

Received December 22, 2019, accepted January 14, 2020, date of publication January 21, 2020, date of current version January 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968514

# SGDB: A Sports Gene Database for Visualization of Sports Effects on Human Skeletal Muscle Gene Expression

QINGLEI CAO<sup>1</sup>, ZHONGYUAN DENG<sup>1,4</sup>, JUAN LIU<sup>2</sup>, AND XIAOLAN LI<sup>3</sup>

<sup>1</sup>Department of Physical Education, University of Science and Technology Beijing, Beijing 100091, China

<sup>2</sup>College of Life Science, Capital Normal University, Beijing 100048, China

<sup>3</sup>Department of Physical Education, Henan Normal University, Xinxiang 453007, China

<sup>4</sup>School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Qinglei Cao (cqlei72@sina.com)

This work was supported by the School Fundamental Research Funds of the University of Science and Technology, Beijing, for the central universities of the PRC Ministry Education under Grant FRF-BR-16-003B.

**ABSTRACT** The relationship between physical exercise and human health has been widely recognized by sports-related professionals and the public. However, the molecular mechanism behind sports remains unclear. A microarray sequencing technology has been widely used to detect gene changes after physical exercise and researchers have accumulated a great deal of data. These data provide a large amount of information for researchers to understand the relationship between genetics and exercise and study changes in genes after different forms of exercise, and their relations with human, gender and age. We integrate eight published gene expression datasets from human skeletal muscle and construct an online web server called SGDB standing for Sports Gene DataBase. SGDB can be used to search for genes that are expressed after exercise and no exercise. One can determine the effects of physical exercise on gene expression while allowing for visualization of changes in female and male. In addition, different forms of exercise and relationships with age are also available. Currently, SGDB contains 54613 annotated gene probes based on the Affymetrix Human Genome U133 Plus 2 Affay (GPL570). It has become one of the most important resources for researchers to study sports and genes. The online access address for SGDB is [www.sportsgene.top](http://www.sportsgene.top).

**INDEX TERMS** Exercise, gene expression, database, SGDB, sports gene.

## I. INTRODUCTION

It is widely known that proper physical exercise is needed for a healthy human body and has much positive impact on human health. It can promote the growth of bones and muscles and enhance the healthy condition of a human body. Proper physical exercise can enhance human immunity and human body's ability to adapt to the environment [1]. In addition to physiological benefits, physical exercise can also improve and adjust our mental state. It can thus promote health, depending on physical fatigue, to facilitate learning and working efficiency. Thus, a collective project can foster a unity of people's consciousness of cooperation [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Mengchu Zhou<sup>1</sup>.

Inappropriate physical exercise has its associated side effects, such as an increase in the probability of sport injuries, especially intense antagonism. Long and excessive exercise can cause sport fatigue, which is one of the most common phenomena among athletes [3]–[5].

Different results can be produced by movements and are based on the regulation of gene expression. However, an understanding of the molecular regulatory network of muscles during different movements remains incomplete. For example, with long-term aerobic and anaerobic exercise time, the body's reactions and associated molecular mechanisms are different. In aerobic exercise, there is an increase in the synthesis of oxygen-related molecules and proteins in comparison with anaerobic exercise [6]. It is well-known that muscle growth is generally associated with activation of the mTOR (mammalian Target of Rapamycin) signaling pathway

and an increase in related proteins, although this is one of many changes [7]. Changes in gene expression caused by exercise are related to a complex regulatory network, and high-throughput gene expression sequencing provides the possibility of full understanding of changes in movement and gene expression.

With the development of a high-throughput sequencing technology, an increasing number of studies have attempted to explore the mRNA change in response to exercise or the transcriptional regulation change in muscle tissue. Specifically, microarray analysis allows researchers to assess large amounts of muscle transcription involved in the response to acute and chronic movements. Many studies have shown that many factors can affect the gene expression response to exercise, including age, gender, exercise style and so on. The studies [8]–[11] show that young people have more than 80 different signature transcripts that differ from those of older adults. Another study finds that a different exercise style (acute and chronic exercise) also has many differentially expressed genes, most of which are related to muscle development and cell oxidation metabolism. Further analysis [12] has shown that chronic exercise contributes more to muscle development. However, most of the described research has considered one factor only. There has not been a study based on systematic research about the molecular mechanism of the skeletal muscle response to physical exercise and all the related factors.

In this article, we present an online accessible database, called SGDB (the abbreviation of Sports Gene DataBase), that allows for the visualization of the relationship between motion and gene expression. The database integrates a total of eight studies of sequencing samples. Gene expression levels in a control group and exercise one can be searched for by gene symbols or gene ID. In addition, we also demonstrated the effects of different sport styles and genders on gene expression. There has not been much research focused on the effect of age on the gene expression in response to exercise. Thus, we integrate the correlation between age and gene expression into our web server. Currently, SGDB contains 54613 annotated transcripts based on the platform GPL570. It could serve as a reliable and convenient tool for researchers to study the gene expression response to exercise and plays important roles as other databases used their respective fields, e.g., cyber-physical systems and bio-informatics [13]–[19]. Its web server is available online at [www.sportsgene.top](http://www.sportsgene.top).

## II. MATERIALS AND METHODS

### A. DATA SOURCES

The experimental data consist of a total of 603 samples, with the source of expression from eight published microarrays. All the expression data are from Affymetrix Human Genome U133 Plus 2 Platform based on NCBI Gene Expression in Omnibus (GEO). To download data for different genders or ages, or other specific information, please refer to the attached sample of the expression of the human

skeletal muscle gene based on different movements (shown in S1 Table).

### B. DATA STANDARDIZATION AND PROCESSING

SGDB contains 603 samples from the Affymetrix HG-U133 Plus 2 gene chip platform, which is based on the transcriptome that includes most people. The current SGDB contains 54613 annotated gene probes as the formal gene name and number. The standard data are used to normalize the original RMA chip data and the final results of the log<sub>2</sub> transformation. Then, we use boxplot drawing to detect whether the standardized data are similar.

One-way analysis of variance (ANOVA) followed by Tukey's HSD tests were performed to test the significance of differences in the relative gene expression levels in Fig.4. When there were only two treatments (e.g., pre-exercise vs. post-exercise Fig.2 and Fig.3), we conducted independent t-tests to compare the significance of differences between the two treatments. All of the statistical tests were completed by SPSS software (version 19.0) and GraphPad (version 5.0 software).

### C. MAP

The gene expression levels of an exercise group and control one are compared by using an unpaired t-test. All the pictures in SGDB are drawn by a program written in R language.

### D. SGDB WEB SERVER CONSTRUCTION

An SGDB Web server is constructed by using the Hypertext Preprocessor (PHP) programming language, and a Linux system (centos 6.5) is used for all the data stored in the MySQL database (5.1.46) (<http://www.mysql.com>). All genetic information, probe data, and expression data can be directly downloaded.

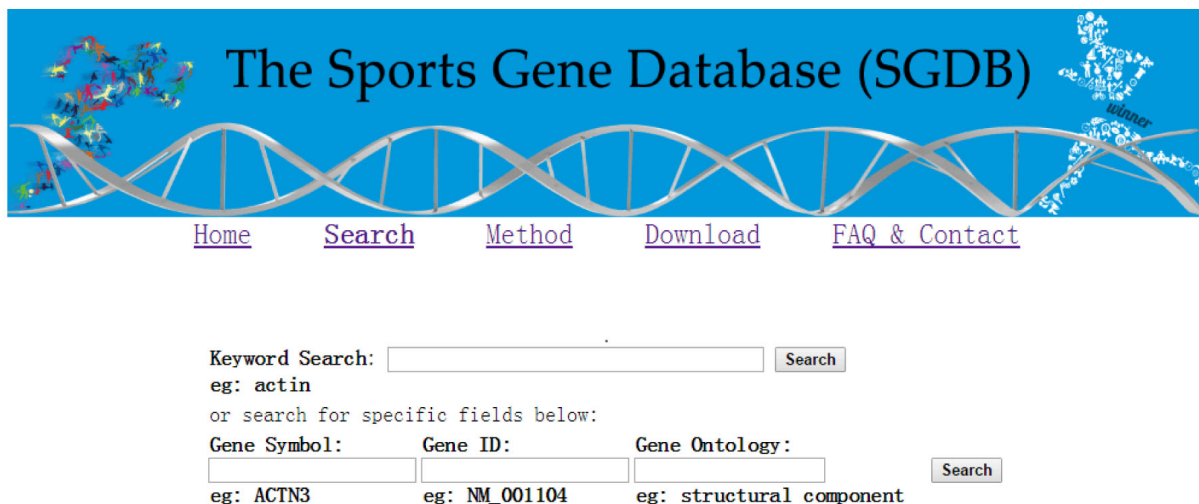
## III. RESULTS

### A. QUERY SGDB

SGDB is a database based on a hypertext preprocessor (PHP). It consists of five core parts: homepage, query, method, data download and user feedback as shown in Fig.1A. Users can access its homepage at via [www.sportgene.top](http://www.sportgene.top), which includes a brief introduction to SGDB. The method includes a page frame data processing method and database. Data download part can be used to download the original gene expression data. User feedback is to collect the user questions and feedback.

A query page is the core part of this database and is shown in Fig.1A. Browsing allows users to find detailed information on various records, including the encoded RNA's sequence ID and annotation. Users can click on a gene name to browse the corresponding gene's basic description as illustrated in Fig.1B. In addition, the data source can allow for relevant information to be downloaded. The detailed information of each record is displayed under the "sequence ID", and the organization expression features of each record

### A: The SGDB search page



### B: The SGDB search result



**FIGURE 1.** SGDB search page and search result. (A) The SGDB web site search page and result. (B) An example of a search result.

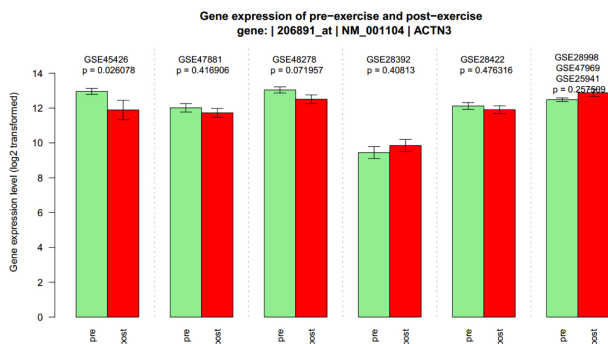
are displayed in the information page. There are two kinds of retrieval methods. One is a keyword that can be input, which is a fuzzy search that searches all of the gene entries associated with the keyword. Another way is to perform accurate search for gene symbols (for example, *ACTN3* and *GAPDH*). One can enter the accession number from the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/gene>).

Gene Ontology (gene annotation information) or GO can be used for such search. After searching, all associated results with this gene probe data are shown, and the details of each probe can be seen by clicking on the name of the gene. The expression and statistical information in different states of genes are stored in a PDF format file. Each gene corresponds to four PDF files. To describe the relationship between gene

and movement, the relationship between different genders, the effects of different exercise styles on gene expression, and the relationship between movement and SGDB provides user-friendly retrieval capabilities. Users can perform a vague search, such as *ACTN3*, and any gene containing these characters can be displayed. Lastly, we can also do a combination search to narrow a search scope.

**B. GENE CHANGES BEFORE AND AFTER PHYSICAL EXERCISE**

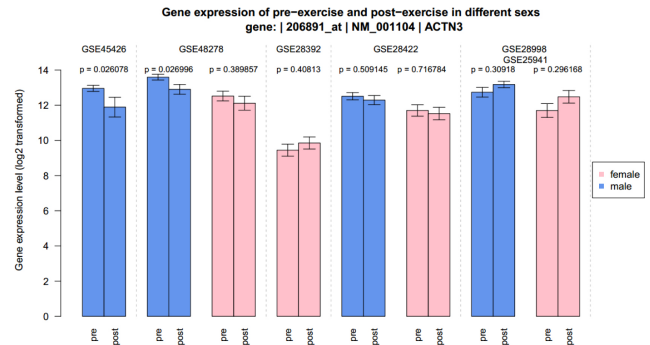
The *ACTN3* gene, for example, is retrieved in SGDB. It encodes the protein actinin-3 [20], [21], which is a part of the contractile (sarcomeric) unit working in the skeletal muscle V fibers to generate rapid and forceful contractions in speedy activities. Its probe number in the Human Genome U133 plus 2.0 platform is 219727. After searching for it (by the gene ID, the gene symbol or other information about *ACTN3*), the related results of *ACTN3* are shown in Fig 2. We compare the expression of *ACTN3* with genes in active and inactive individuals. Users can clearly see the difference in the expression of this gene before and after physical exercise in eight independent studies.



**FIGURE 2. Gene expression of pre-exercise and post-exercise. The gene expression level was calculated by log<sub>2</sub> transformation. There are 6 studies. The difference between the two groups is calculated by a t-test.**

**C. CHANGES OF GENE EXPRESSION BEFORE AND AFTER EXERCISES FOR DIFFERENT GENDER**

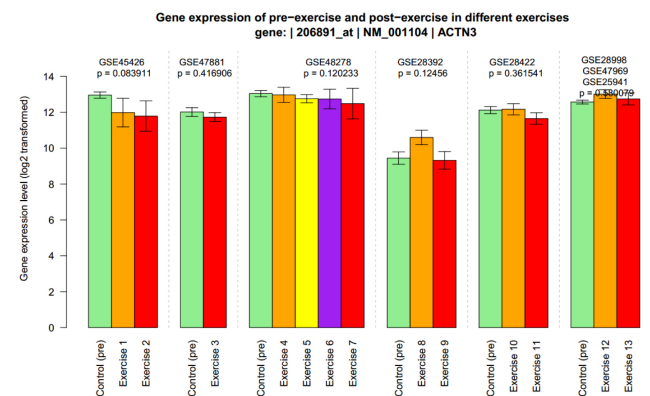
To explore whether the gender of an individual would affect the gene expression response to exercise, we calculate the gene expression in different genders for exercise and non-exercise groups. SGDB can help us understand the differences in response to exercise due to gender. Fig 3 displays the results (the query gene: *ACTN3*). The gender differences cause different gene responses to exercise in the exercise group. The results show that there is no significant difference between female and male regarding the gene response to exercise. The output PDF files are generated by ANOVA (more than two groups) or a t-test (two groups) in an R program. The gene expression levels were calculated and drawn by the R program. In Fig 3, we compare the target gene’s expression levels in male (blue) and female (pink) in terms of response to exercise. We have found that in the



**FIGURE 3. Gene expression of pre-exercise and post-exercise in different gender. The gene expression level was calculated by log<sub>2</sub> transformation. There are 5 studies. The GSE45426 group lacks female, while the GSE28392 group lacks male. The difference between the two groups is calculated by a t-test.**

first and second datasets, there was a significant difference in *ACTN3* in male post-exercise ( $p=0.02608$  and  $p=0.026996$ ), whereas the remaining groups display no significant changes. The result reveals that *ACTN3* expression tends to display no difference.

Due to the gender effect on the gene expression response to exercise, we try to explore whether a different physical exercise style would affect gene expression. From eight studies, we pick six containing various exercise styles, with a total of 13 different exercise styles (the exercises are shown in S2 Table). In six studies (Fig. 4), only the first group (Exercises 1 and 2) displays a large difference ( $p=0.083911$ ), whereas the others present no significant difference. Exercises 1 and 2 share a same common ground of being an aerobic and short-duration physical exercise.



**FIGURE 4. Gene expression of pre-exercise and post-exercise in different exercise styles. The gene expression level was calculated by log<sub>2</sub> transformation. There are 6 studies, and every study has a control group. Exercises 1 to 13 are correlated to the 13 kinds of exercise styles. The difference between two groups is calculated by a t-test, and the multi-group difference is calculated by the ANOVA test.**

We also examine the effects of age on the gene expression response to physical exercise. However, the study only results in a small amount of expression data, which is not enough for any significant conclusions regarding the *ACTN3* gene.

#### IV. CONCLUSION

The proposed SGDB is a useful tool for physical exercise related human health research. It has integrated eight transcriptome studies to systemically study the effects of gender and exercise styles in terms of gene expression responses to exercise. Previous studies on the effects of physical exercise on muscle transcriptome are relatively rare. They focused on different tissues or specific genes [22]–[24]. In our database, we have collected 603 samples from eight public studies' transcriptome data in human skeletal muscle in the context of physical exercise. Based on these data, we have calculated 64513 gene or transcript expression levels. We have further analyzed the gene expression response to physical exercise and the effect of various factors, including gender and exercise styles. We have also explored the correlation between gene expression and exercise. SGDB is expected to provide us a new outlook on the regulation of genes in the process of exercise, especially elucidating how exercise influences the muscles' gene expression response. SGDB is also a useful tool to explore the molecular mechanisms of exercise's influence on the human body while helping us discover the treatment and prevention methods about exercise-induced fatigue. Our web server ([www.sportsgene.top](http://www.sportsgene.top)) can be freely accessed, and anyone who is interested in it can easily search the target gene related to exercise. In the future, we plan to further expand our database and collect more public microarray datasets, in addition to related sequencing data that include miRNA, lncRNA, ciRNA and others. Furthermore, we hope to include proteomics and metabolomics as well. At the same time, more statistical analysis and model algorithms will be applied to the analysis data in the future, and strive to get more effective information in high-throughput sequencing data to assist sports biology research. We believe that with its continuous improvement, our SGDB web server should be increasingly useful in the future.

#### ACKNOWLEDGMENT

(Qinglei Cao and Zhongyuan Deng contributed equally to this work.)

#### REFERENCES

- [1] R. Coll, J. Izquierdo, and G. Salto, "Benefit of physical exercise in medicine," *Anal. Med. Int.*, vol. 8, no. 2, p. 101, Feb. 1991.
- [2] F. J. Penedo and J. R. Dahn, "Exercise and well-being: A review of mental and physical health benefits associated with physical activity," *Current Opinion Psychiatry*, vol. 18, no. 2, pp. 189–193, Mar. 2005.
- [3] B. J. Mutch and E. W. Banister, "Ammonia metabolism in exercise and fatigue: A review," *Med. Sci. Sports Exerc.*, vol. 15, no. 1, pp. 41–50, 1983.
- [4] K. Sahlin, "Metabolic factors in fatigue," *Sports Med.*, vol. 13, no. 2, pp. 99–107, 1992.
- [5] M. L. Schillings, W. Hoefsloot, D. F. Stegeman, and M. J. Zwartz, "Relative contributions of central and peripheral factors to fatigue during a maximal sustained effort," *Eur. J. Appl. Physiol.*, vol. 90, nos. 5–6, pp. 562–568, Nov. 2003.
- [6] S. Nioka, D. Moser, G. Lech, M. Evengelisti, T. Verde, B. Chance, and S. Kuno, "Muscle deoxygenation in aerobic and anaerobic exercise," *Adv. Exp. Med. Biol.*, vol. 45, no. 4, pp. 63–70, 1998.
- [7] S. C. Bodine, "mTOR signaling and the molecular adaptation to resistance exercise," *Med. Sci. Sports Exerc.*, vol. 38, no. 11, pp. 1950–1957, Nov. 2006.
- [8] M. M. Bamman, J. K. Petrella, J.-S. Kim, D. L. Mayhew, and J. M. Cross, "Cluster analysis tests the importance of myogenic gene expression during myofiber hypertrophy in humans," *J. Appl. Physiol.*, vol. 102, no. 6, pp. 2232–2239, Jun. 2007.
- [9] U. Raue, T. A. Trappe, S. T. Estrem, H.-R. Qian, L. M. Helvering, R. C. Smith, and S. Trappe, "Transcriptome signature of resistance exercise adaptations: Mixed muscle and fiber type specific profiles in young and old adults," *J. Appl. Physiol.*, vol. 112, no. 10, pp. 1625–1636, May 2012.
- [10] J. Stockman, "Myostatin mutation associated with gross muscle hypertrophy in a child," *Yearbook Pediatrics*, vol. 2006, pp. 319–321, Jan. 2006.
- [11] J.-S. Kim, J. K. Petrella, J. M. Cross, and M. M. Bamman, "Load-mediated downregulation of myostatin mRNA is not sufficient to promote myofiber hypertrophy in humans: A cluster analysis," *J. Appl. Physiol.*, vol. 103, no. 5, pp. 1488–1495, Nov. 2007.
- [12] B. E. Phillips, J. P. Williams, T. Gustafsson, C. Bouchard, T. Rankinen, S. Knudsen, K. Smith, J. A. Timmons, and P. J. Atherton, "Molecular networks of human muscle adaptation to exercise and age," *PLoS Genet.*, vol. 9, no. 3, Mar. 2013, Art. no. e1003389.
- [13] X. Guan, B. Yang, C. Chen, W. Dai, and Y. Wang, "A comprehensive overview of cyber-physical systems: From perspective of feedback system," *IEEE/CAA J. Autom. Sinica*, vol. 3, no. 1, pp. 1–14, Jan. 2016.
- [14] Y. Guo, X. Hu, B. Hu, J. Cheng, M. Zhou, and R. Y. K. Kwok, "Mobile cyber physical systems: Current challenges and future networking applications," *IEEE Access*, vol. 6, pp. 12360–12368, 2018.
- [15] H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 3, pp. 703–715, May 2019.
- [16] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu, "Review on cyber-physical systems," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 1, pp. 27–40, Jan. 2017.
- [17] W. Luan, G. Liu, C. Jiang, and L. Qi, "Partition-based collaborative tensor factorization for POI recommendation," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 3, pp. 437–446, Jul. 2017.
- [18] X. X. Na and D. J. Cole, "Modelling of a human driver's interaction with vehicle automated steering using cooperative game theory," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 5, pp. 1095–1107, Sep. 2019.
- [19] S. M. M. Rahman, "Cyber-physical-social system between a humanoid robot and a virtual human through a shared platform for adaptive agent ecology," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 190–203, Jan. 2018.
- [20] M. DG and N. KN, "ACTN3: A genetic influence on muscle function and athletic performance," *Exerc. Sport Sci. Rev.*, vol. 35, no. 1, pp. 30–34, Jan. 2007.
- [21] R. A. Scott, R. Irving, L. Irwin, E. Morrison, V. Charlton, K. Austin, D. Tladi, M. Deason, S. A. Headley, F. W. Kolkhorst, N. Yang, K. North, and Y. P. Pitsiladis, "ACTN3 and ACE genotypes in elite Jamaican and US sprinters," *Med. Sci. Sports Exerc.*, vol. 42, no. 1, pp. 107–112, Jan. 2010.
- [22] J. Padilla, N. T. Jenkins, P. K. Thorne, J. S. Martin, R. S. Rector, J. W. Davis, and M. H. Laughlin, "Transcriptome-wide RNA sequencing analysis of rat skeletal muscle feed arteries. II. Impact of exercise training in obesity," *J. Appl. Physiol.*, vol. 116, no. 8, pp. 1033–1047, Apr. 2014.
- [23] T. R. Lundberg, R. Fernandez-Gonzalo, P. A. Tesch, E. Rullman, and T. Gustafsson, "Aerobic exercise augments muscle transcriptome profile of resistance exercise," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 310, no. 11, pp. R1279–R1287, Jun. 2016.
- [24] K. Vissing and P. Schjerling, "Simplified data access on human skeletal muscle transcriptome responses to differentiated exercise," *Sci. Data*, vol. 68, no. 1, Nov. 2014, Art. no. 140041.



**QINGLEI CAO** is currently a Professor with the Department of Physical Education, University of Science and Technology Beijing. His main research interests include volleyball and sports fatigue, database, and big data analytics.



**ZHONGYUAN DENG** received the Ph.D. degree. He is currently with the School of Agricultural Sciences, Zhengzhou University. His main research interest includes bioinformatics.



**XIAOLAN LI** is currently a Faculty Member with the Department of Physical Education, Henan Normal University. Her main research interest includes teaching and research of sports theory.

• • •



**JUAN LIU** is currently pursuing the master's degree with the College of Life Science, Capital Normal University, China. Her main research interests include bioinformatics and data analysis methods.