

Genomik Veri Tabanları

Do. Dr. Deniz İlhan Topcu

İzmir Şehir Hastanesi

Tıbbi Biyokimya Bölümü



Kısa Özgeçmiş

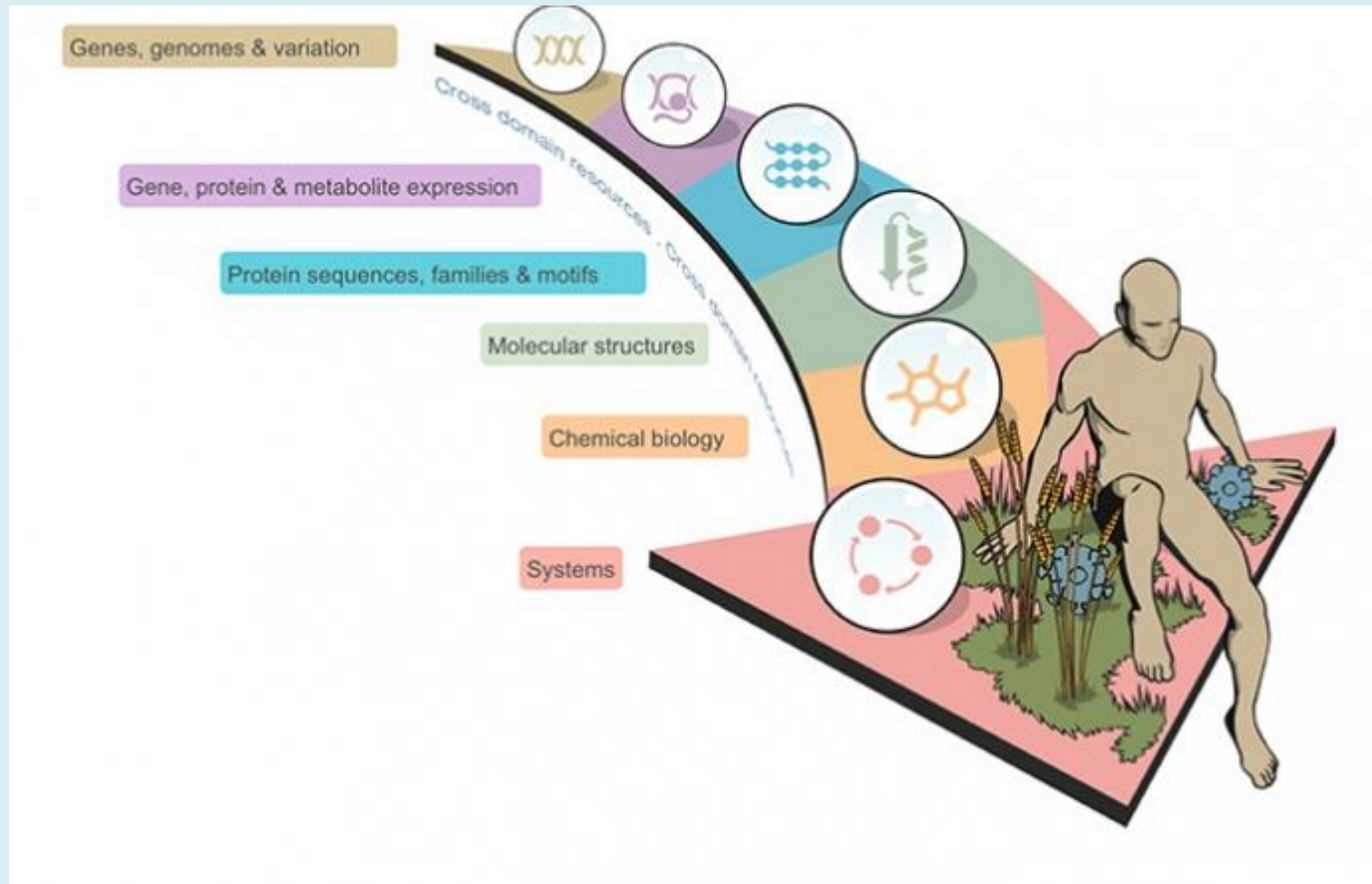
- Tıp Doktoru (2004)
- Biyokimya Uzmanı (2009)
- Bilgisayar Mühendisi (2011)
-  kullanıcısı (2014 - ...)
- Tıbbi Biyokimya Doktora (2018)



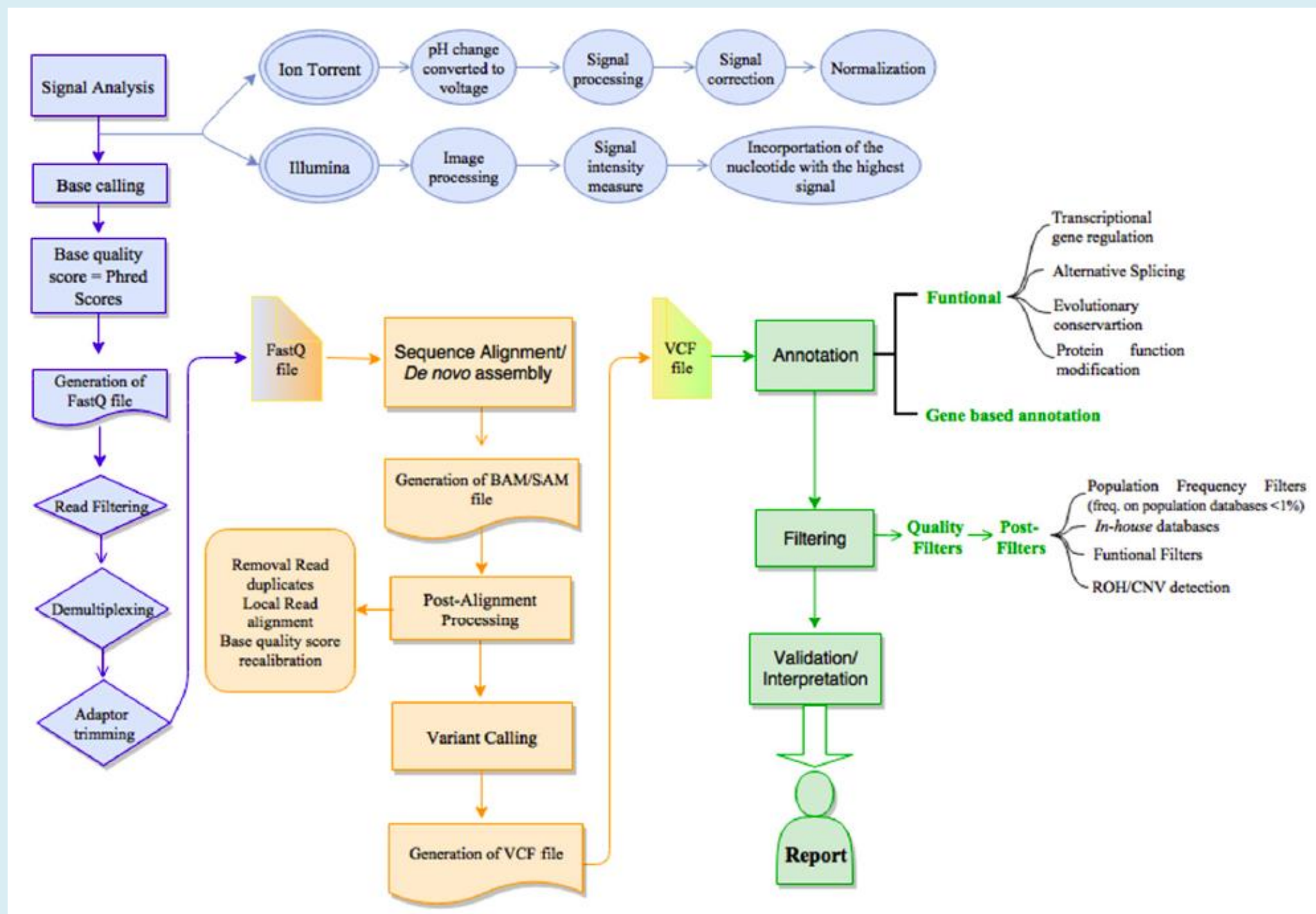
Öğrenme Hedefleri

- Genomik veri tabanlarında bulunan farklı veri tipleri ve bunların biyoinformatikteki rolünü anlamak.
- Genomik veri tabanlarının temel işlevlerini ve türlerini öğrenmek.
- Transplantasyon immünolojisi kapsamında önemli genomik veri tabanlarını tanıyıp bu veri tabanlarından nasıl yararlanılacağını öğrenmek.
- Web tarayıcı üzerinden ve Bioconductor paketleri aracılığıyla genomik veri tabanlarından veri çekme becerilerini edinmek.

Biyoinformatik



Genomik Veri Türleri



Genomik Veri Türleri

Sequence Data

DNA

RNA

Protein

Annotation

Nucleotide-level annotation

Protein-level annotation

Process-level annotation

Sekans Verisi için «FASTA» Dosyası

- FASTA dosya formatı, biyolojik sekans verilerini saklamak için en popüler formatlardan biridir.
- Bu metin tabanlı dosyalar,
 - Amino asit dizilerini (peptitler)
 - nükleotid dizilerini (DNA veya RNA) depolamak için kullanılabilir.
- Rutin olarak dizi açıklaması, veritabanı aramaları ve çoklu dizi hizalaması için kullanılırlar.
- İnsanlar ve bilgisayarlar tarafından anlaşılabilir
 - IUB/IUPAC amino asit kodlarını ve nükleik asit kodlarını kullanır
- “.fasta”, “.fa”, “.ffn”, “.frn”, “.fna”, and “.faa”
- «>» ile başlar.

FASTA Dosyası Örneği

DNA

```
>NC_012920.1 Homo sapiens mitochondrion, complete genome
GATCACAGGTCTATCACCTATTAACCACTCACGGGAGCTCTCCATGCATT
TGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGG
AGCCGGAGCACCTATGTCGCAGTATCTGTCTTTGATTCCTGCCTCATCCT
ATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAG
TGTGTTA...
```

RNA

```
>Mus_musculus_tRNA-Ala-AGC-1-1 Ala (AGC)
GGGGGUGUAGCUCAGUGGUAGAGCGCGUGCUUAGCAUGCACGAGGCCCUUGGU
UCGAUCCCCAGCACCUCCA
```

Protein

```
>sp|P01210|PENK_HUMAN Proenkephalin-A OS=Homo sapiens OX=9606
GN=PENK
MARFLTLCTWLLLLGPGLLATVRAECSQDCATCSYRLVRPADINFLACVMECEG
KLPSLKIWETCKELLQLSKPELPQDGTSTLRENSKPEESHLLAKRYGGFMKRY
GGFMKKMDELYPMEPEEEANGSEILAKRYGGFMKKDAEEDDSLANSDDLKE
LLETGDNRERSHHQDGS DNEE
```

FASTQ Dosya Türü

- FASTQ dosyaları FASTA'ya benzer ancak sekans verilerinin kalite skorunu da içerir.
- İki dosya arasındaki temel fark, FASTQ formatının ham sekanslama bilgilerini, özellikle de baz çağrılarıyla ilgili kalite puanlarını içermesidir.
- «@» ile başlar.

FASTQ Dosyası Örneği

```
1 @ERR000589.41 EAS139_45:5:1:2:111/1
2 CTTTCCTCCCTGCTTTCCTGGCCCCACCATTTCCAGGGAACATCTTGTCAT
3 +
4 3IIIIIIIIIIII>1IIIFF9BG08E00I%IG+&?(4)%00646.C1#&(
5 @ERR000589.42 EAS139_45:5:1:2:1293/1
6 AGTTGTTAAAATCCAAGCCAATTAAGATAGTCTTATCTTTTAAAAGAAAT
7 +
8 IIIIIGII.AIIII=?I9G-/II=+I=4?761BA2C9I+5A711+&>1$/I
```

Anotasyon (Annotation)

GENOME ANNOTATION: FROM SEQUENCE TO BIOLOGY

Lincoln Stein

The genome sequence of an organism is an information resource unlike any that biologists have previously had access to. But the value of the genome is only as good as its annotation. It is the annotation that bridges the gap from the sequence to the biology of the organism.

The aim of high-quality annotation is to identify the key features of the genome — in particular, the genes and their products. The tools and resources for annotation are developing rapidly, and the scientific community is becoming increasingly reliant on this information for all aspects of biological research.

Anotasyon (Annotation)

- Gene locations
- Genetic variation
- Transcription factor binding sites
- Structural features of proteins

Anotasyon (Annotation)

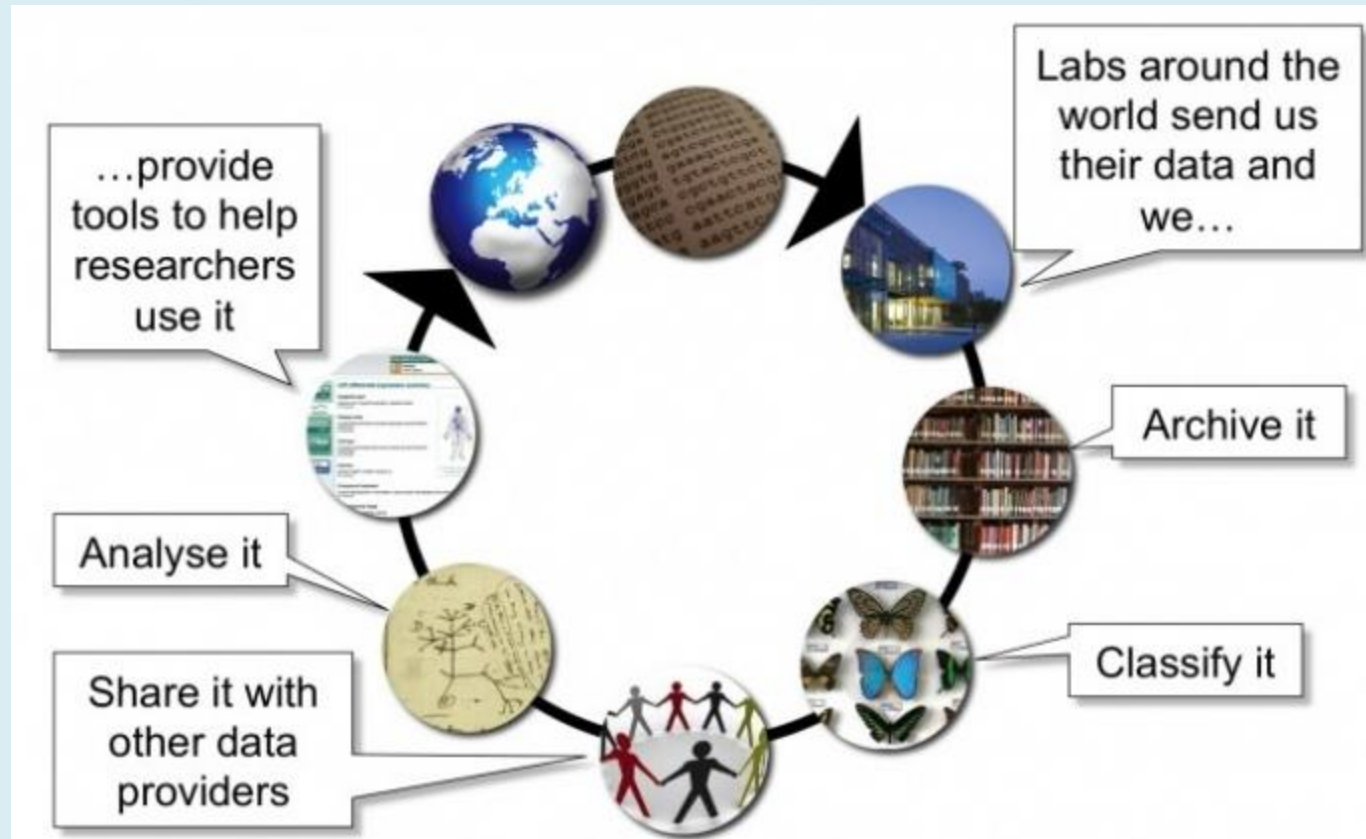
- Tahmin edilen bir genin ürününün işlevinin tanımlanmasından oluşur (in-silico yaklaşımla).
- Bu tanımlamalar çeşitli biyoinformatik yazılımları aracılığıyla gerçekleştirilebilir:
 - (1) **Sinyal analizi** (TATA box, start and stop kodonları, or poly-A sinyal tespiti)
 - (2) **İçerik analizi** (for G 1 C content, codon usage, or dicodon frequency detection),
 - (3) **Benzerlik tespiti** (benzer organizmalardaki proteinler, aynı organizmadaki mRNA referans genomları)

Anotasyon Örneği

Annotation					
Chromosome	Start coordinates	End coordinates	Name (RefSeq ID)	Score	Strand
chr 1	9969621	10009136	NM_145692.2	0	-
chr 1	10024599	10038159	NM_013715.2	0	-
chr 1	10038217	10136768	NM_026493.3	0	+
chr 1	10046492	10136768	NM_001368816.1	0	+
chr 1	10137506	10232670	NM_001102430.1	0	-

Genomik Veri Tabanları

Genomik Veri Tabanları



Genomik Veri Tabanlarının Kullanımı

- Bilgi deposu görevi görür.
 - Genomik verinin organize edilmesi ve düzenlenmesini sağlar.
- Araştırmacıların mevcut verileri incelemesine olanak sağlar.
- Veri türleri, bireyler ve organizmalar arasında verilerin paylaşılmasına ve karşılaştırılmasına olanak tanır.

EMBL's European

EMBL

Unleashing

Find a gene

Example search

Find data



National Library of Medicine

National Center for Biotechnology Information

Log in

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine



Research Organization of Information and Systems

National Institute of Genetics

ACCESS

JP

EN

About NIG

Research

PhD Program

Research
Infrastructure
& Collaboration

Bioresources

Outreach

Career
Development



Research Organization of Information and Systems

National Institute of Genetics

Primer ve Sekonder Veri Tabanları

	Primary database	Secondary database
Synonyms	Archival database	Curated database; knowledgebase
Source of data	Direct submission of experimentally-derived data from researchers	Results of analysis, literature research and interpretation, often of data in primary databases
Examples	ENA, GenBank and DDBJ (nucleotide sequence) ArrayExpress and GEO (functional genomics data) Protein Data Bank (PDB; coordinates of three-dimensional macromolecular structures)	InterPro (protein families, motifs and domains) UniProt Knowledgebase (sequence and functional information on proteins) Ensembl (variation, function, regulation and more layered onto whole genome sequences)

Primer Veri Tabanları Örnekleri

- Genbank sequence database, NCBI
- Ensembl genome database, EMBL-EBI
- UCSC Genome Browser Database

NCBI: NCBI stands for the National Center for Biotechnology Information

EMBL-EBI: European Molecular Biology Laboratory's European Bioinformatics Institute

UCSC: University of California, Santa Cruz



All Databases ▾

All Databases
Assembly
Biocollections
BioProject
BioSample
Books
ClinVar
Conserved Domains
dbGaP
dbVar
Gene
Genome
GEO DataSets
GEO Profiles
GTR
Identical Protein Groups
MedGen
MeSH
NLM Catalog
Nucleotide

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

NCBI

Center for Biotechnology Information advances science and health by providing access to genomic information.

[Home](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Submit manuscripts
and databases

Download

Transfer NCBI data to your
computer

Learn

Find help documents, attend a
class or watch a tutorial

Develop

Use NCBI APIs and code
libraries to build applications

Analyze

Identify an NCBI tool for your
data analysis task

Research

Explore NCBI research and
collaborative projects

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

MedGen Users, We Want Your
Feedback!

28 Mar 2024

Do you use NCBI's MedGen? If so, then
you probably know it's NCBI's one-stop-

Changes to SRA Data Access on the
Google Cloud Platform (GCP)

20 Mar 2024

Sequence Read Archive (SRA) data
available via the Google Cloud Platform

Preview Upcoming Improvements to
PubMed Central (PMC)

14 Mar 2024

Updated article display and journal list
Since announcing the launch of a more

-> PubMed <-

Ensembl Genome Database

 [BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [More ▾](#)

Login/Register

 Search all species... 

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 111 (January 2024)

- MANE Select (v1.2) GRCh38.p14 patch annotation
- Human variation data updated to dbSNP156
- Updated genome assembly and annotation for sheep and cattle
- Regulatory annotation of open chromatin regions and promoters in common carp and rainbow trout (a collaboration with the

Search

All species ▾

for

Go

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

UCSC Genome Browser Database

UNIVERSITY OF CALIFORNIA
SANTA CRUZ Genomics Institute

UCSC **Genome Browser**

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About



Search genes, data, help docs and more... Search

Tools

hg38 hg19 mm39

- **Genome Browser** - Interactively visualize genomic data
- **BLAT** - Rapidly align sequences to the genome
- **In-Silico PCR** - Rapidly align PCR primer pairs to the genome
- **Table Browser** - Download and filter data from the Genome Browser
- **LiftOver** - Convert genome coordinates between assemblies
- **REST API** - Returns data requested in JSON format
- **Variant Annotation Integrator** - Annotate genomic variants
- **More tools...**

Sekonder Veri Tabanı Örnekleri

- clinVar

<https://www.ncbi.nlm.nih.gov/clinvar/>

- COSMIC (Catalogue Of Somatic Mutations In Cancer)

- <https://cancer.sanger.ac.uk/cosmic>

- CFTR2

<https://cftr2.org/>

- Human Hemoglobin Variants and Thalassemias (HbVar)

<https://globin.bx.psu.edu/hbvar/>

Hibrid (Composite) Veri Tabanları

- Hem primer hem de sekonder genomik verileri bir araya getiren bir veritabanı türüdür.
- Ham DNA dizilimleri, RNA dizilimleri gibi doğrudan elde edilen verileri ve bu veriler üzerinde yapılan analizler sonucu elde edilen gen anotasyonları, protein etkileşimleri ve fonksiyonel bilgiler gibi işlenmiş verileri birlikte sunar
 - The Cancer Genome Atlas (TCGA)

[Home](#) > [Research](#) > [Genome Sequencing](#) > The Cancer Genome Atlas Program (TCGA)

TCGA

Program History >

TCGA Cancers Selected for Study

Publications by TCGA

The Cancer Genome Atlas Program (TCGA)

The Cancer Genome Atlas (TCGA), a landmark cancer genomics program, molecularly matched normal samples spanning 33 cancer types. This joint effort between NCI and the National Cancer Institute began in 2006, bringing together researchers from diverse disciplines and

Spesifik Veri Tabanları

- IPD-IMGT/HLA veri tabanı
- dbMHC
- Immune Epitope Database (IEDB)
- EPIMHC

IPD-IMGT/HLA Veri Tabanı



IPD-IMGT/HLA

[Overview](#)[IMGT/HLA](#)[KIR](#)[MHC](#)[NHKIR](#)[HPA](#)[ESTDAB](#)

IPD / IMGT/HLA

Welcome to IPD-IMGT/HLA

Release 3.55 (2024-01) Version Report - Build 9768104

The IPD-IMGT/HLA Database provides a specialist database for sequences of the human major histocompatibility complex (MHC) and includes the official sequences named by the WHO Nomenclature Committee For Factors of the HLA System. The IPD-IMGT/HLA Database was originally part of the international ImMunoGeneTics project (IMGT). For more information about the database and what data and tools are available please see our [i about](#) page.

IPD-IMGT/HLA Veri Tabanı

- MHC ile ilişkili HLA genlerinin ayrıntılı genetik dizilerini içerir.
- Bu veri tabanı, transplantasyon immünolojisi ve otoimmün hastalıklar gibi alanlarda önemli olan HLA gen varyasyonlarını araştırmak için kullanılır.
- Dünya genelindeki popülasyonlarda HLA alel frekanslarını karşılaştırma imkanı sunar.
- İmmünoloji ve kişiselleştirilmiş tıp çalışmalarında referans bir kaynak olarak kullanılabilir.

IPD-IMGT/HLA Veri Tabanı



IPD-IMGT/HLA

[Overview](#)[IMGT/HLA](#)[KIR](#)[MHC](#)[NHKIR](#)[HPA](#)[ESTDAB](#)[IPD](#) / [IMGT/HLA](#) / [ALLELE LIST](#)

Allele Query Tool

The IPD-IMGT/HLA Allele Query Tool provides advanced query interface allowing users to generate customizable queries to the IPD-IMGT/HLA database. This is available via a web interface or a POST API. Please see our [? help pages](#) for documentation on the use of this tool. The default output displays all alleles and can be searched by typing a query in the box provided. More advanced searches can be performed by clicking the Advanced tab. The fields displayed in the output can also be changed by clicking the Fields tab.

dbMHC

- MHC ile ilgili DNA ve klinik verilerine açık erişim sağlayan bir platformdur.
- Bu veritabanı, dünya genelinde 90'dan fazla popülasyonda HLA alelleri ve haplotipleri gibi verilerin yanı sıra, kök hücre transplantasyonu gibi klinik veri setleri sunmaktadır.

dbMHC

NCBI Resources How To

 We are sorry, but the page you requested is

dbMHC and IHWG data

The dbMHC database was an open, publicly accessible p
Complex (MHC). Data from IHWG workshops were provid

Data previously on the site are now available at <ftp://ftp.ncbi.nlm.nih.gov/pub/mhc/mhc/Final Archive>

If you have any specific questions, **Index of /pub/mhc/mhc/Final Archive**

Name	Last modified	Size
Parent Directory		-
CN3D/	2018-02-02 13:18	-
Database/	2018-01-11 11:41	-
IHWG/	2018-01-11 11:48	-
Sequence Viewer/	2018-01-09 12:08	-
alleles/	2018-01-09 12:09	-
dbMHC_FINAL_kit.xml.gz	2018-01-11 11:42	287K
dbMHC_FINAL_tube.xml.gz	2018-01-11 11:42	241K
dbMHC_Sequence_Alignment_Viewer.tar.gz	2018-01-11 11:42	4.0M

[HHS Vulnerability Disclosure](#)

01000100011000010111010001100001
IMMUNOGENOMICS DATA
ANALYSIS WORKING GROUP
GCTAATGCACTGTACAGTATATCA

The immunogenomics data analysis working group (IDAWG) is an international collaboration of [histocompatibility and immunogenetics investigators](#) who share the goal of facilitating the sharing of immunogenomic data (HLA, KIR, etc.) and fostering the consistent analysis and interpretation of those data by the immunogenomics community and the larger genomics communities.

[The International HLA and Immunogenetics](#)
present its recommendations on topics of
W and Conference in 2012. IDAWG projects

Immune Epitope Database (IEDB)

- Immune Epitope Database (IEDB), enfeksiyon hastalıkları, alerjiler, otoimmün hastalıklar ve organ nakli gibi durumlarda antikor ve T hücresi epitoplarının kapsamlı verilerini içerir.
- İmmün tepkilerin ve aşı geliştirme süreçlerinin daha iyi anlaşılmasına yardımcı olacak epitop tahmin araçları sağlar.
- Veritabanı, epitopların yapısını, MHC ligand deneylerini ve T hücresi deney ayrıntılarını içerir.

Immune Epitope Database (IEDB)

**IEDB**
Immune Epitope Database & Tools

Help | More IEDB

Home | Specialized Searches | Analysis Resource

Check out our new IEDB updates! (1) Learn how to [customize your database exports](#) and (2) test out the new [Next-generation Tools site](#) for all your analysis and prediction needs.

Welcome

The Immune Epitope Database (IEDB) is a freely available resource funded by NIAID. It catalogs experimental data on antibody and T cell epitopes studied in humans and other animal species in the context of infectious disease, allergy, autoimmunity and transplantation. The IEDB also hosts epitope prediction and analysis tools, and has a companion site, [CEDAR](#) (funded by NCI), which houses cancer epitopes.

[Learn More](#)

Upcoming Events & News

Virtual User Workshop	Nov 1-3, 2023
* recordings here	
AACR 2024	Apr 5-10, 2024
Festival of Biologics	Apr 15-17, 2024
* free pass registration	
AAI 2024	May 3-7, 2024

Summary Metrics

Peptidic Epitopes	1,612,323
Non-Peptidic Epitopes	3,188
T Cell Assays	511,823
B Cell Assays	1,403,605
MHC Ligand Assays	4,802,904
Epitope Source Organisms	4,449
Restricting MHC Alleles	1,005
References	24,676

START YOUR SEARCH HERE

Epitope ?

☒ Any
☐ Linear peptide
Exact M Ex: SIINFEKL
☐ Discontinuous
☐ Non-peptidic

Assay ?

☒ T Cell
☒ B Cell
☒ MHC Ligand
Ex: neutralization
Outcome: ☒ Positive ☐ Negative

Epitope Source ?

Organism
Ex: influenza, peanut
Antigen
Ex: core, capsid, myosin

MHC Restriction ?

☒ Any
☐ Class I
☐ Class II
☐ Non-classical
Ex: HLA-A*02:01

Host ?

☒ Any
☐ Human
☐ Mouse
☐ Non-human primate
Ex: dog, camel

Disease ?

☒ Any
☐ Infectious
☐ Allergic
☐ Autoimmune
Ex: asthma

Epitope Analysis Resource

T Cell Epitope Prediction ?

Scan an antigen sequence for amino acid patterns indicative of:

- [MHC I Binding](#)
- [MHC II Binding](#)
- [MHC I Processing \(Proteasome, TAP\)](#)
- [MHC I Immunogenicity](#)

B Cell Epitope Prediction ?

Predict linear B cell epitopes using:

- [Antigen Sequence Properties](#)

Predict discontinuous B cell epitopes using antigen structure via:

- [Discotope](#)
- [Ellipro](#)

Epitope Analysis Tools ?

Analyze epitope sets of:

- [Population Coverage](#)
- [Conservation Across Antigens](#)
- [Clusters with Similar Sequences](#)

www.iedb.org ; www.niaid.nih.gov

EPIMHC

- EPIMHC, proteinlerde gözlemlenen MHC-binding peptidler ve T hücreleri epitoplarını içeren ilişkisel bir veri tabanıdır.
- Veri tabanı, 84 tümör ilişkili antijen dahil olmak üzere çeşitli kaynaklardan 4867 farklı peptid sekansına ait bilgi içerir.
- Aşı geliştirme ve özel immünolojik araştırmalar için çeşitli araçlar sağlar.

EPIMHC



UNIVERSIDAD
COMPLUTENSE
MADRID



Immunomedicine Group

[Home](#) [Research](#) [Services](#) [Technology](#) [Contact](#)

TOOLS

- Alignment Analysis
- Databases
- Computational Immunology
- Modelling & 3D-structure Analysis
- Sequence Manipulation & Analysis
- Similarity Searches
- View all Tools

Tools >> EPIMHC Database



EPIMHC

A Database of Naturally Processed MHC-restricted Peptide Ligands and Epitopes for Customized Computational Vaccinology

MHC SELECTION

- All
- BOLA-A11
- DR52
- ELA-A5
- ELA-A9
- H2-AB
- H2-AD
- H2-AG7
- H2-AK
- H2-AS

AND ▼

SEQ ⓘ

AND ▼

LENGTH ⓘ

All ▲678▼

AND ▼

CLASS ⓘ All ▼

AND ▼

MHC SOURCE ⓘ

All ▲BONOBO CHIMPANZEECHIMPANZEECOTTON-TOP TAMARIN ▼

Diğer Veri Tabanları



Allele Frequency Net Database

[About us](#) [Links](#) [Publications](#) [Automated Access](#) [FAQs](#) [Contact](#)

[Home](#) [Populations](#) [HLA](#) [KIR](#) [Other polymorphisms](#) [HLA-ADR](#) [KDDB](#) [EUROSTAM](#)

News

www.hlacovid19.org website

Our friends (who are always very supportive of the AFND website, Jill A. Hollenbach @UCSF, Steven J. Mack @UCSF and Martin Maier @NMDP) have formed the COVID-19 HLA & Immunogenetics Consortium to unite the global community of HLA and immunogenetics experts and leaders in support of these efforts. They invite you to share your expertise as they launch this endeavour by joining the consortium. A project for the next international Histocompatibility Workshop is also planned.

Introduction

The **Allele Frequency Net Database** (AFND) provides the scientific community with a freely available repository for the storage of immune gene frequencies in different worldwide populations. Users can contribute the results of their work into one common database and can perform database searches on information already available. We have currently collected data in [allele](#), [haplotype](#) and [genotype](#) format. However, the success of this website will depend on you to contribute your data.

Login

You are logged as: guest
[Click here to use your account](#)

SUBMIT YOUR DATA

- ☐ Add HLA genotype raw data for SPR
- ☐ Add HLA frequency data
- ☐ Add non-HLA frequency data

Sponsors

- ☐ BAG Health Care 
- ☐ Fujirebio Europe 
- ☐ CareDx 

Genomik Veri Tabanlarına Nasıl Ulaşılır?

Genomik Veri Tabanlarına Nasıl Ulaşılır?

- Web Browser
- REST API
- Bioconductor

Web Browser ile Ulaşım Örneği

- <https://iastate.pressbooks.pub/molecularplantbreeding/chapter/introduction-to-bioinformatics/>

REST API

Ensembl REST API Endpoints

Archive

Resource	Description
GET archive/id/:id	Uses the given identifier to return its latest version
POST archive/id	Retrieve the latest version for a set of identifiers

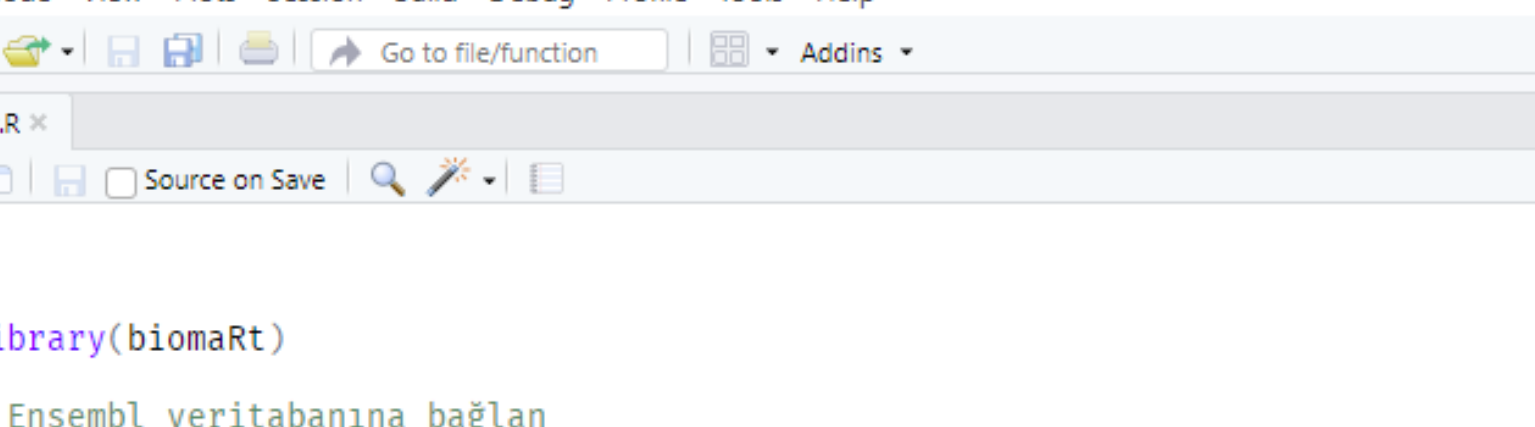
Comparative Genomics

Resource	Description
GET cafe/genetree/id/:id	Retrieves a cafe tree of the gene tree using the gene tree stable identifier
GET cafe/genetree/member/id/:id	Retrieves the cafe tree of the gene tree that contains the gene / transcript / translation stable identifier. DEPRECATED: use cafe/genetree/member/id/:species/:id instead.
GET cafe/genetree/member/symbol/:species/:symbol	Retrieves the cafe tree of the gene tree that contains the gene identified by a symbol
GET cafe/genetree/member/id/:species/:id	Retrieves the cafe tree of the gene tree that contains the gene / transcript / translation stable identifier in the given species
GET genetree/id/:id	Retrieves a gene tree for a gene tree stable identifier
GET genetree/member/id/:id	Retrieves the gene tree that contains the gene / transcript / translation stable identifier. DEPRECATED: use genetree/member/id/:species/:id instead.

Bioconductor

- Veri tabanlarına erişim için Bioconductor kütüphaneleri
 - **BiomaRt** Ensembl and other databases
 - **EnsemblDb** Ensembl database
 - **GenomicFeatures** UCSC Genome Browser
 - **ipdDb** PD-IMGT/HLA Database
 - **GEOquery** Gene Expression Omnibus (GEO)
 - **TxDb.Hsapiens.UCSC.hg19..**
 - **AnnotationHub**
 - **AnnotationDbi**

Bioconductor Örneği: BiomaRt



The screenshot shows the RStudio interface with a script editor open. The script is titled 'biomart.R' and contains the following R code:

```
1  
2  
3  
4 library(biomaRt)  
5  
6 # Ensembl veritabanına bağlan  
7 ensembl ← useMart("ensembl", dataset = "hsapiens_gene_ensembl")  
8  
9  
10 hla_genes ← getBM(attributes = c('ensembl_gene_id', 'external_gene_name',  
11                                'chromosome_name', 'start_position', 'end_position'),  
12                    filters = 'external_gene_name',  
13                    values = c('HLA-A', 'HLA-B', 'HLA-C'),  
14                    mart = ensembl)  
15  
16 head(hla_genes)  
17
```

Bioconductor Örneği: BiomaRt

```
>
> # Ensembl veritabanına bağlan
> ensembl ← useMart("ensembl", dataset = "hsapiens_gene_ensembl")
> hla_genes ← getBM(attributes = c('ensembl_gene_id', 'external_gene_name',
+                                 'chromosome_name', 'start_position', 'end_position'),
+                   filters = 'external_gene_name',
+                   values = c('HLA-A', 'HLA-B', 'HLA-C'),
+                   mart = ensembl)
> head(hla_genes)
```

	ensembl_gene_id	external_gene_name	chromosome_name	start_position	end_position
1	ENSG00000231834	HLA-A	HSCHR6_MHC_MANN_CTG1	1195801	1200442
2	ENSG00000224320	HLA-A	HSCHR6_MHC_COX_CTG1	1369022	1489345
3	ENSG00000235657	HLA-A	HSCHR6_MHC_DBB_CTG1	1144955	1265390
4	ENSG00000224608	HLA-B	HSCHR6_MHC_MANN_CTG1	2578541	2665857
5	ENSG00000237022	HLA-C	HSCHR6_MHC_MANN_CTG1	2577801	2581172
6	ENSG00000228299	HLA-C	HSCHR6_MHC_DBB_CTG1	2526549	2529920

```
> |
```

Bioconductor Örneği: ipdDb

```
hla - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ + + + + Go to file/function + Addins
biomart.R* x
Source on Save
19
20 library(ipdDb)
21
22 hla <- loadHlaData()
23 ## get all loci stored in the db
24 available_loci <- hla$getLoci()
25
26 ## get all alleles of a locus
27 alleles <- hla$getAlleles(available_loci[1])
28 alleles <- hla$getAlleles("HLA-A")
29
30 ## get all sequences of a bunch of alleles as DNASTringSet
31 sequences <- hla$getReference(alleles)
32 sequences <- hla$getReference(c("HLA-A*01:01:01:01", "HLA-A*01:01:01:03" ))
33
34 ## get the closest complete reference for ONE allele as DNASTringSet
35 closest_complete <- hla$getClosestComplete(alleles[1])
36 closest_complete <- hla$getClosestComplete("HLA-A*01:01:01:01")
37
38 ## Get the gene structure for a bunch of alleles as GRanges object
39 structures <- hla$getStructure(alleles)
40 structures <- hla$getStructure(c("HLA-A*01:01:01:01", "HLA-A*01:01:01:03" ))
```

Bioconductor Örneği: ipdDb

```
> closest_complete
DNAStringSet object of length 1:
  width seq
[1] 3503 CAGGAGCAGAGGGGTCAGGGCGAAGTCCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGC ... TCTCCATCTCTGTCTCAACTTCATGGTGCAGCTGAGCTGTAACTTCTTCCTTCCCTATTAAAA names
HLA-A*01:01:01:01
> closest_complete
DNAStringSet object of length 1:
  width seq
[1] 3503 CAGGAGCAGAGGGGTCAGGGCGAAGTCCCAGGGCCCCAGGCGTGGCTCTCAGGGTCTCAGGC ... TCTCCATCTCTGTCTCAACTTCATGGTGCAGCTGAGCTGTAACTTCTTCCTTCCCTATTAAAA names
HLA-A*01:01:01:01
> structures
GRanges object with 34 ranges and 4 metadata columns:
      seqnames      ranges strand |      type      name complete      frame
      <Rle> <IRanges> <Rle> | <character> <character> <logical> <integer>
HLA-A*01:01:01:01 HLA-A*01:01:01:01 1-300      + |      UTR      5' UTR      TRUE        0
HLA-A*01:01:01:01 HLA-A*01:01:01:01 301-373     + |      Exon      Exon 1      TRUE        1
HLA-A*01:01:01:01 HLA-A*01:01:01:01 374-503     + |      Intron     Intron 1    TRUE        0
HLA-A*01:01:01:01 HLA-A*01:01:01:01 504-773     + |      Exon      Exon 2      TRUE        3
HLA-A*01:01:01:01 HLA-A*01:01:01:01 774-1014    + |      Intron     Intron 2    TRUE        0
...
HLA-A*01:01:01:03 HLA-A*01:01:01:03 2840-2981   + |      Intron     Intron 6    TRUE        0
HLA-A*01:01:01:03 HLA-A*01:01:01:03 2982-3029   + |      Exon      Exon 7      TRUE        3
HLA-A*01:01:01:03 HLA-A*01:01:01:03 3030-3198   + |      Intron     Intron 7    TRUE        0
HLA-A*01:01:01:03 HLA-A*01:01:01:03 3199-3203   + |      Exon      Exon 8      TRUE        3
HLA-A*01:01:01:03 HLA-A*01:01:01:03 3204-3300   + |      UTR      3' UTR      TRUE        0
seqinfo: 2 sequences from an unspecified genome; no seqlengths
```

Diğer İlişkili R Kütüphaneleri

- BIGDAWG
- HLAminer
- immunoClust
- ComplexHeatmap

Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG)

BIGDAWG

Derek Pappas, Ph.D. (dpappas@chori.org)

2021-10-23

Overview

'Bridging ImmunoGenomic Data-Analysis Workflow Gaps' ('BIGDAWG') is an integrated analysis system that automates the manual data-manipulation and trafficking steps (the gaps in an analysis workflow) normally required for analyses of highly polymorphic genetic systems (e.g., the immunological human leukocyte antigen (HLA) and killer-cell Immunoglobulin-like receptor (KIR) genes) and their respective genomic data (immunogenomic) (Pappas DJ, Marin W, Hollenbach JA, Mack SJ. 2016. 'Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline.' [Human Immunology. 77:283-287](#)). Starting with unambiguous genotype data for case-control groups, 'BIGDAWG' performs tests of Hardy-Weinberg equilibrium, and carries out case-control association analyses for haplotypes, individual loci, specific HLA exons, and HLA amino acid positions.

<https://cran.r-project.org/web/packages/BIGDAWG/vignettes/BIGDAWG.html>

HLAminer

HLAminer



HLAminer is a software for HLA predictions from next-generation shotgun (NGS) sequence read data and supports direct read alignment and targeted *de novo* assembly of sequence reads.

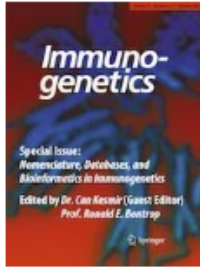
It has been used **successfully** to predict expressed HLA genes in the large TCGA (The Cancer Genome Atlas) cancer cohort.

HLAminer uses **TASR** as its main assembly engine, which in turn is derived from **SSAKE**, the first *de novo* genome assembler published. **A manuscript describing the methodology is published in the peer-reviewed journal *Genome Medicine*.**

Sonuç ve Öneriler

- Genomik veri tabanları, transplantasyon immünolojisi araştırmalarında kritik öneme sahiptir.
- Buradaki bilgiler uygun donör eşleşmelerini ve greft reddi mekanizmalarını anlamada kullanılabilir.
- Veri tabanlarına ulaşmada günümüzdeki bir çok farklı araç bulunmaktadır.
- Bioconductor gibi araçlar, özel veri analizi için bilimsel çalışmaları desteklemektedir.

Kaynaklar - 1



Volume 72, Issue 1-2

February 2020

Special Issue: Nomenclature, Databases, and Bioinformatics in Immunogenetics

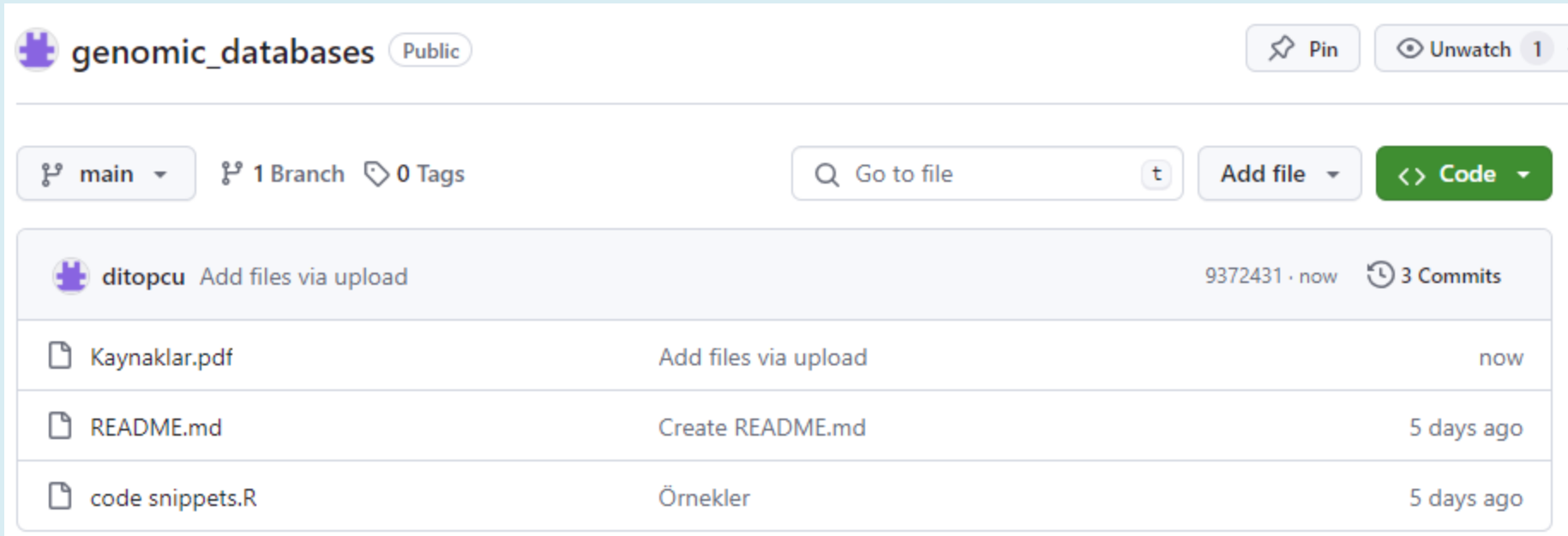
Issue Editors: Can Kesmir, Ronald E. Bontrop

Review

Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics

Rute Pereira ^{1,2,*} , Jorge Oliveira ^{2,3,†}  and Mário Sousa ^{1,2,†}

Kaynaklar - 2



The screenshot shows the GitHub interface for the repository 'genomic_databases' by user 'ditopcu'. The repository is public and has 1 branch (main) and 0 tags. The file list includes 'Kaynaklar.pdf', 'README.md', and 'code snippets.R'. The 'code snippets.R' file has a commit history of 3 commits, with the latest commit made 5 days ago.

genomic_databases Public

Pin Unwatch 1

main 1 Branch 0 Tags

Go to file t Add file <> Code

ditopcu Add files via upload 9372431 · now 3 Commits

Kaynaklar.pdf	Add files via upload	now
README.md	Create README.md	5 days ago
code snippets.R	Örnekler	5 days ago

https://github.com/ditopcu/genomic_databases

Kaynaklar - 2



https://github.com/ditopcu/genomic_databases

Teşekkürler !

Soru ?

Katkı +

