The recent boom in high-throughput "–omics" technologies has revolutionized biomedical research. The suffix –omics is an au courant term in biomedical research meaning ‘a study of the totality of something’ and connotes a well-equipped toolbox for the comprehensive assessment of entire genomes and their products. It embraces several disciplines such as epigenomics, genomics, transcriptomics, proteomics, metabolomics, microbiomics, glycomics, expressomics and interactomics. These specific platforms are capable of interrogating transcripts, genes, proteins and metabolites and their interactions in a systematic high-throughput manner. It is beyond the scope of this Editorial to cover the entire plethora of –omics at reasonable depth given their complexity and diversity. Therefore, I intend to highlight here the basic concepts, opportunities and critical barriers in undertaking transcriptomics studies in musculoskeletal research with a special focus on RNA sequencing (RNA-seq) and to some extent single cell (sc) RNA-seq. I also intend to summarize the recent contributions of these technologies in defining osteoarthritic phenotypes in various contexts.

Transcriptomics refers to the methodologies used to study the transcriptome of an organism or a cell, i.e., the sum of all of its tens of thousands of RNA transcripts. In biology, the term transcript refers to a fragment of RNA or DNA, which has been transcribed from a DNA or RNA template. Transcript is often used interchangeably, and erroneously, with gene. Traditionally, “gene” refers largely to a genomic region along the chromosomal DNA producing a polyadenylated mRNA that encodes a protein. This gene may code for one or more alternative splice isof orm transcripts that collectively make up the parent gene. Transcript may also refer to non-coding RNA’s that are not polyadenylated or translated into proteins but are still biologically active and often serve regulatory functions. Each cell in the human body contains about 25,000 protein-coding genes and 34,214 transcripts [1]. Therefore, the outcome of transcriptomics studies such as RNA microarrays and bulk RNA-seq represents the expression of transcripts but not necessarily genes.

Currently, we have a limited understanding of the etiology and mechanisms of numerous musculoskeletal conditions such as osteoarthritis and osteoporosis, and processes such as degeneration, inflammation, repair, and healing. Transcriptomics analyses provide important initial insights into understanding these disease processes and offer strategies for deep phenotyping and stratification of a heterogeneous population of patients with osteoarthritis [2]. Hence, there is a high demand and interest in transcriptomics studies among the musculoskeletal community. With dynamic progress in transcriptomics platforms, it is desirable to keep ourselves abreast of the fast pace of developments in these technologies and wisely choose the most relevant and up to date platform. For instance, microarrays that pioneered the analysis of the entire transcriptome are now considered ‘obsolete’ and have been replaced with bulk RNA-seq. While there is a strong concordance between the two [3], RNA-seq has several advantages over microarrays [4]. Since it is based on sequencing of cDNA libraries prepared from mRNA and counting reads that map to each gene, it is more sensitive and precise than hybridization-based microarrays. RNA-seq provides an improved dynamic range for expression level quantification and improved gene sequence information down to a single base resolution. Moreover, RNA-seq can detect gene splicing (transcript variants), mutations, gene alternative initiation and termination mapping, and novel transcript identification. It also provides more accurate measurement of various transcript biotypes such as microRNAs, IncRNAs, lincRNAs, and snoRNAs.

RNA-seq analysis provides a number of opportunities to better understand musculoskeletal biology and pathology and has already identified new genes and pathways active in many musculoskeletal tissues such as articular cartilage [5], meniscus [6], ligament [7], tendon [8,9], skeletal muscle [10], bone [11] and synovium [12]. A majority of RNA-seq studies are based on an a priori hypothesis to discover prognostic markers of a disease or a process rather than a fishing expedition as they may be viewed at times. The data obtained from RNA-seq allow researchers to gain a deeper insight into what constitutes a specific cell type, how that cell functions, and how changes in the normal level of gene expression reflects or contributes to a disease. RNA-seq is a robust and powerful tool with unmatched reproducibility to provide a snapshot of a specific time period in the cell, tissue, or organ. While transcriptomics analyses reveal elusive details, there are a number of challenges that hamper successful execution of such studies and fall under three major categories: technical, statistical or bioinformatical, and biological or interpretational.

Technical challenges are generally related to sample preparation and handling. The integrity of RNA is a critical criterion for acquiring high quality data. The common notion “garbage in, garbage out” perfectly applies to this scenario. Each mammalian cell contains approximately 10–30 pg of total RNA depending on cell type and cell state. Cells in musculoskeletal tissues are in a quiescent state unless injured or stimulated. Since healthy tissues are rarely accessible from patients, RNA extraction from injured and/or degenerated tissues is challenging owing to the amount of residual tissue and whether or not the tissue is scarred, fibrotic or otherwise contaminated. Patient-derived specimens need to be collected in RNase-free containers containing RNA stabilization solution (e.g., RNalater) and should be transported to the laboratory in a considerably short period of time under refrigerated conditions. In addition, most musculoskeletal tissues have dense matrix and fewer cells posing a challenge in terms of both amount (due to cell scarcity) and purity (owing to dense matrix with high protein content and enzymatic
degradative activities) of total RNA. To overcome these issues, researchers isolate cells from tissues for expansion in culture to wash away any effects of medication and to gain higher yields of RNA. Unfortunately, this does not represent the true snapshot of a tissue as enzymatic digestion influences transcript expression. New platforms now use low amounts of RNA and even (somewhat) degraded RNA, providing more opportunity to use RNA-seq. While this is encouraging, the samples within a batch need to be of similar degradation or within same range of RNA integrity number to avoid biases related to RNA integrity. Apart from that, RNA integrity number should be added as confounding factor in the analysis. Low quality RNA generates a lot of noise in the data.

A statistical or bioinformatical challenge pertains to the nature and amount of the raw data generated in transcriptomics experiments. RNA-seq data are complex, extensive, and in a format that needs to be carefully processed. For this, various critical steps are needed for visualization, quality control (QC), and streamlining of the data prior to their reporting and interpretation. With the introduction of new platforms, there is also a rapid surge in the development of statistical and bioinformatical software and pipelines for efficient, precise, and swift analysis. QC steps generally cover these parameters: normalization, the total number of reads sequenced, GC content, base quality scores, alignment end bias, splice junction saturation, and gene biotypes detected (protein-coding, lncRNAs, miRNAs, lincRNAs, snoRNAs, pseudogenes, and processed transcripts) [13]. During QC, RNA-seq data are visualized to appreciate distinct grouping of samples (principal component analysis), in presenting individual genes in each sample in terms of both fold-change and/or P values (heatmaps, unsupervised and unbiased clustering and volcano plots). The amount of data generated can be overwhelming, and these ‘big data’ require a considerable amount of storage space on servers or on the cloud. In addition, the raw and, in some cases, slightly processed data should be deposited in a data repository such as Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/). The purpose is to allow access to other investigators to retrieve, analyze, and compare their findings with published work as well as to foster cross-disciplinary collaboration.

Interpretation of large RNA-seq datasets is indeed a daunting task. While these datasets often yield positive, relevant and novel findings, the findings are often inadvertently over-interpreted. RNA-seq differential expression analysis provides a list of transcripts that are differentially expressed between two experimental conditions which can then be evaluated further with gene set analysis. Gene sets such as Gene Ontology (GO) biological processes and disease pathways are based on prior literature from musculoskeletal and non-musculoskeletal disciplines such as cancer, immunology and developmental biology. Because these gene sets are created through known biological or statistical associations in varying biological contexts, gene sets are not necessarily specific to the biological context of a given experiment and often yield significant results that are out of context or are only loosely affiliated with the known biological context of a given experiment. It is therefore not desirable to infer that transcriptomics data provide “evidence” for the involvement of genes and pathways but rather it informs about their differential expression and association to known pathways or processes. Moreover, transcriptomics studies only inform about mRNA (and other RNA types), not proteins or metabolites, so conclusions cannot be generalized and should be limited to the transcript. Inferences on proteins and metabolites would require further studies with their own respective “omics” and further mechanistic studies are often needed to determine the role of genes, proteins and pathways in disease processes. Conversely, transcriptomics data are sometimes under interpreted as well, for example, when the focus is placed on the number of differentially expressed transcripts or on their magnitude of expression instead of the function of individual transcripts. In certain cases, data are not analyzed correctly when confounding factors such as age, sex, activity level, chronicity of injury and stage of the disease or stage of the healing are not included in the statistical model. Lastly, RNA-seq provides information not only on mRNA expression but also about mutations, transcript variants as well as on miRNA and lncRNAs, thus elucidating the epigenetic landscape of the genome.

RNA-seq studies have provided important insights into disease pathogenesis of osteoarthritis in human knee or hip joint tissues (Table 1). These studies have identified new pathways and biological processes in cartilage, synovium and meniscus. Similarly, RNA-seq analysis of murine models of osteoarthritis or osteoarthritis susceptibility in chondrocytes or in the entire joint revealed important information on genetic heterogeneity, epigenetics, metabolism, disease severity and aging (Table 2).

Bulk RNA-seq reveals the transcriptomics profiles of all cells in a sample but obscures fine scale single cell level differences. To gain insights at a single cell resolution, scRNA-seq is employed, which provides this information at the resolution of a single cell in addition to identifying new cell types. The total number of cells in the human body is about 37.2 trillion [29], comprising 411 distinct types [30]. It is the power of scRNA-seq to measure the expression of thousands of transcripts at the resolution of a single cell, in addition to cataloguing cells based on their dynamic transcript profile from a heterogeneous population in a tissue.

### Table 1

| Tissue | Comparison (sample size) | Gene ontology/Pathway | Reference |
|--------|--------------------------|-----------------------|-----------|
| Cartilage | Normal (n = 18) vs. OA (n = 20) | Early growth response genes; transcription factors | [14] |
| | NOF (n = 6) vs. OA (n = 6) | Long intergenic non-coding RNAs | [15] |
| | Non-OA (n = 10) vs. OA (n = 44) | Mechanoreceptors; calcium signaling; ion channels; cytoskeletal organizers | [16] |
| | intact OA cartilage (n = 8) vs. damaged OA cartilage (n = 8) | Non-chondrogenic response of altered matrix protein and secreted proteinase expression | [17] |
| | APM (n = 10) vs. OA (n = 10) | Mesenchymal cell apoptosis; epithelial morphogenesis; canonical glycolysis; extracellular matrix organization; cartilage development; glucose catabolic process | [5] |
| Synovium | Low OA pain (n = 5) vs. High OA pain (n = 5) | Neuronal survival under cellular stress; calcium-dependent synaptic exocytosis; pain in synovial microenvironment | [18] |
| | Failed MoM (n = 6) vs. THA (n = 5) | Inflammation; redox homeostasis; metal ion binding and transport; macrophage activation; apoptosis | [19] |
| | Failed MoM (n = 16) vs. THA (n = 16) | Immune response; extracellular matrix remodeling; epidermal growth factor receptor signaling pathway | [20] |
| | Early RA, OA and others (multiple groups) | Leukocyte activation; T cell activation; B cell mediated immunity; osteoblast differentiation; bone remodeling | [21] |
| Meniscus | APM (n = 12) vs. OA (n = 12) | Response to external stimuli; cell migration and localization; histone deacetylase activity; skeletal development | [6] |
| Facet joint tissues | Moderate degeneration (n = 20) vs. severe degeneration (n = 28) | Wnt signaling; NFκB signaling | [22] |

OA = osteoarthritis; NOF = neck of femur fracture; APM = arthroscopic partial meniscectomy; MoM = metal on plastic; THA = total hip arthroplasty; MoM = metal on metal; RA = rheumatoid arthritis; Wnt = Wntless and the name Int-1; NFκB = nuclear factor kappa-light-chain-enhancer of activated B cells.

---
of rheumatoid arthritis pathogenesis [43]. A study reported that synovial fibroblasts revealed a transition pattern in rheumatoid arthritis, which is akin to osteoarthritis suggesting an overlap in the pathogenetic mechanism between the two diseases [44]. Further work in this area is sorely needed to elucidate the reciprocal contributions of all joint tissues in conjunction with other factors such as age, obesity, injury, and sex.

In summary, transcriptomics platforms have become indispensable tools for biomedical research, and are gaining momentum in the musculoskeletal field. Technological innovations in transcriptomics have progressed from characterizing bulk tissue or cells to single cells and nuclei. Aforementioned critical barriers have to be overcome in order to make use of these powerful technological innovations in economical and practical manners and to generate reliable data for cell subtypes and disease phenotyping. The quickening technological advances in resolution and sensitivity of ‘omics’ have fueled new discoveries in cell types for further research and hopefully will provide new therapeutic targets for complex diseases such as osteoarthritis in the near future.

Author contribution

I am the sole contributor.

Declaration of competing interest

None declared.

Acknowledgments

I acknowledge with thanks Eric D. Tycksen (Genome Technology Access Center, McDonell Genome, Institute, Washington University School of Medicine, St. Louis, MO, USA) for providing important technical input.

References

[1] International Human Genome Sequencing C, Finishing the euchromatic sequence of the human genome, Nature 431 (2004) 931–945.
[2] A. Mobasher, M. Kapoor, S.A. Ali, A. Lang, H. Madhy, The future of deep phenotyping in osteoarthritis: how can high throughput omics technologies advance our understanding of the cellular and molecular taxonomy of the disease? Osteoarthritis Cartilage Open (2021) https://doi.org/10.1016/j.joarc.2021.100144.
[3] M.F. Rai, E.D. Tycksen, L.J. Sandell, R.H. Brophy, Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears, J. Orthop. Res. 36 (2018) 484–497.
[4] R. Hrdlickova, M. Toloue, B. Tian, RNA-seq Methods for Transcriptome Analysis, vol. 6, Wiley Interdiscip Rev RNA, 2017.
[5] M.F. Rai, E.D. Tycksen, L. Cai, J. Yu, R.W. Wright, R.H. Brophy, Distinct degenerative phenotype of articular cartilage from knees with meniscus tear compared to knees with osteoarthritis, Osteoarthritis Cartilage 27 (2019) 945–955.
[6] R.H. Brophy, B. Zhang, L. Cai, R.W. Wright, L.J. Sandell, M.F. Rai, Transcriptome comparison of meniscus from patients with and without osteoarthritis, Osteoarthritis Cartilage 26 (2018) 422–432.
[7] R.H. Brophy, E.D. Tycksen, L.J. Sandell, M.F. Rai, Changes in transcriptome-wide gene expression of anterior cruciate ligament tears based on time from injury, Am. J. Sports Med. 44 (2016) 2064–2075.
[8] Z. Ge, H. Tang, J. Lyu, B. Zhou, M. Yang, K. Tang, et al., Conjoint analysis of IncRNA and mRNA expression in rotator cuff tendinopathy, Ann. Transl. Med. 8 (2020) 335.
[9] M.J. Pfeffer, Y. Fang, K. Cheung, T.K. Wei, P.D. Clegg, H.L. Birch, Transcriptome analysis of ageing in uninjured human Achilles tendon, Arthritis Res. Ther. 17 (2015) 33.
[10] M. Llanos-Diez, W. Fury, H. Okamoto, Y. Bai, J. Gromada, L. Larsson, RNA-sequencing reveals altered skeletal muscle contraction, E3 ligases, autophagy, apoptosis, and chaperone expression in patients with critical illness myopathy, Skeletal Muscle 9 (2019) 9.
[11] B.A. Goates, J.A. McKenzie, E.G. Buettmann, X. Liu, P.M. Gontarz, B. Zhang, et al., Transcriptional profiling of intramembranous and endochondral ossification after fracture in mice, Bone 127 (2019) 577–591.
[12] H.L. Carr, J.D. Turner, T. Major, D. Scheel-Toeller, A. Filer, New developments in transcriptomic analysis of synovial tissue, Front. Med. 7 (2020) 21.
[13] Q. Sheng, K. Vickers, S. Zhao, J. Wang, D.C. Samuels, O. Koues, et al., Multi-perspective quality control of Illumina RNA sequencing data analysis, Brief Funct. Genomics 16 (2017) 194–204.
[14] K.M. Fisch, R. Gamini, O. Alvarez-Garcia, R. Akagi, M. Saito, Y. Muramatsu, et al., Identification of transcription factors responsible for dysregulated networks in
human osteoarthritis cartilage by global gene expression analysis, Osteoarthritis Cartilage 26 (2018) 1531–1536.

[15] B. Ajegbile, K. Cheng, V. Xu, A.J. Skelton, A. Panagiotopoulos, J. Soul, et al., Identification of long non-coding RNAs expressed in knee and hip osteoarthritic cartilage, Osteoarthritis Cartilage (2019).

[16] J. Soul, S.L. Dunn, S. Anand, X. Duan, J. Dai, D. Yu, Y. Wang, Y. Chen, H. Sun, X. Zhang, et al., Kdm6b regulates cartilage homeostasis by altering metabolism, JCI Insight 2 (2017).

[17] J. Soul, S.L. Dunn, S. Anand, F. Serracino-Inglott, J.M. Schwartz, R.P. Boot-Handford, et al., Striatification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage, Ann. Rheum. Dis. 77 (2018) 423.

[18] S.L. Dunn, J. Soul, S. Anand, J.M. Schwartz, R.P. Boot-Handford, T.E. Hardingham, Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses, Osteoarthritis Cartilage 24 (2016) 1431–1440.

[19] A. Bratus-Neuenschwander, F. Castro-Giner, M. Frank-Bertoncelj, S. Alur, S.F. Fucentese, R. Schlapbach, et al., Pain-associated transcriptome changes in synovium of knee osteoarthritis patients, Genes 9 (2018).

[20] A. Pemmari, T. Leppanen, E.L. Paukkeri, A. Eskelinen, T. Moilanen, E. Moilanen, C.G. Salib, E.A. Lewallen, C.R. Paradise, M.E. Tibbo, J.X. Robin, W.H. Trousdale, et al., Comparative transcriptomics identifies age-related differences in knee joint osteoarthritis, Ann. Rheum. Dis. 77 (2018) 423.

[21] A.M. Mandelin 2nd, P.J. Homan, A.M. Shaffer, C.M. Cuda, S.T. Dominguez, J. Keefe, et al., RNA-seq reveals the progression of human osteoarthritis, Ann. Rheum. Dis. 78 (2019) 100–110.

[22] M.F. Rai, Osteoarthritis and Cartilage Open 3 (2021) 100166

[23] M. Attur, X. Duan, L. Cai, T. Han, W. Zhang, E.D. Tycksen, et al., Periostin loss-of-function protects mice from post-traumatic and age-related osteoarthritis, Arthritis Res. Ther. x (2021) 15.

[24] A. Sebastian, D.K. Muruges, M.E. Mendez, N.R. Hum, N.D. Rina-Arce, J.L. McCool, et al., Global gene expression analysis identifies age-related differences in knee joint transcriptome during the development of post-traumatic osteoarthritis in mice, Int. J. Mol. Sci. (2020) 21.

[25] A. Sebastian, J.C. Chang, M.E. Mendez, D.K. Muruges, S. Hatsell, A.N. Economides, et al., Comparative transcriptomics identifies novel genes and pathways involved in post-traumatic osteoarthritis development and progression, Int. J. Mol. Sci. 19 (2018).

[26] E. Bianconi, A. Piovesan, F. Facchin, A. Berardi, R. Casadei, F. Frabetti, et al., An estimation of the number of cells in the human body, Ann. Hum. Biol. 40 (2013) 463–471.