“Bringing Bioinformatics into the Biology Classroom”

Marie-Claude Blatter & Patricia Palagi

Marie-Claude.Blatter@isb-sib.ch

SIB Swiss Institute of Bioinformatics

SIB Swiss Institute of Bioinformatics

- academic, non-profit foundation established in 1998
- coordinates **research** and **education** in bioinformatics throughout Switzerland
- provides high quality **bioinformatics services** to the national and international research community.
- helps shape the future of life sciences
- 52 groups, more than 600 scientists
- GOBLET member from the beginning

Global Organisation for Bioinformatics Learning, Education & Training

SIB
Swiss Institute of
Bioinformatics

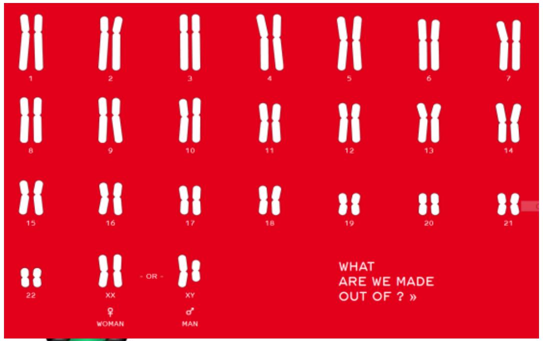
SIB outreach activities

education & public at large

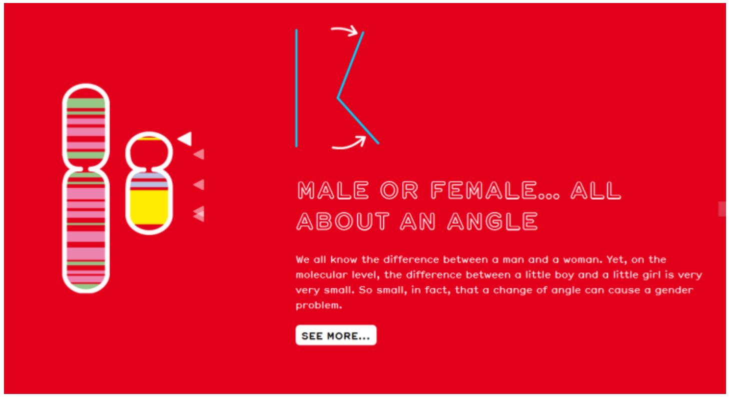
- Since 2000 (science fairs, electronic publication, exhibitions, hands on workshops, high school (HS) teacher continuing education training, etc.)
- Collaboration with public laboratories, didacticians and HS teachers
- **www.chromosomewalk.ch** (EN, FR, DE)
- **Protein Spotlight** (EN): http://web.expasy.org/spotlight/
- **'Ateliers de bioinformatique'** (FR): http://education.expasy.org/bioinformatique/
- *New project*: Drug Design and personalized medicine

www.ChromosomeWalk.ch

- **a saunter along the human genome**
- …take a walk and discover the world of **genes**, **proteins** and **bioinformatics**.
- quizzes, videos, links to databases and bioinformatics tools



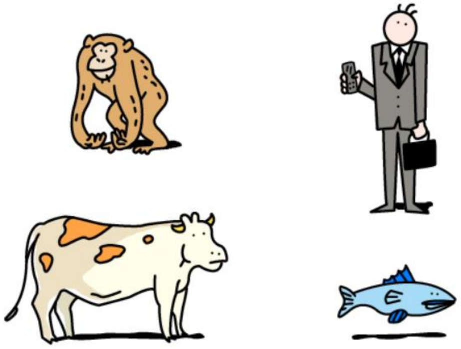
WHAT ARE WE MADE OUT OF ? »



MALE OR FEMALE... ALL ABOUT AN ANGLE

We all know the difference between a man and a woman. Yet, on the molecular level, the difference between a little boy and a little girl is very very small. So small, in fact, that a change of angle can cause a gender problem.

SEE MORE...



```
1   LWPPPPARAFVN
2   LWGPDPASAFVN
3   LWGPDPAAAFVN
4   FSGPGTSYAAAN
```

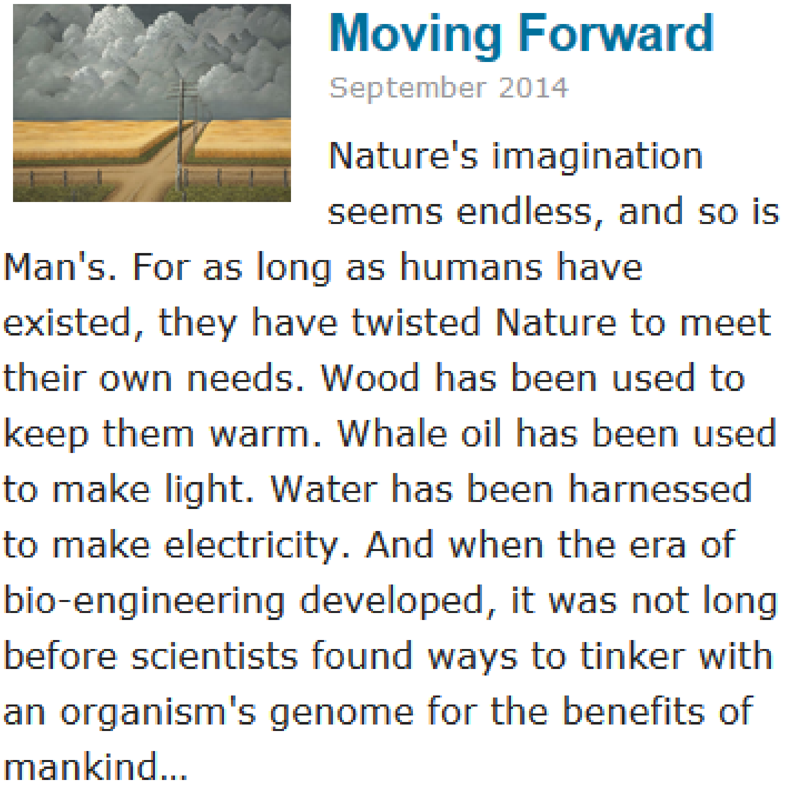
Global Organisation for Bioinformatics Learning, Education & Training

proteinspotlight

> ONE MONTH, ONE PROTEIN <

- http://web.expasy.org/spotlight/
- above 160 articles, informal tone (V. Gerritsen)

> *«The German inventor Nikolaus Otto is credited with having invented the first automobile engine that ran on alcohol.»*



Moving Forward

September 2014

Nature's imagination seems endless, and so is Man's. For as long as humans have existed, they have twisted Nature to meet their own needs. Wood has been used to keep them warm. Whale oil has been used to make light. Water has been harnessed to make electricity. And when the era of bio-engineering developed, it was not long before scientists found ways to tinker with an organism's genome for the benefits of mankind...

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics

‘Ateliers de Bioinformatique’

http://education.expasy.org/bioinformatique/ (FR)

Understanding a genetic disease thanks to Bioinformatics

http://education.expasy.org/bioinformatique/Diabetes.html
(Atelier 7: L’insuline de A à Z ; English version)

additional documents are available here:
http://education.expasy.org/cours/Toronto

Global Organisation for Bioinformatics Learning, Education & Training

SIB
Swiss Institute of Bioinformatics

Bioinformatics

This field of science designs software tools and databases for research in the life sciences.

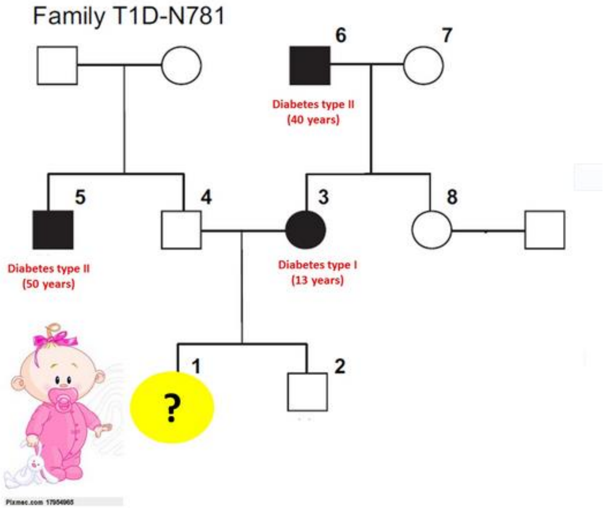
Today, the quantity of biological data accumulated by laboratories is daunting.

As a result, the data can no longer be dealt with ‘manually’ and **bioinformatics** has become an essential ally.

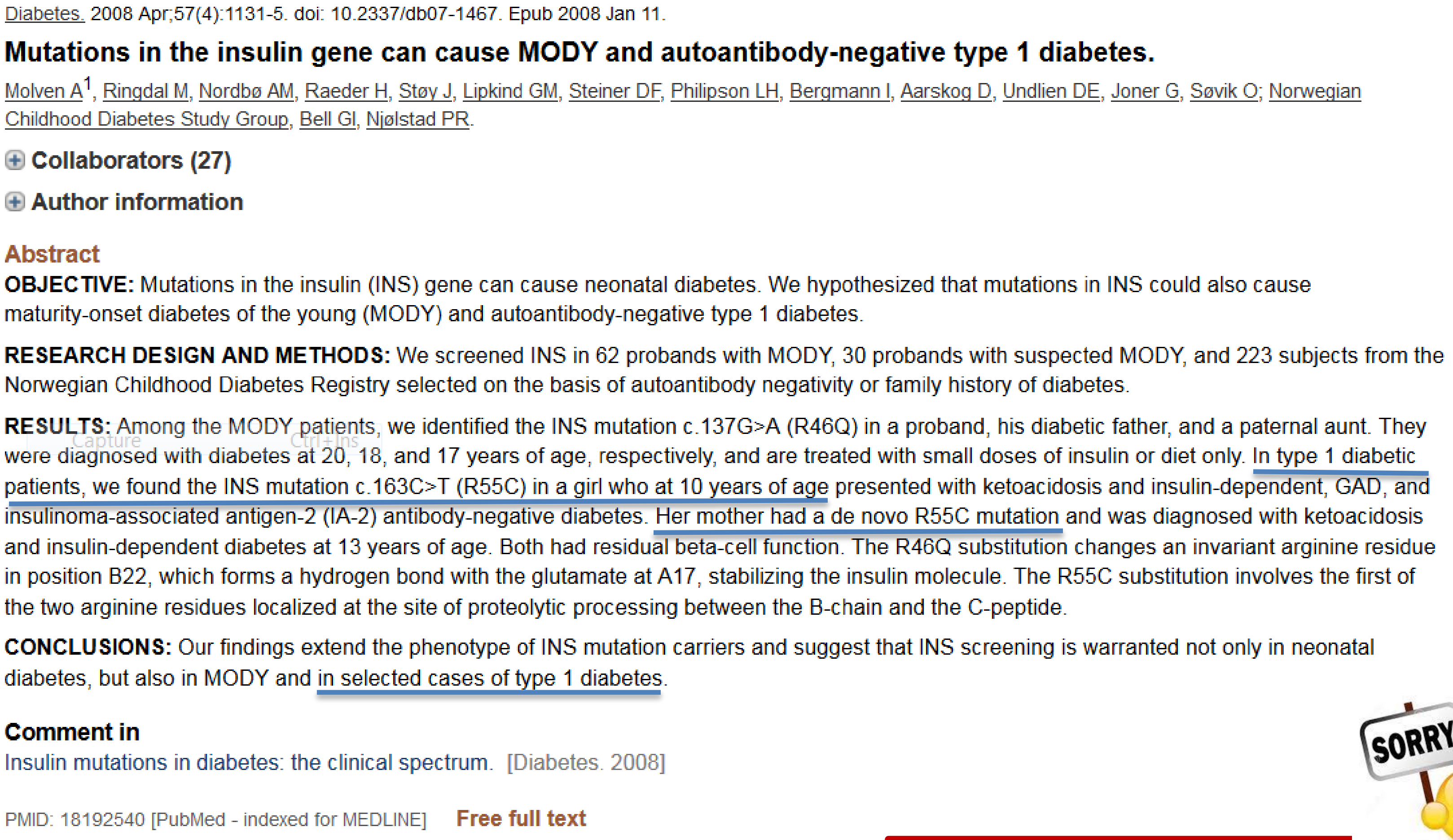
http://www.chromosomewalk.ch/en/we-need-bioinformatics-to/

Context

In a **special case** of type I diabetes described in a Norwegian family, a genetic variation has been found, leading to the production of inactive insulin



http://www.ncbi.nlm.nih.gov/pubmed/18192540



Diabetes. 2008 Apr;57(4):1131-5. doi: 10.2337/db07-1467. Epub 2008 Jan 11.

Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes.

Molven A[1], Ringdal M, Nordbø AM, Raeder H, Støy J, Lipkind GM, Steiner DF, Philipson LH, Bergmann I, Aarskog D, Undlien DE, Joner G, Søvik O; Norwegian Childhood Diabetes Study Group, Bell GI, Njølstad PR.

⊕ Collaborators (27)

⊕ Author information

Abstract

OBJECTIVE: Mutations in the insulin (INS) gene can cause neonatal diabetes. We hypothesized that mutations in INS could also cause maturity-onset diabetes of the young (MODY) and autoantibody-negative type 1 diabetes.

RESEARCH DESIGN AND METHODS: We screened INS in 62 probands with MODY, 30 probands with suspected MODY, and 223 subjects from the Norwegian Childhood Diabetes Registry selected on the basis of autoantibody negativity or family history of diabetes.

RESULTS: Among the MODY patients, we identified the INS mutation c.137G>A (R46Q) in a proband, his diabetic father, and a paternal aunt. They were diagnosed with diabetes at 20, 18, and 17 years of age, respectively, and are treated with small doses of insulin or diet only. In type 1 diabetic patients, we found the INS mutation c.163C>T (R55C) in a girl who at 10 years of age presented with ketoacidosis and insulin-dependent, GAD, and insulinoma-associated antigen-2 (IA-2) antibody-negative diabetes. Her mother had a de novo R55C mutation and was diagnosed with ketoacidosis and insulin-dependent diabetes at 13 years of age. Both had residual beta-cell function. The R46Q substitution changes an invariant arginine residue in position B22, which forms a hydrogen bond with the glutamate at A17, stabilizing the insulin molecule. The R55C substitution involves the first of the two arginine residues localized at the site of proteolytic processing between the B-chain and the C-peptide.

CONCLUSIONS: Our findings extend the phenotype of INS mutation carriers and suggest that INS screening is warranted not only in neonatal diabetes, but also in MODY and in selected cases of type 1 diabetes.

Comment in

Insulin mutations in diabetes: the clinical spectrum. [Diabetes. 2008]

PMID: 18192540 [PubMed - indexed for MEDLINE] **Free full text**

This publication is not available as free 'full text' in PubMed Central (PMC).
For full text:
http://education.expasy.org/cours/Toronto/

Global Organisation for Bioinformatics Learning, Education & Training

http://education.expasy.org/bioinformatique/Diabetes.html

Activity 1: The insulin gene and the human genome

(Genome browser (USCS), BLAT)

Activity 2: Comparing DNA sequences - Diagnosing a rare genetic disease

(alignment tool, database dbSNP)

Activity 3: DNA translation -> protein

(translate tool)

Activity 4: 3D structure of insulin

(database PDB, 3D visualization tool)

Activity 5: Is insulin specific to humans?

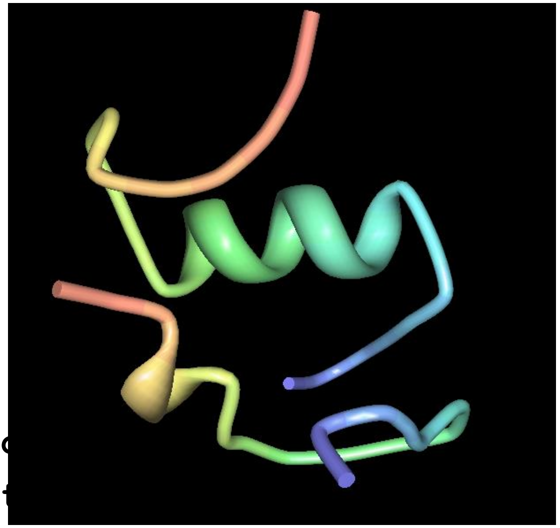
(similarity search (BLAST), dabatase UniProtKB, alignment tool)

Global Organisation for Bioinformatics Learning, Education & Training

&

http://thumbs.dreamstime.com/z/cartoon-man-working-computer-13780903.jpg

biological function



protein
(amino acid)

MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVE
ALYLVCGERGFFYTPKTCREAEDLQVGQVELGGGPGA
GSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

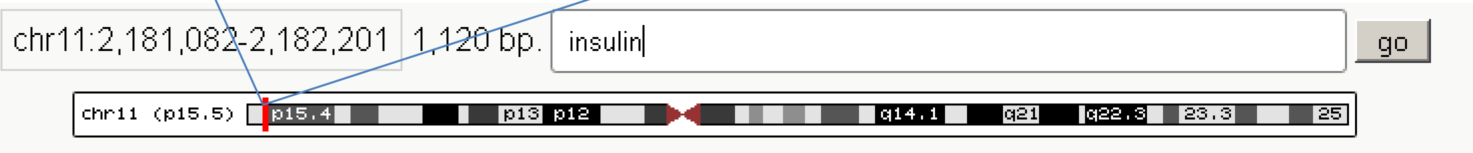
gene
(DNA; nucleic acid)

atggccctgtggatgcgcctcct cgtgctggcg
ccagccgcagcctttgtga gtgcggct

mutation

tagtgtgcggggaacgag acacacccaagacccgtcgggaggcagaggac
tgcaggtggggcaggtggagctgggcgggggccctggtgcaggcagcctgcagcccttg
gccctggaggggtccctgcagaagcgtggcattgtggaacaatgctgtaccagcatctgc
tccctctaccagctggagaactactgcaactag

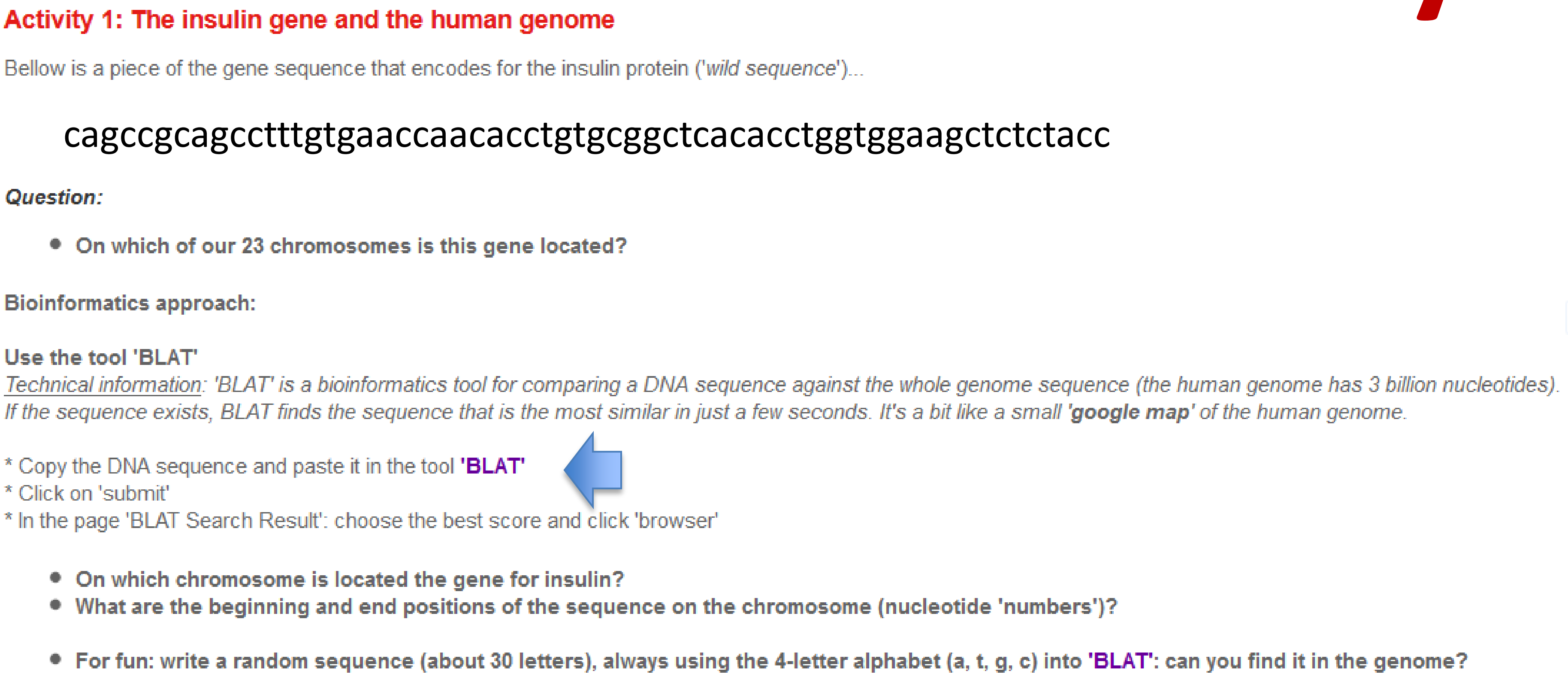
genome



chr11:2,181,082-2,182,201 1,120 bp. insulin

Global Organisation for Bioinformatics Learning, Education & Training

Activity 1



Activity 1: The insulin gene and the human genome

Bellow is a piece of the gene sequence that encodes for the insulin protein ('*wild sequence*')...

cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc

Question:

- **On which of our 23 chromosomes is this gene located?**

Bioinformatics approach:

Use the tool 'BLAT'
Technical information: 'BLAT' is a bioinformatics tool for comparing a DNA sequence against the whole genome sequence (the human genome has 3 billion nucleotides). If the sequence exists, BLAT finds the sequence that is the most similar in just a few seconds. It's a bit like a small ***'google map'*** *of the human genome.*

* Copy the DNA sequence and paste it in the tool **'BLAT'**
* Click on 'submit'
* In the page 'BLAT Search Result': choose the best score and click 'browser'

- **On which chromosome is located the gene for insulin?**
- **What are the beginning and end positions of the sequence on the chromosome (nucleotide 'numbers')?**

- **For fun: write a random sequence (about 30 letters), always using the 4-letter alphabet (a, t, g, c) into 'BLAT': can you find it in the genome?**

http://education.expasy.org/bioinformatique/Diabetes.html

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics approach:

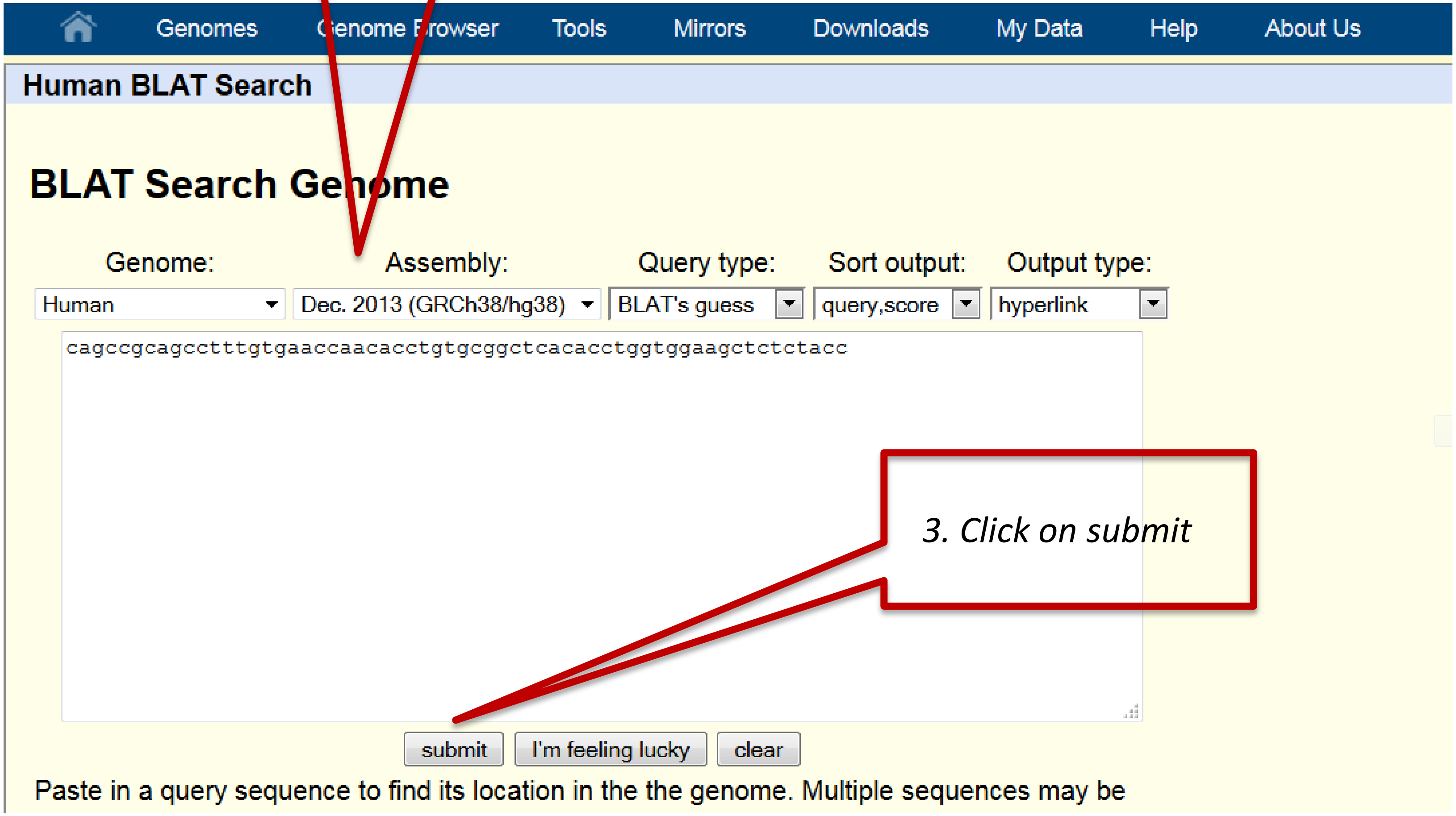
Use the tool 'BLAT‘ @ USCS

Technical information: 'BLAT' is a bioinformatics tool for comparing a DNA sequence against a whole genome sequence.

*If the sequence exists, BLAT finds the sequence that is the most similar in just a few seconds. It's a bit like a small **'google map'** of the human genome.*

1. *Google: look for 'BLAT UCSC'*

2. *Choose the latest release of the human genome (GRCh38)*



Human BLAT Search

BLAT Search Genome

Genome: Human

Assembly: Dec. 2013 (GRCh38/hg38)

Query type: BLAT's guess

Sort output: query,score

Output type: hyperlink

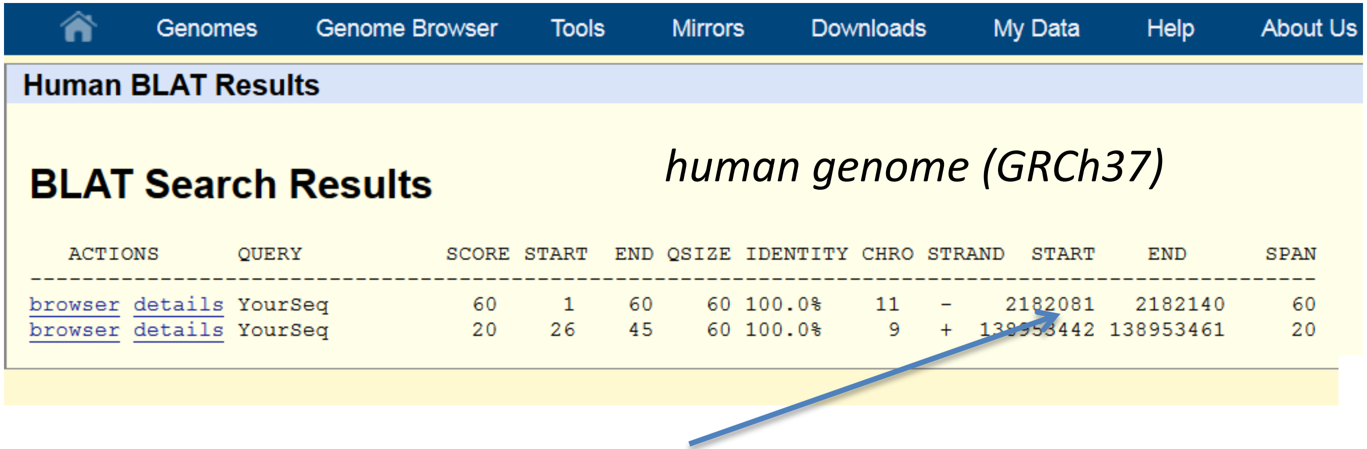
```
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
```

Paste in a query sequence to find its location in the the genome. Multiple sequences may be

3. *Click on submit*

Bioinformatics – Genome Browser (Blat USCS)

http://genome.ucsc.edu/cgi-bin/hgBlat



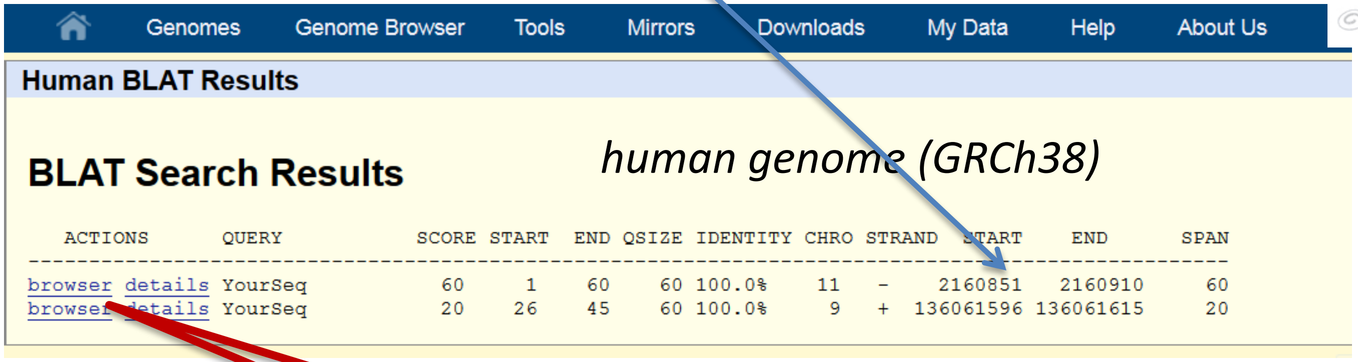
Human BLAT Results

BLAT Search Results

human genome (GRCh37)

```
   ACTIONS      QUERY           SCORE START  END QSIZE IDENTITY CHRO STRAND  START    END      SPAN
---------------------------------------------------------------------------------------------------
browser details YourSeq           60     1    60    60 100.0%    11   -     2182081   2182140    60
browser details YourSeq           20    26    45    60 100.0%     9   +   138953442 138953461    20
```

At each genome release the positions may change



Human BLAT Results

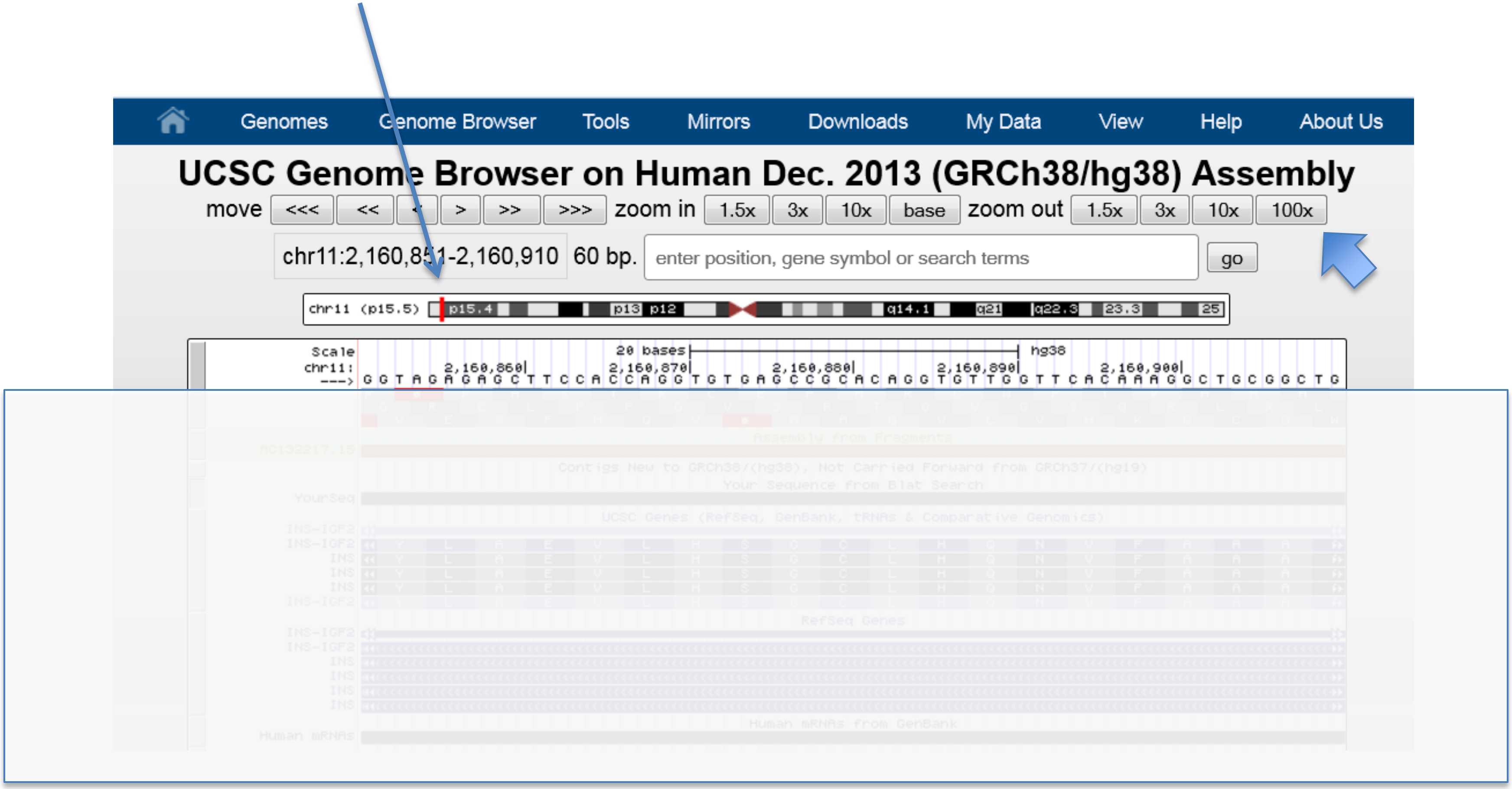
BLAT Search Results

human genome (GRCh38)

```
   ACTIONS      QUERY           SCORE START  END QSIZE IDENTITY CHRO STRAND  START    END      SPAN
---------------------------------------------------------------------------------------------------
browser details YourSeq           60     1    60    60 100.0%    11   -     2160851   2160910    60
browser details YourSeq           20    26    45    60 100.0%     9   +   136061596 136061615    20
```

By default, choose the best score
Click on 'browser'

The insulin DNA sequence is located on **chromosome 11 (11p15.5)**
(positions: 2,160,851-2’160,910 (GRCh38))


Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
go
enter position, gene symbol or search terms
chr11 (p15.5) p15.4 p13 p12 q14.1 q21 q22.3 23.3 25
Scale
20 bases
hg38
Assembly from Fragments
AC132217.15
Contigs New to GRCh38/(hg38), Not Carried Forward from GRCh37/(hg19)
Your Sequence from Blat Search
YourSeq
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
INS-IGF2
INS
RefSeq Genes
Human mRNAs from GenBank

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

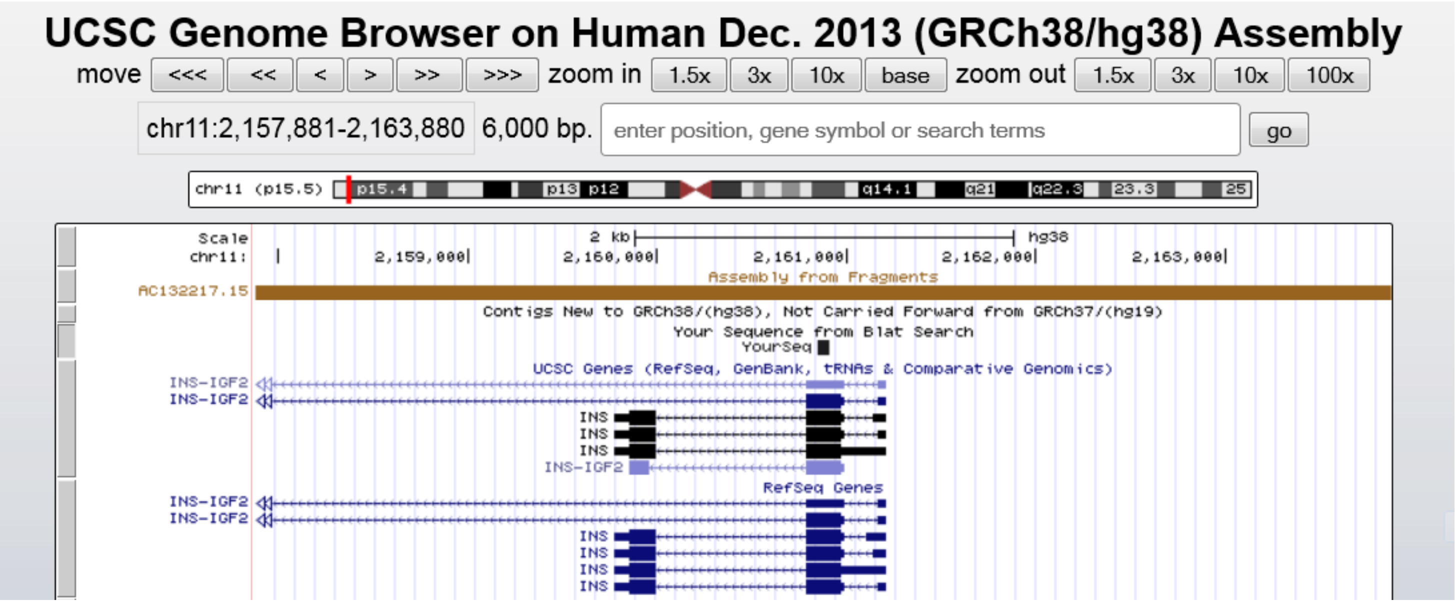
chr11:2,160,851-2,160,910 60 bp.

```
         2,160,860|          2,160,870|          2,160,880|          2,160,890|          2,160,900|
chr11:
---> G G T A G A G A G C T T C C A C C A G G T G T G A G C C G C A C A G G T G T T G G T T C A C A A A G G C T G C G G C T G
```

Zoom out 100 x

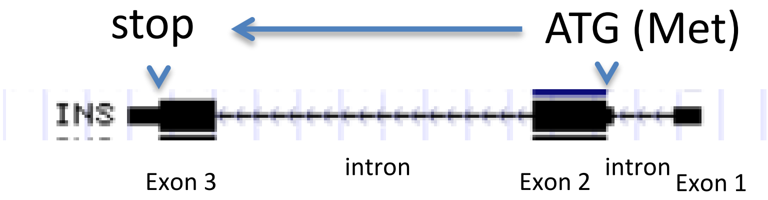
Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics



UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

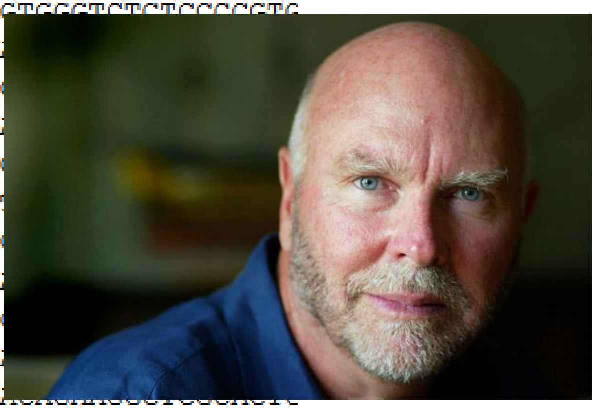
chr11:2,157,881-2,163,880 6,000 bp.



The insulin gene consists of 3 exons and 2 introns

A readthrough transcript INS-IGF2 involves INS and IGF2 genes (neighboring genes)

```
TCCACACGCTCCTTGCGGCCTCATGGGTGTAGGGTCCAGCCCCACAGGGTCGGTGGGTCTCTCCCCGTC
GCAGAGACGAGAGAGTGTAGAAATAAAGACACAAGACAAAGAGATTAAAAAA
GGACCACTACCACCAATGCGCGGAGACCGGTAGTGGCCCCGAATGTCTGGCT
TACAAGGCAAAAGGGGCAGGGTAAAGAGTGTGAGTCATCTCCAATGATAGAT
TGTCCACTGGACAGGGGGCCCTTCCCTGCCTGGCAGCCGAGGCAGAGAGGGA
ATAGCTTACGCCATTATTTTTGTATATTAGAGACGTTTAGTACTTTCACTAA
AAGGCAGAGCCAGGTGCACAGGATGGAACATGAAGGAGGACTAGGAGCGTGA
ACAGGGAGACGGTTAGGCCTCCGGATAACTGCGGGCAGGTCTGACTGATGTC
GGAGGAGCAGAGTCTTCTCTAAACTCCCCCGGGGAAAGGGAGCCCCTCCTTT
GGGTGTTTTTCCTTGACACTTACGCTACCGCTAGACCACGGTCCGCTTGGCA
GCTGGCATCACCGCTAGACCAAGGAGCCCTCTAGTGGCCTTGTCCGGGCATA
TGTCTTCTGGTCACTCCTCACTATGTCCCCTCAGCTCCTATCTCTGTATGGCCTGGTTTTTCCTAGTTT
TGATTATAGAGCGAGGATTGTTATAATATTGGAATAAAGAGTAATTGCTACAAACTAATGATTAATGAT
TTCATATATAATCATATCTAAGATCTATATCTGGTGTAACTATTTTTATTTTATATTTTATTATACTGG
ACAGCTCGTGTCCTCAGTCTCTTGCCTCGGCACCTGGGTGGCTTGCCGCCCACAATGGGCAGCTTATTC
TCAGGGAAGGCCTTTGTCTCCACACCTGTGGGTGAAGACCATCGGGATGCTTTGCCTTCAACAGGCAAG
CCAACAATTCACCTTCACTTCCCTCCCTCCAGGAACACCAGCTCCCAGCTCAGAGTCATCGGCCTCGCT
ACAGGGACGTCACACTACCCGCTCTGTGGGGGGCATCGTGTGGTCTGGACTTGCTGAGCAGAAAGTAGC
GCTGCCCTCAACACCTCCCTAGAGCATCTGCGAGCCGAACACCTGGGGCCCACGCCTCCGGCACGTCTA
GGACCCAGTGGTCCATCCCTTCCCAAGCACAAGGCAAGTGGCTACCTCAGTCCCTTCCTCCACGAAGAA
GAGGCACGATGCCTAGTGCTGTAGGTCCCATGTTATTTGGGAAGCAACTTTTGCCCTATTTGGAAGTGC
```

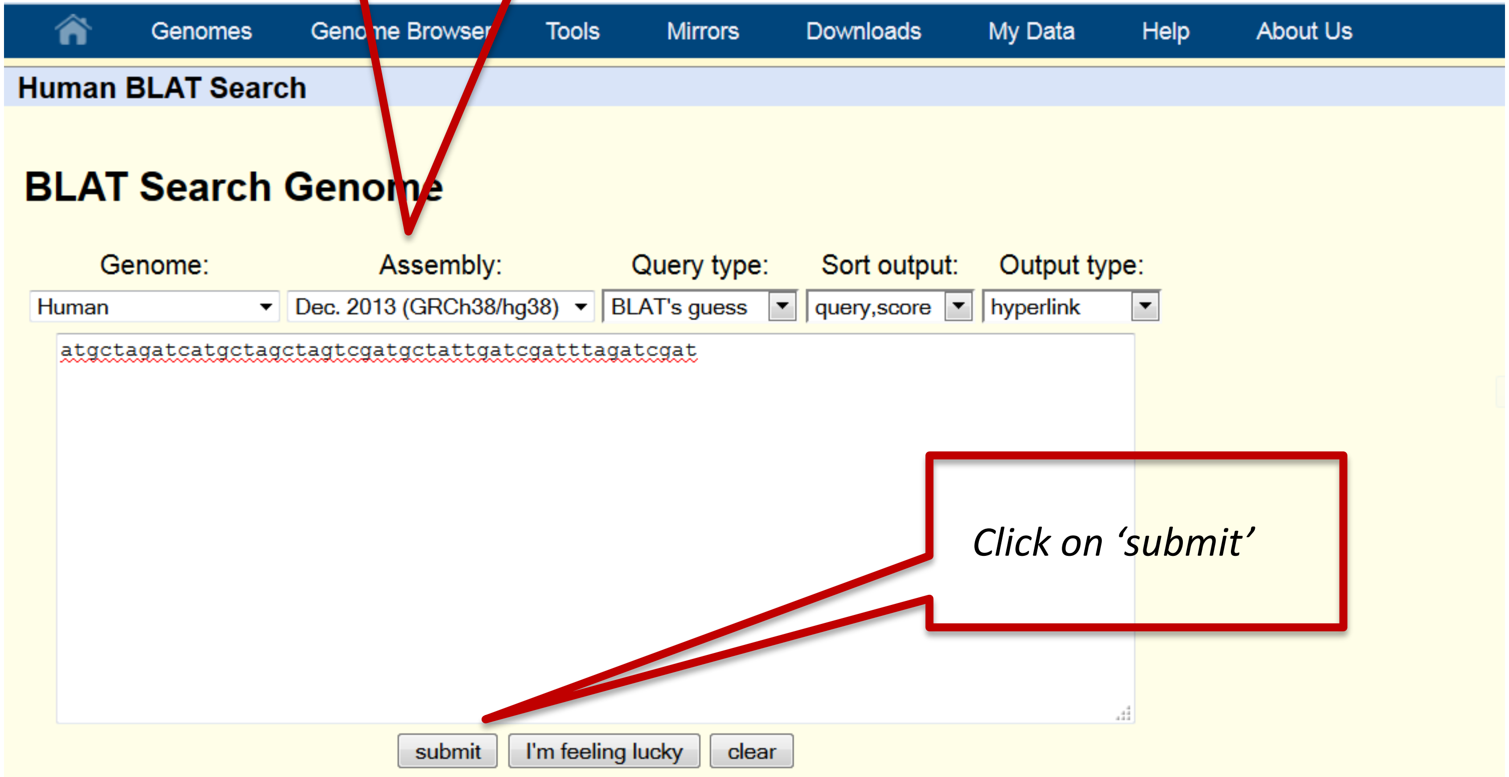


http://www.ncbi.nlm.nih.gov/nuccore/71514639?report=fasta

Click on the link
This is part of the sequence of Craig Venter chromosome 11
(GenBank database; 3’852’046 bp over 135'006'516 bp)

Select short sequences (about 40 bp)
and check with BLAT that they are located on chromosome 11 ….

Write a random sequence (about 30 letters), always using the 4-letter alphabet (a, t, g, c)



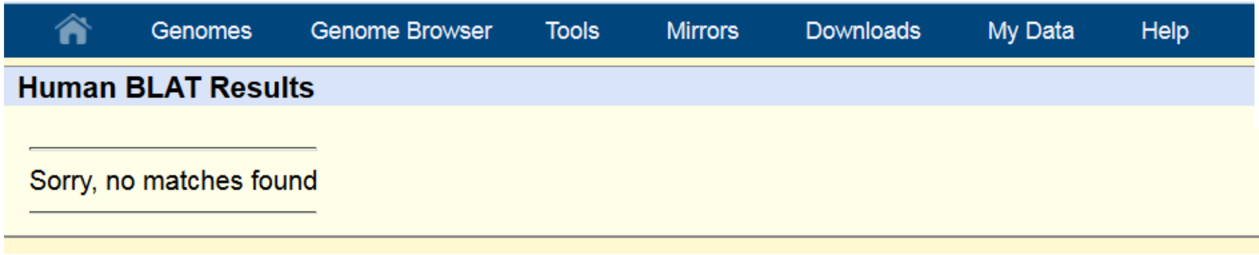
Human BLAT Search

BLAT Search Genome

Genome:	Assembly:	Query type:	Sort output:	Output type:
Human	Dec. 2013 (GRCh38/hg38)	BLAT's guess	query,score	hyperlink

```
atgctagatcatgctagctagtcgatgctattgatcgatttagatcgat
```

Click on 'submit'



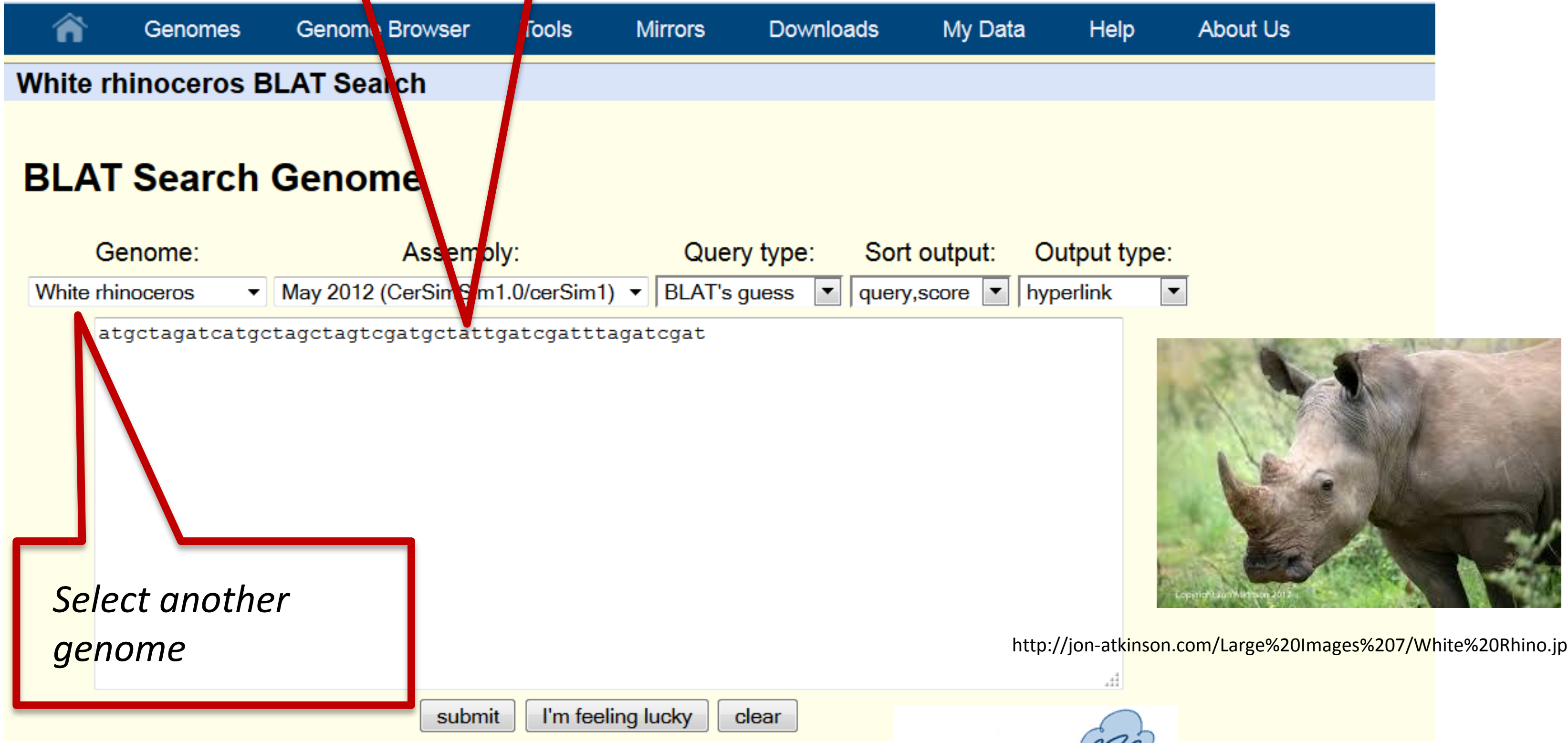
Human BLAT Results

Sorry, no matches found

It is virtually impossible to match a randomly typed sequence (ATGC, n=30) on the human genome sequence, even on «junk» DNA regions (Application: PCR and primer selection)

Randomly selected letters (i.e. n=5) rarely create a correct word….

Write a random sequence (about 30 letters), always using the 4-letter alphabet (a, t, g, c)



White rhinoceros BLAT Search

BLAT Search Genome

Genome:	Assembly:	Query type:	Sort output:	Output type:
White rhinoceros	May 2012 (CerSimSim1.0/cerSim1)	BLAT's guess	query,score	hyperlink

```
atgctagatcatgctagctagtcgatgctattgatcgatttagatcgat
```

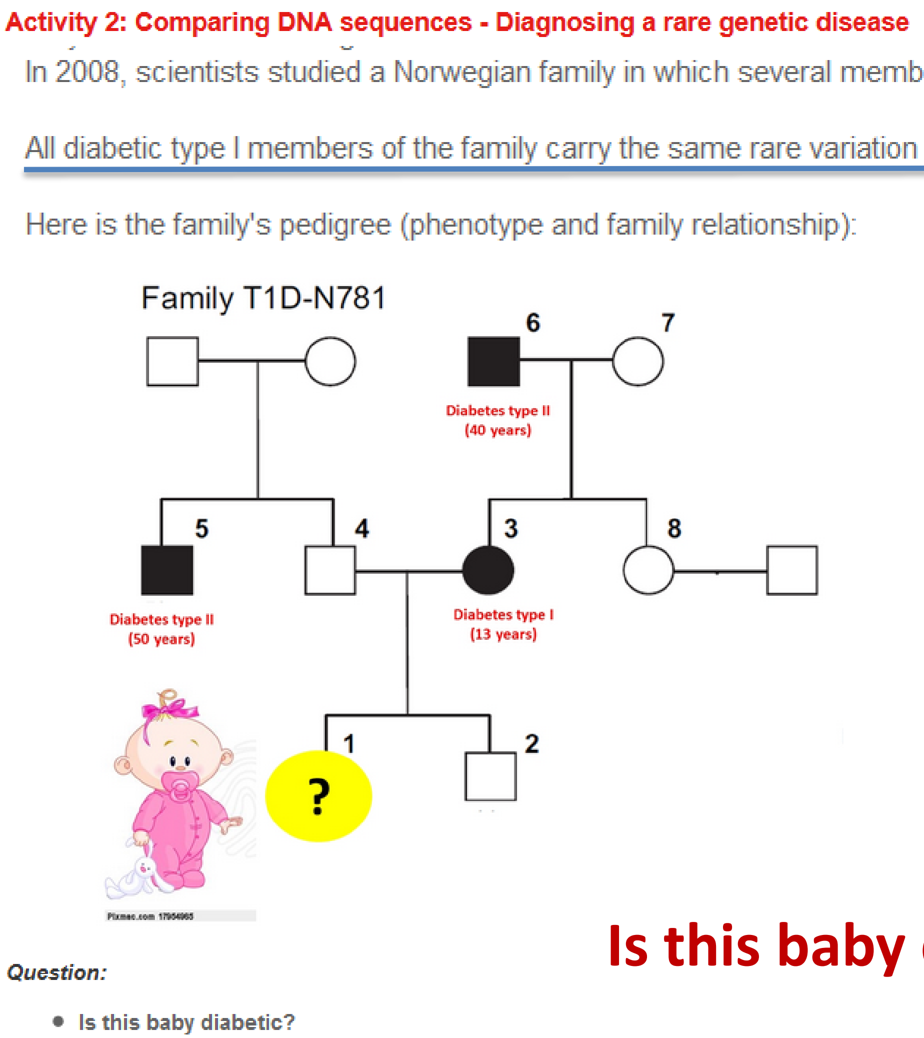
Select another genome

http://jon-atkinson.com/Large%20Images%207/White%20Rhino.jpg

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics

Activity 2



Activity 2: Comparing DNA sequences - Diagnosing a rare genetic disease

In 2008, scientists studied a Norwegian family in which several members had diabetes (type I or type II) (Molven et al., 2008).

All diabetic type I members of the family carry the same rare variation in the gene which encodes for insulin.

Here is the family's pedigree (phenotype and family relationship):

Question:

- **Is this baby diabetic?**

■	affected male
□	male
●	affected female
○	female

Is this baby diabetic (type I) ?

http://education.expasy.org/bioinformatique/Diabetes.html

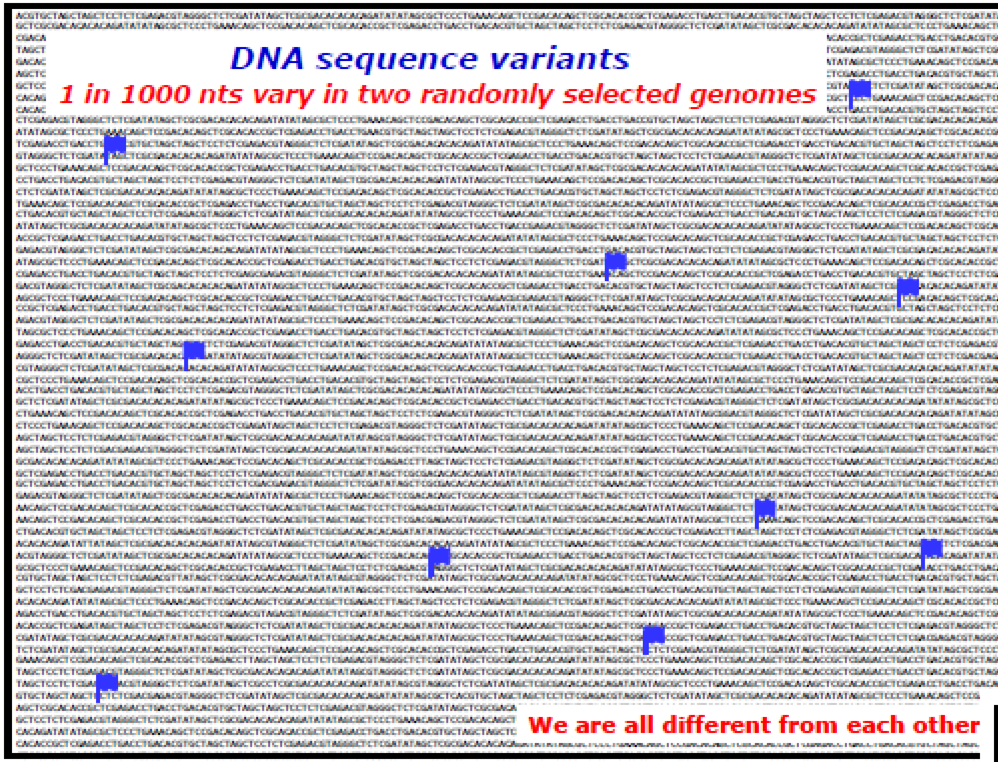
Global Organisation for Bioinformatics Learning, Education & Training

Biological context: Human genome & variations

The human genome = a text of 3’000’000’000 pb

= a reference sequence

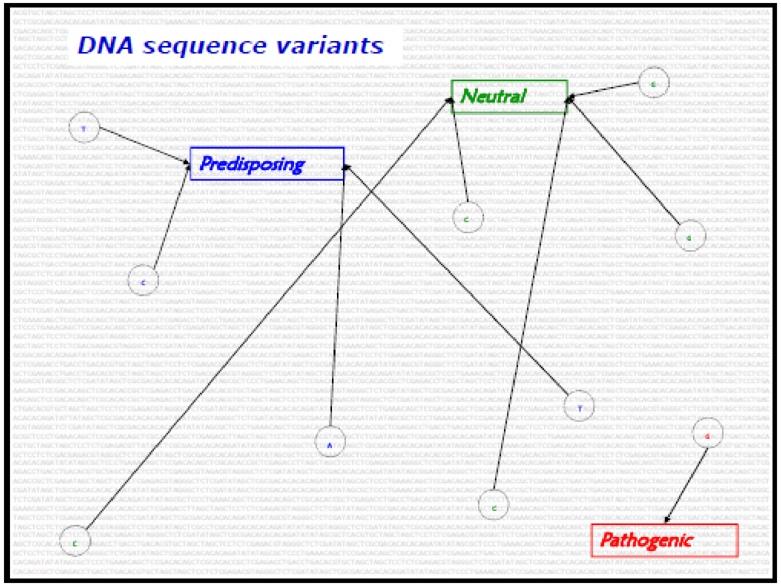
All the differences (also called *variations*, *variants*, *Single Nucleotide Polymorphisms (SNPs*), ~*mutations*, …) between human subjects are described on the basis of this ‘text’


DNA sequence variants
1 in 1000 nts vary in two randomly selected genomes
We are all different from each other

~ 3.3 millions of SNPs between 2 individuals

~10 millions of SNPs in the human population
neutral (the majority)
associated with a particular phenotype…
associated with a predisposition
associated with a genetic disease

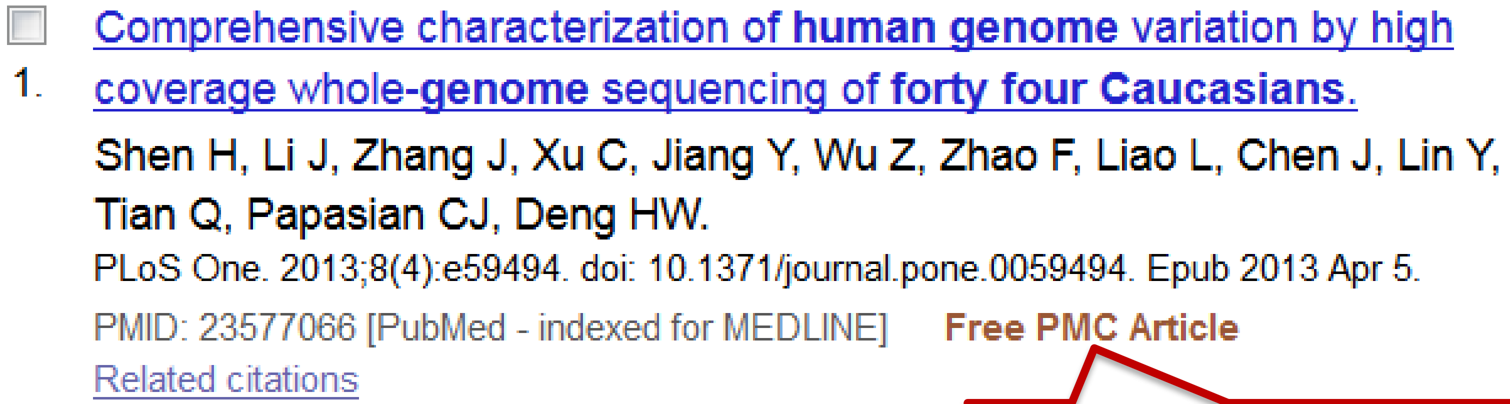
~10 – 30 new mutation at each new generation



Pr S.E. Antonarakis (UNIGE)

Global Organisation for Bioinformatics Learning, Education & Training

Biological context: Human genome & variations



□ 1. [Comprehensive characterization of **human genome** variation by high coverage whole-**genome** sequencing of **forty four Caucasians**.](#)

Shen H, Li J, Zhang J, Xu C, Jiang Y, Wu Z, Zhao F, Liao L, Chen J, Lin Y, Tian Q, Papasian CJ, Deng HW.

PLoS One. 2013;8(4):e59494. doi: 10.1371/journal.pone.0059494. Epub 2013 Apr 5.

PMID: 23577066 [PubMed - indexed for MEDLINE] **Free PMC Article**

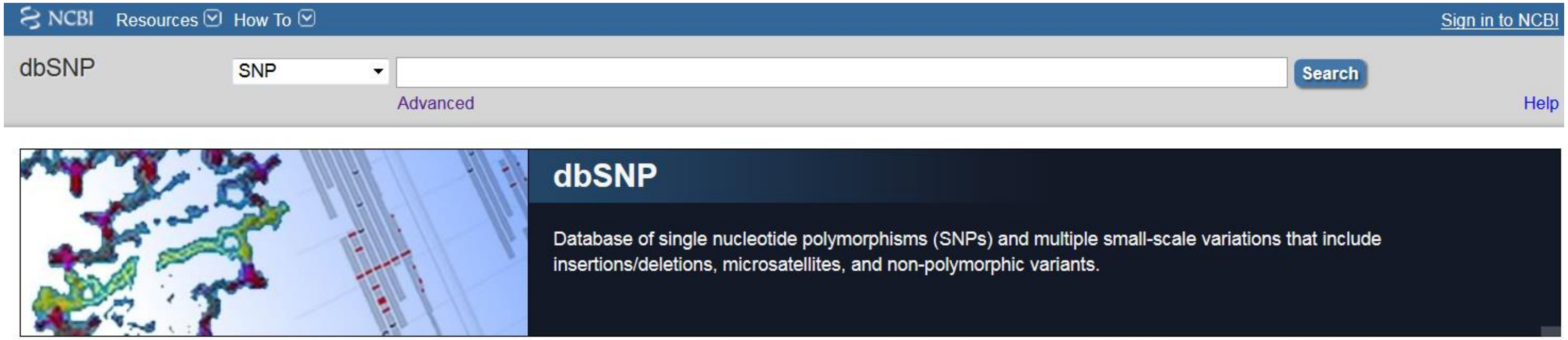
Related citations

Publication (free full text) in PubMed Central (PMC @NCBI) are freely available for everyone

"On average, each individual genome carried ~3.3 million SNPs and ~492,000 indels/block substitutions, including approximately 179 variants that were predicted to cause loss of function of the gene products. "

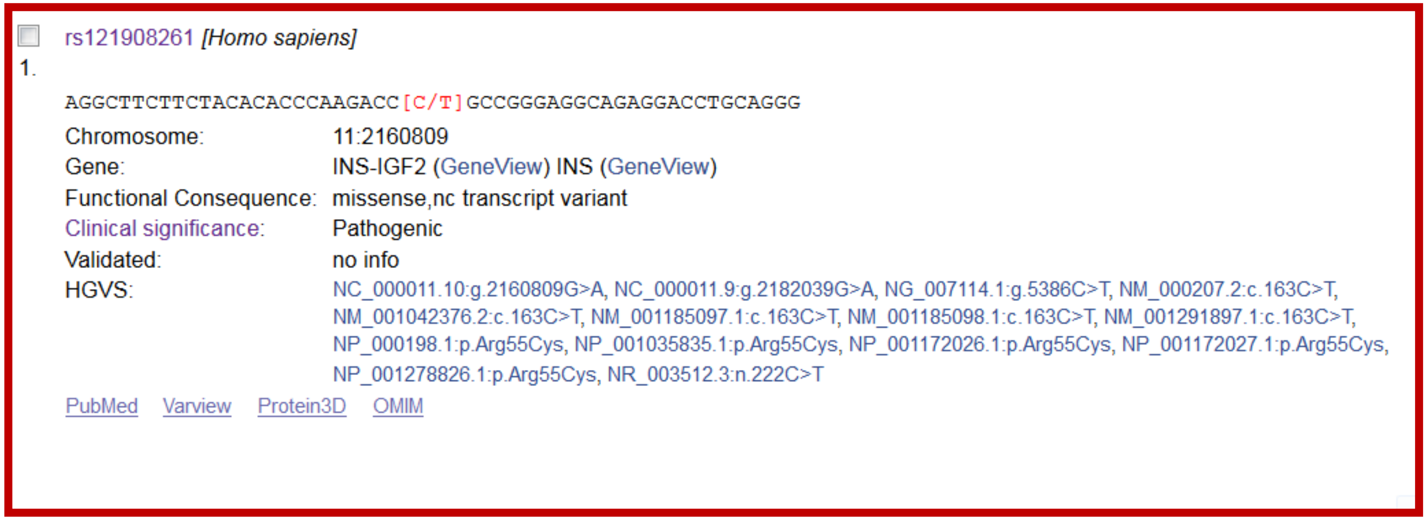
PMID: 23577066

SNPs are stored in the dbSNP database



dbSNP

Database of single nucleotide polymorphisms (SNPs) and multiple small-scale variations that include insertions/deletions, microsatellites, and non-polymorphic variants.

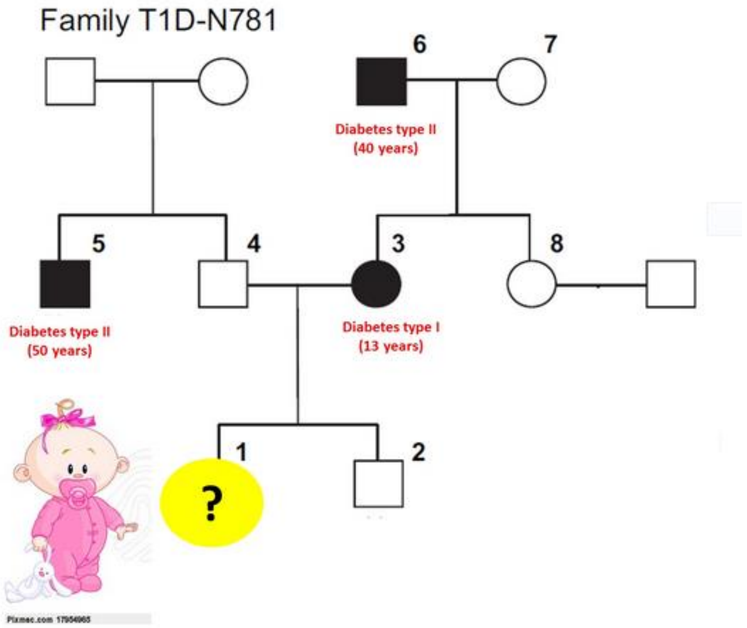


□ 1. rs121908261 *[Homo sapiens]*

```
AGGCTTCTTCTACACACCCAAGACC[C/T]GCCGGGAGGCAGAGGACCTGCAGGG
```

Chromosome:	11:2160809
Gene:	INS-IGF2 (GeneView) INS (GeneView)
Functional Consequence:	missense,nc transcript variant
Clinical significance:	Pathogenic
Validated:	no info
HGVS:	NC_000011.10:g.2160809G>A, NC_000011.9:g.2182039G>A, NG_007114.1:g.5386C>T, NM_000207.2:c.163C>T, NM_001042376.2:c.163C>T, NM_001185097.1:c.163C>T, NM_001185098.1:c.163C>T, NM_001291897.1:c.163C>T, NP_000198.1:p.Arg55Cys, NP_001035835.1:p.Arg55Cys, NP_001172026.1:p.Arg55Cys, NP_001172027.1:p.Arg55Cys, NP_001278826.1:p.Arg55Cys, NR_003512.3:n.222C>T

Bioinformatics – Biological Database: dbSNP @ NCBI
http://www.ncbi.nlm.nih.gov/SNP/



Question: is the baby diabetic ?

To answer this question, researchers extracted DNA from 8 members of the Norwegian family and sequenced part of the gene that encodes for insulin.

http://www2.grifil.com/album.html

Global Organisation for Bioinformatics Learning, Education & Training

Compare these sequences, and locate the common variation for diabetes.

'Paper and pencil' approach:

... You can do it **manually which will help you better understand the principle of sequence comparison and alignment.**
Take into account all the given clues and play with our strips of DNA sequences...

- 8 family members - 4 DNA sequences - one allele
- 8 family members - 1 DNA sequence - two alleles
- 8 family members - 2 DNA sequences - two alleles (not easy)

Bioinformatics approach:

Build an alignment of these 8 sequences using a bioinformatics tool and look out for the common variation among those with diabetes

* Copy these 8 sequences (including the lines starting with '>1') and paste them into the align tool
* Click on the *Run Align* button.
* On the results page, on the lefthand column 'Highlight': select 'Similarity'

'Paper and pencil' approach:

... You can do it **manually which will help you better understand the principle of sequence comparison and alignment.**

Take into account all the given clues and play with our strips of DNA sequences...

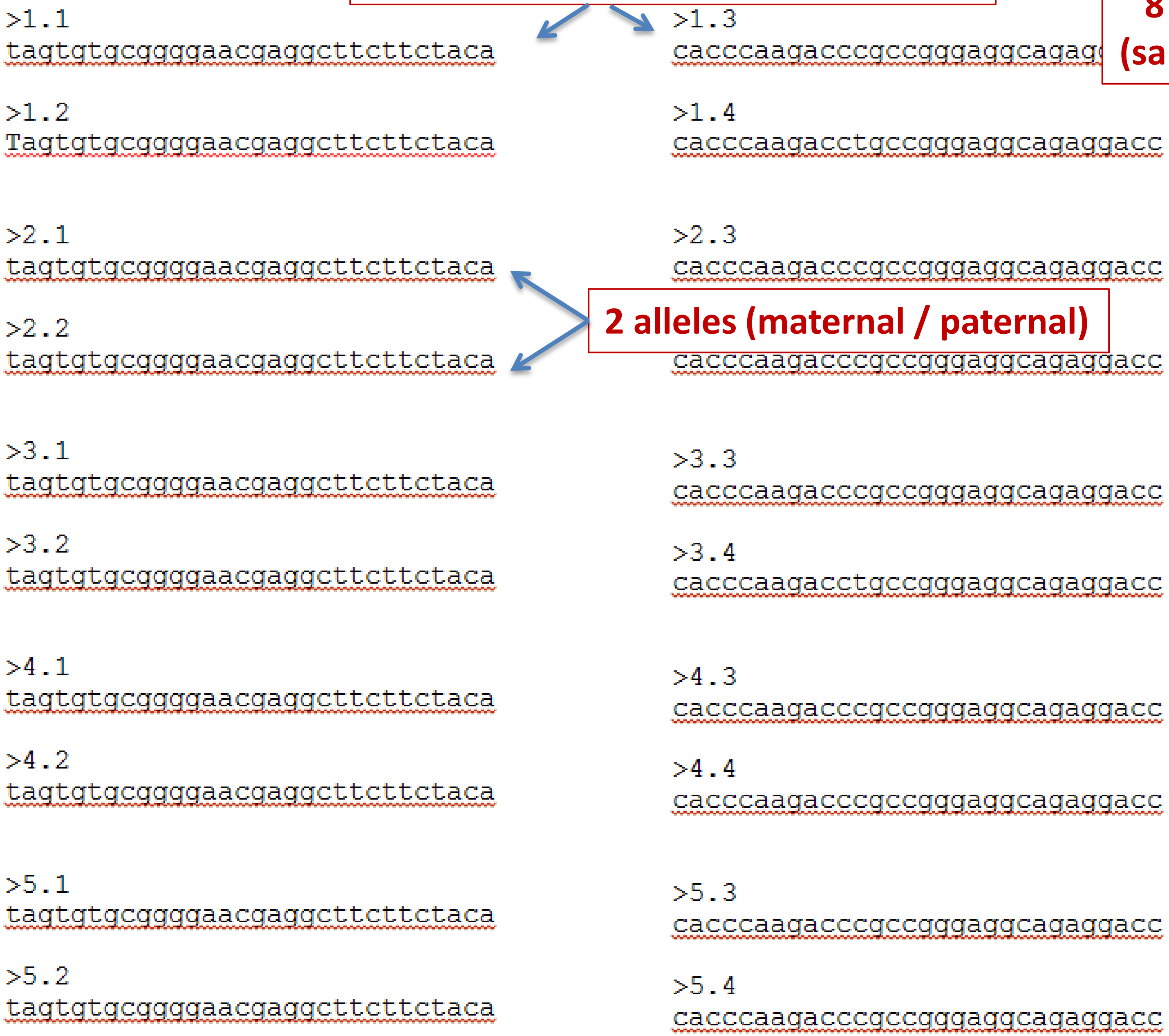
- 8 family members - 4 DNA sequences - one allele
- 8 family members - 1 DNA sequence - two alleles
- 8 family members - 2 DNA sequences - two alleles (not easy)

Global Organisation for Bioinformatics Learning, Education & Training

2 different DNA sequences (INS gene)

8 subjects (same family)

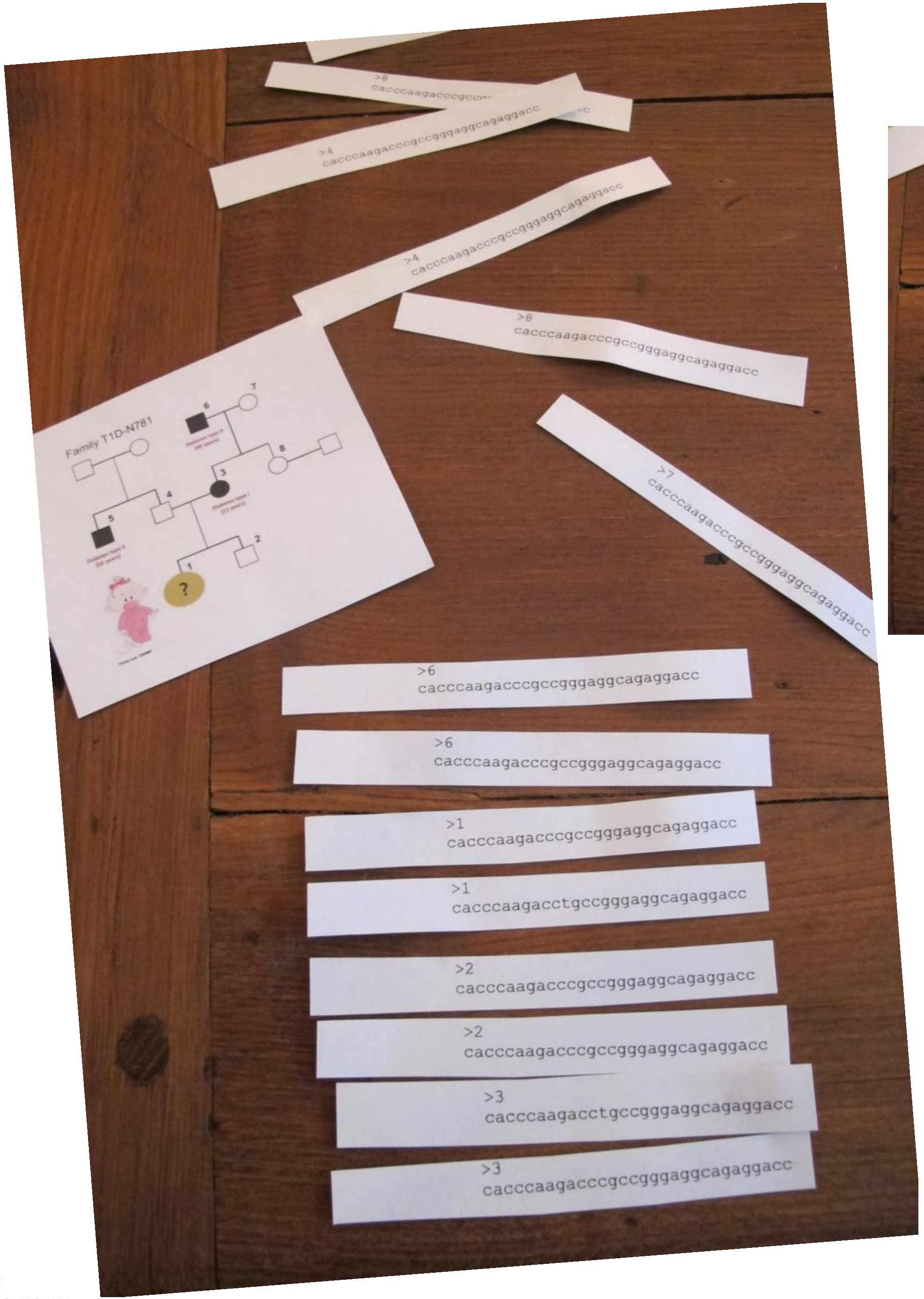
2 alleles (maternal / paternal)

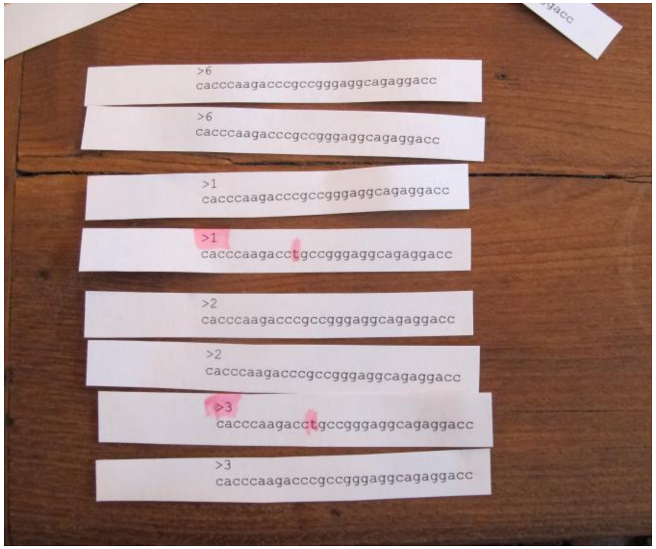


```
>1.1
tagtgtgcggggaacgaggcttcttctaca

>1.2
Tagtgtgcggggaacgaggcttcttctaca

>2.1
tagtgtgcggggaacgaggcttcttctaca

>2.2
tagtgtgcggggaacgaggcttcttctaca

>3.1
tagtgtgcggggaacgaggcttcttctaca

>3.2
tagtgtgcggggaacgaggcttcttctaca

>4.1
tagtgtgcggggaacgaggcttcttctaca

>4.2
tagtgtgcggggaacgaggcttcttctaca

>5.1
tagtgtgcggggaacgaggcttcttctaca

>5.2
tagtgtgcggggaacgaggcttcttctaca
```

```
>1.3
cacccaagacccgccgggaggcagag

>1.4
cacccaagacctgccgggaggcagaggacc

>2.3
cacccaagacccgccgggaggcagaggacc

cacccaagacccgccgggaggcagaggacc

>3.3
cacccaagacccgccgggaggcagaggacc

>3.4
cacccaagacctgccgggaggcagaggacc

>4.3
cacccaagacccgccgggaggcagaggacc

>4.4
cacccaagacccgccgggaggcagaggacc

>5.3
cacccaagacccgccgggaggcagaggacc

>5.4
cacccaagacccgccgggaggcagaggacc
```

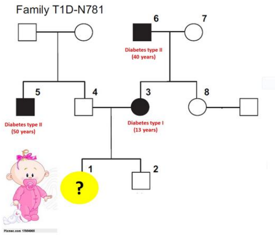
Global Organisation for Bioinformatics Learning, Education & Training


>8
cacccaagacccgcc
>4
cacccaagacccgccgggaggcagaggacc
>4
cacccaagacccgccgggaggcagaggacc
>8
cacccaagacccgccgggaggcagaggacc
>7
cacccaagacccgccgggaggcagaggacc
Family T1D-N781
6
7
8
3
4
5
1
2
?
>6
cacccaagacccgccgggaggcagaggacc
>6
cacccaagacccgccgggaggcagaggacc
>1
cacccaagacccgccgggaggcagaggacc
>1
cacccaagacctgccgggaggcagaggacc
>2
cacccaagacccgccgggaggcagaggacc
>2
cacccaagacccgccgggaggcagaggacc
>3
cacccaagacctgccgggaggcagaggacc
>3
cacccaagacccgccgggaggcagaggacc


>6
cacccaagacccgccgggaggcagaggacc
>6
cacccaagacccgccgggaggcagaggacc
>1
cacccaagacccgccgggaggcagaggacc
>1
cacccaagacctgccgggaggcagaggacc
>2
cacccaagacccgccgggaggcagaggacc
>2
cacccaagacccgccgggaggcagaggacc
>3
cacccaagacctgccgggaggcagaggacc
>3
cacccaagacccgccgggaggcagaggacc

Global Organisation for Bioinformatics Learning, Education & Training

Sequence 1



```
1.1        tagtgtgcggggaacgaggcttcttctaca
1.2        tagtgtgcggggaacgaggcttcttctaca
2.1        tagtgtgcggggaacgaggcttcttctaca
2.2        tagtgtgcggggaacgaggcttcttctaca
3.1        tagtgtgcggggaacgaggcttcttctaca
3.2        tagtgtgcggggaacgaggcttcttctaca
4.1        tagtgtgcggggaacgaggcttcttctaca
4.2        tagtgtgcggggaacgaggcttcttctaca
5.1        tagtgtgcggggaacgaggcttcttctaca
5.2        tagtgtgcggggaacgaggcttcttctaca
6.1        tagtgtgcggggaacgaggcttcttctaca
6.2        tagtgtgcggggaacgaggcttcttctaca
7.1        tagtgtgcggagaacgaggcttcttctaca
7.2        tagtgtgcggagaacgaggcttcttctaca
8.1        tagtgtgcggggaacgaggcttcttctaca
8.2        tagtgtgcggggaacgaggcttcttctaca
```

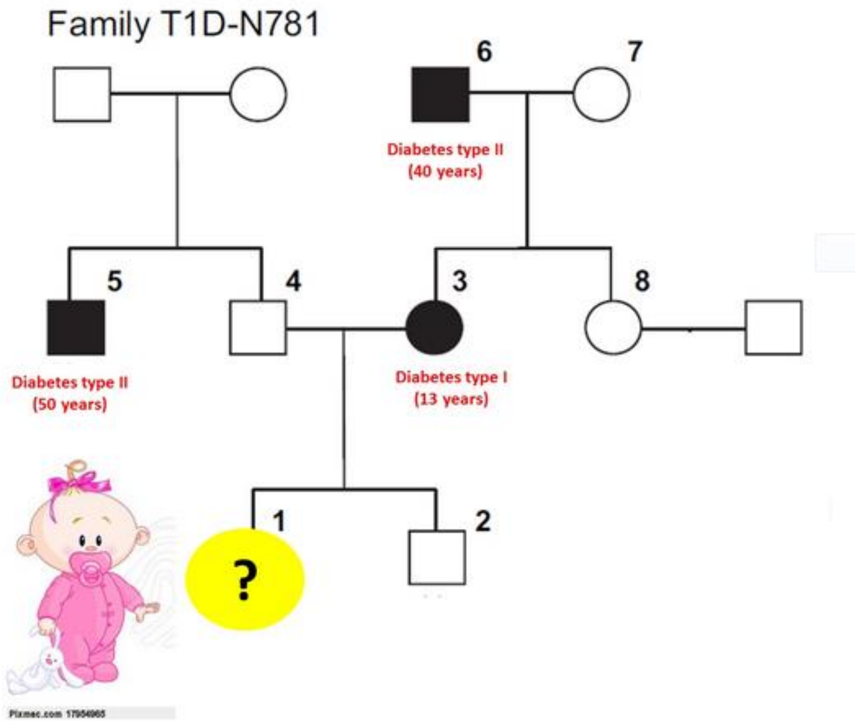
http://www2.grifil.com/album.html

Where are the differences ?

Global Organisation for Bioinformatics Learning, Education & Training

Sequence 1

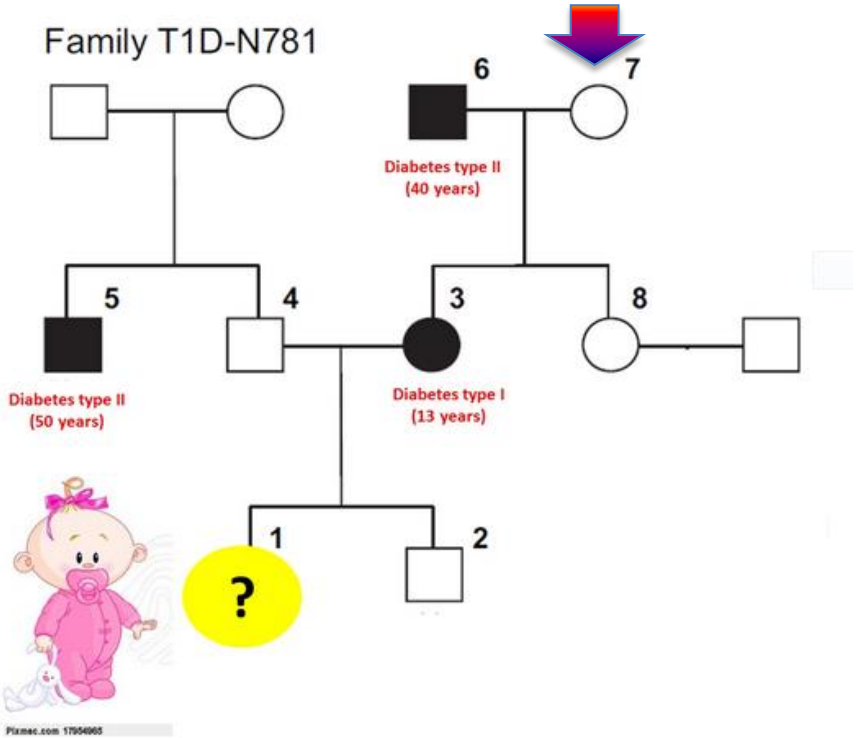
1.1	tagtgtgcggggaa
1.2	tagtgtgcggggaa
2.1	tagtgtgcggggaa
2.2	tagtgtgcggggaa
3.1	tagtgtgcggggaa
3.2	tagtgtgcggggaa
4.1	tagtgtgcggggaa
4.2	tagtgtgcggggaa
5.1	tagtgtgcggggaa
5.2	tagtgtgcggggaa
6.1	tagtgtgcggggaacgaggcttcttctaca
6.2	tagtgtgcggggaacgaggcttcttctaca
7.1	tagtgtgcggagaacgaggcttcttctaca
7.2	tagtgtgcggagaacgaggcttcttctaca
8.1	tagtgtgcggggaacgaggcttcttctaca
8.2	tagtgtgcggggaacgaggcttcttctaca



Global Organisation for Bioinformatics Learning, Education & Training

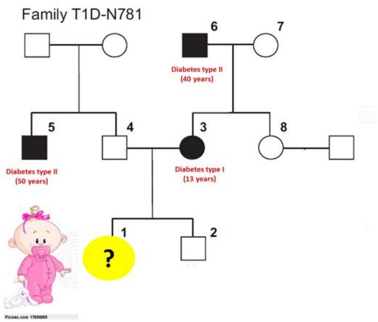
Sequence 1

```
1.1        tagtgtgcggggaa
1.2        tagtgtgcggggaa
2.1        tagtgtgcggggaa
2.2        tagtgtgcggggaa
3.1        tagtgtgcggggaa
3.2        tagtgtgcggggaa
4.1        tagtgtgcggggaa
4.2        tagtgtgcggggaa
5.1        tagtgtgcggggaa
5.2        tagtgtgcggggaa
6.1        tagtgtgcggggaacgaggcttcttctaca
6.2        tagtgtgcggggaacgaggcttcttctaca
7.1        tagtgtgcggagaacgaggcttcttctaca
7.2        tagtgtgcggagaacgaggcttcttctaca
8.1        tagtgtgcggggaacgaggcttcttctaca
8.2        tagtgtgcggggaacgaggcttcttctaca
```



The SNP g -> a (homozygous; subject 7) is not associated with diabetes (neutral)

Sequence 2

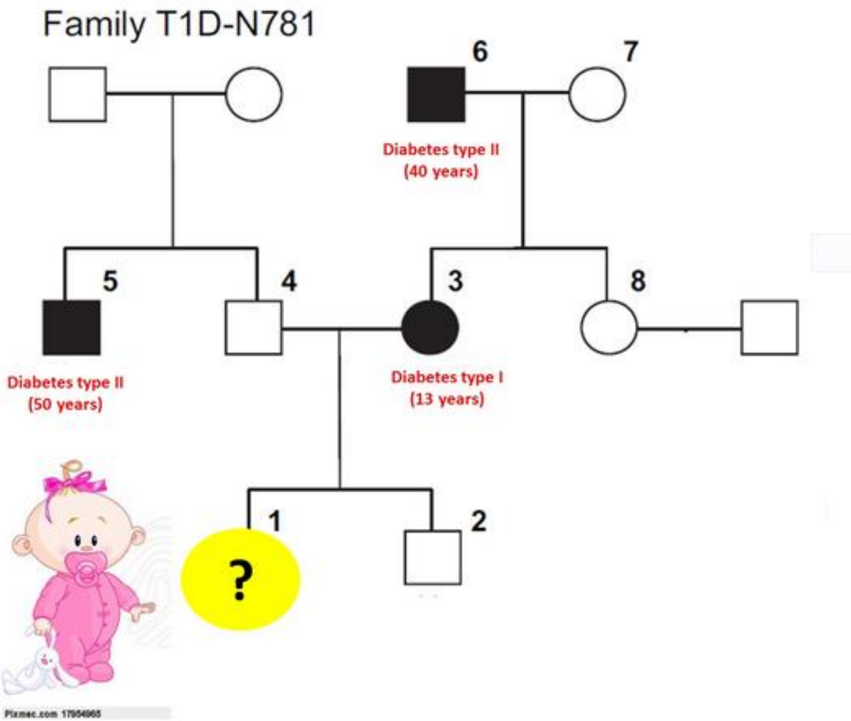

Family T1D-N781
6
7
Diabetes type II (40 years)
5
4
3
8
Diabetes type II (50 years)
Diabetes type I (13 years)
1
?
2

11	cacccaagacccgccgggaggcagaggacc
12	cacccaagacctgccgggaggcagaggacc
21	cacccaagacccgccgggaggcagaggacc
22	cacccaagacccgccgggaggcagaggacc
31	cacccaagacccgccgggaggcagaggacc
32	cacccaagacctgccgggaggcagaggacc
41	cacccaagacccgccgggaggcagaggacc
42	cacccaagacccgccgggaggcagaggacc
51	cacccaagacccgccgggaggcagaggacc
52	cacccaagacccgccgggaggcagaggacc
61	cacccaagacccgccgggaggcagaggacc
62	cacccaagacccgccgggaggcagaggacc
71	cacccaagacccgccgggaggcagaggacc
72	cacccaagacccgccgggaggcagaggacc
81	cacccaagacccgccgggaggcagaggacc
82	cacccaagacccgccgggaggcagaggacc

Where are the differences ?

Global Organisation for Bioinformatics Learning, Education & Training

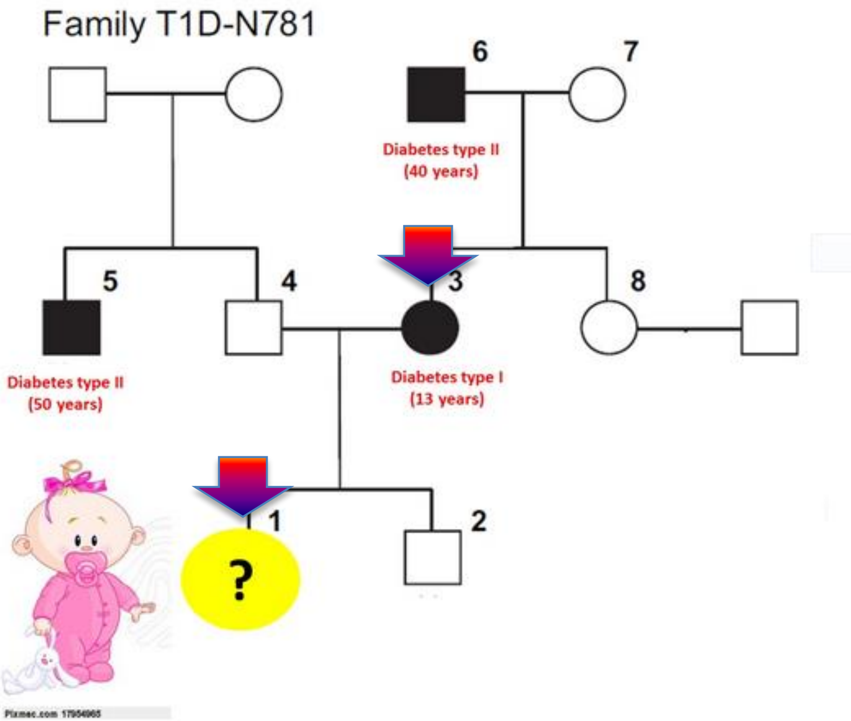
Sequence 2



```
11    cacccaagacccgccgg
12    cacccaagacctgccgg
21    cacccaagacccgccgg
22    cacccaagacccgccgg
31    cacccaagacccgccgg
32    cacccaagacctgccgg
41    cacccaagacccgccgg
42    cacccaagacccgccgg
51    cacccaagacccgccgg
52    cacccaagacccgccgggaggcagaggacc
61    cacccaagacccgccgggaggcagaggacc
62    cacccaagacccgccgggaggcagaggacc
71    cacccaagacccgccgggaggcagaggacc
72    cacccaagacccgccgggaggcagaggacc
81    cacccaagacccgccgggaggcagaggacc
82    cacccaagacccgccgggaggcagaggacc
```

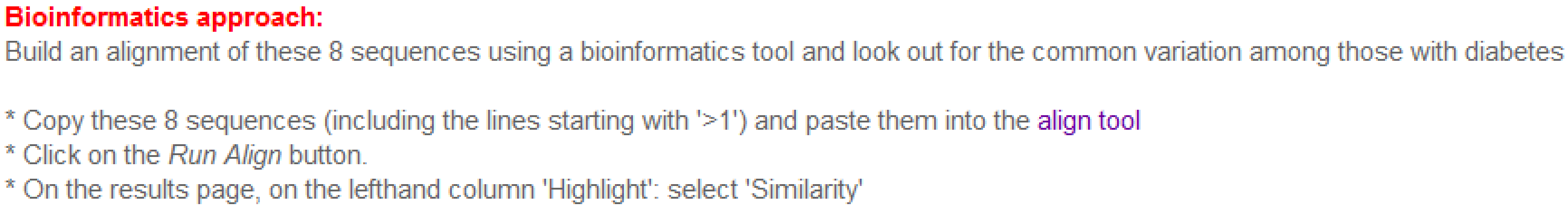
Sequence 2

```
11    cacccaagacccgccggg
12    cacccaagacctgccggg
21    cacccaagacccgccggg
22    cacccaagacccgccggg
31    cacccaagacccgccggg
32    cacccaagacctgccggg
41    cacccaagacccgccggg
42    cacccaagacccgccggg
51    cacccaagacccgccggg
52    cacccaagacccgccgggaggcagaggacc
61    cacccaagacccgccgggaggcagaggacc
62    cacccaagacccgccgggaggcagaggacc
71    cacccaagacccgccgggaggcagaggacc
72    cacccaagacccgccgggaggcagaggacc
81    cacccaagacccgccgggaggcagaggacc
82    cacccaagacccgccgggaggcagaggacc
```



ANSWER: The SNP c -> t is present in subjects 3 and 1 (heterozygous) and is associated with Type I Diabetes (‘all type I diabetic members carry the same variation in the INS gene’)

Global Organisation for Bioinformatics Learning, Education & Training



Bioinformatics approach:

Build an alignment of these 8 sequences using a bioinformatics tool and look out for the common variation among those with diabetes

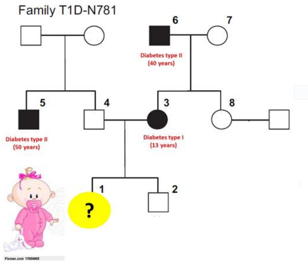
* Copy these 8 sequences (including the lines starting with '>1') and paste them into the align tool
* Click on the *Run Align* button.
* On the results page, on the lefthand column 'Highlight': select 'Similarity'

Bioinformatics – Alignment tool (UniProt)
http://www.uniprot.org/align/

This tool is used to align protein sequences, but it can also properly align short DNA sequences

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics



Question: is the baby diabetic ?

To answer this question, researchers extracted DNA from 8 members of the Norwegian family and sequenced part of the gene that encodes for insulin.

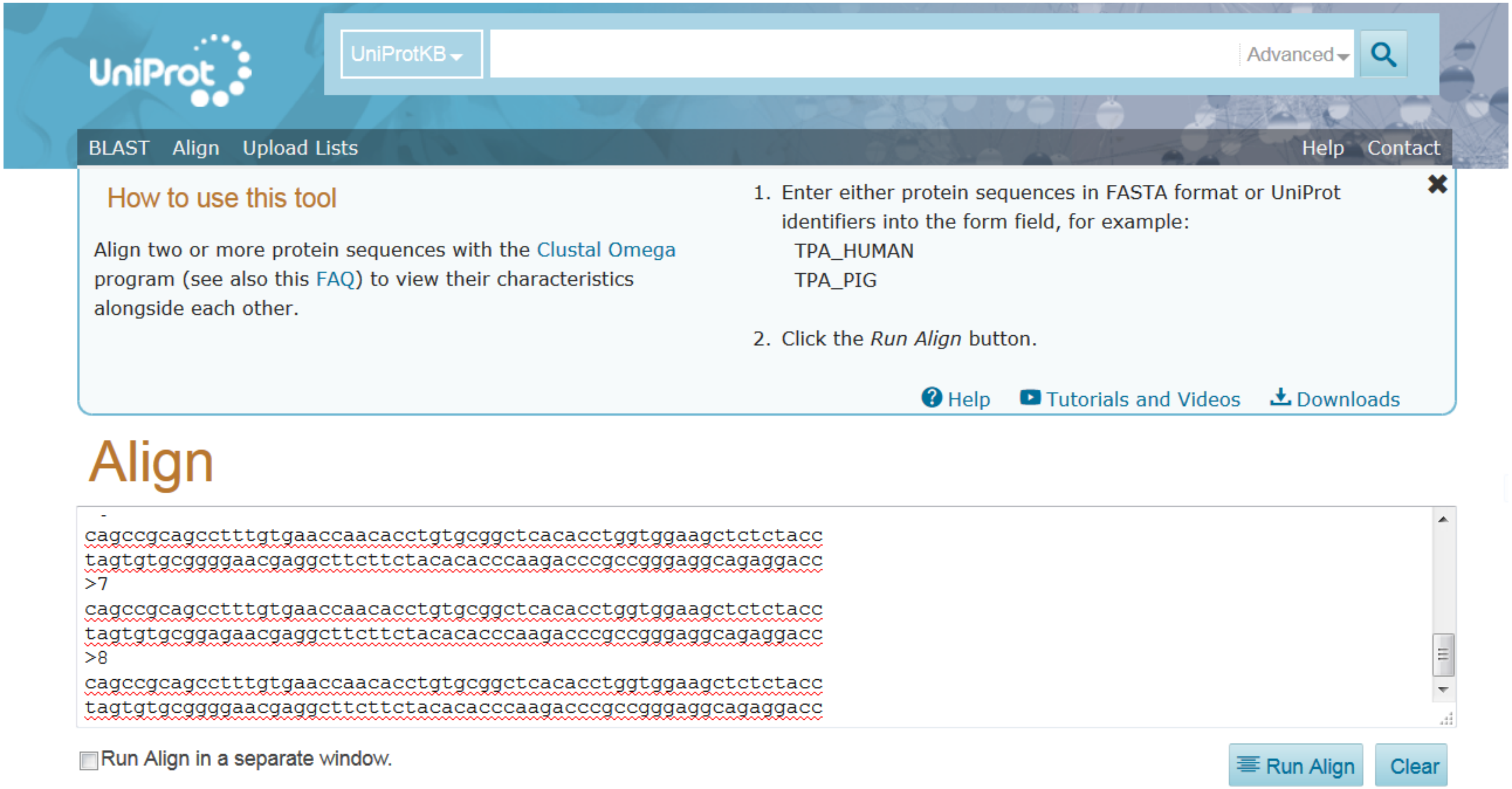
```
>1
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc
>2
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>3
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc
>4
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>5
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>6
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>7
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggagaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>8
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
```

Where are the differences ?

http://www2.grifil.com/album.html

Global Organisation for Bioinformatics Learning, Education & Training

http://www.uniprot.org/align/



How to use this tool

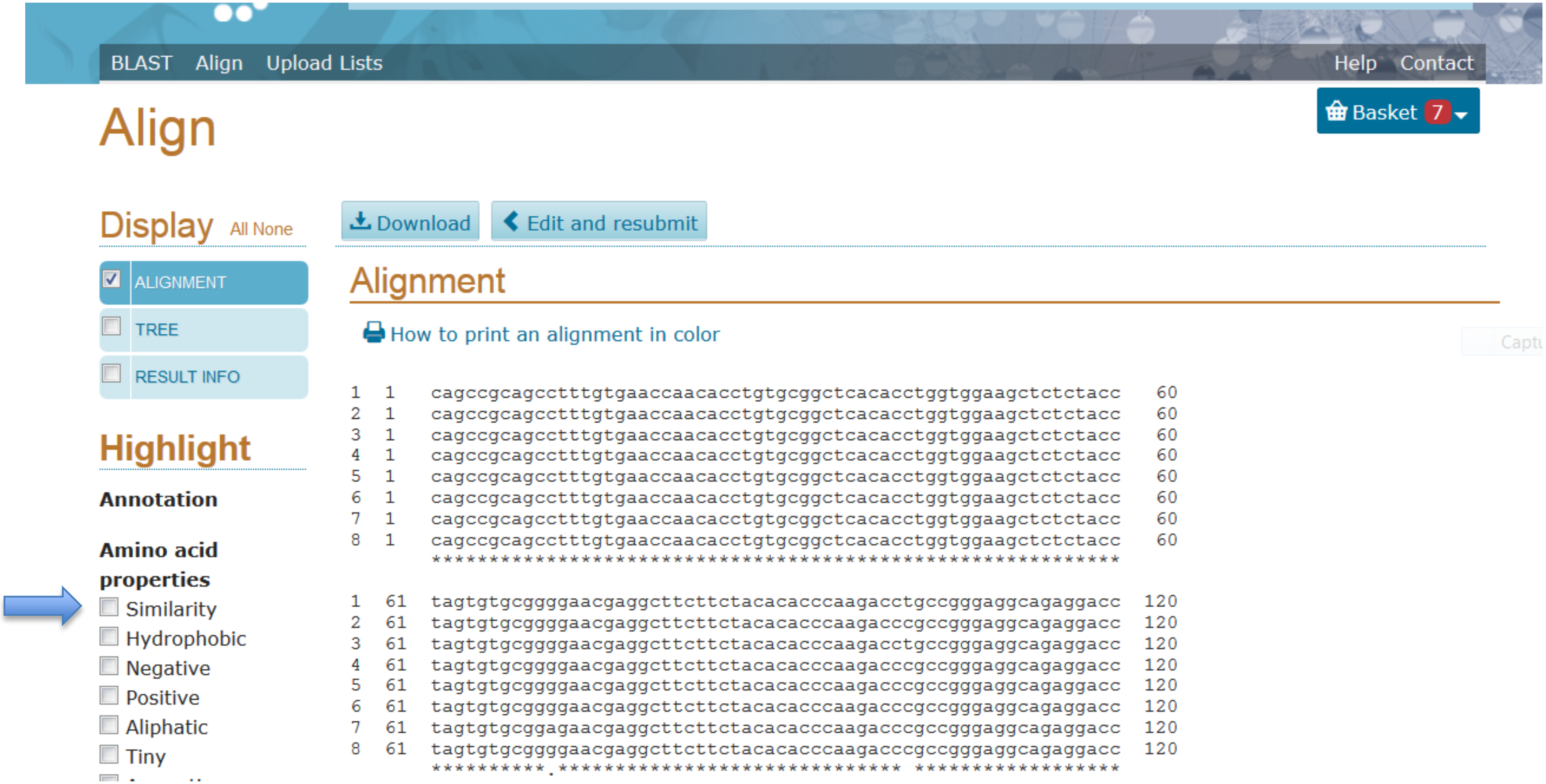
Align two or more protein sequences with the Clustal Omega program (see also this FAQ) to view their characteristics alongside each other.

1. Enter either protein sequences in FASTA format or UniProt identifiers into the form field, for example:
   TPA_HUMAN
   TPA_PIG
2. Click the *Run Align* button.

Align

```
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>7
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggagaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
>8
cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc
tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc
```

☐ Run Align in a separate window.



Align

Display

All None

- ☑ ALIGNMENT
- □ TREE
- □ RESULT INFO

Highlight

Annotation

Amino acid properties

- □ Similarity
- □ Hydrophobic
- □ Negative
- □ Positive
- □ Aliphatic
- □ Tiny

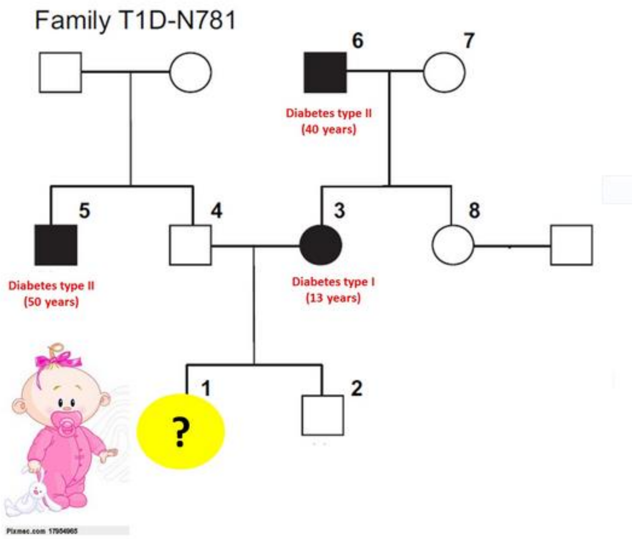
Alignment

How to print an alignment in color

```
1  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
2  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
3  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
4  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
5  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
6  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
7  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
8  1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
        ************************************************************

1  61   tagtgtgcggggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc  120
2  61   tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
3  61   tagtgtgcggggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc  120
4  61   tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
5  61   tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
6  61   tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
7  61   tagtgtgcggagaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
8  61   tagtgtgcggggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc  120
        **********.***************************** *******************
```

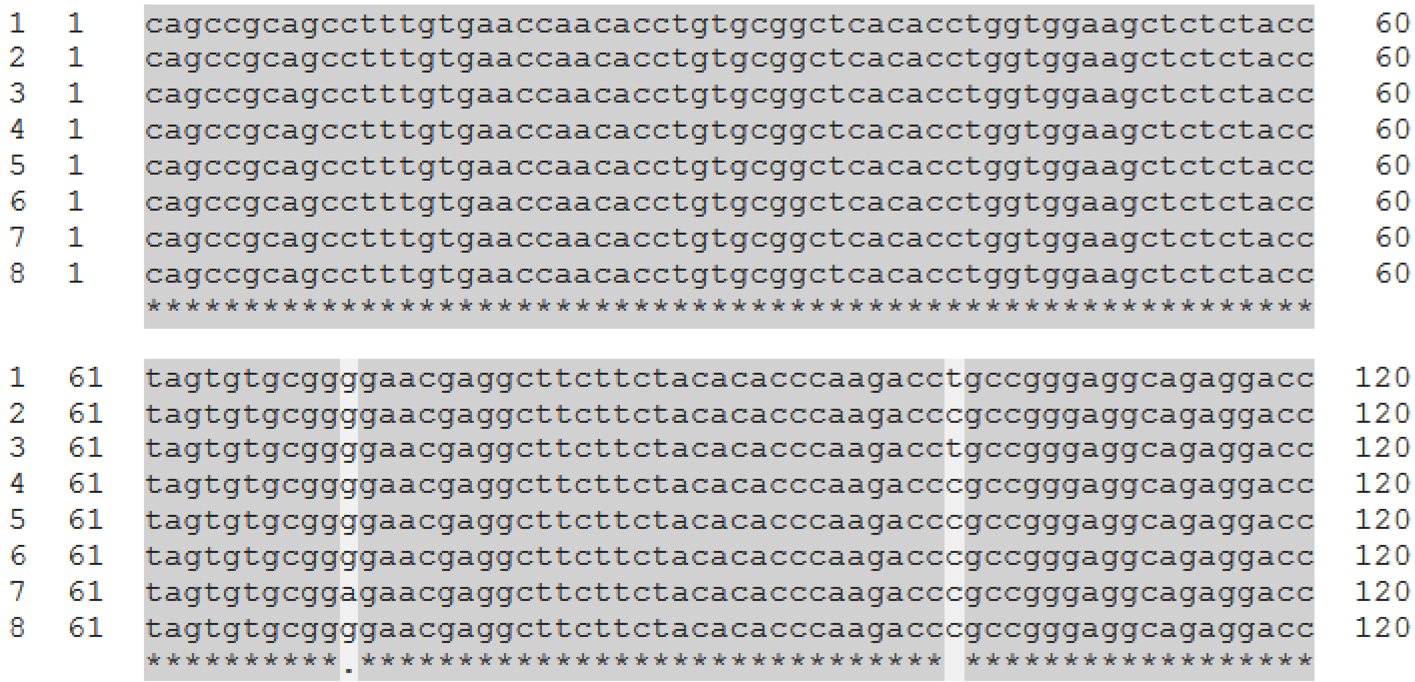
Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics

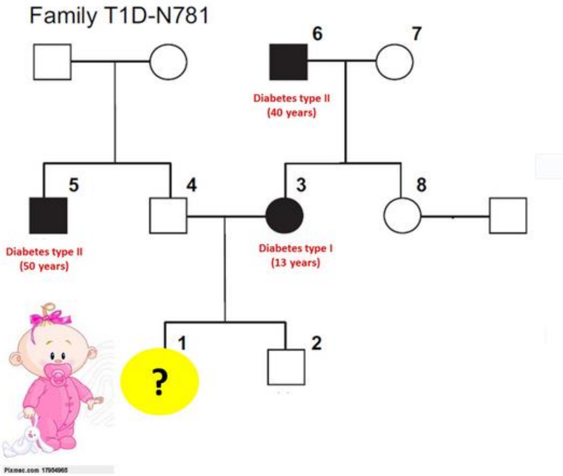


Bioinformatics – Alignment tool (UniProt) http://www.uniprot.org/align/

(only one DNA sequence (allele) per subject)



```
1   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
2   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
3   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
4   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
5   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
6   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
7   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
8   1    cagccgcagcctttgtgaaccaacacctgtgcggctcacacctggtggaagctctctacc   60
         ************************************************************

1   61   tagtgtgcgggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc   120
2   61   tagtgtgcgggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
3   61   tagtgtgcgggaacgaggcttcttctacacacccaagacctgccgggaggcagaggacc   120
4   61   tagtgtgcgggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
5   61   tagtgtgcgggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
6   61   tagtgtgcgggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
7   61   tagtgtgcggagaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
8   61   tagtgtgcgggaacgaggcttcttctacacacccaagacccgccgggaggcagaggacc   120
         **********.****************************** ******************
```



- The subject **(1)** with the **c -> t** mutation (heterozygous mutation) is a girl who presented type I diabetes at the early age of 10.
- The girl's mother **(3)** has type I diabetes that was diagnosed when she was 13. Currently, she is being treated with insulin. She also carries the heterozygous mutation **c -> t** .
- The girl's maternal grandfather **(6)** has type 2 diabetes, which was diagnosed at the age of 40. He is currently being treated with insulin. Neither he nor the healthy maternal grandmother carry mutations.
- **-> Thus, the girl's mother is carrying a *de novo* c -> t mutation, which must be a germline mutation since it has been inherited by her daughter.**

(Molven et al., 2008)

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics

Biology / Statistics / Bioinformatics

Molven et al., 2008

“We screened INS (*gene coding for insulin*) in 92 probands, and 223 subjects from the Norwegian Childhood **Diabetes** Registry selected on the basis of autoantibody negativity or family history of diabetes.”

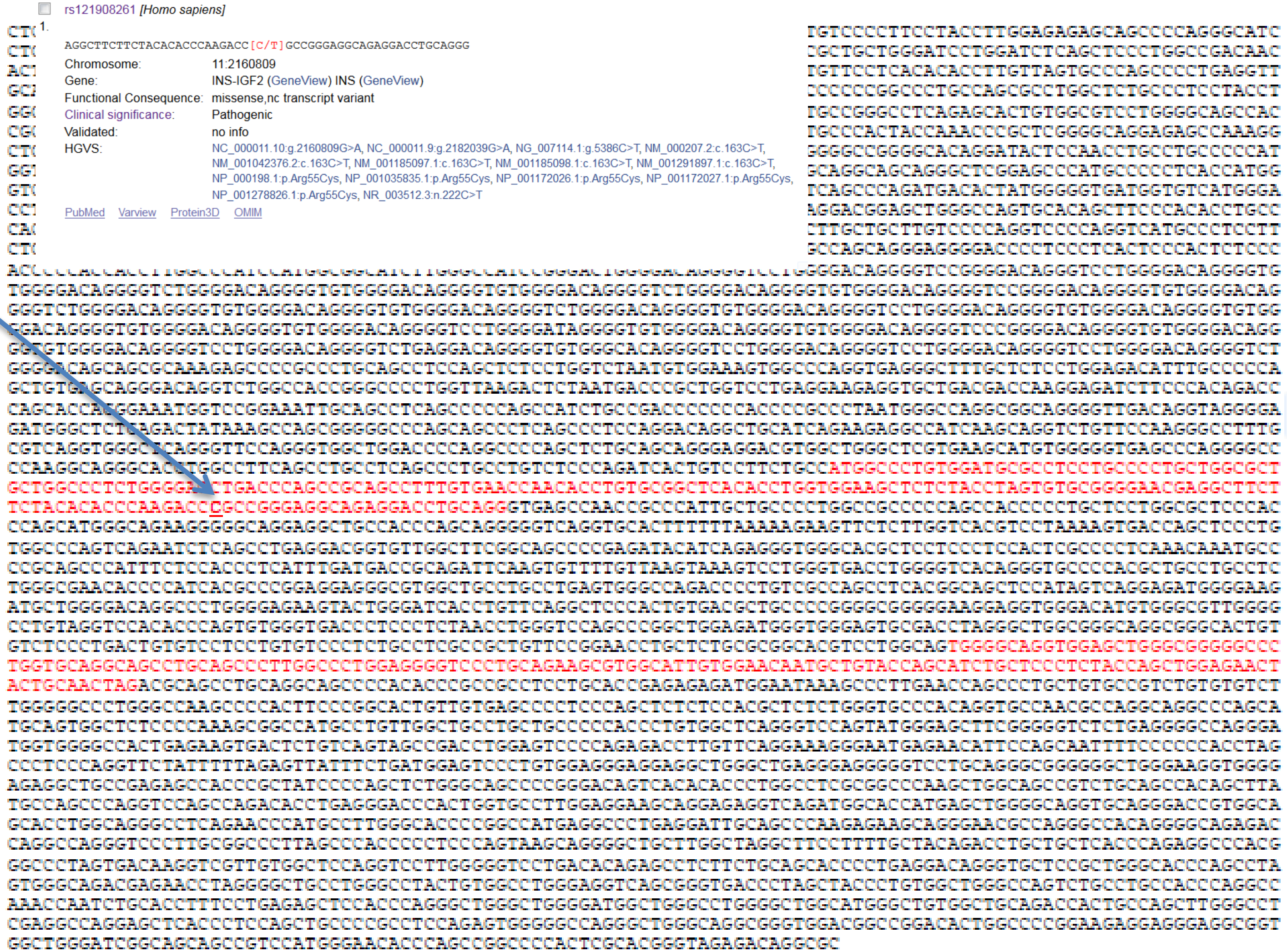
Identification of a rare genetic variation - rs121908261 – in the human INS gene which is the cause of type I diabetes in a Norwegian family

Insulin gene
(4992 pb ; chromosome 11)

In red, part of the DNA sequence translated into the protein sequence

rs121908261

c -> t



□ rs121908261 *[Homo sapiens]*

1.

AGGCTTCTTCTACACACCCAAGACC[C/T]GCCGGGAGGCAGAGGACCTGCAGGG

Chromosome:	11:2160809
Gene:	INS-IGF2 (GeneView) INS (GeneView)
Functional Consequence:	missense,nc transcript variant
Clinical significance:	Pathogenic
Validated:	no info
HGVS:	NC_000011.10:g.2160809G>A, NC_000011.9:g.2182039G>A, NG_007114.1:g.5386C>T, NM_000207.2:c.163C>T, NM_001042376.2:c.163C>T, NM_001185097.1:c.163C>T, NM_001185098.1:c.163C>T, NM_001291897.1:c.163C>T, NP_000198.1:p.Arg55Cys, NP_001035835.1:p.Arg55Cys, NP_001172026.1:p.Arg55Cys, NP_001172027.1:p.Arg55Cys, NP_001278826.1:p.Arg55Cys, NR_003512.3:n.222C>T

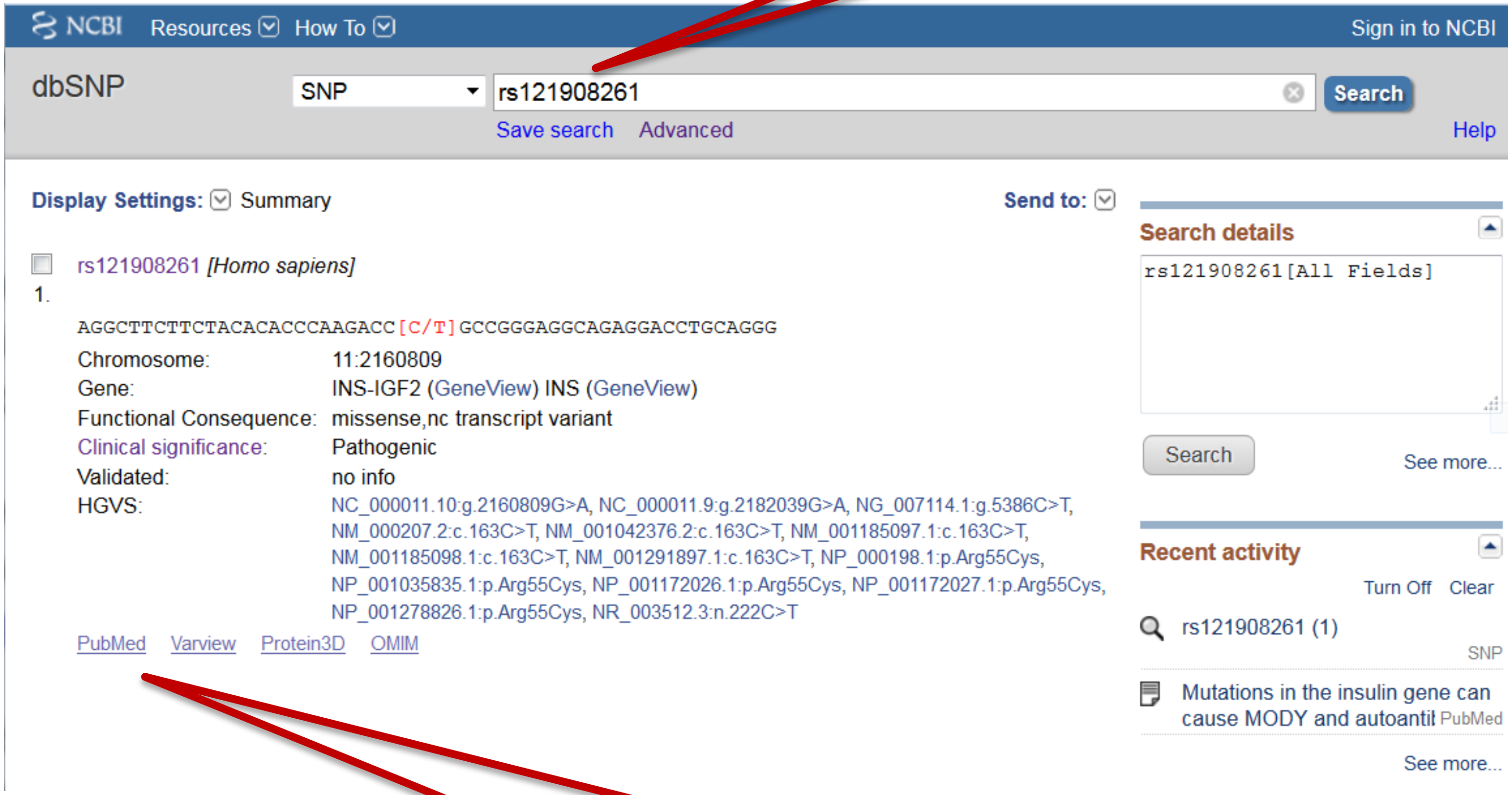
PubMed Varview Protein3D OMIM

```
CT   TGTCCCCTTCCTACCTTGGAGAGAGCAGCCCCAGGGCATC
CT   CGCTGCTGGGATCCTGGATCTCAGCTCCCTGGCCGACAAC
AC   TGTTCCTCACACACCTTGTTAGTGCCCAGCCCCTGAGGTT
GC   CCCCCGGCCCTGCCAGCGCCTGGCTCTGCCCTCCTACCT
GG   TGCCGGGCCTCAGAGCACTGTGGCGTCCTGGGGCAGCCAC
CG   TGCCCACTACCAAACCCGCTCGGGGCAGGAGAGCCAAAGG
CT   GGGGCCGGGGCACAGGATACTCCAACCTGCCTGCCCCCAT
GG   GCAGGCAGCAGGGCTCGGAGCCCATGCCCCCTCACCATGG
GT   TCAGCCCAGATGACACTATGGGGGTGATGGTGTCATGGGA
CC   AGGACGGAGCTGGGCCAGTGCACAGCTTCCCACACCTGCC
CA   CTTGCTGCTTGTCCCCAGGTCCCCAGGTCATGCCCTCCTT
CT   GCCAGCAGGGAGGGGACCCCTCCCTCACTCCCACTCTCCC
ACCCCCACCACCTTGGCCCATCCATGGCGGCATCTTGGGCCATCCGGGACTGGGGACAGGGGTCCTGGGGACAGGGGTCCGGGGACAGGGTCCTGGGGACAGGGGTG
TGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCCGGGGACAGGGGTGTGGGGACAG
GGGTCTGGGGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTCTGGGGACAGGGGTGTGGGGACAGGGGTCCTGGGGACAGGGGTGTGGGGACAGGGGTGTGG
GGACAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTCCTGGGGATAGGGGTGTGGGGACAGGGGTGTGGGGACAGGGGTCCCGGGGACAGGGGTGTGGGGACAGG
GGTGTGGGGACAGGGGTCCTGGGGACAGGGGTCTGAGGACAGGGGTGTGGGCACAGGGGTCCTGGGGACAGGGGTCCTGGGGACAGGGGTCCTGGGGACAGGGGTCT
GGGGACAGCAGCGCAAAGAGCCCCGCCCTGCAGCCTCCAGCTCTCCTGGTCTAATGTGGAAAGTGGCCCAGGTGAGGGCTTTGCTCTCCTGGAGACATTTGCCCCCA
GCTGTGAGCAGGGACAGGTCTGGCCACCGGGCCCCTGGTTAAGACTCTAATGACCCGCTGGTCCTGAGGAAGAGGTGCTGACGACCAAGGAGATCTTCCCACAGACC
CAGCACCAGGGAAATGGTCCGGAAATTGCAGCCTCAGCCCCCAGCCATCTGCCGACCCCCCCACCCCGCCCTAATGGGCCAGGCGGCAGGGGTTGACAGGTAGGGGA
GATGGGCTCTGAGACTATAAAGCCAGCGGGGGCCCAGCAGCCCTCAGCCCTCCAGGACAGGCTGCATCAGAAGAGGCCATCAAGCAGGTCTGTTCCAAGGGCCTTTG
CGTCAGGTGGGCTCAGGGTTCCAGGGTGGCTGGACCCCAGGCCCCAGCTCTGCAGCAGGGAGGACGTGGCTGGGCTCGTGAAGCATGTGGGGGTGAGCCCAGGGGCC
CCAAGGCAGGGCACCTGGCCTTCAGCCTGCCTCAGCCCTGCCTGTCTCCCAGATCACTGTCCTTCTGCCATGGCCCTGTGGATGCGCCTCCTGCCCCTGCTGGCGCT
GCTGGCCCTCTGGGGACCTGACCCAGCCGCAGCCTTTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAGCTCTCTACCTAGTGTGCGGGGAACGAGGCTTCT
TCTACACACCCAAGACCCGCCGGGAGGCAGAGGACCTGCAGGGTGAGCCAACCGCCCATTGCTGCCCCTGGCCGCCCCCAGCCACCCCCTGCTCCTGGCGCTCCCAC
CCAGCATGGGCAGAAGGGGGCAGGAGGCTGCCACCCAGCAGGGGGTCAGGTGCACTTTTTTAAAAAGAAGTTCTCTTGGTCACGTCCTAAAAGTGACCAGCTCCCTG
TGGCCCAGTCAGAATCTCAGCCTGAGGACGGTGTTGGCTTCGGCAGCCCCGAGATACATCAGAGGGTGGGCACGCTCCTCCCTCCACTCGCCCCTCAAACAAATGCC
CCGCAGCCCATTTCTCCACCCTCATTTGATGACCGCAGATTCAAGTGTTTTGTTAAGTAAAGTCCTGGGTGACCTGGGGTCACAGGGTGCCCCACGCTGCCTGCCTC
TGGGCGAACACCCCATCACGCCCGGAGGAGGGCGTGGCTGCCTGCCTGAGTGGGCCAGACCCCTGTCGCCAGCCTCACGGCAGCTCCATAGTCAGGAGATGGGGAAG
ATGCTGGGGACAGGCCCTGGGGAGAAGTACTGGGATCACCTGTTCAGGCTCCCACTGTGACGCTGCCCCGGGGCGGGGGAAGGAGGTGGGACATGTGGGCGTTGGGG
CCTGTAGGTCCACACCCAGTGTGGGTGACCCTCCCTCTAACCTGGGTCCAGCCCGGCTGGAGATGGGTGGGAGTGCGACCTAGGGCTGGCGGGCAGGCGGGCACTGT
GTCTCCCTGACTGTGTCCTCCTGTGTCCCTCTGCCTCGCCGCTGTTCCGGAACCTGCTCTGCGCGGCACGTCCTGGCAGTGGGGCAGGTGGAGCTGGGCGGGGGCCC
TGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAACAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACT
ACTGCAACTAGACGCAGCCTGCAGGCAGCCCCACACCCGCCGCCTCCTGCACCGAGAGAGATGGAATAAAGCCCTTGAACCAGCCCTGCTGTGCCGTCTGTGTGTCT
TGGGGGCCCTGGGCCAAGCCCCACTTCCCGGCACTGTTGTGAGCCCCTCCCAGCTCTCTCCACGCTCTCTGGGTGCCCACAGGTGCCAACGCCAGGCAGGCCCAGCA
TGCAGTGGCTCTCCCCAAAGCGGCCATGCCTGTTGGCTGCCTGCTGCCCCCACCCTGTGGCTCAGGGTCCAGTATGGGAGCTTCGGGGGTCTCTGAGGGGCCAGGGA
TGGTGGGGCCACTGAGAAGTGACTCTGTCAGTAGCCGACCTGGAGTCCCCAGAGACCTTGTTCAGGAAAGGGAATGAGAACATTCCAGCAATTTTCCCCCCACCTAG
CCCTCCCAGGTTCTATTTTTAGAGTTATTTCTGATGGAGTCCCTGTGGAGGGAGGAGGCTGGGCTGAGGGAGGGGGTCCTGCAGGGCGGGGGGCTGGGAAGGTGGGG
AGAGGCTGCCGAGAGCCACCCGCTATCCCCAGCTCTGGGCAGCCCCGGGACAGTCACACACCCTGGCCTCGCGGCCCAAGCTGGCAGCCGTCTGCAGCCACAGCTTA
TGCCAGCCCAGGTCCAGCCAGACACCTGAGGGACCCACTGGTGCCTTGGAGGAAGCAGGAGAGGTCAGATGGCACCCATGAGCTGGGGCAGGTGCAGGGACCGTGGCA
GCACCTGGCAGGGCCTCAGAACCCATGCCTTGGGCACCCCGGCCATGAGGCCCTGAGGATTGCAGCCCAAGAGAAGCAGGGAACGCCAGGGCCACAGGGGCAGAGAC
CAGGCCAGGGTCCCTTGCGGCCCTTAGCCCACCCCCTCCCAGTAAGCAGGGGCTGCTTGGCTAGGCTTCCTTTTGCTACAGACCTGCTGCTCACCCAGAGGCCCACG
GGCCCTAGTGACAAGGTCGTTGTGGCTCCAGGTCCTTGGGGGTCCTGACACAGAGCCTCTTCTGCAGCACCCCTGAGGACAGGGTGCTCCGCTGGGCACCCAGCCTA
GTGGGCAGACGAGAACCTAGGGGCTGCCTGGGCCTACTGTGGCCTGGGAGGTCAGCGGGTGACCCTAGCTACCCTGTGGCTGGGCCAGTCTGCCTGCCACCCAGGCC
AAACCAATCTGCACCTTTCCTGAGAGCTCCACCCAGGGCTGGGCTGGGGATGGCTGGGCCTGGGGCTGGCATGGGCTGTGGCTGCAGACCACTGCCAGCTTGGGCCT
CGAGGCCAGGAGCTCACCCTCCAGCTGCCCCGCCTCCAGAGTGGGGGCCAGGGCTGGGCAGGCGGGTGGACGGCCGGACACTGGCCCCGGAAGAGGAGGGAGGCGGT
GGCTGGGATCGGCAGCAGCCGTCCATGGGAACACCCAGCCGGCCCCACTCGCACGGGTAGAGACAGGCGC
```

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics – Biological Database: dbSNP @ NCBI
http://www.ncbi.nlm.nih.gov/SNP/

Variant accession number in dbSNP



rs121908261

Display Settings: Summary

Send to:

1. ☐ rs121908261 *[Homo sapiens]*

AGGCTTCTTCTACACACCCAAGACC[C/T]GCCGGGAGGCAGAGGACCTGCAGGG

Chromosome: 11:2160809
Gene: INS-IGF2 (GeneView) INS (GeneView)
Functional Consequence: missense,nc transcript variant
Clinical significance: Pathogenic
Validated: no info
HGVS: NC_000011.10:g.2160809G>A, NC_000011.9:g.2182039G>A, NG_007114.1:g.5386C>T, NM_000207.2:c.163C>T, NM_001042376.2:c.163C>T, NM_001185097.1:c.163C>T, NM_001185098.1:c.163C>T, NM_001291897.1:c.163C>T, NP_000198.1:p.Arg55Cys, NP_001035835.1:p.Arg55Cys, NP_001172026.1:p.Arg55Cys, NP_001172027.1:p.Arg55Cys, NP_001278826.1:p.Arg55Cys, NR_003512.3:n.222C>T

PubMed Varview Protein3D OMIM

Search details

rs121908261[All Fields]

Search
See more...

Recent activity

Turn Off Clear

rs121908261 (1) SNP

Mutations in the insulin gene can cause MODY and autoantil PubMed

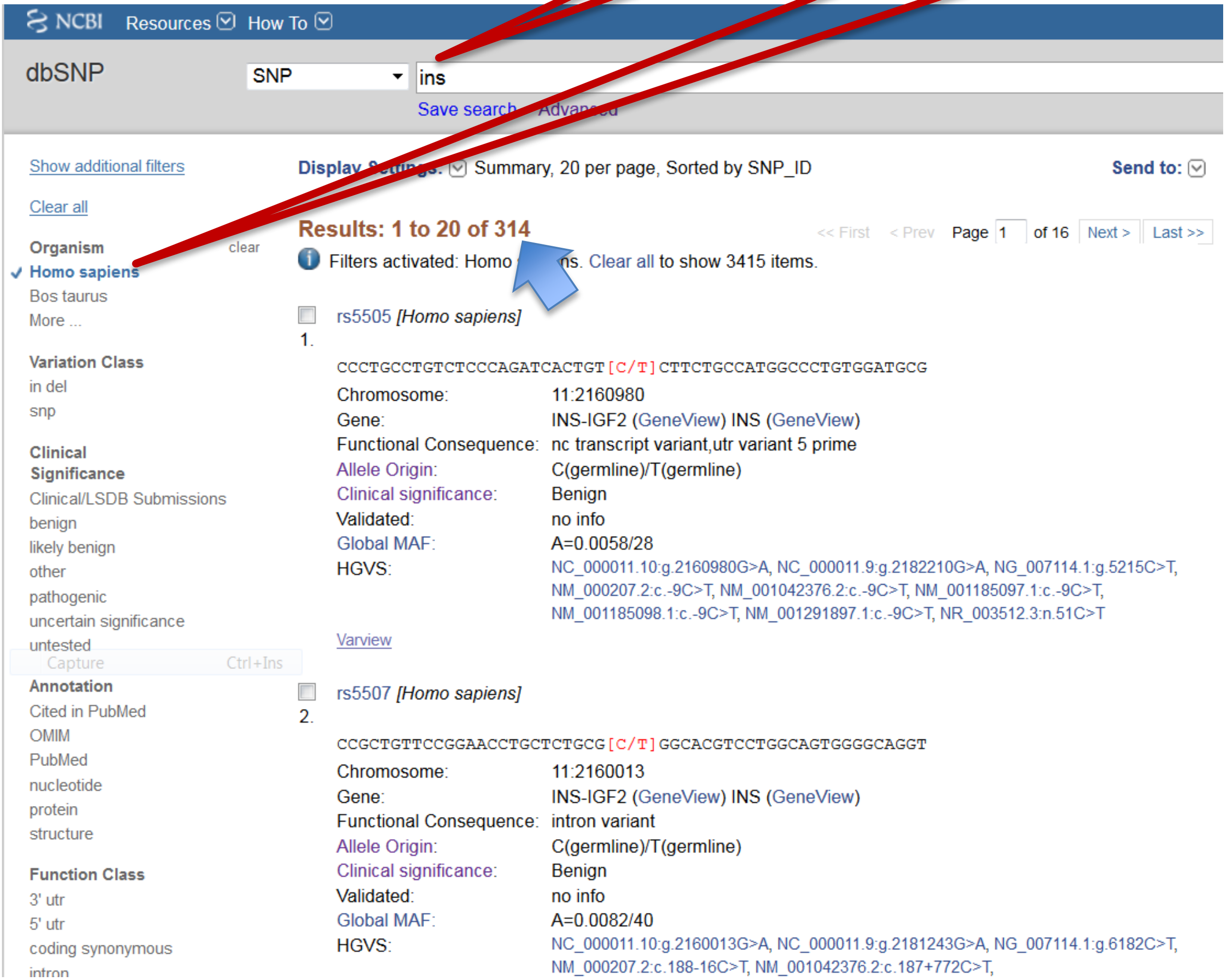
See more...

Link to the publication of Molven et al (2008)

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics – Biological Database: dbSNP @ NCBI
http://www.ncbi.nlm.nih.gov/SNP/

Look for all the SNPs located within human INS gene



ins

Organism
✓ Homo sapiens
Bos taurus
More ...

Variation Class
in del
snp

Clinical Significance
Clinical/LSDB Submissions
benign
likely benign
other
pathogenic
uncertain significance
untested

Annotation
Cited in PubMed
OMIM
PubMed
nucleotide
protein
structure

Function Class
3' utr
5' utr
coding synonymous
intron

Display Settings: Summary, 20 per page, Sorted by SNP_ID

Results: 1 to 20 of 314

Filters activated: Homo ns. Clear all to show 3415 items.

□ 1. rs5505 *[Homo sapiens]*

```
CCCTGCCTGTCTCCCAGATCACTGT[C/T]CTTCTGCCATGGCCCTGTGGATGCG
```

Chromosome:	11:2160980
Gene:	INS-IGF2 (GeneView) INS (GeneView)
Functional Consequence:	nc transcript variant,utr variant 5 prime
Allele Origin:	C(germline)/T(germline)
Clinical significance:	Benign
Validated:	no info
Global MAF:	A=0.0058/28
HGVS:	NC_000011.10:g.2160980G>A, NC_000011.9:g.2182210G>A, NG_007114.1:g.5215C>T, NM_000207.2:c.-9C>T, NM_001042376.2:c.-9C>T, NM_001185097.1:c.-9C>T, NM_001185098.1:c.-9C>T, NM_001291897.1:c.-9C>T, NR_003512.3:n.51C>T

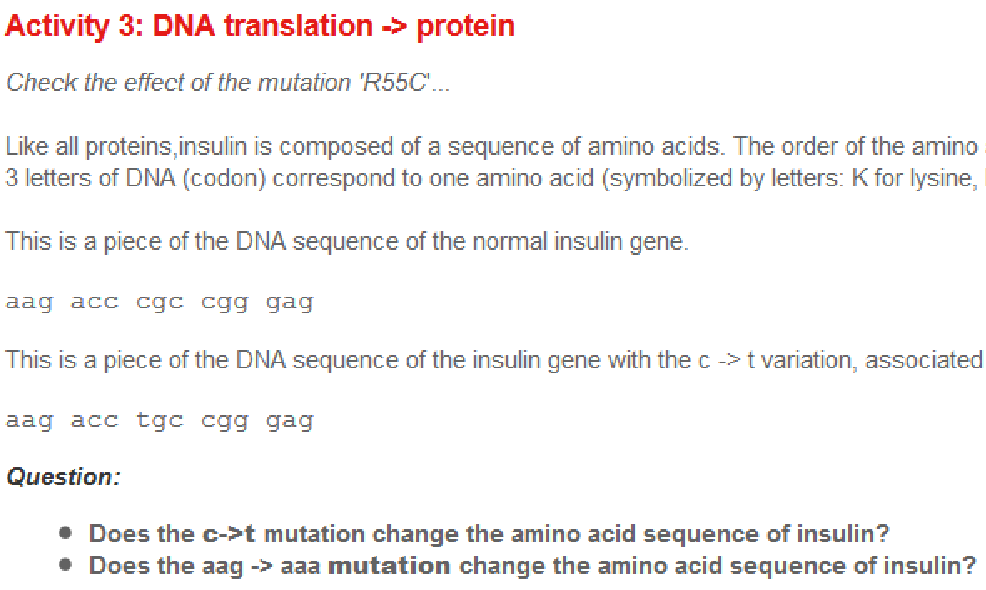
Varview

□ 2. rs5507 *[Homo sapiens]*

```
CCGCTGTTCCGGAACCTGCTCTGCG[C/T]GGCACGTCCTGGCAGTGGGGCAGGT
```

Chromosome:	11:2160013
Gene:	INS-IGF2 (GeneView) INS (GeneView)
Functional Consequence:	intron variant
Allele Origin:	C(germline)/T(germline)
Clinical significance:	Benign
Validated:	no info
Global MAF:	A=0.0082/40
HGVS:	NC_000011.10:g.2160013G>A, NC_000011.9:g.2181243G>A, NG_007114.1:g.6182C>T, NM_000207.2:c.188-16C>T, NM_001042376.2:c.187+772C>T,

Activity 3



Activity 3: DNA translation -> protein

Check the effect of the mutation 'R55C'...

Like all proteins,insulin is composed of a sequence of amino acids. The order of the amino acids is determined by the nucleic acid sequence of the insulin gene. 3 letters of DNA (codon) correspond to one amino acid (symbolized by letters: K for lysine, M for methionine, etc.).

This is a piece of the DNA sequence of the normal insulin gene.

```
aag acc cgc cgg gag
```

This is a piece of the DNA sequence of the insulin gene with the c -> t variation, associated with type I diabetes.

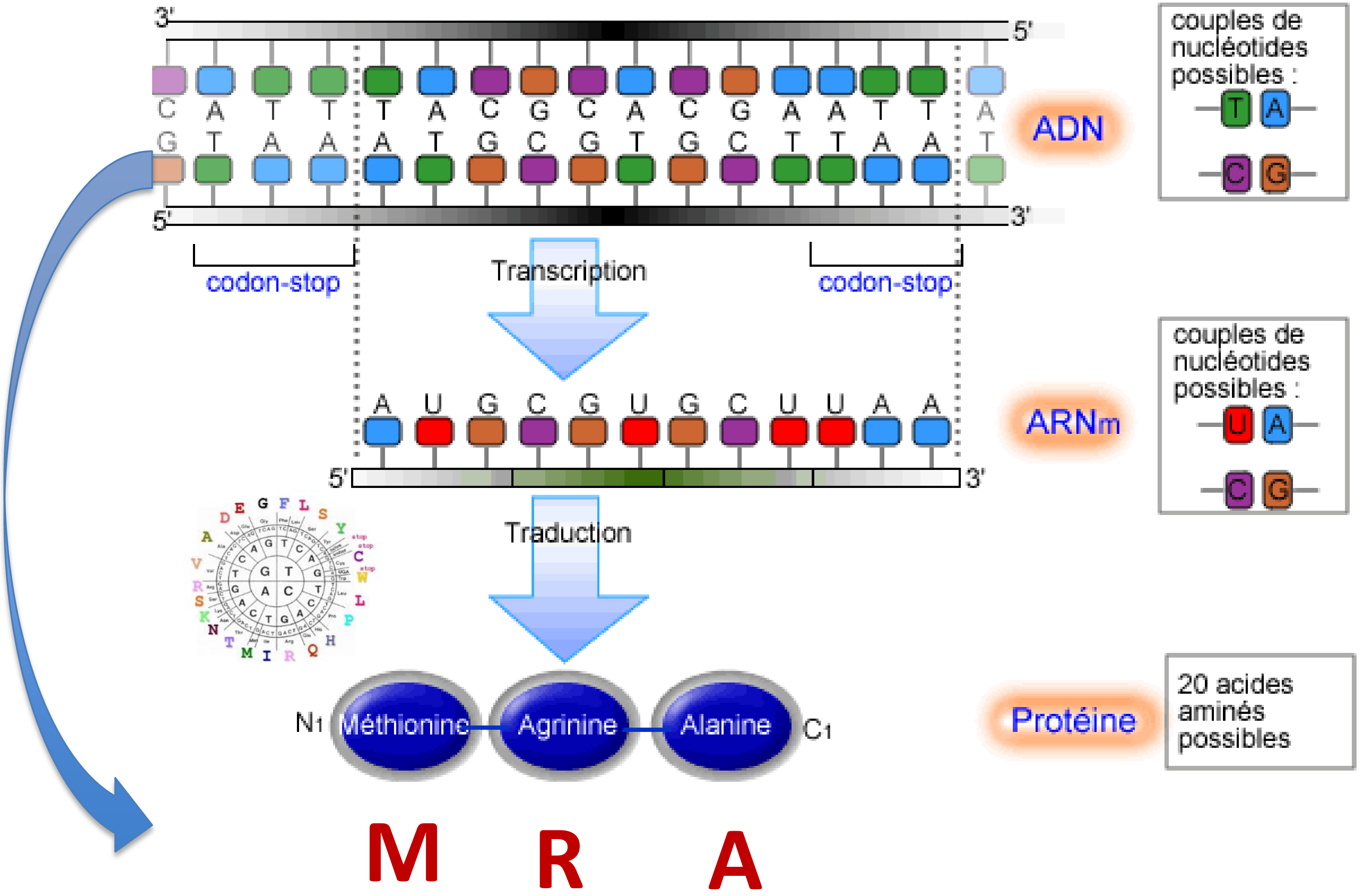
```
aag acc tgc cgg gag
```

Question:

- **Does the c->t mutation change the amino acid sequence of insulin?**
- **Does the aag -> aaa mutation change the amino acid sequence of insulin?**

http://education.expasy.org/bioinformatique/Diabetes.html

Global Organisation for Bioinformatics Learning, Education & Training


3'
5'
C A T T T A C G C A C G A A T T A
G T A A A T G C G T G C T T A A T
5'
3'
ADN
couples de nucléotides possibles :
T A
C G
codon-stop
Transcription
codon-stop
A U G C G U G C U U A A
5'
3'
ARNm
couples de nucléotides possibles :
U A
C G
Traduction
N1
Méthionine
Agrinine
Alanine
C1
Protéine
20 acides aminés possibles
M
R
A

http://fr.wikipedia.org/wiki/Synth%C3%A8se_des_prot%C3%A9ines#mediaviewer/File:Proteines.png

Global Organisation for Bioinformatics Learning, Education & Training

Activity 3: DNA translation -> protein

Check the effect of the mutation 'R55C'...

Like all proteins,insulin is composed of a sequence of amino acids. The order of the amino acids is determined by the nucleic acid sequence of the insulin gene.
3 letters of DNA (codon) correspond to one amino acid (symbolized by letters: K for lysine, M for methionine, etc.).

This is a piece of the DNA sequence of the normal insulin gene.

```
aag acc cgc cgg gag
```

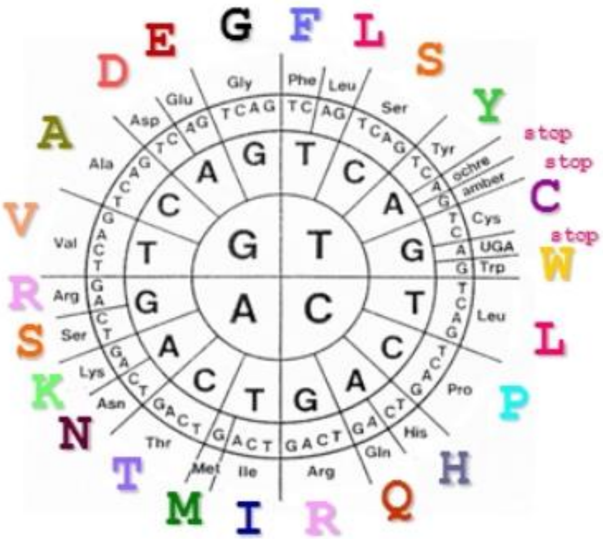
This is a piece of the DNA sequence of the insulin gene with the c -> t variation, associated with type I diabetes.

```
aag acc tgc cgg gag
```

Question:

- **Does the c->t mutation change the amino acid sequence of insulin?**
- **Does the aag -> aaa mutation change the amino acid sequence of insulin?**

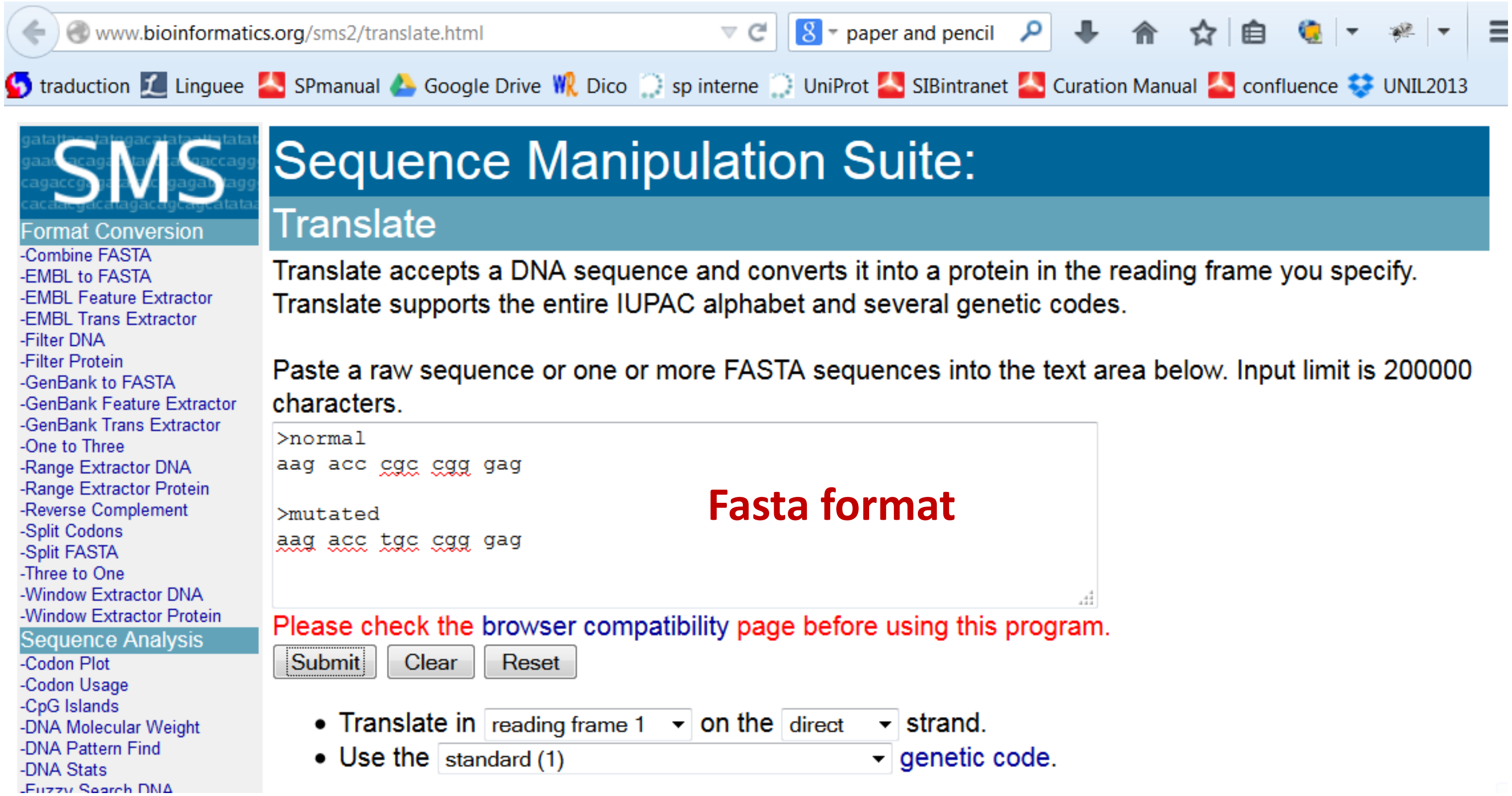
You could manually translate the nucleic acid sequences into amino acid sequences ('1 'letter code) using the genetic code below: :



Global Organisation for Bioinformatics Learning, Education & Training

SIB
Swiss Institute of Bioinformatics

Bioinformatics – Translate tool
http://www.bioinformatics.org/sms2/translate.html



Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

```
>normal
aag acc cgc cgg gag

>mutated
aag acc tgc cgg gag
```

Fasta format

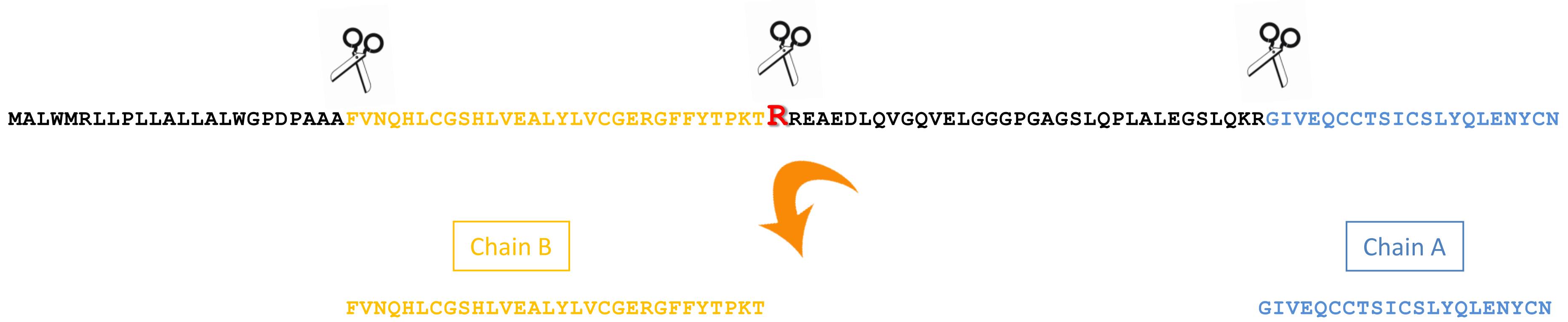
Please check the browser compatibility page before using this program.

- Translate in reading frame 1 on the direct strand.
- Use the standard (1) genetic code.

Translate results

```
>rf 1 normal
KTRRE

>rf 1 mutated
KTCRE
```

Amino acid sequence of the 'normal' insulin



MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

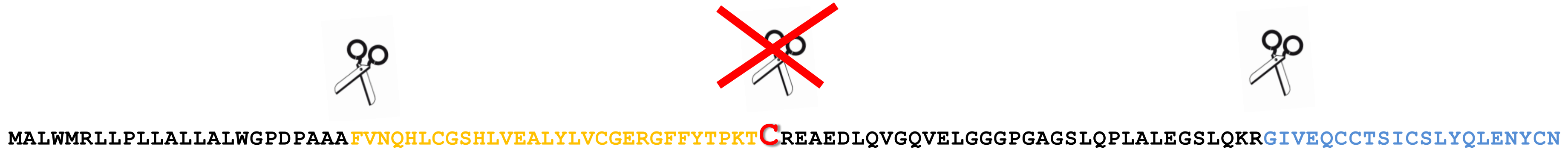
Chain B

FVNQHLCGSHLVEALYLVCGERGFFYTPKT

Chain A

GIVEQCCTSICSLYQLENYCN

Amino acid sequence of the 'mutated' insulin (variant c-> t; R 55 C)



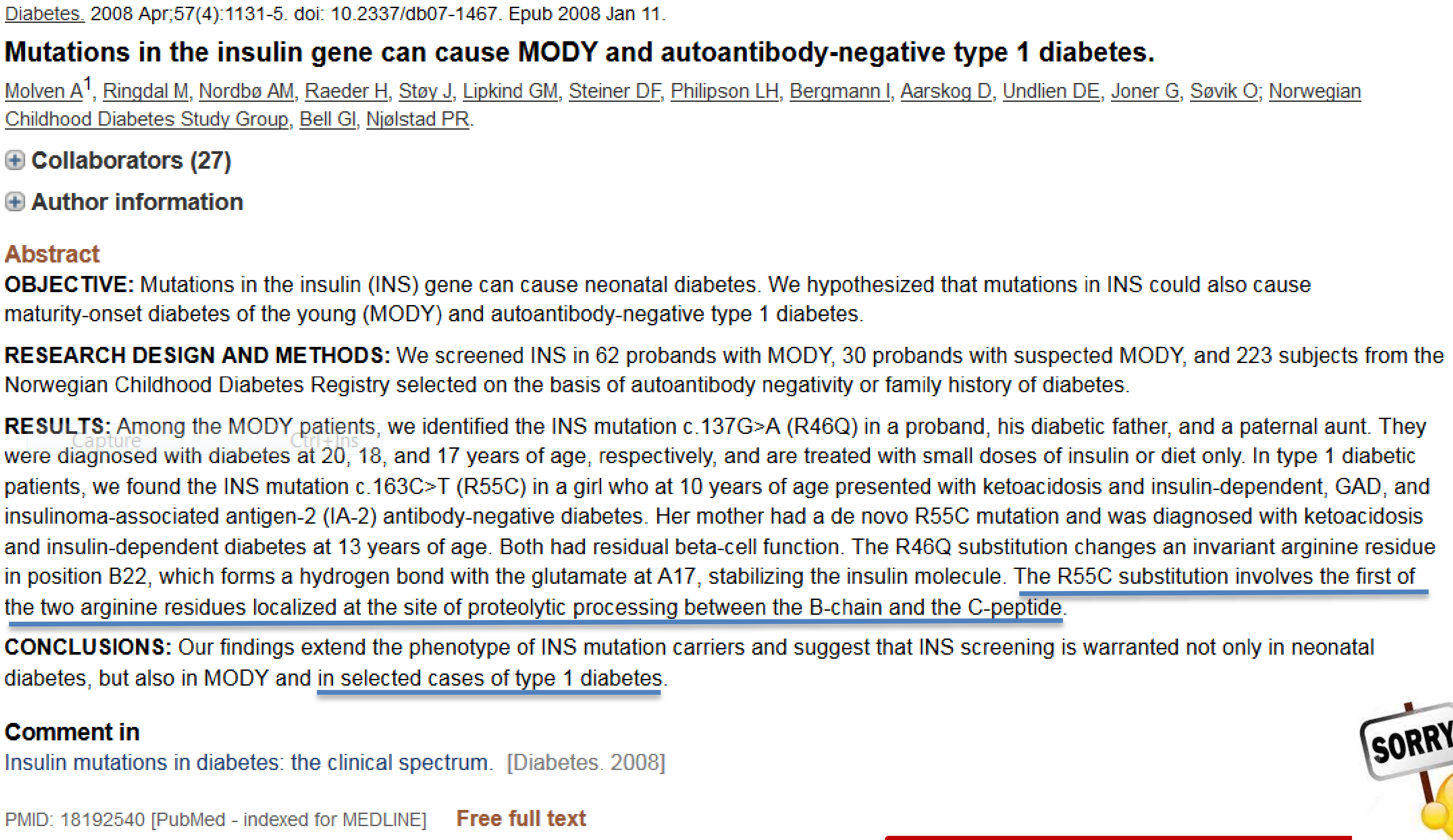
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTCREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

The mutation R -> C prevents insulin from being cut and thus from being biologically active

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics

http://www.ncbi.nlm.nih.gov/pubmed/18192540



Diabetes. 2008 Apr;57(4):1131-5. doi: 10.2337/db07-1467. Epub 2008 Jan 11.

Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes.

Molven A[1], Ringdal M, Nordbø AM, Raeder H, Støy J, Lipkind GM, Steiner DF, Philipson LH, Bergmann I, Aarskog D, Undlien DE, Joner G, Søvik O; Norwegian Childhood Diabetes Study Group, Bell GI, Njølstad PR.

⊕ Collaborators (27)

⊕ Author information

Abstract

OBJECTIVE: Mutations in the insulin (INS) gene can cause neonatal diabetes. We hypothesized that mutations in INS could also cause maturity-onset diabetes of the young (MODY) and autoantibody-negative type 1 diabetes.

RESEARCH DESIGN AND METHODS: We screened INS in 62 probands with MODY, 30 probands with suspected MODY, and 223 subjects from the Norwegian Childhood Diabetes Registry selected on the basis of autoantibody negativity or family history of diabetes.

RESULTS: Among the MODY patients, we identified the INS mutation c.137G>A (R46Q) in a proband, his diabetic father, and a paternal aunt. They were diagnosed with diabetes at 20, 18, and 17 years of age, respectively, and are treated with small doses of insulin or diet only. In type 1 diabetic patients, we found the INS mutation c.163C>T (R55C) in a girl who at 10 years of age presented with ketoacidosis and insulin-dependent, GAD, and insulinoma-associated antigen-2 (IA-2) antibody-negative diabetes. Her mother had a de novo R55C mutation and was diagnosed with ketoacidosis and insulin-dependent diabetes at 13 years of age. Both had residual beta-cell function. The R46Q substitution changes an invariant arginine residue in position B22, which forms a hydrogen bond with the glutamate at A17, stabilizing the insulin molecule. The R55C substitution involves the first of the two arginine residues localized at the site of proteolytic processing between the B-chain and the C-peptide.

CONCLUSIONS: Our findings extend the phenotype of INS mutation carriers and suggest that INS screening is warranted not only in neonatal diabetes, but also in MODY and in selected cases of type 1 diabetes.

Comment in

Insulin mutations in diabetes: the clinical spectrum. [Diabetes. 2008]

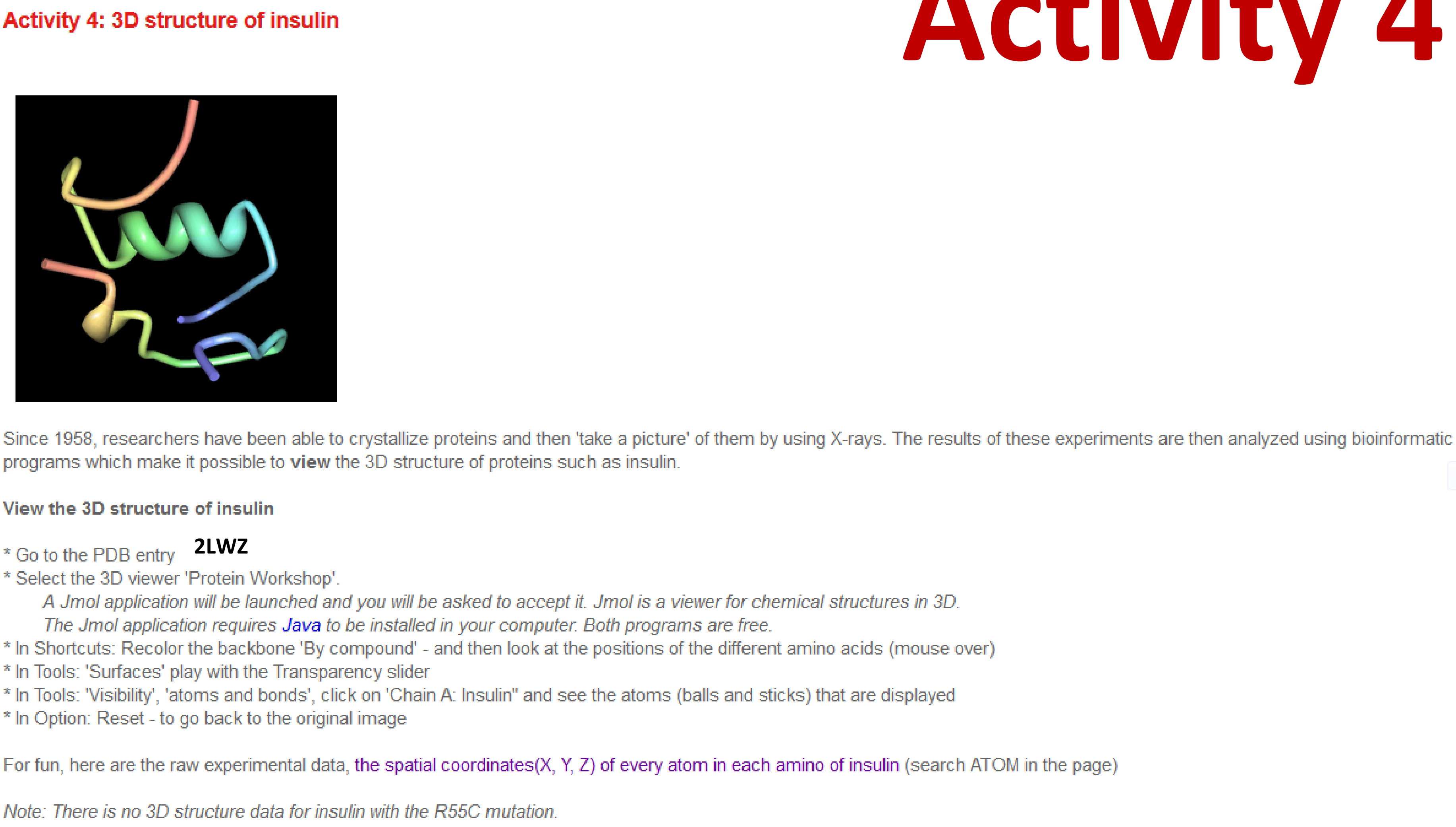
PMID: 18192540 [PubMed - indexed for MEDLINE] **Free full text**

SORRY

This publication is not available as free 'full text' in PubMed Central (PMC).
For full text:
http://education.expasy.org/cours/Toronto/

Global Organisation for Bioinformatics Learning, Education & Training

Activity 4



Activity 4: 3D structure of insulin

Since 1958, researchers have been able to crystallize proteins and then 'take a picture' of them by using X-rays. The results of these experiments are then analyzed using bioinformatic programs which make it possible to **view** the 3D structure of proteins such as insulin.

View the 3D structure of insulin

* Go to the PDB entry **2LWZ**
* Select the 3D viewer 'Protein Workshop'.

A Jmol application will be launched and you will be asked to accept it. Jmol is a viewer for chemical structures in 3D.
The Jmol application requires Java to be installed in your computer. Both programs are free.

* In Shortcuts: Recolor the backbone 'By compound' - and then look at the positions of the different amino acids (mouse over)
* In Tools: 'Surfaces' play with the Transparency slider
* In Tools: 'Visibility', 'atoms and bonds', click on 'Chain A: Insulin" and see the atoms (balls and sticks) that are displayed
* In Option: Reset - to go back to the original image

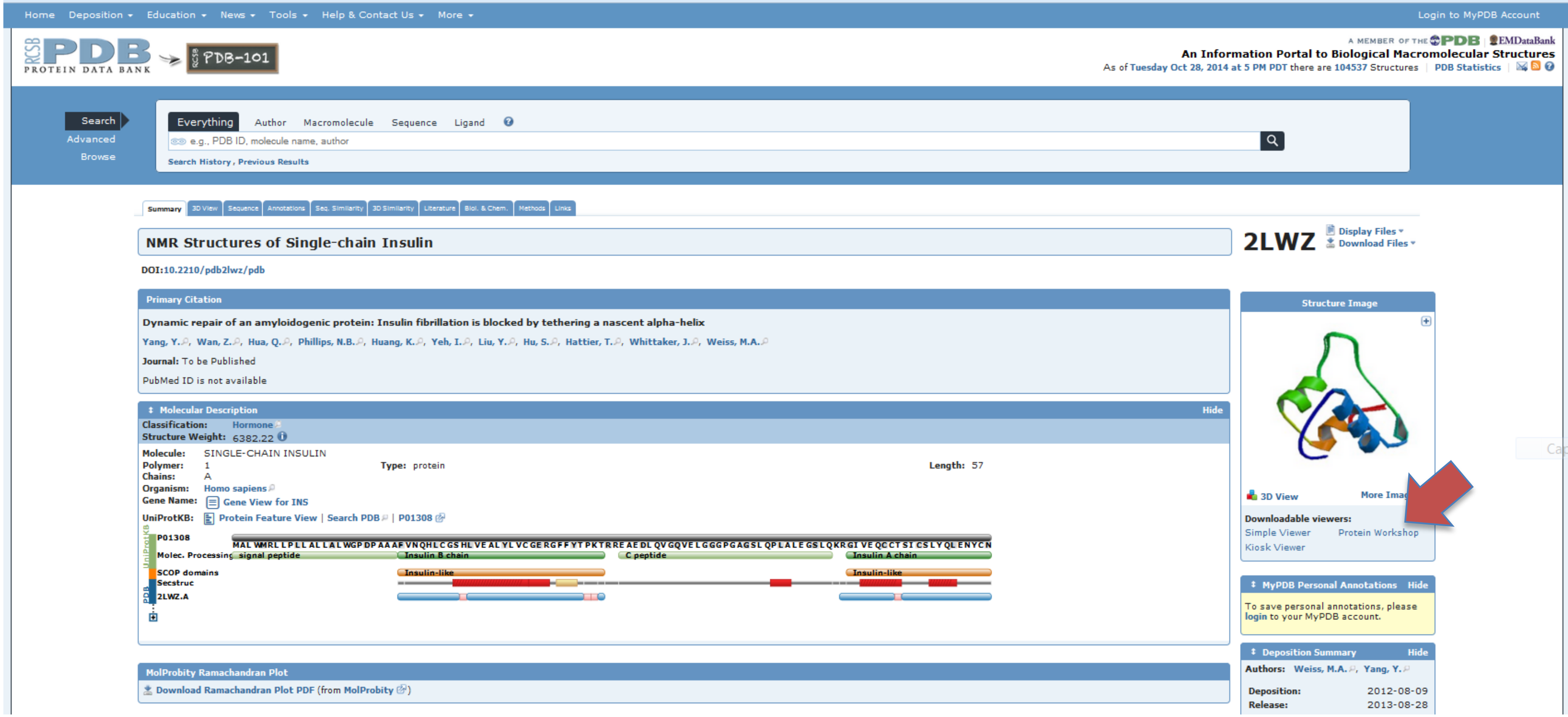
For fun, here are the raw experimental data, the spatial coordinates(X, Y, Z) of every atom in each amino of insulin (search ATOM in the page)

Note: There is no 3D structure data for insulin with the R55C mutation.

Global Organisation for Bioinformatics Learning, Education & Training

http://education.expasy.org/bioinformatique/Diabetes.html

http://www.pdb.org/pdb/explore/explore.do?structureId=2LWZ



An Information Portal to Biological Macromolecular Structures

As of Tuesday Oct 28, 2014 at 5 PM PDT there are 104537 Structures | PDB Statistics

NMR Structures of Single-chain Insulin

2LWZ

DOI:10.2210/pdb2lwz/pdb

Primary Citation

Dynamic repair of an amyloidogenic protein: Insulin fibrillation is blocked by tethering a nascent alpha-helix

Yang, Y., Wan, Z., Hua, Q., Phillips, N.B., Huang, K., Yeh, I., Liu, Y., Hu, S., Hattier, T., Whittaker, J., Weiss, M.A.

Journal: To be Published

PubMed ID is not available

Molecular Description Hide

Classification: Hormone
Structure Weight: 6382.22

Molecule: SINGLE-CHAIN INSULIN
Polymer: 1 **Type:** protein **Length:** 57
Chains: A
Organism: Homo sapiens
Gene Name: Gene View for INS
UniProtKB: Protein Feature View | Search PDB | P01308

Downloadable viewers:
Simple Viewer Protein Workshop
Kiosk Viewer

MolProbity Ramachandran Plot

Download Ramachandran Plot PDF (from MolProbity)

MyPDB Personal Annotations Hide

To save personal annotations, please login to your MyPDB account.

Deposition Summary Hide

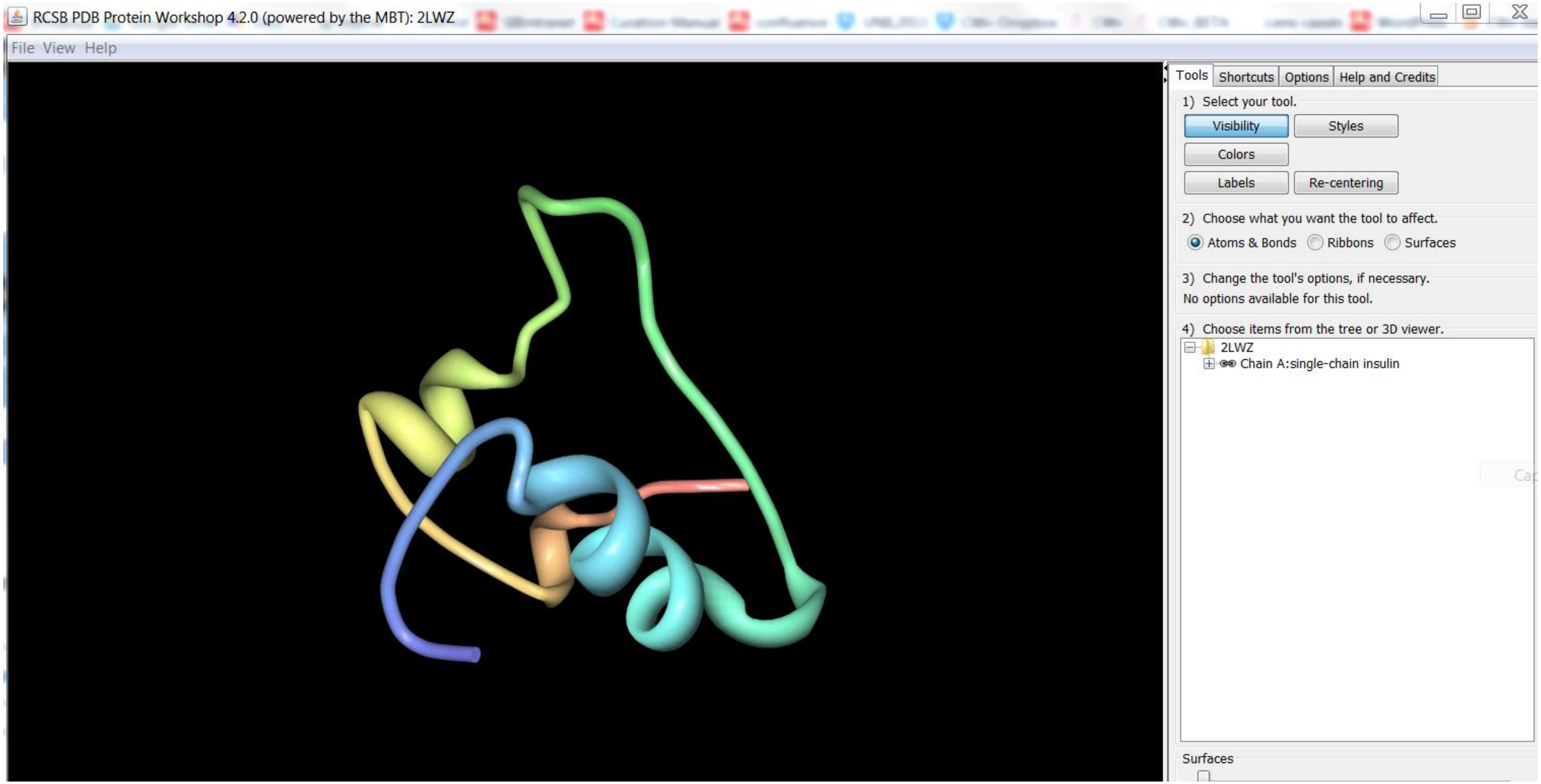
Authors: Weiss, M.A., Yang, Y.

Deposition:	2012-08-09
Release:	2013-08-28

..... requires Java to be installed in your computer.

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics – Insulin 3D structure in PDB database (2LWZ) (Protein Workshop) http://www.pdb.org/pdb/explore/explore.do?structureId=2LWZ


RCSB PDB Protein Workshop 4.2.0 (powered by the MBT): 2LWZ
File View Help
Tools Shortcuts Options Help and Credits
Visibility
Styles
Colors
Labels
Re-centering
2LWZ
Chain A:single-chain insulin
Surfaces

1) Select your tool.

2) Choose what you want the tool to affect.

● Atoms & Bonds ○ Ribbons ○ Surfaces

3) Change the tool's options, if necessary.

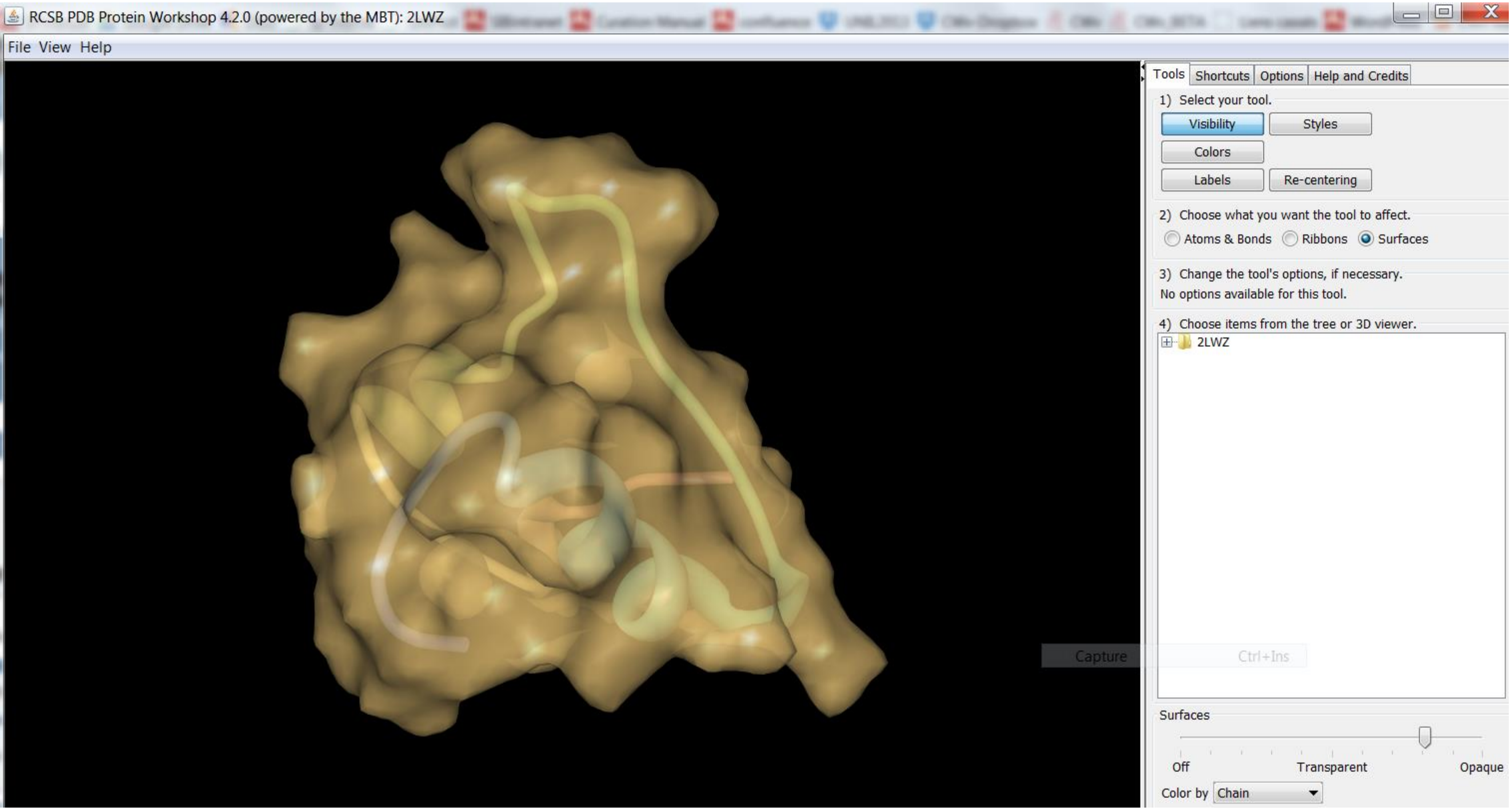
No options available for this tool.

4) Choose items from the tree or 3D viewer.

Visualization tool: Protein Workshop

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics



RCSB PDB Protein Workshop 4.2.0 (powered by the MBT): 2LWZ

1) Select your tool.

2) Choose what you want the tool to affect.

○ Atoms & Bonds ○ Ribbons ● Surfaces

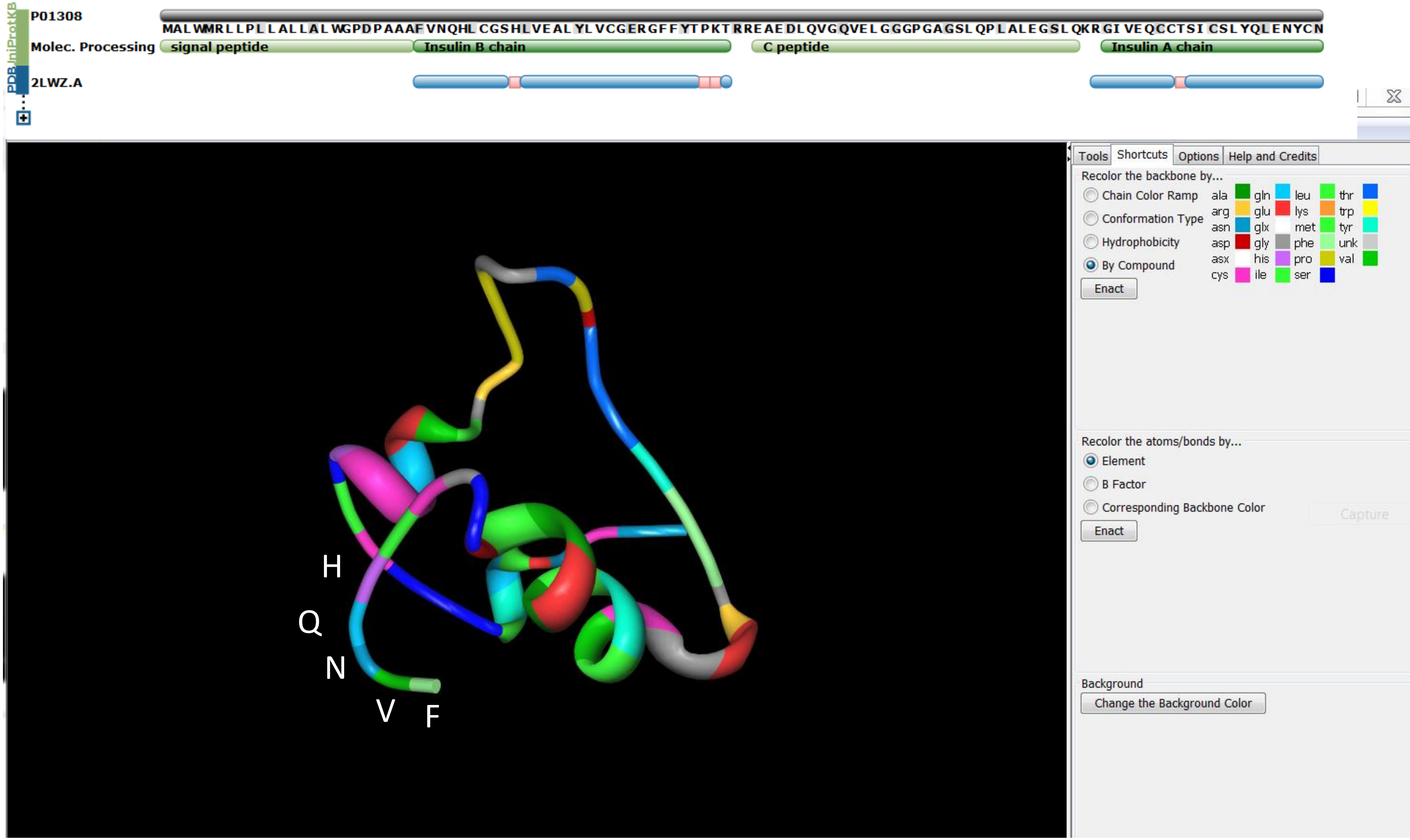
3) Change the tool's options, if necessary.

No options available for this tool.

4) Choose items from the tree or 3D viewer.

Color by: Chain

Protein Workshop: in Tools: 'Surfaces' play with the Transparency slider



Recolor the backbone by...

○ Chain Color Ramp
○ Conformation Type
○ Hydrophobicity
● By Compound

ala	🟩	gln	🟦	leu	🟩	thr	🟦
arg	🟨	glu	🟥	lys	🟧	trp	🟨
asn	🟦	glx	⬜	met	🟩	tyr	🟦
asp	🟥	gly	⬜	phe	🟩	unk	⬜
asx	⬜	his	🟪	pro	🟨	val	🟩
cys	🟪	ile	🟩	ser	🟦		

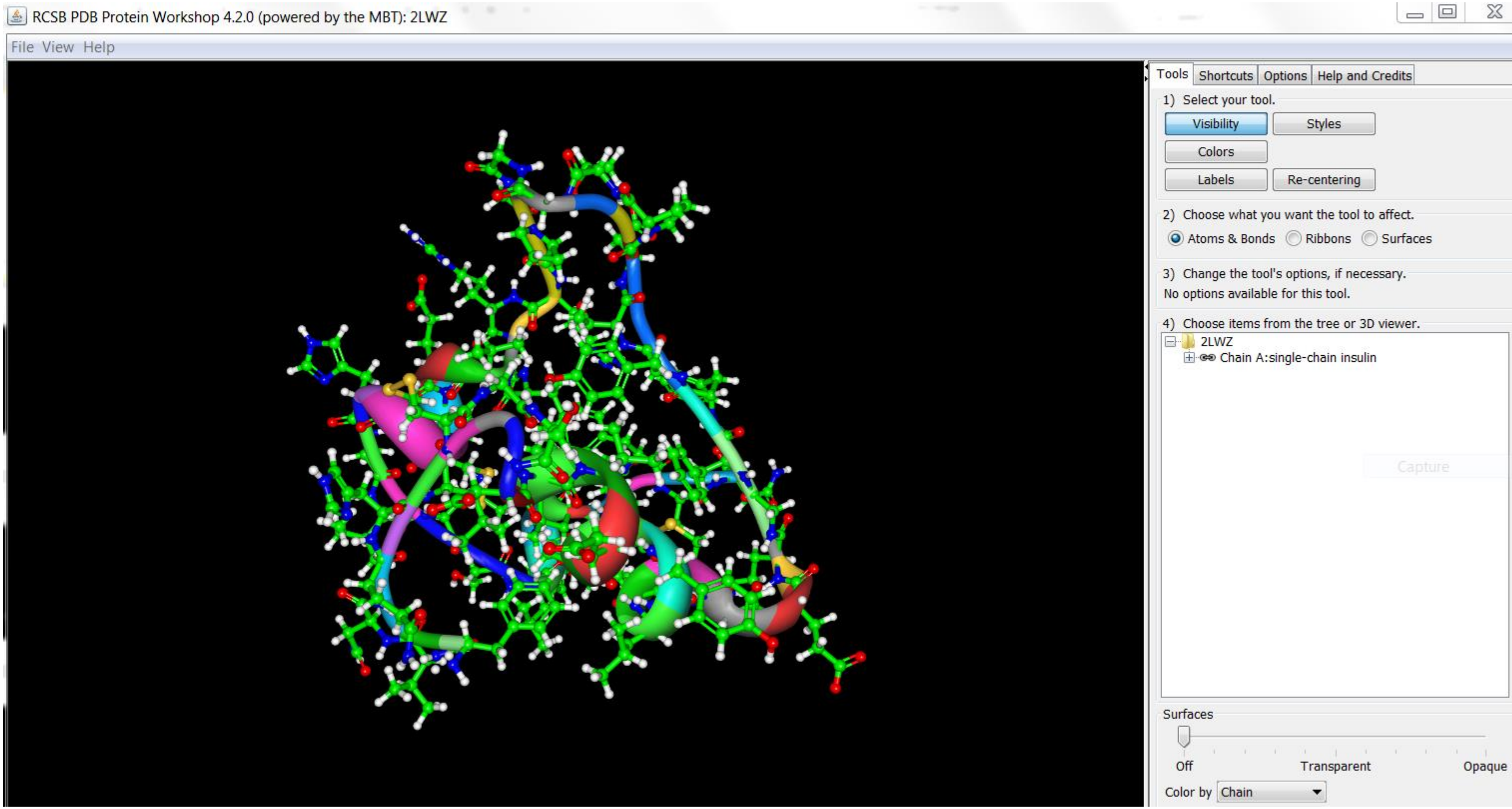
Recolor the atoms/bonds by...

● Element
○ B Factor
○ Corresponding Backbone Color

Background

Protein workshop:
In Shortcuts: Recolor the backbone 'By compound' - and then look at the positions of the different amino acids (mouse over)

Global Organisation for Bioinformatics Learning, Education & Training



1) Select your tool.

2) Choose what you want the tool to affect.

● Atoms & Bonds ○ Ribbons ○ Surfaces

3) Change the tool's options, if necessary.

No options available for this tool.

4) Choose items from the tree or 3D viewer.

2LWZ

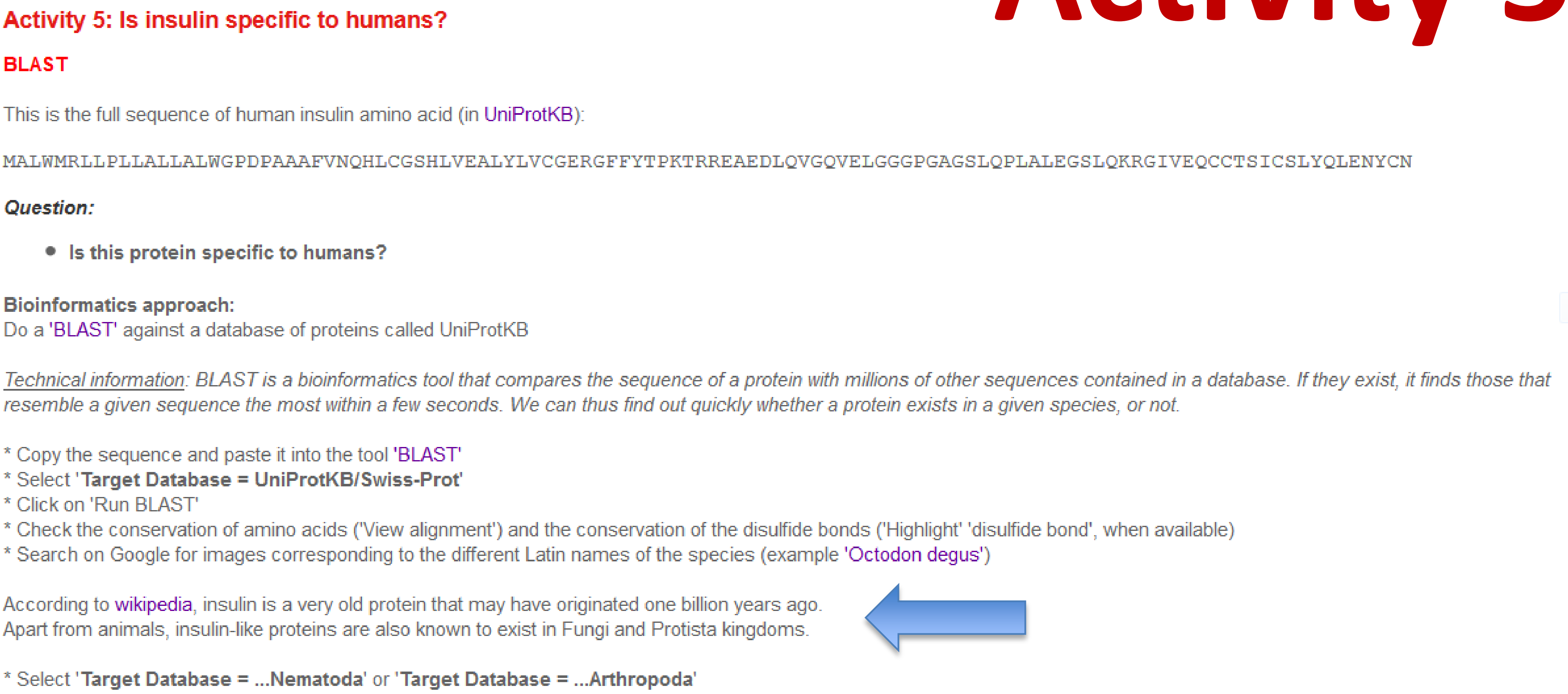
Chain A:single-chain insulin

Color by: Chain

Protein workshop: In Tools: 'Visibility', 'atoms and bonds', click on 'Chain A: Insulin" and see the atoms (balls and sticks) that are displayed

Global Organisation for Bioinformatics Learning, Education & Training

Activity 5



Activity 5: Is insulin specific to humans?

BLAST

This is the full sequence of human insulin amino acid (in UniProtKB):

```
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

Question:

- **Is this protein specific to humans?**

Bioinformatics approach:
Do a 'BLAST' against a database of proteins called UniProtKB

Technical information: BLAST is a bioinformatics tool that compares the sequence of a protein with millions of other sequences contained in a database. If they exist, it finds those that resemble a given sequence the most within a few seconds. We can thus find out quickly whether a protein exists in a given species, or not.

* Copy the sequence and paste it into the tool 'BLAST'
* Select '**Target Database = UniProtKB/Swiss-Prot**'
* Click on 'Run BLAST'
* Check the conservation of amino acids ('View alignment') and the conservation of the disulfide bonds ('Highlight' 'disulfide bond', when available)
* Search on Google for images corresponding to the different Latin names of the species (example 'Octodon degus')

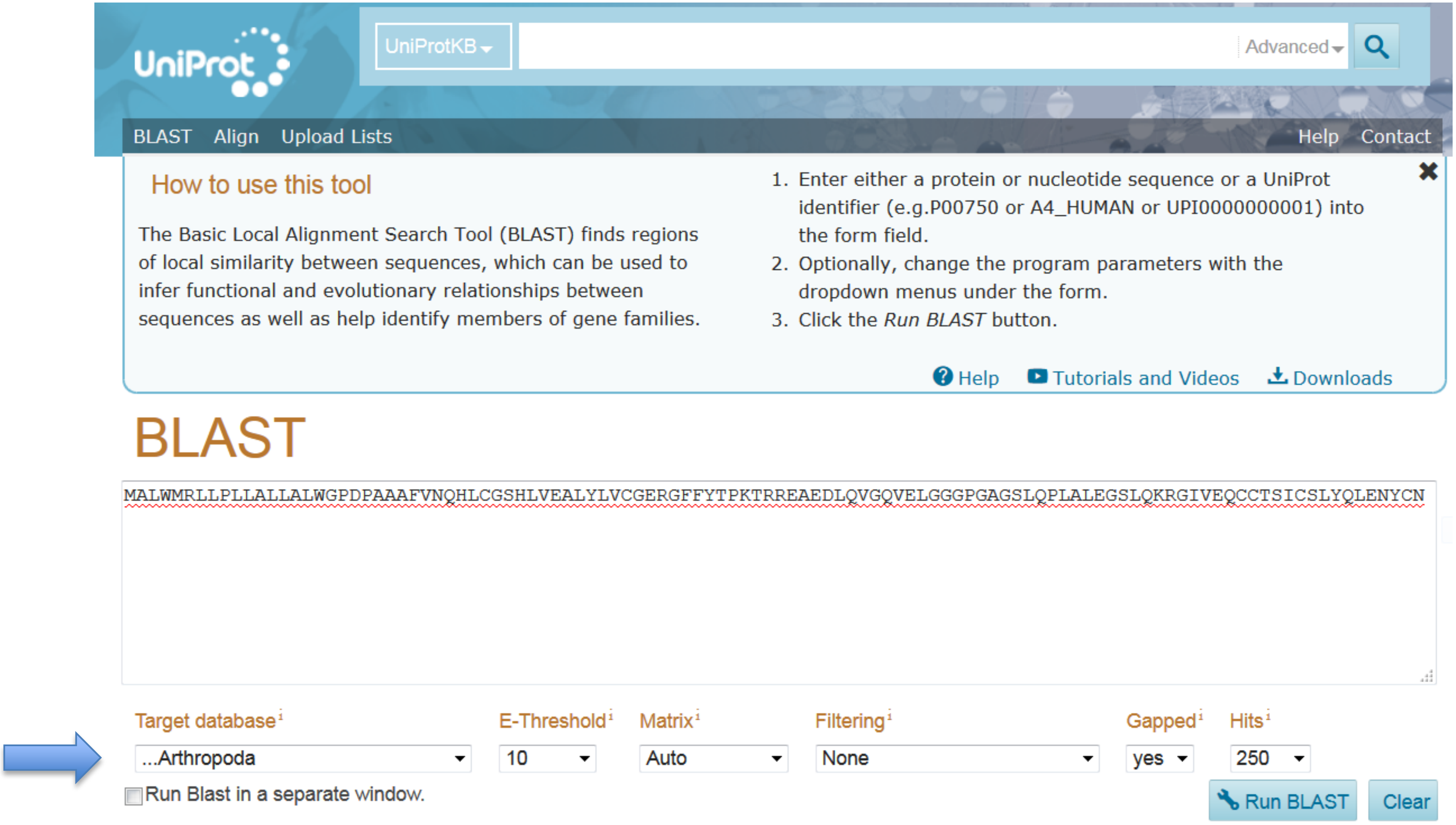
According to wikipedia, insulin is a very old protein that may have originated one billion years ago.
Apart from animals, insulin-like proteins are also known to exist in Fungi and Protista kingdoms.

* Select '**Target Database = ...Nematoda**' or '**Target Database = ...Arthropoda**'

http://education.expasy.org/bioinformatique/Diabetes.html

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics – BLAST Similarity search tool
www.uniprot.org/blast/



How to use this tool

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences, which can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

1. Enter either a protein or nucleotide sequence or a UniProt identifier (e.g.P00750 or A4_HUMAN or UPI0000000001) into the form field.
2. Optionally, change the program parameters with the dropdown menus under the form.
3. Click the *Run BLAST* button.

BLAST

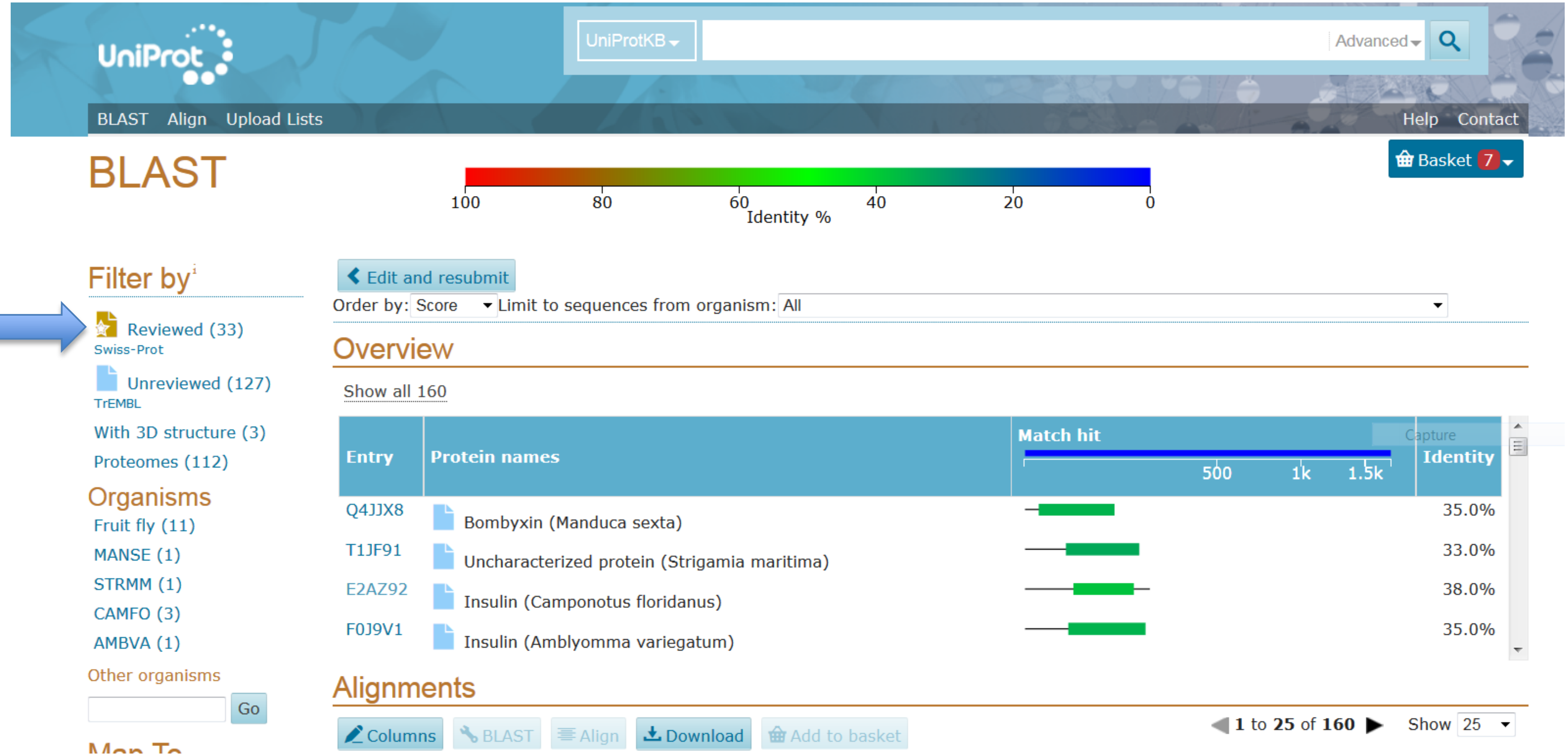
```
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

Target database	E-Threshold	Matrix	Filtering	Gapped	Hits
...Arthropoda	10	Auto	None	yes	250

□ Run Blast in a separate window.

Global Organisation for Bioinformatics Learning, Education & Training

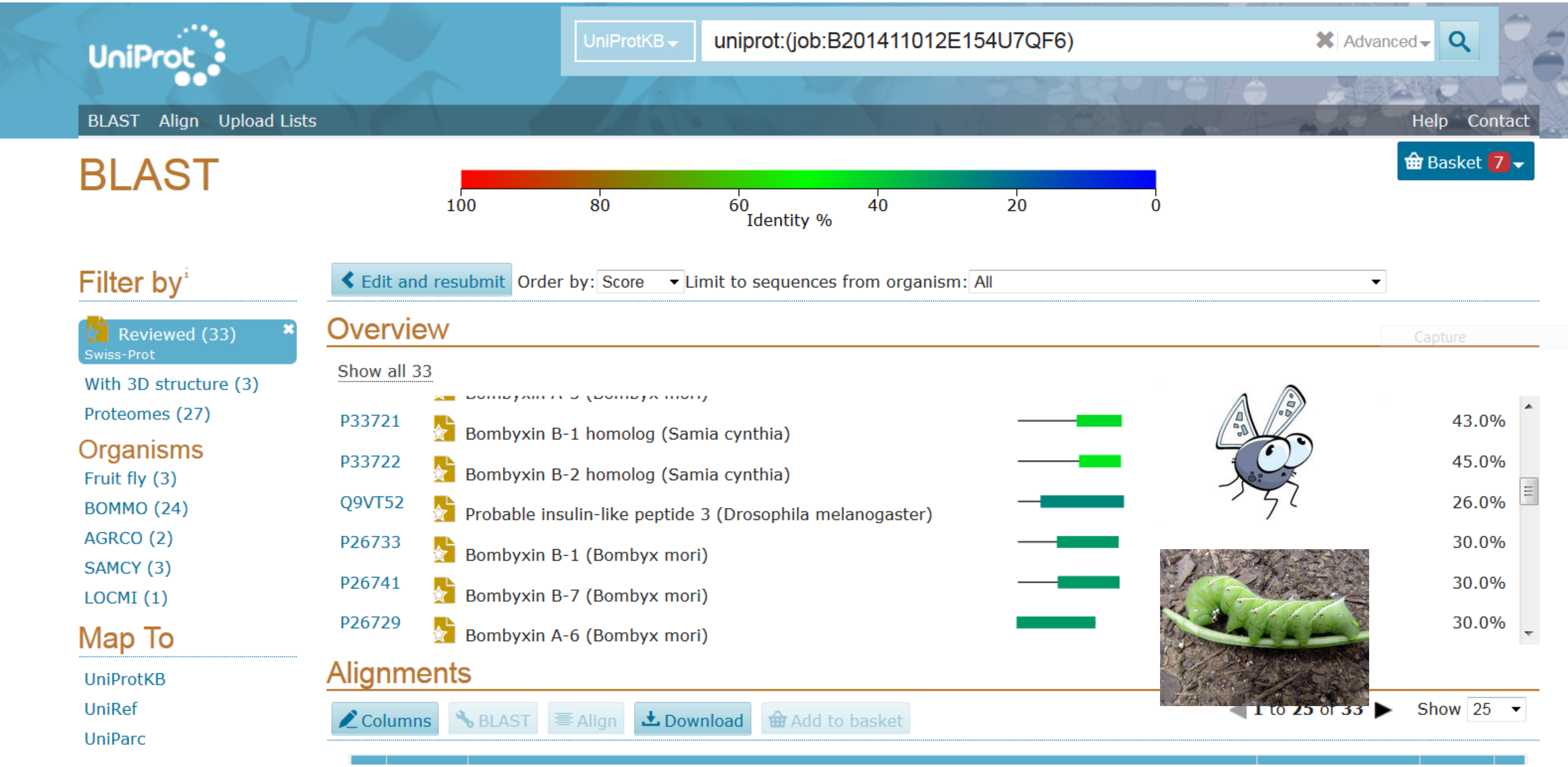
SIB
Swiss Institute of Bioinformatics



Order by: Score Limit to sequences from organism: All

Entry	Protein names	Match hit	Identity
Q4JJX8	Bombyxin (Manduca sexta)		35.0%
T1JF91	Uncharacterized protein (Strigamia maritima)		33.0%
E2AZ92	Insulin (Camponotus floridanus)		38.0%
F0J9V1	Insulin (Amblyomma variegatum)		35.0%

‘reviewed’ entries (UniProtKB/Swiss-Prot section) are manually reviewed
‘unreviewed’ entries (UniProtKB/TrEMBL section) are automatically annotated



uniprot:(job:B201411012E154U7QF6)

BLAST

Order by: Score

Limit to sequences from organism: All

Filter by

Reviewed (33) Swiss-Prot

With 3D structure (3)

Proteomes (27)

Organisms

Fruit fly (3)

BOMMO (24)

AGRCO (2)

SAMCY (3)

LOCMI (1)

Map To

UniProtKB

UniRef

UniParc

Overview

Show all 33

	Bombyxin A-5 (Bombyx mori)	
P33721	Bombyxin B-1 homolog (Samia cynthia)	43.0%
P33722	Bombyxin B-2 homolog (Samia cynthia)	45.0%
Q9VT52	Probable insulin-like peptide 3 (Drosophila melanogaster)	26.0%
P26733	Bombyxin B-1 (Bombyx mori)	30.0%
P26741	Bombyxin B-7 (Bombyx mori)	30.0%
P26729	Bombyxin A-6 (Bombyx mori)	30.0%

Alignments

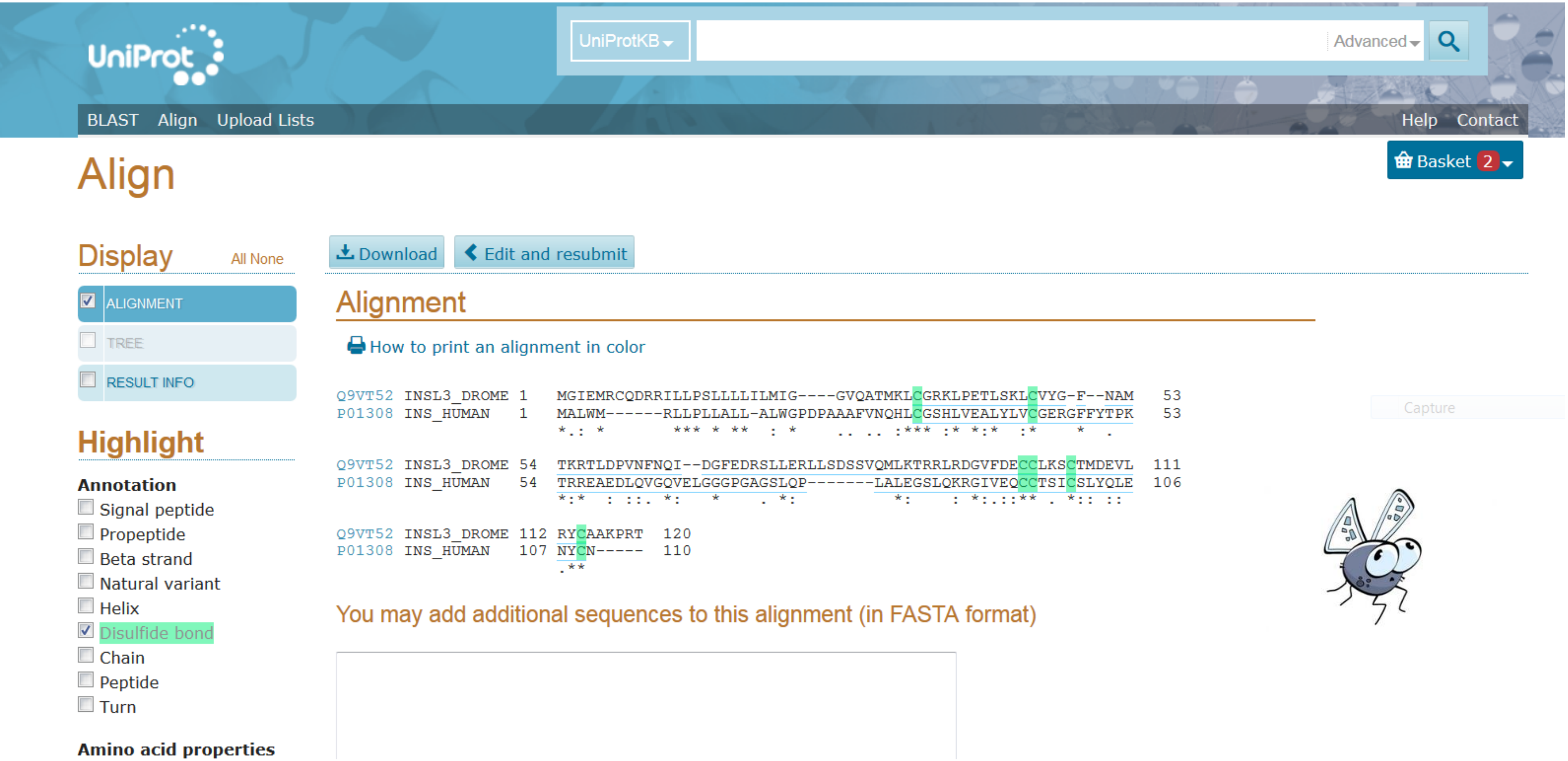
1 to 25 of 33

Show 25

How similar are the human and drosophila sequences ?

```
>sp|Q9VT52|INSL3_DROME Probable insulin-like peptide 3 OS=Drosophila melanogaster
GN=Ilp3 PE=2 SV=2
MGIEMRCQDRRILLPSLLLLILMIGGVQATMKLCGRKLPETLSKLCVYGFNAMTKRTLDP
VNFNQIDGFEDRSLLERLLSDSSVQMLKTRRLRDGVFDECCLKSCTMDEVLRYCAAKPRT

>sp|P01308|INS_HUMAN Insulin OS=Homo sapiens GN=INS PE=1 SV=1
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

Bioinformatics – alignment tool
www.uniprot.org/align/



Align

Display

All None

☑ ALIGNMENT

☐ TREE

☐ RESULT INFO

Highlight

Annotation

☐ Signal peptide
☐ Propeptide
☐ Beta strand
☐ Natural variant
☐ Helix
☑ Disulfide bond
☐ Chain
☐ Peptide
☐ Turn

Amino acid properties

Alignment

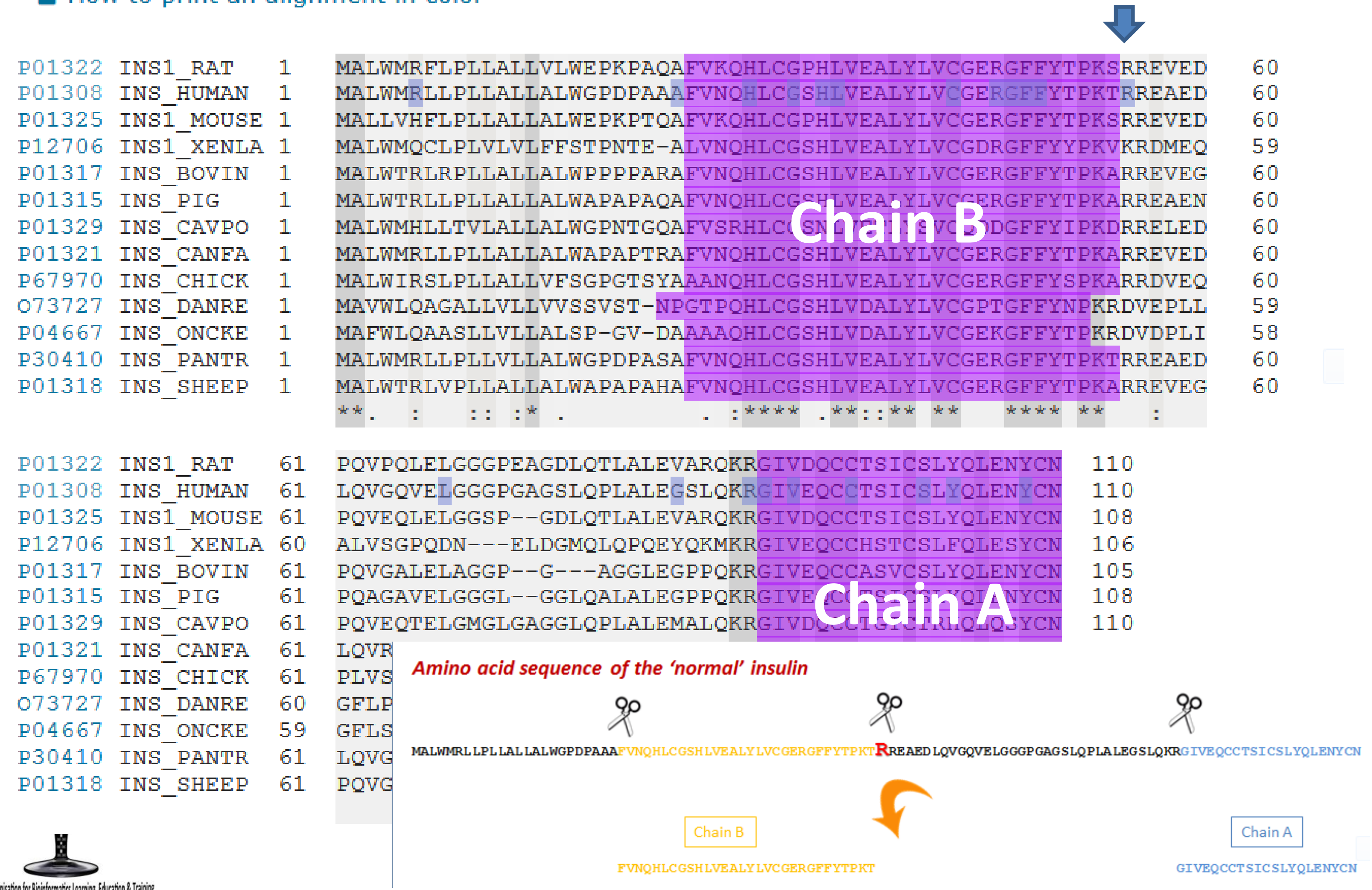
```
Q9VT52 INSL3_DROME 1   MGIEMRCQDRRILLPSLLLLILMIG----GVQATMKLCGRKLPETLSKLCVYG-F--NAM   53
P01308 INS_HUMAN   1   MALWM------RLLPLLALL-ALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPK   53
                       *.: *       *** * **  : *    .. .. :*** :* *:*  :*    *  .

Q9VT52 INSL3_DROME 54  TKRTLDPVNFNQI--DGFEDRSLLERLLSDSSVQMLKTRRLRDGVFDECCLKSCTMDEVL  111
P01308 INS_HUMAN   54  TRREAEDLQVGQVELGGGPGAGSLQP-------LALEGSLQKRGIVEQCCTSICSLYQLE  106
                       *:*  : ::. *:   *    . *:          *:    : *:.::** . *:: ::

Q9VT52 INSL3_DROME 112 RYCAAKPRT  120
P01308 INS_HUMAN   107 NYCN-----  110
                       .**
```

You may add additional sequences to this alignment (in FASTA format)

Alignment

Positively charged amino acids (R @ position 55 in human) seem to be highly conserved during evolution

How to print an alignment in color



```
P01322 INS1_RAT    1   MALWMRFLPLLALLVLWEPKPAQAFVKQHLCGPHLVEALYLVCGERGFFYTPKSRREVED   60
P01308 INS_HUMAN   1   MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED   60
P01325 INS1_MOUSE  1   MALLVHFLPLLALLALWEPKPTQAFVKQHLCGPHLVEALYLVCGERGFFYTPKSRREVED   60
P12706 INS1_XENLA  1   MALWMQCLPLVLVLFFSTPNTE-ALVNQHLCGSHLVEALYLVCGDRGFFYYPKVKRDMEQ   59
P01317 INS_BOVIN   1   MALWTRLRPLLALLALWPPPPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEG   60
P01315 INS_PIG     1   MALWTRLLPLLALLALWAPAPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREAEN   60
P01329 INS_CAVPO   1   MALWMHLLTVLALLALWGPNTGQAFVSRHLCGSNLVETLYSVCQDDGFFYIPKDRRELED   60
P01321 INS_CANFA   1   MALWMRLLPLLALLALWAPAPTRAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVED   60
P67970 INS_CHICK   1   MALWIRSLPLLALLVFSGPGTSYAAANQHLCGSHLVEALYLVCGERGFFYSPKARRDVEQ   60
O73727 INS_DANRE   1   MAVWLQAGALLVLLVVSSVST-NPGTPQHLCGSHLVDALYLVCGPTGFFYNPKRDVEPLL   59
P04667 INS_ONCKE   1   MAFWLQAASLLVLLALSP-GV-DAAAAQHLCGSHLVDALYLVCGEKGFFYTPKRDVDPLI   58
P30410 INS_PANTR   1   MALWMRLLPLLVLLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED   60
P01318 INS_SHEEP   1   MALWTRLVPLLALLALWAPAPAHAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEG   60
                       **.  :   :: :* .         . :**** .**::** **   **** **    :

P01322 INS1_RAT    61  PQVPQLELGGGPEAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN   110
P01308 INS_HUMAN   61  LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN   110
P01325 INS1_MOUSE  61  PQVEQLELGGSP--GDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN   108
P12706 INS1_XENLA  60  ALVSGPQDN---ELDGMQLQPQEYQKMKRGIVEQCCHSTCSLFQLESYCN   106
P01317 INS_BOVIN   61  PQVGALELAGGP--G---AGGLEGPPQKRGIVEQCCASVCSLYQLENYCN   105
P01315 INS_PIG     61  PQAGAVELGGGL--GGLQALALEGPPQKRGIVEQCCTSICSLYQLENYCN   108
P01329 INS_CAVPO   61  PQVEQTELGMGLGAGGLQPLALEMALQKRGIVDQCCTGTCTRHQLQSYCN   110
P01321 INS_CANFA   61  LQVR
P67970 INS_CHICK   61  PLVS
O73727 INS_DANRE   60  GFLP
P04667 INS_ONCKE   59  GFLS
P30410 INS_PANTR   61  LQVG
P01318 INS_SHEEP   61  PQVG
```

Amino acid sequence of the 'normal' insulin

```
MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
```

Chain B

FVNQHLCGSHLVEALYLVCGERGFFYTPKT

Chain A

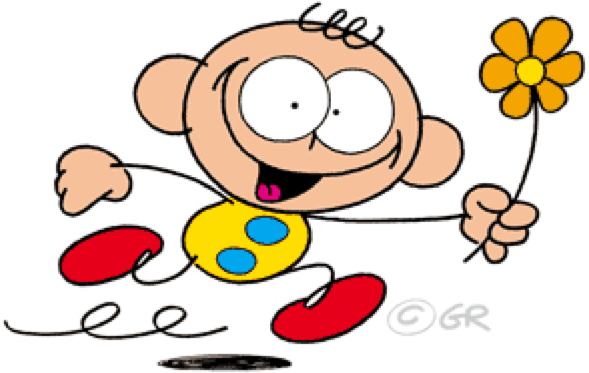
GIVEQCCTSICSLYQLENYCN

Global Organisation for Bioinformatics Learning, Education & Training

Many thanks to all of you

and

to Michelle Brazas



http://www2.grifil.com/album.html

Global Organisation for Bioinformatics Learning, Education & Training

SIB Swiss Institute of Bioinformatics