Methods for combining multiple correlated biomarkers with application to the study of low-grade inflammation and muscle mass in senior horses.

December 18, 2019

Olga A. Vsevolozhskaya¹, Alisa C. Herbst², Amanda A. Adams², Cailey Burns³, Bertsie Cantu⁴, Virginia D. Barker², Dmitri V. Zaykin⁵

¹Department of Biostatistics, University of Kentucky, Lexington, KY, USA

²Gluck Equine Research Center, Department of Veterinary Science, University of Kentucky, Lexington, KY, USA

³Department of Science and Math, Asbury University, Wilmore, KY, USA

⁴College of Veterinary Medicine, Lincoln Memorial University, Harrogate, TN, USA

⁵Biostatistics and Computational Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

Correspondence: Dmitri V. Zaykin, Senior Investigator at the Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, P.O. Box 12233, Research Triangle Park, NC 27709, USA. Tel.: +1 (919) 541-0096 ; Fax: +1 (919) 541-4311. Email address: dmitri.zaykin@nih.gov

Abstract

The simplest analysis of biomarker data is based on a series of single biomarker hypothesis tests, followed by correction for multiple testing. However, it is intuitively plausible that a joint analysis of multiple biomarkers will have higher statistical power and promise improved discrimination over tests based on single markers. In this article, we study analytical properties of the approach for joint analysis of correlated summary statistics based on the test for quadratic forms (TQ). Based on the derivation of the TQ-distribution, we proposed a scale-location approximation of the TQ statistic, which we call approximate TQ. We show that the approximate TQ has very similar power to the traditional TQ test under varying correlation structures among biomarkers. Our application of both the TQ and the approximate TQ test to data on biomarkers for inflamm-aging – an age-related low-grade chronic inflammation – reveals an association between the percentage of IFN γ positive lymphocytes and overall muscle condition in senior horses.

Introduction

The role of biomarkers in biomedical research has been steadily increasing over the recent years. For example, a key aspect of research in animal or human genetics is association of single nucleotide polymorphisms (SNP biomarkers) with complex traits; in large-scale gene expression analyses, one’s goal is to identify differentially expressed gene biomarkers; in cancer screening, increasing efforts are put into identification of “cancer biomarkers” – specific proteins released by tumor into blood; in magnetic resonance imaging (MRI), MRI lesions are studied as potential predictors of further disease progression, etc. In modern large-scale studies, the number of biomarkers collected can greatly exceed the number of samples. Therefore, each biomarker is typically assessed individually for its potential association with disease status. At the same time, the notion that combinations of multiple biomarkers may provide higher diagnosis accuracy is becoming widely accepted.

A number of statistical methods exists that can combine biomarker-specific summary statistics (e.g., Z-scores or P-values) and assess an association between multiple biomarkers and a trait. The Fisher combined probability test is perhaps one of the most popular ones (Fisher, 1992). Despite its wide popularity, the Fisher test relies on the assumption of independence among summary scores, while measured levels of many of the biomarkers are expected to be correlated with each other. For example, minor allele counts are expected to be correlated among SNPs and this correlation is captured by linkage disequilibrium (LD); gene expression levels are expected to be correlated in subsets of informative genes, etc. Therefore, when biomarker-specific association statistics are being combined, the correlation among biomarkers needs to be accounted for, because it induces correlation on the summary statistics.

Multiple correlated statistics can be combined using the test for quadratic forms (TQ). Under TQ, P-values for p individual-level biomarkers are first transformed into Z-scores, $P_i \rightarrow Z_i, i = 1, \dots, p$, and then combined as $TQ = \sum_{i=1}^p Z_i^2$. In this paper, we explore theoretical properties of the TQ test by deriving its distribution under the alternative hypothesis. Based on this derivation, we provide an approximation to the TQ test that is just as powerful as the original test. Our approximation will be very useful to researchers who have summary statistics for individual biomarkers available but no pair-wise correlation values. Finally, we showcase both the TQ and the approximate TQ test by applying them to data on biomarkers for inflamm-aging – an age-associated, low-grade and chronic inflammation (Franceschi et al., 2000), – and muscle mass in

senior horses.

Methods and Materials

The relationship between correlation among biomarkers and estimated biomarker effects

Let $\mathbf{y}' = (y_1, \dots, y_n)$ be an $n \times 1$ vector of trait values and $\mathbf{X}$ be a $n \times p$ matrix of measured biomarkers. Suppose that $\mathbf{R} : p \times p$ is the correlation matrix among biomarkers with $R_{ij} = \text{Cor}(X_i, X_j)$, where X_i and X_j are i th and j th columns of $\mathbf{X}$ respectively. A vector of Z-scores (estimated effect sizes divided by their standard errors) obtained by marginally testing each biomarker one at a time can be expressed in terms of the expected correlations between predictor X_j and the outcome $\mathbf{y}$ as:

$$E(\text{Cor}(\mathbf{y}, \mathbf{X})) = E(\hat{\rho}_{yx})_{(p \times 1)} = \frac{\mathbf{R}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta} + \sigma_e^2}, \quad (1)$$

where $\boldsymbol{\beta} : p \times 1$ is a vector of biomarker effects and σ_e^2 is defined by the linear model $y_i \sim N(X_i\beta, \sigma_e^2)$. If $\mathbf{R} = \mathbf{I}_p$ is an identity matrix (representing an unlikely scenario of uncorrelated biomarkers), expected univariate correlations become proportional to the vector of biomarker effects, $E(\hat{\rho}_{yx}) \propto \boldsymbol{\beta}$. That is, a zero coefficient β_i would imply zero expected correlation between biomarker X_i and the outcome. When correlation among biomarkers is present, expected univariate correlations become proportional to weighted sums of all $\boldsymbol{\beta}$, $E(\hat{\rho}_{yx}) \propto \mathbf{R}\boldsymbol{\beta}$, with weights given by the columns R_j 's. Thus, on the one hand, a biomarker with no effect on the outcome ($\beta_i = 0$) can still be correlated with it ($\rho_{yi} \neq 0$) due to the correlation with causal biomarkers. If a biomarker represents a SNPs, such non-associated SNPs are typically referred to as ‘‘proxy SNPs’’ in literature (Spencer et al., 2009). On the other hand, marginally testing a truly causal biomarkers ($\beta_i \neq 0$) may result in an estimated negligible effect ($\rho_{yi} \approx 0$) if $R_i\boldsymbol{\beta}$ is close to zero.

Construction of multiple-biomarker association P-value requires that test statistics in Eq. (1) be asymptotically distributed as multivariate normal random variables with a known correlation matrix. When none of the biomarkers are associated with a trait, the common correlation matrix among association statistics is a function of correlation among biomarkers. Specifically, if univariate regression models do not include covariates, then correlation among biomarker-level statistics will be equal to the correlation matrix among them, $\text{Cor}(R_{yi}, R_{yj}) \rightarrow \text{Cor}(X_i, X_j) = R_{ij}$. With additional

covariates that may be correlated with biomarkers (e.g., age, sex), $\text{Cor}(R_{yi}, R_{yj})$ will be a Schur complement of the submatrix $\mathbf{X}$ of the matrix of all predictor variables (Conneely and Boehnke, 2007). However, if some of the biomarkers are associated (or even if all of them are only proxies for the causal biomarkers), the problem of finding the common correlation matrix among association statistics becomes more challenging.

The relationship between correlation among biomarkers and correlation among test statistics under the alternative hypothesis

We relate the problem of estimating a common correlation matrix among association statistics to the problem of correlation among correlations, which resurfaces throughout the history of statistics in diverse applications, starting with the 1898 paper by Pearson and Filon (Pearson et al., 1898). To derive the variance-covariance matrix of the test statistics, we assume that our sample consists of n vectors from a $(p+1)$ -variate multinormal distribution, $(y_i, x_{1i}, \dots, x_{pi})'$, with covariance matrix that has the following structure:

$$\Sigma = \left[\begin{array}{c|cccc} \sigma_e^2 & \rho_{y1}\sigma_e\sigma_1 & \rho_{y2}\sigma_e\sigma_2 & \cdots & \rho_{yp}\sigma_e\sigma_p \\ \hline \rho_{y1}\sigma_e\sigma_1 & & & & \\ \rho_{y2}\sigma_e\sigma_2 & & \mathbf{R}_{p \times p} & & \\ \vdots & & & & \\ \rho_{yp}\sigma_e\sigma_p & & & & \end{array} \right], \quad (2)$$

where $(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)' = \text{diag}(\mathbf{R})$ and $(\rho_{y1}, \rho_{y2}, \dots, \rho_{yp})' = \boldsymbol{\rho}_{yx}$ are pair-wise correlations between the outcome and p biomarkers. Since summary statistics can be expressed in terms of the expected correlations (Eq. (1)), to find a covariance matrix among statistics one needs to find the covariance matrix of the correlation estimates $\boldsymbol{\mathcal{V}} = \text{Var}(\hat{\boldsymbol{\rho}}_{yx})$.

Suppose we ignore the structure in Eq. (2) and just calculate the sample covariance matrix $\mathbf{V}$. Diagonal elements of sample covariance matrix, v_{ii} , are the maximum likelihood estimates of σ_i^2 and off-diagonal elements, v_{ij} , are the maximum likelihood estimates of $\rho_{ij}\sigma_i\sigma_j$. Let $\boldsymbol{\mathcal{S}}$ be the covariance matrix of the vectorized sample covariance matrix, $\boldsymbol{\mathcal{S}} = \text{Var}(\text{vec}(\mathbf{V}))$ (Elston, 1975). The properties of a multivariate normal distribution ensures that if s_{ij} is an estimate of the (i, j) th element, c_{ij} , of the covariance matrix, then $\text{cov}(s_{ij}, s_{kl}) = \frac{1}{n}(c_{ik}c_{jl} + c_{il}c_{jk})$ (e.g., Kendall et al.

(1946)). Using this property, we can estimate the elements of $\mathbf{S}$ as follows:

$$\text{var}(v_{ii}) = \text{cov}(v_{ii}, v_{ii}) = \frac{2\sigma_i^4}{n}, \quad (3)$$

$$\text{var}(v_{ij}) = \text{cov}(v_{ij}, v_{ij}) = \frac{\sigma_i^2 \sigma_j^2}{n} (1 + \rho_{ij}^2), \quad (4)$$

$$\text{cov}(v_{ii}, v_{ij}) = \frac{\rho_{ij} \sigma_i^3 \sigma_j}{n}. \quad (5)$$

Finally, we can estimate variances and covariances of correlations using the Delta method as $\mathbf{V} = \mathbf{H}\mathbf{S}\mathbf{H}'$, where the (i, j) 'th element of $\mathbf{H}$ is the derivative of the i th estimated correlation with respect to the j th element of $\text{vec}(\mathbf{V})$, evaluated at the parameter values. For example, the first row of $\mathbf{H}$ will contain partial derivatives of $\hat{\rho}_{y1} = v_{21}/\sqrt{v_{11}v_{22}}$ with respect to all v_{ij} . Note that only 3 of these partial derivatives will have non-zero values: (1) $\frac{\partial(v_{21}/\sqrt{v_{11}v_{22}})}{\partial v_{11}} = -\rho_{y1}/(2\sigma_e^2)$, (2) $\frac{\partial(v_{21}/\sqrt{v_{11}v_{22}})}{\partial v_{22}} = -\rho_{y1}/(2\sigma_1^2)$, and (3) $\frac{\partial(v_{21}/\sqrt{v_{11}v_{22}})}{\partial v_{21}} = 1/(\sigma_e^2 \sigma_1^2)$. Similarly, the last row of $\mathbf{H}$ will contain partial derivatives of $\hat{\rho}_{yp}$ with respect to all v_{ij} . Once again, only three of these partial derivatives will have non-zero values. Therefore, the matrix $\mathbf{H}$ can be constructed as:

$$\mathbf{H} = \left[\begin{array}{cccc|cccc} -\frac{\rho_{y1}}{2\sigma_e^2} & -\frac{\rho_{y1}}{2\sigma_1^2} & \cdots & 0 & \frac{1}{\sigma_e^2 \sigma_1^2} & 0 & \cdots & 0 \\ -\frac{\rho_{y2}}{2\sigma_e^2} & 0 & -\frac{\rho_{y2}}{2\sigma_2^2} & \cdots & 0 & \frac{1}{\sigma_e^2 \sigma_2^2} & \cdots & 0 \\ \vdots & & & & & & & \vdots \\ -\frac{\rho_{yp}}{2\sigma_e^2} & 0 & \cdots & -\frac{\rho_{yp}}{2\sigma_p^2} & 0 & 0 & \cdots & \frac{1}{\sigma_e^2 \sigma_p^2} \end{array} \right].$$

After multiplying matrices and simplification, diagonal and off-diagonal elements of $\mathbf{V}$ are found as:

$$\text{var}(\hat{\rho}_{yi}) = \frac{1}{n} (1 - \rho_{yi})^2, \quad (6)$$

$$\text{cov}(\hat{\rho}_{yi}, \hat{\rho}_{yj}) = \frac{(1 - \rho_{yi}^2 - \rho_{yj}^2)(2\rho_{ij} - \rho_{yi}\rho_{yj}) + \rho_{yi}\rho_{yj}\rho_{ij}^2}{2n}. \quad (7)$$

In our original notation, $\rho_{yi} = b_i$ is the true value of the standardized regression coefficient, and $\rho_{ij} = R_{ij}$ is the correlation matrix among biomarkers. Thus,

$$\text{var}(\hat{\rho}_{yi}) = \frac{1}{n} (1 - b_i)^2, \quad (8)$$

$$\text{cov}(\hat{\rho}_{yi}, \hat{\rho}_{yj}) = \frac{(1 - b_i^2 - b_j^2)(2R_{ij} - b_i b_j) + b_i b_j R_{ij}^2}{2n}. \quad (9)$$

Further, a good approximation to covariances and correlations can be obtained via:

$$\begin{aligned}\text{cov}(\hat{\rho}_{yi}, \hat{\rho}_{yj}) &\approx \frac{1}{n} \left(R_{ij} - R_{ij}(b_i^2 + b_j^2) - (1 - R_{ij}) \frac{b_i b_j}{2} \right), \\ \text{cor}(\hat{\rho}_{yi}, \hat{\rho}_{yj}) &\approx R_{ij} + R_{ij} \frac{b_i b_j}{2} - R_{ij} b_i^2 b_j^2 - \frac{b_i b_j}{2}.\end{aligned}\quad (10)$$

Multiple Biomarker test

Multiple biomarker-based tests aggregate evidence for an association across all p biomarkers by computing a weighted sum of single-biomarker Z-scores (burden tests) or a weighted sum of squared single-biomarker Z-scores (e.g., overdispersion tests). Let $X_i^2 = w_i Z_i^2$, $i = 1, \dots, p$, denote a weighted squared Z-score. Based on the previous section, we know that $\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}_Z, \boldsymbol{\mathcal{V}})$ with $\boldsymbol{\mu}_Z = (\rho_{y1}, \dots, \rho_{yp})'$ and thus $\mathbf{WZ} \sim \text{MVN}(\mathbf{W}\boldsymbol{\mu}_Z, \mathbf{W}\boldsymbol{\mathcal{V}}\mathbf{W}')$, where $\mathbf{W} = \text{diag}(\sqrt{w_i})$. The multiple biomarker-based statistic is defined as $T = \sum_{i=1}^p X_i^2$. The values of X_i^2 's are correlated but following standard theory, the distribution of their sums can be represented as the distribution of the weighted sum of independent one degree of freedom chi-squares:

$$\Pr(\mathbf{X}'\mathbf{X} > t) = \Pr\left(\sum_{i=1}^p X_i^2 > t\right) = \Pr\left(\sum_{i=1}^p \lambda_i \chi_{1, \gamma_i}^2 > t\right), \quad (11)$$

where λ_i is the i th eigenvalue of $\mathbf{W}\boldsymbol{\mathcal{V}}\mathbf{W}'$, and γ_i is a non-centrality parameter defined as $\gamma_i = \frac{1}{\lambda_i}(\boldsymbol{\mu}'_Z \mathbf{u}_i)^2$, where $\mathbf{u}_i$ is the i th orthonormal eigenvector of $\mathbf{W}\boldsymbol{\mathcal{V}}\mathbf{W}'$. The P-value for the statistic $T = \mathbf{X}'\mathbf{X}$, calculate under the global null hypothesis of no association, is obtained by setting $\boldsymbol{\mu}_Z$ to zero, substituting $\mathbf{W}\boldsymbol{\mathcal{V}}\mathbf{W}'$ by $\mathbf{W}\mathbf{R}\mathbf{W}'$ and then calculating the tail probability at the observed value $T = t$.

Power of the multiple Biomarker test

To examine statistical power, in contrast to the null model (that is, $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, \mathbf{R})$) we need to consider the case where causal biomarker have non-zero effects (that is $\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}_Z, \boldsymbol{\mathcal{V}})$). For now, let's assume that $\boldsymbol{\mathcal{V}}$ is equicorrelated with common correlation, $\nu > 0$, for all pairs of statistics.

Then, $\mathbf{V}$ will have only two distinct eigenvalues:

$$\lambda_1 = 1 + (p-1)\nu, \quad (12)$$

$$\lambda_s = 1 - \nu, \quad s = 2, \dots, p. \quad (13)$$

The orthonormal eigenvectors can be obtained by normalizing the orthogonal basis:

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 1/2 & \cdots & 1/p \\ 1 & -1 & 1/2 & \cdots & 1/p \\ 1 & 0 & -1 & \cdots & 1/p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

The corresponding non-centralities can be derived as follows:

$$\gamma_1 = \frac{m_p^2 p}{1 + (p-1)\nu}, \quad (14)$$

$$\gamma_s = \frac{1}{(1-\nu)} \frac{(m_{s-1} - b_s)^2}{s/(s-1)}, \quad s = 2, \dots, p, \quad (15)$$

where $m_s = (1/s) \sum_{i=1}^s b_i$. Consider the behavior of non-centralities as the number of biomarkers to be jointly tested goes to infinity. On the one hand, the first non-centrality increases with p :

$$\lim_{p \rightarrow \infty} \gamma_1 = \lim_{p \rightarrow \infty} \frac{m_p^2 p}{1 + (p-1)\nu} = \frac{m_p^2}{\nu}. \quad (16)$$

On the other hand, the sum of the remaining non-centralities, $\sum_{s=2}^{\infty} \frac{1}{(1-\nu)} \frac{(m_{s-1} - b_s)^2}{s/(s-1)}$, does not diverge, which is easy to show by the ratio test. Therefore, under the null hypothesis, the distribution of the quadratic form $\mathbf{X}'\mathbf{X}$ can be well approximated by the location-scale transformation of the one degree of freedom chi-squared random variable:

$$\Pr \left\{ \frac{\mathbf{X}'\mathbf{X} - (p-1)(1-\nu)}{(p-1)\nu + 1} > \chi_\alpha^2 \right\} \approx \alpha, \quad (17)$$

where χ_α^2 is the $1 - \alpha$ quantile of the one degree of freedom chi-square distribution.

Results

Simulation study

We conducted simulation experiments to study statistical power of the proposed approximation to the TQ test (Eq. (17)) in relation to the traditional TQ test statistic (Eq. (11)). Tables 1 and 2 summarize the results.

Table 1 demonstrates type I simulation results. Simulations were performed under complete null hypothesis (i.e., $\Pr(H_0) = 100\%$) but with different correlation structures $\mathbf{R}$ among biomarkers. The value ν indicates the range of values of the off-diagonal elements of $\mathbf{R}$. For example, $\nu = 0.5$ represents the case of equi-correlated structure with all off-diagonal elements equal to 0.5, $\nu \in (-0.25, 0.75)$ represents a general correlation matrix with the range of the off-diagonal values varying between -0.25 and 0.75, $\nu \in (-0.70, 0.90)$ represents a general correlation matrix with a higher level of heterogeneity among its element. These different general correlation matrices were constructed by adding rank-one matrix $\mathbf{U}'\mathbf{U}$ to the equi-correlated matrix $\mathbf{R}_{\nu=0.5} + \mathbf{U}'\mathbf{U}$, where $\mathbf{U}$ was a vector of random numbers. Different values of p indicate the number of biomarkers that were combined. As Table 1 clear demonstrates, both TQ and its approximation have correct sizes under varying correlation structure conditions.

The first row of Table 2 indicates the first power simulation experiments setting. Specifically, the first row demonstrates that 80% of the simulated values of the test statistic came from the null hypothesis with mean zero vector ($\Pr(H_0) = 0.8$); the remaining statistics were simulated from $\mathbf{Z} \sim \text{MVN}(\mathbf{b}, \mathbf{V})$, where values of mean vector were varied uniformly in the narrow interval between -0.1 and 0.1 and the correlation matrix among biomarkers $\mathbf{R}_\nu$ was assumed to be equi-correlated with the common off-diagonal element $\nu = 0.5$. Note, since we knew the values of $\mathbf{R}$ and $\mathbf{b}$, we employed Eq. (10) to calculate $\mathbf{V}$ directly. The next row of Table 2 shows the number of biomarkers that were combined (i.e., 50, 30, 20, and 10), followed by the empirical power of TQ test (“Empiric. TQ”) based on evaluation of power by computing P-values under the null, and power of test based on the proposed approximation (i.e., Eq. (17), “Approx. TQ”). Since this simulation scenario matches the assumption of equi-correlated biomarkers, it is expected that our modification would result in a test at least as powerful as the original TQ-based test. The first few rows of Table 2 confirm this expectation.

The next set of simulation experiments was conducted under more realistic scenarios. Specifi-

cally, we no longer assumed that the observed biomarker levels were equi-correlated. Instead, we assumed an unstructured correlation matrix with off-diagonal elements varying between -0.25 and 0.75. To calculate power of the proposed Approx. TQ test, we substituted $\nu^* = \sqrt{\bar{\nu}_{ij}^2}$ instead of ν in Eq. (17). Interestingly, as Table 2 indicates, this approximation of the general correlation structure by an equi-correlated one with the off-diagonal elements being equal to the quadratic mean of pairwise correlations did not result in appreciable drop in power. Moreover, if a small number of biomarkers were combined, our approximation resulted in higher statistical power than the traditional TQ test.

To further explore power properties, we assumed that the mean values of the test statistics under the alternative hypothesis are more heterogeneous, varying from -0.2 to 0.2. Once again, the two methods performed comparably. Finally, we increase heterogeneity among off-diagonal elements of the correlation matrix among biomarkers, $\nu \in (-0.70, 0.90)$, but the two tests still had similar power performance.

The results presented in Table 2 have important implications. Specifically, based on Table 2 we can conclude that if a researcher wants to combined statistics over multiple biomarkers but does not have an access to the original data, he or she can use an average correlation value (possibly based on the previously published results) and obtained robust power values after employing our test.

Exploring the relationship between inflamm-aging and muscle mass in senior horses

For real data application, we analyzed data from a pilot study conducted at the University of Kentucky.

It is well established that senior horses, as well as elderly humans, develop an age-related low-grade chronic inflammation termed inflamm-aging (Adams et al., 2008; Franceschi et al., 2000; Horohov et al., 2010). Elevated inflammatory markers, which characterize inflamm-aging (Franceschi et al., 2007), have been associated with diseases in elderly people, including sarcopenia (Singh and Newman, 2011), a condition defined by low muscle mass or quality, and strength (Cruz-Jentoft et al., 2018). In old horses, the relationship between inflamm-aging and such conditions as low muscle mass and strength is currently not well understood. Thus, the goal of this pilot study was

to investigate the association between inflamm-aging biomarkers and muscle mass in senior horses.

Animals

Twelve senior horses (mean = 22 years, range = 15-30 years) of light breed type and mixed sex were enrolled in this study. The horses were housed at the University of Kentucky Woodford County Farm in Versailles, Kentucky, and all animal procedures were approved by the University of Kentucky Institutional Animal Care and Use Committee.

Muscle mass assessment

Muscle mass was determined using a previously designed, subjective, 5-point muscle development scoring system (5=best muscle development) (Walker et al., 2016). Muscle mass was evaluated by three individuals, on the left side of the horse, and scores were recorded for 6 regions: neck, thorax, abdomen, hindlimb, pelvis and lumbosacral region. In addition, body fat was scored by the same persons using a 9-point body condition scoring system (1=emaciated, 9=very fat) (Henneke et al., 1983), as fat was considered an important covariate to be included into statistical analysis related to the muscle scores.

Inflamm-aging markers

To determine the horses inflamm-aging status, pro- and anti-inflammatory cytokines were quantified at the mRNA and protein level, in peripheral blood and in peripheral blood mononuclear cells (PBMCs), respectively. Cytokine mRNA and protein levels have been previously used to demonstrate inflamm-aging in horses (Adams et al., 2008).

For studies of cytokine mRNA, peripheral blood was collected into TempusTM blood RNA tubes (Applied Biosystems, Foster City, CA) in accordance with the manufacturer's instructions. The mRNA was extracted with an iPrep Total RNA Kit (Invitrogen, Carlsbad, CA) in an iPrepTM RNA purification machine (Invitrogen), following the manufacturer's instructions, except for re-suspending the pelleted mRNA in 600 μ l viral lysis buffer (Invitrogen). The extracted mRNA was reverse transcribed into cDNA and gene expression was quantified in the cDNA samples using real-time quantitative polymerase chain reaction (RT-qPCR), following the protocols described in Siard (2017). RT-qPCR was conducted with commercially available TaqMan assays (Thermo Fis-

cher, Scientific, Waltham, MA) for the anti-inflammatory cytokine IL10 (Ec03468647.m1) and the pro-inflammatory cytokines IL1 β (Ec04260298.s1), IL6 (Ec03468678.m1), IFN γ (Ec03468606.m1), TNF α (Ec03467871.m1), and for β -glucuronidase (β -GUS, Ec03470630.m1), a previously described reference gene (Breathnach et al., 2006). Samples were processed in duplicates and relative gene expression (RE) was calculated using the $2^{-\Delta C_T}$ method, where $\Delta C_T = (\text{Average gene of interest } C_T - \text{Average reference gene } C_T)$ (Schmittgen and Livak, 2008). The RE values were used for statistical analysis.

For studies of cytokine production by PBMCs, blood was collected into heparinized blood tubes and PBMCs were isolated, counted and resuspended as previously described in Adams et al. (2008). The PBMCs were then plated in duplicates into 24 well plates at a density of 4×10^6 cells ml $^{-1}$ per well, and one well per horse received 25ng/ml of phorbol 12-myristate 13-acetate (PMA, Sigma-Aldrich, St. Louis, MO) and 1 μ M ionomycin (IONO, Sigma) for stimulation of cytokine production. PBMCs in both wells received 10 μ g of brefeldin A (Sigma) to stop release of protein from the golgi apparatus and thereby induce intracellular accumulation of cytokines, as required for intracellular staining. The plates were incubated for 4h at 5% CO $_2$ and 37°C, and subsequently 2 $\times$ 200 μ l of every well were plated into 96 well plates for intracellular staining of the pro-inflammatory cytokines TNF α and IFN γ , using the protocol and antibodies described in Adams et al. (2008). Once stained, the PBMCs were resuspended in 200 μ l PBS (HyClone Laboratories, Logan, UT) and analyzed on an AttuneTM NxT flow cytometer (Invitrogen). Lymphocytes were identified and gated based on forward scatter (FSC) and side scatter (SSC) and the percentage of TNF α and IFN γ positive lymphocytes after PMA/IONO stimulation was determined by selecting the fluorescence intensity of $1 \pm 0.1\%$ positively stained lymphocytes in the unstimulated wells for each horse and recording the corresponding percent TNF α or IFN γ positive lymphocytes in the stimulated wells. The percentage of TNF α and IFN γ positive lymphocytes recorded in the unstimulated wells was then subtracted from the percentages recorded in the corresponding stimulated wells. This delta value was used for statistical analysis.

Statistical Analysis

The initial discovery analyses focused on individual association between inflamm-aging markers and muscle scores. To explore this question, we conducted 6 regression analyses with primary

outcomes defined by muscle scores at the 6 regions of the horse. To account for correlation among muscle scores induced by three individuals, we fitted linear mixed-effect models with inflamm-aging markers as main predictors, age, sex, and body condition as covariates, and evaluator as a random effect.

Next, we employed the TQ test and the approximate TQ test to combine inflamm-aging statistics across 6 body locations. To determine the appropriate correlation structure among inflamm-aging-level test statistics, we performed an additional set of simulation experiments. These simulations were important because our models at the discovery stage were slightly different from the ones presented in the Methods section. That is, instead of having multiple models with the same outcome (e.g., trait) and multiple distinct but correlated predictors (i.e., biomarkers), the current analyses involved the same set of predictors but 6 different distinct and correlated outcomes. In those simulations (not currently presented) we first confirmed that the correlation among outcomes matched the correlation among statistics obtained by linear mixed-effects models. Second, we performed an additional check of the type I error rates of the TQ and the approximate TQ tests under relatively small sample size observed in this study ($n = 12$ horses). Both the TQ and the approximate TQ tests had satisfactory test sizes (0.041 and 0.043 for the TQ and the approximate TQ, respectively). All statistical analyses were conducted using R statistical software; statistical significance was set to $P = 0.05$.

Results

Table 3 presents results from the discovery stage. Columns labeled “%TNF α ⁺ Lymphocytes” and “%IFN γ ⁺ Lymphocytes” represent the percent TNF α and IFN γ positive lymphocytes after PMA/IONO stimulation and corrected for the percent recorded in the unstimulated wells. The rest of the columns represent the relative expression of pro-inflammatory cytokines (TNF α RE, IFN γ RE, IL6 RE, IL1 β RE) and the anti-inflammatory cytokine (IL10 RE) in whole peripheral blood samples. The results indicate that neck scores were negatively associated with the relative expression of IFN γ ($\beta = -840.818$, $P = 0.003$) and TNF α ($\beta = -298.039$, $P = 0.023$) and positively with the relative expression of IL6 ($\beta = 1041.320$, $P = 0.017$) and IL10 ($\beta = 905.628$, $P = 0.006$). Lumbosacral scores were positively related to the percentage of TNF α positive lymphocytes ($\beta = 0.018$, $P = 0.037$). Hindlimb scores were negatively associated with the percent IFN γ positive

lymphocytes ($\beta = -0.042$, $P = 0.008$) and the relative expression of $\text{TNF}\alpha$ ($\beta = -268.794$, $P = 0.022$), and positively with the relative expression of IL10 ($\beta = 570.252$, $P = 0.045$). Abdomen scores were positively associated with the percent IFN γ positive lymphocytes ($\beta = 0.040$, $P = 0.044$) and the relative expression of IL1 β ($\beta = 8.957$, $P = 0.013$). Pelvis and thorax scores were not significantly associated with any inflamm-aging biomarker.

To check whether inflamm-aging markers are associated with horse’s muscle mass regardless of body location, we employed the TQ and the approximate TQ tests to combine summary statistics obtained at the discovery stage. Table 4 presents the results and indicates that only the percent IFN γ positive lymphocytes is associated with muscle condition regardless of body location. This association is likely inverse, since the individual linear mixed-effects models yielded a negative association between the percent IFN γ positive lymphocytes and muscle mass at five out of six locations (Table 3).

Discussion

In this article, we studied the problem of combining multiple correlated summary statistics. This problem has recently become increasingly important in multiple areas, including problems of pleiotropy in complex traits (Solovieff et al., 2013; Zhu et al., 2015), gene and pathway scores from SNP-phenotype association summary statistics (Lamparter et al., 2016; Liu et al., 2019), prediction of individual’s diagnostic outcome based on combination of multiple biomarkers (Mamtani et al., 2006; Yan et al., 2018), etc. We studied analytical properties of the traditional approach for combining summary statistics, TQ, which includes popular methods for association analysis of genetic variants with a disease (Pasaniuc and Price, 2017), and derived the distribution of the TQ-statistic under the alternative hypothesis. Based on our derivation, we proposed a scale-location approximation of the TQ statistic, which we called approximate TQ. We showed that the approximate TQ has very similar power to the traditional TQ test under varying correlation structures among biomarkers. This finding is not trivial and has practical implications. For example, much of the interest in gene-based tests has recently centered on the simulation-based methods such as VEGAS (Liu et al., 2010). Without access to individual-level data, simulation-based methods incorporate LD information from reference panels (e.g., HamPMap2 (Consortium et al., 2007) or 1,000 Genomes data (Consortium et al., 2012)) to account for the correlation among statistics.

However, sensitivity of these methods to misspecification of the LD pattern were poorly understood. Our results suggest that the use of reference panels in the absence of genotype data or even just a single value of quadratic mean of pairwise correlations results in adequate power and type I error rates of the combination test.

Based on previous studies investigating the relationship between inflammatory markers and muscle mass in other species (Toth et al., 2006; Visser et al., 2002; Wang et al., 2014), we expected to observe a positive association between the anti-inflammatory marker (IL10 RE) and muscle mass, and a negative association between the pro-inflammatory markers (%IFN γ ⁺ and %TNF α ⁺ lymphocytes, and IL6 RE, IL1 β RE, TNF α RE and IFN γ RE) and muscle mass. The initial discovery analysis showed, however, that pro-inflammatory markers were both positively and negatively associated to muscle scores, and that the biomarker associations were not consistent across all scoring locations. This observation might illustrate the complex relationship between muscle mass and inflamm-aging markers, and suggest that this relationship could be indirect or non-causative; especially for blood-level inflamm-aging markers, which might be less intimately linked to muscle mass than inflammatory markers in the muscle (Wilson et al., 2017). In addition, the mRNA and in vitro study-based inflamm-aging markers used in this pilot work might be less closely related to muscle mass than circulating cytokine proteins, which have been found associated to muscle mass in previous studies (Toth et al., 2006; Visser et al., 2002). The observed associations are furthermore complicated by the fact that IL6, considered a pro-inflammatory cytokine in this study, can have both pro- and anti-inflammatory effects, depending on the receptor it engages (Scheller et al., 2011).

Results from TQ and approximate TQ analysis showed a significant, and likely inverse, association between the percent IFN γ positive lymphocytes and muscle mass across all scoring locations. This finding does not corroborate the results of a previous study investigating the relationship of muscle mass and inflamm-aging in senior horses. In this previous study the percent IFN γ positive lymphocytes was tested for correlation with overall muscle mass as well, and no significant association was found (Siard, 2017). These contradicting results might originate from different muscle mass assessment methods used. While we applied a muscle development scoring system leading to six regional scores (Walker et al., 2016), Siard (2017) used a picture based whole body muscle scoring system (Graham-Thiers and Kronfeld, 2005), and ultrasound and deuterium oxide dilution

based fat free mass estimations (Dugdale et al., 2011; Lehnhard et al., 2004). Furthermore, different outcomes of these two studies may be due to analysis of the data using different statistical methods, as well as using different age-range of horses studied (range: 15-30 yrs in the current pilot study, vs range: 18-29 yrs in Siard (2017)).

To our knowledge this is the first study suggesting a potential relationship between inflamm-aging and muscle mass in senior horses. However, given the contradictory findings in another study, and the ambiguous association of markers depending on the scoring location, further investigation is imperative. We also note that the traditional statistical approaches (e.g., linear regression and linear mixed-effects models) could only explore the relationship between inflamm-aging and muscle mass at one assessment location at a time, due to the relatively small sample size (i.e., one would run out of the degrees-of-freedom if were to incorporate all muscle mass assessment locations into a single model). The combination test provided an elegant solution to this statistical issue. We stress that our application is one of many data sets typically encountered in modern research, when the number of biomarkers (usually several thousands) greatly exceeds the number of samples (usually in hundreds or less) and a single full model can not be built. Furthermore, the TQ and the approximate TQ are generally applicable to situations, in which correlated summary statistics need to be combined into an overall score. The power of individual summary statistics can often be weak but they can re-enforce the effect of one another in a combination test, improving the overall statistical power.

Author contributions

DZ and OV performed theoretical derivations; AH, BC, CB, DB and AA conducted the pilot study and DZ, OV, AH and AA drafted the manuscript.

Funding

This research was supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences and by the University of Kentucky Veterinary Science Department.

References

- Adams, A., Breathnach, C., Katepalli, M., Kohler, K., and Horohov, D. (2008). Advanced age in horses affects divisional history of t cells and inflammatory cytokine production. *Mechanisms of ageing and development*, 129(11):656–664.
- Breathnach, C., Sturgill-Wright, T., Stiltner, J., Adams, A., Lunn, D., and Horohov, D. (2006). Foals are interferon gamma-deficient at birth. *Veterinary immunology and immunopathology*, 112(3-4):199–209.
- Conneely, K. N. and Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of P-values for multiple correlated tests. *The American Journal of Human Genetics*, 81(6):1158–1168.
- Consortium, . G. P. et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56.
- Consortium, I. H. et al. (2007). A second generation human haplotype map of over 3.1 million snps. *Nature*, 449(7164):851.
- Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A. A., et al. (2018). Sarcopenia: revised european consensus on definition and diagnosis. *Age and ageing*, 48(1):16–31.
- Dugdale, A., Curtis, G., Milne, E., Harris, P., and Argo, C. M. (2011). Assessment of body fat in the pony: Part ii. validation of the deuterium oxide dilution technique for the measurement of body fat. *Equine veterinary journal*, 43(5):562–570.
- Elston, R. (1975). On the correlation between correlations. *Biometrika*, 62(1):133–140.
- Fisher, R. A. (1992). Statistical methods for research workers. In *Breakthroughs in statistics*, pages 66–70. Springer.
- Franceschi, C., Bonafè, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., and De Benedictis, G. (2000). Inflamm-aging: an evolutionary perspective on immunosenescence. *Annals of the new York Academy of Sciences*, 908(1):244–254.
- Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M. P., Invidia, L., Celani, L., Scurti, M., et al. (2007). Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mechanisms of ageing and development*, 128(1):92–105.
- Graham-Thiers, P. and Kronfeld, D. (2005). Amino acid supplementation improves muscle mass in aged and young horses. *Journal of animal science*, 83(12):2783–2788.
- Henneke, D., Potter, G., Kreider, J., and Yeates, B. (1983). Relationship between condition score, physical measurements and body fat percentage in mares. *Equine veterinary journal*, 15(4):371–372.
- Horohov, D., Adams, A., and Chambers, T. (2010). Immunosenescence of the equine immune system. *Journal of comparative pathology*, 142:S78–S84.
- Kendall, M. G. et al. (1946). The advanced theory of statistics. *The advanced theory of statistics*, 2nd Ed., 3(2).

- Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., and Bergmann, S. (2016). Fast and rigorous computation of gene and pathway scores from snp-based summary statistics. *PLoS computational biology*, 12(1):e1004714.
- Lehnhard, R. A., McKeever, K. H., Kearns, C. F., and Beekley, M. D. (2004). Myosin heavy chain profiles and body composition are different in old versus young standardbred mares. *The Veterinary Journal*, 167(1):59–66.
- Liu, J. Z., Mcrae, A. F., Nyholt, D. R., Medland, S. E., Wray, N. R., Brown, K. M., Hayward, N. K., Montgomery, G. W., Visscher, P. M., Martin, N. G., et al. (2010). A versatile gene-based test for genome-wide association studies. *The American Journal of Human Genetics*, 87(1):139–145.
- Liu, Y., Chen, S., Li, Z., Morrison, A. C., Boerwinkle, E., and Lin, X. (2019). Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics*, 104(3):410–421.
- Mamtani, M. R., Thakre, T. P., Kalkonde, M. Y., Amin, M. A., Kalkonde, Y. V., Amin, A. P., and Kulkarni, H. (2006). A simple method to combine multiple molecular biomarkers for dichotomous diagnostic classification. *BMC bioinformatics*, 7(1):442.
- Pasaniuc, B. and Price, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*, 18(2):117.
- Pearson, K., Filon, L. N. G., et al. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Proceedings of the Royal Society of London*, 62(379-387):173–176.
- Scheller, J., Chalaris, A., Schmidt-Arras, D., and Rose-John, S. (2011). The pro-and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1813(5):878–888.
- Schmittgen, T. D. and Livak, K. J. (2008). Analyzing real-time pcr data by the comparative c t method. *Nature protocols*, 3(6):1101.
- Siard, M. H. (2017). *Lymphocyte-mediated Inflamm-aging in the Horse*. PhD thesis, University of Kentucky.
- Singh, T. and Newman, A. B. (2011). Inflammatory markers in population studies of aging. *Ageing research reviews*, 10(3):319–329.
- Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, 14(7):483.
- Spencer, C. C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS genetics*, 5(5):e1000477.
- Toth, M. J., Ades, P. A., Tischler, M. D., Tracy, R. P., and LeWinter, M. M. (2006). Immune activation is associated with reduced skeletal muscle mass and physical function in chronic heart failure. *International journal of cardiology*, 109(2):179–187.
- Visser, M., Pahor, M., Taaffe, D. R., Goodpaster, B. H., Simonsick, E. M., Newman, A. B., Nevitt, M., and Harris, T. B. (2002). Relationship of interleukin-6 and tumor necrosis factor- α with muscle mass and muscle strength in elderly men and women: the health abc study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 57(5):M326–M332.

- Walker, V., Tranquille, C., Dyson, S., Spear, J., and Murray, R. (2016). Association of a subjective muscle score with increased angles of flexion during sitting trot in dressage horses. *Journal of Equine Veterinary Science*, 40:6–15.
- Wang, B., Yang, G., Liang, X., Zhu, M., and Du, M. (2014). Grape seed extract prevents skeletal muscle wasting in interleukin 10 knockout mice. *BMC complementary and alternative medicine*, 14(1):162.
- Wilson, D., Jackson, T., Sapey, E., and Lord, J. M. (2017). Frailty and sarcopenia: the potential role of an aged immune system. *Ageing research reviews*, 36:1–10.
- Yan, Q., Bantis, L. E., Stanford, J. L., and Feng, Z. (2018). Combining multiple biomarkers linearly to maximize the partial area under the roc curve. *Statistics in medicine*, 37(4):627–642.
- Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., Smith, J. A., Yanek, L. R., Sun, Y. V., Edwards, T. L., et al. (2015). Meta-analysis of correlated traits via summary statistics from gwass with an application in hypertension. *The American Journal of Human Genetics*, 96(1):21–36.

Tables

$\Pr(H_0) = 100\%, \nu = 0.5$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.047	0.051	0.047	0.053
Approx. TQ	0.047	0.054	0.049	0.051
$\Pr(H_0) = 100\%, \nu \in (-0.25, 0.75)$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.044	0.052	0.045	0.045
Approx. TQ	0.046	0.052	0.044	0.046
$\Pr(H_0) = 100\%, \nu \in (-0.70, 0.90)$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.053	0.048	0.055	0.048
Approx. TQ	0.052	0.049	0.055	0.058

Table 1: **Type I error of TQ and approximate TQ methods.** Simulations were performed under complete null hypothesis ($\Pr(H_0) = 100\%$) but with different correlation structures among biomarkers. The value ν indicates the range of values of the off-diagonal elements of the correlation matrix $\mathbf{R}$ among biomarkers. Specifically, $\nu = 0.5$ represents the case of equi-correlated structure with all off-diagonal elements equal to 0.5; $\nu \in (-0.25, 0.75)$ represents a general correlation matrix with the range of the off-diagonal values varying between -0.25 and 0.75; $\nu \in (-0.70, 0.90)$ represents a general correlation matrix with a higher level of heterogeneity among its element. Different values of p indicate the number of biomarkers that were combined.

$\Pr(H_0) = 80\%, b_i \sim \text{Unif}(-0.1, 0.1), \nu = 0.5$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.537	0.440	0.372	0.261
Approx. TQ	0.536	0.448	0.377	0.263
$\Pr(H_0) = 80\%, b_i \sim \text{Unif}(-0.1, 0.1), \nu \in (-0.25, 0.75)$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.594	0.492	0.395	0.254
Approx. TQ	0.602	0.485	0.392	0.267
$\Pr(H_0) = 80\%, b_i \sim \text{Unif}(-0.2, 0.2), \nu \in (-0.25, 0.75)$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.861	0.798	0.717	0.588
Approx. TQ	0.865	0.800	0.729	0.582
$\Pr(H_0) = 80\%, b_i \sim \text{Unif}(-0.2, 0.2), \nu \in (-0.70, 0.90)$				
	$p = 50$	$p = 30$	$p = 20$	$p = 10$
Empiric. TQ	0.879	0.820	0.757	0.616
Approx. TQ	0.876	0.819	0.755	0.609

Table 2: **Power comparison of TQ and approximate TQ methods.** Simulations were performed under the assumption that 80% of the measured biomarkers were not associated with the trait ($\Pr(H_0) = 80\%$). Effect sizes of biomarkers that were truly associated were drawn from a uniform distribution within bounds that either corresponded to little effect size heterogeneity (i.e., -0.1, 0.1) or higher effect size heterogeneity (i.e., -0.2, 0.2). Correlation structure among biomarkers followed the same pattern as in Type I error simulations.

Location	Inflammatory Markers													
	%TNF α ⁺ Lymphocytes		%IFN γ ⁺ Lymphocytes		TNF α RE		IFN γ RE		IL6 RE		IL10 RE		IL1 β RE	
	β	P-value	β	P-value	β	P-value	β	P-value	β	P-value	β	P-value	β	P-value
Neck	-0.014	0.078	-0.03	0.078	-298.039	0.023	-840.818	0.003	1041.32	0.017	905.628	0.006	3.637	0.232
Lumbosacral	0.018	0.037	-0.021	0.228	6.144	0.961	-86.030	0.746	296.655	0.478	39.376	0.900	-1.679	0.581
Hindlimb	0.005	0.502	-0.042	0.008	-268.794	0.022	-441.42	0.066	433.769	0.241	570.252	0.045	2.448	0.362
Abdomen	-0.003	0.741	0.040	0.044	-41.405	0.774	53.068	0.859	534.718	0.256	-146.099	0.680	8.957	0.013
Pelvis	0.014	0.059	-0.023	0.136	6.941	0.951	102.851	0.662	-208.343	0.570	-87.896	0.752	-2.874	0.285
Thorax	0.003	0.717	-0.015	0.361	-214.396	0.076	-380.295	0.130	609.627	0.124	463.021	0.117	3.386	0.237

Table 3: **Inflamm-aging statistics across 6 body locations obtained at the discovery stage.** Individual-level associations indicate that neck scores were negatively associated to the TNF α , and the IFN γ RE values, and positively to the IL6 and IL10 RE values; lumbosacral scores were positively related to the percent TNF α positive lymphocytes; hindlimb scores were negatively associated to the percent IFN γ positive lymphocytes and the TNF α and IFN γ RE values, and positively to the IL10 RE values. Abdomen scores were positively associated with the percent IFN γ positive lymphocytes and the IL1 β RE values.

Marker	TQ	Approx. TQ
%TNF α ⁺ Lymphocytes	0.1278	0.1226
%IFN γ ⁺ Lymphocytes	0.0435	0.0435
TNF α RE	0.0938	0.0912
IFN γ RE	0.0784	0.0767
IL6 RE	0.1311	0.1257
IL10 RE	0.0846	0.0826
IL1 RE/ β	0.1373	0.1313

Table 4: **Overall association results.** Both the TQ and the Approx. TQ tests suggest that only the percentage of IFN γ positive lymphocytes is associated with overall muscle condition, regardless of body location.