

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**MİKRODİZİN KAYNAKLI GENOTİP BİLGİSİ KULLANILARAK KOPYA SAYISI
DEĞİŞİKLİKLERİNDE MOZAIKLİK ORANININ SAPTANMASI**

Hande Melisa ACUN

Biyoinformatik Programı

YÜKSEK LİSANS TEZİ

ANKARA

2023

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

MİKRODİZİN KAYNAKLI GENOTİP BİLGİSİ KULLANILARAK KOPYA SAYISI
DEĞİŞİKLİKLERİNDE MOZAIKLİK ORANININ SAPTANMASI

Hande Melisa ACUN

Biyoinformatik Programı

YÜKSEK LİSANS TEZİ

TEZ DANIŞMANI

Dr. Öğr. Ü. Dr. Arda ÇETİNKAYA

ANKARA

2023

HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
MİKRODİZİN KAYNAKLI GENOTİP BİLGİSİ KULLANILARAK KOPYA SAYISI
DEĞİŞİKLİKLERİNDE MOZAIKLİK ORANININ SAPTANMASI

Öğrenci: Hande Melisa ACUN

Danışman: Dr. Öğr. Üyesi Dr. Arda ÇETİNKAYA

Bu tez çalışması 05/01/2023 tarihinde jürimiz tarafından "Biyoinformatik Programı"nda yüksek lisans tezi olarak kabul edilmiştir.

Jüri Başkanı: Prof. Dr. Yeşim AYDIN SON
(Ortadoğu Teknik Üniversitesi)

Tez Danışmanı: Dr. Öğr. Üyesi Dr. Arda ÇETİNKAYA
(Hacettepe Üniversitesi)

Üye: Dr. Öğr. Üyesi Dr. İdil YET
(Hacettepe Üniversitesi)

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

16 Ocak 2023

Prof. Dr. Müge YEMİŞÇİ ÖZKAN

Enstitü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

05.10.2023

Hande Melisa ACUN

1

¹“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulunun** gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, **tezin yapıldığı kurum** tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, **ilgili kurum ve kuruluşun önerisi** ile **enstitü** veya **fakültenin** uygun görüşü üzerine **üniversite yönetim kurulu** tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
 Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** tarafından karar verilir.

ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, Dr. đr. . Dr. Arda ETİNKAYA danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sađlık Bilimleri Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

Bio. Hande Melisa Acun

TEŞEKKÜR

Başta tez çalışmasının başından sonuna tüm aşamalarında kritik değerlendirmeleri ve yönlendirmeleriyle desteklerini sunan, algoritmanın oluşturulması ve yazılmasında büyük emek harcayan tez danışmanım Dr. Öğr. Ü. Dr. Arda ÇETİNKAYA'ya, tez kapsamında oluşturulan algoritmanın şekillenmesi ve R programına uyarlanması aşamalarında kritik yönlendirmeleriyle kodlamada seviye atlamamı sağlayan değerli arkadaşım Özgün Ozan KILIÇ'a, mikrodizin çalışmalarının dizaynı ve gerçekleştirilmesi konusundaki büyük yardımları ve emeğiyle Uzm. Bio. Serkan KABAÇAM'a; tez çalışması konusundaki kritik katkı ve önerileriyle tüm Biyoinformatik Anabilim Dalı öğretim üyelerine ve tezin gerçekleştirilmesinde desteğini esirgemeyen tüm Tıbbi Genetik Anabilim Dalı öğretim elemanlarına, özellikle kritik katkılarıyla Ar. Gör. Dr. Nesligül IŞIKLI ve Öğr. Gör. Dr. Ceren Damla DURMAZ ÖZDİNÇ ile kritik katkı ve tecrübeye dayalı yapıcı değerlendirmeleriyle Prof. Dr. Nurten AKARSU'ya; deneysel işlemlerin gerçekleştirilmesindeki kolaylaştırıcı katkıları sebebiyle tüm Çocuk Genetik Laboratuvarı çalışanlarına minnetle teşekkür ederim.

Ayrıca, çalışmanın gerçekleştirilebilmesi için genetik veri elde edilmesi amacıyla tez çalışmasına katkı sunan bütün gönüllülere teşekkür ederim.

Tezin tasarımı, ön verilerinin oluşturulması, oluşturulan Copy-fm algoritmasının denenmesi için veri sunulması aşamalarında ve yüksek lisans tez çalışmalarım sırasında sağladığı burs nedeniyle TÜBİTAK destekli EJP-RD projesi RiboEurope (319S062) ve Era-NET projesi EuroDBA (315S192) projeleri ile tez çalışmalarında deneysel işlemlerin gerçekleştirilmesi için kaynak sunması nedeniyle Hacettepe Üniversitesi BAP Koordinasyon Birimi tarafından THD-2021-19532 numaralı hızlı destek projesi kapsamındaki destekleri nedeniyle destekleyen kurumlara teşekkür ederim.

ÖZET

Acun, H. M., Mikrodizin Kaynaklı Genotip Bilgisi Kullanılarak Kopya Sayısı Değişikliklerinde Mozaiklik Oranının Saptanması, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoinformatik Programı Yüksek Lisans Tezi, Ankara, 2023. Kopya sayısı değişiklikleri (*Copy Number Variation, CNV*), genomik bölgelerin normal diploid (2 kopya) kopya sayısında bölgesel bir değişikliği temsil eden, uzunluğu >50 bp olan yapısal varyantlardır. Bu tanıma göre, tek hücre içerisindeki tüm lokuslar için kopya sayıları tam sayı olmak durumundadır; ancak farklı kopya sayılarına sahip alt popülasyonlardan oluşan bir hücre popülasyonu için, tüm hücre popülasyonu açısından bakıldığı zaman, tam sayı olmayan değerlerde kopya sayısı söz konusu olabilir. Bu türden CNV'lere mozaik CNV'ler (mCNV), mCNV'lerin oluşturduğu bu genetik duruma ise genetik mozaizizm denilmektedir. mCNV'lerle doğuştan ya da kanserde olduğu gibi hayatın ileriki dönemlerinde karşılaşılabilir. Özellikle kanser gibi hastalıklarda mCNV'lerin saptanması tanı ve tedavi takibinde kritik öneme sahip olduğundan, birçok araştırmacı ve ticari firma mCNV'lerin saptanmasına yönelik yöntemler geliştirmiştir. Bu yöntemler arasında SNP mikrodizin yöntemi, farklı boyutlarda mCNV'ler için genom boyu taramaya olanak tanıması sebebiyle günümüzde en sık kullanılan yöntemdir. Ancak mCNV'lerin lokalizasyonunun yanı sıra, hastalık seyri ve tedavisinin takibinde asıl kritik role sahip olan hücre popülasyonundaki bulunma oranlarına, yani mozaizizm fraksiyonuna (f_m), yönelik yetkin ve spesifik çalışma sayısı çok azdır. Bu tezde, mCNV'lerin oranını (f_m) SNP mikrodizin verisinden elde edilen BAF ve LRR verileriyle R programlama dilinde Kolmogorov-Smirnov (KS) testi ile hesaplayan, hesaplama için güven aralıklarını belirleyen Copy-fm olarak adlandırılan bir biyoinformatik araç geliştirilmiştir. Copy-fm'in güvenilirliğinin ve fonksiyonlarının test edilmesi, tarafımızca deneysel olarak oluşturulmuş mozaik delesyon örnekleri ve gerçek klinik verilerle birlikte kamuya açık verilerden değerlendirilmiştir. Ek olarak, Copy-fm'in bölgesel olarak f_m hesaplamasının yanında, daha önce denenmemiş bir yaklaşım olan "*f_m-first*" olarak adlandırdığımız genom boyu tarama algoritmasıyla, aracın f_m sonucunu da vererek genom boyu mCNV saptama yeteneği gösterilmiştir.

Bu tez TÜBİTAK destekli EJP-RD projesi RiboEurope (319S062) ve Hacettepe Üniversitesi BAP Koordinasyon Birimi (THD-2021-19532) tarafından desteklenmiştir.

Anahtar kelimeler: Kopya sayısı değişiklikleri, mCNV, Mozaizizm, f_m , SNP Mikrodizin, R programlama dili

ABSTRACT

Acun, H. M., Determination of Mosaicism Ratio in Copy Number Variations Using Microarray Based Genotype Information, Hacettepe University Graduate School of Health Sciences Department of Bioinformatics Master's Thesis, Ankara, 2023. Copy number variations (CNVs) are >50 bp structural chromosomal variants that represent a regional change in the normal diploid (2 copies) copy number (CN) of genomic loci. This definition dictates that all copy numbers in a single cell must be integers, however, for a cell population consisting of various subpopulations with different copy numbers, when the whole cell population is considered there may be non-integer copy numbers. CNVs that cause such copy numbers to have been defined as mosaic CNVs (mCNVs). mCNVs can be encountered at birth or later in life because of conditions such as cancer. Since the detection of mCNVs is critical in diagnosis and treatment monitoring, especially in cancer, many researchers and commercial companies have developed methods for the detection of mCNVs. Among these methods, the SNP microarray method is the most commonly available as it allows genome-wide screening for mCNVs of different sizes. However, despite the rapid development in methods for localization of mCNVs, the number of competent and specific studies for determination of cells with CNV in a population of cells, fraction of mosaicism (f_m), is very limited. Within this thesis, a bioinformatics tool which was named Copy-fm has been developed in R programming language that utilizes BAF and LRR values obtained from SNP microarray data to calculate ratio for f_m by Kolmogorov-Smirnov (KS) test. This allowed us to fulfil the much-neglected need for f_m calculation which may be required in research and clinics. Testing of Copy-fm's reliability and functionality has been achieved by using experimentally-designed mosaic deletion samples along with publicly available and real clinical data. In addition, Copy-fm has been tested on two main microarray platforms, revealing the differences between microarray platforms. In addition, besides the regional f_m calculation, Copy-fm's ability to detect genome-wide mCNVs has been demonstrated by an approach we named "f_m-first" algorithm, which has never been tested before.

This thesis has been supported by TÜBİTAK-supported EJP-RD project RiboEurope (319S062) and Hacettepe University Research Projects Unit (THD-2021-19532).

Keywords: Copy Number Variations, mCNVs, Mosaicism, f_m , SNP Microarray, R programming language

İÇİNDEKİLER

ONAY SAYFASI	iii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYANI	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER VE KISALTMALAR	xiii
ŞEKİLLER	xvii
TABLolar	xix
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. İnsan Genomu ve Genomdaki Varyasyonlar	3
2.1.1. İnsan Genomu	3
2.1.2. Genomik Varyasyonlar	4
2.1.3. Büyük Ölçekteki Varyantlar	6
2.1.4. Küçük Ölçekteki Varyantlar	7
2.1.5. Orta Ölçekteki Varyantlar	9
2.2. Yapısal Varyantlar	10
2.2.1. Kopya Sayısı Değişiklikleri	10
2.2.2. Kopya Nötral Değişiklikler	12
2.2.3. Kopya Sayısı Değişikliklerinin Oluşum Mekanizmaları	13
2.2.4. CNV Saptama Yöntemleri	18
2.3. Heterozigotluk Kaybı (LOH)	31

2.4. Mozaisizm	32
2.4.1. Mozaisizmin Oluşum Mekanizmaları	38
2.5. Uniparental Dizomi (UPD)	41
2.6. Mozaik CNV'lerin İncelenmesi Amacıyla Kullanılan Biyoinformatik Araçlar	45
2.6.1. Mikrodizin Verisi Kullanan Mozaik CNV Araçları	46
2.6.2. NGS Verisi Kullanan Mozaik CNV Araçları	49
2.6.3. Mozaik CNV'lerin İncelenmesinde Model-Örnek Karşılaştırmalarının Kullanılması	50
2.7. R Programlama Dili	51
3. BİREYLER VE YÖNTEM	54
3.1. Bireyler	54
3.2. Çalışmada Kullanılan Deneysel Yöntemler	55
3.2.1. FISH	55
3.2.2. DNA Örneklerinin Elde Edilmesi ve Deneysel Mozaik Delesyon Örneklerinin Oluşturulması	57
3.2.3. SNP Mikrodizin Yöntemi	59
3.3. Copy-fm Algoritmasının Oluşturulması	63
3.3.1. Farklı Mikrodizin Platformlarından Elde Edilen SNP Mikrodizin Verilerinin Ön Veri İşleme ile Analize Hazırlanması	67
3.3.2. Kontrol ve Test Örneklerinde Analiz Edilebilir Varyantların Seçimi	68
3.3.3. Kontrollerden Mozaisizm Modellerinin Oluşturulması	72
3.3.4. fm Değerinin Hesaplanması	75
3.3.5. Mozaisizm Fraksiyonu Güven Aralıklarının Belirlenmesi	76
3.3.6. Copy-fm preprocess Fonksiyonunun Kullanımı	76
3.4. Copy-fm'in İşlevlerinin Deneysel Veride Değerlendirilmesi	76
3.4.1. Farklı Deneysel fm Değerlerinde Copy-fm'in Etkinliğinin Test Edilmesi	77

3.4.2. Affymetrix ve Illumina platformlarından elde edilen verilerin kıyaslanması	77
3.4.3. Copy-fm'in mozaik CNV bölgelerini saptama yeteneğinin denenmesi	77
3.4.4. Copy-fm'in Performansının Diğer Biyoinformatik Araçlarla Kıyaslanması	78
4. BULGULAR	79
4.1. Deneysel Mozaik Delesyon Örneklerinin Elde Edilmesi ve İncelenmesi	79
4.1.1. Test ve Kontrol Örneklerinde X Kromozomu Mozaisizminin FISH ile Değerlendirilmesi	79
4.1.2. Test ve kontrol örneklerinde SNP mikrodizin verilerinin mikrodizin platformunda incelenmesi	81
4.2. Farklı Mikrodizin Platformlarından Elde Edilen LRR ve BAF Verilerinin Analize Hazırlanması	82
4.3. Copy-fm fm fonksiyonu İş Akış Şeması	84
4.4. Copy-fm Algoritmasında Kontrol ve Test Örneklerinin Analiz Edilebilirliğinin İncelenmesi	85
4.5. Gerçek mikrodizin verilerinde Copy-fm'in test edilmesi	88
4.5.1. Kamuya Açık Verilerde Copy-fm'in Test Edilmesi	88
4.5.2. Gerçek Klinik Örneklerle Ait Mikrodizin Verilerinde Copy-fm'in Test Edilmesi	89
4.5.3. Deneysel Mozaik Delesyon Örneklerinde Copy-fm'in Test Edilmesi	93
4.6. Copy-fm'in Mozaik CNV Bölgelerini Saptama Yeteneğinin Denenmesi	97
4.7. Copy-fm'in Diğer Mozaik CNV Saptayan Yazılımlarla Kıyaslanması	100
5. TARTIŞMA	102
6. SONUÇ VE ÖNERİLER	109
7. KAYNAKLAR	110
8. EKLER	

EK-1: Tez çalışmasının etik kurul izni

EK-2: Tez çalışması orijinallik raporu

EK-3: Aydınlatılmış onam formu

EK-4: ChAS'ın Analysis Workflow modülü ile Affymetrix SNP mikrodizin .CEL dosyalarının analizi

EK-5: Affymetrix APT (Analysis Power Tools) ile .cychp dosyalarının .txt dosyalarına dönüştürülmesini sağlayan R kodu

EK-6: Illumina GenomeStudio ve Beeline ile SNP mikrodizin analizi basamakları

EK-7: Copy-fm (*scanfm* fonksiyonu) ile genom boyu tarama

EK-8: f_m hesaplaması için kullanılan diğer programların komutları

9. ÖZGEÇMİŞ

SİMGELER VE KISALTMALAR

A	Adenin
aCGH	Mikrodizin Karşılaştırmalı Genomik Hibridizasyon (<i>array Comperative Genomic Hybridization</i>)
ADTEX	Tümör Ekzomunda Anormallik Tespiti (<i>Aberration Detection in Tumour Exome</i>)
affy2sv	Affymetrix'den Yapısal Varyanta (<i>Affymetrix to Structural Variation</i>)
Alt-EJ	Alternatif Uç Birleştirme (<i>Alternative End-Joining</i>)
APT	Analiz Güç Araçları (<i>Analysis Power Tools, Affymetrix</i>)
AS	İnşa (<i>Assembly</i>)
ASCAT	Tümörlerde Allel Spesifik Kopya Sayısı Analizi (<i>Allelic Specific Copy Number Analysis in Tumors</i>)
ascatNGS	İleri Nesil Dizileme ile Tümörlerde Allel Spesifik Kopya Sayısı Analizi (<i>Allelic Specific Copy Number Analysis in Tumors by Next Generation Sequencing</i>)
BAF	B Allel Frekansı (<i>B Allele Frequency</i>)
BAFsegmentation	B Allel Frekanslarını Ayırma (<i>B Allele Frequency Segmentation</i>)
BeadChip	Mıknatıslı Çip
bp	Baz Çifti (<i>Base Pair</i>)
bpm	Mıknatıs Havuz Manifest (<i>Bead Pool Manifest</i>)
BS	Tekli Ayırma (<i>Binary Segmentation</i>)
C	Sitozin (<i>Cytosine</i>)
CBS	Diresel Tekli Ayırma (<i>Circular Binary Segmentation</i>)
CC	Yaratıcı Ortaklar (<i>Creative Commons</i>)
CDF	Kümülatif Dağılım Fonksiyonu (<i>Cumulative Distribuiton Function</i>)
ChAS	Kromozom Analiz Takımı (<i>Chromosome Analysis Suite</i>)
Chr	Kromozom (<i>Chromosome</i>)
CN	Kopya Sayısı (<i>Copy Number</i>)
CN-LOH	Kopya Nötral Heterozigotluk Kaybı (<i>Copy Neural Loss of Heterozygosity</i>)
CNV	Kopya Sayısı Değişikliği (<i>Copy Number Variation</i>)
conf	Güven (<i>Confidence</i>)
control-FREEC	İleri Nesil Dizileme Kontrol Bağımsız Alelik Kopya Sayısı Değişiklikleri (<i>Control free assessing of copy number and allelic content using next generation sequencing</i>)
Copy-fm	Kopya Sayısı-mozaisizm fraksiyonu (<i>Copy number mosaicism fraction</i>)
CRAN	Kapsamlı R Arşivi Ağı (The Comprehensive R Archive Network)
csv	Virgülle Ayrılmış Değer (Comma Separated Value)
CytoScan	Sitogenetik Tarama (<i>Cytogenetics Scanning</i>)
CytoSNP	Sitogenetik Tek Nükleotid Polimorfizmi (<i>Cytogenetics Single Nucleotide Polymorphism</i>)

DANFIP	Entegre Edilmiş Olasılıkların Uygunluğundan Dağılım Analizi (<i>Distribution Analysis by Fitting Integrated Probabilities</i>)
dgof	Kesikli Sayıların En iyi Uyum Testleri (<i>Discrete Goodness-of-Fit Tests</i>)
DNA	DeoksiriboNükleik Asit
EDTA	Etilendiamin Tetraasetik Asit
EuroDBA	Avrupa Diamond-Blackfan Anemisi Konsorsiyumu
FDR	Yanlış Bulgu Oranı (False Discovery Rate)
FISH	Florasın In Situ Hibridizasyon (<i>Florescent In Situ Hybridization</i>)
f_m	Mozaisizm Fraksiyonu (<i>Fraction of Mosaicism</i>)
fm-first	Önce Mozaisizm Fraksiyonu (<i>fraction of Mosaicism-first</i>)
FoSTeS	Çatalı Durdurarak Kalıp Değişirme (<i>Fork Stalling and Template Switching</i>)
G	Guanin
GA	Güven Aralığı
GADA	Genetik Değişiklik Saptama Algoritması (<i>Genetic Alteration Detection Algoritm</i>)
GATK	Genom Analiz Araç Kiti (<i>Genome Analysis Toolkit</i>)
Gbp	Giga baz çifti (<i>Giga base pair</i>)
GHz	Giga Hertz
gtc	Genotip Çağrı (<i>Genotype Call</i>)
het	Heterozigot
het_int	Heterozigot Sınır (<i>Heterozygous Interval</i> , Copy-fm'in fm fonksiyonu parametresi)
hg	İnsan Genomu (<i>Human Genome</i>)
HR	Homolog Rekombinasyon (<i>Homologous Recombination</i>)
HTML	Hiper Metin İşaretleme Dili (Hypertext Markup Language)
IBD	Kalıtım Yoluyla Özdeşlik (<i>Identity By Decent</i>)
ID	İdeal Dizomi (<i>Ideal Disomy</i>)
idat	Sinyal Verisi (<i>Intensity Data</i>)
ID-CDF	İdeal Dizomi-Kümülatif Dağılım Fonksiyonu (<i>Ideal Disomy-Cumulative Distribution Function</i>)
Kbp	Kilobaz Çifti (<i>Kilobase Pair</i>)
KCl	Potasyum Klorür
KS	Kolmogorov-Smirnov
LCR	Düşük Kopyalı Tekrar (<i>Low Copy Repeat</i>)
LINE1 ya da L1	Uzun Serpiştirilmiş Nükleer Eleman 1 (<i>Long Interspersed Nuclear Element-1</i>)
LOH	Heterozigotluk Kaybı (<i>Loss of Heterozygosity</i>)
LRR ya da log₂R	Logaritma 2 Tabanında R Oranı (<i>Logarithmic R Ratio</i>)
MAD	Mozaik Değişiklik Saptama (<i>Mosaicism Alteration Detection</i>)
MATLAB	Matriks Laboratuvarı (<i>Matrix Laboratory</i>)
Mbp	Megabaz Çifti (<i>Megabase Pair</i>)
mCNV	Mozaik Kopya Sayısı Değişikliği (<i>Mosaic Copy Number Variation</i>)

min(D)	Minimum D Değeri
min_het	Minimum Heterozigot Sayısı (<i>Minimum heterozygous</i>)
MLPA	Multipleks Ligasyon Bağımlı Prob Amplifikasyonu (<i>Multiplex Ligation Dependent Probe Amplification</i>)
MoChA	Mozaik Kromozomal Değişiklikler (<i>Mosaic Chromosomal Alterations</i>)
MONTAGE	Kolayca Mozaik Sayı Etiketleme (<i>Mosaic Number Tag Easily</i>)
mosaicism_dist	Mozaisizm Dağılımı (<i>Mosaicism Distribution</i>)
MSR	Minimum Rezidüeller Toplamı (<i>Minimum Sum of Residuals</i>)
NAHR	Allelik Olmayan Homolog Rekombinasyon (<i>Non-Allelic Homologous Recombination</i>)
NCBI	Ulusal Biyoteknoloji Bilgi Merkezi (<i>National Center for Biotechnology Information</i>)
NextSeq	İleri Nesil Dizileme Sekans Cihazı (<i>Next Generation Sequencing Device</i>)
NGS	İleri Nesil Dizileme (<i>Next Generation Sequencing</i>)
NHEJ	Homolog Olmayan Uç Birleştirme (<i>Non-Homologous End-Joining</i>)
n_loop	n sayıda iterasyon (<i>N count looping</i>)
Nsp	Nişasta Olmayan Polisakkaridler (<i>Non Starch Polysaccharides</i>)
OncoSNP	Onkolojik Tek Nükleotit Polimorfizmi (<i>Oncological Single Nucleotide Polymorphism</i>)
OncoSNP-SEQ	Onkolojik Tek Nükleotit Polimorfizmi – Dizi (<i>Oncological Single Nucleotide Polymorphism – Sequence</i>)
ORF	Açık Okuma Çerçevesi (<i>Open Reading Frame</i>)
PacBio	Pasifik Biyobilimler (<i>Pacific Biosciences</i>)
PAR	Psödootozomal Bölge (<i>Pseudoautosomal Region</i>)
PCR	Polimeraz Zincir Reaksiyonu (<i>Polymerase Chain Reaction</i>)
PMP22	Periferik Miyelin Protein 22 (<i>Peripheral Myelin Protein 22</i>)
pval	p Değeri (<i>p value</i>)
RAM	Rastgele Erişim Belleği (<i>Random Access Memory</i>)
RD	Okuma Derinliği (<i>Read Depth</i>)
RNA	Ribonükleik Asit
RP	Okuma Çifti (<i>Read Pair</i>)
RSS	Rezidüellerin Kareler Toplamı (<i>Residual Sum of Squares</i>)
scanfm	Mozaisizm Tarama (<i>Scanning fraction of mosaicism</i>)
SNP	Tek Nükleotid Polimorfizmi (<i>Single Nucleotide Polymorphism</i>)
SNV	Tek Nükleotid Değişikliği (<i>Single Nucleotide Variation</i>)
SMRTSeq	Tek Molekülün Gerçek Zamanlı Dizilemesi (<i>Single Molecule Real Time Sequencing</i>)
SR	Ayrık Okuma (<i>Split Read</i>)
STR	Kısa Ardışık Tekrar (<i>Short Tandem Repeat</i>)
T	Timin
T2T	Telomerden Telomere (<i>Telomere to Telomere</i>)
Taq	Thermus aquaticus

TdT	Terminal deoksinükleotidil Transferaz
TE	Tris-EDTA
triPOD	Anne-baba-çocuk üçlüsü ile ebeveyn orjini (<i>Trio Analysis with Parent of Origin</i>)
txt	Yazı (<i>Text</i>)
UPD	Uniparental Dizomi (<i>Uniparental Disomy</i>)
VarScan2	Varyant Tarama 2 (<i>Variant Scanning 2</i>)
vcf	Varyant Çağrı Formatı (<i>Variant Call Format</i>)
VNTR	Değişken Sayılı Ardışık Tekrarlar (<i>Variable Number of Tandem Repeats</i>)
WES	Tüm Ekzom Dizileme (<i>Whole Exome Sequencing</i>)
WGS	Tüm Genom Dizileme (<i>Whole Genome Sequencing</i>)
ZMW	Sıfır Modlu Dalga Kılavuzu (<i>Zero Mode Waveguide</i>)

ŞEKİLLER

Şekil	Sayfa	
2.1.	Küçük ölçekli varyant tipleri.	8
2.2.	Yapısal Varyantlar.	10
2.3.	Kopya sayısı değişiklikleri.	11
2.4.	CNV'lerin fenotipik etkileri.	12
2.5.	NAHR mekanizması.	14
2.6.	NHEJ mekanizması.	15
2.7.	FoSTeS mekanizması.	16
2.8.	L1 retrotranspozisyonu mekanizması.	18
2.9.	FISH görüntüleri.	19
2.10.	aCGH yöntemi.	21
2.11.	Bir kromozom boyunca Affymetrix SNP mikrodizin verisine ait LRR ve BAF grafikleri.	23
2.12.	Illumina mikrodizin verisinde R ve θ değerleri.	23
2.13.	MLPA yönteminin analiz çıktısı.	25
2.14.	NGS ile yapısal varyant saptamak için kullanılan yöntemler.	27
2.15.	Üçüncü nesil dizileme yaklaşımları.	30
2.16.	Optik genom haritalama yaklaşımı.	31
2.17.	Affymetrix SNP mikrodizin verisinde LOH.	32
2.18.	Mozaisizm fraksiyonu bilinen duplikasyon mozaisizminin kopya sayısının hesaplanması.	34
2.19.	Mozaisizm fraksiyonu bilinen delesyon mozaisizminin kopya sayısının hesaplanması.	34
2.20.	Mozaik CNV'lerin SNP mikrodizin verisinde görünümü.	36
2.21.	Mayoz I ve Mayoz II aşamalarında ayrılmama.	40
2.22.	Mitozda ayrılmama.	41
2.23.	Uniparental dizomi (UPD).	42
2.24.	Farklı UPD tipleri.	43
2.25.	Trizomiden kurtulma ile oluşan segmental izodizomi/heterodizomi.	44
2.26.	SNP mikrodizin verisinden elde edilen mozaik UPD görüntüsü.	45
3.1.	Affymetrix CytoScan Optima mikrodizin platformunun deneysel uygulama basamakları.	60
3.2.	Copy-fm'de kontrol ve test örneklerinin analize uygunluğunun değerlendirilmesi için geliştirilen iş akışı.	69
4.1.	Kontrol ve test örneklerinde cinsiyet kromozomlarına yönelik yapılan FISH analizleri.	81
4.2.	Kontrol ve test örneklerinde X kromozomuna ait Affymetrix ve Illumina mikrodizin görüntüleri.	82
4.3.	Copy-fm'de Affymetrix verileri için ön veri işleme iş akışı.	83
4.4.	Copy-fm'de Illumina verileri için ön veri işleme iş akışı.	84
4.5.	Copy-fm iş akışı.	85

4.6.	X kromozomunda LOH bölgesi (chrX: 9,620,214-11,377,010) bulunan K5 kontrol örneği analizden çıkarılmadığında elde edilen sonuçlar.	86
4.7.	Kamuya açık verilerde Copy-fm'in denenmesi.	89
4.8.	Klonal mozaik delesyonlar içeren mikrodizin verisinde Copy-fm'in etkinliği.	90
4.9.	Farklı senaryolarda mozaik duplikasyon içeren mikrodizin verisinde Copy-fm'in etkinliği.	91
4.10.	"4'lü BAF şeması" ve "6'lı BAF şeması" şeklinde görünen farklı duplikasyonların Copy-fm tarafından ayırt edilmesi.	92
4.11.	Affymetrix platformundan TO örneği için f_m hesaplaması sırasında kullanılan RSS ve KS yöntemlerinin karşılaştırılması.	95
4.12.	T80 örneği için Affymetrix ve Illumina mikrodizin verilerinde "BAF ters çevirme" sonrası varyant BAF değerlerinin dağılımı.	97
4.13.	Copy-fm ile 5. kromozom boyunca mozaik CNV'lerin taranması.	98
4.14.	Copy-fm içinde yer alan <i>scanfm</i> fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyde 5. ve 12. kromozomun görünümü.	99
8.1.	ChAS'ın <i>Analysis Workflow</i> modülü ile Affymetrix SNP mikrodizin .CEL dosyalarının .cychp dosyalarına dönüştürülmesi.	128
8.2.	Illumina GenomStudio'da proje oluşturulması.	130
8.3.	Illumina GenomeStudio'dan FinalReport.txt alınması.	131
8.4.	Beeline yazılımı ile .idat dosyalarının .gvc dosyalarına dönüştürülmesi.	132
8.5.	Copy-fm içinde yer alan <i>scanfm</i> fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyin (Affymetrix) tüm genom boyu taranması.	133

TABLOLAR

Tablo	Sayfa
2.1. Farklı CNV senaryolarına göre beklenen genotip, kopya sayısı ve BAF değerleri.	24
2.2. Affymetrix ve Illumina mikrodizin üreticilerine ait farklı SNP mikrodizin çiplerinin özellikleri.	24
2.3. Mozaik CNV'lerin incelenmesi amacıyla kullanılan biyoinformatik araçlar.	46
3.1. Anne-oğul DNA örneklerinin farklı mozaik delesyon modeli örneklerinin oluşturulması için yapılan karışımlar.	59
3.2. Heterozigot varyantlara ait BAF değerlerinin f_m değerine göre beklenen ortalama değerinin hesaplanması.	73
4.1. Tez örneklemindeki 7 bireyin cinsiyet kromozomlarına yönelik FISH sonuçları ve elde edilen DNA konsantrasyonları.	80
4.2. Kontrollerdeki (K1-5) homozigosite bölgelerinin analizden çıkarılması ve test örneğinde (T0) homozigosite bölgelerinin değerlendirilmesi.	87
4.3. Copy-fm'in deneysel olarak oluşturulan delesyon mozaizizmlerine ait mikrodizin verileri ile f_m hesabı için KS yöntemi kullanılarak test edilmesi.	93
4.4. Copy-fm'in deneysel olarak oluşturulan delesyon mozaizizmlerine ait mikrodizin verileri ile f_m hesabı için RSS yöntemi kullanılarak test edilmesi.	94
4.5. Yüksek oranlı mozaik delesyon örneklerine ait verilerde "BAF ters çevirme" yaklaşımıyla elde edilen sonuçlar.	96
4.6. Bir CNV bulunmayan T0 örneğine ait Affymetrix ve Illumina verilerinde Copy-fm'in 5. kromozom taraması için geçen süreler.	98
4.7. Copy-fm ile MoChA araçlarının f_m hesaplamalarının karşılaştırılması.	100
4.8. Copy-fm'in benzer biyoinformatik araçlar ile karşılaştırılması.	101

1. GİRİŞ

İnsan, diploid genoma sahip bir organizmadır. Dolayısıyla kadınlarda her bir genomik bölge için kopya sayısı 2 iken, erkeklerde otozomlar için 2, X ve Y şeklinde bulunan farklı cinsiyet kromozomları için ise birer kopyadır. İnsanlarda herhangi bir genomik bölgedeki kopya sayısının beklenen normal kopya sayısı olan 2'den farklı olması durumuna kopya sayısı değişiklikleri (CNV, *Copy Number Variation*) adı verilir. Gerek kalıtsal gerek edinsel olarak ortaya çıkabilen CNV'ler, kopya sayısının azalması (delesyon) veya artması (duplikasyon) şeklinde olabilir. CNV'ler için farklı boyut sınırları kullanılmakla beraber günümüzde >50 baz çifti (bp, *base pair*) uzunluğundaki yapısal varyasyonlar olarak tanımlanmaktadır (1). Tespitinde sıkça kullanılan yöntemlerden olan mikrodizin ve ileri nesil dizileme (NGS, *Next Generation Sequencing*) teknolojileri sayesinde pek çok kalıtsal hastalıkla ve başta kanser olmak üzere çeşitli edinsel genetik hastalıklarla ilişkilendirilmişlerdir (2).

Çok hücreli bir organizmadaki hücre popülasyonunun sadece bir kısmının mutasyonlardan etkilenerek farklılaşmasına genetik mozaizm denilmektedir. Genetik mozaizm farklı genetik ölçeklerde olabilmekte ve CNV ölçeğindeki mozaizm olayları, mozaik CNV (mCNV) olarak adlandırılmaktadır (3). Bu tanım itibariyle; mozaik CNV'ler, hücrelerin belli bir kesrinde kopya sayısının 2'den farklılaşmasına yol açmaktadırlar ve bir hücre içerisinde tam sayı olarak ifade edilebilecek genomik kopya sayılarının tüm hücre popülasyonu için kesirli bir sayı olmasına neden olmaktadır. Bu nedenle, bir mozaik CNV bölgesindeki kopya sayısı, değişikliğe uğramış hücre oranını belirten mozaizm fraksiyonunu (f_m) yansıtmaktadır.

Mozaik CNV'ler canlıyı oluşturan ilk hücre olan zigot evresinden sonra ortaya çıktıkları için mutlaka edinsel genetik değişiklikler olmaktadır. Bu değişiklikler, embriyonik gelişim döneminde ortaya çıkabilseler de en sık kanserli hücreler ile normal hücrelerin bir karışım halinde bulunduğu kanser ilişkili dokularda ortaya çıkmaktadırlar (4). Bu dokulardaki mozaizm fraksiyonu; hastalık şiddeti, hastalığın

seyri, kansere sahip bireylerin takibi ve kanserin ilaç yanıtı konularında bilgiler verebildiği için özellikle önem kazanmaktadır.

Mozaik CNV saptamaya yönelik günümüzde farklı yöntemler kullanılmakla beraber SNP (*Single Nucleotide Polymorphism*, Tek Nükleotit Polimorfizmi) mikrodizin yöntemi, uzun yıllardır sitogenetik yöntemlerle saptanamayacak kadar küçük CNV'leri güvenilir biçimde genom boyu taramayı mümkün kılmıştır (5). SNP mikrodizin yöntemlerinin, mozaik CNV'lerin genomik lokalizasyonunu saptayabilmesi daha zor olmakla beraber bunu gerçekleştiren yaklaşımlar da yıllar içerisinde geliştirilmiştir (6). Buna karşın, CNV'ler için mozaik fraksiyonunu hesaplamaya yönelik literatürde çok az yaklaşım bulunmakta ve var olan yöntemlerin hiçbirisi saptadıkları mozaik fraksiyonu için güven aralığı sunmamaktadırlar.

Bu tez çalışması kapsamında, Kolmogorov-Smirnov testini kullanan yeni bir yaklaşım ile SNP mikrodizin verisinin analizi sonucunda CNV'ler için güven aralığıyla beraber mozaik fraksiyonu hesaplayabilen R programlama dilinde yazılmış bir biyoinformatik araç sunulması hedeflenmiştir. Sunulan aracın etkinliği, farklı SNP mikrodizin platformlarında deneysel olarak oluşturulmuş X kromozomu delesyon mozaik fraksiyonu modellerinde ve mozaik CNV'ye sahip bireylerden elde edilmiş verilerde test edilmektedir.

2. GENEL BİLGİLER

2.1. İnsan Genomu ve Genomdaki Varyasyonlar

2.1.1. İnsan Genomu

İnsan genomu, kromozomlar şeklinde paketlenmiş olan 3,2 milyar baz çifti (*Gbp, giga base pair*) uzunluğundaki DNA'yı (deoksiribonükleik asit) içermektedir. Genomun haploid dizi uzunluğu olan 3,2 Gbp eşey hücrelerinde (gamet) yer alırken, 6,4 Gbp olan diploid genom ise gametler haricindeki vücut hücrelerinde (somatik) yer almaktadır (7).

İnsan genomunu barındıran kromozom sayısı 46 olup, otozomlar (cinsiyetler arası farklılık göstermeyen kromozomlar) ve cinsiyet kromozomları olarak 2 grup kromozom bulunmaktadır. Otozomlarda 22 çift kromozom; cinsiyet kromozomlarında ise X kromozomu ortak olmak üzere, kadınlarda (XX) 1 çift X ve erkeklerde (XY) 1 X ile 1 Y kromozomu bulunmaktadır. Buna göre, 23 çift kromozom taşıyan insan genomu somatik hücreler içerisinde diploid (2 kopya) şekilde bulunmaktadır. Gamet hücrelerinde ise kromozomlar haploid (n) olmakta ve tek bir set 23 kromozomdan oluşmaktadır (7). Kromozom çiftleri, mayoz bölünme sırasında 2 ebeveynden yavruya birer kopya olarak aktarılan ve birbirine çok büyük benzerlik gösteren homolog kromozomları ifade etmektedir. Cinsiyeti belirleyen X ve Y kromozomları arasındaki homoloji ise özel bir durum arz etmekte olup bu kromozomlar boyunca tam bir homoloji olmadan, otozomlar gibi yüksek homoloji özelliği gösteren ve daha kısa olan PAR (Psödootozomal Bölgeler, *Pseudo-Autosomal Regions*) adlı bölgeler cinsiyet kromozomlarının birbirinin homoloğu gibi davranmasını sağlamaktadır (8). X ve Y kromozomlarının uçlarında yer alan PAR'lardan kısa kolların (p kolu) uçlarında yer alan PAR1 (2.6 Mbp, Mega baz çifti), uzun kolların (q kolu) uçlarında yer alan ise PAR2 (320 kbp, kilo baz çifti) olarak adlandırılmaktadır.

İnsan genomunun içeriğini, evrimini ve hastalıkların kökenini aydınlatmak üzere 1990 yılında başlatılıp 2003 yılında DNA haploid dizisinde bulunan yaklaşık 3 milyar bazın dizilenmesiyle tamamlanmış İnsan Genom Projesi neticesinde, nükleotit

diziliminde meydana gelen deęişikliklerin anlaşılması, etkileri ve oluşum mekanizmaları gibi yıllardır bilim dünyasının cevap bulmaya çalıştığı konuların araştırılması hız kazanmıştır (9). İnsan Genom Projesinin sunduęu en büyük inovasyonlardan biri, haploid genom dizisinin ilk kez *NCBI Build 34* adıyla (hg16) dijital olarak arşivlenerek tüm dünyayla paylaşılmasıyla herkesin kaynak alabileceęi ve büyük çoęunluęunu içeren insan referans genomunun ortaya çıkışı olmuştur. Dizileme teknolojisindeki ilerlemeyle birlikte ise, 2004 yılında hg17, 2006'da hg18 ve nihayetinde günümüzde en sık kullanılan iki referans olan hg19 ile hg38 sırasıyla 2009 ve 2013 yıllarında oluşturulmuştur (10). 2022 yılında ise Genom Referans Konsorsiyumu'nun bir kolu olan T2T (*Telomere-to-Telomere*) Konsorsiyumu'nun tamamladığı T2T-CHM13 insan genomunun telomer dizilerini de içeren boşlukların daha minimal olduęu tümüyle dizilenmiş insan genomuna ait bir versiyonu doğmuştur (11).

Referans genomun erişilebilir ve yaygın kullanımda oluşu, mikrodizin ve dizileme gibi teknolojilerin de kısa sürede referans genom kullanımını entegre etmesine ve boyutları deęişkenlik gösteren, bu sebeple tarihsel süreç boyunca saptanmalarında zorluk yaşanarak iyi aydınlatılamamış genomik varyasyonlara karşın adeta yeni bir devrin başlamasına sebep olmuştur.

2.1.2. Genomik Varyasyonlar

Canlıların tümünde, evrimsel sürecin bir parçası olarak gerek çevresel gerek genetik birtakım mekanizmalarla tür içi ve türler arası çeşitlilik görülmektedir. Canlılar arası bu çeşitlilięe sunulan katkıların içerisinde genetik varyantlar önemli rol oynamaktadır. Bir canlıda o canlıya ait referans genom dizisinden farklılık gösteren her bir deęişikliğe genetik varyant denilmektedir. İnsanlarda bireyler arasındaki genomik farklılıklar görece az olup, ortalama olarak bir bireyde genomik dizinin %99,9'u referans genomla örtüşmektedir (12). Genetik varyantların oluşturduęu farklılıklar, bireyler arasında fenotipik farklılıklara sebep olmaktadır. Buna karşın, genetik varyantların birçoęunun hiçbir fenotipik yansıması olmazken, fenotipik farklılığa yol açan varyantlar bazen hafif bazen çok belirgin farklılıklar ortaya

çıkarmaktadır. Özellikle belirgin olan fenotipik farklılıklardan bazıları genetik hastalıklar şeklinde karşımıza çıkabilmektedir. İnsan genetiği alanındaki çalışmaların pek çoğunda hastalık sebebi olabilecek bu varyantlar üzerine odaklanılmakta ve bu varyantlar, patojenik varyant veya mutasyon olarak isimlendirilmektedir (13).

Patojenik varyantlar sonucu ortaya çıkan genetik hastalıklar etkilenen gen sayısına göre gruplanabilmektedir. Buna göre, tek bir gendeki mutasyonlar sonucu ortaya çıkan hastalıklar tek gen hastalıkları olarak isimlendirilmektedir (14). Mendelyan kalıtılan hastalıkların neredeyse tamamı bir tek gen hastalığı olarak ortaya çıkmaktadır. Etkilenen gen sayısına göre bir diğer sınıf ise, genomik lokasyon olarak birbirine yakın birden fazla genin aynı anda etkilenmesi (mutasyona uğraması) sonucu ortaya çıkan kromozomal hastalıklardır. Bu hastalıklarda bir kromozomun tamamı veya bir kısmındaki değişiklik sonucu birden fazla genin etkilenmesi ile birden fazla fenotipik özelliğin bir arada bulunduğu sendrom adını alan hastalıklar ortaya çıkmaktadır. Son olarak, genomik lokasyon olarak birbiri ile ilişkisiz olabilen birden fazla genin etkilenmesi ile ortaya çıkabilen multifaktöriyel (kompleks) hastalıklar bulunmaktadır (15). Bu tür hastalıklarda her bir genetik varyantın hastalık riskine katkısı bulunmakta, çevresel etkenlerin de katkısı sonucu hastalık ortaya çıkmaktadır. Toplumda görece yaygın görülen hipertansiyon, obezite ve yarı damak gibi hastalıklar multifaktöriyel hastalıklara örnektir.

Hastalıkla ilişkili veya ilişkisiz olan tüm genomik varyantlar, tek bir nükleotidin etkilendiği nokta mutasyonlarından tüm bir kromozomu etkileyen mutasyonlara kadar geniş bir yelpazede karşımıza çıkmaktadır (16). Varyantların uzunluğuna göre yapılan bu sınıflama bir varyantın hangi laboratuvar yöntemi ile saptanabileceği ile yakın bir ilişki içerisindedir. Buna göre, varyantlar tarihsel olarak fark edilme sıralarına göre büyük ölçekli, küçük ölçekli ve orta ölçekli varyantlar olarak sınıflandırılmış ve ilerleyen bölümlerde bu varyantlar daha detaylı olarak incelenmiştir.

2.1.3. Büyük Ölçekteki Varyantlar

Down Sendromu'nun genetik sebebinin Trizomi 21 olduğunun saptanabilmesinin ardından genetik yapıdaki değişiklikler ile bazı hastalıkların ilişkili olduğunun anlaşılmasıyla beraber 1960'lı yılların başından itibaren hastalıkların sebebini araştırmak amacıyla incelenmeye başlayan genetik varyasyonlar, başlarda sadece mikroskopla gözlenebilen kromozomların incelenmesi ile sınırlıydı (17). Sitogenetik adı verilen bu inceleme yöntemleri sayesinde kromozomlardaki sayısal ve yapısal bazı değişiklikler keşfedilmeye başlanmıştı (18).

Ancak, kromozom düzeyindeki gözlemler insan genomunda sadece >3 Mbp uzunluğundaki değişiklikleri saptamaya elverişliydi. Bu sebeple, kromozomal varyantların mikroskopla gözlenmesi mümkün olduğu için bu grup varyantlar mikroskobik varyantlar veya büyük ölçekli varyantlar olarak sınıflandırılmaktadırlar (19). Büyük ölçekteki varyasyonlar büyük kromozomal bölgelerde hasar meydana getirdiklerinden birden fazla geni ilgilendiren hastalıklarla ilişkilendirilmektedirler (20).

İnsanlarda normal kromozom sayısı olan 46'dan farklı sayıda kromozom bulunması sayısal kromozom anomalilerini oluşturmaktadır. Normalde her kromozomdan 2 çift bulunması dizomi olarak adlandırılırken; bir kromozomun kaybı monozomi, bir kromozomun ekstra bir kopyası olması durumu ise trizomi olarak adlandırılmaktadır (21). En sık görülen sayısal kromozomal anomalilerden olan Trizomi 21 ise, Down Sendromu'na neden olmaktadır (22). Sayısal anomalilere ek olarak kromozomların içerisinde de bazı değişiklikler meydana gelebilmektedir. Yapısal kromozom anomalileri olarak adlandırılan bu değişiklikler; delesyon, duplikasyon, inversiyon, insersiyon, translokasyon gibi değişikliklerdir. Yapısal kromozom anomalilerinden kromozom analizi ile görülebilenler genomik lokasyonuna göre en küçük 3-10 Mbp büyüklüğünde olabilirken daha küçük olan (submikroskobik) yapısal varyantlar da benzer isimler almaktadır (23). Bu varyantlardan Bölüm 2.1.5'te daha detaylı olarak bahsedilmektedir.

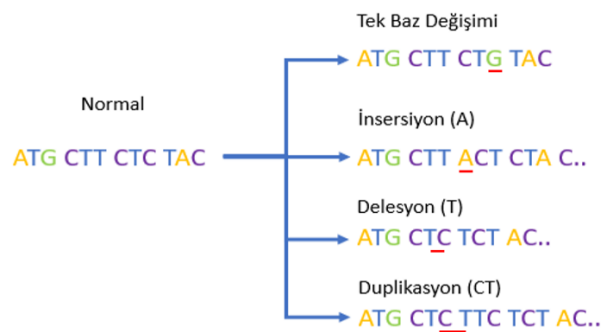
2.1.4. Küçük Ölçekteki Varyantlar

Moleküler biyoloji alanındaki yenilikler sonucu, 1977 yılından itibaren genetik dizileme teknolojileri geliştirilmiş ve insan genom projesi sırasında bu teknolojilerin kullanımının yaygınlaşması ile nükleotit düzeyindeki küçük değişiklikler dikkat çekmeye başlamıştır (24). Küçük ölçekteki varyantlar, insan genomunda sayıca en fazla bulunan ve insan genetik çeşitliliğine büyük katkı sunan varyantlardır. Bu varyantlar arasında tek nükleotit değişiklikleri (SNV, *Single Nucleotide Variation*) çeşitli tekrar dizilerindeki tekrar sayısı değişiklikleri (STR, *Short Tandem Repeat*, Kısa Ardışık Tekrar) ya da VNTR, (*Variable Number of Tandem Repeat*, Değişken Sayıda Ardışık Tekrar), delesyon ve insersiyonlar (indel) gibi kısa değişiklikler bulunmaktadır. Önceleri Sanger dizileme yöntemleri ile ortaya çıkarılmaları sebebi ile küçük ölçekteki değişikliklerin üst sınırı bu yönteminin saptayabildiği yaklaşık en uzun değişiklik olan 1000 bp (baz çifti, *base pair*) olarak kabul edilmekteydi (25). Buna karşın, günümüzde küçük ölçekteki varyantların sıklıkla NGS ile saptanması sebebiyle NGS’de ulaşılan okuma uzunlukları ve veri analizinin getirdiği kısıtlamalar göz önüne alınarak son yıllarda küçük ölçekteki varyantların üst sınırı 50 bp olarak kabul görmektedir (26).

Küçük ölçekteki varyantlar arasında özellikle SNV’ler insan evrimi ve hastalıklarının patogenezi açısından uzun yıllardır araştırma konusu olmuştur ve bu araştırmalar sonucu SNV’lerin farklı insan popülasyonlarında görülme sıklıkları farklı veri tabanlarında kataloglanmıştır (27, 28). Böylece kompleks hastalıklar ve etkin genotipleme açısından daha önemli kabul edilen popülasyonun en az %1’inde görülen görece yaygın varyantlar olan tek nükleotit polimorfizmleri (SNP, *Single Nucleotide Polymorphism*) ortaya konmuştur. Bunun yanında, Mendelyan kalıtılan nadir tek gen hastalıklarının sebebi olabilecek nadir varyantların saptanması da daha kolay hale gelmiştir. SNV’ler tek bir bazı ilgilendiren değişikliklerdir. Belli bir genomik noktada yer alan bir Adenin (A) nükleotidinin Sitozin (C) nükleotidine dönüşmesi SNV’lere örnek teşkil edebilir. Bu türden değişiklikler her zaman protein düzeyinde değişiklik yapmazlar. Eğer protein kodlayan bir bölgedeki SNV sonucu bir amino asit başka bir amino aside dönüşmüşse bu yanlış anlamlı (*missense*) mutasyon; ancak protein düzeyinde herhangi bir değişiklik olmamışsa sinonim olarak adlandırılmaktadır.

Proteinde erken bir “dur” kodonu oluşturan mutasyonlar ise anlamsız (*nonsense* veya *stop-gain*) mutasyon olarak bilinmektedir (29). *Nonsense* mutasyonda ortaya çıkan prematüre stop kodonu nedeniyle kısa bir protein oluşmaktadır. Daha nadir görülen beklenen “dur” kodonunun kaybolması (*stop-loss*) mutasyonu, protein sentezinin durması gereken kodonun ötesinde devam ederek proteine yeni amino asitler eklenmesine sebep olmaktadır. Bir diğer nadir mutasyon olan başlangıç kodonunun kaybolması (*start-loss*) ise proteinin sentezinin başlayacağı kodonun kaybolması ya hiç protein sentezlenememesi ya da başka bir pozisyonun başlangıç kodonu olarak algılanması sonucu yanlış bir proteinin sentezi ile sonuçlanmaktadır.

SNV dışındaki küçük ölçekli varyantlar, genomda bir ya da birkaç bazı kapsayan delesyonlar, duplikasyonlar, insersiyonların meydana gelebildiği varyantlardır. Delesyonlar bir ya da birden fazla nükleotidin DNA dizisinden silindiği; insersiyonlar ise diziyeye yeni bazların eklendiği mutasyonlardır. Duplikasyonlar ise dizi içinde yer alan nükleotit serisinin kendini bir kez daha kopya etmesi sonucu oluşan mutasyonlar olup art arda ya da mesafeli olarak konumlanabilmektedirler. Bu varyantlar protein kodlayan bölge içerisinde bulunduğunda bazı amino asitlerin eklenmesine veya silinmesine neden olabileceği gibi, ortaya çıktığı noktadan sonra tüm amino asit dizilimini bozan çerçeve kayması olarak nitelendirilen ciddi etkilere sebep olabilmektedirler (Şekil 2.1) (21).



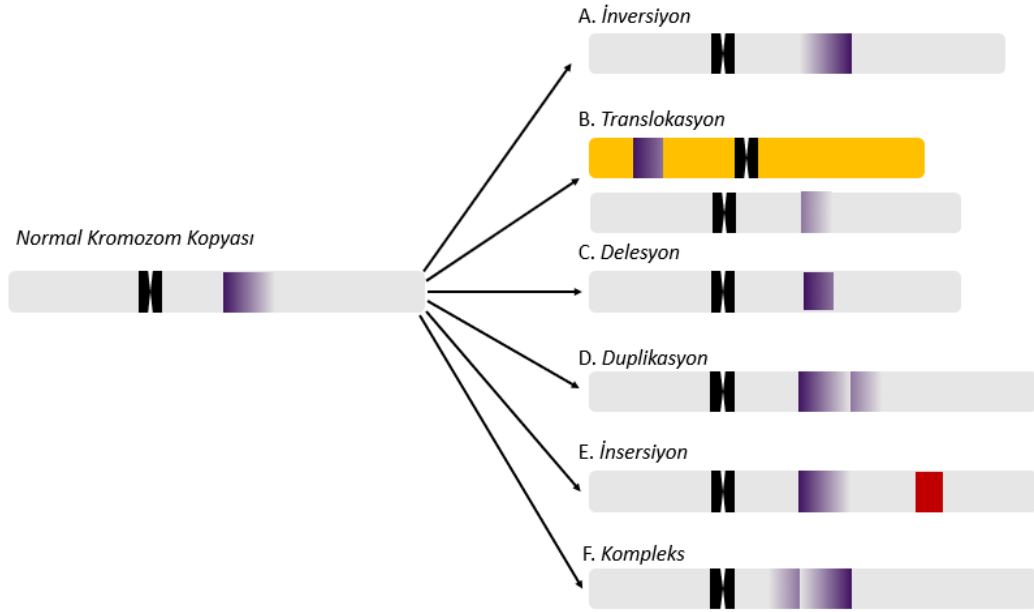
Şekil 2.1. Küçük ölçekli varyant tipleri.

Solda referans nükleotit dizisi, sağda ise sırasıyla tek baz değişimi, delesyon ve duplikasyon değişiklikleri görülmektedir. Her bir varyant kırmızı ile işaretlenmiştir. Delesyon, duplikasyon ve insersiyon varyantlarının üçlü kodon okumalarını kaydırarak çerçeve kaymasına sebep olduğuna dikkat ediniz.

2.1.5. Orta Ölçekteki Varyantlar

Büyük ve küçük ölçekteki varyantların saptanmasına yönelik yöntemler genetik alanında uzun yıllardır kullanılmaktayken bu varyantların dışında kalan ve 50 bp ile 3 Mbp arasındaki uzunluktaki varyantların tespitine yönelik teknolojik gelişmeler son 20 yılda hız kazanmıştır. Bu hızlanma sürecinde, orta ölçekteki varyantların genom boyu saptanmasında kullanılan mikrodizin teknolojisinin yaygın kullanımı önemli bir yer tutmaktadır. Yapısal varyantlar olarak da adlandırılan bu varyantlar, SNV'ler kadar çok sayıda olmasalar da genomda kapladıkları alanlar büyüktür. Sağlıklı bir popülasyonda insan genomunun %4,8 ile %9,5 arasında bir bölümünün yapısal varyantlarla ilişkili olduğu gösterilmiştir (30). Bu nedenle, yapısal varyantlar bireyler arası genomik varyasyona önemli katkı sunmaktadır. Bunun yanında, yapısal varyantların gerek tek gen hastalıkları gerek kromozomal hastalıklar gerekse de kompleks hastalıklarla ilişkisi ortaya konmuştur (31-33). DiGeorge Sendromu gibi bazı mikrodelsiyon sendromları, yapısal varyantlarla ilişkili konjenital genetik hastalıklar arasında önemli bir yer tutmaktadır (34).

Yapısal varyantlar arasında delesyonlar, duplikasyonlar, inversiyonlar, insersiyonlar ve translokasyonlar yer almaktadır. Ayrıca, daha nadir olarak bu varyantlardan birkaçının bir arada bulunduğu kompleks düzenlenmeler de mevcuttur. Büyük ölçekteki varyantlar ile yüksek benzerlik gösteren bu varyantları kromozomal yapısal varyantlardan ayırt etmek için genellikle submikroskopik veya mikro-eki kullanılmaktadır. Örneğin "delesyon" ifadesi genellikle sitogenetik düzeyde saptanabilecek bir değişikliği ifade ederken, "submikroskopik delesyon" veya "mikrodelsiyon" ifadeleri orta ölçekteki yapısal varyantları işaret etmektedir (35). Faklı yapısal varyantların ifade ettikleri genetik değişiklikler Şekil 2.2'de sunulmuştur.



Şekil 2.2. Yapısal varyantlar.

Solda bir kromozoma ait normal durum temsil edilmekte, sağda ise yapısal varyantlar belirtilmektedir. A) İnverson, bir genomik bölgenin ters dönmesi B) Translokasyon, bir genomik bölgenin başka bir kromozoma geçmesi C) Delesyon, bir genomik bölgenin silinmesi D) Duplikasyon, bir genomik bölgenin kendini hemen yanındaki genomik bölgeye kopyalaması olması E) İnsersiyon, kromozoma yeni bir dizinin girmesi F) Bir genomik bölgede birden fazla yapısal varyant (inverson ve duplikasyon) bulunmasıyla sonuçlanan kompleks değişiklik.

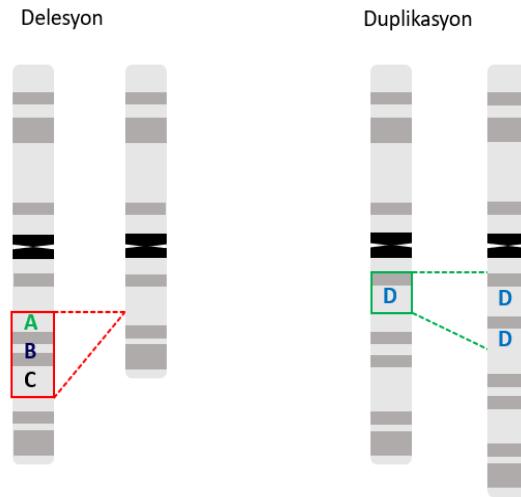
2.2. Yapısal Varyantlar

Orta ölçekteki genomik değişiklikler olarak da ifade edilen yapısal varyantlar, insanlarda bir genomik bölge için beklenen sayıda alelin bulunması veya bulunmamasına göre sırasıyla kopya sayısı değişiklikleri veya kopya nötral değişiklikler olmak üzere iki gruba ayrılmaktadır.

2.2.1. Kopya Sayısı Değişiklikleri

İnsan genomunda >50 bp olan her genomik bölge için beklenen kopya sayısına kıyasla kopya sayısında kazanç ve kayıp şeklinde ortaya çıkan varyasyonlara CNV denilmektedir. Cinsiyet kromozomları için erkeklerde farklılık gösterse de tüm otozomlar ve PAR'lar için beklenen kopya sayısı 2'dir. Kopya sayısı kazanç olayları duplikasyon, kayıpları ise delesyon adıyla da anılmaktadır. Bir genomik bölgenin bir hücre içerisindeki kopya sayısı tamsayı şeklindedir. Buna göre, otozomlarda yer alan

bir genin bir kopyasının delesyonuyla o gen için kopya sayısında kayıp, duplikasyonuyla ise o gen için kopya sayısında kazanç söz konusu olmaktadır (Şekil 2.3).



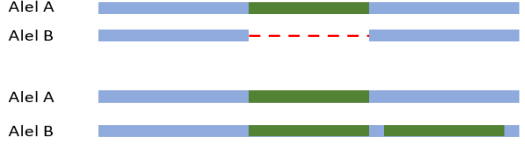





Şekil 2.3. Kopya sayısı değişiklikleri.

Solda bir kromozom kopyasındaki ABC genlerini içeren genomik bölgenin silinmesiyle kopya sayısının 1 olduğu görülmektedir. Sağda ise D genini içeren genomik bölgenin duplike olmasıyla kopya sayısı 3 olmaktadır. Bu durumda, ABC geni için kayıp (delesyon) söz konusuysen D geni için kazanç (duplikasyon) vardır.

Kalıtsal ya da edinsel olabilen CNV'ler her zaman bir fenotipik sonuç ile ilişkilendirilmeyeceği gibi bir kısmı da tıbbi açıdan takip ya da müdahale gerektirecek kadar önemli fenotipik etkilere yol açabilmektedir. CNV'lerin sebep olabildiği bazı hastalıklar arasında DiGeorge Sendromu ve Williams-Beuren Sendromu gibi mikrodelesyon sendromları ile birçok kompleks nöro-gelişimsel hastalık yer almaktadır. Bu kalıtsal hastalıkların yanı sıra birçok çalışmada kanser ilişkili genlerin kalıtsal veya edinsel CNV'lerin içinde veya yakınında bulunduğu gösterilmiştir (36).

CNV'ler, bir ekzon bölgesini etkileyecek kadar küçük veya onlarca geni etkileyecek kadar büyük genomik bölgeleri içerebilirler. Bu genomik olayların fenotipik etkileri kapsadıkları bölgedeki gen ve diğer genomik elemanlar, bunların işlev kaybına veya işlev kazancına duyarlılıkları, ilgili alel üzerindeki mutasyonun hastalıkla ilişkisi gibi faktörlerden etkilenmektedir (37). CNV'ler sonucu ortaya

çıkabilecek fenotipik özellikler Şekil 2.4'te sunulan mekanizmalarla açıklanabilmektedir.

A) Dozaj duyarlı genlerin delesyonu (üstte) veya duplikasyonu (alt) kopya sayısının etkilenmesi	
B) Birden fazla genin silinmesi	
C) Genin dizisinin kesilmesi	
D) Gen füzyonu	
E) Genin regülatör bölgesinin silinmesi	
F) Resesif hastalıkla ilişkili bir mutasyonun tek aktif alel olarak kalması	

Şekil 2.4. CNV'lerin fenotipik etkileri.

Yapısal varyantların hastalıklar ile ilişkili olabilecek çeşitli fonksiyonel sonuçları belirtilmiştir. Solda yapısal varyantın sebep olduğu durum tarif edilirken sağda ise genom üzerindeki etkisi şekilsel olarak açıklanmıştır. A-E başlıklarında gösterilen olaylar bir genomik bölgenin tek alelinde ortaya çıkan yapısal varyantları gösterirken F başlığında tek aleldeki yapısal varyantın karşı aleldeki kısa varyant ile beraber etkisine bir örnek gösterilmektedir.

2.2.2. Kopya Nötral Değişiklikler

Kopya nötral değişiklikler, kopya sayısının 2 olarak korunduğu yapısal varyantlardır. Bu varyantların ortak özelliği kromozomlar üzerindeki genomik bölgelerin sadece sıralamasında bir değişikliğe sebep olmalarıdır. Translokasyon, inversiyon ve insersiyon gibi varyantlar herhangi bir genomik kayıp veya kazanç ile ilişkili olmadıklarından kopya nötral değişiklikler olarak sınıflandırılmaktadırlar (Şekil 2.2).

Kopya nötral deęişiklikler, çoęunlukla sessiz deęişiklikler olsa da kopya nötral deęişiklikler sonucu ortaya çıkan insan hastalıkları bulunmaktadır (38). Bu deęişiklikler, fenotip üzerindeki etkilerini CNV'lerden farklı olarak dozaj etkisi ile göstermezler. Ayrıca, kopya nötral deęişiklikleri genom boyu saptayabilen yöntemler oldukça kısıtlı oldukları için bu varyantların etkisi ile ilgili bilgiler de sınırlı olmaktadır (39).

2.2.3. Kopya Sayısı Deęişikliklerinin Oluşum Mekanizmaları

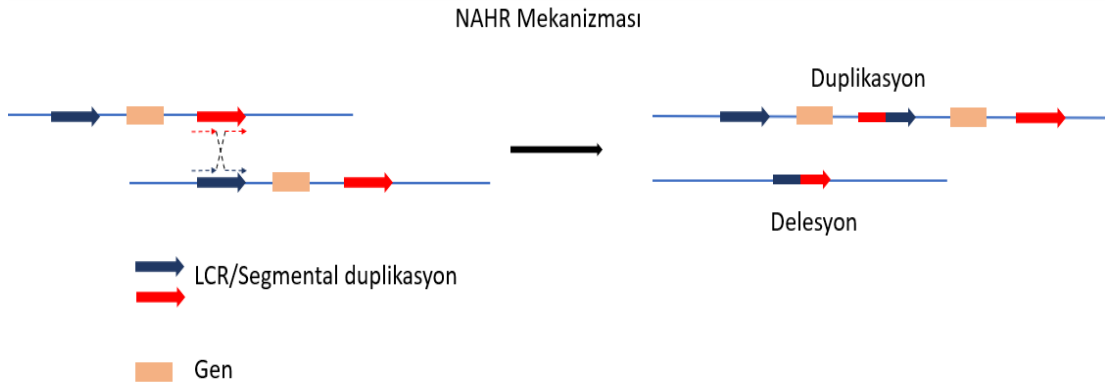
Gerek mayoz sırasındaki rekombinasyon hataları gerekse replikasyon veya replikasyon dışı DNA tamiri sonucu ortaya çıkan hatalar CNV'lere sebep olabilmektedir. CNV'lerin çoęunluğu, aralarında >%95 sekans benzerlięi olan ve >1 kbp uzunluktaki segmental duplikasyon ya da LCR (*Low Copy Repeat*, düşük kopyalı tekrar) olarak adlandırılan bölgelerle ilişkilidir (40). LCR bölgeleri ile ilişkili bu CNV'ler NAHR mekanizması sonucu ortaya çıkarken, LCR'lerden baęımsız ve genelde DNA tamiri ile ilişkili farklı mekanizmalar da CNV oluşumu ile sonuçlanabilmektedir (41).

2.2.3.1. Alelik Olmayan Homolog Rekombinasyon (NAHR)

Aleller, homolog iki kromozomun aynı lokusunda yerleşen ve aynı karakteri temsil eden genomik bölgelerdir. Homolog rekombinasyon (HR), aynı lokustaki bu iki alel arasında gerçekleşir ve bu işlem sonucunda aleller birbirleri arasında belirli bir bölgeyi deęiştirirler (Şekil 2.5). HR, mayoz sırasında kros-over veya mitoz sırasında DNA tamirinde gerçekleşebilmekte, ayrıca çift zincir DNA kırıklarının tamiri için kullanılan neredeyse hatasız bir mekanizma olarak bilinmektedir. Alelik Olmayan Homolog Rekombinasyon (NAHR, *Non-Allelic Homologous Recombination*) ise, rekombinasyonun alelik HR'de olduęu gibi aynı lokustaki aleller arasında olması yerine aynı ya da farklı kromozomlar üzerindeki LCR bölgeleri arasında gerçekleştięi bir HR çeşidini ifade etmekte ve farklı yapısal varyantlar ile sonuçlanabilmektedir (42).

NAHR sonucu yapısal varyantlar; segmental duplikasyonlar arası dizi benzerlięi, konumları ve oryantasyonlarına baęlı olarak farklı şekillerde ortaya çıkabilmektedir. Örneęin, aynı oryantasyondaki ve aynı homolog kromozomlardaki iki

LCR arası NAHR gerçekleşirse bir homolog kromozomda delesyon diğerinde ise duplikasyon ortaya çıkmaktadır (Şekil 2.5). Bir diğer senaryoda, eğer iki LCR zıt oryantasyondaysa, rekombinasyon sonucu inversiyonu da içeren bir yapısal varyant görülmektedir (43).



Şekil 2.5. NAHR mekanizması.

Mavi ve kırmızı oklar aynı yönlü yüksek dizi homolojisi gösteren segmental duplikasyonları temsil etmektedir. Siyah renkli kesikli çapraz çizgiler ise iki LCR arası gerçekleşen homolog rekombinasyonu göstermektedir. Bu olay, LCR'ların yönü ve rekombinasyonun gerçekleştiği genomik bölge itibariyle bir aleldeki genin duplikasyonu ve diğer aleldekinin ise delesyonuna sebep olmaktadır.

NAHR mekanizması, başlangıç ve bitişleri aynı olan ve farklı bireylerde *de novo* mutasyon olarak ortaya çıkabilen rekürren CNV'lerin başlıca sebebi olarak görülmektedir (43). Bu duruma en iyi örneklerden biri olarak *PMP22* gen bölgesinde görülen rekürren CNV sonucu ortaya çıkan Charcot-Marie-Tooth hastalığı gösterilmektedir (44). Buna göre, bu hastalıktan sorumlu CNV'ler diğer CNV'lerden ayırt edici olarak her zaman aynı genomik bölgede yaklaşık 3,4 Mbp uzunluğunda ve benzer bölgelerden başlayıp benzer bölgelerde bitmektedirler. Rekürren CNV'ler, tüm *PMP22* ilişkili Charcot-Marie-Tooth hastalığına sahip bireylerdeki delesyonların %99'unu oluşturmaktadır (45).

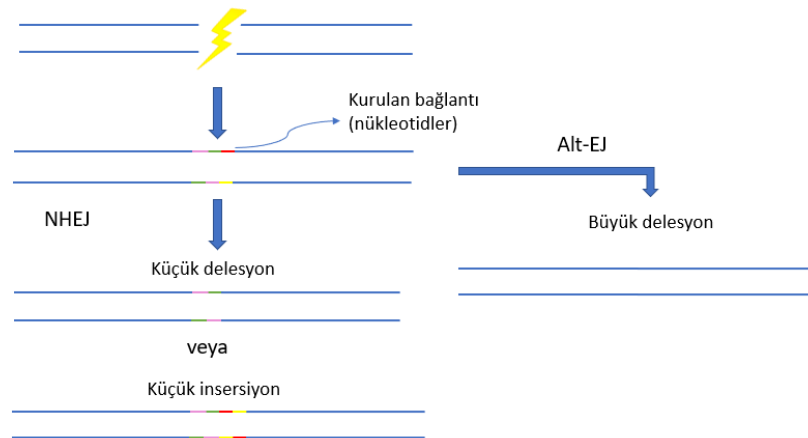
2.2.3.2. Homolog Olmayan Uç Birleştirme (NHEJ, *Non-Homologous End Joining*) ve Alternatif Uç Birleştirme (Alt-EJ)

Çift zincir DNA kırıklarının HR'den başka bir tamir mekanizması da hataya daha açık olan ve genomda "moleküler yara" bırakan mekanizmalardır. Bu mekanizmalar

arasında Homolog Olmayan Uç Birleştirme (NHEJ, *Non-Homologous End Joining*) ve Alternatif Uç Birleştirme (Alt-EJ, *Alternative End Joining*) ön plana çıkmaktadır. Her iki mekanizma da daha çok rekürren olmayan CNV'ler ile ilişkilidir (46).

NHEJ'de bir çift zincir DNA kırığı, kırık uçlar arası köprü oluşturularak tamir edilmekte ve birleşme noktasında küçük delesyon ve insersiyonlar ile çeşitli yapısal varyantlar oluşabilmektedir. NHEJ, homoloji ya da LCR gereksinimi duymasa da tekrar dizilerinin olduğu yerlerde daha sık meydana gelmektedir (47).

Çift zincir DNA kırıklarında uç birleştirmenin bir diğer yolu da NHEJ ile benzerlik gösteren Alt-EJ'dir. Bu mekanizmanın NHEJ'den farkı, birleştirilecek genomik bölgelerin uçlarında yer alan en az 3 bp uzunluğundaki mikrohomojoloji adı verilen küçük dizi benzerliklerinin DNA tamir proteinlerince tanınmasıdır. NHEJ ve Alt-EJ mekanizmaları Şekil 2.6'da gösterilmektedir.



Şekil 2.6. NHEJ Mekanizması.

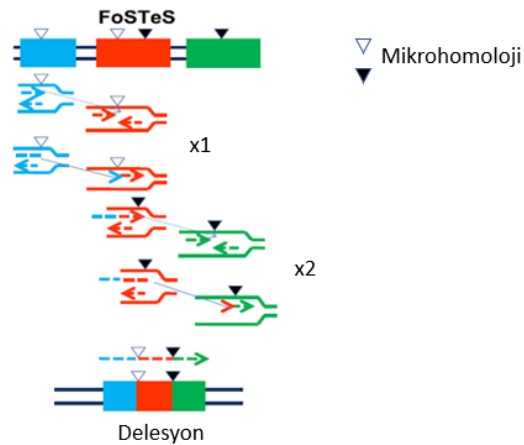
Çift zincirli DNA kırığının ardından devreye giren NHEJ mekanizması sonucunda ortaya çıkan olası sonuçlar gösterilmektedir. NHEJ ile kırık uçlar nükleotit köprüsüyle birleştirildiğinde küçük delesyon veya insersiyon meydana gelebilmektedir. Alt-EJ ile ise mikrohomojoloji veya daha büyük homolojilerin tamir mekanizmasında görev almasından dolayı daha büyük delesyonlarla (>1 kbp) karşılaşmaktadır (48).

2.2.3.3. DNA Replikasyonu Sırasındaki DNA Hasarının Tamir Mekanizmaları

FoSTeS (*Fork Stalling and Template Switch*, Çatalın Duraksamasıyla Kalıp Değişirme) veya MMBIR (*Microhomology-Mediated Breake-Induced Replication*,

Mikrohomoloji-Aracılı Kırıkla İndüklenen Replikasyon) mekanizmaları DNA replikasyonu sırasında oluşan hatalara karşı replikasyonu devam ettirme amaçlı tamir mekanizmalarıdır (49).

Normal şartlar altında, hücrede DNA replikasyonu öncesinde DNA'da hasar tespit edildiğinde hücrenin replikasyona girmesi hücre döngüsünü kontrol eden proteinler tarafından engellenmektedir (50). Ancak eğer DNA replikasyonu sırasında bir hata ile karşılaşırsa replikasyon çatalı durdurulmakta, kıraktan kurtulmak adına yeni sentezlenmekte olan DNA ipliğinin 3' ucu farklı bir DNA kalıbına atlayarak replikasyonun devam etmesi sağlanmaktadır. Atlama işlemi yakın bölgede bir mikrohomoloji dizisinin varlığını gerektirmekte ve yeni sentezlenen dizinin stabilitesini artırmak için birkaç kez tekrarlanabilmektedir (Şekil 2.7). Bu şekilde gerçekleşen atlama işlemi FoSTeS olarak adlandırılmakta ve DNA replikasyonunun devam etmesini mümkün kılmaktadır.



Şekil 2.7. FoSTeS mekanizması.

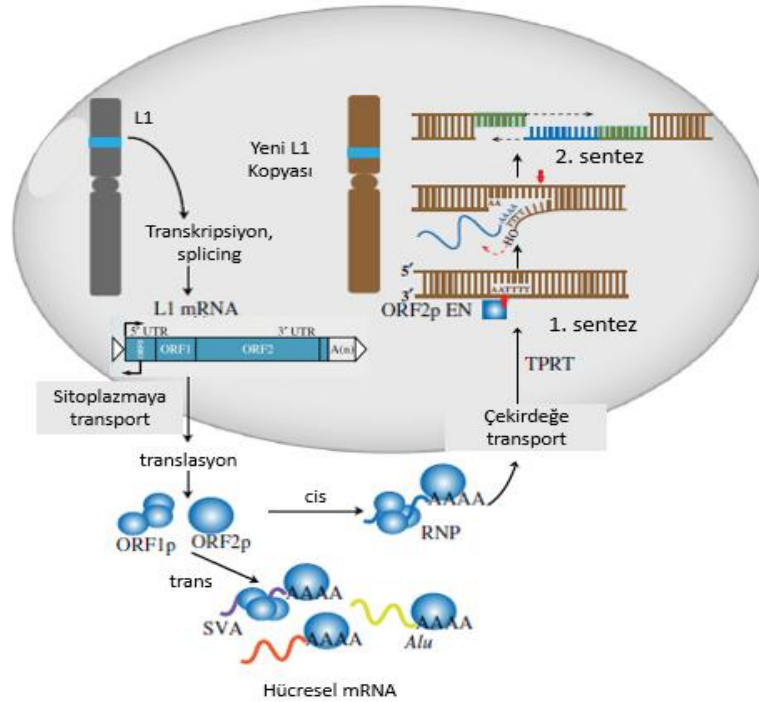
Şekilde içi boş ve dolu aşağı bakan üçgenler mikrohomoloji barındıran 2-5 bp uzunluğundaki dizileri işaret etmektedir. Şekilde verilen örnekte, DNA'nın mavi ile işaretli bölgesinde replikasyon çatalı bir hata ile karşılaştığı zaman duran replikasyonu devam ettirebilmek amacıyla mikrohomoloji gösteren bölgelerden DNA replikasyonuna devam etmektedir. Bu sebeple maviden kırmızı bölgeye ve kırmızıdan yeşil bölgeye olmak üzere iki kez kalıp DNA zinciri değiştirilmekte ve sonuçta kırmızı, mavi ve yeşil kutucuklardan bazı kısımların silindiği kompleks bir delesyon gözlenmektedir. Bu şekil CC BY 2.0 lisansı altında açık erişimde paylaşılan bir makaleden türetilmiştir (51).

FoSTeS'in yarattığı önemli bir sorun mikrohomoloji tutturulan yerin replikasyon çatalının normalde devam etmesi gereken dizinin önünde ya da sonrasında yer alması ve buna bağlı olarak delesyon veya duplikasyonların meydana gelmesidir (52).

Sonraki yıllarda mikrohomoloji kontrolünde olan bu mekanizmanın DNA kırıkları sonucu aktive olan bir türüne MMBIR da denmiştir. MMBIR'da replikasyon sırasında kalıp zincirde karşılaşılan bir DNA tek zincir kırığı sonucu mikrohomoloji içeren bölgelere sıçramalar olur ve FoSTeS'tekine benzer şekilde bu sorun farklı dizilerin sentezi ile aşılır (50). MMBIR mekanizmasında, duraksamış replikasyon çatalının açıkta kalan 3' ucunun herhangi bir tek zincirli DNA kalıbına, ve hatta saç tokası (*hairpin*) gibi ikincil DNA yapılarına atlayabildiği gösterilmiştir (53). Replikasyon çatalındaki kalıp DNA'yı değiştiren MMBIR olayında genellikle kalıp zincirin birden fazla kere değiştirmesi ve mikrohomolojinin rastgele bir yerde olabilmesi sebebiyle bu mekanizma sonucu çoğunlukla kompleks genomik değişikliklerin oluştuğu düşünülmektedir (52).

2.2.3.4. L1 Retrotranspozisyonu

L1 elemanları insan genomunda sınıf 1 (LINE, *Long Interspersed Nuclear Elements*, Uzun Serpiştirilmiş Nükleer Elemanlar) transpozonlar olarak bilinen, bir noktadan bir noktaya 'sıçrayıcı' elemanlardandır (54). Bu elemanların genomun yaklaşık %17'sini oluşturduğu bilinmektedir; fakat çoğu aktif değildir. L1 retrotranspozisyonu mekanizmasıyla, kompleks yapısal değişiklikler ve CNV'ler oluşabilmektedir. L1 retrotranspozisyonunun nasıl çalıştığı tam anlamıyla netlik kazanmamakla birlikte, temel sebebinin L1'lerin dizisinde yer alan ORF1 ve ORF2 kaynaklı olduğu bilinmektedir (53). ORF1'in RNA bağlayıcı protein kodlayabilme, ORF2'nin ise endonükleaz ve ters transkriptaz aktiviteleri bulunmaktadır. Basit olarak, bir L1 elemanı zıpladığı başka bir diziye kodladığı endonükleaz ve ters transkriptaz aktivitesi sayesinde yerleşmekte ve kendi kopyasını sentezlemektedir (Şekil 2.8) (55).



Şekil 2.8. L1 retrotranspozisyon mekanizması.

Bir L1 elemanı transkripsiyon ve translasyonunun ardından L1 tarafından kodlanan proteinlerin endonükleaz ve ters transkriptaz aktiviteleriyle L1 mRNA'ları kendilerini genomdaki başka bir lokusa kopyalayabilmektedirler. Bu şekil CC BY 4.0 lisansı altında açık erişimde paylaşılmış bir makaleden türetilmiştir (55).

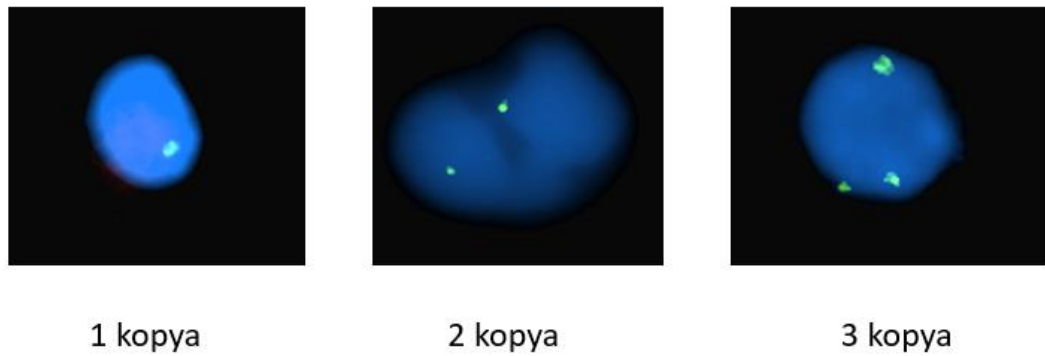
2.2.4. CNV Saptama Yöntemleri

2.2.4.1. Florasan in Situ Hibridizasyon (FISH)

FISH yöntemi, insan genomundaki ilk uygulamalarını içeren 1982 yılını takiben, 1990'lı yılların sonlarında hem rutin tanı hem de araştırma alanlarında kullanımı yaygınlaşarak sitogenetiğin altın standartlarından kabul edilen bir yöntem haline gelmiştir (56). DNA hibridizasyonuna dayanan FISH, prensip olarak yapısal varyantların tespitinde, florasan işaretli problemlerin genomda bilinen spesifik hedef bölgelere bağlanması ve bu problemlerin florasan mikroskop yardımıyla gözlemlenmesine dayanmaktadır (55).

Tanı açısından, önemli olan genomik bölgeler ya da spesifik genler için tasarlanan problemler, hedef bölgeye hibridize olduğunda florasan olarak o bölgenin işaretlenmesini sağlar. Herhangi bir orta ölçekli mutasyon durumunda bölge

tanınmaz ve ışın alınamaz ise bağlanma gerçekleşmeyeceği için eksik sinyal; aynı genomik bölgenin genomda ekstra kopyaları bulunuyorsa fazla sinyaller elde edilebilir ve böylece 2 kopya beklenen bölgelerin farklılıkları floresan mikroskop ile görülebilir (57). Bu yöntem kullanılarak genomda kromozom analizi ile saptanamayacak küçüklükteki delesyon ve duplikasyonlar hücre metafaz veya interfaz aşamasındayken saptanabilmektedir (Şekil 2.9). Buna karşın FISH yönteminin çözünürlüğü 100-200 kbp'den büyük CNV'lerin saptanmasına olanak tanımaktadır (58). FISH ile tek tek her hücrenin ya da hücre çekirdeği içindeki kromozomların fotoğraflanabiliyor olması ve aynı zamanda genomik bölgelerin sayısal bilgisinin elde edilebiliyor olması bu yöntemi günümüzde halen daha bazı durumlarda tercih edilebilir bir yöntem kılmaktadır.



Şekil 2.9. FISH yöntemi ile CNV'lerin saptanması.

FISH yöntemi ile interfaz aşamasındaki hücre çekirdekleri (mavi) içerisinde hedeflenen kromozomal bölgeye yönelik sinyaller (yeşil) görülmektedir. Soldan sağa 3 görüntü incelendiğinde, sinyal sayılarına göre sırasıyla işaretlenen genomik bölgenin monozomik (tek kopya), dizomik (2 kopya) ve trizomik (3 kopya) olduğuna dikkat ediniz.

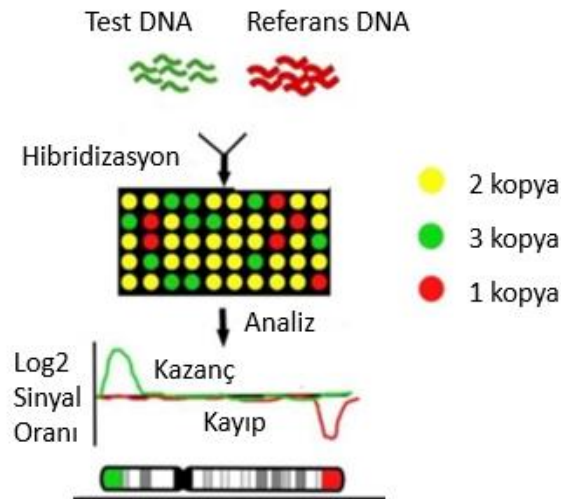
2.2.4.2. Mikrodizin ile Karşılaştırmalı Genomik Hibridizasyon (*array Comparative Genomic Hybridization, aCGH*)

aCGH yöntemi, FISH yöntemi gibi hibridizasyon temelli bir yöntem olup mikrodizin (*microarray*) teknolojisi ile imal edilmiş zeminler üzerinde pek çok hibridizasyon işleminin paralel olarak ve aynı anda yapılması ilkesi ile çalışmaktadır. Özellikle kopya sayısı değişikliklerinin tespitine yönelik olarak kullanılan bu yöntem, 2000'li yıllarda CNV'lerin aydınlatılmasına çok büyük katkı sağlamıştır (5). Adından da

anlaşılabilirliği üzere, bir referans DNA ile hedef DNA'nın hibridizasyon açısından karşılaştırılması kullanılarak CNV'lerin araştırıldığı bir yöntemdir.

aCGH yönteminde referans ve hedef DNA'nın sırasıyla Cy5 (kırmızı) ve Cy3 (yeşil) gibi farklı florasan boyalarla işaretlenmesinin ardından BAC (Yapay Bakteri Kromozomu, *Bacterial Artificial Chromosome*) veya oligonükleotit problemlere "yarışmalı" bir hibridizasyon ile bağlanması sağlanmaktadır (59). Florasan işaretler sayesinde sinyal şiddetlerinin oranı LRR (logaritmik R oranı, *Logarithmic R Ratio*) veya \log_2 oranı olarak hesaplanmakta ve bu sayısal veriye dayanarak CNV'ler delesyon veya duplikasyonlar şeklinde belirlenebilmektedir. LRR, bir genomik pozisyon için elde edilen sinyal şiddetinin, beklenen ve 2 kopyaya eşit olan sinyal şiddetine oranının \log_2 değerini ifade etmektedir (60). Buna göre, bu oranın karşılığı kopya sayısının 2'ye eşit olduğu durumlar için her zaman 0 olmaktadır. Bu nokta referans alınarak 0'ın üstünde gözlenen değerler kopya sayısındaki göreceli artışa (duplikasyon) işaret ederken, 0'ın altındaki değerler kopya sayısındaki azalışlara yani delesyonlara işaret etmektedir (Şekil 2.10) (61).

Çözünürlüğü, tasarlanan prob sayısına ve problemlerin genomda karşılık geldiği lokusların dağılımına göre değişiklik gösteriyor olsa da aCGH yöntemi oligonükleotit problemler ve ekstra problemler kullanıldığında ~50 kbp'a kadar tespit yapılabilmektedir (62). Bu yöntem genom boyu analize imkan sağlasa da alel spesifik sinyal şiddetini saptayamadığı için bir genotip bilgisi oluşturamamaktadır. Bu nedenle, günümüzde yerini çoğunlukla SNP mikrodizirlere bırakmış olsa da daha basit protokolü, validasyon kolaylığı ve CNV spesifik olması açısından tercih sebebi olmaktadır.



Şekil 2.10. aCGH yöntemi.

Yeşil florasanla işaretlenen test örneği ve kırmızı florasanla işaretlenen referans DNA örneğinin birlikte hibridizasyonunun ardından yapılan analize göre kromozom boyu \log_2 sinyal oranları (LRR) referansa kıyasla fazla gelen genomik bölgeler kopya sayısı kazançlarını, düşük gelen genomik bölgeler ise kopya sayısı kayıplarını göstermektedir. Bu resim CC (*Creative Commons*) BY 4.0 lisansı altında üretilen açık erişimli bir makaleden türetilmiştir (63).

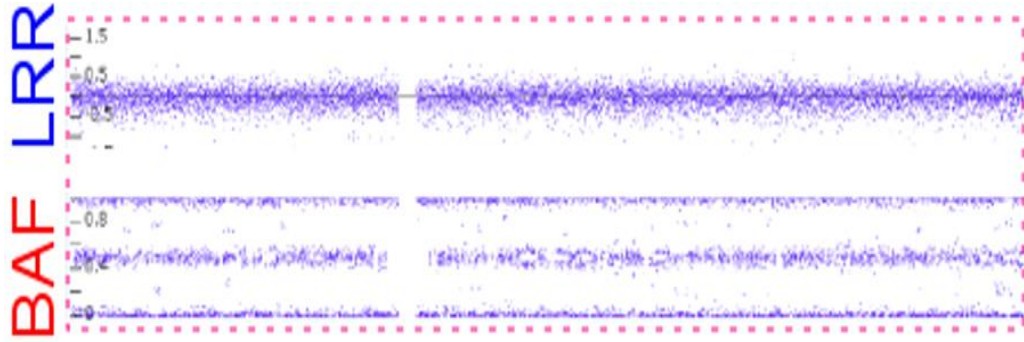
2.2.4.3. SNP Mikrodizin

CNV'lerin saptanmasında günümüzde en sık kullanılan yöntemlerden biri olan SNP mikrodizin yöntemi, hedef SNP pozisyonlarına yönelik tasarlanan yüz binlerce DNA probu kullanılarak genom boyu analiz yapılmasını sağlamaktadır. Prob hibridizasyonuna dayanan bu yöntem, aCGH yönteminin aksine örnek DNA'nın bir referans DNA ile karşılaştırılmasına gerek duymamaktadır. Temel çalışma ilkesi, mikrodizin çipi olarak adlandırılan bir cam ya da polimer yüzey üzerinde her bir noktada spesifik DNA prob dizilerinin bulunduğu ve bu noktalara florasan boyayla etiketlenmiş olan örnek DNA'nın hibridize olması sağlanarak florasan işaretlerden alınan sinyallerin görüntülenip işlenmesidir (64).

Rutin tanı ve araştırmada yaklaşık 20 yıldır kullanılmasıyla güven kazanmış olan SNP mikrodizin yöntemi, aCGH teknolojisine bir yenilik olarak alel spesifik veriler de oluşturabilmekte ve kopya sayısı bilgisinin yanında genotip bilgisinin de edinilmesine imkân vermektedir (65). Genotip bilgisi, aCGH analizlerinde bulunmayan

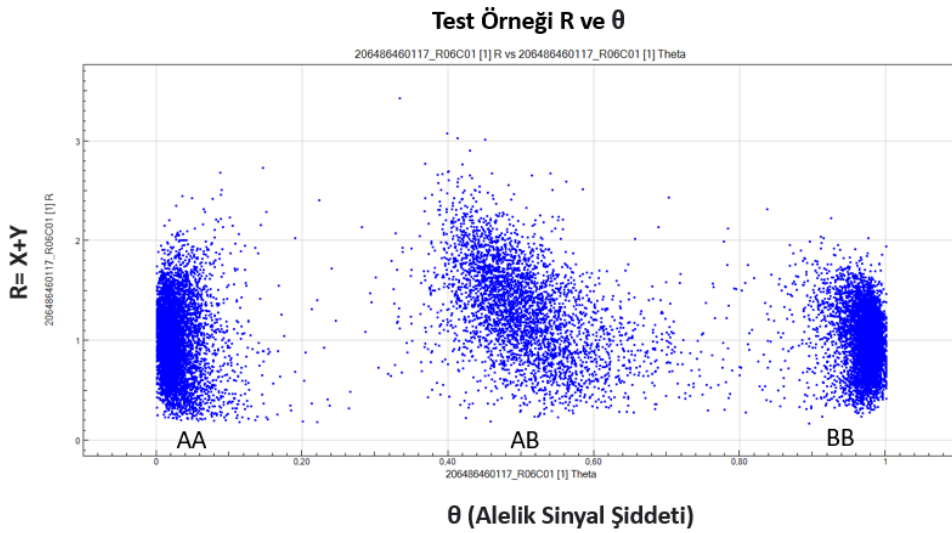
ve SNP mikrodizin analizlerinde B alel frekansı (BAF) olarak bilinen bir veri ile ortaya konmaktadır (Şekil 2.11). BAF değeri, bir genomik pozisyon için B aleli olarak adlandırılan bir alelin tüm aleller (A+B) içindeki oranını ifade etmektedir. Burada B aleli varyant aleli, A aleli ise referansı temsil etmektedir. BAF değeri, 2007 yılında PennCNV algoritması olarak açıklanan bir metoda dayanmaktadır (66). BAF değeri, aynı lokusa denk gelen farklı SNP'ler arasındaki sinyal şiddeti oranına karşılık gelen θ (teta) değeri baz alınarak hesaplanmaktadır. θ değeri, her bir genomik lokus için hesaplanıp daha önceden o lokus için bilinen 3 genotipe ait θ_{AA} , θ_{AB} ve θ_{BB} değerleri ile kıyaslanarak o lokus için genotip tayin edilmektedir. θ değerinden köken alan BAF değeri ise, 0 ile 1 arasındaki tüm değerleri alabilmekte olup bu değer B alelinin tüm aleller içindeki oranını göstermektedir. Genotip dağılımlarının normalize edilmesiyle nihai BAF değerleri elde edilmektedir (67). Bu algoritmaya göre, örneğin eğer B aleli için sinyal şiddeti 0'a yakinken A aleli için yüksek bir değere sahipse B alelinin bulunma durumu θ değeri 0'a yakındır ve B aleli oranı çok düşüktür. Bu nedenle, BAF değeri de 0'a yakın olacaktır. B aleli için alınan sinyal şiddeti yüksekken A aleli için alınan şiddet düşükse θ değeri 1'e yaklaşmakta ve bu nedenle hesaplanan BAF değeri de 1'e yakın olmaktadır. Her iki alele ait sinyal şiddetlerinin birbirine yakın olduğu durumda ise BAF 0,5 değerine yakın olacaktır. Farklı delesyon ve duplikasyon durumlarında oluşabilecek genotiplerin beklenen kopya sayıları ve BAF değerleri Tablo 2.1'de detaylandırılmıştır.

LRR değerleri ise, R (SNP total sinyal oranı) değerinden yapılan hesaplamalarla *Hidden Markov Model* kullanılarak hesaplanmaktadır (68). R değeri temel olarak aynı lokusa ait farklı alellerin sinyal şiddetlerinin toplamı şeklinde hesaplanmakta (Şekil 2.12) ve \log_2 oranı (LRR) aCGH'de olduğu gibi temelde $R_{\text{gözlenen}}/R_{\text{beklenen}}$ şeklinde ortaya konmaktadır (69).



Şekil 2.11. Bir kromozom boyunca Affymetrix SNP mikrodizin verisine ait LRR ve BAF grafikleri.

Mikrodizin verisine göre sağlıklı bir bireyde otozomlarda bulunması beklenen normal kopya sayısı olan 2 için kopya sayısını işaret eden LRR değerleri her genomik lokusta 0 civarında seyretmektedir. Genotip bilgisini gösteren BAF değerleri için ise sağlıklı bir bireyde 3 farklı genotipe sahip lokusların kromozom boyunca rastgele dağılması durumunda BAF değerleri 0 (AA, homozigot), 0,5 (AB, heterozigot) ve 1 (BB, homozigot) civarında dağılım göstermektedir.



Şekil 2.12. Illumina mikrodizin verisinde R ve θ değerleri.

Grafiğin x ekseninde θ değerleri, y eksenindeyse R değerleri yer almaktadır. Beklenen θ değerlerine göre örnek θ değerleri AA, AB ve BB olmak üzere gruplandırılmıştır. Henüz normalize edilmemiş bu değerler nihai BAF değerlerini oluşturmak için kullanılacaklardır. R değerleri ise örnek için hesaplanan SNP sinyal şiddetlerinin toplamıdır. R değerleri de daha sonra çeşitli algoritmalarla geçirilerek mikrodizin platformları tarafından normalize edilen LRR değerlerine dönüştürülmektedirler.

Tablo 2.1. Farklı CNV senaryolarına göre beklenen genotip, kopya sayısı ve BAF değerleri.

CNV	Genotip	Beklenen kopya sayısı (A+B)	Beklenen BAF değeri
Homozigot duplikasyon	AAAA	4	0
	AAAB	4	0.25
	AABB	4	0.5
	ABBB	4	0.75
	BBBB	4	1
Heterozigot duplikasyon	AAA	3	0
	AAB	3	0.33
	ABB	3	0.67
	BBB	3	1
Diploid (normal)	AA	2	0
	AB	2	0.5
	BB	2	1
Hemizigot delesyon	A	1	0
	B	1	1
Homozigot delesyon	-	0	-

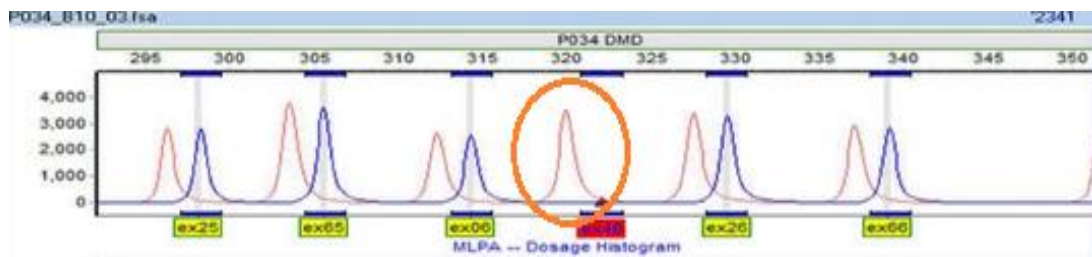
Günümüzde SNP mikrodizin yöntemini için en sık kullanılan çipler iki üretici tarafından ticari olarak tedarik edilmektedir: Affymetrix (ThermoFisher) ve Illumina. Bu iki firma temelde aynı metodolojiyi kullansa da farklı amaçlara yönelik farklı mikrodizin çipleri sunmaktadırlar (70). Her iki mikrodizin üreticisine ait farklı CNV mikrodizin çiplerine ait örnekler Tablo 2.2’de gösterilmiştir. Farklı mikrodizinlerin sunduğu çözünürlük ise tasarlanan problemlerin genomik dağılımına ve prob sayısına göre değişiklik göstermektedir.

Tablo 2.2. Affymetrix ve Illumina mikrodizin üreticilerine ait farklı SNP mikrodizin çiplerinin özellikleri.

Affymetrix Mikrodizinleri	CN Prob Sayısı	SNP Prob Sayısı
Genome wide Human SNP Array 6.0	946,000	906,600
CytoScan HD Array	2,696,550	743,304
CytoScan 750K Array	750,436	200,436
CytoScan Optima Array	18,018	148,450
Illumina Mikrodizinleri	CN Prob Sayısı	SNP Prob Sayısı
HumanCytoSNP-12 v2.1 BeadChip	11,459	289,773
Infinium CytoSNP-850K v1.2 BeadChip	-	844,855
Illumina Human660W-Quad v1	~100,000	550,000
Illumina Human1M-Duo	~60,000	1,100,000

2.2.4.4. Multipleks Ligasyon Bağımlı Prob Amplifikasyonu (MLPA, *Multiplex Ligation Dependent Probe Amplification*)

MLPA yöntemi, CNV tespiti için genom boyu olmayan ancak 20 civarında küçük lokus için kopya sayısının aynı anda tespit edilebildiği bir yöntemdir. Bu yöntem, bölge spesifik yarı-probların ligasyon ile tam prob haline getirildikten sonra polimeraz zincir reaksiyonu ile amplifikasyonu ve oluşan ürünlerin kapiller elektroforez ile incelenmesi esasına dayanmaktadır. MLPA ile test edilen bir örnekte PCR sonucu oluşan ürünlerin kontrol örneklerinde oluşan ürünlere kıyasla miktarlarındaki artış veya azalış test edilen örnekte hedeflenmiş genomik bölgelerde CNV bulunup bulunmadığını test etmek için kullanılmaktadır (Şekil 2.13). Bir ekzon kadar küçük olabilecek CNV'leri tespit edebilecek kadar hassas ve az maliyetli olması özellikle az sayıda genomik bölgenin hedefli olarak kopya sayısını incelemeyi gerektiren uygulamalarda MLPA'yı avantajlı kılmaktadır.



Şekil 2.13. MLPA yönteminin analiz çıktısı.

Şekilde pembe renkli çizginin yüksekliği kontrol örneğine ait ürün miktarını, mavi renkli çizginin yüksekliği ise test örneğine ait ürün miktarını göstermektedir. Buna göre, kırmızı yuvarlak ile işaretli ekzon 46 olarak belirtilen genomik bölgede test örneğine ait bir ürün bulunmaması burada homozigot delesyon olduğunu göstermektedir. Bu resim *Creative Commons Attribution-ShareAlike Licence 3.0* lisansı altında paylaşılan Wikipedia Özgür Ansiklopedi'nin web sayfasından alınmıştır (71).

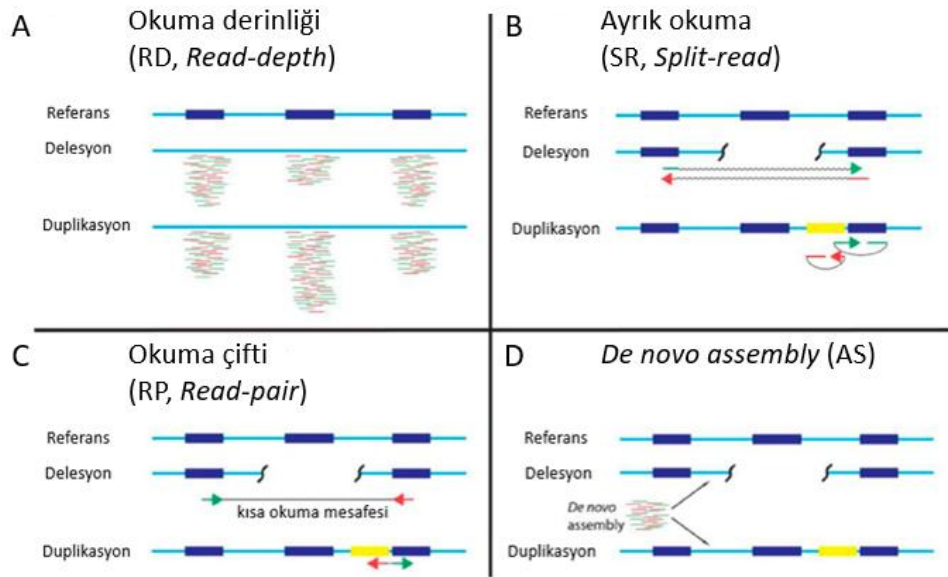
2.2.4.5. İleri Nesil Dizileme (*Next Generation Sequencing, NGS*)

NGS, öncelikle küçük varyantların tespiti için geliştirilmiş bir teknoloji olsa da birtakım algoritmalarla kopya sayısı değişikliklerinin saptanmasında da günümüzde popüler bir hale gelmiştir. NGS yöntemleri arasında tüm ekzom dizileme (*Whole Exome Sequencing, WES*), tüm genom dizilemeye (*Whole Genome Sequencing, WGS*)

göre CNV saptanması açısından daha kısıtlı bilgi sunmakta; ancak özellikle nadir hastalık arařtırmalarında düşük maliyeti nedeniyle tercih edilmektedir (72).

İleri nesil dizileme teknolojisinin git gide geliřtirilmesiyle yapısal varyasyonların saptanmasına dair ciddi bir ilerleme kaydedilmiřtir (73). Hem WES hem de WGS, milyonlarca kısa DNA fragmanının tek ya da çift yönlü olarak paralel dizilemesine dayanmaktadır DNA fragmanlarının dizilenebilmesi için uçlarına takılan adaptör diziler ile her bir fragman sabit bir zemin üzerinde amplifiye edilir ve dizileri ortaya çıkarılır. NGS ile elde edilen dizilerin genomdaki lokalizasyonu ise dizileme işleminin tamamlanması ardından her bir DNA fragmanına ait okumanın referans genomda nereye denk geldiğinin bulunması (hizalama, *alignment*) ile mümkün olur.

Referans genoma hizalanmış NGS verisinde, okumaların referans genoma kıyasla gerek oryantasyonu gerek sayısal olarak farklılık göstermesi yapısal varyasyonları ayırt etmek için kullanılabilir (74) . Genel itibariyle, yapısal varyasyonların tespiti okuma derinliğini (*read-depth*), iki yönlü okumalar için okuma çifti oryantasyon ve uzaklığını (*read-pair*) ve ayrık okumaları (*split-read*) temel alan yaklaşımlar veya *de novo assembly* adı verilen yöntemlerle okumalar kullanılarak genomun sıfırdan inşa edilmesi ile gerçekleştirilmektedir (Şekil 2.14).



Şekil 2.14. NGS ile yapısal varyant saptamak için kullanılan yöntemler.

NGS'den elde edilen farklı veri tipleri ile CNV'lerin saptanması mümkündür. A) *Read-depth* (Okuma derinliği, RD) verisi kullanılarak okumaların diğer genomik bölgelerden elde edilen okuma sayılarına kıyasla beklenenden az ya da çok olmasına göre şekilde ortada yer alan genomik bölge için üstte delesyona, altta ise duplikasyona örnek verilmiştir. B) *Split-read* (Ayrık okuma, SR) yaklaşımında referans genom ile yapısal varyantın sınırı üzerine denk gelen ve ancak farklı genomik bölgelere bölünerek referans genom ile hizalanabilen okumalar yapısal varyantı göstermektedir. C) Çift yönlü okuma yapılan NGS işlemlerinde *Read-pair* (Okuma çifti, RP) analizleri referansa kıyasla okumalar arası yön ve beklenen mesafelerden sapmalara göre yapısal varyantları gösterebilmektedir. D) *De novo assembly* (AS) yaklaşımı ise eşleşmemiş okumalar ile referans genoma hizalama işlemi kullanılmadan okumaların birbirine hizalanması ile bir genomik bölgenin baştan inşası ile yapısal varyantları tespit edebilmektedir. Bu şekil CC BY 4.0 lisansı altında açık erişimde paylaşılan bir makaleden alınmıştır (75).

i. Okuma derinliği

Genomdaki bir lokus birden fazla okuma tarafından temsil edilebilir. Okuma derinliği bir genomik bölgenin tamamı için söz konusu olan toplam okuma sayısını ifade etmektedir. Okuma derinliği yaklaşımı bir genomik bölgeye denk gelen okuma sayısının o bölgenin kopya sayısı ile ilişkili olduğu varsayımına göre çalışmaktadır (74). Buna göre, referans genoma hizalanan okumaların sayısındaki artışlar duplikasyonları, azalışlar ise delesyonları işaret etmektedir (Şekil 2.14.A).

ii. Ayrık Okumalar (*Split-reads*)

Ayrık okumaların incelenmesi ile kırık noktaları (*break points*) tayin edilebilmektedir. Örneğin, bir delesyonun tam sınırına denk gelen bir okuma delesyonun iki ucuna da hizalanmakta, dolayısıyla genomda iki farklı yere ayrılmış olarak hizalandıkları gözlenmektedir (Şekil 2.14.B). Hizalanmanın başlangıç ve bitiş noktaları gözlenebildiği için delesyonu doğuran kırığın pozisyonu tam olarak belirlenebilmektedir (76). Bu analiz, CNV'lerin boyutlarının tek nükleotit çözünürlüğünde sınırlarının belirlenmesi gerektiği durumlarda çok avantajlı olmaktadır.

iii. Okuma Çifti (*Read-pair*)

Çift yönlü okuma yapılan NGS işlemlerinde bir DNA fragmanının iki ucundan iki okuma elde edilir. Çift yönlü okumaların aynı DNA fragmanından köken alması sebebi ile bu okumalar arasında kalan dizi, dizileme işlemi ile ortaya konamasa bile okuma yönleri birbirine doğru ve iki okuma arasında mesafe belli bir aralıktadır (74). Okuma çifti yaklaşımına göre eşli okumaların referans genoma hizalanması sırasında okuma çiftleri arasında beklenen mesafeden veya oryantasyondan sapmaların incelenmesi ile yapısal varyasyonlar belirlenebilmektedir. Örneğin, referans genoma hizalandığında okuma çiftleri arasındaki mesafe beklenenden kısalmışsa delesyonu, tam tersine artmış ise yeni bir insersiyonu göstermektedir (Şekil 2.14.C).

Bu şekilde, okuma çifti yöntemi ile kopya sayısını değiştirmeyen yapısal varyantlar için de çıkarımlar yapılabilmektedir. Örneğin, duplikasyon ve inversiyonlar için, mesafenin yanısıra okuma çiftlerinin yönü de ters olmakta, translokasyon söz konusu olduğundaysa okuma çiftinin biri bir noktadan hizalanmışken diğerinin referans genomun bir başka bölgesinden hizalandığı gözlenmektedir (74).

iv. *De Novo Assembly*

Yeni sekans insersiyonlarına yönelik bilginin iyi bir şekilde edinilmesini sağlayan bir yaklaşımdır. Bu yöntem ile referans genomda hizalanma sırasında herhangi bir yere hizalanamamış okumalar bir referans genomdan faydalanmaksızın sıfırdan birbiri üzerine hizalanmaktadır. Bu yaklaşım sayesinde referans genomun

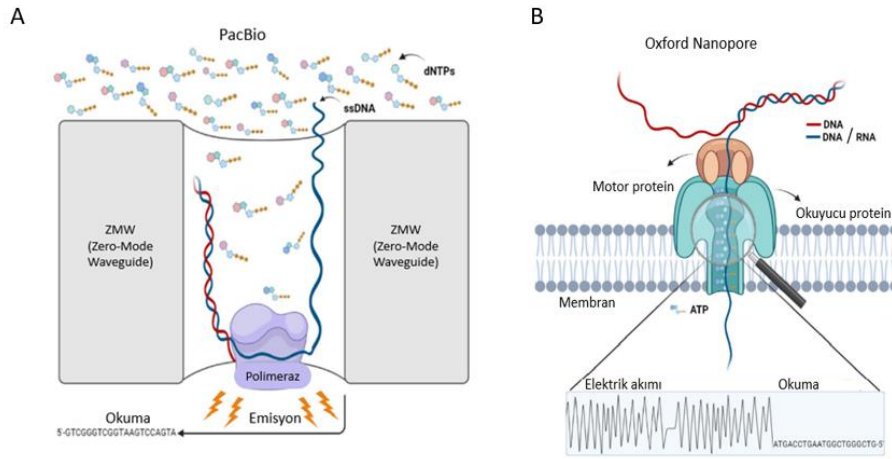
kendi doğasında yer alan sekans boşluklarından kaynaklı hizalama hataları aşıllp aslında referansta yer almayan yeni diziler veya kompleks değışiklikler de ortaya çıkarılabilmektedir (Şekil 2.14.D).

2.2.4.6. Üçüncü Nesil Dizileme

NGS teknolojisiyle kısa okumalar elde edilmesinin dezavantajını ortadan kaldıran Üçüncü nesil dizileme teknolojileri, daha uzun okumalar elde edilmesi sayesinde NGS teknolojisinin iyi tespit edemediđi uzun tekrar dizileri, büyük insersiyonlar ve kompleks yapısal değışiklikler gibi varyasyonların tespitine olanak sağlamaktadır. Günümüzde ticari olarak tedarik edilebilecek iki farklı üçüncü nesil dizileme platformu bulunmaktadır: PacBio ve Oxford Nanopore.

PacBio, SMRTSeq adlı bir yöntem ile ZMW (Zero-Mode Wavelength) teknolojisini kullanarak tek DNA molekülünün sekanslamasını gerçekleştirebilmektedir. DNA Polimeraz içeren bir “*flow cell*” içerisinde DNA’nın her bir nükleotidi polimeraz tarafından replike edilirken yeni sentezlenen DNA zincirine girmekte olan nükleotit üzerindeki florasan sinyalin okunması sayesinde dizileme gerçekleşmektedir (Şekil 2.15.A).

Oxford Nanopore teknolojisi de benzer bir şekilde tek iplikli tek bir DNA molekülünün bu kez küçük bir nanopor içerisinden geçirilerek kanal içerisinden geçmekte olan nükleotitlerin oluşturduđu spesifik elektriksel sinyallerin elektronik bir devre tarafından kaydedilmesi kanal içinden geçen DNA’nın dizisinin saptanması için kullanılmaktadır (Şekil 2.15.B). Üçüncü nesil dizileme yöntemleri amplifikasyon olmadan tek DNA molekülü üzerinden çalıştıkları için sıklıkla kısa dizileme hataları yapsalar da bu yöntem ile uzun okumalar yapılabilmesi sayesinde CNV’lerin tespiti için son derece avantajlıdır.

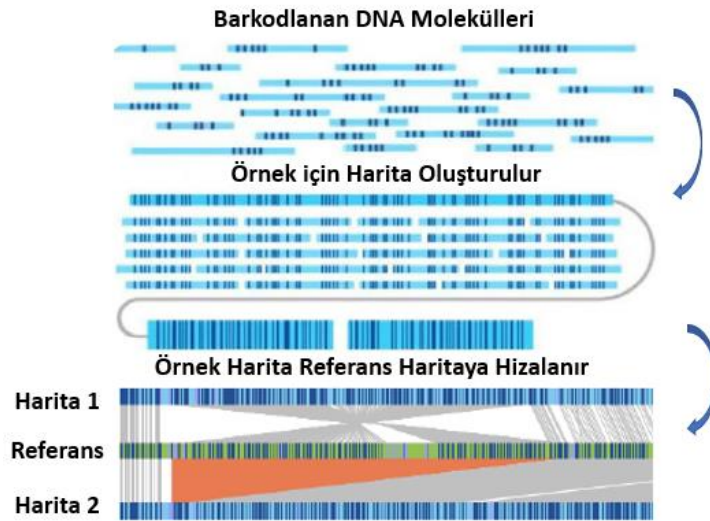


Şekil 2.15. Üçüncü nesil dizileme yaklaşımları.

A) PacBio SMRTseq dizileme yaklaşımı. Teknoloji ZMW adı verilen çok düşük ışık miktarını ölçmeye yarayan nanofotonik bir sisteme dayanmaktadır. DNA Polimeraz ile sentezlenmekte olan her bir florasan işaretli nükleotit için emisyon ölçülerek okumalara dönüştürülmektedir. B) Oxford *Nanopore* dizileme yöntemi. Bu teknoloji, bir membran üzerinde yer alan protein yapısındaki nanopor içerisinden DNA'nın geçirilmesi sırasında kaydedilen elektrik sinyallerinin okumalara dönüştürülmesine dayanmaktadır. Bu şekil CC BY 4.0 lisansı altında açık erişimde paylaşılan bir makaleden türetilmiştir (77).

2.2.4.7. Optik Genom Haritalama

Son yıllarda ortaya çıkan optik genom haritalama teknolojisi, CNV ve diğer yapısal varyasyonların tespitinde çok daha keskin bir yaklaşım sunmaktadır. NGS teknolojisinden farklı olarak, DNA'nın küçük parçalar halinde fragmantasyonu yerine uzun tek molekül DNA fragmanlarının florasan işaretlerle barkodlanıp bu barkodlanmış DNA'ların nano kanallardan geçirilirken fotoğrafları çekilmektedir. Bu şekilde elde edilen DNA barkodlarının birbiri üzerine hizalanması ile genomun sıfırdan inşa edilmesi prensibine dayandığı için NGS'nin kısa okumalar nedeniyle yapısal varyantları saptamakta yaşadığı sorunları ortadan kaldırmaktadır. DNA barkodlama teknolojisi ile uzun okuma yaklaşımını birleştiren bu yöntem sayesinde tüm yapısal varyantlar saptanabilmektedir (Şekil 2.16.).



Şekil 2.16. Optik genom haritalama yaklaşımı.

Optik genom haritalamanın analiz basamakları yukarıdan aşağı sırasıyla gösterilmiştir. Üst panelde barkodlanan tek tek DNA molekülleri gösterilmektedir. Bir sonraki basamakta bu barkodlanmış parçalar birbiri üzerine hizalanarak genom sıfırdan inşa edilmektedir. Sonrasında örnek için oluşturulan harita, referans genom haritasına hizalanarak moleküller arası uyumsuzluklar saptanmakta ve yapısal varyantlar belirlenmektedir. Genellikle ikinci bir harita confirmasyon için kullanılmaktadır. Bu şekil CC BY 4.0 lisansı altında açık erişimde paylaşılan iki farklı makaleden türetilmiştir (78, 79).

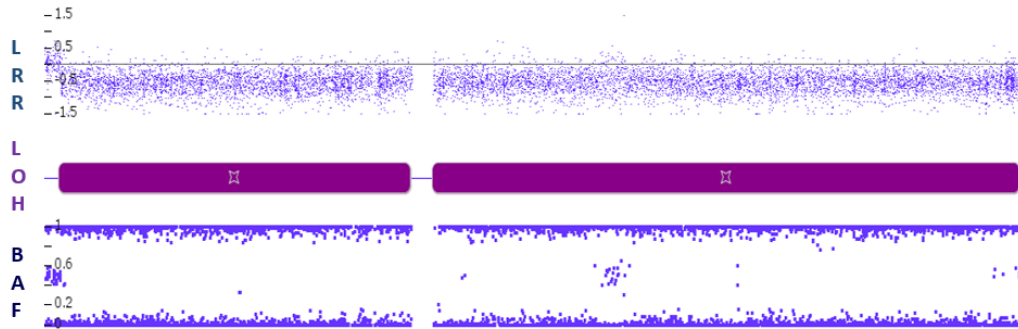
2.3. Heterozigotluk Kaybı (*LOH, Loss of Heterozygosity*)

SNP mikrodizin yöntemlerinin genotip ile ilgili bilgi vermesi sebebiyle bu yöntem ile CNV'ler araştırılırken bir yandan da genotiplerin genom boyu dağılımlarını ve bu dağılımdaki sorunları kolaylıkla gözlemlemek mümkün hale gelmiştir. Heterozigotluk kaybı (LOH) SNP mikrodizin verilerinde genotip dağılımında sıklıkla gözlenen bir değişikliktir. Adından da anlaşılacağı üzere heterozigot kaybında uzun bir genomik bölge için bölgenin uzunluğundan beklenen aksine hiçbir heterozigot genotip ile karşılaşılması durumu görülmektedir (Şekil 2.17).

LOH bölgelerini oluşturan sebeplerden biri, iki alelden birinin delesyonudur. Bu durumda LOH bölgesinde kopya sayısı da azalmış olacaktır. Bu şekilde ortaya çıkan LOH bölgeleri birçok kalıtsal hastalık ve kanserle ilişkilendirilmektedir (80). Kopya sayılarının normal olduğu ancak LOH bölgelerinin ortaya çıkabildiği durumlar da

mevcuttur (81). Bu durumlardan biri, bir ebeveynden iki kromozomun alındığı uniparenteral dizomi (UPD) durumu olup bu konu Bölüm 2.3.2'de detaylandırılmıştır.

Kopya nötral LOH bölgelerinin ortaya çıktığı bir başka durum ise bir bireyin iki farklı ebeveyninden aldığı genomik bölgelerin ebeveynlerin daha üst nesillerinde ortak bir atalarından kaynaklanmasıdır. Bu durumda ebeveynlere ortak atadan kalıtılan aynı haplotipe sahip alellerin bir çocukta buluşması sonucu ebeveynlerinden aynı aleli alan çocukta söz konusu genomik bölge için heterozigot varyant bulunmamaktadır. Kalıtım nedeniyle eş olma (IBD, *Identity By Decent*) adı da verilen bu durum, ebeveynlerin yakın ortak atalarının bulunduğu akraba evliliklerinde veya kapalı toplumlar içinde evliliklerde sık karşılaşılan bir durumdur (82).



Şekil 2.17. Affymetrix SNP mikrodizin verisinde LOH.

Affymetrix SNP mikrodizin verisinde erkek cinsiyetinde bir bireyin X kromozomuna ait üst panel LRR alt panelde ise BAF değerleri görülmektedir. LRR değerleri kromozomun her iki ucundaki PAR bölgeleri dışında tek kopya delesyonu işaret edecek şekilde düşük olup X kromozomu boyu BAF değerlerinde ise heterozigot genotipi gösteren 0,5 değerine sahip BAF neredeyse hiç bulunmamaktadır. Bu durum erkeklerde X kromozomunun doğal olarak bir kopya olmasından kaynaklanmakta ve genomda tek kopya olan bölgelerde LOH ortaya çıktığını göstermektedir. Orta panelde ise PAR'lar arasında kalan genomik bölgenin Affymetrix mikrodizin analiz programı olan *Chromosome Analysis Suite* tarafından bir LOH bölgesi olarak tanıdığı yatay mor çubukla gösterilmektedir.

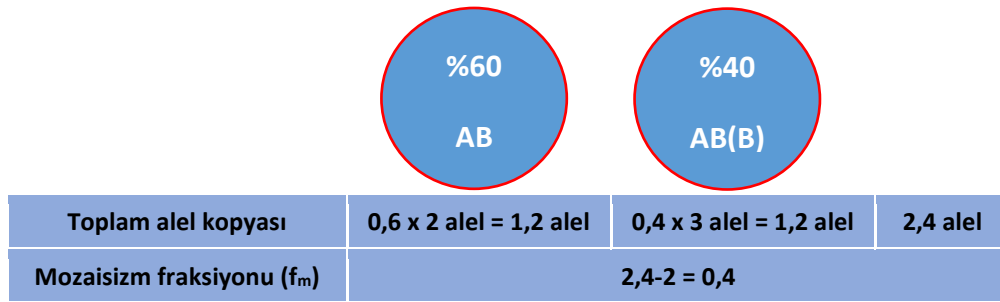
2.4. Mozaisizm

İnsanlar tek bir hücreden köken alan canlılardır. Zigot adı verilen bu tek hücrenin genetik yapısı mitoz bölünmeler zigottan kaynaklanan tüm yavru hücrelere

aynen kopyalanmakta olduğundan bir insanın tüm hücrelerinin genetik yapısı teorik olarak aynıdır. Ancak zigot aşamasından itibaren hücrelerde ortaya çıkan genetik hatalar bu teorik beklentinin gerçekleşmemesine ve insanlarda bazı hücre gruplarının zigottaki orijinal genetik yapıdan farklı varyantlar barındırmasına neden olurlar. Bu varyantlara mozaik genetik varyantlar adı verilmekte olup böyle varyantların bulunması durumu ise mozaisizm olarak adlandırılır. Mozaik genetik varyantlar tek nükleotit değişikliklerinden kompleks düzenlenmelere kadar çeşitli tiplerde ve boyut olarak ise geniş bir genomik skalada yer almaktadır.

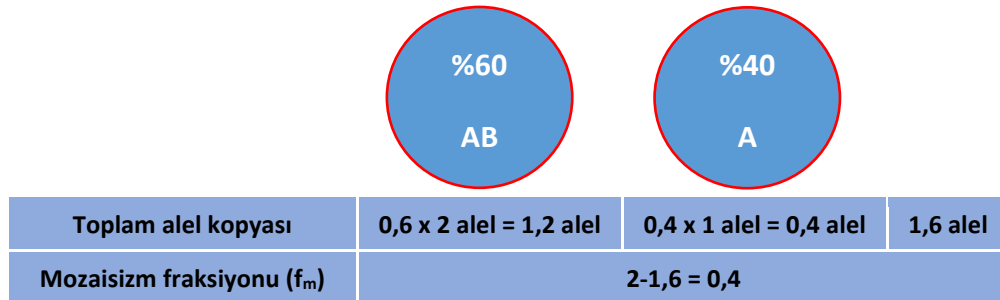
Mozaik varyantlar, elbette ki CNV'leri de kapsamaktadır. CNV'ler tek tek hücreler içerisinde tamsayı olmak durumundadır. Ancak tek tek hücreler içerisinde tamsayı olan değişiklikler, bir hücre popülasyonu incelendiğinde popülasyondaki tüm hücrelerin içinde olabileceği gibi sadece bir kısım hücreyi etkileyen genetik değişiklikler şeklinde de olabilir. Buna göre, farklı sayıda kopya sayılarına sahip bir hücrelerin karışımı ile oluşan bir hücre popülasyonundan bahsedildiğinde popülasyon için kopya sayısı değeri tamsayı olmayan değerler alabilir. Bu durum genellikle kopya sayısının normal değeri olan 2 şeklinde bulunan ve 2'den farklı kopya sayısına sahip hücrelerin karışımı şeklinde gözlenir. Bu şekilde, CNV'lerin sadece bir grup hücre içerisinde yer alıp bu hücrelerin genetik açıdan normal hücreler ile karışık olduğu duruma genetik mozaisizm, mozaisizmi oluşturan CNV'ler ise mozaik CNV (mCNV) olarak adlandırılmaktadır (2).

Mozaik CNV'ler, tipine göre delesyon ya da duplikasyon mozaisizmi şeklinde olabilmektedir. Mozaik CNV'ler sonucu oluşan genetik mozaisizmde, CNV'ye sahip olan hücrelerin tüm hücreler içerisindeki oranı, mozaisizm fraksiyonunu (f_m) olarak adlandırılmaktadır. Popülasyona atfedilen kopya sayısı, f_m değerini yansıtan bir değer olup niteliği bilinen bir CNV söz konusu olduğunda popülasyonun ortalama kopya sayısı kullanılarak f_m hesaplanabilir. Mozaisizm fraksiyonunun delesyon ve duplikasyon mozaisizmi senaryoları için birer örnek hesaplamaları Şekil 2.18 ve 2.19'da gösterilmektedir.



Şekil 2.18. Mozaisizm fraksiyonu bilinen duplikasyon mozaisizminin kopya sayısının hesaplanması.

Yukarıda; tüm hücre popülasyonunda 2 kopyaya sahip normal hücrelerin oranının 0,6 (%60), duplikasyona sahip hücrelerin oranının ise 0,4 (%40) bir senaryoya örnek verilmektedir. Buna göre, anormal hücrelerin oranını gösteren mozaisizm fraksiyonu (f_m) 0,4'tür. Duplikasyon mozaisizminin popülasyonun toplam kopya sayısına katkısı hücrelerin oranı ile 3'ün çarpımı iken normal kopya sayısındaki hücrelerin katkısı 2 ile çarpımıdır. Bu şekilde yapılan hesaplamada tüm popülasyon için kopya sayısı 2,4 olarak hesaplanmakta olup, bu sayıdan diploid kopya sayısı olan 2 çıkarılarak f_m bulunabilir.



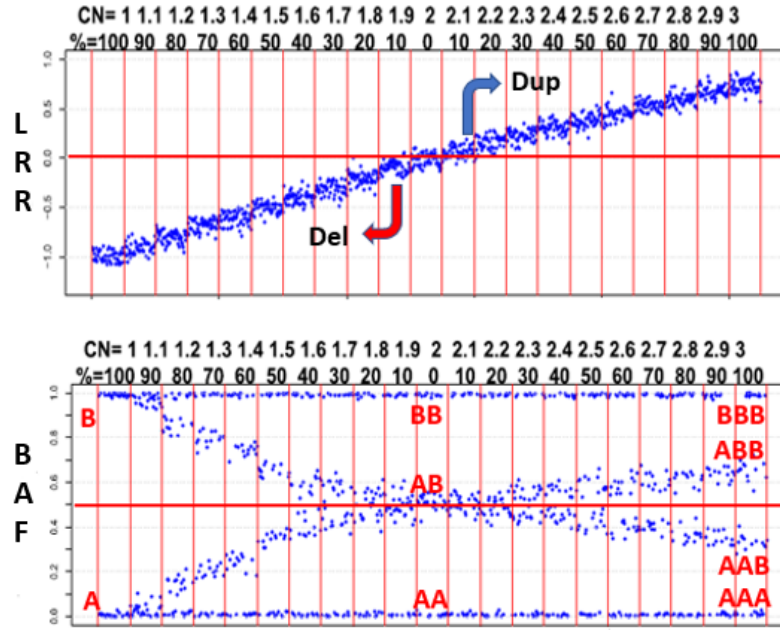
Şekil 2.19. Mozaisizm fraksiyonu bilinen delesyon mozaisizminin kopya sayısının hesaplanması.

Yukarıda; tüm hücre popülasyonunda 2 kopyaya sahip normal hücrelerin oranının 0,6 (%60), delesyona sahip hücrelerin oranının ise 0,4 (%40) bir senaryoya örnek verilmektedir. Buna göre, anormal hücrelerin oranını gösteren mozaisizm fraksiyonu (f_m) 0,4'tür. Delesyon mozaisizminin popülasyonun toplam kopya sayısına katkısı hücrelerin oranı ile 1'in çarpımı iken normal kopya sayısındaki hücrelerin katkısı 2 ile çarpımıdır. Bu şekilde yapılan hesaplamada tüm popülasyon için kopya sayısı 1,6 olarak hesaplanmakta olup, diploid kopya sayısı olan 2'den çıkarılarak f_m bulunabilir.

Mozaik CNV'lerin en sık görüldüğü insan hastalığı somatik dokularda edinsel mutasyonlar sonucu meydana gelen kanserdir. Buna ek olarak, nadir konjenital genetik hastalıklarda da mozaik CNV'ler görülebilmektedir. Bu nedenle, mozaik

CNV'lerin saptanması bu hastalıkların tanısında ve özellikle kanser gibi ilerleyici hastalıkların tanı ve takibinde oldukça önemlidir (36). Ayrıca, mozaik CNV'lerin saptanması kadar CNV içeren hücre oranının göstergesi olan f_m hesaplanması da hastalığın ilerleme veya gerilemesini araştırmak için önemli bir göstergedir (83). Örneğin, mozaik Down Sendromunda bukkal mukozadaki mozaik oranı ile hastalık bulgularının şiddetinin ilişkili olduğu gösterilmiştir (84). Ayrıca, kanser gibi hastalıklarda f_m hastalığın şiddeti, prognozun belirlenmesi, tedaviye yanıtın değerlendirilmesi gibi amaçlarla kullanılmaktadır (83, 85).

Günümüze kadar güvenilirlikle canlıların bütün hücrelerinde olan CNV'lerin saptanması için kullanılmış SNP mikrodizin yöntemi, mozaik CNV'lerin saptanması için de yine BAF ve LRR verilerini kullanarak mozaik CNV'lerin hangi genomik bölgeyi ilgilendirdiğini tespit etme imkanı sunmaktadır (Şekil 2.20)(86). Ancak, bu veriler mCNV türünü ve lokasyonunu gösterse de bu verilerden f_m değerinin hesaplanabilmesi için farklı yaklaşımlar gerekmektedir. Örneğin kopya sayıları $\log_2(\text{gözlenen/beklenen})$ formülüyle LRR değerlerinden de hesaplanabilmektedir. Bu hesaba göre, gözlenen kopya sayısı tam duplikasyona karşılık gelen 3 ise, LRR değeri $\log_2(3/2)=0,585$ şeklinde hesaplanmaktadır. Kopya sayısının bilinmediği ancak LRR değerinin bilindiği durumda ise $2^{2 \cdot \text{LRR}}$ formülü uygulanarak bu işlemin tersi yapılmakta ve $\text{LRR} = 0,585$ değeri kullanıldığında $2^{2 \cdot 0,585}$ ile kopya sayısı 3 olarak bulunmaktadır.



Şekil 2.20. Mozaik CNV'lerin SNP mikrodizin verisinde görünümü.

Şeklin üst tarafında kopya sayısının (CN) azalış veya artış şeklindeki yönünü gösteren LRR değerleri yer almaktadır. mCNV'ler için 0 değerinin üzeri duplikasyonları, altındaki değerler delesyonları işaret etmektedir. CN değeri mozaik olduğunda LRR değerinin tam sayılı kopya sayılarını işaret eden -1,0, 0 ve 0,585 değerlerinden farklı olduğuna dikkat ediniz. Buna göre, LRR değeri -1,0 ile 0 arasında yer alan değerlerin farklı f_m değerlerini işaret eden mozaik delesyonları, 0 ile 0,585 arasında yer alan değerlerin farklı f_m değerlerini işaret eden mozaik duplikasyonları işaret etmektedir. Alt kısımda yer alan grafikte ise farklı f_m değerlerine sahip delesyon ve duplikasyon olaylarına karşılık gelen BAF dağılımları görülmektedir. 2 kopya sayısında heterozigot genotipe sahip genomik lokuslar için BAF değeri 0,5 iken mozaik delesyon ve duplikasyon bulunması durumunda heterozigot genotipler için BAF değerinin 0,5'ten saptığına ve f_m değeri arttıkça 0,5'ten sapmanın her iki yönde gittikçe arttığına dikkat ediniz.

Mozaisizmin fenotipik etkileri ve kalıtlılabirliği, mozaikliğin görüldüğü hücre tipine, değişikliğin lokasyonuna ve etkilenen hücre sayısına bağlı olmaktadır. Kalıtlılabirlik açısından incelendiğinde mozaisizm; somatik mozaisizm, *germline* mozaisizm ve somatik-*germline* mozaisizm şeklinde sınıflandırılmaktadır (87).

Somatik mozaisizm, mozaisizmin meydana geldiği hücre popülasyonunun sadece vücut hücrelerini içerdiği, *germ* hücrelerini içermediği mozaisizm tipidir. Mozaisizmin bu tipinde değişikliğin meydana geldiği embriyonik gelişim dönemi

önemli olmaktadır. Buna göre somatik mozaizm, prenatal ve postnatal dönem mozaizm olarak kabaca ikiye ayrılabilir. Her iki dönemde de ortaya çıkan mozaizm sadece somatik hücreleri etkilemiş olduğundan kalıtsal olarak aktarılmamaktadır (87).

Prenatal somatik mozaizm, post-zigotik dönemde (zigot oluşumundan sonraki embriyonik evreler) ve doğum öncesinde, genellikle embriyonik gelişim sırasında meydana gelmiş bir mozaizmi işaret etmektedir. Bu tür somatik mozaizm sonucu genellikle ortak bir embriyonik yapıdan köken alan dokular veya hücreler topluca etkilenirken diğer doku ve hücreler zigottaki genomik yapılarını korurlar. Bu türden mozaizmlerin fenotipe yansması genellikle etki alanına bağlı olmaktadır. Örneğin mozaik trizomi 13'te, trizomi 13 bulunan ve bulunmayan hücrelerin farklı cilt renklerine sebep olması sonucu embriyonik gelişim şemasını takip eden cilt lekeleri oluşabilmektedir (88).

Postnatal mozaizmde ise, yine post-zigotik olarak mozaizme sebep olan bir somatik değişiklik, bireyin doğumundan sonraki herhangi bir gelişim evresinde ortaya çıkmıştır. Bu tip mozaizmin en iyi örneği kanserdir (4).

Plasental mozaizm, özel bir tip somatik mozaizm olup sadece plasenta dokusuna özgü mozaizmi ifade etmektedir. Bu tip mozaizmde mozaik değişiklikten etkilenmiş hücreler yalnızca ekstra-embriyonik dokularda bulunduğundan dolayı, embriyonun kendisi bu değişikliği taşımamaktadır. Gelişmekte olan bebeğin plasenta aracılığıyla beslenmesinde bir sorun yaratmadığı sürece plasentaya özgü mozaizmlerin birey üzerinde bir fenotipik etkisi bulunmamaktadır (89).

Germline mozaizm, mozaik bir genetik değişiklikten etkilenerek farklılık gösteren hücre popülasyonunun sadece germ (üreme) hücrelerine özgü olduğu mozaizm tipidir. Mozaik CNV'lerin germ hücrelerinde olması, bu değişikliğin bir sonraki nesle de aktarılabilmesine imkan tanımaktadır. Ancak *germline* mozaik CNV'ler bir sonraki nesle aktarırlarsa bu yavrunun zigot aşamasında bulunmaları sebebiyle tüm hücrelerinde söz konusu CNV görülmekte ve mozaizm ortaya çıkmamaktadır. Üstelik değişiklik ebeveynin somatik hücrelerinde olmadığından

ebeveynde fenotipik bir farklılık çoğu zaman ortaya çıkmamakta ve ebeveynde *germline* mozaizm olduğunu anlamak çok mümkün olmamaktadır (90).

Somatik ve *germline* mozaizm, embriyonik erken gelişim döneminde meydana gelen mozaik varyasyonların embriyoda hem somatik hem de germ hücreleri etkilemesi durumudur. Germ hücreleri etkilenmiş olduğundan mozaik değişiklik bir sonraki nesle de aktarılabilmektedir. Somatik hücrelerin etkilenmiş olması sebebiyle bu tür mozaizm varlığında bireyin kendisi, mozaik değişikliğin fenotipik sonuçlarını taşıyabilmektedir. Ancak mozaizm taşıyan birey bu değişikliği çocuğuna aktardığında çocukta non-mozaik bir değişiklik bulunacağı için değişikliğin fenotipik etkisi daha ağır görülebilmektedir (91). Mozaizmin fenotipik etkileri, etkilenen bölgeye ve hücre sayısına bağlı olsa da çoğu hastalıklar için mozaik formlar daha hafif fenotipik etkiye sahiptir ve mozaik varyantlar varlığında da artan mozaizm daha ağır fenotiple sonuçlanabilmektedir (92).

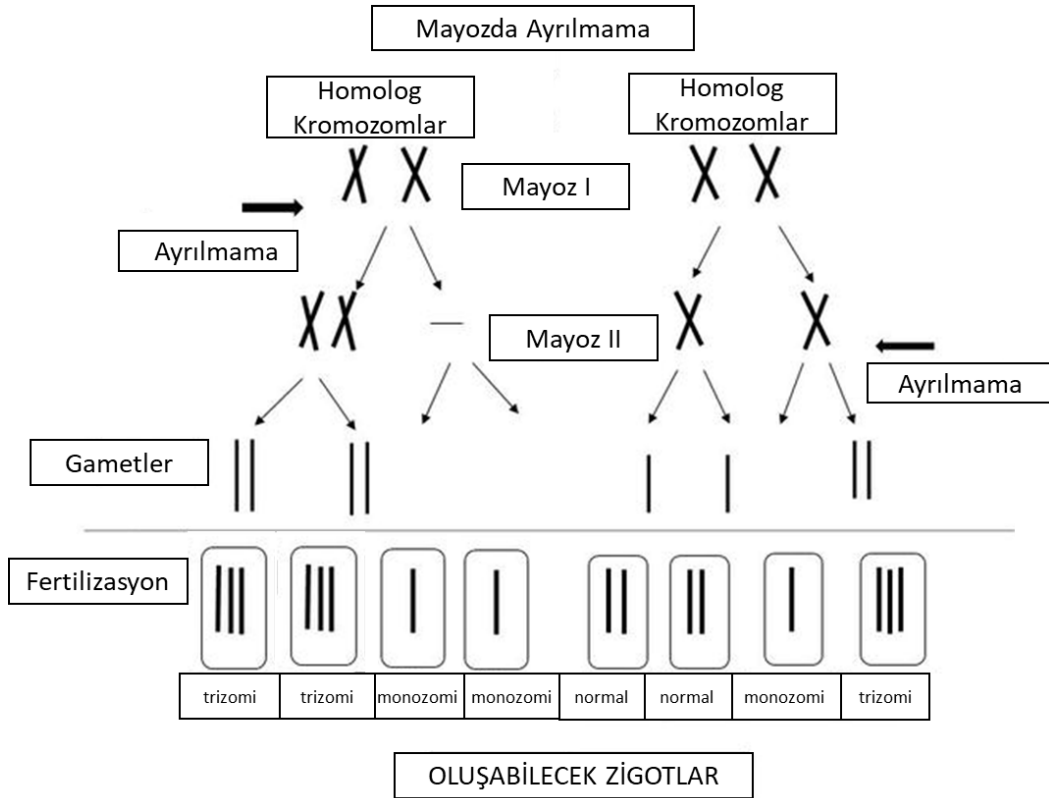
2.4.1. Mozaizmin Oluşum Mekanizmaları

Mozaik CNV'lerin oluşumunda mitotik veya mayotik ayrılmama olayları rol oynayabilmektedir. Ayrılmama sonucu oluşan gamet ile normal genetik yapıya sahip bir gametin birleşmesi sonucu anöploidiye (diploid yapıdan farklı kromozom sayısı) sahip zigotlar ortaya çıkabilmektedir. Mozaizm bu anöploidik zigottan oluşan bazı hücreler içerisinde genetik hatanın düzelmesi sonucu ortaya çıkabileceği gibi normal bir zigotun embriyonik gelişim evresindeki bölünmeler sırasında genetik bir hata yapması sonucu da ortaya çıkabilir (89). Bu iki farklı tür mozaizminin ayırt edilebilmesi için mayoz ve döllenme işlemlerini daha yakından incelemek gerekmektedir.

Germ hücrelerinde meydana gelen ve yavru gametlerin oluşumunu sağlayan mayoz bölünmenin mayoz I ve mayoz II olmak üzere iki aşaması bulunmaktadır. Mayotik ayrılmama, mayozun bu iki aşamasından birinde gerçekleşen, kromozomların olması gerektiği şekilde ayrılmaması sonucu yavru hücrelere eşit sayıda dağılmamalarıyla meydana gelen bir hata olayıdır. Hem mayoz I hem mayoz II için ayrılmamanın anne ya da baba kaynaklı olması mümkündür (93). Mayoz I

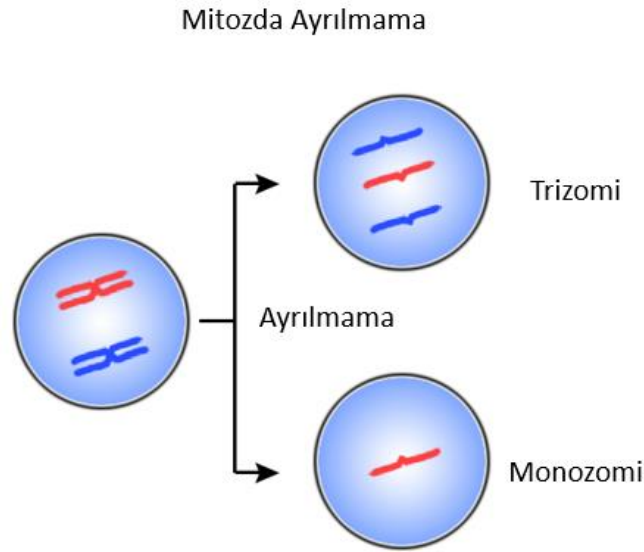
aşamasında ayrılmama söz konusu olduğunda, eşlenmiş homolog kromozomların anafaz I evresinde farklı kutuplara çekilememesi nedeniyle oluşan iki yavru hücreden birinde fazla, diğesinde eksik kromozom bulunmasına yol açmaktadır. Ardından kromozomların mayoz II evresine geçişiyle, ikinci bölünmeden sonra oluşan 4 yavru gametten de ikisi fazla, ikisi eksik kromozom taşımaktadır. Oluşan bu gametlerin ise normal haploid kromozomlu başka bir gametle birleşmesi sonucunda zigot oluşumuna hangi hücrenin katkı sunduğuna bağlı olarak trizomi veya monozomi taşıyan bir zigot meydana gelmektedir (Şekil 2.21). Diğer yandan; mayoz I'de hata olmadan, mayoz II sırasında ayrılmama olayı meydana geldiğinde, kardeş kromatidlerin ayrılmaması sonucunda 4 hücreden ikisi normal olmakta ve diğer ikisinden biri fazla, diğeri ise eksik kromozomlu olmaktadır. Aynı şekilde bu gametlerin de normal bir gamet ile birleşmesi durumunda, öploid veya anöploid barındıran zigotların oluşması mümkündür (Bkz. Şekil 2.21).

Mitotik ayrılmama ise, döllenme sonrası somatik olarak zigot ve zigottan köken alan yavru hücrelerin geçirdiği bölünmeler sırasında meydana gelebilmektedir. Bu durumda da kardeş kromatidler ayrılamaz ve oluşan 2 hücreden biri trizomi diğeri ise monozomi barındırır (Şekil 2.22). Bozuk yapıya sahip bu hücrelerden bir grubu ölere kaybolabilmekte; ancak diğeri bir grubu yaşamakta ve hücreler bölünmeye devam etmekte olduğunda normal hücrelerle birlikte CNV taşıyan hücreler bir arada mozaik bir şekilde bulunabilmektedirler. Mayoz sürecinde ayrılmama sonucu ortaya çıkan bir trizomi durumu da zigot aşamasında veya sonraki embriyonik gelişim döneminde mitotik ayrılmama ile düzeltilebilir. Trizomiden kurtulma adı verilen bu durum da gene mozaizme sebep olabilmektedir (94).



Şekil 2.21. Mayoz I ve Mayoz II aşamalarında ayrılmama.

Şeklin sol tarafında Mayoz I’de ayrılmama olayının sonuçları gösterilmektedir. Mayoz I’de ayrılmama sonucunda 2 kopya içeren 2 gamet oluştuğuna ve normal bir Mayoz II’nin ardından 2 adet trizomik ve 2 adet monozomik hücrenin oluştuğuna dikkat ediniz. Şeklin sağ tarafında ise Mayoz I normalken, Mayoz II’de iki yavru hücreden birinde ayrılmama olduğu durum gösterilmektedir. Bu tür bir ayrılmama ardından fertilizasyon sonucunda 2 adet normal, 1 adet trizomik ve 1 adet monozomik zigot oluşabildiğine dikkat ediniz. Bu şekil “Creative Commons Attribution 3.0 License” altında paylaşılan kitap bölümünden alınarak düzenlenmiştir (95).



Şekil 2.22. Mitozda ayrılmama.

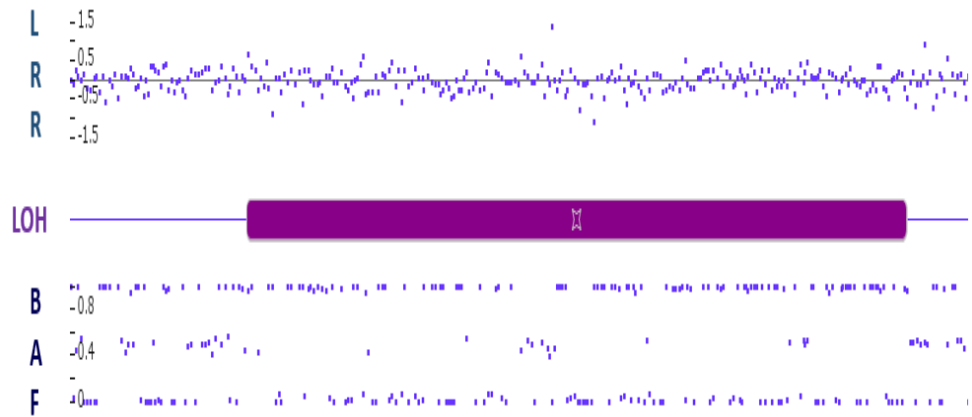
Şekilde mitoz bölünme sırasında kardeş kromatidlerin ayrılmaması sonucunda bir yavru hücrenin trizomik (sağ üst) diğer hücrenin ise monozomik (sağ alt) olduğuna dikkat ediniz. Bu şekil "*Wikimedia Commons, the free media repository*"nin *Creative Commons CCO* lisansı altında paylaştığı şekilden türetilmiştir (96).

2.5. Uniparental Dizomi (UPD)

İnsanlarda homolog kromozom setinden biri anneden diğeri de babadan aktarılmaktadır. Ancak UPD durumu söz konusu olduğunda, biri anneden ve biri babadan aktarılması gereken kromozomların her ikisi de tek bir ebeveynden köken almaktadır. UPD'ler mayotik ayrılma hatalarından köken alan monozomi ve trizomiden kurtulma, post-zigotik hatalar, somatik rekombinasyon veya gen dönüşümü gibi birçok mekanizmadan kaynaklanabilmektedirler. UPD olayında kopya sayısında değişiklik gözlenmemekte olup aynı ebeveynden köken alan 2 kromozom birbiri ile aynı haplotipe sahip olabilmektedir. Bu şekilde ortaya çıkan heterozigotluk kaybı, CN-LOH (*Copy Neutral Loss of Heterozygosity*, kopya nötral heterozigotluk kaybı) olarak da isimlendirilmektedirler (Şekil 2.23). CN-LOH bölgeleri UPD bölgelerinin bir özelliği olmasının yanında IBD bölgelerinin de önemli bir özelliğidir.

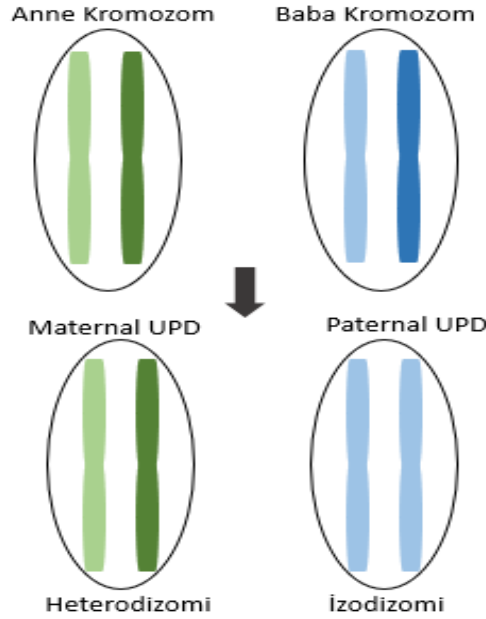
UPD'ler ebeveyn kökeni anne ise maternal UPD, baba ise paternal UPD olarak adlandırılmaktadır. Farklı ebeveyn kökenlerinin yanında UPD olan genomik bölgenin

içerik bakımından ise ebeveyndeki aynı homolog kromozomdan köken alması izodizomi veya ebeveyndeki farklı homolog kromozomlardan köken alması ise heterodizomi olarak sınıflanmaktadır (Şekil 2.24). Kromozomlar mayoz aşamasında *cross-over* yaptıkları için mayotik ayrılmama sonucu ortaya çıkan UPD'ler çoğunlukla hem ebeveyndeki homolog kromozomdan hem de kardeş kromatitten parçalar içerir ve homodizomi ile heterodizominin karışımı bir görüntü sergilerler.



Şekil 2.23. Uniparental dizomi (UPD).

Yukarıdaki grafikte kromozom boyunca LRR değerleri yer almakta ve bu değerlerin 0 olması normal kopya sayısı olan 2'yi işaret etmektedir. Buna karşın, şekil altında kromozom üzerinde BAF değerlerinin dağılımı yer almakta BAF değerleri üzerinde uzun mor çubuk ile işaretlenmiş LOH bölgesinde heterozigot BAF değerlerinin oldukça seyrek olduğu görülmektedir. Bu durum CN-LOH olarak adlandırılmaktadır.



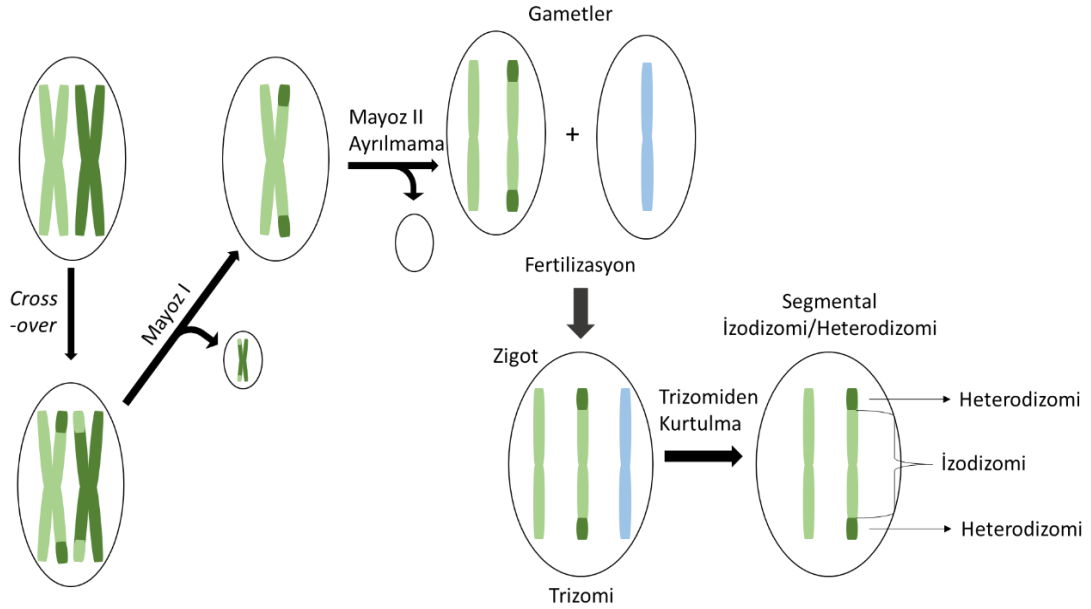
Şekil 2.24. Farklı UPD tipleri.

Şeklin sol tarafında anne kökenli UPD gösterilmektedir. Bireyin her iki kromozomu da anneden aktarıldığı için maternal UPD, iki kromozom annenin farklı homolog kromozomlarından köken aldığı için ise heterodizomi olarak adlandırılmaktadır. Sağ tarafta ise baba kökenli UPD söz konusu olup bu durum paternal UPD ve babanın bir homolog kromozomunun iki kopyası yavrusuna aktarıldığı için izodizomi olarak adlandırılmaktadır. Her 2 UPD senaryosu anne veya baba kökenli olabilmektedir.

İzodizomi, mayoz II veya mitotik hatalardan kaynaklanabilmekte iken heterodizomi ise, Mayoz I'de ayrılmama sonucu ortaya çıkmaktadır (97). Mayotik hatalar sonucu UPD oluşabilmesi için bir gametten 2 homolog kromozom (dizomik) gelirken diğerinden o homolog kromozomdan hiç (nullizomik) gelmemesi gerekmektedir. Bu durum, "gamet tamamlama" (*gamete complementation*) olarak adlandırılır. Öte yandan hem izodizomi hem de heterodizomi ayrılma hataları sonucu oluşan anöploidik bir zigotun kurtulmasıyla da meydana gelebilmektedir. Örneğin, monozomiden kurtulma ile tam izodizomi oluşurken trizomiden kurtulma ile ise her iki tipin birlikte bulunduğu segmental izodizomi/heterodizomi oluşmaktadır (Şekil 2.25).

Trizomiden kurtulma, ebeveynlerin birinden kaynaklı mayoz I veya mayoz II'de gerçekleşen, kromozomların olması gerektiği şekilde ayrılmadığı durumlarda zigotta

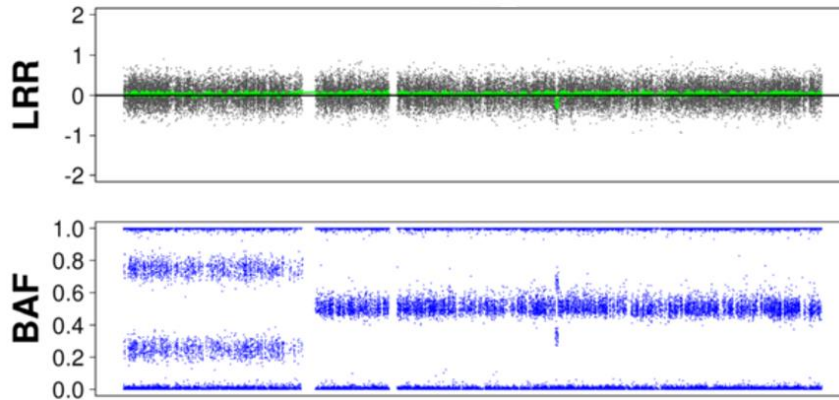
ortaya çıkan trizominin, sonrasında takip eden mitotik olaylarla düzeltilmesidir (94). Buna göre Mayoz I'de gerçekleşen *cross-over* sonucu meydana gelen ayrılmamanın etkisiyle her iki UPD tipi de kromozom üzerinde kısmi olarak görülmektedir (Bkz. Şekil 2.25).



Şekil 2.25. Trizomiden kurtulma ile oluşan segmental izodizomi/heterodizomi.

Mayoz bölünme sırasında gerçekleşen *cross-over* sonrasında olaysız gerçekleşen Mayoz I'in ardından Mayoz II'de ayrılmama olmasının ardından fertilizasyon sonrası oluşan trizomik zigotta trizomiden kurtulma ile oluşan segmental izodizomi/heterodizomi UPD görülmektedir.

Mayotik ayrılmama sonucu oluşan trizomiden kurtulma, mitotik olaylar neticesinde post-zigotik aşamalarda gerçekleştiğinde ise, mozaik UPD'ler ortaya çıkabilmektedir (98) . Bu durumda trizomiden kurtulma sonucunda bazı hücrelerin dizomik, bazılarının trizomik, bazılarının ise UPD barındırması mümkün olacaktır. SNP mikrodizin analizlerinde bu tür mozaik UPD'ler de saptanabilmektedir (93). Mozaik UPD söz konusu olduğunda mikrodizin verisindeki LRR ve BAF değerleri Şekil 2.26'da görülmektedir.



Şekil 2.26. SNP mikrodizin verisinden elde edilen mozaik UPD görüntüsü.

Mikrodizin verisinde bir kromozom için çizdirilmiş mozaik UPD için üstte LRR ve altta BAF değerleri gösterilmektedir. BAF değerlerinin kromozomun başındaki bir genomik bölgede heterozigot 0,5 değerinden sapma gösterdiğine ancak LRR değerinin 0'da kaldığına dikkat ediniz. Bu resim CC BY 2.0 lisansı ile açık erişimde paylaşılmış bir makaleden alınmıştır (6).

Mozaik UPD'ler mayoz ve fertilizasyon sorunları ve bunları takip eden zigotik ve post-zigotik trizomiden kurtulma sonucu ortaya çıkabileceği gibi ebeveynlerden kromozomların normal kalıtılmış olmasına karşın hayatın ilerleyen evrelerinde de ortaya çıkabilirler. Bu şekilde ortaya çıkan UPD'lerin en önemli sebebi homolog rekombinasyon ile DNA tamiri yapılmasıdır. Homolog rekombinasyon bir alelin diğer alelin dizisi kalıp alınarak tamir edilmesi şeklinde gerçekleştiği için bir ebeveynden alınan bir genomik bölgenin diğer ebeveyne ait bir genomik bölgeye dönüşmesine sebep olabilmektedir. Bu durum bir ebeveynden gelen haplotipi ortadan kaldıracağı için mozaik segmental UPD şeklinde görülen izodizomilere yol açabilmektedir (99). Gen konversiyonu olarak da adlandırılan bu durum, kanserli dokularda sık görülmekte ve heterozigot olarak bozulmuş tümör ilişkili gen alellerinin diğer alele de kopyalanmasına sebep olabilmektedirler (100).

2.6. Mozaik CNV'lerin İncelenmesi Amacıyla Kullanılan Biyoinformatik Araçlar

Mozaik CNV saptamaya yönelik geliştirilmiş biyoinformatik araçlar CNV saptamaya yönelik algoritmalarından köken almaktadır. Bu algoritmaların arasında ilklerden yer alan CBS (*Circular Binary Segmentation*) algoritması aCGH için CNV

saptanmasına yönelik modifiye edilmiş BS (*Binary Segmentation*), *Gaussian Markov Model*, FDR (*False Discovery Rate*) gibi istatistiksel yaklaşımları içermektedir. CBS, DNA kopya sayılarının büyük genomik bölgelerde kesintili sayısal değişikliklerle kayıp ya da kazançlar oluşturduğu bilgisiyle gürültülü sinyal şiddetlerinin ($\log_2\text{Ratio}$) hesaba katılarak genomik bölgelerin segmentlere ayrılması esasına dayanmaktadır (101). Algoritmanın 2004 yılında oluşturulmasının ardından geliştirilmiş bir versiyonu ise 2007'den itibaren DNACopy isimli bir R paketi içerisinde kullanılmaktadır (102). Mozaik CNV'lerin incelenmesi amacıyla mikrodizin ve NGS verilerinde kullanılan biyoinformatik araçlar Tablo 2.3'te özetlenmiştir.

Tablo 2.3. Mozaik CNV'lerin incelenmesi amacıyla kullanılan biyoinformatik araçlar.

Biyoinformatik Araç	SNP		Yıl	f_m Hesabı	Yazıldığı Ortam	
	NGS	Mikrodizin				Algoritma
SOMATICS*	-	+	BAF + LRR	2008	E	R
BAFsegmentation	-	+	CBS/BAF	2008	H	R**
ASCAT	+	+	BAF+LRR	2010	E	R
OncoSNP	-	+	BAF+LRR	2010	E	MATLAB
MAD/R-GADA	-	+	BAF/LRR+FDR	2011	H	R
DANFIP	-	+	BAF	2012	E	MS Excel
triPOD	-	+	BAF + LRR	2013	H	Linux + R
OncoSNP-SEQ	+	-	BAF + LRR	2013	E	MATLAB
affy2sv	-	+	MAD+R/GADA	2015	H	R
ascats	+	-	BAF + LRR	2016	E	R
MrMosaic	+	-	MAD+GADA	2017	H	Linux + R
MoChA	+	+	BAF+LRR	2018	E	Linux
MONTAGE	+	+	BAF+LRR (PennCNV)	2021	E	Linux

*Platformdan kaldırılmıştır, **Sonradan R paketine uyarlanmıştır, E: Evet, H: Hayır

2.6.1. Mikrodizin Verisi Kullanan Mozaik CNV Araçları

SOMATICS: 2008 yılında daha önceki hiçbir CNV saptama algoritmasını kullanmadan geliştirilmiş orijinal bir program olan SOMATICS, mozaik CNV oranını saptayabilen ilk algoritmadır (103). Illumina SNP mikrodizin verileri kullanılarak sadece BAF dağılımlarından yola çıkan algoritmalar kullanan; *germline* veya somatik olan, heterozigot, homozigot ve hemizigot CNV'lerle birlikte UPD'leri de saptayabilmektedir. Kontrol veri setinden bağımsız olarak tek bir tümör örneğinden hesaplama yapabilen algoritma, BAF ve LRR değerlerinin olması beklenen

değerlerden sapmalarını kullanarak tahminde bulunmaktadır. Algoritmanın validasyonu kanser hücreleriyle normal hücrelerin karışımlarından hazırlanan ve kanser hücrelerinin oranının %0, %25, %50, %75 ve %100 mozaisizmi oluşturmak üzere deneysel olarak hazırlanan örnekler kullanılarak yapılmıştır. Algoritmanın etkin kullanılabilmesi için f_m değeri yüksek olmalıdır. R'da yazılan kodlarla oluşturulan algoritma günümüzde kodların kaldırılmasından dolayı kullanılamamaktadır.

BAFsegmentation: Temel olarak CBS algoritmasının geliştirilmesiyle, segmentasyonun BAF değerlerine göre uygulanmasına dayanan bir algoritmadır (104). Algoritma sadece eşleşen normal-tümör örnek çiftlerinde denenmiştir. Tümöre özgü olan 0,5 civarındaki heterozigot BAF değerlerine sahip varyantlar, aynı bireyin eşleşen ve tümör bulunmayan bir örneğindeki varyantlar ile karşılaştırılarak analiz gerçekleştirmektedir. Algoritmanın özellikle LOH bölgelerine yönelik geliştirilmesi ve saptama yapması açısından SOMATICS, PennCNV gibi algoritmalara kıyasla yüksek performans sergilemesi ve tetrazomilere yönelik de saptama yapabilmesi öne çıkan özellikleridir. Algoritmanın öncelikli amacı mozaik CNV oranını saptamak olmasa da kanser hücrelerinin segmentasyon algoritması ile saptanan hemizigot delesyon mozaisizmi oranlarını FISH yöntemiyle bulunan oranlarla kıyaslamışlardır. Ancak yalnızca yüksek oranlı mozaisizme yönelik hesaplama yapabilmekte olup sadece yüksek oranlı mozaisizme dair sonuç vermektedir.

ASCAT: BAF ve LRR ile çalışan ve kendi segmentasyon algoritmasını kullanan ASCAT (*Allele Specific Copy Number Analysis of Tumors*) tümör hücrelerinden yapılan deneylerde farklılaşmış hücre oranını, LOH ve kopya nötral değişiklikleri genom boyu saptamaktadır. Normal-tümör karşılaştırması üzerinden hesaplama yapabildiğinden test bireyine ait *germline* veri de sağlanması gerekmektedir (105).

OncoSNP: LRR değerlerini kullanarak Viterbi algoritması üzerinden genom boyu CN hesaplaması yapan OncoSNP programı ise sadece %10'luk dilimler düzeyinde kanser hücrelerinden f_m hesaplayabilmektedir. MATLAB kodu bulunan program Linux sistemi üzerinden çalıştırılabilmektedir (106).

MAD/R-GADA: Özel olarak mozaik CNV saptanmasına odaklanan ilk programlardan biri MAD'dır (*Mosaicism Alteration Detection*)'dir (107). Bu program, mozaizmin saptanması için CBS algoritmasının daha hızlı ve geliştirilmiş versiyonu olan bir başka segmentasyon algoritması GADA (*Genome Alteration Detection Algorithm*)'yı ve beklenen BAF değerlerinden sapmaları (Bdev) kullanmaktadır (108). Buna ek olarak, LRR verisi aracılığı ile mozaik değişikliğin yönünü (kopya artışı, kopya azalışı) ve UPD'leri tespit edebilmektedir. Bu algoritma, mozaik CNV bölgelerini genom boyu saptayabilse de mozaizm fraksiyonunu (f_m) hesaplamamaktadır. Algoritma, R dilinde yazılmış olup sadece Illumina HapMap 1M SNP mikrodizin verileri ile denenmiştir ve aynı zamanda R-GADA isimli R paketi şeklinde sunulmuştur.

DANFIP: Kümülatif dağılım fonksiyonlarından simüle edilen modellerin minimum artıkların toplamı yöntemiyle en uygun f_m hesaplayan modele uydurulmasına dayanan MS Excel'de oluşturulmuş bir algoritmadır. DANFIP yönteminin uygulanmasıyla, araştırmacılar kontrol grubu verilerine dayanan mozaizm simülasyonlarının mozaik CNV'ler içeren hasta örneklerinde f_m hesabı için kullanılabileceğini göstermiştir (85). Sadece Illumina platformundan alınan SNP verilerinin kullanıldığı yöntemde, kromozomal pozisyon araştırmacı tarafından belirlenmekte ve otomatik olarak saptanamamaktadır. MS Excel üzerinden kullanılması nedeniyle hesap kapasitesi sınırlıdır.

triPOD: Mozaizmin tespiti için bir diğer yaklaşım da, anne-baba-çocuk üçlülerinden (trio) elde edilen SNP mikrodizin verisinin haplotip saptanmasında kullanılması ve bu haplotip duyarlı analiz ile mozaiklik saptanmasındaki sensitivitenin güçlendirilmesidir. Bu yaklaşımla çalışan triPOD programında hem BAF hem LRR verileri değerlendirilip mozaik CNV'lerin yanısıra mozaik UPD'ler de saptanabilmektedir (6). Ancak triPOD programı, mozaik CNV'leri mozaik olmayan değişikliklerden ayırmayı amaçlamadığından f_m hesaplamamaktadır. Bunun yanında, anne-baba-çocuk üçlüsüne ait SNP verileri üzerinden çalıştığı için 3 farklı SNP mikrodizin verisine ihtiyaç duyması bu algoritmanın uygulanmasını hem daha pahalı kılmakta hem de bazen ebeveynlere ait DNA örneklerine ulaşamaması sebebi ile

analizi imkansız hale getirmektedir. Program sadece Illumina mikrodizin verisiyle test edilmiştir.

MoChA (*Mosaic Chromosomal Alterations*): MAD'a benzer bir şekilde bdev kullanarak mozaisizmin lokasyonunu ve fraksiyonunu saptayabilen bir diğer biyoinformatik araç da Linux işletim sisteminde çalışan MoChA'dır. Bu araç sadece Illumina ve Affymetrix SNP 6.0 mikrodizinlerinde çalışabilmektedir ve çalıştırabilmek için öncelikle BAF ve LRR bilgilerinin aracın önerdiği şekilde elde edilmesi gerekmektedir (109).

MONTAGE: MoChA gibi sadece Linux içerisinde çalışabilmektedir. Kopya sayısındaki değişiklikleri ve lokasyonu saptayabilmektedir. Asıl olarak PennCNV algoritması üzerine tasarlanmış olup, perl dilinde yazılmıştır. Sadece örneğe ait tek bir BAF ve LRR tablosundan genom boyu hesaplama yapabilmektedir. Bu araç ise platform spesifik değildir (110).

2.6.2. NGS Verisi Kullanan Mozaik CNV Araçları

MAD programının BAF segmentasyonu için uyguladığı algoritma, R tabanlı MrMosaic isimli bir başka programda NGS verilerine de uygulanmıştır. Bu program da MAD programı gibi f_m hesabı yapamamakta, sadece mozaik CNV bölgelerini tespit edebilmektedir. MoChA ise NGS verisinden okuma derinliği (LRR'nin NGS karşılığı olarak) ve BAF değerlerini kullanarak mozaisizmin lokasyonunu ve hücre fraksiyonunu hesaplayabilmektedir. (97). Hem MoChA hem de MrMosaic, Linux tabanlı birtakım başka araçların da kullanımını gerektirmektedir. MONTAGE aracı da mikrodizin verisinden olduğu gibi NGS verisinden elde edilen BAF ve LRR değerlerini kullanarak mozaik CNV'leri saptayabilmektedir (98).

Pek çok program, çok farklı genetik yapıda hücreler içeren kanser dokusunda kanserli hücrelerin belli genetik lokuslarındaki ploidilerini saptamaya odaklanmıştır. Bu programlar LRR ve BAF değerlerini kullanan ve bunların dağılımını da hesaba katan modeller ile f_m tahmini yapabilmektedir. Ancak kanser dokusu için tasarlanmış programların nihai hedefi f_m tahmini olmamakta, f_m tahmini sadece ploid

hesaplamasında kullandıkları bir ara basamak halinde kalarak düşük hassasiyetli hesaplamalara izin vermektedir. Bu türden programlara örnek ascatNGS, OncoSNP-SEQ'dir (111, 112). NGS verisinden f_m hesaplaması yapmadan mCNV lokasyonu tespiti yapan diğer araçlar arasında GATK'nın (*Genome Analysis Toolkit*) somatik CNV iş akışı, control-FREEC, ADTEX (*Aberration Detection in Tumour Exome*), ExomeCNV, VarScan2 sayılabilir (113-118).

2.6.3. Mozaik CNV'lerin İncelenmesinde Model-Örnek Karşılaştırmalarının Kullanılması

Mozaik CNV'lerin saptanması amacıyla üretilen programların birçoğu makine öğrenmesi ve/veya tahmin algoritmaları kullanmaktadır. Diğer araçlardan farklı olarak DANFIP yöntemi deneysel olarak elde edilmiş veriden köken alan kontrol verilerini kullanan bir araçtır. Bu yöntemin çalışması için oluşturulan modellerin olasılık dağılımlarını, test edilen örneğin dağılımı ile karşılaştırmak amacıyla model dağılımlarının birbirine uyumluluğunu test eden çeşitli yöntemler bulunmaktadır. Bu yöntemlerden en sık kullanılanlar hakkında aşağıda bilgi verilmiştir. Bu yöntemlerin yanında dağılımların karşılaştırılmasında kullanılan daha özelleşmiş başka yöntemler de bulunmaktadır (119).

Rezidüel toplamlarının minimumu (*minimum sum of residuals, MSR*):

Deneysel olarak elde edilen veri ile modellenen verinin arasındaki farka dayanan bir yöntemdir. Her bir bağımsız değişken için model ile gerçek veride yer alan bağımlı değişkenler arasındaki farkın, diğer bir deyişle rezidüellerin, toplamı model ile gerçek verinin uyumluluğunun bir göstergesidir. Bu toplamın en küçük olduğu değer ise incelenen modeller arasında gerçek veri ile en uyumlu modeli belirlemek için kullanılabilir ve bu yaklaşım MSR olarak adlandırılmaktadır (120). MSR, veriye en uyumlu modeli saptasa da uyumun şans eseri olma olasılığı ile ilgili bir bilgi vermemektedir. Mozaik CNV'leri saptamak amacıyla kullanılan orijinal DANFIP yöntemi MSR'ye dayanmaktadır (85).

Rezidüel karelerin toplamı (*RSS, Residual Sum of Squares, RSS*):

MSR yaklaşımına çok benzeyen RSS yöntemi, rezidüeller yerine rezidüellerin karelerinin

toplamını kullanarak gerçek veri ile en uyumlu modeli aramaktadır (121). Bu yöntem de veriye en uyumlu modeli saptasa da, uyumun şans eseri olma olasılığı ile ilgili bir bilgi vermemektedir.

Pearson Ki-kare testi: Gerçek veri tarafından temsil edilen “gözlenen” ve model tarafından temsil edilen “beklenen” veri arasındaki farklılıkların incelenmesi için kullanılan ve kategorik verileri karşılaştıran bir istatistiksel yöntemdir. Uyum iyiliği (*Goodness-of-fit*), homojenite ve istatistiksel bağımsızlığın sınanması için kullanılabilir. Bu test aynı zamanda güven aralıklarının hesaplanması için de kullanılabilir (122).

Kolmogorov-Smirnov (KS) Testi: İsmi bu testi hipotez sına algoritması olarak tanıtan Andrey Kolmogorov ve dağılım algoritmasına uyarlanmasını açıklayan Nikolai Smirnov’dan alan KS testi, temelde devamlı verilerden oluşan iki dağılımın istatistiksel olarak farklı olup olmadığını sınıamaktadır. Bir uyum iyiliği testi (*Goodness-of-fit test*) olarak ve güven aralıklarının hesaplanması için de kullanılabilir. Özellikle kümülatif dağılım fonksiyonlarını (*CDF, Cumulative Distribution Function*) hem lokasyon hem de şekil açısından karşılaştırması avantajlarına sahip bir parametrik olmayan testtir. Kesikli ya da sürekli değişkenlere sahip örneklerde, tek ve iki örnek karşılaştırmalarında başarılıdır (123).

2.7. R Programlama Dili

İlk defa 1993 yılında ortaya çıkan R, öncelikle veri analizi için yazılım ortamı sağlayan bir programlama dilidir. Adını geliştiricileri olan iki istatistikçi, Ross Ihaka ve Robert Gentleman’ın baş harflerinden almaktadır (124). İstatistiksel programlama dili olan S dili üzerine C, MATLAB gibi diğer dillerden kodlar eklenerek oluşturulmuş bir dil olup, istatistiksel modeller ve hesaplamalar üzerine pek çok farklı kişi tarafından yazılmış programların kullanılmasına imkan tanımaktadır. R programlama dilinde kullanıcılar tarafından geliştirilen bu ek programlara paket adı verilmektedir (125). Bu sayede R programlama dilinin sunduğu hazır kodların yanı sıra, R’ı herhangi bir amaçla kullanmak isteyen kullanıcılar, diledikleri alana uygun olarak geliştirilmiş paketleri kullanabilmekte, hatta ihtiyaçlarına göre kendileri de paketler oluşturabilmektedirler.

Windows, Mac ve Linux işletim sistemlerinde kullanımı mümkün olan R, açık kaynaklı ve ücretsiz bir yazılıma sahip R konsol ve R Studio olmak üzere iki şekilde masaüstü uygulamasına sahiptir (126). R Studio hem kodun yazılabileceği *script* kısmı hem de çalıştırılabileceği terminal kısmına sahip olan R'in kendine ait bir kod editörüdür. R paketleri, R Studio içerisinde çalıştırılan yükleme komutuyla, CRAN ve Bioconductor gibi çevrim içi depolardan indirilebilmekte veya bilgisayardaki lokal bir kaynaktan yüklenebilmektedir. CRAN, R paketlerinin yanı sıra, R'in indirilebilir kendi yazılımını, kaynak kodlarını, dokümantasyonunu ve arşivlenmiş versiyonlarını da içermektedir (127). Bioconductor ise özellikle biyolojik bilimlerle ilgili paketlerin tutulduğu, yaygın kullanılan bir depodur (128). Bioconductor sunucusunun varlığı ile de görülebileceği gibi, R programlama dili biyolojik verilerin istatistiksel analizi için sıklıkla başvurulmuş ve yeni programların geliştirilebildiği bir dil özelliğini taşımaktadır.

R'da kullanılan paketler içerisinde özellikle program geliştiricilerinin tercih ettiği Shiny paketi ön plana çıkmaktadır (129). Shiny, HTML, Java, SQL gibi web yazılımı geliştirme dillerinin bilinmesine gerek olmadan web arayüzlü uygulamaların yaratılmasına olanak veren bir R paketidir. Bu sayede, Shiny tarafından yaratılmış bir arayüze sahip web uygulamasının, kodlama olmaksızın sadece kullanıcı tarafından girdilerinin sağlanarak çalıştırılması mümkün olmaktadır.

Kullanıcı dostu araçların ve programların geliştirilmesine olanak vermesine karşın, R'in birtakım dezavantajları da mevcuttur. Genellikle Windows masaüstünde kullanmayı tercih eden kullanıcılar için daha çok bireysel amaçlara hizmet etmeye uygundur; ancak özellikle büyük veri içeren ve iç içe kodların kullanılması gereken programlarda hız açısından tatmin edici olmamaktadır. Özellikle birden fazla programın çalışmasını gerektirecek çeşitli iş akışlarının geliştirilmesi için yavaş kaldığından ideal bir ortam sayılmamakta, bu tür amaçlar için genellikle Linux içerisinde C++ gibi diller ve iş akış araçlarının kullanılması tercih edilmektedir. Ek olarak, yoğun hesaplamalı amaçlar için kullanılmak istendiğinde kullanıcının bilgisayarının RAM, disk kapasitesi gibi donanımsal özelliklerinin R'in performansını etkileyeceğinin unutulmaması gerekmektedir. Ancak bu küçük dezavantajlarına

karşın R programlama dili, sıklıkla güncellenen ve geliştirilmeye devam edilen, açık erişimli, paketlerle çalışması sayesinde kullanıcı dostu ve bilimsel camiada kullanımı kabul gören evrensel bir programlama dili olması gibi sebeplerle günümüzde veri analizindeki popülerliğini korumaktadır.

3. BİREYLER VE YÖNTEM

Bu tez kapsamında, CNV'lerin mozaisizm fraksiyonunu saptamaya yönelik R programlama dilinde (Windows versiyon 4.0.4) Copy-fm isimli bir biyoinformatik araç oluşturulmuştur. Bu aracın oluşturulması için 1) Daha önceden elde edilmiş ve deneysel olarak bu çalışmada oluşturulan mCNV içeren mikrodizin verileri elde edilmiş, 2) DANFIP yaklaşımı temel alınarak bir algoritma oluşturulmuş, 3) Algoritmanın etkinliği, deneysel verilerden elde edilen sonuçlara göre test edilmiştir.

3.1. Bireyler

Çalışma kapsamında kullanılmak üzere 2 tür veriden faydalanılmıştır. İlk grup veri tez öncesinde TÜBİTAK destekli Era-Net projesi kapsamındaki EuroDBA (Proje No: 315S192) ve EJP-RD projesi kapsamındaki RiboEurope (Proje No: 319S062) projeleri çerçevesinde elde edilen bazı mikrodizin verileri ile kamuya açık olarak paylaşılan verilerdir. İkinci gruptaki veriler ise tez kapsamında çalışmaya katılmayı kabul eden katılımcılar tarafından sağlanan periferik kan örneklerinden elde edilen mikrodizin verileridir.

Çalışmanın ikinci grup verilerinin elde edilebilmesi amacıyla Hacettepe Üniversitesi Tıbbi Genetik Polikliniği'ne başvuran bireyler çalışmaya davet edilmiştir. Bu kapsamda katkı sunmayı kabul eden ve birbirleri arasında aile bağı olmayan 5 sağlıklı kadından kontrol grubu, bu kişilerle akrabalık bağı olmayan 1 sağlıklı kadın ve oğlu test grubunu oluşturmak üzere çalışmaya dahil edilmiştir. Bu kişilerden elde edilen periferik kan örneklerinin bir kısmı direkt preparasyon ile FISH çalışması için ayrılırken, bir kısmı ise DNA izolasyonu ve mikrodizin çalışmaları için ayrılmıştır. Tez protokolünün Hacettepe Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu tarafından 02.03.2021 tarihli toplantısındaki GO 21/267 nolu oluru EK-1'de sunulmuştur. Çalışmaya katılan bireylerden özellikle cinsiyet kromozomlarına yönelik mikrodizin çalışmaları yapıldığı için cinsiyet katılımcılar aşağıdaki kriterleri karşılayanlar arasından seçilmiştir.

- 1) Katılımcılarda ileri yaşta ortaya çıkabilecek cinsiyet kromozomlarındaki mozaizmin önüne geçebilmek için;
 - a. Kontrol grubunu oluşturan 5 kadın katılımcının yaşları 18-40 arasındadır.
 - b. Test grubunu oluşturan annenin yaşı 35-50, oğlunun yaşı ise 18-25 arasındadır.
- 2) Katılımcılarda kanda mozaik değişikliklerin bulunmamasını sağlamak için;
 - a. Bilinen metastatik kanseri veya lösemisi olanlar çalışmaya alınmamıştır.
 - b. Katılan kadınlarda son 3 ay içerisinde gebelik veya gebelik kaybı bulunmamaktadır.
 - c. Katılımcılarda kromozom sayısı veya CNV ile ilişkili bir hastalık veya şüphesi bulunmamaktadır.
 - d. Katılımcıya ait aile ağacına göre katılımcıda CNV bulunması şüphesi yoktur.
- 3) Katılımcılar çalışmaya katılmayı kabul edip aydınlatılmış onam formunu imzalamıştır (Bkz. EK-3).

3.2. Çalışmada Kullanılan Deneysel Yöntemler

3.2.1. FISH

Tez kapsamında toplanan periferik kan örneklerinde X kromozomlarında %1'den yüksek mozaizm bulunması sonraki basamaklarda yapılacak işlemlerde sorun teşkil edeceği için, mikrodizin çalışmalarına geçmeden önce her bir periferik kan örneği cinsiyet kromozomlarında mozaizm durumunu değerlendirmek amacıyla FISH çalışmaları ile değerlendirilmiştir. Bu amaçla, öncelikle her bir periferik kan örneğinden direkt preparasyon yapılmış, ardından örnekler FISH problemleri ile boyanmış ve florasan mikroskop altında interfazdaki hücrelerin X ve Y kromozom sayıları değerlendirilmiştir.

Direkt Preparasyon: Çalışmaya katılan her bireyin sodyum heparinli tüplerde +4 °C'de 1 gün bekletilmiş kan örneklerine direk preparasyon işlemi yapılmıştır. İşleme başlamadan 1 saat önce 3x metil alkol ve 1x asetik asitle fiksasyon çözeltisi - 20 °C dolapta saklanmıştır. Eritrositlerin hipotonik ortam oluşturularak patlatılması için her bir periferik kan örneğinden 1 ml alınarak 15 ml'lik falkon tüplere aktarılmasının ardından 37 °C sıcaklıktaki KCl (potasyum klorür) çözeltisinden 1 ml vorteksleme işlemi eşliğinde damlalar halinde yavaşça eklenmiş, ardından hızlıca 5-6 sn vorteks yapılmış ve eritrositlerin hipotonik ortamda patlatılmaları sağlanmıştır. Sonrasında 2500 rpm'de 10 dakika santrifüj uygulanmasının ardından süpernatant kısım vakum pompasıyla uzaklaştırılmış ve KCl çözeltisi ile eritrositlerin patlatılması işlemi çökelti kırmızılığını kaybedene kadar en az 2 kez tekrarlanmıştır. Kırmızı rengini tamamen kaybetmiş çözelti önceden hazırlanan fiksatif çözeltisinden (3:1 metanol:asetik asit) 1 ml eklenerek fikse edilmiş ve 2500 rpm'de 10 dk santrifüj edildikten sonra süpernatant vakum pompasıyla uzaklaştırılmıştır. Ardından her tüpteki çökelti üzerine 10 ml fiksatif çözelti eklenerek +4 °C'de 1 gece tamamen fikse olması sağlanmıştır.

Ertesi gün örnekler 2500 rpm'de 10 dk santrifüj edilerek süpernatantlar uzaklaştırılmış; fiksasyon işlemleri 7 ml, 5 ml ve 3 ml fiksatif çözelti ile tekrarlanarak yıkamalar yapılmıştır. Her bir örnek için örnek numaraları ve tarihler yazılmış en az 2'şer tane lam üzerine 3 ml'lik yıkamanın ardından kalan çökeltinin 1 ml fiksatif ile sulandırılması ardından, nemli bir yüzey sağlaması için cam Pastör pipeti ile 1 damla fiksatif damlatılarak, hafifçe yayılarak ıslatılan lam üzerine fikse edilmiş örnek hücrelerini içeren çözeltilerden 3-4 damla yukarıdan damlatılarak yayıldı. Sonrasında, elle sallanarak kurutulan lam üzerindeki hücrelerin patlaması ve hücre çekirdeği içeriklerinin lam üzerinde yayılması sağlandı. Lamaların eskitilmesi için örneklerin iyice lama yapışıp kuruması amacıyla 60 °C'ye getirilmiş etüvde 1 gece boyunca bekletilmiştir.

Ertesi sabah preparatlar ışık mikroskopuyla gözlenerek her bir lam için hücrelerin yoğun olduğu bölge üzerine 10 µl FISH probu (DIAGEN Biyoteknolojik

Sistemler A.Ş., primeFISH® ChrX/Y Sentromerik Prob Kiti) eklenip lamelle üzerleri kapatılmıştır. Arada hava kabarcıklarının kalmadığından emin olunduktan sonra yapıştırıcıyla kenar ve köşeleri tamamen kapsayacak şekilde kapatılmıştır. Ardından HYBrite hibridizasyon fırınında 70 °C'de 7 dk ve 37 °C'de 16 saate ayarlı programda hibridizasyon işlemi tamamlanan lamlar bunu takiben üzerlerindeki lamel kaldırıldıktan sonra 72 °C su banyosu içerisinde 0.4 x SSC çözeltilisinde 30 sn, ardından 2 x SSC + Tween20 çözeltilisi içinde 45 sn bekletilmiştir. Bu işlemden sonra kurutulan lamlar üzerinde FISH problemleri ile boyanan bölgeler üzerine 10 µl DAPI (4',6-diamidino-2-phenylindole) boyası damlatılarak daha büyük bir lamelle hava boşluğu kalmayacak şekilde kapatılmıştır.

Hazırlanan preparatlar ışık almayacak şekilde en az 20 dk boyunca -20 °C'de tutulduktan sonra Nikon Eclipse E800 floresan mikroskopunda her bireye ait örnekten 300'er tane interfaz çekirdeği içerisindeki FISH problemleri sayımı yapılarak CytoVision 4.0.3 (Applied Biosystems) yazılımı kullanılarak analiz ve fotoğraflama gerçekleştirilmiştir. Üretici firmanın talimatlarına göre X kromozomu yeşil Y kromozomu ise kırmızı sinyalle değerlendirilmiştir. Bu sayede 300 hücrenin içerisindeki X ve Y prob sayıları incelenerek örneklerde %1'den fazla X kromozomu mozaizmi bulunup bulunmadığı değerlendirilmiştir.

3.2.2. DNA Örneklerinin Elde Edilmesi ve Deneysel Mozaik Delesyon Örneklerinin Oluşturulması


FISH testleri ile uygun örneklemin toplandığına emin olunduktan sonra çalışmaya katılan bireylerden EDTA'lı tüplerde alınan periferik kan örnekleri +4 °C'deki dolaptan çıkarıldıktan sonra "*Lucigen-MasterPure DNA Purification Kit for Blood Version II*" izolasyon kiti kullanılarak DNA elde edilmiştir. Her bir bireyden alınan kan örneklerinin 3 ml'si 15 ml'lik falkon tüplere aktararak üzerlerine kit içerisinde bulunan 9 ml *Red Cell Lysis* çözeltilisi eklenip alt üst edilerek karıştırılmıştır. Örnekler, buzlu kapta 10 dk boyunca bekletilip birkaç kez vortekslendikten sonra 770 x g'de 3 dk santrifüj yapılmış ve süpernatant uzaklaştırılıp 4,5 ml *Red Cell Lysis* çözeltilisi eklenerek işlemler tekrarlanmıştır. Çökelti üzerine kit içerisindeki *Tissue Cell Lysis* çözeltilisinden 4,5 ml eklenerek karıştırılması ve hücrelerin parçalanması ardından 15

μ l RNA parçalayıcı enzim olan *RNase A* eklenip tekrar karıştırılmış ve 37 °C etüvde 30 dk inkübasyona bırakılarak RNA'lar uzaklaştırılmıştır. İnkübasyon sonrası 5 dk buzda bekletilen örnekler üzerine 2,6 ml MPC protein çöktürücü eklenerek 10 sn vortekslenmiş ve 5500 x g'de 20 dk santrifüj yapıldıktan sonra DNA'yı içeren süpernatant yeni bir tüpe aktarılmıştır. Yeni tüplerdeki örnekler üzerine 7,5 ml +4 °C'deki izopropanol eklendikten sonra 30-40 kez alt üst edilerek DNA'nın presipite olması sağlanmıştır. Örnekler +4 °C'de 10 dk 1350 x g'de santrifüj edilip süpernatantın dökülerek uzaklaştırılmasının ardından 500 μ l %70'lik etanolle yıkama yapılmış ve sonrasında çökelti ile beraber etil alkolün tamamı 2 ml'lik ependorf tüpe aktarılıp pipetajile karıştırılmıştır. Ardından 13000 rpm'de 1 dk santrifüj yapıp etanol uzaklaştırılmış, etanolle yıkama işlemi tekrarlanmış ve etanolün tamamen buharlaşması için 20 dk kurumaya bırakılmıştır. 200 μ l TE (*Tris EDTA*) tampon çözeltisi çökelti üzerine eklenerek 65 °C su banyosunda 10 dk bekletildikten sonra 37 °C'deki etüvde 1 gece bekletilmiş ve DNA'nın tamamen çözünmesi sağlanmıştır. DNA izolasyon işleminin ardından kontrol ve test örneklerine ait DNA konsantrasyonları Nanodrop 2000c (ThermoScientific) spektrofotometrik DNA miktarı ölçüm cihazında 230 nm, 280 nm ve 260 nm dalga boylarında ölçülmüştür. Bu şekilde izole edilen ve yeterli konsantrasyon ve saflığa sahip kontrol grubuna ait DNA örnekleri olduğu gibi kullanılırken test grubunu oluşturmakta kullanılacak anne ve oğluna ait DNA örnekleri X kromozomu mozaik modelleri oluşturmak için farklı oranlarda karıştırılmıştır.

Anneler ve oğulları ortak bir X kromozomu taşımaktadırlar. Oğullarda bulunan X kromozomları doğal olarak annelerinden kalıtıldığı için, annelerine kıyasla tam bir X kromozomu açısından delesyon barındırmakta, ayrıca sahip oldukları X kromozomu ise haplotip açısından annedeki bir kromozomun tam karşılığı olmaktadır. Buna göre, anneye ait her hücrede ikişer X kromozomu bulunması sebebiyle annenin DNA örneği X kromozomunun tam olarak 2 kopyasını içeren %0 mozaik delesyon örneği (T0), oğluna ait örnek ise her hücrede tek X kromozomu bulunması sebebiyle X kromozomunun tam olarak 1 kopyasını içeren %100 mozaik delesyon örneği (T100) olarak tayin edilmiştir. Bu iki örneğin konsantrasyonlarının 100 ng/ μ l değerine ayarlanması ardından farklı oranlarda karıştırılması ile farklı deneysel mozaik

delesyon örnekleri oluşturulmuştur. Buna göre, anne ve oğul DNA'ları kullanılarak X kromozomu delesyon mozaizmine yönelik normalden tam delesyona olmak üzere; T0, T5, T10, T20, T35, T50, T65, T80, T90, T95, T100 ismi verilerek farklı mozaizim oranlarında hazırlanan 11 adet deneysel delesyon örneği oluşturulmuştur (Tablo 3.1).

Tablo 3.1. Anne-oğul DNA örneklerinin farklı mozaik delesyon modeli örneklerinin oluşturulması için yapılan karışımlar.



Örnek adı	fm (delesyon)	T0 (Anne) (100 ng/μl)	T100 (Oğul) (100 ng/μl)
T0	%0	20 μl	0 μl
T5	%5	19 μl	1 μl
T10	%10	18 μl	2 μl
T20	%20	16 μl	4 μl
T35	%35	13 μl	7 μl
T50	%50	10 μl	10 μl
T65	%65	7 μl	13 μl
T80	%80	4 μl	16 μl
T90	%90	2 μl	18 μl
T95	%95	1 μl	19 μl
T100	%100	0 μl	20 μl

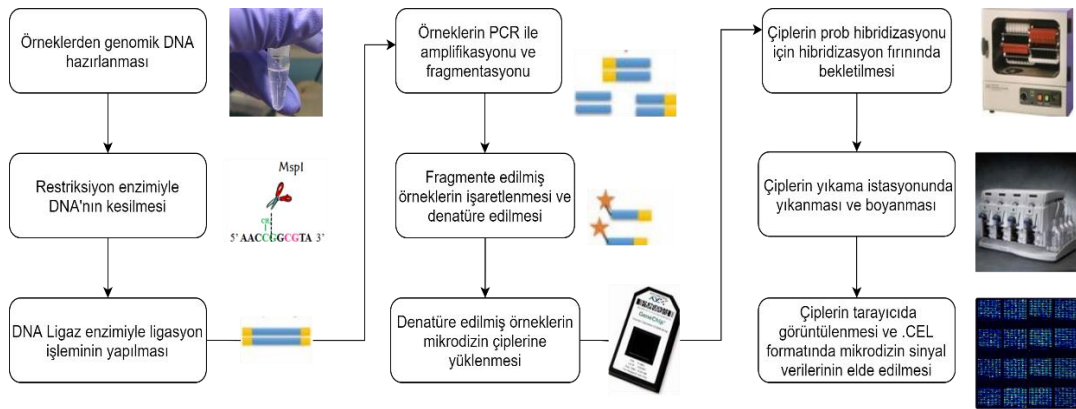
Affymetrix SNP Mikrodizin için örnek başına 5 μl DNA'ya ihtiyaç olup Illumina mikrodizini için de gönderilmek üzere totalde örnek başına 20 μl olacak şekilde hazırlanan DNA'lar, mikrodizin uygulamasına kadar -20 °C'de saklanmıştır.

3.2.3. SNP Mikrodizin Yöntemi

Mikrodizin işlemleri, K1-K5 şeklinde isimlendirilen 5 "kontrol örneği" ve "test örneği" olarak kullanılan ve T0-T100 şeklinde isimlendirilen 11 delesyon mozaizmi örneği olmak üzere aynı anda hazırlanmış toplamda 16 adet DNA örneği Affymetrix ve Illumina olmak üzere iki farklı mikrodizin platformunda çalışılmıştır. Böylece farklı platformlarda gerçekleştirilen deneyler arası varyasyonu en aza indirmek hedeflenmiştir.

Affymetrix için CytoScan Optima mikrodizin çipleri kullanılırken Illumina için Human CytoSNP-12 v2.1 BeadChip mikrodizin çipleri kullanılmıştır. Affymetrix mikrodizini için örnekler Hacettepe Üniversitesi'nde çalışılırken Illumina mikrodizin işlemleri ise Gen Era Diagnostik A.Ş'den hizmet alımı yoluyla gerçekleştirilmiştir. SNP

mikrodizini yapılması amacıyla elde edilen DNA örnekleri Affymetrix CytoScan Optima mikrodizin çipleri kullanılarak üretici firmanın önerdiği *CytoScan® Optima Assay* kitapçığına uygun hazırlandıktan sonra *Affymetrix GeneChip® Scanner 3000* mikrodizin sisteminde çalışılmıştır (130). 5 kontrol ve 11 test örneği olmak üzere toplam 16 örnek 8 örnek çalışmasını bir seferde yapacak şekilde 2 seferde çalışılmıştır. Mikrodizin işlemi için takip edilen basamaklar Şekil 3.1’de özetlenmiş ve mikrodizin işleminde izlenen basamaklar ilerleyen paragraflarda detaylı açıklanmıştır.



Şekil 3.1. Affymetrix CytoScan Optima mikrodizin platformunun deneysel uygulama basamakları.

Şekilde SNP mikrodizin basamakları; örneklerin hazırlanması, DNA'nın kesimi, ligasyon, PCR ve fragmentasyon, işaretleme, mikrodizin çipine yükleme, hibridizasyon, yıkama+boyama ve analiz basamakları özetlenmektedir.

Örneklerin hazırlanması: DNA örnekleri çözüldükten sonra önerilen mikrodizin protokolüne göre 50 ng/µl olacak şekilde sulandırılıp 1,5 ml'lik Ependorf tüplerine aktarılmıştır. Örnekler santrifüj edildikten sonra her birinden alınan 5 µl DNA yeni tüplere aktarılmıştır.

Restriksiyon enzim kesimi: Buz üzerinde protokolde 8 örnek için önerildiği şekilde Nsp I tamponu, nükleaz içermeyen su, 100X BSA ve Nsp I kesim enzimi kullanılarak stok hazırlandıktan sonra hazırlanan DNA örneklerine paylaştırılmıştır. Sonrasında termal döngü cihazında önerilen kesim koşullarında enzim kesimi gerçekleştirilmiştir.

Ligasyon: Buz üzerinde CytoScan Optima protokolüne göre 8 örnek için önerilen miktarlarda DNA ligaz tamponu, Adaptör, Nsp I ve DNA ligaz enzimi karıştırılarak stok hazırlanmış ve 8 tüpe dağıtılmıştır. Örnekler önerilen koşullarda termal döngü cihazında çalıştırılarak adaptör dizilerin ligasyonu tamamlanmıştır.

PCR: Ligasyon işlemi yapılmış DNA örnekleri belirtilen miktarda sulandırılmıştır. Ardından PCR için nükleaz içermeyen su, CytoScan Taq tamponu, 5M Betain, dNTP, PCR primeri, CytoScan Taq DNA Polimeraz ile stok hazırlanmış ve her bir örnek PCR ile amplifiye edilmiştir. Termal döngü cihazında PCR sonrası her bir örneğe ait PCR ürünlerinin agaroz jel elektroforezinde yürütülüp fotoğraflanarak kontrol edilmiştir.

PCR pürifikasyonu: Her bir PCR ürünü 2 ml'lik tüplere aktarıldıktan sonra üzerlerine pürifikasyon *bead* (DNA'ya bağlanan mikroskobik boncuklar) eklenerek karıştırılmıştır. Daha sonra örnekleri içeren tüplerin önüne manyetik stant yerleştirilip DNA'ya bağlanan boncuklar sayesinde PCR ürünlerini içeren çökeltinin bir tarafa toplanması sağlanmış kalan süpernatant pipetle uzaklaştırılmıştır. Manyetik pelet kit içerisinde yer alan etanolle sulandırılmış yıkama tamponu ile iki kez yıkandıktan sonra tüplerde bulunan manyetik boncuklar elüsyon tamponuyla yıkanarak pürifikasyon tamamlanmıştır.

Miktar tayini: Pürifikasyon sonrası örneklerin DNA miktarları NanoDrop ND-3000 spektrofotometresi kullanılarak CytoScan Optima protokolünde tarif edildiği dalga boylarında yapılan ölçümler sonucunda ~2500 ng/ μ l'ye sahip örnekler uygun kalitede kabul edilmiştir.

Fragmentasyon: Fragmentasyon tamponu, fragmentasyon enzimi ve nükleaz içermeyen su ile CytoScan Optima protokolünde belirtildiği şekilde hazırlanıp örnekler paylaştırıldıktan sonra PCR cihazının fragmentasyon gerçekleştirilmiştir. Fragmente edilen örnekler protokolde belirtildiği şekilde jel elektroforezde yürütülerek fragment boyutlarının 25 bp ile 125 bp arasında olduğu doğrulanmıştır.

Etiketleme: Fragmente edilen DNA örneklerinin biyotin ile etiketleme işlemi için, CytoScan Optima Kit içerisinde yer alan TdT tamponu, DNA etiketleme belirteci ve TdT enzimiyle hazırlanan stok örneklere paylaştırılmış ve termal döngü cihazında etiketleme için belirtilen koşullarda çalıştırılmıştır.

Hibridizasyon: Üzerine örnek numaraları yazılan Affymetrix CytoScan mikrodizin çipleri AGCC Portal yazılımı için önerildiği şekilde mikrodizine özgü çalışma adı örnek numaraları ile kaydedilmiş, çipler bilgisayara okutularak eşleşme sağlanmıştır (131). Protokole uygun olarak hibridizasyon stoğunun hazırlanması için 4 adet hibridizasyon tamponu ve oligo kontrol belirteci kullanılmıştır. Hazırlanan stok örneklerle karıştırılmış ve termal döngü cihazında belirtilen denatürasyon koşullarında denatüre edilmiş, ardından örnekler önerilen şekilde pipet yardımıyla çiplere aktarılmıştır. Hava kabarcıklarının giderilmesi için birkaç kez pipetaj yapıldıktan sonra çiplere yüklenen örnekler hibridizasyon fırınına yerleştirilerek 16 saat boyunca hibridizasyon gerçekleştirilmiştir.

Örneklerin yıkanması, boyanması ve taranması: Hibridizasyon sonrası 8 örnek, *Fluidics* 450 yıkama istasyonunda kitin içerisinde yer alan 2 adet yıkama tamponu, mikrodizin tutma tamponu ve boyama tamponu (streptavidin/biyotin boyama) ile yıkanmış ve boyanmıştır. Bu işlemlerin ardından çipler birer birer *Affymetrix GeneChip Scanner 3000 7G* cihazında okutularak .CEL dosya formatında kaydedilip sinyal görüntüleri kontrol edilmiştir.

Mikrodizin analizi: Affymetrix mikrodizini için sinyal şiddetlerine ait bilgileri içeren .CEL formatlı ham veri dosyaları ChAS (Chromosome Analysis Suite) v4.2.1 yazılımı içindeki *Analysis Workflow* modülü ile hg19 referans genomuna uygun olarak ChAS yazılımı tarafından okunabilir formattaki .cychp dosyalarına dönüştürülmüştür (Bkz. EK-4) (132). Ardından Copy-fm ile ilgili işlemlere geçilmeden önce kontrol örnekleri ile test örneklerinden T0 ve T100 ChAS yazılımı ile BAF ve LRR grafiklerinden yararlanılarak X kromozomundaki *germline* CNV'ler, mCNV'ler ve LOH bölgeleri açısından analiz edilmiştir.

Illumina mikrodizin verileri ise hizmet alımı işlemi sonrası .idat formatındaki ham veriler şeklinde yüklenici şirket tarafından teslim edilmiş ve Human CytoSNP-12 v2.1 BeadChip mikrodizinine uygun olarak SNP *manifest* dosyası HumanCytoSNP-12v2-1_LM.bpm ve *cluster* dosyası HumanCytoSNP-12v2-1_LM.egt kullanılarak açık erişimli ve Windows tabanlı Illumina GenomeStudio v2.0.5 yazılımı kullanılarak analiz edilmiştir (Bkz. EK-6)(133, 134). Bu analiz sırasında tüm bireyler için ayrı olmak üzere *FinalReport* adlı .txt formatında nihai veri tabloları elde edilmiştir. Önce Illumina Beeline 2.0 yazılımıyla .idat dosyalarının genotip bilgileri işlenerek .gvc dosyalarına çevrilmesi, ardından BlueFuseMulti v4.5 yazılımıyla BAF ve LRR grafiklerinin görüntülenmesiyle, Affymetrix mikrodizin verilerinin analizinde olduğu gibi, her bir örneğe ait X kromozomlarındaki *germline* CNV, mCNV ve LOH bölgelerinin analizi gerçekleştirilmiştir (135, 136).

3.3. Copy-fm Algoritmasının Oluşturulması

Tez çalışmasında oluşturulan Copy-fm algoritması, R programlama dilinde 3 farklı fonksiyon şeklinde tasarlanmış olup bu fonksiyonlar, işlevleri ve argümanları aşağıda sıralanmıştır:

i. Ön Veri İşleme Fonksiyonu (*preprocess*): Bu fonksiyon Affymetrix ve Illumina tarafından elde edilen ham mikrodizin verilerinden Copy-fm'e uygun formatta kontrol ve test örnekleri için tablolar oluşturmayı sağlamaktadır. Fonksiyon için oluşturulan argümanlar aşağıdaki gibidir:

preprocess(path, data_type, save=T, save_file)

path: Alınan verinin dosya sistemindeki konumunu göstermektedir.

data_type: Verinin hangi mikrodizin platformuna ait olduğunu göstermektedir. "Affymetrix" veya "Illumina" olarak belirtilebilir.

save: Oluşturulan veri tablosunun bilgisayar dosya sistemi içerisinde kaydedilip kaydedilmeyeceğini gösteren mantıksal argümandır. Varsayılan değer T (*true*) olarak belirlenmiştir.

save_file: Dosya kaydedilecekse bunun dosya sistemindeki konumunu göstermektedir.

ii. **Copy-fm Analiz Fonksiyonu (*fm*):** Bu fonksiyon Copy-fm algoritmasını kullanarak belirli bir genomik lokustaki f_m değerini hesaplamaktadır. Fonksiyon için oluşturulan argümanlar aşağıdaki gibidir:

fm(control, sample, chromosome, position_start, position_end, data_type, conf=0.01, het_int=0.35, control_het_check=T, control_het_check_threshold=0.8, seed=42, plot_disomy=F, plot_ks=T, plot_models=T, mosaicism_dist=F, n_loop=NA)

control: *preprocess* işleminden geçmiş csv formatındaki kontrollere ait dosyadır.

sample: *preprocess* işleminden geçmiş csv formatındaki test örneğine ait dosyadır.

chromosome: f_m hesaplaması yapılacak olan genomik lokusun içinde bulunduğu kromozomu belirtmektedir.

position_start: f_m hesaplaması yapılacak olan genomik lokusun kromozomdaki başlangıç pozisyonunu belirtmektedir.

position_end: f_m hesaplaması yapılacak olan genomik lokusun kromozomdaki bitiş pozisyonunu belirtmektedir.

data_type: Verinin hangi mikrodizin platformuna ait olduğunu göstermektedir. "Affymetrix" veya "Illumina" olarak belirtilebilir.

conf: Kontrol ve test örneklerine ait "Confidence" kolonundan seçilecek olan güven skoru değerlerinin sınırını belirlemektedir. Varsayılan değer 0,01 olup, bu parametre sadece *data_type* olarak Affymetrix seçildiğinde işlev görmektedir. Illumina platformu için bu argüman kullanılmamaktadır.

het_int: Kontrol ve test örneklerinden seçilecek olan heterozigot BAF değerlerinin aralığını belirtmektedir. Varsayılan değer olan 0,35 değeri, $0,5 \pm 0,35$ aralığındaki, başka bir deyişle, 0,15 ile 0,85 arasında yer alan BAF değerlerinin seçilmesini sağlamaktadır.

control_het_check: Kontrollerdeki heterozigot BAF değerlerinin yeterli olup olmadığını sorgulayan fonksiyonun uygulanıp uygulanmayacağını belirtmektedir. Mantıksal argüman olup T varsayılan olarak belirlenmiştir.

control_het_check_treshold: *control_het_check* fonksiyonu uygulanacağı zaman kontrollerdeki ortalama heterozigot varyant sayısı ile *control_het_check_treshold* değeri çarpılarak o genomik lokus için bir örneğin sahip olması gereken minimum

heterozigot varyant sayısı limiti belirlenir. Minimum heterozigot varyant sayısının altında heterozigot varyant içeren kontroller analizden çıkarılır. Varsayılan değeri 0,8'dir.

seed: Copy-fm algoritması içerisinde bazı aşamalarda randomizasyon işlemlerine ihtiyaç bulunmaktadır. Copy-fm çalıştırıldığında her seferde aynı sonuçların elde edilmesi için belirli bir "seed" değeri belirlenmektedir. Buna göre, varsayılan değer olarak rasgele 42 belirlenmiştir. Bu argüman kullanıcı tarafından değiştirilebilir; ancak argümandaki değişikliklerin sonuçlarda küçük etkileri olması beklenmektedir.

plot_disomy: Algoritma içerisinde kontrollerden bir ideal dizomi (kopya sayısı=2) CDF modeli oluşturulmaktadır. Bu argüman ideal dizomi grafiğinin çizdirilmesini mantıksal argümanla açmak veya kapatmak için bulunmaktadır. Varsayılan değeri F (false)'dir.

plot_ks: Kolmogorov-Smirnov (KS) test istatistiğinin sonuç grafiğinin çizdirilip çizdirilmeyeceğini belirten mantıksal argümandır. Varsayılan değeri T'dir. Ancak *scanfm* algoritmasında bu parametre F olarak çalıştırılmaktadır.

plot_models: Kontroller ile test örneğinin varyantlarından modellenmiş nihai CDF'lerin çizdirilip çizdirilmeyeceğini belirten mantıksal argümandır. Varsayılan değeri T'dir. Ancak *scanfm* algoritmasında bu parametre F olarak çalıştırılmaktadır.

mosaicism_dist: f_m hesaplamasının çok sayıda iterasyonlar şeklinde yapılmasını sağlayan veya engelleyen mantıksal argümandır. Varsayılan değeri F'dir. Kullanıldığı takdirde R tarafından atanan rasgele *seed* değerlerinde yapılan hesaplamaların ortalaması nihai sonuç olarak bildirilmektedir.

n_loop: *mosaicism_dist* T olarak belirtilmişse bu argümanla f_m hesabının kaç iterasyon ile yapılacağı belirtilmektedir.

iii. Copy-fm Tarama Fonksiyonu (*scanfm*): Bu fonksiyon *fm* fonksiyonunu kullanarak kayan çerçeveler şeklinde belirtilen genomik lokuslar içerisinde mozaik CNV bölgelerini aramaktadır. Fonksiyon için oluşturulan argümanlar aşağıdaki gibidir:
scanfm(control,sample,chr=all,sex,data_type, window_width=10, window_shift=0.5)

control: *preprocess* işleminden geçmiş kontrollere ait csv formatındaki dosyadır.

sample: *preprocess* işleminden geçmiş test örneğine csv formatındaki ait dosyadır.

chr: Kromozom numaralarını alan argümandır. Tamamı taranacak bir kromozomu veya tüm genomu belirtmektedir. Varsayılan değer “all” olarak belirlenmiş olup tüm kromozomları içeren genomun taranması anlamına gelmektedir.

sex: Test örneği için cinsiyet belirten argümandır. Eğer “f” olarak belirtilmişse “X” kromozomu tarama işlemine dahil edilecektir. “m” belirtildiği takdirde sadece 22 otozomal kromozom taranacaktır.

data_type: Verinin hangi mikrodizin platformuna ait olduğunu göstermektedir. “Affymetrix” veya “Illumina” olarak belirtilebilir.

window_width: Taranacak genomik lokusun içerisinden seçilecek pencerelerin genişliğini Mbp (milyon baz çifti) cinsinden belirtmektedir. Varsayılan değeri 10 olup bu değer pencere genişliğinin 10 Mbp olması anlamına gelmektedir.

window_shift: Genomda taranacak pencere boyutunun başlangıç ve bitiş pozisyonları `window_width` ile `window_shift` değerlerinin çarpımı ile belirlenmektedir. Varsayılan değeri 0,5 olup bu değer pencere pozisyonlarının bir sonraki döngüde `window_width` değerinin yarısı kadar kaydırılarak f_m hesaplaması yapılacağı anlamına gelmektedir.

Copy-fm aracının f_m hesaplamasında yaptığı analiz genel bilgiler bölümünde bahsedilmiş olan DANFIP algoritmasına (26) benzemekte olup bu işlemi gerçekleştiren tüm basamaklar f_m fonksiyonu altında toplanmıştır. Bu algoritma; mozaik CNV içerdiğinden şüphelenilen bir test örneği için, belirli bir genomik bölgedeki heterozigot BAF değerlerinden kümülatif dağılım fonksiyonu (CDF, *Cumulative Distribution Function*) oluşturduktan sonra, aynı genomik bölgeden kontrol örnekleri verisine dayanılarak modellenmiş mozaisizm CDF’leri ile kıyaslama yapılması esasına dayanmaktadır. CDF’in temelde bir olasılık dağılımı olması, dolayısıyla 0 ile 1 arasında değerler alabilmesi ve dağılımın sadece bu aralıkta kalmasıyla dağılımların optimum analizi sağlanabilmektedir. CDF’lerin kıyaslanması için uygulanan rezidüel karelerinin toplamı (*residual sum of squares*, RSS) yöntemine ek olarak Kolmogorov Smirnov (KS) testi uygulanmıştır.

Oluşturulan model CDF'ler delesyon ve çeşitli duplikasyon senaryolarına göre farklı hesaplamalar gerektirmekte olup izlenen yöntem Bölüm 3.3.3 ve 3.3.4'te detaylandırılmıştır. Model CDF'ler ile test CDF'si arasındaki kıyaslamalar test edilen örneğe en benzer mozaisizm modelini KS testinde en düşük test istatistiği değerini (D istatistiği) bularak belirlemekte ve bu değer Copy-fm tarafından f_m olarak sunulmaktadır. Ayrıca, yine KS testine dayanılarak %95 güven aralığı da hesaplanmaktadır. Algoritmanın kontrol ve test örneklerinden heterozigot BAF olarak kabul ettiği değerlerin seçimi konusunda Bölüm 3.3.2'de detaylandırılan bazı özellikleri Copy-fm'in analiz algoritmasında hatalara yol açabilecek senaryolarda doğru sonuca ulaşmasını sağlamaktadır.

3.3.1. Farklı Mikrodizin Platformlarından Elde Edilen SNP Mikrodizin Verilerinin Ön Veri İşleme ile Analize Hazırlanması

Affymetrix için kontrol ve test örneklerine ait ChAS yazılımıyla elde edilen cychp uzantılı mikrodizin dosyaları Affymetrix'in APT (*Analysis Power Tools*) v2.11.4 adlı aracı ile R'da çalıştırılarak .txt formatında okunabilen dosyalara dönüştürülmüştür (Bkz. EK-5). Ardından hem Affymetrix hem de Illumina mikrodizininin elde edilen ve "kontrol örneği" olarak kullanılmasına kullanıcı tarafından karar verilen verilere ait .txt dosyaları R'da ön veri işleme basamağında *preprocess* fonksiyonu ile uygun tablo formatında birleştirilmektedir. Gene her iki platformdan elde edilen ve kullanıcı tarafından mCNV'ler açısından test edilmesine karar verilen "test örneği"ne ait veriler *preprocess* fonksiyonu ile aynı ön veri işleme basamağından geçirilmektedir. Bunun sebebi, kontrol ve test örneklerinin verilerinin Copy-fm'in *fm* fonksiyonu içerisine ayrı yollardan girip (*control, sample*) ayrı analiz basamaklarından geçerek işlenmesi gerekliliğidir. Birlikte aynı analizin içerisine girecek tüm kontrol ve test örnekleri için elde edilen veri aynı mikrodizin platformundan elde edilmiş olup BAF ve LRR değerlerini içermelidir. Copy-fm her çalıştırıldığında aynı kontrol örnekleri ve test örneği için ön veri işleme basamağının tekrar çalıştırılması gerekmekte, bu işlemin bir kez yapıp örneklerin kaydedilmesiyle aynı örnekler başka analizler için kullanılabilir.

Ancak ön veri işleme basamağında Affymetrix ve Illumina verilerinin işlenmesinde 2 fark bulunmaktadır. Birincisi, Affymetrix platformunda SNP problemleri için bir güven skoru (*confidence*) değeri sunuluyorken, Illumina için böyle bir verinin kendi analiz platformunda yer almamasıdır. Bu sebeple, ön veri işleme basamağında *preprocess* fonksiyonunda, Illumina verisi için “Confidence” isimli bir sütun oluşturulmamaktadır. İkinci fark ise, Affymetrix platformu direk CN değerini yansıtan “*SmoothSignal*” adlı bir veri sunmasıdır. Bu veri Affymetrix’in kendi algoritmasının CN değeri hesaplamak için LRR değerleri üzerinde yaptığı normalizasyon aşamasından farklı değerlendirme olup daha “öncü” LRR değerlerinden köken almaktadır. Bu sebeple işlenmiş LRR değerleri yerine Affymetrix için “SmoothSignal”, Illumina içinse “LRR” sütunu örneklerin tablosuna dahil edilmektedir.

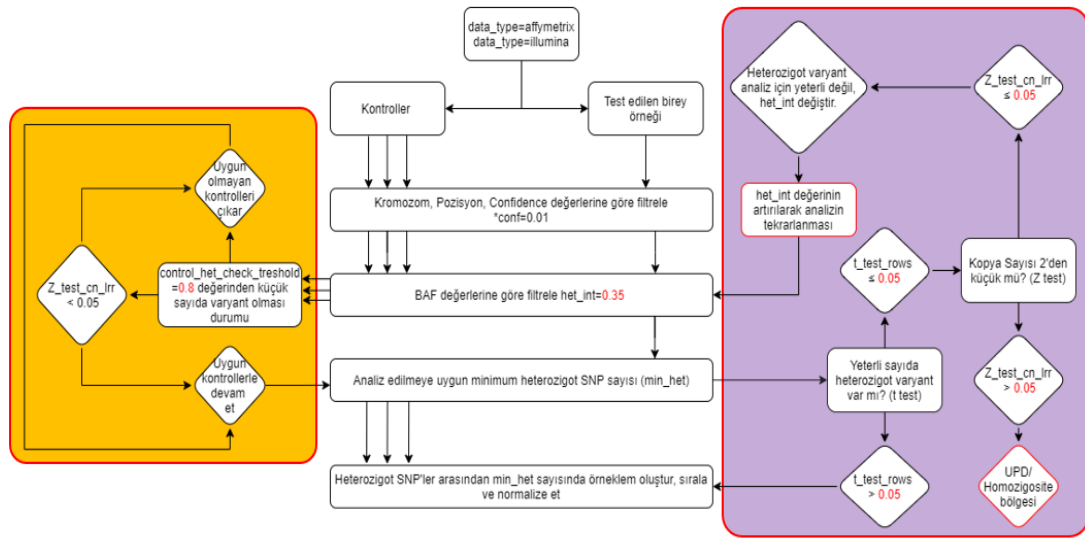
3.3.2. Kontrol ve Test Örneklerinde Analiz Edilebilir Varyantların Seçimi

Copy-fm heterozigot genotipte SNP’lerin BAF değerlerini değerlendirerek tüm analizlerini gerçekleştirmektedir. Bu nedenle, heterozigot SNP’lerden fakir genomik bölgelerde Copy-fm’in çalışmaması veya bazı düzeltmeler yaparak çalışması gerekmektedir. Bu nedenle, Copy-fm algoritmasında f_m hesaplaması ile ilgili işlemlere geçilmeden önce kontrol örnekleri ve test örneğinde heterozigot varyantlar sayılmakta ve farklı senaryolara göre Copy-fm işlemi durdurmakta veya bazı değişikliklerle devam etmektedir.

3.3.2.1. Kontrollerdeki Heterozigot Varyant Sayısının İncelenmesi

Copy-fm, heterozigot varyantlarla çalıştığı için hiç heterozigot varyantın bulunmadığı uzun LOH bölgeleri, *min_het* (analiz edilebilir minimum heterozigot varyant sayısı) değerinin çok düşük olmasına sebep olarak hesaplamaların yanlış sonuç vermesine yol açmaktadır. Bu amaçla, bir kontrol örneğinin LOH bölgesine denk gelmesi veya *germline* CNV barındırması durumunda bu kontrol örneğinin model oluşturma işlemleri başlamadan otomatik olarak analizden çıkarılması için bir algoritma oluşturulmuş ve Şekil 3.2 üzerinde Turuncu Bölge’de gösterilmiştir. Oluşturulan algoritma 2 yaklaşımı içermektedir:

- i. LRR değerleri kullanılarak elde edilen kopya sayısı (CN, *Copy Number*) değerlerinin hem t testi hem de z testi ile CN değerinin 2'den yani normalden farklı olup olmadığının test edilmesi.
- ii. Her bir kontrol örneğine ait heterozigot varyant sayısının kontrollerin ortalama heterozigot varyant sayısının belirlenen eşik değerin altında kalıp kalmadığının test edilmesi.



Şekil 3.2. Copy-fm’de kontrol ve test örneklerinin analize uygunluğunun değerlendirilmesi için geliştirilen iş akışı.

Tez çalışması kapsamında oluşturulan Copy-fm iş akışında verilerin analize uygunluğunun değerlendirme süreci görülmektedir. *Turuncu bölge*: Kontrol örneklerinin uygunluğunun değerlendirilmesine yönelik iş akışı. *Mor bölge*: Test örneğinin uygunluğunun değerlendirilmesine yönelik iş akışı. Kırmızı ile yazılan değerler Copy-fm oluşturulurken belirlenmiş ve bazıları kullanıcı tarafından da değiştirilebilen değişkenlerdir.

İlk test edilen olasılık CNV bölgelerinin tespitine yöneliktir ve heterozigot varyant seçimi yapılmadan önce, eğer platform Affymetrix ($data_type==\text{"affymetrix"}$) ise kontrollere ait *SmoothSignal* değerleriyle, Illumina ise LRR değerlerinden hesaplanan CN değerleriyle normal diploid kopya sayısı olan 2’den anlamlı bir farklılığın varlığını test etmektedir. Bu yaklaşımın işe yararlılığı hem z hem de t testleri ile denenmiş ve uygun olan test ile Copy-fm algoritması oluşturulmuştur. Buna göre,

test istatistiğinde CN değeri normalden farklı olan kontrol örnekleri analize dahil edilmemektedir.

İkinci test edilen olasılık ise, kontrol örneklerinin her birinde yeterli sayıda heterozigot varyant bulunup bulunmadığıdır. Bu amaçla, öncelikle kontrol örnekleri için heterozigot varyant seçimi $0,5 - \text{het_int} \leq \text{BAF} \leq 0,5 + \text{het_int}$ formülüne yerleştirilmek üzere $\text{het_int}=0,35$ parametresine göre yapılmaktadır. Bu değer heterozigot varyantların 0,5 etrafında dağılan BAF değerlerinden alınacak sınır değerlerine yönelik belirlenmiş varsayılan bir değerdir. Buna göre, 0,35 değeri sınır olarak belirlendiğinde, 0,15 ile 0,85 değerleri dışında kalan BAF değerleri homozigot varyantlar bu aradaki BAF değerine sahip varyantlar ise heterozigot varyantlar olarak kabul edilmektedir. Bir genomik lokusta tüm kontrol örneklerinin benzer sayıda heterozigot varyanta sahip olması beklendiğinden, incelenen genomik lokustaki her kontrol örneğinde bulunan heterozigot varyant sayılıp bunların ortalaması hesaplanmaktadır. Ardından her bir kontrol örneğine ait heterozigot varyant sayısı bu ortalamanın $\text{control_het_check_treshold}$ parametresi (0 ile 1 arasında) ile çarpımı sonucu bulunan sayı ile kıyaslanmaktadır. Bu sayının altında heterozigot varyant sayısına sahip kontroller analizden uzaklaştırılmakta, böylece sonraki işlemlere az varyant verisi ile devam etmenin önüne geçilmektedir. “ $\text{control_het_check_treshold}$ ” parametresi varsayılan değeri %80 olarak ayarlanmıştır ve değiştirme seçeneği kullanıcıya sunulmuştur. Her iki yaklaşımla da Copy-fm, atılan kontrol örneğine ilişkin “*Control #(kontrol numarası) is not suitable and will be ignored.*” uyarısı oluşturarak analizleri gerçekleştirmektedir.

Kontrol örneklerinin atılma fonksiyonunun başarı ile çalışıp çalışmadığı belirli bir lokusta LOH bölgesi içerdiği bilinen kontrol örnekleri kullanılarak test edilmiştir. 5 kontrol örneği ile gerçekleştirilen tüm analizlerde kontrol örneklerinden bir tanesinin LOH bölgesi içerdiği 5 farklı genomik lokustaki kontrol örneklerinin uzaklaştırılarak analize devam edilebilmesi fm fonksiyonunun bu özelliğinin test edilmesi için kullanılmıştır.

3.3.2.2. Test Örneğinde Heterozigot Varyant Sayısının İncelenmesi

Test edilen örnekte de heterozigot varyant sayısının seçilmesi het_int parametresine göre belirlenmektedir. Kontrollerle aynı şekilde bu değerin test örneği için de varsayılanı 0,35'dir. Ancak test edilen örnekte yüksek f_m değerlerinde mozaik delesyon bulunması ve het_int değerinin varsayılan 0,35 değeri ile analiz yapılması durumunda, bu örneğe ait aslında mozaik heterozigot olan varyantlar Copy-fm tarafından homozigot olarak değerlendirilmekte ve sonraki aşamalarda analizden çıkarılmaktadır. Yüksek oranlı delesyon mozaisizmlerinde ve test örneğindeki LOH bölgelerinde sorun yaratan bu durum, çok az sayıda heterozigot varyantla güvenilirliği düşük ve yanlış sonuçların elde edilmesine sebep olmaktadır. Bu sorunu aşmak amacıyla, test örneğinde de beklenenden az heterozigot varyant bulunmasının tespiti ve bu durumda Copy-fm'in yaptığı analizlerde kullanıcının uyarılmasını sağlayan bir algoritma oluşturulmuştur.

Bu algoritmaya göre test örneğindeki heterozigot varyantların uygunluğunun incelenmesi için, öncelikle test örneğinde kontrollere kıyasla t testinde anlamlı derecede düşük sayıda heterozigot varyant bulunup bulunmadığı test edilmektedir. Buna amaçla, test örneği için sayılan heterozigot BAF değerleri het_int değerine göre seçildikten sonra t testi kullanılarak kontrollere ait optimum BAF sayısı ile kıyaslanmıştır. Test örneğindeki heterozigot varyant sayısı yeterliyse Copy-fm uygun CNV modeline göre hesaplamalara devam etmektedir. Algoritma hesaplama yapmaya devam edeceği zaman kontrollerin yeterli sayıdaki heterozigot varyant sayısı ile test örneğinde belirlenmiş yeterli sayıdaki heterozigot varyant sayısından minimum olan sayıda varyantı tüm kontrol ve test örneklerinden rasgele seçmektedir. Tüm örneklerden eşit sayıda alınacak olan minimum heterozigot varyant sayısı min_het olarak tanımlanmaktadır.

Yeterli sayıda heterozigot varyant bulunmaması durumunda ise farklı senaryolar söz konusu olabilmektedir. Bunları ayırt etmek için z testi ile test örneğindeki LRR veya SmoothSignal değerlerinden hesaplanan CN değerinin 2'den farklı olup olmadığı araştırılmaktadır. Kopya sayısının 2 olduğu; ancak yetersiz sayıda

heterozigot varyant bulunan durumlar bir LOH bölgesi içinde bulunması olarak değerlendirilmekte ve Copy-fm kullanıcıyı “*Warning: UPD/homozygosity region*” mesajı ile uarmakta ve analizi durdurmaktadır. Kopya sayısının 2’den düşük olduğu ve test örneğinde yetersiz sayıda heterozigot varyant bulunan durumlar ise yüksek oranlı mozaik delesyon ($f_m \geq 0,8$) durumlarını işaret etmektedir. Bu durumda, Copy-fm, önce kullanıcıyı `het_int` parametresini değiştirmesi için “*Warning: heterozygous BAF values are too low to estimate mosaicism, change het_int*” şeklinde uarmaktadır. Ardından `het_int` parametresi 0,35 ile 0,5 arası bir değerle değiştirilerek tekrar hesaplandığında (ideal olarak 0,4, 0,45 ve 0,5) yüksek oranlı delesyon mozaisizmleri hesaplanabilmektedir. Bu amaçla kullanılan algoritma Şekil 3.2 üzerinde Mor Bölge’de gösterilmiştir.

Yüksek oranlı delesyon mozaisizmlerinin `het_int` değeri değiştirilerek hesaplanmasına alternatif bir yaklaşım, test örneğine ait tüm BAF değerlerinin ters çevrilerek (*invert*) homozigot BAF değerlerinin orta 0,5 hattına getirilmesiyle analize dahil edilmesini sağlayan bir algoritmadır. Bu yaklaşıma göre, $BAF > 0,5$ ise $BAF - 0,5$; $BAF \leq 0,5$ ise $BAF + 0,5$ formülü uygulanarak homozigot varyantlarla heterozigot varyantların yer değiştirilmesiyle `het_int` değiştirmeden analizin mümkün olabileceği düşünülmüştür. Copy-fm’e bu algoritma entegre edilerek “`het_int` değiştirme” ile “BAF ters çevirme” işlemlerinin sonuçları karşılaştırılmıştır. Copy-fm’in test örneklerindeki LOH bölgelerini ve yüksek oranlı mozaisizmi bulma başarısı bu bölgeleri içerdikleri bilinen örneklerde incelenmiştir.

3.3.3. Kontrollerden Mozaisizm Modellerinin Oluşturulması

Modellerin oluşturulması kontrol örneklerindeki BAF değerleri seçildikten sonra bunların sıralanması, normalize edilmesi ve satır ortalamalarından oluşturulan “*Point By Point Average*” (`point_by_point_avg`) değerlerinin elde edilmesinin ardından gerçekleştirilen randomizasyonla farklı senaryolara ait beklenen BAF değerlerine uygun simülasyon formüllerinin uygulanması aşamalarını içermektedir.

3.3.3.1. Mozaik Modelleri için BAF Değerlerinin Seçilmesi, Normalizasyon ve “Point By Point Average” Oluşturulması

Copy-fm kontrol örneklerinin analize uygun heterozigot BAF değerleri min_het olarak belirlendikten sonra her birinden BAF değerleri sayıp sıralamaktadır. Ardından kontrollerin BAF değerleri beklenen heterozigot BAF değeri olan 0,5'e normalize edilmekte ve normalize edilmiş BAF değerlerinin bulunduğu tüm satırların ortalaması alınarak point_by_point_avg adında tek bir *dataframe* oluşturulmaktadır. Bu aşamada oluşturulan *dataframe* artık ideal dizomi dağılımını (ID, ideal_disomy_dist) gösteren kümülatif dağılım fonksiyonunu teşkil etmekte ve mozaik olmanın olmadığı ideal dağılımı yansıtmaktadır.

3.3.3.2. Randomizasyon ve Mozaik Formülleri ile Model Kümülatif Dağılım Fonksiyonu Simülasyonlarının Oluşturulması

Mozaik CNV modellerinin oluşturulması işlemi kontrollerin heterozigot varyantları kullanılarak elde edilen ideal dizomi (ID) dağılımındaki BAF değerlerine dayanmaktadır. ID dağılımındaki BAF değerleri rastgele seçilerek yarısının azalma yönünde, yarısının da artma yönünde değiştirilmesi sonucu farklı mozaik modeller oluşturulmaktadır. Delesyon veya duplikasyon senaryoları için mozaik modelleri 0 ile 1 arasında (%0,0 ile %100,0 mozaik) olmak üzere 0.001 artışla 1000 farklı f_m değeri için oluşturulmaktadır. Farklı mozaik senaryolarındaki model CDF dağılımlarının oluşturulması için kullanılan formüller Tablo 3.2'de gösterilmiştir.

Tablo 3.2. Heterozigot varyantlara ait BAF değerlerinin f_m değerine göre beklenen ortalama değerinin hesaplanması.

Mozaik Genotip	CNV tipi	BAF değerindeki değişim yönü	Beklenen Ortalama BAF Değeri
A(B)	Delesyon	Azalma	$(1-f_m)/(2-f_m)$
B(A)	Delesyon	Artma	$1/(2-f_m)$
AB(A)	Duplikasyon (ortak)	Azalma	$1/(2+f_m)$
AB(B)	Duplikasyon (ortak)	Artma	$(1+f_m)/(2+f_m)$
AA(B)	Duplikasyon (heterodizomik)	Azalma	$f_m/(2+f_m)$
BB(A)	Duplikasyon (heterodizomik)	Artma	$2/(2+f_m)$

Mozaik genotip bölümünde parantez içerisinde yer alan alel hücreler arasında farklılık gösteren mozaik aleli işaret etmektedir.

Hem delesyon hem duplikasyon için f_m değerinin hesabını yapan Copy-fm'e eklenen önemli bir işlev, algoritmanın otomatik olarak delesyon veya duplikasyon

hesabı yapmaya karar verme yeteneğidir. Duplikasyon mozaisizmlerinin tespiti için test örneğinin incelenen genomik bölgedeki LRR değerine göre delesyon veya duplikasyon yönünde modeller oluşturulmaktadır. Buna göre, formüllerin delesyon mozaisizmi genotiplerine uygulanmasındaki ön koşul, test örneğinin ortalama LRR değerinin <0 olması, duplikasyon mozaisizmi genotiplerine uygulanmasındaki ön koşul ise >0 olmasıdır.

Duplikasyon mozaisizmi için hangi alelin mozaik duplikasyona katıldığına göre farklı senaryolar bulunmaktadır. Bu sebeple, tez sürecinde farklı çeşitleri olması sebebiyle daha karmaşık olan mozaik duplikasyonlardan model oluşturma formülleri ve hangi CNV türüne göre model geliştirileceğini belirleyen algoritmalar geliştirilmiştir. Bu senaryolardan normal ve izodizomik duplikasyona sahip hücrelerin oluşturduğu mozaisizm, “4’lü BAF şeması” şeklinde bir BAF dağılımı sergilemekte ve bu durum edinilmiş bir anomali veya konjenital trizomiden kurtulma senaryosu şeklinde gelişebilmektedir. Sadece konjenital trizomiden kurtulma sonucu ortaya çıkan, normal ve heterodizomik duplikasyona sahip hücrelerin oluşturduğu mozaisizm ise “6’lı BAF şeması” şeklinde bir BAF dağılımı oluşturmaktadır. Copy-fm tarafından oluşturulan duplikasyon modelleri her iki duplikasyon senaryosunu da dikkate almakta, modelin test örneğine daha yüksek uyum gösterdiği model ile f_m hesaplamasına gidilmektedir. Buna göre, test edilen birey örneğinde izodizomik “4’lü BAF şeması”na ait duplikasyonla heterodizomi gösteren “6’lı BAF şeması”nı ayırt etmek için f_m her iki şema için hesaplanarak, her iki hesaplama sonucu modellerle daha uyumlu CDF oluşturabilen (KS testinde D değerinin daha düşük olduğu) senaryo doğru duplikasyon mozaisizmi olarak kabul edilmektedir. “4’lü BAF” ve “6’lı BAF” şemaları için kullanılan model oluşturma formülleri Tablo 3.2’de sırasıyla izodizomik (ortak) ve heterodizomik şeklinde gösterilmiştir. Burada “6’lı BAF şeması” duplikasyon senaryosuna göre uygunluk tespit edilmiş olduğunda “*Warning: mCNV duplication with heterodisomy detected, change het_int to 0.17 and reanalyze*” uyarısıyla het_int değerinin kullanıcı tarafından 0.17’ye değiştirilmesiyle doğru f_m hesaplamasının yapılabilmesi amaçlanmıştır. Bu sayede duplikasyon için heterozigot

BAF deęerleri analiz edilmesi gereken sınırlar olan 0,33 ile 0,67 arasından seilebilecektir.

3.3.4. f_m Deęerinin Hesaplanması

Copy-fm kontrollerin analiz ařamasında olduęu gibi min_het sayısı kadar heterozigot varyant test rneęinden de seilerek, seilen BAF deęerleri 0,5'e normalize edip sıralamakta ve ardından f_m hesaplamasına hazır olan test rneęinin heterozigot BAF deęerleri kullanılarak test rneęinin CDF grafięi oluřturulmaktadır.

Test rneęine ait normalize BAF deęerleri ile her bir f_m deęeri iin kontrollerden simle edilmiř ID-CDF deęerleri arasındaki farklar (rezideller, *residuals*) kullanılarak models_difference adlı bir tablo oluřturulmaktadır. Buna gre, tam olarak 1000 adet modele ait kolon bulunan bu tablo iin her kolonda rezideller yer almaktadır. Daha sonra bu model farklarının kareleri alınarak kolon deęerlerinin yani rezidellerin toplamı alınmakta ve rezidellerin kareler toplamının minimum deęeri yaklařımıyla (*RSS, Residual Sum of Squares*) dięer adıyla aynı zamanda DANFIP metodunda da yer alan *minimum sum of residuals* (MSR) yntemi uygulanmaktadır (85). RSS yaklařımına gre Copy-fm tarafından sonu olarak verilen f_m deęeri, tm rezidel karesi toplamları arasında en dřk deęerin bulunduęu toplamının yer aldıęı modelin f_m deęeridir. rneęin en kk rezidel kareleri toplamı 0.8 modeli iin hesaplanıyorsa f_m deęeri %80 olarak belirlenmektedir.

Denenen bu yaklařıma ek olarak, ID-CDF modelleriyle test rneęi CDF daęılımlarını kıyaslayarak en iyi uyumun deęerlendirildięi bir uyum iyilięi (*goodness-of-fit*) yaklařımı olan iki rnekli Kolmogorov-Smirnov (KS) testinden yararlanılmıřtır. Bu test iin R'da "dgof" paketinden yararlanılarak test istatistięi uygulanmıřtır. KS testinin istatistiksel deęeri olan D deęeri iki model arasındaki fark deęerini yansittięinden, D deęerinin minimum olduęu dolayısıyla karřılařtırılan test rneęinin CDF'si ile model ID-CDF arasındaki farkın en az olduęu modelin f_m deęeri Copy-fm tarafından sonu olarak sunulmaktadır. Bylece tez kapsamında hem RSS hem de KS testini de uygulayarak iki farklı yaklařımın karřılařtırılması gerekleřtirilmiřtir. İki

kıyaslama arasındaki farklar tez kapsamında elde edilen veriler üzerinde incelenmiştir.

3.3.5. Mozaizm Fraksiyonu Güven Aralıklarının Belirlenmesi

Copy-fm tarafından hesaplanan f_m değeri model ile kıyaslamalarda en uyumlu ve ideal değerdir. Buna karşın her hesaplamalı yaklaşımda olduğu gibi bu ideal değer de bir güven aralığı içerisinde. Copy-fm bu güven aralığını hesaplamak için KS testinin hipotez sına yeteneğinden faydalanmaktadır. Bu şekilde KS testi ile yapılan dağılım karşılaştırması için H_0 hipotezi olarak “iki dağılım arasında fark yoktur” şeklinde kurulan hipotez KS testi ile sınanmıştır. α değeri 0,05 olarak kabul edilip test sonucu $p \leq 0,05$ olan D değerleri iki dağılım arasında fark olan ID-CDF modellerini göstermektedir. Buna göre, $p > 0,05$ olan D değerleri dağılımlar arası farkın istatistiksel olarak anlamlı olmadığı f_m değerleri olarak kabul edilmiştir. Bu yaklaşıma dayanarak H_0 hipotezinin reddedilemeyeceği sınırlar f_m değerlerinin %95 güven aralıkları olarak belirlenmiştir.

3.3.6. Copy-fm *preprocess* Fonksiyonunun Kullanımı

Copy-fm’in *preprocess* fonksiyonundan geçerek analize hazırlanmış veriler *fm* fonksiyonunda analiz edilirken öncelikle algoritmada kullanılacak veri sütunları okunmaktadır. Çalışmada Copy-fm ile yapılan bütün analizlerde analiz edilen mikrodizin platformuna göre farklı veri sütunlarına ihtiyaç bulunduğu için mikrodizin platformu *fm* fonksiyonunun çalışması sırasında bir argüman olarak girilmiştir. Buna ek olarak, BAF değerlerinin seçileceği kromozom ve pozisyon aralıkları ve *data_type* argümanı Affymetrix ise $conf < 0,01$ filtrelenmektedir.

3.4. Copy-fm’in İşlevlerinin Deneysel Veride Değerlendirilmesi

Tez çalışması kapsamında elde edilen deneysel olarak oluşturulan mozaik delesyon ve duplikasyon örneklerine ek olarak, Copy-fm’in etkinliğini göstermek için Affymetrix’in *web* sitesinden alınmış kamuya açık gerçek mozaik delesyon ve mozaik duplikasyon örnekleri de incelenmiştir (137). Ayrıca, trizomiden kurtulma şeklinde oluşmuş mozaik duplikasyon (trizomi 14) bulunan bir mikrodizin verisi Copy-fm’in duplikasyon senaryosunda çalışmasını gözlemlemek için incelenmiştir. Buna ek

olarak, 2 farklı genomik lokusta benzer oranlı mozaik delesyonlar içeren bir örnek ve heterozigot tam delesyona sahip bir birey örneği de delesyonlardan f_m hesaplama konusunda Copy-fm'in etkinliğini incelemek amacıyla değerlendirilmiştir. Bu gerçek klinik kan örnekleri sayesinde, deneysel olarak oluşturulmuş mCNV verisinden de Copy-fm'in performansı ortaya konmuştur.

Tüm istenen işlemlere yönelik kodları tamamlanmış Copy-fm'in performansı, gerçek deneysel veri ile teste tabi tutulmuştur. Öngörülme sorunlar ortaya çıktığı durumlarda, araç algoritması içerisinde bunları düzeltmek için gerekli değişiklikler yapılmıştır. Ayrıca, aşağıda belirtilen 4 yönden, Copy-fm deneysel veriler ile optimize edilmiştir:

3.4.1. Farklı Deneysel f_m Değerlerinde Copy-fm'in Etkinliğinin Test Edilmesi

Copy-fm'in f_m hesaplamalarındaki güvenilirliğini test etmek için, T0-T100 örneklerinde oluşturulan mozaik delesyonlara ait deneysel f_m değerleri ile Copy-fm'in hesapladığı f_m değerleri karşılaştırılmıştır. Bunun için, deneysel f_m değerinin, program tarafından hesaplanan f_m değerine ait güven aralığı içinde kalıp kalmadığı test edilmiştir. Burada özellikle yüksek oranlı mozaisizm için Copy-fm'de yapılan değişikliklerin başarısı değerlendirilmiş ve algoritma içerisinde uygulanan farklı test istatistiklerinin Copy-fm'in işlevlerine olan etkisi ortaya konmuştur.

3.4.2. Affymetrix ve Illumina platformlarından elde edilen verilerin kıyaslanması

Aynı kontrol ve test örnekleri ile 2 platformda çalışılan SNP mikrodizin verileri kıyaslanmıştır. Bu kıyaslama sonucunda, farklı platformlardan hesaplanan f_m değerlerinin birbirinin güven aralıkları içerisinde kalıp kalmadıkları incelenerek Copy-fm'in farklı platformlardaki tutarlılığı sınanmıştır. Kıyaslama yapılırken hem KS hem de RSS yaklaşımlarından elde edilen sonuçlar değerlendirilmiştir.

3.4.3. Copy-fm'in mozaik CNV bölgelerini saptama yeteneğinin denemesi

Tez çalışmasında oluşturulması hedeflenen aracın esas amacı doğru bir f_m değerinin hesaplanmasıdır. Ancak f_m değeri hesaplanması üzerinden mozaik CNV bölgelerinin keşfi denenmemiş bir süreçtir. Biz bu süreci, "*f_m-first*" (önce f_m , sonra

pozisyon) yaklaşımı olarak adlandırmaktayız. Bu amaçla oluşturulan *scanfm* fonksiyonu, kendisine sağlanan pencere boyutları içerisinde otomatik olarak her bir genomik bölge için f_m değeri hesaplamaktadır. İlk hesaplama için başlangıç bölgesi kromozomun 1. bazı olup bitiş bölgesi verilen pencere boyutuna veya kromozomun tamamına eşittir. Sonraki hesaplamalar için ise pencere boyutunun varsayılan değeri olan yarısı miktarında başlangıç ve bitiş pozisyonlarındaki artışlar şeklinde belirlenen genomik bölgelerdeki f_m değerleri ile her bir kromozomun sonuna kadar hesaplama sağlanmıştır. Copy-fm'in *scanfm* fonksiyonunun işlerliğini test etmek amacıyla, 5. ve 12. kromozomlarda bilinen CNV'lere sahip bir örnekte genom boyu mozaik CNV bölgelerinin tespiti denenmiş ve elde edilen sonuçlar incelenmiştir. İnceleme sırasında 3 Mbp, 5 Mbp ve 10 Mbp uzunluğunda olmak üzere farklı pencere boyutları kullanılmıştır. Bu şekilde bilinen mozaik CNV'nin olduğu ve olmadığı bölgeler veya tüm kromozom açısından tarama sağlanıp genom boyu tarama yapıp yapılamadığı test edilmiş, işlemlerin süresi kaydedilmiştir.

3.4.4. Copy-fm'in Performansının Diğer Biyoinformatik Araçlarla Kıyaslanması

Copy-fm'in, f_m hesaplamadaki başarısı, güven aralıklarını sunması ve genom boyu mCNV saptama gibi yetenekleri açısından mCNV'leri SNP mikrodizin verisinden saptayabilen ve erişimde olan DANFIP, MONTAGE ve MoChA biyoinformatik araçlarıyla da performans kıyaslaması yapılmıştır. Karşılaştırmada araçların deneysel olarak oluşturulmuş mozaik delesyon örneklerine ait f_m hesaplamaları ve tutarlılıkları, güven aralıkları sunup sunmadıkları, analiz platformları açısından esneklikleri, NGS verisine uyumları, genom boyu tarama yetenekleri ve analiz süreleri değerlendirilmiştir. Bu araçların ayrıca Linux içerisinde çalıştırılması için kullanılan komutlar EK-10'da yer almaktadır.

4. BULGULAR

4.1. Deneysel Mozaik Delesyon Örneklerinin Elde Edilmesi ve İncelenmesi

Copy-fm'in farklı f_m değerlerine sahip mozaik delesyon örneklerindeki etkinliği, oluşturulan deneysel örnekler kullanılarak Affymetrix ve Illumina platformlarından elde edilen verilerde incelenmiştir. Bu amaçla, öncelikle toplanan örneklerin analize uygunluğunu değerlendirmek için yaşla fizyolojik olarak ortaya çıkabilecek X kromozomu mozaizmi varlığı incelenmiş, Affymetrix ve Illumina platformlarında mikrodizin işlemleri gerçekleştirilmiş ve örnekler X kromozomundaki CNV'ler açısından görsel olarak incelenmiştir.

4.1.1. Test ve Kontrol Örneklerinde X Kromozomu Mozaizminin FISH ile Değerlendirilmesi

Toplanan periferik kan örneklerinden doğrudan preparasyon ile yapılan FISH çalışmalarının sonucu Tablo 4.1'de özetlenmiştir. Buna göre, tüm örneklerde 300 hücrede toplam 597 ile 603 arasında cinsiyet kromozomu bulunduğu saptanmıştır. Böylece, X kromozomu için FISH ile yapılan değerlendirmede delesyon ya da duplikasyon yönündeki f_m değerinin <1 olduğu sonucuna varılmıştır.

Bu durum, örneklerin hepsinin %5 ve üzerindeki f_m değerlerini test etmek için uygun olduğunu düşündürmüştür. Uygun olan örnekler mikrodizin işlemlerinde kullanılmak üzere yapılan DNA izolasyonu sonucunda elde edilen DNA konsantrasyonları da gene Tablo 4.1'de yer almakta ve toplam 7 örneğe ait normal ve anormal sayıda cinsiyet kromozomlarına sahip hücrelerin FISH görüntülerini gösteren birer saha ise Şekil 4.1'da sunulmaktadır.

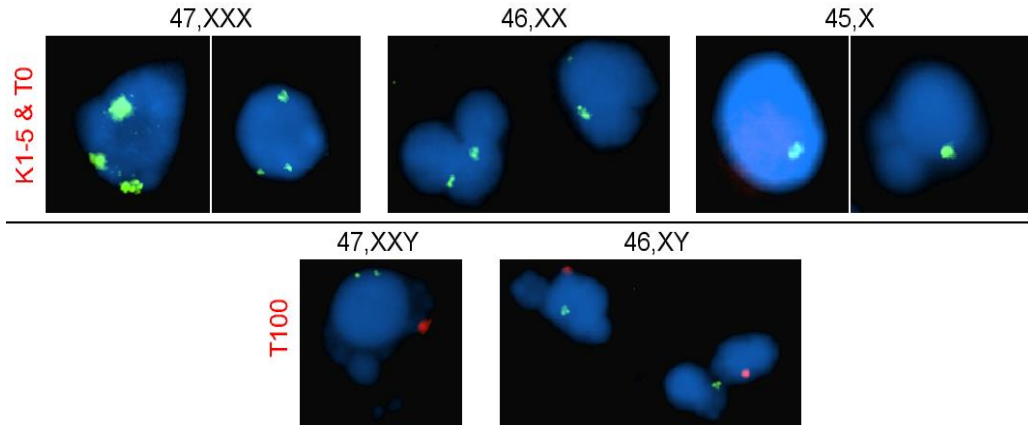
FISH analizlerinden hesaplanan f_m değerleri FISH f_m ve mikrodizin için beklenen f_m olmak üzere değerlendirilmiştir. Buna göre, K1 örneğinde 296 diploid hücre bulunurken 4 hücrede X kromozomu açısından sayısal farklılıklar gözlenmiştir. Bu durumda FISH ile hesaplanan f_m değeri %1,333 (4/300) olsa da bu örnek mikrodizin analizinde kullanılsa hesaplanacak f_m değeri, X trizomisi ve X monozomisi olan hücrelerin karışımı bir arada bulunduğu için farklı olacaktır. Buna göre, K1 örneğinin

FISH sonuçlarına göre X kromozomuna ait “mikrodizin için beklenen f_m ”, sayılan X kromozomları üzerinden $2 \cdot 2^{LRR}$ ($2 \cdot (\text{gözlenen CN} / \text{beklenen CN})$) formülünün uygulanması ardından hücre popülasyonu için hesaplanan kopya sayısının beklenen kopya sayısı olan 2’den ne kadar saptığının hesaplanması ile bulunabilir. K1 örneği için FISH analizinde 300 hücre içerisinde toplam 598 (296x2 (dizomi) + 3x1 (monozomi) + 1x3 (trizomi)) X kromozomu bulunmaktadır. 300 hücrede beklenen toplam X kromozomu kopyası ise 600 (300x2 (dizomi)) olduğu için, hücre popülasyonunun kopya sayısı 1,993’tür ($2 \cdot (598/600)$). Bu sayı beklenen kopya sayısı olan 2’den çıkarıldığında (2-1,993) aradaki fark 0,007 bulunmakta ve yüzde cinsinde ifade edilen mozaizizm oranı %0,7 olarak hesaplanmaktadır.

Tablo 4.1. Tez örneklemindeki 7 bireyin cinsiyet kromozomlarına yönelik FISH sonuçları ve elde edilen DNA konsantrasyonları.

Kromozom Sayısı	K1	K2	K3	K4	K5	T0	T100
Monozomi	3 (X)	3 (X)	1 (X)	2 (X)	2 (X)	2 (X)	-
Dizomi	296 (XX)	296 (XX)	297 (XX)	297 (XX)	296 (XX)	298 (XX)	299 (XY)
Trizomi	1 (XXX)	1 (XXX)	2 (XXX)	1 (XXX)	2 (XXX)	-	1 (XXY)
FISH f_m	%1,3	%1,3	%1,0	%1,0	%1,3	%0,7	%0,3
Mikrodizin için beklenen f_m	%0,7	%0,7	-%0,3	%0,3	%0	%0,7	%99,7
DNA Konsantrasyonu (ng/μl)	133,7	289	61,7	122,4	322	377,4	383,9

K1-5 kontrol örneklerini; test örneklerini oluşturmak için kullanılan örneklerden anneye ait olan T0 (%0 mozaik), oğluna ait olan ise T100 (%100 mozaik) olarak belirtilmiştir. Tüm örnekler için f_m delesyon mozaizmi senaryosuna göre hesaplanmıştır.

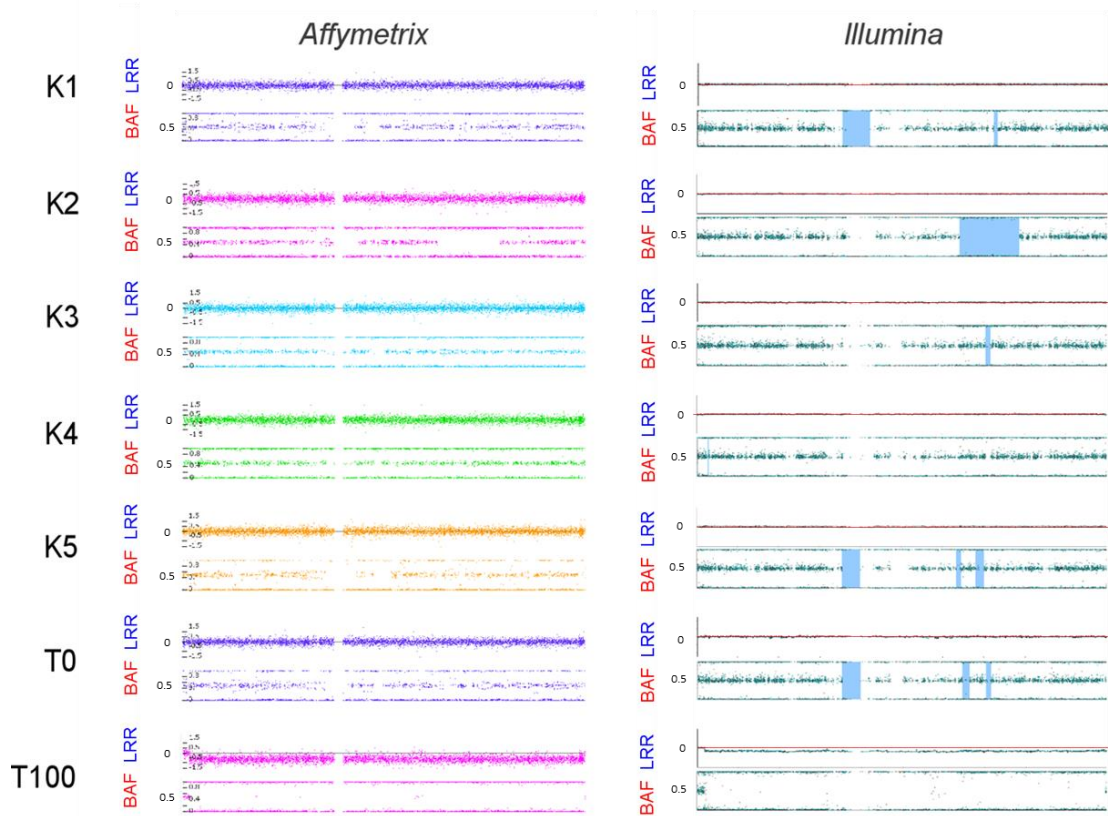


Şekil 4.1. Kontrol ve test örneklerinde cinsiyet kromozomlarına yönelik yapılan FISH analizleri.

K1-5 ile T0 ve T100 örneklerine ait normal sayıda ve temsili olarak anormal sayıda cinsiyet kromozomu içeren sahaları gösteren FISH görüntüleri. Mavi DAPI ile boyanan hücre çekirdeklerini, yeşil Xcen bölgesini, kırmızı ise Ycen bölgesini göstermektedir.

4.1.2. Test ve kontrol örneklerinde SNP mikrodizin verilerinin mikrodizin platformunda incelenmesi

SNP mikrodizin için yapılan DNA izolasyonunun sonucunda 260/230 oranı yaklaşık $\sim 1,5$ olan örnekler SNP mikrodizin analizine uygun olarak kabul edilmiştir. K1-5 şeklinde isimlendirilen kontrol örnekleri ile T0 ve T100 şeklinde isimlendirilen anne ve oğul örneklerine tez kapsamında temin edilebilen Affymetrix CytoScan Optima çipleri ile mikrodizin yapılmış, ayrıca aynı örnekler için Illumina HumanCytoSNP-12 v2.1 BeadChip mikrodizin işlemi için hizmet alımı gerçekleştirilmiştir. Tüm örnekler, X kromozomunda CNV varlıkları açısından incelenmiş; ancak beklendiği üzere klinik açıdan anlamlı bir CNV saptanmamıştır. Çalışmada deneysel olarak oluşturulan mozaik delesyon örnekleri için temel teşkil eden bu 7 bireyin X kromozomlarına ait 2 farklı platformdan elde edilen mikrodizin görüntüleri Şekil 4.2'de sunulmuştur.



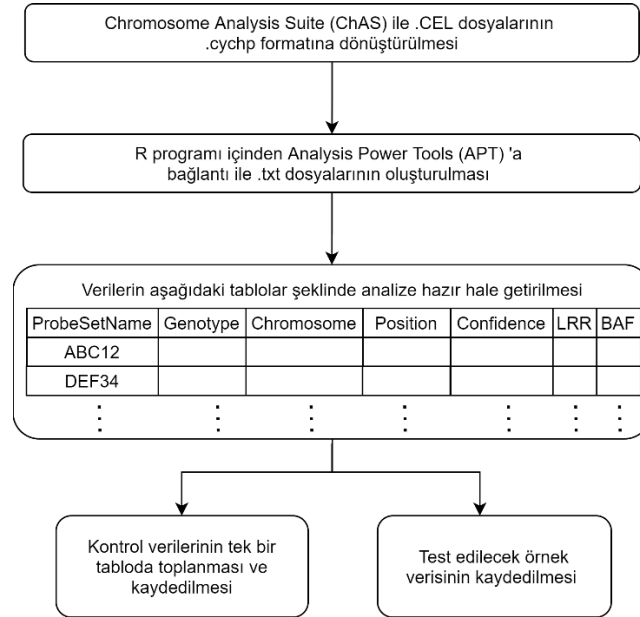
Şekil 4.2. Kontrol ve test örneklerinde X kromozomuna ait Affymetrix ve Illumina mikrodizin görüntüleri.

K1-5 ile T0 (anne) ve T100 (oğul) örneklerinde Affymetrix (solda) ve Illumina (sağda) ile elde edilen X kromozomlarına ait mikrodizin sonuçları. T100'ün erkek cinsiyette bir bireyden elde edilmiş olması sebebiyle LRR değerlerinin her iki platformda da 0'ın altında bulunduğu ve BAF değerlerinin heterozigot genotipi işaret eden bir SNP verisi içermediğine dikkat ediniz. Sağda Illumina'ya ait şekilde bulunan mavi alanlar ise LOH bölgelerini belirtmektedir.

4.2. Farklı Mikrodizin Platformlarından Elde Edilen LRR ve BAF Verilerinin Analize Hazırlanması

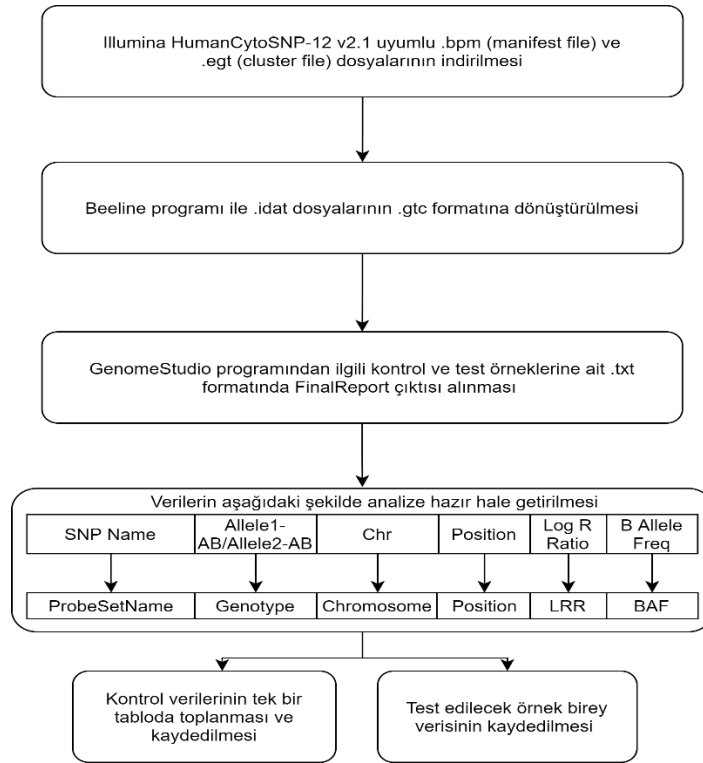
Copy-fm'in *preprocess* fonksiyonuyla ön veri işleme aşamasında işlenmiş Affymetrix ve Illumina mikrodizin verilerini Copy-fm formatına dönüştürmesine yönelik işlemler sırasıyla Şekil 4.3 ve Şekil 4.4'te gösterilmiştir. Elde edilen ham mikrodizin verisinden *preprocess* işlemi öncesinde BAF ve LRR değerlerinin hesaplanması için Affymetrix verilerinde ChAS v4.2.1 *Analysis Workflow* modülü (Bkz. EK-4) ve APT v2.11.4 (Bkz. EK-5) programları, Illumina verilerinde ise Illumina GenomeStudio ve Beeline (Bkz. EK-6) programları kullanılmıştır. Ön verinin işlenmesi

ile hazırlanan veri tabloları takip eden bulguların oluşturulması sırasında sorunsuz çalışan bir veri kaynağı sağlamaktadır.



Şekil 4.3. Copy-fm'de Affymetrix verileri için ön veri işleme iş akışı.

Mikrodizin cihazından elde edilen ham verileri içeren .CEL dosyaları .cychp dosyalarına dönüştürüldükten sonra APT aracılığıyla işlenmiş veri okunabilir formatta .txt dosyaları olarak kaydedilir. Copy-fm'in *preprocess* fonksiyonu bu aşamada R içinde gösterilen formatta "dataframe"ler oluşturur. Bunlar kontrol örnekleri için birleşik halde oluşturulurken, test edilecek örnek için bağımsız olarak elde edilmektedir ve böylece aynı kontroller ile farklı örneklerin test edilebilmesi mümkün kılınmaktadır. Ön veri işleme iş akışı, bu verileri .csv olarak saklamakta ve gelecekte farklı analizler için kullanıma hazır tutmaktadır.

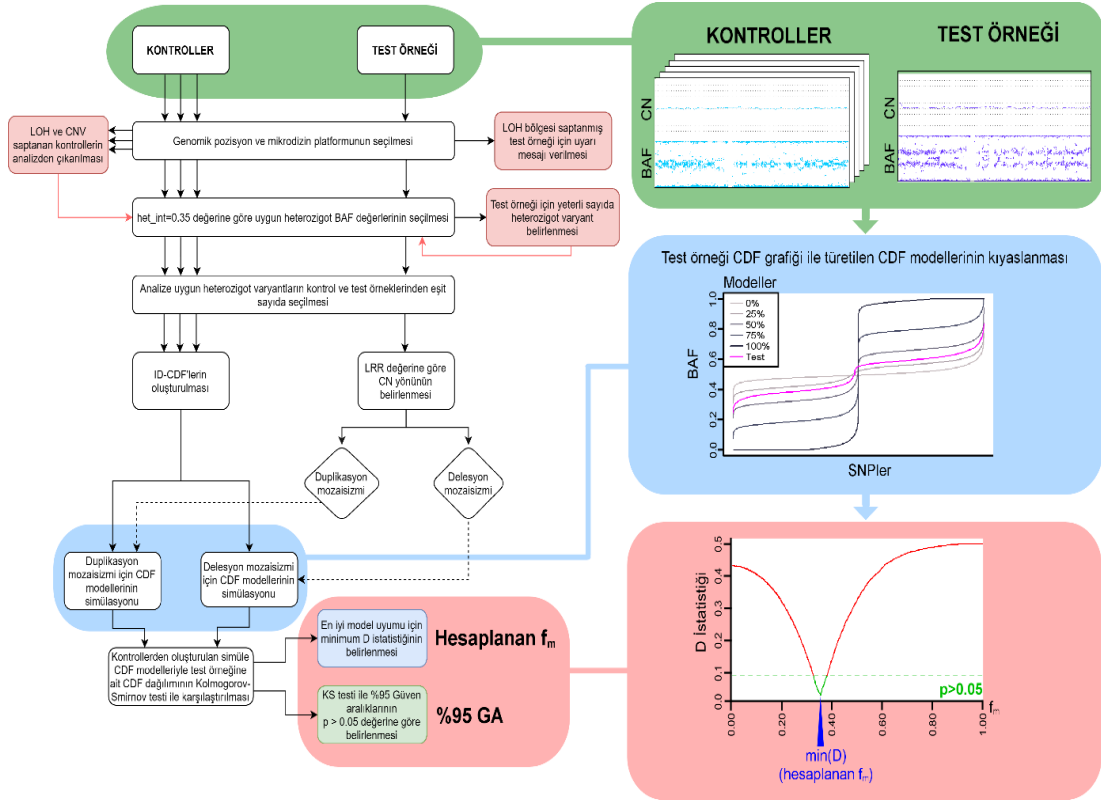


Şekil 4.4. Copy-fm’de Illumina verileri için ön veri işleme iş akışı.

Mikrodizin cihazından elde edilen ham verileri içeren .idat dosyaları .gvc dosyalarına dönüştürüldükten sonra GenomeStudio aracılığıyla işlenmiş veri okunabilir formatta .txt dosyaları olarak kaydedilir. Copy-fm’in *preprocess* fonksiyonu bu aşamada R içinde gösterilen formatta “*dataframe*”ler oluşturur. Illumina verisinde, Affymetrix verisinden farklı olarak *confidence* (güvenilirlik) değerine karşılık bir değer bulunmadığı için böyle bir sütun oluşturulmaz. Oluşturulan “*dataframe*”ler kontrol örnekleri için birleşik halde iken, test edilecek örnek için bağımsız olarak elde edilmekte ve böylece aynı kontroller ile farklı örneklerin test edilebilmesi mümkün kılınmaktadır. Ön veri işleme iş akışı, bu verileri .csv olarak saklamakta ve gelecekte farklı analizler için kullanıma hazır tutmaktadır.

4.3. Copy-fm *fm* fonksiyonu İş Akış Şeması

Copy-fm, kontrol örneklerinin verisini kullanarak bir test örneğindeki belirli bir genomik lokusun duplikasyon veya delesyon senaryolarında f_m değerini ve güven aralığını hesaplamaktadır. Buna yönelik oluşturulan Copy-fm iş akış şeması Şekil 4.5’te özetlenmiş, her aşamadaki verinin görünümü ise şeklin sağ panelinde belirtilmiştir. CDF dağılımları ve test istatistiğine yönelik sunulan grafikler T35 örneği kullanılarak elde edilmiştir.



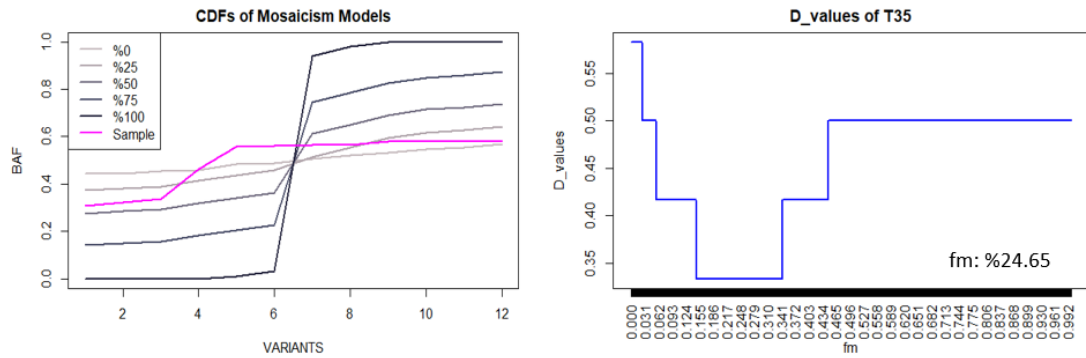
Şekil 4.5. Copy-fm iş akışı.

Copy-fm için oluşturulan iş akışının genel hatları solda görülmektedir. Yeşil ile işaretlenen kutucukta gösterilen ve kontrol ile test örneklerine ait bir genomik lokustaki CN ve BAF değerleri ile başlayan iş akışında kontrol örneklerinden elde edilen CDF'den delesyon veya duplikasyon yönünde CNV modelleri oluşturulmaktadır. Mavi kutucukta delesyon için oluşturulmuş mCNV modelleri (gri tonları) ve test örneğinin (pembe) bir karşılaştırması görülmektedir. Tüm modellere ait f_m 'ler ile test örneğinin Kolmogorov-Smirnov (KS) testi ile karşılaştırılmasından hesaplanan D istatistiği değerleri pembe kutucukta görülmektedir. Minimum D istatistiği, modellerle olan kıyaslamada test örneği için en iyi uyumun saptandığı noktaya karşılık gelmekte ve hesaplanan f_m değeri olarak belirlenmektedir. KS testinin istatistiksel olarak anlamlı fark göstermediği f_m değerleri ise %95 güven aralığını (%95 GA, $p > 0,05$, yeşil kesikli çizginin altında kalan kısım) göstermektedir. T35 için bu değer 38,6 (%95 GA: %38,4-%38,9)'dur.

4.4. Copy-fm Algoritmasında Kontrol ve Test Örneklerinin Analiz Edilebilirliğinin İncelenmesi

Kontrol örneklerinin CNV ve LOH açısından değerlendirilerek analize uygun olmadıklarının tespit edilmemesi halinde düşük sayıda heterozigot varyant kullanılarak Copy-fm algoritmasının hiç çalışmaması veya elde edilen sonuçların

güvensiz olması durumu söz konusudur. Bu duruma örnek teşkil eden bir sonuç, Affymetrix mikrodizin verisi kullanılarak K5 örneğinde bir LOH bölgesi olan X kromozomu üzerindeki chrX: 9,620,214-11,377,010 lokusunda bulunmaktadır. Buna göre, K5 dahil tüm kontrol örnekleri kullanılarak Copy-fm ile analiz gerçekleştirildiğinde K5 örneği analizden çıkarılmadığı için analiz sadece 12 varyant ile ilerlemektedir (Şekil 4.6). Bu durum f_m hesabının yanlış yapılmasına sebep olmakta ve T35 örneğinin f_m değeri %24,65 olarak saptanmaktadır. Ancak bu değer Şekil 4.5'te gösterilmiş olduğu üzere T35 örneği için 38,6 (%95 GA: %38,4-%38,9) olmalıdır. Üstelik az varyant sayısı ile yapılan analiz sonucunda test örneğine ait CDF dağılımının, Şekil 4.5'te sunulan CDF dağılımından oldukça farklı olduğu açıktır (Şekil 4.5). K5 örneği uzaklaştırılmadan yapılan analizin güvenilir olmadığı hesaplanan %95 GA değeri olan %0 - %100 ile de ortaya konmaktadır.



Şekil 4.6. X kromozomunda LOH bölgesi (chrX: 9,620,214-11,377,010) bulunan K5 kontrol örneği analizden çıkarılmadığında elde edilen sonuçlar.

Solda ID-CDF modelleri ve test örneğinin 12 heterozigot varyant kullanılarak oluşturulan CDF grafiği görülmektedir. Test örneğine ait dağılımın şeklinin (pembe) modellere ait dağılım şekilleri (açık gri-koyu gri) ile uyumsuz olduğuna dikkat ediniz. T35 örneği için Copy-fm'den bu şekilde hesaplanan f_m değeri %24.65, %95 GA ise %0 ile %100 arasında olup güvenilmezdir.

LOH bölgelerinden kaynaklanan sorunları aşmak için Copy-fm içerisinde kontrol ve test örneklerinde bu bölgelerden kaynaklı problemleri saptayan ve çözüm sunan algoritmalar oluşturulmuştur. Bu algoritmaların test edilmesi amacıyla LOH bölgesi içerdiği bilinen genomik lokasyonlarda Copy-fm denenmiştir.

Elde edilen Affymetrix mikrodizin verilerinde, 5 farklı kontrol örneğine ait bilinen homozigosite bölgelerinde ve hiçbir kontrol örneğinde homozigosite bölgesine denk gelmeyen bir bölgede, Copy-fm'in yeterli heterozigot varyant bulunmayan kontrol örneklerine ait veriyi analizlerden çıkarma becerisi t ve Z testleriyle test edilmiştir. Kontrollerin CN değerlerinin 2'den farklı olup olmadığı sorgusunda, t testi ile uzaklaştırılmaması gereken kontrollerin de analiz dışında bırakıldığı görülmüştür. Buna karşın, başka bir kıyaslama yöntemi olan ve yine ortalamanın anlamlılık testi için normal dağılan >30 veri noktası içeren örneklerde örneklem büyüklüğünün öneminin azaldığı z testi sadece uzaklaştırılması gereken kontrol örneklerini uzaklaştırmakta başarılı olmuştur (Tablo 4.2).

Tablo 4.2. Kontrollerdeki (K1-5) homozigosite bölgelerinin analizden çıkarılması ve test örneğinde z testi ile (T0) homozigosite bölgelerinin değerlendirilmesi.

Beklenen Sonuç	Genomik Bölge	f _m	%95 Güven Aralığı	Analize Alınmayan Örnekler
Hepsi LOH dışı	chrX: 5,000,000 - 15,000,000	%9,35	%0 - %22,7	-
K1 LOH	chr11: 29,125,237 - 32,268,181	%0	%0 - %18	K1
K2 LOH	chr11: 6,048,292 - 24,625,188	%0,1	%0 - %0,2	K2
K3 LOH	chr13: 95,270,220 - 99,822,327	%0	%0 - %16	K3
K4 LOH	chr2: 201,195,255 - 204,996,793	%0	%0 - %10	K4
K5 LOH	chr10: 72,420,902 - 78,431,966	%0,2	%0 - %8	K5
Test örneği LOH	chr4: 70.768.646 - 77.738.163	<i>Warning: UPD/homozygosity region</i>		-

Analize alınmayan kontrol örneklerinde gerçekten homozigosite bölgeleri bulunduğu Affymetrix mikrodizin platformundan görsel olarak doğrulanmıştır.

Buna göre, z testi kullanılarak algoritma uygulandığında her 5 test örneğinin biri için Affymetrix mikrodizin platformunda LOH bulunduğu bilinen genomik lokasyonlar test edildiğinde beklendiği şekilde LOH bölgesi içeren kontrol örnekleri uzaklaştırılmış, uzaklaştırılması gereken örnek bulunmayan bölge için gereksiz bir uzaklaştırma işlemi yapılmamıştır (Tablo 4.2).

Kontrol örneklerine ek olarak, test örneğinin de herhangi bir analiz sırasında LOH bölgesine denk gelme olasılığı bulunmaktadır. Bu durumu test etmek amacıyla oluşturulan algoritmada, Copy-fm tarafından gerçekleştirilen ve bir test örneğinin LOH bölgesine denk gelen analizde bu durumu açıklayabilecek 2 senaryoyu belirten “*Warning: UPD/homozygosity region*” uyarı mesajı ile f_m hesaplanmasının yapılmasından kaçınıldığı senaryo beklendiği üzere görülmüştür (Tablo 4.2). Bu şekilde düzenlenen program ile UPD ve homozigosite bölgelerinde yanlışlıkla f_m hesabı yapılmasının önüne geçilmektedir.

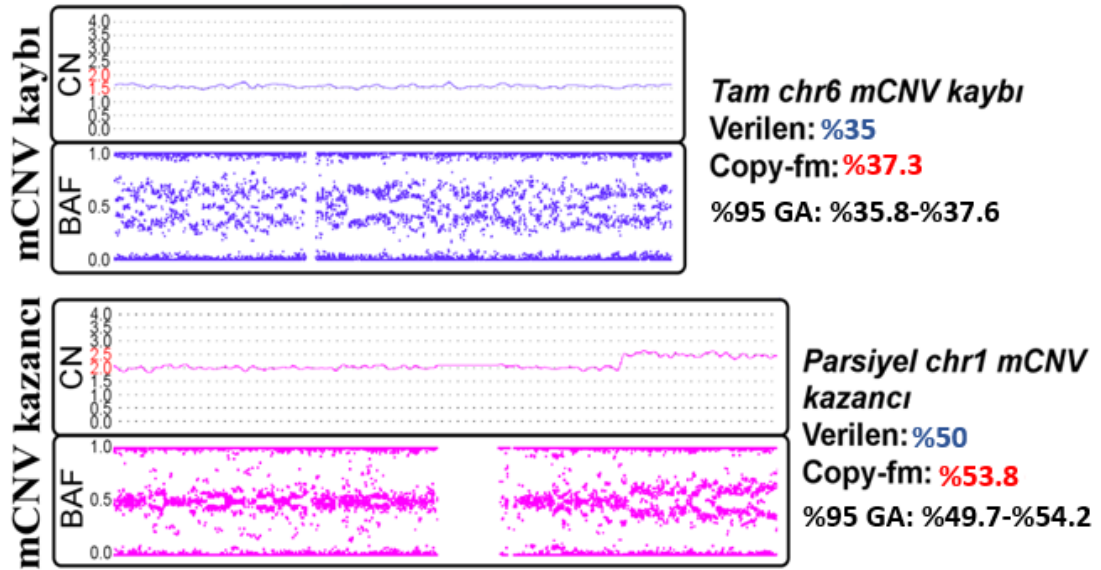
Bu önlemlere ek olarak, bir genomik lokasyonda hem kontrol örnekleri hem de test örneği için BAF sütununda hiçbir heterozigot varyant bulunmadığında Copy-fm hata vermektedir. Bu durum tasarım kökenli olarak SNP probunun denk gelmediği küçük bölgeler ve sentromer bölgeleri için geçerli olabilmektedir.

4.5. Gerçek mikrodizin verilerinde Copy-fm’in test edilmesi

Copy-fm’in işlevleri 3 farklı kaynaktan elde edilen veri kullanılarak incelenmiştir. Bu veriler arasında; 1) Kamuya açık mikrodizin verileri, 2) Farklı projelerde elde edilen ve bir CNV veya mCNV içerdiği bilinen mikrodizin verileri (1 mozaik delesyon ve 1 mozaik duplikasyon), 3) Tez kapsamında elde edilen kontrol mikrodizin verileri ile f_m değeri deneysel olarak oluşturulmuş test örneklerine ait mikrodizin verileri.

4.5.1. Kamuya Açık Verilerde Copy-fm’in Test Edilmesi

Affymetrix’in web sayfasından edinilen mozaik delesyon ve duplikasyon içerdiği bilinen ve Affymetrix tarafından *SmoothSignal* üzerinden hesaplanmış f_m değerleri bulunan mikrodizin verileri Copy-fm ile incelenmiştir. Şekil 4.7’de sunulan sonuçlara göre, Copy-fm tarafından hesaplanan f_m değerleri ile Affymetrix tarafından verilen f_m değerleri arasında %5’ten az farklılık bulunmaktadır.



Şekil 4.7. Kamuya açık verilerde Copy-fm'in denenmesi.

Yukarıda tam mozaik delesyon (chr6:1-170,919,482), aşağıda ise parsiyel mozaik duplikasyon (chr1: 188,843,662-249,224,684) verisinde bilinen f_m değerleri ve Copy-fm'in hesapladığı f_m değerlerinin karşılaştırılması görülmektedir. Copy-fm'in hesapladığı kırmızı ile işaretli değerler ile verilen değerler arasında <%5 fark bulunmaktadır.

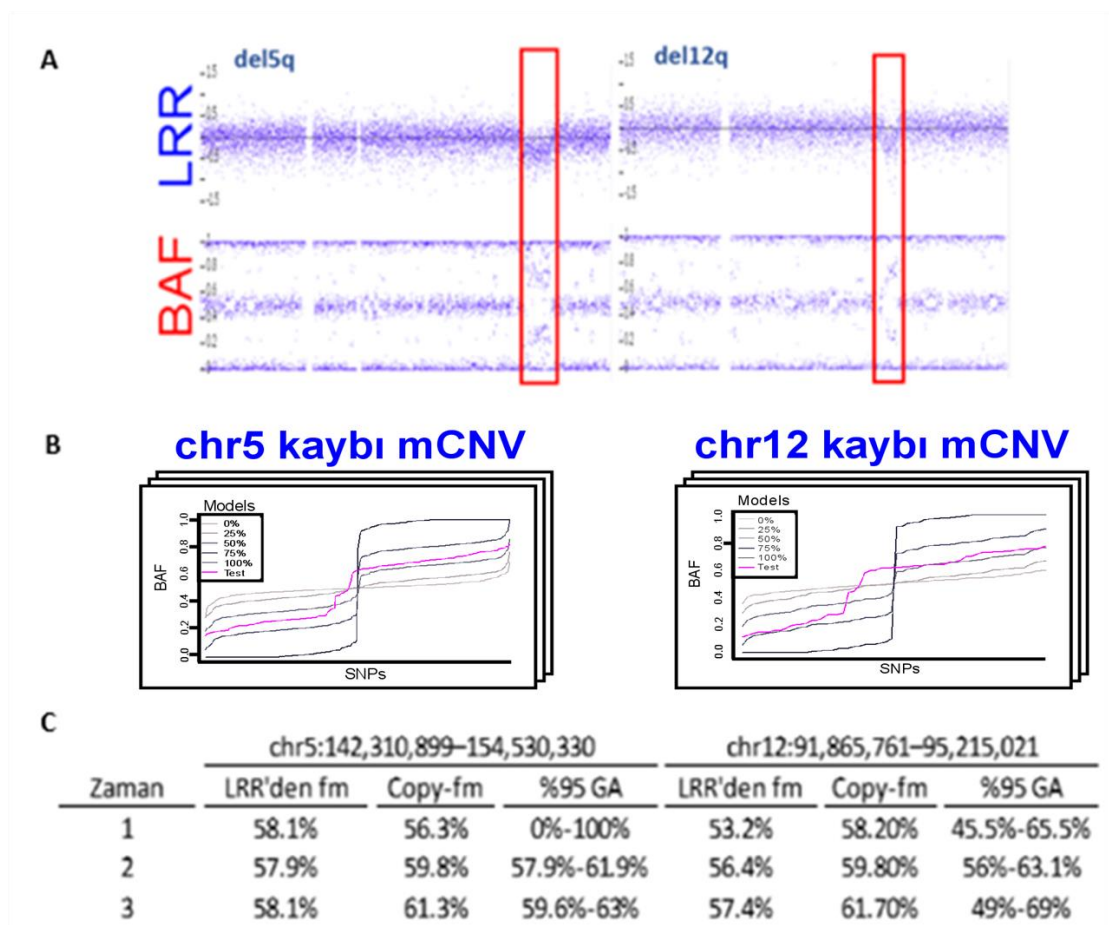
4.5.2. Gerçek Klinik Örneklerle Ait Mikrodizin Verilerinde Copy-Fm'in Test Edilmesi

Tez danışmanlığını yapan D. Öğr. Ü. Dr. Arda Çetinkaya'nın Türkiye yürütücülüğünü üstlendiği RiboEurope (TÜBİTAK Proje No: 319S062) ve RiboEurope projesinin öncüsü niteliğindeki EuroDBA (TÜBİTAK Proje No: 315S192) projelerinin çalışmaları sırasında elde edilen ve mozaik CNV'ler içeren mikrodizin verilerinde de Copy-fm test edilmiştir.

4.5.2.1. Tekrarlı Ölçümler Yapılan Mozaik Delesyon Örneklerinde Farklı Hesaplamalar Arasındaki Tutarlılığın İncelenmesi

Bu örneklerden bir tanesi 5q (chr5: 142.310.899–154.530.330) ve 12q (chr12: 91.865.761–95.215.021) bölgelerinde mozaik bir delesyon içermekte olup bu delesyonların klonal olarak aynı hücreler içerisinde bulunması sebebiyle mozaisizm oranlarının benzer olması beklenmektedir. Bu bireyden 5 yıl içerisinde 3 farklı zamanda alınmış periferik kandaki mozaik delesyonlara ait f_m değerleri Şekil 4.8'de

görülmektedir. Her 3 sonuçta da 5q ve 12q bölgeleri için hesaplanan f_m değerleri birbirine oldukça yakındır.

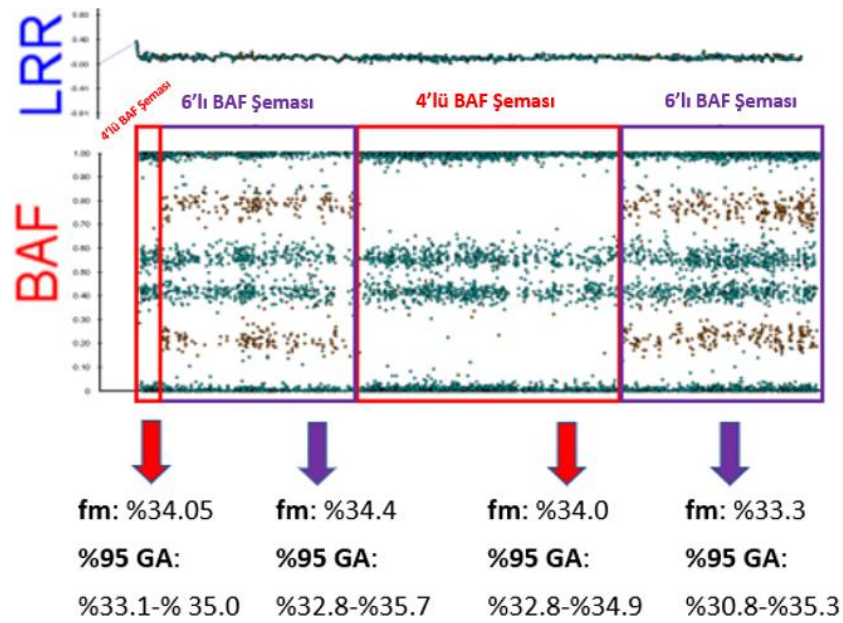


Şekil 4.8. Klonal mozaik delesyonlar içeren mikrodizin verisinde Copy-fm'in etkinliği.

A) Aynı Affymetrix mikrodizin verisinde birlikte görünen 5q ve 12q mozaik delesyonlarını da içeren tüm kromozoma ait LRR ve BAF dağılımları görülmektedir. B) Aynı bireyden ~2,5 yıl arayla 3 farklı zamanda elde edilen mikrodizin verileri Copy-fm ile analiz edilmiş ve CDF dağılımları çizilerek hesaplamalar yapılmıştır. Bu şekilde son mikrodizin verisinden oluşturulan CDF dağılımları görülmektedir. C) 5q (sol) ve 12q (sağ) bölgeleri için Copy-fm tarafından hesaplanan f_m değerleri ve güven aralıkları (GA) tablo şeklinde görülmektedir. Her üç veride de hesaplanan f_m değerlerinin birbirine %2'den yakın olduğuna dikkat ediniz. Buna ek olarak, veri kalitesinin düşüklüğü nedeniyle ilk veride %95 güven aralıkları genişken diğer iki veride \pm %3,5 güven aralığında değerler bulunduğu dikkat ediniz.

4.5.2.2. Farklı Mozaik Duplikasyon Örneklerinde Copy-fm'in Test Edilmesi

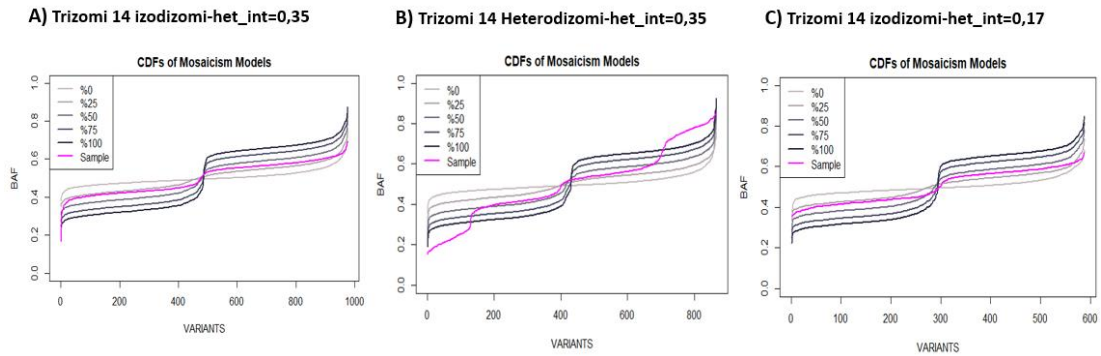
Bir diğer örnekte ise doğuştan tamamı mozaik trizomi 14'e (chr14:1-107.000.000) sahip bir bireye ait periferik kan örneğinden elde edilen mikrodizin verisi görülmektedir (Şekil 4.9). Bu veride ise birbirini izleyen iki farklı trizomi 14 senaryosunun peş peşe geldiği bir örnek görülmektedir. "4'lü BAF şeması" ve "6'lı BAF şeması" şeklinde tanımlanabilecek bu iki senaryodan "6'lı BAF şeması" sadece "trizomiden kurtulma" durumunda ortaya çıkabileceği için bu veri prenatal somatik mozaik trizomi 14'ü göstermektedir. Şekil 4.9'da görülen veride Copy-fm'in hesapladığı f_m değerleri, tamamı duplike olan mozaik trizomi 14'ün farklı farklı duplikasyon senaryoları ile uyumlu parçaları için beklendiği üzere birbirine oldukça yakındır.



Şekil 4.9. Farklı senaryolarda mozaik duplikasyon içeren mikrodizin verisinde Copy-fm'in etkinliği.

14. kromozomun tamamının duplike olduğu hücrelerin ve öplid hücrelerin karışımından oluşan "6'lı BAF şeması" (mor) ve "4'lü BAF şeması" (kırmızı) şeklinde farklı bölgeler içeren mozaik duplikasyona sahip bir bireyin Illumina mikrodizin verisi görülmektedir. Copy-fm dört farklı bölge için şemalarına uygun f_m ve güven aralığı (GA) hesaplamaları yapmış, bunun için "6'lı BAF şeması olan bölgelerde het_int değerini 0,17 olarak kullanmıştır. Sonuçta her dört bölge için hesaplanan f_m değerlerinin %30,8 ile %35,7 arasında değişerek birbirine çok yakın bulunduğu dikkat ediniz.

Farklı genomik bölgelerde “4’lü BAF” ve “6’lı BAF” şeması şeklinde mozaik duplikasyonların saptanması, Copy-fm’in farklı bölgeler için farklı modeller oluşturmasını gerektirmektedir. Bu nedenle, Copy-fm “4’lü BAF şeması” olan bölgelerde kanserli dokularda sıklıkla karşılaşılan duplikasyon (izodizomi) modelleriyle ve varsayılan het_int değeri ile f_m değerlerini hesaplarken (Şekil 4.10.A), “6’lı BAF şeması” olan bölgelerde CDF dağılımları varsayılan het_int değeri ile doğru f_m hesaplanmasına izin vermemektedir (Şekil 4.10.B). Copy-fm, “6’lı BAF şeması”nı otomatik olarak tespit etmekte ve het_int değerinin 0,17’ye değiştirilmesi konusunda kullanıcıyı uyarılmaktadır. Bu şekilde “6’lı BAF şeması” bulunan bölgeler için “het_int = 0,17” olarak ayarlandıktan sonra, Copy-fm başarılı olarak modellerle şeklen uyumlu CDF dağılımları oluşturmayı ve doğru f_m değeri hesaplamasını gerçekleştirmiştir (Şekil 4.10.C).



Şekil 4.10. “4’lü BAF şeması” ve “6’lı BAF şeması” şeklinde görünen farklı duplikasyonların Copy-fm tarafından ayırt edilmesi.

Test örneğine ait CDF pembe renkle, %25 artışla %0-%100 arası modeller ise sırasıyla açıktan koyuya doğru olan renklerle gösterilmektedir. A) Copy-fm tarafından izodizomi senaryosu seçildiğinde varsayılan het_int değeri olan 0,35 ile hesaplamalar yapılmakta ve test örneği için de modellerle uyumlu izodizomi CDF dağılımı görülmektedir. B) Mozaik duplikasyona ait varsayılan het_int değeri olan 0,35 ile CDF oluşturulduğunda “6’lı BAF şeması” nedeniyle modellere ait CDF dağılımları test örneğine ait CDF dağılımından şeklen farklı olmaktadır. Bu durum f_m ve güven aralığı hesaplarının doğru yapılamaması ile sonuçlanmaktadır. C) Copy-fm tarafından heterodizomi bölgesinin incelenmekte olduğu “6’lı BAF şeması”nın bulunması sayesinde tespit edildikten sonra het_int değerinin 0,17’ye değiştirilmesiyle test ve modeller arasındaki CDF uyumsuzluğunun giderildiğine dikkat ediniz.

4.5.3. Deneysel Mozaik Delesyon Örneklerinde Copy-fm'in Test Edilmesi

Copy-fm algoritmasının güvenilirliğini ve farklı mikrodizin platformlarındaki performansını test etmek amacıyla anne ve oğul çiftinden bir dizi deneysel mozaik delesyon örneği oluşturulmuştur. Çalışma kapsamında Affymetrix ve Illumina firmalarına ait birer mikrodizin platformu kullanılarak yapılan f_m hesaplamaları, Affymetrix platformunun sunduğu *SmoothSignal* verisinden dönüştürülmüş LRR'den hesaplanan f_m ile karşılaştırılmıştır. *SmoothSignal* verisi deneysel modellerin istenen şekilde oluşturulup oluşturulmadığının bir ölçüsü olarak değerlendirilmiş ve Tablo 4.3 ile Tablo 4.4'te "LRR'den f_m " sütunu olarak sunulmuştur. Buna göre, T5 modeli dışında deneysel olarak oluşturulan örneklerin hedeflenen f_m değerlerine yakın oldukları görülmektedir. Her 2 platform için yapılan hesaplamalar hem KS testi hem RSS yaklaşımı uygulanarak gerçekleştirilmiştir (Tablo 4.3 ve Tablo 4.4).

Tablo 4.3. Copy-fm'in deneysel olarak oluşturulan delesyon mozaizmlerine ait mikrodizin verileri ile f_m hesabı için KS yöntemi kullanılarak test edilmesi.

Model	Affymetrix				Illumina		
	LRR'den f_m	Copy-fm	%95 GA	Sapma	Copy-fm	%95 GA	Sapma
%0	%0	%3.35	%0-%5.3	+%3.35	%3.4	%0-%6.5	+%3.4
%5	%10.2	%0.9	%0-%7	-%9.3	%5.8	%5.5-%6.1	-%4.4
%10	%11.5	%3.9	%0-%8.7	-%7.6	%9.5	%8.4-%10.6	-%2.0
%20	%23.9	%15.8	%14.1-%18.9	-%8.1	%20.7	%19.5-%21.3	-%3.2
%35	%39.9	%38.6	%38.4-%38.9	-%1.3	%35.5	%35.2-%35.7	-%4.4
%50	%54.3	%59.6	%58.4-%61.5	+%5.3	%51.0	%47.4-%53.4	-%3.3
%65	%68.3	%71.7	%69.4-%75.2	+%3.4	%68.8	%67.2-%69.8	+%1.6
%80 *	%84	%83.3	NA	-%0.7	%81.4	NA	-%2.6
%90 *	%92.6	%90.5	NA	-%2.1	%91.0	NA	-%1.6
%95 *	%94.6	%94.8	NA	+%0.2	%95.8	NA	+%1.2
%100 *	%97.5	%98.4	NA	+%0.9	%96.1	NA	-%1.4

*Bu test örneklerinde het_int değeri varsayılan değerinden değiştirilerek sunulan sonuca ulaşılmıştır.

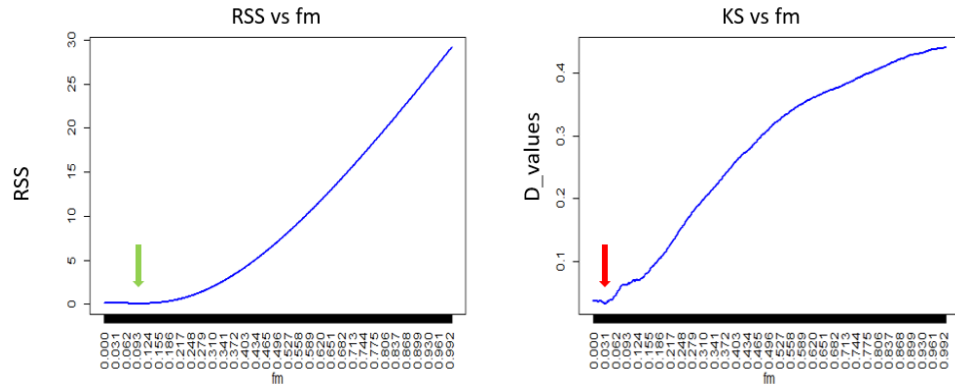
Tablo 4.4. Copy-fm'in deneysel olarak oluşturulan delesyon mozaisizmlerine ait mikrodizin verileri ile f_m hesabı için RSS yöntemi kullanılarak test edilmesi.

Model	RSS						
	Affymetrix			Illumina			
	LRR'den f_m	Copy-fm	%95 GA	Sapma	Copy-fm	%95 GA	Sapma
%0	%0	%9.3	%0-%5.3	+%9.3	%3.5	%0-%6.5	+%3.5
%5	%10.2	%0	%0-%7	-%10.2	%7.5	%5.5-%6.1	-%2.7
%10	%11.5	%3.8	%0-%8.7	-%7.7	%10.3	%8.4-%10.6	-%1.2
%20	%23.9	%17.3	%14.1-%18.9	-%6.6	%21.2	%19.5-%21.3	-%2.7
%35	%39.9	%39.0	%38.4-%38.9	-0.9	%36.2	%35.2-%35.7	-%3.7
%50	%54.3	%58.4	%58.4-%61.5	+4.1	%50.9	%47.4-%53.4	-%3.4
%65	%68.3	%68.4	%69.4-%75.2	+0.1	%67.7	%67.2-%69.8	-%0.6
%80 *	%84	%80.6	NA	-3.4	%80.1	NA	-%3.9
%90 *	%92.6	%89.6	NA	-3.0	%90.9	NA	-%1.7
%95 *	%94.6	%92.0	NA	-2.6	%96.0	NA	+1.4
%100 *	%97.5	%99.5	NA	+2.0	%98.2	NA	+0.7

*Bu test örneklerinde het_int değeri varsayılan değerinden değiştirilerek sunulan sonuca ulaşılmıştır.

4.5.3.1. Copy-fm f_m Hesaplamaları İçin KS ve RSS Yöntemlerinin Karşılaştırılması

RSS ve KS yöntemleri f_m hesaplamaları karşılaştırıldığında hem Affymetrix hem de Illumina platformundan elde edilen sonuçlarda RSS ile hesaplanan f_m değerlerinin %95 güven aralığı (%95 GA) dışında kalabildiği görülmüştür. Affymetrix için T0, T35 ve T65 örneklerinde; Illumina için ise T5 ve T35 örneklerinde bu durum açıkça görülmektedir (Tablo 4.3 ve Tablo 4.4). RSS yaklaşımında kullanıcıda karışıklık yaratabilecek bu durumu önlemek için Copy-fm'in f_m analizlerinde RSS yöntemi terk edilmiş, KS yöntemi ile f_m hesaplanmasına karar verilmiştir. RSS yöntemi ile %95 GA dışında kalan sonuçların önemini araştırmak için T0 örneği için Affymetrix platformundan elde edilen f_m hesabı süreçleri hem KS hem RSS yöntemleri için incelenmiştir (Şekil 4.11). Bu inceleme sonucunda farklı f_m modelleri ile karşılaştırma sonucu elde edilen "rezidüel karelerinin toplamı" değerlerinin daha yatay az bir varyasyon gösterdiği ve böylece ufak farklılıklardan bir sonuca vardığı görülmüştür. Buna karşın, KS istatistiğinin değeri olan D değerleri farklı f_m modelleri için hesaplandığında modeller arası D değeri farklılığı daha yüksek, böylece yapılan hesaplama daha keskindir (Şekil 4.11).



Şekil 4.11. Affymetrix platformundan T0 örneği için f_m hesaplaması sırasında kullanılan RSS ve KS yöntemlerinin karşılaştırılması.

Grafiklerde x ekseninde kontrol verilerine göre oluşturulan farklı f_m modelleri, y ekseninde ise solda her bir model için hesaplanan “rezidüel karelerinin toplamı” sağda ise her bir model için hesaplanan KS istatistiğinin sonucu olan D değerleri yer almaktadır. T0 örneği için çizilen bu grafiklerde beklenen f_m değeri %0 olup KS grafiğindeki keskin değişimler nedeniyle bu değer beklenen %0 değerine daha yakinken (kırmızı ok), RSS grafiğindeki yatay seyir nedeniyle beklenen %0 değerinden daha uzaktır (yeşil ok).

4.5.3.2. Copy-fm’in Farklı Mikrodizin Platformlarındaki Etkinliğinin İncelenmesi

Copy-fm’in her 2 platformda yaptığı hesaplamalara bakıldığında Illumina mikrodizin verisinden elde edilen sonuçların oluşturulan tüm deneysel örneklerde beklenen f_m değeri ile uyumlu f_m değerleri ortaya koyduğu ve %5 hata payı içinde sonuç verdiği görülmektedir (Tablo 4.3 ve Tablo 4.4). Aynı örnekler kullanılarak Affymetrix mikrodizin platformundan elde edilen sonuçlara göre ise, düşük oranlı mozaisizmlerde %5 hata oranının dışına çıkıldığı, maksimum KS için %9,3 (Tablo 4.3), RSS için %10,2 (Tablo 4.4) fark yaratan yakın sonuçlar elde edildiği görülmüştür. Bu nedenle, Affymetrix verisinde istatistiksel yaklaşım ayırt edilmeksizin Copy-fm’in daha az isabetli olduğu sonucuna varılmıştır.

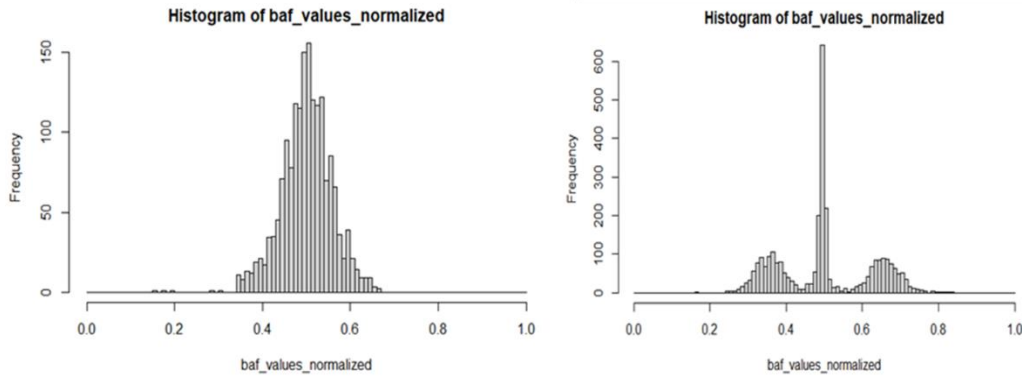
4.5.3.3. Yüksek Oranlı Delesyon Mozaisizmlerinde Copy-fm'in Değerlendirilmesi

Copy-fm tasarlanırken yüksek oranlı delesyon mozaisizmlerinde az sayıda heterozigot varyant sebebiyle analiz yapmanın zorlaşabileceği ve analiz yöntemi değiştirilmediğinde hatalı hesaplamalar yapılabileceği beklenmekteydi. Bu nedenle, Copy-fm algoritması yüksek oranlı delesyon mozaisizmlerini saptadığında kullanıcıyı het_int değerini değiştirerek analiz yapmaya yönlendirmemektedir. Bu sayede, bu uyarının verildiği %80 ve üzerindeki mozaik delesyon örneklerinde het_int değeri 0,4-0,5 arasında değiştirilerek Tablo 4.3 ve Tablo 4.4'te görülen doğru sonuçlara ulaşılmıştır.

Yüksek oranlı delesyon mozaisizmlerinde doğru sonuca ulaşmak için değerlendirilen alternatif "BAF ters çevirme" yaklaşımı, her 2 mikrodizin platformundaki yüksek oranlı mozaisizm delesyonu verilerinde (T80, T90, T95 ve T100) denenmiş ve elde edilen sonuçlar Tablo 4.5'te sunulmuştur. Elde edilen sonuçlar platform arasında tutarlılık arz etmemekte ve beklenen değerden sapmalar Tablo 4.3'te yer alan het_int değiştirme yaklaşımına göre daha fazla olmaktadır. Bu durumun sebebini incelemek için her 2 platformdan elde edilen T80 örneklerine ait BAF değerlerinin histogramları çizilmiş ve BAF değerleri ters çevrilen varyantların dağılımı incelenmiştir. Buna göre heterozigot BAF dağılımının daha belirgin olduğu T80 örneğinde Illumina verisinde Affymetrix verisine kıyasla çok daha fazla varyantın 0,5 değerinde olduğu görülmektedir (Şekil 4.12). Bu durum homozigot varyantlarında dönüştürülmekte ve analize dahil edilmekte olduğunu gösterdiği için "BAF ters çevirme" yaklaşımının kullanımından vazgeçilmiştir.

Tablo 4.5. Yüksek oranlı mozaik delesyon örneklerine ait verilerde "BAF ters çevirme" yaklaşımıyla elde edilen sonuçlar.

Model	Affymetrix				Illumina		
	LRR'den fm	fm	%95 GA	Sapma	fm	%95 GA	Sapma
T80	84%	91.15%	%88,1-%94.4	+%7,15	74.90%	NA	-%9,1
T90	92.60%	99.15%	%93 - %100	+%6,55	%87,65	NA	-%4,95
T95	94.60%	99.50%	NA	+%4,9	%91,7	NA	-%2,9
T100	97.50%	%99,95	NA	+%2,45	%93,6	NA	-%3,9



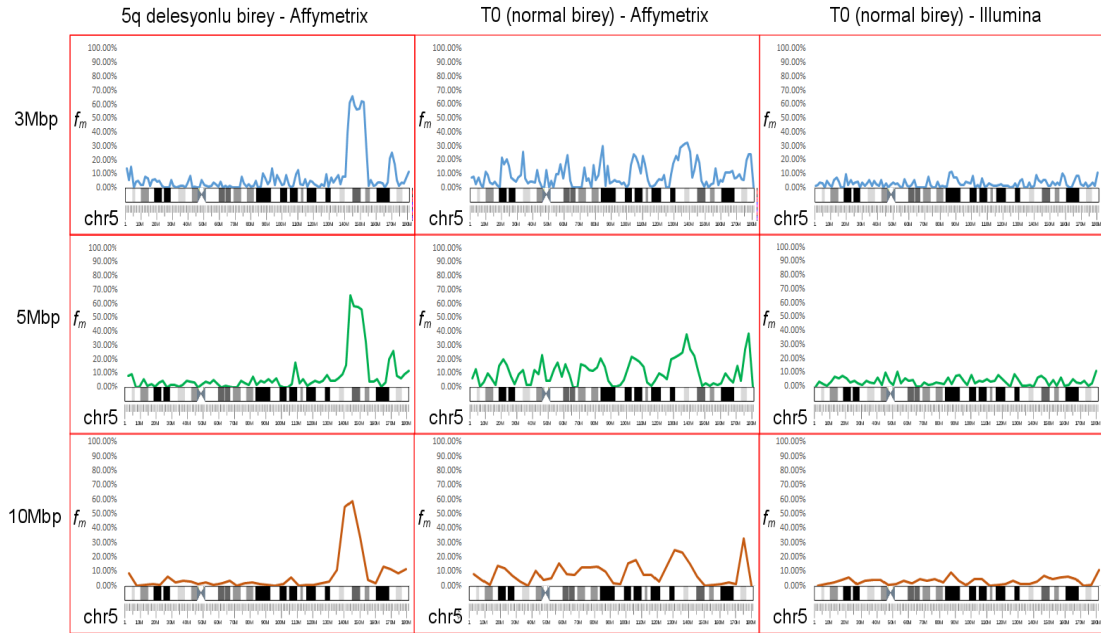
Şekil 4.12. T80 örneği için Affymetrix ve Illumina mikrodizin verilerinde “BAF ters çevirme” sonrası varyant BAF değerlerinin dağılımı.

T80 örneğine ait Affymetrix (solda) ve Illumina (sağda) platformlarından alınan varyantların “BAF ters çevirme” yaklaşımı sonrası BAF değerleri dağılımı histogram olarak görülmektedir. Histogramda 0,5 civarında yer alan varyantlar homozigot varyantları temsil etmekte ve bu varyantların da bir dağılım göstermesinden kaynaklanan zorluk nedeniyle “BAF ters çevirme” yaklaşımında heterozigot varyantların homozigot varyantlardan ayrımı zorlaşmaktadır.

4.6. Copy-fm’in Mozaik CNV Bölgelerini Saptama Yeteneğinin Denenmesi

Copy-fm aracının yazılmasındaki esas amaç, kullanıcı tarafından belirlenen bir genomik bölge için f_m hesabının yapılabilmesidir. Ancak isabetli f_m hesaplayabilen bu yaklaşımın genom boyu tarama yaparak mozaik CNV’leri bulabilmesinin de mümkün olduğu Copy-fm’in yazıldığı ilk andan itibaren düşünülmüştür. Bu nedenle, Copy-fm içerisinde *scanfm* fonksiyonu oluşturularak “önce f_m sonra lokasyon” (f_m -first) yaklaşımı ile 5q’da mozaik delesyonu olan bir bireyin 5. kromozomunun tamamının taranarak bu mozaik delesyonun bulunabilirliği değerlendirilmiştir. Bu amaçla, “kayan pencere” yaklaşımı ile yapılan incelemede farklı pencere boyutlarında (3 Mbp, 5 Mbp ve 10 Mbp) delesyona sahip ve normal bireylere ait mikrodizin verileri incelenmiştir. Şekil 4.13’te görülen sonuçlara göre, delesyon beklenen bireyde *scanfm* fonksiyonu ile beklenen delesyonun farklı pencere boyutlarında bulunabildiği görülmüştür. Bu verinin, delesyona sahip olmayan bir bireyden elde edilen farklı platformdaki veriler ile karşılaştırılması sonucunda da Affymetrix platformundan elde edilen veride arka plan gürültüsünün Illumina platformuna kıyasla genel olarak daha yüksek olduğu

dikkat çekmiştir (Şekil 4.13). Copy-fm'in 2 farklı platforma göre CNV bulunmayan bir bireyde (T0), 5. kromozom boyu tarama yapmak için gereken çalışma süreleri ise Tablo 4.6'da gösterilmiştir. Bu süreler Windows 10 işletim sistemli, Intel(R) Core(TM) i5-10210U, 1,6 GHz, 4 çekirdekli, x64 işlemciye ve 8 GB RAM'e sahip bir bilgisayarda elde edilmiştir.



Şekil 4.13. Copy-fm ile 5. kromozom boyunca mozaik CNV'lerin taranması.

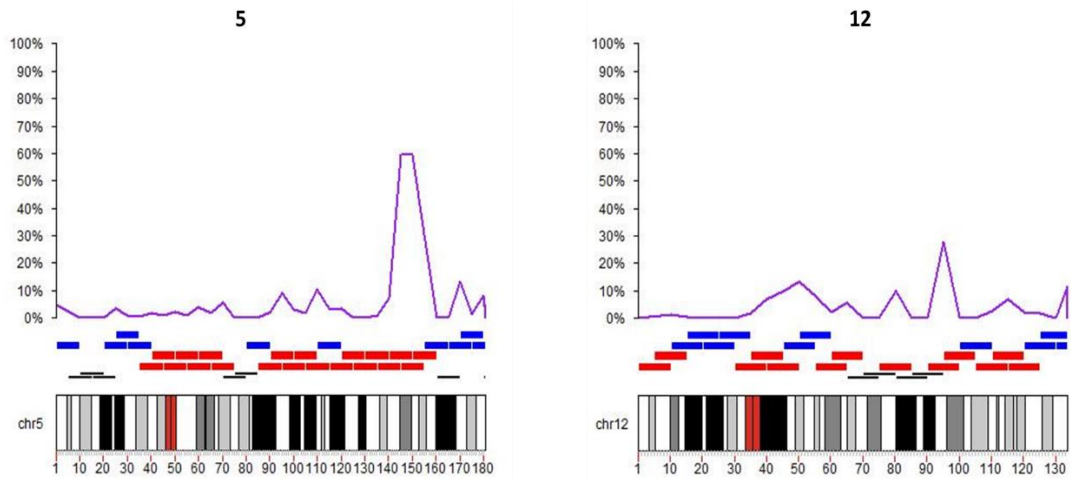
Copy-fm'in *scanfm* fonksiyonu kullanılarak gerçekleştirilen 5. kromozom boyu taramalar için üst, orta ve alt satırlarda sırasıyla 3 Mbp, 5 Mbp ve 10 Mbp pencere boyutları kullanılmıştır. İlk sütundaki grafikler 5q bölgesinde delesyon olan bireye ait Affymetrix verilerinden elde edilmiştir. Orta sütunda 5. kromozomda delesyon bulunmayan bir bireye ait Affymetrix verilerinden elde edilen *scanfm* sonuçları görülmektedirken, son sütunda aynı bireye ait Illumina verilerinden elde edilen *scanfm* sonuçları yer almaktadır. Bu sonuçlara göre chr5:142,310,899-154,530,330 bölgesinde ~%60 delesyon mozaizmi bulunan bireyde bu delesyonun Copy-fm tarafından saptanabildiğine dikkat ediniz.

Tablo 4.6. Bir CNV bulunmayan T0 örneğine ait Affymetrix ve Illumina verilerinde Copy-fm'in 5. kromozom taraması için geçen süreler.

Pencere boyutu	Affymetrix	Illumina
3 Mbp	02:04	02:58
5 Mbp	01:38	02:33
10 Mbp	01:21	02:26

Her 2 platformda T0 örneğine ait 5. kromozom taranmıştır.

Copy-fm'e ait *scanfm* fonksiyonunun genom boyu tarama yapabilme yeteneği de 10 Mbp pencere aralıkları ile 5q delesyonuna sahip bireyin Affymetrix mikrodizin verisi kullanılarak değerlendirilmiştir. Tüm genom boyu tarama işlemi yukarıda bahsedilen donanım alt yapısında 38 dk 23 sn sürmüş olup, 5. ve 12. kromozoma ait sonuçlar Şekil 4.14'te, 1-22 ve X kromozomları ise EK-7'de sunulmuştur. Sonuçların görsel hale getirilmesi için KaryoPlotR isimli R paketi kullanılmıştır (138).



Şekil 4.14. Copy-fm içinde yer alan *scanfm* fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyde 5. ve 12. kromozomun görünümü.

Copy-fm kullanılarak 5q ve 12q delesyonuna sahip birey için 10 Mbp pencere boyutunda *scanfm* fonksiyonu ile gerçekleştirilen tüm genom tarama işleminin sonucu 5. (sol) ve 12. kromozomlara (sağ) ait f_m grafikleri gösterilmiştir. Her iki kromozoma ait grafiğin x ekseninde p ucundan q ucuna genomik lokasyonlar gösterilirken; kromozom ideogramı üzerinde yer alan kırmızı, mavi ve siyah kutucuklar ise sırasıyla delesyon, duplikasyon ve f_m hesabı yapılamayan genomik pencereleri belirtmektedir. Kutucukların üzerinde yer alan mor renkli çizgi grafiği ise y ekseninde yer alan % cinsinden f_m değerlerini göstermektedir. Sonuçlara göre, *scanfm* ile mozaik delesyonun bulunduğu chr5:142.310.899-154.530.330 bölgesini kapsayan pencerelerde ~%60 delesyon mozaizminin saptanabildiğine ve 12q mozaik delesyonunun bulunduğu chr12:91.865.761-95.215.021 bölgesini kapsayan tek pencerede ise arka plan gürültüsünden yüksek bir pik değer (~%30) alındığına dikkat ediniz. 12. kromozomdaki mozaik delesyon küçük olmasına karşın saptanabilmesi; ancak Copy-fm tam sınırlar verilerek çalıştırılmadığında mozaik delesyon oranının gerçek değerinden düşük saptandığı dikkat çekmektedir.

4.7. Copy-fm'in Diğer Mozaik CNV Saptayan Yazılımlarla Kıyaslanması

Copy-fm, mozaik fraksiyonu (f_m) saptaması ile ön plana çıkan bir biyoinformatik araçtır. Literatürde aynı amaçla ön plana çıkan araç DANFIP olup diğer araçların öncelikli amacı mCNV bulunan genomik bölgelerin belirlenmesidir. Bu nedenle, mCNV'leri inceleyen biyoinformatik araçlardan BAFSegmentation, MAD/R-GADA, triPOD, affy2sv araçları f_m değeri hesaplayamamaları nedeniyle bu tez kapsamında karşılaştırma yapmak için kullanılmamıştır. Bunun yanında, triPOD aracı anne-baba-çocuk üçlüsünün verisine ihtiyaç duyması; ASCAT ve OncoSNP spesifik bir mikrodizin platformuna özgü olması ve tümör-kontrol karşılaştırma örnekleri değerlendirebilmesi; OncoSNP-SEQ, ascats ve MrMosaic sadece NGS verisi ile çalışabilmesi; DANFIP MS Excel'den çalışması ve pratik bir karşılaştırma imkanı sunmaması; SOMATICS ise erişilememesi sebebiyle Copy-fm ile karşılaştırılamamıştır. MONTAGE ve MoChA araçları kullanımları için önerilen yöntem (Bkz. EK-8) yardımıyla başarıyla çalıştırılmış ve etkinlikleri Copy-fm ile kıyaslanmıştır.

MoChA kullanılarak bu çalışma kapsamında elde edilen Illumina verilerinin incelenmesi gerçekleştirilerek MoChA'nın etkinliği ve Copy-fm ile kıyaslanması yapılmıştır (Tablo 4.7). Bu araç $<35 f_m$ oranlarında yapılan hesaplamalarda $\pm 5\%$ hata payı içerisinde güvenilir sonuçlar verirken f_m değeri yükseldikçe MoChA tarafından hesaplanan f_m değerleri gerçek f_m değerlerinden büyük sapmalar göstermektedir.

Tablo 4.7. Copy-fm ile MoChA araçlarının f_m hesaplamalarının karşılaştırılması.

Model	Affymetrix	Illumina			
	LRR'den f_m	Copy-fm	Sapma	MoChA	Sapma
%0	%0	%3.4	+%3.4	-	NA
%5	%10.2	%5.8	-%4.4	-	NA
%10	%11.5	%9.5	-%2.0	%7.4	-%4.1
%20	%23.9	%20.7	-%3.2	%19.1	-%4.8
%35	%39.9	%35.5	-%4.4	%33.2	-%6.7
%50	%54.3	%51.0	-%3.3	%40.3	-%14.0
%65	%68.3	%68.8	+%1.6	%50.6	-%17.7
%80	%84	%81.4	-%2.6	%44.4	-%39.6
%90	%92.6	%91.0	-%1.6	%40.0	-%52.6
%95	%94.6	%95.8	+%1.2	%38.4	-%56.2
%100	%97.5	%96.1	-%1.4	%41.0	-%56.5

MONTAGE aracı başarıyla çalıştırılmış; ancak bu çalışma kapsamında deneysel olarak elde edilen Affymetrix ve Illumina mikrodizin verilerinde X kromozomu için fm hesaplaması yapmakta başarılı olamamıştır. Copy-fm ile en uyumlu çalışma ilkelerine sahip 3 biyoinformatik aracın (MoChA, MONTAGE ve DANFIP) yetenekleri ile yapılan kıyaslama Tablo 4.8’de sunulmuştur.

Tablo 4.8. Copy-fm’in benzer biyoinformatik araçlar ile karşılaştırılması.

Araçlar	Deneysel Modelden f_m Hesabı	Güven Aralıklarının Belirtilmesi	Farklı Platformlara Adaptasyon	NGS Verisiyle Uyum	Genom boyu Tarama
Copy-fm	Evet	Evet	Evet	Evet*	Evet
MoChA	Evet**	Hayır	Hayır	Evet	Evet
MONTAGE	Hayır	Hayır	Evet	Evet	Evet
DANFIP	Evet	Hayır	Hayır	Hayır	Hayır

*Copy-fm NGS verisine uyumlu olarak tasarlanmıştır; ancak henüz NGS verisine uygun bir *preprocess* fonksiyonu oluşturulmamıştır. **Sadece Illumina verisinden hesaplama yapabilmektedir.

5. TARTIŞMA

CNV'ler, yapısal varyantlar içerisinde yer alan delesyon ve duplikasyonlarla genomik bölgelerdeki kopya sayılarını tek bir hücre içerisinde tam sayı olacak şekilde etkileyen değişikliklerdir. Bir hücre popülasyonu söz konusu olduğunda ise bir genomik bölge her hücrede farklı kopya sayılarında bulunduğu bu değişiklikler tüm hücre popülasyonu açısından tam sayı dışında kopya sayıları şeklinde ifade edilebilirler. Hücrelerin bir kısmında CNV olması sonucu ortaya çıkan bu durum, mozaik CNV'lerin sebebidir.

Özellikle son 20 yıldır mCNV'ler kanser genetiği çalışmalarında ön plana çıkmaktadır. mCNV'lerin saptanmasının kanser gibi hastalıklarda tanı ve tedavi açısından önemini kavramış olan birçok araştırmacı bu konuda çalışmalar yapmakta ve birçok ticari firma da 2000'li yılların başından beri özellikle CNV'lerin saptanmasına yönelik klinik test, cihaz ve programlar üretmektedirler (139). Ancak mCNV'lerin varlığının ve genomik lokalizasyonunun saptanmasının yanı sıra, tanı ve tedavinin seyrinde önemli role sahip bir değişken de mCNV'lerin oranını yansıtan mozaisizm fraksiyonu (f_m) olmaktadır (83). Literatür taramalarımız gösteriyor ki, f_m hesaplamasını başarabilmiş ve günümüze kadar ulaşabilmiş çok az yöntem mevcuttur (110, 140). Mevcut yöntemlerin tercih edilmelerini sağlayan çeşitli nedenler olsa da; hız, performans, yüksek doğruluk oranı ve kullanım kolaylığı gibi özelliklerin tamamını bir arada barındıran bir araç bulunmamaktadır. Bu sebeple bu tez çalışması kapsamında; güvenilir mikrodizin platformları olan ve yaygın kullanılan Affymetrix ve Illumina SNP mikrodizin verilerinden f_m hesaplayabilen, modifiye edilmiş bir DANFIP algoritmasından faydalanan, genomun farklı özelliklerinden kaynaklı pek çok sorunu analiz aşamasında otomatik olarak çözebilen ve R programlama dilinde yazılmış Copy-fm adı verilen bir biyoinformatik araç geliştirilmiştir.

Copy-fm aracı, temelde 3 fonksiyonun bir arada bulunduğu bir paket olacak şekilde tasarlanmıştır. Bu fonksiyonlardan; 1) *preprocess* fonksiyonu sonraki aşamalarda analiz için farklı kaynaklardan alınan BAF ve LRR değerleri hesaplanmış veriyi Copy-fm için analiz edilmeye uygun hale getirmektedir, 2) *fm* fonksiyonu f_m

değerini ve güven aralığını hesaplamaktadır, 3) *scanfm* fonksiyonu ise tekrarlı olarak *fm* fonksiyonunu kullanarak "*f_m-first*" yaklaşımı ile genom boyu tarama yapmakta ve mCNV bölgelerini aramaktadır.

Copy-fm günümüzde en sık kullanılan 2 mikrodizin platformu üreticisinin verileri kullanılarak denenmiş ve her iki platformda da başarılı olduğu gösterilmiştir. Illumina platformunda daha güvenilir sonuçlar üreten Copy-fm'in bu özelliğinin Illumina platformunda daha az varyasyona sahip bir BAF dağılımından kaynaklandığı düşünülmüştür. Bu durum hem elde edilen deneysel veriden X kromozomu delesyonu için yapılan f_m hesaplamalarında hem de *scanfm* fonksiyonunun denenmesi sırasında gözlenen arka plan gürültüsü incelendiğinde ortaya çıkmaktadır.

Copy-fm algoritması, DANFIP algoritması temel alınarak hazırlanmış; ancak en olası f_m değerini hesaplamak için RSS veya MSR gibi yaklaşımlar yerine KS istatistiğine başvurmuştur. KS istatistiğinin önemli bir özelliği dağılımları sadece ortalamaları ve varyasyonları ile değil dağılım şekilleri açısından da karşılaştırabilmesidir (141). KS testinin kullanılması sayesinde, Copy-fm'in yaptığı f_m değeri hesaplamalarının yanında bir güven aralığı da sunmak mümkün olmaktadır.

RSS yaklaşımı yerine KS testinin kullanılmasının bir avantajı da özellikle Affymetrix verisinde karşılaşılan ve yüksek varyanslı sonuçlar için RSS yaklaşımının güven aralıklarının dışında sonuç vermesinin önüne geçilmesidir. Bu şekilde, RSS yaklaşımı düşük oranlı mozaisizmler için KS istatistiğinden anlamlı sapmalar göstermektedir. Bunun sebebinin CDF dağılımları karşılaştırması yapılırken RSS yaklaşımının dağılım uçlarından ve asimetriden çok etkilenmesi KS'nin ise merkezi değerlere odaklanması olarak görülmüştür (142). Oluşturulan deneysel modeldeki CDF kıyaslamalarında aynı tip özelliği gösteren tüm dağılımlar için bu şekilde sonuç alınması KS'nin bu sebeplerle daha uygun olan yaklaşım olmasına karar verilmesine sebep olmuştur.

Copy-fm'in güven aralıklarını belirlemede yararlandığı KS testinde özellikle yüksek oranlı mozaisizmlerde p değerlerinin anlamsız olduğu dikkat çekmiştir. Bunun bir sebebinin yüksek oranlı mozaisizmlerde heterozigot varyantlarla homozigotların

birbirine karışması ve heterozigot SNP'ler için BAF değerlerinin homozigot SNP'lerin sahip olduğu 1 ve 0 değerlerine yakın olmasıdır. Bu durum dağılımların modellere iyi uyum göstermesine engel olmaktadır. Bu nedenle minimum D değeri saptanabilse bile p değerleri hiçbir mozaisizm modeli için uyumu işaret edemediğinden güven aralıkları belirlenememektedir. Ancak güven aralıklarıyla ilgili hem iterasyon hem de minimum D değerlerinin sınırları gibi çözümler getirmek denenebilecek yöntemlerdir. Bazı verilerde ortaya çıkan ve heterozigot varyantlarda yukarı yönlü değişimler ile aşağı yönlü değişimlerin eşit miktarda olmamasından kaynaklanan asimetri problemini ise SOMATICS algoritmasında yer alan "katlama" (*folding*) yöntemi ile çözmek mümkündür. *Folding* yöntemi, BAF dağılımlarının 0,5 etrafında simetrik olmasını avantaja çevirerek dağılımın 0,5 üstünde ve altında olan parçalarını bir tarafa toplayarak değerlendirmekte ve asimetri problemini ortadan kaldırmaktadır (103).

Copy-fm'i benzer araçlardan ayıran önemli özellikler bulunmaktadır. Bu özelliklerin başında bir kontrol grubu ile kıyaslama yaparak çalışması gelmektedir. Kontrol verileri ile kıyaslama yapması sayesinde Copy-fm farklı platformlara kolaylıkla uygulanma potansiyeli taşımaktadır. Copy-fm kapsamındaki *preprocess* fonksiyonuna küçük eklemeler yapılması ile pek çok farklı veri türü BAF ve LRR (veya CN) değerleri içermeleri kaydıyla Copy-fm'e uyarlanabilmektedir. Tez kapsamında Affymetrix ve Illumina şirketlerinden birer mikrodizin platformu kullanılarak Copy-fm başarıyla çalıştırılmıştır ve mevcut *preprocess* fonksiyonu bu şirketler tarafından üretilen her türlü mikrodizin platformunda analiz gerçekleştirebilecek şekilde tasarlanmıştır. Tez kapsamında kullanılan kontrol örneklerinin verileri frekans verisi şekline getirilerek kullanıcı tarafından bir kontrol verisi sunulmadan, Copy-fm içerisine gömülmüş kontrol verileri ile de analizlerin gerçekleştirilmesi gelecekte mümkün olabilecektir. Bu şekildeki bir yenilik, kullanıcı tarafından sağlanan bir kontrol verisine olan ihtiyacı ortadan kaldıracağı için kullanım kolaylığı sağlayacaktır.

Copy-fm mikrodizin verisi ile çalışan bir araç olarak tasarlanmış olsa da algoritmasının NGS verisinden yapılan analizlere uyarlanması mümkündür. Bunun için gerekli olan en önemli basamak *preprocess* fonksiyonunun NGS verisine yönelik

düzenlenmesidir. Tez sırasında, bir NGS verisi manuel olarak Copy-fm için uygun veri formatına dönüştürülmüş ve Copy-fm'in NGS verisinde de başarısına ilişkin kanıtlar elde edilmiştir. Buna karşın, *preprocess* aşaması manuel olarak gerçekleştirildiği ve Copy-fm'in NGS konusundaki yetkinliği sadece tek bir bireye ait veriden denendiği için bu sonuçlara yönelik veri gösterilmemiştir. Ancak umut vaat eden bu sonuçlar PennCNV yazılımının da yardımı ile Copy-fm'in farklı NGS platformlarından elde edilen verilerde verimli olarak çalışabileceğini işaret etmektedir (143).

Kontrol ve test edilen örneklerdeki heterozigot varyantlar Copy-fm'in çalışması için gerekli veri kaynağını oluşturmaktadır. Doğal genomik olaylar ve edinilmiş mozaik CNV'ler bu kritik veri kaynağının bazı genomik bölgeler için analizi tehlikeye atacak şekilde daralmasına sebep olabilmektedir. Özellikle, UPD veya ortak atadan aynı haplotipin homozigot kalıtılması kaynaklı LOH bölgeleri ile yüksek oranlı mozaik delesyon senaryosunda analiz edilecek varyant sayısındaki azlık sorun yaratmaktadır. Üstelik, bu durum sadece test edilen örneğin verisi ile ilişkili olmamakta, benzer genomik bölgeler CDF çizilmesinde kullanılan kontrol örneklerinin en az birinde bulunduğu da sorun teşkil etmektedir. Copy-fm ile ilgili ilk denemelerde heterozigot varyant azlığı ile ilgili süreç için kontrol basamakları eklenmeden analizler gerçekleştirildiğinde *scanfm* rastgele ve yanlış pozitif mozaik CNV bölgeleri tespit etmekteydi. Hem test edilen örnek hem de kontrol örnekleri için bu kontrol basamakları eklendiğinde ise heterozigot varyant azlığı nedeniyle ortaya çıkan bu sorunların önüne geçilmiş ve *scanfm* fonksiyonunun doğru çalışması sağlanmıştır. Bu sonuçlar, heterozigot varyant miktarı açısından yapılan analiz öncesi eklenen otokontrol basamaklarının önemini açıkça ortaya koymakta olup böyle bir özellik bildiğimiz kadarıyla daha önce literatürde yer alan biyoinformatik araçlarda bulunmamaktadır. Türkiye'de akraba evliliğinin sık olması nedeniyle LOH bölgelerinin genomlarda daha çok karşılaşılan bir durum olduğu düşünüldüğünde ise bu özellik ülkemize özel ayrı bir önem teşkil etmektedir.

Copy-fm ile hem delesyon hem de duplikasyon mozaisizminin saptanması aşamalarında ortaya çıkması öngörülen bazı sorunlara ise tez sürecinde çözümler

bulunmuştur. Mozaik delesyonların incelenmesi sırasındaki önemli bir sorun, yüksek oranlı mozaik delesyonların analizinde ortaya çıkmıştır. Bu bölgelerde het_int sınırları içerisinde az varyant kalması, Copy-fm'in heterozigot varyant kabulünü belirleyen sınır olan het_int değerinin yükseltilmesi ile çözülmüştür. Bu şekilde, Copy-fm'in yaptığı f_m hesapları yüksek oranlı mozaik bölgeler için de yüksek oranla isabetli hale getirilmiş ve bu durum elde edilen deneysel veride gösterilmiştir. het_int değerinin değiştirilmesine ek alternatif bir yaklaşım olan "BAF ters çevirme" algoritması da denenmiş; ancak hem Affymetrix hem de Illumina platformlarında bu yaklaşımla elde edilen sonuçlar isabetli olmadığı için yöntem terk edilmiştir.

Duplikasyon ile ilgili sorun ise farklı duplikasyon senaryolarının farklı BAF dağılımları oluşturabileceği gerçeğidir. Bu durum, kanser dokusundaki mCNV'lerin incelendiği durumlarda trizomiden kurtulma senaryosu mümkün olmadığı için sorun olmamaktadır. Buna karşın, trizomiden kurtulma gibi senaryoların da mümkün olduğu kalıtılan bir anomalinin prenatal somatik mCNV'ler ile düzeltilmesi şeklinde ortaya çıkan genetik değişiklikler ilgilenildiğinde bu tür senaryolara özel çözümler gerekmektedir. Copy-fm, farklı duplikasyon senaryoları sonucu ortaya çıkan farklı BAF dağılım şemalarını tespit edebilmekte ve çözümler sunabilmektedir. Aracın bu yeteneği ve etkinliği, mozaik Trizomi 14'e sahip bir bireye ait veri kullanılarak gösterilmiştir.

Copy-fm'in önemli bir kısıtlılığı mozaik UPD ve mozaik triplikasyon/tetraplikasyon gibi çok daha nadir mozaik değişiklikleri saptayamamasıdır. Ancak bu değişiklikleri analiz etmesi için basit değişikliklerle Copy-fm'i modifiye etmek mümkün olmakla beraber çok nadir değişiklikler olmaları sebebiyle kullanıcı talebi olması ve bu değişiklikleri içerdiği bilinen test örneklerine ait mikrodizin verileri bulunması halinde bu nadir mozaik değişiklikler için ilerleyen versiyonlarda eklemeler yapılabilecektir.

Copy-fm içerisinde yer alan *scanfm* fonksiyonu biyoinformatik aracın genom boyu tarama yeteneğini araştıran ve "kayan pencere" yaklaşımına göre çalışan bir fonksiyon olarak tasarlanmıştır. Tez çalışması sürecinde, tez planında olmamasına

karşın denenen “ f_m -first” yaklaşımını test eden bu fonksiyon, bilinen mozaik delesyonları bulmak konusunda beklentinin üzerinde bir başarı sağlamıştır. Delesyonu olmayan bireylerdeki arka plan gürültüsü dikkate alındığında, bu fonksiyon 3 Mbp uzunluğunda mozaik delesyonları Illumina platformu için $f_m > \%15$, Affymetrix platformu için $f_m > \%25$ değerlerinde saptayabilecek hassasiyette görünmektedir. Üstelik yapılan tüm analizlerde *scanfm* ile 5. kromozomun tamamının mozaik CNV’ler açısından taranması 3 dakikadan kısa, 10 Mbp pencere aralıkları ile genom boyu tarama yapılması ise 38 dakika 23 saniye sürmektedir. Genom boyu tarama Illumina platformunda da değerlendirildiğinde Copy-fm benzer bir hız performansı (42 dakika 39 saniye) göstermiştir. Bu sürelerin pencere boyutunun küçülmesi ve incelenen heterozigot varyant sayısının artması ile uzaması beklenmektedir.

Copy-fm içerisinde yer alan *scanfm* fonksiyonunun başarısının gelecekte eklentilerle iyileştirilmesi hedeflenmektedir. Fonksiyonun iyileştirme gerektiren yönlerinden biri tam sınır tayini için bir “sınır saptama” algoritmasının eklenmesidir. Bunun yanında, 2 senaryoda ihtiyaç duyulan *het_int* değişimlerinin otomatik hale getirilmesi *scanfm* fonksiyonunun çok daha etkin çalışmasını sağlayacak önemli bir geliştirmedir. Fonksiyonun hızlanması sağlayabilecek bir yenilik fonksiyonun birden fazla çekirdekte paralel olarak aynı anda işlev görmesidir ve bu amaca yönelik çeşitli R paketleri mevcuttur (144). Fonksiyonu hızlandırabilecek bir diğer ekleme de genotip bilgisinin elde edilemediği akrosentrik kromozomların p kolları, sentromerler ve gen çölleri gibi genomik bölgelerin analiz öncesinde *scanfm* içerisinde tanımlanarak bu bölgelere yönelik f_m hesaplamasının hiç yapılmamasıdır. Bu şekildeki bir analiz, incelenen bölge sayısını azaltarak *scanfm*’in genom boyu tarama yapma süresini kısaltacaktır.

Copy-fm’in literatürdeki benzer araçlarla karşılaştırılması ise NGS verisinin analizi dışında pek çok açıdan yeteneklerini göz önüne koymaktadır (Bkz. Tablo 4.8). Özellikle platformdan bağımsız analiz gerçekleştirilmesi ve acemi R kullanıcıları için bile basit bir kullanımı olması Copy-fm’i öne çıkaran özellikleri arasındadır. Bu özellikleri ile Copy-fm’in tedavi takibi, prognoz ve hastalık şiddeti belirlenmesinde sık

başvurulabilecek bir araç olarak bilim dünyasının ve tanı ile uğraşan klinik genetik uzmanlarının radarına takılarak kendine geniş bir kullanım alanı bulacağını öngörmekteyiz.

6. SONUÇ VE ÖNERİLER

1. Tez kapsamında mozaik kopya sayısı değişikliklerinin oranını saptamaya yönelik R programlama dilinde Copy-fm isimli bir biyoinformatik araç geliştirilmiştir.
2. Copy-fm mozaik delesyon ve duplikasyonlarda güvenilir bir şekilde f_m hesaplayabilmektedir.
3. Kontrol ve test verisinde heterozigot varyant azlığı nedeniyle ortaya çıkan hesaplama sorunları Copy-fm tarafından otomatik olarak fark edilip analiz buna uygun şekilde yönlendirilebilmektedir.
4. İzodizomik duplikasyonların yanısıra arka planda trizomiden kurtulma gibi karmaşık senaryoların olaylar Copy-fm tarafından kullanıcıya bildirebilmektedir.
5. Copy-fm, f_m değeri %25 kadar düşük olan CNV'lerde başarılı olarak çalışabilmektedir.
6. Copy-fm farklı platformlardan elde edilen mikrodizin verileri ile çalışabilmektedir.
7. NGS verileri ile Copy-fm'in çalışması mümkün olup, gelecekte bu özelliği test edilip geliştirilecektir.
8. Copy-fm içerisinde yer alan *scanfm* fonksiyonu ile " f_m -first" yaklaşımı kullanılarak Copy-fm ile genom boyu mozaik CNV taraması yapılabilmektedir. Bu yöntem mCNV bulunan genomik bölgelerin tam sınırlarının tespiti için geliştirilebilir.
9. Copy-fm kodlarının GitHub'da paylaşılması ve sonrasında bir R paketi haline getirilerek birçok araştırmacıya ulaşması, aracın geliştirilmesi açısından olumlu olacaktır.

7. KAYNAKLAR

1. Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, et al. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomed J.* 2021;44(5):548-59.
2. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med.* 2010;12:e8.
3. Notini AJ, Craig JM, White SJ. Copy number variation and mosaicism. *Cytogenet Genome Res.* 2008;123(1-4):270-7.
4. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature.* 2010;463(7278):191-6.
5. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet.* 2007;39(7 Suppl):S16-21.
6. Baugher JD, Baugher BD, Shirley MD, Pevsner J. Sensitive and specific detection of mosaic chromosomal abnormalities using the Parent-of-Origin-based Detection (POD) method. *BMC Genomics.* 2013;14(1):367.
7. Brown TA. *Genomes*, 2nd edition 2002 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://www.ncbi.nlm.nih.gov/books/NBK21134/>
8. Helena Mangs A, Morris BJ. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics.* 2007;8(2):129-36.
9. Collins FS, Fink L. The Human Genome Project. *Alcohol Health Res World.* 1995;19(3):190-5.
10. Wikipedia contributors. Reference genome [Internet]. Güncelleme: 26 Eylül 2022. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Reference_genome&oldid=1112508652
11. Nurk S, Koren S, Rhie A, Rautiainen M, Bizkadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science.* 2022;376(6588):44-53.
12. Holden AL. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques.* 2002;Suppl: 22-4, 6.
13. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-24.
14. Stranneheim H, Wedell A. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *Journal of Internal Medicine.* 2016;279(1):3-15.
15. Gibson G. Decanalization and the origin of complex disease. *Nature Reviews Genetics.* 2009;10(2):134-40.

16. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-73.
17. Mégarbané A, Ravel A, Mircher C, Sturtz F, Grattau Y, Rethoré M-O, et al. The 50th anniversary of the discovery of trisomy 21: The past, present, and future of research and treatment of Down syndrome. *Genetics in Medicine*. 2009;11(9):611-6.
18. Pös O, Radvanszky J, Styk J, Pös Z, Buglyó G, Kajsik M, et al. Copy Number Variation: Methods and Clinical Applications. *Applied Sciences*. 2021;11(2):819.
19. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7(2):85-97.
20. Harel T, Lupski JR. Genomic disorders 20 years on-mechanisms for clinical manifestations. *Clin Genet*. 2018;93(3):439-49.
21. Jane B. Reece LAU, Michael L. Cain, Steven A. Wasserman, Peter V Minorsky, Robert B. Jackson. *Genetik*. Gündüz E, Türkan İ, Editörler. Campbell Biyoloji. 9. Baskı. Ankara: PALME YAYINCILIK; 2013.
22. Li W, Wang X, Li S. Investigation of copy number variations on chromosome 21 detected by comparative genomic hybridization (CGH) microarray in patients with congenital anomalies. *Mol Cytogenet*. 2018;11:42.
23. Feuk L, Marshall CR, Wintle RF, Scherer SW. Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet*. 2006;15 Spec No 1:R57-66.
24. Kassam S, Meyer P, Corfield A, Mikuz G, Sergi C. Single Nucleotide Polymorphisms (SNPs): History, Biotechnological Outlook and Practical Applications. *Current Pharmacogenomics*. 2005;3:237-45.
25. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, et al. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Invest*. 2020;32(6):767-75.
26. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12(5):363-76.
27. Genome Aggregation Database (gnomAD) 2022 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://gnomad.broadinstitute.org/>
28. dbSNP, The Single Nucleotide Polymorphism Database 2022 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://www.ncbi.nlm.nih.gov/snp/>
29. 4.8: Mutation Types: LibreTexts BIOLOGY; 2022 [Internet]. Güncelleme: 6 Mart 2021. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: [https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_Introductory_Biology_\(CK-12\)/04%3A_Molecular_Biology/4.08%3A_Mutation_Types](https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_Introductory_Biology_(CK-12)/04%3A_Molecular_Biology/4.08%3A_Mutation_Types)
30. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015;16(3):172-83.

31. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 2012;148(6):1223-41.
32. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet*. 2009;18(R1):R1-8.
33. Lupski JR. Structural variation mutagenesis of the human genome: Impact on disease and evolution. *Environ Mol Mutagen*. 2015;56(5):419-36.
34. Du Q, de la Morena MT, van Oers NSC. The Genetics and Epigenetics of 22q11.2 Deletion Syndrome. *Frontiers in Genetics*. 2020;10:1365.
35. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-55.
36. Shlien A, Malkin D. Copy number variations and cancer. *Genome Med*. 2009;1(6):62.
37. Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nature Communications*. 2017;8(1):14366.
38. Mannan A, Muhsen IN, Barragán E, Sanz MA, Mohty M, Hashmi SK, et al. Genotypic and Phenotypic Characteristics of Acute Promyelocytic Leukemia Translocation Variants. *Hematol Oncol Stem Cell Ther*. 2020;13(4):189-201.
39. Hu L, Liang F, Cheng D, Zhang Z, Yu G, Zha J, et al. Location of Balanced Chromosome-Translocation Breakpoints by Long-Read Sequencing on the Oxford Nanopore Platform. *Frontiers in Genetics*. 2020;10:1313.
40. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*. 2005;77(1):78-88.
41. Carvalho CM, Zhang F, Lupski JR. Structural variation of the human genome: mechanisms, assays, and role in male infertility. *Syst Biol Reprod Med*. 2011;57(1-2):3-16.
42. Liu P, Lacia M, Zhang F, Withers M, Hastings PJ, Lupski JR. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am J Hum Genet*. 2011;89(4):580-8.
43. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009;10(8):551-64.
44. Potocki L, Bi W, Treadwell-Deering D, Carvalho CM, Eifert A, Friedman EM, et al. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet*. 2007;80(4):633-49.
45. Lupski JR, Stankiewicz P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet*. 2005;1(6):e49.
46. Bursed B, Zamariolli M, Bellucco FT, Melaragno MI. Mechanisms of structural chromosomal rearrangement formation. *Molecular Cytogenetics*. 2022;15(1):23.

47. Abyzov A, Li S, Kim DR, Mohiyuddin M, Stütz AM, Parrish NF, et al. Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nature Communications*. 2015;6(1):7256.
48. Jourdy Y, Chatron N, Fretigny M, Dericquebourg A, Sanlaville D, Vinciguerra C. Comprehensive analysis of F8 large deletions: Characterization of full breakpoint junctions and description of a possible DNA breakage hotspot in intron 6. *Journal of Thrombosis and Haemostasis*. 2022;20(10):2293-305.
49. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 2007;131(7):1235-47.
50. Moiseeva TN, Bakkenist CJ. Regulation of the initiation of DNA replication in human cells. *DNA Repair (Amst)*. 2018;72:99-106.
51. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *PathoGenetics*. 2008;1(1):4.
52. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet*. 2009;41(7):849-53.
53. Hastings PJ, Ira G, Lupski JR. A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. *PLOS Genetics*. 2009;5(1):e1000327.
54. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451-81.
55. Bodea GO, McKelvey EGZ, Faulkner GJ. Retrotransposon-induced mosaicism in the neural genome. *Open Biology*. 2018;8(7):180074.
56. Kallioniemi A, Visakorpi T, Karhu R, Pinkel D, Kallioniemi OP. Gene Copy Number Analysis by Fluorescence in Situ Hybridization and Comparative Genomic Hybridization. *Methods*. 1996;9(1):113-21.
57. Grigoryan A, Dougherty E, Bubendorf L, Kononen J, Kallioniemi O. Rapid analysis of gene copy numbers by FISH and automated spot counting from stacked confocal images. *Proceedings of SPIE - The International Society for Optical Engineering*. 2000;3924.
58. Levisky JM, Singer RH. Fluorescence in situ hybridization: past, present and future. *J Cell Sci*. 2003;116(Pt 14):2833-8.
59. Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet*. 2003;73(6):1261-70.
60. Khojasteh M, Lam WL, Ward RK, MacAulay C. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics*. 2005;6(1):274.
61. Ali H, Bitar MS, Al Madhoun A, Marafie M, Al-Mulla F. Functionally-focused algorithmic analysis of high resolution microarray-CGH genomic landscapes

- demonstrates comparable genomic copy number aberrations in MSI and MSS sporadic colorectal cancer. *PLOS ONE*. 2017;12(2):e0171690.
62. Vallespín E, Palomares Bralo M, Mori MÁ, Martín R, García-Miñaur S, Fernández L, et al. Customized high resolution CGH-array for clinical diagnosis reveals additional genomic imbalances in previous well-defined pathological samples. *American Journal of Medical Genetics Part A*. 2013;161(8):1950-60.
 63. ATIA T-A. Overview of genetic causes of recurrent miscarriage and the diagnostic approach. *BIOCELL*. 2019;43(4):253-62.
 64. Gresham D, Dunham MJ, Botstein D. Comparing whole genomes using DNA microarrays. *Nat Rev Genet*. 2008;9(4):291-302.
 65. Lin CF, Naj AC, Wang LS. Analyzing copy number variation using SNP array data: protocols for calling CNV and association tests. *Curr Protoc Hum Genet*. 2013;79:Unit 1.27.
 66. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-74.
 67. Staaf J, Vallon-Christersson J, Lindgren D, Juliusson G, Rosenquist R, Höglund M, et al. Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. *BMC Bioinformatics*. 2008;9:409.
 68. Affymetrix. CNAT 4.0: Copy Number and Loss of Heterozygosity Estimation Algorithms for the GeneChip® Human Mapping 10/50/100/250/500K Array Set 2007 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://www.affymetrix.com/support/technical/whitepapers/cnat_4_algorithm_whitepaper.pdf
 69. Sun W, Wright FA, Tang Z, Nordgard SH, Van Loo P, Yu T, et al. Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res*. 2009;37(16):5365-77.
 70. Winchester L, Yau C, Ragoussis J. Comparing CNV detection methods for SNP arrays. *Briefings in Functional Genomics*. 2009;8(5):353-66.
 71. Wikipedia contributors. Multiplex ligation-dependent probe amplification [Internet]. Güncelleme: 8 Eylül 2022. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Multiplex_ligation-dependent_probe_amplification&oldid=1109122151.
 72. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*. 2018;122(1):e59.
 73. Escaramís G, Docampo E, Rabionet R. A decade of structural variants: description, history and methods to detect structural variation. *Brief Funct Genomics*. 2015;14(5):305-14.
 74. Ye K, Hall G. Structural Variation Detection from Next Generation Sequencing. *Journal of Next Generation Sequencing & Applications*. 2015;S1:007.

75. Gabrielaite M, Torp MH, Rasmussen MS, Andreu-Sánchez S, Vieira FG, Pedersen CB, et al. A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. *Cancers*. 2021;13(24):6283.
76. Guan P, Sung WK. Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods*. 2016;102:36-49.
77. Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A. Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life*. 2022;12(1):30.
78. Kriegova E, Fillerova R, Minarik J, Savara J, Manakova J, Petrackova A, et al. Whole-genome optical mapping of bone-marrow myeloma cells reveals association of extramedullary multiple myeloma with chromosome 1 abnormalities. *Scientific Reports*. 2021;11(1):14671.
79. Sahajpal NS, Barseghyan H, Kolhe R, Hastie A, Chaubey A. Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses. *Genes*. 2021;12(3):398.
80. Chen Y, Chen C. DNA copy number variation and loss of heterozygosity in relation to recurrence of and survival from head and neck squamous cell carcinoma: a review. *Head Neck*. 2008;30(10):1361-83.
81. Žilina O, Koltšina M, Raid R, Kurg A, Tõnisson N, Salumets A. Somatic mosaicism for copy-neutral loss of heterozygosity and DNA copy number variations in the human genome. *BMC Genomics*. 2015;16(1):703.
82. Nait Saada J, Kalantzis G, Shyr D, Cooper F, Robinson M, Gusev A, et al. Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nature Communications*. 2020;11(1):6130.
83. Vlachos A, Farrar JE, Atsidaftos E, Muir E, Narla A, Markello TC, et al. Diminutive somatic deletions in the 5q region lead to a phenotype atypical of classical 5q-syndrome. *Blood*. 2013;122(14):2487-90.
84. Papavassiliou P, York TP, Gursoy N, Hill G, Nicely LV, Sundaram U, et al. The phenotype of persons having mosaicism for trisomy 21/Down syndrome reflects the percentage of trisomic cells present in different tissues. *Am J Med Genet A*. 2009;149a(4):573-83.
85. Markello TC, Carlson-Donohoe H, Sincan M, Adams D, Bodine DM, Farrar JE, et al. Sensitive quantification of mosaicism using high density SNP arrays and the cumulative distribution function. *Mol Genet Metab*. 2012;105(4):665-71.
86. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44(6):651-8.
87. Freed D, Stevens EL, Pevsner J. Somatic mosaicism in the human genome. *Genes*. 2014;5(4):1064-94.

88. González-del Angel A, Estandia-Ortega B, Gaviño-Vergara A, Sáez-de-Ocariz M, Velasco-Hernández Mde L, Salas-Labadía C. A patient with trisomy 13 mosaicism with an unusual skin pigmentary pattern and prolonged survival. *Pediatr Dermatol*. 2014;31(5):580-3.
89. Taylor TH, Gitlin SA, Patrick JL, Crain JL, Wilson JM, Griffin DK. The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans. *Hum Reprod Update*. 2014;20(4):571-81.
90. Liu Q, Karolak JA, Grochowski CM, Wilson TA, Rosenfeld JA, Bacino CA, et al. Parental somatic mosaicism for CNV deletions - A need for more sensitive and precise detection methods in clinical diagnostics settings. *Genomics*. 2020;112(5):2937-41.
91. Konnick EQ, Pritchard CC. Germline, hematopoietic, mosaic, and somatic variation: interplay between inherited and acquired genetic alterations in disease assessment. *Genome Medicine*. 2016;8(1):100.
92. Thorpe J, Osei-Owusu IA, Avigdor BE, Tupler R, Pevsner J. Mosaicism in Human Health and Disease. *Annu Rev Genet*. 2020;54:487-510.
93. Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, Zackai EH, et al. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet*. 2010;19(7):1263-75.
94. Balbeur S, Grisart B, Parmentier B, Sartenaer D, Leonard P-E, Ullmann U, et al. Trisomy rescue mechanism: the case of concomitant mosaic trisomy 14 and maternal uniparental disomy 14 in a 15-year-old girl. *Clin Case Rep*. 2016;4(3):265-71.
95. Emine Ikbal A. Ch. 5. What Causes Down Syndrome? Subrata Kumar D, editor. *Down Syndrome and Other Chromosome Abnormalities*. Rijeka: IntechOpen; 2021.
96. Wikimedia Commons contributors. Nondisjunction in Mitosis.jpg [Internet]. Güncelleme: 21 Ağustos 2022. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://commons.wikimedia.org/w/index.php?title=File:Nondisjunction_in_Mitosis.jpg&oldid=441426304
97. Yamazawa K, Ogata T, Ferguson-Smith AC. Uniparental disomy and human disease: an overview. *Am J Med Genet C Semin Med Genet*. 2010;154c(3):329-34.
98. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010;87(1):129-38.
99. Papenhausen PR, Kelly CA, Harris S, Caldwell S, Schwartz S, Penton A. Clinical significance and mechanisms associated with segmental UPD. *Molecular Cytogenetics*. 2021;14(1):38.
100. Takahashi KK, Innan H. Frequent somatic gene conversion as a mechanism for loss of heterozygosity in tumor suppressor genes. *Genome Res*. 2022;32(6):1017-25.

101. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5(4):557-72.
102. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657-63.
103. Assié G, LaFramboise T, Platzer P, Bertherat J, Stratakis CA, Eng C. SNP arrays in heterogeneous tissue: highly accurate collection of both germline and somatic genetic information from unpaired single tumor samples. *Am J Hum Genet*. 2008;82(4):903-15.
104. Staaf J, Lindgren D, Vallon-Christersson J, Isaksson A, Göransson H, Juliusson G, et al. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biology*. 2008;9(9):R136.
105. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences*. 2010;107(39):16910-5.
106. Yau C, Mouradov D, Jorissen RN, Colella S, Mirza G, Steers G, et al. A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. *Genome Biology*. 2010;11(9):R92.
107. González JR, Rodríguez-Santiago B, Cáceres A, Pique-Regi R, Rothman N, Chanock SJ, et al. A fast and accurate method to detect allelic genomic imbalances underlying mosaic rearrangements using SNP array data. *BMC Bioinformatics*. 2011;12:166.
108. Pique-Regi R, Monso-Varona J, Ortega A, Seeger RC, Triche TJ, Asgharzadeh S. Sparse representation and Bayesian detection of genome copy number alterations from microarray data. *Bioinformatics*. 2008;24(3):309-18.
109. Loh P-R, Genovese G, Handsaker RE, Finucane HK, Reshef YA, Palamara PF, et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature*. 2018;559(7714):350-5.
110. Glessner JT, Chang X, Liu Y, Li J, Khan M, Wei Z, et al. MONTAGE: a new tool for high-throughput detection of mosaic copy number variation. *BMC Genomics*. 2021;22(1):133.
111. Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, et al. ascatNgs: Identifying Somatic Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr Protoc Bioinformatics*. 2016;56:15.9.1-15.9.7.
112. Yau C. OncoSNP-SEQ: a statistical approach for the identification of somatic copy number alterations from next-generation sequencing of cancer genomes. *Bioinformatics*. 2013;29(19):2482-4.
113. Broad Institute. (How to part I) Sensitively detect copy ratio alterations and allelic segments [Internet]. Güncelleme: 17 Kasım 2022. [Erişim Tarihi: 20 Kasım 2022]. Erişim adresi: <https://gatk.broadinstitute.org/hc/en-us/articles/360035531092?id=11682>

114. Broad Institute. (How to part II) Sensitive detect copy ratio alterations and allelic segments [Internet]. Güncelleme: 19 Ekim 2022. [Erişim Tarihi: 20 Kasım 2022]. Erişim adresi: <https://gatk.broadinstitute.org/hc/en-us/articles/360035890011?id=11683>
115. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics (Oxford, England)*. 2012;28(3):423-5.
116. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, et al. Inferring copy number and genotype in tumour exome data. *BMC Genomics*. 2014;15(1):732.
117. Sathirapongsasuti JF, Lee H, Horst BAJ, Brunner G, Cochran AJ, Binder S, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*. 2011;27(19):2648-54.
118. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-76.
119. Guatelli S, Mascialino Sudmeier B, Pfeiffer A, Pia MG, Ribon A, Viarengo P. Application of statistical methods for the comparison of data distributions. *IEEE Symposium Conference Record Nuclear Science 2004; 2004; Rome, Italy*. Vol. 4, p. 2086-90.
120. Kærgård N. Estimation criterion, residuals and prediction evaluation. *Computational Statistics & Data Analysis*. 1987;5(4):443-50.
121. Wikipedia contributors. Residual sum of squares [Internet]. [Erişim Tarihi: 15 Aralık 2022]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Residual_sum_of_squares&oldid=1127927489
122. Wikipedia contributors. Chi-squared test [Internet]. [Erişim Tarihi: 15 Aralık 2022]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Chi-squared_test&oldid=1132121646
123. Wikipedia contributors. Kolmogorov–Smirnov test [Internet]. [Erişim Tarihi: 15 Aralık 2022]. Erişim adresi: https://en.wikipedia.org/w/index.php?title=Kolmogorov%E2%80%93Smirnov_test&oldid=1125984976
124. Wikipedia contributors W. R (programming language) [Internet]. Güncelleme: 13 Aralık 2022. [Erişim Tarihi: 15 Aralık 2022]. Erişim adresi: [https://en.wikipedia.org/w/index.php?title=R_\(programming_language\)&oldid=1127197502](https://en.wikipedia.org/w/index.php?title=R_(programming_language)&oldid=1127197502)
125. Peng RD. R Programming for Data Science [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://bookdown.org/rdpeng/rprogdatascience/history-and-overview-of-r.html>
126. The R Project for Statistical Computing [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://www.r-project.org/>

127. The Comprehensive R Archive Network [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://cran.r-project.org/>
128. Bioconductor [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://www.bioconductor.org/>
129. Shiny [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://shiny.rstudio.com/>
130. AppliedBiosystems. CytoScan Optima Assay Manual Workflow 2018 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017734_703280_Cyto_Optima_UG.pdf
131. Affymetrix. Affymetrix® GeneChip® Command Console® (AGCC) 4.0 User Manual 2013 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://tools.thermofisher.com/content/sfs/manuals/agcc_command_console_user_guide.pdf
132. AppliedBiosystems. Chromosome Analysis Suite (ChAS) v4.2 USER GUIDE 2020 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://assets.thermofisher.com/TFS-Assets/GSD/Handbooks/Chromosome_analysis_suite_v4.2_user-guide.pdf
133. Illumina Inc. HumanCytoSNP-12 v2.1 BeadChip Product Files [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://support.illumina.com/array/array_kits/humancyto-snp-12_v2-1_dna_analysis_kit/downloads.html
134. Illumina. GenomeStudio® Genotyping Module v2.0 Software Guide 2016 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2-0/genomestudio-genotyping-module-v2-user-guide-11319113-01.pdf
135. Illumina Inc. Beeline Software v2.0.3 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://support.illumina.com/downloads.html?filters=web%253Asoftware%252Fbeeline>
136. Illumina. BlueFuse Multi v4.5 Software Guide 2018 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bluefuse-multi-software/bluefuse-multi-software-guide-15053620-04.pdf
137. Copy Number and Cytogenetics Sample Data [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi: <https://www.thermofisher.com/tr/en/home/life-science/microarray-analysis/microarray-data-analysis/microarray-analysis-sample-data/copy-number-cytogenetics-sample-data.html>
138. Gel B. karyoploteR: plot customizable linear genomes displaying arbitrary data [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi:

<https://bioconductor.org/packages/devel/bioc/vignettes/karyoploteR/inst/doc/karyoploteR.html>

139. Li W, Olivier M. Current analysis platforms and methods for detecting copy number variation. *Physiol Genomics*. 2013;45(1):1-16.
140. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-43.
141. Dimitrova D, Kaishev V, Tan S. Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous. *Journal of statistical software*. 2020;95.
142. Vrbik J. Deriving CDF of Kolmogorov-Smirnov Test Statistic. *Applied Mathematics*. 2020;11:227-46.
143. Lima L, Wang K. PennCNV in whole-genome sequencing data. *BMC Bioinformatics*. 2017;18.
144. R Core Team. Package 'Parallel' 2022 [Internet]. [Erişim Tarihi: 1 Kasım 2022]. Erişim adresi:
<https://stat.ethz.ch/R-manual/R-devel/library/parallel/doc/parallel.pdf>

8. EKLER

EK-1: Tez çalışmasının etik kurul izni



T.C.
HACETTEPE ÜNİVERSİTESİ
Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu

Sayı : 16969557-491

Konu :

ARAŞTIRMA PROJESİ DEĞERLENDİRME RAPORU

Toplantı Tarihi : 02 MART 2021 SALI
Toplantı No : 2021/05
Proje No : GO 21/267 (Değerlendirme Tarihi: 02.03.2021)
Karar No : 2021/05-16

Üniversitemiz Tıp Fakültesi Tıbbi Genetik Anabilim Dalı öğretim üyelerinden Dr. Öğr. Üyesi Dr. Arda ÇETİNKAYA'nın sorumlu araştırmacı olduğu, Hande Melisa ACUN'un yüksek lisans tezi olan, GO 21/267 kayıt numaralı, "*Mikrodizin Kaynaklı Genotip Bilgisi Kullanılarak Kopya Sayısı Değişikliklerinde Mozaiklik Oranının Saptanması*" başlıklı proje önerisi araştırmamızın gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş olup, 01 Nisan 2021-30 Eylül 2022 tarihleri arasında geçerli olmak üzere etik açıdan **uygun bulunmuştur**. Çalışma tamamlandığında sonuçlarını içeren bir rapor örneğinin Etik Kurulumuza gönderilmesi gerekmektedir.

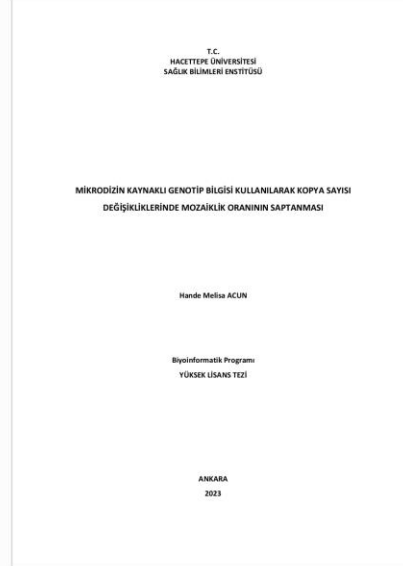
1. Prof. Dr. Ayşe Lale DOĞAN	(Başkan)	7. Doç. Dr. Nüket Paksoy ERBAYDA*	Üye)
2. Prof. Dr. G. Burça AYDIN'	(Üye)	8. Doç. Dr. Betül Çelebi SALTIK	Üye)
3. Prof. Dr. M. Özgür UYANIK	(Üye)	9. Doç. Dr. Hande Güney DENİZ	(Üye)
4. Prof. Dr. Ayşe Kin İŞLER	(Üye)	10. Dr. Öğr. Üyesi Mütge DEMİR	(Üye)
5. Doç. Dr. H. Tuna Çak ESEN	(Üye)	11. Av. Serap MORALIOĞLU	(Üye)
6. Doç. Dr. Can Ebru KURT	(Üye)		

EK-2: Tez çalışması orijinallik raporu**Dijital Makbuz**

Bu makbuz ödevinizin Turnitin'e ulaştığını bildirmektedir. Gönderiminize dair bilgiler şöyledir:

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen: Hande Melisa Acun
Ödev başlığı: TEZ_FİNAL
Gönderi Başlığı: Mikrodizin Kaynaklı Genotip Bilgisi Kullanılarak Kopya Sayısı ...
Dosya adı: H.Melisa_Acun_TEZ.pdf
Dosya boyutu: 4.66M
Sayfa sayısı: 160
Kelime sayısı: 35,177
Karakter sayısı: 230,215
Gönderim Tarihi: 14-Oca-2023 03:01ÖS (UTC+0300)
Gönderim Numarası: 1992712163



Mikrodizin Kaynaklı Genotip Bilgisi Kullanılarak Kopya Sayısı Değişikliklerinde Mozaiklik Oranının Saptanması

ORIJİNALLIK RAPORU

% 4	% 3	% 0	% 2
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	% 2
2	github.com İnternet Kaynağı	% 1
3	Submitted to Hacettepe University Öğrenci Ödevi	<% 1
4	openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<% 1
5	acikbilim.yok.gov.tr İnternet Kaynağı	<% 1
6	Submitted to University of Glasgow Öğrenci Ödevi	<% 1
7	iksadyayinevi.com İnternet Kaynağı	<% 1
8	www.biorxiv.org İnternet Kaynağı	<% 1
9	Submitted to TechKnowledge Turkey Öğrenci Ödevi	<% 1

EK-3: Aydınlatılmış onam formu

**GENETİK ARAŞTIRMALARDA KATILIMCILAR İÇİN
AYDINLATILMIŞ ONAM FORMU**

Sayın.....

Hacettepe Üniversitesi Tıbbi Genetik Anabilim Dalı **“Mikrodizin Kaynaklı Genotip Bilgisi Kullanılarak Kopya Sayısı Değişikliklerinde Mozaiklik Oranının Saptanması”** isimli bir araştırma yapmaktadır. Bu çalışma ile bazı hastalıkların rutin tanısında mikrodizin adı verilen özel bir genetik yöntemin analizini kolaylaştıracak bir bilgisayar programı geliştirmeyi hedefliyoruz. Bu bilgisayar programı yeni bir laboratuvar yöntemi olmayıp elde edilen verinin içinden farklı analizler yaparak özellikle bir kişinin hücrelerinin sadece bir kısmını ilgilendiren “mozaik” dediğimiz genetik değişiklikleri saptamaya yöneliktir. Katılımcı olarak sizde bu mikrodizin testi ile bir hastalık sebebi olacak bir sonuç bulmayı beklemiyor, bilgisayar programını doğrulamak ve kontrol etmek amacıyla, eğer kabul ederseniz, sizi de araştırmaya katmak istiyoruz.

Bu araştırmanın sonuçları yalnızca bilimsel amaçlarla kullanılacak ve kimliğiniz her zaman gizli tutulacaktır. Bu araştırmaya katılmanızdan dolayı sizden herhangi bir para talep edilmeyecektir. Aynı şekilde size de herhangi bir ödeme yapılmayacaktır.

Araştırmaya katılmak isterseniz Dr. Öğr. Ü. Dr. Arda ÇETİNKAYA tarafından aile ağacı bilgileriniz çıkarılacaktır ve sizden çalışma kapsamında 1 tüp (5ml) kan alınıp DNA elde edilecektir. DNA örneği sizin genetik yapınızda bulunan ve toplumda yaygın olarak görünen değişiklikleri (polimorfizm) belirleme ve bu değişikliklerin kompozisyonlarını değerlendirerek genetik materyalde kayıplar ya da kazanımların olup olmadığının araştırılmasında kullanılacaktır. İlgili örneklerin yetersiz kalması halinde sizden 1 tüp (5ml) daha kan alınması gerekli olabilir.

Kan alınırken iğne batması nedeniyle hafif bir acı duyabilirsiniz. Çok düşük bir ihtimal olsa da kan alırken kanamanın uzaması ya da enfeksiyon gelişmesi riskleri olabilir.

Kişiye ait genetik bilgiler maddi ve sosyal açıdan istismar edilebilecek bilgilerdir. Örneğin genetik testte biyolojik anne ve babayı belirlemek mümkündür. O nedenle, araştırma sonuçlarının yalnızca bilimsel amaçlarla ve kimliğinizi gizli tutarak kullanılacağını tekrar vurgulamak istiyoruz. Araştırma sonucunda hastalık sebebi bir değişiklik gözlemlemeyi beklemiyor olsak bile hastalığa yol açabilecek genetik

değişiklikleri taşıdığınız tesadüfen ortaya çıkabilir. Bu bilgileri herhangi bir ücret talep etmeden size bildirebiliriz. Bu konuda bilgi edinmek isteyip istemediğinize yönelik tercihinizi lütfen yazılı formun altında belirtiniz.

Tarafınızdan alınan örneğin saklanması ve ileride yapılacak diğer çalışmalarda kullanımı ancak sizin izninize tabidir. Bu örnekler uzun yıllar saklanabilir ve başka genetik araştırmalarda örneğin kontrol örneği olarak kullanılabilir. Benzer şekilde sizden elde edilen mikrodizin verisi de veri bankalarına kimliğiniz ile ilişkilendirilmeden eklenip başka araştırmacıların kullanımına açılabilir. Lütfen yazılı formun altındaki seçeneklerden size uygun olan bir tanesini işaretleyerek örneklerin ve/veya verilerin saklanması ve ileride başka araştırmalarda kullanılması yönündeki görüşünüzü belirtiniz.

Sizinle ilgili genetik bilgiler gizli tutulacak, ancak çalışmanın kalitesini denetleyen görevliler, etik kurullar ya da resmi makamlarca gereği halinde incelenebilecektir. Sizinle ilgili tıbbi bilgiler kimlik belirtilmeden tıp öğrencilerinin eğitiminde veya bilimsel nitelikte yayınlarda kullanılabilir.

Araştırmaya katılmak zorunda olmadığınız gibi araştırmaya katılmayı kabul ettiğinizde, istediğiniz anda ayrılma hakkına da sahipsiniz. Ancak bu kararınızı bize önceden bildirirseniz araştırmanın bozulmasına meydan vermemiş olursunuz. Katılmak istemediğinizde şu anda sürdürülen tedavi işlemleri bundan etkilenmeyecektir.

Katılımcının Beyanı

Dr. Öğr. Ü. Dr. Arda ÇETİNKAYA tarafından yapılması planlanan genetik çalışma konusunda bilgi verildi ve genetik yapıdaki tüm genlerin incelenmesine yönelik bir araştırma yapılabileceği hakkında bilgilendim. Araştırmanın amacı, uygulama biçimi ile riskleri ve kimlik bilgilerimle ilgili gizliliğin sağlanacağı konusunda yeterli açıklama yapıldı. Araştırma ile ilgili sorularım için Dr. Öğr. Üye Arda ÇETİNKAYA ile temas edebileceğim bana bildirildi. İstedğim zaman araştırmadan çekilebileceğimi biliyorum. Araştırmaya katılımımın tamamen gönüllü olduğu, katılmam ya da katılıp daha sonra araştırmadan çekildiğim durumda tedavi ve tetkiklerimin bundan etkilenmeyeceği belirtildi. Bu araştırmaya kendi gönüllü onayım ile katılmaya olurum vardır.

Katılımcı	Adı Soyadı	Doğum Tarihi	Adres/Tel	İmza
-----------	------------	--------------	-----------	------

Tanık	Adı Soyadı	Doğum Tarihi	Adres/Tel	İmza
-------	------------	--------------	-----------	------

Hekim:

Adı Soyadı :

Adres/Tel :

İmza :

Görüşme tarih ve saati:

Bu çalışmada elde edilecek kendimle ilgili bilgileri,

Öğrenmek istemiyorum ()

Tanı ve tedaviye yol açmasa bile öğrenmek istiyorum ()

Sonuçların önleyici veya tedaviye yönelik olduğu durumlarda öğrenmek istiyorum ()

Biyolojik örneğin başka araştırmalarda kullanımı hakkında olur beyanı

Tarafımdan alınan kodlanmış* örneğin yalnızca önerilen çalışma için kullanımını onaylıyorum; ileride yapılması olası diğer çalışmalar için onay vermiyorum.

Tarafımdan alınan kodlanmış örneğin, araştırma konusuyla bağlantılı diğer çalışmalarda kullanımını onaylıyorum, ancak farklı çalışmalar için tekrar bilgilendirilmek ve yeni onay vermek istiyorum.

Tarafımdan alınan kodlanmış örneğin gelecekte her türlü genetik çalışmada (kimliğim ile bağlantısız) olarak kullanılmasını onaylıyorum.

*Kodlanmış örnek: Sizden alınan örneğe bir kod numarası verilir. Kod numarasını yalnızca araştırmacı bilir ve sizin kimlik bilgilerinize yalnızca araştırmacı ulaşabilir. Böylece kimlik bilgileriniz gizli tutulmuş olur

Genetik Bilgilerin anonim olarak başka arařtırmalarda kullanılması hakkında olur beyanı

- Kendime ait genetik bilgilerinin kimlik bilgilerimle iliřkilendirilmeden veri bankalarına eklenmesini, başka arařtırmalarda kullanılmak üzere veri deęiřimi yapılmasını ONAYLAMİYORUM.
- Kendime ait genetik bilgilerinin kimlik bilgilerimle iliřkilendirilmeden veri bankalarına eklenmesini, başka arařtırmalarda kullanılmak üzere veri deęiřimi yapılmasını ONAYLIYORUM.

EK-4: ChAS'ın Analysis Workflow modülü ile Affymetrix SNP mikrodizin .CEL dosyalarının analizi

The screenshot displays the ChAS Analysis Workflow interface. At the top, there are navigation tabs: Analysis setup, Workflow dashboard, and QC results. Below this, the Array type is set to 'CytoScanOptima_Array', Genome Version to 'hg19', and Analysis workflow to 'CytoScanOptima_Array Single Sample Analysis: NA33'. The Reference model is 'CytoScanOptima_Array.na33.REF_MODEL' and the Annotation is 'CytoScanOptima_Array.na33.annot.db'. A table lists the intensity files to analyze, with 'T10_(CytoScanOptima_Array).CEL' selected. The 'Add...' button is highlighted with a red box and an upward arrow. The Output result information section shows the output folder as 'C:\Affymetrix' and a suffix field. The Optional section includes checkboxes for 'Set Gender Manually', 'Generate Results Summary File', and 'Use Manual Recentering', along with a 'Report format' dropdown set to 'PDF' and 'Submit' and 'Reset' buttons at the bottom right.

Şekil 8.1. ChAS'ın Analysis Workflow modülü ile Affymetrix SNP mikrodizin .CEL dosyalarının .cychp dosyalarına dönüştürülmesi.

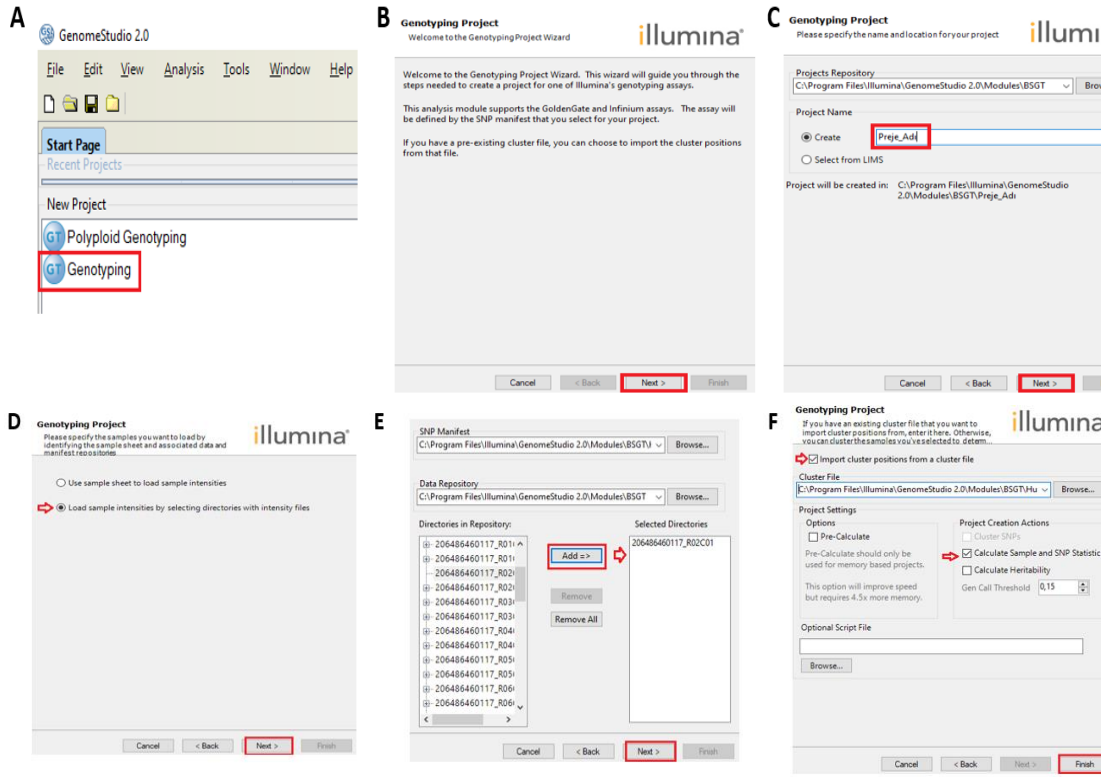
Analysis Workflow yazılımında en üstteki panelde mikrodizin tipi CytoScanOptima_Array, Genome Version hg19 ve Analysis workflow NA33 seçilmelidir. Hemen altındaki seçeneklerde ise na33 REF_MODEL ve na33.annot.db anotasyon versiyonları seçilmiş olmalıdır. Ardından sağ tarafta "Add" seçeneği ile .CEL dosyaları bilgisayarda ChAS yazılımının bulunduğu Affymetrix klasörü içerisinden eklenmelidir. Analiz sonucunda dönüştürülecek .cychp dosyalarının yer alacağı klasör ise ortada yer alan "..." kutucuğu ile seçilerek yine aynı Affymetrix klasörü olmalıdır. En altta sağda bulunan "submit" butonu seçilerek dosya dönüşümü başlatılmaktadır.

EK-5: Affymetrix APT (Analysis Power Tools) ile .cychp dosyalarının .txt dosyalarına dönüştürülmesini sağlayan R kodu

```
setwd("C:/Users/melis/Documents") # it must be the system file that APT uses
system2("C:/Program Files/Thermo Fisher Scientific/Analysis Power Tools/APT-2.11.4/bin/apt-chp-to-
txt.exe", args = c("C:/Users/melis/Documents/sample.cyoapt.cychp")) # run APT with system2 function
library(filesstrings);
file.move("sample.cyoapt.cychp.txt", "C:/Users/melis/Desktop/TEZ/R/Data") # move output file to
analysis folder
```

Yukarıdaki kod sayesinde APT yazılımı R içerisinde system2 fonksiyonu ile çalıştırılarak girdi olarak verilen .cychp dosyasının .txt uzantılı çıktısı alınmaktadır. APT için argümanlara verilen girdinin Windows içerisinde sistem dizini olan “C” dizini altında bulunan bir klasörde yer alması gerekmektedir. Daha sonra çıktı dosyaları Copy-fm’in çalışma klasöre taşınmaktadır.

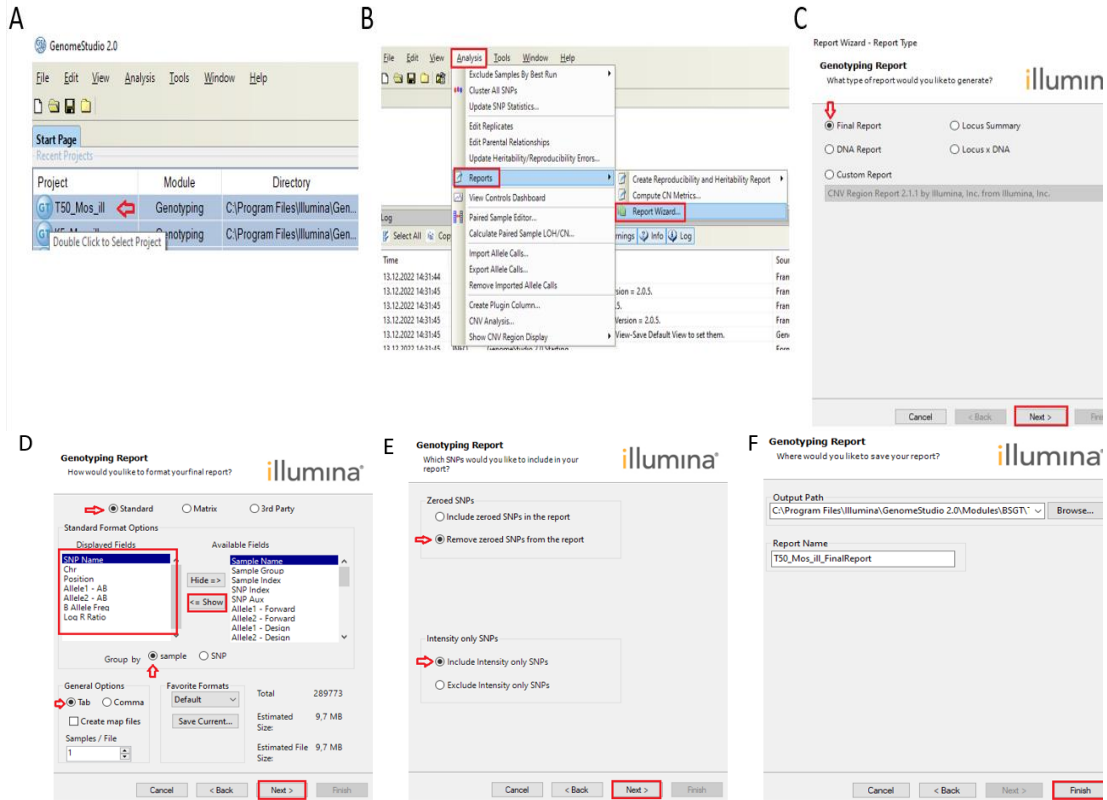
EK-6: Illumina GenomeStudio ve Beeline ile SNP mikrodizin analizi basamakları



Şekil 8.2. Illumina GenomStudio’da proje oluşturulması.

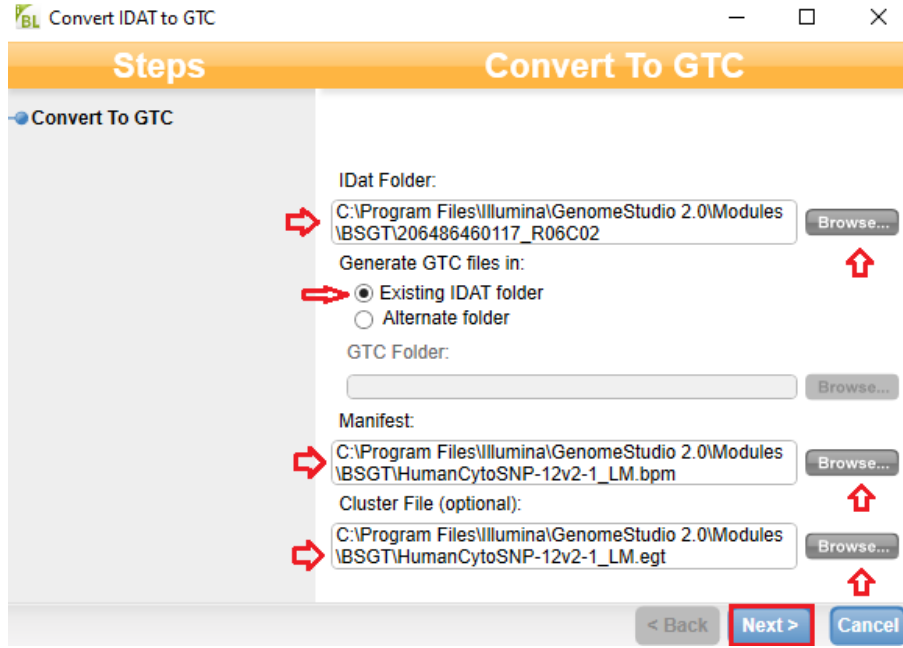
Şekilde GenomeStudio içerisinde tıklanması gereken butonlar kırmızı kutucukla işaretlenmiş ve işaretlenmesi gereken yerler kırmızı oklarla belirtilmiştir.

A) Program açıldıktan sonra karşılaşılan ekranda “Genotyping” seçilmelidir. B) Ardından açılan yazıda “Next” seçilmelidir. C) Bir sonraki pencerede oluşturulacak projenin yer alacağı klasör ve proje adı belirtilerek “Next” seçilmelidir. D) Alttaki seçenek seçilerek SNP mikrodizin örnekleri sinyal dosyaları seçilerek eklenmelidir. E) Ardından .idat dosyalarının bulunduğu klasör ve dosyalar sol tarafta tıklanarak “Add” ile sağ tarafa eklenmeli ve “Next” ile ilerlenmelidir. F) Son aşamada .bpm ve .egt uzantılı manifest ve cluster dosyaları seçilmeli ve SNP istatistiklerinin hesaplanması seçeneği işaretlenerek “Finish” ile proje hazırlanması bitirilmektedir.



Şekil 8.3. Illumina GenomeStudio'dan FinalReport.txt alınması.

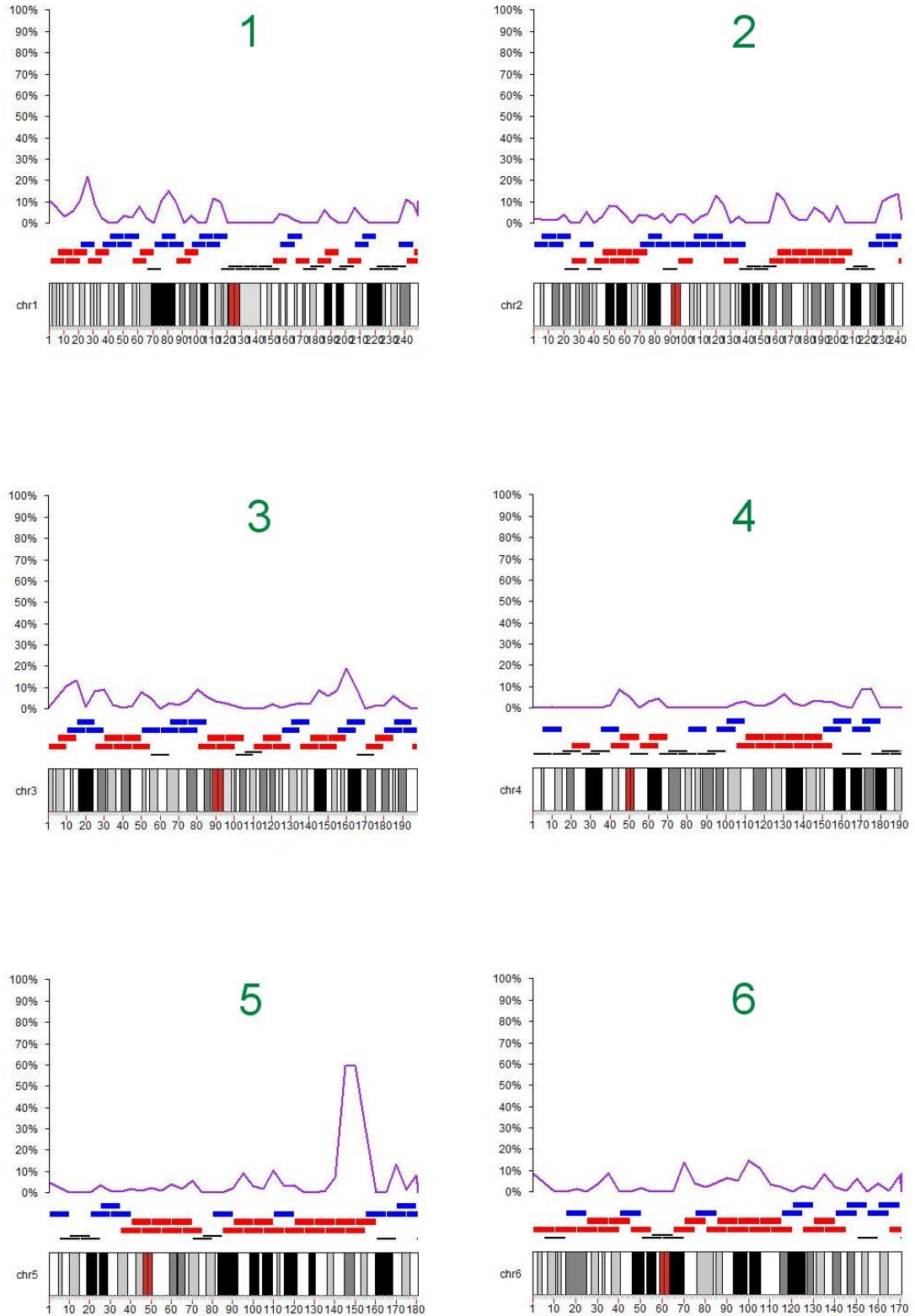
Şekilde A-F aşamalarında seçilmesi gereken yerler kırmızı kutucuk ve oklarla işaretlenmiştir. A) Bir önceki aşamada oluşturulan proje GenomeStudio'da açılır. B) Üst panelde bulunan "Analysis" sekmesinde "Reports" ve sağında açılan sekmede "Report wizard" seçilir. C) Bir sonraki pencerede FinalReport seçilerek "Next" denir. D) Solda bulunan "Standard" işaretlendikten sonra raporda yer alması gereken başlıklar sağ taraftan seçilerek "Show" seçeneği ile sol tarafa eklenmelidir. Altında ise "Group by" seçeneği "sample" olarak işaretlenmeli, sol tarafta "Tab" işaretlenerek "Next" seçilmelidir. E) "Zeroed SNPs" isimli kısımda alttaki seçenek işaretlenerek hesaplanamamış SNP istatistikleri dahil edilmemelidir. Diğer seçenekte ise "include intensity only SNPs" işaretlenmelidir. F) Tablonun kaydedileceği klasör ve tablo ismi belirtilerek "Finish" ile FinalReport.txt uzantılı tablo elde edilmektedir.



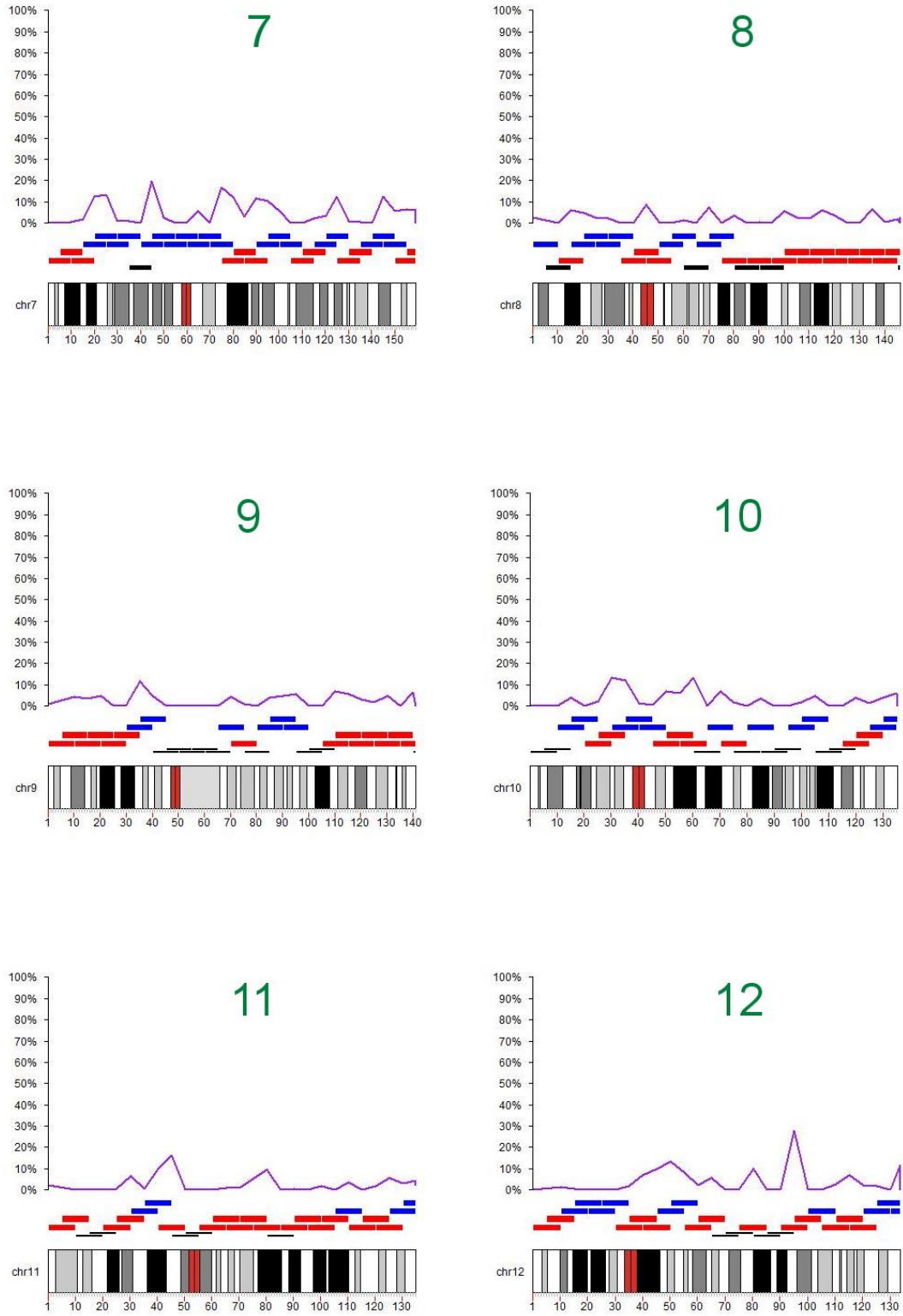
Şekil 8.4. Beeline yazılımı ile .idat dosyalarının .gtc dosyalarına dönüştürülmesi.

Seçilmesi gereken kutucuklar kırmızı oklar ile belirtilmiştir. İlk olarak örneğe ait .idat dosyasının bulunduğu klasör seçilmelidir. Altındaki ilk seçenek seçilerek çevrilen .gtc dosyasının .idat dosyaları ile aynı yerde olması sağlanmalıdır. Manifest ve cluster dosyaları da seçildikten sonra "Next" tıkladığında arkaplanda Beeline çalışacak ve .gtc dosyalarının belirtilen klasörde oluşturulmuş olduğu görülecektir.

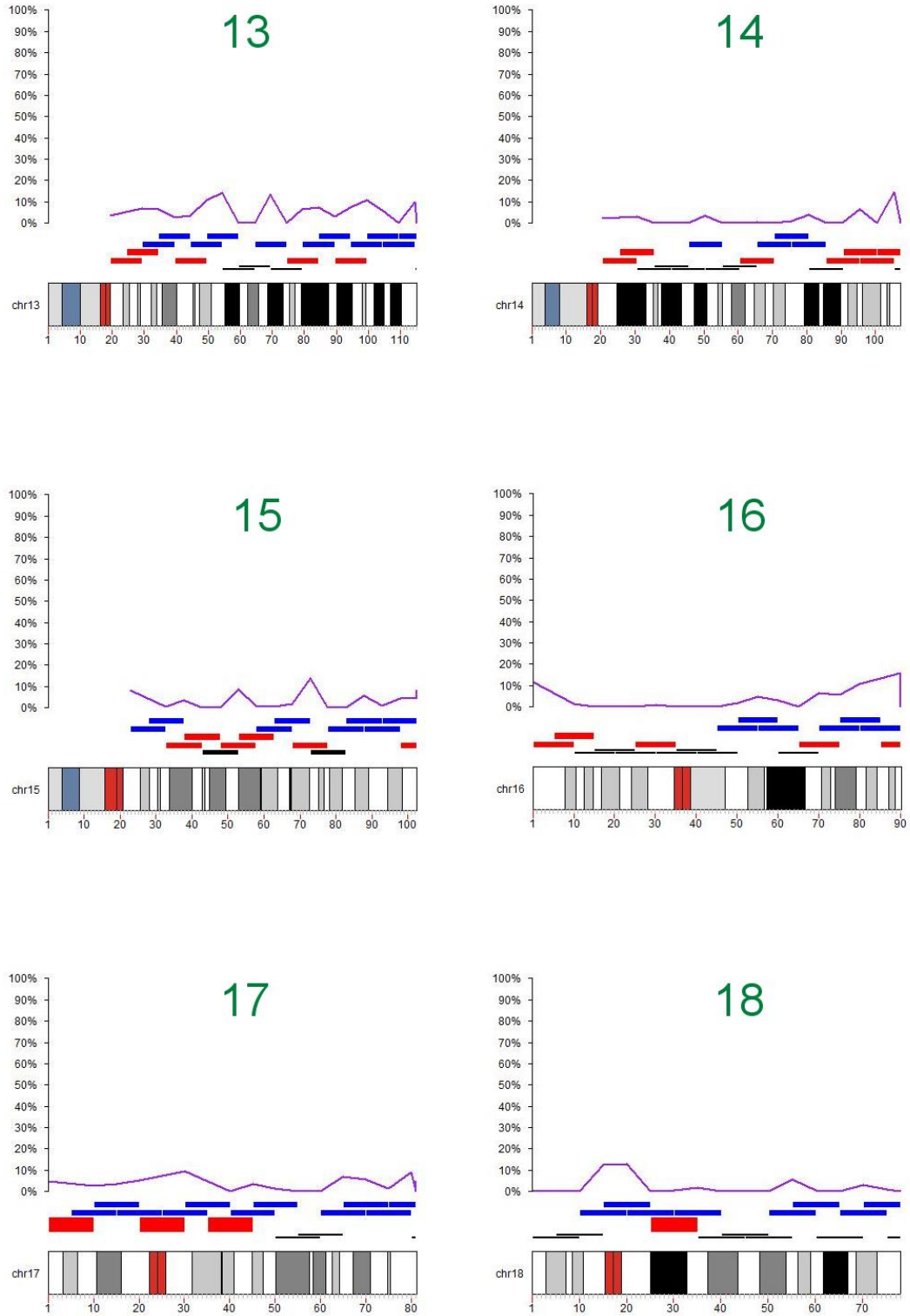
EK-7: Copy-fm (*scanfm* fonksiyonu) ile genom boyu tarama



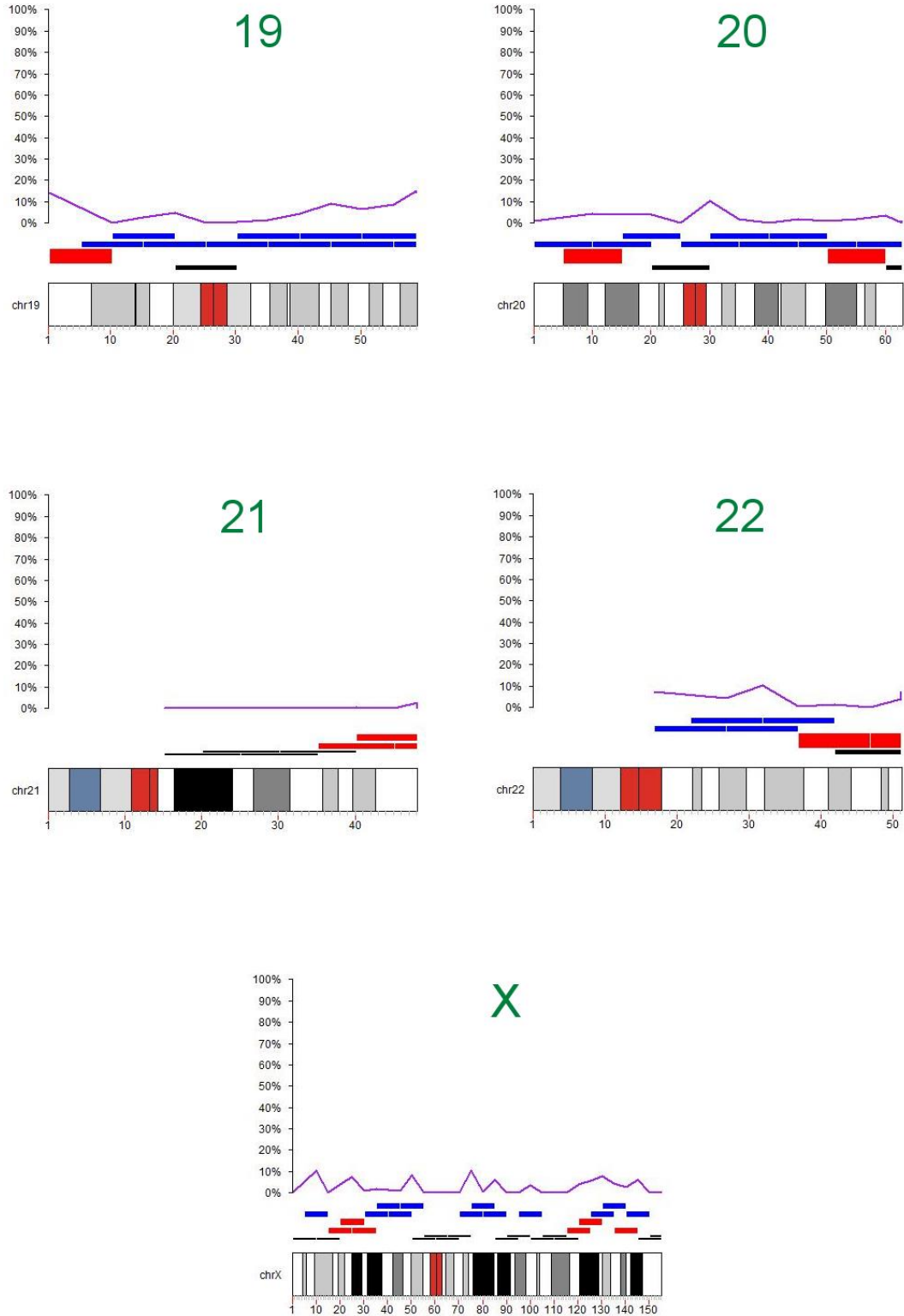
Şekil 8.5. Copy-fm içinde yer alan *scanfm* fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyin (Affymetrix) tüm genom boyu taraması



Şekil 8.5. (devamı) Copy-fm içinde yer alan *scanfm* fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyin (Affymetrix) tüm genom boyu taranması



Şekil 8.5. (devamı) Copy-fm içinde yer alan *scanfm* fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyin (Affymetrix) tüm genom boyu taraması



Şekil 8.5. (devamı) Copy-fm içinde yer alan *scanfm* fonksiyonu ile 5q ve 12q mozaik delesyonlu bireyin (Affymetrix) tüm genom boyu taraması

EK-8: f_m hesaplaması için kullanılan diğer programların komutları

MoChA

Bu program için uygulanan komutlar sadece Illumina mikrodizini için geçerli olan komutlardır. Program için önerildiği şekilde linux içerisinde Illumina.sh, vcf.sh ve eagle.sh isimli bash dosyaları oluşturulmuştur. MoChA'nın çalışabilmesi için gereken aşamalarda bcftools v1.13, htlib v1.15.1 ve Eagle v2.4.1 kullanılmıştır. Opsiyonel olarak vcftools v0.1.13 kullanılmıştır.

```
#!/bin/bash

bpm_manifest_file="HumanCytoSNP-12v2-1_LM.bpm"
egt_cluster_file="HumanCytoSNP-12v2-1_LM.egt"
path_to_gtc_folder="/$HOME/MoChA"
ref="$HOME/ ref="$HOME/refseq/humanG1Kv37_hg19/human_g1k_v37.fasta"
out_prefix="T50"
bcftools +gtc2vcf \
  --no-version -Ou \
  --bpm $bpm_manifest_file \
  --egt $egt_cluster_file \
  --gtcs $path_to_gtc_folder \
  --fasta-ref $ref \
  --extra $out_prefix.tsv | \
  bcftools sort -Ou -T | \
  bcftools norm --no-version -Ob -c x -f $ref -o $out_prefix.bcf | \
  bcftools index --force --output $out_prefix.bcf.csi
```

Illumina .gtc dosyasından .bcf elde edilmiştir.

```
#!/bin/bash

cd "$HOME/ MoChA/T50_ill
pfx="T50"
dir="$HOME/ MoChA/T50_ill

bcftools annotate --no-version -Ob -o $dir/$pfx.unphased.bcf $bcf \
-x ID,QUAL,^INFO/ALLELE_A,^INFO/ALLELE_B,^FMT/GT,^FMT/BAF,^FMT/LRR && \
bcftools index -f $dir/$pfx.unphased.bcf
bcftools convert -O v -o $dir/$pfx.unphased.vcf $dir/$pfx.unphased.bcf
bash vcf.sh

grep -vP '^#' $pfx.unphased.bcf.vcf > data.vcf
grep -P '^#' $pfx.unphased.bcf.vcf > header.vcf
gawk -i inplace '/GT:/' data.vcf
cat header.vcf data.vcf > $pfx.unphased.bcf.gt.vcf

bcftools convert -O u -o $pfx.unphased.gt.bcf $pfx.unphased.gt.vcf
bcftools index -f $pfx.unphased.gt.bcf
```

Phasing aşaması için Eagle v2.1.4 kullanılmıştır.

```
#!/bin/bash
cd /home/arda/MoChA/T50_ill

ref="$HOME/refseq/humanG1Kv37_hg19/human_g1k_v37.fasta"
map="$HOME/refseq/genetic_map_hg19_withX.txt.gz"
kgp_pfx="$HOME/MoChA/GRCh37/ALL.chr"
kgp_sfx=".phase3_integrated.20130502.genotypes"
rule="GRCh37"
cnp="$HOME/MoChA/GRCh37/cnps.bed"
dup="$HOME/refseq/segdups.bed.gz"
cyto="$HOME/refseq/cytoBand.txt.gz"
dir="/home/arda/MoChA/T50_ill"
pfx="T50"

for chr in {1..22} X; do
  eagle \
  --geneticMapFile $map \
  --outPrefix $dir/$pfx.chr$chr \
  --numThreads 22 \
  --vcfRef $kgp_pfx${chr}$kgp_sfx.bcf \
  --vcfTarget $dir/$pfx.unphased.gt.bcf \
  --vcfOutFormat b \
  --outputUnphased \
  --chrom $chr \
  --noImpMissing \
  --pbwtilers 3 && \
  bcftools index -f $dir/$pfx.chr$chr.bcf
done

bcftools concat --no-version -Ob $dir/$pfx.chr{{1..22},X}.bcf -o $dir/$pfx.pgt.bcf
bcftools index -f $dir/$pfx.pgt.bcf
bcftools convert -O v -o $pfx.pgt.vcf $pfx.pgt.bcf
bash eagle.sh
$HOME/vcftools_0.1.13/bin/vcftools --vcf $pfx.pgt.vcf --out $pfx.phased --phased --recode -
-recode-INFO-all
bcftools convert -O u -o $pfx.phased.bcf $pfx.phased.recode.vcf
bcftools index -f $pfx.phased.bcf
bcftools +mochatools $pfx.phased.bcf -Ob -o $pfx.phased.gc.bcf -- -t GC -f
$HOME/refseq/humanG1Kv37_hg19/human_g1k_v37.fasta
bcftools sort $pfx.phased.gc.bcf -O u -o $pfx.phased.gc.sorted.bcf
bcftools index -c $pfx.phased.gc.sorted.bcf
bcftools +mocha -g GRCh37 -p $HOME/MoChA/GRCh37/cnps.bed -c $pfx.calls.tsv -z $pfx
.stats.tsv $pfx.phased.gc.sorted.bcf
```

MONTAGE

```
perl ./Montage.pl sample.txt
```

Yukarıdaki komutla çalıştırılan MONTAGE için girdi olarak örneklere ait Name, Chr, Position, B Allele Freq, Log R Ratio başlıklarından oluşan .txt dosyaları kullanılmıştır.

ÖZGEÇMİŞ