

Brief Report

Not peer-reviewed version

Sex Age Annotated Omics Data: Enabling Powerful Omics Studies

[Pietro Hiram Guzzi](#)^{*}, [Francesca Cortese](#), [Raffaele Giancotti](#), [Ugo Lomoio](#), [Barbara Puccio](#), [Patrizia Vizza](#), [Giuseppe Tradigo](#), [Pierangelo Veltri](#)

Posted Date: 13 March 2023

doi: [10.20944/preprints202303.0207.v1](https://doi.org/10.20944/preprints202303.0207.v1)

Keywords: Annotation Proteome



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Brief Report

Sex Age Annotated Omics Data: Enabling Powerful Omics Studies

Pietro Hiram Guzzi ^{1,*}, Francesca Cortese ¹, Raffaele Giancotti ¹, Ugo Lomoio ¹, Barbara Puccio ¹, Patrizia Vizza ¹, Giuseppe Tradigo ² and Pierangelo Veltri ³

¹ Department of Surgical and Medical Sciences, Universiti of Catanzaro; hguzzi@unicz.it (PHG), francesca.cortese001@unicz.it (FC), raffaele.giancotti@unicz.it (RG), ugo.lomoio@unicz.it (UL), Barbara Puccio (BP) puccio@unicz, vizza@unicz.it (PV),

² University Ecampus, Novedrate, Como, Italy; giuseppe.tradigo@uniecampus.it

³ University of Calabria, Department of Informatics, Modelling, Electronics and Systems (DIMES); pierangelo.veltri@dimes.unical.it

* Correspondence: Pietro Hiram Guzzi hguzzi@unicz.it; Tel.: +39-09613694148 (P.H.G.)

Abstract: There is increasing evidence that many molecular processes exhibit differences with age and sex. Such differences produces also differences in the insurgence and progression of many complex diseases. For instances, demographic data on insurgence of comorbidities of mellitus diabetes, on lethality of COVID-19, and on some cancers, shows differences between sex and age groups. Therefore, the growing interest on such area requires the management of related data as well as the development of algorithms and tool for the analysis. The availability of omics data annotated with metadata related to age and sex is mandatory for building pipeline of the analysis. The number of databases containing data related to age and sex is henceforth growing. We here show some databases and tools storing such data. Finally, future research direction are highlighted.

Keywords: annotation; biological databases; aging analysis

1. Introduction

In healthy individuals molecular mechanisms regulate normal pathophysiology. Diseases usually cause alteration of such mechanisms which in turn cause functional outcomes. Omics studies aims to fill the gap in our knowledge of understanding of these distinct and common mechanisms

Recently, it has been noted that male and female sex share some common molecular mechanisms and differ by some others [1]. Moreover, such mechanisms change with age [2–8] and changes present peculiar distinctive traits with sex [9] and this characteristics belong to many species. In particular, in humans ageing represents the progressive insurgence and accumulation of changes at genomic, proteomic, and epigenomics level related to these changes [10–15]. The whole disclosing of such changes may help to develop novel therapies for many diseases which present different characteristics with age and sex [2–8,16,17].

Consequently, omics studies in individuals should consider as factors both age (e.g. age groups) and sex. Unfortunately, many independent research projects utilizes sex and age usually in an aggregated manners, so results are not aware of these differences [18].

To disclose molecular mechanism related to sex and age differences, there is the need to integrate heterogeneous data produced by experiments of different laboratories (e.g. omics, epigenomics, and medical images) and, mainly, to provide metadata concerning age and sex [19–21].

Research in such area is based on some key points: (i) introduction of publicly available omics databases annotated with age and sex information; (ii) the introduction of standards related to data modelling and exchange; (iii) the development of methods and models for data integration and analysis, also leveraging capabilities from deep learning and artificial intelligence [22–25].

Unfortunately, to the best of our knowledge, there exist few databases providing annotated information. A Google query reports such results: GTEx [26] data portal contains RNA-Seq data

annotated with tissue of provenance and sex and age (grouped into six classes) of patients, SAGD contains sex-associated genes differential expression [27] in multiple species.

Consequently there is the need for the introduction of annotated databases and annotation aware algorithms for the analysis of such data. More recently, the GTEx-Visualizer platform enable [] the query, visualization and analysis at age, sex, and tissue level of GTEx data. In this paper we report the use of GTEx-Visualiser to a general purpose experiment of age and sex analysis.

1.1. Related Work

We here report state of the art databases storing age sex annotated omics data.

The GTEx data portal [26] is a public available resource which collect data related to whole-genome sequencing and RNA-seq in individuals. It provides metadata such as tissue of provenance, sex, and age (grouped into six classes) of the patients. The current version of the GTEx database (accessed on February 22th, 2023) stores 17382 samples of 54 tissues of 948 donors, see at <https://gtexportal.org/home/tissueSummaryPage>. GTEx has a web interface which offers query interfaces and visualisation. In recent years many independent studies used GTEx data to perform ageing-based analysis [18,28–30].

GTEx data portal presents some limitation in age/sex studies since users cannot query data grouped by age or sex, and data are not integrated with existing protein interaction databases. For example, the used who needs to perform analysis at sex/age level has to download the whole database to extract data with own scripts, and this is a significant limitation for the inexperienced user. 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 [8,12,31–34].

SAGD (Sex-Associated Genes Database) [27] is a public database of sex-associated genes available at <http://bioinfo.life.hust.edu.cn/SAGD>. SAGD (accessed on March, 04 2023) contains data of RNA-SEQ of genes presenting difference between males and females in different species. Whenever available, the database also present annotation related to tissue and age (child, fetal, adult). It integrates curated public RNA-seq datasets from multiple species, presenting differential expressio of paired female and male biological replicates from the same condition. It stores identified 2,828 samples of 21 species.

Users can browse SAGs by gene, species, drug and dataset. Main limitations of SAGD is lack of depth in age groups, limitated query possibilities and absence of raw data. Main strengths of the database are the annotations related to targeting drugs, homologs, ontology and related RNA-seq datasets of SAGs.

NOMA-DB [35] is a framework allowing the query of age related genes based on the GTEx database. Current version of NOMA-DB, available at <https://noma-db.org/> allows to query and navigate the database by using sex and age information to perform data analysis of genes related to diabetes comorbidities. The framework wraps the GTEx data and it is based on an application logic level on top of such data. The current version enables the analysis of genes by tissue, gene and age, thus it may be used to analyse aging/sex-related molecular mechanisms based on the analysis of expression data.

The AgingAtlas [36] 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 the 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 [31] <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 [37]. Similarly to AgingAtlas it does not contain tissue level data.

2. Results

In this section we show some case studies of analysis of sex-age related genes.

2.1. Using NOMA-DB for studying gene changes with age.

In this section we show the use of NOMA-DB for the study of changes of gene expression. We use gene related to type 2 diabetes mellitus (T2DM) comorbidities listed in the T2DiACoD database [38]. The code of NOMA-DB is available at <https://github.com/roccoscicchitano001/DjangoAPI/tree/main/ui/Gtex-ui>, accessed at March 08th 2023. T2DM is a chronic disease [39] and often presents at least one comorbidity in patients [40–43]. Demographic data shows the prevalence of T2DM in male adults over 65 years [39,44–46]. This may suggest the presence of a differential behavior which may be explained at molecular level.

Consequently, we selected a list of 650 genes associated with type 2 diabetes from T2DiACoD database. Then we used NOMA-DB framework for retrieving expression data at tissue level stratified for age and sex.

We accessed the web interface of the NOMA-DB framework and we performed queries for all the tissues by setting following parameters: Tissue - liver, Age Interval - all, Sex - Both. Each parameter has a range of values corresponding to the information of GTEx data portal. Thus user can select one of the tissues present in such database, a specific age interval (or all the age intervals), and the desired sex (or both).

```
select *
from gtex
where
    Tissue='Liver',
    and sex ='M'
    and Age='60-69';
```

Alternatively, the user may retrieve the same data by using the web interface, available at <https://github.com/roccoscicchitano001/DjangoAPI/tree/main/ui/Gtex-ui> for a easy installation via Docker interface. First, user may select the genes and the tissue he/she want to analyse as reported in Figure 1

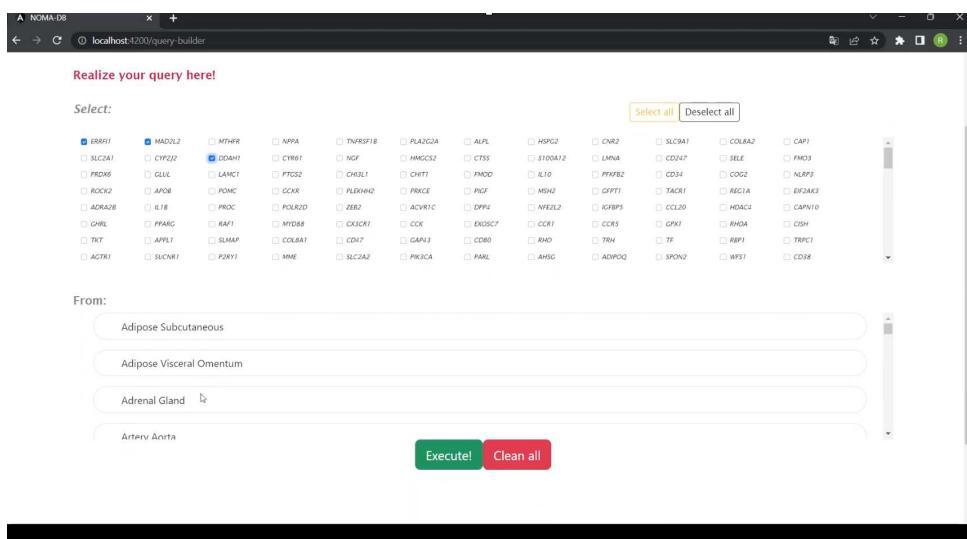
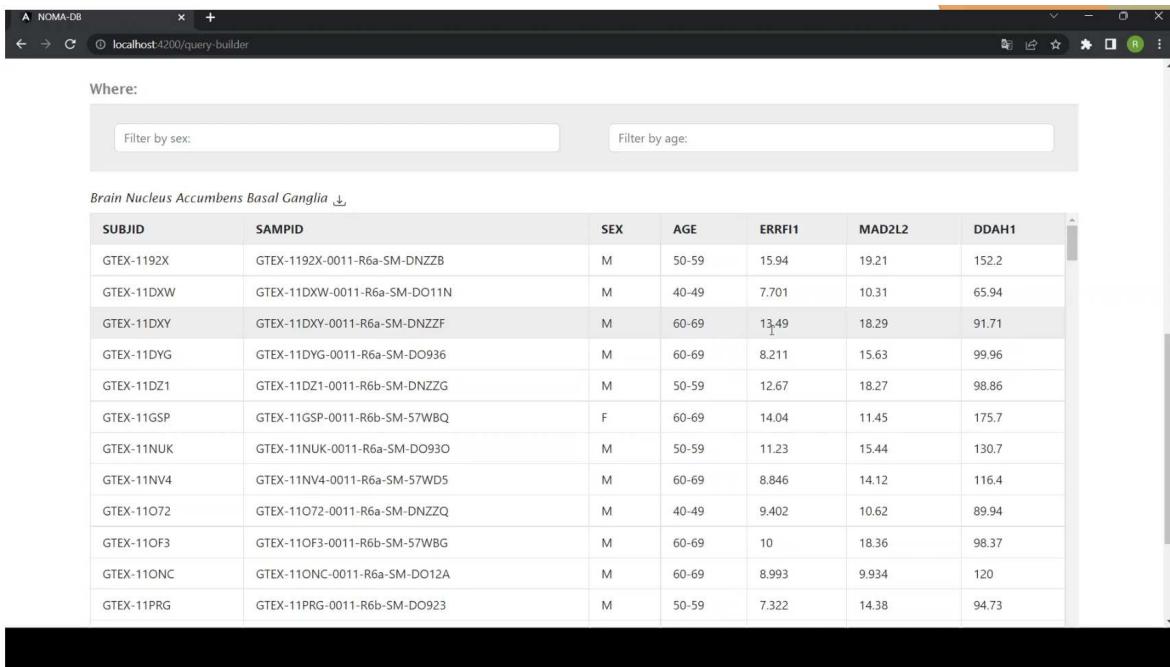


Figure 1. Figure depicts the user interface of NOMA-DB for selecting genes and tissue of interest.

NOMA-DB returns as output the list of the genes and their expression values and the annotation related to age, tissue and sex as reported in Figure 2.



SUBJID	SAMPID	SEX	AGE	ERRF11	MAD2L2	DDAH1
GTEX-1192X	GTEX-1192X-0011-R6a-SM-DNZZB	M	50-59	15.94	19.21	152.2
GTEX-11DXW	GTEX-11DXW-0011-R6a-SM-DO11N	M	40-49	7.701	10.31	65.94
GTEX-11DXY	GTEX-11DXY-0011-R6a-SM-DNZZF	M	60-69	13.49 ₁	18.29	91.71
GTEX-11DYG	GTEX-11DYG-0011-R6a-SM-DO936	M	60-69	8.211	15.63	99.96
GTEX-11DZ1	GTEX-11DZ1-0011-R6b-SM-DNZZG	M	50-59	12.67	18.27	98.86
GTEX-11GSP	GTEX-11GSP-0011-R6b-SM-57WBQ	F	60-69	14.04	11.45	175.7
GTEX-11NUK	GTEX-11NUK-0011-R6a-SM-DO93O	M	50-59	11.23	15.44	130.7
GTEX-11NV4	GTEX-11NV4-0011-R6a-SM-57WD5	M	60-69	8.846	14.12	116.4
GTEX-11O72	GTEX-11O72-0011-R6a-SM-DNZZQ	M	40-49	9.402	10.62	89.94
GTEX-11OF3	GTEX-11OF3-0011-R6b-SM-57WBG	M	60-69	10	18.36	98.37
GTEX-11ONC	GTEX-11ONC-0011-R6a-SM-DO12A	M	60-69	8.993	9.934	120
GTEX-11PRG	GTEX-11PRG-0011-R6b-SM-DO923	M	50-59	7.322	14.38	94.73

Figure 2. Figure depicts the user interface of NOMA-DB showing the obtained results.

Starting from this dataset (which cannot be retrieved by using the query interface of GTEx), user, after downloading it, can perform subsequent analysis, such as retrieving differential expression networks, or finding pattern of co-evolution of genes.

Author Contributions: “Conceptualization, PHG and PV; software, UL.; validation, GT BP RG PV; writing—original draft preparation, PHG.; writing—review and editing, PHG PV.; visualization, X.X.; supervision, X.X.; project administration, PV; funding acquisition, PV. All authors have read and agreed to the published version of the manuscript.”

Funding: “This research was funded by PON-VQA MISE

Acknowledgments: Authors thank Rocco Scicchitano for his work in developing NOMA-DB.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bond, K.M.; McCarthy, M.M.; Rubin, J.B.; Swanson, K.R. Molecular omics resources should require sex annotation: a call for action. *Nature methods* **2021**, *18*, 585–588.
2. Partridge, L. Intervening in ageing to prevent the diseases of ageing. *Trends in Endocrinology & Metabolism* **2014**, *25*, 555–557.
3. Childs, B.G.; Gluscevic, M.; Baker, D.J.; Laberge, R.M.; Marques, D.; Dananberg, J.; Van Deursen, J.M. Senescent cells: an emerging target for diseases of ageing. *Nature reviews Drug discovery* **2017**, *16*, 718–735.
4. Rockwood, K.; Howlett, S.E. Age-related deficit accumulation and the diseases of ageing. *Mechanisms of ageing and development* **2019**, *180*, 107–116.
5. Leonardi, G.C.; Accardi, G.; Monastero, R.; Nicoletti, F.; Libra, M. Ageing: from inflammation to cancer. *Immunity & Ageing* **2018**, *15*, 1–7.
6. De Magalhães, J.P. How ageing processes influence cancer. *Nature Reviews Cancer* **2013**, *13*, 357–365.
7. Bi, S.; Liu, Z.; Wu, Z.; Wang, Z.; Liu, X.; Wang, S.; Ren, J.; Yao, Y.; Zhang, W.; Song, M.; others. SIRT7 antagonizes human stem cell aging as a heterochromatin stabilizer. *Protein & cell* **2020**, *11*, 483–504.

8. Gallo Cantafio, M.E.; Grillone, K.; Caracciolo, D.; Scionti, F.; Arbitrio, M.; Barbieri, V.; Pensabene, L.; Guzzi, P.H.; Di Martino, M.T. From single level analysis to multi-omics integrative approaches: a powerful strategy towards the precision oncology. *High-throughput* **2018**, *7*, 33.
9. Bond, S.T.; Calkin, A.C.; Drew, B.G. Sex differences in white adipose tissue expansion: emerging molecular mechanisms. *Clinical Science* **2021**, *135*, 2691–2708.
10. Spinelli, R.; Parrillo, L.; Longo, M.; Florese, P.; Desiderio, A.; Zatterale, F.; Miele, C.; Raciti, G.A.; Beguinot, F. Molecular basis of ageing in chronic metabolic diseases. *Journal of Endocrinological Investigation* **2020**, *43*, 1373–1389.
11. Teruya, T.; Goga, H.; Yanagida, M. Human age-declined saliva metabolic markers determined by LC–MS. *Scientific reports* **2021**, *11*, 1–11.
12. Nassa, G.; Tarallo, R.; Guzzi, P.H.; Ferraro, L.; Cirillo, F.; Ravo, M.; Nola, E.; Baumann, M.; Nyman, T.A.; Cannataro, M.; others. Comparative analysis of nuclear estrogen receptor alpha and beta interactomes in breast cancer cells. *Molecular BioSystems* **2011**, *7*, 667–676.
13. Kerber, R.A.; O'Brien, E.; Cawthon, R.M. Gene expression profiles associated with aging and mortality in humans. *Aging Cell* **2009**, *8*, 239–250.
14. He, X.; Memczak, S.; Qu, J.; Belmonte, J.C.I.; Liu, G.H. Single-cell omics in ageing: a young and growing field. *Nature Metabolism* **2020**, *2*, 293–302.
15. Guzzi, P.H.; Cannataro, M. μ -CS: An extension of the TM4 platform to manage Affymetrix binary data. *BMC bioinformatics* **2010**, *11*, 315.
16. Boheler, K.R.; Volkova, M.; Morrell, C.; Garg, R.; Zhu, Y.; Margulies, K.; Seymour, A.M.; Lakatta, E.G. Sex-and age-dependent human transcriptome variability: implications for chronic heart failure. *Proceedings of the national academy of Sciences* **2003**, *100*, 2754–2759.
17. Oosenbrug, E.; Marinho, R.P.; Zhang, J.; Marzolini, S.; Colella, T.J.; Pakosh, M.; Grace, S.L. Sex differences in cardiac rehabilitation adherence: a meta-analysis. *Canadian Journal of Cardiology* **2016**, *32*, 1316–1324.
18. Mercatelli, D.; Pedace, E.; Veltri, P.; Giorgi, F.M.; Guzzi, P.H. Exploiting the molecular basis of age and gender differences in outcomes of SARS-CoV-2 infections. *Computational and Structural Biotechnology Journal* **2021**, *19*, 4092–4100.
19. Valdes, A.M.; Glass, D.; Spector, T.D. Omics technologies and the study of human ageing. *Nature Reviews Genetics* **2013**, *14*, 601–607.
20. Zierer, J.; Menni, C.; Kastenmüller, G.; Spector, T.D. Integration of 'omics' data in aging research: from biomarkers to systems biology. *Aging cell* **2015**, *14*, 933–944.
21. Lorusso, J.S.; Sviderskiy, O.A.; Labunskyy, V.M. Emerging omics approaches in aging research. *Antioxidants & Redox Signaling* **2018**, *29*, 985–1002.
22. Hühne, R.; Thalheim, T.; Sühnel, J. AgeFactDB—the JenAge Ageing Factor Database—towards data integration in ageing research. *Nucleic acids research* **2014**, *42*, D892–D896.
23. Fabris, F.; Magalhães, J.P.d.; Freitas, A.A. A review of supervised machine learning applied to ageing research. *Biogerontology* **2017**, *18*, 171–188.
24. Ai, R.; Jin, X.; Tang, B.; Yang, G.; Niu, Z.; Fang, E.F. Ageing and Alzheimer's Disease: Application of Artificial Intelligence in Mechanistic Studies, Diagnosis, and Drug Development. In *Artificial Intelligence in Medicine*; Springer, 2021; pp. 1–16.
25. Miloulis, S.T.; Kakkos, I.; Anastasiou, A.; Matsopoulos, G.K.; Koutsouris, D. Application of Artificial Intelligence Towards Successful Ageing: A Holistic Approach. In *Modern Challenges and Approaches to Humanitarian Engineering*; IGI Global, 2022; pp. 172–193.
26. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; others. The genotype-tissue expression (GTEx) project. *Nature genetics* **2013**, *45*, 580–585.
27. Shi, M.W.; Zhang, N.A.; Shi, C.P.; Liu, C.J.; Luo, Z.H.; Wang, D.Y.; Guo, A.Y.; Chen, Z.X. SAGD: a comprehensive sex-associated gene database from transcriptomes. *Nucleic Acids Research* **2019**, *47*, D835–D840.
28. Stanfill, A.G.; Cao, X. Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database. *Biological research for nursing* **2021**, *23*, 533–540.
29. Pressler, M.P.; Horvath, A.; Entcheva, E. Sex-dependent transcription of cardiac electrophysiology and links to acetylation modifiers based on the GTEx database. *Frontiers in Cardiovascular Medicine* **2022**, *9*.

30. Ortuso, F.; Mercatelli, D.; Guzzi, P.H.; Giorgi, F.M. Structural genetics of circulating variants affecting the SARS-CoV-2 spike/human ACE2 complex. *Journal of Biomolecular Structure and Dynamics* **2021**, pp. 1–11.

31. de Magalhaes, J.P.; Toussaint, O. GenAge: a genomic and proteomic network map of human ageing. *FEBS letters* **2004**, *571*, 243–247.

32. Gu, S.; Jiang, M.; Guzzi, P.H.; Milenković, T. Modeling multi-scale data via a network of networks. *Bioinformatics* **2022**, *38*, 2544–2553.

33. Li, Q.; Newaz, K.; Milenković, T. Improved supervised prediction of aging-related genes via weighted dynamic network analysis. *BMC bioinformatics* **2021**, *22*, 1–26.

34. Agapito, G.; Guzzi, P.H.; Cannataro, M. Parallel extraction of association rules from genomics data. *Applied Mathematics and Computation* **2019**, *350*, 434–446.

35. Guzzi, P.H.; Lomoio, U.; Sicchitano, R.; Veltri, P. NOMA-DB: a framework for management and analysis of ageing-related gene-expression data. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2022, pp. 1905–1910.

36. Aging Atlas: a multi-omics database for aging biology. *Nucleic Acids Research* **2021**, *49*, D825–D830.

37. Tacutu, R.; Thornton, D.; Johnson, E.; Budovsky, A.; Barardo, D.; Craig, T.; Diana, E.; Lehmann, G.; Toren, D.; Wang, J.; Fraifeld, V.E.; de Magalhães, J.P. Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research* **2017**, *46*, D1083–D1090, [<https://academic.oup.com/nar/article-pdf/46/D1/D1083/23162151/gkx1042.pdf>]. doi:10.1093/nar/gkx1042.

38. Rani, J.; Mittal, I.; Pramanik, A.; Singh, N.; Dube, N.; Sharma, S.; Puniya, B.L.; Raghunandanan, M.V.; Mobeen, A.; Ramachandran, S. T2DiACoD: a gene atlas of type 2 diabetes mellitus associated complex disorders. *Scientific Reports* **2017**, *7*, 1–21.

39. Nowakowska, M.; Zghebi, S.S.; Ashcroft, D.M.; Buchan, I.; Chew-Graham, C.; Holt, T.; Mallen, C.; Van Marwijk, H.; Peek, N.; Perera-Salazar, R.; others. The comorbidity burden of type 2 diabetes mellitus: patterns, clusters and predictions from a large English primary care cohort. *BMC medicine* **2019**, *17*, 1–10.

40. Igray, K.; Hannachi, H.; Joseph Howie, P.; Xu, J.; Li, X.; Engel, S.S.; Moore, L.M.; Rajpathak, S. Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus. *Current medical research and opinion* **2016**, *32*, 1243–1252.

41. Succurro, E.; Marini, M.A.; Fiorentino, T.V.; Perticone, M.; Sciacqua, A.; Andreozzi, F.; Sesti, G. Sex-specific differences in prevalence of nonalcoholic fatty liver disease in subjects with prediabetes and type 2 diabetes. *Diabetes Research and Clinical Practice* **2022**, *190*, 110027.

42. Bellary, S.; Kyrou, I.; Brown, J.E.; Bailey, C.J. Type 2 diabetes mellitus in older adults: clinical considerations and management. *Nature Reviews Endocrinology* **2021**, *17*, 534–548.

43. Pirillo, A.; Casula, M.; Olmastroni, E.; Norata, G.D.; Catapano, A.L. Global epidemiology of dyslipidaemias. *Nature Reviews Cardiology* **2021**, *18*, 689–700.

44. Pearson-Stuttard, J.; Holloway, S.; Polya, R.; Sloan, R.; Zhang, L.; Gregg, E.W.; Harrison, K.; Elvidge, J.; Jonsson, P.; Porter, T. Variations in comorbidity burden in people with type 2 diabetes over disease duration: A population-based analysis of real world evidence. *EClinicalMedicine* **2022**, *52*, 101584.

45. Guerrero-Fernández de Alba, I.; Orlando, V.; Monetti, V.M.; Mucherino, S.; Gimeno-Miguel, A.; Vaccaro, O.; Forjaz, M.J.; Poblador Plou, B.; Prados-Torres, A.; Riccardi, G.; others. Comorbidity in an older population with type-2 diabetes mellitus: identification of the characteristics and healthcare utilization of high-cost Patients. *Frontiers in pharmacology* **2020**, *11*, 586187.

46. Dworzynski, P.; Aasbrenn, M.; Rostgaard, K.; Melbye, M.; Gerds, T.A.; Hjalgrim, H.; Pers, T.H. Nationwide prediction of type 2 diabetes comorbidities. *Scientific reports* **2020**, *10*, 1–13.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.