Collagen gene interactions and endurance running performance

K O'Connell,[1] BSc (Hons); **M Posthumus**,[1] PhD; **M Collins**,[1,2] PhD

[1] *MRC/UCT Research Unit for Exercise Science and Sports Medicine, University of Cape Town, South Africa*
[2] *Medical Research Council, Cape Town, South Africa*

Corresponding author: *M Collins (malcolm.collins@uct.ac.za)*

Background. Although variants within genes that encode protein components of several biological systems have been associated with athletic performance, limited studies have investigated the collagen genes that encode the structural components of connective tissues.
Objective. To investigate the association of variants within collagen genes with endurance performance in South African (SA) Ironman triathletes.
Methods. A total of 661 white, male participants were recruited from four SA Ironman triathlon events for this genetic case-control association study. All participants were genotyped for *COL3A1* rs1800255 (G/A) and *COL12A1* rs970547 (A/G).
Results. No independent associations were identified between *COL3A1* rs1800255 and *COL12A1* rs970547 and overall finishing time or time to complete any of the individual components (3.8 km swim, 180 km bike or 42.2 km run) of the 226 km event. The major G+A-inferred pseudo-haplotype, constructed from *COL3A1* rs1800255 and *COL12A1* rs970547, was, however, significantly (p=0.010 and p=0.027) over-represented in the fast run tertile (58.7%) compared with the middle (53.5%) and slow (49.5%) run tertiles, respectively. The major G+T+A-inferred pseudo-haplotype, constructed from *COL3A1* rs1800255, *COL5A1* rs12722 (T/C) and *COL12A1* rs970547, was again significantly (p=0.022) over-represented in the fast run tertile (35.2%) compared with the slow run tertile (28.9%).
Conclusion. Our main novel finding was that the *COL3A1* rs1800255 and *COL12A1* rs970547 variants interacted to modulate endurance running performance in the four SA Ironman triathlons investigated. In addition, the interaction between these variants and *COL5A1* rs12722 appeared to modulate endurance running performance.

S Afr J SM 2014;26(1):9-14. DOI:10.7196/SAJSM.523

The *COL5A1* and *COL6A1* genes encode the α1 chains of types V and VI collagen, respectively.[1-3] Both types V and VI collagen are known to regulate collagen fibrillogenesis.[5-7] Furthermore, the *COL5A1* TT genotype of single nucleotide polymorphism (SNP) rs12722 C/T and the *COL6A1* TT genotype of SNP rs35796750 T/C have been associated with improved endurance running and endurance cycling performance, respectively, during the South African (SA) Ironman triathlon.[4] The association between the *COL5A1* rs12722 TT and rs71746744 (-/AGGG) AGGG/AGGG genotypes and improved endurance running performance was later replicated in a road running event.[4] In addition, it has been proposed that both *COL5A1* variants, located in a functional region of the COL5A1 3'-untranslated region (UTR), regulate type V collagen production.[8] Specifically, the rs12722 T and rs71746744 AGGG allele of *COL5A1* are associated with increased COL5A1 mRNA stability, which may lead to increased levels of type V collagen α1 chain synthesis.[8] Increased type V collagen production may affect normal collagen fibrillogenesis and alter the mechanical properties of the tissue, leading to improved endurance performance.[9]

Similarly to types V and VI collagen, types III and XII are also implicated in fibrillogenesis.[6,7,10-12] The α1 chains of types III and XII collagen are encoded by the *COL3A1* and *COL12A1* genes, respectively. The non-synonymous *COL3A1* rs1800255 A/G and *COL12A1* rs970547 A/G variants within these genes are also associated with a number of multifactorial soft tissue phenotypes.[4,13,14] Furthermore, *COL3A1* rs1800255 and *COL12A1* rs970547 are both proposed to be functional.[14,15] Specifically, the alanine to threonine change at position 698 of the α1(III) chain, as a result of *COL3A1* rs1800255, could affect the tensile strength of type III collagen fibres.[14] In addition, functional bioinformatics analysis of *COL12A1* rs970547 revealed that the resulting glycine to serine change is potentially damaging to the α1(XII) chain.[15] Therefore, since types III and XII are implicated in fibrillogenesis like types V and VI, it may be proposed that common, potentially functional variants within the *COL3A1* and *COL12A1* genes may also be associated with athletic endurance performance.

Objectives

The primary objective of our study was to determine whether *COL3A1* rs1800255 and *COL12A1* rs970547, like *COL5A1* rs12722 and *COL6A1* rs35796750, are associated with athletic endurance performance in the participants of four SA Ironman triathlon events. We hypothesised, due to the proposed functional effects of these variants, that the *COL3A1* rs1800255 GG and *COL12A1* rs970547 AA genotypes are associated with improved endurance performance.

The secondary objective was to investigate gene-gene interactions between *COL3A1* rs1800255 and *COL12A1* rs970547, and previously associated collagen genes where appropriate, and endurance performance. We hypothesised that the G+A pseudo-haplotype is associated with improved endurance performance, and that the *COL5A1* rs12722 T and *COL6A1* rs35796750 T alleles, if included in gene-gene interactions

with *COL3A1* rs1800255 and *COL12A1* rs970547, contributes to interactions for endurance running and cycling, respectively.

Methods

A total of 661 white, male participants were recruited from four SA Ironman triathlon events for this genetic case-control association study, using previously outlined recommendations.[16,17] Participants were recruited at the registration of either the 2000 (n=96) and 2001 (n=294) events held in Gordon's Bay (~50 km from Cape Town) or the 2006 (n=219) and 2007 (n=52) Port Elizabeth (PE) events (~750 km east of Cape Town). All participants were required to complete the event for inclusion in the study. For participants who entered more than one event, only data from one race year was used, since their overall finishing times were similar (data not shown).

Race results were obtained from the race organisers and participants were divided into three equal tertiles based on their finishing times for the 3.8 km swim, 180 km cycle, 42.2 km run and overall race. The fastest triathletes were placed into the fast tertile, those who finished in the mid-field were placed in the middle tertile, and the slowest triathletes were placed into the slow tertile.

Study approval was granted by the Human Research Ethics Committee, Faculty of Health Sciences, University of Cape Town, and the race organisers. All participants completed informed consent forms and a physical activity questionnaire. Participants of the PE subgroup completed training history questionnaires; this was not documented at the events in Gordon's Bay. Since training data were obtained during the PE events, the event priority for the participants who had completed more than one event was 2006, followed by 2007 and finally 2001, which had a larger, more complete dataset than the 2000 event.

Blood collection and DNA extraction

At event registration, ~4.5 ml of venous blood was collected from each participant into an ethylenediaminetetraacetic acid vacutainer tube by venipuncture of a forearm vein. Samples were stored at 4°C until DNA was extracted, as previously described, with minor modifications.[18] All analyses were performed at the UCT/MRC Research Unit for Exercise Science and Sports Medicine, University of Cape Town.

COL3A1 rs1800255 genotyping

Genotyping of *COL3A1* rs1800255 was performed using a custom-designed, fluorescence-based Taqman polymerase chain reaction (PCR) assay (Applied Biosystems, USA). Allele-specific probes and flanking primer sets (sequences available on request) were used along with a pre-made PCR mastermix containing ampliTaq DNA polymerase Gold (Applied Biosystems, USA) in a final reaction volume of 8 µl. The PCR cycling comprised a 10 min heat activation step (95°C) followed by 40 cycles of 15 s at 92°C and 1 min at 60°C. The reactions were performed using a XP Thermal Cycler (Block model XP-G, BIOER Technology Co., Japan). Genotypes were determined by end-point fluorescence using a 7900 HT Fast Real-Time PCR System and SDS software (version 2.3).

COL12A1 rs970547 genotyping

COL12A1 rs970547 was genotyped as previously described.[19] Briefly, fragments containing *COL12A1* rs970547 were amplified by PCR. The PCR products were digested with *Alu*I to produce 599 and 16 bp fragments for the G allele and 460, 139 and 16 bp fragments for the A allele. The fragments were resolved, together with a 100 bp DNA ladder, on a 6% non-denaturing polyacrylamide gel and visualised by SYBER Gold staining (Invitrogen Molecular ProbesTM, USA). The gels were photographed under ultraviolet light using a Uvitec photo-documentation system (Uvitec Limited, UK).

Statistics

Continuous variables were compared between genotype groups using one-way analysis of variance (ANOVA) tests. Chi-squared or Fisher's tests were used to compare categorical variables. Basic descriptive statistical analysis and frequencies were determined using Statistica (version 11) and GraphPad InStat (version 6). Inferred pseudo-haplotypes between gene variants were tested using Hapstat (version 3.0). Hardy-Weinberg equilibrium status was determined using Genepop (version 4.0.10; http://genepop.curtin.edu.au). Statistical significance was assumed at $p<0.05$.

Results

Participant characteristics

Mean ± standard deviation (SD) participant age, height, weight and body mass index (BMI) were 36.1±8.3 years (n=659), 180.5±6.6 cm (n=559), 78.6±9.4 kg (n=586) and 24.0±2.3 kg/m^2 (n=555), respectively. Approximately 65% were SA-born and 81% were SA residents at the time of recruitment. The general characteristics of participants from each event are reported in Table S1 (online supplementary material). Participants from the Gordon's Bay events (2000 and 2001; mean±SD age 34.7±7.9 years; n=390) were significantly younger ($p<0.001$) than those recruited from the PE events (2006 and 2007; mean±SD age 38.2±8.4 years; n=269) (Table S1). Significantly fewer ($p=0.002$ and $p=0.037$) SA-born participants competed in the 2000 (53%; n=50) and 2001 (64%; n=185) events than in the 2006 event (73.3%; n=118) (Table S1). Significantly fewer ($p=0.002$) SA-resident participants competed in the 2001 event (75.7%; n=215) than in the 2006 event (86.8%; n=190) (Table S1). No genotype effects were identified between any of the participant characteristics and the *COL3A1* rs1800255 or *COL12A1* rs970547 variants (data not shown). Both *COL3A1* rs1800255 ($p=0.428$) and *COL12A1* rs970547 ($p=0.062$) were in Hardy-Weinberg equilibrium.

Participant training history

Table 1 summarises self-reported training history data, characterising the 15 weeks prior to each event, collected at the 2006 and 2007 PE SA Ironman triathlon events. Although probably not biologically relevant, the *COL3A1* rs1800255 variant was significantly ($p=0.002$) associated with swim training duration (h/week). Participants with a *COL3A1* rs1800255 GA genotype (3.4±1.6 h/week) trained significantly ($p=0.001$) more than participants with a *COL3A1* rs1800255 GG (2.8±1.0 h/week) or AA (2.3±0.9 h/week) genotype. The distance (km/week) and duration (h/week) trained for the cycle, run and combined components (swim, cycle and run) were not significantly associated with *COL3A1* rs1800255 (Table 1). Furthermore, no significant associations were identified between *COL12A1* rs970547 and distance or duration trained for the swim, cycle, run or combined tertiles (Table 1).

COL3A1 rs1800255 and *COL12A1* rs970547 and performance

The *COL3A1* rs1800255 and *COL12A1* rs970547 variants were not significantly associated with overall finishing time or time taken to

Table 1. Self-reported training history for the *COL3A1* rs1800255 and *COL12A1* rs970547 genotypes of the PE subgroup

		COL3A1 rs1800255			
Variable*	**All (*N*=187)**	**GG (*N*=97)**	**GA (*N*=73)**	**AA (*N*=17)**	***p*-value**
Training (km/week), mean±SD (*n*)					
Swim	6.4±3.0 (185)	6.4±2.8 (95)	6.5±3.2 (73)	6.1±3.1 (17)	0.857
Cycle	224.3±84.9 (170)	218.2±92.9 (85)	232.0±76.5 (69)	223.8±76.2 (16)	0.606
Run	45.7±18.0 (182)	47.0±20.4 (92)	44.8±13.3 (73)	42.4±21.6 (17)	0.535
Combined†	236.9±85.8 (155)	230.2±95.6 (76)	245.3±75.2 (63)	235.3±77.3 (16)	0.588
Training (h/week), mean±SD (*n*)					
Swim	3.0±1.3 (184)	2.8±1.0 (97)	3.4±1.6 (70)	2.3±0.9 (17)	0.002*
Cycle	8.1±2.9 (171)	8.1±3.2 (89)	8.2±2.5 (66)	7.8±2.9 (16)	0.875
Run	4.5±1.7 (174)	4.6±1.9 (92)	4.5±1.4 (65)	4.1±1.8 (17)	0.575
Combined†	15.4±4.8 (162)	15.5±4.8 (85)	15.6±3.7 (61)	14.3±4.5 (16)	0.568
		***COL12A1* rs970547**			
Variable*	**All (*N*=207)**	**AA (*N*=120)**	**AG (*N*=82)**	**GG (*N*=5)**	***p*-value**
Training (km/week), mean±SD (*n*)					
Swim	6.4±3.0 (207)	6.3±2.8 (120)	6.7±3.2 (82)	5.6±1.5 (5)	0.531
Cycle	222.7±83.3 (190)	216.9±83.6 (110)	231.6±82.8 (75)	218.7±90.4 (5)	0.499
Run	46.3±17.8 (203)	44.9±16.6 (117)	48.5±19.6 (81)	44.0±9.6 (5)	0.371
Combined†	236.5±83.6 (174)	229.8±82.6 (97)	246.1±84.8 (72)	229.5±89.9 (5)	0.449
Training (h/week), mean±SD (*n*)					
Swim	3.0±1.5 (206)	3.0±1.7 (119)	3.1±1.4 (82)	2.3±0.7 (5)	0.513
Cycle	8.9±12.8 (190)	9.5±16.6 (110)	8.2±2.9 (75)	7.6±2.1 (5)	0.752
Run	4.9±3.1 (194)	4.9±3.7 (113)	4.8±2.0 (76)	4.3±0.4 (5)	0.896
Combined†	15.5±4.4 (180)	15.3±4.3 (103)	15.8±4.7 (72)	14.2±2.3 (5)	0.607

PE = Port Elizabeth; SD = standard deviation.

* Statistically significant ($p<0.05$).

† Combined = swim, cycle and run.

Table 2. Finishing times for the *COL3A1* rs1800255 and *COL12A1* rs970547 genotypes in the 3.8 km swim, 180 km cycle, 42.2 km run and overall

		COL3A1 rs1800255 genotype			
Triathlon component	**All (*N*=642)**	**GG (*N*=333)**	**GA (*N*=265)**	**AA (*N*=44)**	***p*-value**
3.8 km swim (min), mean±SD (*n*)	77.2±17.4 (629)	77.4±17.3 (326)	76.6±17.6 (259)	79.7±17.2 (44)	0.535
180 km cycle (min), mean±SD (*n*)	393.7±42.0 (615)	394.0±42.7 (317)	392.2±41.6 (254)	399.4±39.9 (44)	0.565
42.2 km run (min), mean±SD (*n*)	288.4±49.1 (620)	285.9±50.2 (324)	289.9±47.4 (253)	297.8±51.1 (43)	0.264
Overall (min), mean±SD (*n*)	767.8±95.2 (642)	765.5±96.7 (333)	767.4±93.6 (265)	787.1±94.2 (44)	0.369
		***COL12A1* rs970547 genotype**			
Triathlon component	**All (*N*=629)**	**AA (*N*=344)**	**AG (*N*=255)**	**GG (*N*=30)**	***p*-value**
3.8 km swim (min), mean±SD (*n*)	78.5±17.6 (614)	78.8±17.1 (334)	78.2±17.9 (251)	76.9±20.4 (29)	0.800
180 km cycle (min), mean±SD (*n*)	395.2±41.5 (600)	393.8±39.8 (330)	397.6±43.2 (243)	391.6±47.3 (27)	0.504
42.2 km run (min), mean±SD (*n*)	290.1±49.7 (609)	289.0±50.7 (331)	290.8±49.3 (249)	297.0±40.7 (29)	0.681
Overall (min), mean±SD (*n*)	771.8±94.7 (629)	768.8±91.4 (344)	775.3±98.9 (255)	776.5±97.8 (30)	0.677

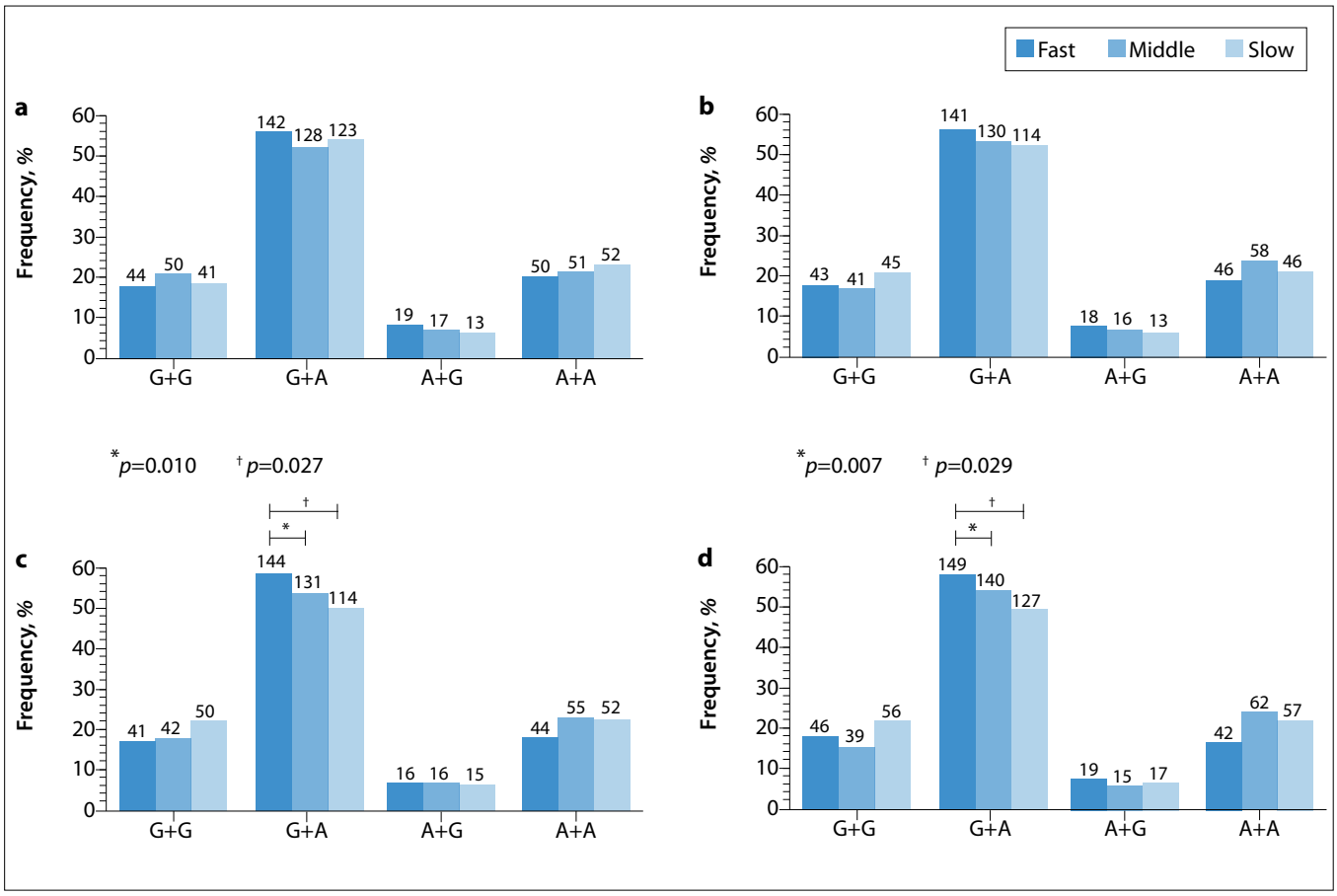

Fig. 1. Frequency distributions of inferred pseudo-haplotypes constructed from COL3A1 *rs1800255 and* COL12A1 *rs970547 between the fast, middle and slow tertiles in terms of: (a) time taken to complete the swim component of the triathlon; (b) time taken to complete the cycling component of the triathlon; (c) time taken to complete the run component of the triathlon; and (d) overall time taken to complete the triathlon. The number of participants is indicated above each column. (* Fast v. slow tertile; † Fast v. middle tertile.)*

complete any of the individual components (3.8 km swim, 180 km cycle or 42.2 km run) of the 226 km triathlon (Table 2). Furthermore, when participants were grouped into performance tertiles, no significant differences were identified for *COL3A1* rs1800255 or *COL12A1* rs970547 genotype distributions between the groups in terms of the overall finishing time or time taken to complete any of the individual components of the triathlon (Table 3).

Gene-gene interactions and performance

Since there were no independent associations of the *COL3A1* and *COL12A1* variants with endurance performance, inferred pseudo-haplotypes between *COL3A1* rs1800255 G/A and *COL12A1* rs970547 A/G were constructed. All four inferred pseudo-haplotypes were identified for the overall finishing time, as well as for the time taken to complete the individual components of the triathlon. For the overall tertiles, the major G+A-inferred pseudo-haplotype was significantly (p=0.007 and p=0.029) over-represented in the fast tertile (58%; n=149) when compared with the middle (55%; n=140) and slow (50%; n=127) tertiles, respectively (Fig. 1d). When the individual components of the triathlon were analysed, the major G+A-inferred pseudo-haplotype was significantly (p=0.010 and p=0.027) over-represented in the fast run tertile (58.7%; n=144) when compared with the middle (54%, n=131) and slow (50%; n=114) run tertiles, respectively (Fig. 1c). No significant associations were identified between the inferred pseudo-haplotypes and the swim (Fig. 1a) or cycling (Fig. 1b) components of the triathlon.

Since this association was identified for the run component of the triathlon, and *COL5A1* rs12722 was previously associated with the run component in this cohort,[4] inferred pseudo-haplotypes between *COL3A1* rs1800255 G/A, *COL5A1* rs12722 T/C and *COL12A1* rs970547 A/G were constructed (Fig. 2). All eight inferred pseudo-haplotypes were identified. The major G+T+A-inferred pseudo-haplotype was again significantly (p=0.022) over-represented in the fast run tertile (35%; n=86) compared with the slow run tertile (29%; n=67) (Fig. 2).

Furthermore, when the cycling component of the triathlon was investigated with inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547, no significant associations were identified (Fig. 3).

Discussion

The main novel finding of this study was that the *COL3A1* rs1800255 (G/A) and *COL12A1* rs970547 (A/G) variants interacted to modulate endurance running performance in the four SA Ironman triathlon events. No significant independent associations were identified between these gene variants and the time taken to complete the overall race, or the 3.8 km swim, 180 km cycle or 42.2 km run components.

Table 3. Performance tertiles for the *COL3A1* rs1800255 and *COL12A1* rs970547 genotypes in the 3.8 km swim, 180 km cycle, 42.2 km run and overall

	COL3A1 rs1800255 genotype, % (*n*)				*COL12A1* rs970547 genotype, % (*n*)			
3.8 km swim	**GG (*N*=333)**	**GA (*N*=265)**	**AA (*N*=44)**	***p*-value**	**AA (*N*=344)**	**AG (*N*=255)**	**GG (*N*=30)**	***p*-value**
Fast	52.7 (119)	40.7 (92)	6.6 (15)	0.803	54.8 (115)	41.0 (86)	4.2 (9)	0.704
Middle	50.7 (104)	43.4 (89)	5.9 (12)		50.8 (102)	44.2 (89)	5.0 (10)	
Slow	52.0 (103)	39.4 (78)	8.6 (17)		57.6 (117)	37.4 (76)	5.0 (10)	
	COL3A1 rs1800255 genotype, % (*n*)				*COL12A1* rs970547 genotype, % (*n*)			
180 km cycle	**GG (*N*=333)**	**GA (*N*=265)**	**AA (*N*=44)**	***p*-value**	**AA (*N*=344)**	**AG (*N*=255)**	**GG (*N*=30)**	***p*-value**
Fast	53.0 (115)	41.9 (91)	5.1 (11)	0.506	55.9 (114)	39.2 (80)	4.9 (10)	0.796
Middle	48.4 (103)	42.7 (91)	8.9 (19)		56.7 (118)	39.9 (83)	3.4 (7)	
Slow	53.5 (99)	38.9 (72)	7.6 (14)		52.4 (98)	42.3 (80)	5.3 (10)	
	COL3A1 rs1800255 genotype, % (*n*)				*COL12A1* rs970547 genotype, % (*n*)			
42.2 km run	**GG (*N*=333)**	**GA (*N*=265)**	**AA (*N*=44)**	***p*-value**	**AA (*N*=344)**	**AG (*N*=255)**	**GG (*N*=30)**	***p*-value**
Fast	56.4 (119)	38.4 (104)	5.2 (11)	0.565	56.4 (114)	40.1 (81)	3.5 (7)	0.461
Middle	50.0 (104)	41.8 (87)	8.2 (17)		56.5 (117)	39.1 (81)	4.4 (9)	
Slow	50.3 (101)	42.3 (85)	7.5 (15)		50.0 (100)	43.5 (87)	6.5 (13)	
	COL3A1 rs1800255 genotype, % (*n*)				*COL12A1* rs970547 genotype, % (*n*)			
226 km overall	**GG (*N*=333)**	**GA (*N*=265)**	**AA (*N*=44)**	***p*-value**	**AA (*N*=344)**	**AG (*N*=255)**	**GG (*N*=30)**	***p*-value**
Fast	56.6 (120)	39.2 (83)	4.2 (9)	0.277	54.2 (109)	41.3 (83)	4.5 (9)	0.185
Middle	48.4 (103)	43.1 (92)	8.5 (18)		60.6 (126)	36.1 (75)	3.3 (7)	
Slow	50.7 (110)	41.5 (90)	7.8 (17)		49.6 (109)	44.1 (97)	6.3 (14)	

Previously, we showed the association of *COL5A1* rs12722 (T/C) and *COL6A1* rs35796750 (T/C) with endurance running and endurance cycling performance, respectively, in the SA Ironman triathlon.[4] Furthermore, variants within the *COL5A1* 3'-UTR, including rs12722, are proposed to alter the expression of type V collagen, thereby modulating normal fibrillogenesis and resulting in changes to the collagen fibril architecture, structure and mechanical properties.[9] Similarly, *COL6A1* rs35796750 is proposed to result in aberrant splicing of COL6A1 mRNA, which may also affect the role of type VI collagen in normal fibrillogenesis.[20] Both type III and XII collagens are also implicated in fibrillogenesis.[10,12] Like *COL5A1* and *COL6A1*, common variants within the *COL3A1* and *COL12A1* genes are associated with soft tissue phenotypes[13,14,19] and the proteins that these genes encode are implicated in fibrillogenesis.[6,7,10-12] Therefore, we proposed that common variants within the *COL3A1* and *COL12A1* genes, namely rs1800255 and rs35796750, could be associated with endurance performance in the SA Ironman triathlons, in a similar manner proposed for *COL5A1* rs12722 and *COL6A1* rs35796750.

Despite the rationale outlined above, no independent associations were identified between *COL3A1* rs1800255 or *COL12A1* rs970547 and endurance swimming, cycling, running and overall performance in the triathlons. However, when inferred pseudo-haplotypes were constructed from *COL3A1* rs1800255 and *COL12A1* rs970547, significant gene-gene interactions were identified. Specifically, participants with the major G+A pseudo-haplotype were significantly over-represented in the fast tertile, compared with the middle and slow tertiles, for overall finishing time, as well as for the running component of the triathlon.

Furthermore, since the *COL5A1* rs12722 variant was previously associated with endurance running,[4] additional gene-gene interactions between *COL3A1* rs1800255, *COL5A1* rs12722 and *COL12A1* rs970547 were investigated. Again, participants with the major G+T+A pseudo-haplotype were significantly over-represented in the fast tertile, compared with the slow tertile, for only the running component of the triathlon. This implicates *COL3A1* and *COL12A1*, as well as their interaction with *COL5A1*, as potential markers for endurance running performance. Additional studies should investigate these genes in true endurance running events, such as marathons, to confirm the findings of our study. Furthermore, since no single variant-independent associations were identified for *COL3A1* rs1800255 and *COL12A1* rs970547, these findings highlight the importance of gene-gene interactions when investigating multigenic complex traits such as endurance performance.

Finally, no significant associations were identified between the cycling component of the triathlon and inferred pseudo-haplotypes constructed from *COL3A1* rs1800255, *COL6A1* rs35796750 and *COL12A1* rs970547.

Study limitations

Study limitations include the lack of training data for the 2000 and 2001 Gordon's Bay events, as well as the lack of data on other important

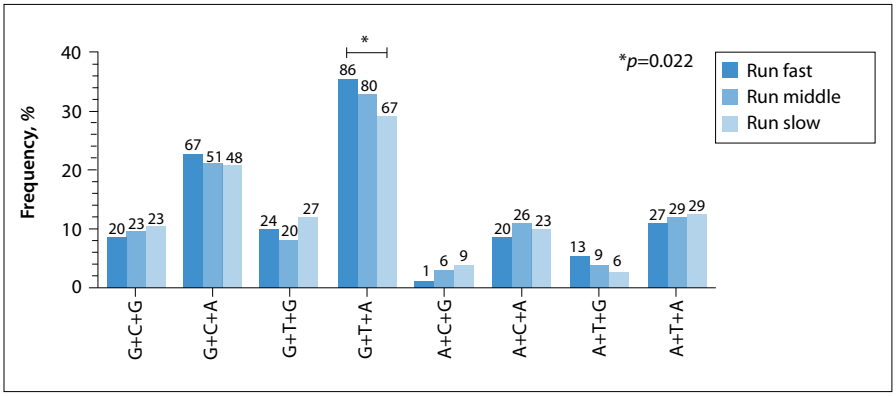

Fig. 2. Frequency distributions of inferred pseudo-haplotypes constructed from COL3A1 *rs1800255,* COL5A1 *rs12722 and* COL12A1 *rs970547 between the fast, middle and slow tertiles in the time to complete the run component of the triathlon. The number of participants is indicated above each column. (* Fast v. slow tertile.)*

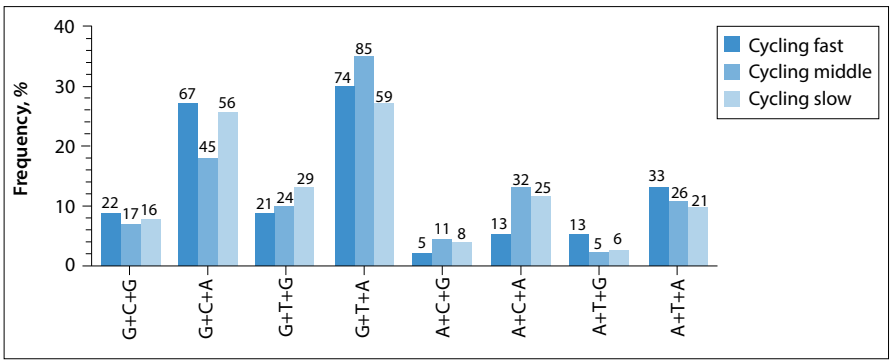

Fig. 3. Frequency distributions of inferred pseudo-haplotypes constructed from COL3A1 *rs1800255,* COL6A1 *rs35796750 and* COL12A1 *rs970547 between the fast, middle and slow tertiles in the time to complete the cycling component of the triathlon. The number of participants is indicated above each column.*

extrinsic factors influencing athletic ability, (e.g. diet). The inclusion of data regarding training and intrinsic/extrinsic factors into a multivariate analysis model in future studies, including the gene variants investigated here, would provide additional insight into possible interactions that may further explain the inter-individual differences in endurance performance.

Conclusion

Our main novel finding was that the *COL3A1* rs1800255 and *COL12A1* rs970547 variants interacted to modulate endurance running performance in the four SA Ironman triathlons investigated. Furthermore, these variants also interacted with *COL5A1* rs12722 to modulate endurance running performance. This implicates *COL3A1* and *COL12A1* as potential markers for endurance running performance.

Funding acknowledgements. This research was supported in part by the National Research Foundation (NRF), the Medical Research Council of South Africa and the University of Cape Town. MP was supported by the Thembakazi Trust.

References

1. Macarthur DG, North KN. Genes and human elite athletic performance. Hum Genet 2005;116(5):331-339. [http://dx.doi.org/10.1007/s00439-005-1261-8]
2. Bray MS, Hagberg JM, Perusse L, et al. The human gene map for performance and health-related fitness phenotypes: The 2006 - 2007 update. Med Sci Sports Exerc 2009;41(1):35-73.
3. Eynon N, Ruiz JR, Oliveira J, Duarte JA, Birk R, Lucia A. Genes and elite athletes: A roadmap for future research. J Physiol 2011;589(13):3063-3070. [http://dx.doi.org/10.1113/jphysiol.2011.207035]
4. O'Connell K, Saunders CJ, Collins M. Collagen gene sequence variants in exercise-related traits. CEJSSM 2013;1(1):3-17.
5. Birk DE. Type V collagen: Heterotypic type I/V collagen interactions in the regulation of fibril assembly. Micron 2001;32(3):223-237.
6. Minamitani T, Ikuta T, Saito Y, et al. Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen. Exp Cell Res 2004;298(1):305-315. [http://dx.doi.org/10.1016/j.yexcr.2004.04.030]
7. Olsen BR. New insights into the function of collagens from genetic analysis. Curr Opin Cell Biol 1995;7(5):720-727.
8. Laguette MJ, Abrahams Y, Prince S, Collins M. Sequence variants within the 3'-UTR of the COL5A1 gene alters mRNA stability: Implications for musculoskeletal soft tissue injuries. Matrix Biol 2011;30(5-6):338-345.
9. Collins M, Posthumus M. Type V collagen genotype and exercise-related phenotype relationships: A novel hypothesis. Exerc Sport Sci Rev 2011;39(4):191-198. [http://dx.doi.org/10.1097/JES.0b013e318224e853]
10. Fleischmajer R, Perlish JS, Burgeson RE, Shaikh-Bahai F, Timpl R. Type I and type III collagen interactions during fibrillogenesis. Ann N Y Acad Sci 1990;580:161-175.
11. Liu X, Wu H, Byrne M, Krane S, Jaenisch R. Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. Proc Natl Acad Sci 1997;94(5):1852-1856.
12. Young BB, Zhang G, Koch M, Birk DE. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. J Cell Biochem 2002;87(2):208-220. [http://dx.doi.org/10.1002/jcb.10290]
13. Chou HT, Hung JS, Chen YT, Wu JY, Tsai FJ. Association between COL3A1 collagen gene exon 31 polymorphism and risk of floppy mitral valve/mitral valve prolapse. Int J Cardiol 2004;95(2-3):299-305. [http://dx.doi.org/10.1016/j.ijcard.2003.05.026]
14. Kluivers KB, Dijkstra JR, Hendriks JC, Lince SL, Vierhout ME, van Kempen LC. COL3A1 2209G>A is a predictor of pelvic organ prolapse. Int Urogynecol J Pelvic Floor Dysfunct 2009;20(9):1113-1118. [http://dx.doi.org/10.1007/s00192-009-0913-y]
15. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-1081. [http://dx.doi.org/10.1038/nprot.2009.86]
16. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association studies (STREGA) – an extension of the STROBE statement. Eur J Clin Invest 2009;39(4):247-266.
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Med 2007;4(10):e296. [http://dx.doi.org/10.1371/journal.pmed.0040296]
18. Mokone GG, Schwellnus MP, Noakes TD, Collins M. The COL5A1 gene and Achilles tendon pathology. Scand J Med Sci Sports 2006;16(1):19-26.
19. Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP, Collins M. The association between the COL12A1 gene and anterior cruciate ligament ruptures. Br J Sports Med 2010;44(16):1160-1165. [http://dx.doi.org/10.1136/bjsm.2009.060756]
20. Tanaka T, Ikari K, Furushima K, et al. Genomewide linkage and linkage disequilibrium analyses identify COL6A1, on chromosome 21, as the locus for ossification of the posterior longitudinal ligament of the spine. Am J Hum Genet 2003;73(4):812-822.

Table S1. General characteristics for the SA Ironman triathlon participants recruited at registration of the 2000 and 2001 events in Gordon's Bay or the 2006 and 2007 Port Elizabeth events

Variable	All (*N*=659)	2000 event (*N*=96)	2001 event (*N*=294)	2006 event (*N*=219)	2007 event (*N*=50)	*p*-value
Age (years), mean±SD (*n*)	36.1±8.3 (659)	34.5±7.2 (96)	34.7±8.1 (294)	38.2±8.6 (219)	38.4±7.1 (50)	<0.001
Height (cm), mean±SD (*n*)	180.5±6.6 (559)	180.5±7.4 (85)	180.5±6.5 (267)	180.3±6.4 (158)	181.4±6.8 (49)	0.794
Weight (kg), mean±SD (*n*)	78.6±9.4 (586)	77.5±10.2 (94)	78.8±8.7 (274)	78.2±9.3 (166)	80.9±11.2 (52)	0.196
BMI (kg/m^2), mean±SD (*n*)	24.0±2.3 (555)	23.7±2.4 (85)	24.0±2.1 (264)	24.0±2.2 (157)	24.7±3.2 (49)	0.084
Country of birth, % (*n*) South Africa	65.0 (388)	53.2 (50)	63.6 (185)	73.3 (118)	68.6 (35)	0.011
Country of residence, % (*n*) South Africa	81.2 (468)	85.7 (12)	75.7 (215)	86.8 (190)	86.3 (44)	0.011

SA = South African; BMI = body mass index; SD = standard deviation.