Next Generation Sequencing (NGS) What and Why?

Gil Stelzer and Noa Wigoda

What is Next Generation Sequencing?

Next Generation Sequencing (NGS)

- Massively Parallel Sequencing (MPS)
- Deep Sequencing
- High-Throughput Sequencing (HTS)

- Development was motivated by the Human Genome Project
- Nucleic acids – DNA and RNA(cDNA) can be sequenced much faster and cheaper than previous Sanger sequencing (chain termination sequencing)

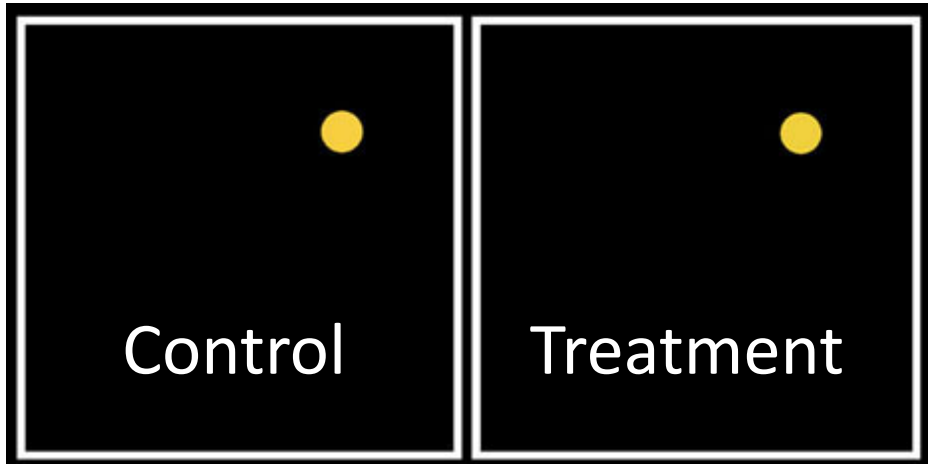
Sanger vs. Next (second) Generation vs. Third Generation Sequencing

	Number of molecules sequenced	Amplification requirement	Length of reads	Single read accuracy	Price per million bases
Sanger Sequencing	Many copies of the **same** DNA molecule	Amplification required (no library)	400-900 bp	99.9%	Very expensive
Next Generation Sequencing (Illumina dominated)	Many random amplified DNA fragments	Amplification required (library)	Machine dependent – 50-600 bp NextSeq 75-300 bp	99.9%	Very cheap
Third Generation Sequencing (PacBio and Nanopore dominated)	Many different DNA fragments	May use native nucleic acids (library)	Up to 2,000,000 bp	~87-97%	Medium

Why Next Generation Sequencing?

Paradigm shift - Olden days

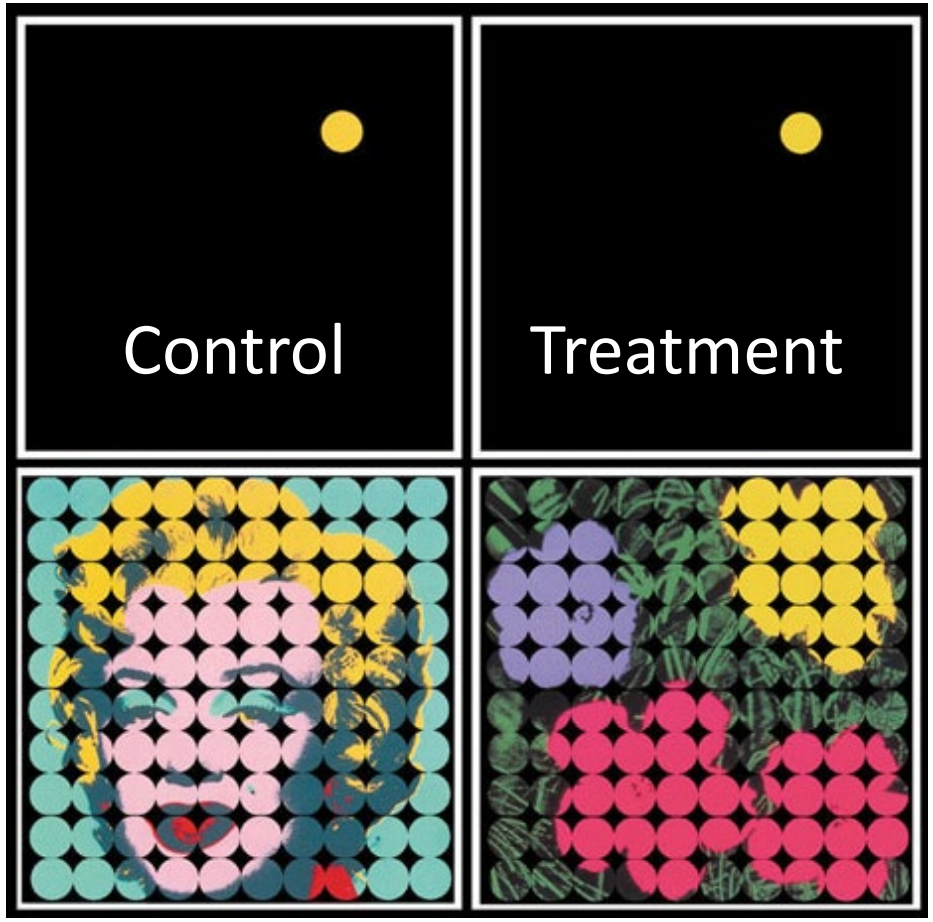
- Single-gene based science
- Classical molecular biology gives information for only a single gene
- Expression level of a single gene might not show “The big picture”



From Yanai *Genome Biology* 2002 4:301
Reproduced from Matthias W. Hentze and Petra Riedinger

Paradigm shift - Modern times

- Systems (biology) based science
- Functional genomics gives information for all the genes in the organism
- We can see “The big picture”



From Yanai *Genome Biology* 2002 4:301
Reproduced from Matthias W. Hentze and Petra Riedinger

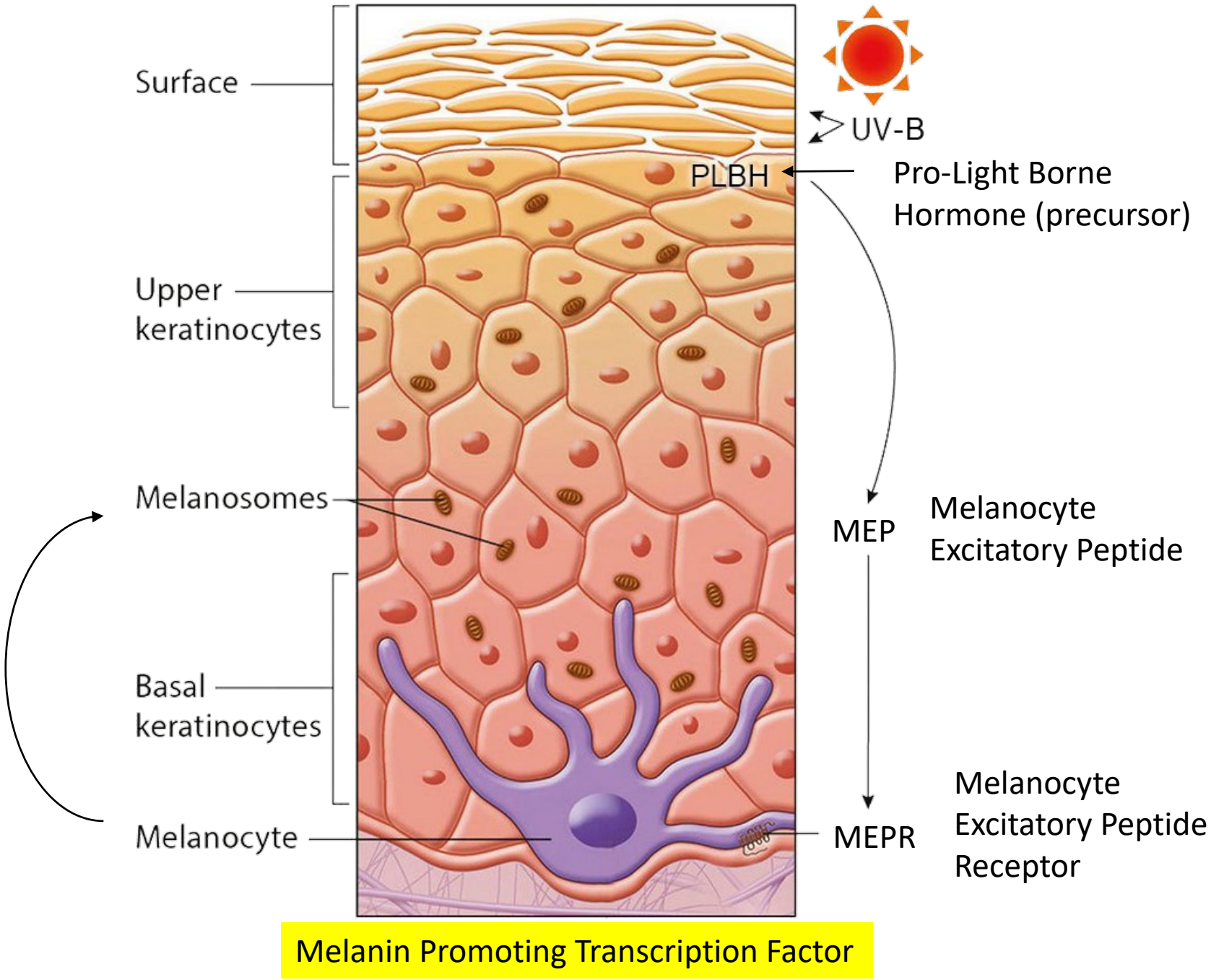
NGS example


FRECKLES

Why are freckles born?

Melanocytes constitute 5%-10% of the basal layer of epidermis, freckle-forming cells even less

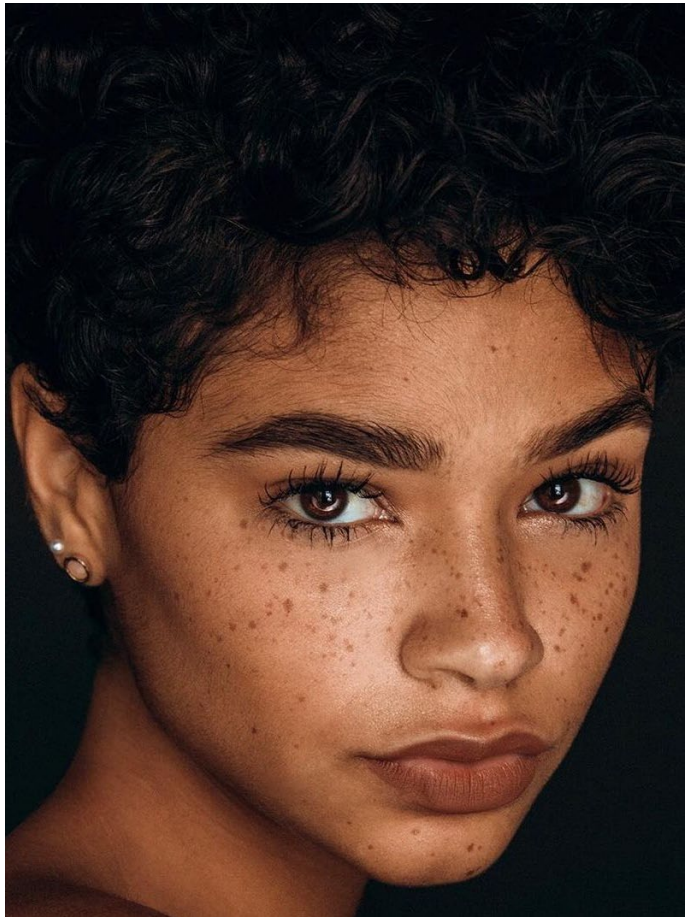
Overproduce melanin granules



Research objectives

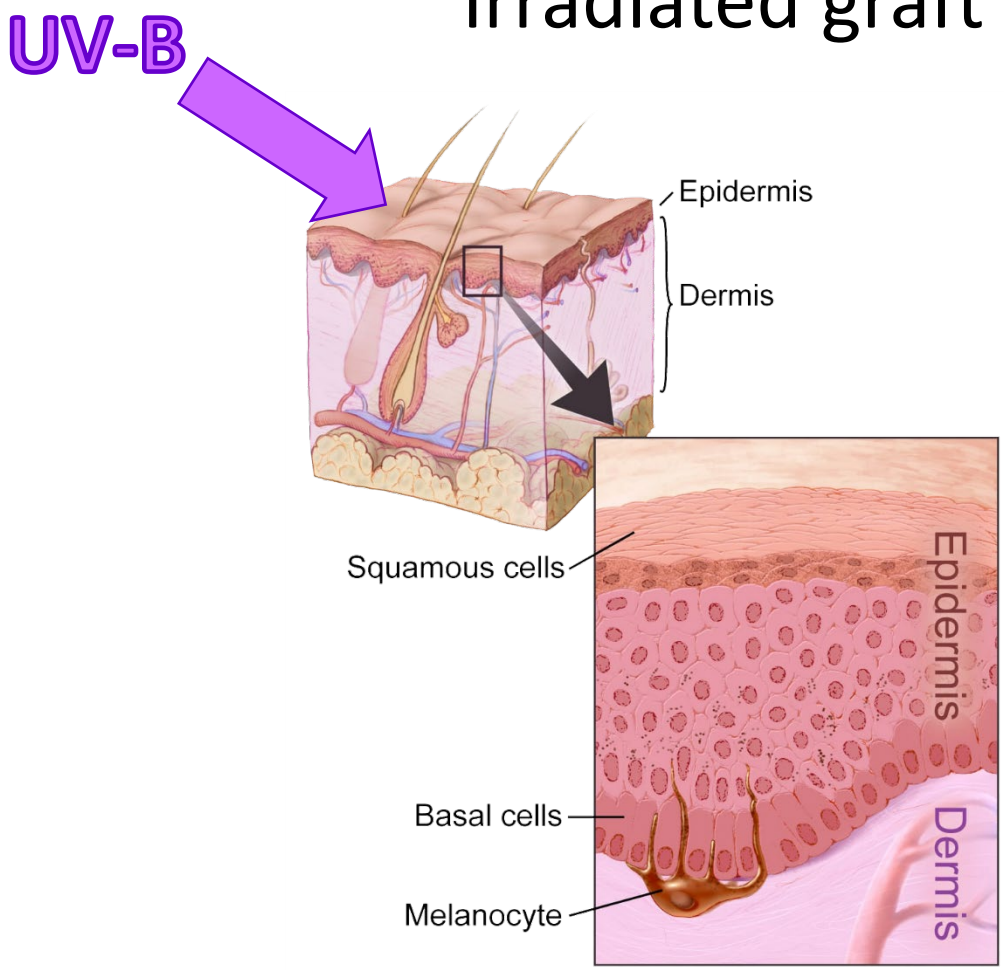
What is the molecular mechanism that explains –

- Only a subset of melanocytes form freckles
- Melanin production persists for long periods of time

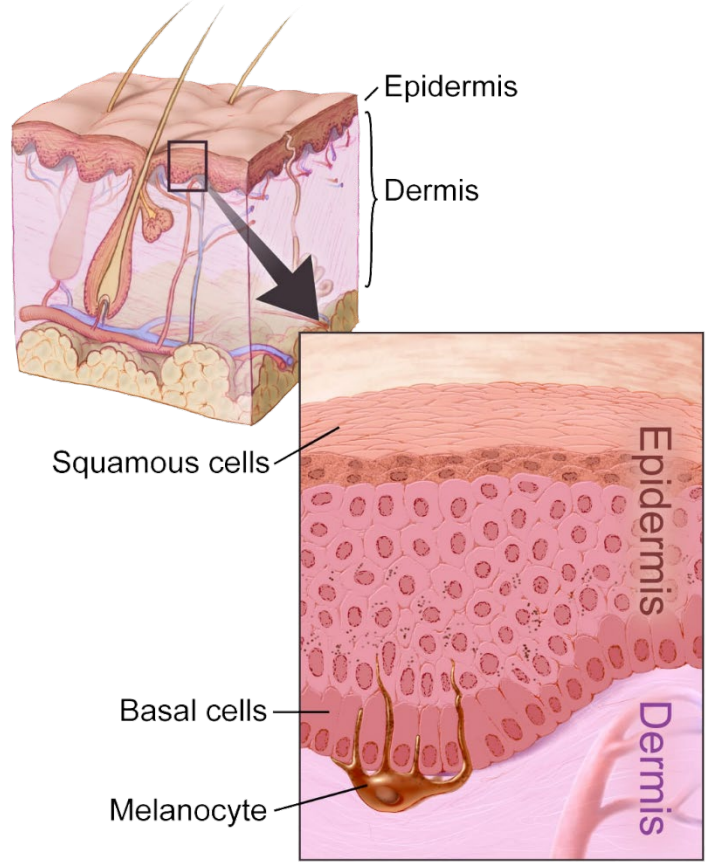


Compared 2 conditions

Irradiated graft

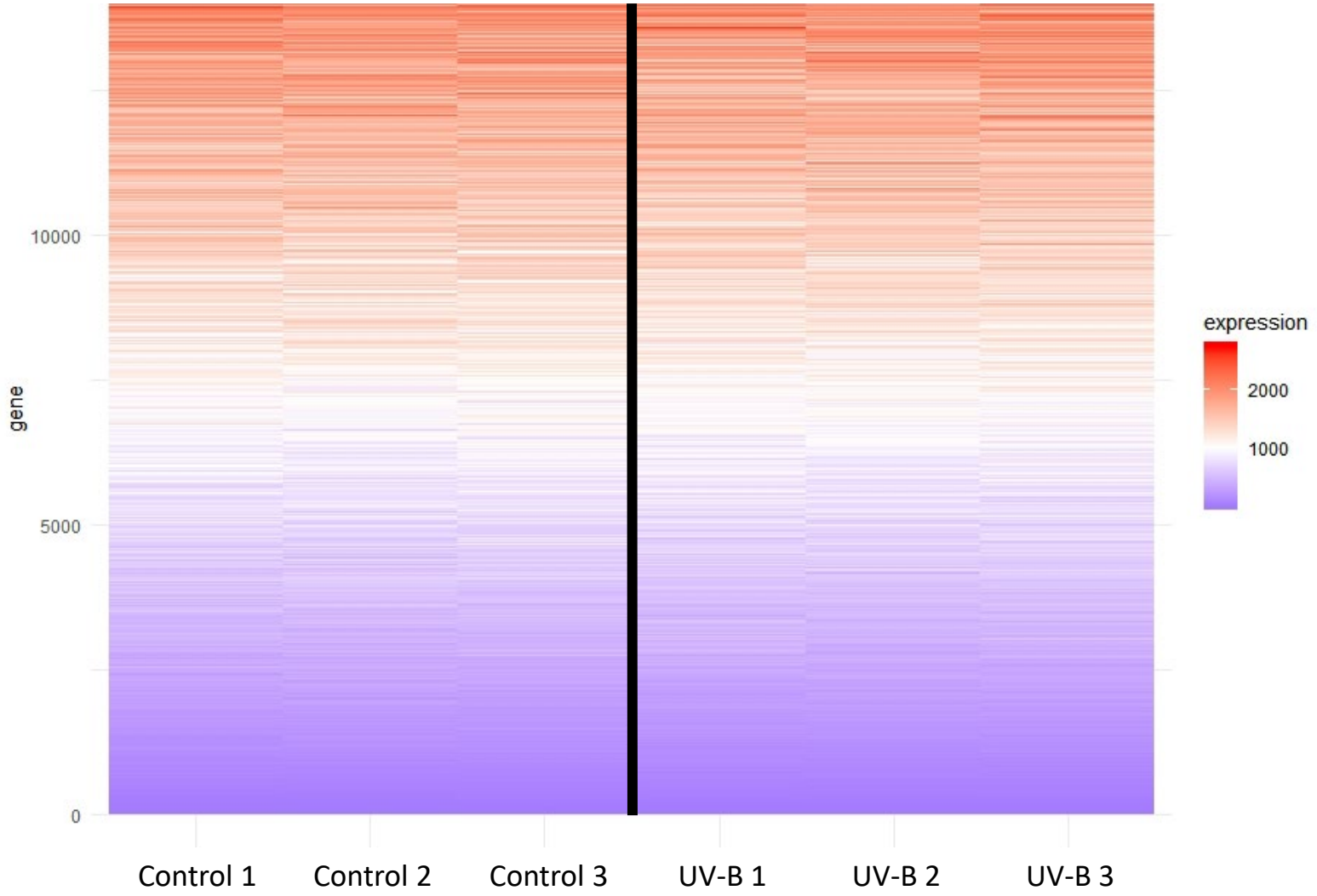


Control graft



RNA-seq of extracted mRNA

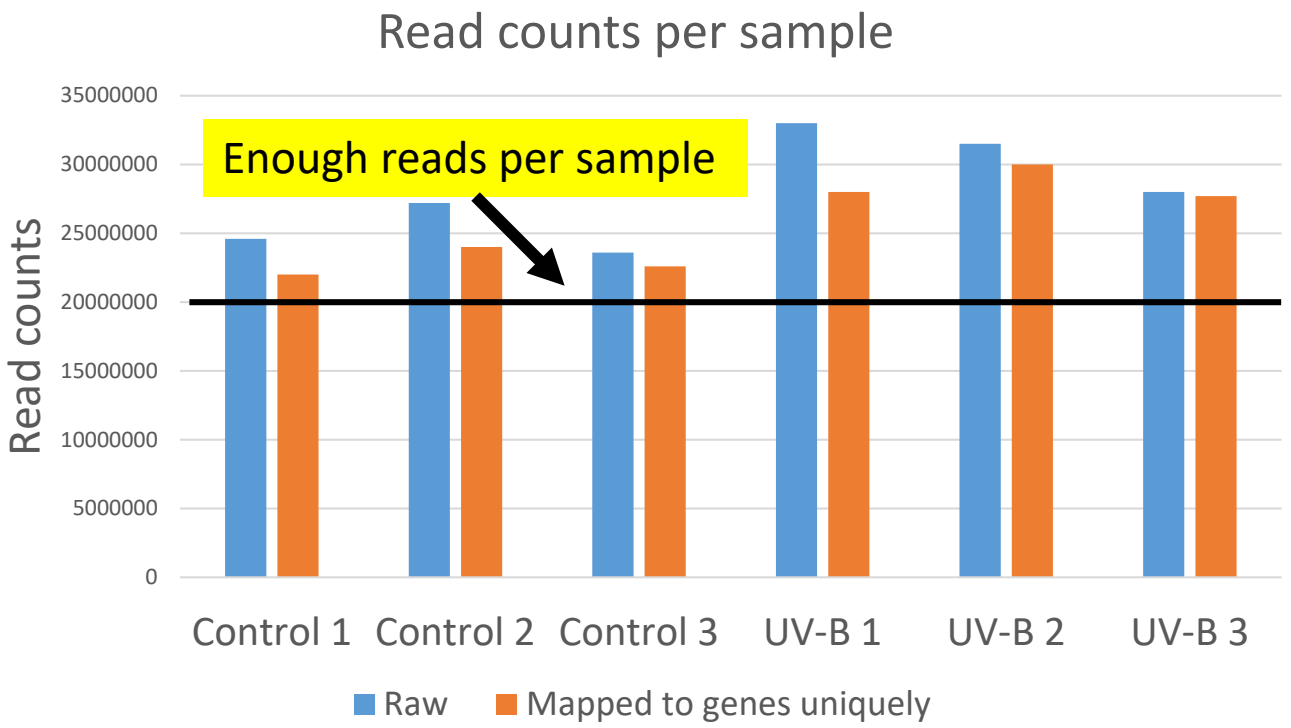
- Over 13,000 genes were observed
- No gene exhibited a significant change (comparing UV-B irradiation to control)

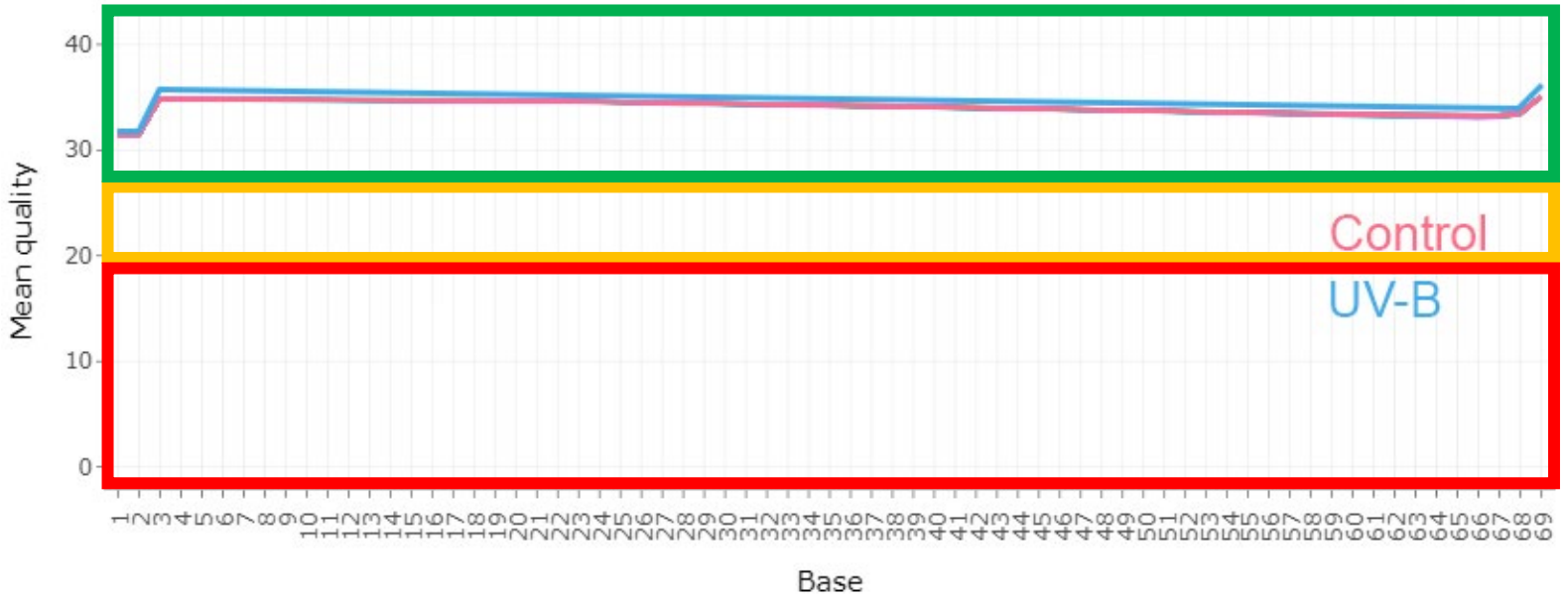


What went wrong?

What was the data quality?

- Looked at quality control (QC) report
- All samples exhibited results indicating high quality





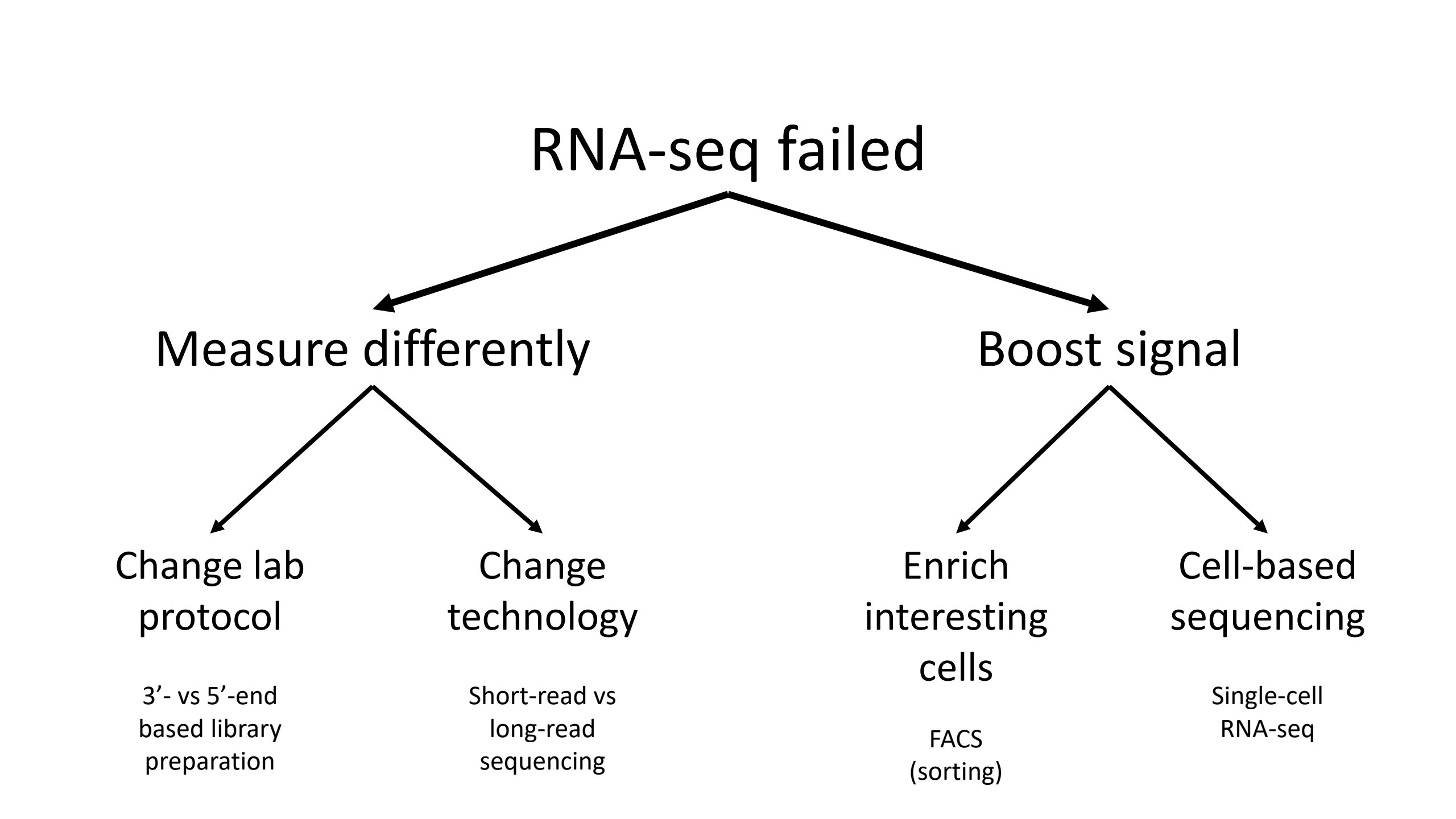
High mean quality of bases that is stable throughout the reads

Threshold selection

What threshold was used?

Threshold parameters were not too stringent

- padj <= 0.05
- |log2FoldChange| >= 1
- baseMean >= 5


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
Change technology
Enrich interesting cells
Cell-based sequencing
3’- vs 5’-end based library preparation
Short-read vs long-read sequencing
FACS (sorting)
Single-cell RNA-seq

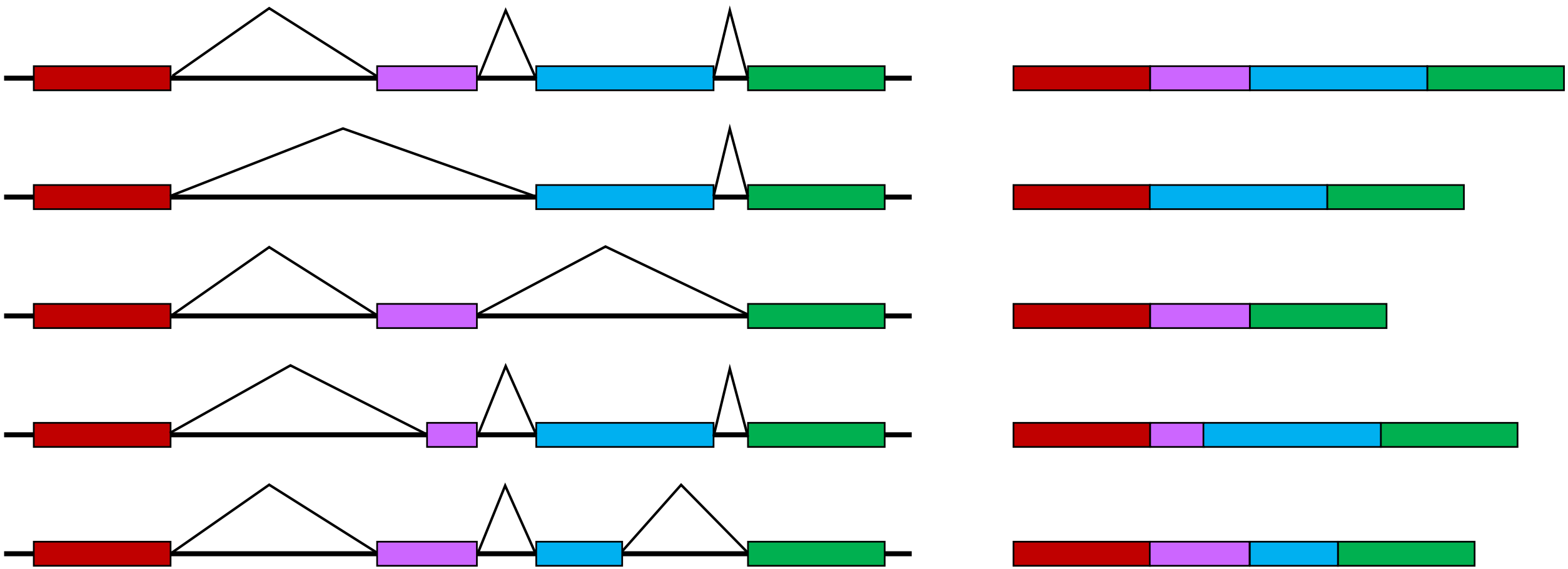
Technological review and considerations

How was the library constructed? How was the expression level calculated?

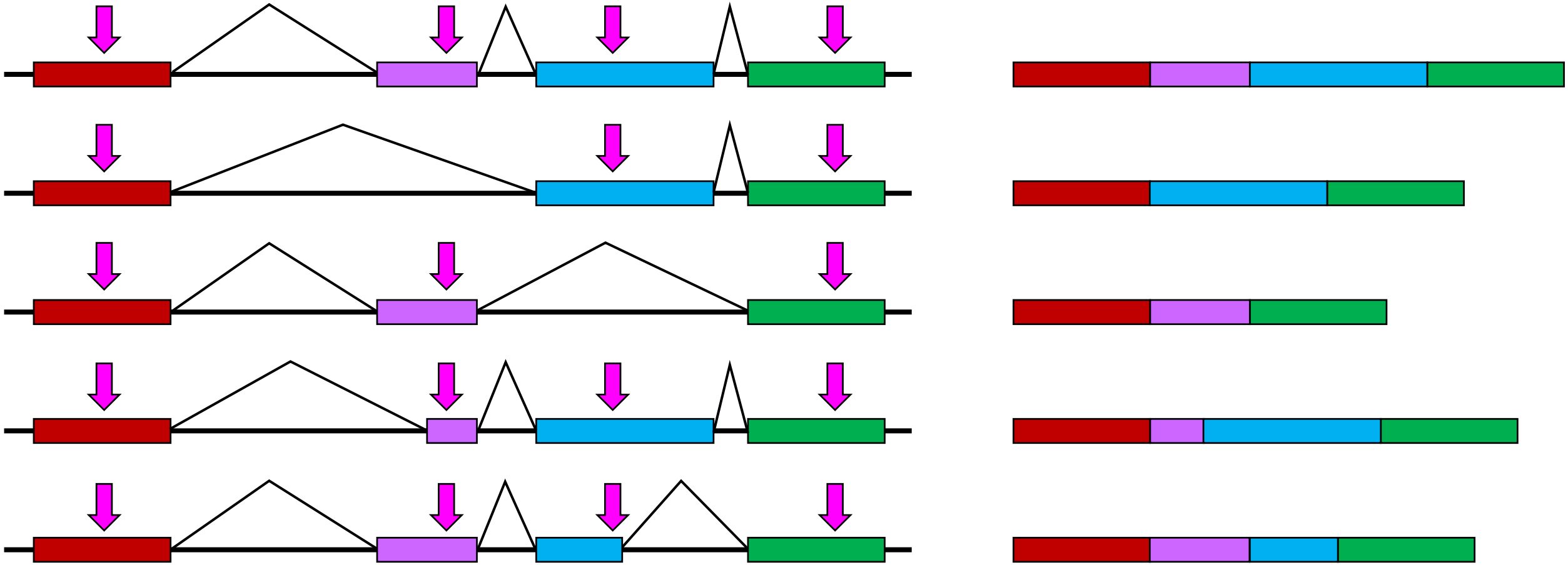
3’-end library construction protocol was used, therefore removing transcript information

This is suitable since gene-based expression levels were of interest

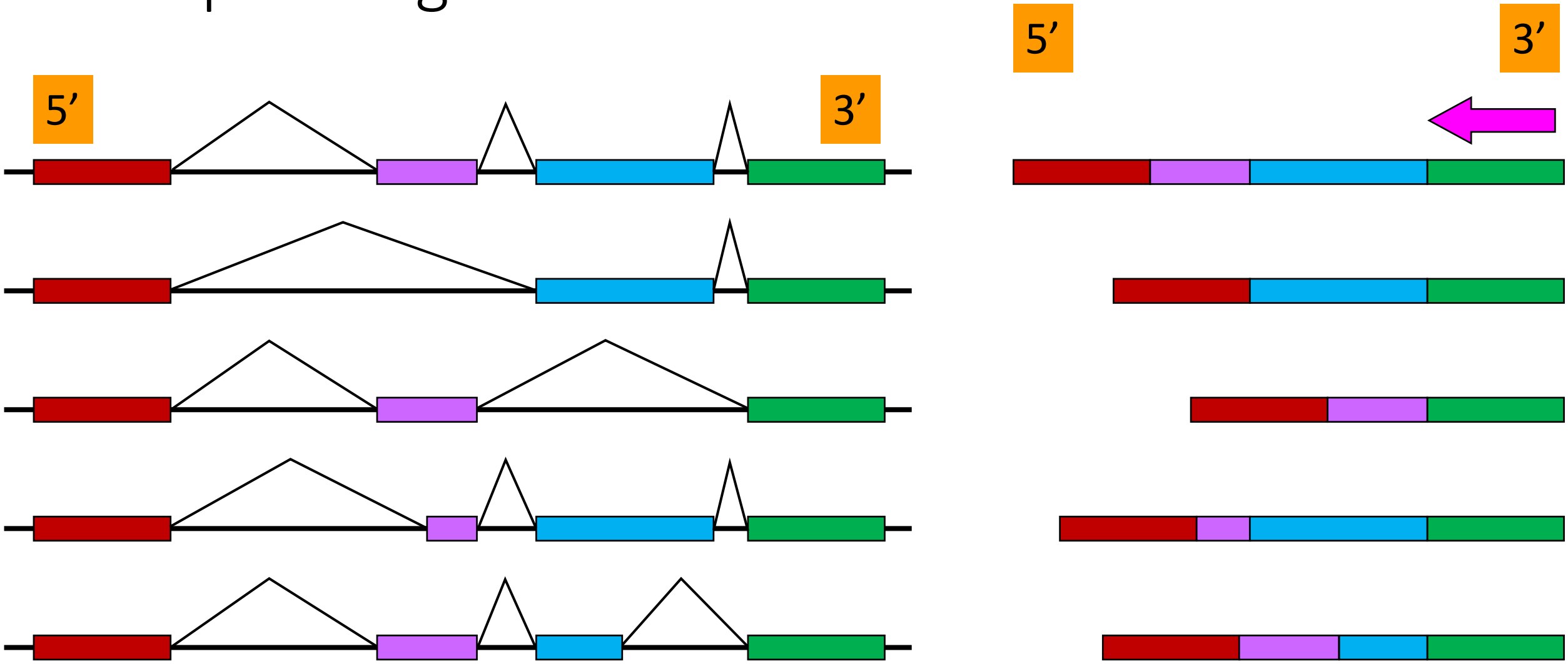
Alternative transcription



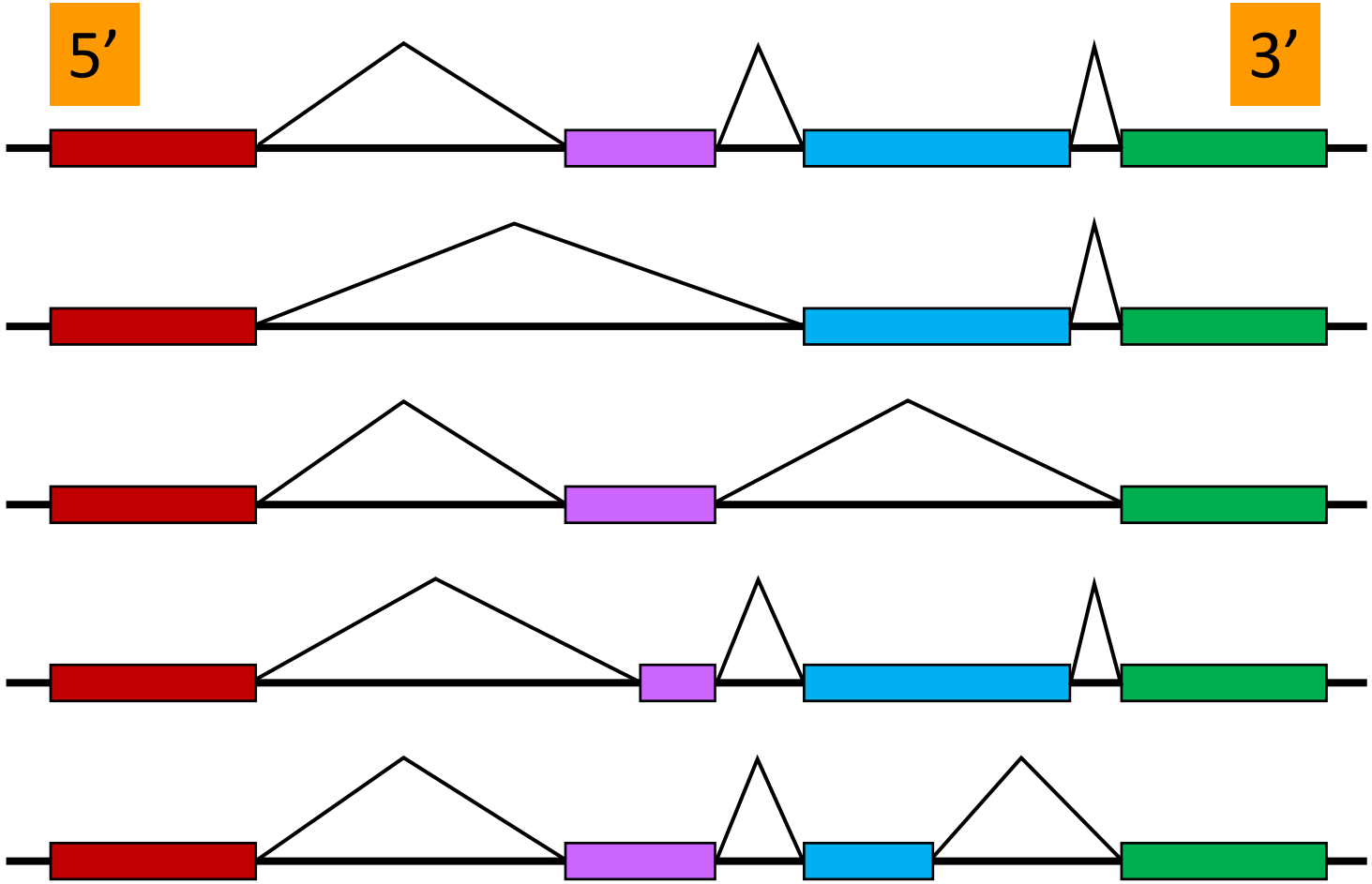
Alternative transcription

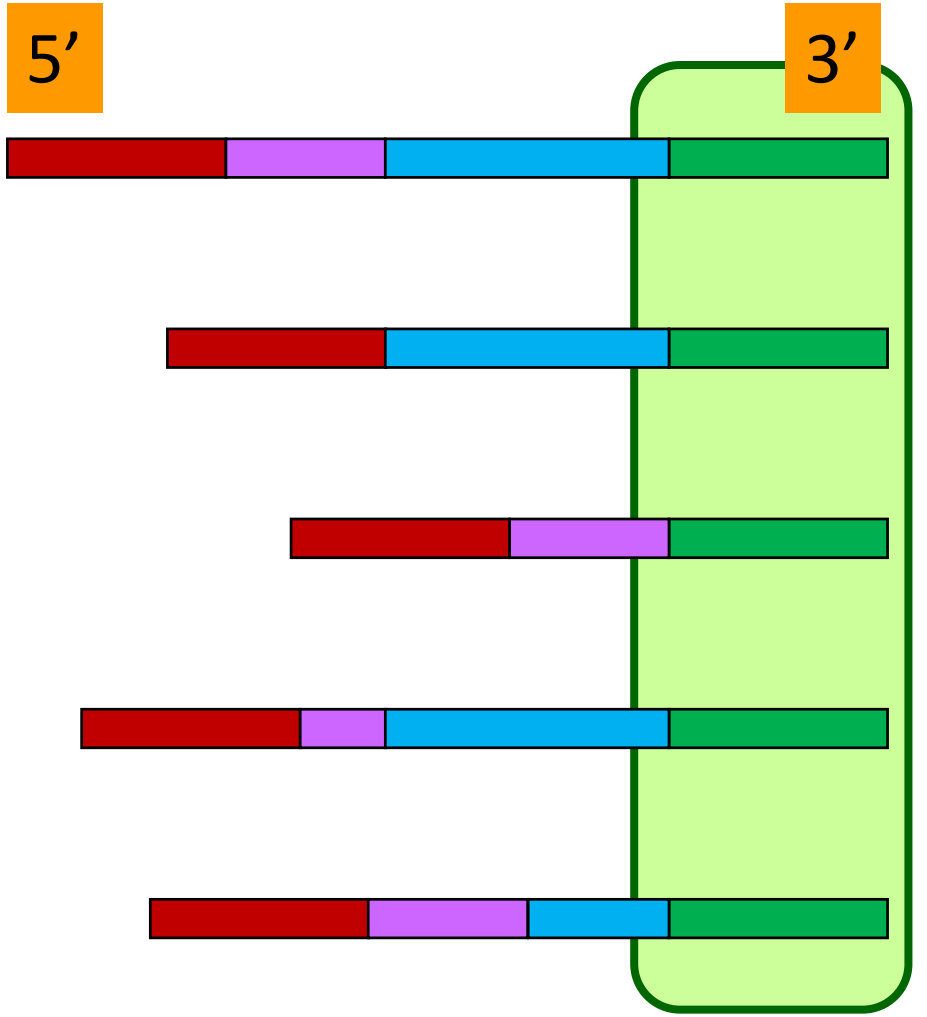


Sequencing 3’ end



Sequencing 3’ end





Technological review and considerations

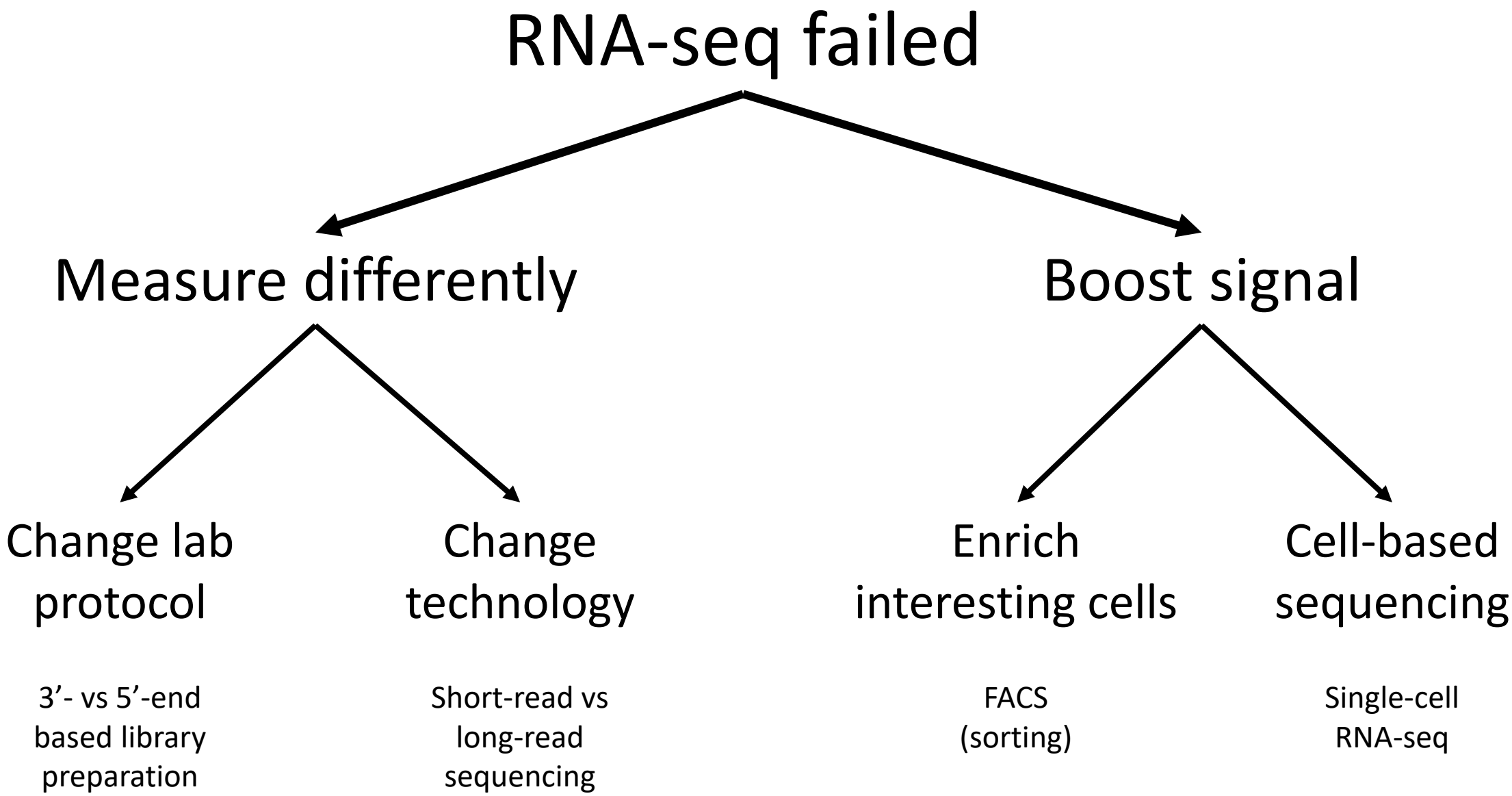
How was the library constructed? How was the expression level calculated?

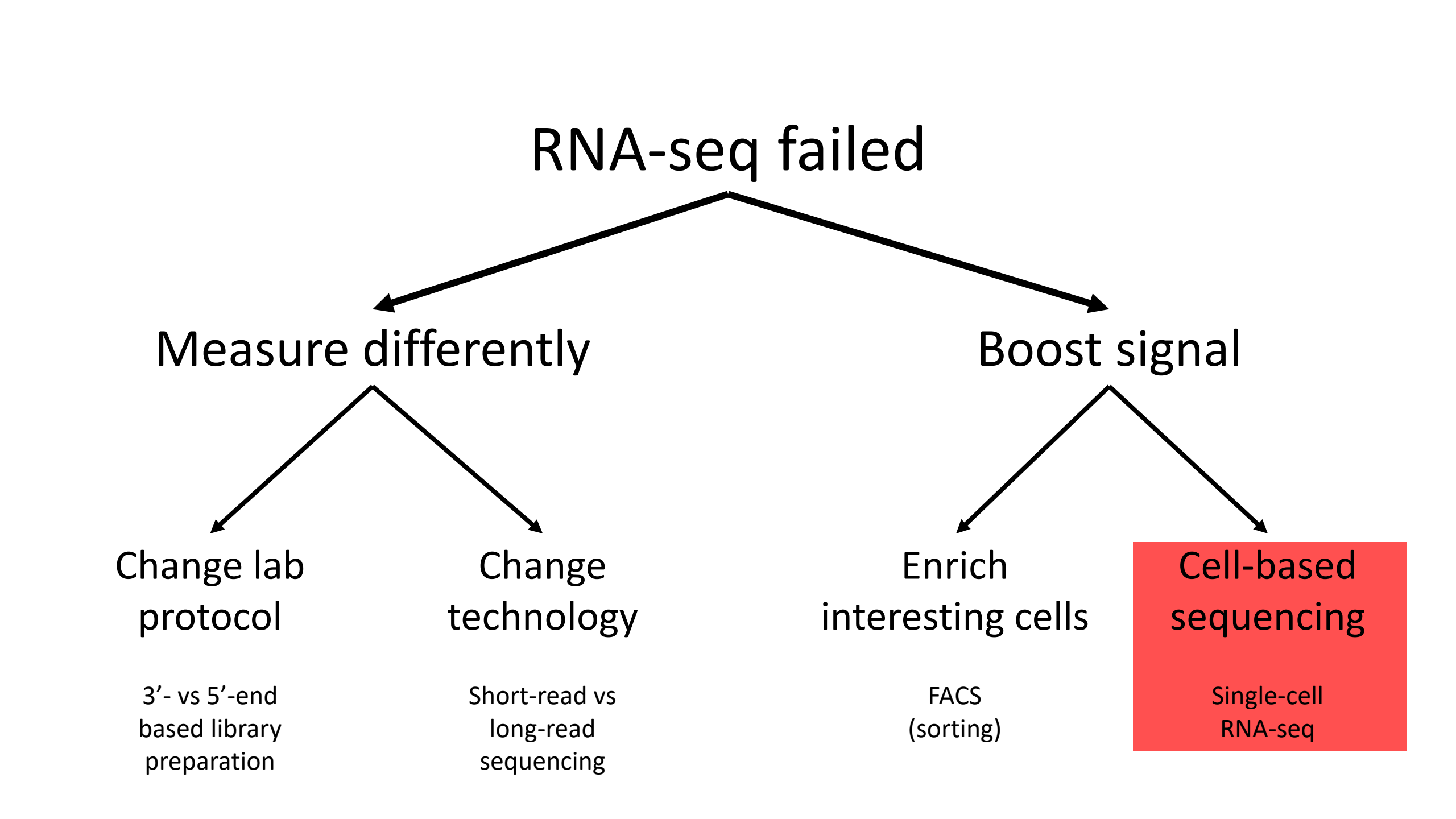
3’-end library construction protocol was used, therefore removing transcript information

This is suitable since gene-based expression levels were of interest

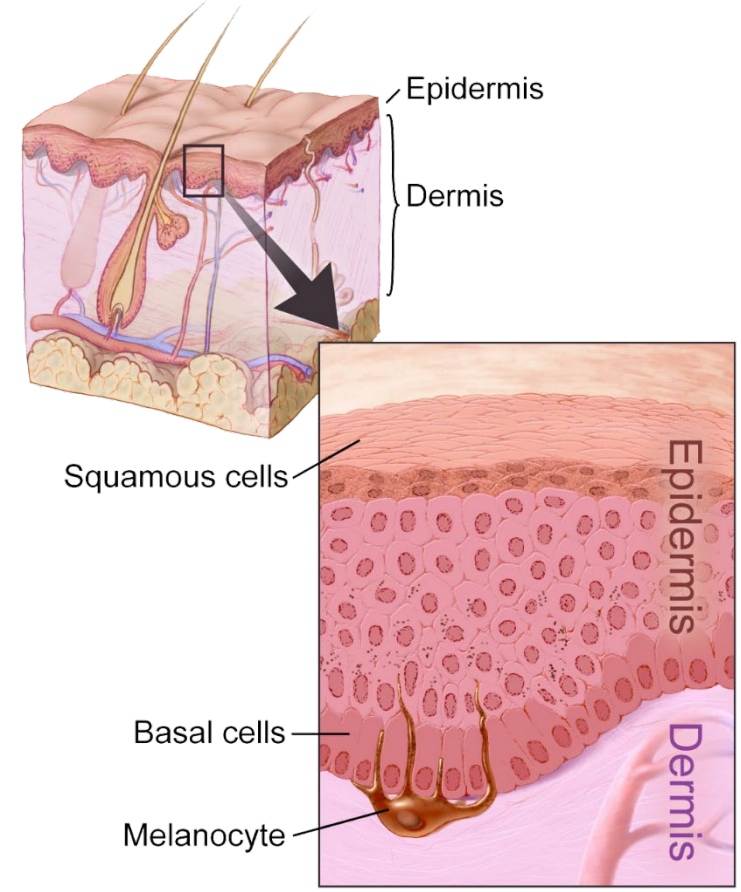
In how many cells were they expecting to see changes?

Few cells (10% at most)


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
Change technology
Enrich interesting cells
Cell-based sequencing
3’- vs 5’-end based library preparation
Short-read vs long-read sequencing
FACS (sorting)
Single-cell RNA-seq


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
Change technology
Enrich interesting cells
Cell-based sequencing
3’- vs 5’-end based library preparation
Short-read vs long-read sequencing
FACS (sorting)
Single-cell RNA-seq

Genes expressed in few cells



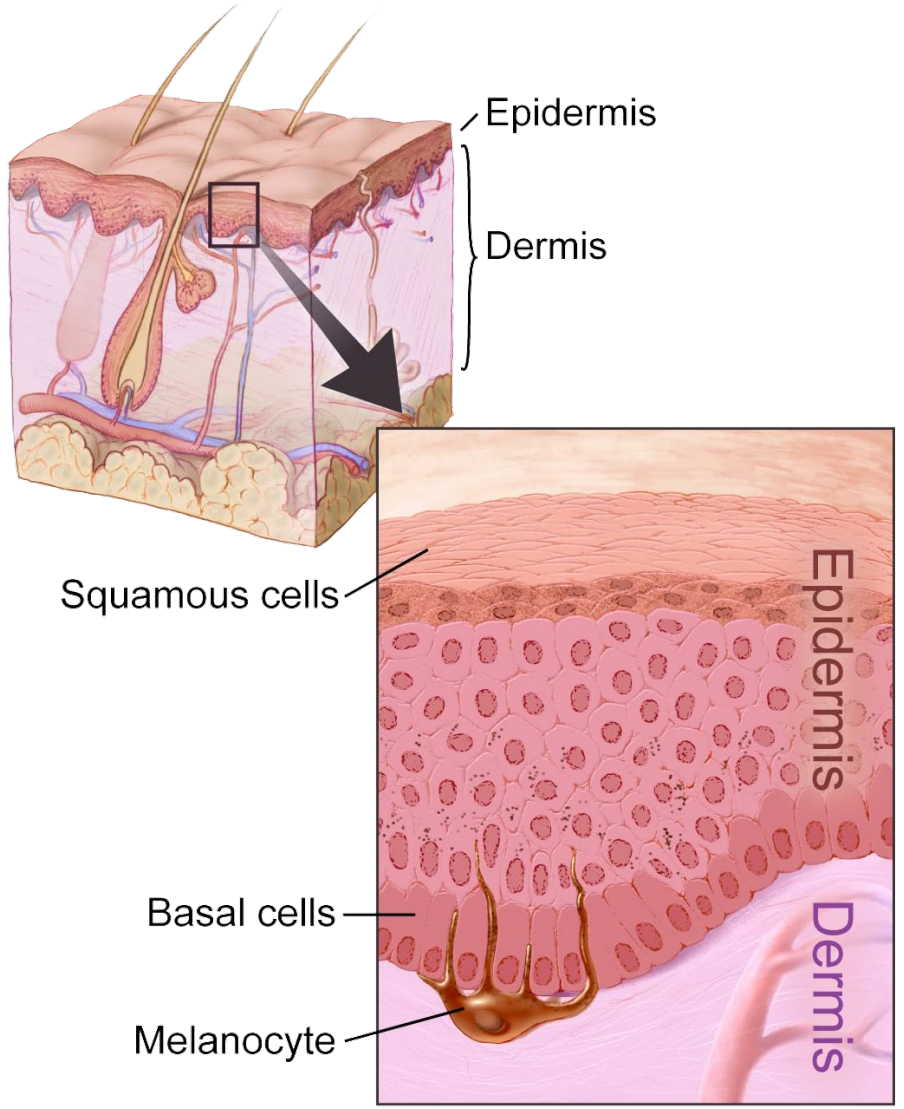
RNA-seq

Genes expressed in few cells

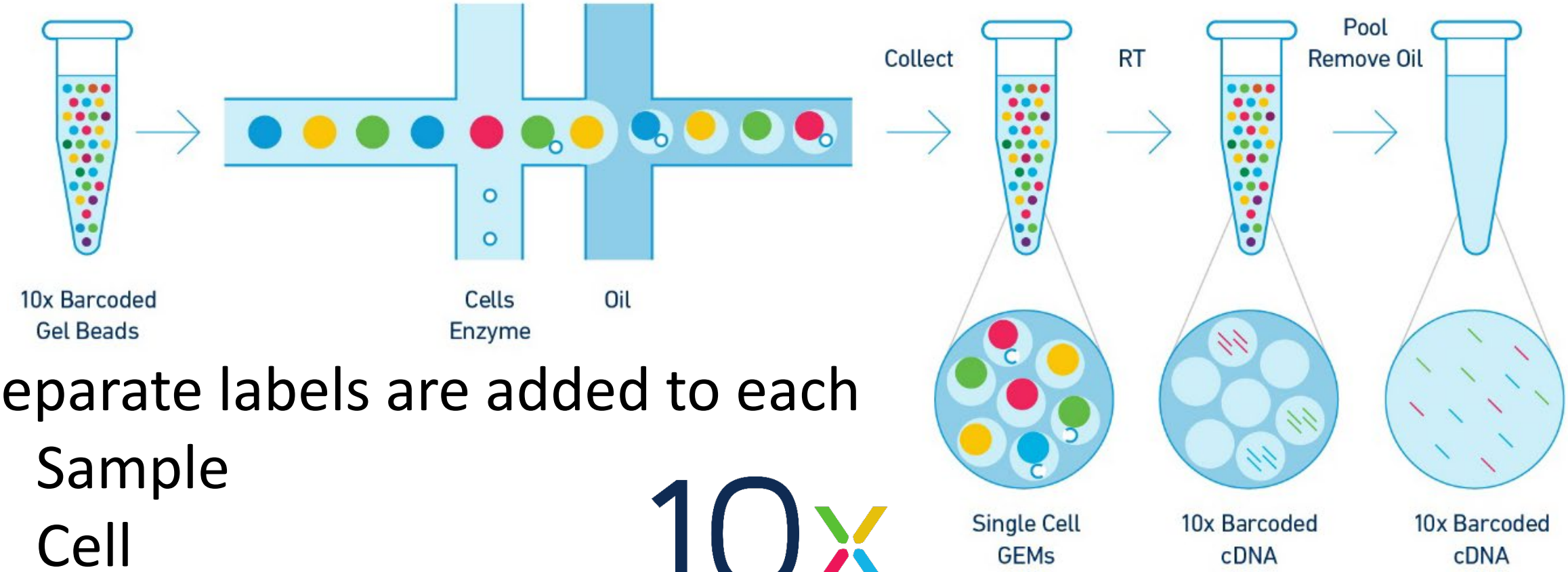
RNA-seq

Single-cell RNA-seq

- Melanocytes constitute 5%-10% of the basal layer of epidermis, freckle forming cells **even less**
- Approx. 1500 cells were isolated from each condition (Approx. 9000 total)
- mRNA was sequenced at the 3’ end using 10x Genomics technology



Droplet-based microfluidics

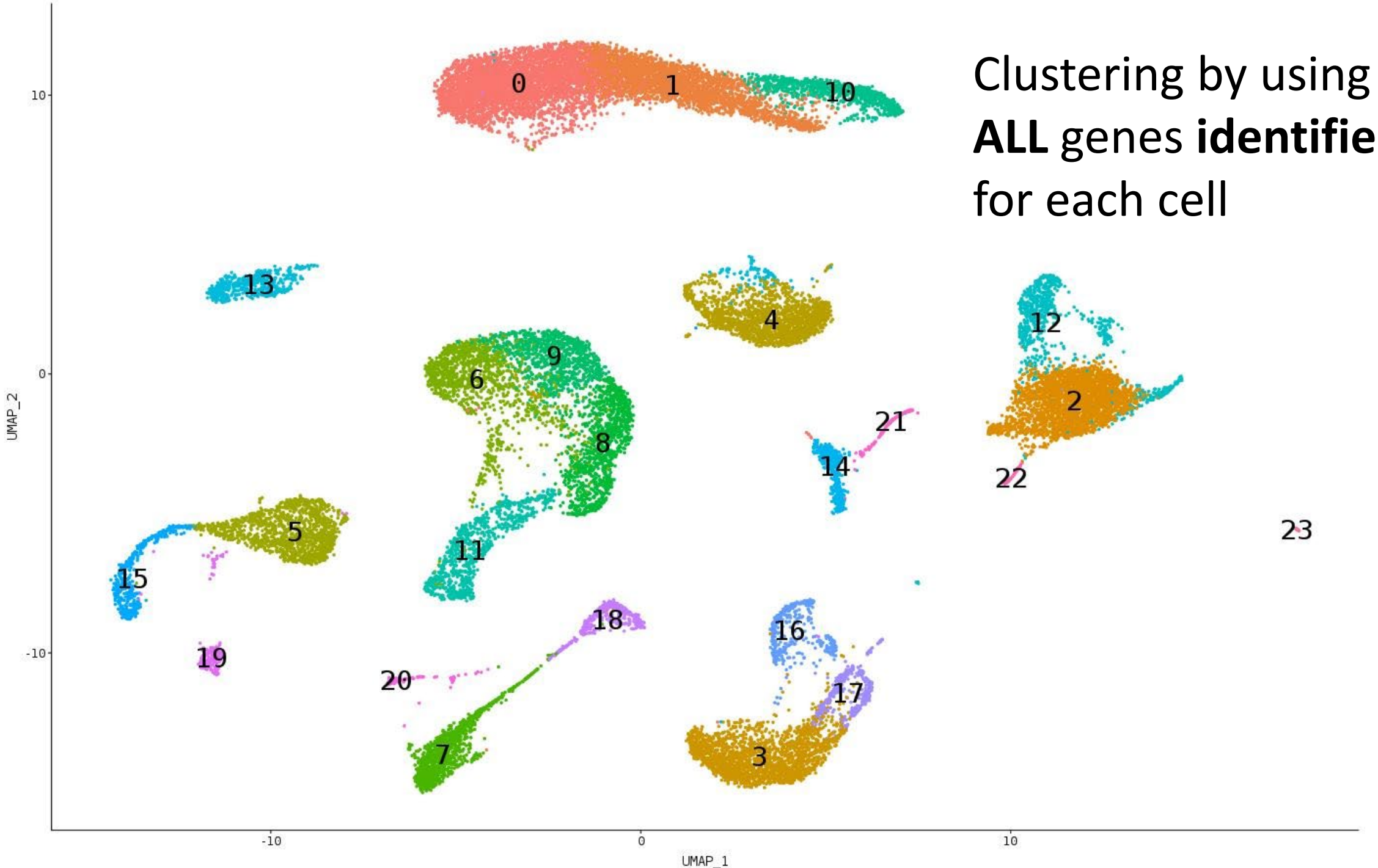


Separate labels are added to each

- Sample
- Cell
- Gene

10x GENOMICS®

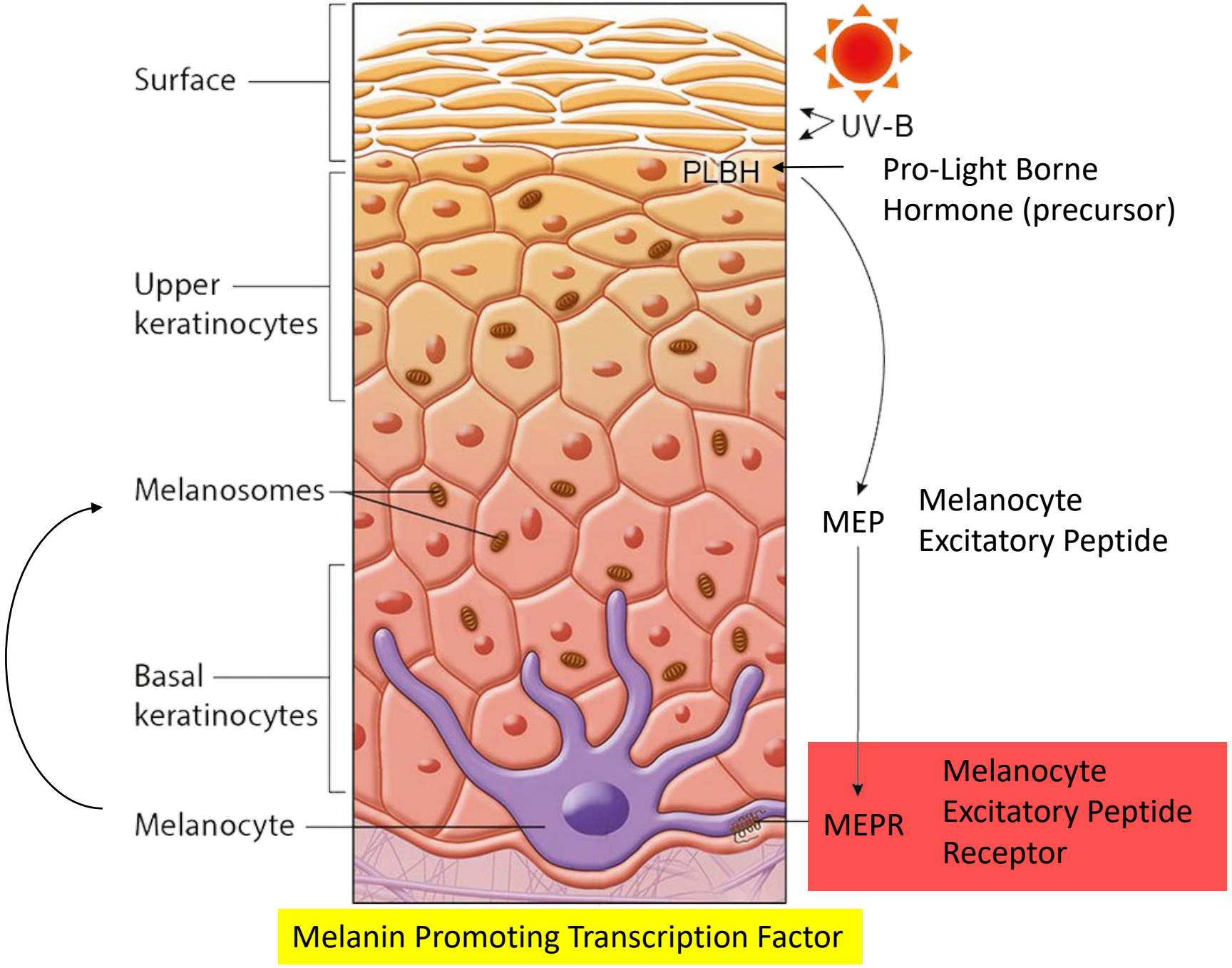
	Control 1				Control 2		Control 3			UV-B 1
	Cell AGT	Cell TTA	Cell ATT	Cell GTG	Cell CCT	Cell ATA	Cell GGG	Cell GTA	Cell AAA	Cell TTT
Gene 1	1				2	3	3	3	1	
Gene 2		4						2		
Gene 3		59		55	106	119	130		82	
Gene 4	3	9				5				
Gene 5					33		44			
Gene 6			19			16		23	15	



Clustering by using **ALL** genes **identified** for each cell

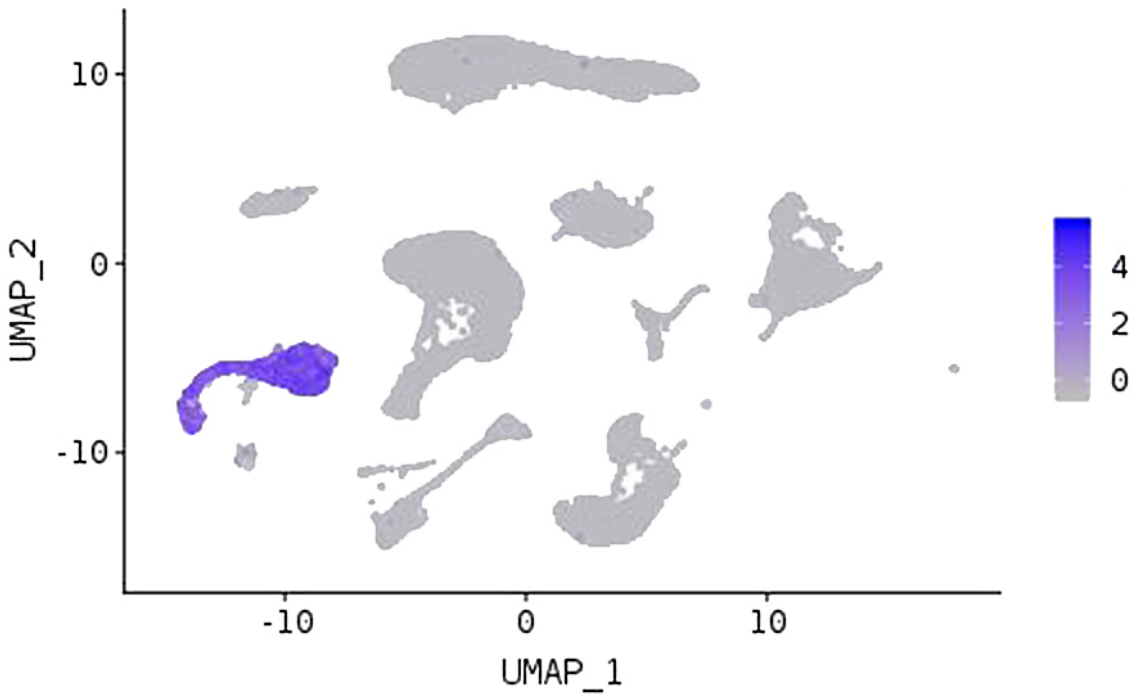
Why are freckles born?

Overproduce melanin granules

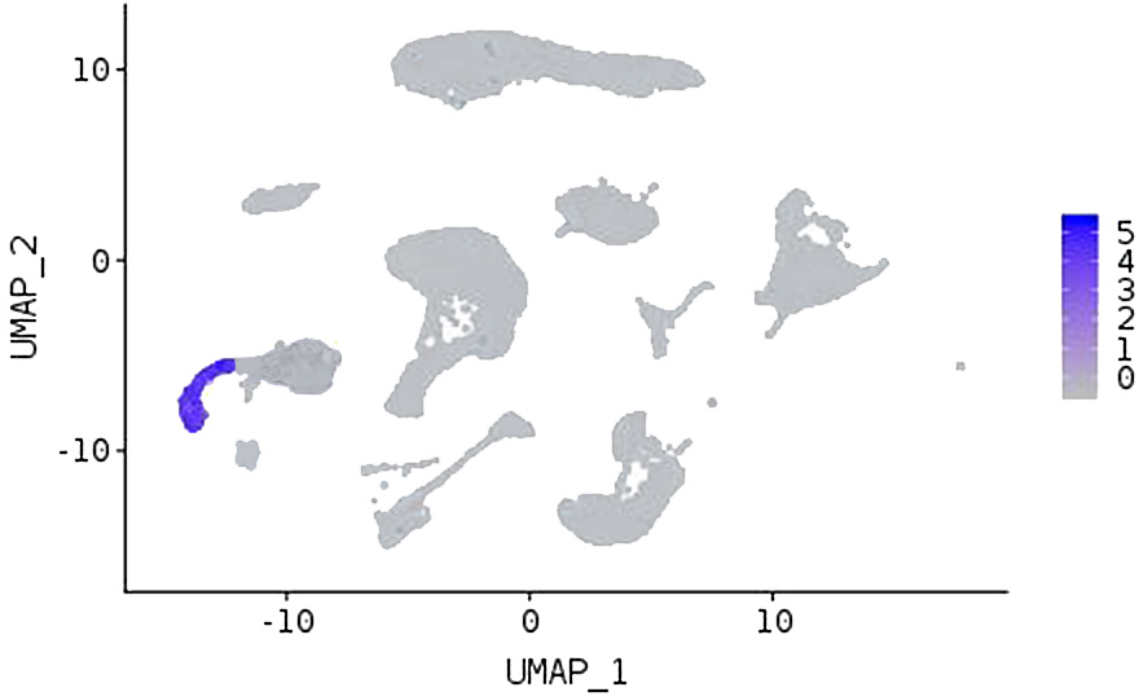


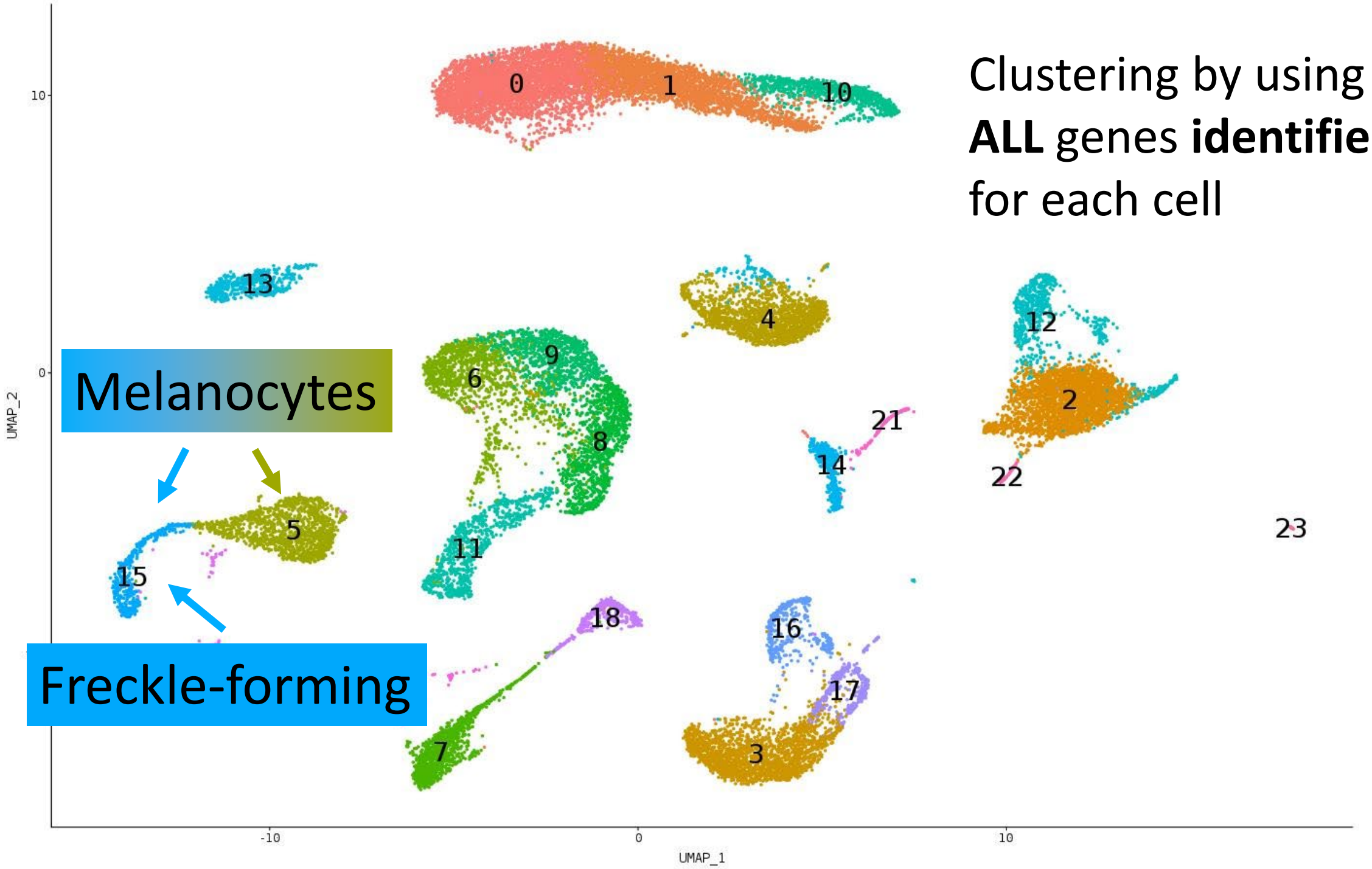
Melanocyte specific markers were used to identify clusters containing them

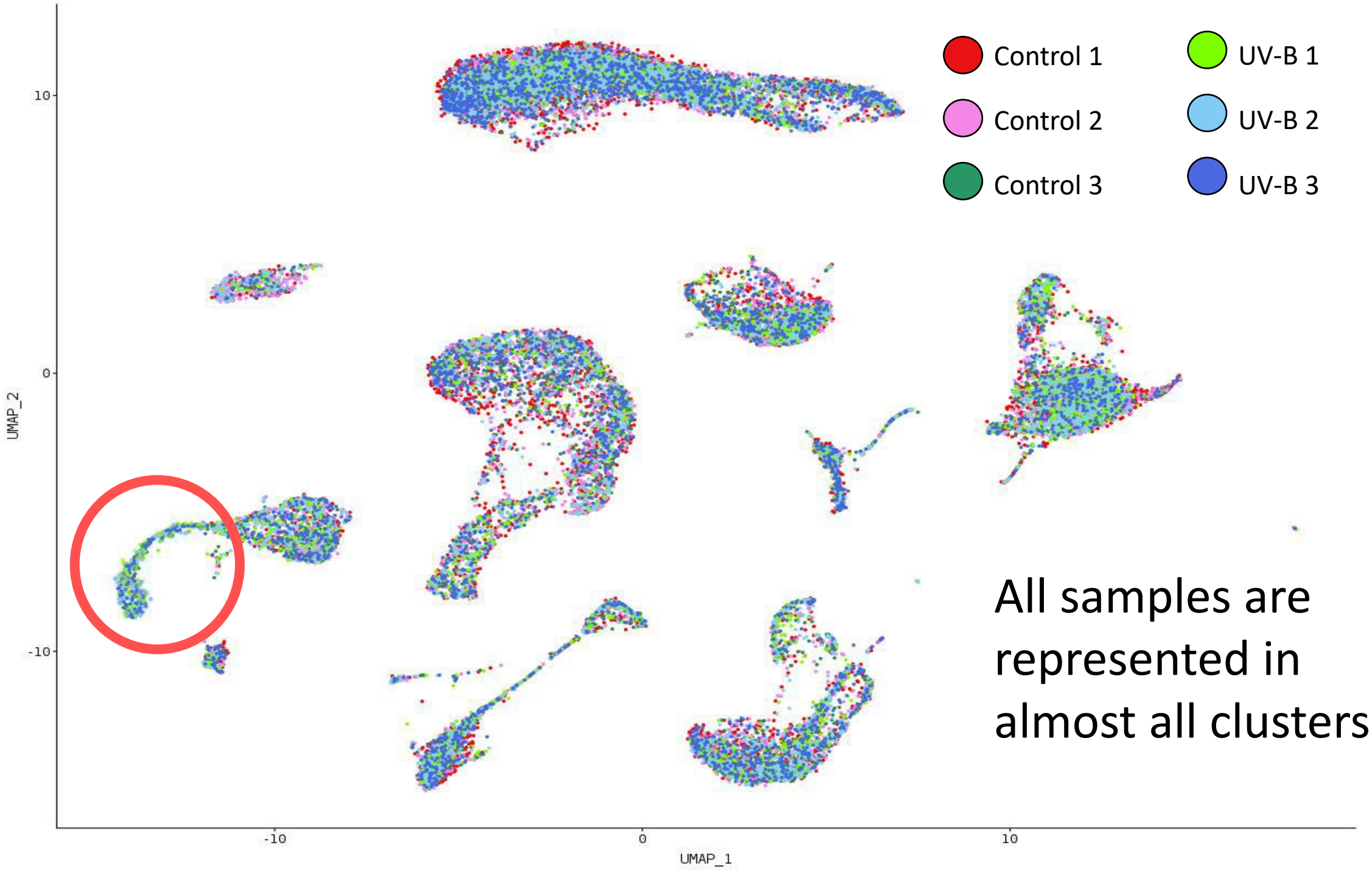
MEPR – Melanocyte Excitatory Peptide Receptor – **All** melanocytes



MLN2 – Melanin synthase – **freckle-forming** melanocytes




Clustering by using **ALL** genes **identified** for each cell
Melanocytes
Freckle-forming
0
1
10
13
4
12
9
6
2
8
21
14
22
5
23
11
15
18
16
17
7
3
10
0
-10
0
10
UMAP_2
UMAP_1



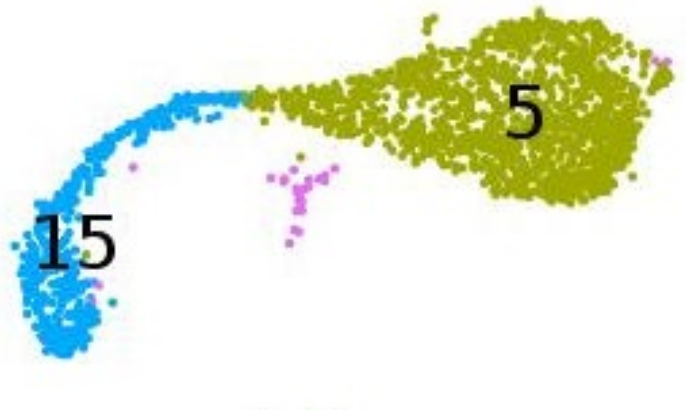
All samples are represented in almost all clusters

Found markers for clusters 5 and 15

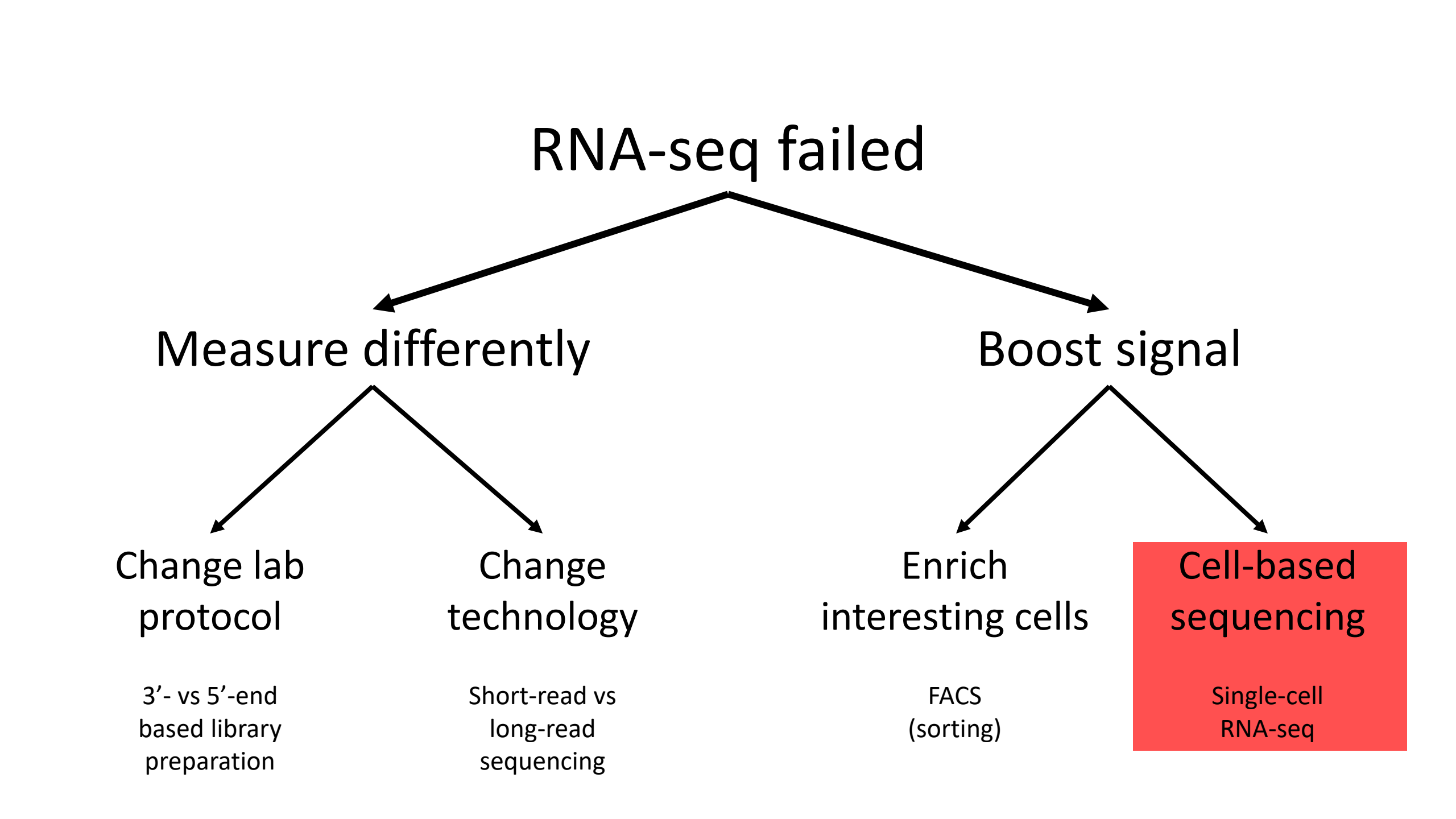
- Cluster 15
  - Produced melanin
  - Comprised of only UV-B samples
  - Smaller

  Probably freckle-forming melanocytes
- Clusters 5 and 15 were compared to all other clusters to find differentially expressed genes

Cluster 5 - 689 cells



Cluster 15 - 191 cells


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
Change technology
Enrich interesting cells
Cell-based sequencing
3’- vs 5’-end based library preparation
Short-read vs long-read sequencing
FACS (sorting)
Single-cell RNA-seq

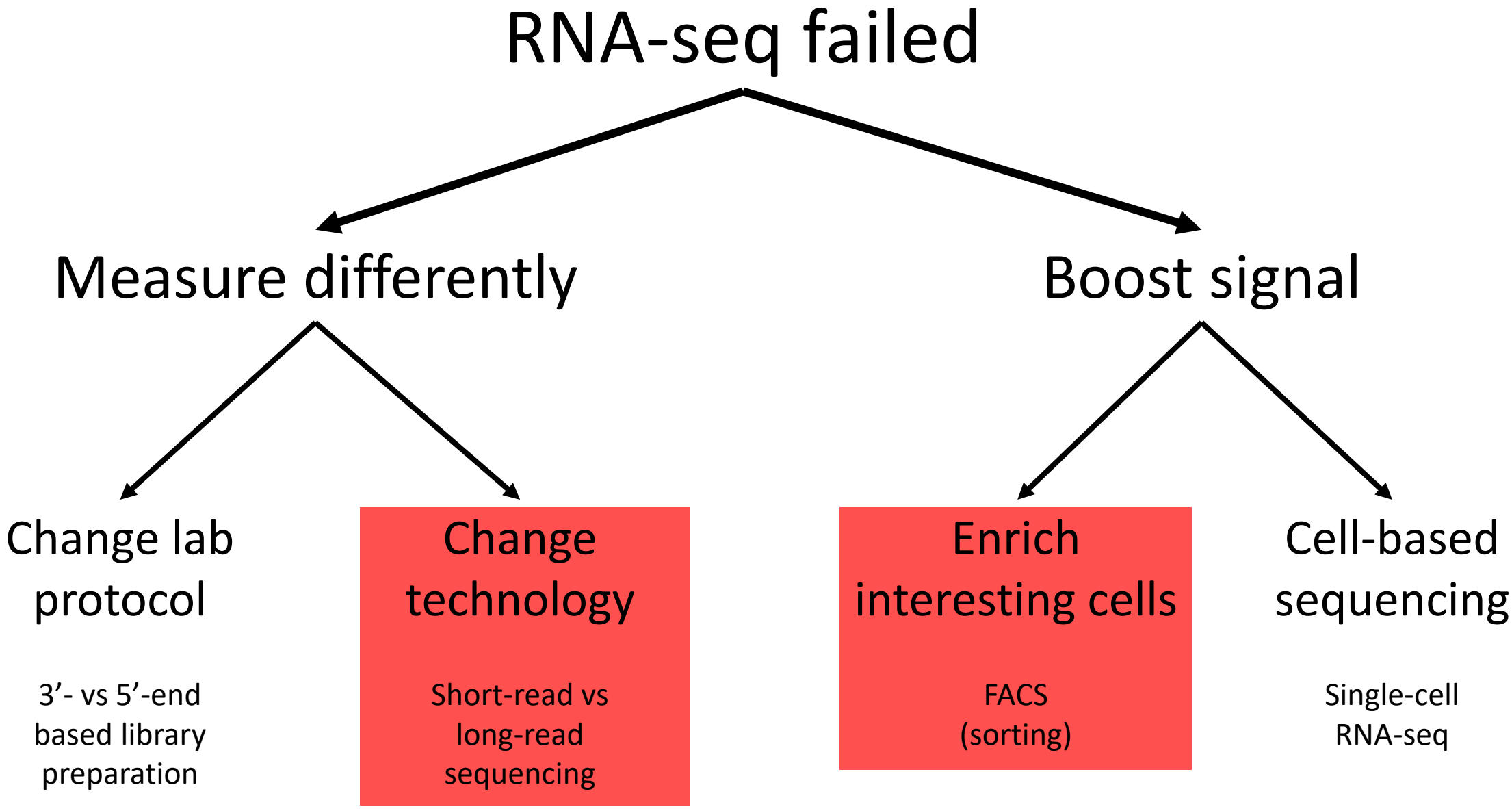
Conclusions so far

- Melanin production and export pathways were elevated
- Transcriptional repressors were elevated
- Cluster 15 expressed
  - Sng1 – a specific transmembrane protein
  - Protein with unknown function has interaction domains
- Cannot easily explain mechanism that **<u>activated</u>** these pathways
- **29 genes to check or maybe we are missing something (wider scope)**

Who is giving the orders?

No obvious candidates

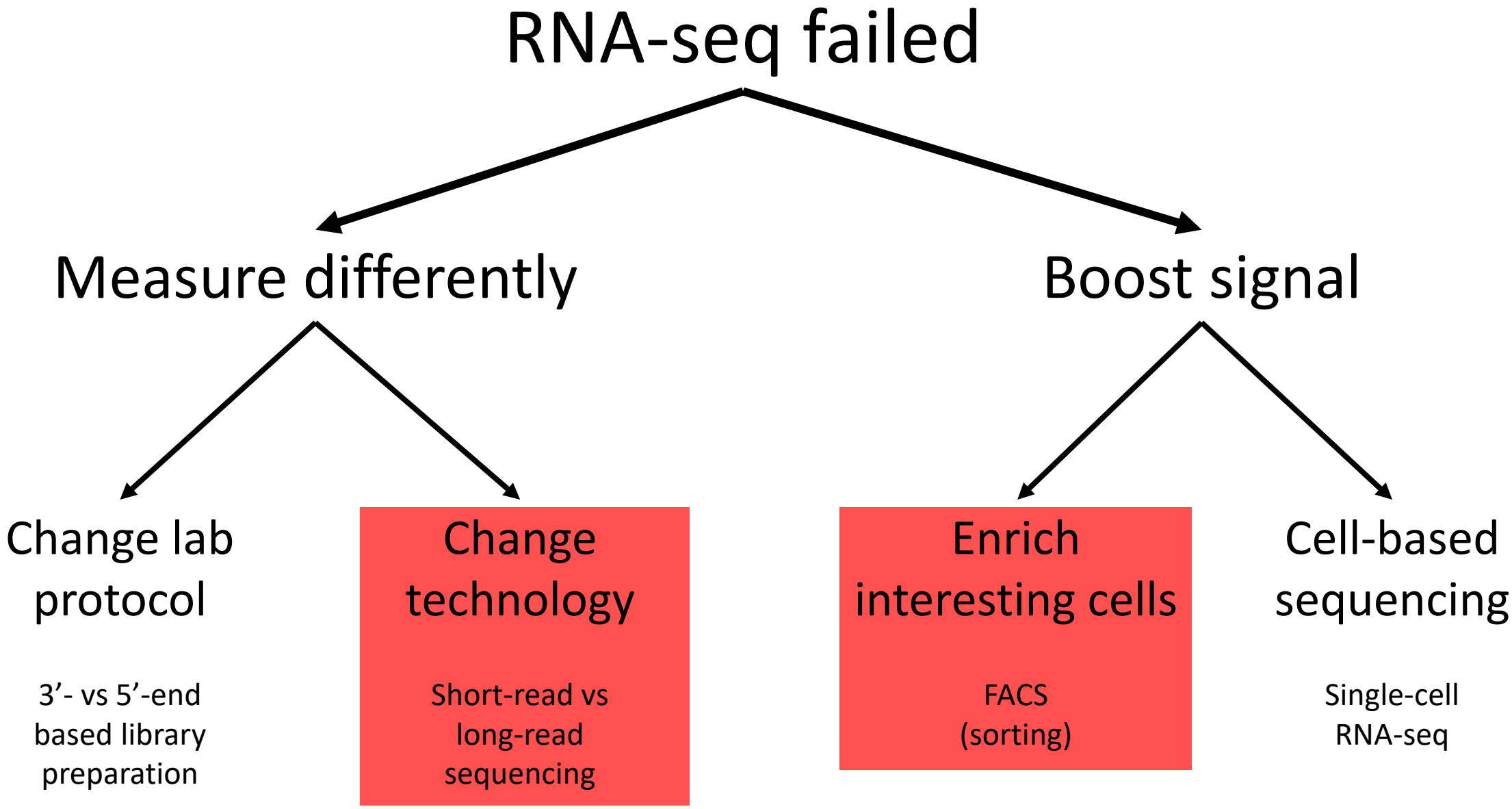
However


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
3’- vs 5’-end based library preparation
Change technology
Short-read vs long-read sequencing
Enrich interesting cells
FACS (sorting)
Cell-based sequencing
Single-cell RNA-seq

Conclusions so far

- Melanin production and export pathways were elevated
- Transcriptional repressors were elevated
- Cluster 15 expressed
  - Sng1 – a specific transmembrane protein
  - Protein with unknown function has interaction domains
- Cannot easily explain mechanism that **<u>activated</u>** these pathways
- **29 genes to check or maybe we are missing something (wider scope)**

Who is giving the orders?


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
3’- vs 5’-end based library preparation
Change technology
Short-read vs long-read sequencing
Enrich interesting cells
FACS (sorting)
Cell-based sequencing
Single-cell RNA-seq

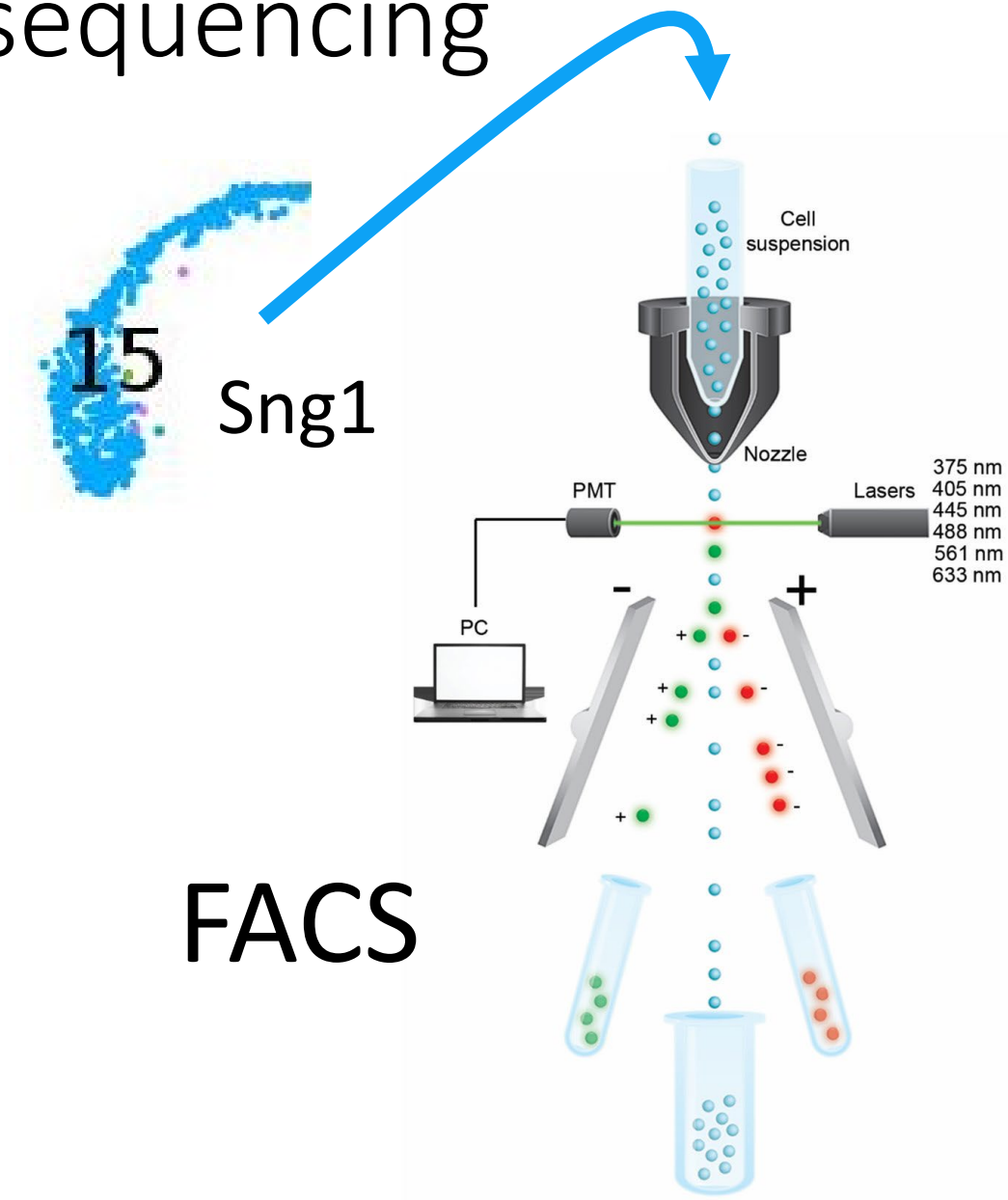
Preparation for long-read sequencing

Few cells

Original experiment conditions

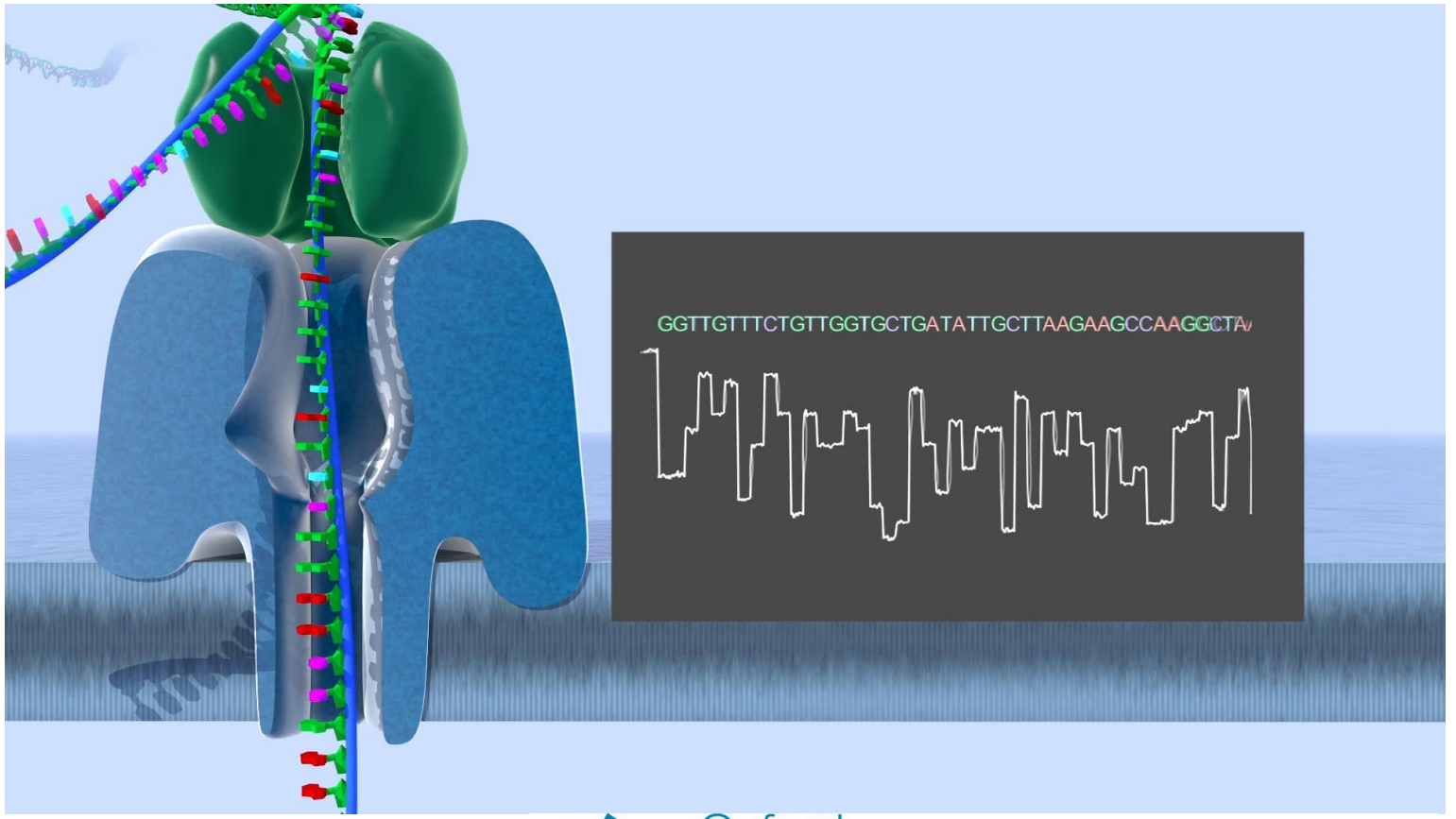
Flow cytometry enrichment of **Sng1** expressed on **cluster 15** cell membranes

Cluster 15 cell enrichment



Long read sequencing (third generation)

- Sequence by measuring current of polymers (nucleic acids) passing through a nanopore
- Length is usually over 1000bp
- Can potentially reveal new transcript structure



Oxford NANOPORE Technologies

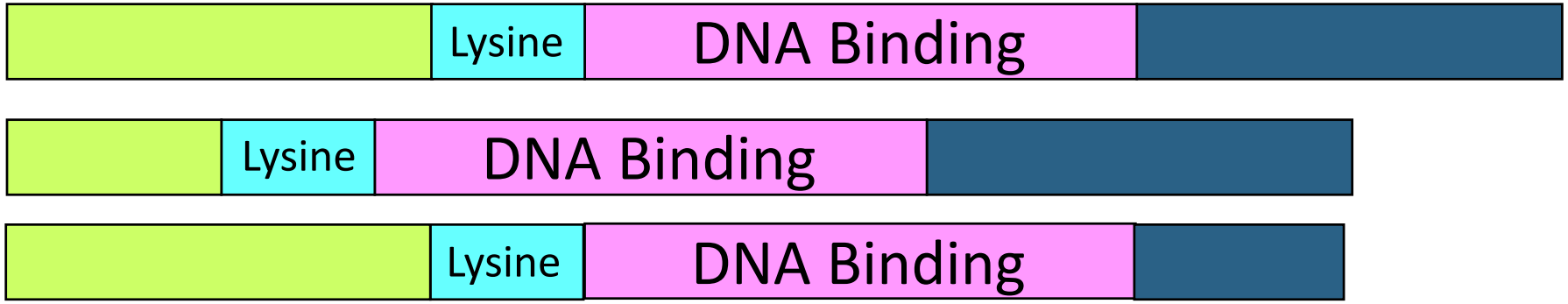
Conclusions so far

- Melanin production and export pathways were elevated
- Transcriptional repressors were elevated
- Cluster 15 expressed
  - Sng1 – a specific transmembrane protein
  - Protein with unknown function has interaction domains
- Cannot easily explain mechanism that **activated** these pathways
- **29 genes to check or maybe we are missing something (wider scope)**

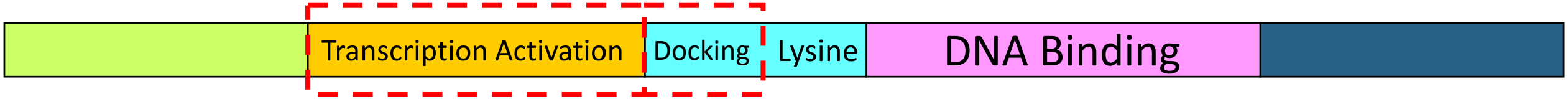
Who is giving the orders?

Nmsh3 – new transcript structure

Transcriptional repressors

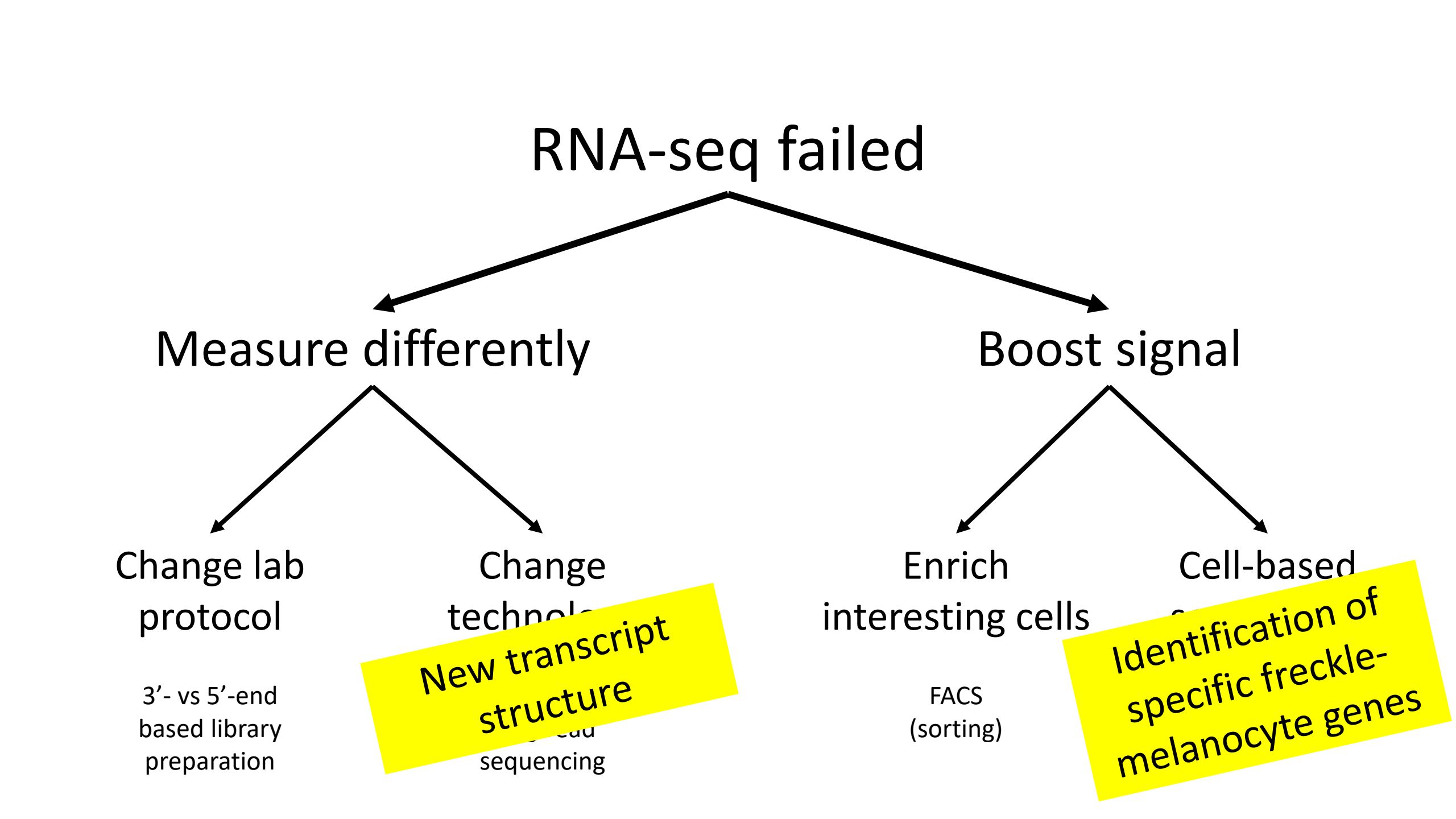


Transcriptional activator



Transcriptional activation domain – attracts basal transcription machinery

Protein interaction domain with Degbr2 (found via scRNA-seq) helps hide lysine rich domain and avoid degradation


RNA-seq failed
Measure differently
Boost signal
Change lab protocol
Change
3'- vs 5'-end based library preparation
sequencing
New transcript structure
Enrich interesting cells
Cell-based
FACS (sorting)
Identification of specific freckle-melanocyte genes

The example story was TOTALLY invented

Summary

Next Generation Sequencing is a robust and flexible set of experimental methods

The experimental technology details are crucial for understanding the obtained results and thus should be inspected carefully

New sequencing technologies are emerging which may answer new questions