ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DISEASE GENE IDENTIFICATION USING LINKAGE AND EXOME ANALYSIS

M.Sc. THESIS

Derya YAVUZ

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DISEASE GENE IDENTIFICATION USING LINKAGE AND EXOME ANALYSIS

M.Sc. THESIS

Derya YAVUZ
(521171119)

Department of Molecular Biology-Genetics and Biotechnology

Molecular Biology-Genetics and Biotechnology Programme

Thesis Advisor: Prof. Dr. Eda TAHİR TURANLI

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

BAĞLANTI VE EKZOM ANALİZİ KULLANARAK HASTALIK GENİ KEŞFİ

YÜKSEK LİSANS TEZİ

Derya YAVUZ
(521171119)

Moleküler Biyoloji-Genetik ve Biyoteknoloji Anabilim Dalı

Moleküler Biyoloji-Genetik ve Biyoteknoloji Programı

Tez Danışmanı: Prof. Dr. Eda TAHİR TURANLI

TEMMUZ 2020

Derya Yavuz, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 521171119 successfully defended the thesis/dissertation entitled "Disease Gene Identification Using Linkage and Exome Analyses", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Eda TAHİR TURANLI** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Arzu KARABAY KORKMAZ** Istanbul Technical University	.....
	Prof. Dr. Sibel Aylin UĞUR İŞERİ Istanbul University	.....

Date of Submission : 15 June 2020
Date of Defense : 16 July 2020

FOREWORD

Firstly, I would like to express my deepest gratitude and thanks to my advisor Prof. Dr. Eda TAHİR TURANLI. I appreciate her constant support and insightful guidance which allowed me to become a more competent researcher. Being in the Human Genetics Lab and to be able to observe and learn from a scientist with such profound knowledge was an inestimable and valued opportunity. I am thankful to her inspiring approach and moral support which enabled me to complete my MSc degree.

Secondly, I would like to thank Dr. İlker KARACAN for sharing his knowledge and his guidance and support throughout this project. I am really grateful to both Elif EVEREST and Merve ÖZKILINÇ ÖNEN not only for sharing their knowledge and moral support but also for being reliable and understanding friends.

I am really grateful to my dear friends Selen ÖZKAN and Doğa GÜNDEM who were always there in time of need and never hesitated to help in every imaginable way. To Mehmet Seyit ZOR, I extend my thanks for his efforts and encouragements. I would also like to thank other lab members, Ayşe BALAMİR and especially Cansın KIRMAÇOĞLU with whom I share many delightful memories.

Finally, I would like to thank my parents Leyla and Saim YAVUZ without whom I would not the person who I am.

June 2020 Derya YAVUZ

TABLE OF CONTENTS

Page

ABBREVIATIONS

CRMO : Chronic Recurrent Multifocal Osteomyelitis
DIRA : Deficiency of Interleukin-1 Receptor Antagonist
CNO : Chronic Non-bacterial Osteomyelitis
TA : Takayasu Arteritis
HLA : Human Leukocyte Antigen
kb : kilobases
LOD : Logarithm of Odds
C : Cytosine
T : Thymine
A : Adenine
G : Guanine
SNP : Single Nucleotide Polymorphism
Sibs : Siblings
MAF : Minor Allele Frequency
IL10 : Interleukin-10
TLR4 : Toll-Like Receptor 10
MAPK : MAP Kinase
GCA : Giant Cell Artertis
MRI : Magnetic Resonance Imaging
mmHg : Milimeter Hg
NIH : National Institute of Health
MZT : Monozygotic Twin
Chr : Chromosome
LD : Linkage Disequilibrium
CRP : C-reactive Protein
CD : Crohn's Disease
kDa : kilo Dalton
IFN-γ : Interferon gamma
TNF-α : Tumor Necrosis Factor alpha
DNA : Deoxyribonucleic acid
RNA : Ribonucleic acid
HGP : Human Genome Project
NGS : Next Generation Sequencing
WGS : Whole Genome Sequencing
HTS : High-throughput Sequencing
WES : Whole Exome Sequencing
PCR : Polymerase Chain Reaction
GWAS : Genome-Wide Association Studies
WGBS : Whole Genome Bisulfite Sequencing
cDNA : Complimentary deoxyribonucleic acid
BWA : Burrows-Wheeler Algorithm
GATK : Genome Analysis ToolKit
VCF : Variant Call Format
SIFT : The Sorting Intolerant from Tolerant

PolyPhen : Polymorphism Pehnotyping
1000G : 1000 Genome Project
GnomAD : The Genome Aggregation Database
CADD : Combined Annotation Dependent Depletion
REVEL : Rare Exome Variant Ensemble Learner
IBD : Identical by Descent
cM : centiMorgan
NPL : Non-parametric Linkage
TE : Tris-EDTA
YCGA : Yale Center for Genome Analysis
LOH : Loss of Heterozygosity
MGI : Mouse Genome Informatics

LIST OF TABLES

Page

LIST OF FIGURES

Page

DISEASE GENE IDENTIFICATION USING LINKAGE AND EXOME ANALYSES

SUMMARY

Rare disorders are described as conditions with prevalence less than 1 in 2000 people. Approximately 300 million people are affected with rare disease worldwide and in Turkey this number is estimated as 5 million. More than 6000 rare diseases are identified to date and 80% of these are monogenic disorders. Disease onset is usually during childhood with chronic, progressive and debilitating nature. Patients life expectancy is lower than unaffected population. Extensive clinical examination is required for definite diagnosis. For these reasons it is crucial to reveal underlying mechanisms of rare disorders.

Chronic recurrent multifocal osteomyelitis (CRMO) is a rare, chronic autoinflammatory disorder that affects children between age of 2-17. It is characterized by aseptic, symmetric and recurrent bone lesions that tend to cluster around metaphysis of long bones such as femur, tibia, vertebral column and pelvic floor. Although there is no definitive data on prevalence of CRMO, it is estimated to be 1-2 cases per million. Though having an unclarified etiology and pathogenesis, families with multiple affected members suggest genetic background on disease development. Identification of disease-causing genes - *LPIN2* and *IL1RA* in Majeed syndrome and deficiency of interleukin-1 receptor antagonist (DIRA) which are defined as chronic nonbacterial osteomyelitis (CNO) further strengthen the involvement of genetic features.

Takayasu arteritis (TA) also known as pulseless disease is a rare, systemic large vessel vasculitis that predominantly affects women younger than 40 years old. Its major clinical symptoms are occlusion and stenosis of aorta and its main branches accompanied with loss of radial pulse whereas involvement of pulmonary, coronary, renal arteries and aneurysmal formations are also encountered. Severity of the clinical progression depends both on the vessels afflicted and ethnicity of the patient. Highest prevalence of TA is reported in East Asia, especially Japan with up to 40 cases per million. In Turkey, it is postulated to be the second mostly encountered vasculitis with a prevalence up to 12.8/million accompanied with 1.11/million incidence.

In spite of undefined etiology, several factors including infections are suggested to be involved in TA pathogenesis. Monozygotic twins diagnosed with TA and familial aggregation observed in some cases indicate the role of genetic factors in disease phenotype. Various genome-wide association studies (GWAS) in cohorts of different size and ethnicity replicated significant association of *HLA-B*52* and *IL12B* to TA. Susceptibility loci *FCGR2A/FCGR3A*, *MLX, HLA-B/MICA* and *HLA-DQB1/HLA-DRB1* also show significant association with TA.

In this study we present three Turkish consanguineous families with multiple members diagnosed with TA and one consanguineous family with two brothers diagnosed with CRMO. Clinical diagnosis and follow up of TA families were done at Cerrahpaşa Medical Faculty whereas CRMO patients were diagnosed and followed in Pamukkale University Medical Faculty. Individuals' were informed and their written consent for

participating in the study was obtained accordingly to the ethics committee approval (MBG.22/2014). The aim of the study is to identify disease causing genes by using gene mapping techniques as linkage and exome analyses. Linkage analysis and homozygosity mapping were used to identify non-recombinant homozygous genomic regions that were shared between affected individuals but not with healthy members. By exome analysis, possible candidate genes were investigated.

Linkage analysis and homozygosity mapping of CRMO1 family revealed 54 regions that are larger than 200 kb with LOD score > 1.50 that were shared among the affected siblings. Exome analysis on these regions did not reveal any homozygous, non-synonymous and rare variants. Therefore, exome analysis was conducted without basis on linkage regions revealing 10 variants out of which only *CARNS*:c.1145C>T had a closer association with CRMO phenotype when compared to the rest. However, the region it lied was homozygous in all family members eliminating the variant from being a candidate.

Two sisters diagnosed with TA are present in all three families under investigation. TA1 and TA2 families indicate autosomal recessive inheritance model. Affected sisters of TA3 family are half siblings (sibs) who share the same mother but different fathers. CRMO1 family two brothers were diagnosed with CRMO. Genomic DNA samples obtained from peripheral blood were used for genotyping via genome-wide single nucleotide polymorphism (SNP) array where ~710,00 fixed markers were genotyped using Illumina Infinium OmniExpress microarray. By performing multipoint parametric linkage analyses under recessive inheritance model and homozygosity mapping using the SNP data, non-recombinant regions that were shared among the affected individuals but not with the healthy members were detected. Exome sequencing data were used to fine map these regions to variant resolution and investigate the presence of a potential candidate variants involved in the TA or CRMO phenotype. Both linkage and exome analyses were conducted under the recessive inheritance model. Variants in the functional regions (exonic and splicing) with minor allele frequency (MAF) < 0.01 were further examined with bioinformatics tools. Expression sites and levels, pathways, interactions, evolutionary conservation of the genes harboring candidate variants were studied to unravel any potential involvement in the immune system or inflammatory response that could lead to the TA or CRMO phenotype.

Linkage analysis of TA1 family revealed 59 regions larger than 200 kilobases (kb) with maximal LOD score ≥ 1.50. Homozygosity mapping confirmed 41 of these to be shared between affected individuals. Co-joint multipoint parametric linkage analysis of TA1 and TA2 families indicated 123 linkage regions larger than 200 kb with LOD score ≥ 1.50. Nine of these were shared between all affected members of both families. Exome analysis of these sites revealed 9 variants in total of which 5 were in TA1-04 which are *RELN*: c.103301976_103301984dup, *ANXA8L1*: c.449C>T, *EHBP1L1:* c.1321C>T, *MYH14*: c.565C>T and *SHANK1:* c.3947G>A. The remaining 4 variants in TA2-01 are *AP4B1*: c.767C>T, *MST1L*: c.811dupG, *AGAP:* c.G8A and *ZNF829:* c.1173A>T. Variant filtration and prioritization of the exomes of TA3 family members, revealed no common variant that are found in affected sibs. There was no gene that got hits that were present in both TA patients.

Although our results were not indicative of a gene involved in the TA development, this is the first study integrating linkage analysis, homozygosity mapping and exome analysis where three consanguineous families with multiple affected members were included to identify possible disease-causing variants. Variants detected in TA patients should be further studied to unravel their possible roles in TA pathogenesis.

BAĞLANTI VE EKZOM ANALİZİ KULLANILARAK HASTALIK GENİ KEŞFİ

ÖZET

Nadir hastalıklar toplumlarda 2000 kişiden birinde görünen hastlıklardır. Dünyada yaklaşık 300 milyon insanın etkilendiği tahmin edilirken Türkiye'de bu sayı 5 milyon olarak öngörülmektedir. 6000'den fazla nadir hastalık tarif edilmişken, bunların %80'I monogenik hastalıklardır. Genellikle çocukluk çağında ortaya çıkan bu hastalıkların doğası kronik, progresif ve sakat bırakıcı olmasıdır. Hastaların beklenen yaşam süresi genel popülasyona göre daha düşüktür. Kesin tanı için çok detaylı ve derinlemesine klinik incelemeler gerekmektedir. Bu sebeplerden ötürü nadir hastalıkların moleküler tabanını ortaya çıkarmak tanı ve tedavisi için çok önemlidir.

Kronik rekürren multifokal osteomiyelit (KRMO) genellikle 2 ile 17 yaş arasındaki çocukları etkileyen enfeksiyöz olmayan nadir ve kronik bir otoinflamatuvar hastalıktır. Aspetik, simetrik şekilde görülen ve tekrarlayan, femur, tibia, omurga gibi kemiklerde lezyonlarının görülmesi ile karakterize edilmektedir. KRMO prevelansı hakkında kesin bir veri olmamakla birlikte, dünya genelinde prevelans 1-2/1000000 olarak tahmin edilmektedir. Aydınlatılmış bir etiyoloji ve patogenezi olmamasına rağmen, birden fazla hastanın bulunduğu aileler hastalığın ortaya çıkmasında genetik faktörlerin etkili olabileceğini düşündürtmektedir. Kronik bakteriyel olmayan osteomiyelit (CNO) olarak sınıflandırılan Majeed sendromu ve interlökin-1 reseptör antagonisti eksikliği (DIRA) hastalıklarının ortaya çıkmasına *LPIN2* ve *IL1RA* genlerindeki mutasyonların sebep olduğunun gösterilmiştir. Literatürde KRMO hastalığına sahip monozigot ikizler raporlanmıştır. Ayrıca, babası başka bir CNO hastalığına sahip KRMO hastası bir çocuk da raporlanmıştır. KRMO patogenizinde genetik bir etmen olduğunu düşündürten bir diğer faktör ise hayvan modelleridir. *pstpip2* geninde kendiliğinden gelişen mutasyonların farelerde KRMO fenotipine yol açtığı ve bunun sonraki jenerasyonlara otozomal resesif bir şekilde aktarıldığı gözlemlenmiştir. Bahsedilen kanıtlar birlikte ele alındığında, KRMO patogenizinde genetik faktörlerin etkisi olduğu hipotezi güçlenmektedir.

Takayasu arteriti (TA) genellikle aort ve aorttan dallanan damarların tutulumu ile karakterize edilen ve nabızsızlık hastalığı olarak da bilinen sistemik bir büyük damar vaskülitidir. Klinik bulguları arteriyel oklüzyon ve stenöz, ekstremitelerde nabız yokluğu ve inflamasyona bağlı anevrizma oluşumudur. Pulmoner, renal veya kardiyak arterlerde de tutulum gözlenebilmektedir. Nadir bir hastalık olan TA çoğunlukla 40 yaş altı kadınlarda rastlanmaktadır fakat prognoz hem tutulan damar tipi hem de etnisite ile ilişkilidir. Hastalığın en çok göründüğü ülkeler Doğu Asya'dadır. En yüksek prevalans 40/milyon ile Japonya'da. Türkiye'de Behçet hastalığından sonra en sık görülen vaskülit olduğu belirtilen TA'nın, Türkiye için prevalansı 12.8/milyon insidansı ise 1.11/milyon olarak gözlenmiştir.

Etiyolojisi kesin olmamakla birlikte TA patogenizinde enfeksiyon dahil çeşitli faktörlerden şüphelenilmektedir. TA tanısı konmuş tek yumurta ikizleri ve bazı olgularda ailesel agregasyon görülmesi, genetik faktörlerin hastalığın ortaya

çıkmasında rol oynayabileceğini önermektedir. Çeşitli etnisite ve büyüklüklerdeki kohortlarda yapılan tüm genom bağlantı analizleri (GWAS), *HLA-B52* ve *IL12B* geni TA ile anlamlı şekilde ilişkilendirilmiştir. Türkiye ve Amerika kohortlarında yapılan GWAS sayesinde *HLA-B52* ilişkisinin gösterilmesinin yanı sıra *FCGR2A/FCGR3A, HLA-B/MICA* ve *HLA-DQB1/HLA-DRB1* lokuslarının TA yatkınlığı oluşturabileceği gösterilmiştir. Japonya kohortunda ise *MLX* genindeki bir varyant risk faktörü olarak gösterilmiş ve hastalığın prognozu ile anlamlı derecede ilintili bulunmuştur. Bu çalışmada birden fazla TA ve KRMO tanılı hastanın olduğu akrabalık evliliği olan 4 Türk ailesi çalışılmıştır. Hastaların TA tanısı Cerrahpaşa Tıp Fakültesi'nde, KRMO tanısı ise Pamukkale Üniversitesi Tıp Fakültesi'nde konulmuştur. Çalışmaya katılacak bireyler bilgilendirilmiş ve yazılı onayları İstanbul Teknik Üniversitesi Etik Kurulu izni doğrultusunda alınmıştır. Bu çalışmanın amacı Takayasu arteriti ve KRMO patogenizinde rol oynayan olası gen veya genlerin bağlantı ve ekzom analizi yöntemleri kullanılarak ortaya çıkarılmasıdır. Genom boyu bağlantı analizi ve homozigotluk haritalandırması kullanılarak hastalar arasında paylaşılan ama sağlıklı bireylerde görülmeyen ortak genomik bölgeler saptanmıştır. Takibinde ise genom-boyu ekzom sekanslama analizi ile aday gen/genler belirlenmesi amaçlanmıştır.

Çalışılan üç ailede 2 kız kardeş TA tanısı almıştır. TA1 ve TA2 aile ağacı hastalığın otozomal çekinik kalıtım gösterdiğini önermektedir. TA3 ailesinde is hasta kız kardeşler ortak anneye sahiplerken babaları farklıdır. KRMO ailesinde ise 2 erkek kardeş tanı almışıtr. Bireylerden alınan kan numunelerinden izole edilen genomik DNA kullanılarak ~710000 marker barındıran Ilumina Infinium OmniExpress mikrodizini ile tek nükletoid polimorifzmleri (SNP) genotiplendirmesi yapılmıştır. Çekinik modelde çok noktalı parametrik bağlantı analizi ve homozigotluk haritalandırması yapılarak, hasta bireylerde ortak homozigot bölgeler tespit edilmiştir. Ekzom datası ise bu bölgelerde aday gen/genler taraması için kullanılmıştır. Yaklaşık 140,000 varyant içinden homozigot, nadir ve fonksiyonel (ekzonik ve splicing) olanlar seçilmiş ve belirlenen bölgelerdeki varyantlar incelenmiştir. Ekzom analizi, bağlantı analizi ve homozigotluk haritalandırması sonucundan bağımsız olarak da yapılmıştır. Belirlenen bölgelere bağlı kalmaksızın bahsedilen kriterler kullanılarak bütün varyantlar arasından aday gen tespiti yapılmıştır. Her iki yöntemden çıkarılan ilk adaylar *in silico* tahmin araçları ile incelenmiş ve gen ifadesi yeri ve seviyesi, yer aldığı yolaklar, etkileşim ortakları, evrimsel korunurluğu gibi faktörler çeşitli veri tabanlarında araştırılmıştır.

TA1 ailesinin bağlantı analizi sonucu 200 kilobazdan (kb) büyük maksimum LOD skor değeri ≥ 1.50 olan 59 bölge belirlenmişti. Homozigotluk haritalandırması bu bölgelerden 41 tanesinin TA hastalarında ortak olduğunu göstermiştir. TA1 ve TA2 ailesinde birlikte gerçekleştirilen bağlantı analizi ise 200 kb'dan büyük LOD skoru ≥ 1.50 olan 123 bölgeye işaret etmiştir. Homozigotluk hairtalandırması ile bu bölgelerden 9 tanesinin TA1 ve TA2 ailesindeki hastalarda ortak paylaşıldığını göstermiştir. Bağlantı analizi ve homozigotluk haritalandırması takibinde yapılan ekzom analizi yapılmıştır. TA1 ailesindeki hastada *RELN*: c.103301976_103301984dup, *ANXA8L1*: c.449C>T, *EHBP1L1:* c.1321C>T, *MYH14*: c.565C>T ve *SHANK1:* c.3947G>A olmak üzere 5, TA2 ailesindeki hastada ise *AP4B1*: c.767C>T, *MST1L*: c.811dupG, *AGAP:* c.G8A ve *ZNF829:* c.1173A>T olmak üzere 4 varyant belirlenmiştir. TA3 ailesinde 2 hasta kız kardeşin ekzom datası ortak varyant veya ortak bir gende varyant varlığı için araştırılmış fakat paylaşılan bir varyant bulunamamıştır. 2 hasta kız kardeş arasındaki ortak varyantlar, sağlıklı anne ile de paylaşıldığından aday olarak seçilmemiş ve sonraki aşamalarda analize dahil edilmemiştir.

CRMO1 ailesinde gerçekleştirilen bağlantı analizi ve homozigotluk haritalandırması 200 kb'dan büyük, LOD skoru > 1.50 olan ve sadece hasta kardeşlerde paylaşılan 54 bölge belirtmiştir. Ekzom datasında bu bölgelerdeki varyantların incelenmesi sonucunda homozigot, nadir varyant bulunamamıştır. Bu yüzden ekzom analizi bu bölgelere bağlı kalmadan yapılmış ve homozigot, nadir ve eş anlamlı olan varyantların ortaya çıkarılması amacıyla gerçekleştirilmiştir. Elde edilen 10 varyanttan, sadece *CARNS*:c.1145C>T KRMO ile ilişkilendirilebilecek bir fonksiyona sahiptir; fakat varyantın bulunduğu bölge SNP datasında incelendiğinde ailedeki herkeste ortak olan bir bölgede olduğu görülmüştür.

Elde edilen sonuçlar, TA veya CRMO hastalarında görülen ortak bir gene işaret etmese de TA ve CRMO patogenizi ile ilgili literatürde gerçekleştirilen ilk genom boyu bağlantı analizi, homozigotluk haritalandırması ve ekzom analizi kullanan ilk çalışma olması açısından önemlidir. TA ve CRMO hastalarında tespit edilen varyantların hastalık patolojisindeki muhtemel rollerinin daha detaylı araştırılması gerekmektedir.

1. INTRODUCTION

1.1 Rare Diseases

Rare diseases (RD) also known as orphan diseases affect a small number of people in the general population. Even though there is no global definition, European Union (EU) Regulations established a rare disease as a condition with prevalence less than 1 in 2000 people. On the other hand, according to the Orphan Drug Act a disease is called an RD in United States when it affects less than 200000 people. Affecting 3.5-5.9% of the population, it is estimated that there are 300 million people worldwide affected with an RD. Number of identified rare diseases to date is between 6000-7000. 72% of these diseases have a genetic basis and thus show a chronic, progressive and degenerative nature. 70% of those start in childhood and lead to an early death [1]. Therefore, it is imperative to inspect their genetic, molecular and biochemical bases for discovery of treatment options. Small number of cases, clinical heterogeneity, reduced reproductive fitness and under- or misdiagnosis elicit further challenges in deducing the true global prevalence of RD. Also, prevalence of a rare disease may differ depending on the geography, population genetic diversity and environmental pressure [2].

With the rise of high-throughput sequencing technologies and reduction of their cost, gene variant discovery of RD has accelerated significantly. However, functional studies lag behind which limit our knowledge on molecular and biochemical bases of RD. A big proportion of rare disorders are monogenic with a Mendelian inheritance pattern. As of July 21st, 2020 Online Mendelian Inheritance of Men (OMIM) enlists 5460 disorders with known molecular basis along with 1419 phenotypes with unknown molecular basis. Therefore, genomic studies are crucial to enlighten the molecular basis of the remaining phenotypes and identify associated variants or genes. Furthermore, the presence of consanguinity, tracing homozygous mutations can aid in identification of disease-causing genes in rare monogenic and especially recessive diseases [3].

1.2 Chronic Recurrent Multifocal Osteomyelitis

Chronic recurrent multifocal osteomyelitis (CRMO) is a chronic autoinflammatory disorder that is characterized by subacute, chronic osteomyelitis. It is a bone disorder identified with culture negative (aseptic), multifocal bone lesions generally emerging symmetrically. Although chronic non-bacterial osteomyelitis (CNO) is used interchangeably, it also refers to Majeed syndrome and deficiency of interleukin-1 receptor antagonist (DIRA) [4] Majeed syndrome is characterized by onset of CRMO generally before 2 years of age and shows skin rashes along with joint contractures [5]. Clinical symptoms include episodes of pain and tenderness that usually worsens at night which can be accompanied by heat, redness and swelling. Due to autoinflammatory nature of the disorder systemic symptoms such as fever, fatigue, malaise and weight loss are also observed. In some cases, skin manifestations such as psoriasis and acne are reported [6]. Due to occurrence of pathologic fractures in severe cases, CRMO may lead to permanent deformities. Lesions commonly occur around metaphysis and can be distributed through both appendicular and axial skeletons. Most common sites that harbor lesions are that of long bones such as femur, tibia, ankle, humerus, vertebral column, clavicle and pelvis.

The first clinical description of CRMO in literature is by Giedion et al in 1972 [7] where they reported four children with sub-acute, chronic and multifocal osteomyelitis. However, the term CRMO was put forward by Probst [8]. It has been shown that CRMO patients with more lesions other autoinflammatory diseases such as inflammatory bowel disorder, arthritis, psoriasis and vasculitis are commonly seen [9]. It is fairly difficult to get the differential diagnosis for CRMO as there are no CRMO specific biomarkers or associated causative gene.

1.2.1 Epidemiology and pathogenesis of CRMO

CRMO is an autoinflammatory disease that affects children aged between 2-17 but generally onset is around 10 years of age. Females are predominantly affected [10]. Even though there are approximately 400 reported cases, absolute prevalence and incidence ratios are not available probably due to rarity and underdiagnosis [11]. Estimated prevalence around the world is 1-2 cases per million but most of the cases are reported from Australia, North America and Europe, including Scandinavia [12,13].

CRMO is a disorder with its etiology and pathophysiology yet to be elucidated. Imbalance of pro- and anti-inflammatory cytokines that are the key components of the innate immune response is the cause of chronic inflammation. Yet, the pathway involved in outburst of this imbalance and disruption of immune homeostasis is not definitely known. Evidence show impaired IL-10 secretion in CRMO patients' peripheral blood monocytes upon TLR4 activation. This disruption was shown to originate from abnormalities in MAPK signaling pathway which lead to decreased levels of transcription factor Sp-1 that regulates IL-10 expression in monocytes. Its lowered recruitment to Sp-1 binding sequence in the *IL10* causes diminished expression of IL10 altering the cytokine balance which result in abnormal innate immune response [14].

1.2.2 Genetics of CRMO

Although CRMO is a very rare disease that has no clear etiology, there are several indications of involvement of genetic features in disease pathogenesis. First of all, both Majeed syndrome and DIRA which are also categorized as CNO disorders are caused by mutations in *LPIN2* and *IL1RA* genes, respectively. Major resemblances in clinical symptoms and pathogenesis suggest a genetic variation in CRMO as well [15]. Secondly, there are two mouse models carrying mutations in *Pstpip2* gene which show the similar features of a CRMO patient. In double mutant *cmo* mouse which is both pstpip2- and Il1β- deficient, aberrant osteomyelitic phenotype was rescued [16,17]. *FBLIM1* is another gene that was suspected to be involved in pathogenesis due to detection of one homozygous missense variant and higher frequency observed for intronic variant rs 114077715 [18]. However, these variants were not replicated in another study [19]. Lastly, reported families with multiple affected individuals and presence of first- and second-degree relatives diagnosed with other autoinflammatory disorders such as inflammatory bowel disease and psoriasis further strengthen the genetic predisposition lying in CRMO pathogenesis.

1.3 Takayasu Arteritis

Takayasu's Arteritis (TA) is a chronic inflammatory disease that mainly affects large blood vessels especially aorta, pulmonary arteries and their branches. It is classified as an idiopathic, granulomatous large vessel vasculitis along with giant cell arteritis (GCA) which predominantly affect women between ages 15-30 [20]. Inflammation

leads to stenotic lesions, thrombus formations and in severe cases destruction of arterial wall that results in aneurysms or rupture of the artery. Carotid, vertebral and renal arteries may also be affected. Blockage in subclavian or brachial arteries leads to a difference in systolic pressure and loss of pulse in upper extremities for which TA is also called pulseless disease. Ischemic disorders such as dizziness, fainting, headaches, retinal hemorrhage, visual impairments and even blindness can emerge due to stenotic nature of the disease. In critical cases where structure of ascending aorta is disrupted, congestive heart failure or strokes may be observed. Due to inflammatory changes observed on intima and adventitia as well as medial layer of arteries, TA is characterized as a panarteritis [21]. Even though TA is considered to develop through 3 stages, depending on the patient and severity of the disease complex symptoms may manifest and make it difficult to make a diagnosis or establish the stage of the disease [22]. Therefore, it can be considered in two stages. In systemic, prevasculitis phase patients develop non-specific, active inflammation symptoms such as fever, fatigue, weight loss, appetite loss and lightheadedness. Progression through vasculitis stage can be characterized by claudication in extremities, diminished or absent pulse, long stenotic lesion, hypertension that is generally associated with renal artery stenosis, vascular bruits, retinopathy and dyspnea. Angiography or MRI that can detect arterial wall thickening is employed to confirm diagnosis. [23].

The first mention of TA in literature is attributed to a report by Rokushu Yamamato in 1830 where he described a 45-year-old male presenting high fever and absent radiant pulse. In fact, Morgagni in 1761 reported a 40-year-old woman large vessel aneurysms and stenotic lesions who lacked radial pulses whose symptoms were in line with TA phenotype [24]. In 1908, Mikito Takayasu presented the case of a 21-year-old woman who had loss of vision accompanied with ocular arteriovenous anastomoses [25]. To establish diagnosis and classification of TA, The American College of Rheumatology 1990 Criteria should be followed according to which, one should comply at least three criteria out of six in order to be diagnosed with TA. These criteria are as follows [26]:

- Disease onset age should be ≤ 40 years.
- Claudication present in extremities.
- Diminished pulse in brachial artery(ies).
- 10 mmHg difference in blood pressure between arms
- Presence of bruit over abdominal aorta or subclavian arteries

- Stenosis or occlusion in aorta, its main branches or subclavian arteries.

1.3.1 Epidemiology of Takayasu Arteritis

Takayasu's Arteritis is a rare vasculitis that predominantly affects women. Although disease onset range was observed as 15-30 years of age, reports suggest that peak incidence is observed at 20-30 years old [27]. TA is not confined to a specific geography but rather observed worldwide with prevalence of 4.7 – 22 cases per million [28, 29]. The highest prevalence is observed in Japan with 40 cases per million [30]. Furthermore, the highest incidence ratio observed is in Kuwait with 2.2 cases per million followed by Japan with 1-2 cases/million. Yet, incidence ratio calculated in Europe is 0.4 – 1.3. Recent studies from Aegean region of Turkey show prevalence of TA as 14.7 [31] whereas in the Northwestern part of the country where population considered was ≥16 years old, this value is up to 33/1,000,000 [32]. Another study conducted in Western Turkey reported an incidence rate of 1.1 cases per million which is similar to those reported from Europe [33].

A sex bias is also found to be significant for TA with females being affected in 12:1, 11:1 and 9:1 ratio in Turkey, North Europe and Japan, respectively. In contrast with these findings, in India and Israel this ratio decreases to 1.6:1 and 1.7:1, respectively [34, 35]. Thus, it may be necessary to consider ethnicity and sex not only while studying TA as prevalence and incidence values but also for estimating severity and progression of TA [36].

As TA is a systemic inflammatory disease, morbidity related to arterial lesions, occlusions or stenosis such as heart failure, stroke or transient ischemic attacks are frequently observed in TA patients. Progressive and relapsing nature of TA generally impairs patients' ability to do daily activities [37, 24]. According to a study administered in NIH, 47% of patients were designated as permanently disable whereas 26% had difficulty conducting their day-to-day activities during active stage of disease [22].

1.3.2 Genetics of Takayasu Arteritis

Although etiology of Takayasu Arteritis is still not clearly elucidated, it is discussed to be multifactorial. Familial aggregation in some cases and monozygotic twins (MZT) affected from TA suggest involvement of genetic factors in pathogenesis [38, 39, 40, 41].

Genetic contribution to TA has been shown primarily for its association with the human leukocyte antigen (HLA) super-locus which encodes 6 *HLA* genes along with 132 protein coding genes whose products are involved in regulation of immune response, antigen processing and presenting. HLA region lies in chr6p21 spanning 3.6 Mb and is associated with various autoimmune disorders such as Behçet's syndrome, diabetes and rheumatoid arteritis [42]. By investigating the segregation of HLA alleles within Japanese familial TA, HLA typing resulted in a haplotype that consist of HLA-A9, -A10 and -B5 [43]. Although HLA-B5 association has been established, the results are heterogenous depending on the serological subtype (-B51 or -B52) observed within the study population [44]. Presence of B52 allele whose association with TA was also confirmed in Korean, Indian and Turkish populations [44, 45, 46] may affect the prognosis of disease which presents higher inflammation and a more severe form of cardiac involvement [47, 48, 49]. However, these findings were not confirmed in North American or Arab populations who presented HLA-A2, A9 and Bw35 [50].

Genome-wide association studies (GWAS) realized with Turkish and North American cohorts revealed suggestive association with *FCGR2A/FCGR3A* and *IL12B* loci. The strongest association was found to be with a variant (rs12524487) in *HLA-B/MICA* intergenic region whose genetic effect was fine mapped to 4 variants located in the region. *HLA-DQB1/HLA-DRB1* locus also showed significant association but fine-mapping in the region was not possible due to strong linkage disequilibrium (LD). The most significant association in *FCGR2A/FCGR3A* locus was found to be with rs10919543 variant which were in high LD with 5 additional variants making it not possible to point out a single candidate variant [51]. Examined risk allele in this locus leads to increased expression of *FCGR2A* which were already shown to be involved in susceptibility to other autoimmune diseases [52].

IL12B association was also revealed in a study where 167 Japanese TA patients and 663 healthy controls were included. It was significantly associated with disease phenotype and variant rs6871626 located in the gene has effect on clinical onset as patients with the risk allele (A) had elevated C-reactive protein (CRP) level and aortic regurgitation. Investigation of interaction between HLA-B52 and *IL12B* locus showed a synergistic relationship [53].

RPS9/LILRB3 locus which resides in leukocyte receptor complex gene cluster that encodes for immunoglobulin (Ig)-like receptors interacting with HLA class-I was shown to be significantly associated to TA. Even though this locus was pinned by

rs1166543 index variant, it was not possible to extract a single associate variant due to high LD in the locus. Yet, the risk allele of the variant was correlated with decreased *LILRB3* expression [54].

Although *MLX* region was shown to have significant association with TA in a study with Japanese cohort, the 17q21 region where *MLX* gene lies was previously associated with various immune-related diseases such as Crohn's disease (CD) or psoriasis leading to inconclusive effect on TA susceptibility. When 96 patients were genotyped for rs665268, ones with more severe aortic regurgitation and higher number of lesions were carrying the risk allele. In vitro studies suggest involvement of *MLX* risk variant in pathogenesis of TA through recruitment of inflammasome complex [54, 55].

Another pro-inflammatory gene to be involved in pathogenesis of TA is *IL6*. Studies have shown genetic association of two promoter region variants with moderate effect along with an intronic variant that lies in an enhancer region. Studies confirmed elevated IL-6 serum levels in TA patients and treatment with antibodies against IL-6 receptor (tocilizumab) improved symptoms of TA [56, 57, 58, 59].

1.3.3 Immunopathogenesis of Takayasu Arteritis

In addition to genetic contributors, infections due to pathogens were suspected to play a role in TA pathogenesis due to cellular response to mycobacterial 65 kDa-heat shock proteins (HSP) which results in granulomatous inflammation in arterial walls. The molecular mimicry between human hsp-65 hypothesized to be the initiating factor [60]. Several cases with previous or concomitant tuberculosis (TB) infections were reported from various countries [61, 62].

Clear immunopathogenesis of TA is still not completely known but research imply involvement of both cell-mediated and humoral immune response. Cellular infiltrations, aberrant thickening, intimal proliferation of media and adventitia along with necrosis of arterial wall accompanying inflammation are indicatives of cellular immune response [63, 64, 65]. Initial infection takes place primarily in vasa vasorum via massive cell infiltration. It is followed by recruitment of lymphocytes, plasma cells and natural killer cells to the site of infection which initiates caspase-dependent apoptosis [66]. CD4+ and CD8+ T cells which co-localize with dendritic cells [67] become a crucial component in pathogenesis due to damage they inflict upon media and intima through secretion of perforin and granzyme B that leads to cell destruction

and necrosis. Upon migration of dendritic cells to the affected vessel wall, various cytokines such as interferon-γ (IFN-γ), IL-12, IL-18 and tumor necrosis factor-α (TNF-α) are secreted leading to granuloma formation [68]. In addition to T-cells, Th17 and Th1 cells that secrete IL-17 and IFN-γ, respectively are discovered to be significantly expanded leading to inflated levels of their respective cytokine [69]. Although reports on B-cell involvement in TA pathogenesis are controversial, many studies prove presence of increased number of antibodies presenting B-cells and CD20+ B-cells around lesions in TA patients [70].

1.4 Next Generation Sequencing

High-throughput sequencing (HTS) Technologies have become an integral part of the omics studies as it allows researchers access to big data, process and analyze it using analytic tools enlightening the underlying complex mechanisms of biological processes. The Human Genome Project (HGP) took more than 10 years to be completed and a Sanger-sequencing based automated capillary electrophoresis method was employed [71]. Development of next generation sequencing (NGS) technologies now allow sequencing of DNA and RNA much faster. Its applications in basic research and clinical studies allow identification of polymorphisms, structural variants, causative genes or susceptibility loci in rare, Mendelian or complex disorders [72]. Different applications such as whole genome sequencing (WGS), whole exome sequencing (WES) or whole genome bisulfite sequencing (WGBS) on NGS platforms are available and widely used [73].

There are multiple platforms that perform NGS and they employ different methods such as amplification based on by polymerase chain reaction (PCR) dependent technique or ligation-based method [74]. In our study NGS was performed by Illumina's platform where workflow is divided into three main stages as library preparation, sequencing and data analysis. In the first step, input DNA either genomic or cDNA is fragmented followed by end-repair and ligation to adaptor sequences which contain barcodes that allow indexing. Amplification step takes place on flow cell where short synthetic oligonucleotides (oligos) are embedded. Input DNA loaded onto the chip is retained through complementation between oligos and adapters. Bridge amplification method where clusters containing multiple copies of a DNA segment is followed by paired-end sequence in which target DNA undergoes massive parallel sequencing from both 5' and 3' ends leading to higher accuracy. Simultaneous

to sequencing, reads are aligned to the reference genome which is step that allows identification of single nucleotide polymorphisms (SNP), deletion/insertions or structural variations [75].

1.4.1 Whole genome SNP genotyping

With the advancements in HTS technologies, WGS became an essential component in human-disease related research. SNPs represent the most common genetic variation in the human genome and 4-5 million SNPs have been known to exist. These may be individual-specific or common and have designated, fixed positions along the genome [76]. SNPs serve as markers allowing genetic mapping where a priori knowledge is not required and can be utilized in various applications such as genome-wide association studies (GWAS), homozygosity mapping and linkage analysis [77]. SNP array genotyping can also identify loss of heterozygosity, mosaiscm and even altered DNA methylation profiles in cancer cell lines [78]. Commercially available SNP chips can genotype almost 4.5 million SNPs with up to 99% accuracy. There are specialized chips developed depending on the type of research such as those specific to a disease or a population [79]. Microarray based genotyping strategy is the most commonly used method where allele specific probes complementary to precisely known SNPs are immobilized on the chip surface. Following the hybridization of fragmented DNA to the probes, signal intensity generated due to fluorescently labelled nucleotides is measured and the output is obtained as a genotype profile over the whole genome [80].

1.4.2 Whole exome sequencing and analysis

Whole exome sequencing (WES) is an HTS technology that allows sequencing of protein coding portions of genome known as exome as well as splicing sites. In contrast to WGS which covers 95%- 98% of the whole genome, WES covers 1.5% of the genome [81]. Its advantages over WGS is that it is cost-effective, reduces the amount of data generated and requires less time. WES is an effective method for identifying causative genes as approximately 85% of Mendelian disorders have genetic basis on the coding region of the genome whereas only 1% of the mutations lie in the regulatory (non-coding) regions [77, 82]. WES offers many advantages as it is unbiased regards to incomplete penetrance, locus or phenotypic heterogeneity and expression variability [83]. Yet, it has limited ability to capture structural variants or variants that lie in GC-rich regions or repetitive regions of the genome [84].

After sequencing is achieved through the process described above, raw data obtained in FASTQ format is handled in three main stages which are mapping, variant calling and annotation. High-throughput reads that pass the quality control threshold are firstly aligned to the reference genome either hg19 or hg38 by using an alignment tools known as Burrows-Wheeler Aligner (BWA) or Genome Analysis ToolKit (GATK) [85, 86] which generates a file in BAM format followed by variant calling step where true variants are differentiated from noise [87].

The last step is the annotation of variants where information concerning chromosomal position, nucleotide change and its impact on the amino acid, frequency of variant in various population is incorporated. The resultant variant call format (VCF) file generally contains 20000 - 50000 variants. In order to detect causative pathogenic variant stringent filtering and prioritization criteria are applied which depend on the disease inheritance model and pedigree under investigation [88]. One of the most crucial parameters is the minor allele frequency (MAF) which is at least set to < 1% for which variants with MAF ≥ 0.01 are considered SNPs. Population frequencies over various databases should be taken into account [89]. Synonymous variants do not alter the amino acid to be incorporated into the final protein product are known to exert minimal effect and hence are also eliminated. Depending on the inheritance model, homozygous or heterozygous variant are filtered out. Outcomes of prediction tools such as SIFT [105], PolyPhen [106], CADD [107] and REVEL [108] may be examined and included into analysis as one piece of suggestive proof. Nevertheless, even after such filters total variant number can be over 100. Thus, additional strategies such as linkage, homozygosity or de novo based should be employed to pinpoint the possible causative variant or gene [90, 91].

1.5 Linkage Analysis

Linkage analysis is a statistical method that can locate disease causing or susceptibility locus for a genetic disorder. By examining a multigenerational pedigree with multiple affected individuals, co-segregation of genetic markers can be traced and loci that are shared among people conferring the disease phenotype can be established [77]. Integrating knowledge from homozygosity mapping, variations that are identical by descent (IBD) can be traced accurately and loci that generated significant linkage signal can be narrowed down to gene(s) resolution by incorporating WES data [83]. The basis of the method relies on the fact that the genes that are physically close to

one another are genetically linked and therefore tend to be inherited together rather than segregating independently [92]. Via tracing recombination events surrounding the disease allele and marker, genetic maps where distance is measured by recombination fraction and 1 centimorgan (cM) is equal to 1 Mb can be constructed. Likelihood of linkage is measured by logarithm of odds (LOD) score [93]. LOD score >3 indicates significant genetic linkage which means that loci are 1000 times more likely to be linked rather than not, whereas LOD score < -2 is accepted as no linkage evidence [94, 95]. Parametric linkage analysis (PL) require appropriate pedigree information, disease allele frequency and inheritance and penetrance model of the disease [96] which can be executed in a multipoint in which haplotype made up of a set of markers are analyzed [97]. On the other hand, non-parametric linkage analysis (NPL) which is more commonly used for complex disorders do not require an inheritance model while estimating common alleles among affected individuals. Even though, linkage analysis generates regions that may bear disease locus and therefore reduce the number of variants to be analyzed in downstream analysis such as WGS or WES, reduced penetrance and presence of phenocopies who show the disease phenotype without having the disease causing variant may diminish the linkage signal [98, 99].

1.6 Homozygosity Mapping

Homozygosity mapping is a technique applied especially when the studied disease trait is recessive and a consanguineous family is the focus. It aims to identify homozygous regions that may contain the suspected disease gene or locus. Its basis lies on the knowledge that in inbred families, closely related individuals would inherit alleles from a founder [100]. This is beneficial when the disease under investigation is inherited in an autosomal recessive manner. The locus containing the disease-causing mutation(s) has a high probability of carrying two same allele sets (haplotypes) that are identical by descent (IBD). These IBD regions are inherited from the ancestor without being recombined through meiosis [101].

1.7 Aim of the Study

Aim of this study was to elucidate underlying genetic cause of TA and CRMO by using linkage analysis followed by exome analysis. CRMO is a chronic autoinflammatory disorder where sterile bone lesions form around metaphysis of

appendicular and axial skeletal bones. Identified genetic features of other CNOs as Majeed syndrome and DIRA along with reported familial cases with multiple affected individuals strongly suggest the genetic contribution in disease pathogenesis. Takayasu arteritis is a rare autoimmune disorder with undetermined etiology. Although highest prevalence is observed in East Asia, TA is also reported in various countries with increasing prevalence including Turkey. Presence of monozygotic twins afflicted with TA and cases where familial aggregation was observed may indicate involvement of genetic components in pathogenesis of TA. Genome-wide association studies aimed to locate disease susceptibility locus have indicated significant association with *HLA-B*, *IL12B* and *FCGR2A/FCGR3A*. Yet, to this date no study where linkage analysis was used to identify disease causing gene or locus was reported. We have studied four consanguineous families with multiple affected members in the same generation. The first TA family consists of healthy mother, father and brother along with two affected sisters. The second TA family is similar to the first as two out of three children are affected. In the third TA family there are two affected half-sisters and their mother. The fourth family consists of two brothers diagnosed with CRMO and their healthy parents and sister. Information on pedigrees are suggestive of autosomal recessive inheritance. Therefore, we performed linkage analyses under recessive model to pin out statistically significant regions along the genome that could harbor the disease-causing gene(s). By integrating the data obtained from whole exome sequencing we have aimed to fine-map the potential causative variants involved in TA and CRMO pathogenesis.

2. MATERIAL AND METHODS

2.1 Material

2.1.2 Patients

Two CRMO patients from a consanguineous family are diagnosed and followed in Pediatric Rheumatology and Pediatric Nephrology Department in Pamukkale University Faculty of Medicine.

Four TA patients from two unrelated consanguineous families are diagnosed and followed in Rheumatology Department of Istanbul University Cerrahpaşa Faculty of Medicine and two patients from one consanguineous family are followed in Child Health and Diseases Department of Istanbul University Cerrahpaşa Faculty of Medicine.

In consanguineous CRMO1 family patients diagnosed with CRMO are two male siblings with a healthy sister whose parents are also healthy. Figure 2.1 shows the family pedigree. Patients were already being followed in Pamukkale University Faculty of Medicine to where non-affected members were invited for obtaining clinical history and physical examination.

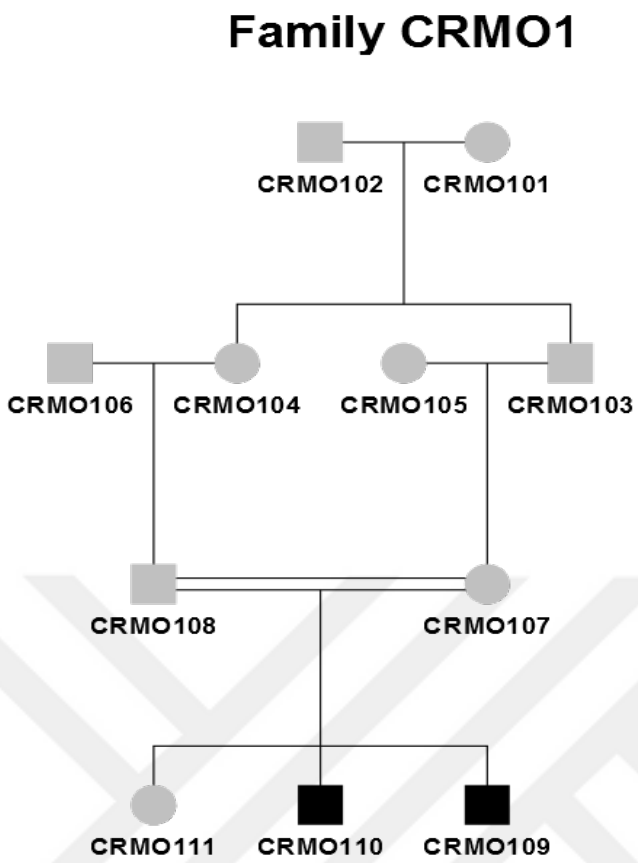

Figure 2.1: Pedigree of CRMO1 family. Individuals included in the study are CRMO109 and CRMO110. Dark shades indicate affected status.

Two sisters TA1-03 and TA1-04 from TA1 family had onset at the ages of 22 and 26, respectively. Their mother has scledorma and their brother has psoriasis. TA1-03 had claudication of left arm and CT showed occlusion of the proximal left artery and the left axillary artery where TA1-04 showed increased FDG uptake in the right carotid artery and abdominal aorta followed by stenosis in the left carotid artery. TA1 family pedigree is given in Figure 2.2.

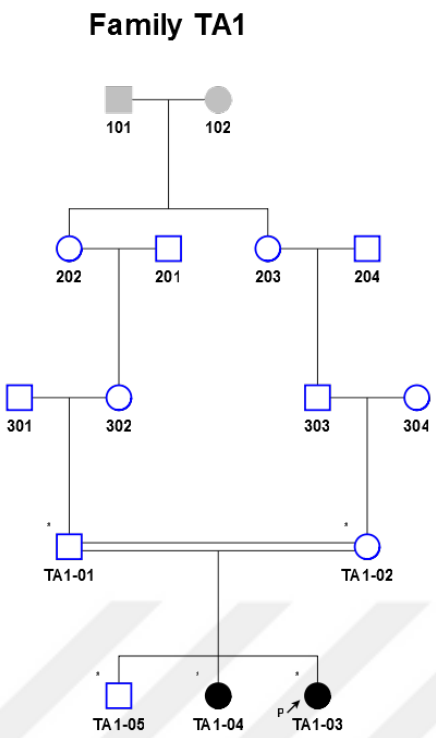

Figure 2.2: Pedigree of the family TA1. Individuals included in the study are indicated with the asterisk and the proband with P. Dark shaded members are patients.

In consanguineous family TA2 two sisters TA2-01 and TA2-02 are diagnosed with TA. TA2-01 had onset at the age of 5 with claudication in arms and carotidynia symptoms. Her elder sister was diagnosed with TA at the age of 32 when she had presented constitutional symptoms along with arthralgia that lasted for 6 months. Their parents are healthy. TA2 family pedigree is given in Figure 2.3.

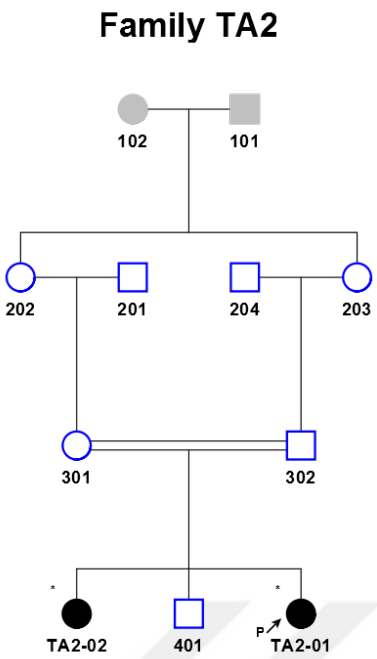

Figure 2.3: Pedigree of TA2 family. Individuals included in the study are indicated with asterisk and the proband with P. Dark shaded members are TA patients.

Two siblings (sibs) that have the same mother but different fathers are diagnosed with TA in family TA3. TA3-02 was 22 years old when she was diagnosed upon detection of severe stenosis in the proximal right subclavian artery and in the left renal artery. CT angiography of thorax and abdomen revealed stenosis throughout the aorta and its main branches leading to TA diagnosis of TA3-03 at the age of 24. TA3 family is given in Figure 2.4.

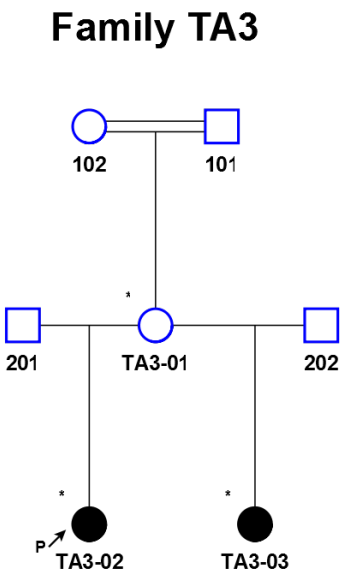

Figure 2.4: Pedigree of TA3 family. Individuals included in the study are indicted with asterisk and the proband with P. Dark shaded members are TA patients.

2.1.3 Equipment

Equipment used in this study is listed under Appendix A.

2.1.4 Chemicals and buffers

Chemicals and buffers used in this study are listed under Apendix B and C.

2.2 Methods

2.2.1 Sample collection

Blood samples of 5 members of TA1 family, 2 members of TA2 family, 3 members of TA3 family and 5 members of CRMO1 family were collected into 10 mL EDTA tubes and were preserved in 4°C. Individuals included were informed accordingly and agreed to participate in the study which was conducted in accordance with the Helsinki Declaration under the ethics committee approval (MBG.22/2014).

2.2.2 DNA isolation

Genomic DNA was obtained from the stored blood sampled of individuals. Blood samples and buffers to be used were warmed to 20 °C. 2 mL of the blood samples were transferred to sterile 50 mL falcon and the remaining was stored in 4 °C for further use. 5 mL of Red Blood Cell Lysis Buffer (RBL) was added onto the blood sample and was incubated on the shaker for 15 minutes at room temperature. Mixture was centrifuged for 10 minutes at 2200 rpm. Supernatant was carefully discarded without

disturbing the pellet which was resuspended with 2 mL of RBL. It was again incubated on shaker for 15 minutes at room temperature and centrifuged for 10 minutes at 2200 rpm. Supernatant was discarded and pellet was disrupted by vortexing. 5 mL of TKM2 buffer and 670 µL of 10% SDS was added to the pellet and vortexed vigorously followed by incubation at 37 °C for 30 minutes. 1.7 mL of 6M NaCl was added into the solution and vortexed after which the solution was partitioned into 2 mL eppendorf tubes. They were centrifuged for 13 minutes under 13000 xg. Supernatant was transferred into 15 mL falcon that contained 5 mL of 100% ethanol. Falcon was inverted gently until cloudy DNA strands were visible which were fished out of ethanol and transferred into a 1.5 mL eppendorf that contained 200 uL of %70 ethanol. The solution was centrifuged for 13 minutes under 14000 xg. Ethanol was carefully removed and tube was left to dry in the hood for 20 minutes. The pellet was resuspended with 100 µL of TE buffer and incubated at 65 °C for 10 minutes while vortexing the tube each 2 minutes. DNA concentration and purity of the samples were assessed in NanoDrop at 260/280 nm which were then stored at 4 °C until use.

2.2.3 SNP genotyping

Whole genome single nucleotide polymorphism (SNP) genotyping was performed for 3 healthy (TA1-01,02 and 05) and 2 affected (TA1-03 and 04) members of TA1 and for 2 affected (TA2-01 and 02) members of TA2. 5 members of CRMO1 family (2 affected, 3 healthy) were also chosen for SNP genotyping. Genotyping was realized in Yale Center for Genome Analysis (YCGA) by using Illumina Infium OmniExpress-24 BeadChip array that consists of ~710,000 fixed markers. Due to the requirements specified by YCGA, 20 µL of 50 ng/ µL of DNA samples were shipped via cold transport chain. Raw genotype data was obtained and call rate and signal ratios were established by using Illumina GenomeStudio V2011.1.

2.2.4 Linkage analysis

Linkage analysis was realized with the SNP genotyping data of TA1 and CRMO families separately along with combined SNP genotyping data of TA1 family and two members (TA2-01 and 02) of TA2 family. For statistical analysis of linkage, GeneHunter v2.1r5 – Multipoint Linkage Analysis package in the EasyLINKAGE Plus software was employed [102]. The analysis was performed under fully penetrant recessive inheritance model with codominant allele frequency algorithm by using all pairs of affected/phenotyped sibs. For CRMO1 family linkage analysis was realized

under the same condition as for TA1 family except a change in allele frequency algorithm which was set as all individuals from marker file. Frequency of disease allele was set to 0.001 and penetrance to 0.99 while 30 markers per set are analyzed. Regions with >1.50 maximal LOD scores were chosen for further analysis. SNP genotyping data was used to confirm shared homozygosity of these regions among the affected individuals

2.2.5 Whole exome sequencing analysis and variant filtration

Whole exome sequencing was performed for 1 affected member from both TA1 (TA1-04) and TA2 (TA2-01), three members of TA3 (TA3-01, 02 and 03) along with 1 affected (CRMO110) from CRMO1 families which was realized in YCGA. NimbleGen SeqCap EZ MedExome Enrichment Kit (Roche) was utilized for sequencing library preparation followed by sequencing via Illumina HiSeq System. Raw sequence data obtained was analyzed following the instructions of GATK Best Practices workflow [86] After realizing the quality check steps, reads were aligned to hg19 human reference genome by using BWA package. SNP and indel realignment and discovery, removal of duplicates by GATK [103] outputs variant call files (VCFs) which were annotated with ANNOVAR. Filtration of the variants were done accordingly to the Figure 2.5 under appropriate mode of inheritance of TA [99].

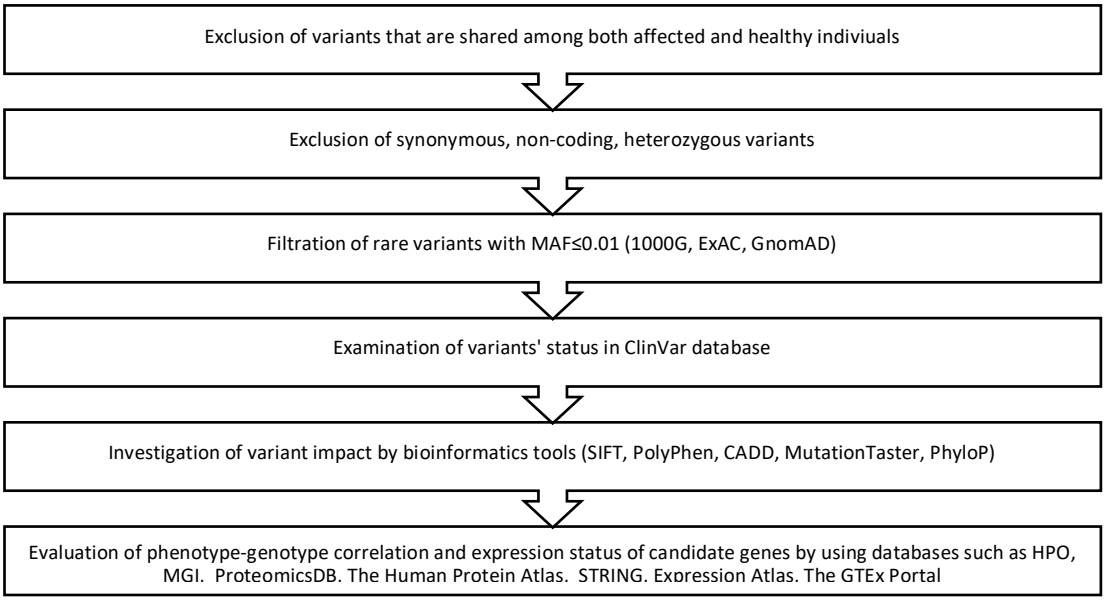

Figure 2.5: Criteria applied for variant filtration and prioritization for autosomal recessive inheritance model.

After filtering and prioritization of variants, candidate genes especially involved in inflammatory response and also in cardiovascular and/or immune system were thoroughly investigated in literature and in the aforementioned databases.

2.2.6 Homozygosity mapping

Homozygosity mapping was realized in order to confirm the regions obtained from linkage analysis and also check for homozygous regions that are IBD in affected sibs that might have been disregarded in the linkage analysis. By using Illumina Genome Studio v2011.1, auto-bookmark analysis was run with homozygosity detector algorithm. Parameters were adjusted as region length ≥ 200kb with minimum 50 SNPs/region where 1 heterozygote is allowed. Obtained results showed regions with loss of heterozygosity (LOH).

In order to ascertain the regions obtained both from linkage analysis and homozygosity mapping, by using Plink [104] IBD estimation was run in which homozygosity runs were calculated under given parameters for selected samples. Parameters in this analysis were set as same as the ones in homozygosity mapping. Outcome of these methods were compared and resultant regions were involved in further analysis.

3. RESULTS

3.1 SNP Genotyping

Whole genome SNP genotyping was performed for 12 individuals in total. Five of them (TA1-01, 02, 03, 04 and 05) are members of TA1 family, two of them (TA2-01 and 02) are from TA2 family and five (CRMO107, 108, 109, 110 and 111) are from CRMO1 family. Samples were analyzed on Illumina OmniExpress-24 BeadChip array which targets more than 710,000 markers. The quality of the raw data was checked on Illumina GenomeStudio V2011.1 software and can be seen in Table 3.1.

Table 3.1: Chip information, call rates and correctly estimated gender of samples by GenomeStudio software.

Sample ID	Gender	Call Rate	Chip Info	Chip Position
TA1-01	Male	0.9995582	200654560033	R05C01
TA1-02	Female	0.9983801	200654560033	R06C01
TA1-03	Female	0.9983282	200654560033	R07C01
TA1-04	Female	0.9983928	200654560033	R08C01
TA1-05	Male	0.9995863	200654560033	R09C01
TA2-01	Female	0.9984727	200654560202	R09C02
TA2-01	Female	0.9984965	200654560202	R10C02

3.2 CRMO

To investigate the genetic contribution CRMO pathogenesis, we firstly conducted linkage analysis in the CRMO1 family for which we had five SNP data from two affected sibs and three healthy individuals. Exome data of the affected member was further investigated in accordance with the regions obtained from linkage that had LOD score ≥1.50. Novel and rare, nonsynonymous and truncating variants mapping to these regions were further investigated for their phenotype-genotype correlation, expression status, animal models and for the impact of the variant on the final protein product.

3.2.1 Linkage analysis of CRMO1 family

After evaluating the quality of data as satisfactory, linkage analysis was performed. As Takayasu Arteritis and CRMO are suspected to be inherited under autosomal recessive model, linkage analyses parameters were set accordingly. For all analyses GeneHunter v2.1r5 algorithm was employed. It was run under fully penetrant recessive model with estimated disease allele frequency set to 0.1% and markers set was chosen as 30 with 0.01 cM spacing between each marker. LOD score threshold was chosen as ≥ 1.50 to be included in further analysis. Regions with smaller LOD score than 1.50 were considered as insignificant and were not included in the following steps. Multipoint parametric linkage analysis of CRMO1 family was conducted under fully penetrant recessive model with disease allele frequency 0.1%. Marker spacing was set to 0.01 cM while set of markers to 100. Table 3.2 presents the maximal LOD scores obtained for each chromosome.

Table 3.2: Maximal LOD score obtained from linkage analysis of CRMO1 family.

Chr	Start	End	Size (bp)	Maximal LOD score
1	242380057	245725714	3345657	1.7979
2	127949198	130943513	2994315	1.7980
3	-	-	-	-
4	188201350	190543118	2341768	1.7980
5	128692695	128718147	25452	1.2761
6	53289583	72300534	19010951	1.7979
7	17144007	18573000	1428993	1.0260
8	-	-	-	-
9	102939193	103338012	398819	0.2339
10	131907568	134607938	2700370	1.1136
11	21339877	24156271	2816394	1.7676
12	-	-	-	-
13	-	-	-	-
14	-	-	-	-
15	102376655	102461162	84507	0.3958
16	33557021	46614595	13057574	0.5680
17	-	-	-	-
18	-	-	-	-
19	-	-	-	-
20	39103371	44365626	5262255	1.7946
21	-	-	-	-
22	22371638	27550764	5179126	1.7978
23	-	-	-	-

As can be seen in Table 3.2 there were seven chromosomes that contain regions with maximal LOD score ≥ 1.50. Maximal LOD score 1.7979 was reached in chromosomes 2 and 4 which are 2994315 and 2341768 base pair (bp) long, respectively. The biggest linkage region is 19010951 bp long with 1.7979 LOD score and is located on chr6. These intervals were manually checked within SNP data to confirm homozygous regions that are only shared among affected sibs.

Homozygosity mapping was conducted in GenomeStudio where the parameters were set as follows: Region length ≥ 200 kb with 50 SNPs per region with 1 heterozygous

allowed. Although there were 79 regions longer than 200 kb, 54 of them were also obtained by linkage analysis. Further analyses were based upon these regions.

3.2.2 Investigation of WES data in the candidate linkage regions

First step of exome analysis was to identify candidate variants by combining the data obtained from linkage analysis with WES. In order to achieve this exome of the affected individual CRMO110 was investigated based upon the regions obtained from linkage and homozygosity mapping. Heterozygous, synonymous and non-coding variant were eliminated leaving 103 variants out of which none had MAF ≤ 0.05.

3.2.3 Whole exome sequencing analysis and variant prioritization

In CRMO1 family, WES data is available for one of the affected sibs, CRMO110. Variants that are synonymous, heterozygous and non-coding are eliminated as CRMO is a rare disorder that is hypothesized to be recessively inherited. Remaining variant with MAF ≤ 0.005 are given in Table 3.3.

Table 3.3: Homozygous, non-synonymous rare variants obtained from filtration and prioritization of CRMO110 WES data.

Location	Gene	dbSNP ID	Aminoacid change	Highest MAF observed
chr1: 109007875	*NBPF6*	rs1262638200	p.Val526Phe	0.0001549
chr1: 225528175	*DNAH14*	rs184094753	p.E3391X	0.003440
chr4: 6864048	*KIAA0232*	rs192490554	p.C647S	0.001623
chr4: 13616966	*BOD1L1*	-	p.E177K	-
chr5: 140741870	*PCDHGB2*	-	p.A723V	-
chr5: 140755502	*PCDHGA6*	-	p.S618A	-
chr7: 6693491	*ZNF316*	-	p.R755L	-
chr10: 134218884	*PWWP2B*	-	p.P294S	-
chr10: 134713136	*CFAP46*	-	p.T987P	-
chr11: 67188111	*CARNS1*	rs12786101	p.P259L	0.0001300

10 remaining variants that are listed in Table 3.3 were further analyzed with bioinformatic tools to determine their impact on protein function and possible contribution in disease pathogenesis. rs12786101 (*CARNS1*: c.776C>T) was identified

as candidate variant but the region the gene lies was found to be homozygous also in healthy members of the family.

Exome data was also examined with regards to the genes listed in Infevers database. Out of 44 genes only in 4 genes contained homozygous but intronic and MAF > 0.05 variants. These are: *NLRP3:* g.247597197_247597238del and *PSMA3*: g. 58724817_58724818insTT. Among these variants non were chosen for further analyses.

3.3 Takayasu Arteritis

To examine the genetic contribution involved in TA pathogenesis linkage analysis in two TA families were conducted. First analysis was done under multipoint, parametric and recessive model by using SNP data of five individuals from TA1 family, 2 affected sibs and their heathy parents and unaffected sister. The second analysis was conducted by combining SNP data of 2 affected sibs from TA2 family with that of TA1 family. Exome data of TA1-04, TA2-01 were investigated for variants that were located in the linkage regions. Lastly, these exomes and exome data from TA3 family were analyzed without being limited to the linkage regions.

3.3.1 Linkage analysis of TA1 and TA2 families

LOD score analysis yielded 135 regions with LOD score ≥ 1.50 distributed among various chromosomes in TA1 family. Table 3.4 shows the intervals where maximal LOD scores obtained within each chromosome. To verify that these regions were only shared among the affected sibs but not with the healthy members, whole genome SNP genotype was checked. 87 false positive regions which were also shared with the healthy members of the family were eliminated. In addition to validation of the regions, SNP data was also scanned for presence of homozygous regions shared between affected sibs that could be overlooked in linkage analysis. 11 more were identified from the SNP data in this manner. As a result, 59 regions longer than 200 kb with maximal LOD score ≥ 1.50 were taken into account.

Table 3.4: Maximal LOD scores obtained for each chromosome from linkage analysis of TA1 family.

Chr	Start	End	Size (bp)	Maximal LOD score
1	156133038	158647312	2514274	2.4956
2	48461933	55316938	6854945	2.5073
3	56493678	58025903	1532225	2.5074
3	131377277	136662529	5285252	2.5074
4	7142427	11861566	4719139	2.5074
5	109904162	110105163	201001	2.5068
6	109665903	112483235	2817332	2.5068
7	-	-	-	-
8	-	-	-	0.1700
9	-	-	-	0.8071
10	52109375	56313403	4204028	2.5074
11	60223018	61711475	1488457	2.5074
12	105442186	11717989	11728803	2.5074
13	-	-	-	0.6473
14	100800103	101204235	40412	2.4759
15	98503023	98211723	291300	2.3766
16	26869896	27163976	294080	2.4901
17	41251931	41773814	521883	2.5074
18	50571241	50937026	365785	1.6086
19	48510553	51393118	2882565	2.5074
20	5024928	5255078	230150	1.9033
21	-	-	-	-
22	37342719	37976791	634072	2.5060
23	-	-	-	-

Maximum LOD score was found to be 2.5074 for chromosomes 2, 3, 4, 5, 6, 10,11, 12, 17 and 19. Boundaries and trueness of regions represented in Table 3.4 were also verified by cross-checking these in the SNP genotyping data. Intervals where genotype of all members are identical were eliminated. Sites where affected sibs shared the same homozygous genotype were chosen for further analysis.

Regions obtained from the linkage analysis and SNP data were cross-checked with the help of homozygosity mapping ran via GenomeStudio. Results revealed 41 sites longer than 200 kb which are shared between affected sibs TA1-03 and -04. These coincide with the linkage regions.

Multipoint parametric linkage analysis within TA2 family could not be realized as only SNP genotyping data of affected sibs were available. In order to examine the shared homozygous regions between affected sibs, SNP genotyping data and pedigree of TA1 and TA2 were analyzed co-jointly in EasyLINKAGE Plus software under GeneHunter v2.1r5 algorithm. The conditions applied were equivalent to that of TA1. The analyses were run under autosomal recessive, fully penetrant model with estimated disease frequency as 0.1% and codominant allele frequency. Spacing between markers were set to 0.01 cM and set of markers to 30. Table 3.5 shows the results of analysis, were maximal LOD scores reached for each chromosome.

Table 3.5: Maximal LOD scores obtained for each chromosome from the co-joint multipoint parametric linkage analysis of TA1 and TA2 families.

Chr	Start	End	Size (bp)	Maximal LOD score
1	35484861	36626006	1141145	3.4098
2	54634288	54905778	271490	3.4727
3	154115627	157567892	3452265	4.0078
4	7120761	11816679	4695918	4.2764
5	109237529	110399033	1161504	4.0080
6	109548713	110007647	458934	4.2942
7	-	-	-	0.8517
8	43589390	43784378	194988	1.5477
9	-	-	-	1.3659
10	21523325	25007306	3483981	4.1730
11	65817307	66608991	791684	4.3006
12	74076692	74666737	590045	3.4073
13	-	-	-	-0.2729
14	105943679	107240779	1297100	4.0021
15	94890010	95062054	172044	3.4909
16	7161598	7282505	120907	2.1148
17	43216281	43836477	620196	4.0646
18	50336979	50544871	207892	3.3691
19	50432190	51186081	753891	3.3430
20	37001761	37243716	241955	1.9902
21	-	-	-	0.8188
22	22281093	22564646	283553	2.7856
X	-	-	-	-

Homozygosity mapping algorithm ran via GenomeStudio showed 117 homozygous regions that were longer than 200 kb which were present in both affected members of TA2 family. These were in correlation with the regions that generated significant (LOD score ≥1.50) linkage signal.

As can be seen in Table 3.5 regions with ≥ 1.50 LOD score were revealed in all chromosomes except 7, 9, 13 and 21. Regions with the highest LOD scores were found to be 4.3006, 4.2942 and 4.2764 in chromosomes 11, 6 and 4, respectively. Intervals obtained from linkage analysis were controlled in SNP genotyping data to validate that these are only common to the affected members rather than the healthy family members. Boundaries of regions were adjusted according to correlation of SNP genotypes which generated 123 regions that are longer than 200 kb. Table 3.6 shows the regions which were shared between all affected members from both families.

Table 3.6: Regions shared among all affected members of both TA1 and TA2 family.

Chr	Start (bp)	End (bp)	Length (bp)	Gene Count	Genes
chr1	48091748	48288016	196268	1	*TRABD2B*
chr1	53631902	53777543	145641	4	*CPT2, C1orf123, MAGOH, LRP8*
chr2	51308075	51513533	205458	-	-
chr2	152881908	153346845	464937	4	*STAM2, CACNB4, FMNL2, NUDCP1*
chr3	4158659	4345766	187107	1	*SETMAR*
chr3	25930262	26297737	367475	-	-
chr3	33375699	33926125	550426	4	*CLASP2, FBXL2, UBP1, PDCD6IP*
chr3	38052725	38417596	364871	6	*PLCD1, DLEC1, ACAA1, MYD88, OXSR1, XYLB*
chr3	154899711	155042640	142929	2	*MME, STRIT1*

As can be seen from Table 3.6 nine regions were common to four affected individuals TA1-03, -04 and TA2-01 and -02. The longest commonly shared region is located on chr3: 33375699 – 33926125 with 550,436 base pairs and contains 4 protein coding genes.

3.3.2 Investigation of WES data in candidate linkage regions

In this study, TA1-04, TA2-01, TA3-01,02, 03 were selected for whole exome sequencing. Output files were sent as .fastQ files from YCGA. After this point all downstream processes took place in our laboratory. Quality checks were done with fastQC in which PCR duplicates, reads shorter than 50 bases or with low coverage depth were removed. Following the trimming of adapters, reads were aligned using GATK which produced .BAM files which then were converted to .VCF files as a result of variant calling step. As the last step .vcf files were annotated via ANNOVAR.

TA1-04 and TA2-01 were filtered stringently using the same conditions for both individuals. Firstly, all the variants mapping to the already appointed linkage regions were filtered out. TA1-01 had 2578 variants mapping to the linkage regions given in Table 3.4. 402 of them were homozygous, exonic and splicing out of which 3 had MAF < 0.01. Resultant variants are given in Table 3.7.

Table 3.7: Homozygous, rare variants in TA1-04 WES data that map to the linkage regions obtained from linkage analysis of TA1 family.

Location	Gene	dbSNP ID	Aminoacid change	Highest MAF observed
chr7: 103301976	*RELN*	rs146986040	splicing	-
chr19: 50726342	*MYH14*	rs200818171	p.R189C	0.0051
chr19: 51171270	*SHANK1*	rs534920931	p.G1316D	0.0068

Variants that are given in Table 3.7 were further investigated with the aid of prediction tools; SIFT, Polyphen, CADD, REVEL and Mutation Taster to assess the impact of aminoacid change to the overall protein structure and function. PhyloP outcome was also examined to understand the effect evolutionary conservation of the aminoacid change.

Lastly, WES variants of TA1-04 and TA2-01 were filtered according to the linkage regions that were extracted from the multipoint parametric linkage analysis ran co-jointly for TA1 and TA2 families. Resultant variants picked up for both individuals are given in Table 3.8.

Table 3.8: Homozygous, rare variants observed in WES data of patients TA1-04 and TA2-01 that reside in the linkage regions obtained from co-joint linkage analysis of TA1 and TA2 families.

Patient ID	Location	Gene	dbSNP ID	Aminoacid Change	Highest MAF observed
TA1-04	chr11: 65352724	*EHBP1L1*	-	p.R1199C	-
TA2-01	chr1: 114442873	*AP4B1*	rs143286419	p.T88I	0.002
TA2-01	chr19: 37382520	*ZNF829*	-	p.K391N	-

It can be seen from Table 3.8, no common variants or genes that reside in the linkage regions were shared among affected individuals TA1-04 and TA2-01.

Due to lack of availability of TA3 family SNP data, linkage analysis was not realized and hence, WES data could not be analyzed accordingly.

3.3.3 Whole exome sequencing analysis and variant prioritization

In the next step of the study WES data of 4 TA affected individuals were analyzed separately in an unbiased manner as there were no shared variants or genes among TA1 and TA2 families in the candidate linkage regions. WES data is also available for members of TA3 family (2 affected sibs and their healthy mother). As the mother is unaffected, TA3-01 exome data was not filtrated according to disease variant prioritization criteria, rather it was used as a control for evaluation of candidate variant by checking its absence or presence in the healthy control. Figure 3.1 shows the number of variants for each individual in each step of filtration and prioritization. Firstly, heterozygous variants were excluded from exome data as TA is studied under autosomal recessive inheritance model. Non-coding and synonymous variants were also eliminated because they are predicted to have minimal effect on the protein structure and/or function. Not to rule out possible candidate variants, MAF was set to ≤ 0.01.

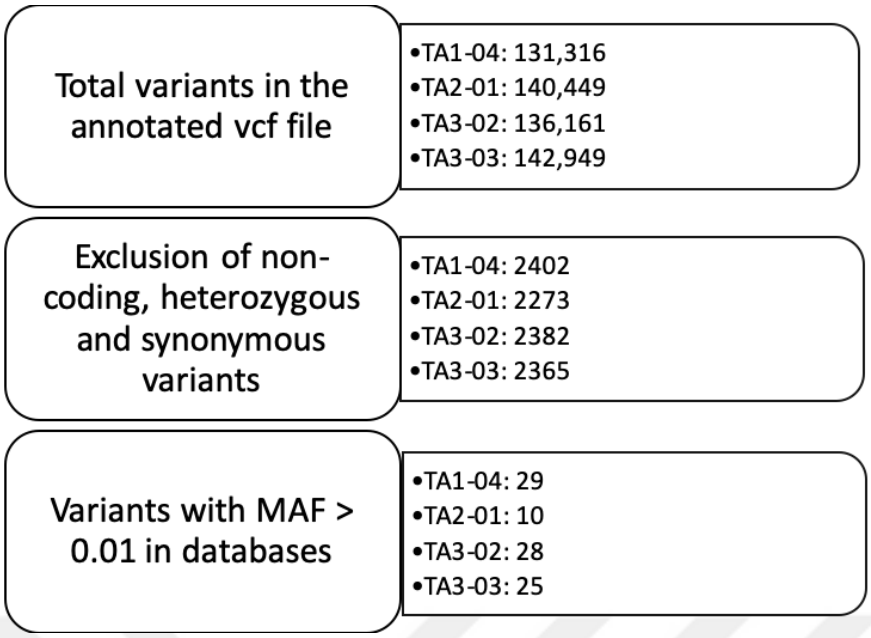

Figure 3.1: Number of total variants are given when exome files were filtrated according to criteria given above.

Following the filtration step, remaining number of variants were between 10-29 for affected members. For the individuals with available SNP data, the segregation of these variants was followed. For the recessive inheritance model the variant should lie in a region which is only homozygous among the affected individuals, variants that were found to be regions with similar genotypes in healthy individual were eliminated from further analysis. This analysis could be applied for TA1-04 and TA2-01 for which yielded 5 and 4 variants, respectively which were already detected in linkage analysis. 4 out of 5 variants for TA1-04 and 2 out of 4 variants in TA2-01 were already uncovered in linkage analyses and are given in Table 3.6 and Table 3.7. Remaining variants are given in Table 3.9.

Table 3.9: Homozygous, non-synonymous, rare variants obtained from filtration and prioritization of WES data of TA1-04 and TA2-01.

Patient ID	Location	Gene	dbSNP ID	Aminoacid Change	Highest MAF observed
TA1-04	chr10:47756706	*ANXA8L1*	rs376366484	p.T150M	0.0042
TA2-01	chr1: 17086085	*MST1L*	rs771454570	p.A271Gfs*25	-
TA2-01	chr10: 46339838	*AGAP4*	rs1401655909	p.W3X	0.000188

These were further analyzed with various bioinformatics tools to weigh the outcome of amino acid substitution on the protein structure, figure out the conservation of amino acid along with the suitability of the phenotype-genotype correlation. Databases such as GnomAD, ClinVar, Mendelian Inheritance of Men (OMIM), Human Protein Atlas (HPA) were checked for entries on these variants. Possible interacting partners on the related pathways were investigated using KEGG, String and PDB.

In family TA3 where WES data of 3 individuals (2 affected sibs and healthy mother) is available, filtration steps were applied only to the affected individuals (TA3-02 and -03). Healthy individual (TA3-01) data was used as an internal control that enabled verification of true variants that were only shared between the affected sib pairs. As can be seen in Figure 3.1, there were 28 and 25 variants left in TA3-02 and TA3-03, respectively. 5 homozygous variants that were also present in TA3-01 were removed from TA3-02 and -03. Out of the resultants, none were common for 2 individuals. Sibs did not contain variants that were located in the same gene.

4. DISCUSSION

Rare disorders are described as conditions with prevalence less than 1 in 2000 people. Approximately 300 million people are affected with rare disease worldwide and in Turkey this number is estimated as 5 million. More than 6000 rare diseases are identified to date and 80% of these are monogenic disorders. Disease onset is usually during childhood with chronic, progressive and debilitating nature. Patients life expectancy is lower than unaffected population.

Takayasu Arteritis (TA) is a rare type of large vasculitis that predominantly affects women younger than 40 years old. Even though it is characterized with inflicted aorta and its main branches. Inflammation on arterial walls are followed by stenosis, occlusion and thrombus formations. In cases with severe prognosis, aneurysms and even rupture of the inflammated artery can occur. Subclavian and brachial artery inflammation lead to disappearance of radial pulses in extremities. East Asian countries have high prevalence especially Japan with 40 cases per million. In Turkey, prevalence is calculated as 12.8 per million with an incidence rate of 1.11 per million. Etiology of TA is yet to be elucidated but both genetic factors and environmental factors such as bacterial or viral infections are hypothesized to take role in pathogenesis.

Reports from many ethnicities, monozygotic twins diagnosed with TA and familial aggregation in some cases suggest there is a genetic background. Genome-wide association studies realized with cohorts of different sample size and ethnicities show significant association to *HLA-B52* and *IL12B.* Results of GWAS in a North American, Turkish cohort revealed *HLA-B/MICA* and *HLA-DQB1/HLA-DRB1* regions to be associated to TA. Results showing association of *FCGR2A/FCGR3A* locus to TA was replicated in several studies. However, no previous study aimed mapping the disease gene by using linkage analysis or homozygosity mapping. Therefore, aim of this study is to identify gene(s) involved in TA pathogenesis by using genetic mapping techniques as linkage analysis and homozygosity mapping followed by exome analysis to further investigate regions detected in these analyses.

Four multiple affected consanguineous Turkish families were involved in the study. TA1 and TA2 families consist of 2 affected sisters and their healthy parents and a brother. TA3 family is made up of 2 affected half-sisters who share their mother but have different fathers. CRMO1 family consists of 2 affected male sibs and their healthy parents and sister. SNP genotyping was realized from the genomic DNA collected from the family members with which linkage analysis and homozygosity mapping were done. Nature of pedigrees suggest autosomal recessive inheritance mode for both TA and CRMO. Therefore, multipoint parametric linkage analyses were realized under fully penetrant recessive model, with 0.1% disease allele frequency. Homozygosity mapping was done tandem to linkage allowing comparison of the results obtained from both analyses and verify the factuality of the regions to be included in further into the study. In TA2 family SNP genotype data of 2 affected sisters enable homozygosity mapping and also performing co-joint linkage with TA1 family. From both analyses 123 homozygous regions that are longer than 200 kb were obtained out of which 9 are shared in affected members from both families. In accordance with these results, WES was done for affected individuals TA1-04 and TA2-01. Exome data was used to examine the variants in the genes that are found within the chromosomal regions appointed from linkage analyses and homozygosity mapping. After elimination of heterozygous, non-functional variants, filtration and prioritization steps of the ones with MAF > 0.01 revealed 4 and 2 variants in TA1-04 and TA2-01, respectively. Furthermore, exome data was analyzed along the whole genome rather than being limited to the linkage regions by applying the same filtration criteria. 1 and 2 variants were added for TA1-04 and TA2-01, respectively.

Remaining variants for TA1-04 are: *RELN*: c.103301976_103301984dup, *ANXA8L1*: c.449C>T, *EHBP1L1:* c.1321C>T, *MYH14*: c.565C>T and *SHANK1:* c.3947G>A. *RELN*: c.103301976_103301984dup (rs146986040) is a splicing variant. The gene encodes for a secreted glycoprotein which is involved in neuronal positioning in developing brain. There are 2 entries in ClinVar on the variant and they are denoted as benign. *RELN* gene is associated with epilepsy. Taken together, these finding suggest neuronal pathway involvement and hence not chosen as a final candidate. *EHBP1L1:* c.1321C>T is a novel variant that is not registered in NCBI's dbSNP. There is no disease associated with neither the variant nor the gene. Aggregated prediction obtained from *in silico* tools is uncertain. Ehbp1l1 may be involved in vesicular trafficking via its role as Rab effector however its function is not clearly defined. Its

high expression in lungs and gastrointestinal tract with previous studies showing epigenetic modifications on *EHBP1L1* relating to renal cell carcinoma prognosis does not make this variant a strong candidate TA. *MYH14*: c.565C>T (rs200818171) is a non-synonymous, rare exonic variant. The gene encodes for a non-muscle myosin and heterozygous mutations in the gene have been associated with peripheral neuropathy and autosomal dominant 4A deafness. These finding taken together with ClinVar entries on the variant implicated as benign and variant of unknown significance (VUS) do not make this variant as a prior candidate. *SHANK1:* c.3947G>A (rs534920931) has no ClinVar entries and is predicted to be a likely benign variant according to ACMG standards. Yet, its differential expression in hippocampul neurons and astrocytes and knockout mouse model with anxiety-related behavior and impaired long term memory supports elimination of this variant from being related to TA phenotype.

There were 4 potential variants left in TA2-01. As there were 5 entries indicated as likely benign about variant *AP4B1*: c.767C>T (rs143286419), and association of the gene *AP4B1* to spastic paraplegia and severe intellectual disability in MIM, rs143286419 is not further studied. No available information was found about variant *MST1L*: c.811dupG (rs771454570) in dbSNP, ClinVar or other databases. Conflicting information about the function of the gene and no previous research does not make this variant as a strong candidate. *AGAP:* c.G8A (rs1401655909) is a stop gain mutation. *In silico* tools predictions are denoted as N/A and fails to give further information. Gene is hypothesized to be involved in endocytosis related pathways but without definitive evidence. No ClinVar or MIM entry about *ZNF829:* c.1173A>T is available but aggregated prediction obtained indicate its benign nature. There is no available protein expression data on the protein but mRNA expression is highly expressed in testis which makes the variant not a prior candidate in TA pathogenesis.

Exome data of 2 affected half-sisters (TA3-02 and -03) and their healthy mother (TA3-01) was available in TA3 family. Following the same conditions for analyses of exomes of TA3 family, presence of variant(s) or gene shared by affected sibs was examined. Out of almost 140,000 variants found in the exomes, only 28 and 25 in TA3-02 and TA3-03, respectively were homozygous, rare and located within functional (exonic, splicing) regions. Homozygous rare variants found common with TA3-01 were eliminated leaving 23 and 20 variants respectively. Individuals did not have any common variants nor did they share a gene. Each exome was also examined for the variants and loci in literature that had been associated with TA by GWAS but

none of them were present in our samples. As we based our analyses on autosomal recessive inheritance model, it was expected to pin out homozygous variants shared between affected sisters in TA3 or to map out the genes that got hits in both patients. Although common variants or genes were detected, presence of these in the healthy mother eliminated the possibility of the said variants' candidacy. One limitation can be addressed is lack of SNP data in this family. With SNP data, linkage analysis and homozygosity mapping can be done which would allow concentrating on the genomic regions that generate significant linkage signal. In this way more accurate mapping could have been achieved leading to identification of associated locus and its thorough examination.

Linkage analysis and following homozygosity mapping in CRMO1 family revealed 54 regions that were larger than 200 kb and were shared only among affected sib pair. Exome of an affected individual did not reveal any variants mapping to the linkage regions and were homozygous, non-synonymous, rare. Therefore, exome data was investigated without prior region specification and also for variations in the genes that were enlisted in Infevers database which yielded 10 and 2 variants, respectively. Further analyses were done with regards to the predicted impact of the variation on the protein function, its interacting partners and pathways they are involved in and its implications on CRMO pathogenesis. All in all, variant found in *CARNS*:c.1145C>T was suggestive of a candidate due to its involvement in synthesis of carnosine which is mainly found in skeletal muscles. But it was shown that it lied in a homozygous SNP region that were common to all members of the family which led to elimination of this variant as well. Although 2 variants were identified within autoinflammatory disease related genes that could have been linked to the autoinflammatory nature of the disease, they were not considered as candidates as they were not rare variants. Reported familial cases and recurrence of various autoinflammatory disorders in family members and/or first-degree relatives suggest a genetic predisposition in CRMO pathogenesis. Therefore, remaining variants might require further study even though genotype-phenotype correlation does not appear to be evident.

This is the first study that employs genome-wide linkage and exome analysis in TA and CRMO families aimed at identifying genetic factors in disease pathogenesis. Families under investigation are valuable for the literature as they are all consanguineous and have multiple affected members in one generation. Although we could not identify variants found in a common gene, candidate variants should be

further examined for their potential involvement in pathogenesis. Expanding these analyses by including exome and SNP data of all the first-degree relatives would allow increasing the detection power of the study and hence more insight to the genetic background of TA and CRMO.

REFERENCES

[1] **Url-1 <***https://www.eurordis.org/***>**, date retrieved 22.07.2020.

[2] **Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., & Rath, A.** (2019). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics, 28(2), 165–173. https://doi.org/10.1038/s41431-019-0508-0

[3] **Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E.** (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nature Reviews Genetics, 14(10), 681–691. https://doi.org/10.1038/nrg3555

[4] **Cox, A. J., & Ferguson, P. J.** (2018). Update on the genetics of nonbacterial osteomyelitis in humans. Current Opinion in Rheumatology, 30(5), 521–525. https://doi.org/10.1097/bor.0000000000000530

[5] **Cox, A. J., Zhao, Y., & Ferguson, P. J.** (2017). Chronic Recurrent Multifocal Osteomyelitis and Related Diseases—Update on Pathogenesis. Current Rheumatology Reports, 19(4). https://doi.org/10.1007/s11926-017-0645-9

[6] **Chen, Z., Cheng, L., & Feng, G.** (2018). Bone inflammation and chronic recurrent multifocal osteomyelitis. European Review for Medical and Pharmacological Sciences, 22(5), 1380–1386. https://doi.org/10.26355/eurrev_201803_14482

[7] **Giedion, A., Holthusen, W., Masel, L. F., & Vischer, D.** (1972). Annales de radiologie, 15(3), 329–342.

[8] **Probst, F. P., Björksten, B., & Gustavson, K. H.** (1978). Radiological aspect of chronic recurrent multifocal osteomyelitis. Annales de radiologie, 21(2-3), 115–125.

[9] **Stern, S. M., & Ferguson, P. J.** (2013). Autoinflammatory Bone Diseases. Rheumatic Disease Clinics of North America, 39(4), 735–749. https://doi.org/10.1016/j.rdc.2013.05.002

[10] **Job-Deslandre, C., Krebs, S., & Kahan, A.** (2001). Chronic recurrent multifocal osteomyelitis: five-year outcomes in 14 pediatric cases. Joint Bone Spine, 68(3), 245–251. https://doi.org/10.1016/s1297-319x(01)00276-7

[11] **Iyer, R. S., Thapa, M. M., & Chew, F. S.** (2011). Chronic Recurrent Multifocal Osteomyelitis:Review. American Journal of Roentgenology, 196(6_supplement), S87–S91. https://doi.org/10.2214/ajr.09.7212

[12] **Catalano-Pons, C., Comte, A., Wipff, J., Quartier, P., Faye, A., Gendrel, D., Duquesne, A., Cimaz, R., & Job-Deslandre, C.** (2008). Clinical outcome in children with chronic recurrent multifocal osteomyelitis. Rheumatology (Oxford, England), 47(9), 1397–1399. https://doi.org/10.1093/rheumatology/ken249

[13] **Jansson, A., Renner, E. D., Ramser, J., Mayer, A., Haban, M., Meindl, A., Grote, V., Diebold, J., Jansson, V., Schneider, K., & Belohradsky,**

B. H. (2007). Classification of non-bacterial osteitis: retrospective study of clinical, immunological and genetic aspects in 89 patients. Rheumatology (Oxford, England), 46(1), 154–160. https://doi.org/10.1093/rheumatology/kel190

[14] **Hofmann SR, Schwarz T, Möller JC, et al.** (2011). Chronic non-bacterial osteomyelitis is associated with impaired Sp1 signaling, reduced IL10 promoter phosphorylation, and reduced myeloid IL-10 expression. Clin Immunol. 141(3):317-327. doi:10.1016/j.clim.2011.08.012

[15] **Ferguson PJ, Chen S, Tayeh MK, et al.** (2005). Homozygous mutations in LPIN2 are responsible for the syndrome of chronic recurrent multifocal osteomyelitis and congenital dyserythropoietic anaemia (Majeed syndrome). J Med Genet. 42(7):551-557. doi:10.1136/jmg.2005.030759

[16] **Lukens, J. R., Gross, J. M., Calabrese, C., Iwakura, Y., Lamkanfi, M., Vogel, P., & Kanneganti, T.-D.** (2014). Critical role for inflammasome-independent IL-1 production in osteomyelitis. Proceedings of the National Academy of Sciences, 111(3), 1066–1071. https://doi.org/10.1073/pnas.1318688111

[17] **Kuemmerle-Deschner, J. B., Welzel, T., Hoertnagel, K., Tsiflikas, I., Hospach, A., Liu, X., … Weber, A. N.** (2020). New variant in the IL1RN-gene (DIRA) associated with late-onset, CRMO-like presentation. Rheumatology. https://doi.org/10.1093/rheumatology/keaa119

[18] **Cox, A. J., Darbro, B. W., Laxer, R. M., Velez, G., Bing, X., Finer, A. L., Erives, A., Mahajan, V. B., Bassuk, A. G., & Ferguson, P. J.** (2017). Recessive coding and regulatory mutations in FBLIM1 underlie the pathogenesis of chronic recurrent multifocal osteomyelitis (CRMO). PLOS ONE, 12(3), e0169687. https://doi.org/10.1371/journal.pone.0169687

[19] **Assmann, G., Köhm, M., Schuster, V., Behrens, F., Mössner, R., Magnolo, N., … Hüffmeier, U.** (2020). Genetic variants in FBLIM1 gene do not contribute to SAPHO syndrome and chronic recurrent multifocal osteomyelitis in typical patient groups. BMC Medical Genetics, 21(1). https://doi.org/10.1186/s12881-020-01037-7

[20] **Gulati, A., & Bagga, A.** (2009). Large vessel vasculitis. Pediatric Nephrology, 25(6), 1037–1048. https://doi.org/10.1007/s00467-009-1312-9

[21] **de Souza, A. W. S., & de Carvalho, J. F.** (2014). Diagnostic and classification criteria of Takayasu arteritis. Journal of Autoimmunity, 48–49, 79–83. https://doi.org/10.1016/j.jaut.2014.01.012

[22] **Kerr, G. S.** (1994). Takayasu Arteritis. Annals of Internal Medicine, 120(11), 919. https://doi.org/10.7326/0003-4819-120-11-199406010-00004

[23] **Url-2** <*https://www.hopkinsvasculitis.org/types-vasculitis/takayasus-arteritis/*>, date retrieved 28.05.2020.

[24] **Maksimowicz-McKinnon, K., & Hoffman, G. S.** (2007). Takayasu Arteritis: What is the Long-Term Prognosis? Rheumatic Disease Clinics of North America, 33(4), 777–786. https://doi.org/10.1016/j.rdc.2007.07.014

[25] **Numano, F., Okawara, M., Inomata, H., & Kobayashi, Y.** (2000). Takayasu's arteritis. The Lancet, 356(9234), 1023–1025. https://doi.org/10.1016/s0140-6736(00)02701-x

[26] **Arend, W. P., Michel, B. A., Bloch, D. A., Hunder, G. G., Calabrese, L. H., Edworthy, S. M., Fauci, A. S., Leavitt, R. Y., Lie, J. T., Lightfoot, R. W., Jr, Masi, A. T., McShane, D. J., Mills, J. A., Stevens, M. B., Wallace, S. L., & Zvaifler, N. J.** (2010). The American College of Rheumatology 1990 criteria for the classification of takayasu arteritis. Arthritis & Rheumatism, 33(8), 1129–1134. https://doi.org/10.1002/art.1780330811

[27] **Subramanyan, R., Joy, J., & Balakrishnan, K. G.** (1989). Natural history of aortoarteritis (Takayasu's disease). Circulation, 80(3), 429–437. https://doi.org/10.1161/01.cir.80.3.429

[28] **Watts, R., Al-Taiar, A., Mooney, J., Scott, D., & MacGregor, A.** (2009). The epidemiology of Takayasu arteritis in the UK. Rheumatology, 48(8), 1008–1011. https://doi.org/10.1093/rheumatology/kep153

[29] **Gudbrandsson, B., Molberg, Ø., Garen, T., & Palm, Ø.** (2016). Prevalence, Incidence, and Disease Characteristics of Takayasu Arteritis by Ethnic Background: Data From a Large, Population-Based Cohort Resident in Southern Norway. Arthritis Care & Research, 69(2), 278–285. https://doi.org/10.1002/acr.22931

[30] **Toshihiko, N**. (1996). Current status of large and small vessel vasculitis in Japan. International Journal of Cardiology, 54, S91–S98. https://doi.org/10.1016/s0167-5273(96)88777-8

[31] **Merih Birlik, Yasin Kücükyavas, Kenan Aksu, Dilek Solmaz, Gercek Can, Ali Taylan, Servet Akar, Ismail Sari, Gokhan Keser, Fatos Onen, Nurullah Akkoc** (2016). Epidemiology of Takayasu's Arteritis in Turkey, Clin Exp Rheumatol, 34(3 Suppl 97), S33-9.

[32] **Saritas, F., Donmez, S., Direskeneli, H., & Pamuk, O. N.** (2016). The epidemiology of Takayasu arteritis: a hospital-based study from northwestern part of Turkey. Rheumatology International, 36(7), 911–916. https://doi.org/10.1007/s00296-016-3445-z

[33] **Onen, F., & Akkoc, N.** (2017). Epidemiology of Takayasu arteritis. La Presse Médicale, 46(7–8), e197–e203. https://doi.org/10.1016/j.lpm.2017.05.034

[34] **Sharma, B. ., & Jain, S.** (1998). A possible role of sex in determining distribution of lesions in Takayasu Arteritis. International Journal of Cardiology, 66, S81–S84. https://doi.org/10.1016/s0167-5273(98)00154-5

[35] **Rosenthal, T., Morag, B., & Itzchak, Y.** (1992). Takayasu arteritis in Israel. Heart and Vessels, 7(S1), 44–47. https://doi.org/10.1007/bf01744543

[36] **Keser, G., Aksu, K., & Direskeneli, H.** (2018), Takayasu arteritis: an update, Turkish Journal of Medical Sciences, 48, 681-697. https://doi.org/doi:10.3906/sag-1804-136

[37] **Richards, B. L., March, L., & Gabriel, S. E.** (2010). Epidemiology of large-vessel vasculidities. Best Practice & Research Clinical Rheumatology, 24(6), 871–883. https://doi.org/10.1016/j.berh.2010.10.008

[38] **Numano, F., Isohisa, I., Kishi, U., Arita, M., & Maezawa, H.** (1978). Takayasu's disease in twin sisters. Possible genetic factors. Circulation, 58(1), 173–177. https://doi.org/10.1161/01.cir.58.1.173

[39] **Heo, J.-H., & Kim, M.** (2009). Familial Takayasu's arteritis in female siblings. Rheumatology International, 31(6), 815–818. https://doi.org/10.1007/s00296-009-1234-7

[40] **Kodama, K., Kida, O., Morotomi, Y., & Tanaka, K.** (1986). Male siblings

with Takayasu's arteritis suggest genetic etiology. Heart and Vessels, 2(1), 51–54. https://doi.org/10.1007/bf02060246

[41] **Deniz, A., Yildiz, F., Aktas, H., Berk, G. I., Erken, E., & Demirtas, M.** (2012). Familial Takayasu arteritis in a mother and daughter: a report of two cases. Herz, 38(1), 93–96. https://doi.org/10.1007/s00059-012-3653-6

[42] **Shiina, T., Hosomichi, K., Inoko, H., & Kulski, J.K.** (2009), The HLA genomic loci map: expression, interaction, diversity and disease, Journal of Human Genetics volume 54, 15–39, https://doi.org/10.1038/jhg.2008.5

[43] **Isohisa, I., Numano, F., Maezawa, H., & Sasazuki, T.** (1978). HLA-Bw52 in Takayasu Disease. Tissue Antigens, 12(4), 246–248. https://doi.org/10.1111/j.1399-0039.1978.tb01332.x

[44] **Mehra, N. K., Jaini, R., Balamurugan, A., Kanga, U., Prabhakaran, D., Jain, S., Talwar, K. K., & Sharma, B. K.** (1998). Immunogenetic analysis of Takayasu arteritis in Indian patients. International Journal of Cardiology, 66, S127–S132. https://doi.org/10.1016/s0167-5273(98)00160-0

[45] **Park, M. H., & Park, Y. B.** (1992). HLA Typing of Takayasu arteritis in Korea. Heart and Vessels, 7(S1), 81–84. https://doi.org/10.1007/bf01744549

[46] **Sahin, Z., Bıcakcıgil, M., Aksu, K., Kamali, S., Akar, S., Onen, F., Karadag, O., Ozbalkan, Z., Ates, A., Ozer, H. T., Yilmaz, V., Seyahi, E., Ozturk, M. A., Cefle, A., Cobankara, V., Onat, A. M., Tunc, E., Düzgün, N., … Aydin, S. Z.** (2012). Takayasu's arteritis is associated with HLA-B*52, but not with HLA-B*51, in Turkey. Arthritis Research & Therapy, 14(1), R27. https://doi.org/10.1186/ar3730

[47] **Origuchi, T., Fukui, S., Umeda, M., Nishino, A., Nakashima, Y., Koga, T., Kawashiri, S., Iwamoto, N., Ichinose, K., Tamai, M., Nakamura, H., & Kawakami, A.** (2016). The Severity of Takayasu Arteritis Is Associated with the HLA-B52 Allele in Japanese Patients. The Tohoku Journal of Experimental Medicine, 239(1), 67–72. https://doi.org/10.1620/tjem.239.67

[48] **Numano, F., Ohta, N., & Sasazuki, T.** (1982). HLA and clinical manifestations in Takayasu disease. Japanese Circulation Journal, 46(2), 184–189. https://doi.org/10.1253/jcj.46.184

[49] **Kasuya, K., Hashimoto, Y., & Numano, F.** (1992). Left ventricular dysfunction and HLA Bw52 antigen in Takayasu arteritis. Heart and Vessels, 7(S1), 116–119. https://doi.org/10.1007/bf01744556

[50] **Sattar, M. A., White, A. G., Eklof, B., & Fenech, F. F.** (1985). Takayasu's disease in Arabs. Postgraduate Medical Journal, 61(715), 387–390. https://doi.org/10.1136/pgmj.61.715.387

[51] **Saruhan-Direskeneli, G., Hughes, T., Aksu, K., Keser, G., Coit, P., Aydin, S. Z., … Sawalha, A. H.** (2013). Identification of Multiple Genetic Susceptibility Loci in Takayasu Arteritis. The American Journal of Human Genetics, 93(2), 298–305. https://doi.org/10.1016/j.ajhg.2013.05.026

[52] **Zhang, C. 'e, Wang, W., Zhang, H. 'e, Wei, L., & Guo, S.** (2016). Association of FCGR2A rs1801274 polymorphism with susceptibility to autoimmune diseases: A meta-analysis. Oncotarget, 7(26), 39436–39443. https://doi.org/10.18632/oncotarget.9831

[53] **Terao, C., Yoshifuji, H., Kimura, A., Matsumura, T., Ohmura, K., Takahashi, M., Shimizu, M., Kawaguchi, T., Chen, Z., Naruse, T. K., Sato-Otsubo, A., Ebana, Y., Maejima, Y., Kinoshita, H., Murakami, K., Kawabata, D., Wada, Y., Narita, I., Tazaki, J., Matsuda, F**. (2013). Two Susceptibility Loci to Takayasu Arteritis Reveal a Synergistic Role of the IL12B and HLA-B Regions in a Japanese Population. The American Journal of Human Genetics, 93(2), 289–297. https://doi.org/10.1016/j.ajhg.2013.05.024

[54] **Renauer, P. A., Saruhan-Direskeneli, G., Coit, P., Adler, A., Aksu, K., Keser, G., Alibaz-Oner, F., Aydin, S. Z., Kamali, S., Inanc, M., Carette, S., Cuthbertson, D., Hoffman, G. S., Akar, S., Onen, F., Akkoc, N., Khalidi, N. A., Koening, C., Karadag, O., … Sawalha, A. H.** (2015). Identification of Susceptibility Loci in IL6,RPS9/LILRB3, and an Intergenic Locus on Chromosome 21q22 in Takayasu Arteritis in a Genome-Wide Association Study. Arthritis & Rheumatology, 67(5), 1361–1368. https://doi.org/10.1002/art.39035

[55] **Tamura, N., Maejima, Y., Matsumura, T., Vega, R. B., Amiya, E., Ito, Y., Shiheido-Watanabe, Y., Ashikaga, T., Komuro, I., Kelly, D. P., Hirao, K., & Isobe, M.** (2018). Single-Nucleotide Polymorphism of the MLX Gene Is Associated With Takayasu Arteritis. Circulation: Genomic and Precision Medicine, 11(10). https://doi.org/10.1161/circgen.118.002296

[56] **Park, M. C., Lee, S. W., Park, Y. B., & Lee, S. K.** (2005). Serum cytokine profiles and their correlations with disease activity in Takayasu's arteritis. Rheumatology, 45(5), 545–548. https://doi.org/10.1093/rheumatology/kei266

[57] **Alibaz-Oner, F., Yentür, S. P., Saruhan-Direskeneli, G., & Direskeneli, H.** (2015). Serum cytokine profiles in Takayasu's arteritis: search for biomarkers. Clinical and experimental rheumatology, 33(2 Suppl 89).

[58] **Nishimoto, N., Nakahara, H., Yoshio-Hoshino, N., & Mima, T.** (2008). Successful treatment of a patient with takayasu arteritis using a humanized anti-interleukin-6 receptor antibody. Arthritis & Rheumatism, 58(4), 1197–1200. https://doi.org/10.1002/art.23373

[59] **Loricera, J., Blanco, R., Castañeda, S., Humbría, A., Ortego-Centeno, N., Narváez, J., Mata, C., Melchor, S., Aurrecoechea, E., Calvo-Alén, J., Lluch, P., Moll, C., Mínguez, M., Herrero-Beaumont, G., Bravo, B., Rubio, E., Freire, M., Peiró, E., González-Vela, C., Rueda-Gotor, J., … González-Gay, M. Á.** (2014). Tocilizumab in refractory aortitis: study on 16 patients and literature review. Clinical and experimental rheumatology, 32(3 Suppl 82), S79–S89.

[60] **Schultz, D. R., & Arnold, P. I.** (1993). Heat shock (stress) proteins and autoimmunity in rheumatic diseases. Seminars in Arthritis and Rheumatism, 22(6), 357–374. https://doi.org/10.1016/s0049-0172(05)80028-9

[61] **Lupi-Herrera, E., Sánchez-Torres, G., Marcushamer, J., Mispireta, J., Horwitz, S., & Espino Vela, J.** (1977). Takayasu's arteritis. Clinical study of 107 cases. American Heart Journal, 93(1), 94–103. https://doi.org/10.1016/s0002-8703(77)80178-6

[62] **Clemente, G., Hilário, M. O., Len, C., Silva, C. A., Sallum, A. M., Campos, L. M., … Terreri, M. T.** (2016). Brazilian multicenter study of 71 patients with juvenile-onset Takayasu's arteritis: clinical and angiographic features. Revista Brasileira de Reumatologia (English

Edition), 56(2), 145–151. https://doi.org/10.1016/j.rbre.2016.01.004
[63] **Giordano J. M.** (1995). Takayasu's disease and temporal arteritis. Seminars in vascular surgery, 8(4), 335–341.
[64] **Seko, Y.** (2000). Takayasu Arteritis. Insights into Immunopathology. Japanese Heart Journal, 41(1), 15–26. https://doi.org/10.1536/jhj.41.15
[65] **Hotchi, M.** (1992). Pathological studies on Takayasu arteritis. Heart and Vessels, 7(S1), 11–17. https://doi.org/10.1007/bf01744538
[66] **Smyth, M. J., Cretney, E., Kelly, J. M., Westwood, J. A., Street, S. E. A., Yagita, H., … Hayakawa, Y.** (2005). Activation of NK cell cytotoxicity. Molecular Immunology, 42(4), 501–510. https://doi.org/10.1016/j.molimm.2004.07.034
[67] **Inder, S.** (2000). Immunophenotypic analysis of the aortic wall in Takayasu's arteritis: involvement of lymphocytes, dendritic cells and granulocytes in immuno-inflammatory reactions. Cardiovascular Surgery, 8(2), 141–148. https://doi.org/10.1016/s0967-2109(99)00100-3
[68] **Keser, G., Aksu, K., & Direskeneli, H.** (2017). Discrepancies between vascular and systemic inflammation in large vessel vasculitis: an important problem revisited. Rheumatology, 57(5), 784–790. https://doi.org/10.1093/rheumatology/kex333
[69] **Deng, J., Younge, B. R., Olshen, R. A., Goronzy, J. J., & Weyand, C. M.** (2010). Th17 and Th1 T-Cell Responses in Giant Cell Arteritis. Circulation, 121(7), 906–915. https://doi.org/10.1161/circulationaha.109.872903
[70] **Hoyer, B. F., Mumtaz, I. M., Loddenkemper, K., Bruns, A., Sengler, C., Hermann, K.-G., Maza, S., Keitzer, R., Burmester, G.-R., Buttgereit, F., Radbruch, A., & Hiepe, F.** (2011). Takayasu arteritis is characterised by disturbances of B cell homeostasis and responds to B cell depletion therapy with rituximab. Annals of the Rheumatic Diseases, 71(1), 75–79. https://doi.org/10.1136/ard.2011.153007
[71] **Adams, J.** (2008) DNA sequencing technologies. Nature Education 1(1):193
[72] **Levy, S. E., & Myers, R. M.** (2016). Advancements in Next-Generation Sequencing. Annual Review of Genomics and Human Genetics, 17(1), 95–115. https://doi.org/10.1146/annurev-genom-083115-022413
[73] **Soto, J., Rodriguez-Antolin, C., Vallespín, E., de Castro Carpeño, J., & Ibanez de Caceres, I.** (2016). The impact of next-generation sequencing on the DNA methylation–based translational cancer research. Translational Research, 169, 1–18.e1. https://doi.org/10.1016/j.trsl.2015.11.003
[74] **Soto, J., Rodriguez-Antolin, C., Vallespín, E., de Castro Carpeño, J., & Ibanez de Caceres, I.** (2016). The impact of next-generation sequencing on the DNA methylation–based translational cancer research.Translational Research, 169, 1–18.e1. https://doi.org/10.1016/j.trsl.2015.11.003
[75] **Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y.** (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics, 13(1), 341. https://doi.org/10.1186/1471-2164-13-341
[76] **Url-3** <*https://ghr.nlm.nih.gov/primer/genomicresearch/snp*>, date retrieved

28.05.2020.

[77] **Botstein, D., & Risch, N.** (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics, 33(S3), 228–237. https://doi.org/10.1038/ng1090

[78] **Mao, X., Young, B. D., & Lu, Y. J.** (2007). The Application of Single Nucleotide Polymorphism Microarrays in Cancer Research. Current Genomics, 8(4), 219–228. https://doi.org/10.2174/138920207781386924

[79] **Url-4** <*https://www.illumina.com/products/by-type/microarray-kits.html*>, date retrieved 28.05.2020.

[80] **Syvänen A. C**. (2005). Toward genome-wide SNP genotyping. Nature genetics, 37 Suppl, S5–S10. https://doi.org/10.1038/ng1558

[81] **Url-5 <**https://www.broadinstitute.org/blog/what-exome-sequencing>, date retrieved 28.05.2020.

[82] **Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., & Lifton, R. P**. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences, 106(45), 19096–19101. https://doi.org/10.1073/pnas.0910672106

[83] **Rabbani, B., Mahdieh, N., Hosomichi, K., Nakaoka, H., & Inoue, I.** (2012). Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. Journal of Human Genetics, 57(10), 621–632. https://doi.org/10.1038/jhg.2012.91

[84] **Biesecker, L. G., Shianna, K. V., & Mullikin, J. C.** (2011). Exome sequencing: the expert view. Genome Biology, 12(9), 128. https://doi.org/10.1186/gb-2011-12-9-128

[85] **Li, H., & Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

[86] **Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., … Banks, E.** (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. https://doi.org/10.1101/201178

[87] **Bohannan, Z. S., & Mitrofanova, A.** (2019). Calling Variants in the Clinic: Informed Variant Calling Decisions Based on Biological, Clinical, and Laboratory Variables. Computational and Structural Biotechnology Journal, 17, 561–569. https://doi.org/10.1016/j.csbj.2019.04.002

[88] **Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J.** (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics, 12(11), 745–755. https://doi.org/10.1038/nrg3031

[89] **Frésard, L., & Montgomery, S. B.** (2018). Diagnosing rare diseases after the exome. Molecular Case Studies, 4(6), a003392. https://doi.org/10.1101/mcs.a003392

[90] **Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A.** (2012). Disease gene identification strategies for exome sequencing. European Journal of Human Genetics, 20(5), 490–497. https://doi.org/10.1038/ejhg.2011.258

[91] **Worthey, E. A.** (2017). Analysis and Annotation of Whole-Genome or

Whole-Exome Sequencing Derived Variants for Clinical Diagnosis. Current Protocols in Human Genetics, 95(1), 9.24.1-9.24.28. https://doi.org/10.1002/cphg.49

[92] **Lobo, I. & Shaw, K.** (2008) Thomas Hunt Morgan, genetic recombination, and gene mapping. Nature Education 1(1):205

[93] **Pulst, S. M.** (1999). Genetic Linkage Analysis. Archives of Neurology, 56(6), 667. https://doi.org/10.1001/archneur.56.6.667

[94] **Nyholt, D. R.** (2000). All LODs Are Not Created Equal**A Microsoft Excel spreadsheet, for performing easy calculations of P values for the LOD scores described in this review, is available on request from the author. The American Journal of Human Genetics, 67(2), 282–288. https://doi.org/10.1086/303029

[95] **Morton N. E.** (1955). Sequential tests for the detection of linkage. American journal of human genetics, 7(3), 277–318.

[96] **Kruglyak, L., Daly, M. J., Reeve-Daly, M. P., & Lander, E. S.** (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. American journal of human genetics, 58(6), 1347–1363.

[97] **Kruglyak, L., Daly, M. J., & Lander, E. S.** (1995). Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. American journal of human genetics, 56(2), 519–527.

[98] **Smith, K. R., Bromhead, C. J., Hildebrand, M. S., Shearer, A. E., Lockhart, P. J., Najmabadi, H., Leventer, R. J., McGillivray, G., Amor, D. J., Smith, R. J., & Bahlo, M.** (2011). Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biology, 12(9), R85. https://doi.org/10.1186/gb-2011-12-9-r85

[99] **Ott, J., Wang, J., & Leal, S. M.** (2015). Genetic linkage analysis in the age of whole-genome sequencing. Nature Reviews Genetics, 16(5), 275–284. https://doi.org/10.1038/nrg3908

[100] **Génin, E., & Todorov, A. A.** (2007). Homozygosity Mapping. In Encyclopedia of Life Sciences. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0005407.pub2

[101] **Thompson, E. A.** (2013). Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics, 194(2), 301–326. https://doi.org/10.1534/genetics.112.148825

[102] **Lindner, T. H., & Hoffmann, K.** (2004). easyLINKAGE: a PERL script for easy and automated two-/multi-point linkage analyses. Bioinformatics, 21(3), 405–407. https://doi.org/10.1093/bioinformatics/bti009

[103] **McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A.** (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

[104] **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC** (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

[105] **Ng, P. C.** (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research, 31(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

[106] **Adzhubei, I., Jordan, D. M., & Sunyaev, S. R.** (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current Protocols in Human Genetics, 76(1), 7.20.1-7.20.41. https://doi.org/10.1002/0471142905.hg0720s76

[107] **Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J.** (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics, 46(3), 310–315. https://doi.org/10.1038/ng.2892

[108] **Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., … Sieh, W.** (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. The American Journal of Human Genetics, 99(4), 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016

APPENDICES

APPENDIX A

Centrifuge	Thermo Scientific MICROCL 17 Allegra 25R Centrifuge Beckman Coulter Beckman Coulter Microfuge
Centrifuge Tubes	Greiner Bio-One Axygen
EDTA Blood Tube	VACUTEST-ARZERGRANDE
Electrophoresis Systems	BIO-RAD Mini-PROTEAN Tetra Cell
Examination Gloves	Beybi
Freezers	Arçelik (-20°C) Nuve (-80°C)
Heat Block	Dri-Block DB-2D-Bibby Scientific Ltd
Ice Machine	Scotsman AF 10
Magnetic Stirrer	Dragon Lab MS-HS
Medical X-ray Processor	Kodak
Microwave	Arçelik MD582
Pipette	Axygen Thermo Scientific
Pipette Tips	Axygen Thermo Scientific
pH meter	Mettler Toledo MP220
Refrigerator	Vestel (+4°C)
Rotator	Stuart
Shaker	Heidolph Doumax 1030
Spectrophotometer	NANODROP – Thermo Scientific
Thermal Cycler	Thermo
Vortex	Herdolph Reax Top

APPENDIX B

Ethanol	Sigma-Aldrich
NaOH	Riedel-de Haen / Qiagen / Roche
Acetic acid	Riedel-de Haen
Agarose	Invitrogen
Glycine	Applichem
$MgCl_2$	Thermo Scientific
$ZnCl_2$	Riedel-de Haen
$CaCl_2$	Merck
HCl	Merck
Isopropanol	AppliChem
DNA ladder	Thermo Scientific
dNTP	Thermo Scientific
Nuclease-free Water	Medifar
DreamTaq polymerase	Thermo Scientific
DreamTaq Green Buffer	Thermo Scientific

APPENDIX C

TAE Buffer:

Tris Base 242 g
Acetic acid 57.1 mL
$Na_2EDTA.2H_2O$ 37.2 g
distilled H_2O up to 1 L
pH 8.0 – 8.5

TKM2 Buffer:

Tris HCl 605 mg
pH 7.6
KCl 370 mg
EDTA 370 mg
$MgCl_2$ 6 g
NaCl 2.335 g
distilled H_2O up to 500 mL

CURRICULUM VITAE

Name Surname **: Derya YAVUZ**

Place and Date of Birth **: Istanbul / 14.06.1991**

E-Mail **: deryayayavuz@gmail.com**

EDUCATION

- **B.Sc.** **:** 2017, Istanbul Technical University (ITU), Faculty Science and Letters, Department of Molecular Biology and Genetics

PROFESSIONAL EXPERIENCE AND REWARDS:

- September 2016 - January 2017, Intern, Istanbul Technical University (ITU), Human Genetics Laboratory Supervisor: Prof. Dr. Eda TAHIR TURANLI
- June 2016 - August 2016, Intern, Middle East Technical University (METU), ER Laboratory, Ankara / TURKEY Supervisor: Prof. Dr. Mesut MUYAN
- January 2016 - May 2016, Intern, Boğaziçi University, Laboratory of Genome Regulation, Istanbul /TURKEY Supervisor: Assoc. Prof. Dr. N.C Tolga EMRE

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Yavuz D., Yüksel S. &, Tahir Turanlı E. 1/04/2019. Whole genome linkage and exome analysis in a CRMO family. International Society of Systemic Auto Inflammatory Diseases (ISSAID), 10th International Congress 2019, Genoa/ITALY.