

ORIGINAL RESEARCH ARTICLE

Open Access



The Association of Aquaporin-1 Gene with Marathon Running Performance Level: a Confirmatory Study Conducted in Male Hispanic Marathon Runners

Miguel A. Rivera^{1*} , Thomas D. Fahey², Juan R. López-Taylor³ and Juan L. Martínez⁴

Abstract

Background: Replication studies are essential for identifying credible associations between alleles and phenotypes. Validation of genotype-phenotype associations in the sports and exercise field is rare. An initial genetic association study suggested that rs1049305 (C > G) in the 3' untranslated region (3'UTR) of the aquaporin-1 (AQP1) gene was associated with marathon running (MR) performance level in Hispanic males. To validate this finding, we conducted a replication analysis in an independent case-control sample of Hispanic male marathon runners ($n = 1430$; cases $n = 713$ and controls $n = 717$). A meta-analysis was utilized to test the extent of the association between the initial results and the present report. It also provided to test the heterogeneity (variation) between the two studies.

Results: The replication study showed a statistically significant ($p \leq 0.05$) association between rs1049305 (C > G) of the AQP1 gene and MR performance level. Association test results using a fixed effect model for the combined, original study and the present report, yielded an odds ratio = 1.28, 95% confidence interval = 1.13–1.45, $p = 0.0001$. The extent of the measures of heterogeneity was Tau-squared = 0, H statistic = 1, I^2 statistic = 0, and Cochran's Q test ($Q = 0.29$; p value 0.59), indicated the variation between studies were due to chance and not to differences in heterogeneity between the two studies. Within the limitations of the present replication, contrast of two studies and its effects on meta-analysis, the findings were robust.

Conclusion: This study successfully replicated the results of Martínez et al. (*Med Sportiva* 13:251-5, 2009). The meta-analysis provided further epidemiological credibility for the hypothesis of association between the DNA rs1049305 (C > G) variation in the 3'UTR of the AQP1 gene and MR running performance level in Hispanics male marathon runners. It is not precluded that a linked DNA structure in the surrounding molecular neighborhood could be of influence by been part of the overly complex phenotype of MR performance level.

Keywords: Meta-analysis, Genetic epidemiology, Marathon, Humans, Men, Cochran's Q test, Exercise

* Correspondence: miguel.rivera10@upr.edu

¹Department of Physical Medicine, Rehabilitation & Sports Medicine, School of Medicine, University of Puerto Rico, Main Building Office A204, San Juan, PR 00936, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Key Points

- Replication studies of genotype-phenotype associations in the sports and exercise field literature are rare.
- Replication of findings from genetic association studies is a fundamental assessment process in research endeavors leading to new knowledge.
- *AQP1* gene coexistent molecular mechanisms and their effects on marathon running performance phenotypes are major areas of inquiry in the science and medicine of sport and exercise.

Introduction

Aquaporins (AQPs) are integral membrane pore proteins, also known as water channels. The aquaporin-1 (AQP1) channel is the best known and most studied of the AQP family [1]. The AQP1 channel is encoded by the aquaporin-1 (AQP1) gene, on chromosome 7, region p14 [2]. The expression of AQP1 occurs in various tissues, including red blood cells, endothelial cells, and smooth, skeletal, and cardiac muscle [1, 3]. The major AQP of the cardiovascular system is AQP1, which regulates water permeability of the heart's capillary networks by mediating the flow of water through the endothelial layer into the blood.

Evidence from a recent systematic review [4] on the association between AQP1 and endurance performance (EP) indicated that the first effort to evaluate the hypothesis of association between the AQP1 gene DNA sequence variant rs1049305 (C > G) in the 3' untranslated region (3' UTR) and marathon running (MR) performance level was conducted by Martinez et al. [5]. Genotypic and allelic frequency distribution revealed significant differences (χ^2 , $P = 0.005$) between cases (fast marathon runners—the top three % of finishers) and controls (slow marathon runners—the low three % of finishers). The calculated odds ratio (OR) = 1.35. Its 95% confidence interval boundary was 1.08 to 1.67. The C-allele was more ($P = 0.005$) prevalent in the cases (i.e., fast runners) than in the controls (slow runners). To date, these data have not been replicated.

The confirmation of a genetic-association finding is vital to identifying true positive genetic variants that may contribute to complex phenotypes, such as MR performance level [6]. To confirm findings and make generalizations, the reproduction of results from genetic-association studies should occur using independent samples [7]. Along those lines, an independent study [8] confirmed the original findings of the association between MR performance level and the rs1049305 (C > G), variant within the 3' UTR of the AQP1 gene. This time, the association was evaluated in South African Caucasian male Ironman triathletes ($n = 504$). The findings revealed that the AQP1 rs1049305 C-variant was associated with the duration of the MR segment in three Ironman events. Triathletes who carried the

C-allele completed the 42.2-km run stage faster (mean 286, $s = 49$ min) than triathletes with the GG genotype (mean 296, $s = 47$ min; $P = 0.032$).

The same association was found in another observational study using a genetic epidemiology model [9]. In that study, elapsed running time in a 10-km event was compared by AQP1 C-allele carrier status, e.g., carriers (homozygous for C-allele (CC) and heterozygous for C-allele (CG); $n = 50$) versus non-carriers (homozygous for G-allele (GG); $n = 41$). That study revealed that AQP1 C-allele carriers ran faster ($P < 0.05$) than non-carriers during the 10-km race. Those findings to a certain extent validated [10] the association between endurance running performance and the rs1049305 (C > G) variant within the 3' UTR region of the AQP1 gene. The 3' UTR is a significant participant in regulating gene expression by controlling nuclear export, subcellular targeting, and rates of translation and degradation of DNA. These functions make the 3' UTR an essential tool for regulating phenotypic diversity of in humans [11]. The genes controlled by the sequence of the 3' UTR are generally regulatory proteins, and their irregular expression may have severe effects on humans [11, 12].

To date, no exact replication [10, 13] of the first reported association [5] between the rs1049305 (C > G), in the 3' UTR of the AQP1 gene and MR performance level in Hispanic male marathon runners that has been reported. The concept of exact replication implies the examination of the same genetic variant in an independent sample drawn from the same population as the initial study, using a similar design and methods [10, 13]. The related literature rarely evidence exact replication studies examining the association of genetic markers and MR performance level using comparable subjects, similar methods, and procedures as those of the original publications. Therefore, the purposes of the present research were (1) to verify the hypothesis of association between the AQP1 rs1049305 (C > G) genetic marker and CE MR performance level in Hispanic male marathon runners [5] and (2) to exactly replicate the original study by Martinez et al. [5]. We adhered to established criteria [14, 15] and gathered a larger sample size (adequate statistical power), recruited, screened, and evaluated in an identical approach as that of the initial study [5].

Methods

This observational study adheres to a genetic epidemiology model using a case-control design. All methods and procedures were identical to those described by Martinez et al. [5].

Subjects

The subjects in the present study ($n = 1430$; cases $n = 713$ and controls $n = 717$) were male biologically

unrelated volunteers, recruited during the activities of marathon (42 km) running events conducted between 2006 and 2012: Maratón de Culiacan, Gran Maratón Pacifico, Mazatlán, and Guadalajara International Marathon all held in México. The physical characteristics of the cases and controls are shown in Table 1.

The recruitment process included an interview in determining the ethnic and geographic area of origin, place of residence, history of endurance events participation, cardiovascular disease, respiratory disease, arthritis activity, history of metabolic disease, and the use of related medications.

Exclusion Criteria

Subjects with a history of cardiovascular, respiratory, arthritis, metabolic disease, and taking-related medications were excluded. Subjects reporting any incident that negatively affected the performance during the race (e.g., significant injury) were also excluded. Subjects that participated in the initial study [5] were also excluded from the analysis.

Subjects' Classification

The official results of the competitions determined the classification as cases or controls. Cases ($n = 713$) were the top 3% of finishers of their respective age group, while controls ($n = 717$) were the lowest 3% of their respective age group.

Ethical Standards

The study was carried out following the World Medical Association Declaration of Helsinki ethical principles for research involving human subjects [16]. The protocol of the study was approved by the Committee on Human Research of the Instituto Sinaloense del Deporte in Culiacan, Mexico and the Medical Commissions from each competition. Each participant provided a voluntary written informed consent to be part of the study.

DNA Isolation and Genotype Determination

The Promega Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI) provided for the purification of genomic DNA from 10 mL of whole blood. The detailed procedures were those in the original study [5]. The amplification of the AQP1 polymorphism (restriction site (rs) #1049305) followed according to the previous procedure [5]. A systematic

random sample of DNA aliquots followed the completion of all the genotyping analyses. The selected DNA samples were submitted to a new round of PCR and genotyping, as described in the original study [5]. We performed a new genotyping cycle analysis if a mismatch occurred between the first and second genotyping. All work (DNA extraction, storage and management, and PCR and genotyping) were performed at the facilities of the Exercise Physiology and Human Genomics Laboratory of the Department of Physical Medicine, Rehabilitation and Sports Medicine at the School of Medicine of the University of Puerto Rico, San Juan.

Statistical Analysis

The statistical analysis of the actual replication study was identical to those in the original study [5]. The polymorphism information content (PIC) and the heterozygosity index (H_I), both measures of polymorphism, followed the procedures described [17, 18]. The Pearson chi-square test was used to examine differences in genotype and allele frequencies between cases and controls and to determine whether the genotypes prevalence distribution differed from that expected under the Hardy-Weinberg equilibrium (HWE). Odds ratios with 95% confidence interval and the chi-square tests were performed using the Epi Info™ 3.5.1 software package [19], which was the software application utilized in the original study [5]. The present and original study statistical analysis were corroborated using the most recent Epi Info™ software package version 7.2.2.6 released on February 2, 2018. A meta-analysis [20] was performed to test the extent of the association between the results of Martinez et al. [5] and the present report and to test for heterogeneity between the two studies. Heterogeneity (H) in meta-analysis refers to the variation in study outcomes between studies. P values ≤ 0.05 were statistically significant.

Results

Subjects

The present replication study involved 1430 subjects (cases $n = 713$ and controls $n = 717$) while the initial study by Martinez et al. [5] included 784 subjects (cases $n = 396$ and controls $n = 388$). This replication study used 646 more subjects than the original study, which reduced the probability of a type II error.

Genotype Frequencies

As shown in Table 2, the three expected AQP1 rs1049305 genotypes (CC, CG, GG) were observed in both cases and control groups. These genotype frequencies agreed with those expected under HWE (χ^2 , $P \geq 0.05$). chi-square test revealed a statistically significant ($p = 0.02$) difference in the genotype frequency distribution between the cases

Table 1 Characteristics of 1430 biologically unrelated marathon runners in the present replication study

| Group | Age (years) | Height (cm) | Mass (kg) |
|------------------------|-------------|--------------|-------------|
| Cases ($n = 713$) | 32 \pm 18 | 173 \pm 15 | 62 \pm 7 |
| Controls ($n = 717$) | 39 \pm 19 | 175 \pm 11 | 68 \pm 12 |

Values are mean \pm standard deviation

Table 2 Allele and genotype frequencies of the *Aquaporin-1* gene rs1049305 (C > G) variant within the 3' untranslated region in cases (fast marathon runners) and controls (slow marathon runners) in the present replication study

| | ^c Genotype frequencies | | | ^d Allele frequencies | | Odds ratio | 95% Confidence interval | P value |
|---------------------------------|-----------------------------------|------------|------------|---------------------------------|------------|------------|-------------------------|---------|
| | CC (%) | CG (%) | GG (%) | C (%) | G (%) | | | |
| ^a Cases (n = 713) | 107 (0.15) | 314 (0.44) | 292 (0.41) | 528 (0.37) | 898 (0.63) | | | |
| ^b Controls (n = 717) | 79 (0.11) | 301 (0.42) | 337 (0.47) | 459 (0.34) | 547 (0.66) | 1.43 | 1.21–1.68 | < 0.05 |

^aHardy-Weinberg equilibrium = 2.1, P = 0.14

^bHardy-Weinberg equilibrium = 0.9, P = 0.34

^c $\chi^2 = 7.7$, df = 2, P = 0.02

^d $\chi^2 = 17.7$, df = 1, P < 0.001

and controls. The present result replicates and validates the original study [5], see Table 3, and corroborates those of others [8, 9].

Allelic Frequencies

Table 2 also shows the observed allelic frequency distribution in cases and controls. The allelic frequency distributions revealed statistically significant (P < 0.001) differences between the cases and controls. The C-allele was the less frequently observed in both cases and controls.

Allelic Frequencies by C-allele Carrier Status

In the present study, the prevalence of C-allele carrier status was 59% (107 + 314/713) for the cases and 53% (79 + 301/717) for the controls. The calculated odds ratio (OR = 1.43) and its 95% confidence interval (1.21–1.68), suggested that the C-allele was more likely prevalent (p < 0.05) in the cases than in the controls. In the original study, 57% of the cases were carriers of the C-allele versus 49% for the controls. The computed OR = 1.35 and its 95% confidence interval extended from 1.08 to 1.67; p < 0.05. These findings also replicated and validated those of the original study [5].

AQP1 Polymorphisms

Table 4 shows the calculated values of the H_I and PIC for the original study [5] and those of the present report were the same for both measures of polymorphism. These data replicate the findings of the original study. H_I is a (expected) probability measure of the likelihood that an individual will be heterozygous at the AQP1 rs1049305. Also, H_I is a central measure in the study of

genetic variation in populations. For two allele loci, like the one represented by the AQP1 gene rs1049305 (C > G), H_I peaks at a value of 0.5.

PIC value is an indicator of the probability that a marker locus (AQP1 rs1049305) is informative. It is estimated from the frequency of heterozygotes, determined by the number of marker alleles (allele C and allele G), and their respective frequencies.

Table 5 shows the meta-analysis association that combined data from Martinez et al. [5] and the present study. The findings in Table 5 are the outcome of a fixed effect model (FE). A principal assumption of FE meta-analysis is that the effect size is alike within the studies examined by the meta-analysis. The unadjusted and adjusted and p values, 0.0001 and 0.0007, for the combined FE association test were highly significant. This finding affirms the association between the AQP1 gene rs1049305 (C > G) and MR performance level. Figure 1 illustrates a forest plot showing the present findings, the findings of the original study, and the combine results of the meta-analysis for the association between the AQP1 gene rs1049305 (C > G) and MR performance.

Meta-Analysis Test of Heterogeneity

The meta-analysis also provided four indicators of heterogeneity (variation) between the two studies under consideration. The extent of these four measures of heterogeneity is shown in Table 6.

Tau-Squared (Tau²)

Tau² value reflects the between-study variance in our meta-analysis. Tau² is a useful measure because it is

Table 3 Allele and genotype frequencies of the *Aquaporin-1* gene rs1049305 (C > G) variant within the 3' untranslated region in cases (fast marathon runners) and controls (slow marathon runners) as reported in the original study [5]

| | ^c Genotype frequencies | | | ^d Allele frequencies | | Odds ratio | 95% Confidence interval | P value |
|---------------------------------|-----------------------------------|------------|------------|---------------------------------|------------|------------|-------------------------|---------|
| | CC (%) | CG (%) | GG (%) | C (%) | G (%) | | | |
| ^a Cases (n = 396) | 59 (0.15) | 167 (0.42) | 170 (0.43) | 285 (0.36) | 506 (0.64) | | | |
| ^b Controls (n = 388) | 39 (0.10) | 151 (0.39) | 198 (0.51) | 229 (0.30) | 547 (0.70) | 1.35 | 1.08–1.67 | < 0.005 |

^aHardy-Weinberg equilibrium = 0.09

^bHardy-Weinberg equilibrium = 0.21

^c $\chi^2 = 6.94$, df = 2, P = 0.03

^d $\chi^2 = 7.55$ df = 1, P = 0.005

Table 4 Heterozygosity (H_i) and polymorphism information content (PIC) from original [5] and present exact replication studies

| Group | | Original study | Present study |
|----------|-------|----------------|---------------|
| Cases | H_i | 0.46 | 0.47 |
| | PIC | 0.35 | 0.36 |
| Controls | H_i | 0.42 | 0.43 |
| | PIC | 0.33 | 0.34 |

sensitive to the effect size and not sensitive to the number of studies in the analysis. We can infer the present calculated Tau^2 suggested that the total heterogeneity between studies = 0.

H Statistic

Within the context of meta-analysis, H statistic (H) is the proportion of total variability due to heterogeneity. H does not depend on sample size and theoretically it ranges from a lowest value of one to infinity. A value of $H = 1$ represents complete homogeneity. In the present study, $H = 1$.

I^2 Statistic (I^2)

The I^2 statistic is a descriptive value and not a measure of the real variation across the underlying true effects. Its scale ranges from 0 to 100%. I^2 statistic is a choice test to quantify heterogeneity and calculates the percentage of variation across studies that is due to heterogeneity rather than chance. The present result of $I^2 = 0\%$ indicated that all variability in effect size estimates is due to sampling error (chance) and does not require further explanation.

Cochran’s Q Test (Q)

The purpose of the Cochran’s Q test (Q) statistic is to demonstrate heterogeneity but does not reveal its extent. It only indicates the statistical significance. The I^2 statistic quantifies the magnitude of such heterogeneity. Q value is used to determine if variations are due to chance or they are due to differences between studies. Q is also called chi-square test for heterogeneity or homogeneity. A weakness of the Q statistic is its reduced power to detect true heterogeneity among studies when the meta-analysis includes a few studies. The present replication study is based on two studies. As a counterweight for the low power of Q , we opted to test for heterogeneity using an alpha level of 0.10, rather than 0.05, which

Table 5 Meta-analysis association test results for the combined study of Martinez et al. [5] and the present report

| Model | OR | 95% CI | p value | adjusted p value |
|--------------|------|-----------|-----------|--------------------|
| Fixed effect | 1.28 | 1.13–1.45 | 0.0001 | 0.0007 |

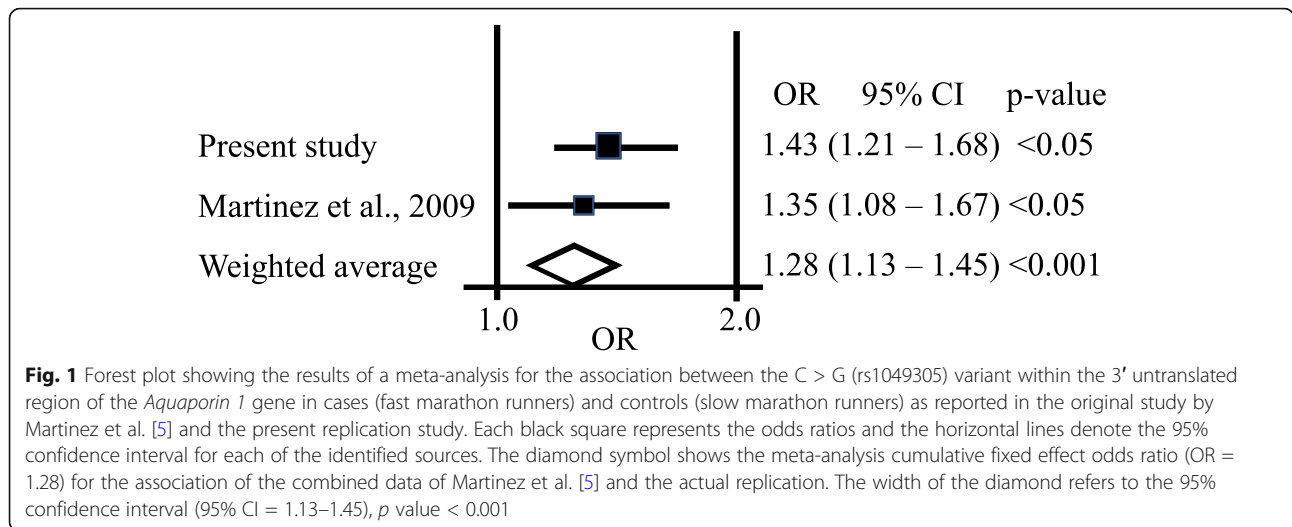
increases the probability of finding heterogeneity. In the present study, the observed value for $Q = 0.29$; p value 0.59. Since the observed p value was higher than 0.10, then we assume variations are due to chance and not differences between studies.

Discussion

The strong association between AQP1 gene sequence variant rs1049305 (C to G) and marathon running is impressive, and we can infer that it promotes whole-body fluid regulation that affects many aspects of performance level. Regulation of water flow across cell membranes is essential for maintaining an appropriate fluid balance within the cell. Physical activity and exercise disturb muscle cell homeostasis. Depending on the exact nature of the stress, skeletal muscle may adapt by increasing size, improving neuromuscular performance, or increasing endurance [21]. If each cell is to complete the reactions demanded for marathon running performance at optimal speed, movement, and endurance, the body must have adequate access to fluid [22]. AQP1 is present in the plasma membrane of muscle fibers and promotes water transport between the blood and myofibrils during muscle activity [23]. This water transport is essential since some of the functions of water include transporting nutrients and aiding in their digestion and absorption, repairing and replacing old tissues, flushing the system of toxic wastes, maintaining a constant body temperature through sweating, thermal capacitance, and blood flow regulation as well as lubricating the joints and tissues [24].

An adequate body fluid balance may improve blood flow through the smallest vessels, which will enhance oxygen delivery to the working muscles [25]. Metabolic adaptations in skeletal muscle, such as optimizing fluid balance, are critical for marathon running performance. Deficiencies in cell fluid management are likely detrimental to marathon running capacity and performance level. Therefore, the fine regulation of water flow by the AQP1 channel as influenced by DNA sequence variations in the AQP1 gene could explain the observed differences in the genotype prevalence between fast and slow marathon runners.

Replication of findings from genetic association studies is a fundamental assessment process in research endeavors leading to new knowledge in genetic epidemiology and related scientific fields. The main functions of replication are the verification of a research finding or piece of knowledge, control of sampling error, control of artifacts, control of fraud, the generalization of results to a different or broader population, and the verification of the initial research hypothesis [26, 27]. The aim is to obtain the consistent results. The process of acceptance of new knowledge considers the reproducibility of new



findings rather than the “authority of their originators” [26]. Some have suggested that the criteria for a successful replication study should include the same genetic variant, the same direction of the association, the same definition of the phenotype, and the same ethnic group as reported in the original study [14, 15]. Other factors include the assurance of adequate statistical power, a sound epidemiological design aimed at minimizing biases, highly trained technicians, and topic area familiarized investigators.

Some have argued that there is no such thing as an exact replication [26]. The main argument is there are always differences between the original study and the attempted replication. For instance, the date, setting of the experiment, the experimenters, ethnic group, minor differences in reagents, and the execution of experimental protocols, these differences are classified as obvious or subtle. Therefore, efforts to repeat a methodology do not represent an exact replication [26].

The main finding of the present study was the replication of the association between the DNA sequence variant rs1049305 (C > G) in the 3'UTR of the AQP1 gene and MR performance level in Hispanic male marathon runners. The present exact replication process observed the same rigor, methods, means, ways, ethnic group, and personnel as described in the original study [5]. Compliance with the latter observations is critical in the exact replication of genetic associations [13, 15]. Previous studies [8, 9] provided support to the notion that the AQP1 gene could be part of a highly intricate network

of molecular, cellular, and biological components that trigger functions and processes specific to CE running performance in humans. Perhaps the extreme complexity of such network interaction, within the human biology and environment variability domain, had been a factor in the lack of consistent results examining candidate genes by others [28, 29], while examining the association of genetic markers and MR as well as endurance performance domain across populations.

Replication or corroboration studies using comparable subjects and research design as those used in the initial study are rarely conducted in exercise and sports science research. Replication and corroboration attempts have violated the principle of specific physiological adaptations by training modality and or sports discipline. For instance, attempts to replicate genetic associations with candidate genes observed in endurance modalities (e.g., long-distance running) by examining subjects in team sports or activities with different metabolic demands. In the latter cases, such factor could lead to failure to replicate or even corroborate findings [28, 29]. Failure to replicate or corroborate findings is not necessarily an indicator for rejecting the hypothesis of genetic associations. The literature is rich in those arguments [30–33].

The present replication study assessed the validity of the findings of Martinez et al [5]. To enhance the power of comparisons, we performed a meta-analysis combining the results of both studies in addition to performing individual study comparisons. This approach provided four indicators of heterogeneity, which is a critical factor leading to the validation of our results. Factors like statistical power and sample size are always an issue in genetic epidemiology studies and in other fields. Within the limitations of the present replication study, we can argue that the findings are robust.

Table 6 Test of heterogeneity (variation) results between the study of Martinez et al. [5] and the present report

| Tau-squared (tau ²) | <i>H</i> statistic (<i>H</i>) | <i>I</i> ² statistic (<i>I</i> ²) | Cochran's <i>Q</i> test (<i>Q</i>) | <i>p</i> value |
|---------------------------------|---------------------------------|---|--------------------------------------|----------------|
| 0 | 1 | 0 | 0.29 | 0.59 |

Conclusion

The exact replication of the initial findings of Martinez et al. [5] is highly relevant and provides strong support for an association between the AQP1 gene and MR performance level. The findings further contribute to the body of information suggesting the possibility that AQP1 or close molecular genetic mechanism is part of the complex biological processes that determines MR performance level and long-distance running performance. The greater sample size in the present replication study versus the original publication ($\Delta = 646$) results in a greater statistical power than in the initial study. This study examined a highly specific human sample and a favorable statistical power. We are in the early stages of deciphering the complex polygenic biology (molecular, biochemical, physiological, and mental) underlying MR performance level. The study of AQP1 deserves further detailed systematic scrutiny.

Abbreviations

3'UTR: 3' Untranslated region; AQP1: Aquaporin-1 channel; AQP1: Aquaporin-1 gene; AQPs: Aquaporins; C > G : C to G nucleotide change; CC: Homozygous for C-allele; CG: Heterozygous for C-allele; CI: Confidence interval; CO₂: Carbon dioxide; DNA: Deoxyribonucleic acid; FE: Fixed effect model; GG: Homozygous for G-allele; H: Heterogeneity; HgCl₂-: Magnesium chloride; H_i: Heterozygosity index; HWE: Hardy-Weinberg equilibrium; I^2 : I^2 statistic; MR: Marathon running; OR: Odds ratio; PCR: Polymerase chain reaction; PIC: Polymorphism information content; Q: Cochran's Q test; rs: Reference single nucleotide polymorphism cluster identification number; SNPs: Single nucleotide polymorphisms; Tau²: Tau-squared statistic

Acknowledgements

Not applicable

Authors' Contributions

JLT, JLM, and MAR contributed to the study design, implementation, and data collection. MAR coordinated the blood sample analysis and genotyping. Also undertook the data management and analysis. TDF contributed to the interpretation of the results. MAR conceived the paper which was critically reviewed and revised by TDF. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Funding

The authors declare that they have no funding sources to disclose.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate

The study received the approval of the Medical Commissions from Maratón de Culiacán, Gran Maratón Pacífico, Mazatlán, and Guadalajara International Marathon. All participants provided written informed consent. The study followed the ethical standards for sport and exercise science research, as described in Harriss DJ, Macsween A, Atkinson G. Standards for ethics in sport and exercise science research: 2018 update. *Int J Sports Med.* 2017;38(14):1126-31.

Consent for Publication

Not applicable

Competing Interests

The authors, Miguel A. Rivera, Thomas D. Fahey, Juan R. López-Taylor, and Juan L. Martínez, declare that they have no competing interests.

Author details

¹Department of Physical Medicine, Rehabilitation & Sports Medicine, School of Medicine, University of Puerto Rico, Main Building Office A204, San Juan, PR 00936, USA. ²Department of Kinesiology, California State University, Chico, CA, USA. ³Physical Activity and Applied Sport Sciences Institute, Universidad de Guadalajara, Guadalajara, Jalisco, México. ⁴Hospital Civil de Culiacán, Culiacán, Sinaloa, México.

Received: 8 October 2019 Accepted: 19 February 2020

Published online: 20 March 2020

References

- Day RE, Kitchen P, Owen DS, Bland C, Marshall L, Conner AC, et al. Human aquaporins: regulators of transcellular water flow. *Biochim Biophys Acta.* 2014;1840(5):1492–506.
- Deen PMT, Weghuis DO, Geurs van Kessel A, Wieringa B, van Os CH. The human gene for water channel aquaporin 1 (AQP1) is localized on chromosome 7p15-p14. *Cytogenet Cell Genet.* 1994;65(4):243–6.
- Au CG, Cooper ST, Lo HP, Compton AG, Yang N, Wintour EM, et al. Expression of aquaporin 1 in human cardiac and skeletal muscle. *J Mol Cell Cardiol.* 2004;36:655–62.
- Rivera MA, Fahey TD. Association between aquaporin-1 and endurance performance: a systematic review. *Sports Med Open.* 2019;5(40):1–12.
- Martínez JL, Carrión A, Florián ME, Martín JA, López-Taylor JR, Fahey TD, et al. Aquaporin-1 gene DNA variation predicts performance in Hispanic marathon runners. *Med Sportiva.* 2009;13(4):251–5.
- Karaesmen E, Rizvi AA, Preus LM, McCarthy PL, Pasquini MC, Onel K, et al. Replication and validation of genetic polymorphisms associated with survival after allogeneic blood or marrow transplant. *Blood.* 2017;130(13):1585–96.
- Colhoun HM, McKeigue PM, Davey SG. Problems of reporting genetic associations with complex outcomes. *Lancet.* 2003;361(9360):865–72.
- Saunders CJ, Posthumus M, O'Connell K, September AV, Collins M. A variant within the AQP1 3'-untranslated region is associated with running performance, but not weight changes, during an Ironman Triathlon. *J Sports Sci.* 2015;33(13):1342–8.
- Rivera MA, Martínez JL, Carrion AA, Fahey TD. AQP-1 association with body fluid loss in 10-km runners. *Int J Sports Med.* 2011;32(3):223–9.
- König IR. Validation in genetic association studies. *Brief Bioinform.* 2011;12(3):253–8.
- Mayr C. Regulation by 3'-untranslated regions. *Annu Rev Genet.* 2017;51(1):171–94.
- Conne B, Stutz A, Vassalli J-D. The 3' untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat Med.* 2000;6(6):637–41.
- Baldassarre MT, Carver J, Dieste Tubio O, Juristo Juzgado N. Replication types: towards a shared taxonomy. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014). New York: Association for Computing Machinery; 2014.
- Aslibekyan S, Claas SA, Arnett DK. To replicate or not to replicate: the case of pharmacogenetic studies: establishing validity of pharmacogenomic findings: from replication to triangulation. *Circ Cardiovasc Genet.* 2013;6(4):409–12.
- Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. *Nat.* 2007;447(7145):655–60.
- World Medical Association Declaration. of Helsinki ethical principles for medical research involving human subjects. *JAMA.* 2013;310:2191–4.
- Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet.* 1980;32(3):314–31.
- Ott J. Analysis of human genetic linkage. 3rd ed. Baltimore: John Hopkins University Press; 1999.
- Dean AG, Arner TG, Sunki GG, Friedman R, Lantinga M, Sangam S, et al. Epi Info™, a database and statistics program for public health professionals. 7.2 ed. Atlanta, GA, USA: CDC; 2011.
- Martorell-Marugan J, Toro-Dominguez D, Alarcon-Riquelme ME, Carmona-Saez P. MetaGenyo: A web tool for meta-analysis of genetic association studies. *BMC Bioinformatics.* 2017;18(563):1–6.
- Kraemer WJ, A. SB. Skeletal muscle physiology: plasticity and responses to exercise. *Hormone Res.* 2006;66:2–16.

22. von Duvillard SP, Braun WA, Markofski M, Beneke R, Leithäuser R. Fluids and hydration in prolonged endurance performance. *Nutrition*. 2004;20(7-8):651–6.
23. Frigeri A, Nicchia GP, Balena R, Nico B, Svelto M. Aquaporins in skeletal muscle: reassessment of the functional role of aquaporin-4. *FASEB J*. 2004; 18(7):905–7.
24. National Center for Biotechnology Information. Reference SNP cluster report rs1049305. 2018. Available from: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1049305#Diversity. Accessed 15 Feb 2019.
25. Brooks G, Fahey T. K. B. *Exercise Physiology. Human Bioenergetics and its Applications*. 5th ed. New York, NY: Amazon; 2019.
26. Nosek BA, Errington TM. Making sense of replications. *eLife*. 2017;6(6): e23383.
27. Schmidt S. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev Gen Psychol*. 2009;13(2):90–100.
28. Puthuchery Z, Skipworth JRA, Rawal J, Loosemore M, Someren KV, Montgomery HE. The ACE gene and human performance: 12 years on. *Sports Med*. 2011;41(6):433–48.
29. Zarebska A, Sawczyn S, Kaczmarczyk M, Ficek K, Maciejewska-Karłowska A, Sawczuk M, et al. Association of rs699 (M235T) polymorphism in the AGT gene with power but not endurance athlete status. *J Strength Cond Res*. 2013;27(10):2898–903.
30. Hart AB, de Wit H, Palmer AA. Candidate gene studies of a promising intermediate phenotype: failure to replicate. *Neuropsychopharmacology*. 2013;38(5):802–16.
31. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295–300.
32. Tucker HC. The significance of no-replication of gene-phenotypes associations. In: Hood KE, Tucker Halpen C, Greenburg G, Lerner RM, editors. *Handbook of developmental science, behavior and genetics*. Hoboken: Wiley-Blackwell; 2010. p. 466–90.
33. Wilkening S, Chen B, Lorenzo Bermejo J, Canzian F. Is there still a need for candidate gene approaches in the era of genome-wide association studies? *Genomics*. 2009;93:415–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
