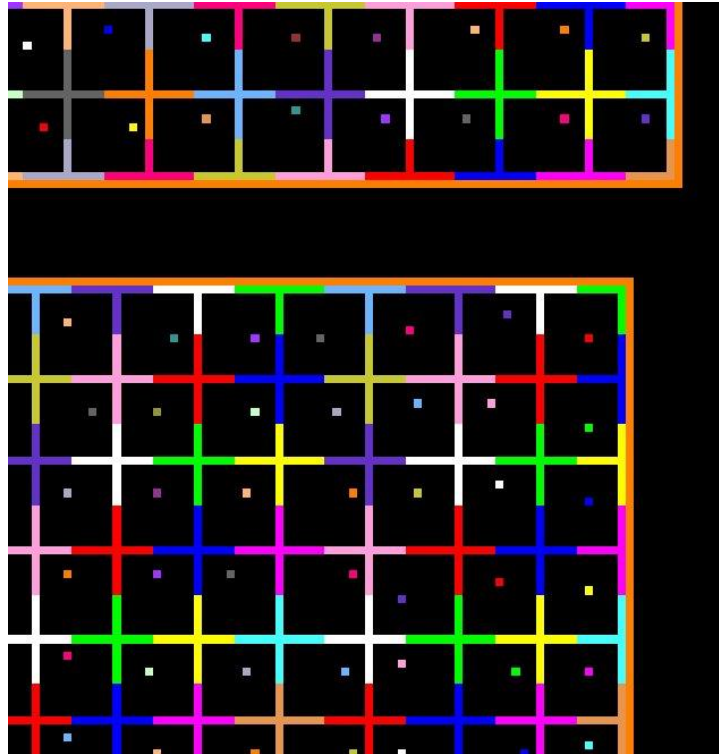
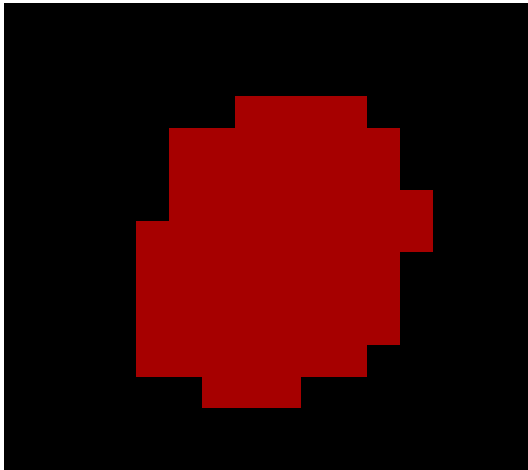


Fixed circle

Specification of starting points or seeds

Bonus: already know geometry of array

Regions grow outwards from the seed points preferentially according to the difference between a pixel's value and the running mean of values in an adjoining region



Segmentacja na podstawie histogramu

Choose target mask larger than any spot

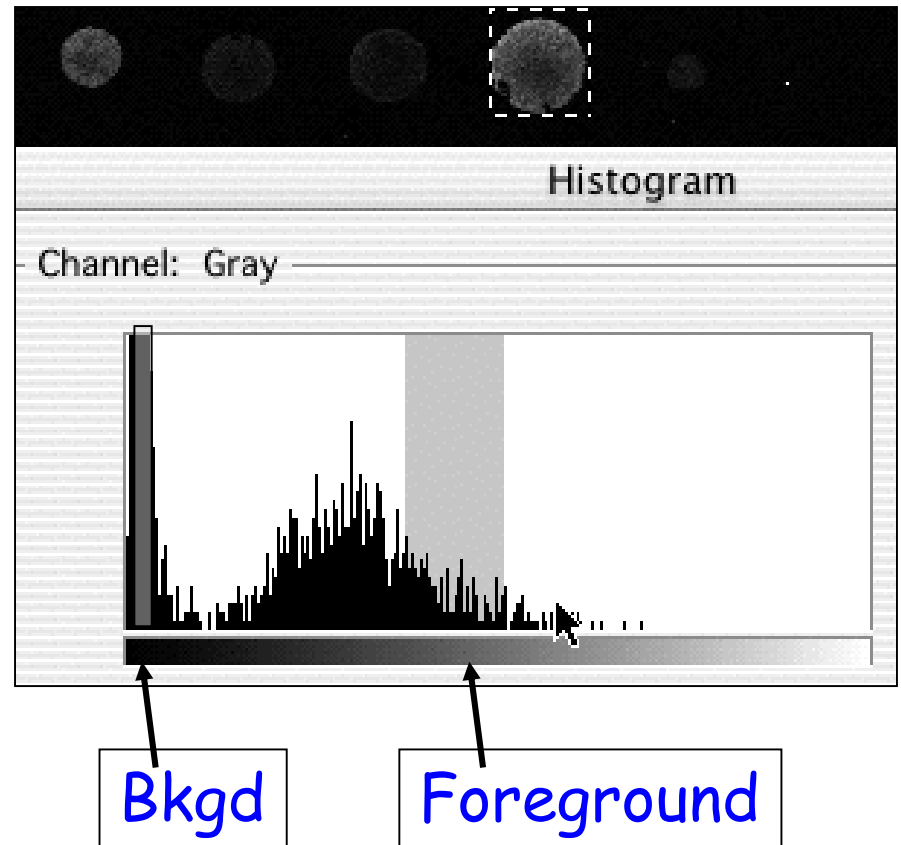
Fg and bg intensities determined from the histogram of pixel values for pixels within the masked area.

Example : QuantArray

Background : mean between
5th and 20th percentile

Foreground : mean between
80th and 95th percentile

May not work well when a large
target mask is set to compensate
for variation in spot size



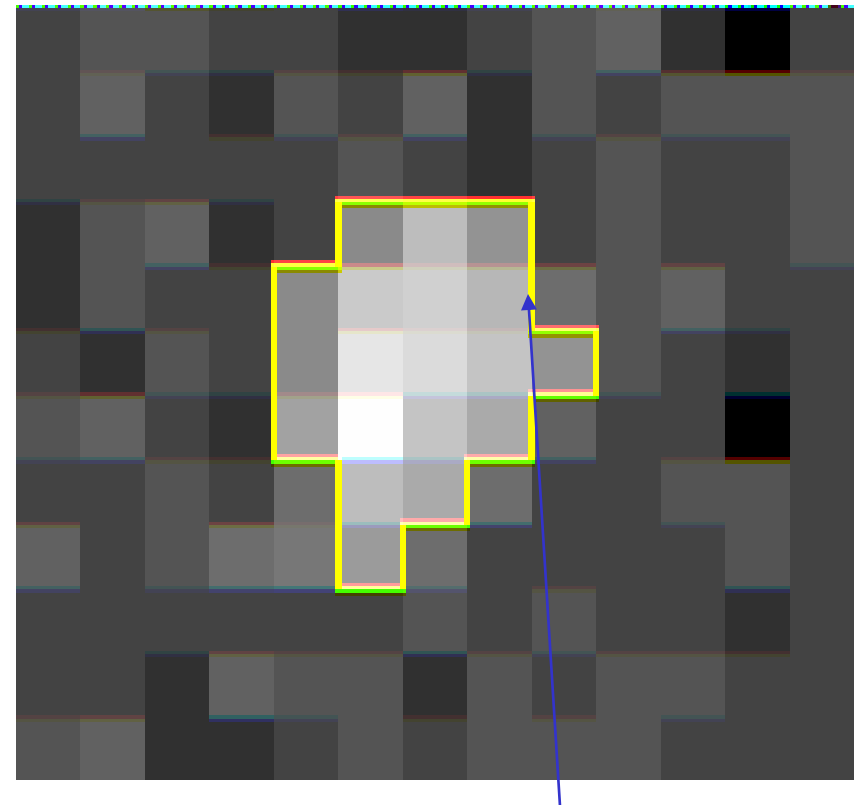
Spot Intensities

- mean of pixel intensities
- median of pixel intensities
- pixel variation (e.g. IQR)

Background values

- none
- local
- constant (global)
- morphological opening

Quality Information



Take the average

Mierzony w eksperymencie poziom fluorescencji uwzględnia również wpływ niespecyficznych hybrydyzacji i zanieczyszczeń na płytce.

Poziom fluorescencji z regionów nie zawierających DNA powinien być inny niż z regionów badanych.

Rozwiązanie: stosowanie kontroli negatywnej: fragment DNA, co do którego wiemy, że nie będzie

None: nie uwzględniamy tła

Probably not accurate in many cases, but may be better than some forms of local background determination

Local:

- Focusing on small regions surrounding the spot mask
- Median of pixel values in this region
- Most software package implement such an approach
- By not considering the pixels immediately surrounding the spots, the background estimate is less sensitive to the performance of the segmentation procedure

Global method which subtracts a constant background for all spots.

Some evidence that the binding of fluorescent dyes to 'negative control spots' is lower than the binding to the glass slide.

More meaningful to estimate background based on a set of negative control spots.

If no negative control spots :
approximation of the average background =
third percentile of all the spot foreground values

Non-linear filtering, used in Spot.

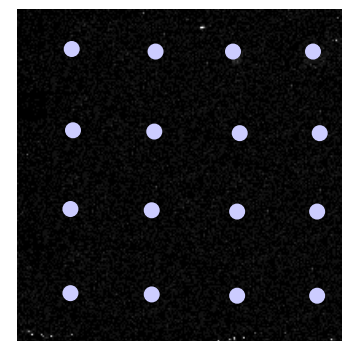
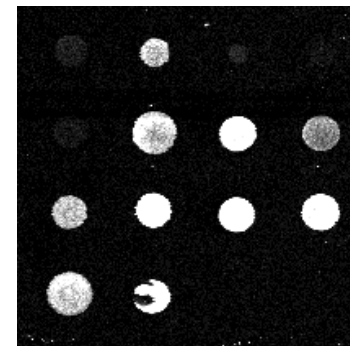
Use a square structuring element with side length at least twice as large as the spot separation distance.

Compute local minimum filter, then compute local maximum filter.

This removes all spots and generates an image that is an estimate of the background for the entire slide.

For individual spots, the background is estimated by sampling this background image at the nominal center of the spot.

Lower, less variable bg estimate .



Array

- Correlation between spot intensities
- Percentage of spots with no signals
- Distribution of spot signal area

Spot

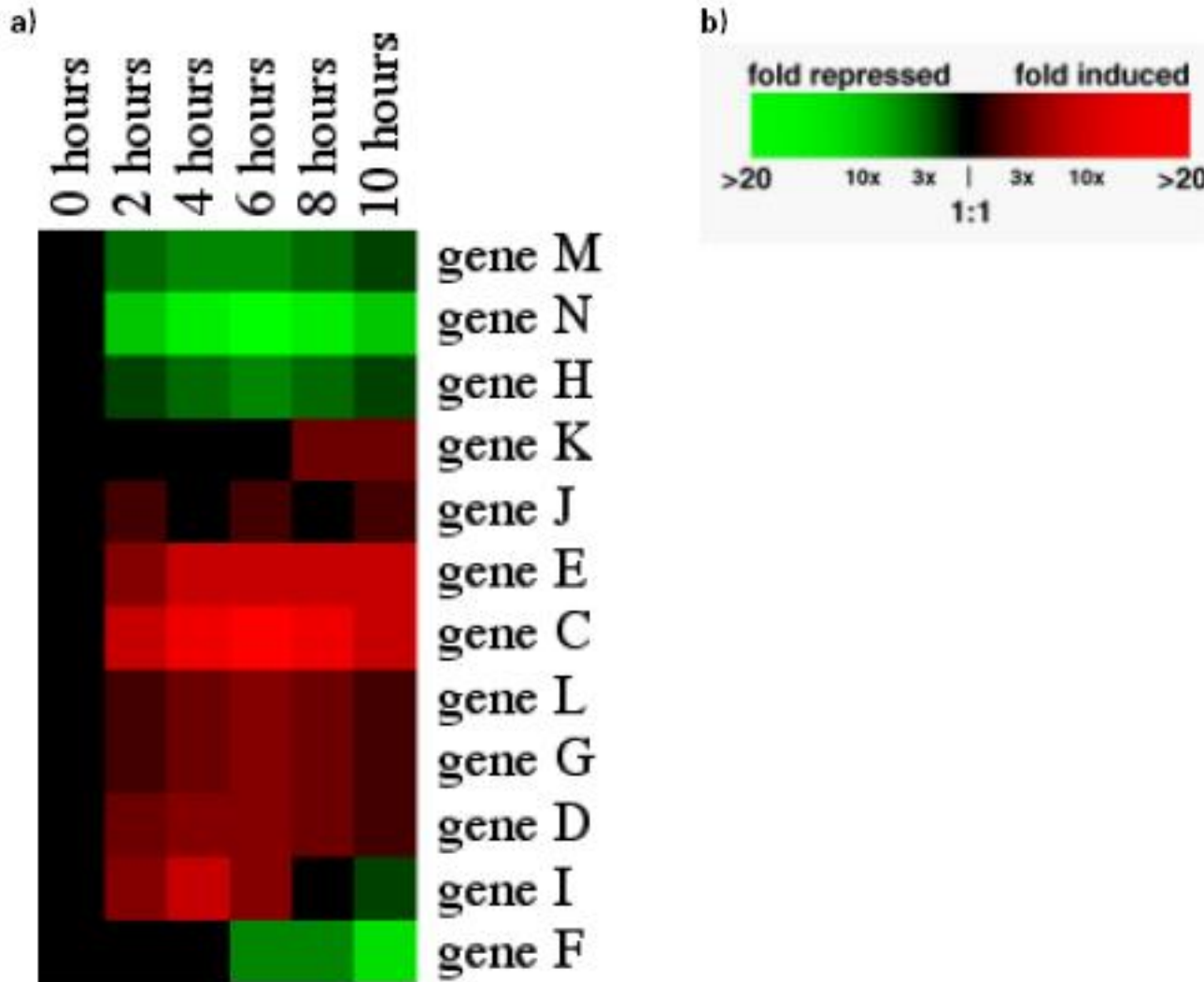
- Signal / Noise ratio
- Variation in pixel intensities
- Identification of “bad spot” (spots with no signal)

Ratio (2 spots combined)

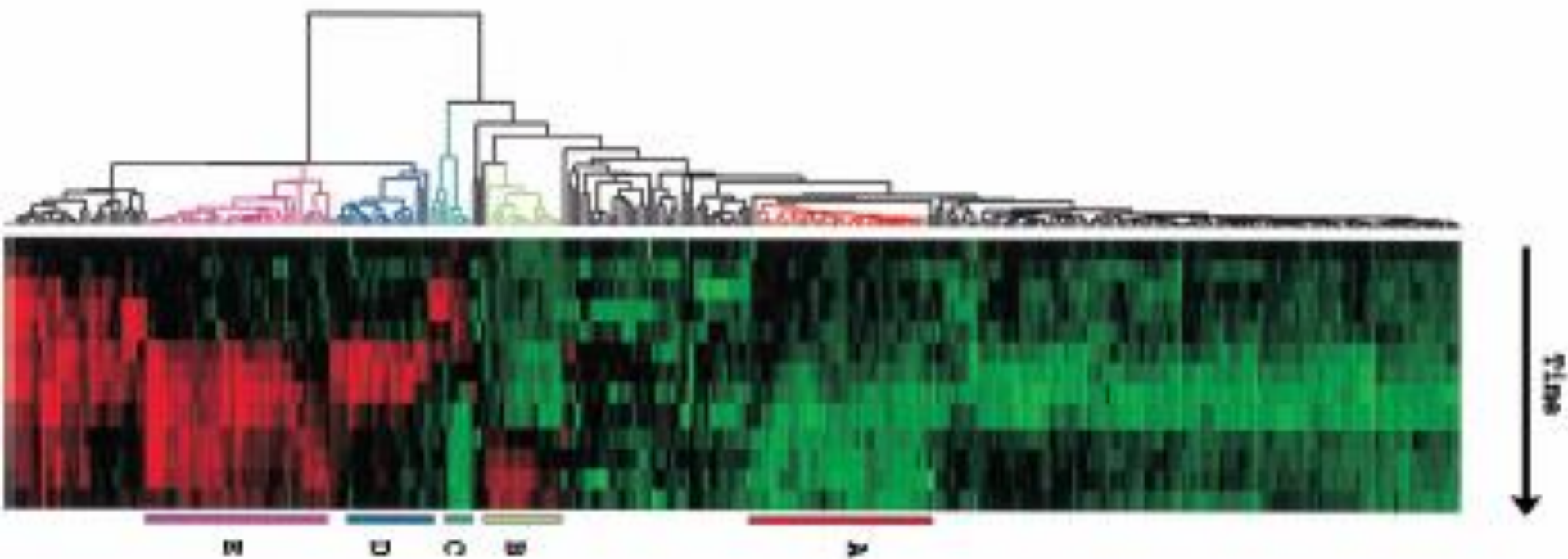
- Circularity

Flag or weight spots based on these (or other appropriate) criteria

Grupowanie genów wg stopnia aktywności w poszczególnych próbach.



Klastrowanie hierarchiczne





Gene discovery.

Disease diagnosis: classify the types of cancer on the basis of the patterns of gene activity in the tumor cells.

Pharmacogenomics = is the study of correlations between therapeutic responses to drugs and the genetic profiles of the patients.

Toxicogenomics – microarray technology allows us to research the impact of toxins on cells. Some toxins can change the genetic profiles of cells, which can be passed on to cell progeny.

<http://www.ncbi.nlm.nih.gov/geo/>



Gene Expression Omnibus

[GEO Publications](#) | [FAQ](#) | [MIAME](#) | [Email GEO](#) | [Login](#)

NCBI > GEO

Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

QUERY

DataSets

Gene profiles

GEO accession

GEO BLAST

DataSets

GEO accessions

Platforms

Samples

Series

Submitter login

User id:

Password:

[» New account](#)

[» Recover password](#)

Site contents

Public data

Platforms	9,983
Samples	726,842
Series	29,463
DataSets	2,720

Documentation

[Overview](#) | [FAQ](#) | [Find](#)
[Submission guide](#)
[Linking & citing](#)
[Journal citations](#)
[Construct a Query](#)
[Programmatic access](#)
[DataSet clusters](#)
[GEO announce list](#)
[Data disclaimer](#)
[GEO staff](#)

Query & Browse

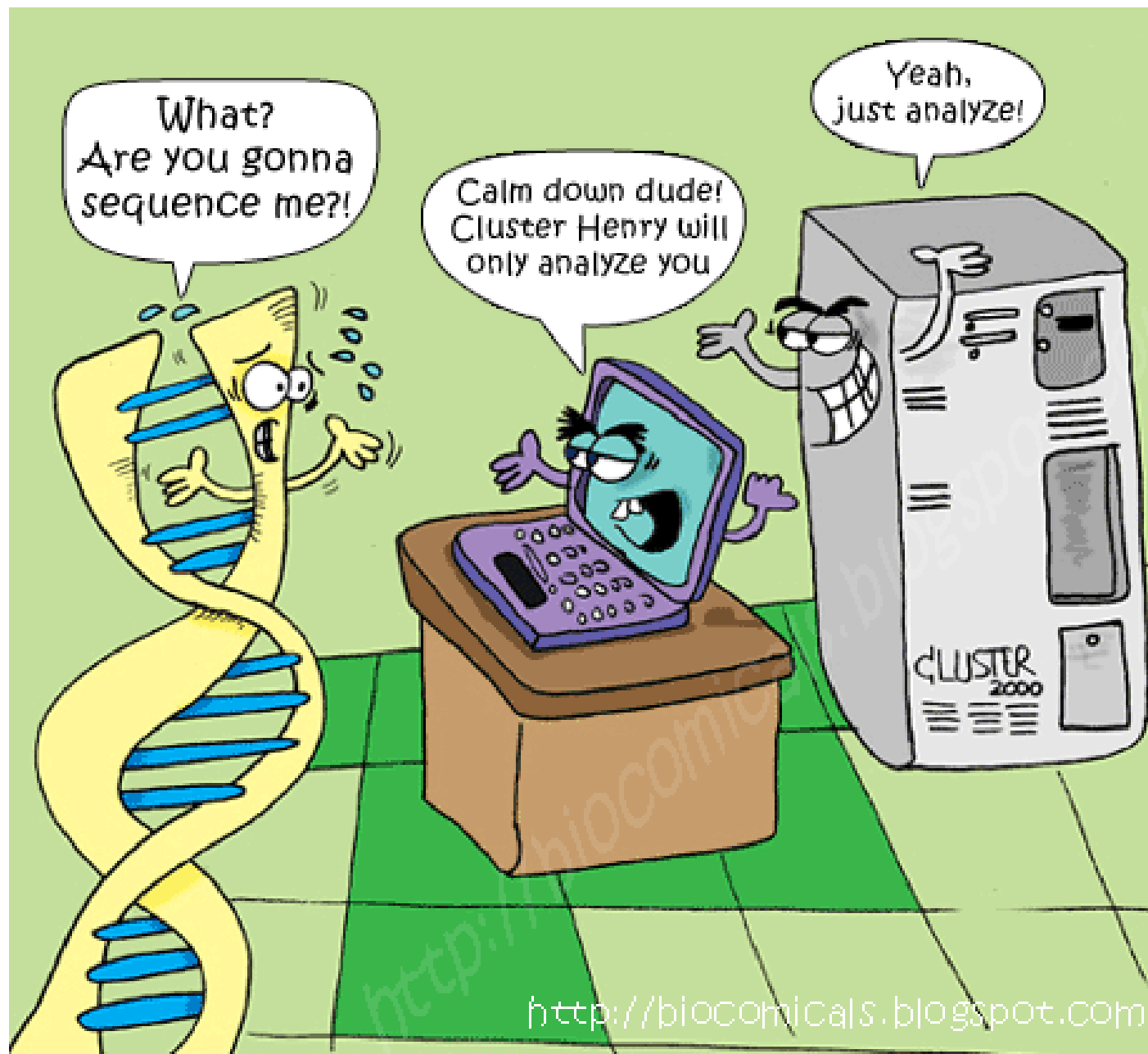
[Repository browser](#)
[GEO2R](#)
[FTP site](#)
[GEO Profiles](#)
[GEO DataSets](#)

Submit

[New account](#)

RNA Seq

Analiza RNA dzięki wykorzystaniu wysokoprzepustowych technik sekwencjonowania.



What?
Are you gonna
sequence me?!

Calm down dude!
Cluster Henry will
only analyze you

Yeah,
just analyze!