Multidimensional gene expression

Mikael Kubista

Leipzig, Germany

TATAA Biocenter

Clustering methods

- Genes/samples that behave similarly are arranged in clusters.

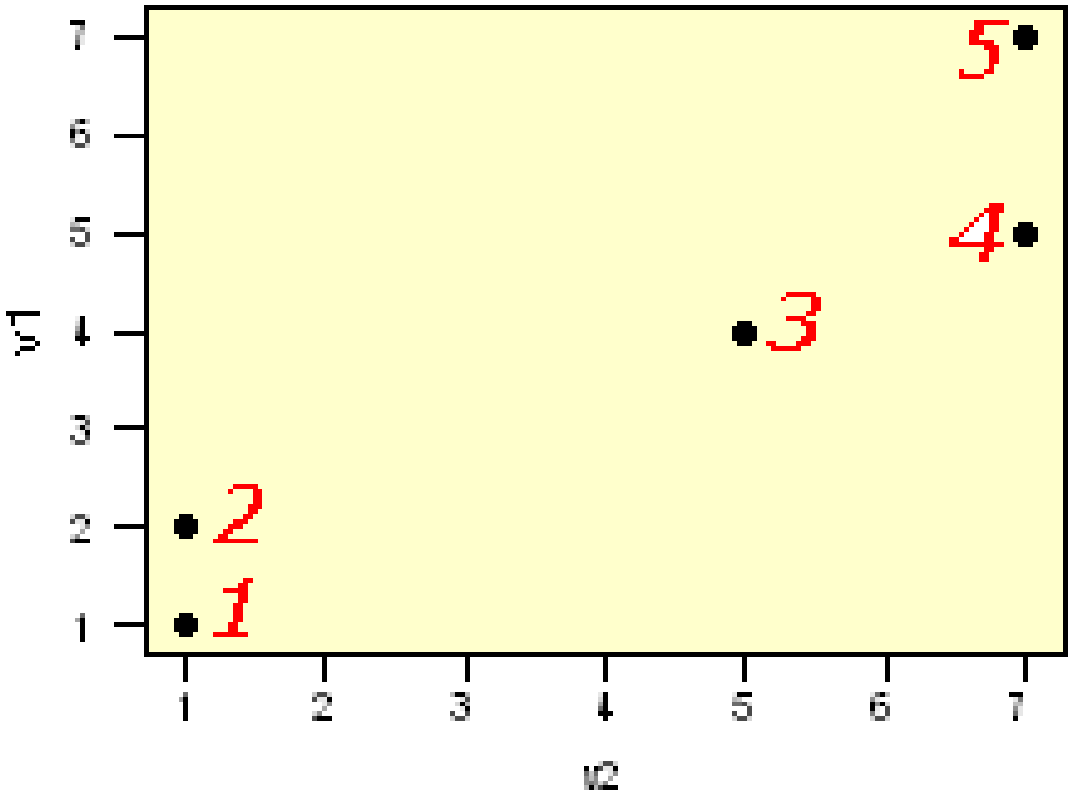
  - » Principle component analysis
  - » Support vector machines
  - » Hierarchical clustering
  - » K-means
  - » Procrustes rotation

Classification methods

- Training data set available with classifiers to train predictive algorithm
  - » Neural networks
  - » Partial Least Square (PLS) analysis
  - » Machine learning algorithms

How to define similarity?

case	v1	v2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7


y1
m2
1
2
3
4
5
6
7

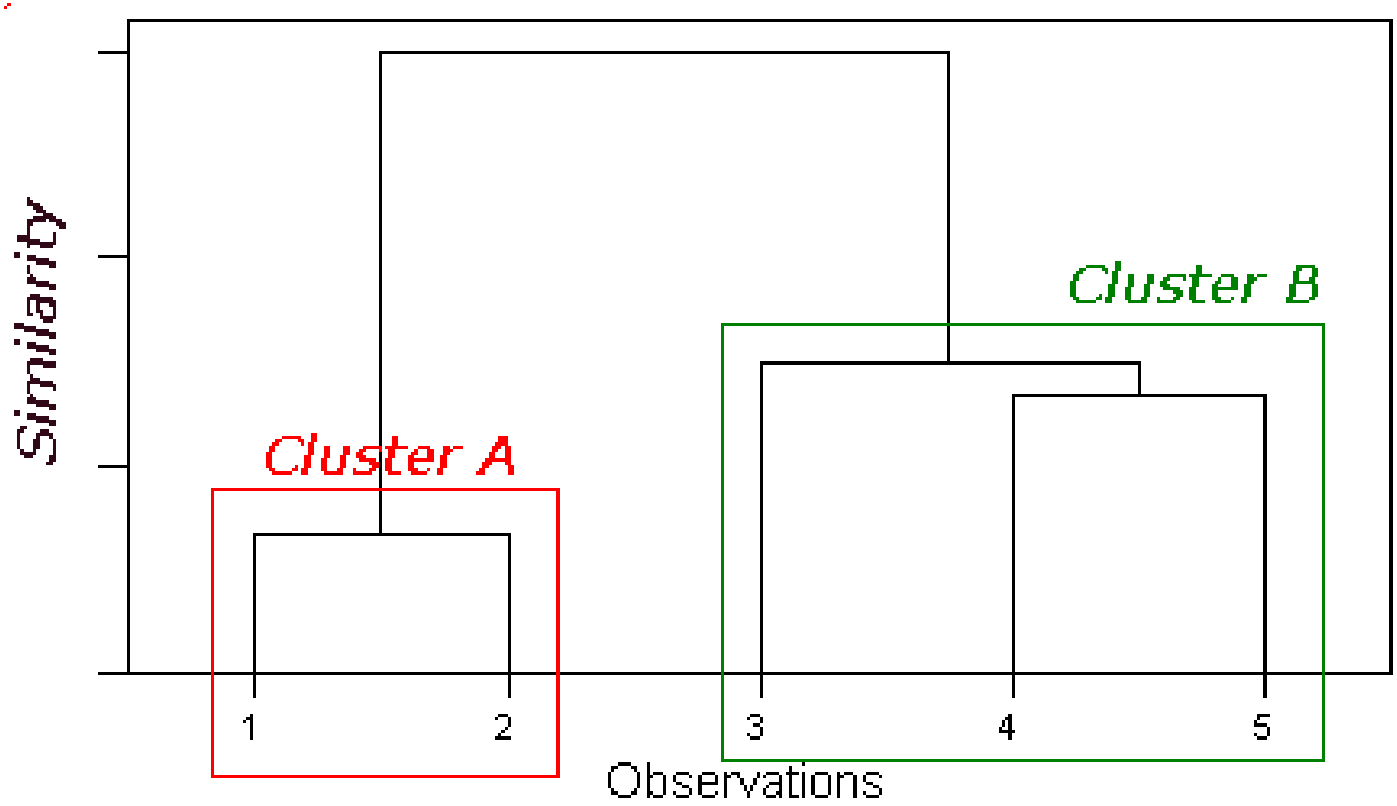
	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

	1	2	3	4	5
1	0.0				
2	1.0	0.0			
3	5.0	4.5	0.0		
4	8.5	7.8	3.6	0.0	
5	7.2	6.7	2.2	2.0	0.0

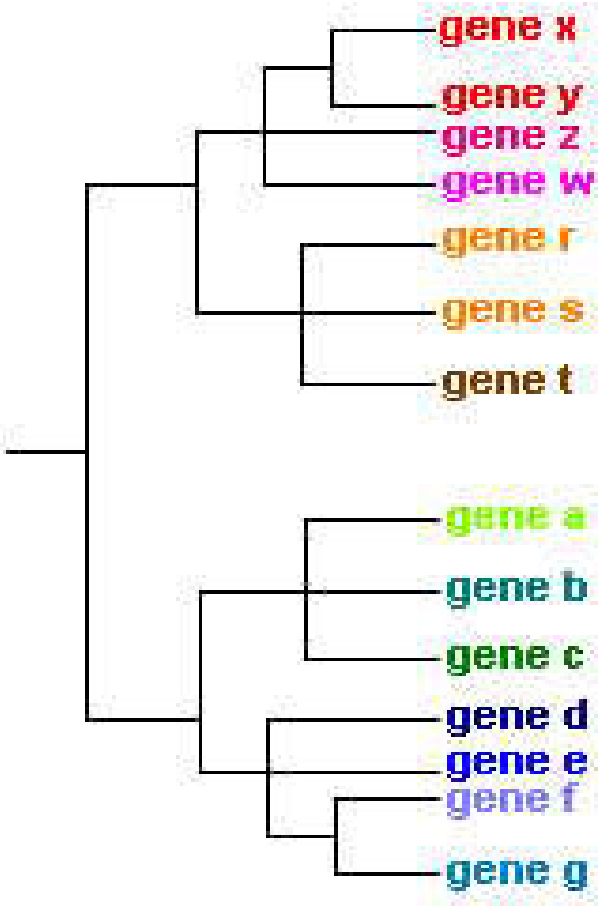
	A	3	4	5
A	0.0			
3	4.7	0.0		
4	8.1	3.6	0.0	
5	6.9	2.2	2.0	0.0

	A	B	3
A	0.0		
B	7.5	0.0	
3	4.7	2.8	0.0

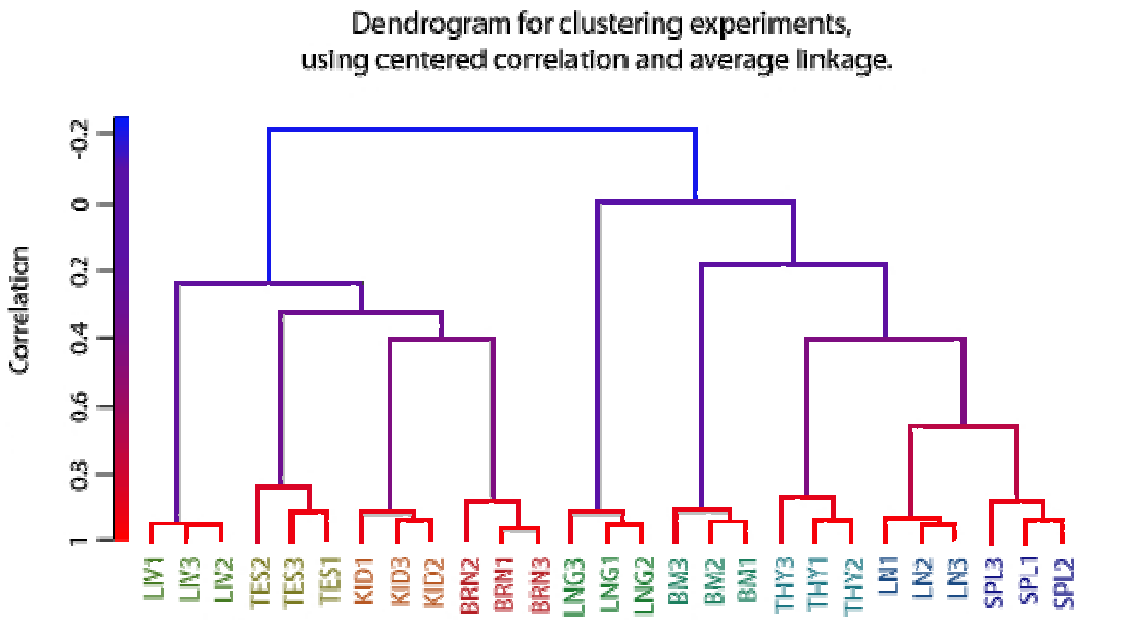
	A	B
A	0.0	
B	6.4	0.0


Similarity
Cluster A
Cluster B
1
2
3
4
5
Observations

Clustering genes

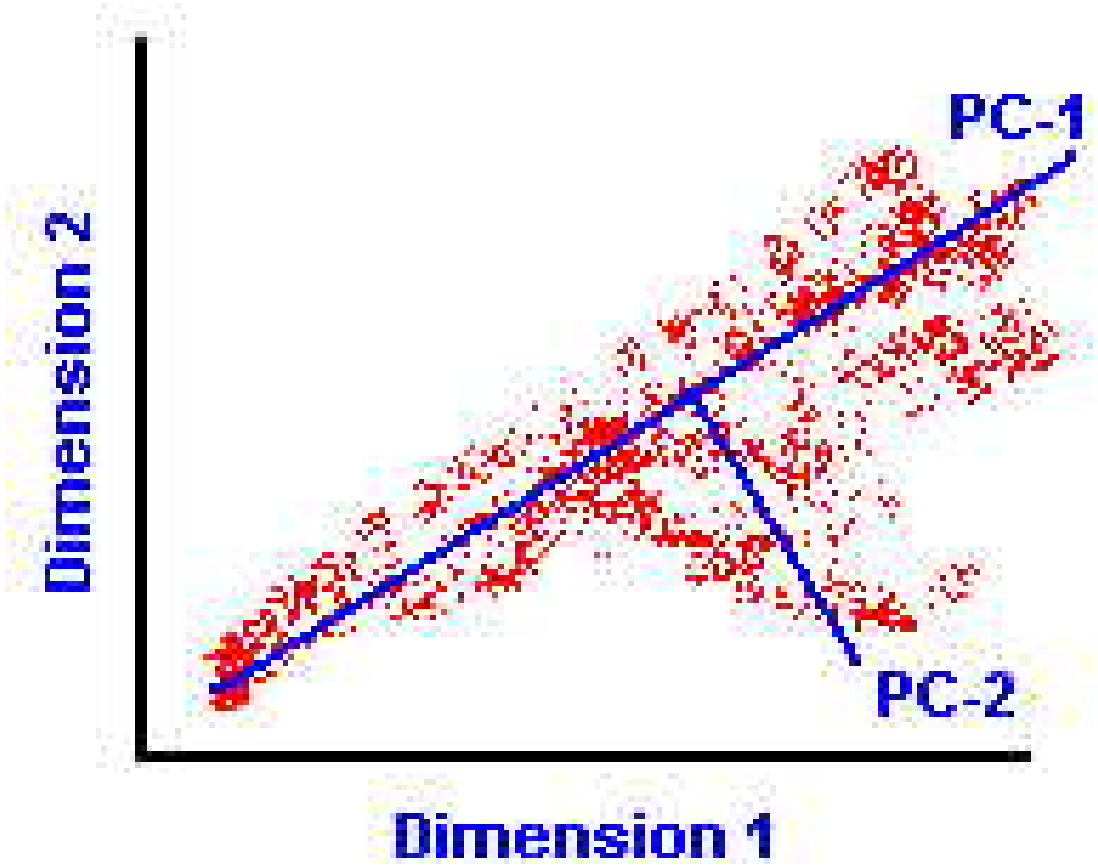


Clustering samples

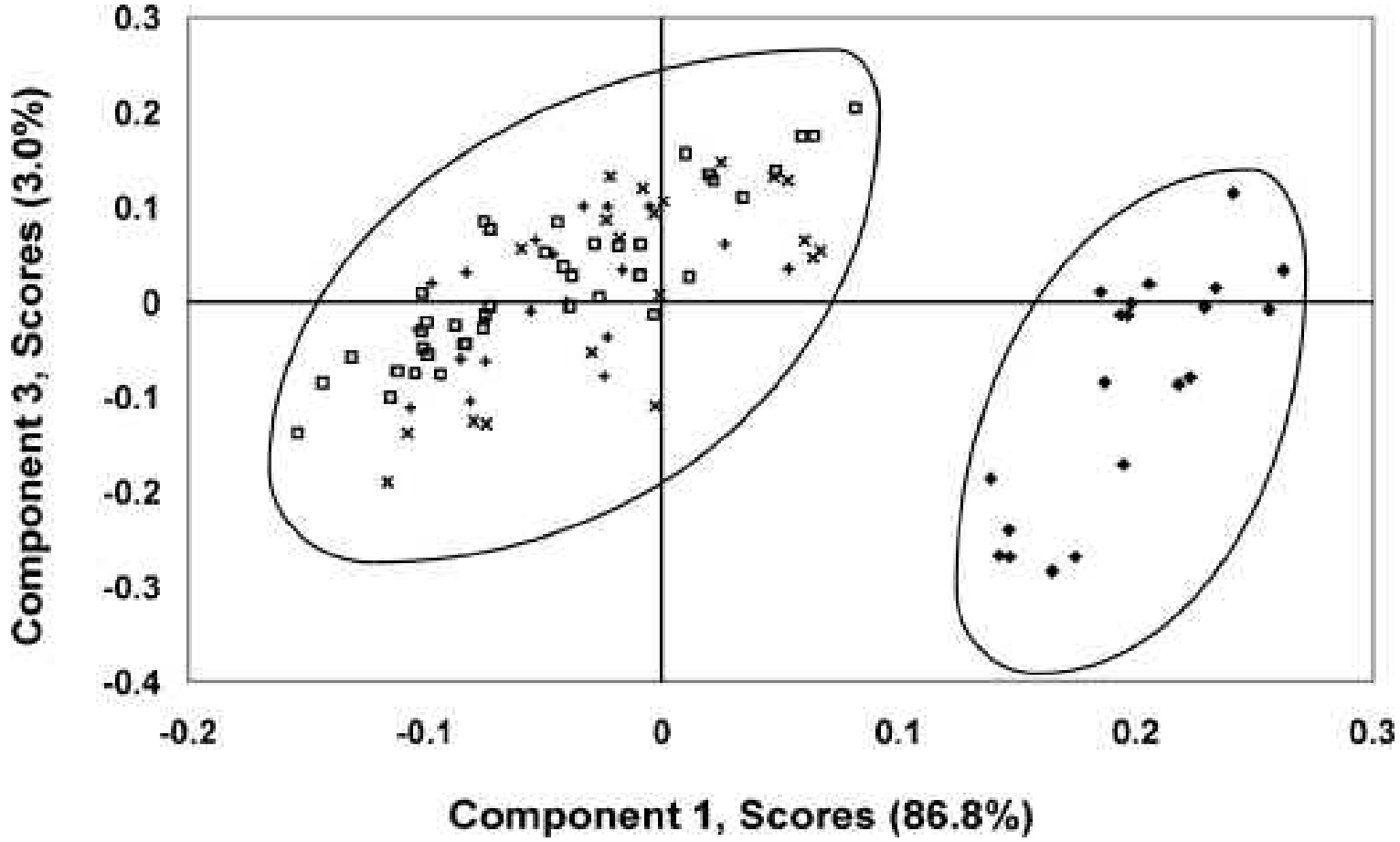


Unsupervised hierarchical clustering of wild type mouse tissue. Cluster for liver (LIV), testes (TES), kidney (KID), brain (BRN), lung (LNG), bone marrow (BM), thymus (THY), lymph node (LN) and spleen (SPL) is shown. Only mouse genes on the GLYCOv1 array were used for clustring.

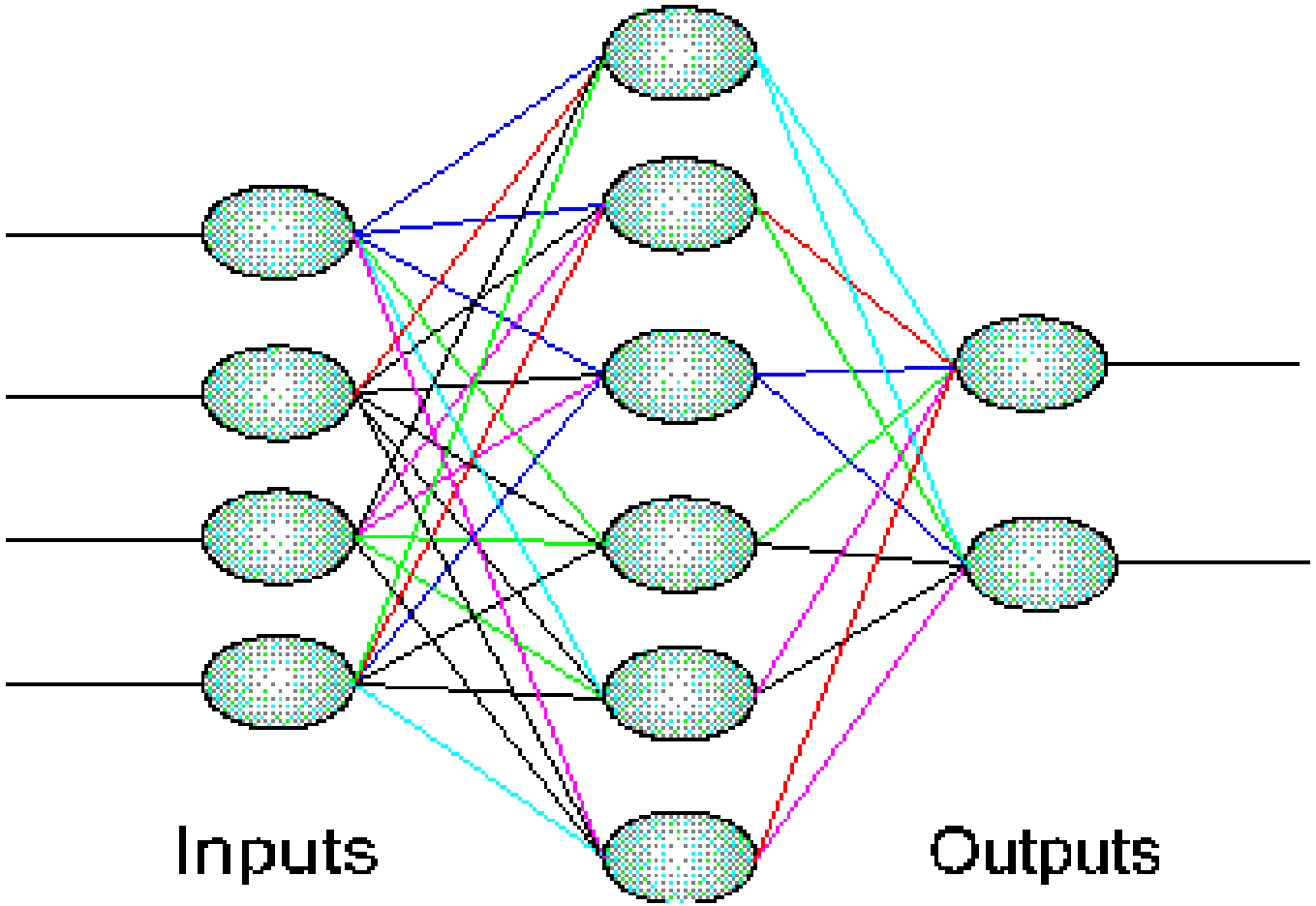
Principle components



Scores plot


0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.2
-0.1
0
0.1
0.2
0.3
Component 3, Scores (3.0%)
Component 1, Scores (86.8%)

Neural network

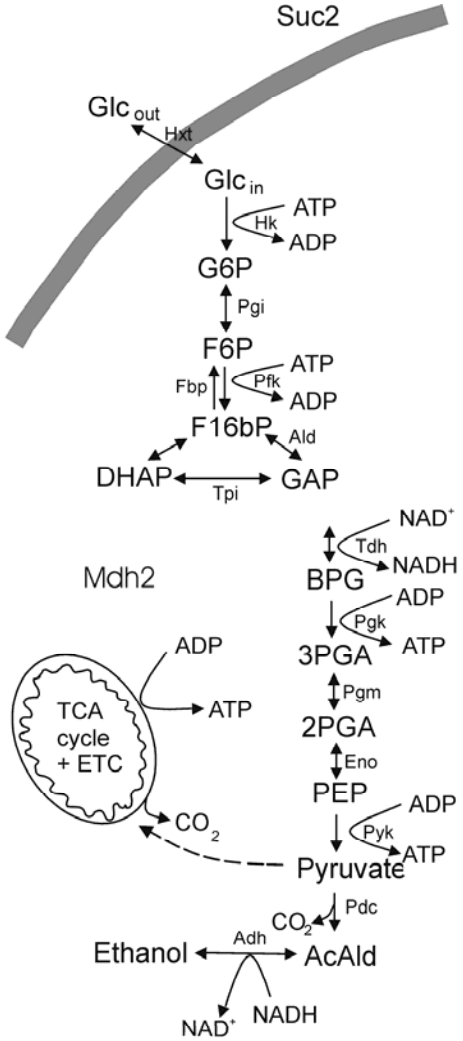


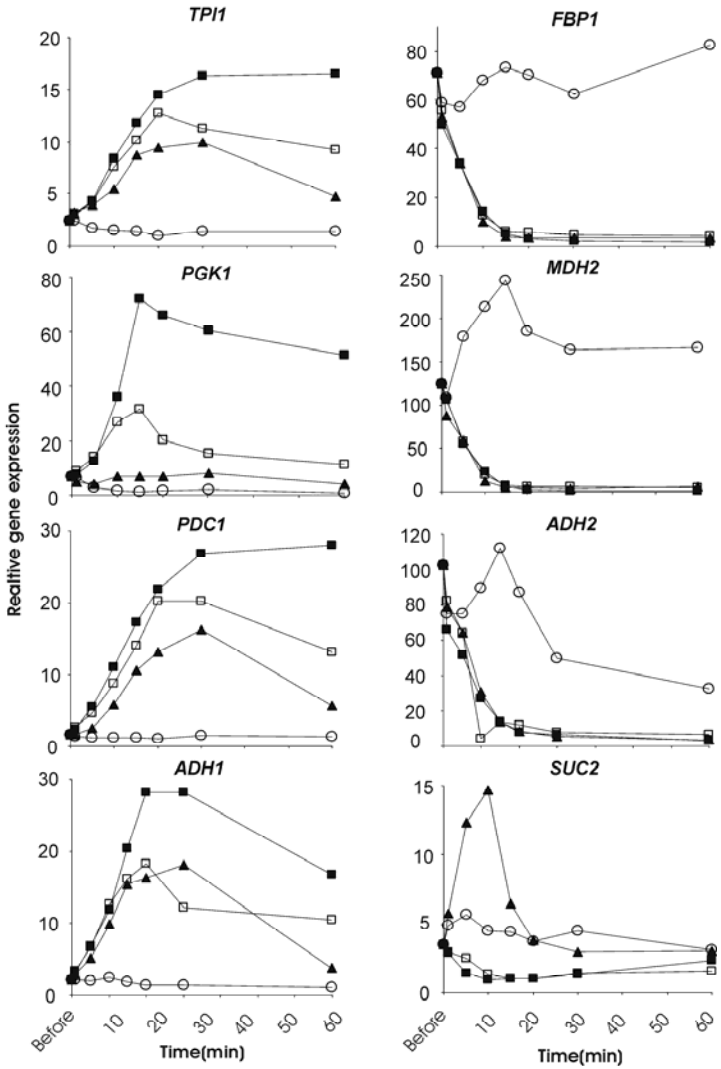
Experimental System

- Four strains of yeast: Wt, Hxt7, Tm6 and Null
- Expression over time after glucose addition: 0 – 60 min
- Expression of genes:
- Genes:
  - » Ref: ACT, IPPI, PDA
  - » Glycolysis: TPI, PGK, PDC, ADH1
  - » Glycogenesis: FBP, MDH2, SUC2, ADH2
  - » Unknown: ADH3, AD H4, ADH5, ADH6
  - » HSP = heat shock
  - » CYC = Cytochrom c

Time dependent gene expression

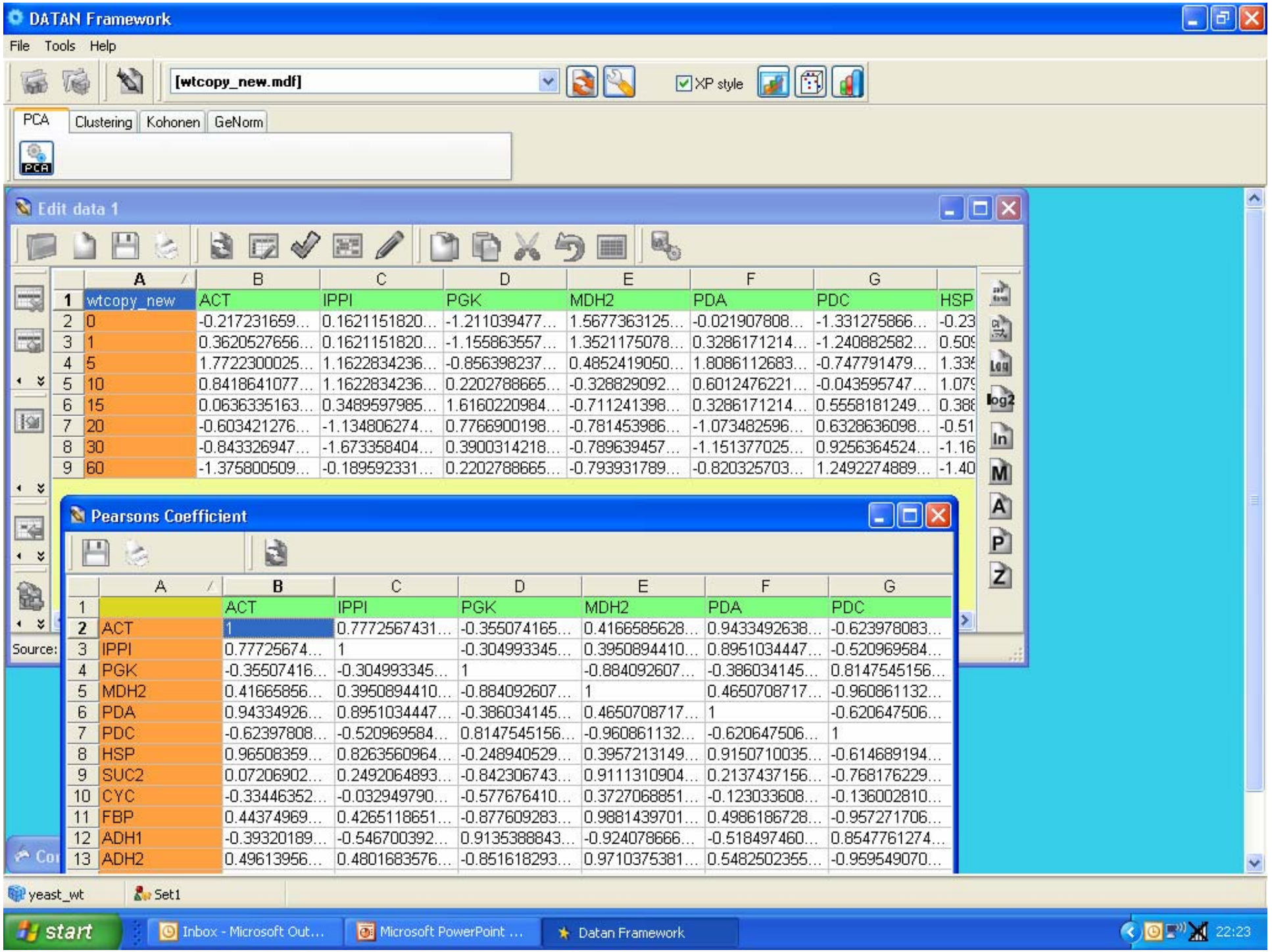
Four different yeast strains with altered glucose uptake





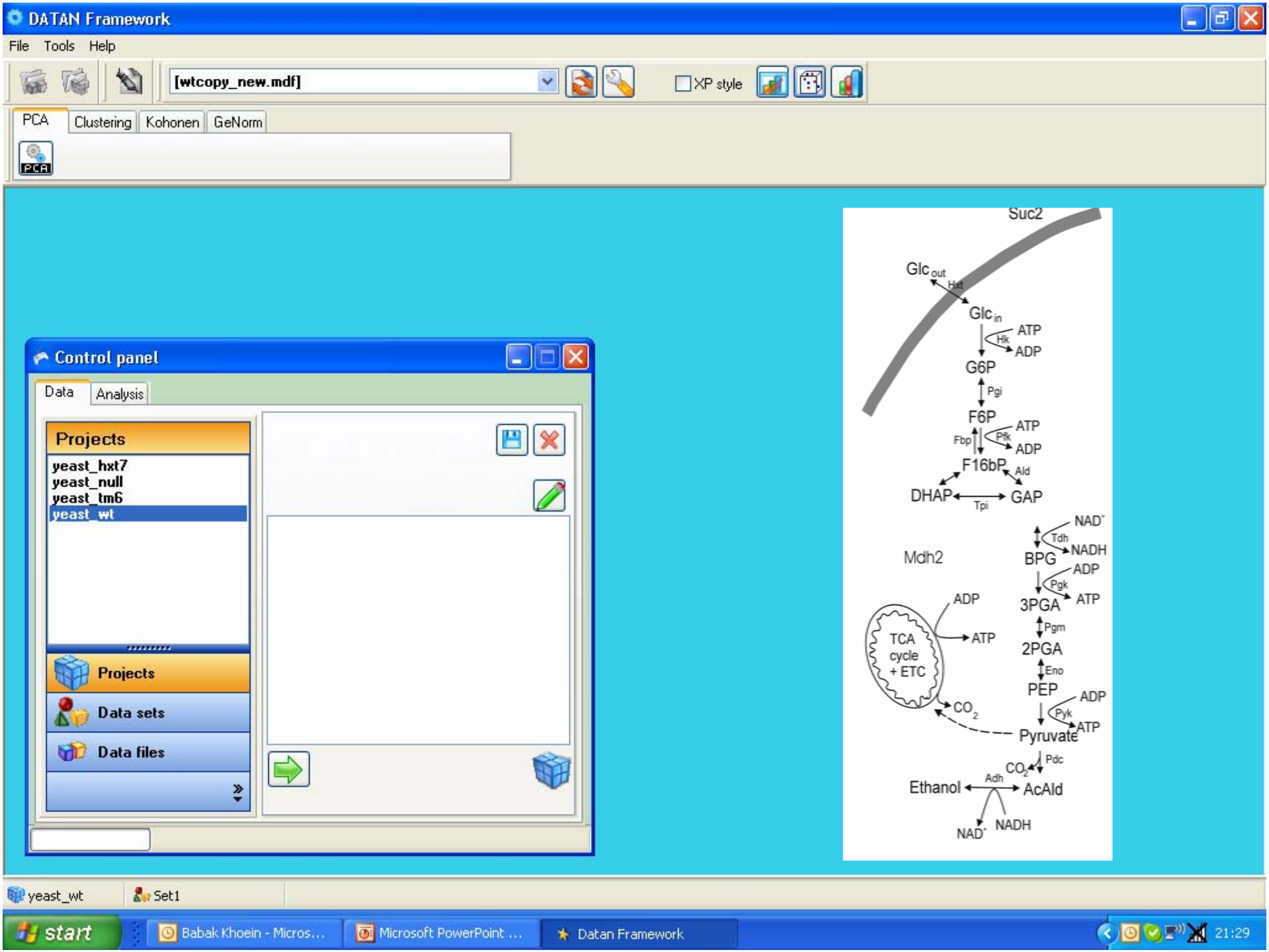
TATAA Biocen

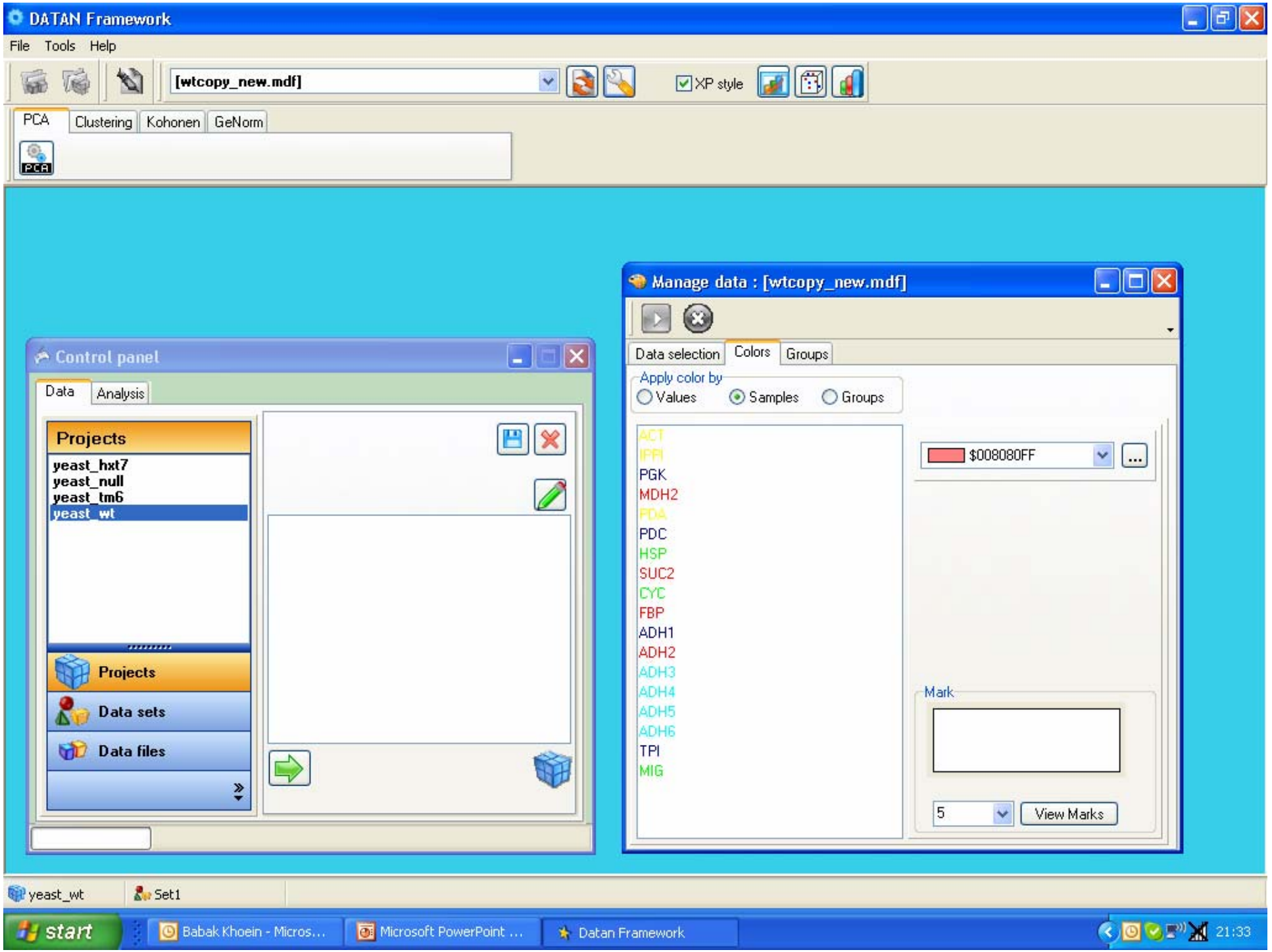

DATAN Framework
File Tools Help
XP style
PCA Clustering Kohonen GeNorm
PCA
Control panel
Data Analysis
Projects
Projects
Data sets
Data files
multid
ANALYSES
start
Inbox - Microsoft Out...
Microsoft PowerPoint ...
Datan Framework
22:26

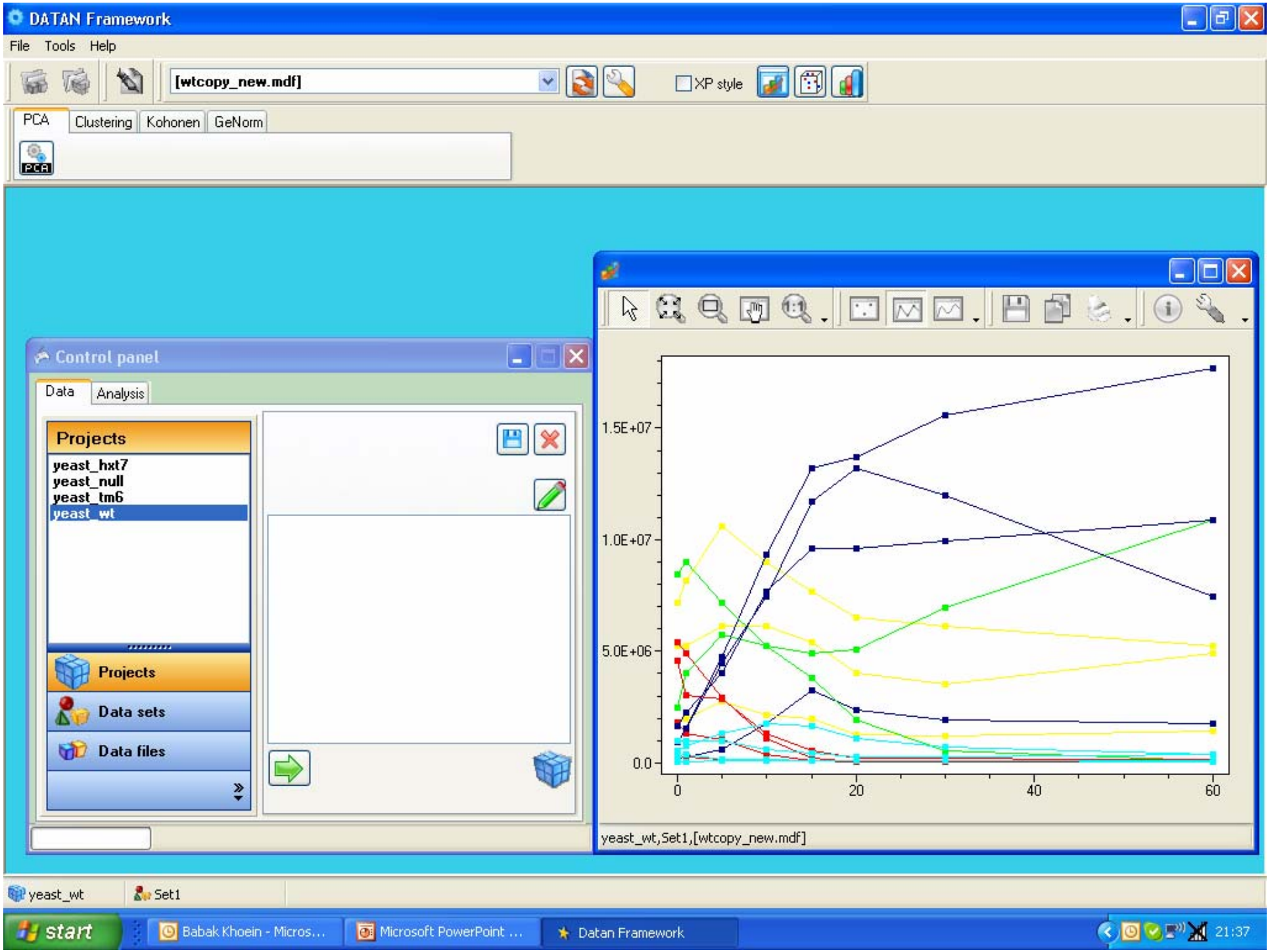


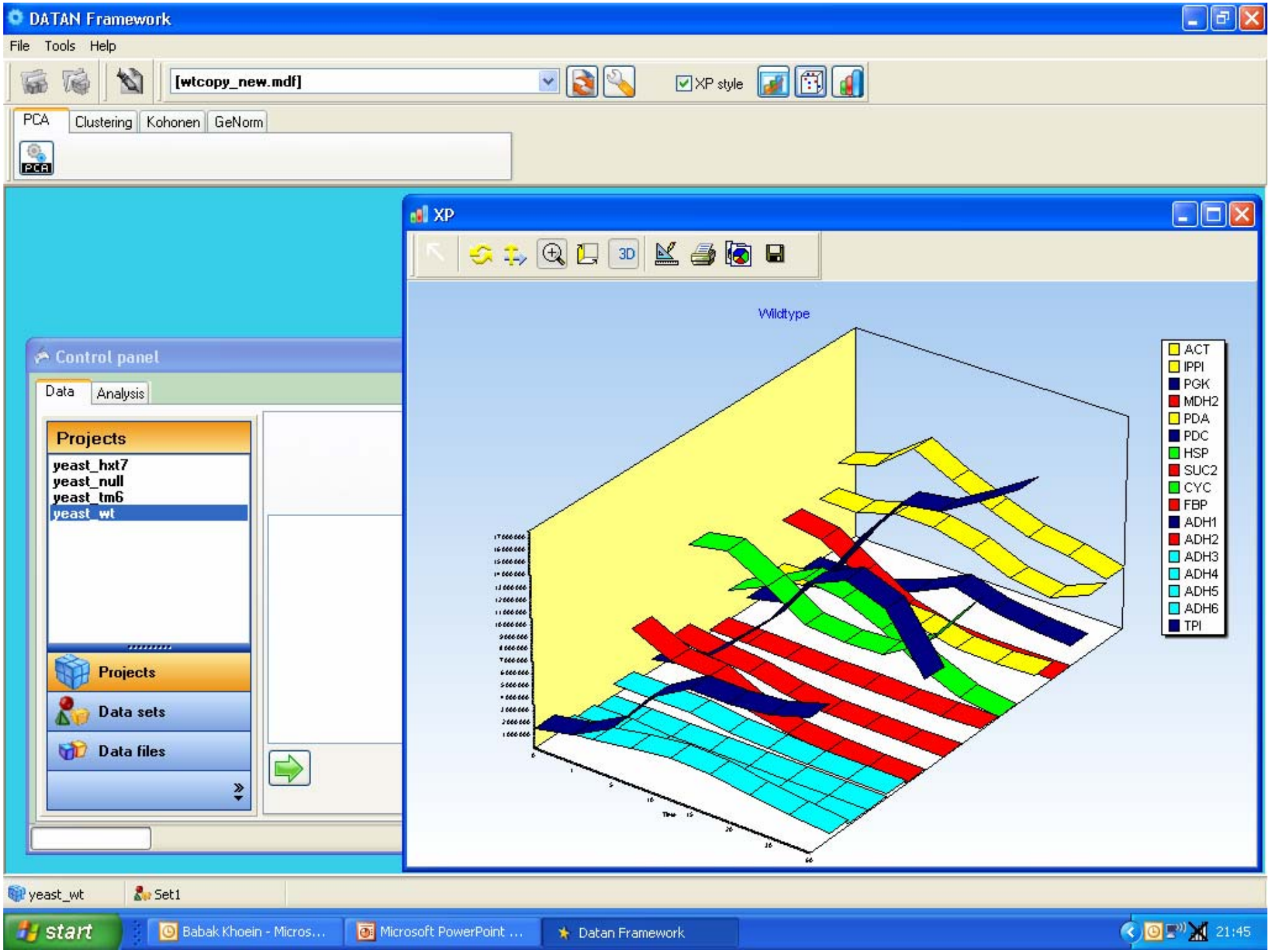
	A	B	C	D	E	F	G	
1	wtcopy_new	ACT	IPPI	PGK	MDH2	PDA	PDC	HSP
2	0	-0.217231659...	0.1621151820...	-1.211039477...	1.5677363125...	-0.021907808...	-1.331275866...	-0.23
3	1	0.3620527656...	0.1621151820...	-1.155863557...	1.3521175078...	0.3286171214...	-1.240882582...	0.509
4	5	1.7722300025...	1.1622834236...	-0.856398237...	0.4852419050...	1.8086112683...	-0.747791479...	1.335
5	10	0.8418641077...	1.1622834236...	0.2202788665...	-0.328829092...	0.6012476221...	-0.043595747...	1.079
6	15	0.0636335163...	0.3489597985...	1.6160220984...	-0.711241398...	0.3286171214...	0.5558181249...	0.388
7	20	-0.603421276...	-1.134806274...	0.7766900198...	-0.781453986...	-1.073482596...	0.6328636098...	-0.51
8	30	-0.843326947...	-1.673358404...	0.3900314218...	-0.789639457...	-1.151377025...	0.9256364524...	-1.16
9	60	-1.375800509...	-0.189592331...	0.2202788665...	-0.793931789...	-0.820325703...	1.2492274889...	-1.40

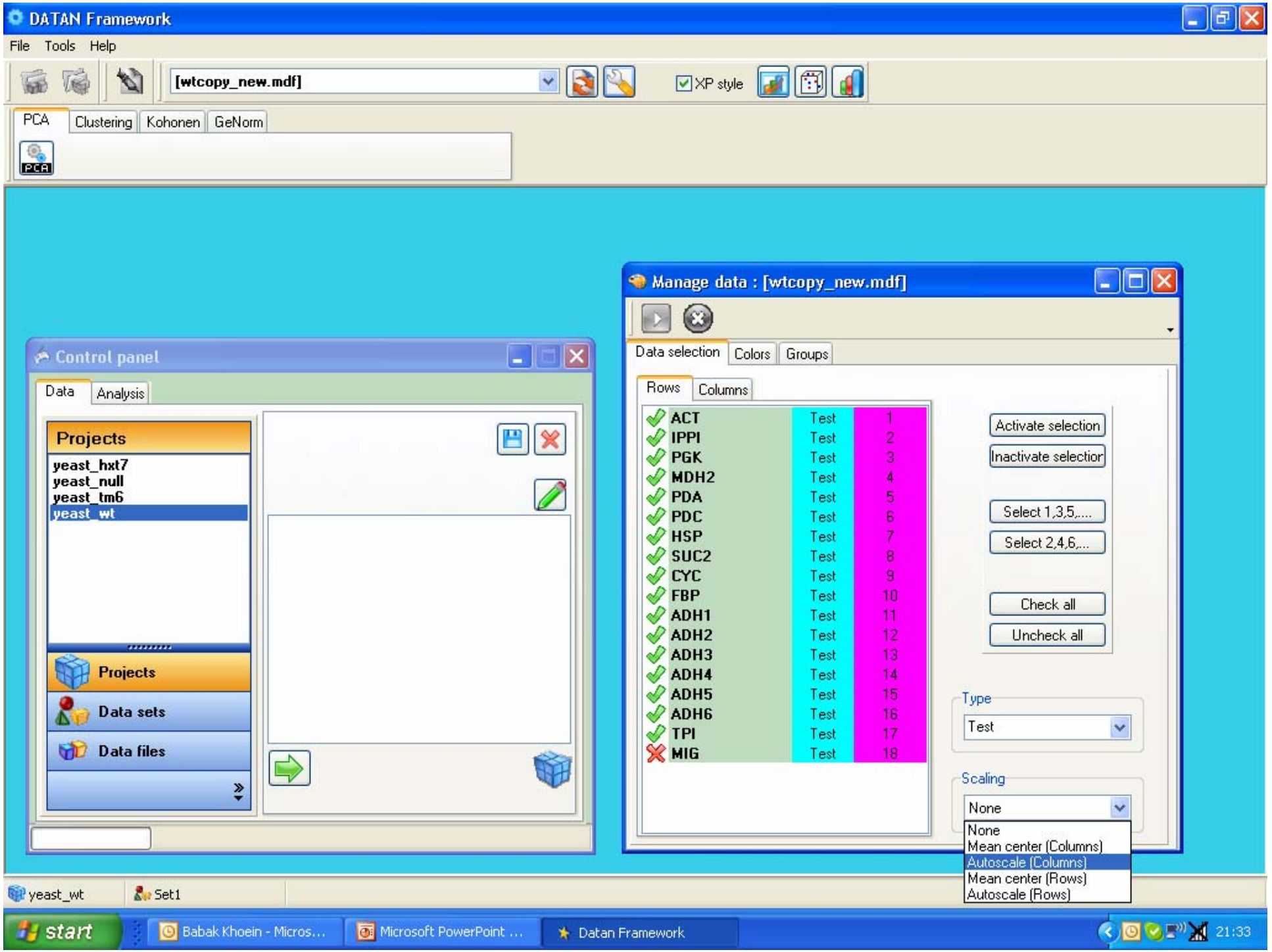
	A	B	C	D	E	F	G
1		ACT	IPPI	PGK	MDH2	PDA	PDC
2	ACT	1	0.7772567431...	-0.355074165...	0.4166585628...	0.9433492638...	-0.623978083...
3	IPPI	0.77725674...	1	-0.304993345...	0.3950894410...	0.8951034447...	-0.520969584...
4	PGK	-0.35507416...	-0.304993345...	1	-0.884092607...	-0.386034145...	0.8147545156...
5	MDH2	0.41665856...	0.3950894410...	-0.884092607...	1	0.4650708717...	-0.960861132...
6	PDA	0.94334926...	0.8951034447...	-0.386034145...	0.4650708717...	1	-0.620647506...
7	PDC	-0.62397808...	-0.520969584...	0.8147545156...	-0.960861132...	-0.620647506...	1
8	HSP	0.96508359...	0.8263560964...	-0.248940529...	0.3957213149...	0.9150710035...	-0.614689194...
9	SUC2	0.07206902...	0.2492064893...	-0.842306743...	0.9111310904...	0.2137437156...	-0.768176229...
10	CYC	-0.33446352...	-0.032949790...	-0.577676410...	0.3727068851...	-0.123033608...	-0.136002810...
11	FBP	0.44374969...	0.4265118651...	-0.877609283...	0.9881439701...	0.4986186728...	-0.957271706...
12	ADH1	-0.39320189...	-0.546700392...	0.9135388843...	-0.924078666...	-0.518497460...	0.8547761274...
13	ADH2	0.49613956...	0.4801683576...	-0.851618293...	0.9710375381...	0.5482502355...	-0.959549070...

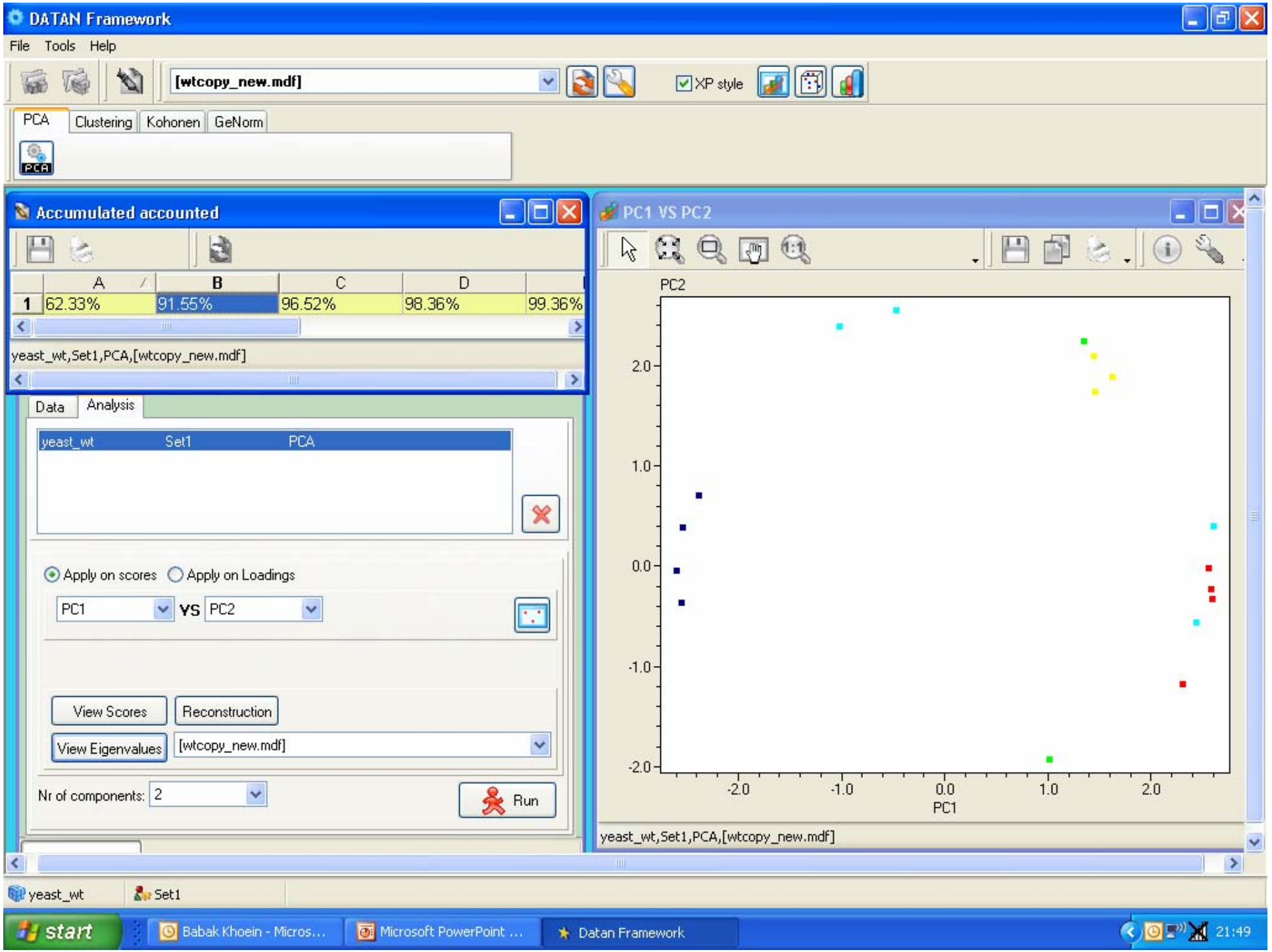

DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
PCA
Control panel
Data Analysis
Projects
yeast_hxt7
yeast_null
yeast_tm6
yeast_wt
Projects
Data sets
Data files
Suc2
Glc out
Hxt
Glc in
ATP
Hk
ADP
G6P
Pgi
F6P
ATP
Fbp
Pfk
ADP
F16bP
Ald
DHAP
Tpi
GAP
NAD+
Tdh
NADH
Mdh2
BPG
ADP
Pgk
ATP
3PGA
ADP
TCA cycle + ETC
ATP
Pgm
2PGA
Eno
PEP
ADP
CO2
Pyk
ATP
Pyruvate
Pdc
CO2
Adh
Ethanol
AcAld
NAD+
NADH
yeast_wt
Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:29

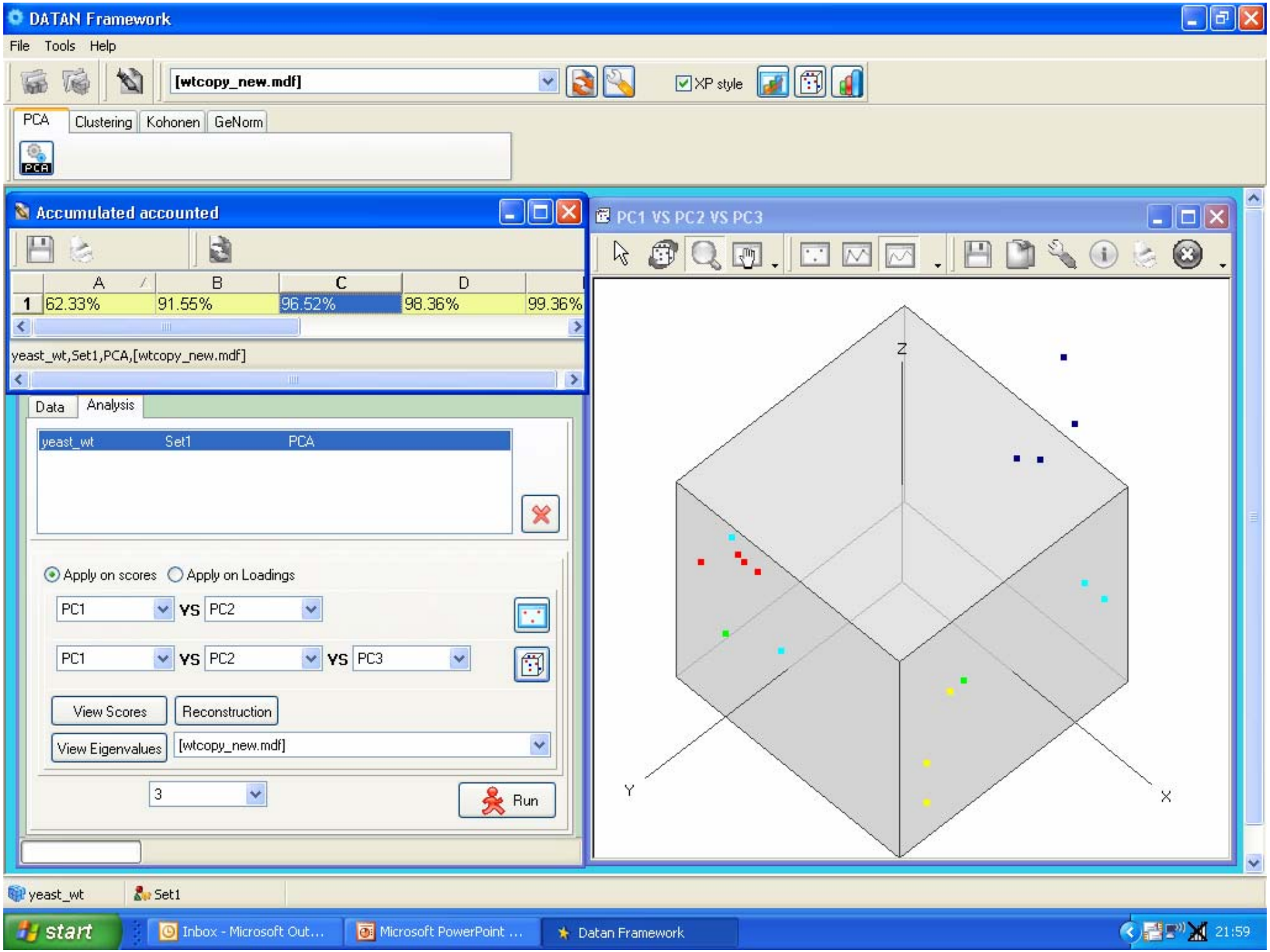

DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
PCA
Control panel
Data Analysis
Projects
yeast_hxt7
yeast_null
yeast_tm6
yeast_wt
Projects
Data sets
Data files
Manage data : [wtcopy_new.mdf]
Data selection Colors Groups
Apply color by
Values Samples Groups
ACT
IPPI
PGK
MDH2
PDA
PDC
HSP
SUC2
CYC
FBP
ADH1
ADH2
ADH3
ADH4
ADH5
ADH6
TPI
MIG
$008080FF
Mark
5
View Marks
yeast_wt
Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:33

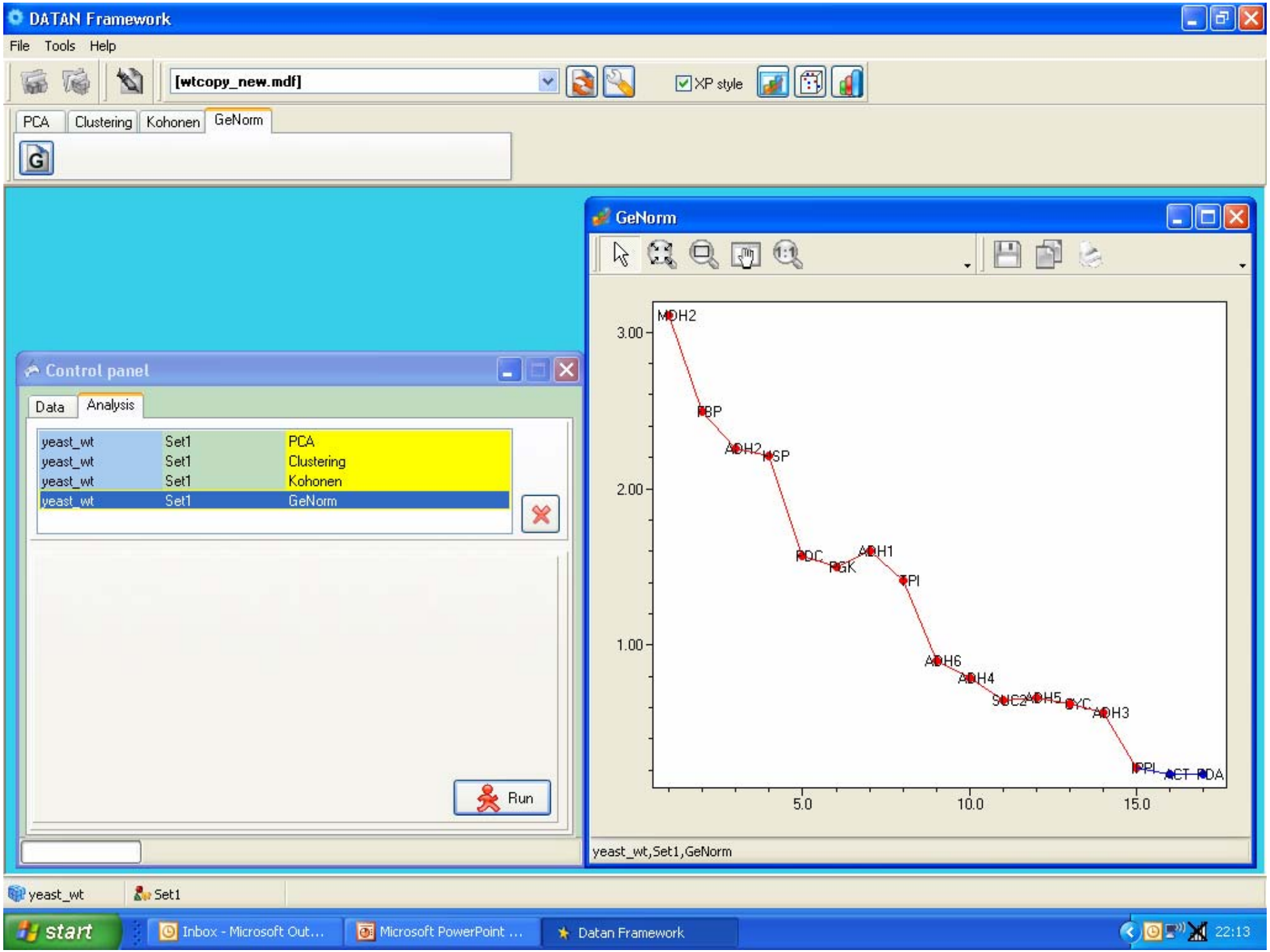

DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Control panel
Data Analysis
Projects
yeast_hxt7
yeast_null
yeast_tm6
yeast_wt
Projects
Data sets
Data files
1.5E+07
1.0E+07
5.0E+06
0.0
0
20
40
60
yeast_wt,Set1,[wtcopy_new.mdf]
yeast_wt
Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:37

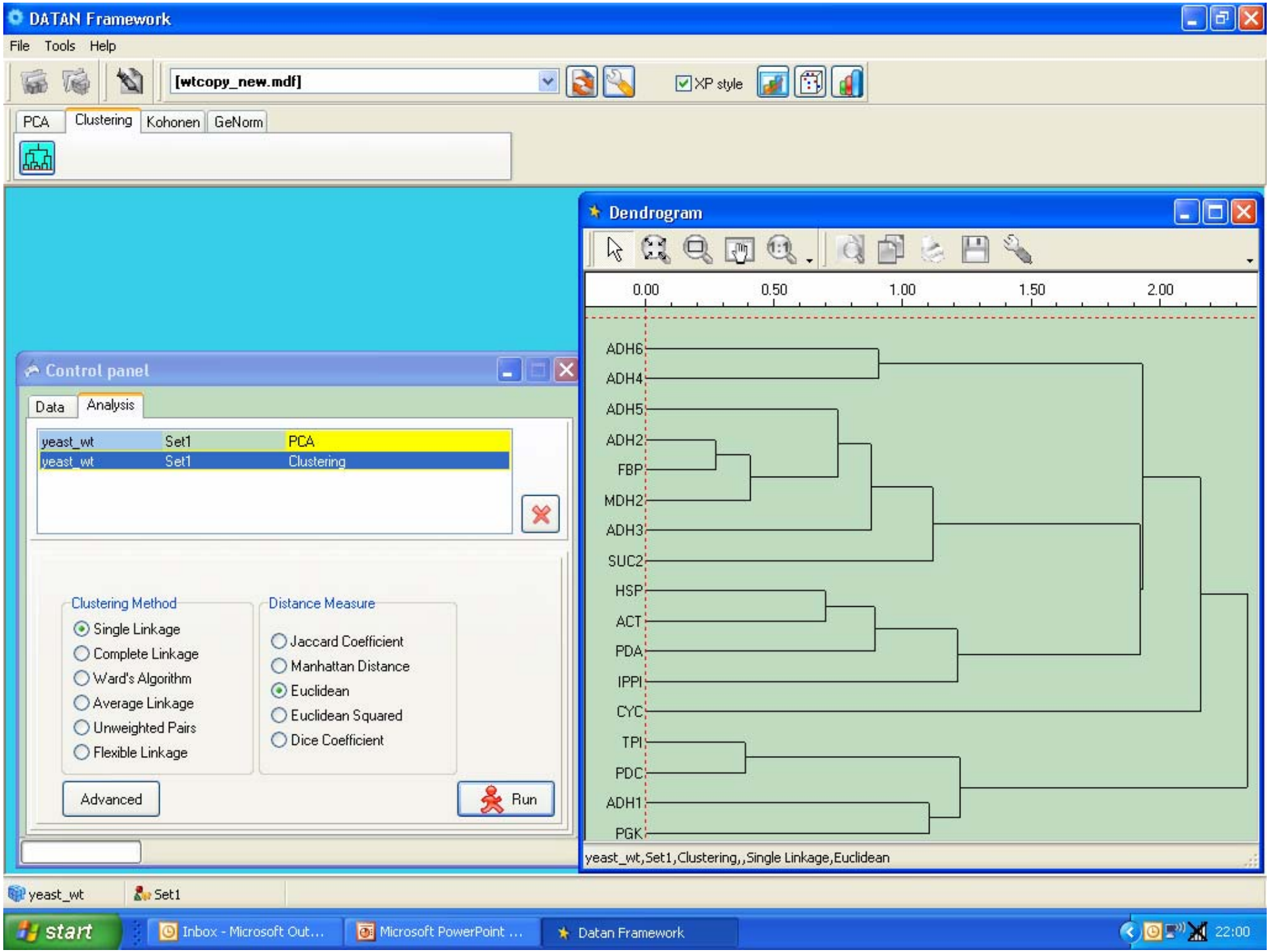

DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
PCA
XP
3D
Wildtype
ACT
IPPI
PGK
MDH2
PDA
PDC
HSP
SUC2
CYC
FBP
ADH1
ADH2
ADH3
ADH4
ADH5
ADH6
TPI
Control panel
Data Analysis
Projects
yeast_hxt7
yeast_null
yeast_tm6
yeast_wt
Projects
Data sets
Data files
yeast_wt
Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:45

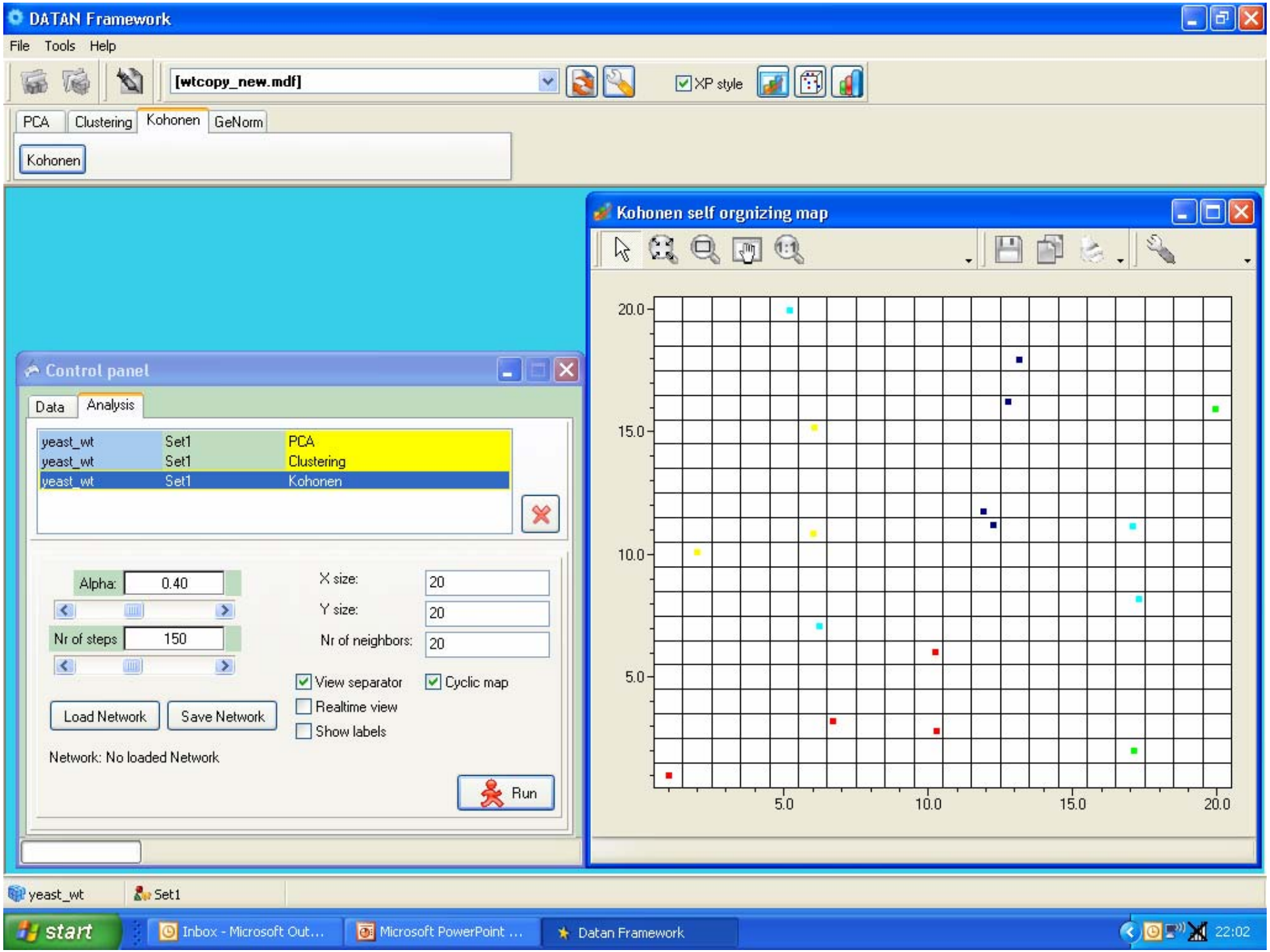

DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Control panel
Data Analysis
Projects
yeast_hxt7
yeast_null
yeast_tm6
yeast_wt
Projects
Data sets
Data files
Manage data : [wtcopy_new.mdf]
Data selection Colors Groups
Rows Columns
ACT Test 1
IPPI Test 2
PGK Test 3
MDH2 Test 4
PDA Test 5
PDC Test 6
HSP Test 7
SUC2 Test 8
CYC Test 9
FBP Test 10
ADH1 Test 11
ADH2 Test 12
ADH3 Test 13
ADH4 Test 14
ADH5 Test 15
ADH6 Test 16
TPI Test 17
MIG Test 18
Activate selection
Inactivate selection
Select 1,3,5,...
Select 2,4,6,...
Check all
Uncheck all
Type
Test
Scaling
None
None
Mean center (Columns)
Autoscale (Columns)
Mean center (Rows)
Autoscale (Rows)
yeast_wt
Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:33


DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Accumulated accounted
A B C D
1 62.33% 91.55% 96.52% 98.36% 99.36%
yeast_wt,Set1,PCA,[wtcopy_new.mdf]
Data Analysis
yeast_wt Set1 PCA
Apply on scores Apply on Loadings
PC1 VS PC2
PC1 YS PC2
View Scores Reconstruction
View Eigenvalues [wtcopy_new.mdf]
Nr of components: 2
Run
PC2
PC1
2.0
1.0
0.0
-1.0
-2.0
yeast_wt,Set1,PCA,[wtcopy_new.mdf]
yeast_wt Set1
start
Babak Khoein - Micros...
Microsoft PowerPoint ...
Datan Framework
21:49


DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Accumulated accounted
A B C D
1 62.33% 91.55% 96.52% 98.36% 99.36%
yeast_wt,Set1,PCA,[wtcopy_new.mdf]
Data Analysis
yeast_wt Set1 PCA
Apply on scores Apply on Loadings
PC1 VS PC2
PC1 VS PC2 VS PC3
View Scores Reconstruction
View Eigenvalues [wtcopy_new.mdf]
3 Run
PC1 VS PC2 VS PC3
Z
Y
X
yeast_wt Set1
start Inbox - Microsoft Out... Microsoft PowerPoint ... Datan Framework 21:59


DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Control panel
Data Analysis
yeast_wt Set1 PCA
yeast_wt Set1 Clustering
yeast_wt Set1 Kohonen
yeast_wt Set1 GeNorm
Run
GeNorm
MDH2
BP
ADH2
SP
PDC
PGK
ADH1
PI
ADH6
ADH4
SUC2
ADH5
CYC
ADH3
PPI
ACT
TDA
3.00
2.00
1.00
5.0
10.0
15.0
yeast_wt,Set1,GeNorm
yeast_wt
Set1
start
Inbox - Microsoft Out...
Microsoft PowerPoint ...
Datan Framework
22:13


DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Dendrogram
0.00 0.50 1.00 1.50 2.00
ADH6
ADH4
ADH5
ADH2
FBP
MDH2
ADH3
SUC2
HSP
ACT
PDA
IPPI
CYC
TPI
PDC
ADH1
PGK
yeast_wt,Set1,Clustering,,Single Linkage,Euclidean
Control panel
Data Analysis
yeast_wt Set1 PCA
yeast_wt Set1 Clustering
Clustering Method
Single Linkage
Complete Linkage
Ward's Algorithm
Average Linkage
Unweighted Pairs
Flexible Linkage
Distance Measure
Jaccard Coefficient
Manhattan Distance
Euclidean
Euclidean Squared
Dice Coefficient
Advanced
Run
yeast_wt
Set1
start
Inbox - Microsoft Out...
Microsoft PowerPoint ...
Datan Framework
22:00


DATAN Framework
File Tools Help
[wtcopy_new.mdf]
XP style
PCA Clustering Kohonen GeNorm
Kohonen
Kohonen self orgnizing map
20.0
15.0
10.0
5.0
5.0 10.0 15.0 20.0
Control panel
Data Analysis
yeast_wt Set1 PCA
yeast_wt Set1 Clustering
yeast_wt Set1 Kohonen
Alpha: 0.40
Nr of steps 150
X size: 20
Y size: 20
Nr of neighbors: 20
View separator
Cyclic map
Realtime view
Show labels
Load Network
Save Network
Network: No loaded Network
Run
yeast_wt
Set1
start
Inbox - Microsoft Out...
Microsoft PowerPoint ...
Datan Framework
22:02

Acknowledgement

- Amin Forootan, MultiD Analyses AB
- José Manuel Andrade, University of a Coruña
- Anders Stålberg, TATAA Biocenter
- Karin Elbing, Department of Chemistry and Biosciences, Chalmers University

www.multid.se

TATAA Biocenter

multid ANALYSES

Special offer!

Leipzig participants can purchase a pre-release of GenEx for 190 Euro by placing order to:

leipzig@multid.se

The pre-release will be updated without charge to GenEx 1.0, which will be released in summer for an introductory price of 395 Euro.

TATAA Biocenter