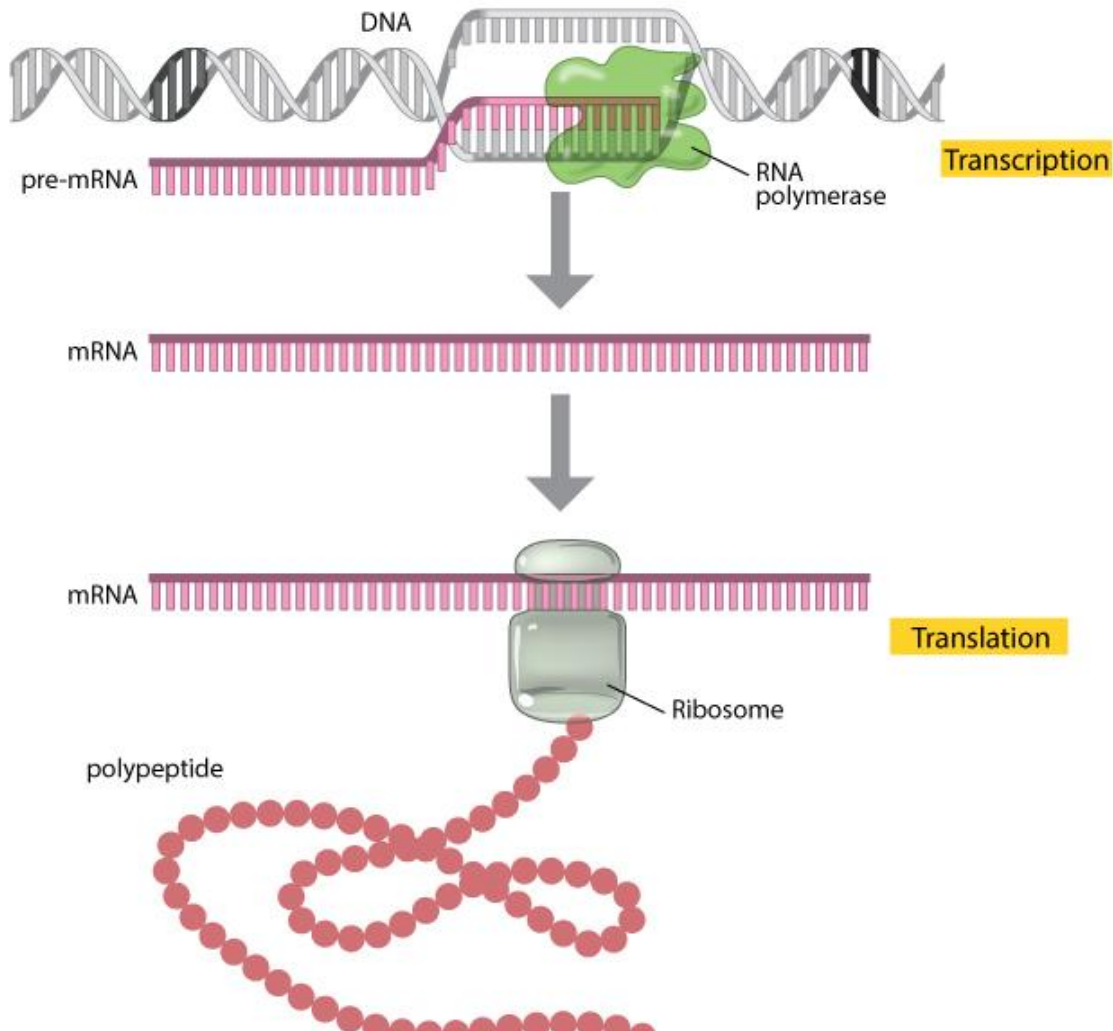


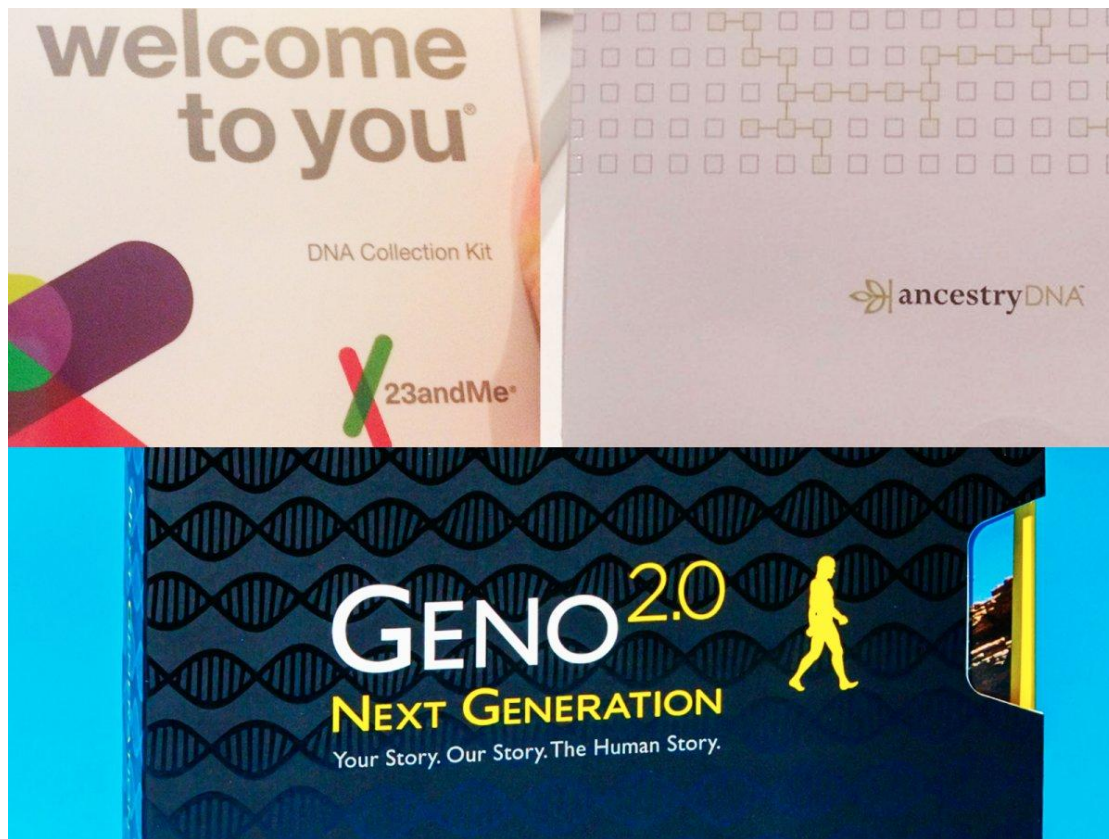
# Wykorzystanie data mining w genetyce



Odczytywanie informacji  
genetycznej w komórce

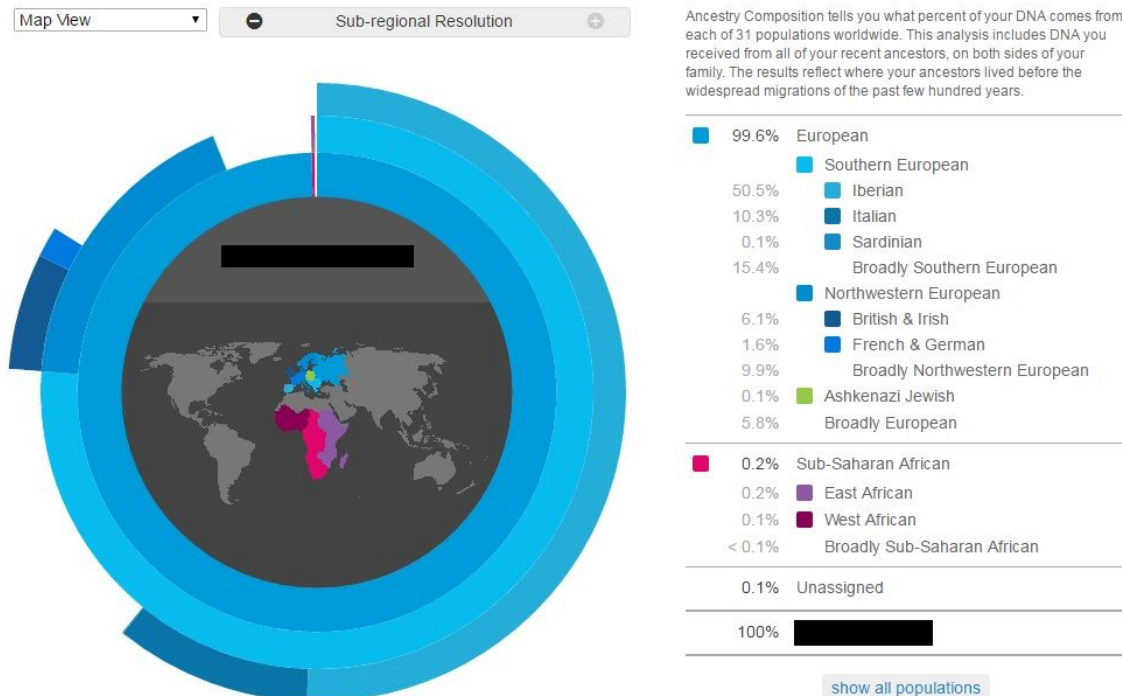
(Clancy, 2008)

# Identyfikacja pochodzenia etnicznego



# Identyfikacja pochodzenia etnicznego

## 23andme przykładowy raport



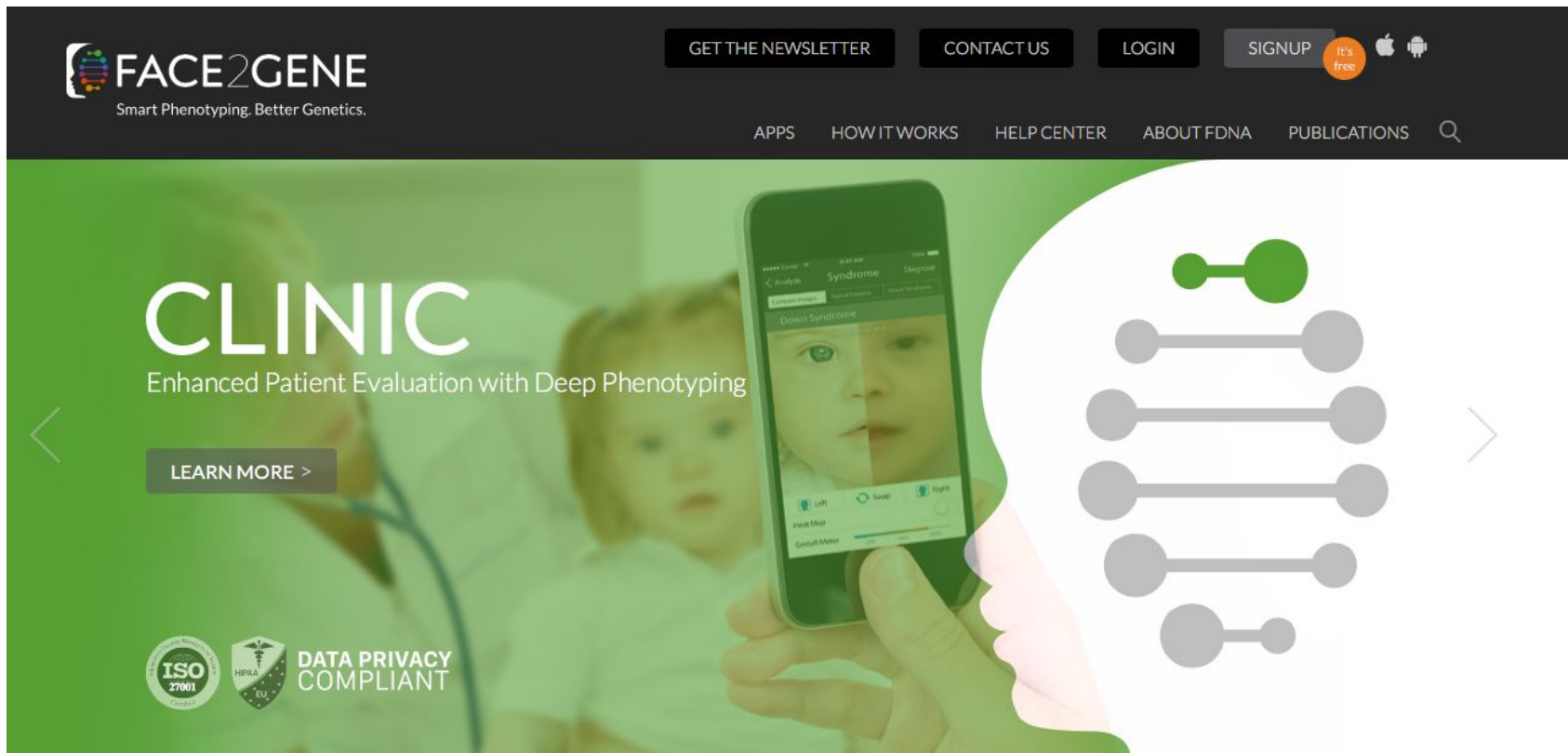
# Diagnostyka chorób rzadkich

## Przykłady chorób rzadkich

Nazwa	Częstość
Mukowiscydoza	1/2500
Syndrom Brugada	1/5000
Neurofibromatoza typu I	1/10000
Miopatia Nonaki	1/100000
Syndrom Warburga Micro 1	1/1000000

# Diagnostyka chorób rzadkich

## Rozpoznawanie dysmorfii twarzy



The banner features a dark header with the FACE2GENE logo and navigation links. The main content area has a green background with a blurred image of a doctor and a child. A hand holds a smartphone displaying the app's interface, which includes a 'Syndrome' section and a 'Down Syndrome' image. To the right is a stylized DNA helix graphic. The bottom left corner contains ISO and HIPAA compliance logos.

**FACE2GENE**  
Smart Phenotyping. Better Genetics.

GET THE NEWSLETTER CONTACT US LOGIN SIGNUP It's free

APPS HOW IT WORKS HELP CENTER ABOUT FDNA PUBLICATIONS

# CLINIC

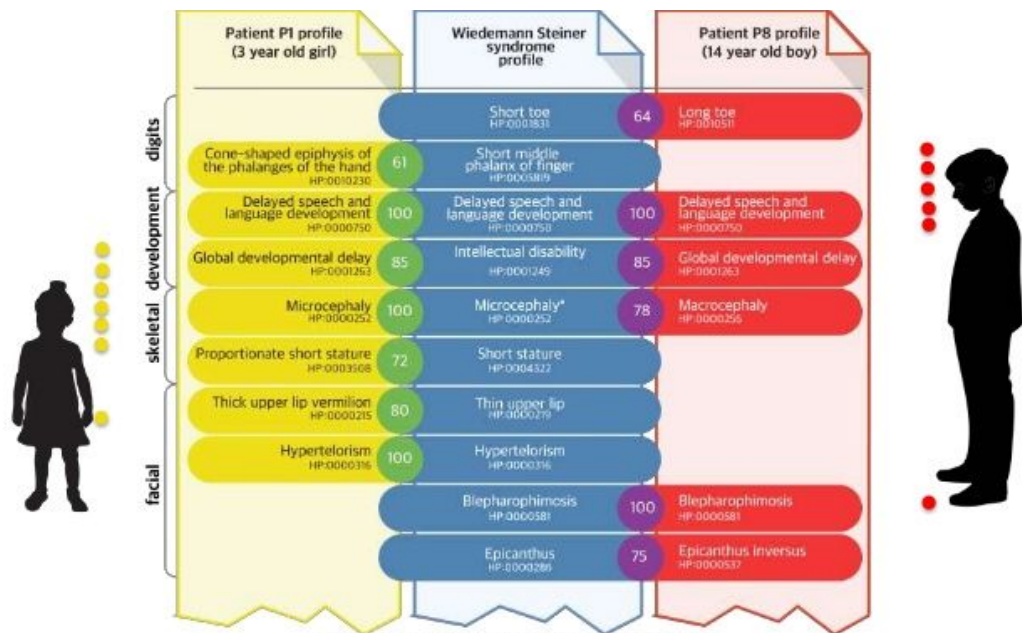
Enhanced Patient Evaluation with Deep Phenotyping

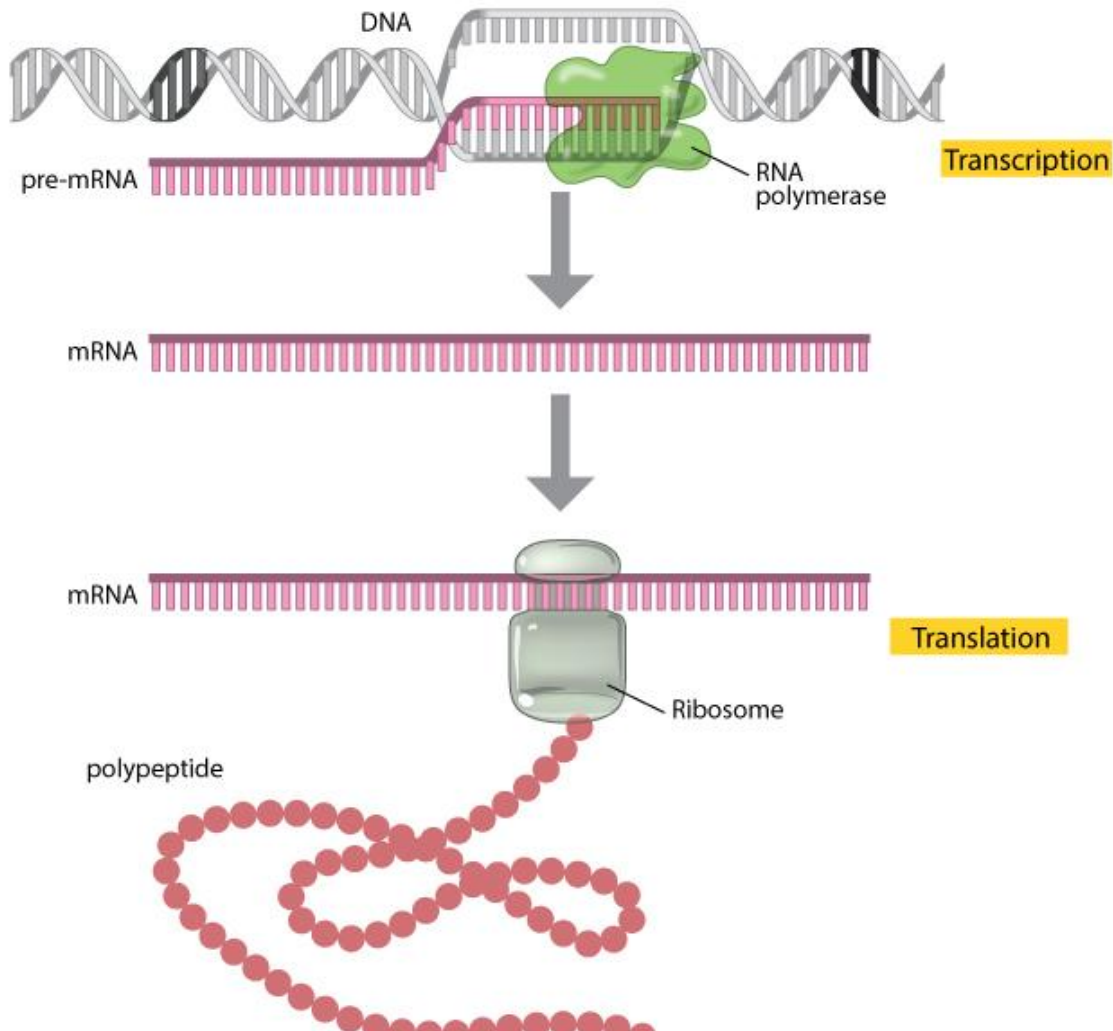
LEARN MORE >

ISO 27001 HIPAA DATA PRIVACY COMPLIANT

# Diagnostyka chorób rzadkich

## Diagnostyka na podstawie fenotypu





Odczytywanie informacji  
genetycznej w komórce

(Clancy, 2008)



# Ludzki genom

3 mld zasad ATGC

Ile możliwych podmian?

Ile bitów na jedną pozycję?

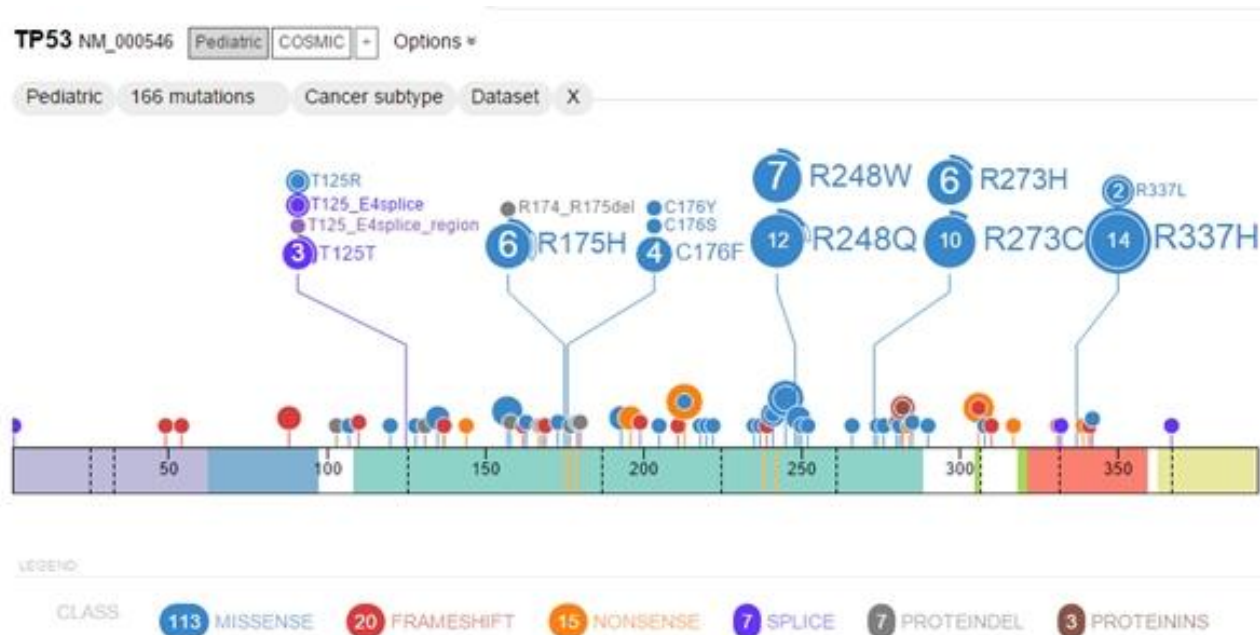
Czy informacja jest rozłożona równomiernie?

Ile genów?

Jaki procent genomu to geny?

# Diagnostyka chorób rzadkich

Zastosowanie data mining: Predykcja patogenności nowych mutacji



# Diagnostyka chorób rzadkich

Omówienie przypadku

Deleted

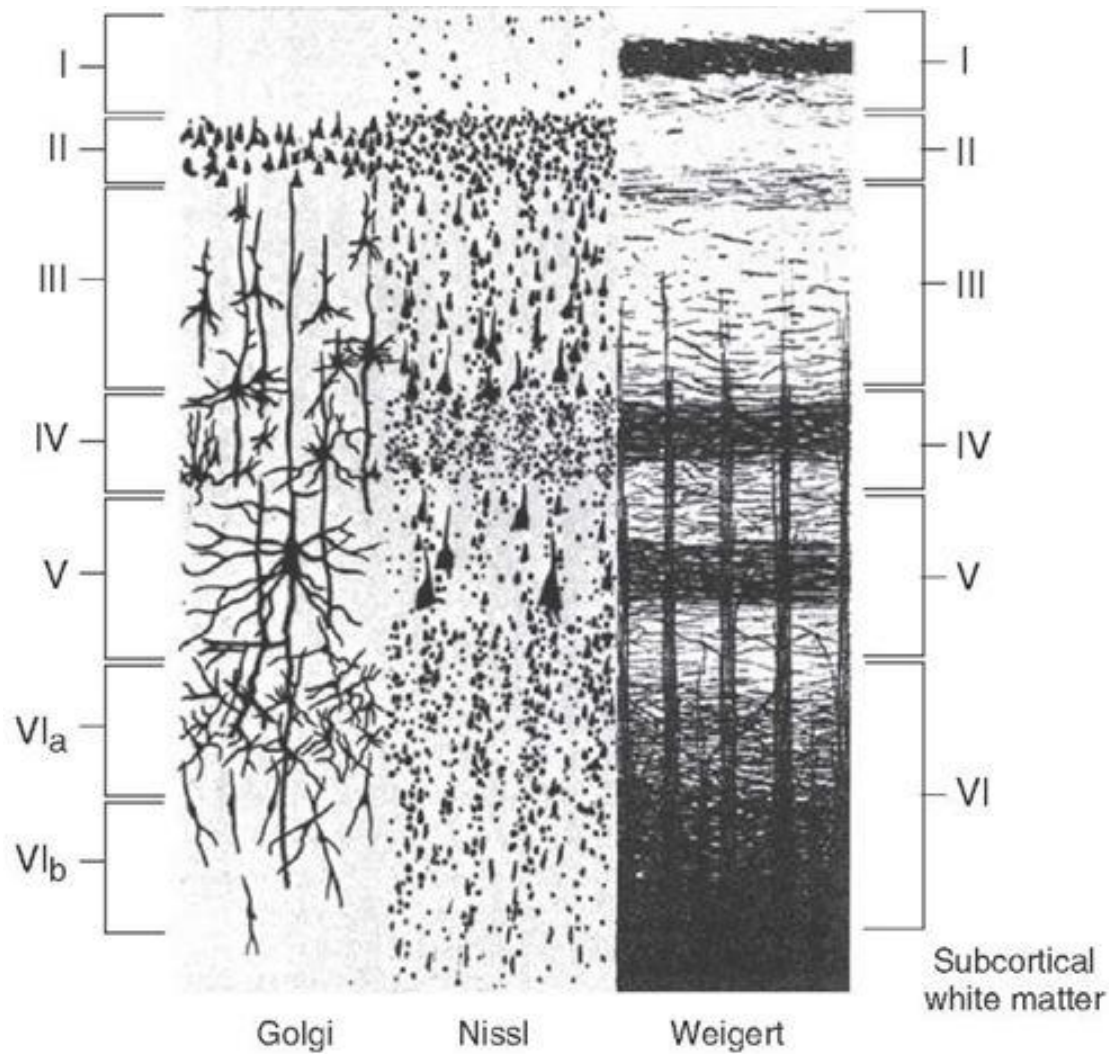
Deleted

Deleted

Deleted

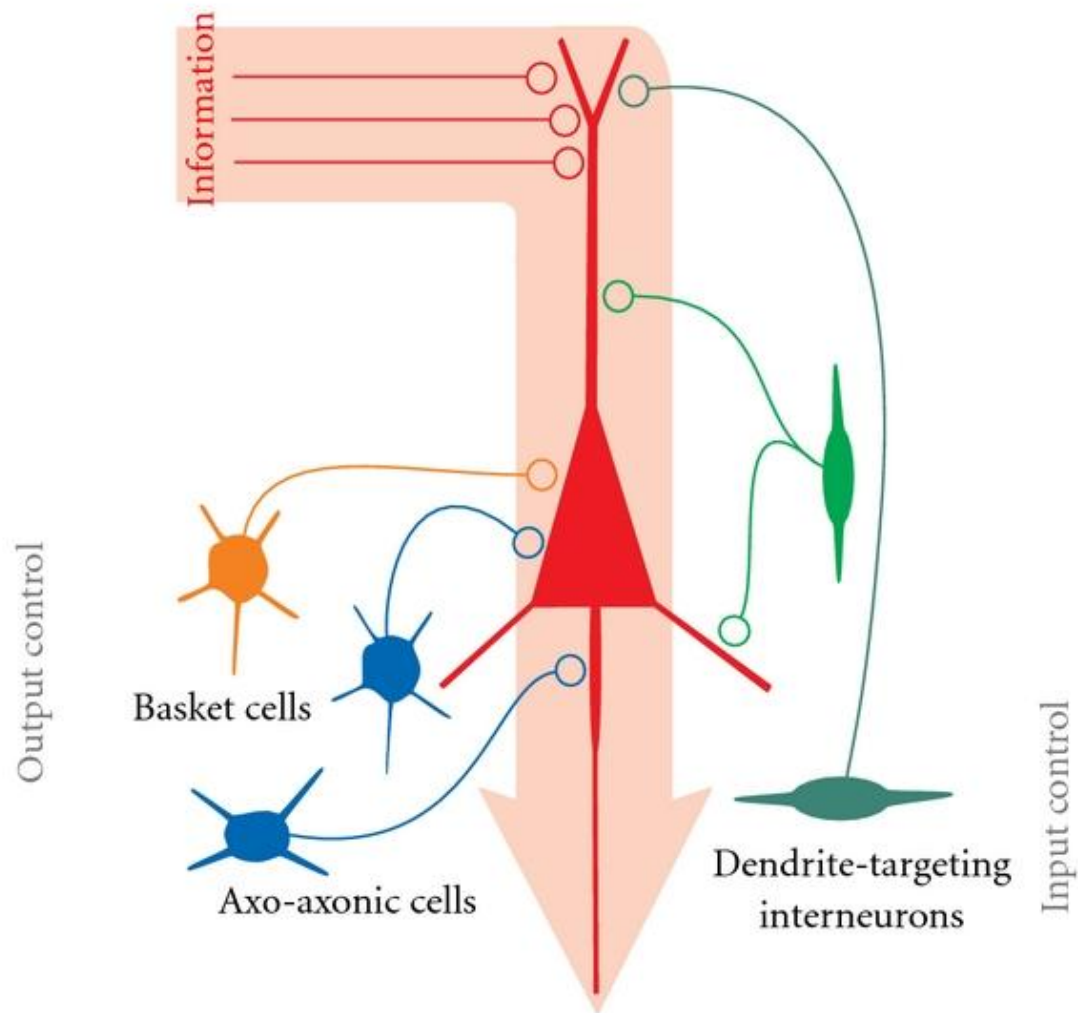
# Klasyfikacja komórek kory mózgowej





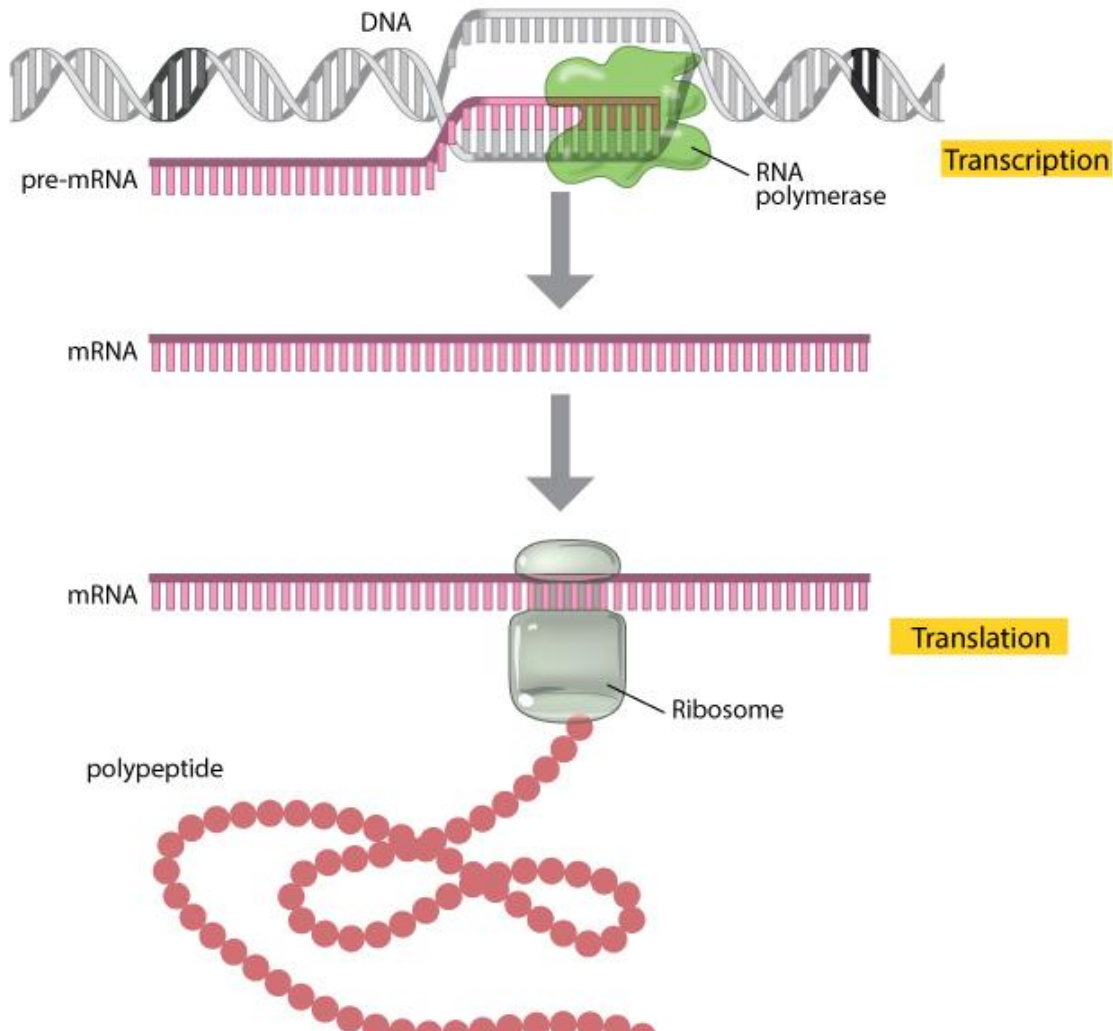
Struktura kory mózgowej

Berne & Levy, Physiology,  
5th Edition, 2003



Uproszczony schemat  
połączeń w korze mózgowej

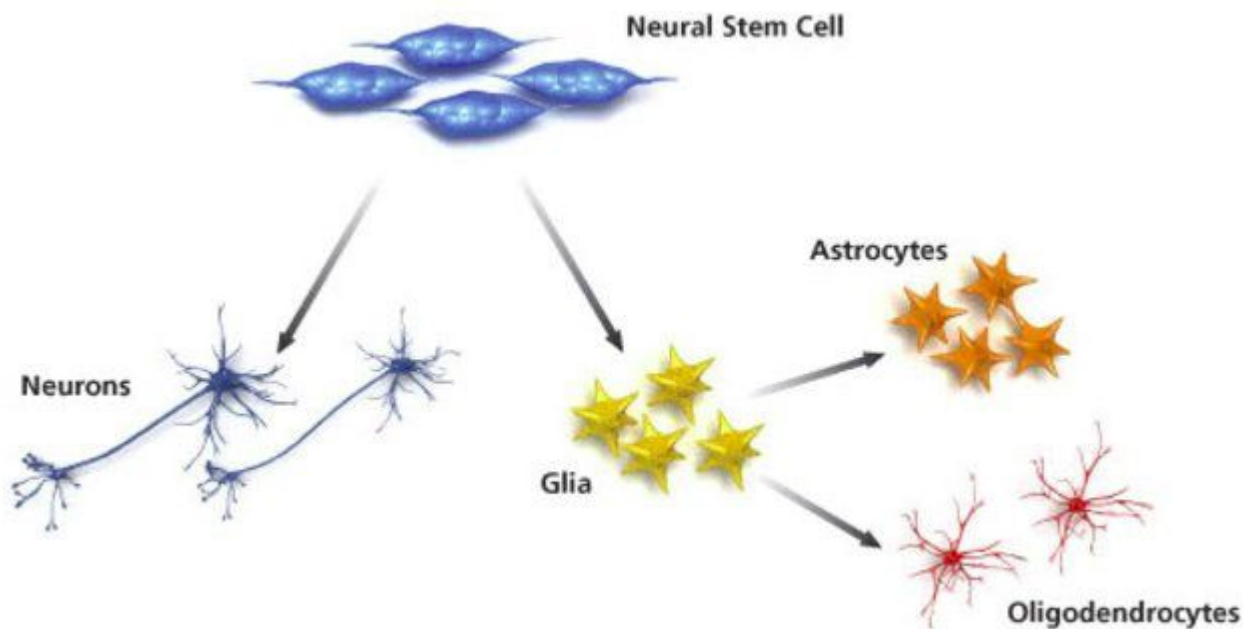
Mendez et al., 2011



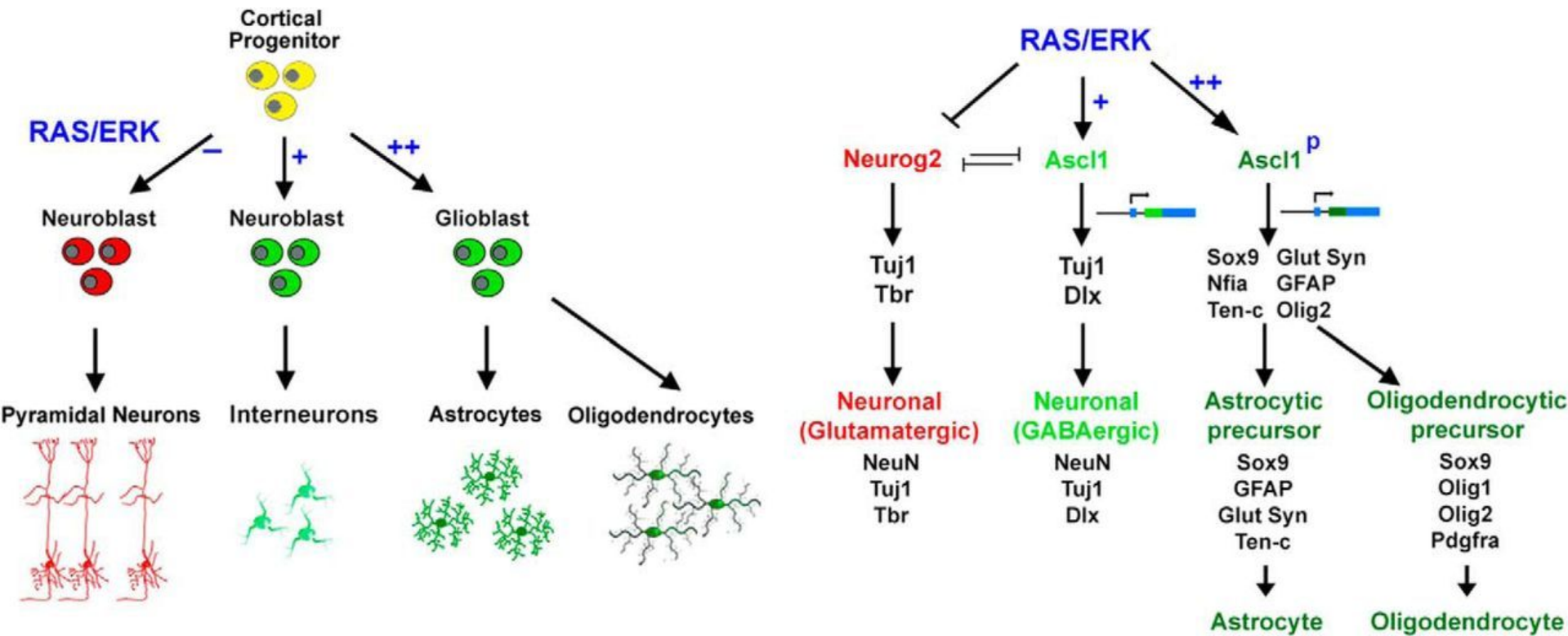
Odczytywanie informacji  
genetycznej w komórce

Clancy, 2008

# Różnicowanie komórek układu nerwowego



# Różnicowanie komórek układu nerwowego



(Li, 2014)

# Zbiór danych

Pomiary ekspresji dla 24057 genów w 1809 komórkach

Zadanie: zidentyfikować klasy komórek na podstawie ekspresji genów i poklasyfikować komórki

Nat Neurosci. 2016 Feb;19(2):335-46. doi: 10.1038/nn.4216. Epub 2016 Jan 4.

## **Adult mouse cortical cell taxonomy revealed by single cell transcriptomics.**

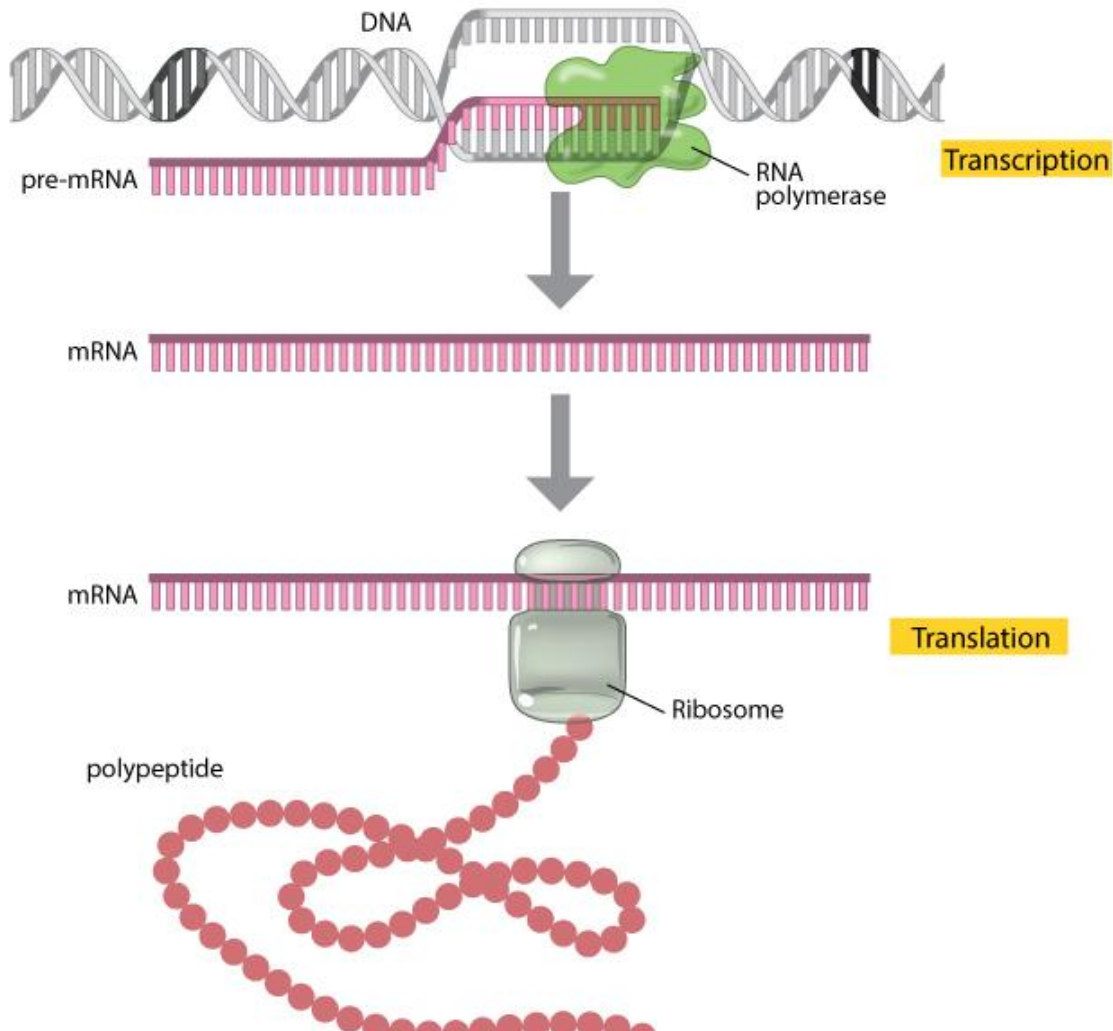
Tasic B<sup>1</sup>, Menon V<sup>1</sup>, Nguyen TN<sup>1</sup>, Kim TK<sup>1</sup>, Jarsky T<sup>1</sup>, Yao Z<sup>1</sup>, Levi B<sup>1</sup>, Gray LT<sup>1</sup>, Sorensen SA<sup>1</sup>, Dolbeare T<sup>1</sup>, Bertagnolli D<sup>1</sup>, Goldy J<sup>1</sup>, Shapovalova N<sup>1</sup>, Parry S<sup>1</sup>, Lee C<sup>1</sup>, Smith K<sup>1</sup>, Bernard A<sup>1</sup>, Madisen L<sup>1</sup>, Sunkin SM<sup>1</sup>, Hawrylycz M<sup>1</sup>, Koch C<sup>1</sup>, Zeng H<sup>1</sup>.

### **Author information**

#### **Abstract**

Nervous systems are composed of various cell types, but the extent of cell type diversity is poorly understood. We constructed a cellular taxonomy of one cortical region, primary visual cortex, in adult mice on the basis of single-cell RNA sequencing. We identified 49 transcriptomic cell types, including 23 GABAergic, 19 glutamatergic and 7 non-neuronal types. We also analyzed cell type-specific mRNA processing and characterized genetic access to these transcriptomic types by many transgenic Cre lines. Finally, we found that some of our transcriptomic cell types displayed specific and differential electrophysiological and axon projection properties, thereby confirming that the single-cell transcriptomic signatures can be associated with specific cellular properties.





Odczytywanie informacji  
genetycznej w komórce

Clancy, 2008

# Zadanie

Poklasyfikować typy komórek na podstawie profili ekspresji genów



# Klasyfikacja nienadzorowana

# Feature selection (zalety)

- Uproszczenie modelu, łatwiejsza interpretacja
- Krótszy czas obliczeń
- Klątwa wielowymiarowości
- Ograniczenie overfittowania

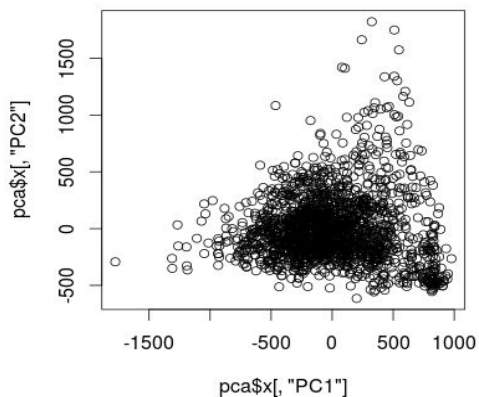
# Feature selection (filtering i wrapping)

- Filtrowanie polega na wybraniu zmiennych na podstawie metryki (np. korelacja ze zmienną zależną)
- Wrapping polega na ocenie modelu zbudowanego na podstawie wybranych zmiennych i iteracyjnym dobieraniu, odejmowaniu zmiennych
- Do wrappingu potrzebujemy funkcję celu i algorytm dokładania oraz usuwania zmiennych

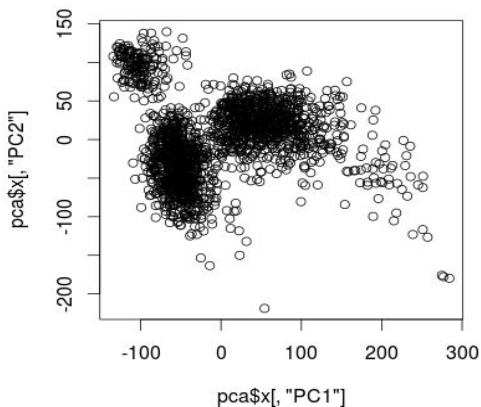
# Feature selection

Wybór zmiennych do modelu ma znaczenie dla klasyfikacji

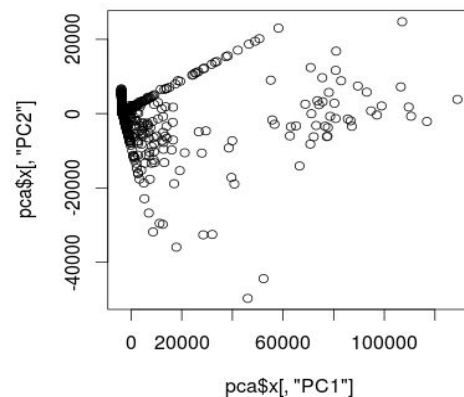
300 zmiennych  
wybranych  
losowo



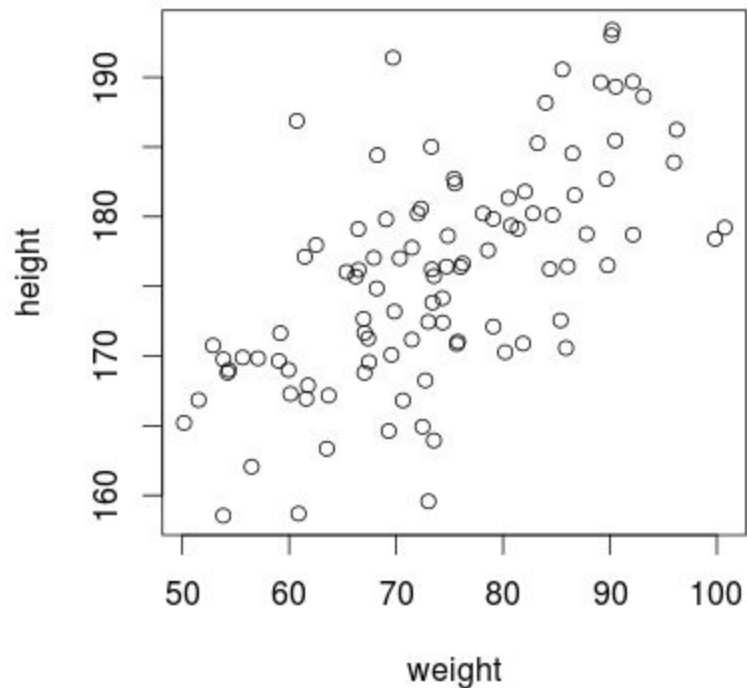
300 zmiennych  
wybranych na podstawie  
rozkładu



300 zmiennych  
o największym  
maksimum



# Redukcja wymiarowości (PCA)



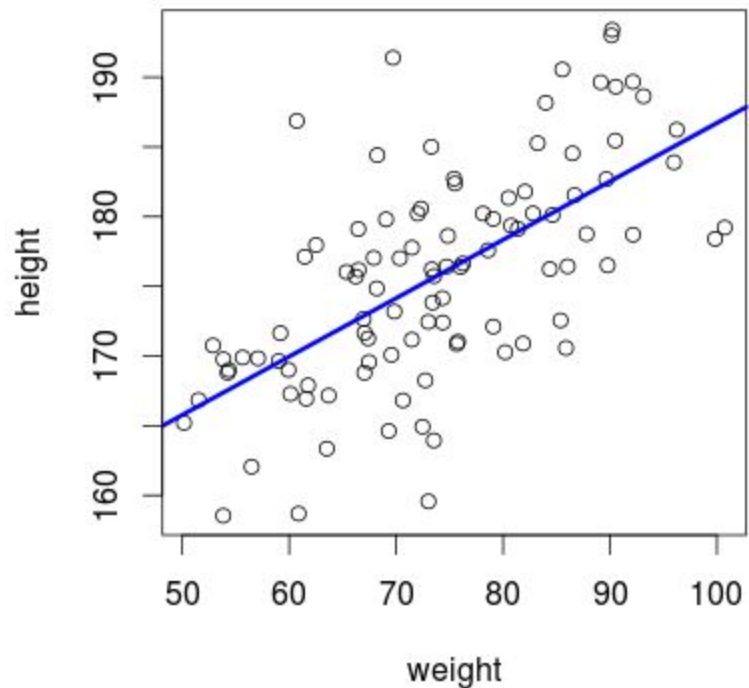
```
height <- rnorm(100) * 9 + 175
```

```
weight <- height - 100 + rnorm(100) * 10
```

```
x = cbind(weight, height)
```

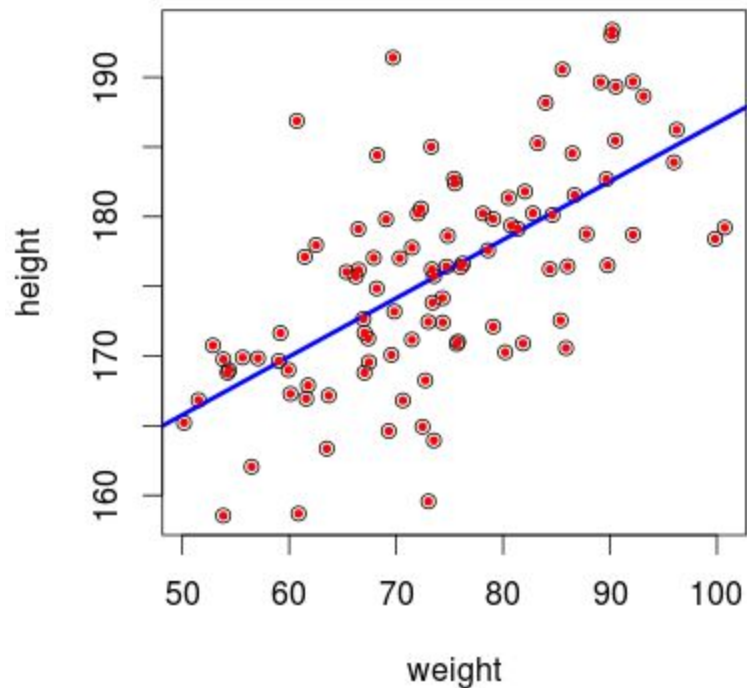
```
plot(x)
```

# Redukcja wymiarowości (PCA)



```
abline(lm(height ~ weight, data =  
data.frame(x)), col = "blue", lwd = 3)
```

# Redukcja wymiarowości (PCA)



```
pca <- prcomp(x)
```

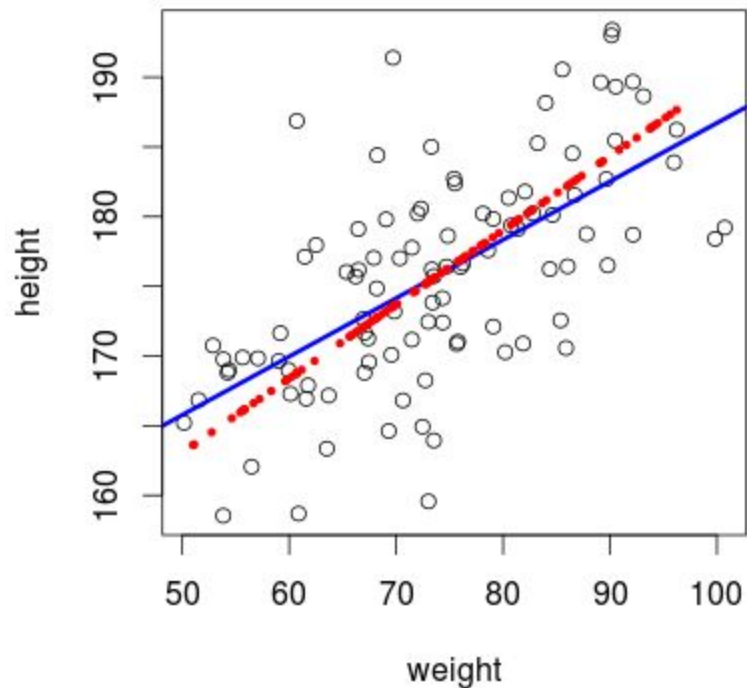
```
pca$rotation
```

```
transformed <- pca$x %*% t(pca$rotation)
```

```
transformed <- scale (transformed, center  
= -pca$center, scale = FALSE)
```

```
points (transformed, col = 2, pch = 19,  
cex = 0.5)
```

# Redukcja wymiarowości (PCA)



```
plot (x)

abline(lm(height ~ weight, data = data.frame(x)), col =
"blue", lwd = 3)

transformed <- -pca$x[,1] %*% t(-pca$rotation [,1])

transformed[,1] = transformed[,1] + pca$center[1]

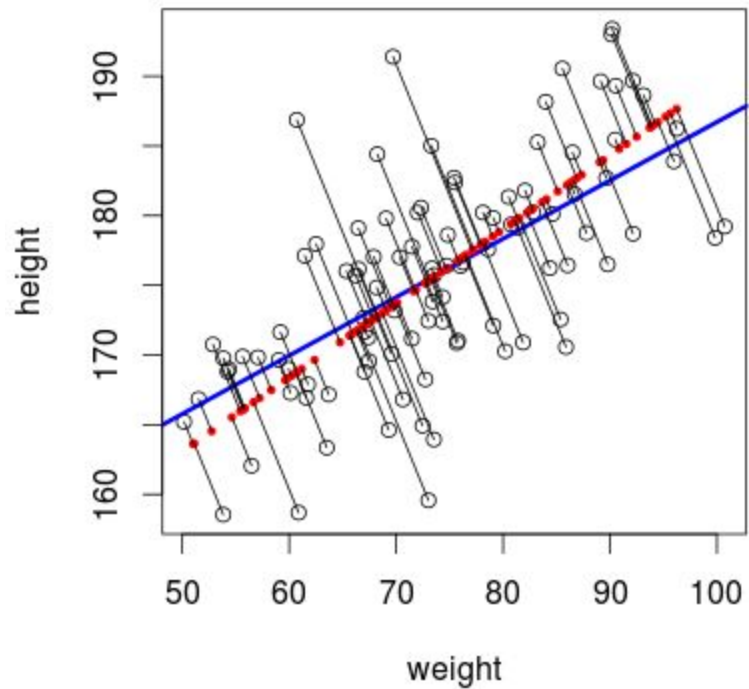
transformed[,2] = transformed[,2] + pca$center[2]

points (transformed, col = 2, pch = 19, cex = 0.5)

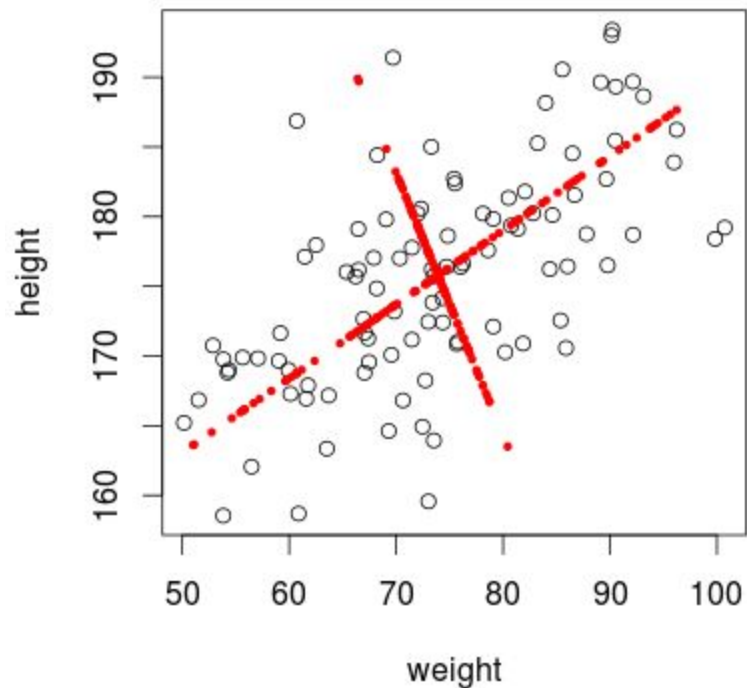
segments (x [,1],x [,2], transformed [,1], transformed [,2])
```



# Redukcja wymiarowości (PCA)



# Redukcja wymiarowości (PCA)



```
> summary(pca)
```

Importance of components:

	PC1	PC2
Standard deviation	13.1044	5.4557
Proportion of Variance	0.8523	0.1477
Cumulative Proportion	0.8523	1.0000

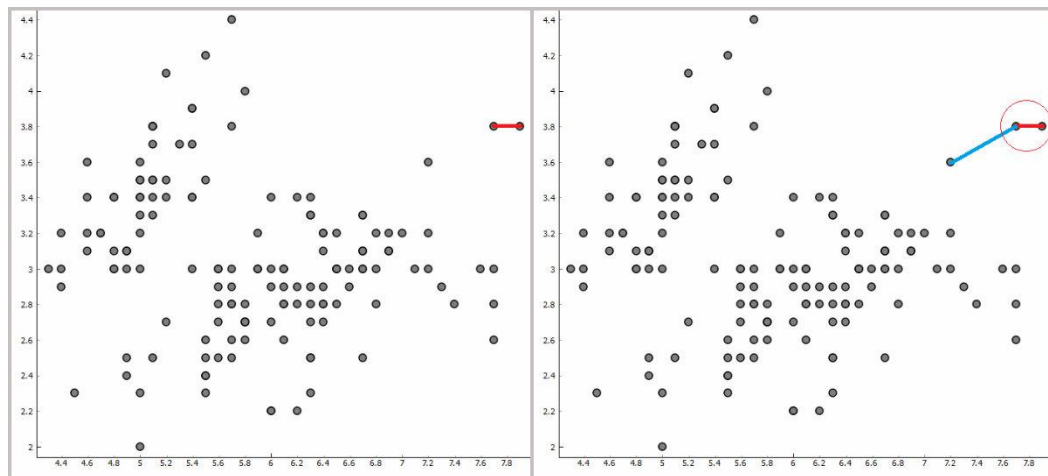
# Grupowanie, Klasteryzacja

Connectivity models - hierarchical clustering

Centroid models - k-means

Inne: biclustering, clique, self-organizing maps

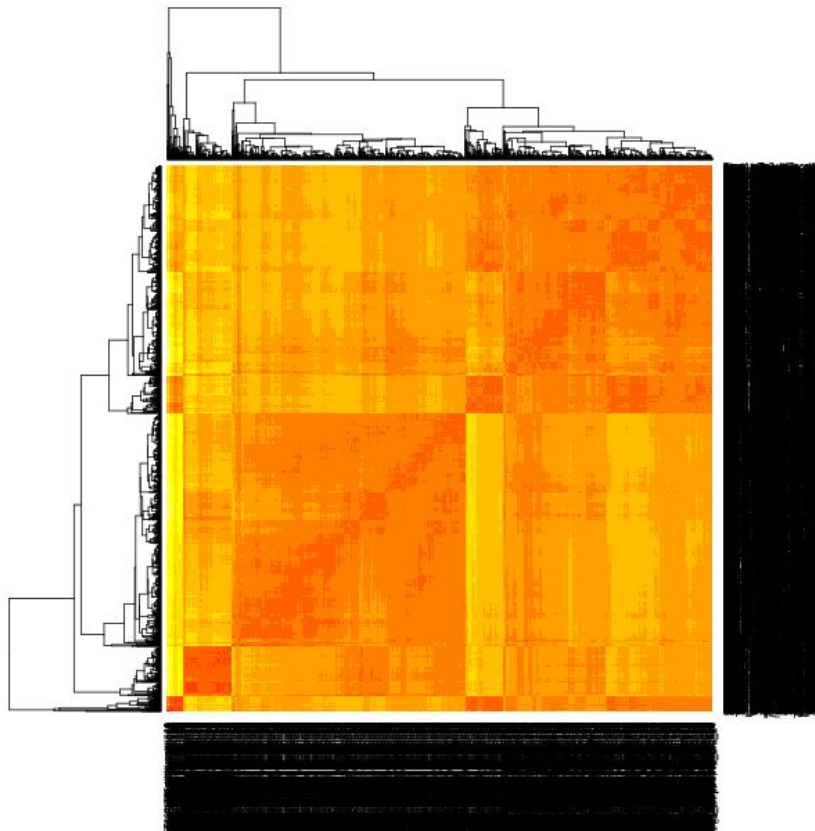
# Klasteryzacja Hirarchiczna - Odległość



Ajda Pretnar, 2015 (Fair share use)

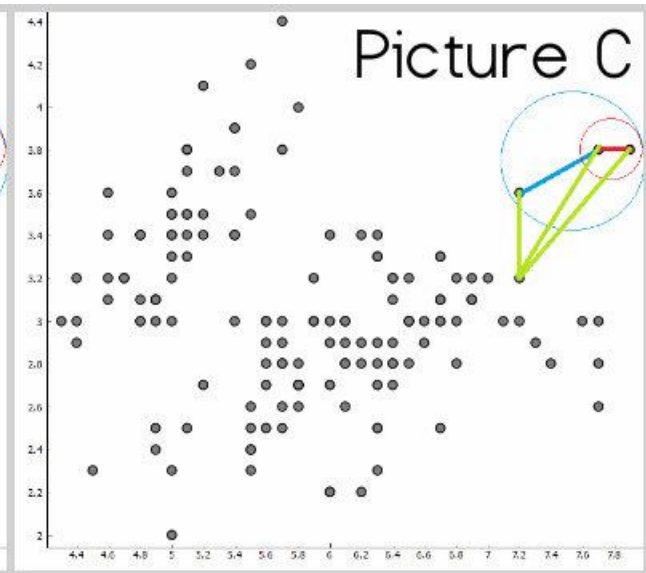
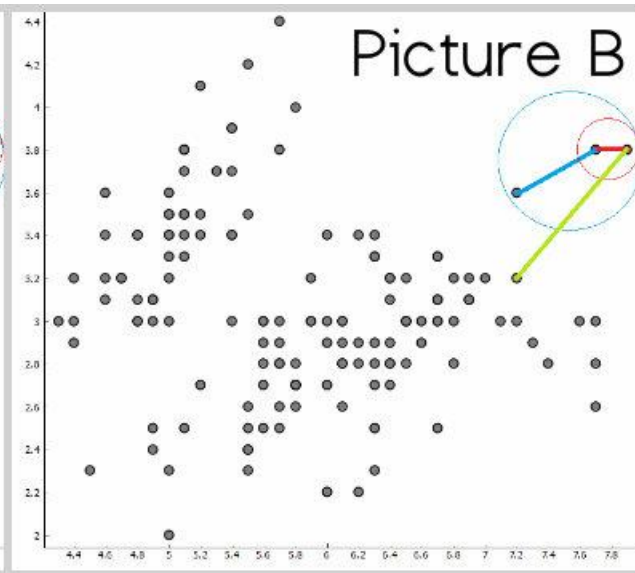
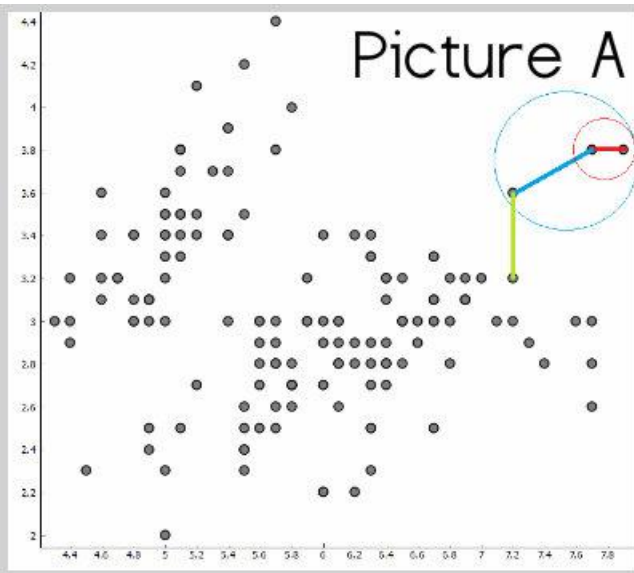
# Hierarchiczna klasteryzacja

Przykład



# Klasteryzacja Hierarchiczna

Sposoby składania drzewa (Single linkage, Complete, Average)



# Klasteryzacja Hierarchiczna

## Single vs complete

