Designing genome-wide MERFISH probes for understudied species

Researchers studying any organism with genomic data can follow this simple walkthrough to create sets of barcoded probes for the multiplexed FISH technique called MERFISH. We're sharing interactive code notebooks that can be adapted to design barcoded FISH probes for any species.

Contributors (A-Z)

Prachee Avasthi, Feridun Mert Celebi, Jase Gehring, Megan L. Hochstrasser

Version 1   ·   Mar 31, 2025

Purpose

We're interested in understanding how RNA trafficking and local translation affect ciliary growth in *Chlamydomonas reinhardtii*. MERFISH is a technology for visually tracking hundreds to thousands of RNAs at once, so it is an appealing option. However, existing tools for designing the barcoded DNA probes used in MERFISH experiments are not very versatile, and many understudied organisms do not have publicly shared probe sets.

We developed a short bioinformatic workflow to design MERFISH probes for *C. reinhardtii*, and we're sharing it here so the community can use it for experiments in

other understudied organisms.

- This pub is part of the **project**, "[Spatial genomics: Tracking biomolecules across space and time](#)." Visit the project page for more background and context.

- We've shared all of our **code** in executable Colab notebooks so users can modify the workflow and run it with their own data on Google's cloud computing network.

Summary

This publication is part of a larger effort at Arcadia to apply [spatial genomics](#) techniques, experiments that will allow us to profile RNAs and other biomolecules from tissue sections, cell lines, and unicellular organisms. One of the first steps in a new spatial transcriptomics experiment is to design barcoded probes complementary to RNA molecules in the cell. Here, we are releasing an executable probe design pipeline as a set of Google Colab notebooks. These notebooks serve as a functional example of probe design, stepping through the process to generate MERFISH probes against the *C. reinhardtii* genome. They allow users to reproduce our work on Google's free, browser-based virtual computing platform, Colab, providing interactive access for those who want to design probes against new genomes or otherwise modify our workflow. Users can upload their own data, download their results, and make custom changes to the codebase.

The strategy

[MERFISH](#) (multiplexed error-robust fluorescence *in situ* hybridization) is a method that allows researchers to simultaneously image hundreds to thousands of RNA transcripts or other molecules within single cells. The protocol relies on sets of barcoded "encoding probes" that hybridizes to target biomolecules and are subsequently identified through the sequential binding of a set of fluorescent "readout" probes.
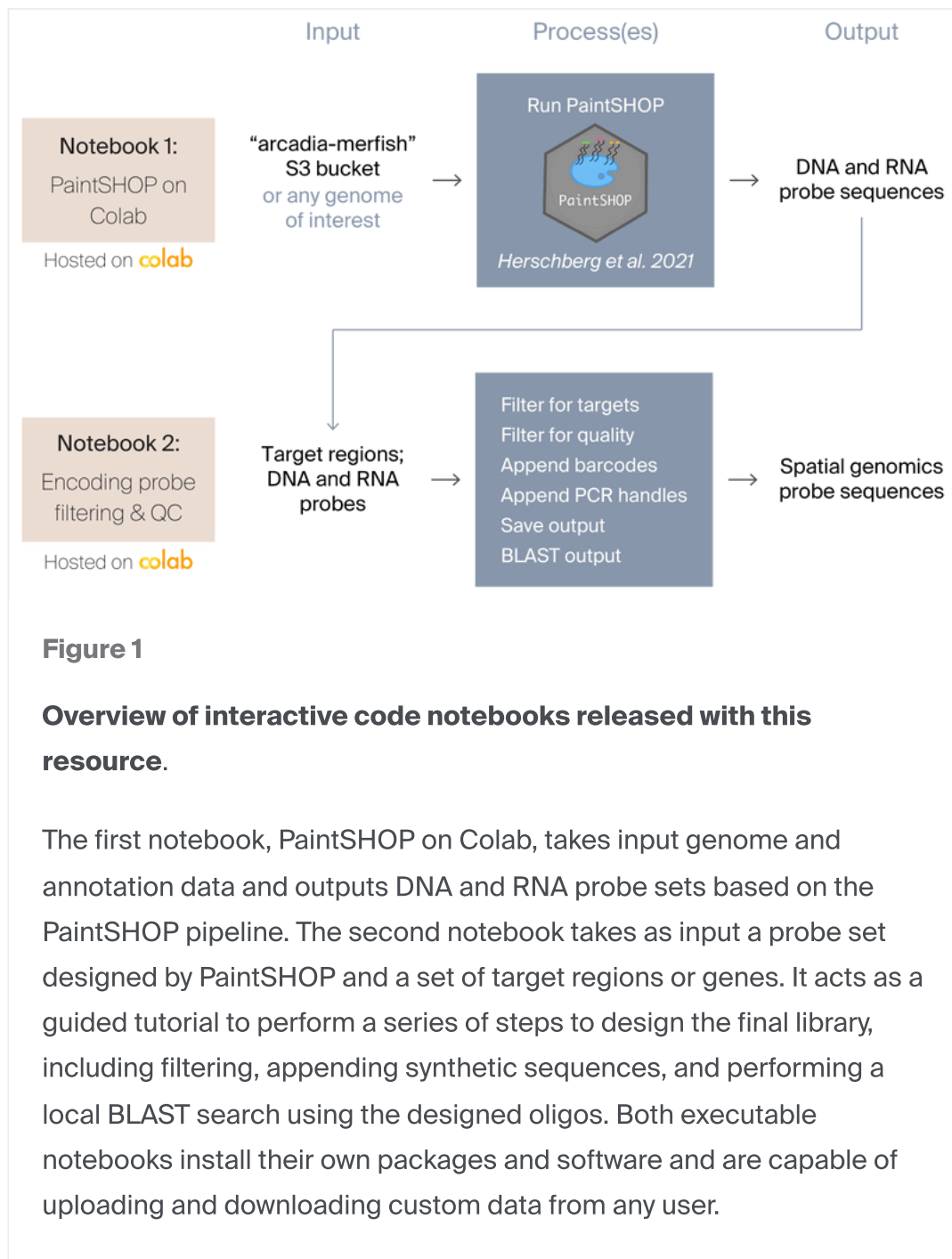
The problem

There are great tools to generate MERFISH probes, but they aren't always set up to work with new genomes. Our first target organism, *Chlamydomonas reinhardtii*, does not yet have a set of designed MERFISH probes.

Our solution

Here, we apply the PaintSHOP pipeline **[1]** for FISH probe design to generate a genome-wide panel of MERFISH probes for *Chlamydomonas reinhardtii*. We provide our workflows in the form of interactive, browser-based code notebooks that can be used by researchers to generate MERFISH probes for any genome of interest. These executable notebooks can be readily modified or improved to support any number of applications in the rapidly developing field of spatial genomics and multiplexed in situ hybridization.

The resource

The process to create gene-specific MERFISH probe sets for *Chlamydomonas* is summarized below. By supplying appropriate genome and genome annotation files (FASTA and GTF, respectively), as well as gene names or target genomic coordinates, users can easily modify these self-contained notebooks to design MERFISH probes against any genome of interest. The core of the workflow is the PaintSHOP pipeline for FISH probe design, but the notebooks also contain all code necessary to download packages and data, install software, execute the probe design workflow, download, and verify the designs. The first of two notebooks runs the PaintSHOP pipeline on a new genome of interest, generating a genome-wide set of DNA and RNA FISH probes (Figure 1). Researchers can use the second notebook for individual MERFISH experiments. It designs probes for a given experiment by filtering for a set of target genes, appending MERFISH barcodes and PCR handles, and finally running BLAST against the target genome to check for off-target binding events.

Figure 1

Overview of interactive code notebooks released with this resource.

The first notebook, PaintSHOP on Colab, takes input genome and annotation data and outputs DNA and RNA probe sets based on the PaintSHOP pipeline. The second notebook takes as input a probe set designed by PaintSHOP and a set of target regions or genes. It acts as a guided tutorial to perform a series of steps to design the final library, including filtering, appending synthetic sequences, and performing a local BLAST search using the designed oligos. Both executable notebooks install their own packages and software and are capable of uploading and downloading custom data from any user.

These notebooks make it easy for us to share our workflows in a more permanent and flexible way than traditional bioinformatics software. They are one example of containerized, cloud-based scientific computing. This model of scientific computing is focused on cloud datasets rather than physical computing hardware. Users attach scalable virtual machines to their datasets and build lightweight environments in which to carry out actions on the data. In this way, data is more readily accessed from different physical endpoints or different users. Clean, reproducible computing environments greatly reduce the effort required to adopt code written by others.
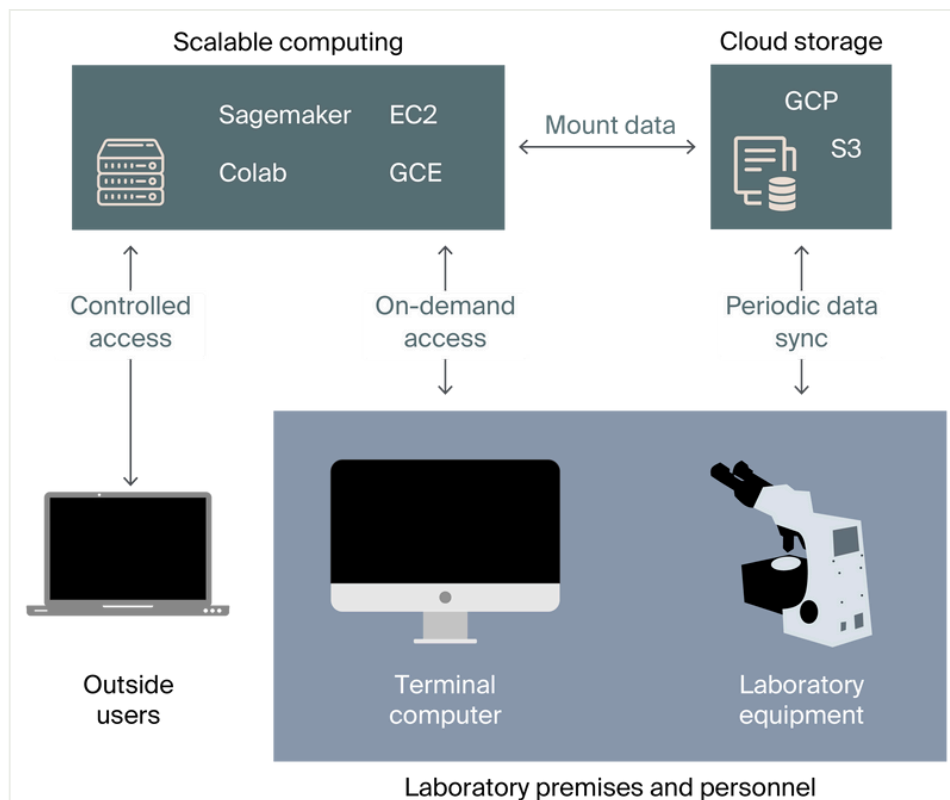
Figure 2

Diagram of scientific cloud computing network.

Computing environments, including compute resources (CPU, GPU, number of cores, etc), packages, and installed software, are established by the user at the beginning of a session. Computing environments can be attached or mounted to cloud datasets. Scientists interact with the system as any remote computer, but it is straightforward to allow outside users to reproduce or modify an analysis on the same computing resource on which it was originally performed. Importantly, scientific data can be organized and backed up almost autonomously.

Process overview

Designing target probes

1. Generate genome-wide probes

First, we generate probe sequences against an entire target genome. Our first Colab notebook uses the PaintSHOP pipeline to generate genome-wide probes using *C. reinhardtii* genome FASTA and GTF files.

1. Select targets

Next, we must choose the set of target genes or genome regions that we'd like to detect in our eventual MERFISH experiment. Our second notebook filters the genome-wide encoding probe set based on user-specified target genes. This example uses a small set of genes associated with ciliary assembly in *Chlamydomonas*. In practice, the 16-bit MHD4 barcode set used in these notebooks can barcode up to 140 separate genes or genomic targets.

1. Append MERFISH sequences

Next, we append MERFISH sequence regions to the oligos we've designed so far. This includes one to four MERFISH readout regions (where fluorescent probes bind) as well as PCR primer binding sites for amplification and manipulation of the pooled MERFISH oligo library.

1. Do quality control

Our second notebook installs the BLAST CLI and builds a local BLAST database against the Chlamy genome. Using BLAST, any off-target binding sites can be identified, and a BED file can be created to visually inspect the designed probes using a genome browser.

Designing readout probes

Readout probes are complementary to the synthetic readout regions appended to the target sequences in Step 3 above. In most cases, we use disulfide-bridged oligos to bind directly to the encoding probe. These can be easily cleaved with a DTT wash between rounds of staining and imaging.

Colab Notebook #1: Designing probes with PaintSHOP on Colab

This interactive notebook generates new FISH probes against a genome of interest. As a demonstration of our own workflow, the notebook downloads data from Arcadia's public Amazon S3 bucket `arcadia-merfish` and generates probes against the *Chlamydomonas reinhardtii* genome. To run the notebook, use the "Open in Colab" button below and then select `Runtime > Run all` or use `Shift+Enter` to execute individual code blocks.

<div align="center">

CO Open in Colab

</div>

You can also view the workflow below:

`Loading`

CO Open in Colab

Introduction: PaintSHOP Pipeline on Colab

This notebook generates MERFISH probes against the Chlamydomonas reinhardtii genome using the PaintSHOP pipeline (Herschberg et al. 2021) available from the Beliveau Lab here.

The notebook wraps the Linux PaintSHOP pipeline within Colab's shareable and extensible environment, making it easy to generate MERFISH probes against any target genome. To process your own data, work through the notebook below. When prompted, you will need to provide **genome and annotation files** as well as a **temperature** to run the PaintSHOP model.

You can download output files from this notebook to a permanent location. An example is shown in the section *Export the PaintSHOP*

Colab Notebook #2: Encoding probes filtering and quality control

This interactive notebook processes FISH probes for individual MERFISH experiments. It performs the target selection, appending, and quality controls steps described under "Process Overview" above. Our demonstration downloads genomic data from our `arcadia-merfish` S3 bucket, and we show how users can download their own designs.

You can also view the workflow below:

Loading

MERFISH Filtering and Appending MERFISH Sequences

This notebook is used to filter and append MERFISH indexes and PCR primers to the probe sets output by the PaintSHOP pipeline. It is the second in a set of three notebooks used at Arcadia to generate MERFISH probes.

To run this notebook on your own data, you will need to provide:

1. a set of genome-wide target probes output by the PaintSHOP pipeline command line interface
2. genome coordinates or Gene IDs for the targeted regions

This notebook illustrates the filtering and appending process using a panel of seven genes involved in ciliary assembly in

At the end of this process, users will have a set of probes that they can order and use in MERFISH experiments.

Key takeaways

Bioinformatics has traditionally relied on large workstations or clusters to store and process individual datasets, leading to massive redundancy and reduced interoperability of scientific software. A new model for scientific computing is emerging, based on cloud storage of datasets and flexible virtual computers, obviating the need for extensive physical computing resources. Here, we present an example of reproducible scientific computing relying on containerized workflows and virtual machines. We have used this resource to generate FISH probes against the *Chlamydomonas reinhardtii* genome at a range of model temperatures. We provide our inputs and outputs, allowing anyone to reproduce our work and modify the workflow.

Next steps

We will continue to develop these workflows to incorporate alternative probe design engines or alternative encoding probe architectures and barcode sets. We're already using these probes in MERFISH experiments at Arcadia as part of our spatial genomics project and cell biology efforts, and will report our results when we have them. The PaintSHOP pipeline does not currently support overlapping oligo designs, although it has been demonstrated that overlaps up to 20 nucleotides can be used, increasing oligo density severalfold. We may explore pipelines with support for overlapping probes in the future.

Do you have any tools that you find useful for spatial genomics probe design? Would you like to contribute to these notebooks? Are you able to use this workflow successfully in your own work? What could make it easier to use? Comment with your feedback!

References

1 Hershberg EA, Camplisson CK, Close JL, Attar S, Chern R, Liu Y, Akilesh S, Nicovich PR, Beliveau BJ. (2021). PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. https://doi.org/10.1038/s41592-021-01187-3