Università Campus Bio-Medico di Roma

Corso di dottorato di ricerca in
Scienze Biomediche Integrate e Bioetica
XXXIV ciclo a.a. 2018-2019

TOWARDS DEFINITION OF THE GENETIC BASIS OF ROTATOR CUFF TEARS: GENOME WIDE ASSOCIATION STUDIES VERSUS EXOME SEQUENCING STUDIES

Dott. Andrea Guarnieri

Coordinatore
Prof. Raffaele Antonelli Incalzi

Tutor
Prof. Umile Giuseppe Longo

15 Giugno 2022

SUMMARY

ABSTRACT

Background: Degenerative rotator cuff tear (RCT) is a musculoskeletal disease characterized by shoulder pain and functional impairment without history of trauma. A vast literature describing RCT suggests that a genetic predisposition likely underlies this condition. Many studies have identified RCT susceptibility loci, but due to genetic and phenotypic heterogeneity observed across individuals, very few loci were subsequently replicated. Research on genetic predisposition to atraumatic RCT has relied on classical linkage studies or genome-wide association studies (GWAS) to identify candidate genes with common susceptibility variants. Although the observation of such common variants is informative, they can only explain a small fraction of the predicted RCT heritability, suggesting a considerable contribution would come from rare and highly penetrant variants.

Purpose: The purpose of this study was to test the hypothesis that rare genetic variants cause or predispose to RCT. The long-term objective is to understand the molecular basis of RCTs to delineate targeted therapies and preventive measures to reduce both direct and indirect costs for the healthcare system.

Methods: We searched for families composed by patient affected by bilateral atraumatic full thickness RCT documented at magnetic resonance imaging, and with at least a 1st degree relative affected by bilateral RCT too. Families had to respect inclusion and exclusion criteria. Also their unaffected 1st degree relatives were studied. All patient suffering RCT underwent arthroscopic repair procedure in Orthopaedic Department of our Institution. One 10-ml blood sample was drawn from all included subjects to perform a whole exome sequencing. Exome sequencing was performed by an external laboratory DanteLab SRL (Pizzoli, AQ),

while genetic analysis was conducted in collaboration with Genetics Department of our Institution.

Results: A family of four individuals was enrolled. Two subjects (affected son and his mother) had bilateral atraumatic RCT. Two subjects (unaffected son and his father) denied shoulder pathology. Patients with atraumatic RCT underwent reparative surgery. Notably, the male patient underwent surgery in 2021, the mother before the start of our study. The healthy brother participated in the exome sequencing study, while the father will be studied in the future with a family segregation analysis by the Sanger method. In our study we found four rare gene variants associated with atraumatic RCT: COL23A1: p.P458S, JARID2: p.Q454P, MDC1: p.G207Pfs*3, HDAC10: p.P426L.

Conclusion: This is the first study to evaluate the genetic predisposition of RCT through whole exome sequency study on a family. It is a first step in an ongoing investigation of genetic basis and to delineate preventive measures and targeted therapy for RCTs. To ensure the validity of RCT susceptibility genes identified through this study, it will be imperative that these findings be validate in a larger external cohort

INTRODUCTION

DNA SEQUENCING

DNA sequencing is the procedure of sequential reading of nucleotides in a given sample of interest. Over the past fifty years, the evolution of sequencing technologies, together with significant advances in their automation and the development of computer programs for analyzing large amounts of sequence data, have allowed the transition from sequencing short stretches of sequence to millions of bases, to rapid sequencing of entire genomes.

In the mid-1970s, two different methods of DNA sequencing were developed: the so-called chemical method, devised by Allan Maxam and Walter Gilbert (1), and the enzymatic method, devised by Frederick Sanger (2). The chemical method is based on marking with radioactive phosphorus the DNA to be sequenced, which is then denatured and divided into four aliquots, each of which is treated with chemical reagents at low concentrations so as to cause the DNA to break at specific bases. This results in a series of labeled fragments that are run on polyacrylamide gels to be separated according to their different sizes, defined by the breakpoint in the DNA. Through an autoradiographic film it is then possible to determine the order of the nucleotides and, therefore, identify the DNA sequence in the starting sample. On the contrary, in Sanger's method, which will be described in detail later because it is still in use today, the fragments are synthesized by a DNA polymerase that uses as a mold the DNA whose sequence is to be determined. Sanger's enzymatic method was initially used less than the chemical method, mainly because of the limitation of having to already know the nucleotide sequence of the region to which to affix the trigger necessary for the sequencing reaction. Over the years, however, Sanger's method has been refined, made more accurate and, above all, proved to be more suitable for an automation of the procedure than the chemical method. In fact, starting from since the 1990s a succession of technological innovations has made it possible to make sequencing almost completely automatic and extremely rapid, lowering the cost of the procedure by several orders of magnitude. Sanger's method was the method used to

sequence the entire human genome. In February 2001, the first complete draft of the human genome sequence was published by the internationally funded Human Genome Project (HGP) (3) and by the private firm Celera (4). The two concurrent, parallel, and competing projects used two different strategies, a so-called hierarchical approach and Celera's so-called "shotgun" method, respectively. In the hierarchical approach, total DNA was broken into relatively large fragments that were first cloned into Bacterial Artificial Chromosome (BAC) vectors and mapped to human chromosomes by Fluorescent In Situ Hybridization (FISH). Sequences cloned into BACs were subsequently randomly fragmented into smaller fragments, then inserted into smaller vectors and sequenced. The obtained sequences (reads) were then analyzed by specific programs that allowed the overlapping of the fragments (overlapping) and finally the assembly of the sequence of the original BAC. In the shotgun method, on the other hand, the whole part based on the cloning of the larger fragments into BAC vectors was bypassed, starting directly from the obtaining of the smaller fragments, whose sequences were overlapped to obtain contigs (stretches of sequence assembled without discontinuity consisting of many overlapping reads) and finally the representative sequence, or consensus sequence. The fundamental contribution of the two projects to the history of DNA sequencing is certainly represented by the publication of the entire sequence of 3.1 billion base pairs of the human genome. This enormous achievement, in fact, was the starting point for the development of subsequent Next Generation Sequencing (NGS) platforms, aimed at sequencing entire genomes on a large scale, in a single laboratory, in a matter of days.

THE SANGER METHOD

The Sanger method, designed by Frederick Sanger in 1977 also known as deoxy-terminators method, is based on the enzymatic synthesis of a new DNA chain from a template strand whose sequence is to be determined. A mixture of deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) of the four DNA bases is used. The ddNTPs differ from common dNTPs in the absence of the -OH group at the 3' carbon in deoxyribose (Figure 1), which is required for the formation of the phosphodiester bond with the next nucleotide.

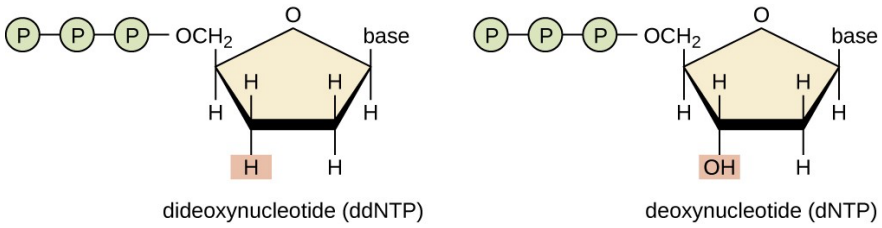

Figure. 1 Comparison of the structure of a dideoxynucleotide and a deoxynucleotide. Note how the deoxynucleotide (left) lacks the -OH group at the 3' carbon in deoxyribose, which is required for elongation in synthesis (www.genomeup.com).

Therefore, random incorporation of a ddNTP causes the polymerase reaction catalyzed by the DNA polymerase enzyme to stop, producing truncated fragments. The concentration ratio of ddNTPs to dNTPs is such that the DNA polymerase enzyme will statistically terminate the nascent chain at all possible positions where a ddNTP can be inserted. Thus, the enzyme produces a series of fragments that all share the same initial end at 5' but are of different lengths, interrupted at the incorporation of the ddNTP at 3'. The DNA polymerase needs a trigger to initiate synthesis, which in the sequencing reaction is represented by an oligonucleotide complementary to the sequence and adjacent to the region to be sequenced. In the original method (2), the sequencing reaction took place separately for each nucleotide in four different tubes, as it was possible to identify the truncated fragments with the same radioactive marking as the phosphate molecules. The DNA sequence was then reconstructed by running the amplification products on polyacrylamide gels, in four different lanes, and visualizing the bands by autoradiography. In the 1990s, the introduction of automated sequencers, in which ddNTPs began to be labeled with different fluorescent molecules for each of the four nitrogen bases, allowed the synthesis reaction to be carried out in a single reaction mixture and the fragments separated in a single electrophoretic lane (5). The newly obtained DNA chains are separated according to their length by capillary electrophoresis in the automated sequencer, and when the fluorescent ddNTP-labeled fragments reach the detector position, they are detected and identified by emission of the fluorescence specific to each nucleotide. A digital camera reports the signal intensity in a graph called electropherogram, characterized by the succession of peaks corresponding to the different fluorochromes.

NEXT GENERATION SEQUENCING PLATFORMS

In recent years, the methodologies and technological platforms for DNA sequencing have undergone an extraordinary evolution that is revolutionizing the possibilities and application horizons in many sectors. The introduction on the market, starting from 2005, of the next generation sequencing platforms has produced an exceptional increase in sequencing capacity and, in parallel, a drastic reduction in costs. Among the main innovations introduced in NGS platforms, cloning of DNA fragments to be sequenced is not required and, at the same time, the use of microreactors, normally immobilized on a solid support, allows a high level of parallelization. Furthermore, sequence determination does not require the step of electrophoretic separation because as nucleotides are incorporated into the sequencing reaction, they are simultaneously identified. DNA is fragmented by chemical or enzymatic approaches into fragments of appropriate size, in depending on the technology used, obtaining libraries of fragments to the ends of which adaptor oligonucleotides are covalently bound, which are then used both for clonal amplification of the fragments and for the actual sequencing reaction. Currently, the most popular sequencing platforms that involve a clonal amplification step of the fragments are Illumina's Genome Analyzer and Life Technologies' Ion Torrent/Proton. All the technologies that will be described are still subject to a progressive development that will gradually improve their performance, obtaining a greater length and accuracy of sequence reading, a higher number of sequences produced and progressively lower costs.

Illumina Genome Analyzer

The first step of sequencing with the Illumina system, a platform introduced to the market in 2006, is the preparation of a fragment library, obtained by fragmenting the DNA of interest. The DNA fragments of the library, with a length of 500-700bp, are then suitably modified by adding a pair of adaptor oligonucleotides at both ends. Thanks to the presence of the adapters, the DNA fragments can be fixed on the surface of a flow cell, composed of eight independent "lanes", on which oligonucleotides complementary to the adapters of the fragments have been previously bound. In this way it is possible to perform a so-called bridge amplification process (bridge-PCR), in

which the DNA polymerase synthesizes a complementary strand to the template to be sequenced. The molecule obtained is then denatured in such a way that the two strands remain bound to the slide at one end only. The operation is repeated clonally: the two fragments hybridize their free end with a complementary oligonucleotide from the flow cell, reassuming again the bridge shape and allowing the synthesis of a new complementary strand. The immobilization, synthesis, and denaturation steps continue until clonal clusters of the fragments of interest are obtained, each consisting of approximately one million copies and representative of the initial fragment (Figure 2).

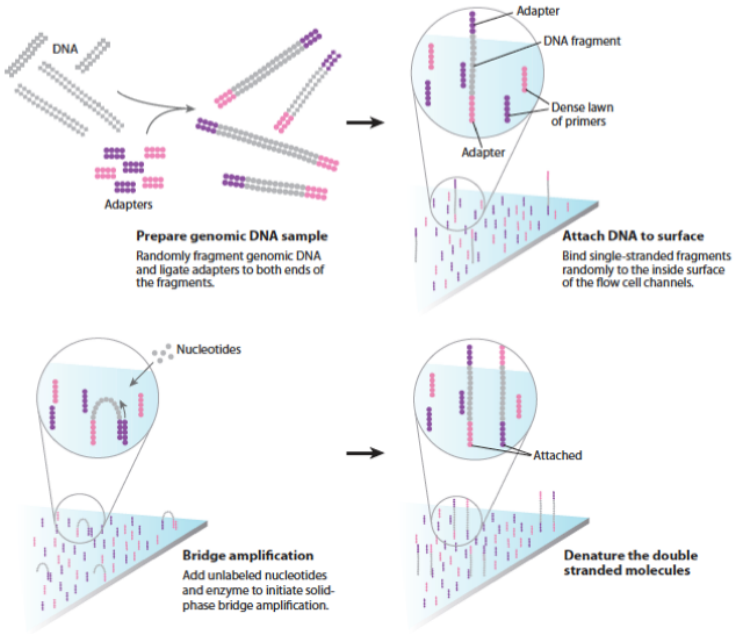

Figure 2 - Bridge-PCR. Preparation of the fragments library of which it is desired to determine the sequence and ligation of the adaptor oligonucleotides, thanks to which the bridge amplification on the slide of a flow cell is possible; denaturation of the neosynthesized molecules in such a way as to ensure the continuation of clonal amplification (Mardis et al. (6)).

The real sequencing step is known as Sequencing by Synthesis (SBS) and involves the use of a DNA polymerase. After removing the strands complementary to the fragments whose sequence is to be determined, the primers for sequencing are paired with the clonal clusters and the four dNTPs are added one at a time, progressively detecting their fluorescent emission. Each dNTP, in fact, is labeled with a different fluorochrome and carries a blocking group at 3'-OH that prevents the incorporation of additional nucleotides. Once the fluorescence is detected, the 3'-OH blocker is removed from the

incorporated dNTP and the process continues with the addition of other dNTPs. The cycle is then repeated a predetermined number of times, each time reading the fluorescence (Figure 3).

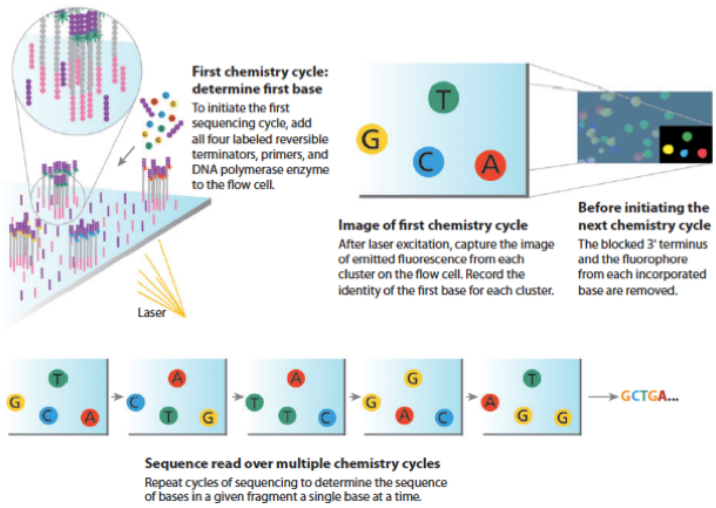

Figura 3 – Sequencing reaction using the Illumina platform. The four differentially labeled deoxynucleotides with a 3'-OH blocking group, the primers for sequencing to begin, and the DNA polymerase are added simultaneously to clonal clusters of single-stranded molecules. As each nucleotide is added, its fluorescence is detected and recorded, the blocking group is removed and the next nucleotides are added, whose fluorescence will be progressively detected with the same procedure. (Mardis et al. (6))

The advantage of using reversible inhibitors consists in the ability to read one base at a time and it doesn't present problems related to the sequence determination of homopolymeric traits, for example. Among the peculiarities of the Illumina platform is the opportunity to obtain the sequences of both ends of the fragments (paired-end sequencing), thus enabling a much more accurate and specific subsequent bioinformatic analysis of the data. The most advanced Illumina instrument is capable of producing 600 Gigabp in a single experiment, with paired-end reads that can be as long as 250 bp for each end of the fragment.

Ion Torrent/Proton from Life Technologies.

The technology used by the Ion Torrent/Proton system, a platform introduced to the market in 2010, exploits the principle that for each nucleotide incorporated into the sequencing reaction, a proton (H+) is released from the 3'-OH group involved in

phosphodiester bond formation. DNA fragments to be sequenced are first immobilized on agarose microbeads, by taking advantage of the complementarity between fragment adapters and oligonucleotides present on the surface of the beads, and then amplified clonally by emulsion PCR. Clonal amplification generates approximately one million molecules on the surface of each microbead. After the PCR reaction, the emulsion is destroyed and the marbles with the amplification products and DNA polymerase are inserted into the microwells of a chip capable of housing millions of microbeads. During the sequencing reaction, dNTPs are added simultaneously in successive cycles and an electrical system attached to the chip, known as an Ion-Sensitive Field-Effect Transistor (ISFET), is able to detect pH changes that are proportional to the number of nucleotides sequentially incorporated into the nascent DNA chain. The electrical pulses transmitted from the chip to the computer connected to it are then translated into a DNA sequence, with no intermediate signal conversion step required (Figure 4).

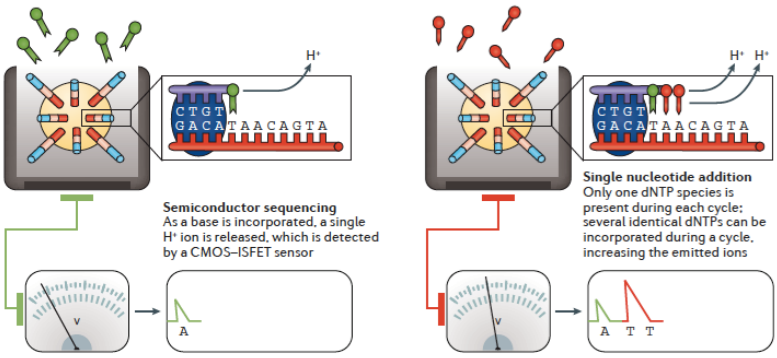

Figure 4 - Sequencing reaction using the Ion Torrent/Proton platform. At the time of nucleotide incorporation into the nascent chain, an H+ ion is released resulting in a pH change, detected by an ISFET device sensitive to changes in ion concentration in the solution. After insertion of the nucleotide that is complementary to the mold, excess unincorporated nucleotides are removed and those for the next cycle are added (Goodwin et al (7)).

An advantage of this technology compared to the Illumina platform, with whom it shares the sequencing process by synthesis, lies in its greater simplicity since an optical apparatus for signal detection, normally more complex and expensive, is not required. In addition, the Ion Torrent/Proton technology does not require the use of dNTPs modified with fluorochromes, since the signal detection is based on an electrical system that is able to perceive the concentration, and therefore the variation, of H+ ions in solution. A limitation is instead represented by the determination of the

sequence of homopolymeric stretches. In fact, if in the filament of which you want to determine the sequence is present a repetition of a given nucleotide, more nucleotides will be incorporated into a single cycle and will be released proportionally more H+ ions resulting in a higher signal, then more susceptible to error. This platform is capable of generating up to 60-80 million reads of an average length of 200 nucleotides in only about two hours.

Identification of polymorphic sites and genetic variants

The error rates associated with next-generation technologies are higher than sequencing with the traditional Sanger method. For this reason, this method is still used today as an aid to NGS in confirming identified variants. However, the accuracy of the sequencing of next-generation platforms is ensured by a repeated and massive reading of each gene fragment that determines the degree of coverage, given by the sum of reads that are aligned to a reference sequence. The high depth of sequencing of NGS platforms allows, therefore, a genome-wide study of genetic variability, allowing the identification of SNPs (Single Nucleotide Polymorphisms) and genetic variants even in mosaic with low percentage.

These analyses can be conducted on the entire genome (WGS, Whole Genome Sequencing), or focused on its coding regions. In the latter case, before of sequencing, the library of DNA fragments undergoes a process of enrichment by hybridization through probes drawn on known exonic regions of the genome (8). This technique, known as Whole Exome Sequencing (WES), has been and is still widely applied for the identification of causative variants of many diseases. The advantages over whole genome sequencing include the reduced costs and less time required for data generation. The resulting substantial reduction in data complexity means that the data can be analyzed in a more computationally straightforward manner than those obtained from whole genome sequencing.

Bioinformatics pipeline for NGS data analysis

The term pipeline refers to a chain of bioinformatics operations required to obtain information from next-generation platform sequencing. The amount of data produced

by NGS technologies is in the range of Gibabases, representing a difficulty for data analysis and storage. In particular, for example, the assembly and reconstruction of a genome, an exome, or the sequence of target genes of interest, represent a very complex critical point to address from a computational point of view. In addition to the large amount of data to be analyzed, in fact, some of the main obstacles are represented by the quality, not always optimal, of the sequences obtained during sequencing and by the fact that not all of them are uniquely localized in a single point of the genome. The analysis softwares differ according to the NGS technology used, but they all follow a data analysis system that has as first task to convert the signals obtained by the sequencer in nucleotide sequences or reads, which represent the raw data of the sequence. These sequences are collected in files in FASTQ text format (Figure 5), containing in the first line the individual sequence identifier (ID) and an optional description preceded by the "@" character, in the second line the nucleotide sequence, in the third line a description preceded by the character "+" and in the fourth line an estimate of the quality for each nucleotide of the sequence included in the second line (quality score), known as Phred or Q score and coded as a decimal conversion of the ASCII (American Standard Code for Information Interchange) code (Figure 6).

Figure 5 - FASTQ file. The first line reports the individual sequence identifier, along with an optional description, preceded by the "@" character; the second line reports the nucleotide sequence; the third line may include a description preceded by the "+" character; the fourth line reports quality estimates for each nucleotide in the sequence in the second line (Q scores) (www.drive5.com).

Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII	Q	P_error	ASCII
0	1.00000	33 !	11	0.07943	44 ,	22	0.00631	55 7	33	0.00050	66 B
1	0.79433	34 "	12	0.06310	45 -	23	0.00501	56 8	34	0.00040	67 C
2	0.63096	35 #	13	0.05012	46 .	24	0.00398	57 9	35	0.00032	68 D
3	0.50119	36 $	14	0.03981	47 /	25	0.00316	58 :	36	0.00025	69 E
4	0.39811	37 %	15	0.03162	48 0	26	0.00251	59 ;	37	0.00020	70 F
5	0.31623	38 &	16	0.02512	49 1	27	0.00200	60 <	38	0.00016	71 G
6	0.25119	39 '	17	0.01995	50 2	28	0.00158	61 =	39	0.00013	72 H
7	0.19953	40 (	18	0.01585	51 3	29	0.00126	62 >	40	0.00010	73 I
8	0.15849	41)	19	0.01259	52 4	30	0.00100	63 ?	41	0.00008	74 J
9	0.12589	42 *	20	0.01000	53 5	31	0.00079	64 @	42	0.00006	75 K
10	0.10000	43 +	21	0.00794	54 6	32	0.00063	65 A			

***Figure 6 - Q scores, probability of error associated with a base and ASCII code**. The quality score of a base, also known as Phred or Q score, is an integral number representing the probability of error associated with that base. If P is the probability of error, then: P=10-Q/10 and therefore Q=-10 log10(P). Q scores are often associated with ASCII codes, as shown in the table (www.drive5.com).*

Quality scores, which indicate the probability of error associated with each nucleotide, represent an important tool to eliminate, from the analysis process, bases or reads that do not exceed the minimum quality parameters (trimming), thus improving the accuracy of the next step, which is the alignment/mapping of reads to a reference genomic sequence. This step consists of a process of determining the most likely source position of a read within the reference genome. A good alignment software should be able to: handle the large amount of reads produced by sequencing, then align them quickly and without requiring huge computational resources; handle short DNA sequences (35-300bp), then accurately identify their genomic position; show a good ratio between sensitivity and speed of search, then quickly identify the correct genomic position; handle large reference genomes, then quickly search the positions of reads in genomes several Gigabases long, again without requiring huge computational resources. There are various software that can be used in this phase, different for the algorithm on which they are based, for the speed of data processing and mapping reads to the reference genome, then for the accuracy with which they are aligned, and also different for the length of reads that can be mapped by each program. The result of this step is a file in BAM (Binary Alignment Map) format, binary equivalent of the SAM file (Sequence Alignment Map), which provides information about the "coverage" of the sequenced region and is compressed and more suitable to be used by analysis programs that operate the subsequent steps of the pipeline. An optional step, which may follow the alignment step, is the "refinement" of the BAM file by removing duplicate reads to maintain the information contained in the independent reads and to

eliminate multiple copies of the same read. Duplicate reads, in fact, can compromise the subsequent variant calling step and, in addition, affect the actual sequencing depth (9). After mapping reads to their specific address on a reference genome and eventually removing duplicates, we proceed with variant calling, a step aimed at identifying differences between the sequence under examination and the reference genomic sequence. Among the software used in this step, there are programs based on probabilistic models (e.g., Bayesian model) that, for each genomic position, calculate the probability of observing every possible genotype given the alleles present in the reads. Other programs, however, use haplotype-based methods. These software programs divide the genome into small regions, in each of which the most frequent haplotypes are identified and variants are then called based on the observed combinations. The variant calling phase results in a VFC (Variant Call Format) tabular file. As a last step we proceed with the annotation of the variants, a procedure that adds useful information such as the position of the variant (exonic, intronic, 3'/5'-UTR etc.), its functional role if exonic (synonymous, frameshift etc.), the amino acid substitution that determines, if the variant is exonic and if it falls in regions of interest such as genomic regions conserved from the evolutionary point of view. The annotation step can also provide information about the allele frequencies reported by different databases (1000 Genomes, dbSNP, gnomAD etc.) and is also useful to evaluate the scores reported by SIFT and PolyPhen (predicts the possible impact of amino acid substitution on the stability and function of the protein) for non-synonymous variants. From the annotation of the variants we obtain a file in tabular CSV (Comma-Separated Values) format, the first file in the pipeline that can be effectively analyzed from a common PC using the functionality of Microsoft Excel. Starting from this file it is possible to proceed with the filtering and prioritization of the variants, following the parameters of interest chosen by the operator on the basis of the condition under examination, subsequently exploiting information available from various databases on the web.

Reference database and tools used

- gnomAD (Genome Aggregation Database) (www.gnomad.broadinstitute.org) is a database developed with the intent of bringing together sequencing data, both exomes

and genomes, from a great number of large-scale sequencing projects, and making the data collection available to the scientific community (10). The amount of data in the GRCh37/hg19 dataset spans 125,748 exonic sequences and 15,708 genomic sequences from unrelated individuals sequenced as part of population genetics studies on specific diseases. Once the name of the gene of interest is entered, gnomAD, in addition to giving an idea of the coverage of the sequenced regions, provides a huge catalog of variants associated with that gene, the frequencies of which can be appreciated in different populations. Also shown are the structures of the transcripts corresponding to the gene of interest, each with its own "ENST" abbreviation from the Ensembl database. The database also supplies the user with other information, including the ratio between the number of loss-of-function variants observed in a given gene and the number of expected variants (oe value), thereby providing a measure of how well a gene is tolerant to such variants.

- VarSome (www.varsome.com) is a powerful search engine that provides the user informations about human genomic variants (Kopanos et al. 2019). If the searched variant falls within a gene, the browser shows its exonic structure, any multiple transcripts, regions of interest (e.g. functional protein domains, binding sites etc.) retrieved from UniProt and nearby structural variants reported by ClinVar. The browser is also able to show any other known variants in the same genomic region. The pathogenicity of the variant of interest is reported based on an automated variant classifier, which evaluates the searched variant according to the American College of Medical Genetics and Genomics (ACMG) guidelines, classifying it as "pathogenic", "likely pathogenic", "likely benign", "benign" or "Variant of Uncertain Significance" (11). Variant frequency data in the population are retrieved from gnomAD and other databases such as Kaviar3 and the ICGC (International Cancer Genome Consortium), as well as pathogenicity predictions from dbNSFP, a database that collects prediction scores from twenty different algorithms (FATHMM-MKL, MutationTaster, SIFT, LRT etc.), and from DANN. Relevant clinical information (e.g., variant-associated conditions, mode of transmission etc.) is retrieved from the CGD (Clinical Genomic Database).

- ClinVar (www.ncbi.nlm.nih.gov/clinvar) is a public repository of human genetic variants, maintained by the NCBI (National Center for Biotechnology Information), containing relationships between these variants and related phenotypes (12). The database contains more than 331,000 variants, including structural variants (more than 15,000 variants > 1 kb), for which it provides a description, associated phenotypes, interpretations of clinical significance, which is reported according to ACMG guidelines, and evidence for those interpretations.

- **VarElect** (www.varelect.genecards.org) is a tool that works as a prioritizer of genetic variants in a phenotype-dependent manner (13). Based on the GeneCards database, which contains information on all putative and annotated human genes, this software works by matching genes to keywords entered into the tool. VarElect is an efficient software to prioritize lists of genes containing variants, obtained as a result of NGS experiments, by simplifying data analysis based on keywords of interest chosen by the user. Using this software it is possible to enter as input hundreds or thousands of genes and obtain as output lists of few genes, prioritized and annotated, whose variants are potentially associated with particular diseases, phenotypes or biological functions, defined by the keywords.

- **FATHMM (Functional Analysis through Hidden Markov Models)** (www.fathmm.biocompute.org.uk) is an in silico tool that can predict the effect of missense variants on the protein product, based on the level of sequence conservation (14). In addition to the functions of FATHMM, the FATHMM-MKL version is also able to predict functional consequences on non-coding regions by integrating ENCODE (ENCyclopedia Of DNA Elements) annotations.

- **MutationTaster** (www.mutationtaster.org) is an in silico prediction tool, based on a Bayesian model, that evaluates variants in a DNA sequence for their potential pathogenicity (15). It takes into account not only evolutionary conservation and protein product, but also splicing sites, messenger RNA, and any domains with regulatory function.

- SIFT (Sorts Intolerant From Tolerant) (www.sift.bii.a-star.edu.sg) is an in silico prediction tool for non-synonymous variants that can predict whether an amino acid

substitution affects protein function (16). It is based on the degree of sequence conservation and physicochemical properties of amino acids.

- **LRT (Likelihood Ratio Test)** (www.genetics.wustl.edu/jflab/lrt_query.html) predicts whether a variant is "deleterious" or "tolerated" through the identification of highly conserved amino acid regions, using a genomic dataset of 32 vertebrate species (17).

ATRAUMATIC ROTATOR CUFF TEARS

Atraumatic rotator cuff tears (RCTs) consist of progressive atrophy, fatty infiltration, and surface fibrillation of one or more tendons. These types of injuries mostly occur in adult patients 35 to 65 years of age without previous shoulder trauma. All traumatic or degenerative RCTs represent a risk factor for other shoulder pathologies. RCTs have been shown to predispose to proximal humeral head migration and osteoarthritis (OA) over time (18). Patients complain of chronic shoulder pain, more typically during the night, and functional impairment. Diagnosis is based on a thorough history, physical examination with function testing, and imaging. The history is critical to rule out a history of prior trauma and correctly define the degenerative nature of the tear. Physical examination consists of active and passive mobilization to assess shoulder range of motion (ROM) and various RCT tests. Plain radiographs can be used for evaluation of degenerative changes in the shoulder joint, but magnetic resonance imaging (MRI) is mandated for evaluation of tendon status and potential fatty degeneration.

Atraumatic RCTs also differ from traumatic injuries from an anatomopathological perspective. Studies of ruptured rotator cuff tendons have revealed that degenerative changes also appear medially to the tear and have been seen even in patients who had not been diagnosed with RCTs. Therefore, it appears that degeneration is present before the tear occurs (19-21). Even if traumatic or atraumatic lesion can present the same macroscopic morphological characteristics, their management couldn't be the same due to their different histological patterns. Since that RC integrity is important for normal glenohumeral kinematics and thus joint stability, its compromission in form and function can affect the shoulder joint's long-term health. RC tendons functionality is heavily dependent on its extracellular matrix properties. Collagen fibers, primarily

type I, define the tissue's tensile strength. Over the last two decades, there has been much research that has helped clinicians better understand the natural history of degenerative rotator cuff disease. These studies, often relying on abnormalities in the asymptomatic shoulder, suggest that atraumatic RCT histological changes could be part of the normal aging process in many individuals (22). In a systematic review investigating the prevalence of rotator cuff disease with increasing age, Teunis et al. showed that abnormalities ranged from 9.7% in patients aged 20 years and younger and increased to 62% in patients of 80 years and older (23). This increase persisted regardless of the presence of symptoms.

Patients with unilateral shoulder pain are also at greater age-related risk of possessing a RCT in their contralateral asymptomatic shoulder. In a comparative study of patients with unilateral shoulder pain, the average age for a patient without a RCT was 48.7 years (24). The age increased to 58.7 years for a unilateral tear, and 67.8 years for those with a bilateral tear. After the age of 66 years individuals have 50% likelihood of bilateral tear. Notably, the study correlated the age with the presence and type of RCT, but not with the size of the tear. Kim et al., analyzed 360 shoulders and they found that small full-thickness tears most commonly involved an area 15 mm posterior to the biceps tendon (25). This suggested that degenerative cuff tears may initiate in a region 13 to 17 mm posterior to the biceps tendon. From this position, tears subsequently spread as the disease proceeds. This study highlighted common tear location to be within the rotator crescent and often sparing the anterior cable insertion of the supraspinatus. The authors showed that only 30% of full-thickness RCTs involved the most anterior aspect of the supraspinatus tendon footprint.

Tear size and tear location have been shown to correlate closely with fatty degeneration of rotator cuff muscles. In a study investigating these factors in 413 shoulders, fatty degeneration was seen on ultrasound imaging almost exclusively in full-thickness tears (26). In shoulders with fatty degeneration, tear length and width were significantly higher. In addition, as the distance measured from posterior to the biceps tendon and to the anterior edge of the RCT decreased, the number of fatty infiltrated muscles increased. Regression analysis demonstrated that tear size was the most important predictor of infraspinatus degeneration, whereas disruption of the

anterior supraspinatus tendon was the most important predictor of supraspinatus muscle degeneration.

Proximal migration is commonly seen in rotator cuff deficient shoulders via alteration of normal glenohumeral kinematics. It has been shown that the size of the tear correlates strongly with humeral migration and is the strongest predictor of migration in symptomatic shoulders (18). A tear area threshold of 175 mm^2 was identified as a critical size, which demonstrated a positive correlation to migration. Smaller tears did not demonstrate significant correlation with migration regardless of the presence or absence of pain. Greater involvement of the infraspinatus leads to significantly greater alteration of glenohumeral kinematics.

On the basis of these and other studies, the natural history of atraumatic rotator cuff disease follows a predictable clinical path. As patients age, asymptomatic full-thickness RCTs increase in incidence. These tears initiate at the junction of the posterior supraspinatus and anterior infraspinatus within the rotator crescent, propagating from that point. As the size of the tear increases, the amount of fatty degenerative changes and the number of muscles involved increase. When the tear size progresses beyond a critical threshold, proximal humeral migration is observed. More recent research from the last 3–5 years has focused on continuing to describe and understand the natural history of atraumatic rotator cuff disease, including symptom progression, tear enlargement, and the development of arthritis. Additionally, studies have investigated, with long-term healing data, whether the natural history of atraumatic RCTs can be altered with surgical intervention (27).

Pathology

Tendons have a water content of 70%, while type I collagen accounts for 85% of their dry weight. Cells are the minor part and those present are mostly (90–95%) fibroblasts. The histopathological changes associated with RCTs include thinning and disorganization of collagen fibers, the presence of granulation tissue, infiltration of glycosaminoglycans, fibrocartilaginous metaplasia, calcification, fatty degeneration, and necrosis of the tendon margin with cell apoptosis (28, 29). These changes are also present in macroscopically intact tendons in the early stage of the degenerative process

(28). The degenerative changes at the tendon margin can also explain the high rate of recurrence after surgery, even in cases with less than grade 2 fatty infiltration (21). Necrotic tendon tissue, fibrin and fibronectin appear to be signs of tendon degeneration, while fibrosis and thinning of fascicles have also been found in controls without RCTs (30). Studies in pathologic tendons have demonstrated an increase in the expression of MMPs and a decrease in TIMP mRNA expression in tenocytes (31-33).

Castagna et al. (19), in a study of 13 patients, found that levels of MMPs 1, 2, and 3 were altered not only at the edge of the torn supraspinatus tendon, but also in areas far from the lesion and in the arthroscopically and histologically healthy subscapularis tendon. In tendinopathy with chronic RCT, MMPs are increased and TIMPs are decreased (34, 35), while in small RCTs and in young people both MMPs and TIMPs are increased. In another study conducted by Castagna et al. (36) tendon samples were harvested from 13 patients who underwent arthroscopic repair of an isolated full thickness tear of the supraspinatus tendon (< 2 cm). Three specimens were harvested en bloc from each patient: from the lateral edge of the supraspinatus tear (L), from the arthroscopically intact middle portion of the tendon, more than 1 cm lateral to the edge of the tear (M), and from the macroscopically intact superior margin of the subscapularis (S) tendon. This latter specimen was used as control. At histochemical analysis the detected chondroid metaplasia (Fig. 7A) of the medial area of the supraspinatus tendon (M) in 10 of 13 patients, thinning and disorientation of collagen fibers, with mucoid and lipoid degeneration, in 9 patients, single micro calcification (Fig. 7B) in 11 of 13 patients. The subscapularis specimens (S), used as controls, showed normal appearance in all patients, with no lesions, fatty infiltrations, necrosis, scar lesions, and micro calcifications.

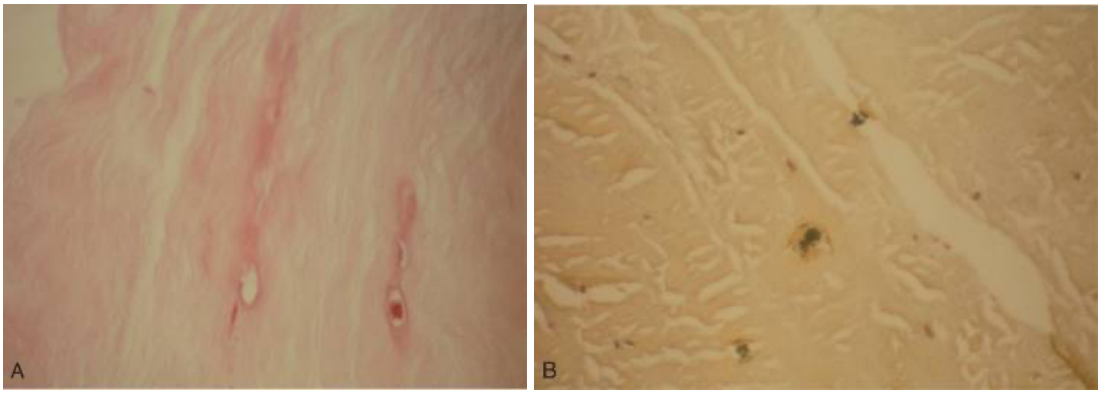

Figure 7 - From A. Castagna et al. Matrix metalloproteases and their inhibitors are altered in torn rotator cuff tendons, but also in the macroscopically and histologically intact portion of those tendons. Muscles Ligaments Tendons J. 2013 Jul-Sep; 3(3): 132–138 (36)

Epidemiology

Rotator cuff disease is a widespread musculoskeletal pathology, a major cause of shoulder pain and is among the most common musculoskeletal disorders (37). It is the third cause of musculoskeletal disease (16%), after the back (23%) and the knee (19%) (38). Its relevance is correlated not only to its high prevalence rate, but also to the fact that it is disabling, causing high direct and indirect costs in industrialized countries (39). The incidence of RCTs increases in the elderly population (40-42), being approximately 6 and 30% in those under and over 60 years, respectively (42). Partial thickness tears are more frequent than full thickness, with a prevalence of 13 versus 7% (43). Symptomatic disease is generally localized in the dominant arm, and in 28% of patients in the non-dominant shoulder only. Bilateral disease is present in 36% of patients (24). Not all RCTs are symptomatic, thus influencing the reported incidence of the disease. The incidence of RCTs (RCT) increases with aging, affecting approximately 30% to 50% of individuals older than 50 years, and more than 50% of individuals older than 80 years. The impact of this condition on earnings, missed workdays, and disability payments is relevant. Rotator cuff disease is a common health concern among working populations (44), causing high healthcare costs in Western industrialized countries and relevant costs for workers' compensation systems (39). It represents the costliest problem in Workers' Compensation Systems after low-back pain. Epidemiologic data concerning the association between mechanical overuse and RCTs are scanty. The prevalence rate of rotator cuff syndrome in workers range

between 29% in those highly exposed to repetitive work, and 16% in those weakly exposed to repetitive work (45). However, other studies show that 70% of patients with full-thickness tears undertake light work only (46). Epidemiologic studies have helped to identify factors potentially associated with rotator cuff disease.

Classification and MRI evaluation

RC disease includes a spectrum of pathology ranging from tendinopathy to partial or complete tears (47). RCTs can be classified by etiology (traumatic or atraumatic), location (bursal surface, articular surface), size and extension (partial thickness, full thickness, massive). Acute (or traumatic) lesions are described as sudden onset of pain and functional impairment after a traumatic event, while chronic (degenerative or atraumatic) lesions as a variable level of weakness and pain without a traumatic event for more than 3 months. MRI can provide information about RCTs such as tear dimensions, tear depth or thickness, tendon retraction, and tear shape that can influence treatment selection and help to determine the prognosis. In addition, tear extension to adjacent structures, muscle atrophy, size of muscle cross-sectional area, and fatty degeneration have implications for the physiologic and mechanical status of the rotator cuff. DeOrio and Cofield (48) classified full thickness RCTs according to size as either small (<1 cm), medium (1–3 cm), large (3–5 cm), or massive (>5 cm or involving 2 or more tendons). Partial thickness tears are divided in articular surface and bursal surface lesions. Both patterns are graded according to their depth as either grade 1 (<3 mm), grade 2 (3– 6 mm), or grade 3 (>6 mm). The normal rotator cuff is 10 –12 mm thick; thus, grade 3 tears are considered significant tears involving more than 50% of the cuff thickness (49, 50). Tears can be classified arthroscopically into three basic shapes according to the tear geometry as viewed from the tendon surface: crescentic, U shaped, and L shaped (51, 52). In crescentic tears, the tendon pulls away from the greater tuberosity but typically does not retract far medially and therefore can be reattached to bone with minimal tension. U-shaped tears are massive RCTs that may extend medially to the level of the glenoid fossa. L-shaped tears are massive tears with a longitudinal component along the orientation of the rotator cuff fibers and a transverse component along the cuff insertion. Also, the degree of tendon retraction is an important information obtained with MR imaging. Optimally, in primary repair, the

tendon stump should be adjacent to the attachment site so that reattachment is free of tension. When MR imaging shows retraction of the tendon edge medial to the glenoid fossa it has been suggested that a tear is suspected to be irreparable. Tear extension is important for the evaluation of joint stability in the axial plane, that is the result of a balance of forces between the anterior and posterior cuff muscles ("transverse force couple" resulting from integrity of infraspinatus and subscapularis tendons) (53-55). If supraspinatus tendon tear extension involves the subscapularis and infraspinatus tendons there is a strong association with arthritic degeneration effects on glenohumeral joint stability and function (53, 56). The importance of the transverse force couple concept in shoulder bio-mechanic may explain why some patients with isolated full-thickness tears of the supraspinatus tendon have acceptable shoulder function and are candidate for a conservative approach in selected cases (54, 56). MRI is used also to assess muscle atrophy and fatty degeneration. Both can reduce muscle strength and can result in worse functional outcome after surgical repair. Supraspinatus muscle atrophy is evaluated measuring muscle cross-sectional area, the scapular ratio and the tangent sign. The muscle cross-sectional area may provide important information, since this measurement correlates with muscle strength (57). The cross-sectional area of the supraspinatus muscle can be measured with MRI. The scapular ratio is calculated in the sagittal oblique plane at the level of the medial coracoid process, where the supraspinatus fossa is largely encompassed by osseous boundaries. If the ratio of the cross-sectional area of the supraspinatus muscle to the area of the supraspinatus fossa (occupation ratio) is less than 50% in the sagittal oblique plane, supraspinatus muscle atrophy is indicated. In a study by Thomazeau et al (58), supraspinatus muscle atrophy as determined with this method correlated with extent of tendon tear and was associated with tear recurrence after surgical repair. The tangent sign consists in the use of an MRI sagittal plane and bone landmarks similar to those used in the scapular ratio method, and a normal supraspinatus muscle should cross superior to a line drawn through the superior borders of the scapular spine and the superior margin of the coracoid process (Fig. 8a) This finding is not present with atrophy (Fig. 8b) (59). There is a significant correlation between occupation ratio, tangent sign, and improved strength and mobility (60).

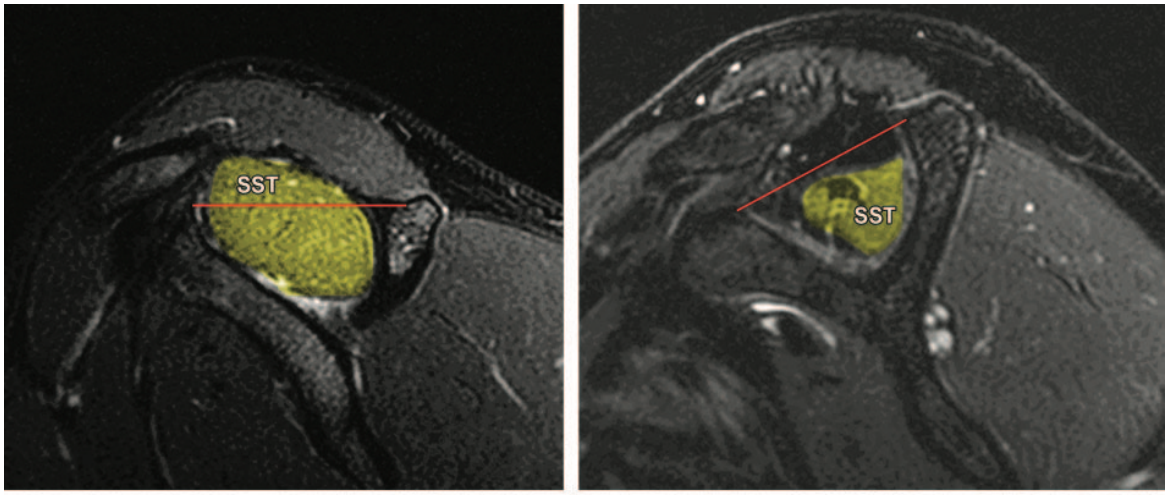

Figure 8 - Tangent sign. **(a)** Sagittal fat-saturated T2-weighted MR image shows the belly (green) of a normal supraspinatus muscle *(SST)* crossing a tangent (red line) drawn between the superior borders of the scapular spine and the superior margin of the coracoid process. **(b)** Sagittal fat-saturated T2-weighted MR image (3000/60) shows atrophy of the belly (green) of the supraspinatus muscle *(SST)*, which now lies entirely below the tangent (red line). *From Yoav Morag et al. MR imaging of rotator cuff injury: what the clinician needs to know. Radiographics Jul-Aug 2006;26(4): 1045-65. doi: 10.1148/rg. 264 055087 (61).*

Fatty degeneration usually occurs around tendon fibers and blood vessels (Fig. 9), and the true pathologic process may, in fact, be fatty infiltration rather than fatty degeneration (62, 63).

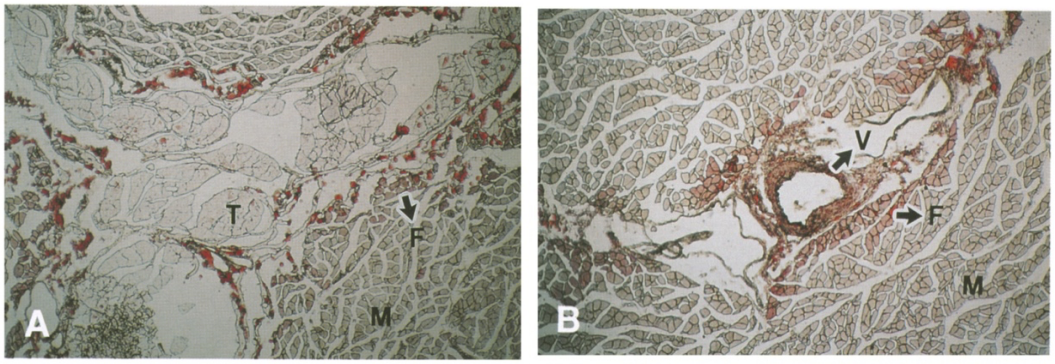

Fig. 9 Section of supraspinatus muscle associated with cuff tear. Sudan three stain; F, fatty degeneration tissue; T, tendon fiber; V, vessel; M, muscle fiber. A, Fatty degeneration of muscle near tendon fiber (original magnification x20). B, Fatty degeneration of muscle near vessels (original magnification x20). *From Nakagaki K, Ozaki J, Tomita Y, Tamai S. Fatty degeneration in the supraspinatus muscle after rotator cuff tear. J Shoulder Elbow Surg 1996; 5(3):194 –200 (62).*

There is an association between worsening fatty degeneration and alteration in muscle function. Fatty degeneration is an important predictive factor for surgical outcome. Both conventional MR imaging techniques and more advanced techniques such as MR spectroscopy have been used to evaluate fatty degeneration of rotator cuff muscles (Fig. 10).

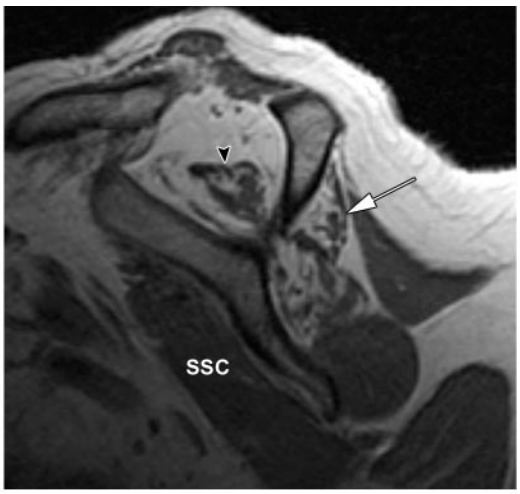

Fig 10: Sagittal T1-weighted MR image shows severe fatty degeneration and volume loss in the supraspinatus muscle belly (arrowhead) and infraspinatus muscle belly (arrow). The volume of the subscapularis *(SSC)* muscle belly appears to be unchanged. *From Yoav Morag et al. MR imaging of rotator cuff injury: what the clinician needs to know. Radiographics Jul-Aug 2006;26(4):1045-65. doi: 10.1148/rg.26405508 (61).*

Goutallier et al (64) described the fatty infiltration grades. They defined grade 1 (completely normal muscle, without any fatty streak), grade 2 (muscle contains some fatty streaks, fatty infiltration is important, but there is more muscle than fat), grade 3 (equal amounts of fat and muscle) and grade 4 (more fat than muscle is present). Multiple studies have suggested that muscle infiltration beyond grade 2, according to the Goutallier classification, represents an inflection point at which healing becomes less consistent (65, 66).

Etiology

The cause of atraumatic RCTs has only been studied by a limited number of investigators and remains unknown. The pathophysiology of RCT is multifactorial (67-71) composed by extrinsic factors (sports activity or occupation, smoking), intrinsic factors (age or gender) and biological factors as defects of tendons (69, 72, 73), including increased tendon cell death, a higher proportion of fat composition,

aberrant microstructure of structural fibers, and abnormal nutrient vessels (72, 74). This suggests that atraumatic RCTs are not purely due to repetitive micro-trauma or overuse. In a recent case-control study conducted on 2738 patients Song et al (75) identified advancing age, male sex, higher BMI, and diagnosis of carpal tunnel syndrome as risk factors significantly associated with an increased risk of degenerative RCTs. It is possible that the biological changes are regulated by genes. Identifying genes associated with rotator cuff disease can help early recognition of individuals at higher risk of developing this pathology. This could warrant application of primary or secondary prevention strategies for this specific population.

MANAGEMENT OF ROTATOR CUFF TEARS

Full-thickness tears do not heal spontaneously and the majority (36–50%) progress in size gradually. Arthroscopic surgical repair is the gold standard treatment for RC tears but the percentage of retears range from 11% and 94%. In the United States in 2006 there were an estimated 272,148 ambulatory surgeries performed for RCTs (76). In 2017 Longo U.G. et al performed a retrospective 14-year period study on the Italian population. They described 390,001 RC repairs performed in Italy, 65% of which in patients <65 years (working population), with an incidence of 62.1 RC procedures for every 100,000 Italian inhabitants over 25 years old (67) (Fig. 11). The number of RC repair procedure is gradually increasing over the years, affecting mostly the male gender in younger population, and females in older population (Fig. 12).

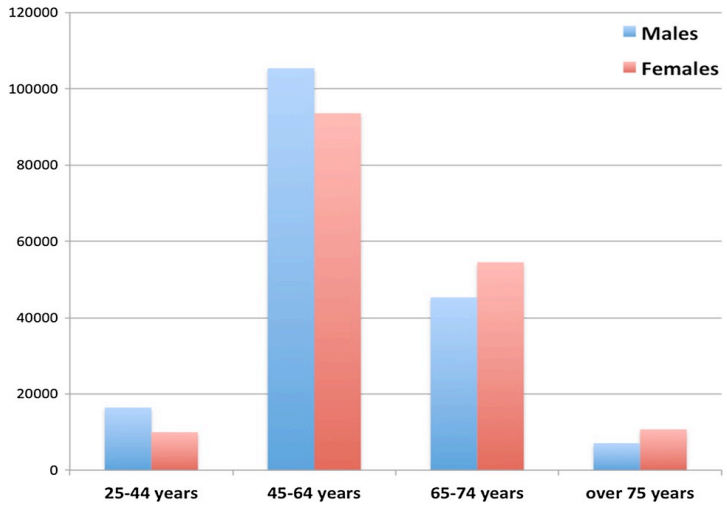

Figure 11 - Distribution of total RC repair procedures by

From Longo et al. The burden of rotator cuff surgery in Italy: a nationwide registry study. Arch Orthop Trauma

2610-x. Epub 2016 Dec 8.

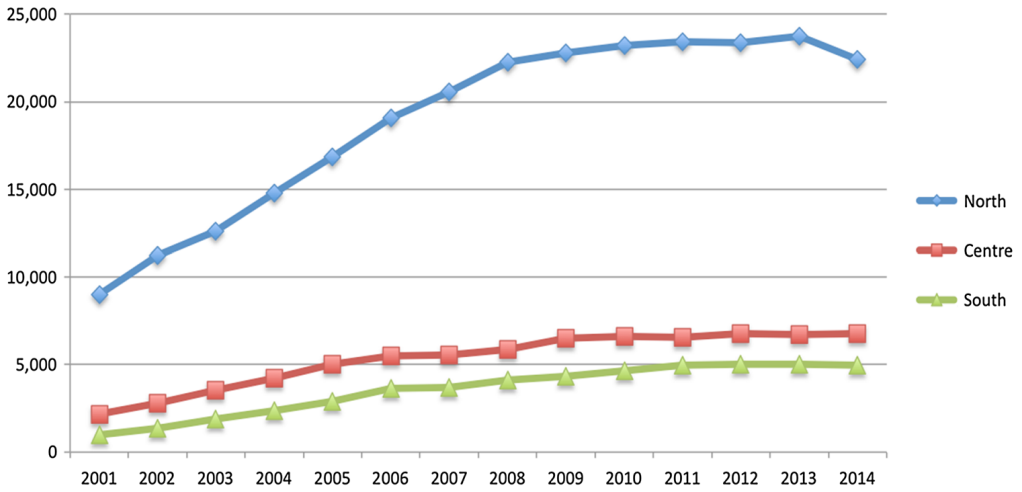

Figure 12 - Number of RC repair procedures for each year during the study period. From Longo et al. The burden of rotator cuff surgery in Italy: a nationwide registry

016-2610-x. Epub 2016 Dec 8.

Several factors can affect the risk of retears after RCT repair procedures. Many studies showed that the success of cuff repair decreases with advancing age and retears rates may be higher after 65 years. Chronicity of tears may play an indirect role as it affects

the quality of muscle and tendon unit with a higher degree of atrophy, fatty infiltration, and retraction. Tears characteristics directly affect the repair outcome. Small- and medium-sized tears carry greater chance of healing as compared with large and massive tears, and if two or more tendons are involved, it reduces the chance of healing and leads to poor outcome.

The last half-decade has seen significant advances in our understanding of the natural history of rotator cuff disease. Studies have continued to provide valuable insight into the clinical, radiographic, and anatomic features of these atraumatic tears. Healing data in postoperative patients has allowed us to refine surgical indications based on predictive factors for successful or unsuccessful repair.

One study has provided an algorithm for surgical indications by classifying RCTs into three categories based on the relative risks of nonoperative management and potential benefits of surgical management (77). Group I, or the early operative repair group, includes shoulders with a full-thickness tears as a result of a distinct, acute event as well as younger patients (under 62–65 years) with small to medium full-thickness tears with minimal or no atrophy. Given recent natural history data, an increased benefit to early surgical intervention may be amplified if the tear is greater than 15–20 mm in size or when there is disruption of the anterior cable attachment of the supraspinatus. Consideration for early surgical intervention is given based on the established risks of tear enlargement and progression of muscle atrophy in a patient still able to heal the repair. Group II includes partial thickness tears and atraumatic degenerative tears with signs of chronicity. These tears warrant a trial of initial nonoperative management and may respond to surgery if nonoperative management fails. Group III includes tears that are unlikely to heal. Nonoperative treatment should be maximized. These tears include tears in older patients (> 70 years), chronic and retracted full-thickness tears with muscle degeneration, and tears with fixed proximal humeral migration. Conservative treatment should be maximized for these patients. In a recent review A.A. Narvani et al (78) concluded that a non-surgical route with an appropriate physiotherapy program has a role in management of degenerative RCTs. This is especially the case in patients with significant risk factors for surgery, those who do not wish to go through a surgical treatment, as well as those with small, partial and irreparable tears.

ROLE OF GENETIC PREDISPOSITION

Genetics has recently been proposed as a factor involved in the pathogenesis of tendinopathies (79-82). Several studies show its potential role in RCTs, but evidence is still limited (83). Several studies suggest a familial genetic predisposition in the pathogenesis of rotator cuff disease (24, 84-86). Yamaguchi et al. (24) observed a strong trend between family history of RCT and risk of being affected by RCT (10.9% compared with 6.9%; $p = 0.06$), in 586 patients who underwent diagnostic shoulder ultrasound for shoulder pain. Family members of patients with RCT have a significantly higher risk to suffer RCT than general population (24, 87) . According to Tashjian et al. the increased risk of tears in family members of patients with RCTs extends out and beyond third-cousin relationships (88). The analysis of relatedness from this population-based, multi-generational level study shows significantly elevated risks for first- and second-degree relatives (88). It is particularly evident in relatives of young patients, who represent the best group to study genetic predisposition (88). Harvie et al. (85) compared the relative risk of full-thickness tears in siblings and spouses of 205 patients diagnosed as having full-thickness RCTs by ultrasound. Genetic susceptibility was demonstrated by the higher risk of full-thickness tears in siblings (2.42 compared with controls, 95% confidence interval 1.77–3.31). Moreover, siblings had a higher risk of experiencing symptoms (4.65 compared with controls, 95% confidence interval 2.42–8.63). Accordingly, patients stratified by family history and matched for age, gender and environmental conditions show a different prevalence of RCTs (86). Genetic factors also seem to be involved in symptom presentation (85, 86, 89, 90) and tear progression (86). Heritable components of pain may present at any point of the sensorineural pathway of the cuff (85). Pain may be associated with enlargement of the tear. Specific gene abnormalities have not been identified (84). As rotator cuff disease is a multifactorial condition (29, 70, 71, 91-98), it is difficult to select a single gene involved in its pathology. The main difficulty in genetic studies is to understand the interactions between gene products with each other and with the environment.

The phenotypic expression of genetic susceptibility may manifest in the ultrastructure of the tendon. Genetic factors have been hypothesized to influence structural components of the tendon. Research on genetic predisposition to atraumatic RCT has

relied on classical linkage studies or genome-wide association studies (GWAS) to identify candidate genes with common susceptibility variants.

AIMS

The aim of this research is to identify genes implicated in the determination of complex genetic disorders such as tendinopathies. In particular, our study will focus on atraumatic RCT, a disease for which no causative genes have been identified so far in a significant fraction of patients. For this reason, we decided to study a family in which RCT segregates with an autosomal dominant pattern with apparently complete penetrance and in which the presence of variants in known genes implicated in shoulder diseases has been excluded. The goal, therefore, is to identify a genetic variant in a new gene that can be held responsible for the phenotype. To achieve this goal will be analyzed and compared the exomes of two affected subjects and a healthy subject and then, through procedures of filtering and prioritization of variants common to affected subjects, we will obtain a list of variants in candidate genes. In future, the presence of these variants in family members will be confirmed by Sanger sequencing and through a family segregation analysis we will be able to identify variants that co-segregate in affected subjects of the family. The segregating variant will be the starting point for subsequent studies aimed at identifying the role of the identified gene in a larger cohort of patients with RCT and to try to highlight the molecular pathogenetic mechanisms underlying this disease.

MATERIALS AND METHODS

STUDY DESIGN

The study is an observational study. The study was approved by the local ethics committee (Ethics Committee Campus Bio-Medico University of Rome, 02/2010 prot. INT ComEt CBM). All patients signed informed consent before inclusion.

INCLUSION AND EXCLUSION CRITERIA

Patients were included in the study if they presented the following conditions: male and female > 18 years old of age with atraumatic rotator cuff lesions documented at MRI, without any history of significant acute trauma of the shoulder; bilateral disease and at least an affected 1st degree relative; caucasian patients of Italian origin. Patients were excluded if they presented the following conditions: radiographic evidence of osteoarthritis of glenohumeral joint; unilateral disease and absence of affected relatives; other surgical treatment of the shoulder; frozen shoulder; neurological deseases or language barriers.

PARTICIPANTS

Figure 13 shows the family tree of the family under investigation in which patients I-2 and II-1 were diagnosed with atraumatic bilateral RCT. Exome sequencing was performed on two affected subjects and a healthy subject. Genes shared between affected mother and son were selected, excluding those in common with the healthy son. Rotator cuff repair has been performed on son with RCT.

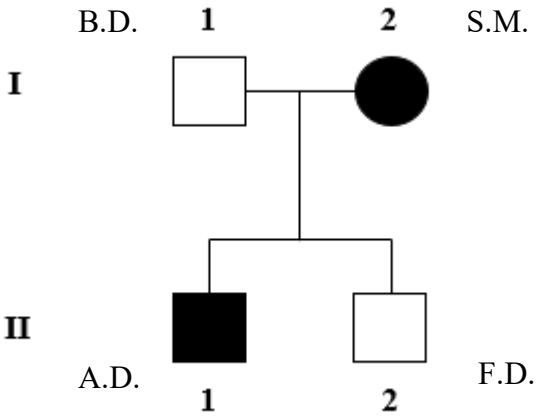

Figure 13 - Pedigree of the family enrolled. The symbols in black indicate the subjects affected by bilateral RCT. The presence of multiple cases in the same family, as indicated in the pedigree, suggests an autosomal dominant transmission of the disease.

The clinical picture of the patients under study is given below:

- Subject I-1 didn't report shoulder pathology during his life.
- Subject I-2 reported bilateral atraumatic RCT at a young age and underwent RCT repair procedure for each side.
- Subject II-1 reported bilateral atraumatic RCT at a young age and underwent RCT repair procedure in our Istitution for each side. During the surgery portions of healthy and torn tendon have been taken.
- Subject II-2 didn't report shoulder pathology during his life.

DNA extraction from peripheral blood

After signing informed consent for participation in the present study, DNA from patients I-2, II-1 and II-2 was extracted from peripheral blood and prepared by the Genetics Department of our institution according to the following protocol:

1) Subject blood to hypotonic lysis with RBC lysis buffer;
2) Centrifuge at 1500 rpm for 15 minutes and remove the supernatant without disturbing

the white blood cell pellet at the bottom (repeat steps 1 and 2 again);

3) Resuspend the pellet in 5 mL of SE with 100 µL Proteinase K (10 mg/mL) and 250 µL

of 20% SDS;

4) Incubate overnight at 37°C or 55°C for 4 hours;

5) Add 2 mL of 5M NaCl and 7 mL of Chloroform;

6) Shake for 30 minutes and then centrifuge at 2500 rpm for 30 minutes;

7) Transfer the supernatant and add 70% Isopropanol to perform an alcoholic precipitation of the DNA by dehydration;

8) Transfer the pellet to a 1.5 mL eppendorf and add 1 mL of 70% Ethanol;

9) Wash for 10 min, remove the Ethanol and air dry or in the lyophilizer, then resuspend in TE at pH 8;

After extraction, DNA samples were subjected to measurement and quality control using a spectrophotometer. Protein traces show a spectrum with an additional peak at 280 nm. For good quality DNA the ratio of absorption values at 260 nm and 280 nm should be about 1.8. The extracted DNA samples, after assessing their quality and concentration, were stored appropriately coded at +4°C.

RBC lysis buffer

- 8.3 g NH_4Cl
- 1 g $KHCO_3$
- 200 µL 0.5M EDTA at pH 8
- Make up to a volume of 1 L with double-distilled H_2O

SE

- 4.38 g NaCl
- 50 mL 0.5M EDTA at pH 8

TE (Tris-EDTA) at pH 8

- 1 mL 1M Tris-Cl at pH 8
- 0.2 mL 0.5M EDTA at pH 8

- Make up to a volume of 100 mL with double distilled H2O

Exome analysis by NGS

Exomes from patients I-2, II-1, and II-2 were sequenced by DanteLab SRL (Pizzoli, AQ), requiring a depth of approximately 100X on an Illumina sequencing platform, resulting in an average coverage of approximately 60X. The obtained file, in tabular CSV format with the annotated variants, was filtered using the functions present in Microsoft Excel. Variants in common between the two affected subjects (I-2 and II-1) were subsequently selected and those in common with the healthy subject (II-2) discarded, then prioritized using the online software VarElect. In particular, the literature was queried with the keywords "tendon" and "tendinopathy". To obtain information on the frequency of each variant, a search of the gnomAD database was performed. The potential pathogenicity of the variants was investigated by exploiting the predictions of several in silico tools present in VarSome.

FLOW CHART

Enrolment: inclusion and exclusion criteria

Blood samples

Surgical repair and tissue sample

Whole exome sequencing and genetic analysis

RESULTS

In the family, the diagnosis of bilateral atraumatic RCT was made in the mother (I-2), and son (II-1) while her brother (II-2) is in apparent good health. The exomes of patients I-2, II-1, and II-2 were sequenced, and the approximately 600000 variants obtained for each patient were filtered for coding and splicing regions and excluded synonymous variants, based on the autosomal dominant transmission pattern of RCT in this family. Variants were also filtered taking into account their frequency in the population. The 1000 Genomes database was taken as reference and variants with a frequency greater than 1% were excluded, obtaining an average of about 5000 variants for each patient. Their frequency was subsequently checked in the gnomAD database to exclude common variants (>1%). Variants in common with the healthy subject (II-2) were discarded, obtaining a total of 348 rare variants among 248 genes present only in the affected mother and son. To prioritize them, the literature was queried through the software VarElect with the keywords "tendon" OR "tendinopathy", to evaluate if and which genes obtained from the filtering were implicated in the onset of tendon pathologies. From this search, a list of about 57 candidate genes was generated, whose variants were manually checked in the BAM file to include the real ones, with a good coverage and that did not fall into polymorphic regions. Potential association with disease development was investigated according to the OMIM database (https://omim.org/), and those already associated with a disease were excluded. The risk of damaging was calculated using the VarSome Software (https://varsome.com/), and those found not to be damaging according to the prediction were excluded. Thanks to these additional steps, a list of four genes and four variants was obtained (Table 1).

gnomAD %	Tools Varsome	OMIM	Chr	Start	End	Reference	Alteration	Genome sequence	Gene	Exonic Function	Genes' reference
0,062%	8/12 damaging	N/A	5	177673296	177673296	G	A	exonic	COL23A1	Non-synonymous SNV	COL23A1:NM_173465:exon24:c.C1372T:p.P458S
absent	8/12 damaging	N/A	6	15496817	15496817	A	C	exonic	JARID2	Non-synonymous SNV	JARID2:NM_001267040:exon7:c.A845C:p.Q282P
0,21%	1/1 damaging	N/A	6	30681083	30681101	AAAAGGCGGCCCAAGGCCG	-	exonic	MDC1	Frameshift deletion	MDC1:NM_014641:exon5:c.618_636del:p.G207Pfs*3
absent	VUS/CADD 25	N/A	22	50686379	50686379	G	A	exonic	HDAC10	Non-synonymous SNV	HDAC10:NM_001159286:exon12:c.C1217T:p.P406L

Table 1 *- List of candidate variants.*

The presence of the variants shown in Table 1 will need to be confirmed by Sanger sequencing in unaffected subject I-1 to conduct a familial segregation analysis. The allele frequency in gnomAD of the variant c.C1372T:p.P458S in the COL23A1 gene is 120/191178 (0.0006277), of the variant c.618_636del:p.G207Pfs*3 in the MDC1 gene is 574/266980 (0,0021), while the variant c.C1217T:p.P406L of the gene HDAC10 and the variant c.A845C:p.Q282P in the JARID2 gene have never been reported before. The analysis of the effect of the variants on the protein products of the respective four genes was conducted on the VarSome database, which integrates the results of several in silico tools to predict their functional effect: the variant c.C1372T:p.P458S in the COL23A1 gene, and the variant c.C1217T:p.P406L in the HDAC10 gene is associated with four "damaging" predictions reported by the in silico prediction tools FATHMM-MKL, MutationTaster, DANN and SIFT; the variant c.A845C:p.Q282P in the JARID2 gene is associated with five "damaging" predictions reported by DANN, FATHMM-MKL, MutationTaster, SIFT and LRT; the variant c.618_636del:p.G207Pfs*3 in the MDC1 gene is absent in gnomAD. In Table 2 it is possible to appreciate, for each of the four variants, the "damaging" and "conservation" scores, respectively reported by the in silico prediction tools FATHMM-MKL, MutationTaster, DANN, SIFT, LRT and GERP.

	COL23A1: p.P458S	**JARID2: p.Q454P**	**MDC1: p.G207Pfs*3**	**HDAC10: p.P426L**
FATHMM-MKL	Damaging	Damaging	/	Damaging
MutationTaster	Disease causing	Disease causing	/	Disease causing
DANN	0.9967	0.9952	/	0.9945
SIFT	Damaging	Damaging	/	Damaging
LRT	Neutral	Deleterious	/	Neutral
GERP	4.9499	5.38	4.6758	4.3299

Table 2 - In silico predictions of the variants identified in the study*. The range of values provided by DANN is from 0 to 1, with 1 considered as the value of maximum damage of the variant on the protein product. The range of values provided by GERP is from -12.3 to 6.17, with 6.17 considered as the value of maximum degree of preservation of the amino acid residue affected by the variant.*

The different in silico prediction tools agree in defining the effect of the variant in the COL23A1, JARID2, HDAC10 genes as detrimental to their protein product, while for the variant MDC1:p.G207Pfs*3 the predictions are conflicting. Based on the scores provided by GERP, however, the MDC1 variant alters a well-conserved residue, increasing the likelihood that it may have functional significance.

DISCUSSION

The aim of this study was to shed light on the heritability in RCT, a multifactorial orthopaedic disease characterized by pain and loss of function of the shoulder. Only a small part of the predicted heritability have been accounted through previous linkage and association studies. It's impact on patients, their families, and society is significant. According to the prediction model by Longo et al (67), hospital costs sustained by the national health care system in Italy for RC procedures are expected to be over 1 billion euros by 2025. These facts highlitght the importance of understanding the genetic susceptibility factors involved so that better prevention, detection, and treatment strategies may be developed. Several genes are responsible for rotator cuff disease. Genetic susceptibility may affect the ultrastructure of the tendon. Achilles tendinopathy has been associated with polymorphisms of tenascin C and collagen type Va (99). Similar mechanisms could play a role in the pathogenesis of rotator cuff disease. The genetic basis of this condition may also result from aberrations in the normal cell regulation of apoptosis and tissue regeneration.

J. Jancuska et al (100) recently performed a summary of systematic reviews on the topic of the rotator cuff. They found 4 reviews focused on histology and genetics of RCT (101-104) . Dean et al (101) systematically reviewed 101 studies that examined cellular and molecular changes involved in pathogenesis of RC disease (Tables 1 and 2).

Changes to the Extracellular Matrix Components and Enzymes in Rotator Cuff Disease[a]

Matrix Components	Matrix Enzymes
Type I collagen ↑[116]	MMP-1 ↑[19,63,89,103,105,156,234] ↓[108]
Type II collagen ↑[57,161]	MMP-2 ↑[156] ↑ (ftRCT vs ptRCT[224])
Type III collagen ↑[10,87,100,116,180,182] ↑ (RCT vs non-RCT[82])	MMP-3 ↑[87,156] ↓[103,105,108,117] ↑ (ftRCT vs ptRCT[247])
Type X collagen ↑[161]	MMP-9 ↑[19,103,105,209,234] ↑ (ftRCT vs ptRCT[224])
Type I collagen α1 ↓[8] ↑ (ftRCT vs ptRCT[209])	MMP-13 ↑[83,108,117,156,209]
Type I collagen α2 ↓[6,85]	TIMP-1 ↓[117]
Type II collagen α1 ↑[6,8,85]	TIMP-2 ↓[117]
Type III collagen α1 ↑[8,231] ↓[6]	TIMP-3 ↓[6,85]
Type VI collagen ↑[7] α1 ↑[8]	ADAM10 ↓[6]
Collagen crosslinking ↑[10]	Transglutaminase 2 ↓[152]
Total collagen content ↓[10,180,182]	Other Enzymes
Calcium phosphate ↑[180]	COX-1 ↑[19,234]
Aggrecan ↑[6,8,85,116]	COX-2 ↑[19,169,209,234]
Biglycan ↑[8]	Cathepsin D ↑[63]
Decorin ↑[8] ↓[6,116]	iNOS ↑[209,222]
Clusterin ↑[6,135]	eNOS ↑[222]
Elastin ↓[6]	Transcription Factors
Fibronectin ↑ (RCT vs non-RCT[230])	SOX9 ↑[6,8]
Osteopontin ↑[225]	FOXO1A ↑ (massive tears[199])
Tenascin-C ↑[57,80]	FOXO3A ↑ (in tears greater than one-third[199])
Versican ↑[8]	
GAG content ↑[8,180]	
Chondroitin sulphate ↑[7,57,180,181]	
Dermatan sulphate ↑[7,180,181]	
Hyalauronan ↑[181]	
Hyaluronic acid ↑[181]	
α-Skeletal muscle actin and of myosin heavy polypeptide 1 ↑[55]	

Tab.1 [a]Includes changes to other enzymes and transcription factors. ↑, increased; ↓, decreased. Used with permission from Dean et al. ADAM, a disintegrin and metalloproteinases; COX, cyclooxygenase; FOX, forkhead box protein; ftRCT, full-thickness rotator cuff tear; GAG, glycosaminoglycan; MMP, matrix metalloproteinase; NOS, nitric oxide synthase; ptRCT, partial-thickness rotator cuff tear; RCT, rotator cuff tear; SOX9, sex-determining region Y-box 9 protein; TIMP, tissue inhibitor of metalloproteinase (101).

Changes to Cytokines, Growth Factors, Neuronal Factors, Apoptosis/Cell Cycle–Related Factors, and Other Factors in Rotator Cuff Disease[a]

Cytokines/Growth Factors	Apoptosis/Cell Cycle Related
IL-1α ↑[19,234]	HIF-1α ↑[15,103,135] ↓[136,137,140]
IL-1ra ↑[60–62]	BNIP-3 ↑[15]
IL-1β ↑[18,19,60–62,99,191,209,234]	BCL-2 ↑[135]
IL-2 ↓[136,140]	Caspase 3 ↑[136,137]
IL-6 ↑[18,19,99,136,209,234]	Caspase 8 ↑[136,137,140]
IL-11 ↑[136,140]	Heat shock protein 27 ↑[136,137,140]
IL-15 ↑[136]	Heat shock protein 70 ↑[136,137,140]
IL-18 ↑[136]	Poly(ADP-ribose) polymerase ↑[136,140]
Stromal derived factor 1α ↑[18,94]	Type-2 angiotensin II receptor ↓[136,140]
TNFα ↑[19,99,191,209,234]	cFLIP ↑[137]
VEGF ↑[103,104,135,169,209] ↑ (associated with motion pain[242])	cFLIP receptor ↑[137]
IGF-1 ↓[6,85]	p-53 induced gene I, cell division cycle 25A, Max protein, meiotic recombination 11 homolog A ↑[140]
TGF-β ↑[161,191]	Peroxiredoxin 5 ↑[237]
bFGF ↑[161,191]	P53 ↑[121,140]
FGF 18 ↑[140]	P53 inhibitors ↓[121]
BMP2 and BMP7 ↑[161]	NF-KB ↓[121]
Small inducible cytokines ↑[19]	Receptor activator of NF-KB ↑[121]
Macrophage inhibitory factor ↑[136]	
Heparin affinity regulatory peptide ↑[8]	
Five-lipoxygenase activating protein ↑[169]	
Hepatocyte growth factor ↓[140]	
Neuronal Factors	**Others**
Substance P ↑[89] (higher in nonperforated RCTs vs perforated[62])	Ubiquitin proteasome pathway UBE2A and UBE3A ↑ (massive tears vs small/controls[199])
β-endorphin ↑[89]	Calpain (CAPN1) and CTSB (lysosomal enzyme) ↑ (massive tears vs small/controls[199])
Anti-NGF30 ↑[136,140]	vWF ↑[169]
PGP9.5, GAP43 ↑[241]	T-cell receptor variable βchain ↑[136]
6, glutamate receptor inotropic 3A, GABA receptor α1 ↑[140]	Ig chain, T-cell receptor α chain ↓[136]
AMPA1, glutamate receptor interacting protein 1/2 ↓[140]	Insulin induced gene 1, FGFr1, nuclear receptor coactivator 2, G protein coupled receptor 54, Ephrin A1, Thyrotroph embryonic factor, Odd Oz/ten-m homolog 2, POU domain, TNF 11, TGF-β binding protein 3, T-cell receptor β chain, cytochrome b-245, CD3 γ chain, polyprotein 1-microglobulin, Fc receptor IgE, solute carrier family 2, adenosine deaminsae, integrin-linked kinase ↑[140]
	Dynein, nuclear receptor subfamily 2 group F member 1, Homeobox A1, FGF receptor 3, MHC class I-like sequence, T-cell receptor β chain, killer cell lectin-like receptor, strain T-cell receptor ↓[140]
	T-cell receptor ↓[136,140]

Tab. 2 [a]↑, increased; ↓, decreased. Used with permission from Dean et al. AMPA, alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; ADP, adenosine diphosphate; BCL, B-cell lymphoma; BMP, bone morphogenetic protein; cFLIP, cellular FLICE (FADD-like IL-1beta-converting enzyme)-inhibitory protein; CTSB, cathepsin B; FGF, fibroblast growth factor; GABA, gamma-aminobutyric acid; GAP43, growth-associated protein 43; HIF, hypoxia-inducible factor; IGF, insulin-like growth factor; IL, interleukin; NF-KB, nuclear factor kappa light chain enhancer of activated B cells; NGF, nerve growth factor; PGP9.5, protein gene product 9.5; RCT, rotator cuff tear; TGF, transforming growth factor; TNF, tumor necrosis factor; UBE, ubiquitin conjugating enzyme; VEGF, vascular endothelial growth factor; vWF, Von Willebrand factor (101)

Degenerative RCT histologic changes consists in inflammation and tendon healing. The progressive formation of collagen fibers, mainly types II and III, are associated with loss of the normal tendon structure and myxoid degeneration. Proinflammatory cytokines accelerate remodeling, promote tenocyte apoptosis, creating an imbalance between the catabolic and anabolic systems. In the catabolic state, there is a decrease in tissue inhibitors of metalloproteinases (TIMPs), resulting in higher levels of metalloproteinase, a matrix-remodeling protein. Furthermore, TIMPs may play a role in the degradation of proinflammatory cytokines, and their relative decrease may contribute to the local inflammatory state. As the total collagen content decreases,

smaller and more disorganized fibrils form. The increase in tenascin-C and fibronectin is consistent with the wound-healing process. Two systematic reviews examined gene expression and protein composition in RC tendons. (102, 103). In 2017, Dabija et al. (103) reviewed the literature on this topic describing the results of 4 studies investigating familiar predisposition. These results demonstrated that siblings of an individual with a RCT were more likely to develop a full-thickness tear and more likely to be symptomatic. A 5-year follow-up showed that the relative risks were increased for the siblings to have a full-thickness tear, for a tear to progress in size, and for being symptomatic. A significantly higher number of individuals with tears had family members with a history of tears or surgery than those without tears did. The other 3 studies investigated whether a genetic predisposition to rotator cuff disease exists and found significant association in seven candidate genes: in DEFB1, FGFR1, FGF3, ESRRB, and FGF10 and 2 single-nucleotide polymorphisms within SAP30BP and SASH1 (105-107).

DEFB1 (Defensin, Beta 1) encodes the protein antimicrobial peptide defensin β-1, which aids in preventing epithelial surfaces from being colonized by microbes. The rs1800972 C>G variant was significantly more frequent in individuals with rotator cuff disease (105).

ESRRB (estrogen-related receptor beta) encodes a protein similar to the estrogen receptor and is believed to have an inhibitory effect on estrogen signaling (108). In vitro studies have demonstrated a correlation between estrogen deficiency and poor tendon healing (109), implying a possible role ESRRB may have in rotator cuff disease.

FGF3 (fibroblast growth factor 3) and FGF10 (fibroblast growth factor 10) encode fibroblast growth factor proteins and are involved in a number of processes such as cell growth and tissue repair, including tendons, and could thus be associated with the pathogenesis of rotator cuff disease. FGFR1 encodes one of the receptors also associated with fibroblast growth factor; however, this gene is more specific to limb development.

SAP30BP is implicated in cell death. SASH1 is a tumor suppressor gene implicated in a number of cancers (110, 111). Thus, many of the single nucleotide polymorphisms

(SNPs) have a potential biologic mechanism for their association with rotator cuff disease, but further research is needed in this area.

Sejersen et al (102) found that RC tears were associated with increased BNip3 in 1 study and increased expression of hypoxia inducible factor 1α, vascular endothelial growth factor, and metalloproteinases 1 and 9 in 4 studies. Moreover, the authors found that 2 studies cited no correlation between apoptotic and cytokine gene expression and tear size or histologic grade among patients with supraspinatus tears, while 1 study did not find a correlation between tear size and apoptotic markers.

8 studies investigating genes variations associated with RCTs were included in a review performed by Longo et al.. Six of these studies were focused on candidate genes (105, 106, 112-115) and two studies were GWASs (116, 117). The following candidate genes were investigated: DEFB1, DENND2C, ESRRB, FGF3, FGF10, FGFR1, MMP-1, MMP-2, MMP-3, MMP-9, MMP-13, Col5A1, TNC, FOXP3, FCRL3. A significant association between SNPs and rotator cuff disease was found for DEFB1, FGFR1, FGFR3, ESRRB, FGF10, MMP-1, TNC, FCRL3. Contradictory results were reported for MMP-3 (112, 113). The two GWASs identified the following locus associated with RCTs: SASH1- rs12527089, SAP30BP - rs820218, rs71404070 located next to cadherin8 (116, 117). Recent studies on candidate genes and GWASs draw attention to SNPs associated with rotator cuff disease (91-93). Genetically susceptible patients experience symptoms more often (85, 89, 90, 118).

Tashjian et al's 2014 study (n = 92) demonstrated that a significantly higher number of individuals with RCTs (32.3%) also had family members with a history of RCTs or surgery when compared to individuals without RCTs (18.3%) (119). S.E. Gwilym et al evaluated the progression of a RCT, in terms of its size, from a group of siblings of patients with a RCT and of controls subjects without a first-degree family history of rotator cuff disease. According to them, a tear in the sibling of a patient with a painful tear has a relative risk of being painful of 1.44, compared to one in a control subject. Moreover, the progression of a tear over a five-year period, is greater in siblings than in controls (tear size increased in 16.1% of siblings, compared with 1.5% of control group). Comparing the presence and size of the tears to five years previously, the relative risk progression in the sibling group compared to the control group was 2.08.

Genetic factors have a role, not only in the development but also in the progression of full-thickness tears of the rotator cuff (86).

Preliminary evidences of genetic and familiar predisposition to RCTs provided the basis for further studies that better highlight the importance of the genetic component in the pathogenesis of rotator cuff disease.

SNPs are the most common type of genetic variation among people. Each SNP represents a difference in a single DNA building block, called a nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA. Although certain definitions require the substitution to be present in a sufficiently large fraction of the population (e.g., 1% or more), many publications (3, 120, 121) do not apply such a frequency threshold. Rare variants are alternative forms of a gene that are present with a minor allele frequency of less than 1%. One of main contributions of SNPs in clinical research is GWAS. Genome-wide genetic data can be generated by multiple technologies, including SNP array and whole genome sequencing. GWAS has been commonly used in identifying SNPs associated with diseases or clinical phenotypes or traits. Since GWAS is a genome-wide assessment, a large sample site is required to obtain sufficient statistical power to detect all possible associations.

Two methods, whole exome sequencing and whole genome sequencing, are increasingly used in healthcare and research to identify genetic variations. With next-generation sequencing, it is now feasible to sequence large amounts of DNA, for instance all the pieces of an individual's DNA that provide instructions for making proteins. These pieces, called exons, are thought to make up 1% of a person's genome. Together, all the exons in a genome are known as the exome, and the method of sequencing them is known as whole exome sequencing. This method allows variations in the protein-coding region of any gene to be identified, rather than in only a select few genes. Because most known mutations that cause disease occur in exons, whole exome sequencing is thought to be an efficient method to identify possible disease-causing mutations.

However, researchers have found that DNA variations outside the exons can affect gene activity and protein production and lead to genetic disorders that whole exome sequencing would miss. Another method, called whole genome sequencing,

determines the order of all the nucleotides in an individual's DNA and can determine variations in any part of the genome. Whole genome sequencing is able to identify several significant hits within intergenic regions.

Nevertheless, is still expensive to perform whole genome sequencing on a large cohort of samples, so reducing the cost by sequencing the most informative regions is a desirable approach.

The human exome consists of 1% of the human genome but harbors 85% of disease-related variants (122). Therefore, the cost of exome sequencing is typically only one-sixth that of whole genome sequencing (123). Several commercial exome-capture platforms are currently available, each with a different design focus (124-126). In a study conducted by Sean Lacey et al (127) they compared the performance of whole genome sequencing with exome sequencing in a family-based association study. After correcting for multiple testing, they did not find great benefit from whole genome sequencing compared to exome sequencing. These results suggest that exome sequencing is a cost-effective way to capture disease-related variants. Given the lower cost, exome sequencing allows a larger number of samples to be sequenced, which would significantly increase the statistical power for association studies.

Though many genetic loci have been associated with RCT, very few of these have been replicated across studies. The main reason for this is believed to be the phenotypic and genotypic heterogeneity of the disorder. Thus, we have designed a study that aims at reducing heterogeneity of the studied sample through a family-based approach as well as focusing on a more homogenous subtype of the disorder: atraumatic or degenerative bilateral lesions in a family with specific criteria. Previous evidence justifies the examination of genetic predisposition in families, where the genetic susceptibility is more homogeneous than in unrelated cases. In this study, we took specific steps to reduce the effects of intrafamilial heterogeneity by excluding subjects with other shoulder diseases like osteoarthritis or frozen shoulder. Part of our strategy to reduce heterogeneity has been to focus on a particular form of RCT (atraumatic and reparable tears) and on selected family clusters as we deemed such an endophenotype should enrich the likelihood that affected individuals from a family would all share the same susceptibility variant(s). Our choice to pursue a high-throughput exome-sequencing approach enabled us to look for the possibility of rare variants in a number

of different genes across the family and across the cohort. Thus, in this context sequencing the exome should prove to be most valuable. Though not as comprehensive as whole genome sequencing, it allows for a much more feasible experiment, both from the financial and data management points of view. A limitation of exome sequencing is that it does not identify the structural and noncoding variants found by whole genome sequencing. In addition, most disease-predisposing variants identified so far tend to be in coding regions, suggesting that for new RCT susceptibility genes identified by our study, the mutations will also be in the coding regions.

In this study, to identify genes whose variants could be thought to be responsible for the disease, the exomes of two patients diagnosed with bilateral RCT were analyzed. In particular, following filtering procedures and prioritization of variants common to the two affected subjects, four missense variants in the COL23A1, JARID2, MDC1 and HDAC10 genes were highlighted, whose segregation, will be assessed by Sanger sequencing using a family segregation analysis, and will be confirmed to be consistent with the phenotype of the family members under investigation. The variants are, in fact, present in heterozygosity in all affected subjects and absent in one healthy subject. COL23A1 is a gene located in position 5q35 that encode for a transmembrane protein. Is a member of the transmembrane collagens, a subfamily of the nonfibrillar collagens that contain a single pass hydrophobic transmembrane domain. Using a differential display PCR analysis, Banyard et al. (128) identified an mRNA transcript that was highly overexpressed in a metastatic subline of rat prostate adenocarcinoma cells. By database searching with the full-length rat COL23A1 sequence, they identified a homologous human EST from retina. Using PCR-based methods, they amplified human COL23A1 cDNA from heart, brain, and a leukemia cell line. The deduced 540-amino acid protein is predicted to be a type II transmembrane protein (Figure 14). Using immunofluorescence cell localization of truncation mutants, Banyard et al. (2003) determined that rat COL23A1 is oriented with the C terminus on the cell surface. The N-terminal cytoplasmic and transmembrane domains are required for cell surface targeting. Using truncation mutants, Banyard et al. (128) demonstrated that the predicted furin-cleaved secreted portion of human COL23A1 also formed multimers and displayed low affinity binding to heparin in vitro.

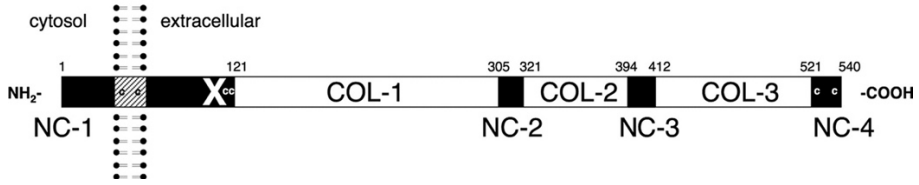

Fig. 14. Schematic representation of Collagen XXIII structure. *The sequence predicted domain structure of collagen XXIII is shown. Collagenous domains (*COL*) are* white boxes. *Noncollagenous domains (*NC*) are* black boxes. *The transmembrane domain is indicated with* cross-hatching. *The amino acid positions (human collagen XXIII) are listed across the* top. *The positions of cysteine residues are marked (*C*). The potential cleavage site is marked (*X*). From Banyard, J., Bao, L., Zetter, B. R. Type XXIII collagen, a new transmembrane collagen identified in metastatic tumor cells. J. Biol. Chem. 278: 20989-20994, 2003 (128)*

JARID2 gene was mapped to chromosome 6p24-p23 by Berge-Lefranc et al. (129) using autoradiographic in situ hybridization. They reported that the Human Jumonji (JMJ) and mouse Jmj gene products are 90% identical. Northern blot analysis revealed a low level of expression of JMJ in all human embryonic and adult tissues analyzed. In situ hybridization studies on embryonic slices revealed high levels of expression in dorsal root ganglia neurons. The authors detected high levels of expression in adult cerebral cortex. Toyoda et al. (130) determined that JMJ is expressed as a 160-kD protein by Western blot analysis. Immunofluorescence and Western blot analysis demonstrated that JMJ specifically localizes to the cell nucleus. Pasini et al. (131) showed that the Polycomb repressive complex-2 (PRC2) forms a stable complex with JARID2. Using genomewide location analysis, they found that JARID2 binds to more 90% of previously mapped Polycomb group target genes. The authors revealed that JARID2 is sufficient to recruit Polycomb group proteins to a heterologous promoter, and that inhibition of JARID2 expression leads to a major loss of Polycomb group binding and to a reduction of histone H3 lysine-27 trimethylation (H3K27me3) levels on target genes. Pasini et al. (131) concluded that JARID2 is essential for the binding of Polycomb group proteins to target genes and, consistent with this, for the proper differentiation of embryonic stem cells and normal development. In an animal model Toyoda et al. (132) found that Jmj was highly expressed in developing mouse cardiac ventricles. Jmj deficiency in mice caused hyperproliferation of embryonic trabecular myocytes. Jmj-deficient embryos showed enhanced expression of cyclin D1 (CCND1;

168461), but no other cyclin examined. They concluded that JMJ downregulates cardiac cell proliferation by repressing cyclin D1 expression. Takahashi et al [117] found that Jmj-deficient mice show several morphologic abnormalities, including neural tube and cardiac defects, and die in utero around embryonic day 11.5. They concluded that at least 2 lethal periods exist in Jmj mutant mice, with cardiac abnormalities causing the earlier lethality.

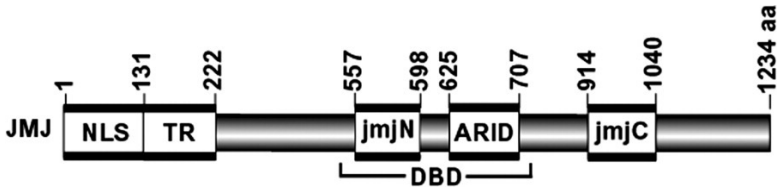

Figure 15. Schematic diagram of the structure of JUMONJI protein (JMJ). *The amino acid sequence of JMJ reveals the jmjN and jmjC domains and AT-rich interaction domain (ARID) motif as indicated. Structural and functional analyses by Kim et al. (2003) showed functional domains including the domains for nuclear localization signal (NLS), transcriptional repression (TR), and DNA binding (DBD) that mediates binding to DNA sequence (AT-rich) as well as protein–protein interaction. From Jung J, Mysliwiec MR, Lee Y. Roles of JUMONJI in mouse embryonic development. Dev Dyn. 2005 Jan;232(1):21-32. (133)*

Mediator of DNA damage checkpoint 1 (MDC1) is a gene located in position 6p21.33 that encode for a scaffold protein involved in the early steps of the DDR (DNA damage response), a process necessary to maintain genome integrity and to prevent the accumulation of oncogenic mutations. Consequently, proteins involved in the DDR often serve as tumor suppressors, carrying out the crucial task of keeping DNA fidelity intact. MDC1 interacts directly with γ-H2AX, the phosphorylated form of H2AX, a commonly used marker for DNA damage. It then propagates the phosphorylation of H2AX by recruiting ATM kinase. MDC1 is an emerging tumor suppressor because loss of MDC1 is associated with increased tumor frequency in mice (Minter-Dykhouse et al. (134)), and reduction or lack of MDC1 is observed in a significant proportion of carcinomas (135).

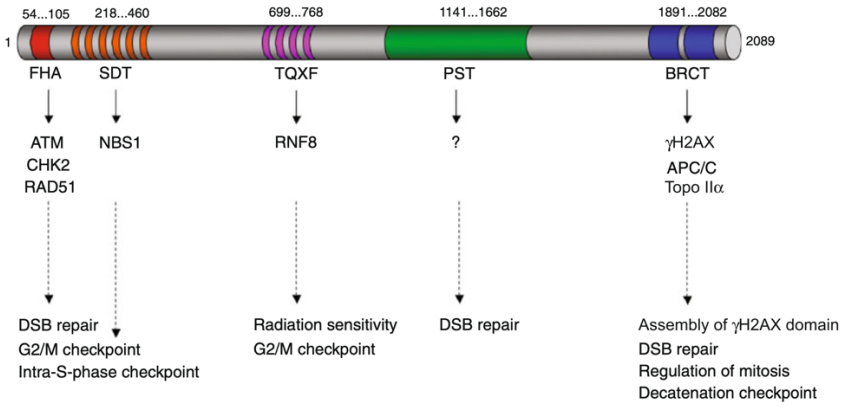

Figure 16. Schematic representation of the domain architecture of MDC1 and its interaction partners*. MDC1 is composed of several distinct domains and regions that either interact with phosphorylated proteins or are themselves phosphorylated and serve as docking sites for other proteins that contain phospho-specific interaction modules such as FHA and BRCT domains. Each of these domains and regions appears to be functionally relevant for the DDR and/or for control of the cell cycle. From Jungmichel S, Stucki M MDC1: The art of keeping things in focus. Chromosoma. 2010 Aug;119(4):337-49. (136)*

While the function of MDC1 in the DDR has been reviewed previously, its role in cancer has not been reviewed, and numerous studies have recently identified a link between MDC1 and carcinogenesis. This includes MDC1 functioning as a tumor suppressor, with its loss serving as a biomarker for cancer and contributor to drug sensitivity. Studies also indicate that MDC1 operates outside of its traditional role in DDR, and functions as a co-regulator of nuclear receptor transcriptional activity, and that mutations in MDC1 are present in tumors and can also cause germline predisposition to cancer (137).

Human HDAC10 is a gene located in position 22q13.33that contains 20 exons and encodes at least two and possibly more differentially spliced isoforms. The largest HDAC10 isoform encodes a protein of 669 amino acids and contains an amino-terminal histone deacetylase catalytic domain. Similar to other class II HDACs, HDAC10 can localize to either the nucleus or cytoplasm. Northern blot analyses indicated that the expression of HDAC10 may be ubiquitous with the highest expression in the liver, spleen, and kidney (138). It also has potential clinical utility for targeting tumors and non-tumor diseases, such as renal cell carcinoma, prostate cancer, immunoglobulin A nephropathy (IgAN), intracerebral hemorrhage, human

immunodeficiency virus (HIV) infection and schizophrenia. It may be involved also in other non-tumor diseases affecting respiratory, urinary and nervous systems and hormone regulation and lipid metabolism (139).

With the exome sequencing performed so far, the four variants in the COL23A1, JARID2, MDC1, HDAC10 genes segregated with the phenotype of the family members. Regarding the four genes, variant c.A845C:p.Q282P in the JARID2 gene appears to be the most promising as it is considered "damaging" by the in silico prediction tools DANN, FATHMM-MKL, MutationTaster, SIFT and LRT.
In addition, the "oe value" (ratio of the number of loss-of-function variants observed in a given gene over the expected ones) reported by gnomAD for the COL23A1 (0,63), JARID2 (0,09) AND MDC1 (0,39), indicates that the expected loss-of-function variants are indeed greater than those observed and suggesting, therefore, that the genes are intolerant to such variants, it supports our theory that a haploinsufficiency condition linked to loss of function of COL23A1 (0,63), JARID2 (0,09) and MDC1 may be involved in the determination of the RCT. HDAC10 "oe value" of 1,17 suggests, therefore, that this gene could be tolerant to such variant. Through subsequent functional studies on COL23A1: p.P458S, JARID2: p.Q454P, MDC1: p.G207Pfs*3 and HDAC10: p.P426L, aimed at identifying pathogenetic mechanisms related to these variants, and their analysis in a larger cohort of patients, it will be possible to support our proposed hypothesis that COL23A1, JARID2, MDC1 and HDAC10 may be implicated in the pathogenesis of RCTs. Furthermore, proposed susceptibility genes will have to be validated functionally for their role in the tendon as well as the impact of identified rare mutations on protein structure and function. To this purpose, at a later stage, a histological analysis of the tendon sampled during surgery will be performed. The long-term aim is to define the molecular basis of RCT to delineate preventive measures and targeted therapy and in second hand to decrease direct and indirect costs to the health care system.

BIBLIOGRAPHY

1. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977;74(2):560-4.
2. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
5. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science. 1987;238(4825):336-41.
6. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402.
7. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51.
8. Teer JK, Mullikin JC. Exome sequencing: the sweet spot before whole genomes. Hum Mol Genet. 2010;19(R2):R145-51.
9. Ebbert MT, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics. 2016;17 Suppl 7:239.
10. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
11. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
12. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.
13. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. BMC Genomics. 2016;17 Suppl 2:444.
14. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013;34(1):57-65.
15. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-2.
16. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7.
17. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009;19(9):1553-61.

18. Keener JD, Wei AS, Kim HM, Steger-May K, Yamaguchi K. Proximal humeral migration in shoulders with symptomatic and asymptomatic rotator cuff tears. J Bone Joint Surg Am. 2009;91(6):1405-13.
19. Castagna A, Cesari E, Gigante A, Conti M, Garofalo R. Metalloproteases and their inhibitors are altered in both torn and intact rotator cuff tendons. Musculoskelet Surg. 2013;97 Suppl 1:39-47.
20. Garofalo R, Cesari E, Vinci E, Castagna A. Role of metalloproteinases in rotator cuff tear. Sports Med Arthrosc Rev. 2011;19(3):207-12.
21. Goutallier D, Postel JM, Van Driessche S, Voisin MC. [Histological lesions of supraspinatus tendons in full thickness tears of the rotator cuff]. Rev Chir Orthop Reparatrice Appar Mot. 2005;91(2):109-13.
22. Milgrom C, Schaffler M, Gilbert S, van Holsbeeck M. Rotator-cuff changes in asymptomatic adults. The effect of age, hand dominance and gender. J Bone Joint Surg Br. 1995;77(2):296-8.
23. Teunis T, Lubberts B, Reilly BT, Ring D. A systematic review and pooled analysis of the prevalence of rotator cuff disease with increasing age. J Shoulder Elbow Surg. 2014;23(12):1913-21.
24. Yamaguchi K, Ditsios K, Middleton WD, Hildebolt CF, Galatz LM, Teefey SA. The demographic and morphological features of rotator cuff disease. A comparison of asymptomatic and symptomatic shoulders. J Bone Joint Surg Am. 2006;88(8):1699-704.
25. Kim HM, Dahiya N, Teefey SA, Middleton WD, Stobbs G, Steger-May K, et al. Location and initiation of degenerative rotator cuff tears: an analysis of three hundred and sixty shoulders. J Bone Joint Surg Am. 2010;92(5):1088-96.
26. Kim HM, Dahiya N, Teefey SA, Keener JD, Galatz LM, Yamaguchi K. Relationship of tear size and location to fatty degeneration of the rotator cuff. J Bone Joint Surg Am. 2010;92(4):829-39.
27. Codding JL, Keener JD. Natural History of Degenerative Rotator Cuff Tears. Curr Rev Musculoskelet Med. 2018;11(1):77-85.
28. Chillemi C, Petrozza V, Garro L, Sardella B, Diotallevi R, Ferrara A, et al. Rotator cuff re-tear or non-healing: histopathological aspects and predictive factors. Knee Surg Sports Traumatol Arthrosc. 2011;19(9):1588-96.
29. Longo UG, Franceschi F, Ruzzini L, Rabitti C, Morini S, Maffulli N, et al. Histopathology of the supraspinatus tendon in rotator cuff tears. Am J Sports Med. 2008;36(3):533-8.
30. Tillander B, Franzén L, Norlin R. Fibronectin, MMP-1 and histologic changes in rotator cuff disease. J Orthop Res. 2002;20(6):1358-64.
31. Dalton S, Cawston TE, Riley GP, Bayley IJ, Hazleman BL. Human shoulder tendon biopsy samples in organ culture produce procollagenase and tissue inhibitor of metalloproteinases. Ann Rheum Dis. 1995;54(7):571-7.
32. Gardner K, Arnoczky SP, Caballero O, Lavagnino M. The effect of stress-deprivation and cyclic loading on the TIMP/MMP ratio in tendon cells: an in vitro experimental study. Disabil Rehabil. 2008;30(20-22):1523-9.
33. Osawa T, Shinozaki T, Takagishi K. Multivariate analysis of biochemical markers in synovial fluid from the shoulder joint for diagnosis of rotator cuff tears. Rheumatol Int. 2005;25(6):436-41.
34. Arnoczky SP, Lavagnino M, Egerbacher M, Caballero O, Gardner K. Matrix metalloproteinase inhibitors prevent a decrease in the mechanical properties of stress-deprived tendons: an in vitro experimental study. Am J Sports Med. 2007;35(5):763-9.
35. DE Giorgi S, Saracino M, Castagna A. Degenerative disease in rotator cuff tears: what are the biochemical and histological changes? Joints. 2014;2(1):26-8.
36. Castagna A, Cesari E, Garofalo R, Gigante A, Conti M, Markopoulos N, et al. Matrix metalloproteases and their inhibitors are altered in torn rotator cuff tendons, but also in the

macroscopically and histologically intact portion of those tendons. Muscles Ligaments Tendons J. 2013;3(3):132-8.
37. Goodier HC, Carr AJ, Snelling SJ, Roche L, Wheway K, Watkins B, et al. Comparison of transforming growth factor beta expression in healthy and diseased human tendon. Arthritis Res Ther. 2016;18:48.
38. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Ann Rheum Dis. 1998;57(11):649-55.
39. Meislin RJ, Sperling JW, Stitik TP. Persistent shoulder pain: epidemiology, pathophysiology, and diagnosis. Am J Orthop (Belle Mead NJ). 2005;34(12 Suppl):5-9.
40. Codman EA, Akerson IB. THE PATHOLOGY ASSOCIATED WITH RUPTURE OF THE SUPRASPINATUS TENDON. Ann Surg. 1931;93(1):348-59.
41. Fukuda H, Hamada K, Yamanaka K. Pathology and pathogenesis of bursal-side rotator cuff tears viewed from en bloc histologic sections. Clin Orthop Relat Res. 1990(254):75-80.
42. Lehman C, Cuomo F, Kummer FJ, Zuckerman JD. The incidence of full thickness rotator cuff tears in a large cadaveric population. Bull Hosp Jt Dis. 1995;54(1):30-1.
43. Fukuda H, Mikasa M, Yamanaka K. Incomplete thickness rotator cuff tears diagnosed by subacromial bursography. Clin Orthop Relat Res. 1987(223):51-8.
44. Roquelaure Y, Ha C, Leclerc A, Touranchet A, Sauteron M, Melchior M, et al. Epidemiologic surveillance of upper-extremity musculoskeletal disorders in the working population. Arthritis Rheum. 2006;55(5):765-78.
45. Leclerc A, Franchi P, Cristofari MF, Delemotte B, Mereau P, Teyssier-Cotte C, et al. Carpal tunnel syndrome and work organisation in repetitive work: a cross sectional study in France. Study Group on Repetitive Work. Occup Environ Med. 1998;55(3):180-7.
46. Harryman DT, 2nd, Hettrich CM, Smith KL, Campbell B, Sidles JA, Matsen FA, 3rd. A prospective multipractice investigation of patients with full-thickness rotator cuff tears: the importance of comorbidities, practice, and other covariables on self-assessed shoulder function and health status. J Bone Joint Surg Am. 2003;85(4):690-6.
47. Neer CS. Impingement lesions. Clin Orthop Relat Res. 1983(173):70-7.
48. DeOrio JK, Cofield RH. Results of a second attempt at surgical repair of a failed initial rotator-cuff repair. J Bone Joint Surg Am. 1984;66(4):563-7.
49. Ellman H, Kay SP. Arthroscopic subacromial decompression for chronic impingement. Two- to five-year results. J Bone Joint Surg Br. 1991;73(3):395-8.
50. Snyder SJ, Pachelli AF, Del Pizzo W, Friedman MJ, Ferkel RD, Pattee G. Partial thickness rotator cuff tears: results of arthroscopic treatment. Arthroscopy. 1991;7(1):1-7.
51. Burkhart SS, Danaceau SM, Pearce CE. Arthroscopic rotator cuff repair: Analysis of results by tear size and by repair technique-margin convergence versus direct tendon-to-bone repair. Arthroscopy. 2001;17(9):905-12.
52. Lo IK, Burkhart SS. Arthroscopic repair of massive, contracted, immobile rotator cuff tears using single and double interval slides: technique and preliminary results. Arthroscopy. 2004;20(1):22-33.
53. Thompson WO, Debski RE, Boardman ND, Taskiran E, Warner JJ, Fu FH, et al. A biomechanical analysis of rotator cuff deficiency in a cadaveric model. Am J Sports Med. 1996;24(3):286-92.
54. Aluisio FV, Osbahr DC, Speer KP. Analysis of rotator cuff muscles in adult human cadaveric specimens. Am J Orthop (Belle Mead NJ). 2003;32(3):124-9.
55. Parsons IM, Apreleva M, Fu FH, Woo SL. The effect of rotator cuff tears on reaction forces at the glenohumeral joint. J Orthop Res. 2002;20(3):439-46.
56. Goutallier D, Postel JM, Zilber S, Van Driessche S. Shoulder surgery: from cuff repair to joint replacement. An update. Joint Bone Spine. 70. France2003. p. 422-32.

57. Maughan RJ, Watson JS, Weir J. Strength and cross-sectional area of human skeletal muscle. J Physiol. 1983;338:37-49.
58. Thomazeau H, Rolland Y, Lucas C, Duval JM, Langlais F. Atrophy of the supraspinatus belly. Assessment by MRI in 55 patients with rotator cuff pathology. Acta Orthop Scand. 1996;67(3):264-8.
59. Zanetti M, Gerber C, Hodler J. Quantitative assessment of the muscles of the rotator cuff with magnetic resonance imaging. Invest Radiol. 1998;33(3):163-70.
60. Schaefer O, Winterer J, Lohrmann C, Laubenberger J, Reichelt A, Langer M. Magnetic resonance imaging for supraspinatus muscle atrophy after cuff repair. Clin Orthop Relat Res. 2002(403):93-9.
61. Morag Y, Jacobson JA, Miller B, De Maeseneer M, Girish G, Jamadar D. MR imaging of rotator cuff injury: what the clinician needs to know. Radiographics. 2006;26(4):1045-65.
62. Nakagaki K, Ozaki J, Tomita Y, Tamai S. Fatty degeneration in the supraspinatus muscle after rotator cuff tear. J Shoulder Elbow Surg. 1996;5(3):194-200.
63. Gerber C, Meyer DC, Schneeberger AG, Hoppeler H, von Rechenberg B. Effect of tendon release and delayed repair on the structure of the muscles of the rotator cuff: an experimental study in sheep. J Bone Joint Surg Am. 2004;86(9):1973-82.
64. Goutallier D, Postel JM, Bernageau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. Clin Orthop Relat Res. 1994(304):78-83.
65. Voigt C, Bosse C, Vosshenrich R, Schulz AP, Lill H. Arthroscopic supraspinatus tendon repair with suture-bridging technique: functional outcome and magnetic resonance imaging. Am J Sports Med. 2010;38(5):983-91.
66. Cho NS, Rhee YG. The factors affecting the clinical outcome and integrity of arthroscopically repaired rotator cuff tears of the shoulder. Clin Orthop Surg. 2009;1(2):96-104.
67. Longo UG, Salvatore G, Rizzello G, Berton A, Ciuffreda M, Candela V, et al. The burden of rotator cuff surgery in Italy: a nationwide registry study. Arch Orthop Trauma Surg. 2017;137(2):217-24.
68. Del Buono A, Oliva F, Longo UG, Rodeo SA, Orchard J, Denaro V, et al. Metalloproteases and rotator cuff disease. J Shoulder Elbow Surg. 2012;21(2):200-8.
69. Longo UG, Petrillo S, Rizzello G, Candela V, Denaro V. Deltoid muscle tropism does not influence the outcome of arthroscopic rotator cuff repair. Musculoskelet Surg. 2016;100(3):193-8.
70. Longo UG, Forriol F, Campi S, Maffulli N, Denaro V. Animal models for translational research on shoulder pathologies: from bench to bedside. Sports Med Arthrosc Rev. 2011;19(3):184-93.
71. Longo UG, Franceschi F, Ruzzini L, Rabitti C, Morini S, Maffulli N, et al. Light microscopic histology of supraspinatus tendon ruptures. Knee Surg Sports Traumatol Arthrosc. 2007;15(11):1390-4.
72. Longo UG, Berton A, Papapietro N, Maffulli N, Denaro V. Epidemiology, genetics and biological factors of rotator cuff tears. Med Sport Sci. 2012;57:1-9.
73. Linaker CH, Walker-Bone K. Shoulder disorders and occupation. Best Pract Res Clin Rheumatol. 2015;29(3):405-23.
74. Chaudhury S, Carr AJ. Lessons we can learn from gene expression patterns in rotator cuff tears and tendinopathies. J Shoulder Elbow Surg. 2012;21(2):191-9.
75. Song A, Cannon D, Kim P, Ayers GD, Gao C, Giri A, et al. Risk factors for degenerative, symptomatic rotator cuff tears: a case-control study. J Shoulder Elbow Surg. 2021.
76. Jain NB, Higgins LD, Losina E, Collins J, Blazar PE, Katz JN. Epidemiology of musculoskeletal upper extremity ambulatory surgery in the United States. BMC Musculoskelet Disord. 2014;15:4.
77. Hsu J, Keener JD. Natural History of Rotator Cuff Disease and Implications on Management. Oper Tech Orthop. 2015;25(1):2-9.

78. Narvani AA, Imam MA, Godenèche A, Calvo E, Corbett S, Wallace AL, et al. Degenerative rotator cuff tear, repair or not repair? A review of current evidence. Ann R Coll Surg Engl. 2020;102(4):248-55.
79. Lippi G, Longo UG, Maffulli N. Genetics and sports. Br Med Bull. 2010;93:27-47.
80. Longo UG, Fazio V, Poeta ML, Rabitti C, Franceschi F, Maffulli N, et al. Bilateral consecutive rupture of the quadriceps tendon in a man with BstUI polymorphism of the COL5A1 gene. Knee Surg Sports Traumatol Arthrosc. 2010;18(4):514-8.
81. Longo UG, Lamberti A, Maffulli N, Denaro V. Tendon augmentation grafts: a systematic review. Br Med Bull. 2010;94:165-88.
82. Longo UG, Lamberti A, Maffulli N, Denaro V. Tissue engineered biological augmentation for tendon healing: a systematic review. Br Med Bull. 2011;98:31-59.
83. Magra M, Maffulli N. Genetic aspects of tendinopathy. J Sci Med Sport. 2008;11(3):243-7.
84. September AV, Schwellnus MP, Collins M. Tendon and ligament injuries: the genetic component. Br J Sports Med. 2007;41(4):241-6; discussion 6.
85. Harvie P, Ostlere SJ, Teh J, McNally EG, Clipsham K, Burston BJ, et al. Genetic influences in the aetiology of tears of the rotator cuff. Sibling risk of a full-thickness tear. J Bone Joint Surg Br. 2004;86(5):696-700.
86. Gwilym SE, Watkins B, Cooper CD, Harvie P, Auplish S, Pollard TC, et al. Genetic influences in the progression of tears of the rotator cuff. J Bone Joint Surg Br. 2009;91(7):915-7.
87. Tashjian RZ, Hollins AM, Kim HM, Teefey SA, Middleton WD, Steger-May K, et al. Factors affecting healing rates after arthroscopic double-row rotator cuff repair. Am J Sports Med. 2010;38(12):2435-42.
88. Tashjian RZ, Farnham JM, Albright FS, Teerlink CC, Cannon-Albright LA. Evidence for an inherited predisposition contributing to the risk for rotator cuff disease. J Bone Joint Surg Am. 2009;91(5):1136-42.
89. Foulkes T, Wood JN. Pain genes. PLoS Genet. 2008;4(7):e1000086.
90. Buskila D. Genetics of chronic pain states. Best Pract Res Clin Rheumatol. 2007;21(3):535-47.
91. Longo UG, Berton A, Khan WS, Maffulli N, Denaro V. Histopathology of rotator cuff tears. Sports Med Arthrosc Rev. 2011;19(3):227-36.
92. Denaro V, Ruzzini L, Longo UG, Franceschi F, De Paola B, Cittadini A, et al. Effect of dihydrotestosterone on cultured human tenocytes from intact supraspinatus tendon. Knee Surg Sports Traumatol Arthrosc. 2010;18(7):971-6.
93. Denaro V, Ruzzini L, Barnaba SA, Longo UG, Campi S, Maffulli N, et al. Effect of pulsed electromagnetic fields on human tenocyte cultures from supraspinatus and quadriceps tendons. Am J Phys Med Rehabil. 2011;90(2):119-27.
94. Longo UG, Oliva F, Olivia F, Denaro V, Maffulli N. Oxygen species and overuse tendinopathy in athletes. Disabil Rehabil. 2008;30(20-22):1563-71.
95. Longo UG, Franceschi F, Ruzzini L, Rabitti C, Morini S, Maffulli N, et al. Characteristics at haematoxylin and eosin staining of ruptures of the long head of the biceps tendon. Br J Sports Med. 2009;43(8):603-7.
96. Maffulli N, Franceschi F, Longo UG, Ruzzini L, Denaro V. Clinical evidence for suture anchor repair of rotator cuff tears does add up: some just do not want to see it. Arthroscopy. 2010;26(12):1568-9; author reply 9-70.
97. Maffulli N, Longo UG, Berton A, Loppini M, Denaro V. Biological factors in the pathogenesis of rotator cuff tears. Sports Med Arthrosc Rev. 2011;19(3):194-201.
98. Maffulli N, Longo UG, Denaro V. Novel approaches for the management of tendinopathy. J Bone Joint Surg Am. 2010;92(15):2604-13.
99. Longo UG, Ronga M, Maffulli N. Acute ruptures of the achilles tendon. Sports Med Arthrosc Rev. 2009;17(2):127-38.

100. Jancuska J, Matthews J, Miller T, Kluczynski MA, Bisson LJ. A Systematic Summary of Systematic Reviews on the Topic of the Rotator Cuff. Orthop J Sports Med. 2018;6(9):2325967118797891.
101. Dean BJ, Franklin SL, Carr AJ. A systematic review of the histological and molecular changes in rotator cuff disease. Bone Joint Res. 2012;1(7):158-66.
102. Sejersen MH, Frost P, Hansen TB, Deutch SR, Svendsen SW. Proteomics perspectives in rotator cuff research: a systematic review of gene expression and protein composition in human tendinopathy. PLoS One. 2015;10(4):e0119974.
103. Dabija DI, Gao C, Edwards TL, Kuhn JE, Jain NB. Genetic and familial predisposition to rotator cuff disease: a systematic review. J Shoulder Elbow Surg. 2017;26(6):1103-12.
104. Longo UG, Candela V, Berton A, Salvatore G, Guarnieri A, DeAngelis J, et al. Genetic basis of rotator cuff injury: a systematic review. BMC Med Genet. 2019;20(1):149.
105. Motta GaR, Amaral MV, Rezende E, Pitta R, Vieira TC, Duarte ME, et al. Evidence of genetic variations associated with rotator cuff disease. J Shoulder Elbow Surg. 2014;23(2):227-35.
106. Teerlink CC, Cannon-Albright LA, Tashjian RZ. Significant association of full-thickness rotator cuff tears and estrogen-related receptor-β (ESRRB). J Shoulder Elbow Surg. 2015;24(2):e31-5.
107. Tashjian RZ, Granger EK, Farnham JM, Cannon-Albright LA, Teerlink CC. Erratum to "Genome-wide association study for rotator cuff tears identifies two significant single-nucleotide polymorphisms" [J Shoulder Elbow Surg 2016;25:174-179]. J Shoulder Elbow Surg. 2016;25(10):1731.
108. Tanida T, Matsuda KI, Yamada S, Hashimoto T, Kawata M. Estrogen-related Receptor β Reduces the Subnuclear Mobility of Estrogen Receptor α and Suppresses Estrogen-dependent Cellular Function. J Biol Chem. 2015;290(19):12332-45.
109. Torricelli P, Veronesi F, Pagani S, Maffulli N, Masiero S, Frizziero A, et al. In vitro tenocyte metabolism in aging and oestrogen deficiency. Age (Dordr). 2013;35(6):2125-36.
110. Rimkus C, Martini M, Friederichs J, Rosenberg R, Doll D, Siewert JR, et al. Prognostic significance of downregulated expression of the candidate tumour suppressor gene SASH1 in colon cancer. Br J Cancer. 2006;95(10):1419-23.
111. Zeller C, Hinzmann B, Seitz S, Prokoph H, Burkhard-Goettges E, Fischer J, et al. SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is downregulated in breast cancer. Oncogene. 2003;22(19):2972-83.
112. Assunção JH, Godoy-Santos AL, Dos Santos MCLG, Malavolta EA, Gracitelli MEC, Ferreira Neto AA. Matrix Metalloproteases 1 and 3 Promoter Gene Polymorphism Is Associated With Rotator Cuff Tear. Clin Orthop Relat Res. 2017;475(7):1904-10.
113. Kluger R, Burgstaller J, Vogl C, Brem G, Skultety M, Mueller S. Candidate gene approach identifies six SNPs in tenascin-C (TNC) associated with degenerative rotator cuff tears. J Orthop Res. 2017;35(4):894-901.
114. Longo UG, Margiotti K, Petrillo S, Rizzello G, Fusilli C, Maffulli N, et al. Genetics of rotator cuff tears: no association of col5a1 gene in a case-control study. BMC Med Genet. 2018;19(1):217.
115. Salles JI, Lopes LR, Duarte MEL, Morrissey D, Martins MB, Machado DE, et al. Fc receptor-like 3 (-169T>C) polymorphism increases the risk of tendinopathy in volleyball athletes: a case control study. BMC Med Genet. 2018;19(1):119.
116. Tashjian RZ, Granger EK, Farnham JM, Cannon-Albright LA, Teerlink CC. Genome-wide association study for rotator cuff tears identifies two significant single-nucleotide polymorphisms. J Shoulder Elbow Surg. 2016;25(2):174-9.
117. Roos TR, Roos AK, Avins AL, Ahmed MA, Kleimeyer JP, Fredericson M, et al. Genome-wide association study identifies a locus associated with rotator cuff injury. PLoS One. 2017;12(12):e0189317.

118. Roberti M, Mustich A, Gadaleta MN, Cantatore P. Identification of two homologous mitochondrial DNA sequences, which bind strongly and specifically to a mitochondrial protein of Paracentrotus lividus. Nucleic Acids Res. 1991;19(22):6249-54.
119. Tashjian RZ, Saltzman EG, Granger EK, Hung M. Incidence of familial tendon dysfunction in patients with full-thickness rotator cuff tears. Open Access J Sports Med. 2014;5:137-41.
120. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. 1999;9(8):677-9.
121. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
122. Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, et al. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. Hum Mutat. 2010;31(8):918-23.
123. Biesecker LG, Shianna KV, Mullikin JC. Exome sequencing: the expert view. Genome Biol. 2011;12(9):128.
124. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods. 2007;4(11):903-5.
125. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet. 2007;39(12):1522-7.
126. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182-9.
127. Lacey S, Chung JY, Lin H. A comparison of whole genome sequencing with exome sequencing for family-based association studies. BMC Proc. 2014;8(Suppl 1 Genetic Analysis Workshop 18Vanessa Olmo):S38.
128. Banyard J, Bao L, Zetter BR. Type XXIII collagen, a new transmembrane collagen identified in metastatic tumor cells. J Biol Chem. 2003;278(23):20989-94.
129. Bergé-Lefranc JL, Jay P, Massacrier A, Cau P, Mattei MG, Bauer S, et al. Characterization of the human jumonji gene. Hum Mol Genet. 1996;5(10):1637-41.
130. Toyoda M, Kojima M, Takeuchi T. Jumonji is a nuclear protein that participates in the negative regulation of cell growth. Biochem Biophys Res Commun. 2000;274(2):332-6.
131. Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature. 2010;464(7286):306-10.
132. Toyoda M, Shirato H, Nakajima K, Kojima M, Takahashi M, Kubota M, et al. jumonji downregulates cardiac cell proliferation by repressing cyclin D1 expression. Dev Cell. 2003;5(1):85-97.
133. Jung J, Mysliwiec MR, Lee Y. Roles of JUMONJI in mouse embryonic development. Dev Dyn. 2005;232(1):21-32.
134. Minter-Dykhouse K, Ward I, Huen MS, Chen J, Lou Z. Distinct versus overlapping functions of MDC1 and 53BP1 in DNA damage response and tumorigenesis. J Cell Biol. 2008;181(5):727-35.
135. Bartkova J, Horejsí Z, Sehested M, Nesland JM, Rajpert-De Meyts E, Skakkebaek NE, et al. DNA damage response mediators MDC1 and 53BP1: constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours. Oncogene. 2007;26(53):7414-22.
136. Jungmichel S, Stucki M. MDC1: The art of keeping things in focus. Chromosoma. 2010;119(4):337-49.
137. Ruff SE, Logan SK, Garabedian MJ, Huang TT. Roles for MDC1 in cancer development and treatment. DNA Repair (Amst). 2020;95:102948.
138. Kao HY, Lee CH, Komarov A, Han CC, Evans RM. Isolation and characterization of mammalian HDAC10, a novel histone deacetylase. J Biol Chem. 2002;277(1):187-93.

139. Cheng F, Zheng B, Wang J, Zhao G, Yao Z, Niu Z, et al. Histone deacetylase 10, a potential epigenetic target for therapy. Biosci Rep. 2021;41(6).