

Explanation of the input and output files

Radboudumc SNPCheck

Input file

You can use a text editor such as Notepad or Excel to create your input file. The file consists of four columns that are explained below in more detail. An example is also given below. Please note that files should always be saved as tab delimited text file. Next to the upload you need to enter your name and e-mail address. Please make sure your e-mail address is correct because the results are send through mail.

Column 1 = Name amplicon

This is your own chosen unique name of the primer pair.

Column 2 = Sequence primer 1

Sequence of the first primer. Please note that primers should have a minimal sequence length of 16 bps.

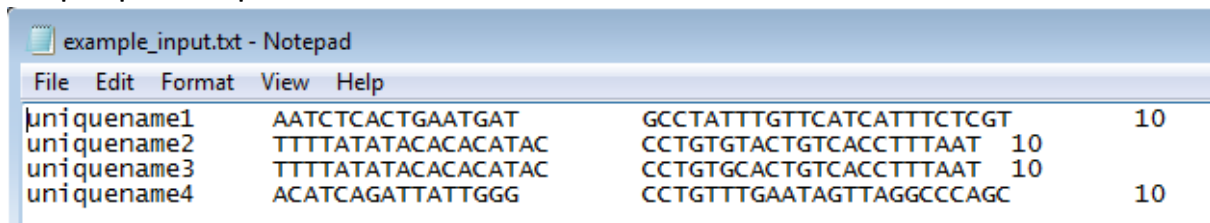
Column 3 = Sequence primer 2

Sequence of the second primer. Please note that primers should have a minimal sequence length of 16 bps.

Column 4 = Chromosome

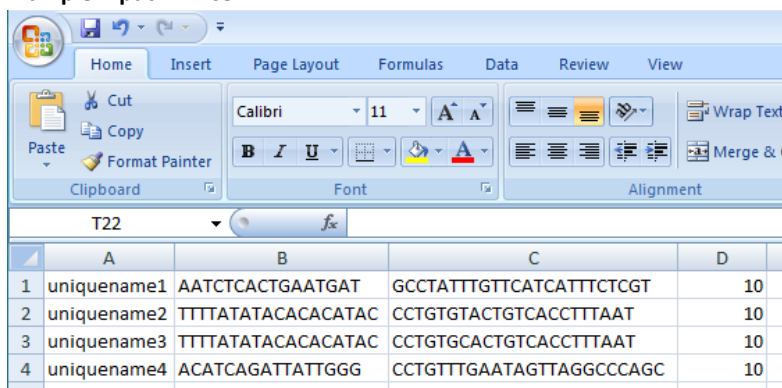
The chromosome you expect to blast to. For example you can use *chr1*, or just *1*. Valid chromosomes are chr1 until chr22, chrX and chrY

Example input in Notepad:



```
example_input.txt - Notepad
File Edit Format View Help
uniquename1      AATCTCACTGAATGAT      GCCTATTTGTTTCATCATTTCCTCGT      10
uniquename2      TTTTATATACACACATAC    CCTGTGTACTGTCACCTTTAAT      10
uniquename3      TTTTATATACACACATAC    CCTGTGCACTGTCACCTTTAAT      10
uniquename4      ACATCAGATTATTGGG      CCTGTTTGAATAGTTAGGCCCAGC      10
```

Example input in Excel:



	A	B	C	D
1	uniquename1	AATCTCACTGAATGAT	GCCTATTTGTTTCATCATTTCCTCGT	10
2	uniquename2	TTTTATATACACACATAC	CCTGTGTACTGTCACCTTTAAT	10
3	uniquename3	TTTTATATACACACATAC	CCTGTGCACTGTCACCTTTAAT	10
4	uniquename4	ACATCAGATTATTGGG	CCTGTTTGAATAGTTAGGCCCAGC	10

Output files

1. Blastn file

Primer Pair Name

Column 1 of the input file.

Primer 1

Column 2 of the input file.

Primer 2

Column 3 of the input file.

Mismatching bases Primer1

In case the primer has no mismatches and has at least one 100% hit to the chromosome that was given as input, a "0" is given as output. In case there is a mismatch, a different number than "0" is given and means there is at least 1 mismatch in the primer. However, when a number higher than 1 is given, it does not necessarily mean there are more than 1 mismatches (e.g. "4" can still mean there is only 1 mismatch in the primer). Note that the Radboudumc SNPCheck will search for the most likely position of the primer at the chromosome that is given as output. In case the primer aligns more than once at the chromosome, the position is chosen that results in the smallest in silico PCR-product." -1" Means that the Radboudumc SNPCheck cannot properly align the primer to the chromosome, and there is at least 1 mismatch in the primer.

Mismatching bases Primer2

See "Mismatching bases Primer 1".

100% Hits in genome Primer 1

Number of 100% hits of Primer 1 in the genome. The Radboudumc SNPCheck aligns the primer to the entire genome, irrespective the chromosome number that is given as input.

100% Hits in genome Primer 2

Number of 100% hits of Primer 2 in the genome. The Radboudumc SNPCheck aligns the primer to the entire genome, irrespective the chromosome number that is given as input.

Both primers have at least 2x a 100% hit in the genome (indication for a pseudogene)

When Primer 1 and Primer 2 both have more than 1 100% hit in the genome "true" is given as output. This might indicate that the given region has more than two copies in the genome. We recommend to perform a BLAST of the gene/region to ensure there is no pseudogene or other homologous region in the genome, to prevent amplification of unwanted regions. Note that when "false" is given, it does not exclude that a pseudogene or other homologous region is present in the genome.

100% Hits in genome AND same chromosome Primer 1

The number of 100% hits of Primer 1 on the chromosome number that is given as input.

100% Hits in genome AND same chromosome Primer 2

The number of 100% hits of Primer 2 on the chromosome number that is given as input.

Amplicon size (bp)

The calculated size of the amplicon based on the 100% aligned primers (and the chromosome number given as input). In case more than one PCR-product can form in silico, the smallest PCR-product is given.

PCR sizes

In case primers bind more than one time on the chromosome number that is given as input, more than 1 PCR product can form in silico. The sizes of potential PCR products is presented in an order from small to large.

Primers align at same strand

In case primers do not align on the same strand “false” is given (this is usually the case with primers designed for normal PCR-amplification). When both primers align at the same strand “true” is given as output (this is for example the case with Molecular Inversion Probes).

Other Hits in genome with 1 mismatch Primer 1

The number of hits of primer 1 in the genome with 1 mismatch.

Other Hits in genome with 1 mismatch Primer 2

The number of hits of primer 2 in the genome with 1 mismatch.

Nr SNPs found in primer 1

The number of SNPs found in primer 1. This number corresponds with the snpcheck-file. In case a SNP is found in dbSNP, Gnomadexomes and Gnomadgenomes, this will be counted as 3 SNPs.

Nr SNPs found in primer 2

The number of SNPs found in primer 2. This number corresponds with the snpcheck-file. In case a SNP is found in dbSNP, Gnomadexomes and Gnomadgenomes, this will be counted as 3 SNPs.

Chromosome

The chromosome number that is given as input.

Amplicon start

Genomic location of the start of the amplicon.

Amplicon end

Genomic location of the end of the amplicon.

Primer 1 Start

Genomic location of the start of primer 1.

Primer 1 End

Genomic location of the end of primer 1.

Primer 2 Start

Genomic location of the start of primer 2.

Primer 2 End

Genomic location of the end of primer 2.

2. SNPCheck file

Primer Pair Name

The name of the primer pair that is provided in the input file.

Primer

The primer that is provided in the input file and contains the respective SNP.

Sequence

The sequence of the primer. The nucleotide that is capitalized is the nucleotide that contains a SNP at the position. In case the variant is a deletion of several nucleotides, the respective nucleotides are all capitalized. In case of an insertion there will be no nucleotide capitalized. The location of the insertion can be derived from column “Position”.

Source

The source in which the SNP has been found. This can be either derived from the dbSNP database, or the Gnomad database which is divided into exome sequencing data (gnomAdExomes) and whole genome sequencing data (gnomAdGenomes). See <http://gnomad.broadinstitute.org/> and <https://www.ncbi.nlm.nih.gov/projects/SNP/> for more details.

SNP ID

The SNP ID that has been given to the respective SNP.

Position

The position of the SNP in the genome (Build GRCh37). This position also indicates the start of a deletion or position of an insertion.

Reference

The reference nucleotide(s) in the reference genome (Build GRCh37) given as it is in the plus strand, regardless whether the respective primer binds at the minus strand.

When this cell is empty it means there is an insertion with the variant that is given in column "Variant". The position of the insertion can be inferred from the column "Position".

Variant

The variant that has been found and differs with the reference genome (Build GrCh37). The respective nucleotide(s) are presented as they are annotated on the plus strand of the reference genome.

AF_MAX

The highest frequency in percentages of the variant found in one of the population (columns AF_AFR, AF_AMR, AF_ASJ, AF_EAS, AF_FIN, AF_NFE, AF_OTH, AF_SAS).

AF

The average allele frequency in percentages of the variant calculated over all populations. See table below for the total number of genomes and exomes that have been sequenced.

AF_AFR

The average allele frequency in percentages in the African/African American population. See table below for the total number of genomes and exomes that have been sequenced.

AF_AMR

The average allele frequency in percentages in the Admixed American population. See table below for the total number of genomes and exomes that have been sequenced.

AF_ASJ

The average allele frequency in percentages in the Ashkenazi Jewish population. See table below for the total number of genomes and exomes that have been sequenced.

AF_EAS

The average allele frequency in percentages in the East Asian population. See table below for the total number of genomes and exomes that have been sequenced.

AF_FIN

The average allele frequency in percentages in the Finnish population. See table below for the total number of genomes and exomes that have been sequenced.

AF_NFE

The average allele frequency in percentages in the Non- Finnish European population. See table below for the total number of genomes and exomes that have been sequenced.

AF_OTH

The average allele frequency in percentages in Other populations (populations not assigned). See table below for the total number of genomes and exomes that have been sequenced.

AF_SAS

The average allele frequency in percentages in Other populations (populations not assigned). See table below for the total number of genomes and exomes that have been sequenced.

Chromosome

The chromosome number the primer is aligned to (this will always be the same chromosome number that is given as input).

Primer Start

The genomic position of the start of the primer.

Primer End

The genomic position of the end of the primer.

POPULATION	DESCRIPTION	GENOMES	EXOMES	TOTAL
AFR	African/African American	4,368	7,652	12,020
AMR	Admixed American	419	16,791	17,210
ASJ	Ashkenazi Jewish	151	4,925	5,076
EAS	East Asian	811	8,624	9,435
FIN	Finnish	1,747	11,150	12,897
NFE	Non-Finnish European	7,509	55,860	63,369
SAS	South Asian	0	15,391	15,391
OTH	Other (population not assigned)	491	2,743	3,234
	Total	15,496	123,136	138,632

Source: <http://gnomad.broadinstitute.org/>