

Sprawozdanie z przedmiotu

Bioinformatyka

Grupowanie danych z ekspresji oraz analiza funkcjonalna uzyskanych grup.

Część 1 i 2

21.02.2017, 28.02.2017

Szymon Kocot
Wojciech Opoczyński
Łukasz Witek
Alexandra Zając

1. Przebieg ćwiczenia:

Lab 1

- a. Wczytano przykładowe dane pochodzące z eksperymentu Eisen - przebiegi czasowe ekspresji dla drożdży.

Za pomocą polecenia 'whos' wyświetlono informacje na temat wczytanych danych:

| Name | Size | Bytes | Class | Attributes |
|-------------|--------|--------|--------|------------|
| genes | 6400x1 | 806522 | cell | |
| times | 1x7 | 56 | double | |
| yeastvalues | 6400x7 | 358400 | double | |

- Zmienna 'yeastvalues' zawiera liczbowe wartości oznaczające poziom ekspresji badanych genów
- Zmienna 'genes' zawiera listę wszystkich genów drożdży przebadanych na mikromacierzy
- Zmienna 'times' zawiera wartości czasowe, w których mierzono poziom ekspresji genów

- b. Filtracja danych. Usunięcie pustych pól oraz zapamiętanie zmiennej genes do dalszych operacji.

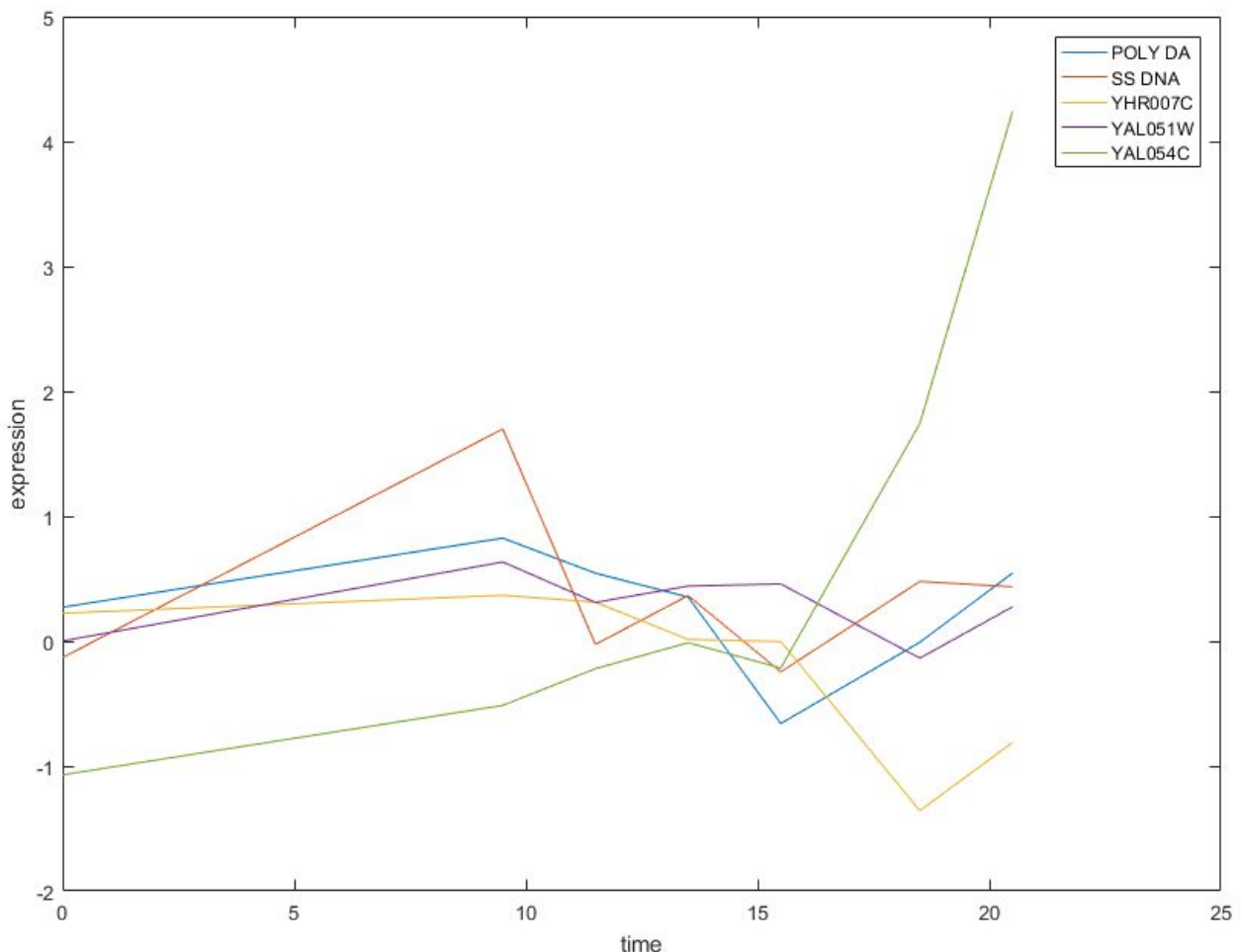
```
emptyspots = strcmp('EMPTY', genes);  
genes(emptyspots) = [];  
yeastvalues(emptyspots, :) = [];  
chipgenes = genes;
```

- c. Usunięcie brakujących wartości NaN, wykorzystując funkcję knnimpute, która zastępuje brakujące wartości najbliższą wartością pobraną z sąsiedniej kolumny. Jeżeli odpowiednia wartość z kolumny najbliższego sąsiedztwa jest NaN, używana jest następna najbliższa kolumna.

```
yeastvalues = knnimpute(yeastvalues);
```

- d. Wyświetlono wykres zawierający przebieg ekspresji 6-10 genów z mikromacierzy, korzystając z funkcji `legend`, aby nazwać przebiegi.

```
figure; plot(times, yeastvalues(6:10, :));  
xlabel('time'); ylabel('expression'); legend(genes(6:10));
```



Na wykresie można zaobserwować oscylowanie genów 6:9 dookoła wartości '0', oznaczającej niewielki poziom ekspresji lub nawet jej brak. Gen YAL054C jako jedyny z wymienionych ulega znaczącej ekspresji w 15 sekundzie. Poziom tej ekspresji narasta aż do zakończenia eksperymentu osiągając wartość bliską '4'.

- e. Usunięto geny o niskiej wartości absolutnej ekspresji (nie wykazujących żadnej aktywności w trakcie całego eksperymentu) z wykorzystaniem funkcji `genelowvalfilter`, parametru `absval=log2(3)`. Otrzymywany jest wektor logiczny zawierający wartości logiczne "1" dla wierszy macierzy reprezentujące geny, które wykazały się ekspresją, również macierz danych oraz listę nazw genów niezawierającą usuniętych wartości.

```
[mask, yeastvalues, genes] = genelowvalfilter(yeastvalues, genes, 'AbsValue', log2(3));
```

Pozostało 834 genów, 5480 zostało odfiltrowanych.

- f. Usunięto geny których przebiegi profili ekspresji mają niską entropię dla wartości percentyla równego 15%, poniżej którego dane są usuwane.

```
[mask2, yeastvalues, genes] = geneentropyfilter(yeastvalues, genes, 'Percentile', 15);
```

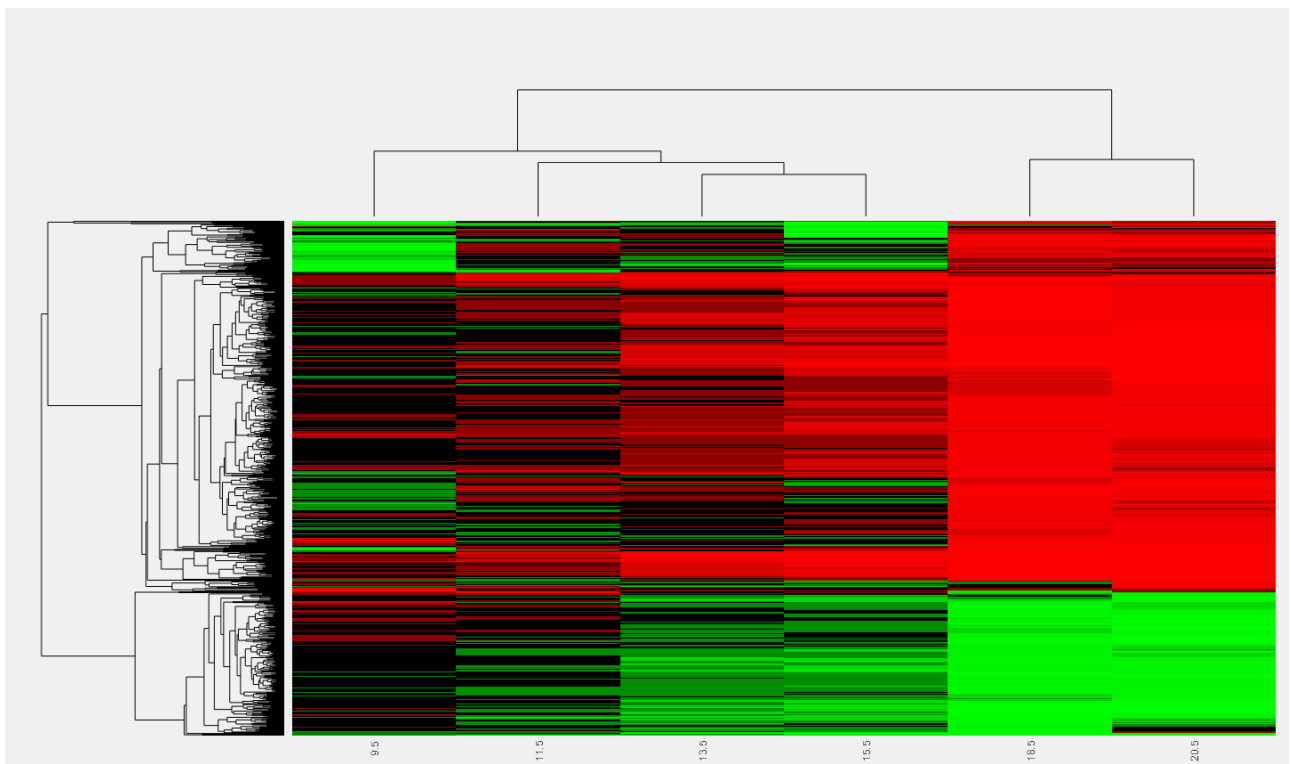
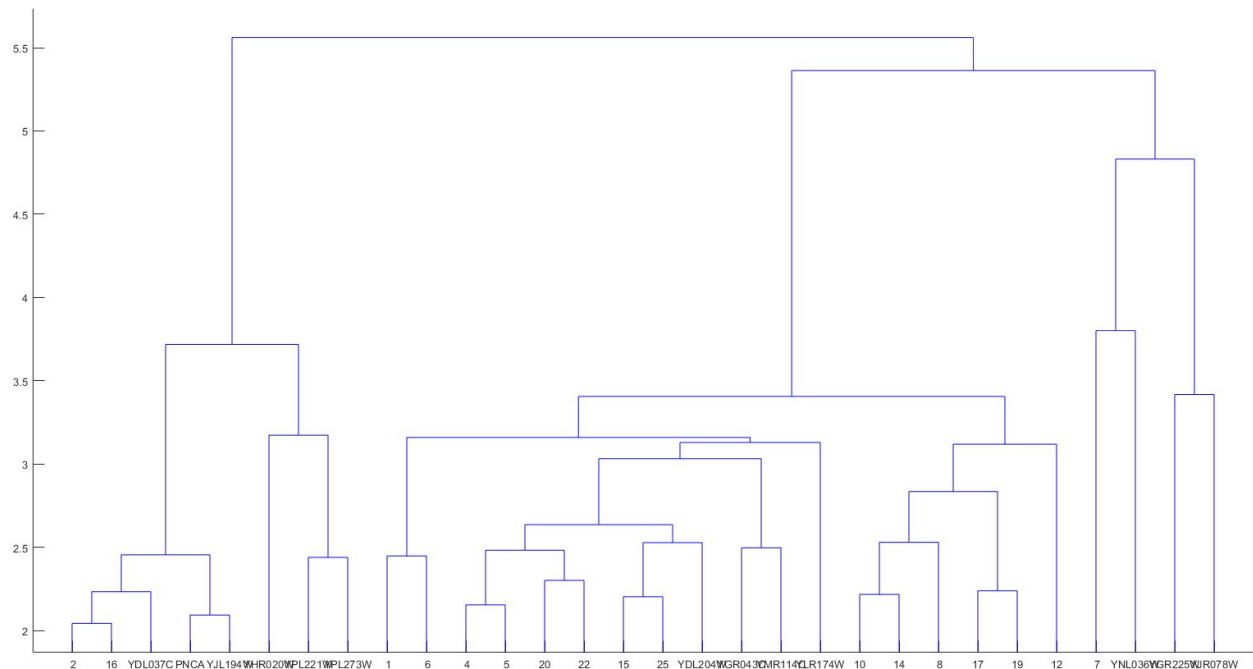
Pozostało 621 genów, 213 zostało odfiltrowanych.

Analiza nienadzorowana

dla 10 klastrów

a. Klastrowanie hierarchiczne.

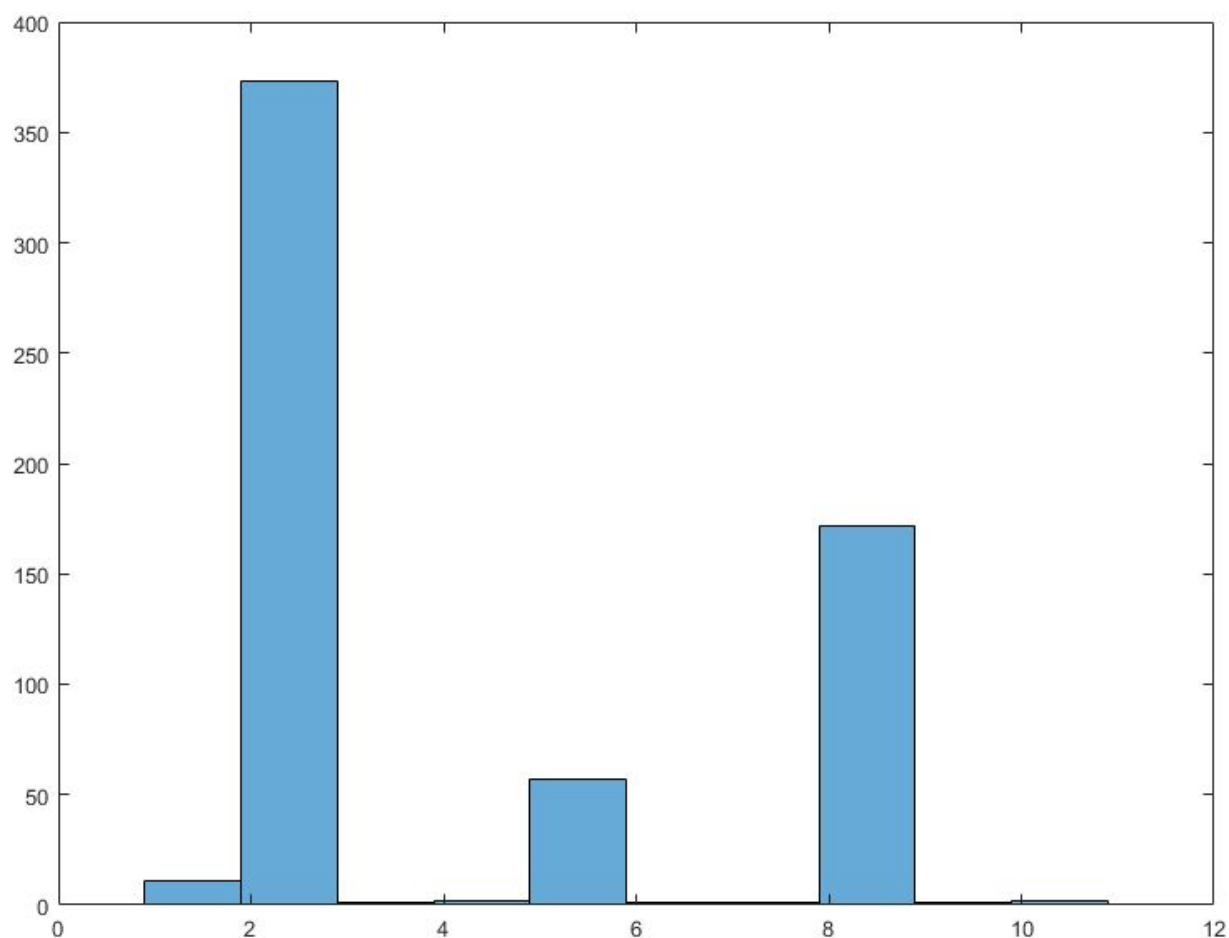
```
clusterTree = linkage(yeastvalues, 'average');  
clusters = cluster(clusterTree, 10);  
figure; set(gcf,'color','w');  
dendrogram(clusterTree, 'labels', genes); set(gcf,'color','w');  
cg = clustergram(yeastvalues(:,2:end), 'RowLabels', genes, 'ColumnLabels', times(2:end),  
'Linkage', 'average'); set(gcf,'color','w');
```



Zaprezentowane grafiki odzwierciedlają wynik klasteryzacji genów metodą klastrowania hierarchicznego. Na dendrogramie można zaobserwować wszystkie badane geny pogrupowane na zasadzie podobieństwa profili ekspresji. Im bliżej siebie są położone, tym większe między nimi podobieństwo. Duża odległość (oraz dłuższa gałąź łącząca) oznacza znaczącą różnicę w uleganiu ekspresji pomiędzy danymi genami. Na wykresie heatmapy można zaobserwować granicę dzielącą grupy genów na dwie części: zieloną oznaczającą wzrost ekspresji genu w danym momencie oraz czerwoną oznaczającą brak ekspresji genu. Ponownie, geny położone najbliżej siebie wykazują znaczne podobieństwo w profilach ekspresji, stąd widoczna granica pomiędzy grupą 'zieloną' i 'czerwoną'.

Histogram

```
figure; histogram(clusters, 10); set(gcf,'color','w');
```



Histogram przedstawia ilość genów objętych danym klastrem. Geny objęte jednym klastrem wykazują większe podobieństwo między sobą, niż do genów z pozostałych klastrów. Najwięcej genów (373) zostało objętych klastrem drugim. Większe ilości genów występują jeszcze w klastrach: ósmym - 172 geny, piątym - 57 genów.

b. Analiza k-średnich.

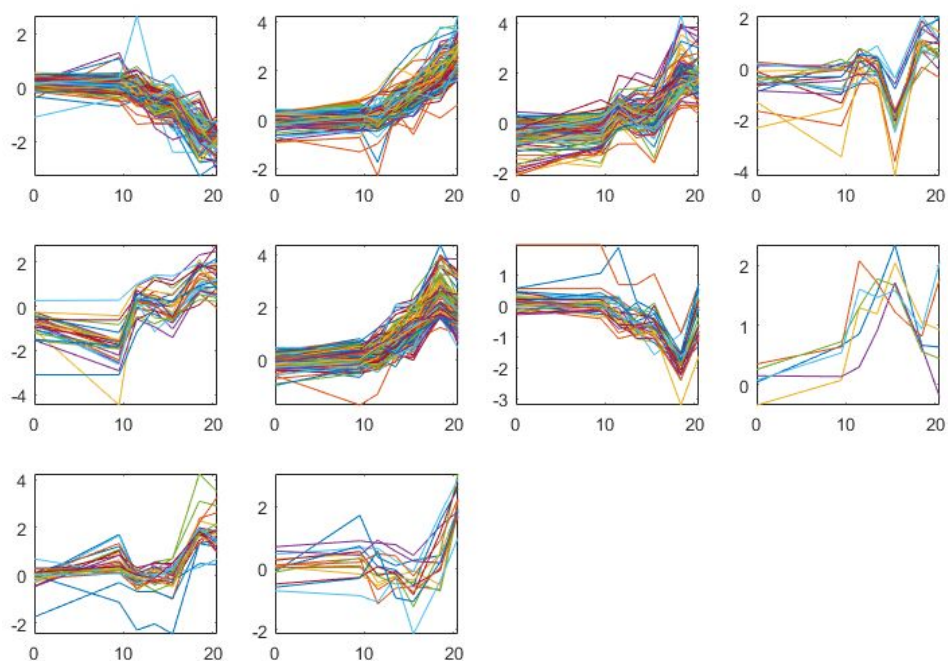
```
[cidx, ctrs] = kmeans(yeastvalues, 10, 'Distance', 'correlation', 'Replicates', 5);  
figure  
hist(cidx, 10)
```

```
figure; set(gcf,'color','w');  
for c = 1:10  
    subplot(3, 4, c);  
    plot(times, yeastvalues((cidx == c), :));  
    axis('tight');  
end  
suptitle('K-Means Clustering of Profiles');
```

```
figure; set(gcf,'color','w');  
for c = 1:10  
    subplot(3, 4, c);  
    plot(times, yeastvalues((cidx == c), :), 'r:', times, ctrs(c, :), 'kp-');  
    axis('tight');  
end  
suptitle('K-Means Clustering of Profiles with Centroids');
```

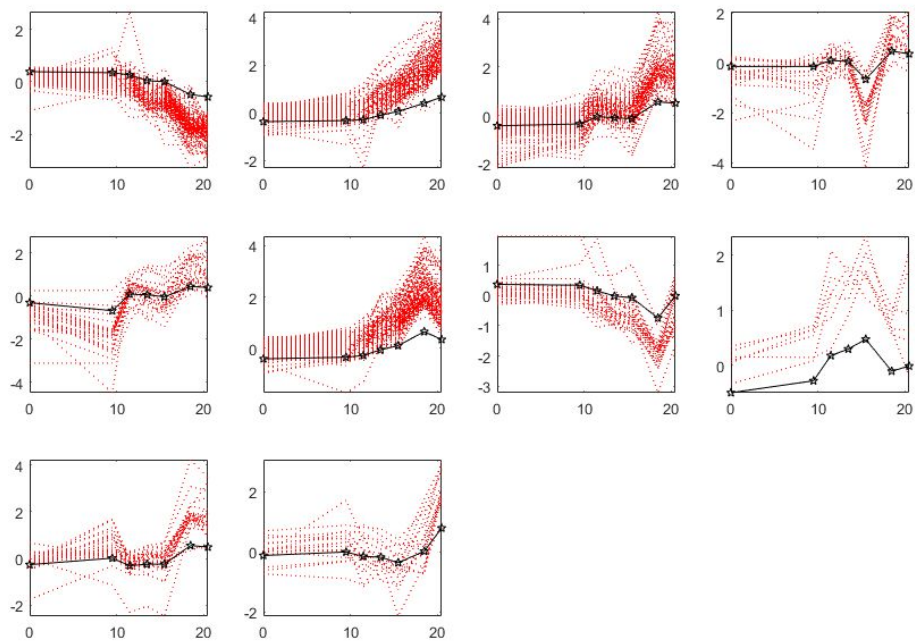
```
figure; set(gcf,'color','w');  
for c = 1:10  
    subplot(3, 4, c);  
    plot(times, ctrs(c, :));  
    axis('tight');  
end  
suptitle('K-Means Centroids');
```

K-Means Clustering of Profiles



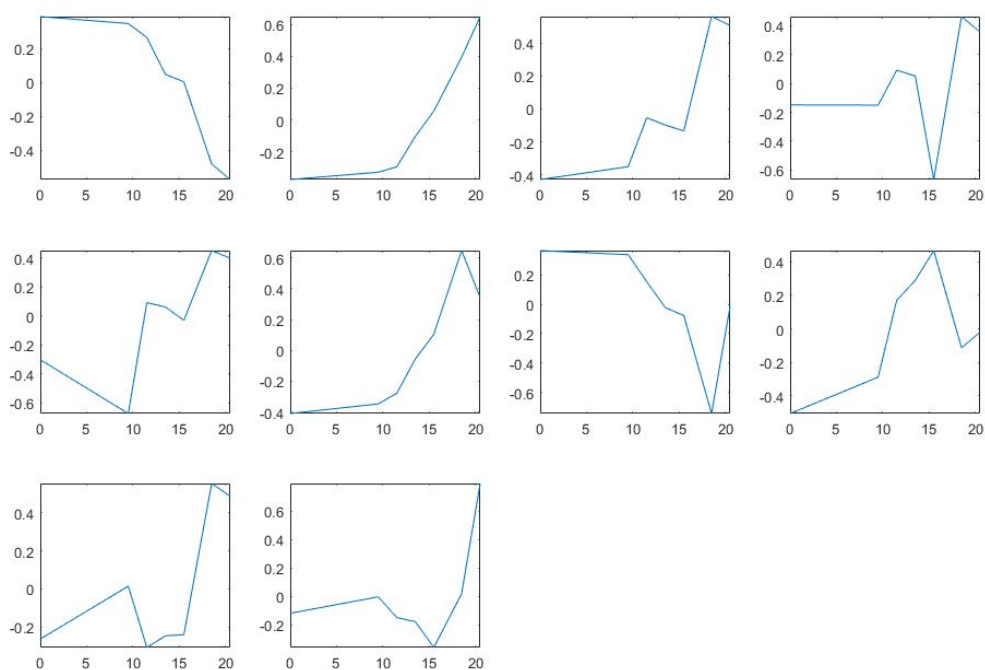
Wykres przedstawia profile ekspresji wszystkich genów objętych danym klastrem. Najmniej genów znajduje się w klastrze nr 6 (6 genów), najwięcej w klastrze nr 5 (150 genów). Można zauważyć, iż geny pogrupowane są na zasadzie podobieństwa ekspresji w czasie (w pewnym momencie ekspresja wzrasta, w innych maleje). Dzięki temu podobieństwu geny są pogrupowane w taki, a nie inny sposób.

K-Means Clustering of Profiles with Centroids



Wykres przedstawia profile ekspresji genów w klastrach wraz z ich centroidami, czyli profilami charakterystycznymi - takimi, do których zostają przyrównane pozostałe profile ekspresji.

K-Means Centroids



Na danym wykresie można zauważyć dokładny kształt centroidów dla każdego klastra, czyli pewnych średnich profili ekspresji, dla każdego klastra, wokół których gromadziły się pozostałe profile ekspresji genów.

Metoda k-średnich jest metodą należącą do grupy algorytmów analizy skupień tj. analizy polegającej na szukaniu i wyodrębnianiu grup obiektów podobnych (skupień) . Reprezentuje ona grupę algorytmów niehierarchicznych. Główną różnicą pomiędzy niehierarchicznymi i hierarchicznymi algorytmami jest konieczność wcześniejszego podania ilości skupień.

Przy pomocy metody k-średnich zostanie utworzonych k różnych możliwie odmiennych skupień. Algorytm ten polega na przenoszeniu obiektów ze skupienia do skupienia tak długo aż zostaną zoptymalizowane zmienności wewnątrz skupień oraz pomiędzy skupieniami. Oczywistym jest, iż podobieństwo w skupieniu powinno być jak największe, zaś osobne skupienia powinny się maksymalnie od siebie różnić.

Lab 2.

1. Ze strony konsorcjum Gene Ontology pobrano plik 'go-basic.obo' zawierający bazę danych ontologii genowych.
2. Wczytano bazę ontologii do środowiska Matlab za pomocą funkcji:

```
GO = geneont('File', 'go-basic.obo')
```

3. Za pomocą funkcji 'get' otrzymano między innymi informacje o aktualności bazy oraz o ilości terminów GO aktualnie zarejestrowanych w bazie:

```
data_version: 'releases/2017-02-27'  
Terms: [46473x1 geneont.term]
```

4. Pobrano z internetu plik zawierający bazę adnotacji terminów GO dla organizmu *Saccharomyces cerevisiae*. Zawartość pliku wczytano za pomocą funkcji:

```
yeastAnnot = goannotread('gene_association.sgd')
```

5. Do osobnych zmiennych wczytano listę genów, listę terminów powiązanych z tymi genami oraz typ ontologii:

```
yeastGenes = { yeastAnnot.DB_Object_Symbol}  
yeastGO = [yeastAnnot.GOid]  
yeastAspect = {yeastAnnot.Aspect}
```

Odfiltrowano tylko terminy GO pochodzące z grafu 'Proces Biologiczny':

```
mask = strcmp({yeastAnnot.Aspect}, 'P')
```

Wykorzystano otrzymany indeks logiczny do usunięcia odpowiednich wierszy z listy genów oraz z listy terminów GO:

```
yeastGenes_BP = yeastGenes(mask);  
yeastGO_BP = yeastGO(mask);
```

Otrzymano 40213 genów adnotowanych terminami należącymi do procesu biologicznego.

6. Wczytano dane, uzyskane na poprzednim laboratorium. Dla każdego klastra przeprowadzono analizę częstości występowania terminów GO (osobno dla metody klastrowania hierarchicznego oraz K-means):

Do znalezienia odpowiednich wartości użyto funkcji:

```
[GOTerms, clusterGOTermsNo, chipGOTermsNo] =  
getClusterGOTerms(GO, yeastGenes_BP, yeastGO_BP, genes, chipgenes, clusterNo,  
clusters)
```

Za jej pomocą uzyskano:

GOTerms - listę terminów GO genów tworzących dany klaster

clusterGOTermsNo - liczbę wystąpień danego terminu GO w analizowanym klastrze

chipGOTermsNo - liczbę wystąpień danego terminu GO w całym zbiorze genów analizowanych na mikromacierzy

Przykładowo, dla metody k-means, użyto następującego algorytmu, zapisującego w kolejnych komórkach wartości uzyskane podczas badania danego klastra:

```
for clusterNo = 1:10  
    [GOTerms_k{clusterNo} clusterGOTermsNo_k{clusterNo} chipGOTermsNo_k{clusterNo}] =  
    getClusterGOTerms(GO, yeastGenes_BP, yeastGO_BP, genes, chipgenes, clusterNo,  
    cidx);  
end
```

Następnie przeprowadzono analizę statystyczną umożliwiającą wybór istotnych dla eksperymentu genów, za pomocą funkcji 'hygecdf', z pomocą której otrzymano p-wartości dla każdego genu w klastrach.

Przykładowo, algorytm obliczania p-wartości dla genów klastrowanych metodą k-means prezentuje się następująco:

```
for i = 1:10  
  
    x_k = clusterGOTermsNo_k{i};  
    M_k = size(chipgenes);  
    M_k = M_k(1);  
    K_k = chipGOTermsNo_k{i};  
    N_k = length(clusters(clusters == i));
```


$pVal_{k\{i\}} = 1 - hygecdf(x_{k-1}, M_k, K_k, N_k);$

end

gdzie x , M , K , N są parametrami funkcji hipergeometrycznej, koniecznymi do wyznaczenia wszystkich p-wartości

7. Za pomocą wybranego terminu GO (5975), odczytano informacje na jego temat funkcją `GO(5975).terms`:

```
GO(5975).terms
  id: 5975
  name: 'carbohydrate metabolic process'
  ontology: 'biological process'
  definition: '"The chemical reactions and pathways involving carbohydrates, any of a gr...'
  comment: ''
  synonym: {'synonym' '"carbohydrate metabolism" EXACT []'}
  is_a: [2x1 double]
  part_of: [0x1 double]
  obsolete: 0
```

gdzie:

definition - wyjaśnienie do jakiego szlaku należy i jakie procesy uwzględnia dany termin

name - nazwa procesu jaki obejmuje termin

synonym - inna nazwa procesu w jaki jest zaangażowany termin

is_a - inne terminy GO określające pochodzenie wybranego terminu (44238, 71704)

part_of - inne terminy GO określone wybranym terminem

Wywołano funkcję `'getancestors(GO, 5975)'`, za pomocą której otrzymano następujące terminy: 5975, 8150, 8152, 44238, 71704. W porównaniu do metody `'is_a'` wskazuje ona dodatkowo na samą siebie (5975), każdy proces biologiczny (8150) oraz każdy proces metaboliczny (8152). Zawiera zatem szerszą przynależność terminu, aniżeli metoda `'is_a'`. Metoda `'part_of'` w tym przypadku nie daje żadnego rezultatu.

8. Wygenerowano raport, zawierający informacje o istotnych statystycznie genach podzielonych na metody klastrowania i klastry. Zawiera informacje dotyczące nazwy procesu, opisu, numeru klastra oraz p-wartość. Raport został posortowany według wzrastających p-wartości:

```
C:\Users\wojte\Dropbox\studia\bioinformatyka\semestr VI\lab 1+2\raport - Notepad++
Plik Edycja Szukaj Widok Format Skladnia Ustawienia Tools Makra Uruchom Wylczki Okno ?
raport
1 Raport z Laboratorium Bioinformatyki nr 1 i 2
2
3 Imię i Nazwisko: Szymon Kocot
4 Imię i Nazwisko: Wojciech Opoczyński
5 Imię i Nazwisko: Łukasz Witek
6 Imię i Nazwisko: Alexandra Zając
7 Data 14-Mar-2017 20:53:02
8
9 Metoda klastrowania: klastrowanie hierarchiczne
10
11 Numer klastra: 1
12 GO Term p-val clusterNo/chipNo geneName definition
13
14
15 Numer klastra: 2
16 GO Term p-val clusterNo/chipNo geneName definition
17
18 GO:0000422 0.0035 1 / 2 mitophagy "The autophagic process in which mitochondria are delivered to the vacuole and deg
19 GO:0005975 0.0193 1 / 7 carbohydrate metabolic process "The chemical reactions and pathways involving carbohydrates, i
20 GO:0006000 0.0376 1 / 1 fructose metabolic process "The chemical reactions and pathways involving fructose, the ketohe
21
22 Numer klastra: 3
23 GO Term p-val clusterNo/chipNo geneName definition
24
25 GO:0006412 0.0002 1 / 6 translation "The cellular metabolic process in which a protein is formed, using the sequenc
26 GO:0006418 0.0003 1 / 3 tRNA aminoacylation for protein translation "The synthesis of aminoacyl tRNA by the formati
27 GO:0006433 0.0005 1 / 1 prolyl-tRNA aminoacylation "The process of coupling proline to prolyl-tRNA, catalyzed by proly
28 GO:0006450 0.0010 1 / 2 regulation of translational fidelity "Any process that modulates the ability of the translat
29
30 Numer klastra: 4
31 GO Term p-val clusterNo/chipNo geneName definition
32
33
34 Numer klastra: 5
35 GO Term p-val clusterNo/chipNo geneName definition
36
37 GO:0006278 0.0059 2 / 13 RNA-dependent DNA biosynthetic process "A DNA biosynthetic process that uses RNA as a template
38 GO:0006508 0.0059 2 / 22 proteolysis "The hydrolysis of proteins into smaller polypeptides and/or amino acids by cle
39 GO:0006515 0.0059 1 / 1 misfolded or incompletely synthesized protein catabolic process "The chemical reactions and
40 GO:0006623 0.0090 1 / 1 protein targeting to vacuole "The process of directing proteins towards the vacuole, usually
41 GO:0006810 0.0090 1 / 27 transport "The directed movement of substances (such as macromolecules, small molecules, ions
>
Normal text file length: 22 241 lines: 161 Ln: 1 Col: 1 Sel: 0 | 0 Unix (LF) ANSI INS
21:18
14.03.2017
```

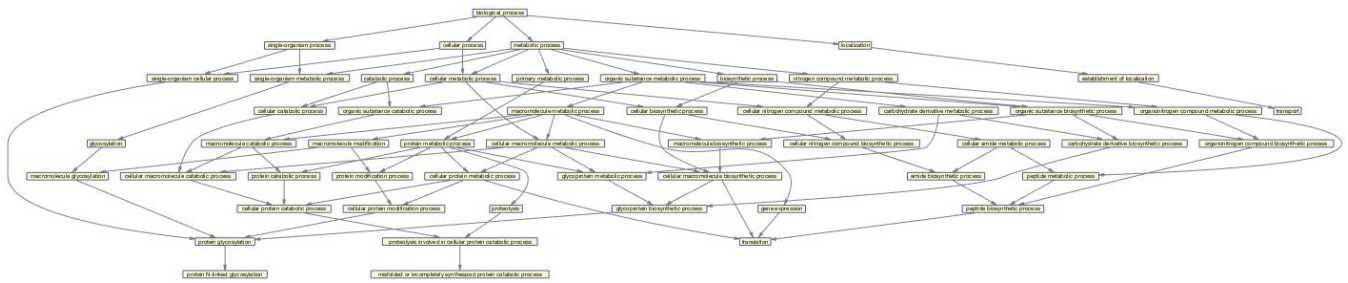
```
C:\Users\wojte\Dropbox\studia\bioinformatyka\semestr VI\lab 1+2\raport - Notepad++
Plik Edycja Szukaj Widok Format Skladnia Ustawienia Tools Makra Uruchom Wylczki Okno ?
raport
41 GO:0006810 0.0090 1 / 27 transport "The directed movement of substances (such as macromolecules, small molecules, ions)
42 GO:0006820 0.0090 1 / 2 anion transport "The directed movement of anions, atoms or small molecules with a net negat
43 GO:0007039 0.0165 1 / 4 protein catabolic process in the vacuole "The chemical reactions and pathways resulting in t
44 GO:0008150 0.0179 9 / 517 biological process "Any process specifically pertinent to the functioning of integrated living
45 GO:0032197 0.0180 2 / 23 transposition, RNA-mediated "Any process involved in a type of transpositional recombination
46 GO:0043161 0.0268 1 / 3 proteasome-mediated ubiquitin-dependent protein catabolic process "The chemical reactions and
47 GO:0045721 0.0356 1 / 1 negative regulation of gluconeogenesis "Any process that stops, prevents, or reduces the frequ
48 GO:0055085 0.0405 1 / 24 transmembrane transport "The process in which a solute is transported across a lipid bilaye
49
50 Numer klastra: 6
51 GO Term p-val clusterNo/chipNo geneName definition
52
53
54 Numer klastra: 7
55 GO Term p-val clusterNo/chipNo geneName definition
56
57
58 Numer klastra: 8
59 GO Term p-val clusterNo/chipNo geneName definition
60
61 GO:0006412 0.0272 1 / 6 translation "The cellular metabolic process in which a protein is formed, using the sequenc
62 GO:0006418 0.0272 1 / 3 tRNA aminoacylation for protein translation "The synthesis of aminoacyl tRNA by the formati
63
64 Numer klastra: 9
65 GO Term p-val clusterNo/chipNo geneName definition
66
67
68 Numer klastra: 10
69 GO Term p-val clusterNo/chipNo geneName definition
70
71
72 *****
73
74 k-srednich: k-srednich
75
76 Numer klastra: 1
77 GO Term p-val clusterNo/chipNo geneName definition
78
79 GO:0006278 0.0071 1 / 13 RNA-dependent DNA biosynthetic process "A DNA biosynthetic process that uses RNA as a template
80 GO:0006414 0.0071 1 / 1 translational elongation "The successive addition of amino acid residues to a nascent polype
81 GO:0006508 0.0071 1 / 22 proteolysis "The hydrolysis of proteins into smaller polypeptides and/or amino acids by cle
>
Normal text file length: 22 241 lines: 161 Ln: 1 Col: 1 Sel: 0 | 0 Unix (LF) ANSI INS
21:20
14.03.2017
```

Normal text file

| GO Term | p-val | clusterNo/chipNo | geneName | definition |
|------------------|--------|------------------|---|--|
| GO:0006414 | 0.0071 | 1 / 1 | translational elongation | "The successive addition of amino acid residues to a nascent polypeptide" |
| GO:0006508 | 0.0071 | 1 / 22 | proteolysis | "The hydrolysis of proteins into smaller polypeptides and/or amino acids by cleavage" |
| GO:0006623 | 0.0071 | 1 / 1 | protein targeting to vacuole | "The process of directing proteins towards the vacuole, usually by a specific signal" |
| GO:0006749 | 0.0071 | 1 / 1 | glutathione metabolic process | "The chemical reactions and pathways involving glutathione, usually in the cytoplasm" |
| GO:0006810 | 0.0071 | 1 / 27 | transport | "The directed movement of substances (such as macromolecules, small molecules, ions, or groups of molecules) from one location to another" |
| GO:0007039 | 0.0095 | 1 / 4 | protein catabolic process in the vacuole | "The chemical reactions and pathways resulting in the breakdown of proteins in the vacuole" |
| GO:0008150 | 0.0212 | 9 / 517 | biological_process | "Any process specifically pertinent to the functioning of integrated living organisms" |
| GO:0010257 | 0.0282 | 1 / 1 | NADH dehydrogenase complex assembly | "The aggregation, arrangement and bonding together of individual subunits to form a functional complex" |
| GO:0032197 | 0.0489 | 1 / 23 | transposition, RNA-mediated | "Any process involved in a type of transpositional recombination" |
| Numer klastra: 2 | | | | |
| GO Term | p-val | clusterNo/chipNo | geneName | definition |
| GO:0006412 | 0.0154 | 1 / 6 | translation | "The cellular metabolic process in which a protein is formed, using the sequence of a messenger RNA as a template" |
| GO:0006418 | 0.0305 | 1 / 3 | tRNA aminoacylation for protein translation | "The synthesis of aminoacyl-tRNA by the formation of an ester bond between the amino group of an amino acid and the 3' terminal adenosine of a transfer RNA" |
| GO:0006420 | 0.0454 | 1 / 1 | arginyl-tRNA aminoacylation | "The process of coupling arginine to arginyl-tRNA, catalyzed by arginyl-tRNA synthetase" |
| Numer klastra: 3 | | | | |
| GO Term | p-val | clusterNo/chipNo | geneName | definition |
| GO:0000422 | 0.0211 | 1 / 2 | mitophagy | "The autophagic process in which mitochondria are delivered to the vacuole and degraded" |
| GO:0006000 | 0.0211 | 1 / 1 | fructose metabolic process | "The chemical reactions and pathways involving fructose, the ketohexose" |
| GO:0006003 | 0.0211 | 1 / 1 | fructose 2,6-bisphosphate metabolic process | "The chemical reactions and pathways involving fructose 2,6-bisphosphate" |
| GO:0006110 | 0.0211 | 1 / 1 | regulation of glycolytic process | "Any process that modulates the frequency, rate or extent of glycolysis" |
| GO:0006508 | 0.0211 | 2 / 22 | proteolysis | "The hydrolysis of proteins into smaller polypeptides and/or amino acids by cleavage" |
| GO:0007039 | 0.0211 | 1 / 4 | protein catabolic process in the vacuole | "The chemical reactions and pathways resulting in the breakdown of proteins in the vacuole" |
| GO:0007049 | 0.0417 | 1 / 4 | cell cycle | "The progression of biochemical and morphological phases and events that occur in the life of a cell" |
| GO:0008150 | 0.0417 | 7 / 517 | biological_process | "Any process specifically pertinent to the functioning of integrated living organisms" |
| Numer klastra: 4 | | | | |
| GO Term | p-val | clusterNo/chipNo | geneName | definition |
| GO:0006487 | 0.0028 | 1 / 1 | protein N-linked glycosylation | "A protein glycosylation process in which a carbohydrate or carbohydrate derivative is attached to a protein by an N-glycosidic bond" |
| GO:0006508 | 0.0220 | 1 / 22 | proteolysis | "The hydrolysis of proteins into smaller polypeptides and/or amino acids by cleavage" |
| GO:0006620 | 0.0220 | 1 / 1 | posttranslational protein targeting to membrane | "The targeting of proteins to a membrane through a process that occurs after translation" |
| GO:0006725 | 0.0220 | 1 / 2 | cellular aromatic compound metabolic process | "The chemical reactions and pathways involving aromatic compounds" |
| GO:0006810 | 0.0220 | 1 / 27 | transport | "The directed movement of substances (such as macromolecules, small molecules, ions, or groups of molecules) from one location to another" |
| GO:0008150 | 0.0220 | 10 / 517 | biological_process | "Any process specifically pertinent to the functioning of integrated living organisms" |
| GO:0008361 | 0.0220 | 1 / 2 | regulation of cell size | "Any process that modulates the size of a cell." [GOC:go_curators] |
| GO:0008311 | 0.0435 | 1 / 1 | oligosaccharide metabolic process | "The chemical reactions and pathways involving oligosaccharides" |
| GO:0016050 | 0.0435 | 1 / 6 | vesicle organization | "A process that is carried out at the cellular level which results in the formation of a vesicle" |

length: 22 241 lines: 161 Ln: 1 Col: 1 Sel: 0 0 Unix (LF) ANSI INS

9. Dla metody klastrowania k-means znaleziono pięć najbardziej istotnych statystycznie terminów, za pomocą których wygenerowano podgraf ontologii:



Terminami o najniższej p-wartości okazały się:

1)

```
id: 6487
name: 'protein N-linked glycosylation'
ontology: 'biological process'
definition: '"A protein glycosylation process in which a carbohydrate or carbohydrate ...'
comment: ''
synonym: {3x2 cell}
is_a: 6486
part_of: [0x1 double]
obsolete: 0
```

Rodzic: GO: 6486, Name: protein glycosylation

2)

```
id: 6515
name: 'misfolded or incompletely synthesized protein catabolic process'
ontology: 'biological process'
definition: '"The chemical reactions and pathways resulting in the breakdown of misfol...'
comment: ''
synonym: {4x2 cell}
is_a: 51603
part_of: [0x1 double]
obsolete: 0
```

Rodzig: GO: 51603, Name: proteolysis involved in cellular protein catabolic process

3)

```
id: 6412
name: 'translation'
ontology: 'biological process'
definition: '"The cellular metabolic process in which a protein is formed, using the s..."'
comment: ''
synonym: {9x2 cell}
is_a: [3x1 double]
part_of: 10467
obsolete: 0
```

Rodzice: GO: 34645, Name: cellular macromolecule biosynthetic process
GO: 43043, Name: peptide biosynthetic process

GO: 44267, Name: cellular protein metabolic process

4)

```
id: 6508
name: 'proteolysis'
ontology: 'biological process'
definition: '"The hydrolysis of proteins into smaller polypeptides and/or amino acids ...'
comment: 'This term was intentionally placed under 'protein metabolic process ; GO:...'
synonym: {2x2 cell}
is_a: 19538
part_of: [0x1 double]
obsolete: 0
```

Rodzic: GO: 19538, Name: protein metabolic process

5)

```
id: 6810
name: 'transport'
ontology: 'biological process'
definition: '"The directed movement of substances (such as macromolecules, small molec...'
comment: ''
synonym: {6x2 cell}
is_a: 51234
part_of: [0x1 double]
obsolete: 0
```

Rodzic: GO: 51234, Name: establishment of localization

Otrzymane wyniki pozwalają oszacować, które z procesów komórkowych okazały się istotnymi podczas przeprowadzania eksperymentu na mikromacierzy. Są to procesy dotyczące tworzenia się pewnych białek (translation), ich modyfikacji (protein glycosylation), transportu (transport) oraz degradacji (proteolysis).

10. Porównując wyniki z różnych metod klastrowania z wynikami otrzymanymi przez Eisen, widzimy, że te drugie warunkują się większym statystycznym znaczeniem. Możemy to stwierdzić na podstawie wartości P-value.

11. Na końcową listę otrzymanych terminów GO mają wpływ czynniki takie jak regulacja genów kodujących białka rybosomalne, wiemy również, że geny syntezy białek mitochondrialnych uległy ekspresji w połączeniu z pewną liczbą genów biorących udział w oddychaniu co na pewno ma wpływ na otrzymaną listę. Ponadto, istnieje wiele przykładów koekspresji genów, które mają wspólną funkcję, ale nie należą do ogromnego białkowego kompleksu kodującego enzymy glikolityczne. Czynniki te działają głównie na poziomie funkcji komórkowej.