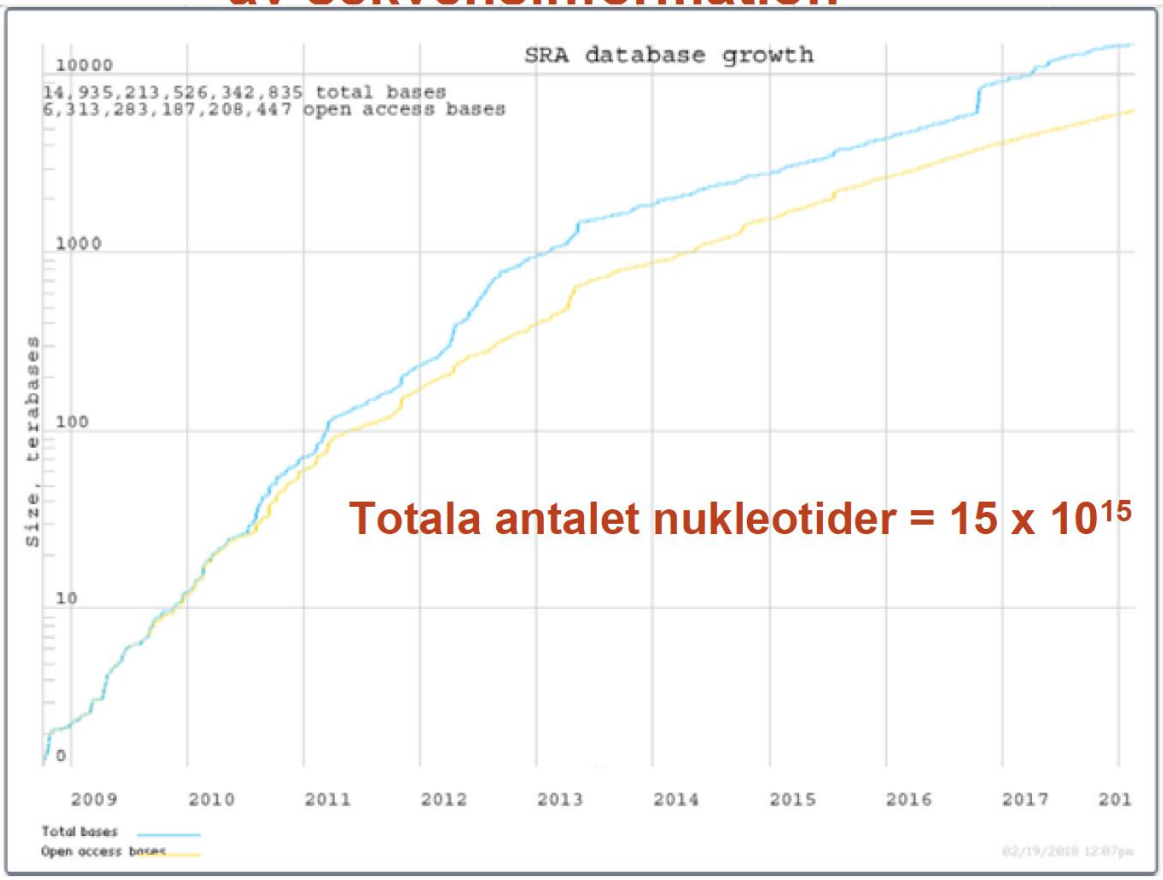
GÖTEBORGS UNIVERSITET

Bioinformatics
Core Facility

BMA Molekylärbiologisk metodik
Bioinformatik

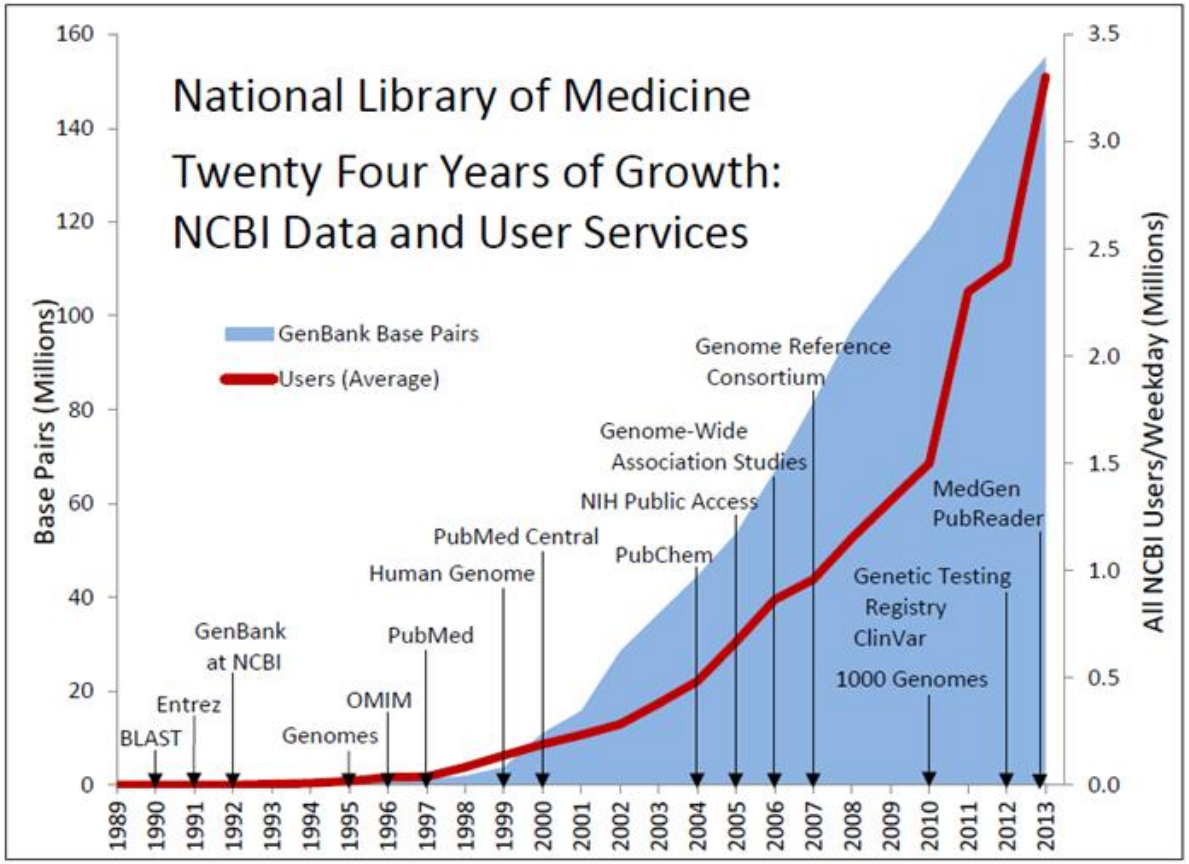
Feb/2019

Explosiv tillväxt av sekvensinformation



Totala antalet nukleotider = 15 x 10^{15}

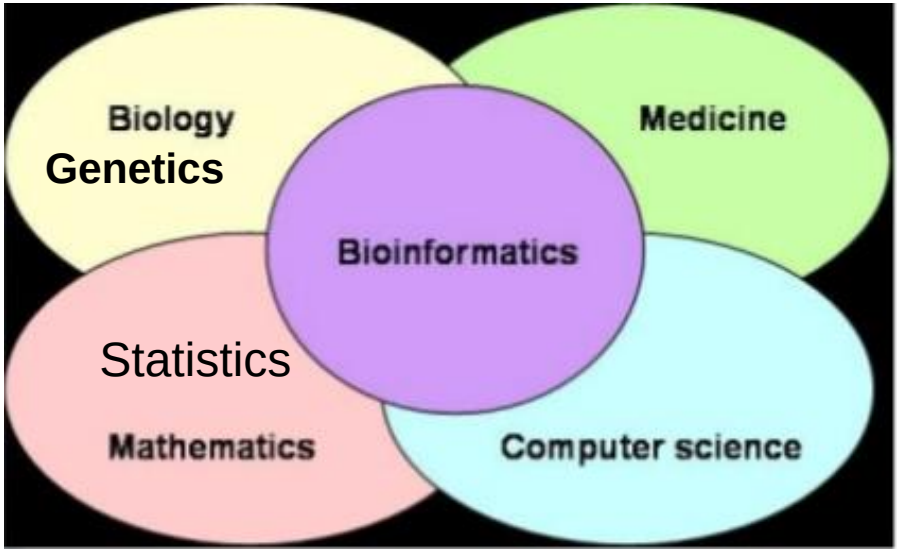
Growth of GenBank



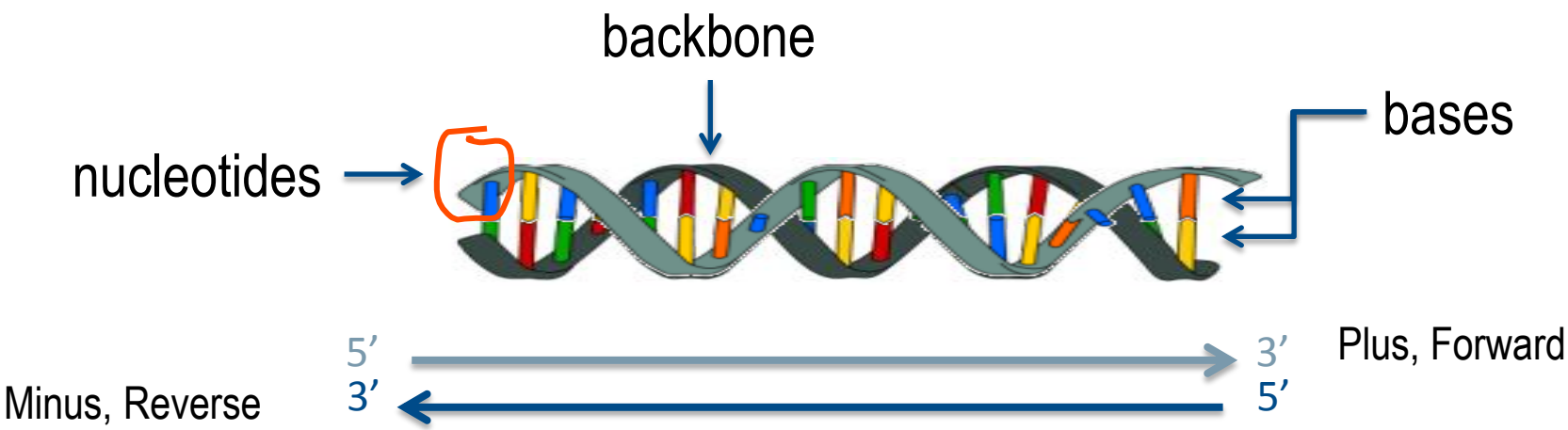
Report from NCBI 2013
GenBank world's largest collection of DNA sequences in collaboration with UK/EMBL and Japan/DDBJ (DNA data bank of Japan)

Bioinformatik

Hantering och analys av information
från genteknologisk forskning
Tvärvetenskaplig disciplin



DNA




```
AGCTGACGATGGGCAGATACACAGTAAC
TCGACTGCTACCCGTCTATGTGTCATTG
```

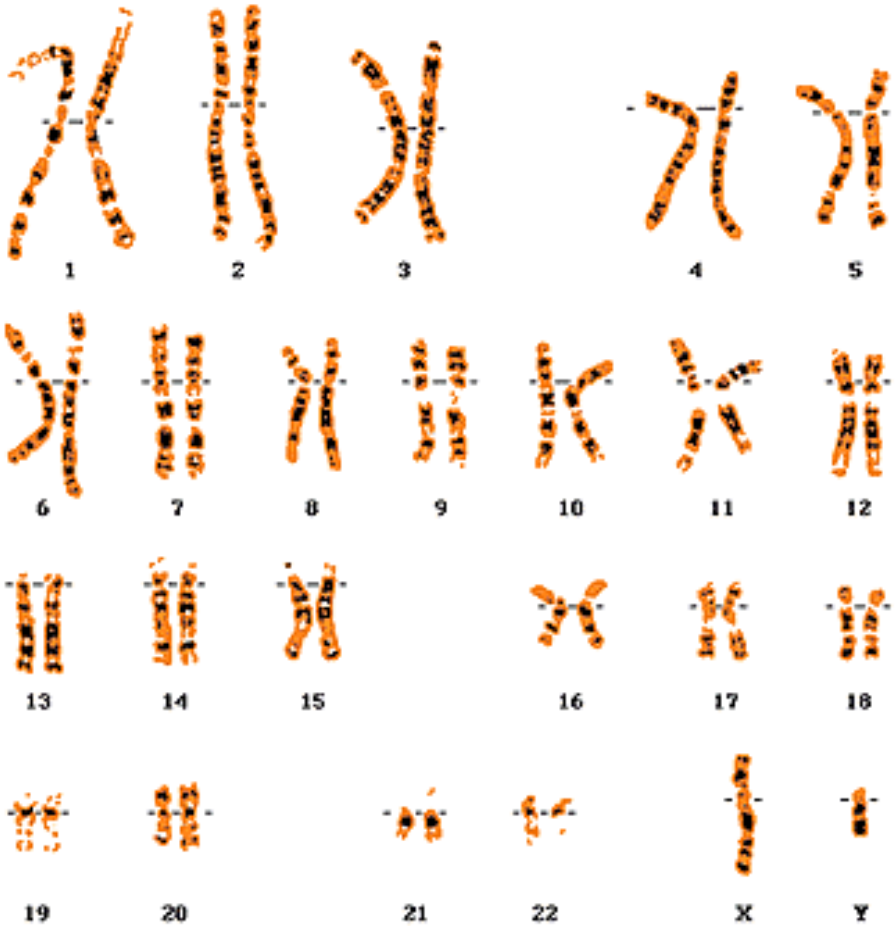
Genes can be located on both strands. So the mRNA sequence always corresponds to the 5-3 coding sequence of a gene.

- Double stranded
- Antiparallel
- Complementary

A=T adenine=thyamine

C≡G cytosine≡guanine

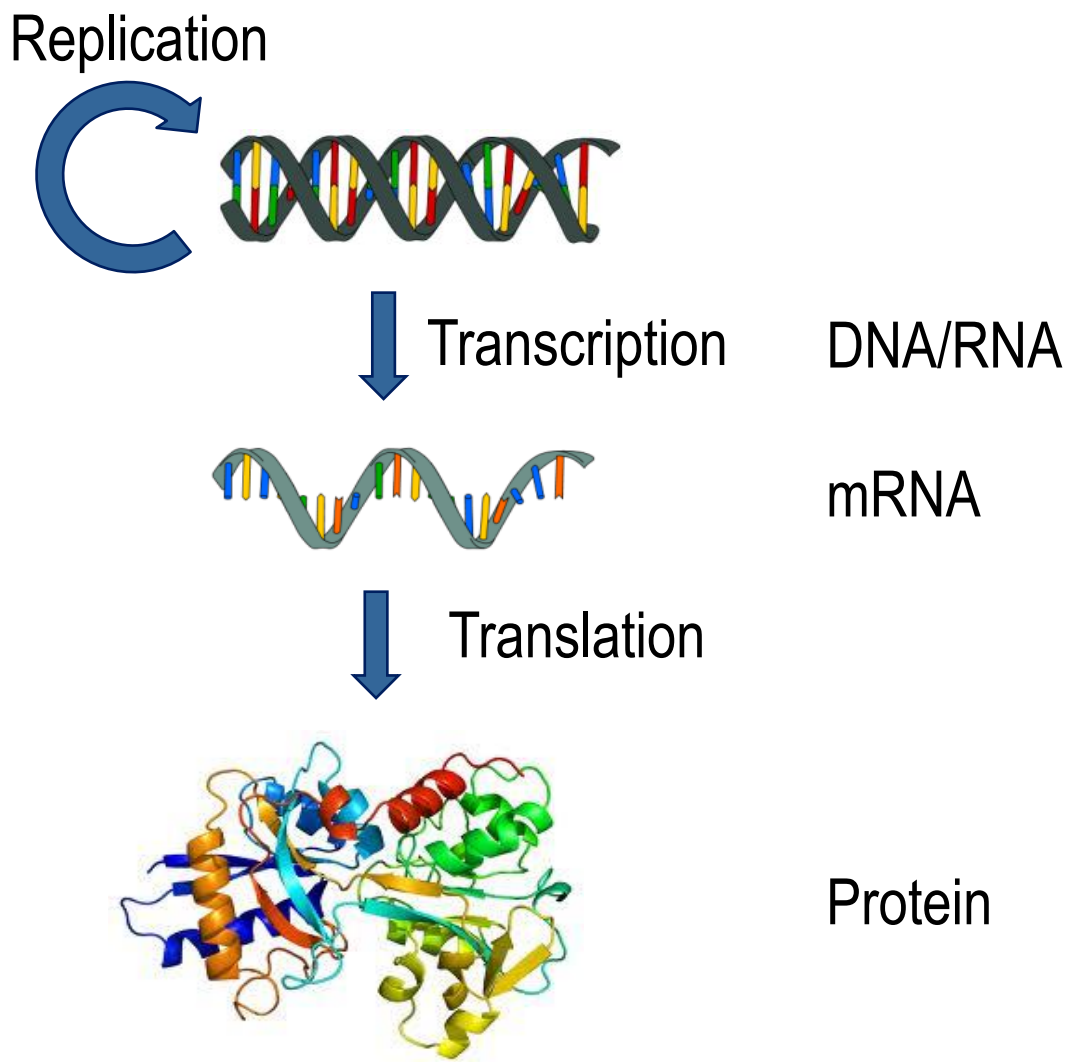
Genomes split into chromosomes



Human Genome
22 autosomal pairs
1 allosome pair
Diploid

Species	# Chr	Ploidy
A.thaliana	10	Diploid
Bread Wheat	42	Hexaploid
Tobacco	48	Tetraploid
Fruit fly	8	Diploid
Earth worm	36	Diploid
Mouse	40	Diploid
Human	46	Diploid
Dog	78	Diploid
Goldfish	100-104	Diploid

Central dogma - Flow of genetic information



The Genetic Code

20 amino acids Proteins - long chain of amino-acids

nonpolar | polar | basic | acidic | (stop codon)

Standard genetic code

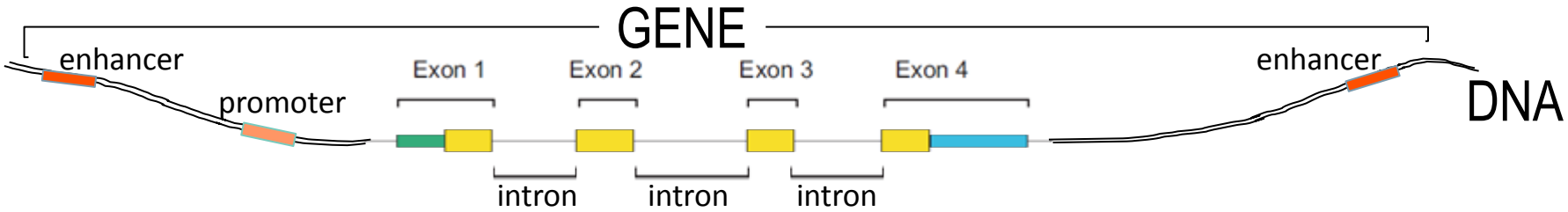
1st base	2nd base: U		2nd base: C		2nd base: A		2nd base: G		3rd base
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

[A] The codon AUG both codes for methionine and serves as an initiation site: the first AUG in an mRNA's coding region is where translation into protein begins.[30]

Several cellular functions: regulation, structure, movement, catalysis, transport, signaling

Genes

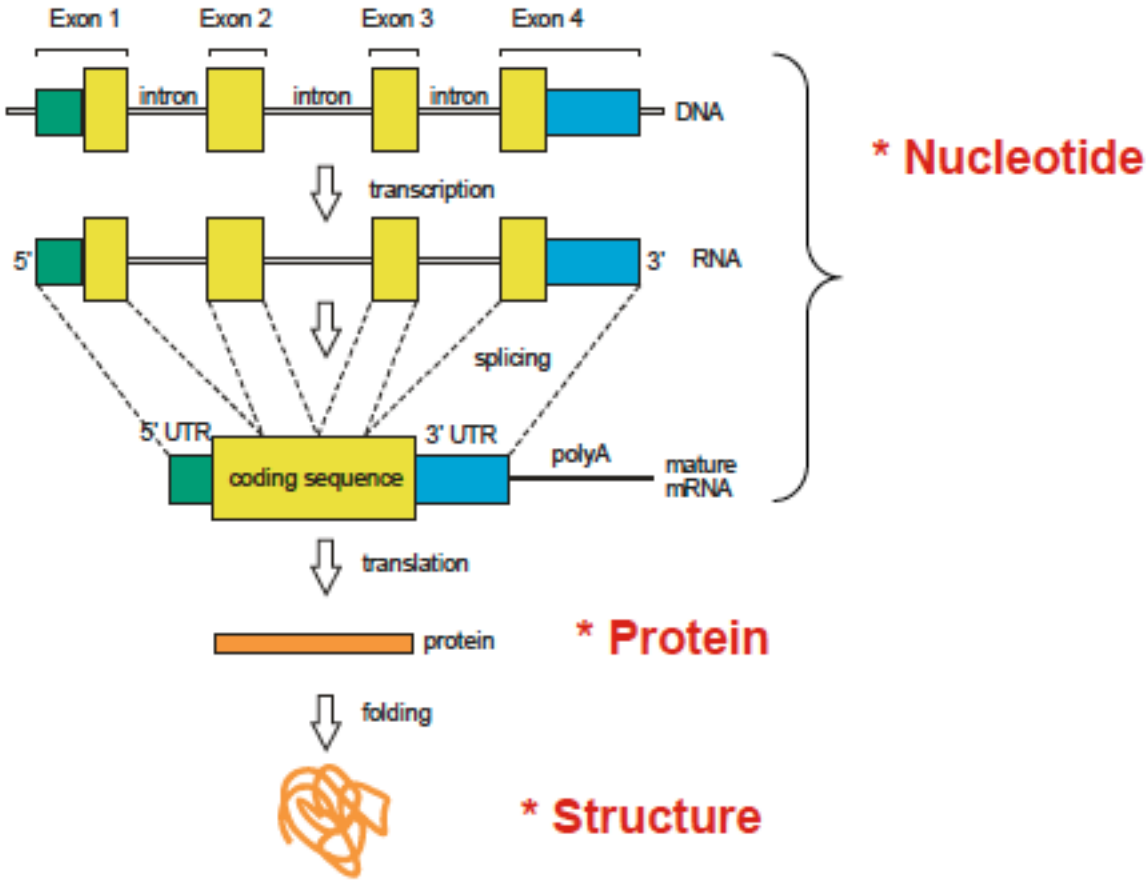
A gene is a segment of DNA that encodes for a function.
Below portion of the DNA that contains all the information for production of a protein



20,000 protein-coding genes (~1.0%)
Unevenly distributed across the chromosomes

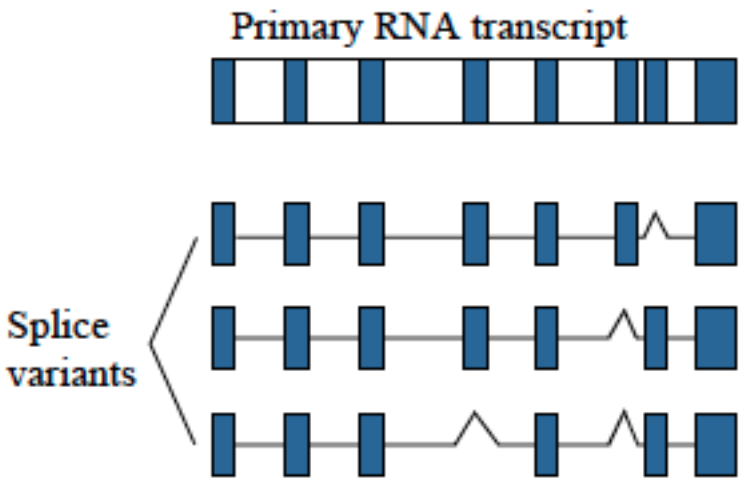
HIST1HIA	781 nt	lack introns
DMD	2.2Mb	largest gene
TTN	80Kb (364 exons)	longest coding sequence

Coding sequence



One gene - multiple transcripts and proteins

Alternative :
- splicing
- promoters
- polyA / termination sites



Variation mellan individer
Single nucleotide polymorphism (SNPs)

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

Functional Consequences

Type	Consequence
SNPs in coding area that alter aa sequence	Cause of most monogenic disorders, e.g: Hemochromatosis (HFE) Cystic fibrosis (CFTR) Hemophilia (F8)
SNPs in coding areas that don't alter aa sequence	May affect splicing
SNPs in promoter or regulatory regions	May affect the level, location or timing of gene expression
SNPs in other regions	No direct known impact on phenotype Useful as markers

Practical applications of genomic variation

- Disease diagnosis
- Association studies
- Pharmacogenomics
- Forensic testing
- Population genetics and evolutionary studies

Comparative Genomics

```
M.sp            UGCAUAACGAA.AGGACAGAACUUUGGCA..AUCCCCUUGAGAAGGAC.GGAAU..GAGC.UUUUGCUCaaacU...UA.CA...GC....A.CCU..CGU.G.UGUUUGa..AGGC..ACCGCUCU.....U.AGGAGCAUCUAggcauauugCCACCAAUUUUUUUU~
R.oryzae        GGUUAAUAGAA.AGGACAGAACUUUGGCA..AUCCCCUUGAGAAGGAC.GGAAU..GAGC.UUUUGCUC....Aa..CA.UA...AC....A.CAC..CGU.G.UGUGUUugaAGGC..AUCGCUCC.....U.CGGAGCAUCUU....ggcauAUUGCCACUAUUUUU~
P.blakesleeanus AACAUAUAGAA.AGGACAGAACUUUGGCA..AUCCCCUUGAGAAGGAC.GGAAC..GAGC.UUUUGCUC....A...CA.CU...AC....A.ACA..CCU.C.GGUGUUugaAGGC..AUUGCUCU.....A.CGGAGCGCUCUU....ggcauAUUGCCACCCAAUUU~
H.sapiens       UGUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAAGA.GGCC.CUCGGGCC....U...GA.CA...AC....A.CGC..AUA.C.GGUUA....AGGC..AUUGCCACCUACUU.CGUGGCA...U.........CUAACCAUCGUUUUU~
G.gallus        UGUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGAUAAGGAGuGGAAGAGGCCC.UAGUGGCC....Ua..UA.CA...AC....A.CGC..AUA.C.GGUUA....AGGC..ACUGCCACCUACUUgCGUGGCA...U.........CUAACCA..UGUUUUU
D.rerio         UGUUGUAUGAAuAGGACAGAAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAAGAGGUCC.UAGUGGCC....U...UC.AA...CA....U.CAC..AUA...UGGUUA...AGGC..AUUGCCACCUACUU.CGUGGCAUCUC.........UAACCAA..GUUUUUU
T.nigroviridis  UGUUGUAUGAA.AGGACAGAAGGUUAGCAccAUCCCCUUGACAAGGAU.GGAAGAGGCCC.CUGAGGCC........UU.CA...AC....A.CAC..AUA.C.GGUUA....AGGCcaAUUGCCACCUACUU.UGUGGCAUCUC.........UAACC.AAGUUUUUU
F.rubripes      ~GUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAAGAGGCCC.CUGUGGCC........UU.CA...AC....A.CAC..AUA.C.GGUUA....AGGC..AUUGCCACCUACUU.CGUGGCAUCUC.........UAACC.AAGUUUUUU
B.floridae      UGUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGAUAAGGAC.GGAAGA.CCCG.CUUGCGGG....U...UUaCA...AC....A.CAU..AUA.C.GGUUA....AGGC..AUUGCCGCCUACUU.CGCGGCA...U.........CUAACC.AAUUUUUUA
X.laevis        UGUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAAGAGGUCC.CAGUGGCC....U...UUaCA...AC....A.CAU..AUA.C.GGUUA....AGGC..AUCGCCACCUACUU.CGUGGCAUCUC.........UAACC.AUGUUUUUU
C.intestinalis  UGUUGAAUAAA.AGGACAGCUGGUUAGCA..CUCCCCUUGAUAAGGAC.GGACAUGUUGA.UUGGUUCA....GccaUU.CA...AU....A.CAU..AUA.A.UGGAUA...AGGC..ACUGUCACCAU.....GUGGCAU..C.........UAUCCAU..GUUUUGU
S.purpuratus    UGUUGUAUGAA.AGGACAGAAGGUUAGCA..CUCCCCUUGAUAAGGAC.GGAAUU.CACC.CUCGGGUG.........A.CA...AC....A.CAC..AAA.U.GGUUA....AGGC..ACUGCCACUUACUU.UGUGGCA...U.........CUAACCACGUUUUUA~
T.spiralis      UGCGGAAUGAA.AGGACAGCUAGUUAGCA..CUCCCCUUCACAAGGAC.GGAGAAU...................U...GU.CAg..AC....A.AGC..ACA.A.CAGCAA...AG...........AUCGCCUA.CCU.................UUGCUCCCCAUUUUUU
A.aegypti       UGUCUUUUGAA.AGGACAGCAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAA..................................C....A.CAU..AAG...GGGCC............GGACAGGCACAA.AAUUUCC...............GGCCC.UAGUCGAUU
A.gambiae       UGUCAUUUGAA.AGGACAGCAGGUUAGCA..CUCCCCUUGACAAGGAU.GGAA..................................C....A.CAU..ACC.C.UCAGAU...CAGU..GG.....GCACCCA.GAGAAAACCUG.........AUCUCGGCCAUUUUU~
A.mellifera     UGUAGUAUGAA.AGGACAGACGUUUAGCA..CUCCCCUAGACAAGGAU.GGAAUAC...............................AC....A.UAU..AAA.U.GGUAA.....GGC..ACUGUCGU.CCUUU.GGCGACA...C.........UUACC.AAAUUUUUU
B.mori          UGUCUAUUGAA.AGGACAGACGGUUAGCA..CUCCCCUUGACAAGGAU.GGAA..................................C....A.CAC..AAU.U.UGGUUA...AGGC..AC...........U.A.........U.........UAACCAAACAUUUU~~
D.melanogaster  UGUUGUUUGGA.AGGAGAGCAAGUUAGCA..CUCCCCUAGACAAGGAU.GGAA..................................C....A.CAU..AAA.C.GGUCG............GCUAGGCACAGA.CAAAAGC..............CGUCCACAAAUUUU~
D.pseudoobscura UGUUGUUUGGA.AGGAGAGCAAGUUAGCA..CUCCCCUAGACAAGGAU.GGAA..................................C....A.CAU..AAA.C.GAGCG............GCUGGC.ACAAG.AGAAAGC..............CGUUCACCAGUUUU~
T.castaneum     UGUAGUAUGAA.AGGACAGACGGUUAGCA..CUCCCCUAGACAAGGUU.GGAA..................................AC....A.CAA..AUA.C.GGUUCA...........GC........GU.C....GCACAC.........UCAACC.CCAUUUUUU
D.pulex         UGUAGUACAAA.AGGAGAGCUGUUUAGCA..CUCCCCUUGAUAAGGAC.GGAA..................................C....A.CAU..AAU...GAGUCA...CGGC..ACAACCUU.UGCG...AAGGU....G.........UGA.UCAUAAUUUUU~
T.adhaerens     UGUUGUAUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGAUAAGGAC.GAGAAA.AGUC.UCCGGACU....U...UC.AAcauAC....A.CAC..AAA.U.GGUUA....AGGC..AUUGCCACCUACUU.UGUGGCA...U.........CUGACCAUCUUUUAA~
H.magnipapillata UGUUGUAUGAG.CCGAGAGCCGGUUAGCA..UCUCCCUCGAUAAGGAC.GAGAAUAGACC.CUGA.GUC....U...AU.CA...ACcacaA.CAC..GCA.U.GGUUA....AGGC..UAUGUCACUUACUU.UGUGGCA...U.........CUAACC.AAUUUUUUG
N.vectensis     UGUUGUAUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGAUAAGGAC.GAGAAA.AGUC.UUCGGACU....U...UC.AA...CC....U.ACA..ACA.CgUGGUUA...AGGC..UAUGCCACCUACUU.CGUGGCAU..C.........UAACCA...UUUUUUG
R.sp            UGUUGUAUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGAUAAGGAC.GAGAAAAGUCU.UCN.GGAC....U...UU.CA...ACuu..A.CAA..CAC.A.UGGUUA...AGGC..UAUGCCAUCUACUU.UGUGGCAU..C.........UAACCAG..UUUUUUU
P.polycephalum  UGUUCGUUGAG.CCGAGAGAAGGUUAGCA..UGCCCCUCGAUAAGGAC.GGCA..AAGCUcUAUGAGCU....U...CA.AA...UU....A.CAC..AUC.C.GUGUUA...AGGC..UAUGACUG........CGGUCAUA.U.........UAACACU..UUUUUUU
A.castellanii   UGCUGGUUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUGCAUAAGGAC.GGGAAA.AGAC.UCCGGUCU.........U.CA...AC....U.CAC..AUC...GUGUUA...AGGC..UA.........................G.........UAACAC.UAACUUUU~
A.thaliana      UGUUCGUAGAA.AGGAGAGAUGGUUGGCAu.CUCCUCU.GACAGAGAC.GGGAUUUGACC.UUCGGGUC....U...UU.GA...AC....A.C....AUC.C.GGUU.A...AGGC..UCUCCACA.UUCG...UGUGGAUCUA...........AACC.CAAUUUUUU
O.sativa        UGUUCGUAGAA.AGGAGAGACGGUUGGCAu.CUCCCCU.GACAGGGAC.GGGAUC.AGCU.UUCGGGCU....U...UC.CU...CC....AuCAC..AUUgC.GGUU.A...AGGC..UACCCGCU....U.CGGCGGAUCUU..........ACCCA..AAUUUUU
S.moellendorffii UGUUCGUAGAA.UGGAGAGAUGUUUGGCA..UCUCCCUCGACAAGGAC.GGGAAU.AACC.GUUCGGGC....U...UU.GA...AC....A.CACagAUG.U.GGUUAG...AGGC..AUCCCACC.CUCG...GGUGGA...U.........CUUACCCCCACUUUU~
P.patens        UGUUCGUUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGAUAAGGAC.GAGAAU.AGCC.UUCGGGCU....U...........U....A.CAC..AUG.A.GGUU.A...AGGC..UUCCCGCGC.......UAUGGG...C.........AACCAUCGCUUUU~
H.parasitica    UGUUCGUGGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGACAAGGAC.GGGAUU.GCAC.GAAAGUGU............A...UC....C.AGU..AUC.U.GGAAAG...UCUA..AGCUCUGUCCUCCA.CAGUGAGGAUA.........UCUAUCAUAGUUUUC~
P.sojae         UGUUCGUUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGACAAGGAC.GGGAUU.GCGC.GUAUGCGU....A...UC.CA...AC....C.ACU..GGA.U.GGUUUA...AGCU..CUGUCCUCCUCCGU..GAAGAC...A.........UCUACCAGUUUUUUU~
P.ramorum       UGUUCGUUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGACAAGGAC.GGGAUU.GCGC.GUAUGCGU....A...UC.CA...AC....C.ACU..GGA.U.GGUUUA...AGCU..CUGUCCUCCUUCGU..GAAGAC...A.........UCUACCAGUUUUUUU~
P.infestans     UGUUCGUUGAG.CCGAGAGAAGGUUAGCA..UCUCCCUCGACAAGGAC.GGGAUU.GCGC.GUUUGCGU....A...UC.CA...AC....C.ACU..GGA.U.GGUUUA...AGCU..CUGUCAUCCUUCU..GGAAGAC...A.........UCUACCAGUUUUUUU~
SS_cons         ::::::::::.::::::::::::::::..::::::::::::::::::.::::::<<<<.____>>>>.....................................<<<<<<.............<<<<<<_______._>>>>>>.............>>>>>>::::::::::
```

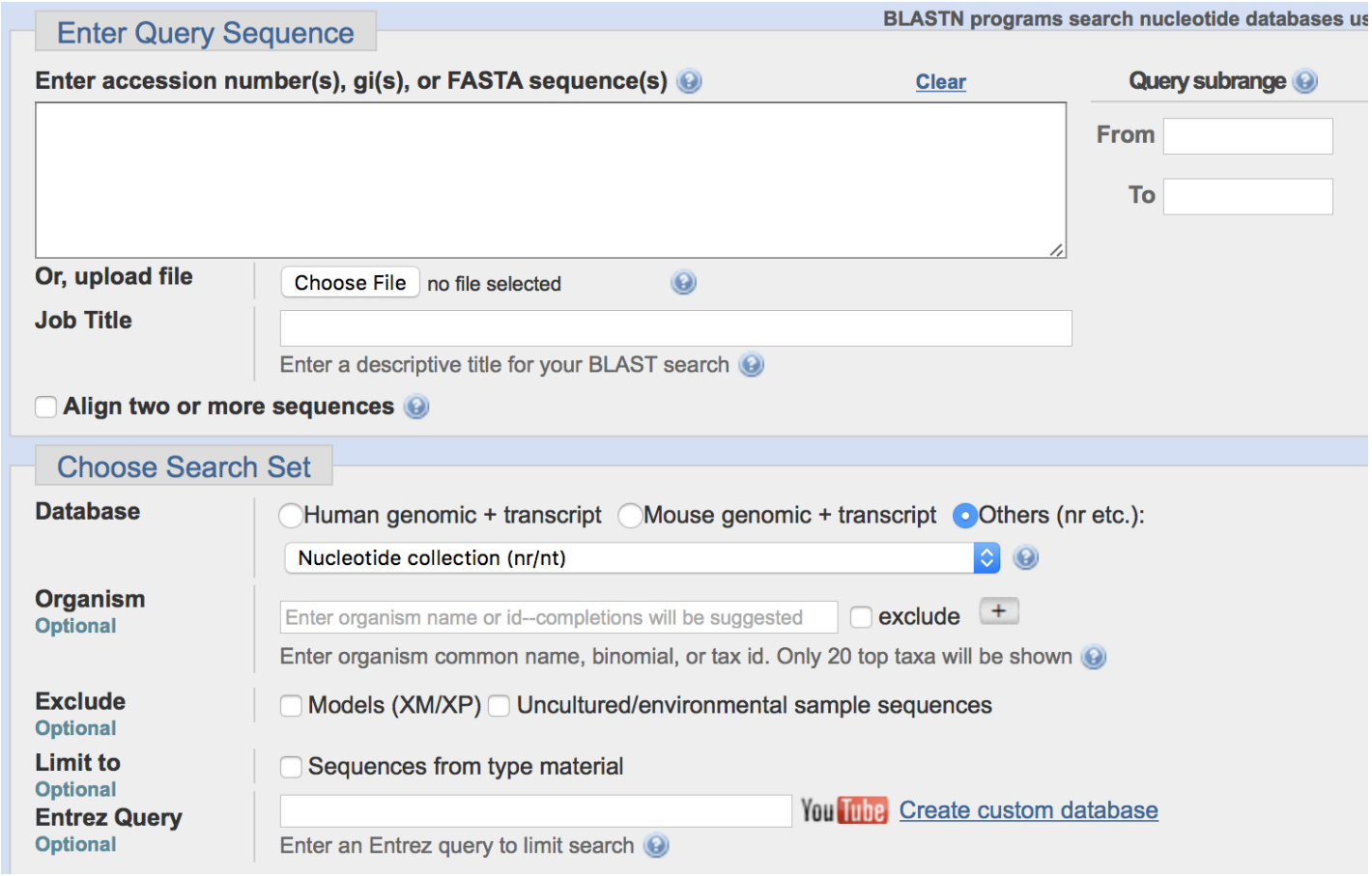
Davila Lopez, M., Rosenblad, MA and T. Samuelsson, (2008)

BLAST

Bioinformatics
Core Facility

BLAST (basic local alignment search tool) är en algoritm* som används för att jämföra sekvens information.

Klistra in en sekvens och hitta liknande sekvenser



*Algoritm är en mängd väl definierade instruktioner för att lösa en uppgift

GÖTEBORGS UNIVERSITET

BLAST results

Bioinformatics
Core Facility

Sequences producing significant alignments:

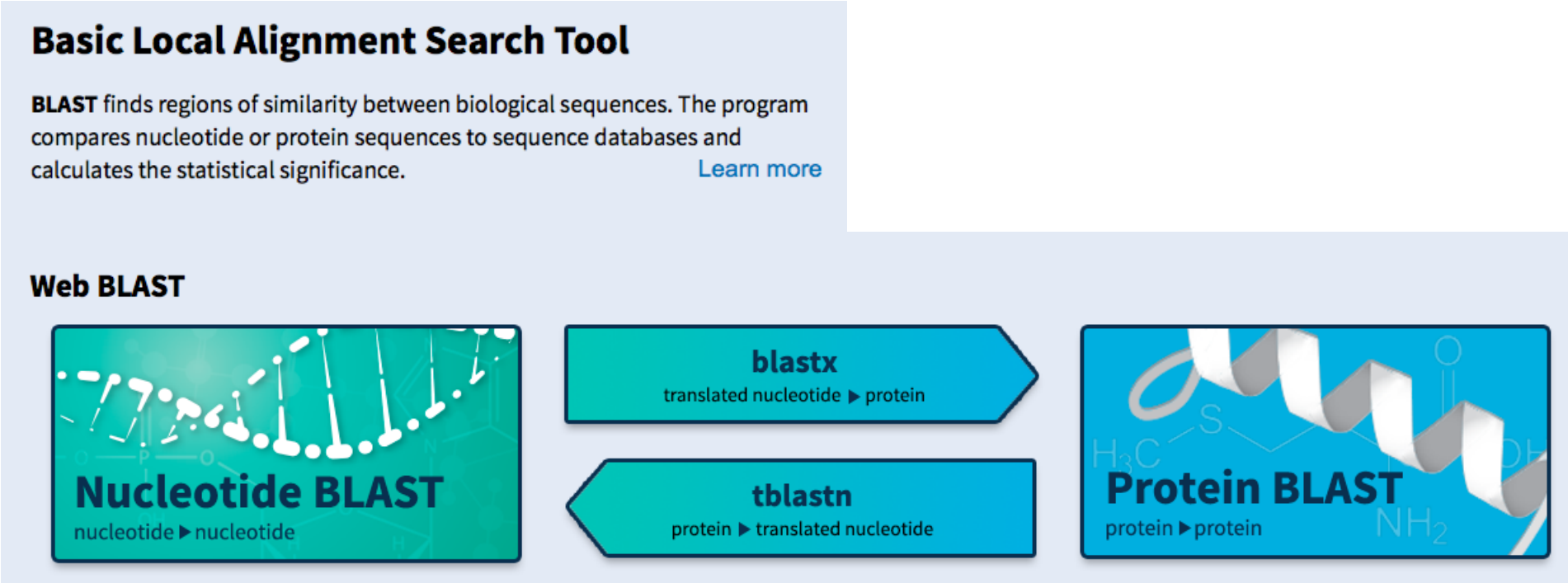
Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
PREDICTED: Nomascus leucogenys rhodopsin (RHO), mRNA	187	187	100%	4e-44	100.00%	XM_003265030.3
PREDICTED: Pan troglodytes rhodopsin (RHO), mRNA	182	182	100%	2e-42	99.01%	XM_516740.7
PREDICTED: Pongo abelii rhodopsin (RHO), mRNA	182	182	100%	2e-42	99.01%	XM_002813145.4
PREDICTED: Pan paniscus rhodopsin (RHO), mRNA	182	182	100%	2e-42	99.01%	XM_003829435.3
Homo sapiens rhodopsin (RHO) gene, RHO-930CG allele, complete cds	182	182	100%	2e-42	99.01%	KP718610.1
Homo sapiens rhodopsin (RHO) gene, RHO-284TC allele, complete cds	182	182	100%	2e-42	99.01%	KP734176.1

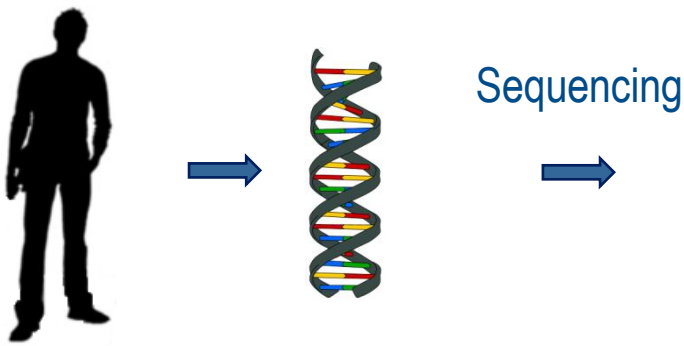
Mest lik sekvens i toppen av listan. Hög score och E-värde nära 0 betyder att det är liten risk att sekvensen av ren slump är lik din fråge-sekvens

Olika typer av BLAST



Man kan söka med både DNA, RNA och protein sekvens

Sequencing & identification of SNPs



illumina · Roche · Oxford NANOPORE Technologies · Applied Biosystems · Helicos BioSciences Corporation · Affymetrix · PACIFIC BIOSCIENCES

Variant, med allele G eller T

Reads
CGATCGACGCAGCA
ACGTGCAGGGACGT
CGACGCAGCTACTACACG
CTACGATCGATCGA
TACTACACGTGCAGG

Reference CTACGATCGATCTACGCAGCTACTACACGTGCTGGGACGC

SNP/Indels

GÖTEBORGS UNIVERSITET

Bioinformatics
Core Facility

Next Generation Sequencing (NGS) Workflow

Ut från sekvenseringsmaskin - Rådata miljontals “reads”
t.ex. exom 30 miljoner reads, 150 baspar långa
fil-format FASTQ

Mappa mot referens
t.ex. BWA (Burrows-Wheeler Aligner)
Output bamfiler binärt format

“Calling” output: identifierade varianter, filer i s.k. VCF format

Annotera, filtrera ut kandidat-varianter som kan förklara fenotyp

FASTQ format

```
@HWI-D00457:83:C6EAMANXX:8:1101:6854:2245 1:N:0:CGCTCATTGGCTCTGA
GAAGCAGGTGCCATGCCTGCACGTTTGTGGCTTAATGACCAAGGAGGGCCGATGAGCAGCCATGGT
GGTGATCACTGC+
B<<BBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FF
```

Varje “read” består av fyra rader. Första raden börjar med @ och har ett unikt namn.
Andra raden visar nukleotid sekvensen
Tredje raden har ett + tecken
Fjärde raden visar kvalitén för varje läst nukleotid

VCF format

```
##fileformat=VCFv4.0
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=.,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF ALT    QUAL FILTER INFO                              FORMAT      NA00001
20    14370   rs6054257 G   A      29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP    0|0:48:1
20    17330   .         T   A      3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP    0|0:49:3
20    1110696 rs6040355 A   G,T    67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP    1|2:21:6
20    1230237 .         T   .      47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP    0|0:54:7
20    1234567 microsat1 GTCT G     50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4
```

Header

Body

VCF format

Alternate allele

dbSNP ID

Defined in the header

Sample values

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	NA00001
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP	0|0:48:1
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017	GT:GQ:DP	0|0:49:3
20	1110696	rs6040355	A	G,T	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT:GQ:DP	1|2:21:6
20	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T	GT:GQ:DP	0|0:54:7
20	1234567	microsat1	GTCT	G	50	PASS	NS=3;DP=9;AA=G	GT:GQ:DP	0/1:35:4

Reference allele

How reliable is The SNP

GT: Genotype - 0/0 homozygous ref
- 0/1 heterozygous
-1/1 homozygous alt

DP: Depth

Variant annotering

- 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?

Genome Browsers

- A genome browser is a graphical interface to display an integrated picture of data from several databases including e.g.
  - Genes
  - Proteins
  - Expression
  - Regulation
  - Variation
  - Comparative analyses

Big Genome Browser:

Ensembl Genome browser

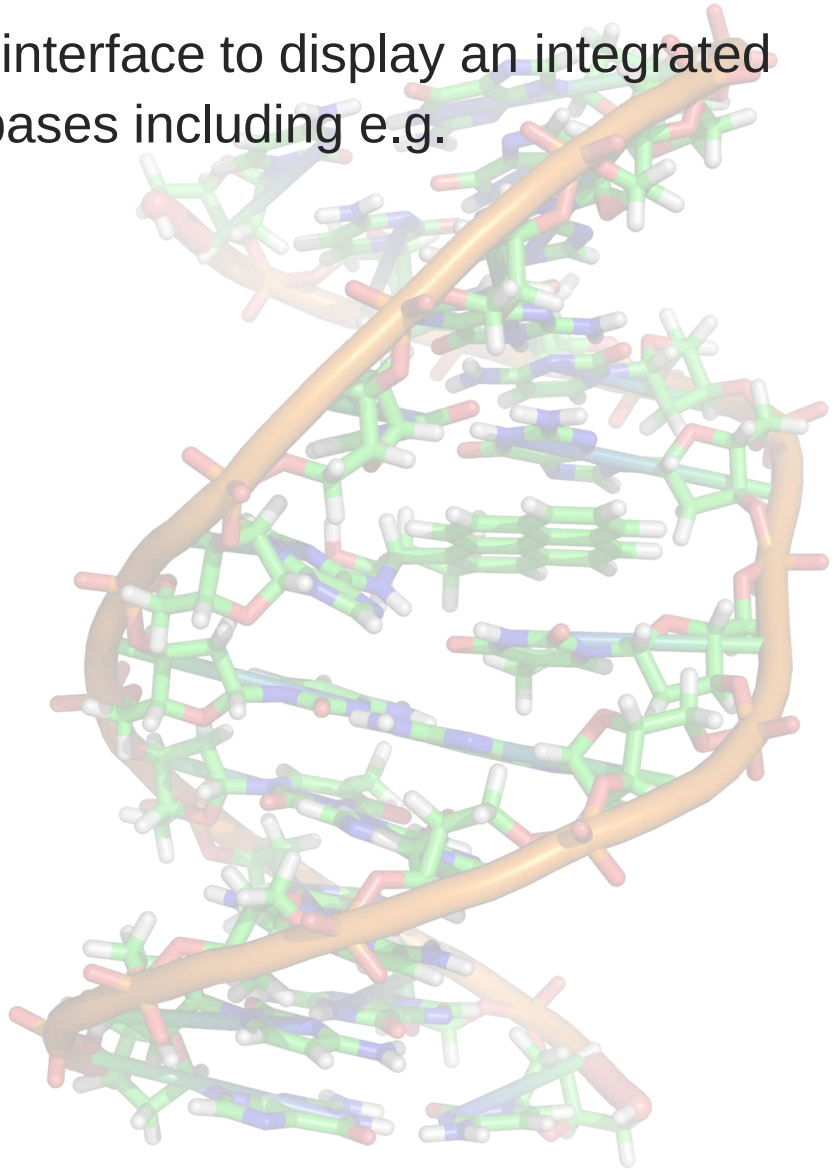
http://www.ensembl.org

NCBI

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

UCSC Genome Browser

https://genome.ucsc.edu



UCSC Browser https://genome.ucsc.edu

On-line genome browser with access to genome sequence data from a variety of organisms integrated with a large collection of aligned annotations.

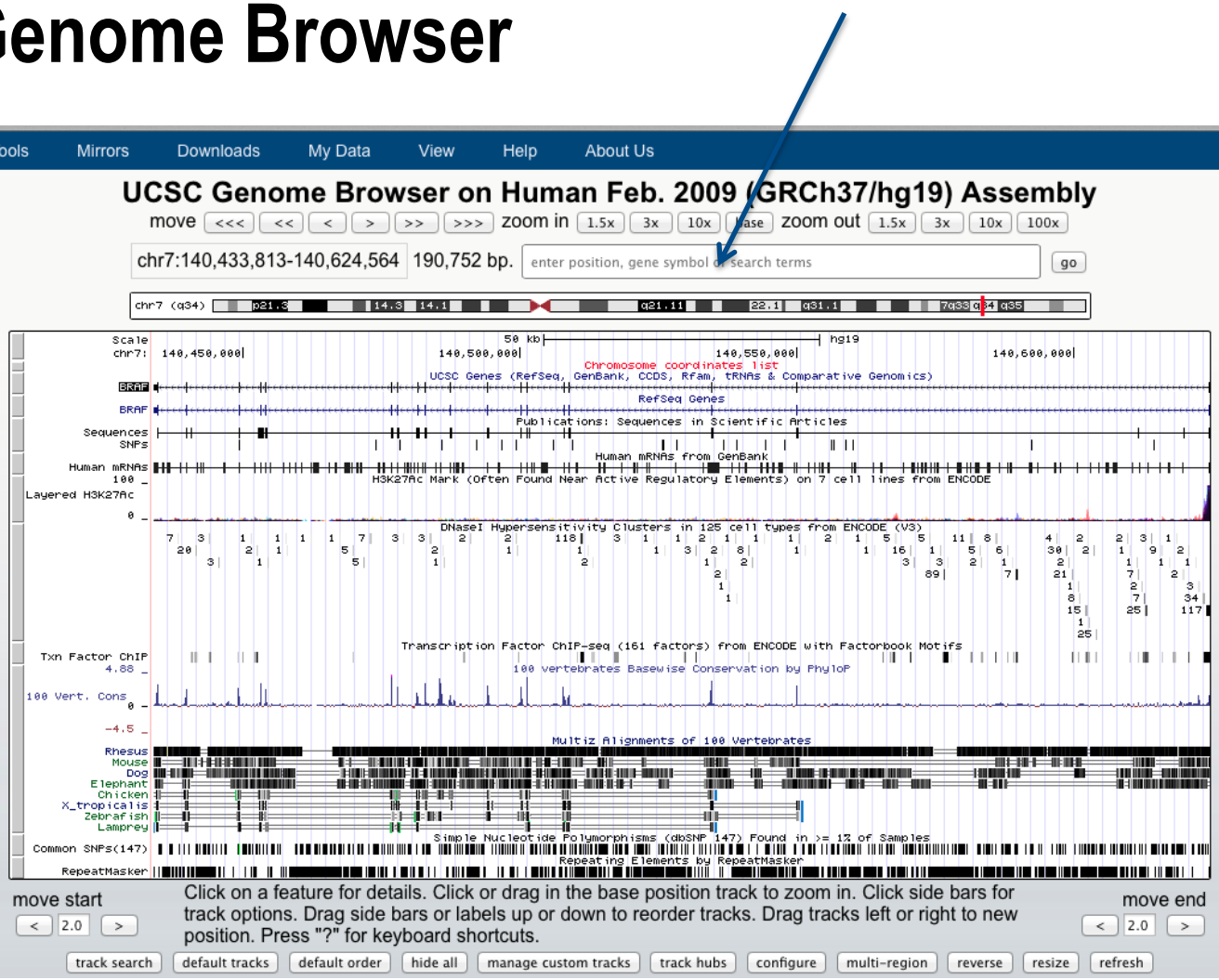
Choose species, assembly, and go to Genome Browser

NOTE! Genomic positions differ between assemblies

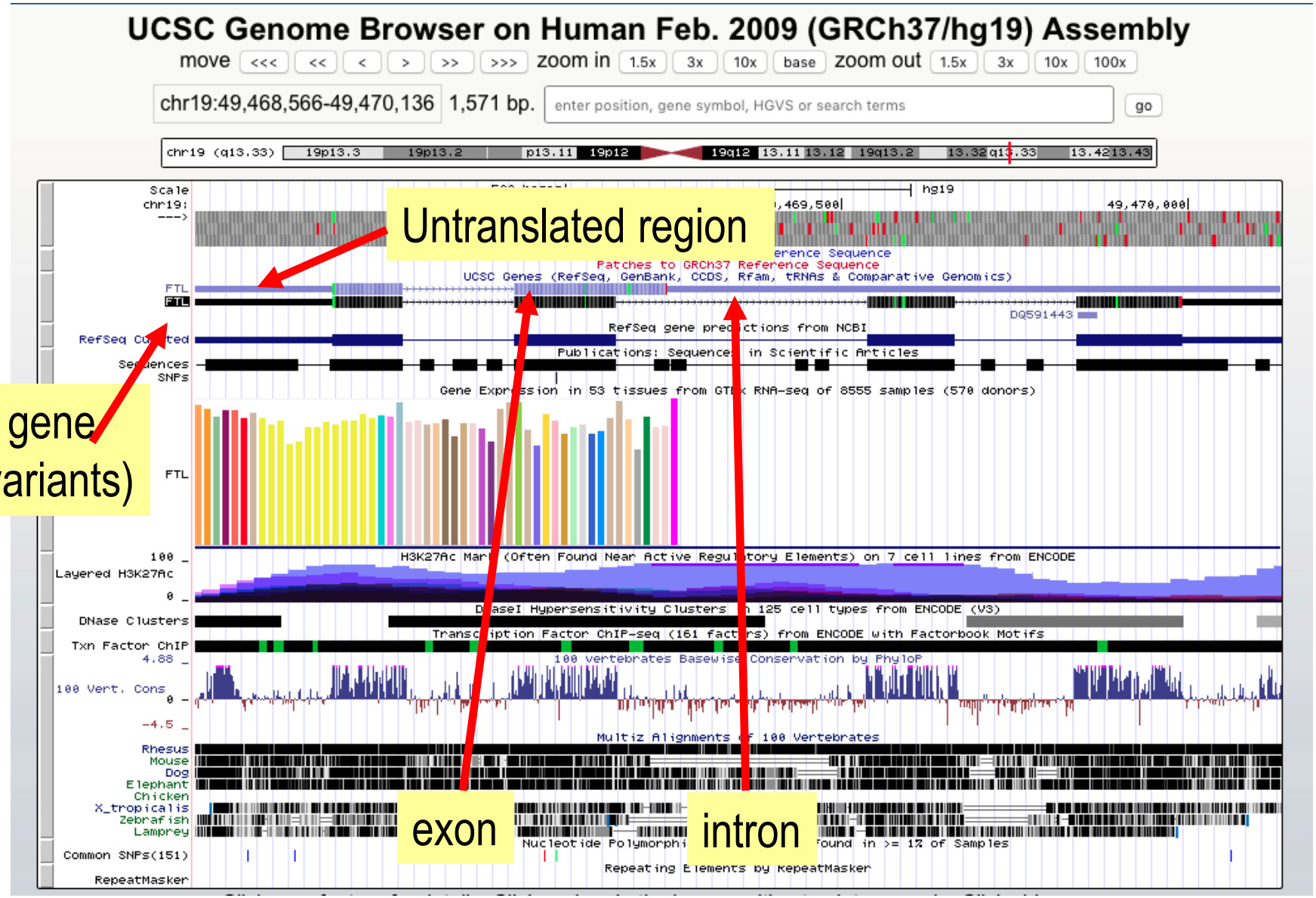
GÖTEBORGS UNIVERSITET

Bioinformatics
Core Facility

UCSC Genome Browser

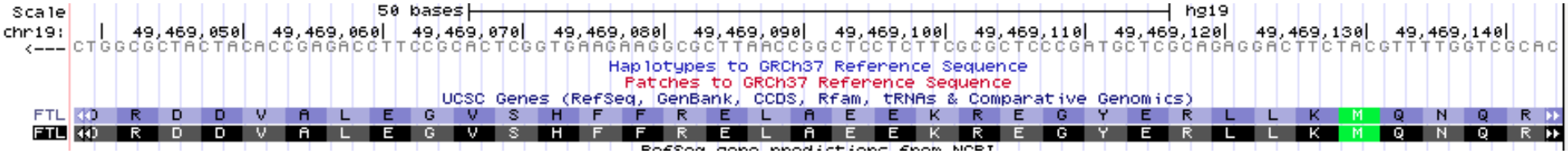


UCSC Genome Browser

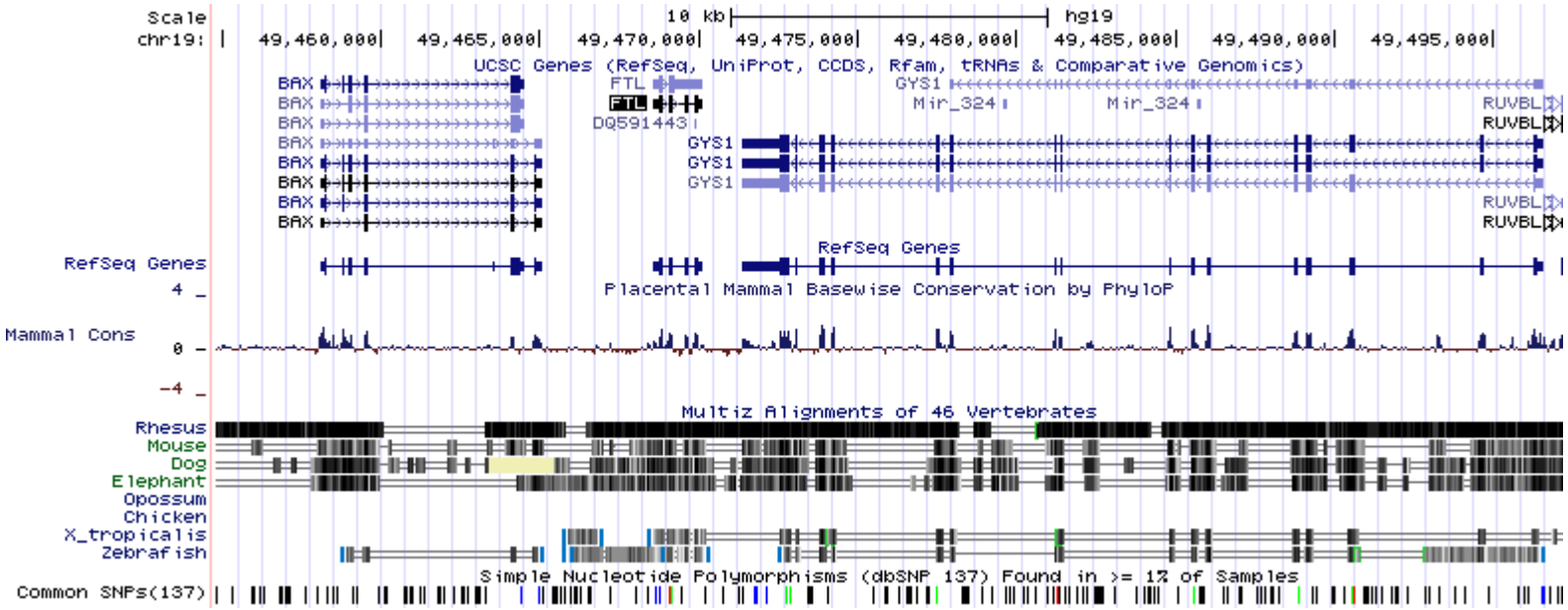


UCSC Genome Browser

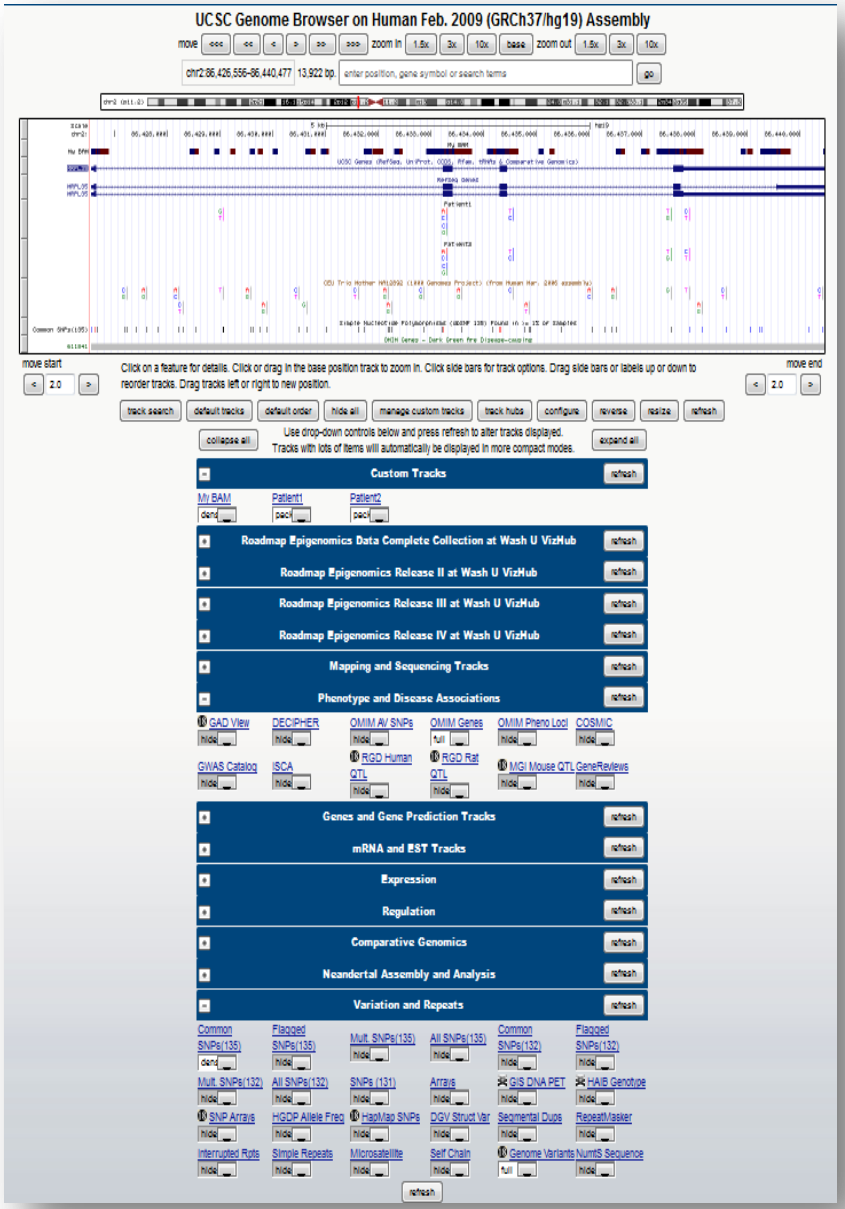
Zoom in:



Zoom out:



UCSC Genome Browser



Genome viewer

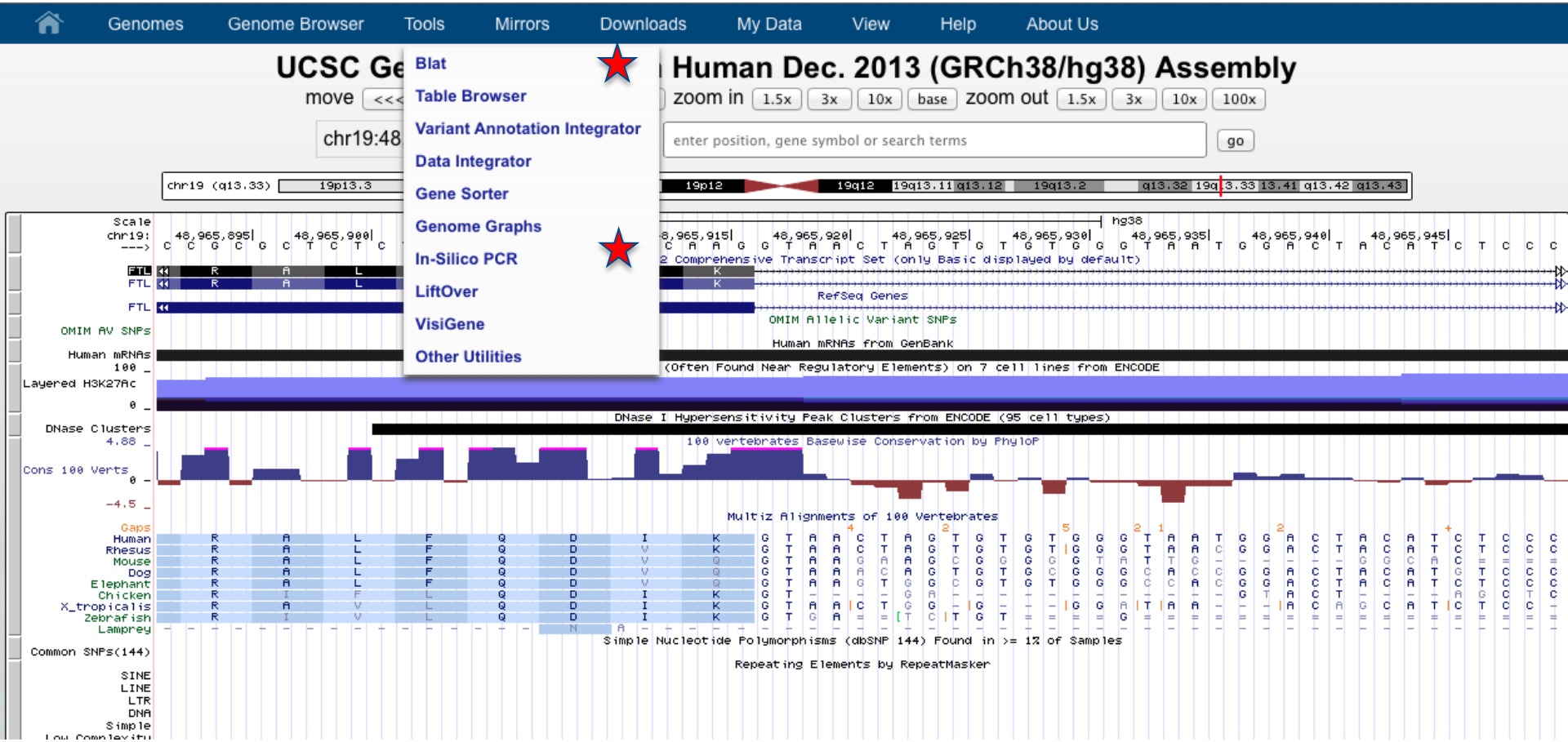
Custom Annotation tracks
(groups of data)

- Mapping and sequencing
- Gene predictions
- Variation and Repeats
- Comparative genomics
- Custom tracks (BAM, VCF)

GÖTEBORGS UNIVERSITET

Bioinformatics Core Facility

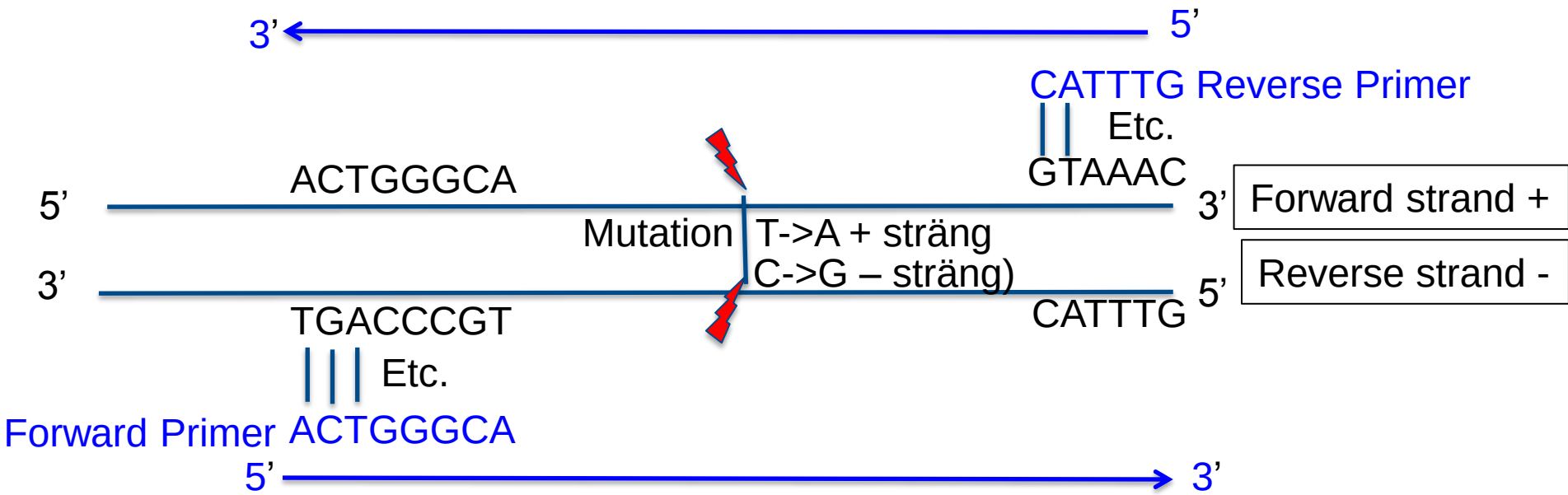
UCSC Genome Browser



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

Polymerase Chain Reaction (PCR)

Bioinformatics
Core Facility



Polymeras läser DNA mall i 3'-> 5' men bildar ny sträng med riktning 5'->3'

Männsikan läser alltid i riktning 5' -> 3'

Alltså läser vi forward primer: ACTGGGCA
Men reverse primer: GTTTAC

OBS! Det är viktigt att man hittar primers som binder unikt till ett ställe i genomet

Sammanfattning

- Bioinformatik är ett tvärvetenskapligt fält i rörelse
- Central dogma:
  - Replikation, Transkription, Translation
- Online verktyget Blast
- Identifiering av SNPs
- Genome Browsers
- PCR för att snabbt identifiera kända mutationer i patienter