Identifying Gender Independent Biomarkers Responsible for Human Muscle Aging Using Microarray Data

Emmanouil G. Sifakis, *Member, IEEE*, Ioannis Valavanis, *Member, IEEE*, Olga Papadodima, and Aristotelis A. Chatziioannou, *Member, IEEE*

***Abstract*—The scope of this study is the identification of gender-independent muscle transcriptional differences between younger and older subjects using skeletal muscle gene expression profiles. Towards this end, a combination of statistical methods, functional analyses, and machine learning techniques were exploited, and applied on an integrative dataset of publicly available microarray data obtained from healthy males and females. Through the proposed framework, a set of 46 reliable genes was identified that comprise a candidate gender-independent aging signature in human skeletal muscle. The identification was based on differential expression, information gain content, and significance regarding their central regulatory role in the underlying active molecular networks in the GO. The resulted gene subset was also tested for its generalization potency regarding the classification task, through the use of a series of classifiers, and results show that high classification accuracies could be obtained. Therefore, the selected genes comprise a promising group of biomarkers of ageing in human skeletal muscle to be evaluated in future studies.**

I. Introduction

AGEING is arguably the most familiar yet least well-understood aspect of human biology [1]. It is a complex, physiological process that is probably influenced by various genetic and environmental factors. Throughout the years, ageing has been associated with macromolecules impairments (e.g. DNA, RNA, proteins) [1]. However, the molecular pathways involved are not fully elucidated [2].

Ageing of skeletal muscle comprises a significant issue, since muscle accounts for approximately half of the cell mass of the human body, and sarcopenia is a key feature of age-related frailty [3], [4]. Many of the age-related changes in skeletal muscle appear to be influenced by sex, however controversy exists regarding how sex influences each aspect of the aging process of skeletal muscle [5].

High-throughput transcript profiling methods using DNA microarrays have been successful in producing an unbiased global view of the molecular events that occur with age in various tissues, including human skeletal muscle. Specifically, DNA microarrays has been used to study primarily sex-related differences in gene expression in human skeletal muscle, e.g. [5]–[7], while few of them have focused on gender independent profiling , e.g. [8]. This can be due to that men and women differ in hormone levels and certain muscle characteristics [9], [10], so that it cannot be assumed that they have identical age-related changes in muscle [11].

In this present study, we combined transcriptional profiling data of human skeletal muscle from young and elderly, male and female subjects. Using statistical methods along with functional analysis and data mining techniques, we objected a common, gender-independent aging signature. To the authors' knowledge no study has so far applied this combination of statistical and artificial intelligence based pipeline to document the gender-independent differences in the expression profiling of aged skeletal muscle between young and elderly adults. Specifically, the scope of the present study was the selection of the most reliable biomarkers, and the assessment of their classification power in discriminating between the two classes: Young and Old. The dataset used here comes from different, publicly available microarray data experiments. All the statistical analyses were performed using the statistical programming language R and Bioconductor, which provides tools for the analysis and comprehension of high-throughput genomic data [12]. The functional analyses were carried out through StRAnGER [13] and GORevenge [14] web applications. The data mining was performed through Rapidminer, a freely available open-source platform that integrates fully the machine learning WEKA library, and permits easily data mining algorithms integration, process and usage of data and metadata [15].

II. Dataset

The dataset used in the present study contains microarray gene expression data from vastus lateralis biopsies obtained from healthy young (20-29 years old) and old (65-75 years old), male and female subjects. The dataset was formed using data from different, MIAME-compliant [16] experiments that are publicly available at the Gene Expression Omnibus (GEO) database with accession numbers GDS287, GDS288, GDS472 and GDS473 (see [3], [11] for more detailed information regarding these hybridizations). In total, the dataset encompassed 30 samples, 14 from young male (N_{Ym}=7) and female (N_{Yf}=7),

Manuscript received July 30, 2013. This work was supported by the MIS 377001 THALES Project entitled "Development of Systems Biology and Bioinformatics tools to study the dynamics of cell aging", co-financed by the European Union (European Social Fund) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework.

E. G. Sifakis, I. Valavanis, O. Papadodima, and A. A. Chatziioannou (corresponding author) are with the Institute of Biology, Medicinal Chemistry & Biotechnology, National Hellenic Research Foundation, Athens, Greece (phone: +30-210-7273768; fax: +30-210-7273751; e-mail: achatzi@eie.gr).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

and 16 from old male (N_{Om}=8) and female (N_{Of}=8) subjects. All experiments were used the Affymetrix® (Santa Clara, CA) Human Genome U133A and U133B oligonucleotide arrays (HG-U133 Set). The HG-U133 Set has about 44,000 probe sets that measure the expression of about 33,000 genes.

III. Methods

A. Data Pre-Processing

For each experiment, no raw data were available and the already normalized data were used. Missing value imputation was applied in the normalized data using the nearest neighbor averaging [17]. Then, the expression values were filtered out based on a threshold fraction of the Present detection calls (derived by the Affymetrix' s MAS5 algorithm), hence increasing the ratio of true positives to false positives [18]. Specifically, we filtered out probes characterised as not Present by the MAS5 detection call in at least 60% of the samples in at least one group. After filtering, each expression value was divided by the mean of all expression values in each array series, in order to be comparable between the experiments.

B. Statistical Selection

To preselect a wide subset of candidate biomarkers prior to the functional analyses or application of the data mining framework, we utilized well-established differential expression analysis. Specifically, the approach described in [19] was followed, hence a linear model was fitted to the log expression values for each gene, and then empirical Bayes shrinkage is used in order to calculate moderated *t*-statistics. Using the *p*-values produced by the empirical Bayes method, statistically significant probe IDs were selected for each experiment. In particular, the probe IDs presenting a *p*-value lower than 0.01 were chosen per experiment. The resulted subsets of statistically significant probe IDs of each experiment were unified. The unification of these subsets yielded a total of 1,507 probe IDs, corresponding to 1,218 unique gene symbols, according to the latest Affymetrix HG-U133 Set annotation files. This unified set comprises of genes that are potentially significant to the ageing process. Thus, it was used as input to both the independent and the coupled with entropy-based information gain (IG) estimation functional analyses.

C. Independent Functional Analysis

To better understand the molecular mechanisms implicated in the promotion of muscle ageing, functional analysis was performed through the use of the StRAnGER web application [13]. By employing established statistical tests coupled with bootstrapping, StRAnGER manages to derive a final population of statistically significant terms that comprise a set of over-represented terms, compared to all other terms of the ontology utilized. By feeding the StRAnGER platform with the unified gene list, the over-represented Gene Ontology (GO) terms were identified.

D. Functional Analysis Coupled with Entropy-Based Information Gain Estimation

GO was further used for the exploration of the underlying functional content beneath the unified gene set. To this end, the GORevenge algorithm [14], freely available through the web, was applied. GORevenge exploits graph-theoretical algorithmic methodologies and systematically exploits the GO tree in order to output a series of functionally related genes/GO terms that may not be included in the list of genes/GO terms initially submitted to the algorithm. Thus, it can aid the elucidation of hidden functional regulatory effects among genes and can therefore promote a system's level interpretation. GORevenge was applied separately twice, once for the 'Molecular Function' and once for the 'Biological Process' aspect, utilizing the same parameters (the Resnik semantic similarity metric, the Bubble genes algorithm, and a relaxation equal to 0.15). Out of the GORevenge's output list with the most important genes, only those that were simultaneously contained in the original unified gene set were selected for further processing.

To further increase our confidence in the context of the systemic analysis, the entropy-based information gain (IG), also known as Kullback–Leibler divergence [20], was also calculated for the initial unified gene set. Specifically, the genes of the unified set were ranked based on their IG values, and genes with an IG value higher than 0.4 were selected as the most informative ones in terms of variability.

Genes with relatively high IG and genes with a central regulatory role in the underlying active molecular networks (outputted from the GORevenge analysis) represent genes with mutually independent (orthogonal) characteristics. Therefore, an intersection of the high IG valued gene set

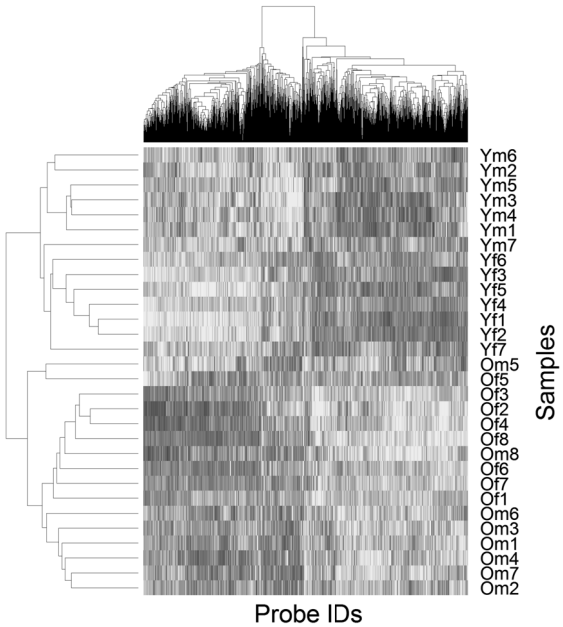

Fig. 1. Clustered display of the expression values of the 1,507 unified probe ID set. Each row corresponds to a sample either of a (Y)oung or (O)ld, (m)ale or (f)emale, with the number designating its corresponding replicate. Each column corresponds to a probe ID. Higher expression values are represented with darker shades of gray.

with the functionally related GORevenge gene sets may possibly reveal the highly informative critical molecular players involved in the ageing process in both genders. Thus, the two GORevenge output lists (based on 'Molecular Function' or 'Biological Process' aspect) were searched for potential genes that also presented high IG.

E. Classification

The resulted gene subset was also tested for its generalization potency regarding the classification task. To this effect, various classifiers were constructed and their performance was measured, in order to validate the relevance of the selected genes. Specifically, three weighted *k*-nearest neighbor (*k*-NN) classifiers (*k*=1, 6, 10) [21], a Decision Tree (DT) [22], and the Random Forest (RF) algorithm [23] were used, while their performance was measured using the leave-one-out resampling technique.

IV. Results and Discussion

A. Cluster analysis

The clustering method applied to unified set of 1,507 probe IDs was the average-linkage hierarchical clustering algorithm, with the Unweighted Pair-Group Method with Arithmetic mean (UPGMA) as the linkage rule, and the Pearson product-moment correlation coefficient as the similarity metric. As illustrated in Fig. 1, hierarchical clustering was able to separate completely the two classes (Young and Old).

TABLE III
Classification Accuracy of the Different Classifiers

1-*nn*	6-*nn*	10-*nn*	DT	RF
93.33 %	96.67 %	96.67 %	63.33 %	90.00 %

B. Independent Functional analysis

To relate the individual genes with consistent biological processes, the StRAnGER algorithm was utilized to perform statistical enrichment analysis, taking as input the unified set of 1,218 unique gene symbols. The output comprised a list of GO terms including 75 significant over-represented GO terms (hypergeometric test $p \leq 0.05$, 90% cut-off percentage, 10^4 bootstrap iterations). The most significant ($p \leq 10^{-5}$) GO terms of the 'Biological Process' domain are presented in Table II. These biological processes are mainly related with cellular metabolism and RNA splicing. The finding regarding to metabolism-related processes is in line with the fact that processes, like tricarboxylic acid cycle, mitochondrial electron transport, and ATP metabolism, are known key mechanisms implicated to ageing-related pathways [24], [25]. In addition, it is really interesting that significantly enriched GO terms indicate alteration in mechanisms governing RNA splicing. In agreement with this finding, recent evidence suggests that alternative RNA splicing is one of the mechanisms that regulate gene expression during lifespan [26], [27].

C. Functional Analysis Coupled with Entropy-Based Information Gain Estimation

In the direction of identifying the most critical molecular players that also embody high information content, the two

TABLE I
Over-Represented GO Terms of the Unified Set of 1,218 Genes

GO Term	Term Description	Hypergeometric *p*-value	Enrichment Score	Bootstrap-corrected *p*-value
GO:0006099	tricarboxylic acid cycle	2.09E-11	15/27	3.24E-07
GO:0022904	respiratory electron transport chain	2.78E-11	40/100	4.24E-07
GO:0044281	small molecule metabolic process	1.17E-10	136/1029	8.14E-07
GO:0008380	RNA splicing	1.11E-09	44/274	1.22E-06
GO:0006810	transport	2.62E-09	87/753	1.41E-06
GO:0006120	mitochondrial electron transport, NADH to ubiquinone	2.66E-09	15/44	1.50E-06
GO:0000398	nuclear mRNA splicing, via spliceosome	1.12E-08	31/171	1.79E-06
GO:0042776	mitochondrial ATP synthesis coupled proton transport	4.00E-08	8/15	1.89E-06
GO:0015986	ATP synthesis coupled proton transport	7.10E-08	9/20	1.99E-06
GO:0006397	mRNA processing	7.38E-08	34/213	2.09E-06
GO:0006412	translation	1.51E-06	38/283	2.93E-06
GO:0010510	regulation of acetyl-CoA biosynthetic process from pyruvate	1.84E-06	6/12	3.01E-06
GO:0006915	apoptosis	2.43E-06	68/638	3.22E-06
GO:0006090	pyruvate metabolic process	2.69E-06	8/22	3.31E-06
GO:0055114	oxidation reduction	3.31E-06	56/498	3.40E-06
GO:0006102	isocitrate metabolic process	4.73E-06	4/6	3.49E-06
GO:0010467	gene expression	6.80E-06	65/620	3.67E-06
GO:0006006	glucose metabolic process	7.21E-06	19/108	3.88E-06
GO:0006103	2-oxoglutarate metabolic process	7.61E-06	7/19	3.96E-06
GO:0006107	oxaloacetate metabolic process	1.78E-05	5/11	4.06E-06
GO:0043484	regulation of RNA splicing	2.73E-05	7/22	4.16E-06
GO:0045926	negative regulation of growth	3.45E-05	6/17	4.27E-06
GO:0006094	gluconeogenesis	3.84E-05	10/43	4.42E-06
GO:0046034	ATP metabolic process	3.97E-05	7/23	4.56E-06
GO:0006200	ATP catabolic process	4.05E-05	15/84	4.66E-06

The enrichment score equals to the number of times a GO term appears due to its relation to a differentially expressed gene, divided by the number of times the GO term appears due to all genes in the human background annotation available to the StRAnGER platform. All GO terms belong to the 'Biological Process' domain.

GORevenge output lists with the most important genes involved in the 'Molecular Function' and/or 'Biological Process' aspect were overlapped with genes with an IG value higher than 0.4. Genes resulted from the GORevenge analysis, which were simultaneously contained in the original differentiated gene set, comprised a total of 77 and 75 genes for the two aspects, respectively. Genes presented IG≥0.4 were equal to 371 in total. The union of the GORevenge's resulted gene lists of the two aspects comprised a total of 118 unique genes. From these genes, the ones with high informative content in terms of IG were equal to 46 unique genes. This set of genes was also tested for its generalization potency regarding the classification task.

Table II presents the GORevenge's most important genes for the 'Biological Process' aspect that were simultaneously included in the original differentially expressed gene set and also demonstrated relatively high IG values. Notably, two genes of this list are known to be involved in age-related processes; the *DLD* gene (IG=0.75) has been found to play a unique role in modulating length of life [28], and the *TFRC* gene (IG=0.45) has been associated with age-dependent mechanisms [29].

D. Classification

The final set of the 46 unique genes were used as features in a series of classification algorithms. These genes achieved high accuracies when evaluated by the constructed classifiers by the use of the leave-one-out strategy (Table III). The 6-*nn* and 10-*nn* weighted classifiers outperformed all other classifiers. In general, high classification power in discriminating between the vastus lateralis biopsies from young and old subjects was provided by the selected features (only DT is featured an accuracy less than 90%).

Overall, the final list comprises a selection of differential expressed genes that also present high informative content, and high classification power. These point to the direction that an underlying functional association with muscle ageing may exist regardless of gender.

TABLE II
GOREVENGE'S MOST IMPORTANT GENES, CORRESPONDING TO THE 'BIOLOGICAL PROCESS' ASPECT, WITH IG>0.4

Gene	GO Count [a]	Prune [0.6] [b]	Prune [0.9] [c]	IG
VEGFA	60	42	7	0.45
SMAD4	35	23	3	0.54
ALDH1A2	35	21	2	0.50
HDAC4	30	18	5	0.63
ABCA1	29	21	4	0.45
PPARA	27	18	5	0.50
CDH13	26	19	7	0.45
MKKS	25	19	6	0.45
TGFBR3	22	18	7	0.71
FOXO3	22	19	6	0.56
MDM2	21	15	8	0.40
NDEL1	20	16	10	0.45
PSMA1	19	10	3	0.40
PIK3R1	19	11	5	0.65
NR2F2	19	16	4	0.45
NRP1	18	16	5	0.40
GOT1	18	8	3	0.71
FKBP1A	18	15	5	0.43
PSMD4	17	10	3	0.45
GOT2	17	9	3	0.43
FOXO1	17	15	3	0.56
DLD	17	13	3	0.72
TFRC	16	12	6	0.45
FLT1	16	13	4	0.40
VLDLR	15	13	6	0.72
UBE2D3	15	7	3	0.40
CEBPB	15	11	5	0.45
HOXA3	14	10	3	0.56

[a] Refers to the number of original GO terms. [b,c] Refer to number of GO terms remaining after GOrevenge pruning in two steps, reflecting the centrality of each gene.

V. CONCLUSION

In the present study, a combination of statistical methods, functional analyses and data mining techniques were proposed, in order to identify a common, gender-independent aging signature in human skeletal muscle. This framework was applied to transcriptional profiling data from different vastus lateralis biopsies from young and elderly, male and female subjects. The final list of genes identification was based on differential expression, information gain content, and significance regarding their central regulatory role in the underlying active molecular networks in the GO tree. The 46 selected genes, when used as input to various learning algorithms, defined classifiers with a moderate till up to a very good classification performance. In whole, the final set of genes comprises a promising set of biomarkers that is to be further investigated for their systemic role in muscle ageing.

REFERENCES

[1] T. B. L. Kirkwood, "Understanding the odd science of aging.," *Cell*, vol. 120, no. 4, pp. 437–47, Mar. 2005.

[2] J. Vijg and J. Campisi, "Puzzles, promises and a cure for ageing.," *Nature*, vol. 454, no. 7208, pp. 1065–1071, 2008.

[3] S. Welle, A. I. Brooks, J. M. Delehanty, N. Needler, and C. a Thornton, "Gene expression profile of aging in human muscle.," *Physiological genomics*, vol. 14, no. 2, pp. 149–59, Jul. 2003.

[4] K. S. Nair, "Aging muscle.," *The American journal of clinical nutrition*, vol. 81, no. 5, pp. 953–63, May 2005.

[5] D. Liu, M. a Sartor, G. a Nader, E. E. Pistilli, L. Tanton, C. Lilly, L. Gutmann, H. B. Iglayreger, P. S. Visich, E. P. Hoffman, and P. M. Gordon, "Microarray Analysis Reveals Novel Features of the Muscle Aging Process in Men and Women.," *The journals of gerontology. Series A, Biological sciences and medical sciences*, no. 14, pp. 1–10, Mar. 2013.

[6] S. Welle, R. Tawil, and C. a Thornton, "Sex-related differences in gene expression in human skeletal muscle.," *PloS one*, vol. 3, no. 1, p. e1385, Jan. 2008.

[7] D. Liu, M. a Sartor, G. a Nader, L. Gutmann, M. K. Treutelaar, E. E. Pistilli, H. B. Iglayreger, C. F. Burant, E. P. Hoffman, and P. M. Gordon, "Skeletal muscle gene expression in response to resistance exercise: sex specific regulation.," *BMC genomics*, vol. 11, no. 1, p. 659, Jan. 2010.

[8] J. M. Zahn, R. Sonu, H. Vogel, E. Crane, K. Mazan-Mamczarz, R. Rabkin, R. W. Davis, K. G. Becker, A. B. Owen, and S. K. Kim, "Transcriptional Profiling of Aging in Human Muscle Reveals a Common Aging Signature," *PLoS Genetics*, vol. 2, no. 7, p. e115, 2006.

[9] J. A. Simoneau and C. Bouchard, "Human variation in skeletal muscle fiber-type proportion and enzyme activities," *American*

Journal of Physiology Endocrinology And Metabolism, vol. 257, no. 4, pp. E567–E572, 1989.

[10] R. S. Staron, F. C. Hagerman, R. S. Hikida, T. F. Murray, D. P. Hostler, M. T. Crill, K. E. Ragg, and K. Toma, "Fiber Type Composition of the Vastus Lateralis Muscle of Young Men and Women," *Journal of Histochemistry & Cytochemistry*, vol. 48, no. 5, pp. 623–629, May 2000.

[11] S. Welle, A. I. Brooks, J. M. Delehanty, N. Needler, K. Bhatt, B. Shah, and C. a Thornton, "Skeletal muscle gene expression profiles in 20-29 year old and 65-71 year old women.," *Experimental gerontology*, vol. 39, no. 3, pp. 369–77, Mar. 2004.

[12] R. Doerge, "Bioinformatics and Computational Biology Solutions Using R and Bioconductor," *Biometrics*, vol. 62, no. 4, pp. 1270–1271, 2006.

[13] A. A. Chatziioannou and P. Moulos, "Exploiting Statistical Methodologies and Controlled Vocabularies for Prioritized Functional Analysis of Genomic Experiments: the StRAnGER Web Application," *Frontiers in neuroscience*, vol. 5, no. January, p. 14, 2011.

[14] K. Moutselos, I. Maglogiannis, and A. Chatziioannou, "GOrevenge: A novel generic reverse engineering method for the identification of critical molecular players, through the use of ontologies.," *IEEE Transactions on Biomedical Engineering*, no. c, 2011.

[15] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler, "YALE: Rapid Prototyping for Complex Data Mining Tasks," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, vol. 2006, pp. 935–940.

[16] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron, "Minimum information about a microarray experiment (MIAME)-toward standards for microarray data," *Nature genetics*, vol. 29, no. 4, pp. 365–371, 2001.

[17] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," *Bioinformatics (Oxford, England)*, vol. 17, no. 6, pp. 520–525, 2001.

[18] J. N. McClintick and H. J. Edenberg, "Effects of filtering by Present call on analysis of microarray experiments.," *BMC bioinformatics*, vol. 7, p. 49, Jan. 2006.

[19] G. K. Smyth, "Limma: linear models for microarray data," in *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, vol. 1st, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, Eds. New York: Springer, 2005, pp. 397–420.

[20] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[21] Y. Wu, "Improved k-nearest neighbor classification," *Pattern Recognition*, vol. 35, no. 10, pp. 2311–2318, 2002.

[22] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 19, no. Book, Whole. Wadsworth, 1984, p. 368.

[23] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] A. Navarro, J. M. López-Cepero, and M. J. Sánchez del Pino, "Skeletal muscle and aging.," *Frontiers in bioscience : a journal and virtual library*, vol. 6, pp. D26–44, Jan. 2001.

[25] E. Carmeli, R. Coleman, and A. Z. Reznick, "The biochemistry of aging muscle.," *Experimental Gerontology*, vol. 37, no. 4, pp. 477–489, 2002.

[26] J. R. Tollervey, Z. Wang, T. Hortobagyi, J. T. Witten, K. Zarnack, M. Kayikci, T. A. Clark, A. C. Schweitzer, G. Rot, T. Curk, B. Zupan, B. Rogelj, C. E. Shaw, and J. Ule, "Analysis of alternative splicing associated with aging and neurodegeneration in the human brain," *Genome Research*, vol. 21, no. 10, pp. 1572–1582, 2011.

[27] P. Mazin, J. Xiong, X. Liu, Z. Yan, X. Zhang, M. Li, L. He, M. Somel, Y. Yuan, Y.-P. Phoebe Chen, N. Li, Y. Hu, N. Fu, Z. Ning, R. Zeng, H. Yang, W. Chen, M. Gelfand, and P. Khaitovich, "Widespread splicing changes in human brain development and aging.," *Molecular systems biology*, vol. 9, p. 633, Jan. 2013.

[28] J. A. Butler, R. J. Mishur, S. Bhaskaran, and S. L. Rea, "A metabolic signature for long life in the Caenorhabditis elegans Mit mutants.," *Aging cell*, vol. 12, no. 1, pp. 130–8, Feb. 2013.

[29] M. Hadziahmetovic, Y. Song, P. Ponnuru, J. Iacovelli, A. Hunter, N. Haddad, J. Beard, J. R. Connor, S. Vaulont, and J. L. Dunaief, "Age-Dependent Retinal Iron Accumulation and Degeneration in Hepcidin Knockout Mice," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 1, pp. 109–118, 2011.