

MATH:7450 (22M:305) Topics in Topology: Scientific and Engineering Applications of Algebraic Topology

Sept 18, 2013: javaPlex

Fall 2013 course offered through the
University of Iowa Division of Continuing Education

Isabel K. Darcy, Department of Mathematics
Applied Mathematical and Computational Sciences,
University of Iowa

<http://www.math.uiowa.edu/~idarcy/AppliedTopology.html>

http://bioinformatics.nki.nl/data.php



Bioinformatics and Statistics

Division of Molecular Carcinogenesis, Netherlands Cancer Institute



About

People

Research

Publications

Software

Data

Vacancies

Student Projects

Courses

Meetings

Intranet

Contact

Data

- **Gene expression profiling predicts clinical outcome of breast cancer**
van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH
Nature 2002 Jan 31;415(6871):530-6.

Title: Software
Address: <http://bioinformatics.nki.nl/software.php>

Patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70-80% of patients receiving this treatment would have survived without it. None of the signatures of breast cancer gene expression reported to date allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of BRCA1 carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and angiogenesis. This gene expression profile will outperform all currently used clinical parameters in predicting disease outcome. Our findings provide a strategy to select patients who would benefit from adjuvant therapy.

- Data can be downloaded [here](#).

3 columns = patient
middle column (ratio) = data point

log10(Intensity)	Log10(ratio)	P-value
-1.66	-0.299	6.72E-01
-1.55	0.093	8.93E-01
-1.71	-0.215	8.36E-01
-1.46	-0.566	2.83E-01
-1.08	-0.596	1.17E-01
-1.61	-0.195	8.14E-01
0.69	0.039	5.25E-01

rows = genes

Create Data Matrix

```
load_javaplex
```

```
C = csvread('Array5yr.csv',2,1,[2,1,3,21])
```

```
C(1, 2)
```

```
for i = 1:7 D(:,i) = C(:,3*i-1); end
```

```
R = transpose(D)
```

```
size(R)
```

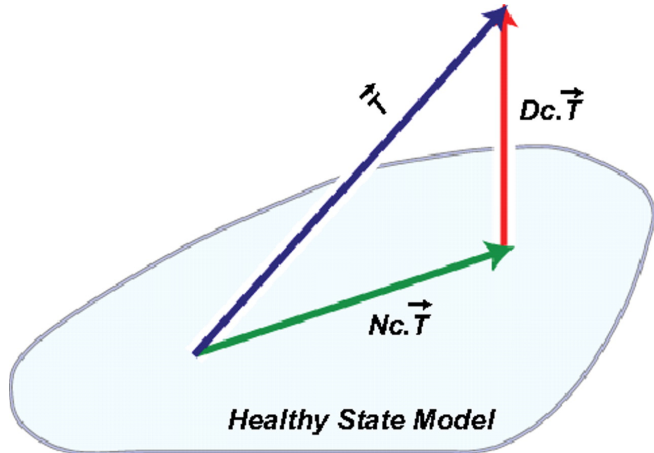
Use standard Euclidean Metric:

```
m_space = metric.impl.EuclideanMetricSpace(R);
```

```
m_space.getPoint(0)
```

```
m_space.distance(m_space.getPoint(0), m_space.getPoint(1))
```

```
sqrt([R(1,1) - R(2, 1)]^2 + [R(1,2) - R(2,2)]^2)
```



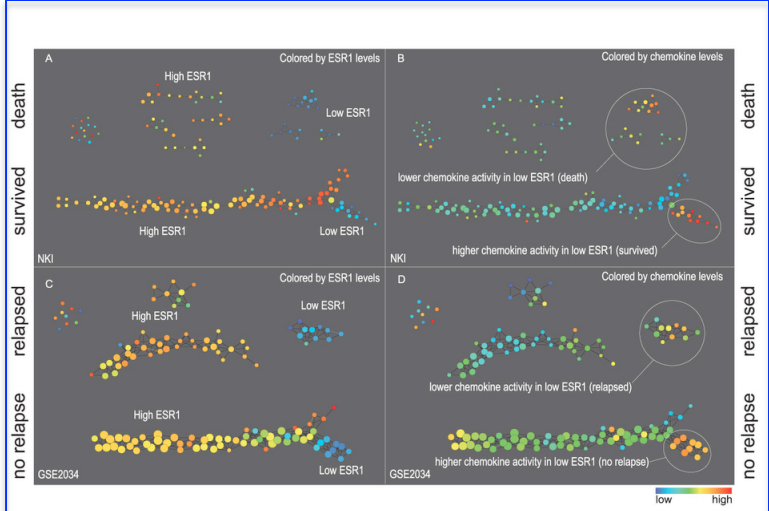
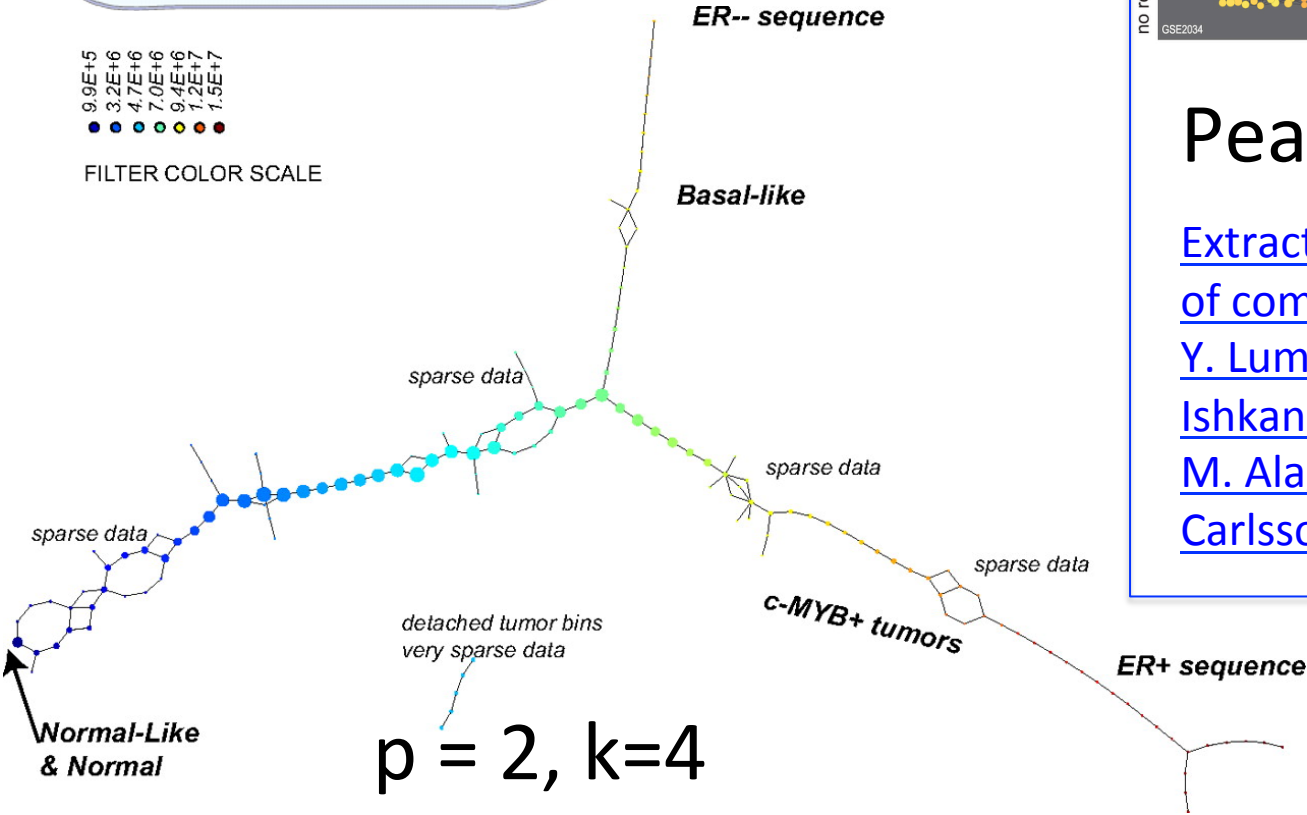
$$[\sum |x_i|^p]^{k/p}$$

$$k = 1...10$$

$$p = 1...5$$

- 9.9E+5
- 3.2E+6
- 4.7E+6
- 7.0E+6
- 9.4E+6
- 1.2E+7
- 1.5E+7

FILTER COLOR SCALE



Pearson correlation

[Extracting insights from the shape of complex data using topology P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, G. Carlsson \(2013\)](#)

Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival
 Monica Nicolau, Arnold J. Levine¹, and Gunnar Carlsson, PNAS 2011

Choose your own distance matrix:

```
dist = ones(7) - eye(7)
```

```
dist_space = metric.impl.ExplicitMetricSpace(dist);
```

```
dist_space.distance(0,1)
```

Calculate Vietoris Rips Complex

max_dimension = 3;

max_filtration_value = 4;

num_divisions = 100;

```
stream = api.Plex4.createVietorisRipsStream(R,  
max_dimension,max_filtration_value, num_divisions);
```

$$t \in \left\{ 0, \frac{t_{max}}{N-1}, \frac{2t_{max}}{N-1}, \frac{3t_{max}}{N-1}, \dots, \frac{(N-2)t_{max}}{N-1}, t_{max} \right\}$$

Calculate Persistence

```
persistence
```

```
=api.Plex4.getModularSimplicialAlgorithm(max_dimension, 2);
```

```
intervals = persistence.computeIntervals(stream)
```

```
intervals = persistence.computeAnnotatedIntervals(stream)
```

```
betti_numbers_array = infinite_barcodes.getBettiSequence()
```

```
betti_numbers_string = infinite_barcodes.getBettiNumbers()
```

```
options.filename = 'small_data'  
options.max_filtration_value = max_filtration_value  
options.max_dimension = max_dimension - 1  
plot_barcodes(intervals, options)
```

Run on entire set:

```
load_javaplex;
clear C; clear D; clear R;
C = csvread('Array5yr.csv',2,1);
for i = 1:35 D(:,i) = C(:,3*i-1); end
R = transpose(D);
stream = api.Plex4.createVietorisRipsStream(R,
max_dimension,max_filtration_value, num_divisions);

persistence
=api.Plex4.getModularSimplicialAlgorithm(max_dimension, 2);

intervals = persistence.computeIntervals(stream)
options.filename = 'data';
options.max_filtration_value = max_filtration_value;
options.max_dimension = max_dimension - 1;
plot_barcodes(intervals, options)
```