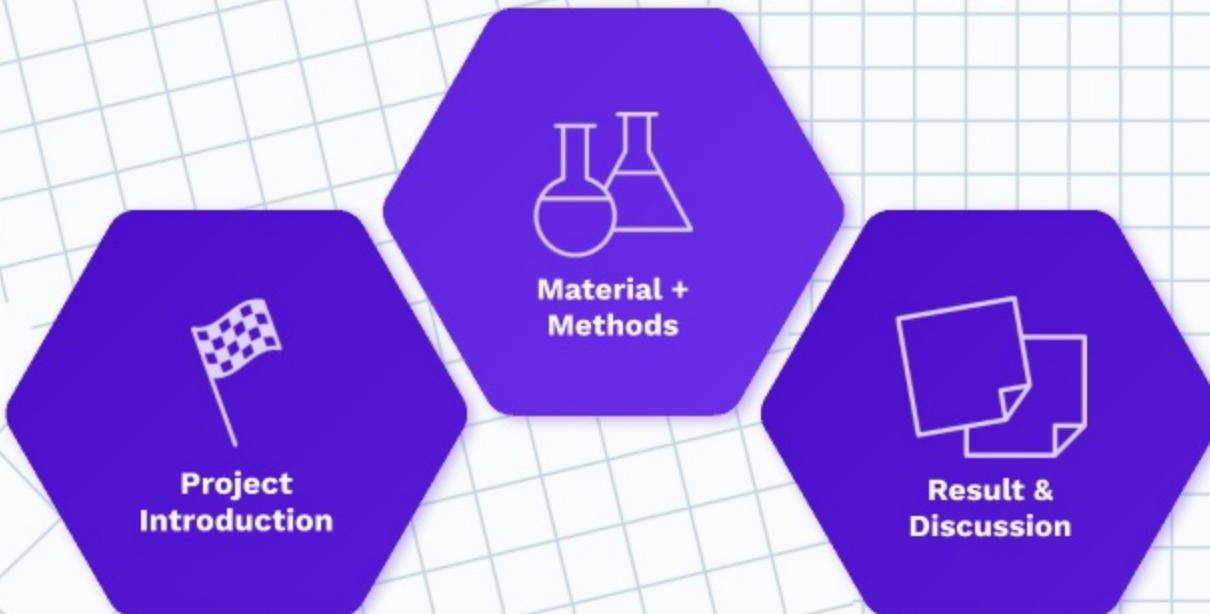


Expanding the DNA-probe toolbox for molecular profiling of tissues and their microbiomes



Project Introduction

- Sequencing



- Spatial Transcriptomics (ST)



- MERFISH

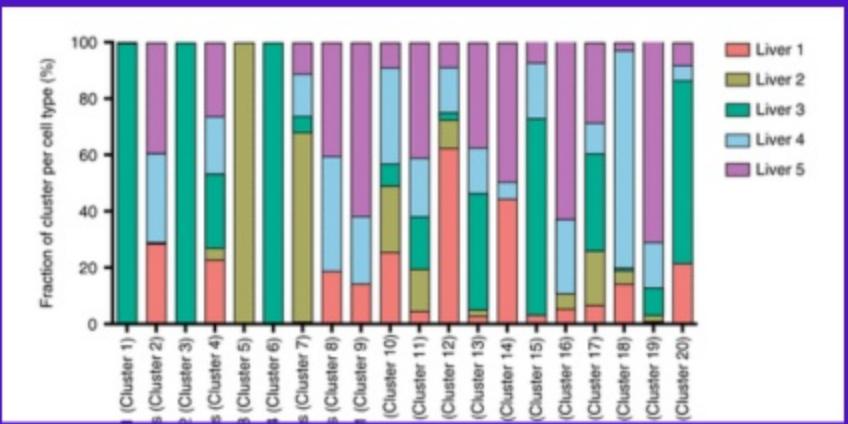


Problem Statement

Sequencing: Determining the order of nucleotides in a DNA or RNA molecule.

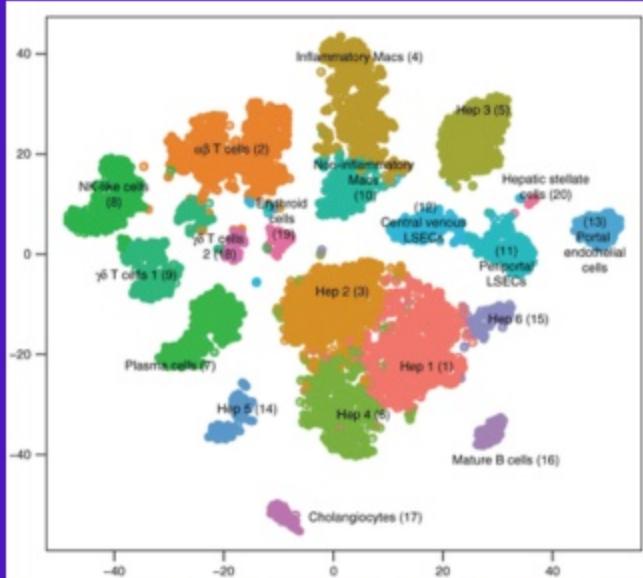
Single-cell sequencing: analysing the genetic material of individual cells.

scRNA-seq: analysing the gene expression profiles of individual cells at the RNA level.



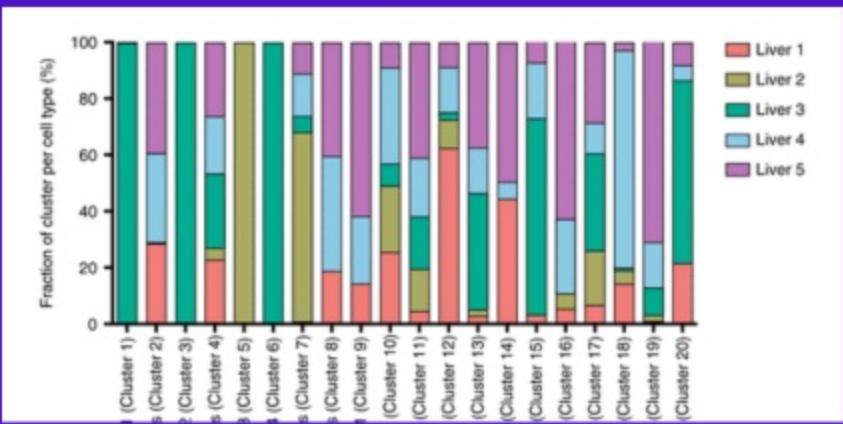
The proportion of cells that contributed to each cluster by liver sample.

(Contribution of Cells to Each scRNA-seq Cluster by Sample And... n.d.)



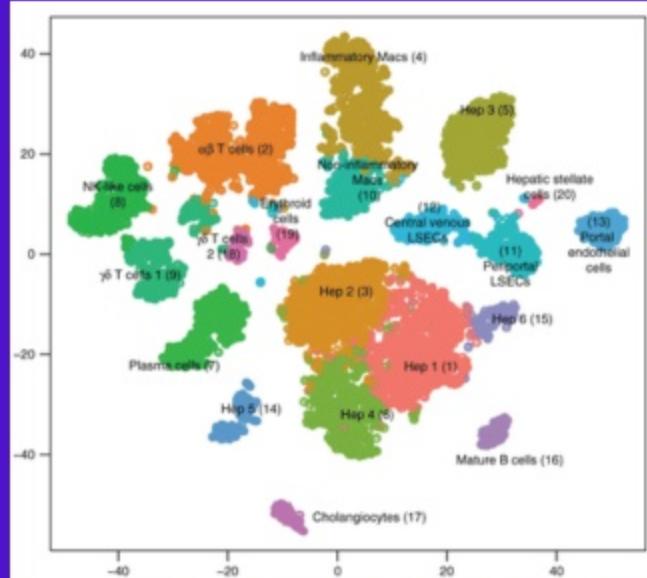
Cluster map showing the assigned identity for each cluster defined cells that share similar transcriptome

(20 Distinct Cell Populations Were Revealed in Healthy Human Livers.
And... n.d.)



The proportion of cells that contributed to each cluster by liver sample.

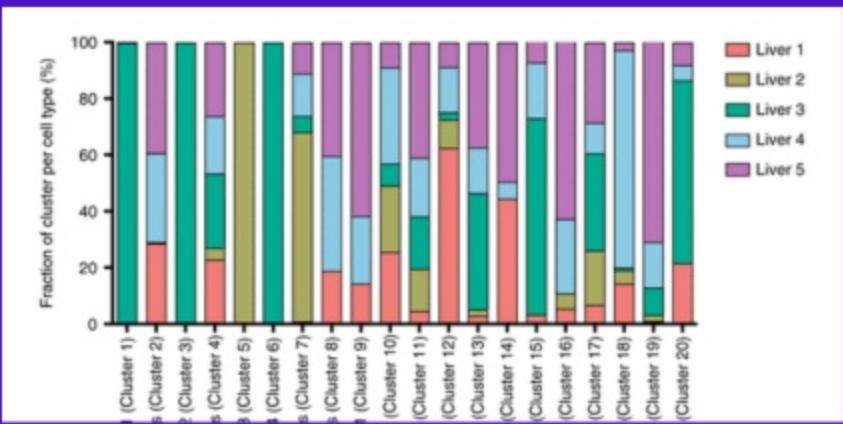
[\(Contribution of Cells to Each scRNA-seq Cluster by Sample And... n.d.\)](#)



Cluster map showing the assigned identity for each cluster defined cells that share similar transcriptome

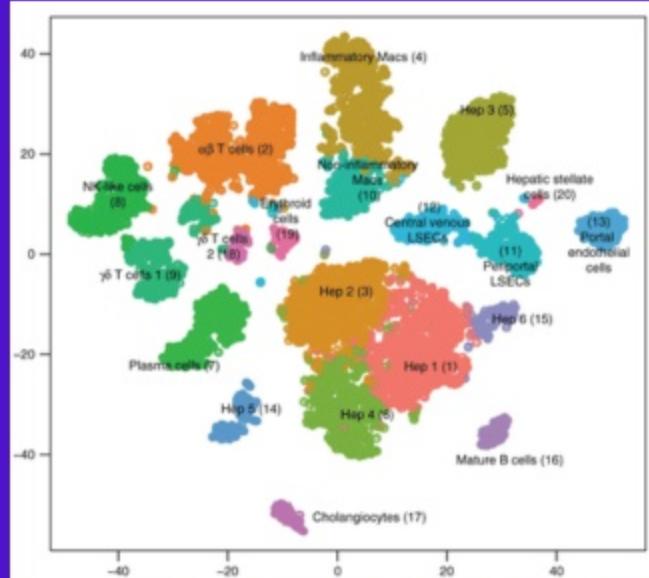
[\(20 Distinct Cell Populations Were Revealed in Healthy Human Livers. And... n.d.\)](#)

Spatial information is missing



The proportion of cells that contributed to each cluster by liver sample.

(Contribution of Cells to Each scRNA-seq Cluster by Sample And... n.d.)



Cluster map showing the assigned identity for each cluster defined cells that share similar transcriptome

(20 Distinct Cell Populations Were Revealed in Healthy Human Livers.
And... n.d.)

Spatial information is **missing**
Why?

- sequencing + spatial data
- localisation of cell types and their associate gene expression
- two primary approaches:
 - + sequencing-based
 - + imaging-based

involves isolating mRNAs from tissue samples while preserving spatial information and profiling mRNA species

Sequencing methods

Arrays

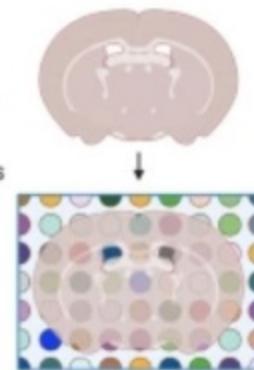
1. Array of spatially barcoded probes



2. Image barcode locations via ISS



4. NGS of captured probes



3. Overlay sample on array. Ligate mRNA to probes.

Microdissection



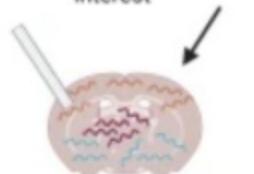
1. Hybridise probes to target mRNAs



2. Image region of interest



4. NGS of captured probes

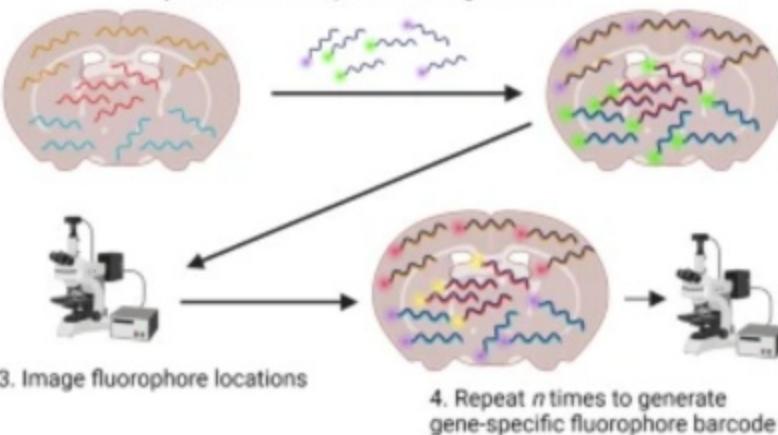


3. Release probes from ROI with UV. Capture via capillary.

Imaging methods

ISH

1. Hybridise labelled probes to target mRNA



visualising the **spatial distribution** of RNA molecules using **microscopy images**

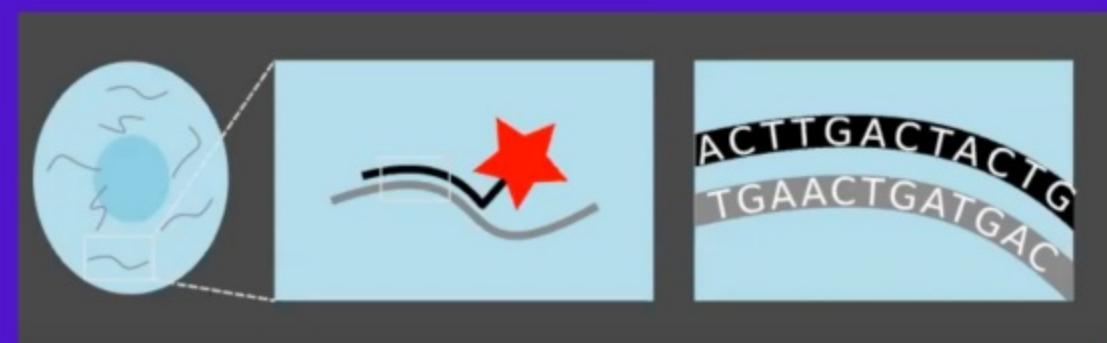
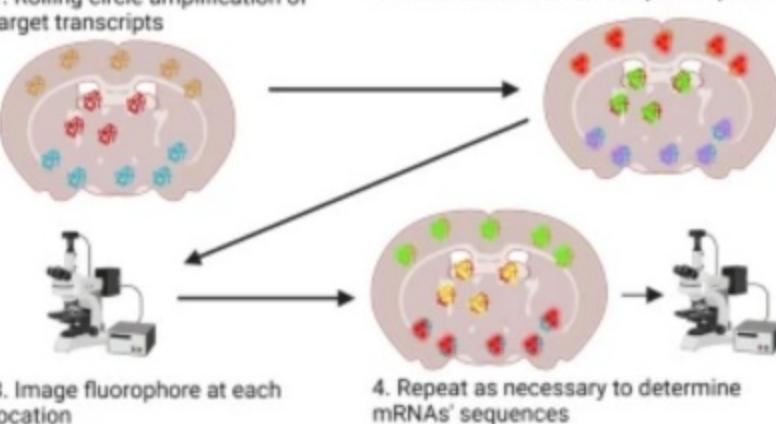
two main types:

- + in situ sequencing (ISS): identify RNA molecules by sequencing it while it's in the tissue
- + in situ hybridisation (ISH):

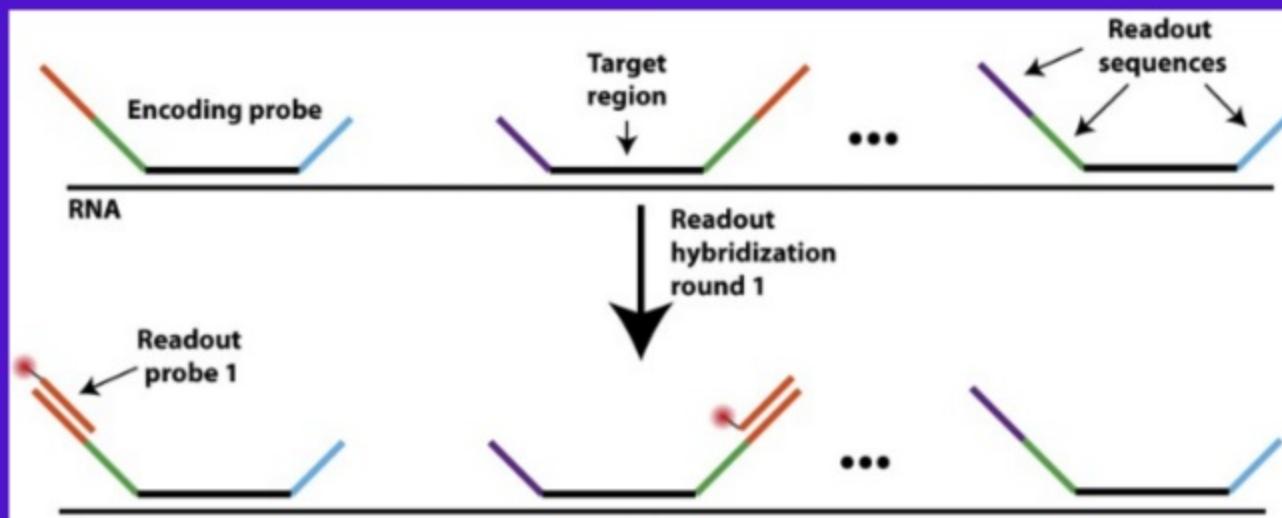
ISS

1. Rolling circle amplification of target transcripts

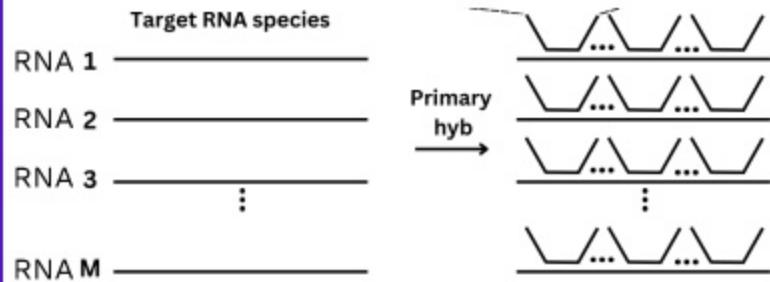
2. Hybridise short, labelled probes to determine 1-2nt of transcript's sequence



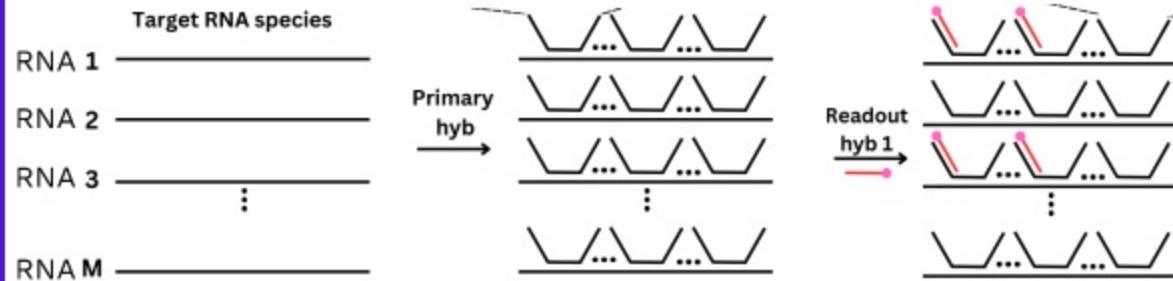
Multiplexed Error-Robust Fluorescence *in situ* Hybridisation



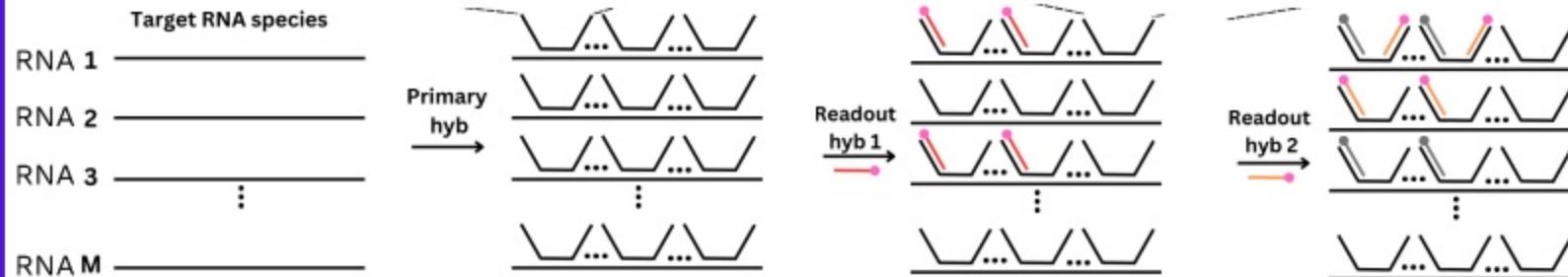
Multiplexing



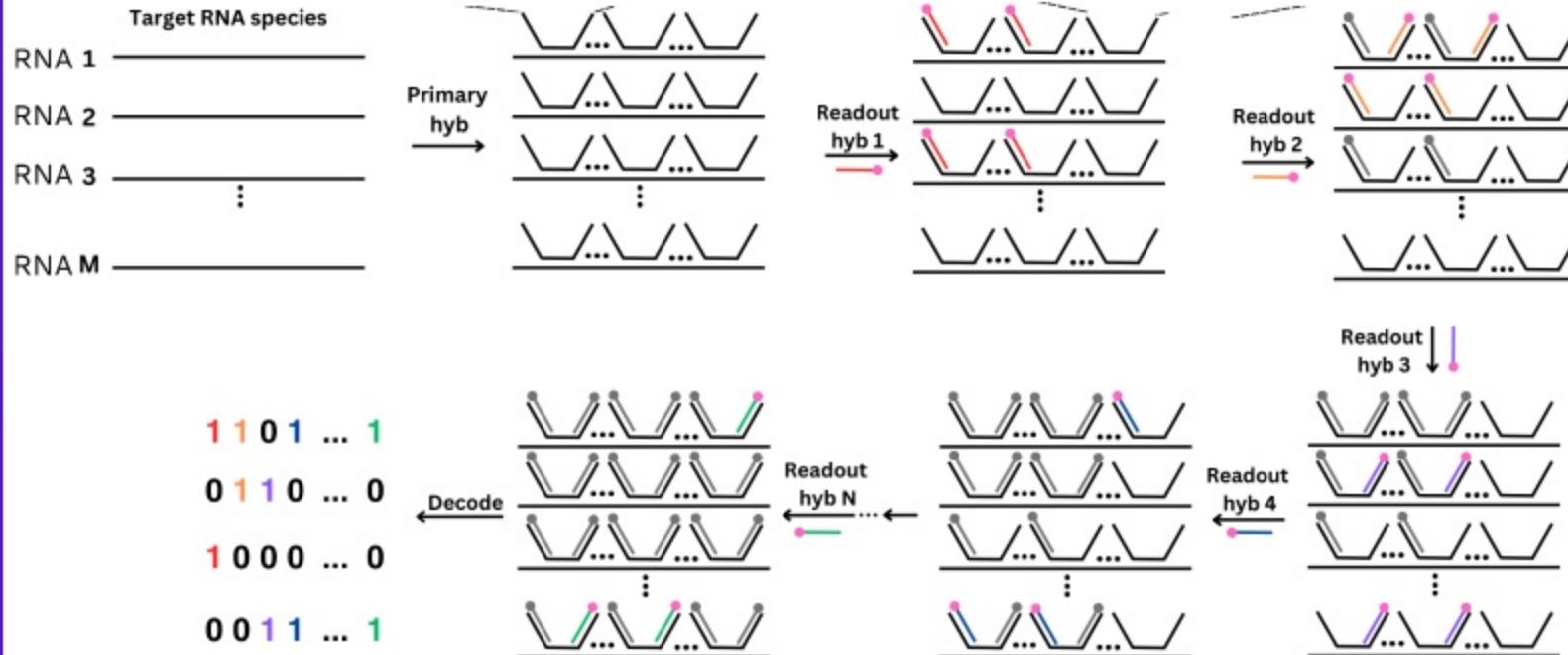
Multiplexing



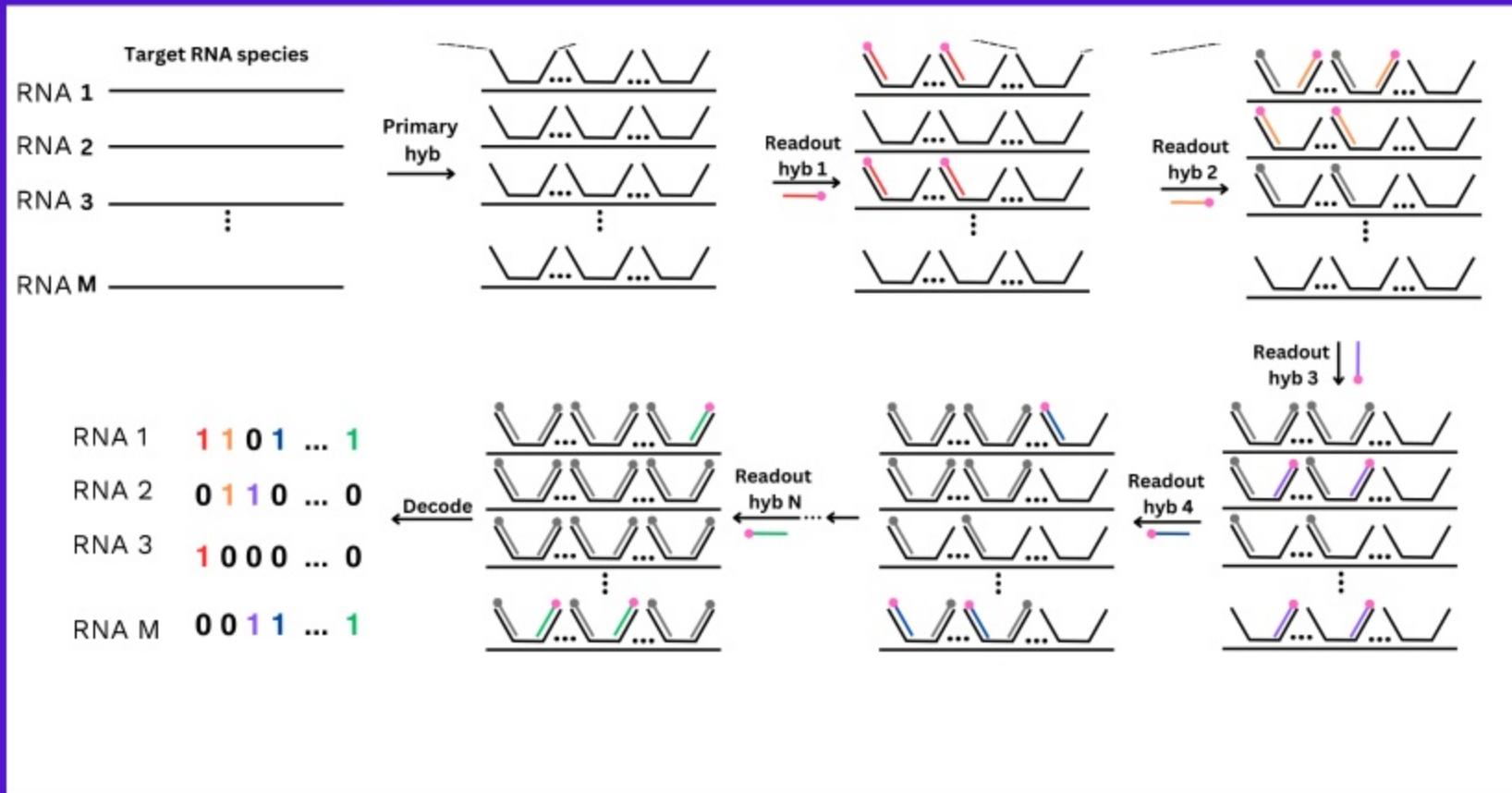
Multiplexing



Multiplexing



Multiplexing



Error - robust

Single error correction

A 15-bit binary sequence is shown in two rows. The top row is 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1. The bottom row is identical except the 5th bit from the left is highlighted in red as 0. A red arrow points down to this red '0'. A green curved arrow starts at the end of the sequence and loops back to the 5th bit.

0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1

Double error detection

A 15-bit binary sequence is shown in two rows. The top row is 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 1. The bottom row is identical except the 5th and 11th bits from the left are highlighted in red as 0. Two red arrows point down to these red '0's.

0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1

Thesis objective

I aim to develop a **pipeline to design probes** for MERFISH experiment aiming to investigate microbes in lungs tissue.

Why the lungs?

a diverse and dynamic community of microbes in the lungs

ST techniques could provide insights into the spatial distribution of the microbes in lung tissue

Project Introduction

- Sequencing



- Spatial Transcriptomics (ST)

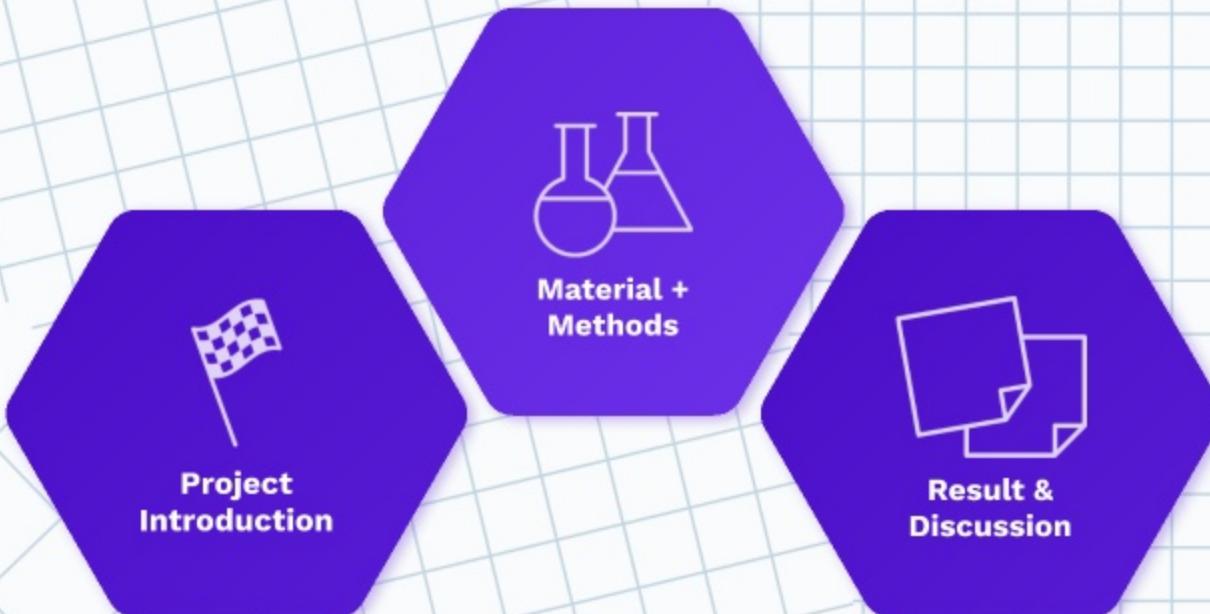


- MERFISH



Problem Statement

Expanding the DNA-probe toolbox for molecular profiling of tissues and their microbiomes



Research pipeline



Classification of microbes

Data: metagenomic DNA sequences obtained from bronchoalveolar lavage (BAL) fluid samples (Campos et al., 2023)

Tools: FastQC + Trimmomatic + Kraken2 + Bracken + NCBI Datasets

- Classify microbes from data
- Download their **references genome** (use for target probe design)

Probe design

Codebook

0	1	0	0
0	0	1	0

Hamming weight: 1

Hamming distance 2

4-bit

Probe design

Codebook

0	1	0	0
0	0	1	0

Hamming weight: 1
Hamming distance 2
4-bit

La Jolla Covering Table

C(v=16, k=4, t=3)

1	2	3	4
1	2	5	6
1	2	7	8
1	2	9	10
1	2	11	12
1	2	13	14
...			

Hamming weight: 4

Hamming distance 4

16-bit

16 readout sequence will be assigned for
16 position of the "1" in each code

Probe design

Codebook

0	1	0	0
0	0	1	0

Hamming weight: 1
Hamming distance 2
4-bit

La Jolla Covering Table

C(v=16, k=4, t=3)

1	2	3	4
1	2	5	6
1	2	7	8
1	2	9	10
1	2	11	12
1	2	13	14
...			

Hamming weight: 4
Hamming distance 4
16-bit

16 readout sequence will be assigned for
16 position of the "1" in each code

Encoding probe composition



Target sequence

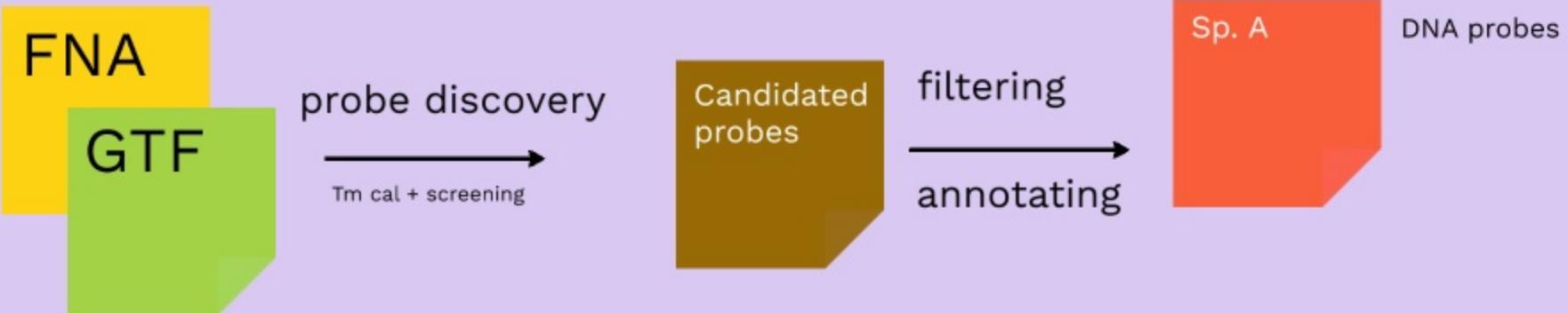
Data: Reference genomes of the lungs' microbiome

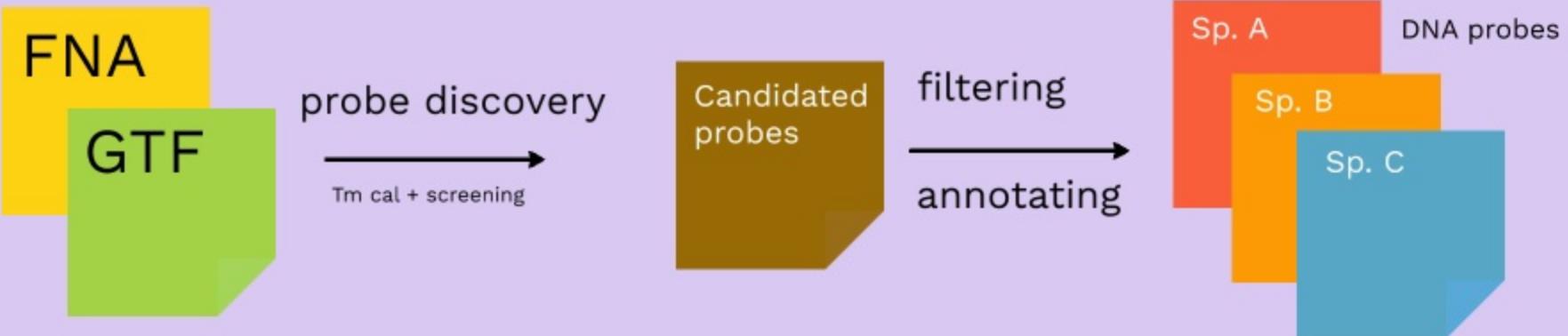
Tools: PaintSHOP & BLAST+

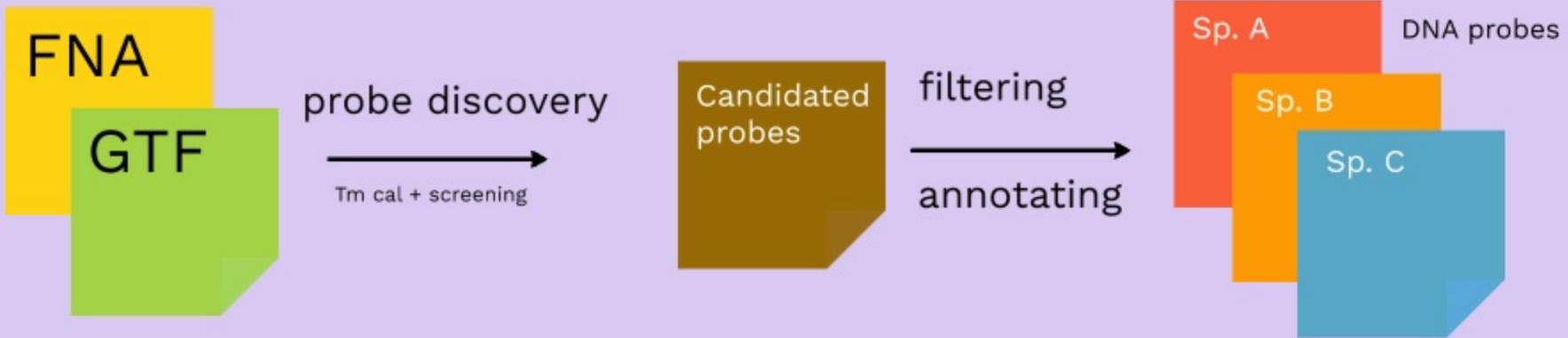
Input: FNA + GTF + probe's parameters

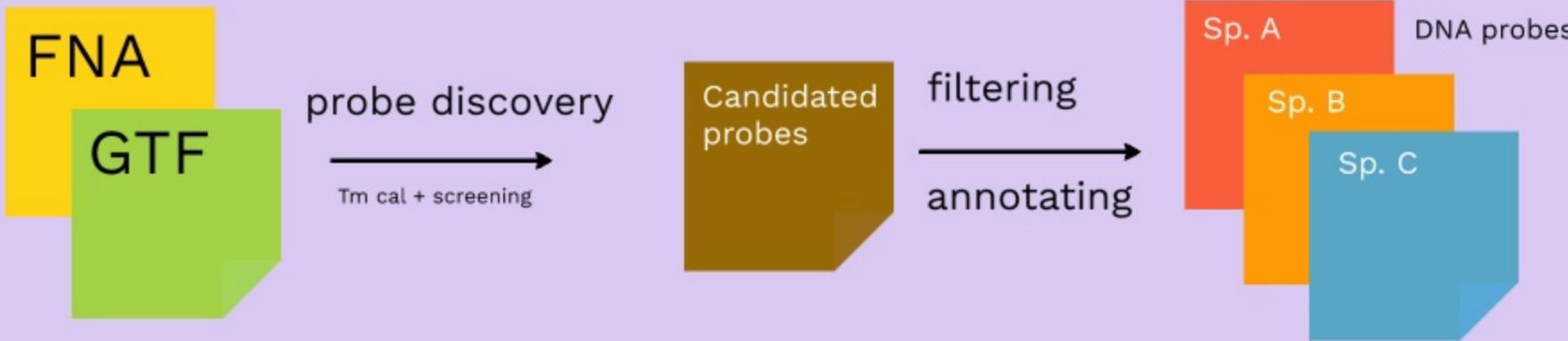


Output: "Unique" probes









Readout probe & sequences

Input: 20-nt ATG
G~50%

Primers

25-nt orthogonal
sequence (Xu et
al., 2009)

Preliminary filtering
+
Screening against
human RNA and
microbes genome

Readout probe & sequences

Input: 20-nt ATG
G~50%

Primers

25-nt orthogonal
sequence (Xu et
al., 2009)

Preliminary filtering
+
Screening against
human RNA and
microbes genome

Output: 16 readout
sequences + probes

20nt Primers

Readout probe & sequences

Input: 20-nt ATG
G~50%

Primers

25-nt orthogonal
sequence (Xu et
al., 2009)

reduce # of sequences

Preliminary filtering
+
Screening against
human RNA and
microbes genome

Output: 16 readout
sequences + probes

20nt Primers

Readout probe & sequences

Input: 20-nt ATG
G~50%

Primers

25-nt orthogonal
sequence (Xu et
al., 2009)

reduce # of sequences

Preliminary filtering
+

Screening against
human RNA and
microbes genome

Tm
GC clamp
hairpin formation
GC content

Output: 16 readout
sequences + probes

20nt Primers

microbes' list

primers

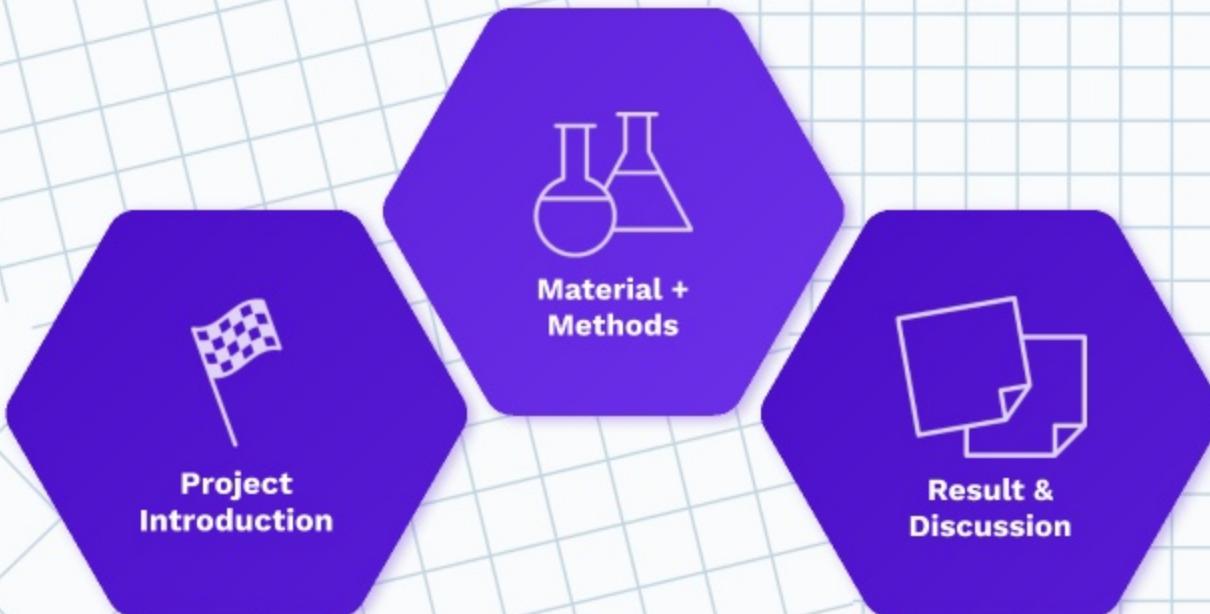
codebook

encoding
probe
table

readouts

target probes

Expanding the DNA-probe toolbox for molecular profiling of tissues and their microbiomes



Encoding probe table

91 species - over 4300 encoding probes

Pipeline

The end?

Specie	Chr	Primer_1	Readout_1	Re
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT		AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT		AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AC
Acinetobacter_baumannii	NZ_CP043953.1	CGTGTAGTGGCCCGGGTCT		AC

Readout_1	Readout_2	Target	Read
	AGATGGATAGGGTTATGAGT	AAAGAAGCCAGACAATTCTGCCGGATGGC	AGAT
AGATTAGGAGGGTAGTTATG		ACATCCGGAGTCGAGGTTTCGCTTACACCC	AGAT
AGATTAGGAGGGTAGTTATG	AGATGGATAGGGTTATGAGT	TGTTCTTAGTCTCGTGTAGGGTCCGGGCT	
AGATTAGGAGGGTAGTTATG	AGATGGATAGGGTTATGAGT	TTACAGAGCCCGTACTTAACCGCAGCTC	AGAT
	AGATGGATAGGGTTATGAGT	ATACACCGCCATGTGGAGTCACAAGCTCAC	AGAT
AGATTAGGAGGGTAGTTATG		CAAGTGCCTAGGCACATCCCTGTATGCG	AGAT
AGATTAGGAGGGTAGTTATG	AGATGGATAGGGTTATGAGT	AGCAATTGCAACTACGGTCGTGCTTGCTCA	
AGATTAGGAGGGTAGTTATG	AGATGGATAGGGTTATGAGT	TGACGTTGCCACACATCGCTAAACTGCGG	AGAT
	AGATGGATAGGGTTATGAGT	AGGGTATACTGGTCCAACTATGCGGTGT	AGAT

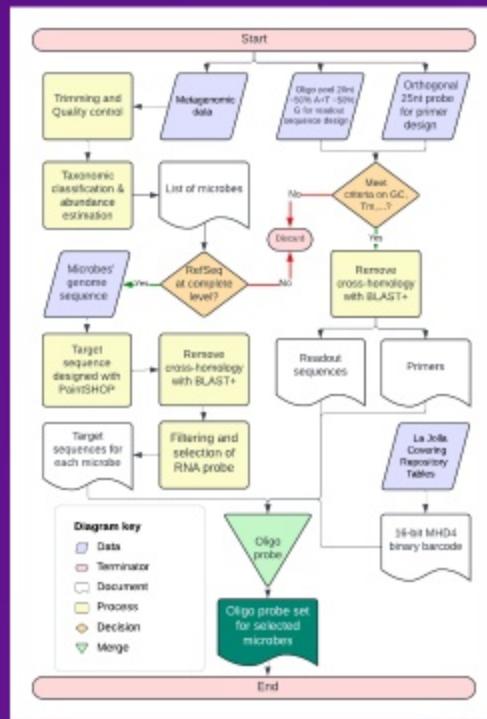
	Readout_3	Readout_4	Primer_2
CCGGATGGC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
CTTACACCC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
TCCGGGCT		GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
CGCAGCTC	AGATGTAGGTTAGGTGAGAG		GGCCGCGACTAGGTAAGCCT
AAGCTCAC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
TGTATGCG	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
GCTTGCTCA		GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT
AACTGCGG	AGATGTAGGTTAGGTGAGAG		GGCCGCGACTAGGTAAGCCT
ATGCGGTGT	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGAGGGTTT	GGCCGCGACTAGGTAAGCCT

Encoding probe table

Species	Chr	Primer_1	Readout_1	Readout_2	Target	Readout_3	Readout_4	Primer_2
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT		AGATGGATAGGTTATGAGT	AAAGAAGCCAGACAAATTGTCGGATGGC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG		ACATCGGAGTCAGGTTGCCTAACGCC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	TGTTCTAGTCGTTAGGGTCGGGGCT	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	TTACAGAGCGCGTACTTTAACCGCAGCTC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	ATACACGCCATGTGGAGTCACAAGCTAC	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	CAAGTGCGCTAGCACATCCCTGTATGCG	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	AGCAATTGCAACTACGGTGTGCTGCTCA	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT	AGATTAGGAGGGTAGTTATG	AGATGGATAGGTTATGAGT	TGACGTTGCCACACATCGCTAAACCTGCGG	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT
Acinetobacter_baumannii	NZ_CPO43953.1	CGTGTAGTGGCCCGGGTCT		AGATGGATAGGTTATGAGT	AGGGTATACTGTCGGTCCAACTATCGCGTGT	AGATGTAGGTTAGGTGAGAG	GTTAGAGAGAGAGGGTTT	GGCCGGGACTAGGTAAAGCT

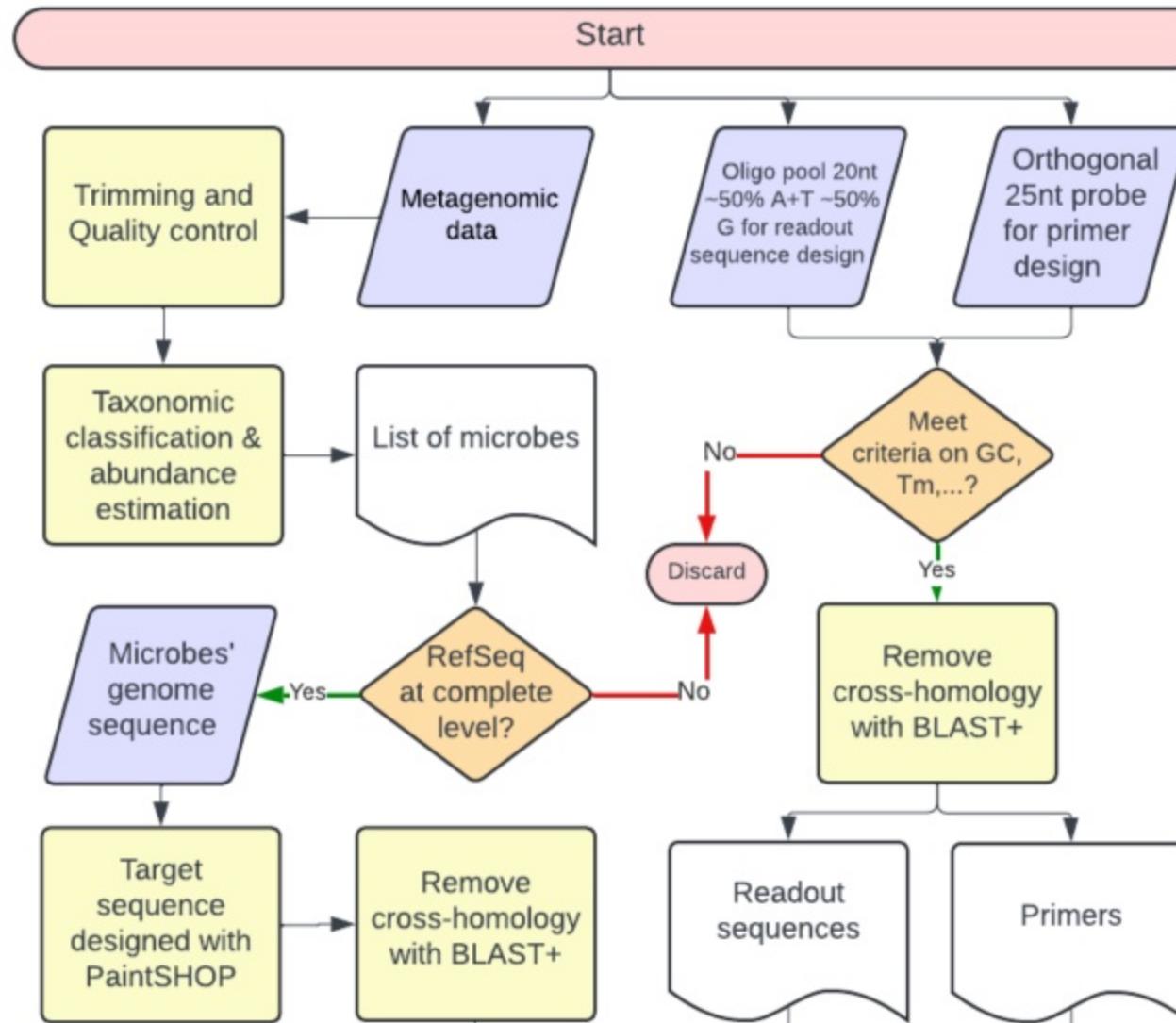
91 species - over 4300 encoding probes

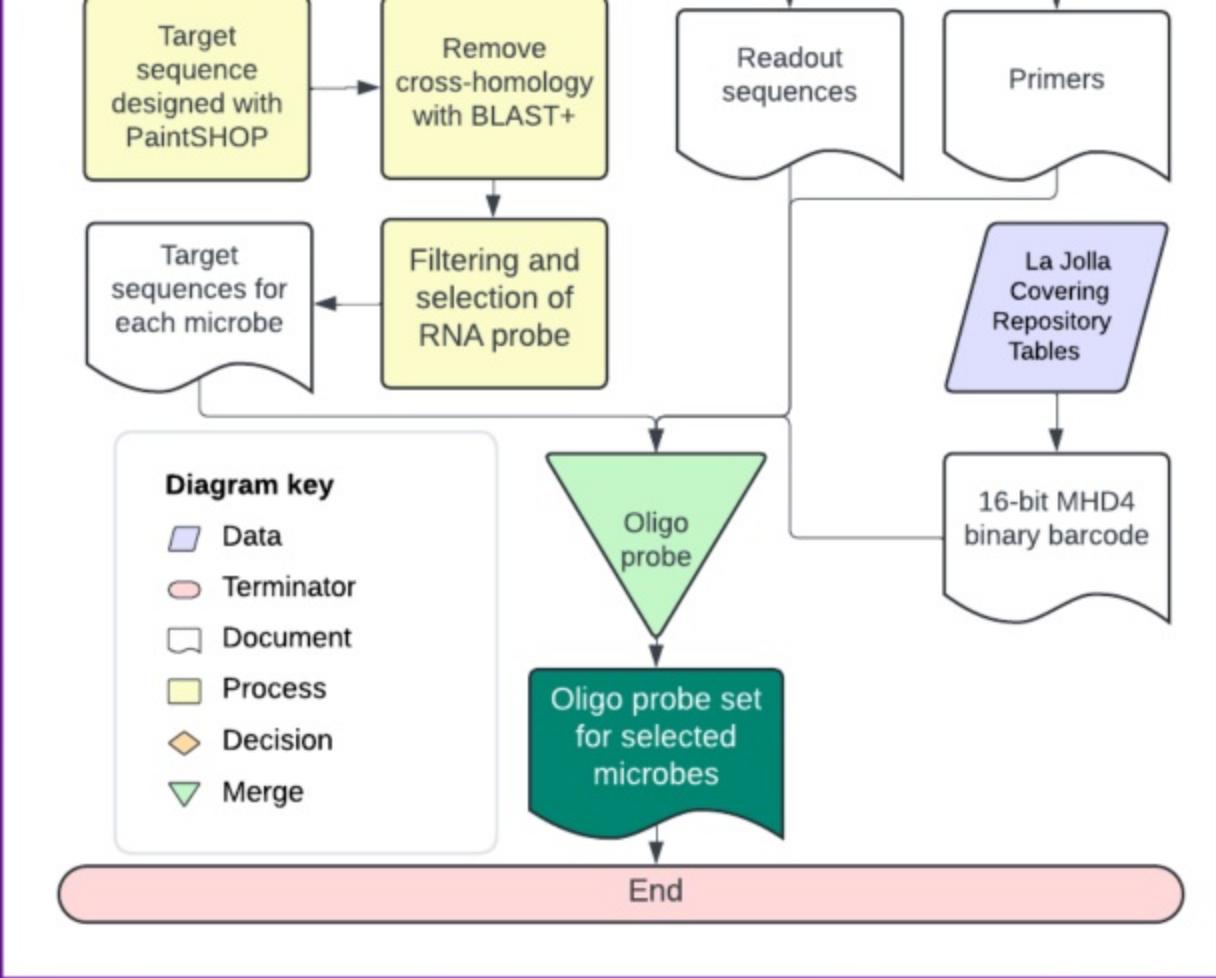
Pipeline of probe design



https://github.com/npxhuy/MeREFISH_probeDesign

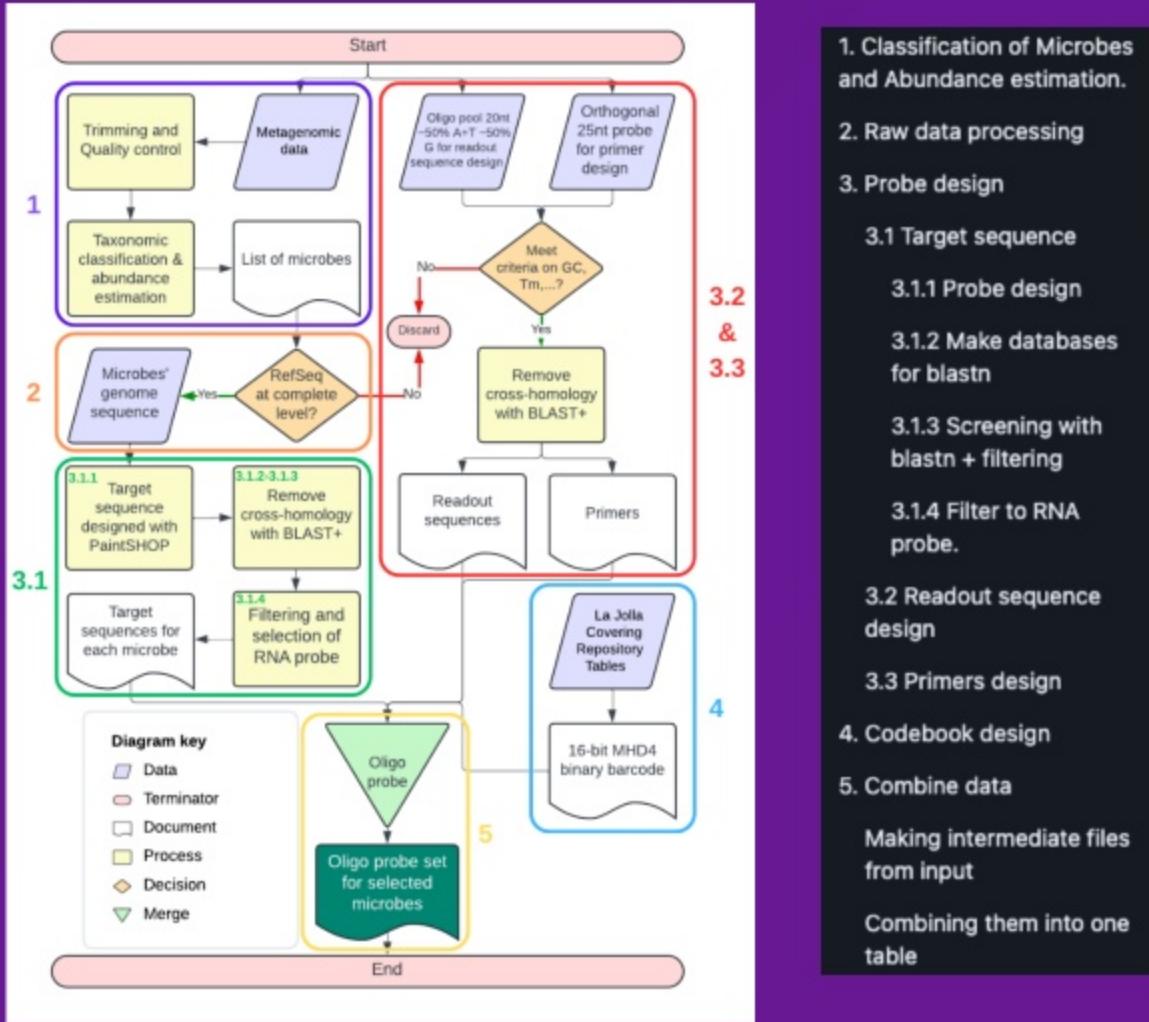






[https://github.com/npxhuy/
MeRFISH_probeDesign](https://github.com/npxhuy/MeRFISH_probeDesign)





Encoding probe table

91 species - over 4300 encoding probes

Pipeline

The end?

Acknowledgement

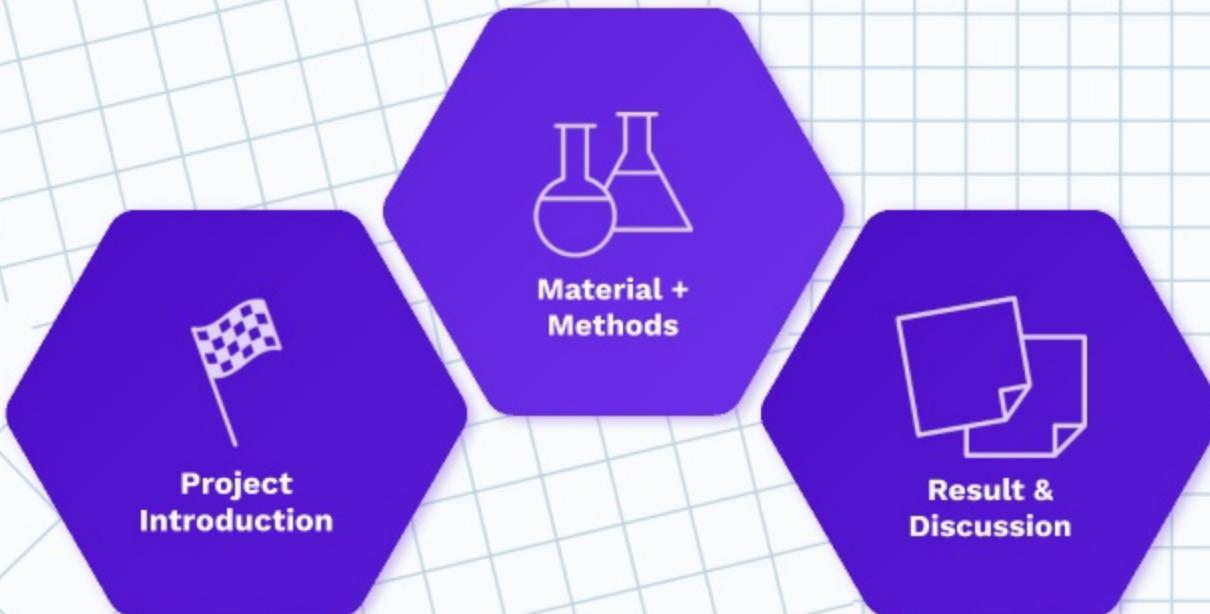
Encoding probe table

91 species - over 4300 encoding probes

Pipeline

The end?

Expanding the DNA-probe toolbox for molecular profiling of tissues and their microbiomes



Expanding the DNA-probe toolbox for molecular profiling of tissues and their microbiomes

