



Database Tool

MuscleAtlasExplorer: a web service for studying gene expression in human skeletal muscle

Olof Asplund^{1,*}, Johan Rung², Leif Groop^{1,3}, Rashmi Prasad B¹
and Ola Hansson^{1,3}

¹Genomics, Diabetes and Endocrinology Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Jan Waldenströms gata 35, Malmö 20502, Sweden, ²SciLifeLab, BMC, Husargatan 3, Uppsala University, Uppsala 751 22, Sweden and ³Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Tukholmankatu 800290 Helsinki, Finland

*Corresponding author: Olof Asplund. Phone: +46 40 39 12 27; Fax: +46 40 39 12 22; Email: Olof.Asplund@med.lu.se

Citation details: Asplund, O., Rung, J., Groop, L., *et al.* MuscleAtlasExplorer: a web service for studying gene expression in human skeletal muscle. *Database* (2020) Vol. 2020: article ID baaa111; doi:10.1093/database/baaa111

Received 3 July 2020; Revised 13 October 2020; Accepted 14 December 2020

Abstract

MuscleAtlasExplorer is a freely available web application that allows for the exploration of gene expression data from human skeletal muscle. It draws from an extensive publicly available dataset of 1654 skeletal muscle expression microarray samples. Detailed, manually curated, patient phenotype data, with information such as age, sex, BMI and disease status, are combined with skeletal muscle gene expression to provide insights into gene function in skeletal muscle. It aims to facilitate easy exploration of the data using powerful data visualization functions, while allowing for sample selection, in-depth inspection and further analysis using external tools.

Availability: MuscleAtlasExplorer is available at <https://mae.crc.med.lu.se/mae2>.

Introduction

Microarrays have been widely used for analyzing gene expression in biological samples. More recently, RNA sequencing (RNA-Seq) has become the prevailing technology, but is still limited by high running costs compared to array-based platforms.

Since the introduction of microarrays more than 20 years ago (1), a large amount of raw data have been archived in public databases. As of January 2020, there are almost three times more human gene expression samples from microarrays (970 530) in the Gene Expression Omnibus than from RNA-Seq data (333 074). Combination of these data across studies would allow for the analysis of a very large number of samples without any new experimental cost. A prerequisite is, of course, that the

phenotype data is highly structured and detailed. To enable this type of analysis, Su *et al.* (2) manually curated the phenotypic information from 44 microarray studies of human skeletal muscle, allowing analysis of all samples as a single dataset. This dataset (European Bioinformatics Institute (EBI) project E-MTAB-1788) (3) contains expression data from 19 603 genes from a total of 1654 samples connected with 40 different phenotypes.

The dataset has already been utilized in several ways (4–24). Here we make it even more accessible by the creation of MuscleAtlasExplorer (MAE). This web service facilitates fast lookups of gene expression using a simple and powerful interface, which allows for analyses that can go far beyond the scope of the original article.

Materials and methods

Data preprocessing

The dataset consists of two parts. The patient metadata contains the phenotypes of individuals, such as age, gender and BMI, as well as experimental protocol information. To make the application easy to use, we have limited the available metadata to a set of 16 variables deemed to be non-redundant and particularly relevant.

The expression data contain information on the relative amount of expression from a single normalized analysis for each sample and gene (2). The curated dataset has data from two major microarray platforms, i.e. Affymetrix ‘HGU133Plus2’ and ‘HGU133-A’. The results from each platform are plotted and presented separately to avoid platform-specific biases.

Web interface

Computations and plotting are performed using R statistical software (25) and ggplot2 package (26). The Shiny framework (27) provides a graphical interface in the browser of the user.

Results

The application has four major parts: (i) a gene lookup view that gives a quick overview of a chosen gene, (ii) a phenotype view that focuses on phenotype data, (iii) a gene-centric view that focuses on the gene expression data and (iv) an association view with focus on associations between phenotypes and gene expression.

Gene lookup

The user can search for information regarding a specific gene in the Gene Lookup view. Upon selection of the gene, a short summary of associations between phenotypes and gene expression is shown for each phenotype including plots. The gene glycogen synthase 1 is shown as an example in Figure 1A.

Phenotypes

MAE has several functions to characterize and select patients. The Phenotype tab allows users to filter patients based on phenotype. Filters applied in this tab are used in most other parts of the web application. The table of patients can be exported as a comma-separated file, for further analysis using other programs.

The ‘visualize samples’ tab allows the user to investigate phenotypes in more detail. The phenotypes can be visualized either using a table or using different plotting functions.

The ‘Custom Plot’ (Figure 1B) function allows the user to plot any phenotype.

Two different phenotypes can also be plotted in relation to each other. This allows for more detailed analysis. In addition, each data point can be colored depending on a third phenotype.

Genes

The ‘Genes’ view show data on gene expression. Users can search for genes of interest using symbols, Ensembl IDs or trivial names. Clicking a gene in this table will select that gene for plotting under the ‘Gene plots’ tab. Only patients that have been previously selected under the ‘Samples’ view are shown.

Associations

The ‘Associations’ view allows users to explore the relationship between gene expression and phenotypes. Five linear models have been precomputed, using a similar methodology as in Su *et al.* (2), to identify genes that are differentially expressed in relation to age, gender, type 2 diabetes status, acute training and BMI.

Discussion

Data repositories such as Gene Expression Omnibus (28,29), ArrayExpress (30,31) and Sequence Read Archive (32) provide massive amounts of raw data, but the phenotypes are not easy to standardize. As a result, these services have limited analysis capabilities within the websites. Services such as Genotype Tissue Expression project (GTEx) (33) and Ensembl (34) provide tools for analysis, but do not contain detailed phenotype data. Skeletal muscle-specific tools and datasets are also available. Examples of this include MetaMEX (35), which gives detailed information regarding gene expression in association with intervention studies as well as gene-gene correlations in skeletal muscle. NeuroMuscleDB (36) provides a database of genes active during different stages of muscular development, and SysMyo (37) provides gene sets of genes associated with a vast variety of neuromuscular conditions and experimental intervention studies. We hope that MAE will work in a complementary manner for these datasets, providing in-detail visualization of skeletal muscle gene expression. For this dataset (2), we have leveraged the highly structured metadata to provide information on associations between gene expression and phenotypes. This way, we hope to augment the wealth of information from other databases with functional association data in regard to skeletal muscle gene expression.

The Shiny framework was selected for development because it allows the developer to leverage the powerful

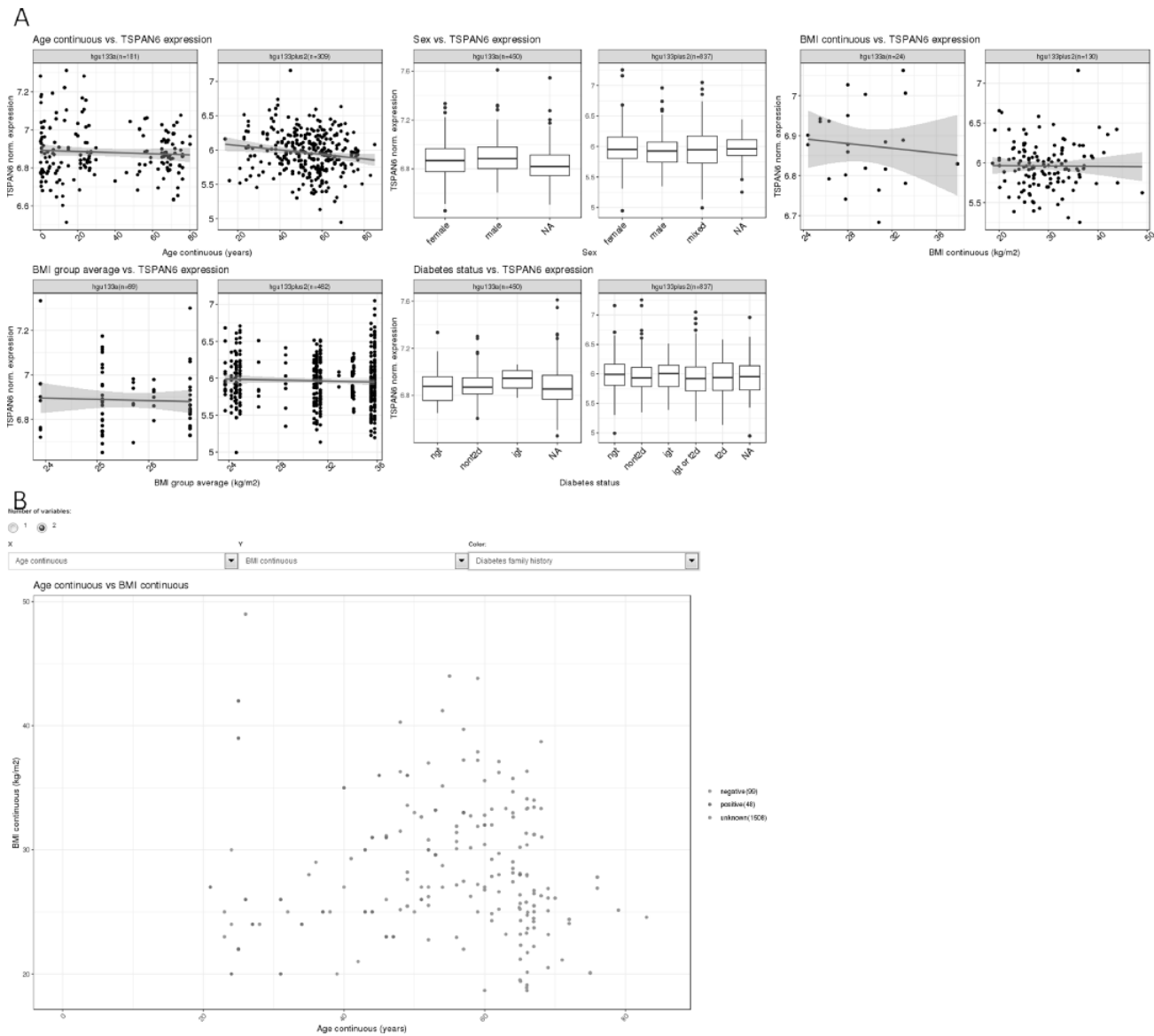


Figure 1. (A) Example of the gene summary view, showing expression data for the gene H3 histone, family 3B (H3F3B) in relation to important phenotypes. (B) One example of two-variable plots produced by MuscleAtlasExplorer, i.e. scatter plots for two different continuous variables (exemplified by age and BMI, with coloring set to signify diabetes family history).

plotting and processing capabilities of R, while providing an easy-to-use interface.

MAE bases its plots only on microarray gene expression data. In contrast, in the modern day, high-throughput sequencing is the prevailing technology for gene expression analysis. There are some technical disadvantages to microarrays for gene expression analysis. For instance, the restriction of gene coverage to the targeted probe regions can decrease the sensitivity to detect certain genes. Furthermore, the nature of microarray technology can decrease the sensitivity to detect genes with low expression. On the other hand, the much lower cost of microarrays and vast

availability of open datasets increase the power to detect differentially expressed genes.

It is expected that users will need to extend the analysis using external software. For this reason, we have included functions to export the relevant data. The openly available source code, available at <https://github.com/olof-a-bio/muscle-atlas-explorer>, allows users to extend and modify the plotting and analysis functions in the web application for their own use.

We thereby consider the MAE as a new resource to facilitate research needing information on gene expression in human skeletal muscle.

Acknowledgements

We thank Science for Life Laboratory (SciLifeLab) for their technical and scientific support.

Funding

This study was supported by the Swedish Research Council (Dnr 2018-02635), Strategic Research Area Exodiab (Dnr 2009-1039), project grant (Dnr 2018-02635) and Linnaeus grant (Dnr 349-2006-237); the Swedish Foundation for Strategic Research (Dnr IRC15-0067); Crafoord foundation; Wallenberg foundation; Novo Nordisk foundation; Pahlsson foundation; Diabetes Wellness and Swedish Diabetes Research Foundation.

Conflict of Interest

The authors declare no competing interests.

References

- Lockhart, D.J., Dong, H., Byrne, M.C., *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.
- Su, J., Ekman, C., Oskolkov, N., *et al.* (2015) A novel atlas of gene expression in human skeletal muscle reveals molecular changes associated with aging. *Skelet Muscle*, **5**, 35.
- E-MTAB-1788 < Browse < ArrayExpress < EMBL-EBI. <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1788/> (12 June 2018, date last accessed).
- Mielcarek, M. and Isalan, M. (2015) A shared mechanism of muscle wasting in cancer and Huntington's disease. *Clin. Transl. Med.*, **4**, 34.
- Yu, H., You, X., Li, J., *et al.* (2016) Genome-wide mapping of growth-related quantitative trait loci in orange-spotted grouper (*Epinephelus coioides*) using double digest restriction-site associated DNA sequencing (ddRADseq). *Int. J. Mol. Sci.*, **17**, 501.
- Rudolf, R., Deschenes, M.R., Sandri, M., *et al.* (2016) Neuromuscular junction degeneration in muscle wasting. *Curr. Opin. Clin. Nutr. Metab. Care*, **19**, 177–181.
- Pollard, A., Shephard, F., Freed, J., *et al.* (2016) Mitochondrial proteomic profiling reveals increased carbonic anhydrase II in aging and neurodegeneration. *Aging*, **8**, 2425–2436.
- Ren, Y.-Y., Koch, L.G., Britton, S.L. *et al.* (2016) Selection-, age-, and exercise-dependence of skeletal muscle gene expression patterns in a rat model of metabolic fitness. *Physiol. Genomics*, **48**, 816–825.
- Huffman, K.M., Jessee, R., Andonian, B. *et al.* (2017) Molecular alterations in skeletal muscle in rheumatoid arthritis are related to disease activity, physical inactivity, and disability. *Arthritis Res. Ther.*, **19**, 12.
- Gonzalez-Freire, M., Semba, R.D., Ubaida-Mohien, C. *et al.* (2016) The Human Skeletal Muscle Proteome Project: a reappraisal of the current literature. *J. Cachexia Sarcopenia Muscle*, **8**, 5–18.
- Marck, A., Berthelot, G., Foulonneau, V. *et al.* Age-related changes in locomotor performance reveal a similar pattern for *Caenorhabditis elegans*, *Mus domesticus*, *Canis familiaris*, *Equus caballus*, and *Homo sapiens*. *J. Gerontol. Ser. A*. <https://academic.oup.com/biomedgerontology/article/doi/10.1093/gerona/glw136/2630042/Age-Related-Changes-in-Locomotor-Performance> (26 January 2017, date last accessed).
- Cutler, A.A., Dammer, E.B., Doung, D.M. *et al.* (2017) Biochemical isolation of myonuclei employed to define changes to the myonuclear proteome that occur with aging. *Aging Cell*, **16**, 738–749.
- Ingenbleek, Y. (2017) Lean body mass harbors sensing mechanisms that allow safeguarding of methionine homeostasis. *Nutrients*, **9**, 1–13.
- Pollack, R.M., Barzilai, N., Anghel, V. *et al.* (2017) Resveratrol improves vascular function and mitochondrial number but not glucose metabolism in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.*, **72**, 1703–1709.
- Gonzalez-Freire, M., Scalzo, P., D'Agostino, J. *et al.* (2018) Skeletal muscle ex vivo mitochondrial respiration parallels decline in vivo oxidative capacity, cardiorespiratory fitness, and muscle strength: the Baltimore longitudinal study of aging. *Aging Cell*, **17**, 1–11.
- Han, A., Bokshan, S.L., Marcaccio, S.E. *et al.* (2018) Diagnostic criteria and clinical outcomes in sarcopenia research: a literature review. *J. Clin. Med.*, **7**, 70.
- D'Amico, D., Mottis, A., Potenza, F. *et al.* (2019) The RNA-binding protein PUM2 impairs mitochondrial dynamics and mitophagy during aging. **73**, 775–787.
- Trajanoska, K., Rivadeneira, F., Kiel, D.P. *et al.* (2019) Genetics of bone and muscle interactions in humans. *Curr. Osteoporos Rep.*, **17**, 86–95.
- Mahmassani, Z.S., Reidy, P.T., McKenzie, A.I. *et al.* (2019) Age-dependent skeletal muscle transcriptome response to bed rest-induced atrophy. *J. Appl. Physiol.*, **126**, 894–902.
- Ingenbleek, Y. (2019) Plasma transthyretin as a biomarker of sarcopenia in elderly subjects. *Nutrients*, **11**, 4.
- Zhou, J., So, K.K., Li, Y. *et al.* (2019) Elevated H3K27ac in aged skeletal muscle leads to increase in extracellular matrix and fibrogenic conversion of muscle satellite cells. *Aging Cell*, **18**, e12996.
- Balliu, B., Durrant, M., Goede, O.D. *et al.* (2019) Genetic regulation of gene expression and splicing during a 10-year period of human aging. *Genome Biol.*, **20**, 230.
- Gheller, B.J.F., Riddle, E.S., Lem, M.R. *et al.* (2016) *Understanding Age-Related Changes in Skeletal Muscle Metabolism: Differences Between Females and Males*. <http://www.annualreviews.org/doi/10.1146/annurev-nutr-071715-050901> (26 January 2017, date last accessed).
- van den Borne, J.J., Kudla, U. and Geurts, J.M. (2016) Translating novel insights from age-related loss of skeletal muscle mass and phenotypic flexibility into diet and lifestyle recommendations for the elderly. *Curr. Opin. Food Sci.*, **10**, 60–67.
- R Core Team. (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2018-2020, last accessed).
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. <https://cran.r-project.org/web/packages/ggplot2/index.html> (2018-2020, last accessed).
- Chang, W., Cheng, J., Allaire, J.J. *et al.* (2016) *Shiny: Web Application Framework for R*. <https://CRAN.R-project.org/package=shiny> (2018-2020, last accessed).

28. Barrett, T., Wilhite, S.E., Ledoux, P. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
29. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
30. Athar, A., Füllgrabe, A., George, N. *et al.* (2019) ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res.*, **47**, D711–D715.
31. Kolesnikov, N., Hastings, E., Keays, M. *et al.* (2015) ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
32. Leinonen, R., Sugawara, H. and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
33. Lonsdale, J., Thomas, J., Salvatore, M. *et al.* (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
34. Zerbino, D.R., Achuthan, P., Akanni, W. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.
35. Baig, M.H., Rashid, I., Srivastava, P. *et al.* (2019) NeuroMuscleDB: a database of genes associated with muscle development, neuromuscular diseases, ageing, and neurodegeneration. *Mol. Neurobiol.*, **56**, 5835–5843.
36. Thorley, M., Malatras, A., Mazza, E. *et al.* (2017) SysMyo: tailored bioinformatics tools for omics data exploration in muscular dystrophy and other neuromuscular disorders. *Neuromuscular Disord.*, **1**, S8.
37. Pillon, N.J., Gabriel, B.M., Dollet, L. *et al.* (2020) Transcriptomic profiling of skeletal muscle adaptations to exercise and inactivity. *Nat. Commun.*, **11**, 470.