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

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

[1]Department of Applied Mathematics and Laboratory of Systems Medicine of Ageing, Lobachevsky University, Nizhny Novgorod, 603950, Russia
[2]Department of Mathematics and Institute for Women's Health, University College London, London, WC1H 0AY, UK
[3]Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, 40126, Italy
[4]Centre for Analysis of Complex Systems, Sechenov First Moscow State Medical University, 119146, Moscow,Russia
[†] Equally contributed
[*] Corresponding author: t.nazarenko@ucl.ac.uk

ABSTRACT

Despite the cause of Down Syndrome (DS) is well established, the underlying molecular mechanisms that contribute to the syndrome and to the phenotype of accelerated ageing remain largely unknown. DNA methylation profiles are largely altered in DS but it remains unclear how different methylation regions and probes are structured into a network of interactions. We develop and generalize the Parenclitic Networks approach that enables to find correlations between distant CpG probes (and not pronounced as stand-alone biomarkers) and quantifies hidden network changes in DNA methylation. DS and a family-based cohort (including healthy siblings and mothers of persons with DS) is used as a case study. Following this approach, we constructed parenclitic networks and obtained different signatures informative for (i) the differences between individuals with DS from healthy individuals; (ii) differences between young and old healthy individuals and determining the place of DS individuals on this scale; (iii) differences between DS individuals from their age-matched siblings, and (iv) the difference between DS and the adult population (their mothers). Conducted Gene Ontology analysis showed that the CpG network approach is more powerful than the single CpG approach in identifying biological processes related to DS phenotype, like those involved in central nervous system, skeletal muscles, disorders in carbohydrate metabolism, cardiopathology, and oncogenes. Our open-source software implementation is accessible to all researchers. The software includes a complete workflow to construct Parenclitic Networks with any machine learning algorithm as a kernel to build edges. We anticipate a broad applicability of the approach to other diseases.

Introduction

Epigenetic modifications are chemical and/or physical changes in chromatin that can be preserved across cell divisions, and DNA methylation is the most studied among them [1]. This is in part due to the fact that this molecular layer tends to be relatively stable and can be assessed by cost-effective approaches, like the Infinium microarrays, that allow to measure DNA methylationof hundreds of thousands of CpG loci with single base resolution at the genome-wide level [2]. In addition, a bulk of literature demonstrated that DNA methylation changes occurs both in physiological (development, sex differentiation or aging) and in pathological (cancer, age-related diseases or genetic syndromes) conditions [1,3,4].

So far, most of the studies that have investigated DNA methylation changes associated to a certain condition focused on single CpG sites (differentially methylated positions) or groups of chromosomally adjacent CpG sites (differential methylated regions) [5,6]. However, the epigenetic regulation of the genome is a complex and highly integrated system, in which DNA methylation modules are not isolated [7]. Analytical approaches that consider the cross-talk between DNA methylation alterationsin different regions of the genome could therefore provide new insights on the epigenetic remodelling occurring in various physiological and pathological conditions, as it has been recently demonstrated [8-11].

In this light, the methods based on network analysis have the potential to provide deeper insights into the regulation of DNA methylation profiles. Among others, the parenclitic network approach [12] is of particular interest, as it enables to

representmethylation data in the form of a network even if functional links between different methylation probes or regions are not established.

Briefly, this approach implements networks, whose nodes correspond to the relevant parameters, and assigns edges between nodes, if their values significantly deviate from the appropriately defined Control set (group). A measure of deviation is usually introduced on a plane of two coordinates that correspond to the pair of nodes, e.g. Zanin et al. implemented the distance between a data point to the linear regression line build on a Control group [13]. Since that pioneering work, parenclitic networkshave been successfully applied to detecting key genes and metabolites in the context of different diseases [13-16], see [12] for a review and [17] or a discussion of applications in brain research. In [18] we have applied this methodology to realize machine learning classification of human DNA methylation data carrying signatures of cancer development. Later [19], acknowledging that the interactions of two features (at least in biological systems of biomarkers) often cannot be described by linear model, itwas proposed to use 2-dimensional kernel density estimation (2DKDE) to model the Control distribution.

The main difficulty in using such approaches is implementing an end-to-end analysis package (incorporating many technically-complex nested steps), in particular, avoiding *ad hoc* choices of deviation measures and subjective definition of "cut-off thresholds" to discriminate small and large deviations in edge assignment. Here we present a solution of these issues that employs machine learning algorithms for parenclitic network construction.

We demonstrate the validity of the method analyzing DNA methylation changes for patients with Down Syndrome (DS). Despite the fact that the cause of DS is well established, the underlying molecular mechanisms of associated epigenetic modifications remain poorly understood. Previous studies have showed that Down Syndrome (DS), which is caused by a full or partial trisomy of chromosome 21, is characterized by a profound remodeling of DNA methylation patterns, which involves not only chromosome 21 but is widespread across the genome [20-23]. DS is a segmental progeroid syndrome, as it is characterized by a phenotype of premature/accelerated aging that occurs in a subset of organs and systems, including the immune and the nervous systems [24-25]. Based on epigenetic clock of 353 CpG sites [9], Horvath et al demonstrated accelerated epigenetic aging in DS against age-matched Controls, manifested in blood and brain tissues [26].

Among the different cohorts investigated so far, the GSE52588 dataset is of particular interest, as a family-based cohort, that allows to evaluate not only the epigenetic remodelling in DS, but also the potential contribution of genetic background and environment, which are at least in part shared by the members of a family [20]. The analysis that focused on differential methylation probes and regions on this dataset have highlighted epigenetic alterations in genes involved in the developmental functions (including haematological and neuronal development), metabolic functions and regulation of chromatin structure, in line with the results on independent cohorts [20-23].

In this paper we employ the parenclitic network analysis to get further insight and quantify complex topological changesof the epigenome in disease, in other words, changes, which can be very different for particular DNA methylation probes for individuals, but share a similar topological profile, distinct from that of the healthy subjects. We also identify genes and molecular pathways associated with the differential network DNA methylation in Down Syndrome, not found in single or local region CpG analysis. The paper is structured as follows. Firstly, we apply our algorithm to investigate the age-dependent network signatures of Down Syndrome on methylation data. In particular, we report the signature of the pure DS disease (regardless of age); the signature of age-related changes in a healthy population (which manifests the proximity of DS patients cohort to an older healthy population); the signature, which changes with age in a healthy population more slowly than in DS patients; and a signature, which changes in a healthy population with age faster than in DS patients. We also conducted the Gene Ontology and KEGG analysis and found associations with the nervous system, cell fate and oncogenesis, cellular communication, development of metabolic syndrome, and female sex hormones.

Secondly, (this information is mainly presented in Supporting Information) we present the implementation of an end-to-end workflow to construct Parenclitic Networks, that is freely available for download; we demonstrate that machine learning algorithms can be chosen as a kernel to build edges in the network using geometric and probabilistic approaches as examples. We also present a new approach (PDF-adaptive) that allows an automated choice of the “cut-off threshold” to remove insignificant edges.

Materials and Methods

Down Syndrome methylation data

As an application and validation of our approach, we consider a publicly available dataset (GSE52588), where the 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).

Age distribution of subjects from the 3 groups is shown in Fig.1(B). As depicted in Fig.1(A), this family-based model (which minimizes potential confounding effects, as members of the same family tend to share the same habits/environment and their genetic background) allows 4 different comparisons, in which 2 phenotypes (DS and aging) are differently combined:

1. comparing between DS against both DSM and DSS yields a healthy phenotype;
2. comparing between DSS vs DSM is informative of epigenetic remodeling occurring during ageing in euploid subjects;
3. comparing between DS vs DSS is informative of epigenetic remodeling associated to the syndrome in the age-matched groups;
4. comparing between DS and DSM discriminates the effects of the syndrome and aging.

Preprocessing

Raw *.idat* files extracted and pre-processed using *minfi* package [27]. Data were normalized using the *preprocessFunnorm* function and probes with detection p-value higher than 0.05 in more than 1% of samples were removed. Furthermore, cross-reactive and polymorphic probes, as reviewed by Zhou et al [28], and probes on the X and Y chromosomes were removed. Finally, the procedure left 412993 probes from the original array. In our study, we are focusing only on a subset HM450 probes from Islands and Shores regions, that includes 114674 CpGs. Additionally, we exclude sex-specific CpGs to decrease possible bias, relying on the list of sex-specific CpGs [29] with 5665 items. Lastly, the list contains 113630 CpGs (full list of such CpGs can be found in SM-files: in *background_no_sex_specific_cpgs.xlsx* file).

Generalized Parenclitic Network Analysis

The implementation of a complete workflow to construct Parenclitic Networks is available at https://github.com/mike-live/parenclitic (see Supporting information for a detailed description). The analysis is intended to grasp co-occurring differential changes in DNA methylation of the pairs of probes, and we first opt out the probes, that are individual good separators of the Control and Case groups. That is, for each CpG probe we build a split border classifier based on maximizing the information gain, and move those providing accuracy of more than 75% to separate lists (they can be found in SM-files: in *1d_(age/ds/m/s)_control_model_cpgs.xlsx* files). Such probes are further excluded from the set of nodes of parenclitic networks, and the results are further considered through the prism of single CpG based and co-occurring differential CpG methylation.

Overall, we consider four model types of Parenclitic networks, setting different Control-Case combinations (see illustration on Fig.5 in Supporting information for more details):

1. **DS-Control Network** — the Parenclitic network model of healthy phenotypes, where DS group is *Control* group, DSS and DSM groups are *Case* group;
2. **AGE-Control Network** — the Parenclitic network model of aging, where DSS group is *Control* group, DSM group is *Case* group, DS group is a *Test* group;
3. **S-Control Network** — the Parenclitic network model of the syndrome, where DSS group is *Control* group, DS group is *Case* group, DSM group is *Test* group;
4. **M-Control Network** — the Parenclitic network model of aging and syndrome, where DSM group is *Control* group, DS group is *Case* group, DSS group is *Test* group.

Parenclitic networks for individual subjects are constructed according to the following rules. For each pair of CpG sites (i, j) we point subjects on a corresponding two-dimensional plane according to beta-values, and mark Control and Case groups. Fig.1(C,D) illustrates the case, when DSM is a Control group and DS and DSS constitute a Case group. We build classifiers for the Case and Control groups, making use of one of the three different classification methods: PDF reconstruction, SVM classification and a PDF-adaptive (the best threshold) method. If the separation accuracy between the Control and Case groups is less than 90%, the edge between the network nodes (i, j) is not assigned for all subjects. Otherwise, the edge (i, j) is assigned for a particular subject if it is classified as the Case group. Finally, having repeated the procedure for all pairs of CpG probes, we arrive at individual networks for each subject such that nodes correspond to probes and edges between them designate that a subject falls out of the Control group with regard to methylation of a particular pair of CpGs (Fig.1(D)).

Where appropriate, we consider the third phenotypic group, that we coin “Test”, and investigate the closeness of its parenclitic networks to the Cases and Controls.

In addition to the obtained networks for each patient separately, we also build "general" networks for each DS/AGE/S/M-Control construction. In each model, we leave only those edges that correspond to separation between Cases and Controls with at least 90% accuracy. In the Fig.2(B,C,D,E) we present reduced copies such general DS/AGE/S/M-networks (enlarged images can be found by the links given in the caption of this figure; constructions respectively are given SM-files: in *(ds/age/s/m)_union_network.xlsx* and lists of vertices of networks are given SM-files: in *(ds/age/s/m)_union_network_list.xlsx*). The size of each vertex in this representation is associated with the degree of the vertex in the network (that is, the sum of the edges included in it). Thus, the larger the vertex, the more often this CpG-site has

participated in the "successful" separation of cases and controls in a pair with another CpG-site. Fig.2(A) shows the Venn Diagram for the vertices (CpG sites) of networks, which demonstrates that each general network is predominantly unique. Further, for each of these generalizing networks (based on the vertices selected in them), we will carry out GO and KEGG analysis, and recover biological processes that corresponding to differential methylation in Cases and Controls.

DNAmAGE and epigenetic age acceleration

DNAmAGE (DNA methylation age) was calculated using the online tool available at https://dnamage.genetics.ucla.edu/. Epigenetic age acceleration values (residuals) were calculated as the difference between the Horvath's epigenetic age and chronological age.

Functional enrichement analysys

Gene Ontology enrichment was using the *methylgometh* function implemented in the methylGSA Bioconductor package [30]. In the *Generalized Parenclitic Network Analysis* section, we describe the lists of one-dimensional separators (those CpGs, which in themselves allow distinguishing between the groups of cases and controls) and CpGs lists that fall into the general parenclitic networks. We conducted GO analysis for each approach (1D and parenclitic) separately. The results of this analysis can be found in Table 2 (for Parenclitic) and Table 3(for 1D), Supporting Information. As the parenclitic approach does not return an explicit p-value, the input file for *methylgometh* function was created by assigning 0.001 to all the CpG sites selected by the parenclitic analysis, and 0.9 to the other CpG sites. A significance cut-off 0.01 was then used. REVIGO(http://revigo.irb.hr/) was used to remove redundant GO terms. The function enrichement analysis was used for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on genes corresponding to CpGs in the parenclitic approach. Significantly enriched KEGG pathways with Benjamini < 0.05 and $P < 0.05$ were selected. Results of KEGG analysis are presented in the Table 1, Supporting Information.

Results

Networks analysis and identification of network-age-dependent effects

We previously demonstrated [20,26] that individuals with DS tend to show a higher epigenetic age compared to age-matched Controlsby Horvath's clocks, and found some signatures of DS, based on the local differential DNA methylation analysis. In this work, we conduct for the first time a networks analysis of differential methylation in DS, taking into account genome-wide associations, and provide network DNA methylation signatures of DS based on co-occurrence of probe methylations modifications.

The proposed method allows for obtaining individual parenclitic networks for each subject. In this way, we analyzed the statistical properties of such networks (betweenness, pagerank, closeness, eigenvector centrality, nodes degrees, number of edges and nodes, maximum connected component, efficiency, robustness) and demonstrated that most of them allow for a high quality of class separability (see, for example, Figures 6, 9,12 and 15 in Supporting Information).

This method also determines signatures (specific CpG sites) associated with each of DS/AGE/S/M-Control models (see Fig.2(B,C,D,E)). These signatures are essentially unique (see Fig.2(A)) for each model, and specificity is also observedin different aging trends (for example, in the characteristics of the M-control network, the DS group is closer to young representatives healthy population, and in the characteristics of the S-network - on the contrary, closer to the elderly) (see Fig.3).

We illustrate findings focusing on the number of non-zero degree edges in the parenclitic network. This quantity characterizes the degree of difference of DNA methylation patterns in the individual from the Control group in terms of differential network CpG methylation. The observed trends in each model case (especially, in the behavior of the *Test* group), allows the following generalizations (Fig.3):

- Healthy phenotype model network (DS-control): here DSM and DSS show similar networking properties and are very different from DS (see, for example number of non-zeros degree nodes characteristic Fig.3(A) and other characteristics on Fig.6 in Supporting Information). We assume that the signature of this network (the involved nodes) is responsible for the pure (regardless of age) difference between the healthy population and patients with DS;

- Parenclitic model of aging (Age-control): DS network signatures are closer to DSM than to DSS and get closer with age (see. Fig.3(B), and also Fig.9 in Supporting Information), this fact corroborates with the established concept of accelerated aging in DS. Examining the behavior of network characteristics against age acceleration (see, for example number of non-zeros degree nodes characteristic Fig.4 and other characteristics on Fig.10 in Supporting Information), we observe that the greater is absolute age acceleration, the closer patients with DS are to mothers. Notably, this

trend is seen for both the positive and negative age accelerations. The latter could indicate the limitations of Horvath's epigenetic clocks for some special phenotypes of DS patients, where formally negative age acceleration does not reflect actual accelerated biological aging. In this way, the parenclitic network signatures behave more robustly, consistently manifesting the difference between the healthy elderly population and the healthy young one, and the proximity of DS patients to the mothers - the elderly group (indicating a strong effect of age-accelerating for DS patients);

- Parenclitic model of DS (S-control): DSM network signatures are different from DSS, and get closer to DS with age (see, Fig.3(C), and other characteristics on Fig.12 in Supporting Information), also in line with the concept of age-acceleration in DS. We infer that the signature of this network (the involved nodes) captures the difference betweenthe DS patients and their siblings. Moreover, the trend in the group of mothers suggests that in a healthy population this signature changes with age slower (and only with increasing age does it approach the state of the DS group);

Parenclitic model of both aging and syndrome (M-control): DSM network signatures are different from DS and DSS group trend shows that this group moves away from DS peers and gets closer to mothers with age (see, for example the number of non-zeros degree nodes, Fig.3(D), and other characteristics, Fig.15 in Supporting Information). We assume that the signature of this network (the nodes selected into it) is changing with age faster in healthy population than in DS patients.

Gene Ontology and KEGG Results

In this section we identify genes responsible for age related trajectories of methylation network signatures and make an enrichment analysis in Gene Ontology (GO) terms. Note, that we make a separate analysis of the lists of CpGs selected by the parenclitical approach and the lists of CpGs identified as good individual (1D) classifiers. The results are summarized in Table 2 (for parenclitic approach) and Table 3 (for 1D approach) in Supporting Information. While the DS-Nodes comparison did not return any significantly enriched GO, 23, 37 and 77 GO terms were enriched in the lists of CpG probes resulting from the S-Nodes comparison, M-Nodes comparison and AGE-Nodes comparison, respectively. A substantial overlap was observed between these 3 cases, with 19 terms in common. To gain more insights on the shared and specific gene ontology, we applied REVIGO to the lists of GO terms to remove redundant terms and focused on the "Biological Processes" description (Table 2, Supporting Information). We found 5 GO terms (pattern specification process, regionalization, cell fate commitment, skeletal system development, appendage development) common to all the comparisons, and another 5 GO terms (embryonic organ morphogenesis, epithelial cell differentiation, forebrain development, camera-type eye development, neuron migration) shared between the S-Nodes comparison and the M-Nodes comparison. Notably, the same GO analysis on the results of 1D comparison (Table 3, Supporting Information) returned a substantially lower number of enriched GO in the S-Nodes comparison and in the M-Nodes comparison, and no overlap between these two comparisons and the AGE-Nodes comparison.

The earlier study based on the local probe differential DNA methylation analysis employing the same data set [20] identifiedsome specific methylation changes in individuals with Down syndrome, as compared to their healthy siblings and mothers. Those differences mainly concerned the genes responsible for morphogenesis and intracellular processes. Notably, the use of parenclitic analysis made it possible to identify more complex patterns of epigenetic changes occurring in Down syndrome. Analysisin GO revealed significant differences in the methylation of loci associated with the functioning of the nervous system anda large number of processes (8 out of 36) related to the development of the genitourinary system and the musculoskeletal system. Biological processes common to all groups are associated with cell differentiation (GO:0007389 pattern specification process, GO:0003002 regionalization, GO:0045165 cell fate commitment) and morphogenesis (GO:0001501 skeletal systemdevelopment, GO:0048736 appendage development).

Identified signatures show the molecular mechanisms of changes associated with the age development of the Down Syndrome. Analysis in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (see (Table 1, Supporting Information)) shows a large number of signaling pathways associated with i) the nervous system (Neuroactive ligand-receptor interaction, Circadian entrainment, Axon guidance, Glutamatergic synapse, Calcium signaling pathway, GABAergic synapse, Morphine addiction, Nicotine addiction, Retrograde endocannabinoid signaling, Oxytocin signaling pathway, Cocaine addiction, Dopaminergic synapse, Glycosaminoglycan biosynthesis - heparan sulfate / heparin, Cholinergic synapse); ii) Cell fate and oncogenesis (Wnt signaling pathway, Signaling pathways regulating pluripotency of stem cells, Pathways in cancer, cAMP signaling pathway, Basal cell carcinoma, Proteoglycans in cancer); iii) Cellular communication (Rap1 signaling pathway, Calcium signaling pathway, Gap junction, PI3K-Akt signaling pathway, Glycosaminoglycan biosynthesis -heparan sulfate / heparin); iv) Development of metabolic syndrome (Adrenergic signaling in cardiomyocytes, Type II diabetes mellitus, Dilated cardiomyopathy, Insulin secretion, Maturity onset diabetes of the young, Arrhythmogenic right ventricular cardiomyopathy (ARVC)); and v) Female sex hormones (Oxytocin signaling pathway, Estrogen signaling pathway). These groups of genes were not identified via local differential DNA methylation analysis [20].

Discussion

In this study, we have shown how parenclitic networks can be used to find ageing trajectories of methylation network markers in the Down Syndrome. Three cohorts of the analysed data set, Down Syndrome individuals, their Mothers and Siblings, allowed for constructing different Parenclitic Networks, depending on the choice of Control, Case and Test groups. The pipeline starts with identifying CpG sites, which are good individual classifiers, and pairs of CpGs sites, which are good two-dimensional classifiers, but fail when taken alone. The latter set is used to build a parenclitic network, with edges corresponding to significant differential paired CpG methylation. By construction, the Case networks are much richer in nonzero degree nodes and edges than Control networks. Test networks do not follow such preset, and one investigates, which other groups they are closer to by network properties, and how does it change with age. On top of that, we build "general" networks that include only nodes and edges common to all individuals from a group, thus defining an epigenetic network signature and an entry set for further GO analysis.

DS has many well-described clinical manifestations, including mental retardation, congenital heart defects, gastrointestinal and hematological abnormalities, and other specific signs that ultimately form complex unique phenotypes. Here, we decomposed the epigenetic signature of DS patients into categories based on the hidden links between the covariates: the signature of the pure DS disease (regardless of age); the signature of age-related changes in a healthy population (which defines a plurality of DS patients as being close to an older population); the signature, which changes with age in a healthy population more slowly than in DS patients; and a signature, which changes in a healthy population with age faster than in DS patients. This could help, in the future, to identify molecular targets for medical treatment of DS accompanying diseases.

Our study identified network molecular changes associated with the DS development. The molecular changes emerging in the DS support the fact that genes do not act as autonomous units of the genome, but they are a part of spatially coordinated regulatory networks; accordingly, epigenetic remodeling does not interest only chromosome 21, but the entire genome. We showed that patientswith DS have different epigenetic changes in CpG sites of genes related to the clinical manifestations of DS. These are primarily biological processes responsible for the functioning of the central nervous system, skeletal muscles, disorders in carbohydrate metabolism, cardiopathology, and oncogenes. Instructively, most of the identified epigenetic changes in DS involve the same processes as biological pathways undergoing normal age-dependent epigenetic changes. Unique biological processes not associated with age are related to the brain development and organogenesis. DS is characterized by marked changes in brain formation during embryogenesis and in the early postnatal period (reduced cell proliferation, disorganized patterns of cortical stratification, pronounced neurodegeneration after birth and reduced number of dendritic spines). Most likely, the same molecular mechanisms continue to play a key role in the formation of the phenotype of the central nervous system in adulthood (ex. diminished brain size, smoothed gyrus).

We believe that the networks built in this work are not the end of the study, but only the beginning, as the new information obtained in the course of network analysis can be now widely analysed by biologists and clinicians in order to identify molecular mechanisms resulting in and accompanying DS. Additionally, this network analysis may help to identify new molecular targets for treatment of patients with DS to prevent their accelerated ageing. The networks built in this study require a further detailed analysis and can help researchers involved in DS research to discover new interpretations based on the interactions detected. We particularly highlight here that the main idea of Parenclitic Network approach, the separation of Case-Control group states, can be also applied for analyzing the transition in time, by age or other continuous scale, from one state to another. Introducing the Test set, which is located between the Case and Control states, can become an additional methodological improvement in the parenclitic study of complex systems.

We presented our algorithm as an open-source implementation of Generalized Parenclitic Network analysis to make it more accessible to all researchers. One of the main methodological advances is introducing new machine learning methods and kernels and discussing the possibility of their selection based on the goals of the problem. We believe that a simple integration into the overall implementation will allow researchers to use not only the proposed methods, but also join forces and test their own ideas.

This design can be applied to any disease, including, e.g., cancer, in which, in addition to critical states, such as healthy or diagnosed patient, there are data on intermediate conditions. These conditions can be, for example, the diagnosed patients analyzed at earlier time, when they could be considered as healthy people. We believe that the results of such an analysis based on Generalized Parenclitic Networks will not only help in the early diagnosis of the disease, e.g., by identifying critical transition marks and risk assessment, but also would shed light on the process itself, through attributes involved in it.

References

[1] Greenberg MVC., Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease.*Nat. Rev. Mol. Cell Biol*. 2019; 20, 590–607. https://doi.org/10.1038/s41580-019-0159-6.

[2] Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth G.P., Gunderson KL, Fan J-B,

Shen R. High density DNA methylation array with single cpg site resolution. *Genomics*. 2011; 98 (4), 288 – 295, DOI: https://doi.org/10.1016/j.ygeno.2011.07.007.

[3] CeCerrato F, Sparago A, Ariani F, Brugnoletti F, Calzari L, Coppedè F, De Luca A, Gervasini C, Giardina E, Gurrieri F, Lo Nigro C, Merla G, Miozzo M, Russo S, Sangiorgi E, Sirchia SM, Squeo GM, Tabano S, Tabolacci E, Torrente I, Genuardi M, Neri G, Riccio A. DNA Methylation in the Diagnosis of Monogenic Diseases. *Genes*. 2020; 11(4):355. https://doi.org/10.3390/genes11040355.

[4] Robertson KD. DNA methylation and human disease. *Nat. Rev. Genet*. 2005; 6, 597–610. https://doi.org/10.1038/nrg1655.

[5] Li D, Xie Z, Le Pape M, Dye T. An evaluation of statistical methods for DNA methylation microarray data analysis. *BMC Bioinforma*. 2015; 16, 217. https://doi.org/10.1186/s12859-015-0641-x.

[6] Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data. F1000Res. 2016 Jun 8; 5:1281. 10.12688/f1000research.8839.3. PMID: 27347385; PMCID: PMC4916993.

[7] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature.1999; 402, 47–52. https://doi.org/10.1038/35011540.

[8] Cui Z-J, Zhou X-H, Zhang H-Y. DNA Methylation Module Network-Based Prognosis and Molecular Typing of Cancer. *Genes*. 2019; 10(8):571.https://doi.org/10.3390/genes10080571.

[9] Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks, MPM, Eijk K, Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol*. 2012; 13, R97. https://doi.org/10.1186/gb-2012-13-10-r97 .

[10] Lund JB, Li S, Baumbach J, Christensen K, Li W, Mohammadnejad A, Pattie A, Marioni RE, Deary IJ, Tan. Weighted gene co-regulation network analysis of promoter DNA methylation on all-cause mortality in old-aged birth cohorts finds modules of high-risk associated biomarkers. *J Gerontol A Biol Sci Med Sci*. 2020; 75(12), 2249-2257. https://doi.org/10.1093/gerona/glaa066.

[11] Yuan L, Huang, D-S. A network-guided association mapping approach from DNA methylation to disease. *Sci Rep*.2019; 9, 5601. https://doi.org/10.1038/s41598-019-42010-6.

[12] Zanin M, Papo D, Sousa P, Menasalvas E, Nicchi A, Kubik E, Boccaletti S. Combining complex networks and data mining: Why and how. Phys. Reports. 2016; 635, 1–44. https://doi.org/10.1016/j.physrep.2016.04.005

[13] Zanin M, Boccaletti S. Complex networks analysis of obstructive nephropathy data.*Chaos: An Interdiscip. J. Nonlinear Sci*. 2011; 21, 033103 https://doi.org/10.1063/1.3608126.

[14] Zanin M., Menasalvas E, Sousa P, Boccaletti S. Preprocessing and analyzing genetic data with complex networks: An application to obstructive nephropathy. *Networks & Heterog. Media*. 2012; 7, 473. http://dx.doi.org/10.3934/nhm.2012.7.473

[15] Zanin, M., Menasalvas, E., Boccaletti S, Sousa P. Feature selection in the reconstruction of complex network representations of spectral data. *PloS ONE. 2013;* 8(8): e72045. https://doi.org/10.1371/journal.pone.0072045

[16] Zanin M, Papo D, Solís JL, Espinosa JC, Frausto-Reyes C, Anda PP, Sevilla-Escoboza R, Jaimes-Reategui R, Boccaletti S, Menasalvas E, Sousa P. Knowledge discovery in spectral data by means of complex networks. *Metabolites*. 2013 Mar 11;3(1):155-67. https://doi.org/10.3390/metabo3010155. PMID: 24957895; PMCID: PMC3901251.

[17] Papo D, Buldú JM, Boccaletti S, Bullmore ET. Complex network theory and the brain. *Phil. Trans:R.Soc.B*. 2014; 369, 1–7. https://doi.org/10.1098/rstb.2013.0520

[18] Karsakov A, Bartlett T, Ryblov A, Meyerov I, Ivanchenko M, Zaikin A. Parenclitic network analysis of methylation data for cancer identification. *PloS ONE*. 2017; 12, e0169661https://doi.org/10.1371/journal.pone.0169661

[19] Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. *Oncotarget*. 2018; 9, 22717–22726. https://doi.org/10.18632/oncotarget.25216 .

[20] Bacalini MG, Gentilini D, Boattini A, Giampieri E, Pirazzini C, Giuliani C, Fontanesi E, Scurti M, Remondini D, Capri M, Cocchi G, Ghezzo A, Del Rio A, Luiselli D, Vitale G, Mari D, Castellani G, Fraga M, Di Blasio AM, Salvioli S, Franceschi C, Garagnani P. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. Aging (Albany NY). 2015 Feb;7(2):82-96. https://doi.org/10.18632/aging.100715. PMID: 25701644; PMCID: PMC4359691.

[21] Do C, Xing Z, Yu YE, Tycko B. Trans-acting epigenetic effects of chromosomal aneuploidies: lessons from down syndrome and mouse models. *Epigenomics*. 2017; 9, 189–207. https://doi.org/10.2217/epi-2016-0138.

[22] Henneman P, Bouman A, Mul A, Knegt L, van der Kevie-Kersemaekers A-M, Zwaveling-Soonawala N, Meijers-

Heijboer HEJ, Trotsenburg ASP, Manners MM. Widespread domain-like perturbations of DNA methylation in whole blood of Down syndrome neonates. *PLoS ONE*. 2018; 13(3): e0194938. https://doi.org/10.1371/journal.pone.0194938

[23] Laufer BI, Hwang H, Vogel Ciernia A, Mordaunt CE, LaSalle JM. Whole genome bisulfite sequencing of down syndrome brain reveals regional DNA hypermethylation and novel disorder insights. *Epigenetics*. 2019, 14, 672–684. https://doi.org/10.1080/15592294.2019.1609867.

[24] Franceschi C, Garagnani P, Gensous N, Bacalini MG, Conte M, Salvioli S. Accelerated bio-cognitive aging in Down syndrome: State of the art and possible deceleration strategies. *Aging Cell*. 2019 Jun;18(3):e12903..https://doi.org/10.1111/acel.12903 Epub 2019 Feb 15. PMID: 30768754; PMCID: PMC6516152.

[25] Martin GM. Genetic syndromes in man with potential relevance to the pathobiology of aging. Birth Defects *Orig Artic Ser.* 1978;14(1):5-39. PMID: 147113.

[26] Horvath S, Garagnani P, Bacalini MG, Pirazzini C, Salvioli S, Gentilini D, Di Blasio AM, Giuliani C, Tung S, Vinters HV, Franceschi C. Accelerated epigenetic aging in Down syndrome. *Aging Cell*. 2015 Jun;14(3):491-5. https://doi.org/10.1111/acel.12325. Epub 2015 Feb 9. PMID: 25678027; PMCID: PMC4406678.

[27] Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014 May 15;30(10):1363-9. https://doi.org/10.1093/bioinformatics/btu049. Epub 2014 Jan 28. PMID: 24478339; PMCID: PMC4016708.

[28] Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017 Feb 28;45(4):e22. https://doi.org/10.1093/nar/gkw967. PMID: 27924034; PMCID: PMC5389466.

[29] Singmann P, Shem-Tov D, Wahl S, Grallert H, Fiorito G, Shin S-Y, Schramm K, Wolf P, Kunze S, Baran Y, Guarrera S, Vineis P, Krogh V, Panico S, Tumino R, Kretschmer A, Gieger C, Peters A, Prokisch H, Relton CL, Matullo G, Illig T, Waldenberger M, Halperin E. Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenetics & Chromatin*. 2015; 8, 43. https://doi.org/10.1186/s13072-015-0035-3

[30] Ren X, Kuan PF. methylGSA: a Bioconductor package and shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics*. 2019; 35, 1958–1959. https://doi.org/10.1093/bioinformatics/bty892

Acknowledgements

Authors acknowledge support by the grant of the Ministry of Education and Science of the Russian Federation Agreement No. 074-02-2018-330. AZ, TN acknowledge support by the MRC grant MR/R02524X/1.

Author contributions statement

AZ, CF, MI designed the study, MK developed software, TN developed methodology of age-related analysis, MK and TN conducted the graph analysis, MK, TN, MGB and MV conducted the data analysis. All authors have written and reviewed the manuscript.

Additional information

Authors declare no competing interests.

Figures

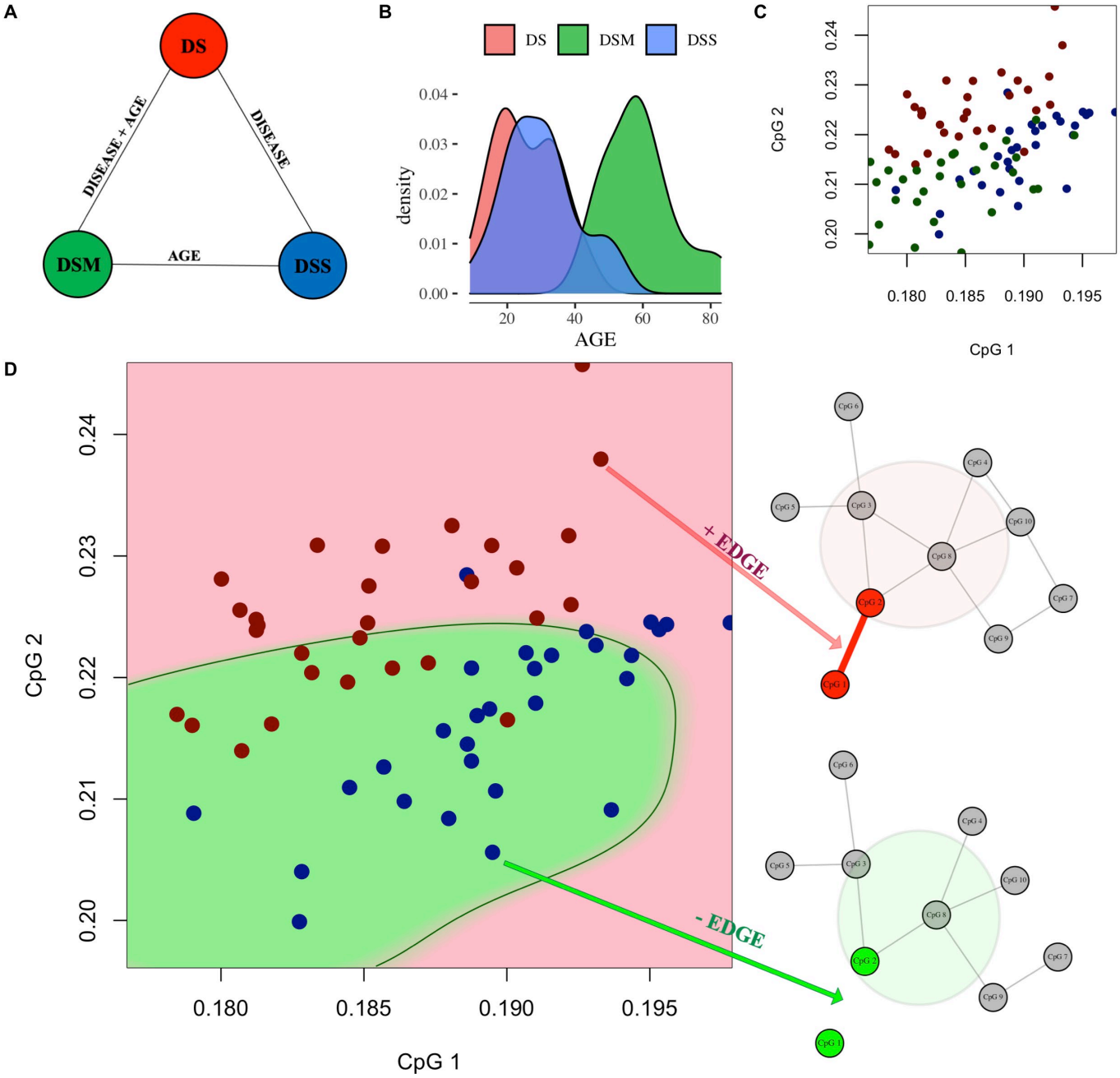

Figure 1. Features of the initial data and the scheme of networks construction **(A)** Schematic representation of the three groups in the GSE52588 dataset and their phenotypic difference; **(B)** Age-distribution for DSM/DSS/DS groups; **(C)** Example of the three groups representation on the plane of a pair of CpGs loci; **(D)** Illustration of the parenclitic network construction: defining the PDF area for Controls (here, DSM), assigning the edgebetween two CpGs loci if the point is“outside” the Controls-area (red) and no edge if the point is “inside” the Controls-area(green).

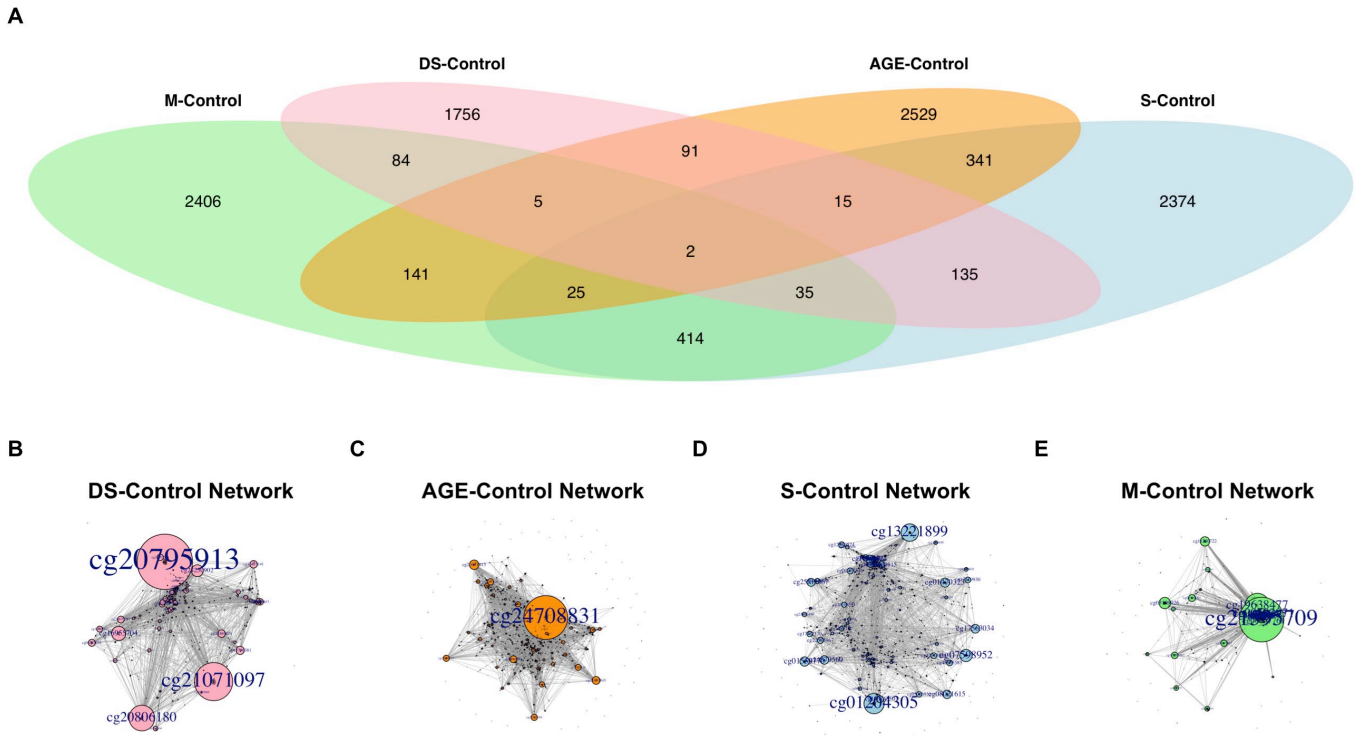

Figure 2. Commonality and individuality of the networks of the four considered models **(A)** Venn Diagram for the vertices (CpG sites) of generalizing networks; size of nodes in each image **(B)** (DS-Control generalizing Network), **(C)** (AGE-Control generalizing Network), **(D)** (S-Control generalizing Network) and **(E)** (M-Control generalizing Network)) are associated with their degree (the greater the degree of the vertex, the more connections it has). Enlarged presentations of networks can be found on the links on https://tatiananazarenko.github.io/PN–DS/.

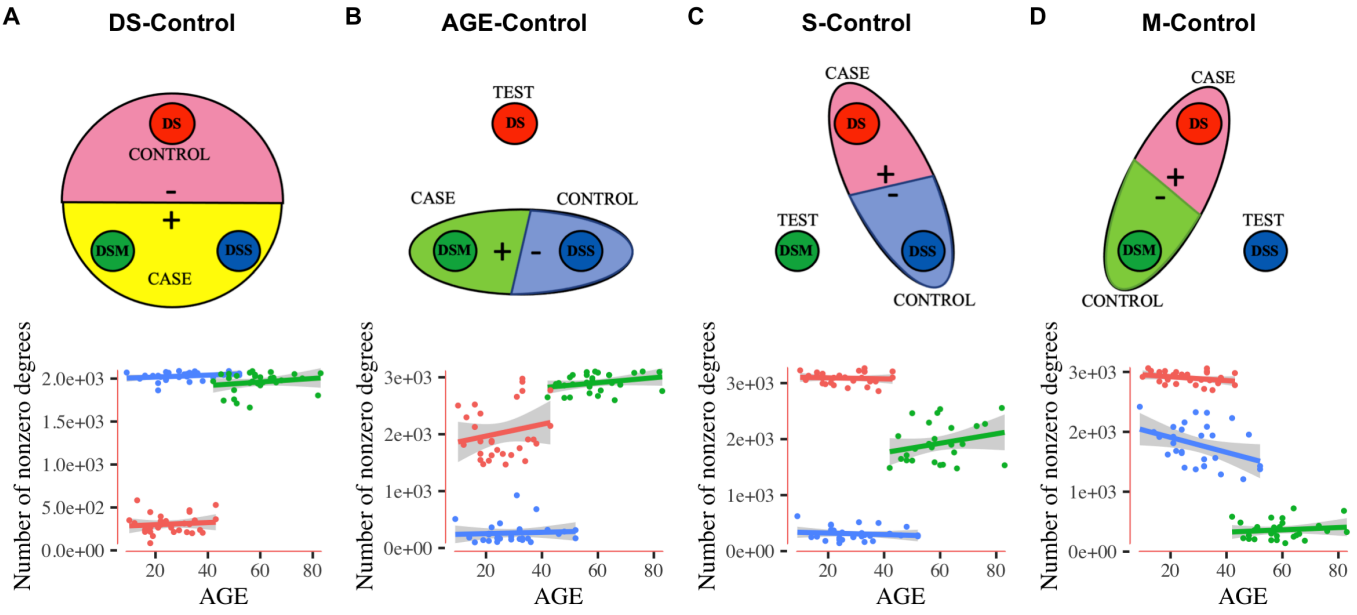

Figure 3. Age dependence of network characteristics Top panel: Assigning the cohorts for network construction. Bottom panel: Considering the Test group against theControls and Cases reveals age-related trajectories, quantified by network signatures (B-D). Here we exemplify it in the numberof nonzero degrees nodes (the number of CpG probes, that show differential network methylation) in individual networks forDS/AGE/S/M-network design versus AGE (A–D labels respectively). Commonly, the Case group networks are characterizedby a considerable number of co-occurring differential CpG methylation changes, as compared to the Control group (as followsfrom the network construction algorithm). Specifically, for different Control, Case and Test groups: B) Parenclitic network model of aging (AGE-Control): DS samples are closer to DSM than to their DSS and get even closer with age; C) Parencliticnetwork model of the syndrome: DSM samples get closer with age to DS; D) Parenclitic network model of aging and syndrome: DSS samples get closer with age to DSM. Cf. Supporting Information for the plots for the other network indexes,that confirm the described effects.

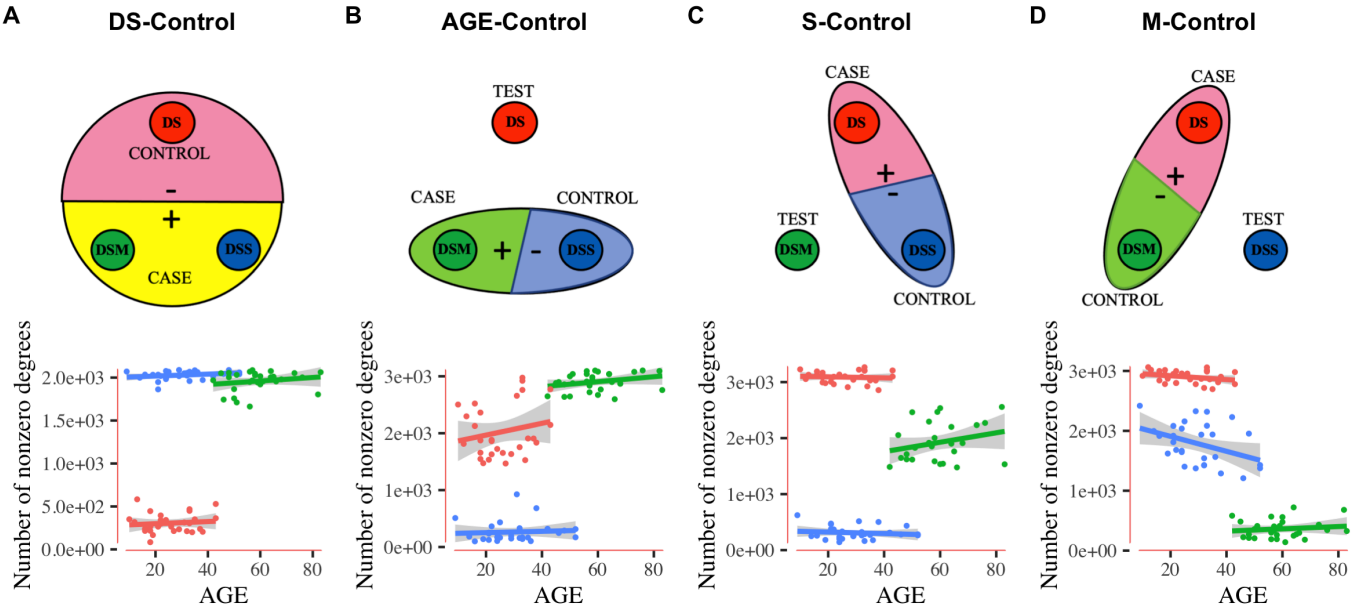

Figure 4. Age-acceleration dependence of network characteristics. The number of nonzero degree network nodes (the number of CpG probes, that show differential network methylation) vs age acceleration (horizontal axis). In all cases, the DS methylation network signature is closer to mothers. For the other network characteristics cf. Supporting Information.