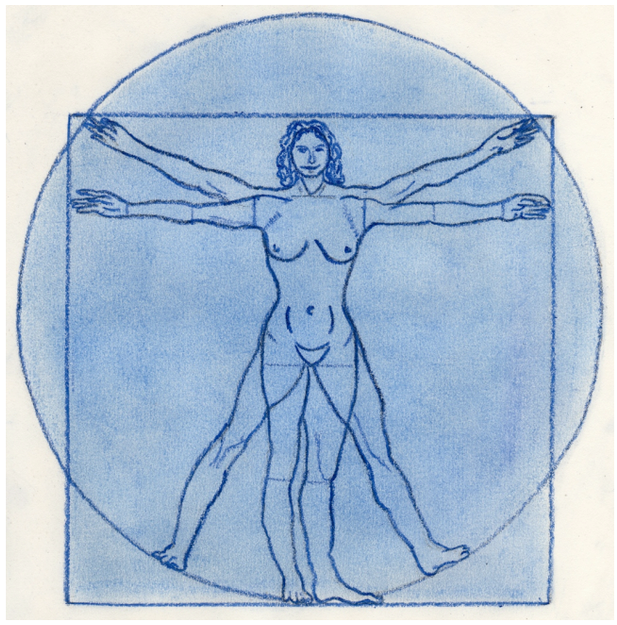
Серийные, сетевые и глубинные маркеры

(и немного про генетический интеллект)



Проф Алексей Заикин

Institute for Women's Health and Department of Mathematics

University College London

www.zaikinlab.com

alexey.zaikin@ucl.ac.uk

УНИВЕРСИТЕТ ЛОБАЧЕВСКОГО

ЦАСС Центр анализа сложных систем

Сеченовский Университет

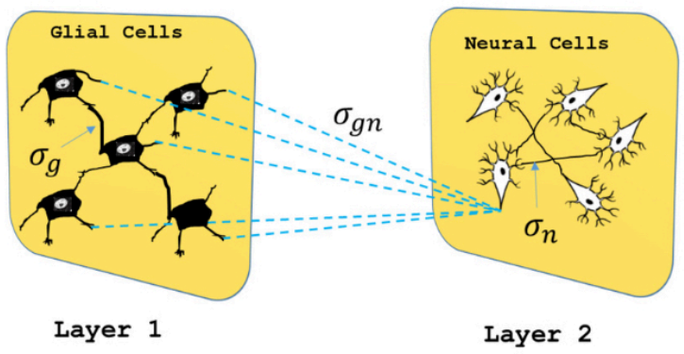
UCL

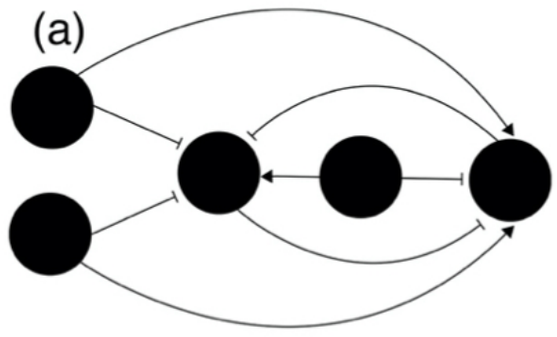
www.zaikinlab.com

Data analysis

Зам глав врача: искусственный интеллект пока приносит лечащему врачу больше хлопот чем пользы

Mathematical modelling





Content and aim: review of possible collaborations

Data analysis:

Longitudinal analysis

Network analysis

Deep learning analysis

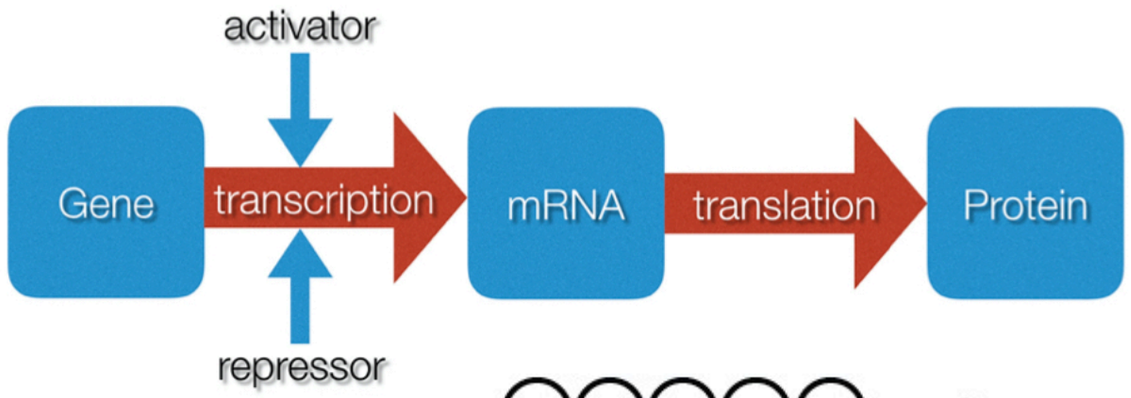
Proteome

Epigenome

Images

Transcriptional regulation

- Basic paradigm of molecular biology. Kind of a "Newton's law" for living systems:



Omic Information: Small review on molecular biology

Epigenetics and MicroRNAs

JODY C. CHUANG, AND PETER A. JONES

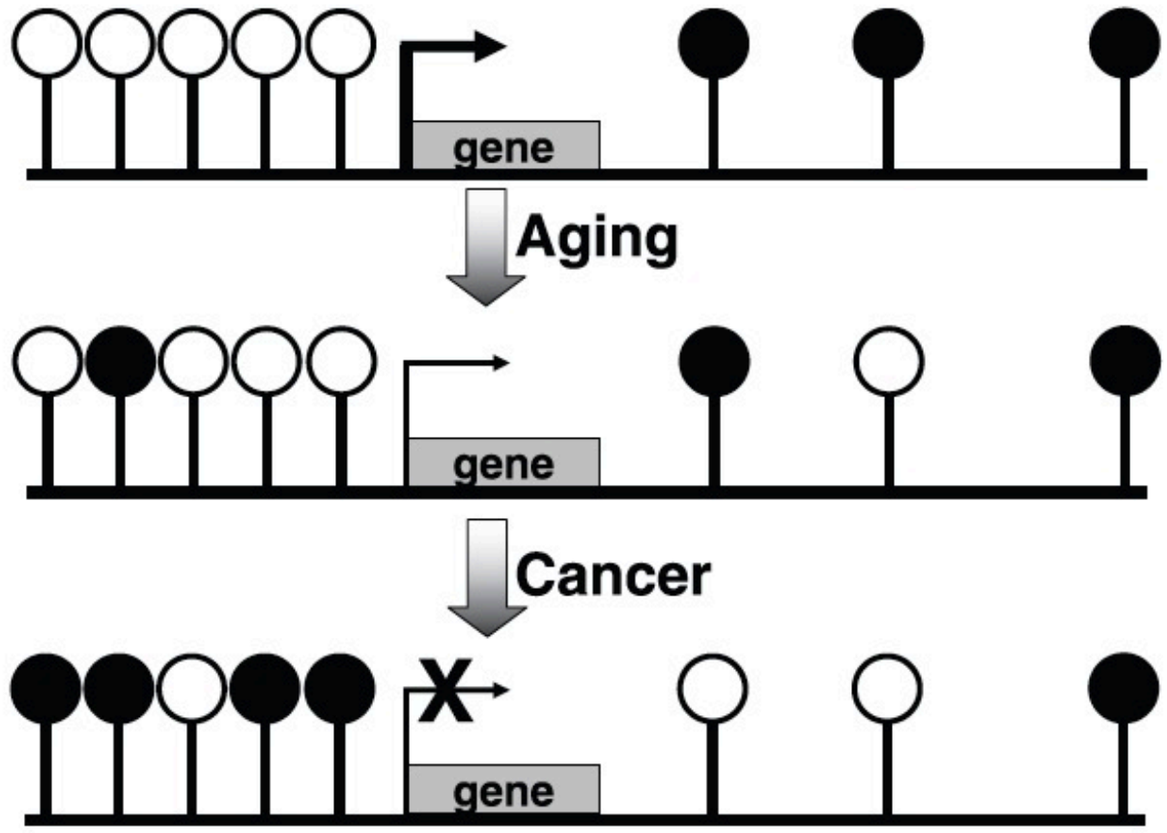


Figure 3. CpG islands normally remain unmethylated, whereas the sporadic CpG sites located in the rest of the genome often are methylated. With aging, there is a gradual reversal of this phenomenon. During carcinogenesis, this change is much more dramatic, leading to a global hypomethylation and hypermethylation of CpG islands. The results are chromosomal instability and silencing of some important tumor-suppressor genes.

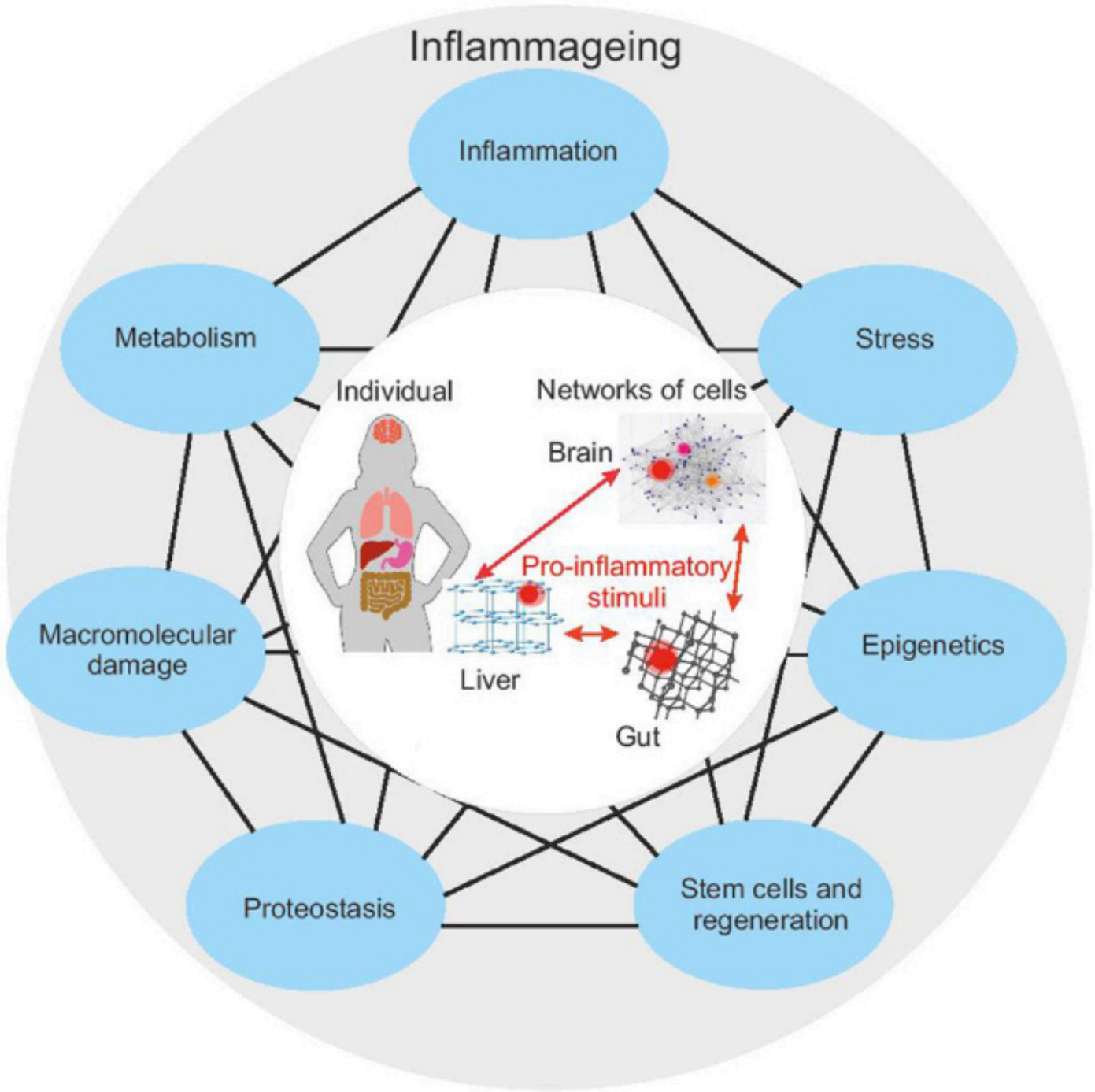
Personalised medicine

frontiers
in Aging Neuroscience

REVIEW
published: 25 May 2020
doi: 10.3389/fnagi.2020.00136

The Human Body as a Super Network: Digital Methods to Analyze the Propagation of Aging

Harry J. Whitwell[1], Maria Giulia Bacalini[2], Oleg Blyuss[3,4], Shangbin Chen[5], Paolo Garagnani[6], Susan Yu Gordleeva[7], Sarika Jalan[8,9], Mikhail Ivanchenko[10], Oleg Kanakov[7], Valentina Kustikova[10], Ines P. Mariño[11], Iosif Meyerov[10], Ekkehard Ullner[12], Claudio Franceschi[7,10] and Alexey Zaikin[4,10,13]**



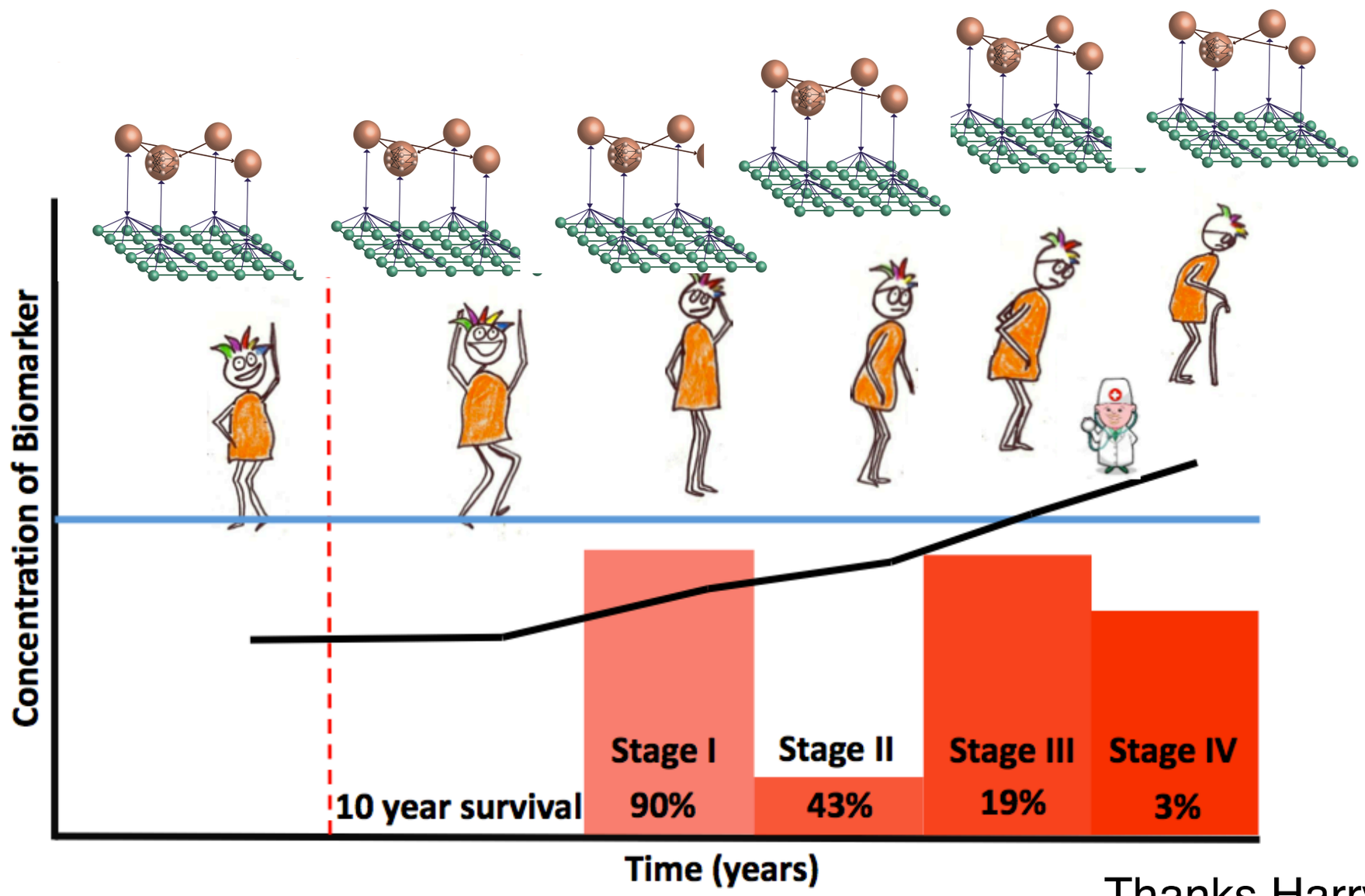
OM&P

MAMMALIAN BRAIN AS A NETWORK OF NETWORKS

Veronika Samborska [1], Susanna Gordleeva [2], Ekkehard Ullner [3], Albina Lebedeva [2], Viktor Kazantsev [2], Mikhail Ivanchenko [2] and Alexey Zaikin [2,4]

Search for Network, Longitudinal and Deep Learning Biomarkers

Multi-omic data: Genetic, Epigenetic and Proteomic



Thanks Harry for picture

Longitudinal and deep learning markers

a First-Line Screen for Ovarian Cancer

- Ovarian cancer remains the leading cause of death from gynecological cancer among women and accounts for 5% of all female deaths from cancer, corresponding to annual deaths of around 4,100 in the United Kingdom, 42,700 in Europe, 22,280 in the United States and 152,000 worldwide.
- Most women are diagnosed in advanced stage (Stage III–IV) with reported 5-year survival rates of 19% (Stage III) and 3% (Stage IV) respectively. The higher survival rates of 70%–90% in earlier stage (Stage I–II) disease has driven international screening efforts to detect the disease earlier.
- In the United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), women in the multimodal (MMS) arm had a serum CA125 test (first-line), with those at increased risk, having repeat CA125/ultrasound (second-line test). CA125 was interpreted using the "Risk of Ovarian Cancer Algorithm" (ROCA).
- Experimental Design: 50,083 post-menopausal women who attended 346,806 MMS screens were randomly split into training and validation sets, following stratification into cases (ovarian/tubal/peritoneal cancers) and controls.

Longitudinal analysis of biomarkers

Markov Chain Monte-Carlo Approximate Bayesian Computation and Bayesian Change Point Models

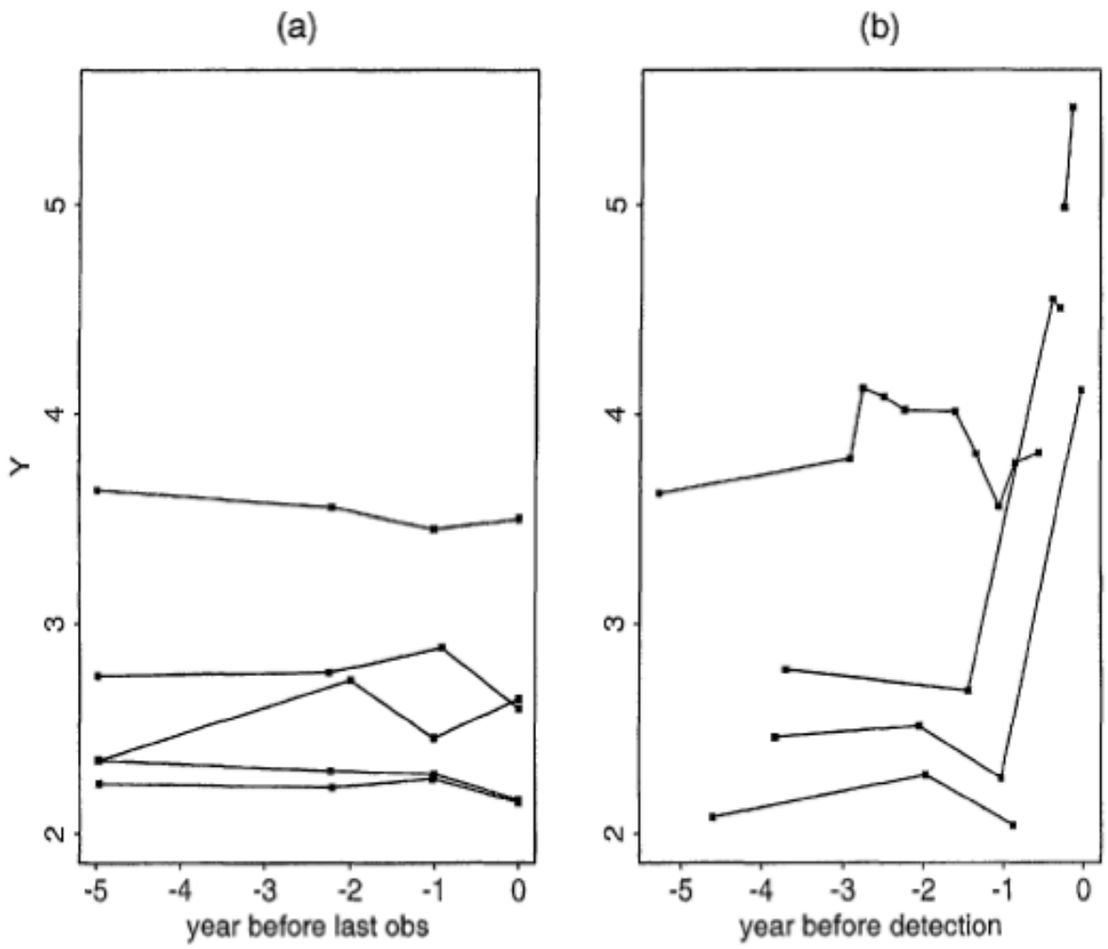
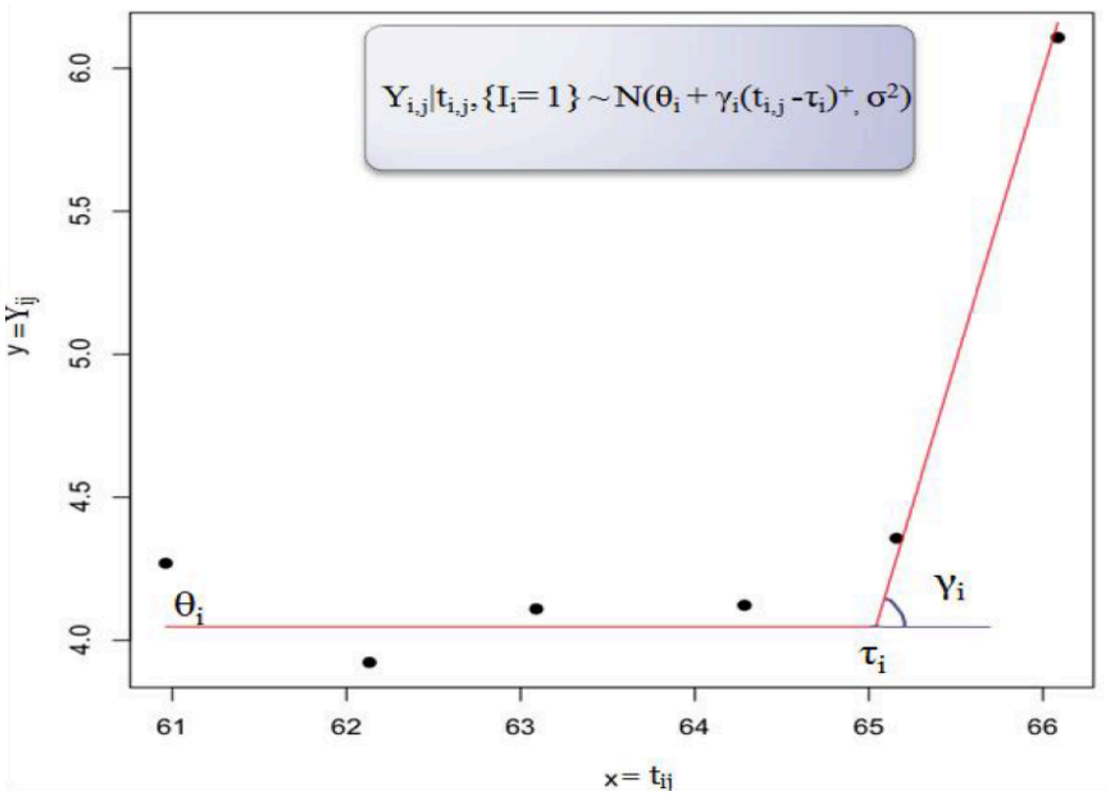
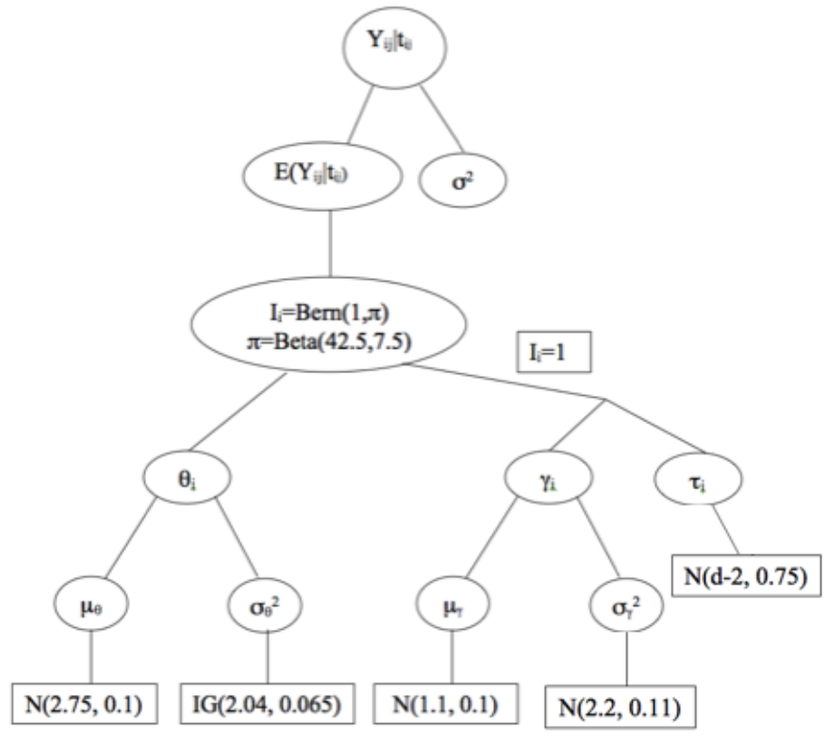


Figure 1. Transformed CA125 Values (Y) From Five Controls (a) and Five Cases (b) in the U.K. Screening Trial (Jacobs et al. 1993).





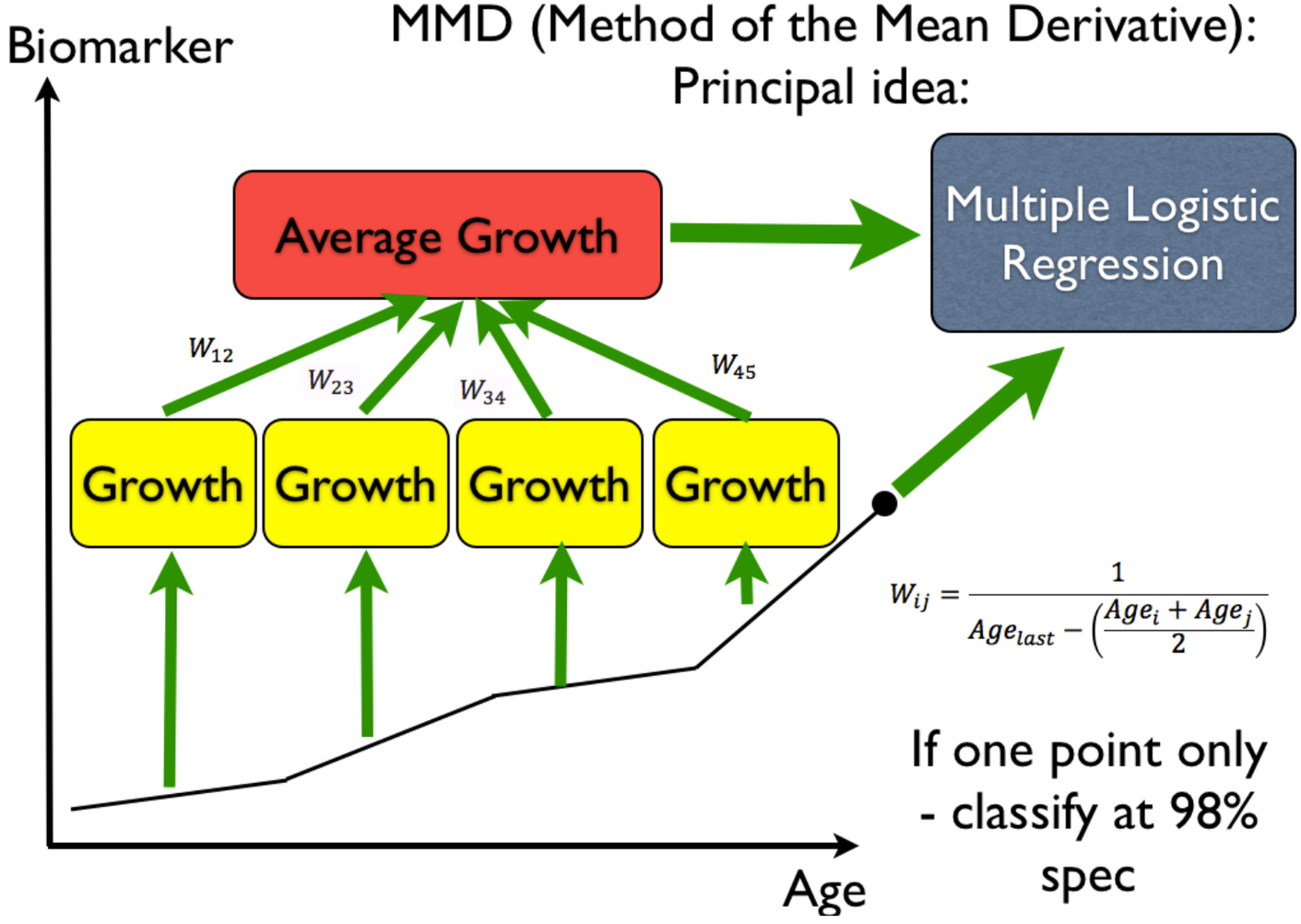
Change-point of multiple biomarkers in women with ovarian cancer

Inés P. Mariño[a,b,*,1], Oleg Blyuss[b,1], Andy Ryan[b], Aleksandra Gentry-Maharaj[b], John F. Timms[b], Anne Dawnay[c], Jatinderpal Kalsi[b], Ian Jacobs[b,d], Usha Menon[b,2], Alexey Zaikin[b,2]

Biomedical Signal Processing and Control 33 (2017) 169–177

Longitudinal analysis of biomarkers

Method of Mean Trends



Comparison of longitudinal CA125 algorithms as a first line screen for ovarian cancer in the general population

Oleg Blyuss[1], Matthew Burnell[1], Andy Ryan[1], Aleksandra Gentry-Maharaj[1], Inés P. Mariño[1,2], Jatinderpal Kalsi[1], Ranjit Manchanda[1,6], John F. Timms[1], Mahesh Parmar[3], Steven J. Skates[4], Ian Jacobs[1,5,8], Alexey Zaikin[1,7*], and Usha Menon[1*].

Indicators used (for every i-th patient):

- Last measurement
- Trend 1 (Mean derivative)

$$\left(\sum_{j=1}^{k_i-1} \frac{y_{j+1}-y_j}{t_{j+1}-t_j} \frac{1}{t_{k_i}-(t_{j+1}+t_j)/2}\right)/(k_i-1)$$

- Trend 2

$$\left(\sum_{j=1}^{k_i-1} \frac{(y_{j+1}-y_j)(t_{j+1}-t_j)}{2}\right)/(k_i-1)$$

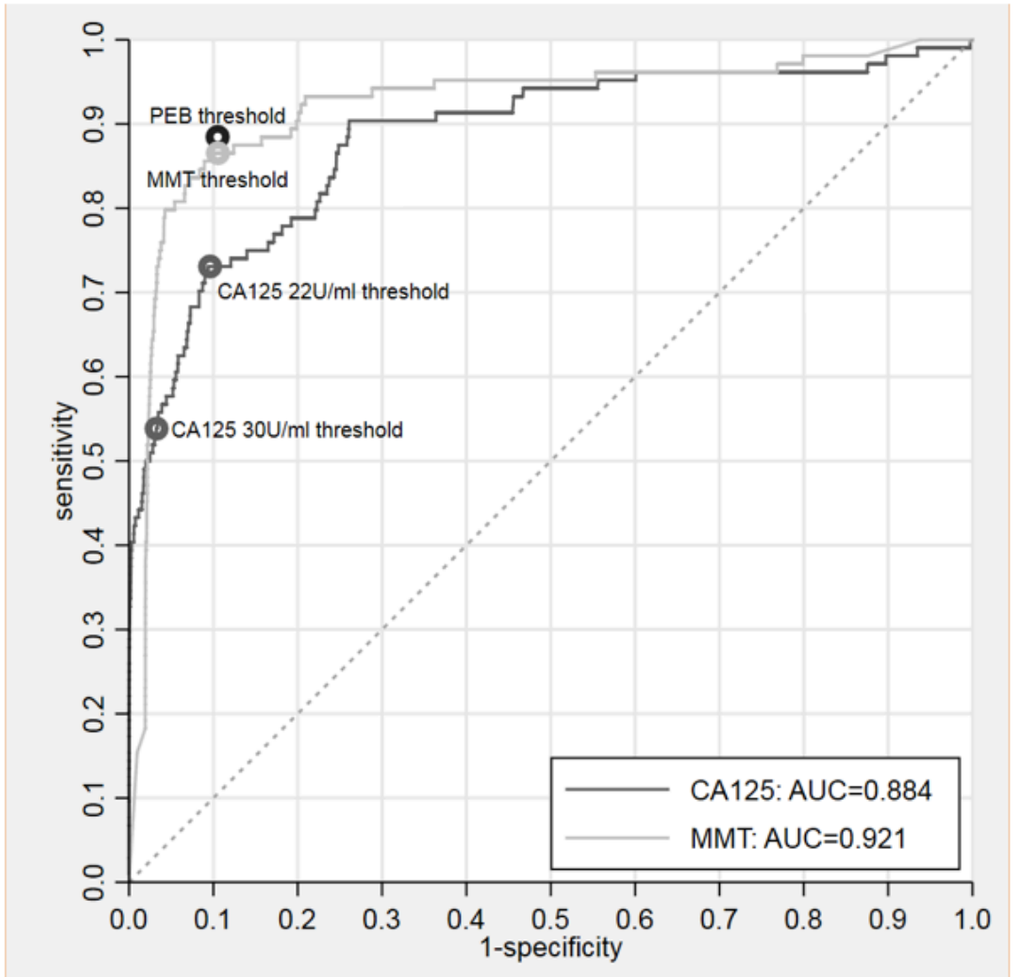
- Trend 3

$$\sqrt{\frac{\sum_{j=1}^{k_i}(y_j-\bar{y})^2}{k_i}}/\bar{y}$$

- Trend 4

$$\frac{\sum_{j=1}^{k_i} y_j t_j}{\sum_{j=1}^{k_i} t_j}$$

- Trend 5 (Variance)



- Key papers:
- finding longitudinal oncomarkers- License obtained!

Published OnlineFirst July 3, 2018; DOI: 10.1158/1078-0432.CCR-18-0208

Precision Medicine and Imaging

Clinical Cancer Research

Comparison of Longitudinal CA125 Algorithms as a First-Line Screen for Ovarian Cancer in the General Population

Oleg Blyuss[1], Matthew Burnell[1], Andy Ryan[1], Aleksandra Gentry-Maharaj[1], Inés P. Mariño[1,2], Jatinderpal Kalsi[1], Ranjit Manchanda[1,3], John F. Timms[1], Mahesh Parmar[4], Steven J. Skates[5], Ian Jacobs[1,6,7], Alexey Zaikin[1,8], and Usha Menon[1]

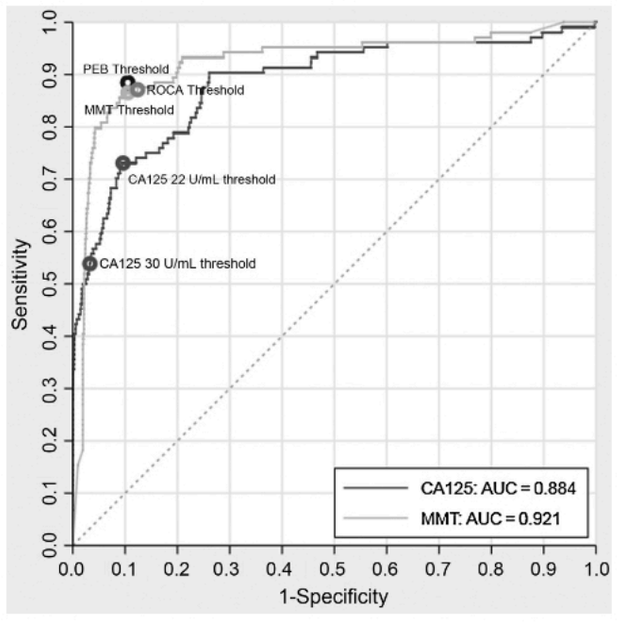
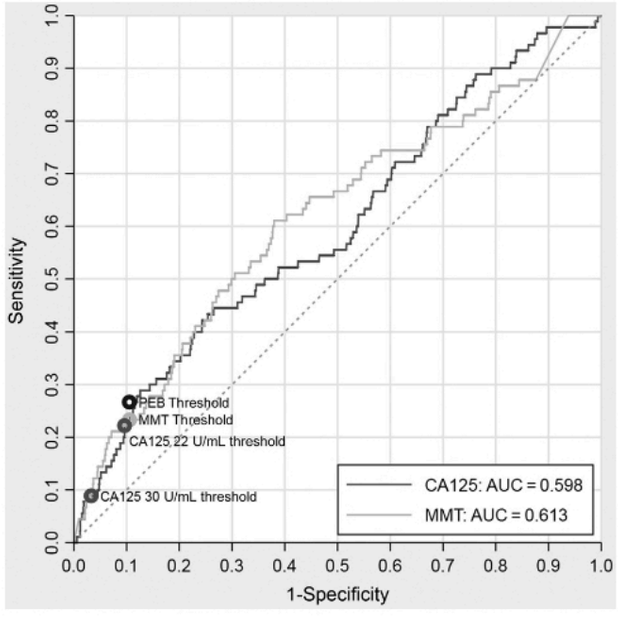


Figure 1.
Performance characteristics of CA125 interpreted using MMT, threshold rules, PEB and ROCA for detection of iEOC/PPC cases. Circle points give particular values of sensitivity and specificity provided by MMT and PEB corresponding to cutoff values obtained from the training set (MMT and PEB), CA125 using 22 and 30 U/mL cutoff values and ROCA as reported in ref. (6). Abbreviations: PEB, parametric empirical Bayes; MMT, method of mean trends; CA125, cancer antigen 125; AUC, area under roc-curve.



igure 2.
econdary analysis ROC curves for CA125 interpreted using MMT, threshold rules nd PEB for detection of iEOC/PPC cases. Circle points on the ROC curves give articular values of sensitivity and specificity provided by MMT and PEB orresponding to cutoff values obtained from the training set (MMT and PEB). bbreviations: PEB, parametric empirical Bayes; MMT, method of mean trends; A125, cancer antigen 125; AUC, area under roc-curve.

Deep oncomarkers

Deep Learning Neural Networks Recurrent Neural Networks with GRU

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A quantitative performance study of two automatic methods for the diagnosis of ovarian cancer

Manuel A. Vázquez[a,b,1], Inés P. Mariño[c,d,*,1], Oleg Blyuss[e,d], Andy Ryan[d], Aleksandra Gentry-Maharaj[d], Jatinderpal Kalsi[d], Ranjit Manchanda[d,f], Ian Jacobs[d,g,h], Usha Menon[d,2], Alexey Zaikin[d,i,j,2]

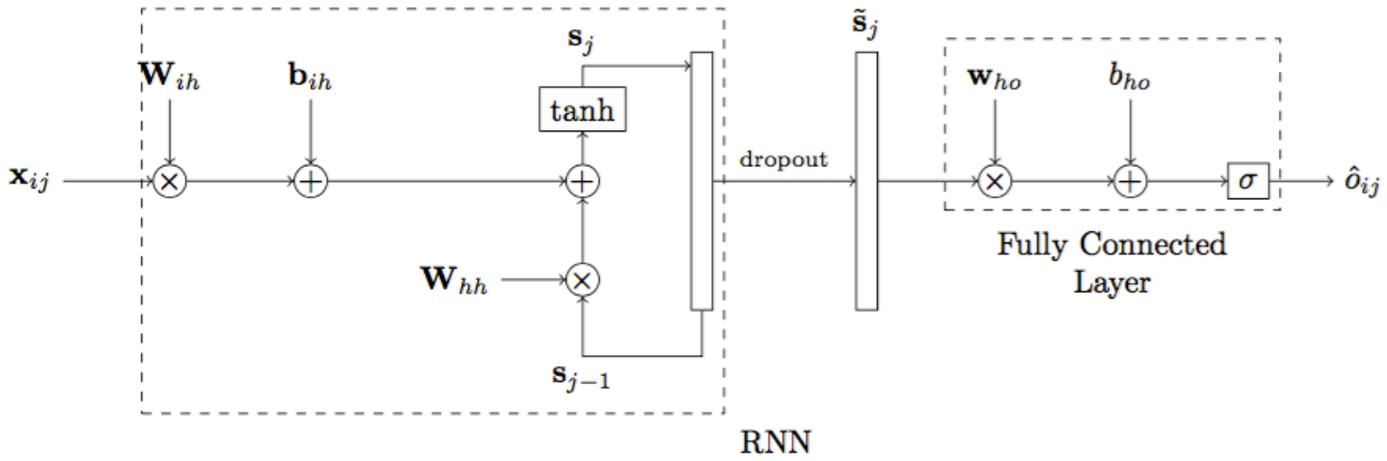
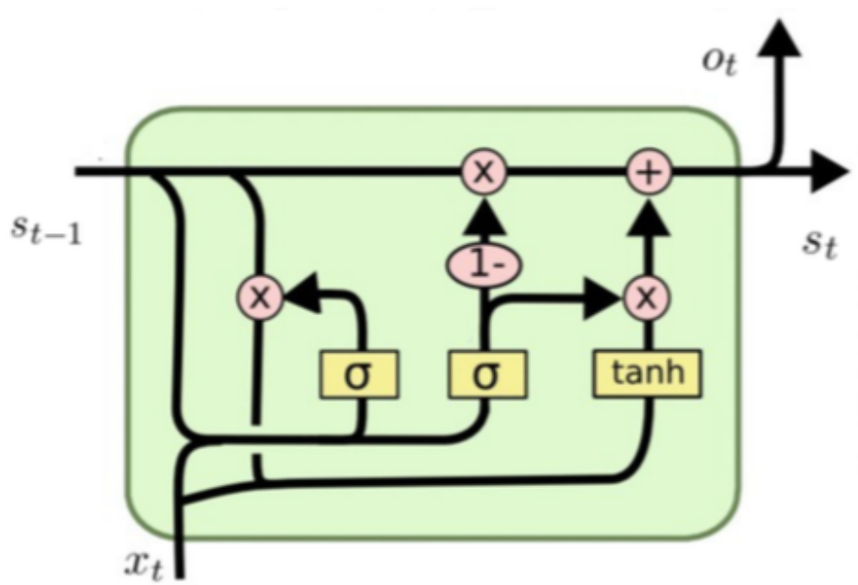


Fig 2. Network architecture for a single biomarker.

Development of neural network architecture for the early diagnosis analysis of oncomarker data in womens cancer

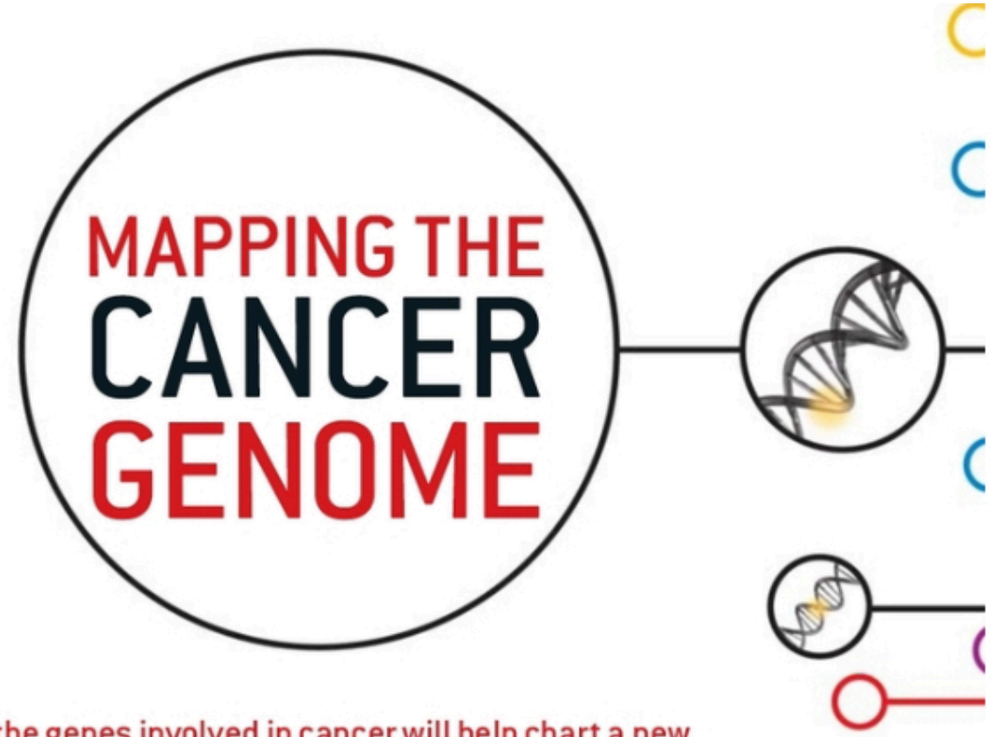


V. Cherepanova et al.

Work in progress!

Network markers

Constructing network biomarkers



Scientific American
Vol. 296, No. 3 (MARCH 2007), pp. 50-57 (8 pages)

Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies

By Francis S. Collins and Anna D. Barker

> *"One difficulty in interpreting this data for defining clinically useful information is that multiple different changes may be responsible for the onset of a disease, as exemplified by the efforts of The Cancer Genome Atlas project"*

- Network biomarkers for methylation: using correlations between intra-gene profiles
- Parenclitical networks: constructing networks when links are unknown

DNA Methylation Analysis

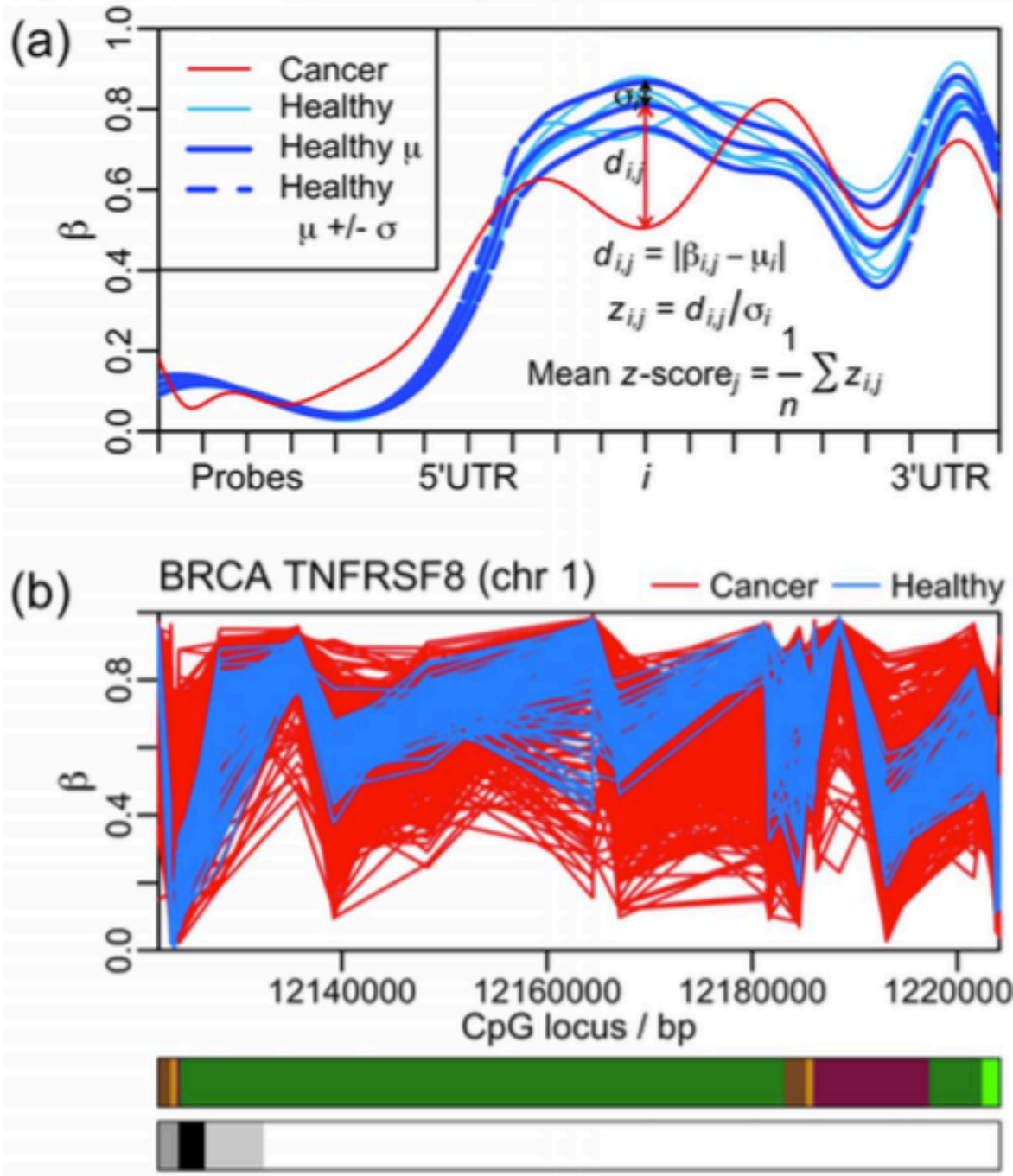
Misha Ivanchenko, Zaikin

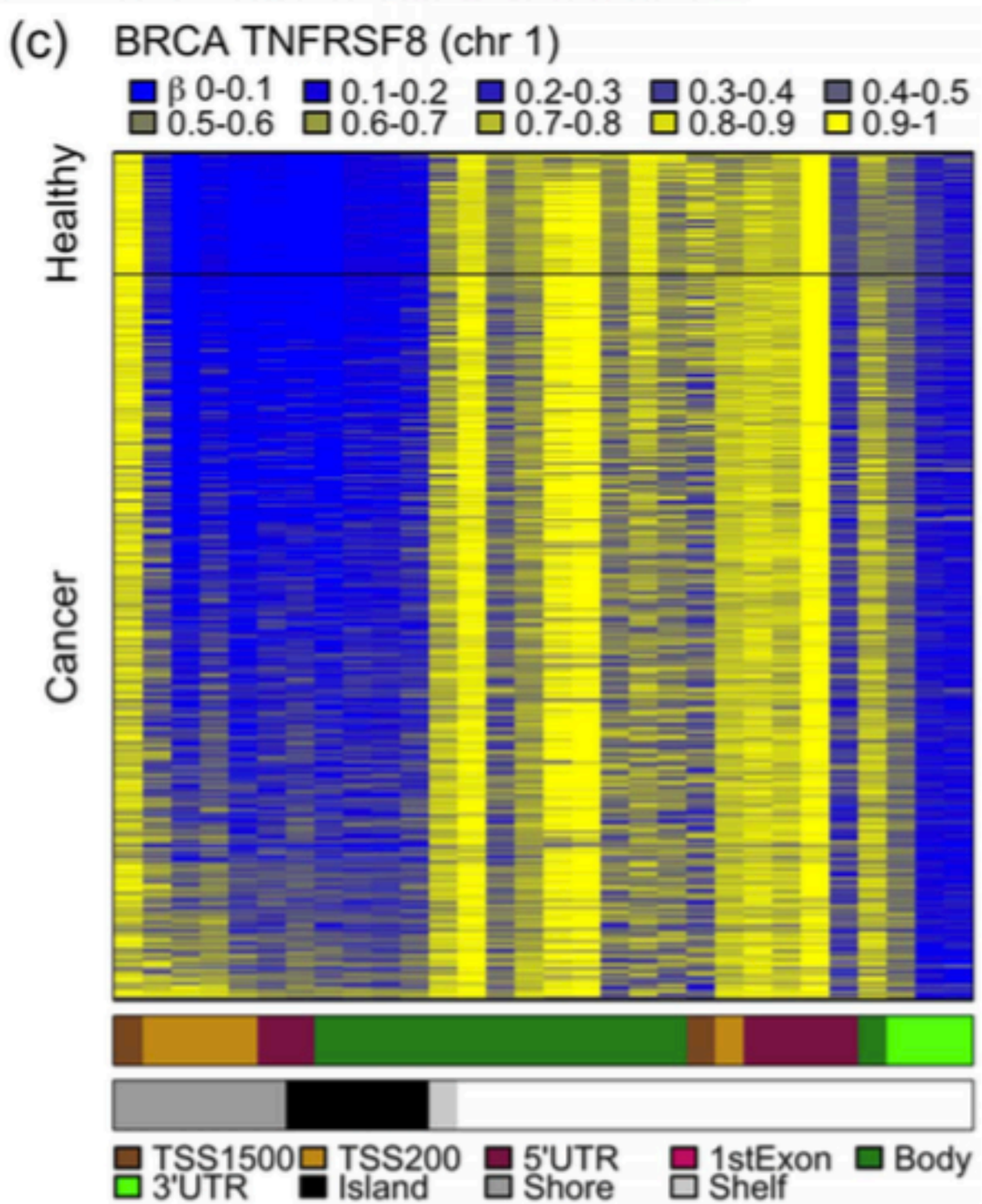
OPEN ACCESS Freely available online

PLOS ONE

Corruption of the Intra-Gene DNA Methylation Architecture Is a Hallmark of Cancer

Thomas E. Bartlett[1,2], Alexey Zaikin[2], Sofia C. Olhede[1,3], James West[1,4], Andrew E. Teschendorff[4], Martin Widschwendter[5]*





DNA Methylation

OPEN ACCESS Freely available online

PLOS ONE

A DNA Methylation Network Interaction Measure, and Detection of Network Oncomarkers

Thomas E. Bartlett[1,3]*, Sofia C. Olhede[2,3], Alexey Zaikin[1]

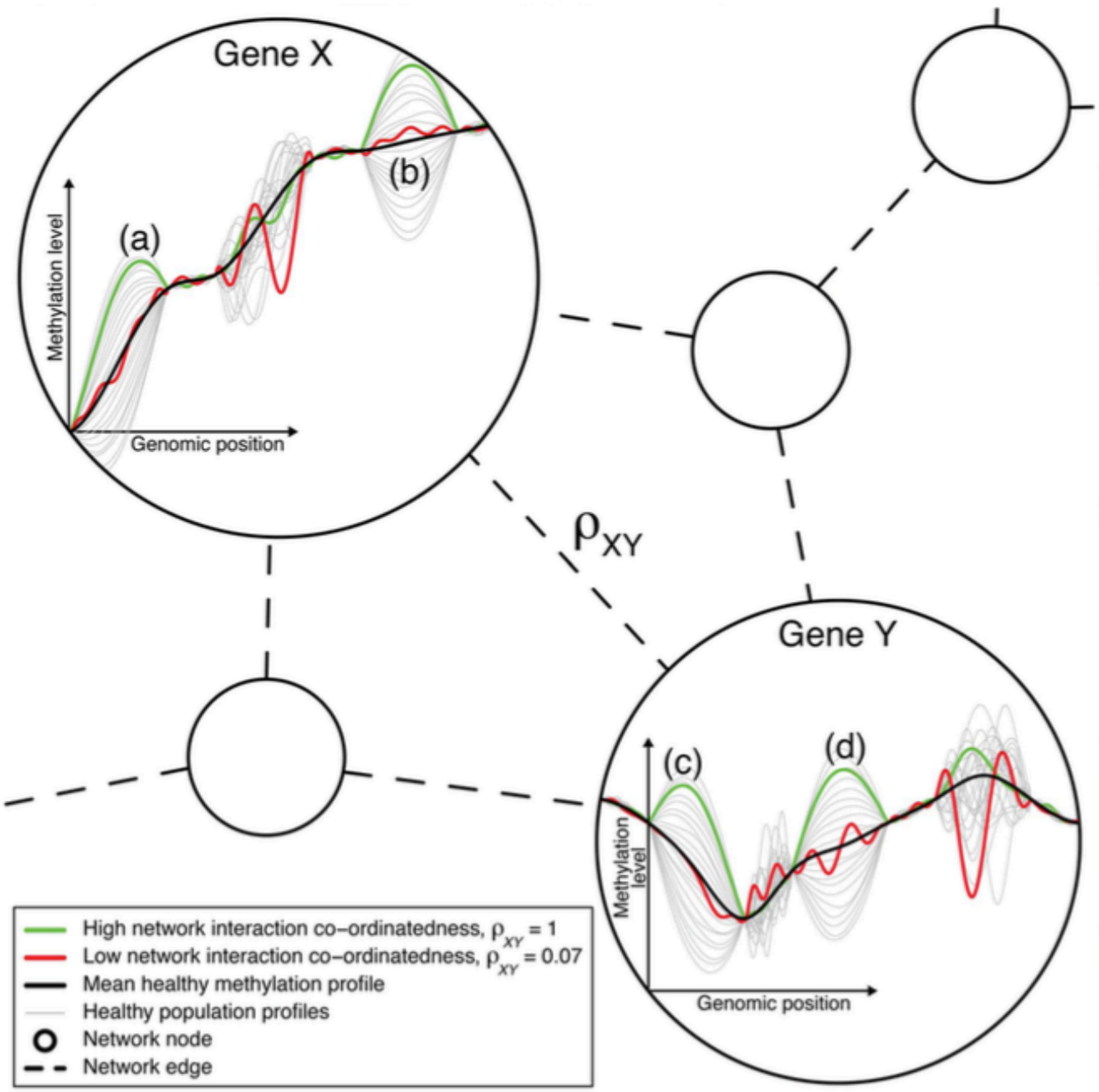


Figure 1. The DNA methylation network interaction measure. A combination of the variation of the healthy methylation profiles in regions (a) and (b) of gene X explains well/is well-explained by a combination of the variation of the healthy methylation profiles in regions (c) and (d) of gene Y. The green cancer sample varies by a large amount about the mean methylation profile and in a typical way in these regions in both genes. Hence, the green sample corresponds to a high level of network interaction co-ordinatedness, as measured by the DNA methylation network interaction measure, $\rho_{XY} = 1$. The variation in the other regions of these genes do not well-explain each other, and so the red sample, which varies by a large amount in these other regions and varies less and in an atypical way in regions (a)–(d), corresponds to a low level of network interaction co-ordinatedness, $\rho_{XY} = 0.07$. Genes X and Y are likely to have different numbers of methylation measurement locations (i.e., variables X and Y are of different dimension). The ordering of the measurement locations has no influence on the calculation of ρ, as long as the ordering is consistent across samples.

OPEN ACCESS Freely available online

PLOS ONE

A DNA Methylation Network Interaction Measure, and Detection of Network Oncomarkers

Thomas E. Bartlett[1,3]*, Sofia C. Olhede[2,3], Alexey Zaikin[1]

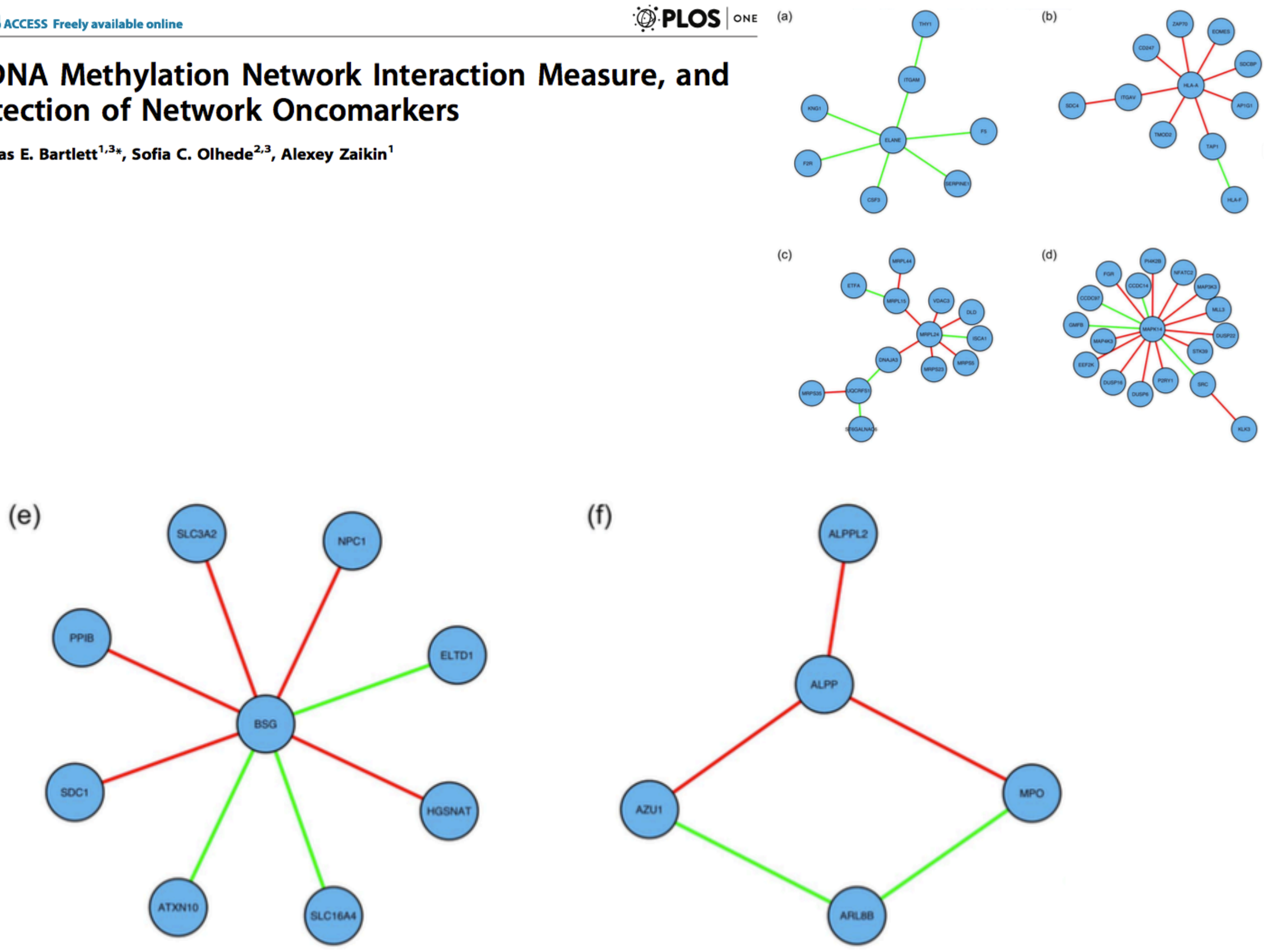


Figure 3. Smaller significant network modules: network diagrams. Network edges displayed in green and red indicate positive and negative hazard ratios, respectively, for the DNAm network correlation measure corresponding to that interaction; these correspond, respectively, to an

DNA Methylation Analysis

OPEN ACCESS Freely available online

PLOS ONE

A DNA Methylation Network Interaction Measure, and Detection of Network Oncomarkers

Thomas E. Bartlett[1,3]*, Sofia C. Olhede[2,3], Alexey Zaikin[1]

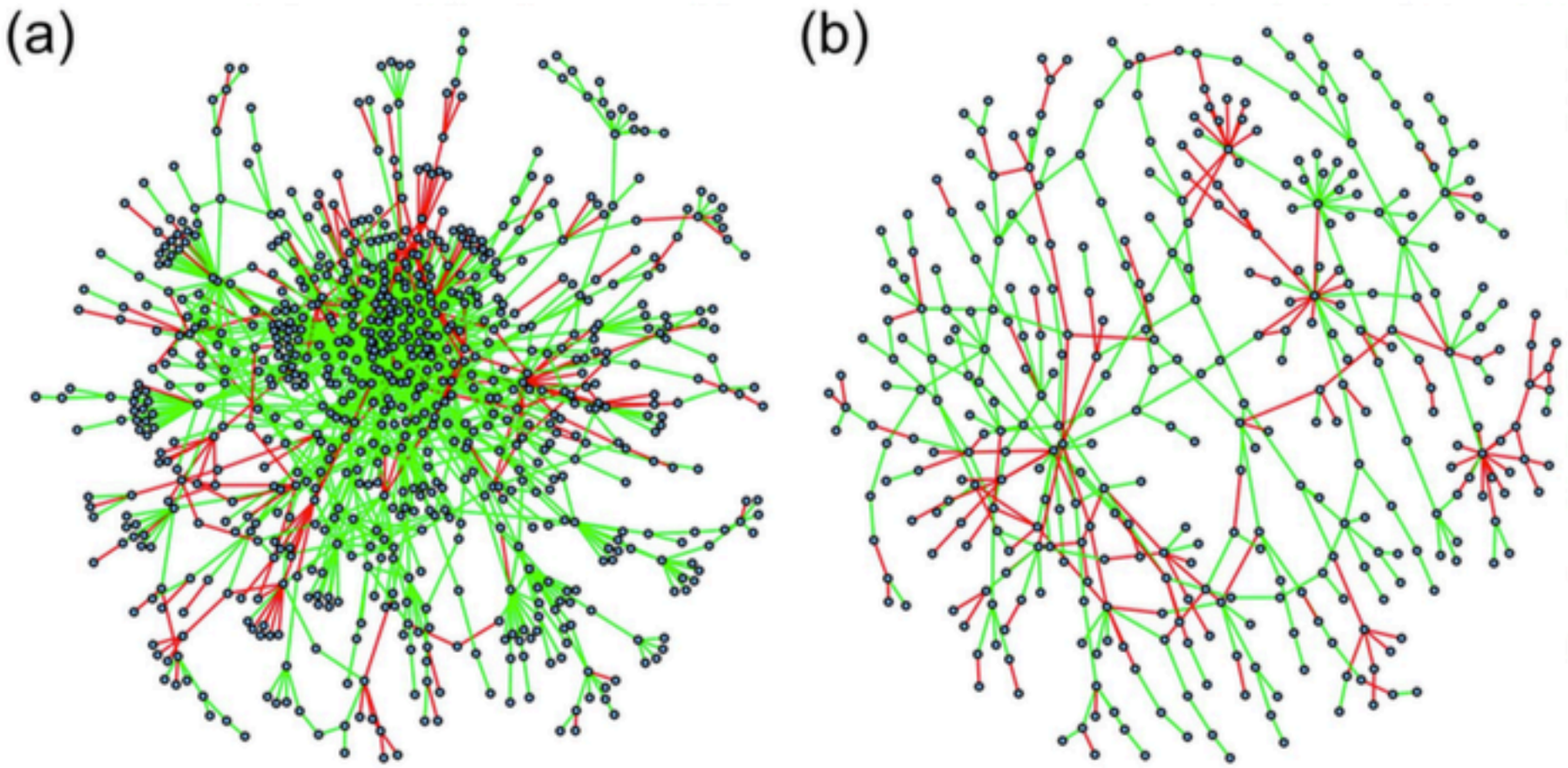


Figure 4. Larger significant subnetworks: network diagrams. Network edges displayed in green and red indicate positive and negative hazard ratios, respectively, for the DNAm network correlation measure corresponding to that interaction; these correspond, respectively, to an increase and decrease in 'network interaction co-ordinatedness' for worse disease prognosis. (a) the KIRC large subnetwork. (b) the LUAD large subnetwork. Further details about the corresponding network nodes (genes) for the top 5% of the degree distribution and top 25 significantly enriched gene sets appear in tables S5–6.

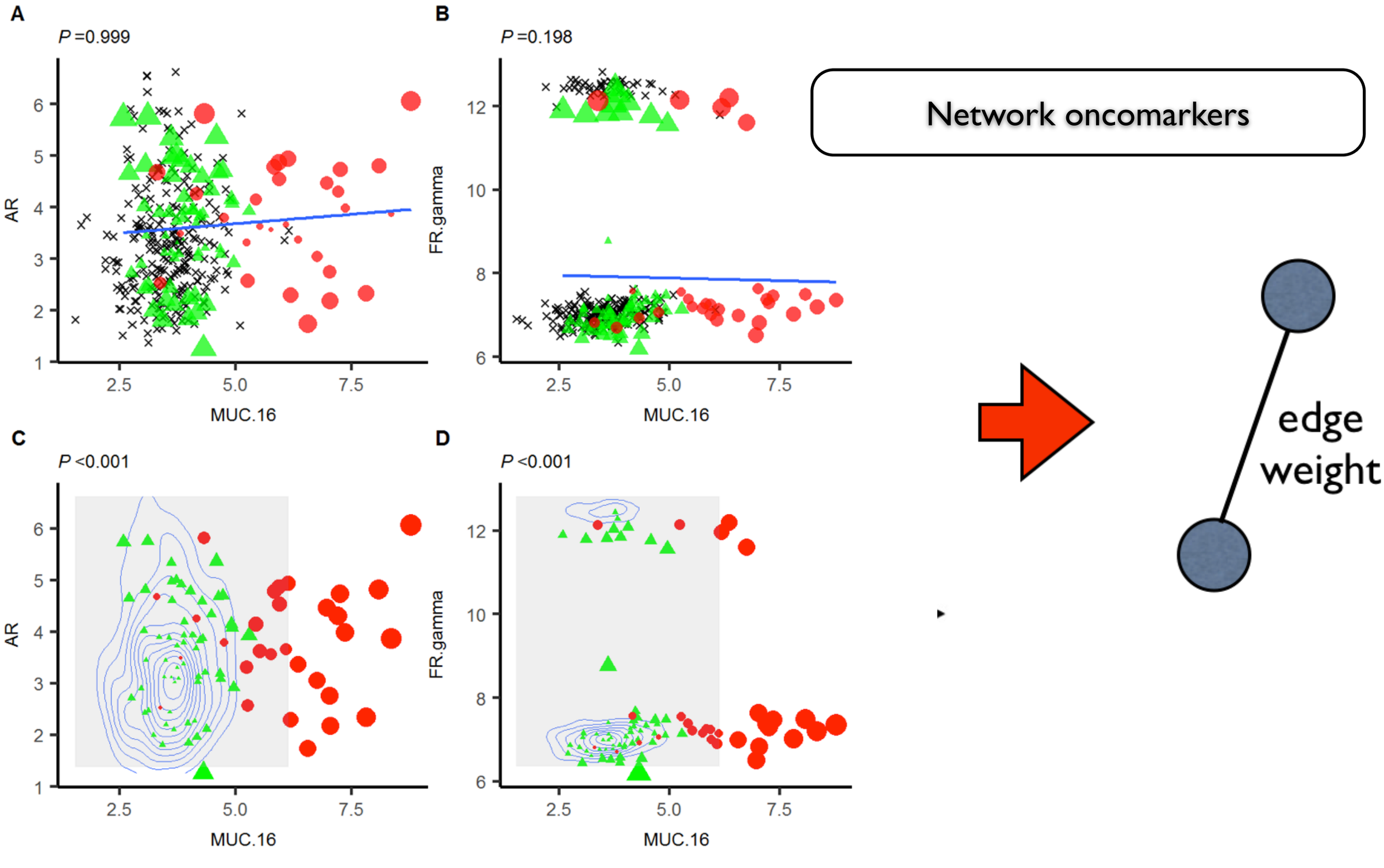
How to build a network if links are unknown?

www.oncotarget.com Oncotarget, 2018, Vol. 9, (No. 32), pp: 22717-22726

Research Paper

Parenclitic networks for predicting ovarian cancer

Harry J. Whitwell[1], Oleg Blyuss[2], Usha Menon[3], John F. Timms[3] and Alexey Zaikin[3,4]



PLOS ONE

RESEARCH ARTICLE

Parenclitic Network Analysis of Methylation Data for Cancer Identification

Alexander Karsakov[1], Thomas Bartlett[2], Artem Ryblov[1], Iosif Meyerov[3], Mikhail Ivanchenko[1], Alexey Zaikin[1,2]*

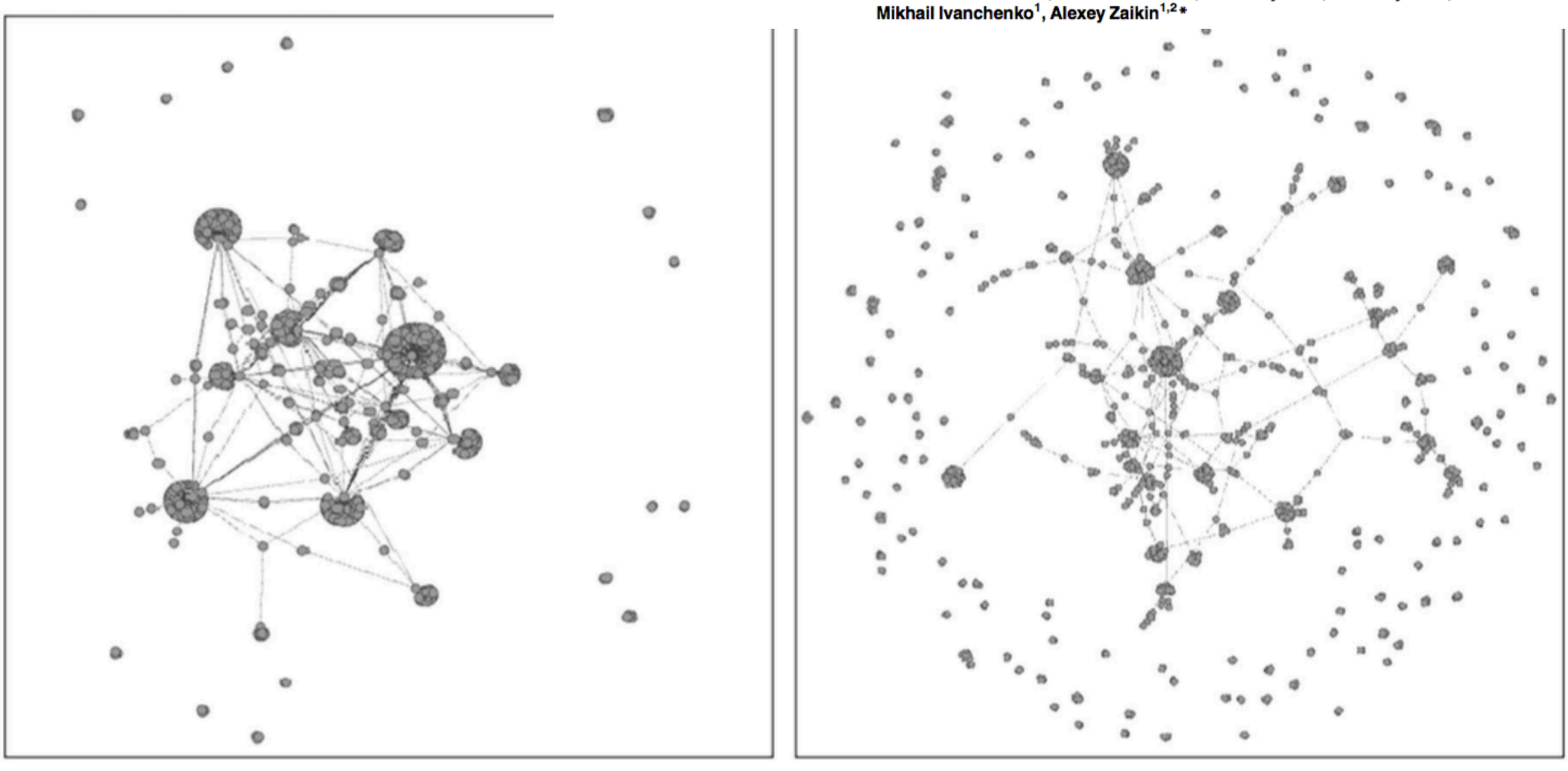


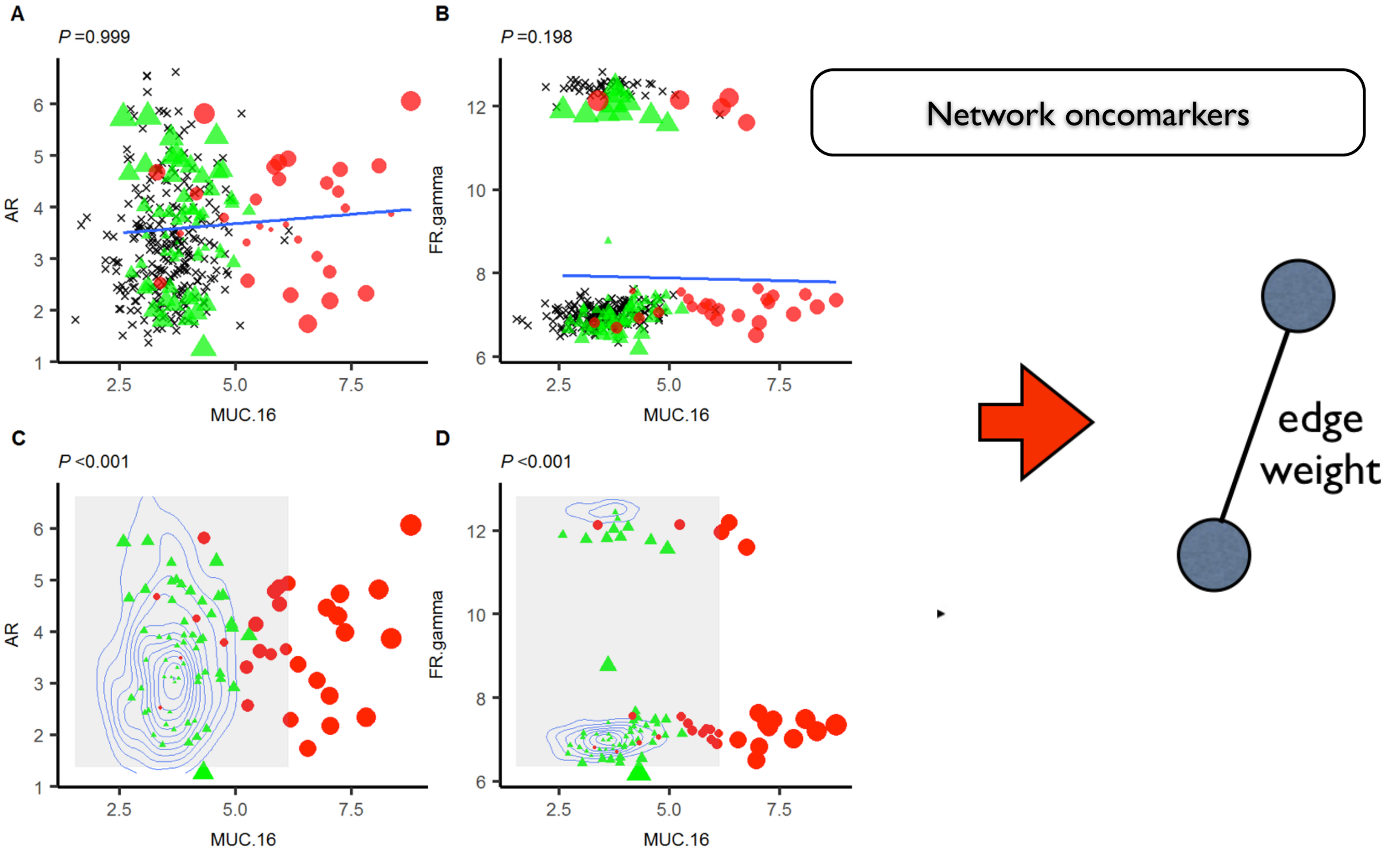
Fig 3. Typical examples of parenclictic networks constructed from gene methylation profiles for cancer (left) and normal (right) samples from BRCA data. Only a 1000 of the strongest edges and their incident nodes are shown. Note the pronounced modular structure for the cancer network.

www.oncotarget.com Oncotarget, 2018, Vol. 9, (No. 32), pp: 22717-22726

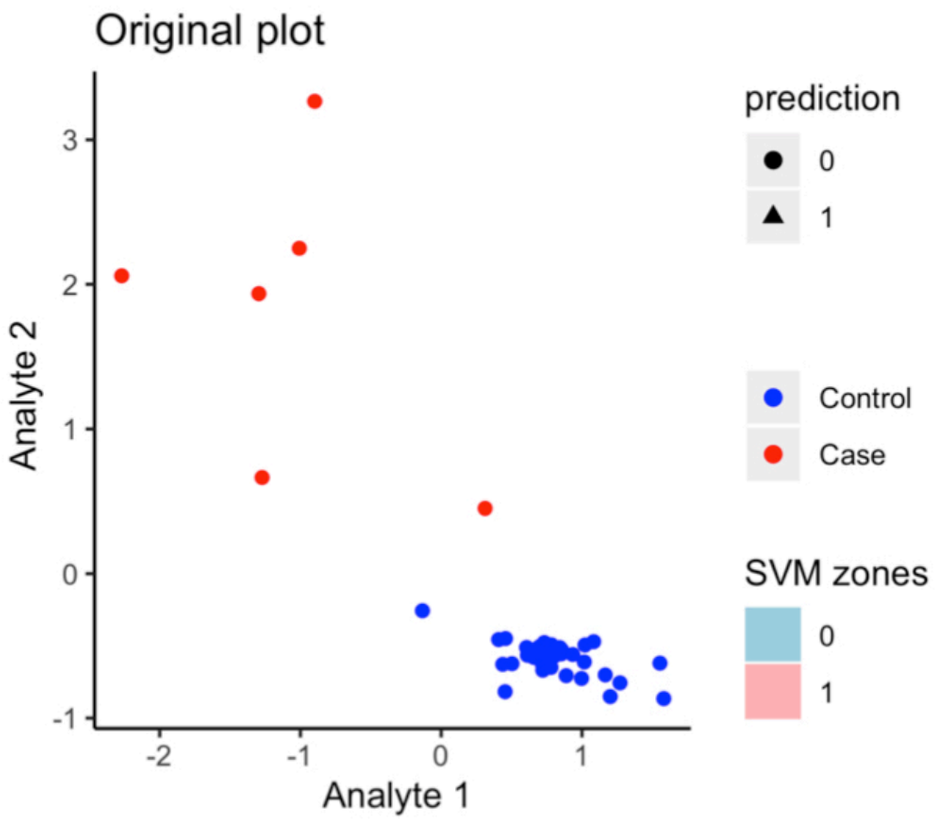
Research Paper

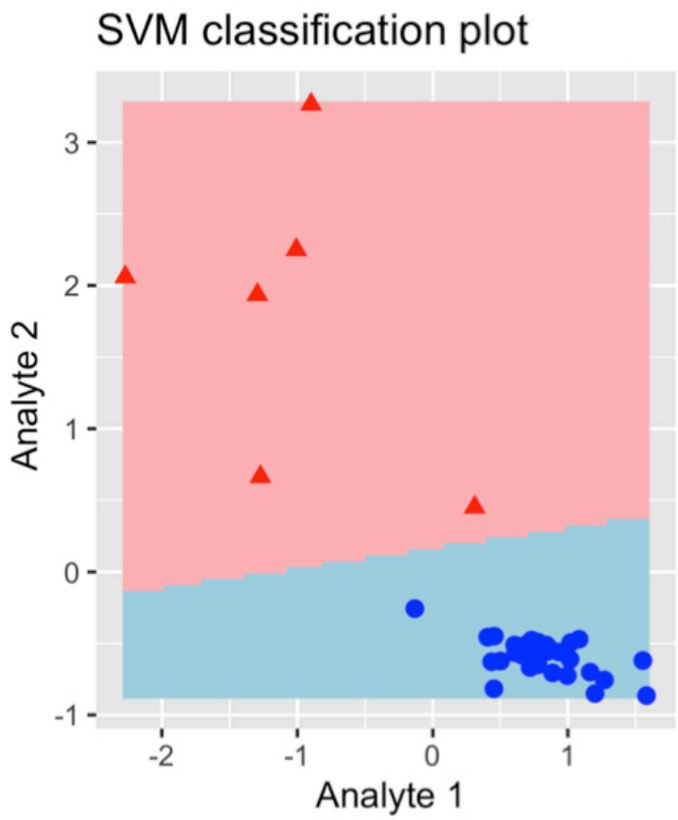
Parenclitic networks for predicting ovarian cancer

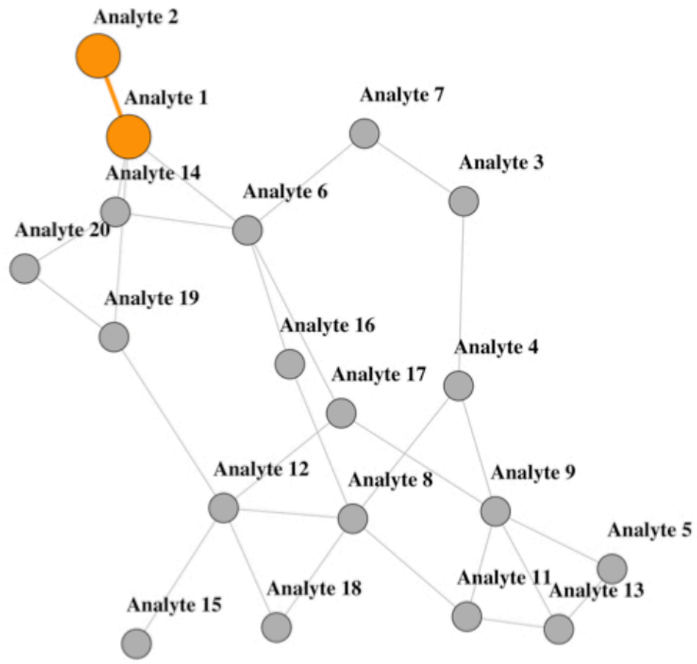
Harry J. Whitwell[1], Oleg Blyuss[2], Usha Menon[3], John F. Timms[3] and Alexey Zaikin[3,4]



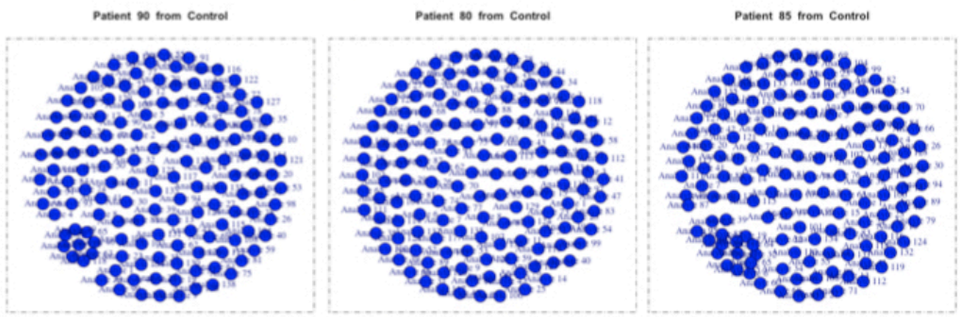
Parenclitic Network Construction (SVM approach as an example)



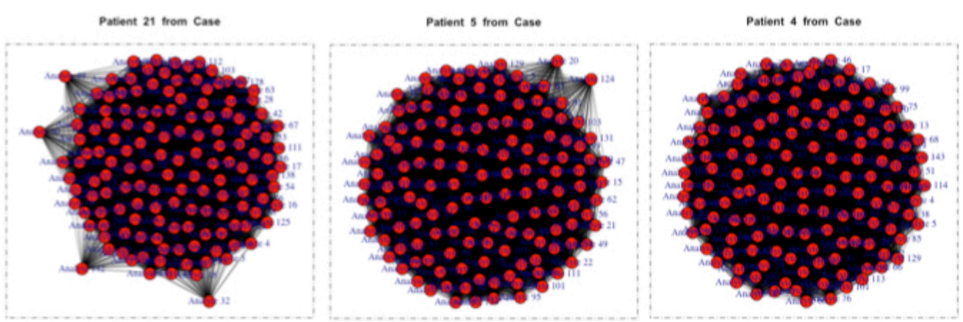




Examples of Controls

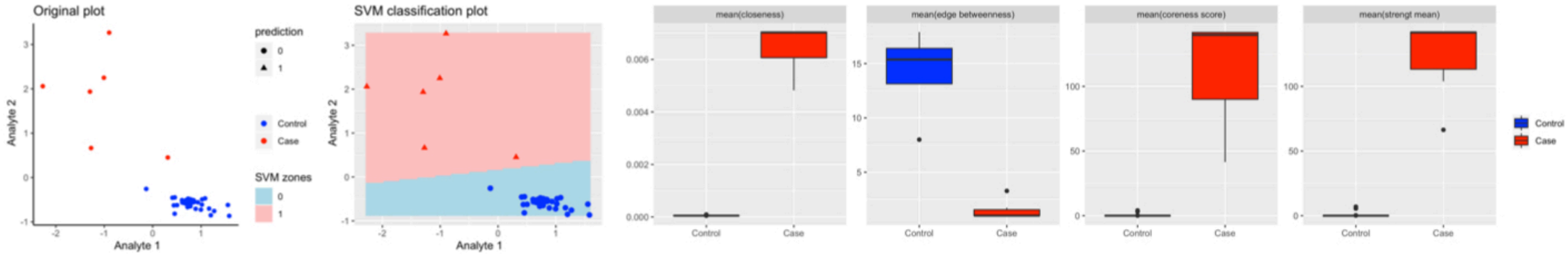


Examples of Cases

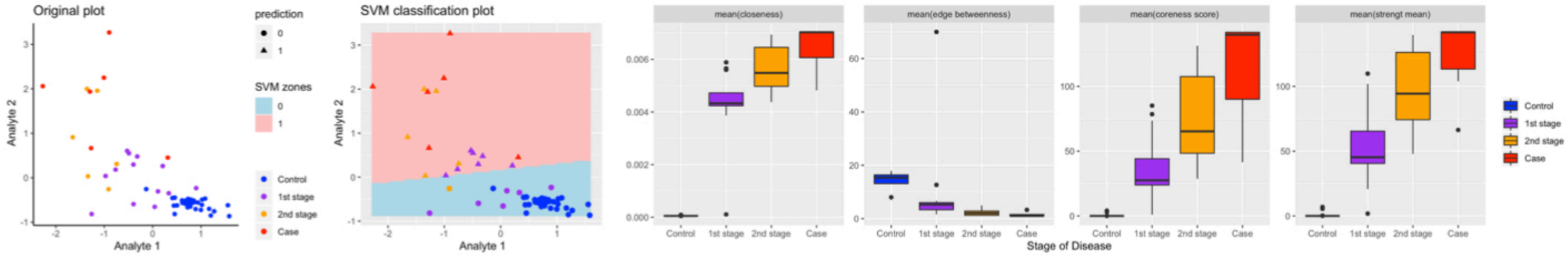


Examples of problems that can be solved using networks characteristics

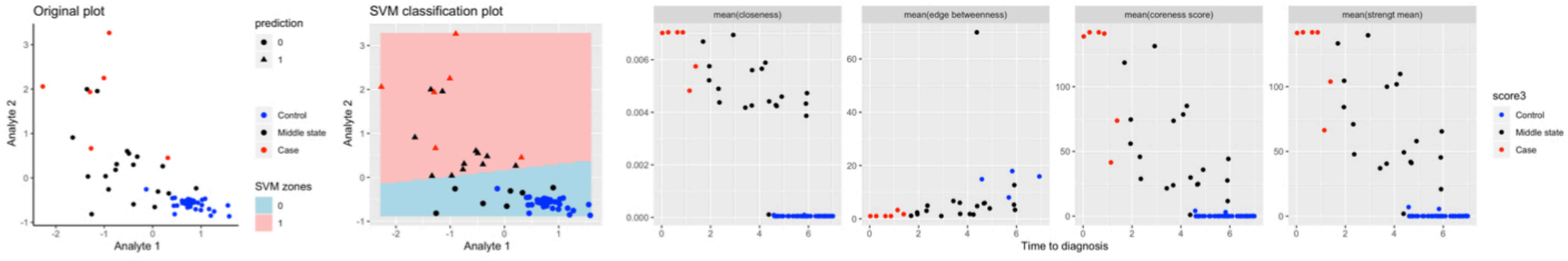
- Case/Control Classification (e.g. Disease/Health)



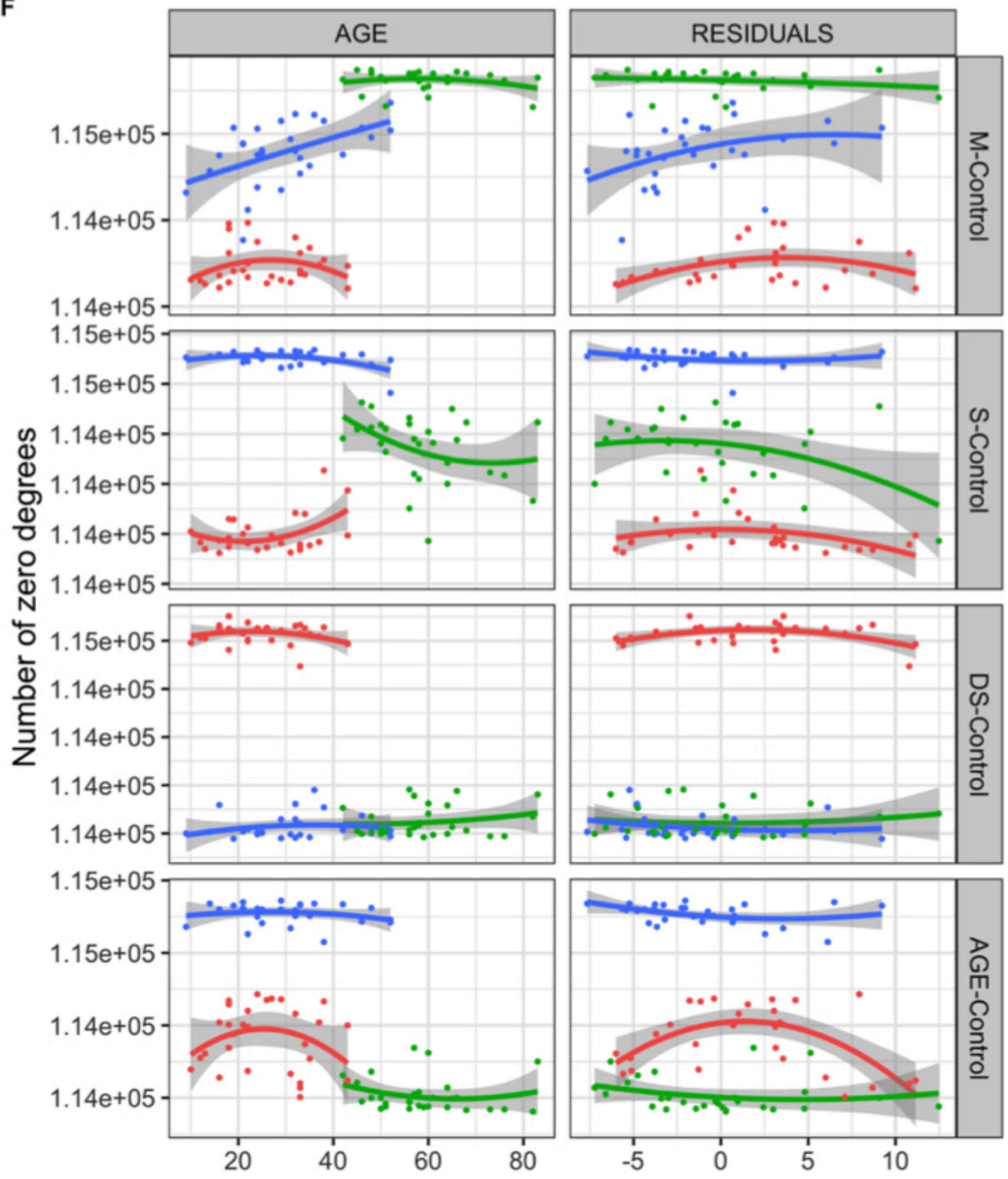
- Detection of *discrete state between* Case and Control (e.g. Stage of Disease)



- Detection of *continuous state between* Case and Control (e.g. time to Diagnosis)



F



Age-related changes in the network topology of DNA methylation probes: a parenclitic network approach to a family-based cohort of patients with Down Syndrome

M. Krivonosov[1], T. Nazarenko[2,*], M.G. Bacalini[3], C. Franceschi[1,3], A. Zaikin[1,2,4], and M. Ivanchenko[1]

1. **M-Control Network** — DSM group is *Control* group, DS group is *Case* group, DSS group is *Test* group;
2. **S-Control Network** — DSS group is *Control* group, DS group is *Case* group, DSM group is *Test* group;
3. **AGE-Control Network** — DSS group is *Control* group, DSM group is *Case* group, DS group is *Test* group;
4. **DS-Control Network** — DS group is *Control* group, DSS and DSM groups sre *Case* group.

Dynamic predictive model for baseline early detection and follow- up re-evaluation of the risk of prostate cancer progression on active surveillance (PROGRESS Prostate)

Prostate cancer (PCa) is the second commonest male cancer worldwide, with 43% of patients being offered active surveillance (AS) as an alternative to radical treatment. However, a five-year dropout rate of 44% with pathological upgrading of 27% of re-biopsied cases within the first year of AS highlight the lack of robust risk-stratification models enabling both early detection and continuous re-evaluation of individualised progressive potential of PCa. With this in mind, here we present a CRUK ACED collaboration that seeks to develop a personalised dynamic predictive model able to estimate the risk of PCa progression throughout the whole AS continuum starting from the initial appointment. To develop the model, we will use a range of novel high-performance modelling methodologies utilising serial biomarker measurements, which we have previously developed and trialed as part of UKCTOCS trial.

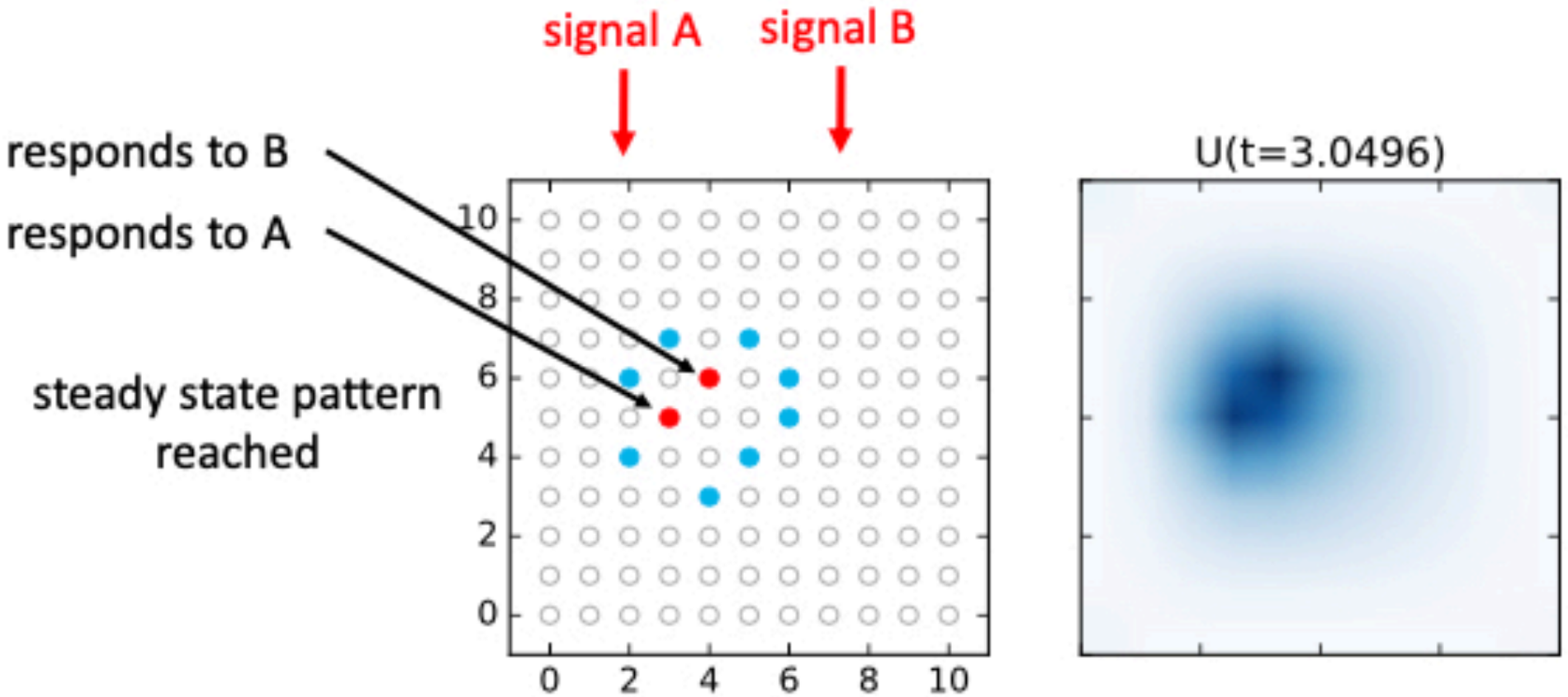
Deep Learning with Recurrent Artificial Neural Networks. We have recently shown that one of the most widely used deep learning techniques, recurrent neural networks (RNN), is capable of predicting the risk of developing ovarian cancer based on measurements of serial CA125. (26) To use RNN as part of this project, we have developed a neural network architecture based on state- of-the-art AI findings and using Long Short Term Memory RNNs, especially RNNs with Gated Recurrent Unit, and shown further improvement of cancer early detection (unpublished data). In the current project, we will further adapt this architecture for a further improvement of a risk- stratification model, potentially integrating findings from the MMT and BCP approaches with the RNN.

Genetic intelligence (student project)

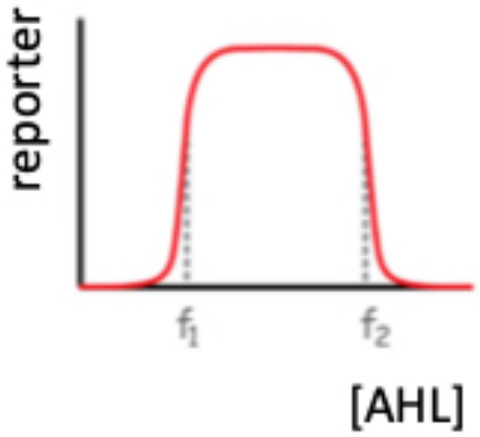
Building biological computers from bacterial populations

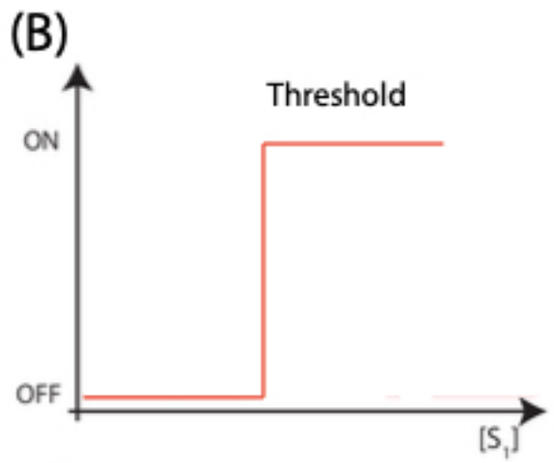
UCL

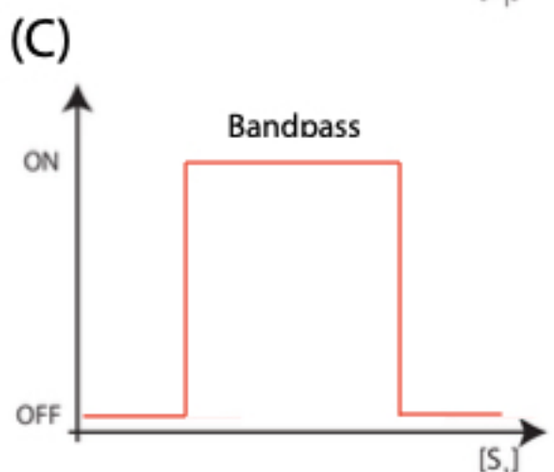
Neythen Treloar
Supervisor: Chris Barnes



receiver colonies contain a band detector that responds to AHL and switches on/off



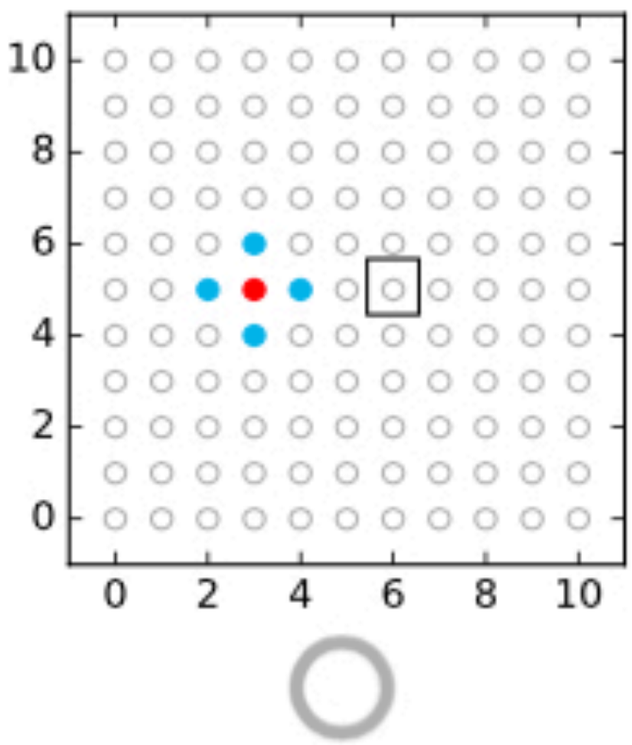




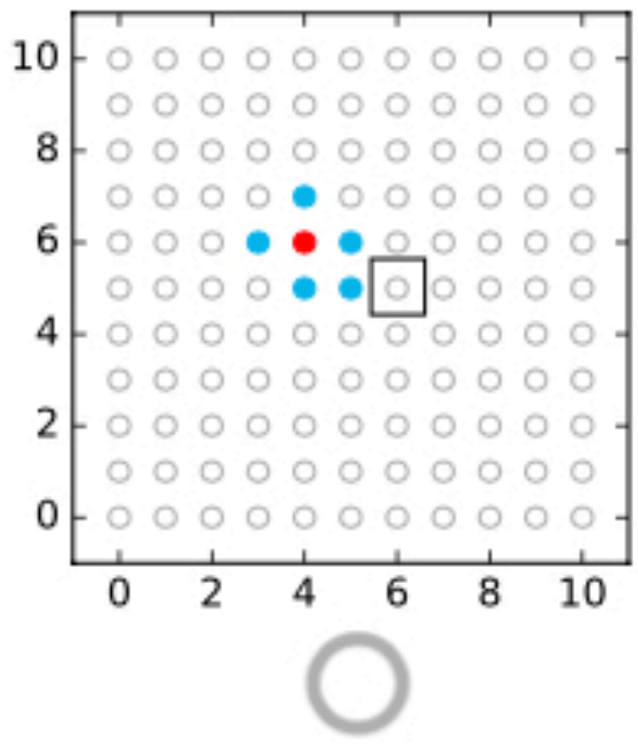
Main idea:

Spatial pattern can be used for signal integration and computation

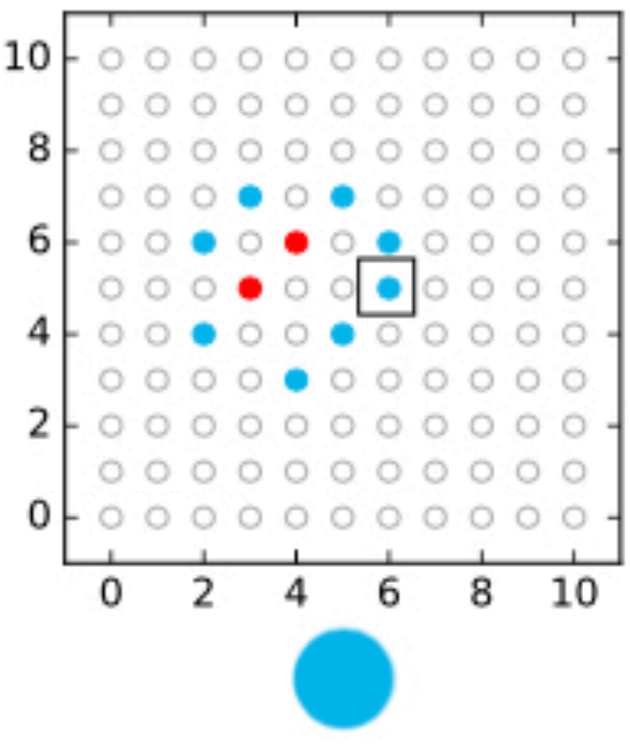
B on, A off



B off, A on



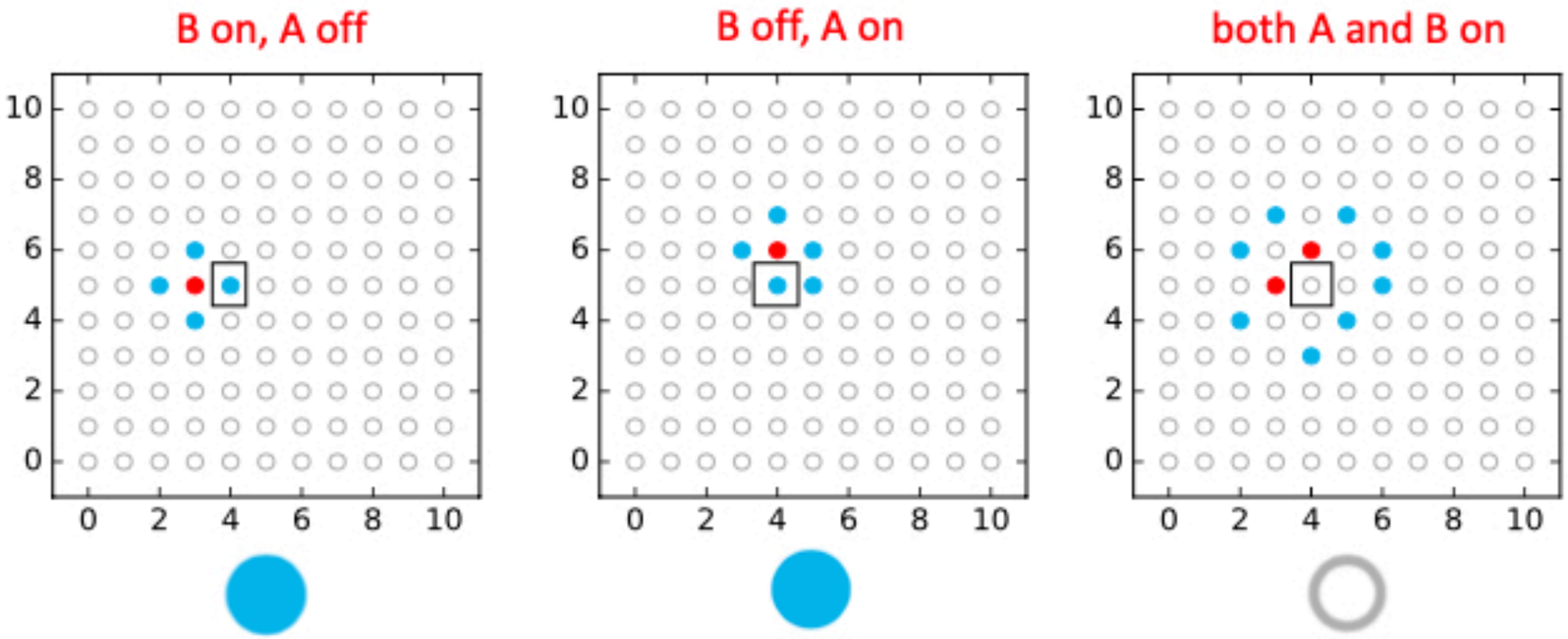
both A and B on



A	B	AND
0	0	0
0	1	0
1	0	0
1	1	1

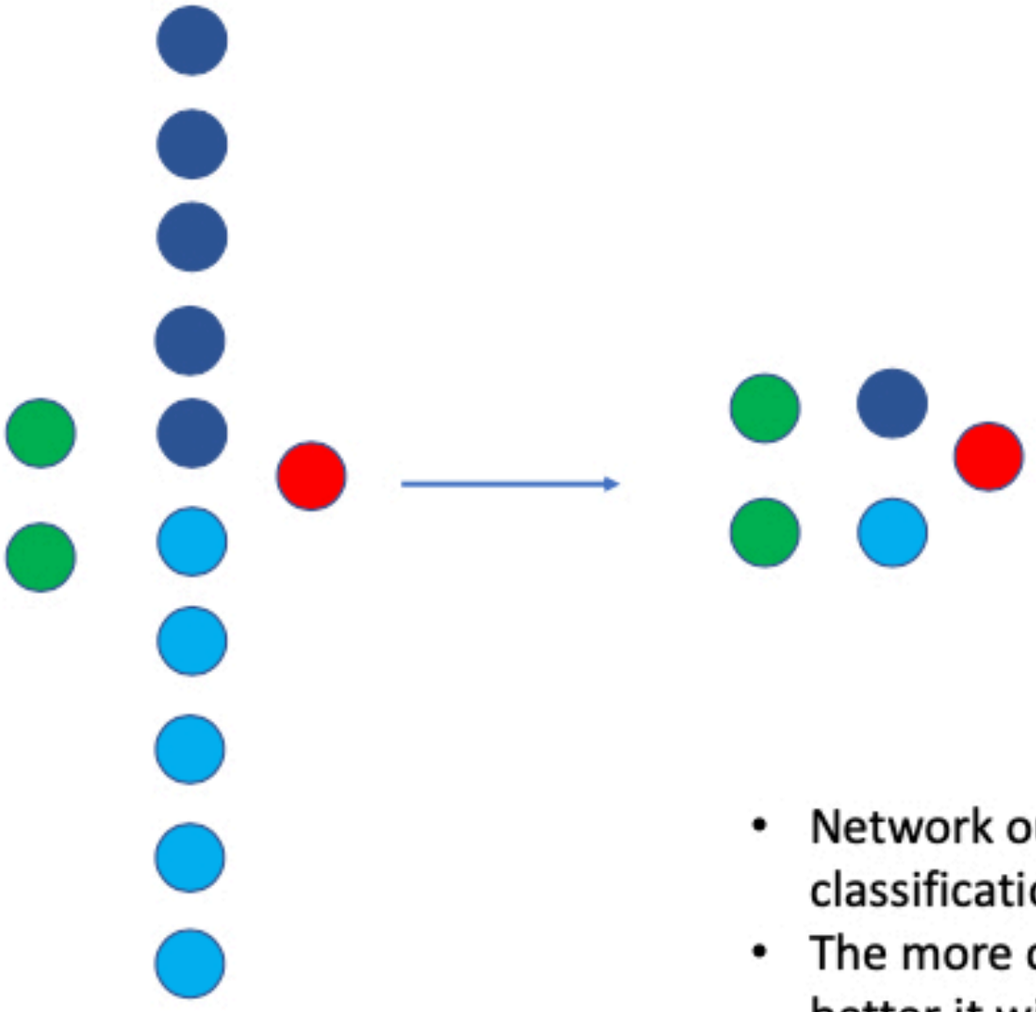
Main idea:

Spatial pattern can be used for signal integration and computation

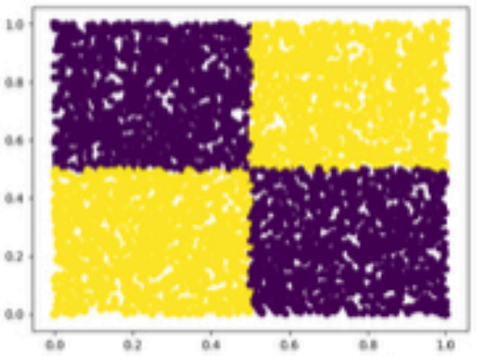


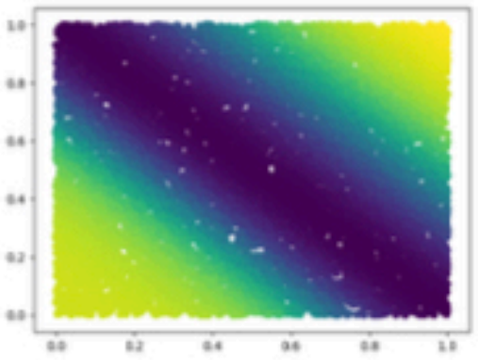
A	B	XOR
0	0	0
0	1	1
1	0	1
1	1	0

Network trained with these constraints using regularization to encourage sparsity.



XNOR

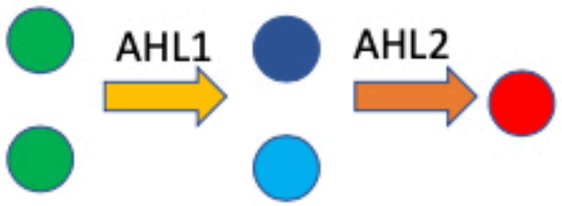


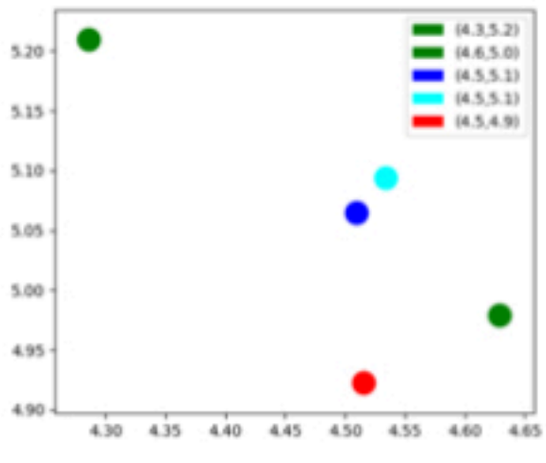


- Network output: enough for digital classification
- The more constraints we can overcome the better it will perform

Weights extracted from trained network and evolutionary algorithm used to find node position

- This assumes we have two AHLs





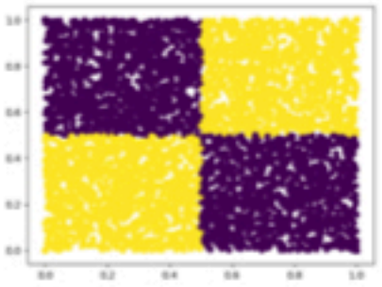
Target:
[[3.6815872, 3.7029088],
[6.8024187, 6.783618]]

[7.0231166,5.7998238]

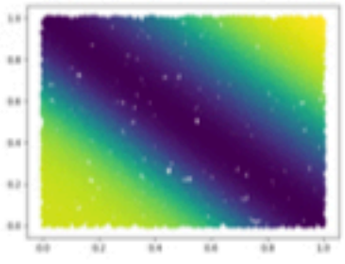
Solution:
[[3.6642652], [3.6659336],
[6.80611], [6.792463]]

[7.0273304], [5.795528]

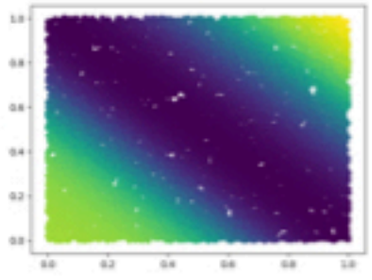
XNOR



Network trained with constraints



Weights recalculated from positions



Diffusion system

?

Main open question: longitudinal data analysis of multiplex networks with different space and time scales