Tendon ‘Omics: A Survey-Based Perspective on Obstacles and Opportunities

Anne E.C. Nichols[1,#], Stephanie G. Dakin[2,#], Dianne Little[3,#], Sarah J. B. Snelling[4,#], Alayna E. Loiselle[1,#]

[1] Center for Musculoskeletal Research, Department of Orthopaedics & Physical Performance, University of Rochester Medical Center, Rochester, NY, 14642, USA
[2] Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[3] Department of Basic Medical Sciences and Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, 47906, USA
[4] The Botnar Institute of Musculoskeletal Sciences, Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[#] Denotes equal contribution of all authors

Acknowledgements: The authors would like to thank and acknowledge all of the participants in the Tendon ‘Omics breakout discussion at the ORS Tendon Satellite Meeting at the Mayo Clinic:
Varun Arvind, Hani Awad, Emily Camilleri, Hoi Pan (Harry) Cheung, Brianne Connizzo, Giulia Crosio, Michael David, Nathaniel Dyment, Sushmitha Durgam, Dylan Easley, Pernille Eliasson, Ross Fitzsimmons, Anne Gingery, Eric Gracey, Alice Huang, Tommy Jenkins, Megan Killian, Emily King, Drew Koch, Rita Marqueti, Ruth McCarrick-Walmsley, Ashley Potter, Tracy Boakye Serebour, Hua Shen, Katie Sikes, Alexander Stigall, Herb Sun, Emily VanZeeland, Dan Michelle Wang, Farbod Yousefi

Key words: Tendon, transcriptomics, single cell RNA sequencing

Abstract

Until recently, our fundamental understanding of tendon biology, from development to post-natal growth, aging, and response to injury has been limited to bulk transcriptomic profiling and low-throughput molecular biology techniques. Recent advances in ‘omics technologies have provided a critical opportunity to define and interrogate the cellular and molecular landscape across different tissue states, with the ultimate goal of informing translational strategies to retain or restore tendon health. However, there are clear gaps, including infrastructure and expertise that have limited the ability to fully leverage the potential of these datasets. Therefore, in preparation for the 2024 Orthopaedic Research Society Tendon Section Satellite Meeting at the Mayo Clinic, a survey of the tendon community was circulated in Fall 2023 to understand the most-pressing challenges in tendon ‘omics that should be addressed at this meeting. Through this survey, three dominant themes emerged: (1) data sharing, emphasizing the need for open-source tools to query existing datasets; (2) data reporting, with a focus on the need for consistent reporting of associated metadata; and (3) data use and annotation, identifying the need for consensus molecular definition of different cell populations. These priorities then informed both the plenary and breakout sessions at the Satellite meeting. This perspectives article summarizes and synthesizes the survey data, plenary sessions, and discussion groups that developed community-driven priorities and actionable items to generate a community roadmap to enhance the impact of tendon ‘omics research.

Introduction

Given the frequency of both acute tendon injuries and clinical presentations of chronic, degenerative changes in the tendon, translational strategies are needed to retain and restore tendon structure-function to maintain or return patient quality of life. To that end, understanding the fundamental processes that govern the pathogenesis of different tendon diseases, how the tendon responds to acute vs. chronic injury, and how other factors (e.g., age, co-morbidities) effect these processes have been critical gaps in knowledge that recent advances have aimed to fill. More specifically, while comprehensive molecular profiling of the tendon has historically been limited by low-throughput gene by gene, or protein by protein analyses, recent work using high-dimensional molecular profiling, broadly termed ‘omics’ has facilitated a more robust, unbiased, and comprehensive characterization of the cell and molecular environment of different tendon processes across species. For example, technologies such as a single cell RNA sequencing (scRNA-seq), spatial transcriptomics, and proteomics have established the diversity of the tendon cell environment and highlighted spatial heterogeneity within a given tendon in terms of cell and ECM composition. Additionally, these platforms have demonstrated compositional differences across tendon types, established conserved and discrete populations and processes across species, during aging, and in the context of different injuries and disease states [1-11].

While part of the potential of these and other ‘omic techniques has been demonstrated by providing an unprecedented understanding of the cell environment in multiple contexts, there are key considerations, hurdles, and opportunities that need to be addressed at the level of the tendon research community to ensure we can leverage the full potential of these techniques. Therefore, at the 2024 Orthopaedic Research Society Tendon Section Satellite Meeting at the Mayo Clinic, a full plenary session and three breakout discussion sessions were focused on tendon ‘omics. In preparation for this session, in late 2023 a survey was circulated through the tendon community to define the current state of tendon ‘omics, existing datasets, as well as pressing needs, hurdles, and concerns about how to best utilize these important datasets. Over 40 individual researchers (88% Academia, 7% Industry, 5% Other) from various career stages (80% PI/Team Lead/Project Manager, 10% Postdoc, 5% Staff, 5% Other) and backgrounds (49% Biology, 37% Engineering, 12% Medicine, 2% Other) responded to the survey. Of the respondents, over 90% indicated that they currently

produce or plan to produce transcriptomic datasets in the near future. Datasets include a variety of tendons, species, and tissue states (**Figure 1**) as well as technologies (**Figure 2**), reflecting widespread adoption of these techniques across the community. Through this survey, three major themes emerged: data sharing, data reporting, and data use, which also included data annotation and integration. These three themes informed development of both the ‘omics plenary session and breakout sessions at the ORS Tendon Section Satellite meeting at the Mayo Clinic, which were conducted with the goals of: 1) highlighting the current state of the art of ‘omics technologies and their potential to inform translational approaches in other tissues, 2) understanding the current efforts for community-driven annotation, curation and sharing of ‘omics datasets in other tissues, and how the tendon community may partner with these resources, 3) defining the current state of the tendon ‘omics field, 4) identifying the most pressing needs within the tendon ‘omics community, and the hurdles to fully leveraging the power of these datasets for both accurate definition of fundamental tendon processes, and translational gain, and 5) establish short and long-term tangible tasks that the tendon community can tackle. In this perspective article we will summarize these sessions, outline the community-developed priorities and guidelines, and highlight actionable tasks and opportunities for the community to move the field forward.

To highlight the potential for ‘omics to inform translation, Dr. Ilya Korsunsky provided an overview of efforts within the rheumatology community to define the spatial-molecular environment of disease pathogenesis in rheumatoid arthritis and inflammatory disease [12, 13], and how different spatial cell/molecular profiles may be helpful in predicting treatment responsiveness [14]. Dr. Jennifer Westendorf then presented on the efforts of the International Federation of Musculoskeletal Research Societies (IFMRS) to develop the MSK knowledge portal (MSK-KP; https://msk.hugeamp.org/), an online repository and query-able data viewer that facilitates data sharing, data re-use, and hypothesis generation across musculoskeletal tissues [15].

Within the broad umbrella of ‘omics, the tendon researchers that participated in the breakout session identified more than 15 different ‘omics technologies that are currently being used including: scRNA-seq, bulk RNA sequencing, spatial transcriptomics, bulk proteomics, metabolomics and lipidomics. Given that transcriptomic (RNA sequencing) datasets currently account for the vast majority of datasets that have been generated in the tendon community, and thus currently represent the greatest opportunity for advancement,

this article is primarily focused on transcriptomics, particularly scRNA-seq. The two most commonly used species for tendon 'omic profiling were mouse and human, although datasets from other species (e.g., rat, horse, zebrafish, dog) provide additional important information, particularly when integrated across species, about the conserved and unique aspects of the cell/molecular environment of the tendon.

Data Sharing and Reporting

A key point raised frequently during the discussion was the need for tendon transcriptomic data sets to be shared openly and widely among investigators to advance our understanding of tendon biology. Transcriptomic datasets are most frequently generated by individual investigators to test a specific hypothesis or to characterize a particular process in tendon; however, these technologies generate a wealth of additional data that could be valuable in other research contexts if effectively shared. Reusing and reanalyzing existing datasets not only fosters collaboration between research groups but also makes data more accessible to those who might be deterred by the high costs or specialized expertise required to generate new datasets. Additionally, large-scale efforts to create comprehensive atlases of tendon markers and cell populations across different tendons, species, sexes, and disease states can only succeed if open data sharing is adopted by the entire tendon research community.

Currently, many tendon transcriptomic datasets are stored in publicly available data repositories including the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home) and the NIH Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/gds). However, the choice of repository is often dictated by the funding source, including the 2024 NIH Public Access Policy, and therefore varies between countries and funding agencies which can make finding and accessing datasets difficult. As such, development of an online resource to inventory and link to existing datasets may further enhance data access.

In addition to making the raw data from tendon transcriptomic studies widely available, there is a pressing need for the development of more user-friendly platforms for data visualization and quick query or analysis to make the data accessible to a broader range of researchers. This approach has already been taken with scRNA-seq data generated from uninjured mouse Achilles tendon [4] (https://mendiaslab.shinyapps.io/Tendon_SingleCell_Atlas/) available as interactive web-based "apps". These

apps require no specialized technical knowledge to access and interrogate and thus provide a great way for other investigators to screen datasets for useful information before committing to more time-intensive analysis. Potential drawbacks to this approach include the need for continued app support by individual investigators or institutions, which poses challenges for long-term data availability, and the need for computational expertise and/or collaborators to generate and publish the apps. Despite these potential issues, individually-hosted interactive datasets provide an important “middle-ground” to improve data accessibility in the short term and are strongly encouraged.

The use of publicly available datasets is further complicated by a lack of standardized metadata reporting. Immediate suggestions from the discussion to address these data accessibility challenges included a call for community consensus on which data repositories should be prioritized for consistent upload and standardization of which type of data are made publicly available in repositories (e.g. raw data vs. processed data). More specifically, the adoption of standards for metadata reporting by the tendon community is imperative to ensure that reused data is identified appropriately (sample identification, experimental conditions, patient characteristics, etc.) and that appropriate measures are taken to eliminate any technical artifacts upon re-analysis or integration into larger databases. For example, standardized reporting metrics, such as those developed the Human Cell Atlas (https://data.humancellatlas.org/metadata), should be included as supplemental information in manuscripts and uploaded to data repositories to facilitate the most efficient use of ‘omics datasets.

In addition to efforts by individual groups or investigators to make tendon datasets available, a major area of agreement in the group discussion was the recognition that it will be crucial for the advancement of the tendon field as a whole for there to be concerted efforts 1) to increase the representation of tendon in existing “big data” consortium efforts and 2) to aggregate the data from multiple studies into comprehensive, integrated (and ideally) interactive tendon databases that can be accessed by the entire community. For example, the MSK-KP genomic data mining platform enables the analysis of genetic/genomic data and potential linkages to human musculoskeletal phenotypes or pathologies. To date, no tendon datasets are included and therefore tendon researchers are unable to leverage the data available through the MSK-KP. Therefore, to improve tendon representation in existing projects, tendon investigators to represent the tendon community in working groups and who, in the short term, could facilitate integration of existing tendon datasets into established platforms,

should be identified, nominated, and supported. Long-term, having representatives of the tendon community present in these discussions will also facilitate the development of tendon-specific cellular atlases by giving crucial insight into the funding mechanisms, logistics, and governance of existing efforts. As an example, LungMAP [16] (https://www.lungmap.net/) is an integrated data portal that hosts standardized, multi-modal datasets from a broad group of investigators. Additionally, the Brain Initiative Cell Census Network (BICCN; https://biccn.org/) has compiled a multimodal cell atlas of the mammalian primary motor cortex, the result of a coordinated effort among neuroscience researchers [17]. Consortium efforts like LungMAP and BICCN will provide an important blueprint for the tendon community as we seek to build our own data portals and expand the impact of our individual datasets. Importantly, these consortium efforts require substantial funding, for example, the successful Accelerating Medicines Partnerships (AMP) programs are supported by public-private partnerships between NIH, FDA, and multiple pharma companies. Therefore, while smaller-scale efforts may yield some success, the full potential of these datasets, including leveraging the critical combination of cell/molecular tendon biology, clinical perspective, and computational biology, require substantial extramural support, which should be a priority for the community to advocate for. Moreover, it is also important to increase awareness of newly published tendon 'omic datasets, which could be accomplished through mechanisms such as a 'year in review' article in the *Journal of Orthopaedic Research* to highlight new datasets.

Finally, an additional important aspect of data sharing brought up during the discussion was the determination of authorship and the recognition that this is not a straightforward pursuit. The group agreed that an open and on-going discussion of what should constitute authorship within the community will be critical to ensure that investigators receive proper recognition for their efforts in generating, analyzing, and sharing data, while also ensuring adherence to the International Committee of Medical Journal Editors (ICMJE) guidelines on authorship.

Data use, annotation, and integration

Despite the massive increase in data generation, there has yet to be community-level discussion and efforts to reach consensus on how best to annotate these datasets in the context of our current knowledge. As such, researchers have annotated datasets according to their own best practice, resulting in difficulty in understanding one dataset in relation to another. The bulk of the group discussion focused on how to tackle this

important challenge. While there was discussion that different levels of annotation (e.g. function or location) will be important future work, there was general agreement that identifying a consensus set of genes that define different cell clusters in healthy tendon is a critical first step to the most informed and robust use of existing and future datasets. Moreover, while no single population was identified as a priority, it was recognized that some populations in the tendon may be more amenable to transfer-annotation approaches, including tendon-resident immune cell populations that have a comprehensive suite of markers that can be employed for highly-resolved annotation. In contrast, consistent definition of other tendon cell populations can provide important insights into whether or not there is a tendon-cell-specific signature, how expression of putative tendon markers (e.g., SCX, TNMD, MKX) vary across tendon cell subpopulations or between different anatomic sites, or if conserved fibroblast markers of healthy [18] or diseased [13] tissue are also present in healthy or diseased tendon. Finally, while the need for clear and consistent reporting on the cell source was agreed on, there was also discussion on the importance of developing a precise definition for the use of different anatomic descriptors (e.g., epitenon vs. paratenon vs. synovial sheath).

In the short-term, substantial progress on this front can be made by routine sharing of the complete gene lists that define individual cell clusters, rather than the standard reporting of the top 5 differentially expressed genes. Moreover, anatomical (e.g., interfascicular, fascicular, endotenon) and functional (matrix producing) descriptors could also be useful tools in annotating tendon cell sub-populations, particularly within the context of the highly organized topography of tendon tissues. Spatial ‘omics datasets will be particularly helpful in clarifying the anatomical localization of tendon sub-populations in tissue sections.

Long-term, working groups or collaborative efforts to integrate multiple datasets either within a given species, or across species will yield the most impactful understanding of the diversity of the tendon environment(s), and the tendon response to aging, injury, or other insults. While some level of community-driven efforts can happen organically, there is a clear need for large-scale funding opportunities to support consortium efforts given the magnitude and complexity of this task on both the computational and biological fronts, with recent work highlighting the challenges and opportunities of this work [19]. Finally, while it is critical to understand the cellular diversity within discrete tendon states, siloing tendon cell biology would be a missed opportunity to understand how tendon cell biology intersects with other tissues and disease contexts (e.g., tissue development, growth,

degeneration, or fibrosis) in order to both inform and be informed by broader tissue atlasing efforts, further highlighting the importance of adopting broadly-used reporting guidelines such as those from the HCA.

As a final example of the important overlap between the pillars of data sharing, reporting, and annotation, there was robust discussion on the value of developing computational bioinformatics expertise within the tendon community. Initiatives that support knowledge gain and transfer within the tendon community will train the next generation of tendon ‘omics experts, strengthening this skillset within the tendon community. Indeed, the sharing of tendon NGS datasets will also be a helpful training resource. Amongst the ways this could be accomplished include an ORS Pre-Course, or online training modules as part of the ORS education platform, LearnORS.

In closing, the tendon community survey, plenary session, and breakout sessions highlighted the tremendous energy and opportunities within and for the tendon community related to ‘omics. We hope this article serves as a roadmap and open-invitation to cultivate community-driven work to enhance the rigor, translational potential, and impact of the tendon field.

References

1. Ackerman, J.E., et al., *Defining the spatial-molecular map of fibrotic tendon healing and the drivers of Scleraxis-lineage cell fate and function.* Cell Rep, 2022. **41**(8): p. 111706.
2. Korcari, A., et al., *Scleraxis-lineage cells are required for tendon homeostasis and their depletion induces an accelerated extracellular matrix aging phenotype.* Elife, 2023. **12**.
3. Nichols, A.E.C., et al., *Epitenon-Derived Progenitors Drive Fibrosis and Regeneration After Flexor Tendon Injury in a Spatially-Dependent Manner.* Nat Commun, 2025.
4. De Micheli, A.J., et al., *Single-cell transcriptomic analysis identifies extensive heterogeneity in the cellular composition of mouse Achilles tendons.* Am J Physiol Cell Physiol, 2020. **319**(5): p. C885-C894.
5. Mimpen, J.Y., et al., *Single nucleus and spatial transcriptomic profiling of healthy human hamstring tendon.* FASEB J, 2024. **38**(10): p. e23629.
6. Akbar, M., et al., *Single cell and spatial transcriptomics in human tendon disease indicate dysregulated immune homeostasis.* Ann Rheum Dis, 2021. **80**(11): p. 1494-1497.
7. Ng, M.T.H., et al., *A single cell atlas of frozen shoulder capsule identifies features associated with inflammatory fibrosis resolution.* Nat Commun, 2024. **15**(1): p. 1394.
8. Thorpe, C.T., et al., *Anatomical heterogeneity of tendon: Fascicular and interfascicular tendon compartments have distinct proteomic composition.* Sci Rep, 2016. **6**: p. 20455.
9. Peffers, M.J., et al., *Proteomic analysis reveals age-related changes in tendon matrix composition, with age- and injury-specific matrix fragmentation.* J Biol Chem, 2014. **289**(37): p. 25867-78.
10. Kendal, A.R., et al., *Multi-omic single cell analysis resolves novel stromal cell populations in healthy and diseased human tendon.* Sci Rep, 2020. **10**(1): p. 13939.
11. Choi, H., et al., *Heterogeneity of proteome dynamics between connective tissue phases of adult tendon.* Elife, 2020. **9**.
12. Baglaenko, Y., et al., *Making inroads to precision medicine for the treatment of autoimmune diseases: Harnessing genomic studies to better diagnose and treat complex disorders.* Camb Prism Precis Med, 2023. **1**: p. e25.
13. Korsunsky, I., et al., *Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases.* Med, 2022. **3**(7): p. 481-518 e14.
14. Zhang, F., et al., *Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes.* Nature, 2023. **623**(7987): p. 616-624.
15. Kiel, D.P., et al., *The Musculoskeletal Knowledge Portal: Making Omics Data Useful to the Broader Scientific Community.* J Bone Miner Res, 2020. **35**(9): p. 1626-1633.
16. Gaddis, N., et al., *LungMAP Portal Ecosystem: Systems-level Exploration of the Lung.* Am J Respir Cell Mol Biol, 2024. **70**(2): p. 129-139.
17. Callaway, E.M., et al., *A multimodal cell census and atlas of the mammalian primary motor cortex.* Nature, 2021. **598**(7879): p. 86-102.
18. Buechler, M.B., et al., *Cross-tissue organization of the fibroblast lineage.* Nature, 2021. **593**(7860): p. 575-579.
19. Hemberg, M., et al., *Insights, opportunities, and challenges provided by large cell atlases.* Genome Biol, 2025. **26**(1): p. 358.

FIGURES

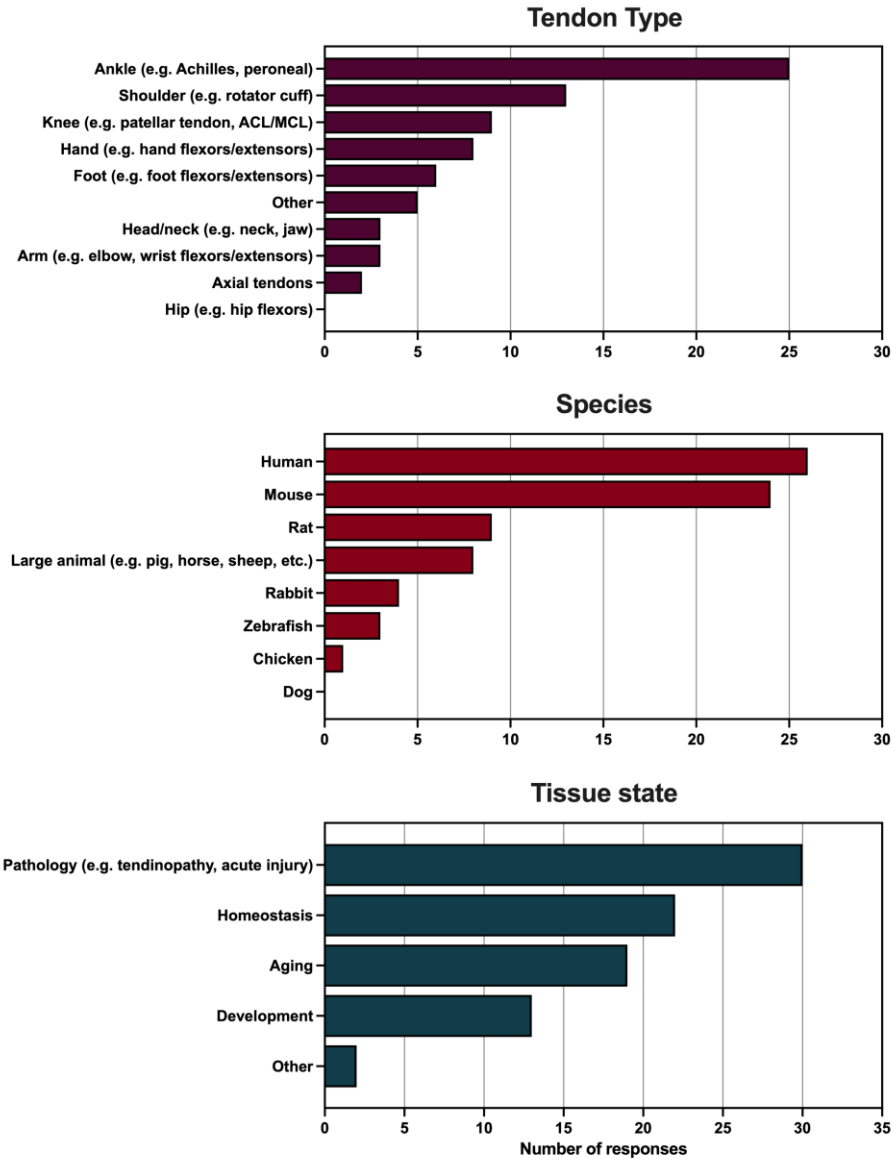

Figure 1. Attributes of experimental systems used in transcriptomic studies. Survey participants who responded “Yes” to the question “Do you currently produce or plan to produce transcriptomic datasets?” (37 total) were asked to report the tendons (by general anatomic site), species, and tissue states for transcriptomic studies in their research.

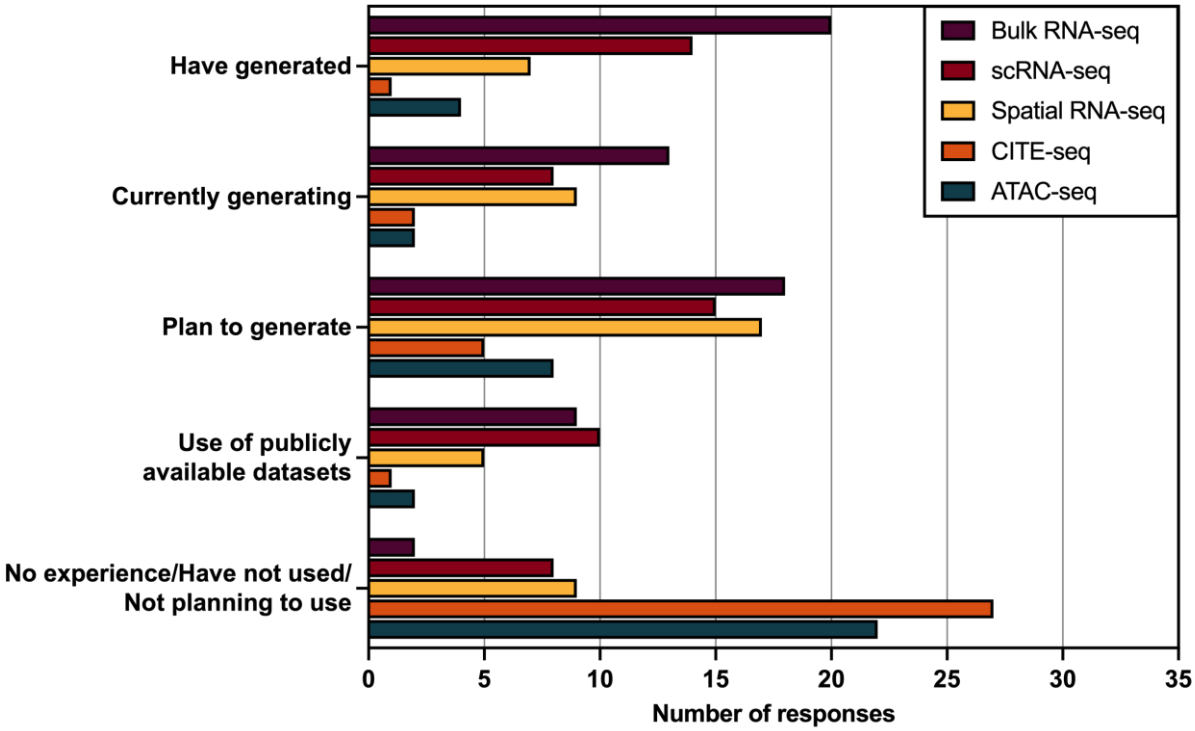

Figure 2. Reported use of transcriptomic technologies by the tendon community. Survey participants who responded “Yes” to the question “Do you currently produce or plan to produce transcriptomic datasets?” (37 total) were asked to report their current and planned use of the following transcriptomic techniques: bulk RNA-sequencing, single cell RNA-sequencing (scRNA-seq), spatial RNA-sequencing, Cellular Indexing of Transcriptomes and Epitopes (CITE)-seq, and Assay for Transposase-Accessible Chromatin (ATAC)-seq (238 total responses).