

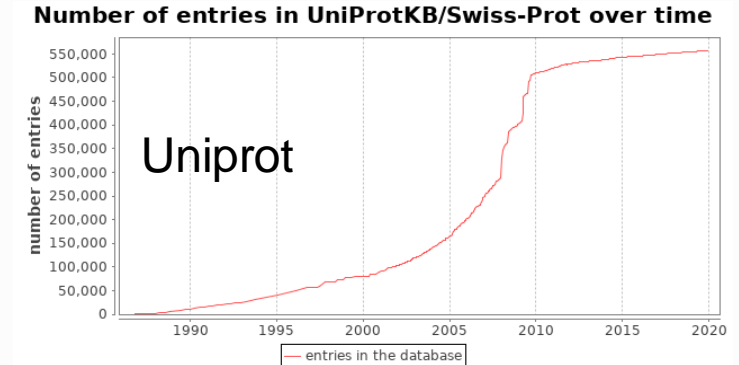
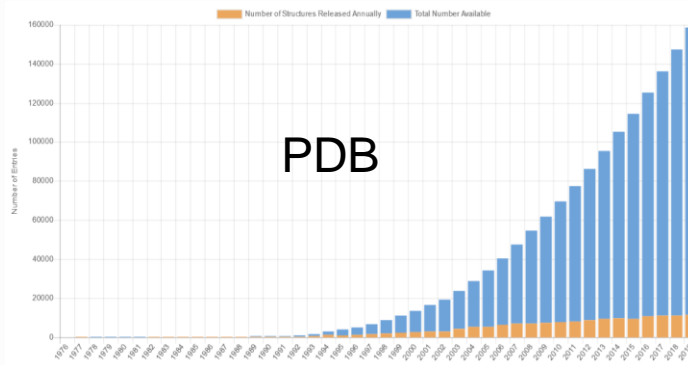
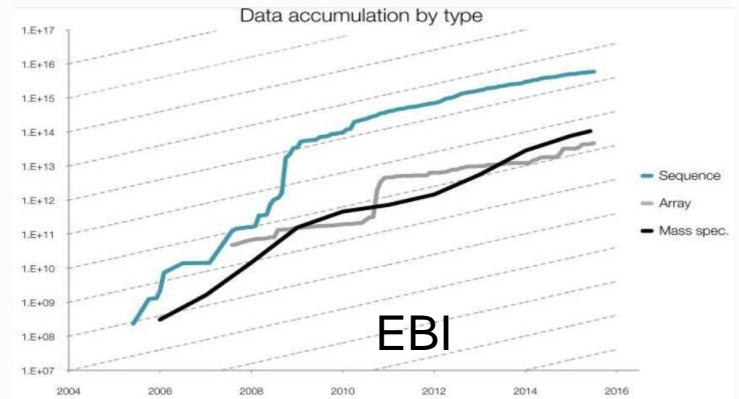
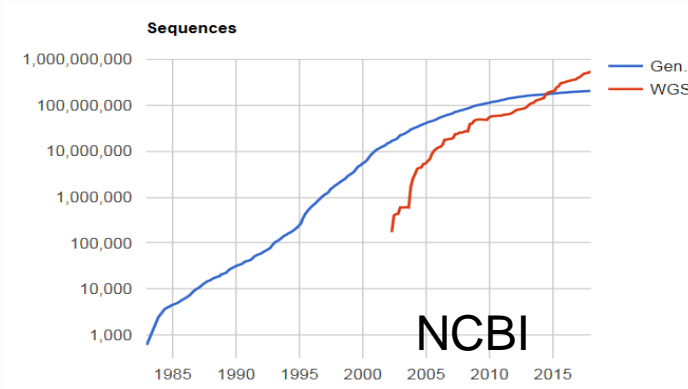


UNIVERSITY OF
GOTHENBURG

BMA Molekylärbiologisk metodik Bioinformatik

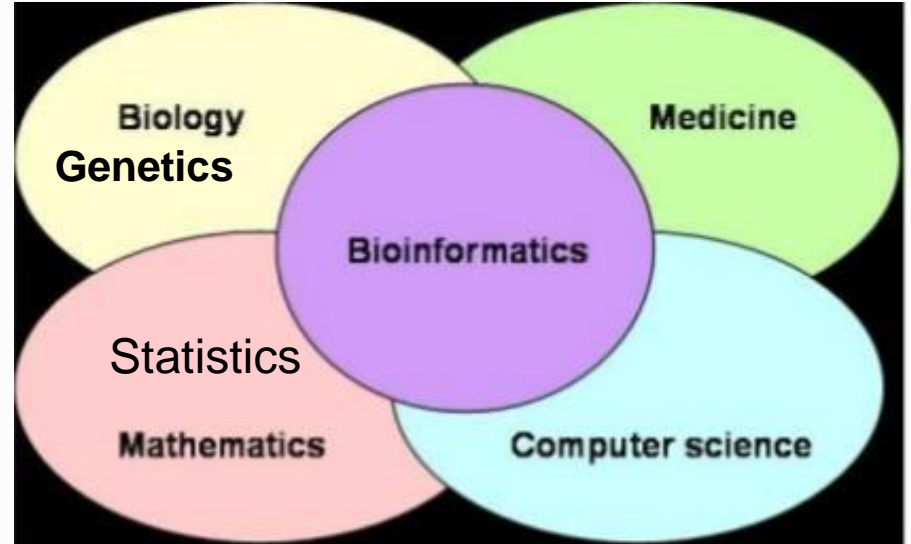
Mars/2021

Explosiv tillväxt av sekvensinformation



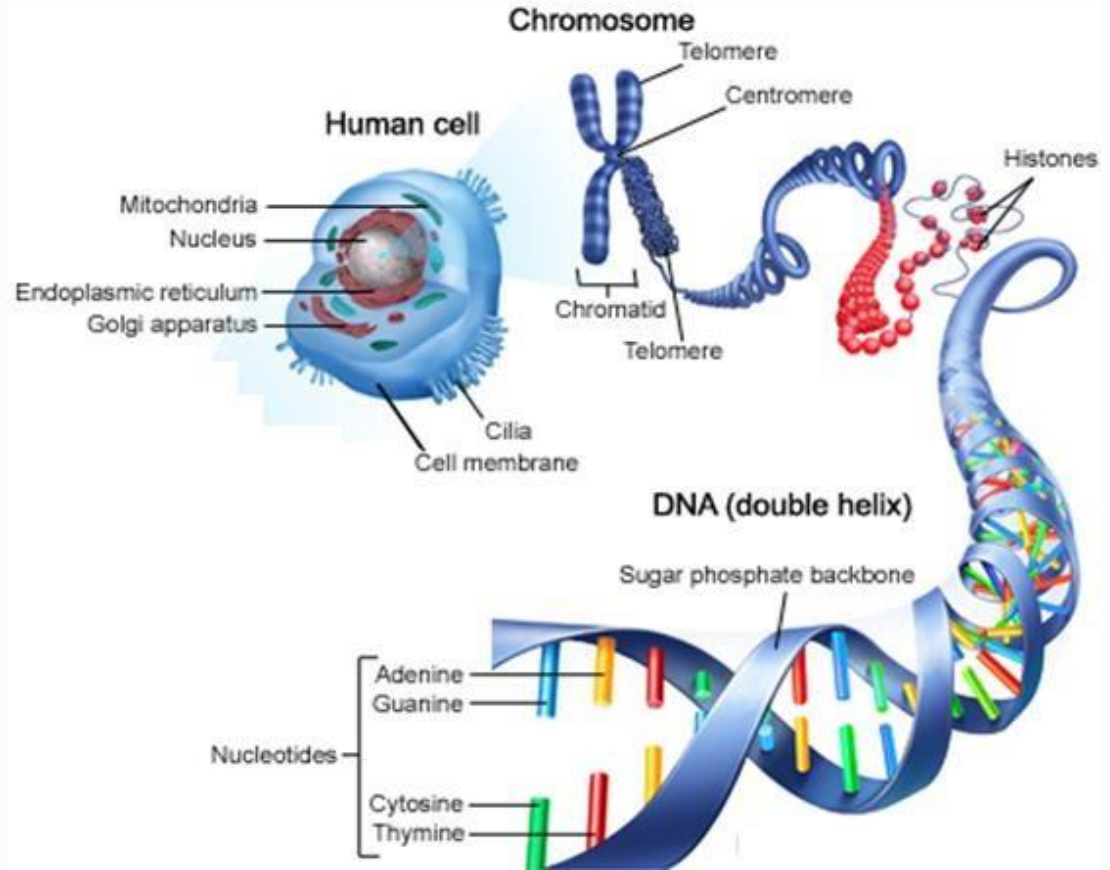
Bioinformatik

Hantering och analys
av information från
genteknologisk
forskning.
Tvärvetenskaplig
disciplin

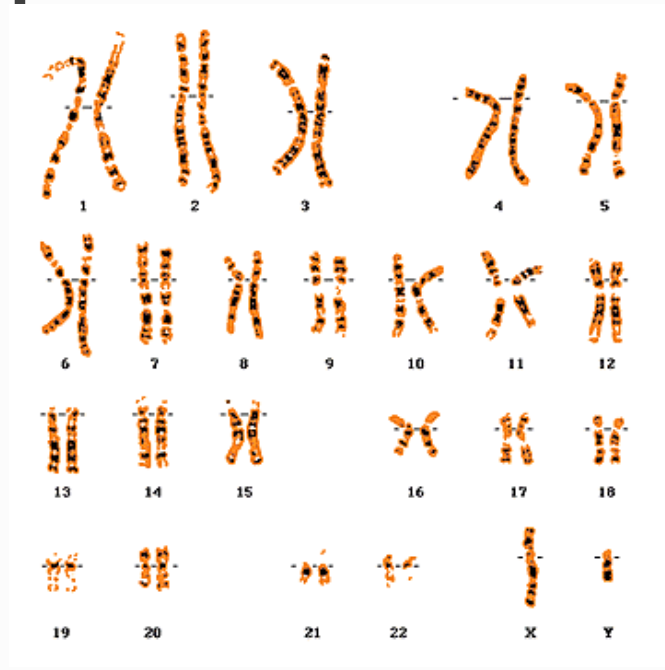


Genom

Genetiskt material
som bestämmer
de ärvda
egenskaperna för
en organism



Genomet är uppdelat i



22 autosomala par & 1 allosom par
Diploid

Species	# Chr	Ploidy
Thale cress	10	Diploid
Bread wheat	42	Hexaploid
Tobacco	48	Tetraploid
Fruit fly	8	Diploid
Earth worm	36	Diploid
Mouse	40	Diploid
Human	46 (2x23)	Diploid
Dog	78	Diploid
Goldfish	100-104	Diploid

Jumper jack



XY -> 1chr
XX -> 2 chr



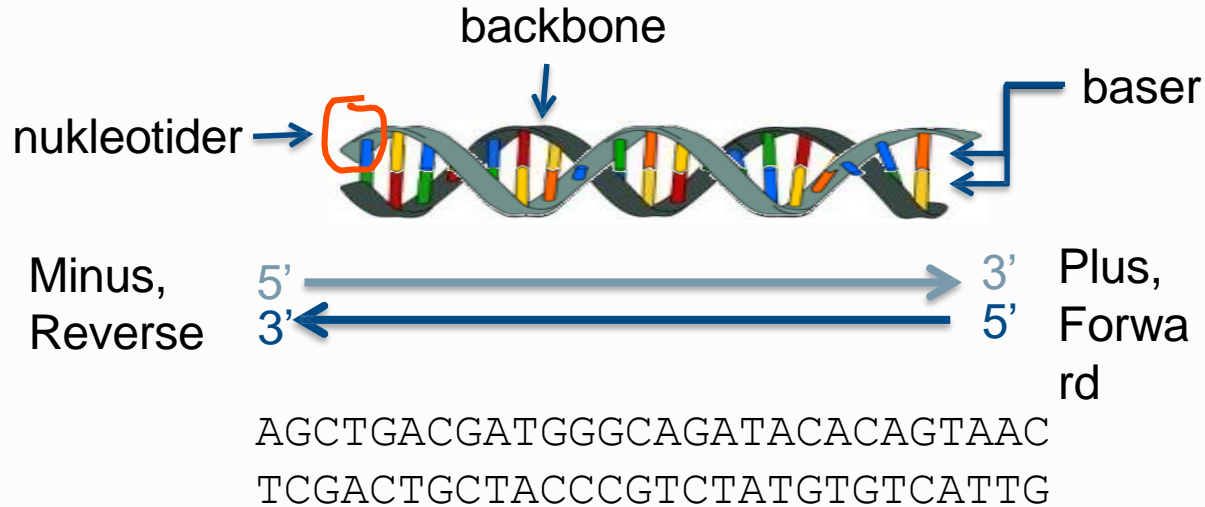
Adder's Tongue
1260 chr

Andra

Type	Organism	Genome size
Virus	P. circovirus	1.8Kb
Virus	Megavirus	48Kb
Bacterium	N. deltocephalinicola	112Kb
Bacterium	E. coli	4.6Mb
Plant	G. margaretae	63Mb
Plant	P. japonica	150GB
Moss	P. patens	480Mb
Yeast	S. cerevisiae	12.1Mb
Nematode	C. elegans	100Mb
Insect	D. melanogaster	130Mb
Fish	T. nigroviridis	390Mb
Fish	P. aethiopicus	130Gb
Mammal	M. musculus	2.7Gb
Mammal	H. sapiens	3.2Gb



DNA



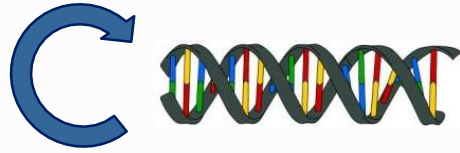
Gener kan ligga på båda strängarna. mRNA sekvensen motsvarar alltid 5- 3' kodningen av en gen

- Dubbelsträngat
- Antiparallellt
- Komplementärt

A=T adenine=thymine

C≡G cytosine≡guanine

Central Dogma: Flödet av genetisk information



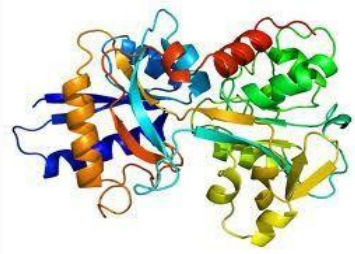
Replikation: DNA till DNA



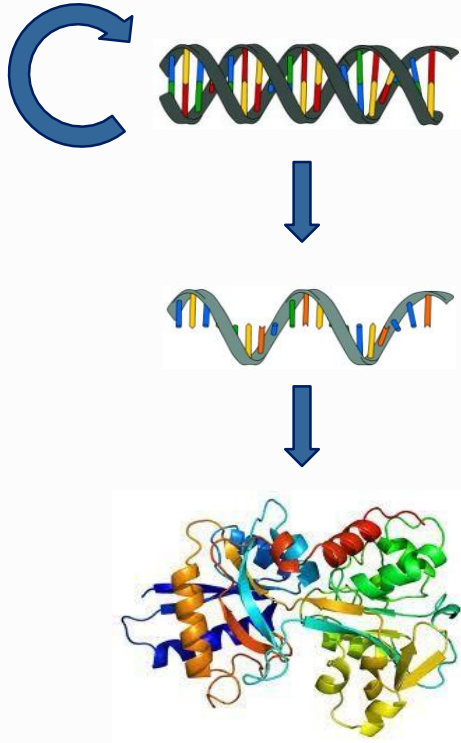
Transkription: DNA till mRNA



Translation: mRNA till protein



Från mRNA till protein



5' AGTACCATACTTCTGGCCTGGGCGAAAGAATAA 3'
3' TCATGGTATGAAGACCGGACCCGCTTTCTTATT 5'

5' AGUACCAUACUUCUGGCCUGGGCGAAAGAAUAA 3'

S T I L L A W A K E *

CODON: 3 baser av ett reads som läses
av tRNA för att bli translaterat till en
aminosyra

Den genetiska koden

UUU Phe F UUC Phe F UUA Leu L UUG Leu L	UCU Ser S UCC Ser S UCA Ser S UCG Ser S	UAU Tyr Y UAC Tyr Y UAA Stop UAG Stop	UGU Cys C UGU Cys C UGA Stop UGG Trp W
CUU Leu L CUC Leu L CUA Leu L CUG Leu L	CCU Pro P CCC Pro P CCA Pro P CCG Pro P	CAU His H CAC His H CAA Gln Q CAG Gln Q	CGU Arg R CGC Arg R CGA Arg R CGG Arg R
AUU Ile I AUC Ile I AUA Ile I AUG Met M	ACU Thr T ACC Thr T ACA Thr T ACG Thr T	AAU Asn N AAC Asn N AAA Lys K AAG Lys K	AGU Ser S AGC Ser S AGC Arg R AGG Arg R
GUU Val V GUC Val V GUA Val V GUG Val V	GCU Ala A GCC Ala A GCA Ala A GCG Ala A	GAU Asp D GAC Asp D GAA Glu E GAG Glu E	GGU Gly G GGC Gly G GGA Gly G GGG Gly G

Translation

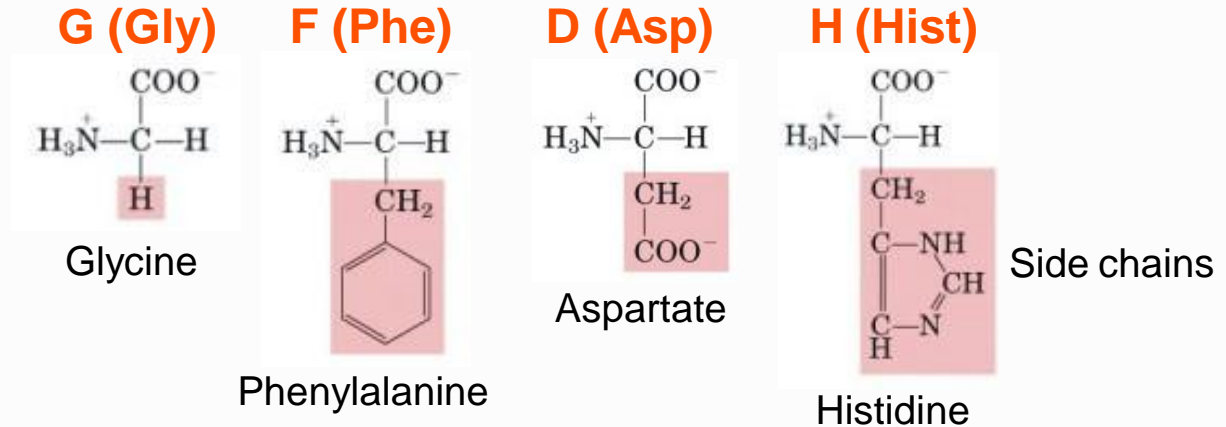
- Diff. nt sekvenser kan översättas till identiska aa sekvenser
- Substitutioner, deletioner, insertioner

AUG	UUG	GGU	UGA
M	L	G	*
AUG	CUA	GGA	UAA
M	L	G	*

AUG	UUC	GGU	UGA
M	F	G	*
AUG	UAG	GGU	UGA
M	*	G	*
AUG	UGG	GUU	GAU
M	W	V	N
AUG	CUU	GGG	UUG
M	L	G	L

Proteiner

Lång kedja av aminosyror som är nödvändiga för levande celler

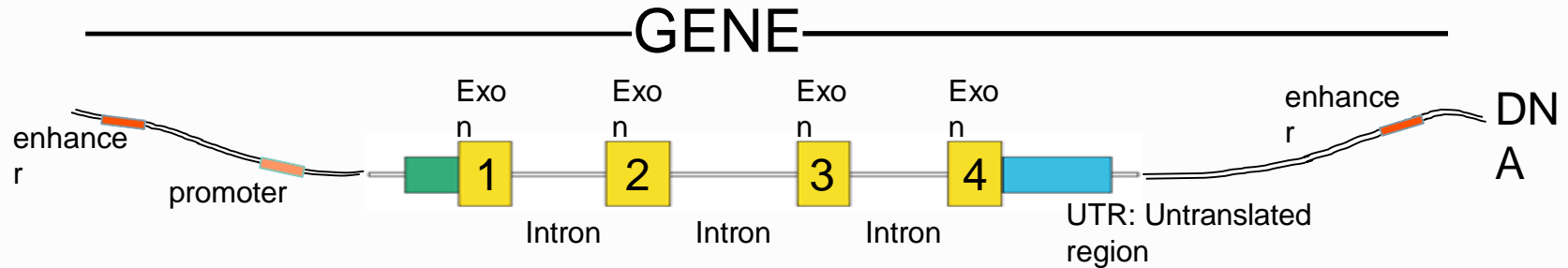


20 standard aminosyror

Olika cellulära funktioner: Regulering, struktur, rörelse, katalys, transport, signalering

Gener

Delen av DNAt som innehåller all information för att producera ett protein.

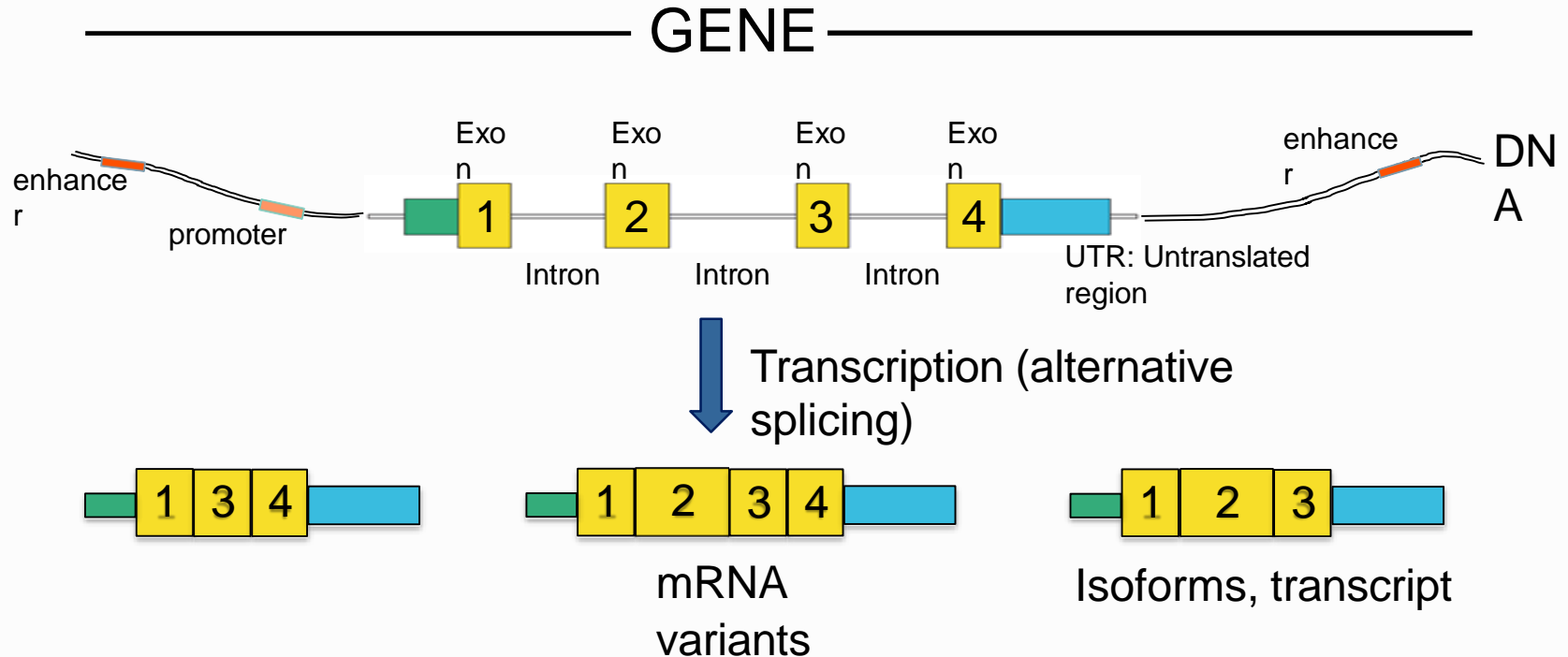


20,000 protein-kodande gener (~1.5%)

Ojämnt distruberade över alla kromosomer

HIST1H1A	781 nt	brist på introner
DMD	2.2 Mb (79 exons)	största genen
TTN	80 Kb (364 exons)	längsta kodande sekvensen

Kodande sekvensen



Komperativ genomik

Är vi verkligen så olika?



```

M.sp          UGCAUAAACAA AGGACAGAACTUUUGCCA . AUCCCCUUGACAAAGGAC. GGAUU. . GAGC. UUUUGCUCaaacU. . UA. CA. . . . . CC. . . . . A. CCU. C. UGUUUGA. .
R.oryzae      GGUUAAUAACAA AGGACAGAACTUUUGCCA . AUCCCCUUGACAAAGGAC. GGAUU. . GAGC. UUUUGCUC. . . . . AA. CA. UA. . . . . A. CAC. CGU. G. UGUUUGA. .
P.blakesleeanus AACAUUAACAA AGGACAGAACTUUUGCCA . AUCCCCUUGACAAAGGAC. GGAAC. . GAGC. UUUUGCUC. . . . . A. CA. CU. . . . . A. ACA. CCU. C. GGUUUGA. .
H.sapiens     UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
G.gallus      UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
D.rerio       UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
T.nigroviridis UUUUGUAUCAA AGGACAGAACTUUUGCCA . AUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
F.rubripes    -UUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
B.floridiae   UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. CCCC. CUUGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
X.laueis      UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
C.intestinalis UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGA. GAGC. CUCGGGCC. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
S.purpuratus  UUUUGUAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. CACC. CUCGGGCC. . . . . A. CA. . . . . A. CAC. AUA. C. GGUUA. .
T.spiralis    UUGGCAUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
A.aegypti     UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
A.gambiae     UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
A.mellifera    UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
B.moris       UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
D.melanogaster UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
D.pseudobscura UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
T.castaneum    UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
D.pulex        UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
T.adhaerens    UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
H.magnipapillata UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
N.vectensis    UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
R.sp           UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
P.polycephalum UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
A.castellani   UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
A.thaliana     UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
O.sativa       UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
S.moeilendorffii UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
P.patens       UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
H.parasitica   UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
P.sojae        UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
P.ramorum      UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
P.infestans    UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .
..            UGUUUUUUCAA AGGACAGAACTUUUGCCA . CUCCCCUUGACAAAGGAC. GGAAGU. . . . . UA. UA. CA. . . . . A. CAC. AUA. C. GGUUA. .

```

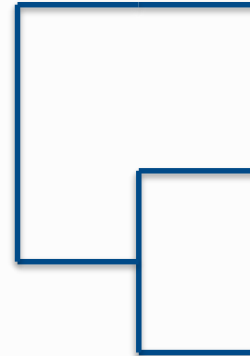
Multiple sequence alignment

```

quagga  GGATTCGTTCACTGATTCCCTCTATTCTCAGGATACACACTCAACCAAACCTGAGCAAAA
zebra   GGATTCGTTCACTGATTCCCTCTATTCTCAGGATACACACTCAACCAAACCTGAGCAAAA
horse   GGATTTGTACACTGATTCCCTCTATTCTCAGGATACACACTCAACCAAACCTGAGCAAAA
***** ** *****

quagga  ATTCACTTTACAATTATATTCGTAGGGGTCAACATAATTTTCTTCCCA
zebra   ATTCACTTTACAATTATATTCGTAGGGGTCAATATAACTTTCTTCCCA
horse   ATCCACTTTACAATTATATTCGTAGGGGTAAATATAACTTTCTTCCCA
-       ** ***** ** *****

```



Viktiga data repos

Nukleotid:

GenBank - NCBI

EMBL - EBI

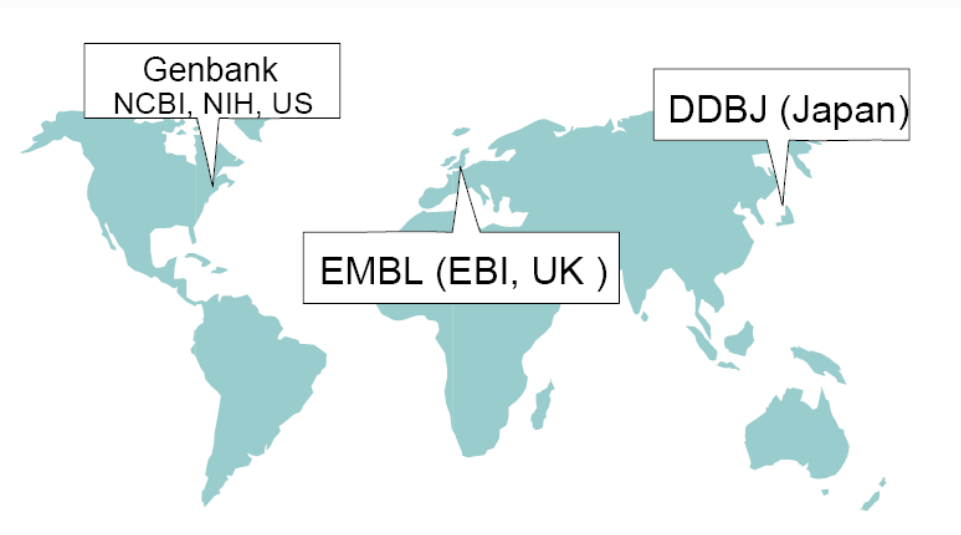
DDBJ - NIG

Protein:

UniProt – SIB, EBI and PIR

3D strukturer

PDB - RCSB



Ensembl

MGI, SGD, PlasmoDB,

JGI, ... TCGA, GEO, 1000GP,

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

Welcome to NCBI

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

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

Submit

Deposit data or manuscripts
into NCBI 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

BLAST+ 2.10.0 now available with
improved composition-based statistics

18 Dec 2019

The BLAST+ 2.10.0 release is now
available from our FTP site. The new

Genome Workbench is now in the cloud!

17 Dec 2019

If you're interested in visualizing and
analyzing genomic data, then you'll want
to check out a new way to run Genome

BLAST –Basic Local Alignment Search Tool

Ett verktyg som används
för att jämföra
sekvensinformation.

Klistra in en sekvens och
hitta liknande sekvenser

BLASTP program

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

[Clear](#) **Query subrange**

From

To

Or, upload file No file selected.

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism
Optional Enter organism name or id—completions will be suggested ☐ exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude
Optional ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental s

Program Selection

Algorithm

- ☐ Quick BLASTP (Accelerated protein-protein BLAST)
- ☒ blastp (protein-protein BLAST)
- ☐ PSI-BLAST (Position-Specific Iterated BLAST)
- ☐ PHI-BLAST (Pattern Hit Initiated BLAST)
- ☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search **database nr** using **Blastp (protein-protein BLAST)**

☐ Show results in a new window

BLAST

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show

100



☒ select all 100 sequences selected

[GenPept](#)

[Graphics](#)

[Distance tree of results](#)

[Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	ferritin, heavy polypeptide 1, gene 1 [Xenopus tropicalis]	368	368	100%	7e-129	100.00%	NP_989008.1
<input checked="" type="checkbox"/>	hypothetical protein XENTR_v10019363 [Xenopus tropicalis]	360	360	100%	1e-125	95.65%	KAE8593888.1
<input checked="" type="checkbox"/>	hypothetical protein XELAEV_18037841mg [Xenopus laevis]	358	358	100%	3e-125	97.16%	OCT70916.1
<input checked="" type="checkbox"/>	ferritin heavy chain B [Xenopus laevis]	357	357	100%	1e-124	96.59%	NP_001079580.1
<input checked="" type="checkbox"/>	ferritin heavy chain 1 a [Xenopus laevis]	356	356	100%	3e-124	96.02%	NP_001090207.1
<input checked="" type="checkbox"/>	ferritin heavy chain A [Xenopus laevis]	355	355	100%	1e-123	96.02%	NP_001084057.1
<input checked="" type="checkbox"/>	unnamed protein product [Xenopus laevis]	342	342	100%	8e-119	94.32%	CAA35760.1
<input checked="" type="checkbox"/>	ferritin heavy chain A [Microcaecilia unicolor]	342	342	100%	1e-118	91.48%	XP_030053805.1

Mest lik sekvens i toppen av listan. Hög score och E-värde nära 0 betyder att det är liten risk att sekvensen av ren slump är lik ditt query

BLAST resultat

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view

Pairwise

 ?

100 sequences selected ?

Download

GenPept

Graphics

ferritin, heavy polypeptide 1, gene 1 [Xenopus tropicalis]
Sequence ID: [NP_989008.1](#) Length: 176 Number of Matches: 1
[See 2 more title\(s\)](#)

Range 1: 1 to 176

GenPept

Graphics

Next Match

Previous Match

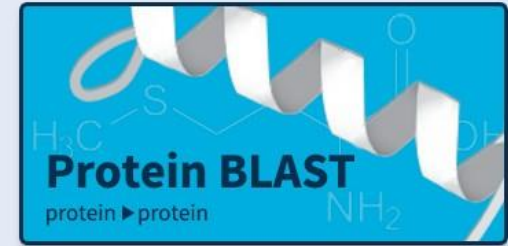
	Score	Expect	Method	Identities	Positives	Gaps
	368 bits(944)	7e-129	Compositional matrix adjust.	176/176(100%)	176/176(100%)	0/176(0%)
Query 1	MQSQVRQNFNSDCEAAINRMVMNELYASYVYLSMSYYFDRDDVALHHVAKFFKEQSHEER					60
Sbjct 1	MQSQVRQNFNSDCEAAINRMVMNELYASYVYLSMSYYFDRDDVALHHVAKFFKEQSHEER					60
Query 61	EHA EKFLKYQNKRGGRAVLQDIKKPERDEWGNTLEAMQAALQLEKTVNQALLDLHKLASD					120
Sbjct 61	EHA EKFLKYQNKRGGRAVLQDIKKPERDEWGNTLEAMQAALQLEKTVNQALLDLHKLASD					120
Query 121	KTDPHLCDFLESEYLEEQVKAMKELGDYITNLKRLGVPQNGMGEYLFDKHTLGESS					176
Sbjct 121	KTDPHLCDFLESEYLEEQVKAMKELGDYITNLKRLGVPQNGMGEYLFDKHTLGESS					176

Olika typer av BLAST

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

Web BLAST



Genome Browsers

- Genome browsers visualiserar olika typer av integrerad data.
 - Gener
 - Proteiner
 - Expression
 - Regulering
 - Variation
 - Komperativ genomik

Big Genome Browser:

Ensembl Genome browser

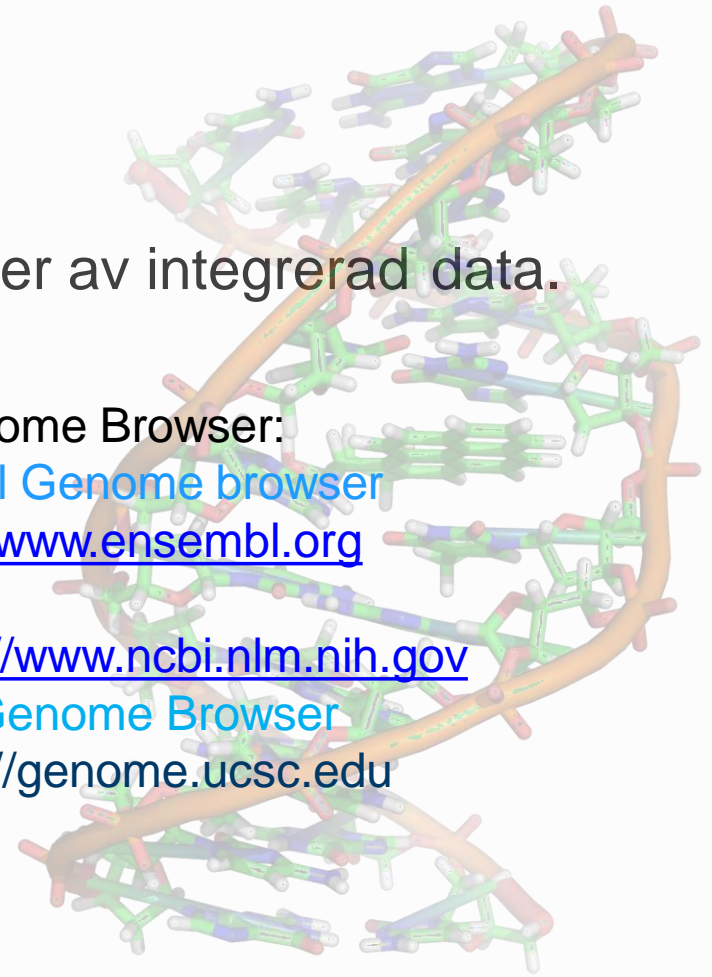
<http://www.ensembl.org>

NCBI

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

UCSC Genome Browser

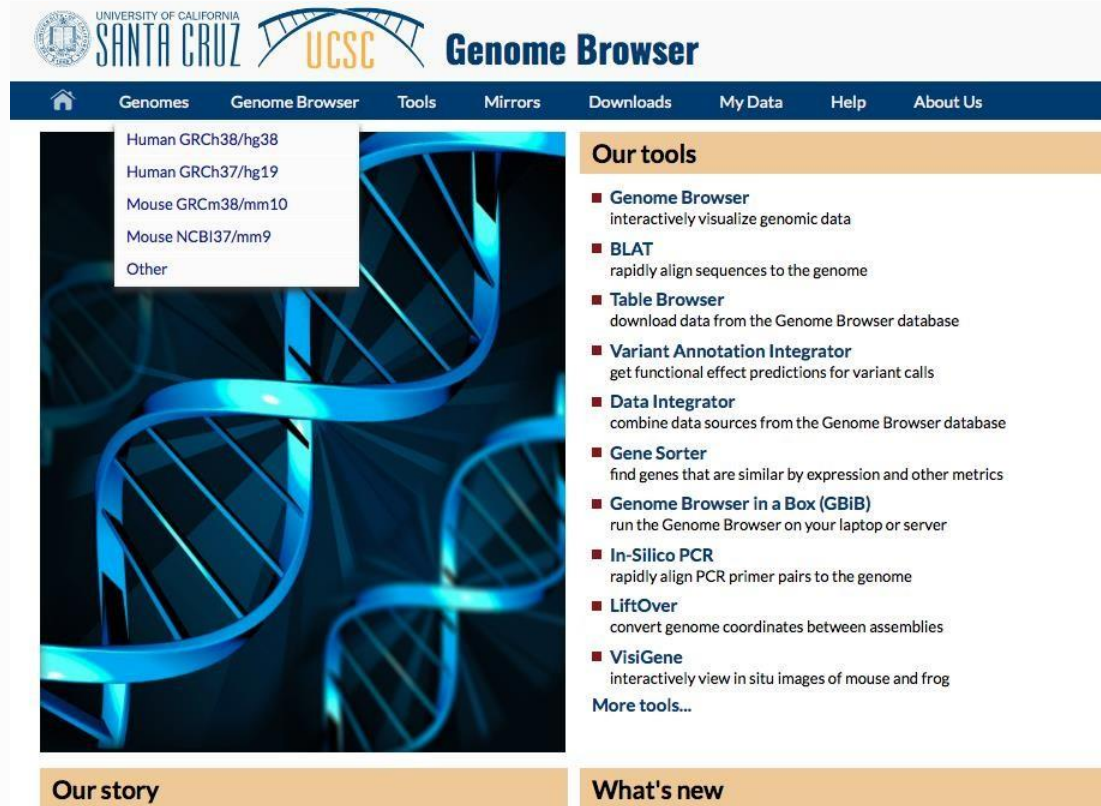
<https://genome.ucsc.edu>



UCSC Genome Browser

Online genome browser med access till genomsekvensdata ifrån många olika organismer.

NOTE! Genom koordinater skiljer sig mellan versioner



The screenshot shows the UCSC Genome Browser homepage. At the top, there is a header with the University of California Santa Cruz logo and the UCSC Genome Browser title. Below the header is a navigation bar with links: Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area is divided into two columns. The left column features a large image of a DNA double helix and a dropdown menu with options: Human GRCh38/hg38, Human GRCh37/hg19, Mouse GRCm38/mm10, Mouse NCBI37/mm9, and Other. The right column is titled 'Our tools' and lists several tools with brief descriptions: Genome Browser (interactively visualize genomic data), BLAT (rapidly align sequences to the genome), Table Browser (download data from the Genome Browser database), Variant Annotation Integrator (get functional effect predictions for variant calls), Data Integrator (combine data sources from the Genome Browser database), Gene Sorter (find genes that are similar by expression and other metrics), Genome Browser in a Box (GBiB) (run the Genome Browser on your laptop or server), In-Silico PCR (rapidly align PCR primer pairs to the genome), LiftOver (convert genome coordinates between assemblies), and VisiGene (interactively view in situ images of mouse and frog). Below the 'Our tools' section are two more sections: 'Our story' and 'What's new'.

UNIVERSITY OF CALIFORNIA SANTA CRUZ UCSC Genome Browser

Home Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Human GRCh38/hg38
Human GRCh37/hg19
Mouse GRCm38/mm10
Mouse NCBI37/mm9
Other

Our tools

- **Genome Browser**
interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
interactively view in situ images of mouse and frog

[More tools...](#)

Our story

What's new

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[View](#)[Help](#)[About Us](#)

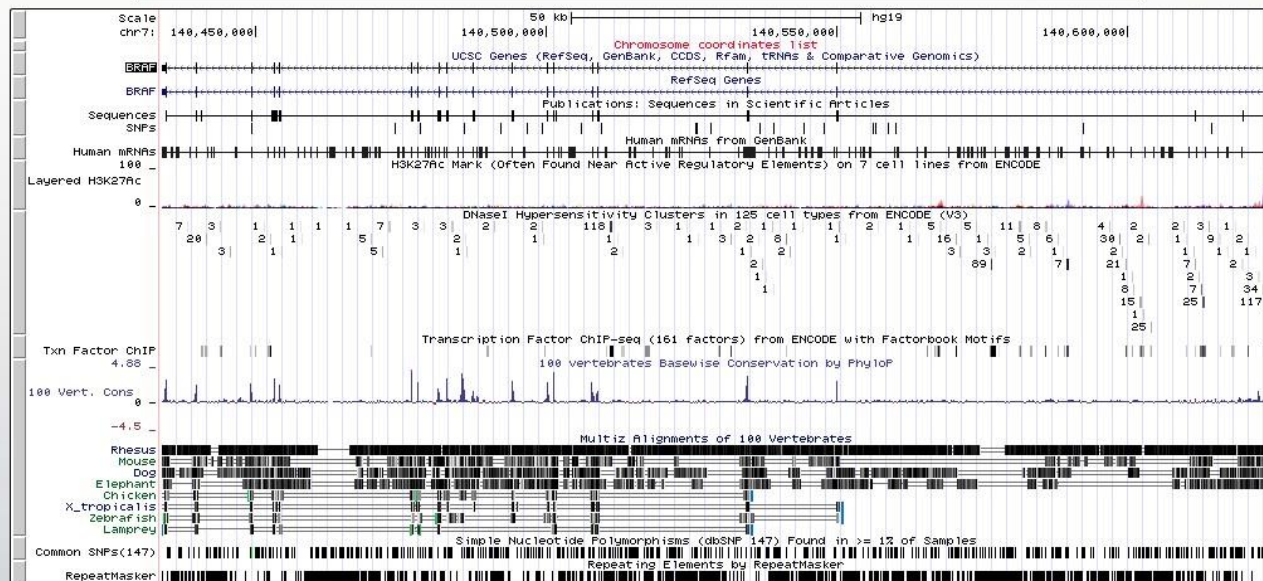
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr7:140,433,813-140,624,564 190,752 bp.

enter position, gene symbol or search terms

go



move start

< 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

move end

< 2.0 >

track search

default tracks

default order

hide all

manage custom tracks

track hubs

configure

multi-region

reverse

resize

refresh

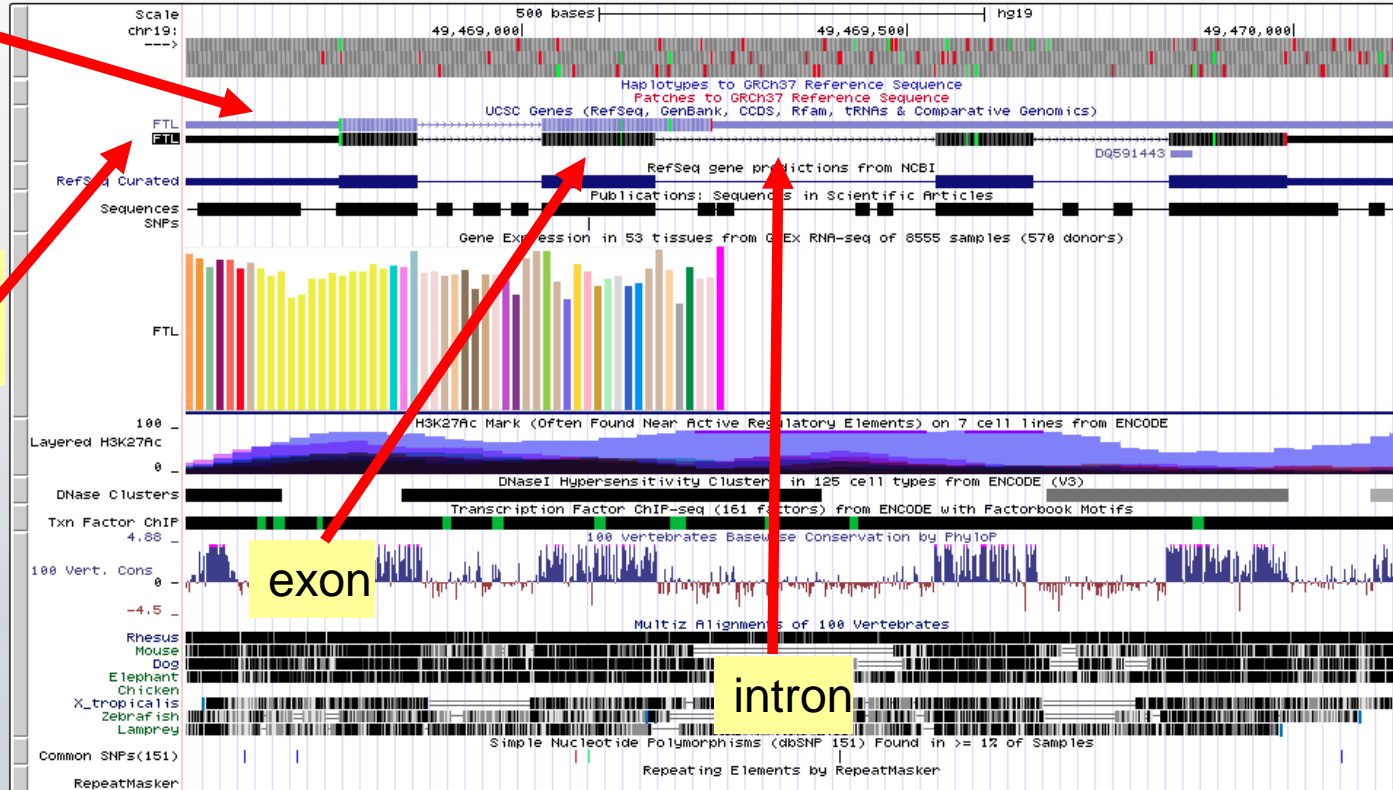
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

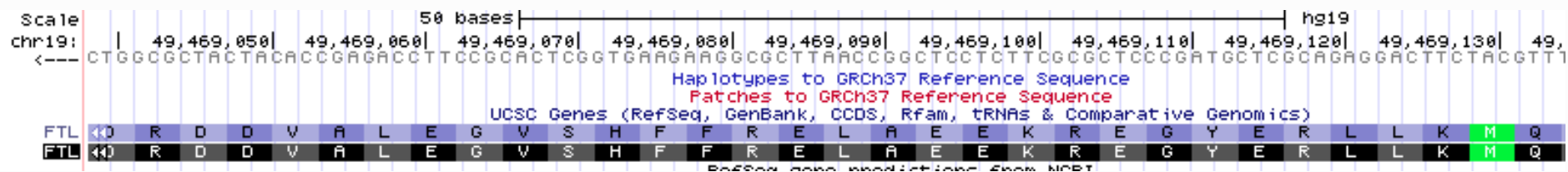
chr19:49,468,566-49,470,136 1,571 bp. enter position, gene symbol, HGVS or search terms

go

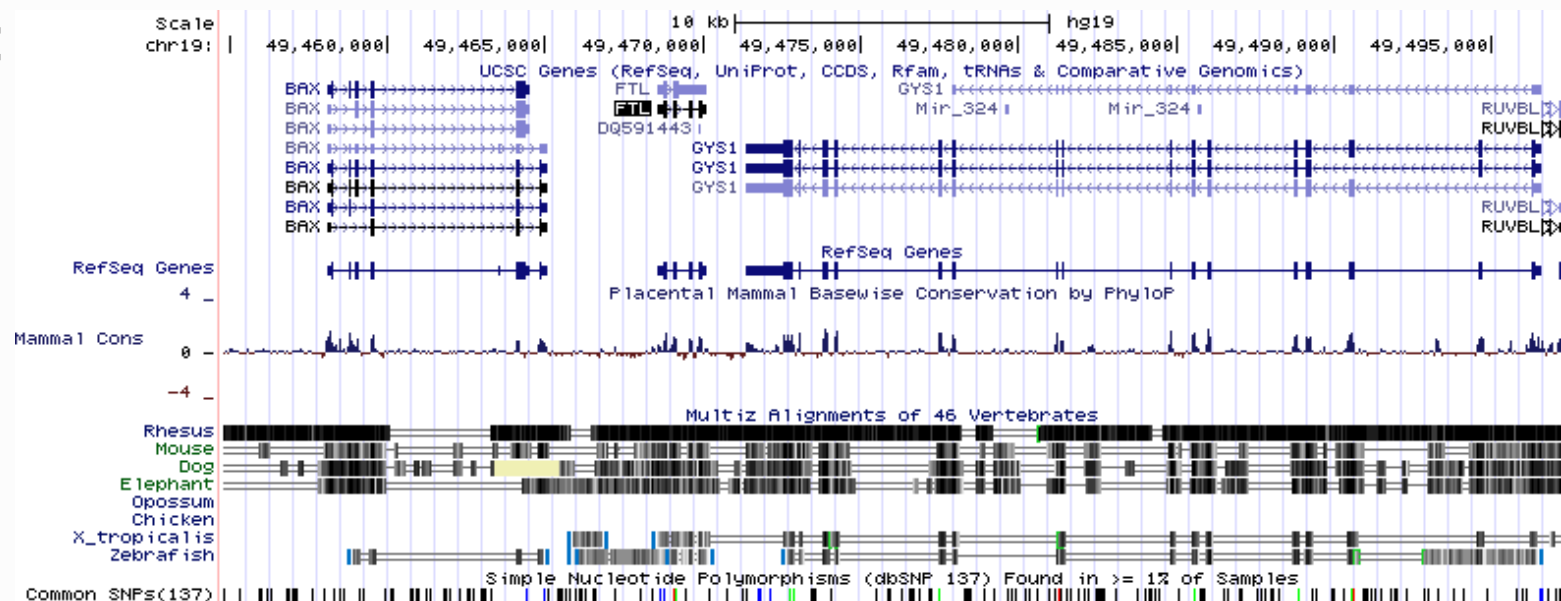
chr19 (q13.33) 19p13.3 19p13.2 p13.11 19p12 19q12 13.11 13.12 19q13.2 13.32 q13.33 13.42 13.43

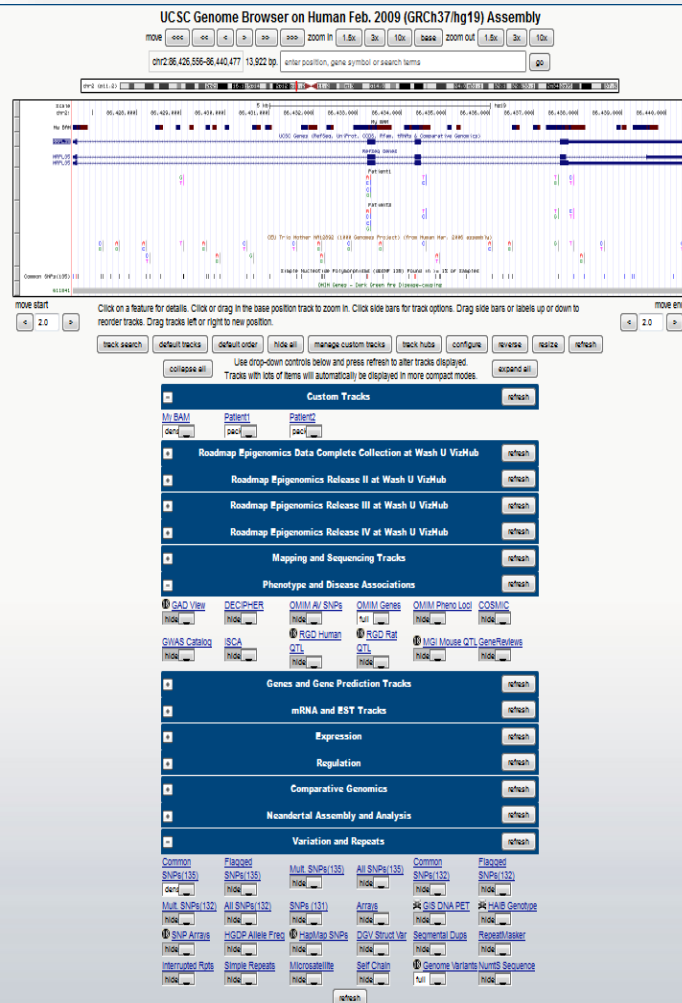


Zooma in:



Zooma ut:

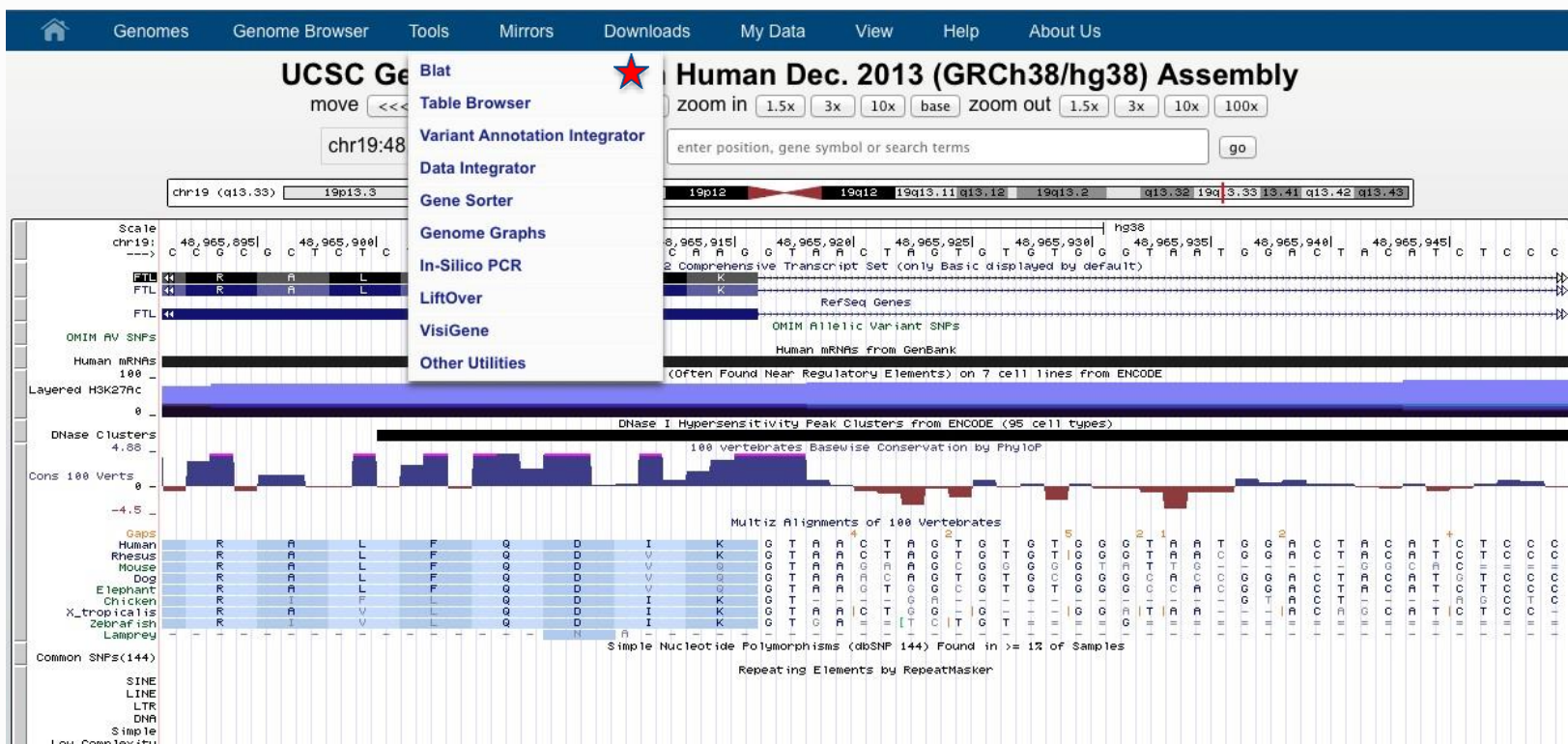




Genome viewer

Custom Annotation tracks (grupper av data)

- Mappning och sekvenser
- Genprediktioner
- Variationer och repeats
- komparativ genomik
- Custom tracks (BAM, VCF)



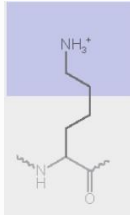
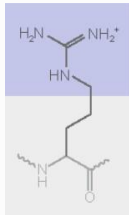
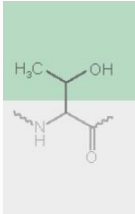
BLAT har en liknande funktion som BLAST, men behöver en nästan exakt match för att hitta rätt. BLAT passar inte om man vill hitta avlägset relaterade sekvenser.

Variation mellan individer

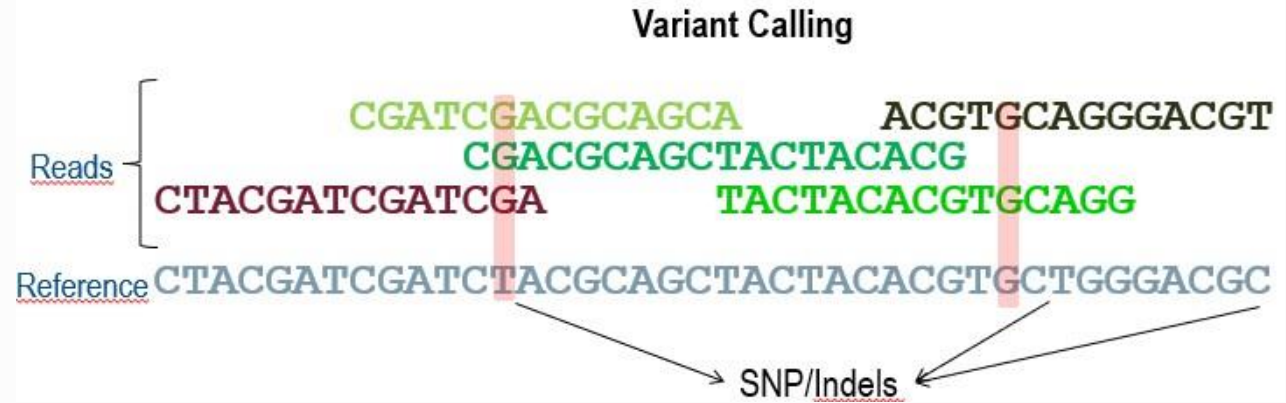
- Single nucleotide polymorphism (SNP)
- Två mänskliga genom är i genomsnitt 99,9% lika
- Genom är stora så innebär ca 3 miljoner skillnader, men de flesta av dessa har ingen funktion
- De flesta varianterna 90% är i form av förändring av en nukleotid SNP



Single nucleotide variants (SNV)

* Substitution	CTGAAG	CTGA---AG	(ref)
* Insertion	CTGA TG	CTGA CTG AG	
* Deletion		CTGAAG CTG--G	(ref)
* Non-coding	AAG	Lys	
* Coding	AA A	Lys	
- Synonymous			
- Non-synonymous			
- Misense	A GG	Arg	
	A CG	Thr	
- Nonsense	UAG	STOP	

Identifiering av SNVs



Variant Annotation

- Är det en känd SNP?
- Hur vanlig är den i befolkningen?
- Ligger varianten i en gen?
- Vilken påverkan kan varianten ha på genens funktion?
- Vad vet vi om den genens funktion idag?

VCF: Variant Call format

```
##fileformat=VCFv4.1
##FILTER=<ID=FSindel,Description="FS > 200.0">
##FILTER=<ID=FSsnp,Description="FS > 60.0">
##FILTER=<ID=LowDP,Description="DP < 10.0">
##FILTER=<ID=LowQual,Description="Low quality">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=BaseCounts,Number=4,Type=Integer,Description="Counts of each base">
##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">
##contig=<ID=chr1,length=249250621,assembly=hg19>
##contig=<ID=chr2,length=243199373,assembly=hg19>
##contig=<ID=chr3,length=198022430,assembly=hg19>
##reference=file:///jumbo/db/GATK/Minusy/ucsc.hg19.minusMY.fasta
```

Header

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	C2463	S1
chr1	1671852	.	G	C	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671853	.	C	T	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671858	.	TCTCGC	T	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671866	.	A	AGAGAT	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671880	.	G	C	143.77	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1684347	.	C	CCCT	391.79	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,9:19:429,0,19	
chr1	1695462	.	C	A	537.77	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:21,21:99:566,0,618	
chr1	1695574	.	G	T	532.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:561,48,0	
chr1	1695742	.	C	T	34.78	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,2:28:63,0,28	
chr1	1695831	.	C	T	61.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:89,9,0	
chr1	1710035	.	C	T	75.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:103,9,0	
chr2	123332	.	C	G	521.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:550,48,0	
chr2	128345	.	C	T	4286.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,127:99:4315,378,0	
chr2	130618	.	G	C	340.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,12:36:369,36,0	
chr2	132469	.	C	A	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,82:99:4917,256,0	
chr2	132490	.	C	G	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,62:99:4917,256,0	
chr2	132696	.	A	G	144.90	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,5:15:173,15,0	

Body

VCF: Filter

Header

```
##fileformat=VCFv4.1
##FILTER=<ID=FSindel,Description="FS > 200.0">
##FILTER=<ID=FSsnp,Description="FS > 60.0">
##FILTER=<ID=LowDP,Description="DP < 10.0">
##FILTER=<ID=LowQual,Description="Low quality">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=BaseCounts,Number=4,Type=Integer,Description="Counts of each base">
##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">
##contig=<ID=chr1,length=249250621,assembly=hg19>
##contig=<ID=chr2,length=243199373,assembly=hg19>
##contig=<ID=chr3,length=198022430,assembly=hg19>
##reference=file:///jumbo/db/GATK/Minusy/ucsc.hg19.minusMY.fasta
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	C2463	S1
chr1	1671852	.	G	C	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671853	.	C	T	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671858	.	TCTCGC	T	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671866	.	A	AGAGAT	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671880	.	G	C	143.77	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1684347	.	C	CCCT	391.79	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,9:19:429,0,19	
chr1	1695462	.	C	A	537.77	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:21,21:99:566,0,618	
chr1	1695574	.	G	T	532.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:561,48,0	
chr1	1695742	.	C	T	34.78	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,2:28:63,0,28	
chr1	1695831	.	C	T	61.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:89,9,0	
chr1	1710035	.	C	T	75.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:103,9,0	
chr2	123332	.	C	G	521.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:550,48,0	
chr2	128345	.	C	T	4286.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,127:99:4315,378,0	
chr2	130618	.	G	C	340.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,12:36:369,36,0	
chr2	132469	.	C	A	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,82:99:4917,256,0	
chr2	132490	.	C	G	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,62:99:4917,256,0	
chr2	132696	.	A	G	144.90	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,5:15:173,15,0	

Body

VCF: Info

Header

```
##fileformat=VCFv4.1
##FILTER=<ID=FSindel,Description="FS > 200.0">
##FILTER=<ID=FSsnp,Description="FS > 60.0">
##FILTER=<ID=LowDP,Description="DP < 10.0">
##FILTER=<ID=LowQual,Description="Low quality">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=BaseCounts,Number=4,Type=Integer,Description="Counts of each base">
##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">
##contig=<ID=chr1,length=249250621,assembly=hg19>
##contig=<ID=chr2,length=243199373,assembly=hg19>
##contig=<ID=chr3,length=198022430,assembly=hg19>
##reference=file:///jumbo/db/GATK/Minusy/ucsc.hg19.minusMY.fasta
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	C2463	S1
chr1	1671852	.	G	C	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671853	.	C	T	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296	
chr1	1671858	.	TCTCGC	T	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671866	.	A	AGAGAT	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1671880	.	G	C	143.77	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296	
chr1	1684347	.	C	CCCT	391.79	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,9:19:429,0,19	
chr1	1695462	.	C	A	537.77	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:21,21:99:566,0,618	
chr1	1695574	.	G	T	532.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:561,48,0	
chr1	1695742	.	C	T	34.78	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,2:28:63,0,28	
chr1	1695831	.	C	T	61.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:89,9,0	
chr1	1710035	.	C	T	75.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:103,9,0	
chr2	123332	.	C	G	521.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:550,48,0	
chr2	128345	.	C	T	4286.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,127:99:4315,378,0	
chr2	130618	.	G	C	340.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,12:36:369,36,0	
chr2	132469	.	C	A	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,82:99:4917,256,0	
chr2	132490	.	C	G	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,62:99:4917,256,0	
chr2	132696	.	A	G	144.90	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,5:15:173,15,0	

Body

VCF: Format

```
##fileformat=VCFv4.1
##FILTER=<ID=FSindel,Description="FS > 200.0">
##FILTER=<ID=FSsnp,Description="FS > 60.0">
##FILTER=<ID=LowDP,Description="DP < 10.0">
##FILTER=<ID=LowQual,Description="Low quality">
##FORMAT=<ID=AD,Number=.,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth (reads with MQ=255 or with bad mates are filtered)">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=BaseCounts,Number=4,Type=Integer,Description="Counts of each base">
##INFO=<ID=BaseQRankSum,Number=1,Type=Float,Description="Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP Membership">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Approximate read depth; some reads may have been filtered">
##contig=<ID=chr1,length=249250621,assembly=hg19>
##contig=<ID=chr2,length=243199373,assembly=hg19>
##contig=<ID=chr3,length=198022430,assembly=hg19>
##reference=file:///jumbo/db/GATK/Minusy/ucsc.hg19.minusMY.fasta
```

Header

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	C2463 S1
chr1	1671852	.	G	C	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296
chr1	1671853	.	C	T	143.77	LowDP;MQD	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:0,1:99:172,0,296
chr1	1671858	.	TCTCGC	T	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296
chr1	1671866	.	A	AGAGAT	134.73	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296
chr1	1671880	.	G	C	143.77	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,1:99:172,0,296
chr1	1684347	.	C	CCCT	391.79	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,9:19:429,0,19
chr1	1695462	.	C	A	537.77	PASS	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:21,21:99:566,0,618
chr1	1695574	.	G	T	532.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:561,48,0
chr1	1695742	.	C	T	34.78	LowDP	AC=1;AF=0.500;AN=2	GT:AD:GQ:PL	0/1:1,2:28:63,0,28
chr1	1695831	.	C	T	61.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:89,9,0
chr1	1710035	.	C	T	75.28	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,3:9:103,9,0
chr2	123332	.	C	G	521.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,16:48:550,48,0
chr2	128345	.	C	T	4286.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,127:99:4315,378,0
chr2	130618	.	G	C	340.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,12:36:369,36,0
chr2	132469	.	C	A	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,82:99:4917,256,0
chr2	132490	.	C	G	4888.77	PASS	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,62:99:4917,256,0
chr2	132696	.	A	G	144.90	LowDP	AC=2;AF=1.00;AN=2	GT:AD:GQ:PL	1/1:0,5:15:173,15,0

Body

Övning

<https://github.com/bcfgothenburg/VT21/wiki/Molekylarbiologisk-metodik>

- Svara på frågorna på övningarna
- Fastnar du finns svaren uppe på github
- Fråga oss
- Ni behöver inte lämna in något



VISIT OUR WEBSITE

<https://cf.gu.se/english/Bioinformatics>