PaintSHOP User Guide

PaintSHOP can be accessed at paintshop.io

Revised: 15 April 2021 (latest version: PDF | HTML)

Contents

1 RNA Probe Design

1.1 Overview of RNA probe design

The RNA Probe Design tab allows you to enter RefSeq annotations in order to retrieve the probes for your target or list of targets. You can either enter the annotations manually into the text box provided, or upload a file with one annotation per line. If entering annotations manually, please separate each entry with a comma and a space, as shown in the example entry. Once you have entered your annotations, select which input option you used, and press the Submit button to retrieve your probes.
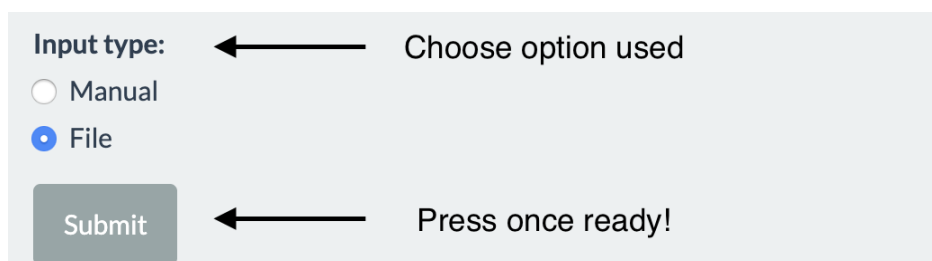
Figure 1: Screenshot of the PaintSHOP target input interface.

PaintSHOP will return a density plot of probe coverage and a dynamic table with your probes. The density plot is intended to provide a visual estimate of how well covered your annotations are. RNA probe tables have the following columns:

Isoform-resolved RNA probe sets:

- **refseq:** The transcript ID of the transcript that the probe targets, stripped of any version suffixes e.g. NM_001180043
- **chrom:** the chromosome of the probe sequence
- **start:** the start coordinate of the probe sequence
- **stop:** the stop coordinate of the probe sequence
- **sequence:** The DNA sequence of the oligo probe
- **Tm:** The melting temperature of the probe sequence
- **on_target:** The on-target score generated by the Homology Optimization Pipeline
- **off_target:** The off-target score generated by the Homology Optimization Pipeline
- **repeat_seq:** Whether or not the probe sequence contains bases flagged as repetitive by RepeatMasker. 0 = False, 1 = True
- **max_kmer:** The maximum number of occurences of all 18-mers in the probe sequence in the genome it targets
- **probe_strand:** The strand orientation of the probe sequence. Plus (+) or minus (-)
- **transcript_id:** The unmodified transcript ID of the transcript that the probe targets e.g. NM_001180043.1
- **gene_id:** The gene ID of the gene whose transcript the probe targets e.g. PAU8

Isoform-flattened RNA probe sets:

- **refseq:** The gene ID of the gene whose transcript the probe targets e.g. PAU8
- **chrom:** the chromosome of the probe sequence
- **start:** the start coordinate of the probe sequence
- **stop:** the stop coordinate of the probe sequence
- **sequence:** The DNA sequence of the oligo probe
- **Tm:** The melting temperature of the probe sequence
- **on_target:** The on-target score generated by the Homology Optimization Pipeline
- **off_target:** The off-target score generated by the Homology Optimization Pipeline
- **repeat_seq:** Whether or not the probe sequence contains bases flagged as repetitive by RepeatMasker. 0 = False, 1 = True
- **max_kmer:** The maximum number of occurences of all 18-mers in the probe sequence in the genome it targets
- **probe_strand:** The strand orientation of the probe sequence. Plus (+) or minus (-)
- **transcripts:** The number of isoforms that this probe targets

Note: If your target is on the + strand, PaintSHOP will automatically return the - strand probe sequence. This is to ensure your FISH experiment works the way you want!

Also note that the table can be searched, resized, and paged through. For more information about the advanced settings and set balancing features, please read those topic descriptions.

1.2   PaintSHOP RNA probe sets

newBalance (RNA)

probe sets for the hg38, hg19, mm10, mm9, dm6, ce11, danRer11, TAIR10, and sacCer3 (new: rn6, galGal5, galGal6) genomes with a length window of 30-37 nucleotides and a Tm of window of 42-47 degrees. These parameters were selected to optimize probe coverage and hybridization. For more information on these new probe sets, please refer to the PaintSHOP manuscript.

OligoMiner (RNA)

the 'balance' probe sets generated by OligoMiner for the hg38 and hg19 reference genomes. These probes have a length window of 35-41 nucleotides, and a Tm window of 42-47 degrees. For more information on these probes, please refer to the OligoMiner manuscript.

2012 Oligopaints (RNA)

the original Oligopaints genome-scale probe set from the Beliveau et al. 2012 PNAS publication. The probes have a length of 32 bases, and have an approximate Tm window of 34-42 degrees. For more information on this probe set, please refer to the 2012 Oligopaints publication.

iFISH4U (RNA)

the full 40-mer probe set from iFISH4U.

All probe sets in the RNA Probe Design tab have been intersected with the RefSeq for their respective reference genomes in order to provide quick retrieval of probes for annotations. To retrieve probes target regions outside of these annotations, please use the DNA Probe Design tab.

We have now added a new set of 'isoform flattened' probe sets for RNA FISH probe design. These 'isoform flattened' annotation sets prioritize shared exonic sequence between isoforms (Methods) in order to maximize the chance of detection and only modestly reduce the coverage of the transcriptome when used for probe intersects. These sets exist with the newBalance probes for the hg38, hg19, mm10, dm6, ce11, danRer11, TAIR10, and sacCer3 (new: rn6, galGal5, galGal6) reference genomes.

2   DNA Probe Design

2.1   Overview of DNA probe design

The RNA Probe Design tab allows you to enter genomic coordinates in order to retrieve the probes for your target or list of targets. You can either enter the coordinates manually into the text box provided, or upload a file with the format shown in the example. If entering coordinates manually, please separate each entry with a comma and a space, as shown in the example entry. Once you have entered your annotations, select which input option you used, and press the Submit button to retrieve your probes.

PaintSHOP will return a density plot of probe coverage and a dynamic table with your probes. The density plot is intended to provide a visual estimate of how well covered your targets are. The table has the following columns:

- **chrom:** the chromosome of the probe sequence
- **start:** the start coordinate of the probe sequence
- **stop:** the stop coordinate of the probe sequence
- **sequence:** The DNA sequence of the oligo probe
- **Tm:** The melting temperature of the probe sequence
- **on_target:** The on-target score generated by the Homology Optimization Pipeline
- **off_target:** The off-target score generated by the Homology Optimization Pipeline
- **repeat_seq:** Whether or not the probe sequence contains bases flagged as repetitive by RepeatMasker. 0 = False, 1 = True
- **max_kmer:** The maximum number of occurences of all 18-mers in the probe sequence in the genome it targets
- **probe_strand:** The strand orientation of the probe sequence. Plus (+) or minus (-)

Note: If you specify probe strand (+ or -) in either the manual or file entry, PaintSHOP will return the probe in the same orientation as what you enter. For example specifying + will return +. You are specifying the probe strand, not the target strand.

Also note that the table can be searched, resized, and paged through. For more information about the advanced settings and set balancing features, please read those topic descriptions.

2.2   PaintSHOP DNA probe sets

newBalance (DNA)

probe sets for the hg38, hg19, mm10, mm9, dm6, ce11, danRer11, TAIR10, and sacCer3 (new: rn6, galGal5, galGal6) genomes with a length window of 30-37 nucleotides and a Tm of window of 42-47 degrees. These parameters were selected to optimize probe coverage and hybridization. For more information on these new probe sets, please refer to the PaintSHOP manuscript.

OligoMiner (DNA)

the 'balance' probe sets generated by OligoMiner for the hg38 and hg19 reference genomes. These probes have a length window of 35-41 nucleotides, and a Tm window of 42-47 degrees. For more information on these probes, please refer to the OligoMiner manuscript.[1]

2012 Oligopaints (DNA)

the original Oligopaints genome-scale probe set from the Beliveau et al. 2012 PNAS publication.[2] The probes have a length of 32 bases, and have an approximate Tm window of 34-42 degrees. For more information on this probe set, please refer to the 2012 Oligopaints publication.

iFISH4U (DNA)

the full 40-mer probe set from iFISH4U.

All probe sets in the DNA Probe Design tab include all probes targeting the entire reference genome that they target. It takes longer to load and search these sets, but probes can be designed against any coordinates in a given genome. You can use the RNA Probe Design for faster retrieval of probes targeting specific RefSeq coordinates.

3   Advanced Probe Settings

In some cases, you may find that you want to increase the number of probes returned for your targets. PaintSHOP[3] provides a set of advanced features to enable this flexibility. In computational probe design there is an inherent trade-off between probe coverage and specifity. Probe specificity is the likelihood that a probe hybridizes to its intended target instead of at another site in the genome. A greater emphasis on probe specificity inevitably filters probes, reducing coverage. In order to have more control over this design choice, PaintSHOP provides three parameters: 1) repeat inclusion, 2) off-target score, and 3) the maximum k-mer count. These three parameters are described below.

Repeat: RepeatMasker is a program that identifies the presence of repetitive elements in a given DNA sequence. The human genome has been annotated by RepeatMasker. Previous

probe design tools have excluded repetitive sequences. PaintSHOP provides the option to allow for probes which contain bases that have been flagged as repetitive, if it is necessary to have enough probes for a target that is challenging to cover. By default, repeat mode is set to off.

Off-Target Score: One important component of PaintSHOP is the Homology Optimization Pipeline (HOP). The pipeline is used when creating new probe sets to create an on-target and off-target score for every probe identified. We have developed a machine learning model to approximate nucleic acid thermodynamics. For the on-target score, the model is used to score the likelihood (0-100) that a probe is likely to hybridize at its intended target. For the off-target score, we start by searching for up to 100 possible alignments for each candidate probe. Any candidate with greater than 100 possible alignments is discarded. Next, we use our model to generate a score for the likelihood of hybridization at each possible site. We sum these scores, generating an off-target score between 0 and 10,000. By default, PaintSHOP sets the default maximum off-target score to 200. The off-target score slider can be used to make this value more or less stringent, depending on the experiment. The probe table and density plot will dynamically update, providing information on how the parameters chosen affect the probe set.

Max K-mer count: A k-mer is a substring of a DNA sequence of length k. We use JELLYFISH to count how many times each 18-mer substring in a given probe occurs in the genome it targets.[4] The maximum value of all 18-mer counts is another way to control probe specifity, and can identify problematic substrings that other alignment approaches may miss. By default, PaintSHOP uses a maximum k-mer count of 5. This can be changed using the slider, and the probe table and density plot will dynamically update.

Note: At any time, the Restore Default Parameters button can be used to reset the default values if you want to remove changes you have made.

4   Optimizing a Probe Set

In some instances, you may want to even out your probe set by selecting a certain number of probes per target. PaintSHOP offers two features for optimizing a probe set: 1) trim, and 2) unify number. For either option, use the slider to choose how many probes you want per target. The trim option simply ranks the probes for each target by off-target score, and keeps however many you need, removing the rest. The unify number option will behave the same way for targets with enough probes, and will automatically relax stringency parameters for targets without enough probes with the currently selected parameters to return enough probes to meet the target, and trim the probes from targets with too many probes. Importantly, this means that changes from the Advanced Probe Settings section will be overriden using the unify number feature.

5 Appending Sequences

5.1 Overview of appending feature

PaintSHOP provides a suite of functionality for appending the necessary sequences to your probes to carry out your experiment. The following diagram shows where you can append sequences to your probes:
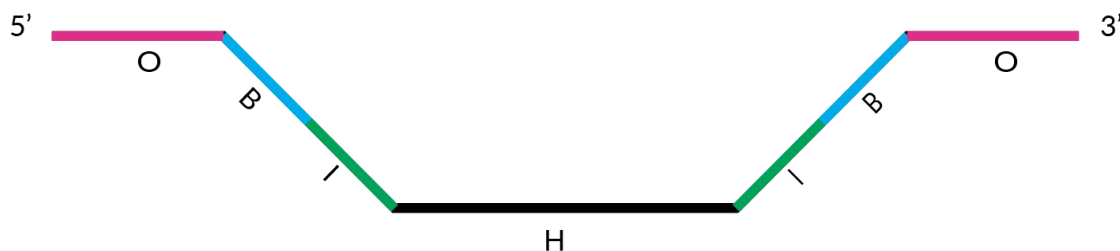
Figure 2: Schematic of a PaintSHOP probe showing with the required homology domain (H) as well as the optional inner (I), bridge (B), and outer (O) domains where sequences can be appended to facilitate design of complex targeting and readout schemes.

The probe sequence itself is the Homology Region (H) in the center. If you are going to amplify your probes from an oligo pool using PCR, you can add forward primers to the 5' I location, and reverse primers to the 3' I location (I stands for Inner Primer). The next regions where sequences can be added are the bridge regions (B) on both the 5' and 3' sides of the probe. Other names for this include read out, ear, and barcode. This sequence can be used as a location for secondary oligos to bind to. Sequences can also be appended to the outer region (O) on the 5' and 3' sides. One useful application of this is to include region/target specific primer pairs to be able to amplify only specific portions of your oligo pool at certain times in your experimental workflow. You can also choose to add SABER concatemer sequences to your probes.[5]

Every appending location is optional.

To append sequences, first select whether you used the RNA Probe Design or DNA Probe Design tab. If you want to append sequences to a specific region, switch the radio button for the region from None to Append. Once you choose to append a sequence, a set of options will appear. For example, once the 5' Outer Primer Sequence is selected, the appending menu will look like this:

Figure 3: Screenshot of the PaintSHOP appending interface.

The **Orientation** option lets you choose between appending the forward (5' to 3') orientation of the sequence, and the reverse complement of the sequence. The default orientation is forward for the 5' regions and reverse complement for the 3' regions. SABER sequences have a specific orientation and don't have this option.

The **Format** orientation is very important. This determines how sequences from whatever set you choose will be used. If you choose 'Same for all probes', the first sequence in the selected sequence set will be appended to all probes for that region. An example of this use case would be appending the same forward and reverse primer pair to your entire probe set. The next option is 'Unique for each target'. If this option is selected, each target in your probe set will have a unique appended to it. For example, if you are appending a primer sequence to a probe set with 10 targets, the first target will have the first sequence in the selected set appended, while the second sequence will be appended to the second target, all the way up to the tenth target receiving the tenth sequence. This can be used to append a unique bridge sequence to each target, for example. 'Multiple per target' allows you to add more than one sequence to each target. Assume you have set [A, B, C] of sequences to append, and your target has probes 1, 2, 3, 4, 5. You select to append 3 sequences to this target. This option will append A1, B2, C3, A4, B5. This option is intended to support oligo FISH technologies such as seqFISH and MERFISH. Choose how many probes per target with the **Number Per Target** slider.

If none of the built-in appending formats are quite right for your experiment, you can use the **Custom Ranges** option. First, select 'Custom ranges' as your **Format** choice. The ranges that you enter in the text box correspond to ranges of your probe set that should get a unique probe. For example, for a probe set with 200 probes, you could add a unique sequence to each half by specifying '1-100, 101-200' as a custom range. You can use the probe table returned in the RNA/DNA Probe Design tab to design custom ranges.

To choose what set of sequences to append for a region, use the **Select Sequence Set** drop-down menu. The menu will be populated with sets specific to the region that you chose to append sequences to. Descriptions of the different sequence sets can be found below. Alternatively, you can choose 'Custom Set' from the drop down and use the **Upload Custom Set** option to use your own sequence file. Upload a text file with one sequence per line.

You can choose between appending Primer/Bridge/Universal sequence(s) or a SABER concatemer sequence to the 3' side of your probe. The Primer/Bridge/Universal sequences have the same options on both the 5' and 3' side. The SABER interface is simpler, since there are fewer choices to make. The two options are the number of copies of the concatemer (1 or 2), and the **Format** that should be used to append the sequences.

Once you have selected the appending options necessary for your experiment, click the **Append** button at the bottom of the side panel. A dynamic table will appear, showing which sequences where appended. The first column of the table is a list of your targets. The other columns are determined by which sequences you append. Each entry in the table has the following format: set_type# where the sets are:

- ps = PaintSHOP
- saber_1x = SABER PER primers
- saber_2x = SABER PER primers with two copies of a given concatemer
- merfish = Sets from MERFISH site[6]
- Kishi2019 = Sets from Kishi et al. 2019[5]
- Mateo2019 = Sets from Mateo et al. 2019[7]
- Xia2019 = Sets from Xia et al. 2019[8]

And the types are:

- of = outer forward primer
- or = outer reverse primer
- if = inner forward primer
- ir = inner reverse primer
- fpb = 5' bridge sequence
- tpb = 3' bridge sequence
- primer = primer with no specified forward/reverse
- bridge = bridge set that can be used on either 5' or 3' side or used to make a custom set

For example: ps_of1 is the first outer forward primer of the PaintSHOP set. The actual sequence is also shown for each table entry as well.

Note: Because the appending sequences are generally designed to be orthogonal to many genomes and have minimal secondary structure propensity, these generally can be added at will to any probe set. That said, it is always possible that in rare cases specific probe + appending sequence combinations can lead to *de novo* secondary structure or drive off-target binding. Secondary structure formation propensity can be checked with the OligoMiner structureCheck.py script and off-target binding potential can be checked with BLAST against nr/nt.

5.2   PaintSHOP appending sequence sets

PaintSHOP 5' Outer Primer Set

a set of 10 forward primers for the 5' O position. This set is paired with the PaintSHOP 3' Outer Primer Set.

MERFISH Primer Set

a set of 318 primer sequences from the MERFISH resources. This set is available for use at all primer locations.

PaintSHOP Full Bridge Set

a set of 800 new bridge sequences provided with PaintSHOP that are designed to be orthogonal to all other sequences in the set. This set is available for use on both the 5' and 3' Bridge regions.

PaintSHOP 5' Bridge Set

400 of the PaintSHOP bridge sequences that can be used for the 5' Bridge region if you are going to append different bridge sequences to the 5' and 3' sides of your probe.

PaintSHOP 3' Bridge Set 400 of the PaintSHOP bridge sequences that can be used for the 3' Bridge region if you are going to append different bridge sequences to the 5' and 3' sides of your probe.

MERFISH Bridge Set

a set of 16 bridge sequences from the MERFISH resources. This set is available for use on both the 5' and 3' Bridge regions.

Kishi et al. 2019 Bridges

a set of 84 bridge sequences provided with the Kishi et al. 2019 publication in Nature Methods.[5] These sequences are available for use on both the 5' and 3' Bridge regions.

Mateo et al. 2019 Bridges

a set of 199 bridge sequences provided with the Mateo et al. 2019 publication in Nature.[7] These sequences are available for use on both the 5' and 3' Bridge regions.

Xia et al. 2019 Bridges

a set of 70 bridge sequences provided with the Xia et al. 2019 publication in PNAS.[8] These sequences are available for use on both the 5' and 3' Bridge regions.

PaintSHOP 5' Inner Primer Set

a set of 74 forward primers for the 5' I position. This set is paired with the PaintSHOP 3' Inner Primer Set.

PaintSHOP 3' Inner Primer Set

a set of 74 reverse primers for the 3' I position. This set is paired with the PaintSHOP 5' Inner Primer Set.

PaintSHOP 3' Outer Primer Set

a set of 10 reverse primers for the 3' O position. This set is paired with the PaintSHOP 5' Outer Primer Set.

It is important to note that PaintSHOP will start at the top of the list of whatever set you choose at each location. If you would like to use different primers from the same set at different locations, you may want to download the set from the **Resources** tab and create custom files with the sequences that you would like.

If you have sequences that you would like to be included in the PaintSHOP application, please reach out to us!

6 Downloading Designs

Once you have finished designing your probe set and appending any sequences necessary for your experiment, you can visit the **Download** tab. The first step when downloading is to select whether or not you appended sequences to your probe set, and whether you used the RNA or DNA Probe Design tab. Next, you can choose the type of file you would like to download from the drop-down menu. PaintSHOP provides four different files for download:

1. Order File
2. Appending File
3. Full Probe File
4. Citation File

The **Order File** is formatted to be ready to be sent directly to a company as an order for an oligo pool/library. The format is simple. The file has two columns. The first contains a unique identifier for each probe sequence, generated by concatenating the chromosome and the start coordinate of the probe. The second column is the actual sequence.

The **Appending File** is of the same format as the table returned in the appending tab. This can be useful to download as documentation of your final appending designs.

The Full **Probe File** is of the same format as whichever design tab you used to select your probes. This file can serve as a reference with full information about the probe sequences in your set.

The **Citation File** contains formatted citations and BibTeX entries for the publications associated with all resources used in the application. The BibTex entries can be imported directly into most reference managers. Please cite the publications that are associated with the resources that you used. This recognizes the effort involved in making these resources freely available to accelerate research.

7   Resources Available

As part of our effort to provide open resources for the oligo-FISH community, all probe sets and appending sets used in the backend of PaintSHOP can be downloaded at paintshop.io.

Our goal in designing PaintSHOP was to make easy things easy, and hard things possible. A key use case of the resources we have made available is to download the sets of sequences that can be appended, create custom appending files, and use those files in PaintSHOP. With this flexibility, practically all appending design schemes should be possible.

References

[1] Brian J. Beliveau, Jocelyn Y. Kishi, Guy Nir, Hiroshi M. Sasaki, Sinem K. Saka, Son C. Nguyen, Chao ting Wu, and Peng Yin. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences*, 115(10):E2183–E2192, feb 2018.

[2] B. J. Beliveau, E. F. Joyce, N. Apostolopoulos, F. Yilmaz, C. Y. Fonseka, R. B. McCole, Y. Chang, J. B. Li, T. N. Senaratne, B. R. Williams, J.-M. Rouillard, and C. t. Wu. Versatile design and synthesis platform for visualizing genomes with oligopaint FISH probes. *Proceedings of the National Academy of Sciences*, 109(52):21301–21306, dec 2012.

[3] Elliot A. Hershberg, Jennie L. Close, Conor K. Camplisson, Sahar Attar, Ryan Chern, Yuzhen Liu, Shreeram Akilesh, Philip R. Nicovich, and Brian J. Beliveau. PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. jul 2020.

[4] Guillaume Marçais and Carl Kingsford. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6):764–770, jan 2011.

[5] Jocelyn Y. Kishi, Sylvain W. Lapan, Brian J. Beliveau, Emma R. West, Allen Zhu, Hiroshi M. Sasaki, Sinem K. Saka, Yu Wang, Constance L. Cepko, and Peng Yin. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nature Methods*, 16(6):533–544, may 2019.

[6] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 348(6233):aaa6090–aaa6090, apr 2015.

[7] Leslie J. Mateo, Sedona E. Murphy, Antonina Hafner, Isaac S. Cinquini, Carly A. Walker, and Alistair N. Boettiger. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, 568(7750):49–54, mar 2019.

[8] Chenglong Xia, Jean Fan, George Emanuel, Junjie Hao, and Xiaowei Zhuang. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences*, 116(39):19490–19499, sep 2019.