 doi:10.20944/preprints202209.0172.v1

NOMA-DB: A Framework for Management and Analysis of Ageing-related Gene-Expression Data

Pietro Hiram Guzzi*
Department of Surgical and Medical Sciences
University of Catanzaro
Catanzaro, Italy
hguzzi@unicz.it

Rocco Scicchitano
Department of Surgical and Medical Sciences
University of Catanzaro
Catanzaro, Italy

Pierangelo Veltri
Department of Surgical and Medical Sciences
University of Catanzaro
Catanzaro, Italy
hguzzi@unicz.it

*Corresponding author: hguzzi@unicz.it

ABSTRACT

Recently, many independent research project focus on the study of the molecular basis of aging processes. I parallel, the different progression of many diseaes between sex is a hot topic research area. These studies require many data, models and tools for inferring aging and sex specific molecular machinaries. Among the others, the Genotype-Tissue Expression (GTEx) database is one of the preminent resources for the analysis of expression data related to tissues, sex and age. The current version of the database has a lot of querying interface that enable many analysis centered on the expression of genes on tissues. Despite this the database lacks on the analysis at sex/age level, thus the researcher has to download data and then write queries by hand. Nevertheless it lacks on the integration with existing protein interaction data. Therefore, the need for the introduction of tools enabling easy access and powerful analysis capabilities (i.e. state of the art network based analysis and integration), arises. We here present NOMA-DB a framework for the analysis of age related genes based on the GTEx database that enable easy querying at sex/age level, network based analysis. The framework is based on wrapping the GTEx database and on building an application logic level on top of existing data. The current version enables the analysis of genes by tissue, gene and age, thus it may be used in potentially future directions of analysis towards better comprehension of aging/sex-related molecular machineries based on the analysis of expression data.
Keywords: ageing GTEx analysis

1 Introduction

Ageing is commonly defined as the process of becoming older. In humans, ageing represents the progressive insurgence and accumulation of changes over time [1, 2]. Ageing studies focus on disclosing genomic and epigenomics changes related to these changes [3, 4, 5, 6]. The rationale of these studies is that the elucidation of these mechanisms can also help develop novel therapies for many diseases that seem related to ageing [7, 8, 9, 10, 11, 12].

The accumulation of results of different studies has led to the consideration that ageing may be explained as a set of *ageing clocks*, i.e. a set of molecules whose changes may be predictors of the age [13, 14]. Consequently, from a theoretical point of view, it could be possible to measure a *biological age* for each individual who may differ from the *anagraphical age*, providing more information about the health status and the risk of insurgence and development of diseases [15].

To shed light on ageing processes, the researcher needs to integrate and mine heterogeneous data produced by experiments of different laboratories (e.g. omics, epigenomics, and medical images) [16, 17, 18].

Thus, it is possible to evidence some main key points enabling research in this area: (i) availability of data in public databases; (ii) data standards for easy integration and comparison; (iii) methods and models for data integration and analysis, also leveraging capabilities from deep learning and artificial intelligence [19, 20, 21, 22]. Among the others, the GTEx data portal [23] represents one of the most used resources for collecting data related to whole-genome sequencing and RNA-seq in individuals. For each sample, GTEx provides some information such as tissue of provenance and sex and age (grouped into six classes) of the patient. The current version of the GTEx database (v8 accessed on August 22th) stores 17382 samples of 54 tissues of 948 donors, see at `https://gtexportal.org/home/tissueSummaryPage`. GTEx is available through a web interface that offers easy to use query interface and visualisation of data in tissues, and it is used in many ageing-related studies [24, 25, 26, 27]. Despite this, to our knowledge, the GTEx data portal has some limitations, such as (i) users cannot query data grouped by age or sex, and (ii) data are not integrated with existing protein interaction databases.

In particular, the user who wants analyse data at sex /age level has to download the database has to write his own script, so this is a significant limitation for the inexperienced user, and in general, it requires time to download and prepare data. The second limitation is related to the possibility of reconstructing and studying ageing processes at a network level, which is promising, as demonstrated in some recent works [28, 29, 30, 31, 32, 31, 33]. Consequently, we designed NOMA-DB (Network Omics Ageing Database), a framework able to gather data from the GTEx database and existing protein interaction databases. Such data are then made available to the user through a web interface. The user can query the GTEx database and analyse data by sex and age through the web interface. Moreover, expression data are then integrated into the existing protein interaction network to enable network-based analysis and visualisation. Thus, the NOMA-DB framework extends the existing resources, and it enables tissue-level ageing studies.

We here describe the design and development of NOMA-DB. We also present the first implementation and some case studies demonstrating the approach's effectiveness.

2 Related Work

The recent advances in high-throughput omics technologies have accelerated the development of databases, methods and tools for analysing biological data with the ultimate goal of measuring biological age at the molecular level. We here present state-of-the-art databases that constitute the main data sources for machine learning algorithms to identify new biomarkers of biological ageing [34].

Ageing-Related Databases State-of-the-art databases for ageing studies share a common principle: storing multi-omics datasets that gather information from almost all the omics aspects, i.e. genomics, epigenomics, transcriptomics, proteomics, metabolomics and pharmacogenomics. All the databases are available through a web interface offering an easy query interface. Moreover, they provide a limited analysis interface.

The AgingAtlas [35] database, available at `https://ngdc.cncb.ac.cn/aging/index` offer data coming from five different experimental platforms: RNA-seq, single-cell transcriptomics, epigenomics, proteomics , and pharmacogenomics. The web interface provides the user to analyse changes in expression profiles at age level. The database is also available for download. Despite the presence of a protein interaction module, it provides only a search of interactions related to a gene without the possibility of analysis of the networks. Finally, AgingAtlas does not contain tissue-level data.

GenAge [28] `https://genomics.senescence.info/genes/index.html` is a curated database of genes related to ageing in humans. The database, available through a web interface at `https://genomics.senescence.info/genes/index.html` offer the possibility to search and analyse genes and related studies. It allows users to analyse the genetic network of a gene or associated pathways. GenAge is a reference database for ageing-related studies, but it does not offer the possibility of discovering other age-related genes or expression profiles. GenAge is part of Human Ageing Genomic Resources (HAGAR), which collects databases and tools for studying ageing [36]. Similarly to AgingAtlas it does not contain tissue level data.

Aging Studies In parallel to the accumulation of data into public databases many efforts have been done to identify genes, proteins and molecules related to the ageing process [36]. Since wet lab experiments are difficult and expensive, researchers leveraged computational methods to discover such genes. These experiments integrate gene expression and protein interaction data and analyse the expression of these molecules during the age in order to evidence changes that are correlated or caused by age progression. In particular the use of protein interaction data enable to evidence changes at system level scale, by highlighting network modifications [37, 38, 39].

3 A framework for extending GTEx database.

The NOMA-DB architecture comprises the following modules as depicted in Figure 1: (i) the GTEx Manager, (ii) the GTEx (local) DB, (iii) the PPI-DB integrator, (iv) the PPI-DB (local(database, (v) the network analysis module, (vi) the query builder, and (vii) the web interface.

The GTEx DB and the PPI-DB store, respectively, a local copy of the GTEx database and Protein interaction databases. The current version of the framework uses MySQL 8.0 database to store data.

The GTEx Manager and the PPI DBI Integrator are responsible for updating data crawled from the GTEx data portal and other web resources. Both use HTTP protocol to access data on the web. After connection, they download data from the local repository. Furthermore, since the system may access different protein interaction databases, it integrates them into a single repository avoiding redundancies. The current version of NOMA-DB uses python libraries for accessing and downloading GTEx data.

The Query Builder module is responsible for the interaction with the user interface. It receives as input the user formulated queries (e.g. genes expressed in tissue X at age Y) and it extracts data into the local databases. Moreover, it accesses PPI-DB to retrieve protein interactions related to the user query. The network analysis module implements network mining modules, i.e. differential expression of genes of the network, network parameters etc. [40]. The user access to the framework through a web interface based on a responsive HTML5-based client.

The rest of the Section shows the functionalities of NOMA-DB through a case study.

4 Conclusion

Recently the interest for investigating sex and age based differences on disease insurgence and progession is increasing. These studies need a deep knowledge of molecular mechanisms and their differences among indivuduals. These data

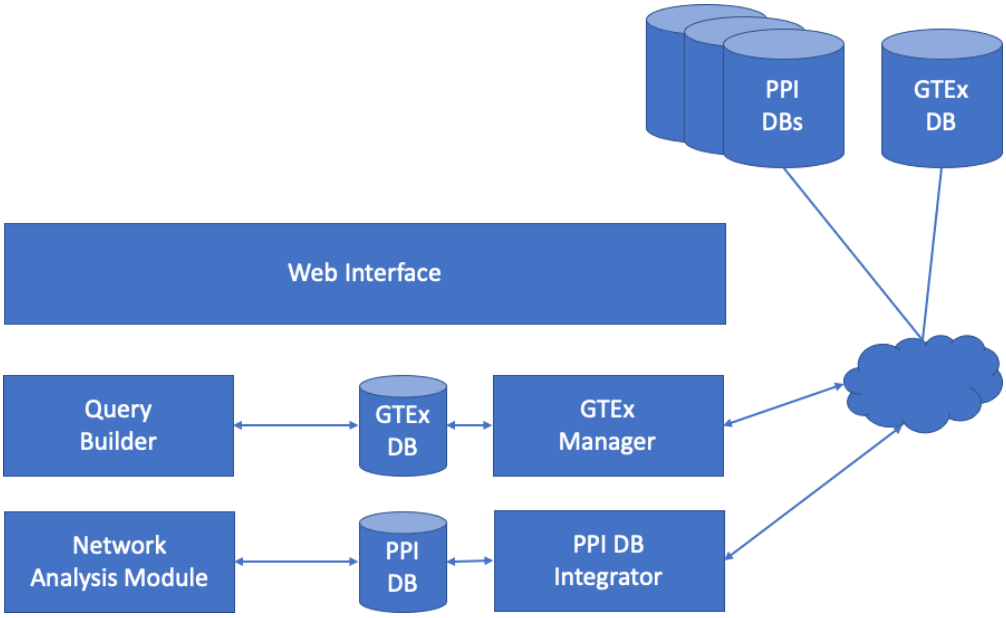

Figure 1: The NOMA-DB architecture comprises the following modules: (i) the GTEx Manager, (ii) the GTEx (local) DB, (iii) the PPI-DB integrator, (iv) the PPI-DB (local(database, (v) the network analysis module, (vi) the query builder, and (vii) the web interface. The GTEx DB and the PPI-DB store, respectively, a local copy of the GTEx database and Protein interaction databases. The GTEx Manager and the PPI DBI Integrator are responsible for the update of data crawled from the GTEx data portal and other web resources. The Query Builder is responsible for the accession to the local DB and extracting data, while the network analysis module implements network mining modules. The web interface realises the interface to the system.

should be contained in public available databases. There exists some efforts in this direction such as the GTEx database. Despite this, there is a need for some specialised query interfaces and for the integration with existing protein interaction data. We here presented NOMA-DB a framework for the analysis of age related genes based on the GTEx database that enable easy querying at sex/age level, network based analysis. Future work will regard from one side on the extesion of this database and from another side the use of them for a better comprehension of aging/sex-related molecular machineries based on the analysis of expression data.

Acknowledgements

PNG and PV were partially funded by the MISE PON-VQA project.

References

[1] Lauren N Booth and Anne Brunet. The aging epigenome. *Molecular cell*, 62(5):728–744, 2016.

[2] Adiv A Johnson, Maxim N Shokhirev, Tony Wyss-Coray, and Benoit Lehallier. Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing research reviews*, 60:101070, 2020.

[3] R Spinelli, L Parrillo, M Longo, P Florese, A Desiderio, F Zatterale, C Miele, G Alexander Raciti, and F Beguinot. Molecular basis of ageing in chronic metabolic diseases. *Journal of Endocrinological Investigation*, 43(10):1373–1389, 2020.

[4] Takayuki Teruya, Haruhisa Goga, and Mitsuhiro Yanagida. Human age-declined saliva metabolic markers determined by lc–ms. *Scientific reports*, 11(1):1–11, 2021.

[5] Richard A Kerber, Elizabeth O'Brien, and Richard M Cawthon. Gene expression profiles associated with aging and mortality in humans. *Aging Cell*, 8(3):239–250, 2009.

[6] Xiaojuan He, Sebastian Memczak, Jing Qu, Juan Carlos Izpisua Belmonte, and Guang-Hui Liu. Single-cell omics in ageing: a young and growing field. *Nature Metabolism*, 2(4):293–302, 2020.

[7] Linda Partridge. Intervening in ageing to prevent the diseases of ageing. *Trends in Endocrinology & Metabolism*, 25(11):555–557, 2014.

[8] Bennett G Childs, Martina Gluscevic, Darren J Baker, Remi-Martin Laberge, Dan Marquess, Jamie Dananberg, and Jan M Van Deursen. Senescent cells: an emerging target for diseases of ageing. *Nature reviews Drug discovery*, 16(10):718–735, 2017.

[9] Kenneth Rockwood and Susan E Howlett. Age-related deficit accumulation and the diseases of ageing. *Mechanisms of ageing and development*, 180:107–116, 2019.

[10] Giulia C Leonardi, Giulia Accardi, Roberto Monastero, Ferdinando Nicoletti, and Massimo Libra. Ageing: from inflammation to cancer. *Immunity & Ageing*, 15(1):1–7, 2018.

[11] João Pedro De Magalhães. How ageing processes influence cancer. *Nature Reviews Cancer*, 13(5):357–365, 2013.

[12] Shijia Bi, Zunpeng Liu, Zeming Wu, Zehua Wang, Xiaoqian Liu, Si Wang, Jie Ren, Yan Yao, Weiqi Zhang, Moshi Song, et al. Sirt7 antagonizes human stem cell aging as a heterochromatin stabilizer. *Protein & cell*, 11(7):483–504, 2020.

[13] David Melzer, Luke C Pilling, and Luigi Ferrucci. The genetics of human ageing. *Nature Reviews Genetics*, 21(2):88–101, 2020.

[14] João Pedro De Magalhães, João Curado, and George M Church. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*, 25(7):875–881, 2009.

[15] Mario Cannataro, Pietro Hiram Guzzi, and Alessia Sarica. Data mining and life sciences applications on the grid. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(3):216–238, 2013.

[16] Ana M Valdes, Daniel Glass, and Tim D Spector. Omics technologies and the study of human ageing. *Nature Reviews Genetics*, 14(9):601–607, 2013.

[17] Jonas Zierer, Cristina Menni, Gabi Kastenmüller, and Tim D Spector. Integration of 'omics' data in aging research: from biomarkers to systems biology. *Aging cell*, 14(6):933–944, 2015.

[18] Jared S Lorusso, Oleg A Sviderskiy, and Vyacheslav M Labunskyy. Emerging omics approaches in aging research. *Antioxidants & Redox Signaling*, 29(10):985–1002, 2018.

[19] Rolf Hühne, Torsten Thalheim, and Jürgen Sühnel. Agefactdb—the jenage ageing factor database—towards data integration in ageing research. *Nucleic acids research*, 42(D1):D892–D896, 2014.

[20] Fabio Fabris, João Pedro de Magalhães, and Alex A Freitas. A review of supervised machine learning applied to ageing research. *Biogerontology*, 18(2):171–188, 2017.

[21] Ruixue Ai, Xurui Jin, Bowen Tang, Guang Yang, Zhangming Niu, and Evandro F Fang. Ageing and alzheimer's disease: Application of artificial intelligence in mechanistic studies, diagnosis, and drug development. In *Artificial Intelligence in Medicine*, pages 1–16. Springer, 2021.

[22] Stavros-Theofanis Miloulis, Ioannis Kakkos, Athanasios Anastasiou, George K Matsopoulos, and Dimitris Koutsouris. Application of artificial intelligence towards successful ageing: A holistic approach. In *Modern Challenges and Approaches to Humanitarian Engineering*, pages 172–193. IGI Global, 2022.

[23] John Lonsdale, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, Gary Walters, Fernando Garcia, Nancy Young, et al. The genotype-tissue expression (gtex) project. *Nature genetics*, 45(6):580–585, 2013.

[24] Ansley Grimes Stanfill and Xueyuan Cao. Enhancing research through the use of the genotype-tissue expression (gtex) database. *Biological research for nursing*, 23(3):533–540, 2021.

[25] Daniele Mercatelli, Elisabetta Pedace, Pierangelo Veltri, Federico M Giorgi, and Pietro Hiram Guzzi. Exploiting the molecular basis of age and gender differences in outcomes of sars-cov-2 infections. *Computational and Structural Biotechnology Journal*, 19:4092–4100, 2021.

[26] Michael P Pressler, Anelia Horvath, and Emilia Entcheva. Sex-dependent transcription of cardiac electrophysiology and links to acetylation modifiers based on the gtex database. *Frontiers in Cardiovascular Medicine*, 9, 2022.

[27] Francesco Ortuso, Daniele Mercatelli, Pietro Hiram Guzzi, and Federico Manuel Giorgi. Structural genetics of circulating variants affecting the sars-cov-2 spike/human ace2 complex. *Journal of Biomolecular Structure and Dynamics*, pages 1–11, 2021.

[28] Joao Pedro de Magalhaes and Olivier Toussaint. Genage: a genomic and proteomic network map of human ageing. *FEBS letters*, 571(1-3):243–247, 2004.

[29] Shawn Gu, Meng Jiang, Pietro Hiram Guzzi, and Tijana Milenković. Modeling multi-scale data via a network of networks. *Bioinformatics*, 38(9):2544–2553, 2022.

[30] Qi Li, Khalique Newaz, and Tijana Milenković. Improved supervised prediction of aging-related genes via weighted dynamic network analysis. *BMC bioinformatics*, 22(1):1–26, 2021.

[31] Maria Eugenia Gallo Cantafio, Katia Grillone, Daniele Caracciolo, Francesca Scionti, Mariamena Arbitrio, Vito Barbieri, Licia Pensabene, Pietro Hiram Guzzi, and Maria Teresa Di Martino. From single level analysis to multi-omics integrative approaches: a powerful strategy towards the precision oncology. *High-throughput*, 7(4):33, 2018.

[32] Giuseppe Agapito, Pietro Hiram Guzzi, and Mario Cannataro. Parallel extraction of association rules from genomics data. *Applied Mathematics and Computation*, 350:434–446, 2019.

[33] Giovanni Nassa, Roberta Tarallo, Pietro H Guzzi, Lorenzo Ferraro, Francesca Cirillo, Maria Ravo, Ernesto Nola, Marc Baumann, Tuula A Nyman, Mario Cannataro, et al. Comparative analysis of nuclear estrogen receptor alpha and beta interactomes in breast cancer cells. *Molecular BioSystems*, 7(3):667–676, 2011.

[34] Jarod Rutledge, Hamilton Oh, and Tony Wyss-Coray. Measuring biological age using omics data. *Nature Reviews Genetics*, pages 1–13, 2022.

[35] Aging atlas: a multi-omics database for aging biology. *Nucleic Acids Research*, 49(D1):D825–D830, 2021.

[36] Robi Tacutu, Daniel Thornton, Emily Johnson, Arie Budovsky, Diogo Barardo, Thomas Craig, Eugene Diana, Gilad Lehmann, Dmitri Toren, Jingwei Wang, Vadim E Fraifeld, and JoÃ£o P deÂ MagalhÃ£es. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, 46(D1):D1083–D1090, 11 2017.

[37] Fazle E Faisal and Tijana Milenković. Dynamic networks reveal key players in aging. *Bioinformatics*, 30(12):1721–1729, 2014.

[38] Csaba Kerepesi, Bálint Daróczy, Ádám Sturm, Tibor Vellai, and András Benczúr. Prediction and characterization of human ageing-related proteins by using machine learning. *Scientific reports*, 8(1):1–13, 2018.

[39] Gustavo Daniel Vega Magdaleno, Vladislav Bespalov, Yalin Zheng, Alex A Freitas, and Joao Pedro De Magalhaes. Machine learning-based predictions of dietary restriction associations across ageing-related genes. *BMC bioinformatics*, 23(1):1–28, 2022.

[40] Tijana Milenković, Han Zhao, and Fazle E Faisal. Global network alignment in the context of aging. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, pages 23–32, 2013.