

HOW MANY GENES FROM WNT-CASCADE IDENTIFY *tnbc*-PATIENTS?

E. A. Martynova, A. A. Feller

V. F. Voino-Yasenetsky Krasnoyarsk state medical university

February, 4th, 2020

Very briefly

Mammalian cancer

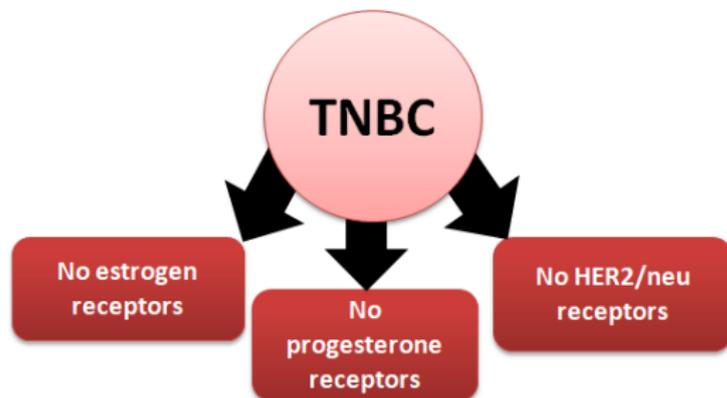
- mammalian cancer is # 1 in female mortality caused by cancer, both worldwide (24.2%), and at Russia (20.9%);
- average age of the first diagnostics of mammalian cancer in 2018 was 61.5 years;
- mammalian cancer is among the leading mortality factors (16.2%) in women subpopulation;
- average age of death caused by mammalian cancer in 2018 was 66.1 years.

Cancer is a severe social problem

tnbc – triple negative breast cancer

tnbc as is

- This type of cancer is “invisible” by three standard techniques of diagnostics;
- It is hard in early diagnostics;
- It has unfavorable prognosis.



Cancer is a severe social problem

tnbc – an attempt to involve genetic profiling

In 2018, Vladimir Katanaev from Geneva university gathered the database of human genes expression profiles from the patients with *tnbc*, *non-tnbc* and healthy women, to reveal the difference between *tnbc* and *non-tnbc* patients.

tnbc – an attempt to involve genetic profiling

In 2018, Vladimir Katanaev from Geneva university gathered the database of human genes expression profiles from the patients with *tnbc*, *non-tnbc* and healthy women, to reveal the difference between *tnbc* and *non-tnbc* patients.

Very few progress with linear statistics (correlation, regression, etc.) has been reported.

Cancer is a severe social problem

tnbc – an attempt to involve genetic profiling

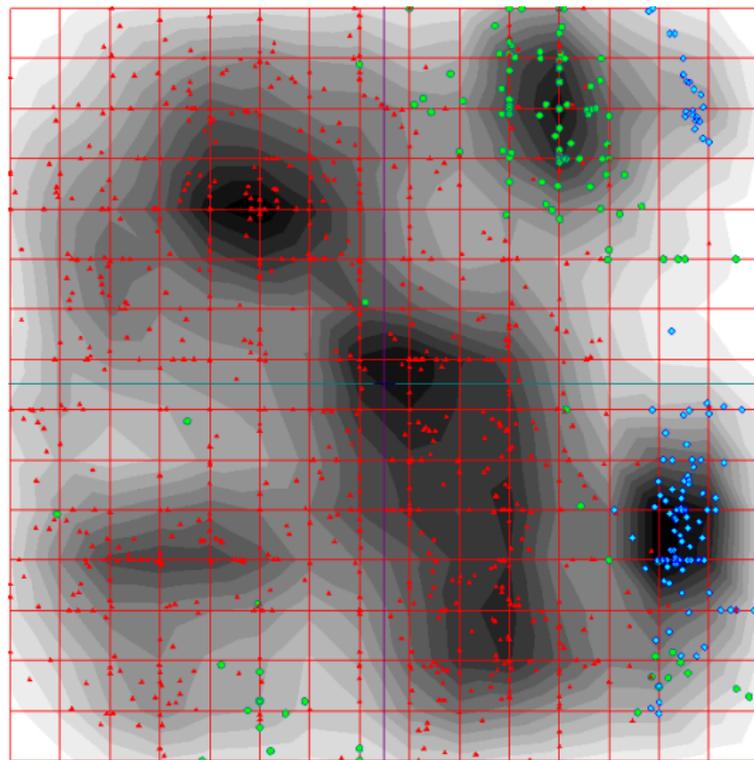
In 2018, Vladimir Katanaev from Geneva university gathered the database of human genes expression profiles from the patients with *tnbc*, *non-tnbc* and healthy women, to reveal the difference between *tnbc* and *non-tnbc* patients.

Very few progress with linear statistics (correlation, regression, etc.) has been reported.

An implementation of elastic map technique followed in great progress in discrimination of *tnbc* patients from other ones.

Cancer is a severe social problem

title



blue – normal

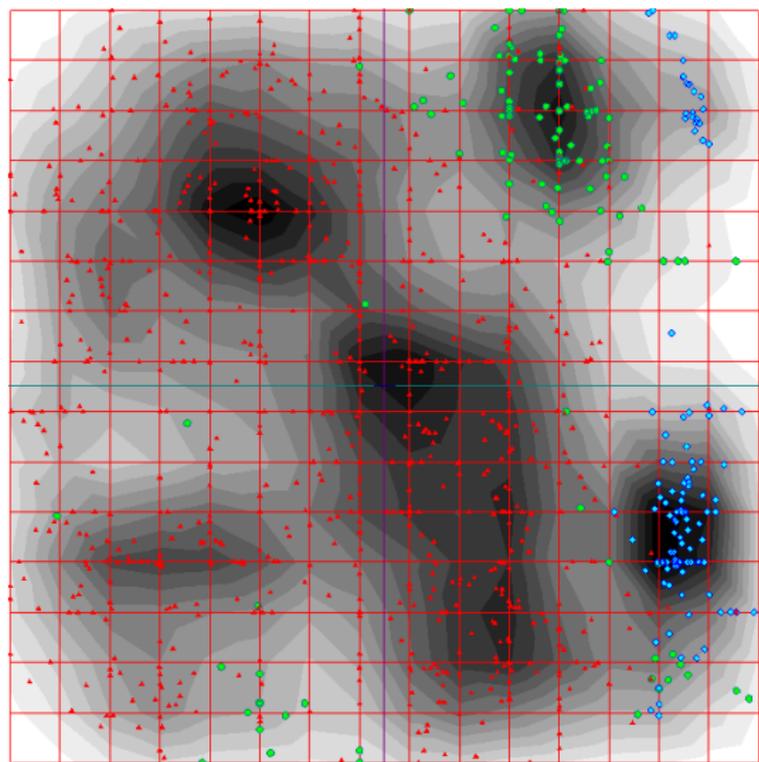
nontnbc

tnbc

Gray scale represents
the local density of
points on the elastic
map.

Cancer is a severe social problem

title



blue – normal

red *nontnbc*

green *tnbc*

Gray scale represents the local density of points on the elastic map.

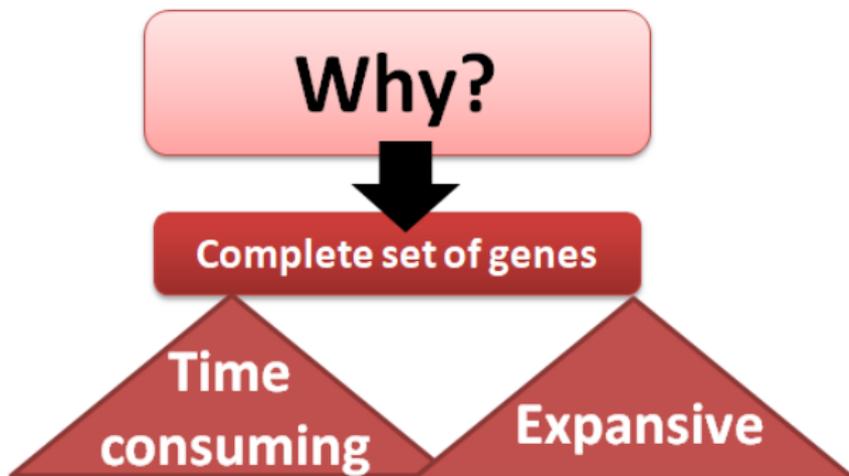
Surprisingly, it was found that a part of the set of genes still provides a good discrimination of sick and healthy patients.

Cancer is a severe social problem

The aim of the study

For the genes of Wnt-cascade

- to reveal the minimal subset of genes providing reliable discrimination of *non-tnbc*, *tnbc* and healthy patients, over the expression profiles.



Tasks

Following are the specific tasks:

- To develop random sets of 34, 17, 8 and 4 genes, 50 entities each;
- To find out the minimal subset of genes from Wnt-cascade that provides stable segregation of *non-tnbc*, *tnbc* and healthy patients;
- To estimate the impact of each gene from the minimal set into the separability of the patients;
- To check the stability of the found minimal subsets.

The database has 1082 entries

Among them are:

- 118 entries of *tnbc* patients;
- 112 entries of healthy patients;
- 852 entries of *non-tnbc* patients.

Each patient is characterized by

- the figures of the expression of 68 genes from Wnt-cascade.

The database

- The database is provided by the Translational Research Centre in Onco-haematology headed by prof. V. Katanaev).

The first steps

For each expression level figure the following analysis has been done:

- descriptive statistics: mean values, standard deviations, modes, variance;
- correlation analysis for all the genes.

We clustered the patients by elastic map basing on Wnt-cascade expression data

Elastic map technique

- Elastic map is the approximation of multidimensional bulky data with manifolds of low (two, indeed) dimension.
- The method is based on the approximation of the data with specially “jammed” manifolds that fits the data best of all.
- We used two dimensional manifold: square and sphere. Both manifolds were transformed to fit the data, and then stretched back to a plane, in order to visualize cluster pattern.

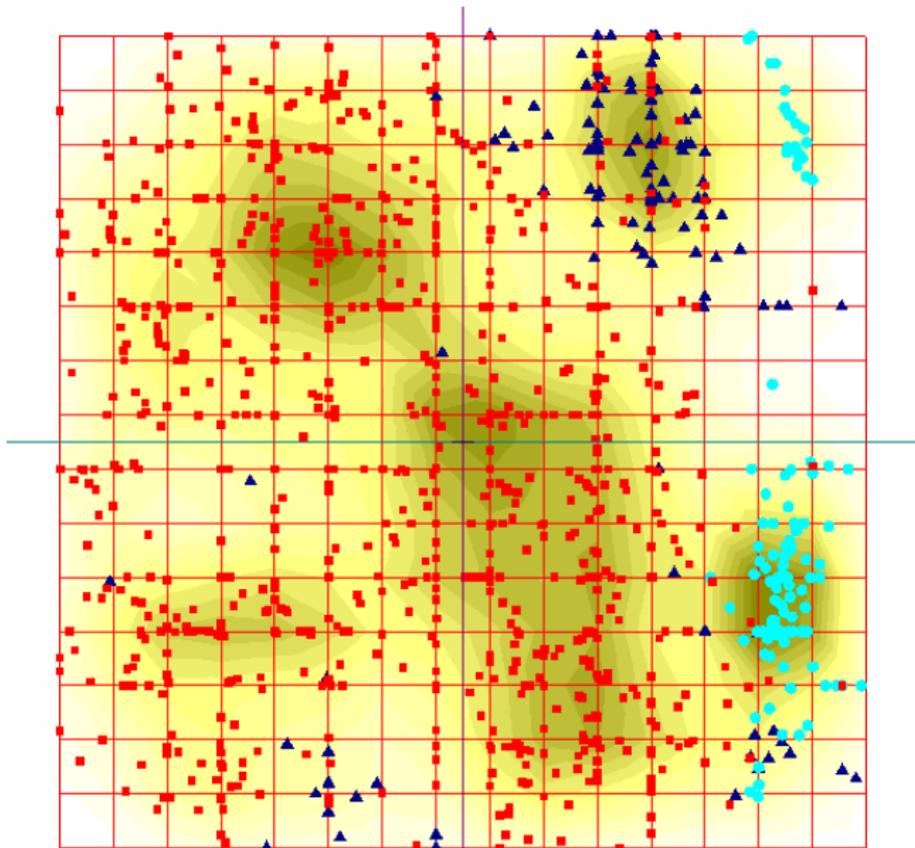
Any transformations **must** conserve the topology!

Elastic map clustering

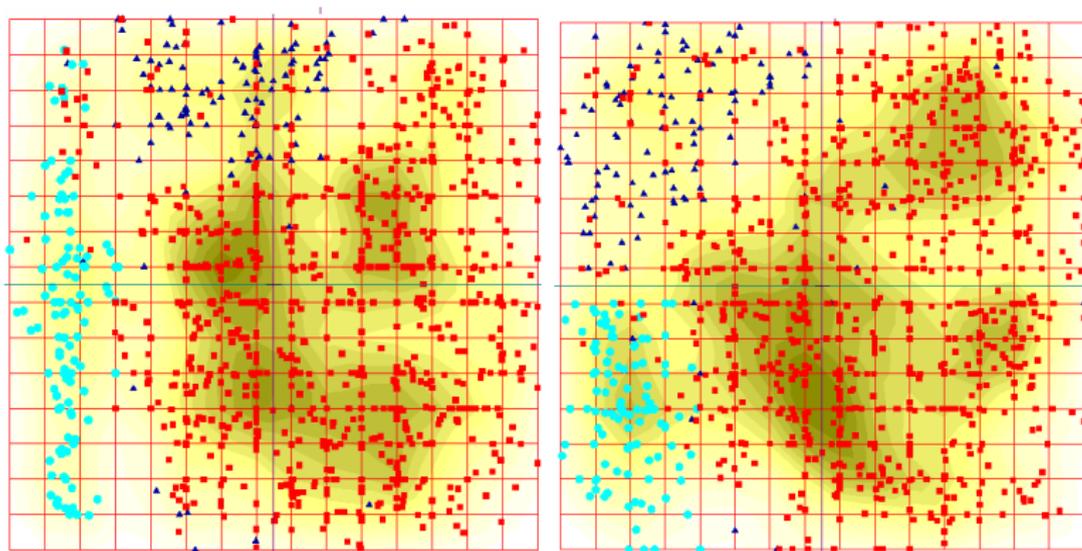
We tested 50 random sets of genes comprising:

- 34 genes;
- 17 genes;
- 8 genes, and
- 4 genes.

Clustering over 68 genes of Wnt-cascade

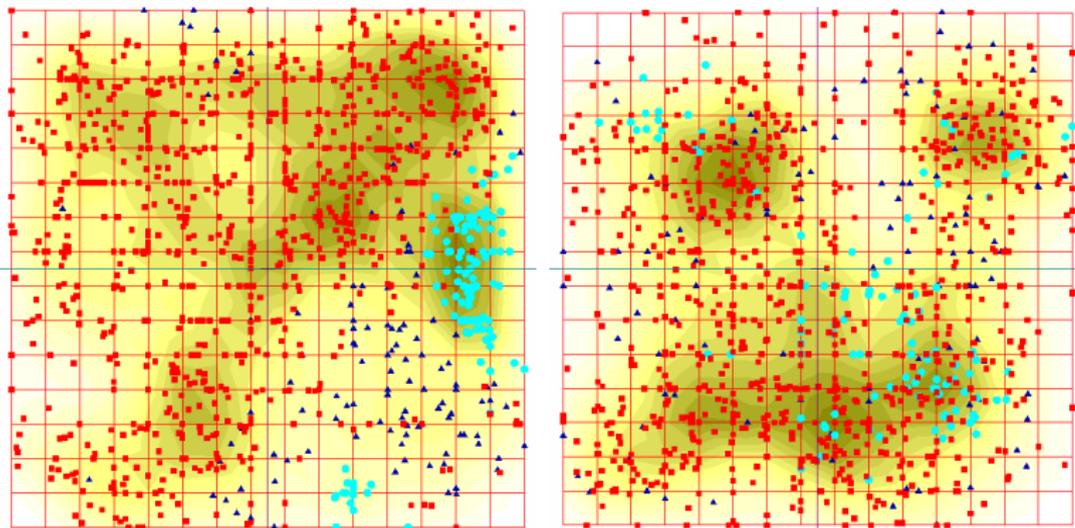


Clustering over 34 and 17 genes of Wnt-cascade



Clustering: left 34 genes, right 17 genes, correspondingly.

Clustering over 8 and 4 genes of Wnt-cascade



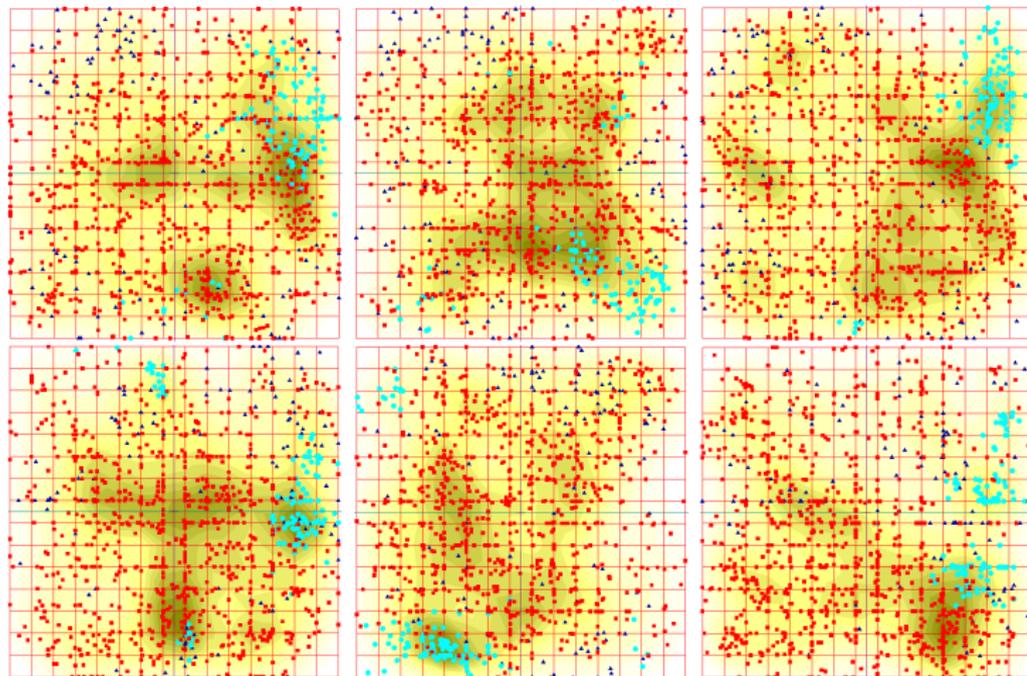
Clustering: left 8 genes, right 4 genes, correspondingly.

Results

So, it is found that

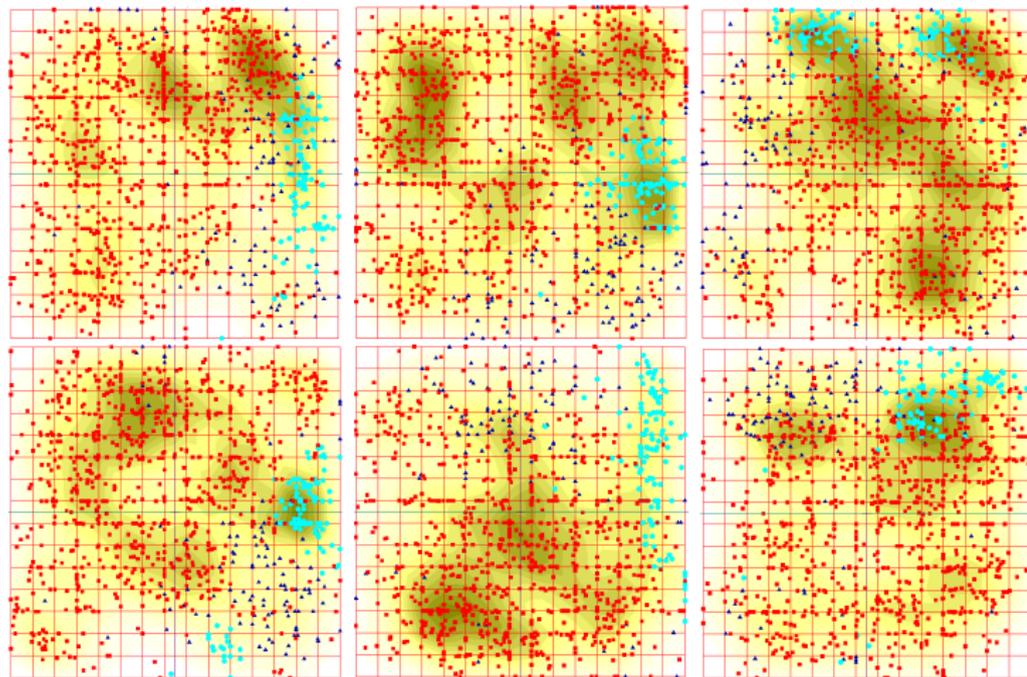
- *tnbc*, *non-tnbc* and healthy patients highly reliably separated into clear clusters using the expression level data of Wnt-cascade genes, with elastic maps;
- the cascade seems to be redundant, from the point of view of an abundance of genes involved into the clustering provided with elastic map technique;
- four octets of genes provide reliable and stable segregation of patients, while others do not.

Relevance of each gene from the efficient octet. A gene from efficient octet was added with 7 random genes.



Genes up to down, left to right: *DKK2*, *DKK3*, *FZD4*, *FZD6*, *GNA11*, *NKD2*.

Relevance checking – 2. A gene from each efficient octet was excluded.



Genes up to down, left to right: *DKK3*, *GNA14*, *GNAI1*, *GNAZ*, *SFRP1*, *TLE1*.

Relevance checking: results

It was found that

- there are no single specific gene determining the stability of clustering;
- any 7 genes kit provides significantly poorer clustering, in comparison to 8 genes.

The results prove the efficient clustering provided by 8 genes from Wnt-cascade. Not any set of 8 genes supports good stratification; the role of each gene from the efficient octet awaits for further study.

Conclusions

The main results of the work are following.

- Four octets of genes from Wnt-cascade supporting the reliable stratification of the patients with various mammalian cancer type are found.
- Some genes are tested for its relevance for clear clustering.
- The found clear cluster patterns are found to be stable.

We are thankful to



Professor Vladimir Katanaev from Geneva university, for the formulation of the problem of research, and for the database on cancer patients.



Professor Michael Sadovsky from Institute of computational modeling of SB RAS, for strong encouragement and permanent interest to the work.

Thank you for your attention!

Questions, please!