From Samples to Insights – RNA sequencing

A powerful solution that works for everyone on NGS-data analyses & annotations

CLC 次世代定序序列分析軟體
IPA 生物路徑分析軟體

Nereus Chang | 張嘉瑋
Senior Engineer | 資深工程師

Cleo Hsi | 席佩妤
Sale Specialist | 業務專員

Christine Hsiung | 熊嘉妮
Project Supervisor | 專案主任

BIOINFORMATICS POWERED BY

OUTLINE

- **Bulk RNA-seq Analysis**
  - Case study : Dynamic gene expression changes in the mouse brain during pregnancy and the postpartum period
  - Hands-on Session
- **Single-cell RNA-seq Analysis**
  - Case study : Key benefits of dexamethasone and antibody treatment in COVID-19 hamster models
  - Hands-on Session
- **Spatial single-cell RNA-seq Analysis**
  - Case study : Spatial transcriptomics reveals distinct tumour core and edge architectures
  - Hands-on Session

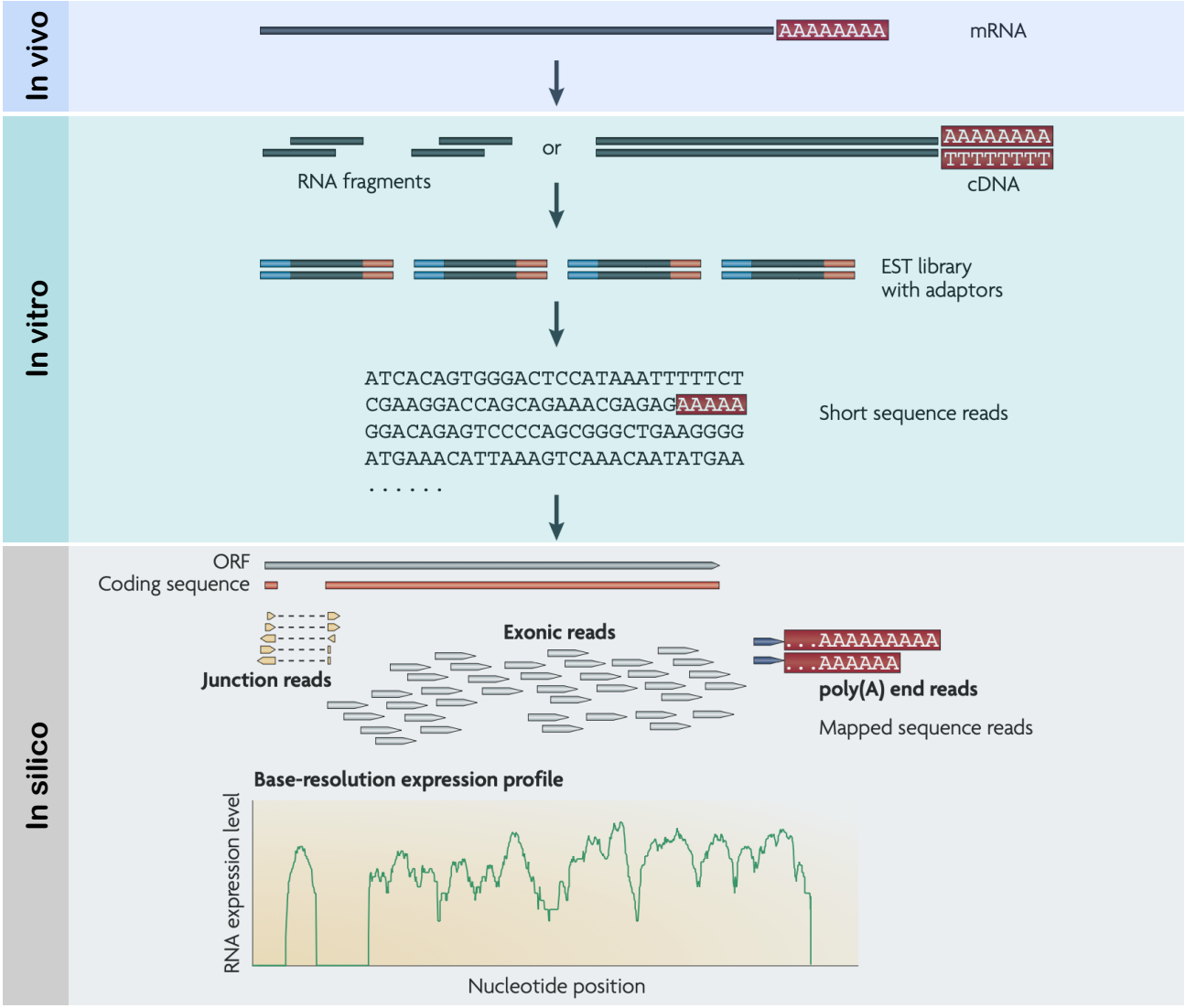
CLC Genomics Workbench

Bulk RNA Sequencing Analysis

BIOINFORMATICS POWERED BY

Sample to Insight

Typical Bulk RNA-seq Overview



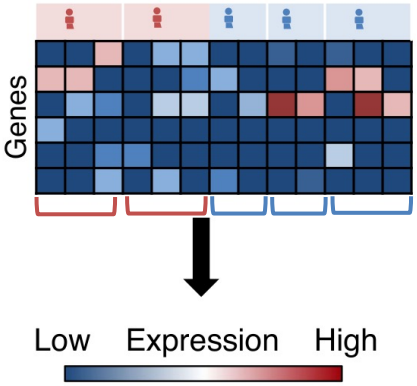
Sample to Insight

Source : Zhong Wang, et al., 2009. https://doi.org/10.1038/nrg2484.

Conventional RNA-seq Analysis Study Design

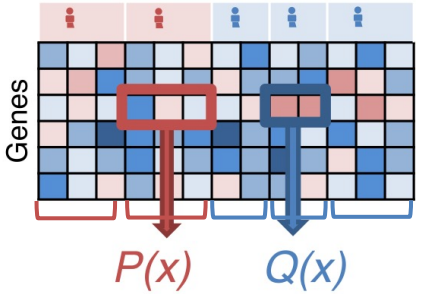
Raw counts

Normalization, log transformed, and/or denoising

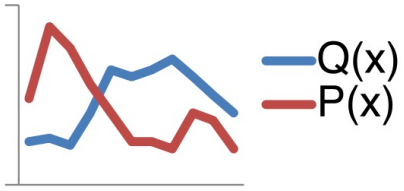


Processed gene expression

Estimate expression distribution given cell-level covariates



Gene expression distribution for each gene and each individual



READ MAPPING

Find original read source within the reference genome or transcriptome.

	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	15	40
gene 3	33	40	35	70
…	…	…	…	…
gene N	100	120	105	220

COUNTS COMPUTATION

Estimate gene expression with "counts", i.e. with the number of reads mapped on each gene.

	sample A1	sample A2	sample B1	sample B2
gene 1	8	10	100	200
gene 2	14	15	115	40
gene 3	33	40	35	70
…	…	…	…	…
gene N	100	120	105	220

Tot. reads: 5 millions

Tot. reads: 10 millions

COUNTS NORMALIZATION

Eliminate biases to make expression levels comparable between samples (e.g. different sequencing depths of samples A1 and B2) and within samples (e.g. different lengths of gene 1 and gene 2).

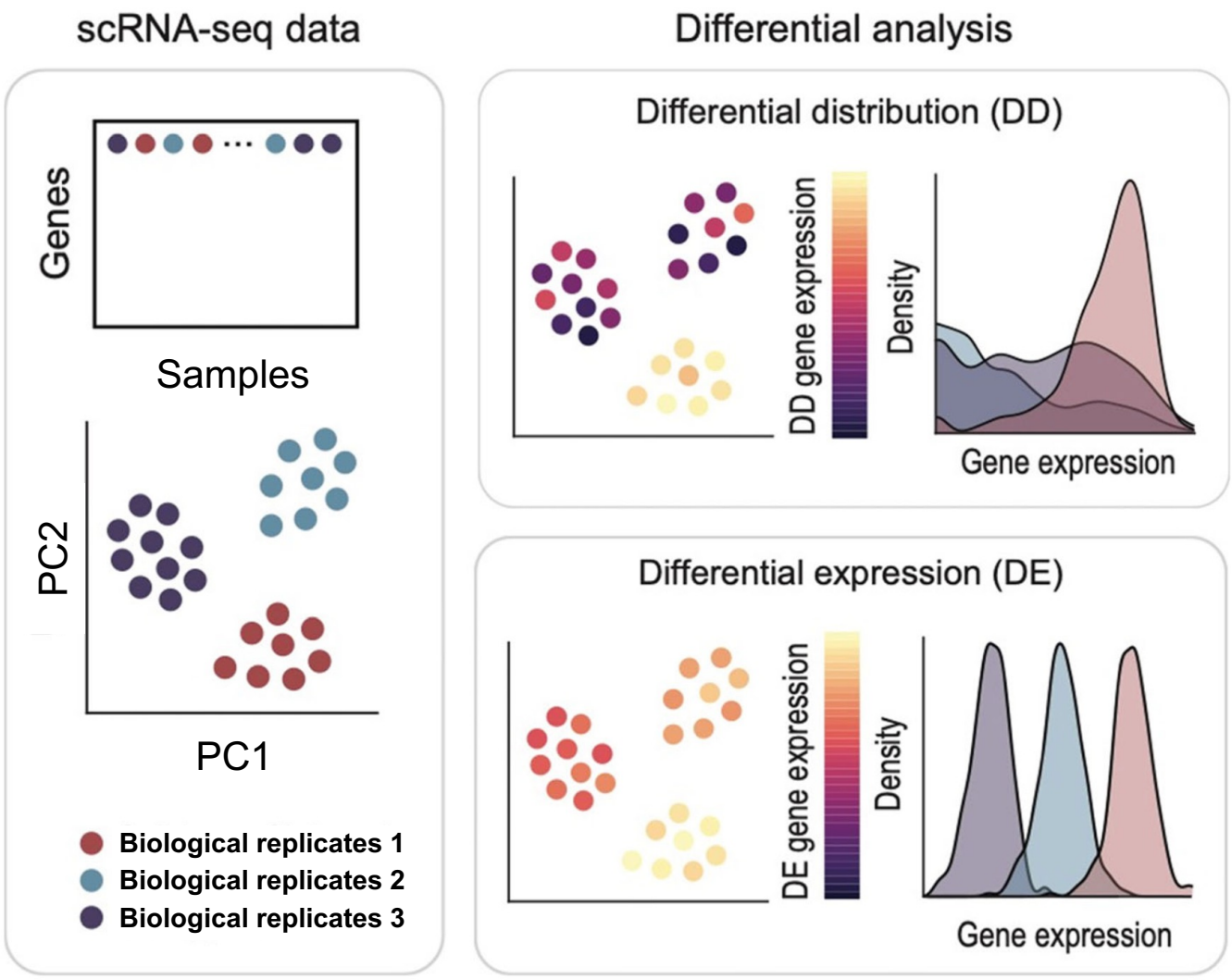
	sample A1	sample A2	sample B1	sample B2
gene 1	0.16	0.20	2.00	2.00
gene 2	0.28	0.30	0.30	0.40
gene 3	0.66	0.80	0.70	0.70
…	…	…	…	…
gene N	2.00	2.40	2.10	2.20

DIFFERENTIAL EXPRESSION ANALYSIS

Identify genes with statistically different expression levels in the compared conditions (e.g. A and B).

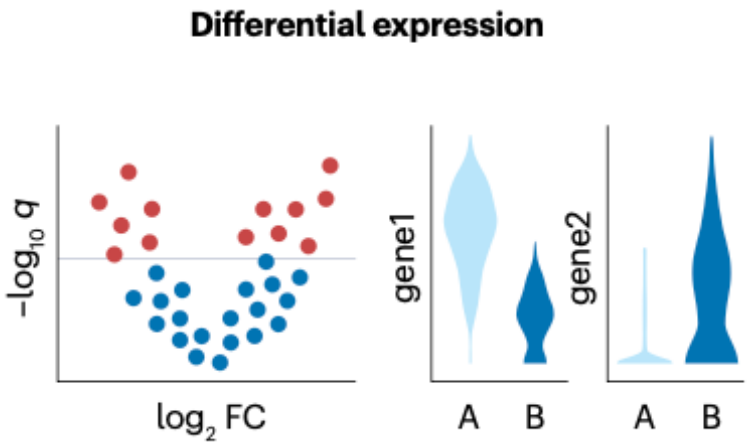
Source : Zhang, Mengqi, et al., 2022. https://doi.org/10.1186/s13059-022-02605-1.; Finotello, F., et al., 2014. https://doi.org/10.1093/bfgp/elu035.

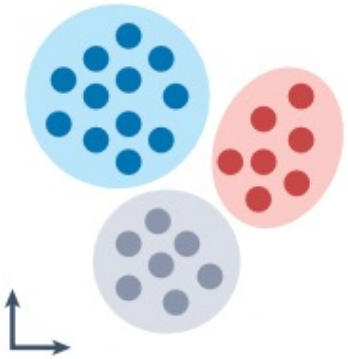
Differential Gene Expression Analysis



Source : Kim, Hani Jieun, et al., 2021. https://doi.org/10.1186/s13619-021-00083-7.

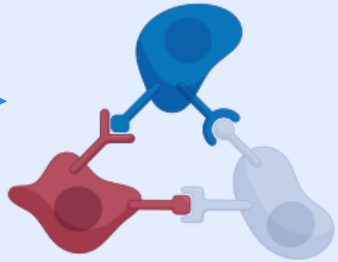
Bulk RNA-seq Downstream Analyses





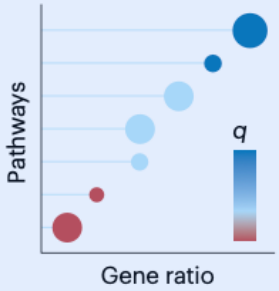
Differential expressed genes

Gene ontology (GO) analysis



- Cellular Component
- Biological Process
- Molecular Function

Gene set enrichment analysis (GSEA)



Gene regulatory networks (e.g. KEGG)

IPA

Sample to Insight

Source : Heumos, L., et al. 2023, https://doi.org/10.1038/s41576-023-00586-w

CLC Genomics Workbench

Hands-on session: RNA-seq analysis

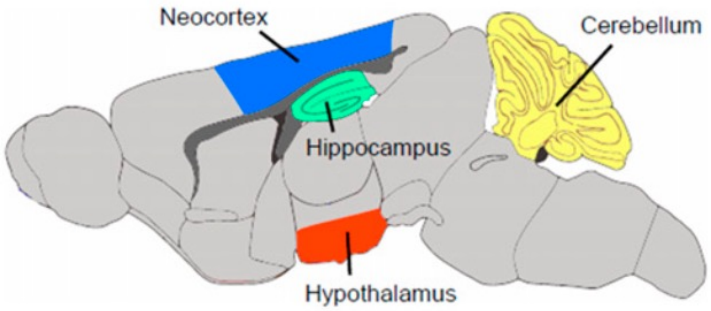
BIOINFORMATICS POWERED BY

Sample to Insight

Study of Sample Data and Study Design

Ray, S., et al. (2015). **An examination of dynamic gene expression changes in the mouse brain during pregnancy and the postpartum period.** *G3 (Bethesda, Md.)*, *6*(1), 221–233. https://doi.org/10.1534/g3.115.020982

A



B

Virgin	*Female mouse - unmated*
PC14	*Dam - 14 days post-conception*
PC16	*Dam - 16 days post-conception*
PP1	*Dam - 1 day postpartum*
PP3	*Dam - 3 days postpartum*
PP10	*Dam - 10 days postpartum*

Three samples for each tissue (4) and time points (6), 72 samples in total.

Source : Ray S, et al., 2015 . https://10.1534/g3.115.020982.

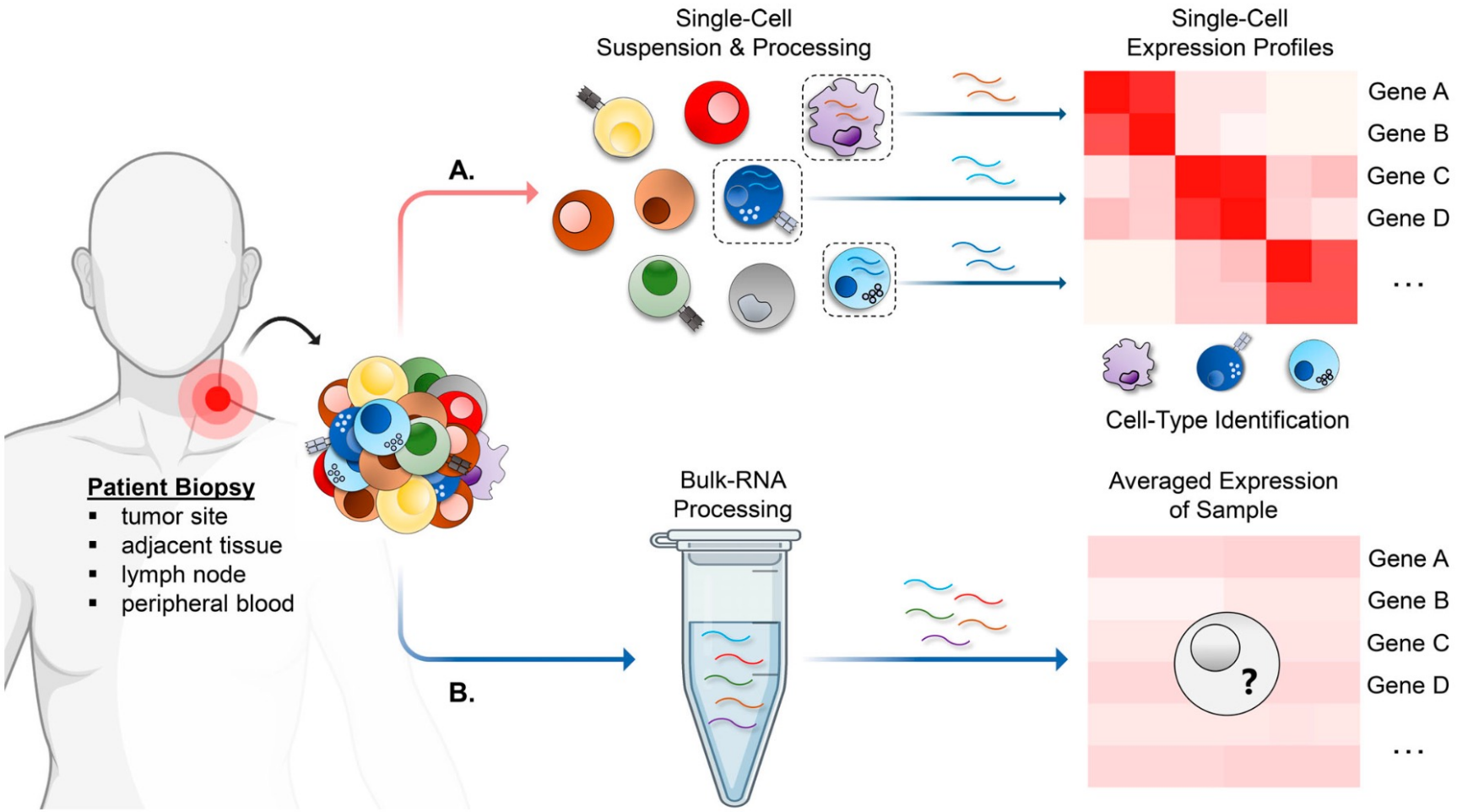
CLC Genomics Workbench

Single-cell RNA-Seq Analysis

BIOINFORMATICS POWERED BY

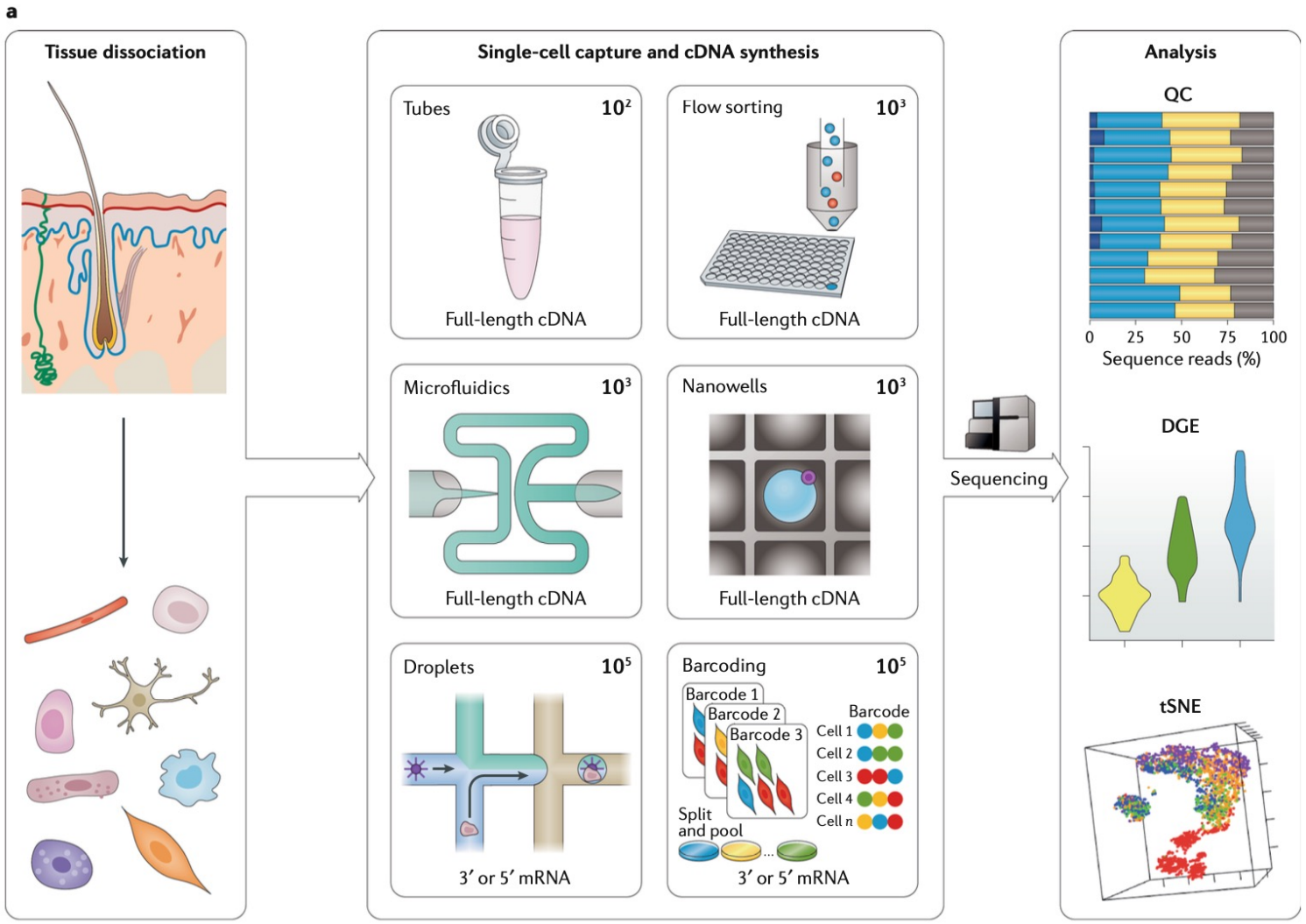
Sample to Insight

Bulk RNA-seq vs. Single-Cell RNA-seq (scRNA-seq)



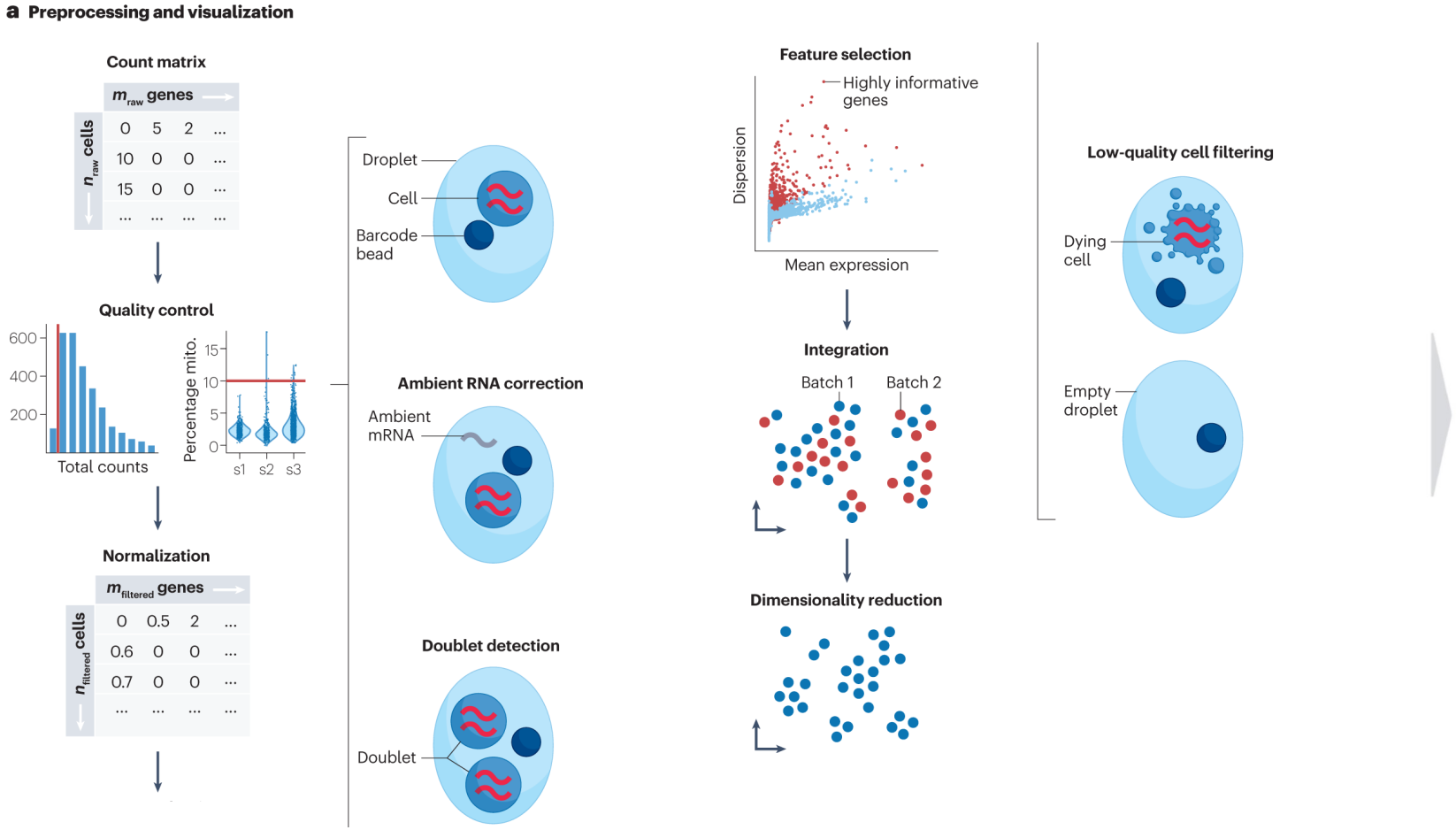
Source : Guruprasad, Puneeth, et al., 2020. https://doi.org/10.1084/jem.20201574.

scRNA-seq with droplet-based methods



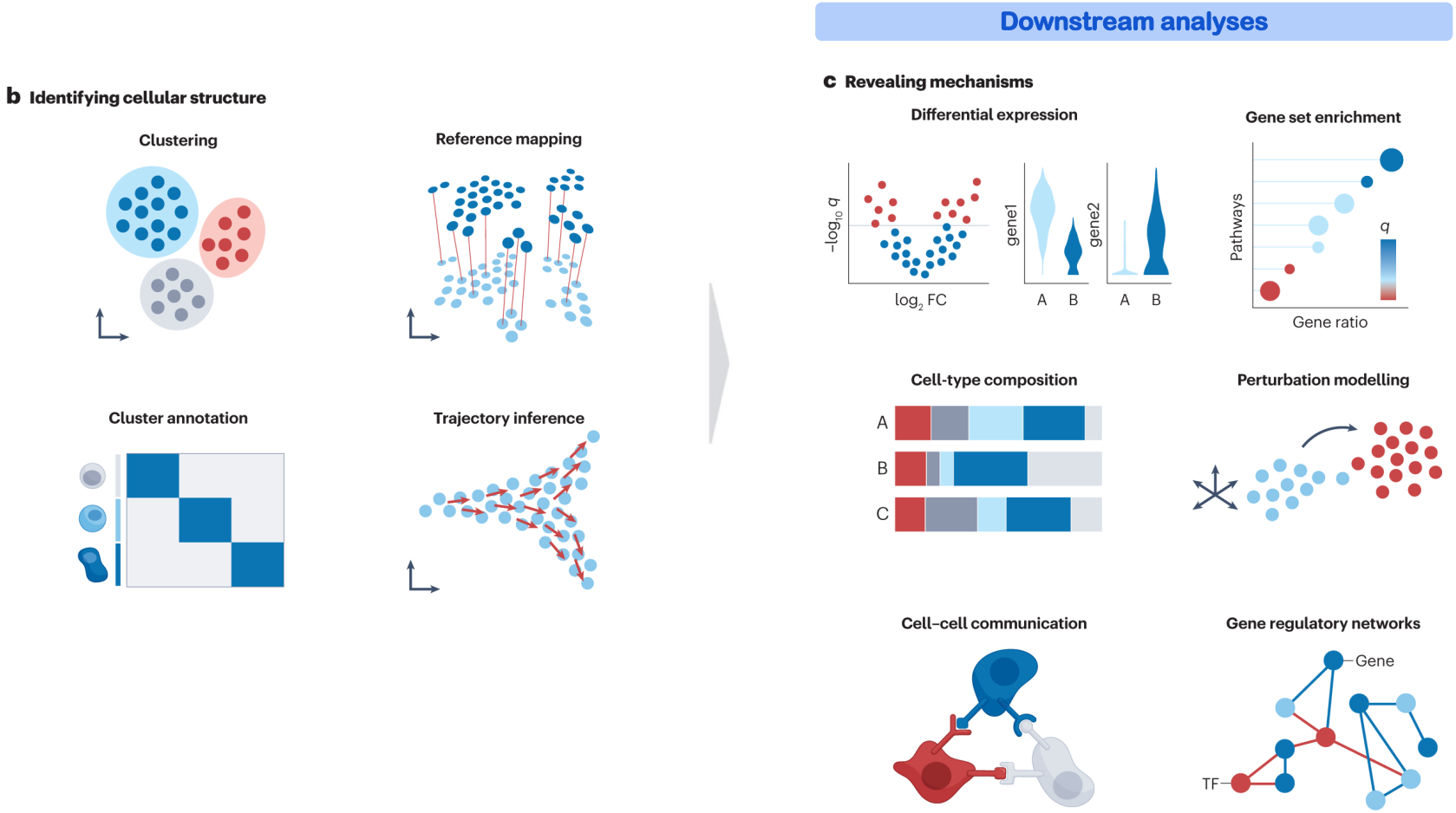
Source : Stark, Rory. 2019. https://doi.org/10.1038/s41576-019-0150-2.

Spatial scRNA-seq Analysis (1/2)



Source : Heumos, L., et al. 2023, https://doi.org/10.1038/s41576-023-00586-w

Spatial scRNA-seq Analysis (2/2)



Sample to Insight

Source : Heumos, L., et al. 2023, https://doi.org/10.1038/s41576-023-00586-w

CLC Genomics Workbench

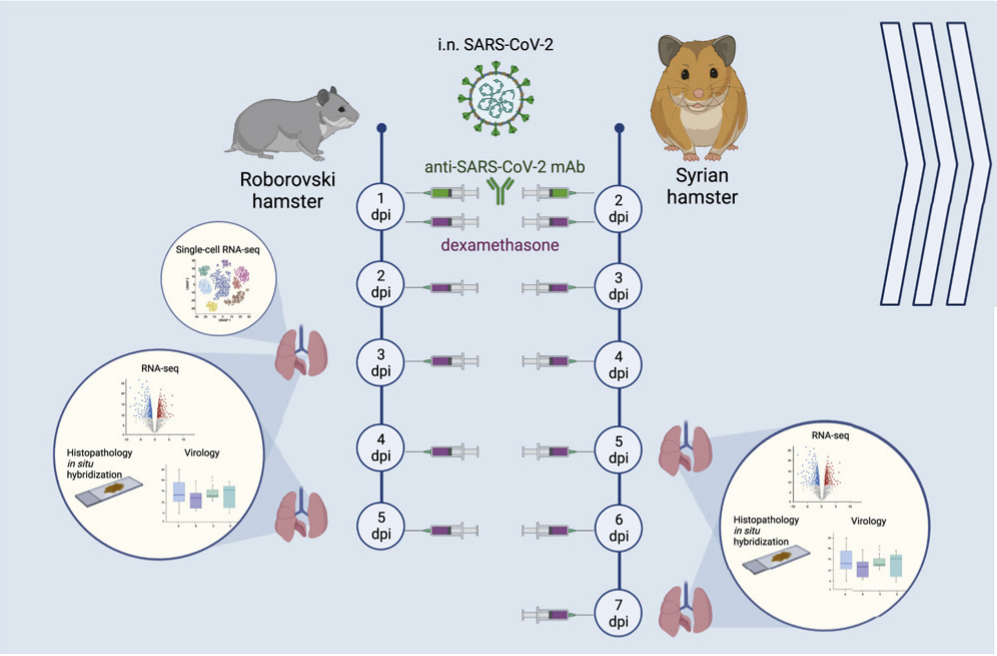
Hands-on session: scRNA-seq analysis

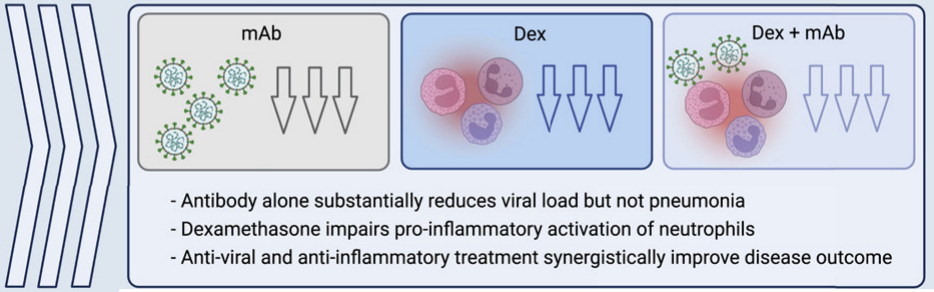
BIOINFORMATICS POWERED BY

Sample to Insight

Study of Sample Data and Study Design

Wyler, E., et al. (2022). **Key benefits of dexamethasone and antibody treatment in COVID-19 hamster models revealed by single-cell transcriptomics.** Molecular therapy, 30(5), 1952–1965. https://doi.org/10.1016/j.ymthe.2022.03.014

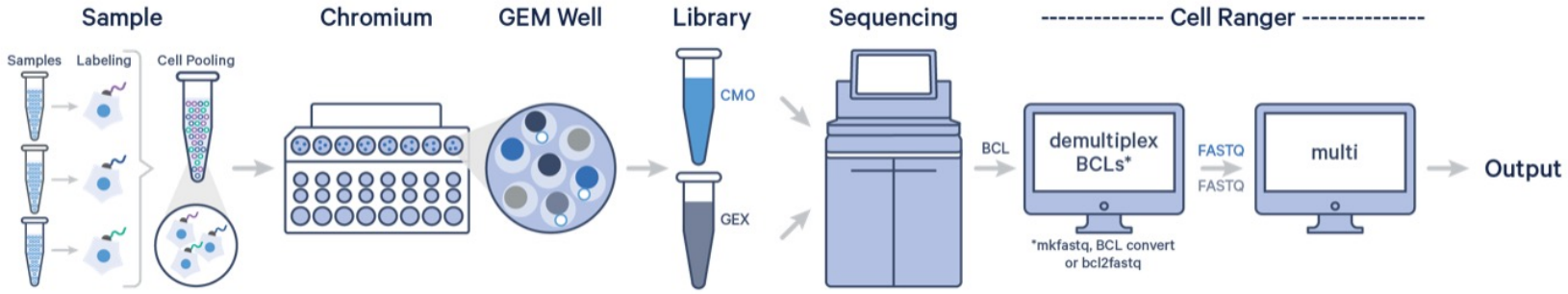




Source : Wyler E, et al., 2022 . https://10.1016/j.ymthe.2022.03.014.

Analysing 3’ CellPlex data with *Cell Ranger*

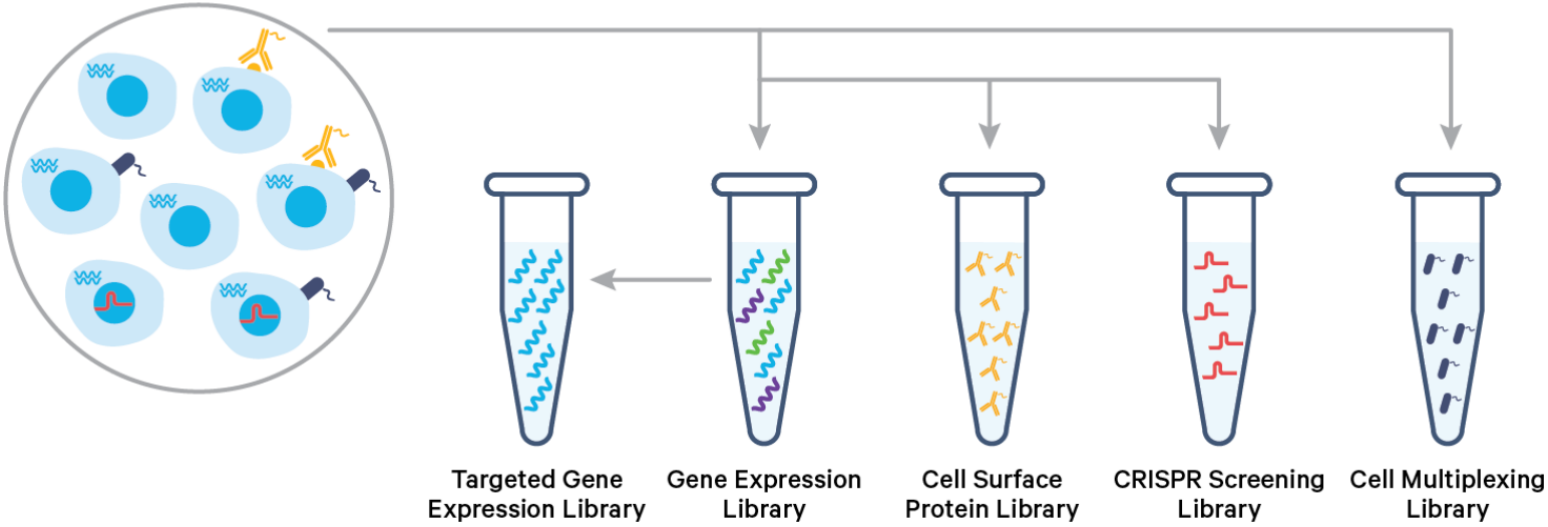
Cell Ranger v6.0 or later is required to analyze Cell Multiplexing data. Multiple samples are uniquely tagged with CMOs prior to pooling in a single GEM well, resulting in a CMO and Gene Expression (GEX) library for each GEM well. After demultiplexing the BCL files, run the cellranger multi pipeline on the combined FASTQ data for the CMO and GEX libraries to obtain separate per-sample output files for each CMO.



Source : https://www.10xgenomics.com/support/software/cell-ranger/latest/getting-started/cr-3p-what-is-cellplex

10x Genomics 3’ CellPlex Multiplexing Soul.

The 10x Genomics 3’ CellPlex Multiplexing Solution is a Feature Barcode technology, similar to existing 10x Genomics Cell Surface Protein and CRISPR Screening assays. 3’ CellPlex is enabled by SingleCell 3’ v3.1 gelbeads and reagents. Cells or nuclei labeled with Cell Multiplexing Oligos (CMOs) are pooled prior to loading onto a 10x Genomics chip. cDNA from poly-adenylated mRNA and DNA from the CMO Feature Barcode are generated simultaneously from the same single cells inside of a Gel Bead-in-Emulsion (GEM). 3’ CellPlex reagents are compatible with samples whose cell surface proteins have been labeled with TotalSeq-B antibody-oligonucleotide conjugates or samples transduced with Feature Barcode technology compatible sgRNA constructs.



Source : https://www.10xgenomics.com/support/software/cell-ranger/latest/getting-started/cr-3p-what-is-cellplex

Data Repository - Gene Expression Omnibus

Series GSE191080	Query DataSets for GSE191080
Status	Public on Dec 18, 2021
Title	Insights into standards of care – dexamethasone and antibodies against COVID-19 in hamster models
Organisms	Mesocricetus auratus; Phodopus roborovskii
Experiment type	Expression profiling by high throughput sequencing
Summary	We describe the cellular response to SARS-CoV-2 infections combined with antibody and/or dexamethasone treatment in Syrian and Roborovski dwarf hamsters
Overall design	Histopathology and bulk RNA-seq was performed from the lung of animals both species, single-cell RNA-sequencing was performed from lungs of Roborvski dwarf hamsters, upon SARS-CoV-2 infections combined with the indicated treatments.
Contributor(s)	Wyler E, Adler JM, Eschke K, Teixeira G, Peidli S, Pott F, Kazmierski J, Postmus D, Michalick L, Kershaw O, Hoppe J, Andreotti S, Pennitz P, Goffinet C, Kreye J, Reincke SM, Prüss H, Blüthgen N, Gruber AD, Witzenrath M, Landthaler M, Nouailles G, Trimpert J
Citation(s)	35339689

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191080

Data Repository - Gene Expression Omnibus

Samples (54)
Less...

GSM5738371	RNA_aaUntr_5dpi_ha1
GSM5738372	RNA_aaUntr_5dpi_ha2
GSM5738373	RNA_aaUntr_5dpi_ha3
GSM5738374	RNA_aaUntr_7dpi_ha1
GSM5738375	RNA_aaUntr_7dpi_ha2

⋮

GSM5738419	Hamster1_CELLPLEX
GSM5738420	Hamster1_GEX
GSM5738421	Hamster2_CELLPLEX
GSM5738422	Hamster2_GEX
GSM5738423	Hamster3_CELLPLEX
GSM5738424	Hamster3_GEX

scRNA-seq data

Relations

BioProject PRJNA789576

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE191080

Data Repository - Gene Expression Omnibus

Sample GSM5738420	Query DataSets for GSM5738420
Status	Public on Dec 18, 2021
Title	Hamster1_GEX
Sample type	SRA
Source name	multiplexed sample
Organism	Phodopus roborovskii
Characteristics	tissue: Lung treatment: Multiplexed sample (mixed treatments) sampling timepoint: 3dpi
Treatment protocol	At 6-10 weeks of age hamsters were infected intranasally as previously described (PMID 32698441, 33271063). Briefly, hamster received 1 x 105 pfu SARS-CoV-2 Variant B1 (BavPat1) intranasally under anesthesia.
Growth protocol	Hamsters were weighted daily and monitored for signs of disease twice-daily. Severely sick animals were euthanized according to defined humane endpoints including body temperature <33 °C, acute respiratory distress or weight loss >20%. Hamsters were selected randomly for all scheduled take-out time points (day 3 and 5 for Roborovski hamsters, day 5 and 7 for Syrian hamsters). Euthanasia prior analysis occurred by exsanguination under anesthesia as previously described (PMID 32698441, 33271063). Peripheral blood was collected in EDTA-coated syringes and lung lobes were collected for follow-up analyses, in which the left lobe was used for histopathology, the right caudal lobe for single-cell analysis, the right cranial lobe for virological assessments and the right medial as backup to repeat virological analyses as needed.
Extracted molecule	polyA RNA

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5738420

Data Repository - Gene Expression Omnibus

Library strategy	RNA-Seq
Library source	transcriptomic
Library selection	cDNA
Instrument model	Illumina NovaSeq 6000

Description	Processed data files include DexAb*rds and DexAb.loom files on series record. 10xGenomics Chromium v3.1 single cell 3' gene expression kit
Data processing	For bulk RNA-sequencing, alignments were done using hisat2 (Kim et al, 2015) Syrian hamster: Sequencing reads were aligned to the MesAur 1.0 version of the Mesocricetus auratus genome using standard parameter, using the Refseq gtf file described in GEO accession number GSE162208 Roborovski hamster: Sequencing reads were aligned to the draft assembly and gtf file described previously (doi: 10.1101/2021.10.02.462569) For single-cell RNA-sequencing data, fastq files were processed using a merged Roborvski hamster/SARS-CoV-2 genome by applying CellRanger 6.0.2 with standard parameters Seurat R and loom objects were processed as described on the github page https://github.com/Berlin-Hamster-Single-Cell-Consortium/Dwarf-Hamster-Dexamethasone-Antibody Genome_build: MesAur 1.0 Genome_build: Phodopus roborvskii draft assembly Genome_build: SARS-CoV-2 MN908947 Supplementary_files_format_and_content: tab-separated values Supplementary_files_format_and_content: h5 matrix Supplementary_files_format_and_content: Seurat R object Supplementary_files_format_and_content: loom object

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6339632

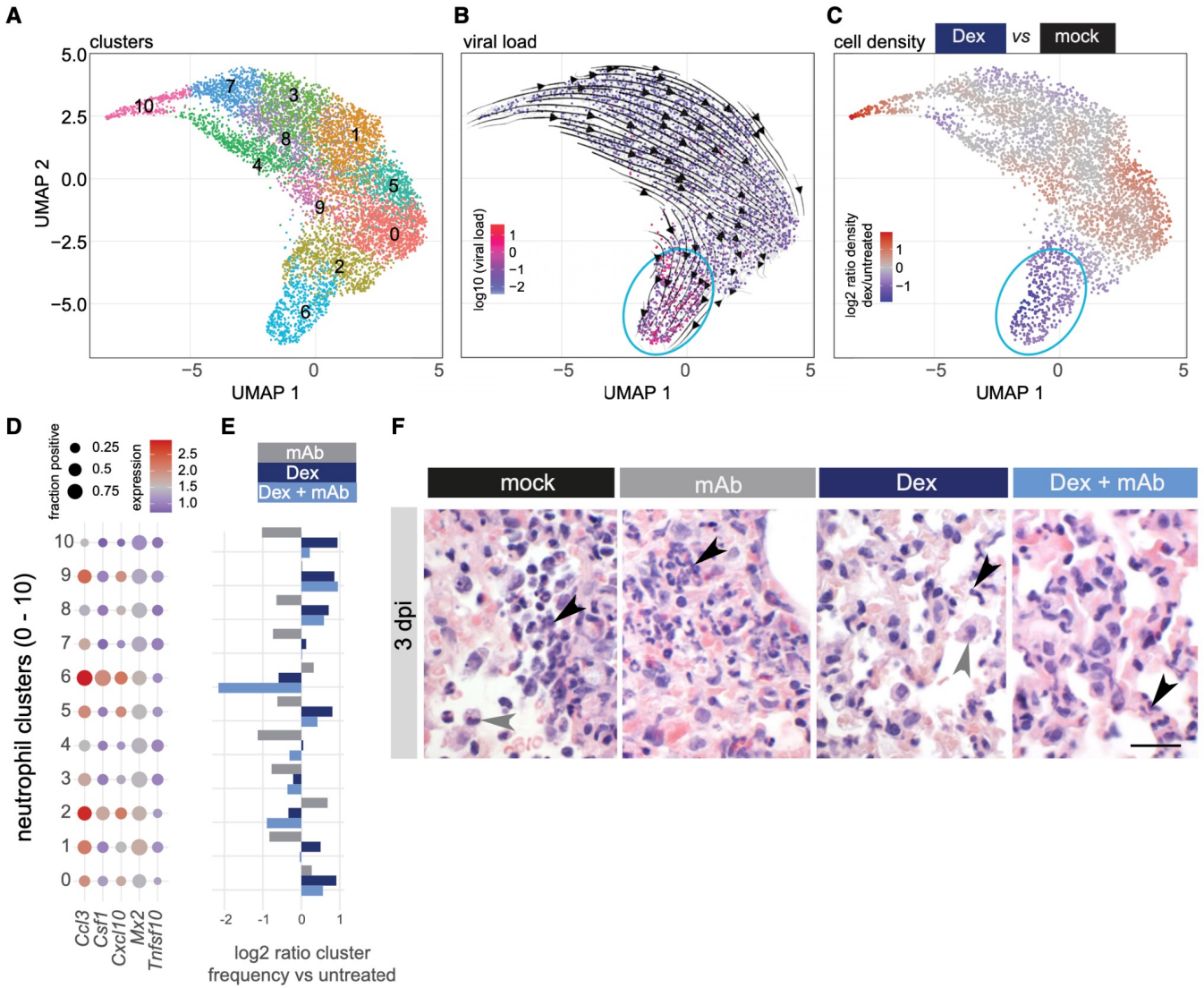
Data Repository - Gene Expression Omnibus

BioSample SAMN24144554
SRA SRX13428793

Supplementary file	Size	Download	File type/resource
GSM5738420_Hamster1_AB1_sample_feature_bc_matrix.h5	2.2 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_AB2_sample_feature_bc_matrix.h5	3.9 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_AB3_sample_feature_bc_matrix.h5	3.8 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_ABD1_sample_feature_bc_matrix.h5	2.7 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_ABD2_sample_feature_bc_matrix.h5	2.1 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_ABD3_sample_feature_bc_matrix.h5	4.1 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_D1_sample_feature_bc_matrix.h5	2.3 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_D2_sample_feature_bc_matrix.h5	3.9 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_D3_sample_feature_bc_matrix.h5	4.5 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_U1_sample_feature_bc_matrix.h5	3.6 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_U2_sample_feature_bc_matrix.h5	2.8 Mb	(ftp)(http)	H5
GSM5738420_Hamster1_U3_sample_feature_bc_matrix.h5	2.2 Mb	(ftp)(http)	H5

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6339632

Treatment Effects Revealed by scRNA-seq



Source : Wyler E, et al., 2022 . https://10.1016/j.ymthe.2022.03.014.

CLC Genomics Workbench

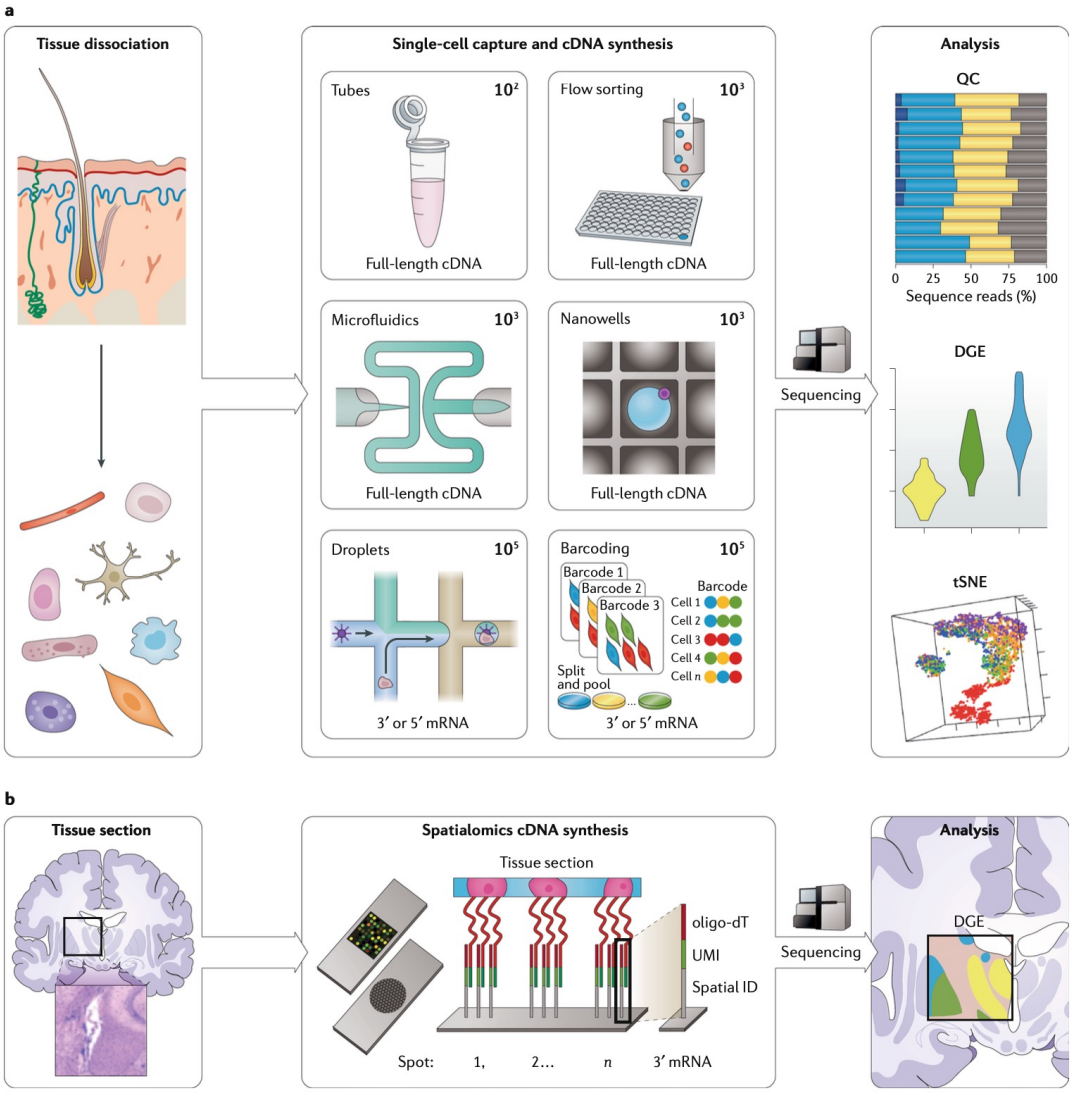
Spacial scRNA-Seq Analysis

BIOINFORMATICS POWERED BY

INGENUITY SYSTEMS · CLCbio · BIOBASE BIOLOGICAL DATABASES · OmicSoft Corporation · n of one

Sample to Insight

scRNA-seq to Spatial scRNA-seq



Sample to Insight

Source : Stark, Rory. 2019. https://doi.org/10.1038/s41576-019-0150-2.

Technologies of Spatial Transcriptomics

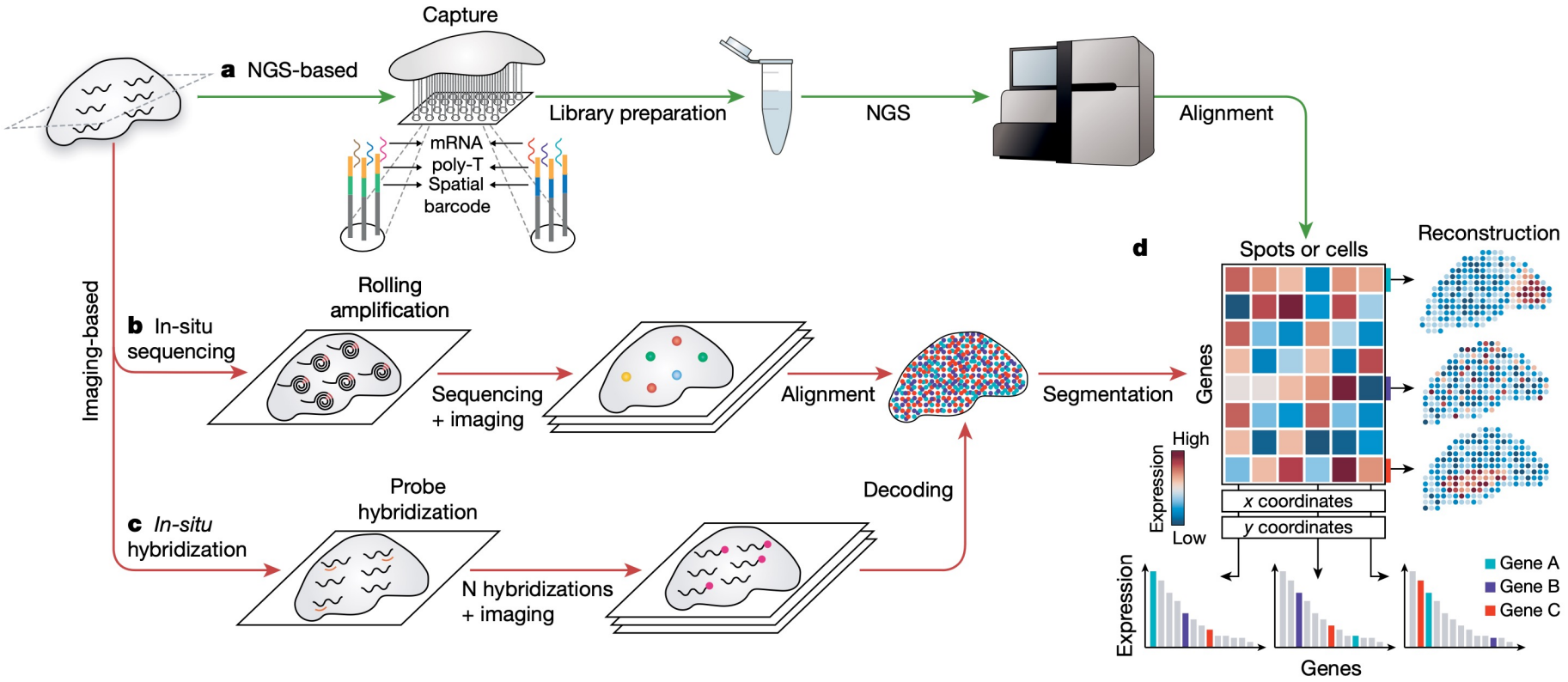


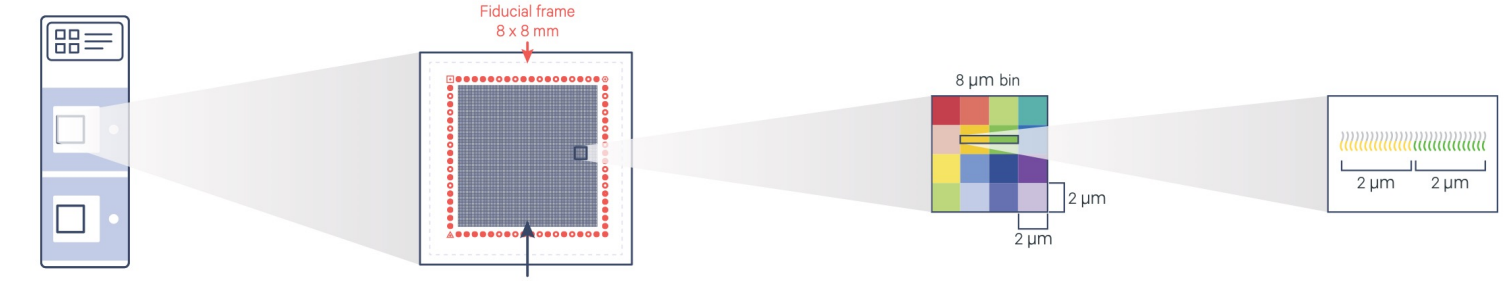
Fig. 1 | The technologies of spatial transcriptomics provide a gene-expression matrix. a, NGS-based spatial transcriptomic methods barcode transcripts according to their location in a lattice of spots. **b**, ISS approaches directly read out the transcript sequence within the tissue. **c**, ISH methods detect target sequences by hybridization of complementary fluorescent probes. **d**, The product of spatial transcriptomics is the gene-expression matrix, in which the rows and columns correspond to genes and locations.

Source : Rao, Anjali. 2021. https://doi.org/10.1038/s41586-021-03634-9.

10x Genomics Visium HD

Spatial Gene Expression

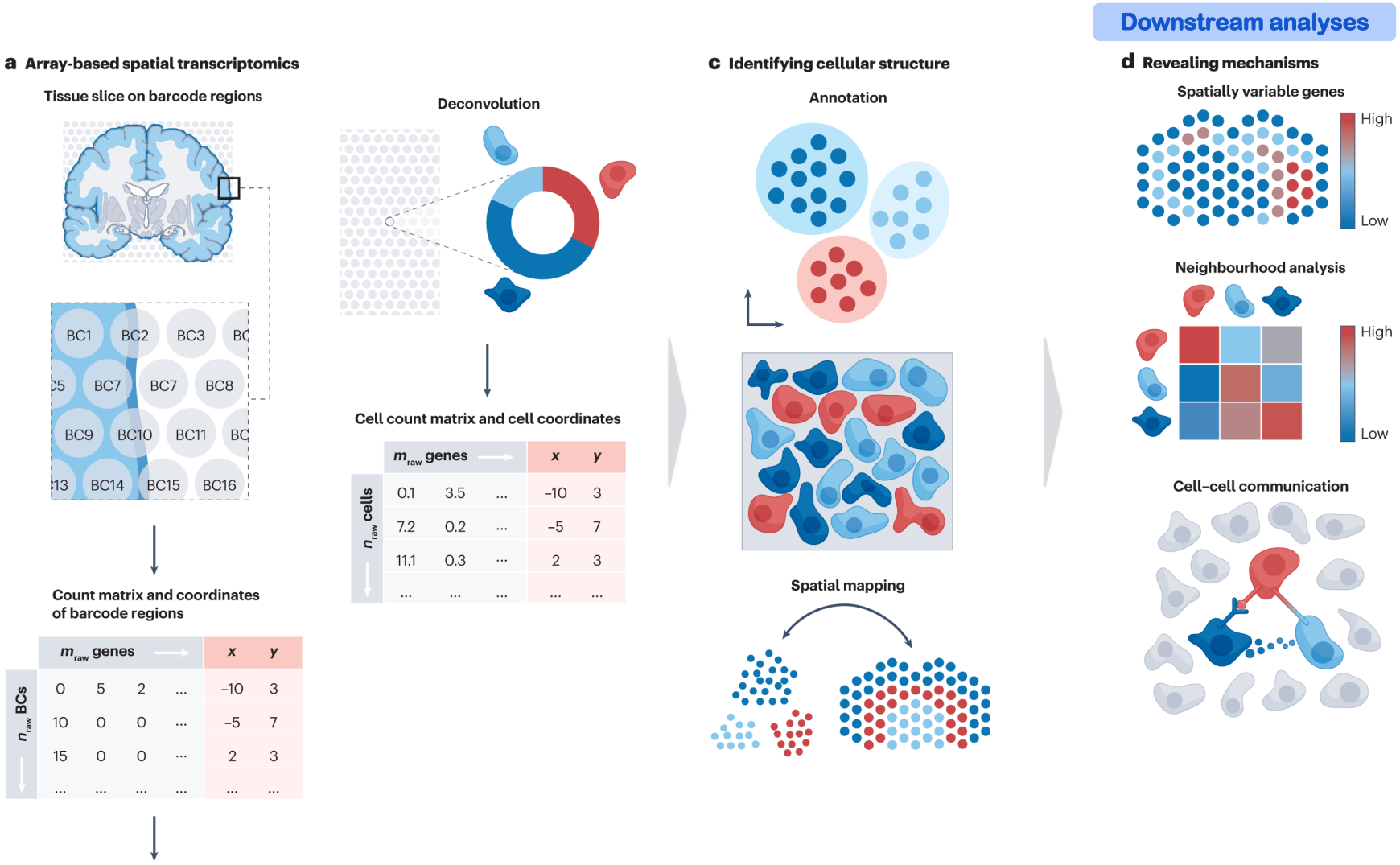
Next-generation slide architecture enables single cell–scale resolution



Visium HD Spatial Gene Expression slides contain two 6.5 x 6.5 mm Capture Areas with a continuous lawn of oligonucleotides arrayed in ~11 million 2 x 2 µm barcoded squares without gaps, achieving single cell–scale spatial resolution. The data is output at 2 µm, as well as multiple bin sizes. The 8 x 8 µm bin is the recommended starting point for visualization and analysis.

Source : https://pages.10xgenomics.com/rs/446-PBO-704/images/SAM000733_Visium_HD_Flyer_Launch_Letter_Digital.pdf

Spatial scRNA-seq Analysis



Sample to Insight

Source : Heumos, L., et al. 2023, https://doi.org/10.1038/s41576-023-00586-w

CLC Genomics Workbench

Hands-on session: Spatial scRNA-seq

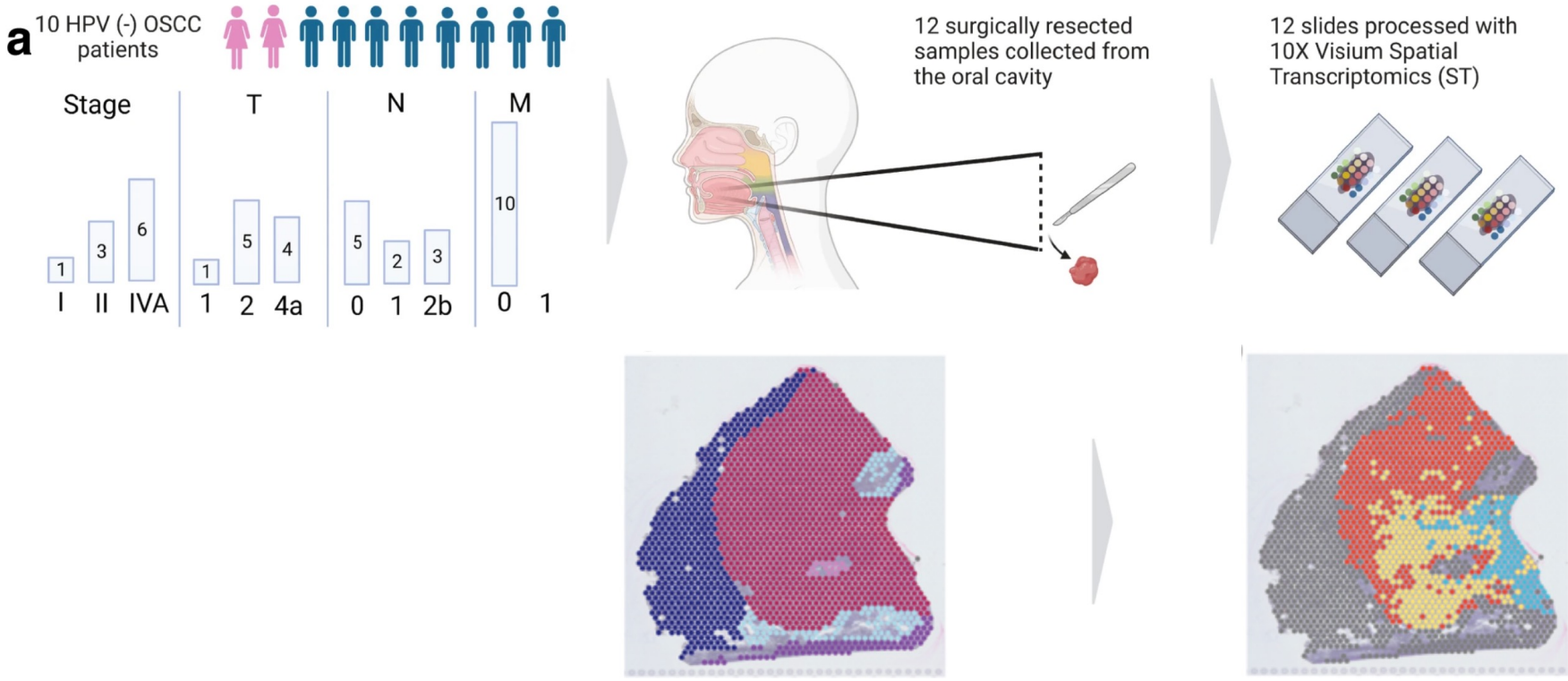
BIOINFORMATICS POWERED BY

INGENUITY SYSTEMS · CLCbio · BIOBASE BIOLOGICAL DATABASES · OmicSoft Corporation · n of one

Sample to Insight

Study of Sample Data and Study Design

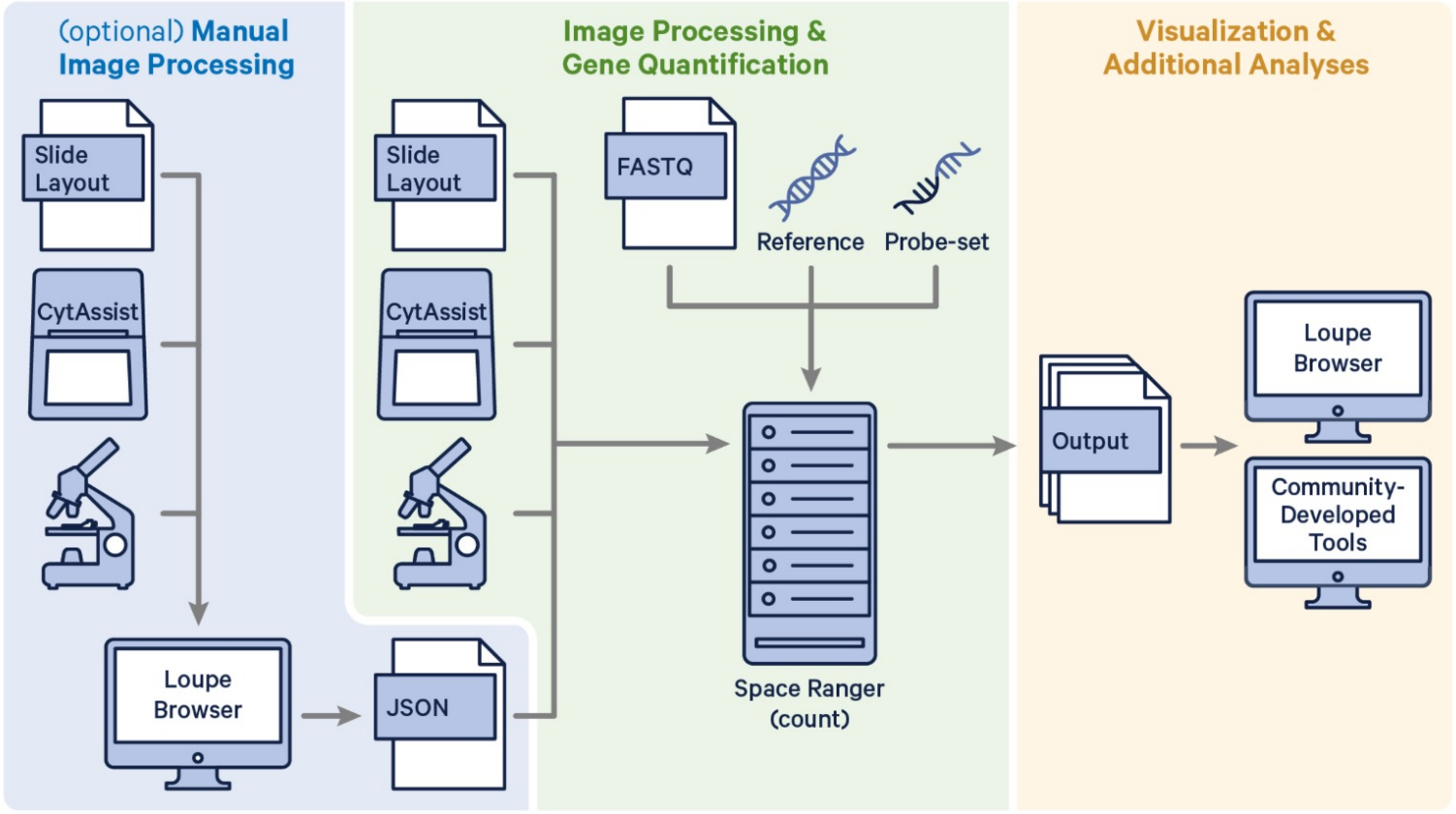
Arora, R., et al. (2023). **Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response.** Nature communications, 14(1), 5029. https://doi.org/10.1038/s41467-023-40271-4



Source : Arora, R., et al. (2023). https://doi.org/10.1038/s41467-023-40271-4

Visium HD Analysis with *spaceranger* count

Space Ranger v3.0 and later can analyze Visium HD Spatial Gene Expression datasets generated from formalin fixed paraffin embedded (FFPE) human or mouse samples and the Visium CytAssist instrument. The `spaceranger count` pipeline is necessary to analyze these data.



Source : https://www.10xgenomics.com/support/software/space-ranger/latest/analysis/running-pipelines/count-visium-hd

Data Repository - Gene Expression Omnibus

Sample GSM6339632	Query DataSets for GSM6339632
Status	Public on Jun 25, 2023
Title	sample_2
Sample type	SRA
Source name	OSCC
Organism	Homo sapiens
Characteristics	Stage: Stage IVA tissue: Oral squamous cell carcinoma (OSCC) library preparation: 10X Genomics Visium Spatial Gene Expression Slide & Reagent kit
Extracted molecule	polyA RNA
Extraction protocol	Banked fresh-frozen OSCC samples were obtained from the in collaboration with the Ohlson Research Institute and the Department of Pathology and Laboratory Medicine. This study was reviewed and approved by the Health Research Ethics Board of Alberta – Cancer Committee (reference number: HREBA.CC-16-0644) A subset of samples were then chosen for downstream processing based on tissue quality. OSCC samples were permeabilized enzymatically according to the Visium Protocol CG000240. Samples were then processed using the Visium Spatial Gene Expression Reagent Kit according to the manufacturer's instructions (10x Genomics, Pleasanton, CA, USA). OSCC samples were permeabilized enzymatically according to the Visium Protocol CG000240. Samples were then processed using the Visium Spatial Gene Expression Reagent Kit according to the manufacturer's instructions (10x Genomics, Pleasanton, CA, USA).

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6339632

Data Repository - Gene Expression Omnibus

Library strategy	OTHER
Library source	transcriptomic
Library selection	other
Instrument model	Illumina NovaSeq 6000
Data processing	All raw FASTQ reads were aligned to Homo_sapiens : GRCh38 reference using the spaceranger count algorithim and quantified through the 10X Genomics space ranger spatial transcriptomics pipeline, with default and recommended parameters. Spaceranger outputs were combined and read depth normalized using the spaceranger aggr function, producing aggregated feature-barcode matrices used in downstream analysis. Assembly: Homo_sapiens : GRCh38 Supplementary files format and content: Filtered feature - barcode HDF5 file output from spaceranger pipeline, Spatial tissue position list CSV, high resolution tissue image as a PNG, scalefactors for tissue positions as a JSON Library strategy: Spatial Transcriptomics

Submission date	Jul 14, 2022
Last update date	Jun 29, 2023
Contact name	Pinaki Bose
E-mail(s)	pbose@ucalgary.ca
Organization name	University of Calgary
Street address	HMRB 354, 3330 Hospital Dr. NW
City	Calgary

Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6339632

Data Repository - Gene Expression Omnibus

Platform ID	GPL24676	
Series (1)	GSE208253	Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response

Relations

BioSample	SAMN29758218
SRA	SRX16247929

Supplementary file	Size	Download	File type/resource
GSM6339632_s2_filtered_feature_bc_matrix.h5	5.8 Mb	(ftp)(http)	H5
GSM6339632_s2_scalefactors_json.json.gz	168 b	(ftp)(http)	JSON
GSM6339632_s2_tissue_hires_image.png.gz	6.0 Mb	(ftp)(http)	PNG
GSM6339632_s2_tissue_positions_list.csv.gz	66.7 Kb	(ftp)(http)	CSV

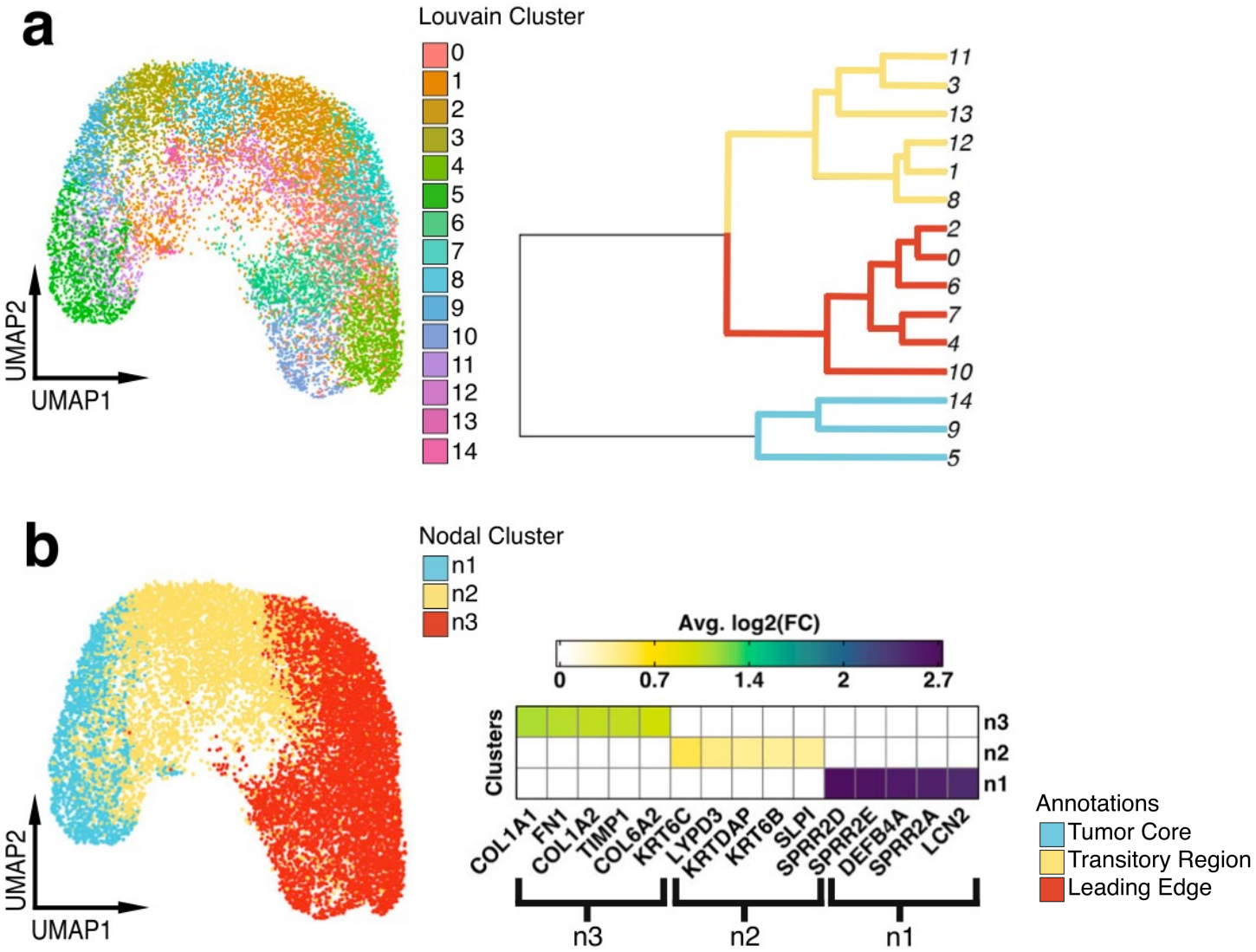
SRA Run Selector

Raw data are available in SRA

Processed data provided as supplementary file

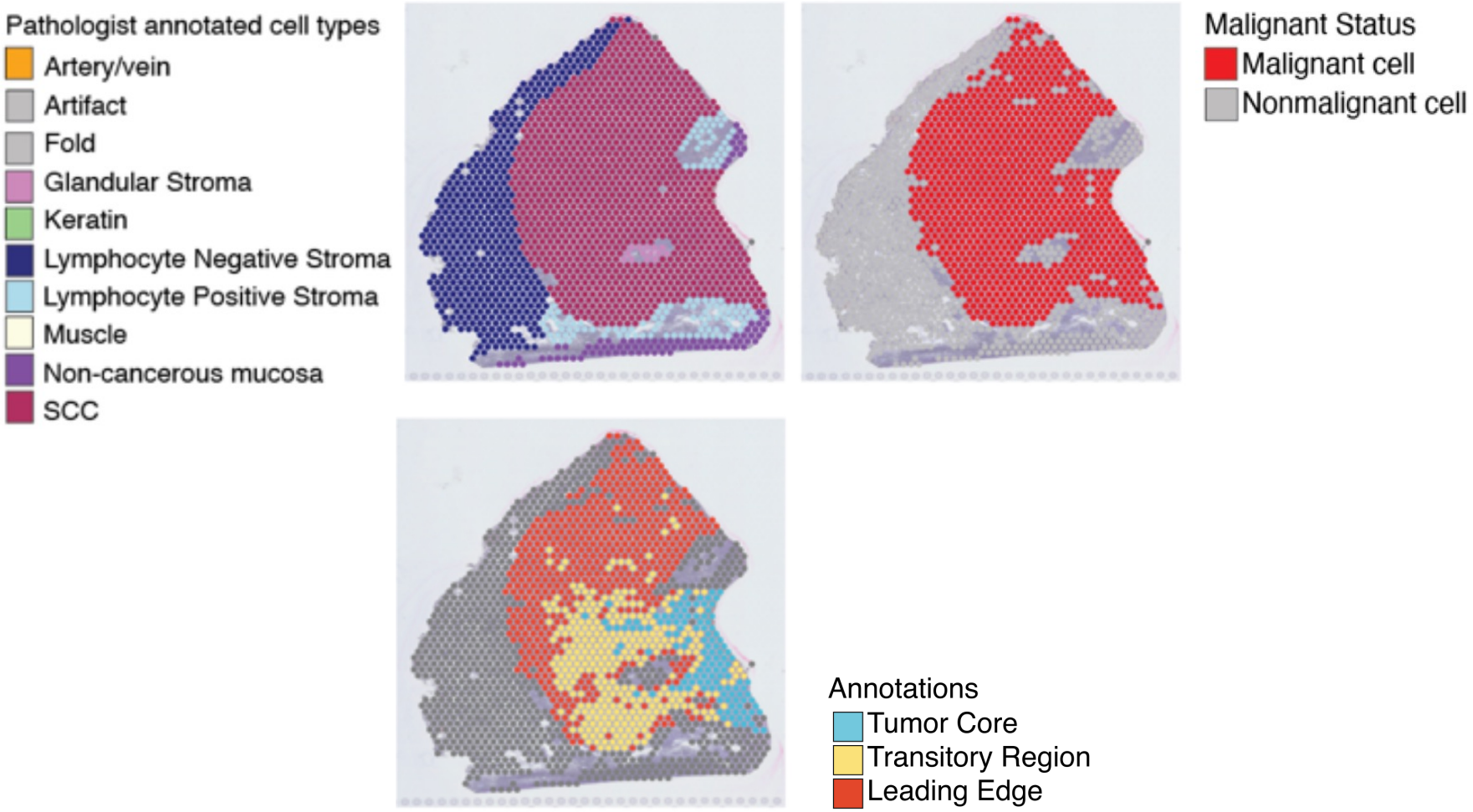
Source : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6339632

Aggregated Inspection on the 12 Samples



Source : Arora, R., et al. (2023). https://doi.org/10.1038/s41467-023-40271-4

Individual Sample Annotation



Sample to Insight

Source : Arora, R., et al. (2023). https://doi.org/10.1038/s41467-023-40271-4

Differential Expressed Genes

	A	B	C	D	E	F	G
1	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	group	gene
157	0	2.706423338	1	0.952	0	n1	SPRR2D
158	0	2.654346543	1	0.972	0	n1	SPRR2E
159	0	2.579461176	0.999	0.955	0	n1	DEFB4A
160	0	2.520689273	1	0.975	0	n1	SPRR2A
161	0	2.449394678	0.999	0.879	0	n1	LCN2
162	0						
163	0						
164	0						
165	0						
166	0						

	A	B	C	D	E	F	G
1	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	group	gene
122	0	0.535880319	0.999	0.997	0	n2	KRT6C
123	0	0.424765471	0.999	0.994	0	n2	LYPD3
124	0	0.392201584	1	0.99	0	n2	KRTDAP
125	0	0.371321607	1	1	0	n2	KRT6B
126	0	0.369927422	0.998	0.986	0	n2	SLPI
127	0						
128	0						
129	0						
130	3.26E-303						
131	2.67E-277						

	A	B	C	D	E	F	G
1	p_val	avg_log2FC	pct.1	pct.2	p_val_adj	group	gene
2	0	1.119188557	0.968	0.952	0	n3	COL1A1
3	0	1.077456695	0.954	0.923	0	n3	FN1
4	0	1.047688083	0.952	0.928	0	n3	COL1A2
5	0	1.040462139	0.983	0.954	0	n3	TIMP1
6	0	0.98234798	0.941	0.877	0	n3	COL6A2
7	0	0.916656628	0.944	0.918	0	n3	COL3A1
8	0	0.902377431	0.966	0.945	0	n3	SPARC
9	0	0.87526439	0.996	0.994	0	n3	VIM
10	0	0.791542974	0.973	0.922	0	n3	HMGN2
11	0	0.778830913	0.986	0.974	0	n3	TNC

Annotations
- Tumor Core
- Transitory Region
- Leading Edge

Functional analysis

Source : Arora, R., et al. (2023). https://doi.org/10.1038/s41467-023-40271-4

For more product details, please refer to :

QIAGEN CLC Genomics Workbench

https://rb.gy/1pnrsj

QIAGEN CLC Genomics Workbench Premium

https://rb.gy/wu3jtd

業務專員

Cleo Hsi 席佩妤

EMAIL | CleoHsi@GGA.ASIA

TELE | 02-2795-1777 # 3014

技術服務

Nereus Chang 張嘉瑋

EMAIL | NereusChang@GGA.ASIA

TELE | 02-2795-1777 # 3027

Thank You for Your Attention