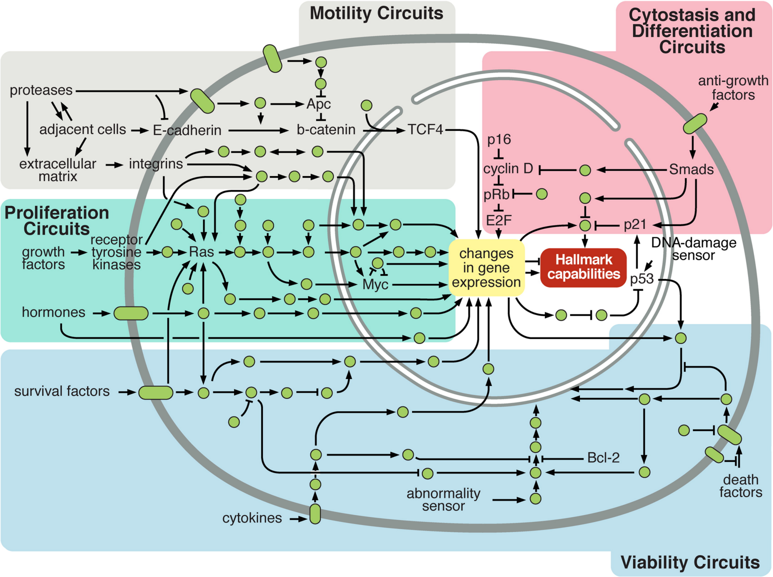
CS 5854: Single-Cell Gene Expression Analysis

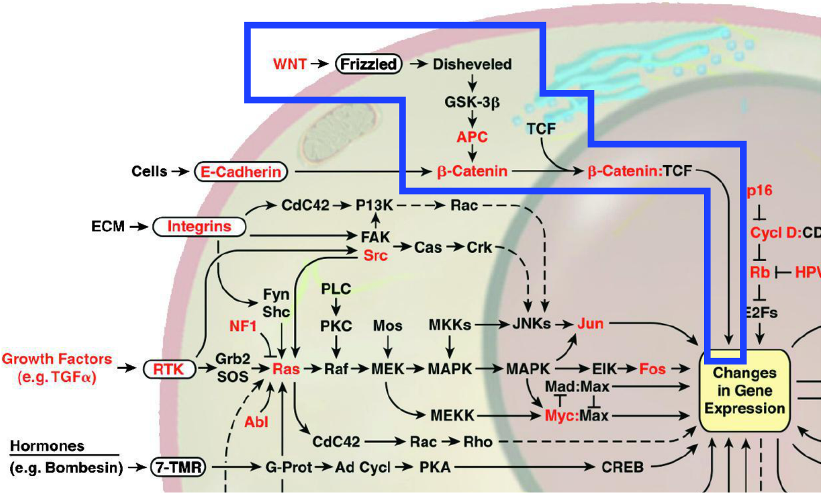
T. M. Murali

March 17, 19, 2025

Signaling Pathways and Gene Expression

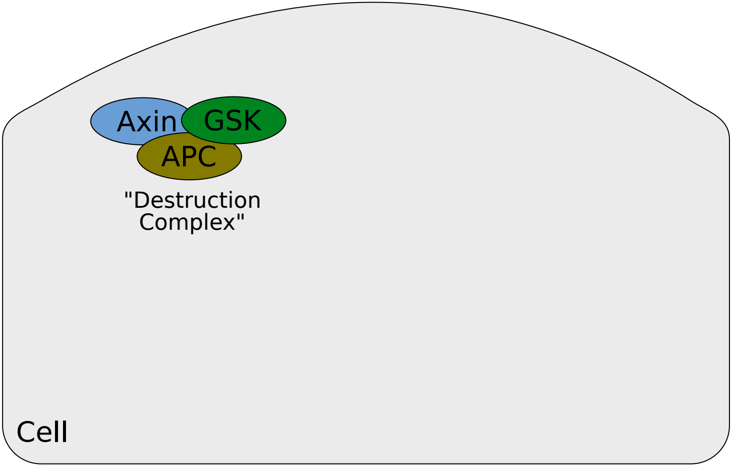


Wnt Signaling

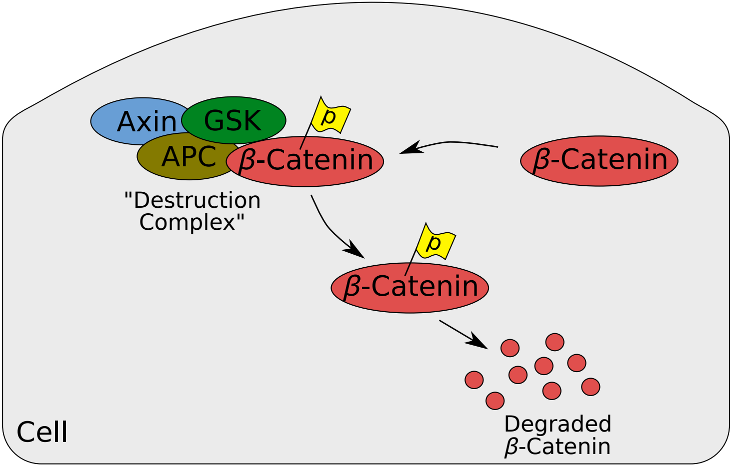


Hanahan and Wienberg. Hallmarks of cancer. Cell, 2000.

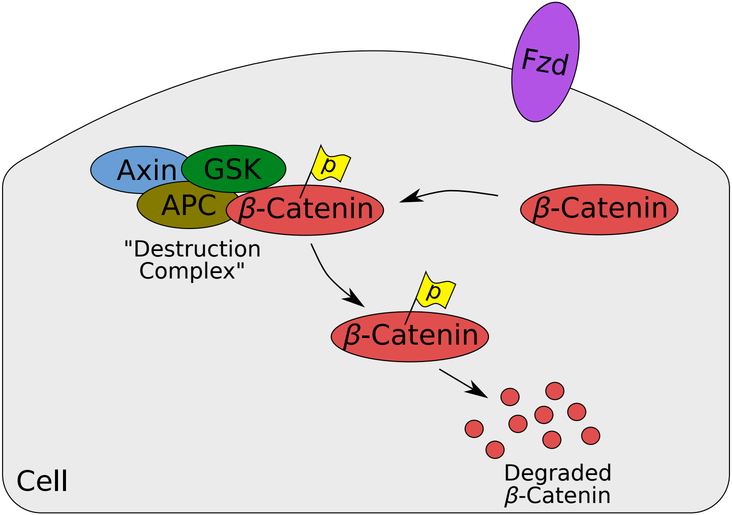
In the Absence of Wnt



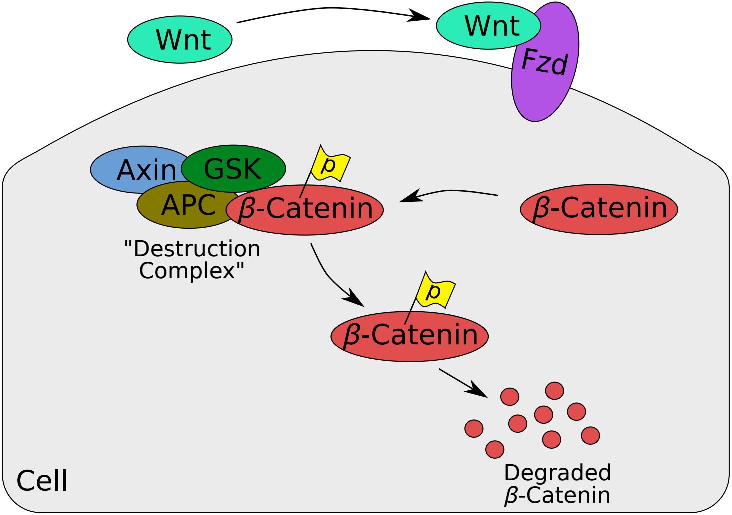
In the Absence of Wnt



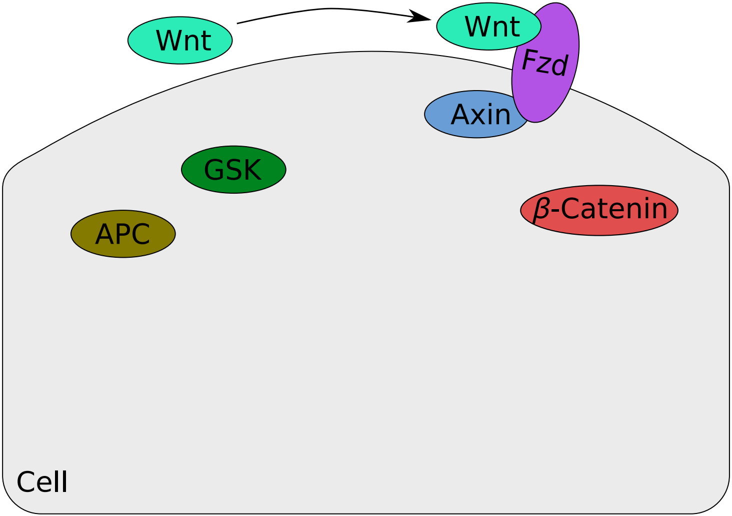
In the Presence of Extracellular Wnt



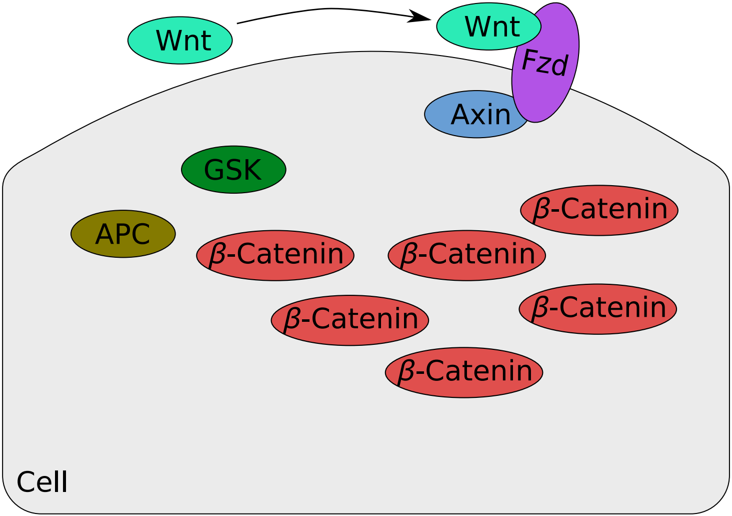
In the Presence of Extracellular Wnt



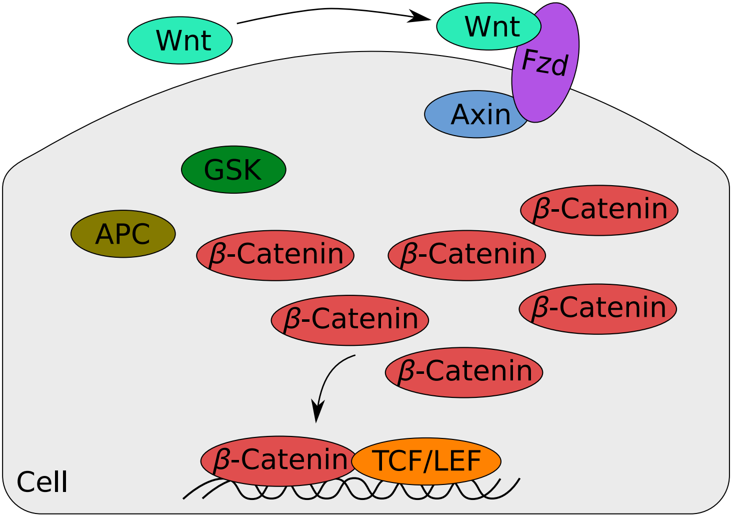
In the Presence of Extracellular Wnt



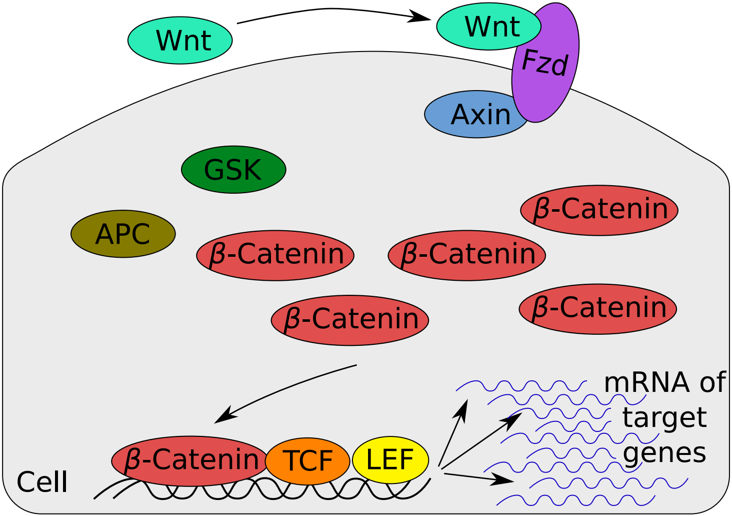
In the Presence of Extracellular Wnt



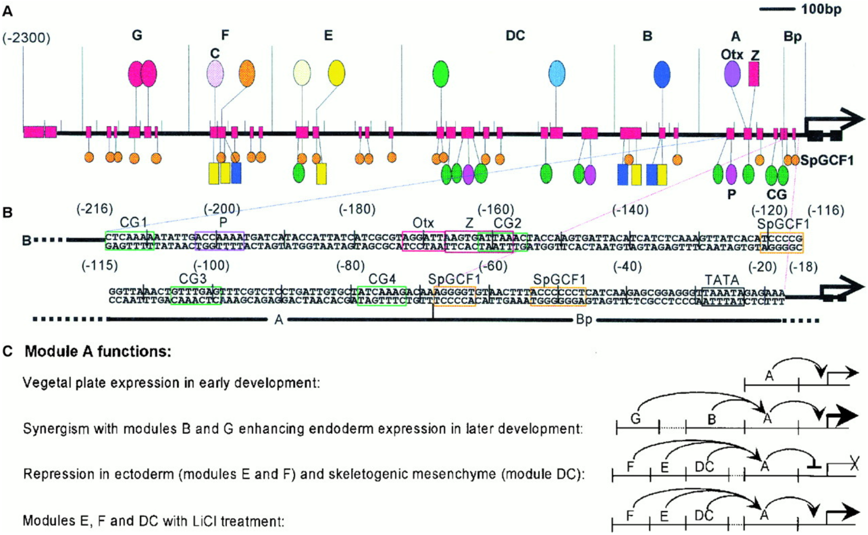
In the Presence of Extracellular Wnt



In the Presence of Extracellular Wnt



Gene Expression is a Dynamic Process



Module A functions:

Vegetal plate expression in early development:

Synergism with modules B and G enhancing endoderm expression in later development:

Repression in ectoderm (modules E and F) and skeletogenic mesenchyme (module DC):

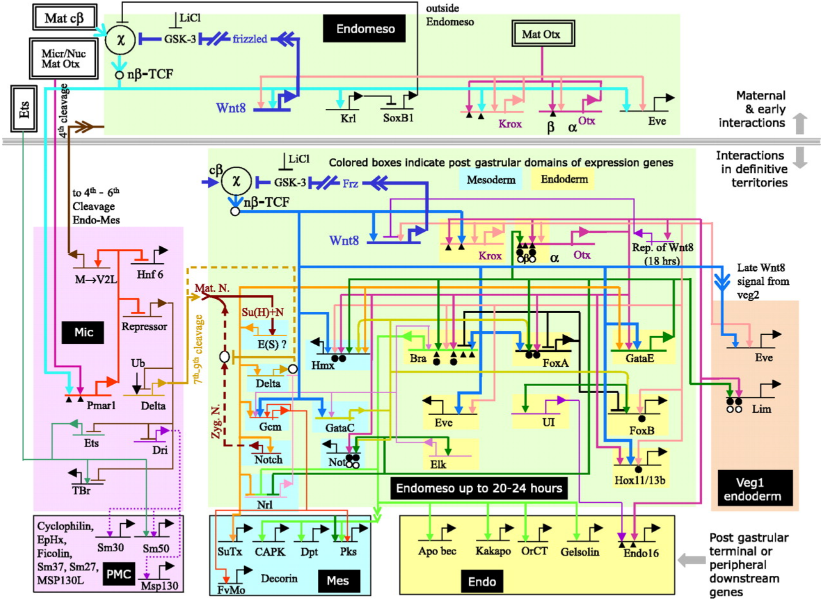
Modules E, F and DC with LiCl treatment:

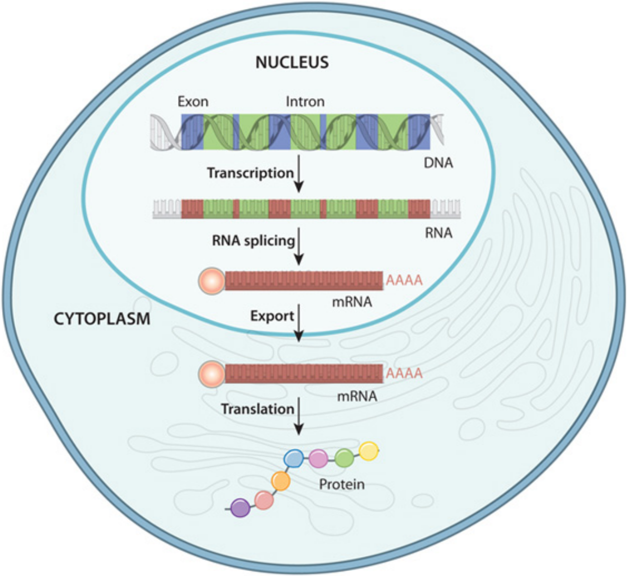
Gene Expression is a Dynamic Process

B

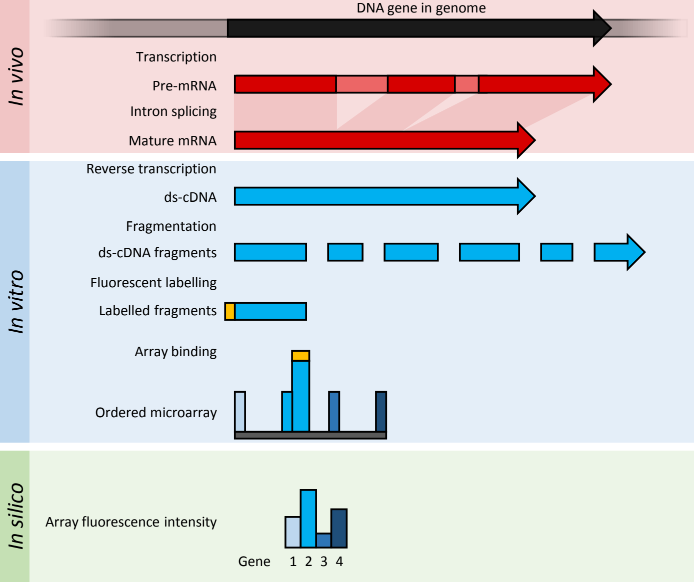
if ($F = 1$ or $E = 1$ or $CD = 1$) and ($Z = 1$) $\alpha = 1$ **else** $\alpha = 0$	Repression functions of modules F, E, and DC mediated by Z site
if ($P = 1$ and $CG_1 = 1$) $\beta = 2$ **else** $\beta = 0$	Both P and CG_1 needed for synergistic link with module B
if ($CG_2 = 1$ and $CG_3 = 1$ and $CG_4 = 1$) $\gamma = 2$ **else** $\gamma = 1$	Final step up of system output
$\delta(t) = B(t) + G(t)$	Positive input from modules B and G
$\varepsilon(t) = \beta * \delta(t)$	Synergistic amplification of module B output by CG_1-P subsystem
if ($\varepsilon(t) = 0$) $\xi(t) = Otx(t)$ **else** $\xi(t) = \varepsilon(t)$	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
if ($\alpha = 1$) $\eta(t) = 0$ **else** $\eta(t) = \xi(t)$	Repression function inoperative in endoderm but blocks activity elsewhere
$\Theta(t) = \gamma * \eta(t)$	Final output communicated to BTA

Gene Expression is a Dynamic Process



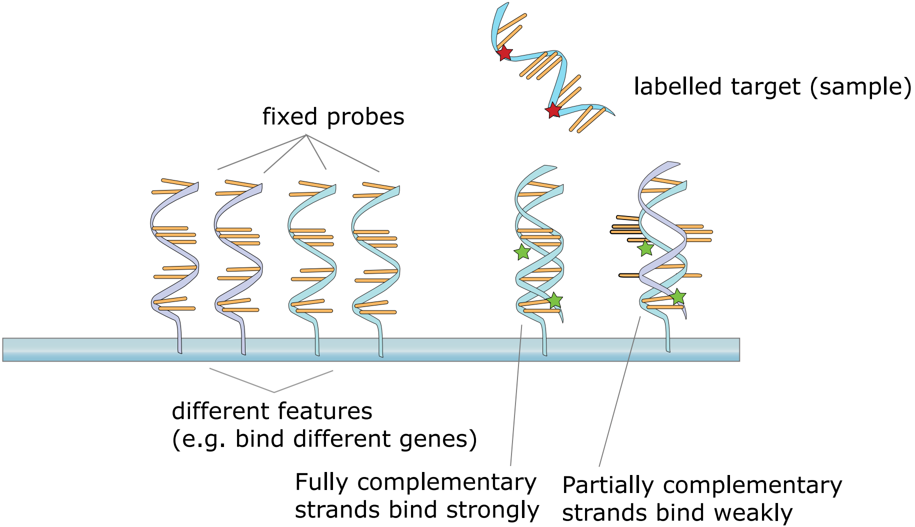

NUCLEUS
Exon
Intron
DNA
Transcription
RNA
RNA splicing
AAAA
mRNA
CYTOPLASM
Export
AAAA
mRNA
Translation
Protein

Measuring Genomewide Gene Expression: DNA Microarrays



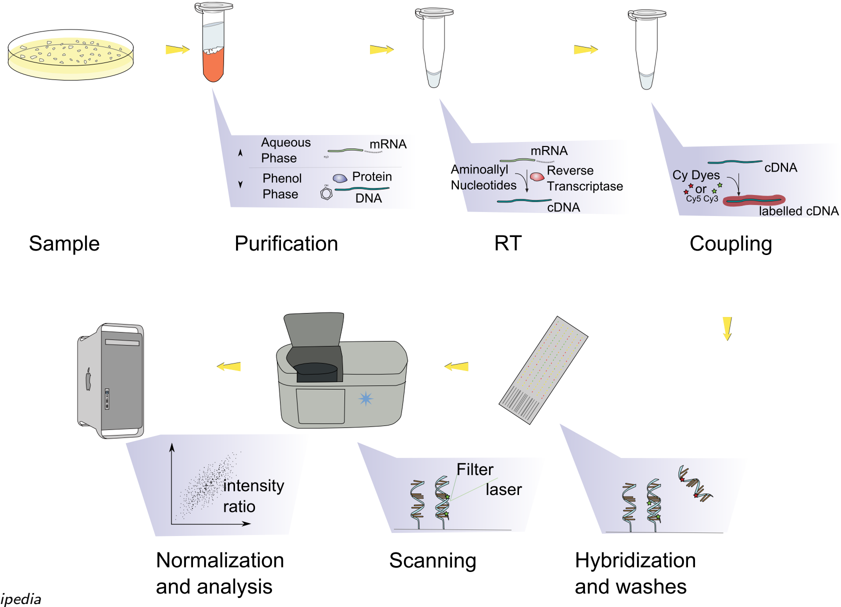
Wikipedia

Measuring Genomewide Gene Expression: DNA Microarrays



Wikipedia

Measuring Genomewide Gene Expression: DNA Microarrays



Wikipedia

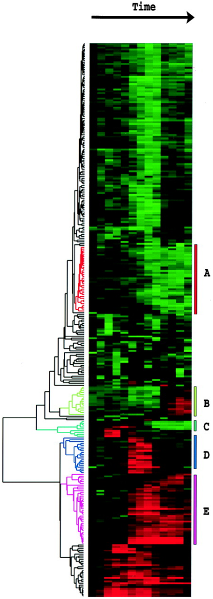
Applications of DNA Microarray Data

RESEARCH ARTICLE

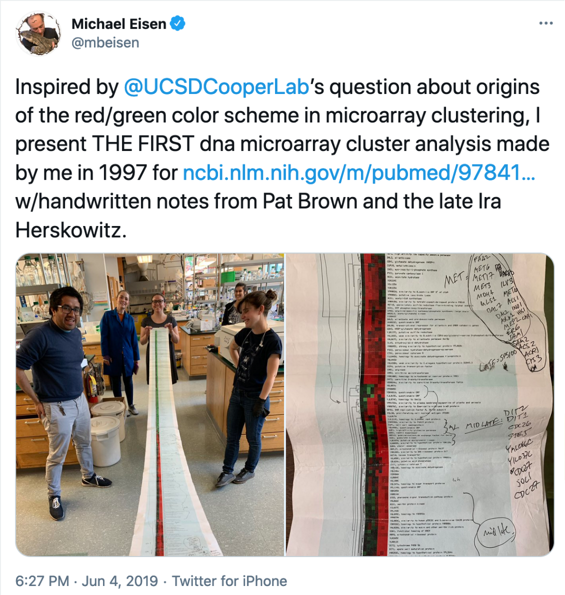
Cluster analysis and display of genome-wide expression patterns

Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein

PNAS December 8, 1998 95 (25) 14863-14868;



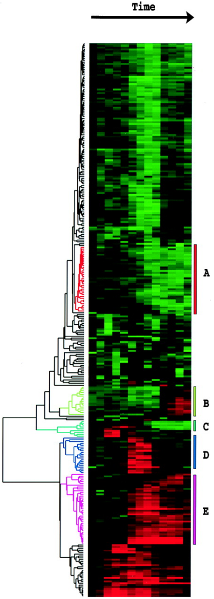
Applications of DNA Microarray Data



Michael Eisen
@mbeisen

Inspired by @UCSDCooperLab's question about origins of the red/green color scheme in microarray clustering, I present THE FIRST dna microarray cluster analysis made by me in 1997 for ncbi.nlm.nih.gov/m/pubmed/97841… w/handwritten notes from Pat Brown and the late Ira Herskowitz.

6:27 PM · Jun 4, 2019 · Twitter for iPhone



Applications of DNA Microarray Data

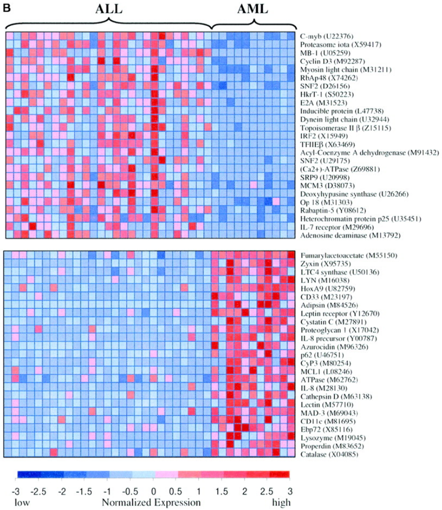
REPORT

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

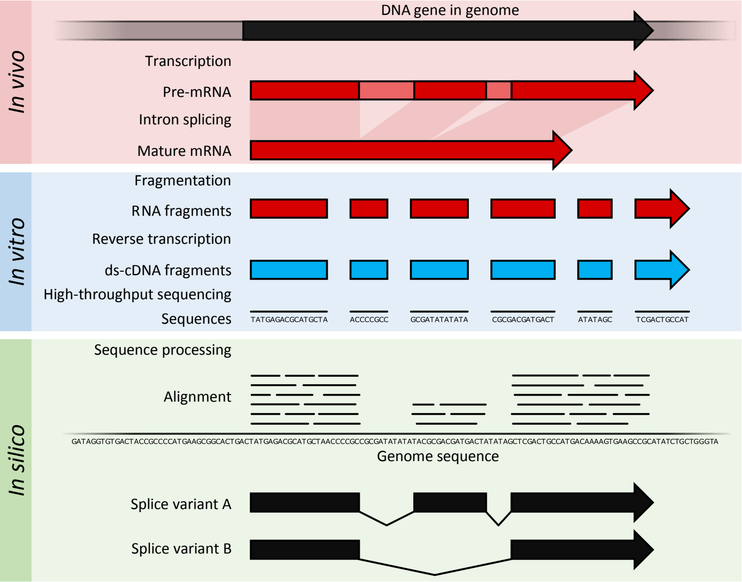
T. R. Golub[1,2,*,†], D. K. Slonim[1,†], P. Tamayo[1], C. Huard[1], M. Gaasenbeek[1], J. P. Mesirov[1], H. Coller[1], M. L. Loh[2], J. R. Downin…

+ See all authors and affiliations

Science 15 Oct 1999:
Vol. 286, Issue 5439, pp. 531-537
DOI: 10.1126/science.286.5439.531



Measuring Genomewide Gene Expression: RNA-seq



In vivo

DNA gene in genome

Transcription

Pre-mRNA

Intron splicing

Mature mRNA

In vitro

Fragmentation

RNA fragments

Reverse transcription

ds-cDNA fragments

High-throughput sequencing

Sequences

TATGAGACGCATGCTA ACCCCGCC GCGATATATATA CGCGACGATGACT ATATAGC TCGACTGCCAT

In silico

Sequence processing

Alignment

GATAGGTGTGACTACCGCCCCATGAAGCGGCACTGACTATGAGACGCATGCTAACCCCGCCGCGATATATATACGCGACGATGACTATATAGCTCGACTGCCATGACAAAAGTGAAGCCGCATATCTGCTGGGTA

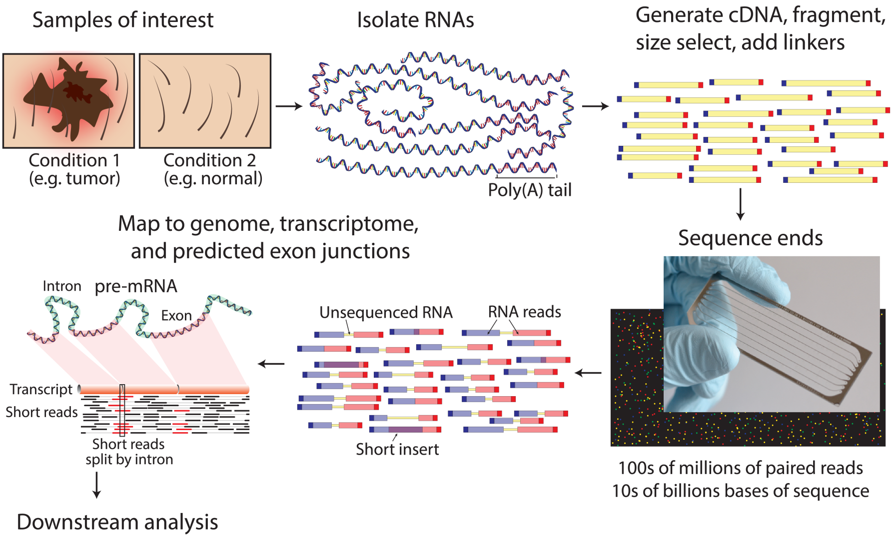
Genome sequence

Splice variant A

Splice variant B

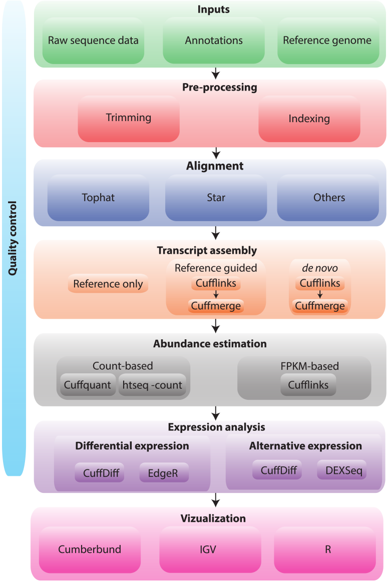
Wikipedia

Measuring Genomewide Gene Expression: RNA-seq



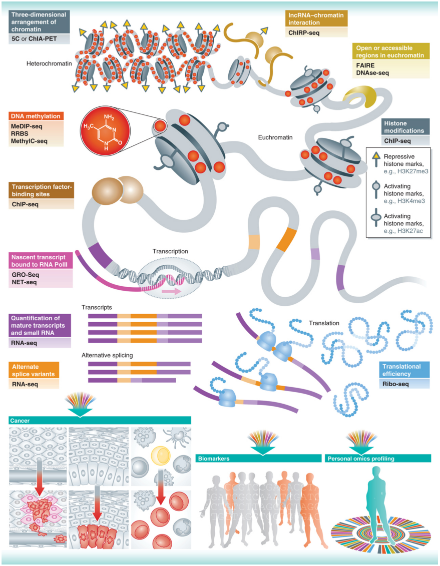
Griffith et al., Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLoS Comput Biol, 2015

Measuring Genomewide Gene Expression: RNA-seq



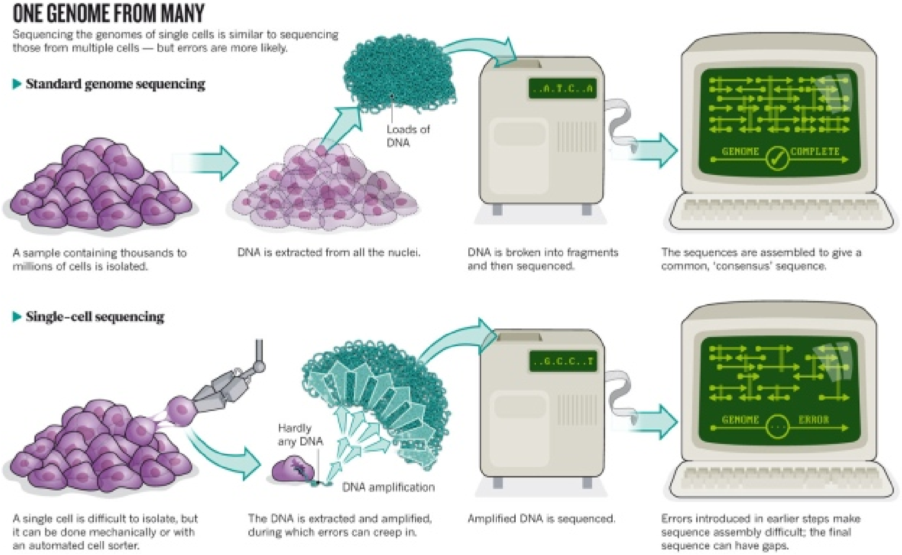
Griffith et al., Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLoS Comput Biol, 2015

*-Seq Techniques



Soon, Hariharan, and Snyder. High-throughput sequencing for biology and medicine. Mol. Sys. Bio, 2013.

Single-Cell RNA-Seq



ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

Standard genome sequencing

A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

The sequences are assembled to give a common, 'consensus' sequence.

Single-cell sequencing

A single cell is difficult to isolate, but it can be done mechanically or with an automated cell sorter.

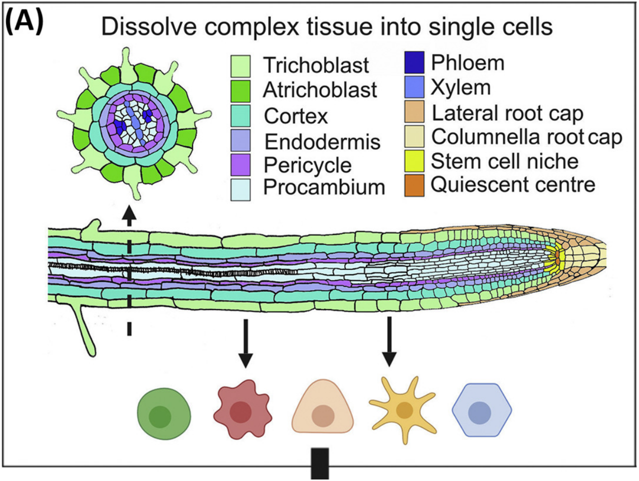
The DNA is extracted and amplified, during which errors can creep in.

Amplified DNA is sequenced.

Errors introduced in earlier steps make sequence assembly difficult; the final sequence can have gaps.

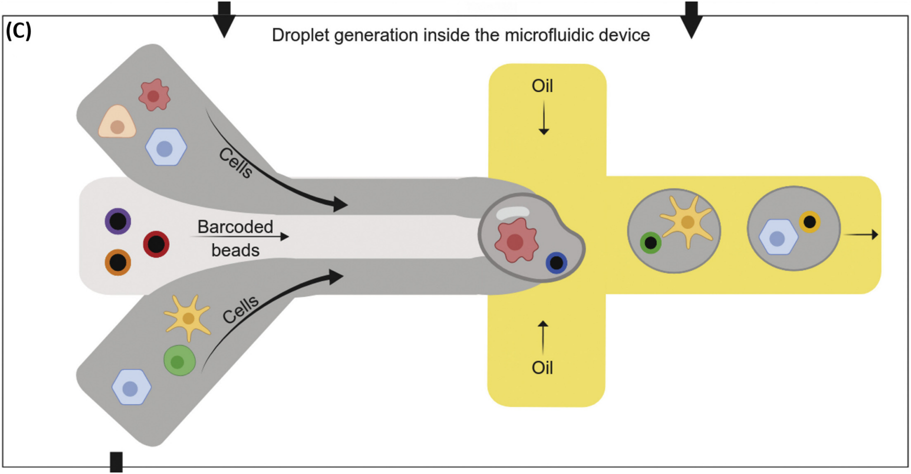
Owens, Genomics: The single life, Nature, 2012.

scRNA-seq Method



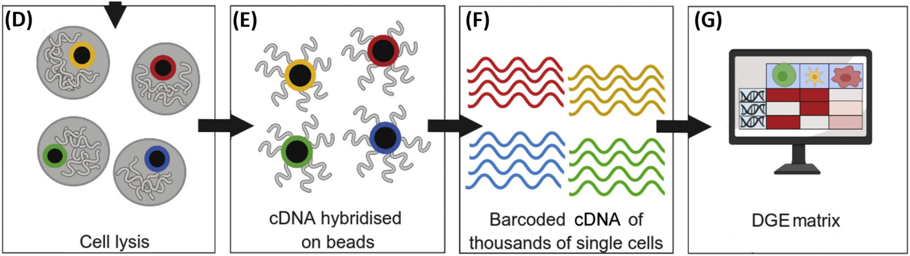
Single-Cell Transcriptomics: A High-Resolution Avenue for Plant Functional Genomics, Rich-Griffin et al., *Trends in Plant Science, 2020*

scRNA-seq Method



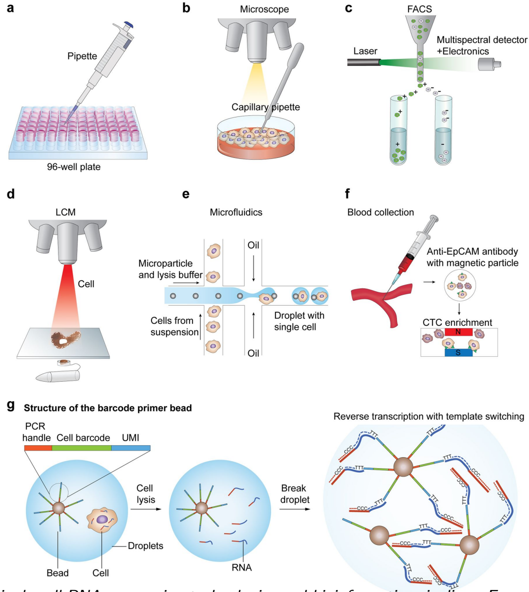
Single-Cell Transcriptomics: A High-Resolution Avenue for Plant Functional Genomics, Rich-Griffin et al., Trends in Plant Science, 2020

scRNA-seq Method



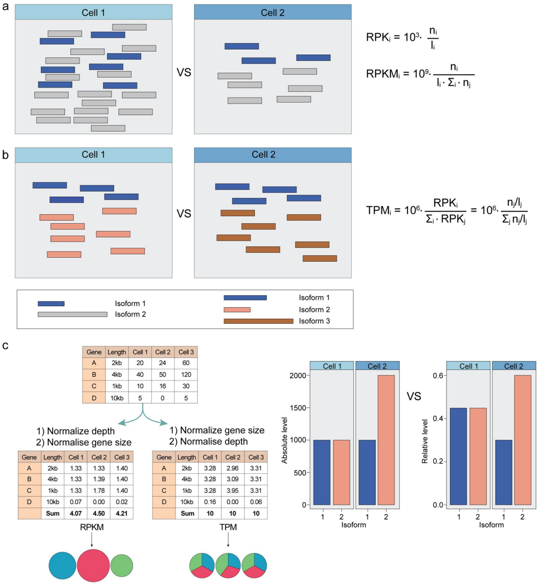
Single-Cell Transcriptomics: A High-Resolution Avenue for Plant Functional Genomics, Rich-Griffin et al., Trends in Plant Science, 2020

Cell Separation Methods



Hwang, Lee, and Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp. Mol. Med., 2018

Quantification of scRNA-seq Data



$$RPK_i = 10^3 \cdot \frac{n_i}{l_i}$$

$$RPKM_i = 10^9 \cdot \frac{n_i}{l_i \cdot \Sigma_i \cdot n_j}$$

$$TPM_i = 10^6 \cdot \frac{RPK_i}{\Sigma_i \cdot RPK_j} = 10^6 \cdot \frac{n_j/l_j}{\Sigma_j\, n_j/l_j}$$

Gene	Length	Cell 1	Cell 2	Cell 3
A	2kb	20	24	60
B	4kb	40	50	120
C	1kb	10	16	30
D	10kb	5	0	5

1) Normalize depth
2) Normalise gene size

Gene	Length	Cell 1	Cell 2	Cell 3
A	2kb	1.33	1.33	1.40
B	4kb	1.33	1.39	1.40
C	1kb	1.33	1.78	1.40
D	10kb	0.07	0.00	0.02
	Sum	**4.07**	**4.50**	**4.21**

RPKM

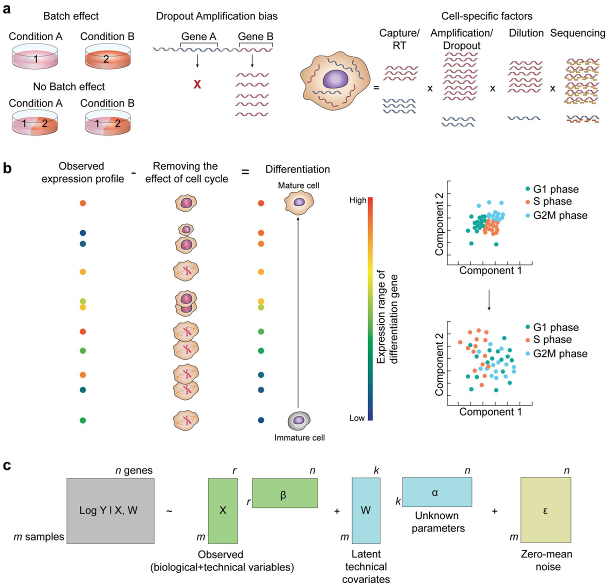
1) Normalize gene size
2) Normalise depth

Gene	Length	Cell 1	Cell 2	Cell 3
A	2kb	3.28	2.96	3.31
B	4kb	3.28	3.09	3.31
C	1kb	3.28	3.95	3.31
D	10kb	0.16	0.00	0.06
	Sum	**10**	**10**	**10**

TPM

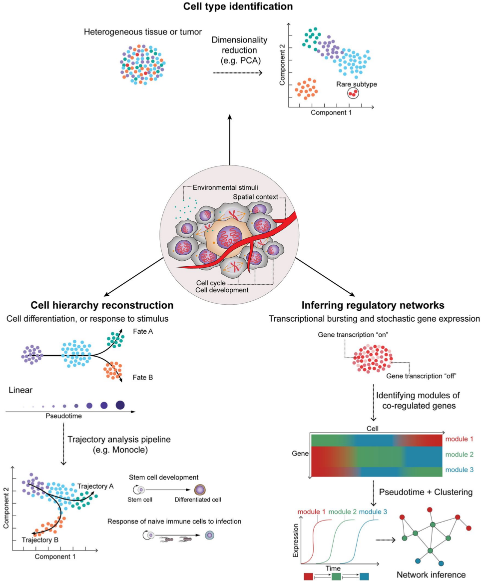
Hwang, Lee, and Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp. Mol. Med., 2018

Confounding Factors in scRNA-Seq Data



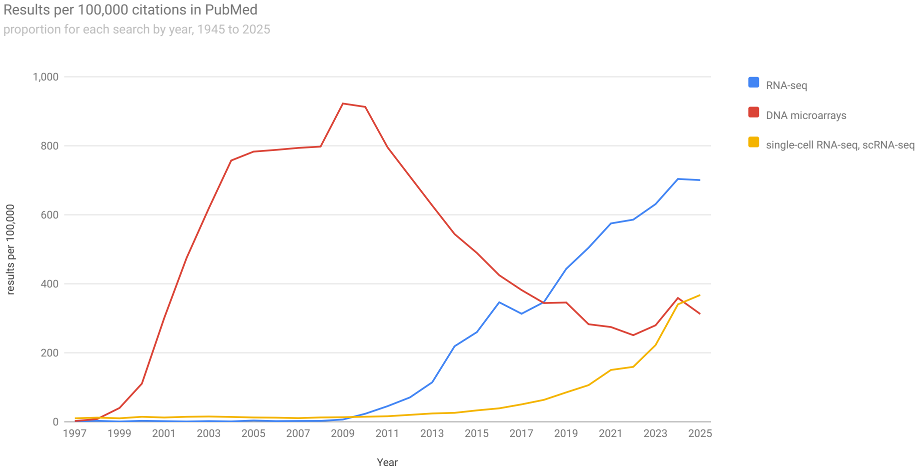
Hwang, Lee, and Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp. Mol. Med., 2018

Applications of scRNA-seq Data

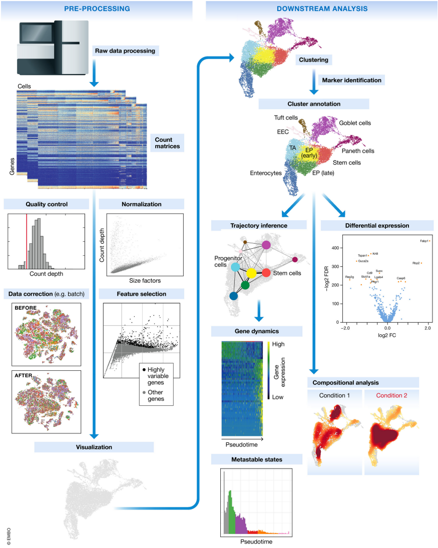


Hwang, Lee, and Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp. Mol. Med., 2018

Technology Trends

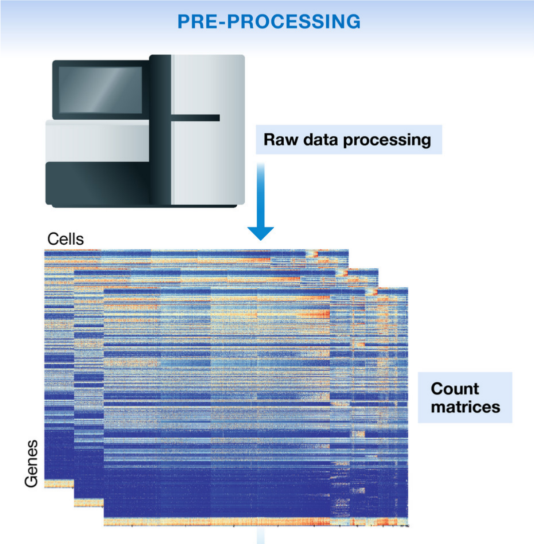


Typical scRNA-seq Analysis Workflow



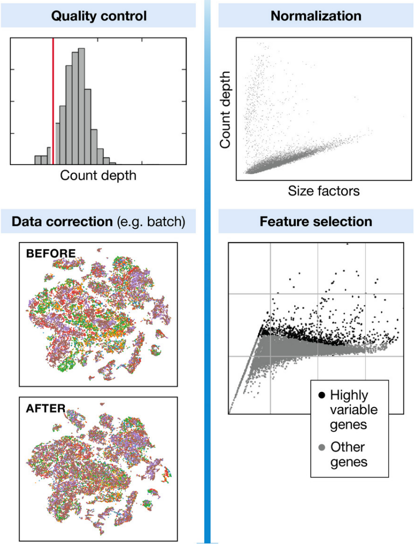
Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, `https://doi.org/10.15252/msb.20188746`

Typical scRNA-seq Analysis Workflow



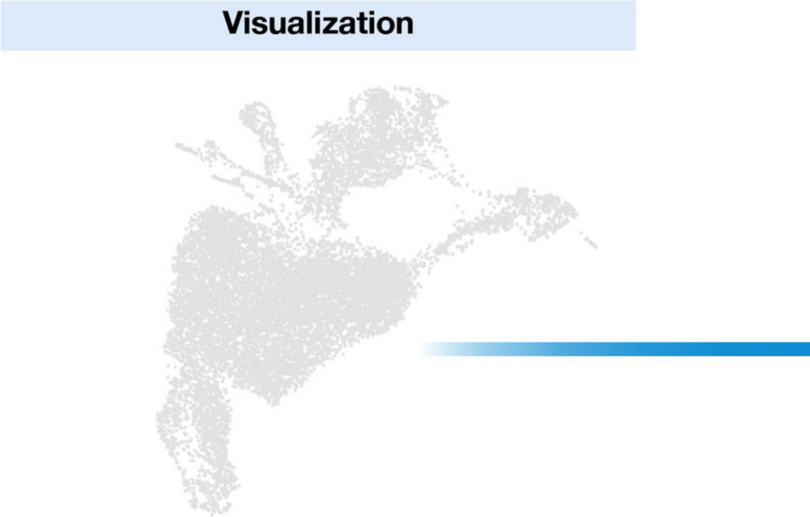
Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, `https://doi.org/10.15252/msb.20188746`

Typical scRNA-seq Analysis Workflow



Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, *`https://doi.org/10.15252/msb.20188746`*

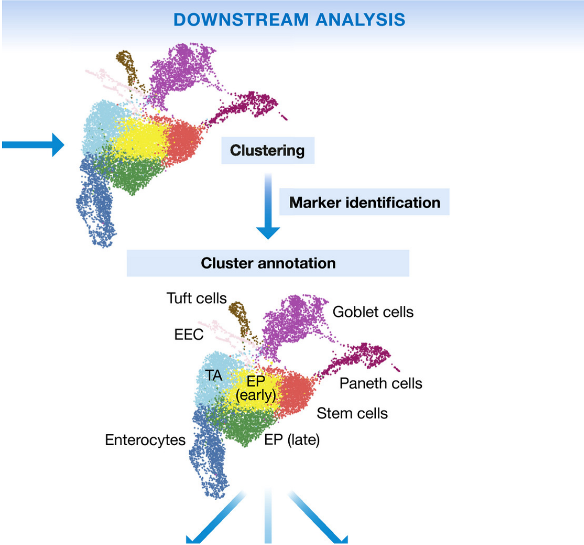
Typical scRNA-seq Analysis Workflow



© EMBO

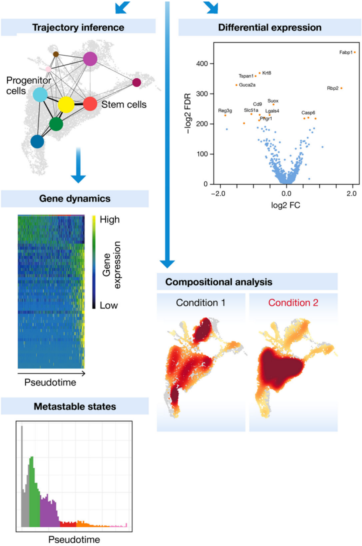
Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, *https://doi.org/10.15252/msb.20188746*

Typical scRNA-seq Analysis Workflow



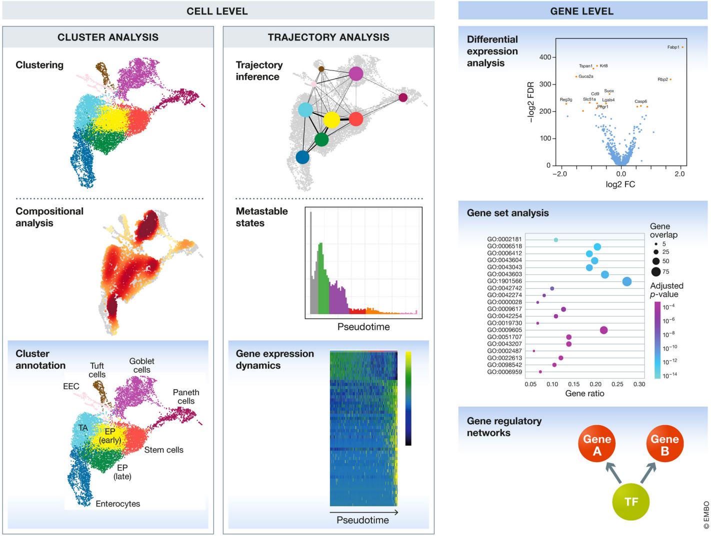
Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, `https://doi.org/10.15252/msb.20188746`

Typical scRNA-seq Analysis Workflow



Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, `https://doi.org/10.15252/msb.20188746`

Downstream Analysis Methods



Leucken and Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Sys. Bio., 2019, `https://doi.org/10.15252/msb.20188746`

Pseudotime Inference in scRNA-Data

- Difficult to experimentally purify cells in intermediate states.
- A population of cells isolated at same time population of cells may correspond to multiple, distinct intermediate differentiation states.

Trapnell et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, Nat. Biotech., 2014.

Pseudotime Inference in scRNA-Data

- Difficult to experimentally purify cells in intermediate states.
- A population of cells isolated at same time population of cells may correspond to multiple, distinct intermediate differentiation states.
- "One RNA-Seq experiment would constitute a time series, with each cell representing a distinct time point along a continuum."

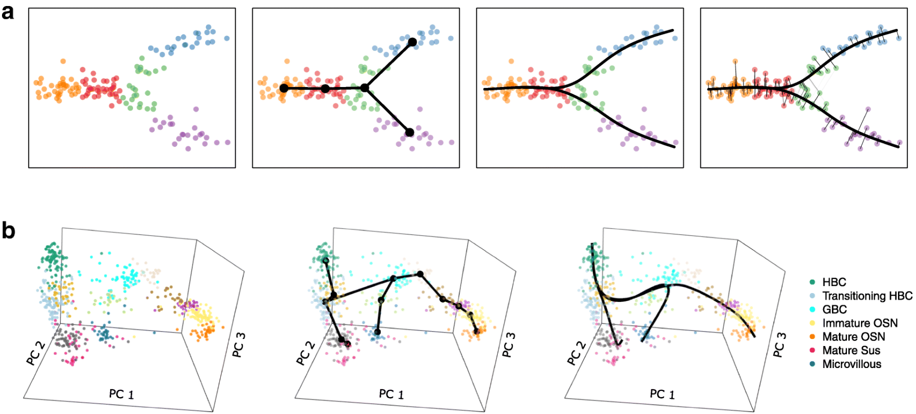
Trapnell et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, Nat. Biotech., 2014.

Pseudotime Inference in scRNA-Data

- Difficult to experimentally purify cells in intermediate states.
- A population of cells isolated at same time population of cells may correspond to multiple, distinct intermediate differentiation states.
- "One RNA-Seq experiment would constitute a time series, with each cell representing a distinct time point along a continuum."
- Pseudotime: "Measure of how much progress an individual cell has made through a process such as cell differentiation."

Trapnell et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, Nat. Biotech., 2014.

~~Pseudotime~~ Trajectory Inference



By Kelly Street, Davide Risso, Russel B. Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, Sandrine Dudoit - https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=75075393