

ExonSurfer: User Guide

Elena Cristina Rusu*
Pablo Monfort Lanzas†

April 11, 2023

*Email: ecrisru@alumni.uv.es

†Email: pablo.monfort@i-med.ac.at

Contents

1 Overview of ExonSurfer and its purpose	3
2 ExonSurfer's workflow	4
2.1 Workflow overview	4
2.2 Target selection	5
2.3 Primer design	5
2.4 Specificity checks	5
2.5 Primer filtering	6
3 User Interface	7
3.1 Input information	7
3.2 Job Status	10
3.3 Results	11
4 Customizing parameters	17
5 Troubleshooting	19
5.1 Why are my primers not being designed?	19
5.2 Contact support	19
6 Acronyms	20
7 Software tools and database versions	21

1 Overview of ExonSurfer and its purpose

ExonSurfer is a web-based tool designed to assist researchers in designing primers for transcript-specific detection and avoiding genomic DNA or hnRNA contamination during reverse transcription uniplex PCR experiments. The tool combines various steps of primer design processes, including target region selection, avoiding of common polymorphisms, and checks for specificity and self-complementarity. ExonSurfer selects primer pairs spanning exon-exon junctions and automatically verifies the primers' specificity using BLAST. The user can customize the primer length, thermal parameters, and maximum amplicon length to suit their amplification procedure. ExonSurfer selects the most specific exon-exon junction for the input target(s), enabling highly accurate and efficient primer design for a wide range of applications. Its purpose is to provide researchers with a reliable and user-friendly tool that streamlines the primer design process and enhances the accuracy and specificity of PCR experiments.

ExonSurfer supports primer design for *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*, for Ensembl's database.

The command line version for ExonSurfer can be accessed at <https://github.com/CrisRu95/ExonSurfer>, and this web tool can be inspected at <https://github.com/pamonlan/ExonSurferWeb>.

2 ExonSurfer's workflow

2.1 Workflow overview

ExonSurfer's workflow consists of 4 modules, as shown in Figure 1.

1. **Target selection:** The user inputs the desired species, gene symbol, and target transcript(s). ExonSurfer chooses the optimal exon junctions depending on the target transcript(s) the user selects. For the human genome, the sequence has polymorphisms with a minor allele frequency larger than 1% masked (optionally).
2. **Primer design:** We use primer3-py¹ to design multiple primer pairs, in order to filter them in the following steps and report the most specific ones. The design can be done either overlapping or flanking the exon junction.
3. **Specificity checks:** Using BLAST², all the primers are aligned against all known transcripts of the selected species and possible off-target amplification is annotated. A second BLAST is performed against genomic DNA in order to ensure transcript-specificity even in the case of DNA contamination.
4. **Primer filtering and reporting:** Only the most specific primers are returned, sorted by the pair score.

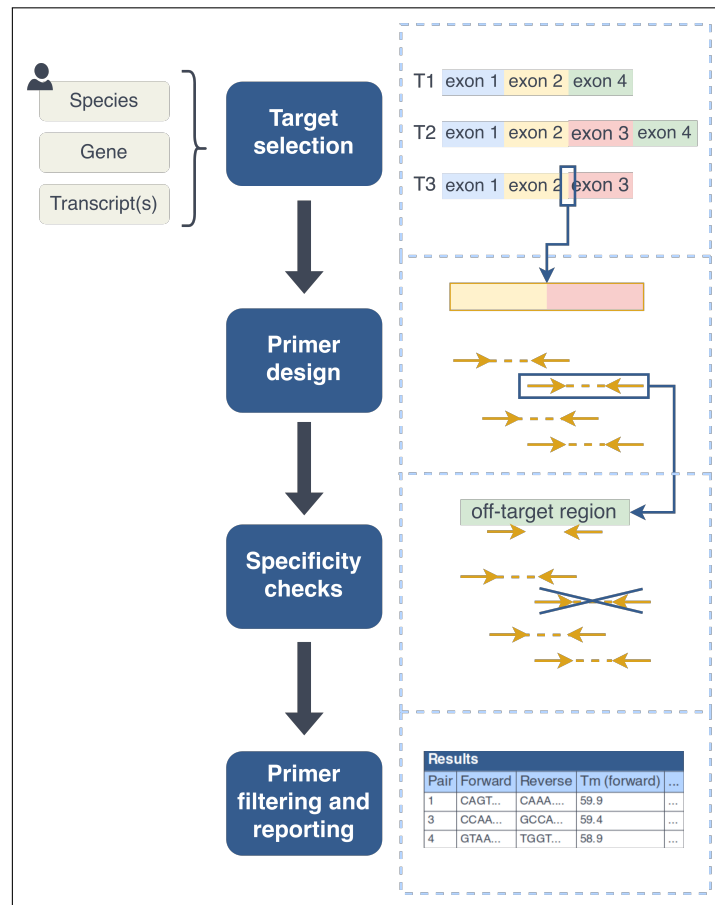


Figure 1: **ExonSurfer workflow.**

¹Untergasser A, Cutcutache I, Koressaar T, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115. doi:10.1093/nar/gks596

²Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. Published 2009 Dec 15. doi:10.1186/1471-2105-10-421

2.2 Target selection

ExonSurfer uses a unique algorithm to determine the best exon junctions on which to design primers. The first step is to build, for each transcript, a list of consecutive exon junctions, where each junction is depicted by a list of ensembl exon identifiers. The number of exons used for each junction is dependant on exon length and the optimum amplicon length specified by the user (200bp by default). These junctions will be further filtered, depending on whether we want the most conserved or the most unique ones. Note that this selection process is done taking into account the exon identifiers. When we target all transcripts from a gene, ExonSurfer will search for the most conserved exon junction among the protein coding transcripts. In many cases, there is no exon junction ubiquitous among all transcripts, and in these case the junctions present in the canonical transcript, that cover the most other transcripts, are prioritized. When we target one or more specific transcripts, ExonSurfer will search for a junction present among all our targets and not present in other transcripts from the same gene. This is often impossible, so ExonSurfer settles for the exon junctions present in all or most of the targets, as a first priority. Secondly, and if possible, it tries to exclude exon junctions that target a large number of not selected transcripts. For *homo sapiens*, ExonSurfer avoids placing primers on polymorphisms with a MAF (minor allele frequency) higher than 1% in any population group. For this, we extracted the SNPs from dbSNP (NCBI) and masked their locations in the chromosome sequences. This reduces the likelihood of a reduced amplification efficiency due to mismatches with SNPs.

2.3 Primer design

Primers are designed using primer3-py using the configuration (primer, amplicon and PCR parameters) provided by the user. ExonSurfer can place the primers on the junctions in two distinct manners (Figure 2). The preferred option is for one of the primers to overlap with the exon junction, amplifying in this way only the cDNA. However, for cases when the experimental setup ensures low or no DNA contamination, ExonSurfer can optionally design primers that flank the exon junction, ensuring a wider variety of oligos.

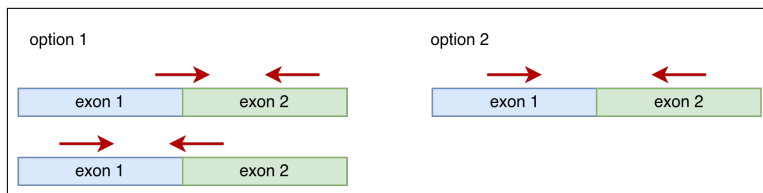


Figure 2: **Design option.** Primers are placed on the exon junction and flanking it. Primer sequences are marked as red arrows.

ExonSurfer uses the default values from primer3-py for discarding primer pairs that form hairpin or dimer-like structures.

2.4 Specificity checks

ExonSurfer uses BLAST in order to map the designed primers to the transcripts from the selected genome. The alignments are filtered according to the E-value and identity values.

For every primer pair, ExonSurfer counts the number of alignments each primers has to genes different from the target. This value is returned in the results table (Fig. 6) as "Individual Alignment Score".

If primers map to the same target, in opposite strand, and close to one another, they lead to off-target amplification. ExonSurfer assesses this event for every possible combination within a primer pair (forward with reverse, reverse with reverse or forward with forward) and annotates the possible off-targets. ExonSurfer classifies as off-target:

- Any possible amplification in a different gene.

- Any amplification in transcripts (of the target gene) not select by the user. This category is not taken into account with the option "ALL". If the number of any of the above categories is higher than 0, we also annotate the number of protein coding off-targets.

2.5 Primer filtering

ExonSurfer takes a flexible approach when filtering the final primers pairs. For some genetic regions it is feasible to expect primers with no alignments to other genes, but for other regions this is hardly possible (e.g. CYP2D6). This flexible approach ensures that the algorithm is as strict as possible for each designed region.

The filtering parameters (FP), sorted according to their importance, are:

- **FP1:** Number of possible off-target amplicons present in different genes.
- **FP2:** Number of protein coding off-target amplicons present in different genes.
- **FP3:** Number of possible off-target amplicons present in different transcripts (only when specific transcripts are selected).
- **FP4:** Number of protein coding off-target amplicons present in different transcripts (only when specific transcripts are selected).
- **FP5:** Number of alignments each primer pair has to other genes (independent of off-targets).
- **FP6:** Type of design (Fig. 8), where option 1 is preferred.

The filtering is done according to the parameter importance. Ideally, for specific transcripts we would have $FP[1-5] = 0$ and $FP6 = 1$, while for "ALL" transcript design type we would disregard FP3 and FP4. However, this level of strictness does not adapt well to distinct genomic regions. Therefore, primer pairs are filtered in a "sorted" fashion, where ExonSurfer tries to keep the $FP1 = 0$ pairs first (independently of the values of the subsequent filters); if this is not possible it disregards FP1 and tries to keep the $FP2 = 0$ primer pairs; if this is possible, it further tries to keep $FP3 = 0$ primer pairs, and so on.

The score reported is a normalization of primer3's pair penalty. For more information on its calculation, please visit this manual.

3 User Interface

In this section we will review how to navigate through the user interface.

3.1 Input information

ExonSurfer requires as input (a) a gene symbol or (b) a genbank or fasta formatted file, as shown in Figure 3.

About ExonSurfer

ExonSurfer is a web-tool for designing high-specificity exon junction binding transcript primers for qPCR, using the Ensembl database and the blastn algorithm. It combines all the steps of transcript-specific primer design, such as target selection to exclude other transcripts of the same gene, avoiding common polymorphic regions, and enables specificity and self-complementarity verification.

ExonSurfer simplifies the primer design process, while providing the user with a variety of options to customize the design to best suit their needs. In conclusion, this tool ensures high specificity and low non-specific binding of the primers, making ExonSurfer an essential tool for transcript-specific detection.

Design Primers Database section

Create now your primers for your favorite species and gene using ExonSurfer.

The program offers fast and accurate primer design, with user specified parameters like primer size, GC content, etc, and a user-friendly web interface to setup search options and visualize results.

Select species

1 → Homo sapiens

Select human gene symbol

2 → Write a human gene...

☒ Use Homo sapiens masked genome

3 → Select Transcript & Primer Parameters

Upload Your Own Transcript Files File upload section

ExonSurfer supports the use of custom transcript information for primer design. If your gene or transcript of interest is not available in the Ensembl database, you can upload your own FASTA or GenBank file to create primers using ExonSurfer. Ensure that the files you upload contain accurate and complete sequence information, as this will impact the primer design results.

Upload a Sequence File

Upload your own GenBank or Fasta format to use with ExonSurfer.

4 → Upload

Figure 3: **Figure 3: ExonSurfer input.** For the database input (green area) you should choose a species (mark "1"), write a gene symbol (mark "2"), and click on "Select Transcript Primer Parameters" (mark "3"). For the file upload input (orange area), you should click on "Upload" (mark "4").

Gene symbol and transcript list

In order to start primer design, you need to select the species and write a gene symbol, then press the "Select Transcript Primer Parameters". This leads you to the page shown in Figure 4, where you obtain a view of the transcripts and exon placement in the chromosome (mark 1), a transcript table (mark 2) where you need to choose your selected targets, the different design parameters (mark 3) (more details on the parameters in section 4) and the "Design Primers" button (mark 4).

Assign design parameters

Transcript Selection

	Species	Symbol	Ensembl ID
Summary	Homo sapiens	VCAM1	ENSG00000162692

Select transcript from gene to design primers

Transcript ID	Transcript Name	Transcript Biotype
<input type="checkbox"/> ENST00000370115	VCAM1-203	protein_coding
<input type="checkbox"/> ENST00000294728	VCAM1-201	protein_coding
<input type="checkbox"/> ENST00000347652	VCAM1-202	protein_coding
<input type="checkbox"/> ENST00000370119	VCAM1-204	protein_coding
<input checked="" type="checkbox"/> ALL		

Primer Parameters

Amplicon Parameters

PCR Parameters

BLAST Parameters

Design Primers

Figure 4: **Transcript selection.** You can check the transcript isoforms from your gene at mark "1". Choose your target(s) in the green area (mark "2"), select the different design parameters at mark "3" and press "Design Primers" (mark "4").

File input

Alternatively, you can also design primers for transcripts not present in Ensembl's database. You can do this by providing a file in any of these two formats: (a) GenBank Flat File Format or (b) FASTA format. This is shown in Figure 5, where you upload a fasta or genbank file by clicking the "Browse" button (area 1), and can modify the design parameters.

Custom Transcript Selection

Select species
Homo sapiens

Upload file
Browse... No file selected.

Indicating exon-exon junctions
Exon-Exon junctions must be indicated in the fasta header using the format:
>SEQUENCE_NAME POS_1 POS_2 POS_3

Download sample files
Click to download sample files in [fasta](#) or [genbank](#) format to check the required header format.

Primer Parameters

Amplicon Parameters

PCR Parameters

BLAST Parameters

Design Primers

Figure 5: **File upload.** Select the target species and upload a genbank or fasta file (mark "1") and choose your design parameters (mark "2").

The GenBank formatted file should include CDS records that indicate exon sequences. The FASTA format file should include the junction positions in the header, separated by a space and after the target name; for e.g., >TARGET_NAME 100 250. Please take into account that these are positions (and not exon length), so in the example the target is formed by 3 exons of 100bp, 150bp and unknown bp.

Design parameters

You can change the design parameters for the primers, amplicon, PCR and BLAST. To toggle the parameters formulars, just click on the light blue areas shown in Figure 6.

Primer Parameters

Parameter	Min	Opt	Max
GC content (%)	20	50	80
Melting temperature (°C)	57	60	63
Length (bases)	18	20	27

Number of primer pairs to design: ☐ High (1500 pairs) ☒ Middle (200 pairs) ☐ Low (50 pairs)

Primer junction design: ☒ Primers spanning exon junctions ☐ Primers spanning or flanking exon junctions

Minimum 3' overlap with junction: 5

Minimum 5' overlap with junction: 6

Amplicon Parameters

PCR Parameters

BLAST Parameters

Design Primers

Figure 6: **Design parameters.** You can modify primer, amplicon PCR and BLAST parameters.

For a detailed information on the parameters descriptions, please check section 4.

3.2 Job Status

Once you click the "Design Primers" button from Figure 6, you get the running job information. Typically, no design requires more than a 2-3 minutes, but this could change in case of high server usage. Click on "Copy URL" and save the link if you want to be able to close the page and revisit you design results later.

Job Status Information

Session Information

Session URL
[Copy URL](#) to check results later

Job Information

ID
23644653-fc8f-41b1-bc2b-48236e84af60

Status
running

Description

Time (s)
37

Estimated time (s)
160

Progress

23.3%

Gene Information

Symbol
AAAS

Ensembl ID
[ENSG00000094914](#)

Species
Homo sapiens

Transcript Information

Selected transcripts
ALL

Check

Cancel

Figure 7: **Job status.** Information on the running job, with estimated time.

3.3 Results

Main results page

Once the design is complete, you are redirected to the main results page (Figure 8 and Figure 9), that include the sections of "Query Summary", "Top primer pairs" and "Primer results".

In Figure 8, if you click in the gene link from the Query Summary (mark 1) you are redirected to Ensembl's gene page. In the section "Top primer pairs" you can get a view of the cDNA and amplicon sequence by clicking on mark 2 (this leads you to the view discussed in Figure 10 and Figure 11) and you can download the information of the specified primer pair by clicking on mark 3.

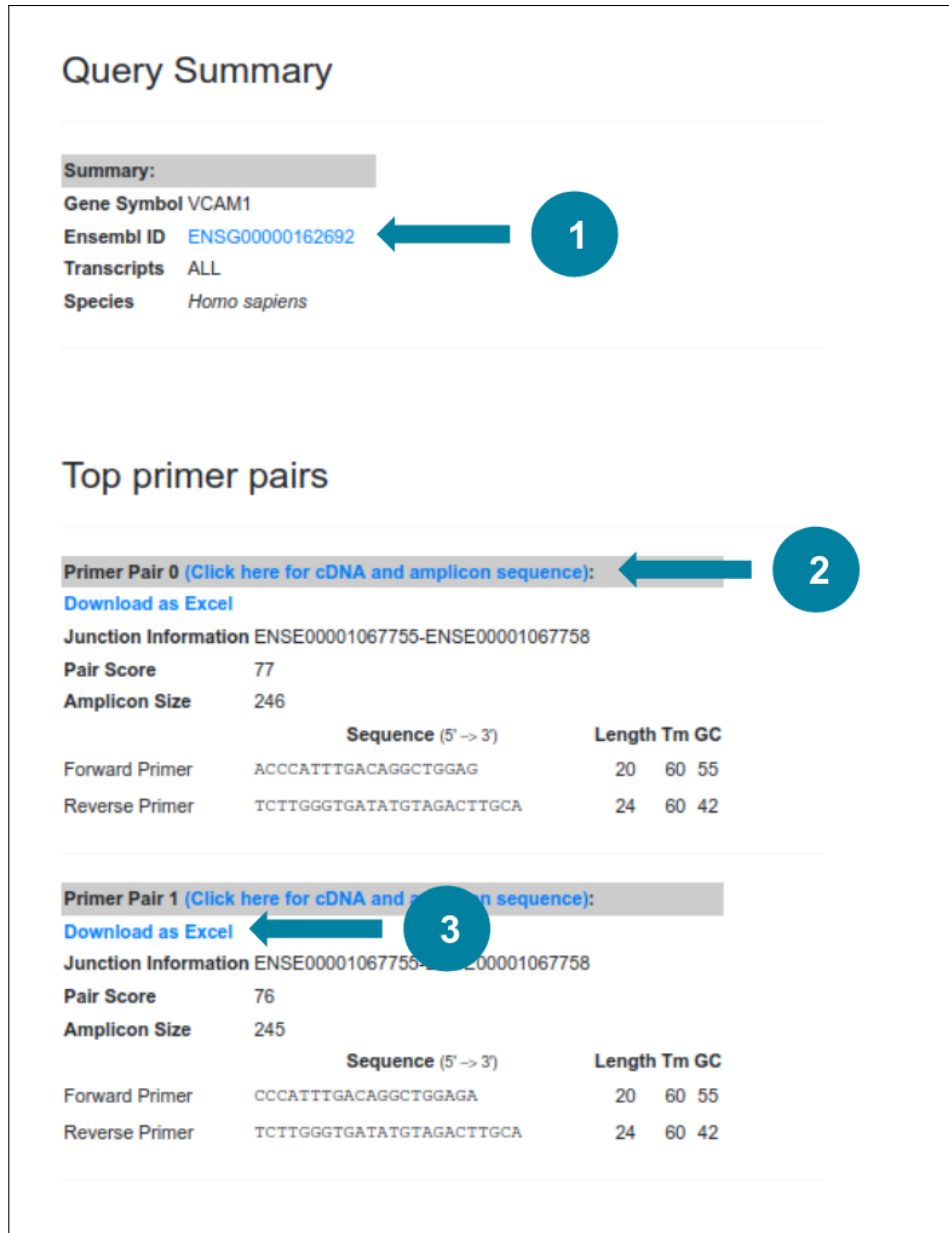


Figure 8: **Main results (part 1)**. Clicking on mark "1" leads you to Ensembl's page on the specified gene, "2" leads you to the view from Figure 10 and "3" downloads the information of the primer pair.

In Figure 9, you can download the table with all the design primer pairs in CSV, Excel or TSV format. Clicking on a specific primer pair number (first column) leads you to the view from Figure 10 and Figure 11. The primer pairs table contains the following columns:

- **Primer Pair:** Pair identifier.
- **Forward Primer:** Forward primer sequence.
- **Reverse Primer:** Reverse primer sequence.
- **Amplicon Size:** Product size (with primer length included).
- **Amplicon Tm:** Amplicon melting temperature, as calculated by primer3-py.

- **Forward Tm:** Forward primer melting temperature, as calculated by primer3-py (using the provided PCR conditions).
- **Reverse Tm:** Reverse primer melting temperature, as calculated by primer3-py (using the provided PCR conditions).
- **Forward GC:** Forward primer GC percentage.
- **Reverse GC:** Reverse primer GC percentage.
- **Individual Alignment Score:** Number of times the primers align to transcripts from other genes (sum of forward and reverse alignments). The closer to 0, the better.
- **Detected Transcripts:** Transcripts of the target gene that the primer pair amplifies.
- **Not Detected Transcripts:** Transcripts of the target gene that the primer pair does not manage to amplify.
- **Pair Score:** Primer pair scoring according to optimal primer length, GC and Tm values, as well as dimer and hairpin interactions. The higher, the better the primer pair is (less dimers and hairpins, closer to optimal values).

Primer Pair 4 ([Click here for cDNA and amplicon sequence](#)):

[Download as Excel](#)

Junction Information ENSE00001067755-ENSE00001067758

Pair Score 76

Amplicon Size 244

	Sequence (5' → 3')	Length	Tm	GC
Forward Primer	ACCCATTGACAGGCTGGAG	20	60	55
Reverse Primer	TTGGGTGATATGTAGACTTGCA	22	57	41

Primer Results

Filtered primers table

CSV

Excel

TSV

Show 10 entries

Search:

Primer Pair	Forward Primer	Reverse Primer	Amplicon Size
Pair32	AATTGCAAGTCTACATATCACCCA	TCTCTGGAGCTGGTAGACCC	120
Pair37	GAATTGCAAGTCTACATATCACCCA	CTCTGGAGCTGGTAGACCCT	120
Pair38	GAATTGCAAGTCTACATATCACCCA	TCTCTGGAGCTGGTAGACCC	121
Pair41	GGAAGTCCCTGGAAACCAAGA	TCTTGGGTGATATGTAGACTTGCA	165
Pair42	AGGAAGTCCCTGGAAACCAAG	TCTTGGGTGATATGTAGACTTGCA	166
Pair51	AGGAAGTCCCTGGAAACCAA	TCTTGGGTGATATGTAGACTTGCA	166
Pair57	TGCAAGTCTACATATCACCCAAGA	TGCATTCCAGAAAGGTGCTG	173
Pair60	AGAATTGCAAGTCTACATATCACCC	CTCTGGAGCTGGTAGACCCT	121
Pair61	AGAATTGCAAGTCTACATATCACCC	TCTGGAGCTGGTAGACCCTC	120
Pair62	AGAATTGCAAGTCTACATATCACCC	TCTCTGGAGCTGGTAGACCC	122

Primer Pair

Forward Primer

Reverse Primer

Amplicon Size

Showing 1 to 10 of 308 entries

Previous

1

2

3

4

5

...

31

Next

Figure 9: **Main results (part 2)**. Includes the table with all the final primer pairs.

Pair-specific results page

The "cDNA and amplicon" results page is specific to each primer pair. In this, you have the specific information of the detected transcripts, primer information, amplicon sequence, and transcript and exon overview, depicted in Figure 10 and Figure 11

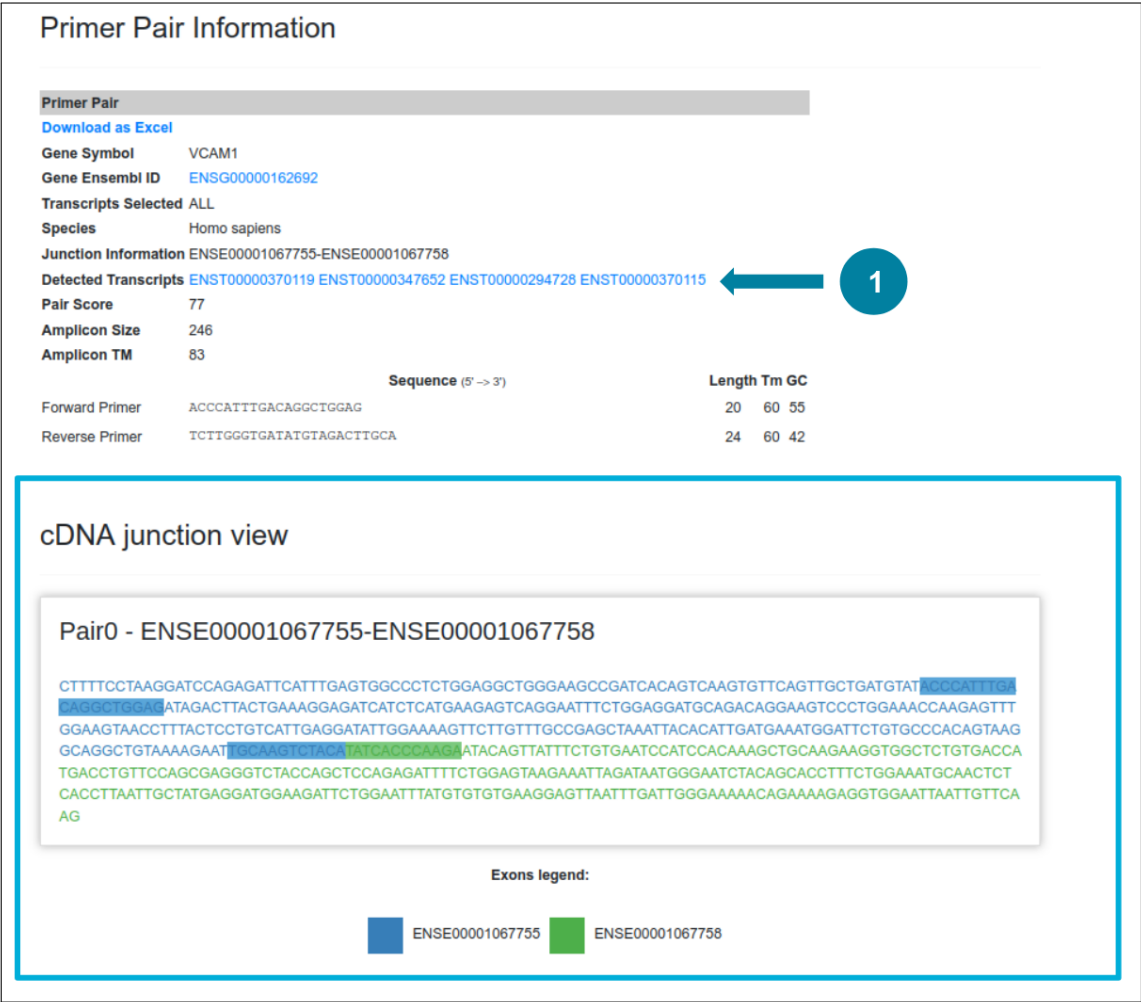


Figure 10: **cDNA and amplicon sequence (part 1)**. Clicking on mark "1" leads you to the Ensembl's page for each of the detected transcripts for that primer pair. The amplicon sequence can be extracted from the cDNA junction view, where each exon is a different color and primers are highlighted.



Figure 11: **cDNA and amplicon sequence (part 2)**. Detected isoforms are colored in green, non-detected isoforms are kept grey, and exons where the primers align to are marked in red.

Warnings

If a primer pair has any off-target amplification, a warning is issued (Figure 12) and we can check the off-target sequences by clicking on it (Figure 13).

It must be noted that, if you choose a specific transcript, all the other transcripts from the same gene will be treated as off-target amplification.

Primer Pair 13 ([Click here for cDNA and amplicon sequence](#)):

[Download as Excel](#)

Junction Information ENSE00003634695-ENSE00003556115

Pair Score 100

Amplicon Size 121

	Sequence (5' → 3')	Length	Tm	GC
Forward Primer	TGAGGAAGCGAGGGTCGTC	19	61	63
Reverse Primer	TTCGCCAACCACCTCCGGAC	19	63	63

Warning possible inespecificities. Check **off-target** [here](#).

Figure 12: **Possible off-target amplification**. A warning is issued if the primer pair has any predicted amplification outside of the target sequence(s).

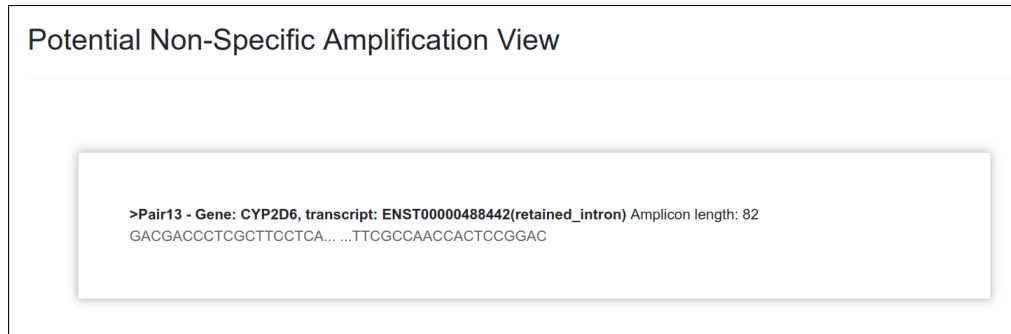


Figure 13: **Off-target amplification sequence.** If there are any mismatches between the primer sequences and the off-target, these will be marked in red.

4 Customizing parameters

Primer parameters

Users can modify the following primer parameters:

Primer length Values between 18-25 are advised. If the target gene has extremely GC rich regions, lower the minimum length to 17. On the contrary, if the region is very poor in GC, primers can be as long as 36 bases, though more than 30 are not advised.

GC content (%) We advise keeping this parameter with the loose default values (20-80), since primer length and melting temperature control it. In addition, it is taken into account to calculate the "primer pair score".

Melting temperature (°C) Primer melting temperature, calculated with the tables of thermodynamic values and the method published in "A unified view of polymer, dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics".

Number of primer pairs to design Total number of primer pairs to design before we do the specificity checks. The more primers, the larger execution time is going to be.

Primer junction design To choose between only designing with a primer in an exon junction or to allow primers also to flank junctions. The second option is preferred when there is no DNA contamination in the sample.

Minimum 3' overlap with junction Minimum number of bases of the 3' end of the primer that need to be in one side of the junction. In order to ensure junction specificity, it is advisable to keep this value lower than "Minimum 5' overlap with junction".

Minimum 5' overlap with junction Minimum number of bases of the 5' end of the primer that need to be in one side of the junction.

Amplicon parameters

Product size (bp) Length of the amplicon (primer sequences included) to generate. Typical values for qPCR are between 120 - 250bp.

Melting temperature (°C) Melting temperature of the product. Calculated using equation (iii) from the paper "Optimization of the annealing temperature for DNA amplification in vitro". This parameter is usually at fault when primers cannot be designed for a specific region (specially in rich GC genes). However, the PCR setup should be adapted if product T_m is higher than 90 °C.

PCR parameters

These parameters are used as specified in the primer3 manual.

Salt Divalent (mM) The millimolar concentration of divalent salt cations in the PCR.

Salt Monovalent (mM) The millimolar (mM) concentration of monovalent salt cations (usually KCl) in the PCR.

dNTP Concentration (mM) The millimolar concentration of the sum of all deoxyribonucleotide triphosphates.

BLAST parameters

Maximun E-value The Expect value is the number of hits that are expected by chance when search a database of a particular size. The smaller the E-value, the smaller is the possibility of finding the sequence in the database by chance (i.e., the smaller the E-value, the more meaningfull the hit is).

For more information, please refer to BLAST's frequently asked questions. Hits with a larger E-value than the specified will not be reported nor used in the off-target assessment.

Minimum percentage of identity (%) The minimum identity to report a hit is the minimum percentage of base pairs that are the same between the query sequence (primer) and the subject sequence (specific mRNA).

Maximum off-target size (bases) The maximum distance between two hits in order to consider them as an off-target.

5 Troubleshooting

5.1 Why are my primers not being designed?

The design parameters are usually at fault for unsuccessful designs. Here we report a list of the most common issues and how to solve them.

- **Excessive amplicon length** Take into account that if your target gene has a very short coding sequence, the amplicon length parameters should be adjusted accordingly.
- **Amplicon T_m outside the expected length** For genes with very extreme GC contents, there might be difficult to design amplicons with a T_m lower than 90 °C. However, take into account that PCR setup should adapt to amplicons with T_m higher than 90 °C.
- **Primer length outside the expected length** Also for genes with very extreme GC contents, you might need to accept primers as long as 32-34 nucleotides, or as short as 17.

5.2 Contact support

If you are having issues with your specific design and these have not been solved by the previous specifications, or you are finding any other issue within the application, please contact pablo.monfort@i-med.ac.at or ecrisru@alumni.uv.es.

6 Acronyms

cDNA Complementary DNA

MAF Minor allele frequency

qPCR Quantitative polymerase chain reaction

SNP Single nucleotide polymorphism

7 Software tools and database versions

Software or database	Version	Usage
primer3-py	1.2.0	Design primers and check of self-complementarity
pyensembl	Ensembl 108	Extract gene and transcript information
BLAST	2.12.0+	Specificity checks
Genome sequences	Ensembl 108	Design primers and specificity checks
dbSNP (NCBI)	build 154	Mask human polymorphisms