Injury, Int. J. Care Injured 47S3 (2016) S41–S50

Contents lists available at ScienceDirect

Injury

journal homepage: www. elsevier.com/locate/Injury

Whole-exome sequencing analysis in twin sibling males with an anterior cruciate ligament rupture

Enrique Caso[a,*], Antonio Maestro[b], Cristina C. Sabiers[a], Manuel Godino[c], Zaira Caracuel[a], Joana Pons[a], F. Jesus Gonzalez[a], Rocio Bautista[d], M. Gonzalo Claros[d], Jaime Caso-Onzain[a], Elena Viejo-Allende[a], Peter V. Giannoudis[e], Sara Alvarez[f], Paolo Maietta[f], Enrique Guerado[c]

[a]Research & Development Unit, Hospital Universitario Costa del Sol, University of Malaga, Marbella, Malaga, Spain
[b]Department of Orthopaedic Surgery, FREMAP, Gijon, Asturias, Spain
[c]Department of Orthopaedic Surgery and Traumatology, Hospital Universitario Costa del Sol, University of Malaga, Marbella, Malaga, Spain
[d]Supercomputing and Bioinnovating Centre (SCBI), University of Malaga, Malaga, Spain
[e]Academic Department of Trauma & Orthopaedic Surgery, University of Leeds, NIHR Leeds Biomedical Research Unit, Chapel Allerton Hospital, Leeds, UK
[f]NIMgenetics, Madrid, Spain

KEYWORDS

Anterior cruciate ligament injury
Soft tissue injuries
Polymorphism
Whole-exome sequencing
Single nucleotide
bioinformatics analysis
INDEL mutation
Genome
Computational biology

ABSTRACT

Familial predisposition is among the major genetic risk factors for non-contact musculoskeletal tissue injuries. Personal genome sequence shows that different polymorphism profiles may account for the number and the degree of injuries and the recovery time. Genotyping studies allow investigation into genome factors with potential impact on pathogenesis of non-contact ligament injuries.

We have studied a family with twin sibling males surgically diagnosed of an anterior cruciate ligament non-contact rupture and non-affected progenitors (father and mother) were subjected to whole exome sequencing (WES) analysis. WES analysis previously carried out on 16 individuals, without ACL injury medical records, were also included in this study for single nucleotide variants (SNVs) and small insertions and deletions detection (indels), variant filtering and to prioritize variants relative to the disease. WES analysis to identify SNVs and indels was performed using open web-based bioinformatics tools.

A set of 11 new variants shared by family members can be associated to ACL non-contact injury, including SerpinA11, ARSI, NOCHT4, EPB41, FDFT1, POMC, KIF26A, OLFML2B, ATG7, FAH and WDR6. All of them, except ATG7 and WDR6, have shown a damaging predictive pattern by combinatorial standard predictive scores. In combination to the identified SNVs of EPB41 and SerpinA11 genes, ACTL7A gene showed a predicted deleterious variant reinforcing the idea these variants impact on of fibroblast-like cells deformability and ECM misbalance, Differential gene expression and RNA sequencing analysis will help to understand the combined participation of these protein coding genes in ACL non-contact injuries.

Introduction

Anterior cruciate ligament injury

The musculoskeletal system is essential for mammal activities of motion, as movement includes not only walking or running but also upper limb activities. The heterogeneous and multiphasic structure of musculoskeletal soft tissues (articular cartilage, ligaments, knee meniscus and intervertebral disk) provides elasticity and capability to guide motion, share loads, and transmit forces in a manner that is unique to each one.

The knee cruciate ligaments (CLs) provide the joint with a combination of robust stability and a wide range of motion for precise locomotor activities. The CLs derived from vascular femoral and tibial mesenchyme are composed of numerous immature fibroblasts having, at the early stage of development, a small and fusiform cytoplasm and nuclei.

Despite their intraarticular location the CLs remain extra-synovial as they are surrounded by a mesentery-like fold of synovium originated at the posterior capsular apparatus of the knee joint. Apparently

* Corresponding author: Research & Development Unit, Hospital, Universitario Costa del Sol, Autovia A-7. Km 187, 29603, Marbella, Malaga, Spain. Tel.: +34 951 97 66 69, +34 951 97 66 70; fax: +34 951 97 62 22.
E-mail address: encape@hcs.es (E. Caso).

0020-1383 / © 2016 Elsevier Ltd. All rights reserved.

the early development of the knee joint is guided by the anterior cruciate ligament (ACL) as it may interact with the resulting shape of the femoral condyles and the tibial plateau during development. The complex ultra-structural organization, the varied orientation of the bundles in the ACL, and the abundant composition of elastic fibres, make it very different from other ligaments and tendons. After birth, the ACL must be able to withstand multiaxial stresses and varying tensile strains, resisting anterior tibial translation and rotational loads. ACL is mainly composed of extracellular matrix (ECM) and relatively few fibroblasts responsible for the synthesis of the surrounded material. A substance composed of protein fiber-rich and proteoglycan-rich ECM and interstitial fluid gives the ACL tissue direction dependent stiffness and strength for mechanical response [1].

ACL disorders include those induced by prior trauma, [2–4] which can lead to partial or complete ACL tears with subsequent tissue deterioration, which include mucoid degeneration by filling the ligament with mucoid or myxoid material. Although ACL ruptures resulting from traumatic injuries are associated with premature knee osteoarthritis (OA), there is a paucity of data regarding the prevalence of degenerative ACL abnormalities among those patients with symptomatic OA and whether such abnormalities are associated with the severity of cartilage degeneration [5].

Family predisposition of ACL rupture

Familial predisposition is among the major genetic risk factors for non-contact musculoskeletal tissue injuries [6]. Personal genome sequence shows that different polymorphism profiles may account for the number and the degree of injuries and the recovery time [7]. For instance, a copy number variation (CNV) deletion of CEP57L1 was detected in a mother and her daughter, both affected by ACL/PCL agnesia and from 3D protein structure analysis it was concluded that functional binding sites related to microtubule attachment were missing [8]. These findings place the genotyping studies at first line of research, to look for genome factors with predictive impact on pathogenesis of non-contact ligament injuries.

Up-to date, results from phenotype clinical studies are not supporting a clear familial predisposition to ACL tears. From a recent retrospective case-control phenotyping study [9] on 120 patients undergoing surgical reconstruction of a torn ACL, when stratified by gender, a significantly higher familial (first degree-relative) prevalence was shown in males in comparison to controls, and a higher prevalence of ACL rupture in first degree-relative of female patients with meniscal injuries was also demonstrated. From a study on 593 elite alpine skiers, an odds ratio of 1.95 was found to suffer an ACL injury if there is a parent with ACL injury compared those having a parent without it [10]. Because these data do not match with previous clinical findings of higher incidence in the immediate-family members, previous to the torn anthropometric differences have been claimed. In twin sibling female players, an increased general joint laxity, decreased hamstrings to quadriceps (H/Q) ratios and femoral intercondylar notch width were considered to be risk factors to familial predisposition to ACL injury increased knee abduction angles, and decreased knee flexion angles [11]. Moreover, for every 1.3-mm increase in side-to-side differences in anterior-posterior knee displacement, the odds of ACL–injured status increased 4-fold (95% confidence interval, 1.68–9.69) and a positive measurement of knee hyperextension increased the odds to 5-fold (95% confidence interval, 1.24–18.44) [12].

The genomic era and the ACL rupture

After the start of the genomic era, whereas Systems Medicine (SM) is taking place, personalized medicine (PM) emerges as a medical model using molecular profiling technologies for tailoring the precise therapeutic strategy for each person at the right time, determining, also, the predisposition to disease at population level, and delivering a timely and stratified prevention.

Establishing a molecular diagnosis and determining genetic causality for disease in clinical practice is challenging. Molecular profiling flows from the patient's bedside to the scientist's bench, thus closing the loop between scientific research and clinical medicine. Molecular profiling interrogates clinical samples for a variety of molecules, including DNA, RNA, proteins, and metabolites. Next Generation sequencing technologies have increased gene discovery, understanding the causes of single-gene disorders among patients with suspected, but previously undiagnosed, genetic disorders.

Whole-exome sequencing (WES) using next generation sequencing technologies brings up the possibility to determine the exomic variation profile, just in a single assay. An individual exome contains 180,000 exons (>85% of protein coding sequences). It constitutes approximately 1% of the whole genome and includes about 20,000 genes. It is believed that the exome contains about 85% of heritable disease-causing mutations. In comparison to traditional Sanger sequencing, the diagnostic rate for WES has been shown to be over 25% on mendelian. In fact, non-targeted WES allows the identification of new candidate disease gene variants that may not be possible to detect by other means.

Patients and methods

First-degree relatives

A family with twin sibling males surgically diagnosed of an anterior cruciate ligament (ACL) non-contact rupture and non-affected progenitors (father and mother) were subjected to WES analysis. WES analysis previously carried out on 16 individuals, without ACL injury medical records, were also included in this study for single nucleotide variants (SNV) and small insertions and deletions (indels) detection, variant filtering and to prioritize variants relative to the disease. Approval from Local Research Ethics Committee Board was obtained and written informed consent for genetic testing and biobank management and storage were obtained from all participants subjected to WES analysis.

Exome sequencing was performed using 100 ng of genomic DNA isolated (MagnaPure, Roche) from whole blood previously stored at −80°C at hospital biobank. Libraries were prepared using the Ion AmpliSeq™ Exome Kit (Life Technologies) and quantified by qPCR. The enriched libraries were prepared using Ion Chef™ and sequenced on PI™ Chip in the Ion Proton™ System (Life Technologies) to provide >95% of amplicons covered with at least 20X, with an average base coverage depth that ranged 85 to 130.

Bioinformatics

WES analysis to identify SNVs and indels was performed using open web-based bioinformatics tools. Gene variants can be located at exones, introns, predictive splice sites, UTRs, Functional variants can be identify as frameshift, synonymous, misssense, nonsense and stoploss. Variants located in exons are more likely to be pathogenic than those located in introns or between genes and these are the bases for bioinformatics applied to WES analysis. The power of this strategy has increased with the access to large numbers of publicly available sequences that allow the controlled comparison of frequencies, as well as the identification of de novo variants and stratification by ethnicity.

Signal processing, base calling, alignment and variant calling were performed on a Proton™ Torrent Server using the Torrent Suite™ Software. Variants were annotated using Ion Reporter™ Software. The bioinformatics workflow includes the elimination of adaptors sequences and sample identification index, the alignment of read sequences to reference genome, the variant identification