

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**BİYOİNFORMATİK TEKNİKLERİ KULLANARAK
YENİ MİKRO RNA'LARIN BULUNMASI VE
VARYANT ANALİZLERİNİN YAPILMASI:
CITRUS MODELİ**

CANKUT ÇUBUK

**TIBBİ BİLİŞİM
YÜKSEK LİSANS TEZİ**

İZMİR-2019

TEZ KODU: DEU.HSI.MSc-2010970024

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**BİYOİNFORMATİK TEKNİKLERİ KULLANARAK
YENİ MİKRO RNA'LARIN BULUNMASI VE
VARYANT ANALİZLERİNİN YAPILMASI:
CITRUS MODELİ**

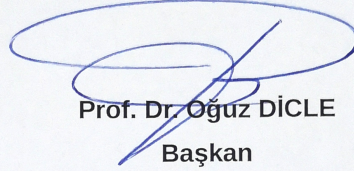
CANKUT ÇUBUK

**TIBBİ BİLİŞİM
YÜKSEK LİSANS TEZİ**

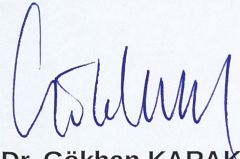
DANIŞMAN: PROF. DR. OĞUZ DİCLE

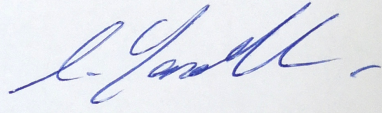
TEZ KODU: DEU.HSI.MSc-2010970024

Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Medikal İnfomatik Anabilim Dalı, Tıbbi Bilişim Yüksek Lisans programı öğrencisi Cankut ÇUBUK “Biyoinformatik teknikleri kullanarak yeni mikro RNA’ların bulunması ve varyant analizlerinin yapılması: *Citrus* modeli” konulu yüksek lisans tezini 15/04/2019 tarihinde başarılı olarak tamamlamıştır.


Prof. Dr. Oğuz DİCLE
Başkan

Dokuz Eylül Üniversitesi Tıp Fakültesi
Radyoloji Anabilim Dalı


Doç. Dr. Gökhan KARAKÜLAH
Üye
İBG Genom Bilimleri ve
Moleküler Biyoteknoloji A.D.


Dr. Öğr. Üyesi Cihangir YANDIM
Üye
İzmir Ekonomi Üniversitesi
Genetik ve Biyomühendislik Bölümü

Prof. Dr. Cüneyt GÜZELİŞ
Yedek Üye
Yaşar Üniversitesi
Elektrik Elektronik Müh. Bölümü

Doç. Dr. Gökhan DALKILIÇ
Yedek Üye
DEÜ Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü

İÇİNDEKİLER

İÇİNDEKİLER.....	I
TABLolar DİZİNİ.....	IV
ŞEKİLLER DİZİNİ.....	V
KISALTMALAR.....	VI
TEŞEKKÜR.....	VI
ÖZET.....	1
ABSTRACT.....	3
1. GİRİŞ VE AMAÇ.....	5
2. GENEL BİLGİLER.....	7
2.1. RNA çeşitleri ve miRNA.....	7
2.2. miRNA biyogenezi.....	8
2.3. miRNA'nın hedef genleri üzerindeki etki mekanizması.....	10
2.4. miRNA'ların biyoteknolojide kullanımı.....	11
2.5. miRNA çalışmalarında kullanılan moleküler teknikler.....	12
2.6. Yeni nesil dizileme.....	13
2.7. Küçük RNA dizileme analizinde kullanılan biyoinformatik araçlar.....	18
2.7.1. Yazılımlar.....	18
2.7.2. Veri tabanları.....	20
2.7.3. Veri formatları.....	21
3. GEREÇ ve YÖNTEM.....	23
3.1. Araştırmanın tipi.....	23
3.2. Araştırmanın yeri ve zamanı.....	23
3.3. Araştırmanın evreni ve örnekleme.....	23
3.4. Çalışma materyali.....	23
3.5. Araştırmanın değişkenleri.....	24
3.6. Veri toplama araçları.....	24
3.6.1. Analiz için gerekli verilerin elde edilmesi.....	24
3.6.1.1. Citrus küçük RNA dizileme verisi.....	24
3.6.1.2. Citrus referans genom dizisi.....	25
3.6.1.3. Citrus transkript dizileri.....	25
3.6.1.4. Filtreleme amacıyla kullanılan diziler.....	26

3.6.1.5. Otuz dört farklı Citrus türünün tüm genom dizilemesi ve tek nükleotit varyantları.....	26
3.6.2. Yeni nesil dizileme verilerini kullanarak, biyoinformatik araçlar yardımı ile yeni miRNA'ların doğru olarak belirlenmesi.....	29
3.6.2.1. Küçük RNA dizilerinin ön işlenmesi.....	29
3.6.2.2. Küçük RNA dizilerinin filtrelenmesi.....	30
3.6.2.3. miRNA belirleme.....	30
3.6.2.4. Yanlış pozitif miRNA'ların çalışmadan çıkarılması.....	33
3.6.3. Belirlenen miRNA'ların sınıflandırılması.....	33
3.6.3.1. Varyant analizi: Otuz dört farklı Citrus türü için ortak olan miRNA'ların belirlenmesi.....	33
3.6.3.2. Varyant analizi: elde edilen miRNA'ların farklı türler arasındaki korunmuşluğunun incelenmesi.....	34
3.6.3.3. Literatüre kazandırılan miRNA'ların belirlenmesi.....	34
3.6.3.4. İsomir dizilerinin kontrolü.....	34
3.6.4. Korunmuş miRNA'ların hedef genleri üzerindeki etkilerinin ve metabolik yollardaki rollerinin belirlenmesi.....	35
3.7. Araştırma planı.....	38
3.8. Verilerin değerlendirilmesi.....	38
3.9. Araştırmanın sınırlılıkları.....	39
3.10. Etik kurul onayı.....	40
4. BULGULAR.....	41
4.1. Ön işlenmesi ve filtrelemesi yapılmış küçük RNA dizileri.....	41
4.2. Belirlenen miRNA'lar.....	41
4.3. Belirlenen miRNA'ların sınıflandırılması.....	42
4.3.1. miRNA'ların sınıflar arası dağılımı.....	42
4.3.2. İsomir dizilerinin kontrolü.....	44
4.4. miRNA'ların etkiledikleri hedef genleri ve metabolik yollar.....	45
5. TARTIŞMA.....	48
6. SONUÇ VE ÖNERİLER.....	53
7. KAYNAKLAR.....	55
8. EKLER.....	61

EK1. Çalışmada belirlenen 1504 tane miRNA'nın tam listesi.....	61
EK2. Hedef genler ve genlerin ait oldukları protein aileleri.....	61
EK3. Etik kurul onayı.....	62
EK4. ARBİS araştırmacı özgeçmişi.....	64

TABLÖLAR DİZİNİ

Tablo 1. Hayvan ve bitki için miRNA biyogenez basamakları.....	9
Tablo 2. Hastalıklarla ilişkisi bulunan bazı miRNA'lar.....	12
Tablo 3. miRNA çalışmalarında kullanılan moleküler teknikler.....	13
Tablo 4a. Yeni nesil dizileme platformlarının kullanım alanlarına göre verimliliği (Birinci kısım).....	15
Tablo 4a. Yeni nesil dizileme platformlarının kullanım alanlarına göre verimliliği (İkinci kısım).....	16
Tablo 4b. Yeni nesil dizileme platformlarının teknik özellikleri.....	17
Tablo 5. Çalışmada kullanılan küçük RNA dizileme verileri.....	25
Tablo 6. Filtreleme öncesi ve sonrası küçük RNA dizi okumalarının miktarları.....	42
Tablo 7. Literatüre kazandırılan yeni miRNA'lar.....	43
Tablo 8. Korunmuş miRNA'ların hedef transkriptlerinin ve genlerinin sayısı.....	46
Tablo 9. Korunmuş miRNA'lar tarafından kontrol edilen metabolik yollar.....	46

ŞEKİLLER DİZİNİ

Şekil 1. Hayvan ve bitki miRNA'sının biyogenez basamakları.....	10
Şekil 2. miRNA ile hedef gen arasındaki dizi eşleşmesi.....	11
Şekil 3. Çalışmada kullanılan 34 farklı Citrus türünün listesi ve filogenetik ilişkileri	28
Şekil 4. Düz metin formatının <i>fasta</i> formatına çevrilmesi.....	29
Şekil 5. miRNA belirlenmesinde kullanılan parametreler.....	32
Şekil 6. miRBase veri tabanından (v20) elde edilen Viridiplantae (yeşil bitkiler) miRNA dizilerinin uzunluk dağılımı	32
Şekil 7. Yapılan analizin algoritmik akış şeması.....	37
Şekil 8. Çalışmada kullanılan analiz akış şeması.....	39
Şekil 9. Belirlenen miRNA'ların sınıflar arası dağılımı	43
Şekil 10. GSM1086751 veri setinde bulunan parçalanmış mRNA dizileri.....	45

KISALTMALAR

3p: Üç üssü

5p: Beş üssü

AOÇ: Açık okuma çerçevesi

bp: 1 nükleobaz çifti (base pair)

BWA: Burrows-Wheeler aligner

cDNA: Tamamlayıcı DNA

(Complementary DNA)

DKM: Destek vektör makineleri

DNA: Deoksiribonükleik asit

GATK: The genome analysis toolkit

GEO: Gene expression omnibus

GM: Genetik materyali

HEN1: RNA-metiltransferaz Hua-Enhancer1

hsa: *Homo sapiens*

IGV: Integrative genomics viewer

isomir: miRNA isoform

kp: 1000 nükleobaz çifti (kilobase pair)

Mb / Gb / Tb: Megabayt / Gigabayt / Terabayt

miRNA: Mikro RNA (Micro RNA)

mRNA: Mesajcı RNA

NGS: Next generation sequencing

piRNA: Piwi-etkileşimli RNA (Piwi-interacting RNA)

pre-miRNA: Öncül miRNA (Precursor miRNA)

pri-mirna: Birincil miRNA (Primary miRNA)

PZR: Polimeraz zincir reaksiyonu

RISC: RNA-indüklenmiş susturma kompleksi (RNA-induced silencing complex)

RNA: Ribonükleik asit

sRNA: Small RNA

rRNA: Ribozomal RNA

siRNA: Küçük susturucu (Small interfering)

snoRNA: Küçük nükleolar RNA (Small nucleolar RNA)

snRNA: Küçük nükleer RNA (Small nuclear RNA)

SNV: Single nucleotide variant

TNV: Tek nükleotit varyantı

tRNA: Taşıyıcı RNA

YND: Yeni nesil dizileme

TEŞEKKÜR

Bir bilim insanından çok daha fazlası olduğunu düşündüğüm ve örnek aldığım tez danışmanım Prof. Dr. Oğuz Dicle'ye yüksek lisans eğitimim boyunca bana sağladığı destek ve bu süreçte bana güven duyarak elinden gelen tüm olanakları sunmuş olduğu için çok teşekkür ederim.

Tez çalışmamın pratik çalışmalarının gerçekleştirilmesine olanak sağlayan ve laboratuvarında bana yer veren Dr. Joaquin Dopazo'ya, araştırma grubunda çalışan tüm arkadaşlarıma ve Centro de Investigación Príncipe Felipe'de beni her daim evimde hissettiren tüm çalışanlarına, maddi desteği için Avrupa Birliği Erasmus Programına çok teşekkür ederim.

Medikal İformatik Anabilim dalı öğrencileri ve öğretim üyelerine, tez projemin analiz aşamalarında benden bilgi ve desteğini hiç esirgemeyen Dr. José Carbonell Caballero ve Dr. Fransico Garcia'ya, programlama alanındaki destekleri için Alırıza Arıbaş ve Emrah Türkmen'e, özel sektörde çalıştığım dönemde yüksek lisans eğitimime olanak sağlayan ve destekleyen Emre Tepedelen ve Selen Şahin'e, yurt dışında olduğum süreçte benden yardımlarını esirgemeyen arkadaşlarıma, özellikle Hakan Cengiz'e ve SBE öğrenci işlerine çok teşekkür ederim.

Eğitim ve öğretim konusunda bana maddi ve manevi her zaman destek olan ve beni motive eden aileme, uzun süren yüksek lisans serüveninde bana her zaman güvenen, inanan ve hayatımı anlamlandıran Hérkèse, beni her gün sevgiyle karşılayan arkadaşım beagle cinsi Mojito'ya çok teşekkür ederim.

Cankut ÇUBUK

İzmir, 2019

BİYOİNFORMATİK TEKNİKLERİ KULLANARAK YENİ MİKRO RNA'LARIN BULUNMASI VE VARYANT ANALİZLERİNİN YAPILMASI: *CITRUS* MODELİ

Cankut ÇUBUK

Medikal İnfomatik Anabilim Dalı, Sağlık Bilimleri Enstitüsü, Dokuz Eylül Üniversitesi
İnciraltı, 35340, İzmir / TÜRKİYE
cankutcubuk [at] gmail [dot] com

ÖZET

Amaç: Bu çalışmanın hedefi mevcut YND verilerini kullanarak, biyoinformatik araçlar yardımı ile yeni miRNA'ların doğru olarak tanımlanması, tanımlanan miRNA'ların farklı türler arasındaki korunmuşluğunun incelenmesi, korunmuş miRNA'ların hedef genleri üzerindeki etkilerinin ve metabolik yollardaki rollerinin belirlenmesidir.

Yöntem: Gerek duyulan tüm genom ve küçük RNA dizileme verileri mevcut olduğu için, *Citrus* (Narenciye) çalışmamızda model organizma olarak kullanıldı. Ön işlem ve filtreleme basamaklarından geçen veriler, The UEA sRNA Workbench yazılımı kullanılarak miRNA değerlendirmesine tabi tutuldu. tRNAscan-SE ve NOVOMIR yazılımları ile yanlış pozitif olarak değerlendirilen miRNA'lar çalışmadan çıkarıldı. 34 farklı *Citrus* türüne ait genom dizileme verileri ve TNV verileri kullanılarak, bu türler arasında ortak olan ve dizisi korunmuş miRNA'lar tespit edildi. Yeni *Citrus* miRNA'larının tespiti için dizisi korunmuş olan miRNA'lar, veri tabanlarından ve literatür taraması sonucunda elde edilen miRNA'lar kullanılarak filtrelendi. Korunmuş miRNA'ların hangi genlerin ifadenme seviyesini düzenlediği ve bu hedef genler üzerindeki etkileri psRNAtarget yazılımı kullanılarak belirlendi. Korunmuş miRNA'ların metabolik yollardaki rollerinin incelenmesinde, hedef genler ve CitrusCyc veri tabanından elde edilen iki farklı *Citrus* türüne ait metabolik yollar kullanıldı.

Bulgular: Çalışmamızda toplamda 1504 miRNA tespit edildi. 270 miRNA'nın 34 farklı *Citrus* türleri arasında ortak ve 111 miRNA dizisinin bu türler arasında korunmuş olduğu tespit edildi. 111 miRNA'nın 36 tanesi ilk kez bu çalışma ile literatüre

kazandırıldı. Başka bir araştırma grubu tarafından literatüre kazandırılan *Citrus* isomir'lerinin, çalışmamızın filtreleme basamaklarında elimine edildiği saptandı. *Citrus* türleri arasında korunmuş miRNA'ların, farklı türlerde farklı hedef genlere ve dolayısıyla da farklı metabolik mekanizmalara etki ettiği tespit edildi.

Sonuç: Bir biyoinformatik uygulaması olan bu çalışmada, miRNA dizi analizine genomik veriler dahil edilerek miRNA alanındaki çalışmalara ve tanımlara farklı bir bakış açısı getirilmiştir.

Anahtar Sözcükler: mikro RNA, yeni nesil dizileme, biyoinformatik.

DISCOVERY OF NOVEL MIRNAS AND THEIR VARIANT ANALYSIS USING BIOINFORMATICS APPROACHES: *CITRUS* MODEL ORGANISM

Cankut ÇUBUK

Department of Medical Informatics, Health Sciences Institute, Dokuz Eylül University,
İnciraltı, 35340, İzmir / TURKEY
cankutcubuk [at] gmail [dot] com

ABSTRACT

Objective: The aim of this study is discovering of the novel micro RNAs (miRNAs) accurately through the analysis of existing NGS data with the bioinformatics methods, obtaining the conserved miRNAs with their variant analysis and identifying the impact of the miRNAs on their target genes and their roles in the metabolic pathways.

Method: Since all the required data for this study such as the whole genome and sRNA sequencing were exist for *Citrus* species, we used it as the model organism in this work. The preprocessed and filtered data were inputed into The UEA sRNA Workbench software for miRNA identification. The identified miRNAs were controlled by tRNAscan-SE and NOVOMIR tools to detect the false positive miRNA results. The conserved miRNAs across 34 *Citrus* species were determined using their whole genome and single nucleotide variant (SNV) data. The conserved miRNAs were searched in the literature and databases to obtain the list of novel *Citrus* miRNAs which were discovered for the first time in our study. psRNAtarget web tool was used for the target gene analysis of the conserved miRNAs. To find the metabolic pathways that are regulated by the conserved miRNAs, we mapped the target genes into the metabolic pathways of two *Citrus* species that were available in CitrusCyc database.

Results: In our study, we identified 1504 miRNAs. 270 miRNAs were found common across 34 *Citrus* species. Sequence of 111 out of 270 miRNAs were conserved and 36 of these conserved miRNAs were novel for *Citrus*. The isomirs which have been

found by another research group were discarded from our study by the filtering steps applied. We found that the conserved miRNAs had different target genes in different *Citrus* species and these miRNAs were regulating different metabolic pathways through their target genes.

Conclusion: In this bioinformatics based study, we integrated the different types of data from varied sources using the bioinformatics algorithms. By using the genomics data in our study, we add a distinctive frame to miRNA research studies.

Keywords: micro RNA, next generation sequencing, bioinformatics.

1. GİRİŞ VE AMAÇ

Mikro RNA'lar (miRNA) ortalama 22-23 nükleotit uzunluğuna sahip, küçük ve protein kodlamayan RNA sınıfındandır. Son yıllarda, miRNA inhibe edici veya miRNA gibi davranan sentetik oligonükleotit dizileri, kanser gibi çoklu genetik değişikliklere bağlı olarak gelişen hastalıkların tedavisinde yaygın bir biçimde kullanılmaktadır [1]. Transkripsiyon sonrası gen ifadenmesindeki önemli rolleri, miRNA'ları fonksiyonel genomik çalışmalarının merkezinde tutmakta ve önemini artırmaktadır [2, 3].

Yeni nesil dizileme (YND) çalışmalarının yüksek maliyeti ve araştırmacılara açık dizileme verinin eksikliği, miRNA tanımlama çalışmalarında farklı veri kombinasyonlarının kullanılmasına yol açmıştır [4, 5, 6]. Fakat günümüzde, YND ile yapılan çalışmaların maliyetinde meydana gelen hızlı azalma, organizmaların tüm genom ve transkriptom çalışmaların miktarını artırmaktadır. Bu artış, mevcut olan küçük RNA dizileme verilerinin tekrardan analizlenmesine olanak sağlamaktadır. Yeni nesil küçük RNA dizileme yöntemi, belli uzunluk aralığında sahip tüm küçük RNA'ları sınıflandırma yapmadan dizilemektedir. Bir çalışma sonunda dizileme cihazından elde edilen ham veri, küçük RNA dizileme verisi olarak adlandırılmaktadır ve küçük RNA terimi farklı RNA türlerini içinde barındıran genel bir isimdir. Küçük RNA dizileme verisi içinde miRNA dizisi olanlar biyoinformatik teknikler ile tespit edilebilmektedir. miRNA tanımlama çalışmalarının analiz mantığı birbirine çok yakındır [7]. miRNA geni, intergenik, intronik ya da ekzonik bölgede yer alabilmektedir [8]. Buradan yola çıkarak, sayısal olarak en çok miRNA miktarının, tüm genom dizilerinin ve türe özgü küçük RNA dizileme verilerinin kullanılması ile bulunacağını söyleyebiliriz.

Yeni nesil dizileme platformlarının teknolojik gelişimi isomir olarak adlandırılan miRNA türevlerinin tespit edilebilmesine olanak sağlamaktadır. miRNA türevlerinde saptanan tek nükleotit varyantları, gelişigüzel meydana gelen dizileme hatalarından kaynaklanabileceği gibi miRNA biyogenezi sırasında da meydana gelebilmektedir. Bu sebeplerden dolayı, bulunan yeni miRNA'ların miktarı kadar bu miRNA'ların doğru pozitif olması diğer bir önemli unsurdur. Doğru miRNA tanımlama kriterlerini belirlemek için birçok farklı bilimsel çalışma yapılmıştır ve bu kriterlerin en önemlileri

tek çatı altında toplanmıştır [9].

Bu çalışmamızda mevcut YND verilerini kullanarak, biyoinformatik araçlar yardımı ile yeni miRNA'ların doğru olarak tanımlanması, tanımlanan miRNA'ların farklı türler arasındaki korunmuşluğunun incelenmesi, miRNA'ların hedef genleri üzerindeki etkilerinin ve metabolik yollardaki rollerinin belirlenmesi hedeflenmiştir. Çalışma genel bir biyoinformatik uygulamasıdır ve organizmaya özgü değildir. Aynı çalışma, gerekli veriler sağlandığı sürece insan için de uygulanabilir. Çalışmada gerek duyulan tüm genom ve küçük RNA dizileme verilerinin uygunluğundan dolayı, *Citrus* (Narenciye) çalışmamızda model organizma olarak kullanılmıştır.

2. GENEL BİLGİLER

2.1. RNA çeşitleri ve miRNA

Ribonükleik asit (RNA) polimerik yapıda olup bir azotlu baz, bir riboz şeker ve bir fosfat içeren nükleotitlerden oluşmaktadır. RNA'lar protein kodlayanlar ve kodlamayanlar olarak iki sınıfa ayrılmaktadırlar. Protein kodlayanlar mesajcı RNA'lardır (mRNA). Protein kodlamayanlar ise taşıyıcı RNA (tRNA), ribozomal RNA (rRNA), küçük susturucu RNA (siRNA), küçük nükleolar RNA (snoRNA), küçük nükleer RNA (snRNA), piwi-etkileşimli RNA (piRNA), mikro RNA (miRNA), vb. olarak kendi içinde çeşitlenmektedirler. Kodlanmayan RNA'ların bir kısmı dizi uzunlukları bakımından küçük RNA'ları oluşturmaktadırlar ve mikro RNA'lar bu kısımda yer alırlar.

miRNA ilk kez 1993 yılında Lee R.C. ve arkadaşları tarafından keşfedilmiştir. Lin-4 geni üzerinde yaptıkları çalışmada, mRNA'dan daha kısa uzunluklara sahip olan ve lin-14 geninin üç üssü çevrilmeyen bölgesindeki farklı kısımlara kısmen tamamlayıcı olan yeni RNA moleküllerini elde etmişlerdir [10]. Tarihte ikinci miRNA, 2000 yılında keşfedilmiş olan let-7'dir. Let-7 canlının gelişim zamanlamasını düzenlemektedir ve birçok canlıda yüksek düzeyde korunmuştur [11, 12]. 2001 yılı itibarı ile küçük RNA alanındaki çalışmalar artmış ve farklı fonksiyonlara sahip olabileceği düşünülen let-7 ve lin-4 benzeri küçük RNA'lar miRNA olarak adlandırılmaya başlamıştır [13].

miRNA'ları diğer küçük RNA'lardan ayıran en belirgin özellikleri; öncül miRNA (pre-miRNA) dizisinin sahip olduğu sap-ilmik (hairpin) yapısı, bu yapının dizi içeriği ve termodinamik özellikleridir [14, 15]. Öncül miRNA dizisinin bir ucunda olgun miRNA bulunur. Sap-ilmik yapısının oluşabilmesi için olgun miRNA dizisine tamamlayıcı olan diğer bir miRNA dizisi, öncül miRNA dizisinin diğer ucunda yer almaktadır. Bu dizi yıldız miRNA olarak adlandırılır. Yıldız miRNA'nın yok denecek kadar az ifadeleniş olması, onun işlevsiz olduğu düşüncesini uzun bir süre gündemde tutmuştur. Fakat günümüzde iki dizinin işlevsel olabileceği bilindiği için öncül miRNA üzerindeki bu

iki miRNA (olgun ve yıldız) üç üssü (3p) ve beş üssü (5p) olarak da adlandırılmaktadır. miRNA adlandırılmasında standart olarak şunlar kullanılmaktadır [16]:

1. Organizma adının kısaltması (Örn: Homo sapiens için hsa).
2. “miR” ya da “mir” ön takısı.
3. Sıralama amaçlı bir numara.
4. Aynı aileden olan miRNA'lar için farklı harfler.
5. Olgun ve yıldız miRNA dizi ayrımı için sırayla 3p ve 5p ekleri veya sadece * simgesi kullanılır.

Aşağıda miRNA adlandırılması ile ilgili bir örnek verilmiştir.

hsa-miR-181a-5p: 181 ailesinden, öncül mikro RNA'nın beş üssü ucunda yer alan *Homo sapiens* yıldız miRNA'sı.

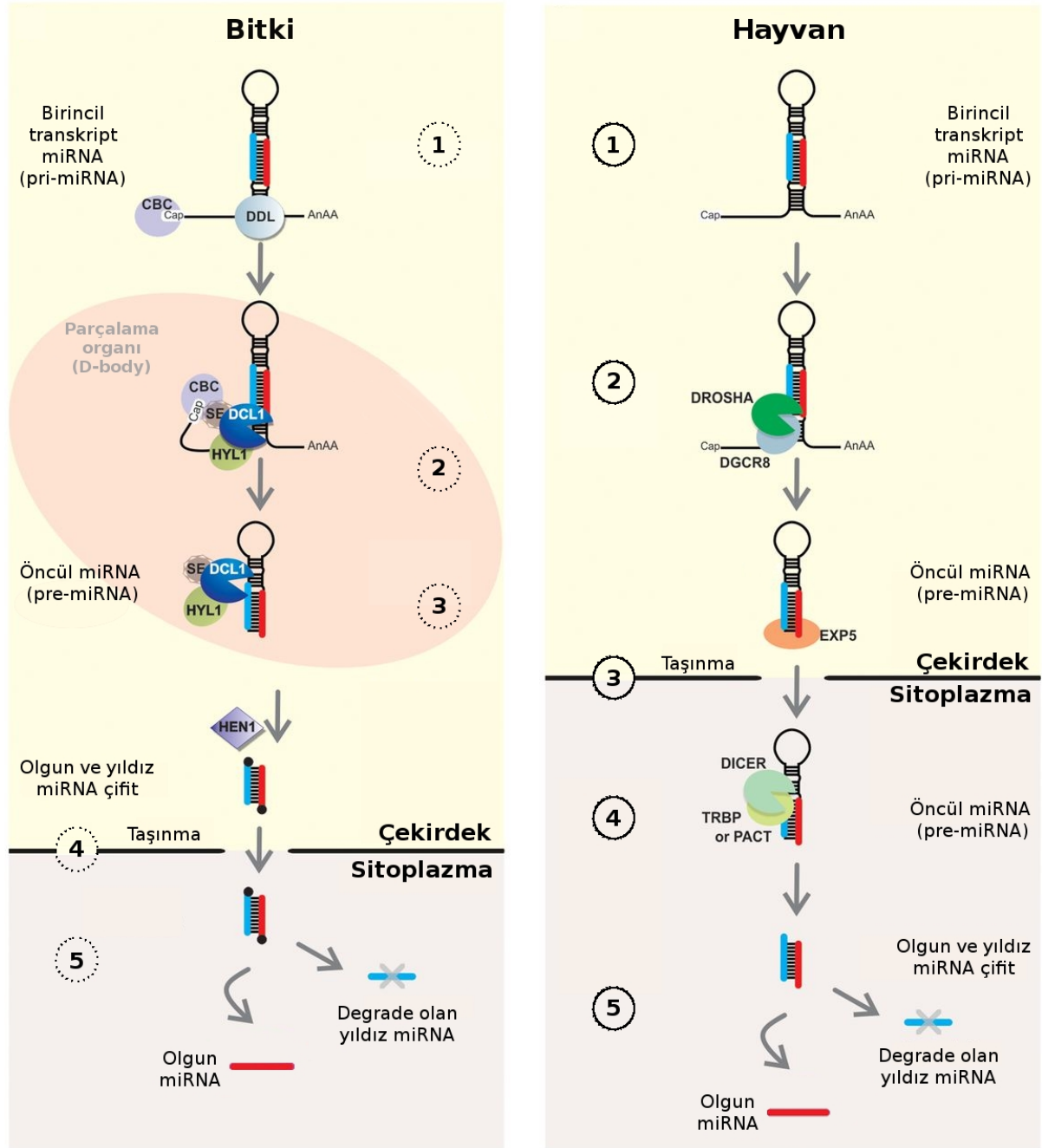
2.2. miRNA biyogenezi

Olgun miRNA'nın oluşumu dört temel aşamada gerçekleşmektedir. Hayvan ve bitkiler aynı biyogenez basamaklarına sahip olmasına rağmen bu basamakların sırası ve bu basamaklarda kullanılan enzim ve protein kompleksleri farklılık göstermektedir [17, 18]. Bu basamaklar hayvan ve bitki için Tablo 1'de ve Şekil 1'de gösterilmektedir. Bitki miRNA biyogenezi hayvan biyogenezinden ayıran temel farklılık, işleyişin hücre çekirdeğinden sitoplazmaya taşınma basamağıdır. Hayvanlarda sitoplazmada gerçekleşen bir enzimatik kesim, bitkilerde hücre çekirdeğinde gerçekleşmektedir. Bitkilerde olgun-yıldız miRNA çifti hücre dışına taşınmadan önce üç üssü ucundan RNA-metiltransferaz Hua-Enhancer1 (HEN1) tarafından metillenir. Bu sayede olgun-yıldız miRNA çifti küçük RNA degrade edici nükleaz enzimlerinden korunmuş olur. Daha sonra bu miRNA çifti, HASTY proteini tarafından sitoplazmaya taşınır ve burada olgun-yıldız miRNA çifti birbirinden ayrılır. Yıldız miRNA genellikle

degradasyona uğrayarak işlevsizleşir. Olgun miRNA ise RNA-indüklenmiş susturma kompleksi (RISC) ile birleşerek transkripsiyon sonrası gen ifadenmesinin düzenlenmesinde rol alır [19, 20].

Tablo 1. Hayvan ve bitki için miRNA biyogenez basamakları.

miRNA biyogenez basamakları	Sıralama Bitki	Sıralama Hayvan
Birincil transkript miRNA'nın (pri-miRNA) miRNA geninden RNA polimeraz II enzimi ile yazılması.	1	1
pri-miRNA'nın ribonükleaz III enzimi ile ~ 60 - 900 [70, 71] nükleotit uzunluğuna ve hairpin yapısına sahip pre-miRNA'ya dönüştürülmesi. ● DCL1 enzimi bitkilerde ● DROSHA enzimi hayvanlarda	2	2
pre-miRNA'nın, olgun ve yıldız miRNA çiftine dönüştürülmesi. ● DCL1 enzimi bitkilerde ● DICER enzimi hayvanlarda	3	4
İşleyişin hücre çekirdeğinden sitoplazmaya taşınması. ● HASTY proteini bitkilerde ● Exportin-5 proteini hayvanlarda	4	3
Olgun-yıldız miRNA çiftinin helikaz enzimi ile birbirinden ayrılması.	5	5

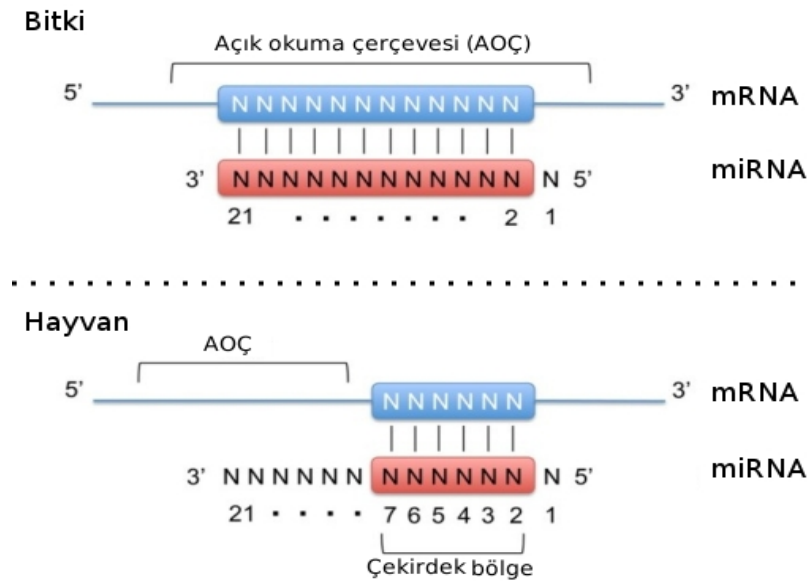


Şekil 1. Hayvan ve bitki miRNA'sının biyogenez basamakları [17].

2.3. miRNA'nın hedef genleri üzerindeki etki mekanizması

miRNA'lar hedef genleri üzerinde gen ifadenmesini azaltıcı yönde bir etkiye sahiptirler. Bitki ve hayvan miRNA'larının işlevselliği ve hedef genlerini tanımaya yardımcı olan dizi bölgeleri arasında farklılıklar vardır. Hayvan miRNA'ları hedef genlerine 5 üssü ucundaki ilk sekiz nükleotit (çekirdek bölge) ile zayıf bir şekilde

bağlanır iken, bitki miRNA'ları hedef genlerine tamamen ve kusursuza yakın bir eşleme ile bağlanır [21]. miRNA ile hedef gen arasındaki dizi eşleşmesi Şekil 2'de gösterilmektedir. Hayvan miRNA'ları hedef genin ifadenmesini azaltıcı etkiye sahip iken, bitki miRNA'ları hedef genin ifadenmesini azaltıcı ya da gene ait mRNA transkriptlerini parçalayıcı etkiye sahiptir. Hayvan miRNA'ları genellikle hedef genin üç üssü çevrilmeyen bölgesine tamamlayıcıdır. Bitki miRNA'ları için öncelikli bir bölge bulunmamakla beraber genin protein kodlayıcı kısmı olan açık okuma çerçevesine tamamlayıcıdır [22, 23, 24].



Şekil 2. miRNA ile hedef gen arasındaki dizi eşleşmesi [25].

2.4. miRNA'ların biyoteknolojide kullanımı

miRNA'ların gen ifadenmesini düzenleyici etkisi onların biyoteknoloji (sağlık, tarım, biyomühendislik) alanında kullanımını artırmıştır. Özellikle miRNA'ların vücut sıvılarından kolaylıkla izole edilebiliyor olması, biyopsi örneği alınmaksızın hastalık teşhisinin konulmasında ve hastalığın seyrinin takibinde biyobelirteç olarak kullanılmalarını yaygınlaştırmıştır [26]. Bunun yanı sıra miRNA inhibe edici ve miRNA gibi davranan sentetik oligonükleotit dizileri, kanser gibi çoklu genetik değişikliklere bağlı olarak gelişen hastalıkların tedavisinde yaygın bir biçimde kullanılmaktadır [1]. Tablo 2'de hastalıklarla ilişkisi bulunan bazı miRNA'ların listesi verilmektedir.

miRNA'lar tarım alanında ürün kalitesini arttırmak, bitkileri dış etkenlere karşı dayanıklı hale getirmek ve biyomühendislik alanında mikroorganizmaların metabolik yollarını modifiye etmek için kullanılmaktadır. Metabolik yolların modifikasyonu ile amaçlanan; mikroorganizmaya az girdi ile kısa sürede, yüksek miktar ve saflıkta istenilen ürünü ürettirmektir. Bu ürünler arasında antibiyotik başta olmak üzere çeşitli ilaçlar yer almaktadır.

Tablo 2. Hastalıklarla ilişkisi bulunan bazı miRNA'lar [27].

Hastalık	Düzensiz çalışan miRNA	miRNA'nın hedef genleri
Obezite	miR-122	AMPK
Obezite	miR-143	ERK5
Tourette sendromu	miR-189	SLITRK1
Kronik lenfositik lösemi	miR-29b; miR-181b	TCL1
Kronik lenfositik lösemi	miR-15; miR-16	BCL2
Akciğer kanseri	miR-17; miR-92	E2F1
Akciğer kanseri	let7	RAS
Tiroid kanseri	miR-221; miR-222; miR-146b	KIT
Gliyoblastom	miR-21	-
Hepatit C virüs enfeksiyonu	miR-122	HCV
Germ hücreli testis tümörü	miR-372; miR-373	LATS2
Kalp hipertrofisi	miR-195	-
B- hücreli lenfoma	miR-17; miR-92	E2F1
B- hücreli lenfoma	miR-155	hAT1R
Şeker hastalığı	miR-375	MTPN

2.5. miRNA çalışmalarında kullanılan moleküler teknikler

Bu alanda yapılan çalışmalar, bilinen miRNA'ların ifadenmesinin (ekspresyon) ölçümü ve yeni miRNA'ların belirlenmesi olarak iki kısma ayrılmaktadır.

miRNA çalışmalarında kullanılan moleküler tekniklerin detayları Tablo 3'te verilmiştir. Tablo 3'te belirtilen yüksek hacimli tekniklerin sonuçları, aynı tabloda verilen düşük hacimli teknikler ile doğrulanmalıdır.

Bu tez çalışması ile amaçlanan, YND verilerini kullanarak farklı kriterlerdeki (yeni, ortak ve korunmuş) *Citrus* miRNA'ların belirlenmesi olduğu için tez çalışmasında sadece YND hakkında ve ıslak laboratuvar deney protokollerini içermeyen genel bilgi verilmektedir.

Tablo 3. miRNA çalışmalarında kullanılan moleküler teknikler.

	Ekspresyon ölçümü	Yeni miRNA belirleme	Örnek hacmi	Maliyet	Sonuç doğrulaması
Northern blot	Evet (sadece göreceli)	Hayır	Düşük	Düşük	Gerektirmez
Polimeraz zincir reaksiyonu (PZR)	Evet (sadece göreceli)	Hayır	Düşük	Düşük	Gerektirmez
Gerçek zamanlı (kantitatif) PZR	Evet	Hayır	Orta	Orta	Gerektirmez
Mikrodizin	Evet	Hayır	Yüksek	Yüksek	Gerektirir
Yeni Nesil Dizileme	Evet	Evet	Yüksek	Yüksek	Gerektirir

2.6. Yeni nesil dizileme

Modern anlamda DNA dizilemesi, 1977 yılında Maxam and Gilbert tarafından Harvard üniversitesinde ve aynı sene Sanger ve arkadaşları tarafından Cambridge, İngiltere'de geliştirilen birbirinden farklı iki metot ile gündeme gelmiştir [28, 29]. Sanger tarafından geliştirilen yöntemin daha kolay olması, daha az toksik ve radyoaktif maddeye ihtiyaç duyması bu metodu daha popüler hale getirmiş ve

dizileme alanındaki gelişmeler bu teknik üzerinden ilerlemiştir. Yıllar içinde geliştirilen yarı otomatik dizileme sistemleri günümüzde Sanger dizileme olarak adlandırılmakta ve dizilemede altın standart olarak kullanılmaktadır. Sanger dizilemenin sahip olduğu düşük örnek analizleme hacmi ve yüksek analiz zamanı günümüzde sıklıkla kullanılan YND yöntemlerine olan ihtiyacın doğmasını sağlamıştır. Yeni nesil dizileme yöntemlerinin ortak özelliği, kılcal elektroforez bazlı Sanger dizileme cihazlarının kapasitesini aşan dizileme reaksiyonlarını tek bir deneyde gerçekleştirebilmektir. Yeni nesil dizileme sistemleri, Sanger yöntemi ile uzun yıllar süren dizileme projelerini haftalar içinde sonuçlandırmayı mümkün kılmaktadır [30]. YND için şu terimler kullanılmaktadırlar; yeni nesil dizileme, ikinci nesil dizileme, yüksek hacimli dizileme veya yüksek miktar paralel dizileme (next generation sequencing, second generation sequencing, high throughput sequencing ve massive parallel sequencing).

Günümüzde değişik firmalara ait yeni nesil dizileme platformları bulunmaktadır. Her platformun kendine özgü dizileme teknolojisi vardır. Bir çalışmada kullanılacak YND platformunun seçimi, o çalışmanın türüne göre değişiklik göstermektedir. Örneğin, referans genomu bilinmeyen organizmaların genom dizilemelerinde uzun okumalar yapan, bilinen genomların yeniden dizilenmesinde kısa okumalar yapan, filogenetik ve biyobelirteç çalışmalarında ise düşük hata payı ile okuma yapan dizileme cihazları kullanılmalıdır [31, 32]. Tablo 4a ve 4b'de bazı YND platformlarının özellikleri verilmiştir [33].

Yeni nesil dizileme platformları farklı dizi okuma teknolojilerine sahiptir. Fakat okunacak olan dizilerin deneysel olan ön hazırlık aşaması genelleştirilebilir. Bu aşamalar;

- Genetik materyalin (GM) izole edilmesi.
- Gerektiği taktirde GM'nin istenilen uzunluklarda parçalara ayrılması.
- Bu parçalara sentetik olan adaptör dizilerinin takılması.
- PZR yöntemi ile GM veya GM parçaların çoğaltılması.
- PZR ürünlerinin saflaştırılması.
- Saf PZR ürünlerinin YND sistemi tarafından okunmasıdır.

Yeni nesil dizileme sistemleri standart olarak dizi okumalarını *fastq* formatında kullanıcıya sunmaktadır.

Tablo 4a. Yeni nesil dizileme platformlarının kullanım alanlarına göre verimliliği (Birinci kısım).

Marka	Model	Büyük Genom	Ekzom	Küçük Genom	Amplikon
Illumina	HiSeq X Ten	YV	DV	DV	DV
	Hi Seq 2500	YV	YV	OV	OV
	NextSeq 500	YV	YV	YV	YV
	MiSeq	DV	YV	YV	YV
PacBio	RS II (P6 C4)	OV	DV	YV	YV
	Sequel	YV*	DV	YV	YV
Ion Torrent (ThermoFisher)	PGM 314	DV	DV	OV	YV
	PGM 316	DV	DV	YV	YV
	PGM 318	DV	DV	YV	YV
	PI	DV	YV	YV	OV
	PII	OV	YV	YV	OV
SOLiD (ThermoFisher)	SOLiD 5500xl	YV	YV	OV	OV
	SOLiD 5500xl Wildfire	YV	YV	OV	OV
	SOLiD 5500	YV	YV	OV	OV
	SOLiD 5500 Wildfire	YV	YV	OV	OV
454 (Roche)	GS FLX +	DV	DV	YV	YV
	GS Jr.	DV	DV	YV	YV
Oxford Nanopore Technologies	MinION	-	-	-	-
	PromethION	-	-	-	-
YV=Yüksek verim, OV=Orta verim, DV=Düşük Verim, *Yüksek maliyet					

Tablo 4a. Yeni nesil dizileme platformlarının kullanım alanlarına göre verimliliği (ikinci kısım).

Marka	Model	Transkriptom	Gen İfadelenmesi	Epigenomik	Metagenomik
Illumina	HiSeq X Ten	DV	DV	DV	DV
	Hi Seq 2500	YV	OV	OV	YV
	NextSeq 500	YV	YV	YV	YV
	MiSeq	DV	YV	YV	DV
PacBio	RS II (P6 C4)	YV*	DV	DV	YV*
	Sequel	YV*	DV	DV	YV*
Ion Torrent (ThermoFisher)	PGM 314	DV	DV	DV	DV
	PGM 316	DV	OV	OV	DV
	PGM 318	DV	OV	OV	DV
	PI	YV	YV	YV	DV
	PII	YV	YV	YV	OV
SOLiD (ThermoFisher)	SOLiD 5500xl	YV	OV	OV	YV
	SOLiD 5500xl				
	Wildfire	YV	OV	OV	YV
	SOLiD 5500	YV	OV	OV	YV
	SOLiD 5500 Wildfire	YV	OV	OV	YV
454 (Roche)	GS FLX +	OV	DV	DV	OV
	GS Jr.	DV	DV	DV	DV
Oxford Nanopore Technologies	MinION	-	-	-	-
	PromethION	-	-	-	-

YV=Yüksek verim, OV=Orta verim, DV=Düşük Verim, *Yüksek maliyet

Tablo 4b. Yeni nesil dizileme platformlarının teknik özellikleri.

Marka	Model	Alt model	Toplam veri miktarı	Koşum süresi	Dizi okuma uzunluğu
Illumina	HiSeq X Ten*		1.8 Tb	3 gün	2 X 150 bp
		HT v4	1 Tb	6 gün	2 X 125 bp
		HT v3	600 Gb	11 gün	2 X 100 bp
	Hi Seq 2500	Rapid	180 Gb	40 saat	2 X 150 bp
		High	129 Gb	29 saat	2 X 150 bp
		Mid	39 Gb	26 saat	2 X 150 bp
	MiSeq		15 Gb	65 saat	2 X 300 bp
PacBio	RS II (P6-C4)		1 Gb	2 saat	10 - 15 kb
	Sequel		10 Gb	2 saat	10 - 15 kb
Ion Torrent (ThermoFisher)	PGM 314		100 Mb	4 saat	400 bp
	PGM 316		1 Gb	5 saat	400 bp
	PGM 318		2 Gb	7 saat	400 bp
	PI		10 Gb	4 saat	200 bp
	PII		32 Gb	4 saat	100 bp
SOLiD (ThermoFisher)	SOLiD 5500xl		95 Gb	6 gün	2 X 60 bp
	SOLiD 5500xl Wildfire		240 Gb	10 gün	2 X 50 bp
	SOLiD 5500		48 Gb	6 gün	2 X 60 bp
	SOLiD 5500 Wildfire		120 Gb	10 gün	2 X 50 bp
454 (Roche)	GS FLX+		700 Mb	23 saat	< 1 kb
	GS Jr.		35 Mb	10 saat	700 bp
Oxford Nanopore Technologies	MinION	Fast mode	84 Gb	2 gün	230 - 300 kb
	PromethION	Fast mode	512 Gb	2 gün	230 - 300 kb
bp = 1 nükleobaz, kp = 1000 nükleobaz Mb = Megabayt, Gb = Gigabayt, Tb = Terabayt					

2.7. Küçük RNA dizileme analizinde kullanılan biyoinformatik araçlar

2.7.1. Yazılımlar

SRA Toolkit: *sra* formatındaki yeni nesil dizileme verilerini; *fastq*, *fasta*, *sam*, *sff*, *qual* gibi okunabilir formatlara dönüştürür. Komut satırında çalışır [34].

Cutadapt: Yeni nesil dizileme çalışmalarında, dizi parçalarının eksiksiz ve doğru okunabilmesi için bu dizi parçalarına üç üssü ve beş üssü uçlarından adaptör adı verilen sentetik oligonükleotitler eklenir. Cutadapt yazılımı, YND dizi okumalarından adaptörler dizilerini uzaklaştırır. Komut satırında çalışır [35].

FastQC: Yeni nesil dizileme platformları dizileri hatalı okuyabilmektedir. Özellikle dizinin sonlarına doğru bu hata oranı reaksiyona giren enzim ve substratların verimliliğini kaybetmesi ile artmaktadır. Cihazdan kaynaklanan hatalarının yanı sıra, ıslak laboratuvarlarda cDNA veya DNA dizi kütüphanelerinin hazırlanmasında meydana gelen kontaminasyonlar ve dizi kütüphanelerinden iyi elimine edilmemiş serbest sentetik oligonükleotit dizileri gibi kalıntılarda hatalara sebebiyet vermektedir. Yeni nesil dizileme çıktı formatı olan *fastq*, bir dizinin her bir nükleotiti için doğruluk değerleri bulundurur. FastQC, *fastq* formatındaki YND verilerini okuyarak, verilerin kalitesi hakkında bilgi verir. Grafiksel kullanıcı arayüzüne sahiptir ve komut satırında da çalışmaktadır [36].

FASTX Toolkit: *fasta* ve *fastq* formatındaki dizi okumalarının işlenmesinde kullanılmaktadır. Bu işlemlerden en yaygın olarak kullanılanları; *fastq* formatındaki dizilerin kullanıcı tarafından belirlenen kalite kriterlerine göre kırılması ve *fasta* formatına çevrilmesidir. Komut satırında çalışır [37].

The UEA sRNA Workbench: Küçük RNA veri setlerinin analizi ve görselleştirilmesinde kullanılmaktadır. Grafiksel kullanıcı arayüzüne sahiptir ve komut satırında da çalışmaktadır [38].

tRNAscan-SE: Bir dizinin kendi üzerine katlanarak oluşturabileceği olası ikincil

yapıları hesaplayarak dizinin taşıyıcı RNA (tRNA) olup olmadığını kontrol eder. Web tabanlı bir yazılımdır [39].

NOVOMIR: Destek vektör makineleri (DKM) yaygın olarak kullanılan bir sınıflandırma algoritmasıdır. NOVOMIR, DKM yöntemini kullanarak bir diziyi bitki miRNA'sı olan ya da olmayan olarak 2 sınıfa ayırır. NOVOMIR, öğrenme veri seti olarak doğru pozitif olduğu bilinen bitki öncül miRNA dizilerini ve doğru pozitif olduğu bilinen miRNA olmayan RNA dizileri kullanmaktadır. Komut satırında çalışır [40].

BEDTools: Yeni nesil dizileme dizi okumalarının referans genoma haritalanması sonucunda elde edilen; dizilerin referans genom üzerindeki konum bilgilerinin, diziler ile referans genom arasında bulunan nükleotit farklılıklarının (tek nükleotit varyantı) ve yüksek boyutlardaki haritalama verilerinin sıkıştırılarak saklandığı değişik veri formatlarının (*bam*, *bed*, *gff*, *gtf*, *vcf* vb.) okunması, birbirine çevrilmesi ve karşılaştırılmasında kullanılmaktadır. Komut satırında çalışır [41].

Burrows-Wheeler Aligner (BWA): Dizilerin referans genom gibi çok uzun dizilere haritalanmasında kullanılmaktadır. Çalışmanın türüne ve kullanılan YND platformuna bağlı olarak, elde edilen dizi okumaları farklı uzunluklar göstermektedir (Tablo 4b). BWA, kısa ve uzun dizi okumalarının haritalanması için *aln* ve *bwasw* olarak iki farklı algoritmaya sahiptir. Komut satırında çalışır [42].

The Genome Analysis Toolkit (GATK): Yeni nesil dizileme verilerinin analizi için kapsamlı bir yazılımdır. Birincil odak noktası, DNA dizi varyantlarının tespiti ve örneklerin genotiplendirilmesidir. Komut satırında çalışır. Bu çalışmada direk olarak kullanılmamıştır [43].

Integrative Genomics Viewer (IGV): Haritalama, varyant, anotasyon vb. genomik veri setlerinin görselleştirilmesinde kullanılmaktadır. Grafiksel kullanıcı arayüzüne sahiptir [44].

psRNATarget: miRNA'nın hedefi olan genlerin bulunmasında ve bu hedef genler

üzerindeki etkisinin incelenmesinde kullanılmaktadır. Web tabanlı bir yazılımdır [45].

2.7.2. Veri tabanları

Gene Expression Omnibus (GEO): GEO, farklı formatlardaki genomik ve transkriptomik veri setlerinin belirli standartlarda ve halka açık olarak kaydedildiği bir veri tabanıdır [46]. Web sitesi: <http://www.ncbi.nlm.nih.gov/geo/>

Phytozome: Bitki genomlarıyla ilgili çeşitli bilgilerin depolandığı ve karşılaştırmalı analizlerin yapılabildiği bir web platformdur [47]. Web sitesi: <https://phytozome.jgi.doe.gov/pz/portal.html>

Citrus Genome Database: *Citrus* türlerine ait genomik, genetik ve ıslah verilerinin tutulduğu bir veri tabanıdır [48]. Web sitesi: <https://www.citrusgenomedb.org/>

CitrusCyc Pathways Database: *Citrus* türlerine ait metabolik yolak bilgilerinin depolandığı ve yolak analizlerinin yapılabildiği bir veri tabanıdır [49]. Web sitesi: <http://ptools.citrusgenomedb.org/>

miRBase: Deneysel olarak veya bilgisayar ortamında belirlenmiş, yüzden fazla organizmaya ait miRNA dizilerinin, hedef genlerinin, yeni nesil dizileme vb. bilgilerinin bulunduğu en geniş kapsamlı miRNA veri tabanıdır [9]. Web sitesi: <http://www.mirbase.org/>

PMDB: Deneysel olarak veya bilgisayar ortamında belirlenmiş bitki miRNA bilgilerinin saklandığı bir veri tabanıdır [50]. Web sitesi: <http://bioinformatics.cau.edu.cn/PMRD/>

Rfam: Tüm RNA çeşitleri hakkında kapsamlı bilginin saklandığı bir veri tabanıdır [51]. Web sitesi: <http://rfam.xfam.org/>

2.7.3. Veri formatları

sra: Dizi okuma arşivi (sequence read archive) adlı veri tabanının, dizileme verilerini saklamak için kullandığı, sıkıştırılmış, ikili dosya biçimidir.

fastq: Dizin ve bu diziyi oluşturan her bir nükleobazın cihaz tarafından doğru okunma (kalite) değerinin tutulduğu metin esaslı bir dosya biçimidir. Birinci satır dizi adını, ikinci satır nükleobaz dizisini, üçüncü satır dizi hakkında açıklamayı (isteğe bağlı olarak boş bırakılabilir) ve dördüncü satır nükleobazların kalite değerini içerir. Birinci satır @, üçüncü satır + karakteri ile başlamalıdır. Başlangıç karakteri (@ ve +) ile sonrasından gelen ikinci karakter (harf veya rakam) arasında boşluk bulunmamalıdır. Yaygın olarak kullanılan bir formattır. Aşağıda *fastq* formatına ait bir örnek verilmiştir.

```
@Dizi adi
ACTACACGATCGACACATATTGCG
+Dizi hakkında açıklama (isteğe bağlı)
!""((((***+))%%%%%%%%+)(%%%%%%%%
```

fasta: Sadece dizi bilgisinin tutulduğu metin esaslı bir dosya biçimidir. Yaygın olarak kullanılan bir formattır. Dizi adı > karakteri ile başlamalıdır ve bu karakter ile sonrasından gelen ikinci karakter (harf veya rakam) arasında boşluk bulunmamalıdır. Aşağıda *fasta* formatına ait bir örnek verilmiştir.

```
>Dizi adi
ACTACACGATCGACACATATTGCG
```

sam: Günümüzde kullanılan, dizi haritalama yazılımlarının standart çıktı biçimidir. Referans genom haritalanan bir dizinin; genom üzerindeki konumu, haritalama kalite değeri, var ise haritalanan bölgedeki dizi eklenme/silme bilgisi, nükleobaz dizisi, bazların doğru okunma değeri vb. bilgilerin sekme aralığı ile saklandığı, dizinlenebilir, metin esaslı bir dosya biçimidir.

bam: *sam* dosya biçiminin sıkıştırılması ile elde edilen ikili dosya biçimidir.

bed, gff ve gtf: DNA, RNA ve protein dizilerinin işlevsel olan veya olmayan kısımlarının ve bu kısımlara ait konum özelliklerinin sekme aralığı ile saklandığı metin esaslı bir dosya biçimidir. Kimliklendirme (anotasyon) dosyası olarak adlandırılmaktadırlar.

vcf: Tek nükleotit varyantı (TNV), eklenme, silinme, yapısal varyant, kopya sayısı varyantları vb. bilgilerin sekme aralığı ile tutulduğu, dizinlenebilir, metin esaslı bir dosya biçimidir.

3. GEREÇ ve YÖNTEM

3.1. Araştırmanın tipi

Yapılan tez çalışması mevcut verilerin biyoinformatik teknikler ile analizlenmesidir. Bu yönüyle tez uygulamalı bir araştırmadır ve tanımlayıcı niteliktedir.

3.2. Araştırmanın yeri ve zamanı

Araştırmanın teorik çalışmaları, yöntem ve algoritma geliştirme kısımları Dokuz Eylül Üniversitesi, Sağlık Bilimleri Enstitüsü, Medikal İnformatik Anabilim Dalı, İzmir, Türkiye’de, 2012-2013 yıllarında yapıldı. Araştırmanın pratik çalışmaları Príncipe Felipe Research Center, Computational Genomics Department, Valensiya, İspanya’da, 2013-2016 yılları arasında yapıldı. Tezin yazım aşaması ve kontrolleri 2016-2019 yılları arasında tamamlandı.

3.3. Araştırmanın evreni ve örnekleme

Araştırmanın evreninde 34 farklı *Citrus* türü bulunmaktadır.

3.4. Çalışma materyali

Bu çalışmada aşağıdaki veriler kullanılmıştır:

- *Citrus* küçük RNA dizileme verileri
- *Citrus* referans genom dizileri
- *Citrus* transkript dizileri
- Filtreleme amacıyla kullanılan miRNA olmadığı bilinen diziler

- 34 farklı *Citrus* türüne ait tüm genom dizilemesi ve tek nükleotit varyantları

Biyoinformatik analizler için, Príncipe Felipe Research Center, Computational Genomics Department, Valensiya, İspanya'da bulunan paralel hesaplamaya olanak sağlayan yüksek performanslı bilgisayar kümesi kullanılmıştır. Küme 16 bilgisayardan oluşmaktadır. Bilgisayarların özellikleri aynı olup şu şekildedir; Debian Linux işletim sistemi, 12 çekirdek, ~130 GB RAM ve ~100 TB veri depolama alanı.

3.5. Araştırmanın değişkenleri

Uygulama tipi bir çalışma olması nedeniyle bu araştırmanın değişkeni bulunmamaktadır.

3.6. Veri toplama araçları

3.6.1. Analiz için gerekli verilerin elde edilmesi

Bu bölümde, çalışmada kullanılan veriler ve kaynaklar hakkında bilgiler verilmektedir.

3.6.1.1. *Citrus* küçük RNA dizileme verisi

Küçük RNA dizileme verileri GEO veri tabanından elde edilmiştir. Çalışmada kullanılan küçük RNA dizileme verileri için anahtar kelimeler (*citron, citrus, cumquats, eremocitrus, fortunella, grapefruit, murcott, kumquat, lime, limon, mandarin, microcitrus, murcott, orange, pomelo, poncirus, pummelo, reticulata, tangerine, tangor, trifoliata, mirna, micro RNA, sRNA, small RNA*) GEO'da sorgulanmıştır. Tablo 5'te GEO'dan elde edilen ve bu çalışmada kullanılan küçük RNA dizileme verilerine ait bilgiler verilmektedir.

Tablo 5. Çalışmada kullanılan küçük RNA dizileme verileri.

GEO erişim kodu	<i>Citrus</i> türü	Doku	Gelişim basamağı	Genotip
GSM712528	<i>Citrus sinensis</i>	Yaprak	Genç ve tamamen açılmış yaprak	Yabani tip
GSM712529	<i>Citrus sinensis</i>	Çiçek	Tamamen açılmış çiçek	Yabani tip
GSM712530	<i>Citrus sinensis</i>	Meyve	Olgun meyve	Yabani tip
GSM455228	<i>Citrus sinensis</i>	Meyve	Çiçeklenmeden 170 gün sonraki meyve	Yabani tip
GSM549098	<i>Citrus trifoliata</i>	Çiçek ve meyve karışık	Çiçek tomurcuğu, yarı açık ve tam açık çiçek, farklı olgunluktaki meyve	Yabani tip
GSM1086751	<i>Citrus aurantifolia</i>	Yaprak	-	Yabani tip

3.6.1.2. *Citrus referans* genom dizisi

Citrus clementine referans genom dizisi, Amerika Birleşik Devletleri Tarım Bakanlığı'na ait *Citrus* genom veri tabanından (Citrus Genome Database) elde edilmiştir. Kullanılan *Citrus clementine* referans genom dizisi; *JGI Citrus clementine assembly, v1.0* (182).

3.6.1.3. *Citrus* transkript dizileri

Citrus clementine ve *Citrus sinensis* transkript dizileri, bitki genomu veri tabanı Phytozome'dan elde edilmiştir. Sırasıyla kullanılan *Citrus clementine* ve *Citrus*

sinensis transkript dizileri; *Cclementine_182_transcript.fa* (v9.0) ve *Csinensis_154_transcript.fa* (v9.0).

3.6.1.4. Filtreleme amacıyla kullanılan diziler

Küçük RNA'ların (miRNA, tRNA, siRNA, vb.) yapısal özellikleri birbirleri arasında yakın benzerlik göstermektedir. Bu sebepten dolayı miRNA belirleme aşamasından önce, kodlanmayan ve dizisi bilinen küçük RNA (mikro RNA hariç) dizileri çalışma veri setinden filtrelenerek uzaklaştırılmalıdırlar. Buna ek olarak, aynı nükleotit veya nükleotit gruplarının tekrarlanması ile oluşan ve karmaşıklığı düşük olan diziler yanlış pozitif sonuç oranını arttıracığı için çalışma öncesinde veri setinden çıkarılmalıdır. Kodlanmayan küçük RNA dizileri, RNA ailesi (familya) veri tabanı Rfam'dan elde edilmiştir. Nükleotit tekrarlarından oluşan diziler bilgisayar ortamında simule edilmiştir.

3.6.1.5. Otuz dört farklı *Citrus* türünün tüm genom dizilemesi ve tek nükleotit varyantları

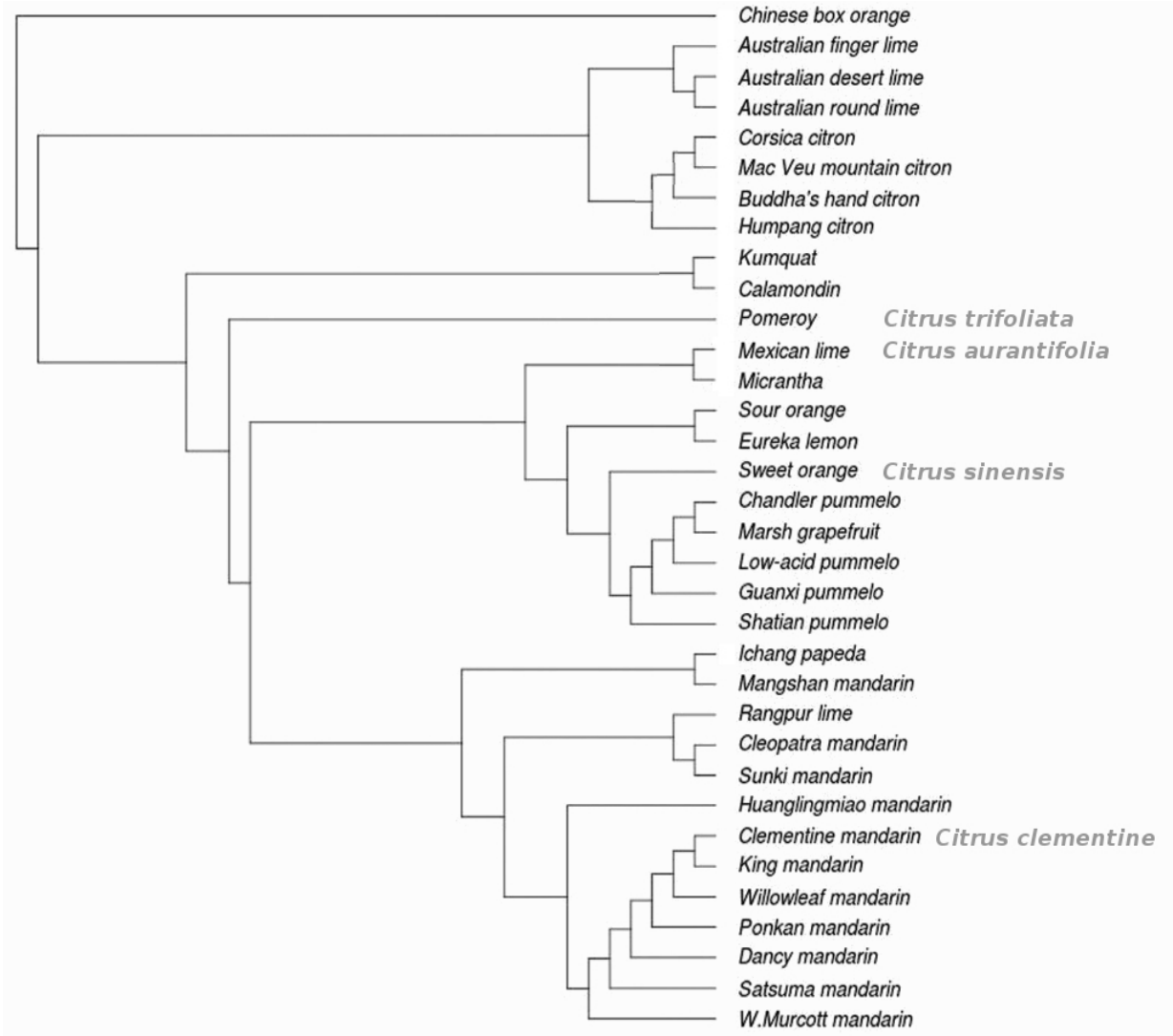
Citrus türüne ait tüm genom dizileme verileri Citrusseq-Citrusgenn [52] projesine ait olup topluma açık durumda olmadığı için işlenmemiş verilere doğrudan ulaşım sağlanamamıştır. Elde edilen veriler, tüm ön işlemleri tamamlandıktan sonra *Citrus* referans genomuna haritalanmış dizileme ve çözümlemesi yapılmış TNV verileridir. Bu veriler, Citrusseq-Citrusgenn proje ortağı Príncipe Felipe Research Center, Computational Genomics Department, Valensiya, İspanya tarafından sağlanmıştır [53, 54].

Referans genomu haritalanmış tüm genom dizileme verileri *bam*, TNV verileri ise *vcf* formatında elde edilmiştir. *Bam* dosyaları, genom okuma dizilerinin *Citrus clementine* referans genomunda hangi bölgelere denk geldiğinin ve bu bölgelerin kaç defa okunmuş olduğunun bilgisini içermektedir. Bir bölgenin fazla okunması (coverage) o bölgedeki TNV'lerin doğru pozitif olarak çözümlenmesinde büyük önem taşımakta, ayrıca referans genomu (örn: *Citrus clementine* referans genomu) haritalaması yapılmış dizilerin ait olduğu türe (örn: *Citrus maxima*) veya organizmaya (örn: *Punica granatum*) özgü referans genomunun çıkarılmasında, yani haritalanmış

olan dizilerin tek bir diziye indirgenmesinde de önemli rol oynamaktadır. VCF dosyaları, bir TNV'nin referans genom üzerindeki yerini ve bu TNV'nin ne kadar doğru pozitif olabileceği ile ilgili bilgiler taşımaktadır.

Tüm genom dizileme verilerinin işlenmesi, dizilerin referans genoma haritalanması ve TNV'lerin çözömlenmesinde kullanılan metotlar Jose Carbonell-Caballero ve arkadaşları tarafından 2015 yılında aynı örnekler kullanılarak kloroplast genomu için yapılan çalışmada detaylı olarak açıklanmıştır [55].

Bu tez çalışmasında tüm genom dizileri ve TNV verileri kullanılan 34 farklı *Citrus* türünün tam listesi Şekil 3'te verilmiştir.



Şekil 3. Çalışmada kullanılan 34 farklı Citrus türünün listesi ve filogenetik ilişkileri [55]: Citrus trioliata (pomeroy), Citrus aurantifolia (mexican lime) ve Citrus sinensis (sweet orange) küçük RNA verileri ile Citrus clementine referans genomu çalışmanın miRNA tanımlama aşamasında kullanılmıştır. Şekilde verilen tüm Citrus türlerine ait genome dizileme ve varyant verileri ortak ve korunmuş miRNA'ların belirlenmesinde kullanılmıştır.

3.6.2. Yeni nesil dizileme verilerini kullanarak, biyoinformatik araçlar yardımı ile yeni miRNA'ların doğru olarak belirlenmesi

Yeni miRNA'ların tanımlanması; küçük RNA dizilerinin ön işlenmesi ve çözümlenmesi olarak iki kısma ayrılmıştır. Ön işleme sonunda elde edilen küçük RNA'lar çalışmadaki çözümlene basamaklarında kullanılmıştır. 3.6.2., 3.6.3. ve 3.6.4. bölümlerinde anlatılan analiz basamaklarının daha anlaşılabilir ve kolay takip edilebilir olması için bu bölümlerin algoritmik akış şeması Şekil 7'de verilmiştir.

3.6.2.1. Küçük RNA dizilerinin ön işlenmesi

Tablo 5'te belirtilen veri setlerinden bazıları işlenmemiş ve *sra* formatında iken bazıları ise işlenmiş ve düz metin dosya formatındadır. Düz metin dosyaları, küçük RNA dizisini ve bu dizinin kaç kez okunduğunu içeren iki kolondan oluşmaktadır. Adaptör dizileri temizlenmiş ve kalite kontrolü yapılmış bu dizi verileri bash betiği kullanılarak *fasta* formatına çevrilmiştir (Şekil 4). *sra* formatındaki dizileme verileri, SRA Toolkit yazılımı ile *fastq* formatına çevrilmiştir. *fastq* formatındaki dizi verileri, Cutadapt yazılımı kullanılarak adaptör dizilerinden arındırılmıştır. Yeni nesil dizileme teknik sebeplerden dolayı doğruluğu düşük okumalar yapabilmekte veya örnek hazırlama basamaklarından gelen safsızlıklar dizileme verilerinin kalitesini düşürebilmektedir. Adaptör dizisi çıkarılmış dizi verilerinin kalite kontrol değerlendirilmesi FastQC yazılımı ile yapılmıştır. FASTX Toolkit yazılımı kullanılarak dizi verileri *fastq* formatından *fasta* formatına çevrilmiştir. *Fasta* formatındaki bu veriler filtreleme aşamasında kullanılmıştır.

Düz metin formatı	Fasta formatı
<i>TTCAGACAGAGAGTATTCATT</i> 2 <i>TTGAAACTAAAGGAAGTATGC</i> 3	>dizi_1 <i>TTCAGACAGAGAGTATTCATT</i> >dizi_2 <i>TTCAGACAGAGAGTATTCATT</i> >dizi_3 <i>TTGAAACTAAAGGAAGTATGC</i> >dizi_4 <i>TTGAAACTAAAGGAAGTATGC</i> >dizi_5 <i>TTGAAACTAAAGGAAGTATGC</i>

Şekil 4. Düz metin formatının *fasta* formatına çevrilmesi.

3.6.2.2. Küçük RNA dizilerinin filtrelenmesi

miRNA olmadığı bilinen, karmaşıklığı düşük veya referans genom (*Citrus clementine*) ile eşleşmeyen diziler küçük RNA veri setinden çıkarılmıştır. Bu aşamada The UEA sRNA Workbench (Filter) yazılımı kullanılmıştır. Filtreleme amacı ile kullanılan diziler, referans genom dizisi ve küçük RNA dizileme verileri bu yazılama *fasta* formatında verilmiştir ve elde edilen filtrelenmiş veriler aynı formattadır. Bu aşamalar, Tablo 5'teki her bir veri seti için ayrı ayrı uygulanmıştır. Bu aşamada yapılan filtreleme işlemleri, hem istenmeyen dizilerin uzaklaştırılması hem de miRNA belirleme aşamasındaki analiz zamanını azaltması açısından önemlidir. Filtreleme işlemi sonrasında elde kalan küçük RNA dizileri, miRNA belirleme aşamasında kullanılmıştır.

3.6.2.3. miRNA belirleme

miRNA'ların belirlenmesi ve kimliklendirilmesinde The UEA sRNA Workbench (miRCat) yazılımı kullanılmıştır. miRNA belirleme aşaması için bu yazılımın seçilmesindeki ana etmen, yazılımın bitki ve insan miRNA'larına özgü seçenekler sunmasıdır.

Çalışmada yazılıma girilen ana parametreler şunlardır;

- miRNA uzunluk aralığı: en kısa 18, en uzun 27 nükleobaz
- Küçük RNA asgari bolluk miktarı: 10
- Diğer parametrelerde bitki için varsayılan değerler kullanılmıştır (Şekil 5).

miRNA uzunluk aralığı miRBase veri tabanından (versiyon 20) elde edilen Viridiplantae (yeşil bitkiler) miRNA dizilerinin uzunluk dağılımına göre belirlenmiştir. Bu dağılım Şekil 6'da gösterilmektedir. Küçük RNA bolluk miktarı bir dizinin kaç kere okunduğunu (saptandığını) göstermektedir. Diğer bir deyişle bu değer küçük RNA transkriptinin ifadenme miktarını göstermektedir. İfadenme miktarı az olan küçük

RNA'ların doğrulanması çok zor olduğu için bolluk miktarı az olan dizilerin miRNA olarak değerlendirilmemesi bu alandaki çalışmalarda kullanılmakta ve bu değer kullanıcıya göre değişiklik göstermektedir [56]. Bu diziler olgun miRNA olarak değerlendirilmemelerine rağmen yıldız miRNA olarak değerlendirilebilmektedirler. Bu sebepten dolayı daha önceki basamaklarda dizilerin bolluk derecesine göre bir filtreleme yapılmamıştır.

miRNA belirleme aşaması her bir küçük RNA dizileme veri seti için ayrı ayrı uygulanmıştır. Analiz sonucunda elde edilen çıktı şu bilgileri içermektedir; belirlenen olgun miRNA'nın dizisi, yıldız miRNA dizisi, öncül miRNA dizisi, dizilerin (olgun, yıldız ve öncül miRNA) referans genom üzerindeki konumları, dizi uzunlukları, miRNA bolluk miktarı, miRNA'nın referans genomun kaç farklı yerinden sentezlendiği, miRNA öncül dizisinin sahip olduğu sap-ilmik yapısının minimum serbest enerji değeri, miRNA öncül dizisinin sahip olduğu guanin nükleobaz miktarının sitozin nükleobaz miktarına oranı (G/C), miRNA'nın miRBase veri tabanına göre kimliklendirilmesi.

MiRCat Information

☒ Show Main Parameters

Set Default Animal Parameters

Set Default Plant Parameters

MiRCAT Parameters

sRNA Sequence Flank Extension: 100.0

Minimum Free Energy Threshold: -25.0

sRNA Sequence Length:

Min: 18 Max: 27

Max Genome Hits: 16

Max Unpaired Bases %: 50

Max Consecutive Gaps: 3

Min GC %: 20

Min sRNA Locus Size: 1

Min sRNA Orientation %: 80

Min Hairpin Length: 60

Min sRNA Abundance (W): 10

Loci Separation Distance: 200

Min Paired Bases: 17

Maximum Overlap %: 80

P_Val: 0.05

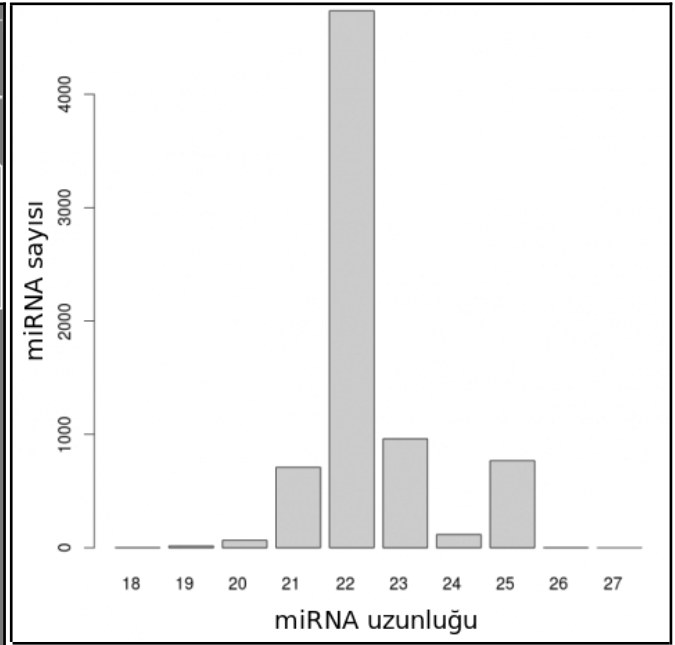
Thread Count: 10

☒ Allow Complex Loops

miRBase

miRBase version: 20

(latest installed version: 20)



Şekil 6. miRBase veri tabanından (v20) elde edilen Viridiplantae (yeşil bitkiler) miRNA dizilerinin uzunluk dağılımı.

Şekil 5. miRNA belirlenmesinde kullanılan parametreler.

3.6.2.4. Yanlış pozitif miRNA'ların çalışmadan çıkarılması

Küçük RNA dizilerinin filtrelenmesi aşamasında, miRNA olmadığı saptanan veya miRNA olma olasılığı düşük olan diziler çalışmadan çıkarılmış olmasına rağmen, miRNA belirleme aşamasında elde edilen sonuçlar yanlış pozitiflik içerebilmektedir.

Bu aşamada uygulanan filtreleme, 3.6.2.3.'te belirlenen miRNA'lara ait öncül miRNA dizilerinin özelliklerini baz almaktadır. Sırasıyla tRNAscan-SE ve NOVOMIR yazılımları kullanılarak yanlış pozitif olabilecek miRNA'lar tespit edilmiş ve çalışmadan çıkarılmıştır.

3.6.3. Belirlenen miRNA'ların sınıflandırılması

3.6.3.1. Varyant analizi: Otuz dört farklı Citrus türü için ortak olan miRNA'ların belirlenmesi

Çalışmada kullanılan küçük RNA dizileme verileri farklı *Citrus* türlerine ait olup kullanılan referans genom ise sadece *Citrus clementine* türüne aittir. Bu sebepten dolayı, belirlenen bir miRNA'nın bazı Citrus türlerinin genomunda bulunmaması olasıdır. Bir miRNA'nın 34 *Citrus* türleri içinde geçerli olduğunu söyleyebilmemiz için bu miRNA'nın öncül dizisinin çalışmada kullanılan tüm *Citrus* örnekleri için dizilenmiş olması gerekmektedir. Bu bilgi, *Citrus clementine* referans genomuna haritalanmış tüm genom dizileme verilerinden (*bam* formatında) elde edilebilmektedir. miRNA öncül dizisinin *Citrus clementine* referans genomu üzerinde bulunduğu bölgelerin *bam* dosyaları içinde aranmasında bedtools yazılımı ve bash betiği kullanılmıştır. Bu aşamada elde edilen bilgi, belirlenen bölgeye haritalanabilmiş genom dizileme verilerinin bolluk miktarıdır. Bir miRNA'nın ilgili *Citrus* türünün miRNA'sı olarak değerlendirilebilmesi için aranan bölgenin %90'ı dizilenmiş ve bu bölgeye en az bir bolluk değerinde genom dizi okumaları haritalanmış olmalıdır. Bolluk derecesinin bir olarak seçilmesinin sebebi referans genom üzerinde birden fazla yere denk gelen genom dizi okumalarının çalışmadan çıkarılmış olmasıdır. Yani referans genoma haritalanan her dizi haritalandığı bölgeye özgüdür. 34 *Citrus* türünün hepsinde genomik konumları dizilenebilmiş olan miRNA'lar ortak miRNA olarak

değerlendirilmiştir.

3.6.3.2. Varyant analizi: elde edilen miRNA'ların farklı türler arasındaki korunmuşluğunun incelenmesi

34 *Citrus* türü için ortak olan miRNA dizilerinin bu türler arasındaki korunmuşluğunun değerlendirilmesinde, olgun miRNA dizilerinin TNV barındırmaması kriter olarak kullanılmıştır. Ortak miRNA'ların referans genom üzerindeki pozisyonlarının başlangıç ve bitiş noktaları arasında kalan bölgelerdeki TNV miktarı *vcf* dosyalarından elde edilmiştir. Tek nükleotit varyantı içermeyen miRNA'lar, *Citrus* türleri arasında korunmuş miRNA'lar olarak değerlendirilmişlerdir. Bu aşamada bedtools yazılımı kullanılmıştır.

3.6.3.3. Literatüre kazandırılan miRNA'ların belirlenmesi

Bilgimiz sınırları içerisinde ilk defa bu çalışma ile tanımlanmış olan *Citrus* miRNA'larını saptamak için daha önceden yayınlanmış olan *Citrus* miRNA çalışmalarının sonuçları kullanılarak korunmuş olarak tanımladığımız miRNA'lar filtrelenmiştir. Çalışmamızın uygulama kısmının tamamlandığı 2016 yılından önce konuyla ilgili yayınlanmış altı tane makale bulunmuştur (PMID: 20398412, PMID: 22952759, PMID: 23824690, PMID: 23292880, PMID: 20626894, PMID: 21796478). Bu makalelerden, toplamda 12508 farklı miRNA verisi elde edilmiştir. Bu veriler, miRNA veri tabanlarındaki tüm kayıtları da kapsamaktadır.

12508 miRNA dizisi bwa yazılımı kullanarak *Citrus clementine* referans genomuna haritalanmıştır. Haritalama sonunda elde edilen miRNA dizilerinin referans genom üzerindeki konum bilgileri, korunmuş miRNA'ların konum bilgileri ile karşılaştırılarak bir filtreleme işlemi gerçekleştirilmiştir. Konum bilgileri olarak kromozom numarası, dizi başlangıç ve bitiş noktaları kullanılmıştır.

3.6.3.4. İsomir dizilerinin kontrolü

Ehya F. ve arkadaşları ve 2013 yılında yaptıkları çalışmada yüksek miktarda (12148 tane) isomir olarak adlandırılan miRNA türevlerini saptadıklarını belirtmişler

ve bu miRNA türevlerinin listesini yayınlamışlardır [57]. Bu türevler, bir miRNA dizisinin üç üssü ve/veya beş üssü uçlarından birkaç nükleotit kaybedilmesi ve/veya kazanılması ile oluşan miRNA benzeri dizilerdir.

Çalışmamızda belirlemiş olduğumuz olgun ve yıldız miRNA dizileri için her satırda bir dizi olacak şekilde iki ayrı düz metin dosyası hazırlanmıştır. Bu dosyalar, 12148 tane isomir'in bulunduğu aynı formattaki dosya ile linux komut satırında *grep* fonksiyonu kullanılarak karşılaştırılmıştır.

Olgun ve yıldız miRNA dosyalarında bulunamayan isomir dizileri, bwa aln haritalama algoritması kullanılarak *Citrus clementine* referans genomuna haritalanmış ve haritalama sonuçları IGV ile görselleştirilmiştir.

3.6.4. Korunmuş miRNA'ların hedef genleri üzerindeki etkilerinin ve metabolik yollardaki rollerinin belirlenmesi

Çalışmanın bu kısmında sadece iki *Citrus* türü kullanılmıştır. Bunun nedeni transkript dizilerinin ve metabolik yolak bilgilerinin sadece *Citrus sinensis* ve *Citrus clementine* için elde edilebilmiş olmasıdır.

miRNA'lar tarafından regüle edilen genlerin belirlenmesinde psRNAtarget yazılımı kullanılmıştır. Bu analiz ile miRNA'ların etki ettiği transkriptler ve etkinin türü hesaplanmıştır.

Analizde kullanılan parametreler şu şekildedir;

- miRNA/hedef dizilerinin birbirine tamamlayıcılığı.
Maximum expectation: 2
- Yukarıdaki parametrenin uygulanacağı dizi uzunluğu.
Length for complementarity scoring (hspsize): 18
- miRNA hedef dizisinin ikincil yapısının kararsızlığı.
Target accessibility (UPE): 25
- RISC'in hedef bölgeye bağlanabilmesi için gerek duyduğu dizinin ikincil yapısına katkıda bulunmayan, üç üssü ve beş üssü uçlarındaki dizilerin

uzunlukları.

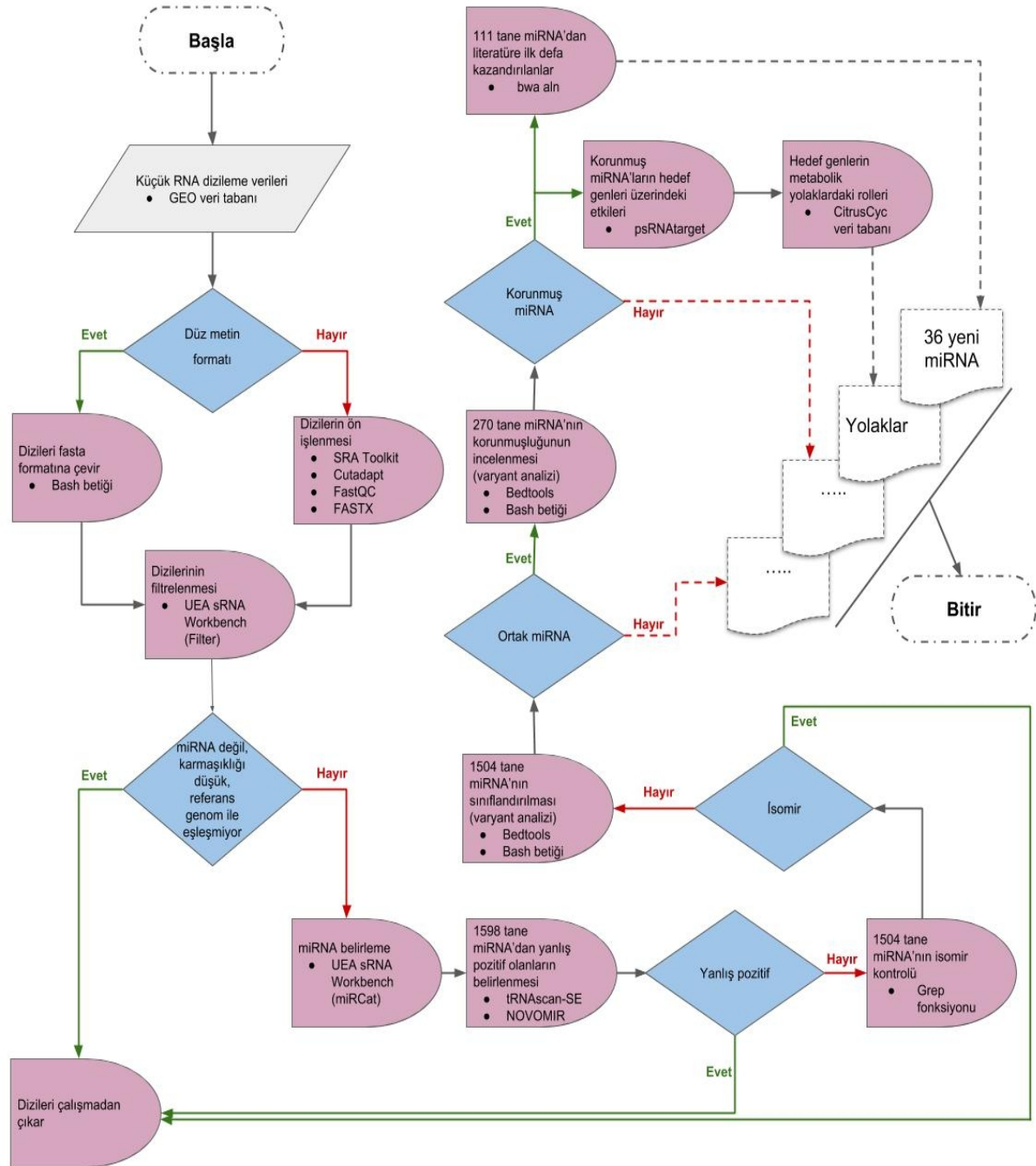
Flanking length around target site for target accessibility analysis:

17-13

- Bu parametrede belirlenen konumlar arasında meydana gelen herhangi bir hatalı dizi eşleşmesi (miRNA/hedef), hedef transkriptin parçalanması yerine inhibe olmasına sebep olur.

Range of central mismatch leading to translational inhibition: 8-12

Citrus sinensis ve *Citrus clementine* metabolik yolları *Citrus* gen adları kullanılarak modellendiği için psRNAtarget ile bulunan hedef transkriptlerin adları, gen adlarına ve protein ailesi kimlik kodlarına dönüştürülmüştür. Bu aşamada biomart (<http://www.phytozome.net/biomart/>) web aracı kullanılmıştır. Elde edilen hedef gen listeleri metabolik yolak genleri ile eşleştirilerek, bu genlerin kontrol ettiği metabolik yollar bulunmuştur. Bu aşamada *Citrus* yolak veri tabanı (CitrusCyc Pathways Database) kullanılmıştır. Her iki tür içinde ayrı ayrı yapılan bu işlemler sonrasında miRNA'lar tarafından kontrol edilen metabolik yollar karşılaştırılmıştır.



Şekil 7. Yapılan analizin algoritmik akış şeması.

3.7. Araştırma planı

Araştırma planı Gantt çizelgesi olarak aşağıda verilmiştir.

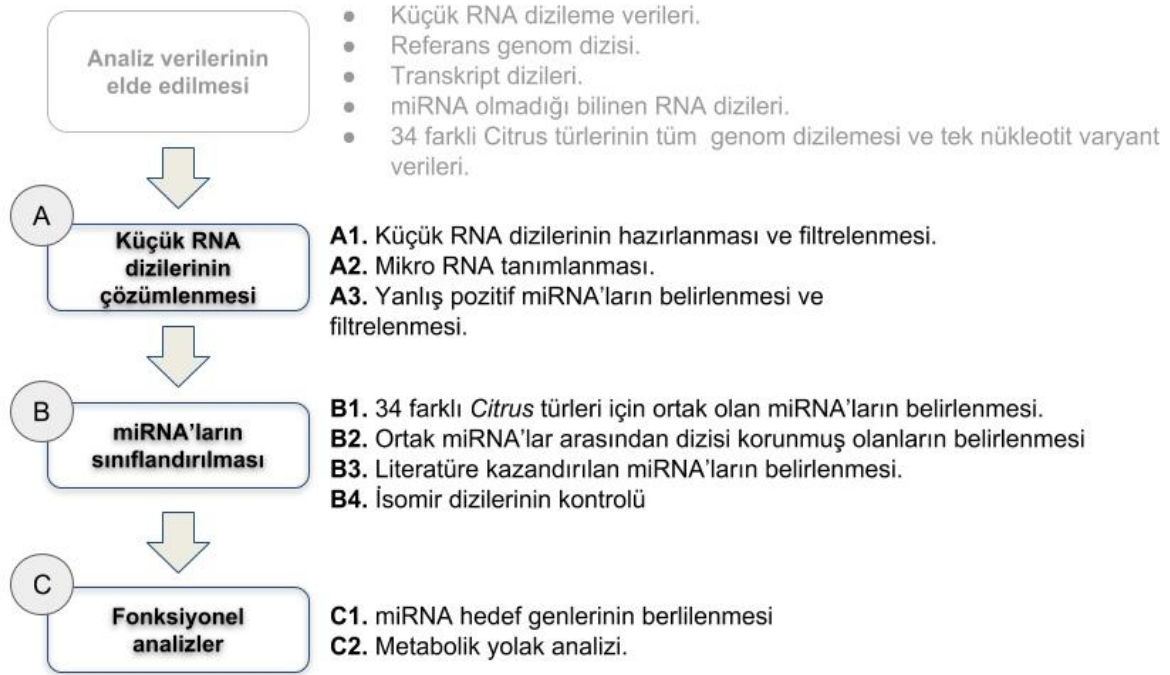
	2012 - 2013	2013 - 2014	2014 - 2015	2015 - 2016	2016 - 2017	2017 - 2018	2018-2019	2018-2019
	Güz - Bahar	Güz - Bahar	Güz - Bahar	Güz - Bahar	Güz - Bahar	Güz - Bahar	Güz	Bahar
Literatür taraması	D	D						
Etik kurul başvurusu ve onayı		D						
Teorik çalışmalar		D	D					
Algoritma seçimi			C	C				
Uygulama				C	C			
Sonuç değerlendirme					C	C		
Tez yazımı						DC	DC	DC
Tezin incelenmesi ve düzeltmeler							DC	DC
Tez savunması								DC
	D	DEU (İzmir, Türkiye)						
	C	CIPF (Valensiya, İspanya)						
	DC	DEU ve CIPF						

3.8. Verilerin değerlendirilmesi

Kullanılan biyoinformatik yazılımlarında gerek duyulan parametreler literatür taraması ile elde edilmiştir. Bu parametrelere göre miRNA tanımlama, sınıflandırma ve fonksiyonel analizler gerçekleştirilmiştir. Çalışmada karşılaştırılacak gruplar olmadığı için karşılaştırmalı istatistiksel analizler yapılmamıştır.

Verilerin değerlendirilmesi üç ana başlık takip edilerek gerçekleştirilmiştir (Şekil 8'de A,B ve C başlıkları). Her ana başlık, kendi içindeki küçük iş paketlerinin izlenmesi ile tamamlanmıştır. A1, B1, B2, B3 ve B4 basamakları diziler arası nükleobaz benzerliğini esas almaktadır. A2 basamağında, şekil 5'te belirtilen

kriterlere uyan küçük RNA okuma dizileri miRNA olarak tanımlanmıştır. Bu basamakta kullanılan program tarafından öncül miRNA dizilerinin ikincil yapılarının tanımlanmasında istatistiksel yöntem olarak hipergeometrik test kullanılmıştır ve anlamlılık düzeyi olarak da 0.05 değeri seçilmiştir (Şekil 5). A3 basamağında makine öğrenmesi ve olasılığa dayanan hesaplamalar ile çalışan yazılımlar kullanılarak, A2 basamağında tanımlanan miRNA'ların sağlaması yapılmıştır. Bu basamakta, miRNA öncül dizisi olma olasılığı düşük bulunan diziler çalışmadan otomatik olarak çıkarılmıştır. C1 basamağında, 3.6.4.'te belirtilen parametreler kullanılmıştır. C2 basamağında hedef genlerin yer aldığı metabolik yolların metin araması yapılarak bulunmuştur.



Şekil 8. Çalışmada kullanılan analiz akış şeması.

3.9. Araştırmanın sınırlılıkları

miRNA'ların belirlenmesinde, *Citrus clementine* referans genomu diğer *Citrus* türleri için de referans genom olarak kullanılmıştır. Çalışmada kullanılan her *Citrus* türüne ait referans genom ve küçük RNA dizileme verilerinin olmaması çalışmanın sınırlılıklarını oluşturmaktadır.

3.10. Etik kurul onayı

Etik kurul kararı Dokuz Eylül Üniversitesi girişimsel olmayan klinik arařtırmalar deęerlendirme komisyonunun 16.05.2013 tarihli toplantısında, 1019-GOA protokol ve 2013/18–13 karar numarası ile alındı (EK3).

4. BULGULAR

4.1. Ön işlenmesi ve filtrelemesi yapılmış küçük RNA dizileri

FastQC yazılımı kullanılarak yapılan kalite kontrol değerlendirilmesinde, adaptör dizilerden arındırılmış olan küçük RNA veri setinde herhangi bir olumsuzluk tespit edilmedi. Bu veri seti *fasta* formatına çevrildikten sonra doğrudan çalışmaya dahil edildi. Küçük RNA dizilerinin filtrelenmesi aşamasında her veri setinden farklı miktarlarda miRNA kriterlerine uymayan diziler uzaklaştırıldı (Tablo 6). Dizinin en çok kaybolduğu aşama referans genom ile eşleşmeyen dizilerin uzaklaştırıldığı kısımdır. Uzaklaştırılan dizilerin, tüm veri setinin yaklaşık olarak %50'lik bir kısmını oluşturduğu gözlemlendi.

4.2. Belirlenen miRNA'lar

miRNA belirleme aşamasında yoğun bilgisayarlı hesaplamalar yapıldığından dolayı, analizin bu aşamasına 70 GB ve üzeri rastgele erişimli hafıza tahsis edilmesi gerektiği ve bu gereksinimin küçük RNA verisinin büyüklüğü ile doğru orantılı olduğu gözlemlendi.

Altı veri setinin filtrelenmiş dizileri kullanılarak toplamda 1598 miRNA belirlendi. tRNAscan-SE ve NOVOMIR yazılımlarıyla yapılan analizde 94 tane miRNA yanlış pozitif olarak tespit edildi ve çalışmadan çıkarıldı. Kalan 1504 tane miRNA'dan sadece 76 tanesinin miRBase veri tabanına göre kimliklendirilebildiği ve bu 76 miRNA'nın 30 farklı miRNA ailesinden oluştuğu gözlemlendi (EK1).

Tablo 6. Filtreleme öncesi ve sonrası küçük RNA dizi okumalarının miktarları.

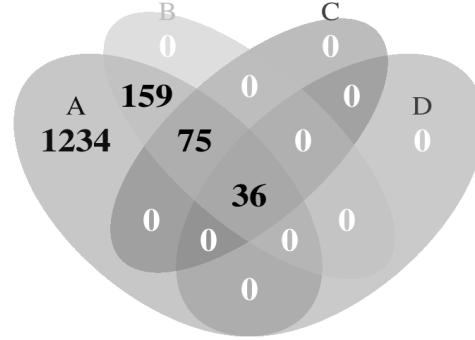
GEO veri tabanı erişim kodu	İşlenmemiş dizilerin miktarı	Dizi uzunluğuna göre filtreleme sonrası (18-27 nükleobaz)	Bolluk derecesi az (<10) diziler uzaklaştırıldıktan sonra **	miRNA olmayan ve karmaşıklığı düşük diziler uzaklaştırıldıktan sonra	Referans genom ile eşleşmeyen diziler uzaklaştırıldıktan sonra ***
GSM455228	4598696 (2787921)	4580242 (2771332)	4468008 (2757947)	4468000 (2757947)	2453144 (1217352)
GSM549098	13106573 (4876395)	13099435 (4872610)	13098472 (4871900)	11593806 (4840460)	5544238 (1446508)
GSM712528	3198491 (1634083)	3137906 (1577154)	3137791 (1577069)	3038128 (1566826)	2043221 (837346)
GSM712529	2770871 (1515486)	2670341 (1420747)	2670245 (1420671)	2534285 (1408163)	1575943 (677994)
GSM712530	2409984 (1053181)	2323751 (973471)	2323656 (973388)	2131195 (960507)	1373912 (509500)
GSM108675 1	15435261 (3252758)	14096535 (2738646)	14078690 (2728026)	10170006 (2682761)	5590925 (845483)
Parantez içinde verilen sayılar, birbirinden farklı dizilerin miktarını göstermektedir. ** Bolluk derecesi direk olarak 3.6.2.2.'de filtreleme amacıyla kullanılmamıştır. Bolluk derecesi (<10) 3.6.2.3.'te parametre olarak kullanılmıştır ve bu tabloda sadece bilgi amaçlı yer verilmiştir. *** Mikro RNA belirleme aşamasında kullanılan dizilerin miktarı.					

4.3. Belirlenen miRNA'ların sınıflandırılması

4.3.1. miRNA'ların sınıflar arası dağılımı

1504 miRNA'dan 270 tanesinin *Citrus* türleri için ortak olduğu belirlendi. *Citrus* türleri için ortak miRNA'ların 67 tanesinin miRBase veri tabanına göre kimliklendirildiği ve bu 67 miRNA'nın 29 farklı miRNA ailesinden oluştuğu gözlemlendi. 270 miRNA'dan 111 tanesinin dizilerinin, bu çalışmada yer alan 34 farklı *Citrus* türü arasında korunmuş olduğu saptandı. Korunmuş miRNA'ların 36 tanesi ilk defa bu çalışma ile literatüre kazandırıldı (Tablo 7). Belirlenen miRNA'ların sınıflar arası

dağılımı Şekil 9'da ve bu bölümde bahsedilen tüm veriler detaylı olarak EK1'de verilmiştir.



- A: Belirlenen tüm miRNA'lar (1504)
- B: Ortak miRNA'lar (270)
- C: Dizisi korunmuş miRNA'lar (111)
- D: Yeni miRNA'lar (36)

Şekil 9. Belirlenen miRNA'ların sınıflar arası dağılımı.

Tablo 7. Literatüre kazandırılan yeni miRNA'lar.

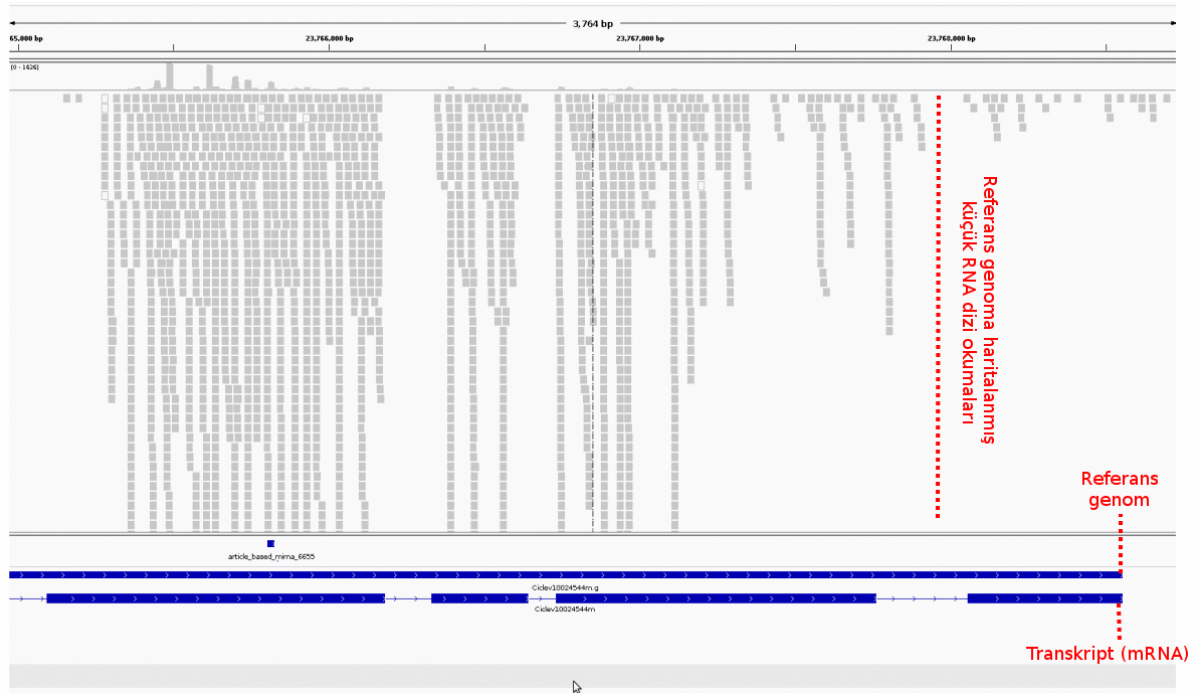
Olgun miRNA dizisi	miRNA'nın bulunduğu veri seti sayısı	Olgun miRNA uzunluğu	miRBase veri tabanı kimliği	Yıldız miRNA
AATGGGACTCTGCAGAGGAGATGT	1	24	-	Tespit edilmedi
CGGGATATTGGTGCGGTTCAA	1	21	-	Var
GGAAAATCGGGCGACAAGCAG	1	21	-	Tespit edilmedi
GTGAAGTCGTAACAAGGTA	1	19	-	Tespit edilmedi
TTCCACAGCTTTCTTGAA	1	18	mir396	Var
ATCGATGCGTGGATCTAAGGCGGA	1	24	-	Tespit edilmedi
ATGTGGTGGAATGCAATTGGGTGG	1	24	-	Tespit edilmedi
CAAAATGGACAAGAAGGAGAT	1	21	-	Tespit edilmedi
TCGGACCAGGCTTCATTCC	1	19	mir166	Var
TGCCAAAGGAGAGTTGCC	1	19	-	Tespit edilmedi
TTAGATTCACGCACAACT	1	19	-	Tespit edilmedi
AAGTTAACGGATTGGTGGACGGTA	1	24	-	Tespit edilmedi
CGAACTCTCTCCCTCAACGG	1	20	-	Var

CTGAAGTGTGTTGGGGGAACTC	1	21	mir395	Var
TGGAATCTTGATGATGCT	1	18	-	Tespit edilmedi
TTGAAGTGTGTTGGAGGAACTC	1	21	-	Tespit edilmedi
CTTGCAATGAGATCTGAGGGCC	1	22	-	Tespit edilmedi
TACACGGTGCAAGAGAAGGAGATG	1	24	-	Tespit edilmedi
TGATTAAATTATGGATGGCTC	1	21	-	Tespit edilmedi
GCAGGTTTGCCCGAGAGGTAA	1	22	-	Var
AGCCAGTATTCTGATCCAT	1	19	-	Var
GGGGGTGTAGCTCATATGG	1	19	-	Tespit edilmedi
TTCAGCTGTGGAAAGGGCAAG	1	21	-	Tespit edilmedi
AGGGATTTAACTAGGGATTGGCAA	1	24	-	Tespit edilmedi
ATATGATTGGATGAGTAAAAA	1	21	-	Tespit edilmedi
TGCCAAAGGAGAGTTGCCC	1	19	-	Tespit edilmedi
TTTGATACTTAGTTGGGATA	1	21	-	Tespit edilmedi
GGGGATGTAGCTCAGATGGT	2	20	-	Tespit edilmedi
TTTTTGGATTGAAGGGAG	2	18	-	Var
GGGGGTGTAGCTCATATGGTA	2	21	-	Tespit edilmedi
AACGGTGATGAGTCTACAGACCT	2	23	-	Tespit edilmedi
TCGGACCAGGCTTCATTCCCC	2	21	mir166	Tespit edilmedi
TCTGGAACAGGAATCGTCATA	3	21	-	Var
TCTGGAACAGGAATCGTCATA	3	21	-	Var
TGAGCTGTTTGGCTATTCTCGC	6	22	-	Var
ACAGGAGGTGGAACAAATATGAAA	6	24	-	Var

4.3.2. İsomir dizilerinin kontrolü

Ehya F. ve arkadaşları tarafından 2013 yılında yapılan çalışmada kullanılan küçük RNA veri setini de kullanmamıza rağmen, onların isomir olarak belirlediği dizilerinin sadece küçük bir kısmının bu çalışmada yıldız miRNA olarak belirlendiği tespit edildi. Yıldız miRNA'ların dizileri EK1'de verilmiştir. Çalışmamızda elde edilen, olgun ve yıldız miRNA listelerinde bulunamayan isomir dizileri ise *Citrus clementine* referans genomuna haritalanarak incelenmiştir. Bu dizilerin özellikle mRNA üzerinde konumlandığı ve parçalanmış mRNA dizilerini yüksek oranda içerdiği tespit edilmiştir

(Şekil 10). Bu durum, RNA izolasyonu aşamasında oluşabilen teknik bir sorundur. Bu istenmeyen mRNA dizilerinin, çalışmamızın filtreleme basamaklarında veri setinden uzaklaştırıldığı tespit edilmiştir.



Şekil 10. GSM1086751 veri setinde bulunan parçalanmış mRNA dizileri: gri bloklar referans genom haritalanmış küçük RNA dizi okumalarını, mavi kalın blok referans genomu, mavi kalın ve ince kesitli blok transkripti (mRNA) göstermektedir.

4.4. miRNA'ların etkiledikleri hedef genleri ve metabolik yollar

Yüz on bir tane korunmuş miRNA'nın etki ettiği *Citrus sinensis* ve *clementine* transkriptlerinin ve genlerinin sayısı Tablo 8'de verilmiştir. Transkript'lerin etkilenme şekli, transkript ve gen adları EK2'de verilmiştir. Her iki *Citrus* türünün gene adlandırmalarının birbirinden farklı olması direk olarak karşılaştırmalarını imkansız kılmıştır. Bu aşamada genlerin protein ailesi kimlik kodları kullanılmıştır. Toplamda 198 protein ailesinden, 63 tanesinin sadece *Citrus clementine* türünde, 31 tanesinin sadece *Citrus sinensis* türünde ve 104 tanesinin her iki türde de etkilendiği belirlendi (EK2).

Yüz on bir tane korunmuş miRNA'nın hedef genlerinin *Citrus sinensis* ve *clementine* metabolik yolları üzerinde farklı mekanizmalara etki ettiği tespit edildi (Tablo 9).

Tablo 8. Korunmuş miRNA'ların hedef transkriptlerinin ve genlerinin sayısı.

	Parçalanmış transkript sayısı	İnhibe olan transkript sayısı	Toplam transkript sayısı	Transkriptlerin ait olduğu genlerin sayısı
C. sinensis	243	88	331	170
C. clementine	213	59	272	207

Tablo 9. Korunmuş miRNA'lar tarafından kontrol edilen metabolik yollar.

<i>Citrus clementine</i> ve <i>Citrus sinensis</i> ortak yolları	Sadece <i>Citrus clementine</i> yolları	Sadece <i>Citrus sinensis</i> yolları
1,4-dihidroksi-2-naftoat biyosentezi II	Oksijenli solunum (sitokrom c)	Absisik asit biyosentezi
Alanin biyosentezi III	Amonyak taşınımı	Betanidin degradasyonu
Askorbat glutatyon döngüsü	Kalsiyum taşınımı	-
Askorbat geri dönüşümü (sitosolik)	Klorofil A degradasyonu	-
Yağ asiti aktivasyonu	Dhurrin biyosentezi	-
Yağ asidi omega-oksidasyonu	Homogalakturon degradasyonu	-
Folik asit poliglutamilasyonu	NAD / NADH fosforilasyonu ve defosforilasyonu	-
Folik asit dönüşümü I	Piridin nükleotit döngüsü	-
Folik asit dönüşümü II	-	-
Glisin biyosentezi I	-	-
Homosistein ve sistein birbirine çevrimi	-	-
O-dikinon biyosentezi	-	-
L-dopakrom biyosentezi	-	-

Linoleat biyosentezi	-	-
Oligomerik uruřyol biyosentezi	-	-
Fosfatidilkolin aıl dzenlemesi	-	-
Iřık solunumu	-	-
Slfonasyon iin slfat aktivasyonu	-	-
Slfat indirgenmesi II	-	-
Slfit oksidasyonu III	-	-
UDP-D-ksiloz ve UDP-D-glukuronat biyosentezi	-	-
ridin-5'-fosfat biyosentezi	-	-

5. TARTIŞMA

Yeni nesil dizileme teknolojisi yüksek çözünürlükte genomik ve transkriptomik bilgi sağlayarak canlıların kompleks sistem biyolojisinin anlaşılmasına büyük katkı sağlamaktadır. Yeni nesil dizileme teknolojisi her geçen gün daha hacimli, kolay kullanılabilir ve ekonomik bir hal almaktadır. Bu durum ile ters orantılı olan tek şey elde edilen verilerin büyüklüğü ve verilerin işleme süresidir [58]. Yeni nesil dizileme ile birlikte biyoinformatik alanındaki araştırmalarda hız kazanmıştır. Artan veri miktarı daha uygun istatistiksel ve biyoinformatik metotların gelişmesine yol açmaktadır. Hatta günümüzde YND ile biyoinformatik gelişmeleri bir döngü halinde birbirlerini beslemektedir. Yeni nesil dizileme ile elde edilen veri sadece araştırmacı tarafından cevaplanmak istenen bilimsel soruya ışık tutacak nitelikte değildir. Mevcut YND verileri, yeni verilere entegre edilerek ve yeni algoritmalar kullanılarak tekrar tekrar işlenebilir ve bu sayede farklı biyolojik mekanizmalar açıklığa kavuşturulabilir.

Bu çalışmada, daha önceden analizlenmiş ve halka açık halde olan *Citrus* küçük RNA verileri, güncel genomik ve transkriptomik verilere entegre edilerek tekrardan analizlenmiş ve güvenilirliği daha yüksek olan bir analiz tasarımının uygulaması yapılmıştır. Başka çalışmalarda analizi yapılmış olan altı tane küçük RNA dizileme verisinin (Tablo 5) kullanıldığı bu çalışmada, yoğun filtreleme basamakları kullanıldığı halde hala yeni miRNA'ların (36 tane) bulunabiliyor olması ilgi çeken bir noktadır. Bunun en önemli sebebi çalışmamızda referans genom dizisinin kullanmış olmasıdır. Tablo 5'teki verilerin analizlendiği çalışmalarda, zamanın şartlarına bağlı olarak *Citrus* referans genomu yerine, *Citrus* ifadelenmiş dizi etiketi (Expressed Sequence Tag) veya transkriptom dizileri kullanılmıştır. miRNA geni, intergenik, intronik ya da ekzonik bölgelerde yer alabildiği için [8] en çok miRNA'nın referans genom dizisinin kullanılması ile belirlenmiş olması gayet normaldir.

Birçok miRNA tanımlama çalışması, miRBase veri tabanındaki diziler ile eşleşen küçük RNA dizilerini korunmuş miRNA, eşleşmeyen fakat miRNA kriterlerine uyan küçük RNA dizilerini ise yeni miRNA olarak adlandırılmıştır. Bu sınıflandırma şekli, miRNA dizilerinin yüksek korunmuşluğa sahip olduğu algısından doğmuştur.

Fakat son yıllarda yapılan çalışmalar bu algıyı değiştirmektedir. Çalışmamızda korunmuşluk kavramını, miRBase dizi eşleştirmesini hesaba katmadan, dizinin tek nükleotit varyantı taşımaması durumunda kullandık. miRBase veri tabanını ise elde ettiğimiz miRNA'ları adlandırmak amacıyla kullandık. Bunun sonucunda 270 *Citrus* miRNA'sından sadece 111 tanesinin 34 *Citrus* türü için korunmuş olduğu tespit edilmiştir. Yani *Citrus* miRNA dizileri yüksek oranda korunmuş değildir. Tek nükleotit varyantı verilerini miRNA dizilerine entegre ederek elde ettiğimiz bu bilgi, miRNA alanında yapılan güncel çalışmaların sonuçlarını destekler niteliktedir. Ryan D. Morin ve arkadaşları 2008 yılında yaptıkları çalışmada miRNA varyantlarından bahsetmiş ve bunları isomir olarak adlandırmıştır [59]. José Carbonell ve arkadaşları 2012 yılında yaptıkları diğer bir çalışmada, insan miRNA'sının popülasyonlar arasındaki TNV'ye bağlı olarak korunmuşluğunu kaybedebildiğini göstermiştir [60]. Ling J. ve arkadaşları 2017 yılında yaptıkları bir çalışmada bazı olgun miRNA'ları ve bu miRNA'ların hedef genlerini dizilmişler ve bu çalışma ile miRNA üzerindeki bir TNV'nin hedef genler ve yolaklar üzerinde farklı etkilere sahip olduğunu ve miRNA dizisinin korunmuş olmadığını deneysel olarak kanıtlamışlardır [72]. Bu bilgiler doğrultusunda gelecekteki miRNA çalışmalarında kullanılacak olan korunmuşluk tanımının çalışmamızda olduğu şekilde kullanılacağı kolaylıkla öngörülebilir.

miRNA varyantları, dizi içerisindeki nükleobaz değişimi ile meydana gelebileceği gibi miRNA dizisinin üç üssü ve/veya beş üssü uçlarından birkaç nükleobaz kazanılması veya kaybedilmesiyle de oluşabilmektedir [61]. Ehya F. ve arkadaşları, 2013 yılında yaptıkları bir çalışmada, 12148 tane birbirinden farklı ve *Citrus aurantifolia* türüne ait isomir yayınlamışlardır. Çalışmamızda aynı küçük RNA dizileme verisini kullanmamıza rağmen benzer sonuçları elde edemedik. Ehya F. ve arkadaşları tarafından isomir olarak tespit edilen dizilerin küçük bir kısmının bizim çalışmamızda yıldız miRNA olarak belirlendiği tespit edildi. Yıldız ve olgun miRNA'lar birbirlerine üç üssü ve beş üssü uçlarında eşleşmeyen çıkıntı diziler kalacak şekilde tamamlayıcıdır. Aslında bu çıkıntılar miRNA dizilerinin DCL1 enzimi (hayvanlarda Drosha ve Dicer enzimleri) tarafından kesilmesi sırasında meydana gelir. Ayrıca üç üssü ucundaki yaklaşık iki nükleobazlık çıkıntı, bu enzim/enzimler tarafından gerek duyulan dizi kesim bölgesidir [62]. Bu açıdan bakıldığında, dikkatli yapılmayan bir

analiz sonucunda yıldız miRNA'ların isomir, isomir'lerin de yıldız miRNA olarak raporlanması olasıdır. İki çalışmanın sonuçları arasındaki büyük farkın sebebi çalışmaların analiz akış şemaları karşılaştırılarak açıklanabilir. Ehya F. ve arkadaşları, miRBase veri tabanında bulunan miRNA'lar ile tam olarak eşleşen dizileri miRNA, maksimum üç hatalı nükleobaz ile eşleşen dizileri ise isomir olarak değerlendirip, veri tabanına göre bu miRNA ve isomir dizilerini adlandırmıştır. Adlandırılması yapılan bu miRNA'lar ve isomir'ler için öncül miRNA dizileri raporlanmamıştır. Geri kalan küçük RNA dizileri ise *Citrus aurantifolia* transkript dizilerine haritalanmıştır. Bu haritalanmış dizilerden miRNA öncül dizisine sahip olanlar, yeni *Citrus aurantifolia* miRNA'sı olarak tanımlanmıştır. miRNA ve isomir belirleme aşamalarında küçük RNA'lar bolluk derecesine göre filtrelenmemiş ve olgun miRNA'nın tamamlayıcısı olan yıldız miRNA'lar raporlanmamıştır. Bizim yaptığımız çalışmada ise Ehya F. ve arkadaşları tarafından yapılan aksine fazla miktarda filtreleme basamakları kullanılmıştır. Küçük RNA filtreleme basamağını başarıyla geçen her dizi referans genoma haritalanmış ve yapısal (öncül miRNA özellikleri) olarak miRNA olduğu düşünülen diziler miRBase veri tabanına göre adlandırılmıştır. Daha detaylı yapılan bir incelemede, hem Ehya F. ve arkadaşları tarafından hem de bu tez çalışmasında kullanılan küçük RNA verisinin parçalanmış mRNA dizilerini yüksek oranda içerdiği tespit edilmiştir (Şekil 10). Bu durum, RNA izolasyonu aşamasında oluşabilen teknik bir sorundur. Parçalanmış mRNA dizilerinin büyük çoğunluğu Ehya F. ve arkadaşları tarafından isomir ve miRNA olarak raporlanırken, bizim çalışmamızda istenmeyen dizi parçaları olarak sınıflandırılmış ve çalışmadan çıkarılmıştır. Bu tespitin ardından çalışmamızda kullanılan analiz akış şemasının ve sonuçların doğruluğu kanıtlanmış, filtreleme basamaklarının ve filtreleme kriterlerinin önemi bir kez daha anlaşılmıştır. Ayrıca isomir ile ilgili yapılacak olan çalışmaların yüksek dikkat gerektirdiği vurgulanmıştır.

Bu tez çalışması bir biyoinformatik analizi olduğu için raporladığımız her miRNA'nın laboratuvar ortamında deneysel olarak doğrulanması gerekmektedir. Her miRNA'nın öncül miRNA dizisi de belirlendiği için sonuçların laboratuvar ortamında doğrulanması mümkündür [63]. Belirlenen her olgun miRNA için tamamlayıcısı olan yıldız miRNA tespit edilememiştir. Bunun sebebinin yıldız miRNA'nın biyogenez sonrasında hemen degradasyona uğruyor olması olarak düşünülmüştür. miRNA'ların

belirlenmesinde, *Citrus clementine* referans genomu diğer *Citrus* türleri için de referans genom olarak kullanılmıştır. Çalışmada kullanılan her *Citrus* türüne ait referans genom ve küçük RNA dizileme verilerinin olmaması çalışmanın sınırlılıklarını oluşturmaktadır. miRNA alanında yapılan çalışmaların daha dikkatli bir şekilde ele alınması ve çalışmaların güncel yöntemler ile tekrarlanması, 2017 yılında Taylor R. ve arkadaşları tarafından yayınlanan bir çalışmada da gündeme gelmiştir. Başlığı “ya doğru yap yada hiç yapma olan” bu çalışmanın amacı miRNA veri tabanlarında çok fazla yer tutan yanlış pozitif sonuçlara dikkat çekmek ve bu sonuçların tekrardan nasıl analizlenebileceğidir. Taylor R. ve arkadaşları örnek veri seti olarak çalışmamızda da kullanılan *Citrus sinensis* veri setini kullanmışlardır (GSM455228) [73]. Taylor R. ve arkadaşları tarafından yapılan çalışmaya atıfta bulunarak onu temel alan, Chavesa I. ve arkadaşları tarafından 2017 yılında yayınlanan diğer bir çalışma, yanlış pozitif sonuçları azaltabilecek bir miRNA analiz şeması ortaya koymaktadır. miRPursuit adını verdikleri bu analiz şeması birbirinden farklı küçük RNA sınıflandırma ve filtreleme programlarını bir arada kullanmaktadır [74]. Alptekin B. ve arkadaşları 2016 yılında yaptıkları çalışmada özellikle bitki miRNA'ları için yanlış pozitif sonuçları daha aza indirebilecek SUMIR serisi programlarını kullanıma açmışlardır [75]. Yeni analiz şeması öneren bu üç çalışma ile çalışmamızın en önemli ortak noktası, dikkatli yapılması gereken dizi filtreleme basamaklarıdır. Bu gelişmeler eşliğinde verilerin tekrardan analizlenmesini kaçınılmaz gören Van Peer G. ve arkadaşları, miRBase Tracker adını verdikleri miRNA bilgi takip sistemini geliştirmişlerdir [76]. Bu sistem ile dizileme verilerin tekrardan analizlenmesi ve miRNA veri tabanlarının elde edilen sonuçlar doğrultusunda güncellenmesi daha verimli ve kolay bir hale almıştır [77].

Bu tez çalışmasında türler arasında korunmuş miRNA'ların, farklı türlerde farklı hedef genlere etki ettiği tespit edildi. Çalışmada kullanılan metabolik yolların her iki *Citrus* türü içinde topolojisi ve gen içeriği aynıdır. Fakat 111 korunmuş *Citrus* miRNA'sının hedef genlerindeki farklılığın, metabolik yollar üzerinde farklı etkilere sahip olduğu gözlenmiştir. miRNA dizileri aynı olsa bile hedef genlerde bulunan nükleobaz farklılıkları veya miRNA hedef dizisinin bir *Citrus* türünde daha fazla miktarda gen tarafından barındırılmış olması, metabolik yollar üzerinde gözlenen farklı etkilere sebep olmaktadır. Bu durum miRNA'nın sadece dizisel olarak

korunmuşluğunun çokta anlamlı olmadığını göstermektedir. miRNA'ların fonksiyonel açıdan korunmuşluğu daha büyük önem taşımaktadır.

6. SONUÇ VE ÖNERİLER

Bir biyoinformatik uygulaması olan bu çalışmada, aynı konuda kullanılmış birbirinden bağımsız çalışma verileri ve algoritmalar birleştirilmiş, çalışmanın başarısı ve önemi model organizma *Citrus* üzerinde kanıtlanmıştır. Literatüre 34 farklı *Citrus* türü için 36 yeni ve 111 korunmuş miRNA dizisi kazandırılmıştır. Analize genomik veriler dahil edilerek miRNA alanındaki çalışmalara ve tanımlara farklı bir bakış açısı getirilmiştir.

Bu tez çalışması, Sağlık Bilimleri Enstitüsünde yapıldığı için çalışmanın sağlık alanına uygulanabilirliği önemlidir. Çalışmada kullanılan parametrelerde küçük değişiklikler yapılarak aynı analiz, insanda da kullanılabilir. Özellikle kanser gibi insan sağlığını ve ekonomisini büyük ölçüde etkileyen, tedavi yöntemleri henüz tam anlamıyla gelişmemiş hastalıkların mekanizmalarının ortaya çıkarılması, günümüzdeki en önemli araştırma konularındandır [78, 79]. Kanser Genom Atlası [64] ve Uluslararası Kanser Genom Konsorsiyumu [65] gibi projeler, verilerini halka açık hale getirerek, translasyonel tıp alanına büyük destek vermektedir. Bu projeler ile elde edilmek istenen, etik kurallar çerçevesinde deneysel veriyi en azami miktarda yararlı bilgiye dönüştürmektir. DREAM ve CAMDA yarışmaları anonimleştirilmiş gerçek verileri paylaşarak, veriyi en doğru ve anlamlı bilgiye dönüştürecek biyoinformatik algoritmalarının doğmasına olanak sağlamaktadır [66, 67]. Yukarıda bahsi geçen ve örneklerini çoğaltabileceğimiz uluslararası proje ve yarışmalar her sene yeni veriler eşliğinde daha da büyümektedir. Yeni nesil dizileme verileri bilginin saklı olduğu büyük bir okyanustur ve canlıların sistem biyolojisinde açıklanması gereken daha birçok mekanizma bulunmaktadır [80]. Bu yüzden verilerin belli aralıklar ile yeni verilere entegre edilerek ve yeni algoritmalar yardımıyla tekrardan analizlenmesi gerektiğini söylemek doğru olacaktır.

Sağlık alanında varılmak istenen nokta, genomik yani kişileştirilmiş tıptır [68]. Günümüzde büyük projeler altında yapılan analizlerin yakın gelecekte hasta yatağından laboratuvara ve laboratuvardan hasta yatağına uygulamalarında kullanılacak rutin testler olması çok yakındır [69]. Bu açıdan bakıldığında,

biyoinformatik analizlerin sađlık bilimlerinin bir parçası olacağı kaçınılmaz bir gerçektir. Bu tez çalışmasında olduğu gibi halka açık veriler ile yapılan biyoinformatik analizler, bilime katkı sağlamakta, ileriye yönelik önemli bilgi ve tecrübe kazandırmaktadır.

7. KAYNAKLAR

1. Price, C., Chen, J. (2014). MicroRNAs in cancer biology and therapy: Current status and perspectives. *Genes & Diseases*, 1(1), pp.53-63.
2. Wirth, S., Crespi, M. (2009). Non-protein coding RNAs, a diverse class of gene regulators, and their action in plants. *RNA Biology*, 6(2), pp.161-164.
3. Bartel, D. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), pp.281-297.
4. Wang, C., Wei, L., Guo, M., Zou, Q. (2013). Computational Approaches in Detecting Non- Coding RNA. *Current Genomics*, 14(6), pp.371-377.
5. Gomes, C., Cho, J., Hood, L., Franco, O. ve ark. (2013). A Review of Computational Tools in microRNA Discovery. *Frontiers in Genetics*, 4, 81.
6. Pundhir, S., Gorodkin, J. (2013). MicroRNA discovery by similarity search to a database of RNA-seq profiles. *Frontiers in Genetics*, 133.
7. Studholme, D. (2011). Deep sequencing of small RNAs in plants: applied bioinformatics. *Briefings in Functional Genomics*, 11(1), pp.71-85.
8. Olena, A., Patton, J. (2009). Genomic organization of microRNAs. *Journal of Cellular Physiology*, 222(3), pp.540-545
9. Kozomara, A., Griffiths-Jones, S. (2010). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39(Database), pp.D152-D157.
10. Lee, R., Feinbaum, R., Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5), pp.843-854.
11. Pasquinelli, A., Reinhart, B., Slack, F., Martindale, M. ve ark. (2000). The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403(6772), pp.901-906.
13. Lee, R. (2001). An Extensive Class of Small RNAs in *Caenorhabditis elegans*. *Science*, 294(5543), pp.862-864.
14. Krol, J., Sobczak, K., Wilczynska, U., Drath, M., ve ark. (2004). Structural Features of MicroRNA (miRNA) Precursors and Their Relevance to miRNA Biogenesis and Small Interfering RNA/Short Hairpin RNA Design. *Journal of*

Biological Chemistry, 279(40), pp.42230-42239.

15. Thakur, V., Wanchana, S., Xu, M., Bruskiwich, R., ve ark. (2011). Characterization of statistical features for plant microRNA prediction. BMC Genomics, 12(1).
16. Ambros, V., Bartel, B., Bartel, DP., Burge, CB., ve ark. (2003). A uniform system for microRNA annotation. RNA, 9(3), pp.277-279.
17. Bologna, N., Schapire, A., Palatnik, J. (2012). Processing of plant microRNA precursors. Briefings in Functional Genomics, 12(1), pp.37-45.
18. Du, T. (2005). microPrimer: the biogenesis and function of microRNA. Development, 132(21), pp.4645-4652.
19. Lelandais-Briere, C., Sorin, C., Declerck, M., Benslimane, A., ve ark. (2010). Small RNA Diversity in Plants and its Impact in Development. Current Genomics, 11(1), pp.14-23.
20. Zhu, J. (2008). Reconstituting plant miRNA biogenesis. Proceedings of the National Academy of Sciences, 105(29), pp.9851-9852.
21. Axtell, M., Westholm, J., Lai, E. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biology, 12(4), p.221.
22. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., ve ark. (2008). Widespread Translational Inhibition by Plant miRNAs and siRNAs. Science, 320(5880), pp.1185-1190.
23. Lim, L., Lau, N., Weinstein, E., Abdelhakim, A., ve ark. (2003). The microRNAs of *Caenorhabditis elegans*. Genes & Development, 17(8), pp.991-1008.
24. Witkos, T., Koscińska, E., Krzyzosiak, W. (2011). Practical Aspects of microRNA Target Prediction. Current Molecular Medicine, 11(2), 93–109.
25. Hou, J., Zhao, D. (2013). MicroRNA Regulation in Renal Pathophysiology. International Journal of Molecular Sciences, 14(7), pp.13078–13092.
26. Mo, M., Chen, L., Fu, Y., Wang, W., Fu, S. (2012). Cell-free Circulating miRNA Biomarkers in Cancer. Journal of Cancer, 3, pp.432-448.
27. Mack, G. (2007). MicroRNA gets down to business. Nature Biotechnology, 25(6), pp.631-638.
28. Maxam, A. M., Gilbert, W. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America,

74(2), pp.560–564.

29. Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp.5463–5467.

30. Hutchison, C. (2007). DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research*, 35(18), pp.6227-6237.

31. Loman, N., Misra, R., Dallman, T., Constantinidou, C., ve ark. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 30(5), pp.434-439.

32. Mardis, E. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), pp.198-203.

33. Allseq. URL: <http://allseq.com/kb-category/sequencing-platforms/>

34. Leinonen, R., Sugawara, H. ve Shumway, M. (2010). The Sequence Read Archive. *Nucleic Acids Research*, 39(Database), pp.D19-D21.

35. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17, pp.10–12.

36. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. URL: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

37. FASTX-Toolkit (2014). URL: http://hannonlab.cshl.edu/fastx_toolkit

38. Stocks, M., Moxon, S., Mapleson, D., Woolfenden, H., ve ark. (2012). The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, 28(15), pp.2059-2061.

39. Schattner, P., Brooks, A. ve Lowe, T. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, 33(Web Server), pp.W686-W689.

40. Teune, J. ve Steger, G. (2010). NOVOMIR: De Novo Prediction of MicroRNA-Coding Regions in a Single Plant-Genome. *Journal of Nucleic Acids*, 2010, pp.1-10.

41. Quinlan, A. ve Hall, I. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), pp.841-842.

42. Li, H. ve Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754-1760.

43. DePristo, M., Banks, E., Poplin, R., Garimella, K., ve ark. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), pp.491-498.
44. Robinson, J., Thorvaldsdóttir, H., Winckler, W., Guttman, M., ve ark. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1), pp.24-26.
45. Dai, X. ve Zhao, P. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research*, 39(suppl_2), pp.W155-W159.
46. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1), pp.207–210.
47. Goodstein, D., Shu, S., Howson, R., Neupane, R., ve ark. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1), pp.D1178-D1186.
48. Citrus Genome Database. URL: <https://www.citrusgenomedb.org/>
49. CitrusCyc Pathways Database. URL: <https://www.citrusgenomedb.org/>
50. Zhang, Z., Yu, J., Li, D., Zhang, Z., ve ark. (2009). PMRD: plant microRNA database. *Nucleic Acids Research*, 38(suppl_1), pp.D806-D813.
51. Nawrocki, E., Burge, S., Bateman, A., Daub, J., ve ark. (2014). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, 43(D1), pp.D130-D137.
52. CITRUSEQ/CITRUSGENN: a Genomic Approach to Citrus Breeding. URL: https://www.researchgate.net/publication/268119805_CITRUSEQCITRUSGENN_a_Genomic_Approach_to_Citrus_Breeding/citations
53. Terol, J., Carbonell-Caballero, J., Alonso, R., Tadeo, F., ve ark. (2015). Sequencing of 150 citrus varieties: linking genotypes to phenotypes. *Acta Horticulturae*. pp.585-589.
54. Wu, G., Terol, J., Ibanez, V., López-García, A., ve ark. (2018). Genomics of the origin and evolution of Citrus. *Nature*, 554(7692), pp.311-316.
55. Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., ve ark. (2015). A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. *Molecular Biology and Evolution*, 32(8), pp.2015-2035.

56. Motameny, S., Wolters, S., Nürnberg, P. ve Schumacher, B. (2010). Next Generation Sequencing of miRNAs – Strategies, Resources and Methods. *Genes*, 1(1), pp.70-84.
57. Ehya, F., Monavarfeshani, A., Mohseni Fard, E., Karimi Farsad, L., ve ark. (2013). Phytoplasma-Responsive microRNAs Modulate Hormonal, Nutritional, and Stress Signalling Pathways in Mexican Lime Trees. *PLoS ONE*, 8(6), p.e66372.
58. Greene, C., Tan, J., Ung, M., Moore, J. ve ark. (2014). Big Data Bioinformatics. *Journal of Cellular Physiology*, 229(12), pp.1896-1900.
59. Morin, R., O'Connor, M., Griffith, M., Kuchenbauer, F., ve ark. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Research*, 18(4), pp.610-621.
60. Carbonell-Caballero, J., Alloza, E., Arce, P., Borrego, S., ve ark. (2012). A map of human microRNA variation uncovers unexpectedly high levels of variability. *Genome Medicine*, 4(8), p.62.
61. Neilsen, C., Goodall, G. ve Bracken, C. (2012). IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, 28(11), pp.544-549.
62. Warf, M., Johnson, W. ve Bass, B. (2011). Improved annotation of *C. elegans* microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. *RNA*, 17(4), pp.563-577.
63. Schmittgen, T., Lee, E., Jiang, J., Sarkar, A., ve ark. (2008). Real-time PCR quantification of precursor and mature microRNA. *Methods*, 44(1), pp.31-38.
64. The Cancer Genome Atlas (TCGA). URL: <https://cancergenome.nih.gov/>
65. International Cancer Genome Consortium (ICGC). URL: <http://icgc.org/>
66. DREAM Challenges. URL: <http://dreamchallenges.org/>
67. CAMDA Challenges. URL: <http://contest.camda.info/>
68. Hood, L. ve Friend, S. (2011). Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature Reviews Clinical Oncology*, 8(3), pp.184-187.
69. Gómez-López, G., Dopazo, J., Cigudosa, J., Valencia, A. ve ark. (2017). Precision medicine needs pioneering clinical bioinformaticians. *Briefings in Bioinformatics*.
70. Thakur, V., Wanchana, S., Xu, M., Bruskiwich, R., ve ark. (2011). Characterization of statistical features for plant microRNA prediction. *BMC Genomics*,

12(1), 108.

71. Xia, J., Li, L., Li, T., Fang, Z., ve ark. (2017). Detecting and characterizing microRNAs of diverse genomic origins via miRvial. *Nucleic Acids Research*, 45(21), p.e176.
72. Ling, J., Luo, Z., Liu, F., Mao, Z., ve ark. (2017). Genome-wide analysis of microRNA targeting impacted by SNPs in cucumber genome. *BMC Genomics*, 18(1), 275.
73. Taylor, R., Tarver, J., Foroozani, A. ve Donoghue, P. (2017). MicroRNA annotation of plant genomes – Do it right or not at all. *BioEssays*, 39(2), p.1600113.
74. Chaves, I., Costa, B., Rodrigues, A., Bohn, A., ve ark. (2017). miRPursuit-a pipeline for automated analyses of small RNAs in model and nonmodel plants. *FEBS Letters*, 591(15), pp.2261-2268.
75. Alptekin, B., Akpinar, B. ve Budak, H. (2017). A Comprehensive Prescription for Plant miRNA Identification. *Frontiers in Plant Science*, 7, 2058.
76. Van Peer, G., Lefever, S., Anckaert, J., Beckers, A., ve ark. (2014). miRBase Tracker: keeping track of microRNA annotation changes. *Database*, bau080, 2014.
77. Backes, C., Fehlmann, T., Kern, F., Kehl, T., ve ark. (2017). miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Research*, 46(D1), pp.D160-D167.
78. Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., ve ark. (2018). Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. *Cancer Research*, 78(21):6059–6072.
79. Cubuk, C., Hidalgo, M.R., Amadoz, A., Rian, K., Salavert, F., ve ark. (2018). Differential metabolic activity and discovery of therapeutic targets using summarized metabolic pathway models. *npj Systems Biology and Applications*, 5, 7.
80. Hidalgo, M.R., Cubuk, C., Amadoz, A., Salavert, F., ve ark. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget* 8, 5160–5178.

8. EKLER

EK1. Çalışmada belirlenen 1504 tane miRNA'nın tam listesi

Bu teze ait EK1'in URL bağlantısı aşağıda verilmiştir.

EK2. Hedef genler ve genlerin ait oldukları protein aileleri

Bu teze ait EK2'in URL bağlantısı aşağıda verilmiştir.

URL: <https://drive.google.com/drive/folders/1UDwFYOWHL6bF5ZOmg1XYSMZujXP3DHEU>

Alternatif URL: goo.gl/NjGTzR

Alternatif URL: bit.ly/2OdCGgs

EK3. Etik kurul onayı

DOKUZ EYLÜL ÜNİVERSİTESİ GİRİŞİMSEL OLMAYAN ARAŞTIRMALAR ETİK KURUL KARARI

ETİK KOMİSYONUN ADI	DOKUZ EYLÜL ÜNİVERSİTESİ GİRİŞİMSEL OLMAYAN ARAŞTIRMALAR ETİK KURULU
AÇIK ADRES	Dokuz Eylül Üniversitesi Tıp Fakültesi Dekanlığı 2. Kat İnciraltı-İZMİR
TELEFON	0 232 412 22 54-0 232 412 22 58
FAKS	0 232 412 22 43
E-POSTA	etikkurul@deu.edu.tr

BAŞVURU BİLGİLERİ	DOSYA NO:	1019-GOA
	ARAŞTIRMA	UZMANLIK TEZİ ☐ AKADEMİK AMAÇLI ☐
	ARAŞTIRMANIN AÇIK ADI	Biyoformatik Teknikleri Kullanarak Yeni MikroRNA'ların Bulunması ve Varyant Analizlerinin Yapılması:Citrus Modeli
	ARAŞTIRMA PROTOKOL KODU	-
	SORUMLU ARAŞTIRMACI ÜNVANI/ADI/SOYADI ve UZMANLIK ALANI	Prof.Dr.Oğuz DİCLE Radyoloji A.D
	DESTEKLEYİCİ VE AÇIK ADRESİ	-
	DESTEKLEYİCİNİN YASAL TEMSİLCİSİ VE ADRESİ	-
	ARAŞTIRMAYA KATILAN MERKEZLER	TEK MERKEZ ☒ ÇOK MERKEZLİ ☐

DEĞERLENDİRİLEN BELGELER	Belge Adı	Tarihi	Versiyon Numarası	Dili
	ARAŞTIRMA PROTOKOLÜ	Mevcut		Türkçe ☒ İngilizce ☐ Diğer ☐
	ARAŞTIRMA İLE İLGİLİ LİTERATÜR	Mevcut		Türkçe ☐ İngilizce ☒ Diğer ☐
	BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU	Mevcut		Türkçe ☒ İngilizce ☐ Diğer ☐
	OLGU RAPOR FORMU	Mevcut		Türkçe ☒ İngilizce ☐ Diğer ☐

KARAR BİLGİLERİ	Karar No:2013/18-13	Tarih: 16.05.2013
	Prof.Dr.Oğuz DİCLE'nin sorumlusu olduğu "Biyoinformatik Teknikleri Kullanarak Yeni MikroRNA'ların Bulunması ve Varyant Analizlerinin Yapılması:Citrus Modeli" isimli klinik araştırmaya ait başvuru dosyası ve ilgili belgeler araştırmanın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş, etik açıdan çalışmanın gerçekleştirilmesinin uygun olduğuna oy birliği ile karar verilmiştir.	

ETİK KURUL BİLGİLERİ	
ÇALIŞMA ESASI	Dokuz Eylül Üniversitesi Girişimsel Olmayan Araştırmalar Etik Kurulu İşleyiş Yönergesi İyi Klinik Uygulamaları Kılavuzu
ETİK KURUL ÜYELERİ	

Unvanı/Adı/Soyadı	Uzmanlık Alanı	Kurumu	Cinsi yet	Araştırma ile ilişkili mi?		İmza
Prof.Dr.Banu ÖNVURAL (Başkan)	Tıbbi Biyokimya	DEU Tıp Fakültesi Tıbbi Biyokimya Anabilim Dalı	Kadın	E ☐	H ☒	
Prof.Dr.Besti ÜSTÜN (Başkan Yardımcısı)	Ph.D.Psikiyatri Hemşireliği	DEU Hemşirelik Fakültesi	Kadın	E ☐	H ☒	
Prof.Dr.Ş.Reyhan UÇKU	Halk Sağlığı	DEU Tıp Fakültesi Halk Sağlığı A.D.	Kadın	E ☐	H ☒	
Prof.Dr.Nejat SARIOSMANOĞLU	Kalp Damar Cerrahisi	DEU Tıp Fakültesi Kalp Damar Cerrahisi Anabilim Dalı	Kadın	E ☐	H ☒	
Prof.Dr.Ece BÖBER	Pediyatrik Endokrinoloji	DEU Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı	Erkek	E ☐	H ☒	
Prof.Dr.Hüseyin BASKIN	Mikrobiyoloji	DEU Tıp Fakültesi Mikrobiyoloji Anabilim Dalı	Erkek	E ☐	H ☒	
Prof.Dr.Vesile ÖZTÜRK	Nöroloji	DEU Tıp Fakültesi Nöroloji Anabilim Dalı	Kadın	E ☐	H ☒	
Prof.Dr.Bilgin CÖMERT	İç Hastalıkları (Yoğun Bakım B.D)	DEU Tıp Fakültesi İç Hastalıkları Anabilim Dalı	Erkek	E ☐	H ☒	
Prof.Dr.Mukaddes GÜNELİ	Tıbbi Farmakoloji	DEU Tıp Fakültesi Tıbbi Farmakoloji Anabilim Dalı	Kadın	E ☐	H ☒	
Prof.Dr.Ayşe Aydan ÖZKÜTÜK	Mikrobiyoloji	DEU Tıp Fakültesi Mikrobiyoloji Anabilim Dalı	Kadın	E ☐	H ☒	
Prof.Dr.Nihal GELECEK	Fizik Tedavi ve Rehabilitasyon	DEU Fizik Tedavi ve Rehabilitasyon Yüksek Okulu	Kadın	E ☐	H ☒	
Doç.Dr.İşıl TEKMEK	Histoloji ve Embriyoloji	DEU Tıp Fakültesi Histoloji ve Embriyoloji Anabilim Dalı	Kadın	E ☐	H ☒	
Doç.Dr.Müge KIRAY	Fizyoloji	DEU Tıp Fakültesi Fizyoloji Anabilim Dalı	Kadın	E ☐	H ☒	
Uzm.Dr.Ahmet Can BİLGİN	Hukuk	DEU Tıp Tarihi ve Etik A.D	Erkek	E ☐	H ☒	
İhsan ÇELİKDEMİR	Sağlık mensubu olmayan üye	75. Yıl Özel İlköğretim Okulu Müdür Yrd.	Erkek	E ☐	H ☒	

EK4. ARBİS araştırmacı özgeçmişi



CANKUT ÇUBUK

Kişisel Bilgiler

İletişim Bilgileri

Kimlik Numarası 37693185080
Doğum Tarihi 12/06/1985
İletişim Adresi Clinical Bioinformatic Area Fundacion Progreso Y Salud Cdca, Hospital Virgen Del
Telefon (530) 512 57 56
633226544
E-posta cankutcubuk@gmail.com
Web Adresi http://www.clinbioinfospa.es/people

Eğitim Bilgileri

01 Ekim 2014 - Şu Anda (3 yıl 10 ay)
Doktora, Bütünleşik Doktora, THE POLYTECHNIC UNIVERSITY OF VALENCIA, İSPANYA
THE POLYTECHNIC UNIVERSITY OF VALENCIA, THE POLYTECHNIC UNIVERSITY OF VALENCIA

01 Eylül 2010 - Şu Anda (7 yıl 11 ay)
Yüksek Lisans, Tezli Program, DOKUZ EYLÜL ÜNİVERSİTESİ, TÜRKİYE
SAĞLIK BİLİMLERİ ENSTİTÜSÜ, MEDİKAL ENFORMATİK (YL) (TEZLİ)

01 Eylül 2005 - 28 Ağustos 2009 (4 yıl)
Lisans, İkinci Öğrenim, EGE ÜNİVERSİTESİ, TÜRKİYE
FEN FAKÜLTESİ, BİYOKİMYA BÖLÜMÜ
Diploma Numarası: 16195

Deneyim / İşyeri Bilgileri

01 Ekim 2017 - Şu Anda (10 ay) (Tam Zamanlı)
ARGE PERSONEL, UZMAN, CLINICAL BIOINFORMATICS AREA, FUNDACION, PROGRESO Y SALUD
İdari Görev: ARGE PERSONEL

01 Kasım 2013 - 01 Kasım 2017 (4 yıl 1 ay) (Tam Zamanlı)
ARGE PERSONEL, ARAŞTIRMA GÖREVLİSİ, CENTRO DE INVESTIGACIONES PRÍNCIPE FELÍPE (CIPF)

01 Ocak 2009 - 01 Nisan 2013 (4 yıl 4 ay) (Tam Zamanlı)
ARGE PERSONEL, DİĞER (Genmar Teşhis Ürünleri Ar-Ge Lab Hizmetleri)

Yabancı Dil Bilgileri

İSPANYOLCA (Okuma: İyi, Yazma: İyi, Konuşma: İyi)

İNGİLİZCE (Okuma: İyi, Yazma: İyi, Konuşma: İyi)

Ar-Ge Yetkinlik

Makaleler

Cubuk C, Hidalgo M, Amadoz A, Pujana MA, Mateo F, Herranz C, Carbonell-Caballero J & Dopazo J (accepted 2018) A pan-cancer metabolic landscape based on gene expression integration into pathway modules. Cancer Research.

Amadoz A, **Cubuk C**, Hidalgo M, Carbonell-Caballero J & Dopazo J (2018) A comparison of mechanistic signaling pathway activity analysis methods. Briefings in Bioinformatics, bby040.

Cubuk C, Salavert F, Hidalgo M, Amadoz A, Carbonell-Caballero J & Dopazo J (2018) Differential metabolic activity and discovery of therapeutic targets using summarized metabolic pathway models. BioRxiv, 367334. (<http://metabolizer.babelomics.org/>)

Ferreira PG, Muñoz-Aguirre M, Reverter F, Godinho C, Sousa A, Amadoz A, Sodaie R, Hidalgo M, Pervouchine D, Carbonell-Caballero J, Nurtudinov R, Breschi A, Oliveira P, **Cubuk C**, Aguet F, Oliveira C, Dopazo J, Sammeth M, Ardlie KG, Guigo R. (2018) The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nature Communications, 9, 1, 490.

Carbonell-Caballero J, Amadoz A, Alonso R, Hidalgo M, **Cubuk C**, Conesa D, López-Quílez A & Dopazo J (2017) Reference genome assessment from a population scale perspective: an accurate profile of variability and noise. Bioinformatics, btx482.

Jiao Y, Hidalgo M, **Cubuk C**, Amadoz A, Carbonell-Caballero J, Vert JP & Dopazo J (2017) Signaling pathway activities improve prognosis for breast cancer, Technical Report, BioRxiv, 132357.

Hidalgo M, **Cubuk C**, Amadoz A, Salavert F, Carbonell-Caballero J & Dopazo J (2017) High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget 8, 5160-5178. (<http://hipathia.babelomics.org/>)

Salavert F, Hidalgo M, Amadoz A, **Cubuk C**, Medina I, Crespo D, Carbonell-Caballero J & Dopazo J (2016) Actionable pathways: interactive discovery of therapeutic targets using signaling pathway models. Nucleic Acids Research 44, W212-W216. (<http://pathact.babelomics.org/>)

Sanghez V, **Cubuk C**, Sebastián-Leon P, Carobbio S, Dopazo J, Vidal-Puig A & Bartolomucci A (2016) Chronic subordination stress selectively downregulates the insulin signaling pathway in liver and skeletal muscle but not in adipose tissue of male mice. Stress 19, 214-224.

Razzoli M, Frontini A, Gurney A, Mondini E, **Cubuk C**, Katz L, Cero C, Bolan P, Dopazo J, Vidal-Puig A, Cinti S & Bartolomucci A (2016) Stress-induced activation of brown adipose tissue prevents obesity in conditions of low adaptive thermogenesis. Molecular Metabolism 5, 19-33.

Alonso R, Salavert F, Garcia-Garcia F, Carbonell-Caballero J, Bleda M, Garcia-Alonso L, Sanchis-Juan A, Perez-Gil D, Marin-Garcia P, Sanchez R, **Cubuk C**, Hidalgo M, Amadoz A, Hernansaiz-Ballesteros R, Alemán A, Tarraga J, Montaner D, Medina I & Dopazo J (2015) Babelomics 5.0: functional interpretation for new generations of genomic data. Nucleic Acids Research 43, W117-W121. (<http://babelomics.org/>)

Diniz G, Aktas S, **Cubuk C**, Ortac R, Vergin C & Olgun N (2013) Tissue expression of MLH1, PMS2, MSH2, and MSH6 proteins and prognostic value of microsatellite instability in wilms tumor: experience of 45 cases. Pediatric Hematology and Oncology 30, 273-284.

Projeler

PEOPLE - MARIE-CURIE ACTIONS, ARAŞTIRMACI, Mlpm ? Machine Learning For Personalized Medicine, Yürütülen Kuruluş: ETH Zürich Basel, Switzerland (Sistemde kayıtlı olmayan kuruluş), Destek Alınan Kuruluş: CENTRO DE INVESTIGACIONES PRÍNCIPE FELIPE (CIPF) (Yurt Dışı) , 01 Haziran 2013, 01 Ocak 2017.

ULUSAL, ARAŞTIRMACI, İZKA-KAMU KURULUSLARINA BAĞLI ÇALIŞAN ARAŞTIRMACILARIN HEDEFLEDİKLERİ ARAŞTIRMA ÇALIŞMALARINI YÜRÜTEBİLECEKLERİ HİZMET AMAÇLI ARGE LABORATUARLARININ KURULARAK ŞİRKET CİROSUNDA ARTIŞIN SAĞLANMASI, Yürütülen Kuruluş: KURULUŞ GÜNCELLENMESİ GEREKİYOR, Destek Alınan Kuruluş: DiĞER (Genmar Teşhis Ürünleri Ar-Ge Lab Hizmetleri) (Yurt İçi) , 18 Ağustos 2009, 31 Mayıs 2010.