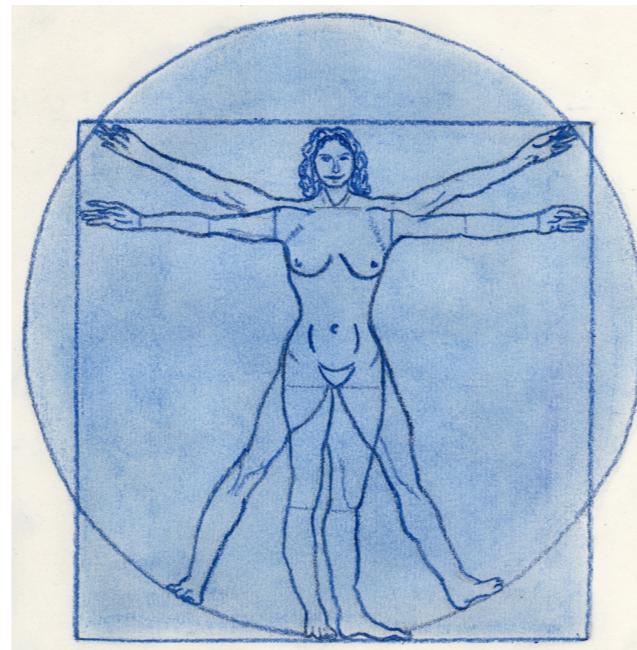


Longitudinal, Deep, and Network Biomarkers: Parenclitic and **Synolitic** Network Analysis



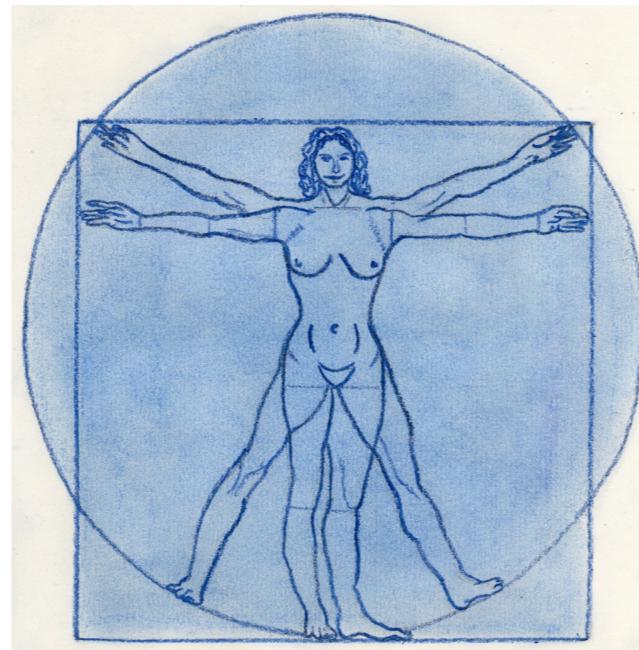
Alexey Zaikin

Institute for Women's Health and Department of Mathematics
University College London

www.zaikinlab.com

alexey.zaikin@ucl.ac.uk

How to Construct Network Biomarkers, if There is No Network??



Alexey Zaikin

Institute for Women's Health and Department of Mathematics
University College London

www.zaikinlab.com

alexey.zaikin@ucl.ac.uk

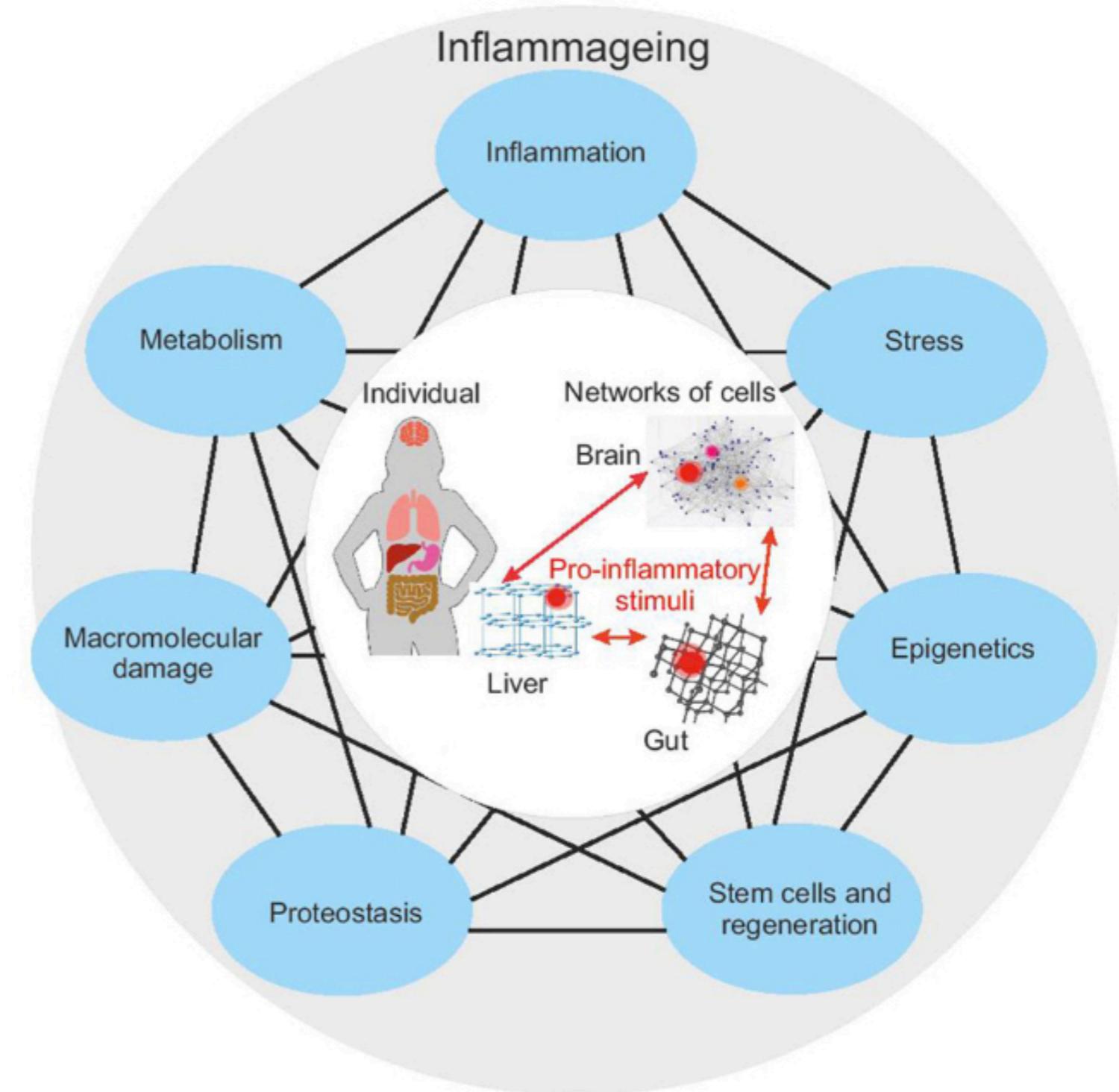


2020

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

Harry J. Whitwell¹, Maria Giulia Bacalini², Oleg Blyuss^{3,4}, Shangbin Chen⁵, Paolo Garagnani⁶, Susan Yu Gordleeva⁷, Sarika Jalan^{8,9}, Mikhail Ivanchenko¹⁰, Oleg Kanakov⁷, Valentina Kustikova¹⁰, Ines P. Mariño¹¹, Iosif Meyerov¹⁰, Ekkehard Ullner¹², Claudio Franceschi^{7,10*} and Alexey Zaikin^{4,10,13*}

Huge amount of -Omic Information: Genome, Chromatome, Methylome, Transcriptome, Proteome



V. Samborska et al. MAMMALIAN BRAIN AS A NETWORK...

OM&P

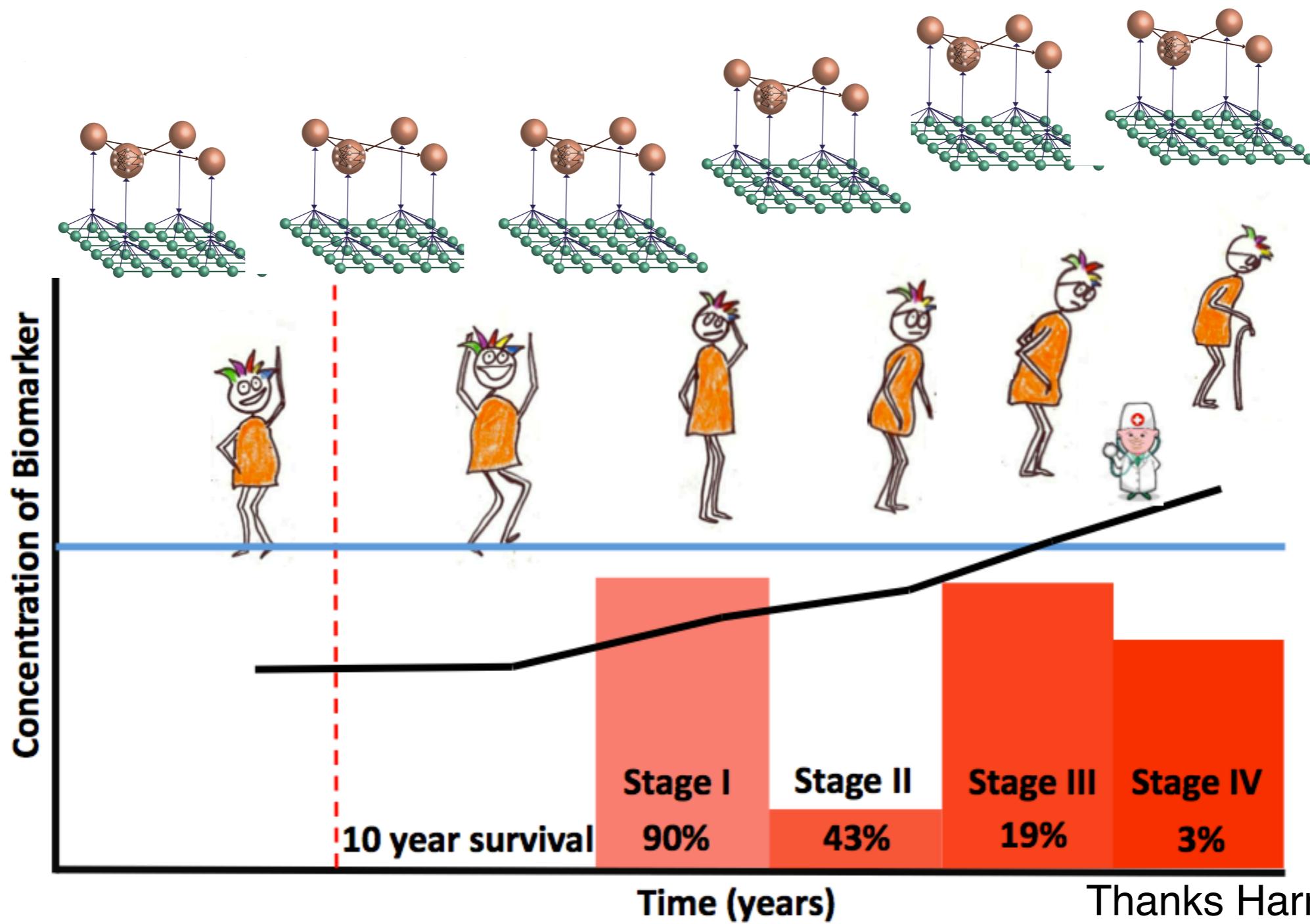
MAMMALIAN BRAIN AS A NETWORK OF NETWORKS

Veronika Samborska¹, Susanna Gordleeva², Ekkehard Ullner³, Albina Lebedeva², Viktor Kazantsev², Mikhail Ivanchenko² and Alexey Zaikin^{2,4}

Search for Network, Longitudinal and Deep Learning Biomarkers



Multi-omic data: Genetic, Epigenetic and Proteomic



Constructing network biomarkers

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

MAPPING THE CANCER GENOME

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”

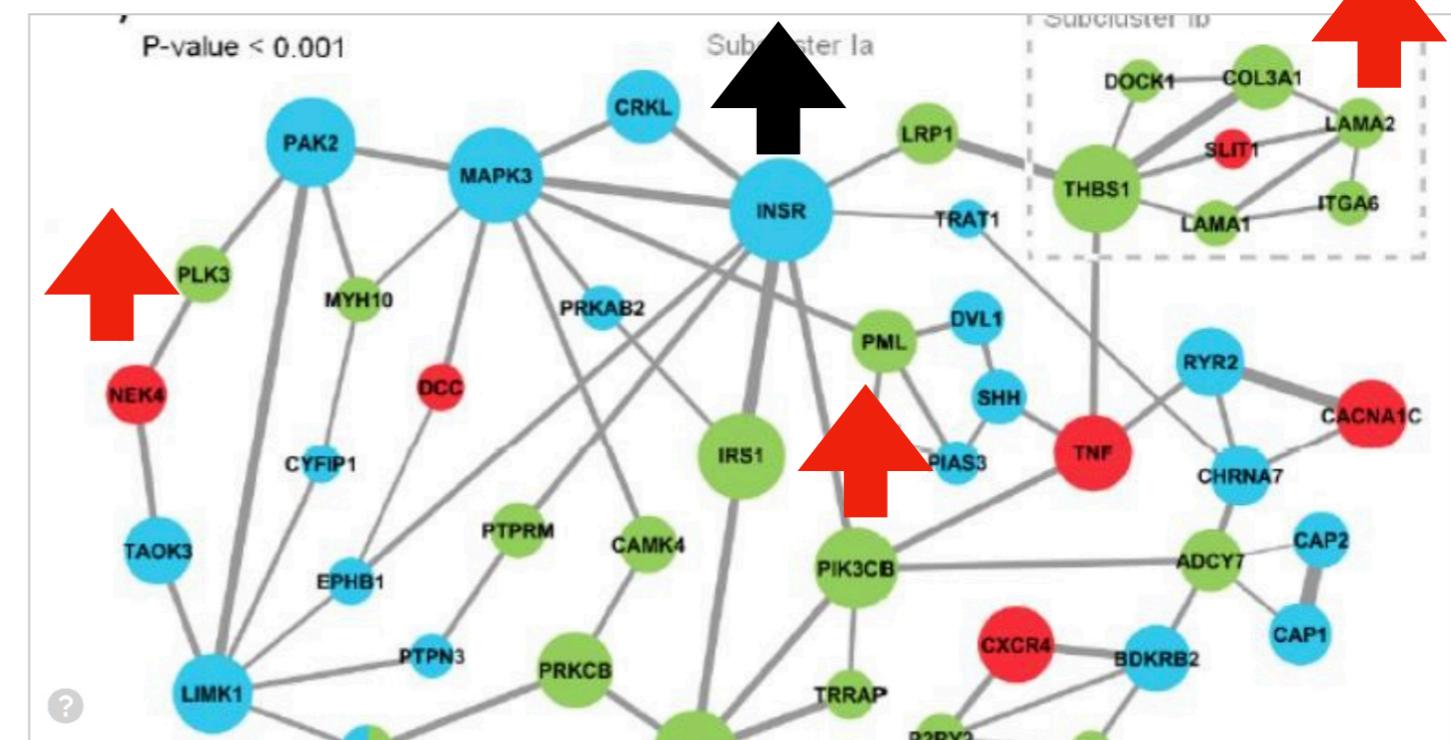
Solution: topological analysis of networks

Global Genetic Landscape Of The Cell (IMAGE) UNIVERSITY OF TORONTO

Heterogeneity

Schizophrenia Gene Networks Found, with Link to Autism

November 11, 2012



Solution: topological analysis of networks

How to construct a network if links are unknown??



ELSEVIER



Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 37 (2021) 17–64

**PHYSICS of LIFE
reviews**

www.elsevier.com/locate/plrev

Correlation graphs

Review

Dynamic and thermodynamic models of adaptation

A.N. Gorban ^{a,b,*}, T.A. Tyukina ^a, L.I. Pokidysheva ^c, E.V. Smirnova ^c



Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 38 (2021) 120–123

**PHYSICS of LIFE
reviews**

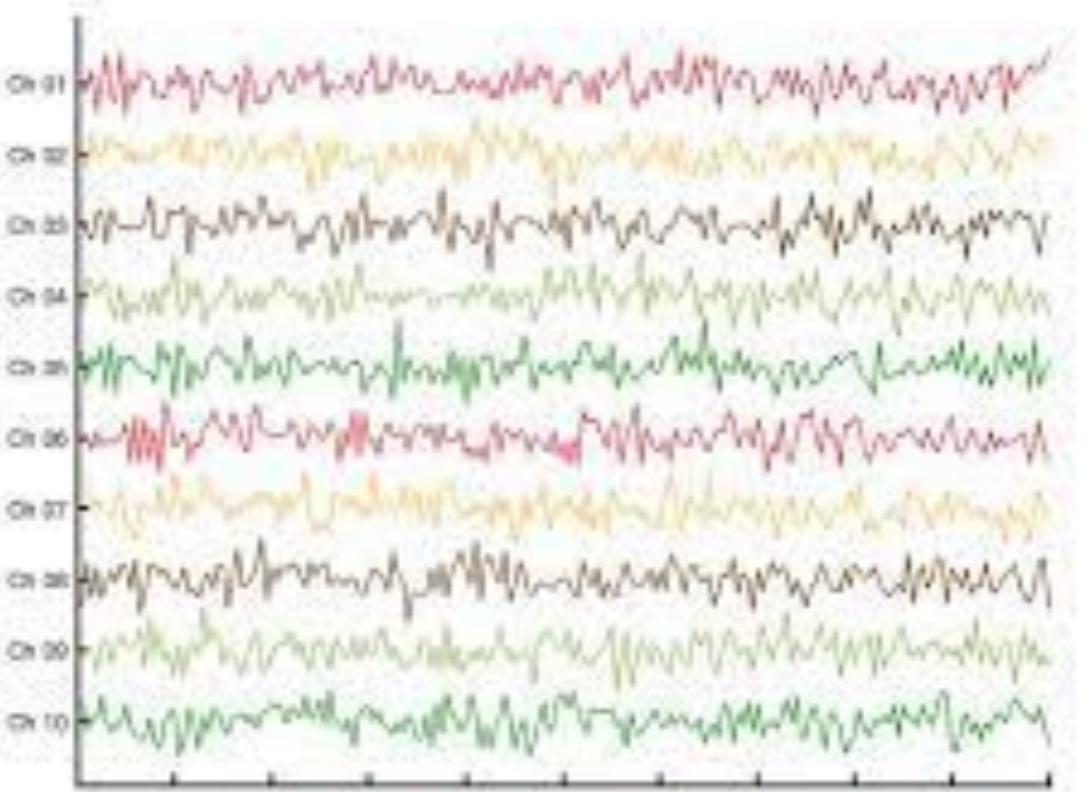
www.elsevier.com/locate/plrev

Comment

Ensemble of correlation, parenclitic and synolitic graphs as a tool to detect universal changes in complex biological systems

Comment on “Dynamic and thermodynamic models of adaptation”
by A.N. Gorban et al.

Tatiana Nazarenko ^a, Oleg Blyuss ^{a,b,e,g}, Harry Whitwell ^{c,d,e,f}, Alexey Zaikin ^{a,e,f,*}



3194

A.N. Gorban et al. / Physica A 389 (2010) 3193–3217

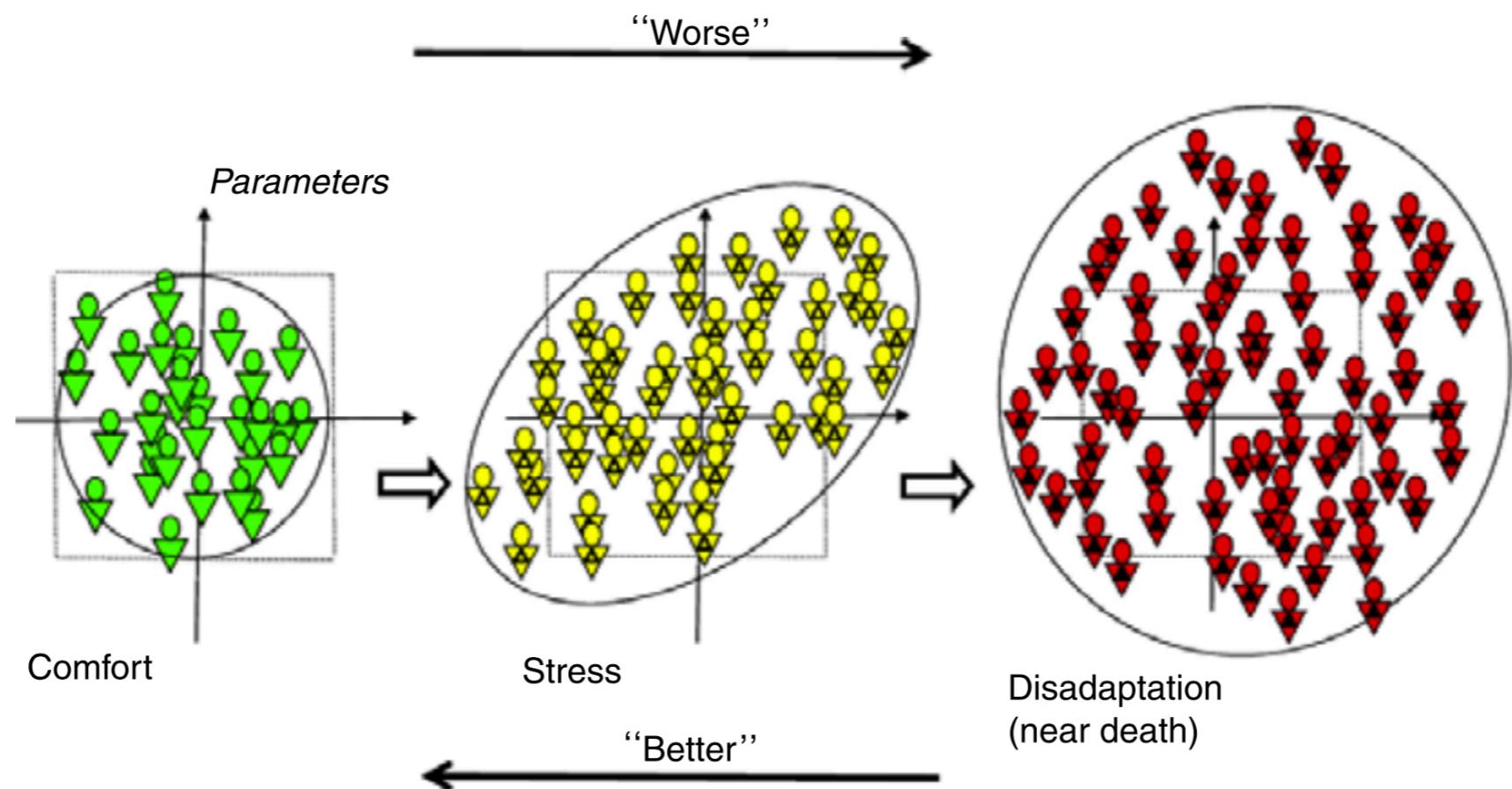
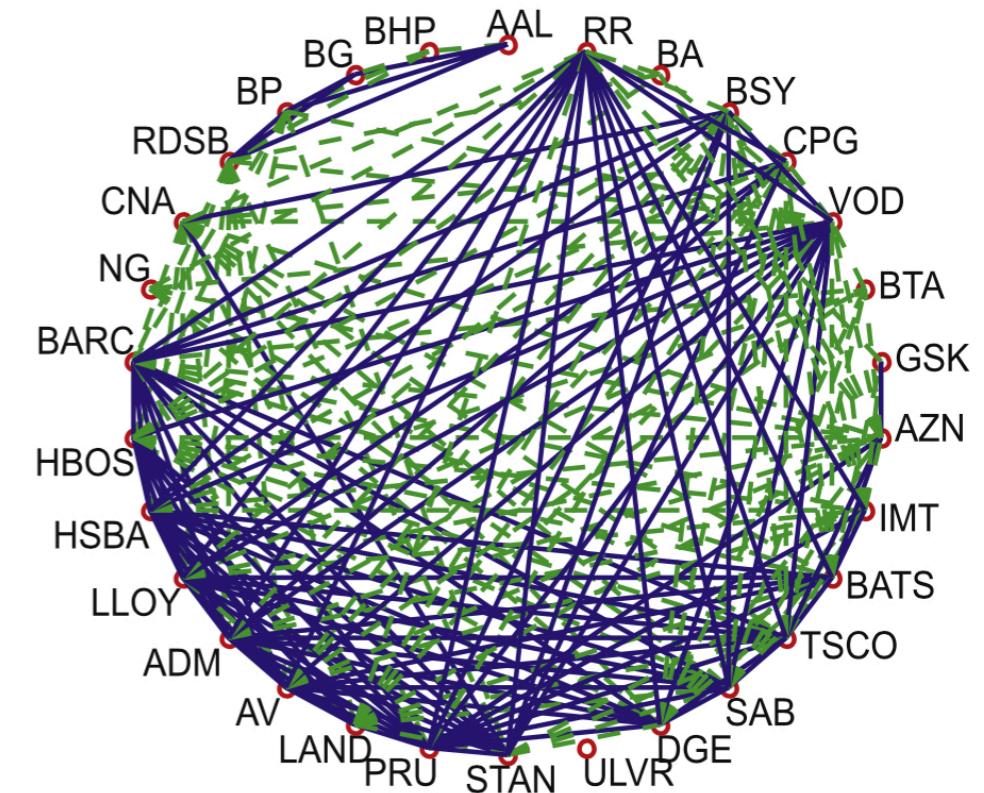


Fig. 1. Correlations and variance in crisis. The typical picture: $\text{Cor} \uparrow; \text{Var} \uparrow$ – stress; $\text{Cor} \downarrow; \text{Var} \downarrow$ – recovering; $\text{Cor} \downarrow; \text{Var} \uparrow$ – approaching the disadaptation catastrophe after the bottom of the crisis. In this schematic picture, axes correspond to attributes, normalized to the unite variance in the comfort state.

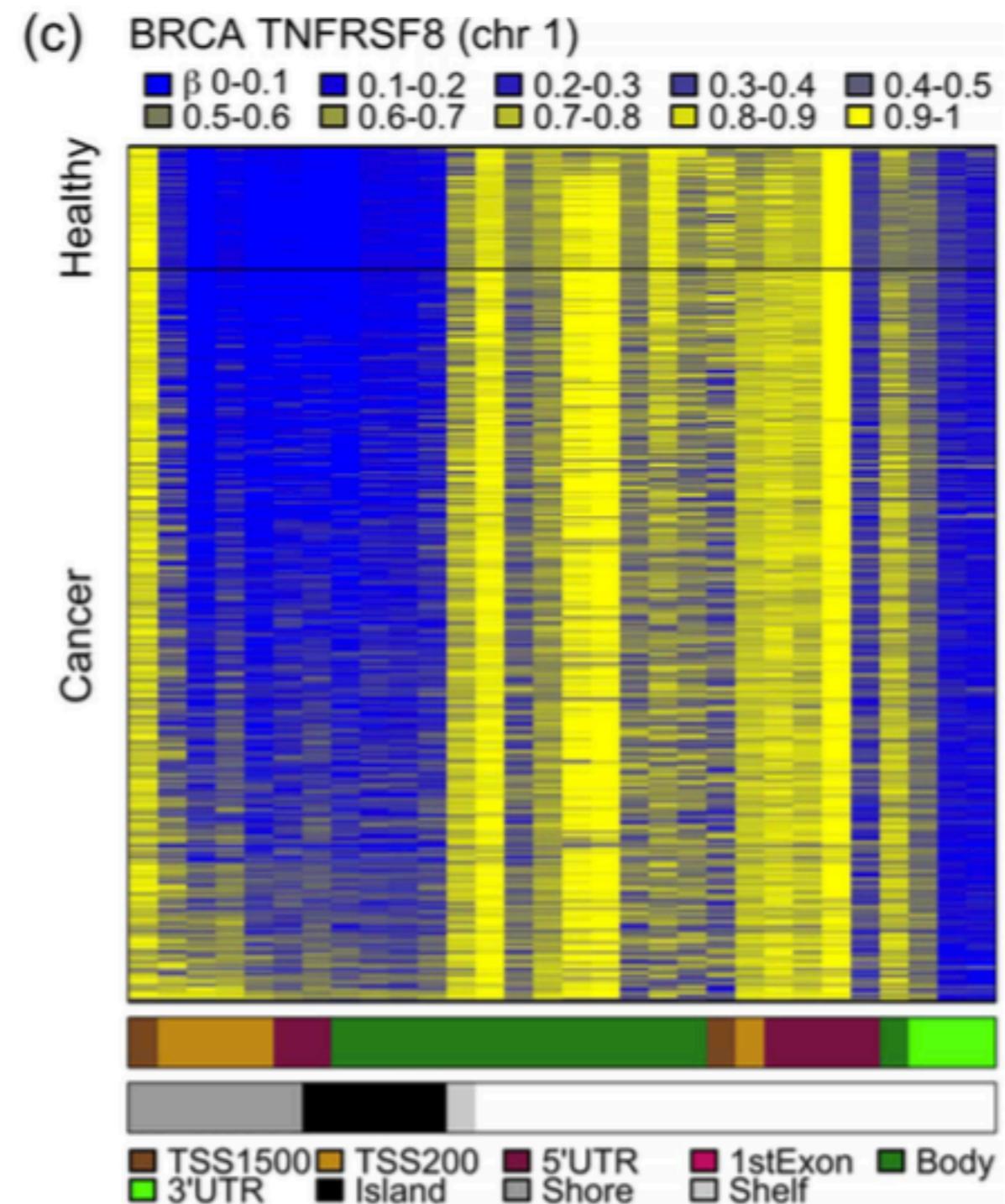
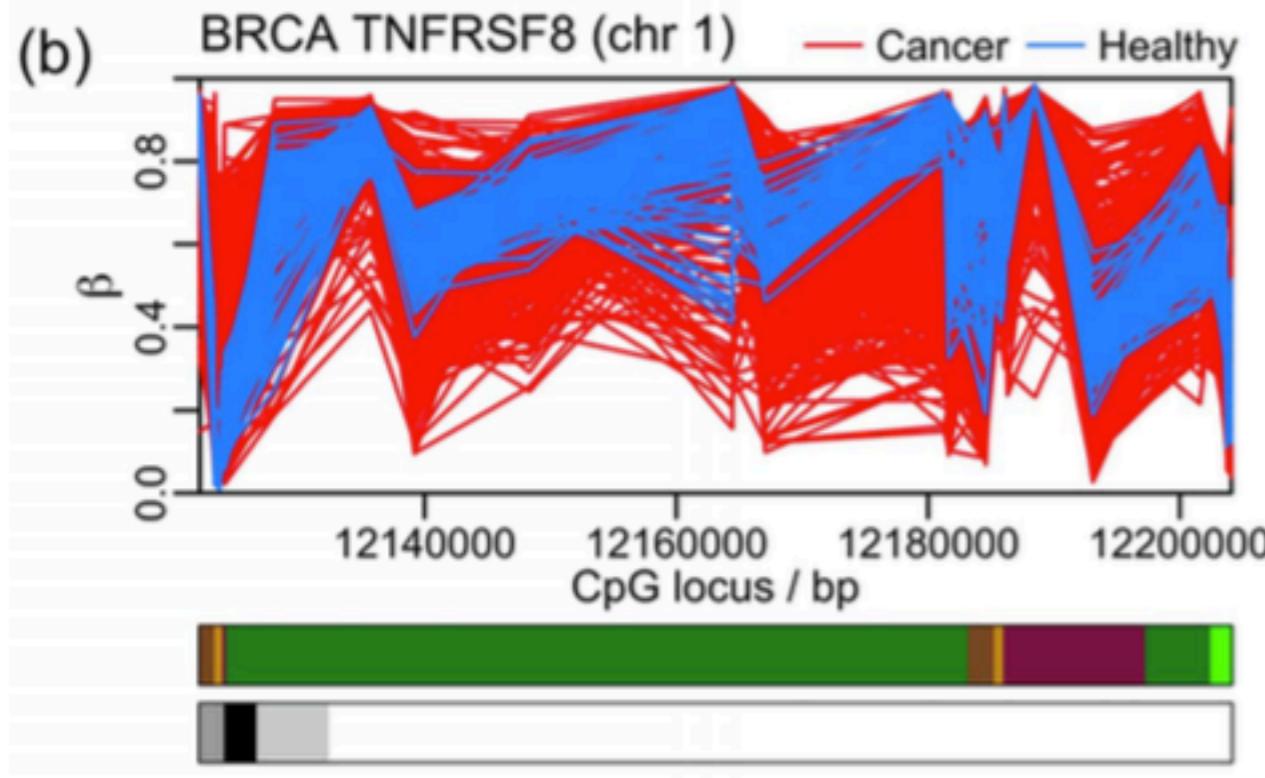
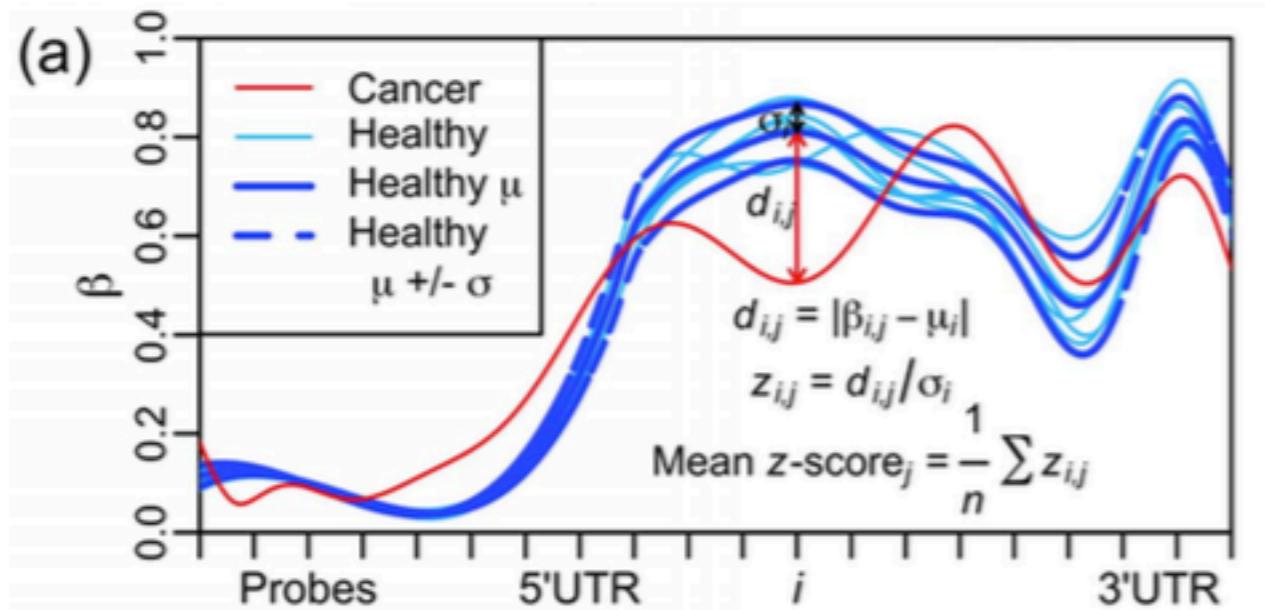
DNA Methylation Analysis

OPEN ACCESS Freely available online

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

Thomas E. Bartlett^{1,2}, Alexey Zaikin², Sofia C. Olhede^{1,3}, James West^{1,4}, Andrew E. Teschendorff⁴, Martin Widschwendter^{5*}

July 2013 | Volume 8 | Issue 7 | e68285



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

Thomas E. Bartlett^{1,3*}, Sofia C. Olhede^{2,3}, Alexey Zaikin¹

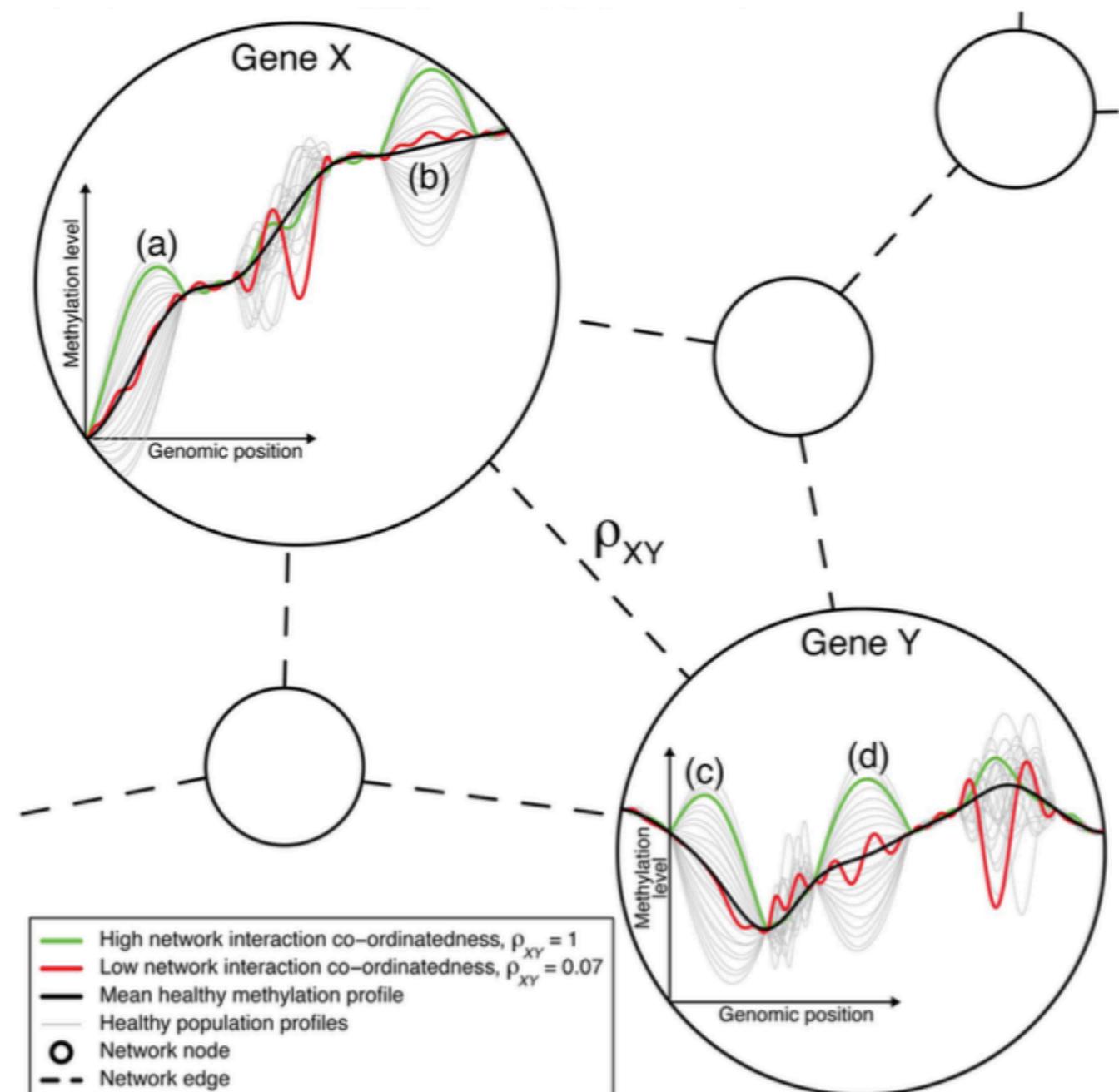


Figure 1. The DNA methylation network interaction measure. A combination of the variation of the healthy methylation profiles in regions (a) and (b) of gene X explains well/is well-explained by a combination of the variation of the healthy methylation profiles in regions (c) and (d) of gene Y. The green 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.

DNA Methylation Analysis

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

Thomas E. Bartlett^{1,3*}, Sofia C. Olhede^{2,3}, Alexey Zaikin¹

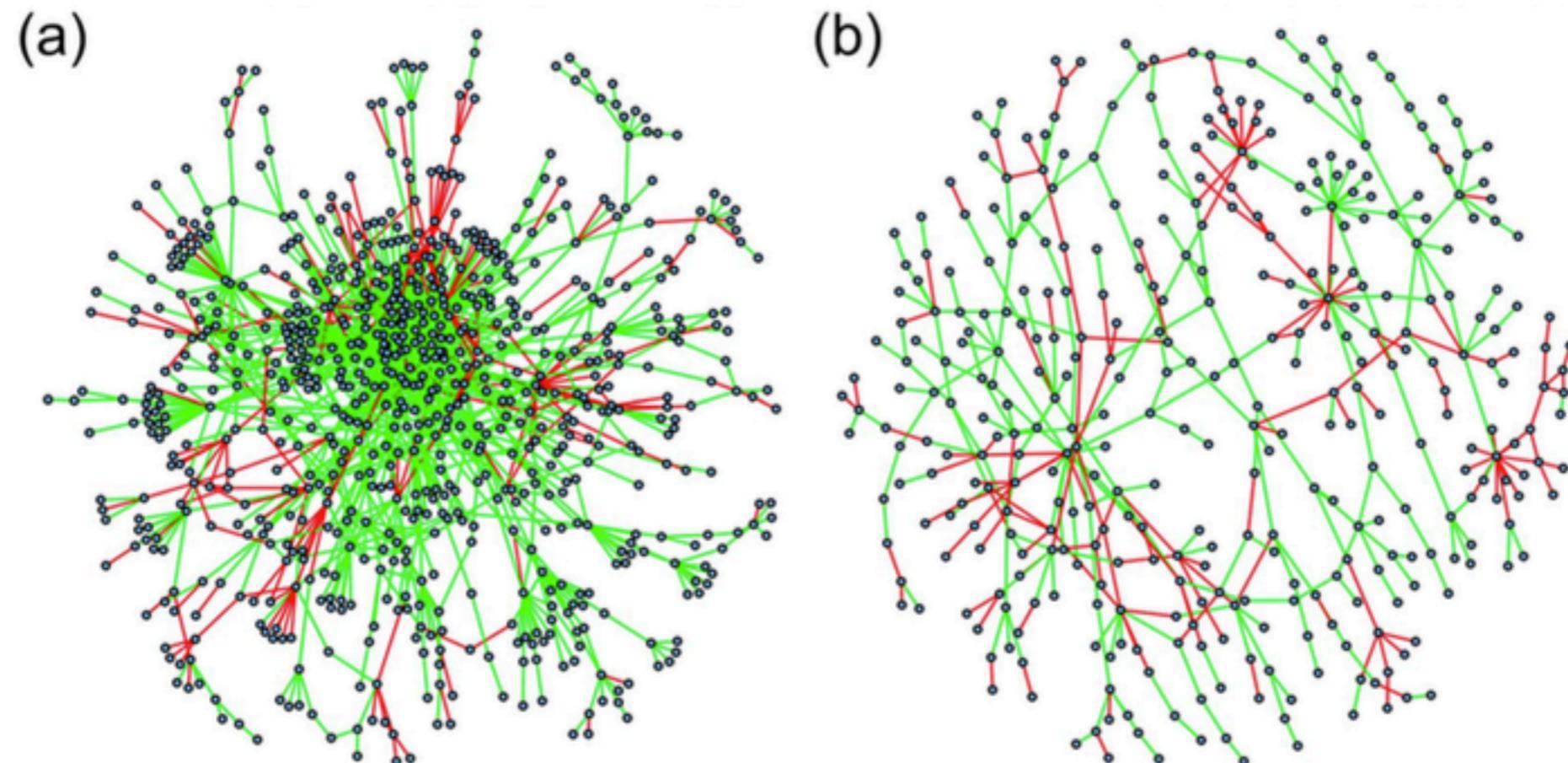


Figure 4. Larger significant subnetworks: network diagrams. Network edges displayed in green and red indicate positive and negative hazard ratios, respectively, for the DNAm network correlation measure corresponding to that interaction; these correspond, respectively, to an increase and decrease in 'network 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?

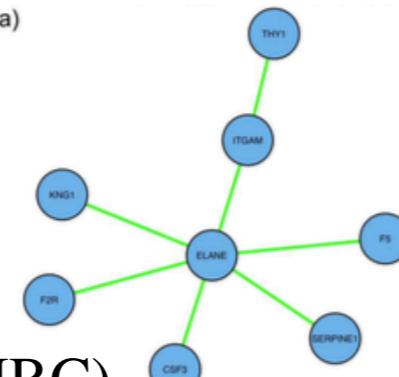
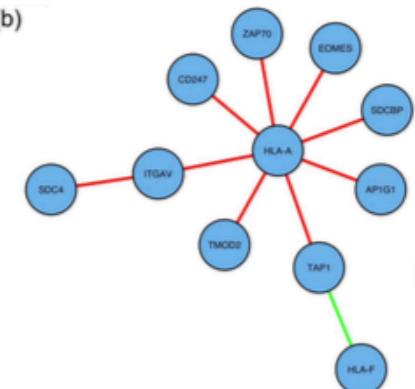
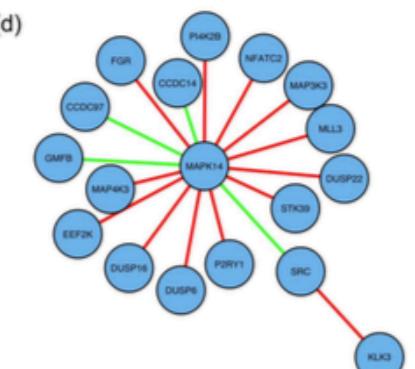
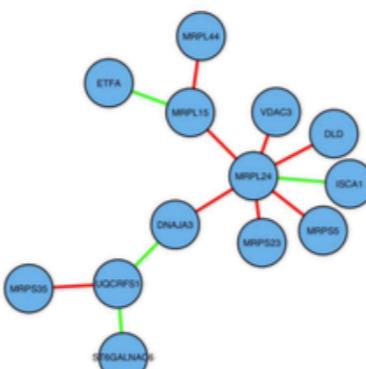
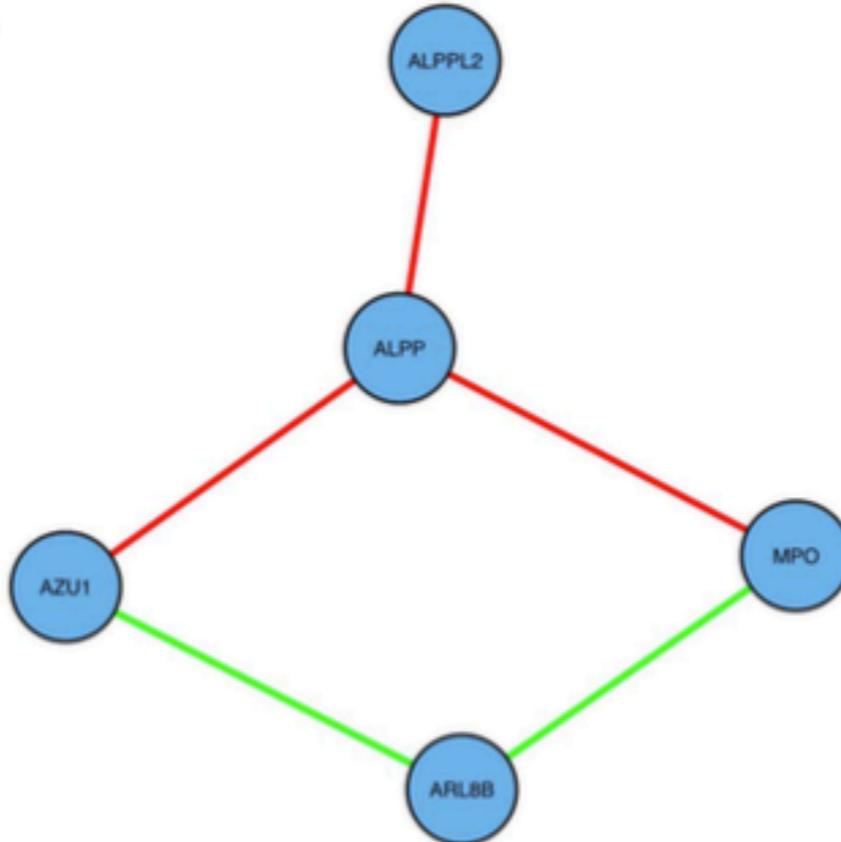
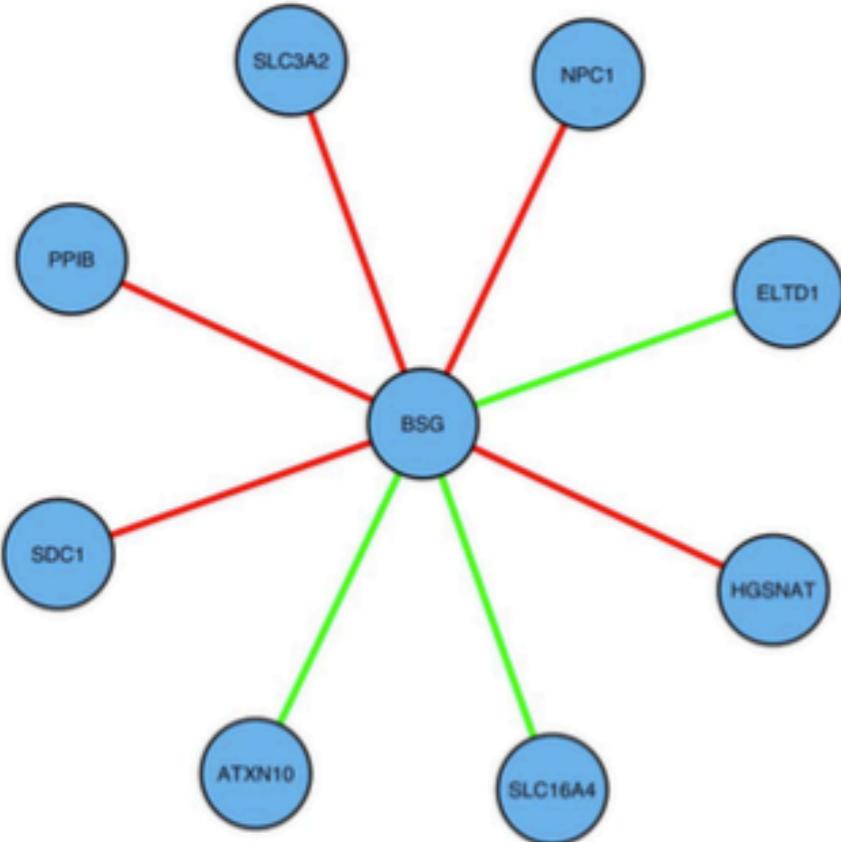
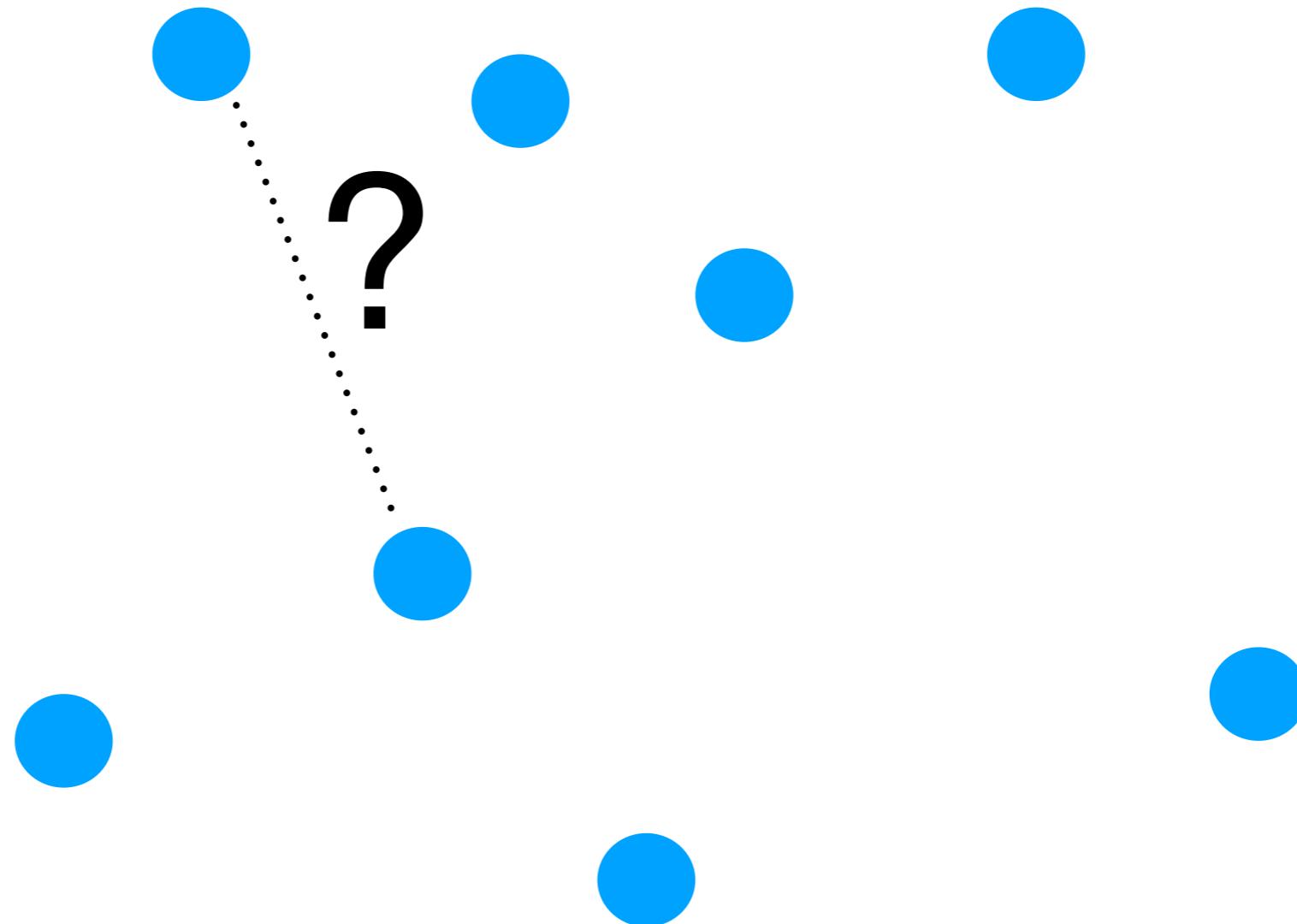
Wound healing module (KIRC).Immune moduleMAP-kinase module (LUSC).

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

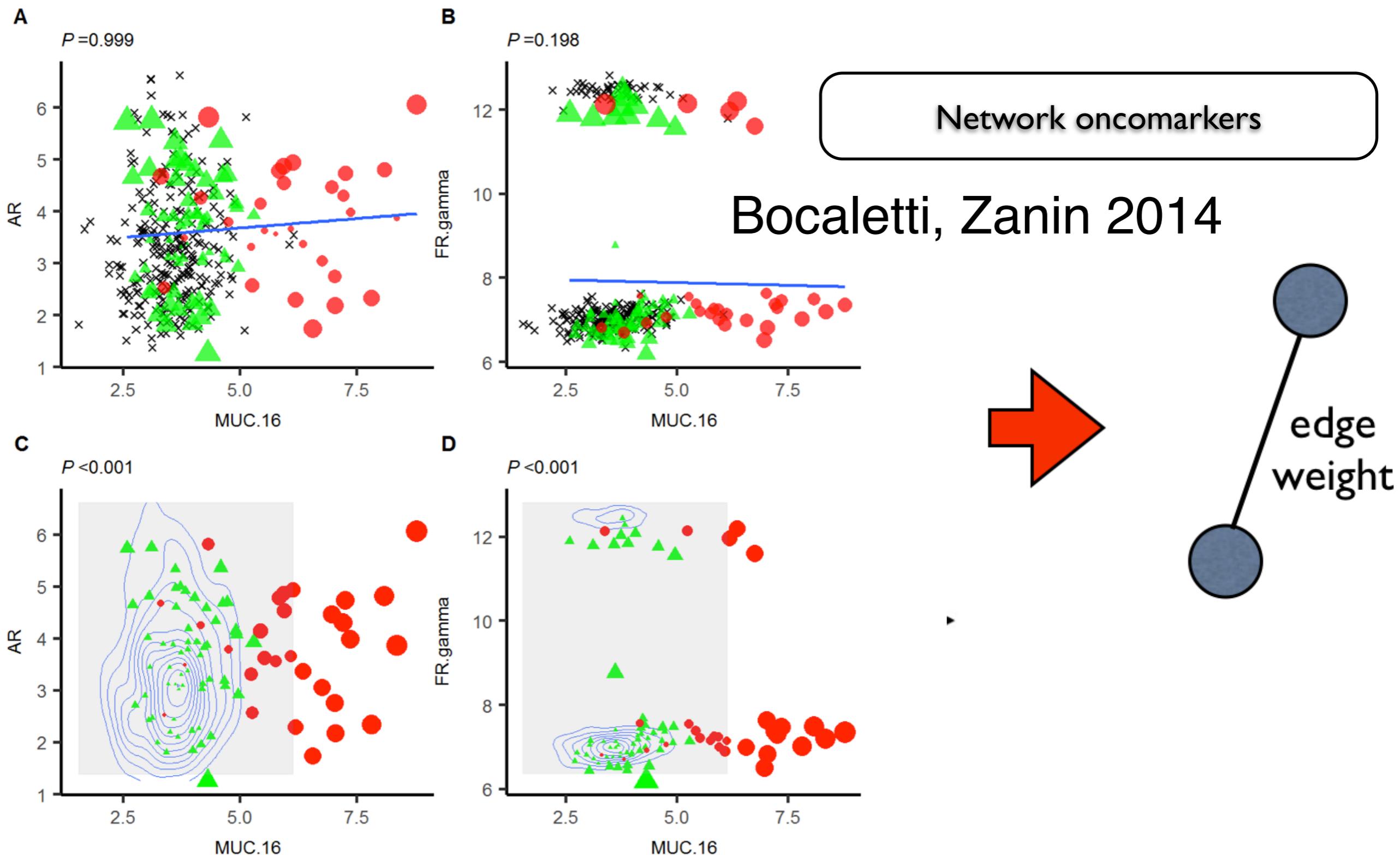
How to construct a network if links are unknown??

Boccaletti, Zanin 2014



Parenclitic networks for predicting ovarian cancer

Harry J. Whitwell¹, Oleg Blyuss², Usha Menon³, John F. Timms³ and Alexey Zaikin^{3,4}



2017

RESEARCH ARTICLE

Parenclitic Network Analysis of Methylation Data for Cancer Identification

Alexander Karsakov¹, Thomas Bartlett², Artem Ryblov¹, Iosif Meyerov³,
Mikhail Ivanchenko¹, Alexey Zaikin^{1,2*}

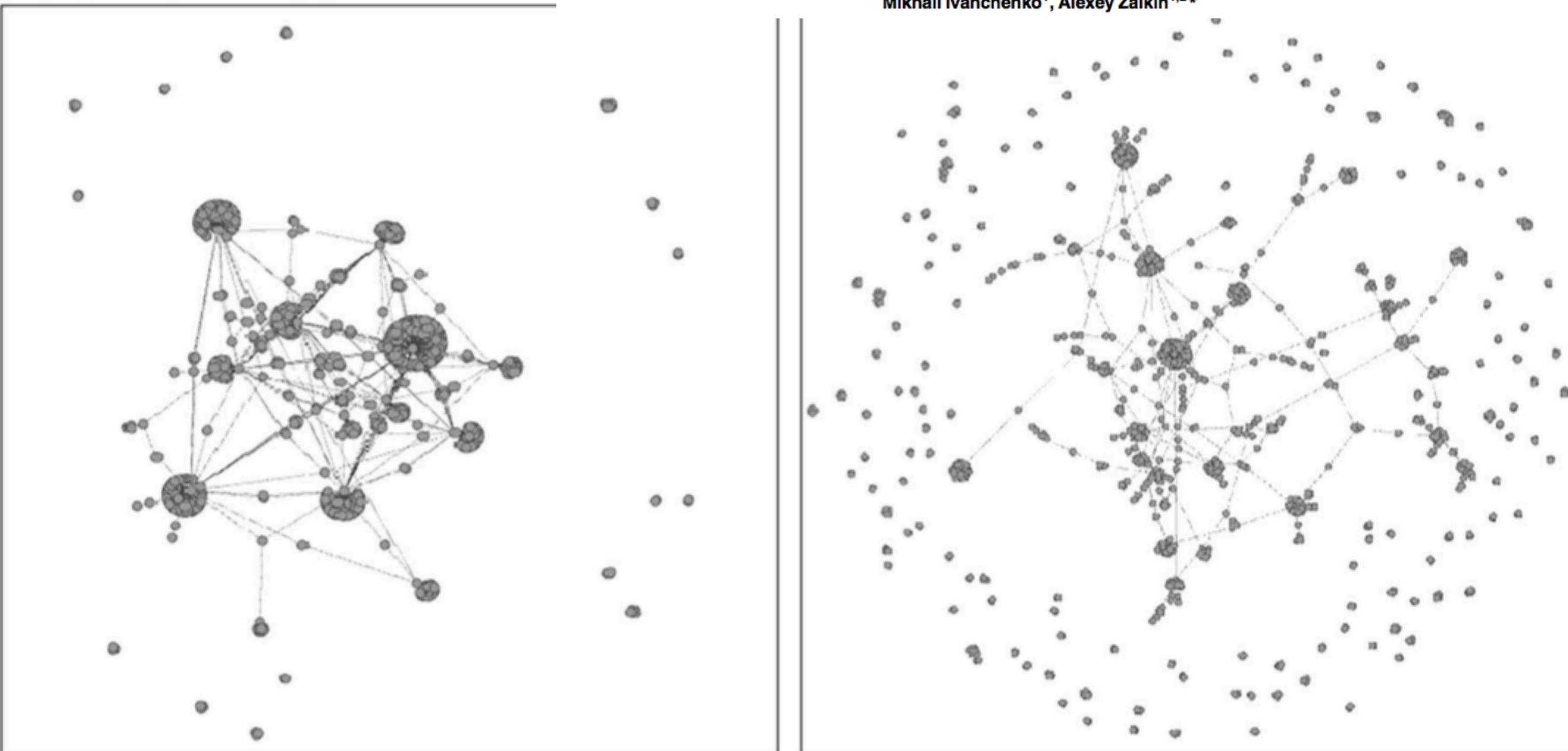
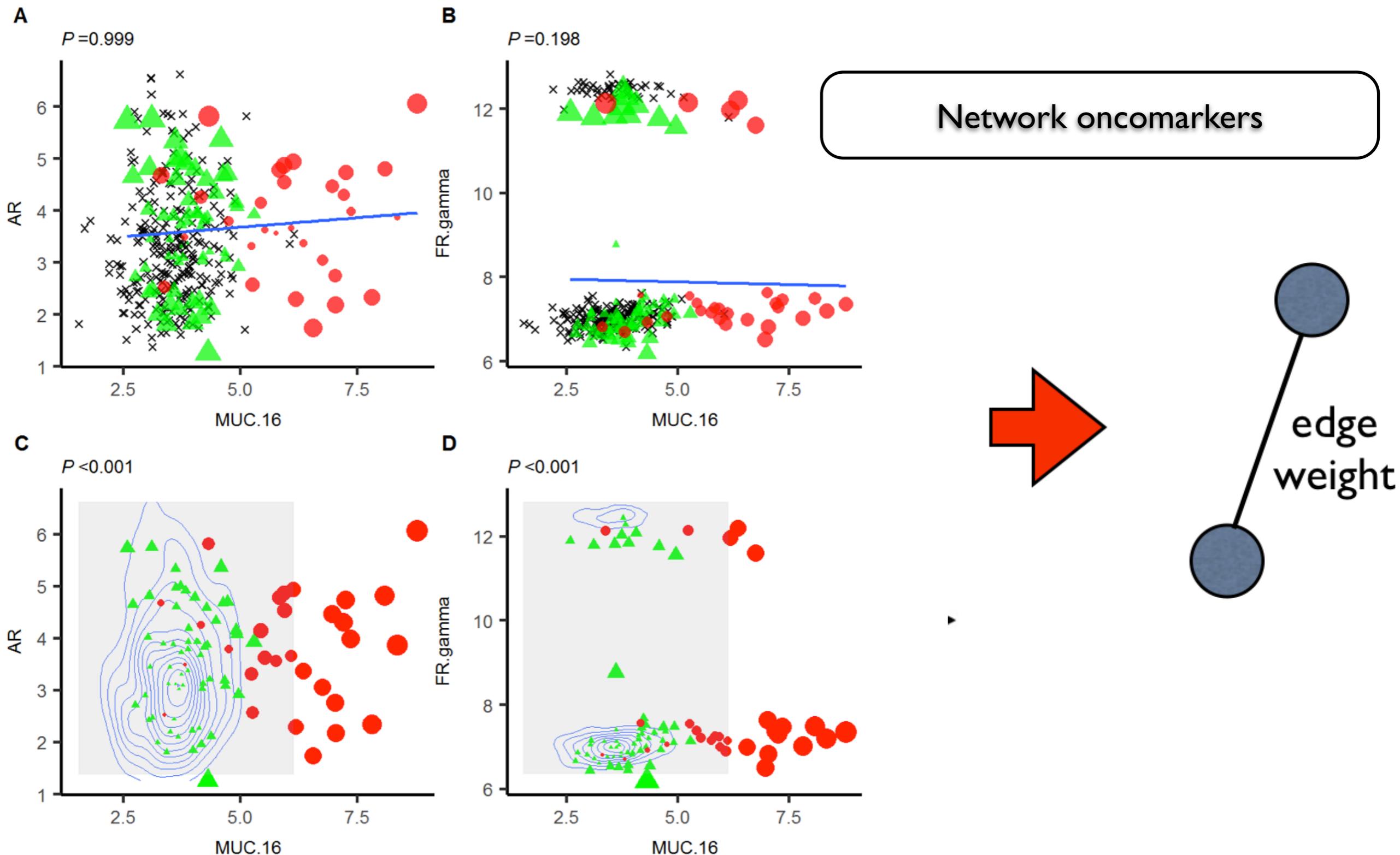


Fig 3. Typical examples of parenclitic networks constructed from gene methylation profiles for cancer (left) and normal (right) samples from BRCA data. Only a 1000 of the strongest edges and their incident nodes are shown. Note the pronounced modular structure for the cancer network.

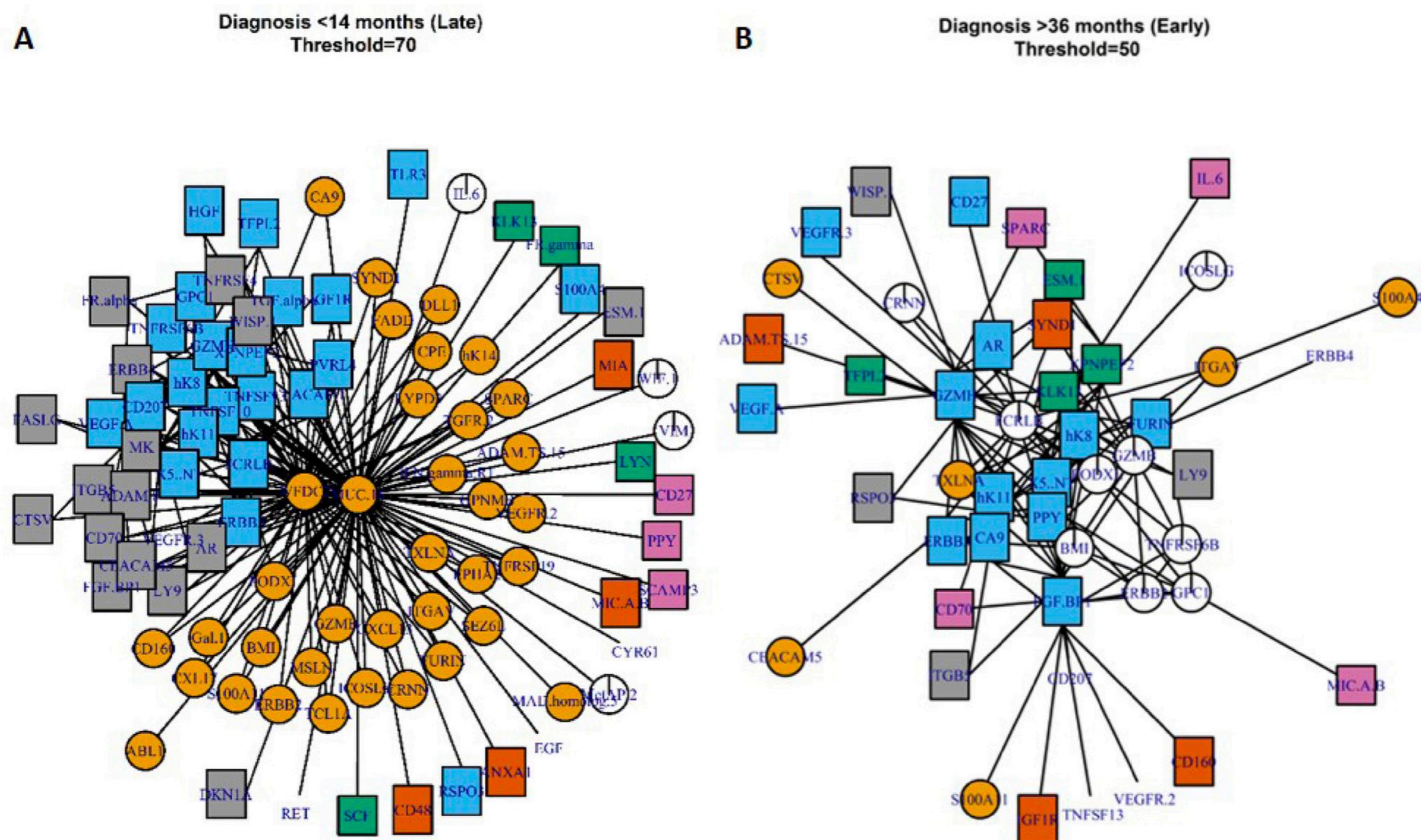
Parenclitic networks for predicting ovarian cancer

Harry J. Whitwell¹, Oleg Blyuss², Usha Menon³, John F. Timms³ and Alexey Zaikin^{3,4}



Parenclitic networks for predicting ovarian cancer

Harry J. Whitwell¹, Oleg Blyuss², Usha Menon³, John F. Timms³ and Alexey Zaikin^{3,4}



Synolic networks
as applied to COVID-19
data

Support us this December

Power vital, open, independent journalism

[Contribute →](#)

[Subscribe →](#)

Search jobs

Sign in

Search

UK edition ▾

The Guardian

News

Opinion

Sport

Culture

Lifestyle

More ▾

World ▶ Europe US Americas Asia Australia Middle East Africa Inequality Global development

Coronavirus

Covid blood test can predict patient survival chances

Protein analysis provides digital picture of immune response and mortality risk, say scientists

- [Coronavirus - latest updates](#)
- [See all our coronavirus coverage](#)

Linda Geddes

Mon 7 Dec 2020 11.26
GMT



707



▲ It is unlikely that a blood test alone would ever be used to dictate which patients are allocated ICU beds.
Photograph: Murtaja Lateef/EPA

A blood test has been developed that can predict whether Covid patients will need intensive care - or are even likely to survive - shortly after they develop symptoms.

The patients admitted at the Charité University Hospital from early March to the end of June

7 Dec 2020

“We can predict which patients will need oxygen support and ventilator support quite accurately, and we also have markers for patients who are not that severely ill initially, but are at high risk of getting worse,” said Ralser, whose research is published as a [preprint](#) but has not yet been peer reviewed.

Cell Systems 2021
PLOS Digital Health 22

A huge-collaborative effort: 61 authors, 28 affiliations

Parenclitic Networks Predict Survival for Severely Ill Covid-19 Patients (Grade WHO = 7) Weeks Before Outcome with Extremely High Predictive Power

a time-resolved deep clinical and molecular phenotyping of 139 adult patients with COVID-19 during hospitalization

Tatiana
Nazarenko

Harry J.
Whitwell

Oleg Blyuss

John F. Timms

Alexey Zaikin



Cell Systems

Cell Systems 12, 1–15, July 21, 2021

Article

A time-resolved proteomic and prognostic map of COVID-19

 CellPress
OPEN ACCESS

+PLOS Digital Health
2022

Vadim Demichev,^{1,2,3,27} Pinkus Tober-Lau,^{4,27} Oliver Lemke,¹ Tatiana Nazarenko,^{8,11} Charlotte Thibeault,⁴ Harry Whitwell,^{9,10,26} Annika Röhl,¹ Anja Freiwald,¹ Lukasz Szrywiel,² Daniela Ludwig,¹ Clara Correia-Melo,²

Biomarkers that classify COVID-19 severity

based on clinical chemistry, enzyme activity, immunoprofile, single cell sequencing, proteomics, and metabolomics.

High-throughput proteomic analysis - 180 samples/day.

309 proteins quantified in undepleted PLASMA using Scanning SWATH with short gradients.

D'Alessandro, A., Thomas, T., Dzieciatkowska, M., Hill, R.C., Francis, R.O., Hudson, K.E., Zimring, J.C., Hod, E.A., Spitalnik, S.L., and Hansen, K.C. (2020). *Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level*. *J. Proteome Res.*

Laing, A.G., Lorenc, A., Del Molino Del Barrio, I., Das, A., Fish, M., Monin, L., Muñoz-Ruiz, M., McKenzie, D.R., Hayday, T.S., Francos-Quijorna, I., et al. (2020). *A dynamic COVID-19 immune signature includes associations with poor prognosis*. *Nat. Med.* 26, 1623-1635.

Liu, Y., Gao, W., Guo, W., Guo, Y., Shi, M., Dong, G., Ge, Q., Zhu, J., and Lu, J. (2020b). *Prominent coagulation disorder is closely related to inflammatory response and could be as a prognostic indicator for ICU patients with COVID-19*. *J. Thromb. Thrombolysis*.

Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., et al. (2020). *Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection*. *Cell Syst* 11, 11-24.e4.

Overmyer, K.A., Shishkova, E., Miller, I.J., Balnis, J., Bernstein, M.N., Peters-Clarke, T.M., Meyer, J.G., Quan, Q., Muehlbauer, L.K., Trujillo, E.A., et al. (2020). *Large-scale Multi-omic Analysis of COVID-19 Severity*. *medRxiv*.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). *Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment*. *Cell* 182, 1419-1440.e23.

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al. (2020). *Proteomic and Metabolomic Characterization of COVID-19 Patient Sera*. *Cell* 182, 59-72.e15.

Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., Zou, X., Yang, Q., Yuan, Y., Bie, Y., et al. (2020). *Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19*. *Immunity*.

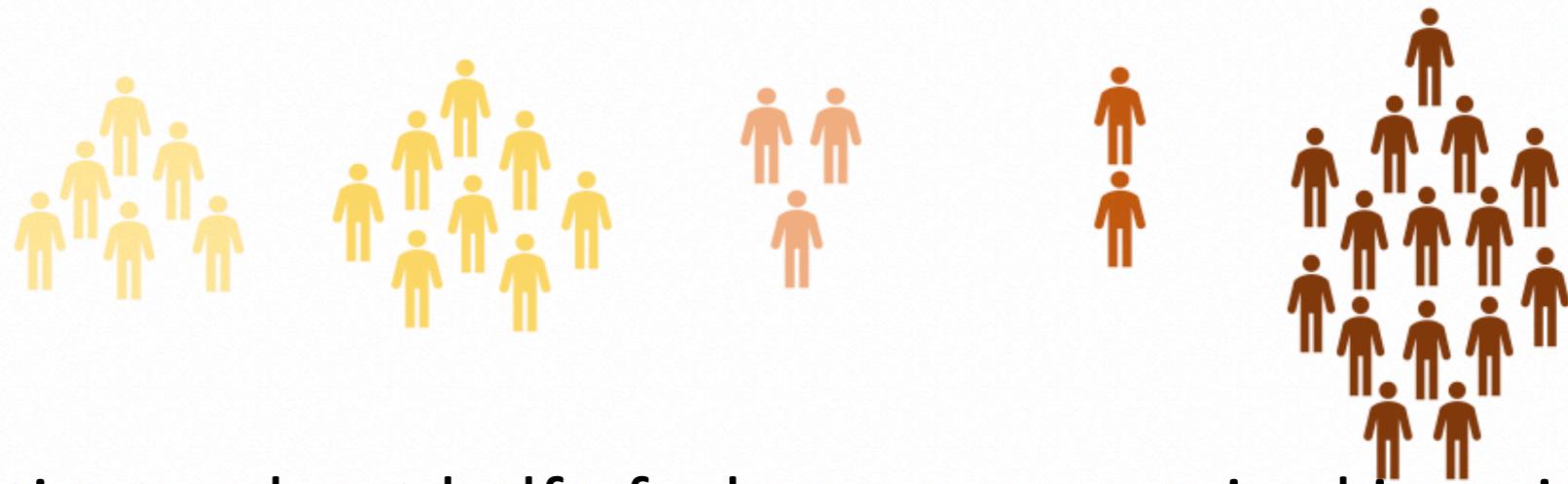
Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., Damen, J.A.A., Debray, T.P.A., De Vos, M., et al. (2020). *Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal*. *BMJ* 369, m1328.

| Patient State | Score | Descriptor |
|-------------------------------|-------|--|
| Uninfected | 0 | No clinical or virological evidence of infection |
| Ambulatory | 1 | No limitation of activities |
| | 2 | Limitation of activities |
| Hospitalised - mild disease | 3 | No oxygen therapy |
| | 4 | Oxygen by mask or nasal prongs |
| Hospitalised - severe disease | 5 | Non-invasive ventilation or high-flow oxygen |
| | 6 | Intubation and mechanical ventilation |
| | 7 | Ventilation + additional organ support (vasopressors, RRT, ECMO) |

WHO ordinal scale for clinical improvement in COVID-19 as used in the study (World Health Organisation 2020)

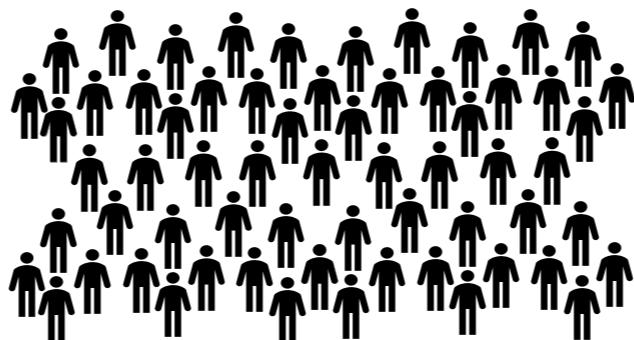
The patients admitted at the Charité University Hospital from early March to the end of June

| All patients | No invasive mechanical ventilation | | | Invasive mechanical ventilation | | |
|--------------|------------------------------------|-----------|-----------|---------------------------------|-----------|--|
| | Max WHO 3 | Max WHO 4 | Max WHO 5 | Max WHO 6 | Max WHO 7 | |
| 139 | 23 | 32 | 15 | 6 | 63 | |
| 100% | 17% | 23% | 11% | 4% | 45% | |



There were 139 patients, about half of whom were required invasive mechanical ventilation and half of whom are not. Among these people, 20 patients died, most of them (namely 17 people) are patients with a grade 7.

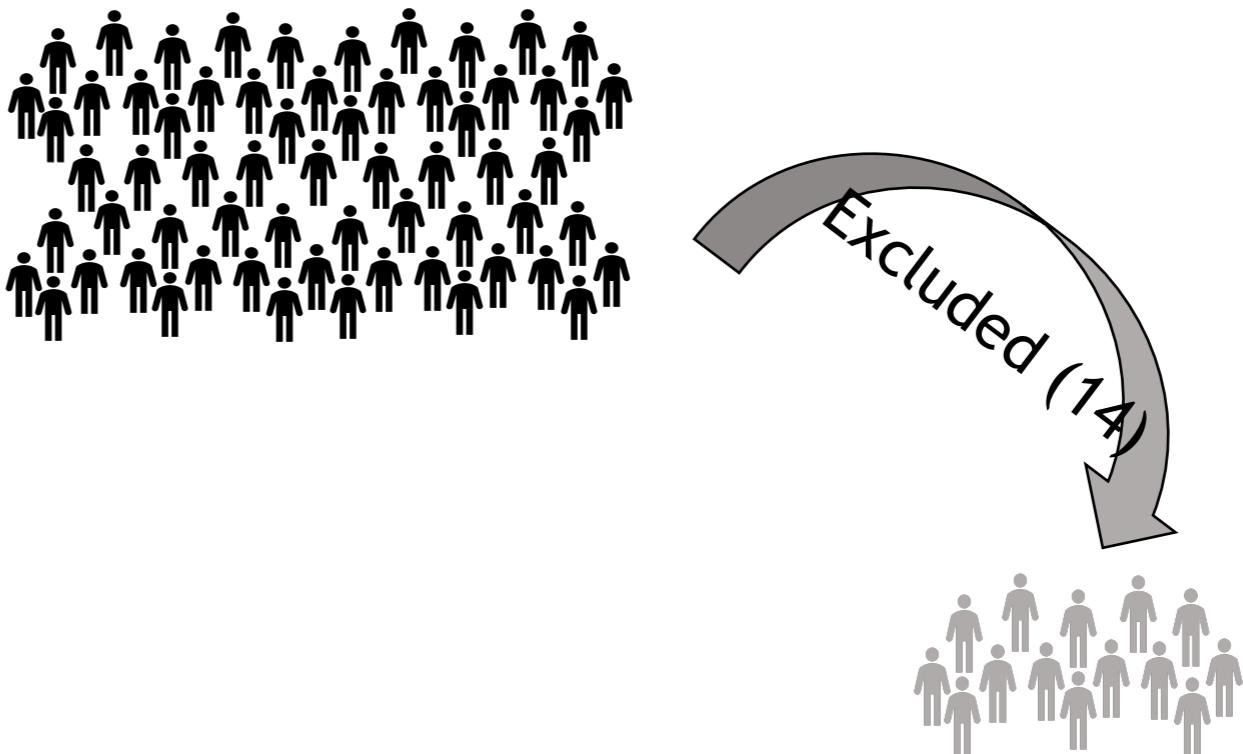
Charite cohort



139 patients from Charite hospital admitted with PCR confirmed CoVID. Serum and clinical diagnostics taken at multiple time points across course of stay – 687 plasma samples. **309 proteins quantified in undepleted PLASMA using Scanning SWATH** with short gradients. Since mortality was predominantly associated with patients with a grade 7 (that is, the most severe patients), we selected only them. There were 63 people.

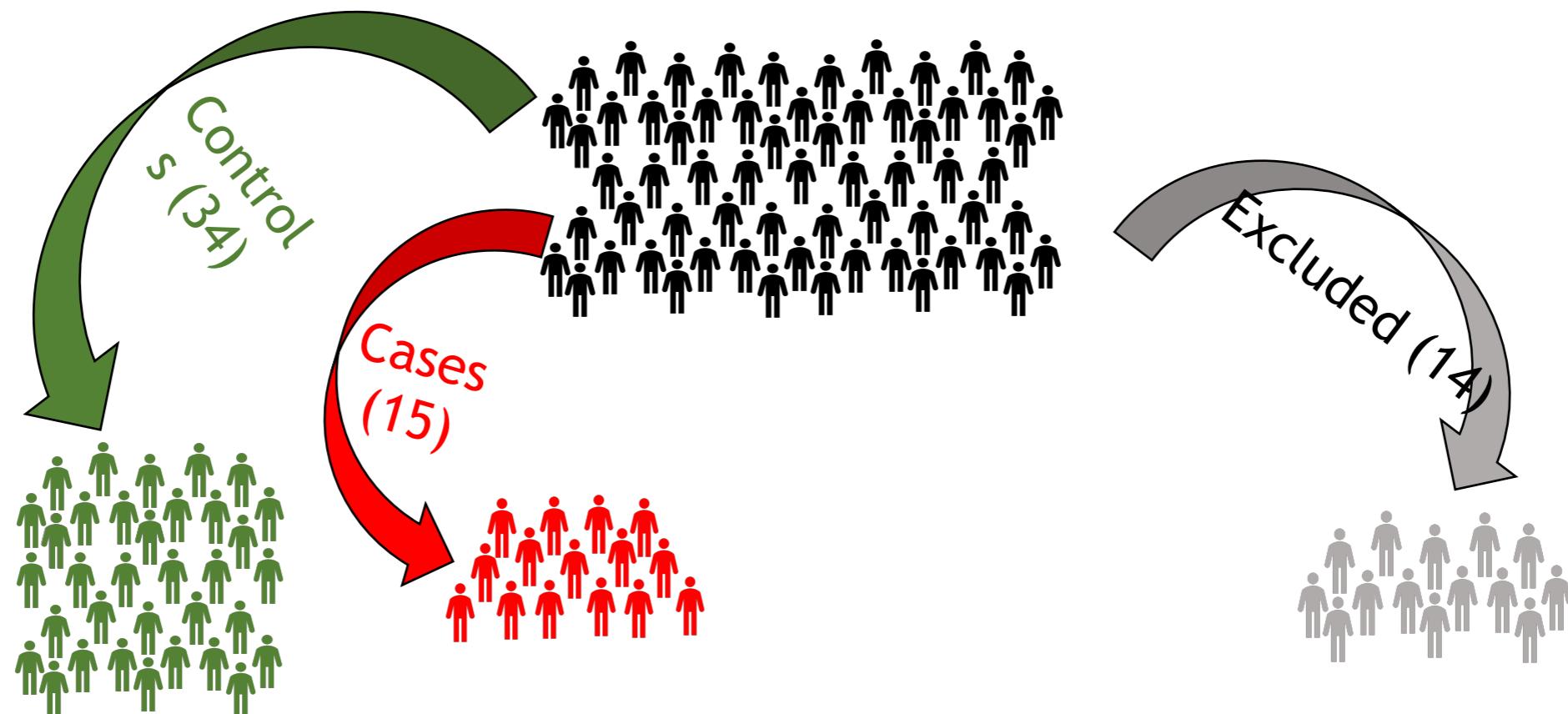
|
WHO grade 7 patients (63)

outcome (i.e. discharge or death) for them was not established



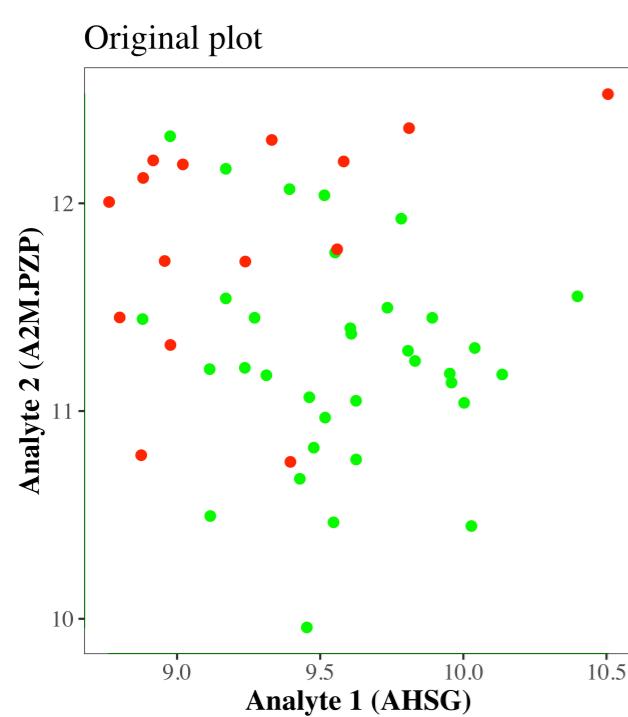
WHO grade 7 patients (63)

The remaining 49 patients were divided into a control group (or discharged patients) and a case group (or deceased patients).



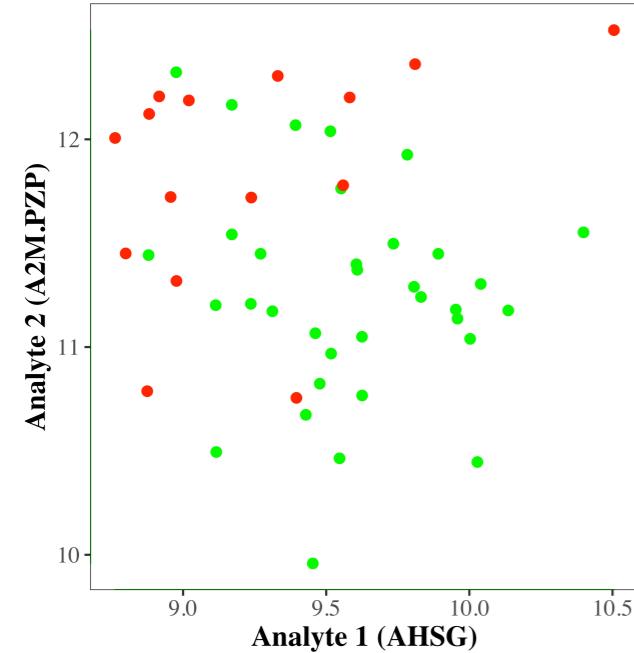
WHO grade 7 patients (63)

It should be especially noted that other machine learning algorithms did not give good quality on this dataset,
and only using of parenclitic networks approach allow us to obtain such a good results.

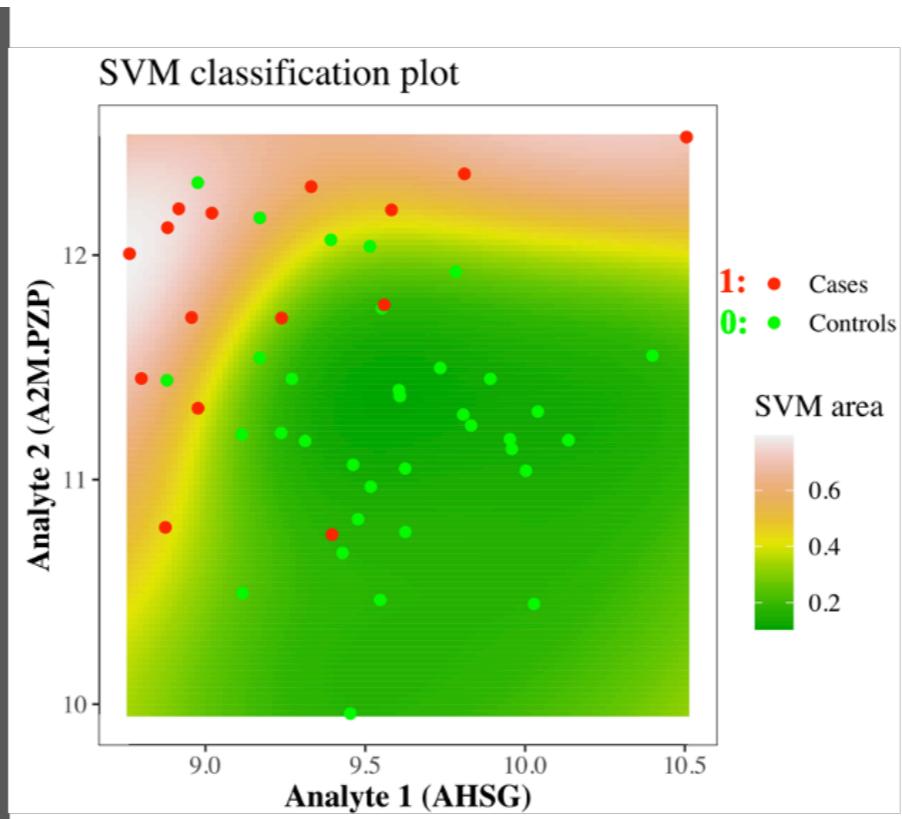


Synolitic Network Construction (SVM approach)

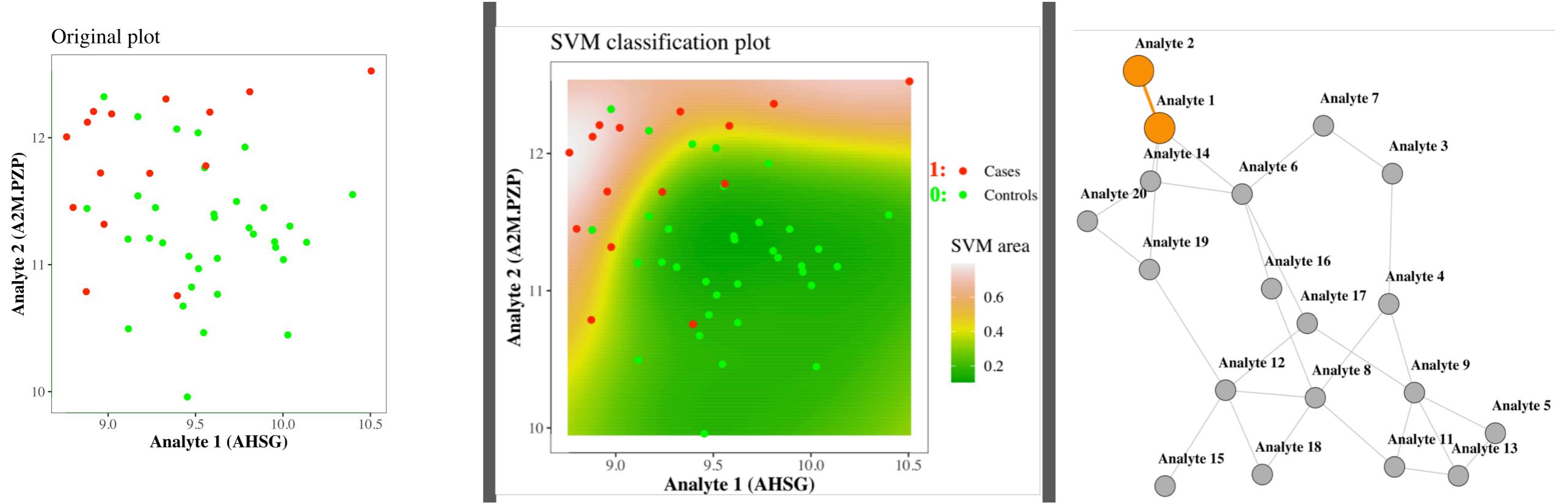
Original plot



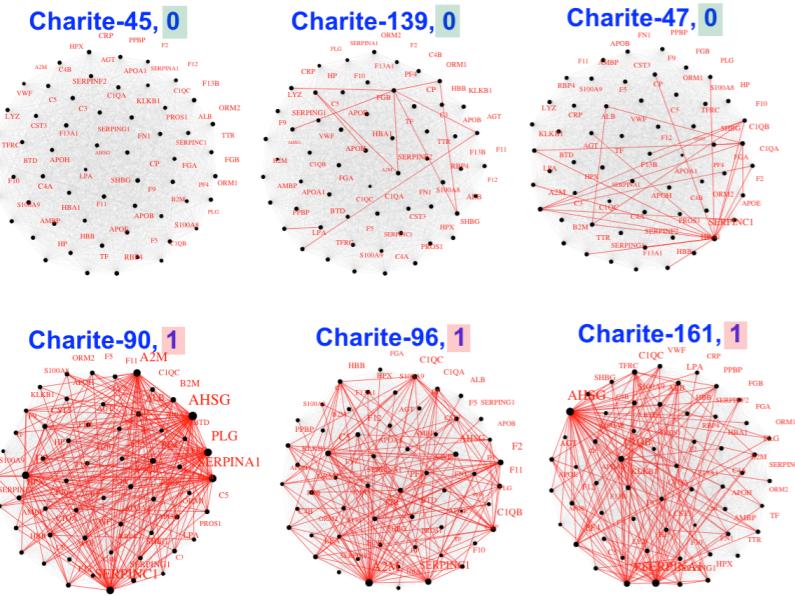
SVM classification plot



Synolitic Network Construction (SVM approach)



Synolitic Network Construction (SVM approach)



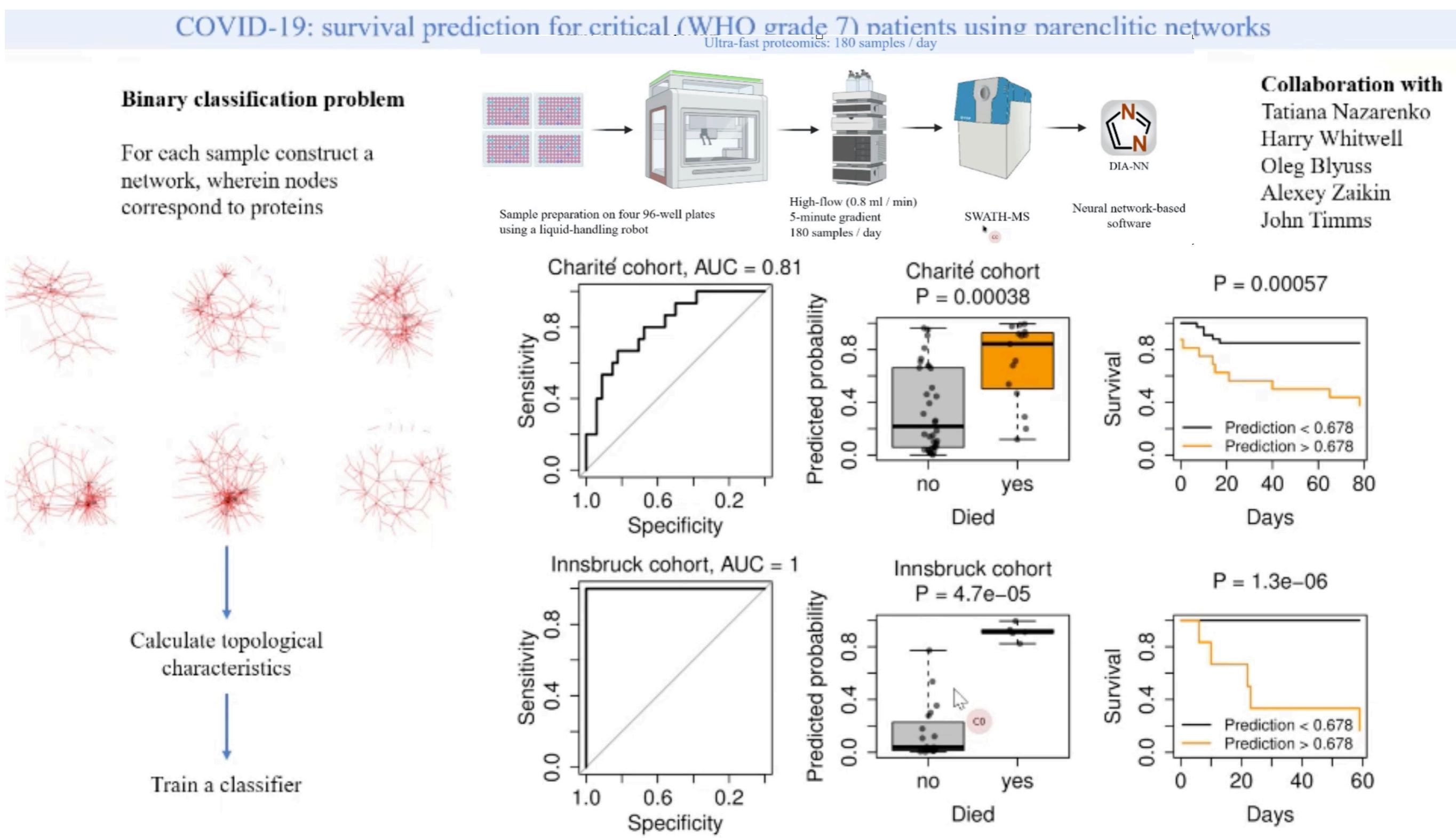
Calculate
topological
characteristics



Train a
classifier

Binary classification problem
on networks characteristics

Following cross-validation, the model showed **excellent accuracy (AUC = 0.81)** despite the median time from sampling to outcome being 39 days. You also can see interquartile range: from 16 to 64 days. Especially amazing results (**with AUC=1**) we had on validation our model on completely independent dataset - 24 Covid-19 patients admitted to the Innsbruck Hospital in Austria.



Age-related trajectories

Age-related trajectories of DNA methylation network markers: a parenclitic network approach to a family-based cohort of patients with Down Syndrome

M. Krivonosov¹, T. Nazarenko^{2,*}, M.G. Bacalini³, M.V. Vedunova¹, C. Franceschi^{1,3}, A. Zaikin^{1,2,4}, and M. Ivanchenko¹

Down Syndrome methylation data

As an application and demonstration of our implementation, we considered a publicly available dataset (GSE52588) in which whole blood DNA methylation was assessed by the Infinium HumanMethylation450 BeadChip in a cohort including persons affected by Down Syndrome (DS), their unaffected siblings (DSS) and their mothers (DSM)²⁰ (29 families in total).

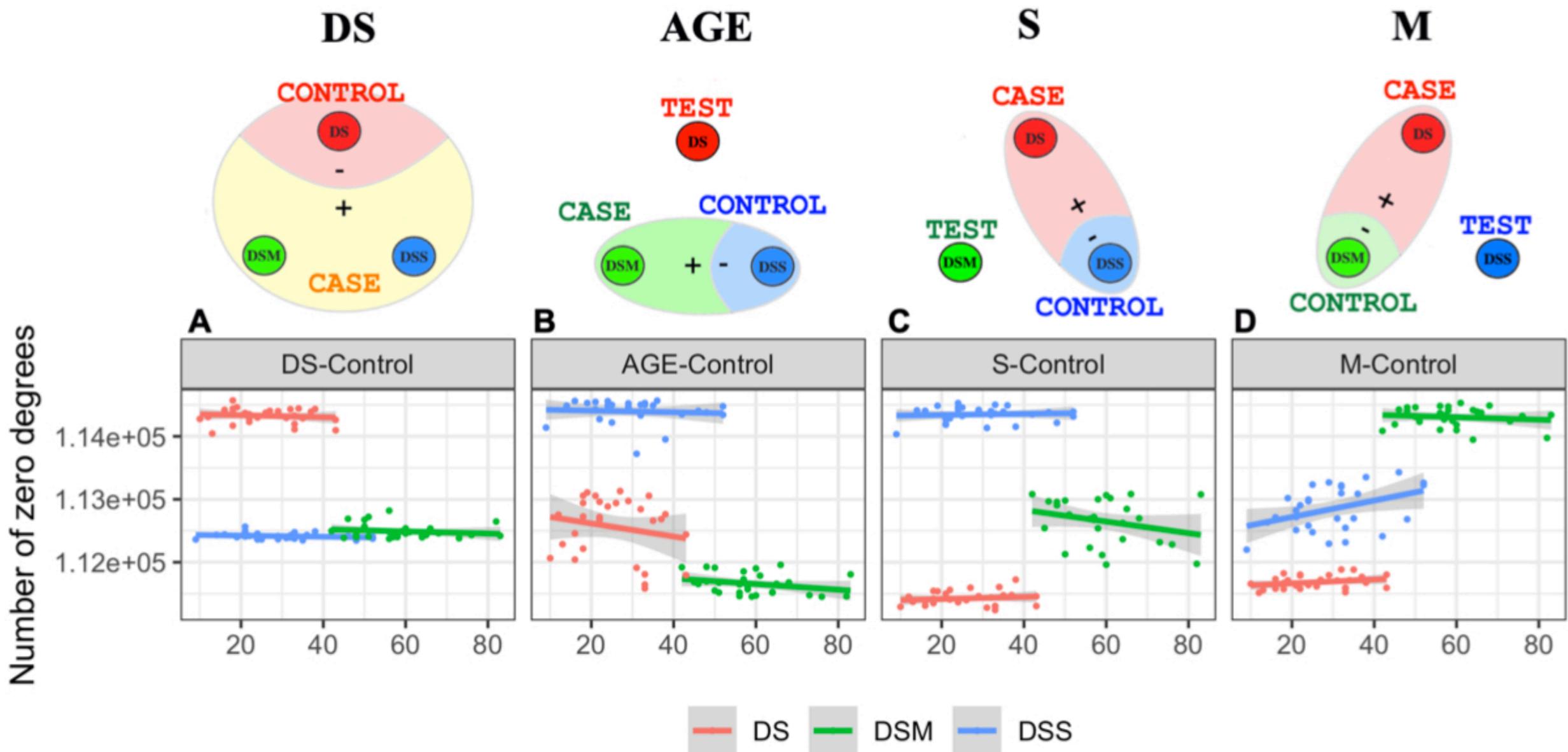


Figure 3. Top panel: Two sets of the data are used to find a boundary necessary to construct a network. This can be used to analyse the age-related trajectories for the third class (B-D). Bottom panel: examples how topological indices plotted vs age enable to find age related trajectories of network signatures. Dependence of characteristics (here, for example, only Number of zero degrees nodes) of individual networks in DS/AGE/S/M-network design versus AGE (A–D labels respectively). A) Down methylation network signature. B) Age-accelerated ageing in Downs. C) Hyper-ageing in DS. D) Divergence of trajectories in Downs and healthy sibs. Plots for all other characteristics can be found in Supporting Information.

But how age accelerated ageing in Downs will depend on a biological age estimated with well-established Horvath's clock? This is shown in the Fig. 4. We found a very surprising behaviour when we plotted topological indices versus residuals, i.e., a difference between a passport age and biological age. DS are closer to mothers, and develop with age towards mothers, i.e., their network features are more similar to mothers, the more is age acceleration. However, we find that even for DS with decelerated ageing, their methylation network signature has a trend towards older mothers. Probably, this can be explained by the fact that Horvath's methylation clock has been developed for healthy people and it does not work so well for DSS. In contrast to it, our approach always shows age acceleration in patients with DS.

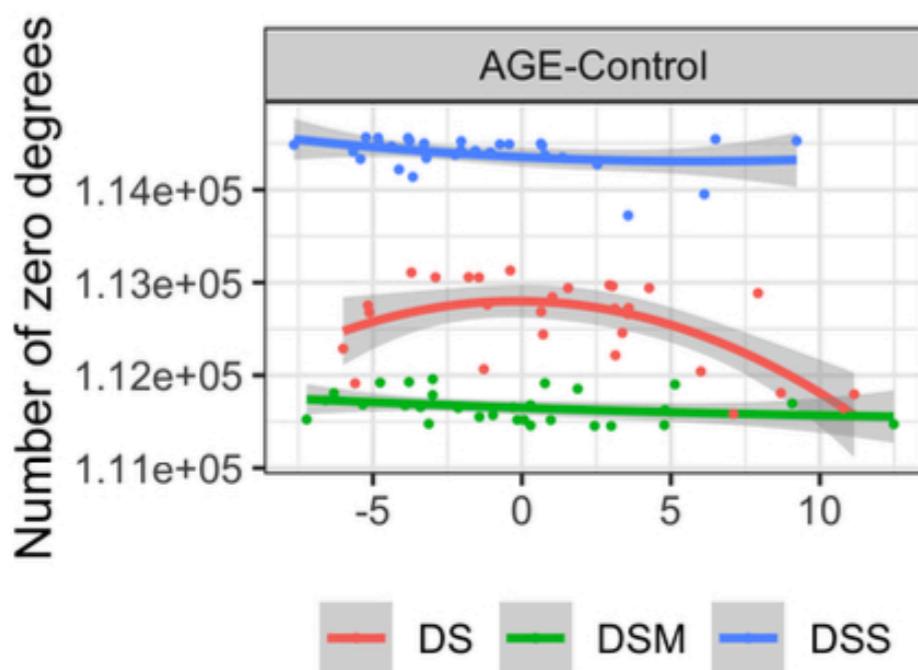


Figure 4. Dependence of characteristics (here, for example, only Number of zero degrees nodes) of individual networks in RESIDUALS, i.e., a difference between passport and biological age. Above zero along x-axis we have accelerated ageing, below - decelerated. Still, in both cases, DS methylation network signature is closer to mothers. Plots for all characteristics can be found in Supporting Information.

How good are Synolitic Network Classification

Impact Factor 4.599 | CiteScore 3.7
More on impact >



Statistical Genetics and Methodology



SECTION ABOUT ARTICLES RESEARCH TOPICS FOR AUTHORS ▾ EDITORIAL BOARD ARTICLE ALERTS

< Articles

THIS ARTICLE IS PART OF THE RESEARCH TOPIC
Prediction and Explanation in Biomedicine using Network-Based Approaches
[View all 10 Articles >](#)

EDITED BY



Alessio Martino
National Research Council (CNR),
Italy

REVIEWED BY



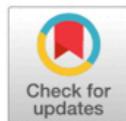
Yuehua Cui
Michigan State University, United
States



Shaoyu Li
University of North Carolina at

ORIGINAL RESEARCH article

Front. Genet., 20 October 2021 | <https://doi.org/10.3389/fgene.2021.733783>



Download Article



357

TOTAL VIEWS

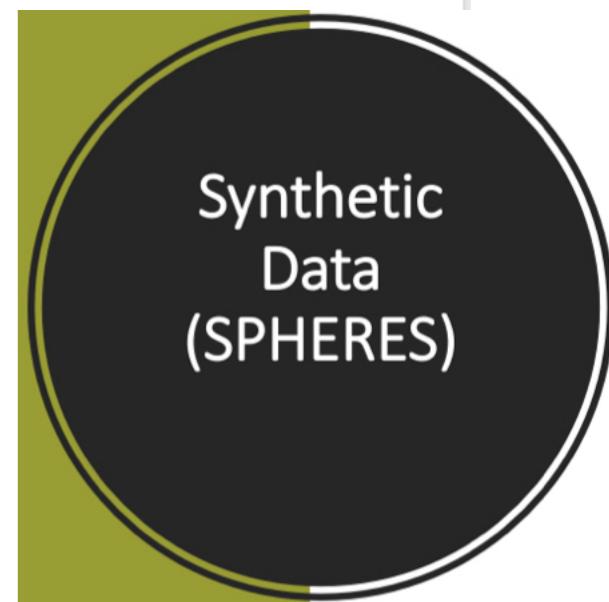
Am score 3

View Article Impact

Parenclitic and Synolytic Networks Revisited

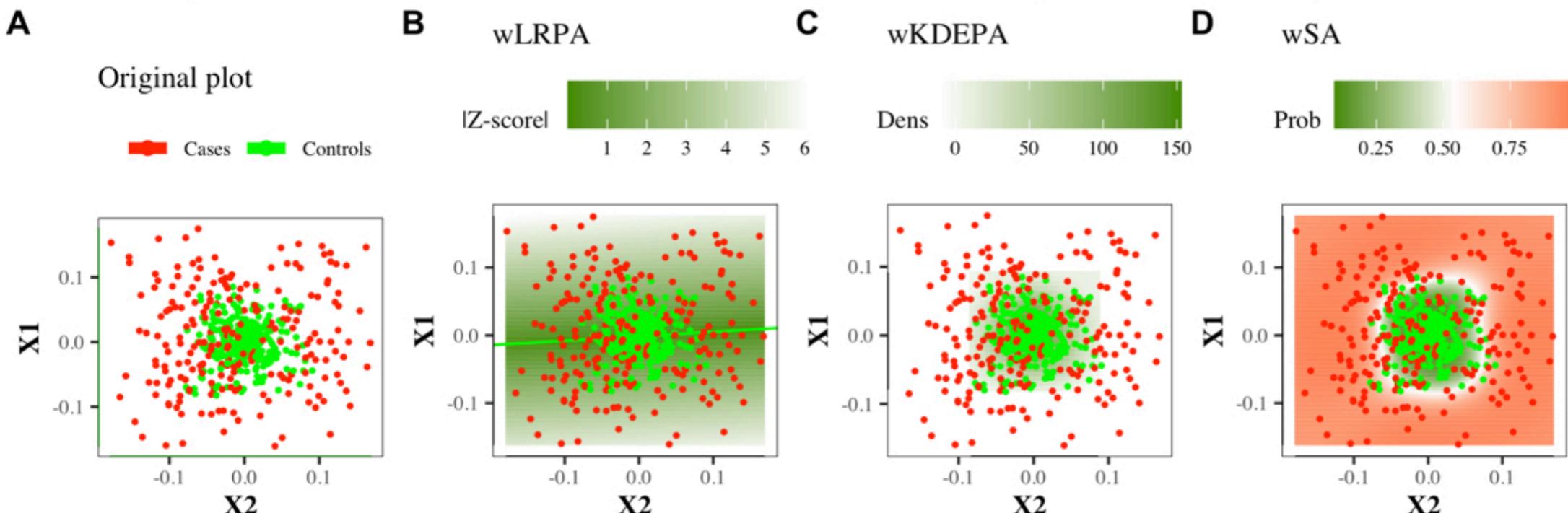
Tatiana Nazarenko^{1*}, Harry J. Whitwell^{2,3,4,5}, Oleg Blyuss^{1,4,6,7} and Alexey Zaikin^{1,4,5}

- Sphere-models: Controls: points with $R < 0.5$, Cases: points with $0.5 < R < 1$)
- Types of Spheres:
 - **Ideal Spheres Model** – all parameters are sphere parameters
 - **Noisy Spheres Model** – 50 noise parameters were added for sphere parameters
 - **Broken Spheres Model** – half sphere parameters were changed by noise parameters



- For all modelling, we considered all possible combinations of
 - *Sphere Dimensions*: (2, 3, 10, 30, 60, 90, 120, 150);
 - number of *Case TRAIN samples*: (15, 65, 115, 165, 215, 265)
 - number of *Controls TRAIN samples*: (15, 65, 115, 165, 215, 265).
- Numbers of *Case TEST samples* and *Controls TEST samples* were calculated as 25% of corresponding *TRAIN* numbers.

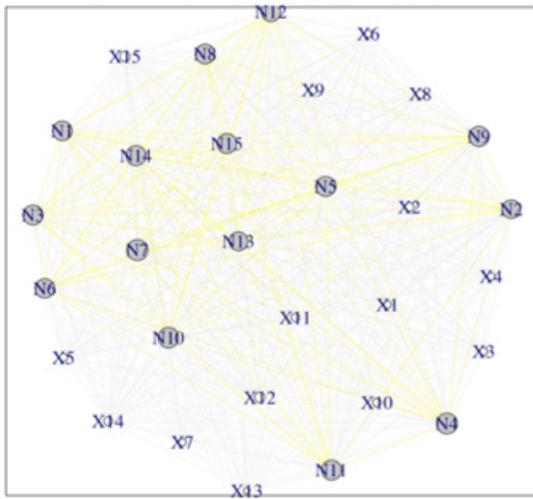
Examples of connections creating by different types of PN approaches



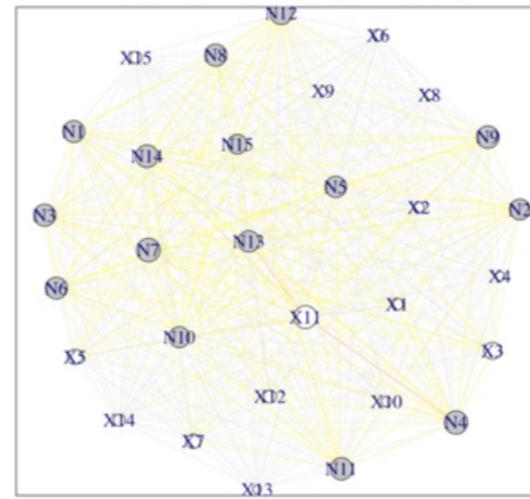
Main advantages:

Visualisation

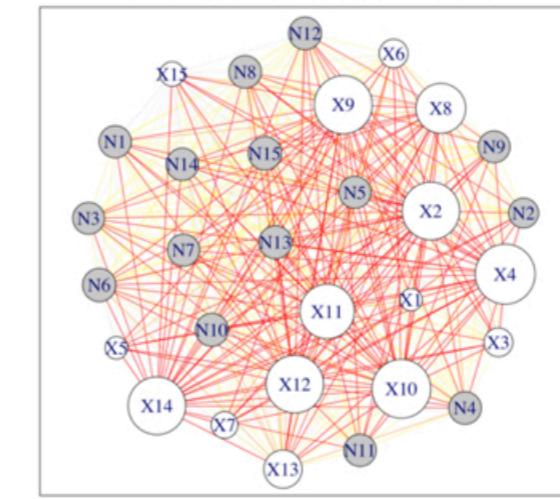
Sample with $R = 0.01$



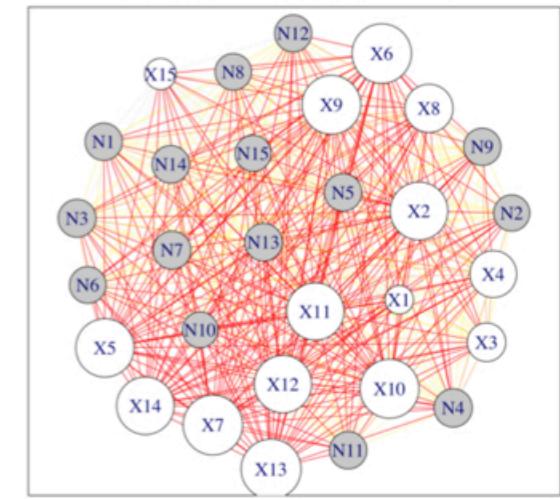
Sample with $R = 0.25$



Sample with $R = 0.75$



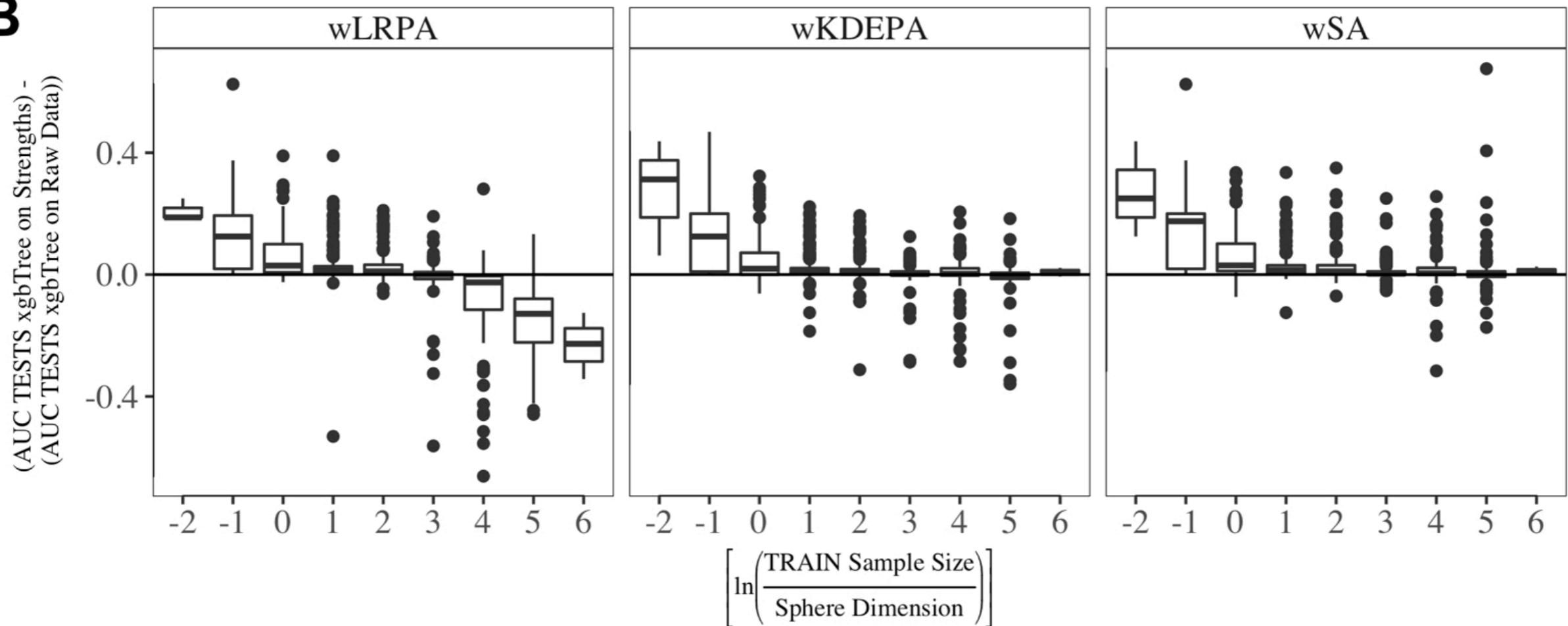
Sample with $R = 0.99$



Main advantages:

Parenclitic approaches on average demonstrate superiority to other ML methods in situations where sample size is small relative to the number of features

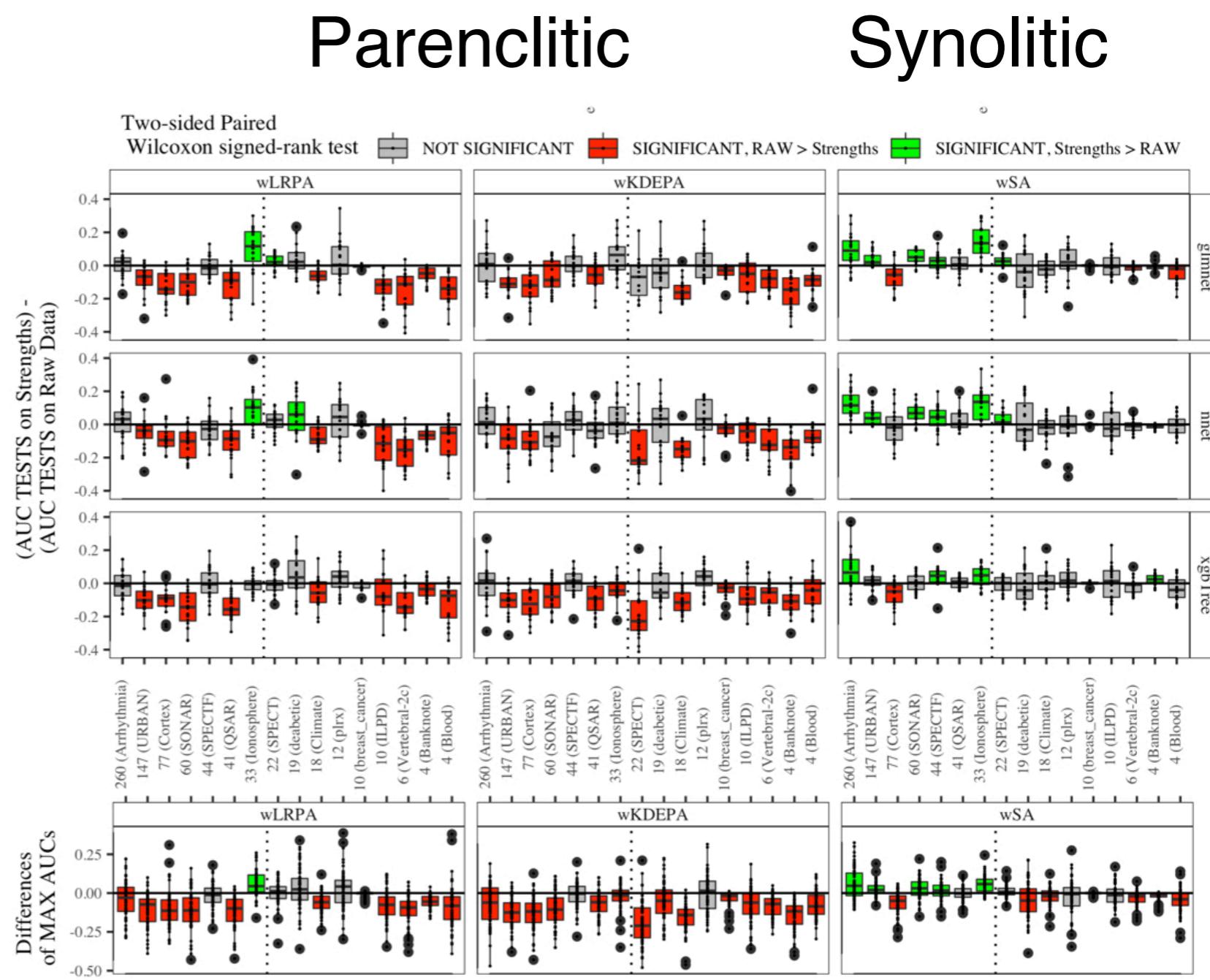
B



REAL DATA

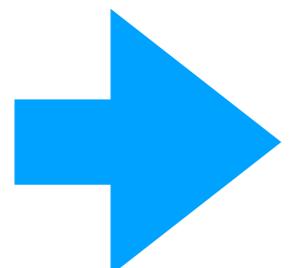
<https://archive.ics.uci.edu>

| Number of | | | | | | Area |
|-----------|---------------------------|----------|---------|-------|----------|-----------------|
| N | Dataset | Features | Samples | Cases | Controls | |
| 1 | Banknote (Ban, 2013) | 4 | 1372 | 610 | 762 | Computer |
| 2 | Blood (Blo, 2008) | 4 | 748 | 178 | 570 | Business |
| 3 | Vertebral-2c (Ver, 2011) | 6 | 310 | 210 | 100 | Medicine |
| 4 | Breast cancer (Bre, 1995) | 10 | 699 | 241 | 458 | Medicine |
| 5 | ILPD (ILP, 2012) | 10 | 583 | 167 | 416 | Medicine |
| 6 | PLRX (PLR, 2012) | 12 | 182 | 52 | 130 | Computer |
| 7 | Climate (Cli, 2013) | 18 | 540 | 494 | 46 | Physical |
| 8 | Diabetic (Dia, 2014) | 19 | 1151 | 611 | 540 | Medicine |
| 9 | SPECT (SPE, 2001) | 22 | 267 | 212 | 55 | Medicine |
| 10 | Ionosphere (Ion, 1989) | 33 | 351 | 126 | 225 | Physical |
| 11 | QSAR (QSA, 2013) | 41 | 1055 | 356 | 699 | Chemical |
| 12 | SPECTF (SPE, 2001) | 44 | 267 | 212 | 55 | Medicine |
| 13 | SONAR (SON, N/A) | 60 | 208 | 97 | 111 | Physical |
| 14 | Cortex (Cor, 2015) | 77 | 1080 | 510 | 570 | Medicine |
| 15 | URBAN (URB, 2014) | 147 | 675 | 122 | 553 | Physical |
| 16 | Arrhythmia (Arr, 1998) | 260 | 452 | 245 | 207 | Medicine |



Open questions and ongoing work:

- 1. Comparison with Correlation graphs**
- 2. Longitudinal Synolic Networks**
- 3. Graph-based Neural Networks**



Personal Patient Tool?

- 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 Blyuss¹, Matthew Burnell¹, Andy Ryan¹, Aleksandra Gentry-Maharaj¹, Inés P. Mariño^{1,2}, Jatinderpal Kalsi¹, Ranjit Manchanda^{1,3}, John F. Timms¹, Mahesh Parmar⁴, Steven J. Skates⁵, Ian Jacobs^{1,6,7}, Alexey Zaikin^{1,8}, and Usha Menon¹

4726 Clin Cancer Res; 24(19) October 1, 2018

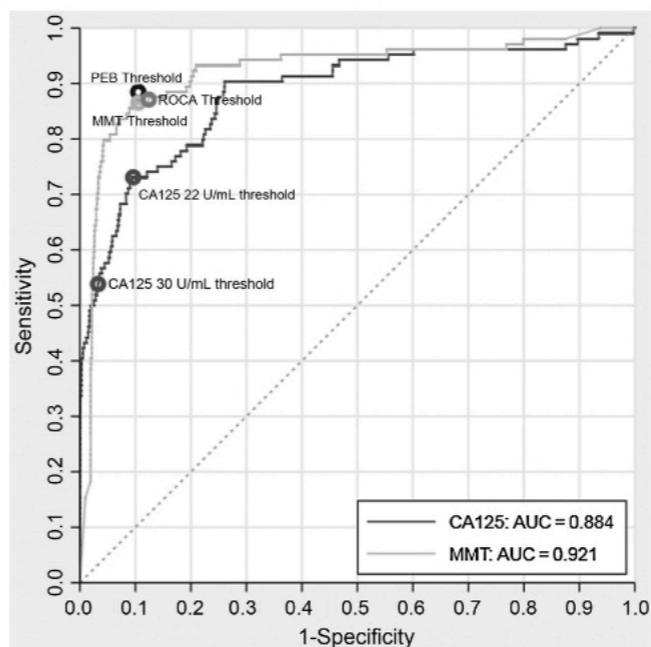


Figure 1.

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.

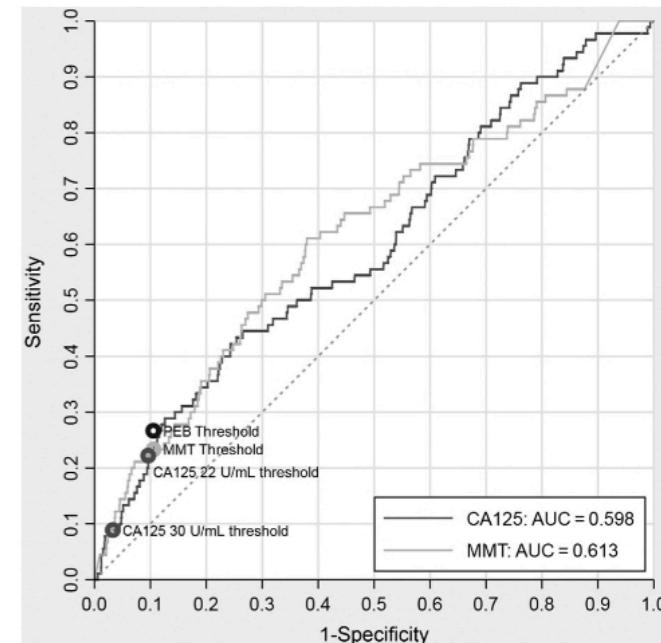
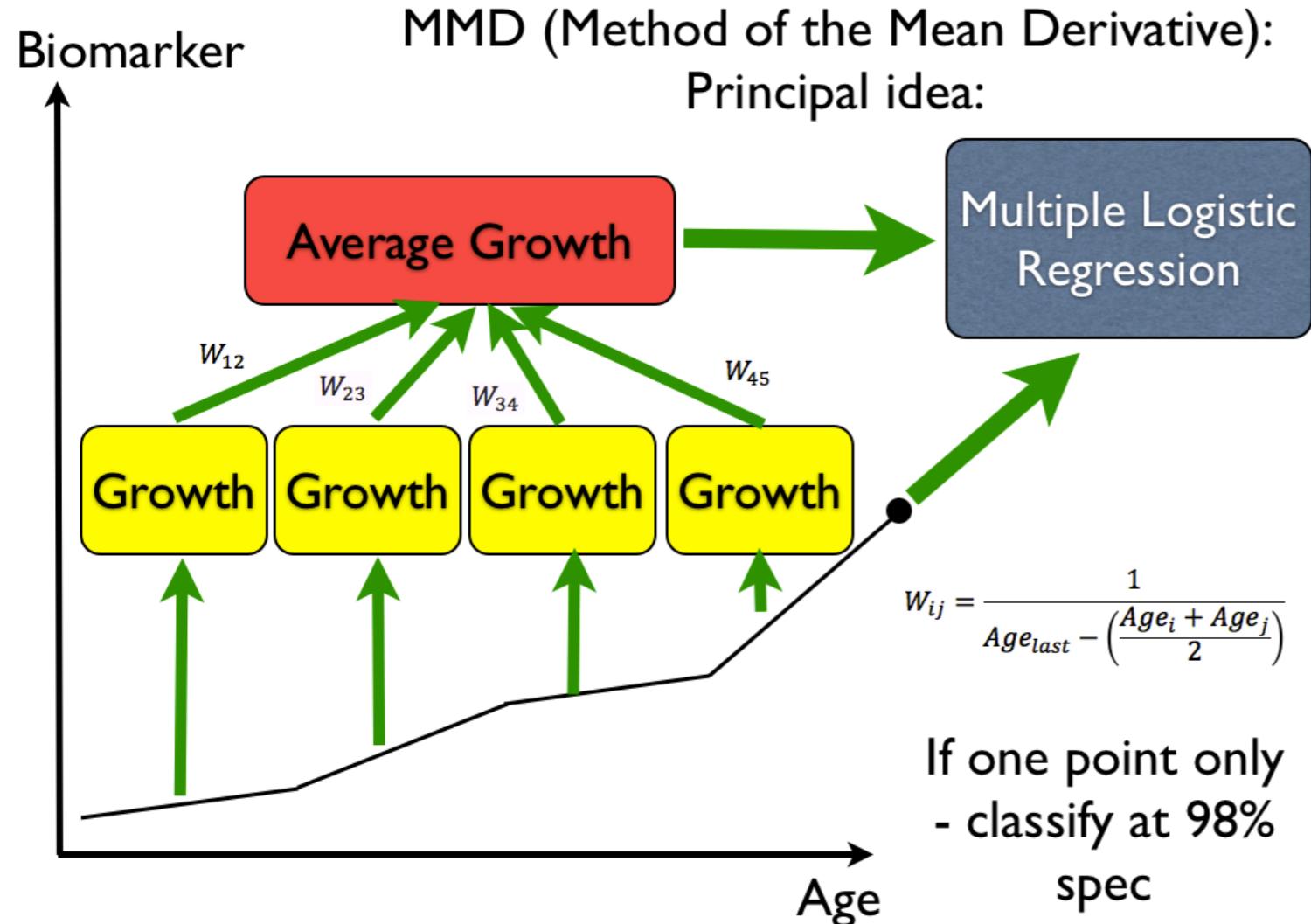


Figure 2.

secondary analysis ROC curves for CA125 interpreted using MMT, threshold rules and 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 corresponding to cutoff values obtained from the training set (MMT and PEB). Abbreviations: PEB, parametric empirical Bayes; MMT, method of mean trends; CA125, cancer antigen 125; AUC, area under roc-curve.

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 Blyuss¹, Matthew Burnell¹, Andy Ryan¹, Aleksandra Gentry-Maharaj¹, Inés P. Mariño^{1,2}, Jatinderpal Kalsi¹, Ranjit Manchanda^{1,6}, John F. Timms¹, Mahesh Parmar³, Steven J. Skates⁴, 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

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

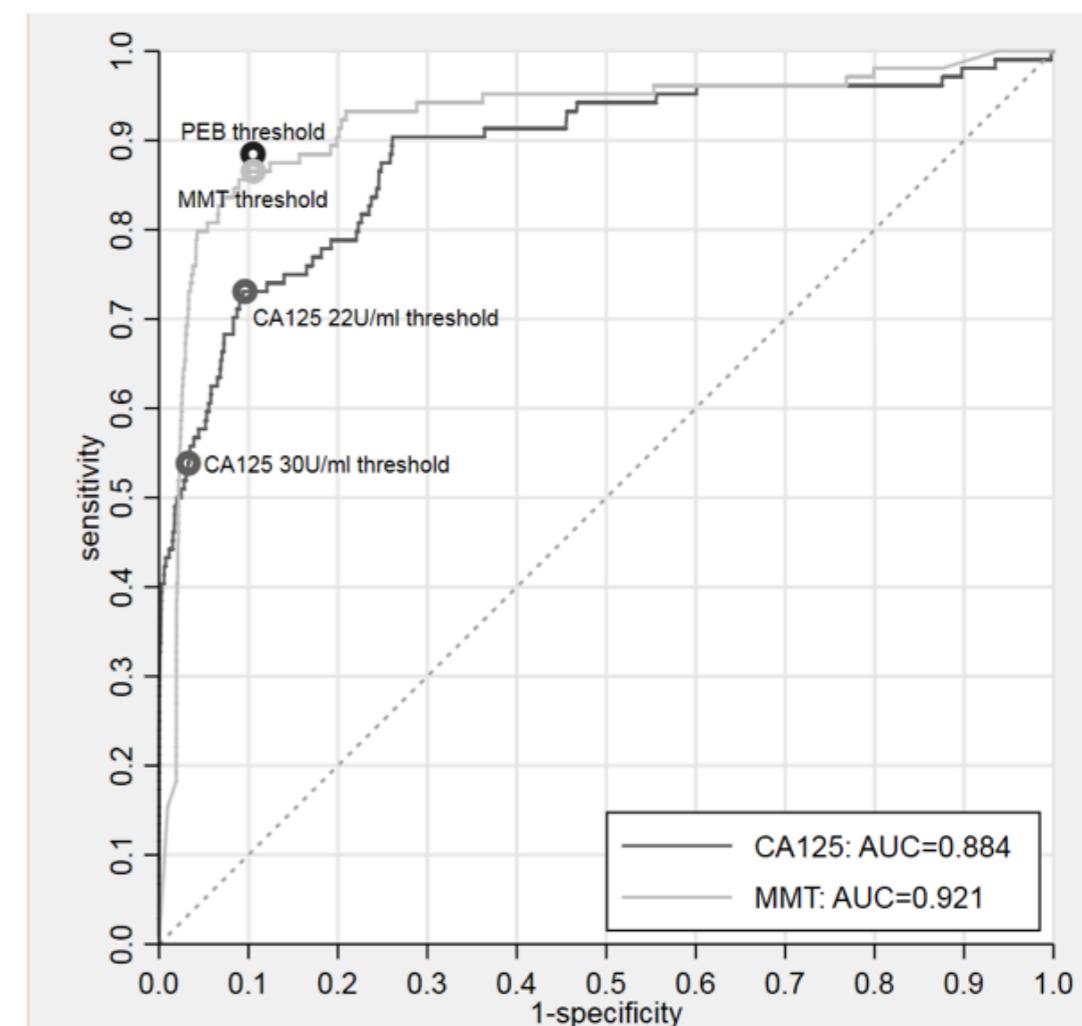
- Trend 3

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

- Trend 4

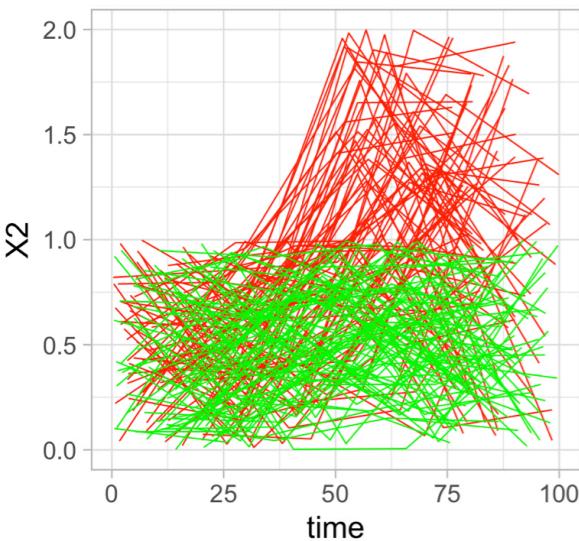
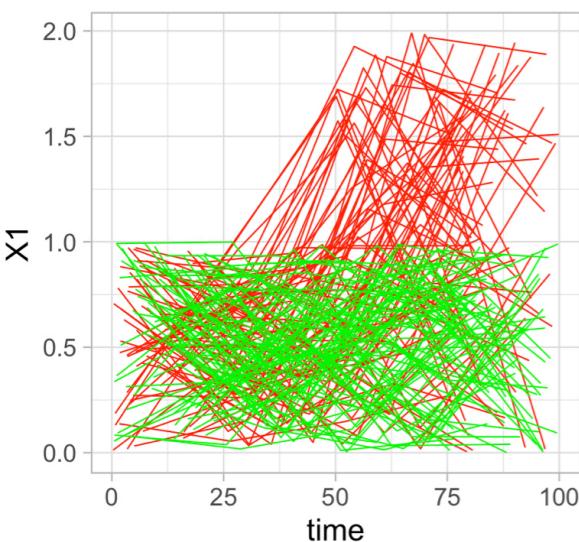
$$\frac{\sum_{j=1}^{k_i} y_j t_j}{\sum_{j=1}^{k_i} t_j}$$

- Trend 5 (Variance)



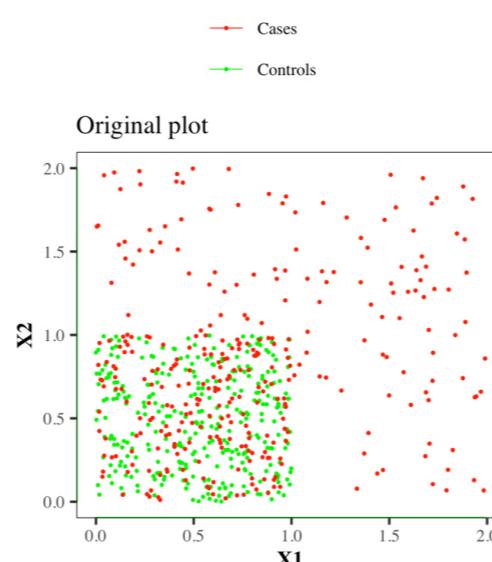
Approach I

Original dependences X1 and X2 on time
class 0 — 1



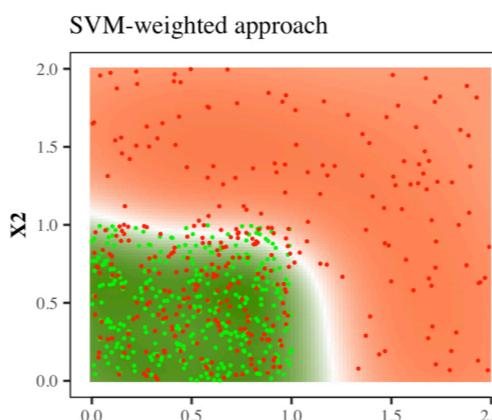
STEP I:

- considering each point as **independent**; ➔



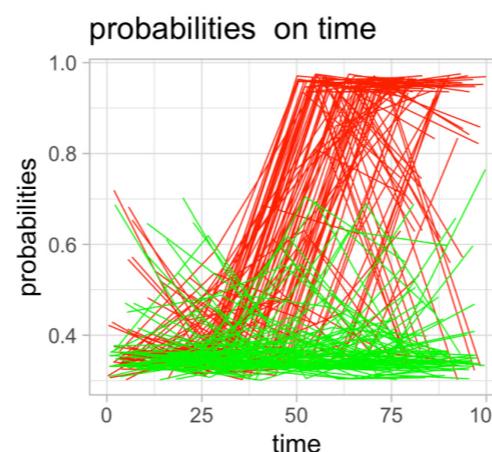
STEP II:

- building simple classification model (SVM, GLM, ...) on X1 and X2; ➔



STEP III:

- get probabilities for each separate point and collect them to longitudinal vectors . ➔



STEP IV:

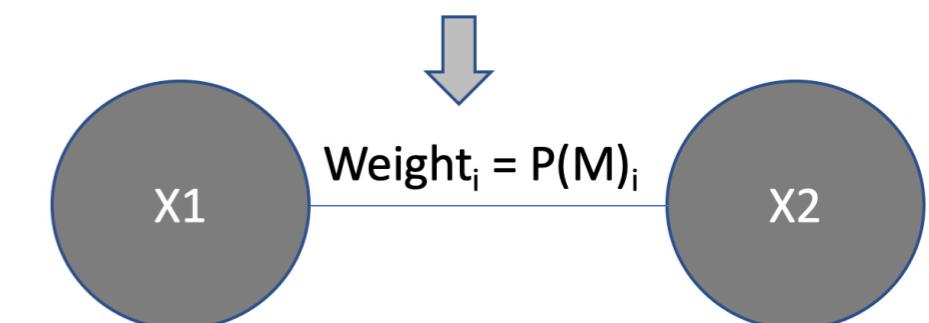
- For each i sample calculating $A_i(P, \text{time})$, $B_i(P, \text{time})$, $C_i(P, \text{time})$, $D_i(P, \text{time})$ indices, where P – vector of probabilities

STEP V:

- Build classification model $M = M(A, B, C, D, \text{score})$, where M – can be any of ML model (xgbTree, glmnet, nnet,...)

STEP VI;

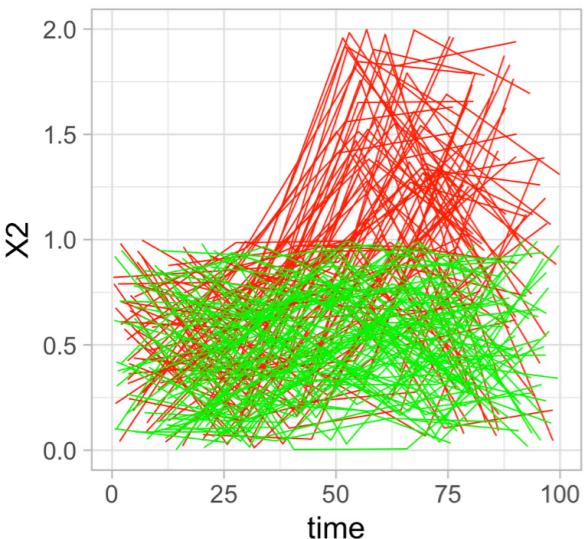
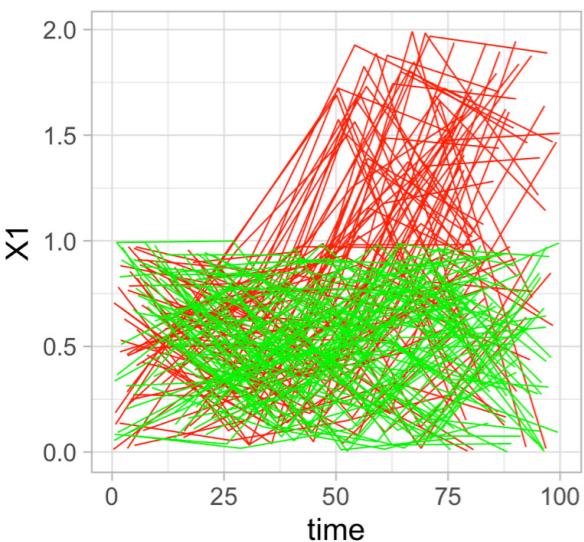
- Use probabilities of M model as weights of edges between X_1 and X_2 vertices



Approach II

Original dependences X1
and X2 on time

class 0 1



STEP I:

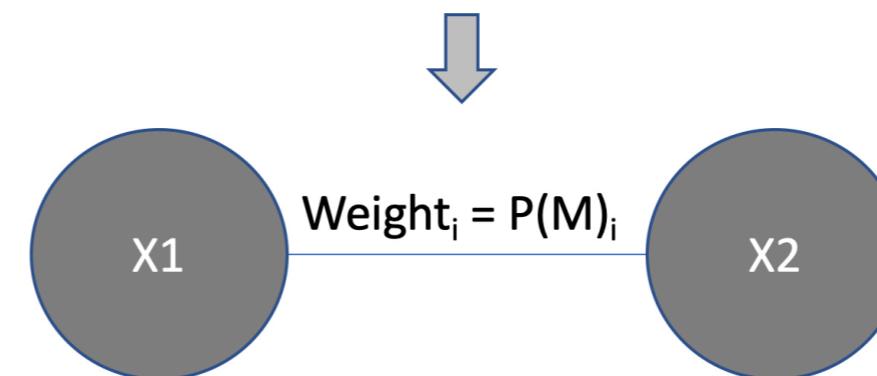
- For each i sample calculating $A_i^{X_1}(X_1, \text{time})$, $A_i^{X_2}(X_2, \text{time})$,
 $B_i^{X_1}(X_1, \text{time})$, $B_i^{X_2}(X_2, \text{time})$,
 $C_i^{X_1}(X_1, \text{time})$, $C_i^{X_2}(X_2, \text{time})$,
 $D_i^{X_1}(X_1, \text{time})$, $D_i^{X_2}(X_2, \text{time})$ indices;

STEP II:

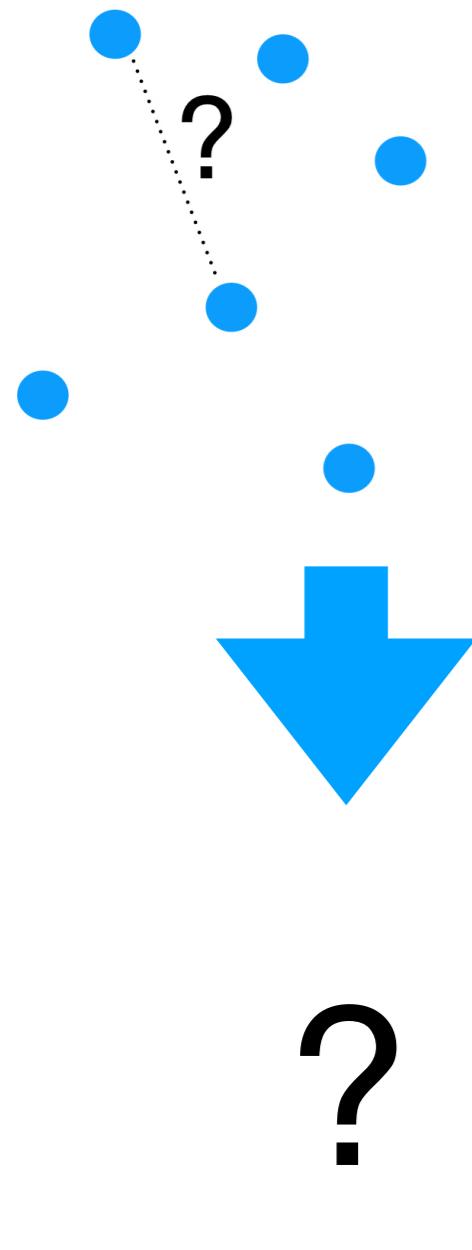
- Build classification model
- $M = M(A^{X_1}, B^{X_1}, C^{X_1}, D^{X_1}, A^{X_2}, B^{X_2}, C^{X_2}, D^{X_2}, \text{score})$, where M – can be any of ML model (xgbTree, glmnet, nnet,...)

STEP III:

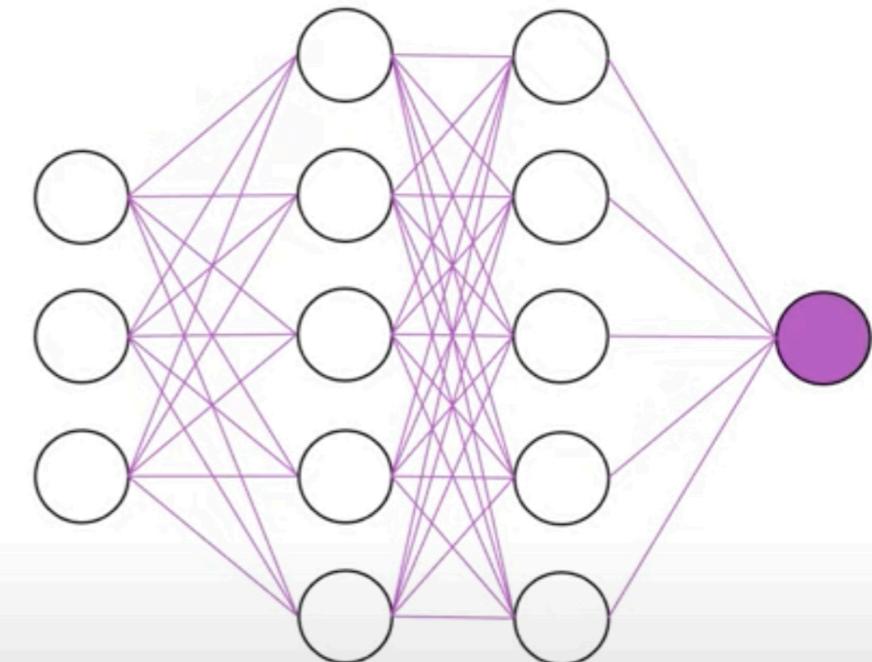
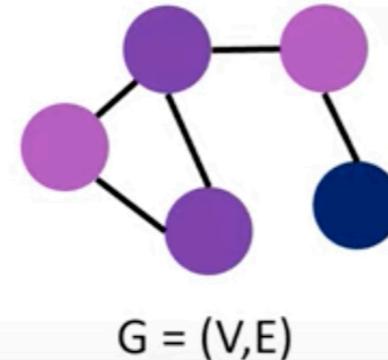
- Use probabilities of M model as weights of edges between X_1 and X_2 vertices



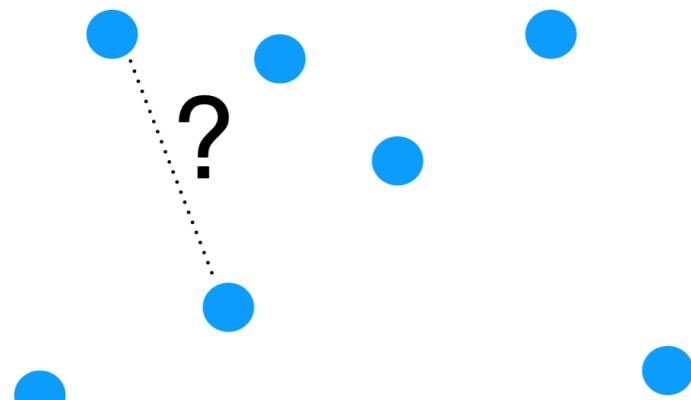
Combination with Graph Neural Networks



Neural Networks

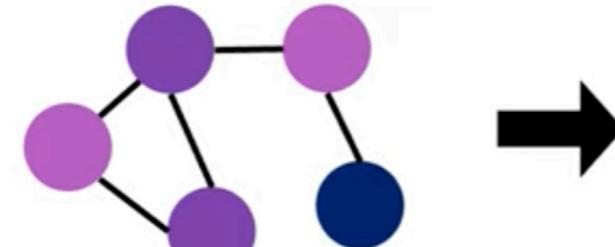
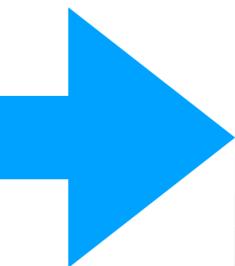
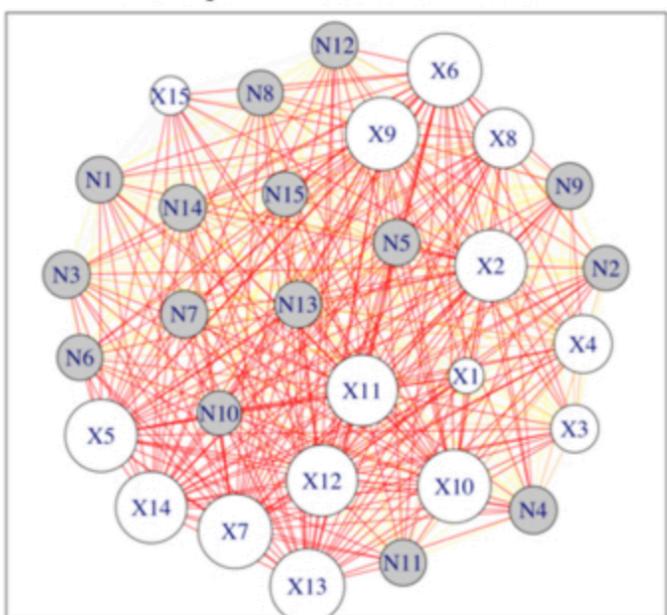


Combination with Graph Neural Networks

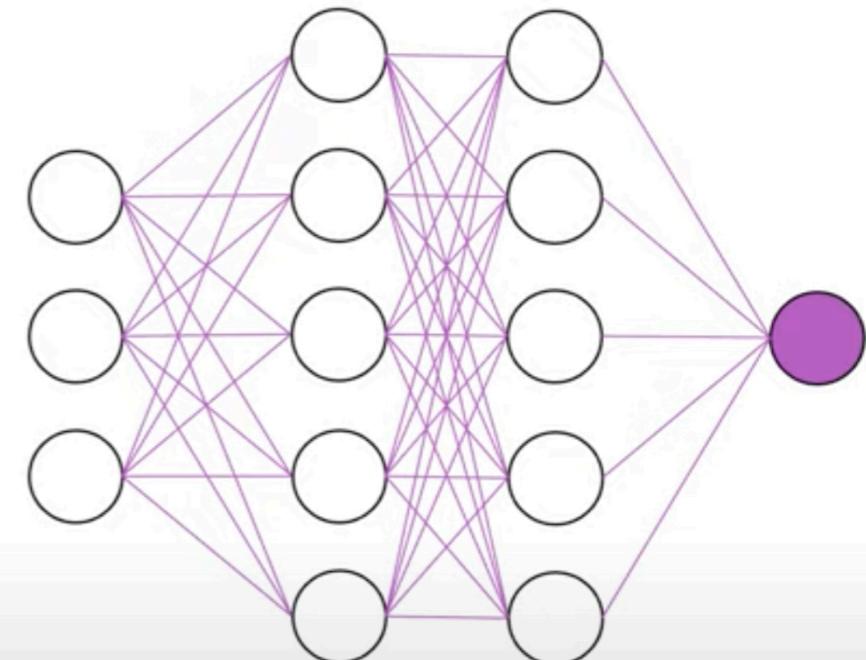


Neural Networks

Sample with $R = 0.99$



$G = (V, E)$



Comparison of the results of parenclitic approaches

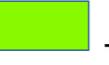
The Graph Convolutional Network (GCN) vs ML models (glmnet, nnet, xgbTree), trained on the strengths of vertices on synthetic data

Color bars — confidence intervals of mean (AUCS TESTS of GCN) - (AUCS TESTS of model on Strengths))

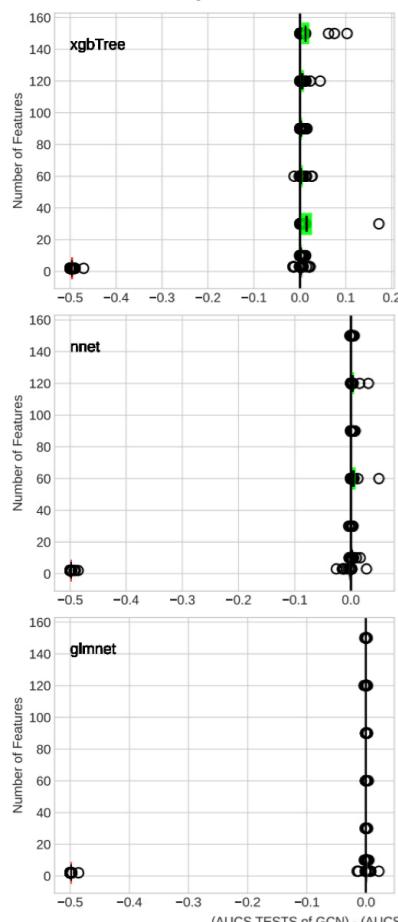
Vertical black line in color bars – the mean

Color of bars, according to two-sided Paired Wilcoxon signed-rank test

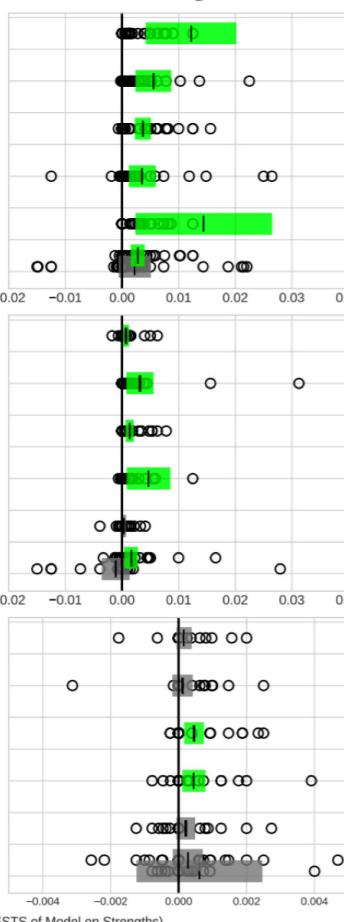
Open circles – individual differences of mean

 - NOT SIGNIFICANT,  - SIGNIFICANT, GCN is better,  - SIGNIFICANT, ML Model on Strengths is better

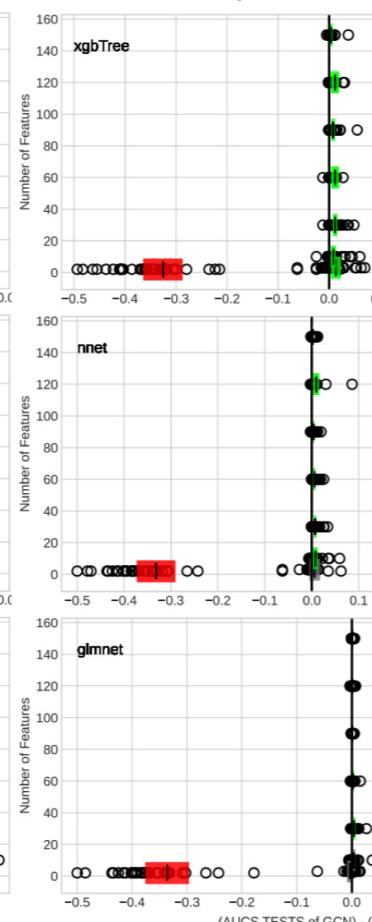
Ideal Spheres



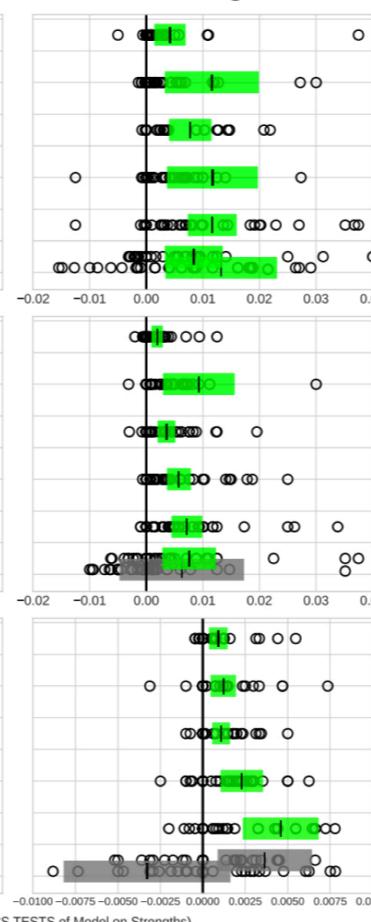
zoomed figure



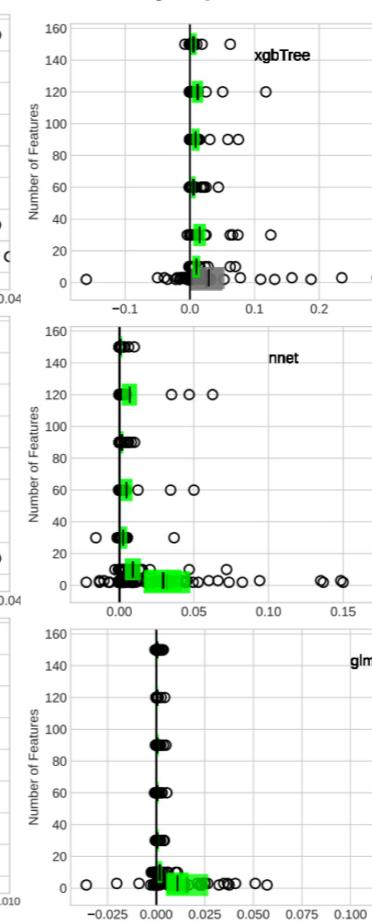
Broken Spheres



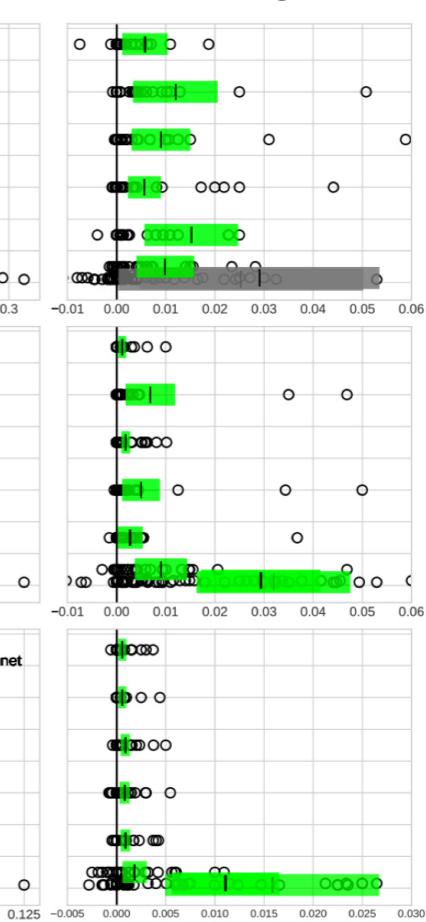
zoomed figure



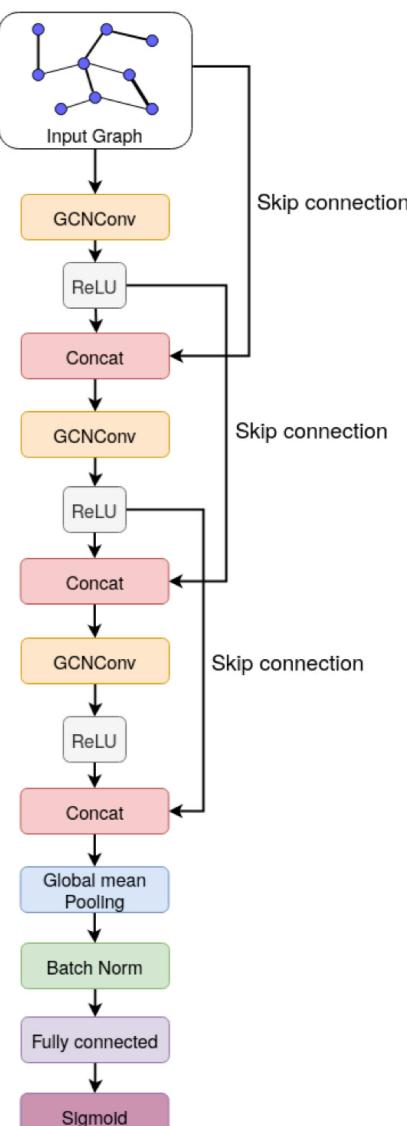
Noisy Spheres



zoomed figure



Network architecture



Tanya Nazarenko

Harry J. Whitwell

Oleg Blyuss

John F. Timms

Alexey Zaikin

Thomas Bartlett

Sofia Olhede

Usha Menon



Vadim
Demichev

Markus Ralser

lab

Clara Correia-
Melo

Anja Freiwald

Oliver Lemke

Christoph
Messner

Annika Röhl

Lukasz
Szyrwiel

Spyros
Vernardis

Matt White

Charité, Berlin

Florian Kurth

group

Pinkus Tober-Lau

Charlotte

Thibeault

Leif Sander group

Michael Mülleder

group

Wolfgang Kübler

group



Italy

Claudio Franceschi

Maria Julia Bacalini

Innsbruck

Ivan Tancevski group

Markus Keller group

NNGU

Misha Krivonosov

Misha Ivanchenko

Maria Vedunova

Spain

Ines Marino

Sechenov University

Vadim Ushakov

Alexander Suvorov

THANK YOU!!