

GÖKÇEN ŞAHİN

← Adınızı soyadınızı giriniz

İSTANBUL ÜNİVERSİTESİ SAĞ. BİL. ENST.

Tez kabul edildikten sonra yapılan **sabit ciltte sırt yazısı** bu şablona göre yazılacak. Yazılar tek satır olacak
Cilt sırtı yazıların yönü yukarıdan aşağıya
(sol yandaki gibi) olacak .

YÜKSEK LİSANS

Tez, Yüksek Lisans'sa, YÜKSEK LİSANS TEZİ;
← Doktora ise DOKTORA TEZİ ifadesi kalacak

İSTANBUL-2020

← Tez Sınavının yapılacağı yılı yazınız

**T.C.
İSTANBUL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

(YÜKSEK LİSANS TEZİ)

**OKSİDATİF FOSFORİLASYON SİSTEMİ
HASTALIKLARINDA MİTOKONDRIYAL DNA
VARYANTLARININ VERİ MADENCİLİĞİ İLE
ARAŞTIRILMASI**

GÖKÇEN ŞAHİN

**DANIŞMAN
PROF. DR. ZEHRA OYA UYGUNER**

**GENETİK ANABİLİM DALI
GENETİK PROGRAMI**

İSTANBUL-2020

İTHAF

Annem ve babam, Serpil ve Nuri Şahin'e ithaf ediyorum.

TEŞEKKÜR

Yüksek lisans eğitimim süresince bilgi, tecrübe ve desteğiyle yolumu açan, tez çalışmamı ilgi duyduğum alanda yapmama imkan veren danışman hocam Sayın Prof. Dr. Oya Uyguner'e,

Bilgi ve tecrübelerine saygı duyduğum değerli hocalarım İÜ Tıbbi Genetik Ana Bilim Dalı Başkanı Sayın Prof. Dr. Seher Başaran ve Sayın Prof. Dr. Birsen Karaman'a

Çalışmam için olgu sağlayan ve desteğiyle yanımda olan Doç. Dr. Asuman Gedikbaşı'na,

Kendisinden öğrenmeyi, düşünmeyi, sormayı, sorgulamayı öğrendiğim, sayesinde akademik dünyaya hazırlandığım, çalışmamın her aşamasında bilgisini hiç esirgemeyen ve desteğini hep hissettiğim çok değerli abim Dr. Çağrı Güleç'e

Laboratuvar teknik bilgi ve becerilerimi kazandığım tüm İÜ Tıbbi Genetik Laboratuvarı çalışanlarına ve destekleri için İÜ İTF Tıbbi Genetik Kliniği çalışanlarına,

Sürecin son evresinde (iyi ki) hayatıma giren, hep yanımda olan Erdem Aybek'e,

Tecrübeleriyle yolumu aydınlatan, daima moral desteği sağlayan, eğitimimi sonuna kadar destekleyen kıymetli annem ve babam Serpil ve Nuri Şahin'e minnetle,

teşekkür ederim

Bu çalışma, İstanbul Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenmiştir. Proje No: TYL-2019-33447

İÇİNDEKİLER

TEZ ONAYI.....	İİ
BEYAN.....	İİİ
İTHAF.....	İV
TEŞEKKÜR.....	V
İÇİNDEKİLER.....	VI
TABLolar LİSTESİ.....	İX
ŞEKİLLER LİSTESİ.....	Xİİ
SEMBOLLER / KISALTMALAR LİSTESİ.....	XIV
ÖZET.....	XVII
ABSTRACT.....	XVIII
1. GİRİŞ VE AMAÇ.....	1
2. GENEL BİLGİLER.....	3
2.1. Mitokondri ve Enerji.....	3
2.1.1. Mitokondrinin Evolüsyonu.....	3
2.1.2. Mitokondriyal Devrim.....	5
2.1.2.1. Mitokondrinin Morfolojisi.....	5
2.1.2.2. Mitokondrinin Hücre İçindeki Yerleşimi.....	7
2.1.2.3. Mitokondrinin Devimsel Düzenlenimi ve Kalite Kontrol.....	7
2.1.3. Mitokondrinin Yapı ve Organizasyon Özellikleri.....	8
2.1.4. Mitokondriyal DNA (mtDNA).....	10
2.1.4.1. Homoplazmi ve Heteroplazmi Kavramları.....	13
2.1.4.2. MtDNA Dizileme Çalışmaları.....	13
2.1.5. Mitokondrinin Fonksiyonları.....	14
2.1.5.1. Enerji Üretimi.....	14
2.1.5.2. Kalsiyum Dengesi.....	17
2.1.5.3. Oksijen Radikallerinin Kontrolü.....	17
2.1.5.4. Apoptoz.....	18
2.1.5.5. Mitokondriyal Metabolitlerin Epigenetik Mekanizmalar Üzerine Etkisi.....	18
2.1.6. Endoplazmik Retikulum ve Mitokondri Arasındaki İlişkisi.....	19
2.1.7. Mitokondriyal Hastalıklar.....	20

2.1.7.1. m.3243A>G Patojenik Varyantı ve MELAS.....	22
2.2. Veri Madenciliği ve Makine Öğrenmesi.....	24
3. GEREÇ VE YÖNTEM.....	26
3.1. Laboratuvar Çalışmaları.....	26
3.1.1. Örnek Grubu.....	26
3.1.2. Örnek Alımı.....	26
3.1.3. DNA İzolasyonu.....	26
3.1.4. DNA Yoğunluk ve Kalite Kontrolü.....	26
3.1.5. Primer Tasarımı.....	27
3.1.6. MtDNA Dizilerinin PZR ile Çoğaltılması.....	27
3.1.7. PZR Ürünlerinin Saflaştırması.....	28
3.1.8. Agarol Jel Elektrophorezi ile Görüntüleme.....	28
3.1.9. Yeni Nesil Dizileme İşlemleri.....	29
3.1.9.1. DNA Konsantrasyonun Hesaplanması ve Örnek Hazırlığı.....	30
3.1.9.2. Fragmentasyon.....	30
3.1.9.3. Adaptör Bağlama ve Barkodlama.....	31
3.1.9.4. Hazırlanan Kütüphanede Boyut-Seçim İşlemi.....	32
3.1.9.5. Kütüphanenin Çoğaltılması ve Saflaştırılması.....	34
3.1.9.6. Oluşturulan Barkodlanmış Kütüphanenin Değerlendirmesi.....	35
3.1.9.7. Otomatize Şablon Hazırlama, Çip Yükleme ve Dizileme.....	36
3.1.10. Dizileme Verilerinin Analizi.....	37
3.1.10.1. Akış Şeması Oluşturma.....	37
3.1.10.2. Ön Analizi Başlatma.....	39
3.1.10.3. Varyant Dosyalarına Erişim ve Düzenleme.....	39
3.1.10.4. Analiz Süreci.....	41
3.2. Veri Madenciliği ve Makine Öğrenmesi Çalışmaları.....	41
3.2.1. Kullanılacak Python Paketleri.....	43
4. BULGULAR.....	45
4.1. Laboratuvar Çalışma Sonuçları.....	45
4.1.1. Örnek Grubu.....	45
4.1.2. PZR Sonuçları.....	45
4.1.3. MtDNA Yeni Nesil Dizileme Sonuçları.....	46
4.1.3.1. Gönüllülerde Saptanan Varyantların Sayısal Dökümleri.....	46

4.1.3.2. Gönüllülerde Saptanan Tüm Varyant Tipleri İçin Sonuçlar.....	48
4.1.3.3. Gönüllülerde Saptanan SNV Tipi Varyantlar İçin Sonuçlar.....	52
4.1.3.4. Gönüllülerde Saptanan Indel Tipi Varyantlar İçin Sonuçlar.....	55
4.1.3.5. Üç Doku Çalışması için Varyant Sonuçları.....	59
4.2. Veri Madenciliği ve Makine Öğrenmesi Çalışmalarının Sonuçları.....	65
4.2.1. Veri Yükleme ve Ön-İşleme Uygulamaları.....	67
4.2.2. Veri Setinde Yapılan Testler ve Sonuçları.....	86
4.2.2.1. Çoklu Eşdoğrusallık (<i>Multicollinearity</i>) Testi ve Sonuçları.....	86
4.2.2.2. Uç Değer (<i>Outliers</i>) Testi ve Sonucu.....	88
4.2.3. Regresyon Modeli Uygulanması.....	89
4.2.3.1. Regresyon Modelinin Değerlendirilmesi, İyileştirilmesi ve Sonuçları.....	91
4.2.3.2. Diğer Model Denemeleri ve Sonuçları.....	106
4.2.3.3. Modelin Kaydedilmesi.....	108
4.2.4. Veri Setinin Hazırlanması.....	109
4.2.4.1. Regresyon Modelinin Uygulanması.....	111
5. TARTIŞMA.....	117
5.1. Laboratuvar Sonuçlarının Değerlendirmesi.....	118
5.2. Veri Madenciliği ve Makine Öğrenmesi Sonuçlarının Değerlendirilmesi.....	121
KAYNAKLAR.....	126
FORMLAR.....	141
ETİK KURUL KARARI.....	144
İNTİHAL RAPORU İLK SAYFASI.....	148
ÖZGEÇMİŞ.....	149

TABLOLAR LİSTESİ

Table 3-1: PZR malzemeleri ve son hacim konsantrasyonları.....	27
Table 3-2: PZR protokolü.....	28
Table 3-3: Ion Xpress™ Plus Fragment Library Kit Malzemeleri ve uygulanan miktarlar	30
Table 3-4: Ion Xpress™ Plus Fragment Library Kit ve Barcode Adapters Kit Malzemeleri ve miktarları.....	31
Table 3-5: Bağlanma (Ligasyon) PZR protokolü.....	31
Table 3-6: Boyut-seçim işlemi için gerekli malzemeler.....	33
Table 3-7: Kütüphane amplifikasyonu malzemeleri ve kullanılan miktarlar.....	34
Table 3-8: 100-500 baz-okuma kütüphanesi için PZR protokolü.....	34
Table 3-9: Seri dilüsyonla hazırlanmış E.Coli DH10B Kontrol Kütüphanesi standartları dilüsyon ve konsantrasyon değerleri.....	36
Table 3-10: Ion Library TaqMan® Quantitation Kit malzemeleri ve miktarlar.....	36
Table 3-11: Gerçek-zamanlı PZR protokolü.....	36
Table 3-12: Ion Reporter Yazılımı'nda akış şemasının oluşturulduğu bölümde bulunan Parameters sayfası, Variant Finding/Main başlığı altında yapılan değişiklikler.....	38
Table 3-13: Ion Reporter Yazılımı'nda akış şemasının oluşturulduğu bölümde bulunan Parameters sayfası, Variant Finding/Advanced başlığı altında yapılan değişiklikler.....	39
Table 4-1: Kan ve kas dokusunda yapılan mtDNA YND çalışmasında elde edilen toplam varyantların tiplerine göre sayısal dökümü.....	47
Table 4-2: Kan ve kas dokusunda yapılan mtDNA YND çalışmasında her bir gönüllüde, doku temelinde ortak olan ve olmayan varyantların tiplerine göre sayısal dökümü.....	47
Table 4-3: Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen mtDNA YND çalışmasında saptanan toplam varyantların tiplerine göre sayısal dökümü.....	48
Table 4-4: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında elde edilen tüm değişimlerin (SNV ve Indel) frekans ortalamaları.....	49
Table 4-5: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında elde edilen tüm değişimlerin sayısal dökümü.....	49
Table 4-6: Kan ve kas örneklerinden edilen mtDNA'da yapılan YND çalışmasında ortak bulunan varyantların frekans ortalama sonuçları.....	50

Table 4-7: Kan ve kas örneklerinden edilen mtDNA’da yapılan YND çalışmasında ortak bulunan varyantların sayısal dökümü.....	51
Table 4-8: mtDNA YND çalışmasında yalnızca bir dokuda görülen, buna karşın diğer dokuda saptanmayan varyantların sayısal dökümü.....	51
Table 4-9: Kan ve kas örneklerine ait mtDNA YND çalışmasıyla saptanan tüm nokta değişimlerin (SNV) frekans ortalama sonuçları.....	52
Table 4-10: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında saptanan tüm nokta değişimlerin (SNV) sayısal dökümü.....	53
Table 4-11: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan tüm nokta değişimlerin (SNV) frekans ortalama sonuçları.....	54
Table 4-12: Kan ve kas örneklerinden edilen mtDNA’da yapılan YND çalışmasında ortak bulunan nokta değişimlerin (SNV) sayısal dökümü.....	54
Table 4-13: Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen buna karşın diğer dokusunda saptanmayan eş nokta değişimlerin (SNV) sayısal dökümü.....	55
Table 4-14: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında saptanan tüm indel tipi değişimlerin frekans ortalama sonuçları.....	56
Table 4-15: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında saptanan tüm indel tipi değişimlerin sayısal dökümü.....	56
Table 4-16: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan indel tipi değişimlerin frekans ortalama sonuçları.....	57
Table 4-17: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan indel tipi değişimlerin sayısal dökümü.....	58
Table 4-18: Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen buna karşın diğer dokusunda saptanmayan eş indel tipi değişimlerin sayısal dökümü.....	59
Table 4-19: Gönüllü 2 (G2)’nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre frekans ortalama sonuçları.....	60
Table 4-20: Gönüllü 2 (G2)’nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre sayısal dökümü ve oranları.....	61
Table 4-21: Gönüllü-2 (G2)’nin kan, kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri.....	62

Table 4-22: Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre ortak varyantların sayısal dökümü ve oranları.....	62
Table 4-23: Gönüllü-2'nin kan ve kas dokularında saptanmış ortak varyantların frekans ortalama sonuçları.....	63
Table 4-24: Gönüllü-2'nin kan ve idrar dokularında saptanmış ortak varyantların frekans ortalama sonuçları.....	63
Table 4-25: Gönüllü-2'nin kas ve idrar dokularında saptanmış ortak varyantların frekans ortalama sonuçları.....	64
Table 4-26: Gönüllü 2 (G2)'nin yalnızca bir dokusunda görülen buna karşın diğer dokularda saptanmayan varyantların tiplerine göre sayısal dökümü ve oranları.....	64
Table 4-27: df veri seti sütunlarındaki kayıp veri miktarları ve yüzdeleri.....	70
Table 4-28: df veri setinden veri sayısı yetersizliği sebebiyle çıkartılan sütunlar.....	72
Table 4-29: df veri setinde yetersiz veri sayısı sebebiyle birleştirilen sütunlar.....	72
Table 4-30: Tablo 4-29'da belirtilen birleştirme işleminin sonucu.....	72
Table 4-31: df veri setinde laktik_asidoz ve dağınık_kırmızı_lifler sütunlarında bulunan kayıp veri miktarları.....	74
Table 4-32: laktik_asidoz ve dağınık_kırmızı_lifler sütunlarındaki farklı durumlar için iskelet kası heterolazmi oranlarının istatistiksel dökümü.....	76
Table 4-33: laktik_asidoz ve dağınık_kırmızı_lifler sütunlarındaki farklı durumlar için kan heterolazmi oranlarının istatistiksel dökümü.....	77
Table 4-34: kas_df veri setindeki kayıp değerlerin miktarları ve yüzdeleri.....	79
Table 4-35: nan_kas_df veri setindeki kayıp değerlerin miktarları ve yüzdeleri.....	79
Table 4-36: Heteroskedastisite testleri ve sonuçları.....	93
Table 4-37: Otokorelasyon testi ve sonucu.....	94
Table 4-38: Model-1 için hata değerleri hesaplama sonuçları.....	97
Table 4-39: Diğer model denemelerinin sonuçları.....	108
Table 4-40: Model-2 için heteroskedastisite testleri ve sonuçları.....	112
Table 4-41: Model-2'nin otokorelasyon testi ve sonucu.....	112
Table 4-42: Model-2 için hata değerleri hesaplama sonuçları.....	114

ŞEKİLLER LİSTESİ

Şekil 3-1: NEB Quick-Load® Purple 1 kb DNA Ladder (N0550G) bant kılavuz görüntüsü.....	28
Şekil 3-2: Thermo Scientific™ Ion GeneStudio S5 System akış şeması.....	29
Şekil 3-3: E-Gel™ SizeSelect™ 2% agaroz jel elektroforezi görüntüsü.....	33
Şekil 3-4: Fragmanların referans çizgisine yaklaşması.....	33
Şekil 3-5: Fragmanların referans çizgisine denk gelmesi.....	34
Şekil 4-1: Dört ayrı parçada çoğaltılan mtDNA'nın kolon saflaştırma işleminden sonra %1'lik agaroz jel elektroforezi ile alınan görüntüsü.....	45
Şekil 4-2: Veri madenciliği ve makine öğrenmesi çalışmalarında izlenen adımlar.....	66
Şekil 4-3: Ham veri setindeki (df) sütunlar (öznitelik) (52 adet).....	67
Şekil 4-4: Ham veri setine (df) uygulanan Missingno görüntüleme kütüphanesi matrix fonksiyonu görüntüleme sonucu.....	69
Şekil 4-5: Ham veri setine (df) uygulanan Missingno görüntüleme kütüphanesi heatmap fonksiyonu görüntüleme sonucu.....	69
Şekil 4-6: df veri setindeki sütunlar (öznitelik) (44 adet).....	73
Şekil 4-7: df veri setine uygulanan Missingno kütüphanesi matrix fonksiyonu kayıp değer görüntüleme sonucu.....	74
Şekil 4-8: df veri setine uygulanan Missingno kütüphanesi heatmap fonksiyonuyla kayıp değerlerin aynı anda veri bulundurma durumunun görüntüleme sonucu. Kırmızı dikdörtgen içine alınmış değerler iskelet_kası ve kan sütunlarının, dağınık_kırmızı_lifler ve laktik_asidoz sütunları ile aynı anda veri bulundurma değerlerini vermektedir.....	75
Şekil 4-9: kas_df veri setindeki sütunlar (öznitelik) (27 adet).....	80
Şekil 4-10: nan_kas_df veri setindeki sütunlar (öznitelik) (26 adet).....	80
Şekil 4-11: Model-1 bağımsız değişkenlerinin katsayı değerlerinin güven aralığı (%95) hesaplama sonuçları (sütun düşürme işleminden önce).....	82
Şekil 4-12: P-değeri hesaplama sonuçları (sütun düşürme işleminden önce).....	83
Şekil 4-13: kas_df veri seti doluluk görüntüsü.....	84
Şekil 4-14: nan_kas_df veri seti doluluk görüntüsü.....	84
Şekil 4-15: kas_df veri setinin korelasyon matrisi görüntüleme sonucu.....	87
Şekil 4-16: kas_df veri setinin çoklu eşdoğrusallık testi sonuçları.....	88

Şekil 4-17: Model-1'in otokorelasyon testi görsel sonucu.....	94
Şekil 4-18: Artık değerlerin regresyon çizgisi etrafındaki dağılımının görsel sonucu (Model-1).....	95
Şekil 4-19: Model-1 bağımsız değişkenlerinin katsayı değerlerinin güven aralığı (%95) hesaplama sonuçları (sütun düşürme işleminden sonra).....	99
Şekil 4-20: P-değeri hesaplama sonuçları (Model-1) (sütun düşürme işleminden sonra)	101
Şekil 4-21: OLS fonksiyonu ile oluşturulan Model-1'in özet tablosu.....	102
Şekil 4-22: GLSAR fonksiyonu ile oluşturulan Model-1'in özet tablosu.....	103
Şekil 4-23: df veri setinin korelasyon matrisi görüntüleme sonucu.....	110
Şekil 4-24: df veri setinin çoklu eşdoğrusallık testi sonuçları.....	111
Şekil 4-25: Model-2'nin otokorelasyon testi görsel sonucu.....	113
Şekil 4-26: Artık değerlerin regresyon çizgisi etrafındaki dağılımının görsel sonucu (Model-2).....	113
Şekil 4-27: df veri seti ile oluşturulan çoklu-regresyon modeli, Model-2'nin sonuçları	115

SEMBOLLER / KISALTMALAR LİSTESİ

12S mtSSU	12S mitoribozomal küçük alt ünite (<i>12S mitoribosomal small subunit</i>)
16S mtLSU	16S mitoribozomal büyük alt ünite (<i>16S mitoribosomal large subunit</i>)
Acetyl-CoA	Asetil koenzim A (<i>acetyl coenzyme A</i>)
ADP	Adenozin difosfat (<i>adenosine diphosphate</i>)
AMP	Adenozin monofosfat (<i>adenosine monophosphate</i>)
ATP	Adenozin trifosfat (<i>adenosine triphosphate</i>)
bç	Baz-çifti (<i>bp, base pair</i>)
BLC2	BLC2 apoptoz regülatörü (<i>BCL2 apoptosis regulator</i>)
CAC	Sitrik asit döngüsü (<i>CAC, citric acid cycle</i>)
coef	Katsayı, eğim (<i>coefficient, slope</i>)
CPEO	Kronik progresif eksternal oftalmopleji (<i>chronic progressive external ophthalmoplegia</i>)
Ct	Eşik döngüsü sayısı (<i>cycle threshold</i>)
D-ilmeği	D-deplase ilmeği/D-ilmeği (<i>D-loop, Displaced heavy strand</i>)
df	Veri çerçevesi (<i>data frame</i>)
DMEM	Dulbecco tarafından modifiye edilmiş, Eagle'ın besi yeri (<i>Dulbecco's modified Eagle's medium</i>)
DNA	Deoksiribonükleik asit (<i>deoxyribonucleic acid</i>)
DNM1L	Dinamin 1 benzeri protein (<i>dynamin-1-like protein</i>)
dNTP	Deoksiribonükleosit trifosfat (<i>deoxyribonucleoside triphosphate</i>)
DRP1	Dinamin-ilşkili protein 1 (<i>dynamin-related protein 1</i>)
EDTA	Etilendiamin tetraasetik asit (<i>ethylenediaminetetraacetic acid</i>)
ER	Endoplazmik retikulum
ETZ	Elektron transfer zinciri (<i>ETC, electron transport chain</i>)
FADH2	Redüksüyona uğramış flavin adenin dinükleotid (<i>reduced flavin adenine dinucleotide</i>)
FIS1	Mitokondriyal fizyon 1 proteini (<i>Mitochondrial fission 1 protein</i>)
FSGS	Fokal segmental glomerülo skleroz (<i>focal segmental glomerulosclerosis</i>)
GLS	Genelleştirilmiş en küçük kareler (<i>generalized least squares</i>)
GLSAR	Otoregresif kovaryans yapısı ile genelleştirilmiş en küçük kareler (<i>generalized least squares with AR (autoregressive) covariance structure</i>)

hg19	<i>Homo sapiens (human) genome assembly GRCh37</i>
HGNC	<i>HUGO Gene Nomenclature Committee</i>
HVR	Çok değişken bölgeler (<i>hypervariable regions</i>)
IGV	<i>Integrative Genomics Viewer</i>
indel	İnsersiyon - delesyon
IRGV	<i>Ion Reporter™ Genomic Viewer</i>
jnlp	<i>Java Network Launching Protocol</i>
Kaspaz	Sistein-bağımlı aspartata yönlenmiş proteaz (<i>Caspases, cysteine-dependent aspartate-directed proteases</i>)
KDD	Veritabanları içindeki bilginin keşfi (<i>knowledge discovery in databases, knowledge extraction</i>)
MAE	Ortalama mutlak hata (<i>mean absolute error</i>)
MAMs	Mitokondri-ilişkili membranlar (<i>mitochondria-associated membranes</i>)
MELAS	Mitokondriyal miyopati, ensefalopati, laktik asidoz ve inme-benzeri epizodlar (<i>mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes</i>)
MERCs	Mitokondri-ER zar etkileşim bölgeleri (<i>mitochondria-ER membrane contact sites</i>)
MERRF	Miyoklonik epilepsi, dağınık-kırmızı liflerle (<i>myoclonic epilepsy with ragged-red fibers</i>)
MFF	Mitokondriyal fizyon faktörü (<i>mitochondrial fission factor</i>)
MFN1	Mitofusin-1
MFN2	Mitofusin-2
MIDD	Maternal kalıtsal diyabet ve sağırılık (<i>maternally inherited diabetes and deafness</i>)
MITOMAP	İnsan mitokondriyal DNA veri tabanı (<i>database of human mitochondrial DNA</i>)
mRNA	Mesajcı Ribonükleik asit (<i>messenger ribonucleic acid</i>)
MSE	Ortalama kare hata (<i>mean squared error</i>)
MT-TL1	Mitokondri tarafından kodlanan tRNA lösin 1 (<i>mitochondrially encoded tRNA leucine 1</i>)
Mt-tRNA ^{Val}	Mitokondri tarafından kodlanmış tRNA valin (<i>mitochondrially encoded tRNA valine</i>)
MtDNA	Mitokondriyal DNA (<i>mitochondrial DNA</i>)
MTERF1	Mitokondriyal transkripsiyon terminasyon faktörü 1 (<i>mitochondrial transcription termination factor 1</i>)
NADH	Redüksüye uğramış nikotinamid adenin dinükleotid (<i>reduced nicotinamide adenine dinucleotide</i>)

NaN	Eksik veri/boşluk (<i>Not a Number</i>)
NLRP3	Pirin domaini içeren NLR ailesi 3 (<i>NLR family pyrin domain containing 3</i>)
OLS	Sıradan en küçük kareler (<i>ordinary least squares</i>)
OPA1	OPA1 mitokondriyal dinamin-benzeri GTPase (<i>OPA1 mitochondrial dynamin like GTPase</i>)
P _i	İnorganik fosfat (<i>inorganic phosphate</i>)
PTMs	Post-translasyonel modifikasyonlar (<i>post-translational modifications</i>)
PZR	Polimeraz zincir reaksiyonu (<i>PCR, polymerase chain reaction</i>)
qPZR	Gerçek-zamanlı PZR (<i>qPCR, real-time PCR, quantitative polymerase chain reaction</i>)
RFLP	Sınırlayıcı enzim parça uzunluk çeşitliliği (<i>restriction fragment length polymorphism</i>)
RMSE	Ortalama kare hatanın karekökü (<i>root mean squared error</i>)
RNA	Ribonükleik asit (<i>ribonucleic acid</i>)
rRNA	Ribozomal Ribonükleik asit (<i>ribosomal ribonucleic acid</i>)
SNHL	Sensörinöral işitme kaybı (<i>sensorineural hearing loss</i>)
SNV	Tek nükleotid varyantı (<i>single-nucleotide variant</i>)
SSE	Artıkların kareleri toplamı (<i>sum of squared error</i>)
std	Standart sapma (<i>standard deviation</i>)
std err	Standart hata (<i>standard error</i>)
TCA	Trikarboksilik asit döngüsü (<i>TCA, tricarboxylic acid cycle</i>)
TE	Tris EDTA solüsyonu
TFAM	Mitokondriyal transkripsiyon faktör A (<i>mitochondrial transcription factor A</i>)
tRNA	Taşıyıcı ribonükleik asit (<i>transfer ribonucleic acid</i>)
tRNA ^{Leu}	tRNA lösin (<i>leucine</i>)
TSV	<i>Tab-separated values</i>
TVT	Tüm varyant tipleri
UPR ^{mt}	Mitokondriyal katlanmamış protein cevabı (<i>mitochondrial unfolded protein response</i>)
VIF	Varyans büyütme faktörü (<i>variance inflation factor</i>)
YND	Yeni nesil dizileme (<i>NGS, next generation sequencing</i>)

ÖZET

Şahin G. Oksidatif Fosforilasyon Sistemi Hastalıklarında Mitokondriyal DNA Varyantlarının Veri Madenciliği ile Araştırılması. İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü, Genetik ABD. Yüksek Lisans Tezi. İstanbul. 2020.

Mitokondrinin halka biçimindeki genomunda (mtDNA) çeşitli patojenik alellerinin heteroplazmik oranı varyant-yükünü belirler ve mitokondriyal hastalıkların klinik heterojenitesinden sorumludur. Bu tez çalışmasının ilk basamağında dokular-arası varyant-yükü karşılaştırmaları yapılarak tanısal testlerde kullanılan örnek etkinliğine öngörü sağlanması amaçlandı. Altı gönüllünün farklı dokularından eş zamanlı elde edilen mtDNA yeni nesil dizi teknolojisiyle araştırıldı. Kasta saptanan varyantların sayısı kana göre yüksek bulundu. Varyant tipi açısından, SNV sayısı kasta daha fazla iken, indellerde sayı yaklaştı. Ortak varyantlarda indel sayısı fazla idi. Ortak varyantların frekansı kasta, 1,2 kat, ortak SNV'lerin 3,7 kat, ortak indellerin 1,1 kat yüksek bulundu. Kandaki varyantların %28'inin kasta, kastakilerin %36'sının kanda saptanamadığı belirlendi. Bulgular, kan gibi mitoz hızı yüksek hücrelerde varyantlı mtDNA'ların eleniyor olması, kas gibi bölünmeyen hücrelerde ise bu avantajın oluşmadığı görüşünü destekledi. Ayrıca sonuçlar, mitokondride indellerin yeni oluşan SNV'lerin tamiri sürecinde oluştuğu, bu nedenle indel tipli varyantların kanda daha fazla beklenebileceği görüşüyle örtüştü. Üç doku karşılaştırması için gönüllü sayısı yetersiz bulundu. İkinci basamakta, klinik çeşitlilik ve kas patojenik varyant yükü arasında ilişki kurmak ve kas heteroplazmi oranı tahmini yapabilen çoklu-doğrusal regresyon modeli oluşturmak hedeflendi. 445 yayın incelenerek, m.3243A>G patojenik varyantına sahip hastaların özellikleriyle oluşturulan veri setinin, eksik heteroplazmi oranı bilgileri sebebiyle, yarısı kullanılarak (421 satır, 26 sütun) Model-1, yarısı Model-1'le tamamlanmış veri setiyle (744 satır, 26 sütun) Model-2 oluşturuldu. Bu modellerin ilişkilendirme başarı ölççeği olan ayarlanmış R-kare değerleri sırasıyla 0,47 ve 0,72'ydi. Bu değerdeki artış, veri sayısının artmasıyla ilişkilendirildi. Sonuç olarak, regresyon modelinin geliştirilebilmesinin, satır sayısının arttırılmasıyla, eksik değerlerin, optimize edilmiş klinik formlarıyla giderilmesiyle mümkün olabileceği düşünüldü.

Anahtar Kelimeler: Mitokondriyal DNA, heteroplazmi, mitokondriyal hastalıklar, m.3243A>G, çoklu-doğrusal regresyon

Bu çalışma, İstanbul Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenmiştir. Proje No: TYL-2019-33447

ABSTRACT

Şahin G. Data Mining Examination of Mitochondrial DNA Variants in Oxidative Phosphorilation System Diseases. İstanbul University, Institute of Health Science, Department of Genetics Thesis of Master. İstanbul. 2020.

The heteroplasmic ratio of various pathogenic alleles in the circular genome (mtDNA) of the mitochondria, determines the variant-burden and is responsible for the clinical heterogeneity of mitochondrial diseases. In the first step of the study, it was aimed to predict the efficacy of the sample used in diagnostic tests by making tissue variant-load comparisons. The mtDNA obtained simultaneously from different tissues of six volunteers was investigated using new generation sequencing technology. The number of variants detected in the muscle was higher than in blood. In terms of variant type, the number of SNVs was greater in muscle, while in indels the number was approximately. The number of indels was high in common variants. The frequency of common variants in the muscle was found to be 1.2 times higher, 3.7 times higher in common SNV and 1.1 times higher in common indel. It was determined that 28.6% of the variants in the blood could not be detected in the muscle, and 36.5% of the variants in the muscle could not be detected in the blood. The findings supported the view that mtDNAs with variants are eliminated in cells with high mitosis, such as blood, and that this advantage does not occur in non-dividing cells such as muscle. In addition, the results coincided with the view that indels in mitochondria are formed during the repair of newly formed SNVs, so indel-type variants can be expected more in blood. The number of volunteers was insufficient for three tissue comparisons. In the second step, it was aimed to establish a relationship between clinical diversity and muscle pathogenic variant-load and to create a multi-linear regression model that can predict the ratio of muscle heteroplasmy. A data set was created with the characteristics of patients with the m.3243A>G pathogenic variant by examining 445 publications. Because of the missing heteroplasmy ratio information in this data set, Model-1 was created using half of the data set (421 rows, 26 columns). Model-2 was created with the data set (744 rows, 26 columns) half completed with Model-1. The adjusted R-squared values of these models, which are the association success scale, are 0.47 and 0.72, respectively. The increase in this value was associated with the increase in the number of data. As a result, it was thought that the development of the regression model would be possible by increasing the number of rows and eliminating the missing values with optimized clinical forms.

Key Words: Mitochondrial DNA, heteroplasmy, mitochondrial diseases, m.3243A>G, multi-linear regression

The present work was supported by the Research Fund of Istanbul University. Project No. TYL-2019-33447

1. GİRİŞ VE AMAÇ

Nükleer genomdan ayrı bir genetik materyal içeren tek hücreyel organel olan mitokondrinin işlev bozukluğuna sebep olabilen mtDNA patojenik varyantları, geniş spektrumlu bir klinik ile kendini göstermektedir. Hücrelerin içerdiği mitokondri sayısındaki farklılık, mtDNA miktarındaki değişkenlik, farklı enerji ihtiyacına sahip hücrelerin oluşturduğu dokuların işlevleri doğrultusunda klinik heterojeniteye ve mitokondriyal hastalıkların tanı güçlüğüne neden olmaktadır. Heteroplazmi kavramıyla ifade edilen, varyantı içeren mtDNA'ların dokulardaki oranıdır ve bu oranın belli bir eşik değeri aştıktan sonra fenotipte etkilerinin gözlenmeye başladığı bilinmektedir. Heteroplazmi durumunun sağlıklı bireylerde dahi gözlemlendiği bildirilmiştir.

Bu durumlar, klinik özellikler ve heteroplazmi oranı arasındaki olası bir ilişkinin varlığının sorgulanmasının gerekliliğini göstermektedir. Ayrıca, dokular arası heteroplazmi karşılaştırmasının yapılması, farklılıkların incelenmesi, tetkiklerde kullanılabilecek dokunun belirlenmesi açısından önemli görülmüştür. Belirtilen unsurlar bu tez çalışmasının konusunu oluşturmaktadır.

İki basamaktan oluşan bu tez çalışmasının birinci basamağında, dokular arası heteroplazmi oran farklılıklarının deneysel olarak araştırılması, bu doğrultuda mtDNA çalışmalarında kullanılabilecek uygun dokunun tespiti ve bu dokunun diğer dokulara ne derecede üstünlük sağladığının belirlenmesi, ikinci basamağında ise, mitokondriyal hastalıklarda klinik özellikleri, demografik özellikleri, laboratuvar tetkikleri, kas histokimyasal patoloji ve beyin görüntüleme sonuçları ile farklı dokuların heteroplazmi oranları arasındaki ilişkinin veri madenciliği yöntemleri ile saptanması, bu ilişki olasılığının heteroplazmi oranını tahmin aracı olarak kullanılabilirliğinin sorgulanması, oluşturulan veri setinin, bu ilişkinin açıklanmasında yeterliliğinin belirlenmesi amaçlandı.

Literatürde, mitokondriyal hastalıkların tanısı amacıyla yapılan moleküler araştırmalarda farklı dokular kullanılmıştır. Farklı dokuların farklı heteroplazmi değerleri verdiği bilinmektedir ve fenotip ile ilişkisi zayıf olan doku seçiminin hastalığın klinik ciddiyeti hakkında bilgi vermemesi nedeniyle tedavi süreci olumsuz yönde etkilenebilmektedir. Bu nedenlerle bu tez çalışmasında, sağlıklı bireylerin farklı

dokularından elde edilen mtDNA'ların varyant yükü ve sayısı karşılaştırması ile, optimal dokunun belirlenmesi amaçlandı ve ileride yapılması planlanan araştırmalar için doku tercihi yönünde katkı sunması hedeflendi.

Araştırma kapsamında, mitokondriyal hastalıklara sebep olan mtDNA varyantları içinden literatürde en sık görülen patojenik varyant seçilerek ardından oluşturulan veri setinde makine öğrenmesi yöntemlerinden çoklu-doğrusal regresyon uygulandı. Bu yaklaşım ile, seçilen patojenik varyantla ilişkili mitokondriyal hastalıklarda görülen klinik heterojenite durumu ile kas heteroplazmi oranı arasındaki ilişkinin belirlenmesi yoluyla oluşturulan tahmin aracının tanı sürecini hızlandırıcı ve kolaylaştırıcı yönde etki etmesi, zaman içinde ortaya çıkabilecek farklı fenotipik özelliklerle, kas dokusundaki varyant yükü değişiminin takibinin mümkün olabilmesi, en uygun zamanda tedavinin olanaklı hale gelmesi, laboratuvar süreçlerinin kısalması yönünde etkilerinin olması beklenmektedir. Tahmin modelinin gücünü, ulaşılacak hasta sayısının, klinik özelliklerdeki ve heteroplazmi oranındaki çeşitliliği açıklamakta yeterli olacak miktarda olması, literatürde verilen bilgilerin eksiksiz olması, eksik veri olması durumunda, oluşturulan veri setinde kayıp veri yönetiminin iyi bir şekilde yapılması ve doğru algoritma seçimi etkilemektedir. Bu tez çalışmasında elde edilen sonuçların, ileride benzer hastalıklar için de uygulanması mümkün olacaktır.

2. GENEL BİLGİLER

2.1. Mitokondri ve Enerji

Hücre içi yapılara ilişkin ilk gözlemlere 1840'lı yıllarda başlanmıştır. Tanımlanan yapılar içinde mitokondrinin de bulunduğu düşünülmektedir [1]. 1890 yılında Richard Altmann, hücre içinde dağılmış halde bulunan yapılar gördüğünü, bu yapıların önemli faaliyetler üstlenen basit organizmalar olabileceğini bildirmiş ve bu yapıları biyoblast (*bioblast*) olarak isimlendirmiştir [1]. 1898 yılında Carl Benda, spermatogenezde gözlemlediği bu yapıları, iplikli ve tanecikli görünüşleri dolayısıyla mitokondrion¹ olarak adlandırmıştır [1].

Canlı organizmaların en küçük yaşamsal birimi olan hücrelerde, replikasyon, büyüme ve gelişme, metabolizma, protein sentezi, mobilite, apoptoz gibi kendiliğinden gerçekleşmeyen olaylar için (*non-spontaneous events*) yüksek miktarda enerji gerekir. Çok hücreli organizmaların enerji gereksinimleri tek hücreli canlılara kıyasla çok daha fazladır. ATP (adenozin trifosfat) hücrelerin faaliyetlerini gerçekleştirmek için kullanabileceği bir enerji formudur. Oksijenli solunum yapamayan tek hücreli canlılar oksijensiz solunumla enerji ihtiyaçlarını karşılarlar. Oksijensiz solunumda tüketilen besin miktarı ile kazanılmak istenen ATP miktarı doğru orantılıdır. Bu sistemde, glikoliz yolağı ile, bir molekül glikozun piruvata kadar yıkılması gerçekleşirken iki molekül ATP elde edilir. Ancak glikoliz yöntemi ile besinlerin tamamı yıkılamaz ve elde edilen enerji miktarı düşük kalır. Oksijenli solunum reaksiyonları ile aerobik hücrelerin dışarıdan aldığı besinlerden yüksek miktarda enerji elde etmek amaçlanır. Mitokondriler besinlerden maksimum ATP sağlayan, hücrenin fonksiyonel birimleridir. Bu işlemler, mitokondri içerisinde, özelleşmiş çok sayıda protein ile gerçekleştirilen bir dizi reaksiyon ile gerçekleşir. Bir molekül glikoz yıkımından ~30 ATP üretilir [3].

2.1.1. Mitokondrinin Evolüsyonu

1883 yılında botanikçi Andreas Schimper mikroskopta yaptığı incelemede bitki hücre kloroplastlarının bölündüğünü ve bu durumun bakterilerin bölünme şekli ile benzer olduğunu bildirmiştir [4]. İlk kez Rus botanikçi Konstantin Mereschkowski 1905 ve 1910 yıllarında yayımladığı çalışmalarda simbiyogenez teorisinden bahsetmiştir [5]-[7], 1950-60'larda mitokondri ve kloroplast organellerinin, hücre çekirdeğinde

¹ mitos: iplik (*thread*), khondrion: küçük tanecik (*little granule*) [2].

mevcut DNA'dan ayrı, kendi DNA'larına sahip oldukları keşfedilmiştir [8]-[10]. Tüm bu bilgiler 1967 yılında Lynn Margulis tarafından bir araya getirilerek *On the Origin of Mitosing Cells* isimli çalışmada ilk kez kanıtlarıyla birlikte *endosimbiyotik teori* ile açıklanmıştır [11].

Endosimbiyotik teoriye göre, karmaşık yapıdaki hücreler, daha az karmaşık yapıdaki hücrelerin simbiyotik ilişkisinden köken alır [5][11]. Basit yapıdaki tek hücreli organizmalar (prokaryotlar), çevre koşullarının etkisiyle gerçekleşen mutualist bir ilişki sonucu yaşamlarını bir arada sürdürebilirler. Oksijenli solunum yapabilen prokaryotun, yapısal olarak kendisinden daha büyük ancak oksijenli solunum yapamayan bir başka prokaryot tarafından fagositoz ile hücre içine alınması (fagozite edilmesi), karşılıklı yarar sağlayan bir ilişkiyi başlatmış ve bu oluşum uygun ortam koşullarına uyum olarak gelişmiştir [11].

Evrin teorisine göre yeni canlı türleri, çevre koşullarındaki köklü değişimler sonucu ortaya çıkar. Aşamalı olarak değişen koşullara uyum, aşamalı olarak adaptasyonla oluşur. Çevreye uyum ne kadar güçlüyse hayatta kalma, çoğalma ve neslin devamı o kadar mümkündür [12].

İlk atasal ökaryot hücreler, atmosferde serbest oksijen molekülünün artış gösterdiği süreçte simbiyotik ilişkiden avantaj sağlamışlardır [11]. Atasal heterotrofik ökaryotlar, oksijenli solunum ile dışarıdan aldığı besinlerden verimli enerji üretimi gerçekleştiren prokaryotlarla simbiyotik ilişki içerisinde [11]. Bu prokaryotlar modern ökaryot hücrenin atasal mitokondrisidir ve *Rickettsiales* proteobakterilerinin akrabası olduğu düşünülmektedir [13]-[16].

Mitokondrilerin prokaryotlarla olan benzerliği, endosimbiyoz teorisinin kanıtları arasında sunulmaktadır [11]. Boyutları bakterilerin boyutları ile benzerlik göstermektedir ($0,75-3 \mu m^2$) [16]. Hücreden bağımsız, bölünerek çoğalabilmektedir. Bu bölünme biçimi (*binary fission*) bakterilerle benzerlik göstermektedir [17][18]. Antibiyotiklere direnç göstermekte veya antibiyotiklerden etkilenmektedir [19][20].

2.1.2. Mitokondriyal Devinim

Mitokondriler dinamik organeller olup şekli, boyutu, hücredeki sayısı, hücredeki konumu, bölünme ve birleşmesi (fizyon ve füzyon), hücrenin metabolik aktivitesine, fiziksel durumuna, dokuya ve organizmaya göre devamlı olarak bir değişim ve dönüşüm gösterir [21][22].

2.1.2.1. Mitokondrinin Morfolojisi

Mitokondriler hücrede birçok farklı şekilde görülebilmektedir; küçük ve büyük granül, halter-görünümlü granül, iğsi granül, çubuk yapısı, iplik yapısı, ilmek, halka yapısı gibi. Ayırt edilemez durumlardaki görüntüsü ağ yapısı şeklindedir [21, p. 353]. Elastik bir yapısı vardır. Tek bir mitokondrinin, dakikalar içinde 15-20 farklı şekil alabildiği gözlemlenmiştir [21, p. 355]. Şekil farklılıkları, hücrelerin farklı yaşam evrelerinden bağımsız olarak, mitoz evrelerinde, bölünmüş hücrelerde, gelişen genç hücrelerde veya dinlenme halindeki yaşlı hücrelerde gözlemlenmiştir [21, p. 357]. Mitokondrinin morfolojisindeki değişimlerin, işlevsel bir özellik taşıdığı düşünülmektedir. Bölünme evresindeki mitokondrilerin halter veya raket görünümü aldığı, halter görünümünün bölünmenin ilk aşamasında, raket görünümünün bölünmenin bitmeye yakın evresinde ya da henüz ayrılmış mitokondrilerde gözlemlendiği bildirilmektedir [23]. Kısalmış kalınlaşmalar ile kristallerin birbirlerine yaklaştığı büzülmüş biçim ise ADP (adenozin difosfat)'ye fosfatın eklendiği, ATP'nin oluştuğu elektron transportu evreleriyle ilişkilendirilmektedir. Morfolojik değişimler, ADP ve ATP'nin mitokondriden çıkması veya mitokondriye girmesi, iç zarda bulunan ADP-ATP translokaz kanal proteinlerinde algılanmanın oluşmasıyla tetiklenmektedir [24][25]. Çubuk veya oval şeklindeki mitokondrilerin kristallerinin dallanmış ve düzenli yapıda olduğu ve toksik maddelere karşı duyarlı olduğu gözlemlenmiştir [26].

Mitokondride morfolojik çeşitliliği ve aktif şekil değişiklikleri, mitokondriyal devinim (*mitochondrial dynamics*) disiplini altında sınıflandırılmaktadır. Devamlı gözlenen kısalmış kalınlaşmalar, uzamalar, incelmeler, büyüme ve küçülmeler farklı aktivitelerin, bölünme ve birleşme süreçlerinin bir göstergesi olduğu, belli sinyaller sonucu gerçekleştiği bildirilmektedir [24].

Çubuk ve iplikli yapılar kısalmış kalınlaşarak küçük/büyük tanecik yapılarına dönüşebilir. İplikli yapıdaki mitokondriler bir araya gelebilir ve ağsı bir görüntü

oluşturabilir. Granül görünümlü yapılar bir araya gelip daha büyük granül yapıları, halka yapıdakiler ise iplik, çubuk veya tanecik yapısına dönüşebilir. Ömrünü tamamlamış, aktivite göstermeyen mitokondrilerin ise veziküllere ayrıldığı gözlenmiştir [21, p. 353].

Mitokondriler dinamik yapılar olduğundan boyutları da devamlı olarak değişebilmektedir [21, p. 359]. Farklı türlerde boyutunun $0,75-3 \mu m^2$ arasında değiştiği gösterilmiştir [27].

Hücrenin metabolik aktivitesine, yaşam evresine, fiziksel durumuna göre içerdiği mitokondri miktarı çeşitlilik göstermektedir. Sayıda artışlar veya azalmalar mitokondrilerin bölünmesi veya birleşmesi ile ilişkilidir [26].

Organları oluşturan farklı doku hücrelerinde, organ ve dokunun aktivitesi ile doğru orantılı sayısal çeşitlilik gözlenmekte, sayıları $\sim 80-2000$ arası değişebilmektedir [28]. Kalp, karaciğer gibi yaşam için önemli fonksiyonlar gerçekleştiren organların hücrelerinde sayısı, enerji ihtiyacının karşılayabilmek için yüksek orandadır [29]. Örneğin, sağlıklı bir insanda kalbin her gün ~ 30 kg ATP tükettiği bildirilmiştir [30]. Farklı memeli türlerinden alınan kalp kası örneklerinde yapılan miktar tayini çalışmasında, mitokondrilerin, kalp kası hücrelerinin %22-37'sini kapladığı görülmüştür [31]. Fareler üzerinde yapılan başka bir çalışmada, mitokondrilerin karaciğer hücrelerindeki hacmi $\sim 20\%$ olarak bildirilmiştir [32]. Yoğun egzersiz programlarında, sayıda önemli artışlar olduğu bilinmektedir [33]. Böylece hücre aktiviteleri için gerekli ATP ihtiyacı karşılanır. Mitokondrilerin sayısındaki azalma, boyutlarında büyüme ile sonuçlanmaktadır. Bu çok sayıda mitokondrinin birleşmesi ile gerçekleşir [34].

Vücuttaki hücre tipleri arasında yalnızca olgun eritrositler mitokondri içermez ve enerji ihtiyaçlarını hücresel solunumun ilk basamağı olan glikoliz ile sağlar. Eritrositler, oksijeni taşıyan hemoglobin molekülüne daha çok yer sağlayabilmek için iç alanı boşalmış hücrelerdir. Kemik iliğinde üretildikten kısa bir süre sonra, olgunlaşmakta olan eritrosit (retikülosit), çekirdeğini, çekirdek DNA'sını, mitokondrilerini ve mitokondriyal DNA'larını ve diğer organellerini kaybeder. Bu nedenle, herhangi bir hasar durumunda kendilerini onaramaz. Ömürleri ortalama 115 gündür, 70-140 gün arasında yaşayabilmektedir. Kısaca, bu hücreler kendilerinin hiçbir zaman kullanmadığı oksijeni tüm diğer dokular için sağlarlar [33]-[35].

2.1.2.2. Mitokondrinin Hücre İçindeki Yerleşimi

Mitokondrilerin hücre içindeki aktif çalışır bölgelerde yoğunluk gösterdiği, yerleşimlerinin devamlı değiştiği, dinlenme halindeyken bile devamlı pozisyon değiştirdikleri, sitoplazmanın geneline dağılmış halde bulundukları ve özellikle aktiviteye bağlı olarak çekirdek ve organellerin çevresinde yoğunlaşma gösterdikleri bildirilmiştir [21, p. 349]. Sinsityum halindeki hücre birleşmelerinde, sitoplazma köprüleri üzerinden bir hücreden diğer hücreye geçebildiği gözlemlenmiştir [21, p. 352].

Hücresinin farklı bölgelerindeki işlemler için devamlı olarak konumlarını değiştirirler. Mitokondrinin hareketleri hücre iskeleti elementleri ve mitokondriyal adaptör proteinleri ile sağlanmaktadır [38]. Çevreye dağılan mitokondriler daha sonra tekrar nükleus çevresinde toplanabilmektedir [21, p. 349]. İnce bağırsakta besinlerin aktif olarak absorbe edildiği lümen çevresini oluşturan epitel hücrelerinin ise apikal yüzeyinde gözlenmiştir. Organizmanın aldığı besinlere göre, hücre içinde farklı lokasyonlara toplanırlar. Yağ ile beslenen organizmalarda hücrenin merkezinde, karbonhidrat ağırlıklı beslenen organizmalarda hücre yüzeyinde toplanmaktadır [39]. Hipoksiye maruz bırakılan pulmoner arter endotel hücrelerinde bulunan mitokondrilerin nükleus çevresinde toplanarak, burada mitokondriyal reaktif oksijen ürünlerinin birikmesini sağladığı ve bu durumun nükleusun hipoksiye yanıt niteliğinde transkripsiyonel düzenlemeye gittiği gösterilmiştir [40]. Lenfosit hücrelerinin hedef yönünde hareketlerini desteklemek için, lenfositlerin polarize olmuş uropodlarında birikme gösterdiği izlenmiştir [41] Hareketini sürdüren lenfositlerin durması ile konum değiştiren mitokondrilerin, lenfositlerin bu kez sabit kalmasını sağladığı gözlenmiştir [42]. Nöronlarda mitokondrilerin akson uçlarına taşınması mikrotübüller ile gerçekleşmektedir [43]. Mitoz bölünme esnasında bölünmenin gerçekleştiği bölgede toplandıkları görülmüştür [44].

2.1.2.3. Mitokondrinin Devimsel Düzenlenimi ve Kalite Kontrol

Mitokondriyal birleşme ve bölünme (füzyon ve fizyon) nükleer genomdan sentezlenen düzenleyici proteinler ile düzenlenir. *Mitofusinler* yani *MFN1* (**mitofusin-1**) ve *MFN2* (**mitofusin-2**) mitokondrinin dış zarında, *OPA1*² ise iç zarında bulunan proteinler olup mitokondriyal birleşmeyi yönetirler. Mitokondriyal bölünme ise dış

²OPA1: OPA1 mitokondriyal dinamin-benzeri GTPase (*OPA1 mitochondrial dynamin like GTPase*) [45], HGNC

zarda bulunan FIS1³, MFF⁴ ve sitoplazmada bulunan ancak fizyon sırasında dış zara geçen DNM1L⁵ proteinleri ile yönetilir [45].

Mitokondrilerin biyoenerji faaliyetlerinin farklı olduğu gösterilmiş ve bu durum farklı zar potansiyelleri ile ilişkilendirilmiş, bunun mitokondrinin kalite kontrolü için önemli olduğu görülmüştür. Bölünmeye uğrayan bir mitokondriden oluşan iki mitokondrinin zar potansiyeli birbirinden farklı olabilir. Zar potansiyeli, ATP sentezinin sürdürülebilmesi için gerekli olan zarlar arası proton gradyanı dolayısıyla oluşan zarlar-arası elektriksel potansiyeldir. Potansiyelin düşük olduğu mitokondri, zar potansiyelini yeniden kazanabilir ya da depolarize edilebilir. Mitokondri zar potansiyelini kazanamazsa, proteolitik aktiviteler göstererek birleşme işlemi kontrol eden OPA1 yıkıma uğrar. Mitokondri, birleşme ile aktif mitokondri ağı yapısına katılamazsa mitofajiye uğrar. Polarizasyon gücü yüksek diğer mitokondri aktivite göstermeye devam eder. Böylece biyogenez ve mitofaji ile mitokondrilerin kalitesi kontrol edilir. Mitokondriyal biyogenezin yaşlanmayla birlikte azaldığı bildirilmiştir [46].

Matrikste ve zarlar-arası bölgede bulunan çözünebilir proteinlerde oksitlenme veya katlanmada bozulmalar olabilir. Böyle durumlarda bu proteinler mitokondriyal proteazlar tarafından yıkılır. Dış zar proteinlerinde meydana gelen hasarlar, bu proteinlerin ubiquitinasyona uğramasına ve sitoplazmada bulunan proteozomlar tarafından parçalanmasına neden olur. Mitokondri içerisinde oluşan atıklar ise dış zarın tomurcuklanması sonucu veziküllerle dışarı atılır ve lizozomlarda parçalanır [47].

2.1.3. Mitokondrinin Yapı ve Organizasyon Özellikleri

Mitokondri, bazı hücrelerin olgunlaşma evresinde yok olsa da tüm ökaryotik hücrelerde bulunur. Enerji üretimi için özelleşmiş çift zar yapıları, çift katmanlı fosfolipitlerden oluşmuştur ve her iki zar yapısına gömülü proteinleri vardır. Dışta bulunan zar *dış zar (outer membrane)*, iç kısımda bulunan zar ise *iç zar (inner membrane)* olarak isimlendirilir.

Dış zarda bulunan ve *porin* olarak adlandırılan proteinler, kanal biçiminde olan yapılardır. Porinlerden, piruvat gibi küçük moleküller ve iyonlar geçebilirken, yarı

³ FIS1: Mitokondriyal fizyon 1 proteini (*mitochondrial fission 1 protein*) [45], HGNC

⁴ MFF: Mitokondriyal fizyon faktörü (*mitochondrial fission factor*) [45], HGNC

⁵ DNM1L: Dinamin-1-benzeri protein (*dynammin-1-like protein*). Sıkça kullanılan diğer sembolü: DRP1, dinamin ilişkili protein 1 (*dynammin-related protein 1*) [45], HGNC

geçirgen özelliği sebebiyle RNA molekülü ve protein gibi daha büyük moleküller geçemez. Sitoplazma ve mitokondri arasındaki alışveriş bu kanallar ile sağlanır. Dış zar, mitokondriyal lipid sentezi ile ilişkili birçok enzim içermektedir ki bu mitokondrinin bölünme işlemi için gereklidir. Zar üretimi sayesinde bölünme sayısını arttırabilmektedir.

İç zar, yüzey alanını arttıran ve *krista* olarak isimlendirilen çok sayıda katlanma gösterir. Porin içermez. İyonlar ve küçük moleküller için seçici-geçirgen (*semi-permeable*) olup yalnızca su (H_2O), oksijen (O_2) ve karbondioksit (CO_2) için geçirgendir. Üzerinde, seçici-geçirgen özelliğine katkı sağlayan kardiyolipin fosfolipitleri yer almaktadır. Kardiyolipinlerin miktarı iç zarda, dış zara oranla çok daha yüksektir [48][49]. İç zar proteinleri üç ana grupta toplanabilir:

- Katalitik Enzimler: Metabolik reaksiyonları katalize eden protein grubudur.
- Transport Proteinleri: Hücre ve mitokondri arasındaki alışverişini sağlayan büyük protein moleküllerinin geçişine imkan veren kanal proteinleridir.
- ATP sentaz kompleksleri: ATP sentezi yapan, özelleşmiş enzim kompleksleridir (F_1F_0 ATPase).

İç ve dış zarın çevrelediği alanlar isimlendirilmiştir. Dış zar ve iç zar arasında kalan alan *zarlar-arası bölge* (*intermembrane space*), iç zarın çevrelediği alan ise *matriks* (*matrix*) olarak tanımlanmıştır. Zarlar-arası bölge yoğunluk açısından sitoplazma ile benzerlik göstermektedir. Yoğunluktaki benzerlik porin kanalları sayesinde. Zarlar-arası bölgede elektron taşıma sisteminde görevli sitokrom-C mevcuttur. Matriks, sitoplazmaya ve zarlar-arası bölgeye göre çözünabilir protein konsantrasyonu yüksek olduğundan daha yoğundur. Hücresel solunumun *Krebs döngüsü*⁶ bu kompartımanda gerçekleşmektedir. Matriks içerisinde, su, bir dizi reaksiyonun gerçekleştiği Krebs döngüsü için gerekli enzimler, koenzimler, ribozomlar ve nükleoid olarak adlandırılan özel yapılar içerisinde bulunan mitokondriyal DNA (mtDNA) bulunmaktadır [48][49].

Mitokondriler enerji üretimi için özelleşmiş hücre kompartımanları olsa da, tamamen bağımsız değildir. Mitokondriyal proteinler büyük ölçüde hücrede

⁶ Krebs döngüsü (*Krebs cycle*), sitrik asit döngüsü (**CAC**, *citric acid cycle*) veya trikarboksilik asit döngüsü (**TCA**, *tricarboxylic acid cycle*) olarak da bilinir [49].

sentezlenmektedir. Sitoplazmadan mitokondriye geçişleri, dış ve iç zarda gömülü olan özelleşmiş protein yapıları, *protein translokasyon kompleksleri* (*protein translocator complexes*) ile sağlanır. Dış ve iç zarda bulunan kompleksler etkileşim halindedir. Hücrede sentezlenen protein bir sinyal peptidi içermektedir. Sinyal peptidi sayesinde gidecekleri organelere ulaşmaları sağlanmaktadır. Şaperon proteinleri sayesinde mitokondri içine girene kadar katlanmaları engellenmektedir. Sinyal proteinleri dış zarda bulunan proteinler tarafından tanınır, mitokondri dış membranında gömülü protein olan translokasyon kompleksinden içeriye alınır ve etkileşimde olunan iç zar kompleksine yönlendirilir. Bu kompleksler multi-protein kompleksleridir. Hücreden mitokondriye gelen proteinlerin çok az bir kısmı yalnızca dış zarı geçmektedir. Çoğu zaman matrikse geçmeyip iç zara gömülürler. Matriks proteinleri ise zarlar-arası bölgeye girmeden, iç ve dış zar translokasyon proteinlerinin etkileşim yolunu kullanarak sitoplazmadan doğrudan matrikse ulaşır. Sinyal peptidi burada *sinyal peptidaz enzimleri* tarafından kırılır. Matriks içinde bulunan şaperon proteinleri ile proteinlere son şekli verilir [48][50][51].

2.1.4. Mitokondriyal DNA (mtDNA)

Memelilerde, nükleus DNA (**nDNA**, **nükleer DNA**)'sı dışında başka genetik materyal içeren tek organel mitokondridir. Mitokondriyal DNA (mtDNA), çift sarmallı ve halkasal yapıda olup bir DNA-protein kompleksi olan nükleoid yapısında, mitokondrinin matriksinde yer almaktadır. Nükleoidin yapısında bulunan 50'den fazla proteinin [52], mtDNA'nın korunmasında, transkripsiyon sürecinde, mitokondrilerin bölünme ve birleşmesinde, apoptozda, nükleer genomla iletişimi sağlayan sinyal yollarında rol aldığı bilinmektedir [53]. Nükleoid yapılarının ~100 nm olduğu gözlemlenmiştir [54]. MtDNA'nın nükleoid yapıları biçiminde paketlenmesi ve organizasyonu, TFAM⁷ başta olmak üzere pek çok nükleoid proteini ile sağlanır [54]. Kültüre edilmiş *insan A2780 hücre hattı* çalışmalarında, her bir mitokondrinin 1-10 ya da daha fazla sayıda nükleoid içerdiği, mtDNA kopya sayısının ise 1-15 arasında değiştiği gözlemlenmiştir, her bir nükleoidin ortalama 1,4 mtDNA içerdiği belirlenmiştir [55].

İnsan mitokondriyal genomu S. Anderson ve ark.'ları tarafından 1981 yılında dizilenmiştir [56]. mtDNA, 16.569 baz-çiftinden (bç) oluşmaktadır ve bç'leri

⁷ TFAM: Mitokondriyal transkripsiyon faktör A (*mitochondrial transcription factor A*) [54], HGNC

D-ilmeğinde bulunan *çok değişken bölgelerden (hypervariable regions) HVR1 ve HVR2* arasında lokalize olan orijin bölgesinden başlanarak saat yönünde sayılır. MtDNA, 12S ve 16S rRNA'yı kodlayan 2 adet rRNA geni, 22 adet tRNA geni ve 13 adet protein kodlayan gen içermektedir [56]. Sezyum klorid (CsCL) gradyanıyla denatürasyonu ile, sarmalların içerdiği guanin baz miktarlarındaki farklılık nedeniyle, iki sarmalın yoğunluğunun farklı olduğu gözlemlenmiştir [48][57]. Yoğunluğu fazla olan dış sarmal, *ağır sarmal (H-sarmalı, heavy strand)*; yoğunluğu az olan iç sarmal, *hafif sarmal (L-sarmalı, light strand)* olarak isimlendirilmiştir. MtDNA'nın kodladığı 37 genden 28'i H-sarmalı üzerinde bulunurken, 9'u L-sarmalı üzerinde bulunmaktadır. Protein kodlayan 13 genin ürünleri, ETZ⁸ ve oksidatif fosforilasyonla ilişkili proteinlerdir. Mitokondride, protein sentezinde kullanılabilecek tRNA'ları kodlayan gen sayısı 22'dir [56]. 64 adet mRNA kodonundan tRNA'ya bağlananların sayısı 61, 3 tanesi ise durma kodonudur. Öyleyse, *Watson-Crick bağlanma* özelliğine göre 61 adet tRNA olmalıdır. Ancak her bir kodon için ayrı ayrı tRNA bulunmadığından, mitokondriyal tRNA'ların bir çoğunda bulunan *wobble (non-Watson-Crick bağlanma)* özelliği sayesinde, özelliğe sahip mt-tRNA'lar kodonun üçüncü pozisyonunda bulunan dört nükleotidin her birini algılayabilmekte, böylece bir mt-tRNA birden fazla kodona bağlanabilmektedir [58] [59].

MtDNA'nın H-sarmalında bulunan *D-deplase ilmeği/D-ilmeği (D-loop, Displaced heavy strand)* nükleotid içeriği açısından sabit değildir ve *çok değişken bölge (hypervariable region)* olarak adlandırılır [60]. D-ilmeği fazladan bir sarmal daha içerir ve bu bölge transkripsiyon ve translasyonun başladığı yerdir. D-ilmeğinde kodlanan dizi bulunmaz. Çok değişken bölgeler HVR1 ve HVR2'yi içeren [60] bu dizideki varyantlar aracılığıyla mitokondriyal köken belirleme çalışmaları yapılabilmektedir. Kodlama yapmadığı için, oluşan varyantlar organizmaya zarar vermez [61].

Mitokondriyal genlerin transkripsiyonu her iki sarmalda da gerçekleşir. H ve L-sarmallarında bulunan promotor bölgeleri transkripsiyonun başlama bölgeleridir. Çift yönlü gerçekleşen transkripsiyon ile oluşan polisistronik transkript ürünler, daha sonra farklı işlemlerden geçerek olgun RNA'lara dönüşür [60][62][63]

İki sarmalın replikasyon orijinleri farklı yerlerde bulunur; H-sarmalı replikasyon orijini (O_H) D-ilmeğinde bulunurken, L-sarmal orijini ise bu bölgeden uzaktadır.

⁸ ETZ: Elektron transfer zinciri (ETC, *electron transport chain*) [56]

Replikasyon, nükleer genom tarafından kodlanan transkripsiyon faktörleri tarafından başlatılır (TFAM (**mtTFA**), *Mitochondrial Transkription Factor A*). TFAM transkripsiyonu tetikler ve H-sarmalında mtRNA polimeraz enziminin oluşturduğu DNA-RNA hibriti, DNA polimeraz gama tarafından üretilen yeni H-sarmalı için şablon görevi görür. Replikasyon saat yönünde gerçekleşir. O_L (L-sarmal orijini)'ne ulaşıldığında replikasyon buradan zıt yönde başlar. H-sarmalının replikasyonu bittiğinde L-sarmalı replikasyonu da sonlanır, süperkoil oluşumu ve yeni bir H-sarmalı replikasyonunun başlaması ile süreç tamamlanır [60][62][63].

mtDNA'nın maternal kalıtım biçimiyle yavru hücrelere aktarıldığı bilinmekle birlikte, nadiren paternal kalıtım gösterilmiştir. İnsan olgun oosit hücresi, 100.000-200.000 mtDNA içerirken [60], sperm hücresi ~100 mtDNA içerir [64]. Oosite penetre olurken, mitokondrilerinin büyük bir çoğunluğunu kuyruğu ile birlikte kaybeden sperm hücresinin baş kısmında bulunan az miktardaki mitokondrilerin içerdiği mtDNA'ların ubiquitinasyon mekanizması ile degrade edildiği bildirilmiştir [65]. Ancak, son yıllardaki çalışmalar bilateral kalıtımı destekler yönde olsa da [66]-[68], embriyogenezde paternal mtDNA'nın seçici yıkıma uğradığı ileri sürülmüştür [69].

Mitokondriyal ribozomlar (mitoribozomlar) matrikste bulunan protein kompleksleridir. MtDNA'dan transkribe edilen mRNA'ların translasyonunu gerçekleştirerek protein sentezinde görev alırlar. Hayvan hücrelerinde bulunan formları, 16S mtLSU⁹ ve 12S mtSSU¹⁰ olmak üzere iki alt-üniteden oluşmaktadır [70]. Yapısında aynı zamanda mt-tRNA^{Val11} ve 80 farklı protein bulunmaktadır. Mitoribozomdaki protein çokluğu protein-protein etkileşimlerini sağlarken aynı zamanda yapıyı reaktif oksijen ürünlerine karşı korunaklı hale getirir [71].

Mitokondri kendi genomuna sahip olsa da işlevi için gerekli çok sayıdaki proteinden sadece 13'ünü kodlamaktadır. Transkripsiyon, translasyon, replikasyon işlemlerine katılan, mitokondri yapısını oluşturan, sinyal yollarında, metabolizma yollarında mevcut proteinlerin neredeyse tamamı nükleer genom tarafından kodlanmakta ve bu proteinlerin mitokondriye taşınması yine nükleer genomdan sentezlenen proteinlerle sağlanmaktadır. 2009'da yayımlanan bir proteom çalışmasında, farenin (*murine*) kalp proteinlerinin tanımlaması yapılmış, 4906 proteinden 647'sinin

⁹ 16S mtLSU: 16S mitoribozomal büyük alt ünite (*mitoribosomal large subunit*) [71]

¹⁰ 12S mtSSU: 12S mitoribozomal küçük alt ünite (*mitoribosomal small subunit*) [71]

¹¹ Mt-tRNA^{Val}: Mitokondri tarafından kodlanmış tRNA-valin (*mitochondrially encoded tRNA-Valine*) [71], HGNC

mitokondri matriksinde, 846'sinin ise mitokondri membran yapısında bulunduğu gösterilmiştir [72]. Calvo ve Mootha (2010) tarafından yayımlanan proteom çalışmasında ise memeli mitokondriyal proteinlerinin ~1000-1500 kadar olduğu belirtilmiştir [73].

2.1.4.1. Homoplazmi ve Heteroplazmi Kavramları

Homoplazmi ve heteroplazmi kavramları¹², tek bir somatik hücre içerisindeki miktarı ~80-2000 [28] arasında olabilen mitokondrilerin genetik materyallerinin ne kadar benzer olduğunu ifade etmek için kullanılmaktadır [74].

Homoplazmi, hücrede bulunan tüm mtDNA dizilerinin aynı olduğu durumu ifade etmektedir [74]. Referans dizi nükleotidleri ve %100 heteroplazmi oranı gösteren varyantlar için homoplazmi durumu mevcuttur. Bazı kanser hastalarının tümör hücrelerinde, mtDNA patojenik varyantlarının homoplazmi durumu bildirilmiştir [78] [79].

Dokudaki hücrelerde bulunan mtDNA'ların genetik içeriklerinin birbirinden farklı olduğu durumlar *heteroplazmi* kavramıyla tanımlanmıştır. Mitokondriyal hastalığa sahip bireylerde, hastalığın şiddeti, ilgili dokudaki heteroplazmi oranıyla ilişkilendirilir. Hastalıkların fenotipte kendini göstermesi, bu oranın belli bir eşik değerini aşması ile gerçekleşmektedir [74].

Mikroheteroplazmi, mitokondriye sahip tüm organizmalarda görülebilen bir durumdur. Rafal M. Smigrodzki ve Shaharyar M. Khan (2005) tarafından, bir organizmanın yüzlerce, kalıtılmış veya *de novo* oluşmuş mitokondriyal patojenik varyantına sahip olabileceği ve her birinin %1-2 oranında saptanabileceği belirtilmiştir [80]. Oran oldukça düşük olsa da, vücutta bulunan mtDNA'ların büyük bir çoğunluğunda bulunan bu patojenik varyantlar, diyabet, kalp hastalıkları, *Parkinson*, *Alzheimer*, kanser gibi yaşa bağlı olarak ortaya çıkan hastalıklar ile ilişkilendirilmiştir [80][81].

2.1.4.2. MtDNA Dizileme Çalışmaları

İnsan mtDNA'sının tüm dizisi 1981 yılında Anderson ve ark.'ları tarafından tanımlanmıştır. Daha sonra mitokondriyal genom boyu ilişkilendirme çalışmaları ile *Alzheimer*, *Parkinson* gibi nörodejeneratif hastalıkların, otoimmün hastalıkların ve

¹² *homo-*: bir, aynı (*one, same*); *hetero-*: farklı (*other (of two), another, different; second, other than usual*); *-plasm*: hücre plazması kastedilmiştir (*a growth, a development; something molded*) [75]-[77].

kanserin de dahil olduğu pek çok diğer hastalığın mtDNA ile ilişkili olduğu anlaşılmaya başlanmıştır [82][83].

Olası patojenik varyantların ve heteroplazmi oranının belirlenmesi için PZR-RFLP analizleri, alel-spesifik PZR analizi, gerçek-zamanlı (*real-time*) amplifikasyon PZR, pirosekanslama gibi pek çok yöntem kullanılmıştır. Bu yöntemlerin kısıtlılığı yalnızca bilinen patojenik varyantları tanımlamak için kullanılabiliyor olmasıdır. Heteroplazmi oranının düşük olduğu durumlarda bu yöntemler, bilinen patojenik varyantları dahi saptamada yetersiz kalabilmektedir [84].

Yeni nesil dizileme teknikleri, eş zamanlı paralel sekanslama sağlamaktadır. MtDNA patojenik varyantlarından kaynaklanan kalıtsal mitokondriyal hastalıkların genetik tanısının belirlenmesi sürecinde, heteroplazmik varyantları da saptayabilen etkili bir yöntemdir [85]. Bu yöntemlerle düşük heteroplazmi oranlarının dahi belirlenebiliyor olması ve doğruluk payının yüksek olması (doğruluk katsayısı = 0,96 [84]) araştırmalara olanak tanımakta ve hastalıkların şiddetinin ve vücuda yayılım boyutunun anlaşılmasına katkı sağlamaktadır [83].

2.1.5. Mitokondrinin Fonksiyonları

2.1.5.1. Enerji Üretimi

Oksidatif fosforilasyon, dışarıdan alınan besinlerden verimli enerji üretiminin sağlandığı bir dizi redoks/indirgenme ve yükseltgenme tepkimelerini ifade etmektedir [86]. 1939 yılında Volodymyr Belitser tarafından adlandırılrsa da, 1900'lerin başından beri farklı bilim insanları tarafından, çeşitli enerji üretim mekanizmaları açıklanagelmiştir. Bu çalışmalardan ikisi, üç bilim insanına Nobel Ödülü kazandırmıştır. Peter Dennis Mitchell (1961) kemiosmotik teori ile ATP üretim mekanizmalarının [3] [87], John Ernest Walker ve Paul Delos Boyer ise *ATP sentaz protein kompleksinin* çalışma mekanizmasının anlaşılmasını sağlayarak ödüle layık görülmüştür. *Oksidatif fosforilasyon* tanımlamasındaki oksidasyon NADH ve FADH₂ kofaktörlerinin oksidasyonunu, fosforilasyon ise ADP molekülünün fosforillenmesini ifade etmektedir [88].

Sitoplazmada oksijen varlığından bağımsız gerçekleşen karbonhidrat, yağ ve protein yıkımının ürünleri, mitokondrinin matriks kompartımanında oksidasyon reaksiyonlarına uğratılır ve krebs döngüsünün bir metaboliti olan *asetil-CoA*

(asetil koenzim A) oluşturulur. Asetil-CoA'nın, matriksteki *oksaloasetat* molekülü ile birleşmesiyle *sitrik asit* oluşur bu da döngüyü başlatan metabolittir [89][90]. *Sitrik asit döngüsü* olarak adlandırılan bu döngüde mevcut metabolitlerin ve gerçekleşen reaksiyonların keşfi 1930'lu yıllarda Albert Szent-Györgyi tarafından yapılmış olsa da, döngünün tanımlanması 1937 yılında Hans Adolf Krebs ve William Arthur Johnson tarafından yapılmış, sonrasında döngü *Krebs döngüsü* olarak da anılmaya başlanmıştır. Sitrik asit yeniden oksaloasetata dönüşüncüye dek birçok reaksiyon gerçekleşir ve bu esnada hidrojen ile indirgenerek 3 molekül NADH^{13} ve 1 molekül FADH_2^{14} ve daha sonra ATP'ye çevrilecek olan 1 molekül GTP üretilir [89]. Hidrojen bir proton olmasına karşın bağlandığı moleküle elektron kazandırmasının sebebi elektronegatifliğinin düşük olmasıdır [91].

Krebs döngüsünde oluşan kofaktörler, mitokondri iç zarında bulunan protein Kompleks-I'e bir çift elektron aktararak okside olur. Elektronlar, elektron transfer zincirini (ETZ) oluşturan protein kompleksleri üzerinde bulunan redoks merkezleri boyunca ilerler ve kademeli enerji oluşumu sağlanır. Oluşan bu enerji, elektron taşıyıcı kofaktörlerin oksidasyonu sonucu serbest kalan ve matrikste biriken hidrojen iyonlarının zarlar arası bölgeye geçmesi için kullanılır. ETZ kompleks proteinleri boyunca elektron aktarımı ve protonların karşı tarafa taşınması işlemleri devam eder. Kompleks-IV'te elektronlar son elektron alıcısı oksijene aktarılır ve oksijen, serbest hidrojen iyonları ile birleşerek su oluşturur. Bu işlemler, zarlar-arası bölgede protonların birikmesi ile sonuçlanır. Zarlar-arası bölge ve matriks arasındaki yoğunluk farkı, elektrokimyasal gradyan oluşturarak protonlara hareket gücü kazandırır (*proton-motive force*) [3][86]. Serbest hidrojenler, ATP sentaz kompleksinin iç zara gömülü F_0 alt ünitesinden geçerek matrikse ulaşır. ATP sentaz, ETZ'nin son protein kompleksi olan Kompleks V'tir. F_0 halka yapısındadır ve hidrojenlerin geçişi bu yapının dönme hareketi ile sonuçlanır. Kompleksin F_1 alt ünitesi matrikse uzanmıştır. F_0 ünitesinin rotasyon hareketi, katalitik alt ünitelerden oluşan F_1 'de konformasyonel değişimlere sebep olur. Bu değişimler F_1 'in ATP ve ADP'ye ilgisini değiştirir. F_1 'in aktif bölgesi sentezin gerçekleştiği yerdir. ADP ve serbest fosfat (P_i) birleşimi burada gerçekleşir ve konformasyonel değişim sonucu ATP salınır [92].

¹³ NADH: Redüksiyona uğramış nikotinamid adenin dinükleotid (*reduced nicotinamide adenine dinucleotide*) [89]

¹⁴ FADH_2 : Redüksiyona uğramış flavin adenin dinükleotid (*reduced flavin adenine dinucleotide*) [89]

Krebs döngüsüne katılan her bir metabolitin belli bir miktarda sabit kalması gereklidir. Hücrede başka reaksiyonlarda da kullanılabilen bu metabolitlerin miktar dengesi, eksikliklerinin giderilmesini sağlayan anaplerotik ve fazlalıklarının uzaklaştırılmasını sağlayan kataplerotik reaksiyonlar ile gerçekleştirilir. Oksaloasetat miktarı, ATP üretim miktarını belirlemektedir. Hücrede artan ATP gereksinimi, oksaloasetat miktarının artırılması ile sonuçlanmaktadır [93].

Hücredeki ATP üretim ve tüketim işlemleri arasındaki dengenin sağlanması, üretim ve tüketim reaksiyonlarının düzenlenmesini gerektirmektedir. Hücre aktiviteleri sonucunda tüketilen ATP, hücrede ADP ve AMP¹⁵ miktarını artırmaktadır. Artan AMP miktarı hücre için bir sinyaldir [94].

AMP-aktif protein kinaz (AMPK), hücrenin enerji sensörüdür. Artan AMP:ATP oranı, AMPK'nın allosterik olarak aktivasyonunu sağlar [95]. AMPK mayalardan memelilere kadar neredeyse tüm ökaryotlarda korunmuş olan protein kompleksleridir [94]. Biri katabolik (α), ikisi düzenleyici (β , γ) üç çeşit alt üniteden (*subunit*) oluşmaktadır. Alfa alt-ünitesi, *Thr172* adı verilen bir kinaz domainine sahiptir. LKB1 ve CaMKK β kinazları tarafından fosforillenerek AMPK aktifleştirilir. Beta alt-ünitesi *karbonhidrat-bağlanma bölgesi* içerir. Gama-alt ünitesi ise AMP:ATP oranı değişimlerinde adenin nükleotidlerine bağlanarak AMPK'nın düzenleyici yönde etki mekanizmasını uyarmaktadır [95]-[97].

Oksijen yetersizliği, açlık, aşırı egzersiz gibi metabolik stres durumlarında, hücrenin var olan enerjiyi verimli kullanması ve enerji yapım mekanizmalarının aktive edilmesi süreçleri AMPK protein kompleksleri tarafından regüle edilir [98]. Kinaz enzim grubundan olan AMPK, metabolik yollardaki enzimlerin fosforillenmesini sağlayarak katabolik yolları aktive eder, glikoz ve yağ asitleri oksidasyonu gerçekleşir ve mitokondrilerin sayısı artırılarak ATP üretimi sağlanır. Bu süreçte ATP tüketimi yine AMPK tarafından engellenir. Hücresel faaliyetler, anabolik yollar inaktive edilir [98] [99].

AMPK'nın, anabolik ve katabolik yollarda reaksiyonların gerçekleşmesini sağlayan enzimlerin transkripsiyonun düzenlenmesinde rolü vardır. Metabolik stres mitokondrilerde hasara neden olabilir. Böyle durumlarda ATP miktarındaki düşüş,

¹⁵ AMP: Adenozin monofosfat (*adenosine monophosphate*) [94]

AMPK enzimlerinin MFF¹⁶ adı verilen *mitokondri bölünme faktörlerini* aktive etmesiyle sonuçlanır. Hasarlı mitokondriler fragmente edilir, ileri hasar durumunda ise apoptoza yönlendirilir. ATP üretimi gerçekleştiremeyen mitokondriler otofajiye uğrar. Organel-spesifik bir otofaji türü olan bu olay mitofaji olarak adlandırılır [98].

Besin depolarının yıkımı veya dışarıdan besin alımının uyarılması, hipotalamusta AMPK aktivasyonu sonucu yemek yeme uyarısı olarak organizma tarafından algılanır. AMPK aktivitesi düşerse yemek yeme isteği azalır. Hipotalamustaki sinyal yolları, sempatik sinir sistemi aracılığıyla uyarıları periferik dokulara gönderir. Böylece, organ ve dokularda glikoz ve yağ asidi oksidasyonu aktive edilir [95].

Metabolik hastalıkların tedavisi için üretilen ilaçlar AMPK protein kompleksinin aktivasyonunu tetikleyici yönde etki etmektedir. İlaçlar ile verilen ETZ inhibitörleri ATP deplasyonuna sebep olmakta ve AMPK aktivasyonunu sağlamaktadır [94]. AMPK aynı zamanda adiposit hücrelerinden salgılanan adipokinez, adiponektin ve leptin hormonları tarafından regüle edilir [95]

2.1.5.2. Kalsiyum Dengesi

Kalsiyum (Ca^{2+}), hücre içinde çeşitli faaliyetlerin gerçekleşmesi için sinyal görevi görür. Çoğalma, sekresyon, hücre ölümü gibi hücre faaliyetleri, kalsiyum miktarı ve kalsiyumun hücre içinde belli bir lokasyonda birikimi ile düzenlenir. Hücrede belli bir seviyede tutulması gereken Ca^{2+} seviyesi, endoplazmik retikulum, mitokondri, golgi aygıtı, lizozom gibi organellerin tampon etkisiyle düzenlenmektedir [100][26].

2.1.5.3. Oksijen Radikallerinin Kontrolü

Elektron transport zincirinde, kofaktörlerden kompleks proteinlere aktarılan enerji, kompleks-IV'te son elektron alıcısı oksijene (O_2) verilir. Elektronların kompleks-I ve III'e aktarımı sırasında O_2 'nin elektron yakalaması, serbest elektronlu reaktif superoksit ($O_2^{\cdot-}$) oluşumuna sebep olur. Kompleks-I, $O_2^{\cdot-}$ 'in matrikse sızmasına neden olurken kompleks-III matrikse ve zarlar-arası bölgeye sızmasına neden olur. Matrikste ve zarlar-arası bölgede bulunan dismutaz enzimleri ile $O_2^{\cdot-}$, hidrojen peroksit (H_2O_2) dönüştürülür. Süperoksit dismutaz-2 (SOD2) matrikste, SOD1 ise zarlar-arası bölgede bulunur. Süperoksitin reaktif formda kalması, nitrik oksitle (NO) birleşerek perosinitrit ($ONOO_2^{\cdot-}$)'i oluşturabilir. Bu reaktif mtDNA'da hasara, kalıcı protein

¹⁶ MFF: Mitokondriyal fizyon faktörü (*Mitochondrial fission factor*) [98]

hasarlarına, mitokondriyal bütünlüğün bozulmasına sebep olabilmektedir. Süperoksit dismutaz enzimleri antioksidan görevi görmektedir. $O_2^{\cdot-}$ 'i, H_2O_2 'ye yani reaktif bir formun elektrofobik bir forma dönüşmesini sağlar ve H_2O_2 , kataliz ve glutatyon peroksidaz enzimleri ile su molekülüne dönüştürülür [101][102].

2.1.5.4. Apoptoz

Apoptoz, hücre ve organizma homeostazı için gereklidir ve gelişim sırasında gerçekleşen parçaların ayrılması sürecinde, DNA hasarlarında, çevre şartlarına uyum için hücrenin maruz kaldığı stres durumlarında, enfeksiyonlarda veya serbest oksijen radikallerinin biriktiği durumlarda hücrelerin programlı ölümünü sağlar. Mitokondri, hücredeki apoptozun başlama sürecinde rol oynar. Hücrede mevcut makromoleküllerin yıkımı yönetilir. Apoptoz mitokondrinin dış zarının geçirgenliğinin artması ile başlar. Pro-apoptotik ve anti-apoptotik proteinler içeren BCL2¹⁷ protein ailesi üyelerinden olan proteinler, geçirgenliği düzenleyen proteinlerdir. Sağlıklı bir hücrede anti-apoptotik proteinler aktiftir ve mitokondrinin apoptozu başlatmasını engeller. Apoptozu gerektirecek sinyaller alındığında pro-apoptotik proteinler aktive olur ve dış zar geçirgenliği artırılır. Geçirgenliğin artması mitokondride, zarlar-arası bölgede bulunan, aynı zamanda ETZ'de rol oynayan sitokrom-C moleküllerinin sitoplazmaya yayılmasına neden olur. Sitokrom-C, sitoplazmada kaspaz¹⁸ enzim ailesinin aktivasyonunu sağlar. Kaspaz, proteinlerin degradasyonunu sağlayan bir proteaz grubudur. Aktif bölgesinde mevcut sistein rezidüleri sayesinde, proteinleri, aspartat rezidülerinin hemen ardından keserek proteinlerin yıkımını sağlamaktadır. Nükleazların aktivasyonu ile nükleer DNA degradasyona uğratılır. Yıkım ürünleri çevredeki hücreler tarafından fagozite edilip yapı taşlarına dönüştürülerek yeniden kullanılabilir [103].

2.1.5.5. Mitokondriyal Metabolitlerin Epigenetik Mekanizmalar Üzerine Etkisi

Mitokondriyal aktivite sonucunda açığa çıkan, asetil koenzim A, ketoglutarat, nikotinamid adenin dinükleotid (NAD⁺), koenzim A gibi metabolitler, DNA ve protein modifikasyonlarında görev almaktadır [104]. Mitokondriyal metabolitler, asetilasyon, metilasyon gibi DNA, histon ve protein modifikasyonlarında, malonilasyon, süksinilasyon, glutarilasyon gibi diğer protein modifikasyonlarında (PTMs¹⁹) görevli

¹⁷ BCL2: BCL2 apoptoz regülatörü (*BCL2 apoptosis regulator*) [103], HGNC

¹⁸ Kaspaz: Sistein-bağımlı aspartata yönelmiş proteaz (**Caspases**, *cysteine-dependent aspartate-directed proteases*) [103]

¹⁹ PTMs: Post-translasyonel modifikasyonlar (*Post-Translational Modifications*) [105]

enzimler için birer kofaktör görevi görmektedir [105]. Epigenetik değişimler ile nükleer gen ekspresyonunun düzenlenmesi ve transkripsiyon-sonrası protein modifikasyonları sağlanır. Mitokondriyal metabolitlerin bu işlevi, kök hücre aktiviteleri ve yaşlanma ile de ilişkilendirilmiştir [106].

Nükleer genom ve mitokondriyal genom tarafından kodlanan mitokondriyal proteinler arasındaki denge, mitokondrinin işlev gösterebilmesi için önemlidir. Dengenin bozulduğu durumlar proteotoksik stres ile sonuçlanabilmektedir. Nükleer genom tarafından kodlanan mitokondriyal proteinlerin katlanmalarındaki hatalar sonucunda, mitokondrilerin içinde yanlış katlanmış proteinler birikmektedir. Bu durumun düzeltilmesi için verilen sinyaller, UPR^{mt} ²⁰ olarak adlandırılmıştır. Bu tepki ile mitokondriyal faaliyetler ve hücre içi denge arasında senkronizasyon sağlanmaktadır. Sinyal yolları ile nükleer gen ekspresyonu düzenlenerek hücrel stresin oluşması engellenmektedir [107].

Mitokondriyal devinim aktiviteleri sonucu oluşan sinyaller ve hücrel stresin oluşturduğu sinyaller, nükleer genomun epigenetik düzenlenmelerinin kontrolünü sağlamaktadır [108]. Ayrıca, apoptozdaki rolü sayesinde, mitokondrinin gen ekspresyonunun düzenlenmesinde de etkili olduğu bilinmektedir [106][109].

2.1.6. Endoplazmik Retikulum ve Mitokondri Arasındaki İlişkisi

Endoplazmik retikulum (ER) ve mitokondriler yan yana geldiği zaman zarları arasında etkileşim meydana gelmektedir [110]. MERCs²¹ olarak isimlendirilen bu etkileşim, organeller arası bağlantı oluşumunu sağlamaktadır [111]. Bu bağlantının farklı işlemlerin düzenlenmesinde rol oynadığı bildirilmiştir [112].

ER-mitokondri bağlantısının mitokondriyal bölünme işleminin başlamasında rol oynadığı gösterilmiştir [110]. Dinamin-ilişkili proteinler tarafından regüle edilen mitokondriyal bölünmenin bölünme merkezi, ER tarafından belirlenmektedir. ER ve mitokondrinin yan yana gelmesi ile ER tübülleri ve mitokondri etkileşim içine girer ve tübüller mitokondriyi sararak sıkma hareketi gösterir [113]. Bu işlemde sonra DNM1L²², heliks yapısı oluşturarak mitokondrinin bölünmesini sağlar [110].

²⁰ UPR^{mt} : Mitokondriyal katlanmamış protein cevabı (*Mitochondrial Unfolded Protein Response*) [107]

²¹ MERCs: Mitokondri-ER zar etkileşim bölgeleri (*Mitochondria-ER Membrane Contact sites*) [111]

²² DNM1L: Dinamin-1-benzeri protein (*dynamitin-1-like protein*). Sıkça kullanılan diğer sembolü: DRP1, dinamin ilişkili protein 1 (*dynamitin-related protein 1*) [110], HGNC

İki organel arasında fosfolipit alışverişi gerçekleşmektedir. Gerçekleşen lipit transferleri, devamlı bölünme ve birleşme faaliyetleri gösteren mitokondrinin zar yapısının lipit kaynağıdır [111]. Etkileşimin gerçekleştiği zar bölgelerinde mevcut proteinler, iki organel arasında Ca^{2+} alışverişini sağlamaktadır. Artan kalsiyum konsantrasyonu, salgılama, kasılma, çoğalma veya apoptoz gibi hücre faaliyetlerini uyarmaktadır [114]. Hücre homeostazı için önemli hücresel fonksiyonlardan otofaji mekanizmasının ilk aşaması olan otofagozom oluşumunu sağlamaktır [115].

Etkileşimde yer alan proteinler bir multi-protein kompleksi olan *inflamazom* oluşumunda rolü olan proteinlerdir [112]. Bu protein kompleksinde bulunan sensör proteini, patojen kaynaklı sinyalleri, mikrobiyal ürünleri ve sitoplazmada gerçekleşebilecek diğer tehlikeleri algılar. Farklı çeşitlerde inflamazomlar bulunsada en iyi bilineni ve ER ile mitokondrinin ilişki halinde olmasını sağlayan, MAMs²³ ile ilişkisi açıklanmış NLRP3²⁴tür [116].

2.1.7. Mitokondriyal Hastalıklar

Rolf Luft ve ark.'ları 1962 yılında, metabolik bir hastalığın, mitokondriyal hücresel solunum hasarından kaynakladığını ortaya koyan bir çalışma yayımlamıştır. Bu çalışmada ilk mitokondriyal hastalık tanımlanmıştır [117]. Wallace ve ark.'ları ise 1988 yılında LHON²⁵ ile ilişkisi bilinen ilk mtDNA patojenik varyantını (m.11778G>A) tanımlamıştır [116].

Mitokondriyal hastalıklar, mitokondrinin yapısal ve/veya işlevsel proteinlerinin ya da RNA'larının yanlış üretimine sebep olan ve mitokondrinin işlev bozukluğuna yol açacak yönde etki eden, nDNA veya mtDNA patojenik varyantları nedeniyle ortaya çıkan genetik hastalıklardır [85]. Hastalıklar, bazı ilaçların ters etkisi, enfeksiyonlar ve bazı çevre faktörleri nedeniyle meydana gelen mitokondriyal işlev bozuklukları ile de ortaya çıkabilmektedir [119][120]. Patojenik varyantlar kalıtsal veya sonradan kazanılmış olabilir. Hastalık, mtDNA patojenik varyantları nedeniyle ortaya çıkmış ise maternal kalıtım, nükleer genom kaynaklı ise otozomal dominant, otozomal resesif veya X'e bağlı kalıtım gösterir [117]. Gorman ve ark.'ları tarafından, İngiltere'nin kuzey doğu bölgesinde, mitokondriyal hastalık şüphesi olan yetişkin bireyler arasında yapılan çalışmada (2015), nDNA ve mtDNA patojenik varyantlarından kaynaklanan

²³ MAMs: Mitokondri-ilişkili membranlar (*Mitochondria-Associated Membranes*) [116]

²⁴ NLRP3: Pirin domaini içeren NLR ailesi 3 (*NLR family pyrin domain containing 3*) [116]

²⁵ LHON: Leber'in hereditör optik nöropatisi (*Leber's Hereditary Optic Neuropathy*) [116]

mitokondriyal hastalıkların prevalansı araştırılmış ve mtDNA ilişkili hastalıklar için minimum prevalans 20/100.000, nDNA ilişkili mitokondriyal hastalıklar için ise 2,9/100.000 olarak belirlenmiştir [121]. MtDNA'da 400'den fazla patojenik varyant bildirilmiştir [122]. En yaygın kalıtsal metabolik hastalık grubudur. Aynı zamanda, kalıtsal nörolojik hastalıkların en yaygın formları arasındadır [83].

Mitokondriyal hastalıklarda görülen klinik özellikler, etkilenmiş mitokondrilerin miktarına ve farklı organ ve dokulardaki dağılımına bağlı olarak çeşitlilik gösterir. Hastalık, vücudun tek bir işlevsel birimiyle sınırlı olarak kendini gösterebileceği gibi, pek çok işlevsel birimin etkilendiği sistemik hastalıklar olarak da görülebilmektedir [85]. Hastalıkların isimlendirilmesi genellikle etkilenen organ veya organların gösterdiği kliniğe göre akronimlerle yapılmaktadır [74].

Yüksek enerji ihtiyacı olan hücre tipleri, mitokondriyal bozukluklardan en çok etkilenen birimlerdir. Enerji yoksunluğu, bu hücrelerin görevlerini yerine getirememesi ile ve bu hücrelerin bir araya gelerek oluşturduğu doku ve organların işlev bozukluğu ile sonuçlanmaktadır. Etkilenen işlevsel vücut birim veya birimleri ve bu birimin veya birimlerin etkilenme derecesi, hastalığın kliniğini belirleyen faktörlerdir. Klinik özellikler, etkilenen bu birim veya birimlerin işlevleri üzerinden kendisini göstermektedir. Etkilenen iskelet kası ve sinir hücreleri, klinikte, mitokondriyal hastalıklarda en sık görülen kas problemleri ve nörolojik problemlere sebep olmaktadır. Beyin, kalp, böbrek, karaciğer, göz, kulak ve pankreas gibi farklı vücut birimlerinin tutulumları görülebilmektedir [85]. MtDNA patojenik varyantlarından kaynaklı hastalıklarda, klinik bulgu ve belirtilerin şiddeti, heteroplazmi oranıyla doğru orantılıdır. Hastalıklar hafif seyredebildiği gibi ağır klinik bulgularla seyredebilir, zamanla hafifleme veya ağırlaşma gösterebilir, hatta ölümcül olabilir [74].

Klinik heterojenite gösteren mitokondriyal hastalıkların tanı süreci zordur. Klinik bulgular ve belirtiler, biyokimyasal testler, kasın histokimyasal analizi, beyin görüntüleme yöntemleri ve genetik test yöntemleri tanı sürecinde değerlendirilmektedir [123].

Hastalıkların kesin bir tedavisi olmamakla birlikte semptomları hafifletici yönde destekleyici tedavileri uygulanmaktadır. MtDNA patojenik varyantlarından kaynaklanan hastalıklar için mitokondriyal replasman teknikleri uygulanabilmektedir [122][124].

2.1.7.1. m.3243A>G Patojenik Varyantı ve MELAS

3243 pozisyonu, mtDNA'nın MT-TL1²⁶ (3230-3304) kodlanmayan geninde, MTERF1'in²⁷ mtDNA'ya bağlandığı 13 bç uzunluğundaki terminasyon sekansı (3237-3249) içerisinde yer almaktadır [125]-[128]. Bu pozisyondaki A nükleotidinin G'ye değişiminin, mitokondriyal hastalığa sebep olan en yaygın patolojik varyant olduğu görülmüş [121] ve bu nokta değişimi MELAS hastalığıyla ilişkilendirilmiştir [129]. Ortaya çıkan patolojik durumun moleküler olarak anlaşılabilmesi için, pozisyonun MTERF1'in bağlanma bölgesine denk gelmesi nedeniyle, değişimin transkripsiyonun sonlanması etkileyebileceği üzerinde durulmuştur. Yapılan *in vitro* çalışmalarda MTERF1'in terminasyon sekansına bağlanma ilgisinin azaldığı görülse de, *in vivo* çalışmalarda transkripsiyon ürünlerinin miktarında bir değişim görülmediği bildirilmiştir [130]-[132]. Bu nedenle bakış açısı değiştirilmiş ve değişimin patolojik etkisinin, 3243 pozisyonunun da içinde bulunduğu dizinin ürünü olan tRNA lösün (tRNA^{Leu} (UUR)) yapısındaki değişimden kaynaklanabileceği düşünülmüştür.

Post-transkripsiyonel modifikasyona uğrayan tRNA bölgelerinden biri de, antikodonun 5' ucundan birinci pozisyonudur. Bu pozisyon *wobble* pozisyonu olarak bilinmektedir. Mitokondriyal *wobble* kuralına göre mitokondriyal tRNA'nın *wobble* pozisyonundaki modifiye edilmemiş üridinin dört farklı bazla sonlanan kodonlara bağlanabildiği, modifiye edilmiş üridinin ise pürin bazlarıyla (A/G) sonlanan kodonlara bağlanabildiği bildirilmiştir. Bu durum post-transkripsiyonel modifikasyonların, translasyonun verimliliği ve doğruluğu açısından önemini göstermiştir. Mitokondriyal tRNA^{Leu} *wobble* pozisyonunda 5-taurinometilüridin (τm⁵U) modifikasyonuna sahiptir [58][133]. tRNA^{Leu} genindeki m.3243A>G patojenik değişiminin sonucu olarak MELAS hastalığına sahip hastalardan alınan örnekler üzerinde yapılan çalışmalarda, tRNA^{Leu}'in *wobble* pozisyonundaki üridinde τm⁵U modifikasyonunun olmadığı görülmüştür [134]. Modifiye olmamış üridine sahip tRNA^{Leu}'in, *wobble* pozisyonunda dört farklı bazla eşleşebileceği için, lösün (UUR, R: pürin bazları, A/G) kodonları yerine fenilalenin (UUY, Y: pirimidin bazları, U/C) kodonları ile eşleşerek hatalı translasyona ve/veya mitokondriyal protein sentezinde azalmaya neden olabileceği düşünülmüştür [135].

²⁶ MT-TL1: Mitokondri tarafından kodlanan tRNA lösün 1 (*mitochondrially encoded tRNA leucine 1* (UUR, R: A/G)) [126]

²⁷ MTERF1: Mitokondriyal transkripsiyon terminasyon faktörü 1 (*mitochondrial transcription termination factor 1*) [127]

MELAS, mitokondriyal miyopati, ensefalopati, laktik asidoz ve inme-benzeri epizodlar (*Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes*) bulgularının birlikte görüldüğü ancak bu bulgularla sınırlı kalmayıp pek çok organın dahil olduğu klinik bulgularla seyredabilen mitokondriyal bir hastalıktır (OMIM 540000) [136]. Bulguların birlikteliği, Steven Pavlakis ve ark.'ları (1984) tarafından, ilk kez bu isimle tarif edilmiştir [137]. Pek çok mtDNA patojenik varyantının bu hastalığa sebep olabildiği belirtilmişken [136], hastaların %80'inde görülen en sık patojenik varyant, m.3243A>G olarak Goto ve ark.ları (1992) tarafından bildirilmiştir [129]. Kafkas popülasyonunda yapılan bir çalışmada hastalığın prevalansı 236/100.000 olarak belirlenmiştir [138].

MELAS ile ilişkili semptomlar, hastaların büyük çoğunluğunda 2-40 yaş arasında ortaya çıkmaktadır. Boy kısalığı, kaslarda zayıflık, egzersiz intoleransı, baş ağrısı ve kusma, erken dönem bulgu ve belirtileridir. Periferik nöropati, işitme kaybı, diyabet, epilepsi, demans, pitoz, CPEO²⁸, katarakt, kortikal körlük, hemianopsi gibi göz tutulumları, kardiyomiyopatiler, kalp iletim bozuklukları, *Wolff-Parkinson-White* sendromu gibi kalp tutulumları, FSGS²⁹, tübüler asidoz, proteinuria gibi böbrek tutulumları, solunum bozuklukları, karaciğer tutulumu, psikolojik hastalıklar ve gastrointestinal problemler, m.3243A>G patojenik varyantının saptandığı hastalarda görülebilen klinik özelliklerdir [136][139][140]. Varyant, MELAS sendromu ile direkt ilişkilendirilse de, klinik çeşitliliğin boyutu, hastaların farklı hastalık tanımlamaları ile tanı alması ile sonuçlanmış, MIDD³⁰, MELAS/MERRF³¹ çakışma sendromu (*overlap syndrome*), yaşa bağlı maküler dejenerasyon (*age-related macular degeneration*), SNHL³², döngüsel kusma sendromu (*cyclic vomiting syndrome*) gibi fenotipik tanımlamalar mevcuttur [141].

²⁸ CPEO: Kronik progresif eksternal oftalmopleji (*chronic progressive external ophthalmoplegia*) [141]

²⁹ FSGS: Fokal segmental glomerülo skleroz (*focal segmental glomerulosclerosis*) [141]

³⁰ MIDD: Maternal kalıtmı diyabet ve sağırılık (*maternally inherited diabetes and deafness*) [141]

³¹ MERRF: Miyoklonik epilepsi, dağınık-kırmızı liflerle (*myoclonic epilepsy with ragged-red fibers*) [141]

³² SNHL: Sensörinöral işitme kaybı (*sensorineural hearing loss*) [141]

2.2. Veri Madenciliği ve Makine Öğrenmesi

Veri madenciliği (*data mining*), veri tabanlarında, veri setlerinde var olan bilgilerin kullanılarak, ortaya çıkartılan ilişkiler sonucunda, var olmayan bilgilerin keşfedilmesi veya öngörülmesi işlemidir (KDD³³) [142][143]. *Data mining* terimi ilk kez Michael Lovell tarafından 1983 yılında kullanılmıştır [144].

Aynı veya farklı veri tabanlarından (*databases*) elde edilen veriler, seçme işlemleri sonucunda bir araya getirilerek veri seti oluşturulur. Oluşturulan veri seti, ön işleme (*pre-processing*) sürecinden geçirilir. Hazır olan veri üzerinde amaca yönelik olarak algoritmalar uygulanır [142].

Makine öğrenmesi (*machine learning*) algoritmaları aracılığıyla, veri setinin içerisinde mevcut örüntüler, birbiriyle hareket eden özellikler arasındaki ilişkilerle ortaya çıkartılır. Veri işlenebilir hale getirildikten sonra, öğrenen algoritmalara bir öğrenme verisi (*training data*) sunulur. Bu veride mevcut özelliklerin birbirleri arasındaki ilişki ve birbirlerine etki eden unsurlar algoritma tarafından öğrenilir. Gelecek hakkında tahminler, öğrenim süreci sonucunda oluşturulan modeller ile yapılır [142][143].

Öğrenme sürecinde kullanılan algoritmalar, veriden çıkarılması istenen bilgiye göre belirlenir. Bu seçimin doğru yönetimi, öğrenme verisinin test edilmesi sırasında açığa çıkan bilginin doğruluk yüzdesini etkilemektedir. Algoritmalar belli makine öğrenmesi yöntemi çerçevesinde geliştirilmektedir. Sınıflandırma, kümeleme ve ilişkilendirme, bağlanım (regresyon) yöntemleri kullanılan algoritma çeşitlerine örnektir [142].

Sınıflandırma, veri setinde mevcut bilgilerin gruplandırılması metodudur. Gruplandırma ile sağlanan öğrenme sonucunda yeni verilerin, verilen özelliklerine göre dahil olduğu grupların belirlenmesidir. Medikal alanda kullanım örnekleri mevcuttur. Hasta bulgularının ve belirtilerinin girdi olarak verilmesi ile çıktı olarak tanının belirlenmesi mümkün olur. Kümeleme, benzer özellikler çerçevesinde bilgilerin ayrılması işlemidir. Veri dağılımının anlaşılmasında kullanılabilir. İlişkilendirme, veri setinde belirlenmiş özellikler (*features*) arasında saklı örüntülerin bulunmasını amaçlayan bir yaklaşımdır. Büyük veri setleri için çok boyutlu ilişkilendirme (*multi-dimentional association*) çalışmaları yapılır. Market sepeti uygulaması, en iyi

³³ KDD: Veri tabanları içindeki bilginin keşfi (*Knowledge Discovery in Databases, knowledge extraction*) [143]

bilinen ilişkilendirme çalışmasıdır. Alışveriş esnasında birlikte alınan ürünlerin saptanması ile satışların arttırılmasına yönelik yolların ortaya çıkarılmasını sağlar. Bağlanım, veride mevcut ya da oluşturulan özellikler arasındaki ilişkilerin ortaya çıkarılması yöntemidir. Elde edilmek istenen bilginin çeşidine göre uygun metot ve belirlenen metot için geliştirilmiş uygun algoritma seçimi sonucunda veri madenciliği/makine öğrenmesi uygulamaları gerçekleştirilir [142][145].

Veri madenciliği, büyük verinin olduğu tüm alanlarda uygulamaları olan işlemlerdir. Büyük veride mevcut ilişkilerin ve modellerin bulunmasını hedefleyen bu işlemler biyolojik bilimler ve medikal alanlarda da kullanılmaktadır (biyoenformatik = *bioinformatics*) [145].



3. GEREÇ VE YÖNTEM

3.1. Laboratuvar Çalışmaları

3.1.1. Örnek Grubu

Bilgilendirilmiş onamları alınan toplam 6 gönüllü birey çalışmaya dahil edildi.

Gereken etik kurul kararı, İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Klinik Araştırmalar Etik Kurulu'ndan alınmış olup, 27.11.2018 tarihinde, 1626 sayılı belge ile çalışmanın onayı bildirilmiştir.

3.1.2. Örnek Alımı

Her bir gönüllüden ortopedik bir işlem amacıyla alınan, gerekli tıbbi işlemleri tamamlandıktan sonra tıbbi atık statüsüne giren kan, kas ve idrar olmak üzere toplam 3 örnek bu çalışma için kullanıldı. Kan örneği 2ml'lik steril EDTA'lı tüplere alınmıştı. İdrar kabında gelen idrar örnekleri ve 50 ml'lik falkon tüplerde, besin medyumunu içinde gelen kas örnekleri 15 ml'lik steril falkon tüplere aktarıldı. Kas örneklerine 15 ml DMEM besin medyumunu eklendi. +4°C buzdolabında bir gün muhafaza edildikten sonra bir sonraki işleme geçildi.

3.1.3. DNA İzolasyonu

Kan örneklerinin DNA izolasyonları *MagPurix 12s Instrument, Automated Nucleic Acid Purification System* (Katalog no: ZP01001) kullanılarak, *MagPurix Blood DNA Extraction Kit 200* ile gerçekleştirildi.

Kas ve idrar örneklerinden DNA izolasyonu için *Thermo Scientific™ Invitrogen™, PureLink™ Genomic DNA Mini Kit* (Katalog no: K182001) kullanıldı. İdrar örnekleri santrifüj edildi ve elde edilen peletten DNA izolasyonu, *Mammalian Cells Lysate*; kas örneklerinden DNA izolasyonu, *Mammalian Tissue and Mouse/Rat Tail Lysate* protokolleri takip edilerek gerçekleştirildi.

3.1.4. DNA Yoğunluk ve Kalite Kontrolü

İzole edilen DNA örneklerinin yoğunluğu ve kalitesi *Thermo Scientific™ NanoDrop 2000/2000c* spektrofotometre kullanılarak belirlendi, yoğunluk (ng/μl) ve absorbans değerleri (A260/280, ~1,8) kaydedildi.

A260/280 değeri 1,8'den az olan örnekler protein degradasyon işlemine alındı. Çöktürme ve saflaştırma işlemleri uygulandı. 2 mM amonyum asetat ($\text{NH}_4\text{CH}_3\text{CO}_2$) ve 2,5 hacim %100 etanol ($\text{C}_2\text{H}_5\text{OH}$) ile çöktürüldü. Pelet, iki kez %80'lik, bir kez %100'lük etanolle yıkandı. Kurutma işleminden sonra, dH_2O 'da çözüldü.

A260/280 değeri 1,8'den fazla olan örnekler RNA degradasyon işlemine alındı. RNAase enzimi kullanılarak RNA'ların yıkımı sağlandı.

3.1.5. Primer Tasarımı

16.569 bç'nden oluşan mtDNA, her biri 4000-6000 baz aralığında 4 parçaya ayrılmıştır. Her bir parça için ileri yönlü ve geri yönlü primerler *UCSC In-Silico PCR*, *NCBI Primer-BLAST* ve *IDT OligoAnalyzer Tool* kullanılarak, her biri 20-22 bç uzunluğunda *in house* tasarlanmıştır.

3.1.6. MtDNA Dizilerinin PZR ile Çoğaltılması

İzolasyonları ve saflaştırma işlemleri tamamlanan örneklerde hedef bölge çoğaltma işlemi steril ortamda rekombinant Taq DNA Polimeraz (5 U/ μL) (*Thermo Scientific*TM Katalog no: EP0401) enzimi ve 10x tamponu, dNTP karışımı (25 mM *each* (her biri)) (Katalog no: R1121), tasarlanan ileri ve geri yönlü primerler kullanılarak uygulandı. Kullanılan malzemeler ve uygulanan protokol Tablo 3-1 ve Tablo 3-2'de belirtildiği gibidir.

Table 3-1: PZR malzemeleri ve son hacim konsantrasyonları

Malzemeler	Final Konsantrasyon (50 μL)
dH_2O (18 M Ω)	-
10xTampon	1x
MgCl_2 (25 mM)	2 mM
dNTP (25mM)	0,2 mM
İleri Primer	0,2 μM
Geri Primer	0,2 μM
DMSO	%5
Taq DNA Polimeraz	0,3 μL
DNA Miktarı	Kan/İdrar: 0,1 μg , Kas: 0,05 μg

Table 3-2: PZR protokolü

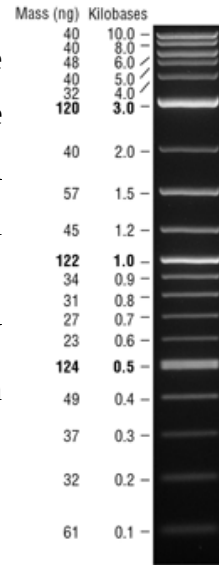
İşlem	Derece	Süre	Döngü
Çift Zincir Ayrılması	95°C	10 dk	1
Çift Zincir Ayrılması	95°C	30 sn	
Primer Bağlanma	60°C	30 sn	10
Uzama	72°C	6 dk	
Çift Zincir Ayrılması	95°C	30 sn	
Primer Bağlanma	58°C	30 sn	30
Uzama	72°C	6 dk	
Son Uzama	72°C	10 dk	1
Saklama	4°C	∞	1

3.1.7. PZR Ürünlerinin Saflaştırması

Hedef bölge çoğaltma işlemi sonrası, mtDNA örneklerin purifikasyon işlemi filtreleme tekniğiyle, *Roche High Pure PCR Product Purification Kit*'i (Katalog no: 11 732 676 001) kullanılarak gerçekleştirildi. Son hacim 30 µL olarak belirlendi.

3.1.8. Agarol Jel Elektroforezi ile Görüntüleme

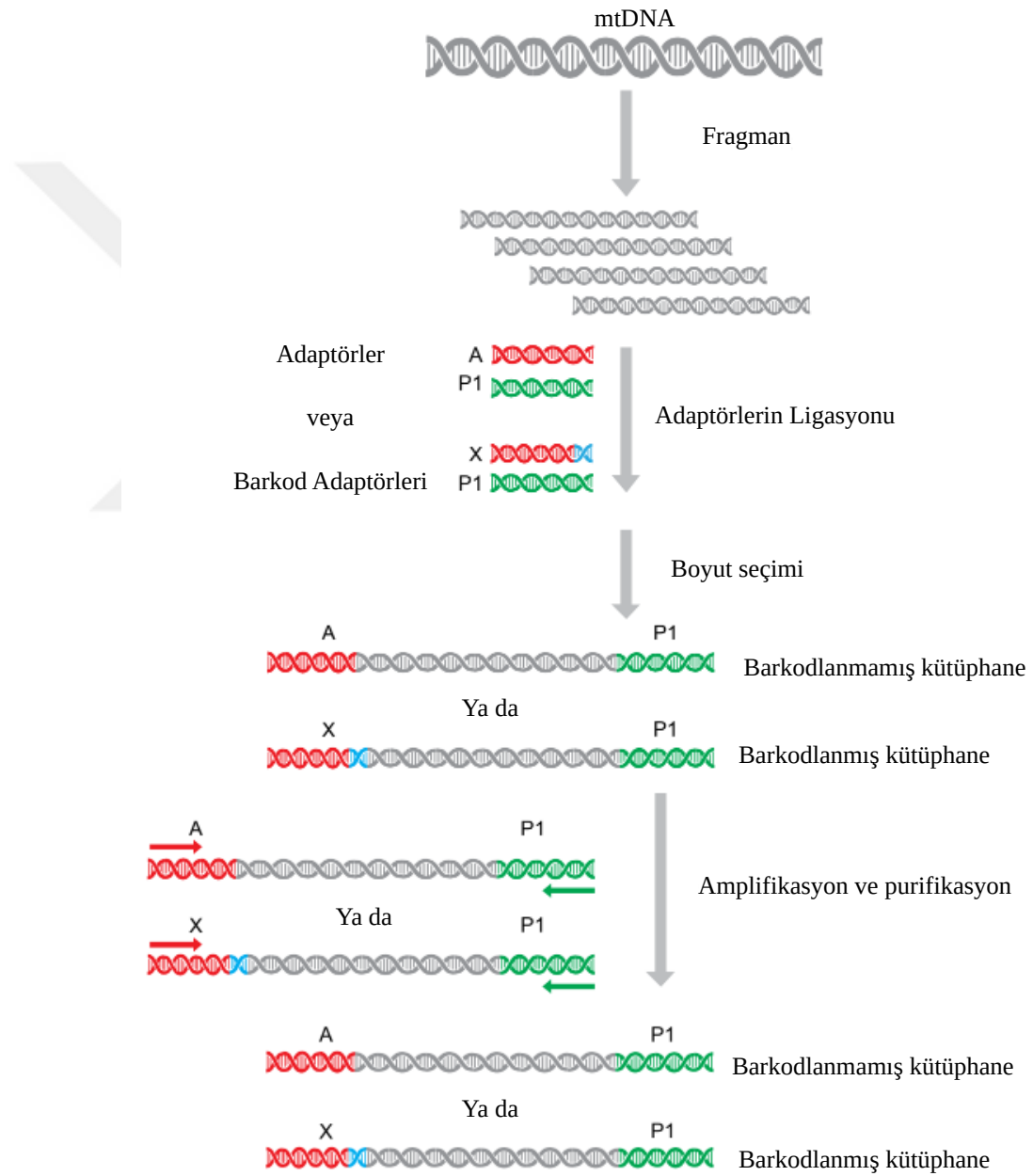
Purifikasyon sonrasında örnekler, kalite tayini için jele yüklendi. *Sigma Agarose A-5093* ve 1X TBE tamponu (tris, borik asit ve EDTA çözeltisi) kullanılarak %0,7'lik, 80 kuyuluk jel hazırlandı. İlk kuyuya *NEB Quick-Load® Purple 1 kb DNA Ladder* (N0550G) yüklendi. UV transilluminatörde görüntüleme sağlandı. Bantların kontrolü Şekil 3-1'de verilen kılavuz ile yapıldı.



Şekil 3-1: NEB Quick-Load® Purple 1 kb DNA Ladder (N0550G) bant kılavuz görüntüsü

3.1.9. Yeni Nesil Dizileme İşlemleri

Elde edilen hedef bölgelerin dizileme işlemi *Thermo Scientific™ Ion GeneStudio S5 System* platformu ile *Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide (Revision: H.0)* ve *Ion 510™ & Ion 520™ & Ion 530™ Kit - Chef User Guide (Revision: E.0)* kılavuzlarına uygun olarak gerçekleştirildi. Platformun akış şeması Şekil 3-2’de verilmiştir.



Şekil 3-2: Thermo Scientific™ Ion GeneStudio S5 System akış şeması

Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide, Sayfa: 18

Heteroplazmi incelemesinde YND tekniđi kullanılarak 13 farklı mitokondriyal genom dizilemesi eş zamanlı çalışıldı ve 16.569 bç uzunluğundaki mtDNA dizilendi. YND tekniğinde kullanılan çiplerin 1000-4000 kuyu kapasitesi, eş zamanlı, çoklu örnek çalışmasına olanak sağladı. YND'nin kullanıldığı, tüm genom dizileme (TGD), tüm ekzom dizileme (TED) ve gen panel çalışmaları hata oranı düşük, hızı yüksek ve maliyeti uygun özellikleri ile tanısal testlerde ve araştırma çalışmalarında tercih edilmektedir [146]. Bu avantajlara paralel olarak YND, mtDNA'da heteroplazmi oranının belirlenmesinde, özellikle %1'lik heteroplazmi oranlarının saptanmasında, benzer metotlara göre, doğruluk oranı yüksek olduğu için tercih edilmektedir [147] [148]. Çoklu paralel dizileme ile güvenilirliği yüksek ve ölçeklenebilir olduğu için bu çalışmada YND tekniđini kullanıldı.

3.1.9.1. DNA Konsantrasyonun Hesaplanması ve Örnek Hazırlığı

Dört ayrı bölge halinde çoğaltılan ve saflaştırılan mtDNA fragmanları, her bir hastanın her bir dokusu için, son konsantrasyon 100 ng olacak şekilde 0,2 mL'lik tüpte birleştirildi. Son hacim 35 µL olacak şekilde her bir tüpe dH₂O (18 MΩ) eklendi.

3.1.9.2. Fragmentasyon

Her bir tüpte bulunan 4.000-6.000 bç uzunluğundaki mtDNA parçalarını 350-450 bç uzunluğunda parçalara kesme işlemi *Ion Xpress™ Plus Fragment Library Kit* kullanılarak yapıldı. Tüplere *Ion Shear™ Plus 10X* reaksiyon tamponu eklendi. *Ion Shear™ Plus* Enzim Karışımı II eklendikten sonra *Veriti™ 96-Well Thermal Cycler* (Katalog no: 4375786) cihazında 37°C'de 2 dk bekletildi. Enzim aktivitesini durdurmak için *Ion Shear™ Plus* durdurma tamponu kullanıldı ve örnekler buza alındı. Protokole uygun olarak yapılan işlemlerde kullanılan miktarlar Tablo 3-3'te belirtildiđi gibidir.

Table 3-3: *Ion Xpress™ Plus Fragment Library Kit* Malzemeleri ve uygulanan miktarlar

Malzeme	Miktar
Ion Shear™ Plus 10X Reaksiyon Tamponu	5 µL
Ion Shear™ Plus Enzim Karışımı II	10 µL
Ion Shear™ Plus Durdurma Tamponu	5 µL
Düşük TE Solüsyonu	25 µL

Fragmente Edilen mtDNA'nın Purifikasyonu

Saflaştırma işlemi protokole uygun olarak gerçekleştirildi. Örnekler buzdan alınarak 1,5 mL'lik tüplere alındı. 500'den az baz-okumaları için uygun olarak belirtilen miktarda (99 µL) *Agencourt™ AMPure™ XP Reagent* manyetik boncuk eklemesi yapıldı ve 5 dk beklendi. Tüpler *DynaMag™-2* magnete alındı ve 3 dk bekletildi. Manyetik boncuklara bağlanan mtDNA'nın tutulumu sağlandı ve süpernatant atıldı. Tüpler magnet üzerindeyken 500 µL %70'lik alkolle iki kez yıkama işlemi gerçekleştirildi. Yıkama esnasında tüpler 180° 4-5 kez döndürüldü. İşlem sonrası alkol atıldı ve alkolün tamamen uçması için tüplerin kapakları açıldı ve 5 dk beklendi. Fragmanların saklanabilmesi için pelete 25 µL düşük Tris EDTA (*Low TE*) solüsyonu eklendi. Örnekler 0,2 µL'lik tüplere alındı. Bir sonraki işleme kadar -20°C'ye kaldırıldı.

3.1.9.3. Adaptör Bağlama ve Barkodlama

25 µL düşük TE solüsyonu içerisindeki mtDNA örneklerine, Tablo 3-4'te belirtilen malzemelerle muamele edildi ve pipetaj yapıldı. Her bir örnek için ayrı *Ion Xpress™ Barcode X* kullanıldı. Barkod ve adaptör bağlama işlemi için Ligasyon PZR uygulandı. Ligasyon PZR protokolü Tablo 3-5'te belirtildiği gibidir.

Table 3-4: *Ion Xpress™ Plus Fragment Library Kit* ve *Barcode Adapters Kit* Malzemeleri ve miktarları

Malzeme	Miktar
10X Ligaz Tamponu	10 µL
Ion Xpress™ P1 Adaptör	2 µL
Ion Xpress™ Barkod X (kütüphane başına 1 barkod adaptörü)	2 µL
dNTP Karışımı	2 µL
dH ₂ O (18 MΩ)	49 µL
DNA Ligaz	2 µL
Nick Repair Polimeraz	8 µL

Table 3-5: Bağlanma (Ligasyon) PZR protokolü

İşlem	Derece	Süre
Karışma	25°C	15 dk
Uzama	72°C	5 dk
Saklama	4°C	∞

Adaptör Bağlanmış Ürünlerin Purifikasyonu

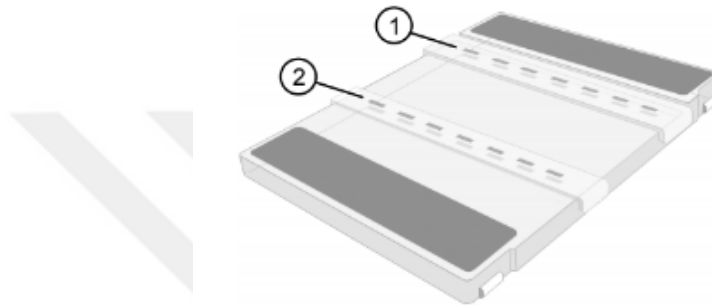
Bağlanmayan barkod ve adaptörlerin uzaklaştırılması amacıyla saflaştırma işlemi protokole uygun olarak gerçekleştirildi. Örnekler 1,5 mL'lik tüplere alındı. 400-600 baz-okumaları için uygun olarak belirtilen miktarda (100 µL) *Agencourt™ AMPure™ XP Reagent* manyetik boncuk eklemesi yapıldı ve 5 dk bekletildi. Tüpler *DynaMag™-2* magnete alındı ve 3 dk bekleme sonrasında süpernatant atıldı. Magnet üzerinde %70'lik alkolle iki kez temizleme işlemi uygulandı. Temizleme süresince tüpler 180° 4-5 kez döndürüldü. Alkol atıldı ve tamamen uçması için tüpler, kapakları açılarak 5 dk bekletildi. Pelet üzerine 20 µL düşük TE eklendi ve örnekler 0,2 µL'lik PZR tüplerine alındı. Bir sonraki işleme kadar -20°C'ye kaldırıldı.

3.1.9.4. Hazırlanan Kütüphanede Boyut-Seçim İşlemi

İşlem *Size-select the library with the E-Gel™ SizeSelect™ Agarose Gel* protokolüne uygun olarak yapıldı. İşlem için gerekli malzemeler Tablo 3-6'da belirtilmiştir. DNA kütüphane boyutu 400 baz-okuması (*base-read*) uzunluğunda belirlendi. *Invitrogen 50-bp DNA Ladder* 1:40 oranında dH₂O (18 MΩ) ile seyreltilti. *E-Gel™ SizeSelect™ 2% Agarose Gel*, *E-Gel™ iBase™* cihazına yerleştirildi. Örnek kuyularından 1. ve 8. kuyuya ve geri kazanım kuyularının tamamına 25'er µL dH₂O (18 MΩ) eklendi. 20'şer µL düşük TE solüsyonu içerisinde bulunan örnekler örnek kuyularına (2-7), 10 µL dilüe edilmiş *Invitrogen 50-bp DNA Ladder* birinci sıradaki markör kuyusuna, 10 µL dH₂O (18 MΩ) ikinci sıradaki markör kuyusuna olmak üzere *E-Gel™ SizeSelect™ 2% Agarose Gel* (Şekil 3-3) yüklemeleri yapıldı. Cihazda jel tipi *SizeSelect 2%* olarak seçildi. Örneklerin yürüme zamanı 15 dk olarak belirlendi ve yürüme başlatıldı. DNA *Ladder* referans bandı olarak kullanıldı. 500 bç uzunluğundaki fragmanlar referans çizgisine yaklaştığında yürüme durduruldu (Şekil 3-4). Mavi renk transilüminatör aktive edildi ve yürüme tekrar başlatılarak referans bandın geri kazanım kuyusuna denk gelmesi kontrollü şekilde sağlandı (Şekil 3-5). Toplam yürüme süresi 20 dk 58 sn olarak kaydedildi. Geri kazanım kuyularından örnekler toplandı. Yıkama işlemi için kuyulara 10'ar µL *UltraPure™ DNase/RNase-Free Distilled Water* eklendi ve kuyularda kalmış olabilecek örnekler toplandı.

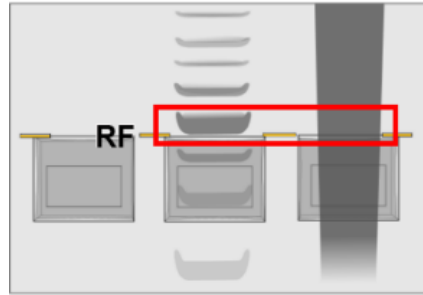
Table 3-6: Boyut-seçim işlemi için gerekli malzemeler

Malzeme
Düşük TE Solüsyonu (Ion Plus Fragment Library Kit ürünü)
E-Gel™ iBase™ unit and E-Gel™ Safe Imager™ transilluminator combo kit
E-Gel™ SizeSelect™ 2% Agarose Gel
Invitrogen 50-bp DNA Ladder - Cat.No. 10416-014 (400-base-read libraries)
UltraPure™ DNase/RNase-Free Distilled Water

**Şekil 3-3: E-Gel™ SizeSelect™ 2% agaroz jel elektroforezi görüntüsü**

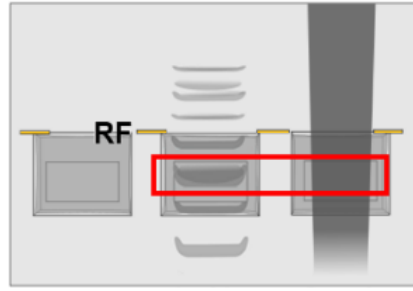
1. Örnek kuyuları, 2. Geri kazanım kuyuları

Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide, Sayfa: 50

**Şekil 3-4: Fragmanların referans çizgisine yaklaşması**

Sırasıyla; ladder, örnek. RF: Referans Çizgisi

Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide, Sayfa: 52



Şekil 3-5: Fragmanların referans çizgisine denk gelmesi

Sırasıyla; ladder, örnek. RF: Referans Çizgisi

Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide, Sayfa: 53

3.1.9.5. Kütüphanenin Çoğaltılması ve Saflaştırılması

İşlemler, *Amplify and purify the library* protokolü takip edilerek gerçekleştirildi. Kütüphane girdi miktarı 50-100 ng için prosedürde belirtildiği gibi 25 µL'dir. Tablo 3-7'de belirtilen malzemeler, belirtilen miktarlarda birleştirilerek karışım oluşturuldu. Son hacim olan 130 µL, iki adet 0,2 mL'lik tüplere paylaştırılarak Tablo 3-8'de belirtilen protokolle, PZR işlemi için *Veriti™ 96-Well Thermal Cycler* (Katalog no: 4375786) cihazına alındı. PZR işlemi tamamlandıktan sonra tüpler 1,5 mL'lik türlerde birleştirildi.

Table 3-7: Kütüphane amplifikasyonu malzemeleri ve kullanılan miktarlar

Malzeme ve ürün	Miktar
Platinum™ PCR SuperMix High Fidelity	100 µL
Library Amplification Primer Mix	5 µL
Unamplified library	25 µL
Son Hacim	130 µL

Table 3-8: 100-500 baz-okuma kütüphanesi için PZR protokolü

İşlem	Derece	Süre	Döngü
Çift Zincir Ayrılması	95°C	5 dk	1
Çift Zincir Ayrılması	95°C	15 sn	
Primer Bağlanma	58°C	15 sn	8
Uzama	70°C	1 dk	
Saklama	4°C	∞	1

Hazırlanan Kütüphanenin Saflaştırılması:

Örnekler 1,5 mL'lik tüplere alındı. Saflaştırmada kullanılan *Agencourt™ AMPure™ XP Reagent* miktarı, protokolde verilen tabloya göre (*Ion Xpress™ Plus gDNA Fragment Library Preparation User Guide*, Sayfa: 65), 50-100 ng örnek, 400-500 baz-okuma kütüphanesine uygun olarak 130 µL seçilmiştir. 130 µL *Agencourt™ AMPure™ XP Reagent* örneklerle eklendi ve 5 dk oda sıcaklığında bekletildi. Tüpler manyetik tablaya (*DynaMag™-2* magnet) yerleştirildi ve 3 dk bekletildi. Süpernatant atıldı. Temizleme işlemi için %70'lik alkol kullanıldı. Tüpler manyetik tabla üzerindeyken alkolle 2 kez yıkandı. Yıkama esnasında tüpler 180° 4-5 kez döndürülerek etkin temizleme sağlandı. Alkol atıldı. Alkolün tamamen uçması için tüpler kapakları açık 5 dk bekletildi. Tüplere 20 µL düşük TE eklendi. Örnekler 0,2 mL'lik tüplere alındı ve bir sonraki işleme kadar -20°C'de muhafaza edildi.

3.1.9.6. Oluşturulan Barkodlanmış Kütüphanenin Değerlendirmesi

İşlemler, *Ion Library TaqMan® Quantitation Kit* (Katalog no: 4468802, Revizyon: B.0) kullanılarak, protokole uygun olarak gerçekleştirildi. Örnek kütüphanelerinin yoğunluk değeri hesaplamasında referans olarak kullanılacak olan *E.Coli DH10B Kontrol Kütüphanesi* standardı seri dilüsyonla hazırlandı (Tablo 3-9). Hazırlanan örnek kütüphanelerin dilüsyonu, Kontrol Kütüphanesi standardı aralığına denk gelecek şekilde 1:2000 oranında yapıldı. Örnek başına 15 µL olacak şekilde karışım hazırlandı (Tablo 3-10) ve PZR tabakasındaki (*PCR plate*) kuyucuklara eklendi. Karışımın üzerine 5 µL dilüe edilmiş örnek kütüphaneleri eklendi. Örnek kütüphanelerin ve Kontrol Kütüphanesi standartlarının Tablo 3-11'de belirtilen protokole göre, *Applied Biosystems™ real-time PCR* cihazı kullanılarak gerçek-zamanlı PZR (qPZR)'ı yapıldı. qPZR'da gözlemlenen standart eğri ile, PZR ürünlerinden alınan floresan sinyalinin, arka plan sinyalinin üzerine çıktığı döngü sayısını veren reaksiyonun eşik döngüsü sayısı (Ct: *cycle threshold*) 0,2 olarak saptandı. Barkodlanmış her bir kütüphanenin pmol/L cinsinden molar konsantrasyonunu tespit etmek için Ct değeri, dilüsyon oranıyla çarpıldı (0,2 x 2000) ve kütüphane konsantrasyonu 200 pM bulundu. Örnek kütüphaneleri, *Ion Chef™ System*'de kullanım için 100 pM'a dilüe edildi.

Table 3-9: Seri dilüsyonla hazırlanmış *E.Coli* DH10B Kontrol Kütüphanesi standartları dilüsyon ve konsantrasyon değerleri

Standart	Kütüphane girdisi	Nükleazdan arındırılmış su	Dilüsyon katı	Konsantrasyon
1	5 µL (dilüe edilmemiş)	45 µL	0,1	6,8 pM
2	5 µL (standart 1'den)	45 µL	0,01	0,68 pM
3	5 µL (standart 2'den)	45 µL	0,001	0,068 pM
4	5 µL (standart 3'ten)	45 µL	0,0001	0,0068 pM
5	5 µL (standart 4'ten)	45 µL	0,00001	0,00068 pM

Table 3-10: Ion Library TaqMan® Quantitation Kit malzemeleri ve miktarlar

Malzeme ve ürün	Miktar
Ion Library TaqMan® qPCR Mix, 2X	10 µL
Ion Library TaqMan® Quantitation Assay, 20X	1 µL
Nuclease-free Water	4 µL

Table 3-11: Gerçek-zamanlı PZR protokolü

İşlem	Derece	Süre
İnkübasyon	50°C	2 dk
Polimeraz aktivasyonu	95°C	20 sn
40 Döngü	95°C	1 sn
	60°C	20 sn

3.1.9.7. Otomatize Şablon Hazırlama, Çip Yükleme ve Dizileme

Çalışmanın yönetim işlemleri *Torrent Suite™ Software* kullanılarak *Ion 510™ & Ion 520™ & Ion 530™ Kit - Chef* kullanıcı rehberine uygun gerçekleştirildi. Plan sekmesindeki şablon hazırlama ekranından çalışılmak istenilen uygulama, mitokondri olarak seçildi ve çalışmaya uygun referans dizi ve BED dosyası eklenerek çalışma planı oluşturuldu. Plana göre 100 pM'e dilüe edilmiş örnek kütüphaneleri *Ion Chef™* sistemi ile çipe yüklendi. Yükleme işlemi gerçekleştikten sonra çip, *Ion S5™ Sequencer* dizileme cihazına alındı. Gerekli solüsyonlar cihaza yerleştirildi ve dizileme işlemi başlatıldı.

3.1.10. Dizileme Verilerinin Analizi

Dizileme sonrasında elde edilen ham veriler, analizi *Ion Reporter™ Software* (yazılım)'de (versiyon: 5.12.2.0) gerçekleştirmek üzere, Torrent Server'dan Ion Reporter™ Server'a aktarıldı.

3.1.10.1. Akış Şeması Oluşturma

Ham verilerin ön analizi için Ion Reporter™ Yazılımı'nda akış şeması (*Create Workflow*) oluşturuldu. İzlenen adımlar şu şekildedir:

1. Araştırma Uygulaması (*Research Application*):
 - Araştırma uygulaması (*Research Application*): DNA
 - Örnek grubu (*Sample Groups*): Tek (*Single*)
2. Referans:
 - Referans dizi: hg19
 - Odak Analiz (*Focus Analysis*):
 - Hedef Bölgeler (*Target Regions*): Mitokondri BED dosyası (*.bed*)
 - Varyant türü belirleme (*Variant Type Detection*): Somatik (*Somatic*)
 - Hotspot Bölgeler (*Hotspot Regions*): Boş
3. Açıklama (*Annotation*):
 - Açıklama Seti (*Annotation Set*): Hepsi (*All*)
4. Filtreler (*Filters*):
 - Filtre Zinciri (*Filter Chains*): Filtre yok (*No Filter*)
5. Kopya Sayısı (*Copy Number*):
 - Ana Hat (*Baseline*): Ana Hat Yok - CNV'leri çağırma
(*No Baseline - Don't call CNVs*)
6. Eklentiler (*Plugins*):
 - Analiz içi eklentiler (*In-Analysis Plugins*): Eklenti kullanılmamıştır.

7. Final Raporu (*Final Report*):

- Final Rapor Şablonu (*Final Report Template*): Ion Varsayılan Final Rapor Şablonu (*Ion Default Final Report Template*)

8. Parametreler (*Parameters*):

Ion Reporter'ın genom analizi için optimize edilmiş, varsayılan olarak bulunan parametreler kullanıldı. Parametrelerde mitokondriyal varyantların heteroplazmik durumu dolayısıyla bir takım değişiklikler yapıldı. Parametrelerdeki değişimlerle, varyantların her tipi için, varyantın her iki sarmalda varlığından emin olunabilir miktarda okuma bulunması ve heteroplazmi oranı düşük varyantların da yakalanabilir olması amaçlandı. Bu amaçlarla tek nükleotid varyantları (SNV), insersiyon - delesyon (indel) tipi değişimler ve çoklu nükleotid yapı değişimlerinin tespiti için minimum okuma sayısı, her bir sarmaldaki en düşük okuma sayısı ve minimum alel frekansı değerlendirildi ve değişiklikler, önerilen değerler göz önünde bulundurularak yapıldı (Tablo 3-12, 3-13).

Table 3-12: Ion Reporter Yazılımı'nda akış şemasının oluşturulduğu bölümde bulunan Parameters sayfası, *Variant Finding/Main* başlığı altında yapılan değişiklikler

Parametre	İzin Verilen Değer	Önerilen Değer	Varsayılan Ayar	Değişiklik
Snp Min Cov Each Strand	$x \geq 0$	$x \geq 3$	4	3
Snp Min Allele Freq	$0 \leq y \leq 1,0$	$0,01 \leq y \leq 0,2$	0,02	0,01
Snp Min Coverage	$x \geq 0$	$5 \leq x \leq 20$	100	20
Indel Min Cov Each Strand	$x \geq 0$	$x \geq 3$	4	3
Indel Min Allele Freq	$0 \leq y \leq 1,0$	$0,05 \leq y \leq 0,2$	0,02	0,01
Indel Min Coverage	$x \geq 0$	$15 \leq x \leq 30$	100	15
Mnp Min Cov Each Strand	$x \geq 0$	$x \geq 3$	4	3
Mnp Min Allele Freq	$0 \leq y \leq 1,0$	$0,01 \leq y \leq 0,2$	0,02	0,01
Mnp Min Coverage	$x \geq 0$	$5 \leq x \leq 20$	100	15

x = Tam sayı, y = Ondalık sayı

Table 3-13: Ion Reporter Yazılımı'nda akış şemasının oluşturulduğu bölümde bulunan Parameters sayfası, Variant Finding/Advanced başlığı altında yapılan değişiklikler

Parametre	İzin Verilen Değer	Önerilen Değer	Varsayılan Ayar	Değişiklik
Allow Complex	True, False	True	False	True
Output MNV	True, False	False	False	True

9. Onay (*Confirm*):

- Akış şemasını isimlendirme (*Workflow Name*)
- Tanımlama (*Description*)
- Akış şemasının kaydı (*Save Workflow*)

3.1.10.2. Ön Analizi Başlatma

Ion Reporter™ Software üzerinde, analiz başlatıldı (*Analysis>Launch*).

Aşamaları şu şekildedir:

1. Akış şeması seçimi (*Workflow*): Oluşturulmuş olan akış şeması eklendi.
2. Örnek seçimi (*Samples*): Toplam 13 örnek seçildi.
3. Eklentiler (*Plugins*): Eklenti kullanılmadı.
4. Onaylama ve analizi başlatma (*Confirm & Launch*): Her bir örnek için analiz isimleri (*Analysis Name*) kontrol edildi ve analiz başlatıldı (*Launch Analysis*).

Analiz tamamlandıktan sonra (*Analysis Status: Successful*) değerlendirme aşamasına geçildi.

3.1.10.3. Varyant Dosyalarına Erişim ve Düzenleme

Yazılım üzerindeki analiz (*Analysis*) sekmesinde bulunan genel görünüm (*Overview*) ekranındaki dosya formatında mevcut analiz (*Analysis*) sütununda, analizi tamamlanmış örnekler üzerine tıklanarak, varyantlar görüntülendi. Tercihler (*Preferences*) ikonu tıklanarak sütunların seçimi yapıldı. Seçimi yapılan sütunlar şu şekildedir:

- Lokus (*Locus*)
- Genotip (*Genotype*)

- Referans Alel (*Ref*)
- Gözlemlenen Alel (*Observed Allele*)
- Tür (*Type*)
- % Frekans (*% Frequency*)
- Alel Kapsama (*Allele Coverage*)
- Kapsama (*Coverage*)
- Referans Alelin ve Varyantın Derinliği (*Ref+/Ref-/Var+/Var-*)
- Alel Oranı (*Allele Ratio*)
- p-değeri (*p-value*)
- Kalite Skoru (*Phred QUAL Score*)

İndirme (*Download*) iconu tıklandı ve verilen seçenekler içinden *Current Results TSV* (güncel TSV sonuçları) seçildi. Ana sayfa altında bulunan bildirimler (*Notifications*) sayfasına yönlendirilen dosya burada seçildi ve indirildi (*Download*). Bu işlem her bir örnek için yapıldı. İlgili sütunlar dosya yerleşiminde ön sıralarda olacak şekilde dosyalar şekillendirilmiş ve indirilmiş oldu.

Örneklere ulaşılan analiz genel görünüm (*Analysis/Overview*) ekranında, aynı gönüllüye ait farklı doku örnekleri seçildi ve görselleştirme (*Visualize*) iconu tıklandı. Doku örneklerinin varyantlarının bir araya getirildiği dosyaya ulaşıldı. İndirme (*Download*) ikonundan *All Variants* (tüm varyantlar) seçeneği tıklandı. Ana sayfada bulunan bildirimler ekranında birleştirilmiş olan dosya seçildi ve indirildi (*Download*). Bu işlem aynı gönüllüye ait tüm doku örnekleri için yapıldı.

Örneklere tıklandığında ulaşılan varyant listesindeki herhangi bir varyanta tıklandı. Yeni sekmede açılan IRGV sayfanın sağ alt köşesinde bulunan *IRGV Export & Preferences* (IRVG dışa aktarım ve tercihler) ikonuna tıklandı ve *Open in IGV (download .jnlp file)* (IGV’de Aç (.jnlp dosyasını indir)) seçildi. Her bir örneğin .jnlp uzantılı dosyası indirildi. Varyantlar, dosyaların IGV 2.8.0 yazılımı kullanılarak açılmasıyla görüntülendi.

3.1.10.4. Analiz Süreci

İndirilen örnek dosyaları açıldı. Manuel incelemeyle verilerin parametrelere uygun olduğu doğrulandı. Homoplazmik ve olası heteroplazmik varyantlar belirlendi. IGV yazılımı üzerinde incelemeler yapıldı. Homoplazmik varyantlar için referans heteroplazmisi durumu sorgulandı. Heteroplazmik varyantlar, varyantın saptandığı bölgedeki diğer baz okumalarının temizliği ve varyantın ileri yönlü ve geri yönlü okumalarının en az 3'er tanesinin temiz okunmuş olması durumu sorgulanarak değerlendirildi. Mevcut parametrelere karşın Ion Reporter çıktısında yer alan, ancak IGV'de temiz okuma göstermediği anlaşılan değişimler elendi. Rapora göre aynı gönüllünün yalnızca bir dokusunda var olduğu belirtilen her bir değişim için diğer doku/dokular IGV üzerinden kontrol edildi ve değişimin diğer dokuda/dokularda da var olduğunun saptandığı durumlarda bu değişimler de frekans hesaplamasına eklendi.

Örnek dosyaları için bu işlemler yapıldıktan sonra her bir gönüllünün farklı dokuların varyantlarının birleştirildiği dosyada homoplazmik, heteroplazmik ve referans heteroplazmisi durumları üzerinden sınıflandırma yapıldı.

Her bir gönüllü için bulgular dosyası oluşturuldu. Homoplazmik varyantlar ve sayıları tablo şeklinde bu dosyaya kaydedildi. Olası her bir heteroplazmik varyant için Ion reporter, IGV ve MITOMAP bilgileri her doku için girildi ve değerlendirme sonucu yazıldı. Değerlendirme sonucu heteroplazmik olan varyantlar tablo biçiminde kaydedildi ve varyantın her bir dokudaki varlığı belirtildi. Venn şeması oluşturularak dokularda ortak olan ve yalnızca tek dokuda bulunan heteroplazmik varyant sayısı gösterildi.

Heteroplazmik varyantlar için, Ion Reporter ve IGV platformları arasında karşılaştırma yapıldı. Varyant, dokular, her bir doku için o varyantın derinlik bilgisi, kalite skoru ve frekans bilgileri ile tablo oluşturuldu. Değerlendirmeler yapılarak sonuçlar kaydedildi.

3.2. Veri Madenciliği ve Makine Öğrenmesi Çalışmaları

MITOMAP (*FoswikiMITOMAP WebWebHome (05 Dec 2019, MarieLott)*) veri bankasından, m.3243A>G patojenik varyantı için verilmiş olan kaynak listesi indirildi. İngilizce makalelerde, aşağıda belirtilen anahtar kelimeler kullanılarak taramalar yapıldı:

- 3243
- Heteroplasmy
- Mutation Load
- Percentage
- %
- Muscle
- Blood
- Case
- Patient
- Clinic
- Mother
- Sister
- Brother
- History

Erişilen yayımlanmış makalelerde m.3243A>G değişimi saptanmış hastaların klinik bilgileri mevcutsa ve heteroplazmi oranı olarak patojenik varyant yükünün belirtimi yapıldıysa, makale ayrıntılı incelemeye alındı. Hastaların demografik özellikleri, klinik bulguları, dokuya özgü heteroplazmi oranları, tüm bu özellikler hesap tablosu oluşturmak için *LibreOffice Calc* programında oluşturulan sütunların altına kaydedildi. Her bir sütun makalelerde karşılaşılan hasta klinik özellikleri ve bulgularından her biri için hazırlandı. Bu sütunların her biri, birer öznitelik (*feature*) olarak değerlendirildi. Tüm makaleler bu özellikler için tarandıktan sonra sütun düzenlemeleri yapıldı. Makine öğrenmesi metotlarının veri setinde uygulanabilmesi için verilerin tamamının sayısal değerlerden oluşması gerektiğinden, klinik bulgular *var* ya da *yok* olarak düzenlendi. MELAS tanı kriteri dikkate alınarak bu durum ile ilişkili hastalıkların her biri tek bir sütun olarak ele alındı. Bir hastalığın alt bulguları niteliğinde olan özellikler sınıflandırılıp tek bir sütunda toplandı.

Klinik bulgular ve belirtiler, biyokimyasal testlerin, kas histokimya analizlerinin ve beyin görüntüleme sonuçları ile doku heteroplazmi oranları arasındaki ilişkinin saptanabilmesi için oluşturulan veri setinde çoklu-doğrusal regresyon analizi yapıldı. Makalelerden elde edilen ham verinin ön-işleme süreci, model oluşturma ve yorumlama adımları için *Python* (versiyon: 3.5.3) program dili kullanıldı. Çalışma alanı (*ide: integrated development environment* = tümleşik geliştirme ortamı) olarak *Jupyter Notebook* tercih edildi.

3.2.1. Kullanılacak Python Paketleri

Kullanılacak olan Python kütüphaneleri çalışma alanına yüklendi [149].

- Pandas
- Numpy
- SciPy
 - Stats
- Math
- Seaborn
- Missingno
- Matplotlib
- PyLab
- StatsModel
 - statsmodel.api
 - statsmodels.stats
 - diagnostic
 - statsmodels.stats.stattools
 - durbin_watson
 - statsmodel.stats.outliers_influence
 - variance_inflation_factor

- SkLearn
 - sklearn.linear.model
 - LinearRegression
 - sklearn.model_selection
 - train_test_split
 - sklearn.metrics
 - mean_squared_error
 - r2_score
 - mean_absolute_error
- Pickle

4. BULGULAR

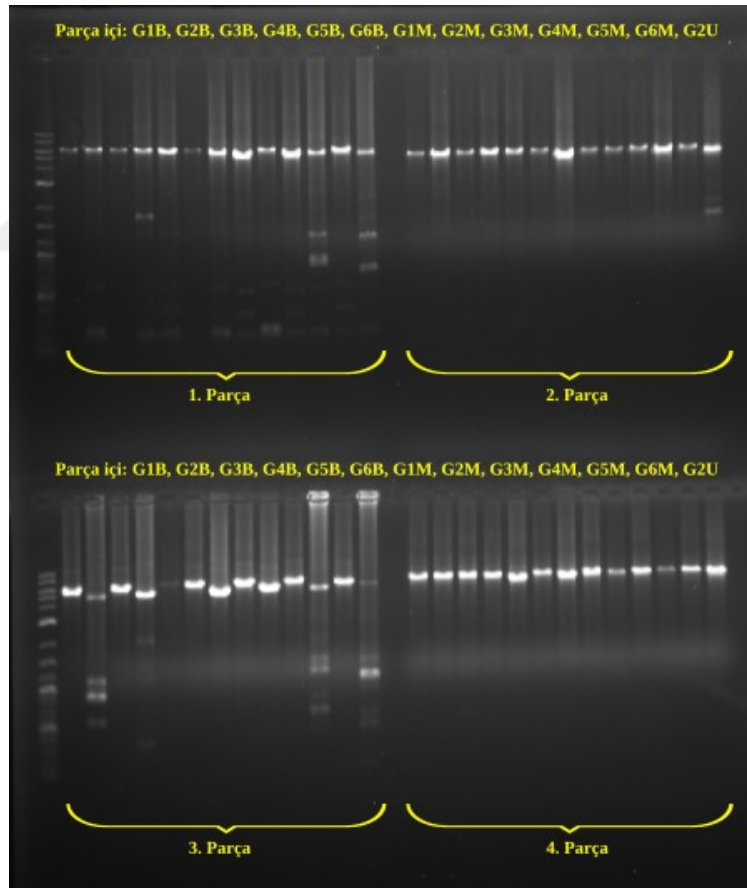
4.1. Laboratuvar Çalışma Sonuçları

4.1.1. Örnek Grubu

Çalışma, 5 sağlıklı gönüllünün atık kan ve kas dokularından, 1 gönüllünün atık kan, kas ve idrar dokularından izole edilen toplam 13 mtDNA (6 x 2 + 1 = 13) örneğinde YND yapılarak gerçekleştirildi.

4.1.2. PZR Sonuçları

Dört parçada çoğaltılmış mtDNA, kolon saflaştırma işleminden sonra %1'lik agaroz jel elektroforezi ile görüntülendi (Şekil 4-1).



Şekil 4-1: Dört ayrı parçada çoğaltılan mtDNA'nın kolon saflaştırma işleminden sonra %1'lik agaroz jel elektroforezi ile alınan görüntüsü

13'er setler halinde 1., 2., 3. ve 4. parça.

13 örnekte sıralama: soldan sağa: G1 → G2 → G3 → G4 → G5 → G6;

6 örnek: kan, 6 örnek: kas, 1 örnek idrar. İdrar örneği yalnızca G2'de çalışılmıştır.

B: Kan, M: Kas, U: İdrar

4.1.3. MtDNA Yeni Nesil Dizileme Sonuçları

Toplam 6 gönüllünün kan ve kas dokusundan elde edilen mtDNA'da yapılan YND çalışmasından elde edilen heteroplazmik varyantların frekans analizi sonuçları ve bu varyantların sayısal dökümü, tüm varyant tipleri (SNV + indel), SNV tipi ve Indel tipi varyantlar için ayrı ayrı verildi. Kullanılan iki yazılımdan (Ion Reporter ve IGV) elde edilen bilgilere göre frekans ortalama değerlerinin yani ortalama heteroplazmi oranlarının sonuçları, çalışılan dokuların her biri ve gönüllüde saptanmış tüm varyantlar için hesaplandı.

Yazılım algoritmalarındaki farklılıklara bağlı olarak, Ion Reporter'dan ve IGV'den ulaşılan varyant derinliklerinde farklılıklar olduğu tespit edildiğinden, nihai veriler, her bir değişim için iki platform üzerinden karşılaştırma yapılarak elde edildi ve sonuçlar her iki yazılımın değerleri verilerek sunuldu. IGV üzerinden yapılan kontrollerde, okuma bölgesinin temiz olmasına, her iki sarmaldan en az üç okumanın bulunuyor olmasına, okumaların haritalama kalitesine, değişimler eğer tek nükleotid farklılıkları ise, bazın kalitesine, indel tipi değişimler için, Ion Reporter'da bildirilen sayıda baz okumasındaki her bir bazın okuma temizliğine ve baz kalitesine dikkat edildi.

Ion Reporter ve IGV karşılaştırmalarında, IGV'nin daha derin okuma verdiği gözlemlendi. IGV'deki fazla varyant miktarı, rapor üzerinden aynı varyant için yapılan doku karşılaştırması sonucunda elde edildi. Bu durum IGV'de yer alan bir varyantın Ion Reporter çıktısında belirtilmemiş olmasının mümkün olacağını gösterdi.

4.1.3.1. Gönüllülerde Saptanan Varyantların Sayısal Dökümleri

Yapılan mtDNA YND çalışmasında, gönüllülerin kan ve kas dokusundan elde edilen toplam varyantların tiplerine göre sayısal dökümü Tablo 4-1'de, her bir gönüllüde ortak olan ve olmayan varyantların tiplerine göre sayısal dökümü Tablo 4-2'de sunulmaktadır.

Toplam varyant miktarındaki varyant çeşitliliğinin (SNV, indel), total dağılımın %50'si kadar olduğu (Tablo 4-1) ve dokular arası çeşitlilik izleminde, indel tipi değişimlerin SNV'ye göre daha fazla olduğu (Tablo 4-1), kan ve kas dokusunun yalnızca birinde bulunan varyantlarda ise SNV yoğunluğunun daha fazla olduğu belirlendi ve özellikle kas dokusundaki SNV yoğunluğu dikkat çekici bulundu

(Tablo 4-2). Bu sonuçlarla, her varyantın her dokuda var olmadığı, birden fazla doku üzerinde çalışma yapılarak daha fazla sayıda varyant bilgisine ulaşılabileceği görüldü.

Table 4-1: Kan ve kas dokusunda yapılan mtDNA YND çalışmasında elde edilen toplam varyantların tiplerine göre sayısal dökümü

GÖNÜLLÜ	VARYANT SAYISI								
	KAN + KAS (+İDRAR)			KAN			KAS		
	TVT	SNV	INDEL	TVT	SNV	INDEL	TVT	SNV	INDEL
G1	14	6	8	11	4	7	11	3	8
G2	18	9	9	12	3	9	12	3	9
G3	12	4	8	11	3	8	10	3	7
G4	11	6	5	7	4	3	10	5	5
G5	13	8	5	7	3	4	9	6	3
G6	14	8	6	8	2	6	11	7	4
TOPLAM	82	41	41	56	19	37	63	27	36
%		50	50		34	66		43	57

TVT: Tüm Varyant Tipleri (SNV + INDEL), TOPLAM: Toplam varyant sayısı

Table 4-2: Kan ve kas dokusunda yapılan mtDNA YND çalışmasında her bir gönüllüde, doku temelinde ortak olan ve olmayan varyantların tiplerine göre sayısal dökümü

GÖNÜLLÜ	VARYANT SAYISI											
	KAN + KAS (+İDRAR)			YALNIZ KAN			YALNIZ KAS			KAN-KAS ORTAK		
	TVT	SNV	INDEL	TVT	SNV	INDEL	TVT	SNV	INDEL	TVT	SNV	INDEL
G1	14	6	8	3	3	0	3	2	1	8	1	7
G2	18	9	9	3	3	0	3	3	0	9	0	9
G3	12	4	8	2	1	1	1	1	0	9	2	7
G4	11	6	5	1	1	0	4	2	2	6	3	3
G5	13	8	5	4	2	2	6	5	1	3	1	2
G6	14	8	6	3	1	2	6	6	0	5	1	4
TOPLAM	82	41	41	16	11	5	23	19	4	40	8	32
%		50	50		69	31		83	17		20	80

TVT: Tüm Varyant Tipleri (SNV + INDEL), TOPLAM: Toplam varyant sayısı

Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre sayısal dökümü Tablo 4-3'te sunulmaktadır. Gönüllülerin tümünde kan ve kas dokusunda, G2'de ayrıca idrar dokusunda saptanan mtDNA varyantlarıyla, her bir

gönüllü için toplam varyant sayısına ulaşıldı. G2’de saptanan toplam varyant sayısı 18’idi (Tablo 4-3). Ancak çalışmalara idrar dahil edilmeseydi toplam varyant sayısı 15 olacaktı. Bu gönüllüde 3 adet heteroplazmik SNV’nin yalnızca idrara dökülen hücrelerdeki mtDNA’da olduğu görüldü. G2’deki 3 farklı dokunun karşılaştırılması ise tek bu gönüllü temelinde yapılmıştır.

Table 4-3: Gönüllü-2 (G2)’nin kan, kas ve idrar örneklerinden elde edilen mtDNA YND çalışmasında saptanan toplam varyantların tiplerine göre sayısal dökümü

G2	VARYANT SAYISI							
	KAN KAS İDRAR	YALNIZ KAN	YALNIZ KAS	YALNIZ İDRAR	KAN-KAS ORTAK	KAN-İDRAR ORTAK	KAS-İDRAR ORTAK	KAN-KAS- İDRAR ORTAK
TVT	18	1	2	3	4	2	1	5
SNV	9	1	2	3	0	2	1	0
INDEL	9	0	0	0	4	0	0	5

TVT: Tüm Varyant Tipleri (SNV + INDEL)

4.1.3.2. Gönüllülerde Saptanan Tüm Varyant Tipleri İçin Sonuçlar

Kan ve kas dokusundan elde edilen mtDNA’da yapılan YND çalışmasından elde edilen heteroplazmik varyantların frekans analizi sonuçları ve bu varyantların sayısal dökümü sırasıyla Tablo 4-4 ve Tablo 4-5’te verilmektedir. Frekans ortalaması, saptanan tüm varyantların ortalama heteroplazmi oranıdır. Ion Reporter ve IGV karşılaştırmalarında, IGV’nin daha derin okuma verdiği gözlenmiştir (Tablo 4-4). IGV’deki fazla varyant miktarı, rapor üzerinden aynı varyant için yapılan doku karşılaştırması sonucunda elde edilmiştir (Tablo 4-5). Bu durum IGV’de yer alan bir varyantın Ion Reporter çıktısında belirtilmemiş olmasının mümkün olacağını gösterdi.

Doku karşılaştırmalarında, gönüllü verilere göre hesaplanan ortalama frekans değerinin (heteroplazmi oranının), kasta, Ion Reporter’a göre 2,1 kat, IGV’ye göre ise 1,4 kat daha yüksek olduğu gözlemlendi. Toplam varyant sayısının, Ion Reporter’a göre, %48’i kanda, %57’si kasta; IGV verilerine göre ise %68’i kanda, %77’si kasta saptandı. Bu sonuçlara göre, genetik testlerde her iki dokunun da tercih edilebilir olduğu görüldü. Bununla beraber, kastan bilgi elde edilebilecek varyant sayısının, toplam varyant sayısına daha yakın olduğu gözlemlendi ve elde edilen frekans bilgileri, kas varyant yükünün, kan varyant yüküne göre daha fazla olduğunu gösterdi.

Table 4-4: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında elde edilen tüm değişimlerin (SNV ve Indel) frekans ortalamaları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
TÜM VARYANT TİPLERİ	G1	14	2,6	6,6	1,7	6,7
	G2	18	0,8	3,1	2,4	3,3
	G3	12	1,8	2,8	3,8	6,7
	G4	11	1,2	3	4,6	7,5
	G5	13	0,6	2,8	2,5	4,2
	G6	14	2,1	5,4	4,4	6,9
TOPLAM		82	1,5	4	3,1	5,7
ORTAMALA		14				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-5: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında elde edilen tüm değişimlerin sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
TÜM VARYANT TİPLERİ	G1	14	8	11	6	11
	G2	18	3	12	10	12
	G3	12	10	11	4	10
	G4	11	6	7	8	10
	G5	13	5	7	9	9
	G6	14	7	8	10	11
TOPLAM		82	39	56	47	63
%			48	68	57	77

TOPLAM: Toplam varyant sayısı

Kan ve kas örneklerinden elde edilen mtDNA'da yapılan YND çalışmasında ortak oldukları saptanan varyantların frekans ortalama sonuçları Tablo 4-6'da, bu varyantların sayısal dökümü Tablo 4-7'de sunulmaktadır.

Heteroplazmi oranının hangi dokuda daha yüksek olduğu sorusuna yanıt bulmak amacıyla dokularda ortak olarak bulunan varyantlar üzerinden frekans değerlendirmesi

yapıldı. Kas dokusundan elde edilen ortalama frekans değerinin, Ion Reporter verilerine göre kanda elde edilenin 1,9 katı, IGV verilerine göre ise 1,2 katı olduğu gözlemlendi (Tablo 4-6).

Bu tez çalışmasında, Ion Reporter ile IGV arasında toplam varyant saptama sayıları farklı olduğu gözlemlendiğinden, ortak varyant sayısı bilgileri sunulurken her iki yazılım sonuçları gözetilerek toplam sayıya ulaşıldı. İki ayrı yazılımdan elde edilen farklı sayıda varyantların tüm varyantlar üzerinden yüzdesi alındığında, ortak varyantlar IGV karşılaştırması sonrasında tespit edildiği için IGV'den %100, Ion Reporter'dan her iki doku için ortalama %64 verim sağlandığı görüldü (Tablo 4-7).

Table 4-6: Kan ve kas örneklerinden edilen mtDNA'da yapılan YND çalışmasında ortak bulunan varyantların frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
TÜM VARYANT TİPLERİ	G1	8	2,4	9,4	2	10,5
	G2	9	1,4	5,8	4,3	6,2
	G3	9	2,1	3,4	5	8,8
	G4	6	1,8	5,2	2,2	5,2
	G5	3	1,3	10,3	1,7	8,7
	G6	5	3	12,8	6,6	14
TOPLAM		40	2	7,1	3,8	8,7
ORTAMALA		7				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-7: Kan ve kas örneklerinden edilen mtDNA’da yapılan YND çalışmasında ortak bulunan varyantların sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
TÜM VARYANT TİPLERİ	G1	8	5	8	4	8
	G2	9	2	9	7	9
	G3	9	8	9	3	9
	G4	6	5	6	4	6
	G5	3	2	3	3	3
	G6	5	4	5	4	5
TOPLAM		40	26	40	25	40
%			65	100	63	100

TOPLAM: Toplam varyant sayısı

Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen, buna karşın diğer dokusunda saptanmayan eş varyantların sayısal dökümü incelendiğinde, %59’unun kasta, %41’inin kanda yapılan çalışmadan ulaşılabileceği ve kas dokusundaki mevcut çeşitliliğin kandan daha yüksek olduğu gözlemlendi (Tablo 4-8).

Table 4-8: mtDNA YND çalışmasında yalnızca bir dokuda görülen, buna karşın diğer dokuda saptanmayan varyantların sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
TÜM VARYANT TİPLERİ	G1	6	3	3	2	3
	G2	6	1	3	3	3
	G3	3	2	2	1	1
	G4	5	1	1	4	4
	G5	10	3	4	6	6
	G6	9	3	3	6	6
TOPLAM		39	13	16	22	23
%			33	41	56	59

TOPLAM: Toplam varyant sayısı

4.1.3.3. Gönüllülerde Saptanan SNV Tipi Varyantlar İçin Sonuçlar

Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında dokularda saptanan tüm nokta değişimlerinde (SNV), dokular için ayrı ayrı, Ion Reporter ve IGV frekans ortalama sonuçlarının birbirine eşit olduğu görüldü (Tablo 4-9). Frekans temelinde ise kas frekans ortalama değerinin, kan frekans ortalama değerinin 2,7 katı olduğu gözlemlendi. Ion Reporter’a göre kanda SNV’lerin saptanma yüzdesi %34, kasta %61, IGV’ye göre ise kanda %46, kasta %66 olarak bulundu (Tablo 4-10). Bu sonuçlar, ulaşılan SNV varyant sayısının kas dokusunda daha yüksek olduğunu göstermiş oldu.

Table 4-9: Kan ve kas örneklerine ait mtDNA YND çalışmasıyla saptanan tüm nokta değişimlerin (SNV) frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
SNV	G1	6	2,8	3	2,5	2,5
	G2	9	0,1	0,3	0,4	0,4
	G3	4	2,3	2,3	10,8	11
	G4	6	1,3	1,3	1,7	1,8
	G5	8	0,4	0,4	1	0,9
	G6	8	0,1	0,3	3,6	3,5
TOPLAM		41	1	1	2,7	2,7
ORTAMALA		7				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-10: Kan ve kas örneklerine ait mtDNA YND çalışması sonrasında saptanan tüm nokta değişimlerin (SNV) sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
SNV	G1	6	3	4	2	3
	G2	9	1	3	3	3
	G3	4	3	3	3	3
	G4	6	4	4	4	5
	G5	8	2	3	6	6
	G6	8	1	2	7	7
TOPLAM		41	14	19	25	27
%			34	46	61	66

TOPLAM: Toplam varyant sayısı

Kan ve kas örneklerinden elde edilen mtDNA'da yapılan YND çalışmasında ortak oldukları saptanan tüm nokta değişimlerin (SNV) frekans ortalama sonuçları Tablo 4-11'de ve sayısal dökümü Tablo 4-12'de sunulmaktadır.

Kan ve kas dokusunun her ikisinde de saptanan ortak SNV varyantların frekansları hesaplandığında, Ion Reporter'a göre kasın frekans toplam değeri, kan frekans toplam değerinin 4,2 katı, IGV'ye göre 3,7 katı olduğunu gözlemlendi (Tablo 4-11). Varyant sayı bilgisi dökümlerine göre (Tablo 4-12), IGV'de %100 olan veri elde etme yüzdesinin, Ion Reporter'a göre kanda %63, kasta %88 olduğu görüldü. Bu sonuçlarla, Ion Reporter verilerinin mutlaka IGV üzerinden kontrol edilmesi gerektiği sonucuna ulaşıldı. Saptanan varyant sayısının, Ion Reporter'a göre düşük olmasına karşın, frekans değerlerinin doku içi karşılaştırmalarda birbirine yakın olduğu görüldü (Tablo 4-11). Ion Reporter'la saptanamayıp IGV ile varlığı saptanan varyantların düşük frekanslı varyantlar olduğu gözlemlendi. Buna karşın, IGV'de saptanan varyantların frekans değerleri ile Ion Reporter parametre ayarı en az %1 olarak düzenlenmişti. Ortak varyantlar üzerinden yapılan frekans değerlendirmesinde, SNV incelemelerinde kasın, kana göre dört kat daha bilgi verici olduğu görüldü.

Table 4-11: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan tüm nokta değişimlerin (SNV) frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
SNV	G1	1	0	1	9	8
	G2	0	0	0	0	0
	G3	2	4	4	21	21,5
	G4	3	2	2	2,3	2,7
	G5	1	0	1	2	2
	G6	1	0	1	1	1
TOPLAM		8	1,8	2,1	7,6	7,8
ORTAMALA		1				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-12: Kan ve kas örneklerinden edilen mtDNA’da yapılan YND çalışmasında ortak bulunan nokta değişimlerin (SNV) sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
SNV	G1	1	0	1	1	1
	G2	0	0	0	0	0
	G3	2	2	2	2	2
	G4	3	3	3	2	3
	G5	1	0	1	1	1
	G6	1	0	1	1	1
TOPLAM		8	5	8	7	8
%			63	100	88	100

TOPLAM: Toplam varyant sayısı

Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen, buna karşın diğer dokusunda saptanmayan eş nokta değişimlerin (SNV) sayısal dökümü Tablo 4-13’te sunulmaktadır.

Yalnızca bir dokuda görülen SNV varyantların sayı analizine göre, kas dokusunda saptananların, toplamın %63’ü olduğu, geri kalanın ise kana ait olduğu

gösterildiğinden, kasta SNV varyant miktarının kana göre daha fazla olduğu desteklendi (Tablo 4-13).

Table 4-13: Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen buna karşın diğer dokusunda saptanmayan eş nokta değişimlerin (SNV) sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
SNV	G1	5	3	3	1	2
	G2	6	1	3	3	3
	G3	2	1	1	1	1
	G4	3	1	1	2	2
	G5	7	2	2	5	5
	G6	7	1	1	6	6
TOPLAM		30	9	11	18	19
%			30	37	60	63

TOPLAM: Toplam varyant sayısı

4.1.3.4. Gönüllülerde Saptanan Indel Tipi Varyantlar İçin Sonuçlar

Kan ve kas örneklerinden elde edilen mtDNA'da yapılan YND çalışmasında ortak oldukları saptanan tüm indel tipi değişimlerin frekans ortalama sonuçları Tablo 4-14'te ve sayısal dökümü Tablo 4-15'te sunulmaktadır.

Tüm indel değişimlerin frekans ortalama değerleri incelendiğinde, Ion Reporter'a göre kastan elde edilen varyant yükü bilgisi kandan elde edilen oranın 1,8 katı, IGV'ye göre ise 1,3 katıdır (Tablo 4-14). Frekans değerleri bu varyant türü için de kasın kana göre daha yüksek oranda indel içerdiğini göstermektedir. Ortak olan ve olmayan tüm indel değişimleri incelendiğinde, Ion Reporter ile indel varyantlarının kanda %61'i, kasta %54'ü hakkında bilgi elde edilebildiği, IGV ile kanda %90'ından, kasta %88'inden bilgi elde edilebildiği gözlenmektedir. Bu sonuçlar, indel değişimlerinin her iki dokuda da birbirine yakın değerlerde saptanabileceğini ancak, kanın kasa göre az bir farkla indel tipi varyantı bulma konusunda daha bilgi verici olduğunu göstermektedir.

Table 4-14: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında saptanan tüm indel tipi değişimlerin frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
INDEL	G1	8	2,4	9,3	1,1	9,9
	G2	9	1,4	5,8	4,3	6,2
	G3	8	1,5	3	0,4	4,5
	G4	5	1	5	8,2	14,4
	G5	5	1	6,8	5	9,4
	G6	6	4,7	12,3	5,3	11,5
TOPLAM		41	2	6,9	3,6	8,8
ORTAMALA		7				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-15: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında saptanan tüm indel tipi değişimlerin sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
INDEL	G1	8	5	7	4	8
	G2	9	2	9	7	9
	G3	8	7	8	1	7
	G4	5	2	3	4	5
	G5	5	3	4	3	3
	G6	6	6	6	3	4
TOPLAM		41	25	37	22	36
%			61	90	54	88

TOPLAM: Toplam varyant sayısı

Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan tüm indel tipi değişimlerin frekans ortalama değerleri Tablo 4-16’da ve sayısal dökümü Tablo 4-17’de sunulmaktadır.

Indel frekans değerlendirmesinin ortak varyantlar üzerinden sonuçları incelendiğinde, varyantların yükünün, Ion Reporter'a göre kasta, kanın 1,3 katı, IGV'ye göre 1,1 katı olduğu gözlenmektedir (Tablo 4-16). Değerler birbirine yakın olsa da, kastaki heteroplazmi oranının daha fazla olduğu görülmektedir. Toplam ortak indel varyant sayısının, Ion Reporter'a göre, %66'sı kanda, %56'sı kasta saptandı. Ortak varyantlar IGV kontrollerine göre alındığından verim oranının iki doku için de %100 olduğunu görülmektedir (Tablo 4-17).

Table 4-16: Kan ve kas örneklerinden elde edilen mtDNA'da yapılan YND çalışmasında ortak oldukları saptanan indel tipi değişimlerin frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
INDEL	G1	7	2,7	10,6	1	10,9
	G2	9	1,4	5,8	4,3	6,2
	G3	7	1,6	3,3	0,4	5,1
	G4	3	1,7	8,3	2	7,7
	G5	2	2	15	1,5	12
	G6	4	3,8	15,8	8	17,3
TOPLAM		32	2,1	8,3	2,8	8,9
ORTAMALA		5				

TOPLAM: Toplam varyant sayısı, ORTALAMA: Ortalama varyant sayısı

Table 4-17: Kan ve kas örneklerinden elde edilen mtDNA’da yapılan YND çalışmasında ortak oldukları saptanan indel tipi değişimlerin sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
INDEL	G1	7	5	7	3	7
	G2	9	2	9	7	9
	G3	7	6	7	1	7
	G4	3	2	3	2	3
	G5	2	2	2	2	2
	G6	4	4	4	3	4
TOPLAM		32	21	32	18	32
%			66	100	56	100

TOPLAM: Toplam varyant sayısı

Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen, buna karşın diğer dokusunda saptanmayan eş indel tipi değişimlerin sayısal dökümü Tablo 4-18’de sunulmaktadır.

Yalnızca bir dokuda görülen indel tipi varyant miktarları incelendiğinde, bu varyantların %56’sının kanda, %44’ünün kasta saptanması, indel açısından kanın daha zengin olduğunu göstermiştir (Tablo 4-18).

Table 4-18: Kan ve kas dokularında yapılan mtDNA YND çalışmasında her bir gönüllünün sadece bir dokusunda gözlenen buna karşın diğer dokusunda saptanmayan eş indel tipi değişimlerin sayısal dökümü

DEĞİŞİM TÜRÜ	GÖNÜLLÜ	VARYANT SAYISI	VARYANT SAYISI			
			KAN		KAS	
			ION REPORTER	IGV	ION REPORTER	IGV
INDEL	G1	1	0	0	1	1
	G2	0	0	0	0	0
	G3	1	1	1	0	0
	G4	2	0	0	2	2
	G5	3	1	2	1	1
	G6	2	2	2	0	0
TOPLAM		9	4	5	4	4
%			44	56	44	44

TOPLAM: Toplam varyant sayısı

4.1.3.5. Üç Doku Çalışması için Varyant Sonuçları

Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre, frekans bilgileri Tablo 4-19'da ve bu varyantların sayısal dökümü ve oranları Tablo 4-20'de sunulmaktadır.

Kan, kas ve idrar olmak üzere üç dokudan yapılan mtDNA YND işlemleri sonucunda, G2'de toplam 18 adet varyant saptanmıştır. Ion Reporter'a göre sayıların kan ve idrarda birbirine yakın olduğu görülmektedir. Bu iki dokunun kas dokusu ile kıyaslanması sonucunda, kas dokusunda saptanan varyantların frekans toplam değerleri, kandan 3 kat, idrardan 3,4 kat yüksek bulunmuştur. IGV sonuçlarına göre ise kanın idrardan 1,6 kat, kasın kandan 1,1 kat, kasın idrardan 1,7 kat yüksek olduğu görülmüştür (Tablo 4-19). Kan ve kas frekans değerleri karşılaştırıldığında ise, kasta az bir farkla yüksek olmakla beraber, birbirine yakın olduğu gözlenmektedir. Genel olarak, idrarın diğer iki dokudan daha az frekans değeri verdiği görülmüştür.

G2'deki üç doku çalışmasında saptanan varyant tipleri incelendiğinde, varyantların yarısının SNV, diğer yarısının indel tipi varyantlar olduğu gözlenmiştir. SNV'lerin toplam frekans doku oranlamalarında, Ion Reporter'a göre, idrarın, kandan 8 kat, kastan 2 kat yüksek olduğu, kas frekans değerinin ise kandan 4 kat yüksek olduğunu göstermiştir. IGV'ye göre, idrar frekans değeri, kandan 3, kastan 2,2 kat

yüksekken, kas frekans değeri kandan 1,3 kat yüksektir. Indel tipi değişimlerin frekans değerleri karşılaştırıldığında, Ion Reporter’a göre kas frekans toplam değeri, kanın 3,1 katı, idrarın 7,2 katı; kan frekans toplam değeri idrarın 2,3 katı, IGV’ye göre ise kas frekans toplam değeri kanın 1,1 katı iken, idrarın 2,1 katı; kan frekans toplam değeri idrarın 2 katı olduğu görülmüştür (Tablo 4-19). Frekans hesaplamasına göre kan ve kasın yakın sonuçlar verdiği görülmüştür. Ancak doğru değerlendirme için ortak varyantlar üzerinden gitmenin daha doğru olacağı düşünüldü.

SNV ve indel tipi varyantlar karşılaştırıldığında indel tipi varyantların, varyant yükünün SNV tipi varyantlara göre yüksek olduğu saptanmıştır. SNV tipi varyantlarda, idrar dokusu frekans değerlendirmesi bakımından en yüksek bulunurken, indel tipi değişimlerde kas öne çıkmıştır (Tablo 4-19).

Table 4-19: Gönüllü 2 (G2)’nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)					
		KAN		KAS		IDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	18	0,8	3,1	2,4	3,3	0,7	1,9
SNV	9	0,1	0,3	0,4	0,4	0,8	0,9
INDEL	9	1,4	5,8	4,3	6,2	0,6	2,9

TOTAL: SNV + INDEL

Ion Reporter çıktısına göre kanda, kasta ve idrarda varyant saptama oranları sırasıyla %17, %56 ve %39’dur. IGV doku karşılaştırma çalışması sonrasında elde edilen bilgilere göre, oranlar kan ve kasta %67 iken, idrarda %61’dir (Tablo 4-20). Ion Reporter, kasın diğer dokular üzerinde varyant saptama açısından üstün olduğunu gösterse de, IGV incelemesi, üç dokunun da tüm varyantlar açısından benzer oranda verimli olduğunu göstermektedir.

SNV tipi varyantların Ion Reporter’a göre kan, kas ve idrarda bulunma yüzdeleri sırasıyla %11, %33, %56, IGV’ye göre, %33, %33, %67’dir (Tablo 4-20). SNV’lerin idrarda saptanma oranının diğer doku tiplerine göre daha yüksek olduğu görülmektedir. Indel tipi varyantların dokularda saptanma yüzdesi kan, kas ve idrar için sırasıyla,

Ion Reporter'a göre %22, %78, %22'dir. IGV'ye göre ise kan ve kasta saptanma yüzdesi %100 iken, idrarda saptanma yüzdesi %56'dır (Tablo 4-20). Indel tipi varyantlar için kan ve kas üzerinden çalışmak, bu sonuçlara göre idrarda çalışmaktan avantajlı görüldü.

Table 4-20: Gönüllü 2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen varyantların tiplerine göre sayısal dökümü ve oranları

DEĞİŞİM TÜRÜ ve YÜZDE	VARYANT SAYISI	VARYANT SAYISI					
		KAN		KAS		IDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	18	3	12	10	12	7	11
%		17	67	56	67	39	61
SNV	9	1	3	3	3	5	6
%		11	33	33	33	56	67
INDEL	9	2	9	7	9	2	5
%		22	100	78	100	22	56

TOTAL: SNV + INDEL

Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri Tablo 4-21'de ve bu varyantların sayısal dökümü ve oranları Tablo 4-22'de sunulmaktadır.

Frekans değerlendirmesinin doğru yapılabilmesi için dokuların hepsinde bulunan varyantlar incelenmiştir. G2'nin toplam beş adet varyantı tüm dokularda tespit edilmiştir ve tamamının indel tipli olduğu görülmüştür. Frekans toplamı oranları incelendiğinde, Ion Reporter'a göre kanın kasa oranı 1,1, idrara oranı 2,6; kasın idrara oranı 2,4'tür. IGV'ye göre ise, veriler sırayla, 1,1, 1,3 ve 1,1'dir (Tablo 4-21). IGV'ye göre dokuların frekans değerleri birbirine yakındır ancak kanın diğer dokulardan az bir farkla yüksek olduğu gözlenmiştir. Ion Reporter'a göre ortak varyantların kan ve idrarda saptanma oranı %40 iken, kasta bulunma oranı ise %60'tır. IGV ile ise bu varyantların tümünün dokularda mevcut olduğu gözlenmiştir (Tablo 4-22).

Table 4-21: Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri

DEĞİŞİM TÜRÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)					
		KAN		KAS		IDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	5	2,6	6,8	2,4	6	1	5,2
SNV	0	0	0	0	0	0	0
INDEL	5	2,6	6,8	2,4	6	1	5,2

TOTAL: SNV + INDEL

Ion Reporter raporunda bulunan varyantların doku karşılaştırmalarının IGV ile yapılması sonrası elde edilen varyant sayıları dökümü, doku bazında oluşturuldu (Tablo 4-22).

Table 4-22: Gönüllü-2 (G2)'nin kan, kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre ortak varyantların sayısal dökümü ve oranları

DEĞİŞİM TÜRÜ ve YÜZDE	VARYANT SAYISI	VARYANT SAYISI					
		KAN		KAS		IDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	5	2	5	3	5	2	5
%		40	100	60	100	40	100
SNV	0	0	0	0	0	0	0
%		0	0	0	0	0	0
INDEL	5	2	5	3	5	2	5
%		40	100	60	100	40	100

TOTAL: SNV + INDEL

Gönüllü-2 (G2)'nin kan ve kas örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri Tablo 4-23'te sunulmaktadır.

G2'nin kan ve kas dokularında ortak olarak bulunan varyantların tamamının indel tipi olduğu, IGV'den edinilen bilgilere göre, frekans toplam sonuçlarında, kas frekans toplamının kandan 1,4 kat yüksek olduğu görülmüştür (Tablo 4-23).

Table 4-23: Gönüllü-2'nin kan ve kas dokularında saptanmış ortak varyantların frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
		KAN		KAS	
		ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	4	0	4,5	6,8	6,5
SNV	0	0	0	0	0
INDEL	4	0	4,5	6,8	6,5

TOTAL: SNV + INDEL

Gönüllü-2 (G2)'nin kan ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri Tablo 4-24'te sunulmaktadır.

Table 4-24: Gönüllü-2'nin kan ve idrar dokularında saptanmış ortak varyantların frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
		KAN		İDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	2	0	1	1	1
SNV	2	0	1	1	1
INDEL	0	0	0	0	0

TOTAL: SNV + INDEL

Gönüllü-2 (G2)'nin kas ve idrar örneklerinden elde edilen ortak varyantların tiplerine göre frekans bilgileri Tablo 4-25'te sunulmaktadır.

G2'deki üç doku çalışmasında ortak SNV saptanamadığından frekans değerlendirmesi yapılamamıştır ancak ikili eşleşmelerde gözlenmiştir. Kan ve idrarda ortak olduğu tespit edilen iki farklı SNV'nin IGV'deki değerlendirmesinde, iki varyantın da aynı oranda frekans verdiği görülmüştür (Tablo 4-24). Kas ve idrarda ortak bulunan SNV sayısı bir olup, IGV'ye göre iki dokudan da aynı oranda frekans değeri elde edilmiştir (Tablo 4-25).

Table 4-25: Gönüllü-2'nin kas ve idrar dokularında saptanmış ortak varyantların frekans ortalama sonuçları

DEĞİŞİM TÜRÜ	VARYANT SAYISI	ORTALAMA HETEROPLAZMİ ORANI (%)			
		KAS		İDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	1	1	1	0	1
SNV	1	1	1	0	1
INDEL	0	0	0	0	0

TOTAL: SNV + INDEL

Gönüllü 2 (G2)'nin kan, kas ve idrar dokularında yapılan mtDNA YND çalışmasında sadece bir dokusunda gözlenen buna karşın diğer dokusunda saptanmayan eş varyantların sayısal dökümü ve oranları Tablo 4-26'da sunulmaktadır.

G2'de, dokuların yalnızca birinde gösterilen varyantların tamamı SNV tiplidir. Bu varyantların %50'si idrarda, %33'ü kasta, %17'si kanda bulunmuştur (Tablo 4-26).

Table 4-26: Gönüllü 2 (G2)'nin yalnızca bir dokusunda görülen buna karşın diğer dokularda saptanmayan varyantların tiplerine göre sayısal dökümü ve oranları

DEĞİŞİM TÜRÜ ve YÜZDE	VARYANT SAYISI	VARYANT SAYISI					
		KAN		KAS		İDRAR	
		ION REPORTER	IGV	ION REPORTER	IGV	ION REPORTER	IGV
TOTAL	6	1	1	2	2	3	3
%		17	17	33	33	50	50
SNV	6	1	1	2	2	3	3
%		17	17	33	33	50	50
INDEL	0	0	0	0	0	0	0
%		0	0	0	0	0	0

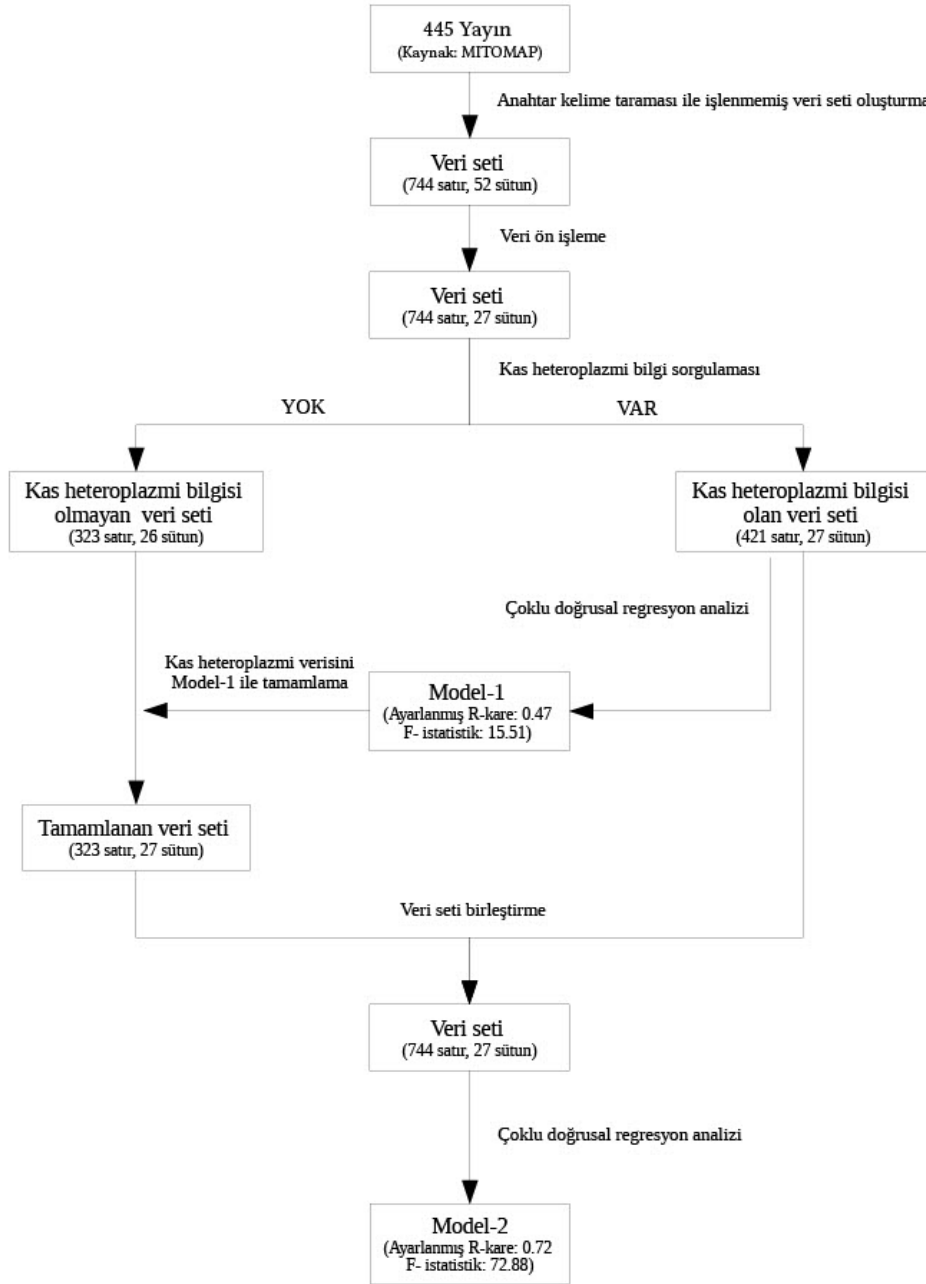
TOTAL: SNV + INDEL

4.2. Veri Madenciliği ve Makine Öğrenmesi Çalışmalarının Sonuçları

MITOMAP veri tabanı aracılığıyla m.3243A>G değişiminin geçtiği akademik yayınların araştırılması sonucunda 445 yayına ulaşıldı. Yayınlardan elde edilen bilgiler (hastaların doku heteroplazmi oranları, klinik özellikleri, aile öyküsü, demografik özellikler, biyokimyasal test sonuçları, beyin görüntüleme sonuçları ve kas histokimya sonuçları) veri setine sütunlar halinde eklendi. Bu incelemeler sonucunda sütun sayısı 52, hasta sayısı 744 olarak kaydedildi.

Veri madenciliği ve makine öğrenmesi çalışmalarının aşamaları Şekil 4-2’de verildiği gibidir.





Şekil 4-2: Veri madenciliği ve makine öğrenmesi çalışmalarında izlenen adımlar

4.2.1. Veri Yükleme ve Ön-İşleme Uygulamaları

1. Veri seti, *Pandas* kütüphanesi *read_csv* fonksiyonu kullanılarak yüklendi ve *df* isimli bir değişkene atandı.
2. Veri seti boyutu, *shape* fonksiyonuyla kontrol edildi.
 - *df* veri seti boyutu: 744, 52 (sıra sayısı, sütun sayısı) (hasta sayısı, öz nitelik sayısı)
3. Sütun isimleri *Pandas* kütüphanesi *rename* metodu ile düzenlendi ve *columns.values* metodu ile yazdırıldı (Şekil 4-3).

```
array(['hasta_no', 'cinsiyet', 'yaş', 'yaşam_durumu', 'miyopati',
      'dağınık_kırmızı_lifler', 'ensefalopati', 'epilepsi',
      'laktik_asidoz', 'inme_benzeri_epizodlar',
      'sensörinöral_işitme_kaybı', 'diyabet',
      'bozulmuş_glukoz_toleransı', 'endokrin_sistem_hastalıkları',
      'görme_bozukluğu', 'maküler_dejenerasyon', 'cpeo', 'pitozis',
      'kardiyomiyopati', 'kalp_ritim_bozukluğu', 'kronik_kalp_hastalığı',
      'iskemik_kalp_ve_koroner_arter_hastalıkları',
      'konjenital_kalp_hastalığı', 'nefropati', 'proteinüri',
      'son_evre_böbrek_hastalığı', 'nöropati',
      'fokal_kortikal_defektler', 'jeneralize_kortikal_defektler',
      'serebellar_defektler', 'hareket_bozuklukları',
      'bazal_ganglion_kalsifikasyonu', 'kognitif_hastalıklar',
      'mental_hastalıklar', 'baş_ağrısı', 'kusma',
      'gastrointestinal_disfonksiyonlar', 'solunum_yetmezliği',
      'gelişme_geriliği', 'hipertansiyon', 'iskelet_kası', 'kan',
      'beyin_kan_akımı', 'kalp', 'beyin', 'karaciğer', 'pankreas',
      'fibroblast', 'idrar', 'böbrek', 'bukkal_mukoza',
      'kıl_folikülleri'], dtype=object)
```

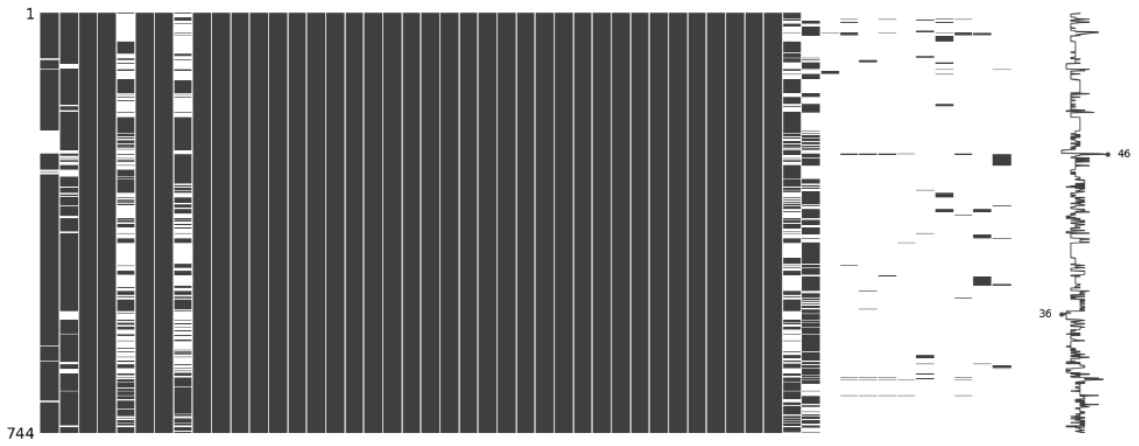
Şekil 4-3: Ham veri setindeki (*df*) sütunlar (öz nitelik) (52 adet)

4. Sayısallaştırma işlemleri yapıldı.
 - *NumPy* kütüphanesi *replace* fonksiyonu kullanılarak:
 - Klinik özellikler sütunlarındaki NaN değerleri sıfırla (0) değiştirildi.
 - *yaşam_durumu* sütunundaki ölü: 0, *hayatta*: 1 değerleriyle değiştirildi.
 - *cinsiyet* sütununda erkek: 0, *dişi*: 1 değerleriyle değiştirildi.

5. Veri, *hasta_no* sütununa göre, *Pandas set_index* fonksiyonuyla dizinlendi (*index*).
 - *df* veri seti boyutu 744, 51 (satır, sütun) olarak kaydedildi (*hasta_no* sütunu indeks olarak kullanıldığı için sütun sayısı 51 (52 → 51) oldu.).
6. Tüm veri türleri ondalık sayı olacak şekilde, *astype(float)* fonksiyonuyla tek tipe ayarlandı.
7. Kayıp değerlerin giderilmesi için bir takım işlemler uygulandı.

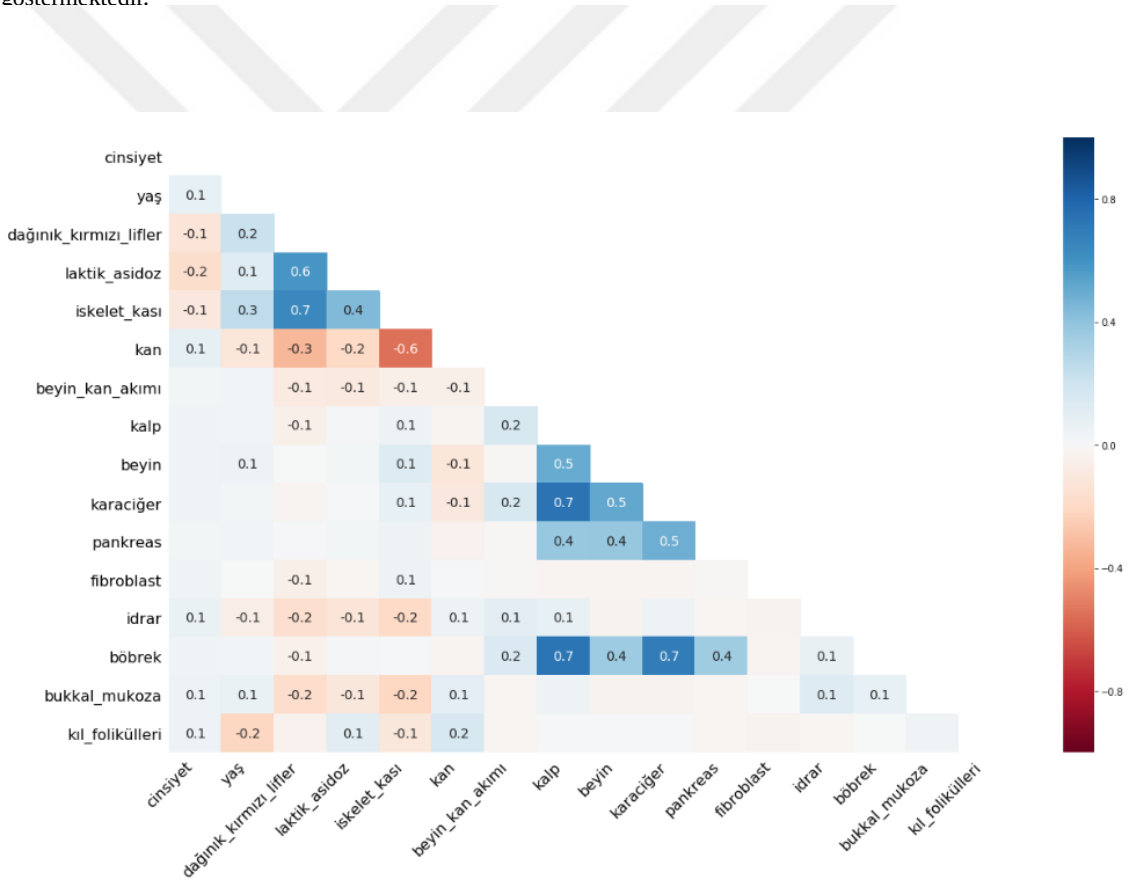
744 hastanın tamamı için bilgilerin tümü mevcut değildi. Bu durum, oluşturulan veri setinde kayıp verilerin varlığına neden oldu. Bu denli klinik çeşitliliğin, ne kadar çok sayıda hasta bilgisi üzerinde çalışılırsa ancak açıklanabileceği düşünülerek veri kaybetmek istenmedi. Olanaklı olduğu durumlarda, bulgular gözetilerek, sütunlardaki veri sayısını arttırmak amacıyla klinik gruplamalar yapıldı. Eksik verilerin olduğu satırları çıkarmak yerine, boşlukların mantıklı şekilde doldurulabilmesi için istatistiksel çalışmalar yürütüldü. Klinik ve beyin görüntüleme sonuçları yayınlarda şayet belirtilmediyse veya eksik belirtildiyse, eksik veri miktarının bilinenden daha da yüksek olabileceği göz önünde bulunduruldu.

- *Missingno* görüntüleme kütüphanesinin *matrix* fonksiyonu kayıp değerlerin genel görünümü için, *heatmap* fonksiyonu ise kayıp değerler arasındaki ilişkinin belirlenmesi için kullanıldı. *Matrix* fonksiyonu sonuç görüntüsü Şekil 4-4'te, *heatmap* (ısı haritası) fonksiyonu sonuç görüntüsü Şekil 4-5'te gösterilmektedir. Isı haritası yorumlamasına örnek olarak:
 - veri setinde *kan x iskelet_kası* verilerinin birlikte veri bulundurma değeri -0,6'dır.
 - *kan* sütununda veri varken *iskelet_kası* sütununda veri bulunmaması durumunun oranı %60'tır.



Şekil 4-4: Ham veri setine (df) uygulanan *Missingno* görüntüleme kütüphanesi *matrix* fonksiyonu görüntüleme sonucu

Bu şekilde veri setinin görüntüsü sunulmaktadır. Beyaz renkler kayıp verileri gösterirken, siyah renkler veri varlığını göstermektedir.



Şekil 4-5: Ham veri setine (df) uygulanan *Missingno* görüntüleme kütüphanesi *heatmap* fonksiyonu görüntüleme sonucu

Bu şekildeki görüntü, yalnızca kayıp veri bulunan sütunları göstermektedir. Sağ tarafta görülen dikey çubukta, lacivert renk, aynı anda veri bulundurma oranının %100 olduğunu belirtmektedir. Boşluk korelasyonu -1 ve 1 arasında değerler alır. -1 negatif korelasyondur ve *bir değişkende veri varsa diğer değişkende veri kesin yoktur* şeklinde yorumlanır. 0 korelasyonun olmadığını gösterir ve *bir değişkende verinin olup olmaması diğer değişkeni etkilemez* şeklinde yorumlanır. 1 pozitif korelasyonu belirtir ve *bir değişkende veri varsa diğer değişkende de kesin veri vardır* şeklinde yorumlanır.

- Her bir sütunun içerdiği NaN değerlerinin toplamı ve yüzdeleri *isnull.sum* ve *len* fonksiyonlarıyla hesaplandı. Sütunlardaki kayıp veri miktarı ve yüzdeleri Tablo 4-27’de sunulmaktadır.

Yayınlarda heteroplazmi çalışmaları hangi dokular üzerinden yapıldıysa, patojenik varyant yükleri veri setine, her bir doku tipi için ayrı sütunlar oluşturularak eklendi. Kan ve kas dışındaki doku tiplerine ait olan sütunlarda, %90’dan fazla kayıp veri olduğu gözlemlendi (Tablo 4-27). Bu sütunlar ilerideki aşamalardan birinde kullanılmak üzere şimdilik kaldırılmadı.

Table 4-27: *df* veri seti sütunlarındaki kayıp veri miktarları ve yüzdeleri

SÜTUN	KAYIP VERİ MİKTARI	KAYIP VERİ YÜZDESİ
cinsiyet	54	7,258065
yaş	92	12,365591
kan	314	42,204301
iskelet_kası	323	43,413978
laktik_asidoz	373	50,134409
dağınık_kırmızı_lifler	406	54,569892
idrar	707	95,026882
bukkal_mukoza	710	95,430108
kıl_folikülleri	711	95,564516
böbrek	724	97,311828
kalp	726	97,580645
fibroblast	726	97,580645
beyin	727	97,715054
karaciğer	727	97,715054
beyin_kan_akımı	737	99,05914
pankreas	738	99,193548

- MELAS hastalığıyla doğrudan ilişkili olmayan klinik özelliklerin (bağımsız değişkenler), öznitelik varken (1) ve yokken (0), iskelet kası heteroplazmi oranına etkileri, *describe* fonksiyonuyla istatistiksel döküm oluşturularak incelendi.

Yayınlarda geçen klinik bulgu ve belirtiler, laboratuvar ve beyin görüntüleme sonuçlarının her biri için oluşturulmuş sütun sayısı çalışmanın başında 51 idi ve bu ana başlık altında yer alan 3. maddede belirtilen başlıklardan oluşmakta idi. Çalışmanın başında belli bir özellik için bilgi elde edilebilecek olgu sayısı bilinmediğinden, hastalarda görülen özelliklerin her biri için sütun oluşumu yapılmış idi. Veri seti tamamlandıktan sonra, toplam olgu sayısının yeterli olmadığı sütunlar veri setinden çıkartıldı (Tablo 4-28). Bazı yakın ilişkili sütunlar birleştirildi (Tablo 4-29, Tablo 4-30). m.3243A>G varyantı ile ilişkilendirilmiş özellikler için oluşturulan sütunlarda sütun birleştirmeye gidilmedi.

Çalışmanın başında *diyabet*, *bozulmuş_glikoz_toleransı* ve *endokrin_sistemi_hastalıkları* sütunları, *endokrin_sistemi_hastalıkları* sütununda birleştirilmek yerine, üç ayrı sütun olarak korundu. Klinik olarak, bozulmuş glikoz toleransına sahip hastaların yalnızca bir kısmında, ilerleyen yaşlarda diyabetin geliştiği bildirilmiştir [150][151]. *diyabet* ve *bozulmuş_glikoz_toleransı* sütunlarının ayrı tutulması ile, diyabet öncesinde görülebilen bu intoleransa sahip hastaların ve diyabet hastalarının heteroplazmi yüklerinin ortaya çıkarılabilecek ilişkisi ile, intoleransın diyabete dönüşmesi durumunun tahmininin yapıp yapılamayacağı araştırılmak istendi. Ancak, bir çok başka klinik bulguları olan, heteroplazmi oranı yüksek raporlanan, glikoz intoleransı görülen ve diyabeti olmayan hastaların varlığı ve glikoz intoleransına sahip olgu sayısının çok düşük olması, bu yaklaşımın doğru tahminlere götürmeyeceğini düşündürdü. Benzer gerekçelerle, bozulmuş glikoz toleransının, heteroplazmi yükü ve diyabet ile ilişkisinin de irdelenmesi olası görülmedi ve *bozulmuş_glikoz_toleransı* sütunu veri setinden çıkartıldı. Diyabet, m.3243A>G varyantı saptanan hastalarda sık görülen bir klinikdir [152] ve olgu sayısı yüksektir. Bu sebeple *diyabet* sütunu korundu. Endokrin sistem hastalıklarının ise bu varyantla direk bir ilişkisi bildirilmemiştir [152] ve veri setindeki olgu sayısı da düşük bulunduğu için *endokrin_sistemi_hastalıkları* sütunu da veri setinden çıkartıldı.

- Veri sayısının yetersiz olması nedeniyle bağımlı değişkeni açıklamakta yeterli görülmeyen bağımsız değişkenler veri setinden *Pandas* kütüphanesi *drop* fonksiyonu ile çıkartıldı (Tablo 4-28). Veri sayısı düşük ancak bağımlı değişkenle ilişkili olduğu bilinen sütunlar bırakıldı.

Table 4-28: *df* veri setinden veri sayısı yetersizliği sebebiyle çıkartılan sütunlar

SÜTUN	YOK	VAR
bozulmuş_glukoz_toleransı	726	18
endokrin_sistem_hastalıkları	712	32
kronik_kalp_hastalığı	730	14
iskemik_kalp_ve_koroner_arter_hastalıkları	741	3
konjenital_kalp_hastalığı	740	4

- Veri sayısı az olan böbrek rahatsızlıkları ile ilgili sütunlar, ortak bir sütunda birleştirildi. *son_evre_böbrek_hastalığı* sütununda bulunan 1 (var) değerlerinin tamamı *nefropati* sütununda zaten 1 idi. *proteinüri* sütunundaki 1 değerlerinin bir kısmı *nefropati* sütununda 1 idi. *proteinüri*'de 1 ancak *nefropati*'de 0 olan beş değer *nefropati* sütununda 1 olarak değiştirildi. Yetersiz veri sayısı nedeniyle *nefropati* sütununda birleştirilen sütunlar Tablo 4-29'da, *nefropati* sütununun son hali ise Tablo 4-30'da sunulmaktadır.

Table 4-29: *df* veri setinde yetersiz veri sayısı sebebiyle birleştirilen sütunlar

SÜTUN	YOK	VAR
nefropati	704	40
proteinüri	723	21
son_evre_böbrek_hastalığı	732	12

Table 4-30: Tablo 4-29'da belirtilen birleştirme işleminin sonucu

SÜTUN	YOK	VAR
nefropati	699	45

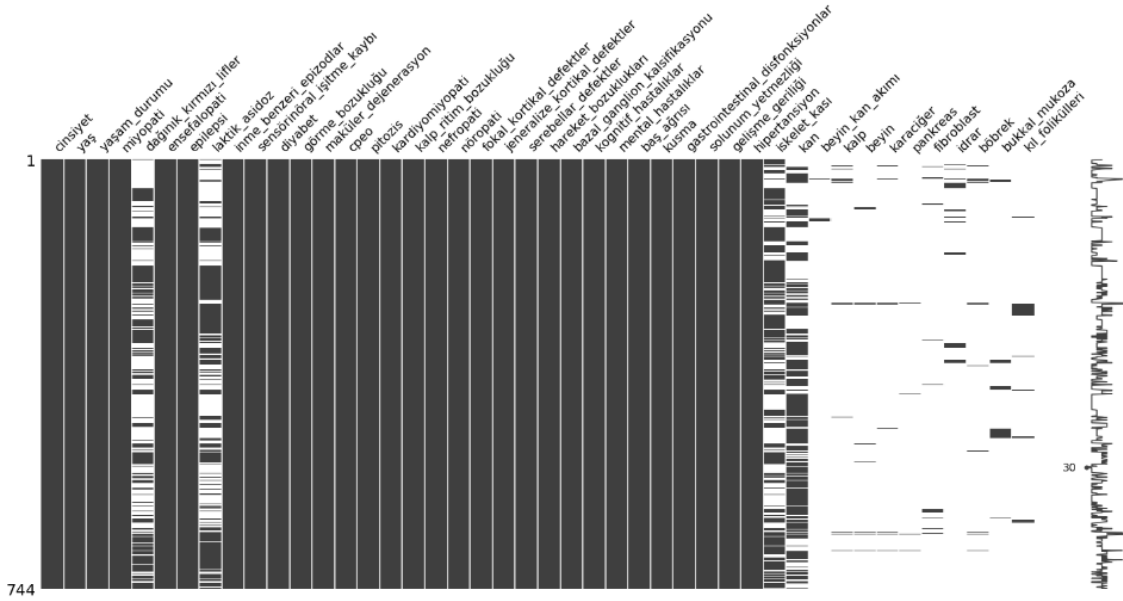
- *df* veri seti boyutu: 744, 44 (satır, sütun) (sütun sayısı: 51 → 44)
- *df* veri setindeki sütunların isimleri Şekil 4-6’da verilmiştir.

```
array(['cinsiyet', 'yaş', 'yaşam_durumu', 'miyopati',
      'dağınık_kırmızı_lifler', 'ensefalopati', 'epilepsi',
      'laktik_asidoz', 'inme_benzeri_epizodlar',
      'sensörinöral_işitme_kaybı', 'diyabet', 'görme_bozukluğu',
      'maküler_dejenerasyon', 'cpeo', 'pitozis', 'kardiyomiyopati',
      'kalp_ritim_bozukluğu', 'nefropati', 'nöropati',
      'fokal_kortikal_defektler', 'jeneralize_kortikal_defektler',
      'serebellar_defektler', 'hareket_bozuklukları',
      'bazal_ganglion_kalsifikasyonu', 'kognitif_hastalıklar',
      'mental_hastalıklar', 'baş_ ağrısı', 'kusma',
      'gastrointestinal_disfonksiyonlar', 'solunum_yetmezliği',
      'gelişme_geriliği', 'hipertansiyon', 'iskelet_kası', 'kan',
      'beyin_kan_akımı', 'kalp', 'beyin', 'karaciğer', 'pankreas',
      'fibroblast', 'idrar', 'böbrek', 'bukkal_mukoza',
      'kıl_folikülleri'], dtype=object)
```

Şekil 4-6: *df* veri setindeki sütunlar (öznitelik) (44 adet)

- Sırasıyla %7,3 ve %12,4 oranında kayıp veri içeren *cinsiyet* ve *yaş* sütunlarındaki eksik değerler, *describe* fonksiyonuyla hesaplanan ortalama değerini (*mean*) bozmayacak şekilde dolduruldu. *yaş* sütunundaki doldurma işleminde hastanın aile ağacındaki konumu, yaşa bağlı heteroplazmi oranları verisinin sağlıklı şekilde değerlendirilebilmesi için göz önünde bulunduruldu. Anne ve birinci çocuk arasındaki yaş farkı ~20 alındı.

Tüm bu işlemleri takiben, kayıp veri görüntü sonucu Şekil 4-7’de, kayıp verilerin ilişki durumu ise ısı haritası şeklinde Şekil 4-8’de gösterilmektedir.



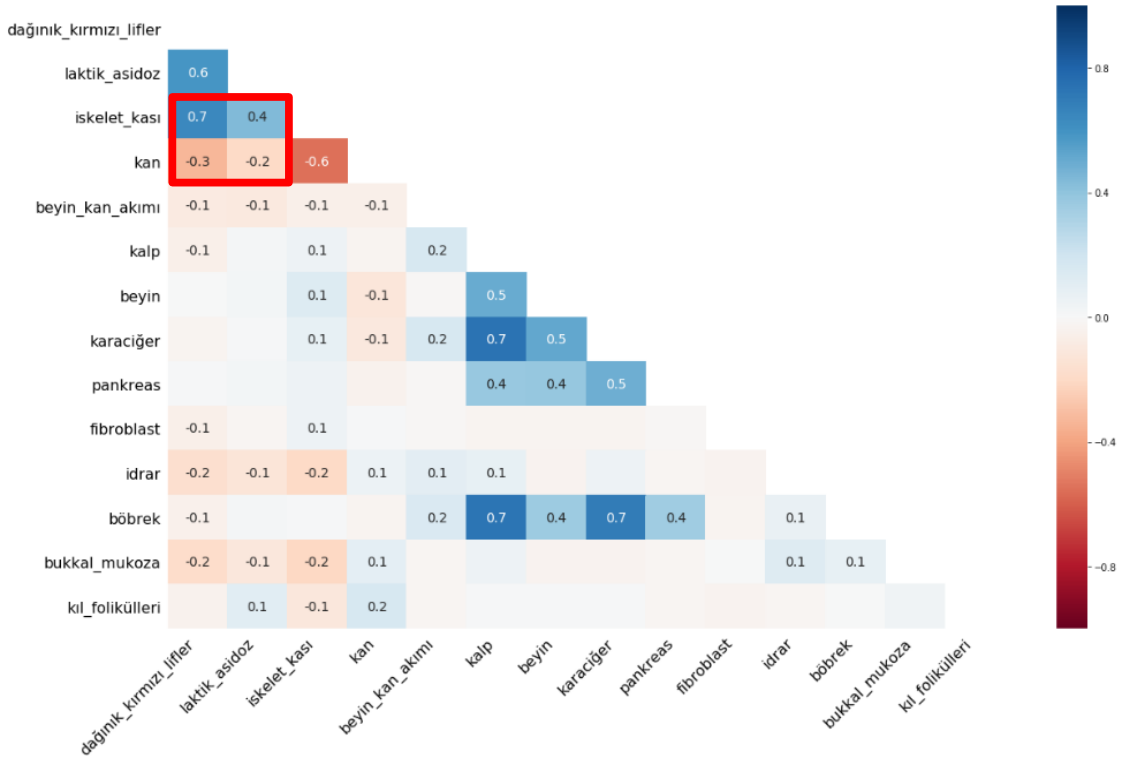
Şekil 4-7: df veri setine uygulanan Missingno kütüphanesi matrix fonksiyonu kayıp değer görüntüleme sonucu

Bu şekilde veri setinin görüntüsü sunulmaktadır. Beyaz renkler kayıp verileri gösterirken, siyah renkler veri varlığını göstermektedir.

- NaN değeri (eksik veri / boşluk) içeren *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunları için *heatmap* fonksiyonunun sonuçları değerlendirildi ve aynı anda veri bulundurma durumu *iskelet_kası* sütununda (sırasıyla 0,7, 0,4) *kan* sütununa (sırasıyla -0,3, -0,2) oranla daha fazla olduğu için boşluk doldurma işleminin, *df* veri setini bölme işleminden sonra yapılmasına karar verildi (Şekil 4-8). *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarında bulunan kayıp veri sayıları Tablo 4-31’de sunulmaktadır.

Table 4-31: df veri setinde laktik_asidoz ve dağınık_kırmızı_lifler sütunlarında bulunan kayıp veri miktarları

SÜTUN	KAYIP VERİ MİKTARI
laktik_asidoz	373
dağınık_kırmızı_lifler	406



Şekil 4-8: *df* veri setine uygulanan *Missingno* kütüphanesi *heatmap* fonksiyonuyla kayıp değerlerin aynı anda veri bulundurma durumunun görüntüleme sonucu. Kırmızı dikdörtgen içine alınmış değerler *iskelet_kası* ve *kan* sütunlarının, *dađınık_kırmızı_lifler* ve *laktik_asidoz* sütunları ile aynı anda veri bulundurma değerlerini vermektedir.

Bu şekildeki görüntü, yalnızca kayıp veri bulunan sütunları göstermektedir. Sağ tarafta görülen dikey çubukta, lacivert renk, aynı anda veri bulundurma oranının %100, bordo renk %0 olduğunu belirtmektedir. Boşluk korelasyonu -1 ve 1 arasında değerler alır. -1 negatif korelasyondur ve *bir değişkende veri varsa diğer değişkende veri kesin yoktur* şeklinde yorumlanır. 0 korelasyonun olmadığını gösterir ve *bir değişkende verinin olup olmaması diğer değişkeni etkilemez* şeklinde yorumlanır. 1 pozitif korelasyonu belirtir ve *bir değişkende veri varsa diğer değişkende de kesin veri vardır* şeklinde yorumlanır.

8. *df* veri setinin bölme işlemi yapıldı.

Kas heteroplazmi oranını içeren ve içermeyen satırlar ayrılarak, *kas_df* ve *nan_kas_df* olarak isimlendirilen iki farklı veri seti oluşturuldu.

- *iskelet_kası* sütununda NaN olmayan değerler *notnull* fonksiyonu ile filtrelendi, *kas_df* isimli değişkene atandı.
 - *kas_df* veri seti boyutu: 421, 44 (sıra, sütun)
- *iskelet_kası* sütunundaki hücrelerin boş olduğu satırlar *isnull* fonksiyonu ile filtrelenerek *nan_kas_df* isimli değişkene atandı.
 - *nan_kas_df* veri seti boyutu: 323, 44 (sıra, sütun)

9. *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki NaN değerler için kas verisi üzerinden doldurma işlemleri yapıldı.

kas_df'te bulunan *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki kayıp değerler için tanımlayıcı istatistik çalışma uygulandı.

- Her iki sütun için (*laktik_asidoz* ve *dağılık_kırmızı_lifler*) *loc* fonksiyonuyla öznitelik varken (1) ve yokken (0), iskelet kası heteroplazmi oranlarının tepe değeri, ortalama değeri ve ortanca değeri (*mode*, *mean*, *median*) *describe* fonksiyonu ile belirlendi (Tablo 4-32).

Table 4-32: *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki farklı durumlar için iskelet kası heteroplazmi oranlarının istatistiksel dökümü

laktik_asidoz = 0		laktik_asidoz = 1		dağılık_kırmızı_lifler=0		dağılık_kırmızı_lifler=1	
count	72,000000	count	219,000000	count	49,000000	count	261,000000
mean	49,895139	mean	72,847489	mean	52,760204	mean	69,543678
std	21,279925	std	18,218484	std	27,550399	std	19,174440
min	0,200000	min	3,000000	min	0,200000	min	3,000000
25%	34,750000	25%	63,500000	25%	30,000000	25%	57,000000
50%	54,000000	50%	77,000000	50%	55,000000	50%	75,000000
75%	68,250000	75%	85,750000	75%	76,000000	75%	84,000000
max	86,000000	max	100,000000	max	95,000000	max	100,000000

Count = miktar, *mean* = ortalama, *std* = standart sapma, *min* = minimum sayı değeri, 25% = %25 oranındaki yoğunluk, 50% = %50 oranındaki yoğunluk, 75% = %75 oranındaki yoğunluk, *max* = maksimum sayı değeri

- Her bir durum için iskelet kası heteroplazmi oranlarının dağılımı incelendi. Değerlendirme sonrasında laktik asidoz için 45, dağılık kırmızı lifler için 60 sınır değer olarak alındı.
- Buna göre, iskelet kası heteroplazmi oranı:
 - ≥ 45 ise *laktik asidoz var* (=1)
 - < 45 ise *laktik asidoz yok* (=0)
 - ≥ 60 ise *dağılık kırmızı lifler var* (=1)
 - < 60 ise *dağılık kırmızı lifler yok* (=0)

olarak kabul edildi ve *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki boş hücreler (NaN) bu şekilde dolduruldu.

10. *nan_kas_df* veri setindeki *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki kayıp değerler için, iskelet kası heteroplazmi oranı bilgisi olmadığından diğer dokulardan elde edilen heteroplazmi bilgileri kullanılarak doldurma işlemleri uygulandı.

- Yalnızca kan dokusu heteroplazmi oranı bulunan satırlarda, *loc* fonksiyonu ile *laktik_asidoz* ve *dağılık_kırmızı_lifler* öznitelikleri varken (1) ve yokken (0) kan dokusu heteroplazmi oranının tepe değeri, ortalama değeri ve ortanca değeri (*mode*, *mean*, *median*) *describe* fonksiyonu ile gösterildi (Tablo 4-33).

Table 4-33: *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki farklı durumlar için kan heteroplazmi oranlarının istatistiksel dökümü

laktik_asidoz = 0		laktik_asidoz = 1		dağılık_kırmızı_lifler=0		dağılık_kırmızı_lifler=1	
count	145,000000	count	142,000000	count	198,000000	count	89,000000
mean	14,786207	mean	39,678873	mean	17,042424	mean	49,483146
std	10,490655	std	16,369305	std	9,224571	std	13,776860
min	0,000000	min	2,000000	min	0,000000	min	28,000000
25%	7,000000	25%	28,125000	25%	9,475000	25%	39,000000
50%	14,000000	50%	36,000000	50%	17,000000	50%	46,000000
75%	19,000000	75%	49,000000	75%	25,000000	75%	58,000000
max	60,000000	max	91,000000	max	35,200000	max	91,000000

count = miktar, *mean* = ortalama, *std* = standart sapma, *min* = minimum sayı değeri, 25% = %25 oranındaki yoğunluk, 50% = %50 oranındaki yoğunluk, 75% = %75 oranındaki yoğunluk, *max* = maksimum sayı değeri

- Her bir durum için kan heteroplazmi oranlarının dağılımı değerlendirildi ve laktik asidoz için 25, dağılık kırmızı lifler için 35 sınır değer olarak alındı.
- Böylece, kan heteroplazmi oranı:
 - ≥ 25 ise *laktik asidoz var (=1)*
 - < 25 ise *laktik asidoz yok (=0)*
 - ≥ 35 ise *dağılık kırmızı lifler var (=1)*
 - < 35 ise *dağılık kırmızı lifler yok (=0)*

olarak kabul edildi ve *laktik_asidoz* ve *dağılık_kırmızı_lifler* sütunlarındaki boş hücreler (NaN) bu şekilde dolduruldu.

- Kan dokusu heteroplazmi oranı bilgisi olmayan satırlar için diğer dokuların heteroplazmi oranı bilgileri kullanıldı.
 - *idrar, bukkal_mukoza, kıl_folikülleri, böbrek, fibroblast, kalp, karaciğer, beyin, beyin_kan_akımı, pankreas* sütunlardaki verinin azlığı nedeniyle istatistiksel bir işlem uygulanamadı.
 - Bu sütunlardaki değerlerle aynı satırda bulunan kas ve kan oranlarının tutarlılığı değerlendirildi.
 - Kas heteroplazmi oranına yakın değerler veren oranlar için madde 9'da belirtilen 45 ve 60 sınır değerleri kullanıldı.
 - Kan heteroplazmi oranına yakın değerler veren oranlar için ise bu maddede belirtilen 25 ve 35 değerleri sınır olarak kullanıldı.
 - *nan_kas_df*te *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarındaki boş kalan son hücreler de böylece dolduruldu.

11. Sütun kaldırma işlemi yapıldı.

Veri kaybı %90'den fazla olan sütunlar, kas verilerinin olmadığı veri setinde (*nan_kas_df*), kan heteroplazmi oranının verilmediği satırlarda, *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarındaki kayıp verilerin doldurma işlemi kullanıldıktan sonra *kas_df* ve *nan_kas_df* veri setlerinden çıkartıldı (Tablo 4-27, Tablo 4-34, Tablo 4-35).

- Veri kaybı %90'den fazla olan sütunlar, *drop* fonksiyonuyla, *kas_df* ve *nan_kas_df* veri setlerinden kaldırıldı.
- *kan* sütunu her iki veri setinden kaldırıldı.
- *iskelet_kası* sütunu, *kas_df*ten oluşturulacak olan regresyon modelinin, *nan_kas_df* veri setinde uygulanabilmesi için, *nan_kas_df*ten kaldırıldı.
 - *kas_df*ten kaldırılan sütunlar:
 - *kan, beyin_kan_akımı, kalp, beyin, karaciğer, pankreas, fibroblast, idrar, böbrek, bukkal_mukoza, kıl_folikülleri*
 - *kas_df* veri seti boyutu: 421, 33 (satır, sütun)

- *nan_kas_df*ten kaldırılan sütunlar:
 - *iskelet_kası, kan, beyin_kan_akımı, kalp, beyin, karaciğer, pankreas, fibroblast, idrar, böbrek, bukkal_mukoza, kıl_folikülleri*
 - *nan_kas_df* veri seti boyutu: 323, 32 (sıra, sütun)

kas_df ve *nan_kas_df* veri setlerindeki kayıp değerlerin miktarı ve yüzdeleri sırasıyla Tablo 4-34'te ve Tablo 4-35'te sunulmaktadır.

Table 4-34: *kas_df* veri setindeki kayıp değerlerin miktarları ve yüzdeleri

SÜTUN	KAYIP VERİ MİKTARI	KAYIP VERİ YÜZDESİ
kan	278	66,033254
beyin	404	95,961995
karaciğer	407	96,674584
kalp	408	96,912114
fibroblast	408	96,912114
böbrek	409	97,149644
kıl_folikülleri	411	97,624703
idrar	416	98,812352
pankreas	416	98,812352
bukkal_mukoza	418	99,287411
beyin_kan_akımı	419	99,524941

Table 4-35: *nan_kas_df* veri setindeki kayıp değerlerin miktarları ve yüzdeleri

SÜTUN	KAYIP VERİ MİKTARI	KAYIP VERİ YÜZDESİ
kan	36	11,145511
idrar	291	90,092879
bukkal_mukoza	292	90,402477
kıl_folikülleri	300	92,879257
böbrek	315	97,52322
fibroblast	318	98,452012
kalp	318	98,452012
beyin_kan_akımı	318	98,452012
karaciğer	320	99,071207
pankreas	322	99,690402
beyin	323	100
iskelet_kası	323	100

12. Anlamlı olmayan bağımsız değişkenlerin kaldırılması işlemi yapıldı.

- MELAS hastalığı ile direk ilişkisi bulunmayan, model denemesinde %95 güven aralığında sıfır (0) değeri bulunan (Şekil 4-11) ve p-değeri 0,05'ten yüksek (Şekil 4-12) olan klinik özellikler *drop* fonksiyonu ile her iki veri setinden çıkartıldı.
 - Çıkartılan sütunlar: *mental_hastalıklar*, *hipertansiyon*, *solunum_yetmezliği*, *yaşam_durumu*, *kalp_ritim_bozukluğu*, *kusma*
 - *kas_df* veri seti boyutu: 421, 27 (satır, sütun) (Şekil 4-9)
 - *kas_df* veri setindeki sütunların isimleri Şekil 4-9'da verilmiştir.

```
array(['cinsiyet', 'yaş', 'miyopati', 'dağınık_kırmızı_lifler',
      'ensefalopati', 'epilepsi', 'laktik_asidoz',
      'inme_benzeri_epizodlar', 'sensörinöral_işitme_kaybı', 'diyabet',
      'görme_bozukluğu', 'maküler_dejenerasyon', 'cpeo', 'pitozis',
      'kardiyomiyopati', 'nefropati', 'nöropati',
      'fokal_kortikal_defektler', 'jeneralize_kortikal_defektler',
      'serebellar_defektler', 'hareket_bozuklukları',
      'bazal_ganglion_kalsifikasyonu', 'kognitif_hastalıklar',
      'baş_ağrısı', 'gastrointestinal_disfonksiyonlar',
      'gelişme_geriliği', 'iskelet_kası'], dtype=object)
```

Şekil 4-9: *kas_df* veri setindeki sütunlar (öznitelik) (27 adet)

- *nan_kas_df* veri seti boyutu: 323, 26 (satır, sütun)
 - *nan_kas_df* veri setindeki sütunların isimleri Şekil 4-10'da verilmiştir.

```
array(['cinsiyet', 'yaş', 'miyopati', 'dağınık_kırmızı_lifler',
      'ensefalopati', 'epilepsi', 'laktik_asidoz',
      'inme_benzeri_epizodlar', 'sensörinöral_işitme_kaybı', 'diyabet',
      'görme_bozukluğu', 'maküler_dejenerasyon', 'cpeo', 'pitozis',
      'kardiyomiyopati', 'nefropati', 'nöropati',
      'fokal_kortikal_defektler', 'jeneralize_kortikal_defektler',
      'serebellar_defektler', 'hareket_bozuklukları',
      'bazal_ganglion_kalsifikasyonu', 'kognitif_hastalıklar',
      'baş_ağrısı', 'gastrointestinal_disfonksiyonlar',
      'gelişme_geriliği'], dtype=object)
```

Şekil 4-10: *nan_kas_df* veri setindeki sütunlar (öznitelik) (26 adet)

- MELAS hastalarında görülen önemli klinik özellikler, güven aralığında sıfır (0) değeri bulursa da ve p-değeri 0,05'ten yüksek olsa da çıkartılmadı.
- Eğitim güven aralığı değerleri, 4.2.3.1. *Regresyon Modelinin Değerlendirilmesi, İyileştirilmesi ve Sonuçları* başlığının 7. maddesi olan *Güven aralığı (Confidence Intervals) hesaplaması* başlığı altında hesaplandı. Belirtilen sütunların çıkartılmasına karar verildikten sonra model revize edildi.
- %95 güven aralığı hesaplama sonuçları Şekil 4-11'de sunulmaktadır.



	0	1
const	31.496059	50.791192
cinsiyet	0.086973	6.759600
yaş	-0.372185	-0.135507
yaşam_durumu	-8.954510	5.837087
miyopati	3.400178	10.628419
dağınık_kırmızı_lifler	11.493242	20.231804
ensefalopati	-3.866792	6.152304
epilepsi	-1.670439	7.766407
laktik_asidoz	14.098088	22.627666
inme_benzeri_epizodlar	-4.450360	4.431822
sensörinöral_işitme_kaybı	-2.051594	5.263655
diyabet	-2.409417	5.691385
görme_bozukluğu	-15.129625	-1.412641
maküler_dejenerasyon	-4.842126	7.175528
cpeo	-13.643913	-1.077083
pitozis	-5.986625	7.405590
kardiyomiyopati	-2.883518	7.017510
kalp_ritim_bozukluğu	-6.558600	6.486773
nefropati	-11.010674	5.027294
nöropati	-3.135621	7.669722
fokal_kortikal_defektler	-1.958450	9.814548
jeneralize_kortikal_defektler	-19.267110	-3.737697
serebellar_defektler	-3.666440	16.960837
hareket_bozuklukları	-6.826971	2.197912
bazal_ganglion_kalsifikasyonu	-3.259379	8.788764
kognitif_hastalıklar	2.028333	11.154453
mental_hastalıklar	-18.393747	9.125811
baş_ ağrısı	-3.393346	7.570133
kusma	-9.838274	3.412750
gastrointestinal_disfonksiyonlar	-18.696593	0.400459
solunum_yetmezliği	-10.848409	11.804656
gelişme_geriliği	-4.673054	3.070625
hipertansiyon	-21.740918	12.304353

Şekil 4-11: Model-1 bağımsız değişkenlerinin katsayı değerlerinin güven aralığı (%95) hesaplama sonuçları (sütun düşürme işleminden önce)

MELAS hastalığı ile direk ilişkisi bulunmayan, model denemesinde %95 güven aralığında sıfır (0) değeri bulunan klinik özellikler: *mental_hastalıklar*, *hipertansiyon*, *solunum_yetmezliği*, *yaşam_durumu*, *kalp_ritim_bozukluğu*, *kusma*

- P-değerleri, 4.2.3.1. *Regresyon Modelinin Değerlendirilmesi, İyileştirilmesi ve Sonuçları* başlığının 8. maddesi olan *P-değeri hesaplaması* başlığı altında hesaplandı. Değerlendirme sonrasında sütun çıkarma işlemi yapıldı ve model revize edildi. P-değeri hesaplama sonuçları Şekil 4-12’de sunulmaktadır.

est.pvalues	
const	9.564992e-16
cinsiyet	4.434718e-02
yaş	3.079878e-05
yaşam_durumu	6.788322e-01
miyopati	1.579491e-04
dağınık_kırmızı_lifler	4.713339e-12
ensefalopati	6.540428e-01
epilepsi	2.048264e-01
laktik_asidoz	5.343889e-16
inme_benzeri_epizodlar	9.967279e-01
sensörinöral_işitme_kaybı	3.885058e-01
diyabet	4.261991e-01
görme_bozukluğu	1.822551e-02
maküler_dejenerasyon	7.028568e-01
cpeo	2.180066e-02
pitozis	8.350899e-01
kardiyomiyopati	4.122005e-01
kalp_ritim_bozukluğu	9.913684e-01
nefropati	4.636919e-01
nöropati	4.098721e-01
fokal_kortikal_defektler	1.903035e-01
jeneralize_kortikal_defektler	3.792800e-03
serebellar_defektler	2.058562e-01
hareket_bozuklukları	3.138629e-01
bazal_ganglion_kalsifikasyonu	3.674431e-01
kognitif_hastalıklar	4.748499e-03
mental_hastalıklar	5.082754e-01
baş_ ağrısı	4.542908e-01
kusma	3.409910e-01
gastrointestinal_disfonksiyonlar	6.036115e-02
solunum_yetmezliği	9.338985e-01
gelişme_geriliği	6.843392e-01
hipertansiyon	5.860941e-01
dtype:	float64

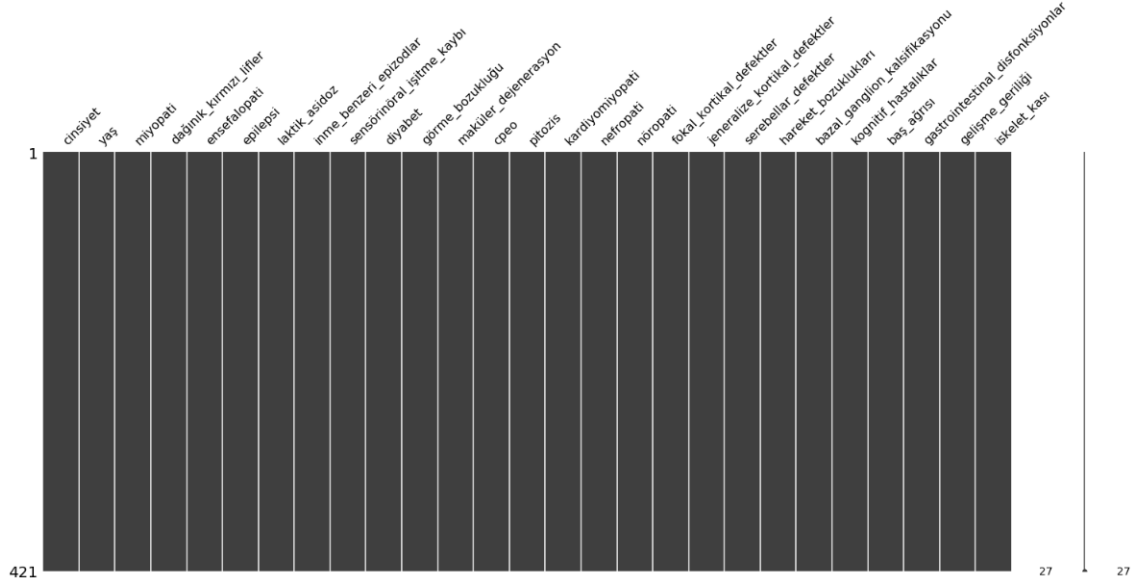
Şekil 4-12: P-değeri hesaplama sonuçları (sütun düşürme işleminden önce)

Model-1 bağımsız değişkenleri ile bağımlı değişken arasındaki ilişkinin sorgulanması için hesaplanan p-değeri 0,05’ten büyükse *ilişki yoktur*, 0,05’ten küçükse *ilişki vardır* şeklinde yorumlanır.

MELAS hastalığı ile direk ilişkisi bulunmayan, model denemesinde %95 güven aralığında sıfır (0) değeri bulunan klinik özelliklerin (*mental_hastalıklar*, *hipertansiyon*, *solunum_yetmezliği*, *yaşam_durumu*, *kalp_ritim_bozukluğu*, *kusma*) p-değerleri de yüksektir.

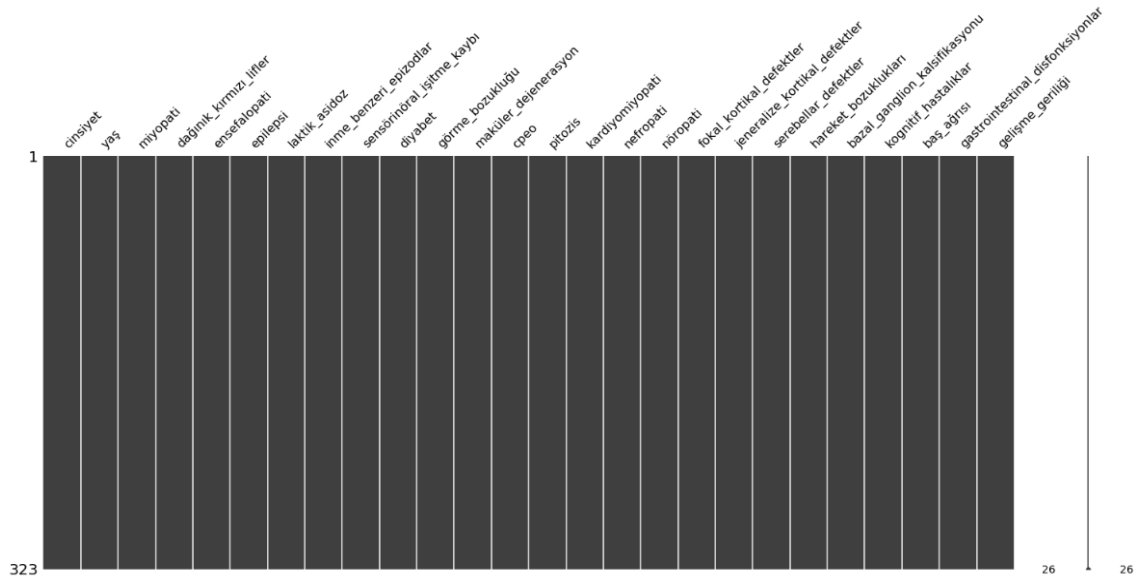
13. Veri dosyası, kayıp verilerden ve fazla sütunlardan arındırıldı. *Missingno* görüntüleme kütüphanesi *matrix* ve *Pandas* kütüphanesi *isnull* fonksiyonları aracılığıyla veri setlerinin NaN değerler içermediği doğrulandı. Uygulanan

işlemler sonucunda eksik verileri giderilen *kas_df* (421 satır, 27 sütun) ve *nan_kas_df* (323 satır, 26 sütun) veri setlerinin doluluk görseli sırasıyla Şekil 4-13 ve Şekil 4-14'te sunulmaktadır.



Şekil 4-13: *kas_df* veri seti doluluk görüntüsü

Bu şekilde veri setinin görüntüsü sunulmaktadır. Beyaz renkler kayıp verileri gösterirken, siyah renkler veri varlığını göstermektedir.



Şekil 4-14: *nan_kas_df* veri seti doluluk görüntüsü

Bu şekilde veri setinin görüntüsü sunulmaktadır. Beyaz renkler kayıp verileri gösterirken, siyah renkler veri varlığını göstermektedir.

Ham veri setinde, toplam 50 olan bağımsız değişken sayısı, son veri setinde 26'ya kadar düşürüldü. Yeterli veri içermeyen sütunların kaldırılmasının yanı sıra, p-değeri yüksek olan yani, bağımlı değişkenle ilişkisi anlamlı olmayan ve klinik açıdan m.3243A>G patojenik varyantı ile direkt ilişkisi bildirilmemiş sütunlar da çıkartıldı. Varyant ile direk ilişkisi bildirilmiş önemli klinik özelliklerin iskelet kası heteroplazmi oranıyla ilişkisi veri setine göre anlamlı bulunmasa da (p-değeri > 0,05), bu durum yetersiz veri sayısına bağlandı ve bu klinik özellikler için oluşturulan sütunlar veri setinden çıkartılmadı (Şekil 4-12). *iskelet_kası* bağımlı değişkenini tanımlamakta yeterli görülen tüm bağımsız değişkenler dışındaki sütunlar çıkartılmış oldu.

Makine öğrenmesi yöntemlerinin uygulanacağı veri setlerinde, aşırı uyum gösterme (overfitting) durumunun önüne geçilmesi için sütun sayısının olabildiğince az tutulması gerekmektedir. Veri setine gereğinden fazla sayıda bağımsız değişken eklemek, bağımlı değişkenin gereğinden fazla bağımsız değişkene bağlı olarak hesaplanması anlamına gelir. Bu durum, değişkenler arasındaki ilişkilerin ortaya çıkarılmasını güçleştirebilir ve oluşturulacak modelin öğrenme sürecine engel olup tahminlerin ezbere dayalı yapılmasına yol açabilir. Çok fazla değişken olduğunda, modelin, var olan ilişkileri öğrenmektense, eğitim verisindeki her bir senaryoyu ezberleyeceği ve test verisinde bu senaryolara denk gelmezse sonuçlara göre tahminde başarısız olacağı belirtilmektedir [153, p. 170]. Bu sebeple, veri setini anlamlı olmayan veya eksik veri içeren sütunlardan arındırmak modelin iyi işlemesine katkı sağlayacaktır.

14. Regresyon modeli oluşturmak üzere *kas_df* veri seti kullanıldı.

Heteroplazmi oranı tahmin modeli temelinde kurgulanan veri madenciliği çalışması sonucunda elde edilen veri seti, istenen tahminlerin hangi doku üzerinden yapılacağının önceden değerlendirilmesinin gerekliliğini gösterdi. Literatürde en fazla veri kas ve kan için sunulmuştu. Oluşturulan veri seti bilgisine göre 744 hastanın 421'inde iskelet kası, 430'unda kan heteroplazmi oranı bilgisi mevcuttu.

Kan hücrelerinin proliferasyonunun yüksek olması nedeniyle heteroplazmi oranlarının yaş ilerledikçe değişebildiği bilindiğinden, klinik ilişkilendirme için ideal bir doku olarak görülmemektedir [154]. Yüksek enerji gereksinimi olan iskelet kası, yüksek mitokondri sayısına sahiptir. Mitokondriyal solunum zincirindeki olası hasar durumunda, hastalığa sebep olan varyantın iskelet kasında yüksek oranda saptanması,

hastalardaki ağır klinik tablolarla ilişkilendirilmiştir [154]. İskelet kasının tanı testlerinde kullanımıyla, heteroplazmi oranı ile klinik arasındaki ilişkinin daha gerçekçi olarak kurulabilmesi sağlanmaktadır. Kas ve idrar dokuları ile kan dokusu karşılaştırmalarında, kan dokusundan elde edilen oranın diğer iki dokudan daha az olduğu görülmüştür [155]. Tüm bu bilgiler değerlendirilerek, heteroplazmi oranı ve klinik özelliklerin ilişkilendirilmesi yoluyla oluşturulacak tahmin programı için kas dokusunun kullanılmasının daha uygun olacağı belirlendi.

744 satırlık *df* veri setinden bir regresyon modeli oluşturabilmek için öncelikle veri setinde hiç boşluk bulunmaması, *iskelet_kası* sütunundaki boşlukların doldurulması gerekti [156]. Bu nedenle, kas heteroplazmi oranlarının verildiği satırlardan oluşturulan veri seti (*kas_df*) kullanılarak oluşturulacak Model-1 ile bu oranın verilmediği satırlardan oluşturulan veri setindeki (*nan_kas_df*) kayıp verilerin doldurulması hedeflendi.

4.2.2. Veri Setinde Yapılan Testler ve Sonuçları

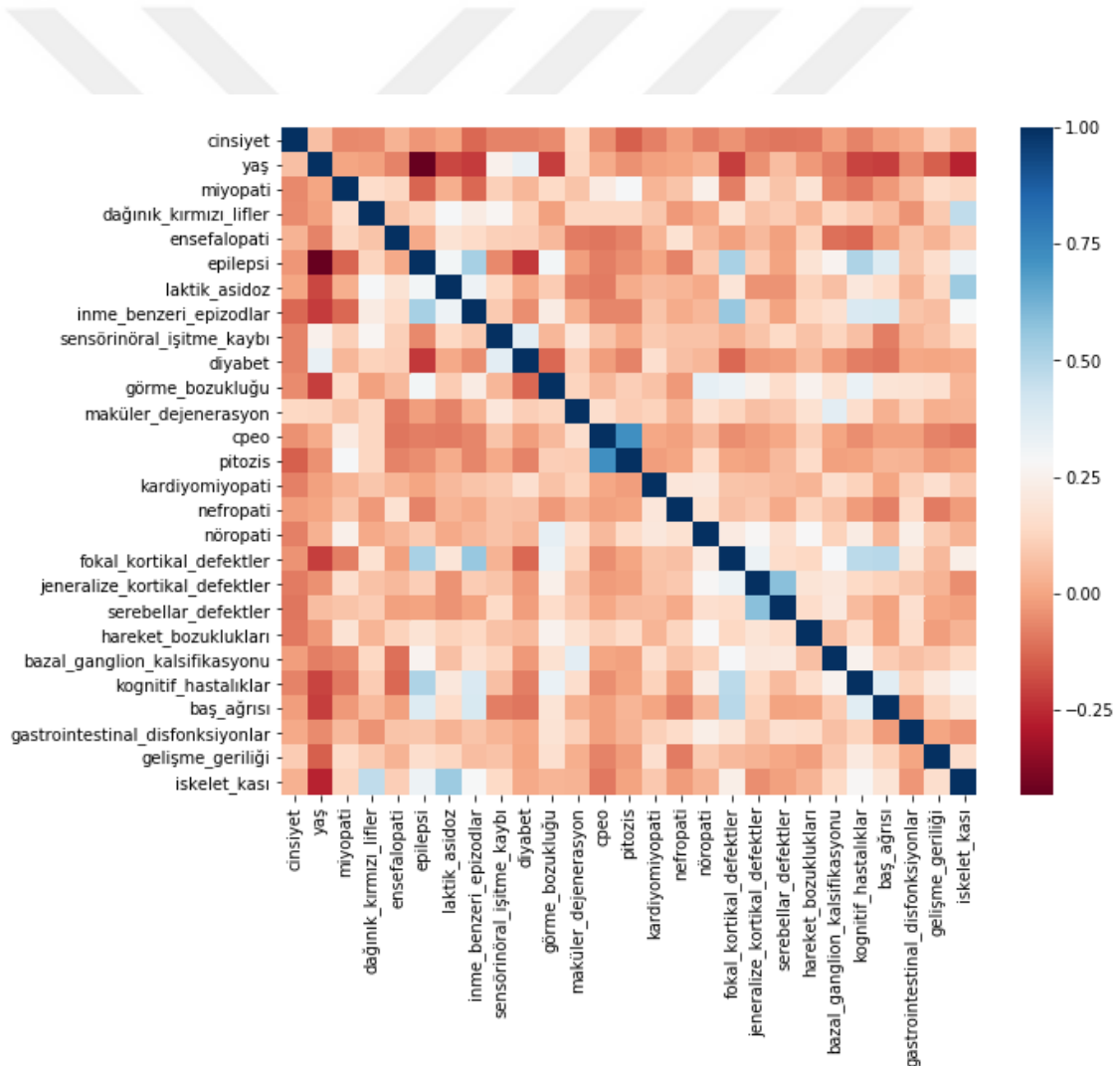
İlk modelin oluşturulmasında *kas_df* veri seti kullanıldı. Model oluşturma işleminden önce, veri setinin model oluşumuna uygunluğunu tespit etmek için, veri seti çoklu eşdoğrusallık ve uç değer testlerinden geçirildi.

4.2.2.1. Çoklu Eşdoğrusallık (*Multicollinearity*) Testi ve Sonuçları

Bağımlı değişkenlerin birbirleri arasındaki doğrusal (lineer) ilişkiyi dışlamak amacıyla, çoklu eşdoğrusallık kontrolü yapıldı. Veri setinde tüm bağımsız değişkenlerin bağımlı değişkenle ilişkisi olmalı ancak bağımsız değişkenlerin birbirleri arasında ilişki olmamalıdır [157]. Bu durumun tespiti için korelasyon matrisi oluşturuldu (Şekil 4-15) ve varyans büyütme faktörü (**VIF**, *variance inflation factor*) ölçeği kullanıldı (Şekil 4-16). Korelasyon değerleri kontrol edilerek, veri noktaları arasındaki olası ilişki hakkında bilgi elde edildi. VIF sonuçlarına göre, bağımsız değişkenlerin birbirleri arasındaki ilişki dışlandı. *iskelet_kası* sütununun, korelasyon matrisinde, diğer sütunlarla ilişkisini renklere bakarak anlamak mümkündür. *iskelet_kası* sütununun *yaş* sütunuyla ilişkisinin en düşük, *laktik_asidoz* sütunuyla ilişkisinin en yüksek olduğu görülmektedir (Şekil 4-15). VIF sonuçları, bağımlı değişken olarak kabul edilecek olan *iskelet_kası*'nın, bağımsız değişkenlerle arasında güçlü bir korelasyon bulunmadığını gösterdi ($2,03 < 5$) (Şekil 4-16). Ancak zayıf olsa da bir ilişkinin olduğu ve veri setinin

çoklu doğrusal (multi-lineer) regresyon analizi için uygun olduğu sonucuna ulaşıldı. Uygulanan işlemler aşağıdaki akış doğrultusunda gerçekleştirildi:

1. Bir değişkenin diğer değişkenlerle ilişkisinin belirlenebilmesi için, *Pandas* kütüphanesi *corr* fonksiyonuyla korelasyon matrisi oluşturuldu.
2. Korelasyon matrisi *Seaborn* kütüphanesi ısı haritası fonksiyonu olan *heatmap* kullanılarak görüntülendi (Şekil 4-15).
3. Değişkenlerin standart hataya etkisi, bir eşdoğrusallık ölçeği olan VIF ile *variance_inflation_factor* fonksiyonu kullanılarak belirlendi.
4. Bulunan VIF değerleri 5'ten küçük ise çoklu eşdoğrusallık dışlandı (Şekil 4-16).



Şekil 4-15: kas_df veri setinin korelasyon matrisi görüntüleme sonucu

(bordo renk) korelasyonun %0 olduğunu, çubuğun ortası (beyaz renk) ise ortalama bir ilişkiyi belirtmektedir. Matriste çapraz geçen lacivert renkler, aynı sütunun x ve y düzlemindeki kesişim noktalarıdır. İskelet kası sütunu y ekseninin en altında yer almaktadır ve x eksenı boyunca diğer sütunlarla olan ilişkisi renklerle anlaşılmaktadır.

VIF < 5 koşulunun sağlandığı öznitelikler için çoklu eşdoğrusallık dışlandı (Şekil 4-16).

const	24.390185
cinsiyet	1.189861
yaş	1.581739
miyopati	1.415862
dağınık_kırmızı_lifler	1.510587
ensefalopati	1.245542
epilepsi	2.185951
laktik_asidoz	1.608523
inme_benzeri_epizodlar	1.933027
sensörinöral_işitme_kaybı	1.403564
diyabet	1.359969
görme_bozukluğu	1.471038
maküler_dejenerasyon	1.395882
cpeo	2.314700
pitozis	2.404471
kardiyomiyopati	1.184748
nefropati	1.191796
nöropati	1.457097
fokal_kortikal_defektler	2.275682
jeneralize_kortikal_defektler	1.903944
serebellar_defektler	1.698647
hareket_bozuklukları	1.294599
bazal_ganglion_kalsifikasyonu	1.396200
kognitif_hastalıklar	1.771069
baş_ ağrısı	1.498262
gastrointestinal_disfonksiyonlar	1.185783
gelişme_geriliği	1.221365
iskelet_kası	2.027436
dtype: float64	

Şekil 4-16: kas_df veri setinin çoklu eşdoğrusallık testi sonuçları

4.2.2.2. Uç Değer (Outliers) Testi ve Sonucu

Uç değer testi ile, veri setinin uç değerler içerip içermediğinin kontrolü yapıldı [158]. Yapılan işlemler şu şekildedir:

1. Regresyon doğrusunda sapmaya neden olabilecek olası uç değerlerin varlığı test edildi.
2. Veri setine *describe* fonksiyonu uygulanarak istatistiksel özet çıkartıldı.

3. Bulunan standart sapma değerleri 3 ile çarpıldı. Sonuç, ortalamadan çıkartıldı ve ortalamaya eklendi. Böylece, taban ve tavan değerleri elde edildi. Taban değerinin altında kalan ve tavan değerinin üstünde kalan değerler, uç değer olarak kabul edildi. Her bir değişken için, uç değer sorgulaması yapıldı.

Kontroller sonucunda veri setinin uç değer içermediği görüldü.

4.2.3. Regresyon Modeli Uygulaması

kas_df veri setinin %80'i kullanılarak, *iskelet_kası* bağımlı değişkeninin bağımsız değişkenlerle (*iskelet_kası* dışındaki tüm değişkenler) olan ilişkisi, çoklu doğrusal (multi-linear) regresyon modeli eşitliğine oturtuldu [158]. Eşitlik Denklem 4-1'de gösterilmektedir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{26} x_{26} + \varepsilon \quad (4-1)$$

y: Bağımlı değişken

β_0 : y-kesen değeri

$x_1, x_2, \dots, x_{26}$: Bağımsız değişkenler

$\beta_1, \beta_2, \dots, \beta_{26}$: Katsayı değerleri

ε : Hata değeri

Eşitlikte β_0 olarak geçen y-kesen değeri, yani tüm bağımlı değişkenler sıfır iken, y değerinin eşit olduğu değer 38,08 bulundu. Bu değer, oluşturulan modele göre, herhangi bir klinik bulgusu olmayan ancak iskelet kasında varyantı taşıyan bir olgunun heteroplazmi oranıdır. Her bir bağımsız değişken ($x_1, x_2, \dots, x_{26}$) (*iskelet_kası* dışındaki tüm sütunlar), değeri sıfır (0) değilse, negatif veya pozitif olabilen katsayı değerleri ($\beta_1, \beta_2, \dots, \beta_{26}$) ile eşitliğe etki edecek ve bağımlı değişken değerini (*iskelet_kası* heteroplazmi oranı) yükseltecek ya da düşürecektir. Katsayı değeri, her bir değişken için özel olup, bir değişkenin bir birimlik değişiminde, diğer tüm değişkenlerin etkisi sıfırken, bağımlı değişkendeki tahmini değişimdir. ε , hata değeridir. Hata, bağımlı değişken birimiyle ifade edilir. Yukarıda verilen denklemdeki y-kesen değeri, bağımsız değişkenlerin değerleri ve katsayı değerleri, y bağımlı değişkeninin bir kısmını açıklamakta yeterli olur. Açıklanamayan varyasyonlar ise hata değerine yansır. Bu değer, veri noktalarının regresyon çizgisine olan ortalama uzaklığıdır. Veri noktalarının, oluşturulan regresyon çizgisi etrafında ne kadar yoğunlaştığının bilgisini verir. Hata

değeri ne kadar düşükse, veri noktaları regresyon çizgisine ne kadar yakınsa, regresyon modeli o kadar iyi çalışıyor anlamına gelmektedir [158].

Yapılan işlemler sırasıyla şu şekildedir:

1. *iskelet_kası* bağımlı değişkenini tahmin etmekte kullanılacak olan bağımsız değişkenler, *iskelet_kası* sütunu *drop* fonksiyonu ile düşürüldükten sonra *X* değişkenine atandı.
2. *iskelet_kası* bağımlı değişkeni *Y* değişkenine atandı.
3. *X* ve *Y* değişkenlerinin eğitim ve test verisi, *sklearn.model_selection.train_test_split* fonksiyonuyla oluşturuldu.
 - Varsayılan olarak, eğitim için *kas_df* veri setinin %80'i, test için geri kalan %20'si kullanıldı.
4. *X* ve *Y* eğitim verileri *sklearn.linear_model.LinearRegression* fonksiyonu kullanılarak regresyon modeline yerleştirildi ve Model_1 oluşturulmuş oldu.
5. Tüm bağımsız değişkenlerin bağımlı değişkene etkisi sıfırken bağımlı değişkenin kesen değeri, *intercept* fonksiyonu ile yazdırıldı.
 - *iskelet_kası* bağımlı değişkenin y-kesen değeri ($\beta_0 = y\text{-intercept}$): **38,08**
6. Diğer bağımsız değişkenler sabitken, her bir bağımsız değişkendeki bir birimlik değişime bağlı olarak bağımlı değişkenin tahmini değişimi olan katsayı değerleri, *coef* fonksiyonu ile yazdırıldı.
7. Oluşturulan regresyon modeli, *predict* fonksiyonuyla *X_test* veri setine uygulandı ve test verisindeki bağımsız değişkenler kullanılarak bağımlı değişken olan kas heteroplazmi oranı tahmin ettirildi.
8. Tahminlerin gerçek veriye yakınlığı, regresyon çizgisine yakınlığı, modelin anlamlılığı durumlarını değerlendirmek üzere sonraki adıma geçildi.

4.2.3.1. Regresyon Modelinin Değerlendirilmesi, İyileştirilmesi ve Sonuçları

Veri noktalarının tamamının üzerinden geçen bir doğrunun çizimi mümkün değildir ve oluşturulan regresyon çizgisi mükemmel olmayacaktır. Bulunan y değerleri (Denklem 4-1) yaklaşık değerler olacaktır, her biri birer tahmin sonucudur. Amaç, gerçek değere en yakın sonucu veren bir regresyon modeli oluşturmak, veri setine en iyi uyum sağlayan doğruyu bulmaktır. Popülasyonun tamamını test etmek olanak dışıdır. Bu nedenle genel uygulamada, eldeki örnek veriyle, popülasyon parametrelerinin tahmini sonuçlarına ulaşmak hedeflenir. Bilinmeyen bu parametrelerin, yani bağımsız değişkenlerin katsayı değerlerinin ($\beta_1, \beta_2, \dots, \beta_{26}$) ve kesen-değerin (β_0) (Denklem 4-1) en optimal değerlerine ulaşmak için, *en küçük kareleri tahmin edici* yöntemlerinden birinin kullanılması gerekmektedir. Amaç, gözlemlenen veri noktaları ile tahmin edilen veri noktaları arasındaki farkların karelerinin toplamının en aza indirilmesini sağlamaktır. Gauss-Markov teoreminin, en düşük örneklem varyansına sahip olduğu için, en iyi doğrusal yansız tahmin aracı olarak değerlendirdiği OLS (*Ordinary Least Squares* = sıradan en küçük kareler) metodu, Model_1 için kullanıldı. Ancak OLS'nin kullanılabilmesi için birtakım ön şartların sağlanmış olması gerekmektedir [153]. Bu ön şartlar aşağıda belirtilen dört madde altında özetlenmektedir:

- Veri setinin homoskedastik olması yani hata değerlerinin yaklaşık olması beklenir. Veri seti heteroskedastik olamamalı yani hata değerleri regresyon çizgisinin belli bölümünde artış ya da azalma göstermemelidir.
- Hataların ilişkili olmaması istenir. Bu yüzden otokorelasyon dışlanmalıdır.
- Artıklar normal dağılımlı olmalıdır.
- Artıkların ortalaması sifıra eşit olmalı veya sifıra yakın bir değer vermelidir.

Yapılan işlemler sırasıyla aşağıda verildiği gibidir:

1. İçerdiği fonksiyonlar sayesinde, modelin değerlendirilmesi sürecinde hesaplamaları kolaylaştıran, Python *statsmodels* paketi kullanıldı.
2. OLS fonksiyonu kullanılarak bağımlı değişken ve bağımsız değişkenler modele yerleştirildi. Sonraki testler bu model üzerinden yapıldı.
3. OLS metodunun varsayımlarının kontrolü yapıldı.
 - Varsayım_1: Heteroskedastisite (değişen varyans sorunu) yok.

Heteroskedastisite, bir değişkenin standart hatasının belli bir zaman aralığında sabit olmaması durumudur. Heteroskedastisite durumunda katsayı tahminlerindeki değişkenlik (varyans) artar. *OLS* fonksiyonu artan değişkenliği tespit edemez ve gerçek olandan daha düşük olan değişkenlik değerini kullanır. Katsayı değerlerinin tahmini değeri, p-değeri, t-değeri ve F-değeri olması gerektiğinden daha düşük bulunmuş olabilir. Model anlamlı değilken anlamlı olarak değerlendirilmesine yol açabilir. Heteroskedastisite durumu varsa, *OLS* fonksiyonu kullanılamaz [153].

- Heteroskedastisite, genellikle yatay kesit verileriyle tahmin yapıldığında ortaya çıkmaktadır, bu sebeple model değerlendirmesi için kontrol edildi.
- Varsayım_1'in kontrolü için *statsmodels.stats.diagnostic* paketi kullanıldı. Bu paket test fonksiyonları içerir. Daha genel durumlar için kullanılan Breusch-Pagan testi için *diag.het_breuschpagan* fonksiyonu, daha spesifik durumlar için kullanılan White testi için *diag.het_white* fonksiyonu kullanılmak üzere iki test üzerinden kontrol yapıldı.
- Sıfır hipotez (H_0) ve alternatif hipotez (H_1) üretildi.
 - H_0 : Standart hatadaki değişkenlikler birbirine eşittir (Denklem 4-2).

$$H_0: \sigma_{2i} = \sigma_2 \quad (4-2)$$

H_0 = Sıfır hipotez, σ_2 = Varyans değeri

- H_1 : Standart hatadaki değişkenlikler birbirine eşit değildir (Denklem 4-3).

$$H_1: \sigma_{2i} \neq \sigma_2 \quad (4-3)$$

H_1 = Alternatif hipotez, σ_2 : Varyans değeri

- Standart hataların eşit olması durumu, heteroskedastisiteyi dışlamak için istenilen durumdur ve p-değerinin 0,05'ten yüksek olması (H_1 hipotezinin dışlanması) beklenir.
- Test sonucunda heteroskedastisite durumunun olmadığı görüldü (Tablo 4-36).

Table 4-36: Heteroskedastisite testleri ve sonuçları

TEST	P-DEĞERİ	SONUÇ
White Testi	0,07699	HETEROSKEDASTİSİTE YOK
Breusch-Pagan Testi	0,09608	HETEROSKEDASTİSİTE YOK

- Varsayım_2: Otokorelasyon yok.

Veri setinde, zamansal veya uzaysal olarak sıralanmış değerler arasında serisel ilişki olabilmektedir. Bu ilişki ile sonraki sayının, oluşturulan modelden bağımsız olarak tahmin edilebilir olması durumu otokorelasyon olarak tanımlanır. Örneğin, sıcaklık değerlerinin bir ay içerisinde birbirini takip eden günlerde yakın değerler vermesi veya belli bir değişkenin coğrafi yakınlık durumunda benzer sonuçlar vermesi gibi. OLS'nin kullanılabilmesi için verideki değerlerin bağımsız olması, değerler arasında bir örüntünün bulunmaması gerekmektedir. Otokorelasyon durumunda, örneklerin bağımsızlığı ön şartı ihlal edilmiş olur ve OLS'nin kullanılması durumunda, bulunan standart hata ve p-değerleri yanıltıcı olabilir. En düşük varyansla katsayı değerlerinin tahmini başarılı olamaz [159, p. 24].

Bu tez çalışmasında oluşturulan veri setinde, örnekler birbirinden bağımsız olduğu düşünülerek otokorelasyon durumu beklenmedi. Yine de aynı değişkenin serisel değerleri arasında korelasyon varlığını dışlamak amacıyla otokorelasyon kontrolü yapıldı.

- Otokorelasyon kontrolü *statsmodels.stats.diagnostic* paketi ile yapıldı. Ljung-Box testi için paketin içerisinde yer alan *acorr_ljungbox* fonksiyonu kullanıldı.
- Ljung-Box testi, sezonsal olmayan zaman serileri için belirlenen gecikme (*lag*) sayısı kullanıldı (Denklem 4-4).

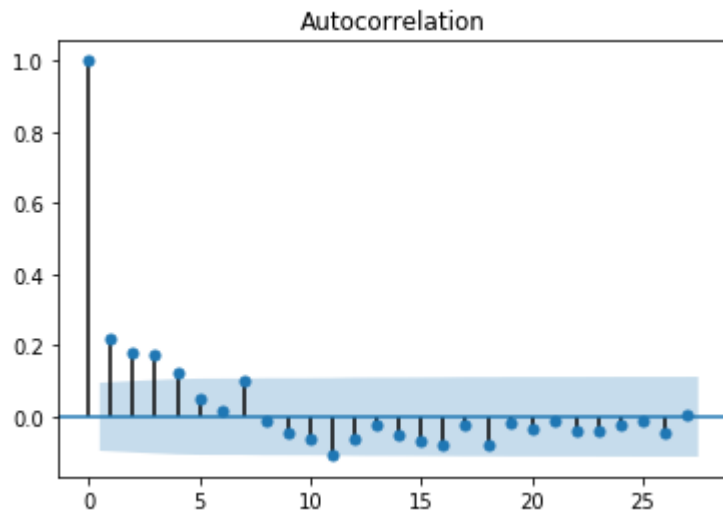
$$\text{test edilen gecikme sayısı} = \text{minimum}(10, (X \text{ değişkeninin uzunluğu} // 5)) \quad (4-4)$$

- *acorr_ljungbox* fonksiyonuna model artık bilgileri (*residuals*) verildi.

- Test kapsamında sıfır hipotez (H_0) ve alternatif hipotez (H_1) üretildi.
 - H_0 : Veri rastgeledir.
 - H_1 : Veri rastgele değildir.
- Verinin rastgele olması (H_0 hipotezi) istenilen durumdur. Otokorelasyonu dışlamak için hesaplanan p-değerinin 0,05'ten yüksek olması (H_1 hipotezinin dışlanması) gerekir.
- Otokorelasyonun görsel testi, *statsmodels.graphics* paketinin *tsa.plot_acf* fonksiyonuna modelin artık değerlerinin (*residuals*) bilgisi verilerek yapıldı.
- Test sonucuna göre otokorelasyonun var olduğu görüldü (Tablo 4-37, Şekil 4-17).

Table 4-37: Otokorelasyon testi ve sonucu

TEST	GECİKME	EN KÜÇÜK P_DEĞERİ	SONUÇ
Ljung-Box testi	10	4,46e-11	OTOKORELASYON VAR



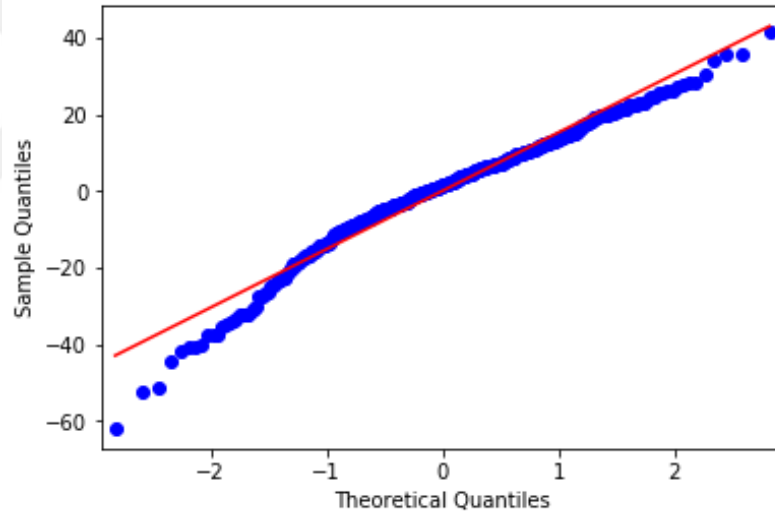
Şekil 4-17: Model-1'in otokorelasyon testi görsel sonucu

x eksen: gecikmeler, y eksen: korelasyonlar

Yapılan testler heteroskedastisitenin olmadığını (Tablo 4-36), ancak otokorelasyonun varlığını gösterdi (Tablo 4-37). Bu sebeple, bu tez çalışmasında OLS yerine, bir GLS³⁴ metodu olan GLSAR³⁵ metodu kullanıldı. Bir tahmin edici olarak GLSAR'ın otokorelasyon ve heteroskedastisite (değişen varyans) durumlarında elverişli bir metot olduğu bildirilmiştir [160].

4. Artıkların normal dağılım kontrolü yapıldı.

- Grafik çizimi için çizim kütüphanesi *PyLab* kullanıldı.
- Artık bilgileri grafiğe *statsmodels* paketinin *sm.qqplot* fonksiyonu ile verildi.
- Verilerin regresyon çizgisine yakın, normal bir dağılım göstermesi beklendi.



Şekil 4-18: Artık değerlerin regresyon çizgisi etrafındaki dağılımının görsel sonucu (Model-1)

Kırmızı çizgi: regresyon doğrusu, mavi noktalar: veri noktaları

Artıkların normal dağılım kontrolü için çizilen grafikte, beklenildiği gibi regresyon çizgisine yakın olan veri noktalarının dağılımı görülmektedir (Şekil 4-18).

³⁴ GLS: Genelleştirilmiş en küçük kareler (*generalized least squares*) [160]

³⁵ GLSAR: Otokoregresif kovaryans yapısı ile genelleştirilmiş en küçük kareler (*generalized least squares with AR (autoregressive) covariance structure*) [160]

5. Artık ortalamasının sıfıra eşit olup olmadığının kontrolü yapıldı.

- Ortalama hesaplaması artıkların toplamının artık sayısına bölünmesi ile bulundu.
- Artık ortalaması, hata değerleri yuvarlama işleminden geçtiği için çoğunlukla sıfıra eşit olmaz. Bulunan değer sıfıra yakın olması beklendi.
- Artık Ortalama Değeri: **-0,00122**

Artık ortalama değeri sıfıra yakın bir değer vermiştir. Bu da istenilen değer olmuştur. Regresyon doğrusunun iki tarafında kalan veri noktalarının tahmin değerlerine yani regresyon doğrusuna olan uzaklıklarının değerlerinin toplamı sıfır olmalıdır. Ancak gerçek dünya verisiyle sıfır elde edilemeyeceği, değer sıfıra yakın bir değerse kabul göreceği bildirilmiştir [153].

6. Hata değerleri hesaplandı.

Regresyonda amaç, veri noktaları boyunca bir doğru oluşturmak ve bunu yaparken artık değerlerini olabildiğince küçük tutmaktır. Artık değer, gözlemlenen veri noktası ve regresyon doğrusuyla elde edilen tahmini veri noktası arasındaki mesafeyi (hata değeri) belirtir. İşlemlerde bu mesafenin karesi alınır. Bunun sebebi fark değerinin tamamını pozitif yapmak ve gözlemlenen veri değerinden büyük oranda sapma gösteren tahmin veri değerlerinin vurgulanmasını sağlamaktır. Tüm veri noktaları için hesaplanan artık değerlerin kareleri toplanır ve modelin uyumu hakkında bilgi verici olan artıkların kareleri toplamı (**SSE, Sum of Squared Error**) değeri elde edilir. Ortalama mutlak hata (**MAE, Mean Absolute Error**), gözlemlenen ve tahmin edilen değerlerin farkının mutlak değerinin, $(n-k)^{36}$ 'ya bölünmesiyle elde edilir. k değeri artıkların serbestlik derecesi olup kurulan regresyon eşitliğinde tahmin edilen tüm parametrelerin sayısıdır. Bu parametreler, katsayı değerleri ($\beta_1, \beta_2, \dots, \beta_{26}$) ve y-kesen (β_0) değeridir. MAE, hatanın ortalama büyüklüğünü verir. Tüm hata değerlerini ortalama üzerinden değerlendireceği için, büyük hata değerlerinin göz ardı edilmesine neden olabilir. Tahmin etme işleminin başarı gücünü belirlemek için ortalama kare hata (**MSE, Mean Squared Error**) değeri kullanılır. Tahmin değerlerin gerçek değerlerden uzaklığının karesinin ortalama değeridir. Büyük sapma gösteren değerlerin karesinin alınması yoluyla, sapma değerlerinin büyüklükleri vurgulanır. Ancak ortalama değer olduğu için büyük hata

³⁶ n = örnek sayısı, k = artıkların serbestlik derecesi (*degrees of freedom*)

değerlerinin cezai değeri hesaba katılmamış olur. Hata büyüklüğünün ölçülmesinde MAE ve MSE'den daha çok kabul gören hata değeri, ortalama kare hatanın karekök değeridir (**RMSE**, *Root Mean Squared Error*) ve bu değer regresyon modelinin ne kadar iyi tahmin yapabildiğinin bilgisini verir. Standart hata olarak da bilinir. RMSE değeri de hatanın ortalama büyüklüğünü verir ancak büyük hata değerlerine ceza değeri ekler. Bu yüzden değeri MAE'den daha yüksektir. RMSE, veri noktalarının regresyon doğrusu etrafında ne kadar yayıldığına bilgisini verir. Hata değerleri sıfırdan sonsuza kadar her değeri alabilir. Bu değerlerin sıfır olması hiç hata olmadığına göstergesidir. İyi bir performans göstergesi için hata değerlerinin sıfıra yakın olması beklenir [158].

Bu bilgiler ışığında, oluşturulan regresyon doğrusunun veriye uyumu, bu hata hesaplamaları ile değerlendirildi. İskelet kası heteroplazmi oranı test verisi üzerinden, gerçek heteroplazmi oranları ve tahmin edilen heteroplazmi oranları karşılaştırıldı.

- Kullanılan metrikler:
 - Ortalama mutlak hata (**MAE**, *Mean Absolute Error*):
 - Gerçek ve tahmin değerleri arasındaki farkın mutlak değeri `sklearn.metrics.mean_absolute_error` fonksiyonuyla hesaplandı.
 - Ortalama kare hata (**MSE**, *Mean Squared Error*)
 - `sklearn.metrics.mean_squared_error` fonksiyonu kullanılarak hesaplandı.
 - Kök ortalama kare hata (**RMSE**, *Root Mean Squared Error*)
 - Hata büyüklüğünün ölçülmesinde kullanılan bu metrik MSE değerinin karekökünün alınması ile hesaplandı. Python *Math* paketinin `sqrt` fonksiyonu kullanılarak bulundu.
- Hata değeri hesaplama sonuçları Tablo 4-38'de sunulmaktadır.

Table 4-38: Model-1 için hata değerleri hesaplama sonuçları

MAE	12,0
MSE	2,78E+02
RMSE	16,7

MAE = Ortalama mutlak hata (*Mean Absolute Error*), MSE = Ortalama kare hata (*Mean Squared Error*), RMSE = Kök ortalama kare hata (*Root Mean Squared Error*)

Model-1 için yapılan hata hesaplamaları sonucunda veri noktaları ve regresyon çizgisi arasındaki mesafenin sıfıra yeteri kadar yakın olmadığı görüldü. Sonuçlar, örnek veri ile oluşturulan regresyon modelinin tahmin edeceği sonuçlar ile gerçek sonuçlar arasında ortalama %12 fark olacağını gösterdi (Tablo 4-38). Ancak büyük hata değerlerinin ağırlık verildiği RMSE değerine göre, veri noktaları ve regresyon doğrusu arasındaki mesafenin ortalama değeri %16,7 idi (Tablo 4-38). Buradaki yüzde, bağımlı değişkenin birimidir. Model-1'in ulaştığı iskelet kası heteroplazmi oranının, gerçek orandan ortalama 16,7 sapma göstereceği anlamına gelir.

7. Güven aralığı (*Confidence Intervals*) hesaplaması yapıldı.

Bağımlı değişkenin her bir bağımsız değişkenle anlamlı doğrusal bir ilişkisinin olup olmadığının test edilmesi gerekmektedir. Bağımlı değişkenlerin anlamlılığının anlaşılması için, her bir bağımsız değişkenin bağımlı değişkeni etkileme derecesi olan eğim değerlerinin sıfıra eşit olmadığı gösterilmelidir. Bunun için eğimin %95 güven aralığı içerisinde verilen en küçük ve en büyük değerleri arasında sıfır (0) değerinin varlığı kontrol edildi. Regresyon çalışmalarında yapılan işlemler örnek veri üzerinden yapıldığı için bulunan değerler tahminden ibaret olur. Bu yüzden bulunan değerlerin etrafında başka olası değerlerin de doğru olabileceğini gösteren bir güven aralığı oluşturulur. Güven aralığı, veri örneklerinin popülasyon içinden seçimi 100 kez yapıldığı durumda 95'inde doğru katsayı değerlerine ulaşılacağı anlamına gelir. %95 oranında verilen aralığın doğru katsayı değerini içerdiğine güven duyulur. %95'lik dilime giren, doğru sayılabilecek tüm katsayı değerlerinin en küçük ve en büyük değerleri aralık şeklinde ifade edilir. Bu aralık içinde sıfır değeri varsa, o bağımsız değişkenin örnek veri üzerinden bağımlı değişkeni hiç etkilemediği sonucuna da ulaşılır. Bu istenmeyen durumdur ve test edilerek dışlanmalıdır [158]. Yapılan katsayı güven aralığı oluşturma işlemi sonucuna göre, birçok bağımsız değişkenin eldeki veri ile bağımlı değişkeni açıklamakta yetersiz olduğu görüldü (Şekil 4-19).

- %95 güven aralığı, *statsmodel* paketi *conf_int* fonksiyonuyla hesaplandı.
- Güven aralığı hesaplama sonuçları Şekil 4-19'da verilmektedir.

	0	1
const	33.363408	46.024458
cinsiyet	-0.103362	6.469347
yaş	-0.367272	-0.138996
miyopati	3.290491	10.434409
dağınık_kırmızı_lifler	11.678571	20.275037
ensefalopati	-4.180472	5.666396
epilepsi	-1.471354	7.849270
laktik_asidoz	14.156573	22.600560
inme_benzeri_epizodlar	-4.351393	4.389448
sensörinöral_işitme_kaybı	-1.859582	5.309522
diyabet	-2.145524	5.835229
görme_bozukluğu	-13.997820	-0.977508
maküler_dejenerasyon	-4.864682	7.041348
cpeo	-13.829709	-1.443478
pitozis	-5.538934	7.579927
kardiyomiyopati	-2.572832	6.885676
nefropati	-11.218036	4.256909
nöropati	-2.825602	7.814218
fokal_kortikal_defektler	-2.455459	8.527877
jeneralize_kortikal_defektler	-18.948736	-4.200351
serebellar_defektler	-3.354457	16.871776
hareket_bozuklukları	-6.707564	2.223093
bazal_ganglion_kalsifikasyonu	-3.598633	8.236871
kognitif_hastalıklar	1.705976	10.538906
baş_ ağrısı	-4.187394	5.369252
gastrointestinal_disfonksiyonlar	-18.054499	0.192975
gelişme_geriliği	-4.929405	2.524701

Şekil 4-19: Model-1 bağımsız değişkenlerinin katsayı değerlerinin güven aralığı (%95) hesaplama sonuçları (sütun düşürme işleminden sonra)

MELAS hastalığı ile direk ilişkisi bulunmayan, model denemesinde %95 güven aralığında sıfır (0) değeri bulunan klinik özellikler (mental_hastalıklar, hipertansiyon, solunum_yetmezliği, yaşam_durumu, kalp_ritim_bozukluğu, kusma) düşürüldükten sonra yapılan güven aralığı hesaplamasıdır (sütunlar düşürülmeden önceki hesaplama değerleri Şekil 4-11'dedir.).

8. P-değeri hesaplaması yapıldı.

- Bulunan tahmini katsayı değerlerinin anlamlılığının kontrolü *statsmodel* paketi *pvalues* fonksiyonuyla yapıldı.
- Sıfır hipotez (H_0) ve alternatif hipotez (H_1) üretildi.
 - H_0 : Bağımlı değişken ve bağımsız değişkenler arasında ilişki yok.
 - H_1 : Bağımlı değişken ve bağımsız değişkenler arasında ilişki var.
- P-değeri 0,05'ten büyükse H_0 kabul edildi. Bağımlı değişken ve bağımsız değişken arasında ilişki yoktur, katsayı değeri sıfıra (0) eşittir, şeklinde yorumlandı.
- P-değeri 0,05'ten küçükse H_1 kabul edildi. Bağımlı değişken ve bağımsız değişken arasında ilişki vardır, katsayı değeri sıfıra (0) eşit değildir, şeklinde yorumlandı.
- Bu aşama kontrol amaçlıdır. İşlem bir kere gerçekleştirildikten sonra, bulunan p-değerleri değerlendirilip, değerin 0,05'ten çok yüksek olduğu sütunların varlığı saptandı ve bu sütunların çıkarılmasına karar verildi (Şekil 4-12). Bu aşamadan, 4.2.1. *Veri Yükleme ve Ön-İşleme Uygulamaları* başlığının 12. maddesi olan *Anlamlı olmayan bağımsız değişkenlerin kaldırılması* bölümüne gidildi. Kararlaştırılan sütunlar çıkartılarak model revize edildi. Kontrol bu aşamaya gelindiğinde yeniden yapıldı.
- P-değerinin hesaplama sonuçları Şekil 4-20'de sunulmaktadır.

est.pvalues	
const	9.564364e-30
cinsiyet	5.761626e-02
yaş	1.661765e-05
miyopati	1.832005e-04
dağınık_kırmızı_lifler	1.528744e-12
ensefalopati	7.668686e-01
epilepsi	1.793018e-01
laktik_asidoz	2.609610e-16
inme_benzeri_epizodlar	9.931751e-01
sensörinöral_işitme_kaybı	3.446820e-01
diyabet	3.639386e-01
görme_bozukluğu	2.429173e-02
maküler_dejenerasyon	7.194677e-01
cpeo	1.579018e-02
pitozis	7.598676e-01
kardiyomiyopati	3.705582e-01
nefropati	3.770328e-01
nöropati	3.572021e-01
fokal_kortikal_defektler	2.777190e-01
jeneralize_kortikal_defektler	2.173300e-03
serebellar_defektler	1.896454e-01
hareket_bozuklukları	3.241405e-01
bazal_ganglion_kalsifikasyonu	4.414859e-01
kognitif_hastalıklar	6.708580e-03
baş_ ağrısı	8.080277e-01
gastrointestinal_disfonksiyonlar	5.502170e-02
gelişme_geriliği	5.262931e-01
dtype: float64	

Şekil 4-20: P-değeri hesaplama sonuçları (Model-1) (sütun düşürme işleminden sonra)

Model-1 bağımsız değişkenleri ile bağımlı değişken arasındaki ilişkinin sorgulanması için hesaplanan p-değeri 0,05'ten büyükse *ilişki yoktur*, 0,05'ten küçükse *ilişki vardır* şeklinde yorumlanır. (Sütun düşürülme işleminden önceki hesaplama değerleri Şekil 4-12'dedir.)

Güven aralığı değerlendirme sonucunda ulaşılan, birçok bağımsız değişkenin bağımlı değişkeni açıklanmakta yetersiz olduğu sonucu, hesaplanan p-değeri sonuçlarında da açık şekilde görülmektedir (Şekil 4-20). P-değeri 0,05'ten yüksek olan bağımsız değişkenlerin bağımlı değişkenle doğrusal bir ilişkisi, örnek veriye göre bulunmamaktadır. Ancak bulunan p-değeri sonuçlarına göre, bağımlı değişkenle arasında ilişki bulunamamış bağımsız değişkenler, m.3243A>G patojenik varyantı ile ilişkili olan belirgin klinik özelliklerden ise, bu sütunların kalmasına karar verildi.

9. Regresyon modelinin özet tablosu oluşturuldu.

- Model-1'in özet tablosunu oluşturmak için *statsmodel.api* paketinin *summary* fonksiyonu kullanıldı. Değerlendirme *OLS* fonksiyonuna göre yapıldı (Şekil 4-21).

OLS Regression Results						
Dep. Variable:	iskelet_kası	R-squared:	0.507			
Model:	OLS	Adj. R-squared:	0.474			
Method:	Least Squares	F-statistic:	15.57			
Date:	Thu, 06 Aug 2020	Prob (F-statistic):	1.09e-45			
Time:	13:49:42	Log-Likelihood:	-1743.5			
No. Observations:	421	AIC:	3541.			
Df Residuals:	394	BIC:	3650.			
Df Model:	26					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	39.6917	3.215	12.346	0.000	33.371	46.013
cinsiyet	3.1844	1.669	1.908	0.057	-0.096	6.465
yaş	-0.2531	0.058	-4.366	0.000	-0.367	-0.139
miyopati	6.8651	1.813	3.788	0.000	3.302	10.429
dağınık kırmızı_lifler	15.9766	2.183	7.317	0.000	11.684	20.269
ensefalopati	0.7412	2.500	0.296	0.767	-4.175	5.657
epilepsi	3.1838	2.361	1.348	0.178	-1.459	7.826
laktik_asidoz	18.3763	2.143	8.573	0.000	14.162	22.590
inme_benzeri_epizodlar	0.0232	2.216	0.010	0.992	-4.334	4.380
sensörinöral_işitme_kaybı	1.7232	1.820	0.947	0.344	-1.855	5.301
diyabet	1.8456	2.027	0.911	0.363	-2.139	5.831
görme_bozukluğu	-7.4830	3.304	-2.265	0.024	-13.978	-0.988
maküler_dejenerasyon	1.0918	3.022	0.361	0.718	-4.849	7.033
cpeo	-7.6373	3.146	-2.428	0.016	-13.822	-1.452
pitozis	1.0209	3.332	0.306	0.759	-5.530	7.572
kardiyomiyopati	2.1586	2.401	0.899	0.369	-2.563	6.880
nefropati	-3.4808	3.931	-0.886	0.376	-11.208	4.247
nöropati	2.4961	2.702	0.924	0.356	-2.816	7.808
fokal_kortikal_defektler	3.0456	2.773	1.098	0.273	-2.406	8.497
jeneralize_kortikal_defektler	-11.5823	3.738	-3.099	0.002	-18.930	-4.234
serebellar_defektler	6.7322	5.065	1.329	0.185	-3.225	16.689
hareket_bozuklukları	-2.2385	2.265	-0.988	0.324	-6.692	2.215
bazal_ganglion_kalsifikasyonu	2.3093	2.989	0.773	0.440	-3.567	8.186
kognitif_hastalıklar	6.1159	2.233	2.738	0.006	1.725	10.507
baş_ ağrısı	0.5917	2.427	0.244	0.808	-4.180	5.364
gastrointestinal_disfonksiyonlar	-8.9286	4.634	-1.927	0.055	-18.040	0.182
gelişme_geriliği	-1.1991	1.890	-0.634	0.526	-4.916	2.517
Omnibus:	41.006	Durbin-Watson:	1.553			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	57.294			
Skew:	-0.696	Prob(JB):	3.62e-13			
Kurtosis:	4.154	Cond. No.	294.			

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Şekil 4-21: OLS fonksiyonu ile oluşturulan Model-1'in özet tablosu

- OLS değerlendirme metodunun ön şartlarından birinin (otokorelasyon) sağlanmadığı görüldükten sonra (Tablo 4-37), değerlendirmenin doğru yapılabilmesi için, heteroskedastisite ve otokorelasyon durumlarında kullanılan GLSAR değerlendirme metodu kullanıldı. Tüm hesaplamalar, *GLSAR* fonksiyonuyla oluşturulan modelle tekrar yapıldı (Şekil 4-22).

GLSAR Regression Results						
Dep. Variable:	iskelet_kası	R-squared:	0.507			
Model:	GLSAR	Adj. R-squared:	0.474			
Method:	Least Squares	F-statistic:	15.51			
Date:	Mon, 13 Jul 2020	Prob (F-statistic):	1.65e-45			
Time:	22:44:53	Log-Likelihood:	-1739.9			
No. Observations:	420	AIC:	3534.			
Df Residuals:	393	BIC:	3643.			
Df Model:	26					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	39.6939	3.220	12.327	0.000	33.363	46.024
cinsiyet	3.1830	1.672	1.904	0.058	-0.103	6.469
yaş	-0.2531	0.058	-4.360	0.000	-0.367	-0.139
miyopati	6.8624	1.817	3.777	0.000	3.290	10.434
dağınık_kırmızı_lifler	15.9768	2.186	7.308	0.000	11.679	20.275
ensefalopati	0.7430	2.504	0.297	0.767	-4.180	5.666
epilepsi	3.1890	2.370	1.345	0.179	-1.471	7.849
laktik_asidoz	18.3786	2.147	8.558	0.000	14.157	22.601
inme_benzeri_epizodlar	0.0190	2.223	0.009	0.993	-4.351	4.389
sensörinöral_işitme_kaybı	1.7250	1.823	0.946	0.345	-1.860	5.310
diyabet	1.8449	2.030	0.909	0.364	-2.146	5.835
görme_bozukluğu	-7.4877	3.311	-2.261	0.024	-13.998	-0.978
maküler_dejenerasyon	1.0883	3.028	0.359	0.719	-4.865	7.041
cpeo	-7.6366	3.150	-2.424	0.016	-13.830	-1.443
pitozis	1.0205	3.336	0.306	0.760	-5.539	7.580
kardiyomiyopati	2.1564	2.405	0.896	0.371	-2.573	6.886
nefropati	-3.4806	3.936	-0.884	0.377	-11.218	4.257
nöropati	2.4943	2.706	0.922	0.357	-2.826	7.814
fokal_kortikal_defektler	3.0362	2.793	1.087	0.278	-2.455	8.528
jeneralize_kortikal_defektler	-11.5745	3.751	-3.086	0.002	-18.949	-4.200
serebellar_defektler	6.7587	5.144	1.314	0.190	-3.354	16.872
hareket_bozuklukları	-2.2422	2.271	-0.987	0.324	-6.708	2.223
bazal_ganglion_kalsifikasyonu	2.3191	3.010	0.770	0.441	-3.599	8.237
kognitif_hastalıklar	6.1224	2.246	2.725	0.007	1.706	10.539
baş_ ağrısı	0.5909	2.430	0.243	0.808	-4.187	5.369
gastrointestinal_disfonksiyonlar	-8.9308	4.641	-1.924	0.055	-18.054	0.193
gelişme_geriliği	-1.2024	1.896	-0.634	0.526	-4.929	2.525
Omnibus:	40.749	Durbin-Watson:	1.552			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	56.721			
Skew:	-0.695	Prob(JB):	4.82e-13			
Kurtosis:	4.144	Cond. No.	295.			

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Şekil 4-22: GLSAR fonksiyonu ile oluşturulan Model-1'in özet tablosu

Model değerlendirme, veri setinin güncellenmesi, GLSAR ile en doğru regresyon doğrusunu belirleme işlemleri sonrasında oluşturulan regresyon modeli için özet tablosu oluşturuldu (Şekil 4-22). OLS regresyon doğrusu ile elde edilmiş özet tablo yalnızca karşılaştırma amaçlı eklendi (Şekil 4-21). GLSAR regresyon sonuç tablosu değerlendirmede esas alındı.

R-kare (*R-Squared*) ve ayarlanmış R-kare (*adjusted R-squared*) değerleri, örnek veri setindeki bağımsız değişkenlerin, bağımlı değişkendeki varyansların ne kadarını açıklayabildiğinin yani veri setinin modele uyumunun bilgisini verir. Değerler 1'e ne kadar yakınsa, modelin o kadar başarılı olduğu, örnek veri ile oluşturulan modelin bağımlı değişken tahminlerini o kadar iyi yaptığı sonucuna ulaşılır [158]. R-kare sonucu, oluşturulan regresyon modelinin, iskelet kası heteroplazmi oranındaki değişkenliklerin %51'ini açıklayabildiğini gösterdi. Ancak R-kare değeri sütun sayısı arttıkça artan bir değerdir. Anlamı olmayan sütunların eklenmesi dahi bu değeri artırır. Ayarlanmış R-kare, yüksek bağımsız değişken sayısının cezai değerinin, R-kare değerinden düşürüldüğü değerdir. Bu yüzden bağımsız değişken sayısının yüksek olduğu veri setlerinden oluşturulan modelin başarısı ayarlanmış R-kare değeriyle ölçülür. Cezai değerin R-kare değerinden düşürülmesiyle, modelin başarı oranı %47 bulundu (Şekil 4-22).

Tablonun alt bölümünde tüm bağımsız değişkenler için katsayı yani eğim (*coefficient (coef) = slope*), standart hata (*standard error (std err)*), t, p ve güven aralığı değerleri verilmiştir. Katsayı değerleri her bir değişkenin diğer tüm bağımsız değişkenlerin etkisi sıfırken bağımlı değişkeni bir birimlik değişimde etkileme değeridir. Standart hata, MSE değerinin kare köküdür, yani RMSE değeridir. Bağımlı değişken birimiyle ifade edilir. Gözlemlenen veri noktalarının regresyon çizgisi ile arasındaki ortalama mesafedir. Bağımlı değişkenin etkilenme derecesi, katsayı değeridir. Bulunan sonucun gerçek değerden sapma değerini veren ise standart hata değeridir. Örneğin hastada miyopati varsa, bu durum kas heteroplazmi oranını 1,8 sapma değeriyle 6,9 değerinde arttırmaktadır. Matematiksel ifade ile $6,9 \pm 1,8$ 'dir. Eksi katsayı değerine sahip olanların ise, heteroplazmi oranını azaltan yönde etki ettiği kurulan ilişki sonucu ortaya çıkmıştır. CPEO klinik durumu mevcutsa, kas heteroplazmi oranı 3,1 sapma ile 7,6 değerinde azalmaktadır. Tablo sonuçlarında, *iskelet_kası* değerini en çok arttıran bağımsız değişkenin, $18,4 \pm 2,1$ katsayı değeriyle *laktik_asidoz* olduğu görülmektedir.

y değerini negatif yönde en çok etkileyen bağımsız değişkense, $11,6 \pm 3,7$ katsayı değeriyle *jeneralize_kortikal_defektler* olmuştur (Şekil 4-17). Her bir değişken için verilen standart hata yani RMSE değeri, modelin ne kadar iyi tahmin yapabildiğinin ölçüsüdür [158].

Bağımlı değişken ve bağımsız değişkenler arasında istatistiksel olarak anlamlı, doğrusal bir ilişki olup olmadığının bilgisini F-istatistik (*F-statistic*) ve t-testi sonuçları verir. F-testi için 0,05 varsayılan sınırdan yüksek değer veren sonuçlar anlamlıdır. Modelin verdiği 15,51 F-istatistik değeri, 0,05'ten yüksektir ve modelin anlamlılığını belirtir. T-testinde, her bir bağımsız değişken ve bağımlı değişken arasındaki ilişki sorgulanır. Her bir bağımsız değişkenin eğim değerinin sıfıra eşit olup olmadığının sorgusu yapılır. T-testi sonucu, eğim değerinin, eğimin sapma değerine bölünmesiyle bulunur. Sıfır hipotezde (H_0), her bir bağımsız değişken için, bağımsız değişkenin regresyon doğrusunun eğimi sıfırdır ve bağımlı değişkenle arasında anlamlı doğrusal bir ilişki yoktur, savunması yapılır ve dışlanması için test sonucunun t-kritik değerinden yüksek olması gerekir. T-test değeri, regresyonun serbestlik derecesi ve t-tablosundan elde edilecek olan t-kritik değeri ile p-değeri belirlenir. P-değeri 0,05'ten küçükse sıfır hipotez dışlanmış olur ve bağımsız değişken ile bağımlı değişken arasında doğrusal bir ilişki olduğu, bağımsız değişken eğim değerinin sıfıra eşit olmadığı sonucuna ulaşılır. Eğim için bulunan güven aralığında sıfır değerinin bulunup bulunmadığı ifade edilmiş olur. Bağımsız değişkenlerin anlamlılığı bu noktada değerlendirilebilir [161].

Özet tabloda belirtilen p-değerlerinin bir çoğu 0,05'ten yüksektir. Bağımlı değişkenle arasında bir ilişki kurulamamış olan bağımsız değişkenler, m.3243A>G varyantı saptanan olgularda gözlenen belirgin klinik özellikler veya bulgu ve belirtilerdir. P-değeri yüksek olan bu sütunların çıkartılması çalışmanın amacını ortadan kaldırmak olacağından bu yönde bir işlem yapılmadı. Gerçek hayatta ilişkili olduğu bilinen tüm özellikler için değişkenlik çok fazlaydı ve bu değişkenlikleri açıklayabilmek için 421 hasta sayısının yeterli gelmediği düşünüldü. Ayarlanmış R-kare değerine göre, var olan örnek veri ile kas heteroplazmi oranındaki değişkenliklerin %47 oranında açıklanabildiği görüldü (Şekil 4-17).

4.2.3.2. Diğer Model Denemeleri ve Sonuçları

df veri setindeki veri sayısını arttırabilmek ve bu veri setiyle bir model (Model-2) oluşturabilmek amacıyla, *kas_df* veri setiyle oluşturulan Model-1 tahmin modeli kullanılarak, *nan_kas_df* veri setinde tamamı boş olan *iskelet_kası* sütununun doldurulup, *nan_kas_df* veri setinin de Model-2 tahmin modeli oluşturma işlemine dahil edilmesi istendi. Bu veri setindeki boş iskelet kası heteroplazmi oranı bilgilerini doğru değerlere en yakın değerlerle doldurabilmek için, özellikler arasındaki ilişkiyi en doğru şekilde kuran, doğruluk oranı daha yüksek bir model araştırması için *kas_df* veri seti kullanılarak iki model denemesi daha yapıldı. Bu denemelerle ayrıca *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarında yapılan doldurma işleminin gerekliliğinin ortaya konulması hedeflendi. Ayrıca kas yerine kan verisi üzerinden gidilerek kan heteroplazmi oranının veri setindeki bağımsız değişkenlerle uyumu araştırıldı ve oluşturulan modelin tahmin gücü ölçüldü. Tüm bu modellerin optimizasyonu *GLSAR* fonksiyonu ile yapıldı. Sonuçları Tablo 4-39’da özetlenmektedir.

Model-3:

kas_df veri setinde, *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarında boşluk doldurma işlemi yapılmadı ancak diğer müdahalelerin tamamı yapıldı. Bu sütunların her ikisinde birden değer bulunan satırlar filtrelendi. Oluşturulan yeni veri setiyle (veri seti boyutu: 258, 27 (satır, sütun)) regresyon modeli oluşturuldu ve sonuçları değerlendirildi.

Model-1’den tek farkı, satır sayısının azalması oldu. Ayarlanmış R-kare değeri 0,38 olarak bulundu. RMSE değerinin artmış olduğu, hata oranı yükseldiği, F-istatistik değerinin yani modelin anlamlılığının düştüğü görüldü. Bu sonuç, iki sütunda yapılan doldurma işlemiyle satır sayısının arttırılmasının, Model-1’in ilişki kurma yeterliliğini arttırdığını göstermektedir. Ayrıca, Model-1’in satır sayısının azlığından dolayı ilişki kurmada yetersiz kalmış olma olasılığını doğrular nitelikte olmuştur.

Model-4:

kas_df veri setinin son halinden, *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunları *drop* fonksiyonu ile düşürülerek (veri seti boyutu: 421, 25 (satır, sütun)) model oluşturuldu ve değerlendirildi.

Ayarlanmış R-kare değerinin büyük bir düşüş gösterdiği görüldü. RMSE değeriyle, hata oranlarının Model-1’e göre yüksek olduğu görüldü. Modelin anlamlılığı

da (F-istatistik) düşük bulundu. Bu sonuçla, *laktik_asidoz* ve *dağınık_kırmızı_lifler* sütunlarının bağımlı değişkeni açıklamakta etkili olduğu görüldü ve eksik verilerinden dolayı bu iki sütunun çıkartılması durumunun doğru bir hamle olmayacağı görüldü.

Model-5:

Kan heteroplazmi oranlarının bağımsız değişkenlerle ilişkisi araştırıldı. *kan* sütununda, boş değer içermeyen satırlar filtrelendi ve ayrı bir değişkene atandı. Bu veri setiyle (veri seti boyutu: 430, 27 (satır, sütun)) bir model oluşturuldu ve sonuçları değerlendirildi.

laktik_asidoz ve *dağınık_kırmızı_lifler* sütunları için boşluk doldurma işlemi, 4.2.1. *Veri Yükleme ve Ön-İşleme Uygulamaları* başlığının 10. maddesine göre yapıldı.

Ayarlanmış R-kare değeri 0,53 olan bu model, kan heteroplazmi oranı ile bağımsız değişkenler arasındaki uyumun, kas oranı ile yapılan ilişkilendirme çalışmasından daha başarılı olduğunu gösterdi. Hata oranının ise 1 puan daha düşük olduğu görüldü. F-istatistik değeri de sonuç olarak daha yüksek çıkmış oldu. Kan heteroplazmi oranının tahmin edildiği Model-5, iskelet kası heteroplazmi oranının tahmin edildiği Model-1'e göre %6 oranında daha iyi sonuç vermiştir. Ancak kan heteroplazmi oranlarının tahmininde nispeten daha başarılı olursa da, kan oranlarının kasa göre daha düşük oranlar verdiği bilinmektedir [154] ve bu tez çalışmasının birinci basamağındaki laboratuvar sonuçlarıyla da benzer sonuçlara ulaşılmıştır. Gerçek uygulamada invazif bir işlem olmasına karşın, kas heteroplazmi oranı hastalığın şiddeti hakkında daha doğru bilgi verici olduğu için, kas heteroplazmi oranı tahmini üzerinden gidilmesine ve *nan_kas_df* veri setindeki tahminlerin kas heteroplazmi oranı üzerinden Model-1 ile yapılmasına karar verildi. *nan_kas_df* veri setinin iskelet kası heteroplazmi oranının tahmini yapıldı. Böylece hasta sayısı *df* veri setinde yaklaşık iki katına çıkartılmış oldu.

Table 4-39: Diğer model denemelerinin sonuçları

Model	3	4	5
y-kesen değeri	50,12	57,15	27,09
Heteroskedastisite	YOK	YOK	VAR
Otokorelasyon	VAR	VAR	VAR
Artıkların Ortalaması	0,009004	0,01472	0,004598
MSE	419	291	250
MAE	15,2	13,9	11,3
RMSE	20,5	17,1	15,8
R-Kare	0,444	0,303	0,558
Ayarlanmış R-Kare	0,381	0,261	0,53
F-İstatistik	7,066	7,165	19,53

4.2.3.3. Modelin Kaydedilmesi

Değerlendirme sonucunda, 0,474 ayarlanmış R-kare ve 15,51 F-istatistik değerine sahip regresyon modeli, Model-1 seçildi (Şekil 4-22).

- *Pickle* paketi *pickle.dump* fonksiyonu kullanılarak *with open* fonksiyonu ile açılan .sav uzantılı dosyaya regresyon modeli kaydedildi.
- .sav uzantılı regresyon modeli dosyası *regresyon_modeli_tahmin* isimli bir değişkene atandı. *predict* fonksiyonu, *X_Test* satırlarına uygulandı. Modelin tahmin ettiği değerler ve gerçek değerler karşılaştırıldı.
- Kaydedilen model, bir sonraki aşamada *nan_kas_df* veri setinin kayıp iskelet kası heteroplazmi oranlarının tahmini için kullanılarak, tüm veri setinin veri eksiği olmadan kullanılabilir hale getirilmesini sağladı.

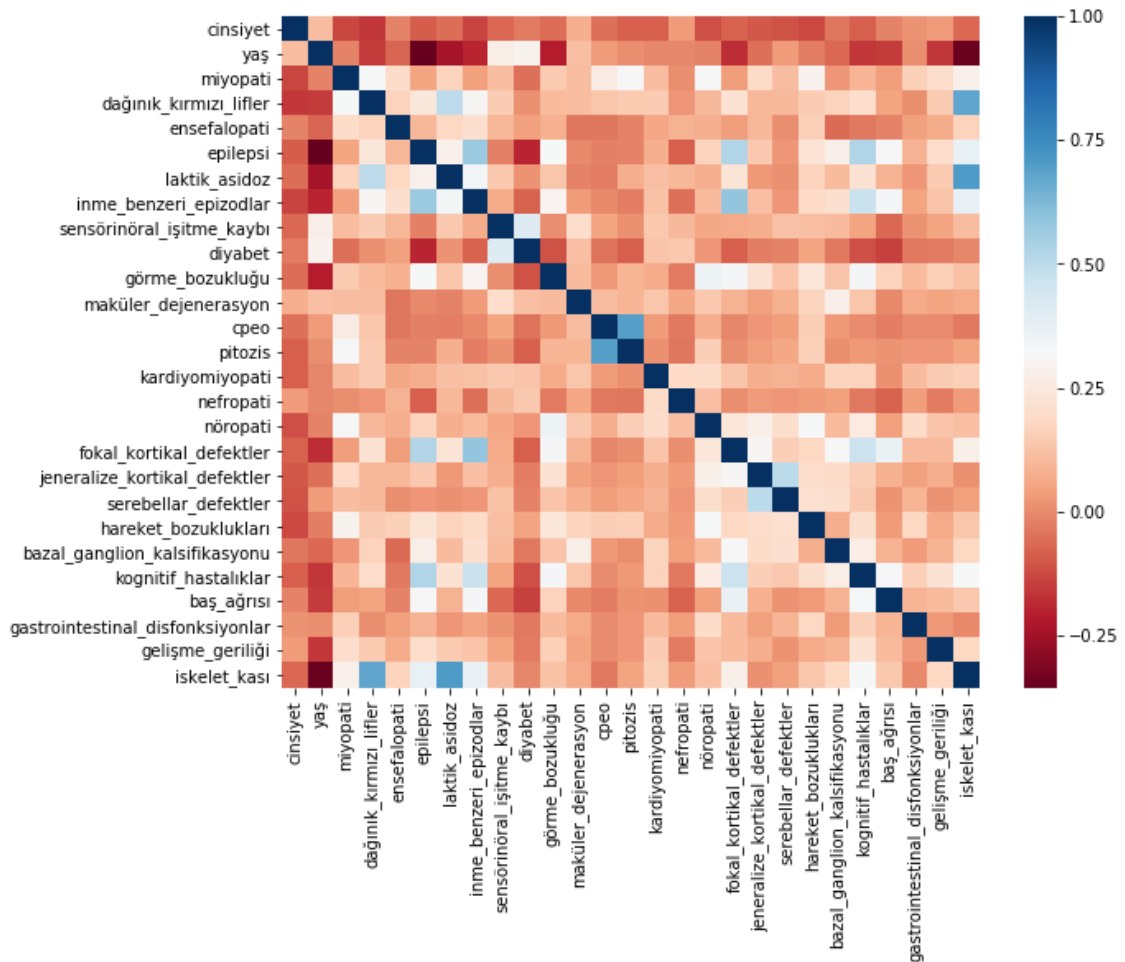
4.2.4. Veri Setinin Hazırlanması

- Kaydedilen regresyon modeli tahmin aracı *nan_kas_df* veri setine uygulandı.
- *kas_df* ve *nan_kas_df* değişkenleri ayrı ayrı *Pandas* kütüphanesi *to_csv* fonksiyonuyla kaydedildi.
- Tahmin aracı sonuçları düzenlenerek, oluşturulan *nan_kas_df* isimli .csv uzantılı dosyada yeniden oluşturulan *iskelet_kası* sütununa yerleştirildi ve kaydedildi.
- *nan_kas_df* dosyası aynı değişken ismiyle, *Pandas* kütüphanesi *read_csv* fonksiyonu kullanılarak yüklendi.
- *hasta_no*, *nan_kas_df* değişkenine *set_index* fonksiyonu ile indeks olarak atandı.
- *kas_df* ve *nan_kas_df* dosyalarının satır ve sütun sayılarının, sütun isimlerinin uyuşup uyuşmadığı kontrol edildi.
- Sütun isimlerinin eşleştiği doğrulanan iki dosya, *Pandas* kütüphanesi *concat* fonksiyonu kullanılarak birleştirildi ve *df* isimli veri seti tipindeki değişkene atandı.
- Elde edilen *df* veri setinin satır ve sütun sayıları kontrol edildi.
 - *df* veri seti boyutu: 744, 27 (satır sayısı, sütun sayısı)
- Çoklu eşdoğrusallık (çoklu bağlantı) testi yapıldı.
 - *Pandas* kütüphanesi *corr* fonksiyonuyla korelasyon matrisi oluşturuldu.
 - Korelasyon matrisi *Seaborn* kütüphanesi *heatmap* fonksiyonu kullanılarak görselleştirildi.
 - Çoklu eşdoğrusallık ölçeği olan varyans büyütme faktörü (**VIF** = **Variance Inflation Factor**) kullanıldı.
 - Çoklu eşdoğrusallığı dışlamak için, $VIF < 5$ koşulu uygulandı.
- Uç değer kontrolü tüm değişkenler üzerinde uygulandı.
 - *Pandas* kütüphanesi *describe* fonksiyonu kullanıldı. Standart sapma değerleri 3 ile çarpılarak ortalamadan çıkartıldı ve ortalamaya eklendi.

Böylece, taban ve tavan değerleri bulundu. Taban değerinin altında kalan ve tavan değerinin üstünde kalan değerler, uç değer olarak kabul edildi.

- Uç değer varlığı kontrol edildi.

df veri setinin korelasyon matrisi ısı haritası Şekil 4-23'te sunulmaktadır.



Şekil 4-23: *df* veri setinin korelasyon matrisi görüntüleme sonucu

Sağ kenardaki renk geçiş çubuğunda: Çubuğun en üstü (lacivert renk), korelasyonun %100 olduğunu, çubuğun en altı (bordo renk) korelasyonun %0 olduğunu, çubuğun ortası (beyaz renk) ise ortalama bir ilişkiyi belirtmektedir. Matriste çapraz geçen lacivert renkler, aynı sütunun x ve y düzlemindeki kesişim noktalarıdır. İskelet kası sütunu y ekseninin en altında yer almaktadır ve x eksenini boyunca diğer sütunlarla olan ilişkisi renklerle anlaşılmaktadır.

VIF < 5 koşulu sağlanıyorsa çoklu eşdoğrusallık dışlandı. Çoklu eşdoğrusallık sonuçları Şekil 4-24'te sunulmaktadır.

const	24.676290
cinsiyet	1.146654
yaş	1.574991
miyopati	1.606121
dağınık_kırmızı_lifler	2.156554
ensefalopati	1.208142
epilepsi	2.124978
laktik_asidoz	2.177426
inme_benzeri_epizodlar	2.193222
sensörinöral_işitme_kaybı	1.406513
diyabet	1.383214
görme_bozukluğu	1.397069
maküler_dejenerasyon	1.238401
cpeo	2.103811
pitozis	2.119590
kardiyomiyopati	1.174062
nefropati	1.121361
nöropati	1.482055
fokal_kortikal_defektler	2.069617
jeneralize_kortikal_defektler	1.611618
serebellar_defektler	1.476280
hareket_bozuklukları	1.368674
bazal_ganglion_kalsifikasyonu	1.321538
kognitif_hastalıklar	1.809977
baş_ ağrısı	1.275822
gastrointestinal_disfonksiyonlar	1.121555
gelişme_geriliği	1.160288
iskelet_kası	3.649129
dtype: float64	

Şekil 4-24: *df* veri setinin çoklu eşdoğrusallık testi sonuçları

4.2.4.1. Regresyon Modelinin Uygulanması

- *iskelet_kası* sütunu *df* veri setinden *drop* fonksiyonu kullanılarak düşürüldü ve *X* değişkenine atandı.
- *df* veri setindeki *iskelet_kası* sütunu seçildi ve *Y* değişkenine atandı.
- *X* ve *Y* değişkenlerinin eğitim ve test verilerine ayrımı *sklearn.model_selection.train_test_split* fonksiyonuyla yapıldı.

- `sklearn.linear_model.LinearRegression` fonksiyonu `regresyon_modeli` isimli değişkene atandı.
- X ve Y eğitim verileri, `fit` fonksiyonu aracılığıyla `LinearRegression` fonksiyonuna uygulandı. Böylece Model-2 oluşturulmuş oldu.
- Kesen değer ve katsayı kontrolü yapıldı.
 - `iskelet_kası` bağımlı değişkenin y -kesen değeri ($\beta_0 = y\text{-intercept}$): **39,53**

Kesen değer ($y\text{-intercept}$) sonucu, patojenik varyanta sahip olup klinik bir bulgu vermeyen bir hastanın heteroplazmi oranının 39,53 olduğunu göstermektedir.

- Oluşturulan model, `predict` fonksiyonu kullanılarak, X_{Test} verisine uygulandı.
- Tahmin ve gerçek değerlerin değerlendirilmesi `GLSAR` fonksiyonuyla yapıldı.
- Model-1 uygulamasında olduğu gibi değerlendirme testleri uygulandı.
 - Heteroskedastisite (değişen varyans sorunu) testi
 - Test sonucunda heteroskedastisite durumunun olduğu görüldü ($P\text{-değeri} < 0,05$) (Tablo 4-40).

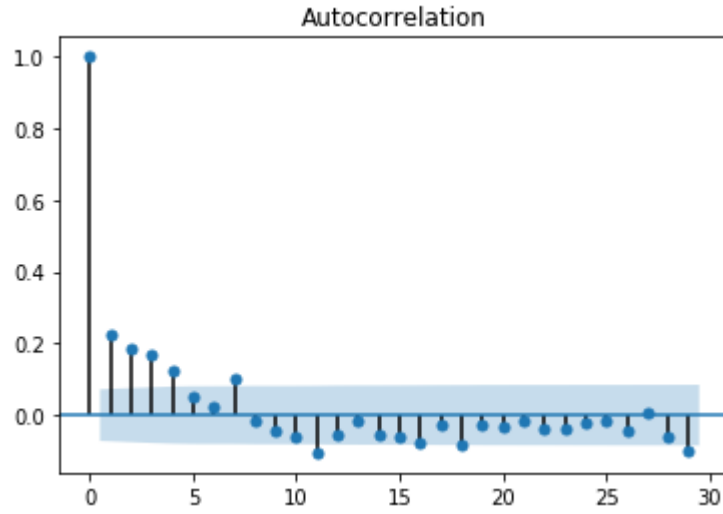
Table 4-40: Model-2 için heteroskedastisite testleri ve sonuçları

TEST	P-DEĞERİ	SONUÇ
White Testi	0,0001065	HETEROSKEDASTİSİTE VAR
Breusch-Pagan Testi	0,03853	HETEROSKEDASTİSİTE VAR

- Otokorelasyon testi
 - Test sonucuna göre otokorelasyonun var olduğu görüldü ($P\text{-değeri} < 0,05$) (Tablo 4-41, Şekil 4-25).

Table 4-41: Model-2'nin otokorelasyon testi ve sonucu

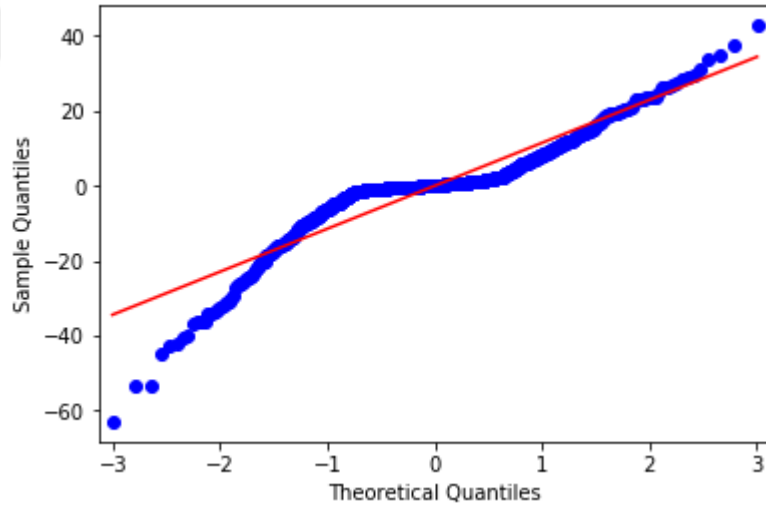
TEST	GECİKME	EN KÜÇÜK P-DEĞERİ	SONUÇ
Ljung-Box testi	10	7,51e-20	OTOKORELASYON VAR



Şekil 4-25: Model-2'nin otokorelasyon testi görsel sonucu

x eksen: gecikmeler, y eksen: korelasyonlar

- Artıkların normal dağılım kontrolü yapıldı (Şekil 4-26).



Şekil 4-26: Artık değerlerin regresyon çizgisi etrafındaki dağılımının görsel sonucu (Model-2)

Kırmızı çizgi: regresyon doğrusu, mavi noktalar: veri noktaları

- Artık ortalamasının sıfıra eşit olup olmadığına bakıldı.
 - Artık Ortalama Değeri: **-0,0006762**
- Verilerin regresyon modeline uyumu hata değerleriyle değerlendirildi. Sonuçlar Tablo 4-42'de sunulmaktadır.

Table 4-42: Model-2 için hata değerleri hesaplama sonuçları

MAE	6,99
MSE	1,34e+02
RMSE	11,6

MAE = Ortalama mutlak hata (*Mean Absolute Error*), MSE = Ortalama kare hata (*Mean Squared Error*), RMSE (Kök ortalama kare hata (*Root Mean Squared Error*))

Hata hesaplamaları sonucu veri noktaları ve regresyon çizgisi arasındaki mesafenin sıfıra, Model-1'e kıyasla, yaklaştığını gösterdi. MAE değeri, tahmin ve gerçek değerler arasında ortalama %7 fark olacağını söylemektedir. Büyük hata değerlerine ağırlık veren RMSE değeri, veri noktaları ve regresyon çizgisi arasındaki mesafenin ortalama değerinin %11,6 olduğunu göstermektedir. Değerdeki yüzde işareti tahmin edilen bağımlı değişkenin birimidir. Bu değer, bağımlı değişkenin yani iskelet kası heteroplazmi oranının tahmininde ortalama %11,6 sapma olacağını belirtmektedir (Tablo 4-42).

- GLSAR fonksiyonu ile optimize edilen Model-2'nin özet tablosu oluşturuldu (Şekil 4-27).

GLSAR Regression Results						
Dep. Variable:	iskelet_kası	R-squared:	0.726			
Model:	GLSAR	Adj. R-squared:	0.716			
Method:	Least Squares	F-statistic:	72.88			
Date:	Tue, 14 Jul 2020	Prob (F-statistic):	1.74e-181			
Time:	00:24:33	Log-Likelihood:	-2868.4			
No. Observations:	743	AIC:	5791.			
Df Residuals:	716	BIC:	5915.			
Df Model:	26					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	38.8143	1.563	24.839	0.000	35.746	41.882
cinsiyet	3.0162	0.926	3.258	0.001	1.199	4.834
yaş	-0.2540	0.031	-8.300	0.000	-0.314	-0.194
miyopati	6.4389	1.081	5.954	0.000	4.316	8.562
dağınık_kırmızı_lifler	17.5302	1.120	15.654	0.000	15.332	19.729
ensefalopati	0.2777	1.720	0.161	0.872	-3.098	3.654
epilepsi	3.2505	1.433	2.268	0.024	0.436	6.065
laktik_asidoz	19.1061	1.168	16.354	0.000	16.812	21.400
inme_benzeri_epizodlar	-0.0964	1.444	-0.067	0.947	-2.932	2.739
sensörinöral_işitme_kaybı	1.1577	1.018	1.138	0.256	-0.840	3.156
diyabet	2.3572	1.093	2.157	0.031	0.212	4.503
görme_bozukluğu	-6.5771	1.908	-3.448	0.001	-10.323	-2.832
maküler_dejenerasyon	1.6091	1.668	0.965	0.335	-1.666	4.884
cpeo	-7.6271	1.967	-3.877	0.000	-11.490	-3.765
pitozis	0.7376	1.964	0.376	0.707	-3.118	4.594
kardiyomiyopati	1.6050	1.447	1.109	0.268	-1.235	4.445
nefropati	-2.9985	1.903	-1.575	0.116	-6.735	0.738
nöropati	2.4150	1.714	1.409	0.159	-0.949	5.779
fokal_kortikal_defektler	2.0143	1.663	1.212	0.226	-1.250	5.279
jeneralize_kortikal_defektler	-10.5395	2.336	-4.512	0.000	-15.125	-5.954
serebellar_defektler	5.3783	3.106	1.732	0.084	-0.719	11.475
hareket_bozuklukları	-2.3795	1.490	-1.597	0.111	-5.306	0.547
bazal_ganglion_kalsifikasyonu	1.8052	1.998	0.904	0.367	-2.117	5.728
kognitif_hastalıklar	6.0973	1.390	4.387	0.000	3.369	8.826
baş_ ağrısı	0.6789	1.254	0.541	0.588	-1.784	3.141
gastrointestinal_disfonksiyonlar	-6.7873	2.417	-2.808	0.005	-11.533	-2.042
gelişme_geriliği	-1.3816	1.098	-1.258	0.209	-3.538	0.775
Omnibus:	162.608	Durbin-Watson:	1.552			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	736.990			
Skew:	-0.926	Prob(JB):	9.22e-161			
Kurtosis:	7.514	Cond. No.	311.			

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Şekil 4-27: df veri seti ile oluşturulan çoklu-regresyon modeli, Model-2'nin sonuçları

Ayarlanmış R-kare değerine göre, modelin kas heteroplazmi oranını tahmin etmekteki gücünün 0,72, istatistiksel olarak anlamlılığını belirten F değerinin ise 72,9 olduğu görüldü. Heteroplazmi oranını pozitif yönde en çok etkileyen sütun *laktik_asidoz* sütunu olmuştur. Laktik asidoz varlığının, heteroplazmi oranını $19,10 \pm 1,2$ birim arttırdığı görüldü. Negatif yönde en çok etkileyen sütun ise *jeneralize_kortikal_defektler* olmuştur, heteroplazmi oranında $10,5 \pm 2,3$ birim eksilmeye neden olduğu görülmektedir (Şekil 4-27). Veri sayısının hala, MELAS

hastalığının önemli bulguları olan ensefalopati ve inme benzeri epizodlar durumlarını açıklamakta yeterli olmadığı, bu iki bağımsız değişkenin p-değerleri, %5 anlamlılık seviyesinden daha yüksek değer verdikleri için anlamlı olmadıklarının savunulduğu sıfır hipotezini haklı çıkarmaktadır.

- Regresyon modeli, *pickle.dump* fonksiyonuyla, *with_open* ile açılan *.sav* uzantılı dosyaya kaydedildi ve kullanıma hazır hale getirildi.
- Sonuç olarak, m.3243A>G patojenik varyantına sahip hastaların verilen özellikleri ile (*features*), kas heteroplazmi oranlarının tahminini yapabilen bir model geliştirilmiş oldu.



5. TARTIŞMA

Mitokondri, nükleus yapısı olan tüm hücrelerin endoplazmik retikulumuna bağlı bir organeli olup, primer fonksiyonu hücrenin biyokimyasal enerjisi için gerekli kimyasal enerjiyi ATP molekülü olarak üretmek ve barındırmaktır. Her bir hücrenin enerji ihtiyacı ile içerdiği mitokondri sayısı doğru orantı gösterir. Sayılarının, somatik hücrelerde ~80-2000 arasında olduğu bildirilmiştir [28]. Düzenli egzersiz yapan kişilerde egzersiz türüne bağlı olarak mitokondri miktarında artış gözlenmiştir [162] [163]. Aynı hücre içerisindeki miktar değişimi, bu dinamik organellerin devamlı olarak bölünme veya birleşme göstermesiyle olur. Bu nedenle farklı dokuların hücrelerinde bulunan mitokondrilerin tam sayısını belirlemek olanaklı değildir [164]. Çalışmalarda, hücrelerin içerdiği mitokondrilerin miktarı, mitokondrilerin hücrelerde kapladığı hacmin oranı ile ifade edilmiştir [31][32].

Mitokondri diğer hücre içi organellerden farklı olarak, nükleer genomdan ayrı bir genoma sahiptir. Her bir mitokondri, yine hücrelerin aktiflikleri ile ilişkili olarak 1-15 arasında mtDNA içerir [55].

Mitokondrilerin dinamik yapısı, farklı hücre tiplerindeki miktarının ve içerdiği mtDNA miktarlarının değişkenliği, mitokondriyal genom varyant çalışmalarında homoplazmi ve heteroplazmi kavramlarını doğurmuştur. Heteroplazmik durumlarda mtDNA'daki tekli veya çoklu nükleotid değişimlerinin frekansından bahsedilir. Doku içinde ve dokular arası frekans çeşitliliği söz konusudur [165]. Mitokondriyal hastalığa sahip, m.3243A>G patojenik varyantı saptanmış yetişkin monozigotik ikizlerde yapılan otopsi çalışması, 23 farklı dokudan alınan örneklerde, dokular arası heteroplazmi farklılıklarını ortaya koymuştur. Benzer klinik özelliklere sahip oldukları görülse de, ikizlerin aynı dokularındaki oranların da farklı olduğu görülmüştür [166]. Çeşitli sebeplerle ölmüş 152 bireyin 12 farklı dokusunda YND ile yapılan mtDNA dizileme çalışmasında, heteroplazmi oranlarının dokularda farklı olduğu gösterilmiş ve farklı dokulardaki farklı mtDNA pozisyonlarındaki yüksek heteroplazmi oranlarına dikkat çekilmiştir. MtDNA'da yaşa bağlı varyant birikimlerinin olduğu ve belli dokuların belli pozisyonlarında saptanan varyantların olumlu yönde etki eden pozitif, patojenik etki gösteren negatif veya hiçbir etkisi olmayacak şekilde nötral seçimler olabileceği belirtilmiştir. Karaciğerde görülen ve mitokondri fonksiyonunu azaltacak yönde etki

eden bir deęişimin, pozitif bir seçim olduęu düşünölmüş ve mitokondri metabolizmasının zararlı yan ürünlerini azaltmak için yapılan bir seçim olarak deęerlendirilmiştir [167]. Yine YND ile 100 bireyin 9 farklı dokusunda gerçekleştirilen başka bir çalışmada, farklı dokularda aynı varyantlar için farklı heteroplazmi oranlarının göröldüęü ve en yüksek oranların kas ve karacięer dokularında olduęu belirtilmiştir [148].

Bu tez çalışmasının amacı, veri madencilięi ve makine öğrenmesi yöntemlerini kullanarak, mitokondriyal genomda günümüze kadar tanımlı hastalık ilişkili deęişimlerin klinik bulgularını, homoplazmi ya da heteroplazmi temelinde dijital olarak ortaya çıkarmak ve aynı zamanda, alınan dokulardaki mtDNA varyantlarının heteroplazmik yapılarını karşılaştırmalı olarak inceleyerek, dokuların bilgi vericilięini araştırmaktır. Günümüzde mtDNA’da hastalık ilişkisi bilinen 300’ün üzerinde varyant tanımlanmış olup her biri için çok sayıda makale ve araştırma çalışması yayımlanmıştır. Bu tez çalışmasında, üzerinde en çok yayın ve araştırma yapılmış olan m.3243A>G varyantı seçildi. Veri madencilięi ve makine öğrenmesi, bu varyantın ilişkili bulguları ve heteroplazmik yapıları temelinde uygulandı. Ayrıca heteroplazmi karşılaştırılması, altı farklı gönüllüden alınan atık dokulardan elde edilen nükleik asit izolatlarında mtDNA’nın YND teknolojisi kullanılarak elde edilen verilerin üzerinde gerçekleştirildi.

5.1. Laboratuvar Sonuçlarının Deęerlendirmesi

Çalışmanın birinci basamaęında, gönüllü bireylerin farklı dokularında mtDNA varyant taramaları yapılarak, çalışılan dokularda bulunan toplam varyant miktarı üzerinden, dokuların varyant sayısının, frekans yoğunluęunun (heteroplazmi oranının) ve ortak varyantlarının frekans (heteroplazmi oranı) farklılıęının gösterilmesi hedeflendi. Ortak olmayan varyantların ise hangi dokuda bulunduęu, bu varyantların türü ve hangi dokunun içerdii varyant miktarının fazla olduęu belirlenmek istendi. Bu analiz gönüllü bireylerin dokuları arasındaki varyant yükü (heteroplazmi oranı) ve sayılarının karşılaştırılmasına imkan sağladı. Mitokondriyal hastalıklarda, genotip-fenotip ilişkilendirme çalışmalarında, elde edilen frekans deęerlerinin ve hangi dokunun daha bilgi verici olduęunun gözlenmesinin, tanısal yaklaşımlarda, kanıta dayalı bilgi sağlaması bakımından önemli olacaęı ve literatüre katkı sağlayacaęı düşünöldü.

Dokular arasında gözlemlenen varyant farklılıkları ve aynı varyantların farklı dokulardaki farklı heteroplazmi oranları, hastalık durumlarında tek bir doku üzerinden çalışma yapıldığında patojenik varyantın bakılan dokudaki frekansı düşük olduğu için hiç tespit edilememesi, tespit edilse dahi düşük frekans durumunun klinik durumla uyuşmadığı durumların önüne geçilmesi amacıyla birden fazla doku üzerinde çalışma yapmanın avantajlarını ortaya koyması açısından önemli görüldü. Farklı dokularda aynı varyant üzerinde yapılan çalışmalar bu yorumu destekler niteliktedir [148][168].

Farklılaşma sürecini tamamlamış olgun hücreler, metabolik olarak aktif hücrelerdir ve genellikle hücre bölünme döngüsünün dışında, dinlenme evresinde (G0) bulunurlar. Bu tür hücrelere post-mitotik hücreler adı verilir [169]. Metabolik olarak aktif olduklarından enerji ihtiyaçları diğer dokulara kıyasla fazladır ve bu yüzden mitokondri sayılarını fizyon ile artırırlar [167][170]. Replikasyon sürecinde, yabani tip mtDNA'nın çoğalmasının yanı sıra, varyant kazanmış bir mtDNA'nın da çoğaldığı gözlenen bir durum olarak bildirilmiştir [171]. Kan gibi devamlı bölünen hücrelerde oluşabilecek mtDNA varyantlarına sahip hücrelerin bölünme aşamasında mitotik segregasyonla mutant mtDNA'larının elenmesi durumunun mümkün olabileceği ancak kas gibi bölünmeyen hücrelerin böyle bir avantajı olmadığı için zamanla varyant birikiminin meydana gelebileceği belirtilmiştir [74][172]. Dokularda saptanan varyant sayılarının fazlalığı literatürde ayrıca yaşla ilişkilendirilmiştir. Yaş ilerledikçe somatik mtDNA varyantlarının birikim gösterdiği belirtilmiştir [173]. Bu bilgiler, erişkin olan gönüllü bireylerde saptanan ve patojenik olmadığı veri tabanları aracılığı ile doğrulanmış tek veya çoklu nükleotid değişimlerinin neden ve nasıl meydana geldiğinin anlaşılmasını sağlamaktadır. Özellikle mitokondri sayısının yüksek olduğu bilinen iskelet kası dokusunda, mtDNA replikasyon işlemi buna bağlı olarak fazla olacağından ve kas hücreleri bölünmeyen hücreler olduğu için mitotik segregasyon ile mutant mtDNA sayısını azaltamayacağından, bu çalışmada saptanan varyant miktarlarının ve heteroplazmi oranlarının diğer dokulara kıyasla kas dokusunda daha fazla olması anlamlı bulunmuştur.

Bu projede kan ve kas dokuları üzerinde yapılan çalışmalar, mtDNA'daki SNV ve indel varyantları temelinde, SNV için kasta çalışmanın, hem sayısal hem de frekans olarak daha yüksek değerlere ulaşabildiğini ve bu bakımdan daha yüksek bilgi vericilik barındırdığını göstermektedir. İndel tipi değişimler için ise, kan dokusunun kas

dokusuna göre varyant adedi olarak hafif bir üstünlüğü olsa da, frekans açısından kas dokusunun küçük bir farkla üstün olduğunu göstermektedir.

Çalışmalar devamlı bölünme gösteren kan hücrelerinde, yabanil tipli mtDNA'ların seçildiğini göstermiştir [174]. Delesyon tipi patojenik varyantların, çoğunlukla mtDNA replikasyonunda oluşan hatalardan değil, diğer değişimlerin tamir işlemleri sürecinde ortaya çıktığı bildirilmektedir [175]. Bu tez çalışmasındaki bulgulara paralel olarak, kanda özellikle yabanil tipli mtDNA'nın seçiminin yapılması için, olası noktasal (SNV) değişimlerin tamir edilmesi ve fakat bu işlemler yapılırken de delesyon tipli varyantların ortaya çıkması, kanda neden indel değişimlerin kasa kıyasla fazla olduğunu destekler niteliktedir. Bu durumda kanda bir aşamada heteroplazmik bir varyant mevcut olsa bile, yabanil mtDNA'nın seçilmesi nedeniyle, olası varyantların frekansı düşecek ve saptama hassasiyetinin altında kalacaktır. Hasta çalışmalarında da kandaki heteroplazmi oranlarının yaş ilerledikçe yabanil tip lehine azaldığı bildirilmiştir [176]. m.3243A>G patojenik varyantına sahip hastaların dahil edildiği bir çalışmada, hastalardan belirli zaman aralıklarında kan örnekleri alınmış, heteroplazmi oranları belirlenmiş ve yılda %2,3 oranında azalmanın gözlemlendiği belirtilmiştir [180].

Üç doku karşılaştırma sonuçlarında, idrarın, kan ve kastan daha düşük frekans değerine sahip olduğu görülmüştür. Ancak çalışmalar, kasın, idrar ve kan üzerinde, idrarınsa kan üzerinde frekans konusunda üstünlüğü olduğunu ve kasa en yakın frekans değerini veren dokunun idrar olduğunu bildirmektedir [177]-[179]. R.G. Whittaker ve ark.'nın (2009), m.3243A>G patojenik varyantına sahip 24 hastada yapılan doku karşılaştırma çalışmasında, kasın idrar ile heteroplazmi oranlarının birbirine yakın olduğu, hatta bazı hastalarda idrar oranının daha yüksek olduğu görülmüştür [177]. Bu tez çalışmasında, üç farklı doku çalışması sadece bir gönüllüde yapılabildiğinden, genel bir kanaate varmak için yetersizdir. Kan ve kas dokusundaki ortak varyantların frekans değerleri karşılaştırmasında IGV'ye göre iki gönüllüde kanın kastan daha yüksek frekans değeri verdiği, ancak toplam gönüllü sayısı üzerinden yapılan hesaplamada kasın daha yüksek frekans değerleri verdiği gözlenmiştir. Bu sebeple, üç doku karşılaştırmasının yapılacağı mtDNA varyant araştırmalarında da, doku örnekleri alınacak gönüllü sayısının artırılması, istatistiksel açıdan doğruluk oranını arttıracak bir unsur olacaktır. Yalnızca bir dokuda tespit edilen varyantların sayıları incelendiğinde en az iki doku ile yapılacak çalışmaların avantajından bahsetmek mümkündür. Grady ve

ark.'larının çalışmasında (2018) kan dokusunda tespit edilemeyen patojenik bir varyantın idrar dokusu ile tespit edilebildiği ve tanı konabildiği belirtilmiştir [180]. Bir başka çalışmada kan dokusunun az sayıda mutant mtDNA içeriyor olması nedeniyle mitokondriyal hastalık şüphesi varken, kanda yapılan patojenik varyant taramasının sonucu negatif ise, taramanın başka bir doku örneği kullanılarak yapılması önerilmiştir [181].

Bu çalışmada elde edilen sonuçlar için iki ayrı yazılımın kullanılması, verilerin güvenilirliğinin artırılması açısından önemli görüldü. Ion Reporter çıktısının mutlaka IGV üzerinden kontrol edilmesinin gerekliliği anlaşıldı. Farklı dokularda mtDNA heteroplazmik varyant karşılaştırmaları yapılırken, aynı gönüllüde, Ion Reporter sonuçlarına göre, bir dokuda heteroplazmi bilgisi verilen bir varyantın diğer dokuda saptanamadığı görüldüğünde, saptanamayan dokuda varyantın IGV üzerinden incelemesi yapıldığında, aslında ilgili varyantın bu dokuda da var olduğu gösterildi. Bu önemli detay, mitokondriyal analizlerde en az iki dokuda inceleme yapılmasının gerekliliğini vurgularken, analizlerde IGV üzerinde de kontrol yapılmasının önemini göstermiştir.

Gönüllülerde genel olarak kas dokusunun, saptanan varyant sayısı ve yükü bakımından en güçlü örneği oluşturduğu ortaya konmuştur. Özellikle SNV tipli varyantların kasta daha yoğun olduğu gözlenmiştir. Üç doku karşılaştırması ise sadece bir gönüllü üzerinde yapılabildiğinden yeterli bir sonuca ulaşılamamıştır. Gönüllü sayısının artırılmasıyla yapılacak olan çalışmalar daha doğru oranlara ulaşılmasını sağlayacaktır.

5.2. Veri Madenciliği ve Makine Öğrenmesi Sonuçlarının Değerlendirilmesi

Mitokondriyal hastalıklar, mitokondrinin işlevsel, sayısal ya da yapısal fonksiyonu için önemli mtDNA'dan ve/veya nDNA'dan kodlanan genlerde gözlenen patojenik değişimler sonucunda ortaya çıkmaktadır. MtDNA ilişkili hastalıklarda, dokulardaki heteroplazmi oranlarının ya da patojenik varyant yüklerinin, gözlenen fenotipik spektrumla ilişkili olduğu bilinmektedir [85][182][154]. Farklı kişilerde, aynı mtDNA patojenik varyantı, aynı dokuda, aynı yükte tespit edilse dahi, hiç bir klinik bulgunun olmaması halinden, çok ağır kliniğe kadar uzanan geniş bir fenotip spektrumu gözlenebilmektedir [183][184]. Bu durumun, diğer dokulardaki patojenik varyant yükünün farklılığı, hastaların yaş, cinsiyet gibi demografik özellikleri, aile öyküsü gibi

birçok etmene bağılı olarak geliştiğı bildirilmekte ve seyreden klinik değıřkenlik, genetik danıřmayı oldukça zorlayan bir durum olmaktadır. Ayrıca bir kiřideki heteroplazmi oranının da yařa bağılı olarak değıřtiğı gösterilmiřtir [123]. Hastalardaki kan örnekleri üzerinde yapılan alıřmalarda yař ilerledike patojenik varyant yükünde meydana gelen değıřimlerin, hastalığın klinik tablosu ile uyumsuzluk oluřturduğı, yükün azalmasına karřın kliniğın ağırlařtığı [182] veya beklendiğı üzere hafiflediğı de izlenebilmektedir [185]. Tüm bu durumlar, hastalık tanı ve takibinde zorluk yaratmaktadır.

Mitokondrinin iřlev kaybı ya da iřlevindeki azalma, majör olarak yařam için enerji metabolizmasına yüksek ihtiya duyulan organları etkilemektedir. Bu nedenle hastalığa sebep olan patojenik değıřimler farklı olsa dahi, etkilenmiř pek ok hastanın bulguları birbirine benzer olabilmekte ve tanı koymak güçleřmektedir [152]. Bu duruma karřın patojenik varyantın saptanması ve saptanan varyantın yükünün incelenmesi hastalığın boyutu hakkında bilgi verebilmektedir. Özellikle etkilenmiř dokudaki mitokondriyal genomun incelenmesi, tanısal olarak yüksek değeri taşısa da, ulařılamayan ya da biyopsisi yapılamayan beyin, kalp, böbrek gibi dokular için bu yaklařım imkansız olmaktadır. Prensipten genetik tanı testlerinin kas gibi enerji ihtiyacı yüksek dokularda yapılması önerilmektedir. Mitokondriyal hastalıkların tanı ve tedavisinde birçok değıřkenin birlikte değeriendirilmesi gerekmektedir. Genetik danıřmanlık verilmesini de zorlařtıran klinik bulgular, demografik özellikler ve dokusal patojenik varyant yükü iliřkisi gibi değıřken unsurları, bir arada değeriendirebilen, akılcı tahminler yapan ya da kolaylařtıran metodolojilere büyük ihtiya duyulmaktadır [123].

Bu tez alıřmasının ikinci basamağında, veri madenciliğı ve makine öğrenmesi metotları kullanılarak, fenotip ve patojenik varyant yükü arasındaki iliřkiyi saptamak yoluyla, hastalardan elde edilen bilgilerle, iskelet kası heteroplazmi oranı tahmini yapabilen bir regresyon modeli geliřtirmek hedeflendi. Bu amaçla, ilk ařamada patojenik mtDNA varyantları için MITOMAP veri tabanında toplanmıř makale sayıları incelendi. En yüksek sayıda makale ieren, Gorman ve ark.'larının alıřmasında, mitokondriyal hastalığa sebep olan en yaygın patojenik varyant olduğı görölmüş [121], MELAS hastalığı ile iliřkilendirilmiř [129], m.3243A>G varyantı seildi. Bu varyant ile ilgili kaynak sayısının ok olması, varyantın saptandığı ok sayıda hastanın arařtırılmıř

ve verisinin hakemli dergilerde yayımlanmış olmasındandır. Aynı zamanda, klinik özelliklerin aynı aile bireylerinde dahi çeşitlilik göstermesi durumunun, bu tez çalışmasının verimliliğini arttırıcı bir unsur olacağı düşünüldü. Tahmin araçlarının geliştirilmesinde, ne kadar çok hastadan bilgi toplanabilirse, patojenik varyant yükü ile hastadan elde edilen bilgiler arasındaki ilişkinin o denli gerçeğe yakın tahmin vereceği öngörüldü. Bilgi elde edilebilecek kaynaklar kısıtlı olduğundan, bilgi sağlanan olguların, klinik bulgularının ve belirtilerinin, biyokimyasal test sonuçlarının, beyin görüntüleme sonuçlarının, kas histoloji sonuçlarının eksiksiz olmasının önemli olduğu düşünüldü. Belirtilen unsurlar doğrultusunda oluşturulan veri seti üzerinde çalışma gerçekleştirildi.

Veri setinde bulunan değişkenlerin birbirleri arasındaki ilişki boyutunu sayısal olarak belirten VIF sonuçlarına göre, *iskelet_kası* sütununun bağımsız değişkenlerle ilişkisinin güçlü olmadığı gözlemlendi. Bu sonucun, bağımsız değişkenlerdeki çeşitliliğin fazla olması durumu ve hasta sayısının az olması durumu ile ilişkili olabileceği düşünüldü. Eksik verilerden arındırılan *df* veri setinde VIF sonucunun yükselmiş olması bu çıkarımı doğrular nitelikte bulunmuştur. Model sonuçları incelendiğinde de varyant ile ilişkili görülen birçok klinik özelliğin iskelet kası heteroplazmi oranları ile ilişkili olmadığı görülebilmektedir. Ensefalopati, inme benzeri epizodlar, sensörinöral işitme kaybı, maküler dejenerasyon, pitoz, kardiyomiyopati gibi bulguların p-değerleri 0,05'ten yüksek bulunmuştur. Mancuso ve ark.'larının, m.3243A>G patojenik varyantı görülen 126 hasta üzerinde yaptığı çalışmada, hastalarda sıklıkla görülen klinik bulguların patojenik varyant yükü ile ilişkisinin olmadığı görülmüştür. İnme benzeri epizodlar, diyabet, işitme kaybı, migren ve kognitif bozukluk bulgularının hastalığın tanısını desteklediği belirtilmiştir ancak heteroplazmi oranları ile bulgular arasında bir ilişki bulunamadığı, patojenik varyantın saptandığı hastaların klinik özelliklerinin tahmin edilebilirliğinin olmadığı belirtilmiştir. Bu durumun, aynı klinik özelliğe sahip hastaların heteroplazmi oranlarının değer aralığının geniş olmasının bir sonucu olabileceği düşünüldü. Değer aralığı geniş olduğu için ve belli bir değerde yoğunluk gözlenmediği için ilişkilendirme çalışması da başarısız olmaktadır [153]. Çalışma sonuçları, bu tez çalışmasında oluşturulan model sonuçlarını destekler niteliktedir [178].

Bir tahmin modeli oluşturmak üzere hazırlanan veri setinde art arda gelen örneklerin aynı merkezden toplanmış olma durumu vardır. Yayınlarda belli klinik

özellikler üzerinde çalışma yapılmış olması ve bu klinik özelliklere sahip hastaların yakın değerlerde heteroplazmi oranı vermiş olması durumunun, yapılan testler sonucunda ortaya çıkan otokorelasyon varlığını (Tablo 4-37, Tablo 4-41) açıklayabileceği düşünüldü.

İkinci model denemesi için kullanılan veri setinde otokorelasyonun yanı sıra heteroskedastisite durumu da görülmüştür. Genellikle uç değer varlığında gözlemlenen bu durum değerler aralığının geniş olması ile de ilişkilendirilmiştir [153]. Benzer klinik özelliklerin görüldüğü hastalarda dahi gözlemlenen en küçük ve en büyük heteroplazmi değerleri arasındaki farkın büyük olması, gözlemlenen ve tahmin edilen değerler arasındaki fark olan hata değerlerinde görülen değişkenliğin sabit olmaması yani heteroskedastisite durumunu açıklayabileceği düşünüldü.

Oluşturulan Model-1 sonuçlarına göre, bağımsız değişkenlerin heteroplazmi oranındaki değişkenlikleri açıklama başarısı %47 (ayarlanmış R-kare)'dir. Grady ve ark.'ları benzer bir çalışma yapmış ve m.3243A>G patojenik varyantı kaynaklı hastalık yükündeki çeşitliliği, çoklu doğrusal (multi-lineer) regresyon analizi uygulayarak kas heteroplazmi yükü, yaş ve mtDNA kopya sayısı bağımsız değişkenleri ile açıklamaya çalışmışlar ve çalışmanın sonucu olarak R-kare değerlerini %40 ($P<0,01$) bulmuşlardır [180]. Belirtilen hastalık yükü bilgisi NMDAS skorlama sisteminden gelmektedir [186]. Bu sonuçlar, çalışmamızdaki sonuçları destekler nitelikte bulunmuştur. Ayarlanmış R-kare değerindeki düşüklük, varyantın sebep olduğu klinik spektrumun genişliğinden kaynaklı olarak yorumlanmıştır [180]. Model-2'de değerlerin yükselmesi (ayarlanmış R-kare değeri) veri sayısının artması ile ilişkili bulunmuştur.

Literatürde verilen m.3243A>G patojenik varyantına sahip olduğu belirlenen hastaların klinik özellikleri toplanarak, elde edilen veri setiyle, kas heteroplazmi oranının tahminini yapabilen bir regresyon modeli üzerine çalışıldı. Modelleme sonuçları ve yapılan değerlendirmeler sonucunda veri sayısının, hastalarda görülen klinik özelliklerdeki, bulgulardaki ve belirtilerdeki çeşitliliği açıklamakta yetersiz kaldığı görüldü. Yine de eldeki verilerle % 47 oranında doğruluk veren Model-1'in tahmin gücü kullanılarak, kas heteroplazmi bilgisinin bulunmadığı (*nan_kas_df*) veri setindeki eksik veri tamamlanmış ve böylece kullanılabilir veri sayısı %100 olan *df* veri setinden oluşturulan Model-2 ile %72'lik doğruluk oranı elde edilmiştir. Doğruluk oranındaki artış, açıkça veri sayısının artması ile ilişkiliyken, eksik veriler için yapılan

müdahalenin başka bir modele bağlı olduğu göz önünde bulundurulmalıdır. Kullanılan yayınlarda eksik veri varlığının fazla olduğu, çok sayıdaki farklı yayında, klinik verinin eş detayda paylaşılmamış olduğu gözlenmiştir. Bu durum, bu tez çalışmasında uygulanan veri madenciliği araştırmasındaki en büyük sorunu oluşturmuştur. Literatürden yapılan metin taramaları ve ulaşılan yayınların titizlikle okunmasıyla elde edilen veriler, gerek standart olmaması gerekse kullanılan laboratuvar yöntemlerinin farklı olması nedeniyle, gerçek iskelet kası heteroplazmi oranına en yakın tahmini üretme işleminde yetersiz kalmıştır. Bu çalışma, geliştirdiğimiz regresyon modelinin, doğruluk oranı yüksek tahminler verebilmesi için ulusal ve uluslararası standartlara uygun ve güncelliğini koruyacak biçimde, verilerin kayıt altında tutulmasının ve veri banka sistemlerinin önemini göstermiştir.

KAYNAKLAR

1. L. Ernster and G. Schatz, "Mitochondria: a historical review.," J. Cell Biol., vol. 91, no. 3, pp. 227s-255s, Dec. 1981.
2. "mitochondria | Origin and meaning of mitochondria by Online Etymology Dictionary." [Online. Available: <https://www.etymonline.com/word/mitochondria>. [Accessed: 06-Nov-2019].
3. P. R. Rich, "The molecular machinery of Keilin's respiratory chain," Biochem. Soc. Trans., vol. 31, no. 6, pp. 1095–1105, Dec. 2003.
4. A. F. W. Schimper, "Über die Entwicklung der Chlorophyllkörner und Farbkörper," Bot. Zeitung, vol. 41, pp. 105–14, 121–31, 137–46, 153–62, 1883.
5. K. Mereschkowsky, "Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen," Biol Centralbl., 1910. [Online. Available: https://archive.org/details/cbarchive_42856_theoriederzweiplasmaartenalsgr1881/page/n2. [Accessed: 21-Nov-2019].
6. K. Mereschkowsky, "Über Natur und Ursprung der Chromatophoren im Pflanzenreiche," Biol Centralbl., 1905. [Online. Available: https://archive.org/details/cbarchive_51353_bernaturundursprungderchromato1881/page/n2. [Accessed: 21-Nov-2019].
7. W. Martin and K. Kowallik, "Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche,'" Eur. J. Phycol., vol. 34, no. 3, pp. 287–295, Aug. 1999.
8. J. T. O. Kirk, "Roots: The discovery of chloroplast DNA," BioEssays, vol. 4, no. 1, pp. 36–38, Jan. 1986.
9. M. M. NASS and S. NASS, "INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. I. FIXATION AND ELECTRON STAINING REACTIONS.," J. Cell Biol., vol. 19, no. 3, pp. 593–611, Dec. 1963.
10. G. Schatz, E. Haslbrunner, and H. Tuppy, "Deoxyribonucleic acid associated with yeast mitochondria," Biochem. Biophys. Res. Commun., vol. 15, no. 2, pp. 127–132, Mar. 1964.
11. L. Sagan, "On the origin of mitosing cells," J. Theor. Biol., vol. 14, no. 3, pp. 225–IN6, Mar. 1967.
12. C. Darwin, On the Origin of Species, Revised Ed. Oxford University Press, 2009.

13. M. Martin, William F.; Müller, Origin of Mitochondria and Hydrogenosomes. Springer Berlin Heidelberg, 2007.
14. V. V. Emelyanov, "Mitochondrial connection to the origin of the eukaryotic cell," *Eur. J. Biochem.*, vol. 270, no. 8, pp. 1599–1618, Apr. 2003.
15. V. V. Emelyanov, "Rickettsiaceae, Rickettsia-Like Endosymbionts, and the Origin of Mitochondria," *Biosci. Rep.*, vol. 21, no. 1, pp. 1–17, Feb. 2001.
16. S. G. E. Andersson et al., "The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria," *Nature*, vol. 396, no. 6707, pp. 133–140, Nov. 1998.
17. M. M. Leger et al., "An ancestral bacterial division system is widespread in eukaryotic mitochondria," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 33, pp. 10239–46, Aug. 2015.
18. I. Scott and R. J. Youle, "Mitochondrial fission and fusion," *Essays Biochem.*, vol. 47, pp. 85–98, 2010.
19. C. L. Bunn, C. H. Mitchell, H. B. Lukins, and A. W. Linnane, "Biogenesis of mitochondria. 18. A new class of cytoplasmically determined antibiotic resistant mutants in *Saccharomyces cerevisiae*," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 67, no. 3, pp. 1233–40, Nov. 1970.
20. S. Kalghatgi et al., "Bactericidal antibiotics induce mitochondrial dysfunction and oxidative damage in Mammalian cells," *Sci. Transl. Med.*, vol. 5, no. 192, p. 192ra85, Jul. 2013.
21. W. H. Lewis, Margaret R.; Lewis, *The American Journal of Anatomy*, vol. v. 17 (191. Baltimore, MD : [publisher not identified, 1915.
22. M. Zhan, C. Brooks, F. Liu, L. Sun, and Z. Dong, "Mitochondrial dynamics: regulatory mechanisms and emerging role in renal pathophysiology," *Kidney Int.*, vol. 83, no. 4, pp. 568–81, Apr. 2013.
23. T. Kuroiwa et al., "Structure, function and evolution of the mitochondrial division apparatus," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1763, no. 5–6, pp. 510–521, May 2006.
24. C. R. Hackenbrock, "ULTRASTRUCTURAL BASES FOR METABOLICALLY LINKED MECHANICAL ACTIVITY IN MITOCHONDRIA," *J. Cell Biol.*, vol. 30, no. 2, pp. 269–297, Aug. 1966.
25. M. Pavelka and J. Roth, "Mitochondria: Crista and Tubulus Types," in *Functional Ultrastructure*, Vienna: Springer Vienna, 2010, pp. 128–129.
26. J. Bereiter-Hahn, "Behavior of Mitochondria in the Living Cell," *Int. Rev. Cytol.*, vol. 122, pp. 1–63, Jan. 1990.

27. L. Wiemerslage and D. Lee, "Quantification of mitochondrial morphology in neurites of dopaminergic neurons using multiple parameters,," *J. Neurosci. Methods*, vol. 262, pp. 56–65, Mar. 2016.
28. L. W. Cole, "The Evolution of Per-cell Organelle Number," *Front. Cell Dev. Biol.*, vol. 4, p. 85, Aug. 2016.
29. H. Lemieux and C. L. Hoppel, "Mitochondria in the human heart," *J. Bioenerg. Biomembr.*, vol. 41, no. 2, pp. 99–106, Apr. 2009.
30. R. Ferrari, A. Cargnoni, and C. Ceconi, "Anti-ischaemic effect of ivabradine," *Pharmacol. Res.*, vol. 53, no. 5, pp. 435–439, May 2006.
31. E. Barth, G. Stämmler, B. Speiser, and J. Schaper, "Ultrastructural quantitation of mitochondria and myofilaments in cardiac muscle from 10 different animal species including man," *J. Mol. Cell. Cardiol.*, vol. 24, no. 7, pp. 669–681, 1992.
32. S. Das et al., "Mitochondrial morphology and dynamics in hepatocytes from normal and ethanol-fed rats," *Pflugers Arch. Eur. J. Physiol.*, vol. 464, no. 1, pp. 101–109, Jul. 2012.
33. H. Wu et al., "Regulation of mitochondrial biogenesis in skeletal muscle by CaMK,," *Science*, vol. 296, no. 5566, pp. 349–52, Apr. 2002.
34. D. Sebastián and A. Zorzano, "Mitochondrial dynamics and metabolic homeostasis," *Curr. Opin. Physiol.*, vol. 3, pp. 34–40, Jun. 2018.
35. "Blood cells." [Online. Available: [https://web.archive.org/web/20160723113019/http://www.biosbcc.net/doohan/sample/htm/Blood cells.htm](https://web.archive.org/web/20160723113019/http://www.biosbcc.net/doohan/sample/htm/Blood%20cells.htm). [Accessed: 06-Nov-2019].
36. A. Mescher, Junqueira's Basic Histology: Text and Atlas, Fifteenth. McGraw-Hill Education, 2018.
37. "Physiology of red blood cells. Erythron. Respiratory pigments. Rheological properties of blood." [Online. Available: [http://intranet.tdmu.edu.ua/data/kafedra/internal/normal_phiz/classes_stud/en/medlik/2 course/2 Cycle Physiology of blood/01 Physiology of red blood.htm](http://intranet.tdmu.edu.ua/data/kafedra/internal/normal_phiz/classes_stud/en/medlik/2%20course/2%20Cycle%20Physiology%20of%20blood/01%20Physiology%20of%20red%20blood.htm). [Accessed: 06-Nov-2019].
38. H. Alshaabi, M. Heininger, and B. Cunliffe, "Dynamic regulation of subcellular mitochondrial position for localized metabolite levels," *J. Biochem.*, Jul. 2019.
39. A. United States and Canadian Academy of Pathology., Laboratory investigation., vol. 24, no. 4. Lippincott Williams & Wilkins, 1971.
40. A.-B. Al-Mehdi et al., "Perinuclear mitochondrial clustering creates an oxidant-rich nuclear domain required for hypoxia-induced transcription,," *Sci. Signal.*, vol. 5, no. 231, p. ra47, Jul. 2012.

41. S. Campello, R. A. Lacalle, M. Bettella, S. Mañes, L. Scorrano, and A. Viola, "Orchestration of lymphocyte chemotaxis by mitochondrial dynamics," *J. Exp. Med.*, vol. 203, no. 13, pp. 2879–2886, Dec. 2006.
42. G. Morlino et al., "Miro-1 Links Mitochondria and Microtubule Dynein Motors To Control Lymphocyte Migration and Polarity," *Mol. Cell. Biol.*, vol. 34, no. 8, pp. 1412–1426, Apr. 2014.
43. W. M. Saxton and P. J. Hollenbeck, "The axonal transport of mitochondria.," *J. Cell Sci.*, vol. 125, no. Pt 9, pp. 2095–104, May 2012.
44. E. G. CHRISTIANSEN, "Orientation of the Mitochondria during Mitosis," *Nature*, vol. 163, no. 4140, pp. 361–361, Mar. 1949.
45. S. Campello and L. Scorrano, "Mitochondrial shape changes: orchestrating cell pathophysiology.," *EMBO Rep.*, vol. 11, no. 9, pp. 678–84, Sep. 2010.
46. M. Liesa and O. S. Shirihai, "Mitochondrial dynamics in the regulation of nutrient utilization and energy expenditure.," *Cell Metab.*, vol. 17, no. 4, pp. 491–506, Apr. 2013.
47. A. Sugiura, G. McLelland, E. A. Fon, and H. M. McBride, "A new pathway for mitochondrial quality control: mitochondrial derived vesicles," *EMBO J.*, vol. 33, no. 19, pp. 2142–2156, Oct. 2014.
48. S. Krauss, "Mitochondria: Structure and Role in Respiration," in *Encyclopedia of Life Sciences*, Chichester, UK: John Wiley & Sons, Ltd, 2001.
49. G. M. Cooper, *The cell : a molecular approach*. ASM Press, 2000.
50. J. Dudek, P. Rehling, and M. van der Laan, "Mitochondrial protein import: Common principles and physiological networks," *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1833, no. 2, pp. 274–285, Feb. 2013.
51. W. Neupert, "PROTEIN IMPORT INTO MITOCHONDRIA," *Annu. Rev. Biochem.*, vol. 66, no. 1, pp. 863–917, Jun. 1997.
52. S. R. Lee and J. Han, "Mitochondrial Nucleoid: Shield and Switch of the Mitochondrial Genome," *Oxid. Med. Cell. Longev.*, vol. 2017, pp. 1–15, Jun. 2017.
53. R. Gilkerson et al., "The mitochondrial nucleoid: integrating mitochondrial DNA into cellular homeostasis.," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 5, p. a011080, May 2013.
54. C. Kukat, C. A. Wurm, H. Spåhr, M. Falkenberg, N.-G. Larsson, and S. Jakobs, "Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 33, pp. 13534–9, Aug. 2011.

55. M. Satoh and T. Kuroiwa, "Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell," *Exp. Cell Res.*, vol. 196, no. 1, pp. 137–140, Sep. 1991.
56. S. Anderson et al., "Sequence and organization of the human mitochondrial genome," *Nature*, vol. 290, no. 5806, pp. 457–465, Apr. 1981.
57. H. Kasamatsu and J. Vinograd, "Replication of Circular DNA in Eukaryotic Cells," *Annu. Rev. Biochem.*, vol. 43, no. 1, pp. 695–719, Jun. 1974.
58. T. Pan, "Modifications and functional genomics of human transfer RNA," *Cell Res.*, vol. 28, no. 4, pp. 395–404, Apr. 2018.
59. "Organelle Structure and Function," *Med. Cell Biol.*, pp. 101–148, Jan. 2008.
60. J. Cummins, "Mitochondrial DNA in mammalian reproduction," *Rev. Reprod.*, vol. 3, no. 3, pp. 172–182, Sep. 1998.
61. "Mitochondrial DNA Polymorphisms and Anthropology," in *Mitochondria*, New York, USA: John Wiley & Sons, Inc., 2002, pp. 326–341.
62. J.-W. Taanman, "The mitochondrial genome: structure, transcription, translation and replication," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1410, no. 2, pp. 103–123, Feb. 1999.
63. I. J. Holt and A. Reyes, "Human mitochondrial DNA replication," *Cold Spring Harb. Perspect. Biol.*, vol. 4, no. 12, Dec. 2012.
64. N. B. Hecht, H. Liem, K. C. Kleene, R. J. Distel, and S. Ho, "Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocyte mitochondrial DNA," *Dev. Biol.*, vol. 102, no. 2, pp. 452–461, Apr. 1984.
65. P. Sutovsky, R. D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly, and G. Schatten, "Ubiquitin tag for sperm mitochondria," *Nature*, vol. 402, no. 6760, pp. 371–372, Nov. 1999.
66. J. St John et al., "Failure of elimination of paternal mitochondrial DNA in abnormal embryos," *Lancet*, vol. 355, no. 9199, p. 200, Jan. 2000.
67. S. Luo et al., "Biparental Inheritance of Mitochondrial DNA in Humans," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 51, pp. 13039–13044, Dec. 2018.
68. T. G. McWilliams and A. Suomalainen, "Mitochondrial DNA can be inherited from fathers, not just mothers," *Nature*, vol. 565, no. 7739, pp. 296–297, Jan. 2019.
69. J. Vissing, "Paternal comeback in mitochondrial DNA inheritance," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 5, pp. 1475–1476, Jan. 2019.

70. B. J. Greber and N. Ban, "Structure and Function of the Mitochondrial Ribosome," *Annu. Rev. Biochem.*, vol. 85, no. 1, pp. 103–132, Jun. 2016.
71. A. Amunts, A. Brown, J. Toots, S. H. W. Scheres, and V. Ramakrishnan, "The structure of the human mitochondrial ribosome.," *Science*, vol. 348, no. 6230, pp. 95–98, Apr. 2015.
72. N. Bousette et al., "Large-Scale Characterization and Analysis of the Murine Cardiac Proteome," *J. Proteome Res.*, vol. 8, no. 4, pp. 1887–1901, Apr. 2009.
73. S. E. Calvo and V. K. Mootha, "The mitochondrial proteome and human disease.," *Annu. Rev. Genomics Hum. Genet.*, vol. 11, pp. 25–44, 2010.
74. P. F. Chinnery, *Mitochondrial Disorders Overview*. University of Washington, Seattle, 1993.
75. "homo- | Search Online Etymology Dictionary." [Online. Available: <https://www.etymonline.com/search?q=homo->. [Accessed: 15-Dec-2019.
76. "hetero- | Origin and meaning of prefix hetero- by Online Etymology Dictionary." [Online. Available: <https://www.etymonline.com/word/hetero->. [Accessed: 15-Dec-2019.
77. "-plasm | Origin and meaning of suffix -plasm by Online Etymology Dictionary." [Online. Available: <https://www.etymonline.com/word/-plasm>. [Accessed: 15-Dec-2019.
78. M. S. Damian, P. Seibel, W. Schachenmayr, H. Reichmann, and W. Dorndorf, "VACTERL with the mitochondrial NP 3243 point mutation," *Am. J. Med. Genet.*, vol. 62, no. 4, pp. 398–403, Apr. 1996.
79. H. A. Coller, K. Khrapko, N. D. Bodyak, E. Nekhaeva, P. Herrero-Jimenez, and W. G. Thilly, "High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection," *Nat. Genet.*, vol. 28, no. 2, pp. 147–150, Jun. 2001.
80. R. M. Smigrodzki and S. M. Khan, "Mitochondrial Microheteroplasmy and a Theory of Aging and Age-Related Disease," *Rejuvenation Res.*, vol. 8, no. 3, pp. 172–198, Sep. 2005.
81. W. D. Parker and J. K. Parks, "Mitochondrial ND5 mutations in idiopathic Parkinson's disease," *Biochem. Biophys. Res. Commun.*, vol. 326, no. 3, pp. 667–669, Jan. 2005.
82. C. J. Carroll, V. Brilhante, and A. Suomalainen, "Next-generation sequencing for mitochondrial disorders.," *Br. J. Pharmacol.*, vol. 171, no. 8, pp. 1837–53, Apr. 2014.

83. T. Huang, "Next Generation Sequencing to Characterize Mitochondrial Genomic DNA Heteroplasmy," *Curr. Protoc. Hum. Genet.*, vol. 0 19, p. Unit19.8, Oct. 2011.
84. S. Tang and T. Huang, "Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system," *Biotechniques*, vol. 48, no. 4, pp. 287–296, Apr. 2010.
85. G. S. Gorman et al., "Mitochondrial diseases," *Nat. Rev. Dis. Prim.*, vol. 2, no. 1, p. 16080, Dec. 2016.
86. Y. Hatefi, "The Mitochondrial Electron Transport and Oxidative Phosphorylation System," *Annu. Rev. Biochem.*, vol. 54, no. 1, pp. 1015–1069, Jun. 1985.
87. H. M. Kalckar, "50 Years of Biological Research—From Oxidative Phosphorylation to Energy Requiring Transport Regulation," *Annu. Rev. Biochem.*, vol. 60, no. 1, pp. 1–38, Jun. 1991.
88. J. B. Taylor, Martha R.; Simon, Eric J.; Dickey, Jean L.; Hogan, Kelly A.; Reece, Campbell *Biology: Concepts & Connections*, 9th ed. Pearson, 2017.
89. H. A. Krebs and W. A. Johnson, "Metabolism of ketonic acids in animal tissues.," *Biochem. J.*, vol. 31, no. 4, pp. 645–60, Apr. 1937.
90. J. M. Berg, J. L. Tymoczko, and L. Stryer, "The Citric Acid Cycle," 2002.
91. S. Horowitz and R. C. Trievel, "Carbon-oxygen hydrogen bonding in biological structure and function.," *J. Biol. Chem.*, vol. 287, no. 50, pp. 41576–82, Dec. 2012.
92. R. A. Capaldi and R. Aggeler, "Mechanism of the F1F0-type ATP synthase, a biological rotary motor," *Trends Biochem. Sci.*, vol. 27, no. 3, pp. 154–160, Mar. 2002.
93. A. Haddad and S. S. Mohiuddin, *Biochemistry, Citric Acid Cycle*. StatPearls Publishing, 2019.
94. D. G. Hardie, "Sensing of energy and nutrients by AMP-activated protein kinase," *Am. J. Clin. Nutr.*, vol. 93, no. 4, pp. 891S-896S, Apr. 2011.
95. D. Carling, M. J. Sanders, and A. Woods, "The regulation of AMP-activated protein kinase by upstream kinases," *Int. J. Obes.*, vol. 32, no. S4, pp. S55–S59, Sep. 2008.
96. M. M. Mihaylova and R. J. Shaw, "The AMPK signalling pathway coordinates cell growth, autophagy and metabolism.," *Nat. Cell Biol.*, vol. 13, no. 9, pp. 1016–23, Sep. 2011.

97. D. G. Hardie, "AMP-activated protein kinase: the guardian of cardiac energy status,," *J. Clin. Invest.*, vol. 114, no. 4, pp. 465–8, Aug. 2004.
98. E. Q. Toyama et al., "AMP-activated protein kinase mediates mitochondrial fission in response to energy stress,," *Science*, vol. 351, no. 6270, pp. 275–281, Jan. 2016.
99. S. Herzig and R. J. Shaw, "AMPK: guardian of metabolism and mitochondrial homeostasis,," *Nat. Rev. Mol. Cell Biol.*, vol. 19, no. 2, pp. 121–135, Feb. 2018.
100. L. Contreras, I. Drago, E. Zampese, and T. Pozzan, "Mitochondria: The calcium connection,," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1797, no. 6–7, pp. 607–618, Jun. 2010.
101. X. Li, P. Fang, J. Mai, E. T. Choi, H. Wang, and X. Yang, "Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers,," *J. Hematol. Oncol.*, vol. 6, p. 19, Feb. 2013.
102. E. Cadenas and K. J. Davies, "Mitochondrial free radical generation, oxidative stress, and aging,," *Free Radic. Biol. Med.*, vol. 29, no. 3–4, pp. 222–30, Aug. 2000.
103. D. R. Green, "Apoptotic pathways: the roads to ruin,," *Cell*, vol. 94, no. 6, pp. 695–8, Sep. 1998.
104. K. J. Menzies, H. Zhang, E. Katsyuba, and J. Auwerx, "Protein acetylation in metabolism-metabolites and cofactors,," *Nature Reviews Endocrinology*, vol. 12, no. 1. Nature Publishing Group, pp. 43–60, 01-Jan-2016.
105. O. Matilainen, P. M. Quirós, and J. Auwerx, "Mitochondria and Epigenetics – Crosstalk in Homeostasis and Stress,," *Trends in Cell Biology*, vol. 27, no. 6. Elsevier Ltd, pp. 453–463, 01-Jun-2017.
106. H. Zhang, K. J. Menzies, and J. Auwerx, "The role of mitochondria in stem cell fate and aging,," *Development (Cambridge)*, vol. 145, no. 8. Company of Biologists Ltd, 15-Apr-2018.
107. V. Jovaisaite, L. Mouchiroud, and J. Auwerx, "The mitochondrial unfolded protein response, a conserved stress response pathway with implications in health and disease,," *Journal of Experimental Biology*, vol. 217, no. 1. Company of Biologists Ltd, pp. 137–143, 2014.
108. N. Sun, R. J. Youle, and T. Finkel, "The Mitochondrial Basis of Aging,," *Molecular Cell*, vol. 61, no. 5. Cell Press, pp. 654–666, 03-Mar-2016.
109. N. Shyh-Chang and H. H. Ng, "The metabolic programming of stem cells,," *Genes and Development*, vol. 31, no. 4. Cold Spring Harbor Laboratory Press, pp. 336–346, 15-Feb-2017.

110. J. R. Friedman, L. L. Lackner, M. West, J. R. DiBenedetto, J. Nunnari, and G. K. Voeltz, "ER tubules mark sites of mitochondrial division.," *Science*, vol. 334, no. 6054, pp. 358–62, Oct. 2011.
111. D. Martinvalet, "The role of the mitochondria and the endoplasmic reticulum contact sites in the development of the immune responses," *Cell Death Dis.*, vol. 9, no. 3, p. 336, Mar. 2018.
112. S. Marchi, S. Patergnani, and P. Pinton, "The endoplasmic reticulum–mitochondria connection: One touch, multiple functions," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1837, no. 4, pp. 461–469, Apr. 2014.
113. S.-H. Shim et al., "Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 35, pp. 13978–83, Aug. 2012.
114. R. Rizzuto et al., "Ca(2+) transfer from the ER to mitochondria: when, how and why.," *Biochim. Biophys. Acta*, vol. 1787, no. 11, pp. 1342–51, Nov. 2009.
115. M. Hamasaki et al., "Autophagosomes form at ER–mitochondria contact sites," *Nature*, vol. 495, no. 7441, pp. 389–393, Mar. 2013.
116. S. Missiroli et al., "Mitochondria-associated membranes (MAMs) and inflammation," *Cell Death Dis.*, vol. 9, no. 3, p. 329, Mar. 2018.
117. R. LUFT, D. IKKOS, G. PALMIERI, L. ERNSTER, and B. AFZELIUS, "A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study.," *J. Clin. Invest.*, vol. 41, no. 9, pp. 1776–804, Sep. 1962.
118. D. C. Wallace et al., "Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy.," *Science*, vol. 242, no. 4884, pp. 1427–30, Dec. 1988.
119. "MeSH Browser." [Online. Available: <https://meshb.nlm.nih.gov/record/ui?ui=D028361>. [Accessed: 20-Dec-2019.
120. J. Neustadt and S. R. Pieczenik, "Medication induced mitochondrial damage and disease," *Mol. Nutr. Food Res.*, vol. 52, no. 7, pp. 780–788, Jul. 2008.
121. G. S. Gorman et al., "Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease.," *Ann. Neurol.*, vol. 77, no. 5, pp. 753–9, May 2015.
122. M. J. Falk, A. Decherney, and J. P. Kahn, "Mitochondrial Replacement Techniques--Implications for the Clinical Community.," *N. Engl. J. Med.*, vol. 374, no. 12, pp. 1103–6, Mar. 2016.

123. N. A. Khan, P. Govindaraj, A. K. Meena, and K. Thangaraj, "Mitochondrial disorders: Challenges in diagnosis & treatment," *Indian J. Med. Res.*, vol. 141, no. 1, p. 13, 2015.
124. L. Craven et al., "Pronuclear transfer in human embryos to prevent transmission of mitochondrial DNA disease," *Nature*, vol. 465, no. 7294, pp. 82–5, May 2010.
125. "Homo sapiens mitochondrion, complete genome - Nucleotide - NCBI." [Online. Available: https://www.ncbi.nlm.nih.gov/nucore/NC_012920. [Accessed: 02-Oct-2020.
126. "MT-TL1 gene symbol report | HUGO Gene Nomenclature Committee." [Online. Available: https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:7490. [Accessed: 02-Oct-2020.
127. "MTERF1 gene symbol report | HUGO Gene Nomenclature Committee." [Online. Available: https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:21463. [Accessed: 02-Oct-2020.
128. S. C. Nam and C. Kang, "DNA light-strand preferential recognition of human mitochondria transcription termination factor mTERF," *J. Biochem. Mol. Biol.*, vol. 38, no. 6, pp. 690–694, 2005.
129. Y. Goto et al., "Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (MELAS): a correlative study of the clinical features and mitochondrial DNA mutation," *Neurology*, vol. 42, no. 3 Pt 1, pp. 545–50, Mar. 1992.
130. J. F. Hess, M. A. Parisi, J. L. Bennett, and D. A. Clayton, "Impairment of mitochondrial transcription termination by a point mutation associated with the MELAS subgroup of mitochondrial encephalomyopathies," *Nature*, vol. 351, no. 6323, pp. 236–239, 1991.
131. E. Yakubovskaya, E. Mejia, J. Byrnes, E. Hambardjieva, and M. Garcia-Diaz, "Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription," *Cell*, vol. 141, no. 6, pp. 982–993, Jun. 2010.
132. A. Chomyn et al., "MELAS mutation in mtDNA binding site for transcription termination factor causes defects in protein synthesis and in respiration but no change in levels of upstream and downstream mature transcripts," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 10, pp. 4221–4225, 1992.
133. T. Suzuki, T. Suzuki, T. Wada, K. Saigo, and K. Watanabe, "Novel taurine-containing uridine derivatives and mitochondrial human diseases," *Nucleic Acids Res. Suppl.*, vol. 1, no. 1, pp. 257–258, Nov. 2001.

134. Y. Kirino, Y. I. Goto, Y. Campos, J. Arenas, and T. Suzuki, "Specific correlation between the wobble modification deficiency in mutant tRNAs and the clinical features of a human mitochondrial disease," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 20, pp. 7127–7132, May 2005.
135. T. Yasukawa, T. Suzuki, T. Suzuki, T. Ueda, S. Ohta, and K. Watanabe, "Modification defect at anticodon wobble nucleotide of mitochondrial tRNAs(Leu)(UUR) with pathogenic mutations of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes," *J. Biol. Chem.*, vol. 275, no. 6, pp. 4251–4257, Feb. 2000.
136. "OMIM 540000." [Online. Available: <https://omim.org/entry/540000>.
137. S. G. Pavlakis, P. C. Phillips, S. DiMauro, D. C. De Vivo, and L. P. Rowland, "Mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes: A distinctive clinical syndrome," *Ann. Neurol.*, vol. 16, no. 4, pp. 481–488, Oct. 1984.
138. N. Manwaring et al., "Population prevalence of the MELAS A3243G mutation," *Mitochondrion*, vol. 7, no. 3, pp. 230–233, May 2007.
139. A. W. El-Hattab, M. Almannai, and F. Scaglia, MELAS. University of Washington, Seattle, 1993.
140. D. M. Sproule and P. Kaufmann, "Mitochondrial Encephalopathy, Lactic Acidosis, and Strokelike Episodes," *Ann. N. Y. Acad. Sci.*, vol. 1142, no. 1, pp. 133–158, Oct. 2008.
141. "OMIM Entry - * 590050 - TRANSFER RNA, MITOCHONDRIAL, LEUCINE, 1; MTTTL1." [Online. Available: <https://omim.org/entry/590050#0001>. [Accessed: 30-Sep-2020.
142. J. Han, M. Kamber, and J. Pei, *Data mining : concepts and techniques*. Elsevier Science, 2011.
143. U. Fayyad, U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, pp. 37--54, 1996.
144. M. C. Lovell, "Data Mining," *Rev. Econ. Stat.*, vol. 65, no. 1, p. 1, Feb. 1983.
145. E. Alpaydin, *Introduction to machine learning*. .
146. "Ion GeneStudio S5 Series for NGS | Thermo Fisher Scientific - TR." [Online. Available: <https://www.thermofisher.com/tr/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-s5-ngs-targeted-sequencing.html>. [Accessed: 15-Aug-2020.

147. S. Dames, K. Eilbeck, and R. Mao, "A high-throughput next-generation sequencing assay for the mitochondrial genome," *Methods Mol. Biol.*, vol. 1264, pp. 77–87, 2015.
148. J. Naue et al., "Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA," *Mitochondrion*, vol. 20, no. 1, pp. 82–94, Jan. 2015.
149. P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020.
150. D. M. Nathan et al., "Impaired fasting glucose and impaired glucose tolerance: Implications for care," in *Diabetes Care*, 2007, vol. 30, no. 3, pp. 753–759.
151. T. O. Krag, S. Hauerslev, T. D. Jeppesen, M. Duno, and J. Vissing, "Muscle regeneration in mitochondrial myopathies," *Mitochondrion*, vol. 13, no. 2, pp. 63–70, Mar. 2013.
152. P. F. Chinnery, *Mitochondrial Disorders Overview*. University of Washington, Seattle, 1993.
153. R. A. Berk, *Statistical Learning from a Regression Perspective*. Springer New York, 2008.
154. P. F. Chinnery, N. Howell, R. N. Lightowlers, and D. M. Turnbull, "Molecular pathology of MELAS and MERRF. The relationship between mutation load and clinical phenotypes," *Brain*, vol. 120, no. 10, pp. 1713–1721, Oct. 1997.
155. Y. S. Ng et al., "Sudden adult death syndrome in m.3243A>G-related mitochondrial disease: an unrecognized clinical entity in young, asymptomatic adults.," *Eur. Heart J.*, vol. 37, no. 32, pp. 2552–9, Aug. 2016.
156. R. Little and D. Rubin, *Statistical Analysis with Missing Data*, vol. 26, no. 3. Wiley, 2019.
157. L. Massaron and A. Boschetti, *Regression Analysis with Python*. Packt Publishing, 2016.
158. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to Statistical Learning*, vol. 7, no. 10. 2000.
159. W. Greene, *Econometric Analysis*, 8th Ed. Pearson, 2018.
160. C. B. Hansen, "Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects," *J. Econom.*, vol. 140, no. 2, pp. 670–694, Oct. 2007.
161. J. Unpingco, *Python for Probability, Statistics, and Machine Learning*. Springer International Publishing, 2019.

162. H. JO., "Biochemical adaptations in muscle. Effects of exercise on mitochondrial oxygen uptake and respiratory enzyme activity in skeletal muscle," *J Biol Chem.*, vol. 242, no. 9, pp. 2278–2282, 1967.
163. J. L. Steiner, E. A. Murphy, J. L. McClellan, M. D. Carmichael, and J. M. Davis, "Exercise training increases mitochondrial biogenesis in the brain," *J. Appl. Physiol.*, vol. 111, no. 4, pp. 1066–1071, Oct. 2011.
164. L. Tilokani, S. Nagashima, V. Paupe, and J. Prudent, "Mitochondrial dynamics: Overview of molecular mechanisms," *Essays in Biochemistry*, vol. 62, no. 3. Portland Press Ltd, pp. 341–360, 20-Jul-2018.
165. N. A. Khan, P. Govindaraj, A. K. Meena, and K. Thangaraj, "Mitochondrial disorders: Challenges in diagnosis & treatment," *Indian Journal of Medical Research, Supplement*, vol. 141, no. JAN 2015. Indian Council of Medical Research, pp. 13–26, 2015.
166. K. Maeda et al., "Clinical Phenotype and Segregation of Mitochondrial 3243A>G Mutation in 2 Pairs of Monozygotic Twins," *JAMA Neurol.*, vol. 73, no. 8, p. 990, Aug. 2016.
167. M. Li, R. Schröder, S. Ni, B. Madea, and M. Stoneking, "Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 8, pp. 2491–2496, Feb. 2015.
168. Y. He et al., "Heteroplasmic mitochondrial DNA mutations in normal and tumour cells," *Nature*, vol. 464, no. 7288, pp. 610–614, Mar. 2010.
169. P. Alberts, Bruce; Johnson, Alexander D.; Lewis, Julian; Morgan, David; Raff, Martin; Roberts, Keith; Walter, *Molecular Biology of the Cell*, 6th ed. W. W. Norton & Company, 2014.
170. B. Westermann, "Bioenergetic role of mitochondrial fusion and fission," in *Biochimica et Biophysica Acta - Bioenergetics*, 2012, vol. 1817, no. 10, pp. 1833–1838.
171. P. F. Chinnery and D. C. Samuels, "Relaxed replication of mtDNA: A model with implications for the expression of disease," *Am. J. Hum. Genet.*, vol. 64, no. 4, pp. 1158–1165, Apr. 1999.
172. D. C. Wallace, "Mitochondrial DNA mutations in diseases of energy metabolism," *J. Bioenerg. Biomembr.*, vol. 26, no. 3, pp. 241–250, Jun. 1994.
173. J. L. Elson, D. C. Samuels, D. M. Turnbull, and P. F. Chinnery, "Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age," *Am. J. Hum. Genet.*, vol. 68, no. 3, pp. 802–806, 2001.

174. M. V. Paz et al., "The Role of Autophagy and Mitophagy in Mitochondrial Diseases," in *Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging*, Elsevier, 2016, pp. 155–172.
175. K. J. Krishnan et al., "What causes mitochondrial DNA deletions in human cells?," *Nature Genetics*, vol. 40, no. 3. Nature Publishing Group, pp. 275–279, 27-Mar-2008.
176. S. Rahman, J. Poulton, D. Marchington, and A. Suomalainen, "Decrease of 3243 A → G mtDNA mutation from blood in MELAS syndrome: A longitudinal study," *Am. J. Hum. Genet.*, vol. 68, no. 1, pp. 238–240, 2001.
177. R. G. Whittaker et al., "Urine heteroplasmy is the best predictor of clinical outcome in the m.3243a>g MTDNA mutation," *Neurology*, vol. 72, no. 6, pp. 568–569, Feb. 2009.
178. M. Mancuso et al., "The m.3243A>G mitochondrial DNA mutation and related phenotypes. A matter of gender?," *J. Neurol.*, vol. 261, no. 3, pp. 504–510, Dec. 2014.
179. A. Fayssol et al., "Prediction of long-term prognosis by heteroplasmy levels of the m.3243A>G mutation in patients with the mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes syndrome," *Eur. J. Neurol.*, vol. 24, no. 2, pp. 255–261, Feb. 2017.
180. J. P. Grady et al., "mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease," *EMBO Mol. Med.*, vol. 10, no. 6, Jun. 2018.
181. C. M. Sue et al., "Detection of MELAS A3243G point mutation in muscle, blood and hair follicles," *J. Neurol. Sci.*, vol. 161, no. 1, pp. 36–39, Nov. 1998.
182. H. Kondo, Y. Fujita, Y. Mizuno, M. Kihara, and K. Murayama, "Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes with severe systemic symptoms: Pathology and biochemistry," *Pediatr. Int.*, vol. 60, no. 3, pp. 300–302, Mar. 2018.
183. A. J. M. Janssen et al., "Muscle 3243A → G mutation load and capacity of the mitochondrial energy-generating system," *Ann. Neurol.*, vol. 63, no. 4, pp. 473–481, Apr. 2008.
184. S. R. Hammans, M. G. Sweeney, M. G. Hanna, M. Brockington, J. A. Morgan-Hughes, and A. E. Harding, "The mitochondrial DNA transfer RNA Leu (UUR) A → G(3243) mutation: A clinical and genetic study," *Brain*, vol. 118, no. 3, pp. 721–734, Jun. 1995.

185. S. Rahman, J. Poulton, D. Marchington, and A. Suomalainen, "Decrease of 3243 A → G mtDNA mutation from blood in MELAS syndrome: A longitudinal study," *Am. J. Hum. Genet.*, vol. 68, no. 1, pp. 238–240, 2001.
186. A. M. Schaefer, C. Phoenix, J. L. Elson, R. McFarland, P. F. Chinnery, and D. M. Turnbull, "Mitochondrial disease in adults: A scale to monitor progression and treatment," *Neurology*, vol. 66, no. 12, pp. 1932–1934, Jun. 2006.



İNTİHAL RAPORU İLK SAYFASI

OKSİDATİF FOSFORİLASYON SİSTEMİ HASTALIKLARINDA MİTOKONDRIYAL DNA VARYANTLARININ VERİ MADENCİLİĞİ İLE ARAŞTIRILMASI

ORIJINALLIK RAPORU

%3	%3	%1	%1
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	webpages.uncc.edu İnternet Kaynağı	<%1
2	Submitted to Yeditepe University Öğrenci Ödevi	<%1
3	Submitted to Haliç Üniversitesi Öğrenci Ödevi	<%1
4	Submitted to Istanbul University Öğrenci Ödevi	<%1
5	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	<%1
6	www.scribd.com İnternet Kaynağı	<%1
7	docplayer.net İnternet Kaynağı	<%1
8	www.scopemed.org	