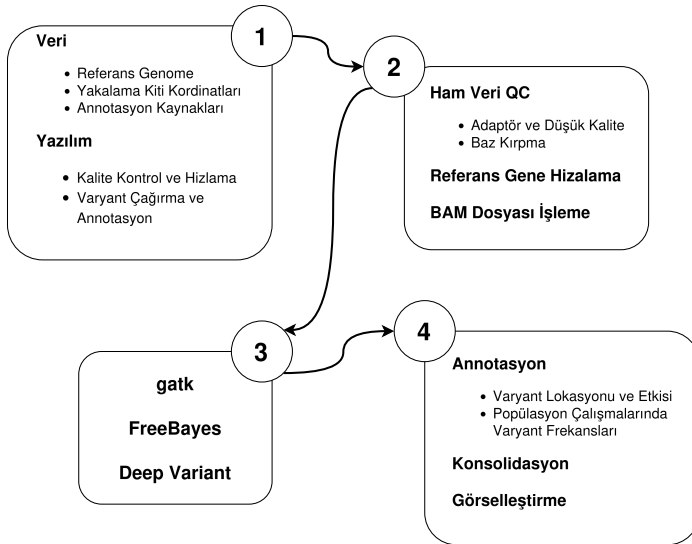


Genetik Veri Analizi

Biyolojik Verilerin Gizemini Çözmek - 8 Kasım 2024

Mustafa Karakaplan, mustafa.karakaplan@inonu.edu.tr

İçerik



“Bu sunumun hazırlanmasında sağladığı destek ve değerli katkıları için OpenAI’nin ChatGPT asistanına teşekkür ederim.”

FASTA

FASTA (“Fast-All” veya “Fast-Adaptive”), biyoinformatikte DNA, RNA ve protein dizilerini depolamak ve paylaşmak için yaygın olarak kullanılan basit bir metin dosyası formatıdır.

FASTA formatındaki bir dosya genellikle iki ana bileşenden oluşur:

- Başlık Satırı (Header Line): “>” karakteriyle başlar ve dizinin açıklamasını veya kimliğini içeren bir satırdır. Bu satırda genellikle dizinin türü, kaynağı, uzunluğu veya id numarası gibi bilgiler bulunur. Örneğin:
- Dizi Satırları (Sequence Lines): Başlık satırının hemen altındaki satırlarda dizi verisi bulunur.

Örnek bir DNA dizisini FASTA formatında şu şekilde görebiliriz:

```
>seq1 Homo sapiens chromosome 1  
ATGCGTACGTAGCTAGCTAGCTGACTGATCGTACGTAGCTAGCTA  
CGTAGCTAGCTGATCGTACGTAGCTAGCATCGATCGTACGTGCT
```

FASTQ

FASTQ (FASTA + Quality) formatı, biyoinformatikte özellikle Next-Generation Sequencing (NGS) verilerinde kullanılan, DNA veya RNA dizilerini saklamak için geliştirilmiş bir dosya formatıdır.

```
@SEQ_ID
ATCGTTAGCTA
+
!''*(((((***+))%%%)++
```

1. Başlık Satırı (Header Line): “@” karakteriyle başlar ve okuma hakkında tanımlayıcı bilgiler içerir.
2. Dizi Satırı (Sequence Line): DNA dizisi verisini içerir.
3. Kalite Başlık Satırı (Optional Description Line): “+” karakteriyle başlar.
4. Kalite Skoru Satırı (Quality Score Line): Dizi satırındaki her bazın okuma kalitesini gösteren ASCII karakterlerinden oluşur.

Phred Quality Score

Kalite skorları, Phred kalite skorlarına dayanır ve ASCII karakter kodlaması kullanılarak kaydedilir. Bu skorlar, sekans okuma cihazının her bir baz için verdiği kalite değerlendirmesini temsil eder ve şu formülle hesaplanır:

Phred Kalite Skoru;

$$Q = -10 \times \log_{10}(P)$$

$$P = 10^{-\frac{Q}{10}}$$

$$P = 10^{-\frac{0}{10}} = 10^0 = 1.0$$

$$P = 10^{-\frac{15}{10}} = 10^{-1.5} = 0.0316$$

Sembol	Phred	Yanlış Okuma Olasılığı, P
	Q	
!	0	1.000
"	1	0.794
#	2	0.631
(7	0.199
+	10	0.100
.	13	0.050
0	15	0.032
1	16	0.025
2	17	0.020
5	20	0.010
?	30	0.001
@	31	0.0008
A	32	0.0006
B	33	0.0005

SAM (Sequence Alignment/Map)

Biyoinformatikte dizileme verilerini metin formatında saklamak için kullanılan bir dosya formatıdır. Dizileme verilerinin genom referansına göre hizalanmış (alignment) halini içerir ve genetik veri analizinde önemli bir rol oynar. SAM formatı özellikle ham dizileme verilerini saklamak ve işlemek için yaygın olarak tercih edilir.

- SAM Dosya Formatının Özellikleri

- ▶ Metin Tabanlıdır: SAM dosyası düz metin formatında olduğu için kolayca okunabilir ve düzenlenebilir. Her bir hizalanmış okuma (read) için bir satır vardır.
- ▶ Dizilim ve Hizalama Bilgilerini İçerir: Her okuma dizisinin referans genoma göre hizalama pozisyonu, eşleşme kalitesi ve diğer özellikleri içerir.
- ▶ Kolonlar ve Bilgi Alanları: SAM dosyası, hizalamaları açıklayan 11 zorunlu alan ile isteğe bağlı alanlardan oluşur.

BAM Formatı (Binary Alignment/Map)

- BAM dosyası, genetik dizileme verilerinin ikili (binary) formatta sıkıştırılmış hali olup, dizileme verilerini verimli bir şekilde saklar.
- BAM dosyaları, dizileme sonucu elde edilen kısa okuma dizilerini (reads) genom referansına göre hizalanmış (alignment) bir şekilde saklar.
- İnsan genomu gibi büyük genomlar için BAM formatı tercih edilir, çünkü sıkıştırma sayesinde dosya boyutları çok daha küçüktür.
- BAM dosyası, ASCII formatında olan SAM (Sequence Alignment/Map) dosyasının sıkıştırılmış ve ikili versiyonudur. Bu nedenle BAM dosyaları SAM dosyalarına göre daha az yer kaplar ve daha hızlı işlenir.

BAI Formatı (BAM Index)

- BAI dosyası, BAM dosyaları için oluşturulan bir indeks dosyasıdır.
- Bu dosya, BAM dosyasındaki verilere hızlı bir şekilde erişim sağlar. BAM dosyasındaki verilerin hizalanmış olduğu genom bölgelerine göre indeksleme yapar.
- İndekslenmiş BAM dosyası sayesinde, araştırmacılar belirli bir genom bölgesine hızla erişebilir ve o bölgedeki okuma dizilerini inceleyebilir.
- BAI dosyası özellikle büyük BAM dosyalarında çok kullanışlıdır, çünkü doğrudan erişim sağlar ve analiz sürecini hızlandırır.

Bu iki dosya formatı, özellikle yüksek hacimli genetik veri analizi, Whole Exome Sequencing (WES) ve Whole Genome Sequencing (WGS) gibi uygulamalarda kullanılır ve biyoinformatik analizlerde yaygın olarak tercih edilir.

VCF

VCF (Variant Call Format), biyoinformatikte genetik varyantların saklanması ve paylaşılması için kullanılan bir dosya formatıdır. Genom dizilerinde bulunan mutasyonlar, varyasyonlar ve varyant çağırma sonuçlarını depolamak için tasarlanmıştır. VCF dosyaları, özellikle DNA veya RNA dizileme sonrası elde edilen varyantların analizinde önemli bir rol oynar.

```
##fileformat=VCFv4.2
```

```
##reference=GRCh38
```

```
chr1 1234567 . G A 50 PASS DP=100 GT:GQ 0/1:35
```

chr1: Varyantın 1. kromozomda bulunduğunu belirtir.

1234567: Pozisyonu.

G: Referans baz (Guanin).

A: Alternatif baz (Adenin).

50: Kalite skoru.

PASS: Filtreleme durumunu geçiyor.

DP=100: Derinlik (okuma sayısı).

GT: Genotip ve

GQ: kalite formatı.

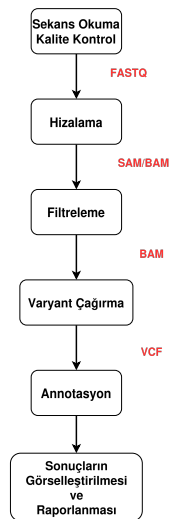
0/1:35: Örnekteki genotip (heterozigot) ve kalite skoru.

NGS

NGS (Next-Generation Sequencing)

Pipeline, büyük ölçekli genomik verilerin işlenmesi, analizi ve yorumlanması için kullanılan otomatik bir iş akışıdır. Bu tür bir boru hattı, yüksek hacimli ham sekans verilerini analiz etmek için bir dizi biyoinformatik adım içerir. Her adım, genetik veriyi anlamlandırmak ve kullanıma hazır hale getirmek için belirli işlemler yapar.

NGS (Next-Generation Sequencing) Akış Şeması



Şekil 1: Next Generation Sequencing

Genome Veri Tabanları

1. **NCBI GenBank:** NCBI'nin GenBank veri tabanı, çeşitli organizmalardan DNA dizilerini içeren bir veritabanıdır. GenBank, genom dizilerine ve genetik çeşitliliğe dair büyük bir veri kaynağı sağlar.
<https://www.ncbi.nlm.nih.gov/genbank/>
2. **Ensembl Genome Browser :** Ensembl, insan ve diğer organizmaların genomlarını içeren geniş bir veri tabanı sunar. Protein kodlayan genler, varyantlar, hastalık ilişkili bölgeler gibi birçok farklı veri içerir.
<https://ftp.ensembl.org/pub/>
3. **UCSC Genome Browser :** UCSC Genome Browser, genomik veri ve ilgili anotasyonlara ulaşabileceğiniz bir kaynaktır. Genetik diziler, gen anotasyonları ve çeşitli biyoinformatik araçlar sunar.
<https://hgdownload.soe.ucsc.edu/downloads.html>
4. **1000 Genomes Project :** 1000 Genomes Projesi, farklı popülasyonlardan insan genetik çeşitliliğini incelemek için geniş çaplı bir projedir. Genomik varyasyonları

Genome Veri Tabanları (ii)

analiz etmek isteyen araştırmacılar için çok değerli bir kaynaktır.

<https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>

5. **Genome Reference Consortium (GRC)** : Genome Reference Consortium, insan genom referans dizisinin en güncel versiyonlarını sağlar. GRC, İnsan Genom Projesi'nin orijinal verilerini ve sonrasında yapılan düzeltmeleri içerir.
<https://www.ncbi.nlm.nih.gov/grc/human>
6. **UCSC Genome Browser** : İnsan Genom Projesi verilerini ve sürekli olarak güncellenen diğer genomları içerir.
<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/>
7. **Ensembl** : İnsan Genom Projesi verilerini kullanarak detaylı anotasyonlar ve güncellemeler sağlar.
https://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/dna/

Klinik Veri Tabanları

Belirli bir genetik varyantı (örneğin, NM_031907.2:c.2088A>G) klinik veritabanlarında bulmak için aşağıdaki kaynakları kullanabilirsiniz:

1. **ClinVar:** Klinik açıdan önemli varyantların ve yorumlarının yer aldığı serbest erişimli bir veritabanıdır.
<https://www.ncbi.nlm.nih.gov/clinvar/>
<https://ftp.ncbi.nlm.nih.gov/pub/clinvar/>
2. **dbSNP:** Tek nükleotid polimorfizmleri ve diğer varyantlar hakkında bilgi sağlar.
<https://www.ncbi.nlm.nih.gov/snp/>
<https://ftp.ncbi.nlm.nih.gov/snp/>
3. **HGMD (Human Gene Mutation Database) :** HGMD, hastalıklarla ilişkili varyantların yer aldığı bir veritabanıdır ancak genelde kurumsal erişim gerektirir.
<http://www.hgmd.org>
4. **gnomAD (Genome Aggregation Database) :** gnomAD, varyantların popülasyon frekansları hakkında bilgi sağlar; nadir varyantların klinik önemini değerlendirmede yardımcı olabilir.

Klinik Veri Tabanları (ii)

<https://gnomad.broadinstitute.org/>

<https://gnomad.broadinstitute.org/downloads>

5. **Ensembl Variant Effect Predictor (VEP)**: Ensembl'in VEP aracı, varyantların yorumlanmasını sağlar.

<https://www.ensembl.org/info/docs/tools/vep/index.html>

<https://www.ensembl.org/downloads.html>

Bu araçların her biri farklı bir perspektif sunar; varyantın klinik önemini kapsamlı şekilde anlamak için birden fazla kaynağı kullanmak faydalı olacaktır.

NGS İçin Programlama Dilleri ve Özellikleri

1. Python

- Avantajları: Python, biyoinformatikte oldukça popülerdir çünkü kolay öğrenilebilir ve geniş bir kütüphane ekosistemine sahiptir. **Biopython**, **scikit-bio**, ve **pandas** gibi biyoinformatik ve veri analizi için güçlü kütüphaneler sunar.
- Kullanım Alanları: Dizileme verilerinin işlenmesi, genomik verilerin analizi, veri görselleştirme ve makine öğrenimi uygulamaları.

2. R

- Avantajları: İstatistiksel analizler için geliştirilmiş olan R, genetik veri analizi için de çok uygundur. Bioconductor kütüphanesi, genomik verilerle çalışma konusunda kapsamlı araçlar sunar.
- Kullanım Alanları: İstatistiksel genetik analizler, farklılaşma analizleri, varyant analizleri ve veri görselleştirme.

NGS İçin Programlama Dilleri ve Özellikleri (ii)

1. Rust

- Avantajları: Rust, bellek güvenliği ve yüksek performans sağlar, bu nedenle biyoinformatikte karmaşık ve yoğun hesaplama gerektiren uygulamalar için tercih edilebilir.
- rust-bio, bio-types, noodles, serde, serde_json, ndarray, linfa ve polars Rust ile genetik analiz yapmak isteyenler için temel bir altyapı sağlar.

2. SQL ve NoSQL Dilleri (Örneğin, MongoDB)

- Avantajları: Genetik verileri depolama ve sorgulama amaçlı kullanılır. SQL, yapılandırılmış veri depolamada iyi iken NoSQL veritabanları büyük ve karmaşık genomik verileri depolamada esneklik sağlar.

Anlatıcının seçimi:  **Rust** ve  **mongoDB**

NGS İçin Yazılım Araçları

- **ANNOVAR**
- **BCFTools**
- **Bowtie ve Bowtie2**
- **BWA**
- **DeepVariant**
- **FastQC**
- **FreeBayes**
- **GATK**
- **HISAT2**
- **IGV (Integrative Genomics Viewer)**
- **SAM Tools**
- **SnpEff**

ANNOVAR

ANNOVAR, biyoinformatik ve genetik araştırmalarda kullanılan bir varyant anotasyon aracıdır. Dizileme verilerinden elde edilen genetik varyantların (örneğin, SNP'ler, indel'ler, yapısal varyantlar) biyolojik anlamlarını belirlemek ve bu varyantların potansiyel etkilerini tahmin etmek amacıyla kullanılır. ANNOVAR, özellikle varyant çağırma (variant calling) işlemi sonrasında elde edilen VCF (Variant Call Format) dosyalarını işleyerek varyantların biyolojik bağlamda anlamlandırılması için geliştirilmiştir.

Kurulum:

<https://annovar.openbioinformatics.org/en/latest/user-guide/download/>

Adresine edu.tr mail adresi ile kaydolunca girince indirme adresini mail adresinize gönderiyorlar.

If you are a commercial user/service provider, you are required to purchase a license to ANNOVAR from QIAGEN.

ANNOVAR'ın Temel İşlevleri

- Genomik Bölgeye Göre Anotasyon: Varyantların genom üzerindeki konumlarına göre tanımlanmasını sağlar (örneğin, ekzon, intron, promotör bölgede olup olmadıklarını belirler).
- Fonksiyonel Anotasyon: Varyantın protein kodlayan bir gen üzerinde olup olmadığını, varsa proteinde hangi değişikliklere yol açtığını belirler. Örneğin, bir varyantın bir amino asit değişimine, çerçeve kaymasına veya durma kodonuna neden olup olmadığını analiz eder.
- Klinik Anotasyon: Varyantların insan hastalıkları ile ilişkisini belirlemek için ClinVar, OMIM, HGMD gibi klinik veri tabanları ile eşleştirme yapar.
- Popülasyon Frekansı Anotasyonu: Varyantların farklı popülasyonlardaki frekanslarını hesaplamak için 1000 Genomes, gnomAD gibi popülasyon veritabanları ile entegrasyon sağlar.
- Pathogenicity Tahmini: SIFT, PolyPhen gibi araçlarla entegre olarak varyantların patojenik (hastalığa neden olma potansiyeli) olup olmadığını tahmin eder.

ANNOVAR'ın Temel İşlevleri (ii)

- Filogenetik Anotasyon: Türler arası varyantların korunma derecesini (conservation score) hesaplayarak varyantın evrimsel açıdan ne kadar önemli olabileceğini analiz eder.

ANNOVAR Kullanım Alanları

- Genetik Hastalık Araştırmaları: Hastalıkla ilişkili olabilecek genetik varyantların tanımlanması.
- Kanser Genomiği: Kanser hücrelerinde ortaya çıkan mutasyonların işlevsel ve klinik etkilerinin değerlendirilmesi.
- Kişisel Genomik: Bireylerin genomlarındaki varyantların biyolojik ve klinik anlamının belirlenmesi.
- Popülasyon Genetiği Çalışmaları: Popülasyonlardaki varyant dağılımlarını ve bu varyantların biyolojik anlamlarını incelemek.

ANNOVAR Kullanımı

```
$ annotate_variation.pl -buildver hg19 -downdb -webfrom annovar refGeneWithVer humandb/  
$ annotate_variation.pl -buildver hg19 -downdb cytoBand humandb/  
$ annotate_variation.pl -buildver hg19 -downdb -webfrom annovar gnomad211_exome humandb/  
$ annotate_variation.pl -buildver hg19 -downdb -webfrom annovar avsn151 humandb/  
$ annotate_variation.pl -buildver hg19 -downdb -webfrom annovar dbnsfp47a humandb/
```

Protocol: refGene, ClinVar, gnomAD, avsn151, dbnsfp47a, cytoBand

Buildver : hg19 or hg38

☞ Gen tabanlı, bölge tabanlı ve filtre tabanlı anotasyon gerçekleştirmek için;

```
$ perl table_annovar.pl example/ex2.vcf humandb/  
-buildver hg19 -out my_first_anno -protocol refGeneWithVer -operation g \  
-remove -polish -vcfinput -nastring .
```

☞ Her varyantın etkisini, yerini, sıklığını ve potansiyel klinik etkisini açıklayan açıklamalı dosyalar oluşturmak için;

```
$ perl table_annovar.pl example/ex1.avinput humandb/  
-buildver hg19 -out ex1_anno  
-protocol refGeneWithVer,cytoBand,gnomad211_exome,avsn151,dbnsfp47a  
-operation gx,r,f,f,f  
-xref example/gene_xref.txt  
-remove -nastring . -csvout -polish
```

BCFTools

BCFTools, biyoinformatik alanında genetik varyantları işlemek ve analiz etmek için kullanılan bir yazılım aracıdır. Özellikle DNA veya RNA dizileme verilerinde bulunan varyantları çağırma, filtreleme, dönüştürme ve analiz etme işlemlerinde sıkça kullanılır. BCFTools, hem VCF (Variant Call Format) hem de BCF (Binary Call Format) dosyaları ile çalışabilen ve genetik varyantları hızlı ve verimli bir şekilde işlemek için optimize edilmiş bir araçtır.

<https://samtools.github.io/bcftools>

veya

```
$ sudo apt install bcftools
```

BCFTools'un Temel İşlevleri

- **Varyant Çağırma (Variant Calling):** Dizileme verisinde bulunan SNP'leri (tek nükleotid polimorfizmleri) ve indel'leri (ekleme/silme varyantları) tespit etmek için kullanılır.
- **Varyant Filtreleme:** Belirli kriterlere göre varyantları filtreleyerek analiz için en uygun olanları seçebilir. Örneğin, belirli bir kalite eşiğinin altındaki varyantları çıkarabilirsiniz.
- **Dönüştürme ve Sıkıştırma:** VCF dosyalarını sıkıştırılmış BCF formatına dönüştürebilir ve veri depolama veya paylaşma sürecinde yerden tasarruf sağlar.
- **Varyant Birleştirme (Merging):** Farklı numunelerden elde edilen VCF veya BCF dosyalarını birleştirerek, çoklu örnek analizleri için ortak bir dosya oluşturur.
- **Popülasyon Genetiği Analizleri:** Genetik varyantların popülasyon düzeyindeki dağılımlarını analiz etmek ve varyant frekanslarını hesaplamak için kullanılır.

BCFTools Örnekleri

Variant Filtreleme;

```
$ bcftools filter -i 'QUAL > 30 && DP > 10' in.vcf -o out.vcf
```

Variant Annotasyonu;

```
$ bcftools annotate -a ano_file.vcf.gz  
-c CHROM,POS,REF,ALT,INFO/ANNOTATED_FIELD input.vcf -o out.vcf
```

Dönüştürme;

```
$ bcftools view -O b input.vcf -o output.bcf
```

Birleştirme;

```
$ bcftools merge sample1.vcf sample2.vcf -o merged_output.vcf
```

İstatistiksel Analiz;

```
$ bcftools stats input.vcf > stats_output.txt
```

Bowtie ve Bowtie2



Bowtie ve Bowtie2, biyoinformatikte ham dizileme verilerinin referans bir genomla hizalanması (alignment) için kullanılan iki popüler yazılımdır. Bu araçlar, dizileme işlemi sonucunda elde edilen kısa DNA veya RNA okuma parçalarını (reads) referans genomdaki uygun konumlara yerleştirir. Bowtie ve Bowtie2, özellikle hızları ve düşük bellek kullanımlarıyla bilinir, bu nedenle büyük verisetleri üzerinde çalışma gerektiren projelerde sıkça tercih edilirler.

<https://github.com/BenLangmead/bowtie2>

Kurulum;

```
$ sudo apt install bowtie2 bowtie
```

Bowtie2 Kullanım

Bowtie2 İndeksi Oluşturma;

```
$ bowtie2-build ref_genome.fasta ref_index
```

Tek Uçlu Okuma (Single-End Read) Hizalama;

```
$ bowtie2 -x ref_index -U reads.fastq -S output.sam
```

Çift Uçlu Okuma (Paired-End Read) Hizalama;

```
$ bowtie2 -x ref_index -1 reads_1.fastq -2 reads_2.fastq  
-S paired_output.sam
```

Çıktıyı BAM Formatında Almak;

```
$ bowtie2 -x ref_index -U reads.fastq | samtools view -Sb - > output.bam
```

Sadece En İyi Eşleşmeleri Çıkarma;

```
$ bowtie2 -x ref_index -U reads.fastq --best -S best_output.sam
```

BWA (Burrows-Wheeler Aligner)

BWA (alignment via Burrows-Wheeler transformation) , biyoinformatikte genetik dizileme verilerini referans bir genomla hizalamak (alignment) için kullanılan hızlı ve hafif bir araçtır. Burrows-Wheeler Transformation (BWT) algoritması kullanılarak çalışır, bu sayede büyük genomik veri setlerini yüksek hız ve düşük bellek kullanımı ile işleyebilir.

BWA'nın Çalışma Prensipleri

BWA, hizalama işlemi için Burrows-Wheeler Transform ve FM-Index algoritmalarını kullanır. Bu algoritmalar, veri sıkıştırma prensipleriyle çalışarak, referans genom ve okuma parçalarını daha hızlı bir şekilde eşleştirir. Burrows-Wheeler Transform (BWT) veriyi sıkıştırarak depolar ve dizinin daha verimli aranmasını sağlar. Bu yöntemle, veride arama ve hizalama işlemleri daha hızlı gerçekleşir.

Kurulum;

```
$ sudo apt instal bwa
```

BWA Kullanımı

İndeks Oluşturma;

```
$ bwa index ref_genome.fasta
```

BWA-MEM ile Tek Uçlu Okuma (Single-End) Hizalama

```
$ bwa mem ref_genome.fasta reads.fastq > output.sam
```

BWA-MEM ile Çift Uçlu Okuma (Paired-End) Hizalama;

```
$ bwa mem ref_genome.fasta  
  reads_1.fastq reads_2.fastq > paired_output.sam
```

BWA'nin Kullanım Alanları

- **Genetik Hastalık Araştırmaları:** Genetik varyantların ve hastalıkla ilişkili mutasyonların saptanmasında kullanılır.
- **Kanser Genomiği:** Kanser hücrelerinde mutasyonları ve yapısal varyantları incelemek için kullanılır.
- **Popülasyon Genetiği:** Genetik varyantların popülasyonlardaki dağılımını incelemek için geniş çapta kullanılır.

BWA'nin Avantajları

- Hızlı ve Verimli: Büyük veri setlerinde hızlı hizalama sunar.
- Düşük Bellek Kullanımı: BWT sayesinde minimum bellekle çalışır.
- Doğruluk: BWA-MEM, çeşitli okuma uzunluklarında yüksek doğruluk sağlar.
- Esneklik: Hem kısa hem de uzun okuma verilerini işleyebilir.

DeepVariant DeepVariant

DeepVariant, Google tarafından geliştirilen ve genomik varyant tespiti için derin öğrenme tekniklerini kullanan bir varyant çağırma (variant calling) aracıdır.

DeepVariant, DNA dizileme verilerini analiz ederek tek nükleotid varyantları (Single Nucleotide Variants - SNV), indeller (insertion-deletion) ve diğer genetik varyasyonları oldukça yüksek bir doğrulukla tespit edebilir. GitHub deposundan indirilebilir.

<https://github.com/google/deepvariant>

DeepVariant'in Özellikleri

- Derin Öğrenme Temelli Yaklaşım
- Evrişimli Sinir Ağları (Convolutional Neural Network - CNN)
- Platform Bağımsız Performans (Illumina, Pacific Biosciences ve Oxford Nanopore gibi farklı DNA dizileme platformları)
- Otomatik Model Eğitimi ve Sürekli Güncellenen Modeller
- Yüksek Doğruluk ve Hassasiyet

DeepVariant Kullanım Alanları

- **Klinik Genetik ve Tanı Çalışmaları**

DeepVariant, hastalık ilişkili genetik varyantların doğru bir şekilde tespiti için kullanılır. Hastalık genetiği çalışmalarında hassas ve güvenilir sonuçlar elde etmeyi sağlar.

- **Kansere Özgü Mutasyon Analizleri**

Derin öğrenme tabanlı analizler sayesinde, DeepVariant düşük frekanslı kanser mutasyonlarının tespiti gibi zorlu analizlerde bile yüksek doğruluğa ulaşabilir.

- **Nadir Genetik Hastalıklar**

Nadir hastalıklara neden olan mutasyonların tespitinde, DeepVariant'ın yüksek doğruluğu önemli bir avantaj sağlar.

DeepVariant Örnek Kullanım

```
BIN_VERSION="1.6.1"
```

```
docker run \
```

```
-v "YOUR_INPUT_DIR":"/input"
```

```
-v "YOUR_OUTPUT_DIR:/output"
```

```
google/deepvariant:"1.6.1"
```

```
/opt/deepvariant/bin/run_deepvariant
```

```
--model_type=[WGS,WES,PACBIO,ONT_R104,HYBRID_PACBIO_ILLUMINA]
```

```
--ref=/input/YOUR_REF \
```

```
--reads=/input/YOUR_BAM \
```

```
--output_vcf=/output/YOUR_OUTPUT_VCF \
```

```
--output_gvcf=/output/YOUR_OUTPUT_GVCF \
```

```
--num_shards=$(nproc) \
```

```
--logging_dir=/output/logs \
```

```
--haploid_contigs="chrX,chrY" \
```

```
--par_regions_bed="/input/GRCh3X_par.bed"
```

FreeBayes

FreeBayes, özellikle genetik araştırmalarda kullanılan bir varyant çağırıcı (variant caller) yazılımıdır. FreeBayes, bir genom ya da bir grup genom üzerinde tek nükleotid varyantları (Single Nucleotide Variants - SNV), indel (insertion-deletion), yapısal varyasyonlar gibi genetik varyasyonları tespit etmek için geliştirilmiştir.

FreeBayes'in Özellikleri

1. Bayesçi Yaklaşım FreeBayes, adından da anlaşılacağı gibi varyantları belirlerken Bayesçi olasılık hesaplamalarına dayanır. Bu yaklaşım sayesinde varyantların güvenilirliği, gözlemler ve olasılıklar üzerinden belirlenir, bu da sonuçların daha doğru olmasını sağlar.
2. Multi-Sample Desteği FreeBayes, birden fazla örneği (örneğin, aynı popülasyonun farklı bireyleri) aynı anda analiz etme yeteneğine sahiptir. Bu sayede popülasyon genetiği çalışmalarında yaygın olarak kullanılabilir.

FreeBayes (ii)

3. Haplotip Bazlı Çıkarımlar FreeBayes, haplotipler üzerinden çalışır ve ardışık varyantları bir araya getirerek daha doğru tahminlerde bulunur. Bu özellikle bağlı varyantların daha hassas bir şekilde belirlenmesine yardımcı olur.
4. Esneklik ve Özelleştirme FreeBayes oldukça esnek bir yazılımdır ve çeşitli parametreler aracılığıyla analizleri özelleştirme olanağı sunar. Kullanıcılar kalite kontrol, filtreleme ve diğer analiz seçeneklerini ayarlayabilir.
5. Düşük Allel Frekansı Duyarlılığı FreeBayes, düşük frekanslı allelleri de algılayabilme kapasitesine sahiptir, bu nedenle tümör heterojenitesi gibi kompleks örneklerde kullanıldığında oldukça faydalıdır.

FreeBayes Kullanımı

- FreeBayes <https://github.com/freebayes/freebayes> indirilebilir.
- FreeBayes MIT Lisansı ile lisanslanmıştır.

Örnek komutlar;

```
# VCF Call
```

```
$ freebayes -f ref.fa aln.bam >var.vcf
```

```
# Sadece chrQ
```

```
$ freebayes -f ref.fa -r chrQ aln.bam >var.vcf
```

```
# chrQ, 1000 ile 2000 arası
```

```
$ freebayes -f ref.fa -r chrQ:1000-2000 aln.bam >var.vcf
```

```
# Filtreleme
```

```
$ freebayes -f ref.fa aln.bam |  
    vcffilter -f "QUAL > 20" >results.vcf
```

GATK (Genome Analysis Toolkit)

GATK, genetik varyantları tespit etmek, analiz etmek ve biyoinformatik verilerle çalışmak için geniş çapta kullanılan, güçlü ve kapsamlı bir araçtır. Broad Institute tarafından geliştirilen GATK, özellikle yüksek verimli dizileme (NGS) verilerinde varyant çağırma (variant calling) işlemleri için optimize edilmiştir. GATK, DNA veya RNA dizileme verilerinden SNP'ler (tek nükleotid polimorfizmleri), indels'ler (ekleme/silme), yapısal varyantlar ve kopya sayısı varyantları (CNV) gibi genetik değişikliklerin tespitinde kullanılabilir.

Web Adresi: <https://github.com/broadinstitute/gatk>

adresinden indirilebilir.

GATK, varyant analiz sürecini kolaylaştırmak ve doğruluğu artırmak amacıyla çeşitli araçlar sunar.

GATK'nin Kullanım Alanları

GATK, geniş özellik yelpazesi sayesinde pek çok biyoinformatik ve genetik araştırmada kullanılır:

- **Klinik Genetik:** Genetik hastalıkların tanısında önemli olan varyantları tespit etmek için.
- **Kanser Genomiği:** Tümörlerdeki somatik mutasyonları belirlemek ve kanser araştırmalarını desteklemek için.
- **Popülasyon Genetiği:** Genetik varyantların popülasyonlardaki dağılımını incelemek ve varyant frekanslarını belirlemek için.
- **Kişisel Genomik:** Bireylerin genomlarındaki varyantların analiz edilmesi ve sağlık risklerinin değerlendirilmesi için.

GATK'nin Avantajları

- **Yüksek Doğruluk:** GATK, çeşitli hataları düzeltecek araçlarla donatılmıştır ve sonuçların doğruluğunu artırmak için kalite kontrol adımları sunar.
- **Esneklik:** Farklı analiz ihtiyaçlarına uygun geniş bir araç yelpazesine sahiptir.

GATK'nin Kullanım Alanları (ii)

- **Yaygın Kullanım:** Biyoinformatikte standart hale gelmiş bir araçtır ve geniş bir kullanıcı topluluğuna sahiptir.

Örnek GATK Kullanımı

1. İndeks Oluşturma;

```
$ gatk CreateSequenceDictionary -R hg38.fasta samtools faidx hg38.fasta
```

2. Okuma Parçalarını Referans Genomla Hizalama

```
$ bwa mem hg38.fasta reads_1.fastq reads_2.fastq > aligned_reads.sam
```

3. SAM Dosyasını BAM Formatına Dönüştürme ve Sıralama

```
$ samtools view -Sb aligned_reads.sam | samtools sort -o sorted_reads.bam
```

4. Çoğaltmaları İşaretleme (Mark Duplicates)

```
$ gatk MarkDuplicates -I sorted_reads.bam -O dedup_reads.bam -M duplicates_report.txt
```

5. Baz Kalitesi Yeniden Kalibrasyonu (BQSR)

```
$ gatk BaseRecalibrator -I dedup_reads.bam -R hg38.fasta  
                        --known-sites known_sites.vcf -O recal_data.table  
$ gatk ApplyBQSR -R hg38.fasta -I dedup_reads.bam  
                --bqsr-recal-file recal_data.table -O recal_reads.bam
```


Örnek GATK Kullanımı (ii)

6. Varyant Çağırma (Variant Calling) - HaplotypeCaller Kullanımı

```
$ gatk HaplotypeCaller -R hg38.fasta -I recal_reads.bam -L target_regions.bed -O  
raw_variants.vcf
```

7. Varyant Kalite Filtreleme

```
$ gatk VariantFiltration -R hg38.fasta -V raw_variants.vcf  
-O filtered_variants.vcf  
--filter-expression "QD < 2.0 || FS > 60.0 || MQ < 40.0"  
--filter-name "basic_filters"
```

Örnek GATK Kullanımı (iii)

8. Varyant Anotasyonu

Son olarak, varyantları biyolojik olarak anlamlandırmak için ANNOVAR veya VEP gibi anotasyon araçları kullanarak varyantlara açıklama ekleyebilirsiniz. Bu araçlar, varyantların hangi genlerde olduğunu, hastalıkla ilişkilerini ve diğer biyolojik bilgileri sağlar.

```
$ table_annovar.pl filtered_variants.vcf humandb/  
-buildver hg38 -out annotated_variants  
-remove -protocol refGene,clinvar_20180603,dbnsfp35a  
-operation g,f,f -nastring .
```

Bu adımlar, GATK kullanarak WES analizinde varyant çağırma ve kalite kontrol işlemlerini kapsar. Sonuç olarak, sadece ekzon bölgelerindeki yüksek kaliteli varyantlar elde edilmiş olur. Bu varyantlar biyolojik bağlamda anlamlandırılabilir ve klinik araştırmalar veya hastalık analizleri için kullanılabilir.

SAM Tools

SAMtools, biyoinformatikte DNA ve RNA dizileme verilerini işlemek için geliştirilmiş bir yazılım paketidir. SAMtools, özellikle SAM (Sequence Alignment/Map) ve BAM (Binary Alignment/Map) formatında hizalanmış verileri işlemek, analiz etmek ve yönetmek için kullanılır. SAM ve BAM dosyaları, genetik dizilimlerin referans bir genomla hizalanmasını temsil eder ve SAMtools bu dosyalar üzerinde çeşitli işlemler yapmayı sağlar.

SAMtools'un Kullanım Alanları

- Varyant Çağırma: BCFtools ile birlikte kullanıldığında SNP ve indel varyantlarını tespit etmek için idealdir.
- Hizalama Kalitesi Değerlendirme: Okuma parçalarının genel kalitesi, hizalama doğruluğu ve çoğaltma oranlarını analiz ederek veriyi değerlendirmeye yardımcı olur.
- Veri Alt Kümesi Oluşturma: Büyük BAM dosyalarından sadece analiz için gerekli olan bölgeleri veya yüksek kaliteli okuma parçalarını ayırarak daha hızlı analiz yapılmasını sağlar.

SAM Tools Kullanım Örnekleri

Dosya Dönüştürme (SAM ve BAM Formatları Arasında):

```
$ samtools view -Sb input.sam > output.bam
```

Sıralama (Sorting):

```
$ samtools sort input.bam -o sorted.bam
```

İndeksleme (Indexing):

```
$ samtools index sorted.bam
```

Büyük BAM dosyalarına hızlı erişim sağlamak için indeksleme yapılır. İndekslenen dosyalar .bai uzantılı bir indeks dosyası oluşturur ve analiz yazılımlarının veriye daha hızlı ulaşmasına olanak tanır.

Filtreleme ve Alt Küme Oluşturma:

```
$ samtools view -q 30 input.bam > high_quality.bam
```

Bölgeye Göre Çıkarma (Region Extraction):

```
$ samtools view -b sorted.bam "chr1:1000000-2000000" > region.bam
```

SnpEff

SnpEff, biyoinformatikte varyant anotasyonu ve varyantların etkilerini tahmin etmek için kullanılan bir yazılım aracıdır. Dizileme verilerinden elde edilen SNP (tek nükleotid polimorfizmleri) ve indel (ekleme/silme) gibi genetik varyantların, genomdaki genler ve proteinler üzerindeki potansiyel etkilerini analiz etmek için tasarlanmıştır. SnpEff, varyantların biyolojik anlamını anlamak ve klinik veya araştırma amaçlı değerlendirmelerini yapmak için yaygın olarak kullanılır.

SnpEff'in Temel İşlevleri ve Özellikleri

1. Varyant Anotasyonu (Annotation): SnpEff, varyantların genomda hangi bölgelere denk geldiğini tanımlar. Varyantın bir genin ekzon veya intron gibi belirli bir bölgesinde olup olmadığını, veya promotör gibi düzenleyici bir bölgede olup olmadığını belirleyerek bu varyantlara açıklama ekler.
2. Varyantların Etki Analizi (Effect Prediction): SnpEff, varyantların protein kodlayan bölgelerde yol açtığı amino asit değişikliklerini ve diğer etkilerini tahmin eder. Örneğin, bir varyantın stop kodon oluşturup oluşturmadığını veya bir çerçeve

Snpeff (ii)

kaymasına yol açıp açmadığını tespit eder. Varyant etkileri, yüksek etki (örneğin, çerçeve kayması, stop kodonu oluşturma), orta etki (örneğin, amino asit değişikliği) veya düşük etki (örneğin, sessiz mutasyonlar) olarak sınıflandırılır.

3. Klinik Veri Tabanları ile Entegrasyon: Snpeff, ClinVar gibi klinik veri tabanlarına erişim sağlayarak varyantların bilinen hastalıklarla ilişkili olup olmadığını da belirtebilir. Bu sayede varyantların potansiyel patojenik etkilerini daha kapsamlı bir şekilde değerlendirir.
4. Genom Destekleri: İnsan, fare, sinek ve diğer birçok model organizmanın genomları için destek sunar. Farklı organizmaların genom versiyonlarına özel veritabanları sağlar.
5. Veri Çıktı Formatı: Snpeff, anotasyon sonuçlarını VCF formatında sunar. VCF (Variant Call Format) dosyasına eklenen bu anotasyon bilgileri, varyantın genomdaki etkilerini özetleyen ek bilgiler içerir. Snpeff ile Sağlanan Bilgiler
6. Anotasyon sürecinde Snpeff, her varyant için aşağıdaki bilgileri sağlayabilir:

SnpEff (iii)

- Varyantın Genomdaki Konumu: Hangi gen veya genomik bölgede bulunduğu.
- Protein Üzerindeki Etkisi: Bir amino asit değişikliğine yol açıp açmadığı veya protein yapısını nasıl etkilediği.
- Varyantın Türü: Örneğin, synonymous (sessiz) varyant, nonsynonymous (amino asit değişikliğine yol açan) varyant, çerçeve kayması, stop-gain (durdurucu kodon oluşturma) veya intronik varyant gibi sınıflandırmalar.
- Klinik Bilgiler: Bilinen patojenik varyantlarla eşleşme olup olmadığı.

7. SnpEff'in Kullanım Alanları

- Genetik Hastalık Araştırmaları: Hastalıklarla ilişkili olabilecek varyantların tanımlanması.
- Kanser Genomiği: Tümör hücrelerinde meydana gelen mutasyonların potansiyel etkilerinin tahmin edilmesi.
- Kişisel Genomik ve Popülasyon Genetiği: Kişisel genomik analizlerde veya popülasyonlardaki varyantların biyolojik anlamlarının değerlendirilmesi.

SnpEff (iv)

- Fonksiyonel Genomik: Genlerin protein yapısı ve işlevi üzerindeki etkilerin değerlendirilmesi.

8. SnpEff'in Avantajları

- Yüksek Hız: SnpEff, büyük varyant dosyalarını hızlı bir şekilde analiz eder.
- Esneklik: Farklı organizma genomları için geniş veritabanı desteği sunar.
- Kapsamlı Anotasyon: SnpEff, sadece varyantları tanımlamakla kalmaz, aynı zamanda onların biyolojik etkilerini de detaylı bir şekilde sınıflandırır.

SnpEff, biyoinformatik analizlerde varyantların biyolojik ve klinik anlamını anlamak için güçlü bir araçtır ve araştırmacılara varyantların potansiyel etkilerini hızlı ve doğru bir şekilde değerlendirme olanağı sunar.

SnEff ile WES Örneği

Kurulum

```
$ sudo apt install snpeff
```

Adım 1: SnEff Veritabanını İndirin

```
$ snpEff databases
```

```
$ snpEff -v -download hg38
```

Adım 2: Varyant Dosyasını (VCF) Anotasyonlama

```
$ snpEff -v hg38 wes_variants.vcf > annotated_wes_variants.vcf
```

Adım 3: Anotasyon Sonuçlarını İnceleme

snpEff tarafından oluşturulan annotated_wes_variants.vcf dosyasında, her varyant hakkında ek bilgiler bulunur. Bu bilgiler arasında şunlar yer alabilir:

- Genetik Bölge: Varyantın hangi gen veya bölgeye denk geldiği (örneğin, ekzon, intron).
- Protein Üzerindeki Etki: Bir amino asit değişikliğine yol açıp açmadığı veya protein yapısını nasıl etkilediği.

SnpEff ile WES Örneği (ii)

- Varyantın Türü: Synonymous (sessiz), missense (amino asit değişikliğine yol açan), nonsense (stop kodonu oluşturan) gibi türler.
- Etkilerin Derecesi: Yüksek, orta veya düşük etki olarak sınıflandırılır.

SnEff ile WES Örneği (iii)

Adım 4: Sadece Yüksek Etkili Varyantları Filtreleme (Opsiyonel)

Anotasyon sonucunda sadece yüksek etkili varyantları analiz etmek istiyorsanız grep veya benzeri bir komut kullanarak annotated_wes_variants.vcf dosyasından bu bilgileri çekebilirsiniz. Örneğin:

```
$ grep "HIGH" annotated_wes_variants.vcf > high_impact_variants.vcf
```

Bu komut, annotated_wes_variants.vcf dosyasındaki yüksek etkili varyantları filtreleyerek high_impact_variants.vcf dosyasına yazar.

Adım 5: Anotasyonları Değerlendirme ve Raporlama

SnEff çıktısında varyantların her birine dair açıklayıcı bilgiler bulunduğundan, klinik ve araştırma açısından önemli olabilecek varyantları seçerek bir rapor oluşturabilirsiniz. Örneğin, şu bilgilere odaklanabilirsiniz:

- Hastalıkla ilişkili genlerdeki varyantlar.
- Yüksek etkili varyantlar (örneğin, çerçeve kayması veya durdurucu kodonlar).

SnpEff ile WES Örneği (iv)

- Literatürde bilinen patojenik varyantlarla eşleşenler.

Bu adımlar sonucunda, SnpEff kullanarak WES verilerindeki varyantları anotasyonlayabilir ve anlamlandırabilirsiniz. Bu süreç, klinik araştırmalar, genetik hastalık analizleri veya kişisel genom analizleri için oldukça yararlıdır.

Yapay Zeka Destekli WES ve CES Analizi

WES ve CES, genomdaki ekzonik bölgelerin (protein kodlayan genlerin) dizilenmesine odaklanır. CES, WES'e göre daha sınırlı bir analiz olup yalnızca klinik açıdan önemli olduğu bilinen belirli genlerin dizilenmesini içerir. Yapay zeka, bu analizlerin her iki türünde de çok miktarda veriyi işleyerek sonuçları daha hızlı ve anlamlı hale getirmeye yardımcı olur.

1. Varyant Filtreleme ve Önceliklendirme

Yapay zeka, WES ve CES verilerinde bulunan binlerce varyant arasından klinik olarak anlamlı olanları seçmek için filtreleme ve önceliklendirme algoritmaları kullanır. Bu süreç, veriyi azaltarak genetik test sonuçlarının hızla yorumlanmasını sağlar.

- **Makine Öğrenimi Modelleri:** SVM (Destek Vektör Makineleri), rastgele ormanlar (random forest) ve k-nearest neighbors (KNN) gibi makine öğrenimi algoritmaları, varyantların biyolojik anlamlarını öğrenebilir ve varyantları klinik önceliğe göre sıralayabilir.

Yapay Zeka Destekli WES ve CES Analizi (ii)

- **Derin Öğrenme Modelleri:** Derin sinir ağları, varyantların patojenik potansiyelini tahmin etmek için genetik verideki kalıpları öğrenebilir. Bu sayede, daha önce bilinmeyen varyantlar hakkında da bilgi edinilebilir.

2. Patojenik Varyant Tahmini

WES ve CES verilerinde bir varyantın patojenik (hastalık yapıcı) olup olmadığını belirlemek, klinik önem taşıyan bir sorudur. Yapay zeka, genetik varyantların patojenik veya benign (zararsız) olup olmadığını tahmin etmek için çeşitli algoritmalar kullanır.

- **Varyant Etki Tahmin Araçları:** PolyPhen, SIFT, ve CADD gibi yapay zeka destekli araçlar, varyantın amino asit değişiklikleri veya protein yapısı üzerindeki etkisini değerlendirir.
- **Yapay Zeka ile Desteklenen Puanlama Sistemleri:** CADD (Combined Annotation Dependent Depletion) gibi araçlar, varyantların işlevsel önemini tahmin etmek için yapay zeka algoritmalarını kullanır. Bu tür puanlama sistemleri, varyantları belirli bir eşik değere göre sınıflandırarak analiz sürecini hızlandırır.

Yapay Zeka Destekli WES ve CES Analizi (iii)

3. Gen-Fenotip İlişkilendirmesi

Yapay zeka, gen varyantlarının belirli hastalık fenotipleriyle olan ilişkisini anlamada da önemli rol oynar. CES analizinde özellikle klinik belirtilerle genetik varyantlar arasında ilişki kurmak önemlidir.

- **Doğal Dil İşleme (NLP):** NLP algoritmaları, literatürde veya veri tabanlarında bulunan gen-hastalık ilişkilerini tarayarak varyantların fenotipik etkilerini belirlemek için kullanılabilir. Bu tür algoritmalar, bilimsel makalelerden hastalıklarla ilişkili varyantları otomatik olarak çıkarabilir.
- **Fenotip-Tabanlı Makine Öğrenimi Modelleri:** Bu modeller, hastalık fenotiplerini gen varyantları ile ilişkilendirmeyi öğrenerek hastalıkla ilişkili yeni varyantları tahmin edebilir. Örneğin, HPO (Human Phenotype Ontology) terimleriyle eşleştirme yaparak analizleri anlamlandırabilir.

4. Klinik Karar Destek Sistemleri

WES ve CES analizlerinde elde edilen sonuçların klinik anlamda yorumlanması, tedavi süreçlerinin planlanması açısından çok önemlidir. Yapay zeka destekli klinik karar

Yapay Zeka Destekli WES ve CES Analizi (iv)

destek sistemleri, hastaya özel genetik bilgiyi kullanarak doktorlara daha hızlı ve kesin kararlar almaları için yardımcı olabilir.

- **Hastalık Tahmin ve Sınıflandırma:** AI algoritmaları, genetik varyantları ve fenotip bilgilerini kullanarak hastalık riskini tahmin eder ve varyantları patojenik veya olası patojenik olarak sınıflandırır.
- **Klinik Raporlama:** AI tabanlı yazılımlar, WES veya CES sonuçlarını kullanarak doktorlara özelleştirilmiş raporlar sunar ve böylece karmaşık verilerin kolayca anlaşılmasını sağlar.

5. Veri Entegrasyonu

Yapay zeka, WES ve CES analizlerinde birden fazla kaynaktan gelen verileri bir araya getirerek bütüncül bir analiz sağlar. Genetik verilerin yanı sıra, klinik veriler, hastalık veri tabanları ve hatta çevresel faktörler gibi çeşitli veri kaynaklarını entegre ederek daha kapsamlı sonuçlar elde edilir.

- **Çok Katmanlı Öğrenme Modelleri:** AI, genetik varyantlar ile çevresel ve klinik faktörleri aynı anda analiz eden çok katmanlı modeller kurabilir. Bu, örneğin

Yapay Zeka Destekli WES ve CES Analizi (v)

hastalığın sadece genetik bir sebebe mi yoksa çevresel bir etkene mi bağlı olduğunu anlamaya yardımcı olur.

Avantajları

- **Hız ve Verimlilik:** Binlerce varyant arasından önemli olanları hızlıca filtreler, zaman kazandırır.
- **Daha Doğru Sonuçlar:** AI, genetik varyantları biyolojik bağlamda daha doğru anlamlandırır.
- **Klinik Anlamda Daha Değerli Veriler:** AI, varyantları klinik fenotiplerle ilişkilendirerek daha kullanışlı raporlar sunar.
- **Yeni Varyantların Keşfi:** Yapay zeka, henüz bilinmeyen veya yeni ortaya çıkan varyantları anlamak için tahmin yapabilir.

Zorluklar

- **Veri Güvenliği ve Gizliliği:** Genetik verilerin korunması ve gizliliği son derece önemlidir.
- **Yüksek Hesaplama Gücü Gereksinimi:** Büyük genom verilerinin işlenmesi yoğun hesaplama gücü gerektirir.
- **Veri Kalitesi ve Çeşitliliği:** AI'nın güvenilir analizler yapabilmesi için çeşitli ve kaliteli veriye ihtiyaç vardır.