NGS High Throughput Genomics Core at BRCAS 新世代基因體定序核心實驗室

基因組學平台解析：次世代定序、單細胞空間技術與三維基因組學

Mei-Yeh Lu 呂美曄

High Throughput Genomics Core

BRC, Academia Sinica

【生物資訊系列】NGS Workshop 1, 20251104

NGS High Throughput Genomics Core at BRCAS 新世代基因體定序核心實驗室

About ▾ | Core News | Equipment ▾ | Publications | Service Flow ▾ | Services & Prices ▾ | Documents | Contact Us | FAQs

Services & Prices

Home / Services & Prices

Sample QC

- Quantification
- Electropherogram
- Pippin Size Selection
- Purification

Illumina System

- Illumina Library Prep
- Illumina HiSeq 2500 Sequencing
- Illumina MiSeq Sequencing
- Illumina NextSeq 2000 Sequencing
- Illumina iSeq 100 Sequencing

PacBio System

- PacBio Sequel Library Prep
- PacBio Sequel Sequencing

Oxford Nanopore System

- ONT Library Prep
- GridION Sequencing

10x Genomics

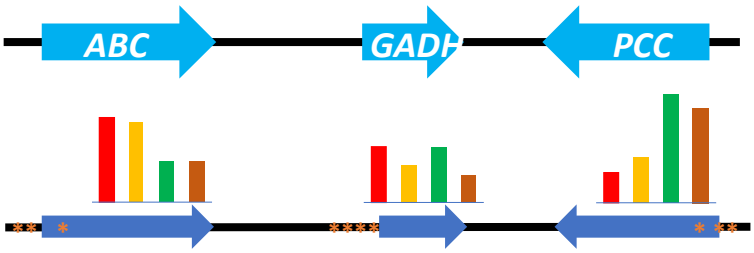
- 10x Genomics – Spatial Transcriptome (Visium)
- 10x Genomics – Single Cell (VDJ, ATAC)

Genomes –big & small

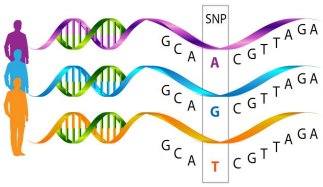
NGS High Throughput Genomics Core at BRCAS

What to learn from meta/genome/transcriptome?

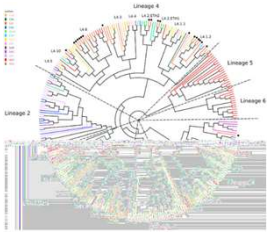
1. Assembly
2. Gene prediction
3. Functional annotation
4. Expression profiling
5. Epigenetic regulation



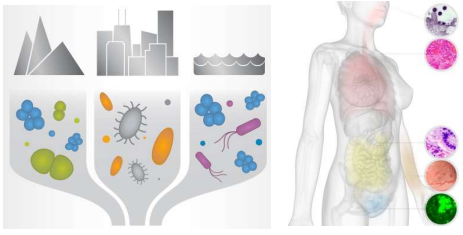
6. Genome variants



7. Population Cohort



8. Metagenome/Community



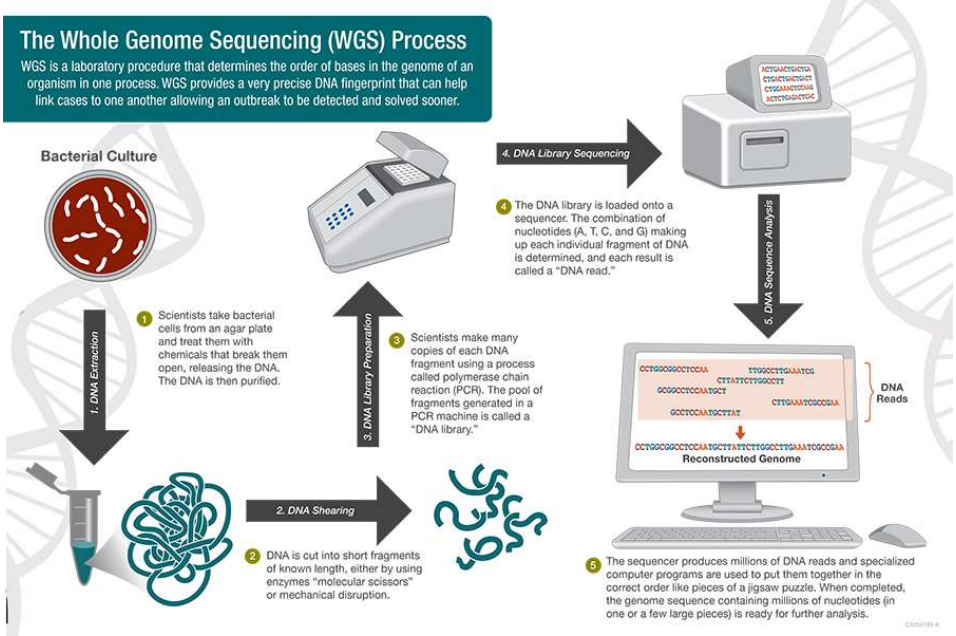
Genomics: important considerations

1. **Defining the research goals:** refine your research questions to ensure the chosen sequencing approach delivers the necessary insights.
2. **Sample preparation recommendations:** Whether it's DNA, RNA, or tissue, find optimal prep methods to maximize data quality and yield.
3. **Platform selection:** With Illumina, PacBio, and Nanopore, choose the platform that best fits your budget, desired read length, and target genome complexity.
4. **Experimental design:** Optimize your library prep, sequencing depth, and replication strategy for reliable and statistically robust results.

The Incredibles (2004)

* Advanced applications*

1. **3D genome analysis:** Delve into the spatial organization of chromosomes and explore higher-order genome structure for a deeper understanding of gene regulation and function.
2. **Single-cell RNA-seq:** Unravel cellular heterogeneity and explore gene expression at the individual cell level, revealing hidden cell populations and dynamic transcriptional landscapes.
3. **Spatial transcriptomics:** Map gene expression patterns across tissues and organs, providing a powerful tool for studying development, disease, and complex biological processes.



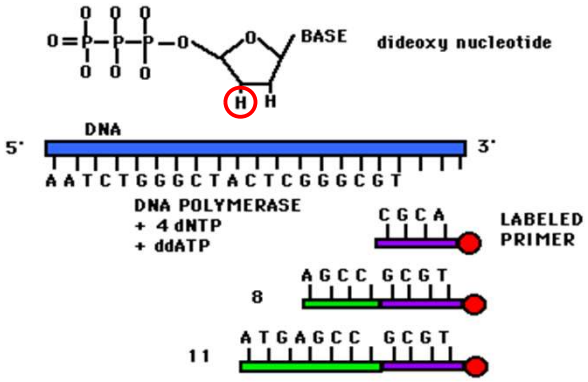
https://www.cdc.gov/pulsenet/pathogens/protocol-images.html#wgs

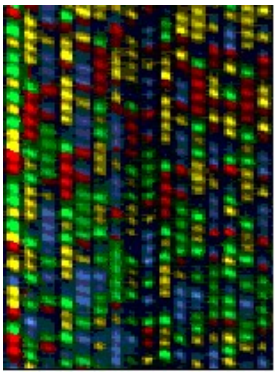
Section I.
Genomics Technology Overview

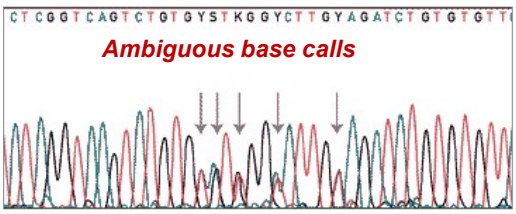
Sanger Seq: chain termination w/ fluorescent ddNTPs

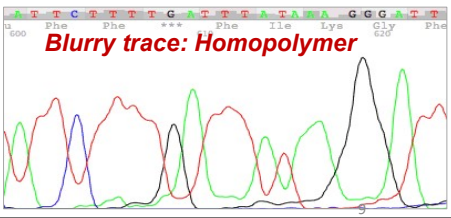
Frederick Sanger

Nobel laureate , 1977









The Evolution of Sequencing:
From Sanger to Short-Read to Long-Read Major NGS & TGS Platforms

Sanger
1970s,
chain termination

Short-read revolution

Illumina
Reversible terminator

New generation short-read innovation

ElementBio
Aviti24

Long-read resolution

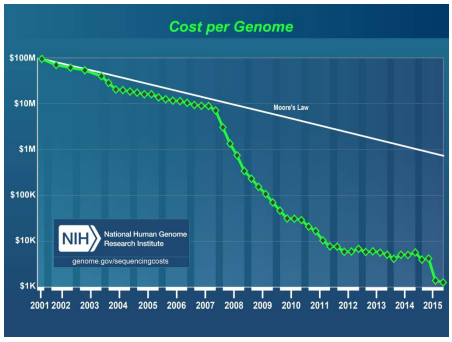
PacBio SMRT HiFi

Oxford NanoPore

10

Now– Large scale NGS sequencing

Cost Dropping

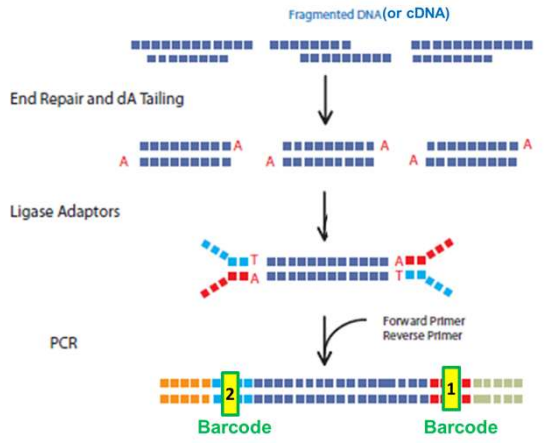


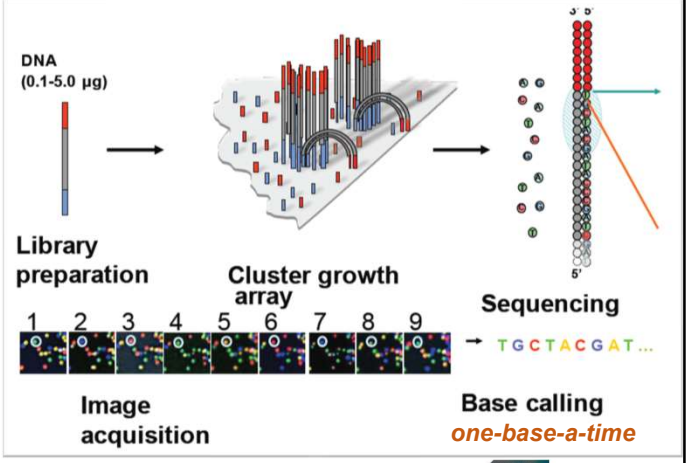
Short Read

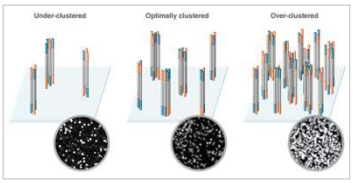
Automation

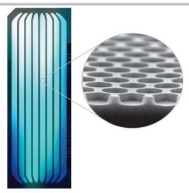
Long Read

Workflow for Illumina: bridge amplify, cyclic reversible seq

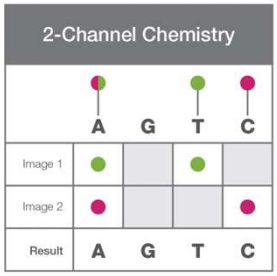


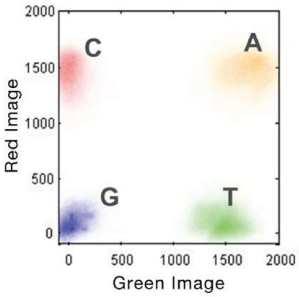


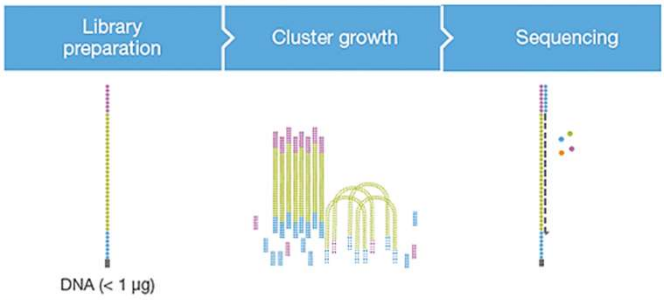


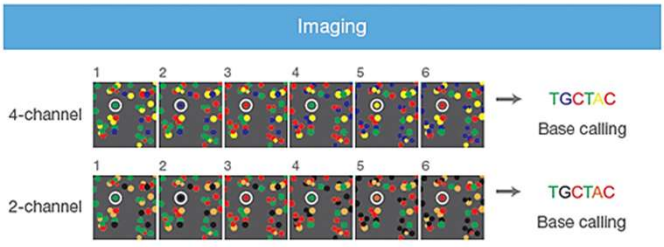


Workflow for Illumina: bridge amplify, cyclic reversible seq





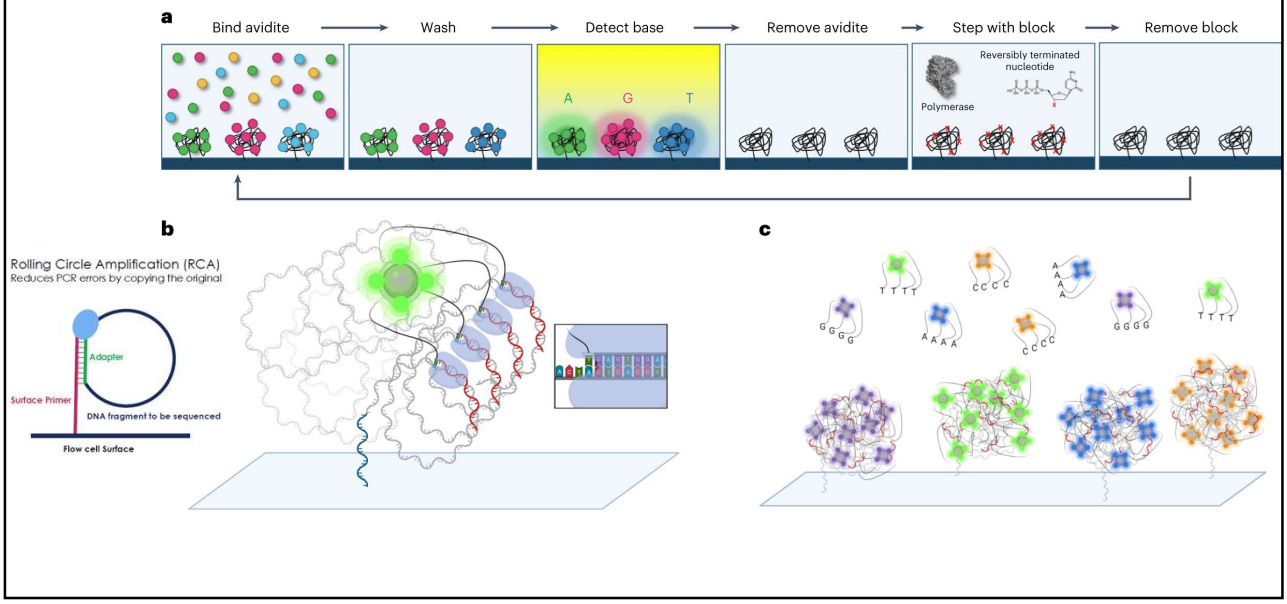


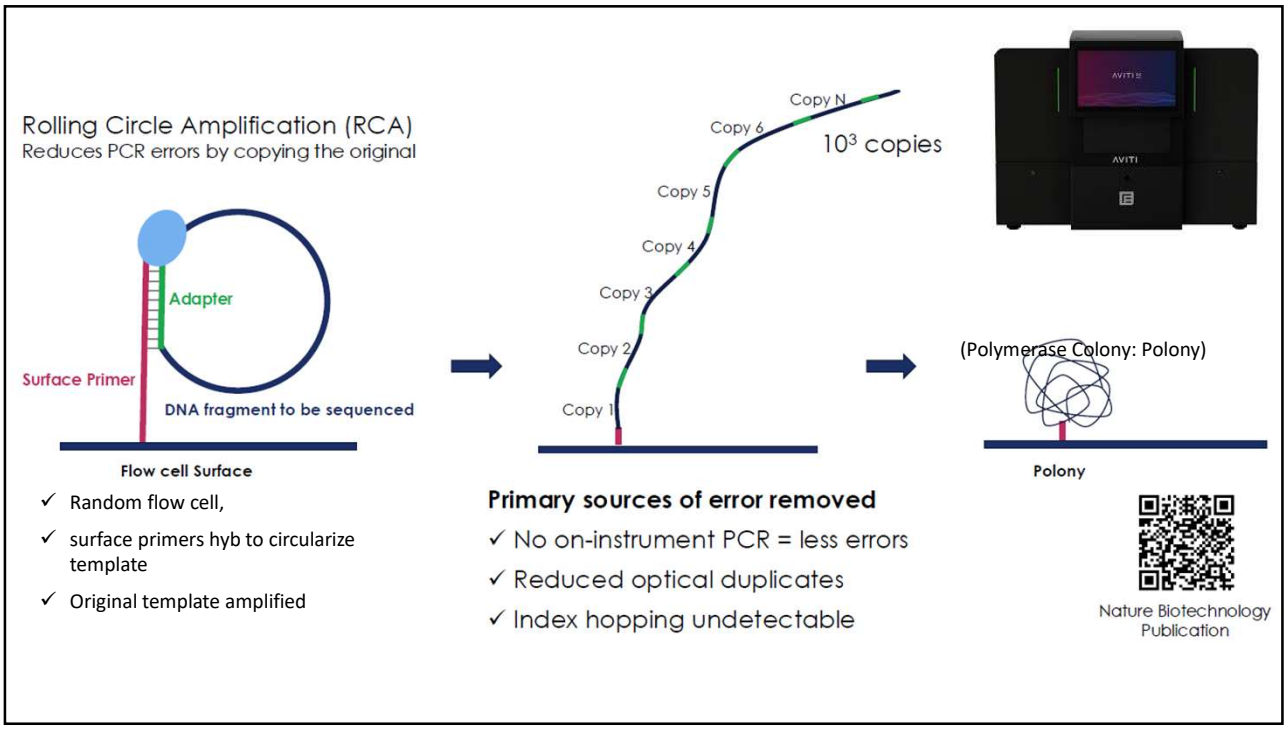


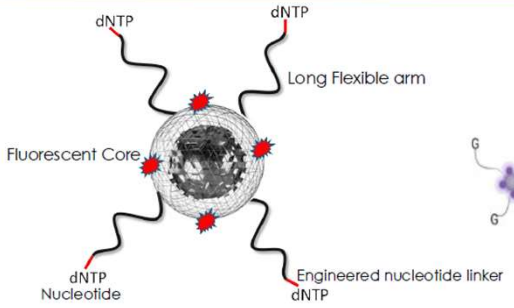
Accuracy may drop at high GC% or poly-G regions.

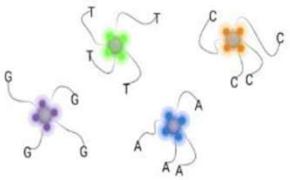
https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html

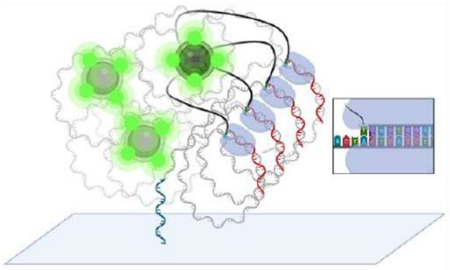
Element Biosience Avidites (Multi-Valent Nucleotides); Higher-intensity signal



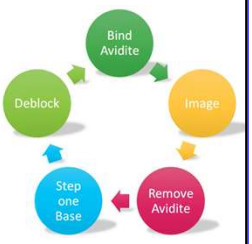








- **Avidites 是一種螢光標記的聚合物；四種Avidite顏色螢光代表四種核苷酸 (A, T, C, G)**
- **Avidite 有獨特的“章魚”多觸手分，每個「觸手」末端上攜帶相同的核苷酸**
- **每個Avidite上的觸手透過可逆性與的template 相對位點結合**
- **「偵測」: Sequencing primer (red) 加入後，特殊設計的聚合酶會把Avidite 加入到 template下一個位置 Avidite上的多個觸手能同時與 polony中的多個聚合酶-DNA複合體結合，雷射激發後產生的螢光信號**
- **價格競爭力：SBS高成本因為需使用高濃度的螢光標記物**



Aviti24: 5D multi-omics
(RNA, protein, morphology, spatial resolution, and dynamic response)

Teton™: Library Prep-Free, System's Biology Multi-Omic Assay

Library prep-free biological studies with 1 day TAT and Hands on time < 45 minutes

30-minute hands on time — 24-hour experiment turnaround time

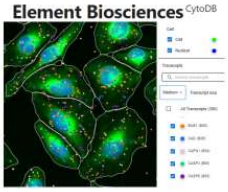
Step 1: Culture/Treat

Step 2: Wash/Fix/Perm — Step 3: Assemble

AVITI24 (Load FC and Cartridge)

Analysis

Element Biosciences CytoDB



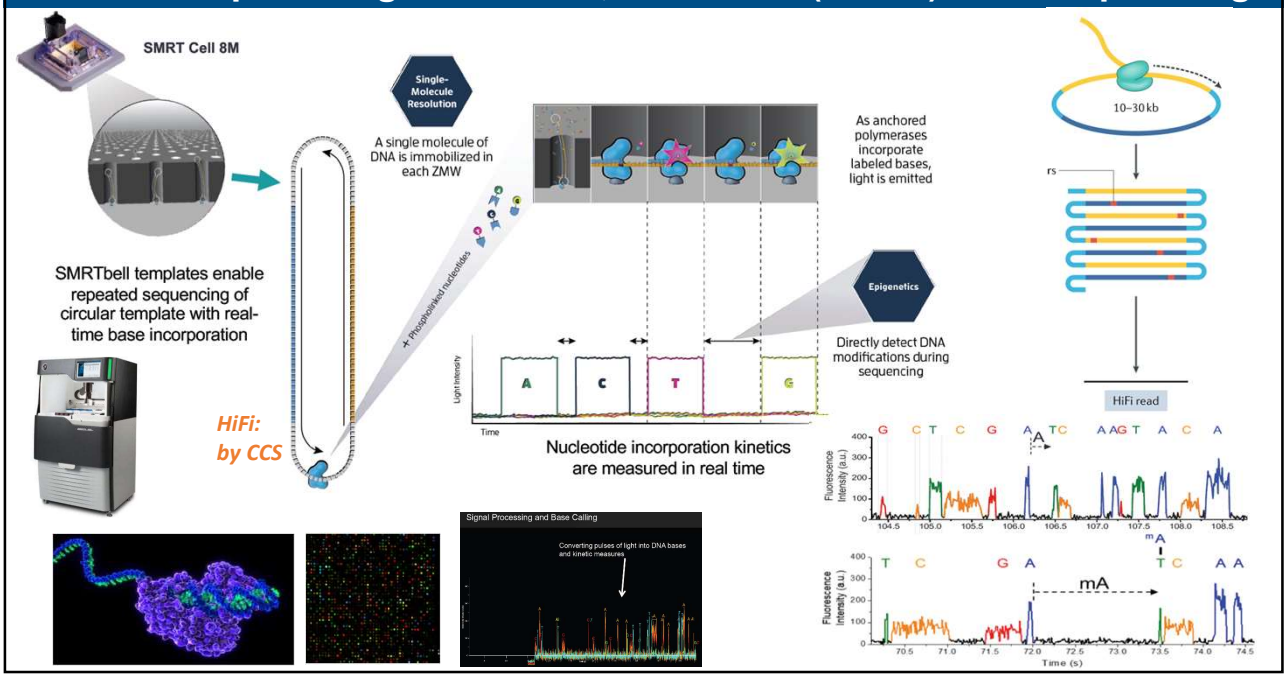
Teton Flow Cell
Step 1: Cell Culture/Capture
Step 2: Wash/Fix/Perm
Step 3: Assemble

Teton Reagent Cartridge
- Cell paint, RNA, Protein
- Phosphorylation detection
- In Vitro ABC 3D Sequencing

Elembio Cloud
- AVITI24 performs analysis and quant.
- Open data format

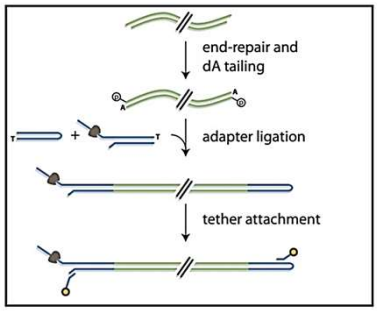
https://omicsomics.blogspot.com/2024/02/post-agbt-both-element-singular-want.html

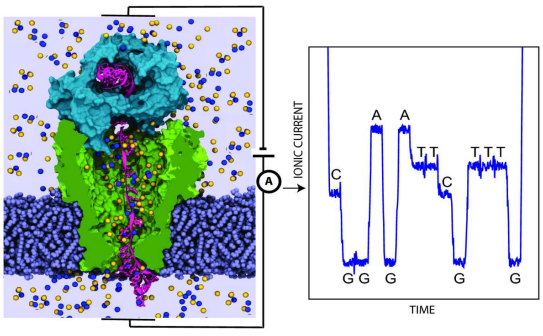
PacBio Sequel: Single Molecule, Real-Time (SMRT) DNA Sequencing



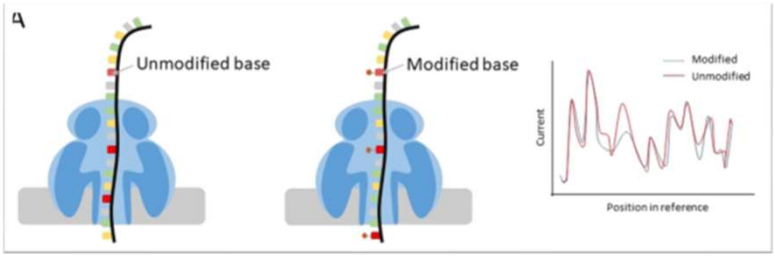
Oxford NanoPore Sequencing

GrinION



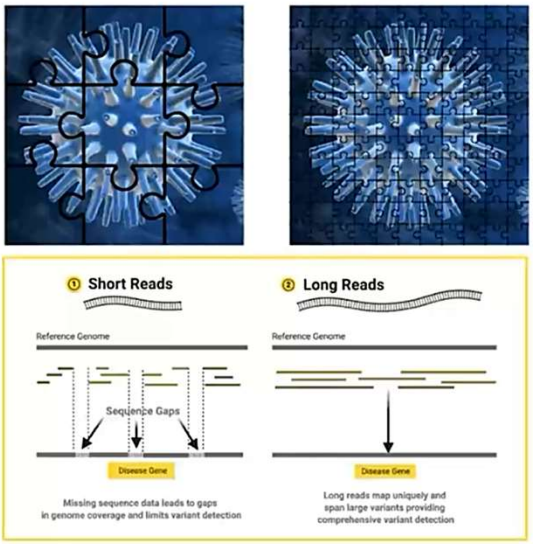


Ebola outbreak, 2016/2
Zika outbreak, 2016/6



Differences Between Long and Short Read Sequencing

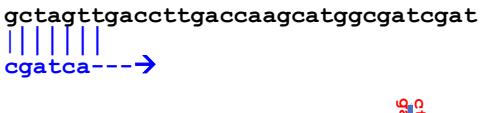
- Short read sequencing requires DNA fragmentation and amplification with subsequent bioinformatic assembly
  - Read Length: 10s-100s of base pairs
  - Sequence gaps, sample bias
  - Examples: Illumina, IonTorrent – massively parallel sequencing by synthesis
- Long read sequencing enables the sequencing of long DNA strands in a single continuous process
  - Read Length: 1,000s-100,000s of base pairs
  - Large span, comprehensive variant detection
  - Works in real time – adjust experiment at any stage
  - Examples:
    - PacBio - High fidelity (HiFi) circular consensus sequencing – rolling loop sequencing of circular DNA molecule by fluorescent signature
    - Oxford Nanopore – DNA sequence deciphered by current disruption signature as it passes through a nanopore

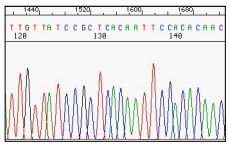


Golden Helix. https://www.goldenhelix.com/resources/webcasts/integrating-long-and-short-read-sequencing-for-comprehensive-NGS-analysis/

Evolution of Sequencing Technologies

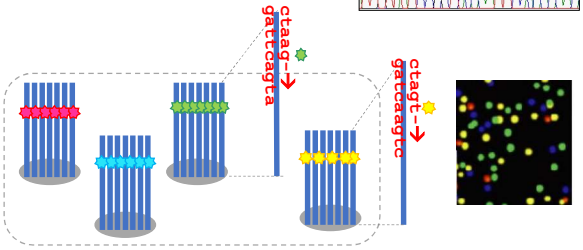
Dideoxy nt.
1 rxn/tube





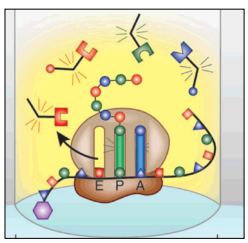
Sanger seq.
(up to 1kb)

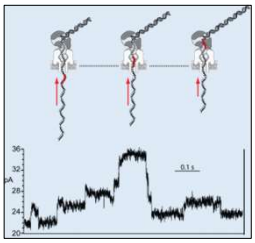
2nd-Gen
Clonal amplification



Illumina
ElemBiosc
(50-1000bp)

3rd-Gen
Single mol. Seq.





PacBio,
Oxford
Nanopore
(0.4-200kb)

Advancements in HTS technologies

Roche 454 GS +

HS2500

NextSeq2000

ChromiumX

CytAssist

2008 2013 2014 2016 2019 2020 2022 2023

GA IIx

MiSeq

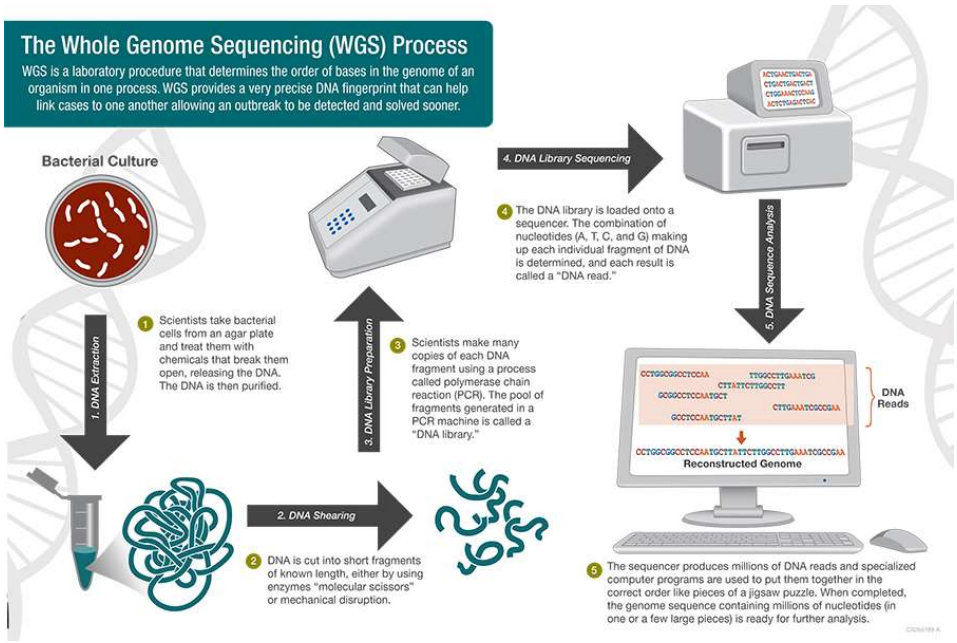
PacBio Sequel

10x Controller

10x Visium

Sequel IIe

Section Ii.
Applications of NGS

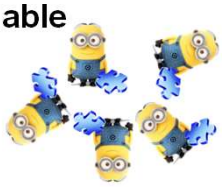


https://www.cdc.gov/pulsenet/pathogens/protocol-images.html#wgs

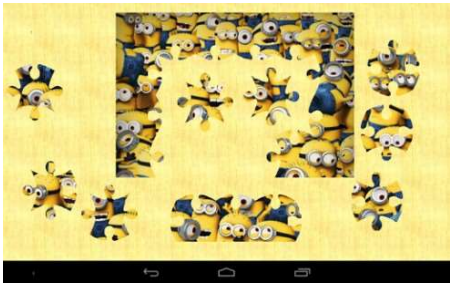
Genome: *De novo* seq. vs re-sequencing

Assembly

***"de novo" sequencing*:**
no reference geneome available



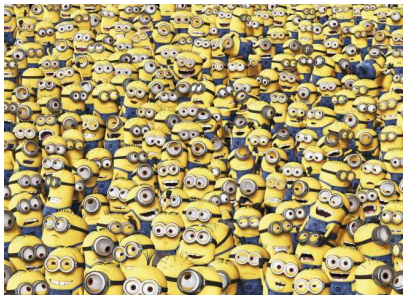
assemble



Mapping

***Re-sequencing*:**
reference genome available

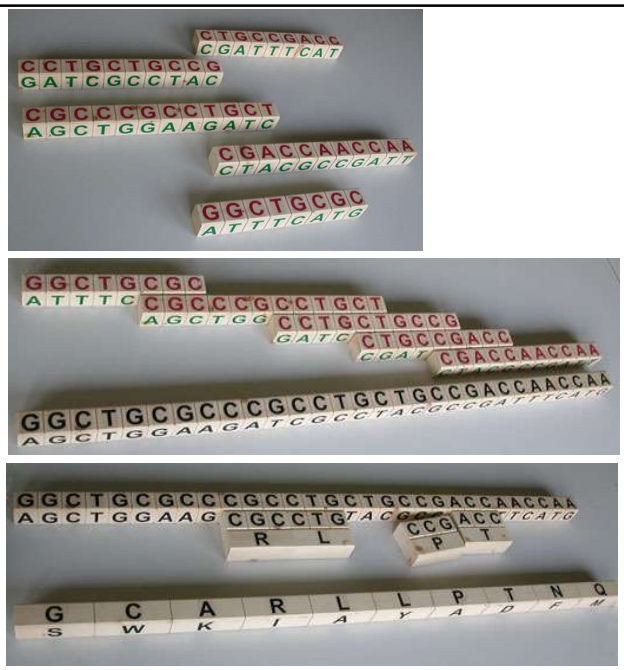
align



Assembly

Align

Annotate



De novo?
Ref-guided?

http://www.scienceinschool.org/2007/issue5/dnapuzzle

Re-sequencing: Variant detection

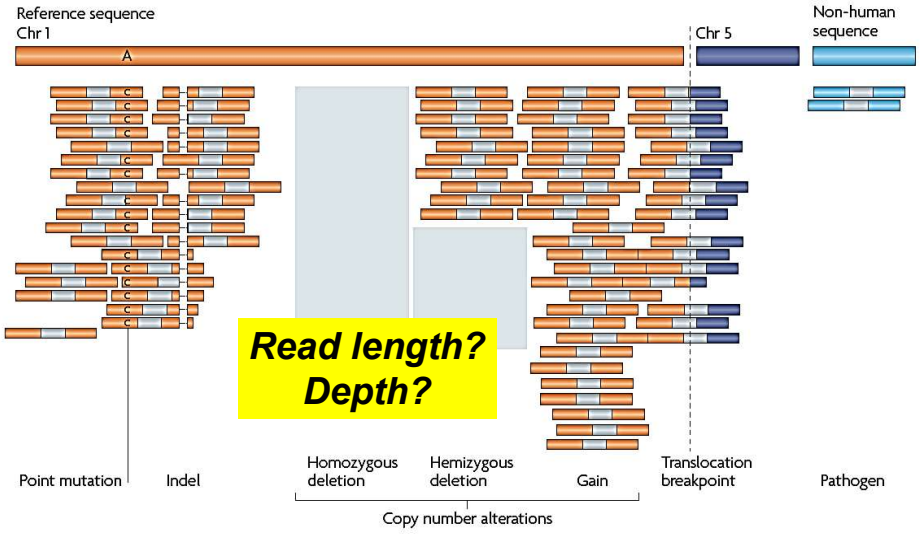


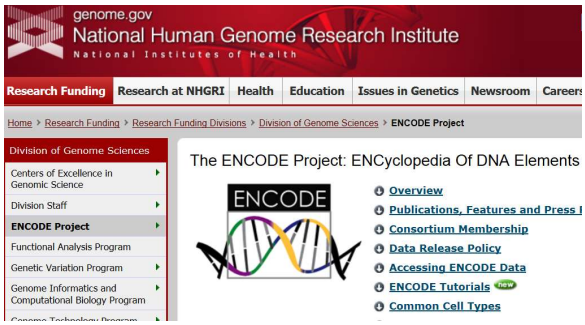
Figure 3 | **Types of genome alterations that can be detected by second-generation sequencing.** Sequenced

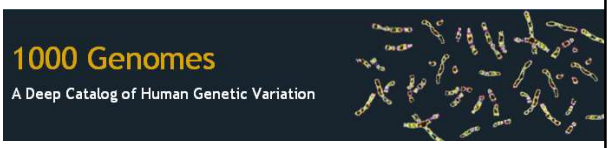
Meyerson et al. NRG 2010

Public NGS Databases

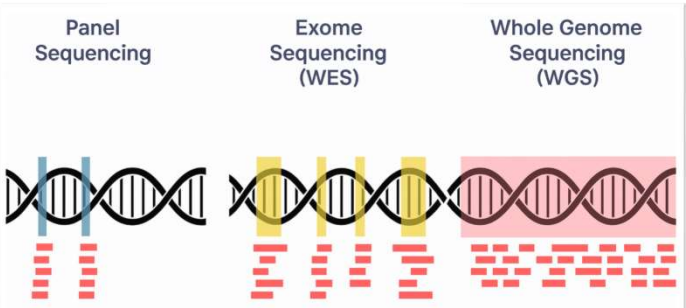
NIH Short Read Archive

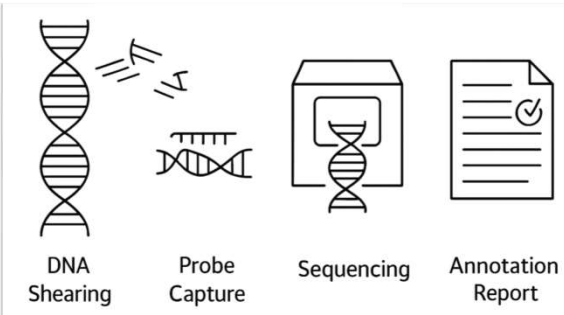
ERA (European Short Read Archive)





Variant Discovery -
Whole Genome (WGS), Exome (WES), and Targeted Panels

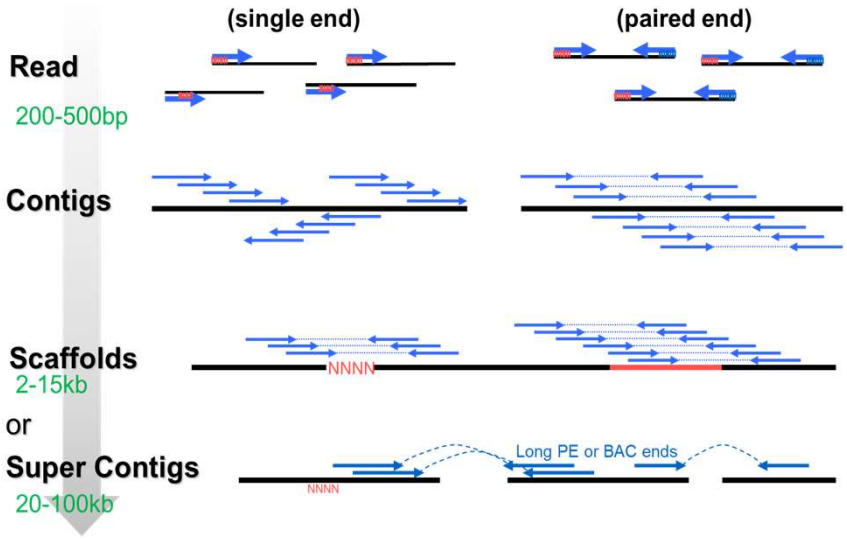




- **WGS: comprehensive, unbiased**
- **WES: protein-coding regions (~1%)**
- **Targeted panels: focused and cost-efficient**

- 基因定序技術比較：從擴增子到全基因組
- 三種定序策略的範圍與應用差異
- 從目標區域到全基因組：定序解析度的進化
- **Amplicon、WES 與 WGS**：基因定序的不同層次
- 基因定序的選擇：精準度與範圍的取捨

De novo Genome Assembly - Hierarchical

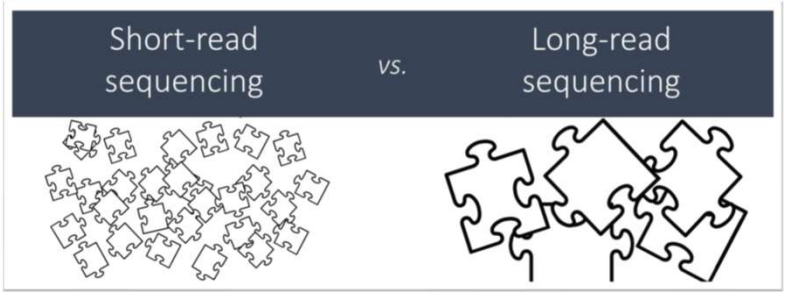


Building Complete Genomes

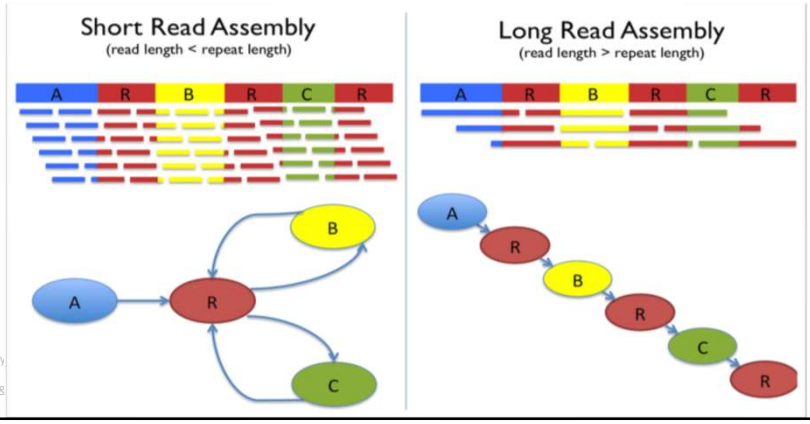
- Short-read: fragmented assembly
- Long-read: improved continuity
- Hybrid assembly: accuracy + contiguity

30

拼圖

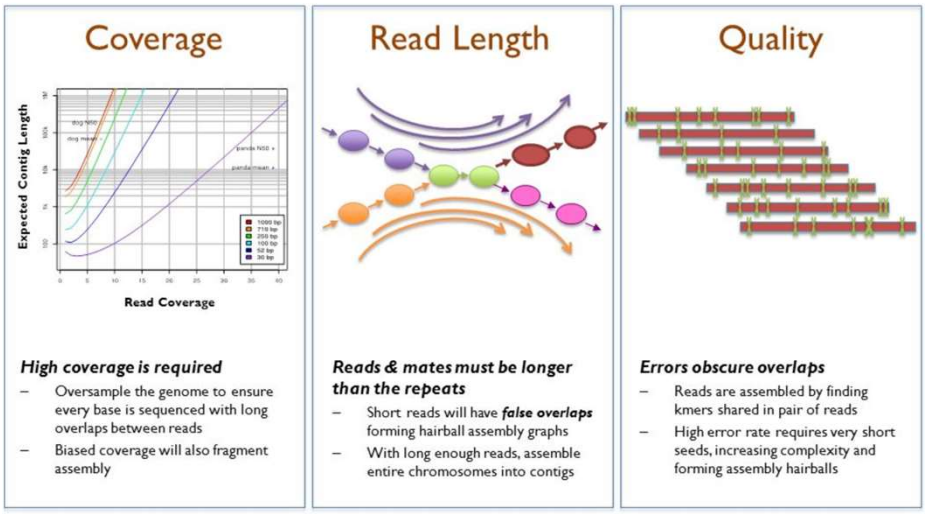


字串接龍



https://media.licdn.com/dms/image/v2/D4D12AQEviW_k5TFy
cover_image-shrink_720_1280/article-cover_image-
shrink_720_1280/0/1658514975459?e=2147483647&v=beta&
X0IrHdVU6Y9yD41HxCUMU_UjzceQgo

2. Ingredients for a good assembly

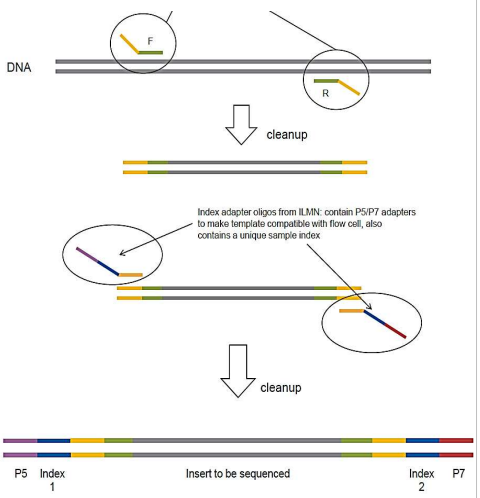


Current challenges in *de novo* plant genome sequencing and assembly
Schatz MC, Witkowski, McCombie, WR (2012) *Genome Biology*. 12:243

Whole Genome Assembly with iPlant
Published by Timothy McGee (https://slideplayer.com/slide/9388705/)

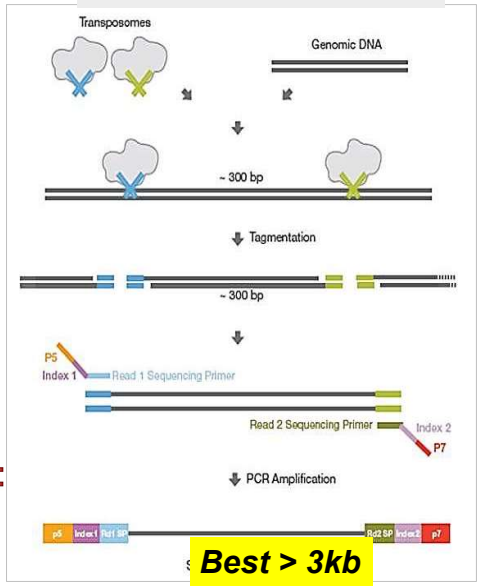
Amplicon sequencing

Indexing PCR



250-600bp

Nextera Tagmentation

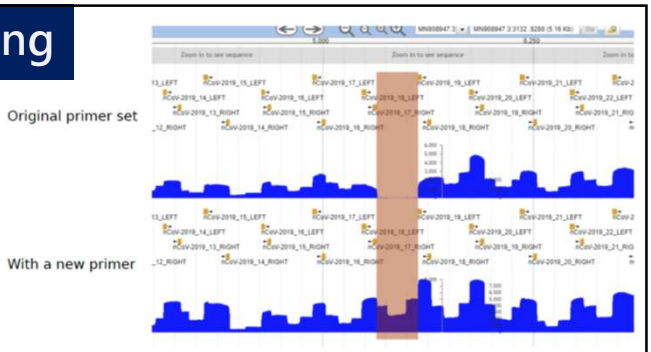


Best > 3kb

COVID-19: amplicon sequencing

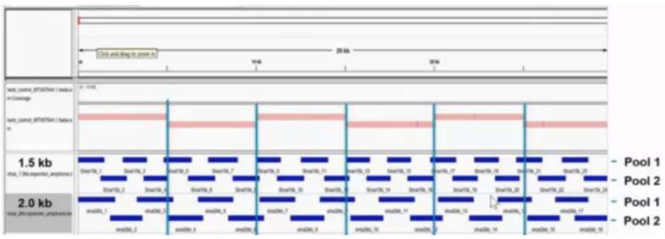
Artic Network:

- 98 PCR primer pairs for 300-400bp amplicons
- Tiling over the 30kb Sars-Cov-2 RNA genome



Long amplicon with 3rd-Gen:

- 1k, 1.5k, 2.5k, etc



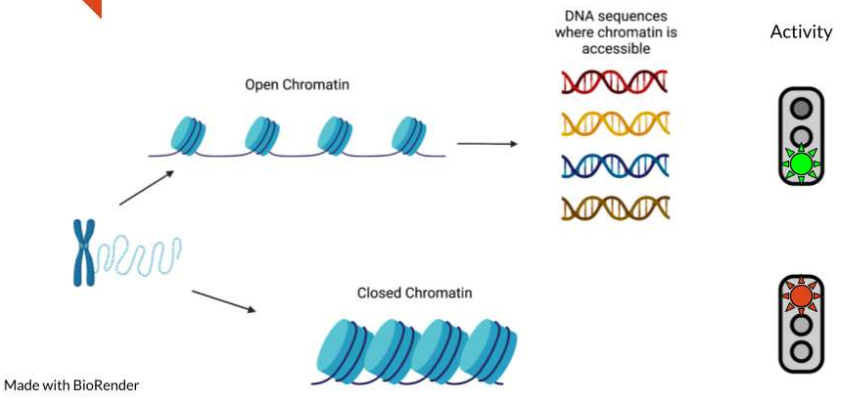
ARTIC primer set: https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V1)
Quick, J, 2020. nCoV-2019 sequencing protocol, protocol.io
dx.doi.org/10.17504/protocols.io.bbmuik6w
https://www.biorxiv.org/content/10.1101/2020.03.15.992818v1.full.pdf

Epigenetic sequencing

- *ChIP-seq*
- *Histone-IP*
- *Bisulfite sequencing*

EpiGenome: 1. ATAC-seq

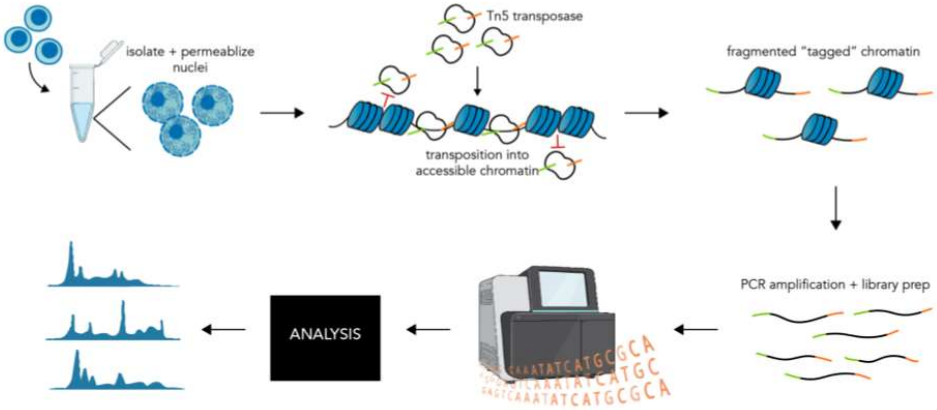
What does accessibility to chromatin represent?



ATAC-seq is a method used to study chromatin accessibility across the genome:
- It can identify which regions of the genome have accessible chromatin.
- The technique can be used to compare chromatin accessibility between different biological samples or over time.
- It helps in finding transcription factor motifs and footprints in accessible regions of interest.

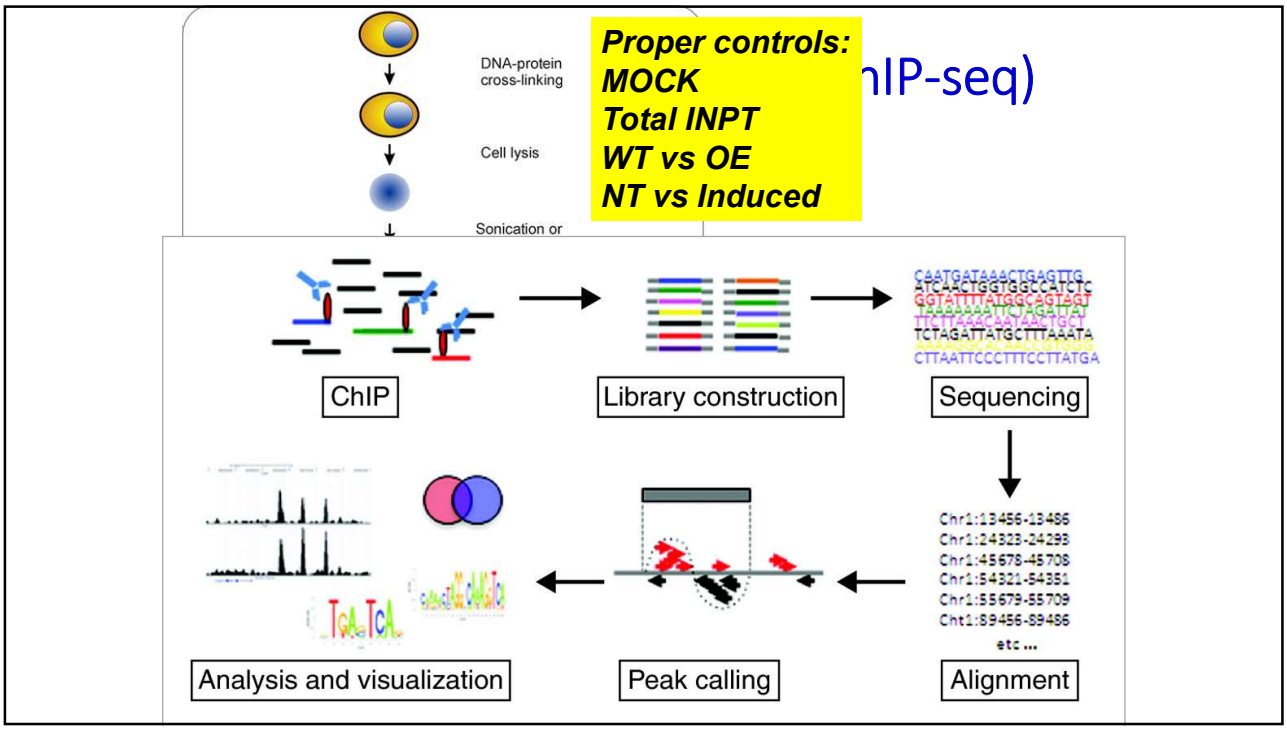
https://hutchdatascience.org/Choosing_Genomics_Tools/atac-seq-1.html

Overview of ATAC-seq



- ATAC-seq data is generated by fragmenting the genome with the Tn5 endonuclease and sequencing the shorter DNA fragments. While most of the genome is associated with protein complexes that preclude the digestion of DNA by Tn5, some regions of the genome have accessible chromatin that can be cleaved by Tn5 resulting in short (<500bp) fragments.
- These regions of the genome are of biological interest as they are likely to harbor transcription factor binding sites and to constitute cis-regulatory elements, genomic regions that are involved in the regulation of gene expression.

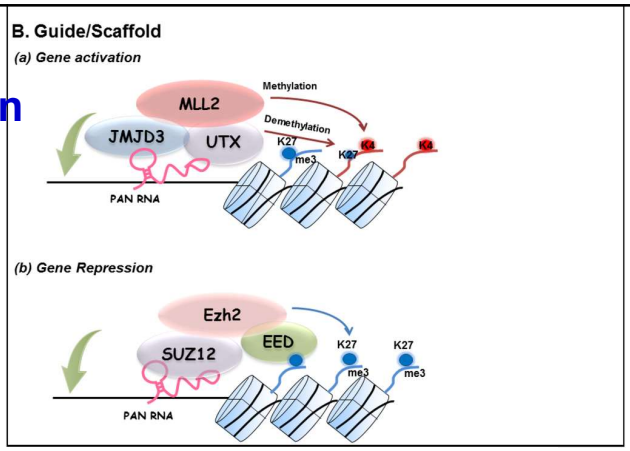
https://hutchdatascience.org/Choosing_Genomics_Tools/atac-seq-1.html

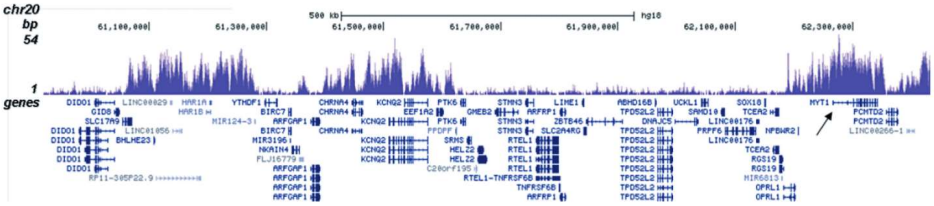


Proper controls:
MOCK
Total INPT
WT vs OE
NT vs Induced

Histone methylation

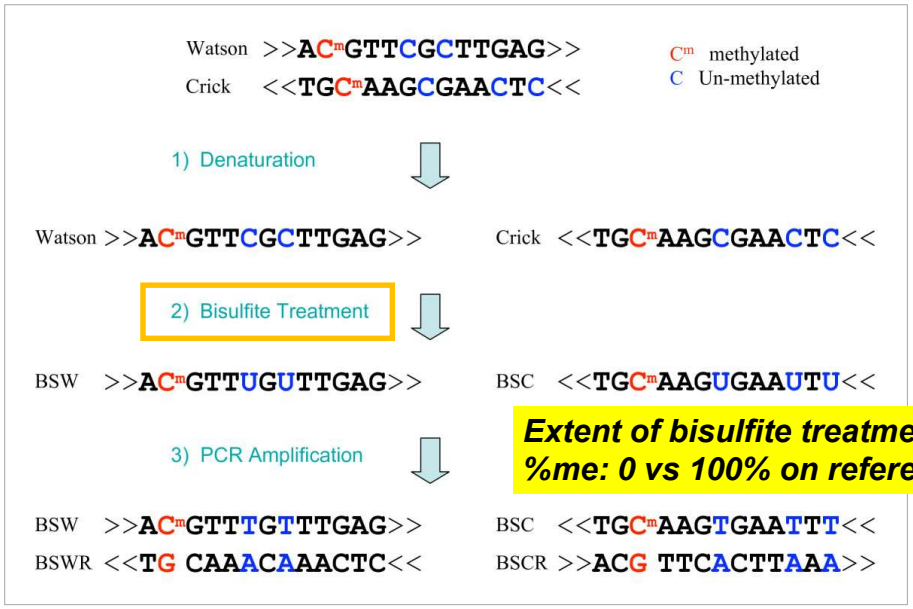
Anti-K27me3





http://www.mdpi.com/1999-4915/6/11/4165/htm

DNA methylation: Bisulfite sequencing

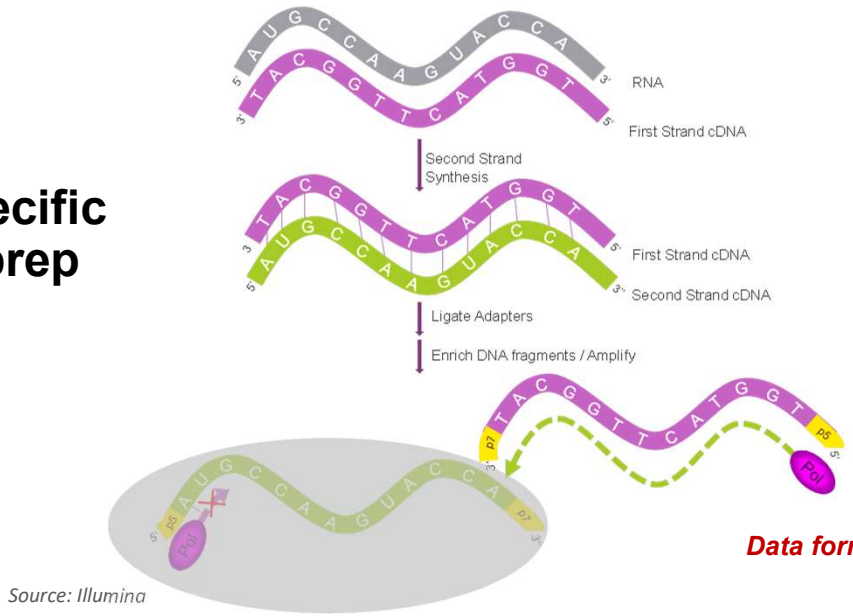


Extent of bisulfite treatment %me: 0 vs 100% on reference

http://www.biomedcentral.com/content/figures/1471-2105-10-232-1-l.jpg

Transcriptome –RNA-seq (short-read)

Strand-specific RNA-seq prep



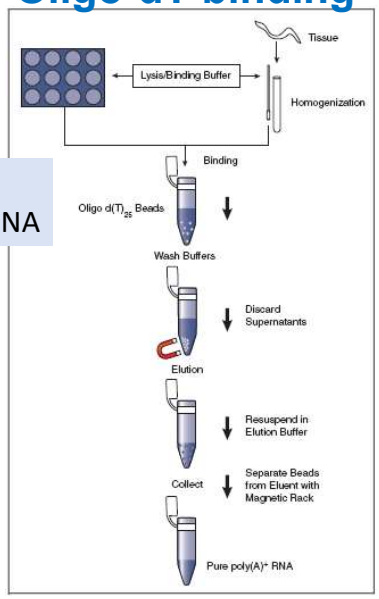
Source: Illumina

Data format: RF

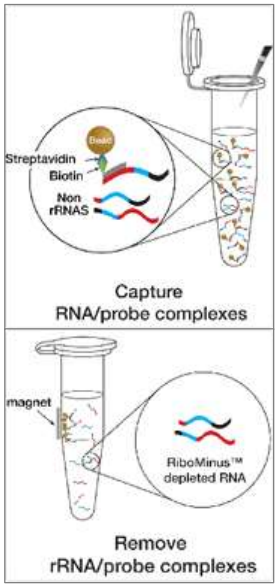
Methods of mRNA enrichment

Oligo-dT binding

Eukaryotes
High quality RNA



rRNA removal



Prokaryotes
Non-A-tailed RNA
Degraded RNA

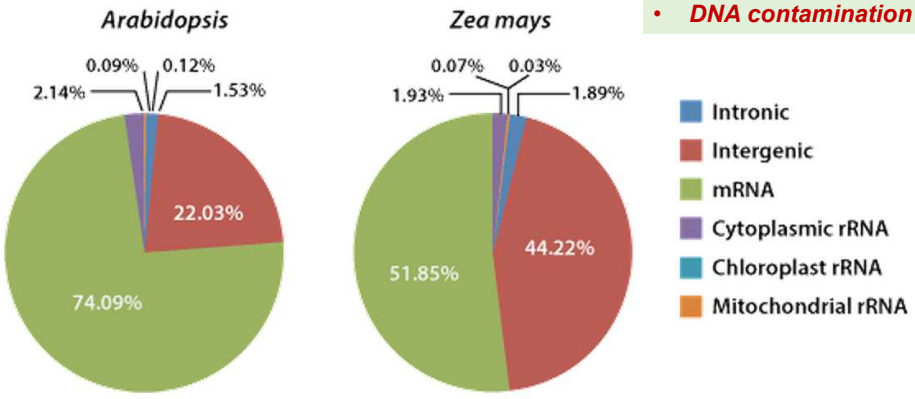
Example of RiboZero purification

LARGE-SCALE BIOLOGY ARTICLE

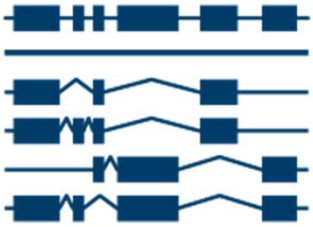
RNA Sequencing of Laser-Capture Microdissected Compartments of the Maize Kernel Identifies Regulatory Modules Associated with Endosperm Cell Differentiation OPEN

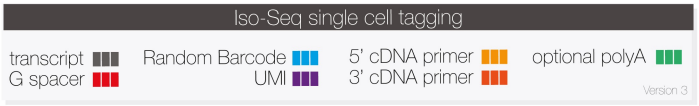
Junpeng Zhan,[a,1] Dhiraj Thakare,[a,1] Chuang Ma,[a,2] Alan Lloyd,[b] Neesha M. Nixon,[b] Angela M. Arakaki,[b] William J. Burnett,[b] Kyle O. Logan,[b] Dongfang Wang,[a,3] Xiangfeng Wang,[a,4] Gary N. Drews,[b] and Ramin Yadegari[a,5]
[a] School of Plant Sciences, University of Arizona, Tucson, Arizona 85721
[b] Department of Biology, University of Utah, Salt Lake City, Utah 84112

- ***Genome divergence***
- ***RNA integrity***
- ***DNA contamination***

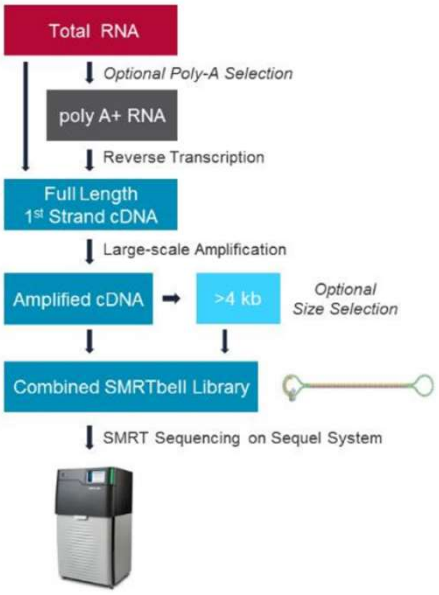


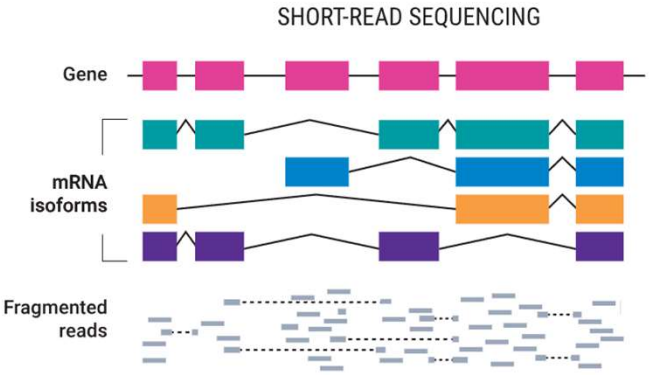
IsoSeq: Full-Length Isoform transcriptome

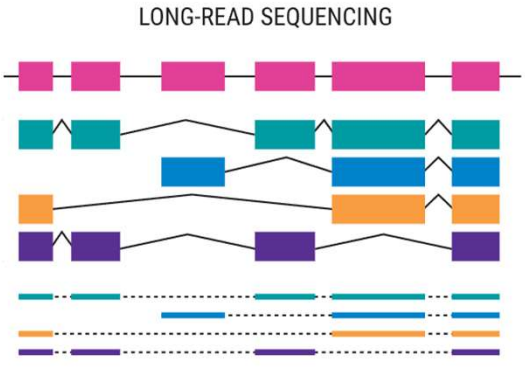




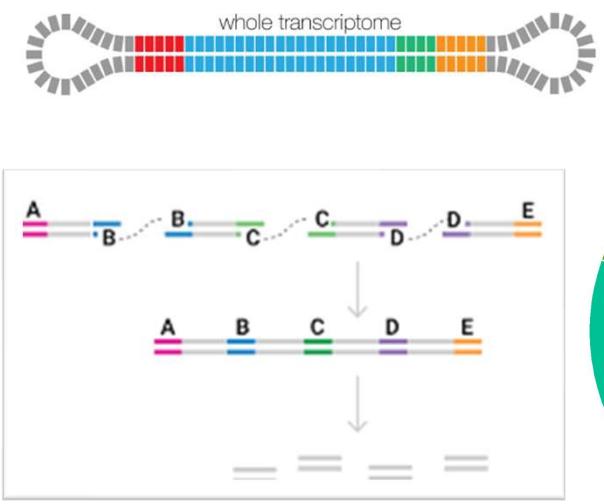
IsoSeq supports a modular tagging design. Possible examples:

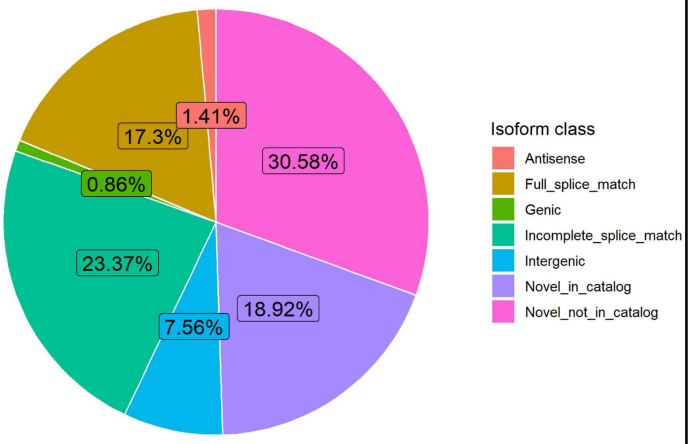






10x + PB: MAS-Sp-Isoform Sequencing





46

Chi-Chih Wu et.al., unpublished

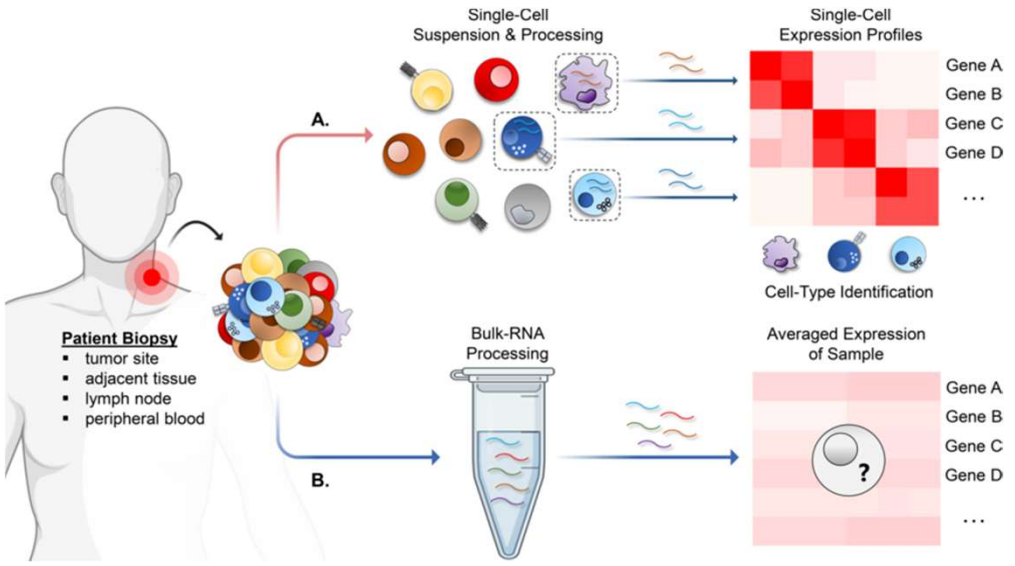
Section III.

High-Resolution Genomics:

Single-Cell & Spatial Technologies

47

scRNA-seq vs Bulk-seq

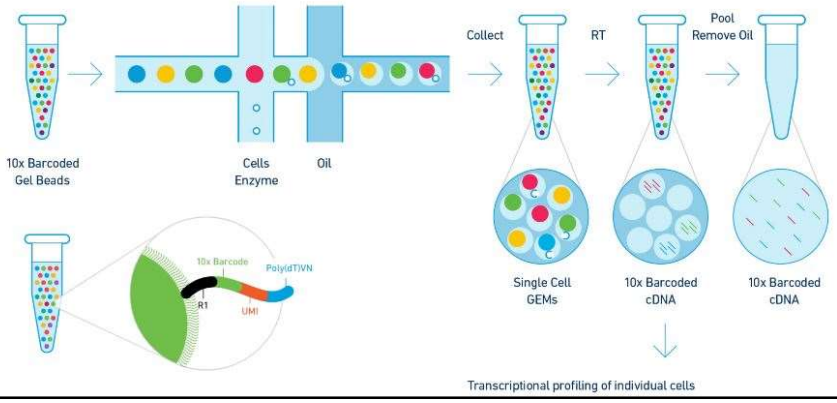


https://www.researchgate.net/figure/scRNA-seq-versus-bulk-RNA-seq-for-profiling-the-TME-A-and-B-Transcriptomic-studies-of_fig1_347823832

10x Genomics: Single-cell applications

Countess III

ChromiumX



10x Single Cell

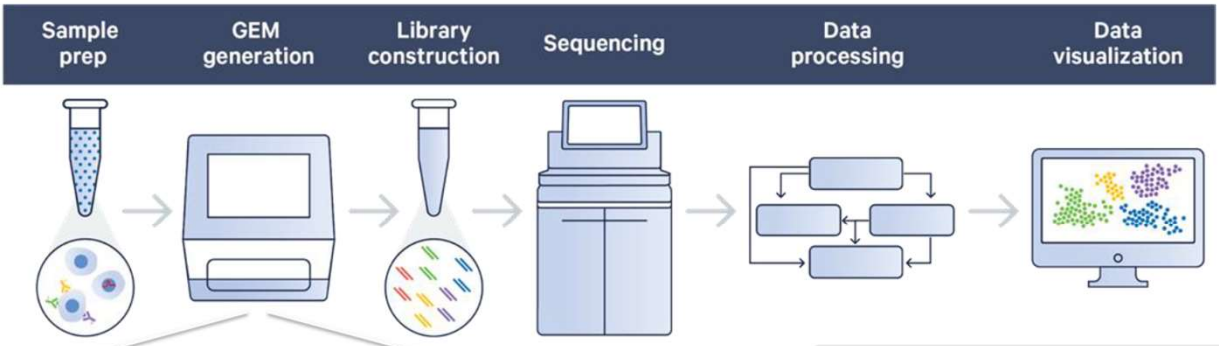
- 10x Chromium 3′ Single Cell RNA Prep
- Single Cell Chip G
- 10x Chromium VDJ 5′ Single Cell RNA Prep
- Single Cell Chip K
- 10x ATAC-seq prep (Chromatin accessibility)

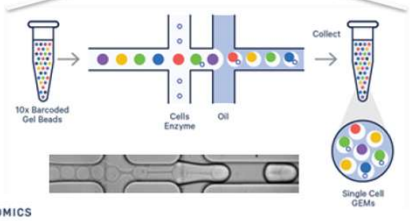
Chromium X

- 10X Chromium X Single Cell Multiome
- 10X Chromium X Single Cell ATAC

Streamlined Single Cell Workflows with Chromium

Next generation solutions on our tried-and-tested Chromium workflow





Multiomicoptions:

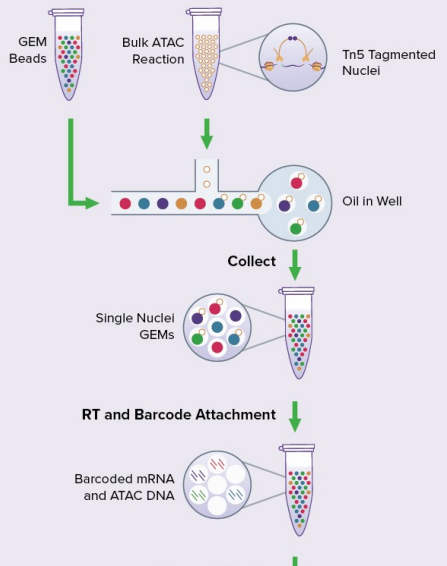
- 3’ and 5’ gene expression
- TCR/BCR
- Protein
- CRISPR
- 3’ gene expression
- Chromatin accessibility

1. Experiment planning & monitoring
2. Data processing & QC
3. Visualization & exploration

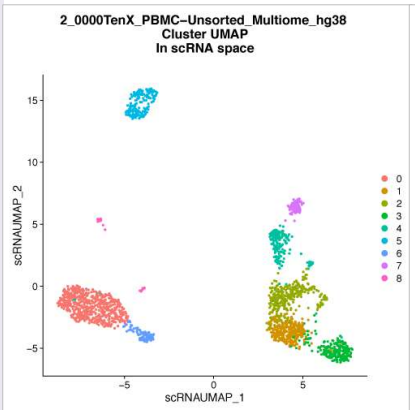
10X GENOMICS

10x sn-Mutiome: snRNA + snATAC

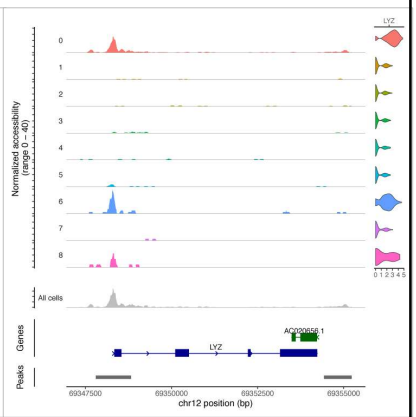
Single-cell gene expression & chromatin accessibility



https://www.activemotif.com/catalog/1355/single-cell-multiome-service



UMAP embedding of PBMC cells colored by multimodal cluster identity.



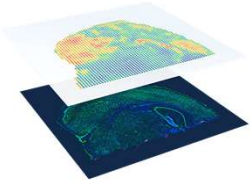
Expression and accessibility of the LYZ gene in PBMC clusters.

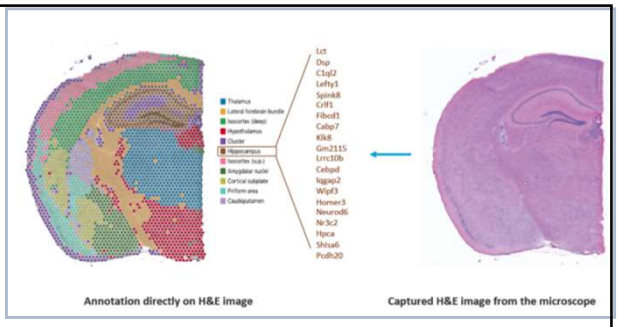
Spatial Transcriptome

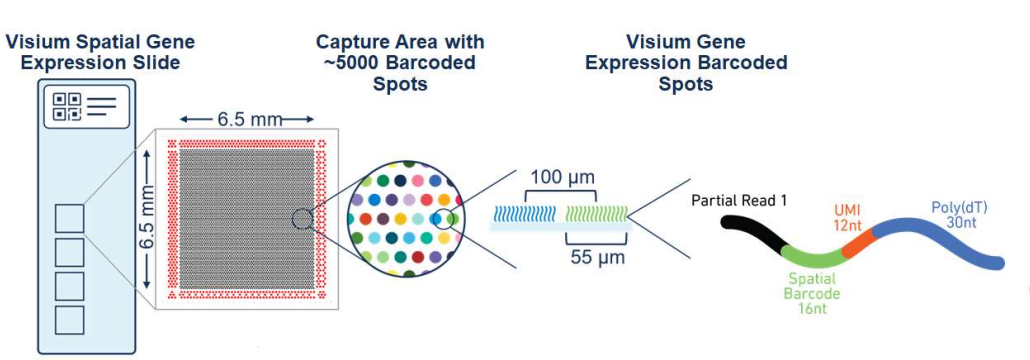
10x Single-Cell vs. Spatial Platforms

Platform	Resolution	Sample Type	Key Use
Chromium	Single cells	Fresh	Cell heterogeneity
Visium	50–55 µm spots	FFPE/FF	Tissue architecture
Visium HD	2 µm	FF	Subcellular mapping
Xenium	Molecular imaging	FFPE	In situ validation

10x Genomics Visium: Spatial Transcriptome



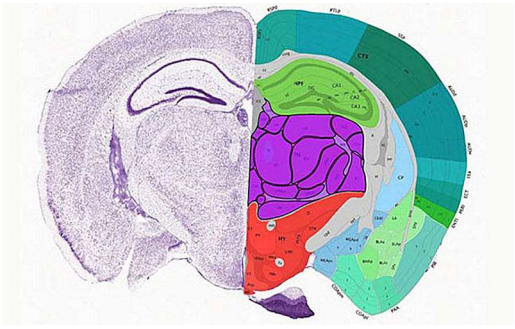


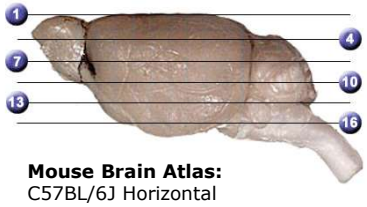


CytAssist

EVOS

ALLEN BRAIN ATLAS



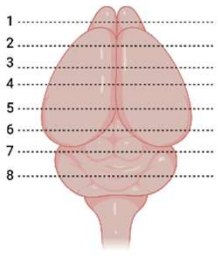


Mouse Brain Atlas:
C57BL/6J Horizontal

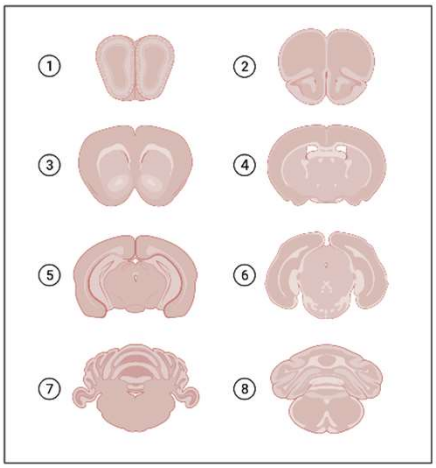
https://www.mbl.org/atlas232/atlas232_frame.html

Dissection direction: Top/Down, Front/Back, Left/Right

Mouse Brain Anatomy
Brain Cross-sections



Dorsal view

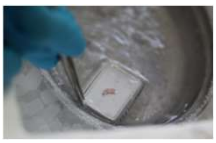


Coronal brain slices

https://app.biorender.com/biorender-templates/t-60db3df7297d8a00a410d7be-mouse-brain-anatomy-coronal-sections

Complete setup for spatial transcriptome workflow (FF, FFPE)

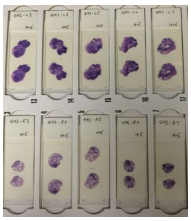
1. OCT embedding



2. Cryosectioning

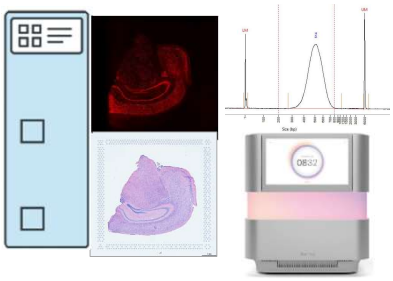
3. H&E staining

4. Tissue catalog



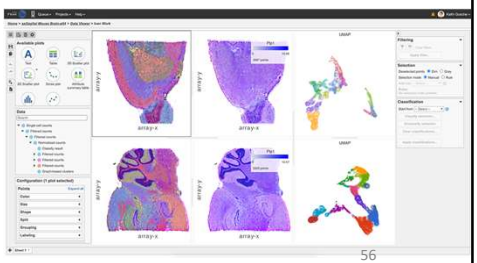
5. RNA QC

6. Visium prep & Sequencing



7. Fluorescent imaging

8. Data analysis with GUI interface

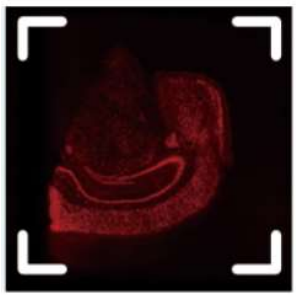


SpaceRanger Analysis:
Clustering of gene expression profiles of spatially barcoded RNA-seq on image

Mouse Brain

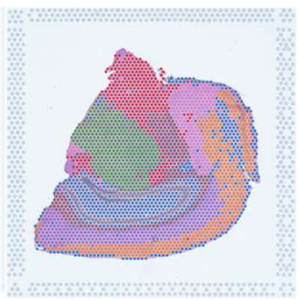
H&E

Capture H&E image from the microscope



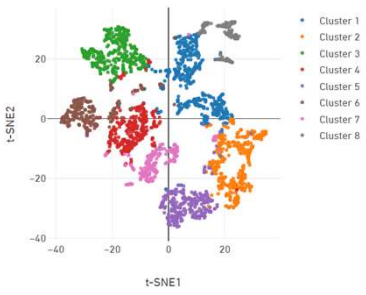
Tissue optimization (cDNA fluorescence)

Tissue lysis on glass slide for RNA capture & cDNA syntehsis



Graphic map overlay

Annotation directly on H&E image

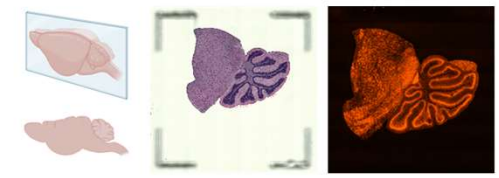


Spatial Transcriptomes Clustering

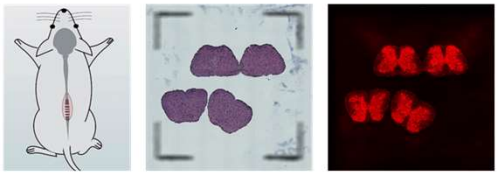
Clustering/classification of glass slide spots based on their gene expression profile

FF Tissue Cryosectioning

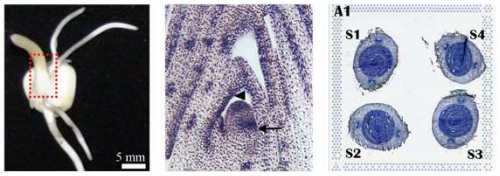
(Mouse Cerebellum)



(Mouse Spinal Cord)

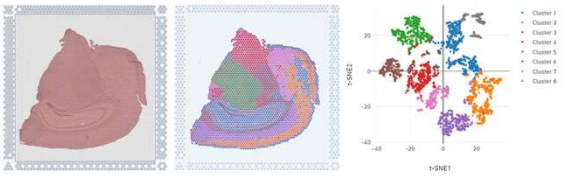


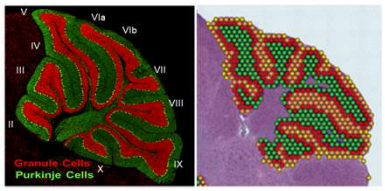
(Embryonic Leaf)

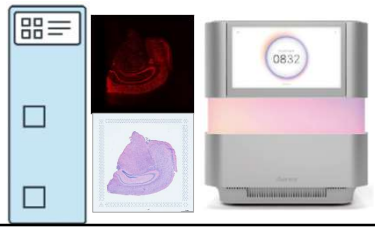


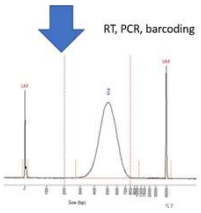
SpaceRanger Analysis

(Mouse Brain)



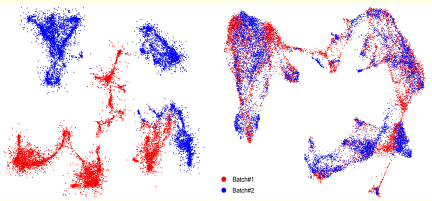






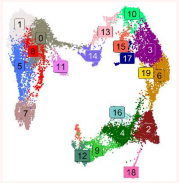
Spatial transcriptomics analysis workflow

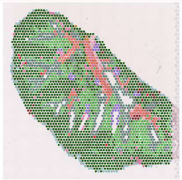
Dataset integration
- Batch-effect removal



Dimension reduction
- UMAP
- TSNE

Unsupervised clustering



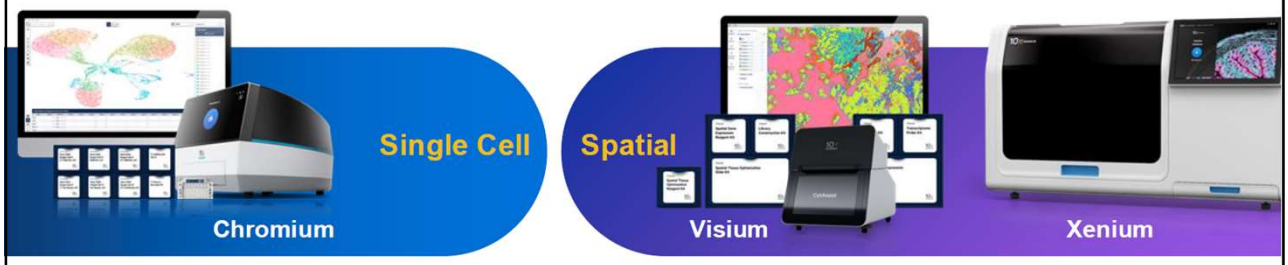


Marker gene identification

Gene	log_2FC	P-val
Cluster 1		
Bcl6b	8.58	< 0.01
Nppa	4.35	< 0.01
Tnfsf10	2.55	< 0.05
⋮	⋮	⋮
Cluster 2		
Lgals7	3.47	< 0.05
Vsig4	3.33	< 0.05
Jchain	2.95	< 0.05
⋮	⋮	⋮
Cluster 3		
Dbp	5.78	< 0.01
Hbb-bt	3.22	< 0.05
Tppp3	3.01	< 0.05
⋮	⋮	⋮

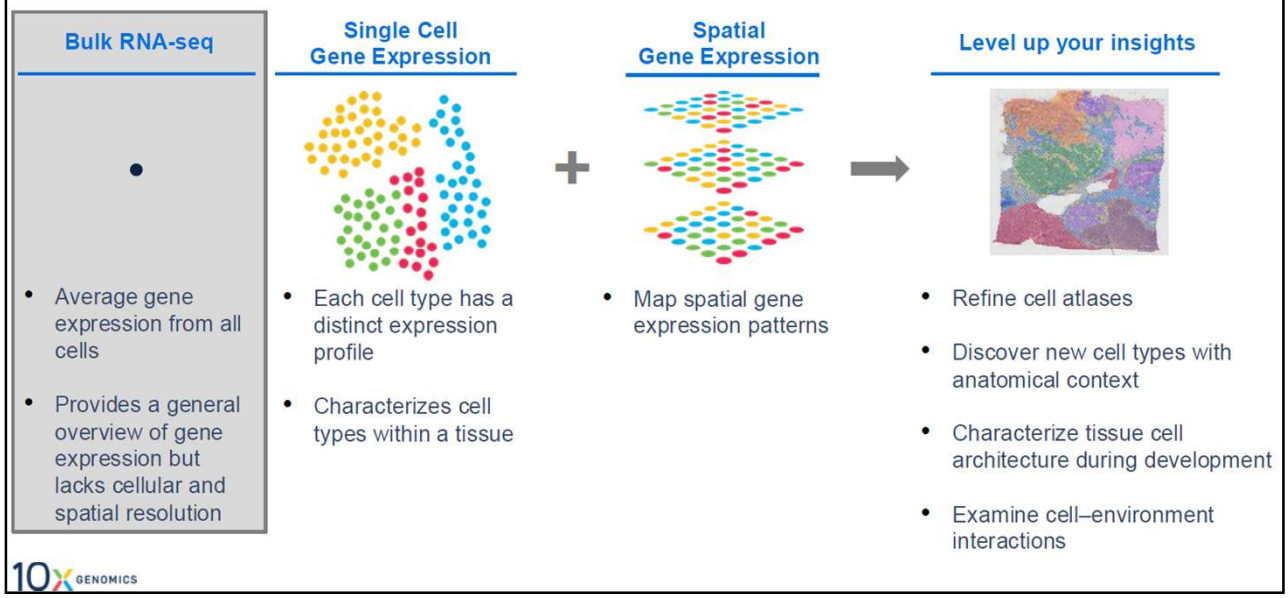
Cell-type annotation
- Cell markers
- Machine learning

10x Genomics Builds Tools to Transform Your Research

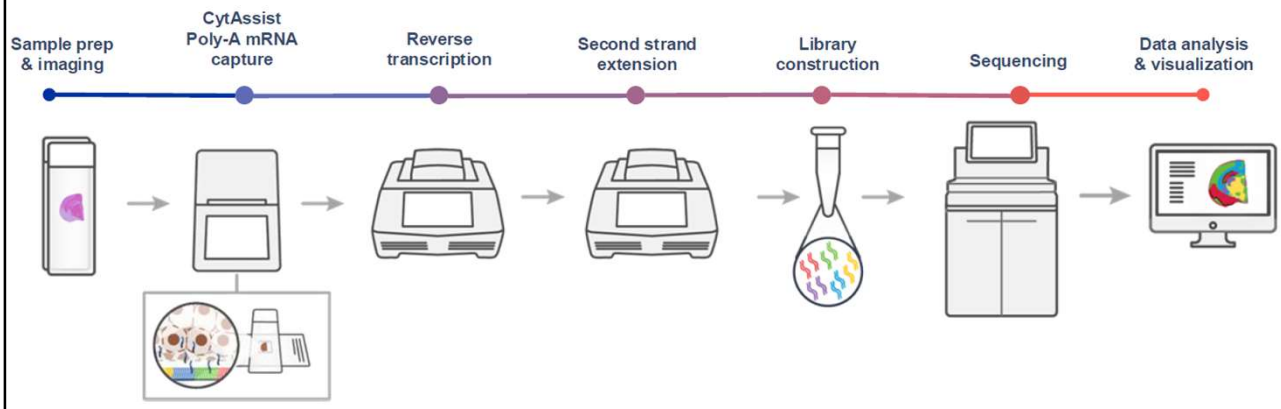


Unbiased Cellular Discovery	Unbiased Spatial Discovery	Precise Single Cell Spatial Insights
Whole transcriptome, V(D)J, CRISPR gRNAs, protein, chromatin	Whole transcriptome	100s–1,000s of genes
Sequencing-based	Transcriptome, protein, histology	Targeted RNA, protein, histology
Single cell resolution	Sequencing-based	High-resolution imaging-based
High per-gene sensitivity and breadth	Single cell scale	Subcellular resolution
	High gene breadth	High per-gene sensitivity
	NGS Sequencing-based	*In situ Imaging-based*

Single Cell and Spatial Omics Are Highly Integratable

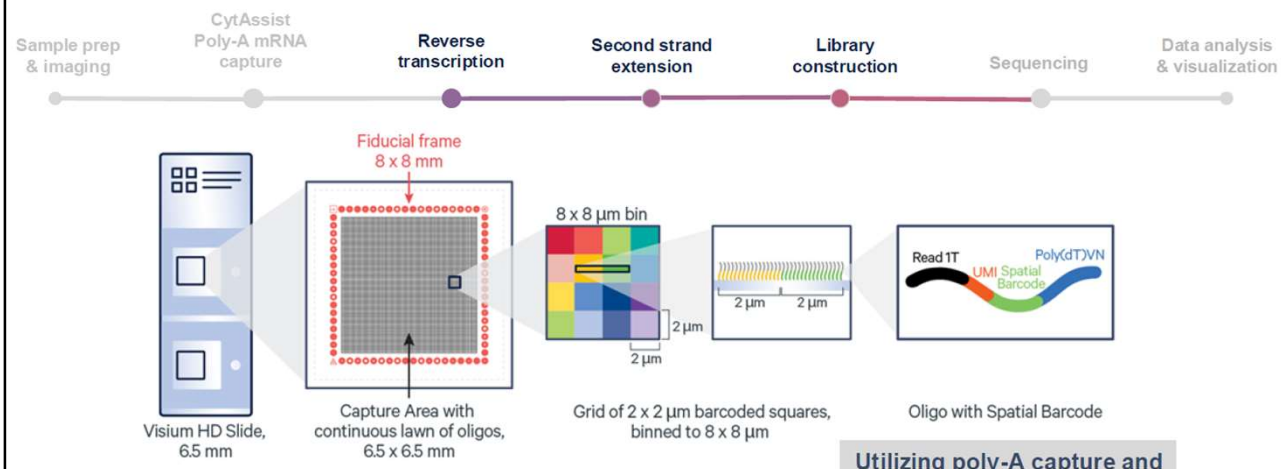


Power by the Visium CytAssist instrument



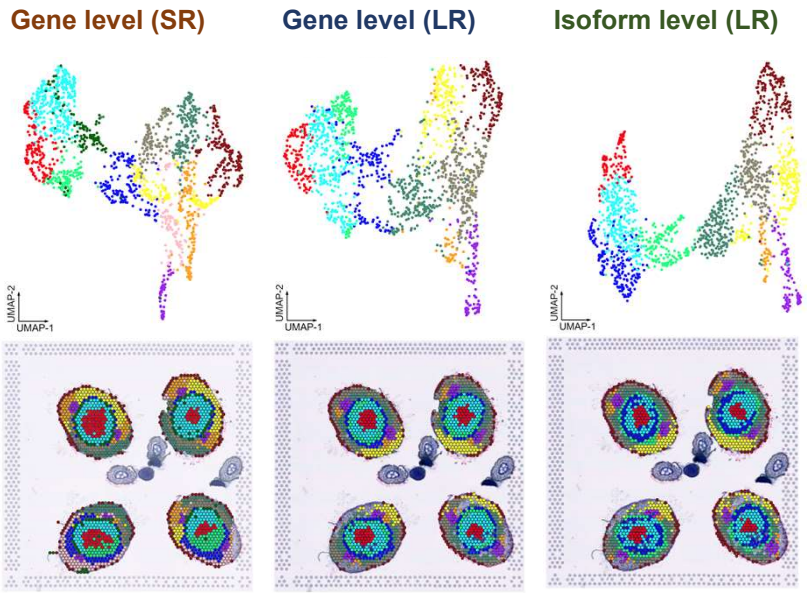
Visium 3'HD

Resolved at Single Cell Scale



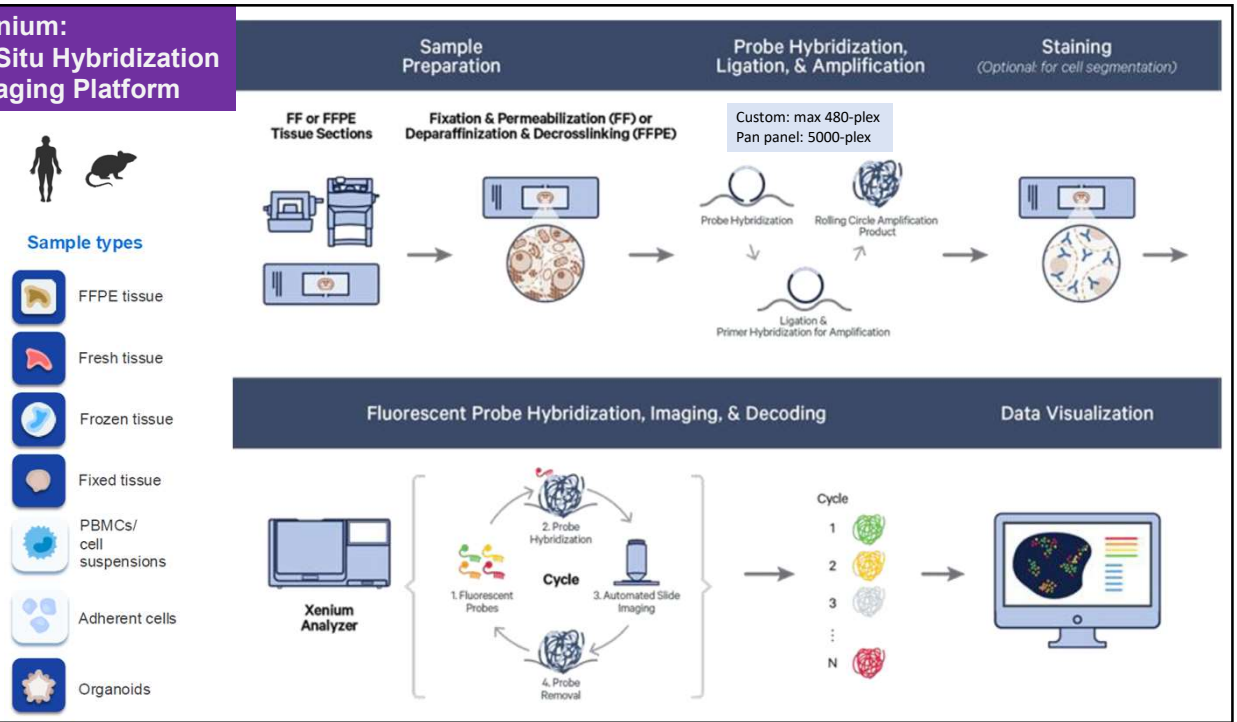
Utilizing poly-A capture and unique spatial barcodes

UMAP clustering topologies maintained b/w different levels



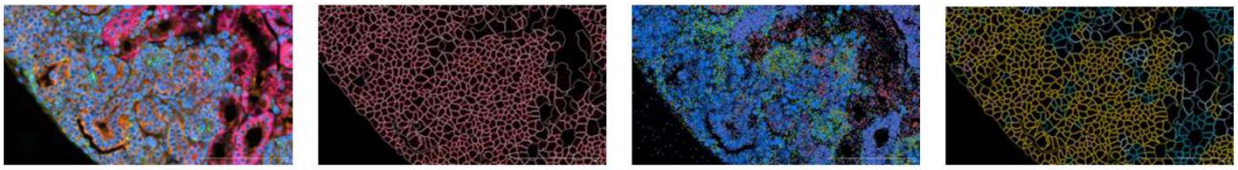
The clustering outputs are similar between the topology of 10xG (SR) gene level, PacBio (LR) gene level, and PacBio (LR) isoform level datasets.

Xenium:
In Situ Hybridization Imaging Platform



Usable Data Immediately After Your Run Ends

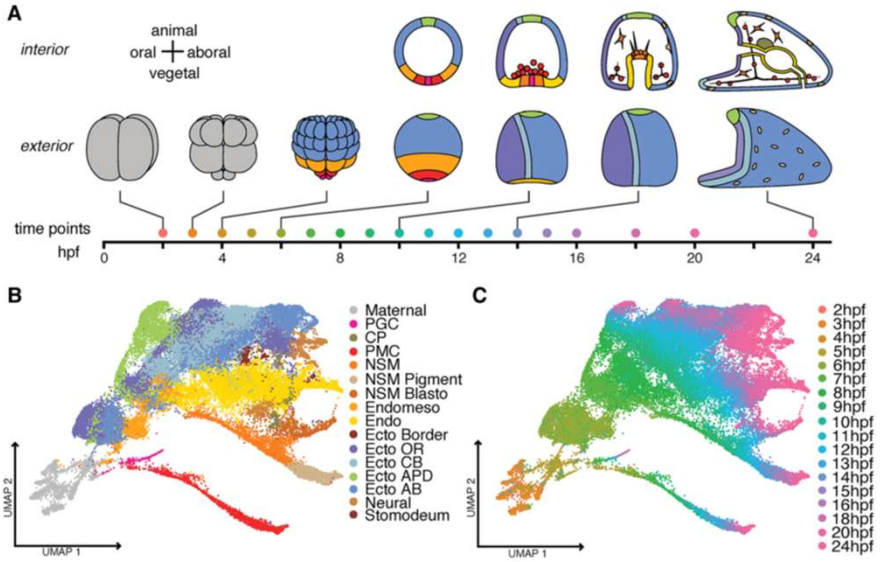
No additional steps required to analyze and visualize your data



Morphology images | Cell segmentation | Localized transcripts | Unbiased Cell Calling

scRNA: Lineage Mapping of Early Sea Urchin Development

- Mapped transcriptional changes over 18 time points during the first 24 hours of sea urchin development
- Identified early lineage divergences and detecting 99% of known developmental gene regulatory network genes in the correct lineages and at the appropriate times
- This high-resolution atlas has advanced the understanding of embryonic cell diversification

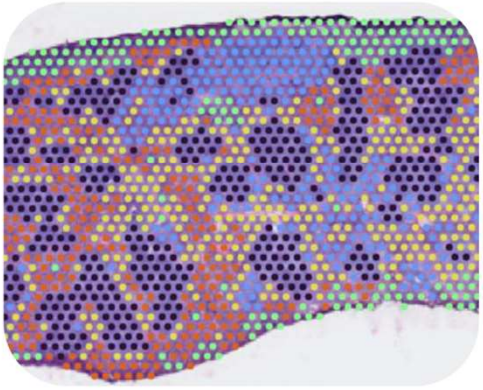


MassriA, Greenstreet L, AfanassievA, et al. *Development* (2021). DOI: 10.1242/dev.198614;

Advances in Technology Bring Visium 3' to High Resolution

Fresh frozen mouse spleen

Visium v1 3'



- Multicell resolution
- Direct placement of tissue
- Tissue optimization

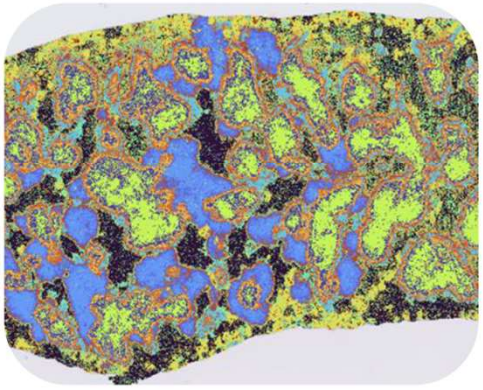
55-um spots, 100-um between centers

Whole transcriptome

Species agnostic

Poly-A based

Visium HD 3'



→ Single cell-scale
→ CytAssist-enabled
→ Single protocol

2-um grids; gapless for cell segmentation; compute @ 2x2 or 8x8 um

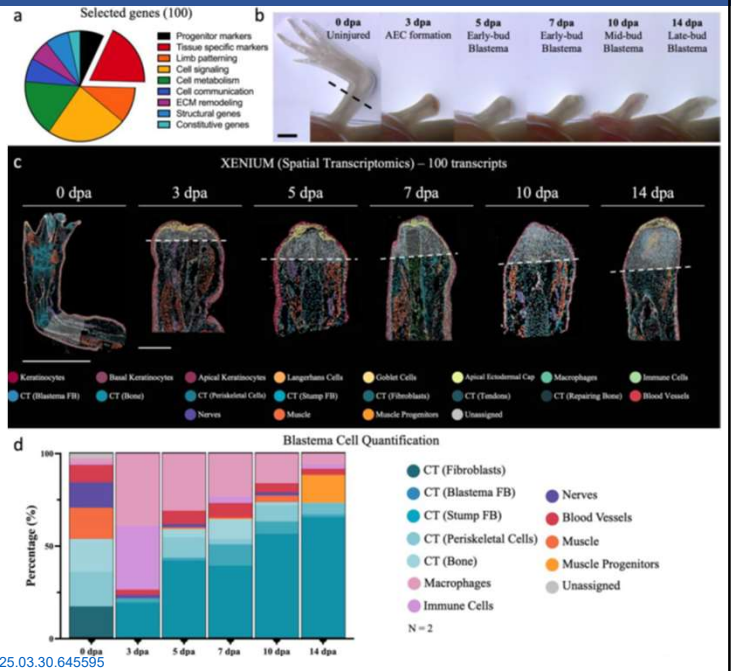
Xenium: Connective tissue (CT)-derived cells are the major constituent of the blastema.

Xenium:

Profiled 100 key genes across regeneration stages

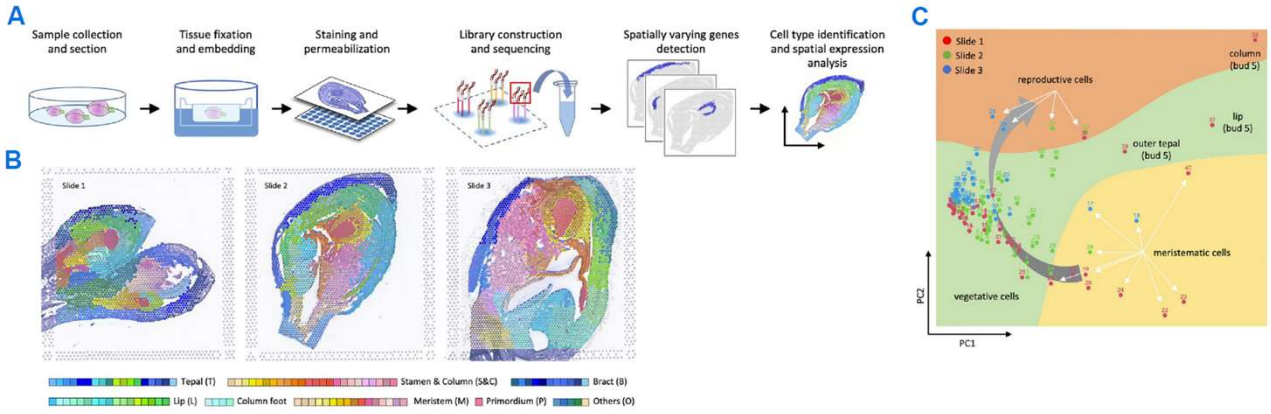
Precisely defined temporal cellular dynamics:

- Early blastemasare enriched with immune cells, particularly macrophages
- Later stages include muscle cells
- Connective tissue-derived cells progressively accumulate and ultimately dominate



García-García D, Knapp D, Kim M, et al. *bioRxiv*(2025). DOI: https://doi.org/10.1101/2025.03.30.645595

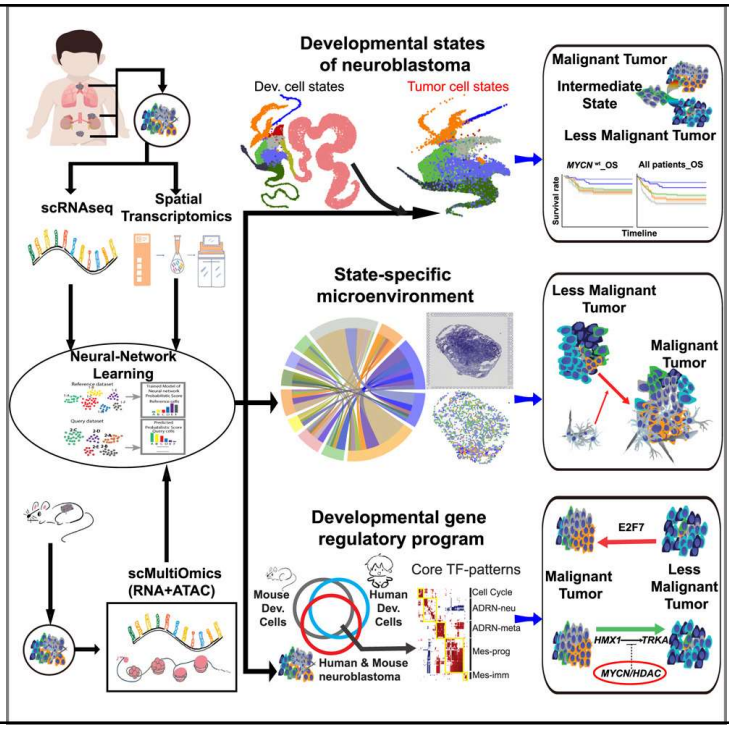
Spatial Transcriptomics Identifies Key Regulators of Orchid Flower Development



Analyzed *Phalaenopsis* with Visium (V1):
(**A**) to identify thousands of tissue-and stage-specific genes
(**B**) PCA grouped 120 cell types into meristematic, vegetative, and reproductive categories
(C) Findings support a model of floral organ initiation and identity determination

Liu C, Leng J, Li Y, et al. *Nucleic Acids Res* (2022). DOI: https://doi.org/10.1093/nar/gkac773

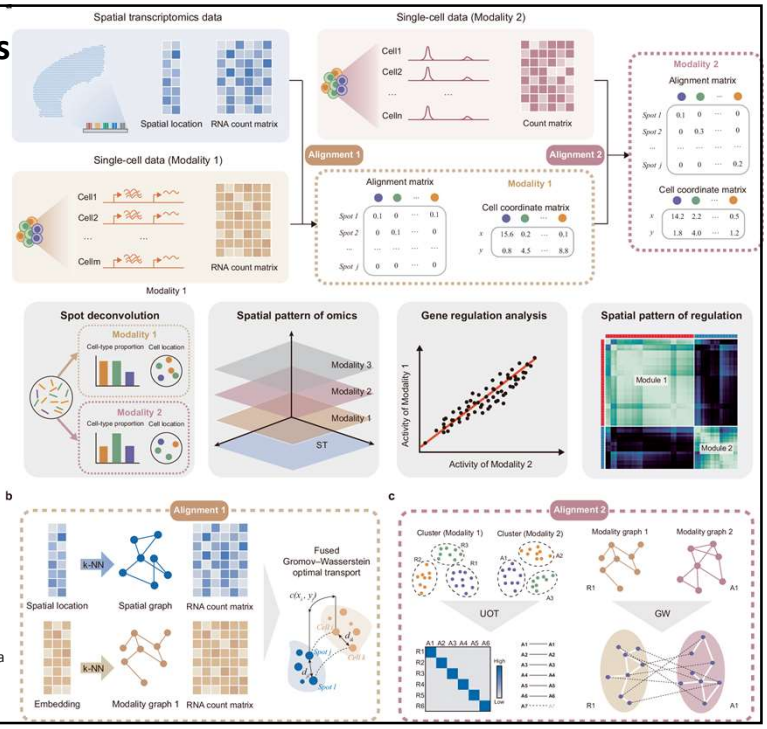
Single-cell MultiOmics and spatial transcriptomics demonstrate neuroblastoma developmental plasticity



Xu, Yunyun, et al. "Single-cell MultiOmics and spatial transcriptomics demonstrate neuroblastoma developmental plasticity." *Developmental Cell* (2025).

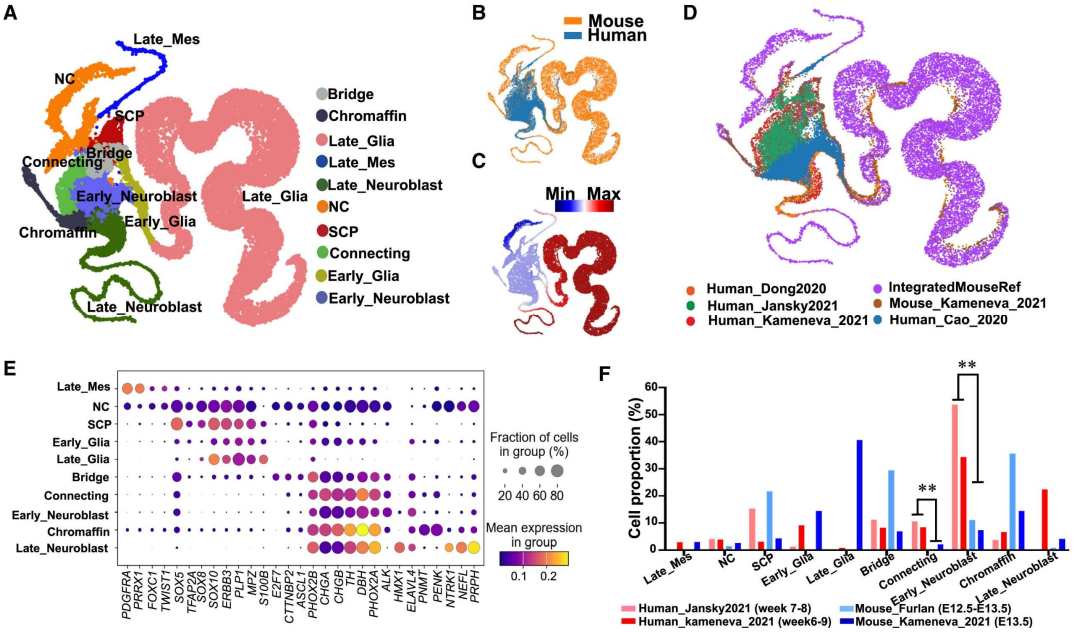
https://www.sciencedirect.com/science/article/pii/S1534580725002515

Spatial integration of multi-omics single-cell data with SIMO



Yang, P., Jin, K., Yao, Y. *et al.* Spatial integration of multi-omics single-cell data with SIMO. *Nat Commun* **16**, 1265 (2025). https://doi.org/10.1038/s41467-025-56523-4

Comparative analysis of chromaffin cell differentiation trajectories in humans and mice



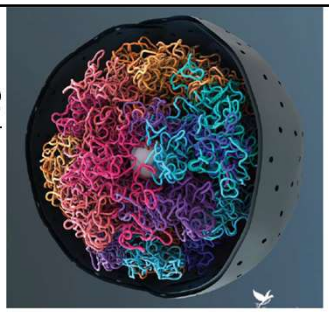
Yang, P., Jin, K., Yao, Y. *et al.* Spatial integration of multi-omics single-cell data with SIMO. *Nat Commun* **16**, 1265 (2025). https://doi.org/10.1038/s41467-025-56523-4

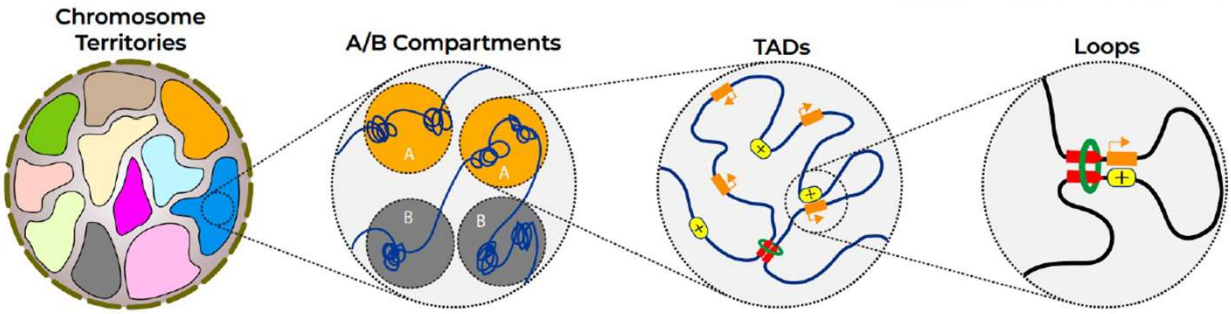
Section IV.
3D Genomics and Chromatin Architecture

Advanced Applications in Genomics:

- ***3D-Genomes: Hi-C, OmniC, microC***
- *10x Genomics: Single-cell & Spatial analyses*
- *PacBio MAS/Kinnex: HT-HiFi*

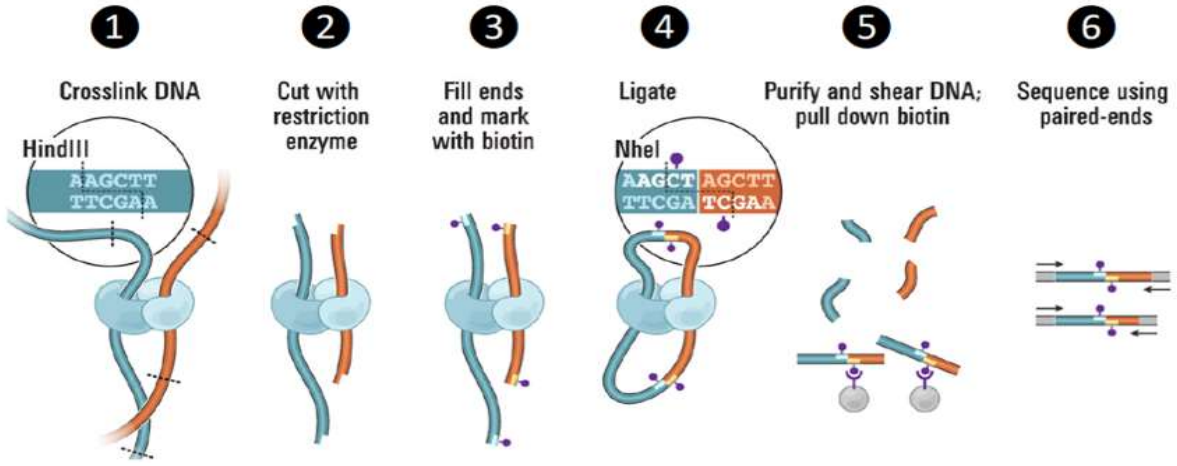
Chromatin Architecture





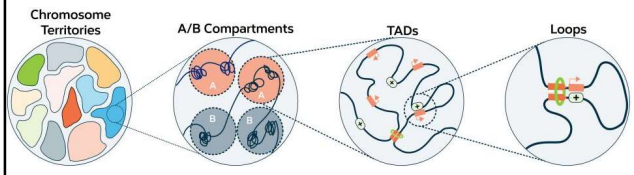
Hierarchy of chromatin topology spanning four major topological features; chromosome territories, A/B compartments, TADs and loops. (Adapted from Wright *et al.*, 2019)

Hi-C: Proximity Ligation

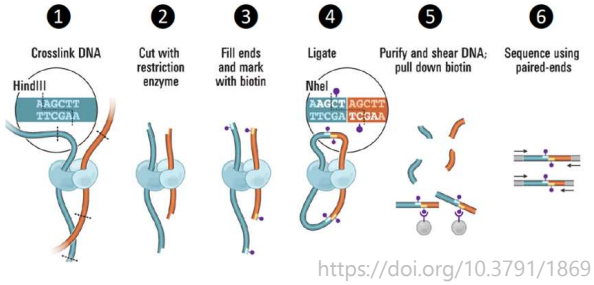


https://doi.org/10.3791/1869

Unlock 3D-Genome Architecture at Nucleosomes



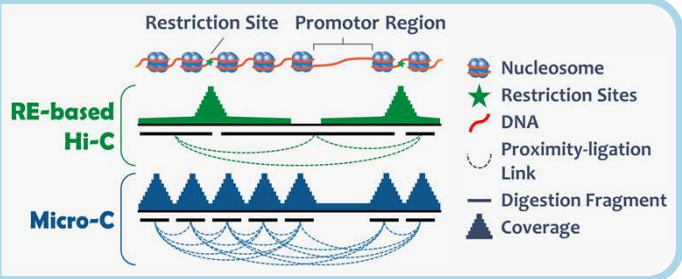
Dovetail Genomics (https://dovetailgenomics.com)



https://doi.org/10.3791/1869

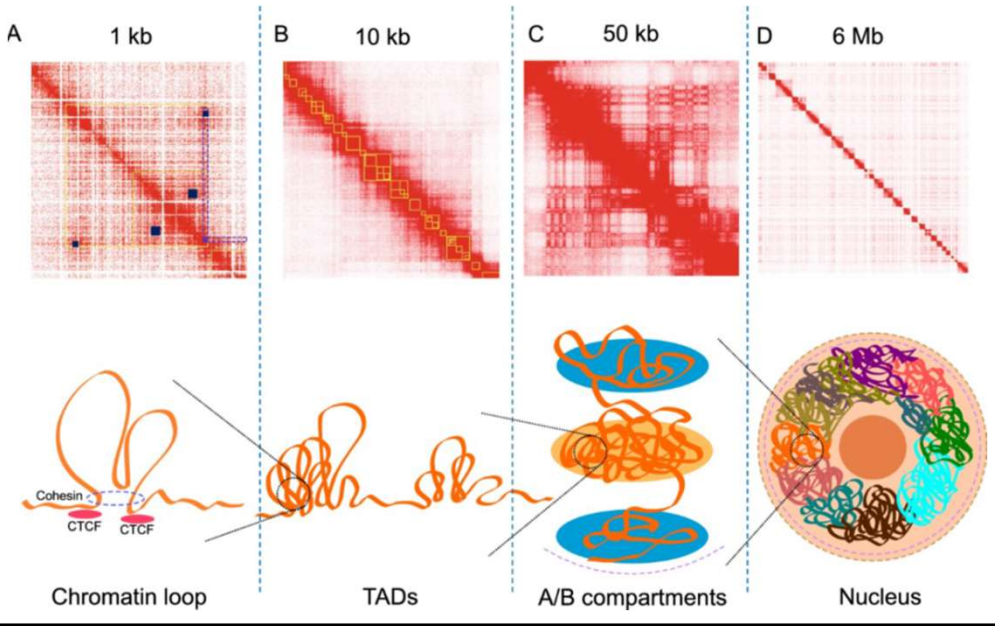
Hi-C Chromatin scaffolding for genome assembly, genome phasing, and/or genome 3D landscape

Micro-C High-resolution interacting domains

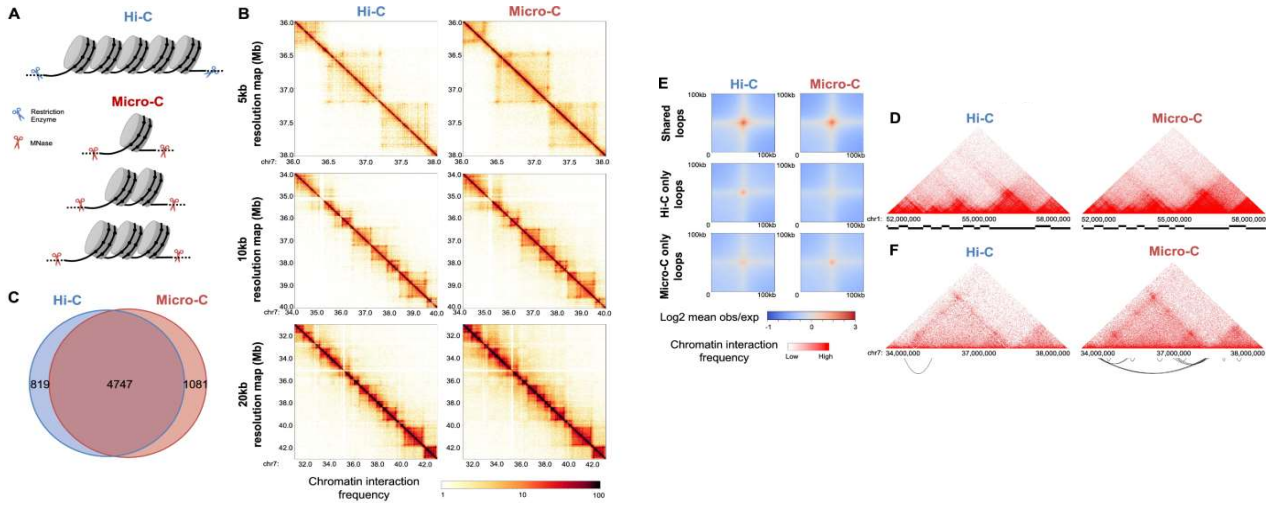


78

3D Genome: assay for chromatin architecture

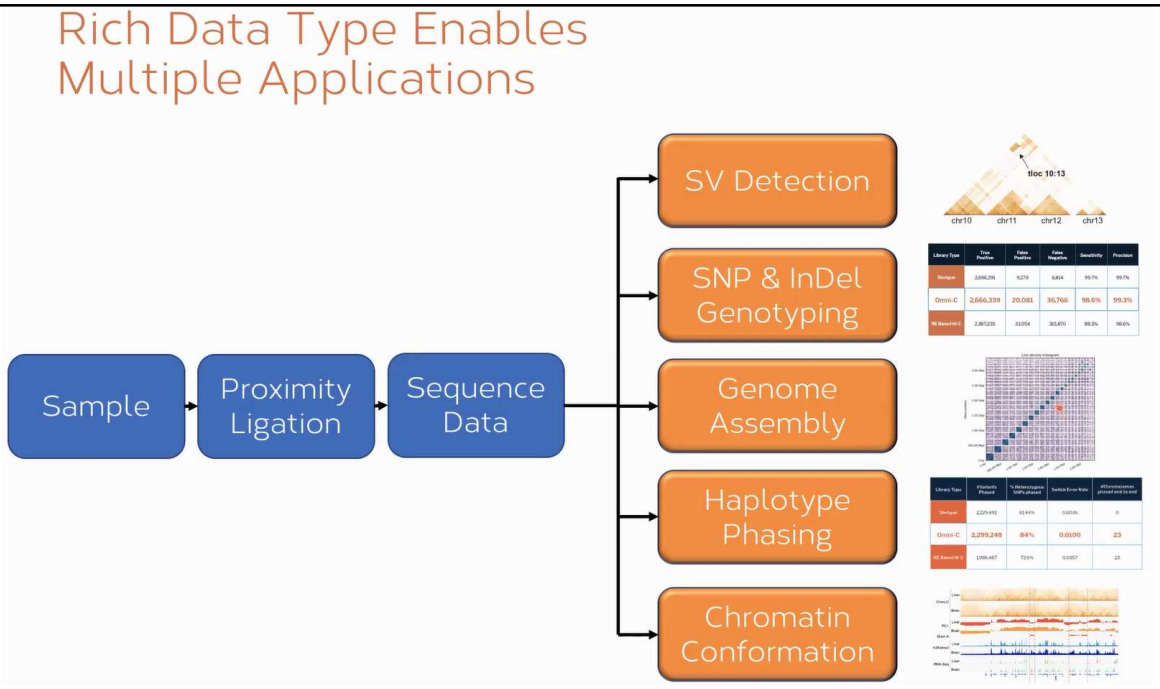


Characterizing chromatin interactions of regulatory elements and nucleosome positions, using Hi-C, Micro-C, and promoter capture Micro-C



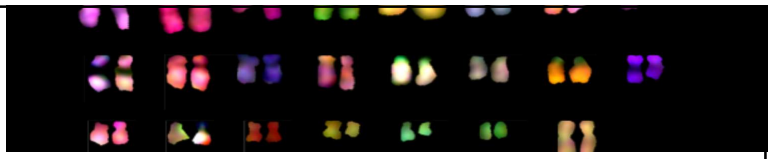
Epigenetics & Chromatin **volume 15**, Article number: 41 (2022)

Rich Data Type Enables Multiple Applications



https://dovetailgenomics.com/

Beyond GRCh38:
T2T genome assembly v1.0



- T2T v1.0: using long-read PacBio and Oxford Nanopore techs
- contain >100Mb novel seq compare to GRCh38
- *Complete* human chr

- Chr X & Chr8: ONT UL backbone; PacBio + ILMN polishing
- Stepwise:
  1. PacBio HiFi reads: construct highly accurate assembly grap
  2. resolve structural ambiguity using Nanopore UL reads
  3. Define complete chr by taking consensus of HiFi reads fron
  4. Nanopore patches several GA-rich repeat gaps in PacBio as
  5. Correct SV and SNP errors with all reads (DeepVariant and
  6. Final polished genome accuracy: error <E10-6
  7. 23 Chr (no Y), 1 Mito.: 3,045,441,522 bp
  8. (only the 5 rDNA arrays remain unfinished: near-identical t

https://genomeinformatics.github.io/CHM13v1/

T2T-CHM13v2.0 - Genome – Assembly 2023

T2T-CHM13v2.0

Description: T2T CHM13v2.0 Telomere-to-Telomere assembly of the CHM13 cell line, with chrY from NA24385
Organism name: Homo sapiens (human)
BioSample: SAMN03255769
BioProject: PRJNA559484
Submitter: T2T Consortium
Date: 2022/01/24
Synonyms: hs1
Assembly level: Complete Genome
Genome representation: full
GenBank assembly accession: GCA_009914755.4 (latest)
RefSeq assembly accession: GCF_009914755.1 (latest)
RefSeq assembly and GenBank assembly identical: no (hide details)
- Only in GenBank: chromosome MT (in non-nuclear assembly-unit)
- Data displayed for RefSeq version

Expected final version: no
Genome coverage: 30x

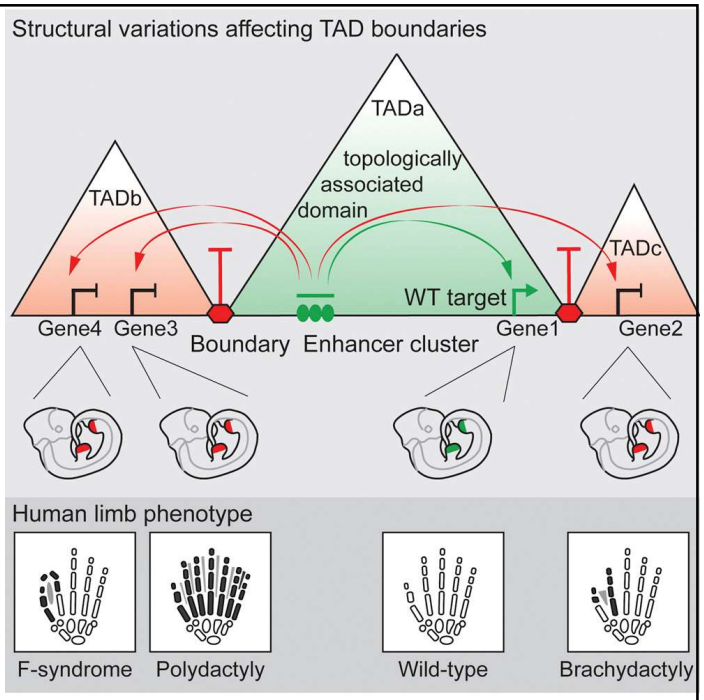
24 chr: 22 autosome+XY

IDs: 11828891 [UID] 31127148 [GenBank] 31865168 [RefSeq]

Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

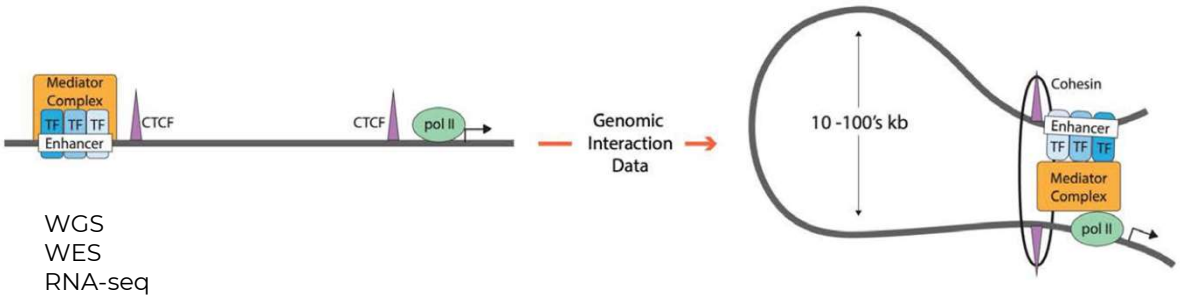
The impact of SVs in inherited disease:

- **Gene Regulation Happens in 3D**
- Genes are regulated by enhancers that may be far away in linear distance but physically close in 3D space.
- Chromatin loops and TADs bring genes and regulatory elements together or keep them apart.
- **Example**: Misfolding of chromatin can lead to **ectopic enhancer–gene interactions**, causing diseases like cancer or limb malformations.



Lupiáñez, Darío G. et al . Cell, Volume 161, Issue 5, 1012 - 1025

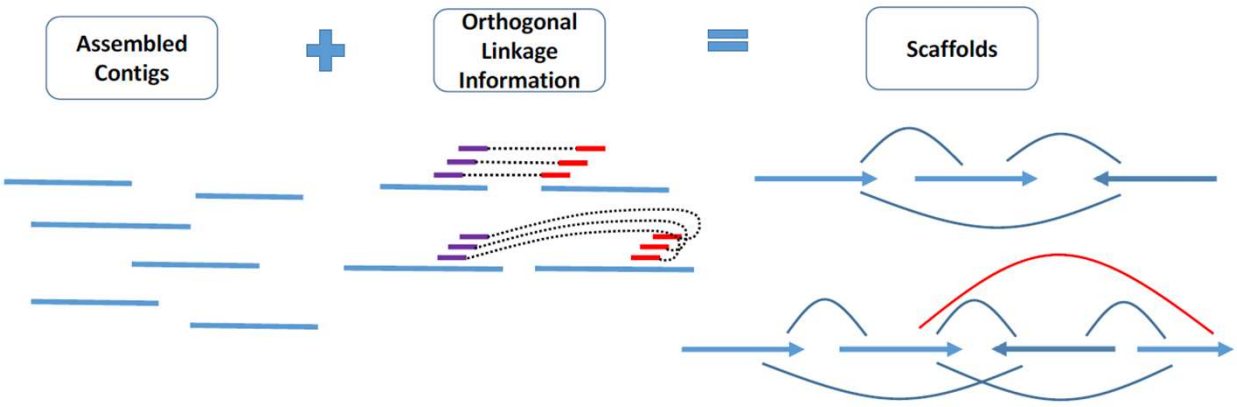
NGS Assays Do Not Capture the *In Situ* Structure of the Genome

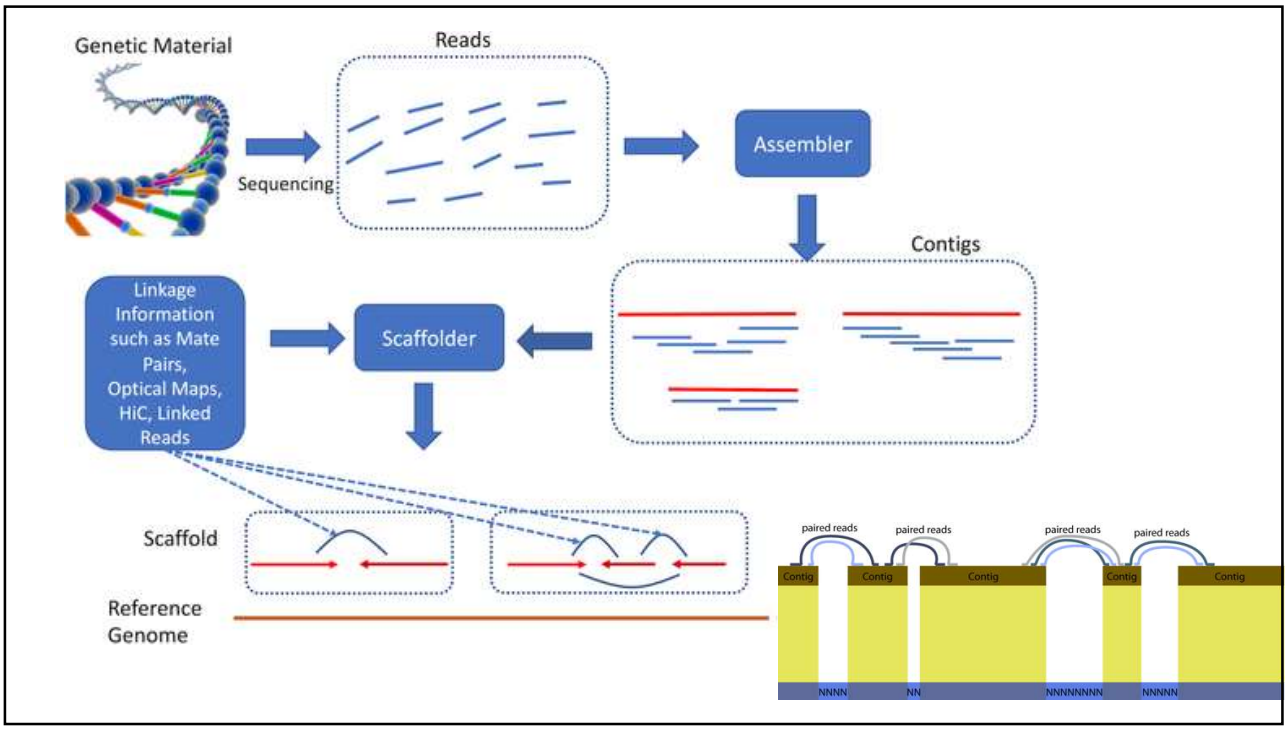


WGS
WES
RNA-seq
ATAC-seq
Methylation seq
ChIP-seq

Dovetail GENOMICS

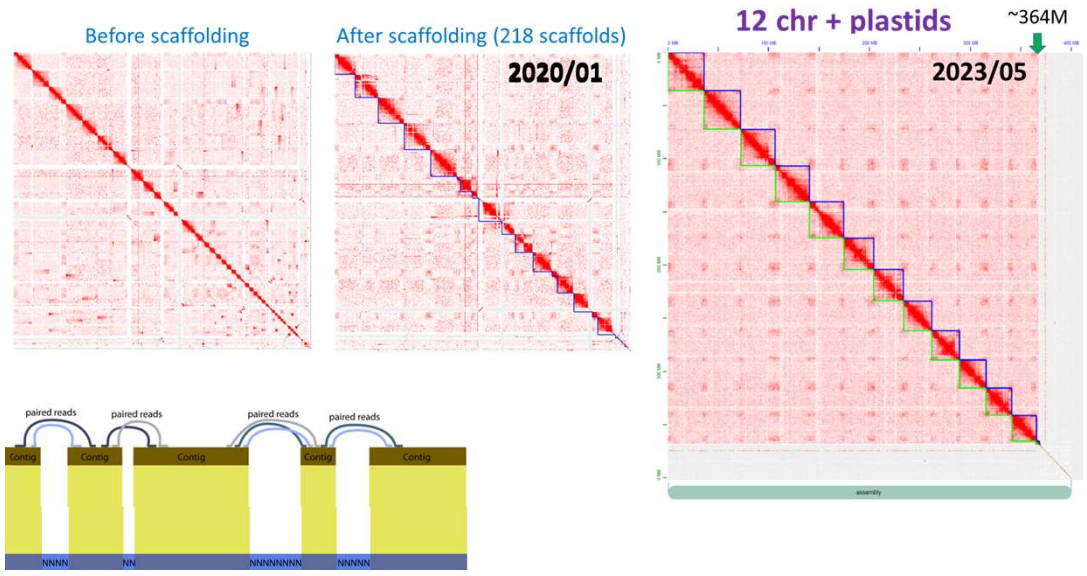
Genome scaffolding





Chromosome-level assembly: LR-assembly + Hi-C scaffolding

Melon genome scaffoding, 2*n* = 2*x* = 24



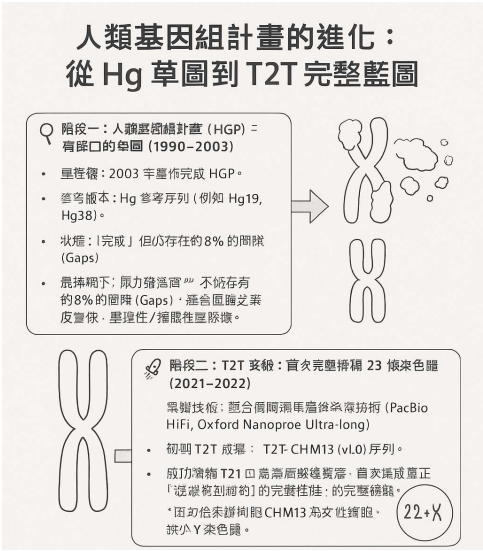
Section V.
Genome Case Studies / Future Perspectives

Human Genome T2T

- **Complete Reference Assembly**
  - T2T-CHM13 fills all previous gaps
  - Adds telomeres, centromeres, and Y chromosome
  - Achieved via PacBio HiFi + ONT ultra-long reads
- -> “Complete Blueprint.”

人類基因組計畫的進化：從 Hg 草圖到 T2T 完整藍圖

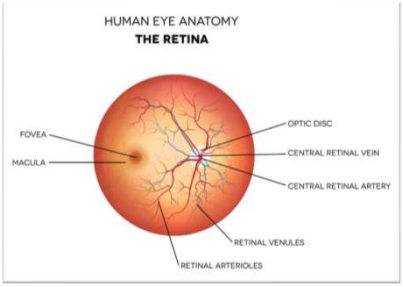
從一個充滿「空白」的草圖，最終進化到一個「從端粒到端粒」的完整藍圖

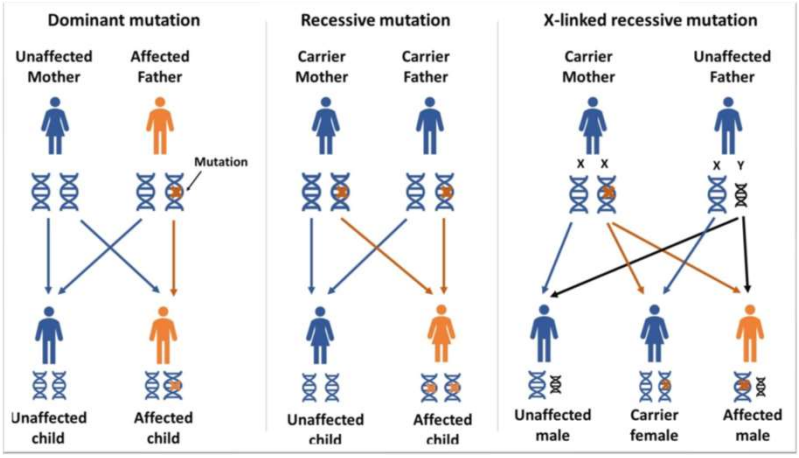


攻克罕見疾病：台灣的基因靶向診斷 SNP HDR panel

- Hereditary Retinal Dystrophy，中文為「遺傳性視網膜失養症」。 HRD 是一種遺傳性視網膜失養症的縮寫，這是一組因光感受器功能異常而導致視力受損的遺傳性疾病。這些疾病會對視力造成嚴重影響，並且可以透過基因檢測來診斷。

- 主要原因： 由於視網膜上的光感受器（感光細胞）出現功能性問題而引起。
- 可能影響： 可能會導致嚴重的視力損害。
- 診斷方法： 基因檢測可以協助診斷這類疾病。





攻克罕見疾病：台灣的基因靶向診斷

SNP: HDR panel

- 「最常見的差異叫做 SNP，就像書上某個單字多了一個字母或少了一個字母。透過高密度分析，我們可以快速找到這些差異，它們與我們的特徵、體質息息相關。」

Inherited Retinal Dystrophy (IRD) – Targeted Panel Study Lu M.-Y. J. & Huang S.-P. et al., *npj Genomic Medicine (2024)*

- Cohort: 493 patients (425 probands), Taiwan
- Panel: 319 IRD genes (1.2 Mb region)
- Coverage: >200×; 94% on-target bases
- Diagnostic yield: 68.5% overall (80% syndromic)
- 87 novel variants discovered
- Top genes: *USH2A, EYS, CYP4V2, ABCA4*, RPGR

Impact: Largest IRD cohort in Taiwan; enables gene therapy readiness & precision diagnosis

診斷挑戰 (Diagnostic Challenge)

- 罕見疾病多樣、複雜，診斷困難且耗時。

台灣方案 (Taiwanese Solution: Designed Capture Panel)

- 目標：快速、經濟、準確的罕見疾病基因診斷。
- 技術：自主研發「遺傳性疾病基因捕獲晶片」，靶向近 4,000 個已知疾病基因。
- 優勢：成本低、分析聚焦，專為台灣設計。

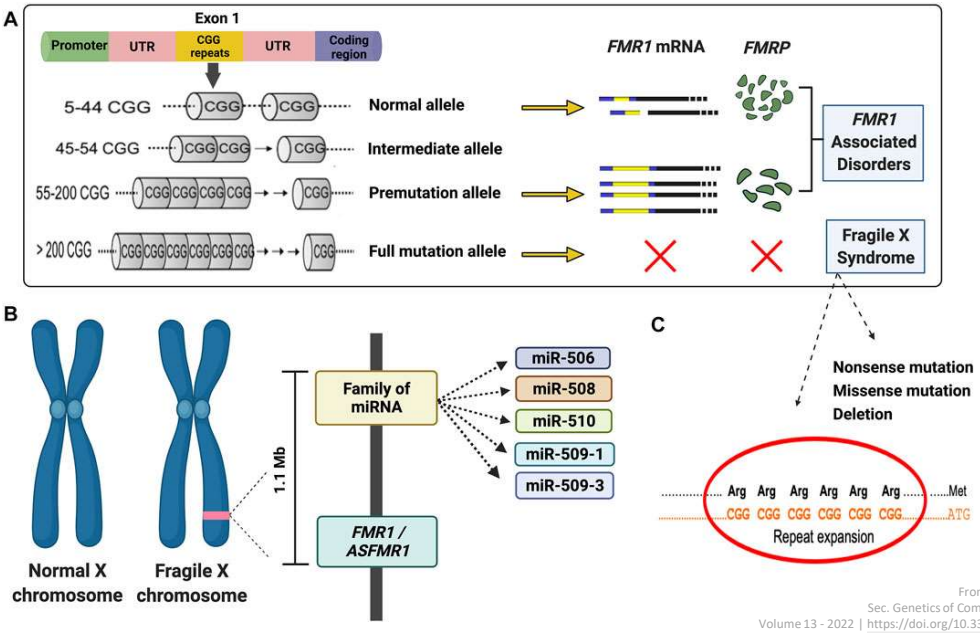
關鍵成果 (Key Achievements)

- >500 名罕病患者成功收案並診斷。
- 達成高診斷率，找到致病基因變異。
- 臨床意義重大： 提供精準診斷，指導治療/遺傳諮詢，避免重複檢查。

精準醫療的曙光 (Precision Medicine's Promise)

- 台灣經驗：將基因定序從研究推向臨床，點亮罕病家庭的希望。

CNV: 脆性X症候群（Fragile X Syndrome）



Front. Genet., 11 May 2022
Sec. Genetics of Common and Rare Diseases
Volume 13 - 2022 | https://doi.org/10.3389/fgene.2022.884424

The Taiwan Precision Medicine Initiative provides a cohort for large-scale studies *Nature* (2025)

Hsin-Chou Yang, Pui-Yan Kwok, Ling-Hui Li, Yi-Min Liu, Yuh-Jyh Jong, Kang-Yun Lee, Da-Wei Wang, Ming-Fang Tsai, Jenn-Hwai Yang, Chien-Hsiun Chen, Erh-Chan Yeh, Chun-yu Wei, Cathy S.-J. Fann, Yen-Tsung Huang, Chia-Wei Chen, Yi-Ju Lee, Shih-Kai Chu, Chih-hsing Ho, Cheng-Shin Yang, Yungling Leo Lee, Hung-Hsin Chen, Ming-Chih Hou, Jeng-Fong Chiou, Shun-Fa Yang, ... Jer-Yuarn Wu + Show authors

TPMI：繪製台灣人的健康地圖

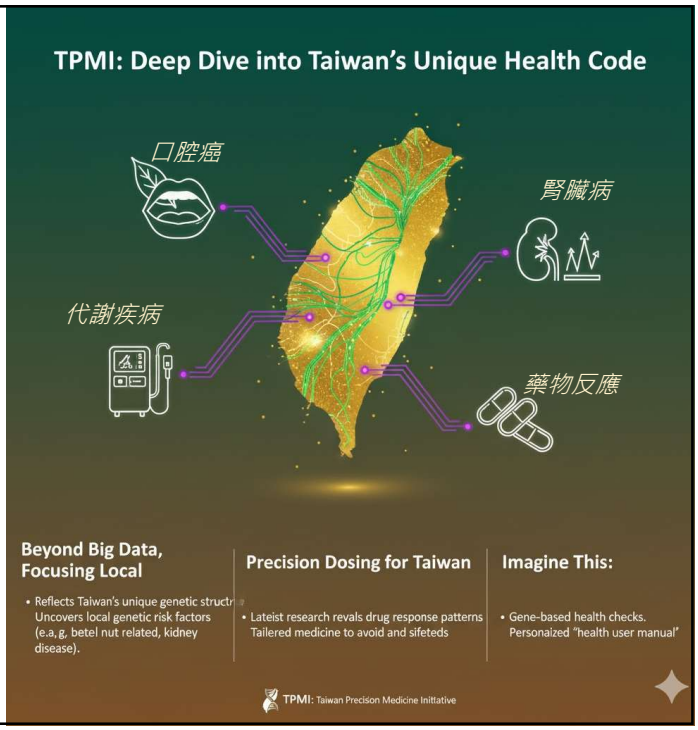
- **台灣的國家級旗艦計畫 (Taiwan's National Flagship Initiative)**：由中研院主導，集合全台數十家醫院與生物資料庫，打造**全球最大規模的漢人基因體資料庫**。
- **填補空白**：過去基因研究多以西方人為主，TPMI 專注於 **>50 萬名台灣民眾的基因型鑑定晶片 (SNP Array)**，提供最完整的漢人基因圖譜。
- **改變未來**：旨在透過整合基因、生活型態與臨床數據，推動疾病預防、精準診斷、個人化用藥及新藥研發。

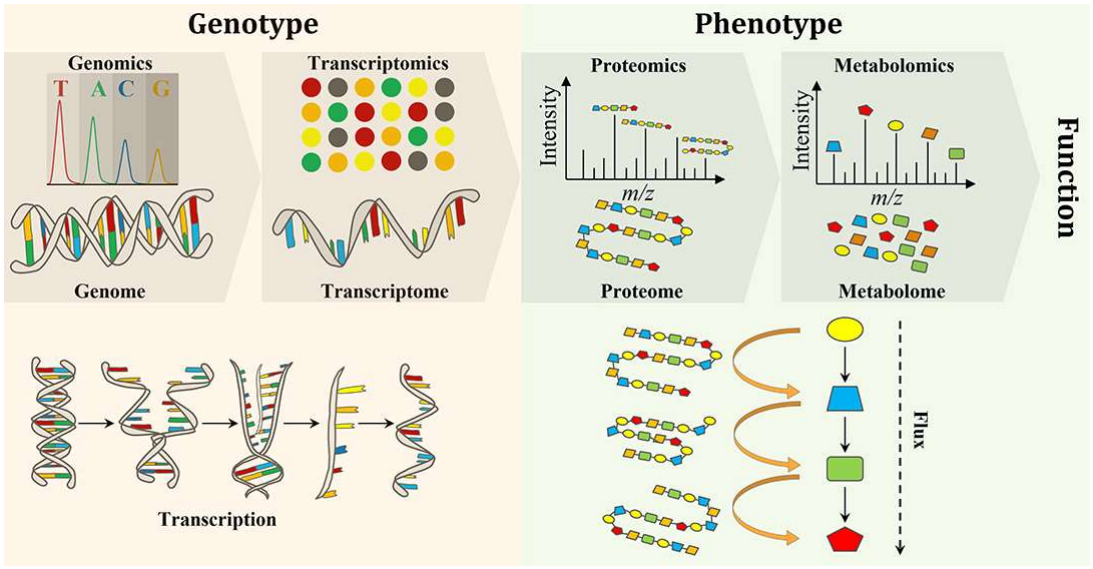
Clinical impact of pharmacogenetic risk variants in a large chinese cohort *Nature Communications* (2025)

Chun-Yu Wei, Ming-Shien Wen, Chih-Kuang Cheng, Yi-Jing Sheen, Tsung-Chieh Yao, Sing-Lian Lee, Jer-Yuarn Wu, Ming-Fang Tsai, Ling-Hui Li, Chun-houh Chen, Cathy S.-J. Fann, Hsin-Chou Yang, Yen-Tsung Huang, Hung-Hsin Chen, Yi-Min Liu, Erh-Chan Yeh, Yu-Ching Peng, Shuu-Jiun Wang, Shih-Pin Chen, Ming-Tsun Tsai, Teh-la Huo, Chien-Wei Su, Der-Cherng Tarng, Chin-Chou Huang, ... Pui-Yan Kwok

TPMI：深入解讀「台灣」的健康密碼

- **超越大數據，聚焦本土**
  - TPMI 不僅建立漢人資料庫，更**精確反映台灣民眾特有的基因結構與變異**。
  - 這有助於理解**台灣高盛行率疾病**（如洗腎、痛風、C 型肝炎、口腔癌等）的**本土基因風險因子**。
- **精準用藥的台灣準則：**
  - 多篇最新研究（例如 2025 年論文中可能涉及的**藥物基因體學**）正在揭示**台灣人對特定藥物的反應模式**。
  - 這能幫助醫生避免藥物副作用，提升治療效果，讓「個人化用藥」在台灣實現。
- TPMI 是**全球最大**的漢人全基因組資料庫，其成果對所有華人社群 (無論身處何地) 都具有參考價值。





Amer B and Baidoo E. *Front BioengBiotechnol*(2021). DOI: 10.3389/fbioe.2021.613307

Key applications: genomics, metagenomics, multi-omics

1. PacBio HiFi:
- De novo assembly
- Haplotype phasing
- Metagenome assembly
- Transcriptome Isoseq: FL- Isoform
- High-resolution profiling: taxonomy, CRISPR-screening

2. Oxford Nanopore:
- Long read / Ultra-Long read assembly
- Genome phasing
- DNA/RNA modifications
- Metagenome assembly

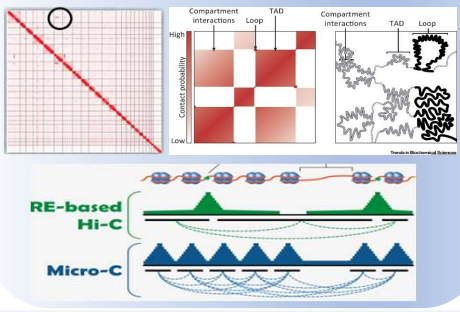
1. 3D genome:
- Hi-C: genome scaffolding, chr phasing
- Micro-C: euchromatin dynamics, chromatin interactions
- Hi-ChIP: ChIP-seq+Hi-C

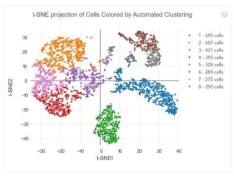
2. 10x Genomics:
- Single-cell: RNA-seq, ATAC-seq, immune repertoire
- Spatial RNA-seq
- HT in-situ hybridization

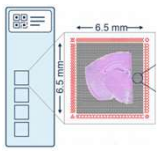
3. Complementary applications:
- Single-cell/Spatial IsoSeq
- Single-cell Hi-C
- LR isoform profiling and clustering

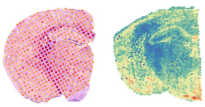
Current Platforms

3D-Genomes









10x Genomics: Single-Cell & Spatial seq

Countess

Chromium X

CytAssist

EVOS

NextSeq2000

MiSeq

ElementBio Aviti24

PacBio Sequel & SQIIe

Nanopore GridION

Promethion P2 Solo