

# Dane wielowymiarowe i ich analiza w badaniach nad roślinami

Paweł Krajewski

Zakład Biometrii i Bioinformatyki  
Instytut Genetyki Roślin PAN w Poznaniu  
[pkra@igr.poznan.pl](mailto:pkra@igr.poznan.pl)



Wykłady Otwarte, UAM, Poznań, 18.04.2018 r.

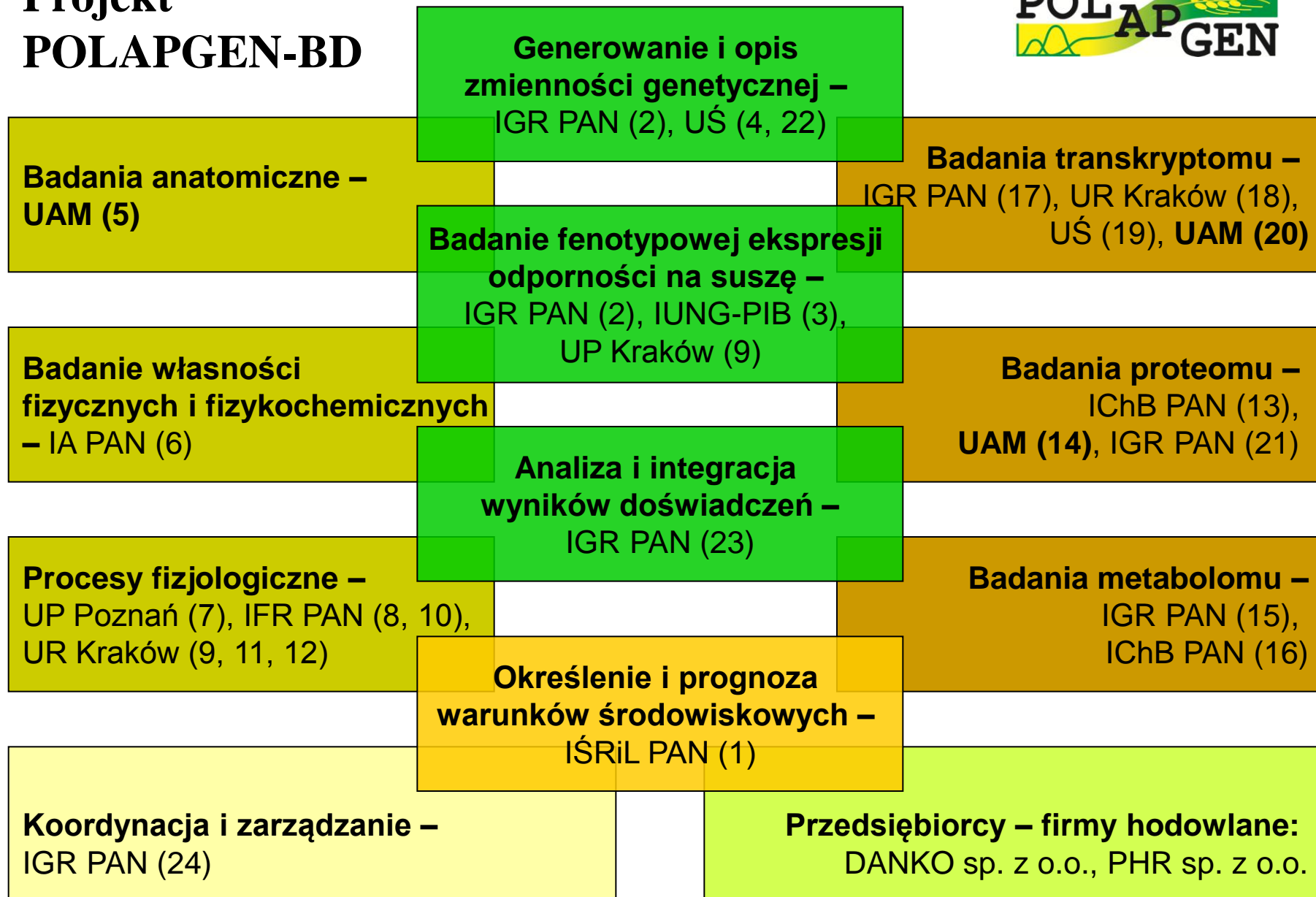
# Tematyka badań Zespołu Biometrii i Bioinformatyki IGR PAN

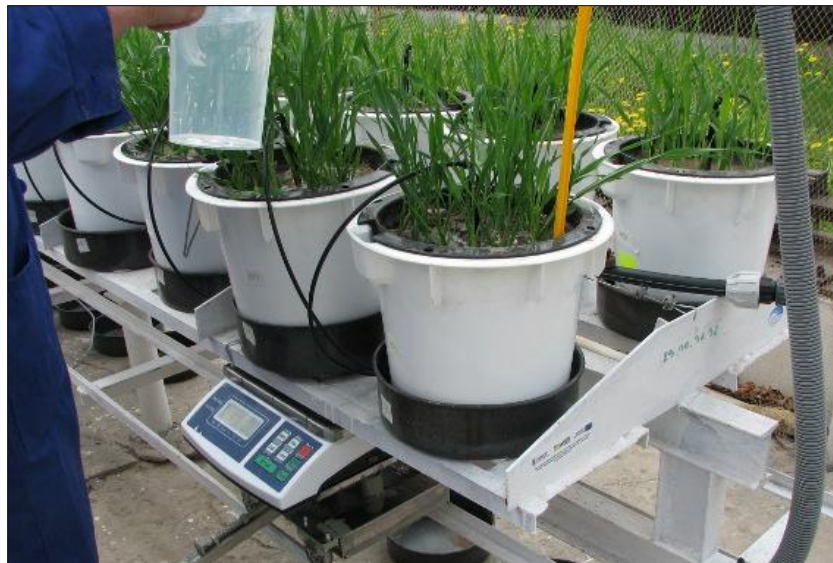
1. Metody analizy danych z sekwencjonowania wysokoprzepustowego DNA (NGS)
2. Metody adnotacji i standaryzacji danych biologicznych
3. Współpraca z biologami – analiza i interpretacja danych doświadczalnych

## **Metody wielowymiarowe:**

- Wykorzystanie na przykładzie projektu POLAPGEN-BD
- Nowe zastosowania w analizie danych NGS

# Projekt POLAPGEN-BD

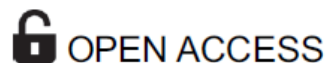




# Doświadczenia polowe z populacjami linii jęczmienia

**Table 1. Agronomic traits observed in the experiments.**

Trait(unit)	Abbreviation	Methods of measurement
Heading stage (days)	HS	Number of days from sowing to the beginning of heading—approximately 50% of spikes in a plot were in the growth stage 51 according to the BBCH scale
Length of main stem (cm)	LSt	Measured from soil surface to the tip of the spike (without awns)
Length of main spike (cm)	LSp	Measured from the base of spike to the tip of the terminal spikelet (without awns)
Number of grains per main spike	NGS	Counted on the basis of 20 randomly selected main spikes from each plot
Grain weight per main spike (g)	GWS	Average weight of hand-threshed grain from 20 randomly selected main spikes from each plot
1000-grain weight (g)	TGW	1000 × weight of one grain averaged for weight of grains from 20 main spikes
Grain yield (g)	GY	Weight of grain com



OPEN ACCESS

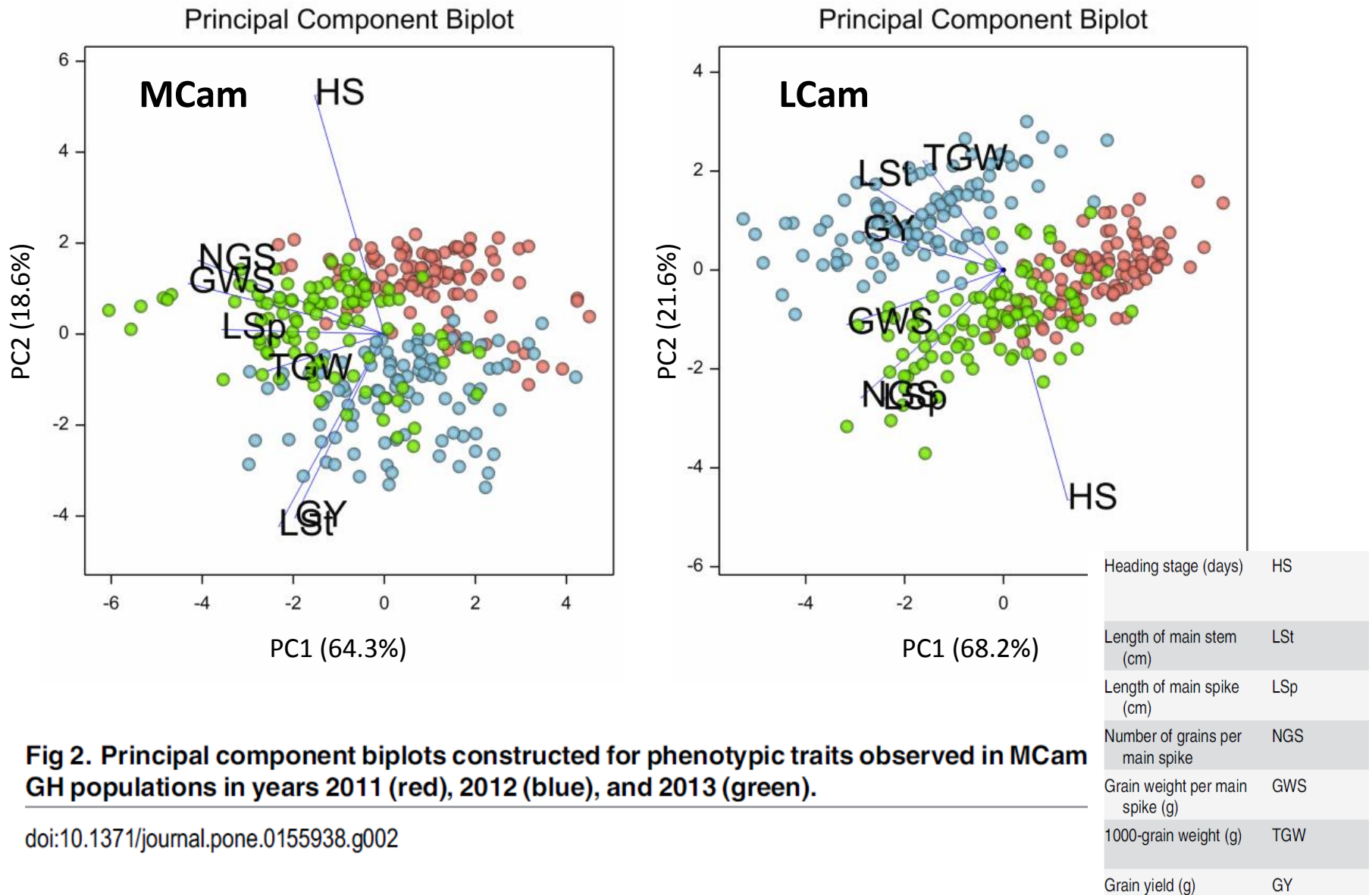
doi:10.1371/journal.pone.0155938.t001

**Citation:** Mikołajczak K, Ogradowicz P, Gudyś K, Krystkowiak K, Sawikowska A, Frohberg W, et al. (2016) Quantitative Trait Loci for Yield and Yield-Related Traits in Spring Barley Populations Derived from Crosses between European and Syrian Cultivars. PLoS ONE 11(5): e0155938. doi:10.1371/journal.pone.0155938

**T. Adamski, M. Surma, A. Kuczyńska**

**IGR PAN**

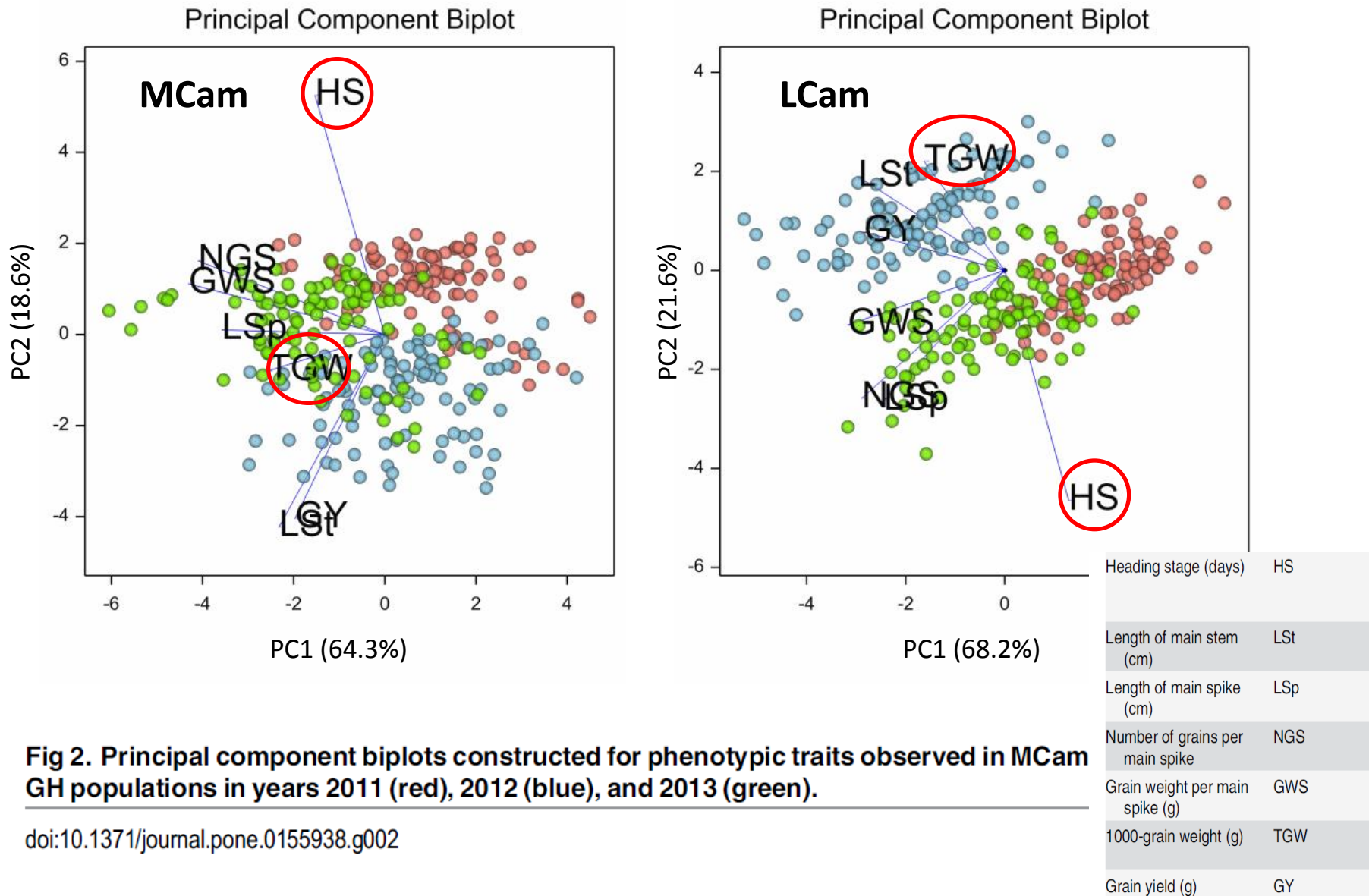
# Doświadczenia polowe z populacjami linii jęczmienia – 3 lata



**Fig 2. Principal component biplots constructed for phenotypic traits observed in MCam GH populations in years 2011 (red), 2012 (blue), and 2013 (green).**

doi:10.1371/journal.pone.0155938.g002

# Doświadczenia polowe z populacjami linii jęczmienia – 3 lata



**Fig 2. Principal component biplots constructed for phenotypic traits observed in MCam GH populations in years 2011 (red), 2012 (blue), and 2013 (green).**

doi:10.1371/journal.pone.0155938.g002

# Doświadczenia szklarniowe z populacjami linii jęczmienia

Table 1. Agronomic traits observed in the greenhouse experiment with the description

Trait (unit)	Phenotype description	Abbrev.
Heading stage (days)	Number of days from sowing to emergence of inflorescence (spike) from the flag leaf (51 BBCH), calculated as number of days from sowing to emergence of spikes for at least 50% of plants	HS
1000-grain weight (g)	Average weight of 1000 grains, calculated as average of 1000 * average weight of one grain for 20 spikes in a pot	TGW
Grain weight per plant (g)	Average weight of grain collected from one plant, calculated as average of measurements of grain weight for 10 plants	GWP
Length of main stem (cm)	Average of measurements of length of stem from ground level to the end of spike (without awns) for 10 main stems in a pot	LSt
Number of productive tillers per plant	Average number of tillers with spikes containing grain in a plant, calculated as average of observed numbers of tillers with spike containing some grain for 10 mature plants in a pot	NPT
Grain weight per main spike (g)	Average weight of grain collected from one spike of the main stem, calculated as average of measurements of grain weight from main spike for 10 main spikes	GWSm
Number of grains per main spike	Average number of grains collected from one spike of main stem, calculated as average of observed numbers of grains in a spike for 10 main spikes in a pot	NGSm
Number of spikelets per main spike	Average number of spikelets in spike of main stem, calculated as average of observed numbers of spikelets in a spike for 10 main spikes in a pot	NSSm
Length of main spike (cm)	Average length of spike from main stem, calculated as average of measurements of length of spike from the base to the top (without awns) of a mature spike for 10 main spikes in a pot	LSp
Grain weight per lateral spike (g)	Average weight of grain collected from one spike of lateral stem, calculated as average of measurements of grain weight from a lateral spike for 10 lateral spikes in a pot	LWSl

J Appl Genetics (2017) 58:49–65  
DOI 10.1007/s13353-016-0363-4

PLANT GENETICS • ORIGINAL PAPER

## QTLs for earliness and yield-forming traits in the Lubuski × CamB barley RIL population under various water regimes

Piotr Ogródowicz<sup>1</sup> • Tadeusz Adamski<sup>1</sup> • Krzysztof Mikołajczak<sup>1</sup> • Anetta Kuczyńska<sup>1</sup> • Maria Surma<sup>1</sup> • Paweł Krajewski<sup>1</sup> • Aneta Sawikowska<sup>1</sup> • Andrzej G. Górny<sup>1</sup> • Kornelia Gudyś<sup>2</sup> • Iwona Szarejko<sup>2</sup> • Justyna Guzy-Wróbelska<sup>2</sup> • Karolina Krystkowiak<sup>1</sup>



# Doświadczenia szklarniowe z populacjami linii jęczmienia – 3 poziomy czynnika „susza”

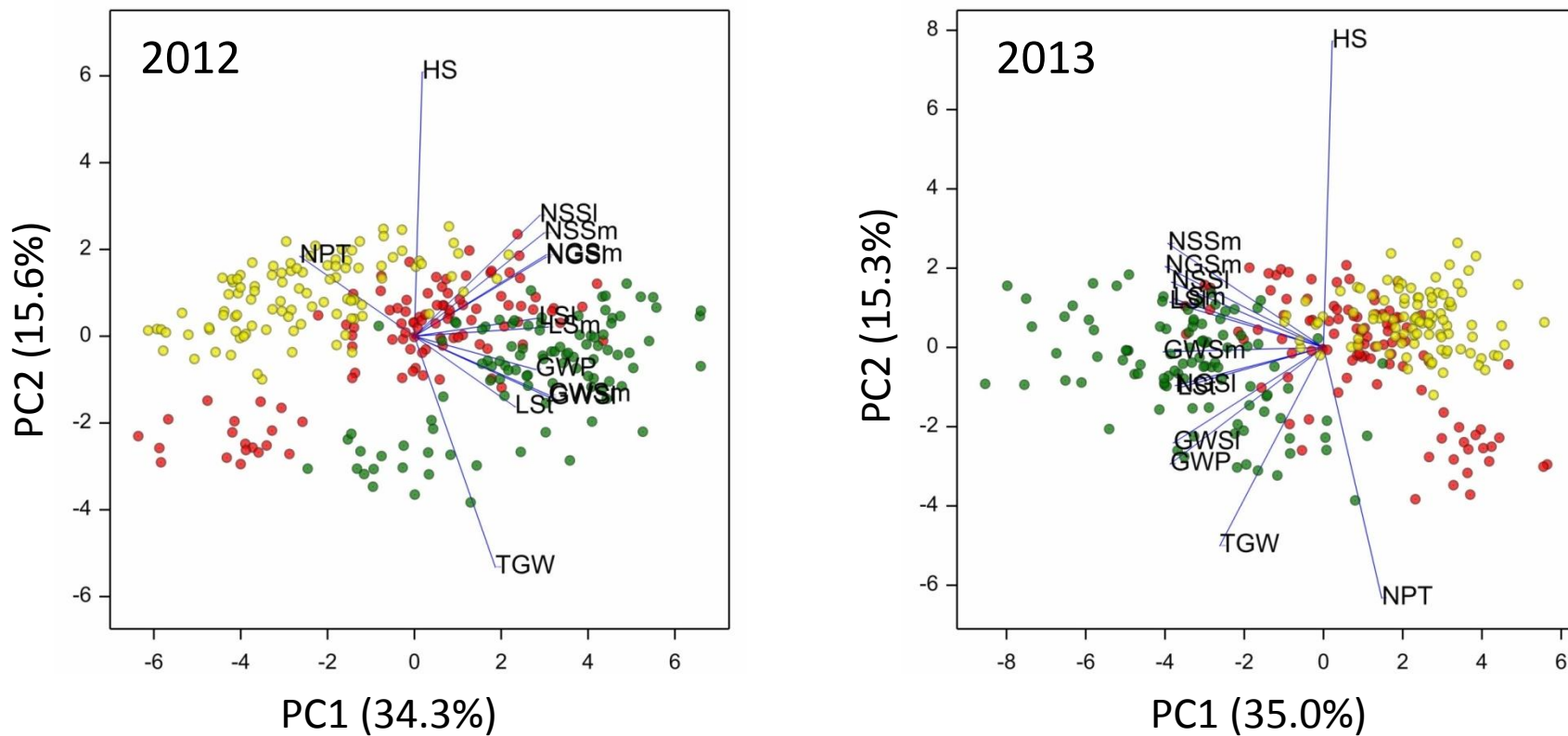


Figure 2. **Principal component biplots**, with dots corresponding to LCam recombinant inbred lines observed in drought DI (red), drought DII (yellow) and in control conditions (green), and vectors corresponding to observed traits, made for data obtained in 2012 and 2013.

# Doświadczenia szklarniowe z populacjami linii jęczmienia – 3 poziomy czynnika „susza”

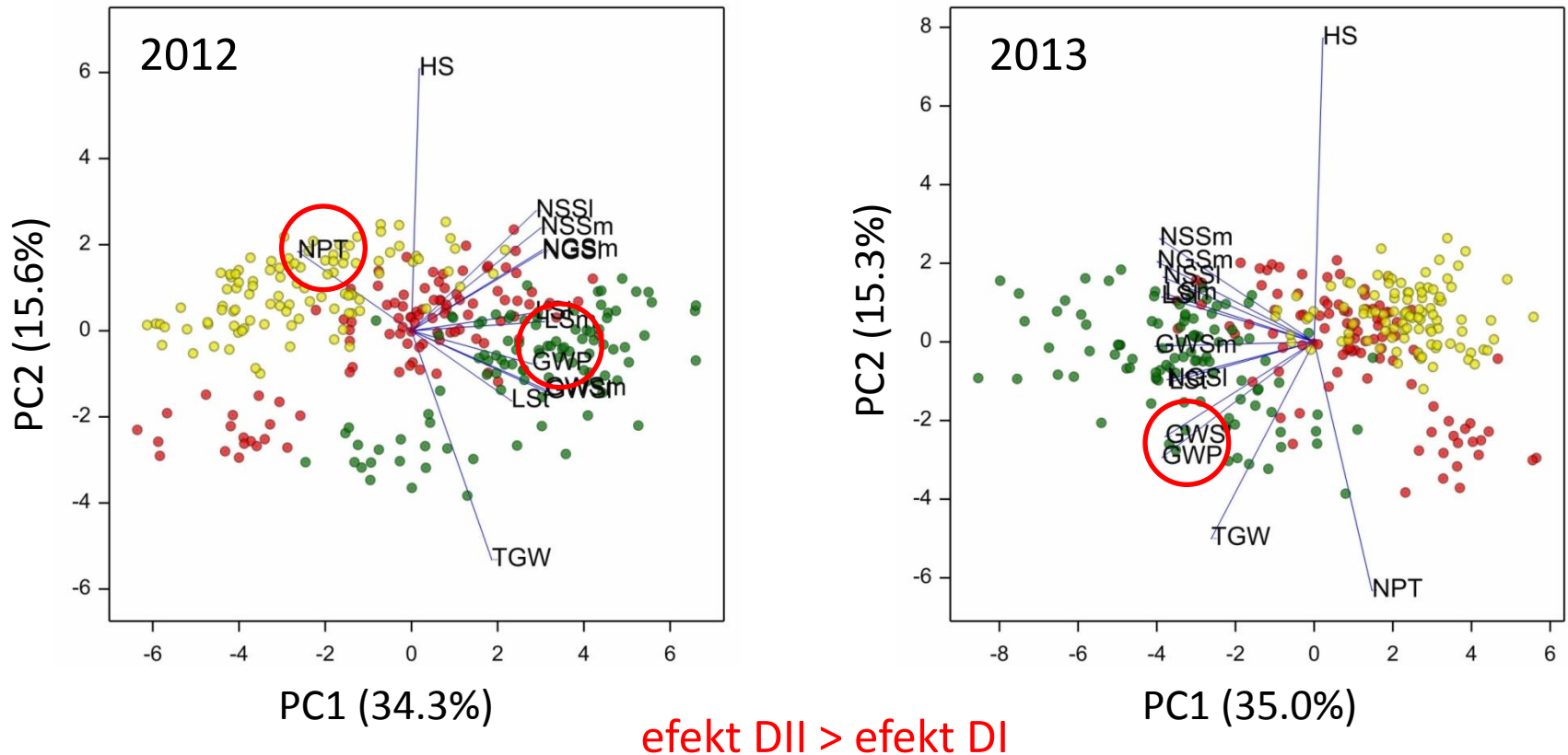


Figure 2. **Principal component biplots**, with dots corresponding to LCam recombinant inbred lines observed in drought DI (red), drought DII (yellow) and in control conditions (green), and vectors corresponding to observed traits, made for data obtained in 2012 and 2013.

# Badanie zawartości metabolitów w liściach (metody chromatograficzne)

Metabolite number	Metabolite name
1	3-Feruloylquinic acid
2	3-p-Coumaroylquinic acid
3	5-Feruloylquinic acid
4	Apigenin 6-C-glucoside-8-C-arabinoside/Apigenin 6-C-arabinoside-8-C-glucoside
5	Apigenin 7-O-arabinosylglucoside
6	Apigenin 7-O-diglucoside
7	Apigenin 7-O-glucosylarabinoside
8	Apigenin 7-O-glucosylrhamnoside
9	Apigenin 7-O-rhamnosylglucoside
10	Blumenol C 2''-O-glucuronylglucoside
11	Blumenol C derivative
12	Blumenol C glucoside
13	Caffeic acid derivative
14	Caffeoyl - hexose
15	Chrysoeriol 7-O-arabinosylglucoside
16	Chrysoeriol 7-O-diglucoside/Isoscoparin
.....	.....
130	UPLC_UV_signal_106
131	UPLC_UV_signal_107
132	UPLC_UV_signal_108
133	UPLC_UV_signal_109
134	UPLC_UV_signal_110
135	UPLC_UV_signal_111

Piotr Kachlicki, IGR PAN

Aneta Sawikowska

the plant journal



The Plant Journal (2017) 89, 898–913

doi: 10.1111/tpj.13430

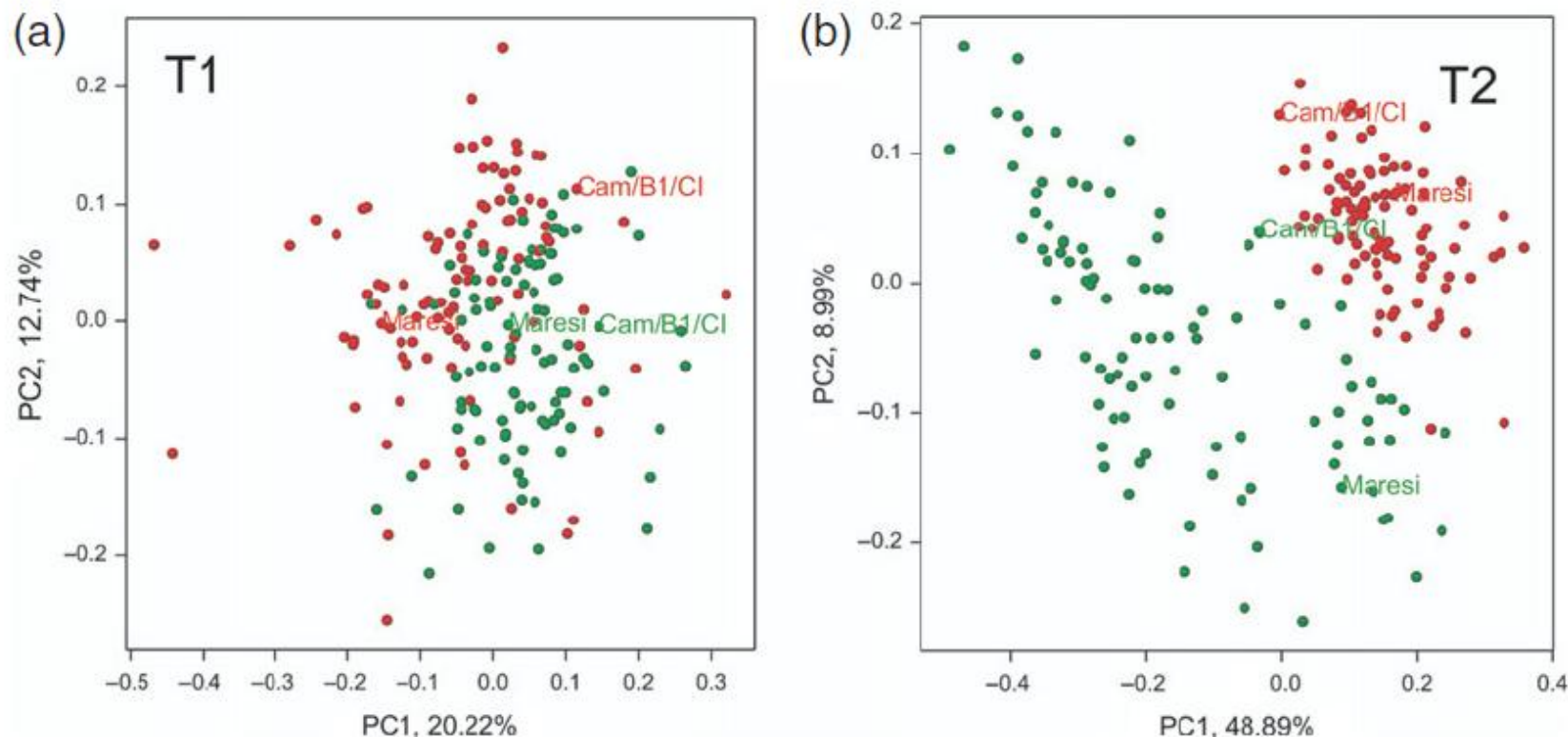
## Drought-related secondary metabolites of barley (*Hordeum vulgare* L.) leaves and their metabolomic quantitative trait loci

Anna Piasecka<sup>1,1</sup>, Aneta Sawikowska<sup>1,1,2</sup>, Anetta Kuczyńska<sup>1</sup>, Piotr Ogródowicz<sup>1</sup>, Krzysztof Mikotajczak<sup>1</sup>, Karolina Krystkowiak<sup>1</sup>, Kornelia Gudys<sup>2,8</sup>, Justyna Guzy-Wróbelska<sup>2</sup>, Paweł Krajewski<sup>1</sup> and Piotr Kachlicki<sup>1,\*</sup>

<sup>1</sup>Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34 60-479, Poznań, Poland, and

<sup>2</sup>Department of Genetics, Faculty of Biology and Environmental Protection, University of Silesia, Jagiellońska 28 40-032, Katowice, Poland

# Badanie zawartości metabolitów w liściach (metody chromatograficzne) – susza/kontrola, 2 terminy badania



**Figure 2.** Principal coordinate visualization of recombinant inbred lines based on observations of 98 metabolites in control (green dots) and drought conditions (red dots) at time points T1 and T2.

**Analiza współrzędnych głównych** (PCO, principal coordinate analysis), macierz podobieństw wyliczona z odległości Euklidesowych pomiędzy liniami

# Badanie zawartości białek w liściach (metoda elektroforezy 2D)

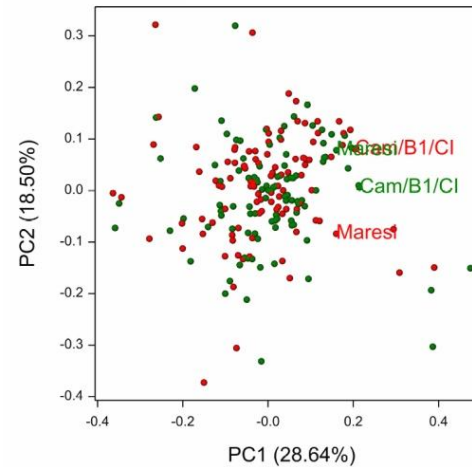
Protein number	Protein name	Number of genotypes in which protein was observed	Function
1	14-3-3-like protein homolog	76	Defence
2	2-Cys peroxiredoxin BAS1	67	Defence
3	2-Cys peroxiredoxin BAS1	64	Defence
4	2-Cys peroxiredoxin BAS1, chloroplastic (Fragment)	73	Defence
5	20 kDa chaperonin, chloroplastic	83	Defence
6	6-phosphogluconate dehydrogenase	70	Carbon metabolism
7	Actin (Fragment)	84	Other
8	Adenylate kinase	91	Carbon metabolism
9	ADP-glucose pyrophosphorylase small subunit, plastid	87	Carbon metabolism
10	Alanyl-tRNA synthetase	68	Gene expression
11	Aminotransferase	69	Nitrogen metabolism
12	Apocytochrome f	71	Photosynthesis
13	Ascorbate peroxidase	95	Defence
14	Ascorbate peroxidase	74	Defence
15	Ascorbate peroxidase	55	Defence
16	ATP synthase subunit alpha, chloroplastic	65	Ion/electron transport
17	ATP synthase subunit alpha, chloroplastic	59	Ion/electron transport
...	...	...	...
253	Unknown protein	40	Unknown
254	Unknown protein	39	Unknown
255	Unknown protein	38	Unknown
256	Unknown protein	32	Unknown
257	Unknown protein	21	Unknown

# Badanie zawartości białek w liściach (metoda elektroforezy 2D)

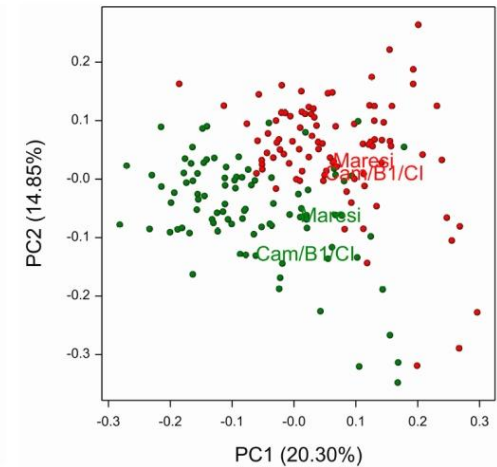
**Analiza współrzędnych głównych** dla grup funkcyjnych białek obserwowanych w liniach RIL i formach rodzicielskich. Zielone – próby kontrolne, czerwone – próby w warunkach niedoboru wody.

Liście

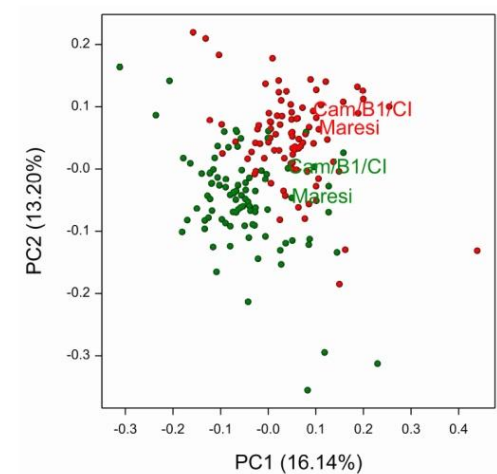
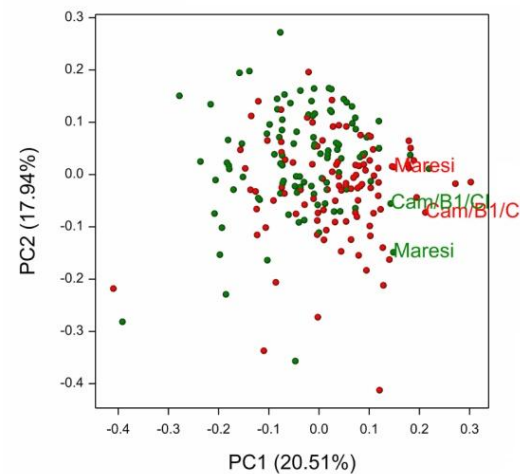
Carbon metabolism



Korzenie

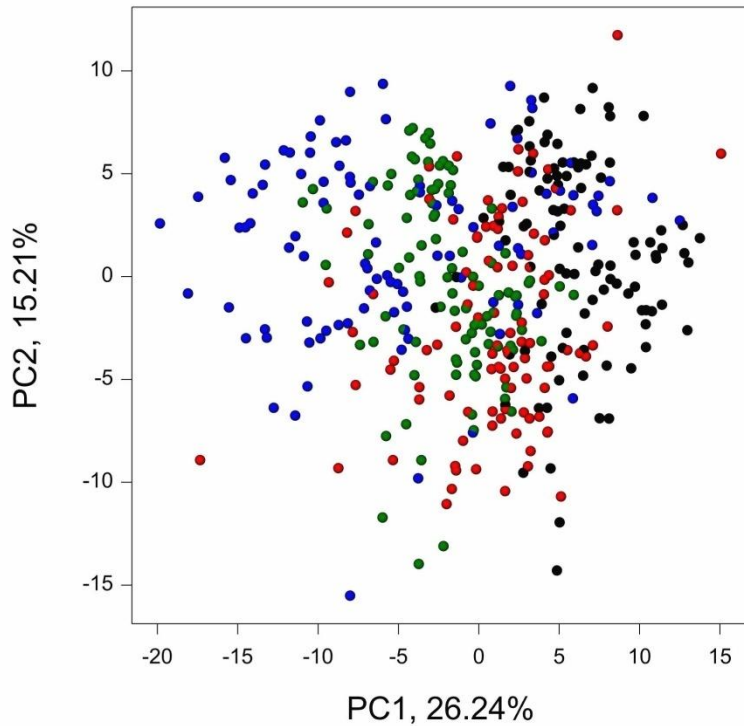


Defence



Dane metabolomiczne – 135 cech, 4 warianty doświadczalne (termin x susza)

## Analiza składowych głównych

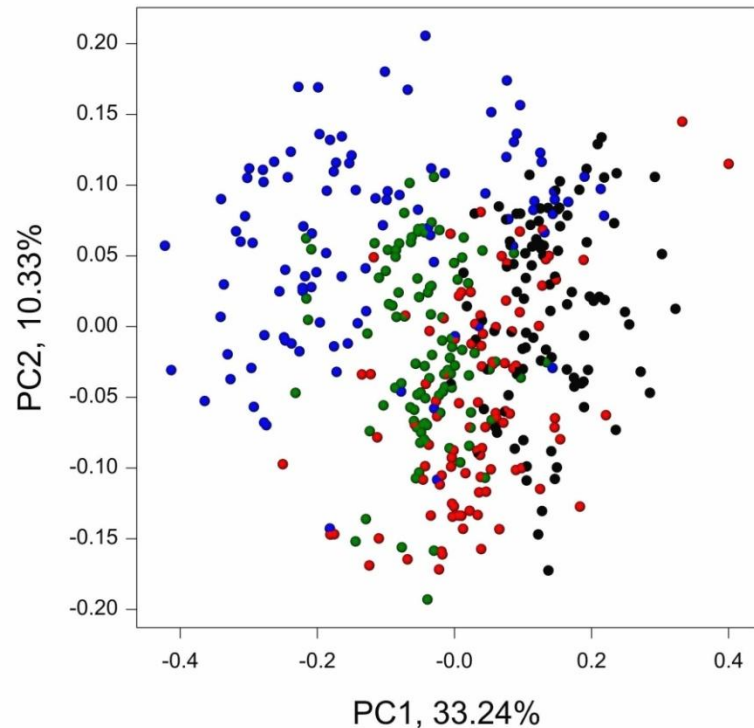


### Principal component analysis

- Na podstawie estymowanej macierzy kowariancji (korelacji) cech

- Możliwy biplot
- Obserwacje brakujące – problem
- Liczba obserwacji > liczby cech

## Analiza współrzędnych głównych



### Principal coordinate analysis (Gower 1966)

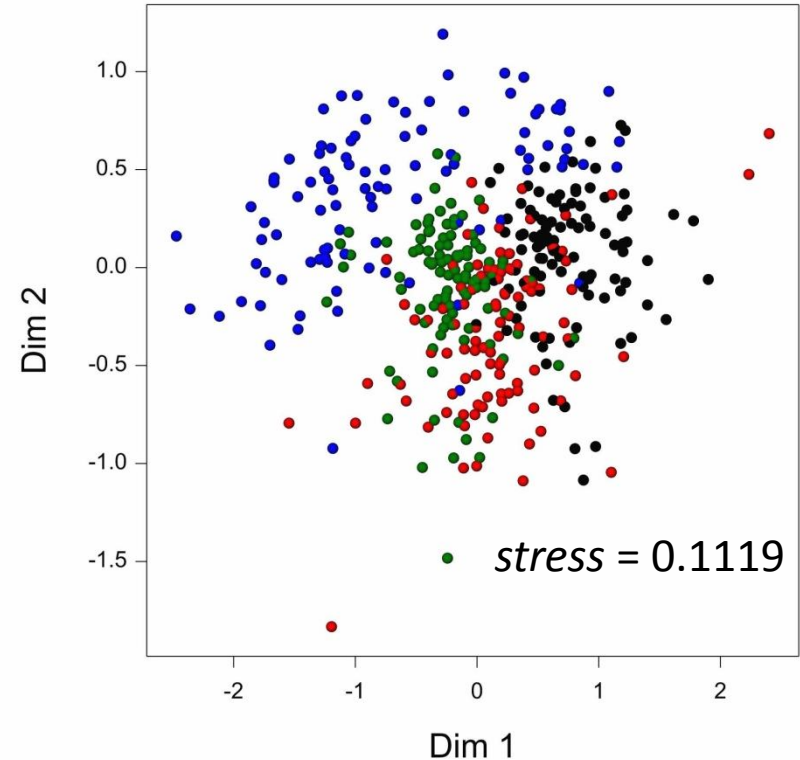
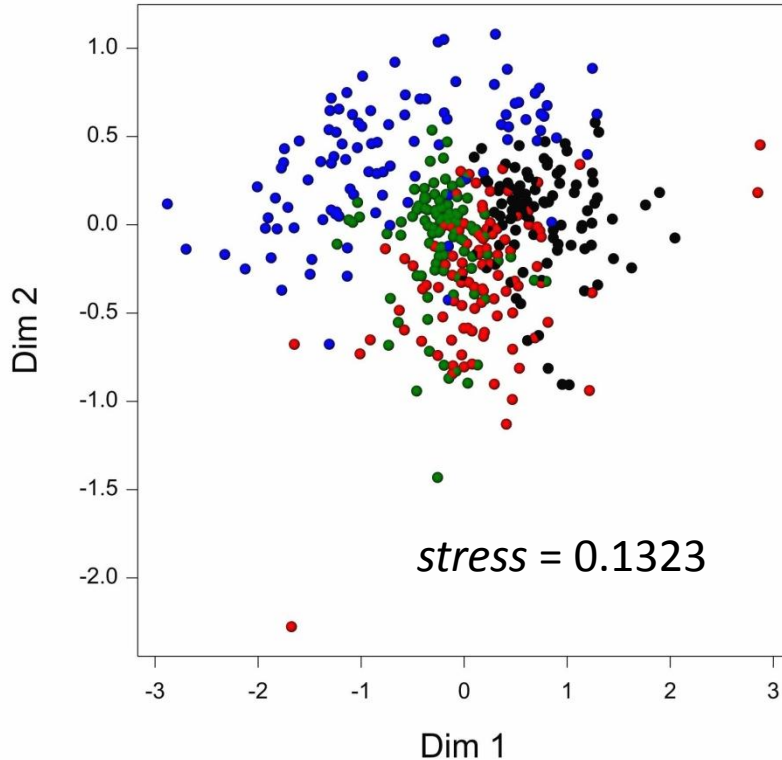
- Na podstawie dowolnej macierzy współcz. podobieństwa pomiędzy obiektami

- Szuka rozmieszczenia, w którym

- Biplot: NIE
- Obserwacje brakujące – dopuszczalne
- Możliwa dla większej liczby cech

$$d_{ij} = a_{ii} + a_{jj} - 2a_{ij}$$

## Skalowanie wielowymiarowe – na podstawie macierzy niepodobieństw



### Liniowe

#### Linear multidimensional scaling

Dopasowanie odległości Euklidesowych w przestrzeni niskowymiarowej do rzeczywistych niepodobieństw za pomocą regresji liniowej

### Niemetryczne

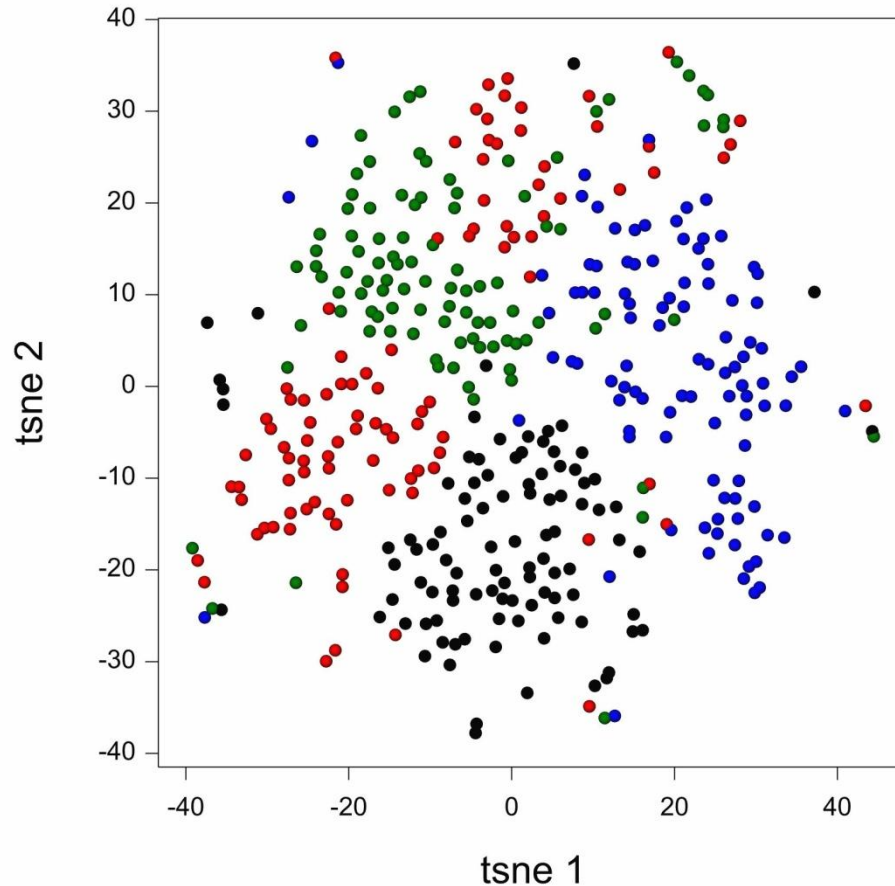
#### Non-metric multidimensional scaling

Dopasowanie odległości Euklidesowych w przestrzeni niskowymiarowej do rzeczywistych niepodobieństw za pomocą regresji monotonicznej

$$stress = \sqrt{\frac{\sum_i \sum_j w_{ij} (d_{ij} - \hat{d}_{ij})^2}{m \times \sum_i \sum_j w_{ij} d_{ij}^2}}$$



## Stochastyczne rozmieszczenie sąsiadów wg rozkładu $t$ -Studenta



Prawdopodobieństwa (podobieństwa) w przestrzeni wysokowymiarowej (z  $N(x_i, \sigma^2)$ )

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2 / 2\sigma^2)},$$

Prawdopodobieństwa (podobieństwa) w przestrzeni niskowymiarowej (z rozkładu  $t$ -Studenta,  $df = 1$ )

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}.$$

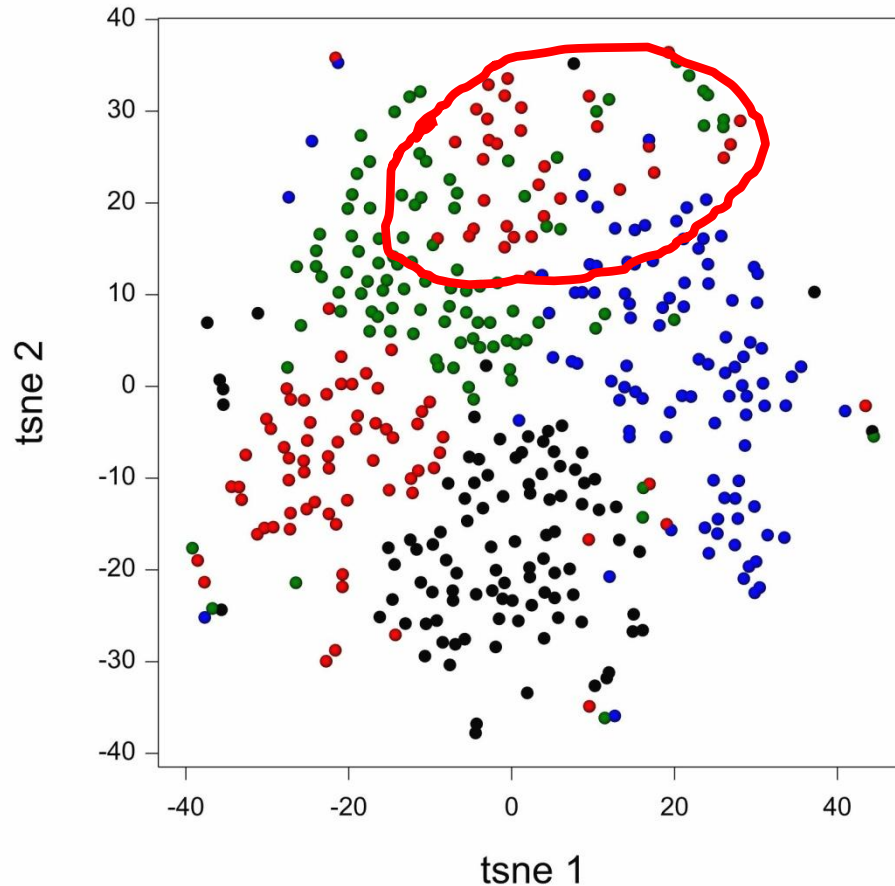
Minimalizacja dywergencji Kullbacka-Leiblera (względnej entropii) pomiędzy  $P$  i  $Q$  (metodą gradientu)

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

### t-Distributed Stochastic Neighbor Embedding

van der Maaten L., Hinton G., Visualizing data using t-SNE. *Machine Learning Res.* 9 (2008): 2579-2605

## Stochastyczne rozmieszczenie sąsiadów wg rozkładu $t$ -Studenta



Prawdopodobieństwa  
(podobieństwa) w przestrzeni  
wysokowymiarowej (z  $N(x_i, \sigma^2)$ )

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|^2 / 2\sigma^2)},$$

Prawdopodobieństwa (podobieństwa)  
w przestrzeni niskowymiarowej  
(z rozkładu  $t$ -Studenta,  $df = 1$ )

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}.$$

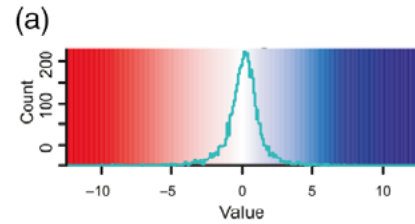
Minimalizacja dywergencji  
Kullbacka-Leiblera (względnej  
entropii) pomiędzy  $P$  i  $Q$  (metodą  
gradientu)

$$C = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}.$$

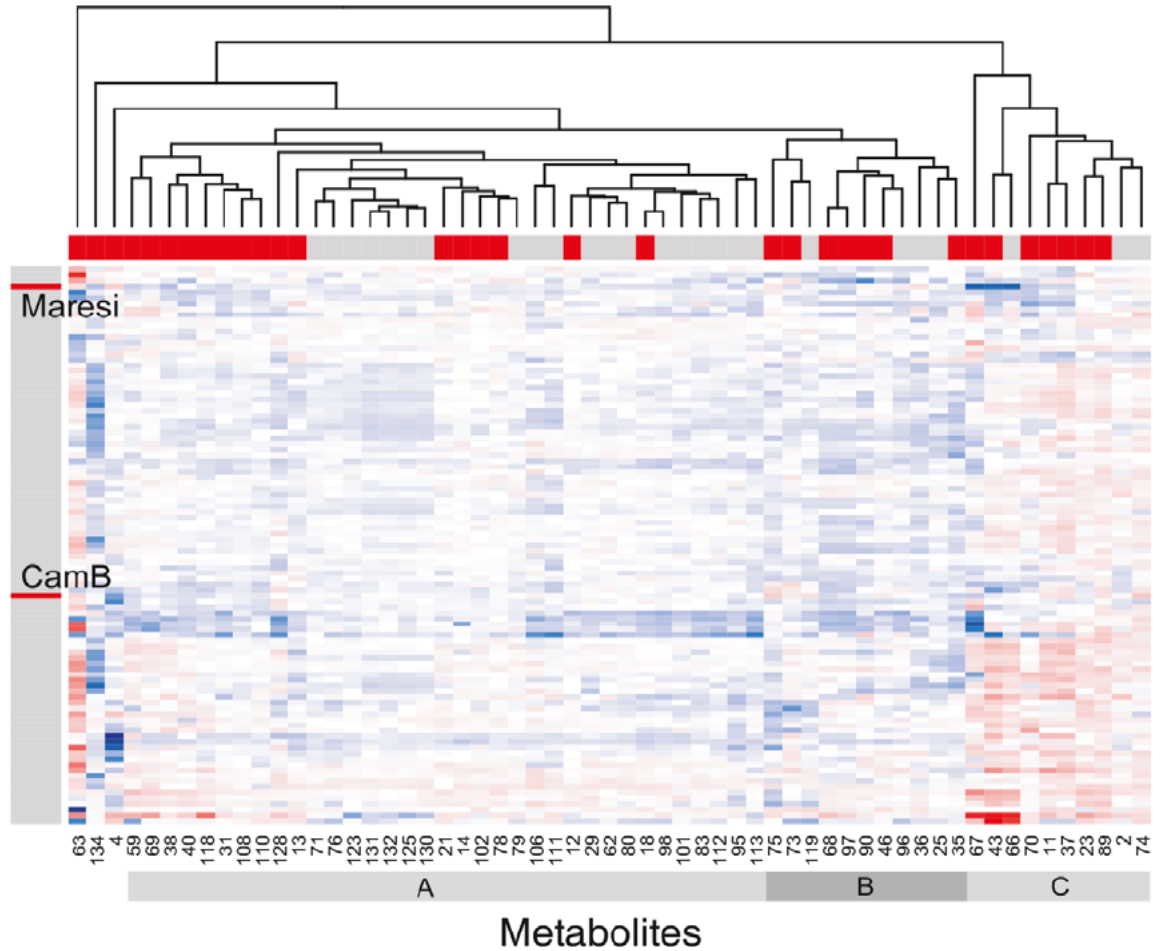
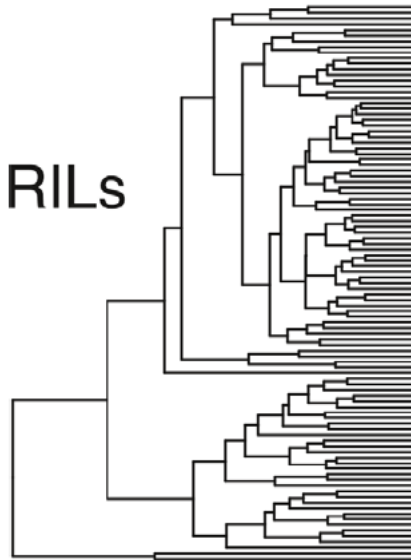
### **t-Distributed Stochastic Neighbor Embedding**

van der Maaten L., Hinton G., Visualizing data using t-SNE. *J. Machine Learning Res.* 9 (2008): 2579-2605

# Grupowanie obiektów i cech – mapa ciepła (heatmap)



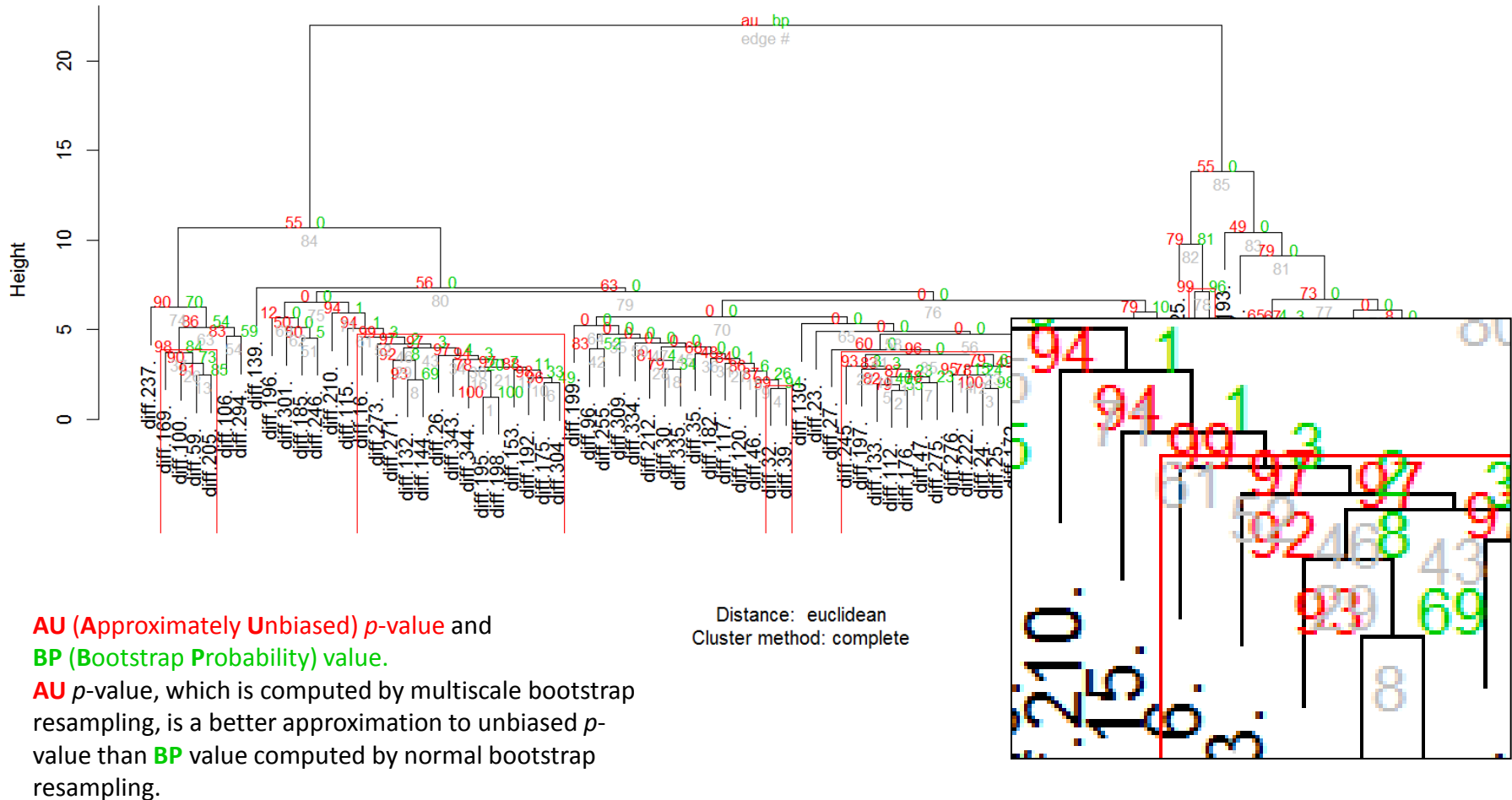
RILs



Wartości cech: odchylenia (susza – kontrola); tylko metabolity wykazujące interakcję linia x traktowanie

# Grupowanie obiektów /cech

Cluster dendrogram with AU/BP values (%)



Grupy białek wyznaczone metodą grupowania hierarchicznego na podstawie odległości Euklidesowej, skupienia istotne wyznaczone metodą bootstrap (1000 iteracji, AU > 95%, funkcja *pvclust* w R)

# Sieci korelacyjne: wizualizacja struktury korelacji cech

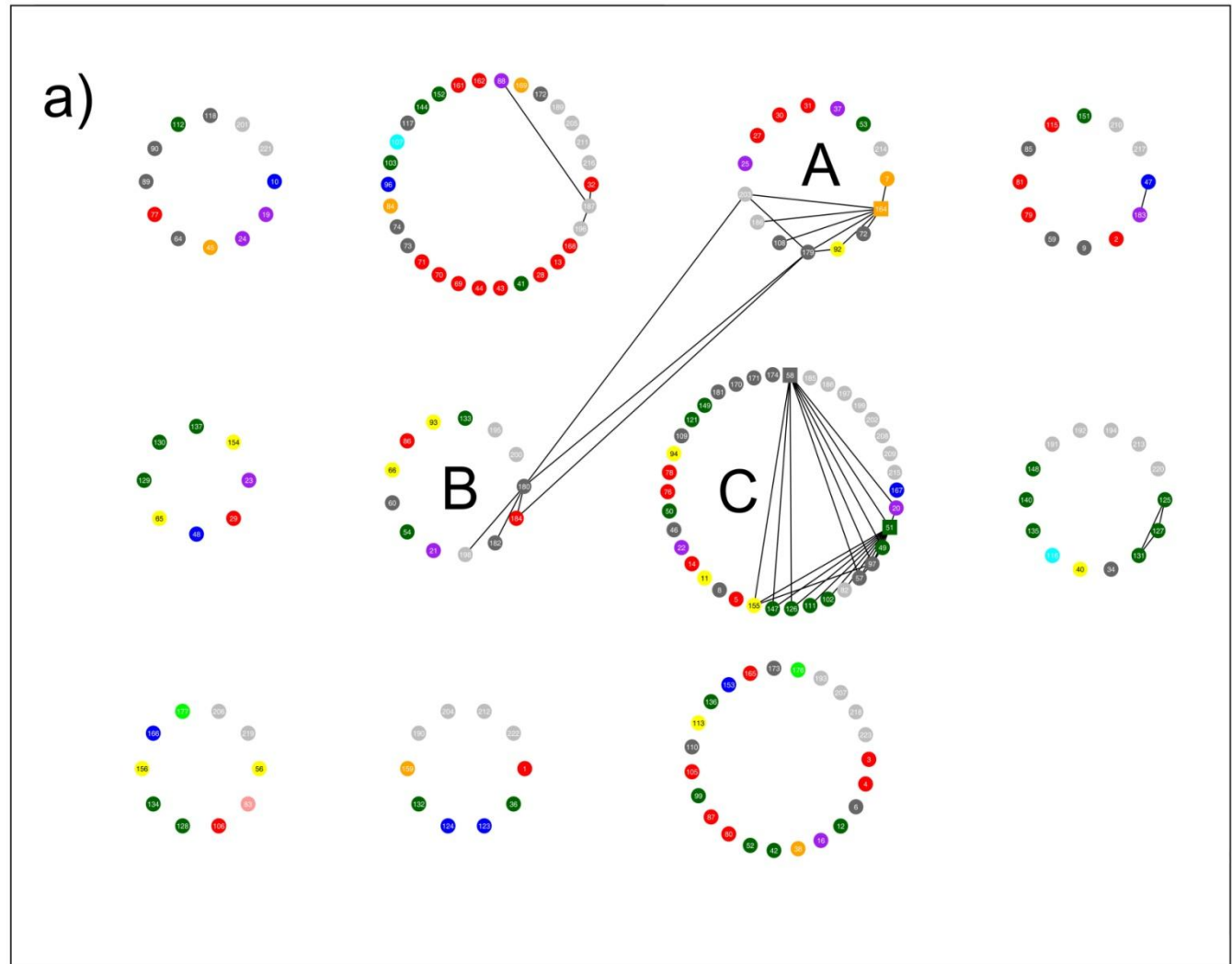
Obliczenia:  
WGCNA (R)

Macierze:  
- korelacji $\beta$   
- TOM – topology  
overlap matrix

-Możliwa ocena  
parametrów  
sieci:  
connectivity, ...

Wizualizacja:  
Cytoscape

Dane: zawartość  
białek, kolory –  
grupy funkcyjne



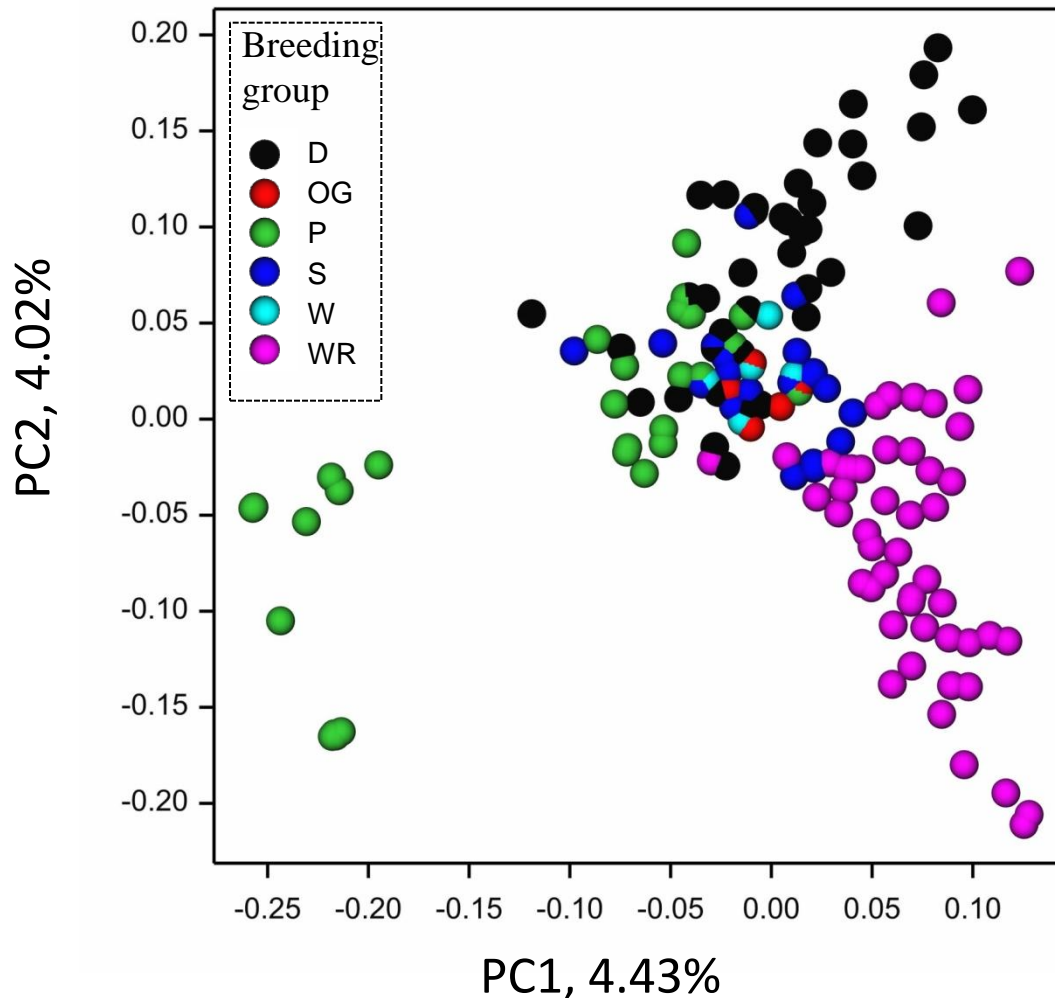
# Genotypowanie: opis genomu obiektów roślinnych za pomocą dużej liczby zmiennych dyskretnych, tzw. markerów molekularnych

CloneID	D1	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19
3341916	3	1	1	1	3	1	3	3	3	1	3
3342043	3	3	3	3	3	3	1	3	3	1	1
3342478	3	3	1	1	1	1	1	1	1	1	1
3342831	3	3	2	3	2	2	2	3	2	2	2
3343340	1	1	1	1	1	1	1	1	1	1	1
3343357	1	1	1	1	3	3	3	3	3	3	3
3344112	3	1	3	1	3	3	1	3	3	1	3
3345097	1	1	1	1	1	1	1	1	1	3	1
3345526	1	1	3	1	1	1	1	1	1	1	1
3345914	1	1	1	1	1	1	1	1	1	1	1
3346247	3	3	3	3	3	3	3	3	3	3	1
3348644	3	2	3	3	3	3	2	2	2	3	2
3349010	3	1	1	1	1	1	1	1	1	1	1
3350003	3	3	3	3	3	3	3	3	3	3	3
3351298	1	1	1	1	3	1	1	3	3	1	3
3353957	1	1	1	1	1	1	1	1	1	1	1
3354576	3	3	1	1	1	3	1	3	1	1	1
3355059	3	1	1	1	1	1	3	1	1	1	1
3355312	1	1	1	3	1	3	3	3	3	3	3
3356023	1	1	1	1	1	3	1	1	1	3	1
3356226	3	3	3	3	3	3	3	3	3	3	3
3356966	1	1	3	1	3	3	1	3	3	1	1

# Genotypowanie: opis genomu obiektów roślinnych za pomocą dużej liczby zmiennych dyskretnych, tzw. markerów molekularnych

CloneID	AlleleSequence	SNP	Snpposition
3341851	TGCAGCAACAATGCTAACCATCGAATTACACTCCTCTGCTACATGCTACTGGTTCATTCCGAGATCGGA	25:T>C	25
3341916	TGCAGATCCCTTCGCTGCTGCGTGCCCATCACCTGCAACGAGCAGAGCAGAGCAAGGAGAGGAGATTAC	40:A>G	40
3342043	TGCAGTACACTAGCGAGGACTAAATCTTTGATCTATGACGTACACCACTAACCACTGTAAGTCCATTCA	14:G>T	14
3342478	TGCAGACTTGTTAAGTTCAGTAGAAAACACATGACAGTCTACCGAGATCGGAAGAGCGGTTACAGCAGGA	18:A>G	18
3342831	TGCAGTCCGCACCAAAGAGCTTCACTATCTATGTAAACAGCGTGCAGCTCAAGGCAGCCCACACCAAGT	26:A>G	26
3343340	TGCAGTAAACCTTCCC GCCGTGCCCTCCAACCACCACACAGTCCCTATCCAAAATCCGAGATCGGAAG	12:T>C	12
3343357	TGCAGATTTCTCATCATCATATCATAACGATAAAAGCTCACAGAGCAGCACACTAGGTTCCGTACCTGCC	42:A>G	42
3344112	TGCAGCGAACAAAAATTCGTTATTTAAAAACAAAATGTCGCTGCACTCGTGTGCACTAGTTAATGAGG	11:A>G	11
3345097	TGCAGACAACGAATTGTTATATATGAAGTGGAGCAGAACTTGAAGACAACAGGCTGCATGTTGTCGAT	54:C>T	54
3345526	TGCAGGGTAACCCTGACATTCCC GCACGCTCTGGCAAACCTCGCCCTAAAGAGGCAGAGATCGCCGAGAT	28:C>G	28
3345914	TGCAGGTTTTTGAGCTCTTGCCGCACATTACTATCATACCAGTACCGAGATCGGAAGAGCGGTTACAGC	22:C>G	22
3346247	TGCAGGGTTGGACGGAGCGCGGAGAGCAAGGCCATGGCATGGCCCTAGCAACTTCTACGGGGATCATCC	13:G>T	13
3348644	TGCAGCTTTTTCTGCTCTGTGTTTGTGTTTAGATCTTGTCGATTGTAAGTCTGCTATACCGTTCAACCTTTA	9:T>G	9
3349010	TGCAGTATTATTAACAACGACAGCACAGGCATCATGCAGAGACCGAGATCGGAAGAGCGGTTACAGCAG	15:C>A	15
3350003	TGCAGGCCTGCATTTTCGTCGTGTAAGGTGGTTCGATCGATCACCTTGTGGAACCACGACCGAGATCGGA	16:G>T	16
3351298	TGCAGAACTTTAACAGTACGATGGAATGCCAAGCACCAGTGATCACACACGCGAACCAGCTGATCATAT	36:C>G	36
3353957	TGCAGATGCACCCTCACAGGCCGCGCAACATCCTGGTCAAGAACAAGAAAAACGGATACGAAACCGAGA	44:A>C	44
3354576	TGCAGTCCCAGCCGCGGCACAGCCCCACGTAGTCTCCAGGTTGAACGCGCCCACCAGCTCCGCCACC	20:C>G	20
3355059	TGCAGCTGCACCATGGATGATGATGGGTGGATTGGCGTCTGCTCCGAGATCGGAAGAGCGGTTACAGCA	13:T>C	13
3355312	TGCAGGATCGCATCGAGGAAACAGCAGCAAGATTTAGGAACCGAGATCGGAAGAGCGGTTACAGCAGGAA	5:G>A	5
3356023	TGCAGTAGCAGGTTGTGGTGCATCCCTAGCTTGTGTACCCTTCCAGGCAGCCGAGATCGGAAGAGCGG	5:T>C	5
3356226	TGCAGCTGGAACCTCAGTTGGGGTGAGGGCAAGGAAGGCGCTGTTGCATGAATGCTCGTTGTAGTATGTG	14:A>T	14
3356966	TGCAGCAACTGCGCGCCATCGTCTCCGTTACGTGTGCGGACGGAAGGACAGAGACGTGCCGAGATCGG	18:A>G	18

# Genotypowanie: opis genomu obiektów roślinnych za pomocą dużej liczby zmiennych dyskretnych, tzw. markerów molekularnych

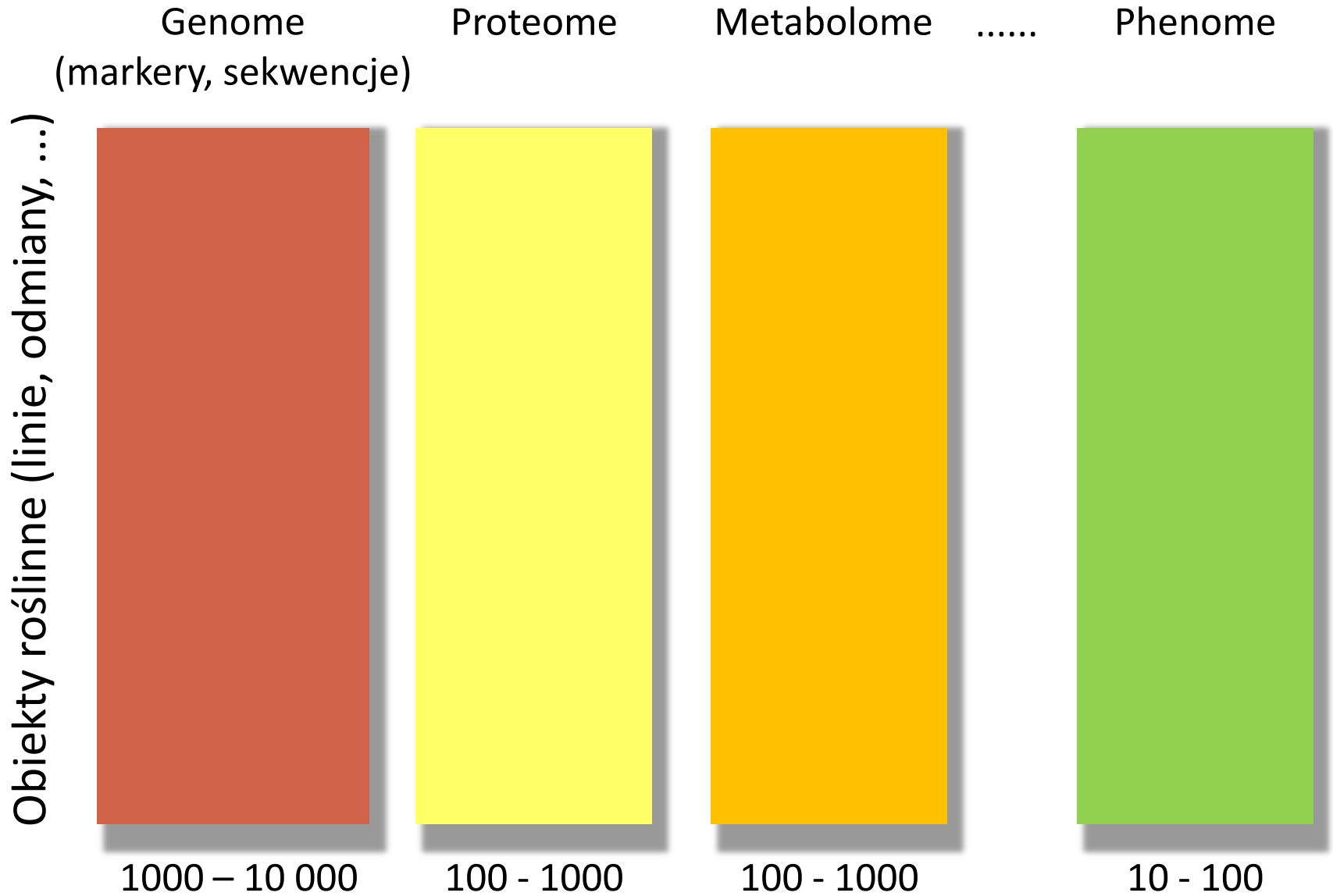


Wizualizacja struktury genetycznej populacji 149 linii żyta za pomocą analizy składowych głównych dla **2330** markerów SNP DartSeq

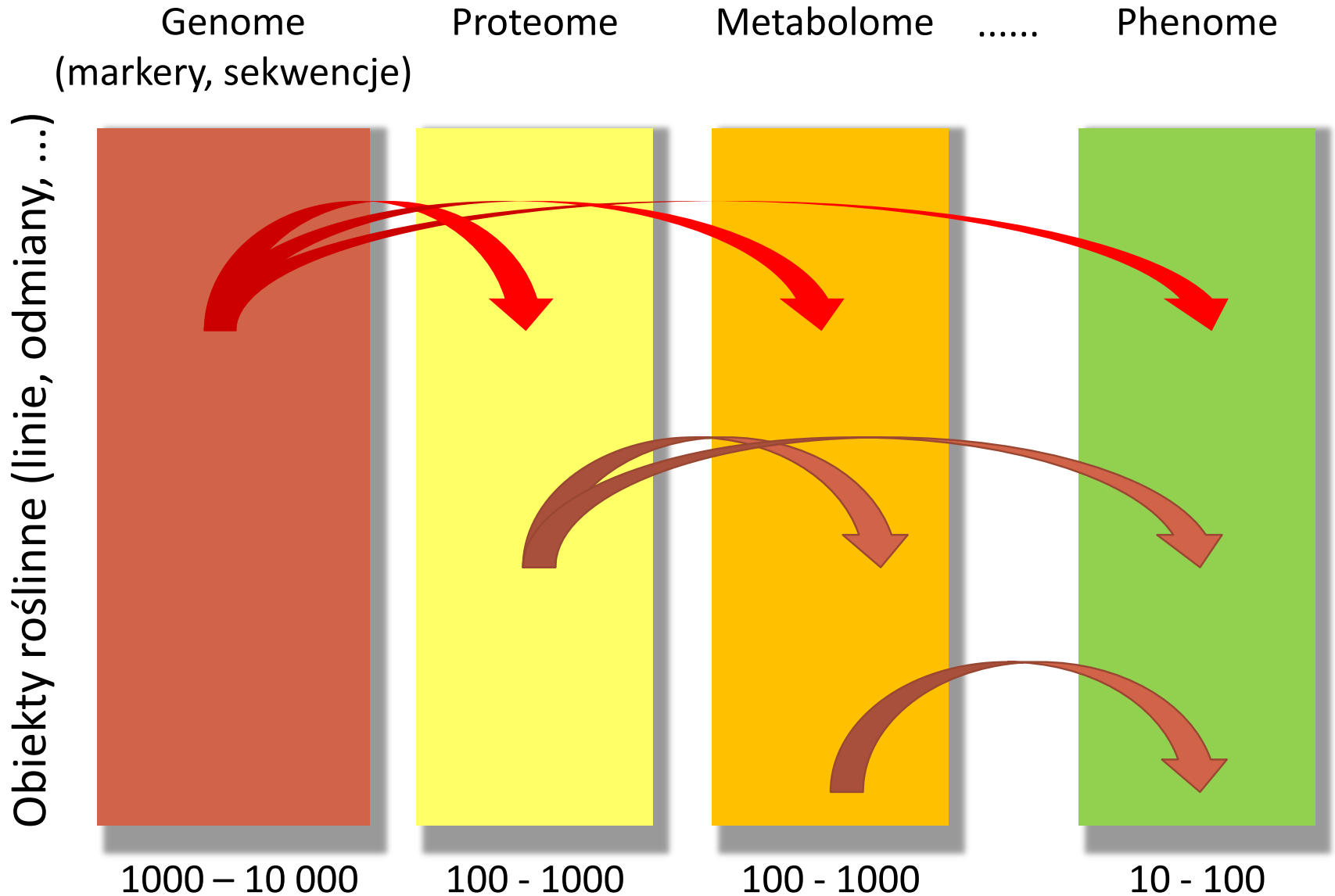
Rakoczy-Trojanowska i in., Plant Mol Biol Rep 35 (2017): 366-378



# „Omika”: analiza wielowymiarowa dla roślin



# „Omika”: analiza wielowymiarowa dla roślin



**Przerwa**

# Analiza danych funkcjonalnych

Madrigal and Krajewski *BioData Mining* (2015) 8:20  
DOI 10.1186/s13040-015-0051-7



BioData Mining

**METHODOLOGY**

**Open Access**



## Uncovering correlated variability in epigenomic datasets using the Karhunen-Loeve transform

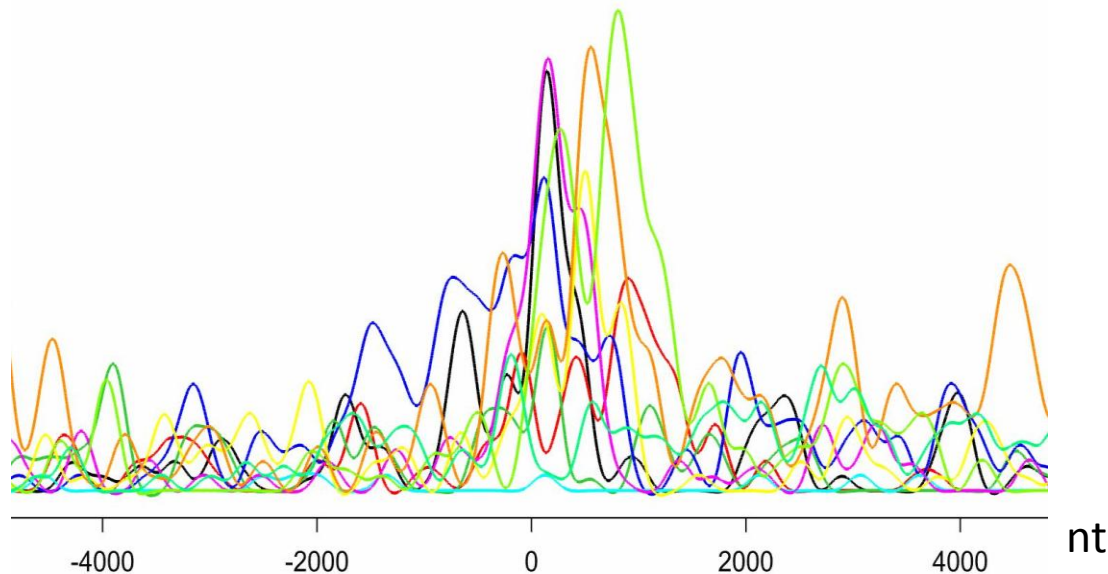
Pedro Madrigal<sup>1,2,3\*</sup> and Paweł Krajewski<sup>1</sup>

**Projekt Marie Curie ITN SYSFLO**





# Analiza modyfikacji histonowych



Sygnal modyfikacji histonu (np. metylacji)

Ilościowy – sygnał z wielu komórek o różnym stanie modyfikacji



Promotor

Miejsce  
początku  
transkrypcji

Wiele genów

HHHHMMHMHMMMHHMMHMMMMMMMMMMMMHHHHHHHHHHHHHHMM

Histon (białko) H normalne /M modyfikowane

# Funkcjonalne składowe główne

**Analiza składowych głównych** - obserwowana  $p$ -wymiarowa zmienna losowa  $\mathbf{x}$  o macierzy kowariancji  $V$ , składowe główne wyznaczane są poprzez rozwiązanie równania:

$$V\xi = \lambda\xi,$$

współrzędne punktów w PC to iloczyny skalarne

$$(\mathbf{x}^T \xi_1, \mathbf{x}^T \xi_2, \dots).$$

**Funkcjonalna analiza składowych głównych** – obserwowana funkcja losowa  $x(t)$ , funkcja kowariancyjna  $v(s, t)$ , składowe główne wyznaczane są z równania

$$\int_a^b v(s, t)\xi(s)ds = \lambda\xi(t).$$

które można zapisać używając operatora kowariancji jako  $V\xi(t) = \lambda\xi(t)$ .

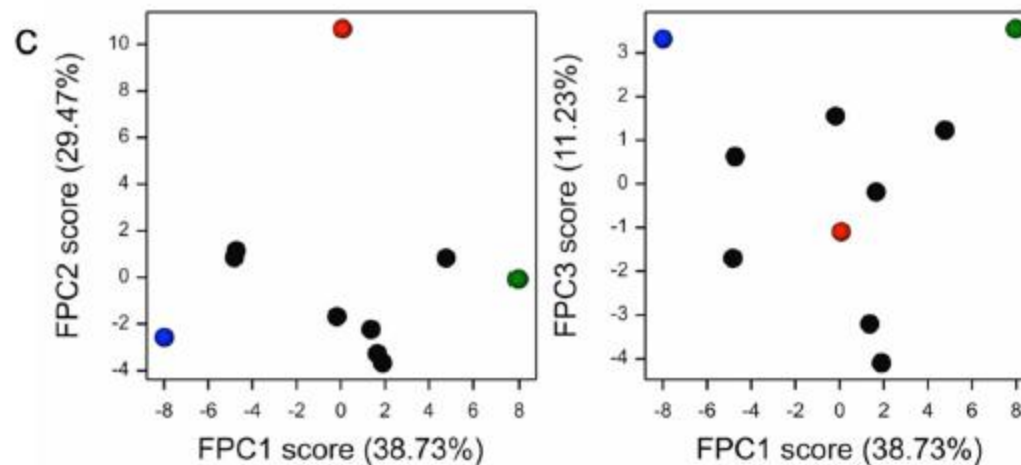
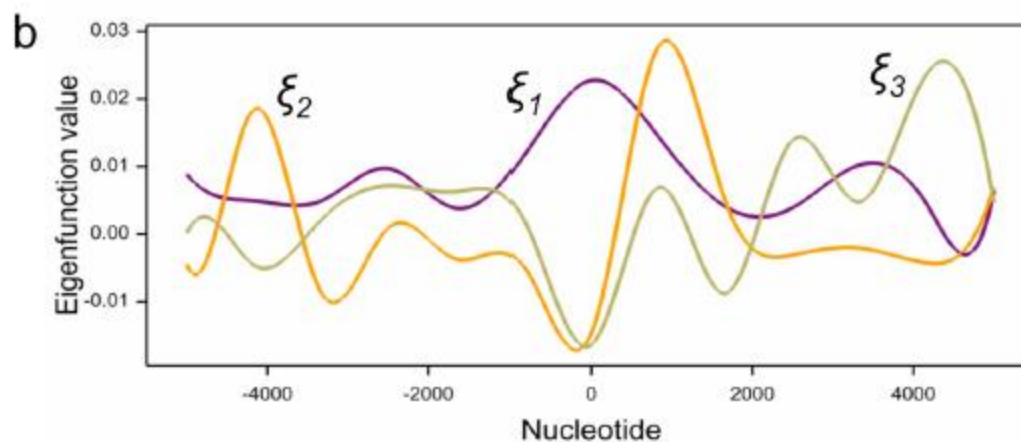
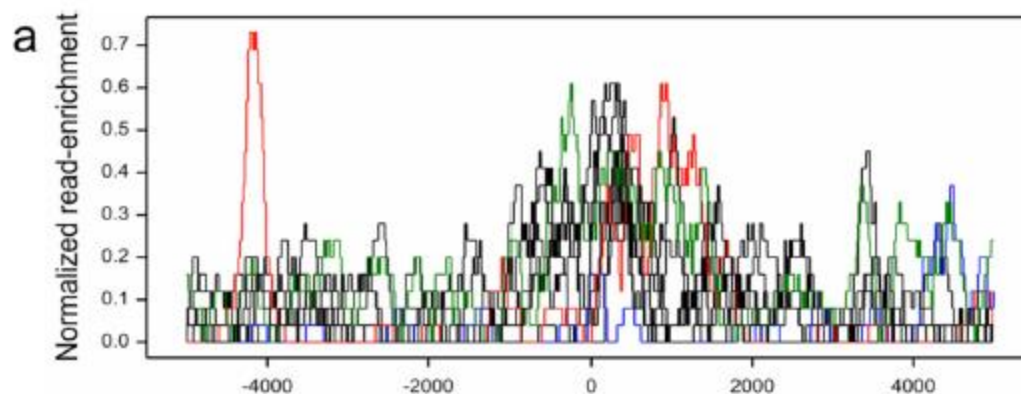
Współrzędne funkcjonalne (FPC) to  $(\int_a^b x(t)\xi_1(t)dt, \int_a^b x(t)\xi_2(t)dt, \dots)$ .

**Rozwiązanie numeryczne** – porze przybliżenie obserwacji i funkcji własnych za pomocą zbioru funkcji bazowych, np. typu spline.

10 obserwacji funkcji

Estymowane funkcje własne – ekstrema funkcji odpowiadają przedziałom największej zmienności między obserwacjami

Położenie 10 obserwacji w układzie pierwszej, drugiej i trzeciej funkcjonalnej składowej głównej.



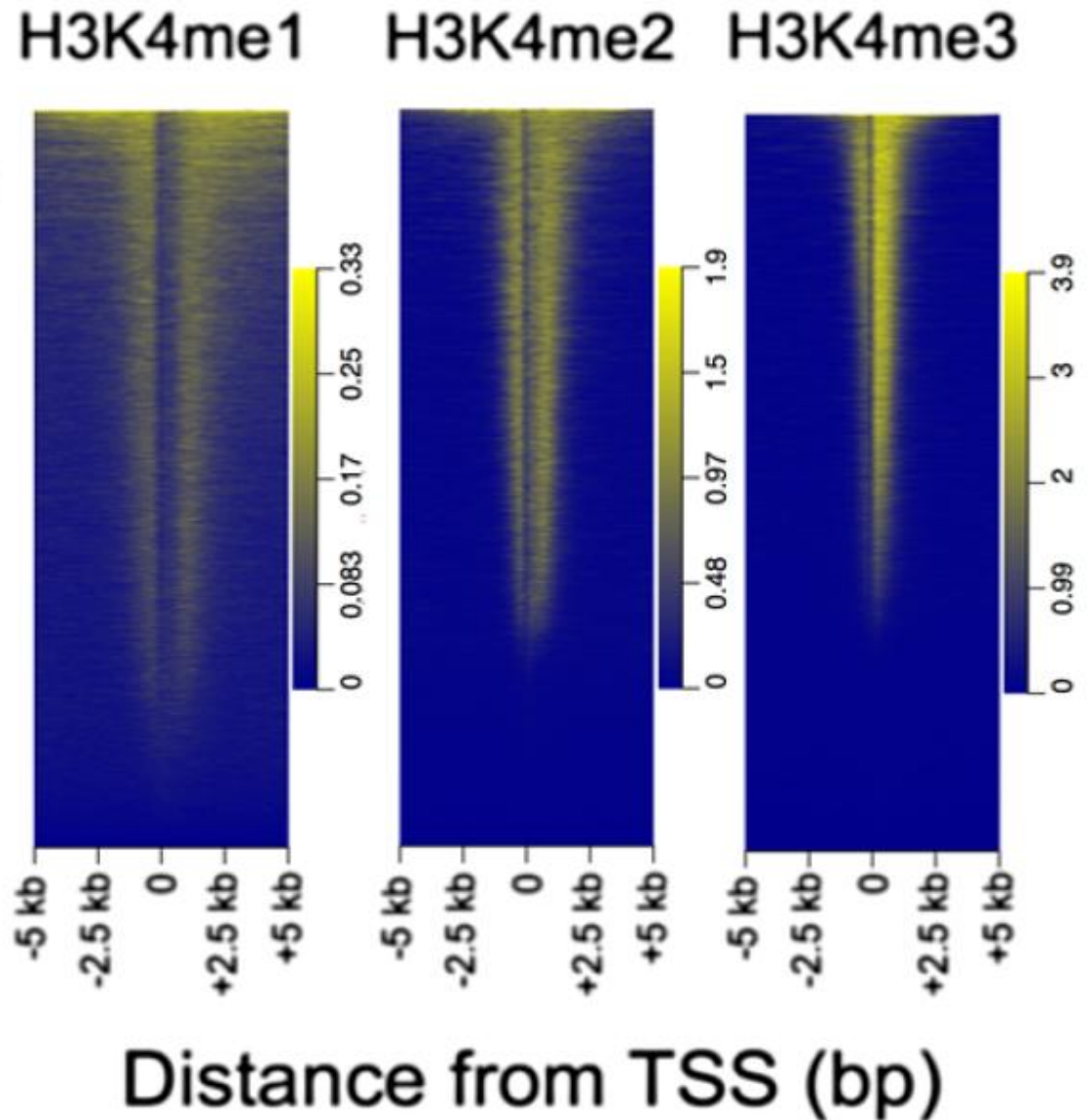


Profile trzech modyfikacji histonowych dla 19762 genów człowieka kodujących białka

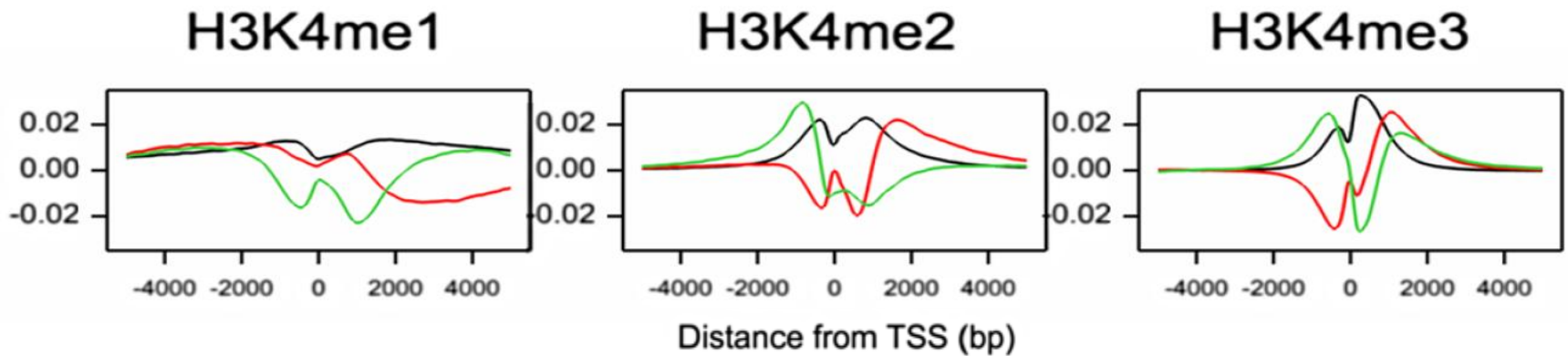
(dane:



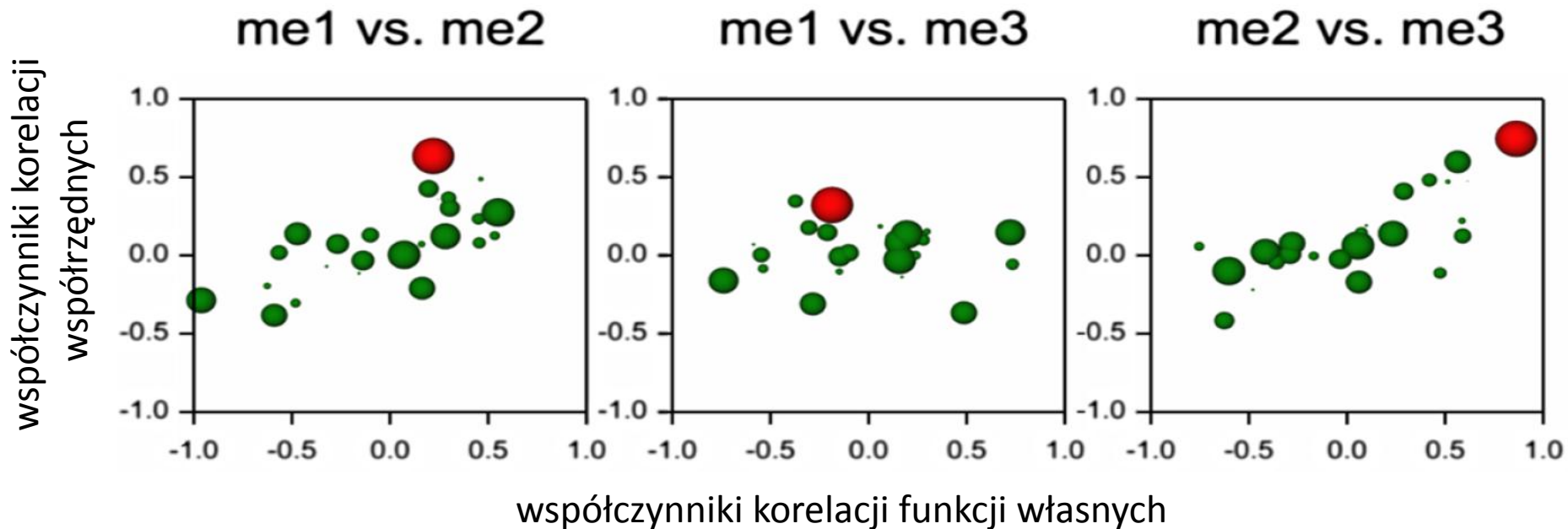
<http://www.roadmapepigenomics.org/>



Pierwsze trzy funkcje własne dla trzech modyfikacji histonowych



Wyliczono współczynniki korelacji pomiędzy współrzędnymi punktów i pomiędzy funkcjami własnymi dla 5 pierwszych FPC dla me1, me2 i me3



# Podsumowanie

- Eksploracja v. testowanie hipotez
- Metody deterministyczne v. zrandomizowane
- Różne metody – różne wyniki?
- Narzędzia
  - Obliczenia: Genstat, R
  - Wizualizacja: Genstat, R, Cytoscape
  - Sprzęt: PC, klastry PCSS



Dziękuję za uwagę