

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!**

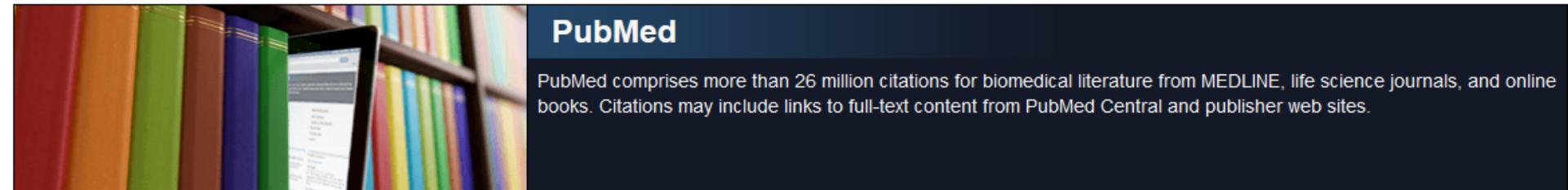
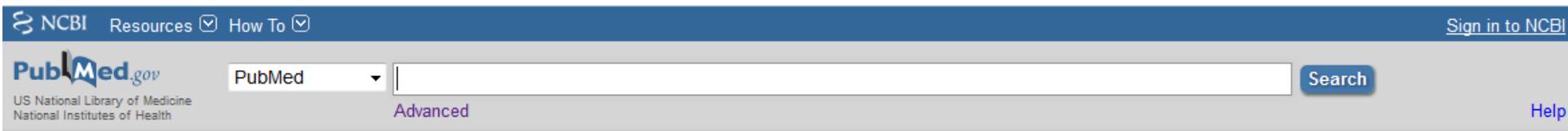
- ❑ **PubMed - Literature search**

- ❑ **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**

Search for **Tiemann-Boege**

NCBI Resources How To

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

PubMed

Create RSS Create alert Advanced

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](#)

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

Send to

Search results

Items: 19

- [The consequences of sequence erosion in the evolution of recombination hotspots.](#)
 1. **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](#)
- [Water transport through the intestinal epithelial barrier under different osmotic conditions is dependent on LI-cadherin trans-interaction.](#)
 2. 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](#)
- [The long zinc finger domain of PRDM9 forms a highly stable and long-lived complex with its DNA recognition sequence.](#)
 3. 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

Format: Abstract ▾

journal and publication date

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 I¹, Schwarz T², Striedner Y², Heissl A².

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.

direct link to online paper

Send to ▾

Full text links



Save items

★ Add to Favorites ▾

author
information

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-I [Chromosome Res. 2017]

The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K9me3 [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...

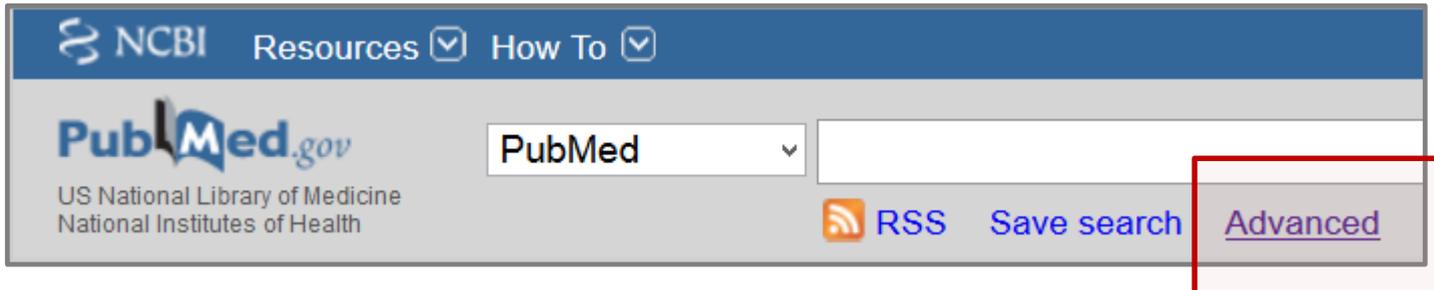
See all...

© 2017 The Authors.

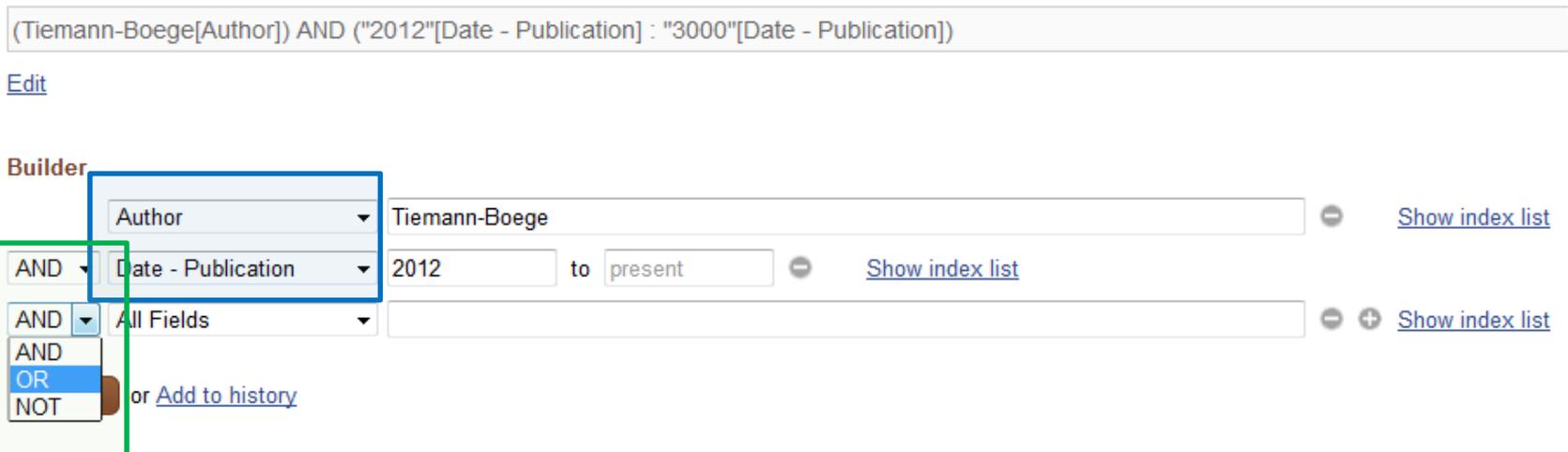
abstract

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

PubMed – Advanced search



PubMed Advanced Search Builder



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

PubMed | lipocalin AND disease |
Create RSS Create alert Advanced

Article types Format: Summary ▾ Sort by: Most Recent ▾
Clinical Trial
Review
Customize ... **Search results**
Text availability **Items: 1 to 20 of 1644**

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

PubMed | lipocalin OR disease |
Create RSS Create alert Advanced

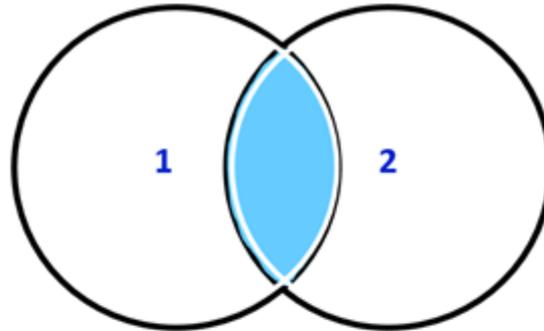
Article types Format: Summary ▾ Sort by: Most Recent ▾
Clinical Trial
Review
Customize ... **Search results**
Text availability **Items: 1 to 20 of 3709258**

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

PubMed | lipocalin NOT disease |
Create RSS Create alert Advanced

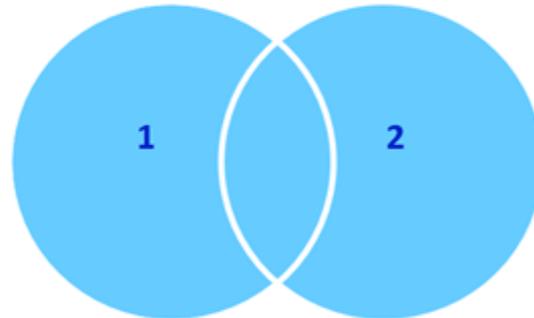
Article types Format: Summary ▾ Sort by: Most Recent ▾
Clinical Trial
Review
Customize ... **Search results**
Text availability **Items: 1 to 20 of 4577**

1 **AND** 2



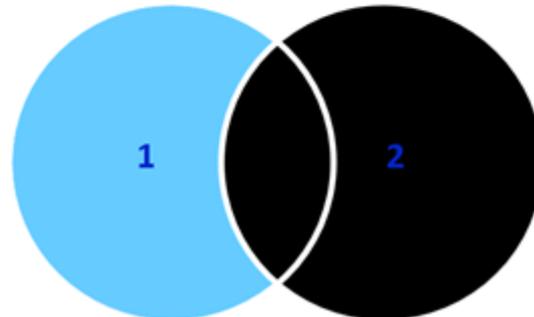
lipocalin **AND** disease
(1644 results)

1 **OR** 2



lipocalin **OR** disease
(3709258 results)

1 **NOT** 2



lipocalin **NOT** disease
(4577 results)

- ❑ PubMed - Literature search

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

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

- ❑ NEB Cutter

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**

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**

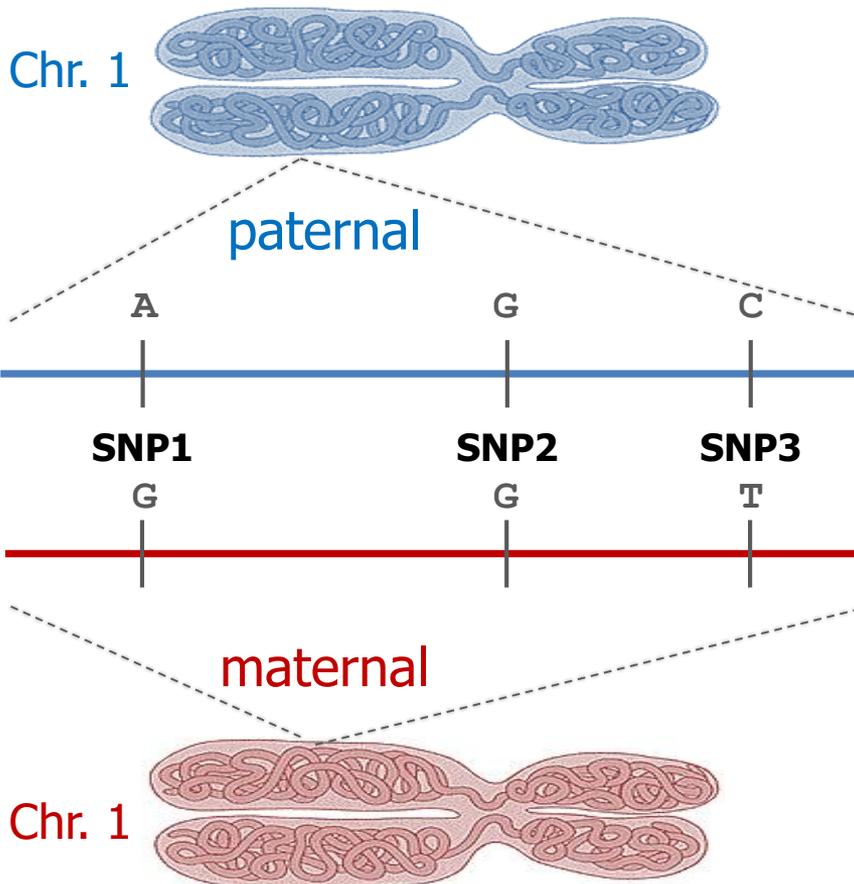
What is a haplotype?

- 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.

The screenshot shows the NCBI website interface. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' dropdown menus. Below this is the NCBI logo and 'National Center for Biotechnology Information'. A left sidebar contains a 'Resource List (A-Z)' with categories like 'NCBI Home', 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', and 'Proteins'. A 'All Databases' dropdown menu is open, listing various databases including PMC, PopSet, Probe, Protein, Protein Clusters, PubChem BioAssay, PubChem Compound, PubChem Substance, PubMed, PubMed Health, **SNP** (highlighted in blue), Sparcle, SRA, Structure, Taxonomy, ToolKit, ToolKitAll, ToolKitBook, ToolKitBookgh, and UniGene. The main content area features a 'Welcome to NCBI' message, a description of the center's mission, and three main sections: 'Submit' (with an upload icon), 'Download' (with a download icon), and 'Learn' (with a book icon).

The screenshot shows the dbSNP search interface. It features a search bar with 'dbSNP' on the left and a dropdown menu set to 'SNP'. Below the search bar, the word 'Advanced' is written in blue text, with a red arrow pointing to it from below.

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

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

Gene [Create RSS](#) [Create alert](#) [Advanced](#)

Gene sources: Genomic, Plasmids
 Categories: Alternatively spliced, Annotated genes, Protein-coding, Pseudogene
 Sequence content: CCDS, Ensembl, RefSeq, RefSeqGene
 Status: Current

Tabular 20 per page Sort by Relevance

Search results
 Items: 1 to 20 of 647
 See also 33 discontinued or replaced items.

Name/Gene ID	Description	Location
<input type="checkbox"/> HBB ID: 3043	hemoglobin subunit beta [<i>Homo sapiens</i> (human)]	Chromosome 11, NC_000011.10 (5225466..5227071, complement)
<input type="checkbox"/> HBA1 ID: 3039	hemoglobin subunit alpha 1 [<i>Homo sapiens</i> (human)]	Chromosome 16, NC_000016.10 (176651..177522)

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

SNP Advanced Search Builder

((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]

[Edit](#)

Builder

Gene Name	HBB
AND Organism	homo sapiens
AND Chromosome	11
AND All Fields	

[Search](#) or [Add to history](#)

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]TGCACCATGGTGTCTGTTGAGGTT

Chromosome: 11:5227013

Find related data

Database: Select

Find items

Search details

((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]

Search

Recent activity

Q ((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]

Q Tiemann-Boege (19)

Q Tiemann. (740)

dbSNP

[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: [Filters: Manage Filters](#)

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,...

Find related data
Database:

Search details

Recent activity
 ((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]
 Tiemann-Boege (19)
 Tiemann. (740)

NG_000007.3: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]TGCACCATGGTGTCTGTTGAGGTT
Chromosome: 11:5227013

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

The screenshot displays the NCBI SNP database interface. On the left, a list of filter categories is visible, including 'Function Class', 'Global MAF', 'Validation Status', and 'by-1000 Genomes'. A modal window titled 'Additional filters' is open, showing a list of filter options with checkboxes. The 'Chromosome Range' and 'Heterozygosity' options are highlighted with green boxes. Below the list is a 'Show' button. At the bottom left of the main interface, a 'Show additional filters' button is also highlighted with a green box. The main content area shows search results for a specific SNP, including its coordinates (11:5227013), gene name (HBB), and various genomic annotations. On the right side, there is a search bar with a magnifying glass icon and a list of search results, including '((HBB[Gene Name]) sapiens[Organism])', 'Tiemann-Boege (19)', and 'Tiemann. (740)'.

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

Global MAF
Custom range...

Validation Status
Paralogous or SND
by-1000 Genomes
by-2hit-2allele
by-cluster
by-frequency
by-submitter
no-info

Clear all

Show additional filters

Additional filters

Variation Class
 Clinical Significance
 Annotation
 Function Class
 Global MAF
 Validation Status
 Chromosomes
 Map Weight
 Chromosome Range
 Variation Allele
 Heterozygosity
 Success Rate
 Method Class
 Individual SNP
 Search fields

Show

NG_042296.1:g.5331>G, NG_046672.1:g.49371>A, NG_046672.1:g.49371>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

sapiens]

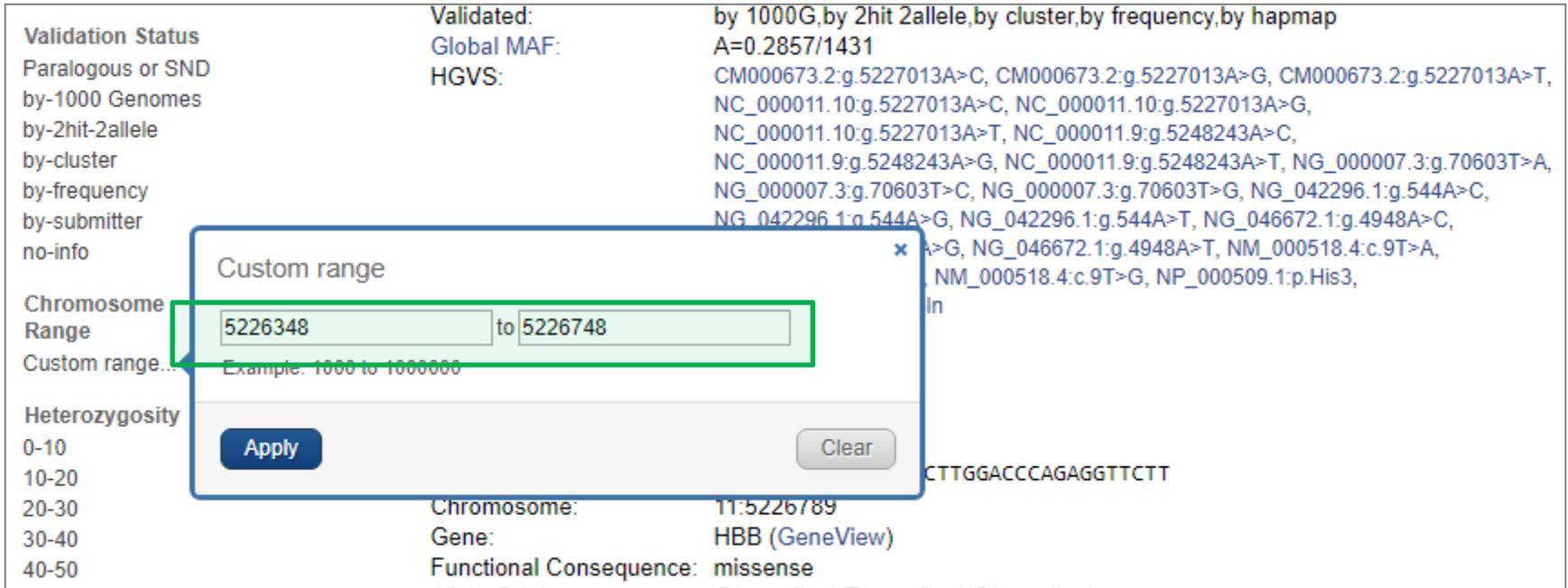
CAGGAGTCAG[A/C/G/T]TGCACCATGGTGTCTGTTTGAGGTT

11:5227013
HBB (GeneView)
Sequence: missense,synonymous codon
T(germline)/A(germline)/C(germline)
other
by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap
A=0.2857/1431
CM000673.2:g.5227013A>C, CM000673.2:g.5227013A>G, CM000673.2:g.5227013A>T,
NC_000011.10:g.5227013A>C, NC_000011.10:g.5227013A>G,
NC_000011.10:g.5227013A>T, NC_000011.9:g.5248243A>C,
NC_000011.9:g.5248243A>G, NC_000011.9:g.5248243A>T, NG_000007.3:g.70603T>A,
NG_000007.3:g.70603T>C, NG_000007.3:g.70603T>G, NG_042296.1:g.544A>C,
NG_042296.1:g.544A>G, NG_042296.1:g.544A>T, NG_046672.1:g.4948A>C,
NG_046672.1:g.4948A>G, NG_046672.1:g.4948A>T, NM_000518.4:c.9T>A,
NM_000518.4:c.9T>C, NM_000518.4:c.9T>G, NP_000509.1:p.His3,
NP_000509.1:p.His3Gln

sapiens]

((HBB[Gene Name]) sapiens[Organism])
Tiemann-Boege (19)
Tiemann. (740)

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



The screenshot displays the NCBI SNP database search results for a specific SNP. A modal window titled "Custom range" is open, showing the range "5226348 to 5226748" entered in the "Chromosome Range" field. The "Apply" button is highlighted. The background shows the search criteria and results for the HBB gene on chromosome 11.

Validation Status	Validated:	by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap
Paralogous or SND	Global MAF:	A=0.2857/1431
by-1000 Genomes	HGVS:	CM000673.2:g.5227013A>C, CM000673.2:g.5227013A>G, CM000673.2:g.5227013A>T, NC_000011.10:g.5227013A>C, NC_000011.10:g.5227013A>G, NC_000011.10:g.5227013A>T, NC_000011.9:g.5248243A>C, NC_000011.9:g.5248243A>G, NC_000011.9:g.5248243A>T, NG_000007.3:g.70603T>A, NG_000007.3:g.70603T>C, NG_000007.3:g.70603T>G, NG_042296.1:g.544A>C, NG_042296.1:g.544A>G, NG_042296.1:g.544A>T, NG_046672.1:g.4948A>C, NG_046672.1:g.4948A>T, NM_000518.4:c.9T>A, NM_000518.4:c.9T>G, NP_000509.1:p.His3,
by-2hit-2allele		
by-cluster		
by-frequency		
by-submitter		
no-info		
Chromosome Range	Chromosome:	11:5226789
Custom range...	Gene:	HBB (GeneView)
Heterozygosity	Functional Consequence:	missense
0-10		
10-20		
20-30		
30-40		
40-50		

dbSNP [Create alert](#) [Advanced](#)

- Variation Class
- in del
- snp
- Clinical Significance
- benign
- likely benign
- likely pathogenic
- other
- pathogenic
- uncertain significance
- Annotation
- Cited in PubMed
- OMIM
- PubMed
- nucleotide
- protein
- structure
- Function Class
- 5' splice site
- coding synonymous
- frame shift
- intron
- missense
- nonsense
- stop gained
- Global MAF
- Custom range...

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

Send to: ▼ [Filters: Manage Filters](#)

Search results

Items: 1 to 20 of 293

<< First < Prev Page of 15 Next > Last >>

 Filters activated: from 5226348 to 5226748. Clear all to show 1776 items.

rs1803195 [*Homo sapiens*]

1.

CTGGCTCACCTGGACAACCTCAAGG[A/G/T]CACCTTTGCCACACTGAGTGAGCTG

Chromosome: 11:5226641
 Gene: HBB (GeneView)
 Functional Consequence: missense
 Allele Origin: G(germline)/A(germline)
 Clinical significance: other
 Validated: by cluster
 HGVS: CM000673.2:g.5226641C>T, 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, NG_042296.1:g.172C>T, NG_046672.1:g.4576C>T, NG_053049.1:g.2962C>T, NM_000518.4:c.251G>A, NM_000518.4:c.251G>T, NP_000509.1:p.Gly84Asp, NP_000509.1:p.Gly84Val

[PubMed](#) [View](#)

rs10768683 [*Homo sapiens*]

2.

AAAGAAGGGGAAAGAAAACATCAAG[C/G/T]GTCCCATAGACTCACCTGAAGTTC
 Chromosome: 11:5226561
 Gene: HBB (GeneView)
 Functional Consequence: intron variant

Find related data

Database: ▼

Search details

((HBB[Gene Name] AND "H [Organism]) AND 11[Chr (5226348[CHRPOS] : 5226748[CHRPOS])

Recent activity

 ((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]

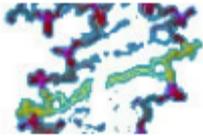
 ((HBB[Gene Name]) AND homo sapiens[Organism]) AND 11[Chromosome]

 Tiemann-Boege (19)

- 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	<input type="text"/>	Go	
Reference SNP (refSNP) Cluster Report rs1803195 ** other **				
RefSNP		Allele		HGVS Names
Organism:	human (<i>Homo sapiens</i>)	Variation Class:	SNV: single nucleotide variation	NC_000011.10:g.5226641C>A
Molecule Type:	Genomic	RefSNP Alleles:	A/G/T (REV)	NC_000011.10:g.5226641C>T
Created/Updated in build:	89/149	Allele Origin:	A:germline G:germline	NC_000011.9:g.5247871C>A
Map to Genome Build:	108/Weight 1	Ancestral Allele:	G	NC_000011.9:g.5247871C>T
Validation Status:		Variation Viewer:		NG_000007.3:g.70975G>A
Citation:	PubMed	Clinical Significance:	other	NG_000007.3:g.70975G>T
		MAF/MinorAlleleCount:	T=0.000008/1 (ExAC)	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:

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 AACTGAGTG AGCTGCACTG TGACAAGCTG CACGTGGATC CTGAGAACTT
CAGGGTGAGT CTATGGGACG CTTGATGTTT TCTTTCCCCT TCTTTTCTAT GGTTAAGTTC
ATGTCATAGG AAGGGGATAA GTAACAGGGT ACAGTTTAGA ATGGGAAACA GACGAATGAT
TGCATCAGTG TGGAAGTCTC AGGATCGTTT TAGTTTCTTT TATTGCTGT TCATAACAAT
TGTTCCTTT
```

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

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**

- 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

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	<i>homozygote</i>	$0.62^2 = 0.3844$	} sum = 1
BB: q^2	<i>homozygote</i>	$0.38^2 = 0.1444$	
AB: $2pq$	<i>heterozygote</i>	$2 * 0.62 * 0.38 = 0.4712$	

$$\mathbf{p + q = 1}$$

$$(p + q)^2 = \mathbf{p^2 + 2pq + q^2 = 1}$$
 (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	} sum = 1
(GAC)4	0.04	
(GAC)5	0.23	
(GAC)6	0.41	

How to calculate genotype frequencies in this case?

→ Use the same mathematical relationship!!

Homozygote: $(GAC)3/(GAC)3 = 0.32^2 = \underline{0.1024}$

Heterozygote: $(GAC)4/(GAC)6 = 2*0.04*0.41 = \underline{0.0328}$

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

dbSNP

SNP

(homo sapiens[Organism]) AND 8[Chromosome]

Search

Variation Class

snp

clear

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

Send to: ▾

Filters: [Manage Filters](#)

Find related data

Database:

Find items

Search details

"Homo sapiens"[Organism] AND 8[Chromosome] AND (missense[Function_Class])

Search

Recent activity

(homo sapiens[Organism]) AND 8[Chromosome] AND (missense[Function_Class])

- 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
- stop gained

Global MAF

Search results

Items: 1 to 20 of 271267

<< First < Prev Page 1 of 13564 Next > Last >>

Filters activated: snp, missense. Clear all to show 35297956 items.

rs268 [Homo sapiens]

TGCAACAATCTGGGCTATGAGATCA [A/G] TAAAGTCAGAGCCAAAAGAAGCAGC
Chromosome: 8:19956018
Gene: LPL (GeneView)
Functional Consequence: missense
Allele Origin: G(germline)/A(germline)
Clinical significance: Pathogenic
Validated: by 1000G,by cluster,by frequency
Global MAF: G=0.0052/26
HGVS: CM000670.2:g.19956018A>G, 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

[PubMed](#) [Varview](#)

rs1124 [Homo sapiens]

2.
CCGGCCTTCTGGGCATGGCCGTGA [A/G] CACCCTGTGTGGCGAGGTGCCGCTC
Chromosome: 8:22164004
Gene: BMP1 (GeneView) SFTPC (GeneView)
Functional Consequence: missense,upstream variant 2KB
Clinical significance: Benign
Validated: by 1000G,by 2hit 2allele,by cluster,by frequency

NCBI – SNP database

Reference SNP (refSNP) Cluster Report: rs268

**** With Pathogenic allele ****

RefSNP	
Organism:	human (<i>Homo sapiens</i>)
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:	
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 [\[here\]](#)

ss#	Sample Ascertainment				Genotype Detail				Alleles	
	Population	Individual Group	Chrom. Sample Cnt.	Source	A/A	A/G	G/G	HWP	A	G
ss132891514	EAS	Populations	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

- 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 [\[here\]](#)

ss#	Sample Ascertainment				Genotype Detail				Alleles	
	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 Beijing, 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	Tosceni 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

- ❑ PubMed - Literature search

- ❑ SNP Database
 - Hardy-Weinberg Equilibrium

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

- ❑ NEB Cutter

Polymerase chain reaction (PCR)

What is the aim of a polymerase chain reaction?

=Amplification of a certain DNA sequence

Which components are needed?

Template DNA

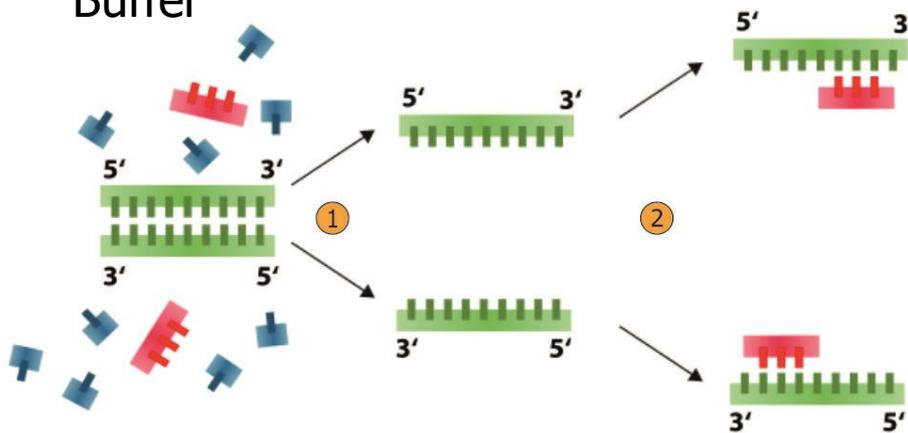
Primer 1 = forward primer

Primer 2 = reverse primer

dNTPs = nucleotides

Polymerase = enzyme

Buffer

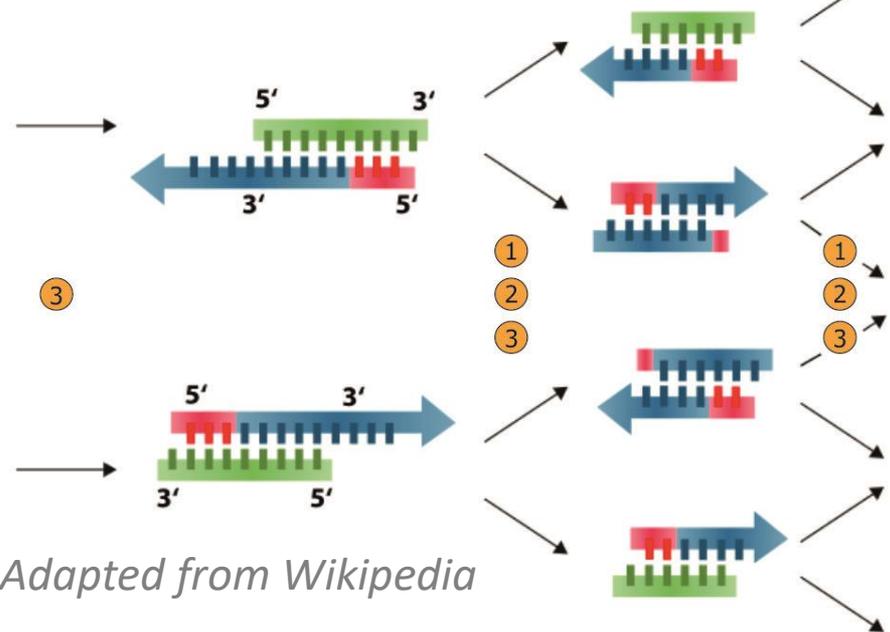


How does a PCR work?

- 1 Denaturation $\sim 94^{\circ}\text{C}$
- 2 Annealing $\sim 60^{\circ}\text{C}$
- 3 Elongation $\sim 70^{\circ}\text{C}$



Use several cycles ($\sim 30x$) to amplify



Adapted from Wikipedia

Template DNA (double-stranded DNA = dsDNA)
(e.g. genomic DNA extracted from blood or saliva)

5' -ATCGGGGCCATATAAATTTGCCCGGGTTTAAAAGATCCATGGATCCCATTTAAGC-3'
|||||
3' -TAGCCCGGGTATATTTAAACGGGCCCAAATTTTCTAGGTACCTAGGGTAAATTCG-5'

- 2 DNA strands

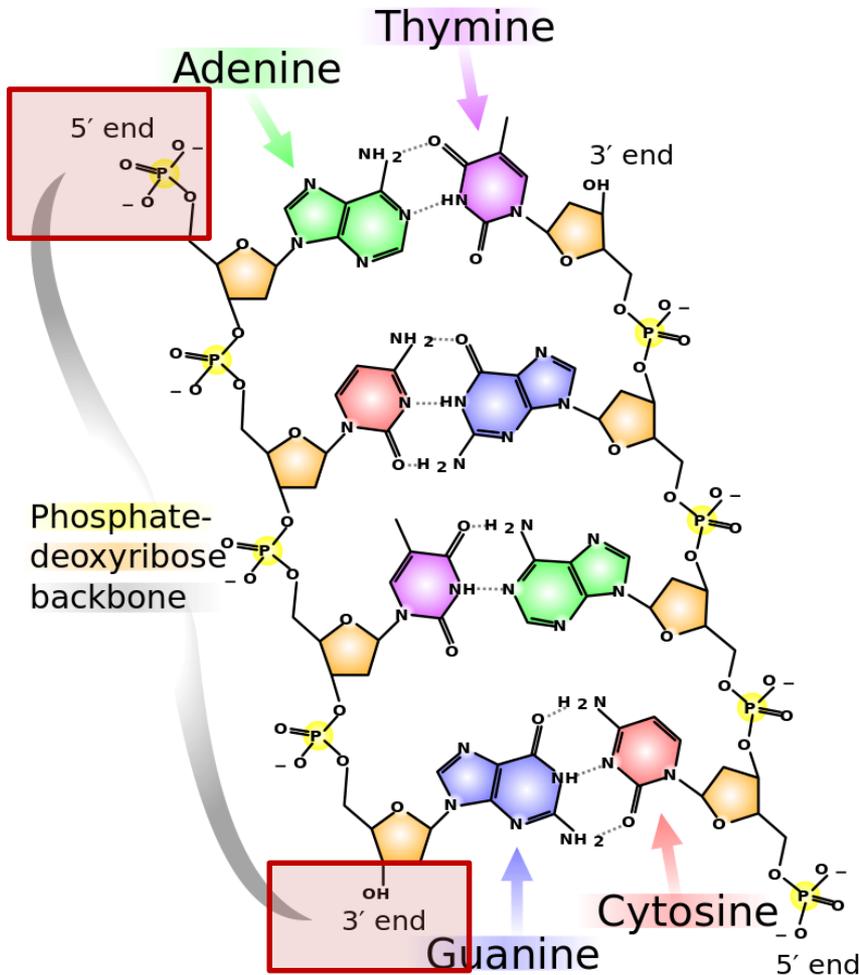
Plus strand	–	Watson strand	–	coding strand
Minus strand	–	Crick strand	–	non-coding strand

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

Polymerase chain reaction (PCR)

5' -ATCGGGGCCCATATAAATTTGCCCGGGTTTAAAAGATCCATGGATCCCATTTAAGC-3'
|||||
3' -TAGCCCGGGTATATTTAAACGGGCCCAAATTTTCTAGGTACCTAGGGTAAATTCG-5'

5' -ATCGGGGCCCATATAAATTTGCCCGGGTTTAAAAGATCCATGGATCCCATTTAAGC-3'
|||||
← 3' -TAGGGTAAATTCG-5'

5' -ATCGGGGCCCAT-3' →
|||||
3' -TAGCCCGGGTATATTTAAACGGGCCCAAATTTTCTAGGTACCTAGGGTAAATTCG-5'

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

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

5' -ATCGGGGCCCATATAAATTTGCCCGGGTTTAAAAGATCCATGGATCCCATTTAAGC-3'



5' -ATCGGGGCCCAT-3'

3' -TAGCCCGGGTATATTTAAACGGGCCCAAATTTTCTAGGTACCTAGGGTAAATTCG-5'

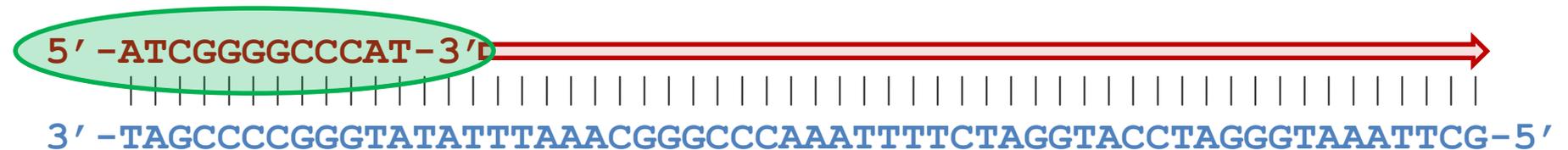
- Fwd primer
- Rvs primer

5' -ATCGGGGCCCAT-3'

5' -GCTTAAATGGGAT-3'

Polymerase chain reaction (PCR)

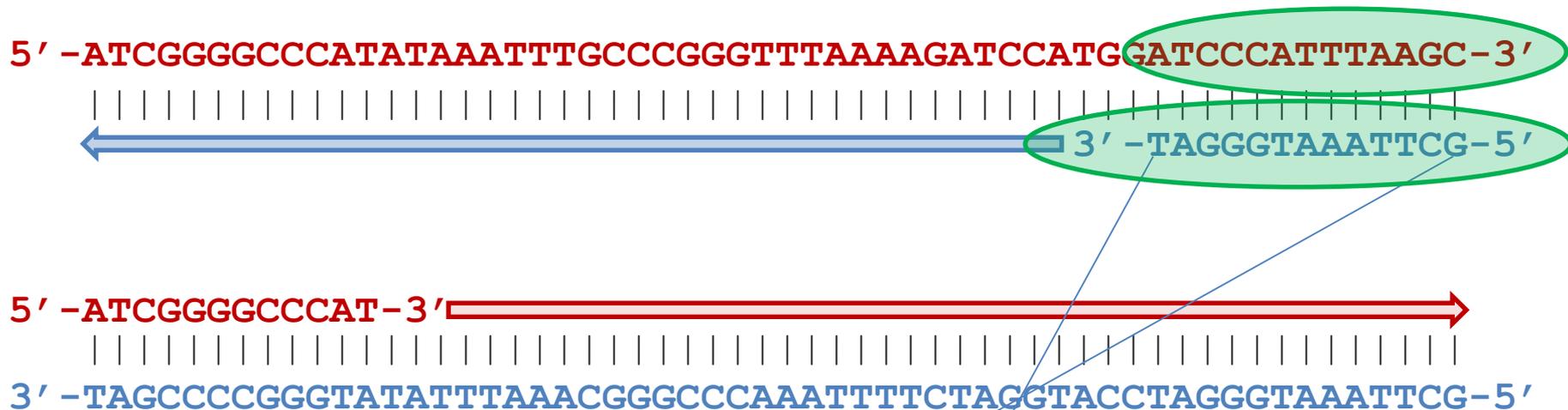
The forward primer has the
same sequence
than plus strand.



- 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'

- 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.

- 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 pick primers from a DNA sequence		Primer3Manager	Help		
		About	Source Code		
Task: Detection	<i>Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.</i>		<input type="button" value="Pick Primers"/> <input type="button" value="Reset Form"/>		
Main	General Settings	Advanced Settings	Internal Oligo	Penalty Weights	Sequence Quality
Sequence Id: <input type="text"/>					
Paste source sequence below Or upload sequence file: <input type="button" value="Durchsuchen..."/> Keine Datei ausgewählt. <input type="button" value="Upload File"/>					
<h2>Paste the sequence of rs1803195</h2> <p>you can download the sequence from MOODLE <i>(info for exercise in computer lab #2)</i></p>					
Mark selected region: <input type="button" value="<>"/> <input type="button" value="[]"/> <input type="button" value="{ }"/> <input type="button" value="Clear"/> <input type="button" value="Save Sequence"/>					
Excluded Regions: < <input type="text"/> >					
Targets: [<input type="text"/>]					
Included Region: { <input type="text"/> }					
<input checked="" type="checkbox"/> Pick left primer or use left primer below.	<input type="checkbox"/> Pick hybridization probe (internal oligo) or use oligo below.	<input checked="" type="checkbox"/> Pick right primer or use right primer below (5'->3' on opposite strand).			
<input type="text"/>	<input type="text"/>	<input type="text"/>			

Task: Detection *Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.* **Pick Primers**

Main **General Settings** **Advanced Settings** **Internal Oligo** **Penalty Weights** **Sequence Quality**

Sequence Id:

Paste source sequence below Or upload sequence file: Keine Datei ausgewählt.

```
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 AACTGAGTG AGCTGCACTG TGACAAGCTG CACGTGGATC CTGAGAACTT
CAGGGTGAGT CTATGGGACG CTTGATGTTT TCTTTCCCCT TCTTTTCTAT GGTTAAGTTC
ATGTCATAGG AAGGGGATAA GTAACAGGGT ACAGTTTAGA ATGGGAAACA GACGAATGAT
TGCATCAGTG TGGAAGTCTC AGGATCGTTT TAGTTTCTTT TATTTGCTGT TCATAACAAT
TGTTTTCTTT
```

Mark selected region:

Targets Region

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

Task: Detection ▾

Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.

Pick Primers

Main

General Settings

Advanced Settings

Internal Oligo

Penalty Weights

Sequence Quality

Product Size Ranges 50-70

Primer Size

Min: 18

Opt: 20

Max: 27

Primer Tm

Min: 57.0

Opt: 60.0

Max: 63.0

Max Tm Difference: 100.0

Primer GC%

Min: 20.0

Opt:

Max: 80.0

Fix the 5 prime end of the primer

Concentration of monovalent cations: 50.0

Annealing Oligo Concentration: 50.0

Concentration of divalent cations: 0.0

Concentration of dNTPs: 0.0

Mispriming/Repeat Library: NONE ▾

Load and Save

Please select special settings here: Default ▾ (use Activate Settings button to load the selected settings)

To upload or save a settings file from your local computer, choose here:

Durchsuchen... Keine Datei ausgewählt.

Activate Settings

Save Settings

Add the product size range of 50-70bp

Primer3Plus

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: **Forward Primer in 5'-3' direction**

Sequence:

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

Right Primer 1: **Reverse Primer in 5'-3' direction = reverse complement**

Sequence:

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      ATAGAACTG  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    DCACCTTGC  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  TTGTTTCTT
501    T
    
```

Blue: forward primer
Yellow: reverse primer
Green: target region

- When you scroll down there are even more results.

Pair 2:

<input type="checkbox"/> 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
<input type="checkbox"/> 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			

Pair 3:

<input type="checkbox"/> 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
<input type="checkbox"/> 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			

- Copy the sequence of your PCR product

< Back

Pair 1:

Left Primer 1:

Sequence:

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

Right Primer 1:

Sequence:

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

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

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

- ❑ **PubMed - Literature search**

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

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

- ❑ **NEB Cutter**

- <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 V2.0** [Program Guide](#) [Help](#) [Comments](#)

This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**
[What's new in V2.0](#) [Citing NEBcutter](#)

Local sequence file: Keine Datei ausgewählt. Standard sequences: # Plasmid vectors
GenBank number: [\[Browse GenBank\]](#) # Viral + phage
or paste in your DNA sequence: (plain or FASTA format)

The sequence is: Linear Circular Enzymes to use: NEB enzymes
 All commercially available specificities
 All specificities
 All + defined oligonucleotide sequences
 Only defined oligonucleotide sequences
[\[define oligos\]](#)

Minimum ORF length to display: a.a.

Name of sequence: (optional)

Earlier projects:
Note: Your earlier projects will be deleted 2 days after they were last accessed. You need to have cookies enabled in your browser for this feature to work.

Disable NEBcutter cookies

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

 **NEBcutter V2.0** [Program Guide](#) [Help](#) [Comments](#)

This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**

[What's new in V2.0](#) [Citing NEBcutter](#)

Local sequence file: Keine Datei ausgewählt. Standard sequences: # Plasmid vectors
GenBank number: [\[Browse GenBank\]](#) # Viral + phage
or paste in your DNA sequence: (plain or FASTA format)

The sequence is: Linear Circular Enzymes to use: NEB enzymes
 All commercially available specificities
 All specificities
 All + defined oligonucleotide sequences
 Only defined oligonucleotide sequences
[\[define oligos\]](#)

Minimum ORF length to display: a.a.

Name of sequence: (optional)

Earlier projects:

Note: Your earlier projects will be deleted 2 days after they were last accessed. You need to have cookies enabled in your browser for this feature to work.

Disable NEBcutter cookies

- 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 V2.0



This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**

[What's new in V2.0](#) [Citing NEBcutter](#)

Local sequence file: Keine Datei ausgewählt.

GenBank number: [\[Browse GenBank\]](#)

or paste in your DNA sequence: *(plain or FASTA format)*

Paste your sequence in here

Standard sequences:
Plasmid vectors ▾
Viral + phage ▾

The sequence is: Linear Circular

Enzymes to use:

- NEB enzymes
- All commercially available specificities
- All specificities
- All + defined oligonucleotide sequences
- Only defined oligonucleotide sequences

[\[define oligos\]](#)

Minimum ORF length to display: a.a.

Name of sequence: *(optional)*

Earlier projects:

Note: Your earlier projects will be deleted 2 days after they were last accessed. You need to have cookies enabled in your browser for this feature to work.

Disable NEBcutter cookies

NEBcutter

5'...GATATC...3'
3'...CTA TAG...5' **blunt end cut**

5'...GAATTC...3'
3'...CTTAAG...5' **5' sticky ends**

5'...GGTACC...3'
3'...CCATGG...5' **3' sticky ends**



Linear Sequence: *unnamed sequence*

[Help](#) [Comments](#)

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

GC=57%, AT=42%

<p>Cleavage code</p> <ul style="list-style-type: none"> ⌵ blunt end cut ⌵ 5' extension ⌵ 3' extension ⌵ cuts 1 strand 	<p>Enzyme name code</p> <ul style="list-style-type: none"> Available from NEB Has other supplier Not commercially available *: cleavage affected by CpG meth. #: cleavage affected by other meth. (enz.name): ambiguous site
---	--



<p>Main options</p> <ul style="list-style-type: none"> New DNA Custom digest View sequence ORF summary Save project Print 	<p>Availability</p> <ul style="list-style-type: none"> All commercial All 	<p>Display</p> <ul style="list-style-type: none"> 2 cutters 3 cutters 	<p>Zoom</p> <ul style="list-style-type: none"> Zoom in More... 	<p>List</p> <ul style="list-style-type: none"> 0 cutters 1 cutters All sites Save all sites Flanking enzymes
--	--	--	---	--

Minimum ORF length to display: aa.

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.

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	Enzyme name code
✂ blunt end cut	Available from NEB
▶ 5' extension	Has other supplier
▶ 3' extension	Not commercially available
▼ cuts 1 strand	*: cleavage affected by CpG meth.
	#: cleavage affected by other meth.
	(enz.name): ambiguous site

1 |-----| 67

BccI HaeIII BpuEI SmaI BspCNI (Bsp1286I) MnlI (BaeGI) BstEII (Tsp45I) BspIMutI DdeI DraIII TspRI

Main options
New DNA
Custom digest
View sequence
ORF summary
Save project
Print

Availability
All commercial
All

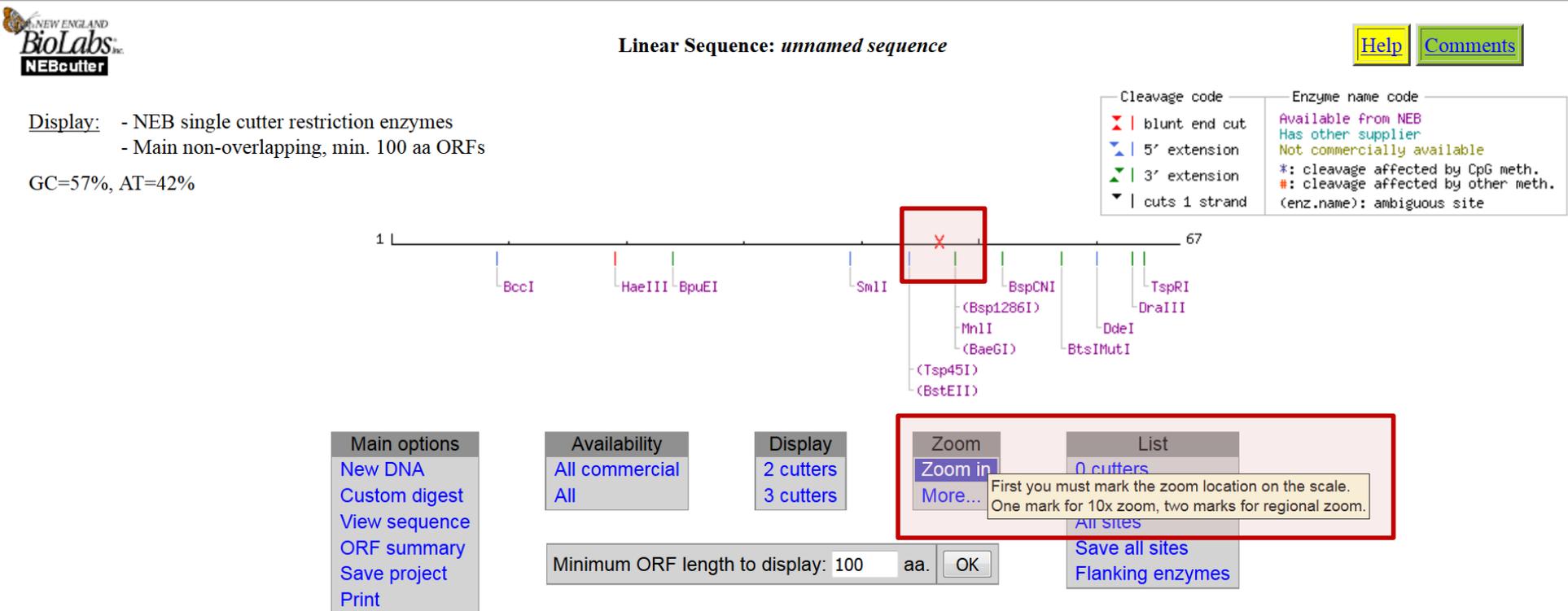
Display
2 cutters
3 cutters

Zoom
Zoom in
More...

List
0 cutters
1 cutters
All sites
Save all sites
Flanking enzymes

Minimum ORF length to display: 100 aa. OK

- Zoom in to the region of interest



Linear Sequence: *unnamed sequence*

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

GC=57%, AT=42%

Legend:

Cleavage code	Enzyme name code
⌵ blunt end cut	Available from NEB
⌵ 5' extension	Has other supplier
⌵ 3' extension	Not commercially available
⌵ cuts 1 strand	*: cleavage affected by CpG meth.
	#: cleavage affected by other meth.
	(enz.name): ambiguous site

Main options: New DNA, Custom digest, View sequence, ORF summary, Save project, Print

Availability: All commercial, All

Display: 2 cutters, 3 cutters

Zoom: Zoom in, More...

List: 0 cutters, All sites, Save all sites, Flanking enzymes

Minimum ORF length to display: 100 aa.

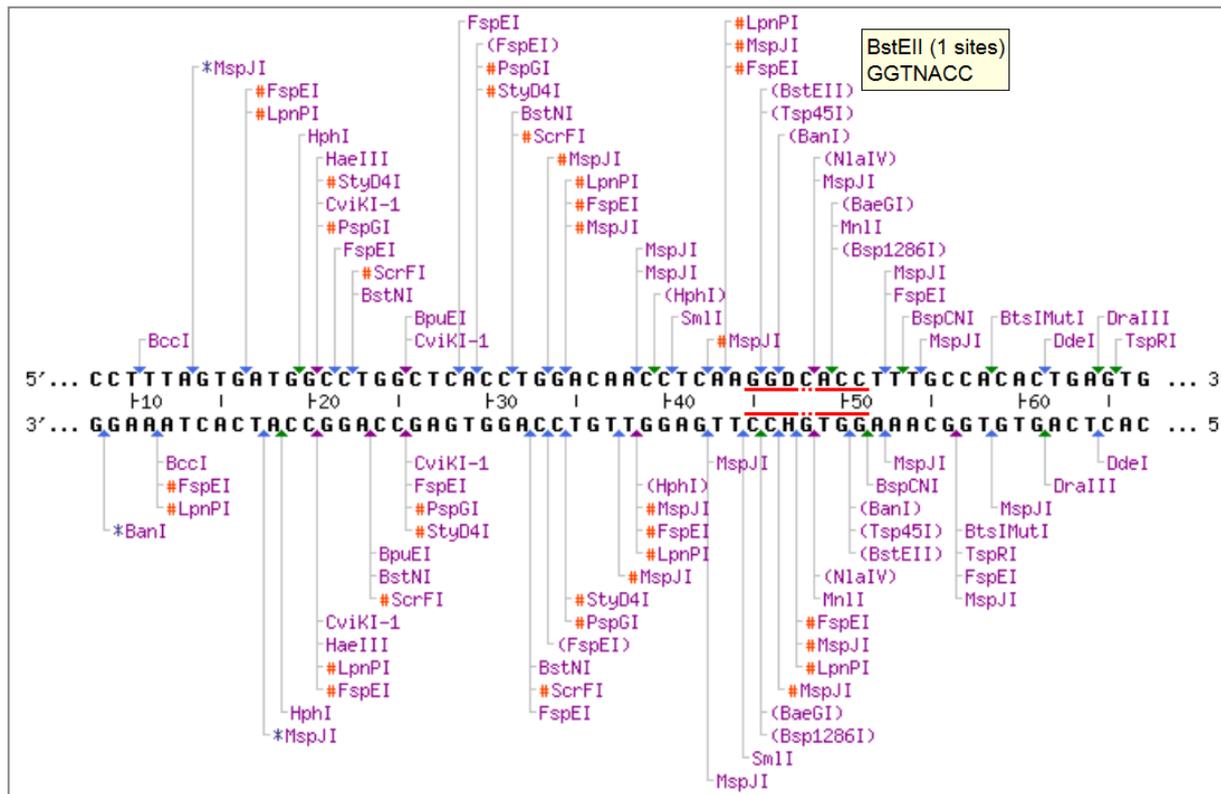
First you must mark the zoom location on the scale. One mark for 10x zoom, two marks for regional zoom.

- 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?**



- Select the enzyme: BstEII

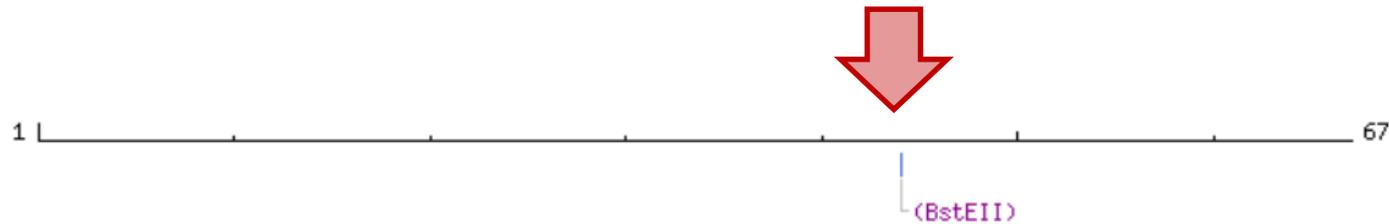
Pick all	Enzyme	Specificity	Cuts	% activity in			
				1.1	2.1	3.1	CS
<input type="checkbox"/>	BaeGI	G _▲ KGCM [▼] C	1	75	75	100	25
<input type="checkbox"/>	BanI	G [▼] GYRC _▲ C	2	10	25	10	100
<input type="checkbox"/>	BccI	CCATC [▼] NNNN [▼] N _▲	1	100	50	10	100
<input type="checkbox"/>	BpuEI	CTTGAG (N) ₁₄ ▲ [▼] NN [▼]	1	50*	100	50*	100
<input type="checkbox"/>	Bsp1286I	G _▲ DGCH [▼] C	1	25	25	25	100
<input type="checkbox"/>	BspCNI	CTCAG (N) ₇ ▲ [▼] NN [▼]	1	100	75	10	100
<input checked="" type="checkbox"/>	BstEII	G [▼] GTNAC _▲ C	1	10	75*	100	75*
<input type="checkbox"/>	BstNI	CC [▼] W _▲ GG	2	10	100	100	75
<input type="checkbox"/>	BtsIMutI	CAGTG _▲ NN [▼]	1	100	50	10	100
<input type="checkbox"/>	CviKI-1	RG [▼] CY	2	25	100	100	100
<input type="checkbox"/>	DdeI	C [▼] TNA _▲ G	1	75	100	100	100
<input type="checkbox"/>	DraIII	CAC _▲ NNN [▼] GTG	1	-	-	-	-
<input type="checkbox"/>	FspEI	CC (N) ₁₂ ▲ [▼] NNNN _▲	9	10	10	10	100
<input type="checkbox"/>	HaeIII	GG [▼] CC	1	50	100	25	100
<input type="checkbox"/>	HobI	GCTGA (N) ₆ ▲ [▼] N [▼]	2	50	50	10	100

Pick previous enzymes

Digest

- 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

Sequence digested with: BstEII



Main options

New custom digest

View gel

Print

Display

Alternative

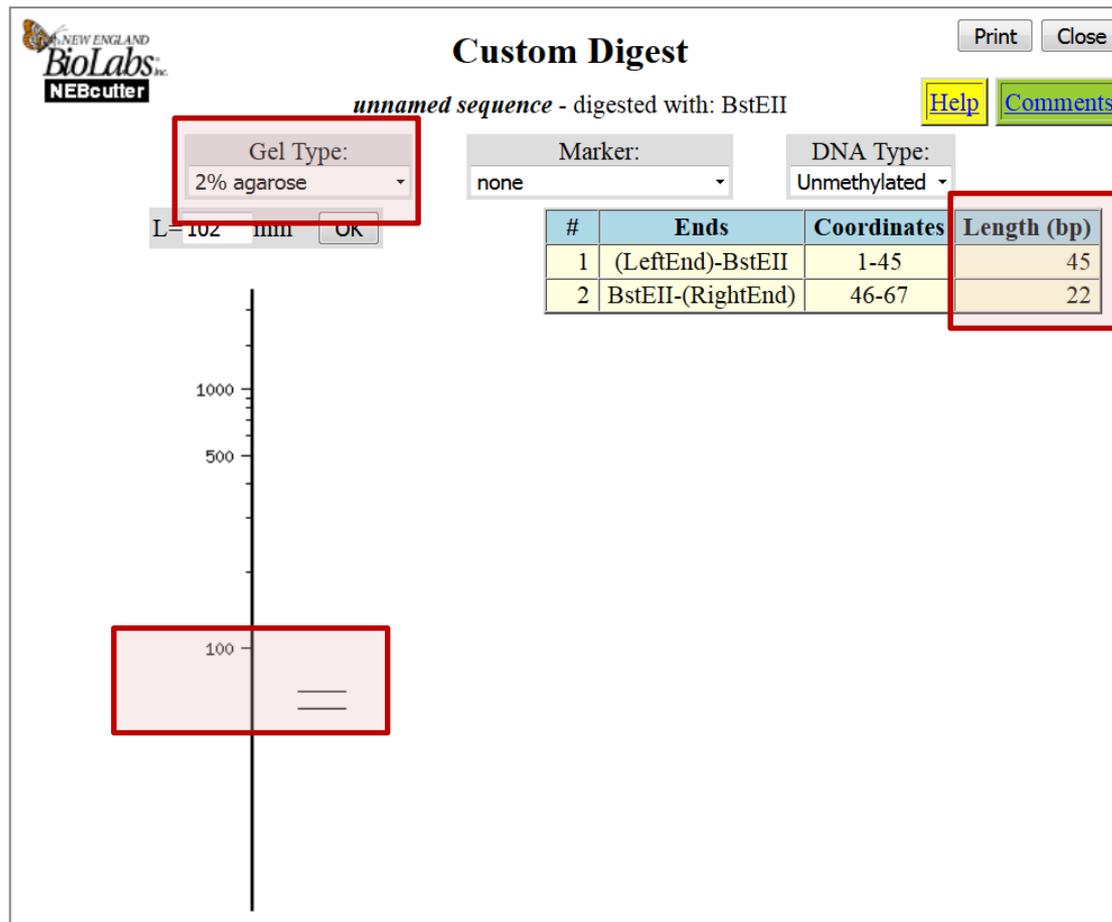
Zoom

Zoom in

More...

NEBcutter

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



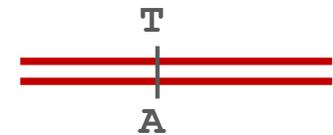
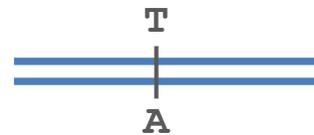
Example of application

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

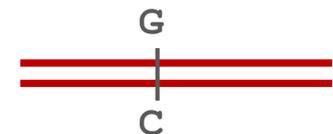
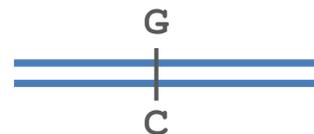
→ to determine the genotype of individuals = genotyping

- Assume there are 3 individuals with the following genotype at a certain SNP

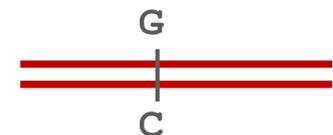
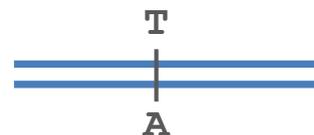
- One is homozygous for T



- 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

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!