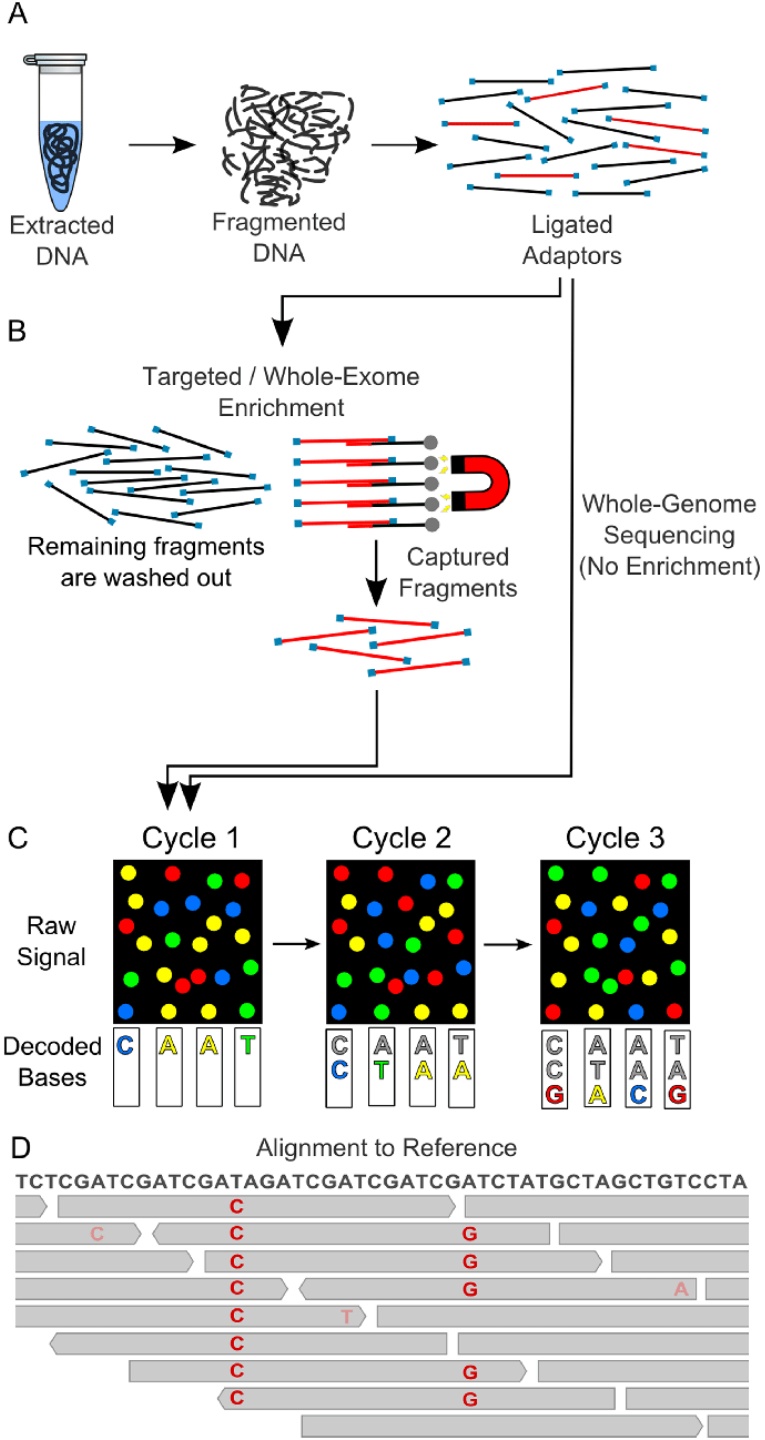
GENOMIK

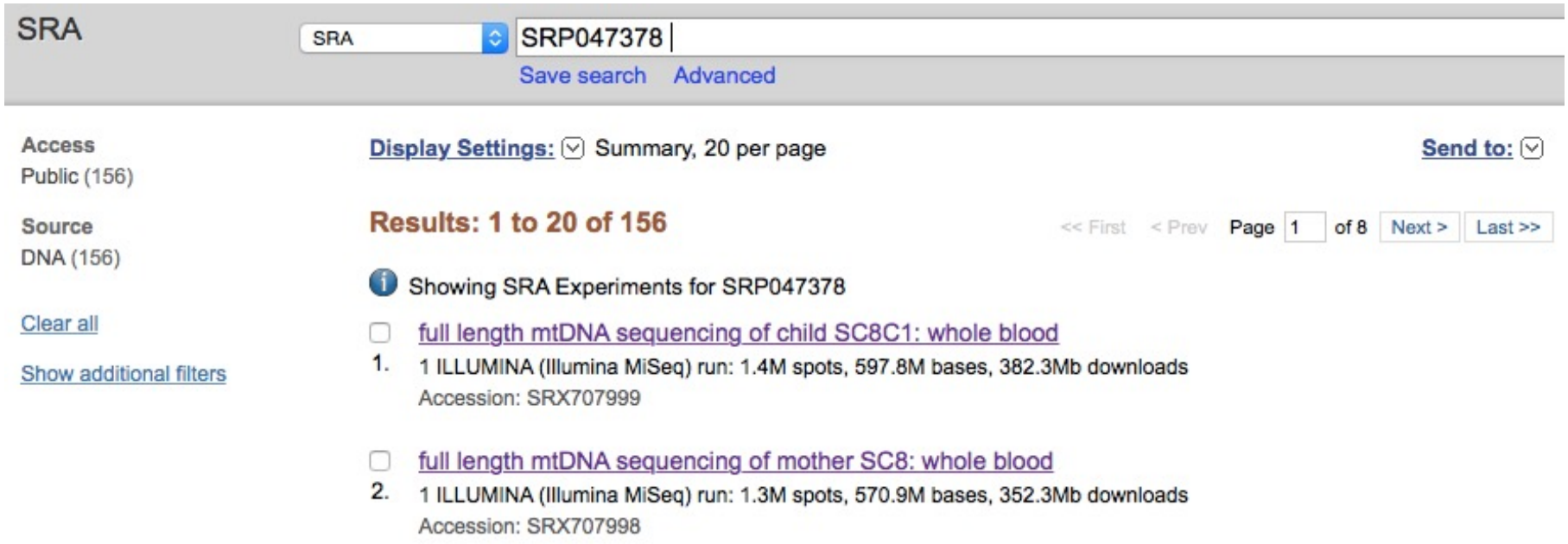
Tugce Bilgin Sonay


A
Extracted DNA
Fragmented DNA
Ligated Adaptors
B
Targeted / Whole-Exome Enrichment
Remaining fragments are washed out
Captured Fragments
Whole-Genome Sequencing (No Enrichment)
C
Cycle 1
Cycle 2
Cycle 3
Raw Signal
Decoded Bases
D
Alignment to Reference
TCTCGATCGATCGATAGATCGATCGATCGATCTATGCTAGCTGTCCTA

- Ilk 35-400 baz ciftini diziliyoruz: okuma
- Bir seferde 6 milyar okuma elde edebiliriz
- Paired-end olursa iki taraftan okuruz.

GTTGAGGCTTGCGTTTTTGGTACGCTGGACTTTGT
GTACTCGTCGCTGCGTTGAGGCTTGCGTTTTTGGT
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
CTTGCGTTTATGGTACGCTGGACTTTGTAGGATAC
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
GCGTTTATGGTACGCTGGACTTTGTAGGATACCCT
GAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGG
GCGTTGAGGCTTGCGTTTATGGTACGCTGGATTTT
CGTTTATGGTACGCTGGACTTTGTAGGATACCCTC
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT
GTTTATGGTACGCTGGACTTTGTAGGATACCCTCG
TCTCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTA
TGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTA
GCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTAC
TATGGTACGCTGGACTTTGTAGGATACCCTCGCTT
TCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTTTG
CGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCT
GTTGAGGCTTGCGTTTATGGTACGCTGGGCTTTTT
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC

SRA
Short **R**ead
Archive

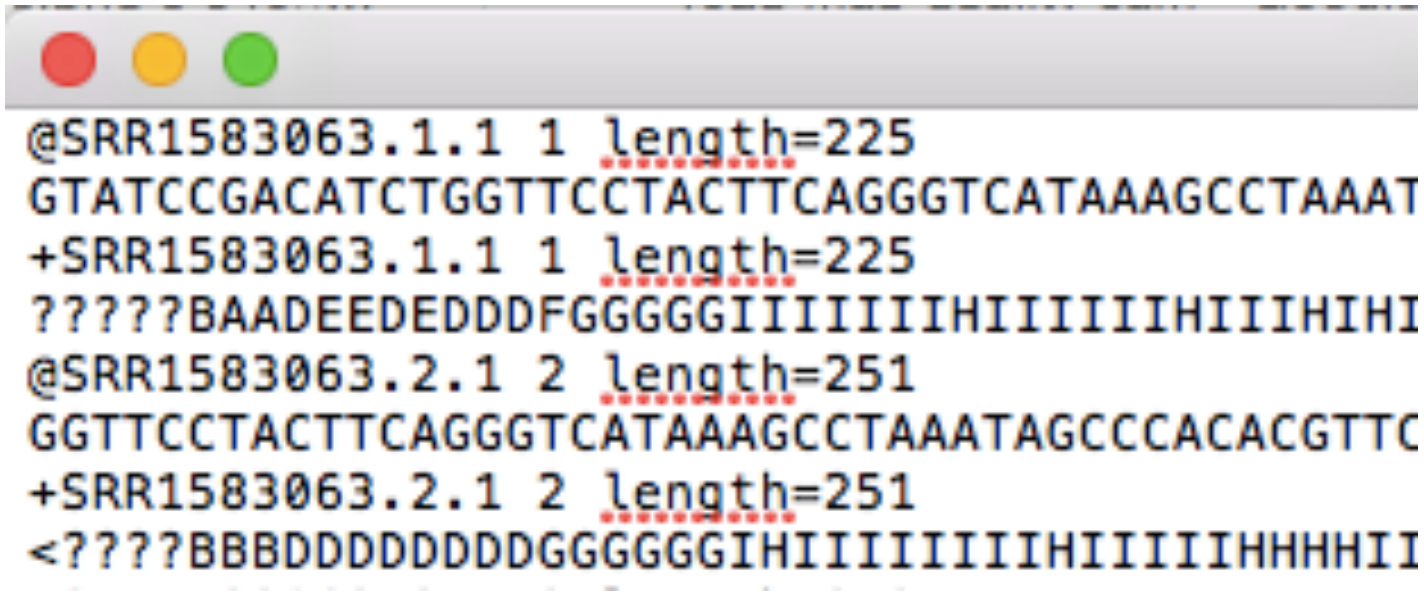


SRA
Short **R**ead
Archive

SRA toolkit

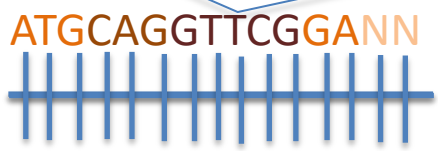
fastQ
fasta ve
Quality data

- Gene ID, uzunluk
- Fasta dizisi
- Gene ID, uzunluk
- Kalite verisi



Dizinin kalite degerleri

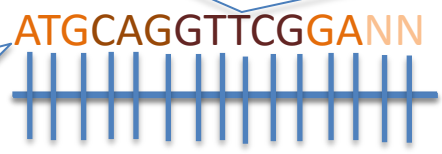
kalite degeri: 30 -> baz %99.9 ihtimal dogru



Dizinin kalite degerleri

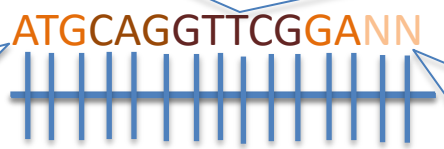
kalite degeri: 30 -> baz %99.9 ihtimal dogru

kalite degeri: 10-> baz %10 yanlis.



Dizinin kalite degerleri

kalite degeri: 30 -> baz %99.9 ihtimal dogru



kalite degeri: 10-> baz %10 yanlis.

bazin ne oldugu belirlenemediyse N

$$Q = -10 \log_{10} P \quad \Longrightarrow \quad P = 10^{\frac{-Q}{10}}$$

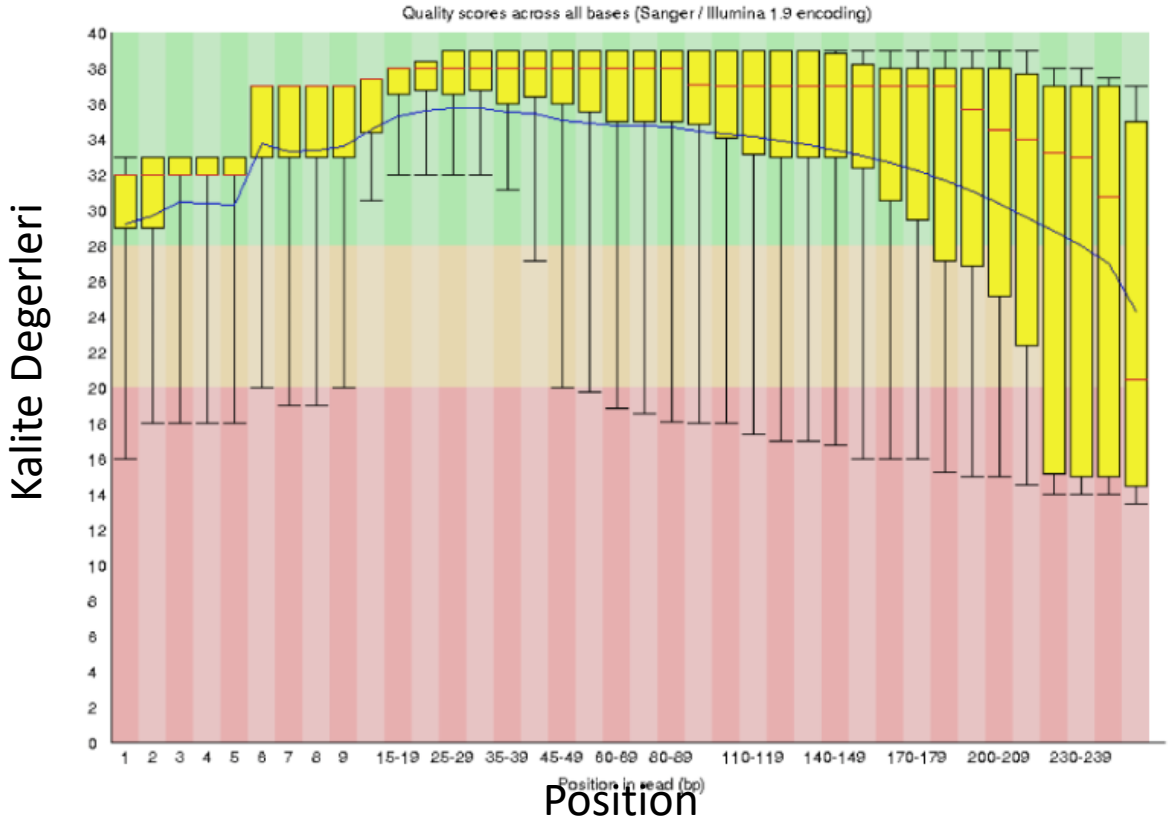
Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

SRA
Short **R**ead
Archive

SRA toolkit

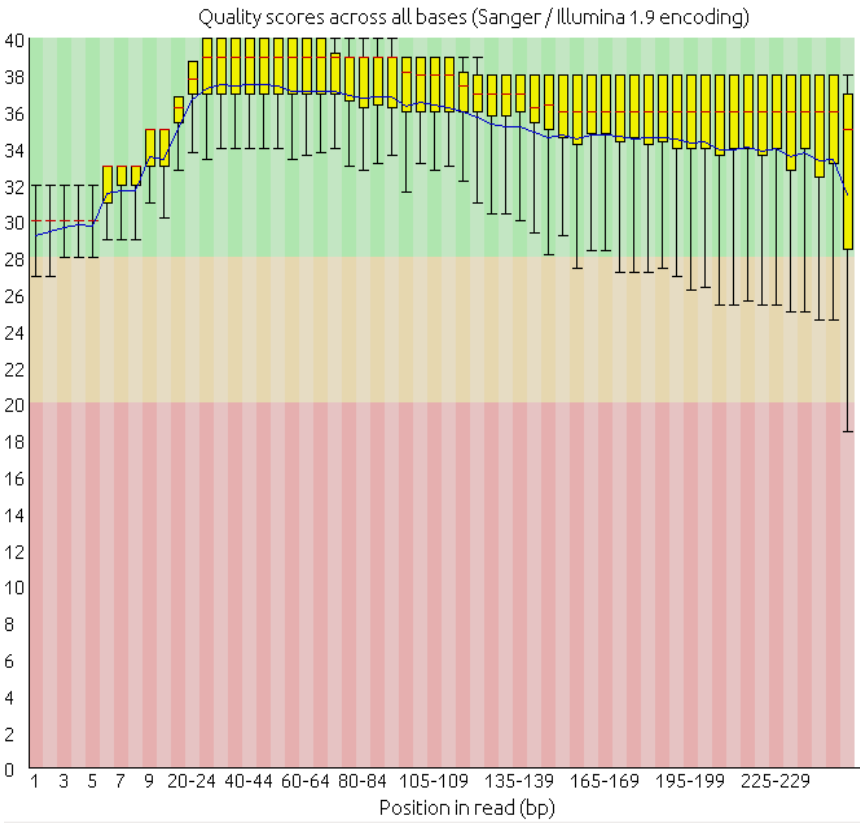
fastQ
fasta ve
Quality data

FastQC

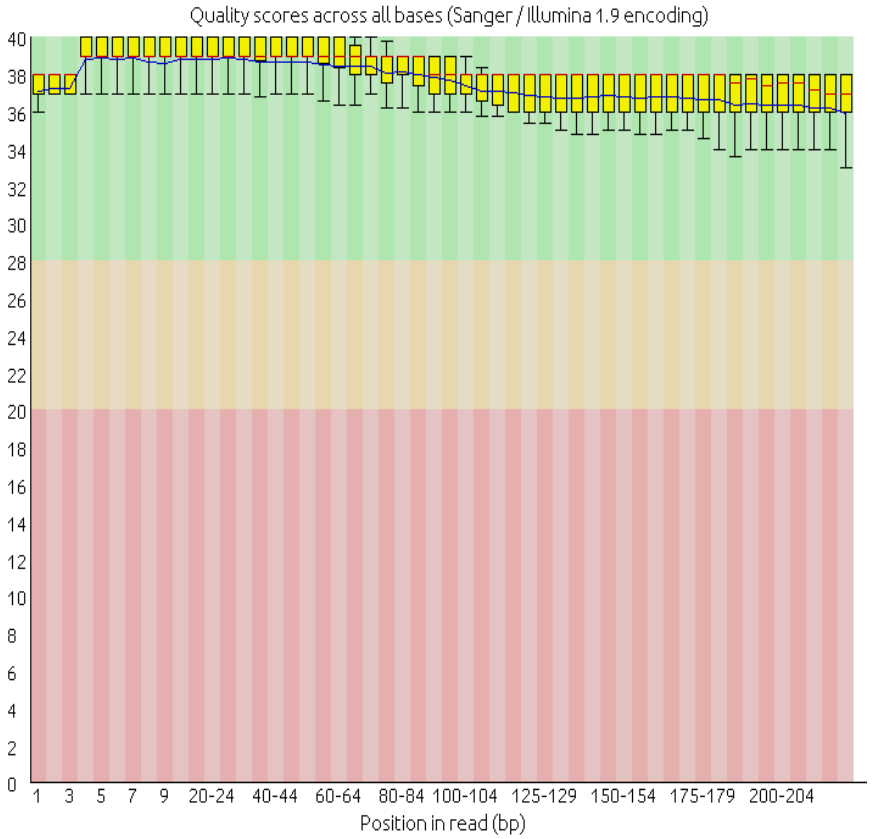


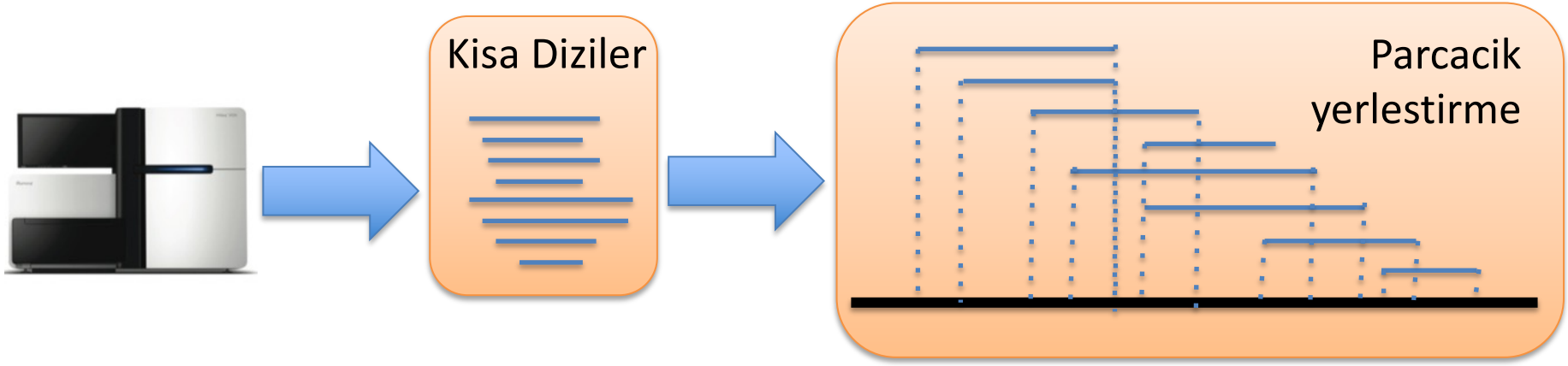
- Bastaki ve sondaki kotu kalite bazlari kes
- Aralardaki kotu kalite bazlari filtrele

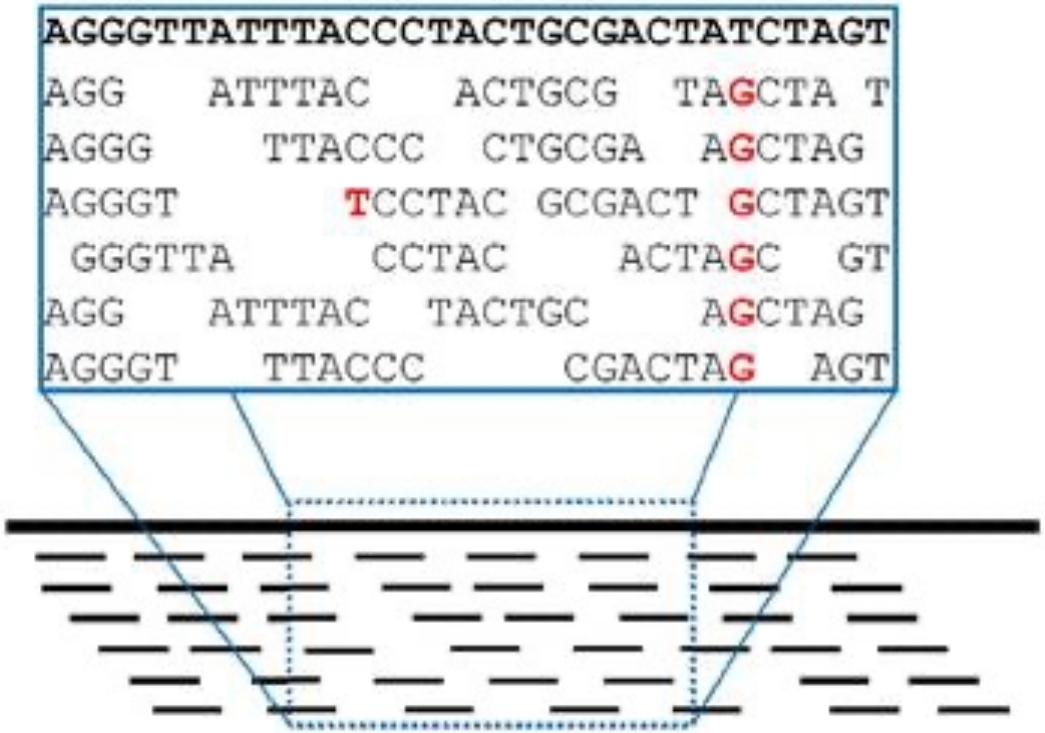
Oncesi



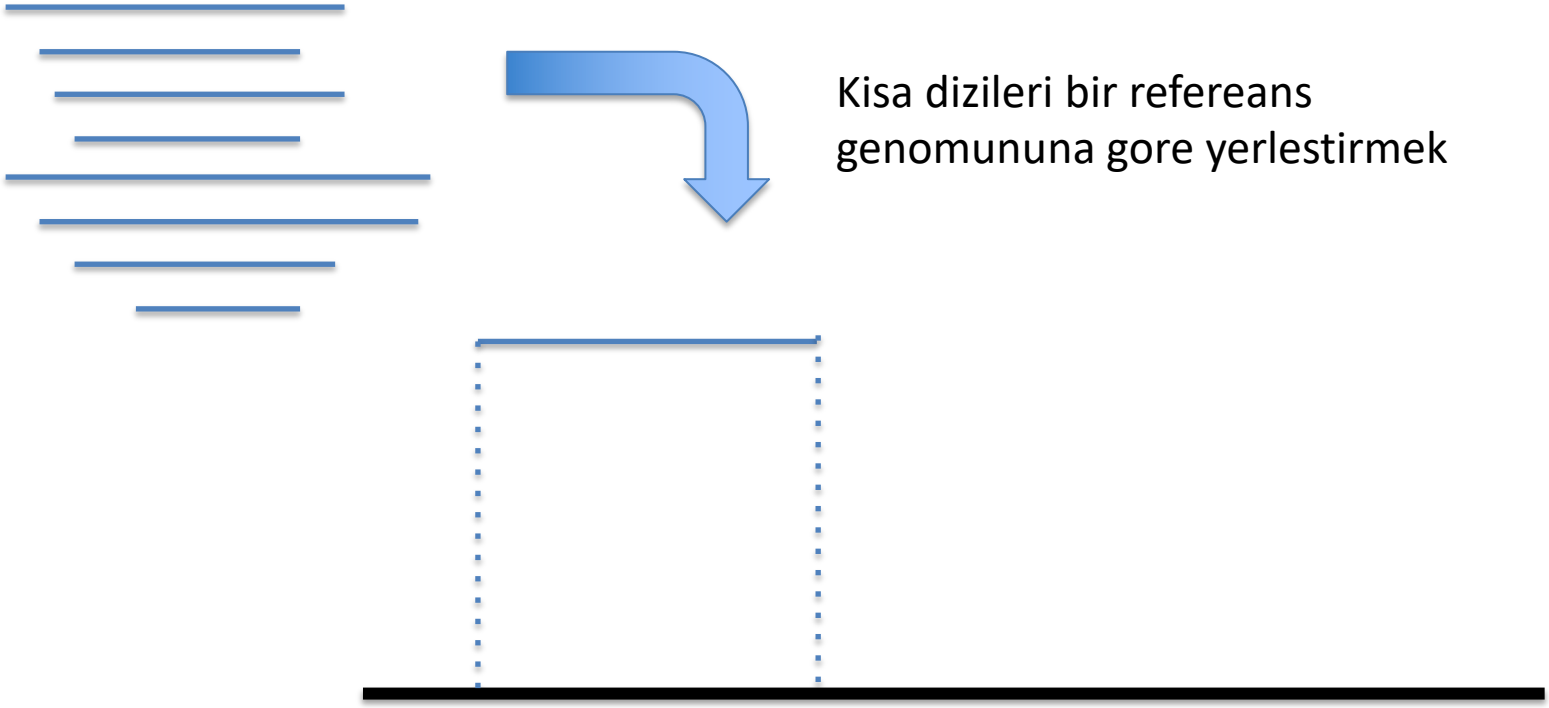
Sonrasi




Kisa Diziler
Parcacik
yerlestirme

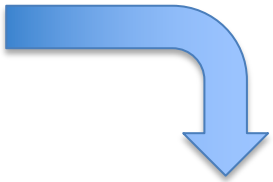

AGGGTTATTTACCCTACTGCGACTATCTAGT
AGG ATTTAC ACTGCG TAGCTA T
AGGG TTACCC CTGCGA AGCTAG
AGGGT TCCTAC GCGACT GCTAGT
GGGTTA CCTAC ACTAGC GT
AGG ATTTAC TACTGC AGCTAG
AGGGT TTACCC CGACTAG AGT

Parcacik Yerlestirme

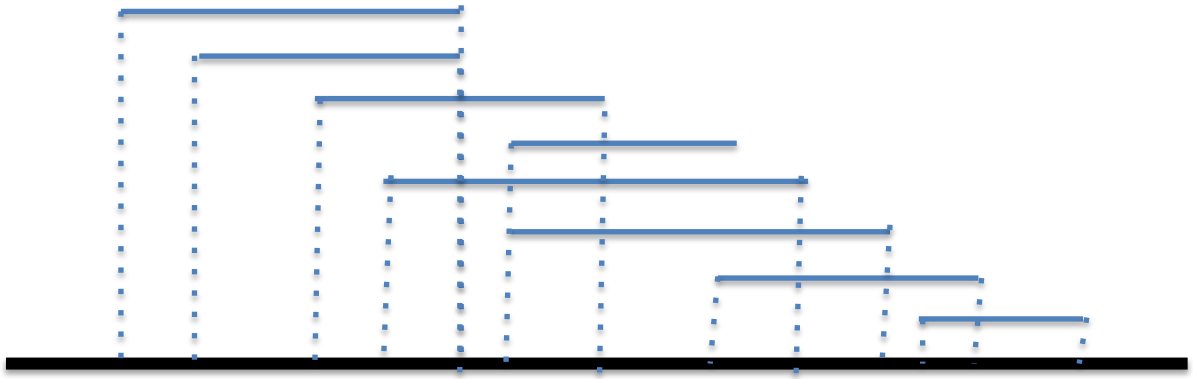


Kisa dizileri bir refereans genomununa gore yerlestirmek

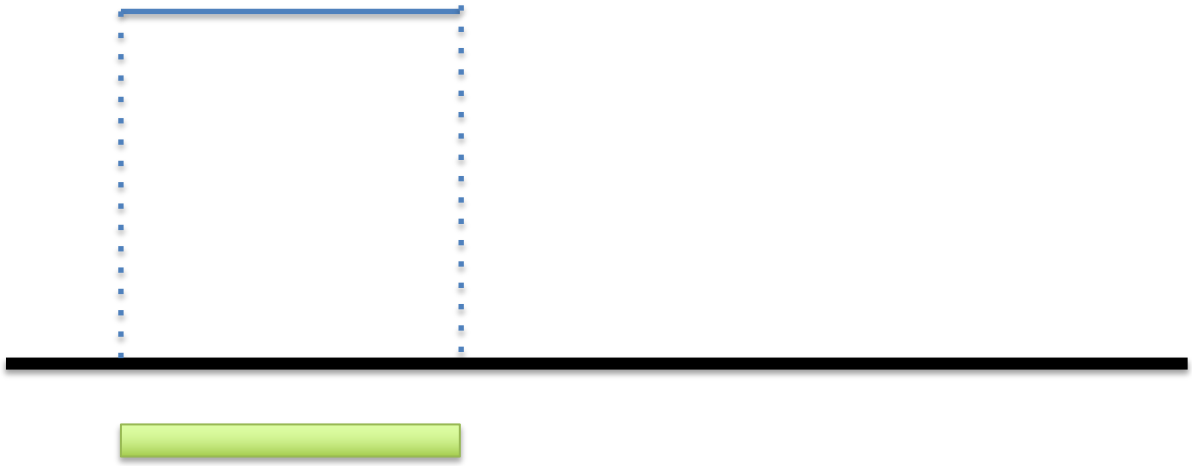
Parcacik Yerlestirme



Kisa dizileri bir refereans genomununa gore yerlestirmek

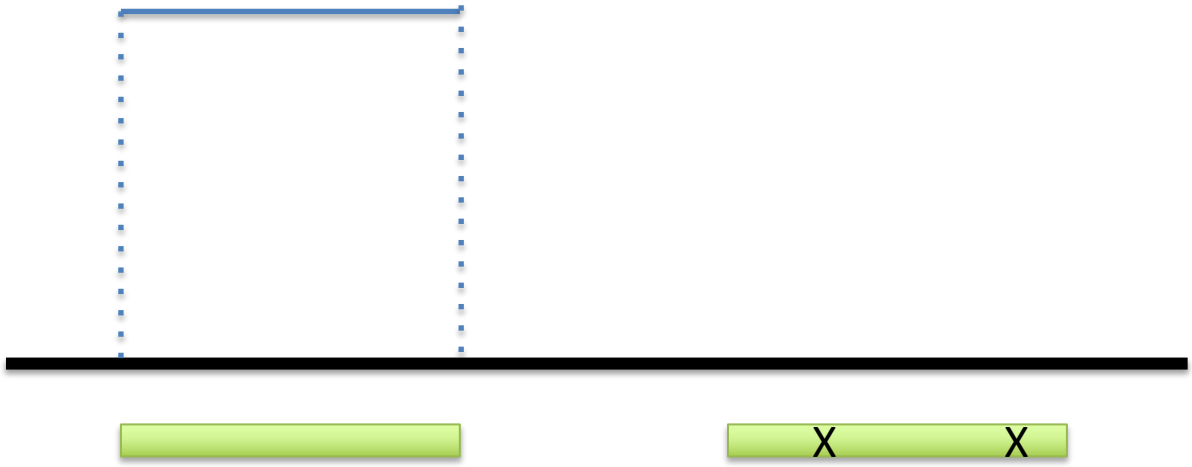


Parcacik Yerlestirme

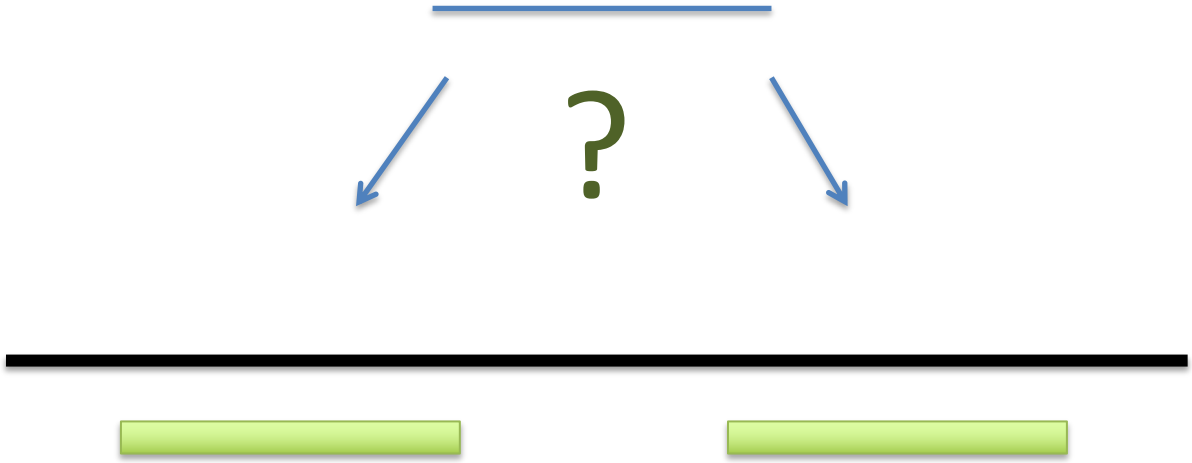


Parcacik Yerlestirme

Parcacik Yerlestirme

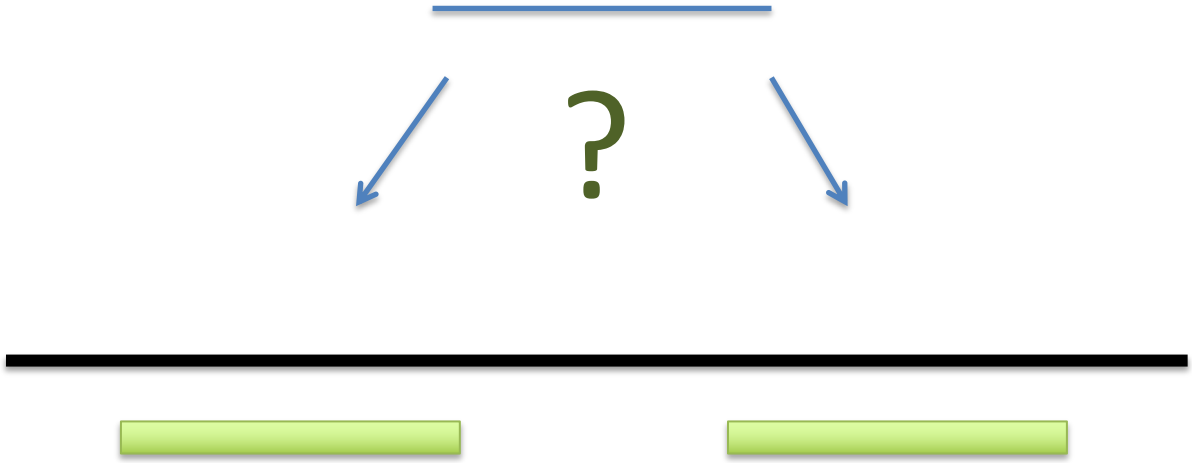


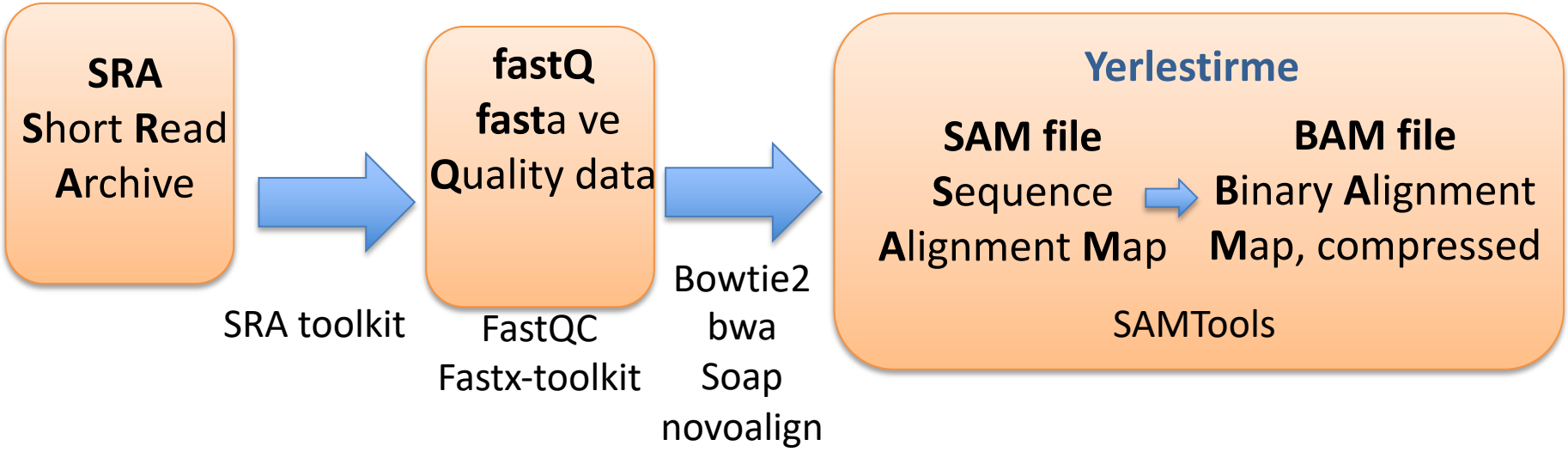
Parcacik Yerlestirme



Yerlestime Kalitesi

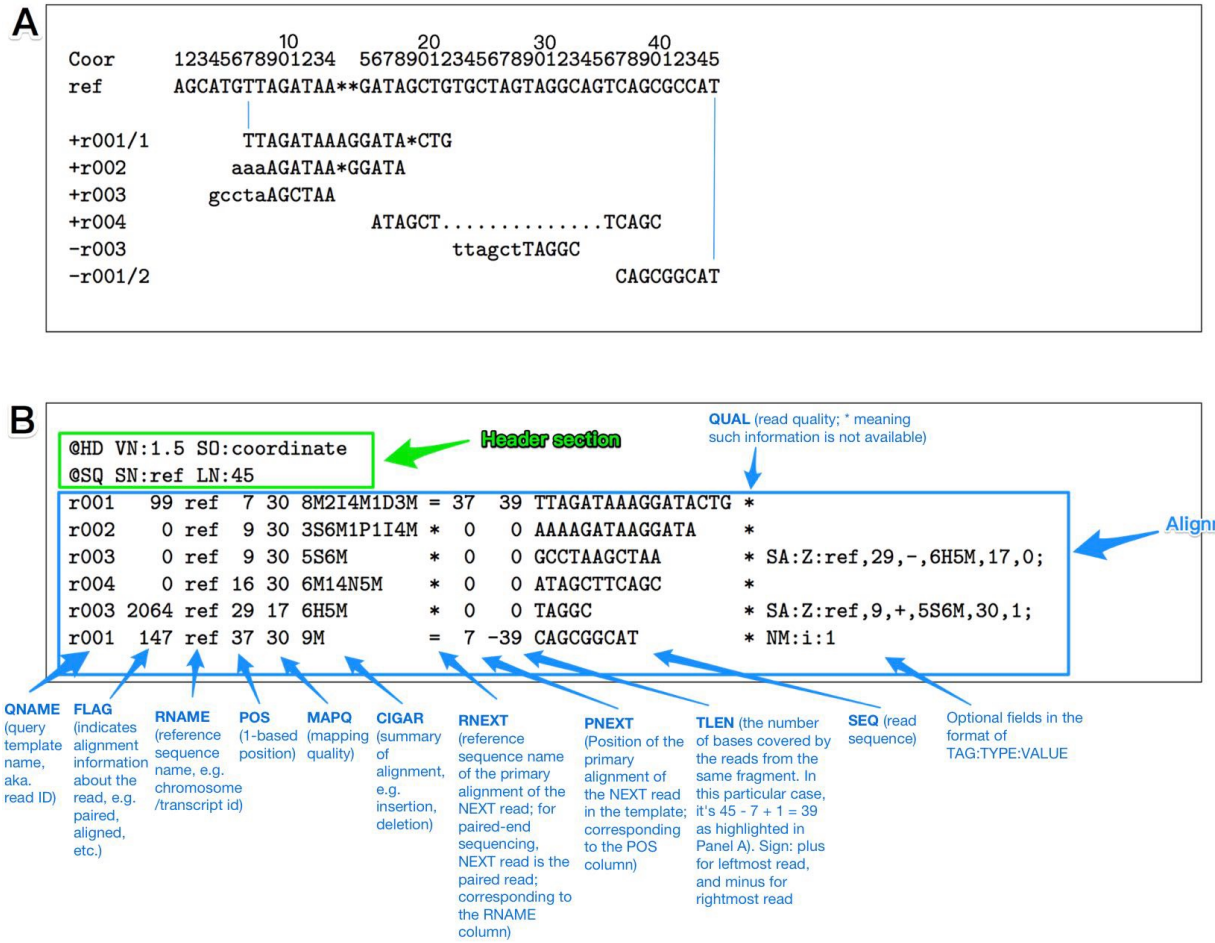
Bu parcacik genoma yerlestirilirken ne kadar kesin konusabiliriz?




SRA
Short Read
Archive
SRA toolkit
fastQ
fasta ve
Quality data
FastQC
Fastx-toolkit
Bowtie2
bwa
Soap
novoalign
Yerlestirme
SAM file
Sequence
Alignment Map
BAM file
Binary Alignment
Map, compressed
SAMTools

Sequence **A**ligment **M**ap (SAM) Dosyasi

Sizce ne kadar büyüktür bu dosyalar?

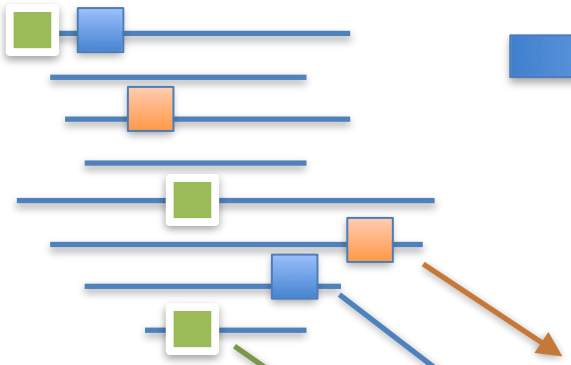


Binary **A**ligment **M**ap (BAM) Dosyasi

Bitwise Flags

Integer	Binary	Description (Paired Read Interpretation)
1	`000000000001`	template having multiple templates in sequencing (read is paired)
2	`000000000010`	each segment properly aligned according to the aligner (read mapped in proper pair)
4	`000000000100`	segment unmapped (read1 unmapped)
8	`000000001000`	next segment in the template unmapped (read2 unmapped)
16	`000000010000`	SEQ being reverse complemented (read1 reverse complemented)
32	`000000100000`	SEQ of the next segment in the template being reverse complemented (read2 reverse complemented)
64	`000001000000`	the first segment in the template (is read1)
128	`000010000000`	the last segment in the template (is read2)
256	`000100000000`	not primary alignment
512	`001000000000`	alignment fails quality checks
1024	`010000000000`	PCR or optical duplicate
2048	`100000000000`	supplementary alignment (e.g. aligner specific, could be a portion of a split read or a tied region)

Indexleme

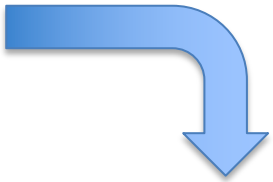


Her islem oncesinde kullandigimiz aracın kendi indeksleme komutuyla dizilerimizi indeksliyoruz.

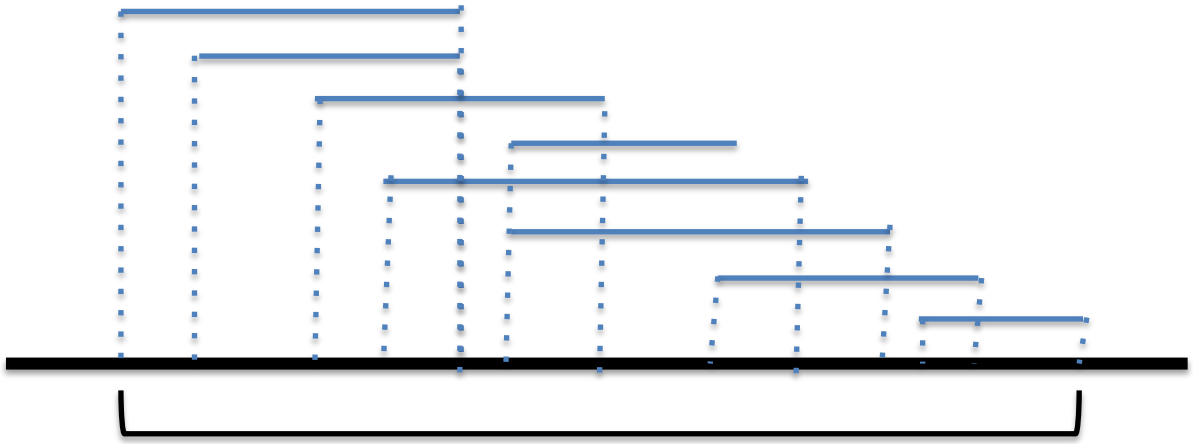
Kok	Pozisyon
AAAA	0,8
ACCT	4,16,24
GTGA	12,20
…	…

Bunun amaci okumaya hizlandirmak. Kitabin icindekiler kısmı gibi dusunebiliriz.

Parcacik Yerlestirme

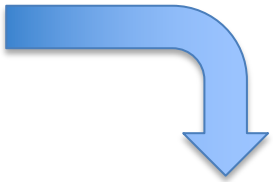


Kisa dizileri bir refereans genomununa gore yerlestirmek

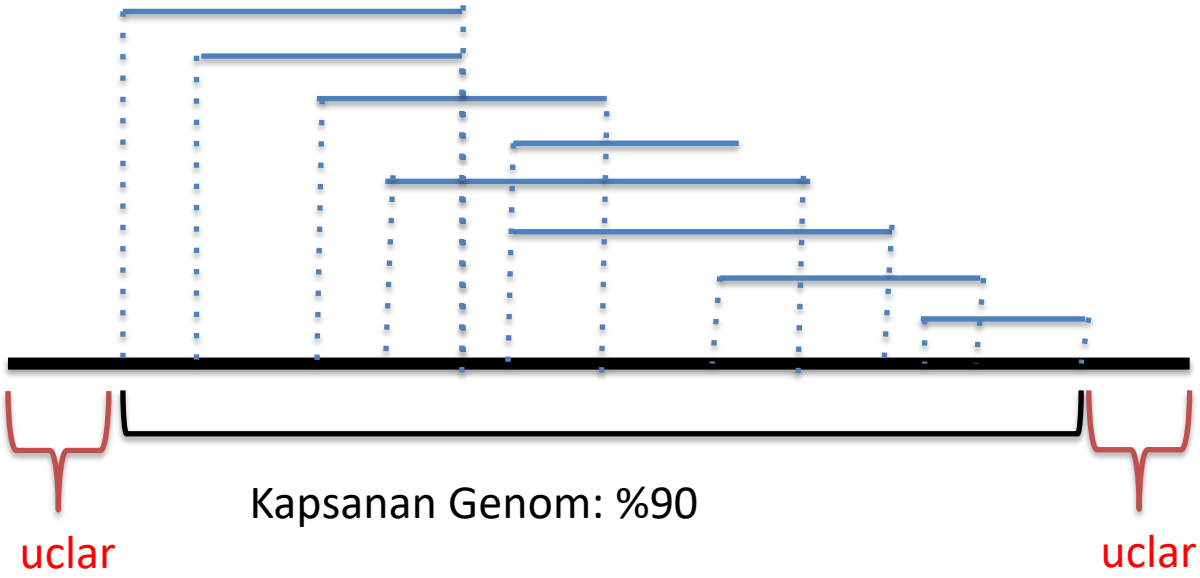


Kapsanan Genom: %90

Parcacik Yerlestirme



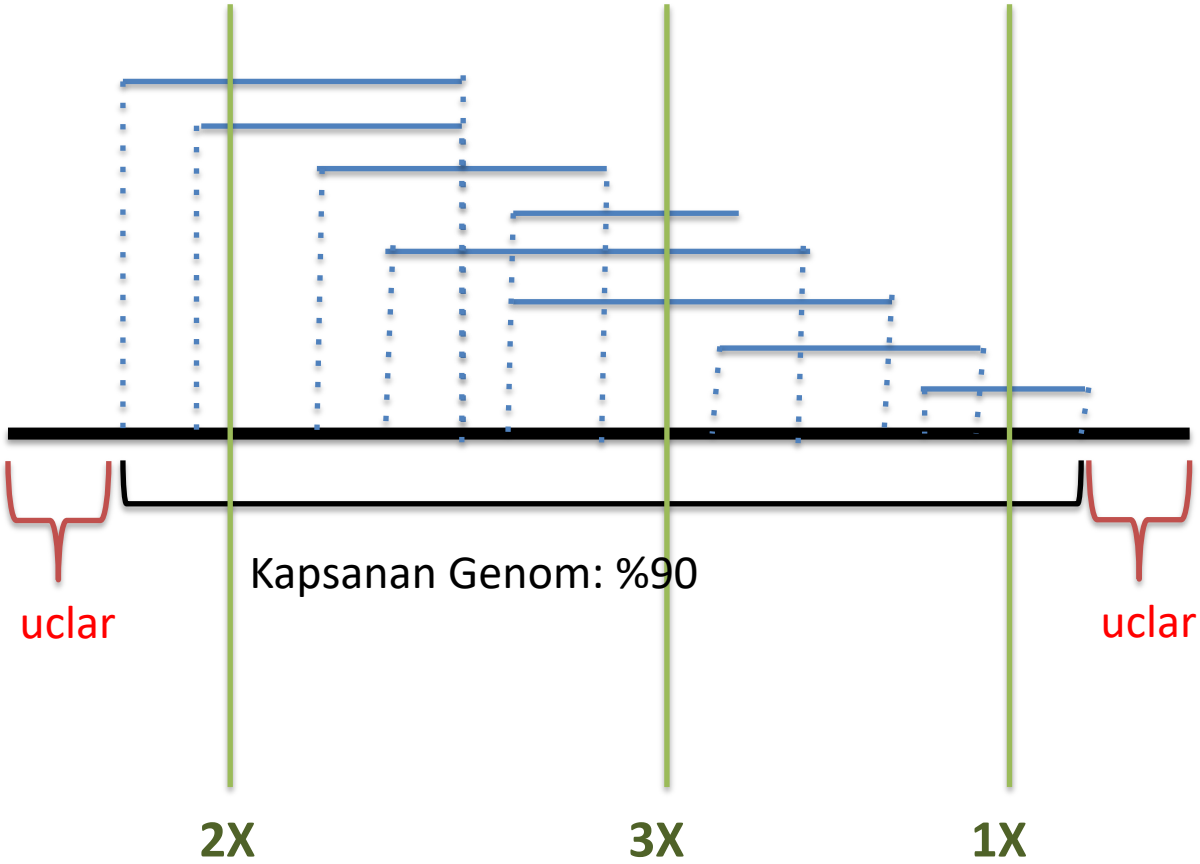
Kisa dizileri bir refereans genomununa gore yerlestirmek



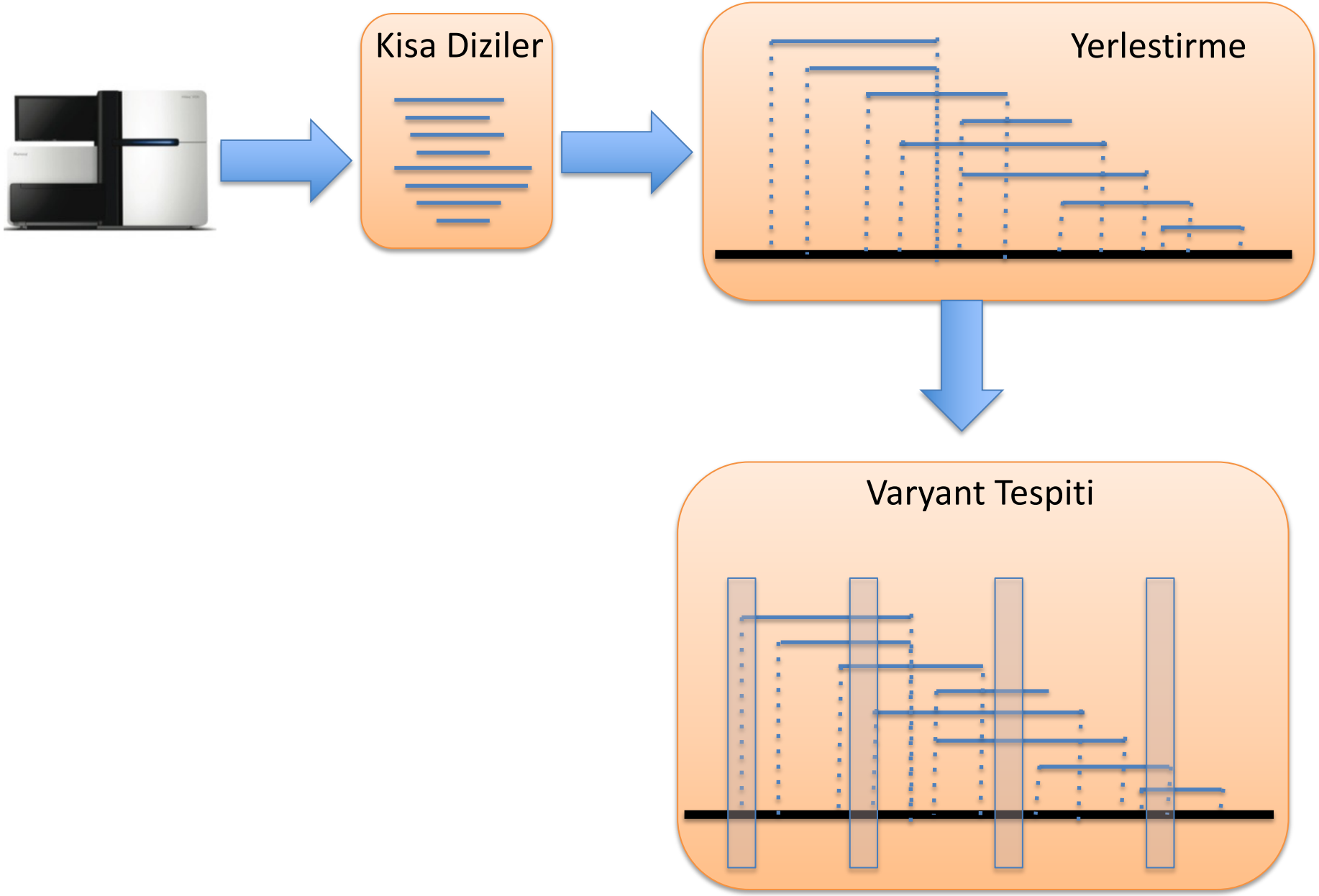
Parcacik Yerlestirme

Kisa dizileri bir refereans genomununa gore yerlestirmek

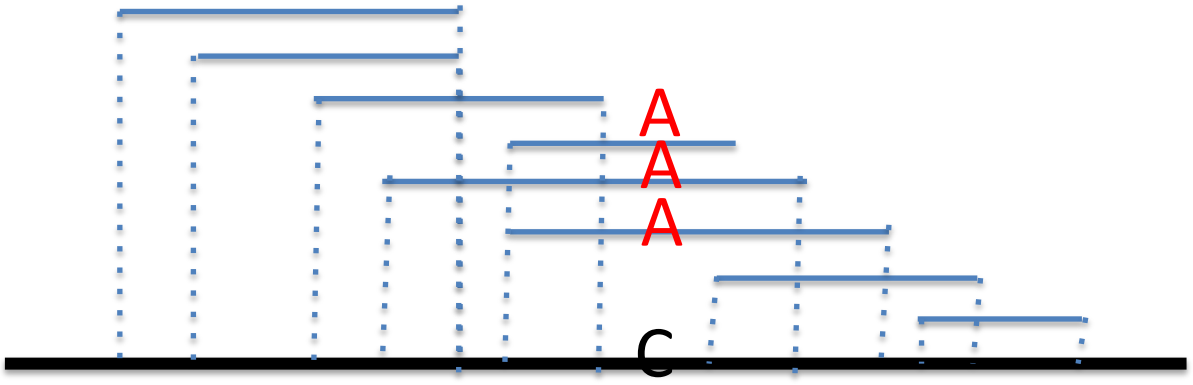
Dizi Derinligi– bir baz kac kere yerlestiildi



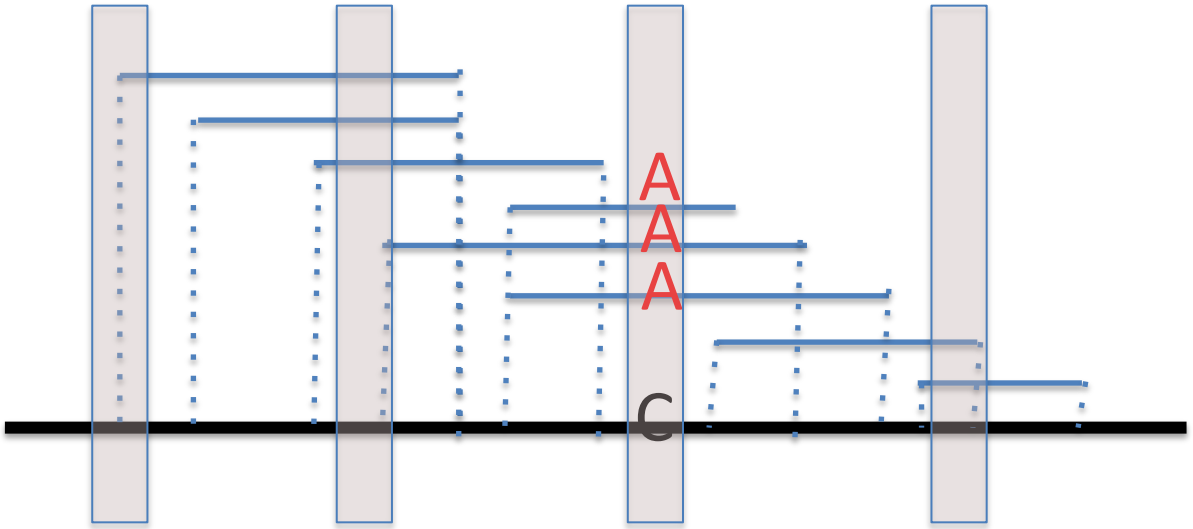
Klonlari temizle


Kisa Diziler
Yerlestirme
Varyant Tespiti

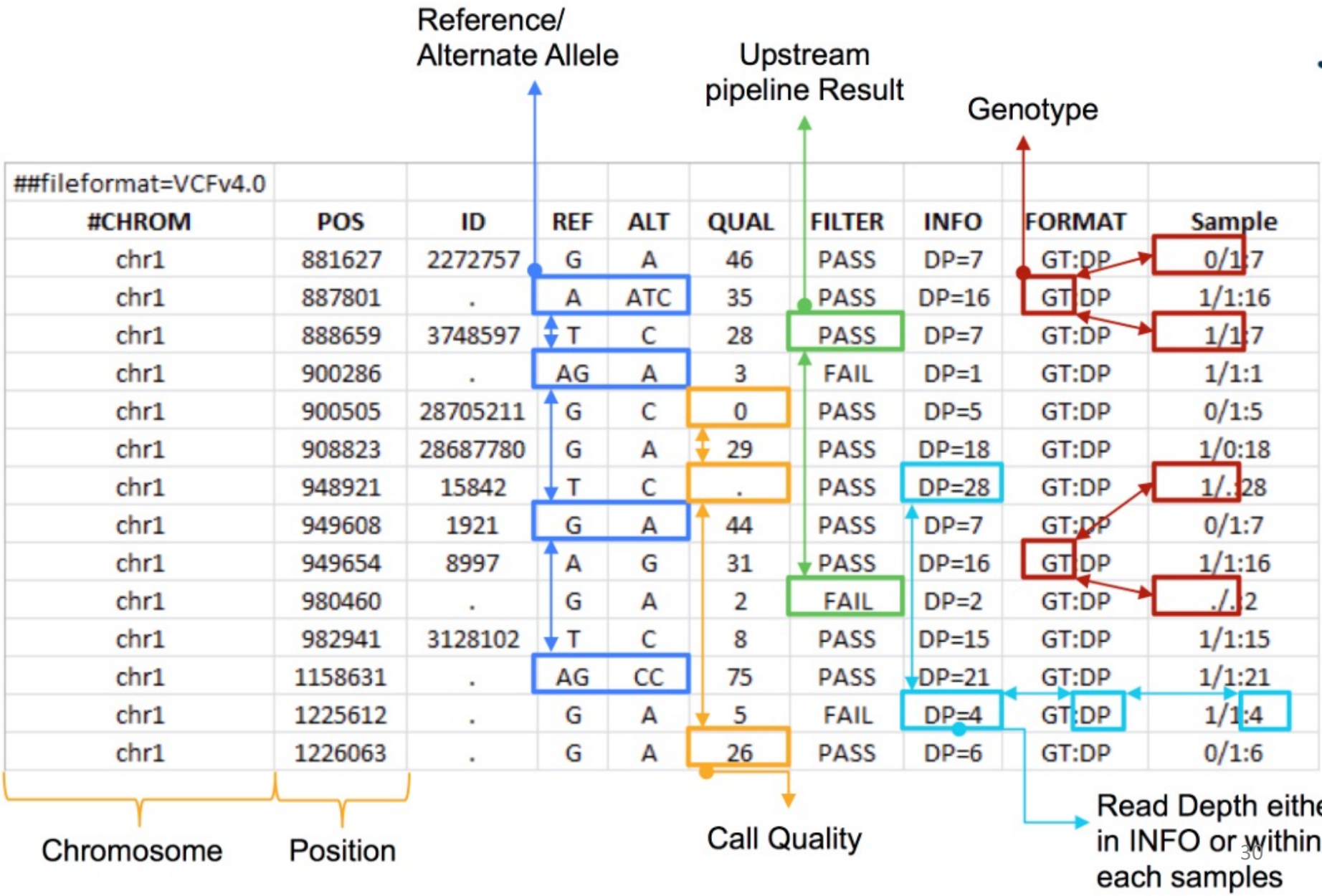
Varyant Tespiti

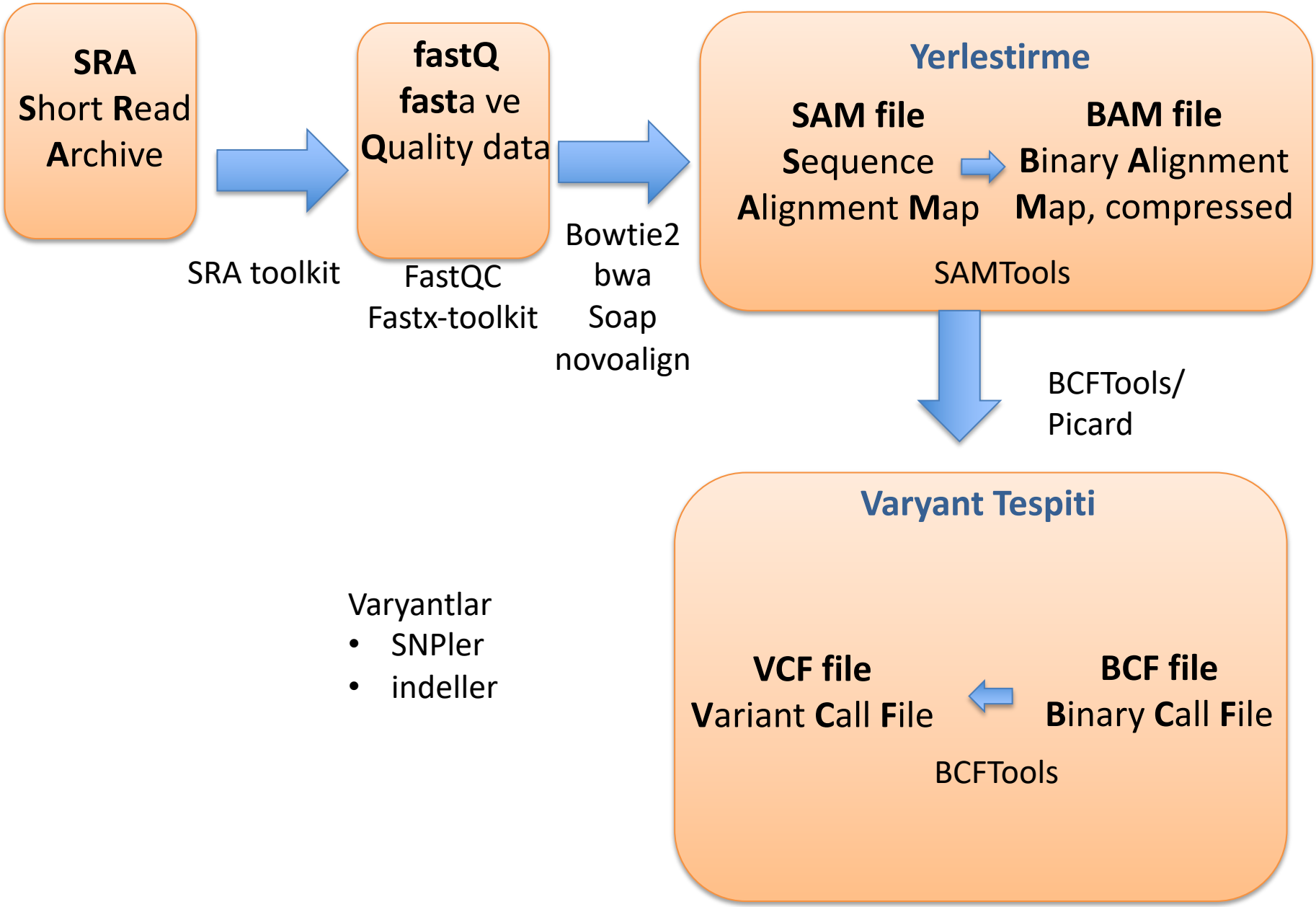


Varyant Tespiti



Variant Call File Format




SRA
Short Read
Archive
SRA toolkit
fastQ
fasta ve
Quality data
FastQC
Fastx-toolkit
Bowtie2
bwa
Soap
novoalign
Yerlestirme
SAM file
Sequence
Alignment Map
BAM file
Binary Alignment
Map, compressed
SAMTools
BCFTools/
Picard
Varyant Tespiti
VCF file
Variant Call File
BCF file
Binary Call File
BCFTools
Varyantlar
• SNPler
• indeller

Websiteleri

SAMtools: http://www.htslib.org

Picard tools: https://broadinstitute.github.io/picard/

BWA: http://bio-bwa.sourceforge.net

Bowtie: http://bowtie-bio.sourceforge.net/index.shtml

SOAP: http://soap.genomics.org.cn/index.html

Novoalign: http://www.novocraft.com/products/novoalign/

FASTX-toolkit: http://hannonlab.cshl.edu/fastx_toolkit/

FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

- 156 kisinin mitokondriyal DNAsi
- Agizdan veya kandan ornek
- 20,000X dizi derinligi

CrossMark
click for updates

Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA

Boris Rebolledo-Jaramillo[a,1], Marcia Shu-Wei Su[b,1], Nicholas Stoler[a], Jennifer A. McElhoe[c], Benjamin Dickins[d], Daniel Blankenberg[a], Thorfinn S. Korneliussen[e,f], Francesca Chiaromonte[g], Rasmus Nielsen[e], Mitchell M. Holland[c], Ian M. Paul[h], Anton Nekrutenko[a,2], and Kateryna D. Makova[b,2]

Departments of [a]Biochemistry and Molecular Biology, [b]Biology, and [g]Statistics, [c]Forensic Science Program, Pennsylvania State University, University Park, PA 16802; [d]School of Science and Technology, Nottingham Trent University, Nottingham NG1 4BU, United Kingdom; [e]Department of Integrative Biology, University of California, Berkeley, CA 94720; [f]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark; and [h]Department of Pediatrics, College of Medicine, Pennsylvania State University, Hershey, PA 17033

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved September 8, 2014 (received for review May 20, 2014)