Research Article

Differential expression of genes associated with sports injuries

Abstract

Candidate gene studies in sports injuries pose a valid new approach to investigate the genetic basis of these. In this context, our objective was to analyze the differential expression and interaction of genes associated with sports injuries using a bio informatic approach. For this study, we analyzed 31 genes associated with sports injuries previously reported in the literature. Expression analysis was performed using the Z-ratio and a protein-protein interaction network was constructed in STRING 10.0. The GO categories associated with outstanding biological processes in the network were also taken from STRING 10.0. The expression network obtained allowed establishing three clusters with a significant number of interactions, the highest interactions were found in the genes COL1A1, COL1A2, COL3A1, COL4A1, COL6A1 and COL5A1. According to the Z-ratio, the most over expressed genes were ITGB6 (z-ratio 4.32), COL1A2 (z-ratio 4.07), COL6A2 (z-ratio 3.99) and TIMM17A. In conclusion, the over-expression of genes is presented, which according to the current literature has been analyzed very little, from the sports field associated with sports injuries, a fact that merits further research. Also, we highlight the importance of bioinformatics as a complementary tool to the analysis of sports genomic data.

Keywords: genes, sport, injuries, interaction

Nancy Janneth Molano Tobar,[1] Alejandra Rocio Rodriguez Ortiz,[2] Felipe García Vallejo[2]

[1]Universidad del Valle, Teacher Universidad del Cauca, Colombia
[2]Universidad del Valle, Santiago de Cali, Colombia

Correspondence: Nancy Janneth Molano-Tobar, PhD Studdent, Universidad del Valle, Teacher Universidad del Cauca, calle 5 #4-70, Popayán, Colombia, Tel 8209800, Email najamoto@unicauca.edu.co

Introduction

Sports genomics had its beginnings since 2011,[1] is based on the study and search for genetic variants that contribute to success in sports disciplines and for this it has been based on different tools such as bioinformatics, whose applications focus on management, simulation, data mining and information analysis, with application also in the prediction of protein structures, sequence studies and other activities derived from biology research,[2] which has formed a way to identify problems, which in this case are associated with sport. Sports injuries have different etiologies and different mechanisms that intervene such as nutrition, hydration and some studies identify it under two aspects that influence such injuries: extrinsic factors, which are associated with the training process,[3] temperature, altitude, type and percentage of workload, equipment,[4] nutrition and hydration, without leaving out the stretching mechanism of the muscle fiber itself.[5]

On the other hand, there are intrinsic factors where it mediates characteristics such as metabolic changes, cell repair processes, age, aspects that until a few decades ago could not be linked and that for a country like Colombia would allow to deepen and take possession in the study of the sports genomics, a situation in which little progress has been made in Latin America.[6] Candidate gene studies provide a valid new approach to investigate the genetic basis of diseases.[7] Currently the presence of polymorphisms is being linked (SNP, del inglés single nucleotide polymorphisms) or dysregulation of genes that are influencing muscle damage; a single nucleotide polymorphism (SNP) is a single variation of the genetic code, although there are multi-allelic SNPs, SNPs are usually biallelic (two alternative bases are produced) and require a minimum frequency (> 1%) in the population.[8]

Rahim[9] raises more than 70 loci that are involved in various lesion profiles, which encode a broad spectrum of matrix proteins that include collagens and non collagens. The vast majority of these studies have followed a case-control study design of candidate genes. A small proportion of these loci have been replicated in independent studies, some of which have included different musculoskeletal injuries. Therefore, the importance of specifying which are the genes associated with sports injuries, since the dysregulation of expression of these genes can lead to disability processes and an increase in time away from the sports field, which is related to important economic losses for clubs as sponsors. In this sense, the objective of this study was to analyze the differential expression and interaction of genes associated with sports injuries using a bio informatic approach.

Material and methods

The approach is based on the tendency of genes associated with biological processes and interactions within a network, which are organized into modules or functional groups. Within these modules, new candidate genes can be identified, and gene interactions can be analyzed with a set of reference genes, which have been cited in various articles.[10–12] For which 31 genes that are related to sports injuries were used: CASP6, COL1A1, COL1A2, COL4A3, COL4A3BP, COL4A4, COL6A2, EFEMP1, EGLN1, EMILIN2, IL11, ITGA8, ITGB1, ITGB6, LTBP2, LTBP4, MACF1, MAP2K5, MAPK14, MBNL2, MEF2D, MFAP1, MFAP3, MFAP5, MORN4, PAQR3, TGFB1, TIMM17A, TIMM44, TNNT2, VCAN.

Data mining

The descriptors and general characteristics of the genes included in this study were obtained from the Geneme Browser of the University of California, Santa Cruz. (UCSC), Information related to loci, ID as a biological process, was also obtained from the NCBI, as shown in Table 1.

Gene expression

All the information was organized in dynamic tables and graphs in Microsoft ™ Excel 2013 for your analysis. The expression values were extracted from the database Gene Expression Omnibus GEO, specifically from the study conducted by Murton et al where a DNA microarray was used. The data matrix is deposited under the serial number GSE45426, platform-based Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133_Plus_2]. Said data matrix contains gene expression data, obtained from muscle biopsies of the vastus lateralis of 16 men, where 8 were controls and the other 8 were the experimental group, taking into account the ethical and legal processes for obtaining said samples, according to the authors.

Table 1 General description of genes

Gene	Full name	ID	Locus	Biological process
COL1A1	Collagen chain type I alfa 1	1277	17q21.33	Extracellular matrix organization
				Collagen biosynthetic process
COL1A2	Collagen chain o type I alfa 2	1278	7q21.3	Development of the skeletal system
				Extracellular matrix assembly
COL4A3	Collagen chain type IV alfa 3	1285	2q36.3	Cell adhesion
				Extracellular matrix organization
COL4A3BP	Collagen, type IV, alpha 3 binding protein	68018	13; 13 D1	Muscle contraction
				Mitochondria morphogenesis
COL4A4	Type IV alpha 4 collagen chain	1286	2q36.3	Extracellular matrix organization
COL6A2	Type VI alpha collagen chain 2	1292	21q22.3	Cell adhesion
				Extracellular matrix organization
EFEMP1	EGF containing fibulin extracellular matrix protein 1	2202	2p16.1	Epidermal growth factor receptor signaling pathway
				Negative regulation of chondrocyte differentiation
EGLN1	Factor 1 inducible by familial hypoxia egl-9	54583	1q42.2	Response to hypoxia
				Regulation of angiogenesis
EMILIN2	Elastin 2 microfibril interface	84034	18p11.32-p11.31	Cell adhesion
ITGA8	Integrin alpha 8 subunit	8516	10p13	Cell matrix adhesion
TIMM17A	Translocase de inner mitochondrial membrane 17A	10440	1q32.1	integral component of the inner mitochondrial membrane
TIMM44	Inner Mitochondrial Membrane Translocase 44	10469	19p13.2	mitochondrial inner membrane
TNNT2	Troponina T2, tipo cardíaco	7139	1q32.1	thin filament of striated muscle
VCAN	Versican	1462	5q14.2-q14.3	extracellular matrix containing collagen

Source:https://www.ncbi.nlm.nih.gov/gene/

Network construction

For the construction of the gene expression network, a data matrix was constructed, in which interaction data of each of the genes to be evaluated were organized; these interactions were collected from the databases, Gene MANIA y STRING 9.1. Subsequently, to the network formed with these interactions, the expression values were added. The data entered into this program were not the intensity values reported in the study from which they were extracted, they were transformed to log10 and then converted into expression levels with the z-ratio index based on the following equation.

The information about the physical interactions of greater weight between the genes evaluated, were extracted through the interface Web GeneMANIA. For the control of the false discovery rate (FDR), according to the criterion exposed by the Benjamini-Hochberg procedure, which determines the expression if the FDR was <0.05 and the fold change (FC) in the expression> 1.5

Ethical aspects

The project has the approval of the Ethics Committee of the University of Cauca through code 4925 called Prevention of sports injuries, taking into account that the databases such as software used are free to use and for free, permits were not required for your use.

Results

The expression network obtained (Figure 1), allowed to establish three clusters with a significant amount of interactions, where the highest were found in genes COL1A1, COL1A2, COL3A1, COL4A1, COL6A1 Y COL5A1, that belong to the collagen chain.
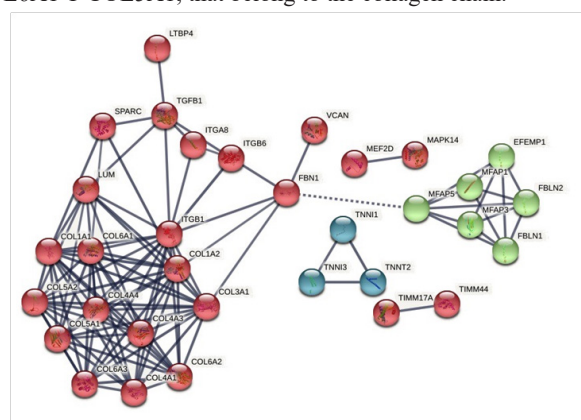
Figure 1 Protein-protein interaction network. The lines represent the confidence level of the protein-protein associations, the lowest are represented in dotted form each node represents all the proteins produced by a single locus of genes that encode proteins. Source https://string-db.org/

The other nodes show the interactions associated with the genes TIMM17A, EFEMP1, MEF2D y la last interaction with TNNT2.

When evaluating the physical interaction obtained in Gene Mania, a high interaction was established between the genes presented, whose weight was less than 0.05, as evidenced in Table 2, managing to determine that the ITGB1 gene is widely related to several genes, noting high affinity with collagen chains (COL).

Among the biological processes to which genes respond, it is shown in Table 2, where it can be seen that they interact in various multicellular processes that are related to the adhesion, architecture and development of the extracellular matrix such as skeletal muscle and growth factor.

The false discovery rate in Table 3, allowed to establish that the genes were found below 0.05, determining activation of genes in biological mechanisms associated with muscle injuries, which involve the integrity of the extracellular matrix, adhesion, healing as organization Of the same. Based on what was found through the

Table 2 Weight of the most relevant physical interactions. Source https://genemania.org/

Gene 1	Gene 2	Weight
ITGB1	COL1A1	0.007366
ITGB1	VCAN	0.009766
ITGB1	COL4A3	0.010506
ITGB1	VCAN	0.014887
COL1A2	COL1A1	0.017644
ITGB1	EMILIN1	0.019752
GFI1B	EFEMP1	0.031929
UBE4B	CASP6	0.037095
MEF2A	MEF2D	0.037797
MEF2A	MEF2D	0.04639

z-ratio, it was possible to highlight the high expression of the genes that code for ITGB6 (z-ratio 4.32), such as COL1A2 (z-ratio 4.07), COL6A2 (z-ratio 3 .99), TIMM17A (z-ratio 3.57), as shown in Table 4.

Table 3 Outstanding biological processes of the network. Source: https://string-db.org/

GO_ID	Term description	False discovery rate
GO:0070208	Protein heterotrimerization	0.00011
GO:0051291	Protein heterooligomerization	0.001
GO:0071560	Cellular response to transforming growth factor beta stimulus	0.0012
GO:0097435	Supramolecular fiber organization	0.0022
GO:0007167	Enzyme linked receptor protein signaling pathway	0.0023
GO:0032836	Glomerular basement membrane development	0.0033
GO:0038063	Collagen-activated tyrosine kinase receptor signaling pathway	0.0033
GO:0043589	Skin morphogenesis	0.0033
GO:0016043	Cellular component organization	0.0037
GO:0007155	Cell adhesion	0.0039
GO:0007179	Transforming growth factor beta receptor signaling pathway	0.0039

Table 4 Z-Ratio of over-expressed genes

Gen	ID_REF	Z-ratio
ITGB6	226535_at	4,32
COL1A2	202404_s_at	4,07
COL6A2	209156_s_at	3,99
TIMM17A	201821_s_at	3,57
MEF2D	225641_at	3,43
ITGB1	216178_x_at	3,20
LTBP2	204682_at	3,10
MBNL2	205018_s_at	3,09
TIMM44	203093_s_at	3,00
COL1A1	217430_x_at	2,99
TGFB1	203085_s_at	2,61
COL4A4	229779_at	2,56
LTBP4	227989_at	2,45
EMILIN2	224374_s_at	2,42

Discussion

The evidence raised in this work allowed us to notice the presence of genes associated with muscle injuries, which is related to their participation in the processes that involve the structural integrity of a complex or its assembly inside or outside a cell as the processes that develop with the extracellular matrix.[13]

The analysis of the expression network allowed to identify 3 nodes, where their expression stands out, finding that ITGB6 acts as an adhesion factor from the signaling of the extracellular matrix, wound healing and fibrosis,[14] since its expression in the keratinocytes at the edge of the wound allows re-epithelialization[15] estudios[16] state that this gene is not detected until after mucular injury, an aspect that requires further review.

It is unquestionable that the extracellular matrix is influenced by different genes, one of which presents the most interaction are those associated with the collagen chain, a fact that is corroborated by Thankam,[17] by reporting that changes in the levels of collagen subtypes promote tendencies to collagen diseases that lead to inflammation and disorganization of the extracellular matrix, which violates the stability of the structure and generates a tendency for repetitive injuries.[18]

The COL1A2 gene is a fibril-forming collagen found in most connective tissues and is abundant in bones, corneas, dermis, and tendons.[19] Several studies demonstrate the genotypic association of the COL1A1 polymorphism with soft tissue lesions, since its alteration would lead to an alteration in the support processes such as resistance to traction between the muscle and the bone.[20]

In the same way, it was possible to show that the COL6A2 gene is a myofibrillar collagen enriched in the pericellular matrix and plays an important role in the repair of tissues such as the tendon and in the cell migration necessary for the recovery of wounds,[21] its association with the muscle occurs with diseases such as muscular dystrophy where there is an excessive accumulation of collagen and other components of the extracellular matrix,[22] fact that denotes the disability in terms of the motor action that the movement entails.

Through the gene expression matrix, it was established that the MEF2D gene presented a high z-ration, taking into account its participation in the control of differentiation and the development of muscle and neuronal cells, it can be considered as a gene important for the marking of the myog gene, which is identified as a master regulator of skeletal myogenesis,[23] It is thus then that the myo gene requires MEF2 for the recruitment of the transcriptionally repressive PcG complex in the muscle-specific promoter at specific stages of development,[24] in the same way Lambert et al.,[25] show that the decrease in this gene affects the elongation as a spiral shape of the muscle fiber, important for the development of physical capacities such as strength and flexibility.[26]

The ITGB1 gene is related to the external integrity of the muscle cell, facilitating cell adhesion through the cell-substrate junction that anchors the cell to the extracellular matrix and forms a termination point for actin filaments. how the intracellular domain of the integrins binds to the cytoskeleton through adapter proteins such as talin, α-actinin, vinculin, focal adhesion kinase, and paxilin, necessary for skeletal muscle development, which is reported by Pang et al.,[27] by indicating that the down regulation of the integrin seriously affects the differentiation of osteoblasts, during the development of skeletal muscle, a fact that impairs the fusion of myoblasts, which is essential for the development and regeneration of skeletal muscles.

t became evident that one of the candidates associated with muscle injuries is the VCAN gene is a member of the family of aggrecan / versican proteoglycans, which is a protein that encodes chondroitin sulfate proteoglycan and is a main component of the extracellular matrix, facilitating functions such as cell adhesion, proliferation, proliferation, migration and angiogenesis and plays a central role in the morphogenesis and maintenance of tissues,[28] fact that it is shown that an alteration of the VCAN interaction leads to an increase in lesions of the musculoskeletal system, since previous studies link this gene with the stimulation of hypoxic macrophages that upregulate a series of hypoxia-inducible transcription factors, which contains increased expression at sites of tissue injury.[28]

The presence of genes such as LTBP2, which presents an important z-ration, allows estimating its link to muscle injuries, since it has a high degree of homology with fibrillins and helps in the formation of elastin microfibers., placing it as a key gene associated with connective tissue from the properties of recovery and resistance, on the other hand its function in cell adhesion reveals its link with the properties associated with tissue repair,[29] a situation that highlights the need to expand research on these genes. The GO categories of biological processes extracted from the network showed the domain of processes such as cell organization processes, response to growth factor, tissue morphogenesis, cell adhesion, response to wounds and development of muscle tissue, which are related to the muscular system. skeletal and injuries, which indicates that an alteration of genes related to these processes can trigger inadequate functions that would correlate with the manifestation of injuries of a muscular nature or chronic pathologies that would limit the activity of athletes.

Conclusion

The present work shows that sports injuries of a muscular nature are associated with various genes, and their identification promotes new challenges when prescribing exercise in order to avoid them as its consequences, in the same way this work presents over-expression of genes that in the current literature have been unknown or worked very little from the sports field associated with sports injuries, a fact that deserves more research, enhancing sports genomics as a field to explore. Additionally, it was established that bioinformatics is a complementary tool to the analysis of sports genomic data.

Acknowledgments

Conflicts of interest

The authors declare there are no conflicts of interest.

References

1. Szalata M, Słomski R, Balkó Š, et al. Advances in athlete genomics in 2019. *Trends in Sport Sciences.* 2019;26(2):55–61.

2. Coltell Ó, Arregui M, Fabregat A, et al. Bioinformatics of biological and clinical data. Rev Med Chile. 2008;136:645–652.

3. Bohm S, Mersmann F, Arampatzis A. Functional adaptation of connective tissue by training. *Deutsche Zeitschrift Fur Sportmedizin.* 2019;70(4):105–109.

4. Pruna R, Artells R. How the genetic component can affect the injuries of athletes. *Apunts Medicina de l'Esport.* 2015;50(186):73–78.

5. Tyler TF, Schmitt BM, Nicholas SJ, et al. Rehabilitation after hamstring-strain injury emphasizing eccentric strengthening at long muscle lengths: Results of long-term follow-up. *Journal of Sport Rehabilitation.* 2017;26(2):131–140.

6. De Las Rivas J, Bonavides Martínez C, Campos Laborie J. Bioinformatics in Latin America and SoIBio impact, a tale of spin-off and expansion around genomes and protein structures. *Briefings in Bioinformatics.* 2019;20(2):390–397.

7. Baird AEG, Carter SD, Innes JF, et al. Genetic basis of cranial cruciate ligament rupture (CCLR) in dogs. *Connective Tissue Research.* 2014;55(4):275–281.

8. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single- nucleotide polymorphisms in the human genome. *Science.* 1998;280(5366):1077–1082.

9. Rahim M, Collins M, September A. Genes and Musculoskeletal Soft-Tissue Injuries. *Medicine and Sport Science.* 2016;61:68–91.

10. Murton AJ, Billeter R, Stephens FB, et al. Transient transcriptional events in human skeletal muscle at the outset of concentric resistance exercise training. *Journal of Applied Physiology.* 2014;116(1):113–125.

11. Kitzler TM, Schneider R, Kohl S, et al. COL4A1 mutations as a potential novel cause of autosomal dominant CAKUT in humans. *Human Genetics.* 2019;138(10):1105–1115.

12. Larruskain JON, Celorrio D, Barrio I, et al. Genetic Variants and Hamstring Injury in Soccer: An Association and Validation Study. *Medicine & Science in Sports & Exercise.* 2018;50(2):361–368.

13. Ferranti Ramos A, Garza Garza G, Bátiz Armenta J, et al. Metalloproteinases of the extracellular matrix and their participation in the healing process. *Revista Médicas UIS.* 2017;30(3):55–62.

14. Meecham A, Marshall J. The ITGB6 gene: its role in experimental and clinical biology. Gene: X, 2020;5:1–6.

15. Jacobsen JN, Steffensen B, HÄkkinen L, et al Skin wound healing in diabetic β6 integrin-deficient mice. *Apmis.* 2010;118(10):753–764.

16. Ducceschi M, Clifton LG, Stimpson SA, et al. Post-transcriptional regulation of ITGB6 protein levels in damaged skeletal muscle. *Journal of Molecular Histology*. 2014;45(3):329–336.

17. Thankam FG, Evan DK, Agrawal K, et al. Collagen type III content of the long head of the biceps tendon as an indicator of glenohumeral arthritis. *Molecular and Cellular Biochemistry.* 2019;454(1–2):25–31.

18. Smith C, Kruger MJ, Smith RM, et al. The inflammatory response to skeletal muscle injury: Illuminating complexities. *Sports Medicine.* 2008;38(11):947–969.

19. Wang C, Li H, Chen K, et al. Association of polymorphisms rs1800012 in COL1A1 with sports-related tendon and ligament injuries: A meta-analysis. *Oncotarget.* 2017;8(16):27627–27634.

20. Goodlin GT, Roos TR, Roos AK, et al. The dawning age of genetic testing for sports injuries. *Clinical Journal of Sport Medicine.* 2015;25(1):1–5.

21. Sardone F, Santi S, Tagliavini F, et al. Collagen VI–NG2 axis in human tendon fibroblasts under conditions mimicking injury response. *Matrix Biology*. 2016;55:90–105.

22. de Greef JC, Hamlyn R, Jensen BS, et al. Collagen VI deficiency reduces muscle pathology, but does not improve muscle function, in the γ-sarcoglycan-null mouse. *Human Molecular Genetics*. 2016;25(7):1357–1369.

23. Tapscott SJ. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development.* 2005;132(12):2685–2695.

24. Aziz A, Liu QC, Dilworth FJ. Regulating a master regulator: Establishing tissue-specific gene expression in skeletal muscle. *Epigenetics.* 2010;5(8):691–695.

25. Lambert M, Bastide B, Cieniewski Bernard C. Involvement of O-GlcNAcylation in the skeletal muscle physiology and physiopathology: Focus on muscle metabolism. *Frontiers in Endocrinology.* 2018;9:1–12.

26. Davegardh C, Broholm C, Perfilyev A, et al. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. *BMC Medicine*. 2017;15(1):1–27.

27. Pang Y, Zhang Z, Wang Z, et al. Platelet endothelial aggregation receptor-1 regulates bovine muscle satellite cell migration and differentiation via integrin beta-1 and focal adhesion kinase. *Cell Adhesion and Migration.* 2019;13(1):192–202.

28. Kenagy RD, Kikuchi S, Evanko SP, et al. Versican is differentially regulated in the adventitial and medial layers of human vein grafts. *PLoS ONE*. 2018;13(9):1–24.

29. Wei T, Kulkarni NH, Zeng QQ, et al. Analysis of early changes in the articular cartilage transcriptisome in the rat meniscal tear model of osteoarthritis: pathway comparisons with the rat anterior cruciate transection model and with human osteoarthritic cartilage. *Osteoarthritis and Cartilage.* 2010;18(7):992–1000.