

High Throughput Sequencing – NGS Core intro & NGS platform principles, applications, and advanced extensions

高通量基因體定序：
NGS平台簡介、應用、和先進拓展

Mei-Yeh Lu, PhD

High Throughput Genomics Facility
Biodiversity Research Center
Academia Sinica

LSL Workshop, 2023/5/24



Outline

I. Evolution of sequencing technologies

- NGS platforms: NGS & TGS
- DNA preprocessing

II. Principles of common NGS applications

- Sample QC
- DNA applications
- RNA applications

III. Advanced applications

- 3D Genomes
- Single-cell and spatial analyses



Missions – provide high quality data and R&D for NGS researches

Establishment:

- 2008: as a research NGS Core
- 2014: promoted as AS service core
- 2019~ : supported as ASCF



NGS technologies:

- SOP: [sample QC](#), [applications & sequencing](#)
- Technology upgrade/acquisition
- Application diversification
- R&D for new apps. & system improvement

Education and R&D expertise:

- Lectures & workshops: [TIGP programs](#), [LSL workshop](#)
- Collaborative Research projects
- Provides NGS consultation on:
 - [Project's need](#)
 - [suitable NGS experimental design](#)
 - [Sample preparation](#)
 - [Cost analysis](#)

Where to Find Us?

中央研究院
生物多樣性研究中心

新世代基因體定序核心實驗室

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

Equipment

Sequencer

QC Related

Library Prep Related

NGS High Throughput Genomics Core at BRCAS

台北市南港區11529研究院路2段128號
路標大樓 A603

Genome is the basis of all living creatures throughout evolution. Massive parallel sequencing can generate huge amounts of data efficiently at low unit cost.

We operate three platforms – **2nd-Gen Illumina (high-output, short-read sequencing)**, as well as **3rd-Gen PacBio (circular consensus Hi-Fi)** and **Nanopore (ultra-long single molecule sequencing)**, and provide applications for 3-D genomics and single-cell spatial applications.

These platforms have their best-suited applications, and can also be customized complementarily for each project's unique needs.

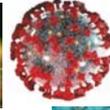
Advancements in HTS technologies



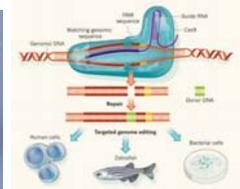
NGS projects of HTG Core in Academia Sinica

Projects:

- Pathogens
- Insect genomes
- Evo-devo: avian species
- C3/C4 plants
- Marine animals
- Human diseases
- Microbiomes
- Single-cell / Spatial



Launching of the 3rd-Gen sequencing service of PacBio Sequel system



I. Evolution of Sequencing Technologies

- From Sanger to Next-Gen Seq.
- Data preprocessing

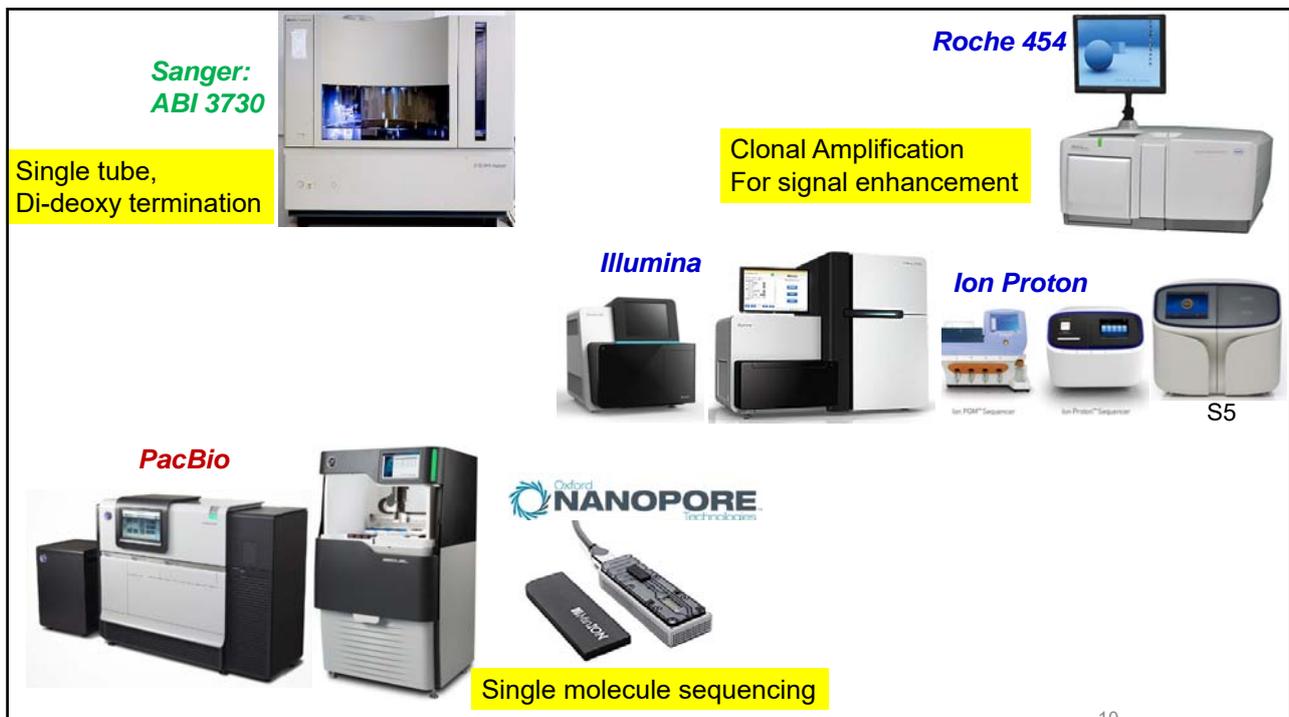
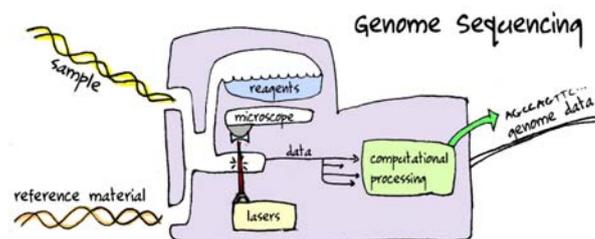
High Throughput Sequencing (HTS) – terminologies & general properties

NGS:

- Next-Gen, New-Gen, 2nd-Gen
- Clonal amplified signals
- Shorter reads
- Higher raw base accuracy
- ILMN cyclic termi., Ion Proton
- SNVs, short INDELS, rare alleles

TGS:

- 3rd-Gen
- Single molecule signal
- Long reads; retain base modifications
- Lower raw base accuracy
- PacBio SMRT, Oxford Nanopore
- Large SVs, translocation, haplotype

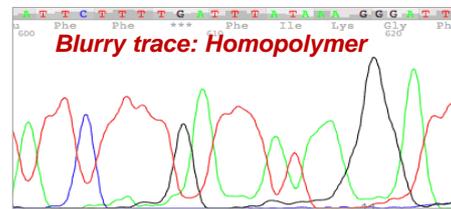
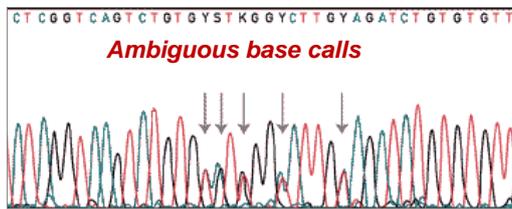
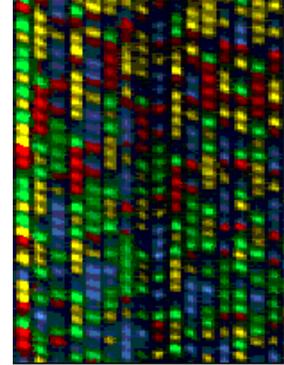
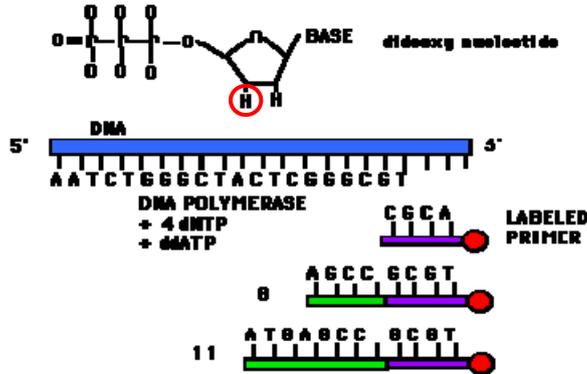


Sanger Seq: chain termination w/ fluorescent ddNTPs

Frederick Sanger

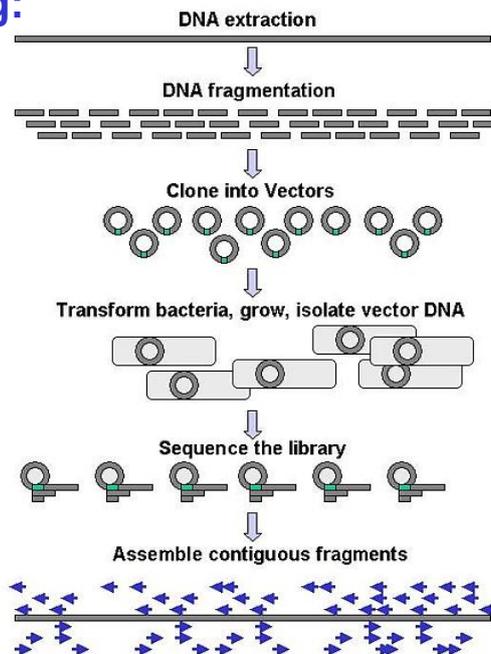


Nobel laureate, 1977



Genome Sequencing: Hierarchical cloning

- BAC
- Cosmid
- Fosmid
- Plasmid



Large scale Cappillary Sequencing

**Library factory -
Whitehead Institute**



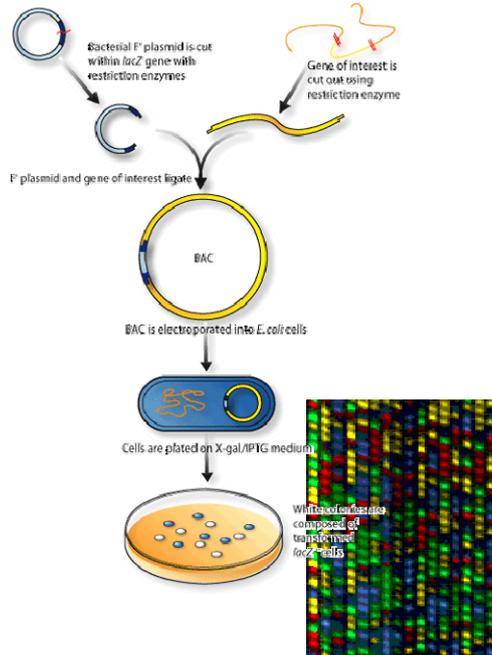
**Sequencing factory -
Sanger Institute**



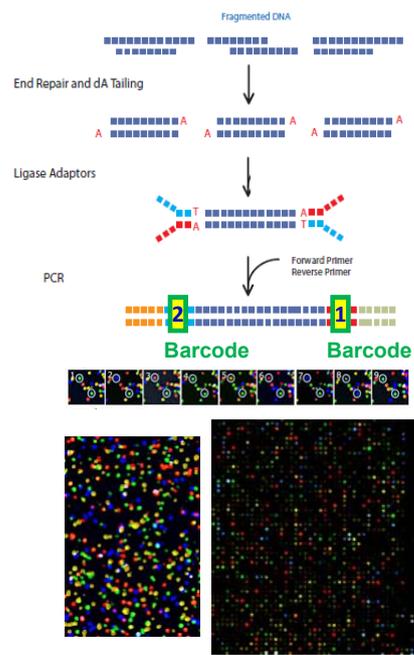
13

Nature 409, 860-921 (15 February 2001)

Cloning-based sequencing



NGS: massive parallel seq.

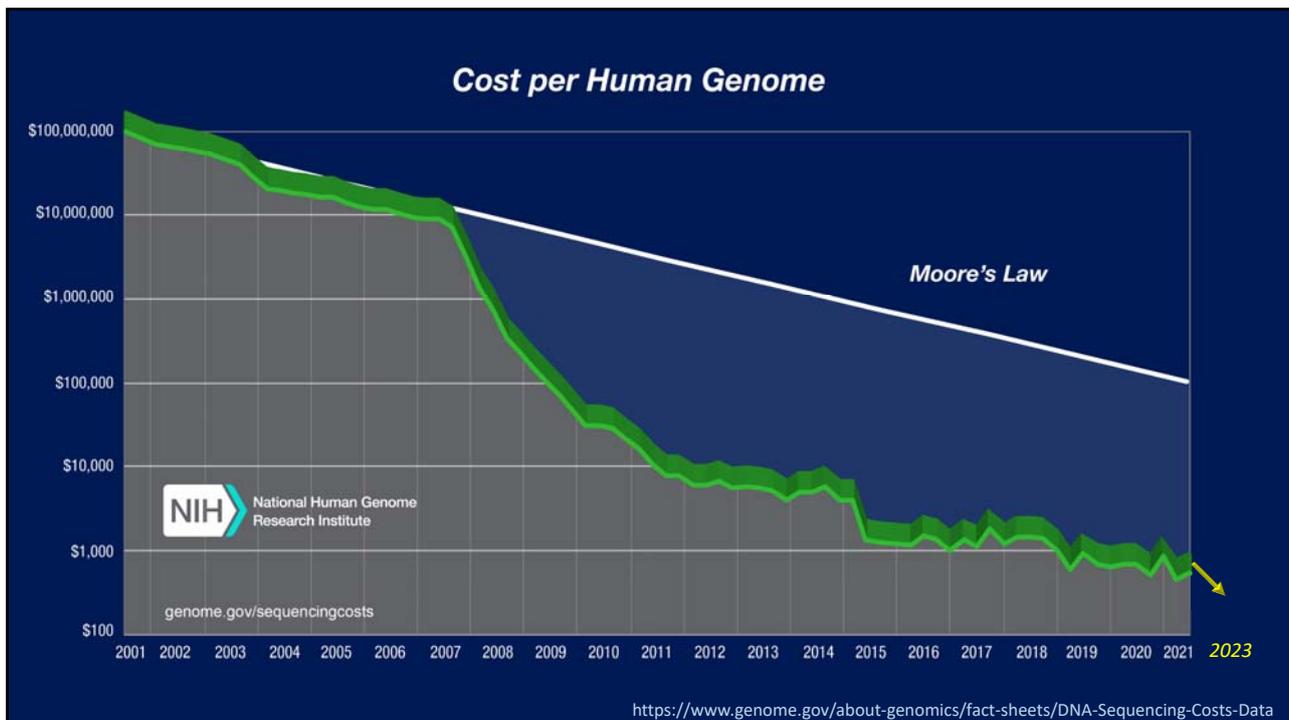


NGS – massive parallel sequencing

Current Popular platforms:

- **2nd-Gen: clonal amplification**
 - Roche 454: GS FLX, , 454 Jr., 454 XL+, 454 Jr.
 - Illumina: GA, [Miseq](#), [HiSeq](#), [NextSeq](#), [NovaSeq](#), [iSeq](#)
 - Life Technologies: SOLiD, Ion Torrent, [Ion Proton](#)
- **3rd-Gen: single molecule sequencing**
 - Pacific Biosciences: PacBio [Sequel](#), [Sequel II](#)
 - Oxford Nanopore: [ONT MinION](#), [GridION](#), [PromethION](#)

15



NGS armies



Current Major NGS Platforms

Illumina
Reversible terminator



HiSeq, MiSeq, NovaSeq, NextSeq

PacBio
SMRT



Sequel (x1~16)



Oxford
NanoPore

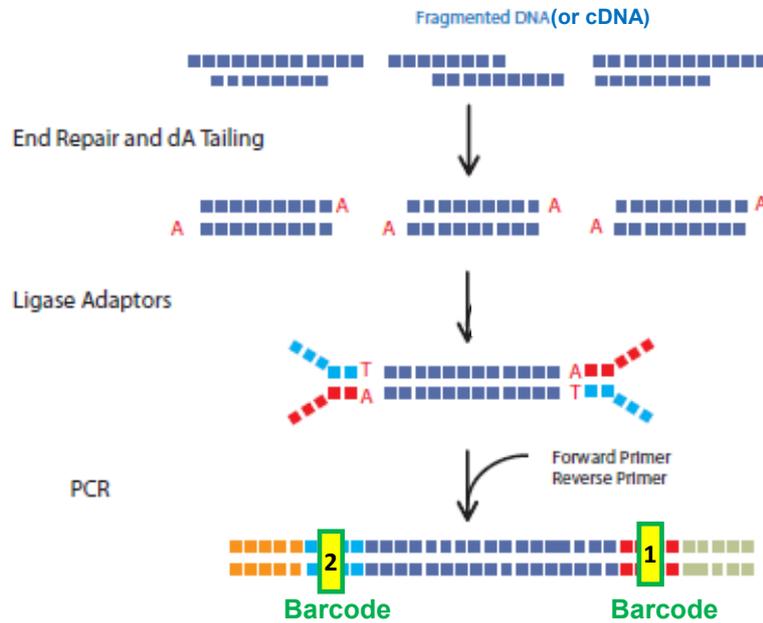


MinION (x1) GrinION (x5)



PromethION (x48)

General workflow for DNA library prep



19

Illumina/Solexa: Cyclic Reversible Terminator (one-base-a-time)

Library prep
(fragment + adaptor)



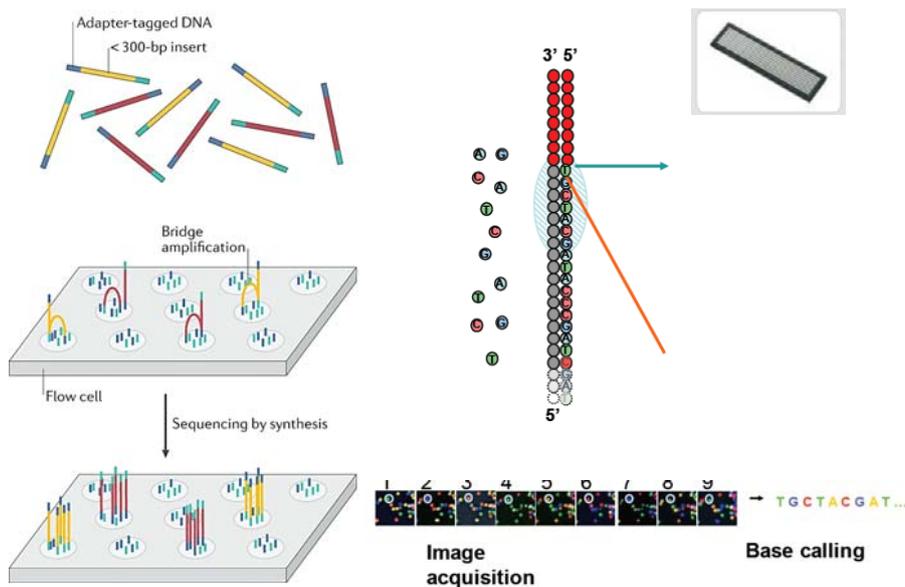
Cluster growth
(bridge formation,
Cluster amplification)



Cyclic Sequencing
(3'-blocked dNTPs)

Base calling

a Illumina short-read sequencing

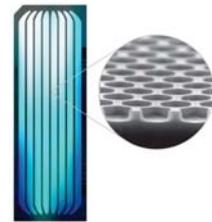
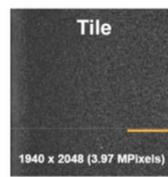
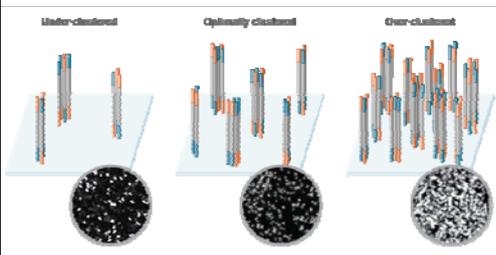


Illumina – Flow cell imaging



**Random flowcell
(4-color chemistry)**

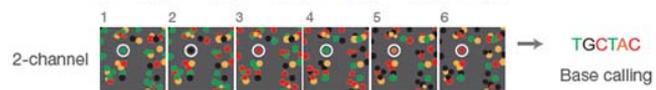
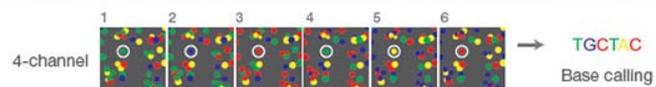
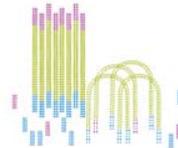
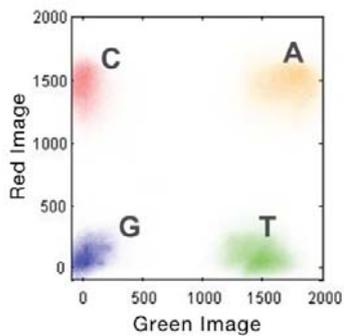
**Patterned flowcell
(2-color chemistry)**



21

2-Channel Chemistry

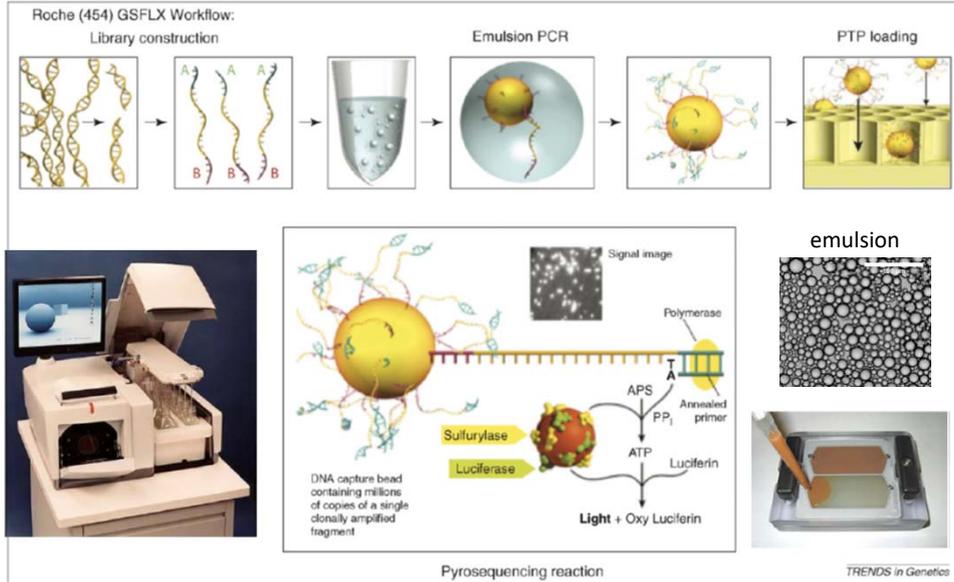
	A	G	T	C
Image 1				
Image 2				
Result	A	G	T	C



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>

454: emPCR & pyrosequencing



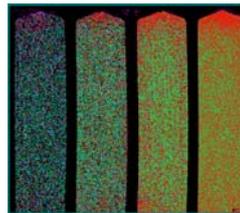
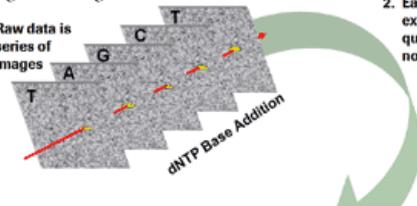
23

454 flowgram and read length profile

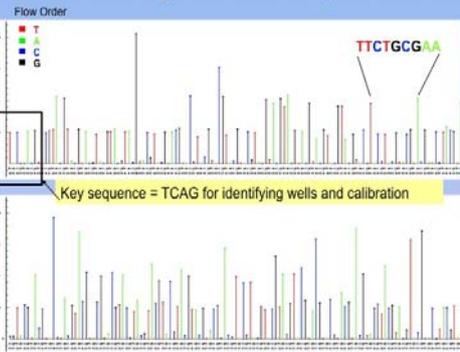
GS FLX Data

Image Processing Overview

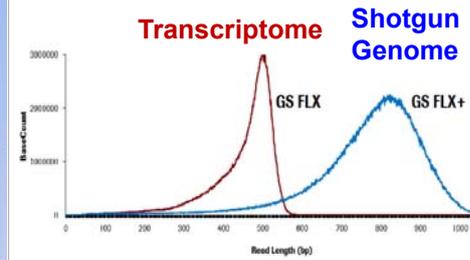
1. Raw data is series of images
2. Each well's data extracted, quantified and normalized



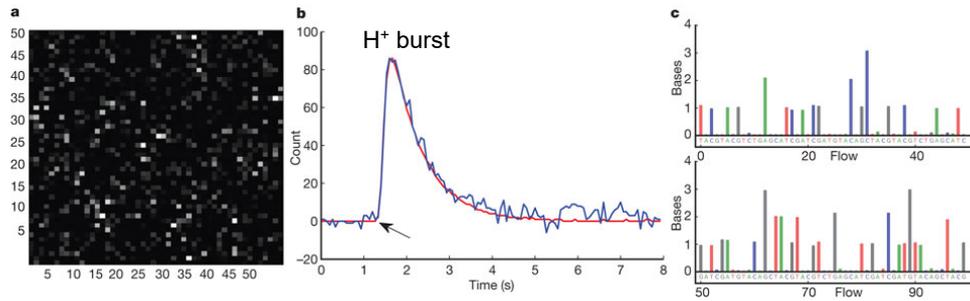
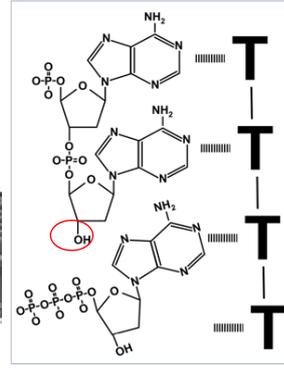
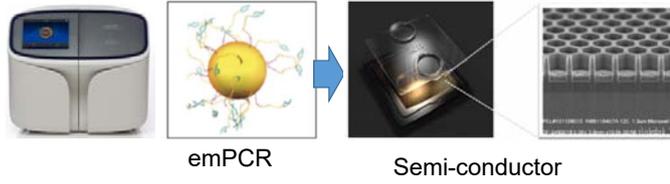
Example of a Flowgram



Significantly more bases from Sanger-like reads

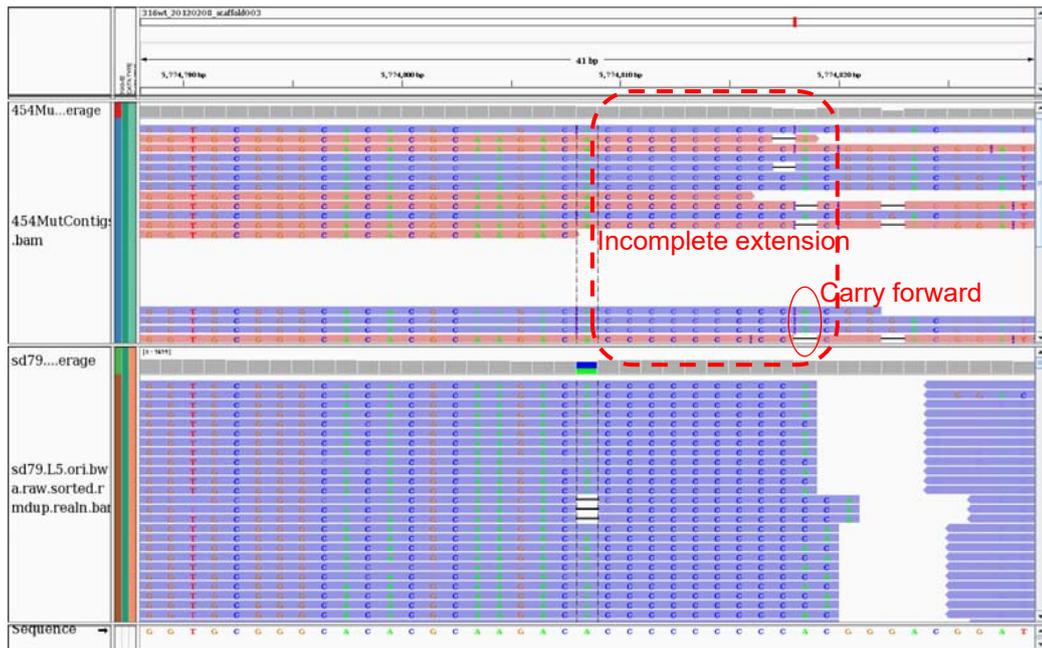


Ion Torrent/Proton: Sensing bulk release of H⁺

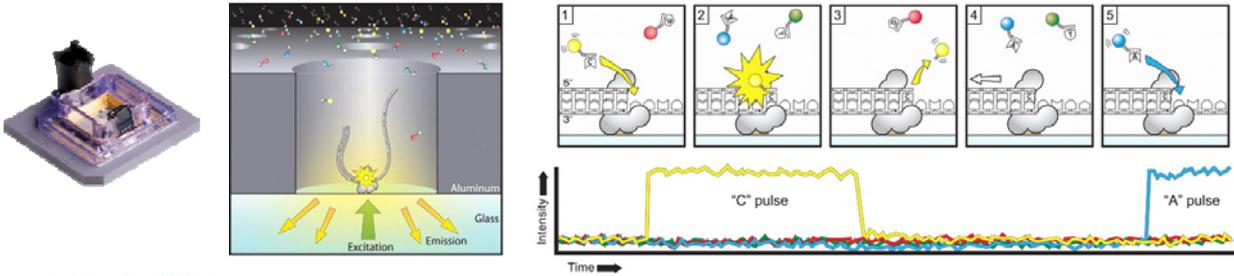


JM Rothberg et al. *Nature* 475, 348-352 (2011) doi:10.1038/nature10242

Homopolymer errors – 9 C's



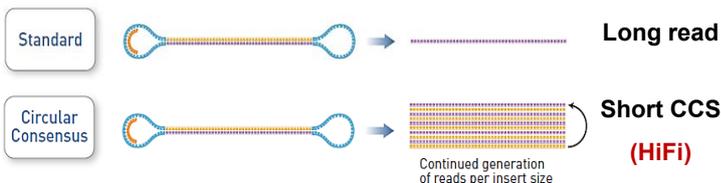
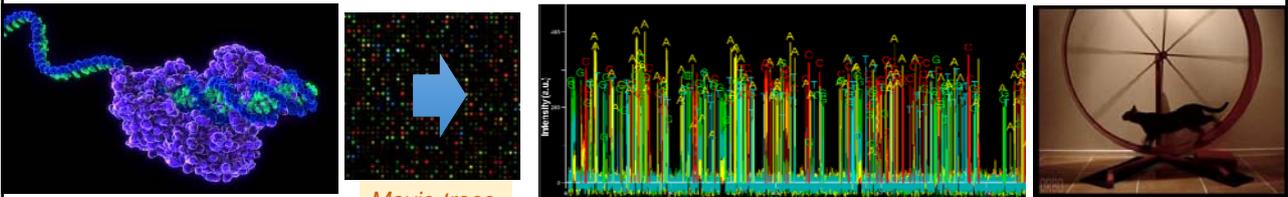
PacBio: 3rd-Gen SMRT Sequencing



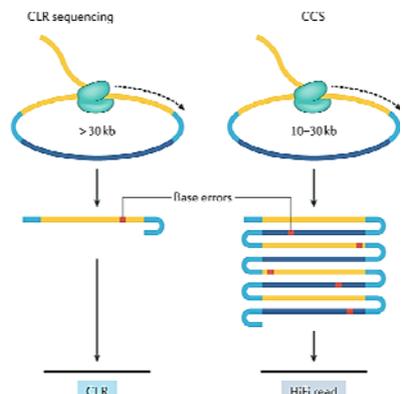
- **Single Molecular Real Time (SMRT) real-time technology**
- ZMW (zero-mode waveguides), a 100-nm hole with DNA/Polymerase complex immobilized at the bottom; recording fluorescence released from P-dNTP upon incorporation



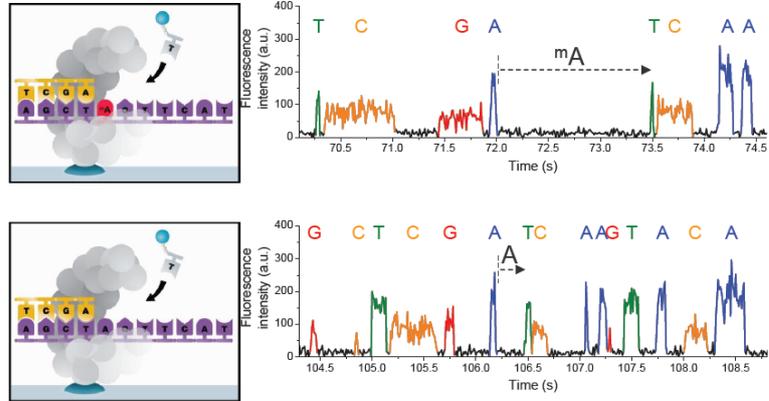
PacBio – HiFi CCS (circular consensus seq.)



CCS length: avg. 1-15kb, max. >100kb
Throughput: 10-30 Gb; CCS for HiFi accuracy (>Q20)

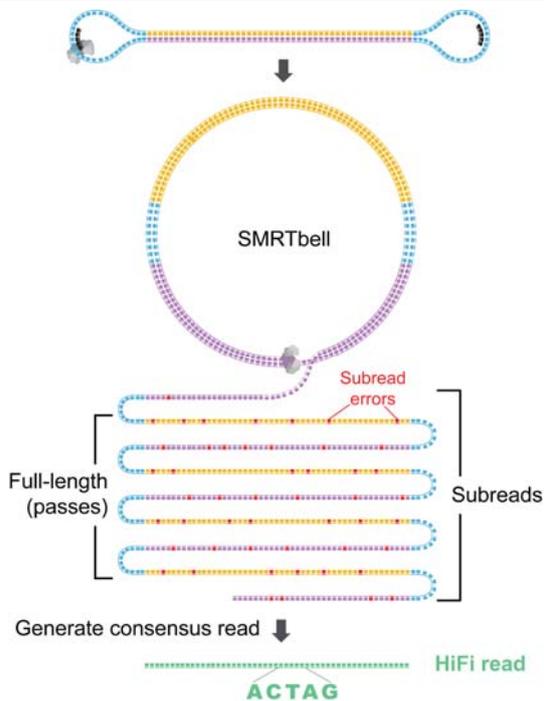


Key Feature: Kinetic Information



- Differentiation between modified and non-modified bases
 - Epigenetics, DNA damage, New, novel modifications
- Direct observation (*e.g.* no bisulfite)

29

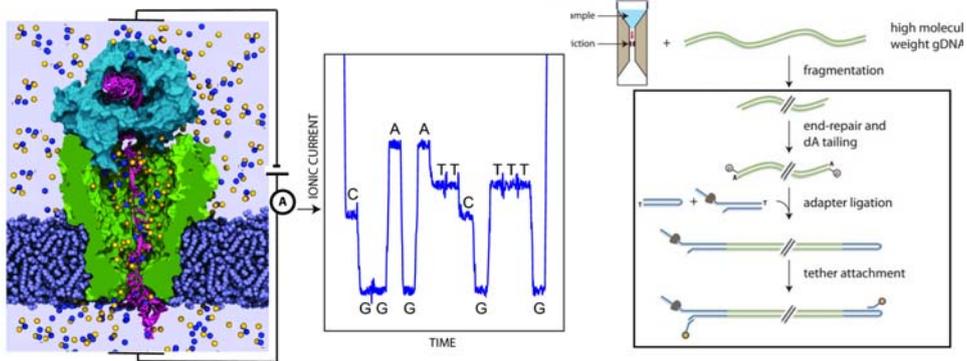


With the CCS / HiFi data...

What can we use it for?

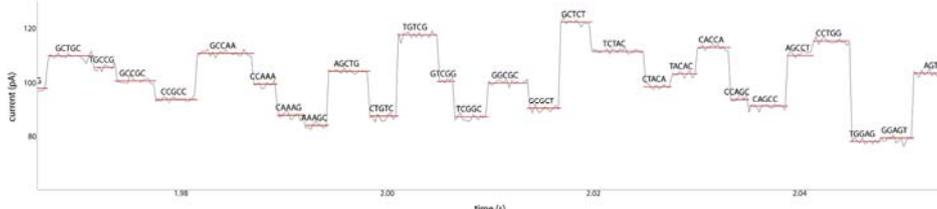
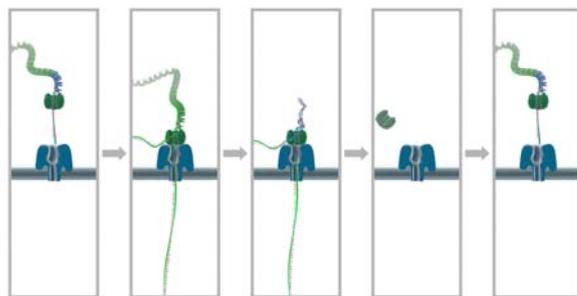
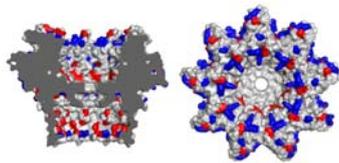
- *Methylation (m5C, m6A, etc)*
- *Resolve repeats*
- *Genome phasing*
- *Haplotyping*
- *RNA isoform*

NanoPore Sequencing Technology



31

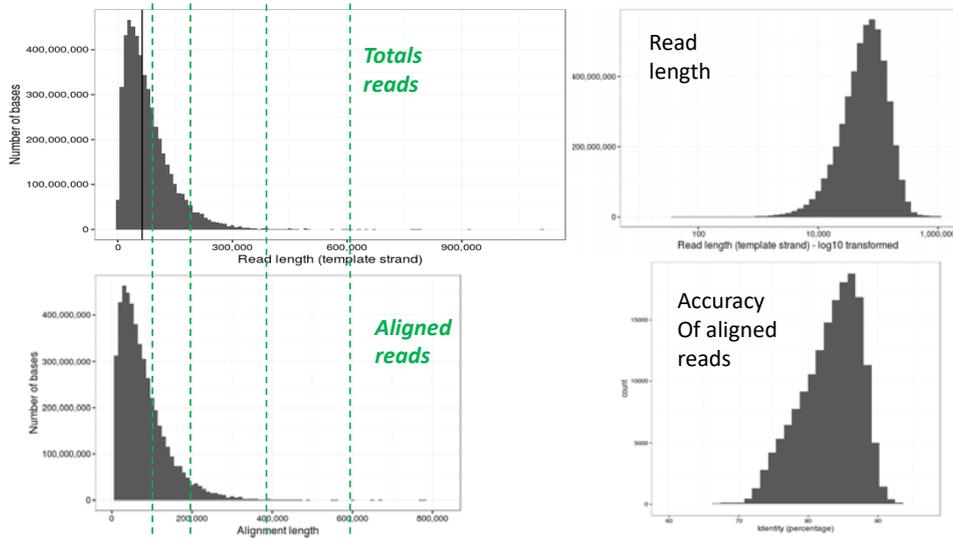
3rd-Gen: Oxford Nanopore (DNA, RNA, protein)



genome assembly using MinION reads [version 1]. F1000Research 2017, 6:1083 (doi: 10.12688/f1000research.12012.1)

F1000Research

E. coli: on MonION flowcell v9.4



Source: Loman Lab

sequencing **Running...**
 Run Statistics
READS: 409.20K ESTIMATED BASES: 1.57GB

Experiment group: 057_Desobinier_amlonon Sample ID: 057_Desobinier_amlonon Flow cell ID: FAH70235 Flow cell product code: FLO-MIN105 Kit ID: BQK-LBK105
 Current output directory: C:\data\reads

Channels Panel

Live status of each channel's state during sequencing

Total Estimated Bases

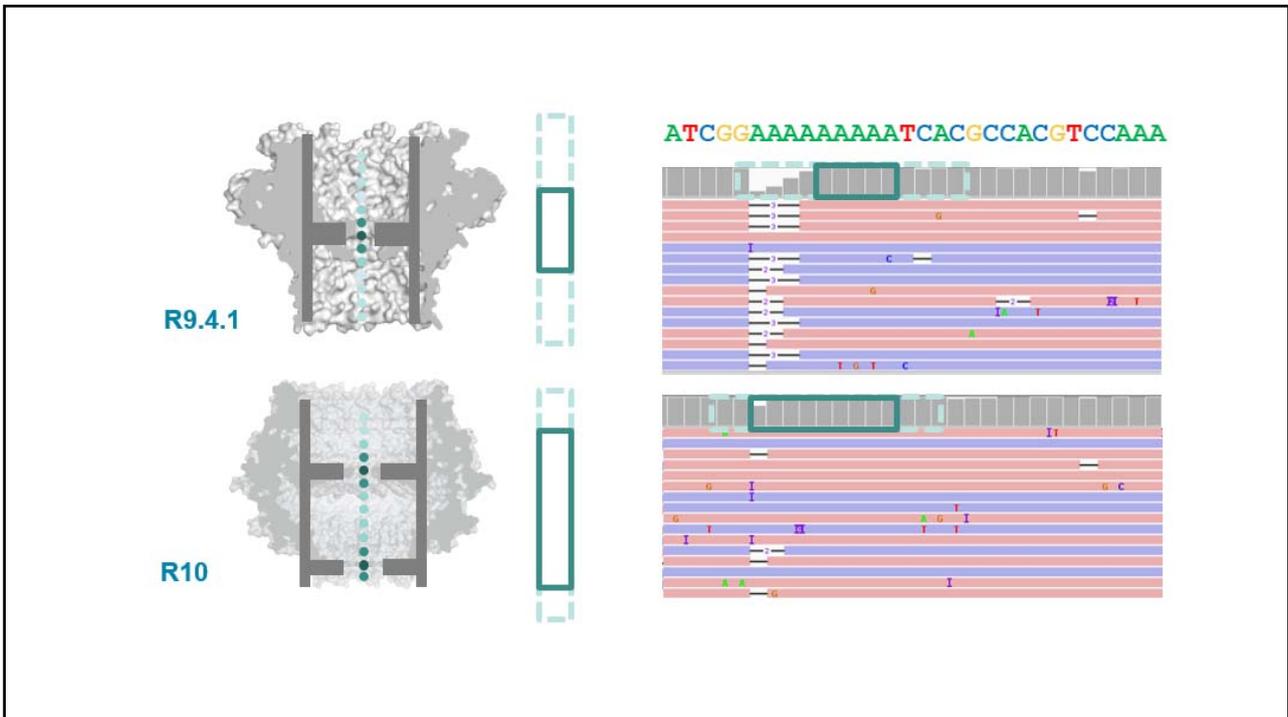
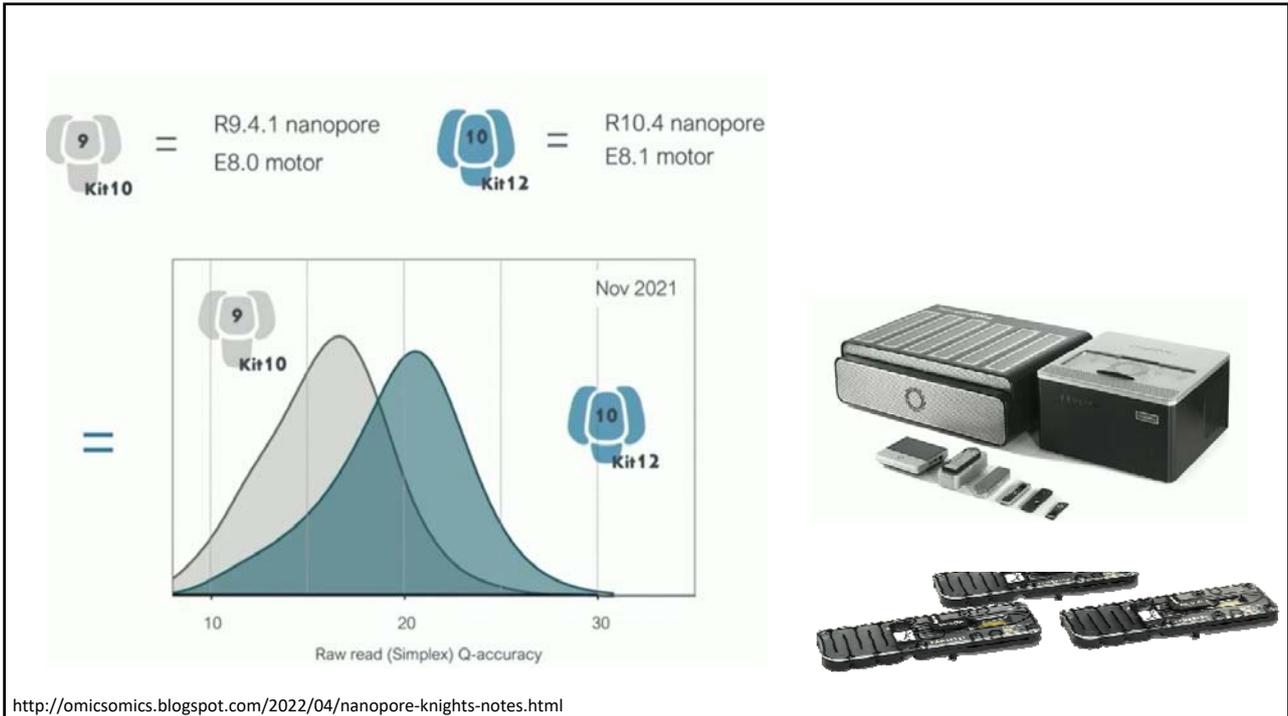
Estimated Read Length

Messages

Device

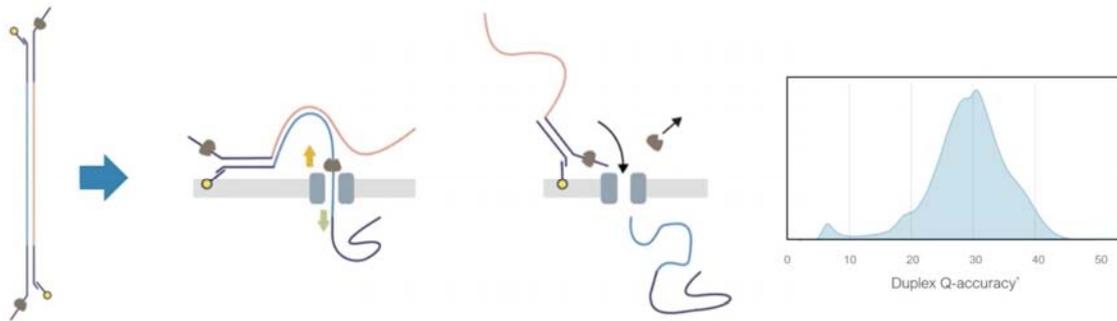
- Starting Sequencing (MN20200) 2 hours ago
- Reached target temperature (MN20200) 2 hours ago
- waiting for temperature to be within acceptable bounds (MN20200) 3 hours ago
- Experimental Parameters Complete (MN20200) 2 hours ago
- Setting Experimental Parameters (MN20200) 2 hours ago
- calibration finished successfully (MN20200) 3 hours ago
- Starting Calibration (MN20200) 2 hours ago
- Experimental Parameters Complete (MN20200) 3 hours ago
- Finished Mux Scan (MN20200) 2 hours ago

Detecting Base Modifications Using Nanopore Sequencing



Maize B73 Oxford Nanopore duplex sequence data release

April 4, 2022

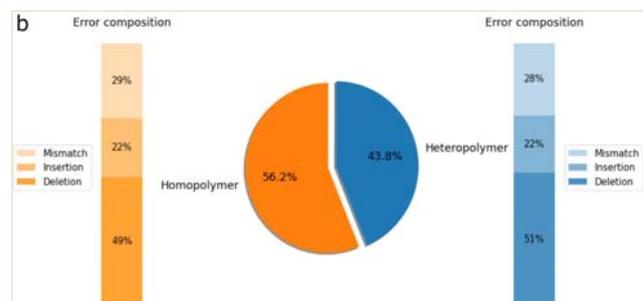
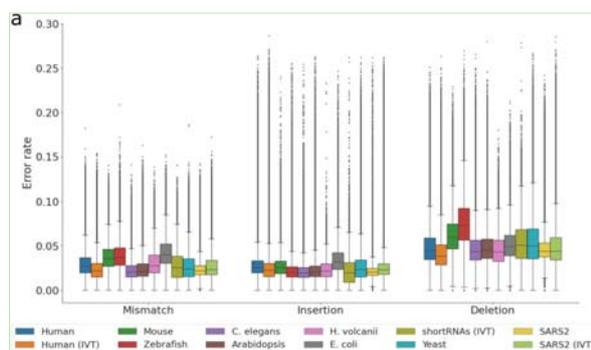


<https://nanoporetech.com/resource-centre/clive-brown-ncm-update-2021>

Figure 1: the principle of nanopore duplex reads resulting in raw read accuracies approaching Q30 accuracy.

Sequencing accuracy and systematic errors of nanopore direct RNA sequencing

Wang Liu-Wei, Wiep van der Toorn, Patrick Bohn, Martin Hölzer, Redmond Smyth, Max von Kleist



<https://doi.org/10.1101/2023.03.29.534691>

Current Major NGS Platforms



Applications

- Genome Assembly
- Variant study
- Transcriptome
- Metagenome
- Population typing
- Disease genomics

Illumina
Reversible terminator



HiSeq, MiSeq, NextSeq2000, iSeq

PacBio
SMRT



Oxford NanoPore



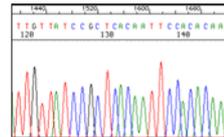
GrinION (x5)



Evolution of Sequencing Technologies

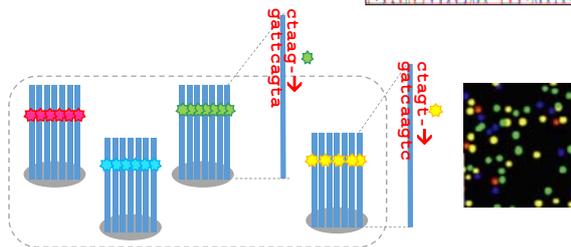
Dideoxy nt.
1 rxn/tube

gctagttgaccttgaccaagcatggcgatcgat
cgatca---->



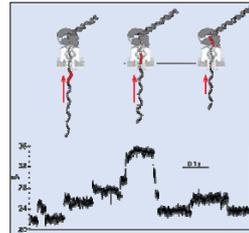
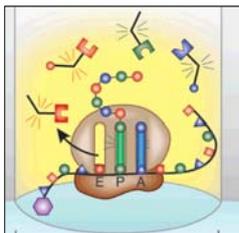
Sanger seq.

2nd-Gen
Clonal amplification



Roche454
Illumina
Ion Proton

3rd-Gen
Single mol. Seq.



PacBio
Oxford Nanopore

NGS Platforms & Features (Y2023)



	Illumina HiSeq 2500	Illumina MiSeq	Illumina NextSeq 2000	PacBio Sequel & SQ IIe	Oxford ONT GridION
Chemistry		Cyclic reversible terminator Of amplified DNA clusters		SMRTbell-tech; DNA polymerization	Electrical current passing through a nanopore channel
Output/run	HT mode: 1.2 Tb Rapid mode: 150 Gb	up to 15 Gb	P2:120 Gb P3*:300 Gb	CLR: 300-650 Gb HiFi: 15-40 Gb	Current: 5-30 Gb
Max Read length	2*250nt	2*300nt	2*150nt	CLR: max>100kb, HiFi: max>50kb	1-50 kb (max>200kb)
# Fragments /Chip	300-400 M (Rapid) 2,000-2,300M (HT)	12-15 M (v2) 20-25M (v3)	P2: 500M P3: 1400M	SQ: 1M SQ IIe: 8M	30-300 K / chip
Data quality	> 99.9%; Tolerate homopolymer; sensitive to high GC	> 99.9%; Tolerate homopolymer; sensitive to high GC		Raw 85-89%; HiFi ~99.9%; Random homopolymeric errors; tolerate high GC%	Raw 80~94%; Systematic homopolymeric errors; tolerate high GC%
Application	De novo assembly; Re-sequencing; RNA-seq	De novo assembly; Re-sequencing; amplicon	De novo assembly; Re-sequencing; RNA-seq	Genome assembly; structural variation; phasing; Iso-Seq	Genome assembly; structural variation; phasing; RNA/DNA-seq

Review

CellPress

Ten years of next-generation sequencing technology

Erwin L. van Dijk¹, H el ene Auger¹, Yan Jaszczyszyn², and Claude Thermes¹¹Centre de G en etique Mol culaire – CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France
²Plateforme Int egr ee IMAGIF – CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France

Trends Genet. 2014 Sep;30(9):418-26. doi: 10.1016/j.tig.2014.07.001.

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin¹, John D. McPherson² and W. Richard McCombie¹[Nature Reviews Genetics](#) volume 17, pages333–351 (2016)

Review of Clinical Next-Generation Sequencing

Sophia Yohe, MD; Bharat Thyagarajan, MD, PhD

Context.—Next-generation sequencing (NGS) is a technology being used by many laboratories to test for inherited disorders and tumor mutations. This technology is new for many practicing pathologists, who may not be familiar with the uses, methodology, and limitations of NGS.

Objective.—To familiarize pathologists with several aspects of NGS, including current and expanding uses; methodology including wet bench aspects, bioinformatics, and interpretation; validation and proficiency; limitations; and issues related to the integration of NGS data into patient care.

Data Sources.—The review is based on peer-reviewed literature and personal experience using NGS in a clinical setting at a major academic center.

Conclusions.—The clinical applications of NGS will increase as the technology, bioinformatics, and resources evolve to address the limitations and improve quality of results. The challenge for clinical laboratories is to ensure testing is clinically relevant, cost-effective, and can be integrated into clinical care.

(*Arch Pathol Lab Med.* 2017;141:1544–1557; doi: 10.5858/arpa.2016-0501-RA)

Arch Pathol Lab Med. 2017;141:1544–1557; doi: 10.5858/arpa.2016-0501-RA

NGS Data processing

- Data types
- Data QC

43

Types and Characteristics of NGS Reads

- Read length:

Short

50-300bp

Long

500-15,000bp

- Read types:

SR

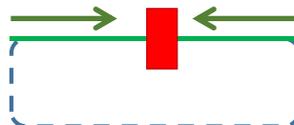


50bp-20kb

PE

50-300 bp;
1~1.5 kb jump

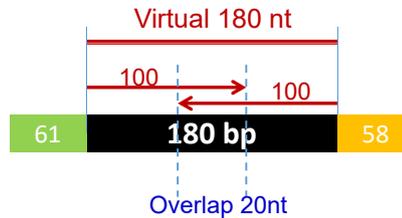
MP

50-300bp;
2~15kb jump

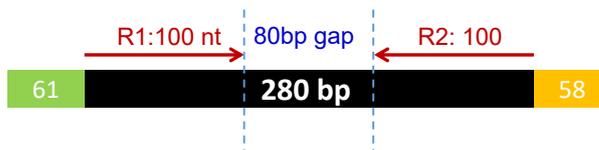
44

Insert size vs Library Fragment Size

300-bp fragment:
Ends overlapped 20 bp



400-bp fragment:
Ends gapped by 80 bp



Illumina Read – fastQ file

```

Sequence header Machine ID, FC ID Lane ID Index sequence
no control
Y/N: failing PF or not
Read1 or Read2
@HWI-D00368:32:H8R31ADXX:2:1101:2034:2140 1:N:0:CAGATC
TTTGNCGAGAACTGGAATTGAACCAATATTTAAGTCTTACAAGGAATTCGTTTTAAC
+
@@@F#2ADFDHHHJJJJGHHIIJIIJJJIJGGJHEIIJIIJIIJIIJJJJJIGI
Q-score header

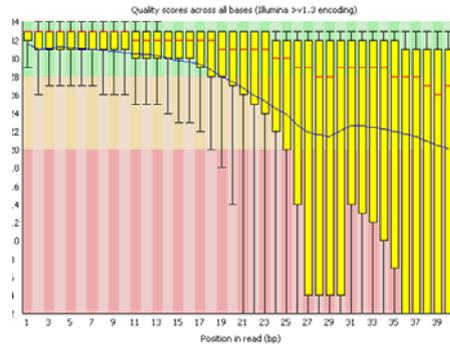
```

Seq. performance assessment – Base Q

Phred quality scores Q: logarithmically related to **error probabilities**

$$P \text{ by } Q = [-10 * \log_{10}(P)]$$

Phred Score Q	Error probability	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%



FastQC Report

Summary

- Basic Statistics
- Per base sequence quality
- Per sequence quality scores
- Per base sequence content
- Per base GC content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Kmer Content

Basic Statistics

Measure	Value
Filename	good_sequence_short.fastq
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Filtered Sequences	0
Sequence length	40
%GC	45

Per base sequence quality



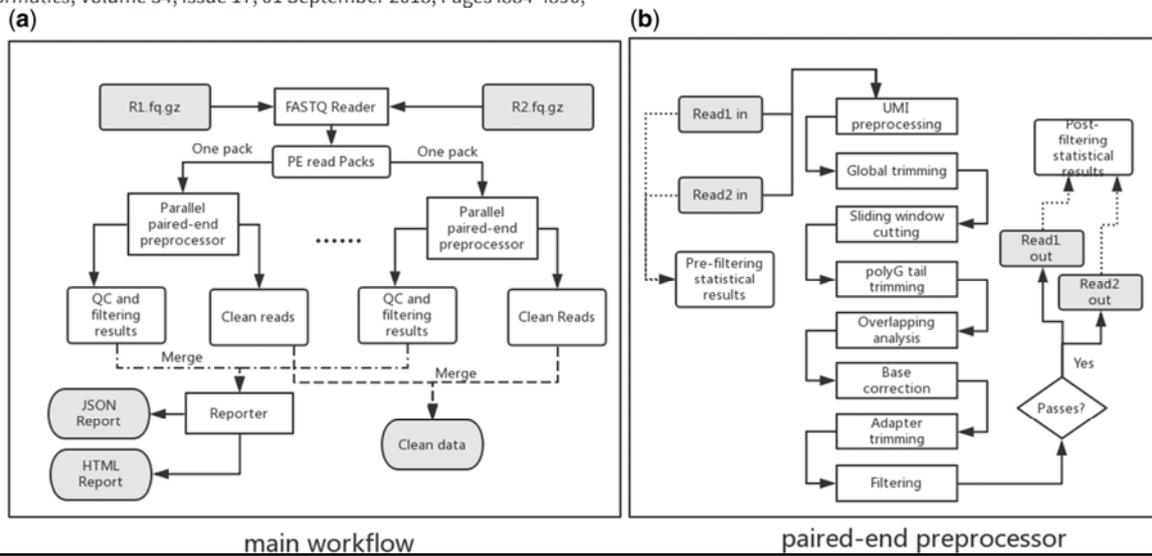
<https://wiki.hpc.cmu.edu/display/Bioinfo/FastQC+Tutorial>

fastp: an ultra-fast all-in-one FASTQ preprocessor

<https://doi.org/10.1093/bioinformatics/bty560>

Shifu Chen ✉, Yanqing Zhou, Yaru Chen, Jia Gu

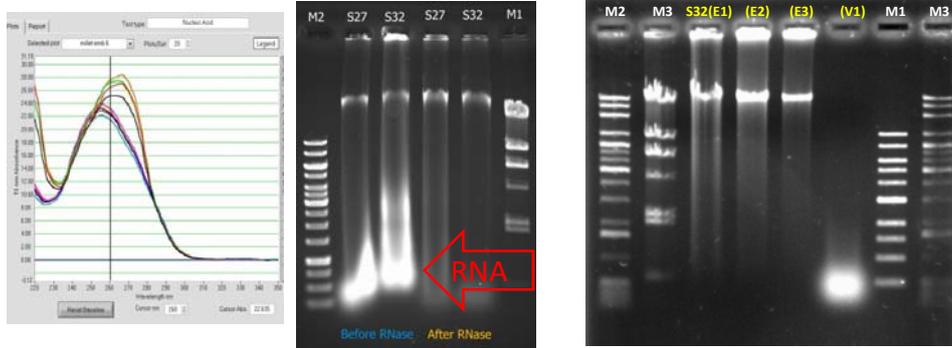
Bioinformatics, Volume 34, Issue 17, 01 September 2018, Pages i884–i890,



II. Principles of common NGS applications

- **Sample QC**
- **DNA applications**
- **RNA applications**

Genomic DNA QC



	OD 260/280	OD 260/230	NanoDrop (ng/uL)	Qubit DNA (ng/uL)	RNA Carry over (NanoDrop/Qubit)
S32-original	2.04	2.05	2250.8	56.0	40.19 X
S32_V1	2.16	2.52	1207.40	7.29	165.62 X
S32_E1	1.77	0.94	394.50	131.00	3.01 X
S32_E2	1.67	0.82	45.06	14.10	3.20 X
S32_E3	1.75	0.75	11.48	4.49	2.56 X

RNA QC

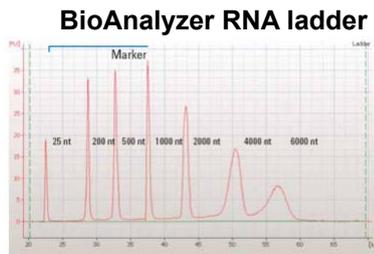
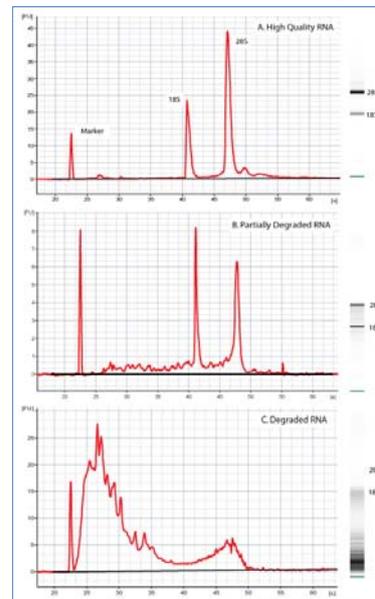
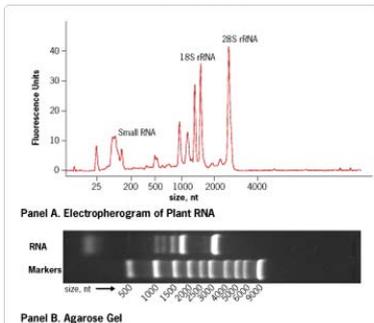


Figure 1 RNA 6000 Nano ladder

Human RNA – various degradation



Plant total RNA



Genome Sequencing

- *De novo*
- *Re-sequencing*
- *Amplicon-seq*

53

Sequencing vs re-sequencing

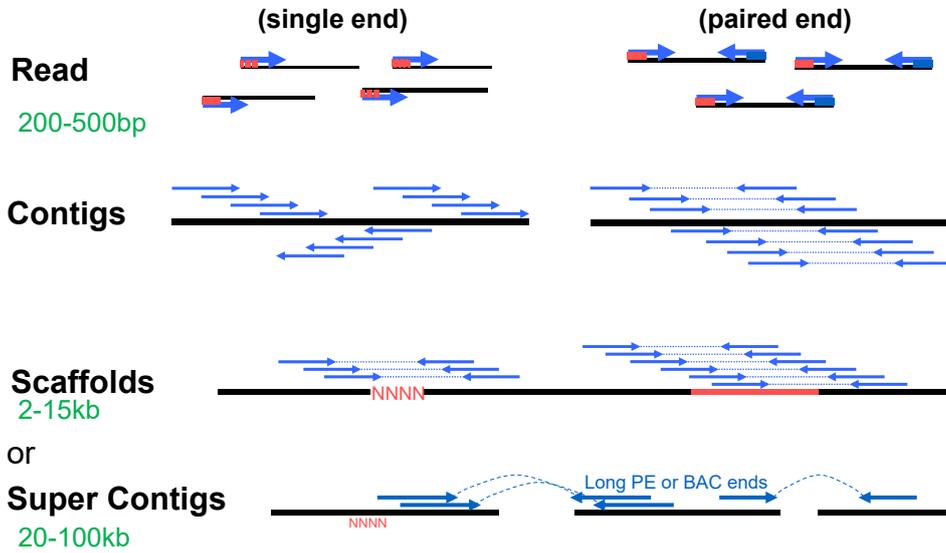
“de novo” sequencing:
no reference genome available



Resequencing:
reference genome available

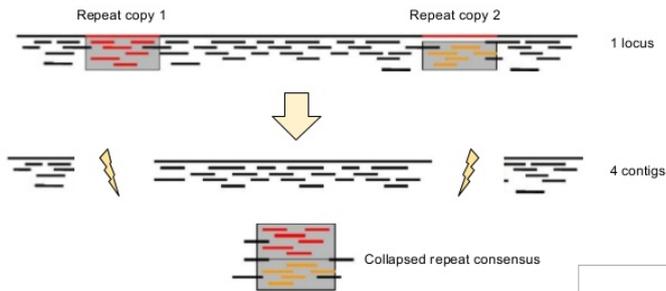


Hierarchical Genome Assembly

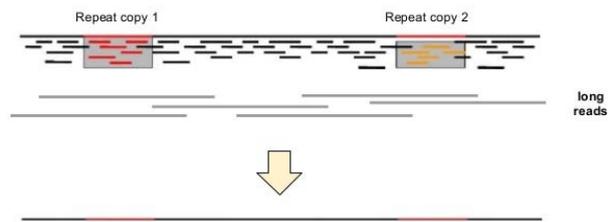


55

Repeats



Long reads can span repeats



https://www.slideshare.net/torstenseemann/long-read-sequencing-wehi-bioinformatics-seminar-tue-16-june-2015?from_action=save

Re-sequencing: Variant detection

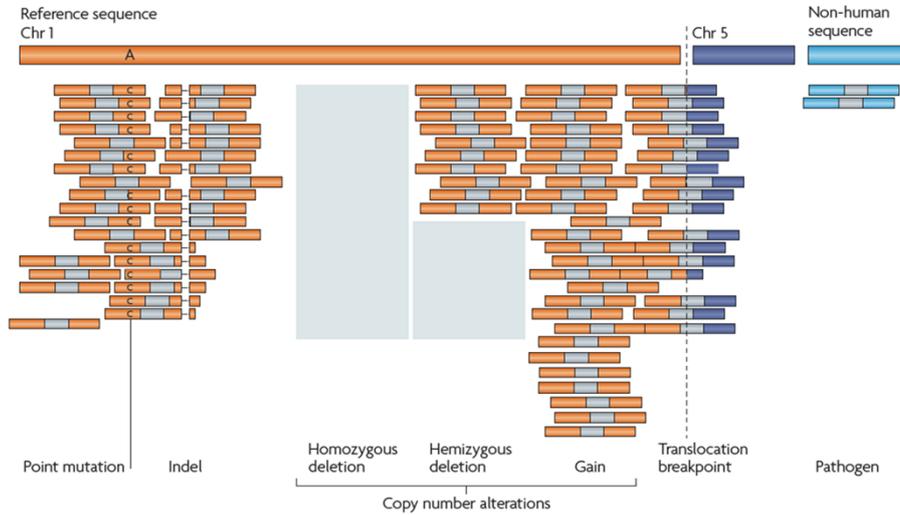
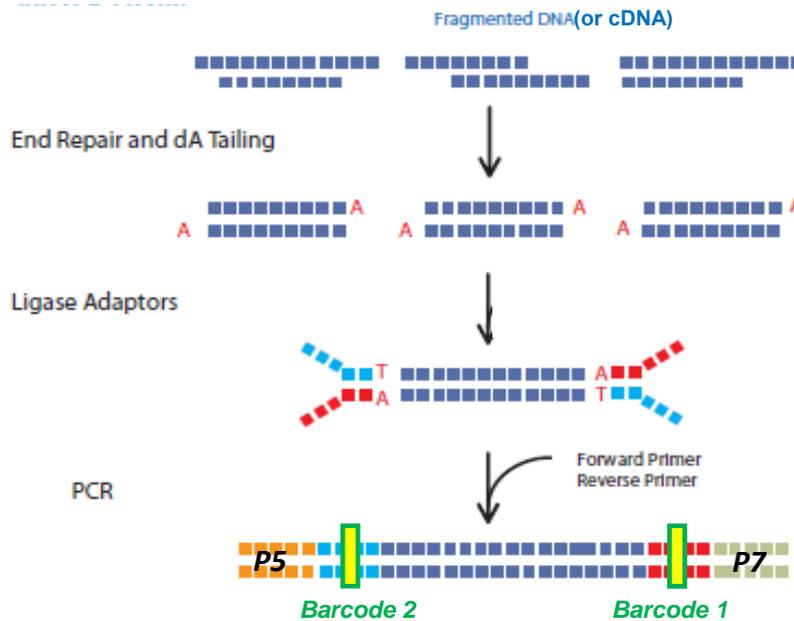


Figure 3 | Types of genome alterations that can be detected by second-generation sequencing. Sequenced

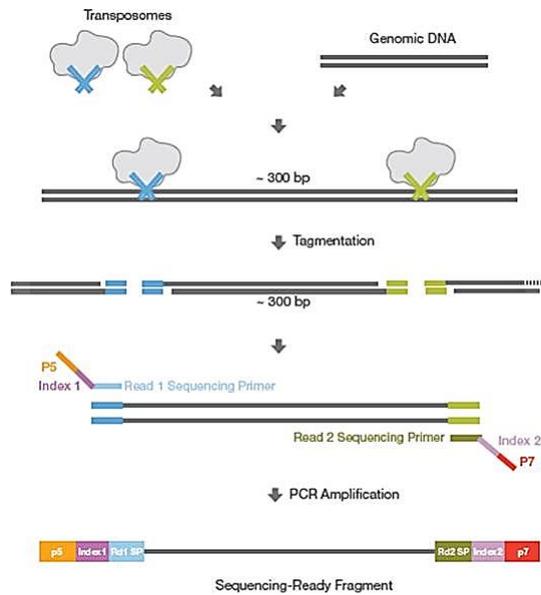
Meverson et al. NRG 2010

1. Shotgun gDNA Library Preparation



58

2. Tn tagging - Nextera Library prep



1. Templated DNA + transposome complex (contain adaptor)

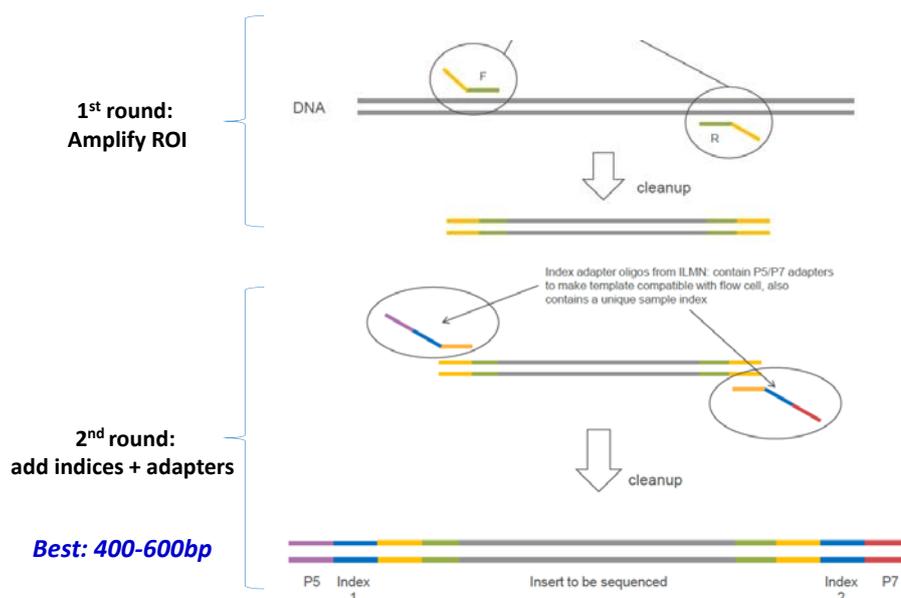
2. Tagmentation breaks DNA and add adaptor to ends

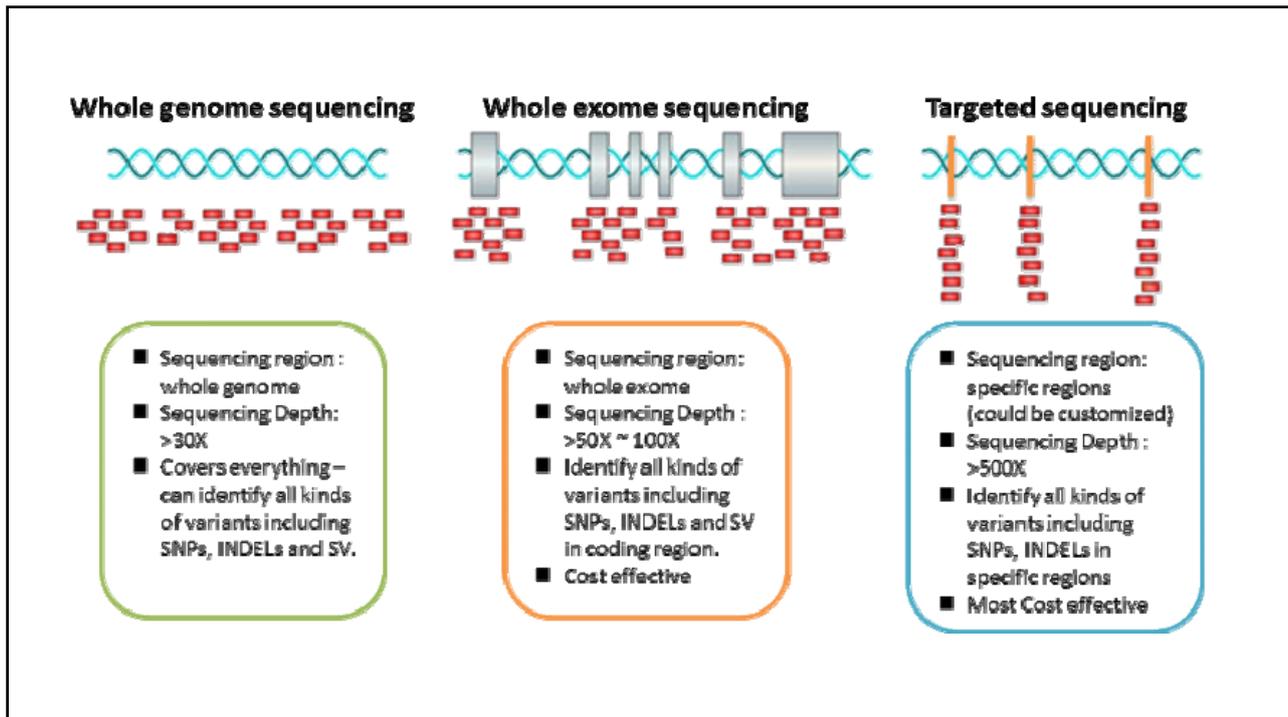
3. PCR amplification to engineer barcode and sequencing primers

http://www.gtbiotech.com.tw/products/Nextera_XT_DNA.asp

59

3. Amplicon library: targeted PCR





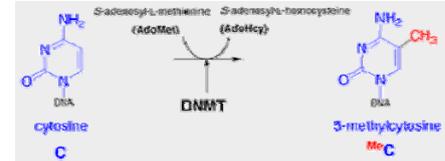
Epigenetic sequencing

- *Bisulfite-seq*
- *ChIP-seq*
- *DAP-seq*
- *ATAC-Seq*

Genome methylation: Bisulfite-seq

Watson >>**AC^mGTT**C**GCTT**G**AG>>
Crick <<**TG**C**^mAAG**C**GAA**C**T**C**<<****

C^m methylated
C Un-methylated



1) Denaturation



Watson >>**AC^mGTT**C**GCTT**G**AG>>**

Crick <<**TG**C**^mAAG**C**GAA**C**T**C**<<**

2) Bisulfite Treatment



BSW >>**AC^mGTT**U**G**U**TT**G**AG>>**

BSC <<**TG**C**^mAAG**U**GAA**U**T**U**<<**

3) PCR Amplification



BSW >>**AC^mGTT**T**GTTT**G**AG>>**

BSC <<**TG**C**^mAAG**T**GAA**T**TT<<**

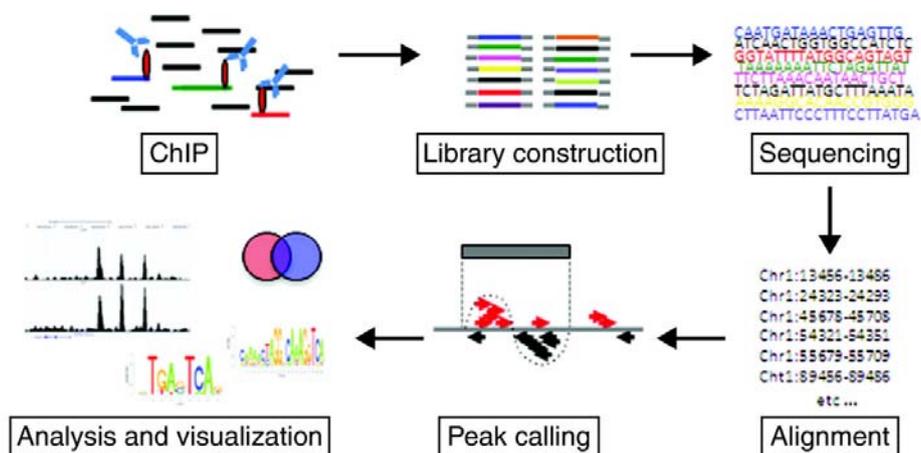
BSWR <<**TG CAA**A**CAA**A**ACTC<<**

BSCR >>**AC**G** TTC**A**CTT**A**AA>>**

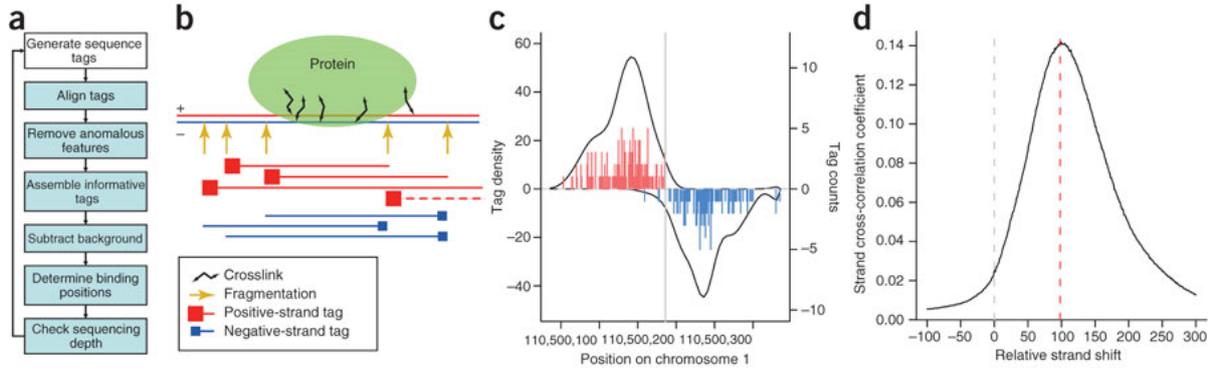
Compare proportion of C/T sequenced at known CpG sites to quantify % of methylated cells in the original sample.

BMC Bioinformatics volume 10
Article number: 232 (2009)

ChIP-seq procedure

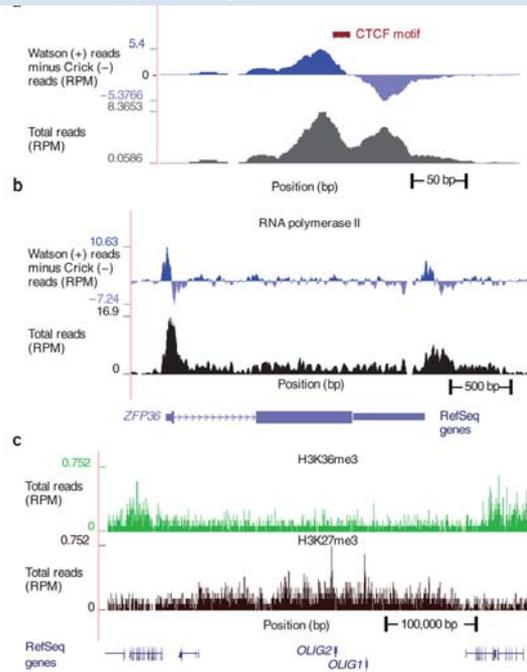


ChIP-Seq: peak calling by looping



Nature Biotechnology 26, 1351 - 1359 (2008)
 Design and analysis of ChIP-seq experiments for DNA-binding proteins

ChIP-seq peak types from various experiments



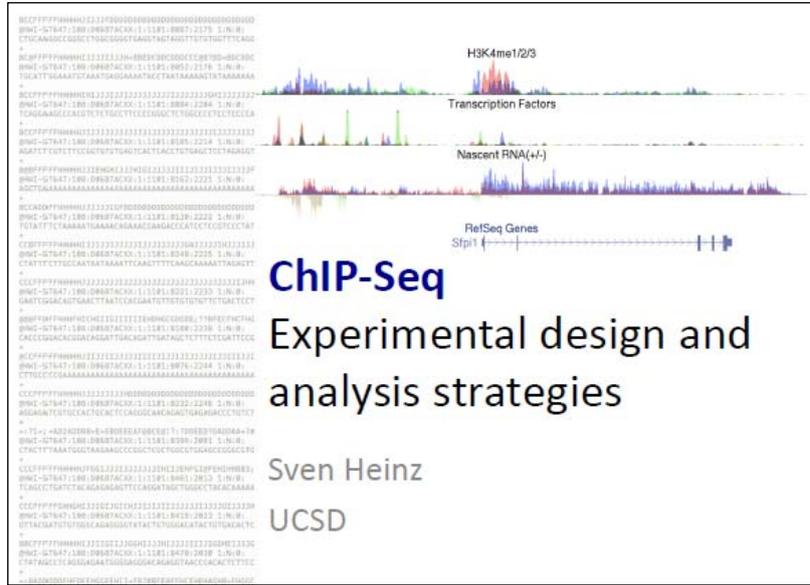
Specific TF

Txn complex

Chromatin remodeling

Nature Methods 6, S22 - S33 (2009)

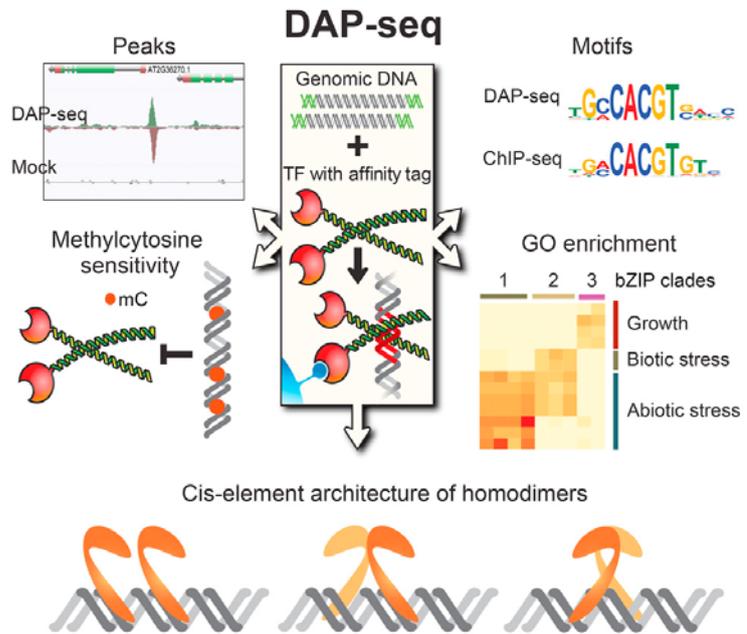
ChIP-seq Experimental Design



http://pharmacology.ucsd.edu/graduate/courseinfo/BIOM231-SP13_8_ChIPseq.pdf

Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape.
Ronan C. O'Malley, et al. Cell (2016)

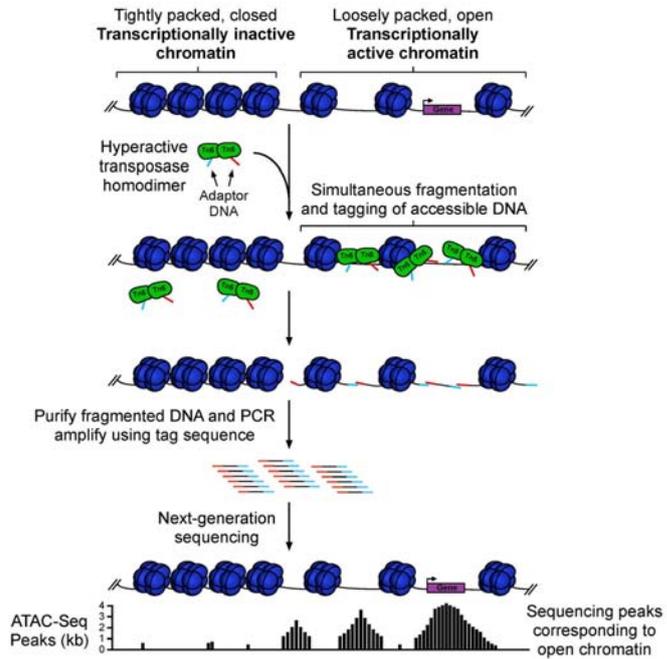
DNA affinity purification sequencing (DAP-seq)



Cell 165.5 (2016): 1280-1292.

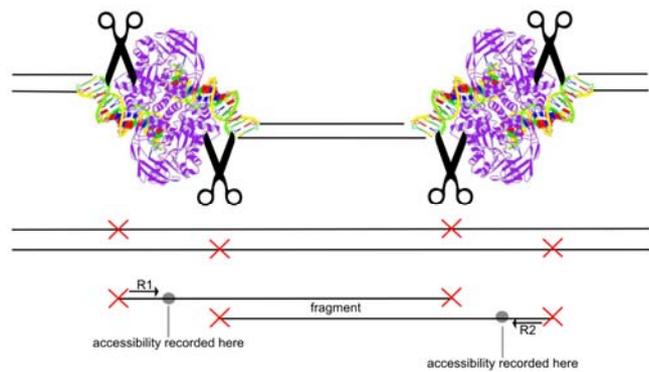
ATAC-seq

Assay for Transposase-Accessible Chromatin using sequencing

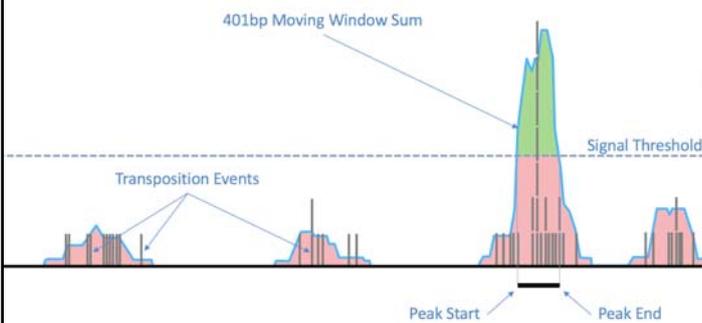


<https://en.m.wikipedia.org/wiki/ATAC-seq>

Epigenomics by ATAC-seq: transposition at open chromatin



Transposase image accessed from the Protein Data Bank, <https://rcsb.org/structure/1MUH>



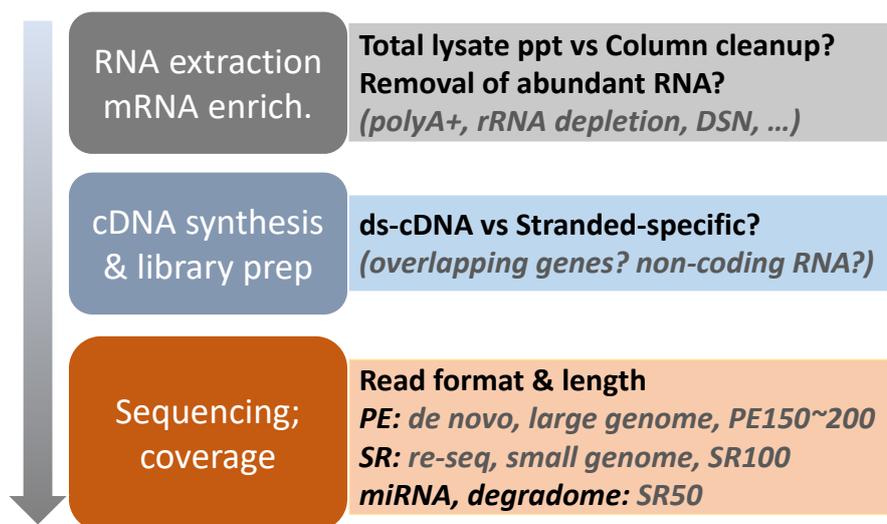
<https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/algorithms/overview>

Transcriptome sequencing

- *mRNA vs stranded*
- *smRNA, non-coding RNA*
- *LR-RNA-seq*

71

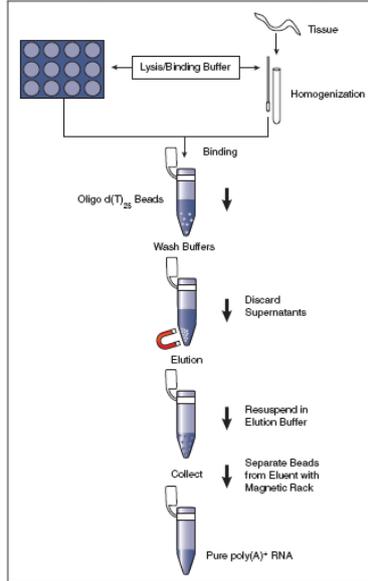
RNA-seq: considerations



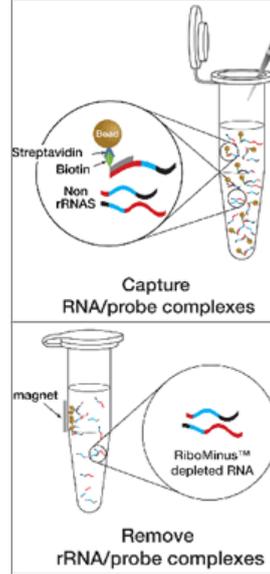
72

mRNA enrichment

Oligo-dT binding

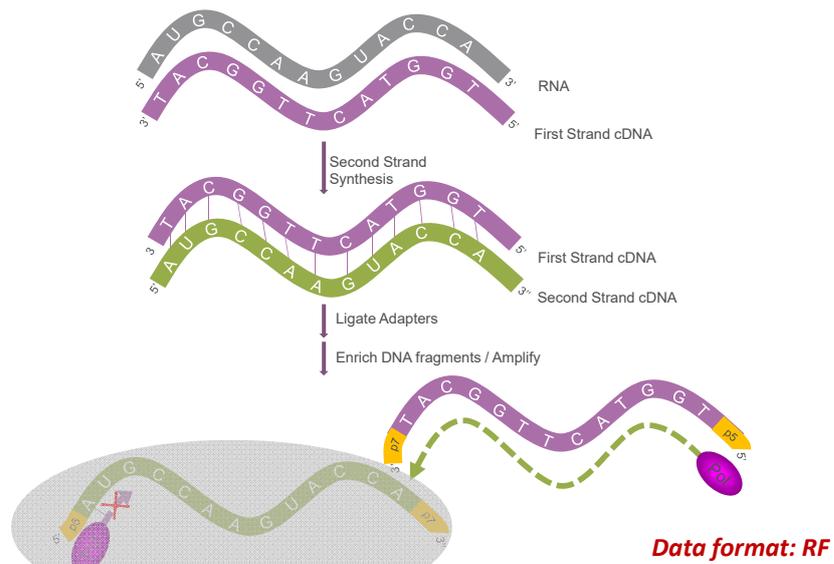


rRNA removal



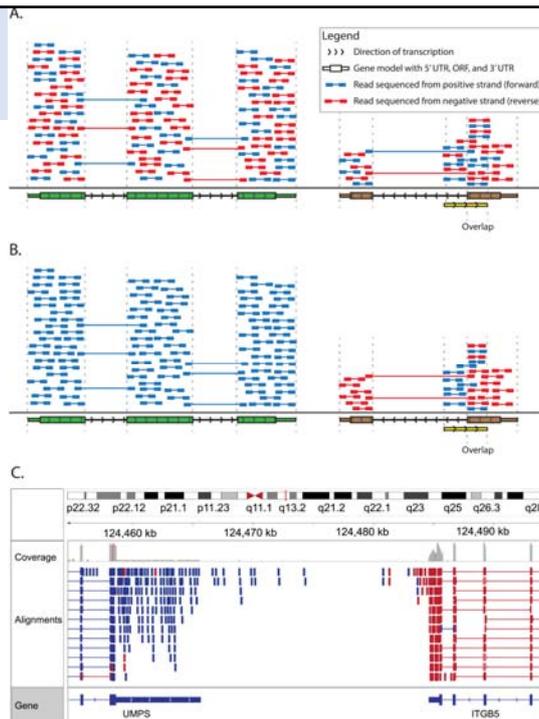
73

Strand-specific RNA-seq prep



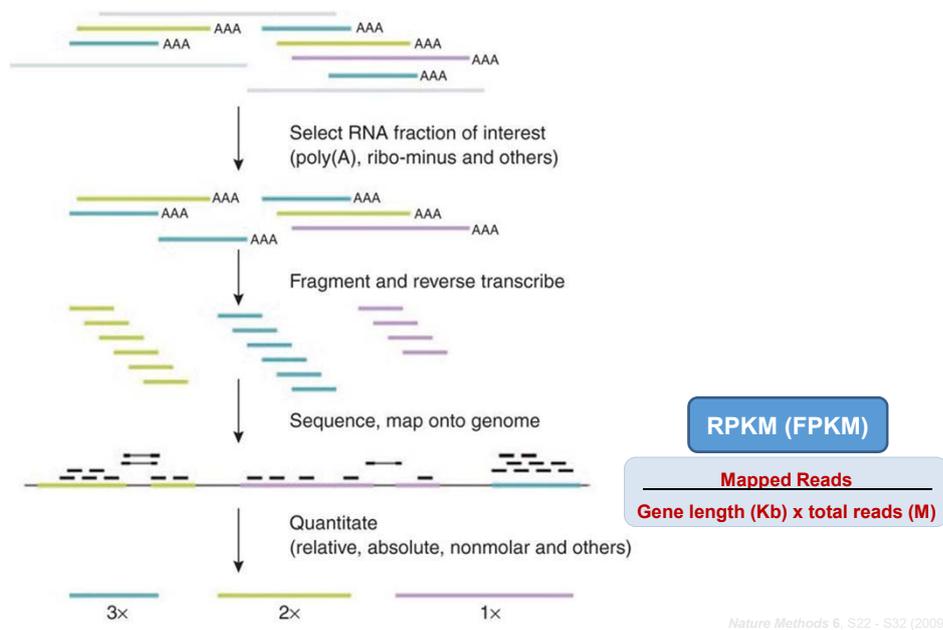
Source: Illumina

Comparison of stranded and unstranded RNA-seq library methods, and their influence on interpretation and analysis.



Griffith M, et al. (2015) Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. PLOS Computational Biology 11(8): e1004393. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004393>

Transcriptome profiling: RNA-seq



RPKM, FPKM, vs TPM

rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/

RPKM, FPKM and TPM, clearly explained

Posted by: RNA-Seq Blog in Data Normalization, Expression and Quantification, Statistical Analysis July 22, 2015
403,845 Views

from StatQuest

RPKM, FPKM and TPM, Clearly Explained!!!

RPKM, FPKM and TPM....

...Clearly Explained!!!

It used to be when you did RNA-seq, you reported your results in RPKM (Reads Per Kilobase Million) or FPKM (Fragments Per Kilobase Million). However, TPM (Transcripts Per Kilobase Million) is now becoming quite popular. Since there seems to be a lot of confusion about these terms, I thought I'd use a StatQuest to clear everything up.

These three metrics attempt to normalize for sequencing depth and gene length. Here's how you do it for RPKM:

- Count up the total reads in a sample and divide that number by 1,000,000 – this is our "per million" scaling factor.

<https://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/>

PUBLICATIONS TREND

RECENT RNA-SEQ PUBS

Prenatal diagnosis of Desbuquois dysplasia Type 1: Utilization of high-density SNP array to map homozygosity and identify the gene.

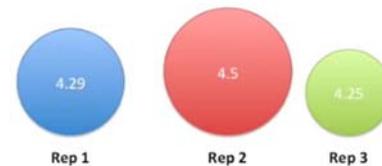
Coordinate regulation of virulence and metabolic genes by the transcription factor

RPKM vs TPM

Gene Name	Rep1 RPKM	Rep2 RPKM	Rep3 RPKM
A (2kb)	1.43	1.33	1.42
B (4kb)	1.43	1.39	1.42
C (1kb)	1.43	1.78	1.42
D (10kb)	0	0	0.009

RPKM:

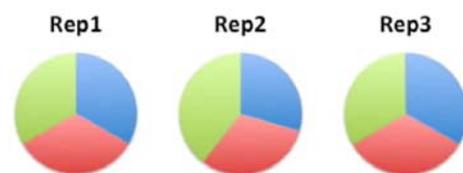
(1) Normalize each gene's read counts to sequencing depth (M), and (2) normalize for gene size (K);
-> each sample has different size of pie



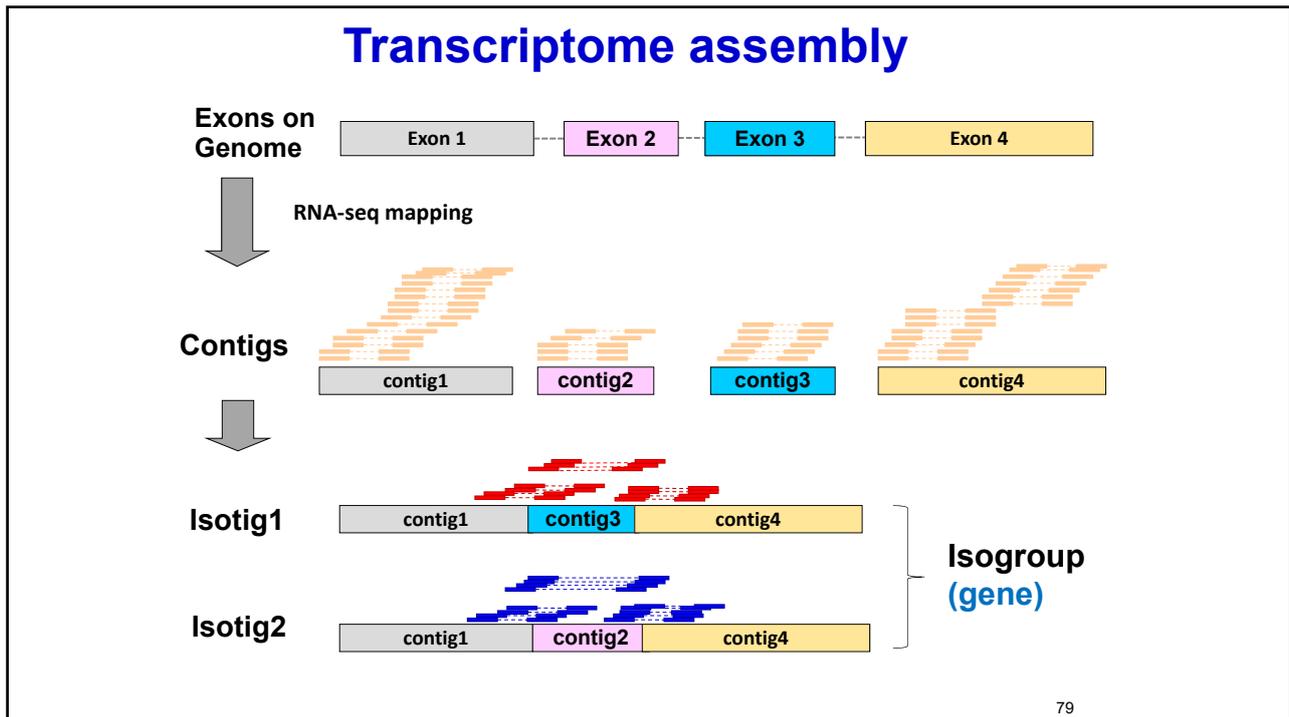
Gene Name	Rep1 TPM	Rep2 TPM	Rep3 TPM
A (2kb)	3.33	2.96	3.326
B (4kb)	3.33	3.09	3.326
C (1kb)	3.33	3.95	3.326
D (10kb)	0	0	0.02

TPM:

(1) Normalize each gene's reads to gene length RPK, (2) sum up all RPK/sample, then (3) divide each RPK/sum
-> each sample has same size of pie, and can compare the relative% of each gene's expn between samples



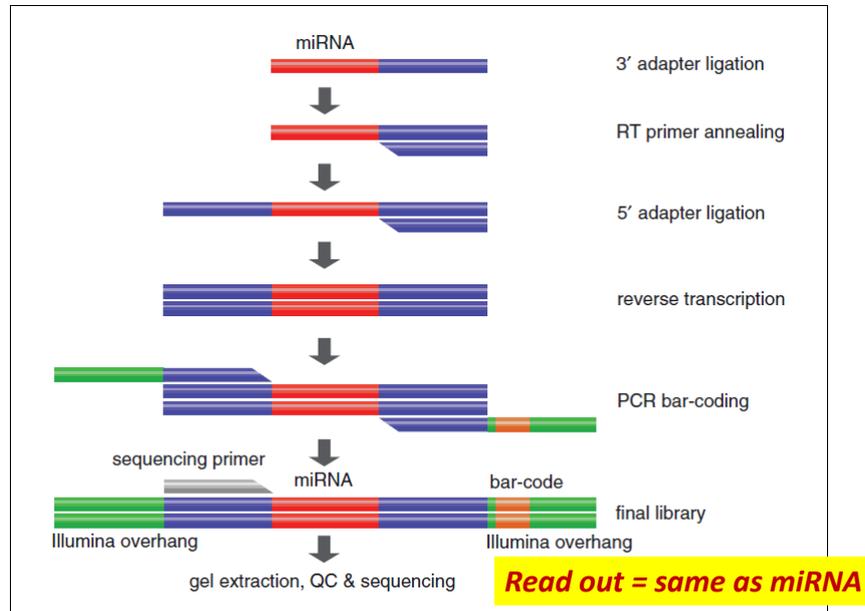
<https://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/>



Transcriptome Sequencing

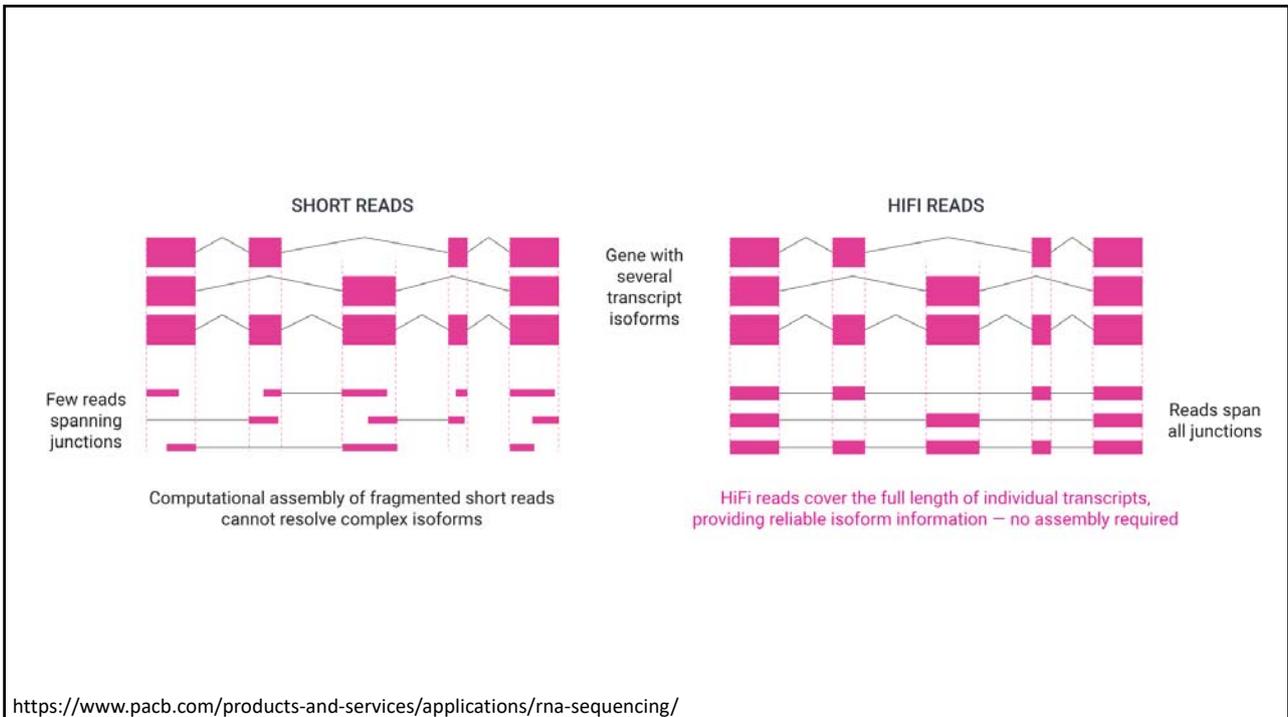
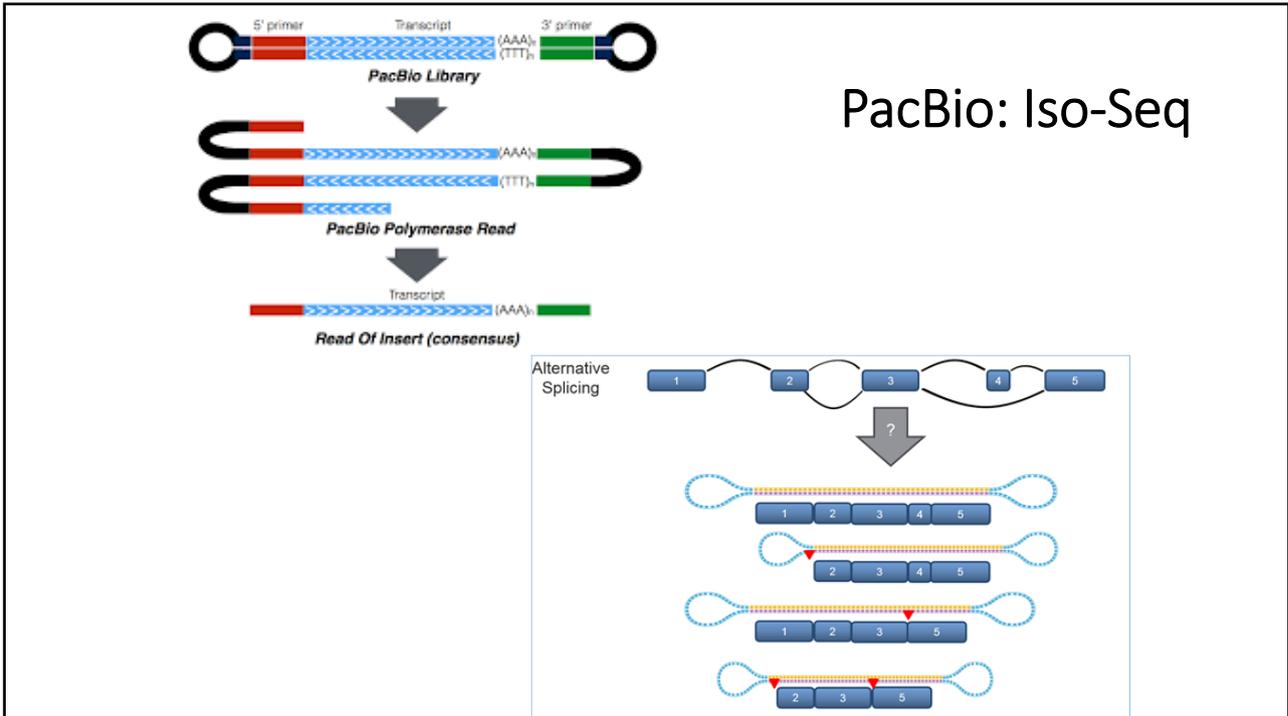
- **Sample:**
 - mRNA to pair with miRNA or lincRNA study?
- **mRNA enrichment method**
 - Oligo-dT vs Ribo-depletion
- **Controls and Biological replicates**
- **Time course?**
- **Re-sequencing (mapping) vs de novo assembly**
- **NGS:**
 - Shotgun: SR or PE
 - PE2*100~200bp
 - Coverage depends on need for detection sensitivity
- **Splicing variants? Fusion junction?**
 - Gel-size selection
 - PE2*150~300

smRNA library prep - Directional

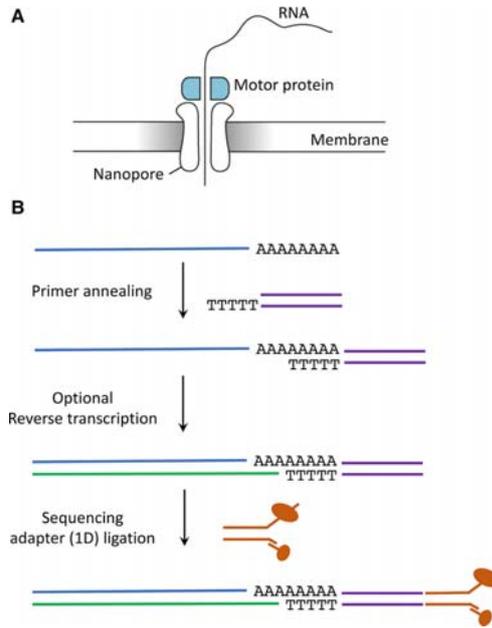


Transcriptome sequencing

- *mRNA vs stranded*
- *smRNA, non-coding RNA*
- ***LR-RNA-seq: PacBio & ONT***

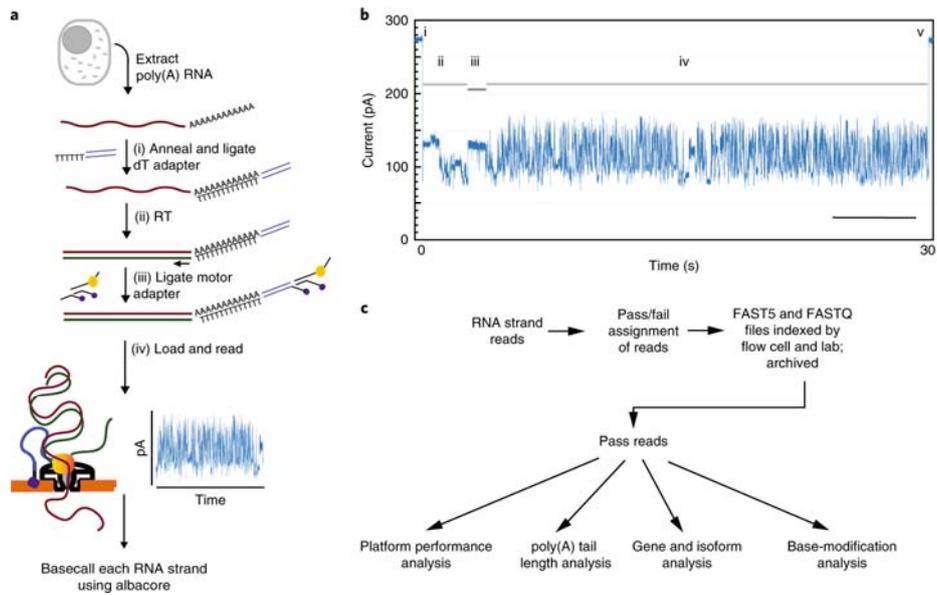


Direct RNA sequencing library preparation steps using Oxford Nanopore Technologies



The RNA modification landscape in human disease

Nanopore native RNA sequencing of a human poly(A) transcriptome



Nature Methods volume 16, pages1297–1305 (2019)

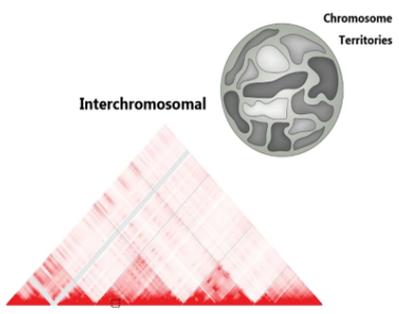
III. Advanced applications

- 3D Genomes
- Single-cell and spatial analyses

Advanced applications

- Chromatin dynamics (3D-DNA)
- Single-cell & Spatial analyses

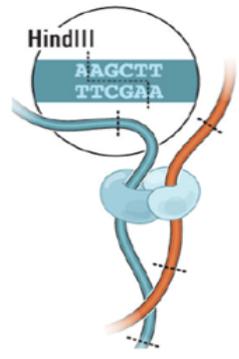
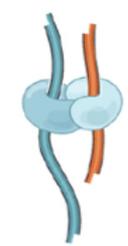
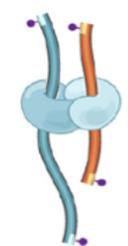
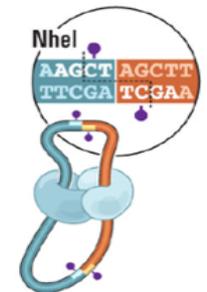
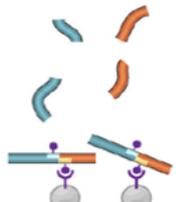
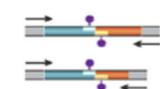
3D Genome Sequencing



Hi-C DNA Lib Prep (ARIMA) Chromosome Proximity Ligation	
Hi-C DNA Lib Prep (Dovetail) Chromosome Proximity Ligation	
Hi-C DNA Lib Prep (Phase Genomics) Chromosome Proximity Ligation	
Micro-C DNA Lib Prep (Dovetail) Chromatin Topology to Nucleosomal Positioning	

<http://ngs.biodiv.tw/NGSCore/illumina-system/>

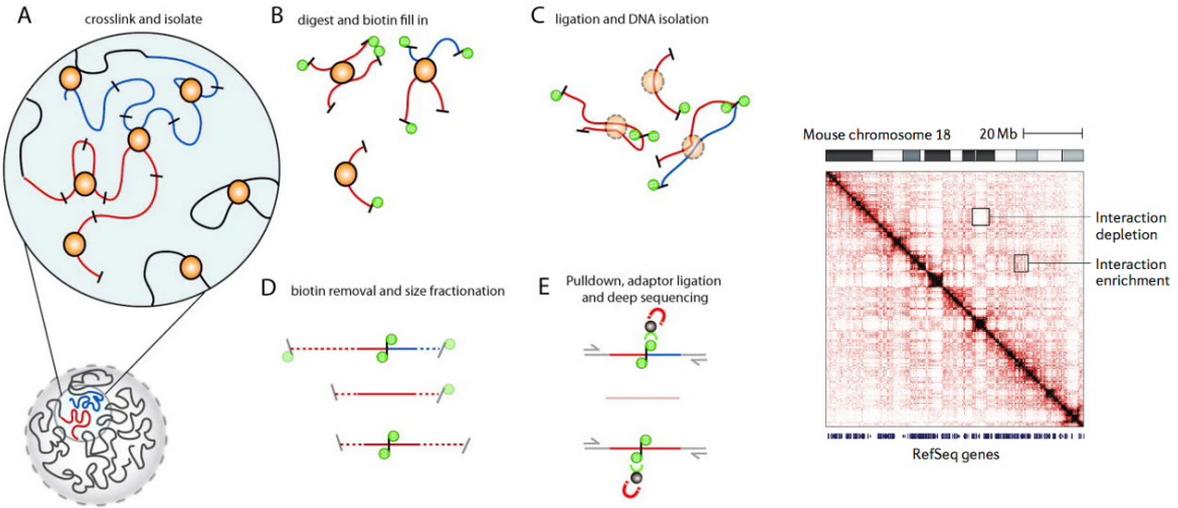
Hi-C: Proximity Ligation

- 1**
 Crosslink DNA

- 2**
 Cut with restriction enzyme

- 3**
 Fill ends and mark with biotin

- 4**
 Ligate

- 5**
 Purify and shear DNA; pull down biotin

- 6**
 Sequence using paired-ends


<https://doi.org/10.3791/1869>

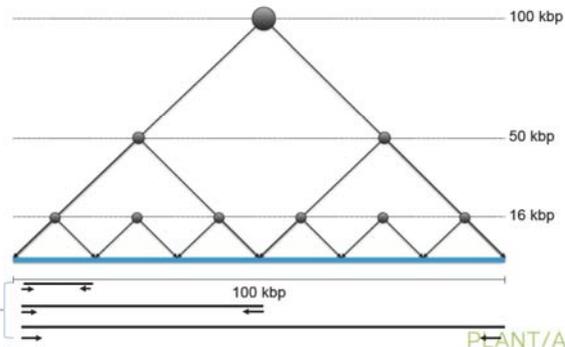
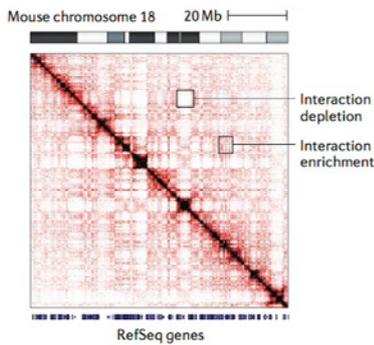
Hi-C: all-by-all interactions

- Note: Biotin labels fragment ends without interfering with folding of bound proteins, DNA



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC187486/>

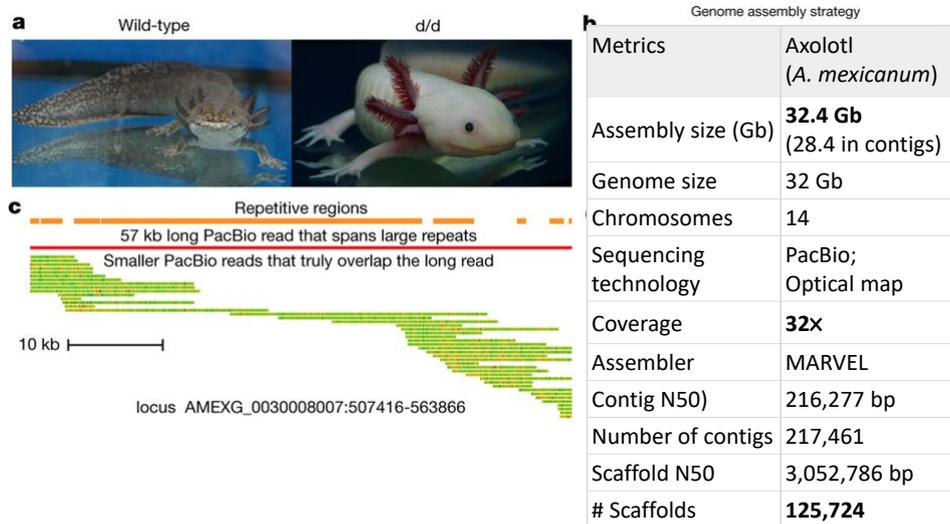
Methods: proximity Ligation Approaches



Read pairs of varying lengths allow you to create a scaffold of the genome

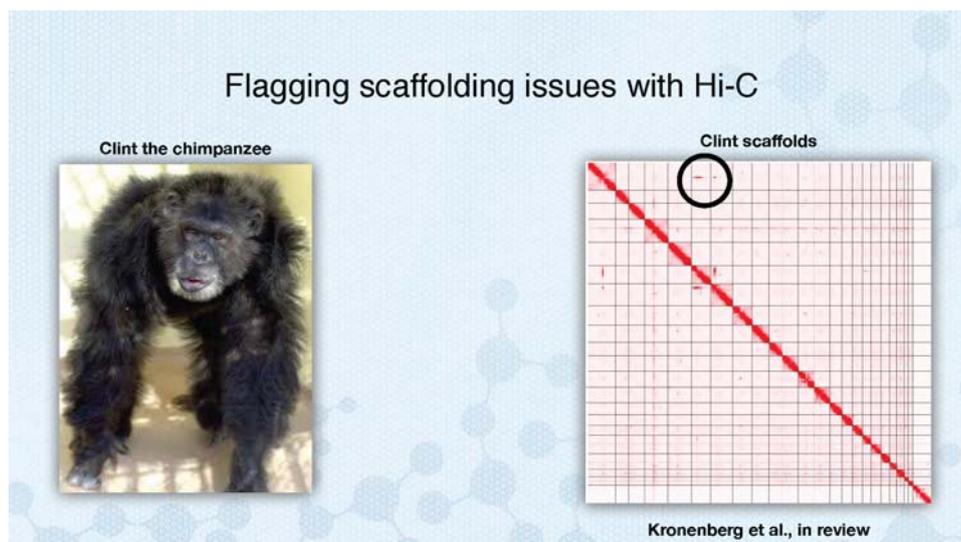
PLANT/ANIMAL

PB - *de novo* large genome: Axolotl



Nature : [The axolotl genome and the evolution of key tissue formation regulators](#)
S Nowoshilow *et al. Nature* **554**, 50–55 (2018)

Hi-C: long range SV; detect assembly error

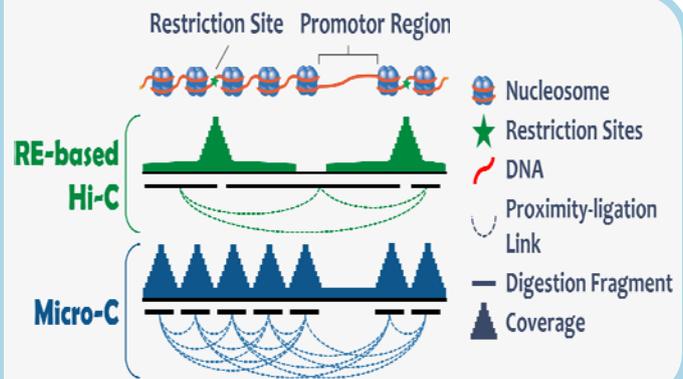


<https://youtu.be/uzlNKcj-p78>

Unlock 3D-Genome Architecture at Nucleosomes

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

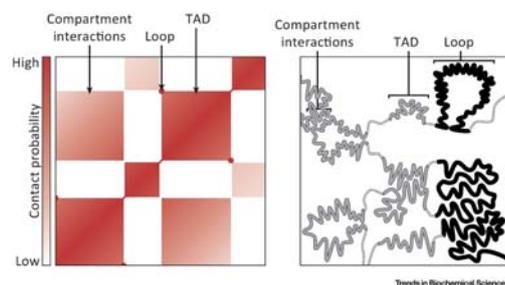
Micro-C High-resolution interacting domains



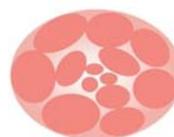
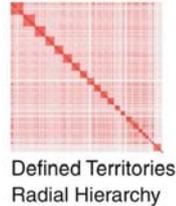
<https://doi.org/10.3791/1869>

Hierarchical Organization of Chromatin Structure

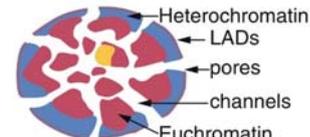
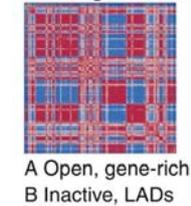
<https://doi.org/10.1038/nrg.2016.112>



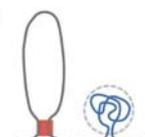
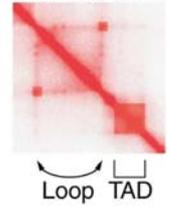
(a) Chromosomes
Hi-C of Genome



(b) Compartments
20Mb Eigenvectors



(c) Looping
5Kb Interactions

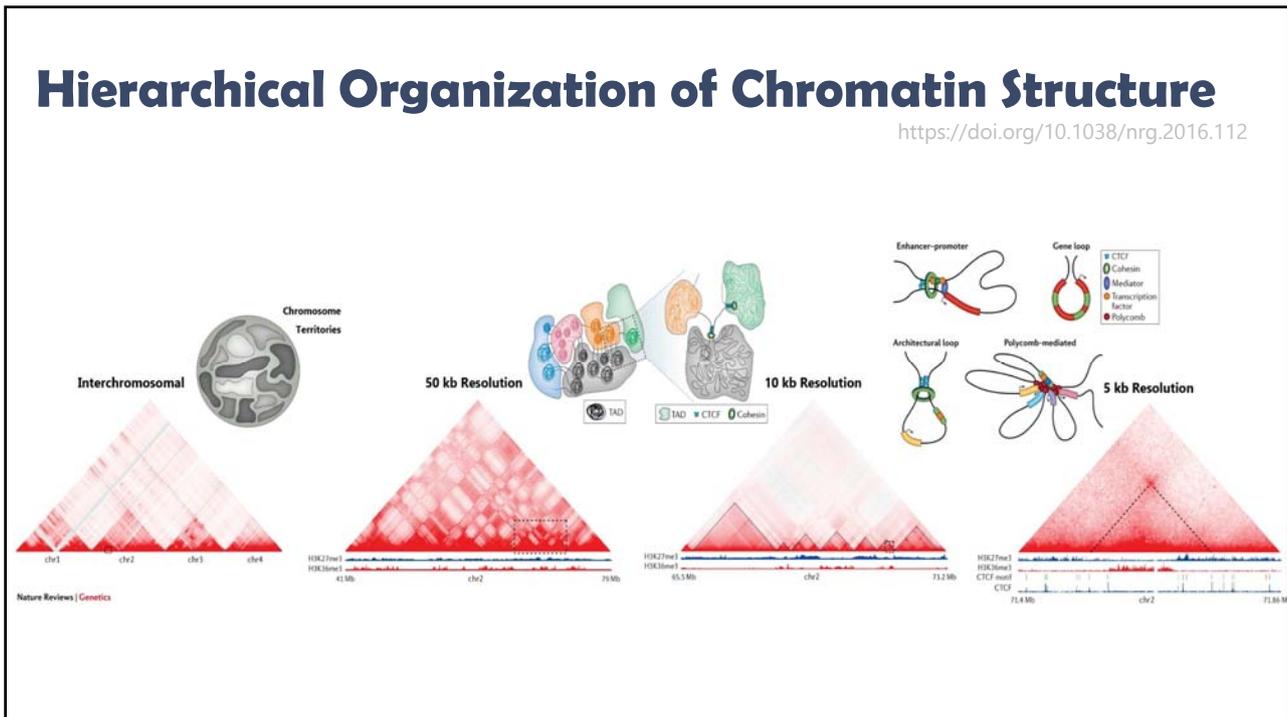


Sinica, A. & Eagen, K. (2018, April 20). Principles of Chromosome Architecture Revealed by Hi-C. Cell. <https://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004%2818%2930060-4>

Goodstadt, M. N., & Marti-Renom, M. A. (2019). Communicating Genome Architecture: Biovisualization of the Genome, from Data Analysis and Hypothesis Generation to Communication and Learning. *Journal of molecular biology*, 431(6), 1071–1087. <https://doi.org/10.1016/j.jmb.2018.11.008>

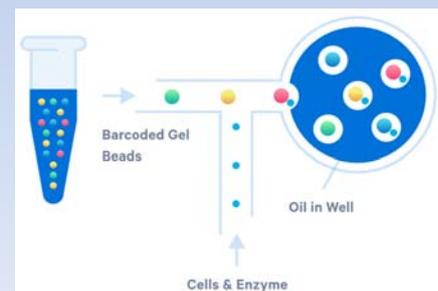
Hierarchical Organization of Chromatin Structure

<https://doi.org/10.1038/nrg.2016.112>



Single-Cell Sequencing

- DNA: sc-ATAC
- RNA: 3'-GEX, 5'-GEX
- Multiome
- VDJ repertoire



10x Single Cell

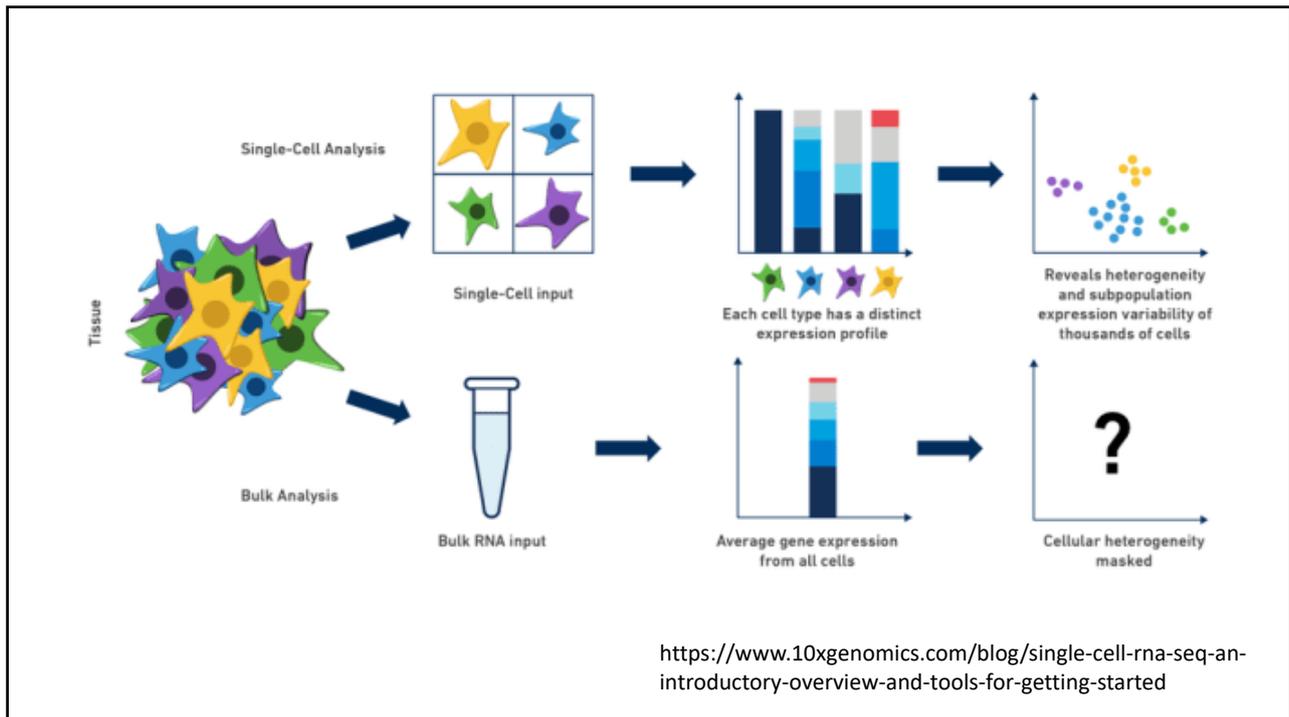
10x Chromium 3' Single Cell RNA Prep

Single Cell Chip G

10x Chromium VDJ 5' Single Cell RNA Prep

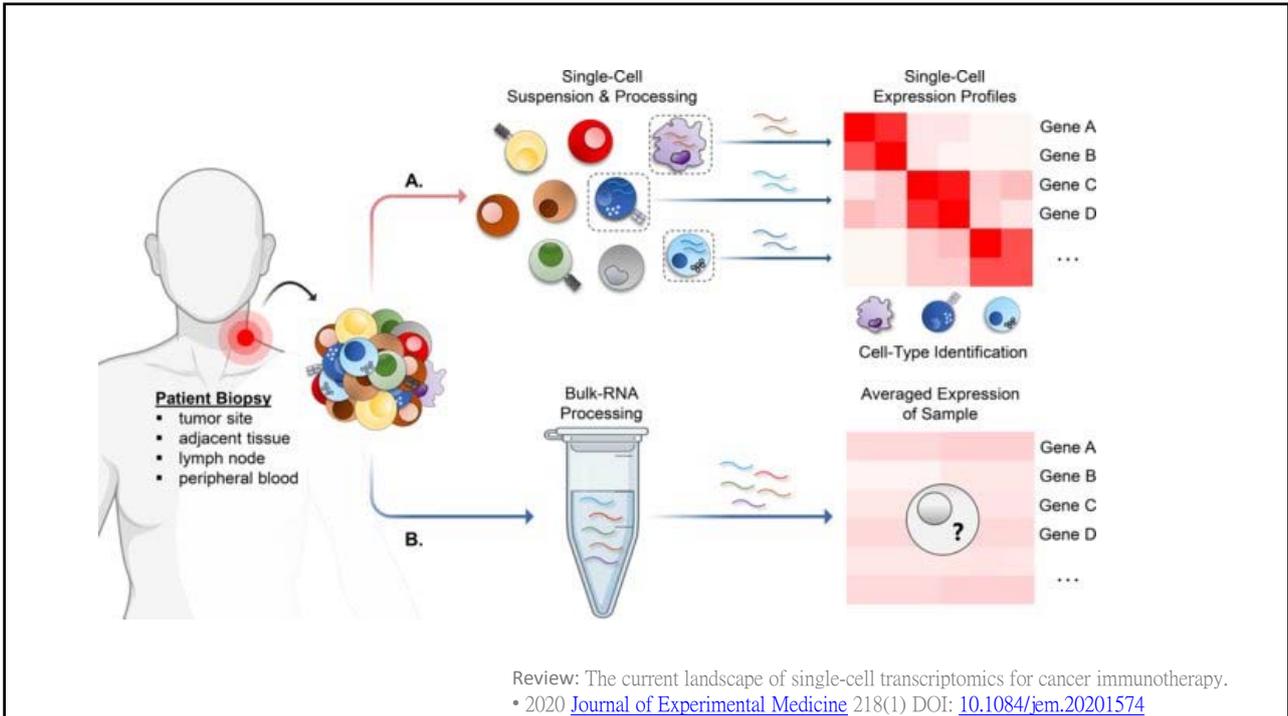
Single Cell Chip K

10x ATAC-seq prep (Chromatin accessibility)



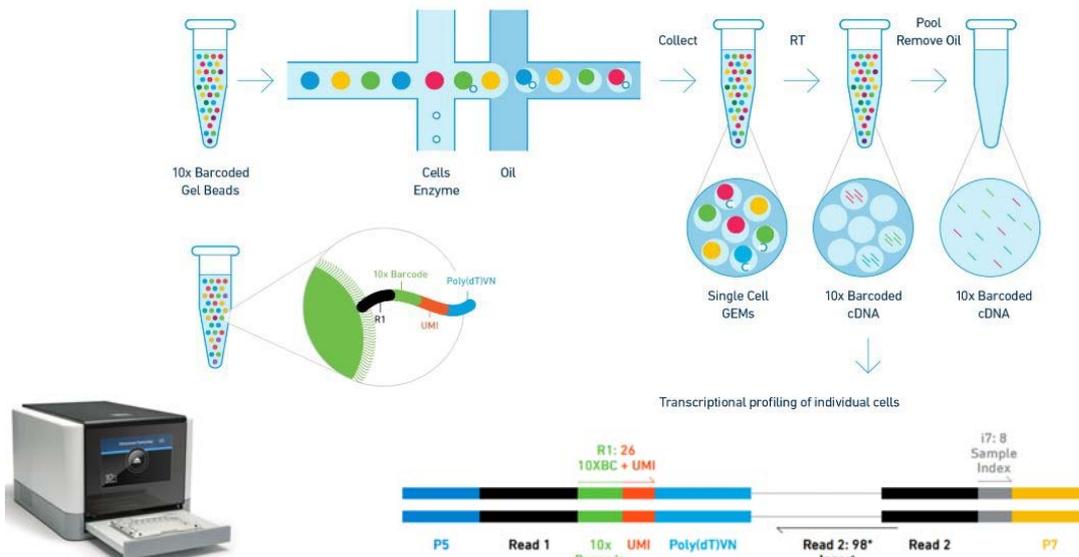
Why single-cell sequencing?

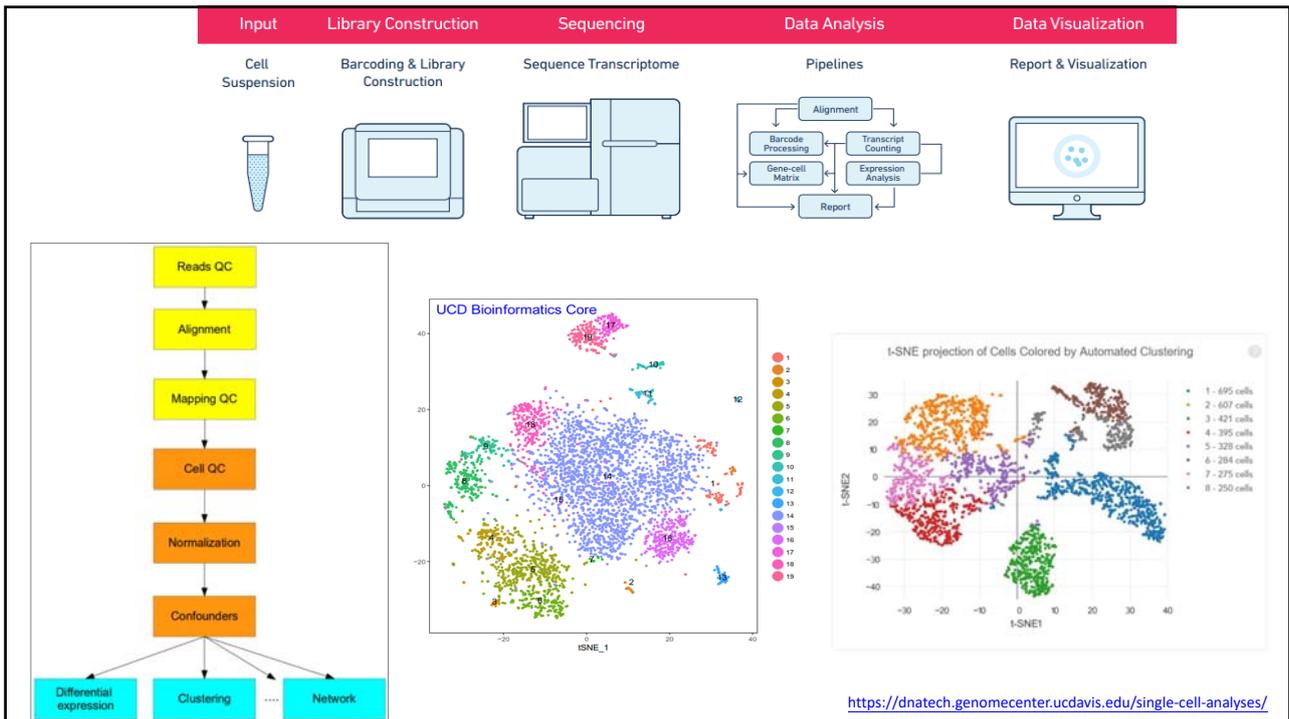
- To study the genetic makeup of individual cells, as opposed to studying an average of many cells as in bulk sequencing. This allows researchers to uncover rare cell populations, detect genomic heterogeneity within a cell population, and study the dynamics of cell differentiation and gene expression at the single-cell level.
- **Important applications:**
 1. **Studying rare cell populations:** Some cell populations in complex tissues or organisms are rare, making them difficult to study using traditional sequencing methods. Single-cell sequencing allows researchers to identify and study these rare cell types in detail.
 2. **Studying genomic heterogeneity:** In a population of cells, there can be genetic variation between individual cells that is masked when studying an average of many cells. Single-cell sequencing allows for the detection of genomic heterogeneity within a cell population, which can have important implications for understanding disease progression and treatment.
 3. **Understanding cell differentiation:** During development or disease progression, cells can differentiate into different cell types with distinct gene expression patterns. Single-cell sequencing can be used to study the gene expression changes that occur during cell differentiation, helping researchers to better understand how different cell types develop and function.
 4. **Discovering new cell types and functions:** Single-cell sequencing can reveal previously unknown cell types and functions that were not detectable using traditional sequencing methods.
 5. **Personalized medicine:** Single-cell sequencing can be used to study the genomic makeup of individual cells from a patient's tissue, providing information about the genetic basis of disease and guiding personalized treatment plans.
- Areas: the genetic and genomic complexity of biological systems, leading to new insights into development, disease, and personalized medicine.



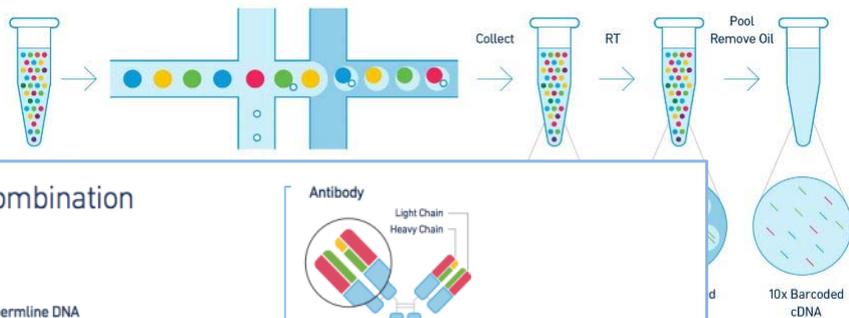
10x Genomics Chromium : barcoded gel beads for single cells

Single-cell embedding

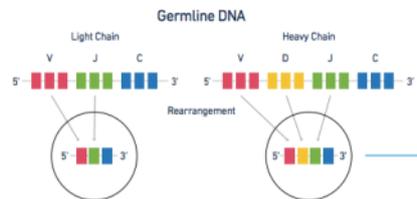




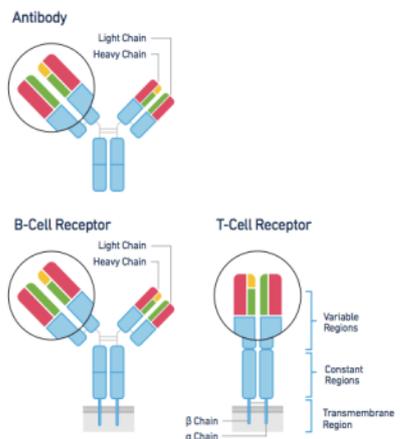
10x Genomics: Single-cell applications



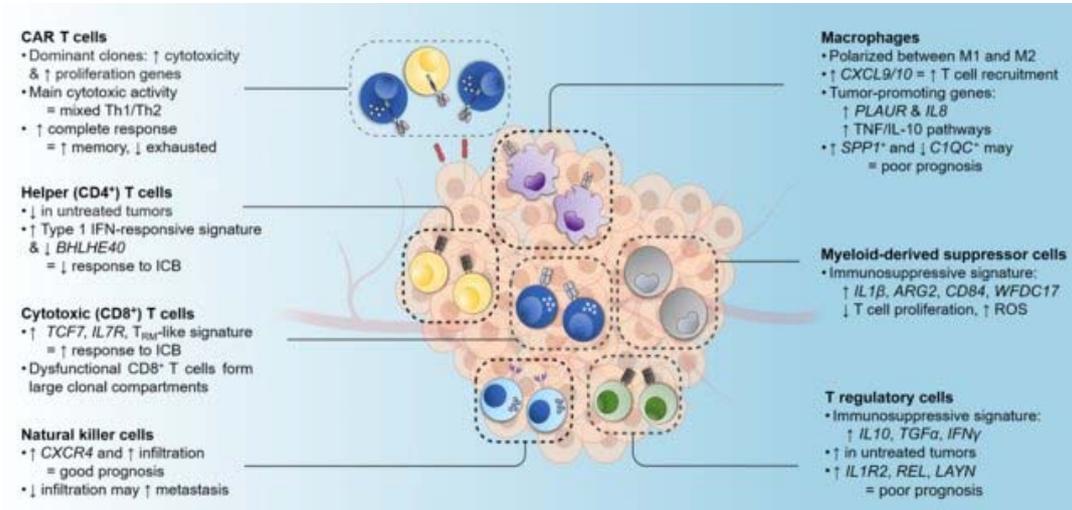
V(D)J Recombination



Single-cell VDJ

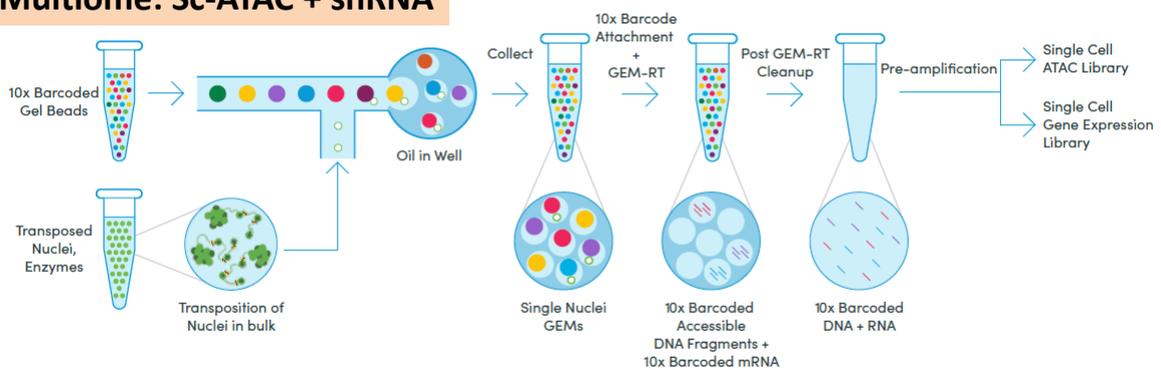


Cell type annotation: immuno marker genes



Review: The current landscape of single-cell transcriptomics for cancer immunotherapy.
 • 2020 *Journal of Experimental Medicine* 218(1) DOI: [10.1084/jem.20201574](https://doi.org/10.1084/jem.20201574)

10x Multiome: Sc-ATAC + snRNA



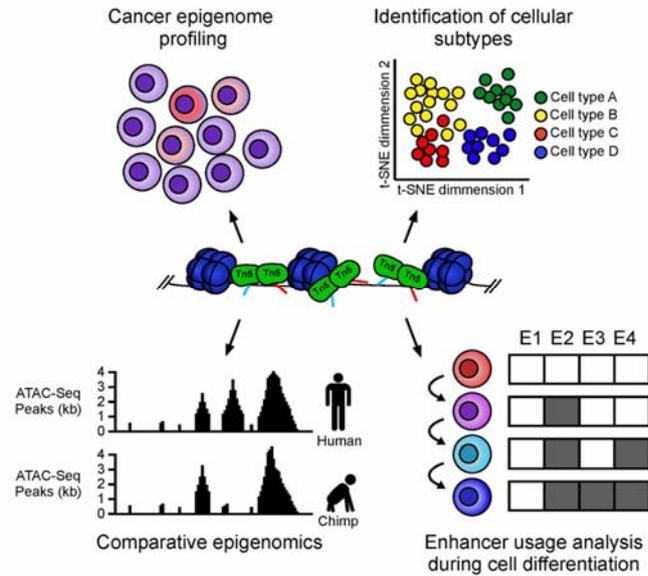
Solution Features

- Integrate gene expression and epigenomic landscape through direct measurement in the same cell, eliminating the need for inferring relationships in silico
- Identify linkages between putative regulatory elements and their target genes
- Simple and robust workflow
- Easy-to-use software for data analysis and visualization

System Features

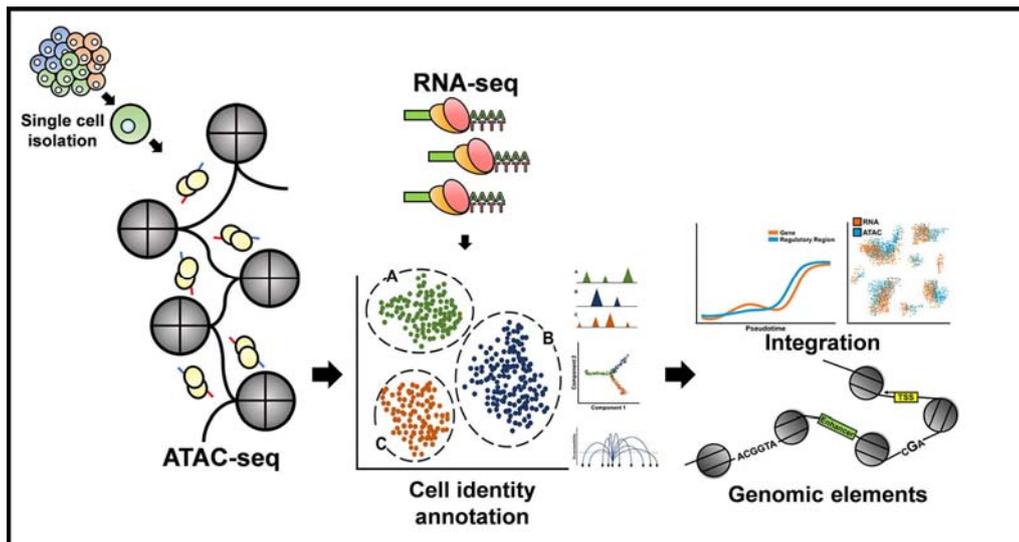
- Efficiently partition 500–10,000 nuclei per channel, for up to 80,000 nuclei per run
- Scalable; run up to 8 samples in parallel
- Recover up to 65% of loaded nuclei
- High sensitivity
- Low microfluidic multiplet rate (<1% per 1000 nuclei)
- Demonstrated with cell lines, primary cells, cryopreserved samples, and fresh and flash-frozen tissue

Single-cell ATAC-seq



<https://en.m.wikipedia.org/wiki/ATAC>

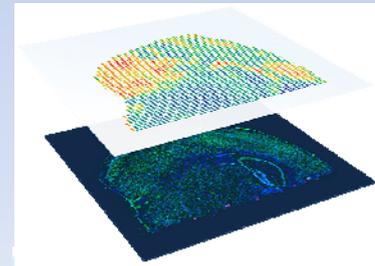
Integration of scRNA & ATAC-seq



<https://doi.org/10.1016/j.csbj.2020.06.012>

Spatial transcriptome

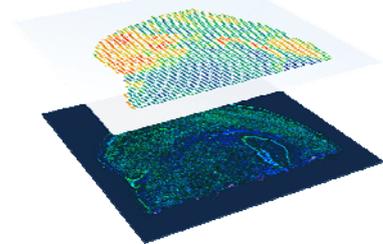
- Fresh frozen
- FFPE
- Symbiosis



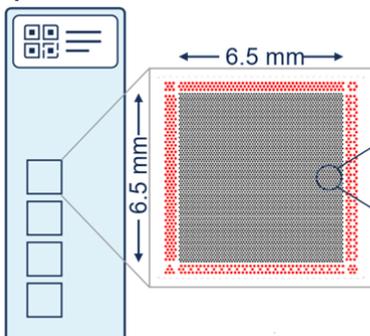
10x Visium

10x Visium Tissue Optimization (per slide)
10x Visium Gene Expression (per sample)
Fresh Frozen Tissue Preparation
Cryosectioning for Fresh Frozen Tissue
RNA Extraction
10x Visium Gene Expression for FFPE (per sample)
Sectioning for FFPE Tissue
FFPE RNA Extraction

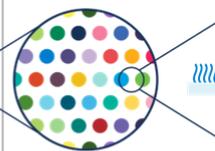
10x Genomics: Spatial Transcriptome



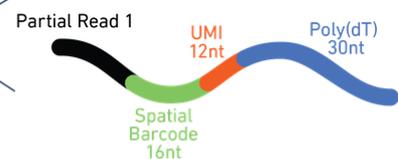
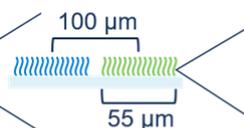
Visium Spatial Gene Expression Slide



Capture Area with ~5000 Barcoded Spots

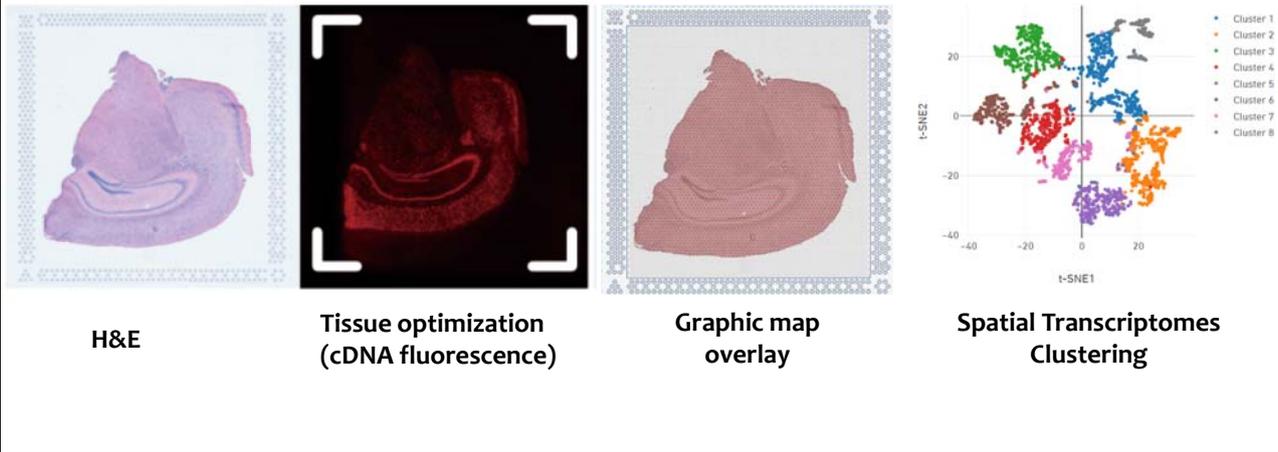


Visium Gene Expression Barcoded Spots



SpaceRanger Analysis

Mouse Brain



SpaceRanger Analysis: mouse brain

2,597
Number of Spots Under Tissue

54,818 **3,550**
Mean Reads per Spot Median Genes per Spot

Sequencing

Number of Reads	142,361,477
Valid Barcodes	97.0%
Valid UMIs	99.9%
Sequencing Saturation	69.8%
Q30 Bases in Barcode	96.0%
Q30 Bases in RNA Read	94.7%
Q30 Bases in UMI	96.4%

Mapping

Reads Mapped to Genome	95.7%
Reads Mapped Confidently to Genome	93.7%
Reads Mapped Confidently to Intergenic Regions	-4.3%
Reads Mapped Confidently to Intronic Regions	1.4%
Reads Mapped Confidently to Exonic Regions	88.0%
Reads Mapped Confidently to Transcriptome	85.9%
Reads Mapped Antisense to Gene	1.2%

Spots

Fraction Reads in Spots Under Tissue	88.6%
Mean Reads per Spot	54,818
Mean Reads Under Tissue per Spot	46,781
Median Genes per Spot	3,550
Total Genes Detected	20,321
Median UMI Counts per Spot	10,140

Sample

Sample ID: XGE21-WH01

UMIs Detected

Tissue Plot with Spots Colored by UMI Count

t-SNE Projection of Spots Colored by UMI Counts

Clustering

Clustering Type: Graph-based

Tissue Plot with Spots Colored by Clustering

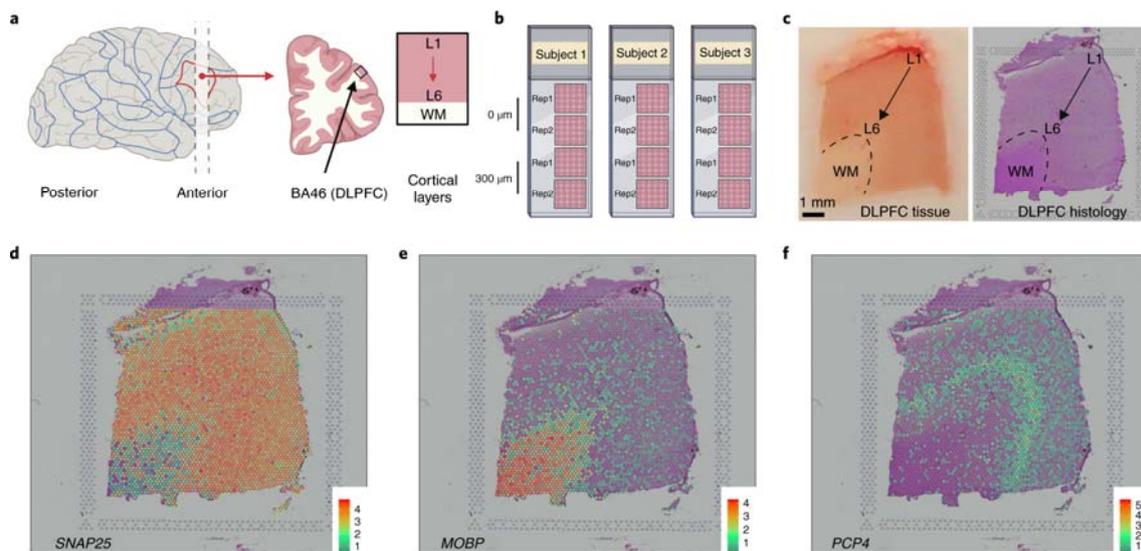
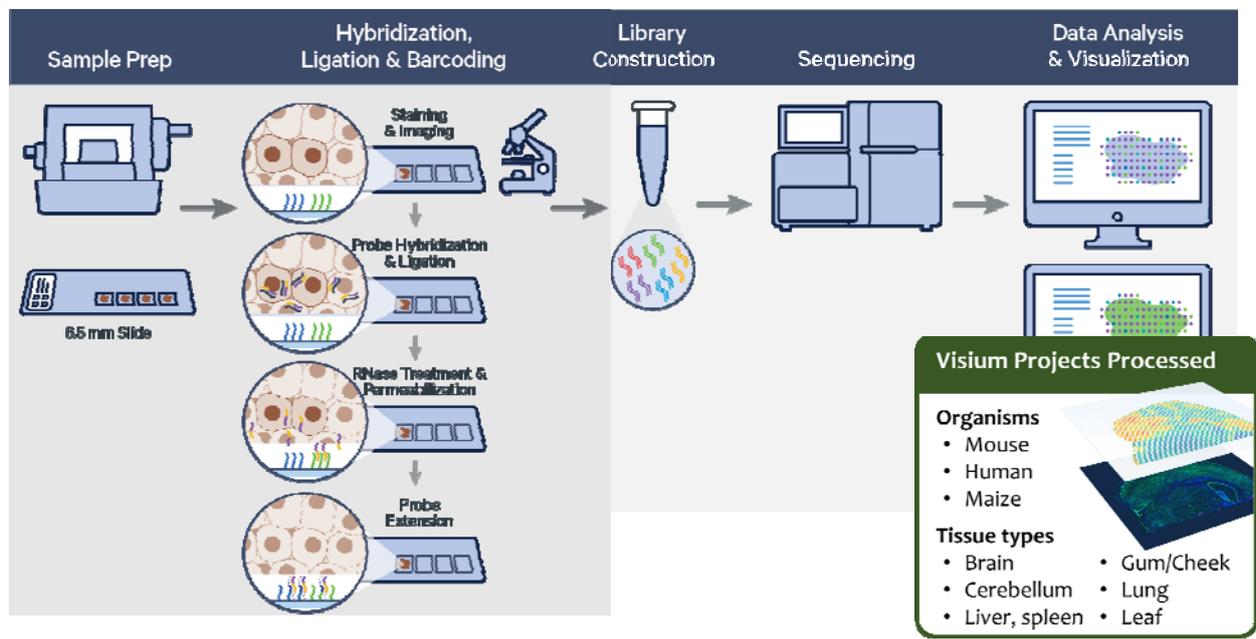
t-SNE Projection of Spots Colored by Clustering

Top Features by Cluster (Log2 fold-change, p-value)

Feature ID	Name	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
		L2FC	p-value										
ENSMUSG0000000618	Tr	4.17	1e-14									2.40	3e-2
ENSMUSG0000000260	Egr3	3.92	2e-16									2.55	1e-3

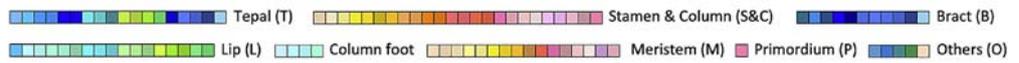
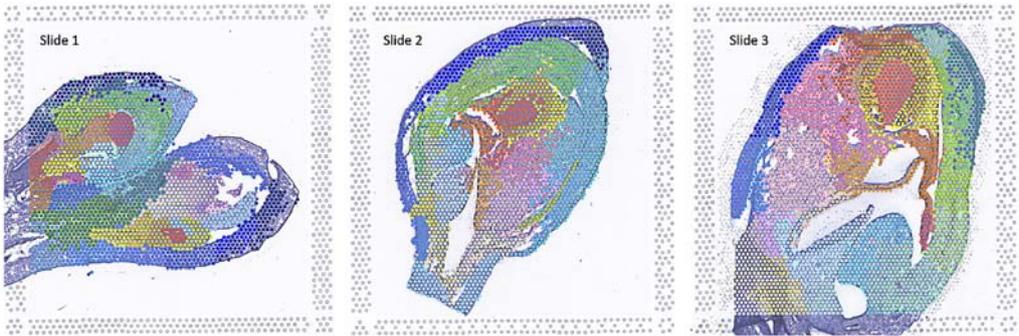
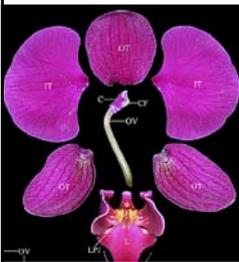
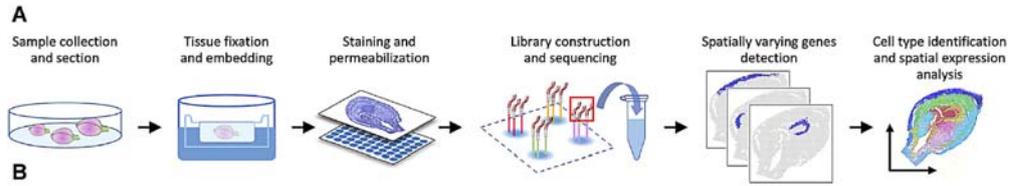


Exploring Spatial Transcriptomics with 10x Genomics Visium



Maynard, K.R., Collado-Torres, L., Weber, L.M. et al.
 Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* 24, 425–436 (2021).
<https://doi.org/10.1038/s41593-020-00787-0>

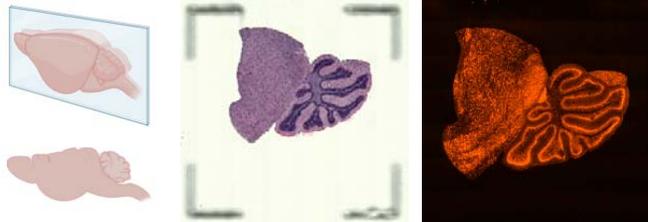
Orchid organogenesis



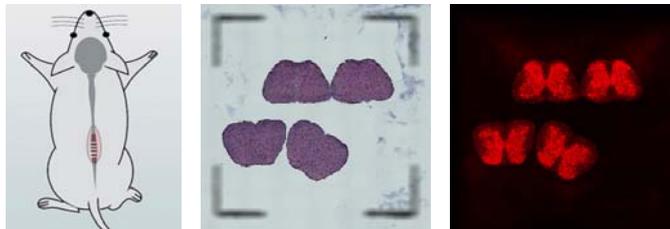
Liu C, et al. A spatiotemporal atlas of organogenesis in the development of orchid flowers. *Nuc Acids Res* gkac773 (2022). doi: 10.1093/nar/gkac773.

Fresh Frozen Cryosectioning

Mouse Cerebellum

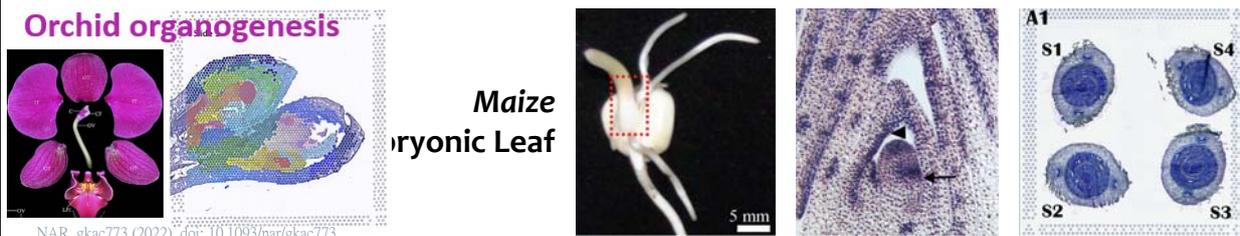


Mouse Spinal Cord

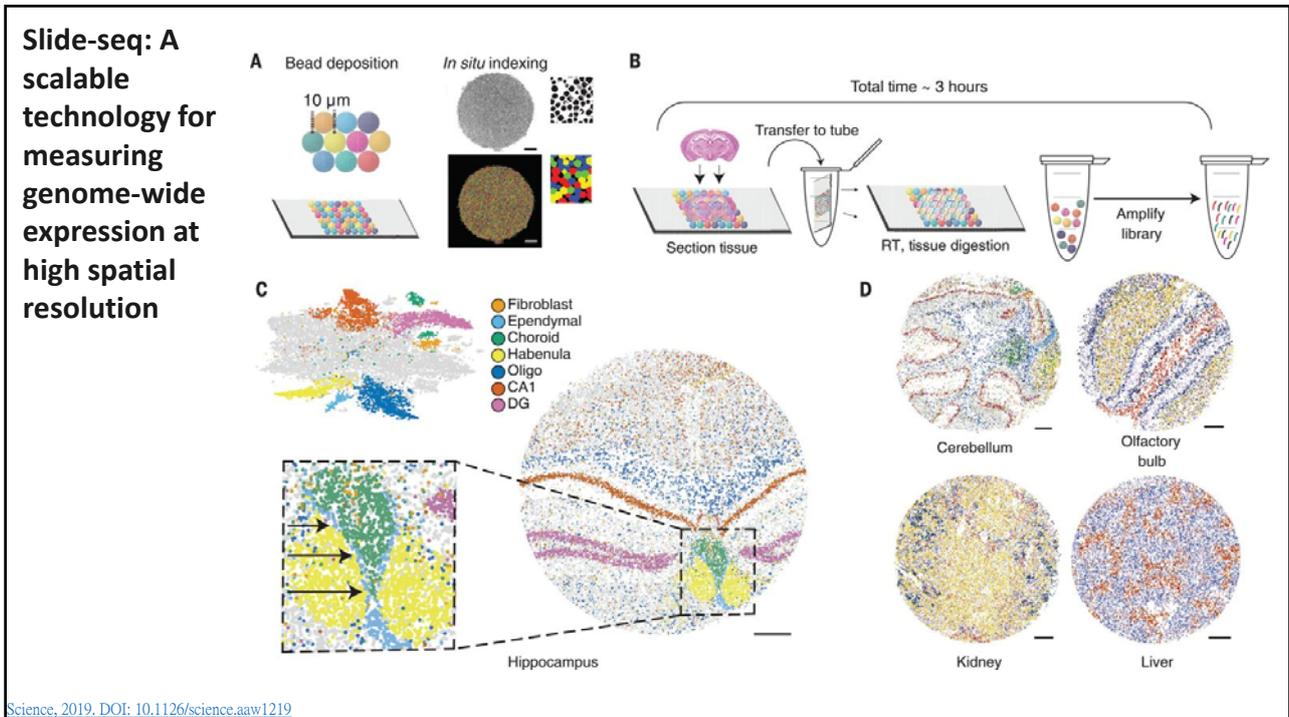
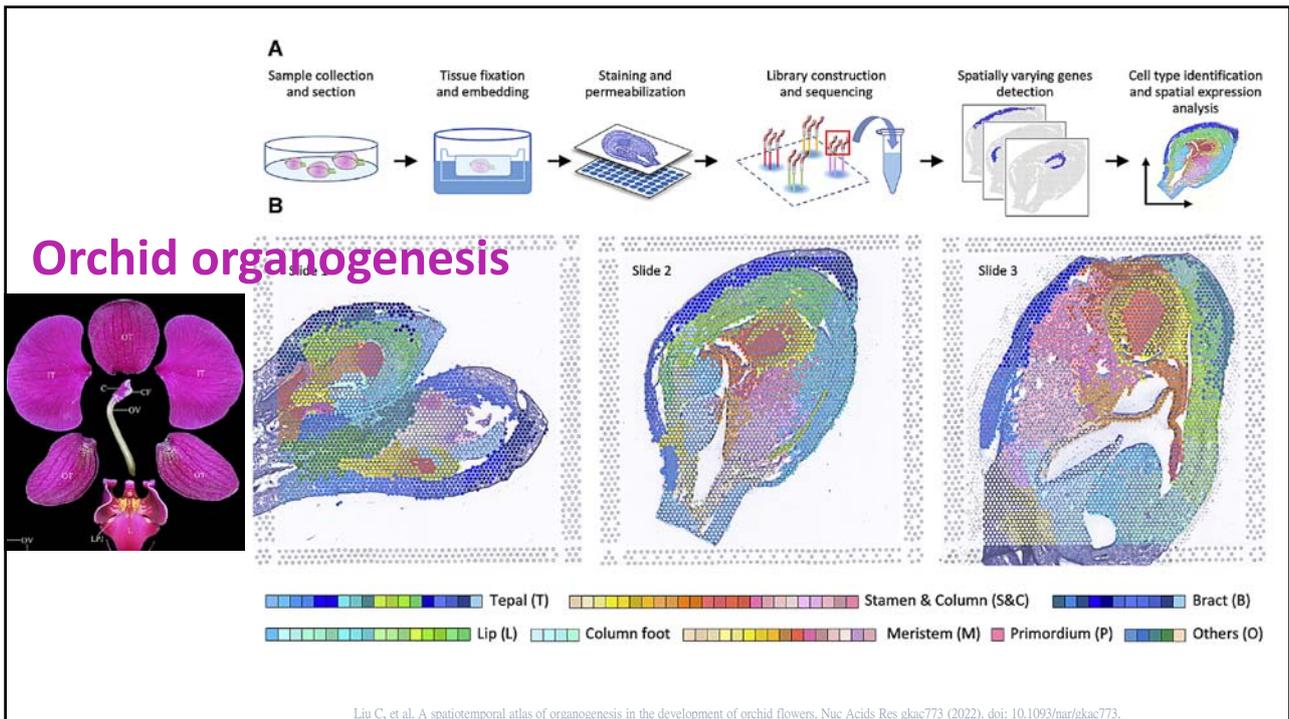


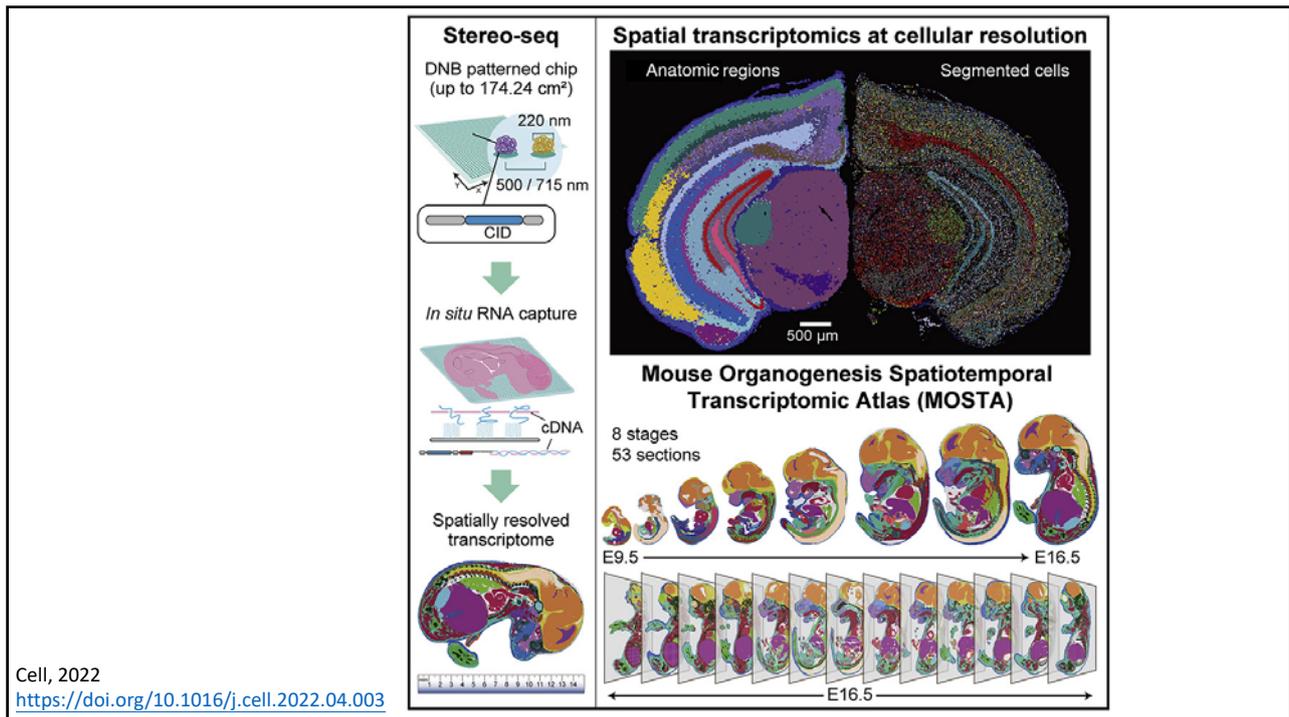
Orchid organogenesis

Maize Embryonic Leaf



NAR, gkac773 (2022). doi: 10.1093/nar/gkac773





The Visium projects we have processed

Organisms

- Mouse
- Human
- Maize

Samples

- Flash frozen (OCT)
- FFPE

Tissue types

- Brain
- Cerebellum
- Liver
- Gum/cheek
- Lung
- Leaf

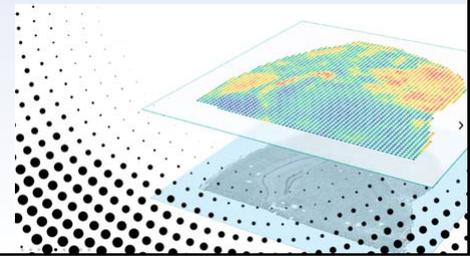
Biological Qs

- Cancer/tumor
- Stress/injury
- Neurology
- Embryonic development
- Infectious disease

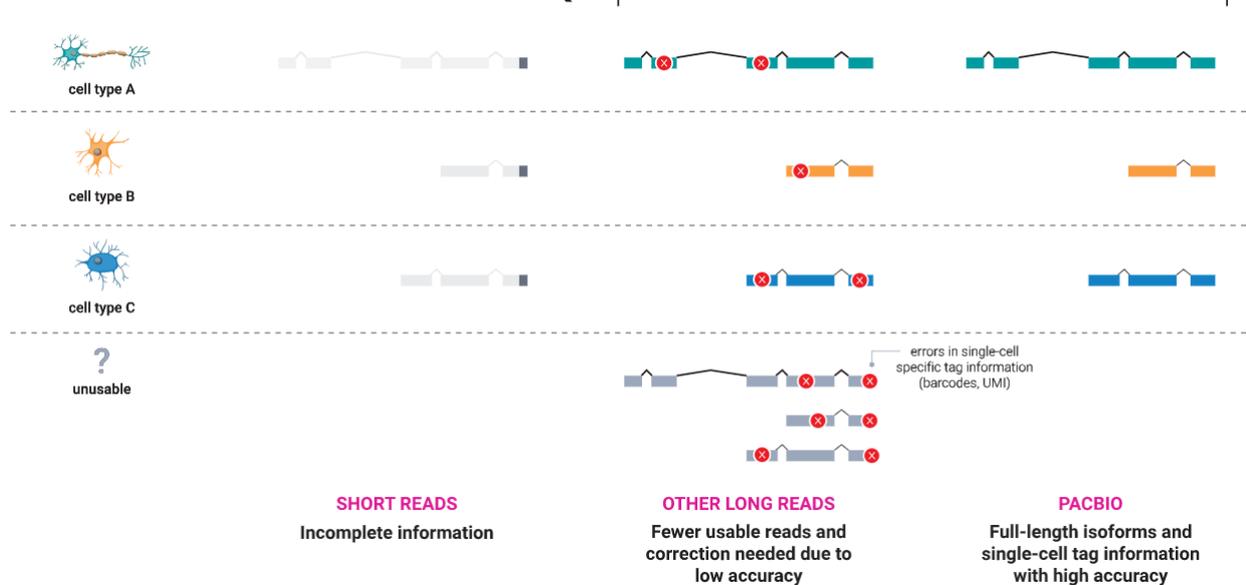
Multiplexed Array Sequencing:

MAS-sclsoseq

- 10X scRNA
- 10X Spatial-Visium



HIFI SEQUENCING ADVANTAGE IN SINGLE-CELL RNA SEQUENCING



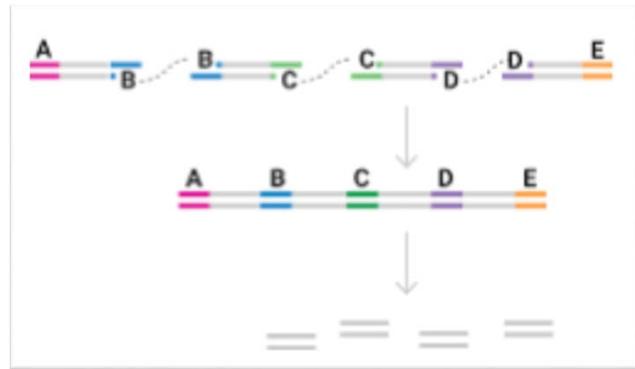
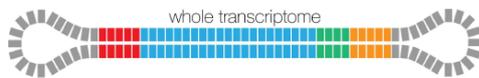
<https://www.pacb.com/products-and-services/applications/rna-sequencing/single-cell-rna-sequencing/>

10x + PB: MAS-Sp-Isoform Sequencing

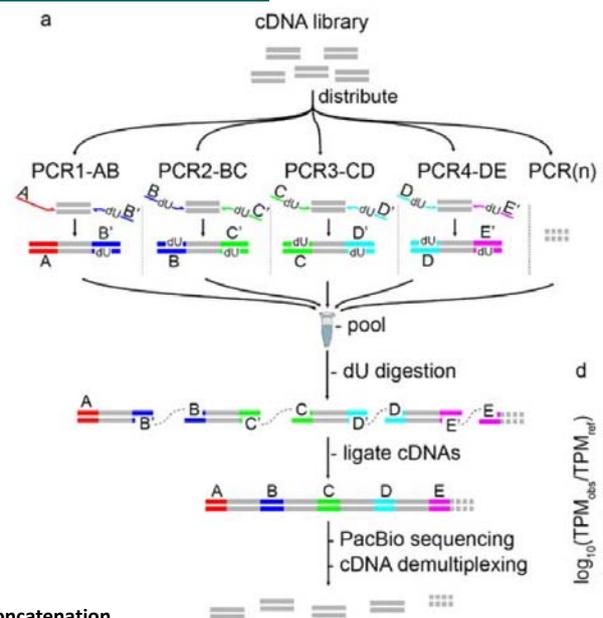
High-throughput RNA isoform sequencing using programmable cDNA concatenation

Aziz M. Al'Khafaji^{1,*,†}, Jonathan T. Smith^{1,*}, Kiran V. Garimella^{1,*,†}, Mehrtash Babadi^{1,*,†}, Moshe Sade-Feldman^{1,2}, Michael Gatzert¹, Siranush Sarkizova¹, Marc A. Schwartz^{1,3,4,5}, Victoria Popic¹, Emily M. Blaum^{1,2}, Allyson Day¹, Maura Costello¹, Tera Bowers¹, Stacey Gabriel¹, Eric Banks¹, Anthony A. Philippakis¹, Genevieve M. Boland⁶, Paul C. Blainey^{1,7,8,†}, Nir Hacohen^{1,9,10,11,†}

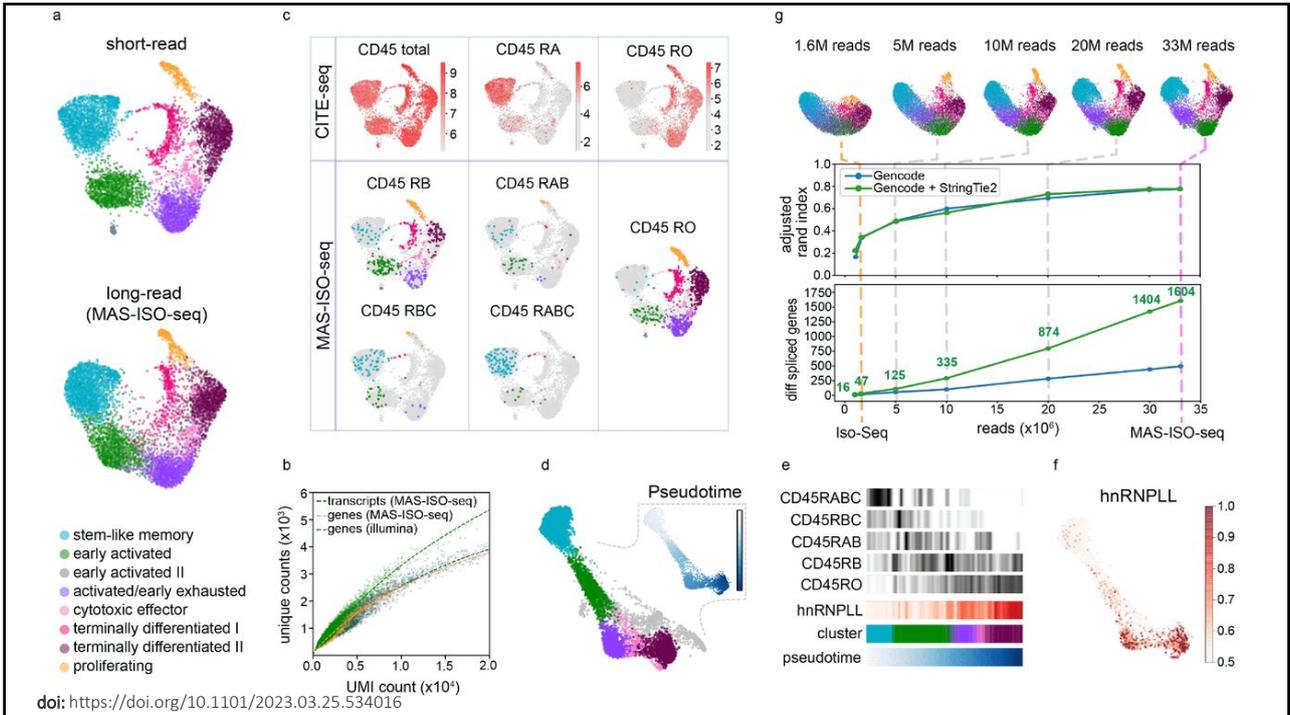
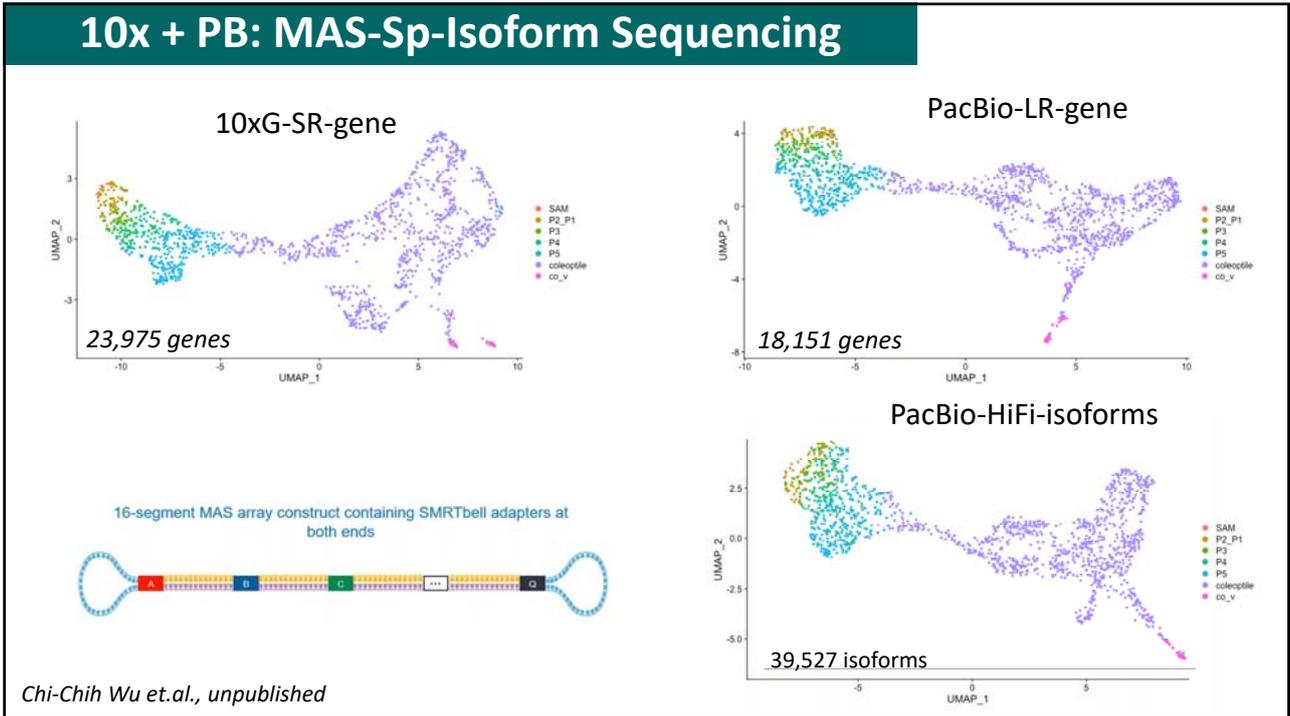
1. Broad Institute of Harvard and MIT, Cambridge, MA, USA
2. Department of Medicine, Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA
3. Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA.



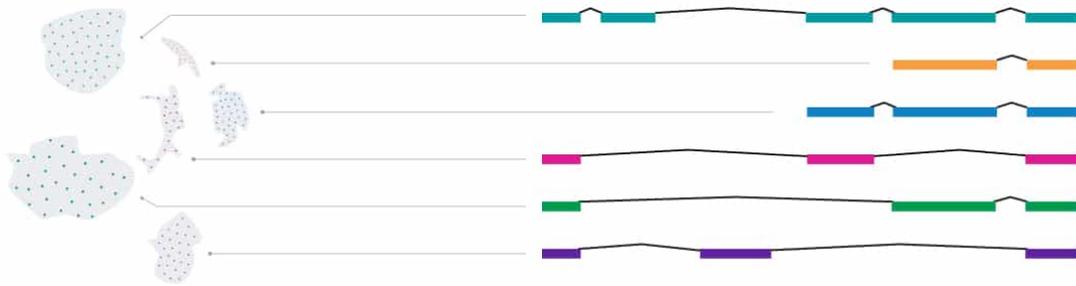
10x + PB: MAS-Sp-Isoform Sequencing



High-throughput RNA isoform sequencing using programmable cDNA concatenation
2121. doi: <https://doi.org/10.1101/2021.10.01.462818>



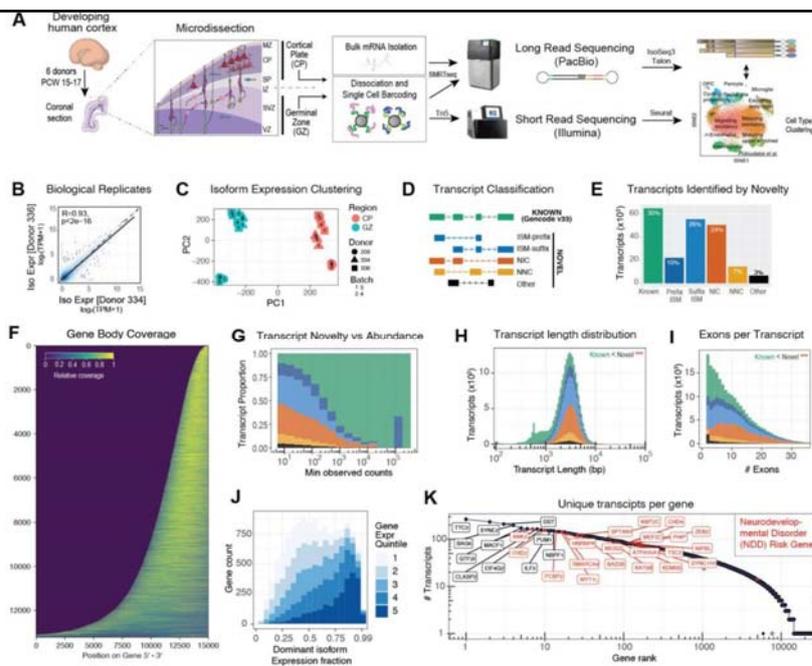
“...Detection of 214,516 unique isoforms covering 22,391 genes,
72.6% of the isoforms are novel.”



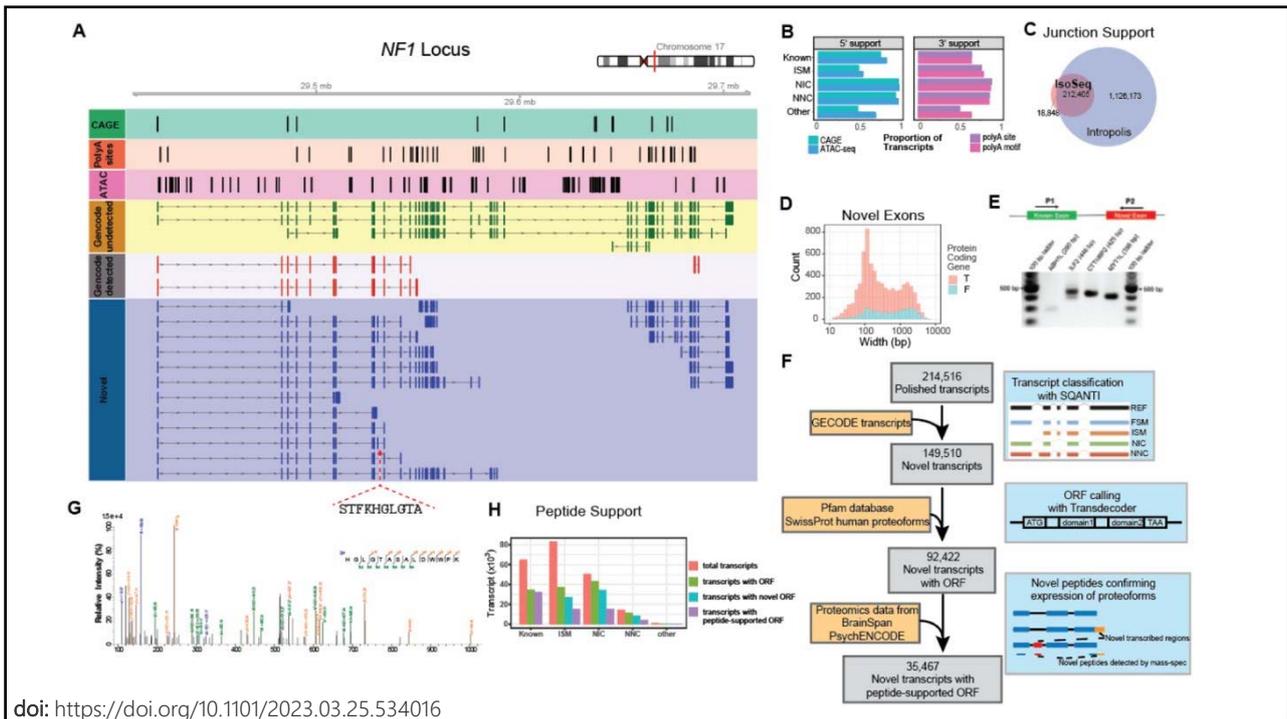
[Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders](https://doi.org/10.1101/2023.03.25.534016)

doi: <https://doi.org/10.1101/2023.03.25.534016>

<https://www.pacb.com/blog/the-hifi-difference-a-better-cell-atlas-with-full-length-isoform-sequencing/>



doi: <https://doi.org/10.1101/2023.03.25.534016>



Seminars /Workshops for advanced NGS Technologies



10x GENOMICS
Technology And Applications
技術新知與產品應用研討會

2018.1.15 (Mon.) at 2-4 pm.
詳情請見研討會大綱
—歡迎踴躍參加—
主持人 呂美蘭 博士

Speaker: **Leo Chen**
Field Applications Scientist, 10x GENOMICS

PacBio Workshop

MEIYEH LU
DE NOVO GENOME & MULTIPLEXED BACTERIAL GENOME

JOAN WONG
METAGENOME ANALYSIS: AMPLICON VS SHOTGUN WMG

BERYL MA
FULL LENGTH ISO-SEQ TRANSCRIPTOME SEQUENCING

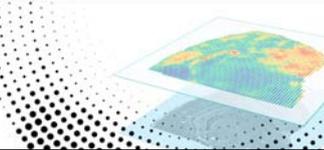
ZUWEI QIAN
APPLICATIONS OF SMRT SEQUENCING

TIME: 2019.5.7 (TUE) 1:30 - 5:10 PM
PLACE: 1F AUDITORIUM, INTERDISCIPLINARY BUILDING

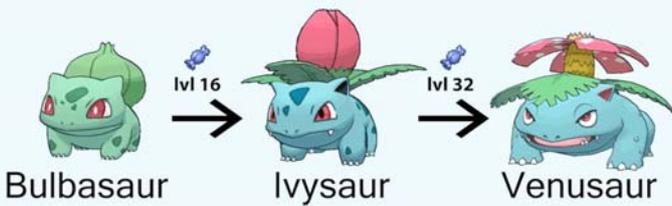
Our New GridION Is Up And Running



Visium Spatial Transcriptome
A Novel Technology Is Now Available At NGS Core



How to evolve Pokemon



Sequencing Evolution

Technology teaming



illumina

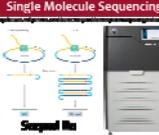
Sequencing By Synthesis



	Illumina HiSeq 2500	Illumina MiSeq	Illumina NextSeq 2000 (new)
Chemistry	Cyclic reversible termination of amplified DNA clusters		
Output per Run	Hi 1.2 Tb Output 1.1T Gb	Up to 15 Gb	1Tc - 1.4T Tb 1.1T - 1.6T Gb
Max Read Length	250bp	250bp	250bp
Fragments per Chip	HT: 2000-2500 M V4: 200-400 M	V2: 12-15 M V3: 20-25 M	R2: 400-500 M V2: 1.2-1.5 TB
Data Quality	~99.9% Tolerates Homopolymers Scalable to High GC		
Application	De novo assembly Re-sequencing RNA-Seq	De novo assembly Re-sequencing Amplification	De novo assembly Re-sequencing Single cell sequencing

PACIFIC BIOSCIENCES

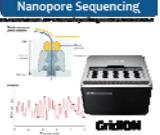
Single Molecule Sequencing



	PacBio Sequel II (Nov. 2022)
Chemistry	SMRT-Seq DNA polymerization
Output per Run	Current: 2.90 Gb
Read Length	1-30 kb (max >100kb)
Fragments per Chip	Up to 1M (100 Reads/10MRT-1hr)
Data Quality	Run 85-90% HRF - 99.9% Random Homopolymers errors Tolerates High GC
Application	Genome assembly/Amplification Structural variation/Phasing Methylation/Genotyping

NANOPORE

Nanopore Sequencing

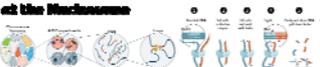


	Oxford ONT GridION
Chemistry	Electrical current passing through a nanopore channel
Output per Run	Current: 2.90 Gb
Read Length	1-50 kb (max >500kb)
Fragments per Chip	10-100 k Reads/Flowcell
Data Quality	Run 82-95% Epitaxial Homopolymers errors Tolerates High GC
Application	Genome assembly/Amplification Structural variation/Phasing Methylation/RNA/TE/SA-Seq

NGS projects of HTG Core in Academia Sinica Internal & Collaborative Projects

- Pathogenic Bacteria
- Insect Genomes
- Human Diseases
- Pathogenic/Medical Fungi
- Marine Animals
- Microbiomes
- Fish/Invertebrate Species
- Cyt4 Plants
- Vitines

Unleash 3D Genome Architecture at the Nucleosome



Hi-C

Chromatin conformation for genome assembly, genome phasing, and/or genome 3D landscape

Micro-C

High-resolution interacting domains

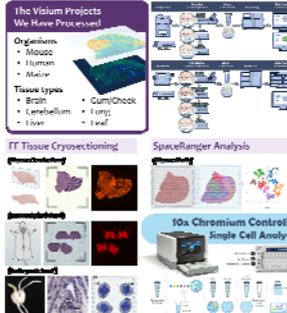
Hi-C

Chromatin conformation for genome assembly, genome phasing, and/or genome 3D landscape

Micro-C

High-resolution interacting domains

Exploring the Extent of Spatial Transcription with the Genomics Vision



The Vision Projects We Have Processed

Organism: Mouse, Human, Medea

Tissue types: Brain, Cerebellum, Liver

IT Tissue Cryosectioning

Immunofluorescence, Immunohistochemistry

Review



Ten years of next-generation sequencing technology

Erwin L. van Dijk¹, Hélène Auger¹, Yan Jaszczyszyn², and Claude Thermes¹

¹ Centre de Génétique Moléculaire – CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France
² Plateforme Intégrée IMAGIF – CNRS, Avenue de la Terrasse, 91198 Gif sur Yvette, France

Trends Genet. 2014 Sep;30(9):418-26. doi: 10.1016/j.tig.2014.07.001.

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin¹, John D. McPherson² and W. Richard McCombie¹

Nature Reviews Genetics volume 17, pages333–351 (2016)

Review of Clinical Next-Generation Sequencing

Sophia Yohe, MD; Bharat Thyagarajan, MD, PhD

Context.—Next-generation sequencing (NGS) is a technology being used by many laboratories to test for inherited disorders and tumor mutations. This technology is new for many practicing pathologists, who may not be familiar with the uses, methodology, and limitations of NGS.

Objective.—To familiarize pathologists with several aspects of NGS, including current and expanding uses; methodology including wet bench aspects, bioinformatics, and interpretation; validation and proficiency; limitations; and issues related to the integration of NGS data into patient care.

Data Sources.—The review is based on peer-reviewed literature and personal experience using NGS in a clinical setting at a major academic center.

Conclusions.—The clinical applications of NGS will increase as the technology, bioinformatics, and resources evolve to address the limitations and improve quality of results. The challenge for clinical laboratories is to ensure testing is clinically relevant, cost-effective, and can be integrated into clinical care.

(Arch Pathol Lab Med. 2017;141:1544–1557; doi: 10.5858/arpa.2016-0501-RA)

Arch Pathol Lab Med. 2017;141:1544–1557; doi: 10.5858/arpa.2016-0501-RA

NGS Reviews

Next-generation DNA sequencing

Jay Shendure¹ & Hanlee Ji²

Nature Biotechnology 26, 1135 - 1141 (2008)



ADVANCES IN WOUND CARE

Many Open Access free publications

Journals

Search

Alerts

Adv Wound Care (New Rochelle) 2015 Jan 1; 4(1): 50-58.
doi: 10.1089/wound.2014.0542

PMCID: PMC4281878

Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research

Brendan P. Hodkinson and Elizabeth A. Grice

Author information Article notes Copyright and License information

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Sequencing technologies — the next generation

Michael L. Metzker^{1*}

Nature Review Genetics 11, 31-46 (2010)

nature REVIEWS GENETICS

nature.com • journal home • archive • issue • review • full text

NATURE REVIEWS GENETICS | REVIEW

ARTICLE SERIES: Applications of next-generation sequencing

Coming of age: ten years of next-generation sequencing technologies

Sara Goodwin, John D. McPherson & W. Richard McCombie

Affiliations | Corresponding author

Nature Reviews Genetics 17, 333–351 (2016) | doi:10.1038/nrg.2016.49
Published online 17 May 2016

NIH Public Access

Author Manuscript

J Genet Genomics. Author manuscript; available in PMC 2011 April 13.

Published in final edited form as:

J Genet Genomics. 2011 March 20; 38(3): 95–109. doi:10.1016/j.jgg.2011.02.003.

The impact of next-generation sequencing on genomics

Jun Zhang^{a,b,*}, Rod Chiodini^c, Ahmed Badr^a, and Genfa Zhang^d

^a COE for Neurosciences, Department of Anesthesiology, Texas Tech University Center El Paso, TX 79905, USA

135

Video Clips

- [Sanger Sequencing of DNA](#) [HD Animation]
 - <https://www.youtube.com/watch?v=nudG0r9zL2M>
- [Pyro Sequencing](#)
 - <https://www.youtube.com/watch?v=nFfgWGFe0aA>
- [Illumina Sequencing Technology](#)
 - <https://www.youtube.com/watch?v=womKfikWlxM>
- [Ion Torrent™ next-gen sequencing technology](#)
 - <https://www.youtube.com/watch?v=WYBzbxIfuKs>
- [Single Molecule Real Time Sequencing - Pacific Biosciences](#)
 - <https://www.youtube.com/watch?v=v8p4ph2MAvI>
- [Oxford Nanopore Technologies](#)
 - <https://www.youtube.com/watch?v=3UHw22hBpAk>
- [Next-Generation Sequencing Technologies - Elaine Mardis \(2014\)](#)
 - <https://www.youtube.com/watch?v=6Is3W7JkFp8>
- [PCR \(Polymerase Chain Reaction\)](#)
 - <https://www.youtube.com/watch?v=iQsu3Kz9NYo>
- [Polymerase Chain Reaction](#) [HD Animation]
 - <https://www.youtube.com/watch?v=0HCWmd7Mv8U>