

# How to **SAVEMONEY** on whole-plasmid sequencing for plasmid verification

*Simple Algorithm for Very Efficient Multiplexing of Oxford Nanopore Experiments for You!*

SAVEMONEY on  
Google Colab



Demo-data



Press to download  
demo-data

Raw

1W

files that are this big right now.)

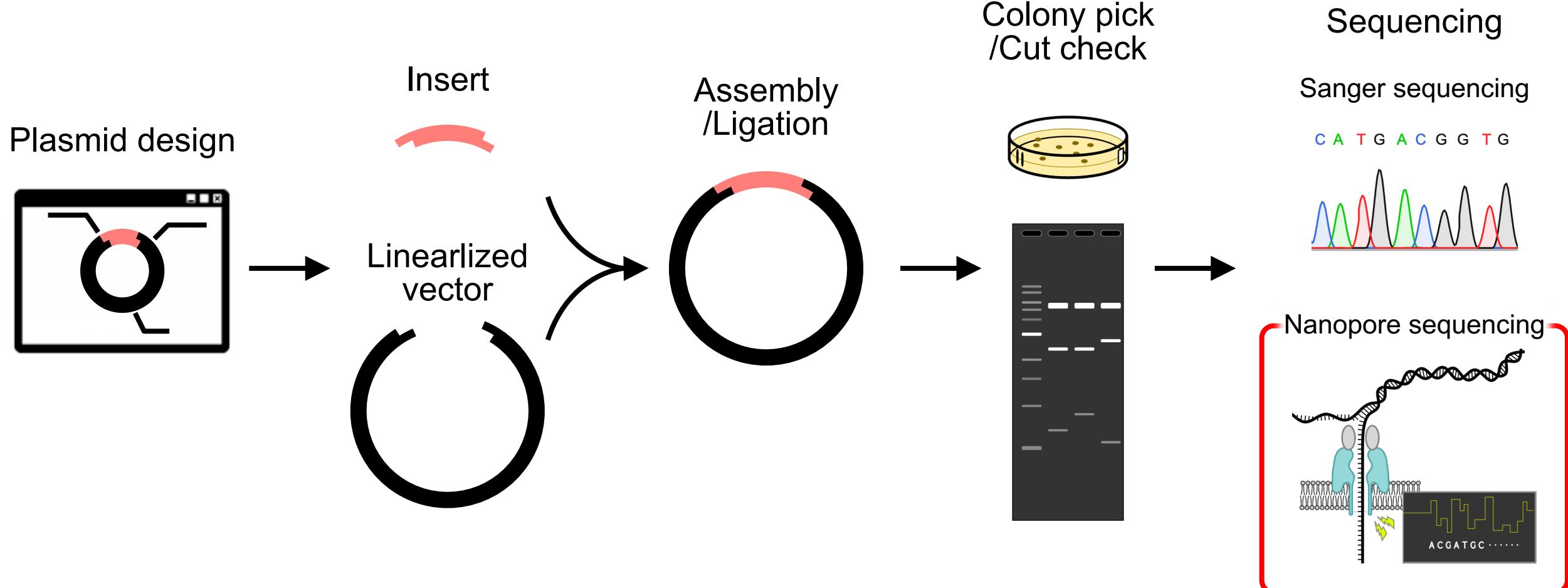
Weill Institute Science Workshop May 22, 2023

Masaaki Uematsu ([mu84@cornell.edu](mailto:mu84@cornell.edu))  
Postdoc, Baskin lab

# Topics

1. Introduction
2. Concept of SAVEMONEY
3. How to execute SAVEMONEY
4. How to interpret outputs

# Plasmid construction



# Sanger sequencing or Nanopore sequencing

plasmidsaurus Services ▾ FAQ Shipping Contact Us

Dashboard

Log out

Why should I sequence the entire plasmid instead of just my region of interest? ^

Lots of reasons!

Nanopore sequencing

Sanger sequencing

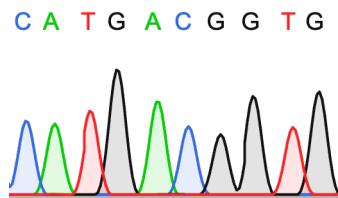
- Probability of mutagenesis in the vector backbone is not zero.
- Robust against repetitive sequences.
- Plasmid can form dimer/multimers.
- It's not much more expensive. ← ??

Nanopore is more rigorous,  
but still Sanger sequencing is chosen because of the COST.

# Baskin Lab's case

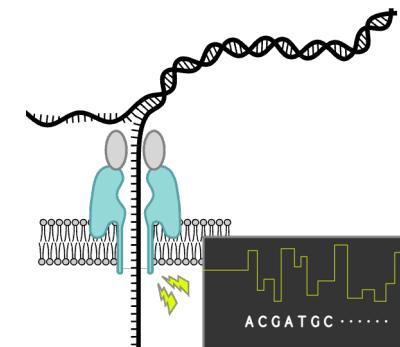
Sanger sequencing

~1000 nt, \$4~5 / run



Nanopore sequencing

whole plasmid, \$15 / plasmid



Previous policy in Baskin Lab

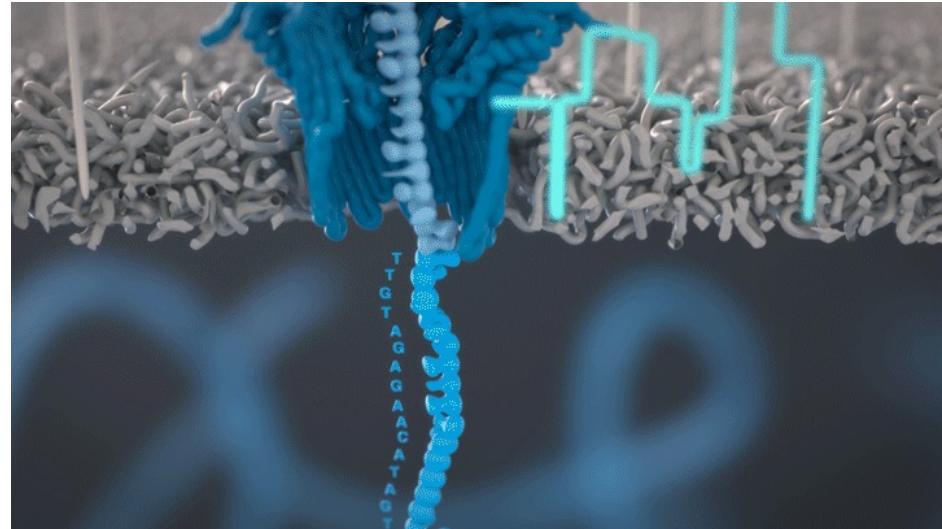
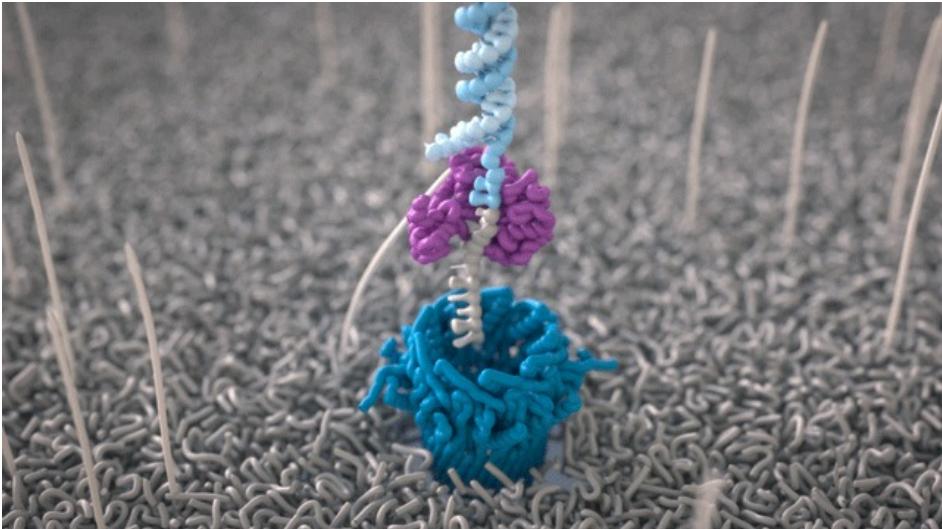
Your insert is large enough to require sanger sequencing runs  $\geq 3$ :

Yes → Nanopore sequencing

No → Sanger sequencing

Cost reduction is required to further democratize Nanopore sequencing.

# How nanopore sequencing works



<https://nanoporetech.com/how-it-works>

- Flow cell contains “nanopores”.
- The current is produced as DNA molecules pass through.
- The produced electric current is decoded into DNA sequences (= **base calling**).
- Different nanopore produces independent sequencing data.

## Long-read sequencing

- 10~50 Kb
- 2.3 Mb (Record in 2020)

## Error prone

- Error rates can be as high as ~20%.  
→ Consensus must be calculated.

# Consensus base calling

Consensus sequence → agtccaaacccggta g acacgacttatcgccactaa

read 1 →	AGTCCAACCCGGTA	A	A : 7 reads
read 2 →	AGTCCAACCCAGTG	G	T : 0 read
read 3 →	AGGTCAACCGGTAG	A	C : 0 read
read 4 →	AGTCCAACCCGGTA	A	G : 1 read
read 5 →	AGTCCAACCCGGTA	A	- : 1 read
read 6 →	AGTCCAACCCGGTA	A	
read 7 →	AGTCCAACCCGGTA	A	
read 8 →	AGTCCAACCCGGTA	-	
read 9 →	AGTCCAACCCGGTA	A	

**Consensus base calling = A**

Plasmidsaurus usually returns 200~2000 reads.

- Excessive reads for verifying sequences
- 30~ reads are sufficient to produce accurate consensus.

→ Sequencing cost can be reduced!

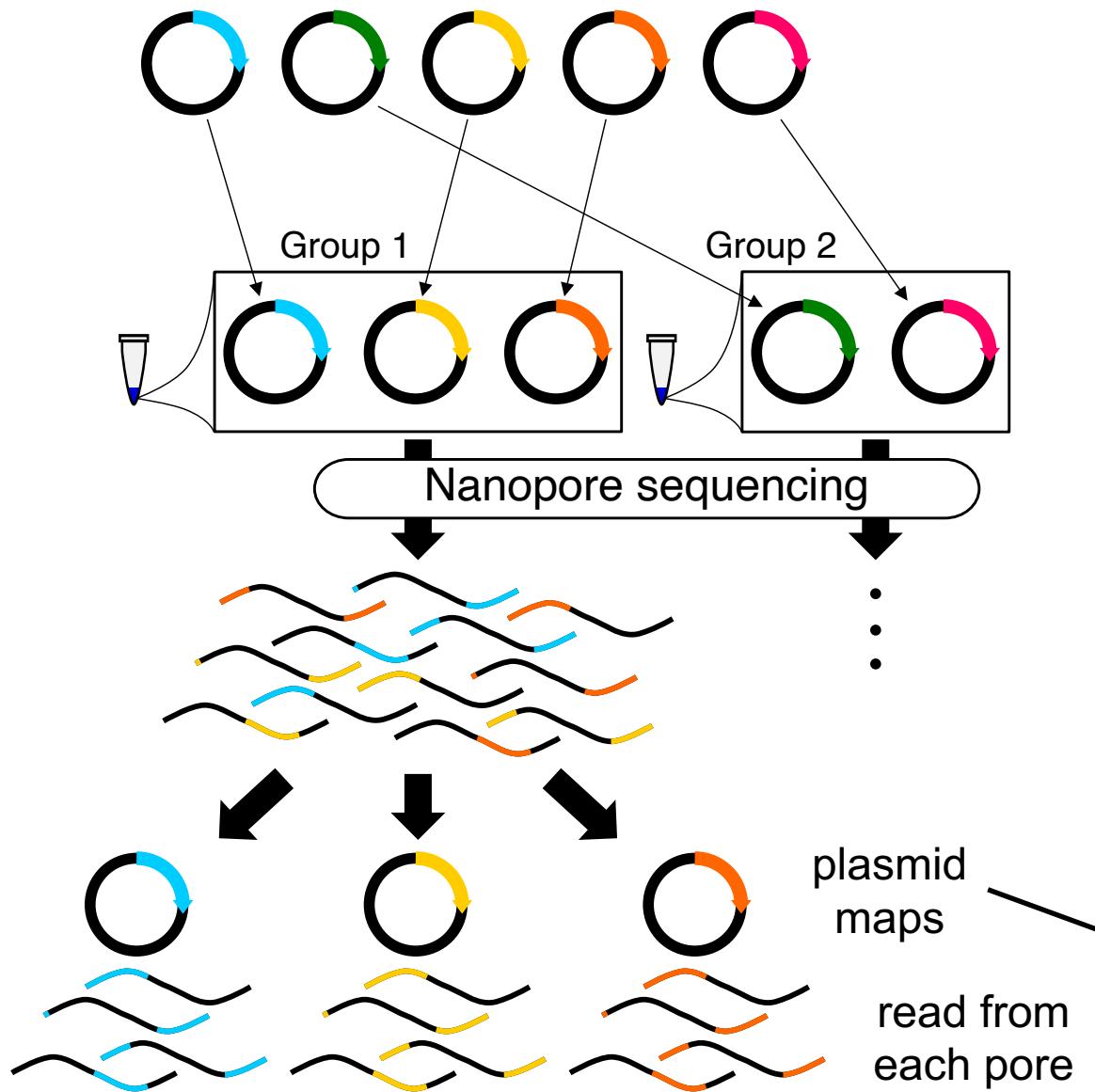
18 errors in 36 nt x 9 reads

: ~95% accuracy without consensus

# Topics

1. Introduction
2. Concept of SAVEMONEY
3. How to use SAVEMONEY
4. How to interpret outputs

# Submit samples together and SAVEMONEY



## Pre-survey

Determine which plasmids are safe to be mixed.  
Plasmid maps are used.

## Sample submission

Mix multiple plasmids together and submit as one sample.

## Post-analysis

Classify each read based on the plasmid maps.  
Obtain consensus sequences.

The differences between each plasmids  
→ Utilized as intrinsic barcodes.  
The use of prior information (plasmid maps, error characteristics, etc.)  
→ High-quality base calling with low number of reads.

# Topics

1. Introduction
2. Concept of SAVEMONEY
- 3. How to use SAVEMONEY**
4. How to interpret outputs

# Prepare data

SAVEMONEY on  
Google Colaboratory



Demo-data



Press to download  
demo-data



Raw

[Raw](#) [Download](#)  
files that are this big right now.)

```
demo_data_1_subset_input/  
  └── my_plasmid_maps_fa/  
      ├── M32_pmNeonGreen-N1.fa  
      ├── M38_mCherry-Spo20.fa  
      └── M42_GFP-PASS_vecCMV.fa  
  └── my_fastq_files/  
      └── Uematsu_n7x_1_MU-test1_subset.fastq
```

# How to use SAVEMONEY

Pre-survey

Sample submission

Post-analysis

# Pre-survey

Determine which plasmids are safe to be mixed before sample submission.

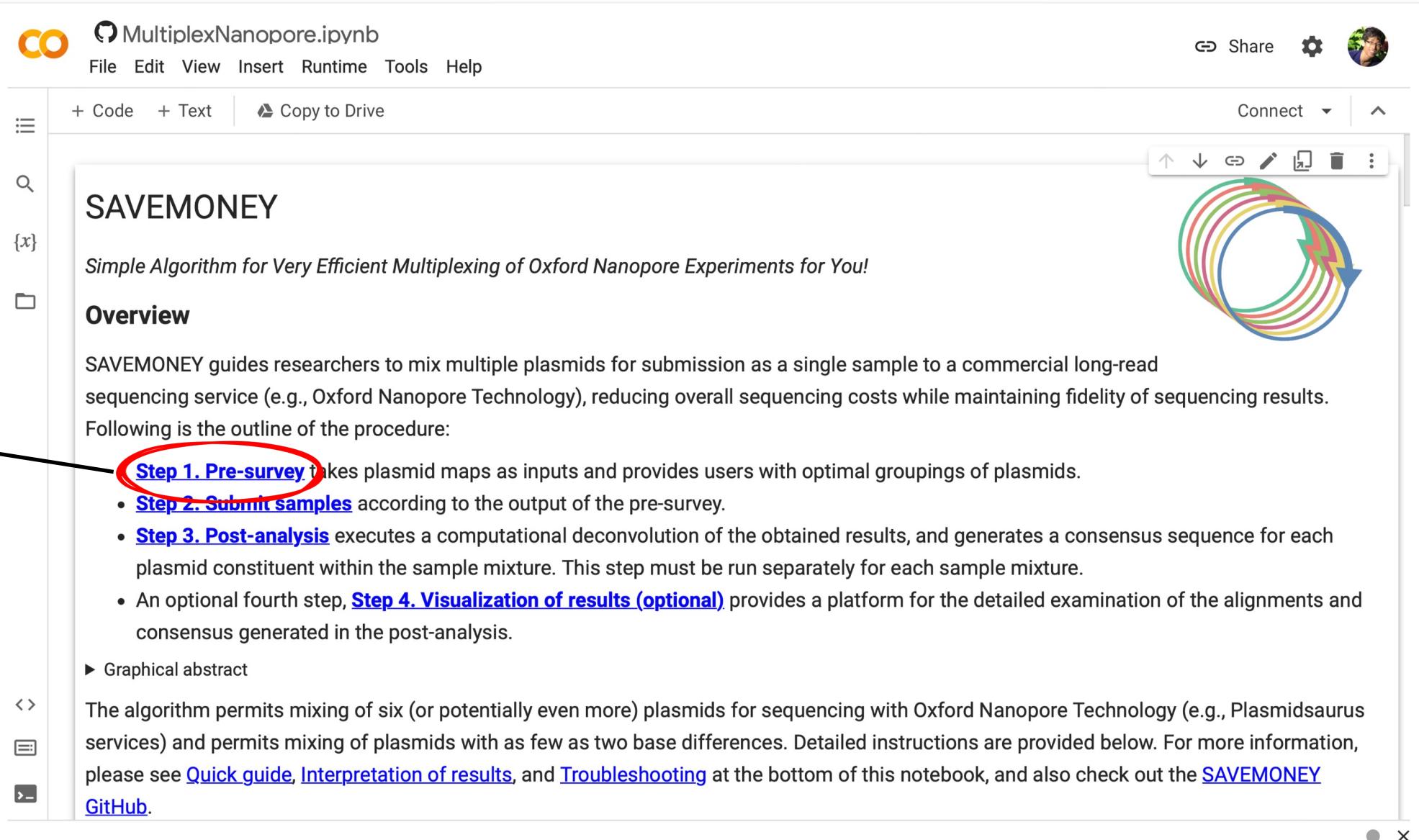
- At least 2 bases difference is required.
- Default threshold is 6 bases.
  - See "[Step 1-2. Advanced settings](#)" to change it.

## Procedure

- Upload plasmid maps
- Execute scripts

# Go to “Step 1. Pre-survey”

Scroll down or press the link



**SAVEMONEY**

*Simple Algorithm for Very Efficient Multiplexing of Oxford Nanopore Experiments for You!*

**Overview**

SAVEMONEY guides researchers to mix multiple plasmids for submission as a single sample to a commercial long-read sequencing service (e.g., Oxford Nanopore Technology), reducing overall sequencing costs while maintaining fidelity of sequencing results.

Following is the outline of the procedure:

- [\*\*Step 1. Pre-survey\*\*](#) takes plasmid maps as inputs and provides users with optimal groupings of plasmids.
- [\*\*Step 2. Submit samples\*\*](#) according to the output of the pre-survey.
- [\*\*Step 3. Post-analysis\*\*](#) executes a computational deconvolution of the obtained results, and generates a consensus sequence for each plasmid constituent within the sample mixture. This step must be run separately for each sample mixture.
- An optional fourth step, [\*\*Step 4. Visualization of results \(optional\)\*\*](#) provides a platform for the detailed examination of the alignments and consensus generated in the post-analysis.

► Graphical abstract

The algorithm permits mixing of six (or potentially even more) plasmids for sequencing with Oxford Nanopore Technology (e.g., Plasmidsaurus services) and permits mixing of plasmids with as few as two base differences. Detailed instructions are provided below. For more information, please see [Quick guide](#), [Interpretation of results](#), and [Troubleshooting](#) at the bottom of this notebook, and also check out the [SAVEMONEY GitHub](#).

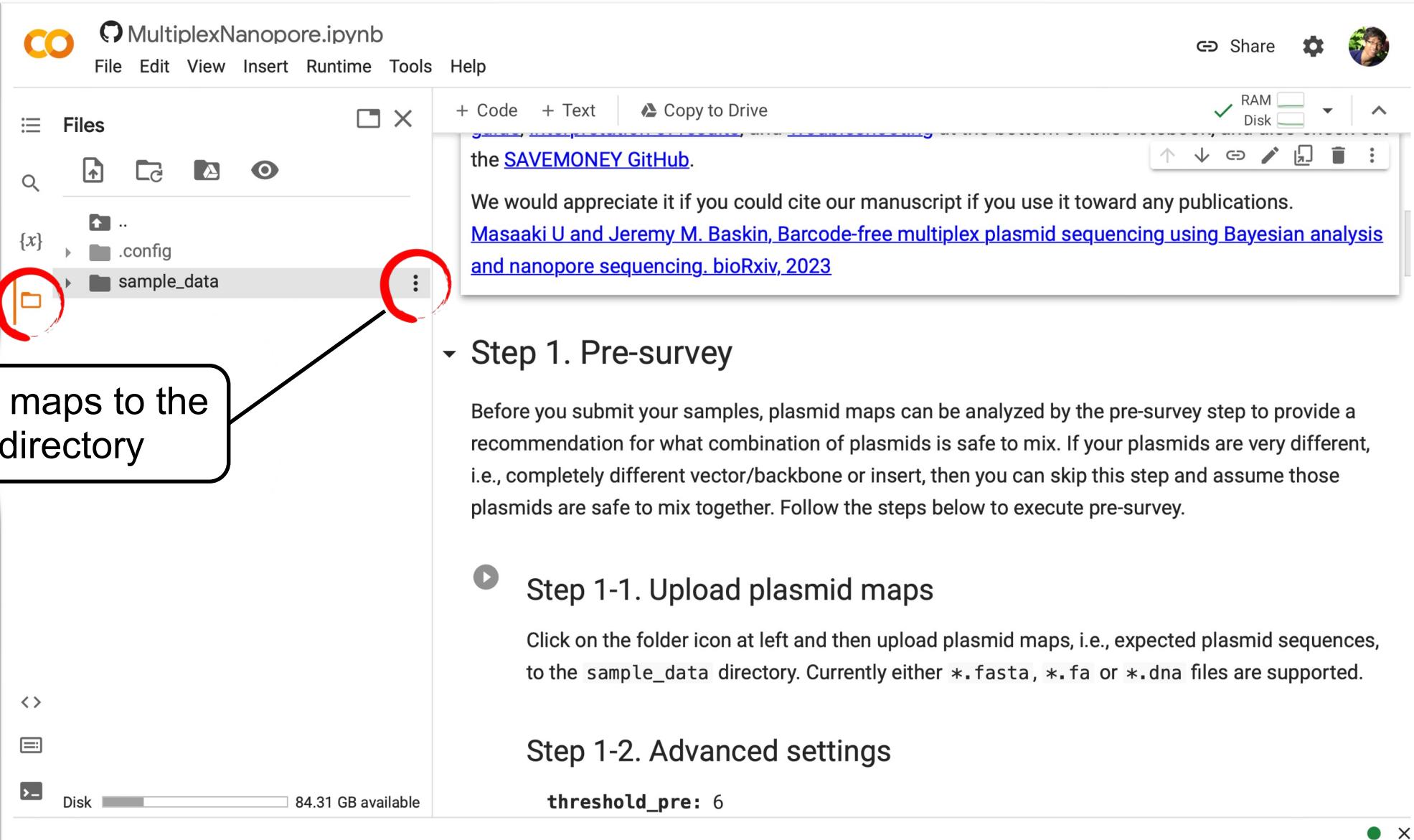
# Upload plasmid maps

The screenshot shows a Jupyter Notebook interface with the following elements:

- Title Bar:** MultiplexNanopore.ipynb, File, Edit, View, Insert, Runtime, Tools, Help, Share, Settings, User Profile.
- Toolbar:** + Code, + Text, Copy to Drive, Connect, Up/Down arrows.
- Left Sidebar:** File browser icons (Browse files, folder, {x}, search, refresh).
- Section Header:** ▾ Step 1. Pre-survey
- Description:** Before you submit your samples, plasmid maps can be analyzed by the pre-survey step to provide a recommendation for what combination of plasmids is safe to mix. If your plasmids are very different, i.e., completely different vector/backbone or insert, then you can skip this step and assume those plasmids are safe to mix together. Follow the steps below to execute pre-survey.
- Step 1-1. Upload plasmid maps:** Click on the folder icon at left and then upload plasmid maps, i.e., expected plasmid sequences, to the `sample_data` directory. Currently either `*.fasta`, `*.fa` or `*.dna` files are supported.
- Step 1-2. Advanced settings:** `threshold_pre: 6`
- Description:** ▶ Description
- Step 1-3. Click this cell and hit Runtime -> Run the focused cell**
- Step 1-4. View pre-survey results**

A red circle highlights the folder icon in the sidebar, and a red arrow points from the text "Browse files" to this icon. A red box highlights the "Browse files" text.

# Upload plasmid maps



The screenshot shows a Jupyter Notebook interface with the title "MultiplexNanopore.ipynb". On the left, there's a file browser titled "Files" showing a directory structure with ".config" and "sample\_data" folders. A red circle highlights the folder icon in the "sample\_data" folder. A callout bubble labeled "Browse files" points to this icon. Another callout bubble labeled "Upload plasmid maps to the sample\_data directory" points to the "sample\_data" folder. The main content area contains text about a GitHub repository and a citation. Below this, a section titled "Step 1. Pre-survey" is expanded. It contains a sub-section titled "Step 1-1. Upload plasmid maps" with instructions to upload plasmid maps to the "sample\_data" directory. A "threshold\_pre: 6" parameter is also visible.

Browse files

Upload plasmid maps to the sample\_data directory

the [SAVEMONEY GitHub](#).  
We would appreciate it if you could cite our manuscript if you use it toward any publications.  
[Masaaki U and Jeremy M. Baskin, Barcode-free multiplex plasmid sequencing using Bayesian analysis and nanopore sequencing, bioRxiv, 2023](#)

▼ Step 1. Pre-survey

Before you submit your samples, plasmid maps can be analyzed by the pre-survey step to provide a recommendation for what combination of plasmids is safe to mix. If your plasmids are very different, i.e., completely different vector/backbone or insert, then you can skip this step and assume those plasmids are safe to mix together. Follow the steps below to execute pre-survey.

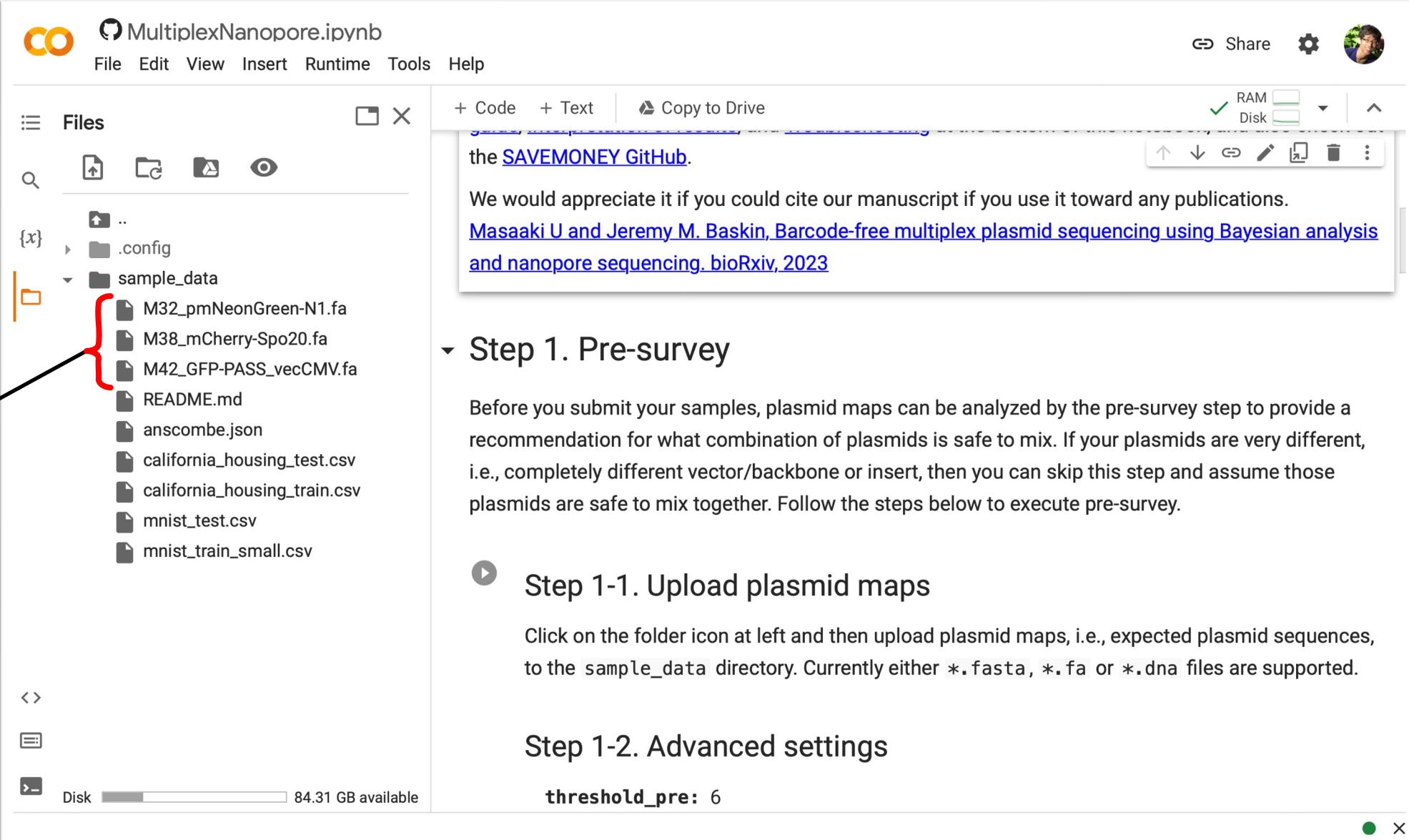
▶ Step 1-1. Upload plasmid maps

Click on the folder icon at left and then upload plasmid maps, i.e., expected plasmid sequences, to the sample\_data directory. Currently either \*.fasta, \*.fa or \*.dna files are supported.

Step 1-2. Advanced settings

threshold\_pre: 6

# Upload plasmid maps



The screenshot shows a Jupyter Notebook interface with the title "MultiplexNanopore.ipynb". On the left, there's a file browser titled "Files" showing a directory structure. A red bracket highlights the "sample\_data" folder, which contains several files: M32\_pmNeonGreen-N1.fa, M38\_mCherry-Spo20.fa, M42\_GFP-PASS\_vecCMV.fa, README.md, anscombe.json, california\_housing\_test.csv, california\_housing\_train.csv, mnist\_test.csv, and mnist\_train\_small.csv. A callout box labeled "Uploaded plasmid maps" points to this folder. The main notebook area contains text about a pre-survey step and a section titled "Step 1-1. Upload plasmid maps".

the [SAVEMONEY GitHub](#).

We would appreciate it if you could cite our manuscript if you use it toward any publications.

[Masaaki U and Jeremy M. Baskin, Barcode-free multiplex plasmid sequencing using Bayesian analysis and nanopore sequencing, bioRxiv, 2023](#)

## ▼ Step 1. Pre-survey

Before you submit your samples, plasmid maps can be analyzed by the pre-survey step to provide a recommendation for what combination of plasmids is safe to mix. If your plasmids are very different, i.e., completely different vector/backbone or insert, then you can skip this step and assume those plasmids are safe to mix together. Follow the steps below to execute pre-survey.

### ▶ Step 1-1. Upload plasmid maps

Click on the folder icon at left and then upload plasmid maps, i.e., expected plasmid sequences, to the sample\_data directory. Currently either \*.fasta, \*.fa or \*.dna files are supported.

### Step 1-2. Advanced settings

`threshold_pre: 6`

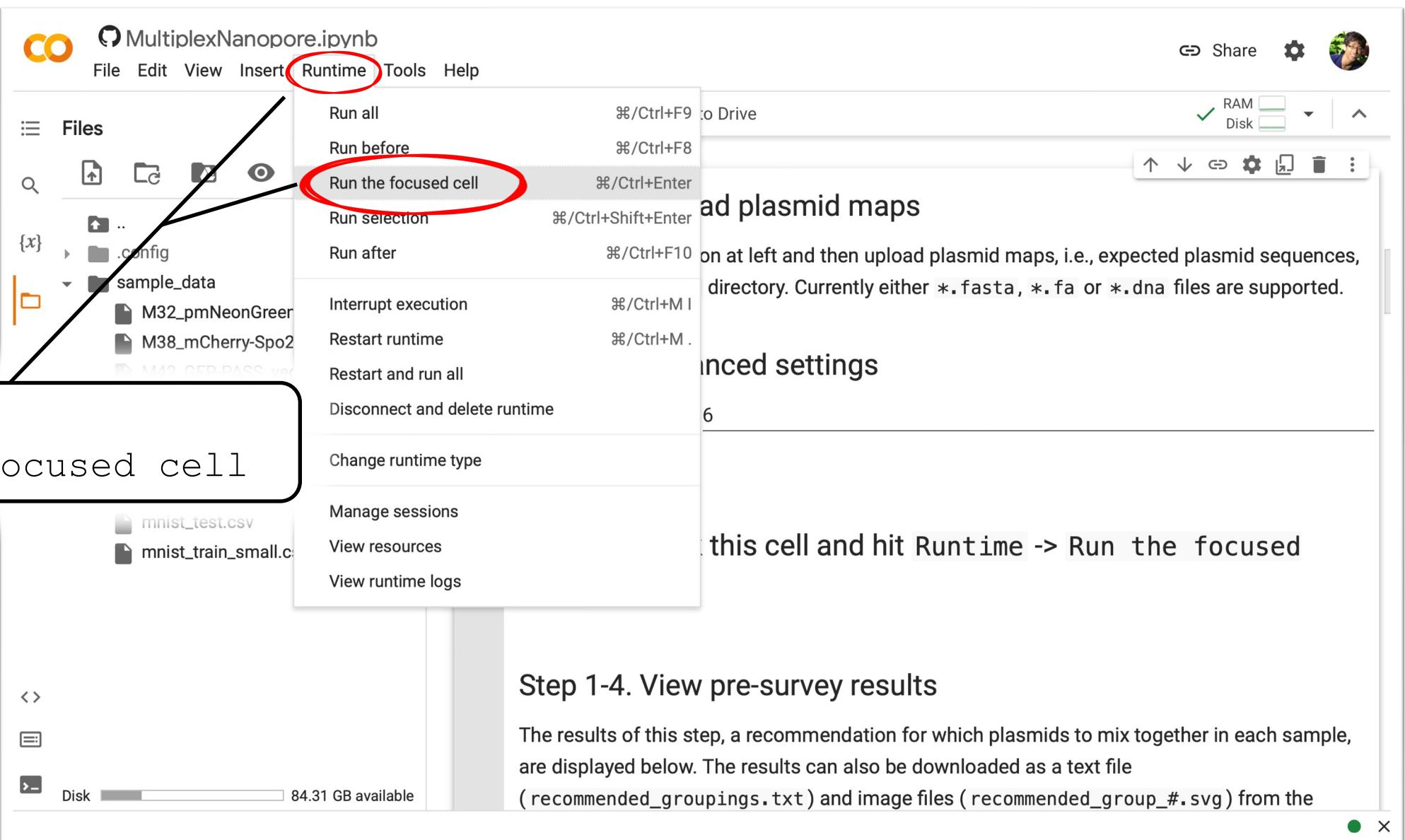
# Execute scripts

The screenshot shows the Google Colab interface with a notebook titled "MultiplexNanopore.ipynb". The sidebar on the left lists files and folders, including a "sample\_data" directory containing several FASTA and CSV files. The main area displays a series of execution steps:

- Step 1-1. Upload plasmid maps**  
Click on the folder icon at left and then upload plasmid maps, i.e., expected plasmid sequences, to the sample\_data directory. Currently either \*.fasta, \*.fa or \*.dna files are supported.
- Step 1-2. Advanced settings**  
threshold\_pre:   
► Description
- Step 1-3. Click this cell and hit Runtime -> Run the focused cell**
- Step 1-4. View pre-survey results**  
The results of this step, a recommendation for which plasmids to mix together in each sample, are displayed below. The results can also be downloaded as a text file (recommended\_groupings.txt) and image files (recommended\_group\_#.svg) from the

A callout box with a black border and white background points to the first step, containing the text "Click anywhere inside this cell".

# Execute scripts



The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** MultiplexNanopore.ipynb
- Menu Bar:** File, Edit, View, Insert, **Runtime**, Tools, Help
- Runtime Menu (highlighted with red circles):**
  - Run all
  - Run before
  - Run the focused cell**
  - Run selection
  - Run after
  - Interrupt execution
  - Restart runtime
  - Restart and run all
  - Disconnect and delete runtime
  - Change runtime type
  - Manage sessions
  - View resources
  - View runtime logs
- File Tree:** Files, sample\_data (containing M32\_pmNeonGreer and M38\_mCherry-Spo2)
- Code Cells:** mnist\_test.csv, mnist\_train\_small.c
- Runtime Status:** RAM 84.31 GB available
- Right Panel:** Contains sections for "ad plasmid maps" (instructions to upload plasmid maps), "Advanced settings" (with a link to step 6), and a note about viewing pre-survey results.

**Hit Runtime  
-> Run the focused cell**

**this cell and hit Runtime -> Run the focused**

**Step 1-4. View pre-survey results**

The results of this step, a recommendation for which plasmids to mix together in each sample, are displayed below. The results can also be downloaded as a text file (recommended\_groupings.txt) and image files (recommended\_group\_#.svg) from the

# Execute scripts

The screenshot shows a Google Colab notebook titled "MultiplexNanopore.ipynb". On the left, there's a sidebar with a "Files" section containing a folder named "sample\_data" which includes several files: M32\_pmNeonGreen-N1.fa, M38\_mCherry-Spo20.fa, M42\_GFP-PASS\_vecCMV.fa, README.md, anscombe.json, california\_housing\_test.csv, california\_housing\_train.csv, mnist\_test.csv, and mnist\_train\_small.csv. Below the sidebar, a "Progress" box contains the text "Processing... 4/9" with a red oval highlighting it. The main workspace shows command-line output for package installations:

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels
Collecting parasail
  Downloading parasail-1.3.4-py2.py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (15.6/15.6 MB 72.8 MB/s eta 0:00:00)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from
Installing collected packages: parasail
Successfully installed parasail-1.3.4
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels
Collecting pycairo
  Downloading pycairo-1.23.0.tar.gz (344 kB) 344.6/344.6 kB 7.0 MB/s eta 0:00:00
Installing build dependencies ... done
Getting requirements to build wheel ... done
Installing backend dependencies ... done
Preparing metadata (pyproject.toml) ... done
Building wheels for collected packages: pycairo
  Building wheel for pycairo (pyproject.toml) ... done
    Created wheel for pycairo: filename=pycairo-1.23.0-cp310-cp310-linux_x86_64.whl
    Stored in directory: /root/.cache/pip/wheels/6d/63/63/4341460df2dca6490f16a239962
Successfully built pycairo
Installing collected packages: pycairo
Successfully installed pycairo-1.23.0
installation. DONE
Processing... 4/9
```

At the bottom of the workspace, a status bar indicates "Executing (1m 18s) <cell line: 627> > main() > pre\_survery() > calc\_distance() > sw\_trace()".

**Step 2. Submit samples**

Mix plasmids according to the [Step 1. Pre-survey](#) results and submit sample(s) for nanopore sequencing.

# Pre-survey results (example 1)

Result files

MultiplexNanopore.ipynb

File Edit View Insert Runtime Tools Help Cannot save changes

Files

sample\_data

- M32\_pmNeonGreen-N1.fa
- M38\_mCherry-Spo20.fa
- M42\_GFP-PASS\_vecCMV.fa
- README.md
- anscombe.json
- california\_housing\_test.csv
- california\_housing\_train.csv
- mnist\_test.csv
- mnist\_train\_small.csv
- recommended\_group\_1.svg
- recommended\_groupings.txt

+ Code + Text Copy to Drive

installation: DONE  
Processing... 9/9  
[[ 0 423 507]  
 [426 0 503]  
 [501 497 0]]  
[2 0 1]  
[[0], [1], [2]]  
[[0, 1, 2]]

#####
# Recommended groupings #
#####

Group1  
P1 : M42\_GFP-PASS\_vecCMV.fa  
P2 : M38\_mCherry-Spo20.fa  
P3 : M32\_pmNeonGreen-N1.fa

query

	P1	P2	P3
P1	0	423	507
P2	426	0	503
P3	501	497	0

reference

RAM Disk

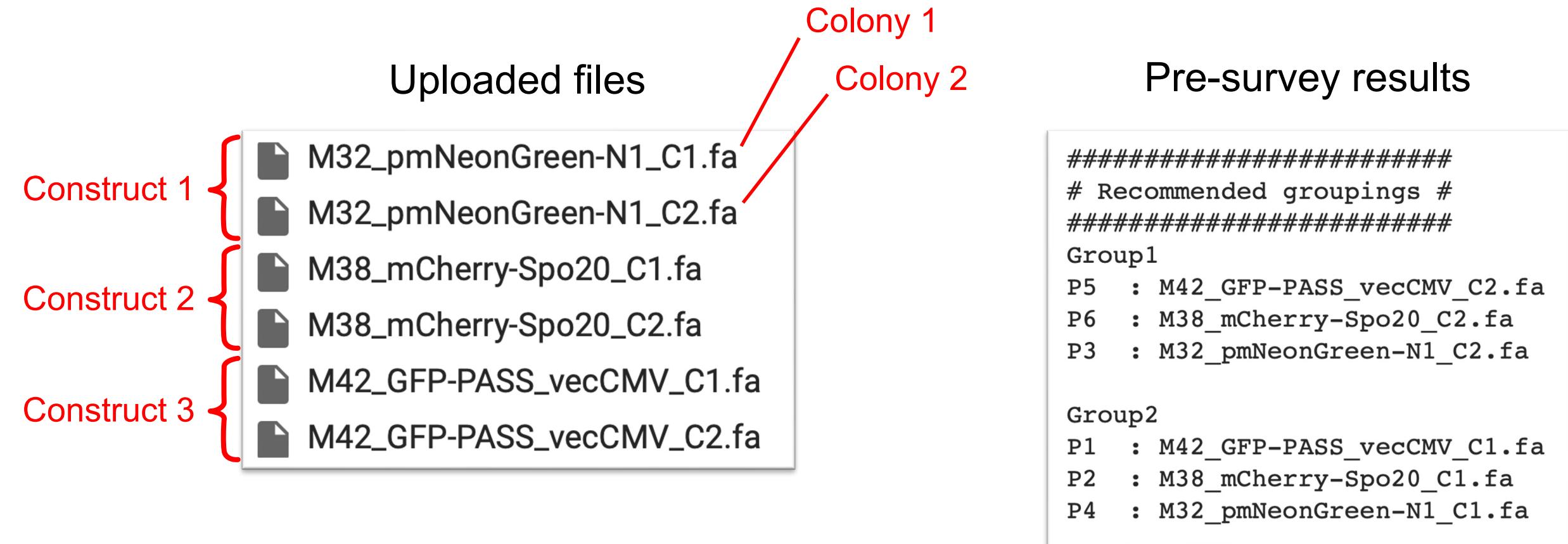
1m 15s completed at 1:22 PM

Recommended grouping  
Only one group is displayed  
→ Safe to mix all together.

Distance matrix  
Small values: similar  
Large values: very different

# Pre-survey results (example 2)

Example of picking multiple colonies from the same plate



Two groups are displayed  
→ Plasmids have to be submitted as two samples.

# How to use SAVEMONEY

Pre-survey

Sample submission

Post-analysis

# Sample submission

Mix plasmids together according to the grouping of the pre-survey step.

e.g.

Sample type	Category	Size	Concentration	Minimum volume	Price per sample	Order now
Plasmid	Standard	2.5 - 25 kb	30 ng/uL	≥10 uL	\$15 or 1 dinocoins	<a href="#">Order</a>

```
#####
# Recommended groupings #
#####
Group1
P1  : M42_GFP-PASS_vecCMV.fa
P2  : M38_mCherry-Spo20.fa
P3  : M32_pmNeonGreen-N1.fa
```

→ Provide 100 ng of each plasmid in 10 µL total of water.

# Request fastq files!

plasmidsaurus

Verify

REQUIRED: Select the highest **Biosafety Level (BSL)** of the organisms from which DNA in this order was extracted

- Select "BSL1" if the samples contain genes from a BSL2 organism but were extracted directly from a non-pathogenic BSL1 host.

BSL1  
 BSL2  
 BSL3 -- please contact us by email before submitting an order

OPTIONAL: I would like a link to download the raw fastq files in my result email (raw reads can now be downloaded from the dashboard as well)

Choose payment method



[\*\*FAQ / DETAILS\*\*](#)

## What data files is delivered?

1. .fasta file (for consensus data): we will provide a clean, complete consensus sequence for each plasmid.
2. .gbk GenBank file (for consensus data): a pLAnnotate map in the GenBank file format.
3. **(OPTIONAL) .fastq file - raw data on reads. Email support to request fastq files.**
4. Histogram file: the histogram file provides a visual representation of the plasmid and raw read data for deeper insight into your samples (image).
5. .html pLAnnotate map (for consensus data): a plasmid map for each sample.

Raw fastq files are required for the post-analysis!

# How to use SAVEMONEY

Pre-survey

Sample submission

Post-analysis

# Post-analysis

Classify each read based on plasmid maps.

Obtain consensus sequences using classified reads.

- Upload plasmid maps
- Upload fastq file
- Execute scripts

# Upload files

\* Make sure there are no unnecessary plasmid maps or fastq files in the `sample_data` directory!

**Uploaded plasmid maps**

**Uploaded fastq file**

**Click anywhere inside this cell**

**MultiplexNanopore.ipynb**

File Edit View Insert Runtime Tools Help Cannot save changes

Files

+ Code + Text Copy to Drive

RAM Disk

**Step 3. Post-analysis**

This step must be performed independently for each experimental sample mixture. Follow **Step 3-1** and **Step 3-2** below to execute the post-analysis. All processes are executed automatically.

After the completion of all processes, a `fastq_file_name.zip` file will appear in the `sample_data` directory. Right-click on the zip file and press `Download` to download it to your local storage. The `fastq_file_name.zip` file will also be automatically uploaded to your Google Drive, unless you uncheck `save_to_google_drive` option at **Step 3-4** before execution.

**Step 3-1. Upload files**

The following two types of files must be uploaded to the `sample_data` directory, which is accessed by clicking on the file folder icon at left:

- the plasmid map for each plasmid within the sample mixture (`*.fasta`, `*.fa`, or `*.dna`)
- the raw nanopore sequencing results (`*.fastq`)

If your results are split into multiple `*.fastq` files, just upload all of them; SAVEMONEY will merge those files and execute the analysis.

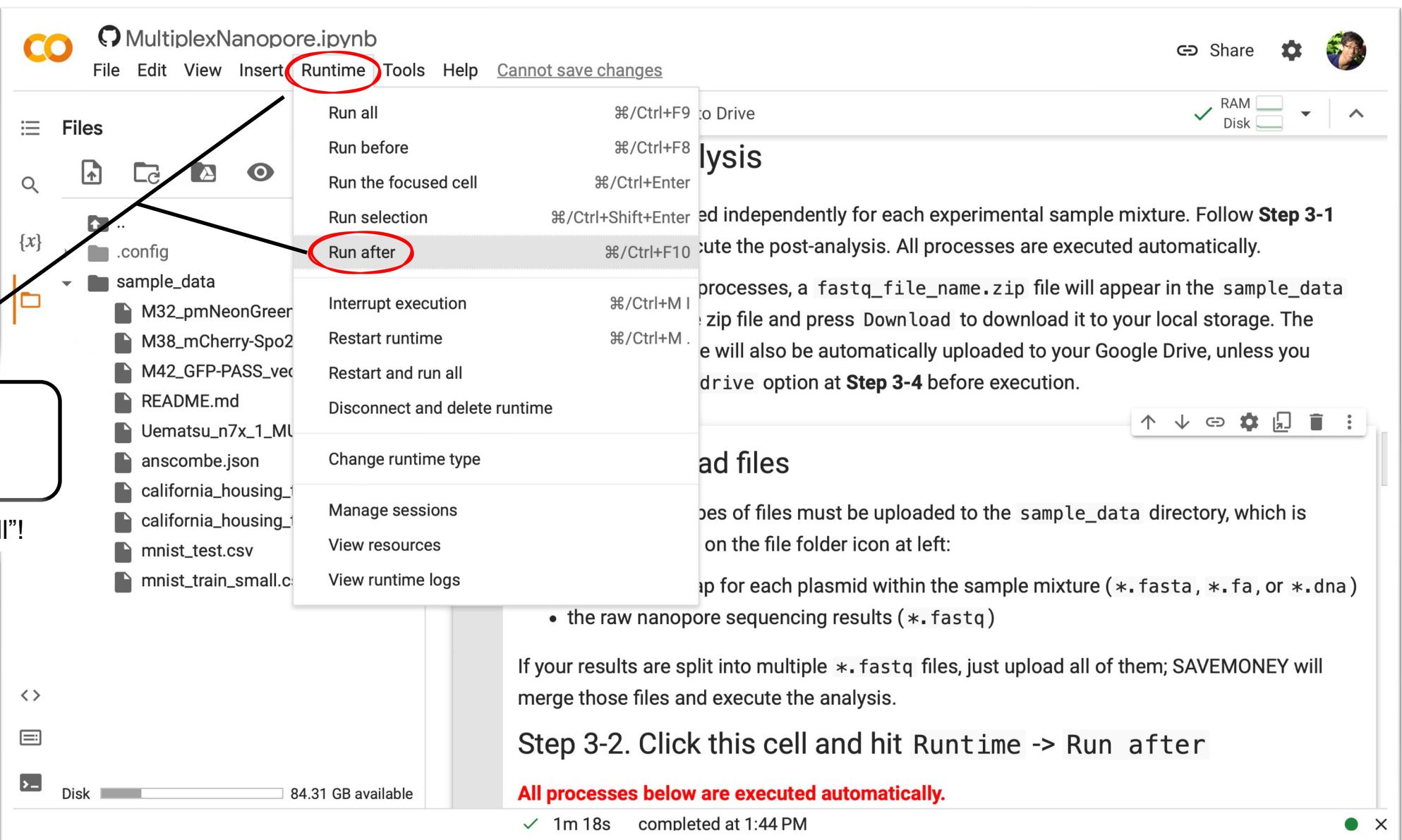
**Step 3-2. Click this cell and hit Runtime -> Run after**

**All processes below are executed automatically.**

84.31 GB available

1m 18s completed at 1:44 PM

# Execute scripts



The screenshot shows the Google Colab interface with a file named "MultiplexNanopore.ipynb". The "Runtime" menu is open, and the "Run after" option is highlighted with a red circle. A callout box on the left says "Hit Runtime -> Run after" with a note "\* Not ‘run focused cell’!". The main content area contains text about running analysis independently for each experimental sample mixture, mentioning Step 3-1 and Step 3-4.

Hit Runtime  
-> Run after

\* Not “run focused cell”!

MultiplexNanopore.ipynb

File Edit View Insert **Runtime** Tools Help Cannot save changes

Files

sample\_data

- M32\_pmNeonGreen
- M38\_mCherry-Spo2
- M42\_GFP-PASS\_vec
- README.md
- Uematsu\_n7x\_1\_ML
- anscombe.json
- california\_housing
- california\_housing
- mnist\_test.csv
- mnist\_train\_small.c

Run all ⌘/Ctrl+F9

Run before ⌘/Ctrl+F8

Run the focused cell ⌘/Ctrl+Enter

Run selection ⌘/Ctrl+Shift+Enter

**Run after** ⌘/Ctrl+F10

Interrupt execution ⌘/Ctrl+M I

Restart runtime ⌘/Ctrl+M .

Restart and run all

Disconnect and delete runtime

Change runtime type

Manage sessions

View resources

View runtime logs

to Drive

RAM Disk

ysis

ed independently for each experimental sample mixture. Follow **Step 3-1** to execute the post-analysis. All processes are executed automatically.

processes, a `fastq_file_name.zip` file will appear in the `sample_data` folder. Download this zip file and press `Download` to download it to your local storage. The zip file will also be automatically uploaded to your Google Drive, unless you choose a different drive option at **Step 3-4** before execution.

ad files

oes of files must be uploaded to the `sample_data` directory, which is indicated by the orange folder icon at left:

- up for each plasmid within the sample mixture (`*.fasta`, `*.fa`, or `*.dna`)
- the raw nanopore sequencing results (`*.fastq`)

If your results are split into multiple `*.fastq` files, just upload all of them; SAVEMONEY will merge those files and execute the analysis.

**Step 3-2. Click this cell and hit Runtime -> Run after**

**All processes below are executed automatically.**

✓ 1m 18s completed at 1:44 PM

# Execute scripts

The progress is displayed at the bottom of Step 3-4 and 3-6.

RESTART RUNTIME

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels
Collecting snapgene-reader
  Downloading snapgene_reader-0.1.20-py3-none-any.whl (6.9 kB)
Collecting biopython (from snapgene-reader)
  Downloading biopython-1.81-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64
  3.1/3.1 MB 41.7 MB/s eta 0:00:00
Collecting xmltodict (from snapgene-reader)
  Downloading xmltodict-0.13.0-py2.py3-none-any.whl (10.0 kB)
Collecting html2text (from snapgene-reader)
  Downloading html2text-2020.1.16-py3-none-any.whl (32 kB)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from snapgene-reader)
Installing collected packages: xmltodict, html2text, biopython, snapgene-reader
Successfully installed biopython-1.81 html2text-2020.1.16 snapgene-reader-0.1.20
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels
Collecting parasail
  Downloading parasail-1.3.4-py2.py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64
  15.6/15.6 MB 29.6 MB/s eta 0:00:00
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from parasail)
Installing collected packages: parasail
Successfully installed parasail-1.3.4
installation: DONE
Executing alignment: 11 out of 336 / Uematsu_n7x_1_MU-test1_13)
```

Step 3-5. Set threshold for assignment

Disk 84.20 GB available

threshold\_pos: 0.5

Executing (1m 43s) <cell line: 562> > execute\_alignment() > align\_all() > sw\_trace()

# Post-analysis results

The screenshot shows a Google Colab notebook titled "MultiplexNanopore.ipynb". The left sidebar displays the file structure:

- ..
- .config
- sample\_data
  - Uematsu\_n7x\_1\_MU-test1\_subset
  - M32\_pmNeonGreen-N1.fa
  - M38\_mCherry-Spo20.fa
  - M42\_GFP-PASS\_vecCMV.fa
  - README.md
  - Uematsu\_n7x\_1\_MU-test1\_subset.fastq
  - Uematsu\_n7x\_1\_MU-test1\_subset.zip
  - anscombe.json
  - california\_housing\_test.csv
  - california\_housing\_train.csv
  - mnist\_test.csv
  - mnist\_train\_small.csv

Annotations on the left side point to specific files:

- "Uploaded plasmid maps" points to the "sample\_data" folder.
- "Uploaded fastq file" points to "Uematsu\_n7x\_1\_MU-test1\_subset.fastq".
- "Result zip file" points to "Uematsu\_n7x\_1\_MU-test1\_subset.zip".

The main content area shows the output of a cell:

```
+ Code + Text Copy to Drive  
refseq No. 0  
[6] 5965 out of 5965  
refseq No. 1  
7507 out of 7507  
refseq No. 2  
6204 out of 6204  
16s Step 3-7. Export results  
See Interpretation of results \(post-analysis\) for the detailed description of result files.  
Show code  
Exporting logs...  
Exporting alignment results...  
Exporting alignment summary...  
Exporting summary...  
Exporting summary svg images...  
Exporting consensus fastq files...  
export: DONE  
Uploaded /content/sample_data/Uematsu_n7x_1_MU-test1_subset.zip to Google Drive
```

**Step 4. Visualization of results (optional)**

Please refer to the following sections to see in which cases this cell must be executed.

- "Alignment TEXT files" section in [Interpretation of results \(post-analysis\)](#)

0s completed at 3:45 PM

# Topics

1. Introduction
2. Concept of SAVEMONEY
3. How to use SAVEMONEY
4. How to interpret outputs

# Post-analysis results

## Output files

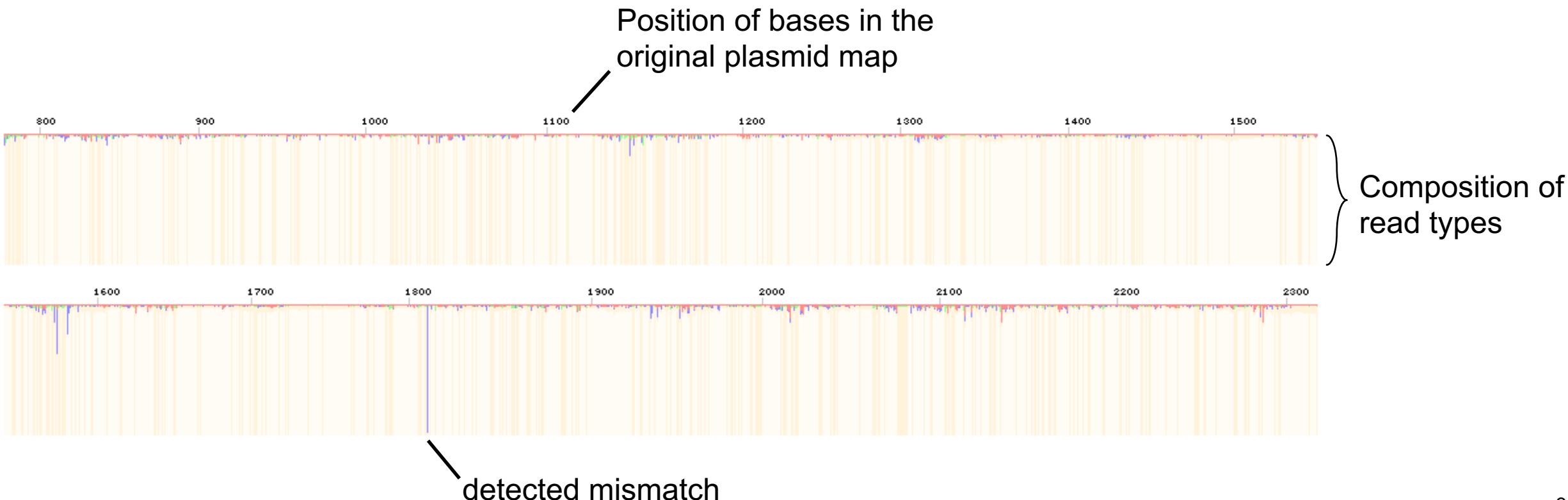
- Overall summary and log files.
    - summary\_distribution.svg
    - summary\_scatter.svg
    - summary\_scores.txt
    - fastq\_file\_name.intermediate\_results.txt
    - log\_with\_prior.txt
    - log\_without\_prior.txt
  - Main result files
    - plasmid\_map\_name.gif
    - plasmid\_map\_name.consensus\_with\_prior.fastq
    - plasmid\_map\_name.consensus\_without\_prior.fastq
  - Files used for “[Step 4. Visualization of results \(optional\)](#)”
    - plasmid\_map\_name.alignment\_with\_prior.txt
    - plasmid\_map\_name.alignment\_without\_prior.txt
- 
- The diagram uses red curly braces to group the output files into three categories:
  - A brace on the right side groups the first two items under the heading "Summary before/during the classification of reads".
  - A brace on the right side groups the next two items under the heading "Results after the read classification".
  - A brace on the right side groups the last two items under the heading "Files used for Step 4. Visualization of results (optional)".

# Main result files (GIF files)

## **plasmid\_map\_name.gif**

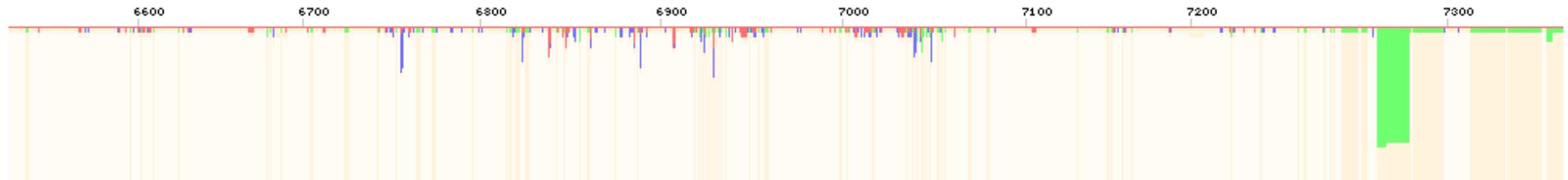
- A summary overview image of each plasmid's results.
- Always check this file to assert there are no significant differences between the plasmid map and the actual sample.

■ Deletion  
■ Insertion  
■ Mismatch  
■ Omitted  
■ Match

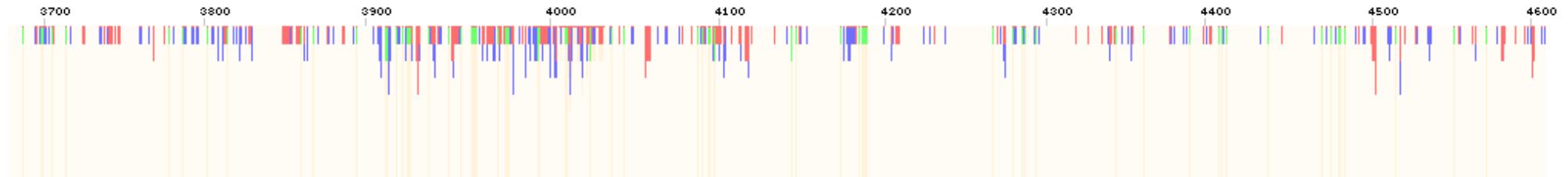


# Example of bad GIF files

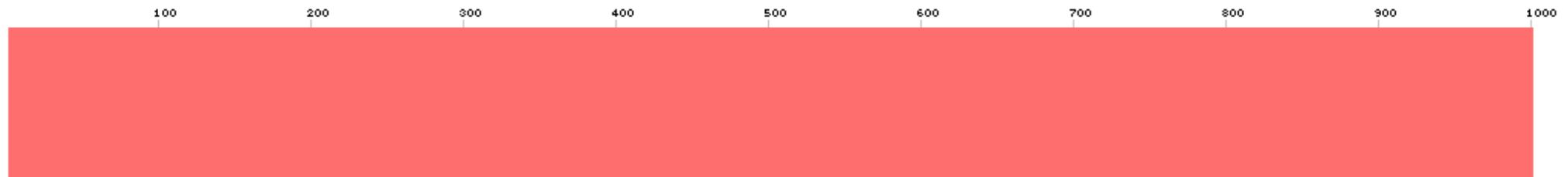
Large insertion



Too noisy



No reads at all



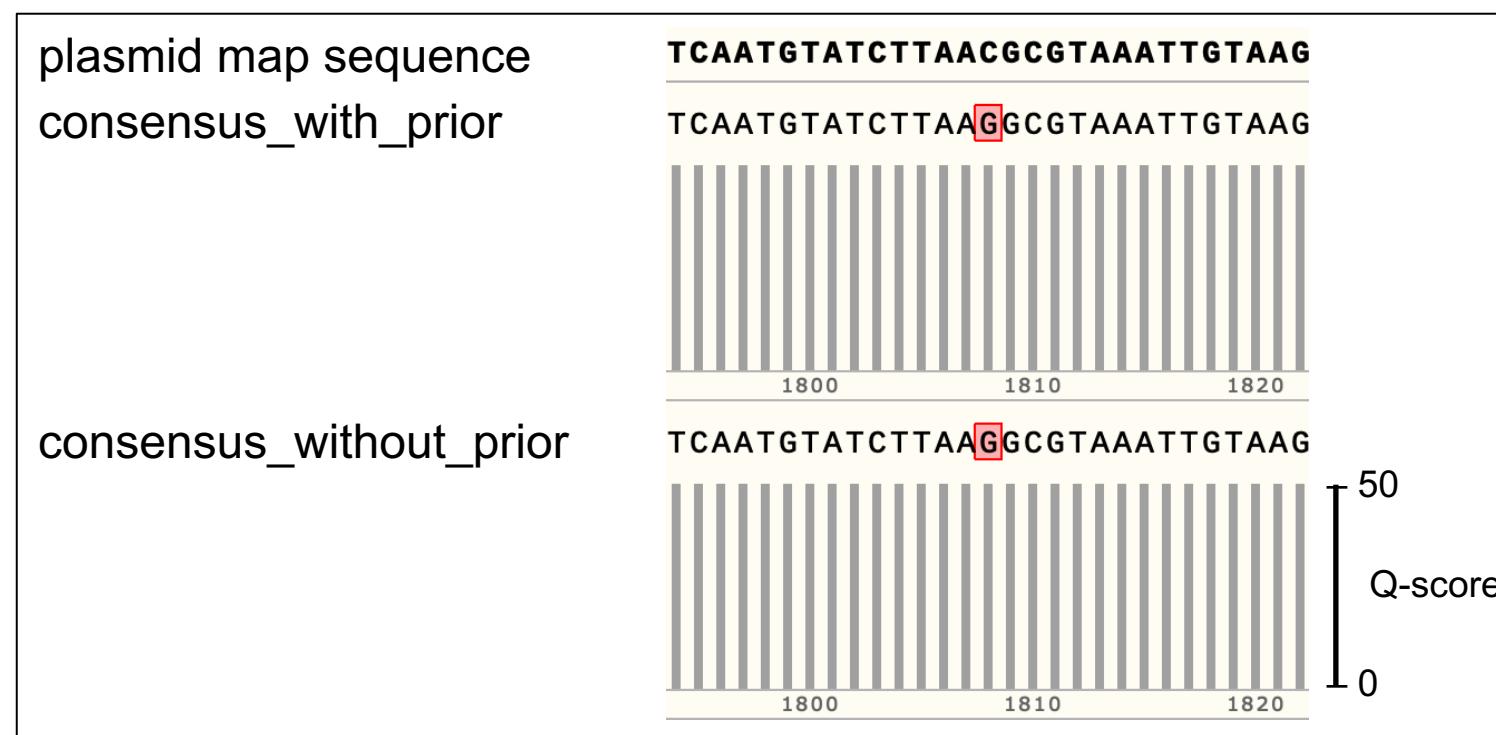
- █ Deletion
- █ Insertion
- █ Mismatch
- █ Omitted
- █ Match

Something is wrong with your sample.  
→ See “[Troubleshooting](#)” section on the Google Colab page.

# Main result files (FASTQ files)

## **plasmid\_map\_name.consensus\_with/without\_prior.fasta**

- Consensus sequences generated for each plasmid maps.
- Use these files to perform alignment with sequence analysis software.
- They usually return the same results.

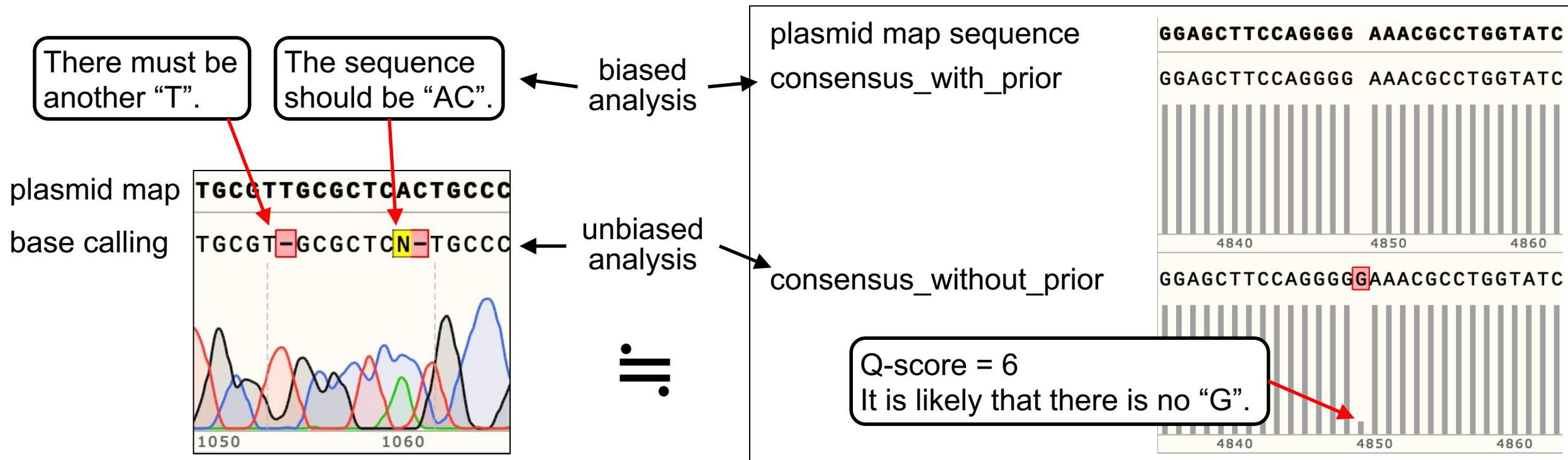


One mutation was found with high (Q-score  $\geq 50$ ) confidence.

# Difference of two FASTQ files

## **plasmid\_map\_name.consensus\_with/without\_prior.fastq**

- “with\_prior” file is statistically biased toward the reference plasmid map.
- It is roughly equivalent to manual inspection of chromatograms in Sanger sequencing.



# Notes

Use SAVEMONEY to confirm that the plasmid sequence is correct.  
It is NOT intended to identify plasmid with unknown sequence.

- The algorithm is designed on the assumption that the plasmid sequence is almost the same as the plasmid map.
- Avoid using SAVEMOENY if:
  - Plasmids in sample may have a large difference from the plasmid map.
  - There are concerns of correctness.
- Restriction Enzyme digestion test is strongly recommended beforehand.

Plasmids with identical maps cannot be mixed.

- Multiple colonies sharing the same plasmid map does not decrease the cost.
- There must be at least 2 nt differences between plasmids.

In practice...

- Coordination between multiple colleagues in a lab may be required.

# More info

SAVEMONEY (Google Colab)

<https://colab.research.google.com/github/MasaakiU/MultiplexNanopore/blob/master/colab/MultiplexNanopore.ipynb>

SAVEMOENY (GitHub)

<https://github.com/MasaakiU/MultiplexNanopore>

Publication

Masaaki U and Jeremy M. Baskin

[Barcode-free multiplex plasmid sequencing using Bayesian analysis and nanopore sequencing. bioRxiv, 2023](#)