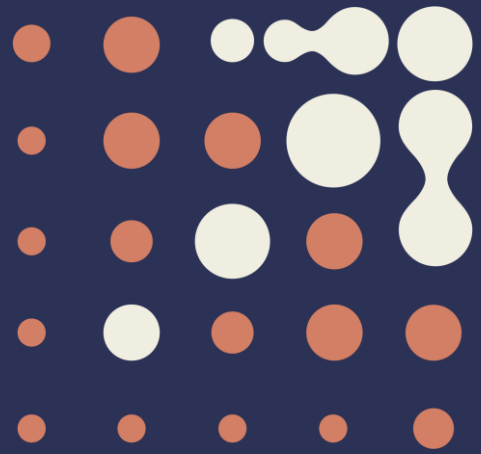
Medical Genomics:

Single-cell & Spatial Transcriptomics

N. Alcala & L. Mangé

Rare Cancers Genomics Team

November 26th 2024



International Agency for Research on Cancer

World Health Organization

RARE CANCERS GENOMICS

Plan

Part I. Transcriptomics

- Concepts
- Resources: databases

Part II. Single-cell transcriptomics

- Concepts
- Techniques
- Analysis: preprocessing & onwards

Part III. Spatial transcriptomics

- Concepts
- Techniques
- Analysis: independent analyses & existing pipelines

Part I. Transcriptomics | *Concepts*

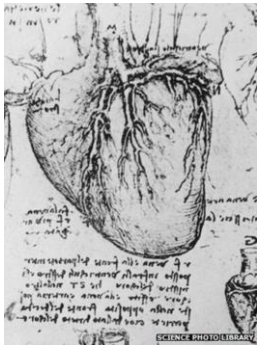
- Transcriptomics: study of the transcriptome -> transcripts made from genes expressed
- Helps us understand:
  - How gene expression influences tumor biology, progression, and treatment response
  - **Tumor tissue heterogeneity**

Part I. Transcriptomics | *Concepts*

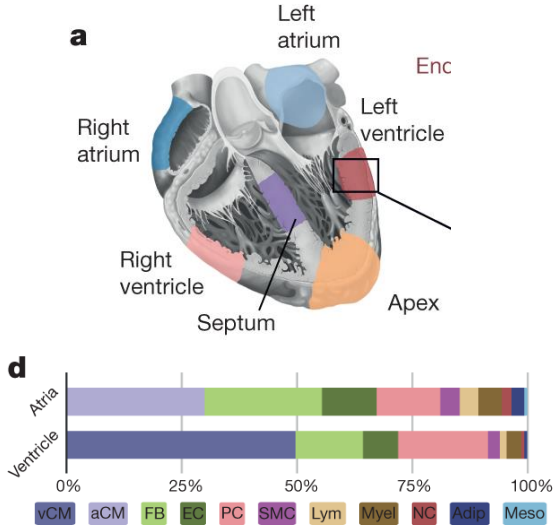
Tissue heterogeneity

- Tissues are made of **mixtures of cells**

- T**he investigation of tissue heterogeneity** gained novel traction with new sequencing technologies



Heart anatomy. Da Vinci circa 1510.



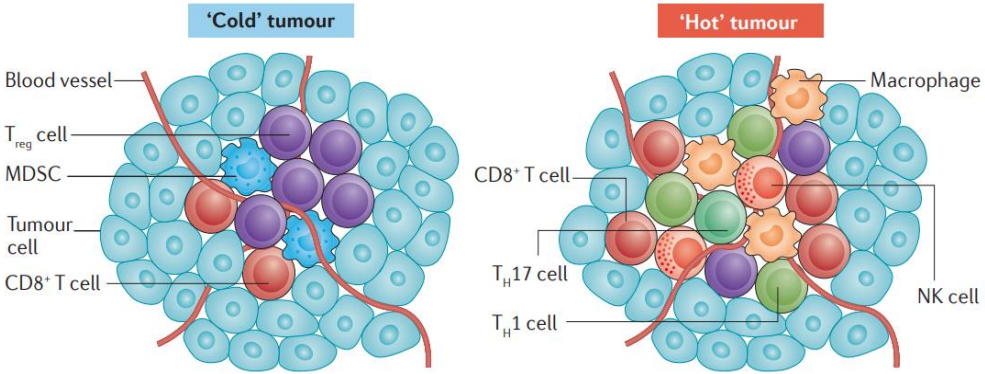
Human heart cell composition. Percentage of cell types estimated from single-cell RNA-seq. *Source: Litviňuková et al. Nature 2020.*

Part I. Transcriptomics | *Concepts*

Tissue heterogeneity: Tumor microenvironment (TME)

- Tumors have various amounts and compositions of **Tumor Infiltrating Lymphocytes (TILs)**

- **TILs influence disease progression**



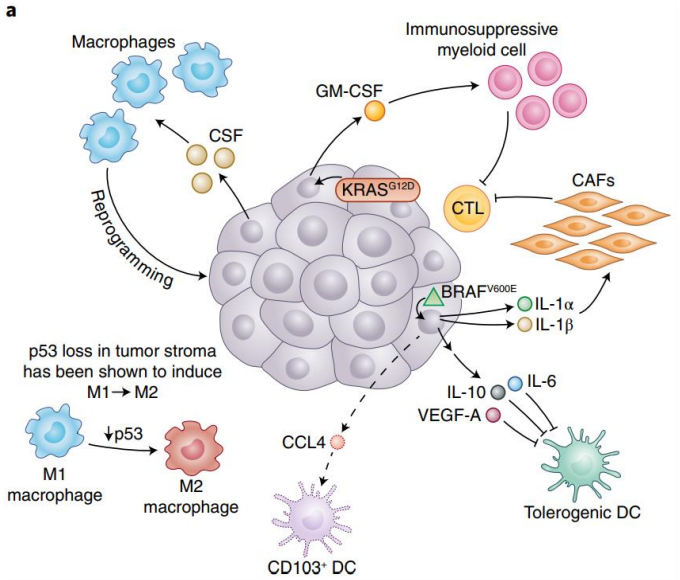
Tumors differ in their level of infiltration. *Source: Nagarsheth et al. Nat Rev Immun 2017.*

Part I. Transcriptomics | *Concepts*

Tumors shape their microenvironment

- **Tumors can establish protumoral and immunosuppressive environments**

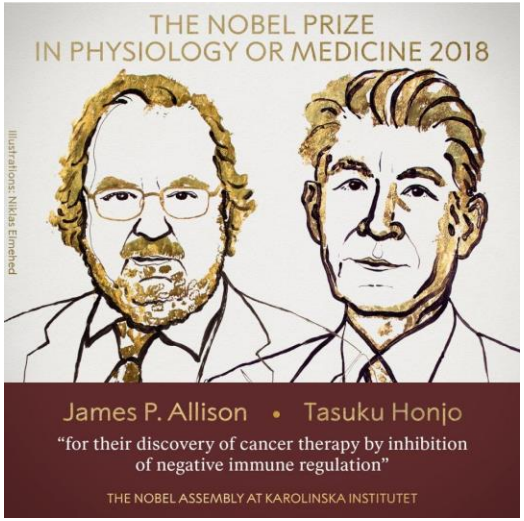
- They **recruit stromal and immune cells** to suppress the immune response (e.g., fibroblasts), promote metastasis (e.g. macrophages) by increasing angiogenesis (blood vessel formation providing nutrients to the tumor)

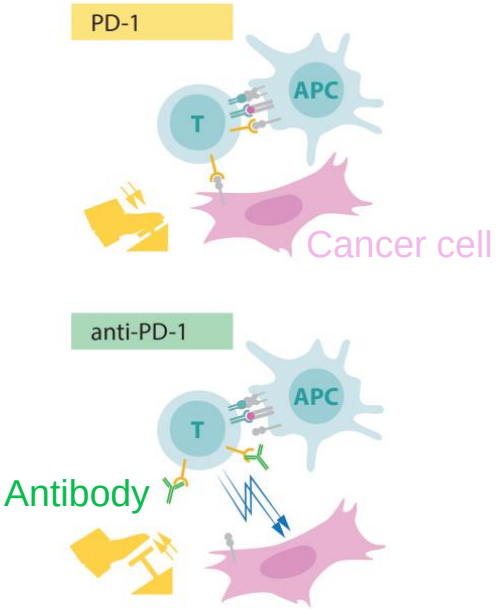


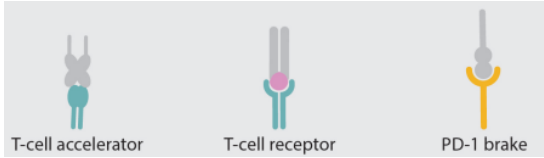
Tumors genotypes and phenotypes shape the TME. In melanoma *KRAS* somatic alterations promote the recruitment of immunosuppressive cells. *Source: Binnewies et al. Nature Medicine 2018.*

Part I. Transcriptomics | *Concepts*

The TME is associated with the tumor genome and response to therapy







"Removing the brakes" on the immune response.

Part I. Transcriptomics | *Concepts*

The TME is associated with the tumor genome and response to therapy

CLINICAL CANCER RESEARCH | CCR DRUG UPDATES

FDA Approval Summary: Pembrolizumab for the Treatment of Tumor Mutational Burden–High Solid Tumors

Leigh Marcus[1], Lola A. Fashoyin-Aje[1], Martha Donoghue[1], Mengdie Yuan[2], Lisa Rodriguez[2], Pamela S. Gallagher[3], Reena Philip[3], Soma Ghosh[3], Marc R. Theoret[4], Julia A. Beaver[4], Richard Pazdur[4], and Steven J. Lemery[1]

ABSTRACT

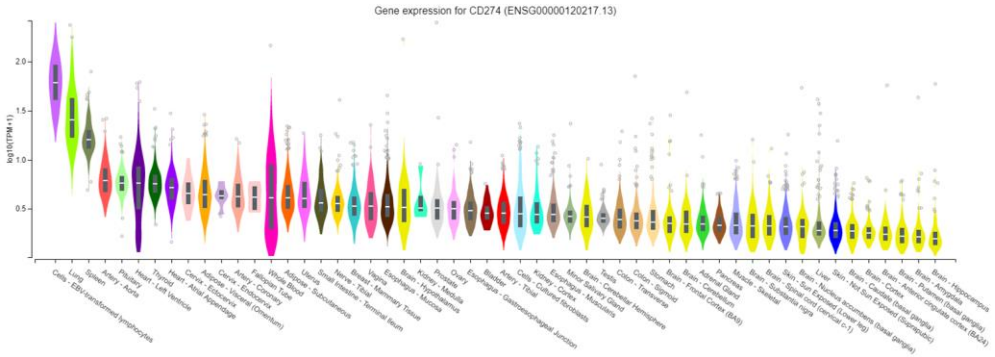
The FDA approved pembrolizumab on June 16, 2020, for the treatment of adult and pediatric patients with unresectable or metastatic tumor mutational burden–high [TMB-H; ≥10 mutations/megabase (mut/Mb)] solid tumors, as determined by an FDA-approved test, that have progressed following prior treatment and who have no satisfactory alternative treatment options. FDA granted the approval based on a clinically important overall response rate (29%; 95% confidence interval, 21–39) and duration of response (57% of responses lasting ≥ 12 months) in the subset of patients with TMB-H solid tumors ($n = 102$) spanning nine different tumor types enrolled in a multicenter single-arm trial (KEYNOTE-158). The efficacy of pembrolizumab was supported by the results of whole-exome sequencing (WES) analyses of TMB in additional patients enrolled across multiple pembrolizumab clinical trials, and a scientific understanding of the effects of PD-1 inhibition. Overall, the adverse event profile of pembrolizumab was similar to the adverse event profile observed in prior trials that supported the approval of pembrolizumab in other indications. This approval of pembrolizumab is the first time that the FDA has approved a cancer treatment for an indication based on TMB, and the fourth based on the presence of a biomarker rather than the primary site of origin.

Part I. Transcriptomics | *Resources*

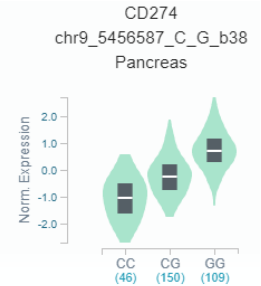
The Genotype-Tissue Expression (GTEx) project

Database of tissue-specific gene expression and regulation

- 54 non-diseased tissue sites for 1000 individuals with WGS/WES, and RNA-Seq
- gene expression, expression quantitative trait loci (eQTL), and histology images



Expression of immune checkpoint gene PD-L1 in 52 tissues.



Example Pancretic eQTL. Gene expression varies as a function of genotype at locus chr9 5456587

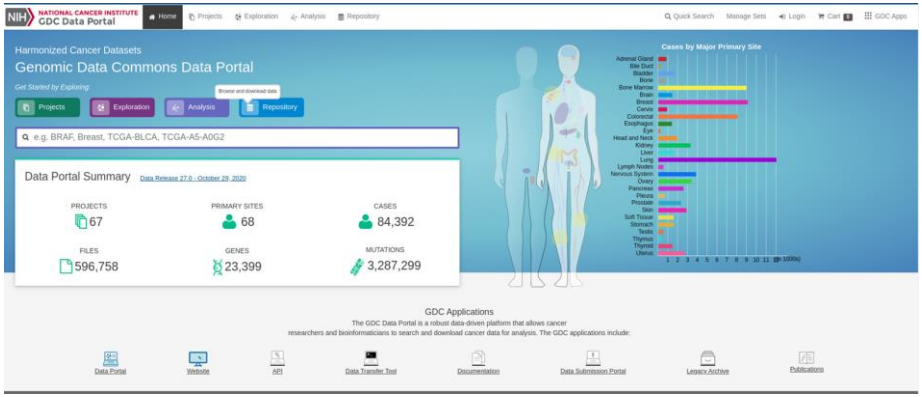
Part I. Transcriptomics | *Resources*

The Cancer Genome Atlas (TCGA) project

Database of cancer multi-omic data for

- Tumors from 33 primary sites

- Genomic, epigenomic, transcriptomic, and proteomic data
- RNA-seq data under controlled access (requires research institute affiliation)
- Processed gene expression data (read counts and FPKM) open-access



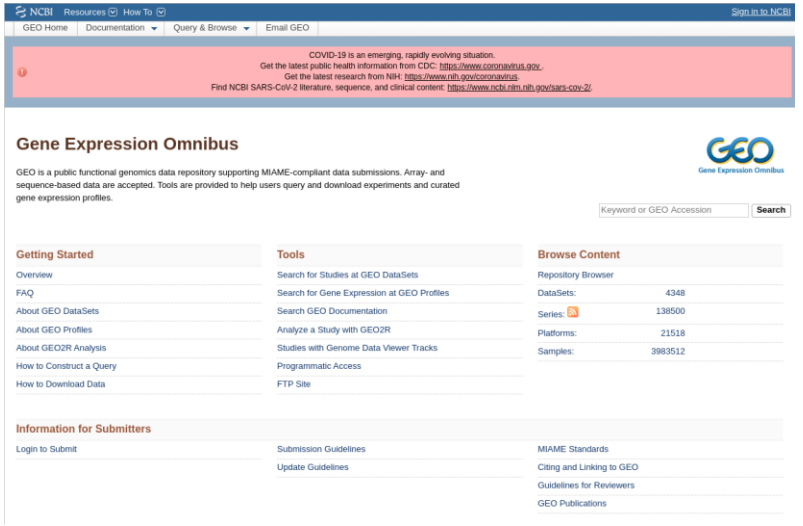
Web interface of the genomic data portal hosting the TCGA data. *Source:* https://portal.gdc.cancer.gov/.

Part I. Transcriptomics | *Resources*

The Gene Expression Omnibus (GEO) repository

Database of expression data (arrays and RNA-seq)

- Includes human data
- Not only cancer
- All data is open-access



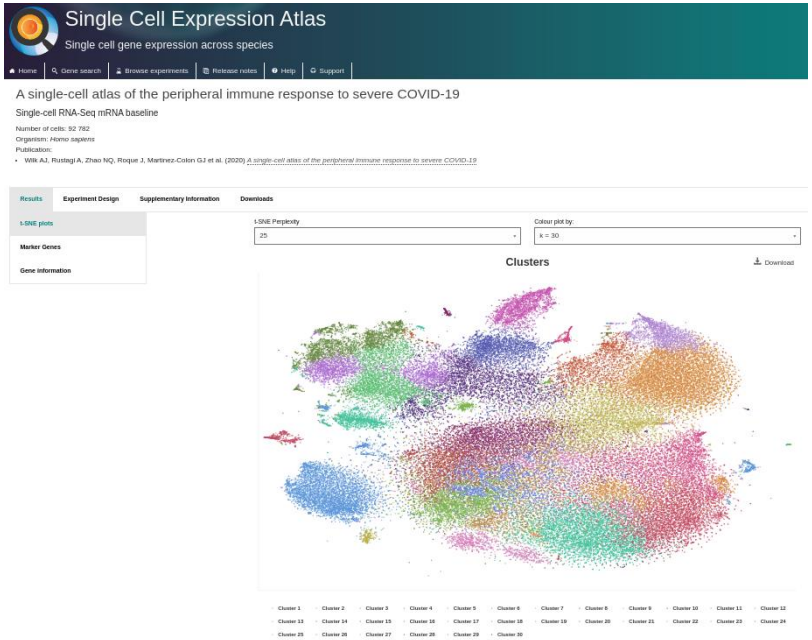
Web interface of the gene expression omnibus repository. *Source:* https://www.ncbi.nlm.nih.gov/geo/.

Part I. Transcriptomics | *Resources*

The Single Cell Expression Atlas

Database of scRNA-seq data

- Processed gene expression data (read counts) open-access



Web interface of the single-cell expression atlas. scRNA-seq of immune response to severe COVID-19 (t-SNE). *Source:* https://www.ebi.ac.uk/gxa/sc/home.

Part II. Single-cell transcriptomics | *Concepts*

Part II. Single-cell transcriptomics | *Concepts*

What: Quantify the level of expression of genes and transcripts of each individual cell of a tissue

FOCUS | EDITORIAL

Method of the Year 2019: Single-cell multimodal omics

Multimodal omics measurement offers opportunities for gaining holistic views of cells one by one.

International Agency for Research on Cancer

World Health Organization

Part II. Single-cell transcriptomics | *Concepts*

What: Quantify the level of expression of genes and transcripts of each individual cell of a tissue

Why:
- Track cell differentiation
- **Quantify tissue heterogeneity**
- Quantify diversity of microbiome…

FOCUS | EDITORIAL

Method of the Year 2019: Single-cell multimodal omics

Multimodal omics measurement offers opportunities for gaining holistic views of cells one by one.

International Agency for Research on Cancer

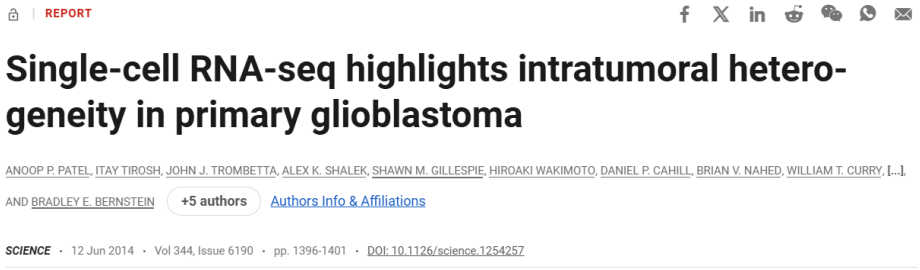
World Health Organization

Part II. Single-cell transcriptomics | *Concepts*

Example:

- GBM is an aggressive brain cancer
- Tumors show a lot of heterogeneity,

but at the time it was unclear if all different cellular profiles coexisted in individual tumors



REPORT

Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

ANOOP P. PATEL, ITAY TIROSH, JOHN J. TROMBETTA, ALEX K. SHALEK, SHAWN M. GILLESPIE, HIROAKI WAKIMOTO, DANIEL P. CAHILL, BRIAN V. NAHED, WILLIAM T. CURRY, [...], AND BRADLEY E. BERNSTEIN +5 authors Authors Info & Affiliations

SCIENCE · 12 Jun 2014 · Vol 344, Issue 6190 · pp. 1396-1401 · DOI: 10.1126/science.1254257

- Authors used scRNAseq:
  - Concluded that individual tumors are a mosaic of the cellular profiles, which explains their resistance to therapy
  - Identified a spectrum of cellular states & which ones could be potential therapeutic targets

Part II. Single-cell transcriptomics | *Concepts*

What: Quantify the level of expression of genes and transcripts of each individual cell of a tissue

Why:
- Track cell differentiation
- Quantify tissue heterogeneity
- Quantify diversity of microbiome…

How:
- **Droplet based (10X genomics) -> most used technique**
- Plate-based with unique molecular identifiers (UMIs): CEL-seq, MARS-seq
- Plate-based with reads: Smart-seq2

Method of the Year 2019: Single-cell multimodal omics

Multimodal omics measurement offers opportunities for gaining holistic views of cells one by one.

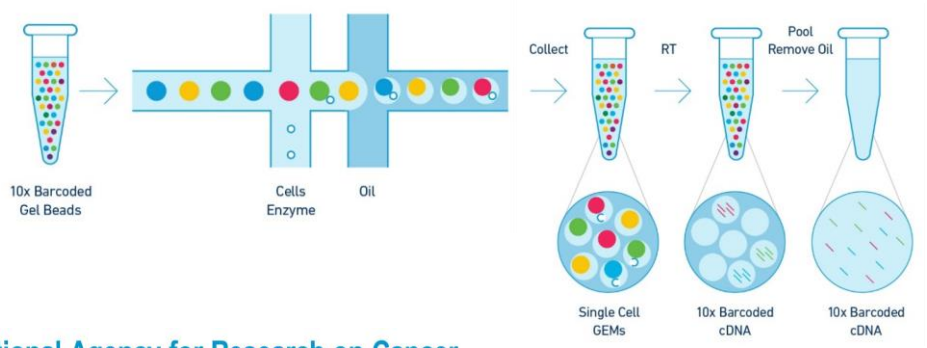
Part II. Single-cell transcriptomics | *Techniques*

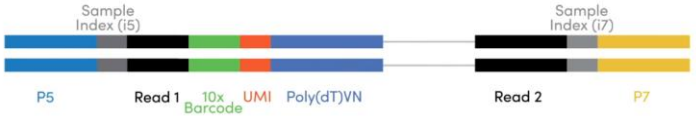
Droplet based (10X genomics)

Part II. Single-cell transcriptomics | *Techniques*

Droplet based (10X genomics)

- Barcoded Gel Beads are attached to each cell to form Gel Bead in EMulsion (GEMs) of nL size



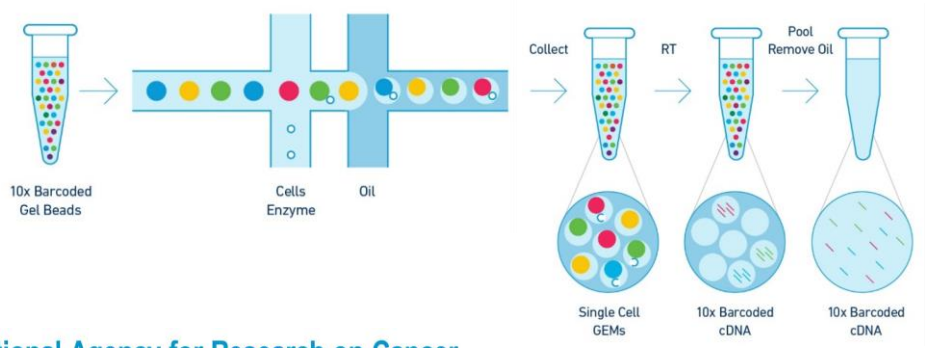


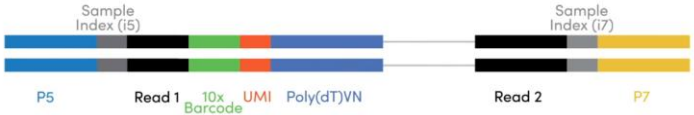
Source: 10X genomics.

Part II. Single-cell transcriptomics | *Techniques*

Droplet based (10X genomics)

- Barcoded Gel Beads are attached to each cell to form Gel Bead in EMulsion (GEMs) of nL size
- Reverse transcription (to obtain cDNA) is performed in each GEM, attaching identifiers to each transcript & cell



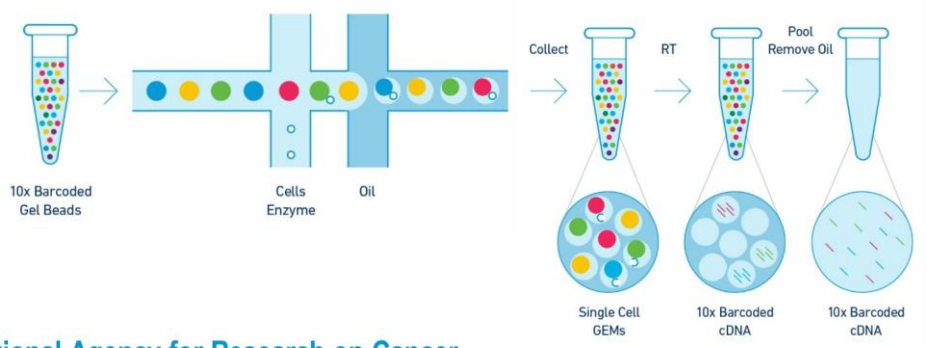


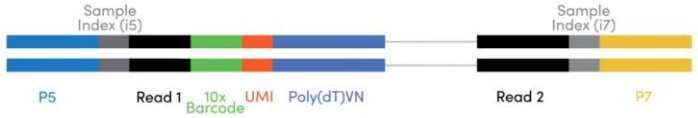
Source: 10X genomics.

Part II. Single-cell transcriptomics | *Techniques*

Droplet based (10X genomics)

- Barcoded Gel Beads are attached to each cell to form Gel Bead in EMulsion (GEMs) of nL size
- Reverse transcription (to obtain cDNA) is performed in each GEM, attaching identifiers to each transcript & cell
- Amplification, library generation, sequencing are performed as in classical RNA-seq experiment



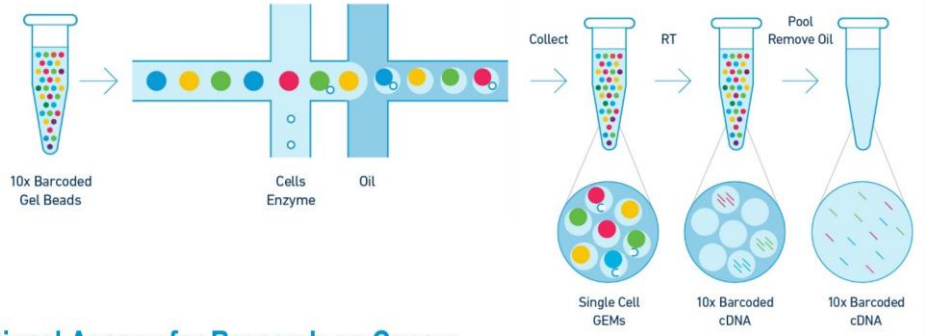


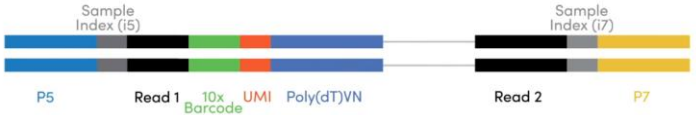
Source: 10X genomics.

Part II. Single-cell transcriptomics | *Techniques*

Droplet based (10X genomics)

- Barcoded Gel Beads are attached to each cell to form Gel Bead in EMulsion (GEMs) of nL size
- Reverse transcription (to obtain cDNA) is performed in each GEM, attaching identifiers to each transcript & cell
- Amplification, library generation, sequencing are performed as in classical RNA-seq experiment
- Each read is paired with a barcode read with cell identifier + Unique Molecular Identifier (UMI)





Source: 10X genomics.

Part II. Single-cell transcriptomics | *Techniques*

Processing:

Part II. Single-cell transcriptomics | *Techniques*

Processing:

- barcode-aware alignment (e.g., **CellRanger**, STARsolo)
  - error-correction and demultiplexing of cell barcodes
  - standard mapping on reference genome
  - deduplication of UMIs

Part II. Single-cell transcriptomics | *Techniques*

Processing:

- barcode-aware alignment (e.g., **CellRanger**, STARsolo)
  - error-correction and demultiplexing of cell barcodes
  - standard mapping on reference genome
  - deduplication of UMIs

- Final output: **Gene expression matrix**
  - Cells x genes/transcripts
  - Values: number of reads of a gene in a given cell
  - Other output files: FASTQ files with the raw sequencing reads, barcodes to link reads to individual droplets, UMIs to eliminate PCR duplicates

Part II. Single-cell transcriptomics | *Techniques*

Depth-cell number trade-off:

- Depth = number of reads per cell

Part II. Single-cell transcriptomics | *Techniques*

Depth-cell number trade-off:

- Depth = number of reads per cell
- Higher depth -> more information about transcriptome, lowly expressed genes, subtle differences in expression between cell types

Part II. Single-cell transcriptomics | *Techniques*

Depth-cell number trade-off:

- Depth = number of reads per cell
- Higher depth -> more information about transcriptome, lowly expressed genes, subtle differences in expression between cell types
- Current technologies can sequence 100 to 100,000 cells, with 1,000 to 100,000 reads/cell

Part II. Single-cell transcriptomics | *Techniques*

Depth-cell number trade-off:

- Depth = number of reads per cell
- Higher depth -> more information about transcriptome, lowly expressed genes, subtle differences in expression between cell types
- Current technologies can sequence 100 to 100,000 cells, with 1,000 to 100,000 reads/cell
- **More cells** help **identify rare cell subtypes**

Part II. Single-cell transcriptomics | *Techniques*

Depth-cell number trade-off:

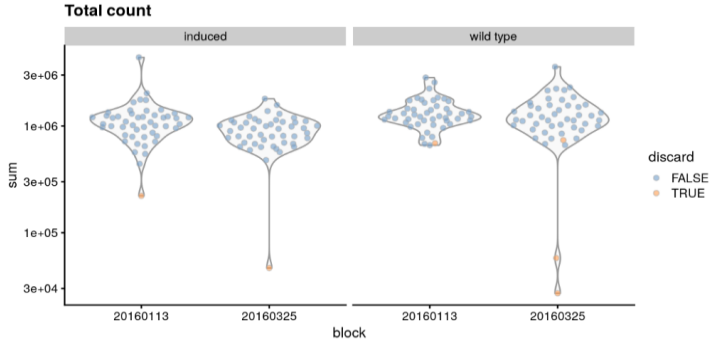
- Depth = number of reads per cell
- Higher depth -> more information about transcriptome, lowly expressed genes, subtle differences in expression between cell types
- Current technologies can sequence 100 to 100,000 cells, with 1,000 to 100,000 reads/cell
- **More cells** help **identify rare cell subtypes**
- Trade-off arises because of **limited sequencing capacity**

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

Preprocessing & visualization:

1. **QC:** Remove low-quality cells (damaged or badly captured), e.g., based on low total counts/cell, proportion of mitochondrial reads and number of non-zero features, correct ambient RNA (*SoupX* R package), detect doublets (*scDblFinder* R package)



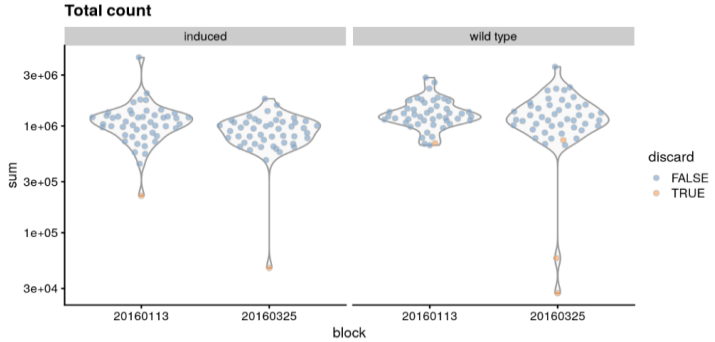
Example QC of scRNA-seq (total count/cell). *Source:* https://osca.bioconductor.org/quality-control.html

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

Preprocessing & visualization:

1. **QC:** Remove low-quality cells (damaged or badly captured), e.g., based on low total counts/cell, proportion of mitochondrial reads and number of non-zero features, correct ambient RNA (*SoupX* R package), detect doublets (*scDblFinder* R package)
2. **Normalization:** counts are normalized for library size differences and transformed to reduce variance



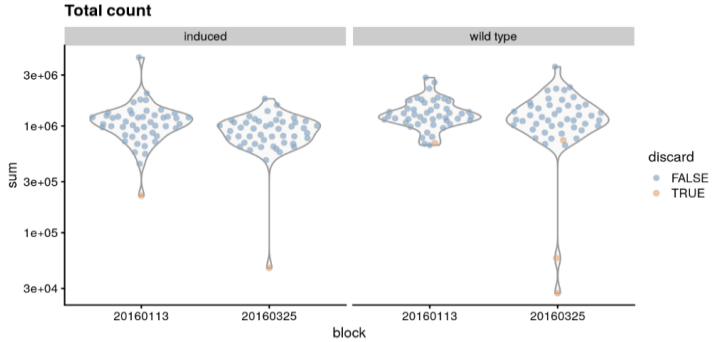
Example QC of scRNA-seq (total count/cell). *Source:* https://osca.bioconductor.org/quality-control.html

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

Preprocessing & visualization:

1. **QC:** Remove low-quality cells (damaged or badly captured), e.g., based on low total counts/cell, proportion of mitochondrial reads and number of non-zero features, correct ambient RNA (*SoupX* R package), detect doublets (*scDblFinder* R package)
2. **Normalization:** counts are normalized for library size differences and transformed to reduce variance
3. **Feature selection**: retaining genes that are highly variable



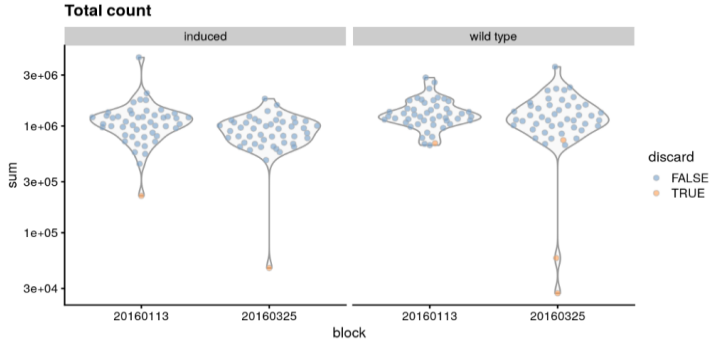
Example QC of scRNA-seq (total count/cell). *Source:* https://osca.bioconductor.org/quality-control.html

Part II. Single-cell transcriptomics | *Analysis*

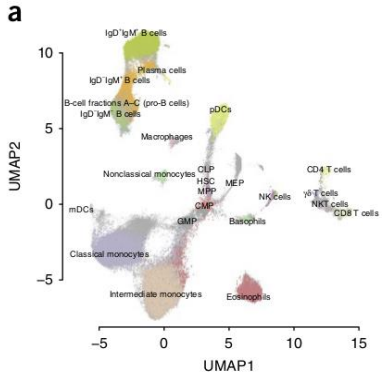
Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

Preprocessing & visualization:

1. **QC:** Remove low-quality cells (damaged or badly captured), e.g., based on low total counts/cell, proportion of mitochondrial reads and number of non-zero features, correct ambient RNA (*SoupX* R package), detect doublets (*scDblFinder* R package)
2. **Normalization:** counts are normalized for library size differences and transformed to reduce variance
3. **Feature selection**: retaining genes that are highly variable
4. **Dimensionality reduction:** compact the data and reduce noise (PCA, or non-linear techniques like UMAP)



Example QC of scRNA-seq (total count/cell). *Source:* https://osca.bioconductor.org/quality-control.html

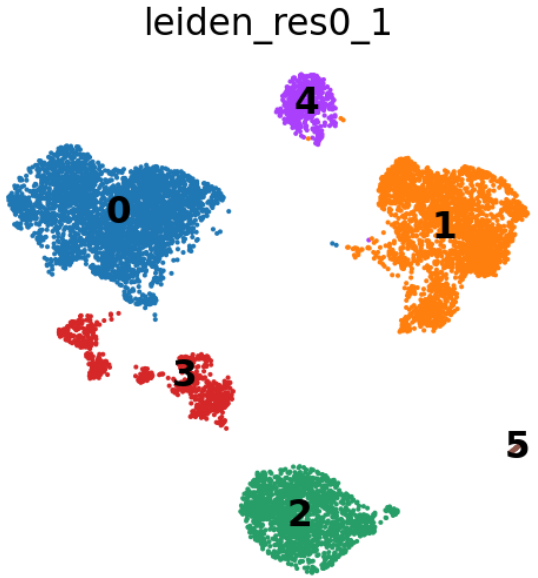


2D embedding of scRNA-seq of immune cell populations. a. UMAP. *Source:* Becht et al. Nature Biotechnology 2019.

Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

CAREFUL with distance interpretation

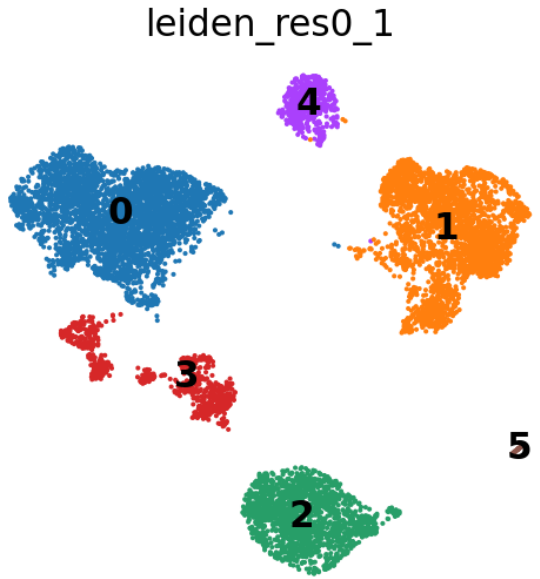


Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

CAREFUL with distance interpretation

Cluster 5 actually closer to 4 than to 1



Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

CAREFUL with distance interpretation

Not only distance interpretation, some say the whole concept is problematic and shouldn't be used

Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

CAREFUL with distance interpretation

Not only distance interpretation, some say the whole concept is problematic and shouldn't be used

OPEN ACCESS

PERSPECTIVE

The specious art of single-cell genomics

Tara Chari, Lior Pachter

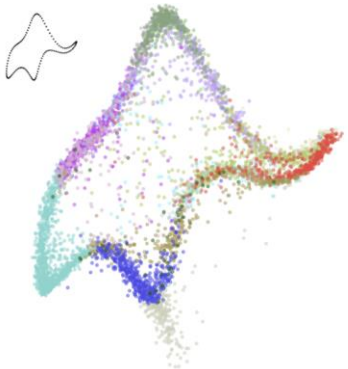
Published: August 17, 2023 • https://doi.org/10.1371/journal.pcbi.1011288

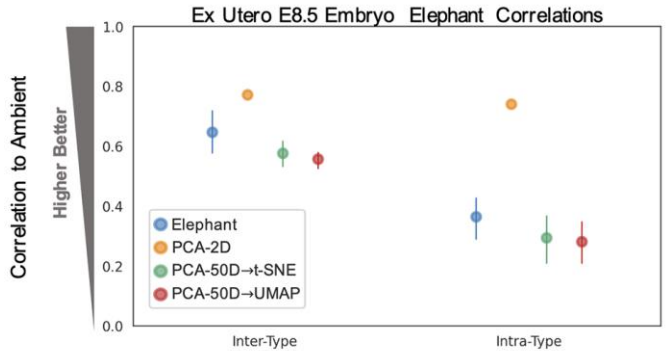
Authors argue that there is little theoretical support for 2D UMAP visualization, that it has little utility and propose other tools instead (ex dendrograms, graph-based diagrams…)

Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

CAREFUL with distance interpretation





Authors show that representing 2D clusters as a “von Neumann elephant” can be as good as/better than the representation given by UMAP (when focusing on inter- and intra-cluster distances)

OPEN ACCESS

PERSPECTIVE

The specious art of single-cell genomics

Tara Chari, Lior Pachter

Published: August 17, 2023 • https://doi.org/10.1371/journal.pcbi.1011288

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

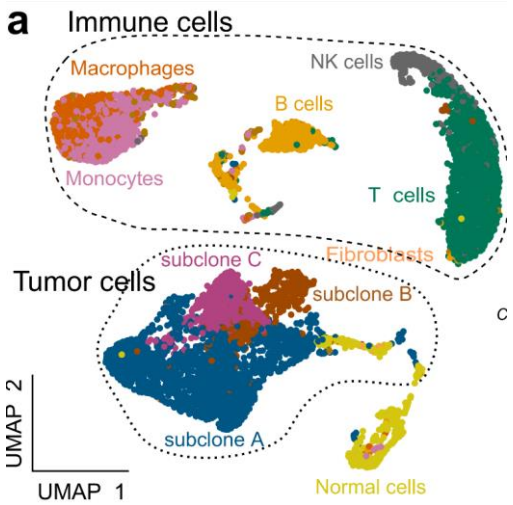
1. **Cell clustering**: group similar expression profiles (biological states)

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1. **Cell clustering**: group similar expression profiles (biological states)
2. **Annotation**: assign cell types to each barcode



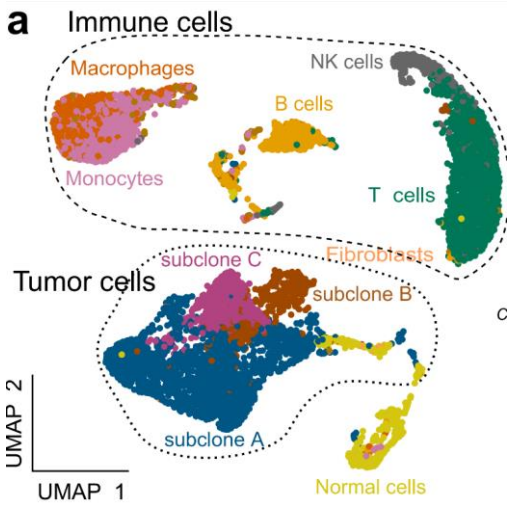
2D embedding of scRNA-seq of mesothelioma. *Source: MESOMICS project, N. Alcala WCR grant application*.

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1. **Cell clustering**: group similar expression profiles (biological states)
2. **Annotation**: assign cell types to each barcode
3. **Copy number variant (CNV) calling**: detect chromosome amplifications & deletions -> identify tumor cells



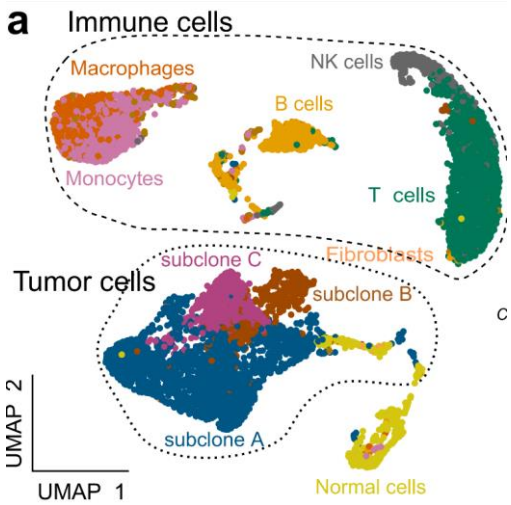
2D embedding of scRNA-seq of mesothelioma. *Source: MESOMICS project, N. Alcala WCR grant application*.

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1. **Cell clustering**: group similar expression profiles (biological states)
2. **Annotation**: assign cell types to each barcode
3. **Copy number variant (CNV) calling**: detect chromosome amplifications & deletions -> identify tumor cells
4. **Differential expression**: similar to annotation, annotate clusters by inferring genes that are statistically significantly over- or underexpressed between groups, uncover genesets which potentially explain observed phenotypes



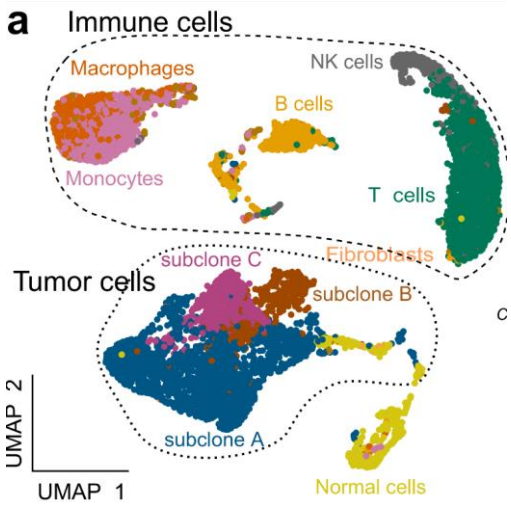
2D embedding of scRNA-seq of mesothelioma. *Source: MESOMICS project, N. Alcala WCR grant application*.

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1. **Cell clustering**: group similar expression profiles (biological states)
2. **Annotation**: assign cell types to each barcode
3. **Copy number variant (CNV) calling**: detect chromosome amplifications & deletions -> identify tumor cells
4. **Differential expression**: similar to annotation, annotate clusters by inferring genes that are statistically significantly over- or underexpressed between groups, uncover genesets which potentially explain observed phenotypes
5. **Cell-cell communication**: understand intercellular signaling pathways (ex ligand-receptor relationships allow tumor cells to escape the immune response) by estimating protein abundance using gene expression information and uncovers which pairs of cell types produce complementary ligand-receptor molecules



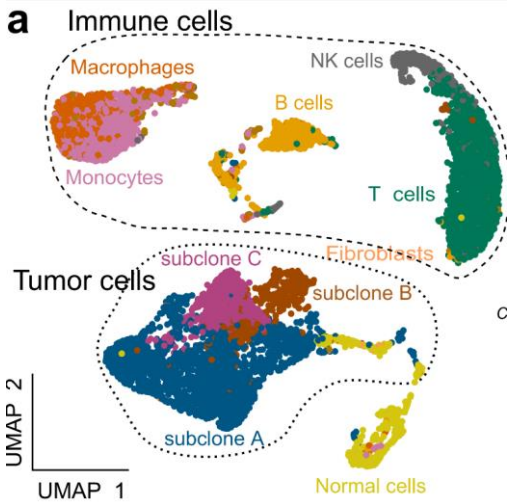
2D embedding of scRNA-seq of mesothelioma. *Source: MESOMICS project, N. Alcala WCR grant application*.

Part II. Single-cell transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing: Used in the GBM study

1. **Cell clustering**: group similar expression profiles (biological states)
2. **Annotation**: assign cell types to each barcode
3. **Copy number variant (CNV) calling**: detect chromosome amplifications & deletions -> identify tumor cells
4. **Differential expression**: similar to annotation, annotate clusters by inferring genes that are statistically significantly over- or underexpressed between groups, uncover genesets which potentially explain observed phenotypes
5. **Cell-cell communication**: understand intercellular signaling pathways (ex ligand-receptor relationships allow tumor cells to escape the immune response) by estimating protein abundance using gene expression information and uncovers which pairs of cell types produce complementary ligand-receptor molecules



2D embedding of scRNA-seq of mesothelioma. *Source: MESOMICS project, N. Alcala WCR grant application*.

Part III. Spatial transcriptomics | *Concepts*

What: similar to single-cell but while keeping the spatial context in the tissue -> allows to map RNA expression patterns within tissue

Part III. Spatial transcriptomics | *Concepts*

What: similar to single-cell but while keeping the spatial context in the tissue -> allows to map RNA expression patterns within tissue

Why: same reasons as single-cell but also:

- Study tumor microenvironment by looking at interactions between tumor, immune & stromal cells
- Identify therapeutic targets via accessibility…

Part III. Spatial transcriptomics | *Concepts*

What: similar to single-cell but while keeping the spatial context in the tissue -> allows to map RNA expression patterns within tissue

Why: same reasons as single-cell but also:

- Study tumor microenvironment by looking at interactions between tumor, immune & stromal cells
- Identify therapeutic targets via accessibility…

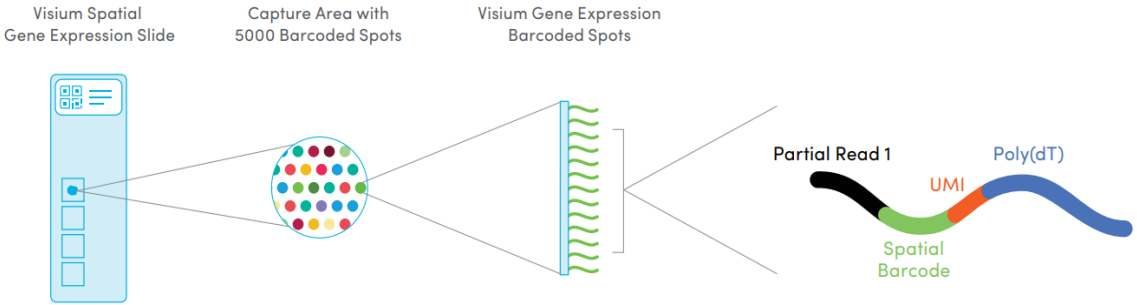
How: Good for imaging & storage Good for precise transcriptomics

- Using **FFPE** (Formalin-Fixed Paraffin-Embedded) or **fresh frozen tissue**
- **10x Genomics Visium & Visium HD -> one of the most popular, easy & cheap to use, high-resolution**
- Slide-seq & Slide-seqV2: higher spatial resolution than Visium but harder to use
- MERFISH (Multiplexed Error-Robust Fluorescence In Situ Hybridization): single-molecule resolution, allows spatially precise quantification of targeted transcripts
- SeqFISH+: single-cell resolution, suitable for imaging specific gene panels

Part III. Spatial transcriptomics | *Techniques*

10x Genomics Visium

- Fresh-frozen tissue section placed on array with capture probes that bind to RNA
- cDNA is synthesized from captured RNA and sequencing libraries prepared
- Libraries are sequenced



Spatial composition of the Visium Spatial Gene Expression slide. Each slide contains four Capture Areas with approximately 5000 barcoded spots, which in turn contain millions of spatially-barcoded capture oligonucleotides. Tissue mRNA is released and binds to the barcoded oligos, enabling capture of gene expression information. *Source: 10X genomics*.

Part III. Spatial transcriptomics | *Techniques*

Visium resolution & output:

Part III. Spatial transcriptomics | *Techniques*

Visium resolution & output:

- Spot ≈ 55 microns -> up to 10 cells per spot
- Sample ≈ 5000 spots
- Visium HD: single-cell or subcellular level

Part III. Spatial transcriptomics | *Techniques*

Visium resolution & output:

- Spot ≈ 55 microns -> up to 10 cells per spot
- Sample ≈ 5000 spots
- Visium HD: single-cell or subcellular level

- Final output: **Gene expression matrix**
  - Spot x gene matrix
  - Spatial coordinates of each spot
  - Tissue histological image

Part III. Spatial transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

Preprocessing:

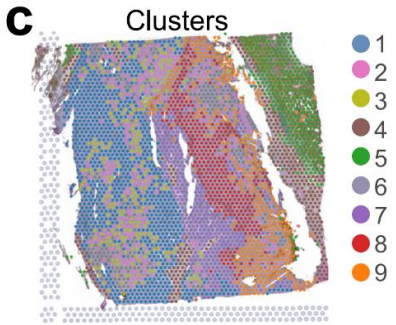
1. **QC:** Remove low-quality *spots*
2. **Normalization**
3. **Feature selection**
4. **Dimensionality reduction**

Part III. Spatial transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1.Spot clustering



Spatial RNA-seq of a lung neuroendocrine tumor. c. molecular clusters. *Source: lungNENomics project, N. Alcala WCR grant application*.

Part III. Spatial transcriptomics | *Analysis*

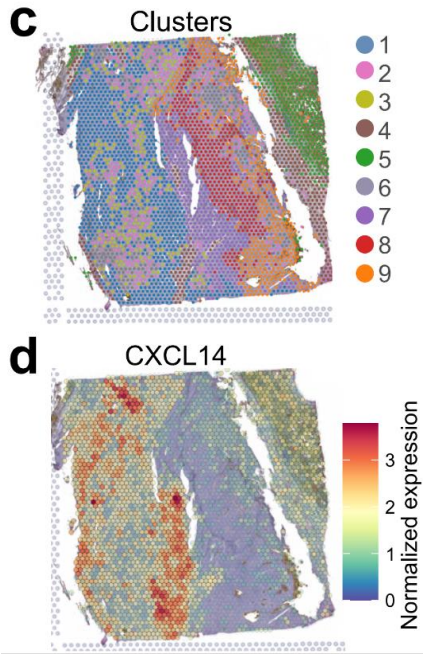
Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

After preprocessing:

1.Spot clustering

2.Spatially variable markers: identifying marker genes that differ across space (HVGs) caused by differences in cell-type composition or cell-cell communication events

Example: marked point process model, where each spot is a point in space associated with a value (gene expression).
-> test if pairs of spots at a distance of *r* have more dissimilar marks than expected by chance



Spatial RNA-seq of a lung neuroendocrine tumor. c. molecular clusters. *Source: lungNENomics project, N. Alcala WCR grant application*.

Part III. Spatial transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

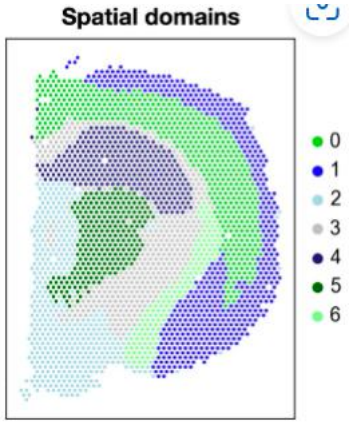
1. **Neighbourhood analysis**: study enrichment of cell types in specific regions of the tissue (local analysis)

Part III. Spatial transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

1. **Neighbourhood analysis**: study enrichment of cell types in specific regions of the tissue (local analysis)

2. **Spatial domains**: identify spatial patterns in tissue: define larger functional or anatomical regions



Source: www.sc-best-practices.org

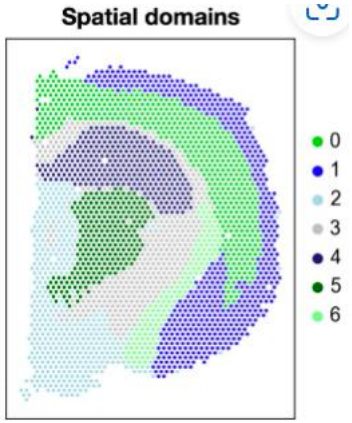
Part III. Spatial transcriptomics | *Analysis*

Single-cell best practices guide (Nat Rev Genet; 2023) -> different methods, python code

1. **Neighbourhood analysis**: study enrichment of cell types in specific regions of the tissue (local analysis)

2. **Spatial domains**: identify spatial patterns in tissue: define larger functional or anatomical regions

3. **Spot deconvolution**: go from spots to single-cells (ex assign proportions of cell types to each spatial location using a single-cell reference)



Source: www.sc-best-practices.org

Part III. Spatial transcriptomics | *Analysis*

Existing pipelines

Banksy

nature genetics

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature genetics › articles › article

Article | Open access | Published: 27 February 2024

BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis

Vipul Singhal, Nigel Chou, Joseph Lee, Yifei Yue, Jinyue Liu, Wan Kee Chock, Li Lin, Yun-Ching Chang, Erica Mei Ling Teo, Jonathan Aow, Hwee Kuan Lee, Kok Hao Chen ✉ & Shyam Prabhakar ✉

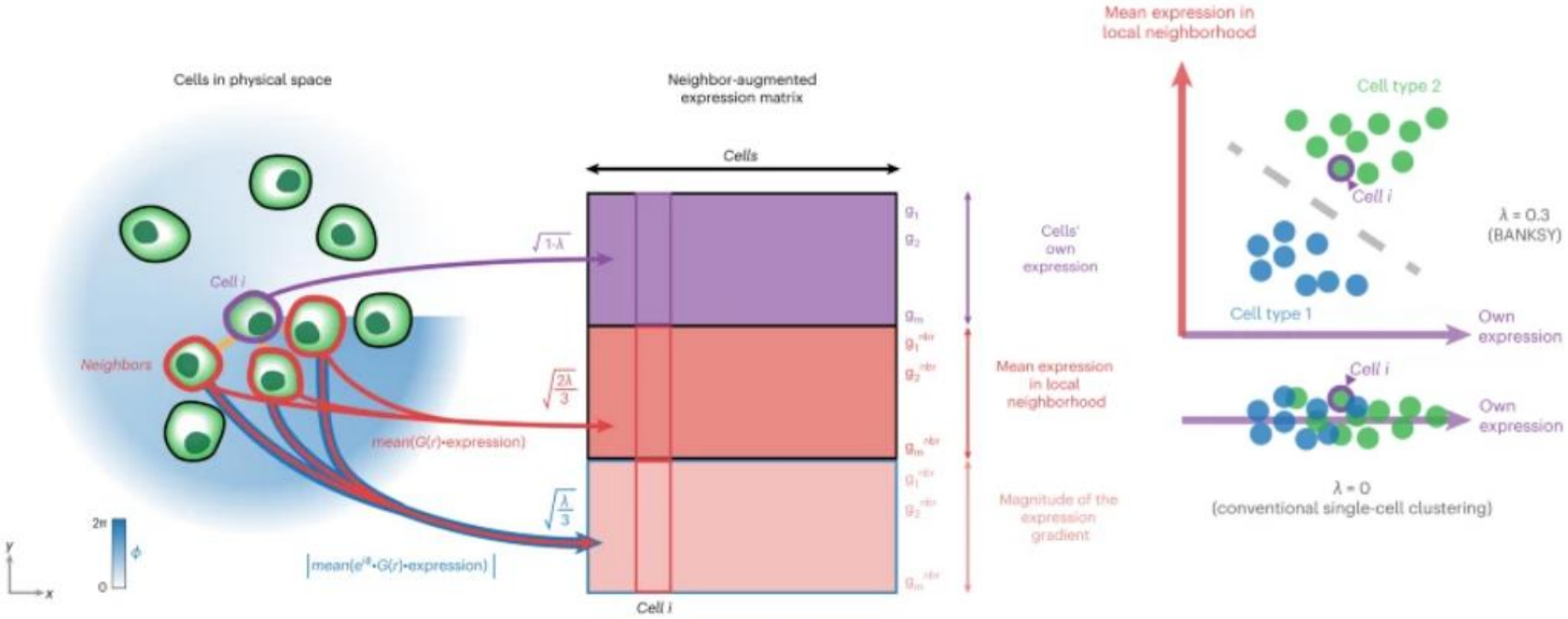
Nature Genetics **56**, 431–441 (2024) | Cite this article

31k Accesses | **16** Citations | **91** Altmetric | Metrics

Part III. Spatial transcriptomics | *Analysis*

Existing pipelines

Banksy



Source: www.nature.com/articles/s41588-024-01664-3#Sec12

Part III. Spatial transcriptomics | *Analysis*

Existing pipelines

iStar

nature biotechnology

Explore content | About the journal | Publish with us

nature > nature biotechnology > brief communications > article

Brief Communication | Published: 02 January 2024

Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology

Daiwei Zhang, Amelia Schroeder, Hanying Yan, Haochen Yang, Jian Hu, Michelle Y. Y. Lee, Kyung S. Cho, Katalin Susztak, George X. Xu, Michael D. Feldman, Edward B. Lee, Emma E. Furth, Linghua Wang & Mingyao Li

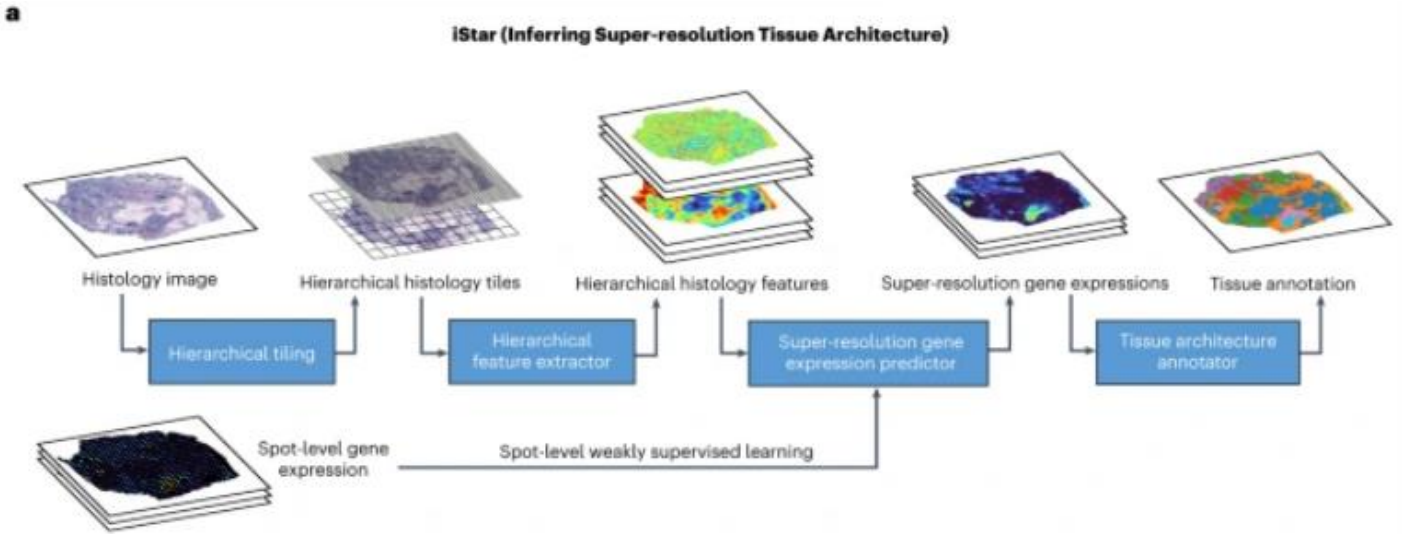
Nature Biotechnology **42**, 1372–1377 (2024) | Cite this article

28k Accesses | **28** Citations | **285** Altmetric | Metrics

Part III. Spatial transcriptomics | *Analysis*

Existing pipelines

iStar



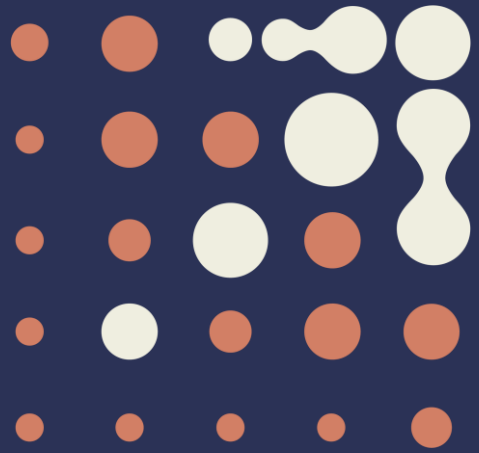
Source: www.nature.com/articles/s41587-023-02019-9

Thank you

N. Alcala & L.Mangé

Rare Cancers Genomics Team

November 26th 2024



International Agency for Research on Cancer

World Health Organization

RARE CANCERS GENOMICS

Appendices

N. Alcala

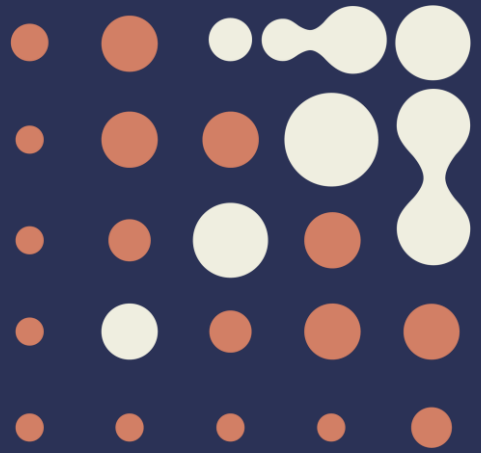
Rare Cancers Genomics Team

November 16th 2022

International Agency for Research on Cancer

World Health Organization

RARE CANCERS GENOMICS



Part I. Transcriptomics | *Concepts*

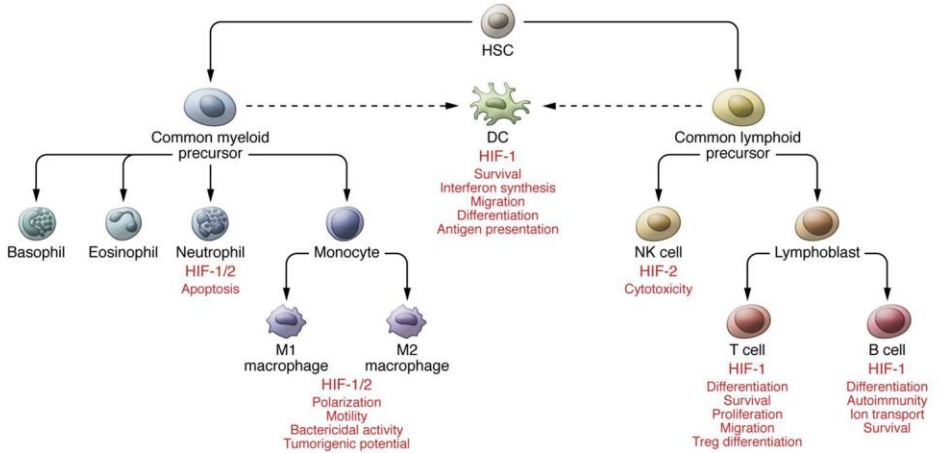
Tissue heterogeneity: Stroma and Microenvironment

Stromal cells (connective tissue cells)

- **Fibroblasts**: synthesize the extracellular matrix and collagen, initiate inflammation and immune response

Immune cells

- **Dendritic cells**: present antigens
- **Macrophages**: perform phagocytosis
- **T cells**: cytotoxic (CD8+), helper (CD4+)
- **Neutrophils**: promote inflammation, phagocytosis



Immune cell differentiation. *Source: Taylor et al. J. Clin Invest 2016.*

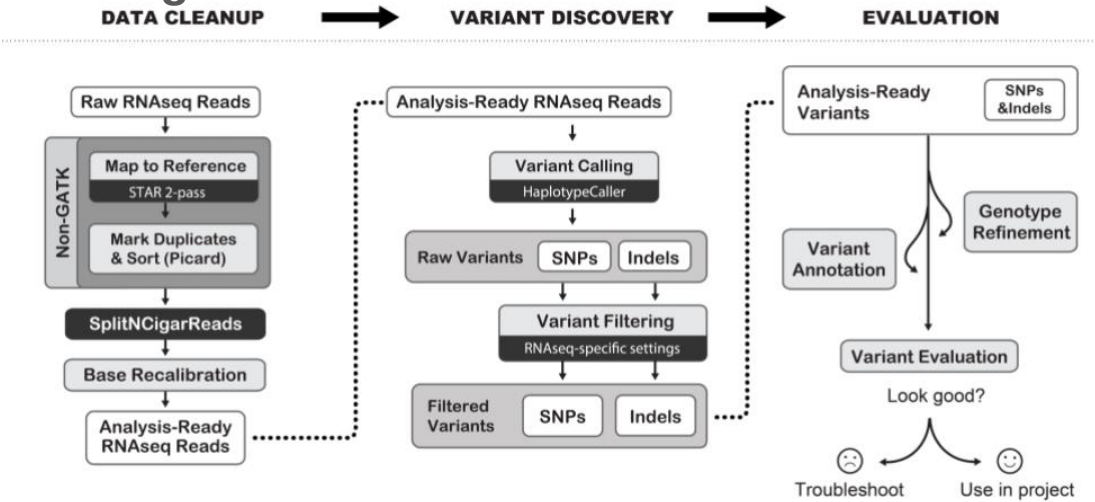
Part I. Transcriptomics | *Analysis*

Variant discovery: small variant calling

Goal: discover (or validate) small somatic variants (single nucleotide polymorphism or indels)

Medical relevance: many diseases are driven by small variants

Methods: Mapping to reference, and heavy filtering using estimated sequencing error rates and databases of known germline variants



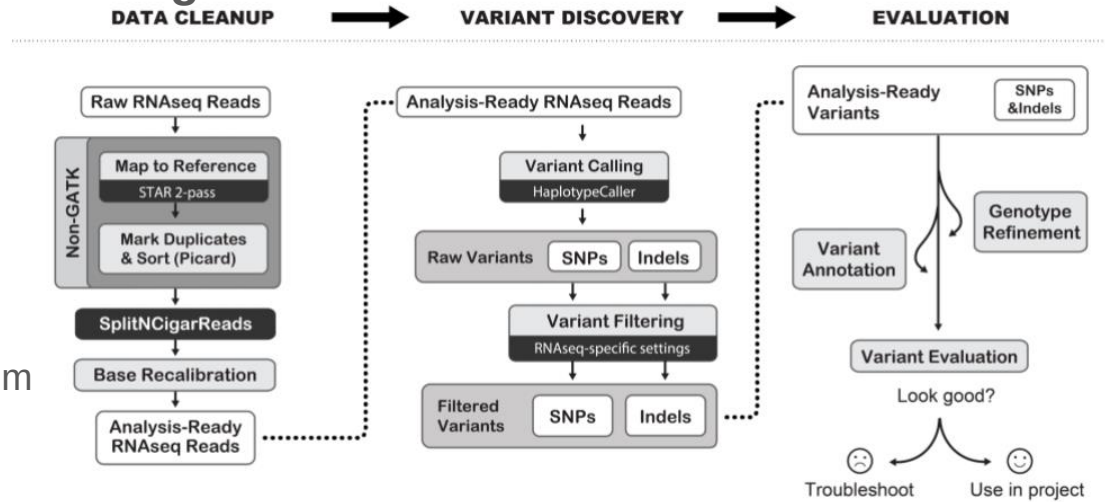
Schematic of the Genome Analysis ToolKit (GATK) best practices for small variant discovery from RNA-seq; *Source: https://gatk.broadinstitute.org*

Part I. Transcriptomics | *Analysis*

Variant discovery: small variant calling

Caveats:

- High false positive rate (due to sequencing error and high depth at some locations)
- High false positive rate (due to variants in low-expression genes)
- Useful for validation of mutations from WGS/WES
- Useful for allele specific expression quantification



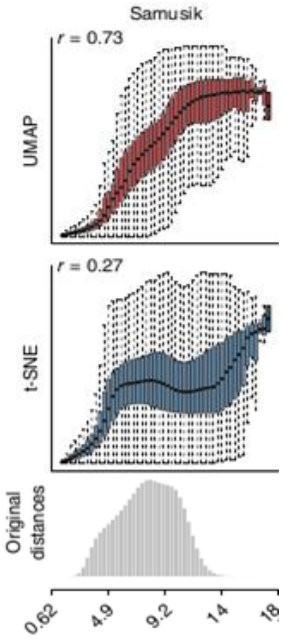
Schematic of the Genome Analysis ToolKit (GATK) best practices for small variant discovery from RNA-seq; *Source: https://gatk.broadinstitute.org*

Part II. Single-cell transcriptomics | *Analysis*

UMAP algorithm:

1. **similarities between points in the original (high-dimensional) space** are computed using fuzzy simplicial sets memberships
2. **Similarities between points in the output (low-dimensional) space** are computed using Student distributions with 1DF
3. **A cost function** (the cross-entropy) is optimized

Notes: UMAP has a faster running time because cross-entropy is easier to optimize, and is claimed to better preserve long distances



Preservation of original distances by UMAP and t-SNE. *Source:* Becht et al. Nature Biotechnology 2019.