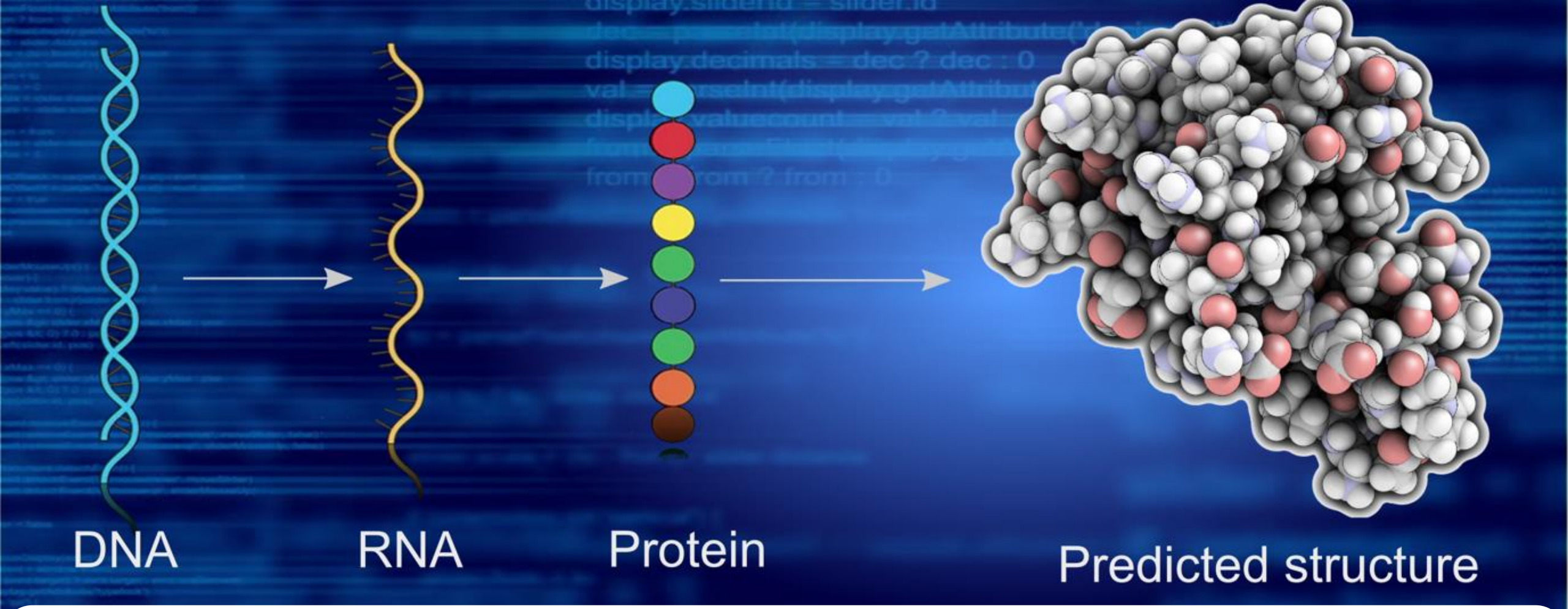


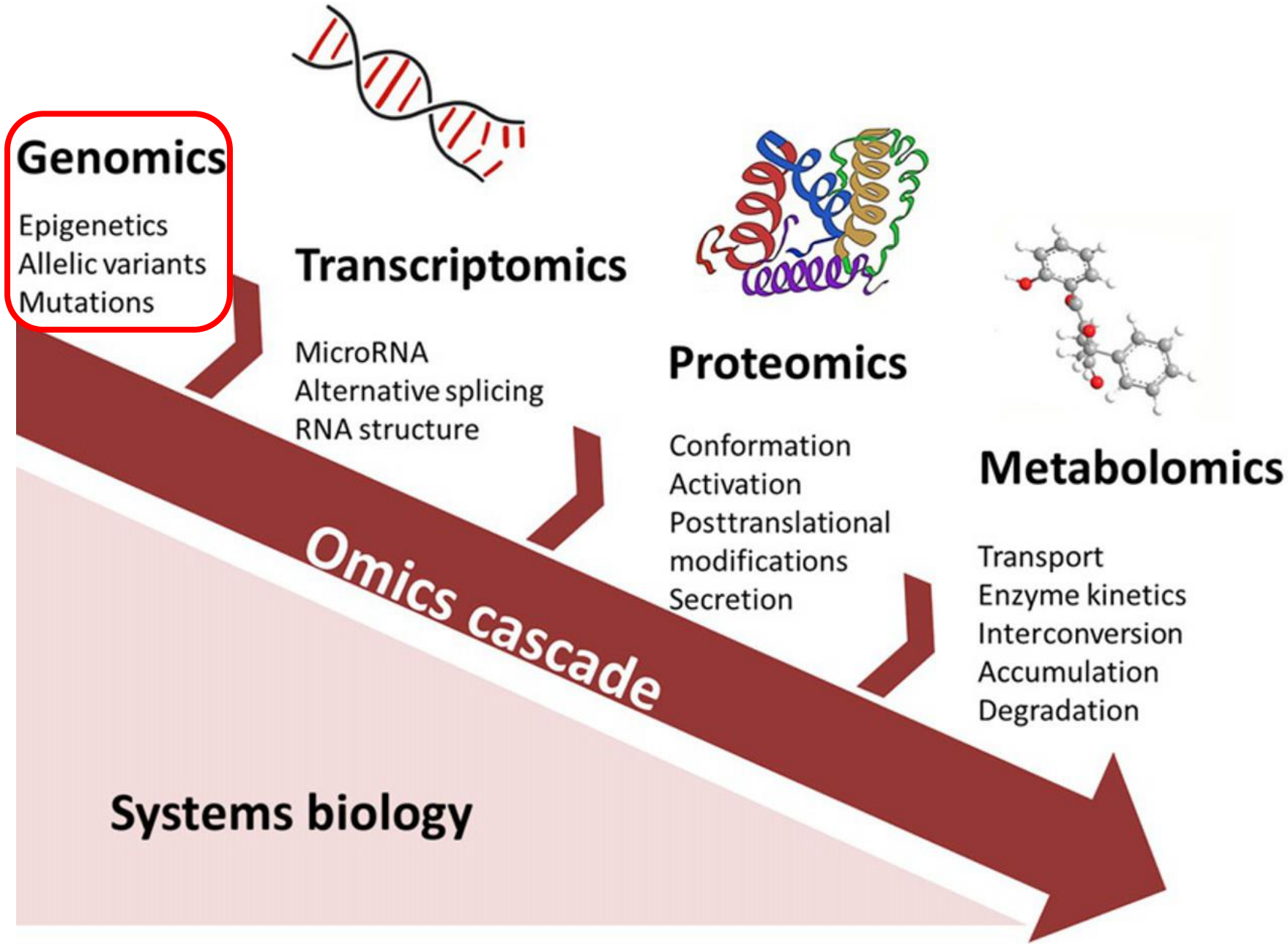
Biyoinformatikte Güncel Konular

Can Koşukcu, M.Sc.

Senior Bioinformatics Application Scientist

QIAGEN

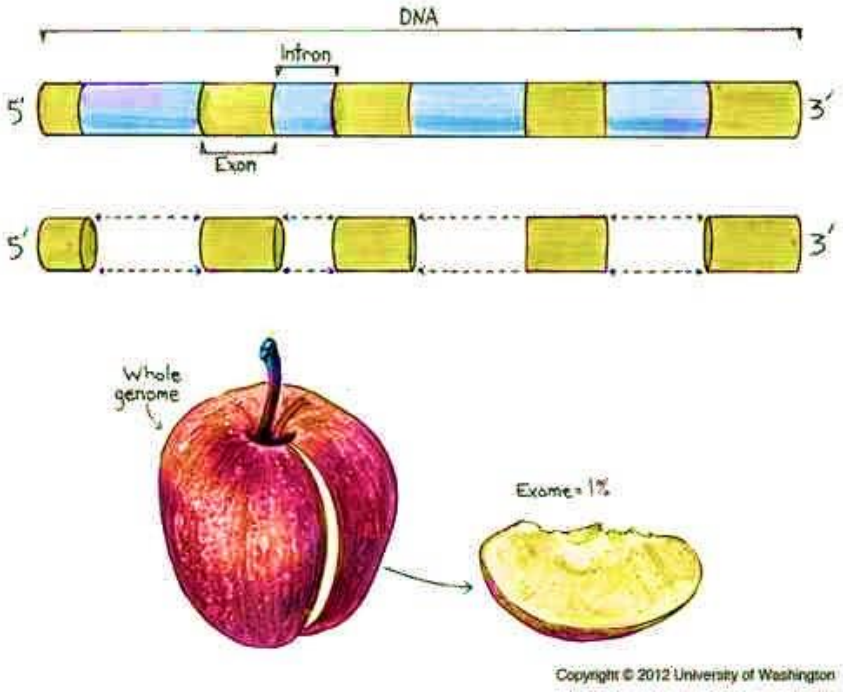
9. Ulusal Transplantasyon İmmünolojisi ve Genetiği Kongresi
18.04.2024

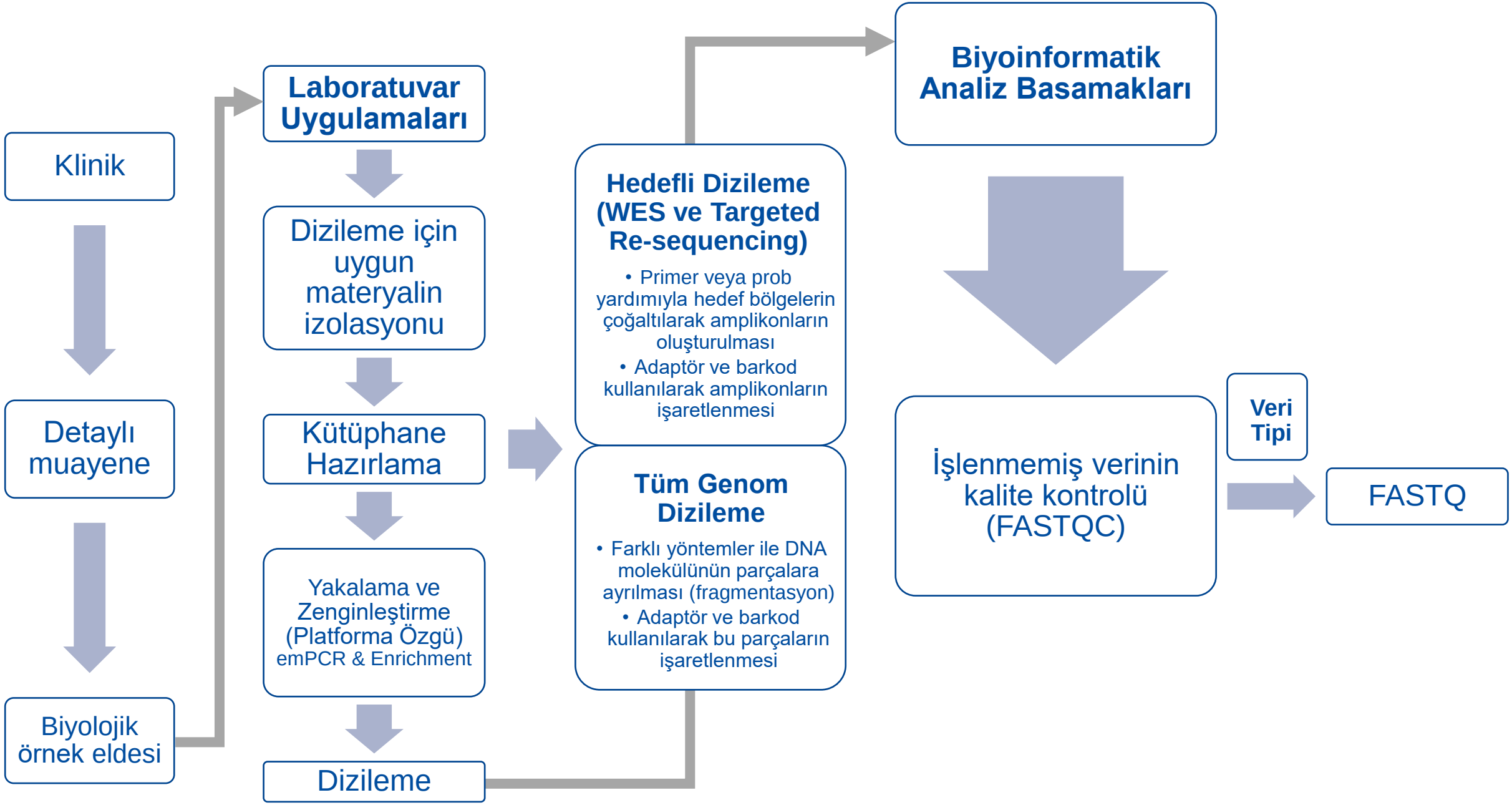


Bedia, C. Experimental Approaches in Omic Sciences (2018) Comprehensive Analytical Chemistry, 82, pp. 13-36.

DNA Dizileme Stratejileri

- **Genom Dizileme (WGS)**
  - ≃ 3.200.000.000 nükleotid çiftinden oluşan insan genomunun tek seferde dizilenmesi
- **Ekzom Dizileme (WES)**
  - Genetik hastalıklar ile ilişkili mutasyonların yaklaşık %90'ı genlerin protein kodlayan bölgeleri olan ekzonlarda yer almaktadır.
  - İnsan genomunun %1-2'lik kısmını oluşturan ekzonlar ≃50 Mb (milyon baz/ megabaz) büyüklüğündedir.
  - Genomdaki bütün genlere ait toplam 180.000 ekzonik bölgenin tek seferde dizilenmesi **Ekzom Dizileme** yöntemi olarak adlandırılır.




Klinik
Detaylı muayene
Biyolojik örnek eldesi
Laboratuvar Uygulamaları
Dizileme için uygun materyalin izolasyonu
Kütüphane Hazırlama
Yakalama ve Zenginleştirme (Platforma Özgü)
emPCR & Enrichment
Dizileme
Hedefli Dizileme (WES ve Targeted Re-sequencing)
• Primer veya prob yardımıyla hedef bölgelerin çoğaltılarak amplikonların oluşturulması
• Adaptör ve barkod kullanılarak amplikonların işaretlenmesi
Tüm Genom Dizileme
• Farklı yöntemler ile DNA molekülünün parçalara ayrılması (fragmentasyon)
• Adaptör ve barkod kullanılarak bu parçaların işaretlenmesi
Biyoinformatik Analiz Basamakları
İşlenmemiş verinin kalite kontrolü (FASTQC)
Veri Tipi
FASTQ

- Veri Tipleri

FASTQ:

FASTQ format is a text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding **quality scores.**

bam (Binary Alignment/Map) file

vcf (Variant Call Format)

Variant Table

Header Sequence Quality

```
@HWI-ST227:389:C4WA2ACXX:7:1204:2272:59979
GGAGGAAGGTCCTCGCTCCTCTTTCATATAAGGGAAATGGCTGAAT
+
FFFFHHHHHHJIJJJJJJJJIJJJIGIGIGGIJJIJIJJJJJJIII
@HWI-ST227:389:C4WA2ACXX:7:1205:15214:42893
GAGGATCCCAGGGAGGAAGGTCCTCGCTCCTCTTTCATCTAAGGGA
+
12BAFB?A:3<AE1@<FF;1*@EG*)?0?DBD>9BF9B*?######
@HWI-ST227:389:C4WA2ACXX:8:2208:2467:44624
AAAGAGGAGAGAGGACCATCCTCCCTGGGATCCTCAGAAGTCTACT
+
BDDA:DB?2AA@FC>F?EEGC<FED>GFD;?GBB?<?F99*/9?9?
```

Explanation on FASTQ Format

Every read is stored in four lines in FASTQ format as follows:

@EAS139:136:FC706VJ:2:2104:15343:197393 1:Y:18:ATCACG

GCTCTTTGCCCTTCTCGTCGAAAATTGTCTCCTCATTCGAAACTTCTCTGT +

@@CFFFDEHHHHFIJJJ@FHGIIIEHIIJBHHHIJJEGIIJJIGHIGHCCF

Line 1 begins with a '@' character which is followed by a sequence identifier and an optional description.

Line 2 shows the sequenced bases.

Line 3 begins with a '+' character and is optionally followed by the same sequence identifier (and any description) again.

Line 4 encodes the sequencing quality for each base in line 2, and must contain the same number of characters as bases in line 2. The ASCII value of every character minus 33 equals to the phred-scaled quality score of the sequenced base.

Explanation on FASTQ Format

@EAS139:136:FC706VJ:2:2104:15343:197393 1:Y:18:ATCACG

GCTCTTTGCCCTTCTCGTCGAAAATTGTCTCCTCATTCGAAACTTCTCTGT +

@@CFFFDEHHHHFIJJJ@FHGIIIEHIIJBHHHIJJEGIIJJIGHIGHCCF

EAS139:	The unique instrument name
136:	Run ID
FC706VJ:	Flowcell ID
2:	Flowcell lane
2104:	Tile number within the flowcell lane
15343:	‘x’-coordinate of the cluster within the tile
197393:	‘y’-coordinate of the cluster within the tile
1:	Member of a pair, 1 or 2 (paired-end or mate-pair reads only)
Y:	Y if the read fails filter (read is bad), N otherwise
18:	0 when none of the control bits are on, otherwise it is an even number
ATCACG:	Index sequence

ASCII Table

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	*[NULL]*	32	20	*[SPACE]*	64	40	@	96	60	`
1	1	*[START OF HEADING]*	33	21	!	65	41	A	97	61	a
2	2	*[START OF TEXT]*	34	22	"	66	42	B	98	62	b
3	3	*[END OF TEXT]*	35	23	#	67	43	C	99	63	c
4	4	*[END OF TRANSMISSION]*	36	24	$	68	44	D	100	64	d
5	5	*[ENQUIRY]*	37	25	%	69	45	E	101	65	e
6	6	*[ACKNOWLEDGE]*	38	26	&	70	46	F	102	66	f
7	7	*[BELL]*	39	27	'	71	47	G	103	67	g
8	8	*[BACKSPACE]*	40	28	(	72	48	H	104	68	h
9	9	*[HORIZONTAL TAB]*	41	29	)	73	49	I	105	69	i
10	A	*[LINE FEED]*	42	2A	*	74	4A	J	106	6A	j
11	B	*[VERTICAL TAB]*	43	2B	+	75	4B	K	107	6B	k
12	C	*[FORM FEED]*	44	2C	,	76	4C	L	108	6C	l
13	D	*[CARRIAGE RETURN]*	45	2D	-	77	4D	M	109	6D	m
14	E	*[SHIFT OUT]*	46	2E	.	78	4E	N	110	6E	n
15	F	*[SHIFT IN]*	47	2F	/	79	4F	O	111	6F	o
16	10	*[DATA LINK ESCAPE]*	48	30	0	80	50	P	112	70	p
17	11	*[DEVICE CONTROL 1]*	49	31	1	81	51	Q	113	71	q
18	12	*[DEVICE CONTROL 2]*	50	32	2	82	52	R	114	72	r
19	13	*[DEVICE CONTROL 3]*	51	33	3	83	53	S	115	73	s
20	14	*[DEVICE CONTROL 4]*	52	34	4	84	54	T	116	74	t
21	15	*[NEGATIVE ACKNOWLEDGE]*	53	35	5	85	55	U	117	75	u
22	16	*[SYNCHRONOUS IDLE]*	54	36	6	86	56	V	118	76	v
23	17	*[END OF TRANS. BLOCK]*	55	37	7	87	57	W	119	77	w
24	18	*[CANCEL]*	56	38	8	88	58	X	120	78	x
25	19	*[END OF MEDIUM]*	57	39	9	89	59	Y	121	79	y
26	1A	*[SUBSTITUTE]*	58	3A	:	90	5A	Z	122	7A	z
27	1B	*[ESCAPE]*	59	3B	;	91	5B	[	123	7B	{
28	1C	*[FILE SEPARATOR]*	60	3C	<	92	5C	\	124	7C	|
29	1D	*[GROUP SEPARATOR]*	61	3D	=	93	5D	]	125	7D	}
30	1E	*[RECORD SEPARATOR]*	62	3E	>	94	5E	^	126	7E	~
31	1F	*[UNIT SEPARATOR]*	63	3F	?	95	5F	_	127	7F	*[DEL]*

Sequence File Formats

fastq file:

Record name; starts with a "@"

DNA sequence

Phred score

```
@sequence_1
TTTCCGGGGCACATAATCTTCAGCCGGGCGC
+
9C;=;=<9@4868>9:67AA<9>65<=>591
@sequence_1
TCAGCCGGGCCTTCAGCCGGGGCACATAATA
+
('%3(&&&%.....
```

Kalite Skoru – FASTQC

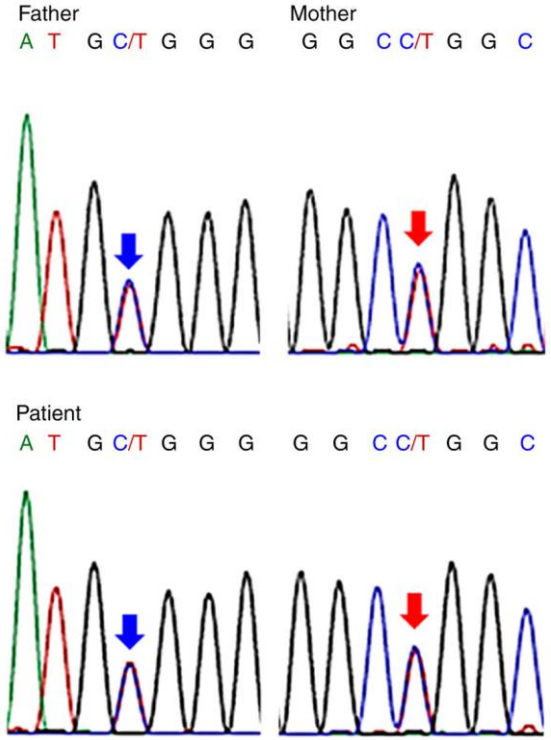
Babraham Bioinformatics

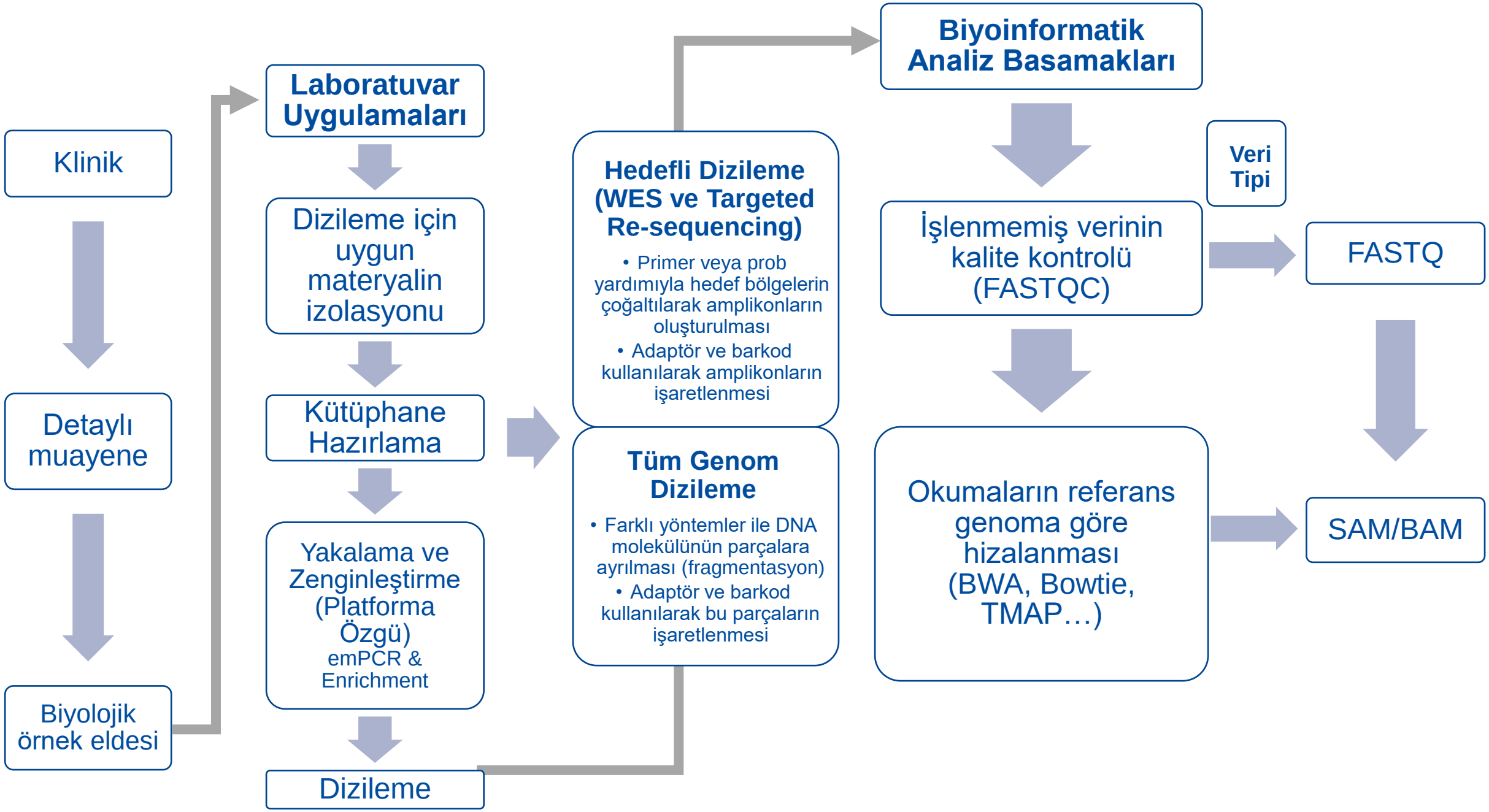
https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

- **Q10**
- **Q20**
- **Q30**
- **Q40**
  - .
  - .
  - .

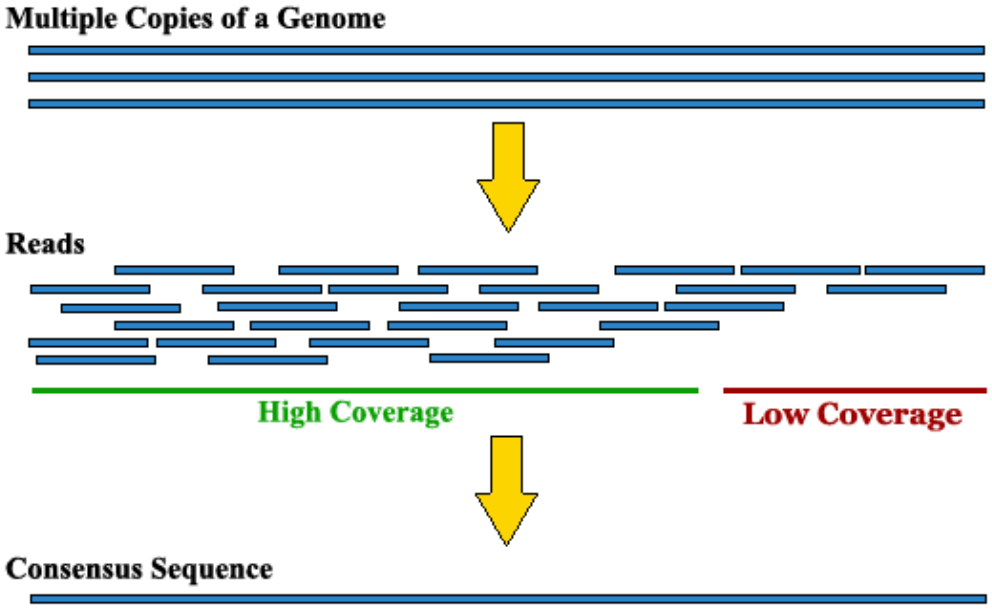
Phred Quality Scores

- Phred is a program that assigns a quality score to each base in a sequence. These scores can then be used to trim bad data from the ends, and to determine how good an overlap actually is.
  - there are much improved algorithms now, but the phred quality score is still widely used.
- Phred scores are logarithmically related to the probability of an error: a score of 10 means a 10% error probability; 20 means a 1% chance, 30 means a 0.1% chance, etc.
  - Q = -10 log P, where Q is the phred score and P is the probability that the base was called incorrectly.
  - A score of 20 is generally considered the minimum acceptable score.




Klinik
Detaylı muayene
Biyolojik örnek eldesi
Laboratuvar Uygulamaları
Dizileme için uygun materyalin izolasyonu
Kütüphane Hazırlama
Yakalama ve Zenginleştirme (Platforma Özgü)
emPCR & Enrichment
Dizileme
Hedefli Dizileme (WES ve Targeted Re-sequencing)
• Primer veya prob yardımıyla hedef bölgelerin çoğaltılarak amplikonların oluşturulması
• Adaptör ve barkod kullanılarak amplikonların işaretlenmesi
Tüm Genom Dizileme
• Farklı yöntemler ile DNA molekülünün parçalara ayrılması (fragmentasyon)
• Adaptör ve barkod kullanılarak bu parçaların işaretlenmesi
Biyoinformatik Analiz Basamakları
İşlenmemiş verinin kalite kontrolü (FASTQC)
Veri Tipi
FASTQ
Okumaların referans genoma göre hizalanması (BWA, Bowtie, TMAP…)
SAM/BAM

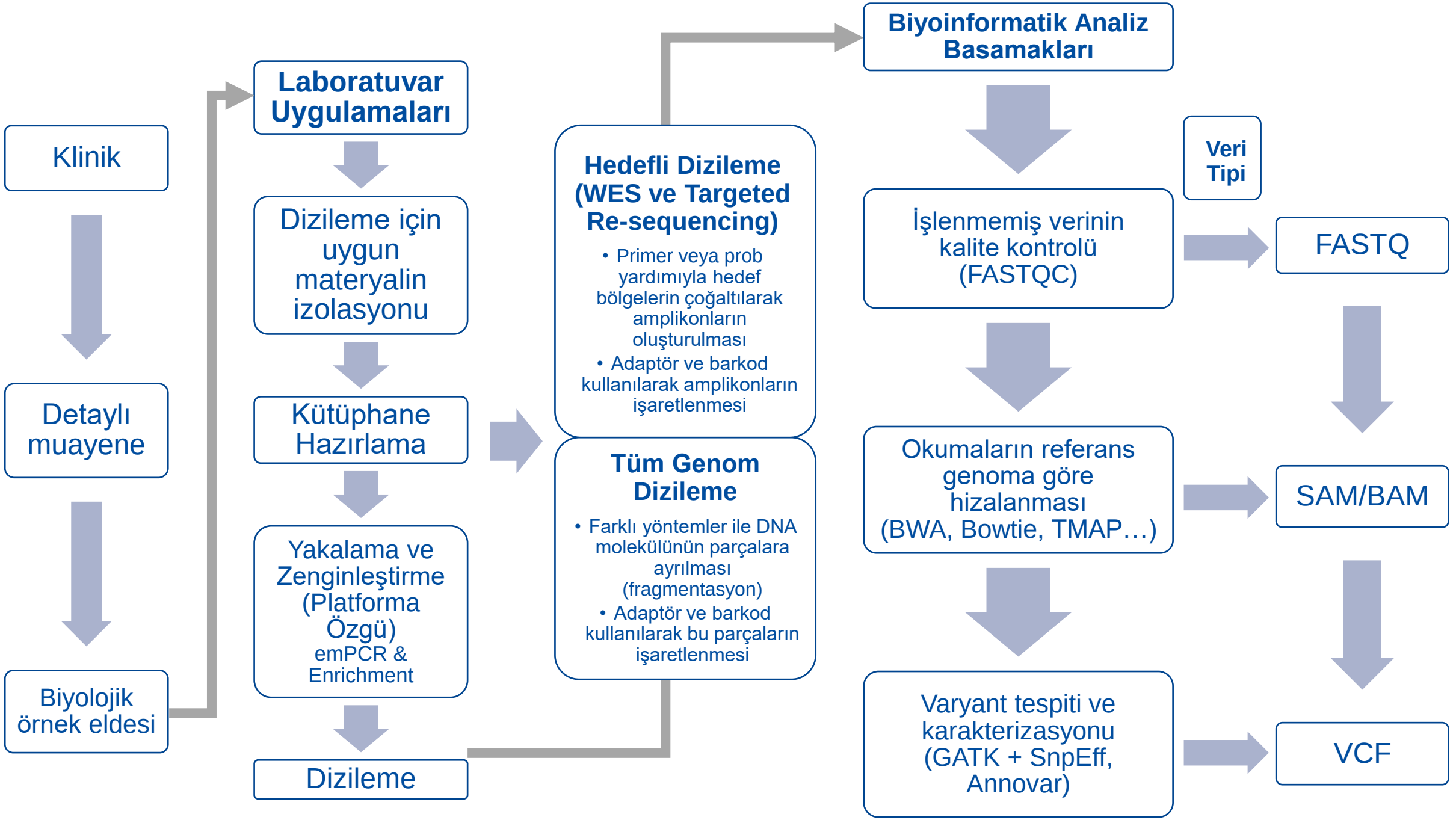
Dizi Hizalaması (Alignment)



Aligned reads

TGAAGTCCTACAGTCATAGTC
AAGTCCTACAGTCATAGTCGA
GTCCTACAGTCATAGTCGATA
CCTACAGTCATAGTCGATATT
TACAGTCATAGTCGATATTT

Consensus contig TGAAGTCCTACAGTCATAGTCGATATTT


Klinik
Detaylı muayene
Biyolojik örnek eldesi
Laboratuvar Uygulamaları
Dizileme için uygun materyalin izolasyonu
Kütüphane Hazırlama
Yakalama ve Zenginleştirme (Platforma Özgü)
emPCR & Enrichment
Dizileme
Hedefli Dizileme (WES ve Targeted Re-sequencing)
• Primer veya prob yardımıyla hedef bölgelerin çoğaltılarak amplikonların oluşturulması
• Adaptör ve barkod kullanılarak amplikonların işaretlenmesi
Tüm Genom Dizileme
• Farklı yöntemler ile DNA molekülünün parçalara ayrılması (fragmentasyon)
• Adaptör ve barkod kullanılarak bu parçaların işaretlenmesi
Biyoinformatik Analiz Basamakları
İşlenmemiş verinin kalite kontrolü (FASTQC)
Okumaların referans genoma göre hizalanması (BWA, Bowtie, TMAP…)
Varyant tespiti ve karakterizasyonu (GATK + SnpEff, Annovar)
Veri Tipi
FASTQ
SAM/BAM
VCF

Varyant Tablolarının Oluşturulması

Açımlama (*Annotation*)

- İlgilenilen hastalıktan sorumlu genetik etkenin saptanması için hizalama sonrası belirlenen değişikliklerin (varyant) karakterize edilerek tablo şekline getirilmesi gerekir.

- Varyantın genomdaki pozisyonu, nükleotid değişikliği, cDNA ve protein yapısına olan etkisine göre adlandırılması gibi birçok bilgiyi içeren bu tabloların oluşturulması için hizalama sonrası oluşturulan vcf dosyası kullanılır.

Veri Tipleri

Kalite Kontrol
(FASTQC)

FASTA & FASTQ:

FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which nucleotides or amino acids are represented using **single-letter codes.**

FASTQ format is a text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding **quality scores.**

Hizalama
(Alignment)

SAM/BAM (*The Sequence Alignment – Binary Alignment/Map*) File

IGV

Açımlama
(Annotation)

VCF (*Variant Call Format*)

Varyant Tablosu

Yorumlama

Veri Tabanları, Üçüncü Parti Yazılımlar ve Tahminleme Programları

Veri Tabanları

gnomAD
genome aggregation database

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Ensembl

NCBI National Center for Biotechnology Information

NIH National Institutes of Health

Üçüncü Parti Yazılımlar

- Platformdan bağımsız veya platforma özgü

TÜBİTAK BİLGEM HomSI

AgileVCFMapper

igv Integrative Genomics Viewer

illumina

QIAGEN QCI Interpret

Tahminleme Programları

Mutation Taster

mutation t@sting

PolyPhen-2 prediction of functional effects
Home | About | Help | Downloads

J. Craig Venter INSTITUTE

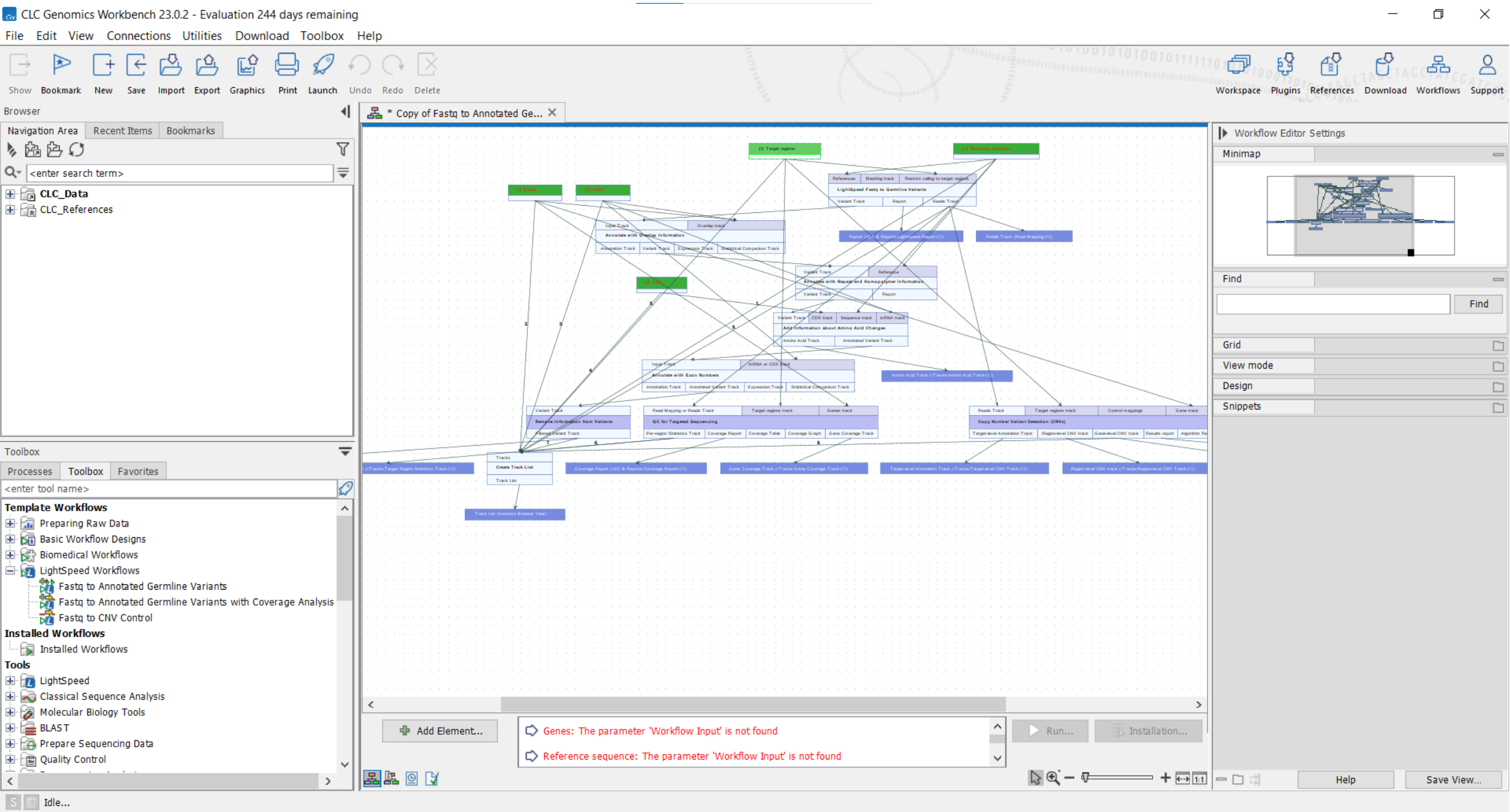
SIFT

PROVEAN

REVEL: Rare Exome Variant Ensemble Learner

VEST (Variant Effect Scoring Tool)

varsome
The Human Genomics Community
VarSome is powered by Saphetor, the Genome Interpreter


CLC Genomics Workbench 23.0.2 - Evaluation 244 days remaining
File Edit View Connections Utilities Download Toolbox Help
Show Bookmark New Save Import Export Graphics Print Launch Undo Redo Delete
Workspace Plugins References Download Workflows Support
Browser
Navigation Area Recent Items Bookmarks
<enter search term>
CLC_Data
CLC_References
Toolbox
Processes Toolbox Favorites
<enter tool name>
Template Workflows
Preparing Raw Data
Basic Workflow Designs
Biomedical Workflows
LightSpeed Workflows
Fastq to Annotated Germline Variants
Fastq to Annotated Germline Variants with Coverage Analysis
Fastq to CNV Control
Installed Workflows
Installed Workflows
Tools
LightSpeed
Classical Sequence Analysis
Molecular Biology Tools
BLAST
Prepare Sequencing Data
Quality Control
* Copy of Fastq to Annotated Ge...
Add Element...
Genes: The parameter 'Workflow Input' is not found
Reference sequence: The parameter 'Workflow Input' is not found
Run...
Installation...
Workflow Editor Settings
Minimap
Find
Find
Grid
View mode
Design
Snippets
Help
Save View...
Idle...

Home > Product Log in

Ingenuity Pathway Analysis

DOWNLOAD CLIENT INSTALLER | SIGN UP | LOG IN

Ingenuity Pathway Analysis for China

DOWNLOAD CLIENT INSTALLER | SIGN UP | LOG IN

QCI Interpret

LOG IN FOR: | USA | CHINA | EUROPE | TURKEY | JAPAN

QCI Interpret Translational

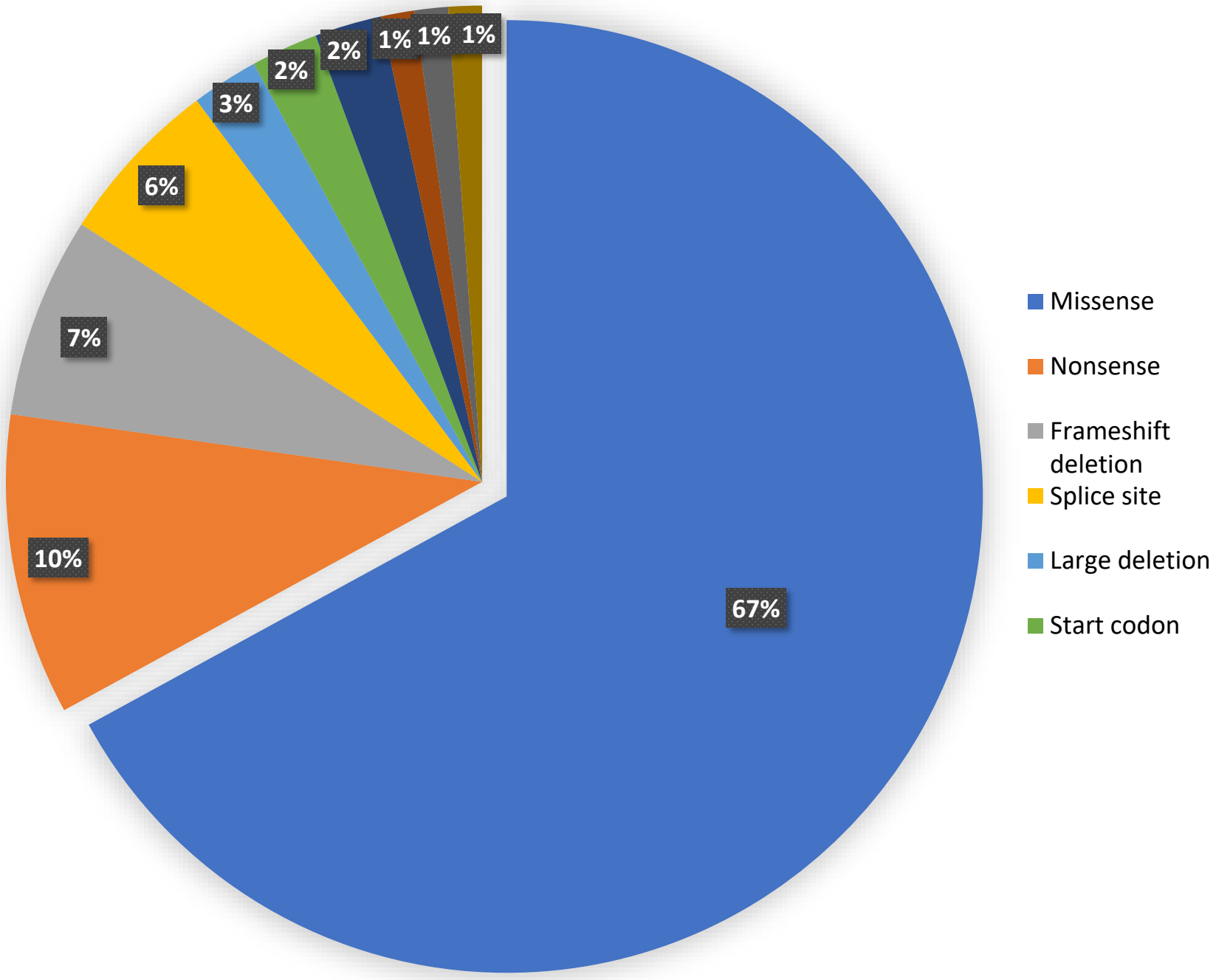
LOG IN FOR: | USA | CHINA

Kalıtsal Metabolik Hastalıklarda Biyoinformatik Analiz, Derin Fenotipleme, Protein Modelleme ve Yolak Analizleri ile Tanı Algoritmalarının Geliştirilmesi

Can Koşukcu

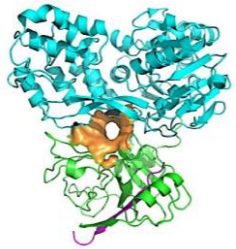
Hacettepe Üniversitesi Moleküler Metabolizma Bölümü

Doktora Öğrencisi

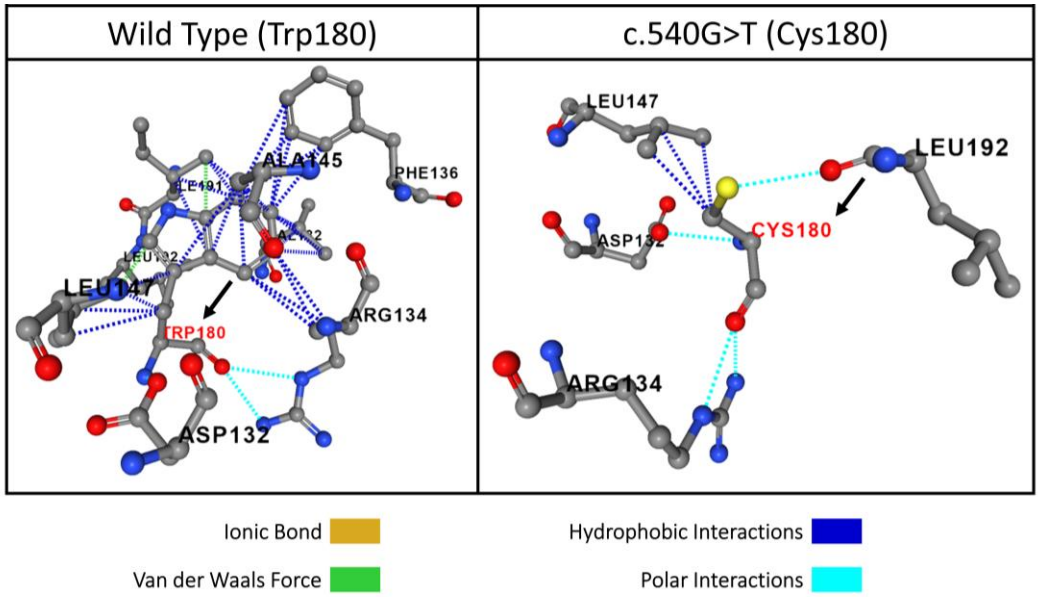

67%
10%
7%
6%
3%
2%
2%
1%
1%
1%
Missense
Nonsense
Frameshift deletion
Splice site
Large deletion
Start codon

Yanlış anlamlı mutasyonlar

- Doğrudan protein-protein etkileşimlerini bozarak,
- Enzim aktivitesini etkileyerek,
- Proteinin yapısal kararsızlığını (destabilizasyon) indükleyerek



Proteinin fonksiyon kaybına (LoF) neden olabilir ve bu durum protein yanlış katlanmasını ve bozulmasını tetikleyebilir (Stein et al., 2019).



Yanlış Anlamlı Mutasyonlara Bağlı Genetik Hastalıkların Anlaşılmasında Protein Kararlılığının Rolü (ΔΔG)

Published online 23 October 2020

Nucleic Acids Research, 2021, Vol. 49, Database issue ***D475–D479***
doi: 10.1093/nar/gkaa925

ThermoMutDB: a thermodynamic database for missense mutations

Joicymara S. Xavier[1,2], Thanh-Binh Nguyen[3], Malancha Karmarkar[3,4], Stephanie Portelli[3,4], Pâmela M. Rezende[2], João P.L. Velloso[2], David B. Ascher[3,4,5,*] and Douglas E.V. Pires[3,4,6,*]

[1]Institute of Agricultural Sciences, Universidade Federal dos Vales do Jequitinhonha e Mucuri, [2]Instituto René Rachou, Fundação Oswaldo Cruz, [3]Bio 21 Institute, University of Melbourne, [4]Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, [5]Department of Biochemistry, University of Cambridge and [6]School of Computing and Information Systems, University of Melbourne

Received August 15, 2020; Revised September 21, 2020; Editorial Decision October 05, 2020; Accepted October 12, 2020

→ Ocak 2021

BioMedInformatics

Article

Gibbs Free Energy, a Thermodynamic Measure of Protein–Protein Interactions, Correlates with Neurologic Disability

Michael Keegan [1], Hava T. Siegelmann [1], Edward A. Rietman [1,2,3], Giannoula Lakka Klement [3] and Jack A. Tuszynski [4,5,6,*]

14 Aralık 2021

International Journal of *Molecular Sciences*

MDPI

Article

Using the Gibbs Function as a Measure of Human Brain Development Trends from Fetal Stage to Advanced Age

Edward A. Rietman [1], Sophie Taylor [2], Hava T. Siegelmann [1], Marco A. Deriu [3], Marco Cavaglia [3,4] and Jack A. Tuszynski [2,3,*]

Şubat 2020

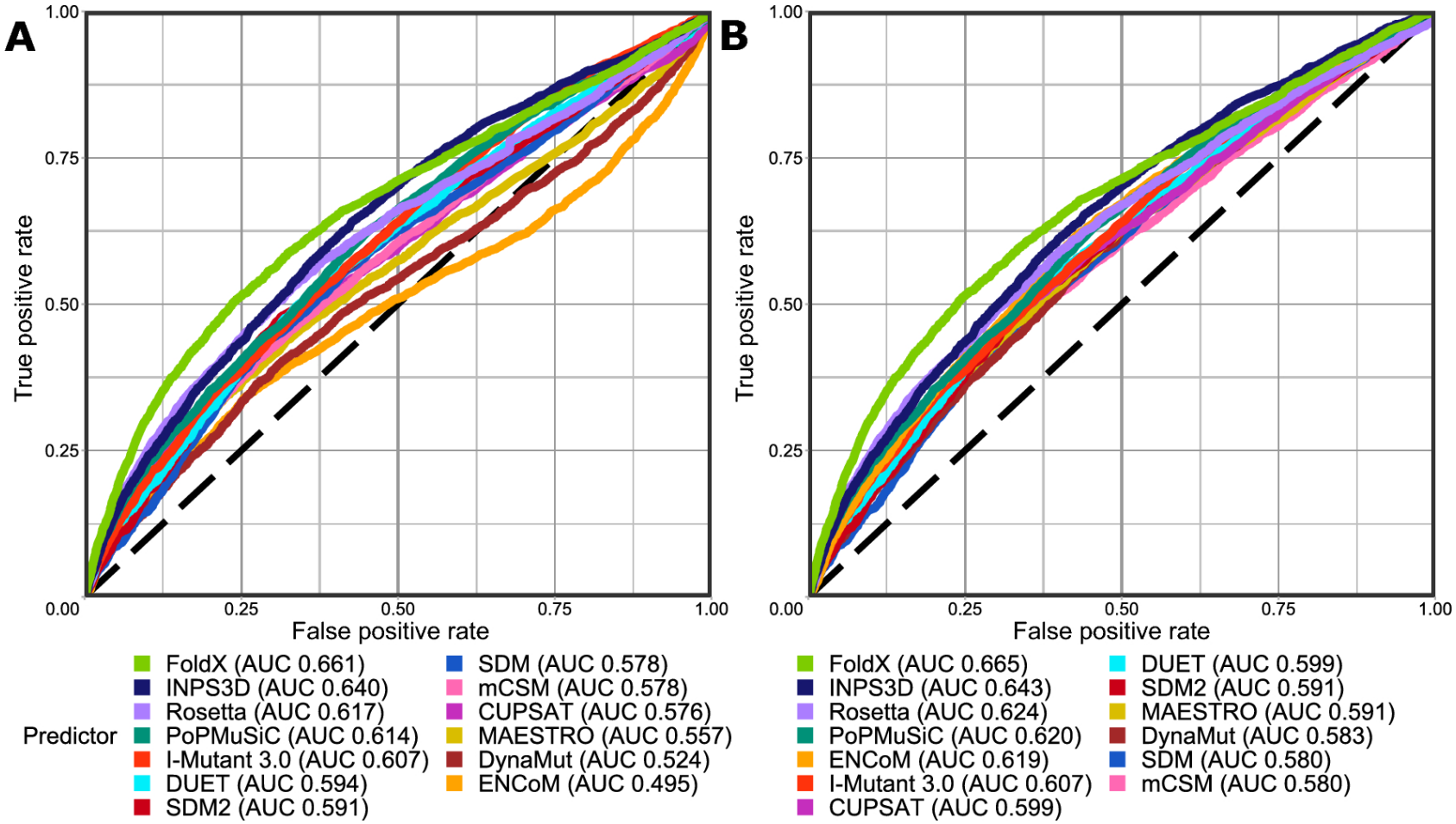
OPEN

Eylül 2020

Identification of pathogenic missense mutations using protein stability predictors

Lukas Gerasimavicius, Xin Liu & Joseph A. Marsh✉

SCIENTIFIC REPORTS
natureresearch



Yanlış Anlamlı Mutasyonların Protein Üzerindeki Etkilerinin Modellenmesi

≅ 120 patojenik varyant

Yanlış anlamlı mutasyon

Proteinin insan PDB dosyası var mı?

Hayır

Proteinin model organizma PDB dosyası var mı?

Hayır

Proteinin üç boyutlu yapısını modelle

Evet

Hayır

Evet

Hayır

Mutasyona uğrayan amino asit modelde mevcut mu?

Sekans homolojisi insan dizisiyle %90 veya üzerinde mi?

Evet

FoldX RepairPDB komutu

Evet

FoldX RepairPDB komutu

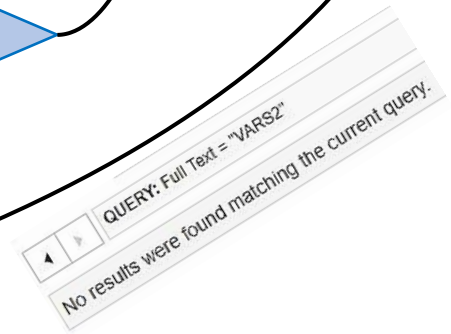
Mutasyonun protein stabilitesi üzerindeki etkisini hesapla (ΔΔG)

2O8N
Crystal Structure of Mouse Apolipoprotein
Shumilin, I.A., Jha, K.N., Zheng, H., Chruszcz, M., Cyr
(2008) Endocrinology 149: 2108-2120

Released	2007-12-25
Method	X-RAY DIFFRACTION 2 Å
Organisms	Mus musculus
Macromolecule	ApoA-I binding protein (protein)
Unique Ligands	CL, SO4

4CHL
Human Ethylmalonic Encephalopathy Protein
Pettinati, I., Brem, J., McDonough, M.A., Schofield, C.J.
(2015) Hum Mol Genet 24: 2458-2469

Released	2014-12-17
Method	X-RAY DIFFRACTION 2.61 Å
Organisms	Homo sapiens
Macromolecule	PERSULFIDE DIOXYGENASE ETHE
Unique Ligands	FE



Display Files | Download Files

2O8N

Crystal Structure of Mouse Apolipoprotein A-I Binding Protein

DOI: 10.2210/pdb2O8N/pdb

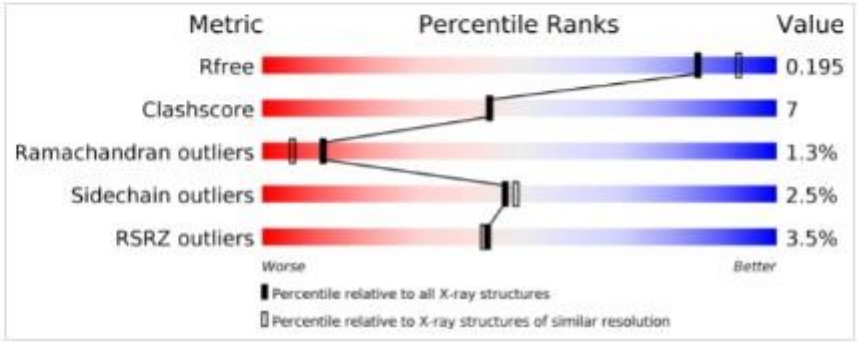
Classification: PROTEIN BINDING
Organism(s): Mus musculus
Expression System: Escherichia coli BL21(DE3)
Mutation(s): Yes

Deposited: 2006-12-12 **Released:** 2007-12-25
Deposition Author(s): Shumilin, I.A., Jha, K.N., Zheng, H., Chruszcz, M., Cymborowski, M., Herr, J.C., Minor, W.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.00 Å
R-Value Free: 0.196
R-Value Work: 0.169
R-Value Observed: 0.171

wwPDB Validation 3D Report | Full Report



FASTA Sequence

PDB Format
PDB Format (gz)

PDBx/mmCIF Format
PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

Biological Assembly 1

Structure Factors (CIF)
Structure Factors (CIF - gz)

Validation Full PDF
Validation XML

fo-fc Map (DSN6)
2fo-fc Map (DSN6)
Map Coefficients (MTZ format)

NAXE: Apolipoprotein A-I Binding Protein

The Protein Data Bank (pdb) dosyası nedir?

https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html

FoldX Komutlar

RepairPDB

- FoldX ile herhangi bir modelleme yapmadan önce yapıların onarılması tavsiye edilmektedir. RepairPDB komutu bozuk burulma açılarını, çakışan VanderWaals bağlarını veya toplam enerjiye etki eden bağ kalıntılarını tespit eder ve onarır.
- RepairPDB komutu şöyle çalışır: Komut önce tüm Asn, Gln ve His kalıntılarını arar ve amino asitleri 180 derece döndürür. Bu, Asn ve Gln karboksamid gruplarının elektron yoğunluğunun hemen hemen simetrik olması ve doğru yerleşimin ancak çevredeki atomlarla olan etkileşimler hesaplanarak fark edilmesi nedeniyle yapıda yanlış rotamer atamasını önlemek için yapılır.
- Aynı durum His amino asidi için de geçerlidir. Komut, VanderWaals çakışmalarını ortadan kaldırmak için yan zincirlerde küçük bir optimizasyon yapar. Bu şekilde son adımda yan zincirlerin stabilizasyonu sağlanır.

AMAÇ: Protein modelini daha kararlı hale getirmek (ΔG Değerini minimize etmek)

Received: 16 April 2021 | Revised: 12 May 2021 | Accepted: 18 May 2021
DOI: 10.1111/cge.14002

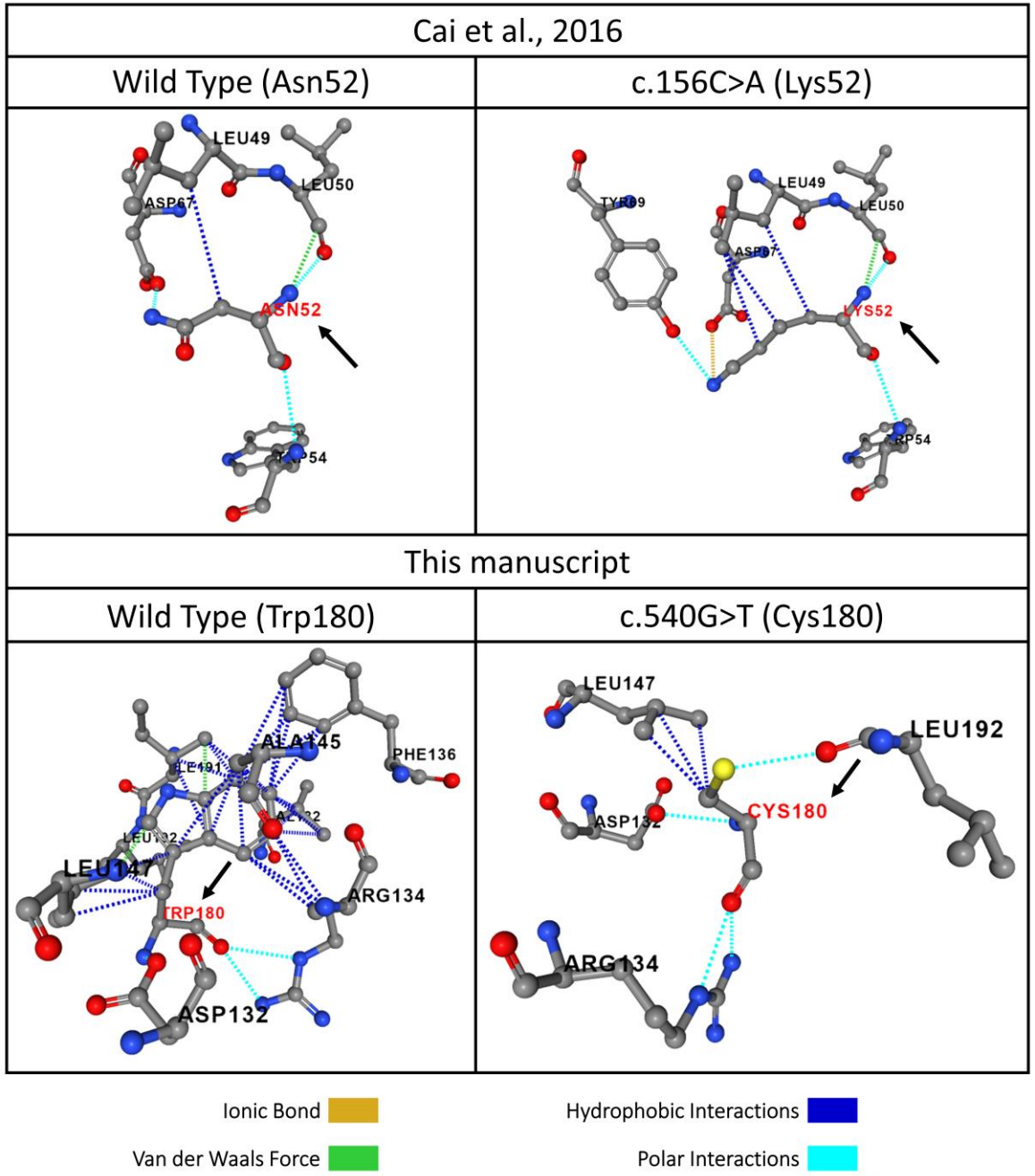
ORIGINAL ARTICLE

CLINICAL GENETICS WILEY

Homozygous missense *VPS16* variant is associated with a novel disease, resembling mucopolysaccharidosis-plus syndrome in two siblings

Yılmaz Yıldız[1,2] | Can Koşukcu[1,3] | Damla Aygün[1] | Meltem Akçaboy[4] | Fatma Zehra Öztek Çelebi[4] | Yasemin Taşcı Yıldız[5] | Gülseren Şahin[6] | Caner Aytekin[7] | Deniz Yüksel[8] | İncilay Lay[9] | Rıza Köksal Özgül[1,10] | Ali Dursun[1]

- The starting free energy ΔG for the VPS16 protein model was 241.395 kcal/mol, and it was lowered to 148.076 kcal/mol to a more stable state after the RepairPDB module was applied.
- The ΔΔG value of the p.Asn52Lys variant reported in late-onset dystonia patients was calculated as 0.43 kcal/mol (neutral). The ΔΔG value of the p.Trp180Cys variant (reported in this study) was calculated as 1.29 kcal/mol (destabilizing).
- Decreased amounts of VPS16 protein disrupts HOPS and CORVET, resulting in the MPS-plus syndrome-like phenotype. In contrast, the p.Asn52Lys substitution is not predicted to result in significant change in protein stability.
- Our 3D protein model suggests that an important functional difference between these two substitutions may be at least partially due to the relatively more severe amino acid substitution from tryptophan to cysteine (where the aromatic indol side chain is replaced by a thiol side chain) compared with the milder amino acid change from asparagine to lysine (both harboring side chains with an amino group), which allows it to retain many of the non-covalent interactions with the surrounding amino acid residues



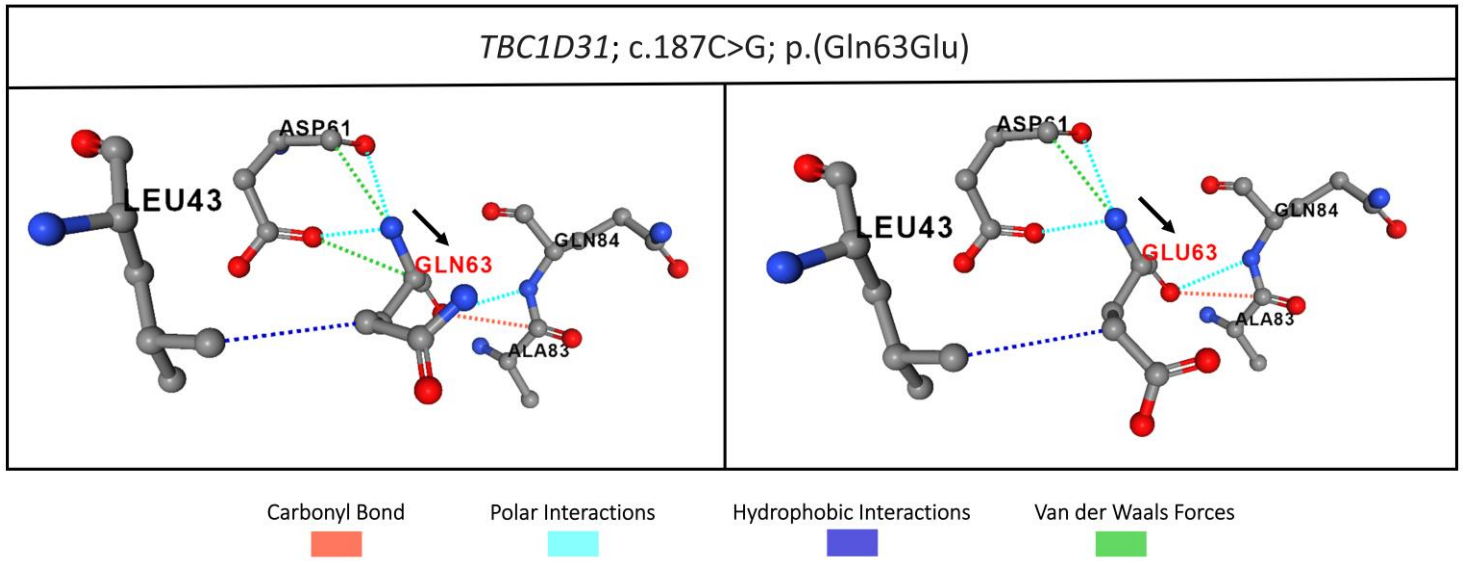
Received: 18 May 2023 | Revised: 28 June 2023 | Accepted: 8 July 2023
DOI: 10.1111/cge.14406

SHORT REPORT

CLINICAL GENETICS WILEY

A novel homozygous missense variant in *TBC1D31* in a consanguineous family with congenital anomalies of the kidney and urinary tract (CAKUT)

Seha Saygılı[1] | Can Koşukcu[2] | Turgut Baştuğ[3] | Özlem Akgün Doğan[4] | Esra Karabağ Yılmaz[1] | Ayşe Uçar Kalyoncu[5] | Ayşe Ağbaş[1] | Nur Canpolat[1] | Salim Çalışkan[1] | Fatih Ozaltin[2,6,7,8]



The starting free energy ΔG for the TBC1D31 protein model was 103.34 kcal/mol, and it was lowered to -22.4663 kcal/mol to a highly stable state after the RepairPDB command of FoldX was applied. 3D protein models of the wild type TBC1D31 and the potential changes on the noncovalent bonds caused by the p.Gln63Glu variant are displayed in Figure 2. Six polar, 1 hydrophobic, 1 carbonyl, 1 Van der Walls interactions changed to 6 polar, 1 carbonyl and 1 Van der Walls interactions after the glutamine at position 63 turned into glutamic acid. The in-silico predictions showed that one hydrophobic interaction was lost on the protein model (Figure 2). The p.Gln63Glu variant decreased protein stability and predicted as "Destabilizing" with all four different protein stability prediction tools upon point mutations.

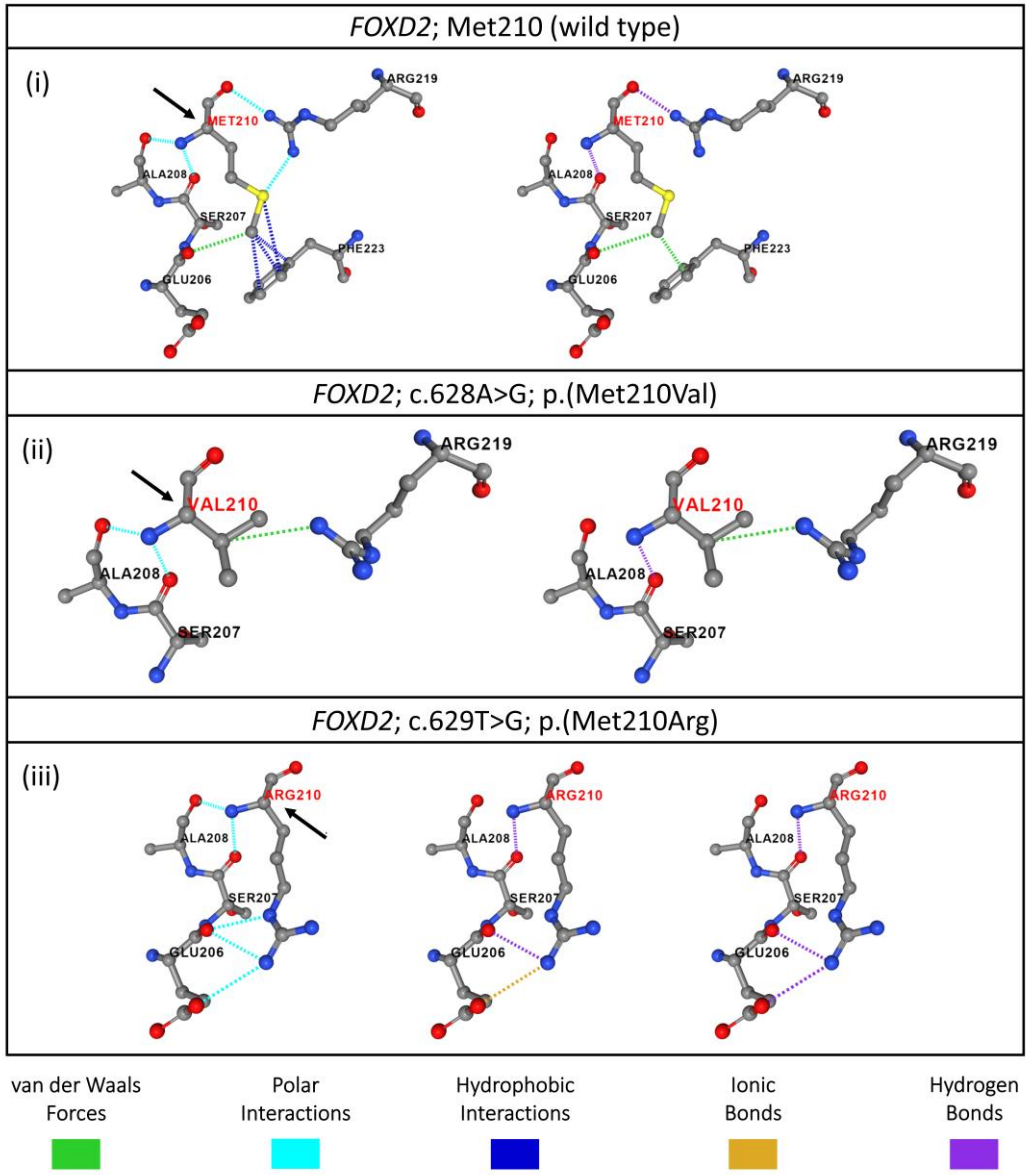
Implication of transcription factor FOXD2 dysfunction in syndromic congenital anomalies of the kidney and urinary tract (CAKUT)

Check for updates

OPEN

Korbinian M. Riedhammer[1,2,37], Thanh-Minh T. Nguyen[3,37], Can Koşukcu[4,37], Julia Calzada-Wack[5], Yong Li[6], Nurit Assia Batzir[7], Seha Saygılı[8], Vera Wimmers[9,10], Gwang-Jin Kim[9], Marialena Chrysanthou[3], Zeineb Bakey[3,10], Efrat Sofrin-Drucker[7], Markus Kraiger[5], Adrián Sanz-Moreno[5], Oana V. Amarie[5], Birgit Rathkolb[5,11,12], Tanja Klein-Rodewald[5], Lillian Garrett[5,13], Sabine M. Hölter[5,13,14], Claudia Seisenberger[5], Stefan Haug[6], Pascal Schlosser[6,15], Susan Marschall[5], Wolfgang Wurst[13,14,16,17], Helmut Fuchs[5], Valerie Gailus-Durner[5], Matthias Wuttke[6], Martin Hrabe de Angelis[5,12,18], Jasmina Ćomić[1,2], Özlem Akgün Doğan[19], Yasemin Özlük[20], Mehmet Taşdemir[21], Ayşe Ağbaş[8], Nur Canpolat[8], Naama Orenstein[7,22], Salim Çalışkan[8], Ruthild G. Weber[23], Carsten Bergmann[24,25], Cecile Jeanpierre[26], Sophie Saunier[26], Tze Y. Lim[27], Friedhelm Hildebrandt[28], Bader Alhaddad[1], Lina Basel-Salmon[22,29,30], Yael Borovitz[22,31], Kaman Wu[3], Dinu Antony[3,10], Julia Matschkal[2], Christian W. Schaaf[1,2], Lutz Renders[2], Christoph Schmaderer[2], Manuel Rogg[32], Christoph Schell[32], Thomas Meitinger[1], Uwe Heemann[2], Anna Köttgen[6,33], Sebastian J. Arnold[9,33], Fatih Ozaltin[4,34,35,36],

The starting free energy ΔG for the FOXD2 protein model was 817.623 kcal/mol, and it was lowered to 719.079 kcal/mol to a more stable state after the RepairPDB command of FoldX was applied. 3D protein models for the wild type FOXD2 and the potential changes on the noncovalent bonds caused by the p.(Met210Val) and p.(Met210Arg) variants are displayed in Figure 1K. -The p.Met210Val variant decreased protein stability and was predicted as "Destabilizing" with all four, the p.Met210Arg variant decreased protein stability and was predicted as "Destabilizing" with three out of four different protein stability prediction tools upon point mutations. The ΔΔG scores calculated for 210VAL and 210ARG variants are displayed in Supplementary Table 3.



TEŞEKKÜRLER…