

# Web applications

- <https://biit.cs.ut.ee/clustvis/>
  - <https://www.ncbi.nlm.nih.gov/gds/?term=SIN3>
- 
- Data import
  - Data pre-processing
  - PCA
  - Heatmap

# Data import

## Enter data

### Choose data input type:

- ☒ Load sample data
- ☐ Upload file
- ☐ Paste data
- ☐ Import public dataset from ArrayExpress
- ☐ Load saved settings
- ☐ Import prepared gene expression matrix

### Choose dataset:

- ☒ NKI breast cancer dataset (PAM50 genes)
- ☐ Wisconsin Diagnostic Breast Cancer
- ☐ Iris flowers

### Data matrix reshape:

- ☐ filter columns

[Introduction](#)[Data import](#)[Data pre-processing](#)[PCA](#)[Heatmap](#)[Export](#)[Help](#)[News](#)

Column annotations (7 rows, 337 columns):

	NKI_4	NKI_6	NKI_7	NKI_8	NKI_9	NKI_11	NKI_12	NKI_13	NKI_14	NKI_17	NKI_23	NKI_24	NKI_25
ER status	1	1	0	0	1	1	0	1	1	1	1	0	1
Histological grade	3	2	1	3	3	3	3	3	1	2	2	3	3
BRCA mutation	0	0	0	0	0	0	0	0	0	0	0	0	0
Metastasis	0	0	0	0	0	0	0	0	0	0	0	0	0
Lymph node status	0	0	0	0	0	0	0	0	0	0	0	0	0
Treatment	0	0	0	0	0	0	0	0	0	0	0	0	0
Study	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI	NKI

Numeric data (56 rows, 337 columns):

	NKI_4	NKI_6	NKI_7	NKI_8	NKI_9	NKI_11	NKI_12	NKI_13	NKI_14	NKI_17	NKI_23	NKI_24	NKI_25
--	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------

# Data pre-processing

showed in rows:

99.99

**Maximum percentage of NAs allowed in columns:**

99.99

☐ remove constant columns

☒ row centering

**Row scaling:**

unit variance scaling

**PCA method:**

SVD with imputation

Percentage	10.1%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	14.3%	5.4%	3.6%	1.8%
Removed	no	no	no	no	no	no	no	no	no	no	no	no

Variance explained by principal components (56 components):

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Individual	0.31	0.12	0.07	0.06	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01
Cumulative	0.31	0.43	0.50	0.56	0.59	0.62	0.65	0.67	0.69	0.71	0.73	0.75	0.76	0.78	0.79	0.8

Principal components (337 data points in rows, 56 components in columns):

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
NKI_4	-1.20	-2.38	1.50	1.95	1.48	-5.28	9.85	-4.55	2.32	1.78	-0.23	1.60	-2.52	1.23	-3.10
NKI_6	-4.26	0.04	-1.21	1.04	1.29	-0.78	0.72	1.22	0.42	0.30	-0.85	0.48	-0.65	1.22	-0.24
NKI_7	-4.35	2.73	0.27	-0.81	0.68	0.48	-0.82	2.23	2.01	0.30	-0.02	-0.42	-0.11	-0.75	-0.24
NKI_8	5.89	-1.17	-1.30	-0.54	1.31	0.28	0.02	0.25	0.81	-0.41	-0.99	0.90	0.75	0.48	-0.59
NKI_9	-3.53	-1.36	-2.02	-2.01	1.78	0.38	-1.19	1.03	-0.66	-1.44	-0.04	-0.35	0.98	-0.75	-1.07
NKI_11	-0.62	-3.66	2.29	-1.40	0.57	-1.05	-0.36	0.88	0.89	-0.56	0.11	-0.17	-0.80	1.62	-0.49
NKI_12	7.42	4.60	2.06	1.52	-0.10	0.17	1.87	-1.91	-2.47	-1.34	1.47	-0.78	-0.48	-0.71	0.14
NKI_13	2.37	-2.51	-0.53	-0.66	-1.55	0.07	1.75	1.51	-0.04	-0.86	-0.39	0.53	0.72	-0.39	0.33
NKI_14	-7.88	2.92	0.78	-0.10	0.59	0.45	0.18	0.88	-0.54	0.44	-0.61	0.81	0.27	0.11	1.15

# Heatmap

## Heatmap options

☒ change data options

### Data options

☒ show imputed values

☐ transpose heatmap

### Clustering distance for rows:

correlation

### Clustering method for rows:

average

### Tree ordering for rows:

tightest cluster first

### Number of clusters in rows:

1

### Clustering distance for columns:

correlation

### Clustering method for columns:

average

### Tree ordering for columns:

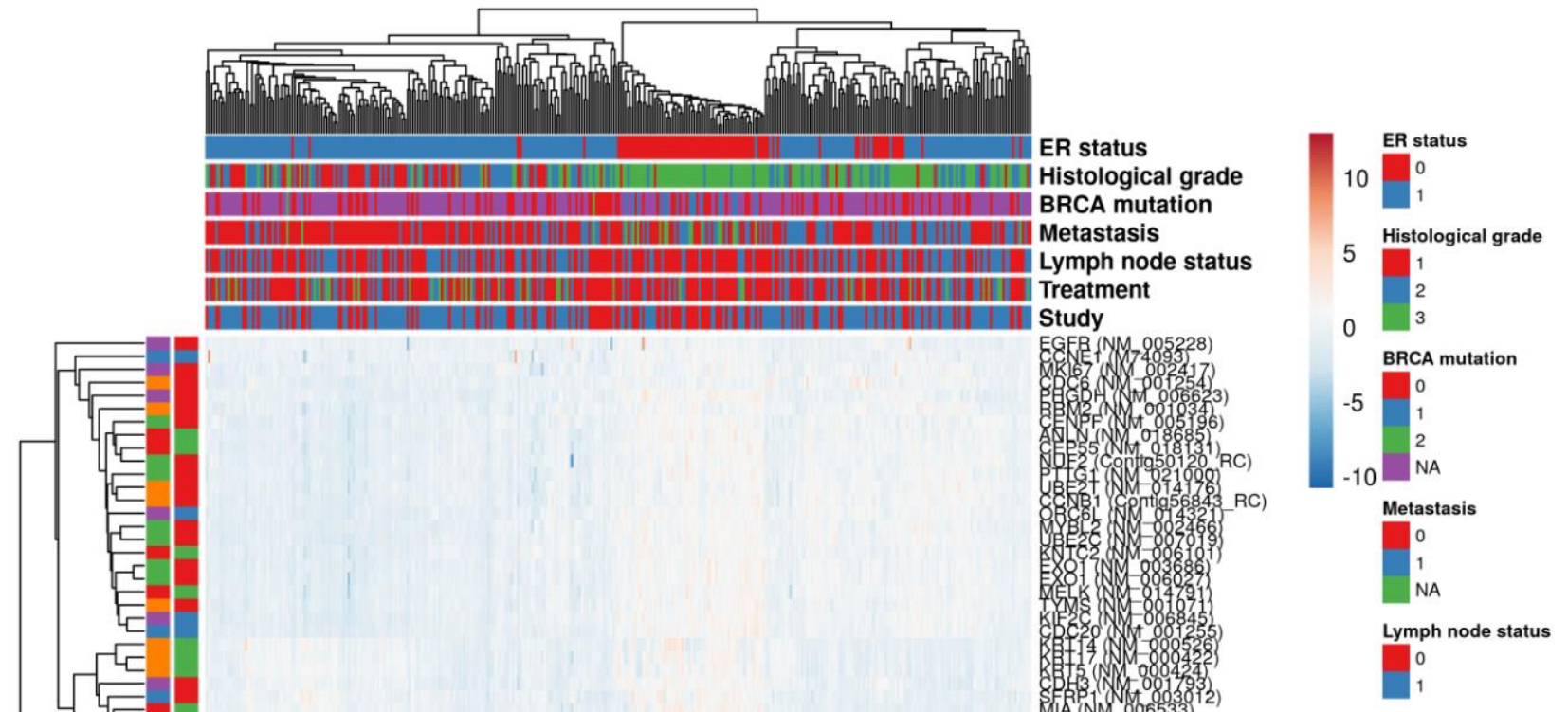
 Download PDF-file

 Download EPS-file

 Download SVG-file

## Caption example

Rows are centered; unit variance scaling is applied to rows. Imputation is used for missing value estimation. Both rows and columns are clustered using correlation distance and average linkage. 56 rows, 337 columns.



# PCA

## PCA options

☒ change data options

### Data options

Principal component on x-axis:

1

Principal component on y-axis:

2

☐ switch direction of x-axis

☐ switch direction of y-axis

☐ change display options

☐ change plot labels

[Download PDF-file](#)

[Download EPS-file](#)

[Download SVG-file](#)

### Caption example

Unit variance scaling is applied to rows; SVD with imputation is used to calculate principal components. X and Y axis show principal component 1 and principal component 2 that explain 31.1% and 11.6% of the total variance, respectively. Prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse. N = 337 data points.

