Network markers of DNA methylation in neurodegenerative diseases

Mikhail Krivonosov
Laboratory of Systems Medicine of Healthy Aging
N. I. Lobachevsky State University
Nizhny Novgorod, Russia
krivonosov@itmm.unn.ru
0000-0002-1169-5149

Tatyana Nazarenko
Department of Mathematics and Institute for Women's Health,
University College London
London, UK
0000-0002-4245-7346

Maria Giulia Bacalini
IRCCS
Istituto delle Scienze Neurologiche di Bologna
Bologna, Italy
0000-0003-1618-2673

Alexey Zaikin
Department of Mathematics and Institute for Women's Health,
University College London
London, UK
0000-0001-7540-1130

Mikhail Ivanchenko
Department of Applied Mathematics and Laboratory of Systems Medicine of Healthy Aging
N. I. Lobachevsky State University
Nizhny Novgorod, Russia
0000-0002-1903-7423

Claudio Franceschi
Laboratory of Systems Medicine of Healthy Aging
N. I. Lobachevsky State University
Nizhny Novgorod, Russia
claudio.franceschi@unibo.it

Abstract—**One of the ways to reveal underlying complexity in high-dimensional data is a parenclitic network approach. Analysis of DNA methylation requires an understanding of interrelations between different genes that can be achieved through reconstructing unknown connectivity. In this paper, we identify network patterns that occur in DNA methylation of Down syndrome patients associated with aging and disease processes.**

Keywords—*complex networks, parenclitic networks, DNA methylation, Down Syndrome*

I. INTRODUCTION

In recent years, experimental biology is generating an enormous amount of omics data [1]. Dealing with such a huge information flow requires novel algorithms in computer science. Standard task for machine-learning field such as classification and regression have at present a handful solutions. However, next-generation biomedical data poses a new challenge of disclosing hidden links between observable features. The lack of prior information on the interdependencies in multi-dimensional data can be addressed by the recently introduced parenclitic approach [2]. It reconstructs a network by inferring connections between measured covariates in order to disclose the systems level information about diseases, for example in cancer [3].

One of the important topics in biomedicine is the identifying molecular basis of complex disorders such as behavior plasticity, memory, cancer, autoimmune disease, as well as neurodegenerative and psychological disorders [4]. The fundamental factors of those diseases are usually investigated at a genetic level. However, the complex interaction between the genome and environment leads to epigenetic modifications, in particular DNA methylation. Patterns of DNA methylation vary with age, cell development, and in disease [5].

High-dimensional data obtained by DNA methylation sequencing pose a problem of extracting clinically useful information from Big Data. In this work, we develop and apply a novel method of parenclitic network construction and analysis of DNA methylation network signatures for patients with Down Syndrome and relate it to the aging process. New insights into the age-dependent epigenetic markers network can give additional information about unknown functional dependencies and possible targets for medical treatment and control.

II. DATA AND ALGORITHMS

A. Dataset

The analysis has been performed on DNA methylation data profiles of blood cells in families with Down syndrome children. Normalized data are publicly available in the repository NCBI Gene Expression Omnibus under accession number GSE52588 [6].

Dataset consists of 29 families: mother and two children - one of them is healthy and the other one has a Down syndrome disease. We distinguished a separate group of mothers (DSM), healthy siblings (DSS), and children with Down syndrome disease (DS).

The HM450 beadchips collect methylation information using and after that, the β-value is calculated. One subject is characterized by 450k scalar numbers β-values of methylation level for considered sites.

B. Algorithms

Enhanced implementation of the parenclitic approach allows us to build networks considering a high number of CpG sites. The idea of the algorithm is to assign subjects that deviate from the control group by pairs of features with a link between those features. Deviation on the two-dimensional plane can be computed by any machine learning classifier. In our work, we use support vector machine (SVM) and probability density function (PDF) approaches to classify subjects into groups. Found links represent the reconstruction of individual networks.

Using computed subsets of network-based CpG sites, we perform biological significance analysis.

III. RESULTS

In this work, we analyze a subset of CpGs, which lie on genes and restricted only to Islands and Shores parts of the gene [6]. Using the developed implementation of the parenclitic approach, we build individual networks. We use three different configurations of groups:

- Mothers as control and subjects with Down syndrome disease as cases.

- Healthy siblings as control and DS subjects as cases.

- Both mothers and siblings as a control group and DS subjects as cases.

Parenclitic network identifies a set of significant pairs of CpGs in each configuration. Each configuration produces its signature of corresponding processes. Those configurations are associated with distinguishing possibilities:

- In this case, we consider differences in methylation patterns induced by two factors: aging and disease.

- Second case defined by disease of Down syndrome subjects versus healthy siblings and not associated with chronological age, but can be associated with biological age.

- The third case attribute only to disease deviations compared to healthy subjects of all ages.

This position allow us to determine CpG sites and corresponding genes that underlie the disease and aging processes. Besides, apart from a more complete list of CpG sites that involved in different disease process we also present hidden interconnections between them. An example of a detected CpG pair is shown on Fig. 1.
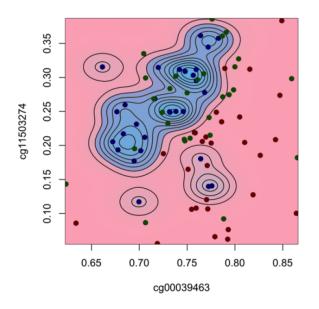

Fig. 1. Pair of CpG sites found based on siblings as control and DS subjects as case.

The next step is performing gene ontology analysis on selected CpG sets. Processes that can be associated with Down syndrome disease from GO presented in Table 1.

TABLE I.　GO ANALYSIS RESULTS.

Ontology	p-value	p-adj	ID
pattern specification process	1.83E-09	2.68E-06	GO:0007389
DNA-binding transcription activator activity, RNA polymerase II-specific	1.34E-08	6.10E-06	GO:0001228
regionalization	1.38E-08	6.10E-06	GO:0003002

IV. CONCLUSION

In the paper we developed a novel method based on parenclitic network approach for identifying network epigenetic signatures in DNA methylation sequencing data. The obtained results are of general interest for the systems biology of Down Syndrome, and offer means to identify novel molecular targets to prevent accelerated aging of patients. We find that it is possible to decompose the epigenetic signature associated with Down syndrome disease into age-dependent and disease-dependent parts based on the hidden links between the covariates.

Developed approaches can be widely used in similar problems with high-dimensional data.

REFERENCES

[1] Hassanien A. Big data in complex systems. Cham: Springer, 2015.

[2] Zanin M. et al. Parenclitic networks: uncovering new functions in biological data // Scientific Reports. 2014. vol. 4, number 1.

[3] Karsakov A., Bartlett T., Ryblov A., Meyerov I., Ivanchenko M., Zaikin A. (2017) Parenclitic Network Analysis of Methylation Data for Cancer Identification. PLoS ONE 12(1): e0169661

[4] Weinhold B. Epigenetics: The Science of Change // Environmental Health Perspectives. 2006. T. 114. № 3.

[5] Jung M., Pfeifer G. Aging and DNA methylation // BMC Biology. 2015. vol. 13, number 1.

[6] Bacalini M. et al. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome // Aging. 2014. vol. 7, number 2. p. 82-96.