

T.C.
İSTANBUL GEDİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



İNSAN KEMİK İLİĞİ HEMATOPOETİK KÖK HÜCRELERİNİN
YAŞA BAĞLI İNCELENMESİ VE KNN İLE
SINIFLANDIRILMASI

YÜKSEK LİSANS TEZİ

Muhammet Burak ŞİMŞEK

Yapay Zekâ Mühendisliği Anabilim Dalı

Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı

EYLÜL 2023
İSTANBUL

**T.C.
İSTANBUL GEDİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**



**İNSAN KEMİK İLİĞİ HEMATOPOETİK KÖK HÜCRELERİNİN YAŞA
BAĞLI İNCELENMESİ VE KNN İLE SINIFLANDIRILMASI**

YÜKSEK LİSANS TEZİ

Muhammet Burak ŞİMŞEK
210039007
0000-0003-3829-5572

Yapay Zekâ Mühendisliği Anabilim Dalı

Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Şerife Esra DİNÇER

İstanbul 2023



T.C.
İSTANBUL GEDİK ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü Müdürlüğü

Jüri Tez Onay Formu

07.09.2023

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ

Bu çalışma 07.09 2023 tarihinde aşağıdaki jüri tarafından Yapay Zekâ Mühendisliği Anabilim Dalı, Yapay Zekâ Mühendisliği (Tezli Yüksek Lisans) Programı Yüksek Lisans Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Dr. Öğr. Üyesi Şerife Esra DİNÇER

Danışman

İstanbul Gedik Üniversitesi

Dr. Öğr. Üyesi Feridun Cemal

ÖZÇAKIR.

Üye (İmza)

İstanbul Gedik Üniversitesi

Dr. Öğr. Üyesi Ümit ÖZTÜRK.

Üye (İmza)

İstanbul Beykent Üniversitesi

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “İnsan Kemik İliği Hematopoetik Kök Hücrelerinin Yaşa Bağlı İncelenmesi ve KNN ile Sınıflandırılması” adlı çalışmanın, tezim proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim.

Muhammet Burak ŞİMŞEK / (07/09/2023).



ÖNSÖZ

Biyoinformatik var olan sağlık bilişiminin en önemli veri tipi olan genetik materyallerin anlaşılması için multidisipliner gelişimini sürdürerek önemini günden güne artırmaktadır. Bu konuya ait en önemli mihenk taşları son 60 yılda karşımıza sıklıkla çıkmaktadır. Adından fazlaca bahsedilen İnsan Genom Projesi ve benzer çalışmalar ile bir dönüm noktası sağlanan bu disiplinde insana ait tüm genetik bilginin derlenmesi sağlanmıştır. Bugün gelinen nokta itibariyle analiz çalışmaları için veri kaynağı adına son derece kritik rol oynamaktadır. Söz konusu verilerin tam olarak anlaşılması ve incelenmesi, insan yaşamındaki süreçlerin işleyişinin anlaşılması yarına ışık tutacaktır.

Biyoinformatiğin önemli araştırma konusu olan gen analizinin istatistiksel yaklaşımlarla incelenmesi ve süreçlerin biyoistatistiksel olarak işleyişini yürüttüğümüz tez çalışmasının bu konuda çalışan araştırmacılar için yol gösterici ve fayda sağlayacağını umuyorum.

Tez çalışmam boyunca katkılarıyla destek olan Dr. Öğr. Üyesi Şerife Esra DİNÇER'e teşekkürlerimi sunarım. Tüm çalışmam boyunca yardımları ve destekleri ile yanımda olan sevgili aileme de sonsuz teşekkürlerimi iletiyorum.

EYLÜL 2023

Muhammet Burak ŞİMŞEK

İÇİNDEKİLER

YEMİN METNİ	İV
ÖNSÖZ.....	V
İÇİNDEKİLER	VII
ŞEKİL LİSTESİ.....	VIII
ÇİZELGE LİSTESİ.....	IX
SİMGELER VE KISALTMALAR	X
ÖZET.....	XII
ABSTRACT	XIII
1. GİRİŞ	133
2. LİTERATÜR İNCELEMESİ	155
2.1 İstatistiksel Yöntemlerle Biyolojik Verilerin İncelenmesi	155
2.1.1 Tek Yönlü Varyans Analizi.....	155
2.1.2 Tek Örneklemli T-Testi.....	166
2.1.3 İki Örneklemli T-Testi	166
2.1.4 Basit Doğrusal Regresyon Testi	177
2.1.5 Çoklu Regresyon Testi	177
2.1.6 Ki-Kare Testi.....	188
2.1.7 Mann-Whitney U-Testi	188
2.1.8 Kruskal-Wallis Testi	19
2.1.9 Lineer Regresyon Algoritması	200
2.1.10 Karar Ağacı Algoritması	200
2.1.11 K-en Yakın Komşu (KNN) Algoritması	211
2.1.12 Destek Vektör Makinesi (DVM) Algoritması	222
2.2 Hematopoetik Kök Hücrelerin Biyolojik Olarak İncelenmesi	233
2.2.1 Kemik İliği	233
2.2.2 Kök Hücreler.....	244
2.2.3 Kanseri Hücreleri.....	277
2.2.4 Lösemi	288
2.2.4.1 Akut Miyeloid Lösemi	288
2.2.5 Hematopoez.....	29
3. MATERYAL VE YÖNTEM.....	300
4. BULGULAR VE TARTIŞMA	400
4.1 Bulgular.....	400
4.2 Miyeloid Özellikli Gen Analizi	411
4.3 Lenfoid Özellikli Gen Analizi	422
4.4 GOrilla Analizi	422
4.5 Makine Öğrenmesi Çıktıları ve Gen İfadelerinin Analizi	466
4.6 Tartışma.....	644
5. SONUÇLAR VE ÖNERİLER	666
KAYNAKLAR	688
ÖZGEÇMİŞ.....	744

ŞEKİL LİSTESİ

Şekil 2.1 : Doğrusal Regresyonun Görsel İfadesi	20
Şekil 2.2 : Karar Ağacı Yapısı	21
Şekil 2.3 : KNN Model Temsili	22
Şekil 2.4 : Destek Vektör Makinesi Model Temsili	23
Şekil 2.5 : Hematopoetik ve Stromal Hücre Farklılaşması	25
Şekil 2.6 : Kan Hücresinin Bileşenleri	27
Şekil 2.7 : Hematopoez Aşamaları	29
Şekil 3.1 : İnsan Gen Ekspresyon Değerlerinin İstatistiksel Dağılımı	31
Şekil 3.2 : KNN Uygulaması	38
Şekil 4.1 : $p < 0.05$ Anlamlı Gen Ekspresyon Değerlerinin Kutu Grafiği.....	40
Şekil 4.2 : Miyeloid Yanlılık Gösteren Genlerin Isı Haritası ve Kutu Grafiği.....	41
Şekil 4.3 : Lenfoid Yanlılık Gösteren Genlerin Isı Haritası ve Kutu Grafiği.....	42
Şekil 4.4 : GORilla Hedef Gen Dizisinde Gerçekleşen Biyolojik Süreçler	43
Şekil 4.5 : Çizelge 3.1'deki Genlerin Isı Haritası.....	45
Şekil 4.6 : Hata Matrisi.....	457
Şekil 4.7 : ABCC1 Geninin Ekspresyon Değerlerinin Grafikleri.....	48
Şekil 4.8 : ABCB1 Geninin Ekspresyon Değerlerinin Grafikleri.....	48
Şekil 4.9 : PRKAA2 Geninin Ekspresyon Değerlerinin Grafikleri.....	49
Şekil 4.10 : SIRPA Geninin Ekspresyon Değerlerinin Grafikleri	49
Şekil 4.11 : DDIT4 Geninin Ekspresyon Değerlerinin Grafikleri.....	50
Şekil 4.12 : TNIP1 Geninin Ekspresyon Değerlerinin Grafikleri.....	50
Şekil 4.13 : DUSP6 Geninin Ekspresyon Değerlerinin Grafikleri	51
Şekil 4.14 : GGNBP2 Geninin Ekspresyon Değerlerinin Grafikleri.....	51
Şekil 4.15 : MECOM Geninin Ekspresyon Değerlerinin Grafikleri	52
Şekil 4.16 : STAT1 Geninin Ekspresyon Değerlerinin Grafikleri	52
Şekil 4.17 : EZR Geninin Ekspresyon Değerlerinin Grafikleri.....	53
Şekil 4.18 : WWC1 Geninin Ekspresyon Değerlerinin Grafikleri	53
Şekil 4.19 : KANK1 Geninin Ekspresyon Değerlerinin Grafikleri.....	54
Şekil 4.20 : MYADM Geninin Ekspresyon Değerlerinin Grafikleri.....	54
Şekil 4.21 : RGS3 Geninin Ekspresyon Değerlerinin Grafikleri.....	55
Şekil 4.22 : DUSP1 Gen Ekspresyon Değerlerinin Grafikleri	55
Şekil 4.23 : TANK Geninin Ekspresyon Değerlerinin Grafikleri	56

Şekil 4.24 : RANBP9 Geninin Ekspresyon Değerlerinin Grafikleri	56
Şekil 4.25 : FNIP1 Geninin Ekspresyon Değerlerinin Grafikleri	57
Şekil 4.26 : SERPINE1 Geninin Ekspresyon Değerlerinin Grafikleri	57
Şekil 4.27 : XIAP Geninin Ekspresyon Değerlerinin Grafikleri	58
Şekil 4.28 : TNRC6A Geninin Ekspresyon Değerlerinin Grafikleri	58
Şekil 4.29 : HEXIM1 Geninin Ekspresyon Değerlerinin Grafikleri	59
Şekil 4.30 : NCOA7 Geninin Ekspresyon Değerlerinin Grafikleri	59
Şekil 4.31 : PPME1 Geninin Ekspresyon Değerlerinin Grafikleri	60
Şekil 4.32 : GPC3 Geninin Ekspresyon Değerlerinin Grafikleri	60
Şekil 4.33 : HEXIM2 Geninin Ekspresyon Değerlerinin Grafikleri	61
Şekil 4.34 : PIP5KL1 Geninin Ekspresyon Değerlerinin Grafikleri	61
Şekil 4.35 : CAPRIN1 Geninin Ekspresyon Değerlerinin Grafikleri	62
Şekil 4.36 : GAPDH Geninin Ekspresyon Değerlerinin Grafikleri	62
Şekil 4.37 : CBEP3 Geninin Ekspresyon Değerlerinin Grafikleri	63
Şekil 4.38 : KNG1 Geninin Ekspresyon Değerlerinin Grafikleri	63
Şekil 4.39 : Örneklemindeki Genlerin Isı Haritası	64

ÇİZELGE LİSTESİ

Çizelge 3.1 : Analiz Sonucu Elde Edilen Genler, ID ve Açıklama Bilgilerinin Bazıları.....	36
Çizelge 3.2 : Makine Öğrenmesi Algoritmasında Kullanılan Veri Setinin İlk 5 Satırı.....	37
Çizelge 4.1 : Şekil 9'daki Biyolojik Süreçlerde Ortaya Çıkan Genler.....	44
Çizelge 4.2 : Hata Matrisi Metrikleri ve Elde Edilen Değerler	47



SİMGELER VE KISALTMALAR

HKH	: Hematopoetik Kök Hücre
KNN	: K-Nearest Neighbor (K-En Yakın Komşu)
DVM	: Destek Vektör Makinesi
ANOVA	: Analysis Of Variance (Varyans Analizi)
AML	: Akut Miyeloid Lösemi
CMP	: Ortak Miyeloid Öncül Hücre
CLP	: Ortak Lenfoid Öncül Hücre
GMP	: Granüloist Makrofaj Öncül Hücre
MEP	: Megakaryosit- Eritrosit Öncül Hücre

ÖZET

İNSAN KEMİK İLİĞİ HEMATOPOETİK KÖK HÜCRELERİNİN YAŞA BAĞLI İNCELENMESİ VE KNN İLE SINIFLANDIRILMASI

İnsanlardaki hematopoetik sistemde yaşa bağlı bir değişim ve bu değişime bağlı olarak kök hücrelerde gen ekspresyon regülasyonu gözlemlenmektedir. Kök hücrelerin bilimsel olarak incelenmesi istatistiksel yöntemler ve çeşitli algoritmalar kullanılarak yapılmaktadır. Yaşlı insan hematopoetik sisteminin özelliklerinin anlaşılması ve değerlendirilmesi programlama ve algoritmalar ile sağlanmaktadır. Bu araştırmada yaş artışı ile birlikte hematopoetik kök hücre (HKH) ekspresyonlarının farklı yaş gruplarında hangi seviyelerde görüldüğü istatistiksel hipotez testi ile incelenmiştir. Araştırma kapsamında yaşlı insan HKH'si miyeloid yanlılığın görüldüğünün ortaya çıkarılması amaçlanmaktadır.

Bu çalışma sürecinde yaşlı, orta yaşlı ve gençlerin oluşturduğu 27 insan örneğine ait 50.000 kemik iliği kök hücre örneği kullanılmıştır. Yaşa bağlı olabilecek hematopoetik fonksiyon bozukluğu göstermeye yatkın hematopoetik kök hücreler ve diğer progenitör popülasyonlar NCBI veri tabanından sağlanan veri seti, R programlama ile analiz edilmiştir. Yaşa bağlı insan hematopoetik kök hücrelerinde ekspresyon seviyelerinin değişimi ANOVA ile test edilmiştir. Yaş grupları arasında ekspresyon seviyesinde istatistiksel olarak anlamlı bir farkın olduğu ($p < 0,05$) saptanmıştır. İstatistiksel olarak anlamlı örneklem üzerinde, Benjamini Hochberg yöntemi ve GOrilla analiziyle veri seti zenginleştirilmiştir. Ekspresyon seviyesinde orta yaş grubuna ait farklı regülasyon durumu miyeloid ve lenfoid yanlılık gösteren hücre popülasyonlarında K-en yakın komşu algoritması kullanılarak sınıflandırılmıştır.

Örnekleme ait gen ekspresyon ifadelerinin, orta yaşlı bireylerde genç popülasyona göre miyeloid yanlılık gösterdiği ve lenfoid yönelimlerinin daha zayıf kaldığı tespit edilmiştir. Makine öğrenmesi algoritması ile regülasyon durumlarına göre sınıflandırılan hematopoetik kök hücrelerinin orta yaşlı gruba ait ekspresyon seviyesinde istatistiksel olarak anlamlı bir farkın olduğu tespit edilmiştir.

Anahtar Kelimeler: ANOVA, makine öğrenmesi, KNN, miyeloid, lenfoid, ekspresyon seviyesi

ABSTRACT

AGE-DEPENDENT EXAMINATION OF HUMAN BONE MARROW HEMATOPOETIC STEM CELLS AND CLASSIFICATION WITH KNN

In the haematopoietic system in humans, an age-related change and gene expression regulation in stem cells is observed. Stem cells are scientifically analysed using statistical methods and various algorithms. Understanding and evaluation of the properties of the old human haematopoietic system is provided by programming and algorithms. In this study, the levels of hematopoietic stem cell (HSC) expressions in different age groups with increasing age were examined by statistical hypothesis testing. Within the scope of the research, it is aimed to reveal that myeloid bias is seen in elderly human HSC.

In this study, 50,000 bone marrow stem cell samples from 27 human samples consisting of elderly, middle-aged and young people were used. Hematopoietic stem cells and other progenitor populations predisposed to age-related hematopoietic dysfunction were analysed using R programming. Age-related changes in expression levels in human hematopoietic stem cells were tested by ANOVA. It was found that there was a statistically significant difference in expression levels between age groups ($p < 0.05$). The data set was enriched by Benjamini Hochberg method and GOrilla analysis on the statistically significant sample. The differential regulation of expression level in the middle-aged group was classified using the K-nearest neighbour algorithm in myeloid and lymphoid biased cell populations.

It was determined that the gene expression of the sample showed a myeloid bias in middle-aged individuals compared to the young population and their lymphoid orientation remained weaker. It was determined that there was a statistically significant difference in the expression level of hematopoietic stem cells classified according to their regulatory status by machine learning algorithm.

Keywords: ANOVA, machine learning, KNN, myeloid, lymphoid, expression level

1. GİRİŞ

Veri analizi, sınırsız uygulamaları ve verileri karar vermeye uygun bir şekilde çıkarmaya yönelik yaklaşımları nedeniyle farklı araştırma alanlarında popülerlik kazanmaktadır (Ahmad ve diğ., 2015). Biyolojik verilerin büyük bir hızla çoğaldığı günümüzde, bu verileri işleme ve kullanıma sunma ihtiyacı bilgi teknolojilerine dayalı disiplinler arası bilimlerin doğmasına neden olmuştur. Biyoinformatik, biyolojik bilgilerin yönetimi ve yorumlanmasına odaklanır. Biyolojik verilerin efektif şekilde kullanımı için çeşitli algoritmalar ve biyoistatistik yöntemler kullanarak büyük veri kümeleri arasındaki ilişkilerin açıklanmasını amaçlamaktadır. Büyük ölçekli biyolojik verilerin yönetimini ve analizini amaçlayan biyoinformatik, istatistiksel yöntem ve veri analizi metotlarını uygulamaktadır (Attwood ve Parry-Smith, 1999).

Modern istatistik, nicel verilerin toplanmasını, analiz edilmesini, yorumlanmasını içeren ve istatistiksel tahminleri konu eden bir bilimdir. Disiplinler arası varlığı ve neredeyse tüm bilimsel çalışmalarda uygulanabilirliği alanında yapılan çalışmaları geliştirmektedir (Bartholomew, 1995). İstatistiksel tahmin, hipotez testleri ve parametrelerin tahmini olmak üzere iki problemle ilgilenir. Örneklem istatistiklerinden popülasyon parametrelerini tahmin etme sürecidir. Çok sayıda olay hakkında belirli bir oran gözlemlenerek nasıl sonuçlar çıkarılabileceğini ve elde edilen sonuçları formüle şekilde standartlaştırmak için teknikler sunmaktadır (Altman ve Bland, 1999).

İstatistik terimi genel olarak hem tanımlayıcı hem de analitik alanları içermektedir. Tıbbi ve insan sağlığı araştırma çalışmalarının tasarımı ve analizinde özellikle yararlı ve yaygın bulunan bu analitik yöntemlere biyoistatistik adı verilmiştir. Biyolojik fenomenlerin analizine uygulanan istatistiksel süreçler ve yöntemlerle ilgilenir. İstatistiksel yöntemlerle sonuçları anlamlı ve kullanışlı şekilde sunar ve karmaşık veri yapılarının altındaki nedenlerin analiz edilerek tahminler yapılmasına olanak sağlar (Weissgerber ve diğ., 2016).

Canlı türlerinin tanımlanması, sınıflandırılması ve türlerin bilimsel olarak farklılaşması istatistiksel teknikler kullanılarak yapılabilmektedir. Hücrelerin işlevleri, fonksiyonları ve ekspresyon seviyeleri birbirlerinden farklı olduğu için bu farklılıklar her bir değişken ve ölçüm için anormal eşikleri belirlemek üzere istatistiksel yöntemler kullanılarak belirlenmektedir. (Sümbüloğlu, 2007).

Hematopoetik kök hücreler yaşam boyunca tüm kan hücrelerinin dengeli bir şekilde üretilmesini sağlar. Yaşlandıkça, HKH'ler kademeli olarak kendi kendini yenileme ve rejeneratif potansiyellerini kaybederken, buna bağlı olarak hücresel problemler ortaya çıkmaya başlamaktadır. Birçok hematolojik patoloji güçlü bir şekilde yaşla ilişkili olduğundan, kök hücre yaşlanma sürecinin araştırılma süreçleri büyük öneme sahiptir (de Haan ve Lazare, 2018). Bu çalışmada insan hematopoetik kök hücrelerinin hücre bazlı regülasyon durumları, istatistiksel testler ve yapay zekâ metodolojisiyle incelenmiştir.

Literatürde ifade edilen yaşın ilerlemesiyle birlikte artan miyeloid yanlılığın miyeloid malignitelere sebep olduğu, yapılan istatistiksel analizler ile ortaya koyularak yerli literatüre katkı sağlanmak istenmiştir. Ayrıca makine öğrenmesi algoritmasıyla miyeloid ve lenfoid hücreleri için sınıflandırma yapan bir tahmin modeli ile yapay zekâ kapsamında özgün bir araştırma yapılması istenmiştir.

Yaşlı insan hematopoetik sisteminin özelliklerinin anlaşılması ve değerlendirilmesi için hemotolojik olarak sağlıklı genç, orta yaşlı ve yaşlı bireylerin kemik iliği örneklerinden HKH ve belirli bazı hematopoetik progenitor popülasyonları değerlendirilmiştir. Bu çalışmada yaş artışı ile birlikte HKH ekspresyonlarının farklı yaş gruplarında hangi seviyelerde görüldüğü incelenmiştir. Yaşlı insan HKH'si miyeloid yanlılık gösterirken, genç insan HKH'sinde bu durumun miyeloid ve lenfoidde dengeli bir seviyede görüldüğünün ortaya çıkarılması amaçlanmaktadır (Bryder ve diğ., 2006).

2. LİTERATÜR İNCELEMESİ

2.1 İstatistiksel Yöntemlerle Biyolojik Verilerin İncelenmesi

Bir araştırma sürecinin doğru bir şekilde yürütülmesi için amaca ve verilere uygun istatistiksel bir test seçilmesi gerekmektedir. Test seçimini belirleyen en önemli parametreler; hipotezin türü, değişkenlerin bağımlılık ilişkisi ve örneklemin normal dağılıma uygunluğu şeklinde değerlendirilmektedir. (Kul, 2014). Hipotez testlerine başlamadan önce bağımlı ve bağımsız değişkenler belirlenir. Bağımlı değişken, araştırmadaki diğer değişken veya değişkenlere bağlı olarak değişim gösterebilen değişkenlerdir. Bağımsız değişken araştırmada değeri başka değişken veya değişkenlere bağlı olarak değişim göstermeyen, bağımlı değişkenin alacağı değerleri belirleyen değişkenlerdir (Hazra ve Gogtay, 2016).

Test seçimini etkileyen diğer faktör örneklemin normallik durumudur. Normal dağılım, homojen varyans dağılımı ve bağımsız gözlemlerin görüldüğü durumlarda parametrik testler kullanılmaktadır. Ayrıca tüm istatistik testlerde kullanılan ‘p değeri’ adında bir olasılık hesaplanmaktadır. P değeri, istatistiksel anlamlılığın ispatını ve varsa bir farkın kanıt düzeyini belirlemek için kullanılır. Çalışmalarda, 0,05'ten küçük p değerinin istatistiksel olarak anlamlı olduğu kabul görmektedir (Kılıç, 2014).

2.1.1 Tek Yönlü Varyans Analizi

Tek yönlü varyans analizi (ANOVA), ikiden fazla bağımsız grubun ortalamaları arasında istatistiksel olarak anlamlı farklılıklar olup olmadığını belirlemek için kullanılmaktadır. ANOVA, farklı örneklerin araçlarını karşılaştırarak bir veya daha fazla faktörün etkisini kontrol eder. Tek yönlü varyans analizi, ilgilenilen gruplar arasındaki ilişkiyi karşılaştırır ve bu gruplardan herhangi birinin istatistiksel olarak diğerlerinden önemli ölçüde farklı olup olmadığını belirlemektedir (Connolly, 2021).

İstatistiksel olarak incelenen hipotez kavramı ANOVA için de geçerlidir. Sıfır hipotezi ve bir alternatif hipotez kullanılmaktadır. ANOVA'daki sıfır hipotezi, tüm örneklem ortalamaları arasında anlamlı bir fark olmadığında geçerlidir. Anlamlı fark

gözlemlendiği durumlarda alternatif hipotez kabul görmektedir (Vijayvargiya, 2009).

2.1.2 Tek Örneklemli T-Testi

Tek bir örneklem ait verilerden hesaplanan ortalamanın, daha önceden belirlenmiş sabit bir değer ile arasında istatistiksel olarak anlamlı bir farkın olup olmadığını belirlemek için kullanılan bir hipotez testidir. Bilimsel gerekçelerle belirlenen bu değer veri içerisinde alınmayan harici bir ifadedir. Genel olarak belirlenen değer, bir örneklem grubunda daha önceden belirlenmiş standart bir ifade veya diğer çalışma bulgularından elde edilmiş bir ortalama olabilir. Tüm hipotez testlerinde olduğu gibi, tek örneklemli t testi, boş hipotezi (H_0) alternatif hipotez ise (H_1) lehine yeterli kanıt olup olmadığını belirlemektedir (Kim, 2015).

Tek örneklem t testi, diğer istatistiksel hipotez testlerinden farklı olarak iki değişken arasındaki ilişkiyi ve iki ayrı grubu incelemeyiz. Popülasyona ait tek bir değişken için toplanan veriler ile önceden belirlenmiş sabit bir değer arasında karşılaştırma yapmaktadır (Ross ve Willson, 2017).

2.1.3 İki Örneklemli T-Testi

İki örneklemli t testi, iki farklı örneklem ortalamasının karşılaştırılmasıyla aralarında anlamlı bir fark olup olmadığını belirlemek için kullanılan istatistiksel bir hipotez testidir. Örneklem popülasyonlarının normal dağılım gösterdiği durumlarda kullanılmaktadır. Popülasyon standart sapmalarının bilinmesini veya örneklem büyüklüğünün büyük olmasını gerektirmeden, daha küçük örneklemle kullanılabilir. Bağımlı ve bağımsız örneklemli t testi olarak iki grupta incelenmektedir (Cressie ve Whitford, 1986).

Bağımsız iki örneklem t testinde karşılaştırılan iki gruptaki örnekler birbirleriyle ilgisizdir. İki farklı popülasyondan alınan örneklem grubu arasındaki fark, bu örneklemelere ait ortalama ve varyans değerleri kullanılarak hesaplanmaktadır. (Ross ve Willson, 2017).

Bağımlı iki örneklem t testinde karşılaştırılan iki örneklem grubu birbirleriyle ilişkilidir. Örneğin, öz nitelikler benzer konularda veya başka şekilde yakından ilişkilendirilen konulara ait alınan ölçüm çiftleri olabilir. Bu durumda, iki grubun ortalamaları arasındaki farklar alınarak, bu farklılıkların tek bir örneklem olarak değerlendirilmesi ile hesaplanmaktadır (Wilkerson, 2008).

2.1.4 Basit Doğrusal Regresyon Testi

Doğrusal regresyon, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan istatistiksel yöntemidir. Doğrusal regresyonda t-testi, yanıt değişkeni ile farklı değişkenler arasındaki korelasyonun doğrusallığına ilişkin hipotezi test etmek için kullanılan istatistiksel bir hipotez test etme tekniğidir (Eberly, 2007).

Birkaç değişken arasındaki ilişkiyi anlamlandırmak ve dağılım grafikleri ile değişken çiftleri arasında potansiyel ilişkileri keşfetmek amacıyla kullanılmaktadır. Regresyon analizinde öznitelikler arasındaki ilişki yapısı incelenerek hangi ilişkilerin istatistiksel olarak anlamlı olduğu ortaya koymaya çalışılmaktadır. Doğrusal regresyon katsayıları, her bir bağımsız değişken ile bağımlı değişken arasındaki matematiksel ilişkiyi tanımlamaktadır. Bağımsız değişken ile bağımlı değişken arasında bir korelasyon olmadığı doğrusal regresyondaki p değeri, sıfır hipotezini test etmektedir. Korelasyon bulunmadığı durumda, bağımsız değişkendeki değişimler ile bağımlı değişkenler arasında bir ilişki yoktur (Maulud ve Abdulazeez, 2020).

2.1.5 Çoklu Regresyon Testi

Çoklu regresyon analizi, basit doğrusal regresyonun bir uzantısıdır. Matematiksel hesaplama ve mantık açısından tek değişkenli regresyon analizi ile benzerlik gösterir. İki veya daha fazla değişkenin değerine dayanarak bir değişkenin değerini tahmin etmek istendiğinde kullanılır. Tahmin etmek istenilen değişkene bağımlı değişken, bağımlı değişkenin değerini tahmin etmek için kullanılan değişkenlere bağımsız değişken adı verilir (Tranmer ve Elliot, 2008).

Çoklu doğrusal regresyon, rastgele değişkenler arasındaki istatistiksel ilişkiyi açıklamak için kullanılmaktadır. Birden fazla bağımsız değişkenin bir bağımlı değişken ile arasındaki ilişkiliyi inceler. Kısaca bağımlı değişken ile birden fazla bağımsız değişken arasındaki korelasyon ilişkisini konu edinmektedir. Ayrıca her bir değişkenin veri setindeki genel durumu ve değişkenler arasındaki toplam varyansa etkisinin belirlenmesini sağlamaktadır. Bağımsız faktörler belirlendikten sonra, çoklu değişkenler ve sonuç değişkeni üzerindeki etki düzeyine ilişkin yapılan analizler bağımlı değişkeni tahmin etmek için kullanılabilir. (Stolzenberg, 2004).

2.1.6 Ki-Kare Testi

Ki-Kare testi bir modelin gerçek gözlemlenen ve beklenen verilerle karşılaştırılmasını ölçen bir testtir. Ki-kare testleri hipotezleri test etmek için, dolayısıyla örneklemin belirli bir beklenen sonuç ifadesi arasındaki ilişkiyi ifade etmek için kullanılmaktadır. Verilerdeki kategorik değişkenlerle ilişkili olup olmadığını belirlemek için de kullanılabilir. İki kategorik değişken arasındaki farkın rastgele mi yoksa aralarındaki ilişkiden mi kaynaklandığını tespit etmeye yardımcı olmaktadır (Ugoni ve Walker, 1995).

Kategorik bir değişkenin dağılımına ilişkin bir hipotezi test etmek için ki-kare testi veya karşılaştırılabilir parametrik olmayan bir test gereklidir. Ki-Kare istatistik testi, ham, rastgele, bağımsız değişkenlerden ve özellikle geniş bir örneklemden alınmış veriler temel alınarak hesaplanmaktadır. Gözlemlenen yanıt modeli, değişkenlerin gerçekten birbirinden bağımsız olması durumunda beklenen modelle karşılaştırılarak iki değişken arasında bir ilişkinin var olup olmadığı değerlendirilmektedir (Zibran, 2007).

2.1.7 Mann-Whitney U-Testi

Mann-Whitney U testi, iki bağımsız grubu karşılaştıran parametrik olmayan bir hipotez testidir. İki örneklem grubun medyanları arasındaki farkın istatistiksel olarak anlamlı olup olmadığını belirlemek için Mann-Whitney U testi kullanılmaktadır. Ancak medyanların aslında Mann-Whitney U test istatistiğinin hesaplanmasında bir karşılığı bulunmamaktadır. İki grup aynı medyana sahip olabilirken aynı zamanda Mann-Whitney U testine göre önemli ölçüde de farklılık gösterebilmektedir (Karadimitriou ve diğ., 2018).

T-testi gibi, bu analiz de iki bağımsız grubun benzer tipik sonuçlara sahip olup olmadığı test edilmektedir. Her iki test de bağımsız olarak örneklenmiş iki grup gerektirir ve iki grubun tek bir sürekli değişken üzerinde farklılık gösterip göstermediği değerlendirilmektedir. Bununla birlikte, iki test varsayılan dağılımda farklılık gösterir. Parametrik olmayan bir test normal bir dağılımı varsaymazken, parametrik bir test belirli bir normal dağılımı varsayar. Bu kapsamda, Mann-Whitney U, örneklemdaki iki grubun tek bir popülasyondan olup olmadığını belirlemek için kavramsal olarak t-testiyle benzerlik göstermektedir (Ruxton, 2006).

2.1.8 Kruskal-Wallis Testi

Kruskal-Wallis testi, ikiden fazla grup arasında bir bağımsız değişken üzerinden istatistiksel açıdan anlamlı farkın olup olmadığını belirlemek için kullanılan parametrik olmayan bir hipotez testidir. İkiden fazla grubu kontrol etmesi yönünden tek yönlü varyans analizinin parametrik olmayan alternatifi, ikiden fazla bağımsız grubun karşılaştırılmasına olanak sağlayan Mann-Whitney U testinin bir uzantısı olarak kabul edilmektedir (Cabral Júnior ve Lucena, 2020).

Normal dağılım görülmediği veri setinde iki ya da daha çok grubun analizinde kullanılmaktadır. Sıfır hipotezi (H_0), medyan değerinin örneklem grupları arasında eşit olduğu veya medyanlar arasındaki farkın sıfır olduğu şeklindedir. Alternatif hipotez (H_1) ise en az bir grubun popülasyon medyan değerinin, diğer gruplardan birinin medyanından farklı olmasıdır. Kruskal Wallis testi, ilgilenilen değişken sürekli olduğunda veya ayrık olduğunda daha çok kullanılmaktadır (Richardson, 2018).

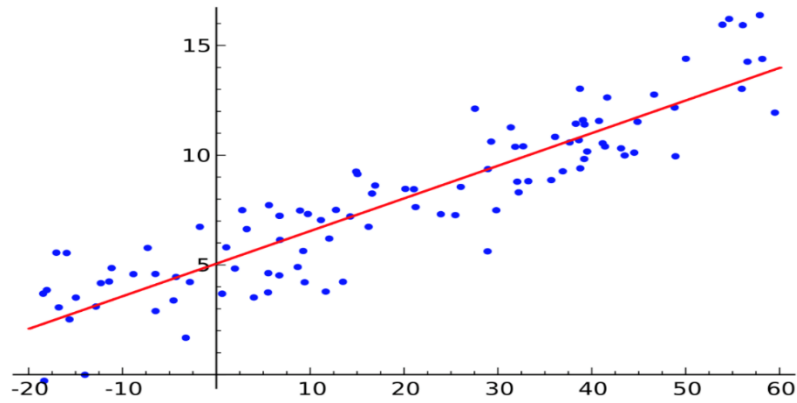
Birçok alanda karşılaşılan problemleri yanıtlamak için yapılan araştırmalarda verileri analiz ederek anlamlı bilgilere ulaşma süreci istatistiksel yöntemler kapsamında incelenmektedir. Makine öğrenmesi ise verilerin çıktı değerlerini tahmin etmek için geçmiş verileri girdi olarak kullanan algoritmalar. Etiketli verilere dayalı tahminler yapmak için bir modeli eğiten algoritmalar denetimli öğrenme modelini oluşturmaktadır (Kotsiantis ve diğ., 2007).

Denetimli öğrenmenin tanımlayıcı özelliği, açıklamalı eğitim verilerinin kullanılabilirliğidir. Makine öğrenmesi algoritmalarının pek çok araştırma alanından sorumludur ve birçok denetimli öğrenme tekniği, hücrelerin işlevlerinin incelenmesinde uygulama bulmuştur (Ayodele, 2010).

Denetimli öğrenme algoritmaları, eğitim verilerinden modeller oluşturur ve bu modeller, diğer etiketlenmemiş verileri sınıflandırmak için kullanılmaktadır. Bir dizi girdi değişkeni ile bir çıktı değişkeni arasındaki eşleşmeyi öğrenmeyi ve bu eşleşmeyi görünmeyen verilerin çıktılarını tahmin etmek için uygulamayı kapsamaktadır. Sınıflandırma problemlerini çözmeye yönelik araştırmalar, makine öğrenmesi ve veri madenciliği literatüründe önemli bir yer tutmaktadır (Cunningham ve diğ., 2008)

2.1.9 Lineer Regresyon Algoritması

Doğrusal regresyon, bir bağımlı değişken ile bir veya daha fazla bağımsız özellik arasındaki doğrusal ilişkiyi hesaplayan denetimli makine öğrenmesi algoritmasıdır. Algoritmanın mantığı, bağımlı değişkenlerin değerini tahmin etmek için bağımsız değişkenlere bağlı en iyi doğrusal fonksiyonu oluşturmaktır. Doğrusal fonksiyon, düz bir çizgi modeli kurar ve bağımlı ve bağımsız değişkenler arasındaki ilişkiyi temsil eder. Doğrunun eğimi, bağımsız değişkendeki değişim durumunda bağımlı değişkenin ne kadar değiştiğini gösterir (Nasteski, 2017).



Şekil 2.1 : Doğrusal Regresyonun Görsel İfadesi

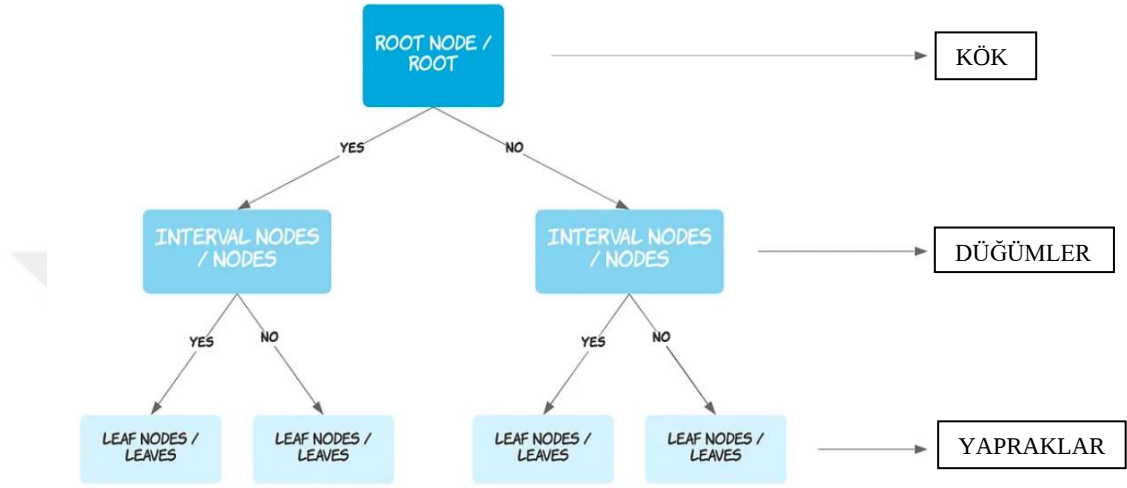
Kaynak: (Nasteski, 2017)

Doğrusal regresyon, belirli bir değişkenin davranışını anlamak ve tahmin etmek için kullanılan en önemli denetimli öğrenme algoritmalarından biridir. Regresyonda, bir işlevi öğrenmek için X ve Y değerlerine sahip bir veri kümesi bulunmaktadır. Bu bağlamda bilinmeyen bir X değişkeninden, Y değişkeni tahmin etmek isteniyorsa bu öğrenilen işlev kullanılmaktadır. Regresyonda Y değişkeninin değerini bulmayı hedefleyen bir fonksiyon gereklidir. Burada Y, bağımlı veya hedef değişkeni olarak adlandırılır ve X değişkeni, Y değişkeninin öngörücüsü olan bağımsız bir değişken olarak adlandırılmaktadır (Rong ve Bao-Wen, 2018).

2.1.10 Karar Ağacı Algoritması

Karar ağaçları, hem sınıflandırma hem de regresyon için kullanılabilen ağaca benzer temelli denetimli makine öğrenme algoritmasıdır. Karar ağacı, örnek uzayının öz yinelemeli bir bölümü olarak ifade edilen bir sınıflandırıcıyı temsil eder. Karar ağacı, kök ağacı oluşturan düğümlerden oluşur ve bu kök, herhangi bir girdisi olmayan temel bir düğüme sahip dağıtılmış bir ağaç yapısındadır. Diğer tüm düğümlerin tam

olarak bir girdi kenarı vardır (Song ve Ying, 2015). Çıktı kenarları olan düğüme iç düğüm veya test düğümü, geri kalan düğümlere de yaprak adı verilir. Bir karar ağacında, her test düğümü, girdi değerlerinin belirli bir ayrık işlevine göre örnek uzayını iki veya daha fazla alt uzaya böler. En basit durumda, her test tek bir özelliği dikkate alır ve böylece örnek alanı özelliğin değerine göre bölümlere ayrılmaktadır (Maimon ve Rokach, 2014).



Şekil 2.2 : Karar Ağacı Yapısı

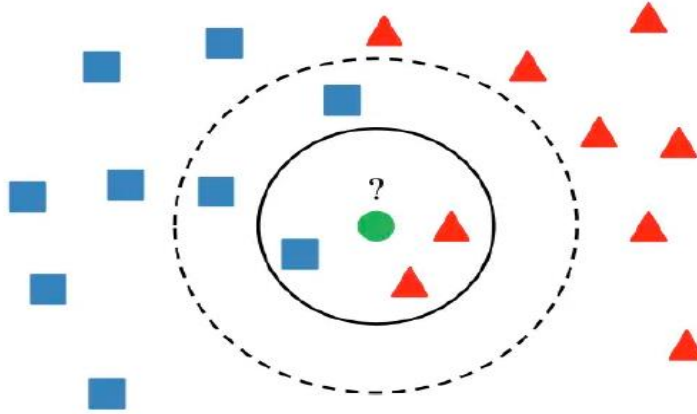
Kaynak: (Zhou ve diğ., 2021)

Her yaprak, en uygun hedef değeri temsil eden bir sınıfa atanır. Yaprak, hedef niteliğin belirli bir değere sahip olma olasılığını gösteren bir olasılık vektörü tutmaktadır. Örnekler, yol boyunca yapılan testlerin sonucuna göre ağacın kökünden yaprağa doğru yönlendirilerek sınıflandırılır. Her düğüm test ettiği öznitelikten etiketlenir ve dalları karşılık gelen değerlere göre şekillenmektedir. (Chauhan, 2014).

2.1.11 K-en Yakın Komşu (KNN) Algoritması

K-en yakın komşu algoritması, sınıflandırma ve regresyon problemlerini çözmek için kullanılan denetimli makine öğrenme algoritmasıdır. Veri noktalarının benzer etiketlere veya değerlere sahip olma eğiliminde olma durumlarını araştırmaktadır. Verinin eğitim aşamasında, KNN algoritması veri setinin belirli bir oranını eğitir. Tahminler yaparken, yeni gelen veri noktası ile tüm eğitim seti arasındaki ilişkiyi k kadar uzaklıktaki komşu mesafelerini kullanarak hesaplar. Algoritma, mesafelere göre girdi veri noktasına en yakın k komşuyu tanımlar. Sınıflandırma durumunda algoritma girdi veri noktası için tahmin edilen etiket olarak k komşuları arasında en

yaygın sınıfı atamaktadır (Zhang ve diğ., 2017).



Şekil 2.3 : KNN Model Temsili

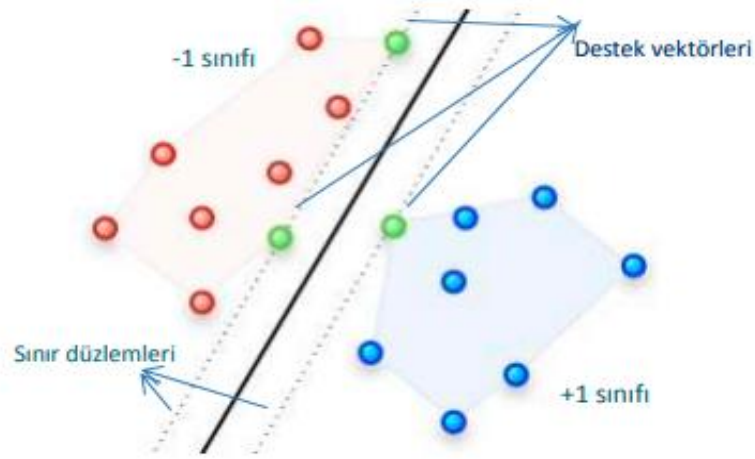
Kaynak: (Altunkaynak ve diğ., 2020)

KNN algoritması, nesneleri ve değer ifadelerini farklı özelliklere göre sınıflandırmak için bir örüntü tanıma görevi gerçekleştirmektedir. Bu özellik benzerliği yüksek doğruluk derecesinde sağlanabilmesi için k değerinin seçilmesi algoritmada kritik bir parametre ayarlamasıdır. K değerinin en iyi derecede seçilmesi çalışılan işe veya veri setine göre değişkenlik göstermektedir. K değeri yüksek doğruluk derecesine göre en optimal değere karşılık gelen komşu sayısını temsil eden ifade olarak seçilmektedir (Hassanat ve diğ., 2014).

2.1.12 Destek Vektör Makinesi (DVM) Algoritması

Destek vektör makinesi, doğrusal veya doğrusal olmayan sınıflandırma, regresyon ve aykırı değer algılama görevleri için kullanılan güçlü bir makine öğrenme algoritmasıdır. Destek vektör makinesi algoritmasının temel amacı, öznitelik uzayında farklı sınıflardaki veri noktalarını ayırabilen n boyutlu uzayda en optimal hiper düzlemi bulmaktır (Dridi, 2021).

DVM algoritması, farklı sınıf gruplarına ait en yakın örnekler arasındaki farkın maksimum olduğu bir hiper düzlem oluşturmaya çalışır. Hiper düzlemin boyutu, özniteliklerin sayısına bağlı olarak değişiklik gösterir. Girdi özelliklerinin sayısı iki ise, hiper düzlem sadece bir çizgidir. Giriş özelliklerinin sayısı üç ise, hiper düzlem 2 boyutlu bir düzlem olur. Özellik sayısı üçten fazla ise düzlem boyutu daha kompleks bir yapıda olmaktadır (Zhang, 2012).



Şekil 2.4 : Destek Vektör Makinesi Model Temsili

Kaynak: (Ayhan ve Erdoğan, 2014)

Sınıflandırma algoritmalarının amacı test edilecek verinin hangi sınıfa dahil edileceğine karar vermektir. Bu sınıflandırmayı yapmak için destek vektörleri çizilir ve -1, +1 sınıfları oluşturulur. Destek vektörleri arasında kalan bölgeye margin adı verilmektedir. Margin ne kadar büyükse sınıfların iyi ayrıştığı yorumu yapılmaktadır (Keerthi ve Gilbert, 2002).

2.2 Hematopoetik Kök Hücrelerin Biyolojik Olarak İncelenmesi

2.2.1 Kemik İliği

Kemik iliği olgunlaşmamış kan hücrelerinde hücre kümesini çevreleyen dokularla bağlantılı temel bir yerdir. İnsan kafatası, femurlar, kaburgalar, pelvis ve göğüs kafesi gibi kemiklerin içinde bulunan süngerimsi yağ dokusudur. Kemik iliği beslenir ve dolaşıma bağlanmaktadır. Kemik iliğinin beslenmesi sinüzoid olarak adlandırılan özelleştirilmiş fenestra kapiller ve uzmanlaşmamış kan damarları aracılığıyla gerçekleşmektedir. Bu sinüzoidler, hücre dışı/ekstraselüler matriks (ECM)' e nüfuz etmektedir. ECM'ler retiküler fibroblastlar tarafından üretilen sünger benzeri matrislerdir. Bu protein ve hücreler hematopoetik hücreleri ayrı bölmelere yerleştirmektedir (Bain ve diğ., 2019).

Kemik iliği hematopoetik hücreler, bidotelyal hücreler, stromal hücreler, ECM, sitokinler, büyüme faktörleri, kemokinleri içeren özel mikro çevreleri bulundurmaktadır. Bu mikro çevre kapsamında gelişen hematopoetik kök hücreler hayatta kalmak, çoğalmak ve farklılaşmak için sinyal almaktadır. Kemik iliği mikro

çevresi lösemi ile birlikte diğer kanser türlerinin de gelişiminde ve ilerlemesinde önemli bir rol oynamaktadır (Karadağ ve diğ., 2007).

2.2.2 Kök Hücreler

Kök hücreler; canlı vücudunda uzun süre bölünebilme, kendini yenileyebilme ve vücudun ihtiyacına göre farklılaşarak diğer doku hücrelerine dönüşebilme özelliklerine sahip hücreler olarak adlandırılmaktadır (İnan ve Özbilgin, 2009). Bir hücrenin kök hücre olarak tanımlanabilmesi için bilim insanları tarafından beş ölçüt belirlenmiştir. Bu ölçütler aşağıdaki gibidir.

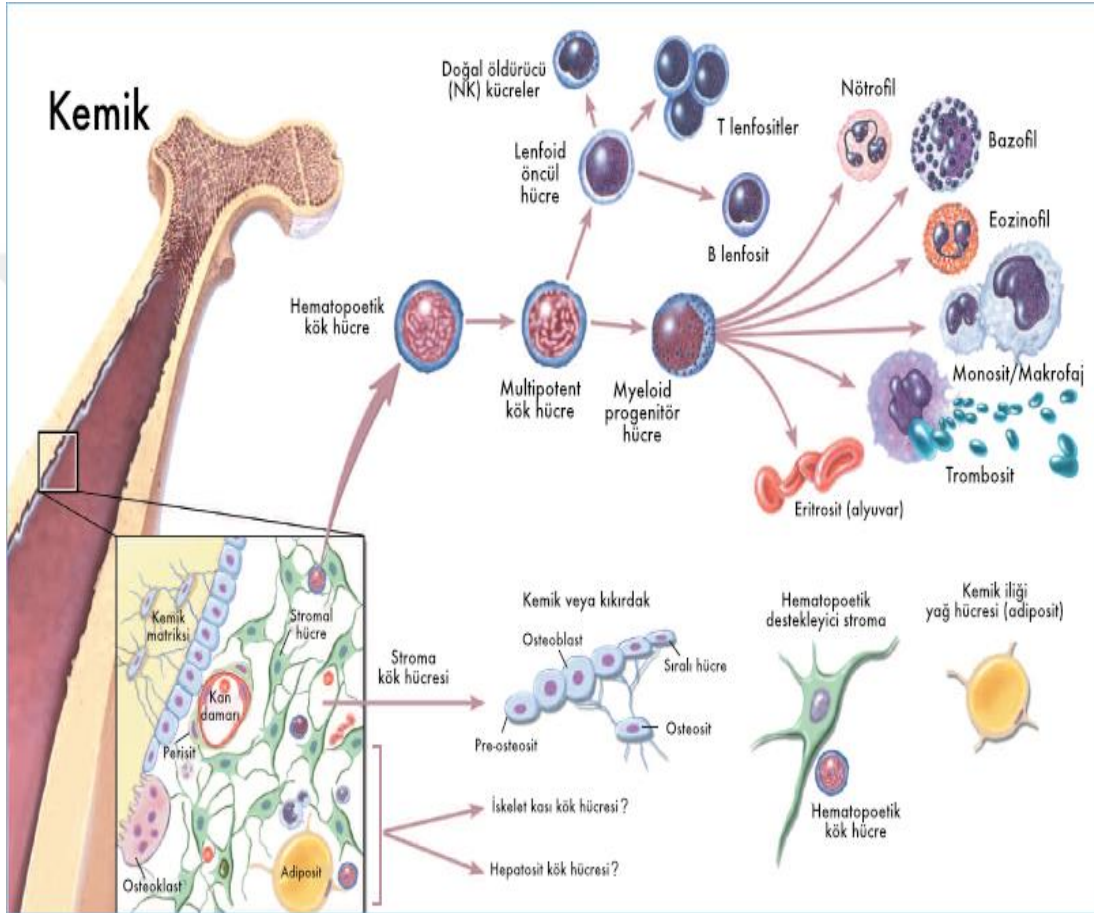
1. Kök hücreler, kendilerini yenileyebilme ve sürekli olarak bölünebilme yeteneğine sahiptir.
2. Kök hücreler, özelleşmemiş hücrelerdir.
3. Kök hücreler, özelleşmiş yavru hücrelere kaynaklık edebilir.
4. Kök hücreler, nakil yapılacak donöre aktarıldıktan sonra kaynak dokuyu yapısal olarak çoğaltabilir.
5. Kök hücreler, doku hasarının olmadığı durumlarda dahi farklılaşmamış kuşaklara in vivo ortamda katkı sağlayabilmektedir (Karaöz ve Ovalı, 2004).

2.2.2.1 Hematopoetik Kök Hücreler

Hematopoetik kök hücreler, kandan ve kemik iliğinden izole edilebilen, kendi kendini yenileyebilen, kan dolaşımı ile aktif hale dönüşebilen, özelleşmiş hücre çeşitlerine farklılaşabilen, programlı hücre ölümüne gidebilen hücreler olarak adlandırılmaktadır (Uşul, 2012).

Hematopoetik hücrelerin özelliklerinin anlaşılması 40 yılı aşkın zaman öncesinde kemik iliği üzerindeki hücrelerin bir alt kümesinin üzerinde makroskopik koloniler oluşturma yeteneğini gösteren bir dizi seminal deney sayesinde gerçekleşmiştir (McCulloch ve Till, 1960).

HKH'ler prospektif olarak dokuya izole edilmiş ilk spesifik kök hücrelerdir ve aynı zamanda rutin klinik kullanımdaki tek kök hücrelerdir. Lösemi ve otoimmün gibi çeşitli hastalıkların tedavisinde yaygın olarak kullanılan greftler bu duruma neden olmaktadır. HKH'lerin hücresel ve moleküller özelliklerinin ayrıntılı olarak incelenmesi, klinik kullanımdaki çalışmaları, kök hücre tanımlamasını ve kullanımını büyük oranda etkilemiştir (Gluckman, 2000).



Şekil 2.5 : Hematopoetik ve Stromal Hücre Farklılaşması

Kaynak: (Terese, 2001)

Şekil 1'de kemik iliği hematopoetik kök hücrenin tüm alt bileşenleri ifade edilmektedir.

Hematopoetik kök hücrenin:

1. Bölünerek kendini yeniden üretme özelliği
2. Farklılaşarak olgun kan hücrelerine dönüşebilme özelliği
3. Farklılaşmaya başlayan hücre yönlendirilmiş fonksiyonel hücreye dönüşmektedir. (eritrosit-lökosit-trombosit) (İnan ve Özbilgin, 2009)

Bir kök hücre lenfoid ve miyeloid kök hücrelerine dönüşmektedir. Bunlar tüm kan

hücrelerinin öncül hücreleridir (Adams ve Scadden, 2006).

Miyeloid Kök Hücreler

Myeloid kök hücreler aşağıdaki olgun kan hücrelerine dönüşebilmektedir:

1. Kırmızı kan hücreleri (alyuvarlar- eritrositler): Bir protein olan hemoglobin ile doludur. Hemoglobinler, oksijeni akciğerlerden vücudun diğer hücrelerine taşımaktadır. Karbondioksitin vücut içinde taşınımını ve dışarı atılmasına kadar olan görevlerden sorumludur.
2. Trombositler (kan pulcukları): Vücut içinde yaralanan dokudaki kanamanın durmasına ve pıhtılaşmaya yardımcı hücrelerdir.
3. Beyaz kan hücreleri (akyuvarlar- lökositler): Üç grupta incelenmektedir;
 - Granülositler (nötrofil, bazofil, eozinofil),
 - Monositler
 - Lenfositler (Altun, 2018)

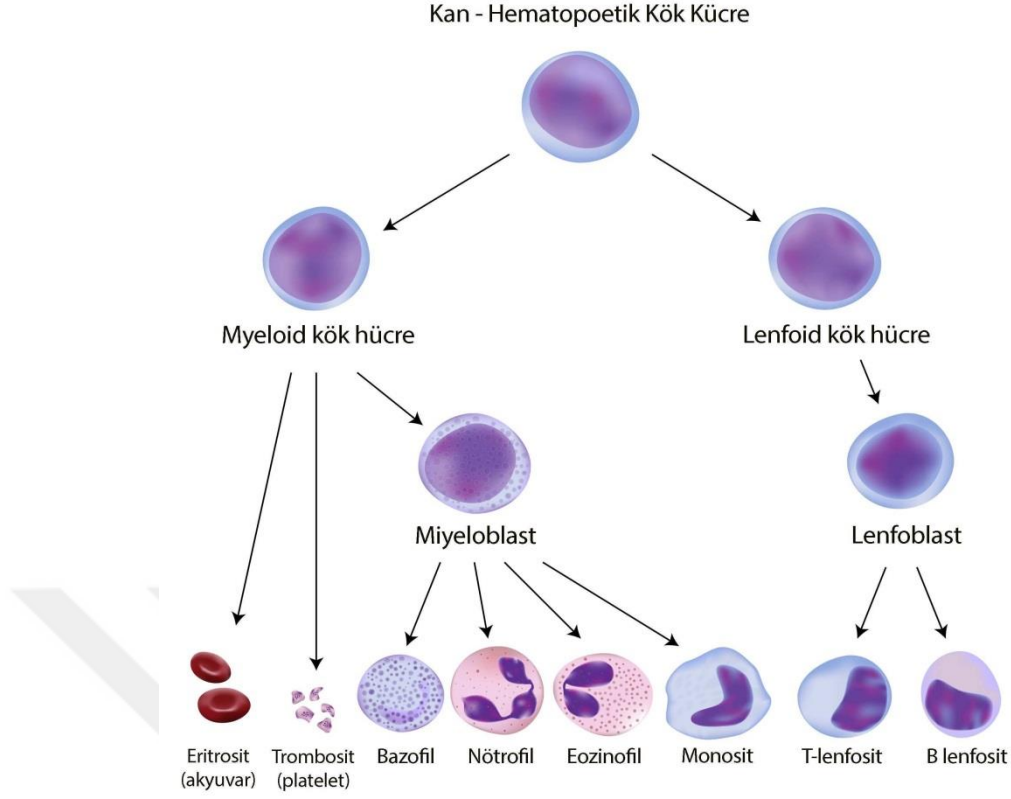
Bu hücrelerin belirli bazı görevleri bulunmaktadır. Nötrofiller ve monositler, bakteri veya mantarları etkisiz hale getirmede kullanılan yabancı maddeleri içine alarak yok etmede görev almaktadır. Monositler, alyuvar ve kan pulcuklarının aksine kandan ayrılıp dokuya girer ve orayı işgal eden organizmaya saldırarak enfeksiyonlarla mücadeleye yardımcı olmaktadır. Eozinofil ve bazofiller vücuda alerjik etki yapan maddelerin ve parazitlerin etkisiz hale getirilmesinde görev alan beyaz kan hücreleridir (Iwasaki ve Akashi, 2007).

Lenfoid Kök Hücreler

Sadece bir kısmı kana geçen lenfositlerin çoğu lenf bezlerinde, dalakta ve lenf kanallarında bulunmaktadır. Bağışıklık sisteminin önemli bir kısmını oluşturur ve üç tipte görülmektedir:

1. B-Lenfositler; Antikor adı verilen koruyucu maddelerin üretilmesinde görev almaktadır. Antikorlar aynı zamanda enfeksiyonlarla mücadele etmektedir.
2. T-Lenfositler; B-Lenfositlere destek sağlamaktadır.
3. NK hücreleri (doğal katil hücreler); kanser ve virüs hücrelerinin etkisiz hale getirilmesinde görev almaktadır.

Kan hücreleri kemik iliğinde üretildikten sonra kana geçer ve taşınımları plazma içinde dağılarak tamamlanmaktadır. (Sudo ve diğ., 2000)



Şekil 2.6 : Kan Hücresinin Bileşenleri

Kaynak: (Tunalı, 2019)

Şekil 2’de Hematopoetik kök hücrenin olgunlaşmamış iki öncül hücresi ve alt bileşenleri görülmektedir.

2.2.3 Kanser Hücreleri

Bir organ veya dokudaki hücrelerin düzensiz bir şekilde bölünüp çoğalması sebebiyle oluşmuş kötü urlara kanser denir. Günümüzde en çok karşılaşılan hastalıklardan biridir. Bu hastalığı anlamak için hücrelerin kanser hücrelerine nasıl dönüştüğünün anlaşılması gerekmektedir. Vücudumuzdaki sağlıklı hücreler (kas ve sinir hücreleri hariç) bölünebilme yeteneğine sahiptir ve ölen hücrelerinin yenilenmesi, yaralanan dokuların iyileşmesi bu özellikleri sayesinde gerçekleşmektedir (Erçakallı ve Polat, 2022).

Hücrelerin bölünebilme özelliği yaşa bağlı olarak farklılık göstermektedir. Yaşamın ilk yıllarında hücrelerin bölünme hızları daha fazla iken, yaşa bağlı olarak bir düşüş meydana gelmektedir. Ayrıca hücrelerin bölünebilme yetenekleri de sınırlıdır. Hücrelerin bir bölünebilme sayıları vardır ve gerektiği zaman programlı ölüm

gerçekleştirmektedir. Bu olaya apoptosis adı verilmektedir (Weiss ve Dahlke, 2019). Sağlıklı bir bireyde yaşamın devamı için hücrelerin kontrollü bir şekilde bölünmesi, büyümesi ve hücre ürettikten bir süre sonra apoptosis olması beklenir. Fakat bazen süreç doğru yoldan sapar ve kontrolsüzce çoğalan kanser hücreleri bilinçsizce bölünmeye başlar ve çoğalırlar. Fazla hücrelerin sayısı arttıkça bu hücre kütleleri bir tümör veya yapısal bozukluğa sahip bir büyüklük oluşturabilir (Huang ve Freter, 2015).

2.2.4 Lösemi

Lösemi diğer bir deyişle kan kanseri, tüm hücrelerin öncülleri olan kök hücrelerden olgun kan hücrelerinin oluşması sırasında meydana gelen sorunlar ile ortaya çıkmaktadır. Öncül hücreler bu olaylar sonucunda anormal, işlevsiz hücrelere dönüşür ve kontrolsüz bölünme gerçekleştirir. Bu da hastalıkların oluşmasına etken rol oynar. Hastalığın vücuda yayılması bu hücrelerin kemik iliğinde ve kanda sayılarının kontrolsüz artması sonucunda oluşmaktadır (Atıcı, 2007).

Hastalığın hızlı veya yavaş seyirli olmasına göre akut veya kronik lösemi olarak adlandırılabilir. Genellikle akut lösemiler çocuklarda ve gençlerde sıklıkla gözlemlenirken, kronik lösemiler erişkin ve yaşlılarda gözlemlenmektedir. Lenfoid ve miyeloid olarak hastalığı oluşturan öncül hücrelerin tipine göre de sınıflandırılmaları bulunmaktadır (Taylor ve diğ., 1992).

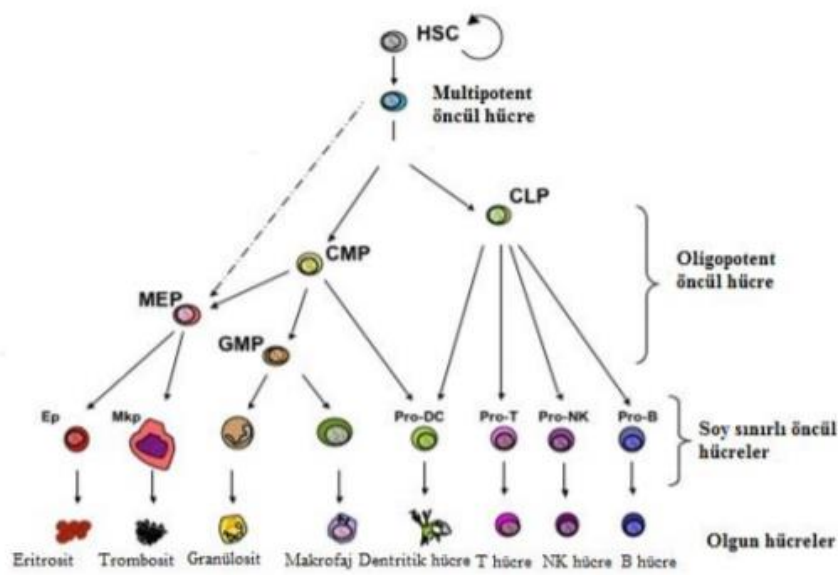
2.2.4.1 Akut Miyeloid Lösemi

Granülosit ve monositler lökositlerin oluşmasını sağlayan hücrelerdir ve bu hücreler blast adı verilen hücrelerin olgunlaşmasından meydana gelir. Oluşan hücrelerin özellikleri blastlarla benzerlik göstermez. Onların aksine mikroplarla mücadele etmektedir (Yiallourous ve diğ., 2022).

AML'de normal olgunlaşma süreci bozulan blast hücreleri kemik iliği ile birlikte kanda birikmeye başlar. Bu duruma beyaz kan hücreleri yeterli düzeyde kalamadığı için vücut savunmasız kalır. Miyeloblastların anormal hızda çoğalması sonucu kemik iliğinde trombosit ve eritrosit yapımı bozulmaktadır. Buna bağlı olarak trombosit ve anemi hücrelerinde azalma gözlemlenmektedir. (Thomas ve Maieti, 2017).

2.2.5 Hematopoez

Hematopoez normal kan hücrelerinin yapımı ve olgunlaşma süreci olarak tanımlanabilir. Hematopoetik gelişimin ilk aşaması, birden çok hücre tipine dönüşme yeteneğine ve kendi kendini yenileme özelliğine sahip hematopoetik kök hücrelerin (pluripotent hücrelerin), lenfoid veya miyeloid öncül hücrelere (multipotent hücrelere) farklılaşmasıyla gerçekleşmektedir. Lenfoid hücreleri T hücre, B hücre ve NK hücrelerini; miyeloid hücreleri ise monosit, eritrosit, bazofil, eosinofil, makrofaj ve megakaryositler oluşturmaktadır (Doulatov ve diğ., 2012).



Şekil 2.7 : Hematopoez Aşamaları

Kaynak: (Altay, 2021)

HSC: hematopoetik kök hücre, CLP: ortak lenfoid öncül, CMP: ortak miyeloid öncül, GMP:granülosit makrofaj öncül, MEP: megakaryosit-eritroid öncül hücre, Ep:eritrosit öncül hücre, MKP: megakaryosit öncül hücre, Pro-DC: öncül dentritik hücre, Pro-T: öncül T hücre, Pro-NK: öncül doğal öldürücü hücre, Pro-B: öncül B hücre

Şekil 3’de multipotent olan hematopoetik kök hücresinin 2 ana bileşeni ifade edilirken bu hücrelerinde kendi içinde soy sınırlı öncül hücreler üzerinden olgun hücrelere dönüşebildiği ifade edilmiştir.

3. MATERYAL VE YÖNTEM

National Center for Biotechnology Information (NCBI) biyomedikal mühendisliği, istatistik ve uygulamalı matematik alanlarındaki yaklaşımların ve yöntemlerin, biyolojik veri analizi için uygulanması olarak çalışılabilen bir veri tabanıdır. Gen dizilerinden protein yapı tahminlerine uzanan genomik bilgilere erişmeyi mümkün kılmaktadır (Ostell ve diğ., 2001). Bu çalışmada NCBI'nin alt veri tabanı Gene Expression Omnibus (GEO) veri kümesi olan insan kemik iliği hematopoetik kök hücre ekspresyon verileri kullanılmıştır.

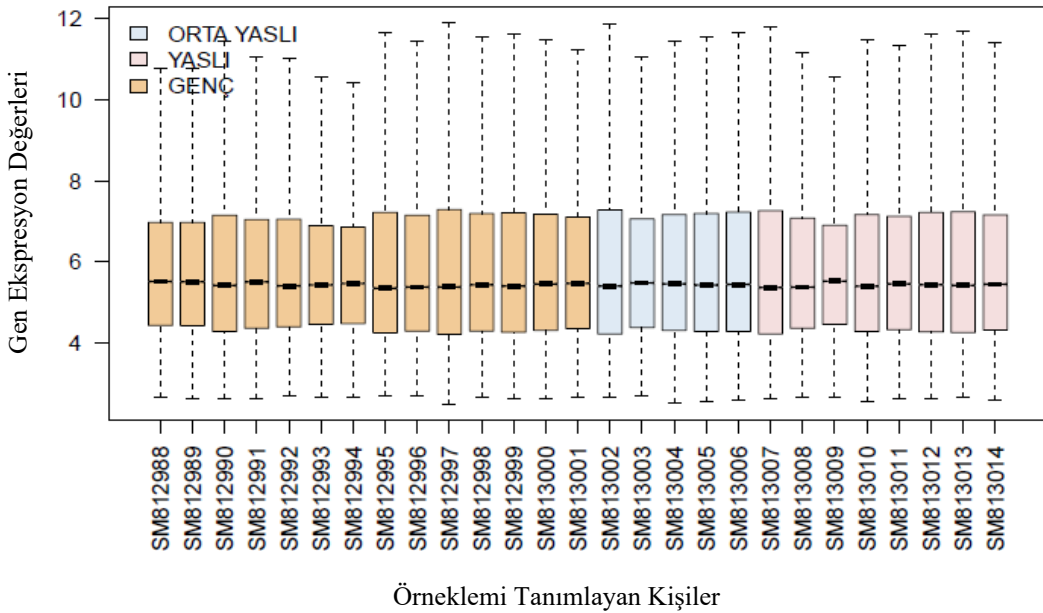
Gene Expression Omnibus (GEO), gen ifadesi veya moleküler verileri için bir veri deposudur. Veri deposu, kullanıcıların ilgilendikleri veri setlerini seçip indirebilecekleri çevrimiçi bir arayüze sahiptir. GEOquery paketi, GEO verilerinin indirilmesini ve bunların çeşitli veri kümelerine ayrıştırılmasını ve analiz için uygun diğer Bioconductor veri yapılarına dönüştürülmesini destekleyen kullanışlı bir arabirim araçları sağlamaktadır (Barrett ve diğ., 2012).

Bu çalışmada NCBI-GEO veri tabanından sağlanan veri seti R programlama ile analiz edilmiştir. R istatistiksel hesaplama ve grafikler için bir yazılım dilidir. Doğrusal ve doğrusal olmayan modelleme, klasik istatistiksel testler, zaman serisi analizi, sınıflandırma ve kümeleme gibi çok çeşitli istatistiksel analizlerin yapılabilirdiği bir programdır. İstatistiksel analizleri ve çıktı grafiklerini çalıştırmak için tasarlanmış bir yazılımdır. R'nin güçlü yönlerinden biri, gerektiğinde matematiksel semboller ve formüller de dahil olmak üzere iyi tasarlanmış görsel çıktıların üretilme kolaylığı sağlamasıdır (Tippmann, 2015).

İstatistiksel hesaplamayı kolayca gerçekleştirmek ve makine öğreniminin matematiksel yönünü kullanmak konusunda R, en popüler dillerden biridir. Verileri herhangi bir otomatik öğrenmeye göndermeden önce keşfetmek ve öğrenme algoritmasının sonuçlarını değerlendirmek için gerekli olan görselleştirme özellikleri de sağlamaktadır. Makine öğrenimi için pek çok R paketi kullanılabilir ve istatistiksel öğrenimdeki birçok modern yöntem, geliştirmelerinin bir parçası olarak R programlamada uygulanmaktadır (Peng, 2016).

İstatistiksel veri analizi yapılırken birçok programlama dili kullanılabilir. Sıkça kullanılan programların başında SPSS, Matlab, SAS ve R programlama dilleri gelmektedir (Jain, 2018). Bu çalışmada NCBI veri tabanı ile entegre güçlü kütüphane seçeneği ve senkronizasyonu sebebiyle R programlama tercih edilmiştir. Diğer programlama dillerinden farklı olarak R programlamada mevcut olan kütüphaneler NCBI veri tabanına ait veri setleri üzerinden çalışmayı daha optimal hale getirmektedir. Bununla birlikte R programlama kütüphane seti, yapılan analizleri görsel açıdan destekleyici histogram ve grafikleri zenginçe ifade etmeye olanak sağlamaktadır. Ayrıca R dili, makine öğrenimi modelleriyle çalışmak için uygun ve çeşitli kütüphanelere sahip bir programdır. Veri setinin makine öğrenmesi algoritmalarıyla analizi için güçlü bir olanak sağlamaktadır (Barr, 2018).

Çalışmada analiz edilen örnekleme, insan kemik iliği hematopoetik kök hücreleri 3 gruba ait olarak kullanılmaktadır. 14 genç (20–31 yaş), 5 orta yaş (42–61), 8 yaşlı (65–85) şeklindedir ve bu gruplara ait toplam 50.000 gen bilgisi bulunmaktadır. Gruplara bağlı kalarak veri seti içindeki veriler ve gen değerleri arasında seviye kontrolü ve yaşlanmaya bağlı genlerdeki değişikliklerin ortaya çıkarılması hedeflenmiştir.



Şekil 3.1 : İnsan Gen Ekspresyon Değerlerinin İstatistiksel Dağılımı

Şekil 8’de örnekleme tanımlayan kişiler 3 grupta incelenerek 50.000 gen ekspresyon ifadesinin tanımlayıcı istatistiksel bilgisi kutu grafiğinde gösterilmiştir.

Bir hipotezin doğruluğunu istatistiksel olarak saptamak için hipotez testleri uygulanmaktadır. Belirli bir grup ya da popülasyon hakkındaki fikirleri test etmenin sistematik bir yoludur. Örnekleme ölçülen verileri kullanarak bir popülasyondaki, belirli bir parametre hakkında ortaya koyulan hipotezi test etmek için kullanılan yöntemlerdir (Klein ve diğ., 2003).

Hipotezi test etmek için izlenen adımlar, hipotezin belirlenmesi, karar için parametrelerin seçilmesi ve test istatistiğinin hesaplanması şeklindedir. Çıkarımsal istatistikte H_0 ve H_1 gibi alternatif hipotezler bulunmaktadır. Sıfır hipotezi, belirli bir gözlem dizisinde değişkenler arasında istatistiksel anlamlılığın bulunmadığını öneren bir yaklaşımdır. Alternatif hipotez ise sıfır hipotezini reddeder ve araştırılan ifadeyi temsil ederek, örnekleme ait değişkenler arasındaki ilişkiyi anlamlı kabul etmektedir (Mallat, 2017). Bu çalışmada yapılan analiz için belirlenen Sıfır hipotezi, örneklem içerisindeki bireylerde yaşlanma arttıkça miyeloid ve lenfoid yanlı genlerin ekspresyon seviyelerinde herhangi bir değişiklik olmadığı şeklinde belirlenmiştir. Alternatif hipotez ise yaşlanma ile birlikte miyeloid ve lenfoid yanlı genlerin ekspresyon seviyelerindeki değişimin anlamlı biçimde farklı olduğunu temsil etmektedir.

Sıfır hipotezinde belirtilen değerin doğru olduğu varsayıldığında, bir örnek ortalamasının elde edilme olasılığı p değeri ile ifade edilmektedir. P değeri bir olasılıktır ve 0-1 aralığında bir değere sahiptir. Bir p değeri, sıfır hipotezinde belirtilen değerin doğru olduğu varsayıldığında, bir örneklem sonucunun elde edilme olasılığıdır. Örneklem sonucunun elde edilmesi için p değeri, anlamlılık düzeyi ile karşılaştırılır. Araştırma çalışmalarında olasılık veya anlamlılık düzeyi %5 olarak belirlenmektedir. Sıfır hipotezi doğru olduğu durumda örneklem ortalamasının elde edilme olasılığı %5'ten azken sıfır hipotezinde belirtilen önerme reddedilmektedir (Ziliak, 2017).

İstatistiksel testler parametrik ve parametrik olmayan testler şeklinde iki ana başlıkta incelenmektedir. Parametrik testlerde normal dağılım ve homojen bir varyans dağılımı aranmaktadır. Parametrik olmayan testlerde anakütle dağılımı incelenmez ve normal bir dağılım olmadığı durumlarda kullanılır. İstatistiksel testlerde analizlerin güçlü olabilmesi için en iyi varsayımların elde edilmesi gerekir. Parametrik testlerin de istatistiksel gücü daha yüksektir. Örnekleme dağılımı, örneklem sayısı, veri grupları ve değişken yapıları gibi parametreler istatistiksel

hipotez testi seçiminde önemli rol oynamaktadır (Harwell, 1988).

ANOVA, ikiden fazla örneklem grubunun ortalamaları arasındaki farkı analiz etmek için kullanılan parametrik bir hipotez testidir. Örneklemeye ait gruplar arasında anlamlı bir biçimde fark olup olmadığı test edilmektedir. Tek yönlü varyans analizi ve iki yönlü varyans analizi şeklinde iki ana tipte incelenmektedir. Tek yönlü varyans analizinde, bağımlı değişken üzerinde tek bir parametrenin etkisi araştırılmaktadır. İki yönlü varyans analizi ise, iki bağımsız değişkenin bir bağımlı değişken üzerindeki etkisi araştırılırken kullanılmaktadır (Speed, 1987). Değişkenler bağımlılık durumlarına göre bağımlı ve bağımsız değişkenler olarak ikiye ayrılır. Bağımsız değişken araştırmada değeri başka değişken veya değişkenlere bağlı olarak değişim göstermeyen, bağımlı değişkenin alacağı değerleri belirleyen değişkenlerdir. Bağımlı değişken, araştırmadaki diğer değişken veya değişkenlere bağlı olarak değişim gösterebilen değişkenlerdir (Hazra ve Gogtay, 2016).

ANOVA testi diğer istatistiksel testler gibi veri grupları arasındaki farkların istatistiksel olarak anlamlı olup olmadığını test eder. Verilerin örneklemdeki gruplar arasındaki varyans düzeylerini analiz ederek çalışmaktadır. Bir örneklem üzerindeki anakütle ortalamaları arasında anlamlı bir fark olup olmadığını kontrol etmek için kullanılmaktadır. Tüm bu unsurlar daha sonra gruplar arasındaki farklılıkların istatistiksel olarak anlamlı olup olmadığına dair bir olasılık (p-değeri) vermektedir (Kim, 2017). Örneklem grubu ikiden fazla grubu içeriyor ve normal bir dağılıma sahipse bağımsız değişken grubuna göre tek yönlü ya da iki yönlü varyans analizi kullanılmaktadır. Örneklem normal dağılım göstermediği durumlarda ANOVA'nın parametrik olmayan karşılığı Kruksal-Wallis sıralamalı tek yönlü varyans analizi yapılmaktadır (Johnson, 2022).

İstatistiksel analiz yapılırken verilerin nasıl dağılım gösterdiği önemlidir. Örneklem normal dağılım veya normale yakın bir dağılım gösterdiği durumlarda parametrik testler kullanılmaktadır. Varyansların homojen bir şekilde dağılması ve örneklem gruplarının birbirlerinden bağımsız olması gerekmektedir (Ali ve Bhaskar, 2008). Şekil 4'te verilen kutu grafiği örneklem gruplarının tanımlayıcı istatistikleri üzerinden normal dağılım ve homojen varyans gösterdiği ifade edilmiştir. Veri setine ait karşılaştırılacak grup sayısı da seçilecek test için çok kritik bir parametredir. İkiden fazla grup incelendiği ve parametrik bir test yaklaşımı seçildiği için bu çalışmada hipotez testi olarak ANOVA testi kullanılmıştır. Analize ait alternatif

hipotez, insan hematopoetik kök hücrelerinde yaşlanmaya bağlı olarak miyeloid ve lenfoid yanlı genlerdeki ekspresyon seviyesinde farklılaşma olduğu varsayımı üzerine kurulmuştur. Burada bağımsız değişken, üç grupta incelenen yaş gruplarıdır. Bağımlı değişken de bu yaş gruplarına bağlı değişen ekspresyon seviyelerini temsil etmektedir. Tek bir bağımsız değişken olduğu için çalışmada tek yönlü varyans analizi (ANOVA) kullanılmıştır.

Parametrik testler, parametrik olmayan testlere göre daha güçlü sonuçlar üretmektedir. Normal dağılıma sahip veri setleri parametrik testler için daha uygundur. Bu sebeple Kruksal-Wallis testinin kullanımına gerek duyulmamıştır. Diğer parametrik testlerden, Ki-Kare ve Mann-Whitney U testi ve parametrik olan t-testleri iki grubun karşılaştırılması için kullanılan hipotez testleri olduklarından dolayı bu çalışmada kapsama dahil edilmemiştir (Johnson, 2022).

ANOVA testi sadece örneklem içerisindeki gruplar arasında anlamlı fark olma durumunu değerlendirmektedir. Bir fark varsa bu farkın hangi gruplar arasında olduğuna ilişkin bir saptamada bulunmaz. Bu bilgiye ulaşmak için post-hoc test yöntemleri kullanılmaktadır. Bu metotlar ANOVA testi sonucunda tespit edilen anlamlı farkın hangi gruplar arasında olduğunu tespit etmek için kullanılmaktadır. Sıfır hipotezi kabul edildiğinde, gruplar arasında bir farkın olmadığı durumda post-hoc testlerine gerek duyulmamaktadır (Sedgwick, 2014).

Post-hoc testi ikiden fazla popülasyon için kullanılan hipotez testlerinin anlamlı olduğu durumlarda kullanılan bir yöntemdir. İstatistiksel olarak anlamlı farkın hangi gruplar arasında gerçekleştiğini anlamak için gruplar ikili şekilde ele alınmaktadır (Homack, (2001). İnsan hematopoetik kök hücrelerinin 3 grupta incelendiği bu çalışmada, alternatif hipotez kabul edilmiş ve gruplar arasında anlamlı bir fark bulunmuştur. Gruplar arasındaki anlamlılık post-hoc testlerinden biri olan Benjamin-Hockberg metodu kullanılmıştır.

Varyansların homojen dağıldığı durumlarda seçilebilecek çoklu karşılaştırma testleri: LSD (Least Significant Difference), Sidak, Bonferroni, Tukey, Benjamin-Hockberg, Gabriel ve Scheffe olarak bilinmektedir. Benjamin-Hockberg testi eşit olmayan örnek boyutlarında çalışmaktadır. Benjamin-Hockberg metodu, yanlış keşif oranını azaltan güçlü bir yöntemdir. Sıfır hipotezi doğru olmasına karşın sıfır hipotezini reddetmek, yanlış pozitifleri ortaya çıkarmaktadır (Ferreira ve Zwinderman, 2006).

Benjamin-Hockberg yöntemi ile bu hata tipindeki ifadelerin azaltılması hedeflenmektedir. Bu çalışmaya ait veri seti 3 gruptan ve farklı örneklem adetlerinden oluştuğu için tercih edilmiştir. Benjamin-Hockberg yöntemi kullanıldıktan sonra anlamlılık seviyesi 0.05'den düşük 570 gen karakteri elde edilmiştir. Gen karakterleri daha sonra veri setindeki anlamlılık seviyesini zenginleştirme amacıyla biyoloji tabanlı GOrilla programında incelenmiştir.

GOrilla, sıralanmış gen listelerinde zenginleştirilmiş GOrilla terimlerini tanımlamaya ve görselleştirmeye yönelik bir araçtır. GOrilla'nın ana kullanımlarından biri, gen setleri üzerinde zenginleştirme analizi yapmaktır. Belirli koşullar altında yukarı regüle edilen bir dizi gen alındığında, bir zenginleştirme analizi, o gen seti için ek açıklamalar kullanılarak hangi GO terimlerinin fazla ya da az temsil edildiğini ortaya çıkarmaktadır (Supek ve Škunca, 2017).

Analiz edilecek genlerin adları satır bazlı sıralandıktan sonra program gen adına göre veri setini işlemektedir. Analiz için moleküler fonksiyon, hücresel bileşen ve organizma türü gibi parametreleri seçilip liste programa yüklenmektedir. Programa ait p değeri, GO terimleri arasındaki anlamlılık ifadesinin değerini temsil etmektedir. Belirli GO teriminin bir grup gen içerisindeki açıklamaların arka plan dağılımıyla karşılaştırılması sonucu p-değerindeki sıfıra yakınlık derecesi anlamlılık kriterini belirlemektedir (Supek ve Škunca, 2017).

Zenginleştirme çalışmaları sağlayan bir dizi farklı araçlar da bulunmaktadır. Bunlardan bazıları web tabanlıdır, diğerleri ise kullanıcının bir uygulama indirmesini veya yerel bir ortam kurmasını gerektirebilir. Araçlar, kullandıkları algoritmalar, gerçekleştirdikleri istatistiksel testler ve temel GOrilla verilerinin güncellenme sıklığı bakımından farklılık göstermektedir. Bu araçlardan bazıları; GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion), Fatiscan, Go-stat, GOEAST, GoMiner, GeneWeaver, gProfiler, Ontologizer şeklindedir (Eden ve diğ., 2009).

Bu çalışmada anlamlı gen dizisini zenginleştirme çalışmaları kapsamında GOrilla aracı kullanılmıştır. Farklı GO zenginleştirme araçları, çok çeşitli istatistikler ve farklı performanslar sağlamaktadır. Her bir GO aracı farklı çalışma sürelerinde çalışırken, ayrıca bazı GO araçları görsel grafikasyon desteği sunmamaktadır. GOrilla zenginleştirme aracının bu çalışmada seçilme amacı, performans açısından yüksek hız desteği ve görsel çıktı ile kolay okunabilir bir arayüz desteğidir. Ayrıca

GORilla kullanılırken herhangi bir program indirme gereksinimi duyulmamaktadır (Eden ve diğ., 2009).

Çizelge 3.1 : Analiz Sonucu Elde Edilen Genler, ID ve Açıklama Bilgilerinin Bazıları

GEN ID	GEN İSMİ	AÇIKLAMA
1554699_at	L3MBTL4	l (3) mbt benzeri 4 (Drosophila)
1558256_at	LINC00662	RNA 662 kodlayan uzun intergenik protein olmayan protein
201041_s_at	DUSP1	çift özgüllüklü fosfataz 1
201324_at	EMP1	epitel membran proteini 1
2028_s_at	E2F1	E2F transkripsiyon faktörü 1

Makine öğrenmesi, bilgisayar yazılımlarına veri türlerine dayalı öğrenmeyi geliştiren bir yapay zekâ tekniğidir. Makine öğrenmesi algoritmaları, bir model olarak önceden belirlenmiş bir denkleme dayanmadan bilgileri doğrudan verilerden öğrenmek için hesaplama yöntemlerini kullanmaktadır (Mahesh, 2020). Makine öğreniminde iki tür teknik kullanılmaktadır. Gelecekteki çıktıların tahmin edilmesi için bilinen girdi ve çıktı verileri üzerinde bir model eğiten denetimli öğrenme ile girdi verilerinde gizli kalıpları veya içsel yapıları bulan denetimsiz öğrenme algoritmaları şeklindedir (Jo, 2021).

Denetimli öğrenme algoritması, hedef yanıtı bilinen veri kümesi üzerinde etiketli bir veri setinin eğitilmesi ile tahmin yapan bir modeli içermektedir. Denetimli öğrenme modelinde veri seti, belirli bir oranda eğitim seti olarak ayrılarak hedef etiketi üzerinde eğitilir. Yeni girdi sağlanan veriler üzerinde tahmin etmeyi sağlayacak bir model kurgulanır. Denetimsiz öğrenmede ise etiketli yanıtlar olmadan verilerdeki gizli kalıpları girdi verilerinden çıkarım yapılması için kullanılmaktadır. Tahmin edilmesi çalışılan çıktı için veriler biliniyorsa denetimli öğrenme algoritmaları kullanılmaktadır (Donalek, 2011).

Sınıflandırma ve regresyon analizi denetimli öğrenme algoritmalarını oluşturan tekniklerdir. Sınıflandırma modelleri, girdi verilerini kategoriler halinde sınıflandırır. Veriler etiketlenebiliyor, kategorilere ayrılabilir veya belirli gruplara ya da sınıflara ayrılabilirse sınıflandırma tekniklerin kullanılmaktadır. Regresyon teknikleri ise bir veri aralığında eğitilen verilerden gelen girdi özelliklerine dayalı olarak sürekli şekilde tahminlerde bulunmaktadır (Thabtah ve diğ., 2019).

Gen ekspresyon seviyelerindeki zenginleştirme metotlarından elde edilen 570 gen karakteri, makine öğrenmesi teknikleriyle miyeloid ve lenfoid yanlılık derecelerine göre sınıflandırılması hedeflenmiştir. Yaş gruplarına göre, genç ve yaşlı bireylerin ekspresyon seviyelerinin ortalaması hesaplanılarak hangi hücre yanlı olduğu makine öğrenmesi metoduyla incelenmiştir. Denetimli öğrenme metotlarından sınıflandırma algoritması, K-en yakın komşu algoritması kullanılmıştır. Etiketli hedef yanıtı bilinen hücreler üzerinden ekspresyon seviyelerine göre tahmin yapan bir model tasarlanmıştır.

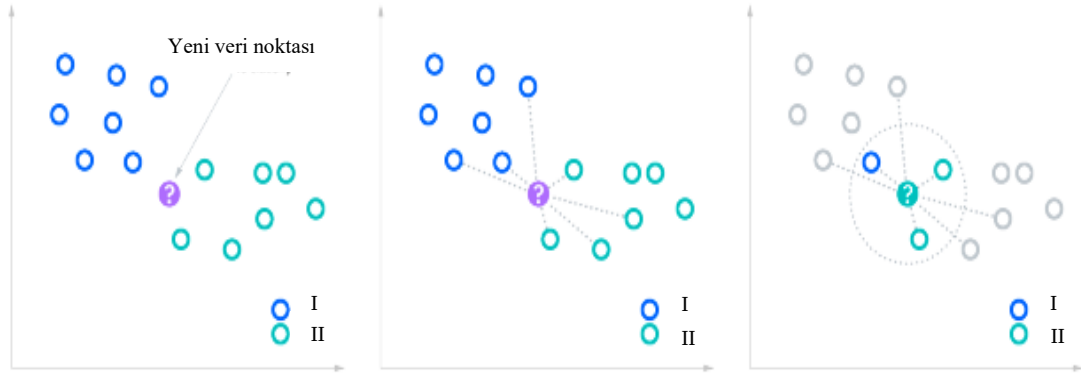
Çizelge 3.2 : Makine Öğrenmesi Algoritmasında Kullanılan Veri Setinin İlk 5 Satırı

GEN ID	GEN İSMİ	AÇIKLAMA	EKSPRESYON GENÇ	EKSPRESYON YAŞLI	MİYELOİD/ LENFOİD
1554699_at	L3MBTL4	l (3) mbt benzeri 4 (Drosophila)	4,52	5,28	Miyeloid
1558256_at	LINC00662	RNA 662 kodlayan uzun intergenik protein olmayan protein	4,22	5,3	Miyeloid
201041_s_at	DUSP1	çift özgüllüklü fosfataz 1	9,69	13,07	Miyeloid
201324_at	EMP1	epitel membran proteini 1	9,44	7,89	Lenfoid
2028_s_at	E2F1	E2F transkripsiyon faktörü 1	8,78	5,61	Lenfoid

K-en yakın komşu algoritması, verilerin gruplandırılması hakkında sınıflandırma tahminleri yapmak için kullanılan bir denetimli makine öğrenme algoritmasıdır. KNN algoritması bilinmeyen bir girdinin, komşuluğunu ve veri kümesindeki uzaklık parametrelerini kullanarak tahmin yapan bir yöntemdir. Veri kümesinde olmayan bir gözlem için algoritma, bu gözleme en yakın çevreyi temel alarak tanımlanan k örnek sayısını aramaktadır (Lesmeister, 2015).

KNN algoritması, sınıflandırma metodunda belirli bir görünmeyen gözleme en çok benzer örneklerle sahip k arasında çoğunluk derecesini oluşturmaya çalışır. Benzerlik iki veri noktası arasındaki mesafe metriğine dayanmaktadır. Algoritmada kullanılacak k değeri doğruluk oranını etkilemektedir. (Yong, 2009). Analiz sırasında en optimal k değerini bulmak için doğruluk ve çapraz geçerlilik (accuracy cross

validation) grafiđi çizdirilmiştir. Bu analiz sonucunda en optimal k değeri 5 olarak tespit edilmiştir.



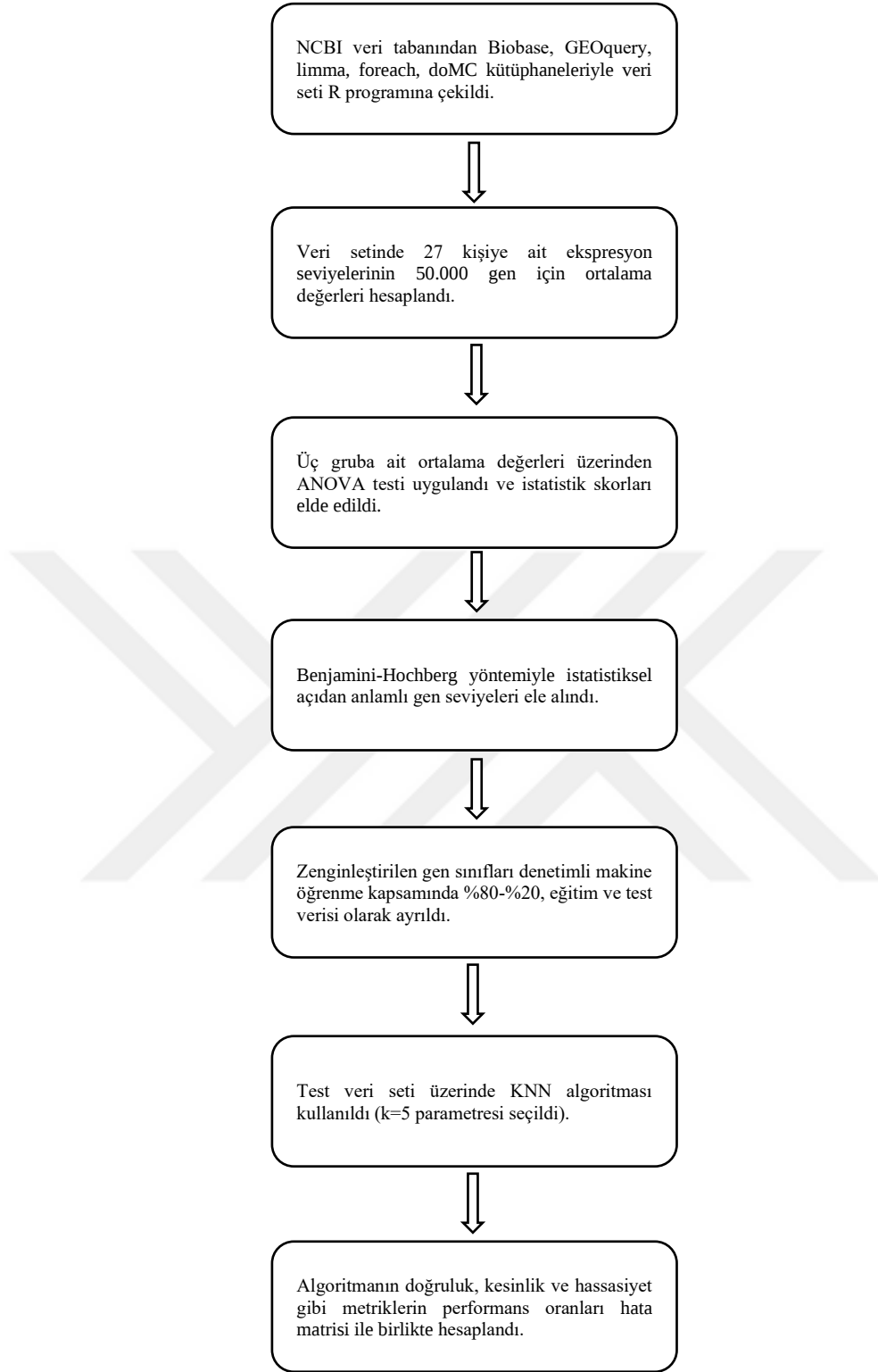
Şekil 3.2 : KNN Uygulaması

Kaynak: (Yong ve diğ., 2009)

Makine öğrenme algoritmalarında tek bir yaklaşım türü yerine çeşitli parametreler doğrultusunda bir yaklaşım izlenmektedir. Verinin boyutu, sınıflandırma mı yoksa regresyon problemi mi ve verinin etiketli olup olmadığı durumlarına göre seçilecek algoritma şekillenmektedir (Sharma ve Kumar, 2017).

Tablo 2’de makine öğrenmesiyle analiz edilecek öznitelikler ifade edilmiştir. Genç ve yaşlı ekspresyon seviyelerinin iki grup için de ortalaması ile gen ekspresyon seviyelerine göre miyeloid ve lenfoid yanlı hücreler etiketlenmiştir. Etiketli veri üzerinden gen ekspresyon seviyeleri bazında sınıflandırma modeli tasarlanmıştır. Veri setinin etiketli bir hedef özniteliđi barındırması denetimli öğrenme algoritması seçimini desteklemektedir. Buradaki etiket, yaş gruplarına ait ekspresyon seviyelerindeki ortalama değeri hangi bilgiye karşılık geldiđini temsil etmektedir.

KNN algoritması bir sınıflandırma problemi için uygun bir algoritmadır. Algoritma kolay uygulanabilir ve anlaşılabilir bir yöntemdir. Algoritma kendi içinde k değeri gibi bir değışken ile optimize edilebilir. Veri setinde az sayıda öznitelik üzerinden bir model kurulması KNN algoritması için daha uygundur. SVM algoritması büyük ve kompleks veri setleri için daha uygun bir algoritmadır. Karar ağaçları yapısı ise birden çok öznitelik için bir hiyerarşı kurmaktadır. Bu çalışmada gen ekspresyonlarının yaş grupları üzerinden etiketlenmesi daha az öznitelikle kurgulanmıştır. (Yahia ve Ibrahim, 2003).



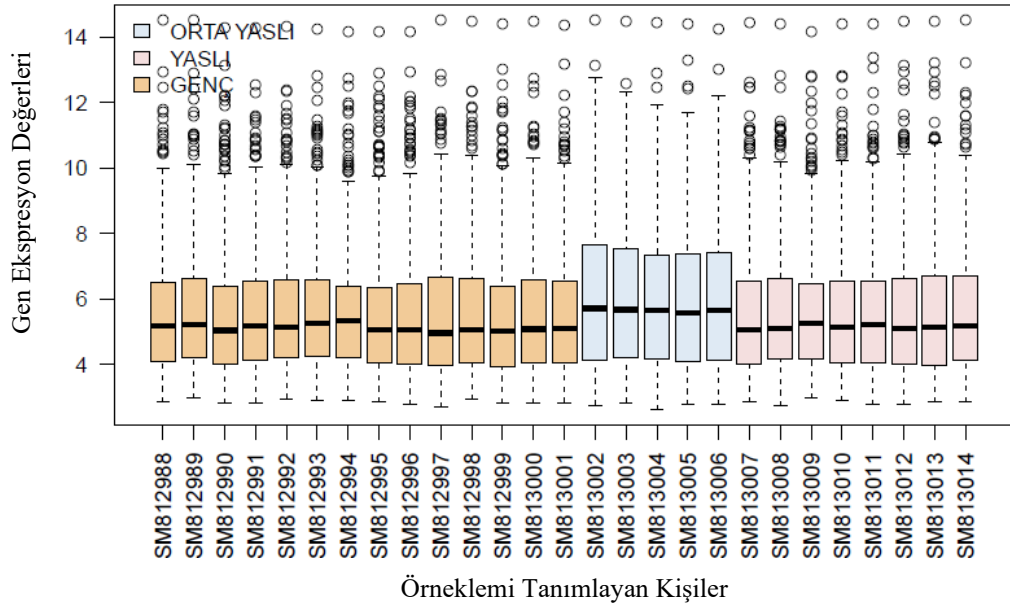
Yukarıda bu çalışmaya ait, materyali R programlama ile gerçekleştirilen akış şeması verilmiştir.

4. BULGULAR VE TARTIŞMA

4.1 Bulgular

Yaşa bağlı hematopoetik fonksiyon bozukluklarına yatkın olabilecek yaşlı insan hematopoetik sisteminin özelliklerini açıklamak için sağlıklı, hematolojik olarak normal genç ve yaşlı insan kemik iliği örneklerinden HKH ve diğer hematopoetik progenitor popülasyonlar değerlendirilerek orta yaşlı HKH'nin miyeloid yanlılığının sıklıkla arttığı istatistiksel olarak tespit edilmiştir. Yapılan ANOVA testi ile örnekleme ait yaş grupları arasında ekspresyon seviyeleri arasında anlamlı bir farkın olduğu $p < 0,043$ ile ortaya koyulmuştur.

Yaşlı HKH'nin çoğunluğu miyeloid yanlılık gösterirken, genç HKH'nin çoğunluğu lenfopoez ve miyelopoezde dengeli olduğu ve orta yaşlı bireylerde miyeloid yanlılık gösterdiği ortaya koyulmuştur. Yapılan testlerin ardından gen örnekleminde anlamlı olarak nitelendirilen genler değerlendirildiğinde, HKH artışı orta yaşlı bireylerde belirgin biçimde görülmektedir.



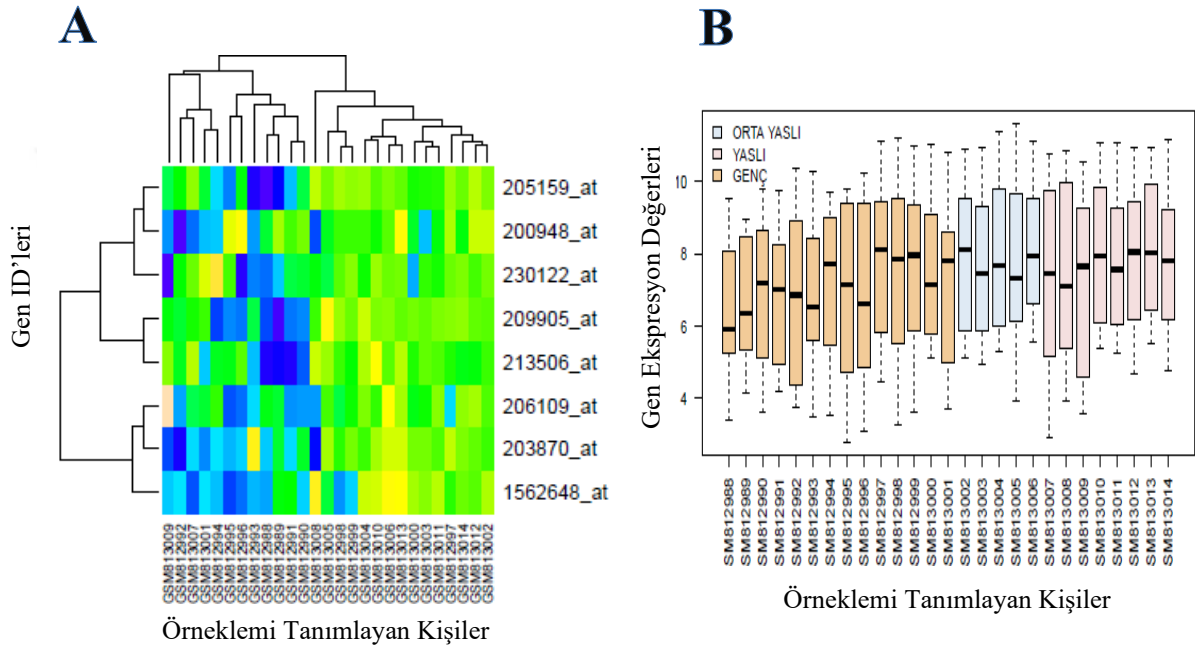
Şekil 4.1 : $p < 0.05$ Anlamlı Gen Ekspresyon Değerlerinin Kutu Grafiği

Şekil 6 'da orta yaşlı grubun gen ekspresyon profili incelendiğinde HKH ekspresyon değerlerinin arttığı görülmektedir. (Bu ifadenin analizi için kutu içerisindeki

ortalama bilgisi veren siyah çizgi referans alınmıştır).Yaşlı grup için ise bu değerler genç gruba yakın değerler seyretmekle beraber artış miktarı daha azdır.

Veri seti üzerinde yapılan çalışmalar sonucunda yaşlı ve genç kemik iliği HKH'nin gen ekspresyon profilinde yaşlı HKH'nin miyeloid soy yanlılığı gösterdiği tespit edilmiştir. Ayrıca lenfoid spesifik, örnekleme yaşa bağlı olarak azaldığı görülmektedir.

4.2 Miyeloid Özellikli Gen Analizi



Şekil 4.2 : Miyeloid Yanlılık Gösteren Genlerin Isı Haritası ve Kutu Grafiği

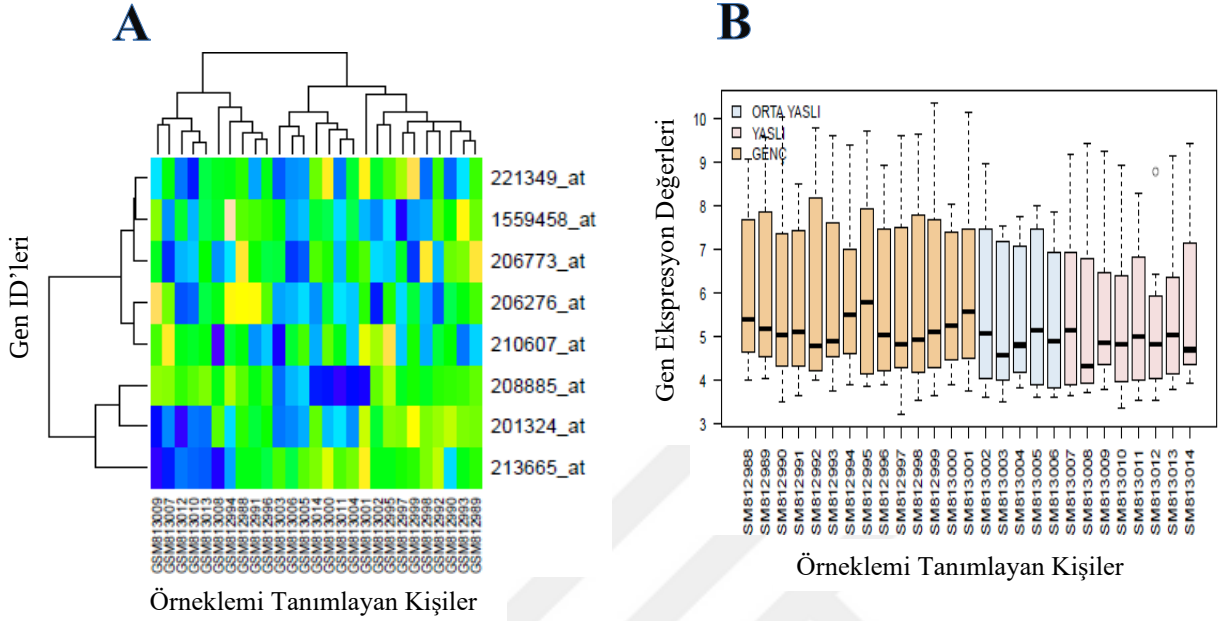
Şekil 7'de miyeloid yanlılık gösteren genlerin ısı haritası ve kutu grafiği A ve B şeklinde ikiye ayrılmıştır. A grafiğinde yer alan renkler ışık spektrumundaki dalga boyu değerleri ile nitelendirilmiştir (kırmızıdan mora azalan gen ekspresyonunu ifade eder). Grafiklerin analizi sonucunda yaşa bağlı olarak artan miyeloid yanlılık gözlemlenmiştir.

Isı haritasında ID'leri ile gösterilen genler sırasıyla, CSF2RB, MLF2, MLLT10, HOXA9, F2RL1, FUT1, USP46, CCDC88A şeklindedir.

Bu bulgular, insan HKH'sinde miyeloid spesifikasyon genleri ifade etmektedir. Belirgin şekilde USP46, CCDC88A ve FUT1 sembolleri ile gösterilen gen ekspresyonlarının yaşla beraber arttığı görülmektedir. B grafiğinde de yukarıdaki gen seti üzerinden incelemeler sonucunda yaş ile birlikte gerçekleşen miyeloid yanlı

regülasyon gözlemlenmiştir.

4.3 Lenfoid Özellikli Gen Analizi



Şekil 4.3 : Lenfoid Yanlılık Gösteren Genlerin Isı Haritası ve Kutu Grafiği

Şekil 8’de lenfoid yanlılık gösteren genlerin ısı haritası ve kutu grafiği A ve B şeklinde ikiye ayrılmıştır. A grafiğinde yer alan renkler ışık spektrumundaki dalga boyu değerleri ile nitelendirilmiştir (kırmızıdan mora azalan gen ekspresyonunu ifade eder). Grafiklerin analizi sonucunda yaşa bağlı olarak azalan lenfoid yanlılık gözlemlenmiştir. HKH’nin yaşlı kemik iliğinde miyeloid farklılaşma kapasitesi ile birlikte genç kemik iliğine oranla azalma ortaya koyulmuştur.

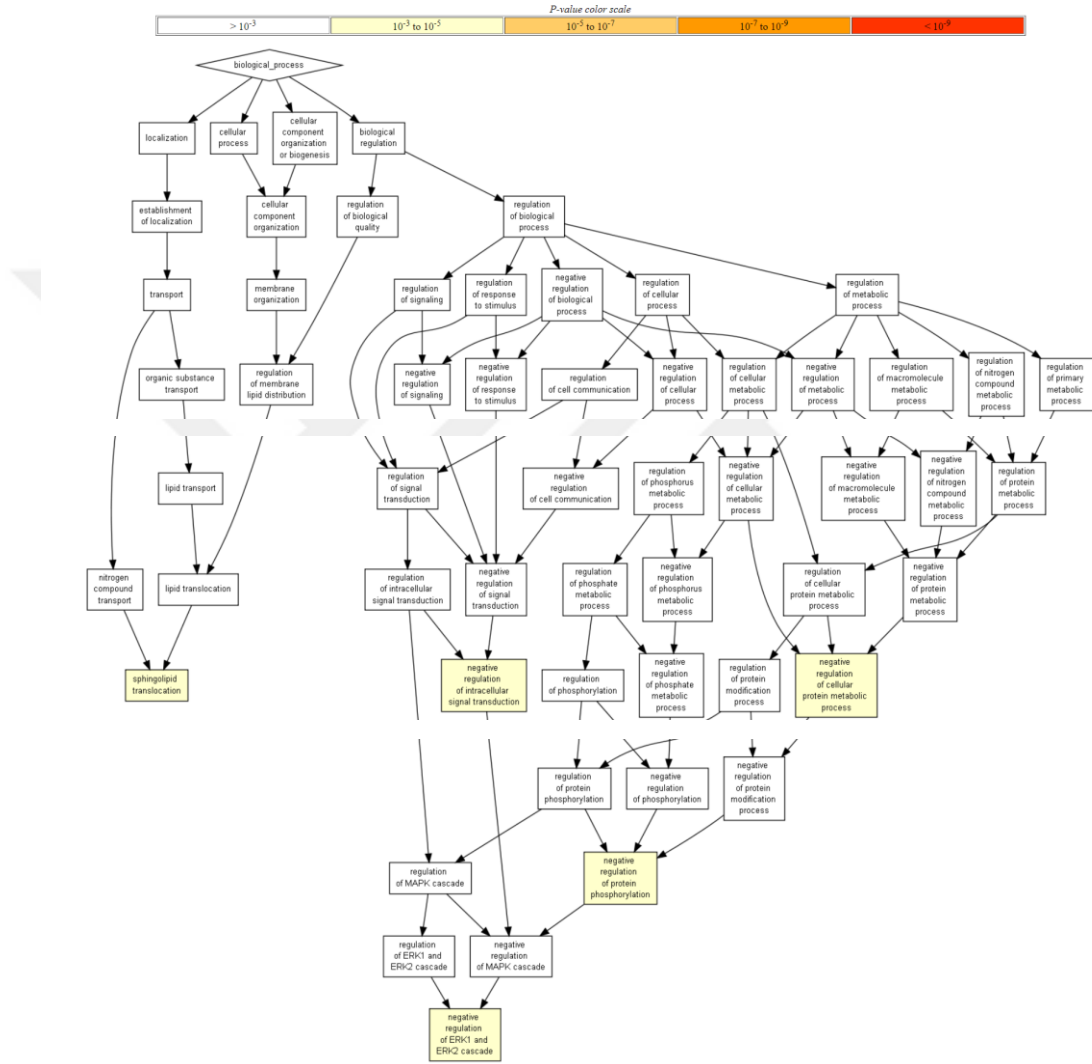
Isı haritasında ID’leri ile gösterilen genler sırasıyla, VPREB1, LTBR, LY6H, LY6D, FLT3LG, LCP1, EMP1, SOX4 şeklindedir.

Bu bulgular, insan HKH’sinde lenfoid spesifikasyon genleri ifade etmektedir. Belirgin şekilde EMP1 ve SOX4 sembolleri ile gösterilen gen ekspresyonlarının yaşla beraber arttığı görülmektedir. B grafiğinde de yukarıdaki gen seti üzerinden incelemeler sonucunda yaş ile birlikte gerçekleşen lenfoid yanlı regülasyon gözlemlenmiştir.

4.4 GOrilla Analizi

Yapılan ANOVA testi sonucunda gen ekspresyon ifadelerinin insan HKH’nin kümülatif karşılaştırılmasıyla anlamlı olduğu ortaya çıkarılmıştır. Benjamini

Hochberg prensibi kullanılarak 50.000 adet gen profili $p < 0.05$ anlamlı ifade değerine sahip olacak şekilde ele alınarak 570 gen sınıfına indirgenmiştir. Anlamlı gen dizisi GOrilla arayüzünde arka plan gen setiyle desteklenerek hedef kümede zenginleştirilme işlemi yapılmıştır. Şekil 9'da GOrilla'da 570 gen sınıfı setinin zenginleştirme işlemi ile p değeri spektrumuna göre biyolojik süreçlerin çıktısı elde edilmiştir.



Şekil 4.4 : GOrilla Hedef Gen Dizisinde Gerçekleşen Biyolojik Süreçler

GOrilla ortaya çıkardığı genleri ve onların biyolojik süreçlerde regüle durumları hakkında bize bilgi sunar. Bu durum tabloda gözüktüğü gibi süreçlerde geçen 96 gen toplamda 32 geni ifade etmektedir.

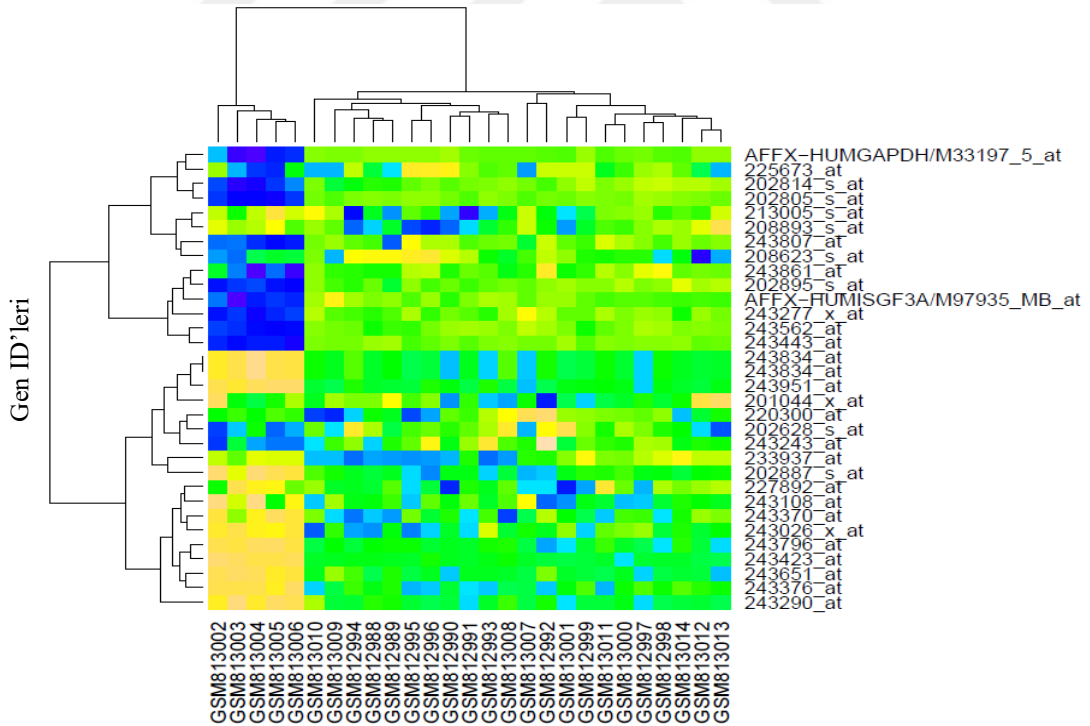
Çizelge 4.1 : Şekil 9'daki Biyolojik Süreçlerde Ortaya Çıkan Genler

GO Term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0099039	<u>s</u> fm ^g olipit <u>t</u> ranslokasyonu	2.11E-4	1E0	68.66 (16546,2,241,2)	ABCC1 - <u>a</u> tp bağlayıcı kaset, alt aile c (<u>c</u> ffr / <u>m</u> rp), üye 1 ABCB1 - <u>a</u> tp bağlayıcı kaset, alt aile b (<u>m</u> dr / <u>t</u> ap), üye 1
GO:0032269	hücrel protein <u>m</u> etabolik sürecinin negatif düzenlenmesi	3.12E-4	1E0	2.02 (16546,951,241,28)	PRKAA2 - Protein kinaz, <u>a</u> mp aktive edilmiş, alfa 2 katalitik alt ünite SIRPA - sinyal düzenleyici protein alfa DDIT4 - DNA hasarı <u>u</u> yarımlı transkript 4 XIAP - <u>x</u> 'e bağlı <u>a</u> poptoz inhibitörü GGBP2 - <u>g</u> ametogenetin bağlayıcı protein 2 TNRC6A - 6a içeren <u>t</u> rinükleotit tekrarı HEXIM1 - indüklenbilir <u>h</u> ekzametilen <u>b</u> is-asetamid 1 NCOA7 - nükleer reseptör koaktivatörü 7 PME1 - protein <u>f</u> osfataz <u>m</u> etilesteraz 1 FNIP1 - <u>f</u> olikülünü etkileyen protein 1 TNIP1 - <u>t</u> nfap3 etkileşimli protein 1 CBL - cbl proto-onkogen, e3 ubiquitin protein ligası GPC3 - glipikan 3 DUSP6 - çift özgülüklü <u>f</u> osfataz 6 MECOM - mds1 ve evl1 karmaşık odağı EZR - <u>e</u> zrin HEXIM2 - <u>h</u> ekzametilen <u>b</u> is-asetamid ile indüklenbilir 2 PIP5KL1 - fosfatidilinositol-4-fosfat 5-kinaz benzeri 1 CAPRIN1 - hücre döngüsü ile ilişkili protein 1 SERPINA1 - serpin <u>p</u> eptidaz inhibitörü, <u>c</u> lade a (alfa-1 <u>a</u> ntiproteinaz, <u>a</u> ntitripsin), üye 1 MYADM - miyeloid ilişkili farklılaşma marker GAPDH - gliseraldehid-3-fosfat <u>d</u> ehidrogenaz KNG1 - kininogen 1 CPEB3 - sitoplazmik poliadensasyon elemanı bağlayıcı protein 3 RGS3 - g-protein sinyalizasyonunun düzenleyicisi 3 SERPINE1 - serpin <u>p</u> eptidaz inhibitörü, <u>c</u> lade e (<u>n</u> exin, <u>p</u> lazminojen <u>a</u> ktivator inhibitörü tip 1), üye 1 DUSP1 - çift özgülüklü <u>f</u> osfataz 1 RANBP9 - bağlayıcı protein 9
GO:0070373	ERK1 ve ERK2 <u>k</u> askadının negatif düzenlenmesi	3.32E-4	1E0	6.44 (16546,64,241,6)	SIRPA - sinyal düzenleyici protein alfa TNIP1 - <u>t</u> nfap3 etkileşimli protein 1 DUSP6 - çift özgülüklü <u>f</u> osfataz 6 DUSP1 - çift özgülüklü <u>f</u> osfataz 1 RANBP9 - bağlayıcı protein 9 EZR - <u>e</u> zrin
GO:0001933	<u>p</u> rotein <u>f</u> osforilasyonunun negatif düzenlenmesi	4.61E-4	1E0	2.72 (16546,379,241,15)	SIRPA - sinyal düzenleyici protein alfa DDIT4 - DNA hasarı <u>u</u> yarımlı transkript 4 TNIP1 - <u>t</u> nfap3 etkileşimli protein 1 CBL - cbl proto-onkogen, e3 ubiquitin protein ligası DUSP6 - çift özgülüklü <u>f</u> osfataz 6 GGBP2 - <u>g</u> ametogenetin bağlayıcı protein 2 MECOM - mds1 ve evl1 karmaşık odağı EZR - <u>e</u> zrin HEXIM2 - <u>h</u> ekzametilen <u>b</u> is-asetamid ile indüklenbilir 2 HEXIM1 - indüklenbilir <u>h</u> ekzametilen <u>b</u> is-asetamid 1 PIP5KL1 - fosfatidilinositol-4-fosfat 5-kinaz benzeri 1 MYADM - miyeloid ilişkili farklılaşma marker RGS3 - g-protein sinyalizasyonunun düzenleyicisi 3 DUSP1 - çift özgülüklü <u>f</u> osfataz 1 RANBP9 - bağlayıcı protein 9
GO:1902532	hücre içi sinyal iletiminin negatif düzenlenmesi	7.26E-4	1E0	2.42 (16546,482,241,17)	PRKAA2 - Protein kinaz, <u>a</u> mp aktive edilmiş, alfa 2 katalitik alt ünite SIRPA - sinyal düzenleyici protein alfa DDIT4 - DNA hasarı <u>u</u> yarımlı transkript 4 TNIP1 - <u>t</u> nfap3 etkileşimli protein 1 DUSP6 - çift özgülüklü <u>f</u> osfataz 6 GGBP2 - <u>g</u> ametogenetin bağlayıcı protein 2 MECOM - mds1 ve evl1 karmaşık odağı STAT1 - sinyal <u>t</u> ransdüseri ve transkripsiyon 1, 91kda'nın <u>a</u> ktivatörü EZR - <u>e</u> zrin WWC1 - 1 içeren <u>w</u> w ve <u>c</u> 2 alanı KANK1 - kn motif ve <u>a</u> nkirin tekrar alanları 1 MYADM - miyeloid ilişkili farklılaşma marker RGS3 - g-protein sinyalizasyonunun düzenleyicisi 3 DUSP1 - çift özgülüklü <u>f</u> osfataz 1

Çizelge 4.1 : Şekil 9'daki Biyolojik Süreçlerde Ortaya Çıkan Genler (devam)

					TANK - traf aile üyesi ile ilişkili nfkb aktivatörü RANBP9 - bağlayıcı protein 9 FNIP1 - folikülünü etkileyen protein 1
GO:0051248	Protein metabolik sürecinin negatif düzenlenmesi	8.4E-4	1E0	1.90 (16546,1013,241,28)	PRKAA2 - Protein kinaz, amp aktive edilmiş, alfa 2 katalitik alt ünite SIRPA - sinyal düzenleyici protein alfa DDIT4 - DNA hasarı uyarımı transkript 4 XIAP - x'e bağlı apoptoz inhibitörü GGBP2 - gametogenezin bağlayıcı protein 2 TNRC6A - 6a içeren trinükleotit tekrarı HEXIM1 - indüklenbilir hekzametilen bis-asetamid 1 NCOA7 - nükleer reseptör koaktivatörü 7 PPME1 - protein fosfataz metilesteraz 1 FNIP1 - folikülünü etkileyen protein 1 TNIP1 - tnfai3 etkileşimli protein 1 CBL - cbl proto-onkogen. e3 ubiquitin protein ligazı GPC3 - glipikan 3 DUSP6 - çift özgülüklü fosfataz 6 MECOM - mds1 ve evl karmaşık odağı EZR - ezrin HEXIM2 - hekzametilen bis-asetamid ile indüklenbilir 2 PIP5KL1 - fosfatidilinositol-4-fosfat 5-kinaz benzeri 1 CAPRN1 - hücre döngüsü ile ilişkili protein 1 SERPINA1 - serpin peptidaz inhibitörü, clade a (alfa-1 antiproteinaz , antitripsin), üye 1 MYADM - miyeloid ilişkili farklılaşma marker GAPDH - gliseraldehit-3-fosfat dehidrojenaz KNG1 - kininojen 1 CPEB3 - sitoplazmik poliadenuilasyon elemanı bağlayıcı protein 3 RGS3 - g-protein sinyalizasyonunun düzenleyicisi 3 SERPINE1 - serpin peptidaz inhibitörü, clade e (nexin , plazminojen aktivatör inhibitörü tip 1), üye 1 DUSP1 - çift özgülüklü fosfataz 1 RANBP9 - bağlayıcı protein 9

Tablo 3'de GOrilla'da yapılan zenginleştirme çalışması sonucunda orta çıkan genler verilmiştir.



Şekil 4.5 : Çizelge 3.1'deki Genlerin Isı Haritası

Şekil 10'da GOrilla'dan elde edilen genlerin pozitif ve negatif regüle oldukları durumlar ifade edilmiştir. Renkler ışık spektrumundaki dalga boyu değerleri ile nitelendirilmiştir (kırmızıdan mora azalan gen ekspresyonunu ifade eder).

4.5 Makine Öğrenmesi Çıktıları ve Gen İfadelerinin Analizi

Makine öğrenmesinde, sınıflandırma algoritmalarının başarımının ölçülmesi için hata matrisi kullanılmaktadır. Çıktının iki veya daha fazla sınıf olabileceği makine öğrenimi sınıflandırma problemi için performans ölçümüdür. Tahmin edilen ve gerçek değerlerin 4 farklı değerlendirmeden oluşmaktadır (Visa ve diğ., 2011). Sınıflandırma probleminin hata matrisi bazı terimleri içermektedir:

Doğru Pozitif (DP): Pozitif sınıfa ait bir örneğin doğru şekilde sınıflandırılmasını ifade eder (Vujović, 2021).

Doğru Negatif (DN): Negatif sınıfa ait bir örneğin doğru şekilde sınıflandırılmasını ifade eder (Vujović, 2021).

Yanlış Pozitif (YP): Negatif sınıfa ait olan ancak yanlış bir şekilde pozitif sınıfa ait olarak sınıflandırılan bir değeri ifade eder (Vujović, 2021).

Yanlış Negatif (YN): Pozitif sınıfa ait olan ancak yanlış bir şekilde negatif sınıfa ait olarak sınıflandırılan bir değeri ifade eder (Vujović, 2021).

Sınıflandırma problemlerinde, doğru tahmin edilen ifadeler ve yanlış tahmin edilen ifadeler şeklinde iki tane sonuç vardır. DP ve DN ifadeleri doğru tahminleri ifade etmektedir. Bu çalışmanın veri setindeki sınıflandırma grupları miyeloid ve lenfoid sınıflarıdır. Gerçekte miyeloid grubuna ait bir gen ifadesi lenfoid sınıfına dahil edildiyse YP ile, lenfoid grubuna ait gen ifadesi miyeloid sınıfta tanımlandıysa YN ile temsil edilmektedir (Vujović, 2021).

Hata matrisinde DP, DN, YP, YN değerlerine bağlı niceliksel değerlendirme metrikleri elde edilmektedir:

Doğruluk: Veri setinde doğru sınıflandırma toplamının, tüm toplama oranı şeklindedir. Doğru pozitif (DP) ve doğru negatif (DN) değerlerin tüm olasılıkların toplamına oranı doğruluk değerini verir (Grandini ve diğ., 2020).

Kesinlik: Gerçekte pozitif sınıfa ait olan örneklerin sayısı, pozitif sınıfa ait olduğu tahmin edilen tüm örneklerle oranı şeklindedir. Doğru pozitif değerlerin (DP), doğru pozitif (DP) ve yanlış pozitiflerin (YP) toplamına oranı kesinlik değerini verir (Grandini ve diğ., 2020).

Hassasiyet: Gerçekte pozitif sınıfa ait olan örneklerin sayısı, pozitif sınıfa ait olduğu doğru tahmin edilen örneklere oranı şeklindedir. Doğru pozitif değerlerin (DP), doğru pozitif (DP) ve yanlış negatiflerin (YN) toplamına oranı duyarlılık değerini verir (Grandini ve diğ., 2020).

Total Observations in Table: 114

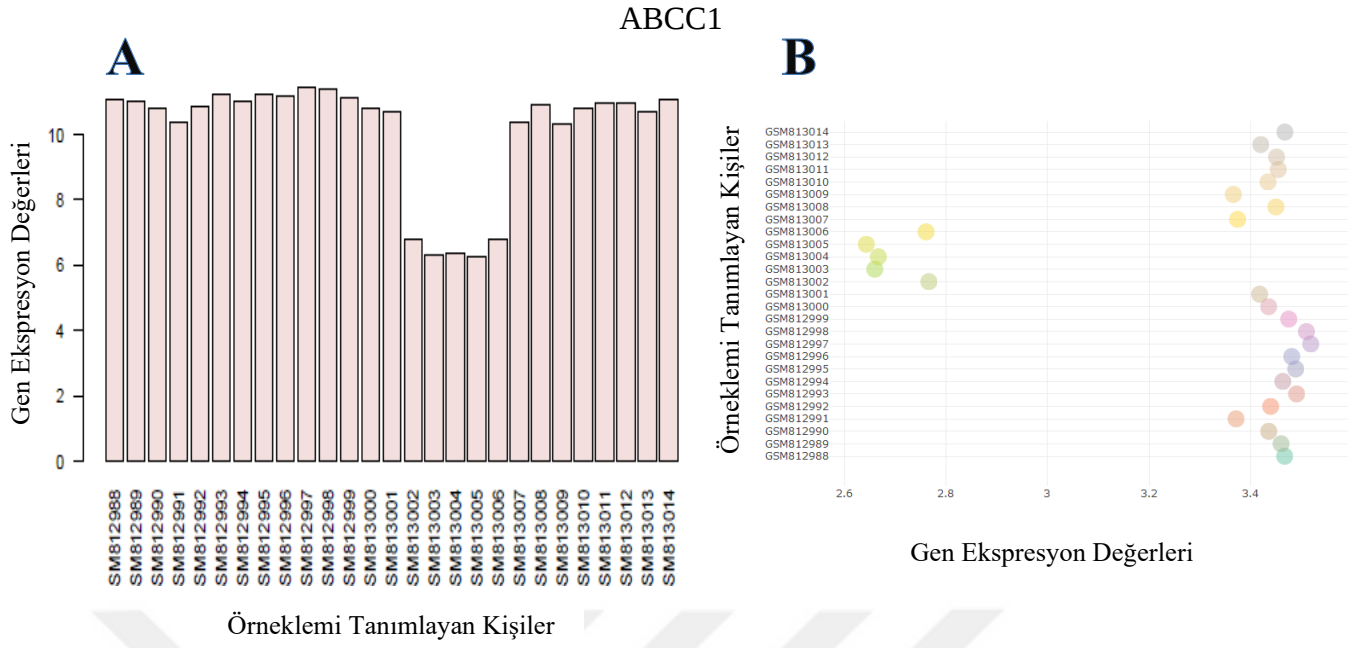
		Tahmin Edilen		
Veri Seti		Miyeloid	Lenfoid	Toplam Satır
Gerçek Değer	Miyeloid	72 11.166 0.960 0.973 0.632	3 20.658 0.040 0.075 0.026	75 0.658
	Lenfoid	2 21.474 0.051 0.027 0.018	37 39.727 0.949 0.925 0.325	39 0.342
	Toplam Kolon	74 0.649	40 0.351	114

Şekil 4.6 : Hata Matrisi

Gen ekspresyon verileri KNN algoritması ile sınıflandırılmıştır. Veri setinin %80'lik kısmı eğitim, %20'lik kısmı ise test için kullanılmıştır. Test verisi üzerinden miyeloid ve lenfoid yanlı hücre durumları makine öğrenmesi metoduyla tahmin edilmiştir. Algoritmanın doğruluk oranı %95,6, kesinlik oranı %97,2 ve hassasiyet oranı %96 olarak hesaplanmıştır.

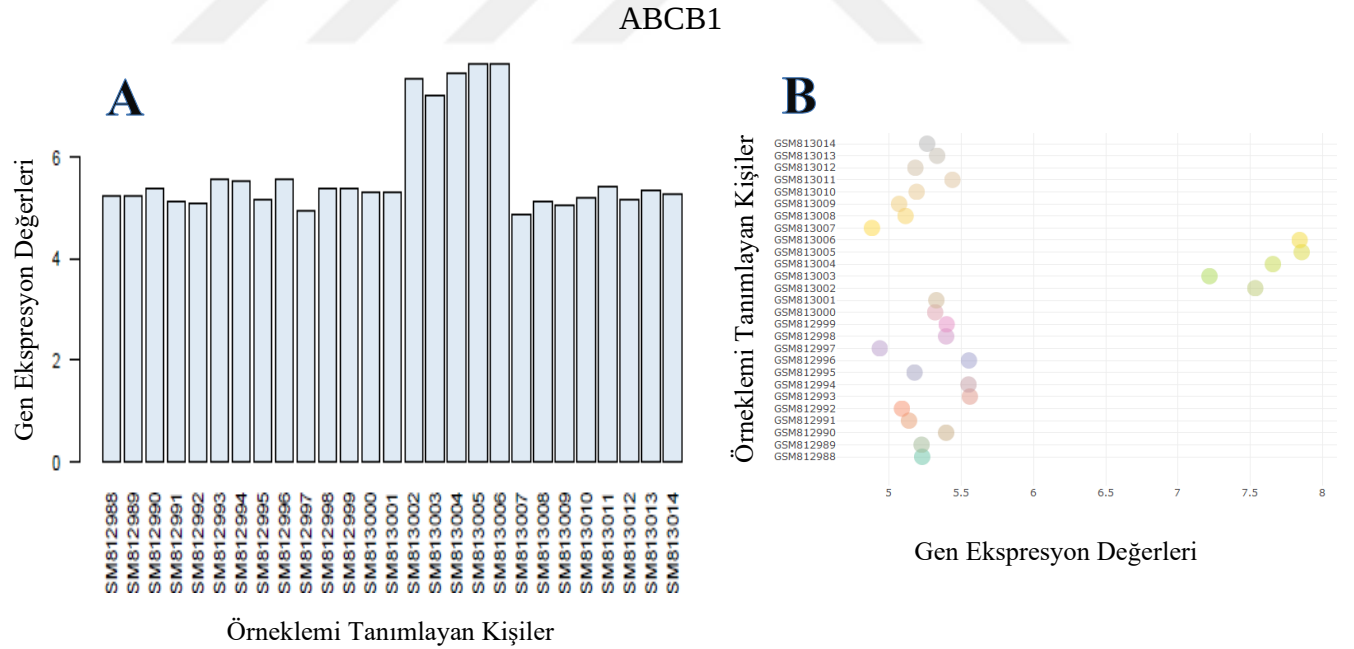
Çizelge 4.2 : Hata Matrisi Metrikleri ve Elde Edilen Değerler

Algoritma	Doğruluk	Kesinlik	Hassasiyet
KNN	%95,6	%97,2	%96



Şekil 4.7 : ABCC1 Geninin Ekspresyon Değerlerinin Grafikleri

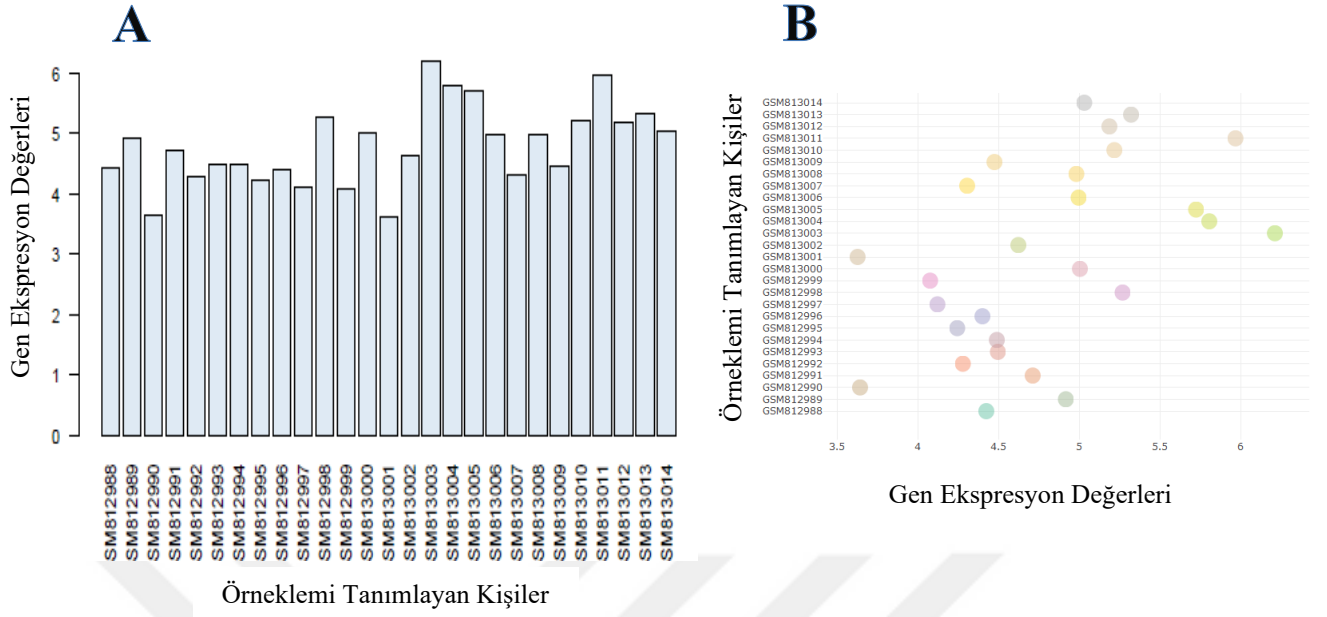
A grafiğinde ABCC1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



Şekil 4.8 : ABCB1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde ABCB1 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.

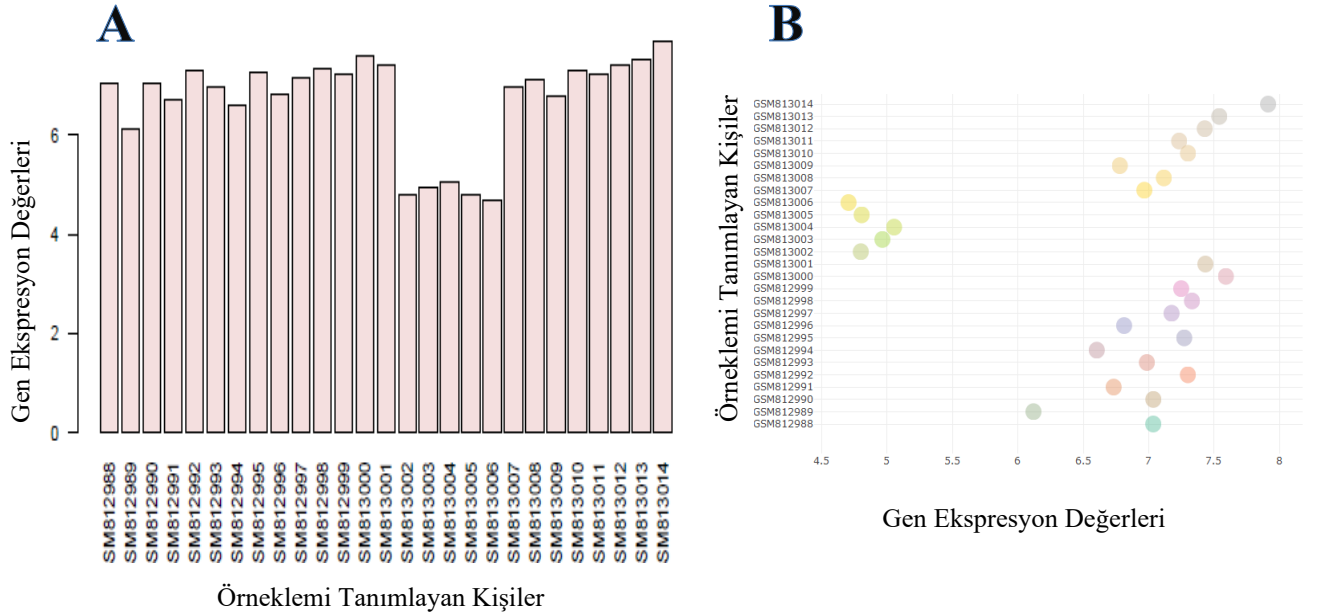
PRKAA2



Şekil 4.9 : PRKAA2 Geninin Ekspresyon Değerlerinin Grafikleri

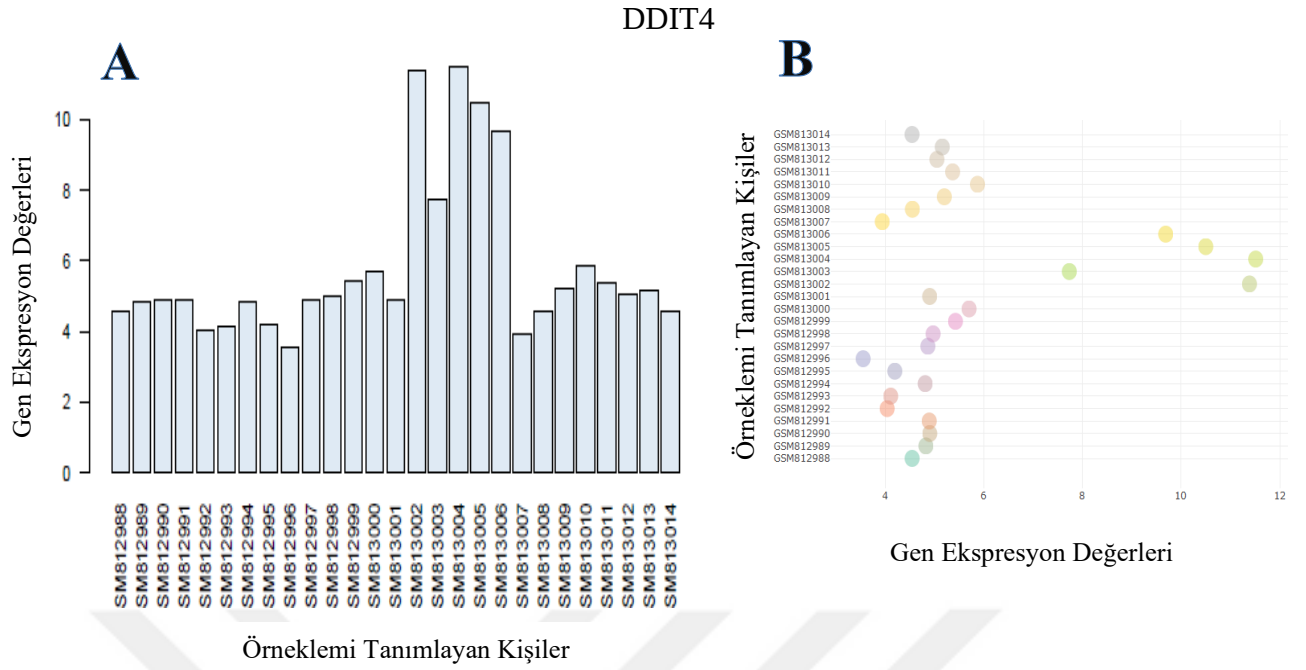
A grafiğinde PRKAA2 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.

SIRPA



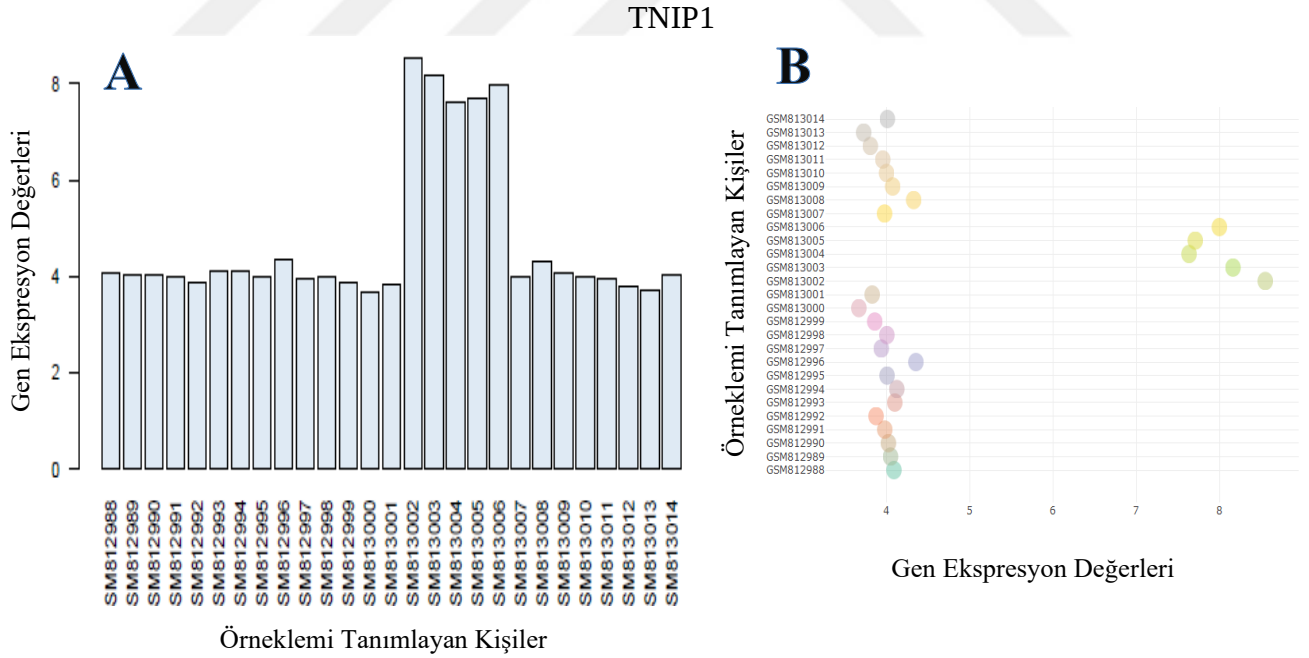
Şekil 4.10 : SIRPA Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde SIRPA geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



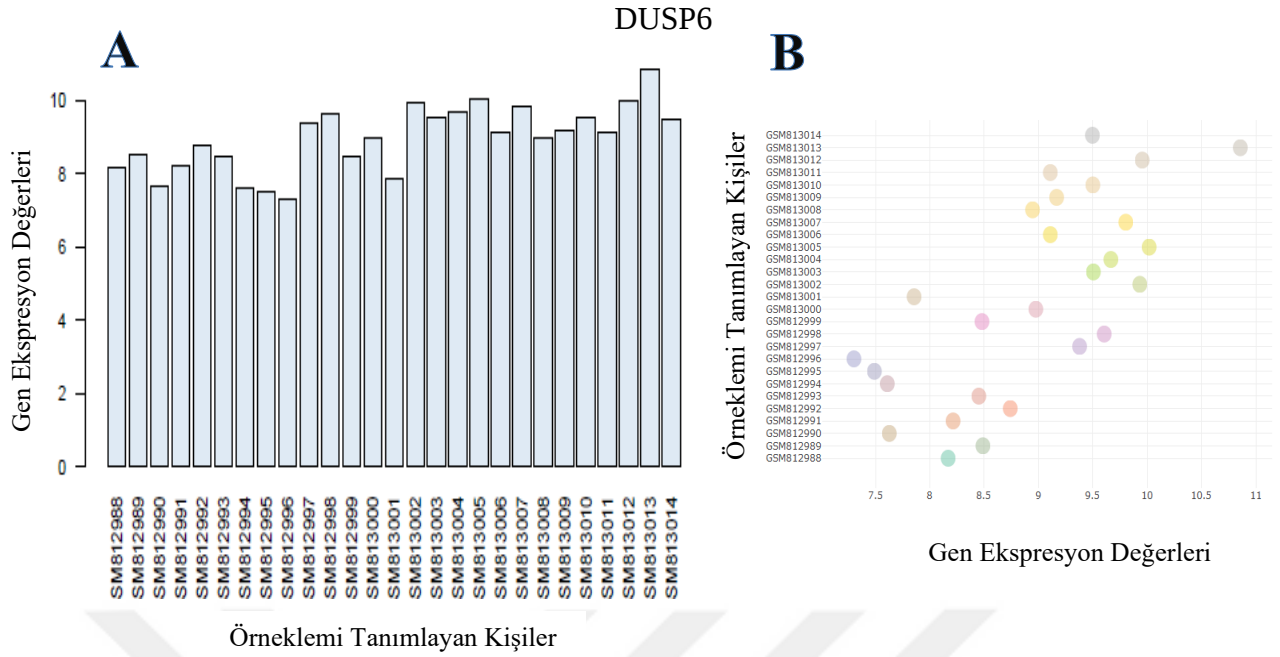
Şekil 4.11 : DDIT4 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde DDIT4 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



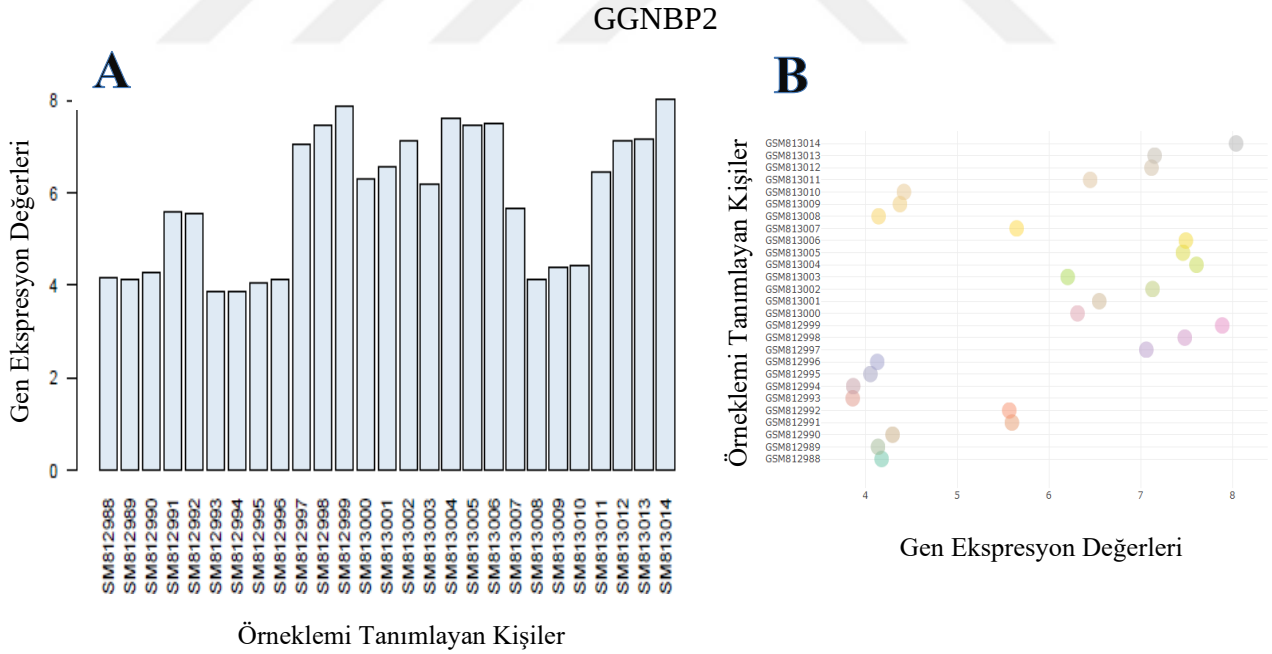
Şekil 4.12 : TNIP1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde TNIP1 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



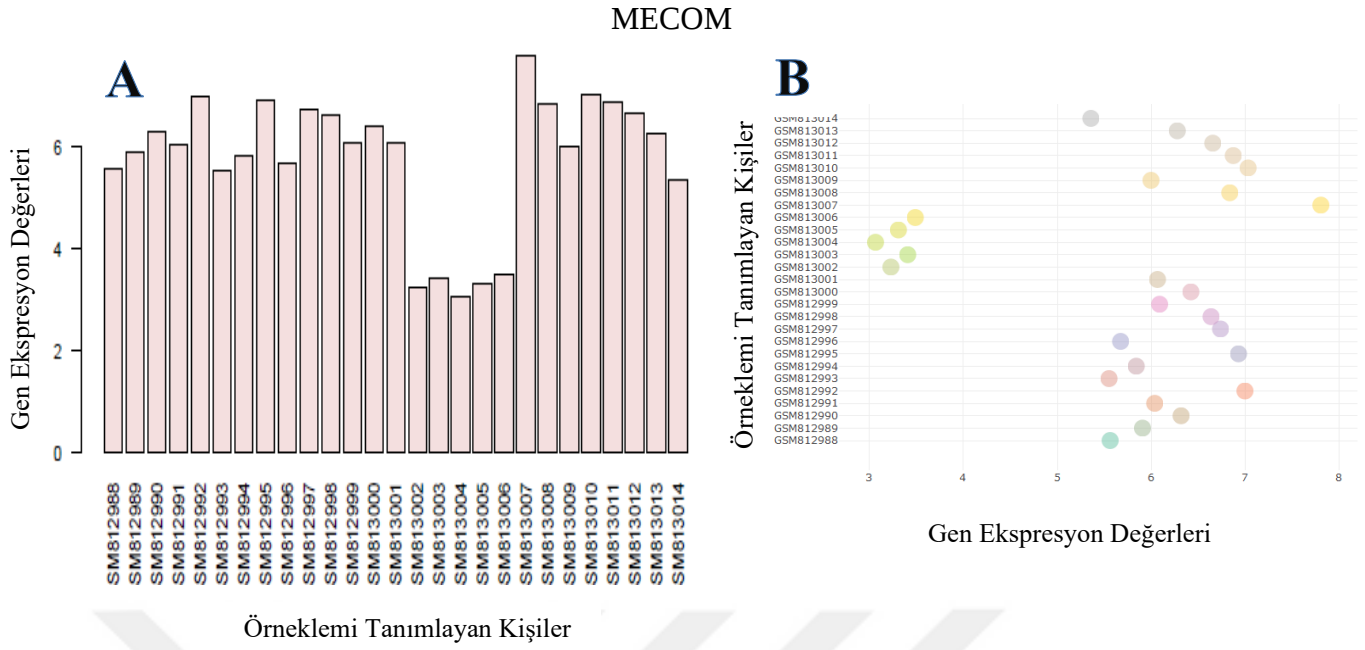
Şekil 4.13 : DUSP6 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde DUSP6 geninin yaşlandıkça yukarı regüle olduğu gözlemlenmiştir. B grafiğinde de yaşa bağlı lineer olarak ekspresyon artışı görülmektedir.



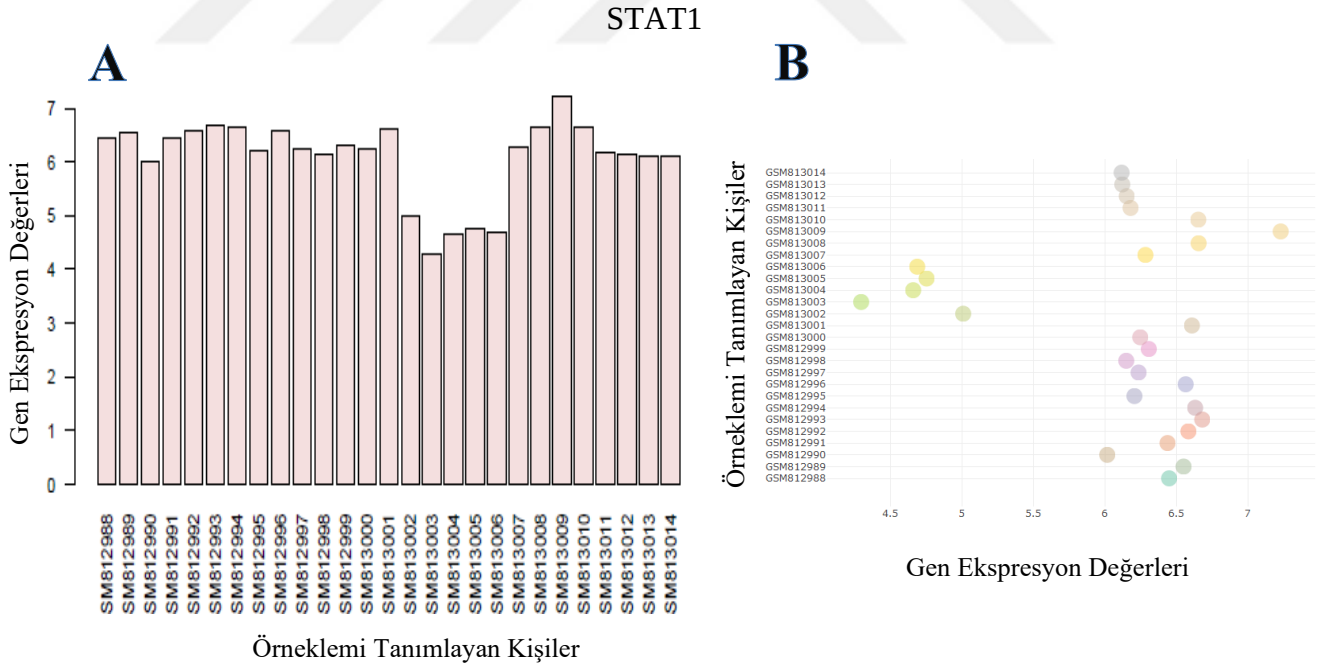
Şekil 4.14 : GGNBP2 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde GGNBP2 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



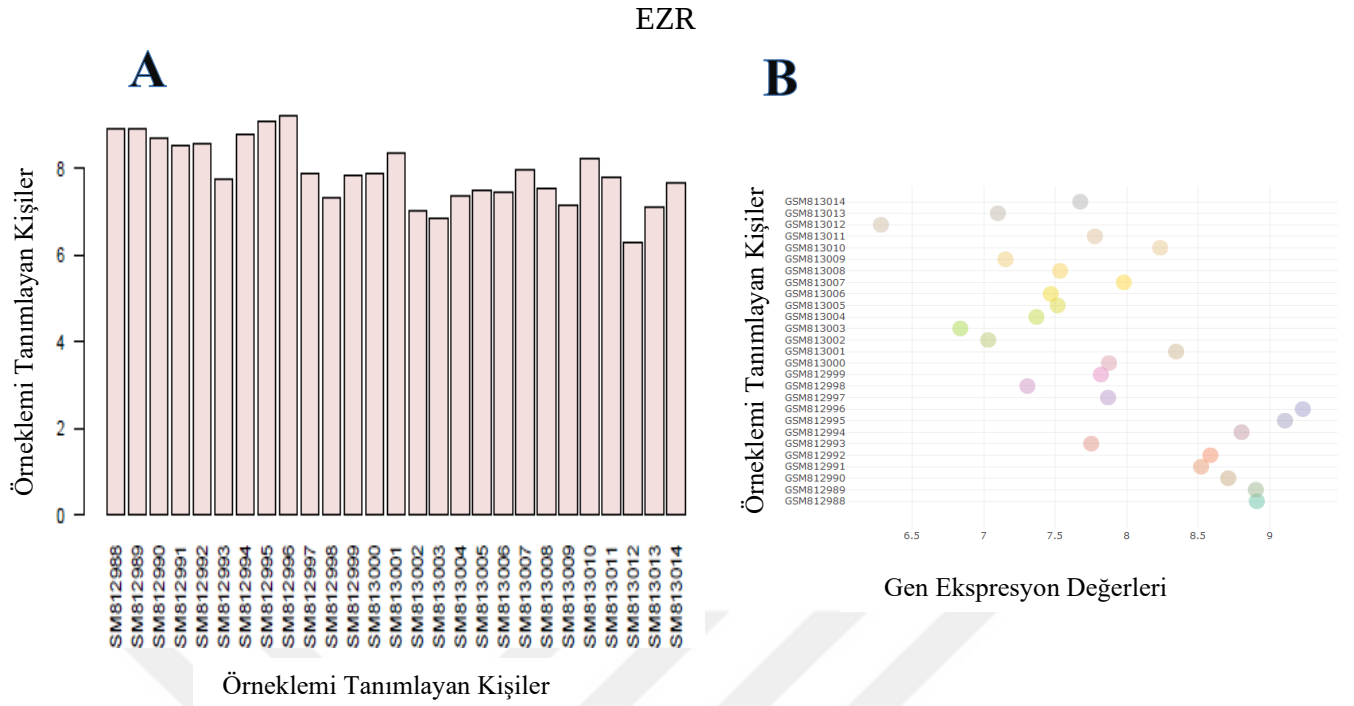
Şekil 4.15 : MECOM Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde MECOM geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



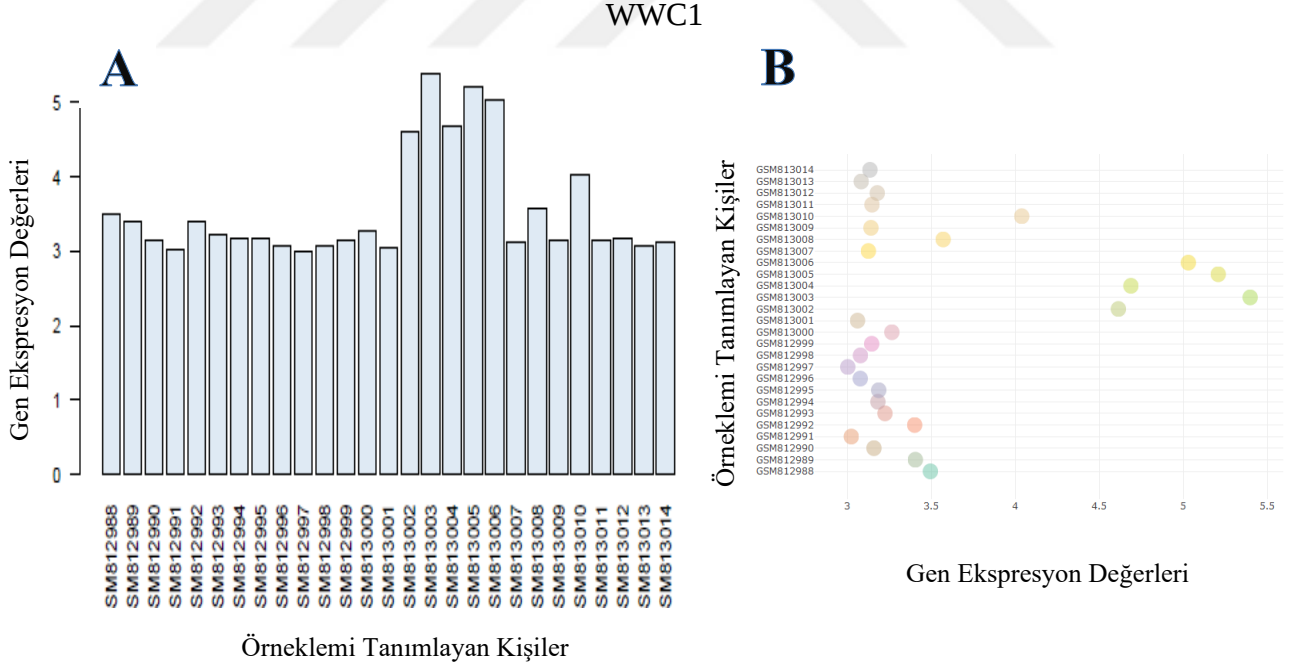
Şekil 4.16 : STAT1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde STAT1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



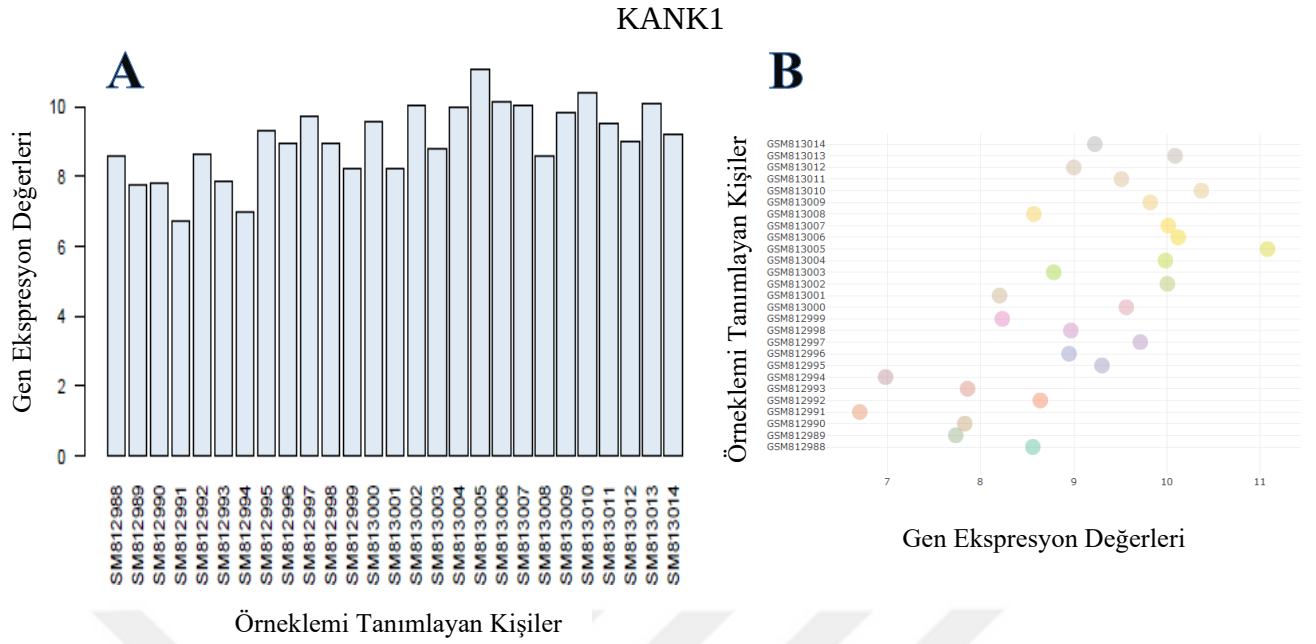
Şekil 4.17 : EZR Geninin Ekspresyon Değerlerinin Grafikleri

EZR geninin A grafiğinde yaşlandıkça aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon azalışı görülmektedir.



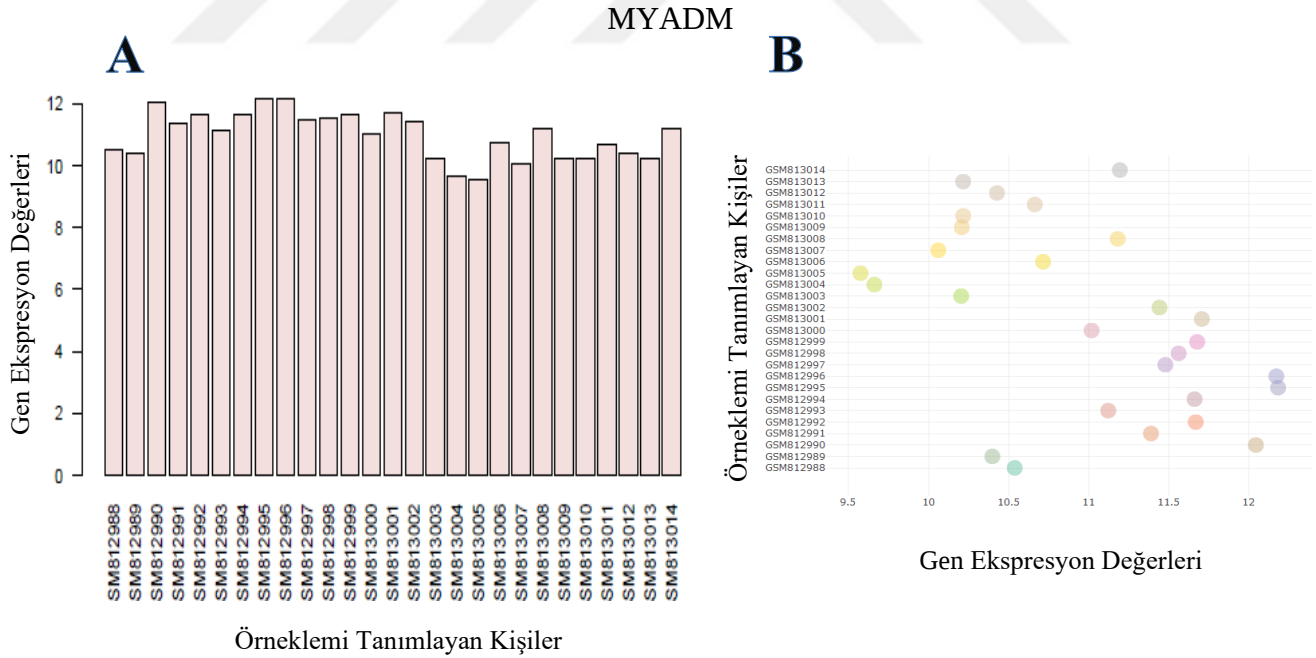
Şekil 4.18 : WWC1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde WWC1 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



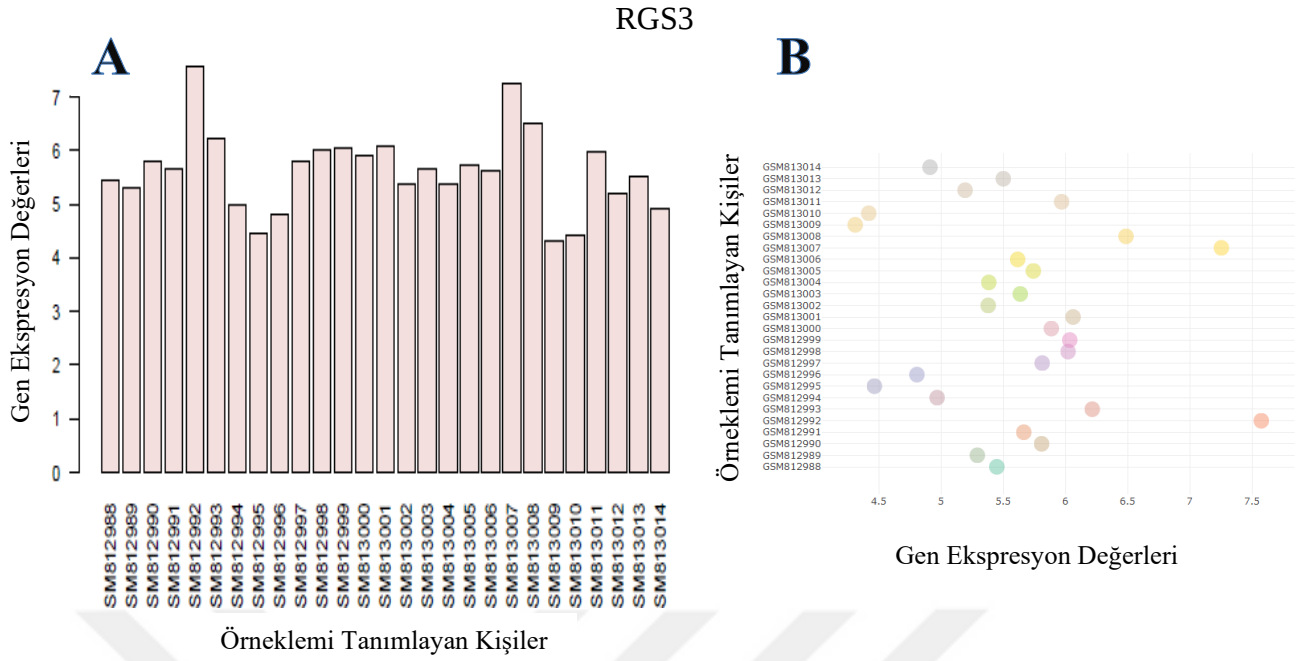
Şekil 4.19 : KANK1 Geninin Ekspresyon Değerlerinin Grafikleri

KANK1 geninin A grafiğinde yaşlandıkça yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon artışı görülmektedir.



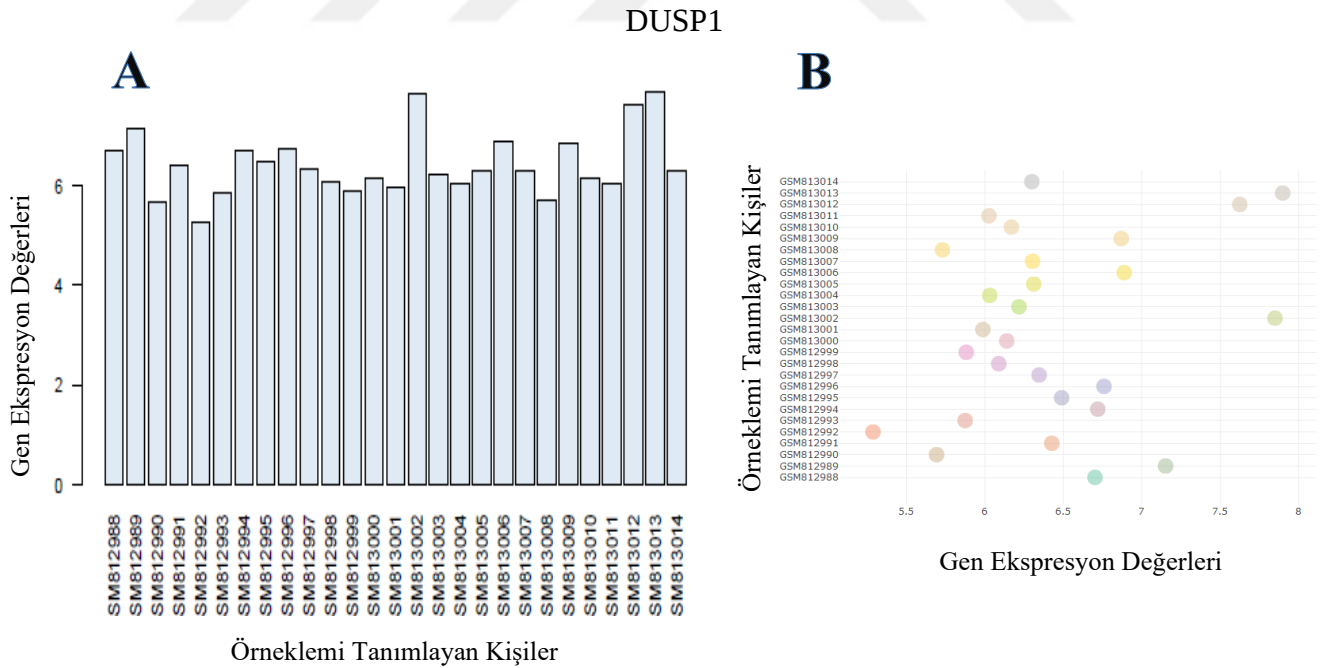
Şekil 4.20 : MYADM Geninin Ekspresyon Değerlerinin Grafikleri

MYADM geninin A grafiğinde yaşlandıkça aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon azalışı görülmektedir.



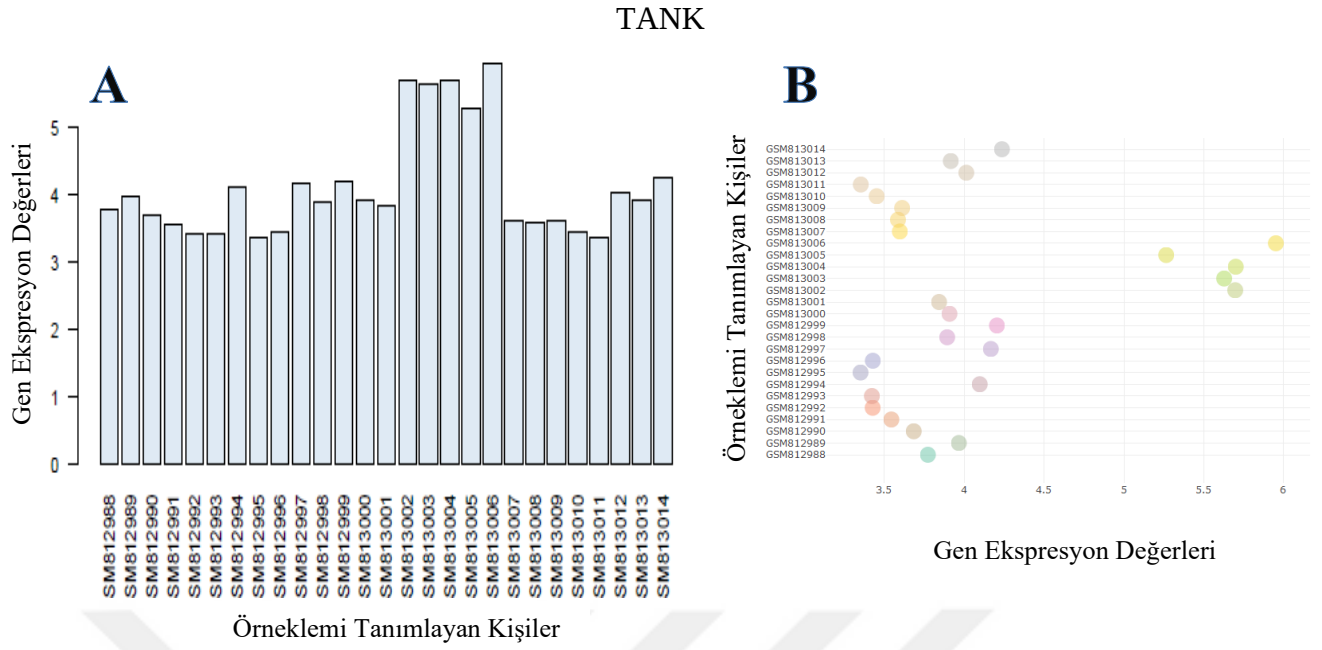
Şekil 4.21 : RGS3 Geninin Ekspresyon Değerlerinin Grafikleri

RGS3 geninin A grafiğinde yaşlandıkça aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon azalışı görülmektedir.

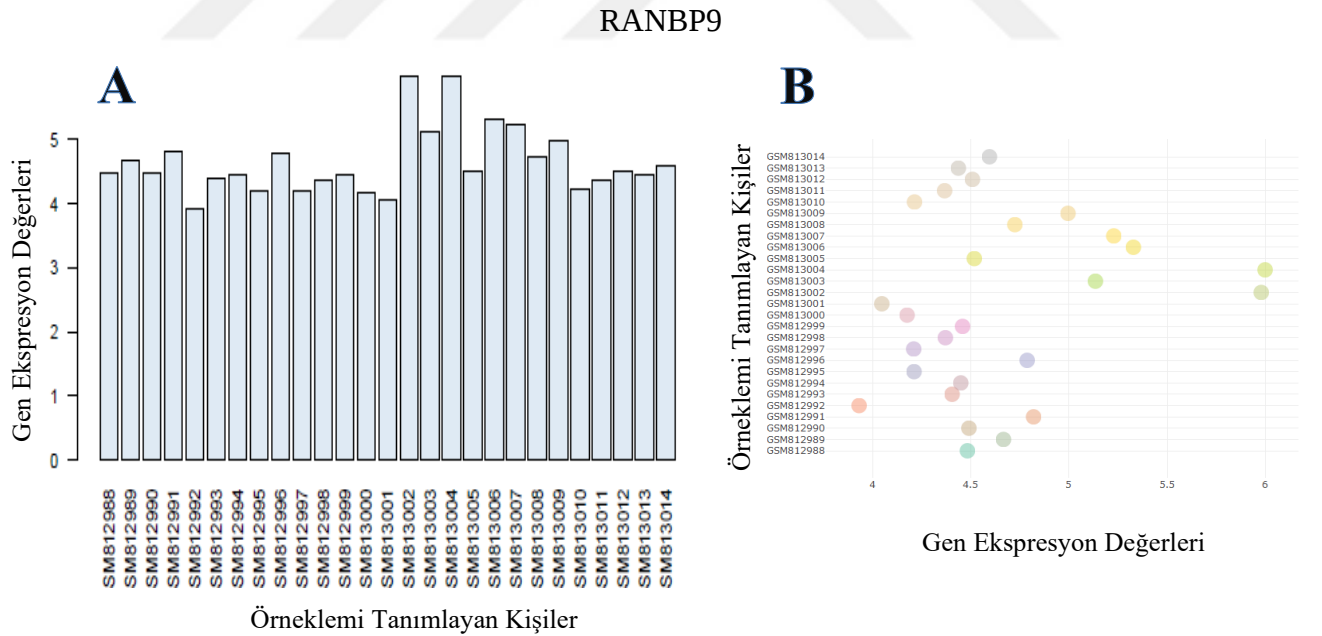


Şekil 4.22 : DUSP1 Gen Ekspresyon Değerlerinin Grafikleri

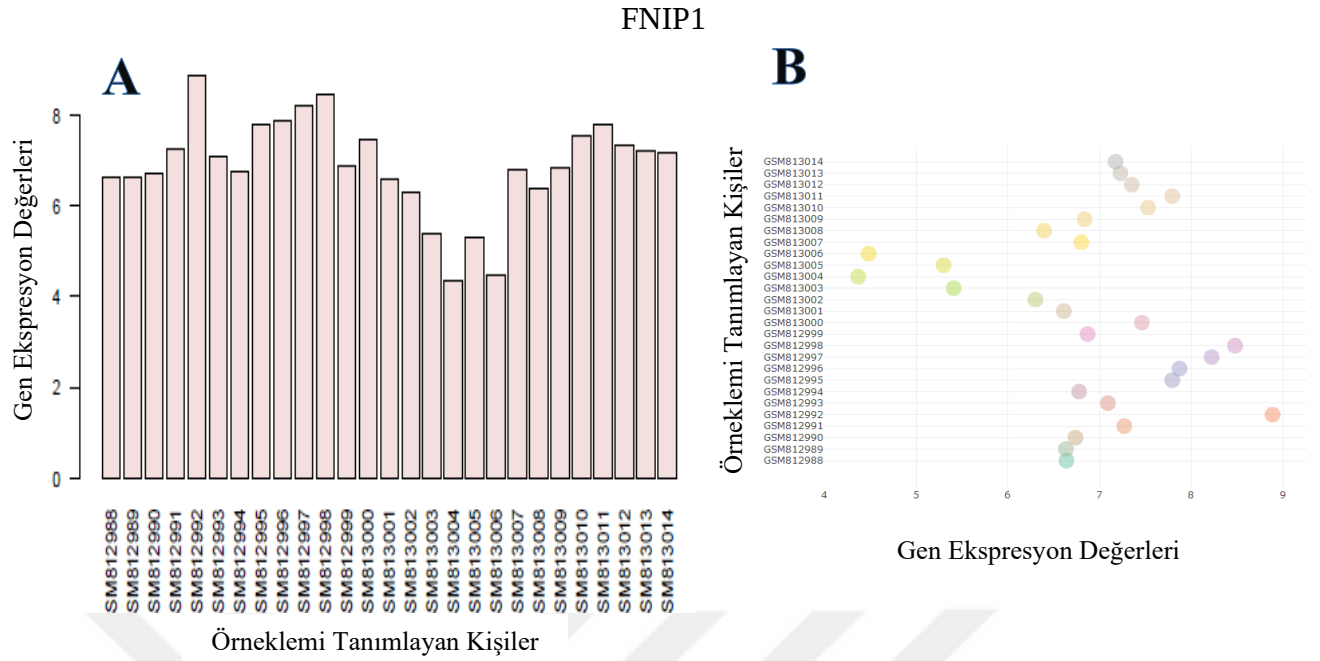
DUSP1 geninin A grafiğinde yaşlandıkça yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon artışı görülmektedir.



A grafiğinde TANK geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.

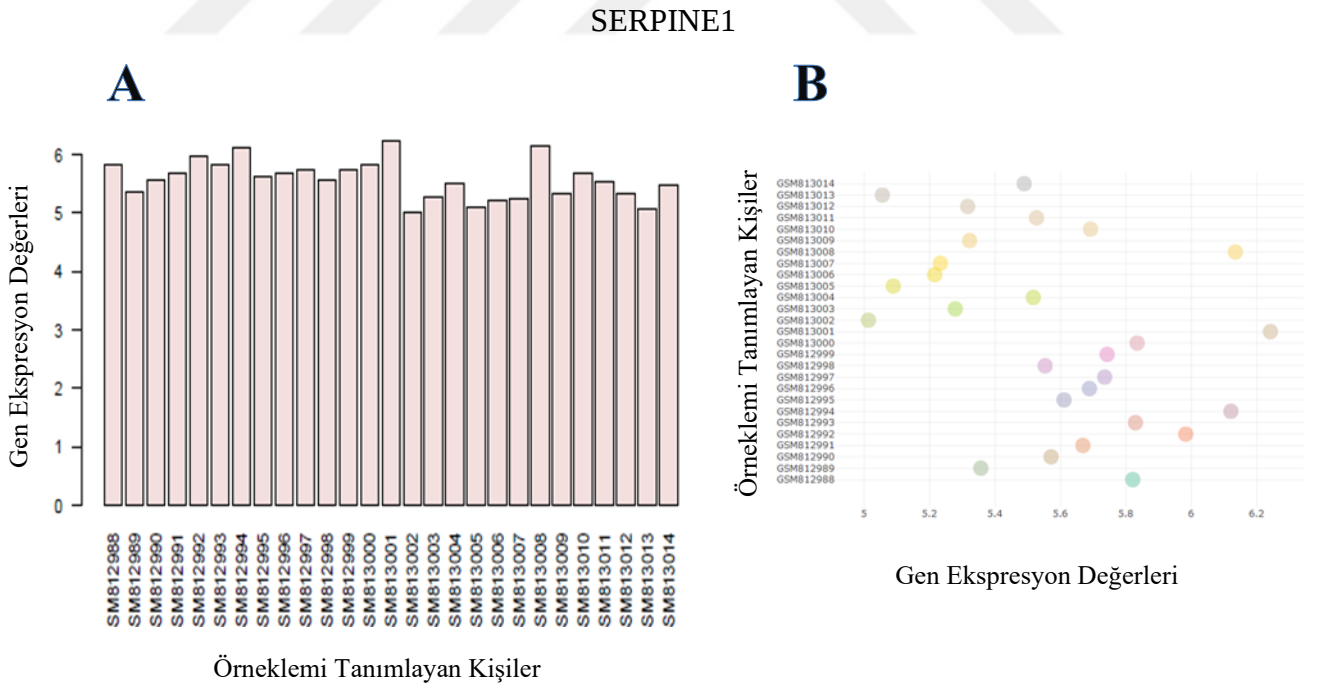


A grafiğinde RANBP9 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



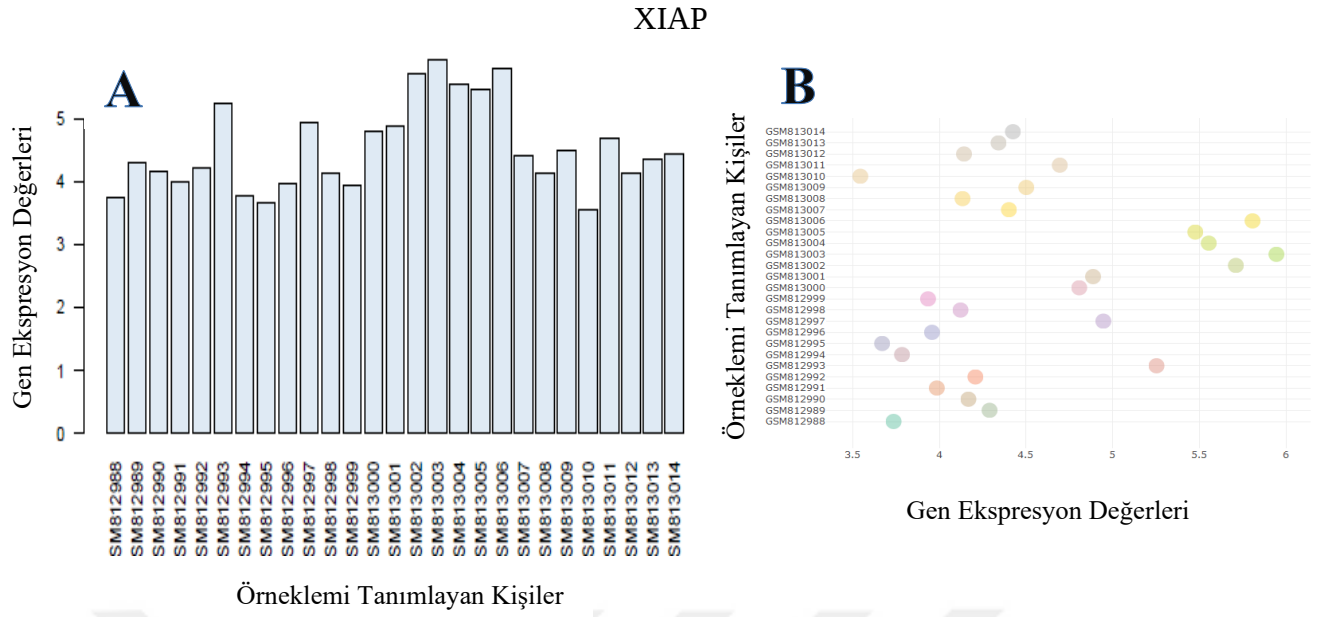
Şekil 4.25 : FNIP1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde FNIP1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



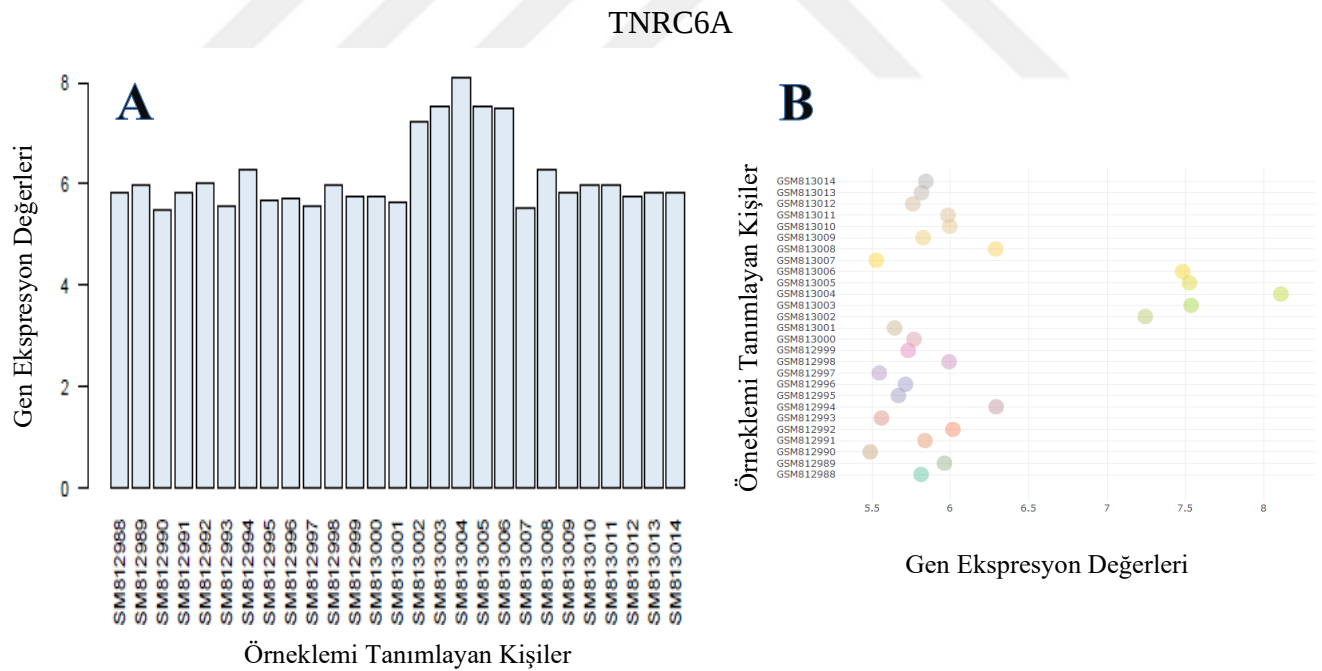
Şekil 4.26 : SERPINE1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde SERPINE1 geninin yaşlandıkça aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı kısmen lineer olarak ekspresyon azalışı görülmektedir.



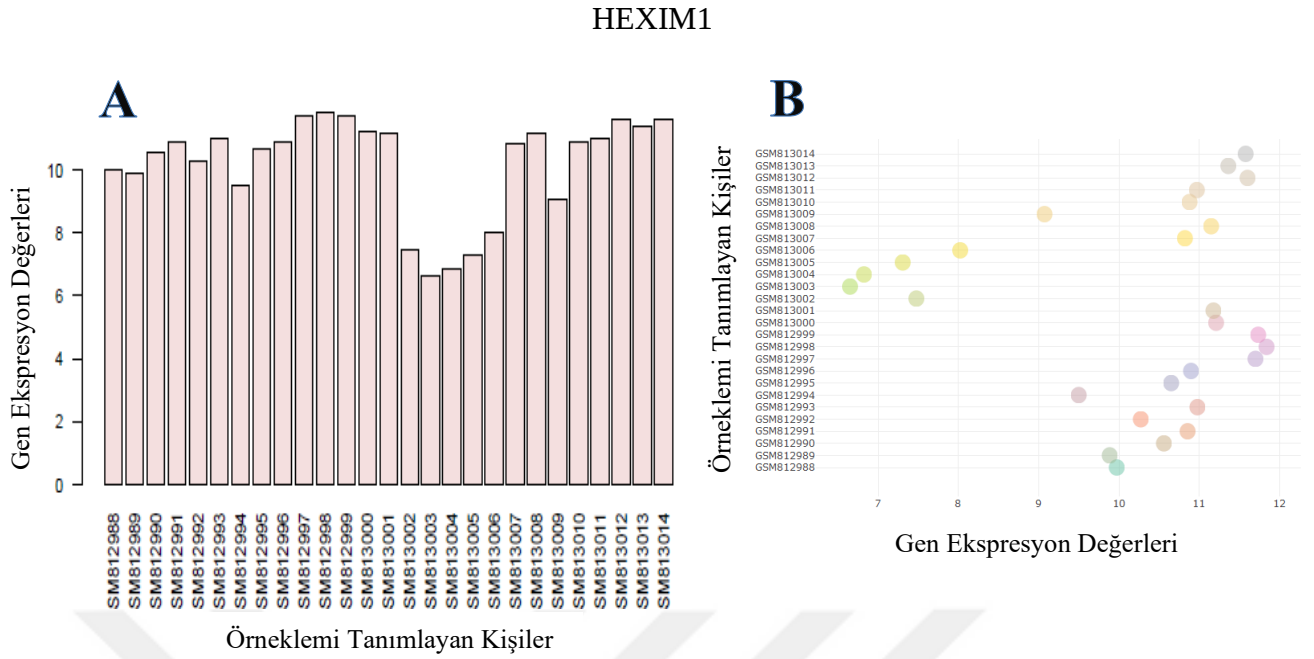
Şekil 4.27 : XIAP Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde XIAP geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



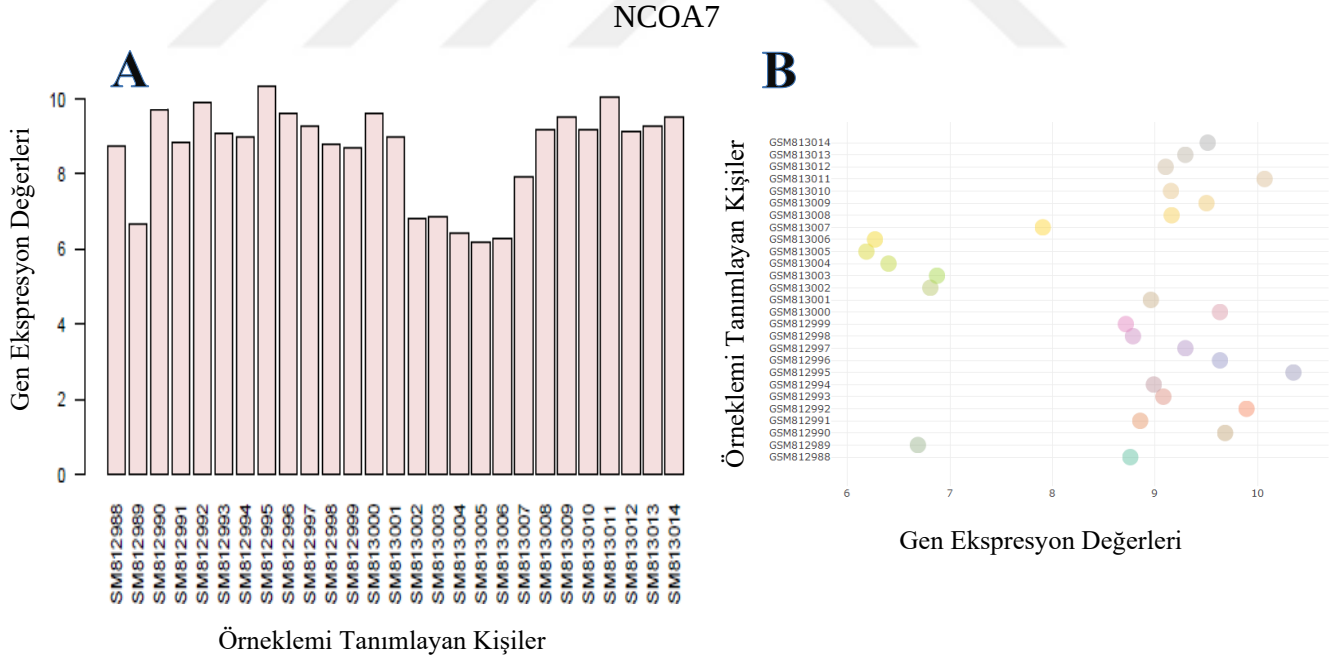
Şekil 4.28 : TNRC6A Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde TNRC6A geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



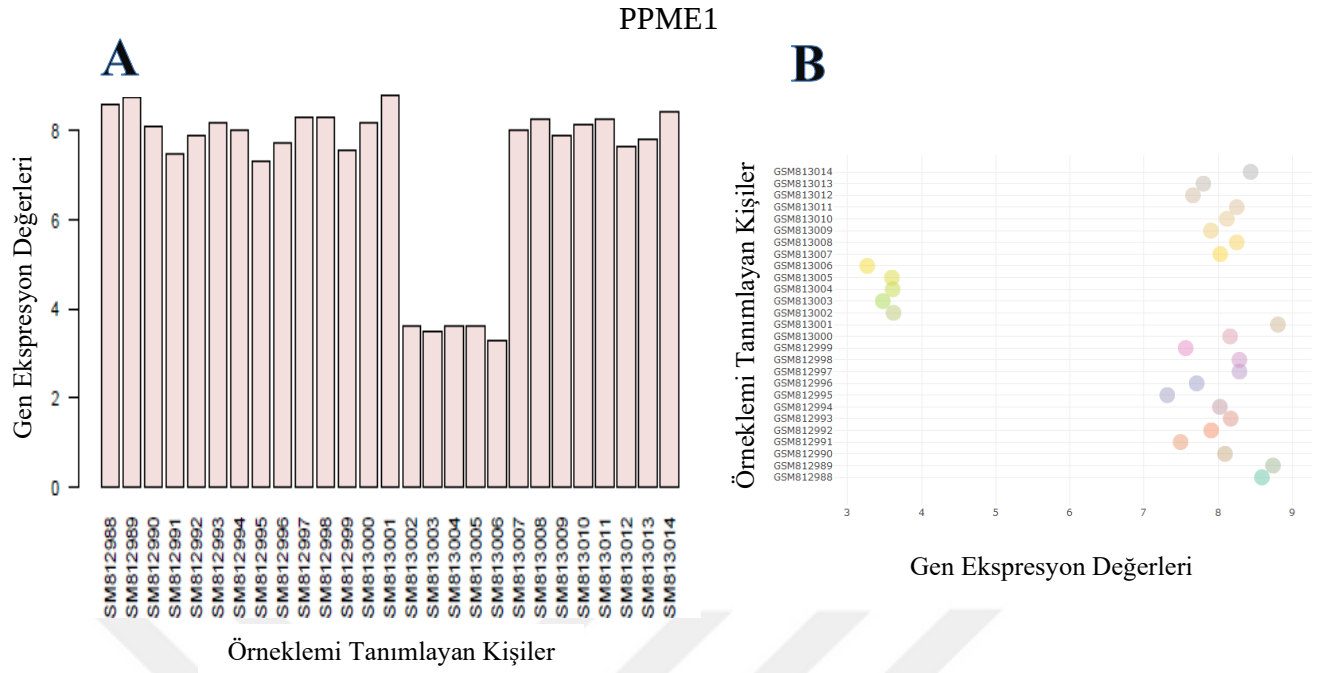
Şekil 4.29 : HEXIM1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde HEXIM1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



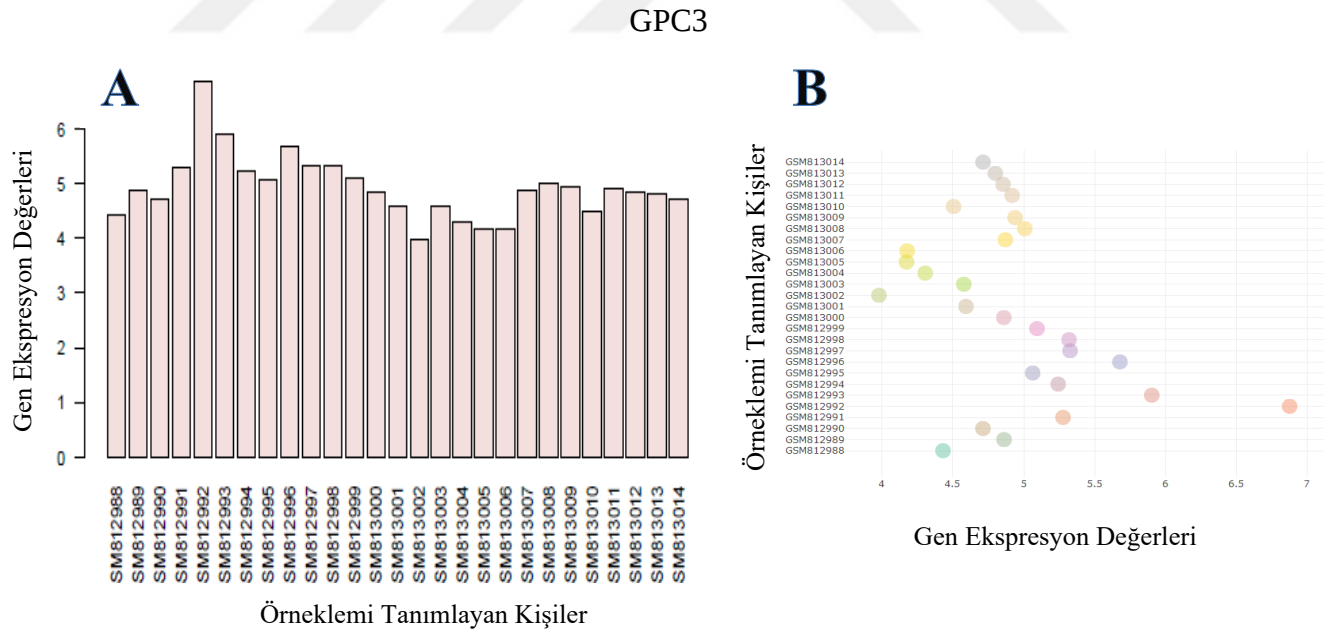
Şekil 4.30 : NCOA7 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde NCOA7 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



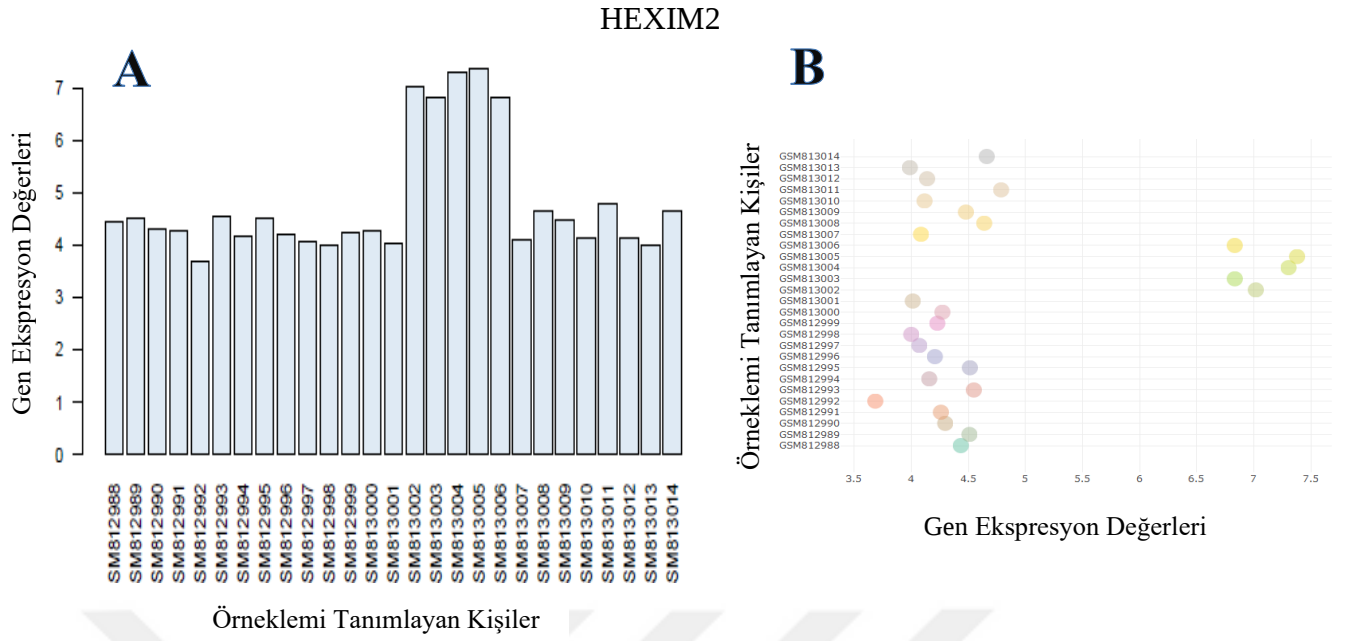
Şekil 4.31 : PPME1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde PPME1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



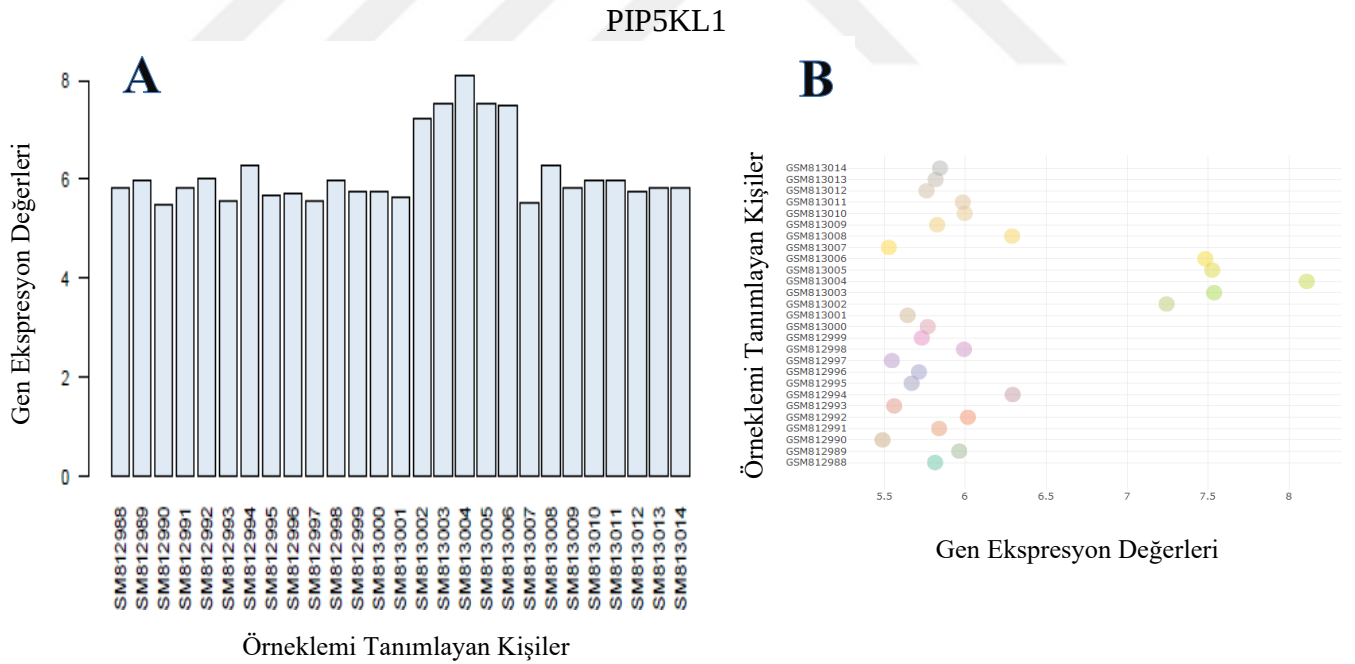
Şekil 4.32 : GPC3 Geninin Ekspresyon Değerlerinin Grafikleri

GPC3 geninin A grafiğinde yaşlandıkça aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise yaşa bağlı lineer olarak ekspresyon azalışı görülmektedir.



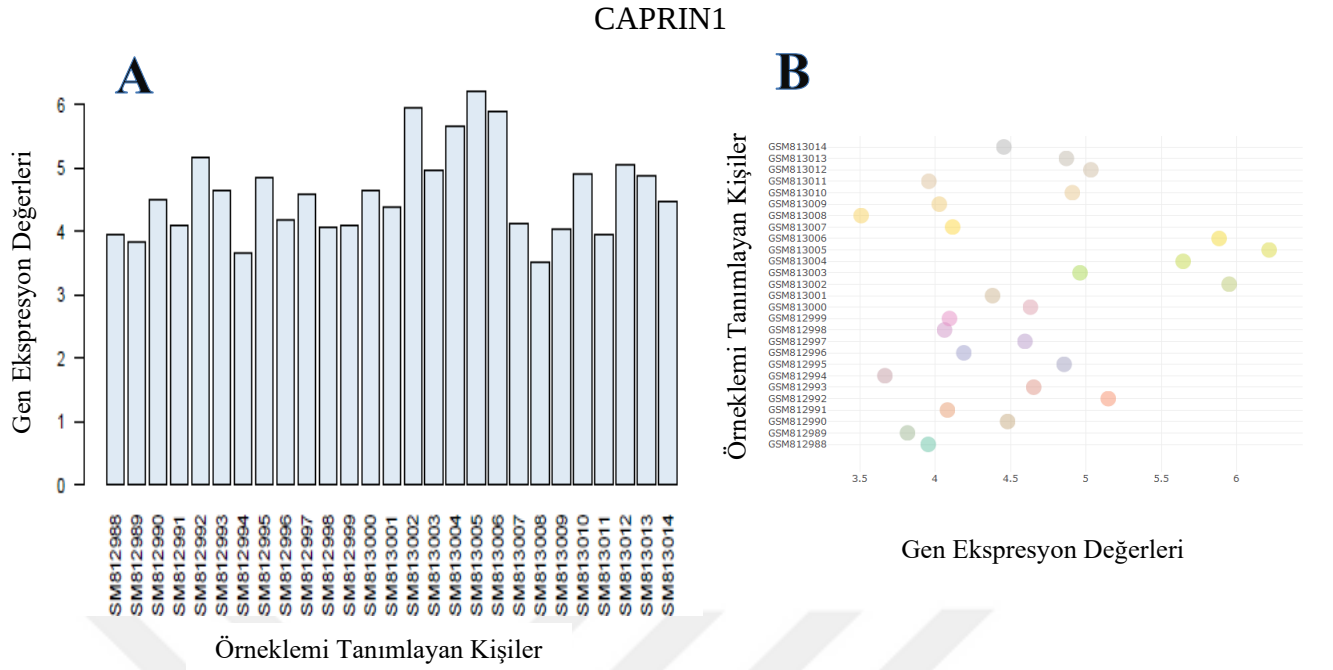
Şekil 4.33 : HEXIM2 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde HEXIM2 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



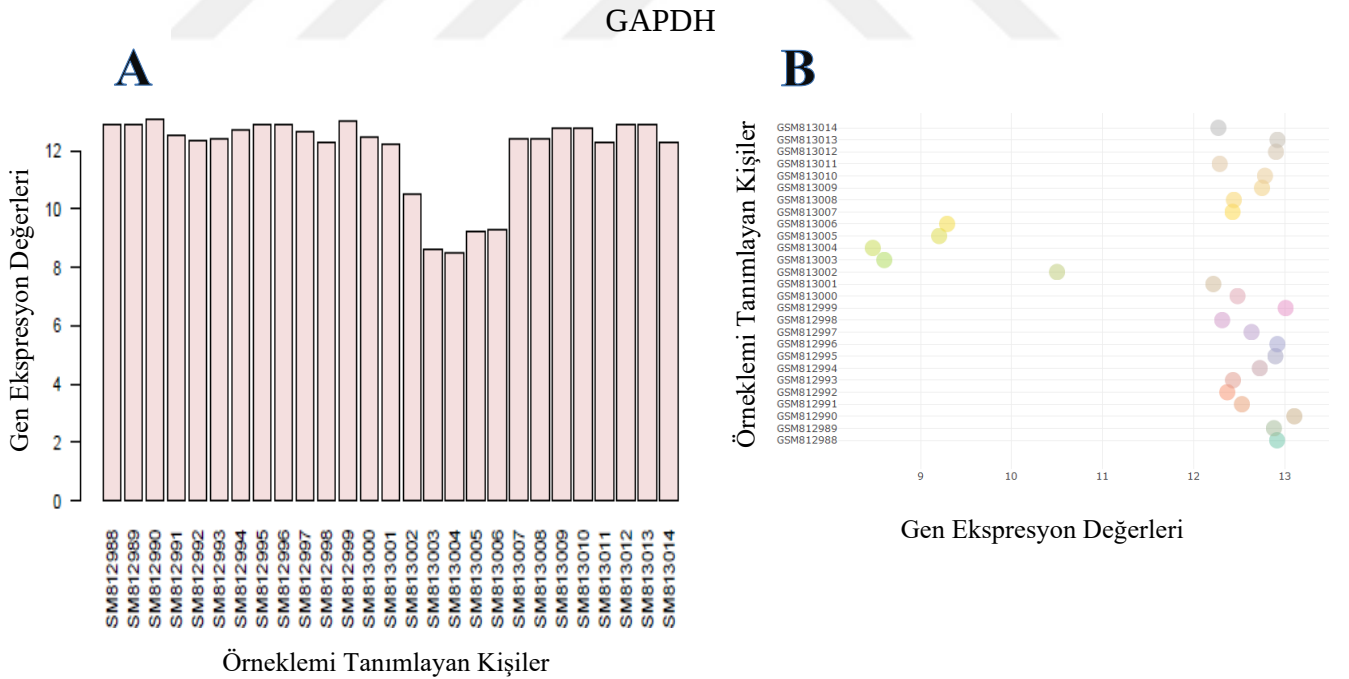
Şekil 4.34 : PIP5KL1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde PIP5KL1 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



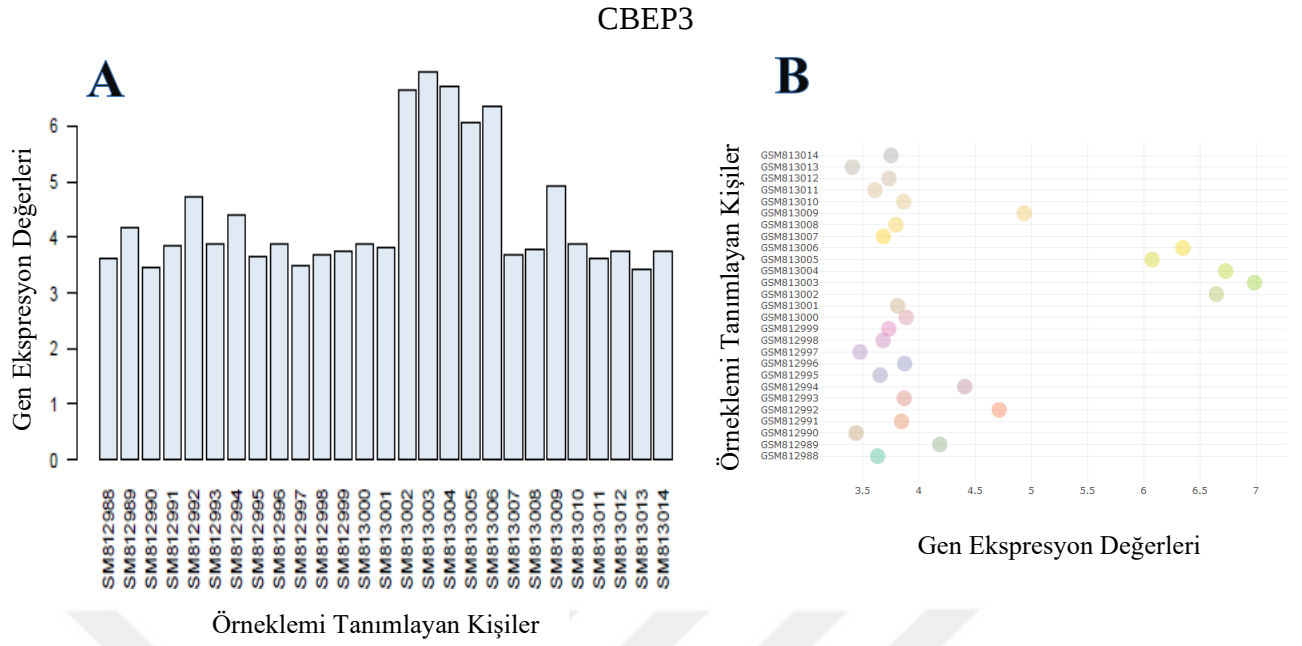
Şekil 4.35 : CAPRIN1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde CAPRIN1 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.



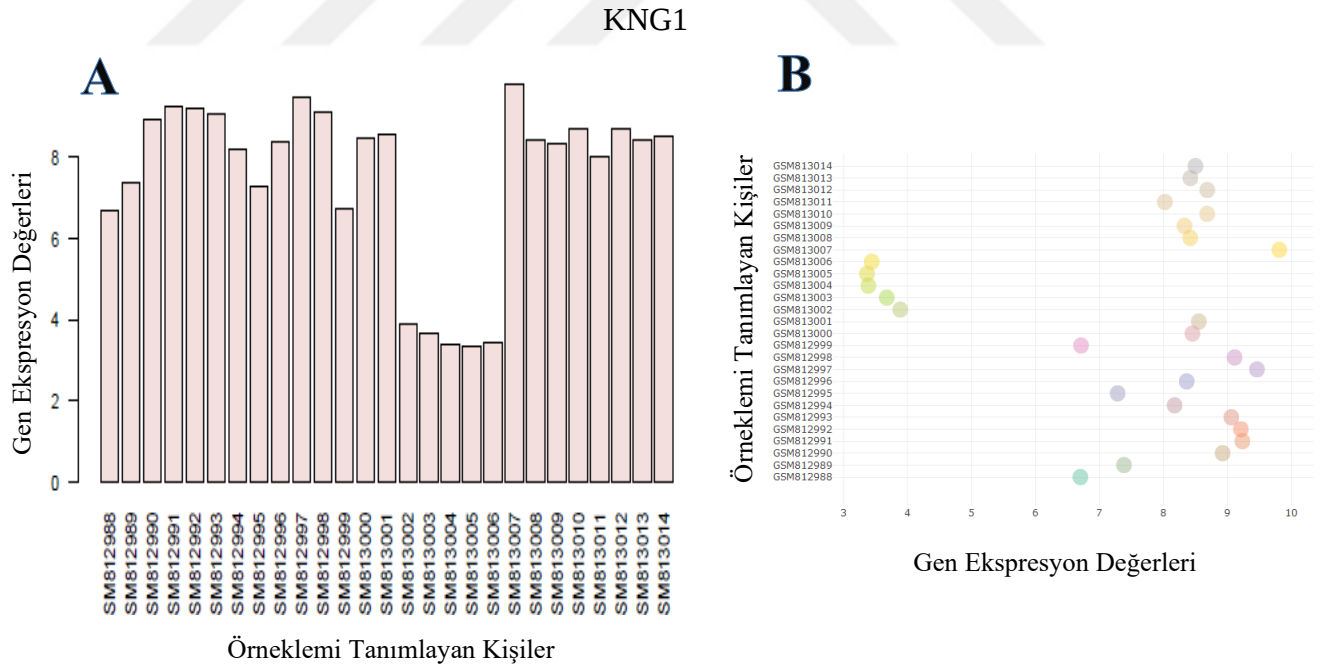
Şekil 4.36 : GAPDH Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde GAPDH geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.



Şekil 4.37 : CBEP3 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde CBEP3 geninin orta yaşlı gruptaki bireylerde yukarı regüle olduğu gözlemlenmiştir. B grafiğinde ise kümelenmiş orta yaşlı grubun yüksek ekspresyon profili görülmektedir.

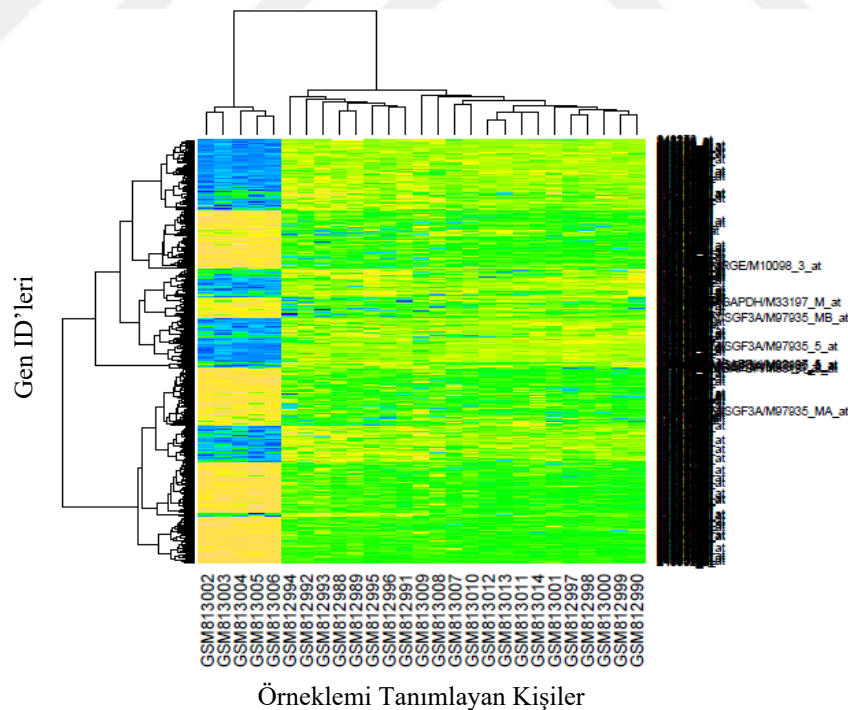


Şekil 4.38 : KNG1 Geninin Ekspresyon Değerlerinin Grafikleri

A grafiğinde KNG1 geninin orta yaşlı gruptaki bireylerde aşağı regüle olduğu gözlemlenmiştir. B grafiğinde kümelenmiş orta yaşlı grubun düşük ekspresyon profili görülmektedir.

Hematopoetik kök hücreler yaşam boyunca tüm kan hücrelerinin dengeli bir şekilde oluşumunu sağlamaktadır. Bu çalışmada, insan hematopoetik hücreleri üzerindeki yaşa bağlı etkileri kök hücre düzeyinde istatistiksel metotlarla analiz edilmiştir.

İmmün fonksiyon kaybı ve artan miyeloid lösemi insidansı, hematopoetik sistemin yaşlanmasının klinik olarak önemli sonuçlarındandır. HKH yaşlanmasında araştırmaların çoğu fareler üzerinde yapılan çalışmaları kapsamaktadır. Yapılan çalışmalar neticesinde kavramsal olarak farede ve insanda kan hücrelerinin oluşumunun aynı olduğunu ve HKH yaşlanmasına sebep olan mekanizmaların benzer olduğunu göstermektedir (Haan ve Lazare, 2019). Farelerde HKH yaşlanmasında artmış kök hücre, kendi kendini yenileme, miyeloid ve lenfoid progenitörlerini üretmek için farklı kapasite ve azalan lenfoid potansiyelini içeren hücre değişikliklerinin eşlik ettiği saptanmıştır (Rossi ve diğ. 2005). Bu çalışma kapsamında insan HKH analiz edildiğinde, yapılan hipotez testi ile örnekleme ait yaş grupları arasında istatistiksel olarak anlamlı ve yaş artışıyla birlikte orta yaşlı grupta miyeloid ekspresyon seviye sıklığı tespit edilmiştir.



Şekil 4.39 : Örneklemindeki Genlerin Isı Haritası

Şekil 43’de 570 gen ekspresyonları üzerinde yapılan ısı haritası sonucunda orta yaşlı gurubun farklılaştığı ve kısmen yukarı kısmen de aşağı regüle durumları

görülmektedir.

HKH yaşlanması hücre içi ve hücre dışı kontrolde derin bozulmalar ve değişimler olduğunu gösterir. HKH yaşlanmasının altında yatan bu bulgular miyeloid lösemogenezde ana protonkogenlerin yaşa bağlı olarak artması yaşlı HKH'lerin artan miyeloid soylarına eğilimleriyle birlikte yaşlandıkça miyeloid lösemi görülme sıklığının artması için bir gerekçe olduğu bilinmektedir (Rossi ve diğ. 2005).

Ekspresyon profili incelendiğinde HKH yaşlanmasına lenfoid spesifikasyonuna aracılık eden genlerin sistemik olarak aşağı regüle durumu miyeloid yanlılığın yukarı regülasyonunu ifade etmektedir. Bu durum özellikle genç ve orta yaşlı insan HKH'sinin ekspresyon analizinden elde edilen sonuçlar ile saptanmıştır. Yaşlı bireylerde ise regülasyon durumu kısmen gençlerle yakın seyretmektedir. Veriler doğrudan insan HKH'sini ve bunların sıklık ve gen ekspresyonu ifadeleri ile ilgili değişikliklerini, insan hematopoetik sistemindeki yaşlanmaya hayati katkı yapan etmenleri doğrudan etkilemektedir (Cho ve diğ., 2008).

Sağlıklı hematopoetik kök hücreler AML hücrelerine belirsiz bir şekilde dönüşebilmekle birlikte AML'de monoblast ve myeloblast gibi genç kök hücrelere dönüşür. Bu hücreler olgun bir hücre yapısında farklılaşmadığı için kanda ve kemik iliğinde birikmektedir. Sonrasında ise enfeksiyonlarla yeterli mücadeleyi veremeyen, görevlerini yapamayan lösemik hücrelere dönüşür. Bu lösemik hücreler kanda ve kemik iliğinde kontrolsüz olarak çok hızlı çoğaldıkça sağlıklı diğer organizmalar da yeteri kadar üretilemez ve görevlerini yapamayan işlevsiz hale gelir. Böylece lenfoidlerin azalmasıyla kolay kanamalar, olgun beyaz kan hücrelerin yenilenmemesi ile enfeksiyon ve diğer maligniteler ortaya çıkmaya başlamaktadır (Bryder ve diğ., 2006).

Örnekleme ait anlamlı gen sınıfları incelendiğinde miyeloid yanlılık gösteren ekspresyon seviyeleri orta yaşlı grupta belirgin bir şekilde artış göstermiş, lenfoid yanlılık gösteren ekspresyon seviyeleri de aynı grup için azalma gözlemlenmiştir. Tespit edilen miyeloid hücre ekspresyon seviyesindeki artış ve lenfoid hücre ekspresyon seviyesindeki bu azalış literatürdeki AML'ye yakalanma riskini istatistiksel olarak desteklemektedir.

5. SONUÇLAR VE ÖNERİLER

Çalışmada yaşa bağlı hematolojik fonksiyon bozukluklarına yatkın olabilecek insan hematopoetik sistemin özelliklerini istatistiksel olarak açıklamak için sağlıklı, hematolojik olarak genç, orta yaşlı ve yaşlı insan kemik iliği örneklerinden HKH ekspresyon profilleri değerlendirilmiştir. Kurulan alternatif hipotezde, yaş grupları arasında ekspresyon seviyeleri bazında anlamlı bir fark olduğu ortaya koyulmuştur. Bu bağlamda anlamlılık derecesi, istatistiksel hipotez testi ANOVA ile p değeri 0,043 değerinde hesaplanmıştır. Yapılan istatistiksel analiz sonucunda anlamlı genler incelendiğinde orta yaşlı bireylerin HKH ekspresyon değerlerinin yüksek frekansı karakterize edilmiştir.

Üç grupta sınıflandırılan örneklem üzerinden miyeloid ve lenfoid yanlılık gösteren genlerin ifadelerinin yaşa bağlı olarak değiştiğini, orta yaşlı bireylerin miyeloid yanlılık gösteren gen ifadelerinde artış gözlemlenirken lenfoid yanlılık gösteren genlerde negatif regüle olduğu ortaya koyulmuştur.

GORilla'da istatistiksel analiz sonucunda anlamlı olan hedef gen dizisi ile arka plan gen setini program dahilinde kullanarak veri setinde optimizasyon çalışması yapılmıştır. Böylece elde edilen gen dizisini azaltarak en anlamlı regüle olmuş gen ifadeleri elde edilmiştir. Grafikasyon işlemleri gerçekleştirilen genlerin miyeloid ve lenfoid yanlı olduğu, sınıflandırma işlemi ile de gruplandırılmıştır.

Veri setinin sınıflandırılması için denetimli makine öğrenmesi algoritması olan K-en yakın komşu algoritması kullanılmıştır. Veri setinin sınıflandırılacak bilgisi gizlenerek algoritmanın doğruluk, kesinlik ve hassasiyet gibi metriklerin performans oranları hesaplanmıştır. Şekil 16'da verilen Hata Matrisi algoritmaya ait performans ölçütlerini göstermektedir. Analiz sonucunda %95,6 oranında doğruluk değeri bulunarak gen ekspresyon seviyelerinin hangi sınıfa ait olduğu tespit edilmiştir.

Spesifik olarak incelendiğinde CSF2RB, MLF2, MLLT10, HOXA9, F2RL1, FUT1, USP46, CCDC88A sembolleri ile ifade edilen genler miyeloid yanlılık gösteren genleri ifade etmektedir. Bu genlerin yaşa bağlı olarak pozitif regülasyonu, VPREB1, LTBR, LY6H, LY6D, FLT3LG, LCP1, EMP1, SOX4 sembolleri ile ifade edilen

genler lenfoid yanlılık gösteren genlerin yaşı ile ilerlemesiyle negatif regüle oldukları belirlenmiştir.

Hematopoez kök hücrelerin biyolojik olarak açıklanması klinik alanda geniş bir yer tutmaktadır. Yaş ile birlikte artan miyeloid yanlılık seviyesi yaşlı bireylerde AML'ye yakalanma riskini de artırmaktadır. Bu çalışmada orta yaşlı grubu temsil eden örneklemdeki anlamlı gen ifadeleri bu sonucu ortaya koymaktadır. İnsan hematopoetik sistemin özelliklerini istatistiksel olarak açıklamak için genç, orta yaşlı ve yaşlı insan kemik iliği örneklerinden HKH ekspresyon profillerinin analizi incelenmiştir. İstatistiksel olarak anlamlı genlere ait veri seti makine öğrenme algoritması ile sınıflandırılmıştır. Sonuç itibarıyla verilenden elde edilen bulgular doğrultusunda örnekleme ait yaş grupları arasında istatistiksel olarak anlamlı bir farkın olduğu ve orta yaşlı grupta belirgin ekspresyon seviye farkı tespit edilmiştir.

Bu doğrultuda, ileri araştırmalara şu önerilerde bulunulabilir:

- Analizde kurulan hipotez testi için tek yönlü varyans analizi (ANOVA) kullanılmıştır. İstatistiksel test kapsamında farklı metot ve yaklaşımlar yapılabilir.
- Bu araştırmada referans alınan bağımlı değişken ekspresyon seviyeleridir. Yapılacak araştırmalarda ek olarak, farklı bağımlı değişkenlerin ilişkisine dair çalışmalar yapılabilir.
- Bu çalışmada makine öğrenmesi kapsamında denetimli öğrenme algoritması olan KNN algoritması kullanılmıştır. Farklı makine öğrenmesi metotları ve yöntemler geliştirilebilir.
- Araştırmanın örnekleme seçilirken veri grupları sayısal olarak artırılabilir ve değişkenler bu bağlamda incelenebilir.

KAYNAKLAR

- Adams, G. B., & Scadden, D. T. (2006). The hematopoietic stem cell in its place. *Nature immunology*, 7(4), 333-337.
- Ahmad, P., Qamar, S., & Rizvi, S. Q. A. (2015). Techniques of data mining in healthcare: a review. *International Journal of Computer Applications*, 120(15).
- Ali, Z., & Bhaskar, S. B. (2016). Basic statistical tools in research and data analysis. *Indian journal of anaesthesia*, 60(9), 662.
- Altay, G. (2021). *Akut myeloid lösemi hastalarında FLT3 ve NPM1 gen mutasyonlarının sıklığının belirlenmesi ve klinik öneminin araştırılması* (Master's thesis, Kocaeli Üniversitesi, Sağlık Bilimleri Enstitüsü).
- Altman, D. G., & Bland, J. M. (1999). Statistics notes Variables and parameters. *Bmj*, 318(7199), 1667.
- Altun, A. (2018). *Thymoquinone ve cytosine arabinoside'in akut miyeloid lösemi hücre hattında apoptozis ve hücre çoğalmasına etkileri* (Master's thesis, Ondokuz Mayıs Üniversitesi, Sağlık Bilimleri Enstitüsü).
- Altunkaynak, A., BAŞAKIN, E. E., & KARTAL, E. (2020). Dalgacık K-En Yakın Komşuluk Yöntemi İle Hava Kirliliği Tahmini. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 25(3), 1547-1556.
- Atıcı, E. (2007). Tıp tarihinde kanser ve lösemi. *Türk onkoloji dergisi*, 22(4), 197-204.
- Attwood T.K., Parry-Smith D.J., Introduction to Bioinformatics, Prentice Hall, Harlow, 1999;240.
- Ayhan, S., & Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201.
- Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, 3, 19-48.
- Bain, B. J., Clark, D. M., & Wilkins, B. S. (2019). *Bone marrow pathology*. John Wiley & Sons.
- Barr, J. R. (2018, September). Machine learning, A tutorial with R. In *2018 First International Conference on Artificial Intelligence for Industries (AI4I)* (pp. 120-121). IEEE.
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., ... & Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research*, 41(D1), D991-D995.
- Bartholomew, D. J. (1995). What is statistics?. *Journal of the Royal Statistical*

Society Series A: Statistics in Society, 158(1), 1-20.

Bryder, D., Rossi, D. J., & Weissman, I. L. (2006). Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *The American journal of pathology*, 169(2), 338-346.

Cabral Júnior, J. B., & Lucena, R. L. (2020). Analysis of precipitation using Mann-Kendall and Kruskal-Wallis non-parametric tests. *Mercator (Fortaleza)*, 19.

Chambers, S. M., Shaw, C. A., Gatz, C., Fisk, C. J., Donehower, L. A., & Goodell, M. A. (2007). Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. *PLoS biology*, 5(8), e201.

Chauhan, H., & Chauhan, A. (2013). Implementation of decision tree algorithm c4. 5. *International Journal of Scientific and Research Publications*, 3(10), 1-3.

Cho, R. H., Sieburg, H. B., & Muller-Sieburg, C. E. (2008). A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood, The Journal of the American Society of Hematology*, 111(12), 5553-5561.

Connelly, L. M. (2021). Introduction to analysis of variance (ANOVA). *Medsurg Nursing*, 30(3), 218-158.

Cressie, N. A. C., & Whitford, H. J. (1986). How to use the two sample t-test. *Biometrical Journal*, 28(2), 131-148.

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. In *Machine learning techniques for multimedia: case studies on organization and retrieval* (pp. 21-49). Berlin, Heidelberg: Springer Berlin Heidelberg.

de Haan, G., & Lazare, S. S. (2018). Aging of hematopoietic stem cells. *Blood, The Journal of the American Society of Hematology*, 131(5), 479-487.

Donalek, C. (2011, April). Supervised and unsupervised learning. In *Astronomy Colloquia. USA* (Vol. 27, p. 8).

Doulatov, S., Notta, F., Laurenti, E., & Dick, J. E. (2012). Hematopoiesis: a human perspective. *Cell stem cell*, 10(2), 120-136.

Dridi, S. (2021). Supervised learning-a systematic literature review.

Eberly, L. E. (2007). Correlation and simple linear regression. *Topics in biostatistics*, 143-164.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics*, 10(1), 1-7

Erçakallı, T. S., & Polat, S. (2022). Kanser Kök Hücrelerine Güncel Yaklaşım. *Arşiv Kaynak Tarama Dergisi*, 31(2), 111-121.

Ferreira, J. A., & Zwinderman, A. H. (2006). On the benjamini–hochberg method.

Gluckman, E. (2000). Current status of umbilical cord blood hematopoietic stem cell transplantation. *Experimental hematology*, 28(11), 1197-1205.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

Hann G, Lazare S. (2019). Aging of Hematopoietic Stem Cells.

- Harwell, M. R. (1988). Choosing between parametric and nonparametric tests. *Journal of Counseling & Development*, 67(1), 35-38.
- Hassanat, A. B., Abbadi, M. A., Altarawneh, G. A., & Alhasanat, A. A. (2014). Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. *arXiv preprint arXiv:1409.0919*.
- Hazra, A., & Gogtay, N. (2016). Biostatistics series module 3: comparing groups: numerical variables. *Indian journal of dermatology*, 61(3), 251.
- Homack, S. R. (2001). Understanding What ANOVA Post Hoc Tests Are, Really.
- Huang C, Freter C. (2015). Lipid metabolism, apoptosis and cancer therapy. *International journal of molecular sciences*. 16(1), 924–949.
- İnan, S., Özbilgin, K. (2009) Kök Hücre Biyolojisi. Sağlıkta Birikim., 1,5:11-23s.
- Iwasaki, H., & Akashi, K. (2007). Myeloid lineage commitment from the hematopoietic stem cell. *Immunity*, 26(6), 726-740.
- Jain, V. K. (2018). *Data Science and Analytics (with Python, R and SPSS Programming)*. Khanna Publishing House.
- Jo, T. (2021). Machine learning foundations. *Supervised, Unsupervised, and Advanced Learning*. Cham: Springer International Publishing.
- Johnson, R. W. (2022). Alternate Forms of the One-Way ANOVA F and Kruskal–Wallis Test Statistics. *Journal of Statistics and Data Science Education*, 30(1), 82-85.
- Kampen, K. R., Ter Elst, A., & de Bont, E. S. (2013). Vascular endothelial growth factor signaling in acute myeloid leukemia. *Cellular and molecular life sciences*, 70, 1307-1317.
- Karadağ, A , Altınok, B , Özkan, T , Hekmatshoar, Y . (2007)"Kemik İliği Stroması: Hücreleri ve Mikroçevresi". *Ankara Sağlık Hizmetleri Dergisi* 16 : 51-62
- Karadimitriou, S. M., Marshall, E., & Knox, C. (2018). Mann-whitney u test. *Sheffield: Sheffield Hallam University*, 4.
- Karaöz, E., Ovalı, E. (2004) Kök Hücreler, Derya kitabevi, Trabzon,187s.
- Keerthi, S. & Gilbert, E., Convergence of a Generalized SMO Algorithm for SVM Classifier Design, Kluwer Academic Publishers. Machine Learning vol. 46, (2002). pp 351–360.
- Kılıç, S. (2014). Etki Büyüklüğü. *Journal of Mood Disorders*, 4(1):44-46.
- Kim, M., MOON, H. B., & Spangrude, G. J. (2003). Major age-related changes of mouse hematopoietic stem/progenitor cells. *Annals of the New York Academy of Sciences*, 996(1), 195-208.
- Kim, T. K. (2015). T test as a parametric statistic. *Korean journal of anesthesiology*, 68(6), 540-546.
- Kim, T. K. (2017). Understanding one-way ANOVA using conceptual figures. *Korean journal of anesthesiology*, 70(1), 22-26.
- Klein, J. P., Moeschberger, M. L., Klein, J. P., & Moeschberger, M. L. (2003). Hypothesis testing. *Survival analysis: techniques for censored and*

truncated data, 201-242.

- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- Kul, S. (2014). Uygun istatistiksel test seçim kılavuzu/guideline for suitable statistical test selection. *Plevra Bülteni*, 8(2), 26.
- Lesmeister, C. (2015). *Mastering machine learning with R*. Packt Publishing Ltd.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9(1), 381-386.
- Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: theory and applications* (Vol. 81). World scientific.
- Mallat, J. (2017). Understanding the null hypothesis (H0) in non-inferiority trials. *Critical Care*, 21(1), 101.
- Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(4), 140-147.
- McCulloch, E. A., & Till, J. E. (1960). The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. *Radiation research*, 13(1), 115-125.
- Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL (1996) The aging of hematopoietic stem cells. *Nat Med* 2:1011–1016.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, 4, 51-62.
- Ostell, J. M., Wheelan, S. J., & Kans, J. A. (2001). The NCBI data model. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 19, 44.
- Pang, W. W., Price, E. A., Sahoo, D., Beerman, I., Maloney, W. J., Rossi, D. J., ... & Weissman, I. L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proceedings of the National Academy of Sciences*, 108(50), 20012-20017.
- Pearce, D. J., Anjos-Afonso, F., Ridler, C. M., Eddaoudi, A., & Bonnet, D. (2007). Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging. *Stem cells*, 25(4), 828-835.
- Peng, R. D. (2016). *R programming for data science* (pp. 86-181). Victoria, BC, Canada: Leanpub.
- Richardson, J. T. (2018). Kruskal–Wallis Test.
- Rong, S., & Bao-Wen, Z. (2018). The research of regression model in machine learning field. In *MATEC Web of Conferences* (Vol. 176, p. 01033). EDP Sciences.
- Ross, A., & Willson, V. L. (2017). One-sample T-test. In *Basic and advanced statistical tests* (pp. 9-12). Brill.

- Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., & Weissman, I. L. (2005). Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proceedings of the National Academy of Sciences*, 102(26), 9194-9199.
- Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4), 688-690.
- Sedgwick, P. (2014). One way analysis of variance: post hoc testing. *Bmj*, 349.
- Sharma, D., & Kumar, N. (2017). A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 6(10), 2278-1323.
- Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.
- Spangrude, G. J., Heimfeld, S., & Weissman, I. L. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science*, 241(4861), 58-62.
- Speed, T. P. (1987). What is an analysis of variance?. *The Annals of Statistics*, 885-910.
- Stolzenberg, R. M. (2004). Multiple regression analysis. *Handbook of data analysis*, 165-208.
- Sudo, K., Ema, H., Morita, Y., & Nakauchi, H. (2000). Age-associated characteristics of murine hematopoietic stem cells. *The Journal of experimental medicine*, 192(9), 1273-1280.
- Sümbüloğlu, K. & Sümbüloğlu, V. (2007). Biyoistatistik. 12. Baskı. Hatiboğlu Yayınevi.
- Supek, F., & Škunca, N. (2017). Visualizing go annotations. *The Gene Ontology Handbook*, 207-220.
- Taylor, P. R., Reid, M. M., Bown, N., Hamilton, P. J., & Proctor, S. J. (1992). Acute lymphoblastic leukemia in patients aged 60 years and over: a population-based study of incidence and outcome.
- Terese, W., Kibiuk, L., & Duckwall, C. (2001). The adult stem cell. *Stem Cell Information*, <http://stemcells.nih.gov/info/scireport/pages/chapter4.aspx>. Accessed, 29(10), 2013.
- Thabtah, F., Abdelhamid, N., & Peebles, D. (2019). A machine learning autism classification based on logistic regression analysis. *Health information science and systems*, 7, 1-11.
- Thomas, D., & Majeti, R. (2017). Biology and relevance of human acute myeloid leukemia stem cells. *Blood, The Journal of the American Society of Hematology*, 129(12), 1577-1585.
- Tippmann, S. (2015). Programming tools: Adventures with R. *Nature*, 517(7532), 109-110.
- Tranmer, M., & Elliot, M. (2008). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, 5(5), 1-5.

- Tunalı Özen, E. (2019). *Kronik lenfostik lösemide NOTCH1 ve XPO1 genlerindeki varyasyonların araştırılması* (Master's thesis, Trakya Üniversitesi, Sağlık Bilimleri Enstitüsü).
- Ugoni, A., & Walker, B. F. (1995). The Chi square test: an introduction. *COMSIG review*, 4(3), 61.
- Usul, F. (2012). *Göbek kordonunda embriyonik kök hücre varlığının belirlenmesi ve tiplendirilmesi* (Master's thesis, Pamukkale Üniversitesi Sağlık Bilimleri Enstitüsü).
- Vijayvargiya, A. (2009). One-way analysis of variance. *Journal of Validation Technology*, 15(1), 62.
- Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *Maics*, 710(1), 120-127.
- Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6), 599-606.
- Weiss, A. R. R., & Dahlke, M. H. (2019). Immunomodulation by mesenchymal stem cells (MSCs): mechanisms of action of living, apoptotic, and dead MSCs. *Frontiers in immunology*, 10, 1191.
- Weissgerber, T. L., Garovic, V. D., Milin-Lazovic, J. S., Winham, S. J., Obradovic, Z., Trzeciakowski, J. P., & Milic, N. M. (2016). Reinventing biostatistics education for basic scientists. *PLoS Biology*, 14(4), e1002430.
- Wilkerson, S. (2008). Application of the Paired t-test. *XULAnEXUS*, 5(1), 7.
- Yahia, M. E., & Ibrahim, B. A. (2003). K-nearest neighbor and C4. 5 algorithms as data mining methods: advantages and difficulties. *Computer Systems and Applications*, 103, 103-109
- Yiallourous, M., Creutzig, U., & Saribeyoglu, E. (2022). Akut Miyeloid Lösemi (AML)-Kısa Bilgiler.
- Yong, Z., Youwen, L., & Shixiong, X. (2009). An improved KNN text classification algorithm based on clustering. *Journal of computers*, 4(3), 230-237.
- Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(3), 1-19.
- Zhang, Y. (2012). Support vector machine classification algorithm and its application. In *Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings, Part II* 3 (pp. 179-186). Springer Berlin Heidelberg.
- Zhou, L., Luo, T., Du, M., Chen, Q., Liu, Y., Zhu, Y., ... & Yang, K. (2021). Machine learning comparison and parameter setting methods for the detection of dump sites for construction and demolition waste using the google earth engine. *Remote Sensing*, 13(4), 787.
- Zibran, M. F. (2007). Chi-squared test of independence. *Department of Computer Science, University of Calgary, Alberta, Canada*, 1(1), 1-7.
- Ziliak, S. (2017). P values and the search for significance. *Nature methods*, 14(1), 3-4.

ÖZGEÇMİŞ

Soyadı, Adı : ŞİMŞEK, Muhammet Burak
Meslek : Veri Mühendisi

Eğitim Bilgileri

- İstanbul Gedik Üniversitesi - Yapay Zekâ Mühendisliği Tezli Yüksek Lisans Programı (2021-2023)
- Düzce Üniversitesi - Biyomedikal Mühendisliği (2015-2020)

Mesleki Deneyim

- Ereteam - Alternatif Bank - Veri Tabanı Analisti (2021-2023)
- Ereteam - Akbank - Veri Tabanı Analisti (2023-)