COMPUTATIONAL BIOLOGY

Single-cell gene regulation across aging tissues

Transcription factors can control cell identity and function in health and disease. However, how they do so during aging is incompletely explored. Maity and colleagues identify age-related changes in gene regulation by analyzing the expression patterns of transcription-factor target genes in single-cell transcriptomics data.

Cyril Lagger and João Pedro de Magalhães

Aging is the main risk factor for many diseases, but it acts in complex ways on a mosaic of molecular and cellular processes that have yet to be fully identified and remain poorly understood[1]. On the one hand, evidence suggests that aging has distinct effects on gene expression and molecular pathways in different cell types and tissues[2]. On the other hand, some studies have found common patterns of gene expression changes in aging across tissues (for example, conserved signatures encompassing changes in genes involved in inflammation and immune response, cell cycle, collagen processing, and metabolism and mitochondrial functions[3,4]. Indeed, one of the holy grails of biogerontology remains the identification of a restricted set of ubiquitous molecular processes that, once targeted, could prevent physiological decline and disease occurrence. In this issue of *Nature Aging*, Maity and colleagues provide us with an exciting bioinformatics analysis that identifies age-related changes in mouse gene regulation[5]. They notably reveal transcription factors whose activity is altered with age similarly across several tissues. The main processes regulated by these transcription factors are T cell differentiation, macrophage polarization, inflammation, antigen processing, collagen processing and the circadian rhythm.

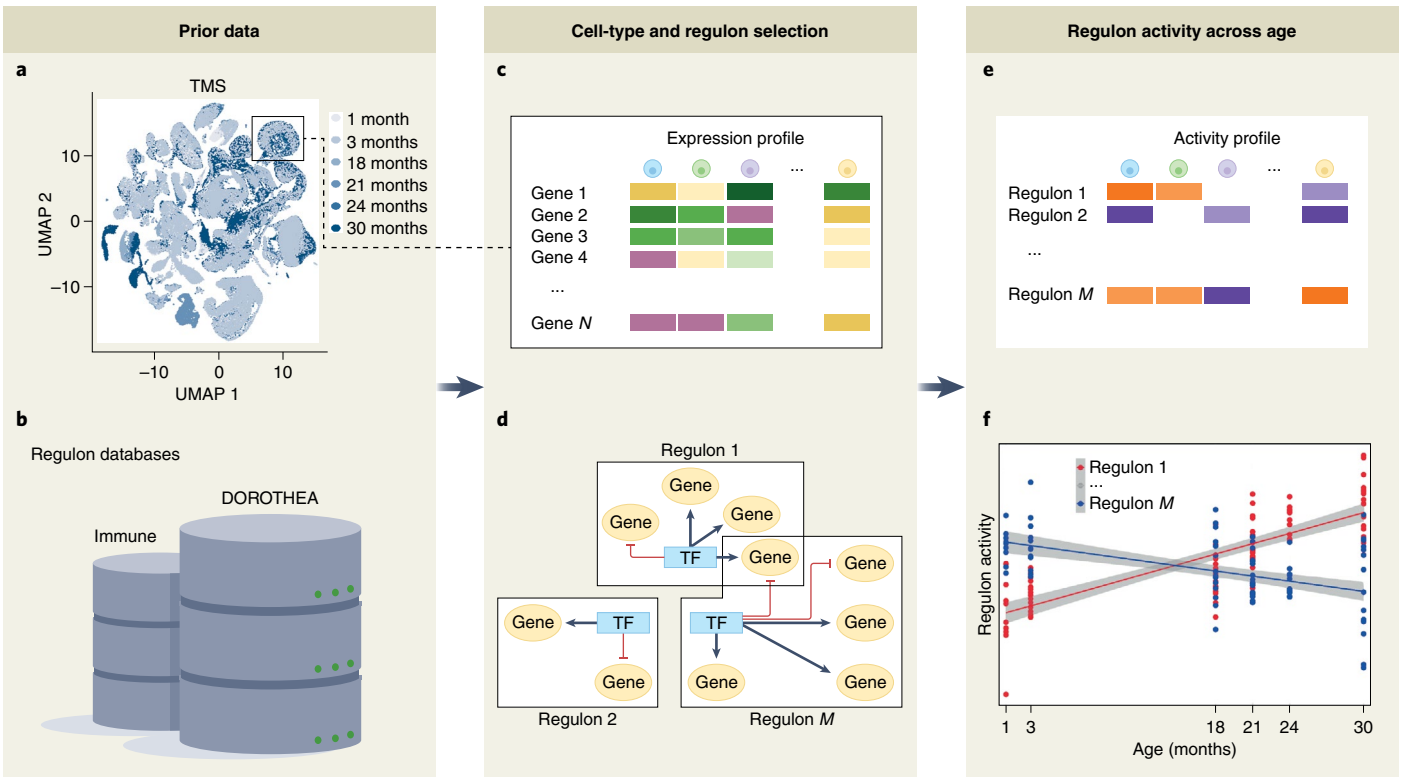

Fig. 1 | Assessing age-related changes in mouse gene regulatory activity across cell types from the TMS single-cell dataset. Maity et al.[5] developed a bioinformatics approach to identify regulon activity changing with age. **a**, To cover as many mouse tissues and cell types as possible, they applied their method to the TMS single-cell dataset. UMAP, uniform manifold approximation and projection. **b**, Immune-specific regulons were built by the authors themselves, whereas generic nonimmune regulons were obtained from the DOROTHEA database. **c**–**e**, For each cell type, the authors combined the expression profile of each cell (**c**) with the internal structure of the relevant regulons (**d**) to obtain the activity profile of these regulons in each cell (**e**). **f**, Fitting a linear regression between such activity scores and age allowed the authors to extract upregulated (red line and points) or downregulated (blue line and points) regulons. Repeating this process across cell types and tissues then shed light on which changes are shared across the entire organism and might be linked to universal age-related processes.

To infer transcription factor activity, the authors used transcription factor regulons. A regulon is defined as a set of genes composed of a transcription factor and its target genes, which can be either activated or repressed by the transcription factor itself. Single-cell transcriptomic data then allowed the authors to compute an 'activity score' for each regulon in a given cell or sample, by mathematically combining the expression of its genes. Recent studies, including that by Maity et al.[5], have shown the advantages of such scores over the use of gene expression alone[6,7]. For instance, when comparing biological conditions, a small change in expression of a transcription factor could easily be overlooked or tagged as non-significant by regular differential expression analysis. By contrast, if both this transcription factor and some of its targets are altered consistently, the activity score might be able to reveal a significant change at the level of the regulon. An important limitation of this method, however, is that regulon databases might be incomplete, and that important regulatory activities might consequently be missed.

The philosophy behind this approach is not dissimilar to multiscale or mean-field methods, which are commonly used in physics and are becoming increasingly popular in biology[8]. When a complex system is made of many individual components interacting together, the study of a simpler or coarse-grained system is often useful to extract behaviors that occur at larger scales. Here, the regulon approach reduces the dimensionality of the problem by at least one order of magnitude — namely, by considering about 1,000 regulons instead of 20,000 or more genes. This is particularly relevant to help to bridge the gap between molecular and higher-level changes that occur with age, as global regulation is probably more relevant for tissue homeostasis and physiological functions than is single gene expression.

However, several questions need to be addressed when working with regulons, including how these structures are constructed and whether they are cell-type- or condition-specific. Maity et al.[5] carefully consider these aspects. One part of their analysis relies on the DOROTHEA database, which encompasses about 1,300 mouse regulons that are not specific to cell types and tissues[6]. One strength of DOROTHEA database is that it is carefully curated: each target of each regulon is assigned a confidence level on the basis of supporting evidence. This includes chromatin immunoprecipitation with sequencing peaks, transcription factor binding motifs on promoters, literature-curated resources and inference from gene expression. Being generic, the second advantage of DOROTHEA is that its regulons can be used with any transcriptomics dataset, regardless of its context. The downside of this lack of cell-type specificity is a potential reduction in the accuracy of the activity scores.

To address this problem, the authors build their own cell-type-specific regulon by using SCIRA[9], one of their previously developed machine learning tools. SCIRA uses a multivariate partial correlation model to infer associations between transcription factors and target genes from bulk transcriptomics data. The difficulty of such an endeavor is having access to cellularly homogeneous data (that is, samples purified for a specific cell type), which are currently limited. Here, Maity et al.[5] rely on mRNA expression from 169 fluorescence-activated cell-sorted blood cell types from the Heamosphere/Haemopedia database[10]. They manage to build hematopoietic-specific regulons that characterize the differentiation activity of various immune cells. The authors validate these regulons by using independent datasets and confirming that their activity is cell-type-specific. For instance, a B cell-specific regulon is validated by confirming that its activity is indeed higher in B cell samples than in samples from other cell types, across several independent datasets. As expected, these regulons have better validation accuracy than those from DOROTHEA. However, each of these cell-type-specific regulons can be applied only to transcriptomics samples originating from the corresponding cell type. For all other cell types, DOROTHEA was used in the study.

Equipped with these two sets of regulons, the authors compute activity scores across as many cell types and tissues as possible (Fig. 1). The authors use the Tabula Muris Senis (TMS) single-cell RNA-sequencing (scRNA-seq) dataset, which contains 350,000 cells distributed across 23 organs, more than 150 cell types, 6 time points and 2 experimental methods for their study — without a doubt, one of the most suitable resources for such analysis to date[11]. Although TMS has already been mined to infer different types of age-related change, Maity et al.[5] are the first to use it to study gene regulation. The authors fit a linear regression model between regulon activity score and age for each cell type, whenever there are enough cells in each age group. This is a limitation of TMS that is sometimes overlooked: a substantial number of cell types are represented only at a few time points.

Overall, Maity et al.[5] manage to compute correlations between age and activity scores for roughly 300 immune-specific regulons in the TMS lymphoid and myeloid cells, and for 1,155 DOROTHEA regulons in 75 nonimmune cell types across 11 tissues. They report decreased activity of *Lef1* and *Ankrd10* regulons in $CD4^+$ T cells as well as increased activity of *Batf* in T cells in both lung and spleen, suggesting a reduction of naive $CD4^+$ T cells with age and a shift toward a more-differentiated state. Such age-related changes in the T cell repertoire have already been well documented in the context of immunosenescence and thymic involution, mostly in peripheral blood and lymph nodes[12]. Here, the authors not only provide further evidence for this pattern but also show its relevance to other tissues, such as the lung.

Another major pattern that the authors identify in their study is the decline in *Klf4* activity in monocytes and macrophages from the lung, kidney and brain. This is interpreted by Maity et al.[5] as a possible shift from M1 to M2 polarization, owing to the observation that TMS alveolar macrophages of the lung contain a lower fraction of M1 macrophages with age. The authors also show that tumor-associated macrophages from an independent human lung scRNA-seq dataset[13] display lower activities of macrophage-specific regulons, including *KLF4*. They argue that regulation of macrophage polarization might, therefore, be a key element in the complex relationship between aging and cancer, as the shift toward M2 polarization with age might create a pro-oncogenic microenvironment. This finding should motivate further investigation into the role of aged macrophages in the onset and development of cancers.

Regarding nonimmune cells, the analysis of Maity and colleagues[5] reveals several regulons that are altered consistently in multiple cell types. These include transcription factors which have received little consideration from the aging community so far. Of interest, the *Zfp746* regulon showed increased activity in 33 TMS cell types. To the best of our knowledge, the ZFP746 transcription factor (or its human ortholog ZNF746, also known as PARIS) has never been directly related to aging, but it has been associated with Parkinson's disease, with oncogenic roles in various tissues and, recently, with the induction of cellular senescence[14]. This transcription factor and its targets deserve to be further studied for their roles in aging and age-related diseases. The same can be said for other observed changes, including an increase in *Rfx5* (potentially reducing collagen production), an increase in *Nfkb1* (a regulator of inflammation) and a

decrease in *Arntl* (a regulator of the circadian rhythm).

Changes in expression in genes related to collagen, immune system and inflammation, cell senescence and other processes identified by Maity et al.[5] have previously been found to be molecular signatures of aging[3], but the regulation and which cell types contribute to these signatures has remained poorly understood in the context of aging. By identifying tissue- and cell-type-independent alterations of master gene regulators, Maity et al.[5] provide important information on these signatures that sheds new light on the organization of the aging mosaic. ❐

Cyril Lagger ✉ and João Pedro de Magalhães ✉

Integrative Genomics of Ageing Group, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK.

✉*e-mail: cyril.lagger@liverpool.ac.uk; jp@senescence.info*

Published online: 17 June 2022
https://doi.org/10.1038/s43587-022-00238-4

References

1. Cevenini, E. et al. *Curr. Pharm. Des.* **16**, 802–813 (2010).
2. Nie, C. et al. *Cell Rep.* **38**, 110459 (2022).
3. de Magalhães, J. P., Curado, J. & Church, G. M. *Bioinformatics* **25**, 875–881 (2009).
4. Palmer, D., Fabris, F., Doherty, A., Freitas, A. A. & de Magalhães, J. P. *Aging* **13**, 3313–3341 (2021).
5. Maity, A. K., Hu, X., Zhu, T. & Teschendorff, A. E. et al. *Nat. Aging* https://doi.org/10.1038/s43587-022-00233-9 (2022).
6. Holland, C. H. et al. *Genome Biol.* **21**, 36 (2020).
7. Van de Sande, B. et al. *Nat. Protoc.* **15**, 2247–2276 (2020).
8. Alber, M. et al. *NPJ Digit. Med.* **2**, 115 (2019).
9. Teschendorff, A. E. & Wang, N. *NPJ Genom. Med.* **5**, 43 (2020).
10. de Graaf, C. A. et al. *Stem Cell Reports* **7**, 571–582 (2016).
11. The Tabula Muris Consortium. *Nature* **583**, 590–595 (2020).
12. Goronzy, J. J., Fang, F., Cavanagh, M. M., Qi, Q. & Weyand, C. M. *J. Immunol.* **194**, 4073–4080 (2015).
13. Lambrechts, D. et al. *Nat. Med.* **24**, 1277–1289 (2018).
14. Bae, J. H. et al. *Cell Death Dis.* **11**, 359 (2020).

Competing interests

J.P.d.M. is an advisor/consultant for the Longevity Vision Fund, NOVOS, YouthBio Therapeutics and the founder of Magellan Science Ltd, a company providing consulting services in longevity science. C.L. has no competing interest to declare.