What’s transcriptomics ?

TRANSCRIPTOMICS

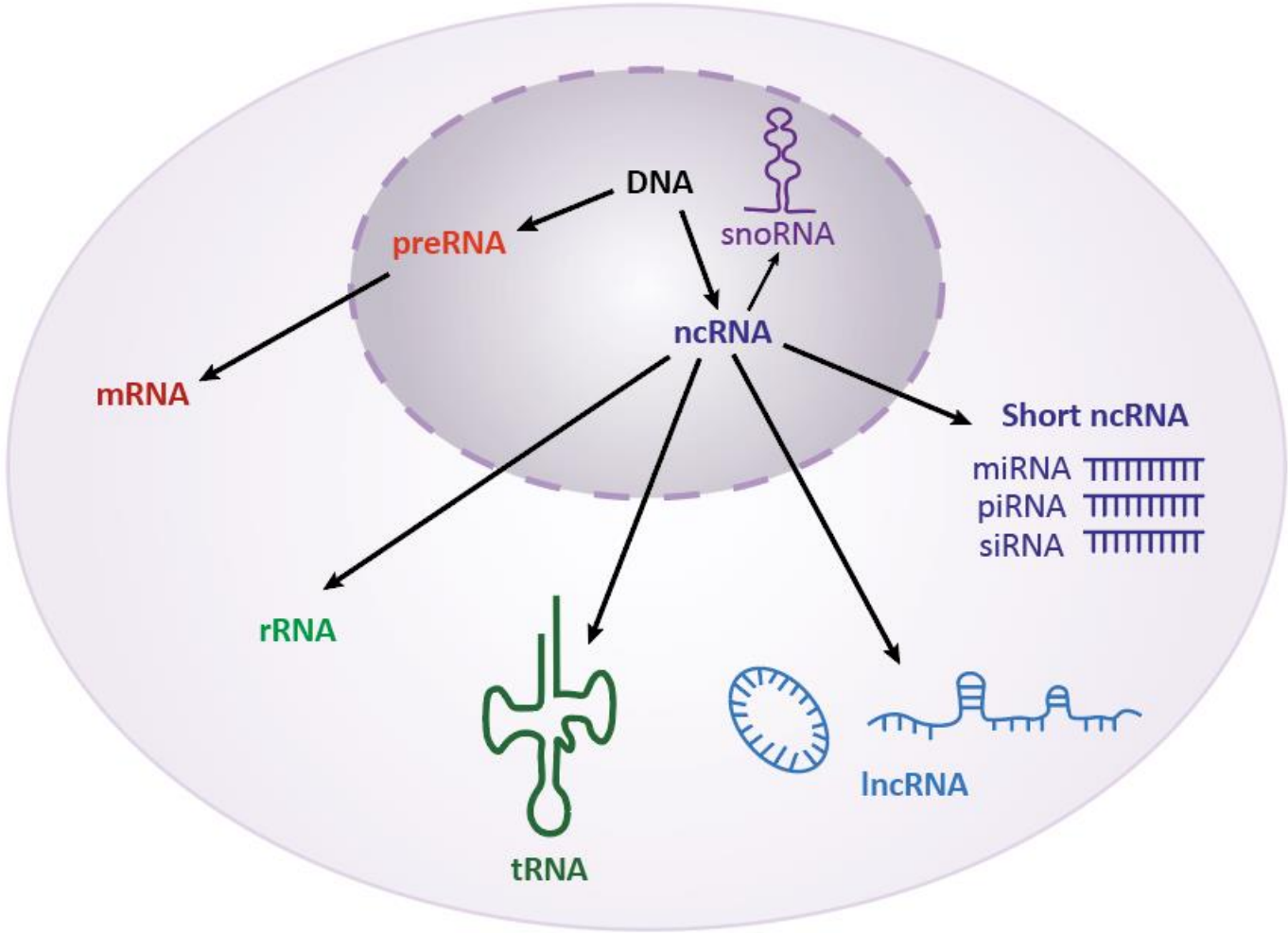
BIGNAUD AMAURY
22/02/2023

What’s transcriptomics ?

- It’s the studied of the **transcriptome**.

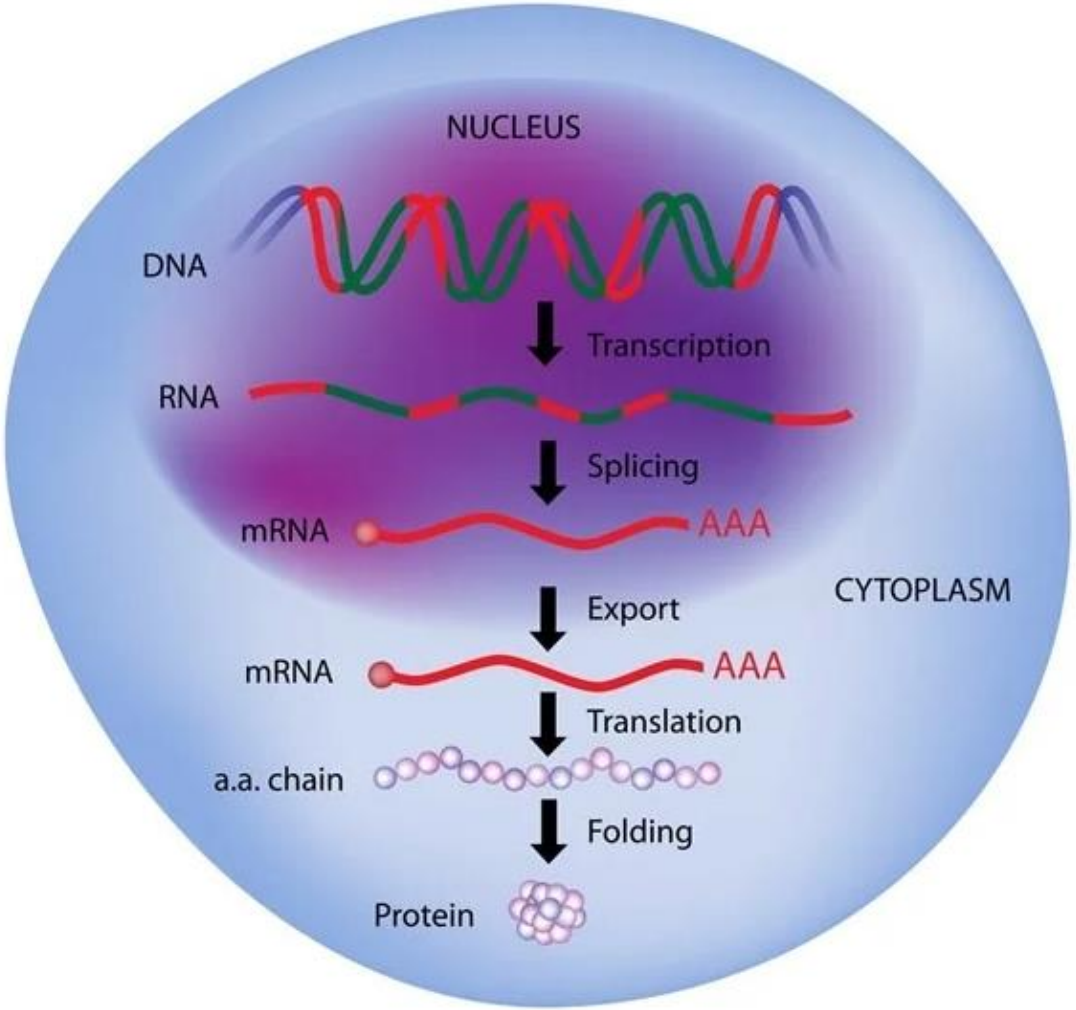
What’s transcriptomics ?

- It’s the studied of the **transcriptome**.



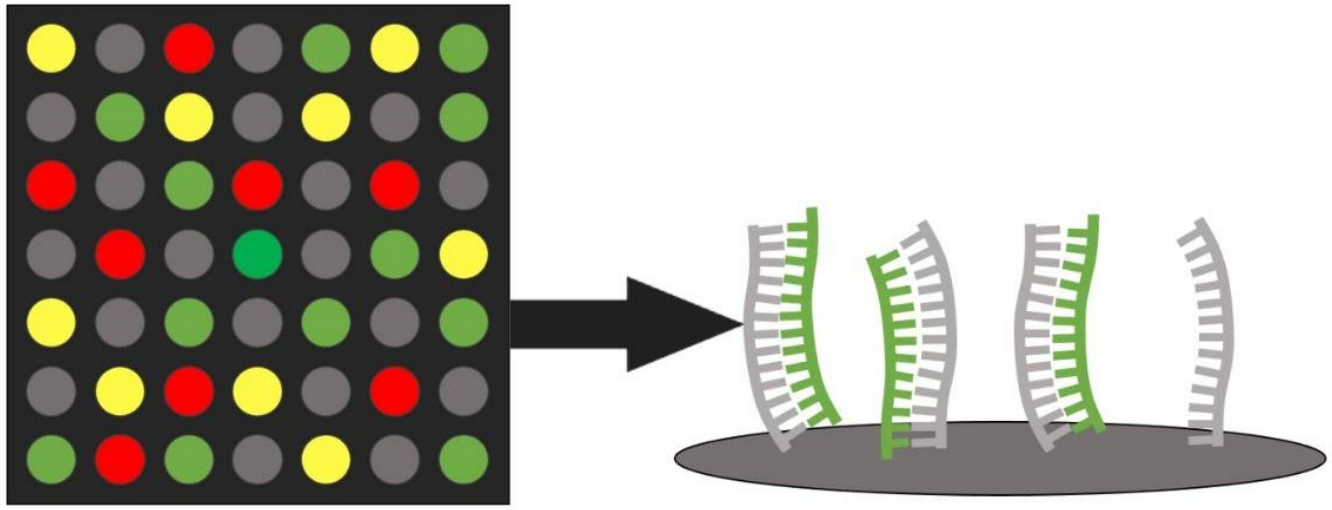
What's transcriptomics ?

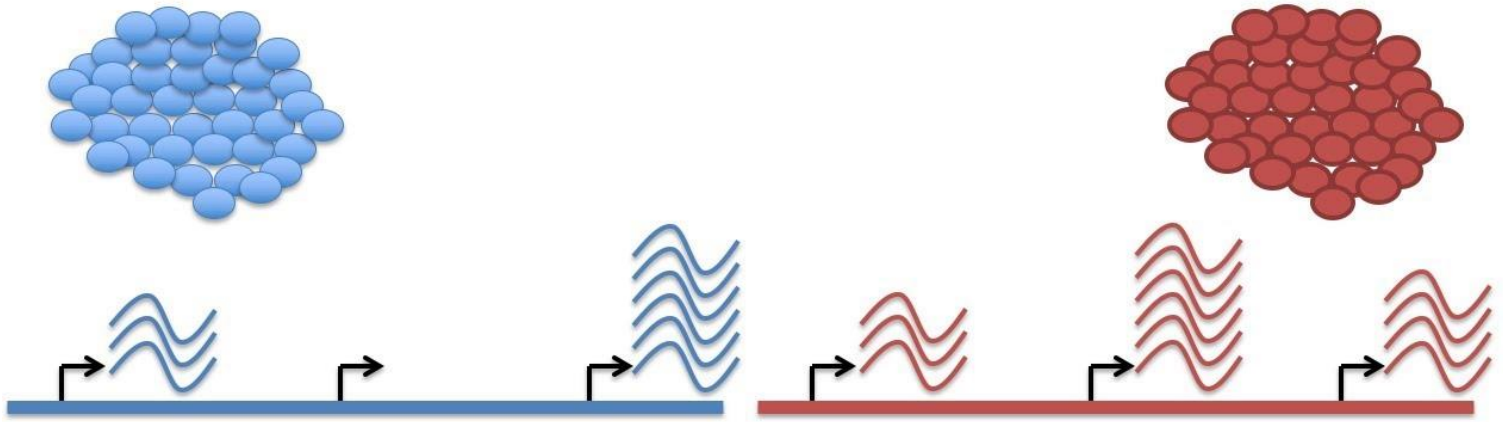
- It's the studied of the **transcriptome**.
- Generally, the goal is to study **gene expression**.



What's transcriptomics ?

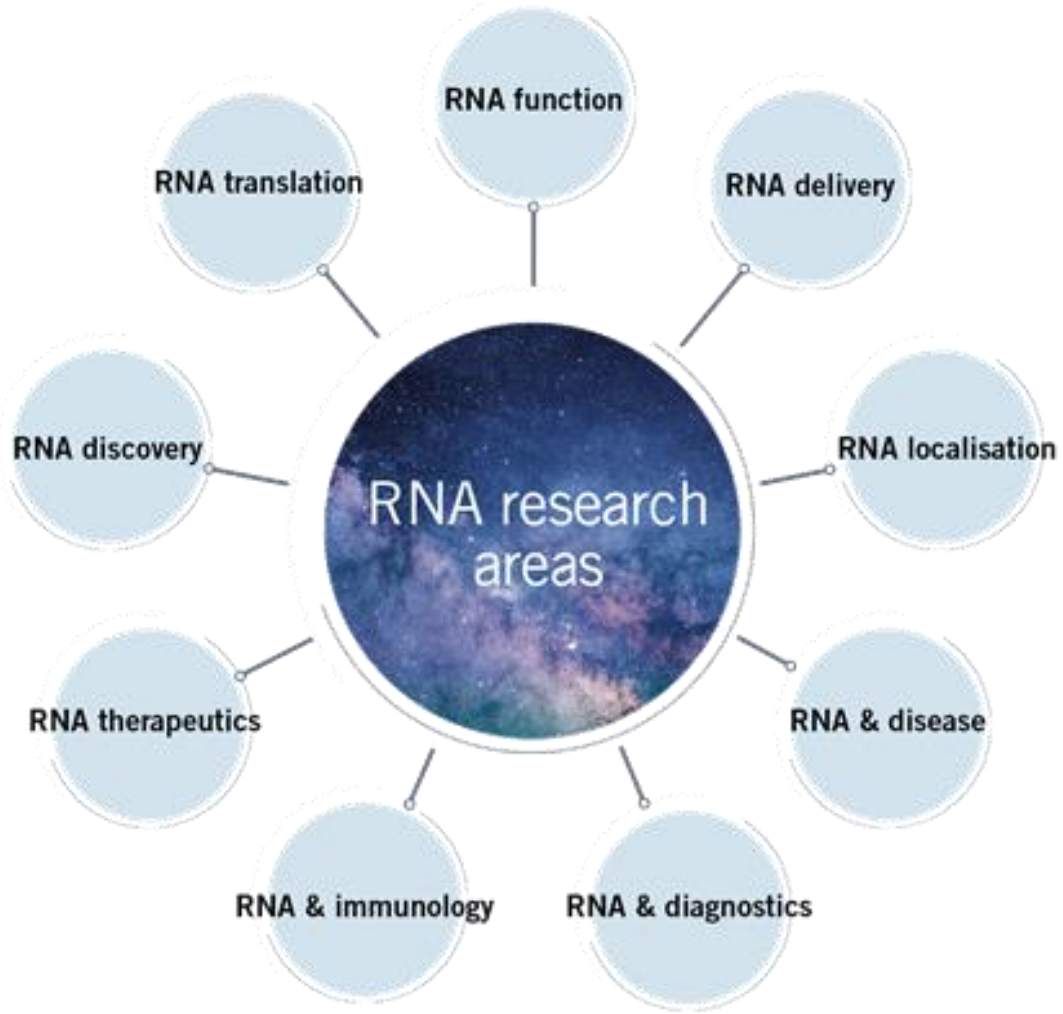
- It's the studied of the **transcriptome**.
- Generally, the goal is to study **gene expression**.
- Based on two main techniques:
  - Micorarrays
  - **RNA-seq**



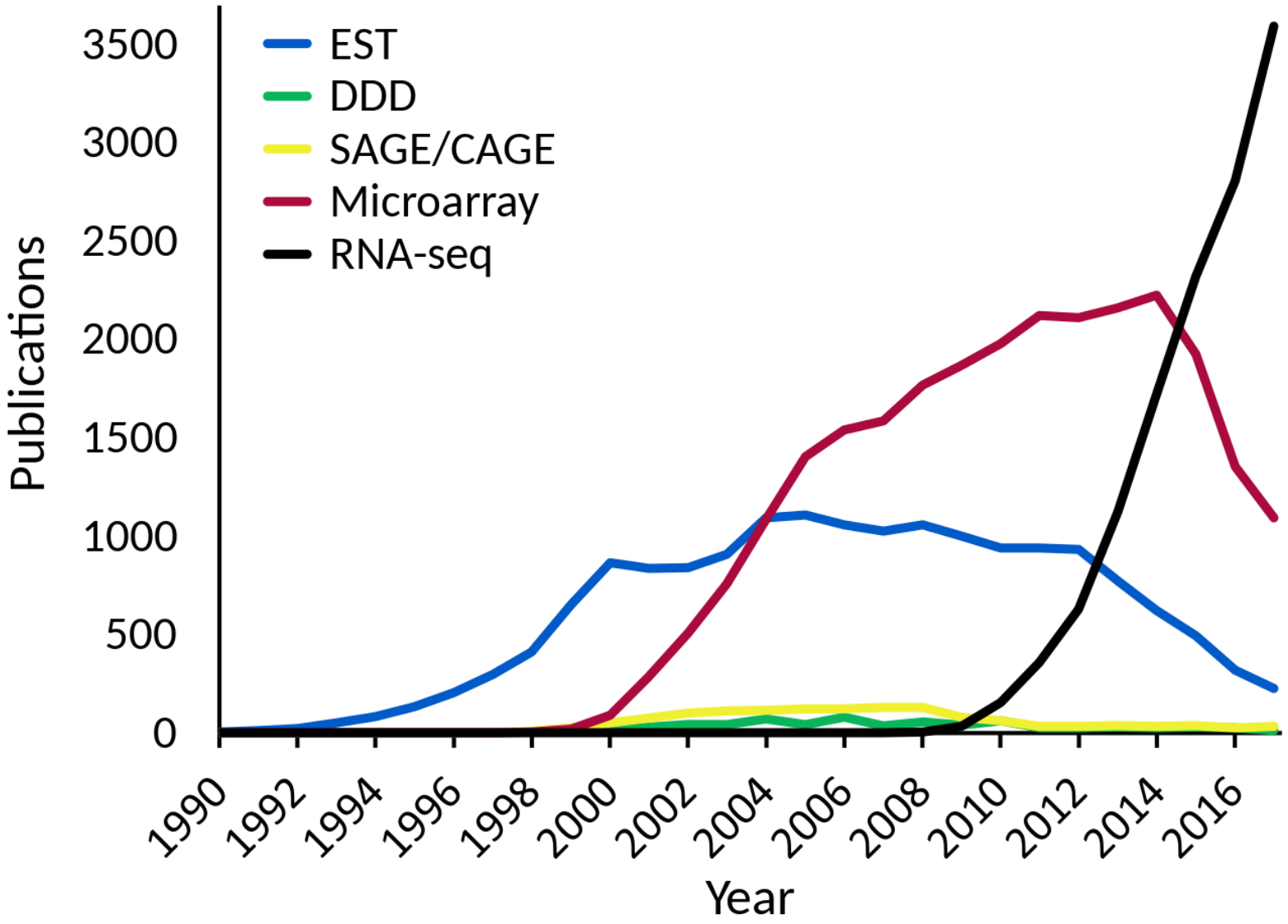


What's transcriptomics ?

- It's the studied of the **transcriptome**.
- Generally, the goal is to study **gene expression**.
- Based on two main techniques:
  - Microarrays
  - **RNA-seq**
- Goals:
  - Measure the gene expression in various conditions (development stade, healthy vs diseased tissues...).
  - Measure the difference of expression between conditions.
  - Understand functions of genes based on detected variations.
  - Understand diseases.

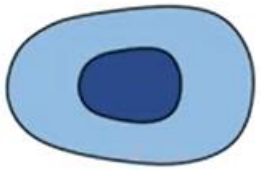


Transcriptomics methods



- First method to do transcriptomics was EST (Expressed sequence tag) invented in 1979.
- The DDD (Differential Display cDNA), SAGE/CAGE and microarrays arrived later.
- Now RNA-seq is the most used method with the apparition of the NGS (Next Generation Sequencing) technologies.

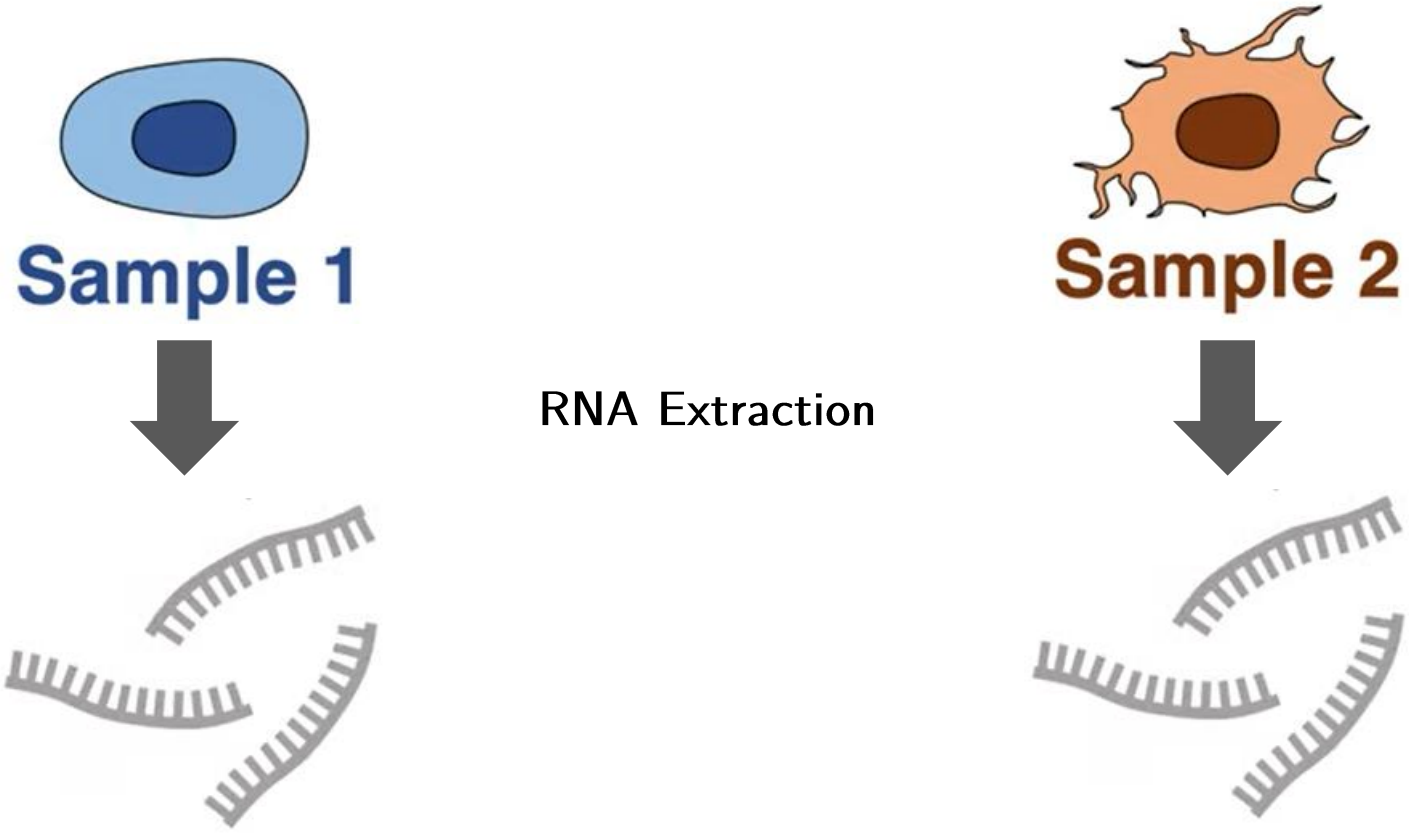
Microarray



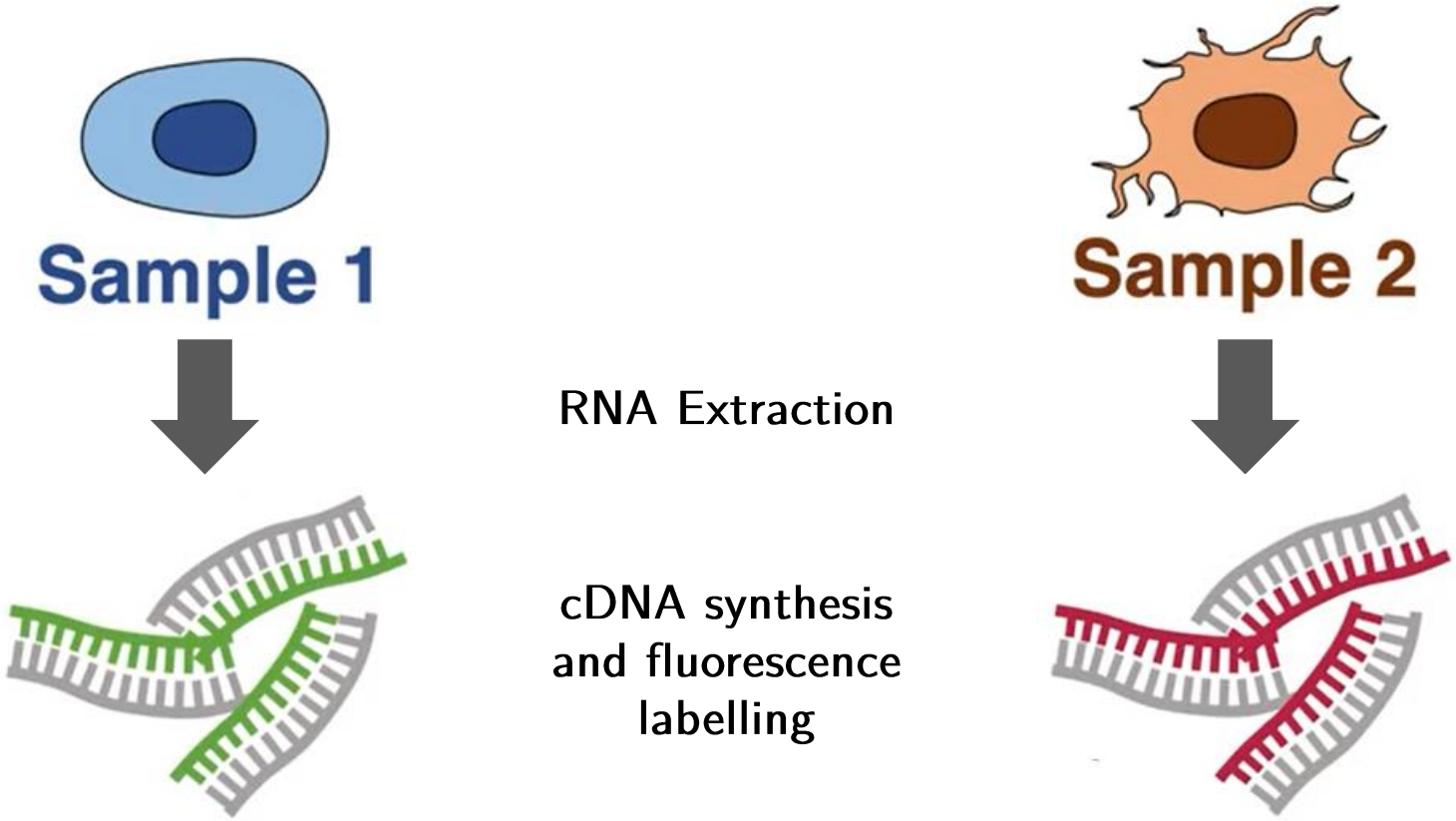
Sample 1

Sample 2

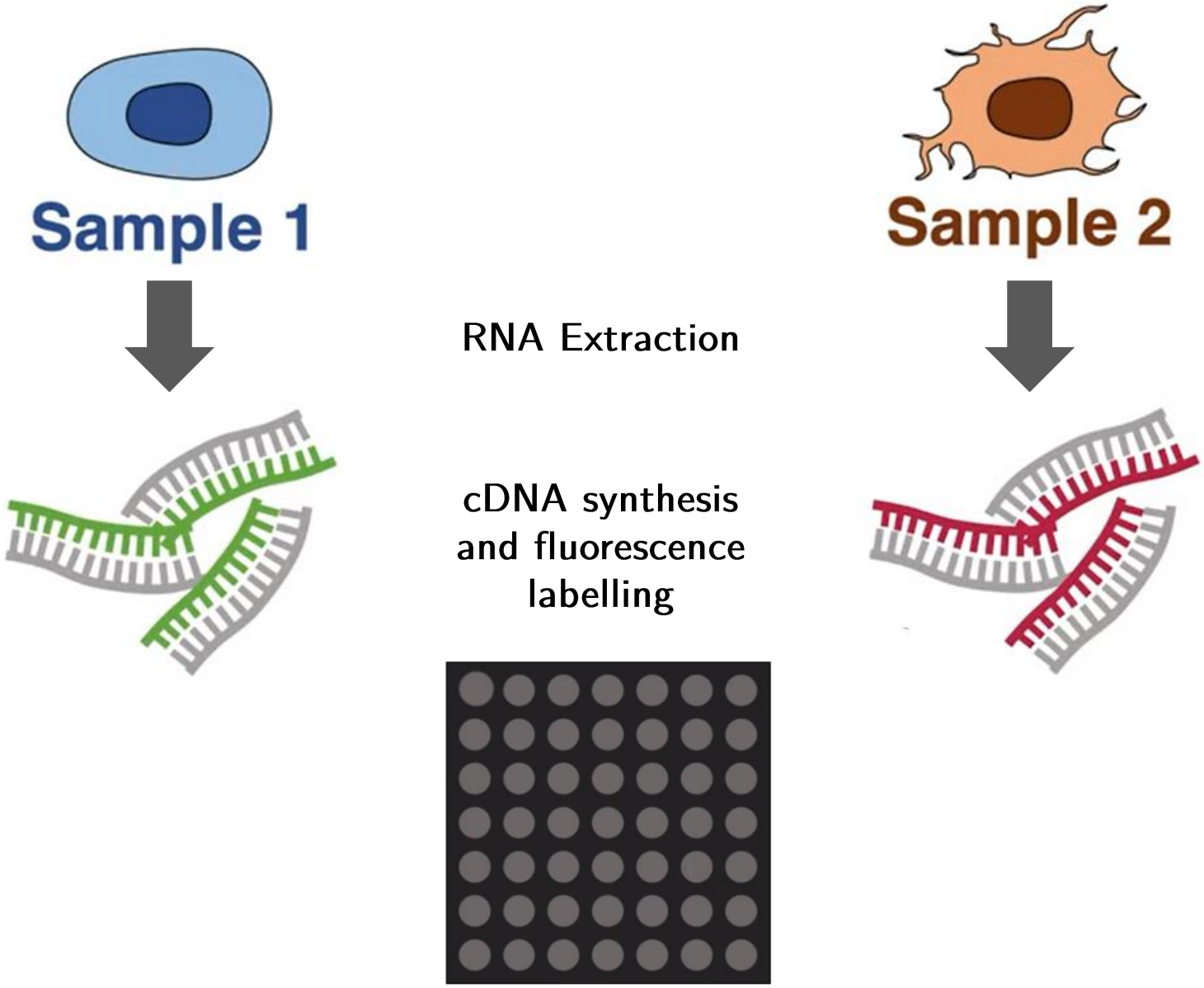
Microarray



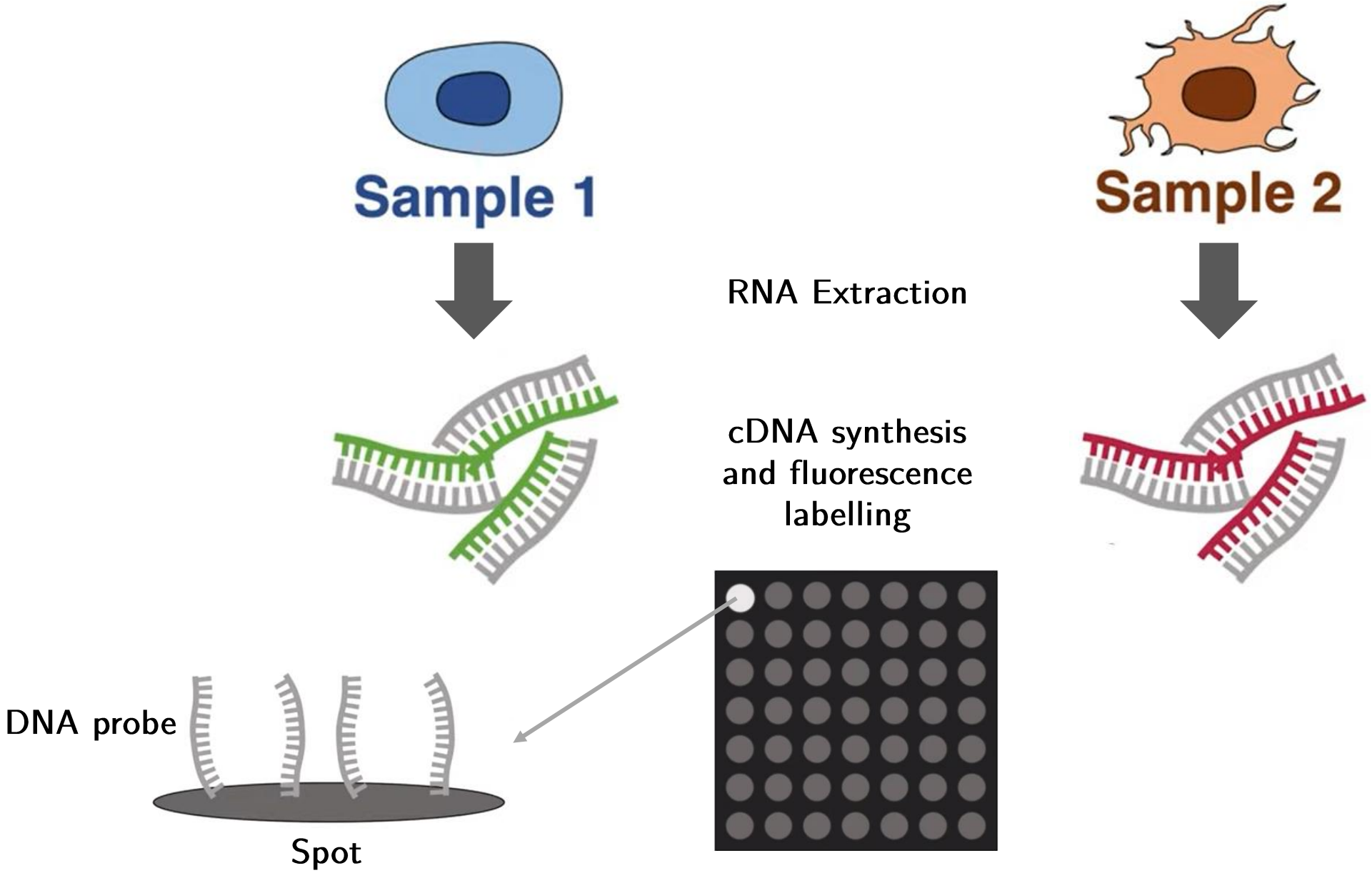
Microarray


Sample 1
Sample 2
RNA Extraction
cDNA synthesis and fluorescence labelling

Microarray


Sample 1
Sample 2
RNA Extraction
cDNA synthesis and fluorescence labelling

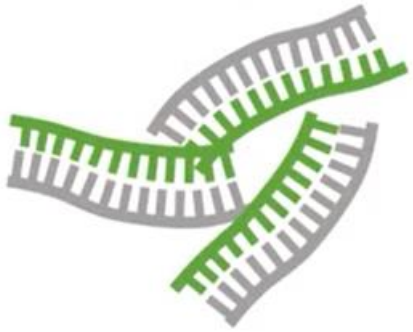
Microarray

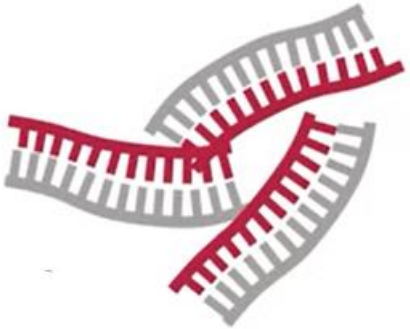

Sample 1
Sample 2
RNA Extraction
cDNA synthesis and fluorescence labelling
DNA probe
Spot

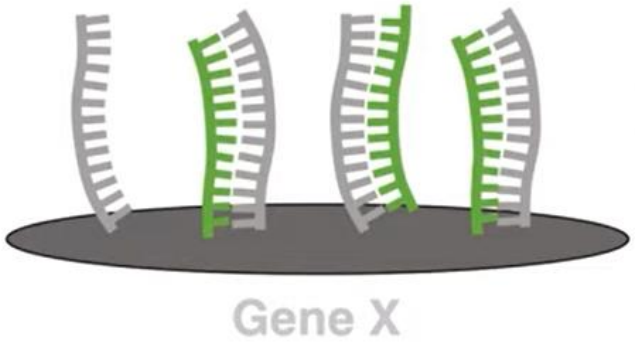
Microarray

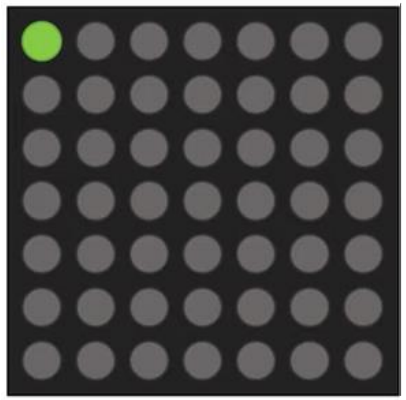
RNA Extraction

cDNA synthesis and fluorescence labelling









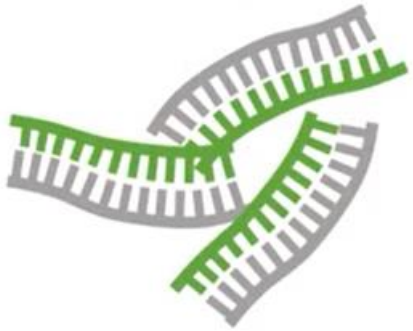
- Gene X expressed only in sample 1.

Microarray

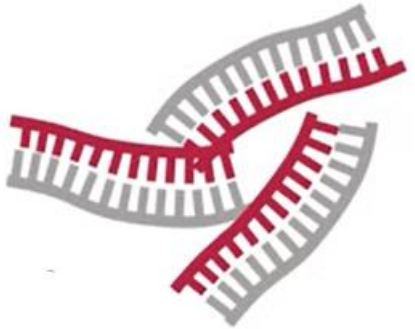
Sample 1

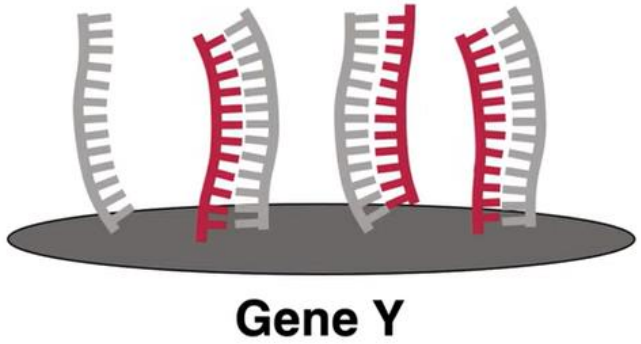
Sample 2

RNA Extraction

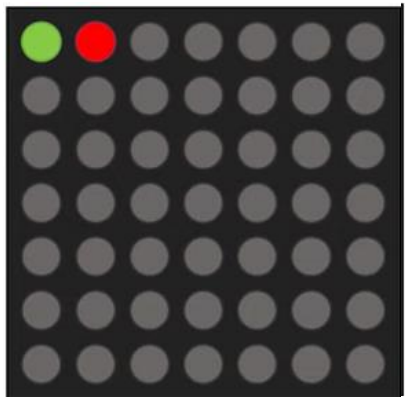


cDNA synthesis and fluorescence labelling





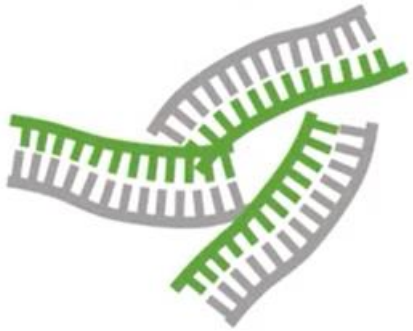
Gene Y



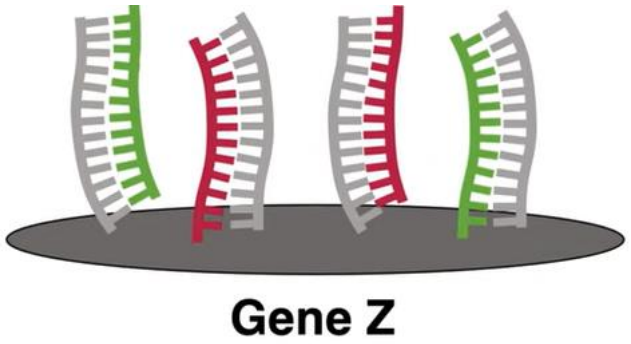
- Gene X expressed only in sample 1.
- Gene Y expressed only in sample 2.

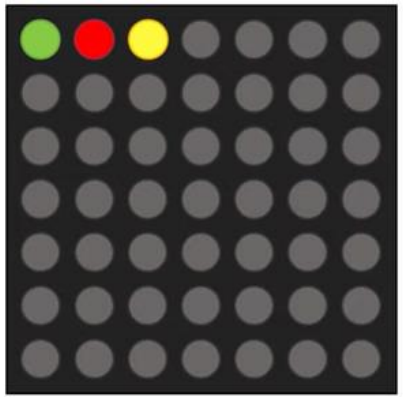
Microarray

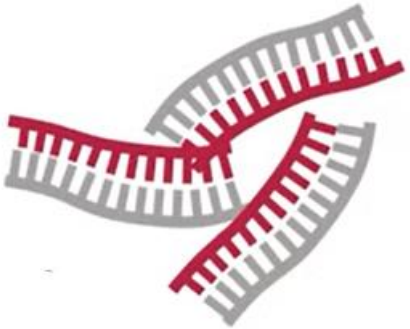
RNA Extraction



cDNA synthesis and fluorescence labelling

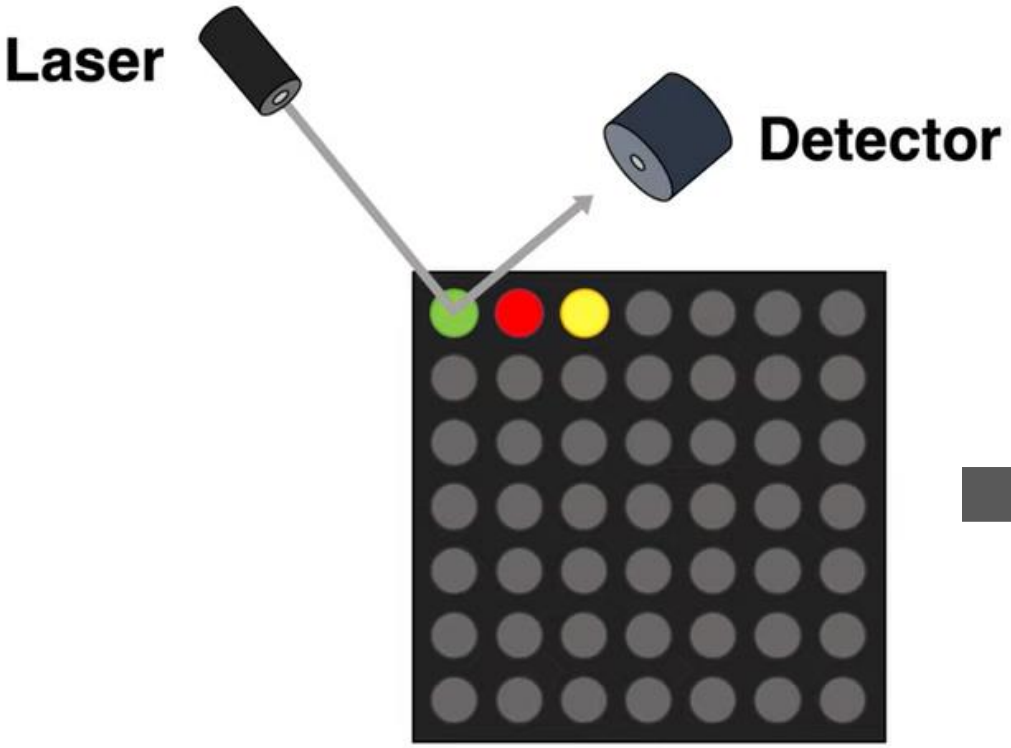




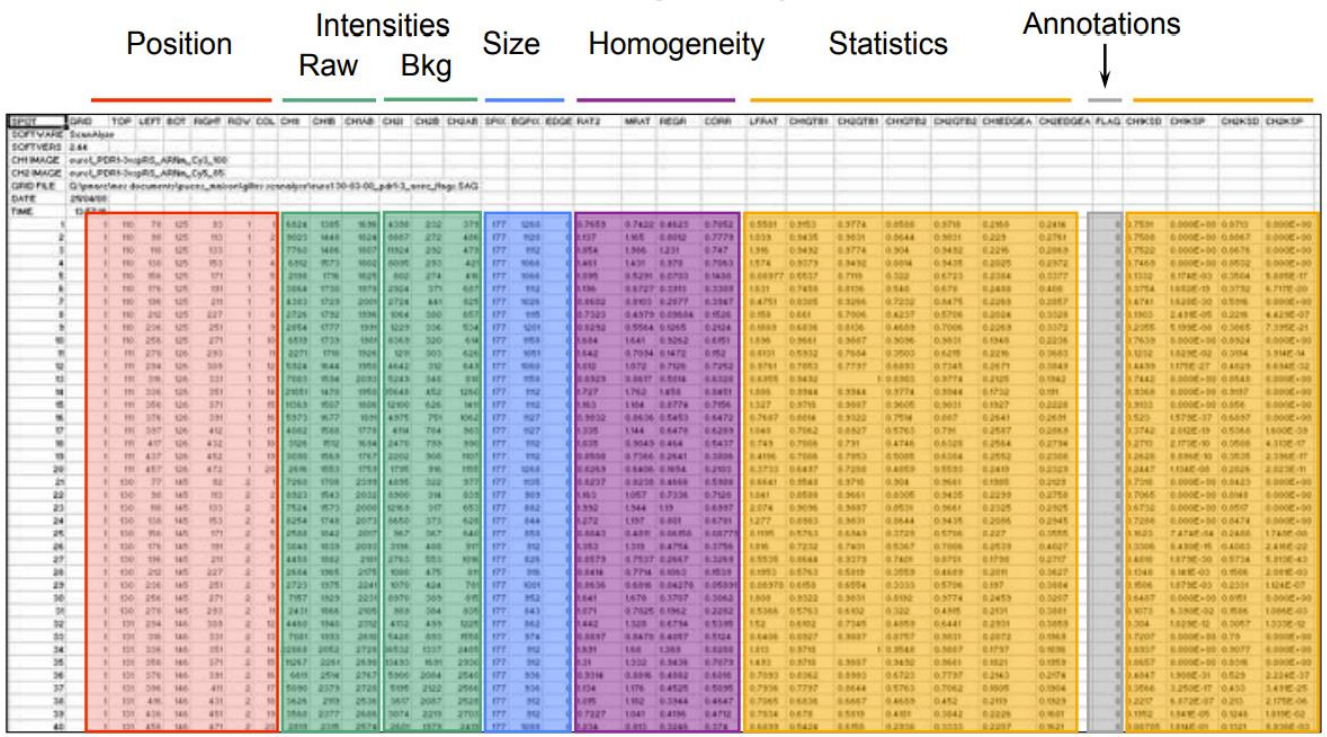


- Gene X expressed only in sample 1.
- Gene Y expressed only in sample 2.
- Gene Z expressed in both samples.

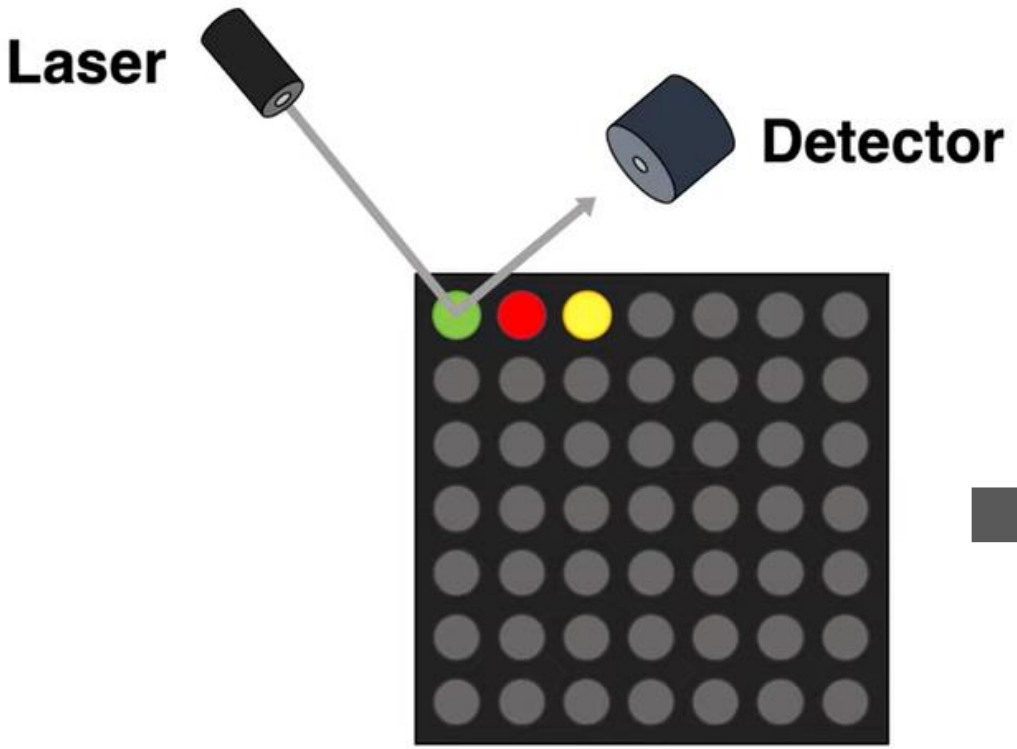
Microarray



- Example of raw output.



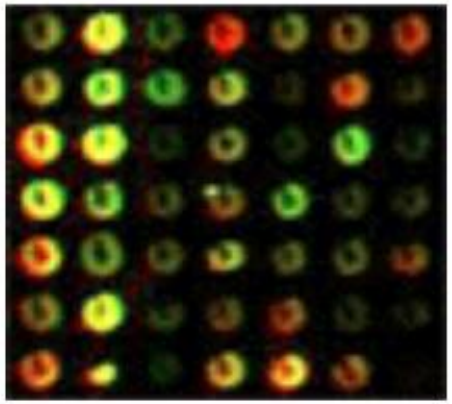
Microarray



- Example of analyzed output.

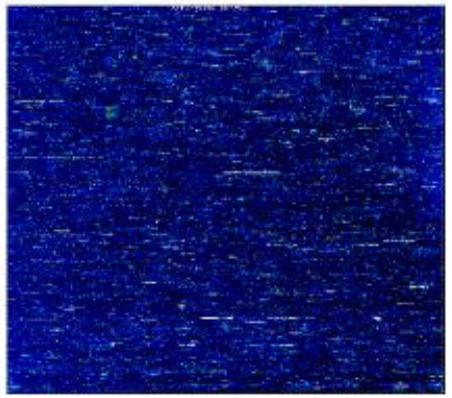
	A	B	C	D	E	F
1	**Probe Set ID**	**Gene**	**logFC**	**AveExpr**	**P.Value**	**adj.P.Val**
2	11725517_x_at	ABCG1	-1.511207274	4.920010314	3.66E-10	2.61E-09
3	11717559_a_at	ABLIM1	-2.94611096	6.235842407	1.67E-39	4.81E-37
4	11747452_a_at	ABLIM1	-2.146802277	3.36859566	9.47E-53	2.85E-49
5	11749709_a_at	ABLIM1	-2.008194865	3.361132409	9.03E-52	2.12E-48
6	11749710_s_at	ABLIM1	-1.658572984	5.485671844	5.10E-38	1.15E-35
7	11724859_s_at	ACER3	1.684093215	4.849239205	1.51E-19	2.73E-18
8	11746650_a_at	ACER3	1.981104906	6.1517831	6.91E-21	1.45E-19
9	11750700_a_at	ACSL1	1.830780762	9.2329024	1.42E-22	3.55E-21
10	11758337_s_at	ADA	-1.754985051	5.528323031	3.28E-22	7.85E-21
11	11755796_a_at	ADAM9	1.844051869	5.861768073	1.08E-19	1.98E-18
12	11755797_x_at	ADAM9	1.864739912	5.8389587	1.55E-19	2.80E-18
13	11732135_a_at	ADCK3	-1.536367954	4.731076102	2.61E-25	8.75E-24
14	11716710_a_at	ADM	1.532511447	8.763865057	2.81E-20	5.50E-19
15	11753618_a_at	ADM	1.742038275	5.077395157	1.96E-22	4.78E-21
16	11756406_x_at	ADM	1.589010274	8.653189419	1.60E-20	3.21E-19
17	11745774_x_at	AES	-1.511524068	7.445269814	1.25E-30	8.27E-29
18	11739696_s_at	AFF3	-1.656908984	2.904329655	2.64E-27	1.12E-25
19	11731943_x_at	AGAP4	-1.698765651	5.026083247	7.57E-38	1.65E-35
20	11723606_a_at	AGFG1	2.098148491	7.372889389	1.38E-35	1.94E-33
21	11723608_a_at	AGFG1	1.912553661	6.868258961	1.78E-23	4.95E-22
22	11755694_a_at	AGFG1	2.137970819	6.371678765	2.11E-37	4.28E-35

The Affymetrix GeneChip



Custom spotted arrays

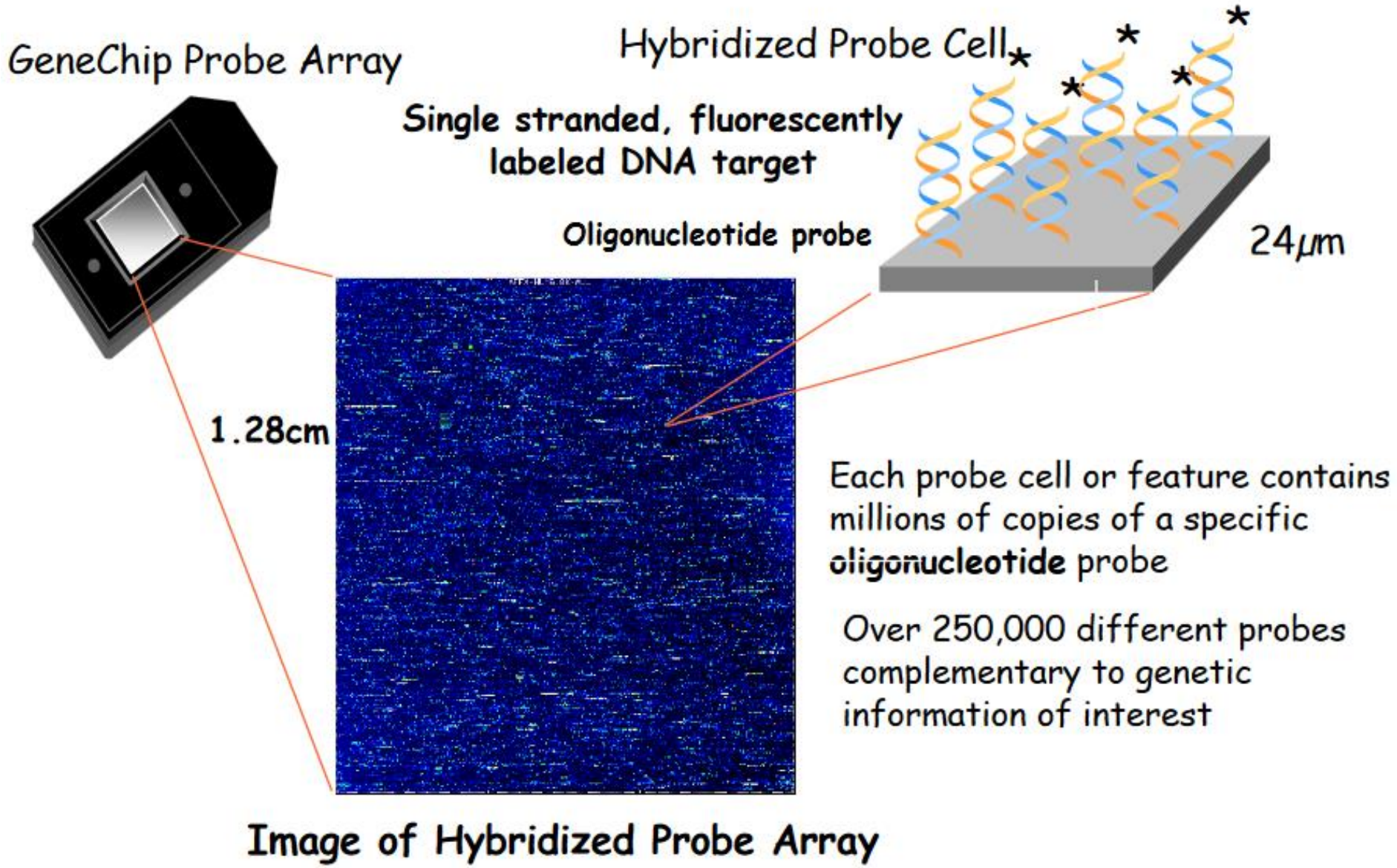
2 channels; yields a ratio between conditions.



Affymetrix GeneChip

1 channel; yields a raw count for one condition.

The Affymetrix GeneChip



Limitations

- High measurement errors

Limitations

- High measurement errors
- Issues of isoforms (same DNA molecule)

Crude model where bases are randomly distributed A, T, C, G:
Probability to have 25 same bases: 10^{-15} on a $3x10^{9}$ human genome it means 0.000003% chance to have it.

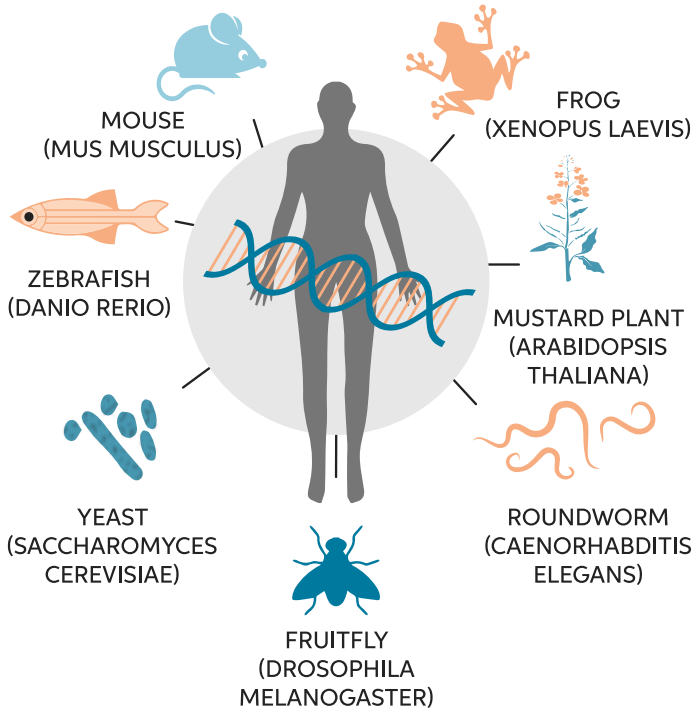
Reality a lot of sequences are more frequents, repeated sequences, gene homologues so a probe can have homologs.

Limitations

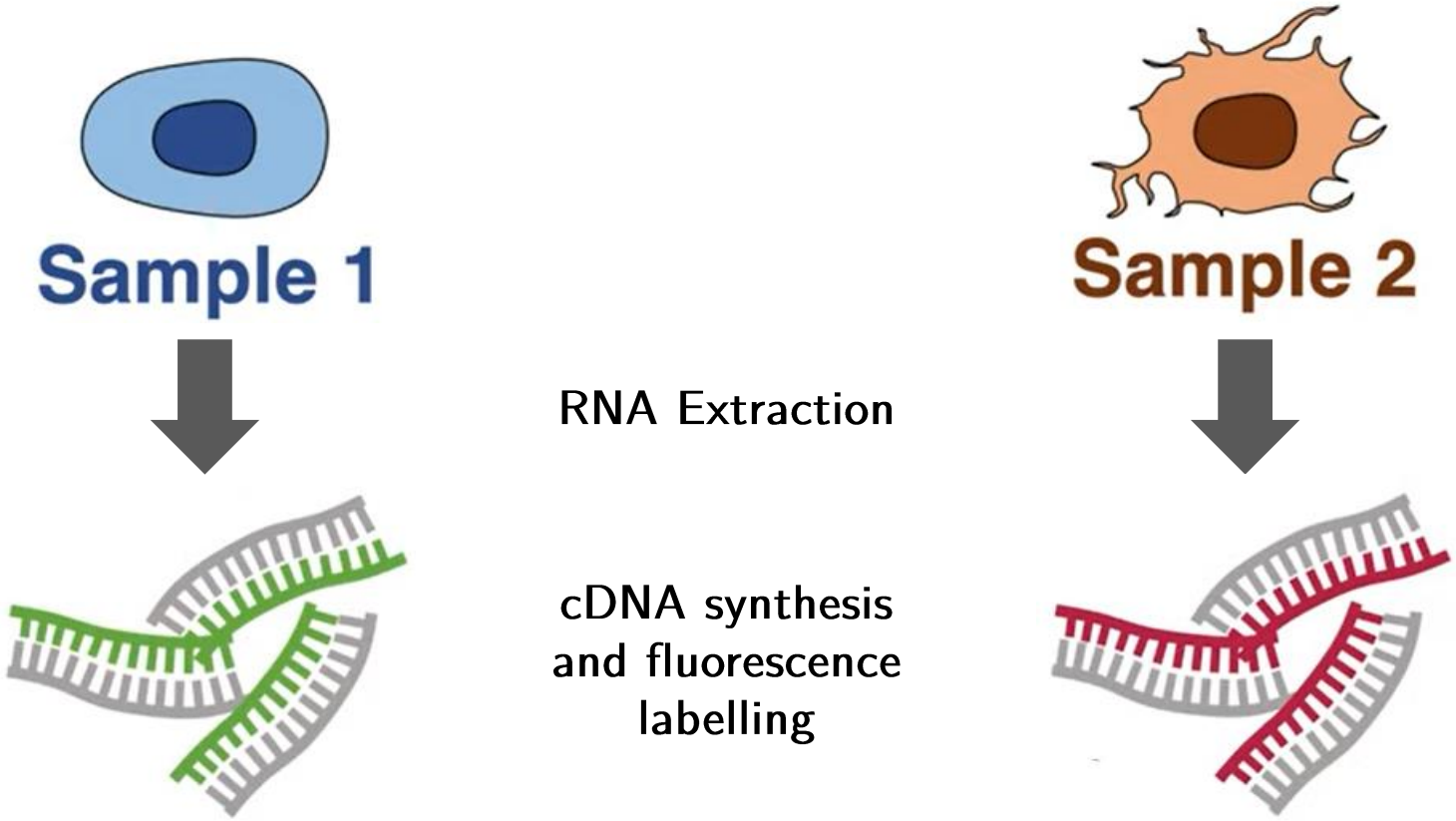
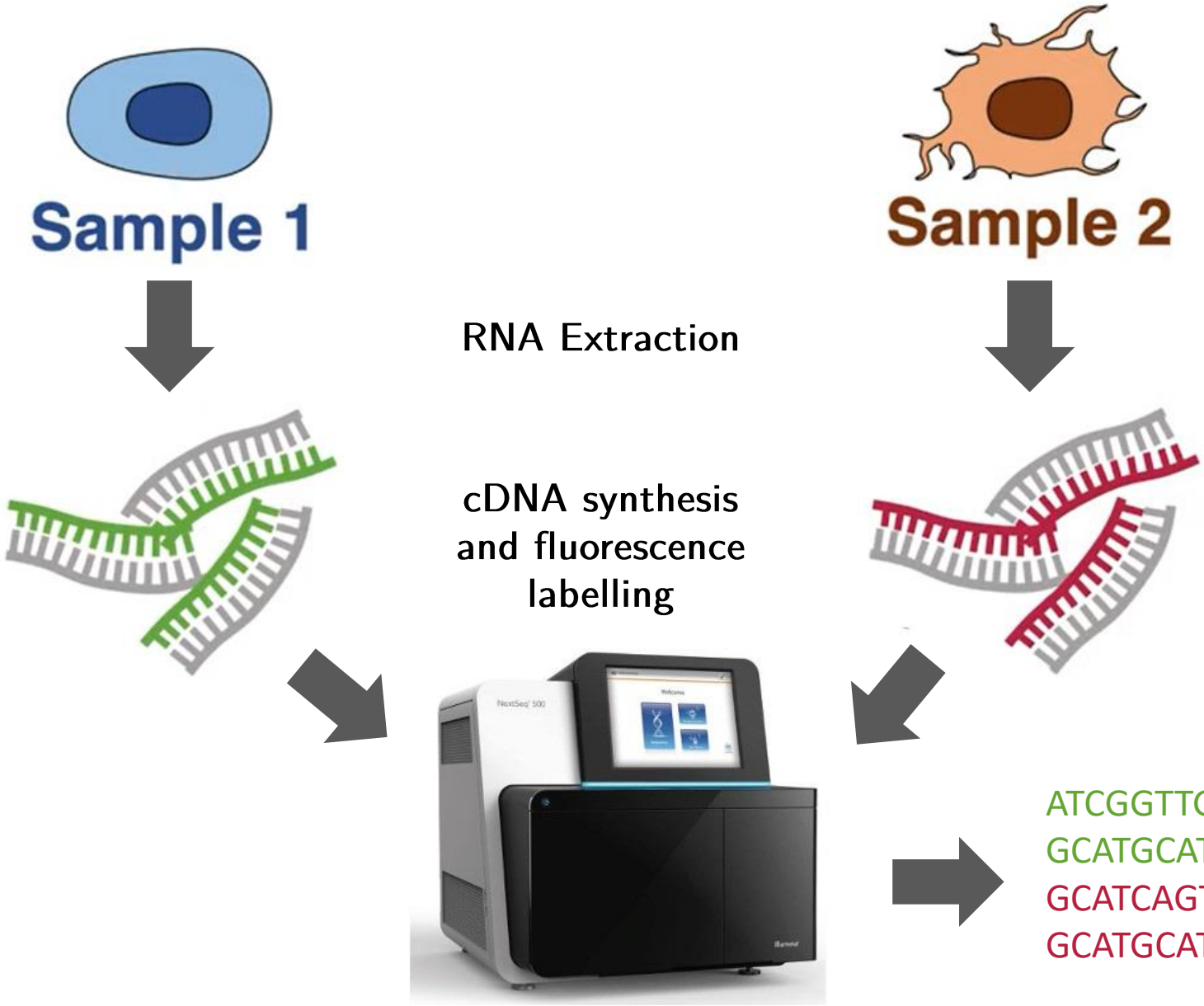
- High measurement errors
- Issues of isoforms (same DNA molecule)
- Higher cost of experiment than a RNAseq as the price of sequencing plummets.

Limitations

- High measurement errors
- Issues of isoforms (same DNA molecule)
- Higher cost of experiment than a RNAseq as the price of sequencing plummets.
- Probes limitations, works for model organisms, very expensive if you want to design personalized probes.



RNA sequencing



RNA sequencing