doi 10.18699/bgrs2024-13-25

# Research of non-monotonic aging changes based on multiomics data

Gorbarenko A.V.[1], Bobkov G.A.[1, 2], Alekseev A.A.[3, 4, 5]*

[1] *Bioinformatics Institute, St. Petersburg, Russia*

[2] *ITMO, St. Petersburg, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

[4] *Kulakov National Medicine Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia*

[5] *Russian Research Institute of Health, Moscow, Russia*

* *alekseev@physics.msu.ru*

**Key words:** aging; aging-related trends; omics data; aging transitions; trends clustering

*Motivation and Aim*: At the age around 37, various molecular alterations occur in human organisms, including dysregulated gene expression, altered metabolite levels, somatic mutations, epimutations, and accumulated molecular damage [7-10]. Age transition in 23±2 years old (y.o.) is also very important for the system biology of aging: first increasing of MDA (malondialdehyde, marker of oxidative stress) occurs at this age [12]. Significant stages in the aging process also occur around ages 24 and 60, where molecular and physiological changes become most pronounced [11-13]. Wide known "hallmarks of aging" involve genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulation of nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication. These mechanisms contribute to aging at molecular, cellular, and systemic levels [6, 14, 15], but interaction between cellular and systemic aging is still unclear. This study aims to obtain the most relevant cellular aspect, associated with metabolic transition of aging around 37 through multiomics analysis, including epigenomics and RNA-sequencing data.

*Methods and Algorithms*: Data was collected from the National Center for Biotechnology Information Gene Expression Omnibus database, accessed via the geoparse library [1]. The complete datasets table can be found on GitHub. All CpG data were selected based on the platform, specifically only using the Illumina 450k array. Due to the large size of the methylation data, only CpGs that significantly differ between the age groups '18-27', '27-30', '30-35', '35-40', '40-45', or '45-60' years were chosen for further investigation. Batch effects were detected using PCA visualization and corrected with pyCombat [2]. The most representative dataset GSE42861was chosen to test methods for calculating central tendency (mean, median, density distribution peak) and clustering algorithms (k-means, hierarchical, DBSCAN, Gaussian Mixture). Correlations of CpG trends were calculated using the Pearson correlation coefficient. The chosen algorithm was applied to all methylation and microarray groups of datasets. Additionally, separate analyses were performed for males and females. The clustering results were visualized using a sliding window to plot samples from each cluster. CpG annotation was performed using Illumina's annotation data [3]. Data from epigenetics and transcriptomics were intersected. Enrichment analysis was conducted using Enrichr (Reactome database) and StringDB tools to identify pathways with the most significant changes [4, 5].

*Results*: Based on the GSE42861 dataset, both the median and density distribution peak methods yielded similar results, but the density distribution peak showed better stability within the methylation group of datasets. Hierarchical clustering and GM produced comparable results, with GM demonstrating better stability. Therefore, the density distribution peak was chosen for central tendency, and GM was selected as the clustering algorithm. The GM covariance matrix was calculated as diagonal due to computational constraints. We chose eight clusters to capture additional points of age-related changes (Fig. 2), exceeding the optimal five. The clusters with changes, obtained as described above, identified 3354 CpGs and 290 RNAs around the age of 24. As a result, we identified 26 genes. The full list of genes is available on GitHub. Using StringDB (Fig. 1), we observed that these genes are responsible for changes in cellular regulation, autophagy, cell cycles, and more. Enrichr analysis showed that the top two enrichments were in FLT3 signaling and cellular senescence. For the male datasets, the shape and age distribution were insufficient to conduct a robust analysis of the methylation data. A separate analysis of the transcriptomic data revealed transitions at ages 37-40 in both males and females, specifically involving the genes PERP, CDKN1A, THBS1, TFPI, and SERPINB.


**Fig. 1.** Genes clustering using StringDB

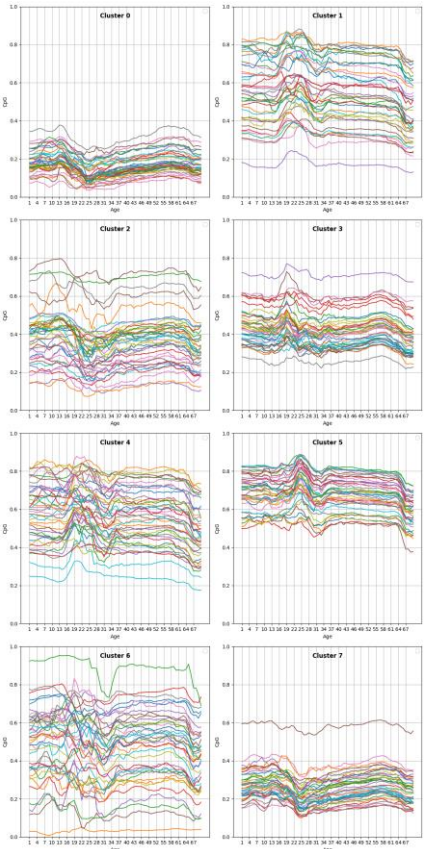


**Fig. 2.** Methylation trends clustering

*Conclusion*: Our study aimed to uncover key molecular pathways and biological processes that undergo significant changes at around 24 and 37 y.o. However, it is important to note that we were unable to identify significant changes at 37 y.o. based on epigenetic analysis alone. This suggests that while transcriptomic data shows clear transitions, epigenetic markers might not be as sensitive or may require different analytical approaches to detect changes at 37 y.o. Identified 26 key genes play distinct roles in cellular processes. CDKN1B regulates cell cycle progression and contributes to cellular senescence. FOXN3 acts as a transcriptional repressor, influencing DNA damage pathways. BCL2L11 induces apoptosis and is crucial for removing damaged cells. HDAC5 regulates gene expression, impacting tissue function and development. PHC3 maintains transcriptional repression, influencing differentiation and senescence. TAF1B is involved in transcription initiation, affecting aging-related processes. Other genes contribute to functions like pre-mRNA splicing and mitochondrial activity, collectively

influencing aging through diverse mechanisms. For future investigations, collecting more datasets, including lipidomics and metabolomics data, is essential for a comprehensive understanding of aging's molecular changes. In the next steps of our research, reviewing our analytical pipeline in transcriptomic data and improving correlation calculations for all data could also enhance our analysis. But even at this stage our study contributes to understanding aging by identifying key genes and pathways undergoing significant changes around 24 y.o.

*Data availability:* A repository was created on GitHub, containing a detailed progress of work and the results obtained. The repository can be found at: https://github.com/CaptnClementine/SODA.

*References*


1. Gene Expression Omnibus (GEO) database, National Center for Biotechnology Information (NCBI). Available at: https://www.ncbi.nlm.nih.gov/geo/
2. Johnson W.E., Li C., Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127
3. Illumina, Inc. (n.d.). Infinium HumanMethylation450 BeadChip Annotation. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanmethylation450.pdf
4. Kuleshov M.V., Jones M.R., Rouillard A.D. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90-W97
5. Szklarczyk D., Gable A.L., Lyon D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-D613
6. Mc Auley M.T., Guimera A.M., Hodgson D. et al. Modelling the molecular mechanisms of aging. *Biosci Rep*. 2017;37(1):BSR20160177
7. Aramillo Irizar P., Schäuble S., Esser D. et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun.* 2018;9(1):327
8. Kabacik S., Lowe D., Fransen L. et al. The relationship between epigenetic age and the hallmarks of aging in human cells. *Nat Aging.* 2022;2(6):484-493
9. Horvath S., Gurven M., Levine M.E. et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* 2016;17:1-23
10. González-Velasco O., Papy-García D., Le Douaron G. et al. Transcriptomic landscape, gene signatures and regulatory profile of aging in the human brain. *Biochim Biophys Acta.* 2020;1863(6):194491
11. Zavorsky G.S., Tomko K.A., Smoliga J.M. Declines in marathon performance: Sex differences in elite and recreational athletes. *PloS One.* 2017;12(2):e0172121
12. Ravera S., Podestà M., Sabatini F. et al. Discrete changes in glucose metabolism define aging. *Sci Rep.* 2019;9(1):10347
13. Lu A.T., Binder A.M., Zhang J. et al. DNA methylation GrimAge version 2. *Aging (Albany NY).* 2022;14(23):9484
14. Yaneske E., Angione C. The poly-omics of ageing through individual-based metabolic modelling. *BMC bioinformatics*. 2018;19:83-96
15. Lehallier B., Gate D., Schaum N. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat med.* 2019;25(12):1843-1850