JOHANNES KEPLER
UNIVERSITÄT LINZ

LVA 320.004

Genome data analysis
Computer lab session 2

Theresa Schwarz MSc

theresa.schwarz@jku.at
Institute of Biophysics, JKU
Gruberstraße 40, 4020 Linz

Hints for Reports

- Prepare Reports in groups of **TWO** and submit the **same file**!
- Show all of your results in **SCREENSHOTS**!

Overview

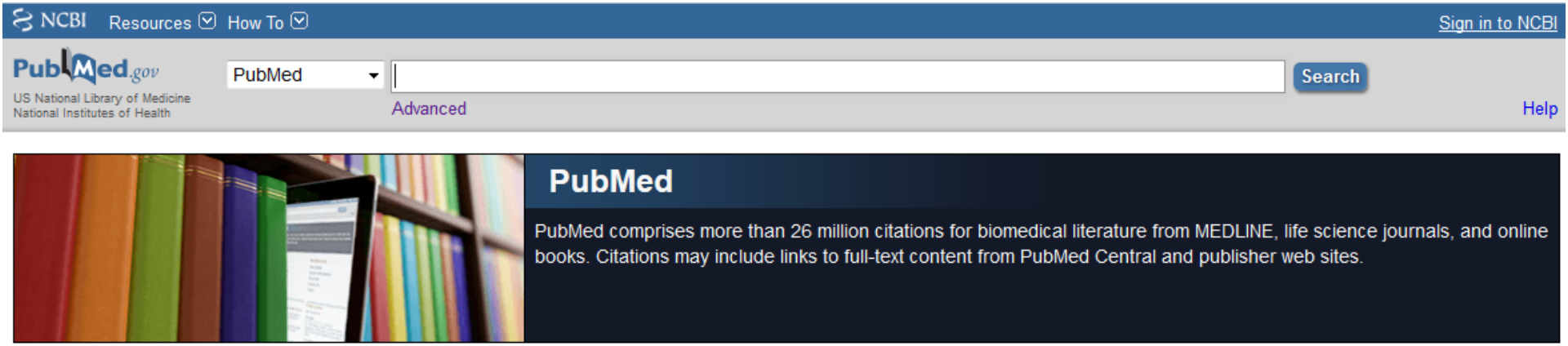
- **PubMed - Literature search**

- **SNP Database**
  - **Hardy-Weinberg Equilibrium**

- **Primer3Plus**
  - **Polymerase-Chain Reaction (PCR)**
  - **Primer design**

- **NEB Cutter**

PubMed – normal search



- http://www.nlm.nih.gov/bsd/disted/pubmed.html
- National Library of Medicine's search service
- Search for **LITERATURE**
- links to **online journals**

PubMed – normal search

JOHANNES KEPLER UNIVERSITÄT LINZ

Search for **Tiemann-Boege**

NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed | Tiemann-Boege

Create RSS Create alert Advanced

Format: Summary Sort by: Most Recent Per page: 20 Send to

Article types
Clinical Trial
Review
Customize ...

Text availability
Abstract
Free full text
Full text

PubMed Commons
Reader comments
Trending articles

Publication dates
5 years
10 years
Custom range...

Species
Humans
Other Animals

Clear all

Show additional filters

Search results

Items: 19

1. The consequences of sequence erosion in the evolution of recombination hotspots.
Tiemann-Boege I, Schwarz T, Striedner Y, Heissl A.
Philos Trans R Soc Lond B Biol Sci. 2017 Dec 19;372(1736). pii: 20160462. doi: 10.1098/rstb.2016.0462. Review.
PMID: 29109225 Free Article
Similar articles

2. Water transport through the intestinal epithelial barrier under different osmotic conditions is dependent on LI-cadherin trans-interaction.
Weth A, Dippl C, Striedner Y, **Tiemann-Boege** I, Vereshchaga Y, Golenhofen N, Bartelt-Kirbach B, Baumgartner W.
Tissue Barriers. 2017 Apr 3;5(2):e1285390. doi: 10.1080/21688370.2017.1285390. Epub 2017 Jan 24.
PMID: 28452574 Free PMC Article
Similar articles

3. The long zinc finger domain of PRDM9 forms a highly stable and long-lived complex with its DNA recognition sequence.
Striedner Y, Schwarz T, Welte T, Futschik A, Rant U, **Tiemann-Boege** I.
Chromosome Res. 2017 Jun;25(2):155-172. doi: 10.1007/s10577-017-9552-1. Epub 2017 Feb 2.
PMID: 28155083 Free PMC Article
Similar articles

PubMed – normal search

JOHANNES KEPLER UNIVERSITÄT LINZ

direct link to online paper

Format: Abstract

journal and publication date

Send to

Philos Trans R Soc Lond B Biol Sci. 2017 Dec 19;372(1736). pii: 20160462. doi: 10.1098/rstb.2016.0462.

title and authors

The consequences of sequence erosion in the evolution of recombination hotspots.

Tiemann-Boege I[1], Schwarz T[2], Striedner Y[2], Heissl A[2].

Author information

author information

1 Institute of Biophysics, Johannes Kepler University, Linz, Gruberstraße 40, 4020 Linz, Austria irene.tiemann@jku.at.
2 Institute of Biophysics, Johannes Kepler University, Linz, Gruberstraße 40, 4020 Linz, Austria.

Abstract

Meiosis is initiated by a double-strand break (DSB) introduced in the DNA by a highly controlled process that is repaired by recombination. In many organisms, recombination occurs at specific and narrow regions of the genome, known as recombination hotspots, which overlap with regions enriched for DSBs. In recent years, it has been demonstrated that conversions and mutations resulting from the repair of DSBs lead to a rapid sequence evolution at recombination hotspots eroding target sites for DSBs. We still do not fully understand the effect of this erosion in the recombination activity, but evidence has shown that the binding of *trans*-acting factors like PRDM9 is affected. PRDM9 is a meiosis-specific, multi-domain protein that recognizes DNA target motifs by its zinc finger domain and directs DSBs to these target sites. Here we discuss the changes in affinity of PRDM9 to eroded recognition sequences, and explain how these changes in affinity of PRDM9 can affect recombination, leading sometimes to sterility in the context of hybrid crosses. We also present experimental data showing that DNA methylation reduces PRDM9 binding *in vitro* Finally, we discuss PRDM9-independent hotspots, posing the question how these hotspots evolve and change with sequence erosion.This article is part of the themed issue 'Evolutionary causes and consequences of recombination rate variation in sexual organisms'.

abstract

KEYWORDS: PRDM9; binding motifs; double-strand breaks; recombination hotspots

Full text links

Royal Society Publishing open access to full text

Save items

Add to Favorites

Similar articles

PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplo [PLoS Genet. 2015]

The long zinc finger domain of PRDM9 forms a highly stable and long-l [Chromosome Res. 2017]

The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 ar [PLoS Genet. 2016]

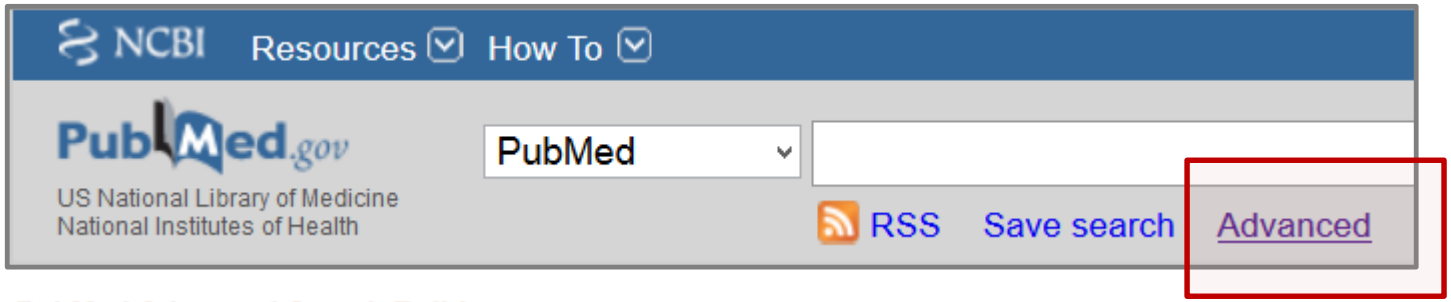
Review The case of the fickle fingers: how the PRDM9 zinc finger protein spec [PLoS Biol. 2011]

Review The spatial regulation of meiotic recombination hotspots: are [Exp Cell Res. 2012]

See reviews...

See all...

PubMed – Advanced search

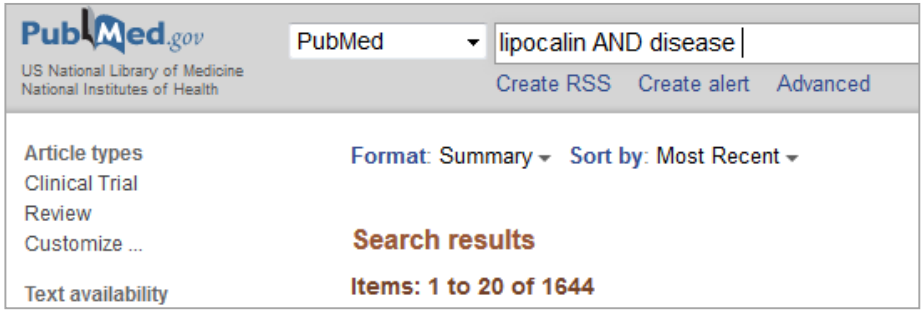


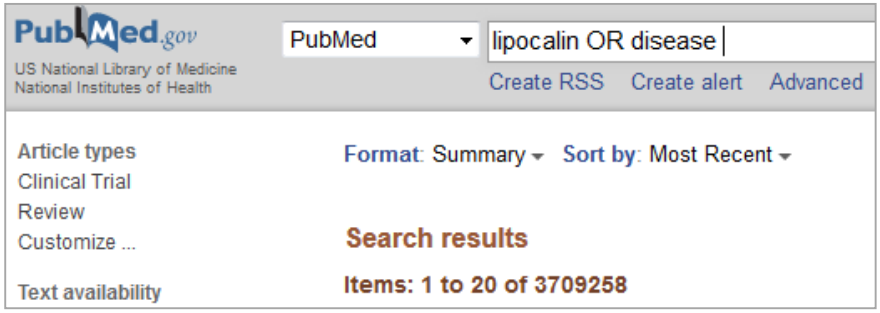
- Make your search more specific (use the pre-defined fields: **Author, Date,…**)

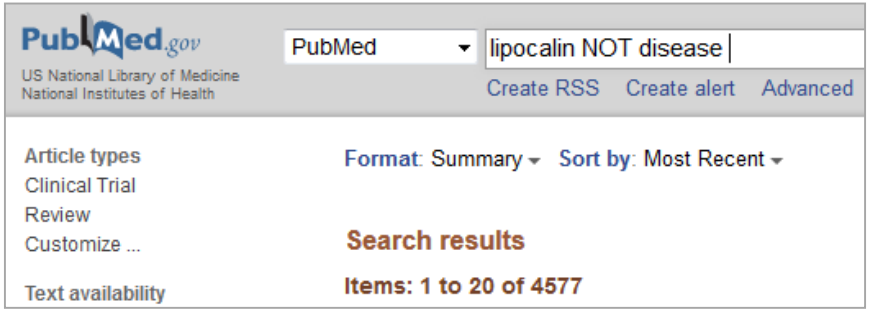
- **AND – OR – NOT** operators help you to limit your search

PubMed - Boolean search

- Boolean searches are carried out using terms like **AND, OR, NOT**. These "operators" specify what words the results of your search **should** or **should not** contain.



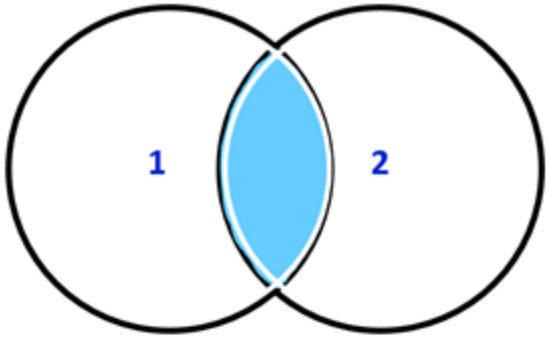




PubMed - Boolean search

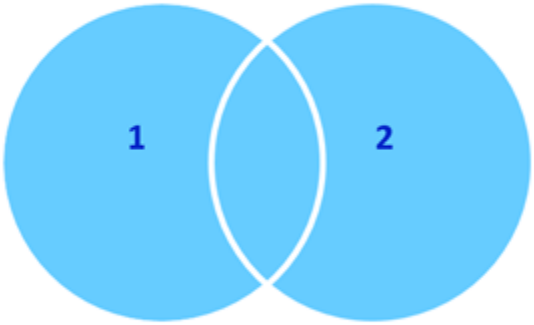
JOHANNES KEPLER UNIVERSITÄT LINZ

1 **AND** 2



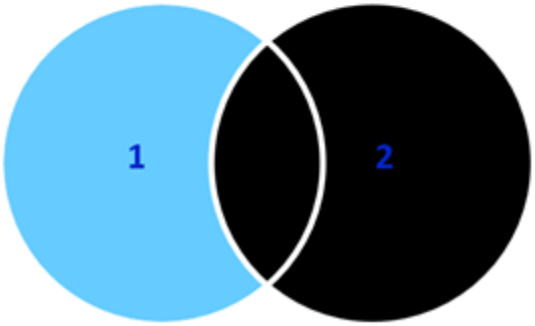
lipocalin **AND** disease
(1644 results)

1 **OR** 2



lipocalin **OR** disease
(3709258 results)

1 **NOT** 2



lipocalin **NOT** disease
(4577 results)

Overview

- PubMed - Literature search
- **SNP Database**
  - **Hardy-Weinberg Equilibrium**
- Primer3Plus
  - Polymerase-Chain Reaction (PCR)
  - Primer design
- NEB Cutter

SNP, allele, genotype, haplotype

JOHANNES KEPLER UNIVERSITÄT LINZ

What is a Single Nucleotide Polymorphism (SNP or SNV)?

- **Variation** at a **single position** in a DNA sequence among individuals
- If **more than 1%** of a population does NOT carry the same nucleotide at a specific position it is classified as SNP
- SNPs can occur in **coding** and **non-coding** regions
- In coding regions, SNPs can lead to variations in the **amino acid sequence**
- If a SNP occurs within a gene, it is described as having more than one **allele**

What is an allele?

- **Variant** from a **gene / genetic locus**
- Since humans are **diploid** organisms, they have **TWO alleles** at each genetic locus (one allele inherited from each parent)
- Each pair of an allele represents the **genotype** of a specific gene /genetic locus
- A genotype can be **homozygous** or **heterozygous**

SNP, allele, genotype, haplotype

JOHANNES KEPLER UNIVERSITÄT LINZ

What is a genotype?

- Combination of the two alleles on a genetic locus
- Homozygous e.g. C/C
- Heterozygous e.g. C/T
- The process of determining a genotype is called **genotyping**

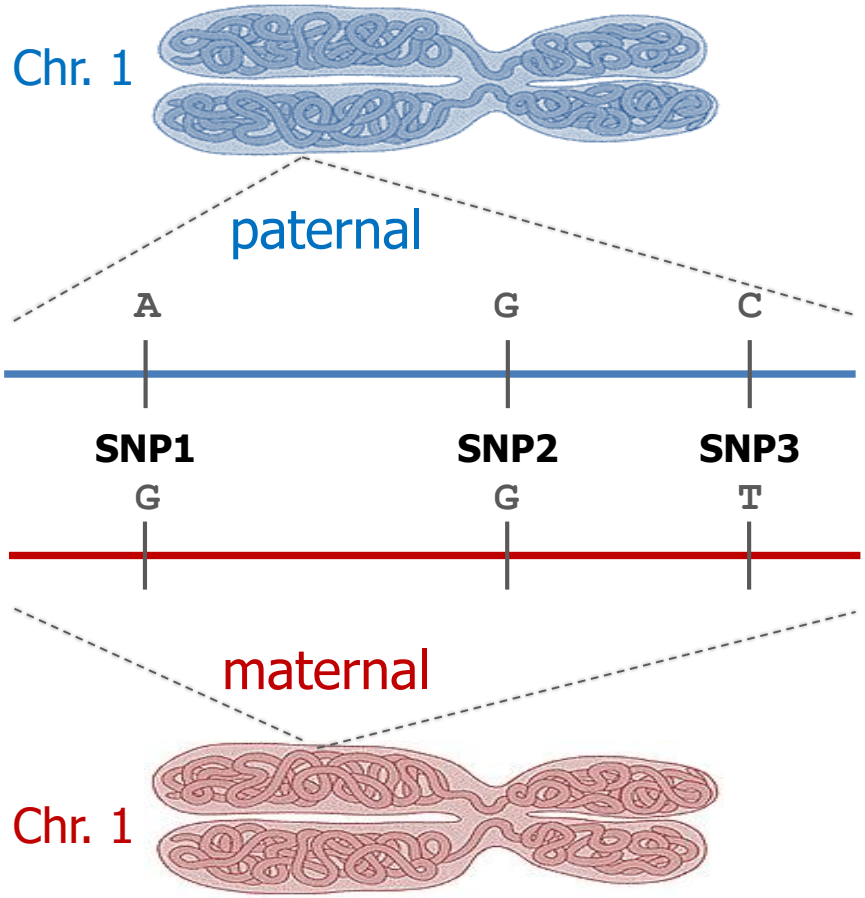
What is a haplotype?

- Combination of genetic markers (e.g. SNPs) on the **same** DNA molecule

SNP, allele, genotype, haplotype

Example

- We are looking at a genetic locus of a diploid cell having 3 SNPs
- To make it easier only one DNA strand is shown from each homolog (paternal and maternal)

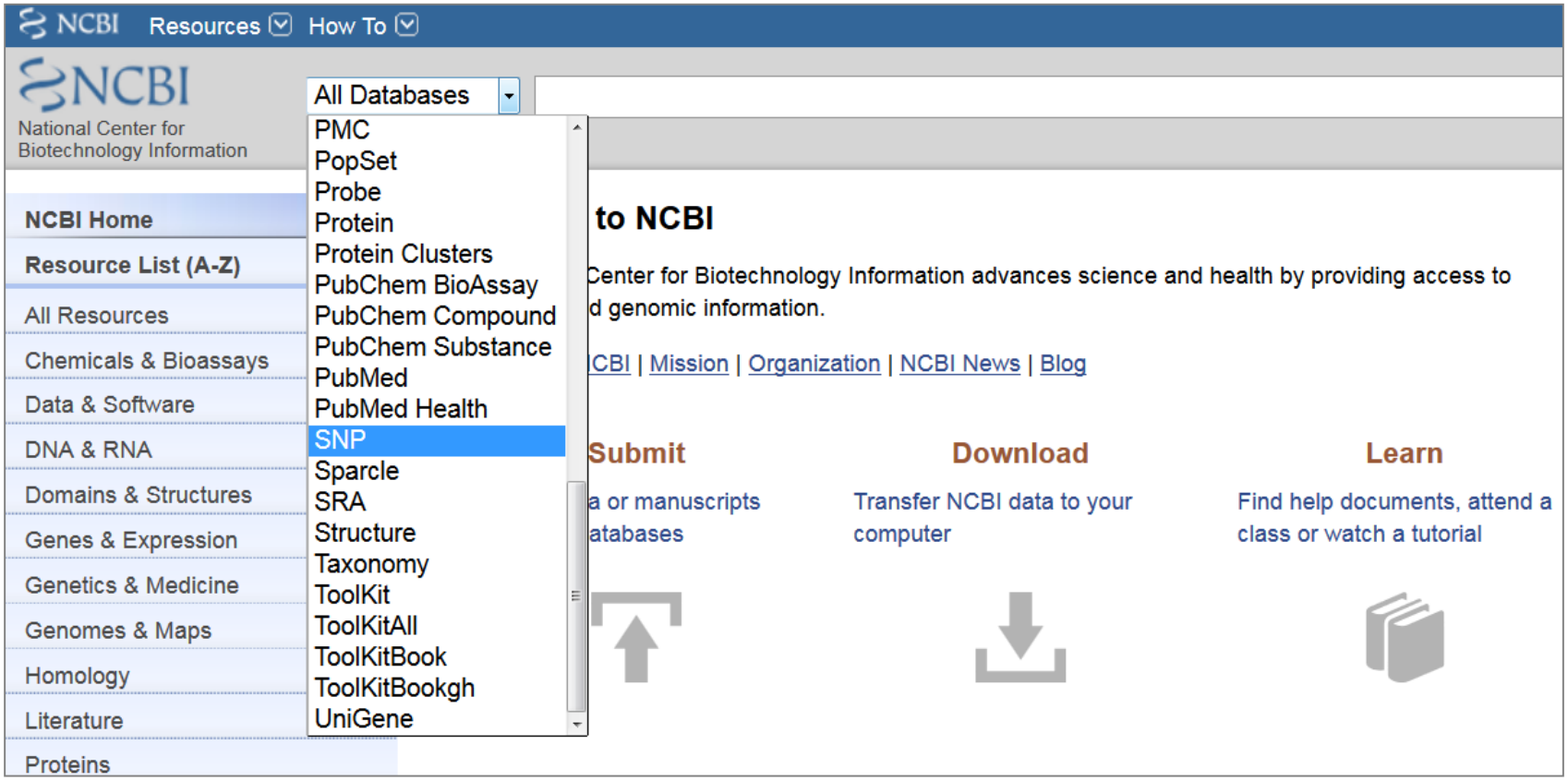


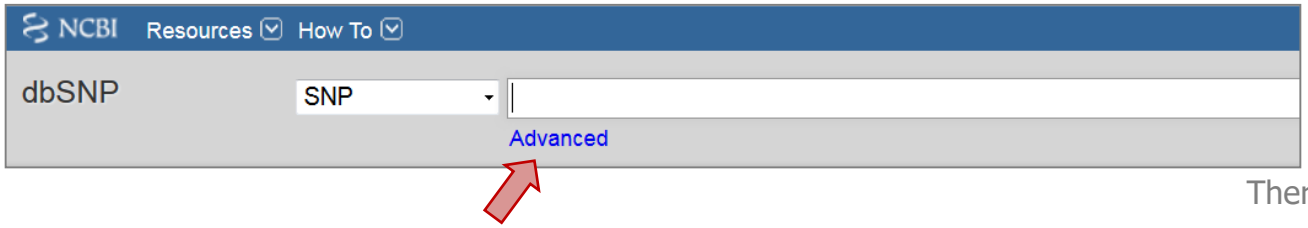
Genotypes:
- SNP1: A/G (heterozygous)
- SNP2: G/G (homozygous)
- SNP3: C/T (heterozygous)

Haplotypes:
- paternal: A-G-C
- maternal: G-G-T

NCBI – SNP database

→ Go to the **SNP** database of **NCBI** and select the **Advanced** search**.**





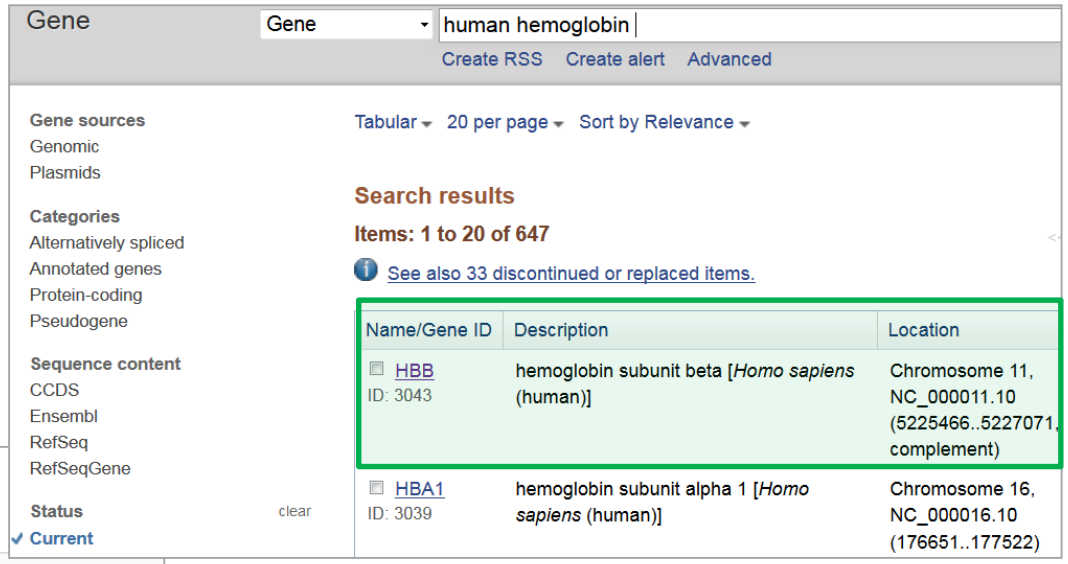
NCBI – SNP database

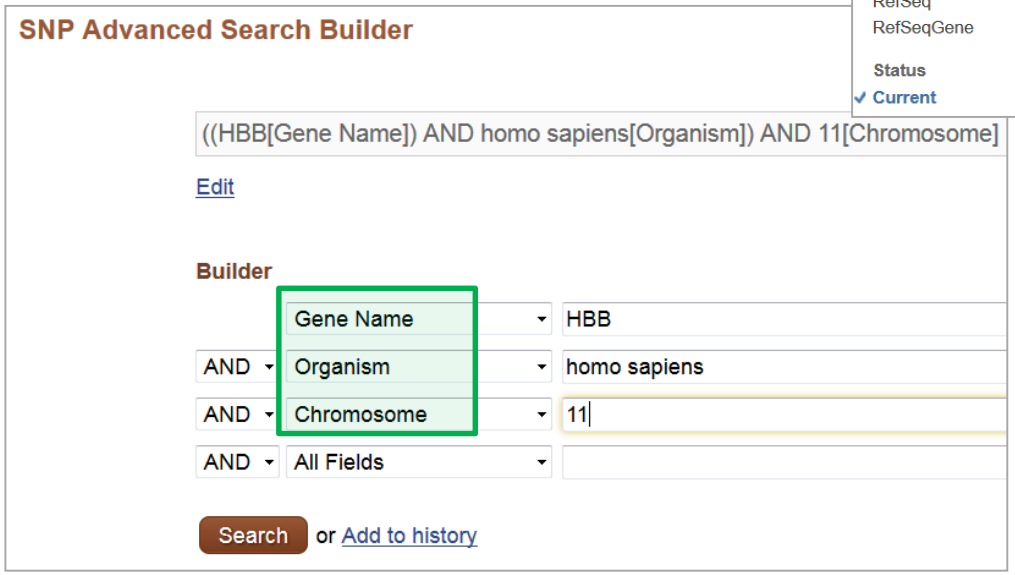
JOHANNES KEPLER UNIVERSITÄT LINZ

Find SNPs in the human hemoglobin beta gene on Chromosome 11.

1. Search for the **gene name** of human hemoglobin in the **Gene databank**

2. Search for **SNPs** using the **Gene name**, **Organism**, and **Chromosome** in the SNP databank Advanced Search Builder





NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ

dbSNP | SNP | ((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome] | Search

Create alert | Advanced

Variation Class
in del
mnp
snp

Clinical Significance
benign
likely benign
likely pathogenic
other
pathogenic
uncertain significance
untested

Annotation
Cited in PubMed
OMIM
PubMed
nucleotide
protein
structure

Function Class
3' splice site
3' utr
5' splice site
5' utr
coding synonymous
frame shift
intron
missense
nonsense

Display Settings: Summary, 20 per page, Sorted by SNP_ID | Send to:

Filters: Manage Filters

Search results

Items: 1 to 20 of 1776

<< First | < Prev | Page 1 of 89 | Next > | Last >>

rs334 *[Homo sapiens]*

1.

GACACCATGGTGCATCTGACTCCTG[A/C/G/T]GGAGAAGTCTGCCGTTACTGCCCTG

Chromosome:	11:5227002
Gene:	HBB (GeneView)
Functional Consequence:	missense
Allele Origin:	G(germline)/T(germline)/A(germline)/C(germline)
Clinical significance:	Pathogenic
Validated:	by 1000G,by cluster,by frequency
Global MAF:	A=0.0274/137
HGVS:	CM000673.2:g.5227002T>A, NC_000011.10:g.5227002T>A, NC_000011.10:g.5227002T>C, NC_000011.10:g.5227002T>G, NC_000011.9:g.5248232T>A, NC_000011.9:g.5248232T>C, NC_000011.9:g.5248232T>G, NG_000007.3:g.70614A>C, NG_000007.3:g.70614A>G, NG_000007.3:g.70614A>T, NG_042296.1:g.533T>A, NG_042296.1:g.533T>C, NG_042296.1:g.533T>G, NG_046672.1:g.4937T>A, NG_046672.1:g.4937T>C, NG_046672.1:g.4937T>G, NM_000518.4:c.20A>C, NM_000518.4:c.20A>G, NM_000518.4:c.20A>T, NP_000509.1:p.Glu7Ala, NP_000509.1:p.Glu7Gly, NP_000509.1:p.Glu7Val

PubMed | Varview

rs713040 *[Homo sapiens]*

2.

CGGCAGACTTCTCCTCAGGAGTCAG[A/C/G/T]TGCACCATGGTGTCTGTTTGAGGTT

Chromosome:	11:5227013

Find related data

Database: Select

Find items

Search details

(HBB[Gene Name] AND
[Organism]) AND 11[Ch

Search

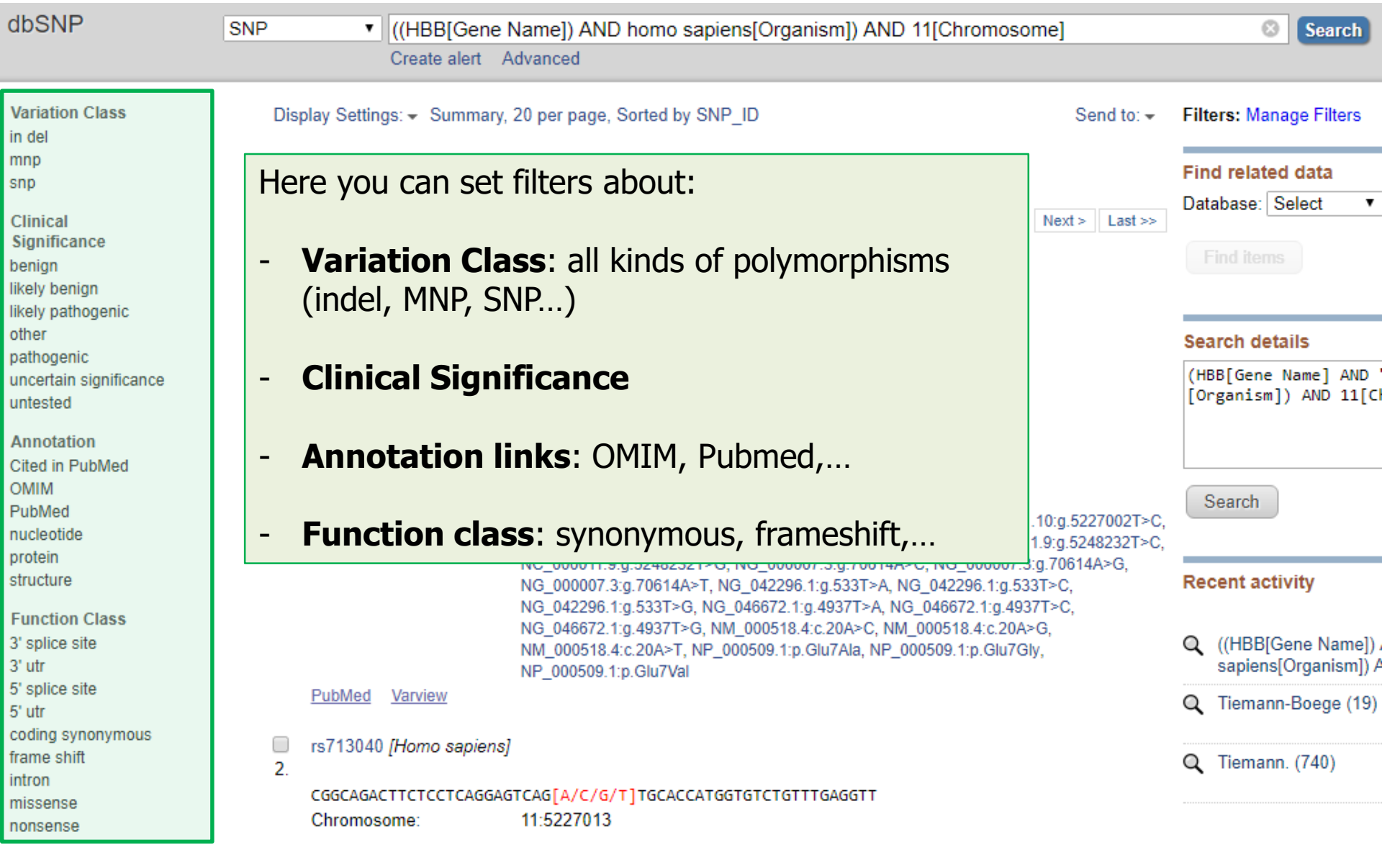
Recent activity

((HBB[Gene Name]) A
sapiens[Organism]) A

Tiemann-Boege (19)

Tiemann. (740)

NCBI – SNP database

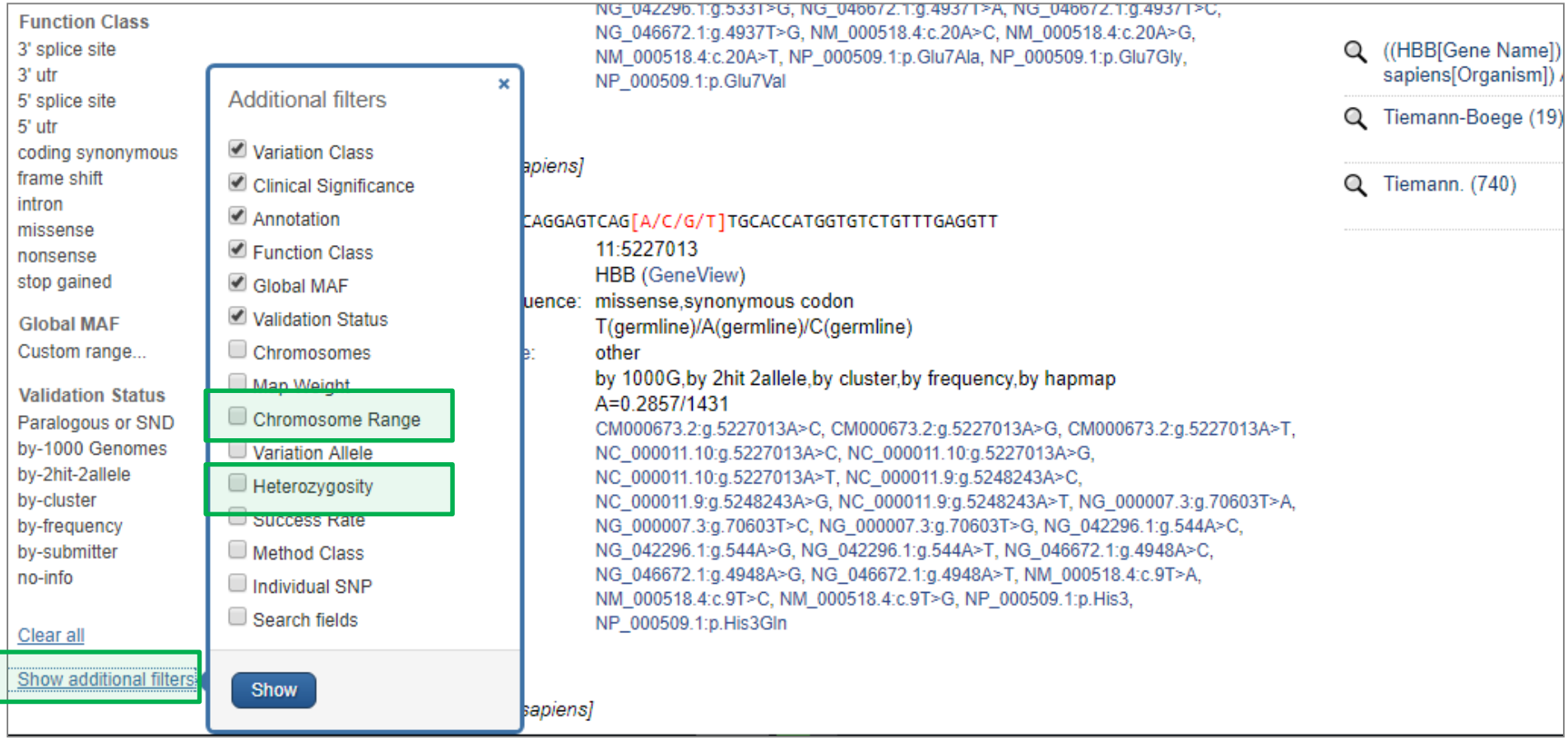


Here you can set filters about:

- **Variation Class**: all kinds of polymorphisms (indel, MNP, SNP…)
- **Clinical Significance**
- **Annotation links**: OMIM, Pubmed,…
- **Function class**: synonymous, frameshift,…

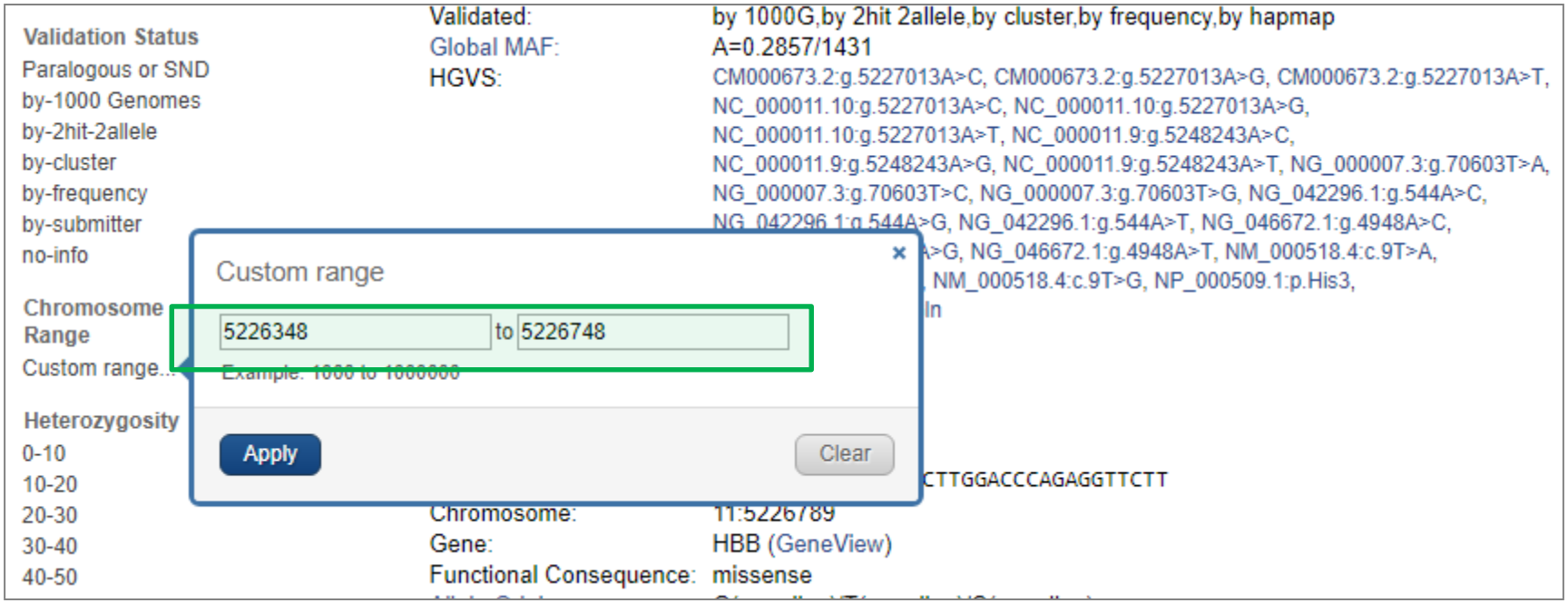
NCBI – SNP database

- When you scroll down you can select additional filters: '**Show additional filters**'.
- There you can select filters like **Chromosome Range**, **Heterozygosity**, …



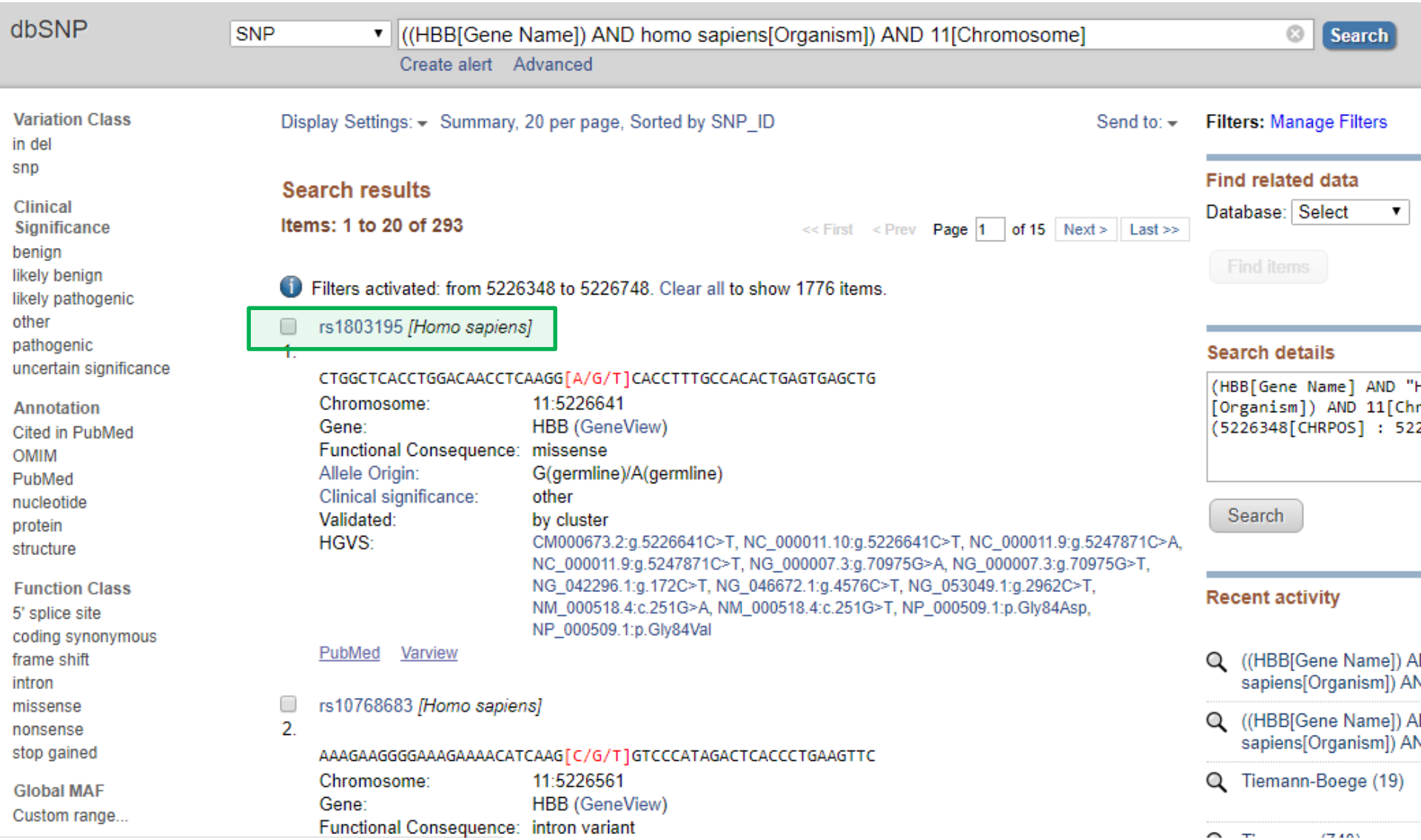
NCBI – SNP database

- Set the **Chromosome range** to **5226348 – 5226748**
- **Apply**



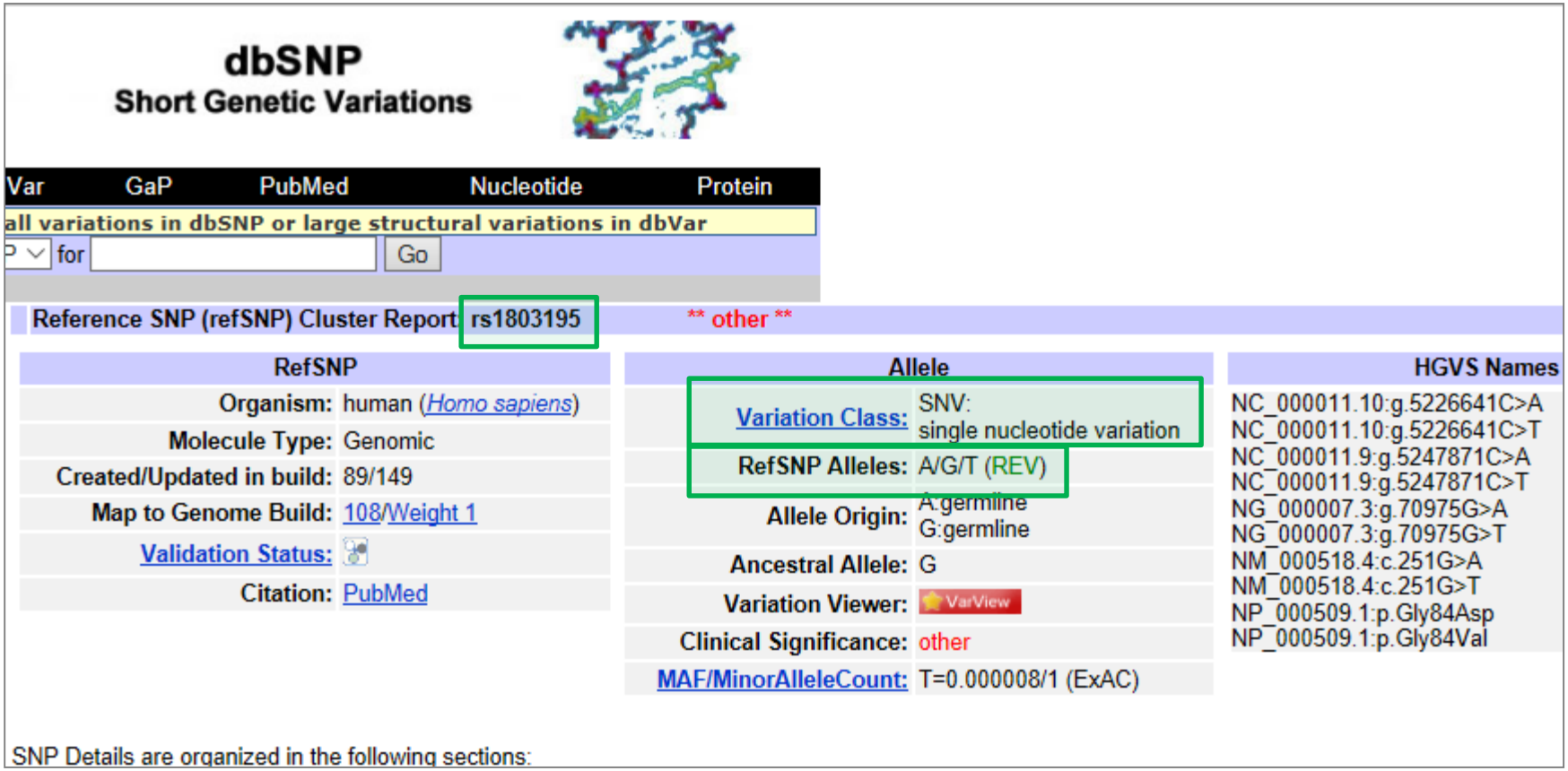
NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ



NCBI – SNP database

- SNP accession number: **rs1803195**
- General information about the polymorphism
- Variation class: SNV = single nucleotide variant
- At this position the nucleotide can be either a **A – G – T**



dbSNP
Short Genetic Variations

Var GaP PubMed Nucleotide Protein

all variations in dbSNP or large structural variations in dbVar

for Go

Reference SNP (refSNP) Cluster Report: rs1803195 ** other **

RefSNP	
Organism:	human (*Homo sapiens*)
Molecule Type:	Genomic
Created/Updated in build:	89/149
Map to Genome Build:	108/Weight 1
Validation Status:	
Citation:	PubMed

Allele	
Variation Class:	SNV: single nucleotide variation
RefSNP Alleles:	A/G/T (REV)
Allele Origin:	A:germline G:germline
Ancestral Allele:	G
Variation Viewer:	VarView
Clinical Significance:	other
MAF/MinorAlleleCount:	T=0.000008/1 (ExAC)

HGVS Names
NC_000011.10:g.5226641C>A
NC_000011.10:g.5226641C>T
NC_000011.9:g.5247871C>A
NC_000011.9:g.5247871C>T
NG_000007.3:g.70975G>A
NG_000007.3:g.70975G>T
NM_000518.4:c.251G>A
NM_000518.4:c.251G>T
NP_000509.1:p.Gly84Asp
NP_000509.1:p.Gly84Val

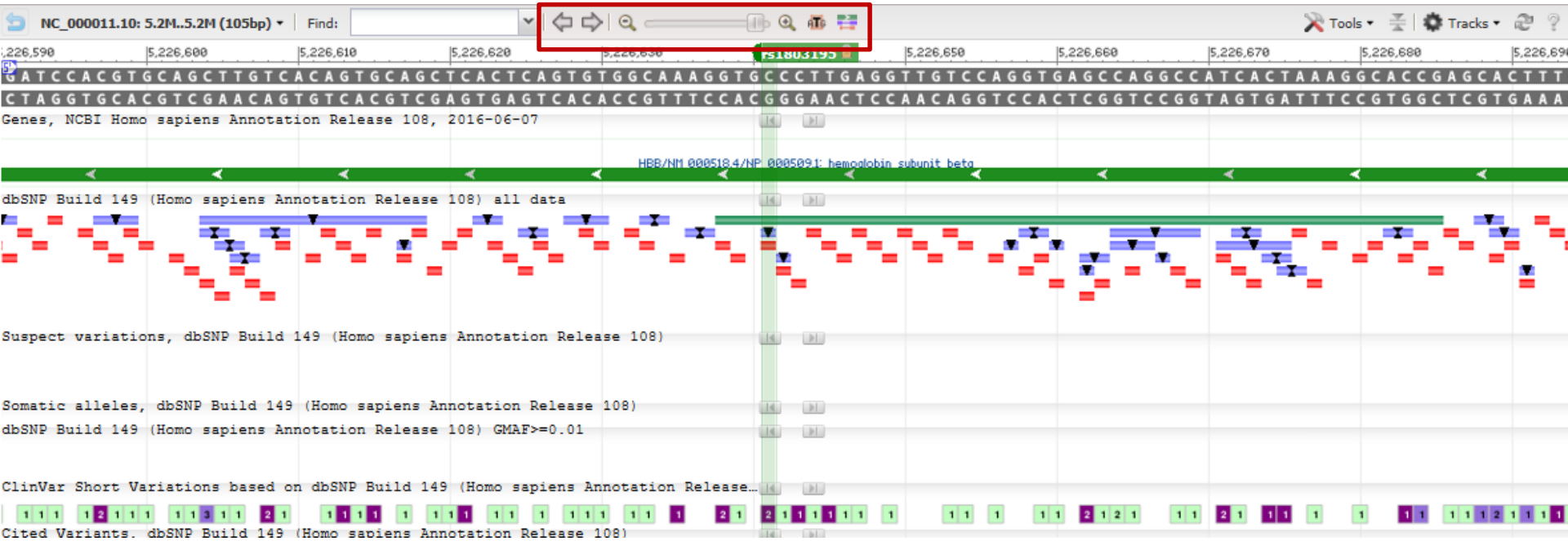
SNP Details are organized in the following sections:

NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ

Scroll down

- Map of the genomic region containing the SNP **rs1803195**
- You can navigate along the map



NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ

Scroll down

- FASTA sequence containing the SNP and genomic region up- and downstream

Fasta sequence (Legend)

```
>gnl|dbSNP|rs1803195|allelePos=251|totalLen=501|taxid=9606|snpclass=1|alleles='A/G/T'|mol=Genomic|build=144

ATAGAAACTG GGCATGTGGA GACAGAGAAG ACTCTTGGGT TTCTGATAGG CACTGACTCT
CTCTGCCTAT TGGTCTATTT TCCCACCCTT AGGCTGCTGG TGGTCTACCC TTGGACCCAG
AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG
GTGAAGGCTC ATGGCAAGAA AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC
AACCTCAAGG
D
CACCTTTGCC ACACTGAGTG AGCTGCACTG TGACAAGCTG CACGTGGATC CTGAGAACTT
CAGGGTGAGT CTATGGGACG CTTGATGTTT TCTTTCCCCT TCTTTTCTAT GGTTAAGTTC
ATGTCATAGG AAGGGGATAA GTAACAGGGT ACAGTTTAGA ATGGGAAACA GACGAATGAT
TGCATCAGTG TGGAAGTCTC AGGATCGTTT TAGTTTCTTT TATTTGCTGT TCATAACAAT
TGTTTTCTTT
```

IUPAC code – bases and mixed bases of DNA/RNA

Source: http://www.bioinformatics.org/sms/iupac.html

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ

Scroll down

- FASTA sequence containing the SNP and genomic region up- and downstream

Fasta sequence (Legend)

```
>gnl|dbSNP|rs1803195|allelePos=251|totalLen=501|taxid=9606|snpclass=1|alleles='A/G/T'|mol=Genomic|build=144

ATAGAAACTG GGCATGTGGA GACAGAGAAG ACTCTTGGGT TTCTGATAGG CACTGACTCT
CTCTGCCTAT TGGTCTATTT TCCCACCCTT AGGCTGCTGG TGGTCTACCC TTGGACCCAG
AGGTTCTTTG AGTCCTTTGG GGATCTGTCC ACTCCTGATG CTGTTATGGG CAACCCTAAG
GTGAAGGCTC ATGGCAAGAA AGTGCTCGGT GCCTTTAGTG ATGGCCTGGC TCACCTGGAC
AACCTCAAGG
D
CACCTTTGCC ACACTGAGTG AGCTGCACTG TGACAAGCTG CACGTGGATC CTGAGAACTT
CAGGGTGAGT CTATGGGACG CTTGATGTTT TCTTTCCCCT TCTTTTCTAT GGTTAAGTTC
ATGTCATAGG AAGGGGATAA GTAACAGGGT ACAGTTTAGA ATGGGAAACA GACGAATGAT
TGCATCAGTG TGGAAGTCTC AGGATCGTTT TAGTTTCTTT TATTTGCTGT TCATAACAAT
TGTTTTCTTT
```

- You can estimate the genotype frequencies for specific polymorphisms using the **Hardy-Weinberg Model**

Hardy-Weinberg Equilibrium

JOHANNES KEPLER UNIVERSITÄT LINZ

- Model to deduce **theoretical predictions** of **genotype frequencies**

- The HW-Model is mainly used in Population Genetics

- We have to assume:
  - large, diploid population
  - mating is random
  - there are NO evolutionary processes going on

Hardy-Weinberg Equilibrium

JOHANNES KEPLER UNIVERSITÄT LINZ

Let's make the following assumptions:

→ On a genomic locus in humans we have the alleles **A** and **B**

→ There are 3 possible genotypes: **AA**, **BB** (*homozygous*), **AB** (*heterozygous*)

→ We know the allele frequencies which are given by **p** and **q**

frequency of **A = p** (0.62)
frequency of **B = q** (0.38)
} sum = 1

→ How can we calculate the genotype frequencies?

Mathematical relation between allele frequencies and genotype frequencies:

AA: p^2	*homozygote*	$0.62^2 = 0.3844$	} sum = 1
BB: q^2	*homozygote*	$0.38^2 = 0.1444$	
AB: 2pq	*heterozygote*	$2*0.62*0.38 = 0.4712$	

p + q = 1

$(p + q)^2 = \mathbf{p^2 + 2pq + q^2} = 1$ (binomial formula)

Hardy-Weinberg Equilibrium

What if there are more than two possible alleles on one locus?

→ This often happens for microsatellites (short tandem repeats = STRs)

(GAC)3	0.32
(GAC)4	0.04
(GAC)5	0.23
(GAC)6	0.41

sum = 1

How to calculate genotype frequencies in this case?

→ Use the same mathematical relationship!!

Homozygote: (GAC)3/(GAC)3 = 0.32^2 = <u>0.1024</u>

Heterozygote: (GAC)4/(GAC)6 = 2*0.04*0.41 = <u>0.0328</u>

NCBI – SNP database

How can we find allele frequencies in the SNP database?

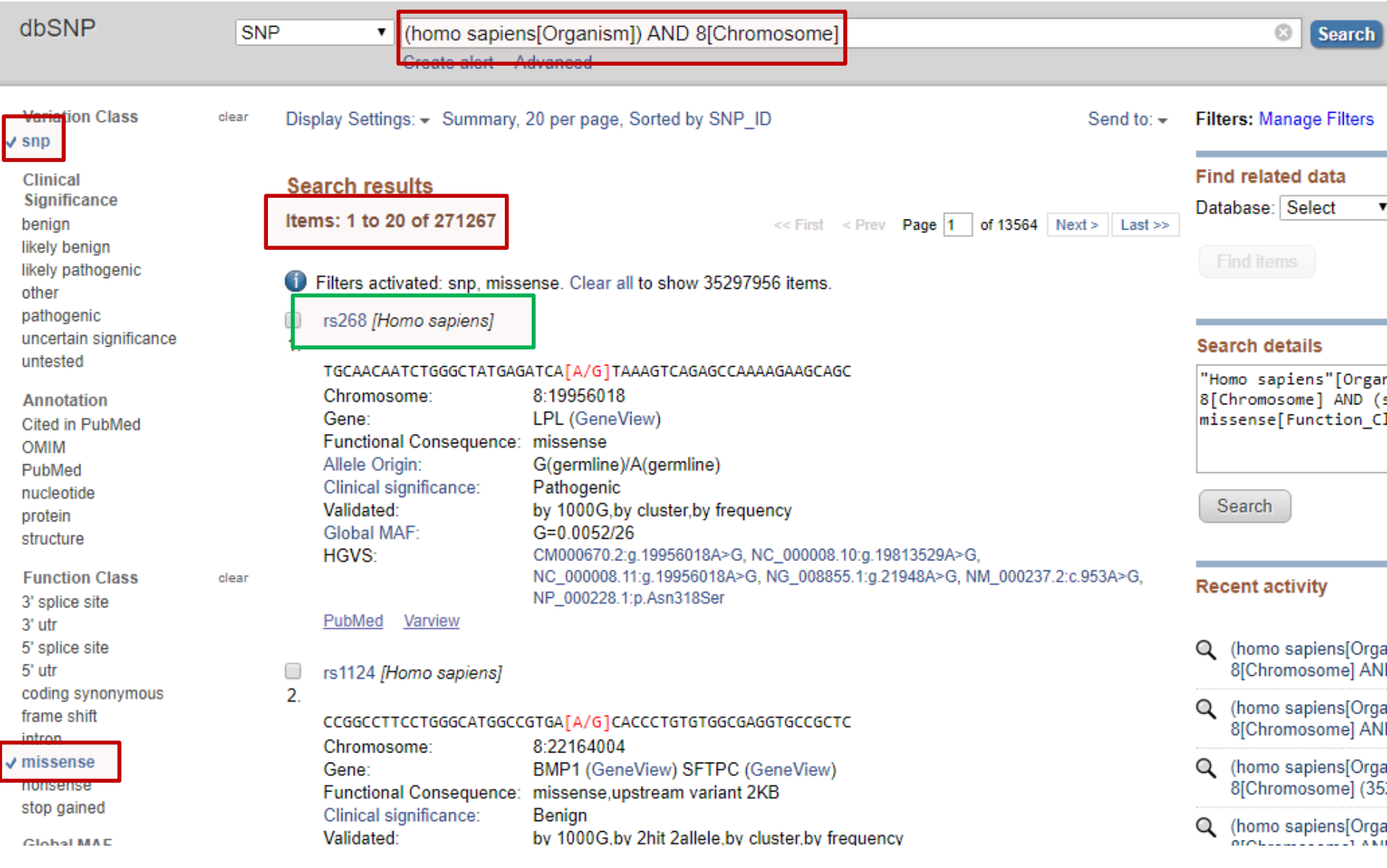
→ Go to **NCBI – SNP database**

→ Search for all **SNPs** in the **human** organism on **Chromosome 8**

→ Select for **missense** SNPs

NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ



NCBI – SNP database

JOHANNES KEPLER UNIVERSITÄT LINZ

Reference SNP (refSNP) Cluster Report: rs268 ** With Pathogenic allele **

RefSNP	
Organism:	human (*Homo sapiens*)
Molecule Type:	Genomic
Created/Updated in build:	36/147
Map to Genome Build:	107/Weight 1
Validation Status:	
Citation:	PubMed
Association:	NHGRI GWAS PheGenI

Allele	
Variation Class:	SNV: single nucleotide variation
RefSNP Alleles:	A/G (FWD)
Allele Origin:	A:germline G:germline
Ancestral Allele:	A
Variation Viewer:	VarView
Clinical Significance:	With Pathogenic allele [ClinVar]
MAF/MinorAlleleCount:	G=0.0134/1622 (ExAC) G=0.0052/26 (1000 Genomes) G=0.0134/174 (GO-ESP)

HGVS Names

NC_000008.10:g.19813529A>G
NC_000008.11:g.19956018A>G
NG_008855.1:g.21948A>G
NM_000237.2:c.953A>G
NP_000228.1:p.Asn318Ser

Scroll down to **Population Diversity**

Allele frequencies

Population Diversity (Alleles in RefSNP orientation) . See additional population frequency from 1000Genome [here]

	Sample Ascertainment				Genotype Detail				Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	A/A	A/G	G/G	HWP	A	G
ss1328915147	EAS		1008	AF					1.00000000	
	EUR		1006	AF					0.98610002	0.01390000
	AFR		1322	AF					0.99919999	0.00080000
	AMR		694	AF					0.98850000	0.01150000
	SAS		978	AF					0.99690002	0.00310000

Populations

NCBI – SNP database

- What is the heterozygous genotype frequency in Europeans?

A : 0.986 (p)

G : 0.014 (q)

AG = 2pq = 2*0.986*0.014 = 0.027608

Population Diversity (Alleles in RefSNP orientation) . See additional population frequency from 1000Genome [here]

	Sample Ascertainment				Genotype Detail				Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	A/A	A/G	G/G	HWP	A	G
ss1328915147	EAS		1008	AF					1.00000000	
	EUR		1006	AF					0.98610002	0.01390000
	AFR		1322	AF					0.99919999	0.00080000
	AMR		694	AF					0.98850000	0.01150000
	SAS		978	AF					0.99690002	0.00310000

Population codes

Population Code	Population Description	Super Population Code
CHB	Han Chinese in Bejing, China	EAS
JPT	Japanese in Tokyo, Japan	EAS
CHS	Southern Han Chinese	EAS
CDX	Chinese Dai in Xishuangbanna, China	EAS
KHV	Kinh in Ho Chi Minh City, Vietnam	EAS
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	EUR
TSI	Toscani in Italia	EUR
FIN	Finnish in Finland	EUR
GBR	British in England and Scotland	EUR
IBS	Iberian Population in Spain	EUR
YRI	Yoruba in Ibadan, Nigeria	AFR
LWK	Luhya in Webuye, Kenya	AFR
GWD	Gambian in Western Divisions in the Gambia	AFR
MSL	Mende in Sierra Leone	AFR
ESN	Esan in Nigeria	AFR
ASW	Americans of African Ancestry in SW USA	AFR
ACB	African Caribbeans in Barbados	AFR
MXL	Mexican Ancestry from Los Angeles USA	AMR
PUR	Puerto Ricans from Puerto Rico	AMR
CLM	Colombians from Medellin, Colombia	AMR
PEL	Peruvians from Lima, Peru	AMR
GIH	Gujarati Indian from Houston, Texas	SAS
PJL	Punjabi from Lahore, Pakistan	SAS
BEB	Bengali from Bangladesh	SAS
STU	Sri Lankan Tamil from the UK	SAS
ITU	Indian Telugu from the UK	SAS

These populations have been divided into 5 super populations

- **AFR**, African
- **AMR**, Ad Mixed American
- **EAS**, East Asian
- **EUR**, European
- **SAS**, South Asian

Overview

- PubMed - Literature search
- SNP Database
  - Hardy-Weinberg Equilibrium
- **Primer3Plus**
  - **Polymerase-Chain Reaction (PCR)**
  - **Primer design**
- NEB Cutter

Polymerase chain reaction (PCR)

JOHANNES KEPLER UNIVERSITÄT LINZ

What is the aim of a polymerase chain reaction?

=Amplification of a certain DNA sequence

Which components are needed?

Template DNA

Primer 1 = forward primer

Primer 2 = reverse primer

dNTPs = nucleotides

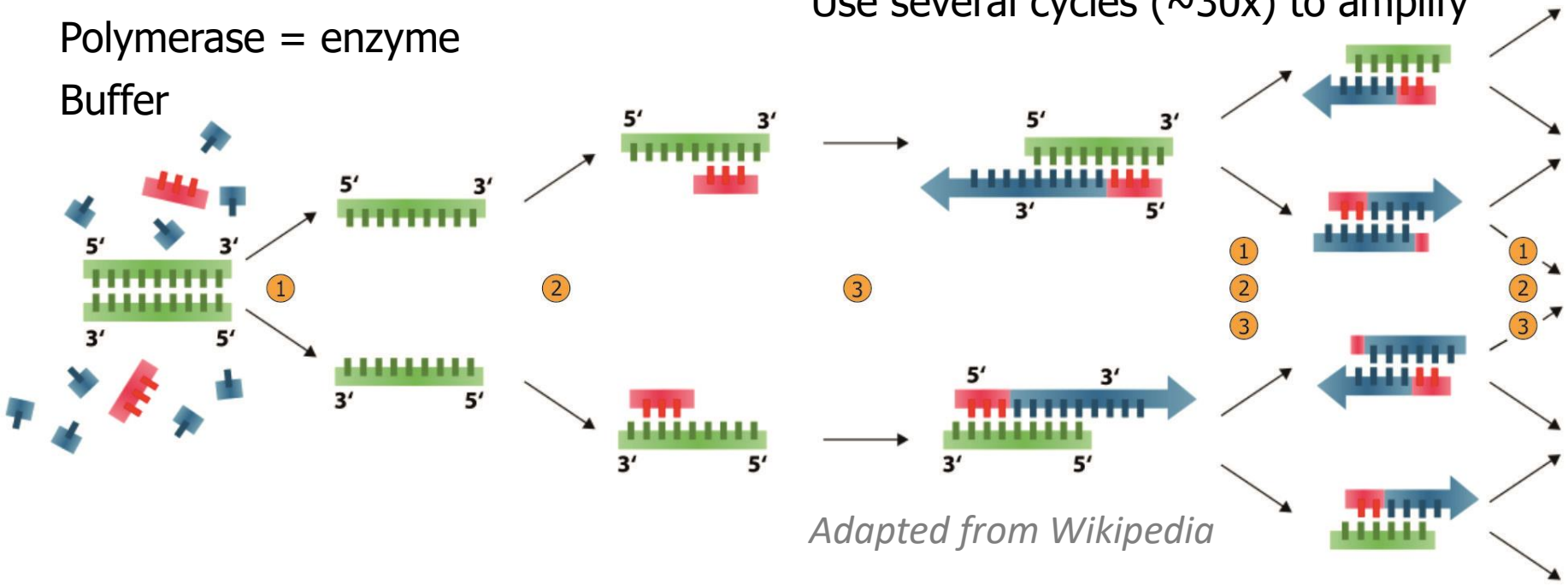
Polymerase = enzyme

Buffer

How does a PCR work?

1. Denaturation ~94°C
2. Annealing ~60°C
3. Elongation ~70°C

Use several cycles (~30x) to amplify



Adapted from Wikipedia

Polymerase chain reaction (PCR)

Template DNA (double-stranded DNA = dsDNA)

(e.g. genomic DNA extracted from blood or saliva)

```
5′-ATCGGGGCCCATATAAATTTGCCCGGGTTTAAAAGATCCATGGATCCCATTTAAGC-3′
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
3′-TAGCCCCGGGTATATTTAAACGGGCCCAAATTTTCTAGGTACCTAGGGTAAATTCG-5′
```

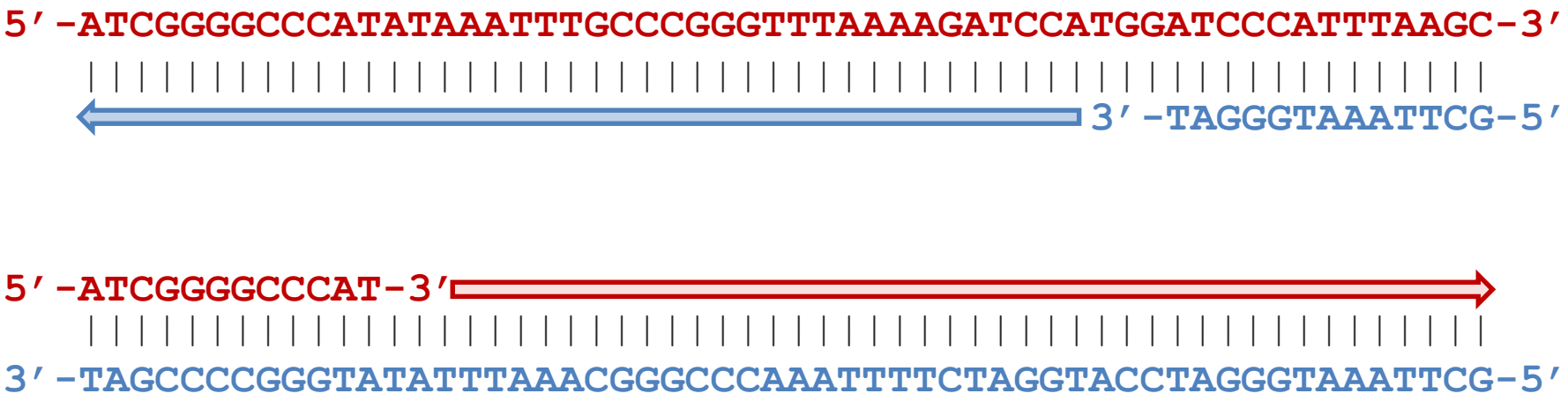
- 2 DNA strands

  Plus strand – Watson strand – coding strand

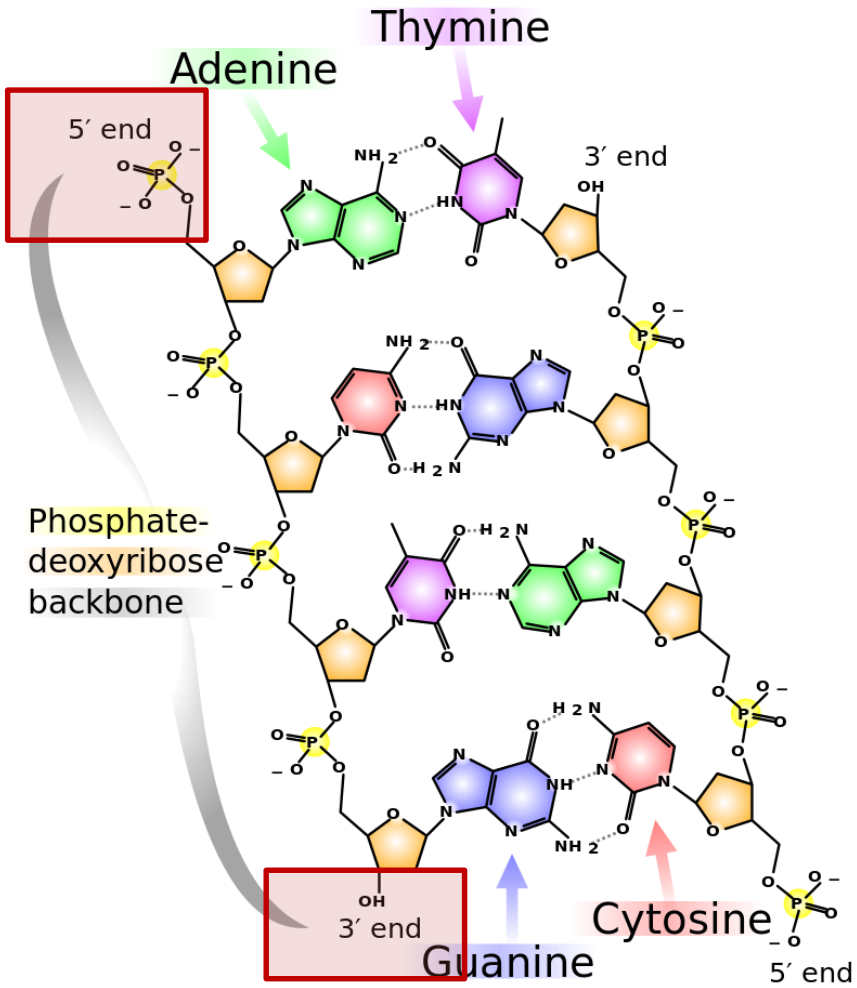
  Minus strand – Crick strand – non-coding strand

Polymerase chain reaction (PCR)

1. Denaturation: dsDNA → ssDNA ~94°C
2. Annealing: Primer binding ~60°C (depends on melting temp. of primer)
3. Elongation: in 5‘-3‘ direction ~70°C (optimal temp. for polymerase)



Polymerase chain reaction (PCR)



- Elongation:

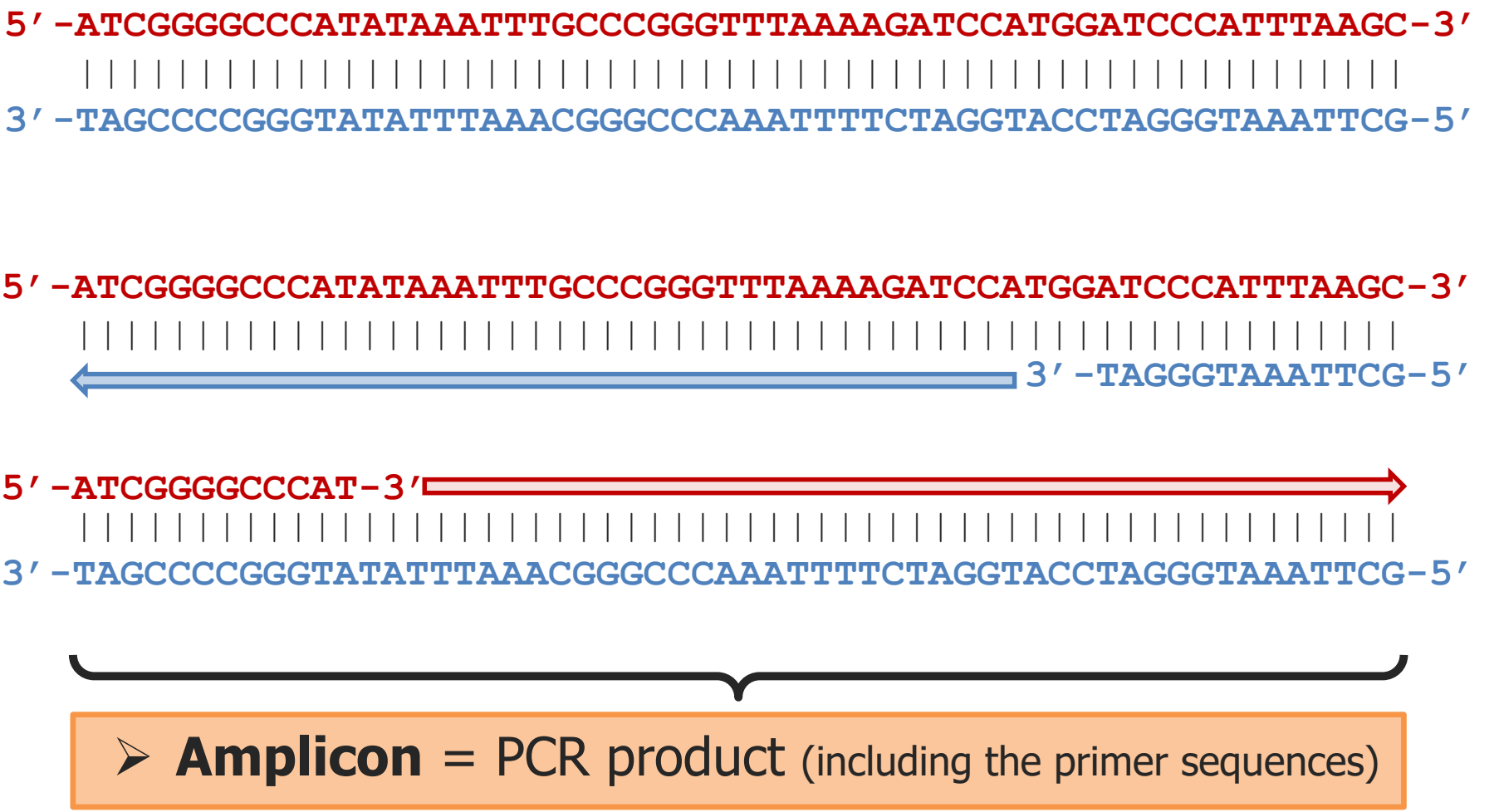
  The polymerase extends the **3'-end of the primer/DNA sequence** by adding nucleotides to the free OH-group and creates phosphodiester bonds

  →the synthesis therefore goes **in 5'-3' direction**.

- **5' end**: phosphate group
- **3' end**: OH group

http://en.wikipedia.org/wiki/DNA#mediaviewer/File:DNA_chemical_structure.svg

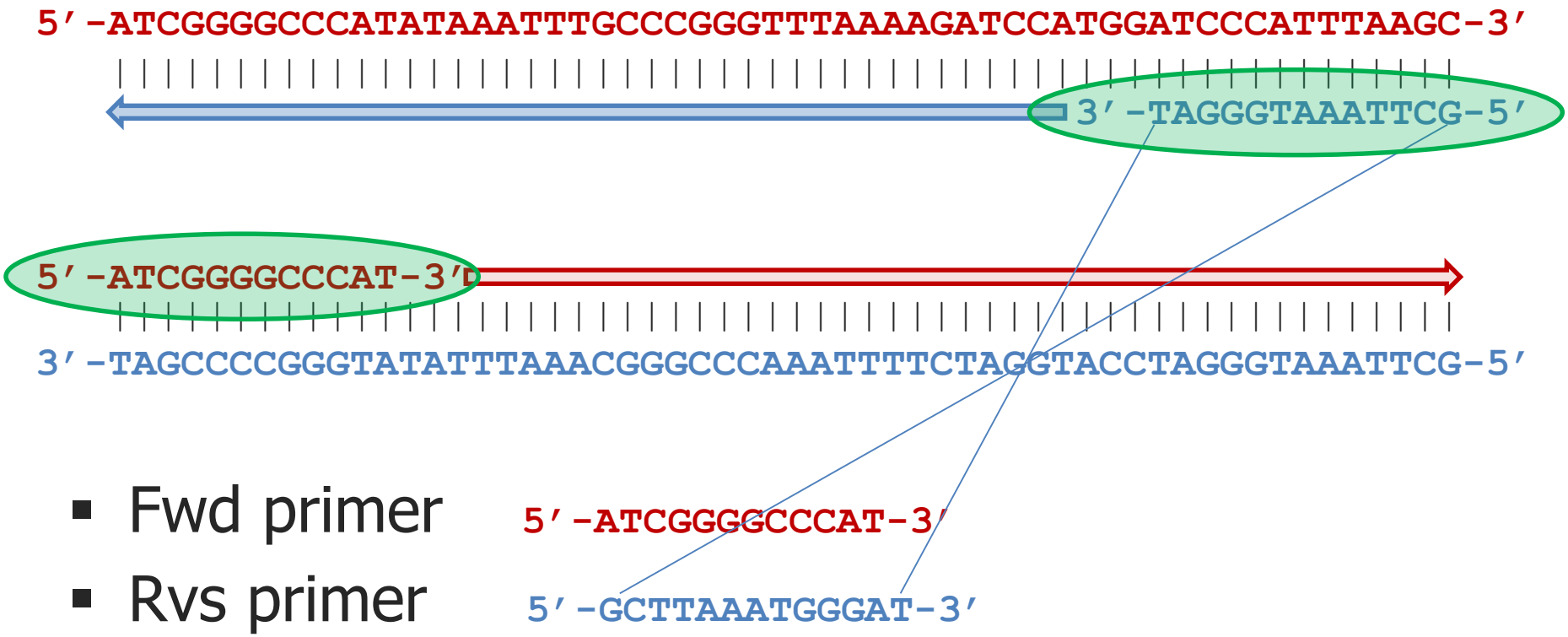
Polymerase chain reaction (PCR)



➢ **Amplicon** = PCR product (including the primer sequences)

Polymerase chain reaction (PCR)

A DNA sequence is always written in 5‘-3‘ direction !



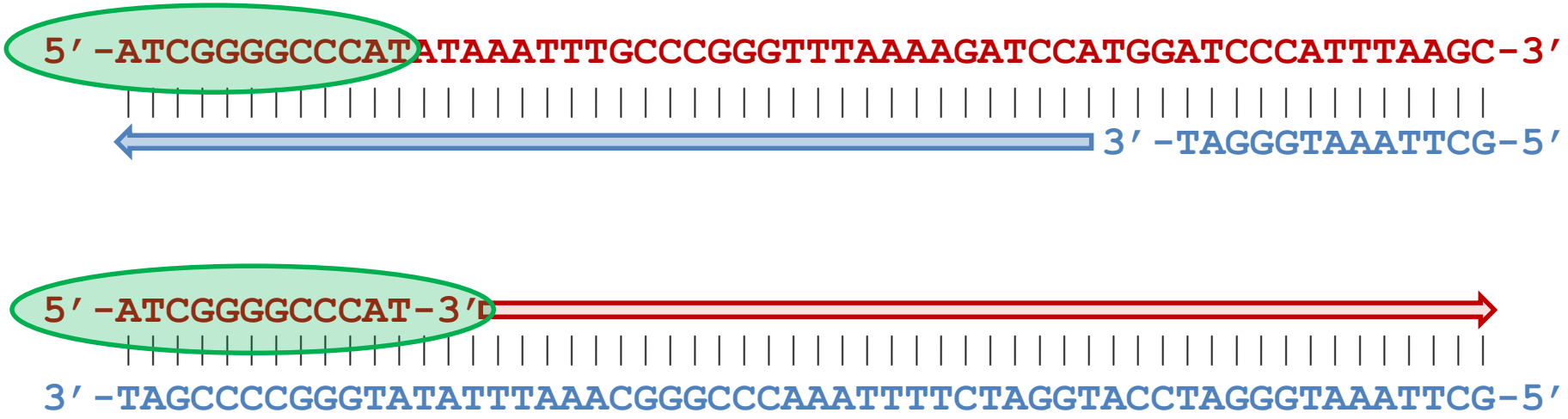
- Fwd primer 5′-ATCGGGGCCCAT-3′
- Rvs primer 5′-GCTTAAATGGGAT-3′

Polymerase chain reaction (PCR)

The forward primer has the

same sequence

than plus strand.



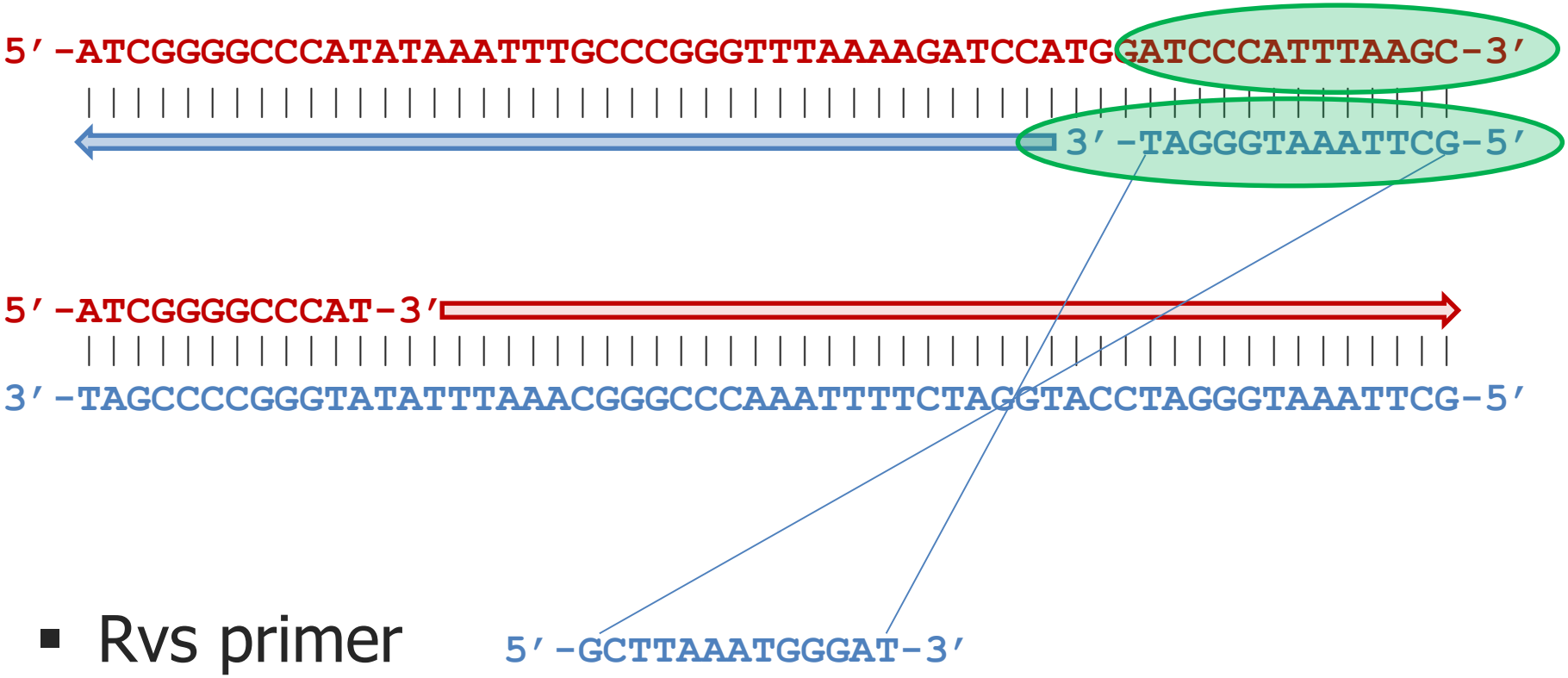
- Fwd primer `5′-ATCGGGGCCCAT-3′`

Polymerase chain reaction (PCR)

The reverse primer is the

reverse complement

to the sequence on the plus strand.



- Rvs primer 5′-GCTTAAATGGGAT-3′

Primer3Plus

JOHANNES KEPLER UNIVERSITÄT LINZ

- Before you can perform your experiment in the laboratory you have to carefully plan all steps and purchase the materials required.

- In case of a PCR you have to **design two primers** for the amplification of a desired DNA sequence.

- **Primer3Plus** is an online tool for primer design.

Primer3Plus

- Task:

We are interested in a SNP (**rs1803195**), which we found in the SNP database, to analyze it in the lab

Therefore, we want to do a PCR to amplify a region of **50-70bp** surrounding this SNP

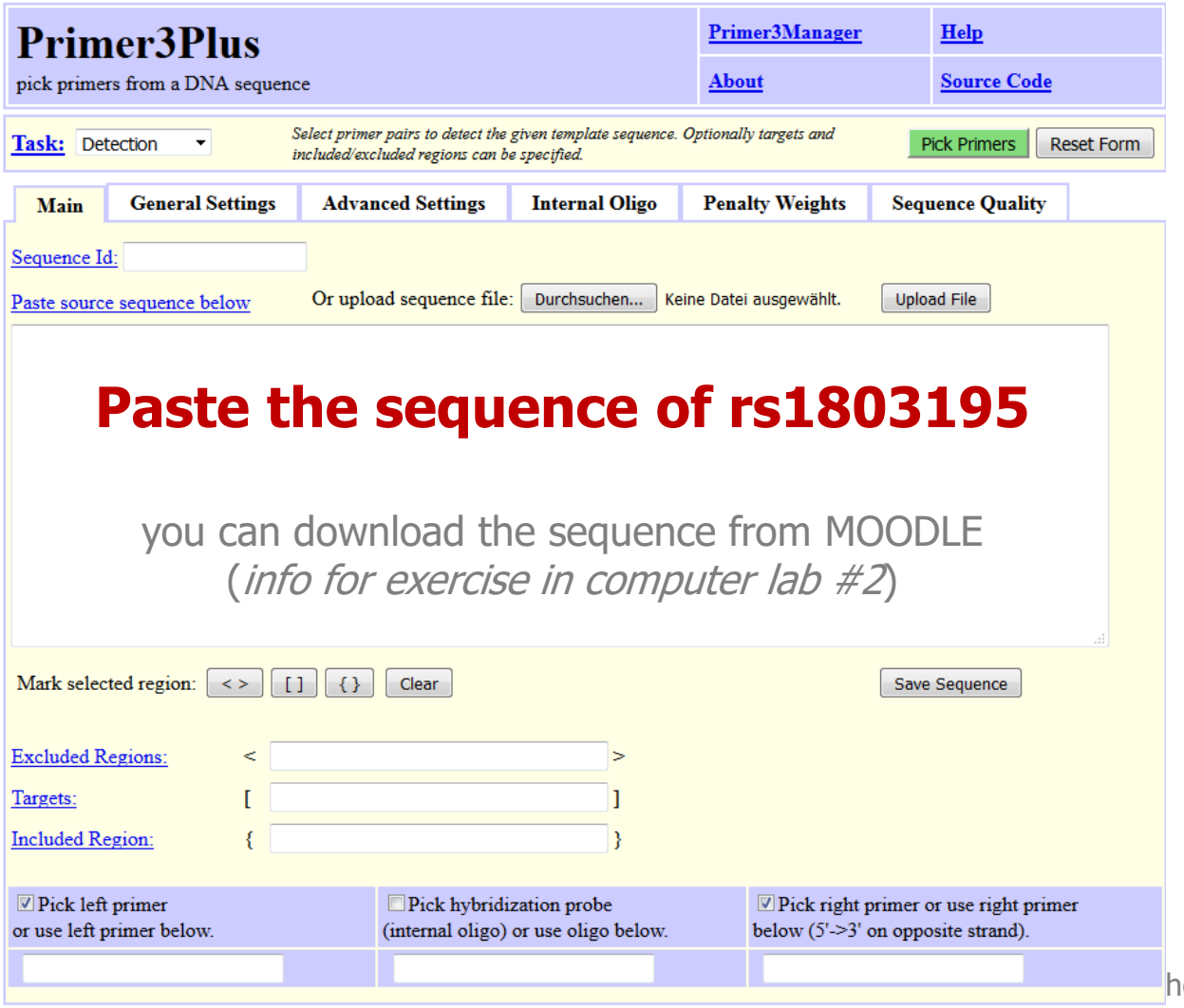
As template we will use e.g. genomic DNA that we extracted from blood.

To carry out the PCR we have to **design primers**.

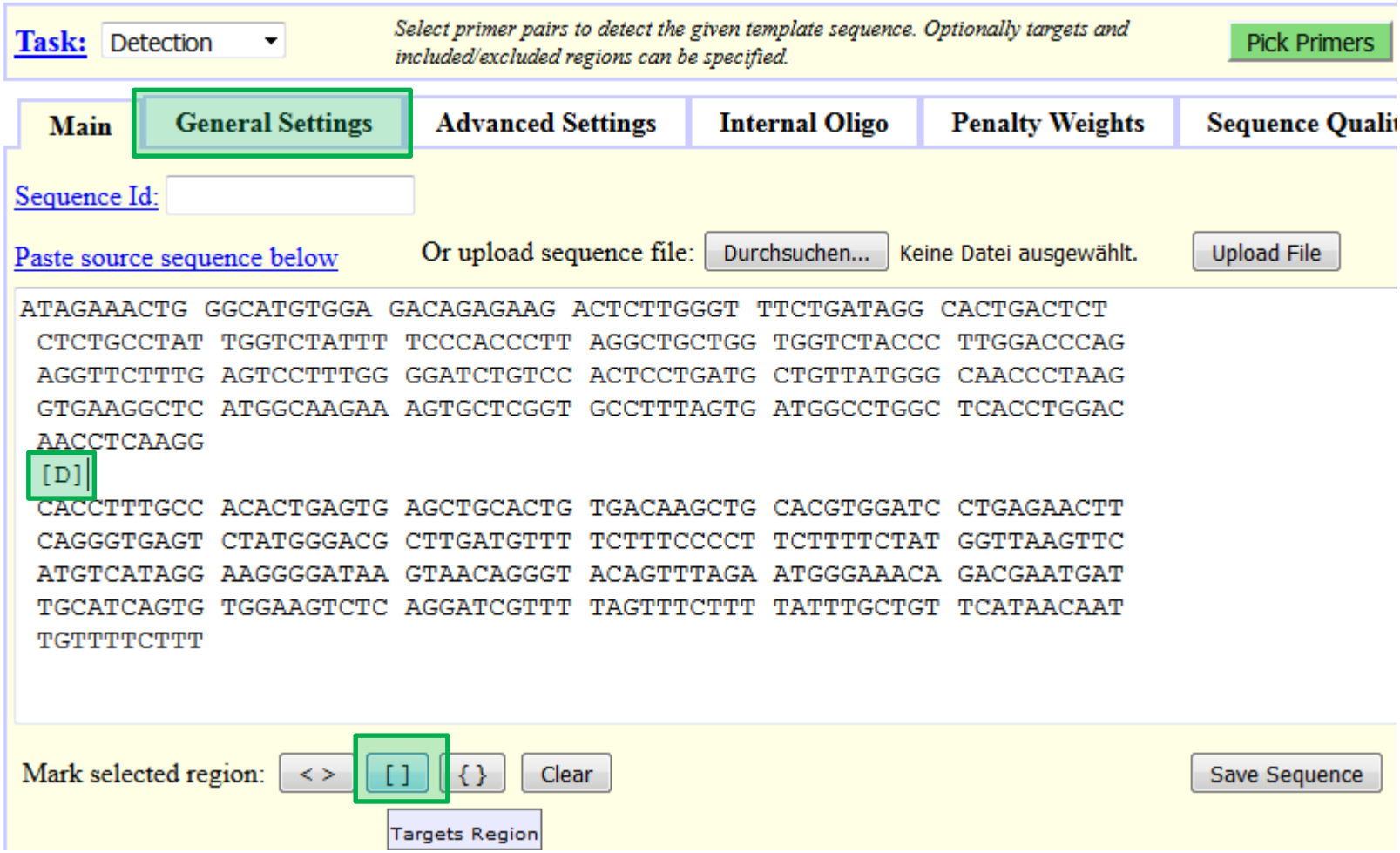
→ **Go to Primer3Plus**

Primer3Plus

- http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
  (search for Primer3Plus in google)

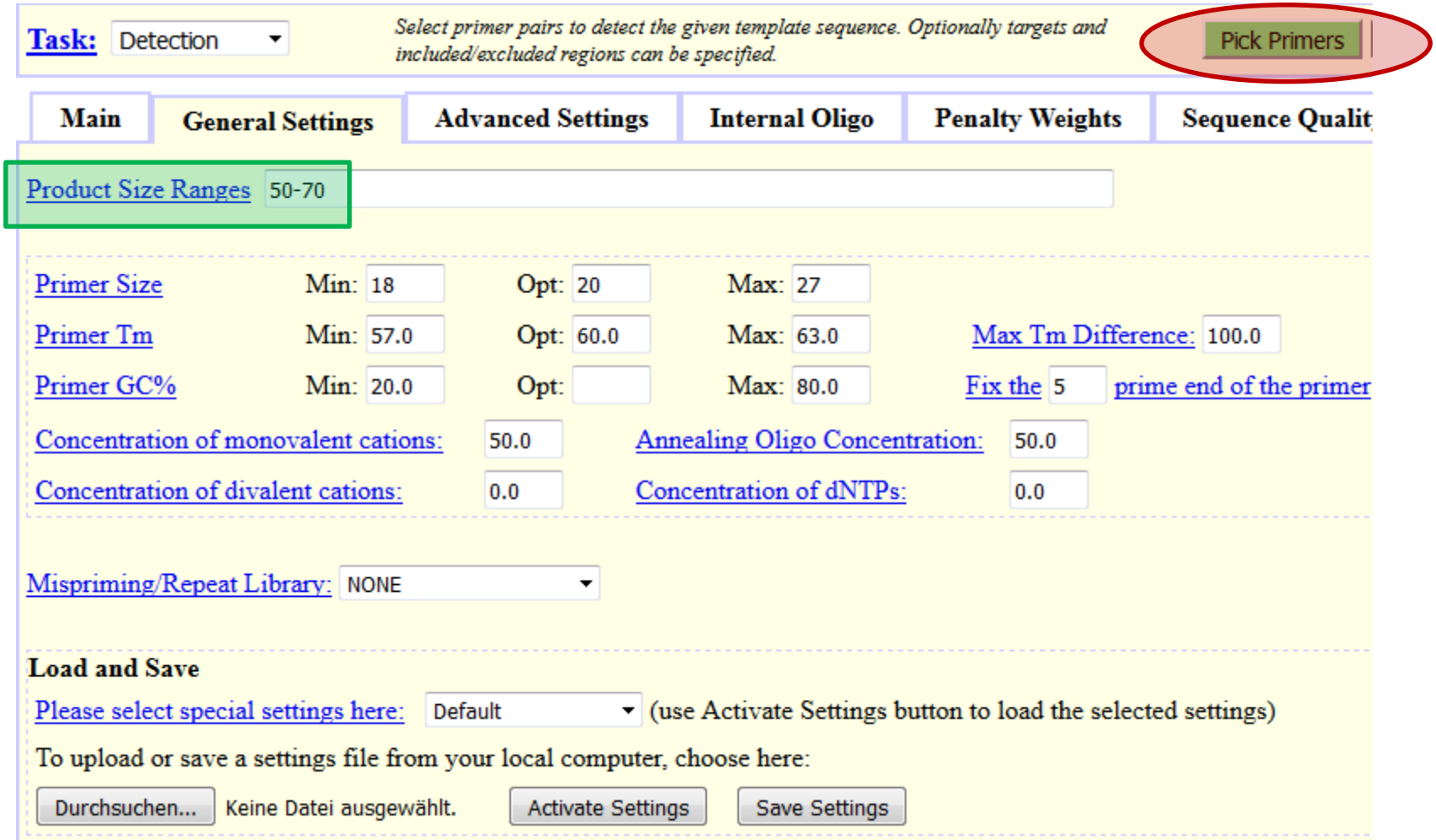


Primer3Plus



We want that the SNP is included in the final PCR product.
Therefore, mark the SNP by using [ ]

Primer3Plus



Add the product size range of 50-70bp

Primer3Plus

JOHANNES KEPLER UNIVERSITÄT LINZ

Primer3Plus

pick primers from a DNA sequence

Primer3Manager | Help | About | Source Code

Unrecognized base in input sequence

Unrecognized base = SNP, but it still works…

< Back

Pair 1:

Left Primer 1: Primer_F

Forward Primer in 5`-3` direction

Sequence: CTCGGTGCCTTTAGTGATGG

Start: 205 Length: 20 bp Tm: 60.6 °C GC: 55.0 % ANY: 3.0 SELF: 3.0

Right Primer 1: Primer_R

Reverse Primer in 5`-3` direction = reverse complement

Sequence: CACTCAGTGTGGCAAAGGTG

Start: 271 Length: 20 bp Tm: 60.3 °C GC: 55.0 % ANY: 7.0 SELF: 1.0

Product Size: 67 bp Pair Any: 5.0 Pair End: 0.0

Final product size = 67bp

Send to Primer3Manager | Reset Form

```
1     ATAGAAACTG  GGCATGTGGA  GACAGAGAAG  ACTCTTGGGT  TTCTGATAGG
51    CACTGACTCT  CTCTGCCTAT  TGGTCTATTT  TCCCACCCTT  AGGCTGCTGG
101   TGGTCTACCC  TTGGACCCAG  AGGTTCTTTG  AGTCCTTTGG  GGATCTGTCC
151   ACTCCTGATG  CTGTTATGGG  CAACCCTAAG  GTGAAGGCTC  ATGGCAAGAA
201   AGTGCTCGGT  GCCTTTAGTG  ATGGCCTGGC  TCACCTGGAC  AACCTCAAGG
251   DCACCTTTGC  CACACTGAGT  GAGCTGCACT  GTGACAAGCT  GCACGTGGAT
301   CCTGAGAACT  TCAGGGTGAG  TCTATGGGAC  GCTTGATGTT  TTCTTTCCCC
351   TTCTTTTCTA  TGGTTAAGTT  CATGTCATAG  GAAGGGGATA  AGTAACAGGG
401   TACAGTTTAG  AATGGGAAAC  AGACGAATGA  TTGCATCAGT  GTGGAAGTCT
451   CAGGATCGTT  TTAGTTTCTT  TTATTTGCTG  TTCATAACAA  TTGTTTTCTT
501   T
```

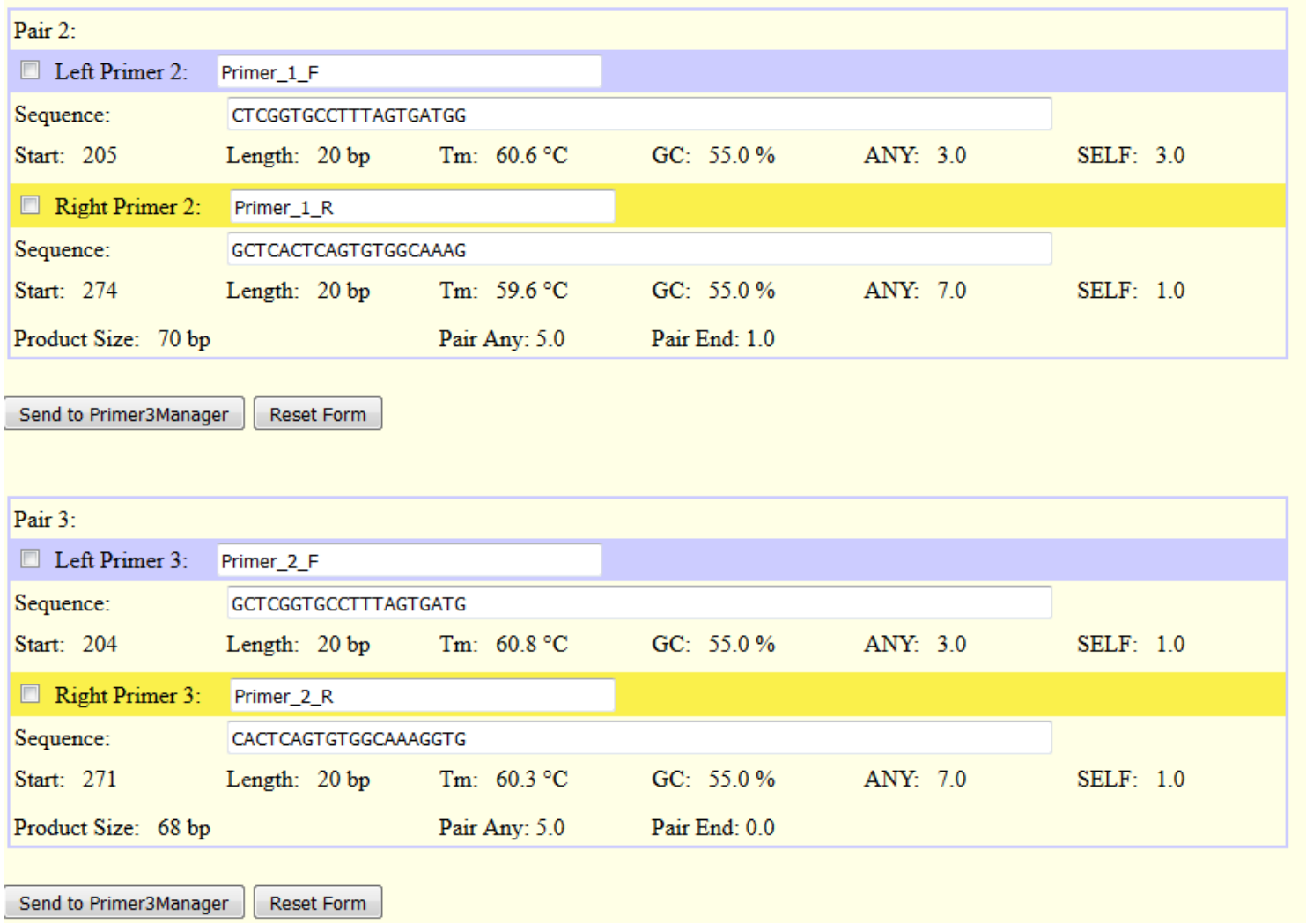
Blue: forward primer

Yellow: reverse primer

Green: target region

Primer3Plus

- When you scroll down there are even more results.



Pair 2:

Left Primer 2: Primer_1_F

Sequence: CTCGGTGCCTTTAGTGATGG

Start: 205 Length: 20 bp Tm: 60.6 °C GC: 55.0 % ANY: 3.0 SELF: 3.0

Right Primer 2: Primer_1_R

Sequence: GCTCACTCAGTGTGGCAAAG

Start: 274 Length: 20 bp Tm: 59.6 °C GC: 55.0 % ANY: 7.0 SELF: 1.0

Product Size: 70 bp Pair Any: 5.0 Pair End: 1.0

Send to Primer3Manager Reset Form

Pair 3:

Left Primer 3: Primer_2_F

Sequence: GCTCGGTGCCTTTAGTGATG

Start: 204 Length: 20 bp Tm: 60.8 °C GC: 55.0 % ANY: 3.0 SELF: 1.0

Right Primer 3: Primer_2_R

Sequence: CACTCAGTGTGGCAAAGGTG

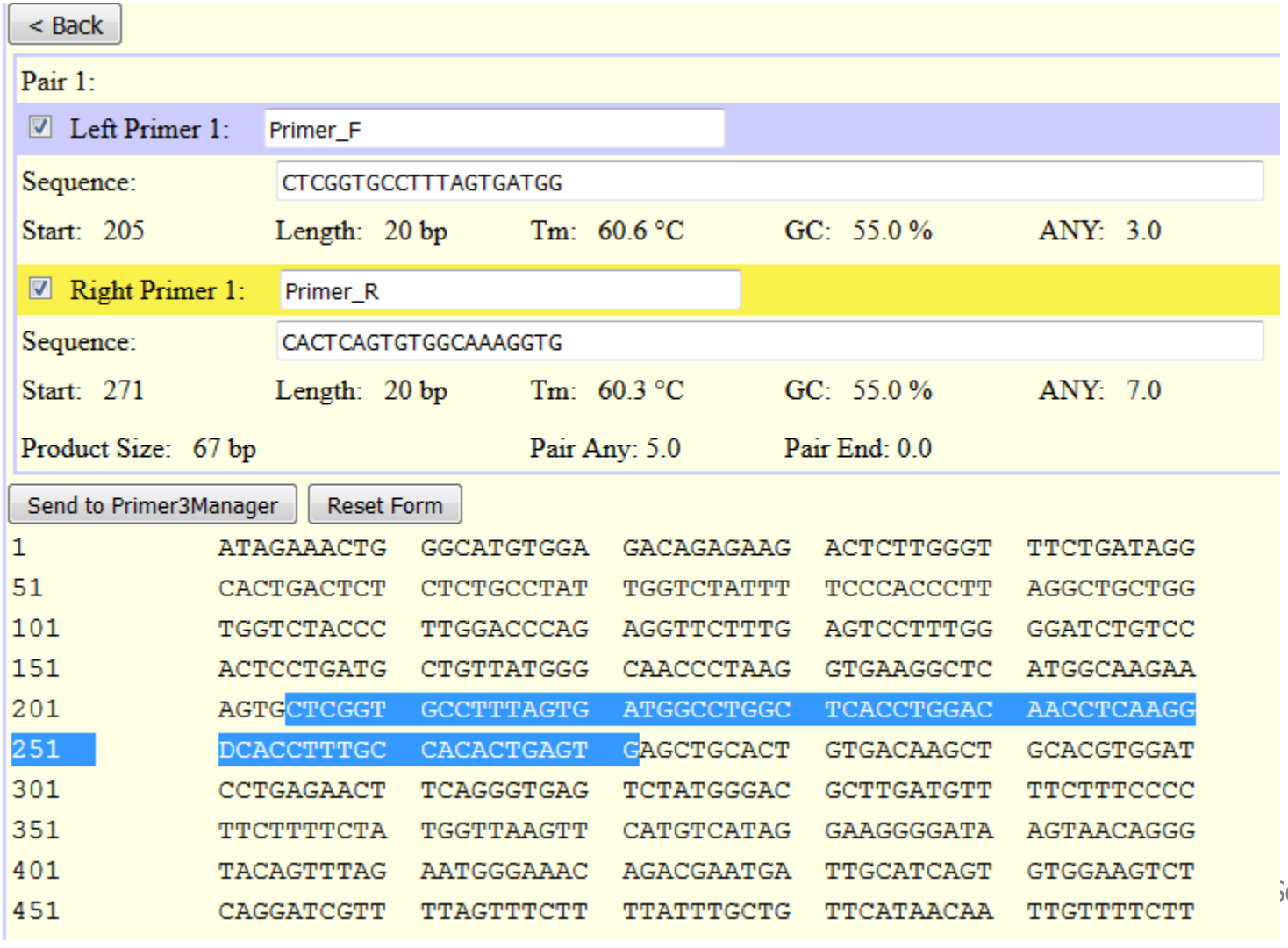
Start: 271 Length: 20 bp Tm: 60.3 °C GC: 55.0 % ANY: 7.0 SELF: 1.0

Product Size: 68 bp Pair Any: 5.0 Pair End: 0.0

Send to Primer3Manager Reset Form

Primer3Plus

- Copy the sequence of your PCR product

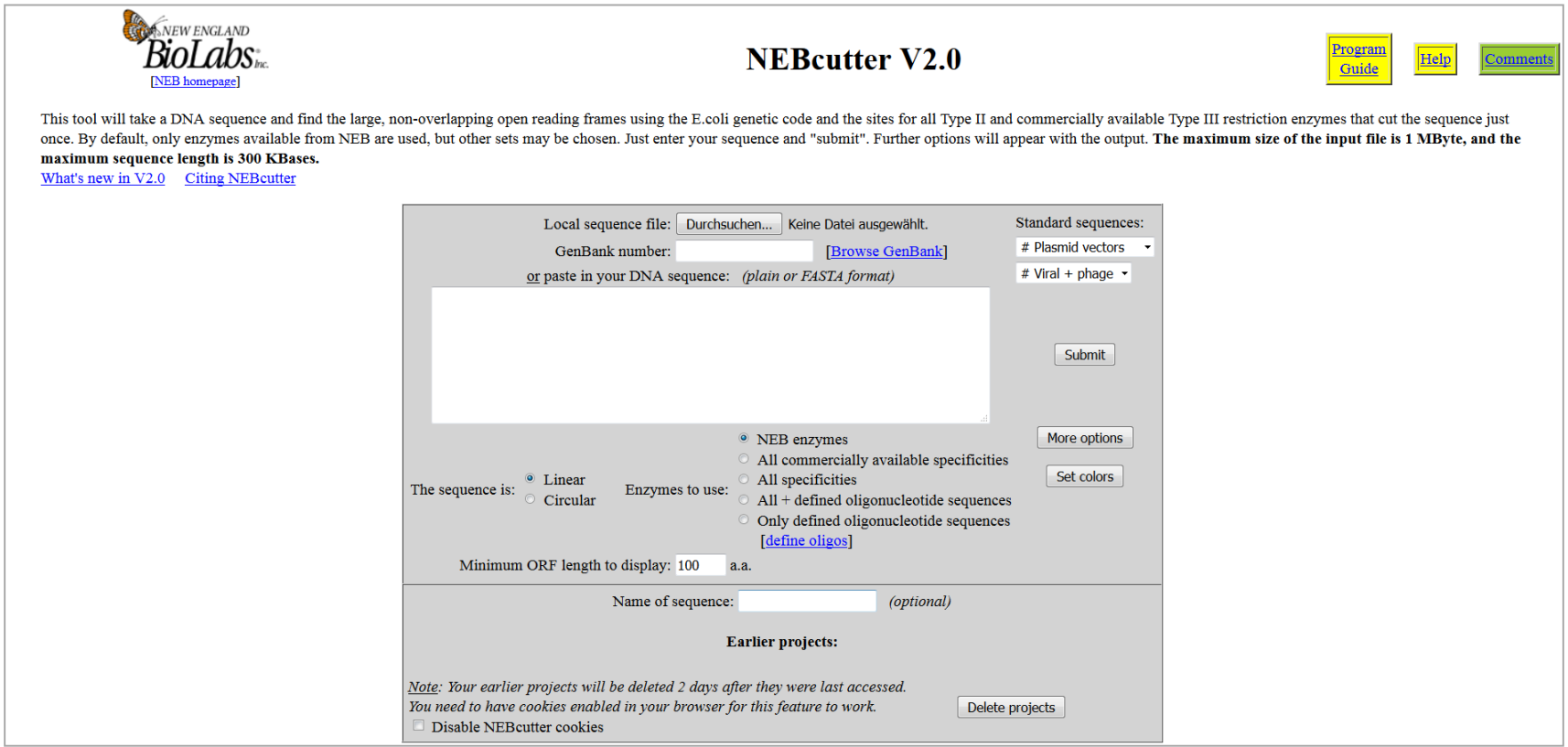


Overview

- PubMed - Literature search
- SNP Database
  - Hardy-Weinberg Equilibrium
- Primer3Plus
  - Polymerase-Chain Reaction (PCR)
  - Primer design
- **NEB Cutter**

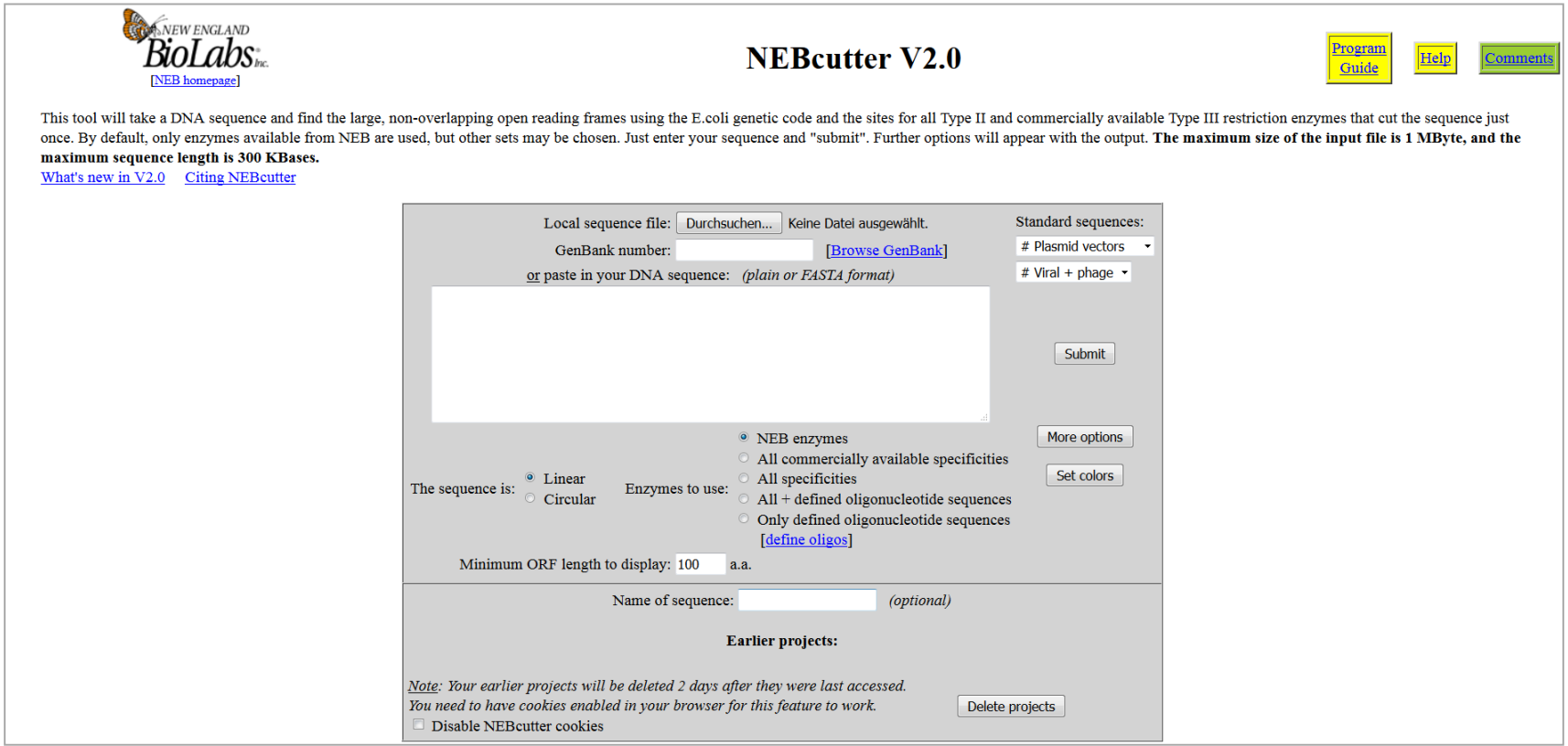
NEBcutter

- http://tools.neb.com/NEBcutter2/
  (search for NEBcutter in google)
- With the NEBcutter you can do **Restriction Enzyme digests**
- This can be used to do a '**Genotyping**' assay = determine the genotype at a certain SNP position



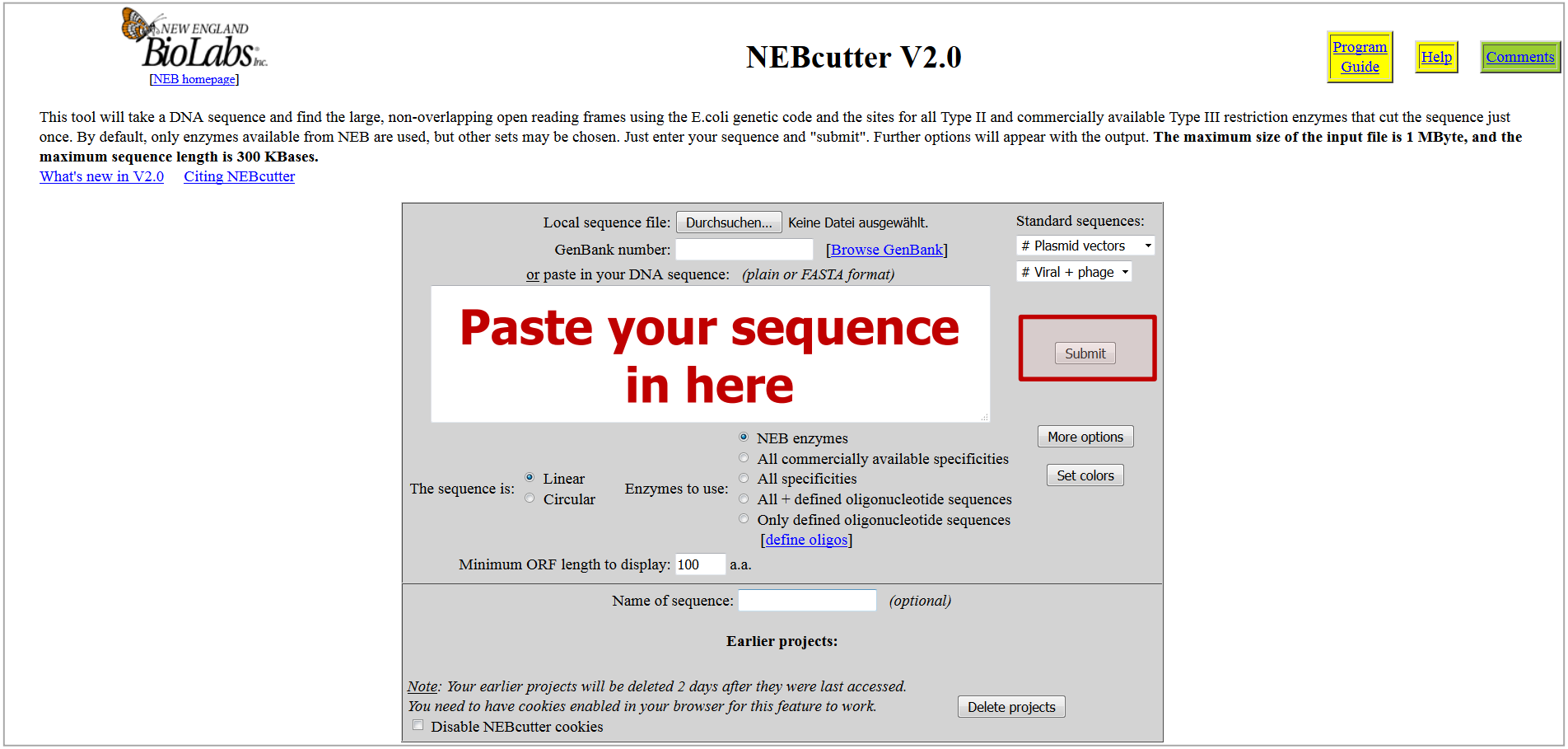
NEBcutter

- Restriction Enzymes can cut the DNA at very specific sites
- These sites often are palindromic sequences: **GGATCC**
  CCTAGG



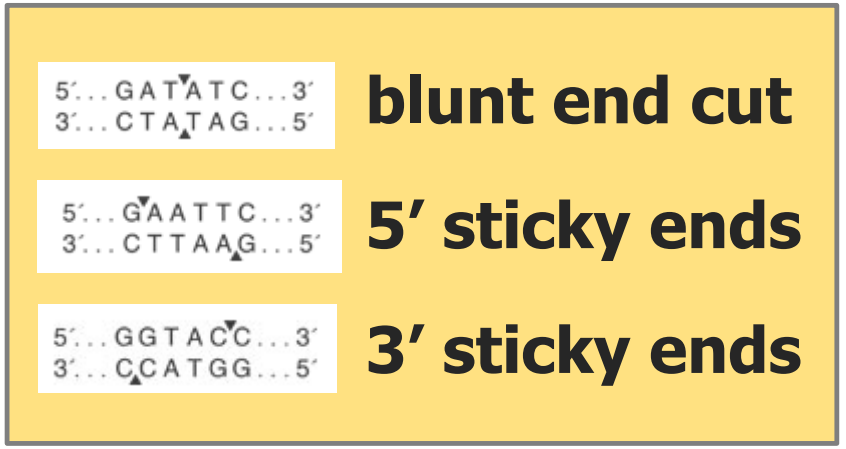
NEBcutter

- To determine the genotype at the position of **our SNP**, we are looking for an enzyme with an recognition sequence including the **SNP** and cuts only **once** within the entire sequence



NEBcutter

JOHANNES KEPLER UNIVERSITÄT LINZ



NEW ENGLAND BioLabs Inc. NEBcutter

Linear Sequence: ***unnamed sequence***

Help | Comments

Display: - NEB single cutter restriction enzymes
- Main non-overlapping, min. 100 aa ORFs

GC=57%, AT=42%

Cleavage code
- blunt end cut
- 5′ extension
- 3′ extension
- cuts 1 strand

Enzyme name code
Available from NEB
Has other supplier
Not commercially available
*: cleavage affected by CpG meth.
#: cleavage affected by other meth.
(enz.name): ambiguous site

1 — 67

BccI, HaeIII, BpuEI, SmlI, (Tsp45I), (BstEII), MnlI (BaeGI), (Bsp1286I), BspCNI, BtsIMutI, DdeI, DraIII, TspRI

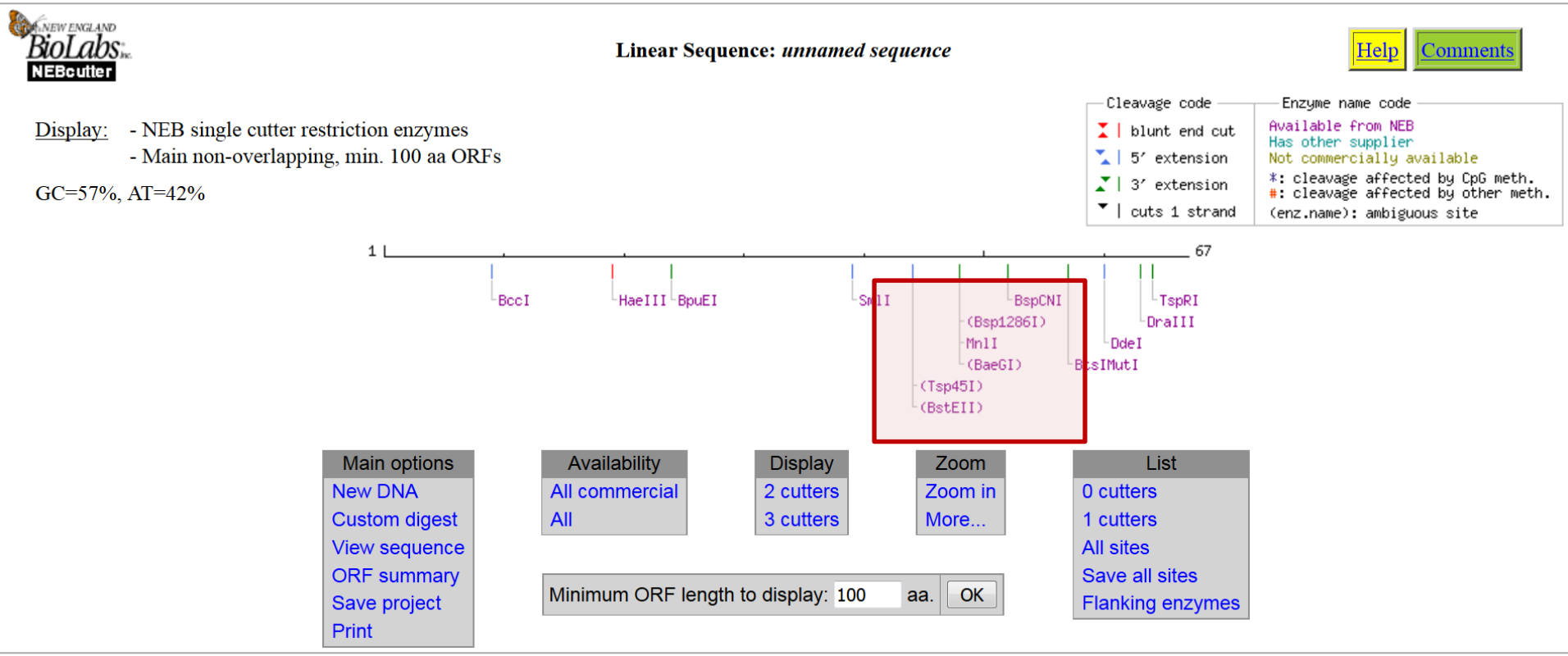
Main options	Availability	Display	Zoom	List
New DNA	All commercial	2 cutters	Zoom in	0 cutters
Custom digest	All	3 cutters	More...	1 cutters
View sequence				All sites
ORF summary				Save all sites
Save project				Flanking enzymes
Print				

Minimum ORF length to display: 100 aa. OK

NEBcutter

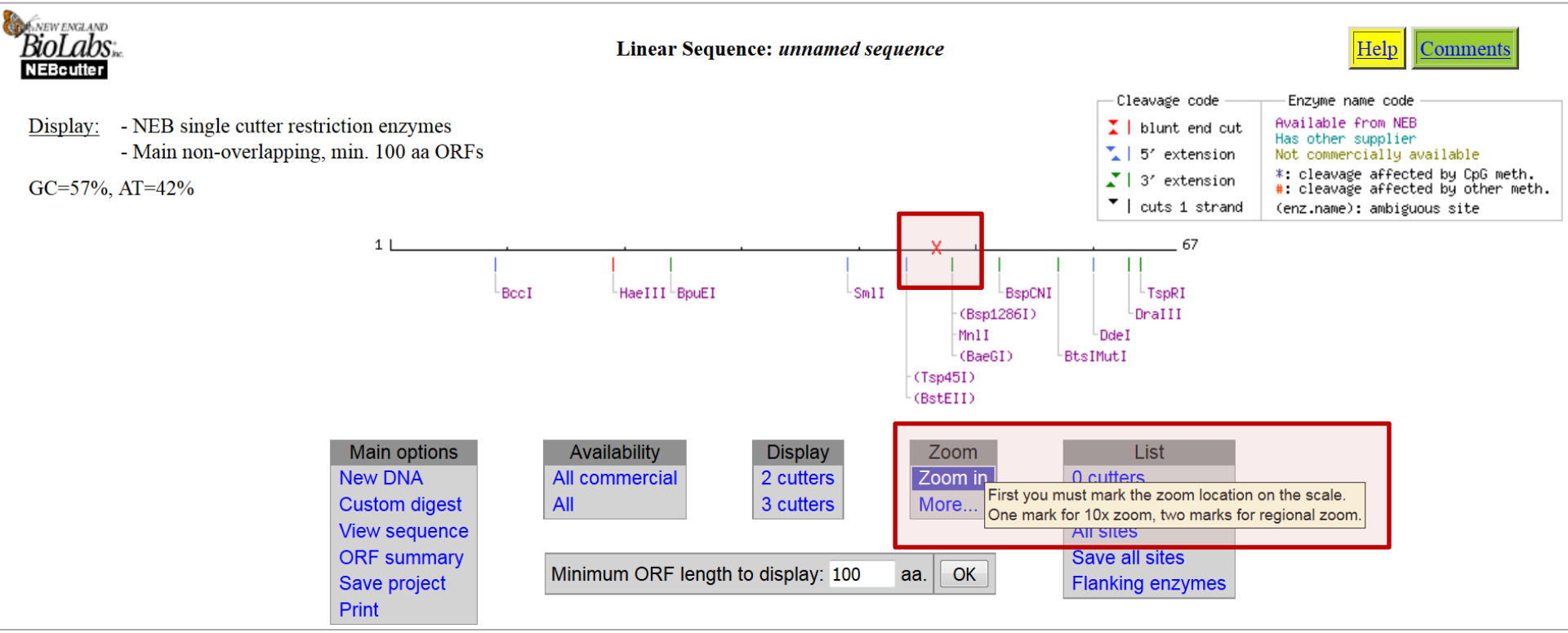
Enzymes in **parenthesis ()** have ambiguous sites, meaning that they need a precise sequence but overlap a degenerate base.

The only degenerate base in the sequence is the **SNP 'D'**, so these enzymes are of special interest.



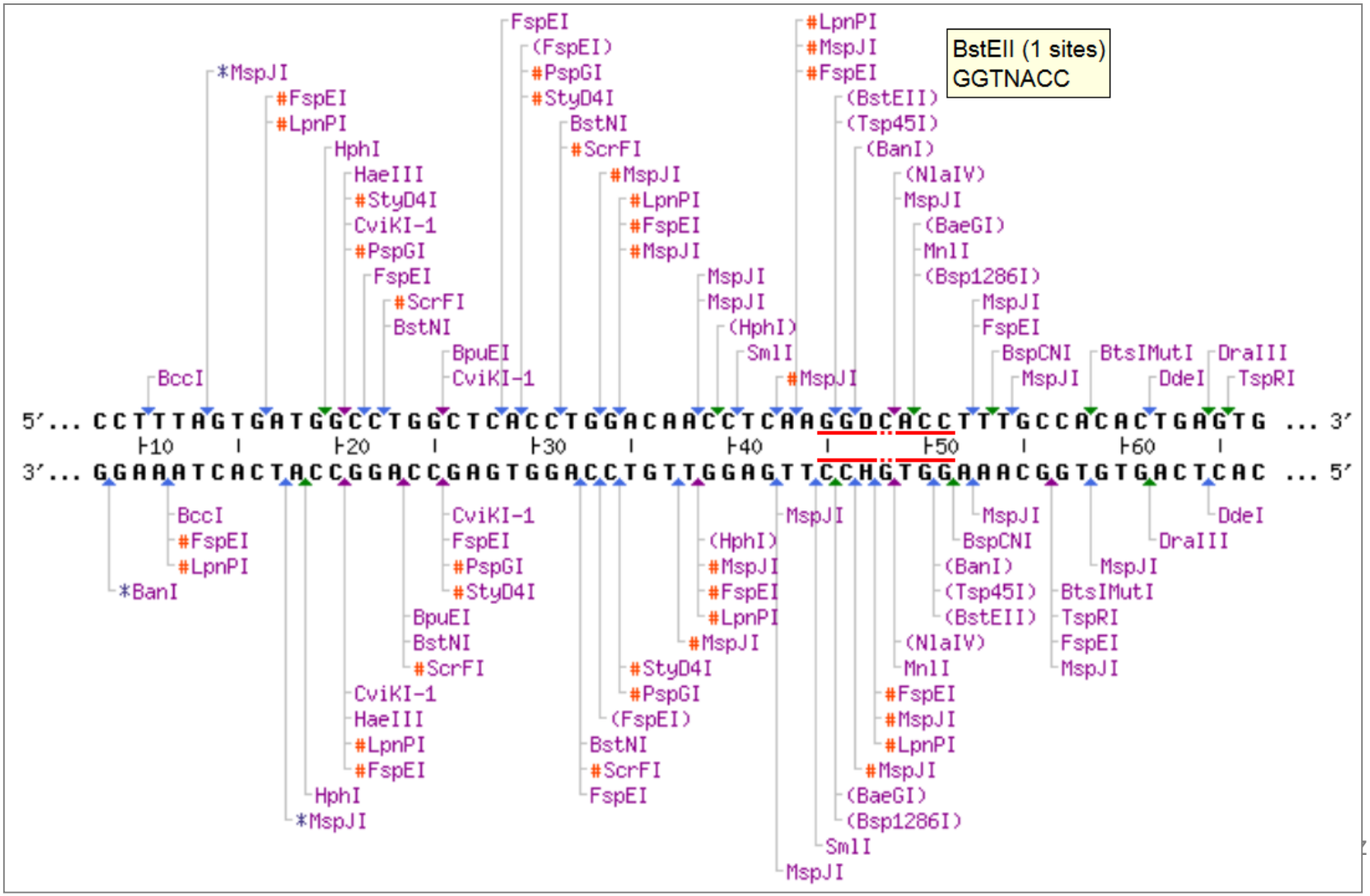
NEBcutter

- Zoom in to the region of interest



NEBcutter

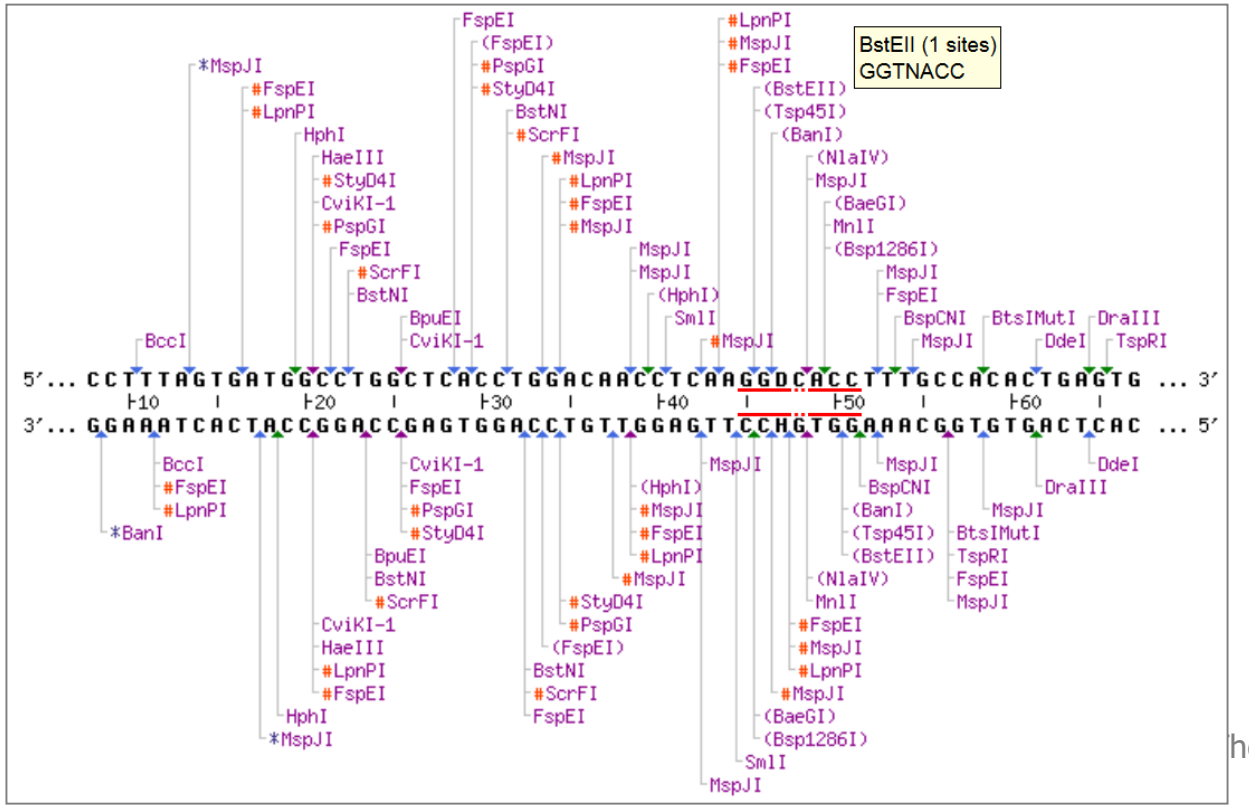
- Find the target DNA sequence of a restriction enzyme (**recognition site**) by moving the mouse cursor over the enzyme names.



NEBcutter

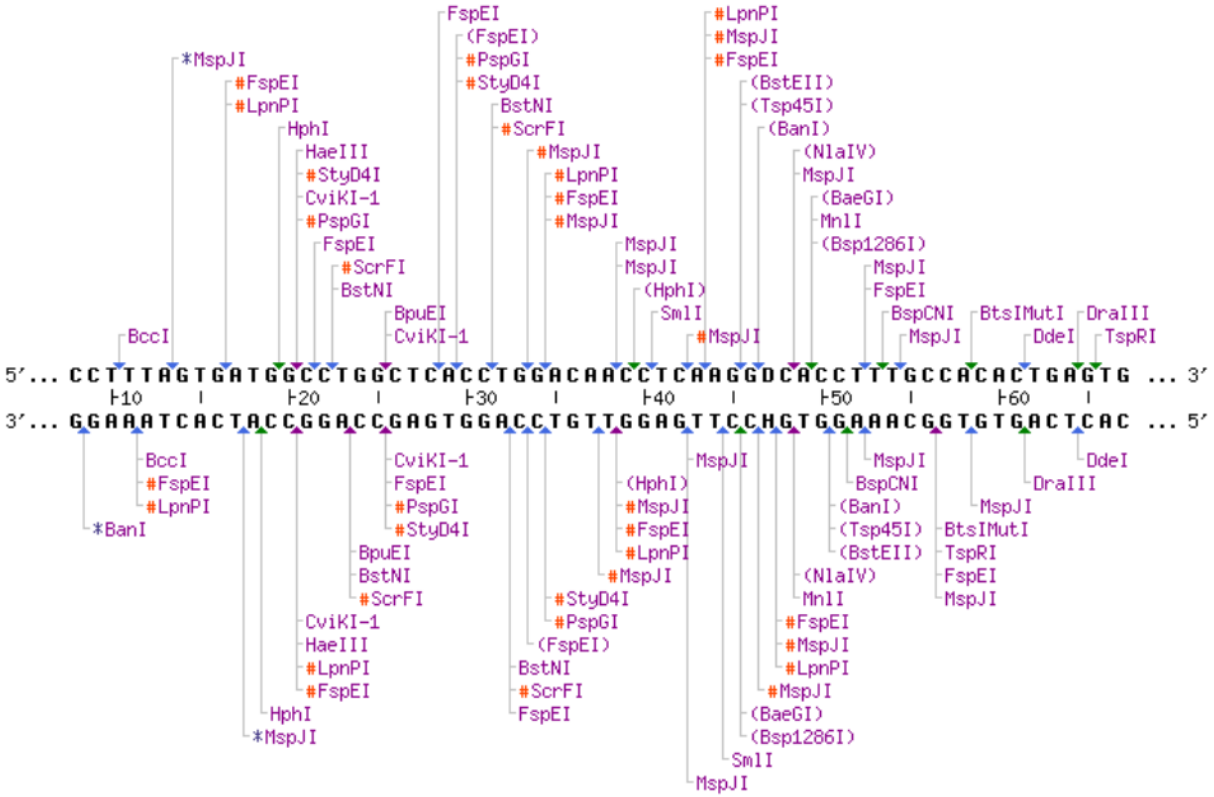
- Find an enzyme that cuts only **once** and **includes the SNP** in its recognition site.
- Example:
  BstEII recognition site = GGTNACC (N = any base)

  remember what kind of SNP 'D' is: **A** or **G** or **T** → **Will it cut?**



NEBcutter

- To theoretically do a restriction enzyme digest, go to **Custom digest**



Main options
New DNA
Custom digest
View sequence
Save project
Print

Availability
All commercial
All

Display
Highlight bases

Zoom
<< UnZoom >>
More...

NEBcutter

- Select the enzyme: BstEII

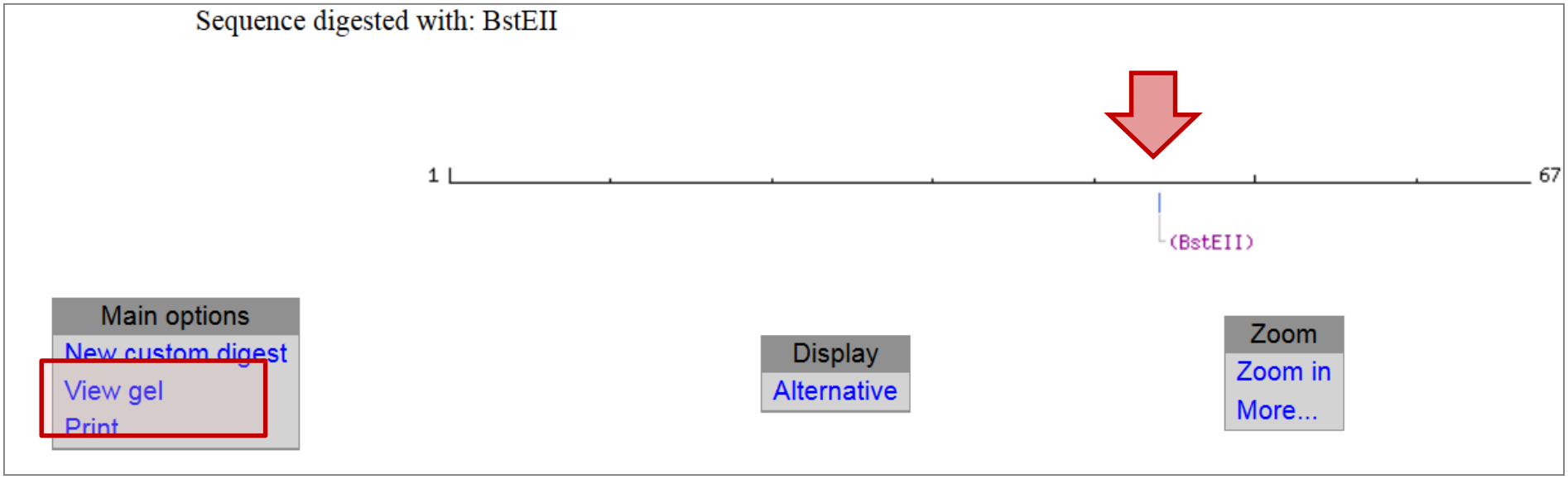
Pick all	Enzyme	Specificity	Cuts	% activity in			
				1.1	2.1	3.1	CS
☐	BaeGI	G▴KGCM▾C	1	75	75	100	25
☐	BanI	G▾GYRC▴C	2	10	25	10	100
☐	BccI	CCATCNNNN▾N▴	1	100	50	10	100
☐	BpuEI	CTTGAG(N)$_{14}$▴NN▾	1	50*	100	50*	100
☐	Bsp1286I	G▴DGCH▾C	1	25	25	25	100
☐	BspCNI	CTCAG(N)$_7$▴NN▾	1	100	75	10	100
☑	BstEII	G▾GTNAC▴C	1	10	75*	100	75*
☐	BstNI	CC▾W▴GG	2	10	100	100	75
☐	BtsIMutI	CAGTG▴NN▾	1	100	50	10	100
☐	CviKI-1	RG▾CY	2	25	100	100	100
☐	DdeI	C▾TNA▴G	1	75	100	100	100
☐	DraIII	CAC▴NNN▾GTG	1	-	-	-	-
☐	FspEI	CC(N)$_{12}$▾NNNN▴	9	10	10	10	100
☐	HaeIII	GG▾CC	1	50	100	25	100
☐	HphI	GGTGA(N)$_8$▾N	2	50	50	10	100

Pick previous enzymes

Digest

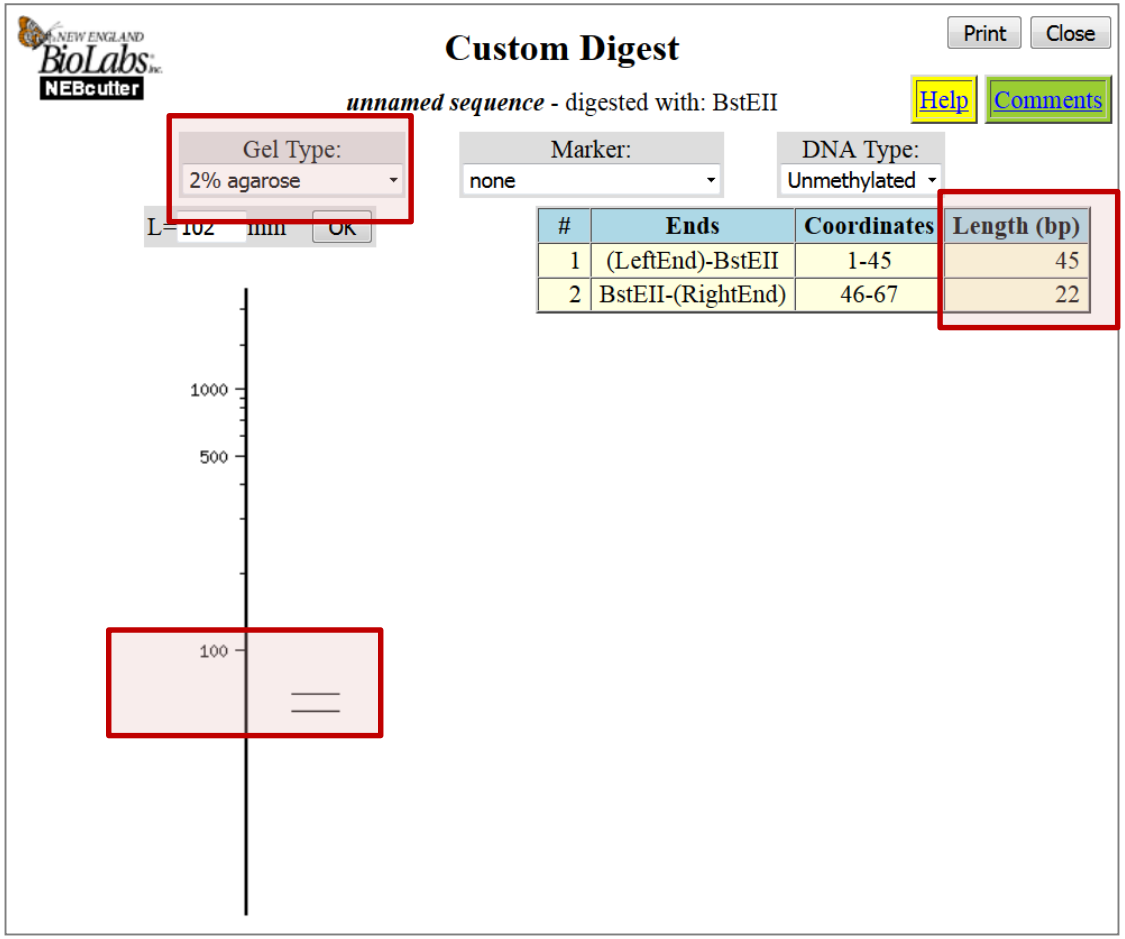
NEBcutter

- You can see where in your sequence the enzyme is cutting
- Select **View gel** to see how your sample would look like when doing gel electrophoresis



NEBcutter

- How many bands can you see on the gel? → **2**
- How long are those DNA fragments? → **45bp and 22bp**

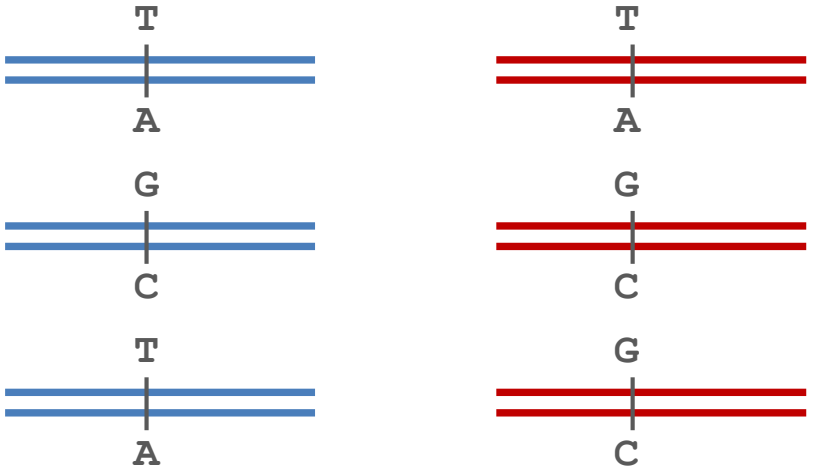


Example of application

- How can such an experiment be used in the lab?

→ to determine the genotype of individuals = genotyping

- Assume there are 3 individuals with the following genotype at a certain SNP
  - ➢ One is homozygous for T
  - ➢ One is homozygous for G
  - ➢ One is T/G heterozygous



- It's your task to find out who has which allele(s) by performing a genotyping assay (PCR + Restriction enzyme digest).

Example of application

JOHANNES KEPLER UNIVERSITÄT LINZ

1. Design and perform a PCR of e.g. **100bp** including the SNP

2. Design and perform a restriction enzyme digest using an enzyme with a recognition site including the SNP
   !! The enzyme must only cut ONE allele !!
   Let's assume that the enzyme cuts **after base 80** only when our **SNP** is a **T**

3. Do gel electrophoresis with your digested PCR product to see the sizes of your PCR fragment(s)

4. How many fragments do you get and how long are they?

	how many fragments	how long are the fragments
T/T homozygous	2	80bp, 20bp
G/G homozygous	1	100bp
T/G heterozygous	3	100bp, 80bp, 20bp

- QUESTIONS?
- Please, download **Report #2** from MOODLE
- Upload the Report until next Monday 8:00 a.m.

GOOD LUCK!