Lec 6. Molecular Diagnostics

Zida Li
Associate professor

深圳大学
SHENZHEN UNIVERSITY

Why?

分子诊断&身体状态

分子诊断&国家政策

十三五
国家科技创新
规划

经李克强总理签批，国务院近日印发《“十三五”国家科技创新规划》（国发〔2016〕43号文，以下简称《规划》），明确提出了未来五年国家科技创新的指导思想、总体要求、战略任务和改革举措。

9、体外诊断产品

突破微流控芯片、单分子检测、自动化核酸检测等关键技术，开发全自动核酸检测系统、高通量液相悬浮芯片、医用生物质谱仪、快速病理诊断系统等重大产品，研发一批重大疾病早期诊断和精确治疗诊断试剂以及适合基层医疗机构的高精度诊断产品，提升我国体外诊断产业竞争力。

利用个体基因信息能有效找到患者病因

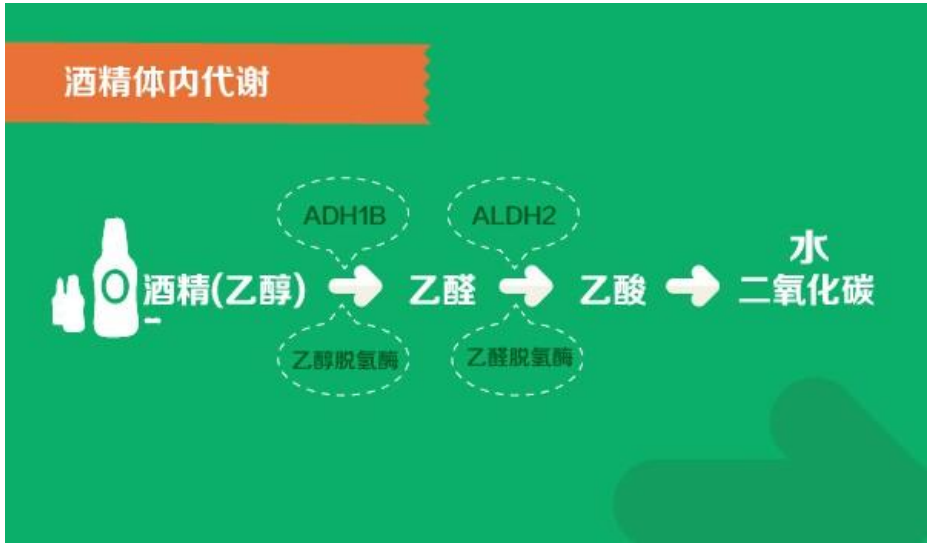
美国不必要治疗和无效治疗耗费：7500 亿美元，30%

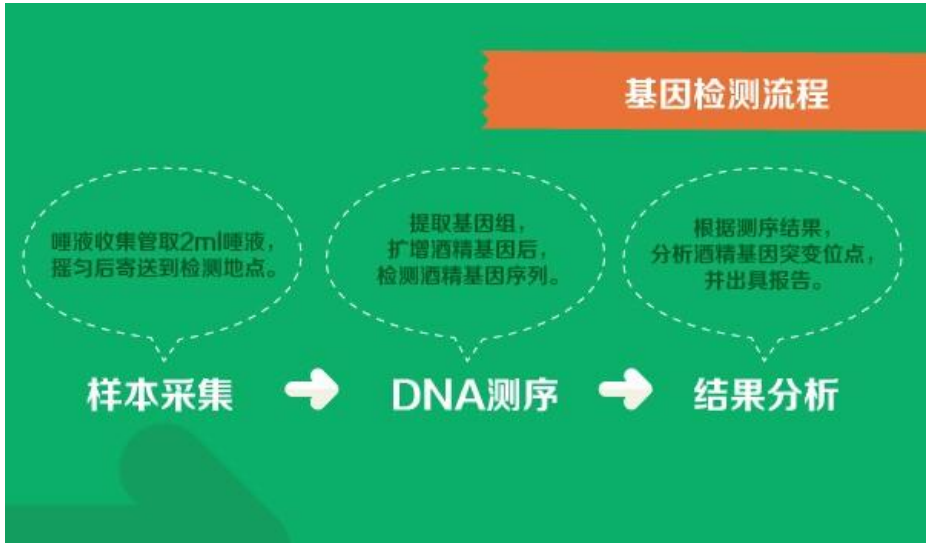
- 精准诊断的核心是分子诊断

酒精基因检测


99元 +一口唾液
便知你是酒神？还是半杯倒？
酒精基因检测
检测基因：
乙醇脱氢酶基因
乙醛脱氢酶基因


你们看看我的“酒神证”，
两个酒精基因活力超强，
喝酒就像喝水一样！
不能喝，真不能喝酒，
基因检测说我乙醛脱氢酶失活，
过量乙醛会致癌！


酒精体内代谢
ADH1B
ALDH2
酒精(乙醇)
乙醛
乙酸
水
二氧化碳
乙醇脱氢酶
乙醛脱氢酶


基因检测流程
唾液收集管取2ml唾液，
摇匀后寄送到检测地点。
提取基因组，
扩增酒精基因后，
检测酒精基因序列。
根据测序结果，
分析酒精基因突变位点，
并出具报告。
样本采集
DNA测序
结果分析

肿瘤&药物

药物类型	检测基因		突变位点
EGFR TKI	耐药	EGFR	Exon 20 Insertion c.2369C>T (T790M)
		KRAS	c.34 (G12C、G12R、G12S) c.35 (G12A、G12D、G12V) c.37 (G13C、G13R、G13S) c.38 (G13A、G13D) c.181 (Q61K) c.182 (Q61L、Q61R) c.183 (Q61H)
		MET	Amplification
	药敏	EGFR	c.2156 (G719A) c.2155 (G719C、G719S) Exon 19 Deletion Exon 19 Insertion c.2290_2291ins (A763_Y764insFQEA) c.2573 (L858R) c.2582 (L861Q)
ALK TKI	耐药	ALK	1151Tins L1152R C1156Y F1174L L1196M G1202R S1206Y G1269A
	药敏	ALK	fusion
		ROS1	Fusions

EGFR TKI治疗的选择

对象	有效率
不加选择的中国患者	30%
根据临床特征选择优势人群	50%
根据EGFR敏感基因选择	80%
+耐药基因（T790M，KRAS、EML4-ALK etc）排除	90-95%

高血压&基因&药物

	传统治疗	精准治疗
模式	不断更换药物，寻找合适药物	明确诊断，找到危险因素，直接选用针对药物
特点	耗费时间，费用	简便、直接、优秀

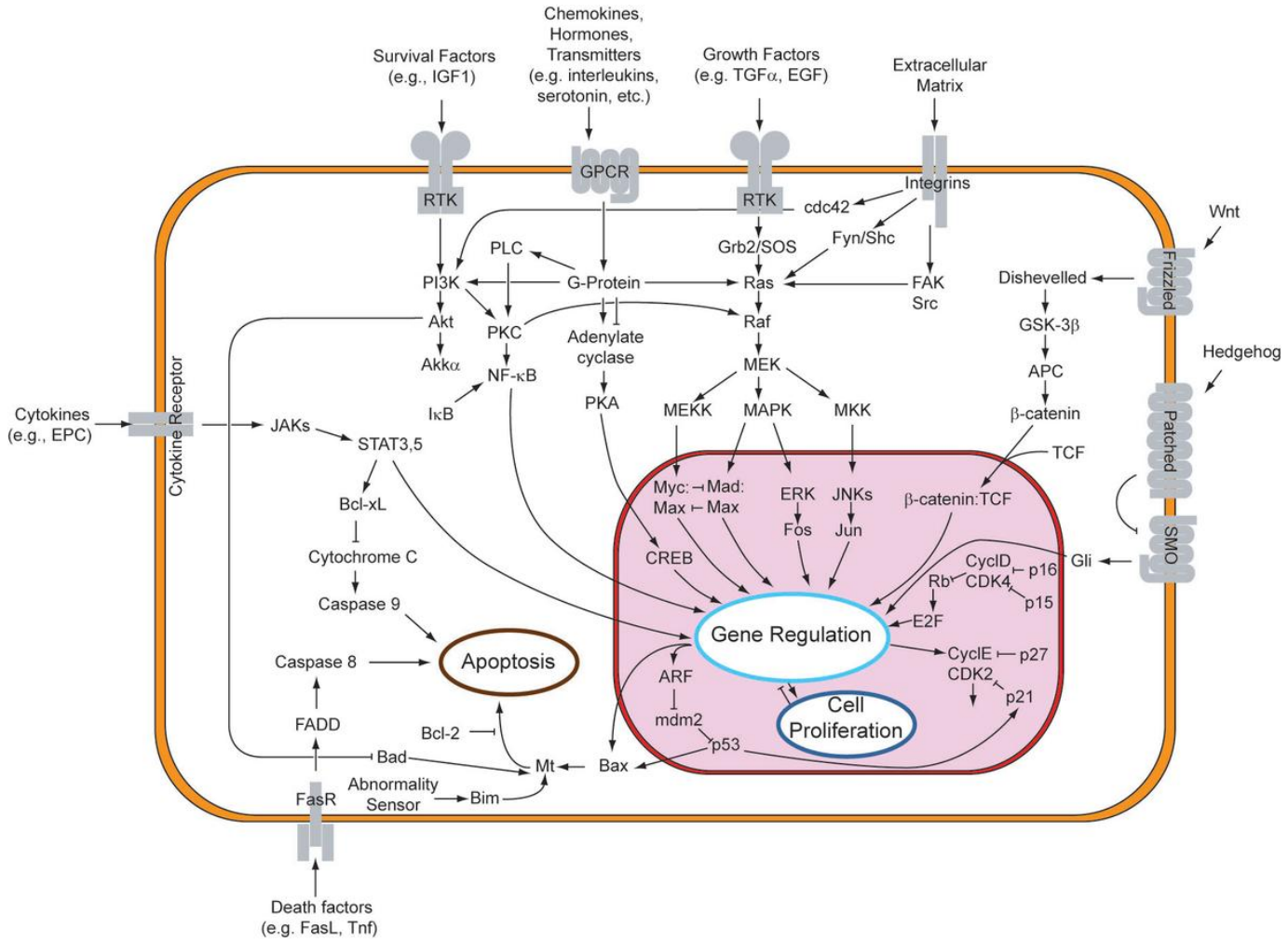
Clinical Applications

- Cancer

Cancer genes

- During the evolution of malignant tumors, a series of gene mutations often accumulate, including those in proto-oncogenes, tumor suppressor genes, cell cycle regulatory genes, apoptosis-related genes, and genes that maintain genomic stability.
  These genes can all potentially be selected as biomarkers for tumor diagnosis.

Signal transduction



- **1. Proto-oncogenes**
- **Normal role:** Proto-oncogenes are normal genes that encode proteins involved in cell growth, division, and survival. They act like "accelerators" for the cell cycle, helping cells respond to growth signals when needed.
- **In cancer:** When mutated or abnormally activated, proto-oncogenes become **oncogenes**, which drive uncontrolled cell proliferation. Even a single gain-of-function mutation can be enough to push cells toward malignancy.
- **Examples:**
  - **RAS family (KRAS, HRAS, NRAS):** Mutations cause continuous signaling for cell growth.
  - **MYC:** Promotes transcription of genes for cell division.
  - **HER2/neu (ERBB2):** Encodes a receptor tyrosine kinase often amplified in breast cancer.

- **2. Tumor Suppressor Genes**
- **Normal role:** These act like the "brakes" of the cell cycle, preventing uncontrolled division. They repair DNA damage, halt the cell cycle if errors are detected, or trigger apoptosis if damage is irreparable.
- **In cancer:** Loss-of-function mutations (often requiring *both alleles* to be inactivated, following Knudson's two-hit hypothesis) remove these safety checks, allowing malignant growth.
- **Examples:**
  - **TP53 ("guardian of the genome"):** Induces cell cycle arrest or apoptosis in response to DNA damage. Mutated in over 50% of cancers.
  - **RB1:** Controls the G1 → S cell cycle checkpoint.
  - **BRCA1/2:** Critical for DNA repair via homologous recombination.

- **3. Cell Cycle Regulatory Genes**
- **Normal role:** These genes encode proteins that directly control progression through the different phases of the cell cycle (G1, S, G2, M). Cyclins, cyclin-dependent kinases (CDKs), and checkpoint regulators maintain orderly division.
- **In cancer:** Mutations or dysregulation allow the cell cycle to continue unchecked, bypassing normal checkpoints.
- **Examples:**
  - **Cyclin D/CDK4 complex:** Drives progression through G1. Overexpression can lead to premature entry into S phase.
  - **CDK inhibitors (e.g., p16^INK4a^):** Normally halt cycle progression; often lost in cancer.
  - **ATM and ATR kinases:** Monitor DNA damage and stop cycle progression if errors are found.

- **4. Apoptosis-Related Genes**
- **Normal role:** These genes control programmed cell death, eliminating damaged or unnecessary cells. Apoptosis prevents accumulation of harmful mutations.
- **In cancer:** Mutations allow cancer cells to evade death, making them resistant to therapies that rely on apoptosis.
- **Examples:**
  - **BCL-2 family:**
    - *Pro-survival:* BCL-2, BCL-XL (often overexpressed in cancer, e.g., follicular lymphoma with BCL2 translocation).
    - *Pro-apoptotic:* BAX, BAK.
  - **Caspases:** Execute apoptosis by degrading cellular proteins.
  - **TP53:** Also acts here by inducing pro-apoptotic genes when DNA damage is irreparable.

- **5. Genes That Maintain Genomic Stability**
- **Normal role:** These safeguard the integrity of DNA by repairing mutations and maintaining chromosome structure. They prevent accumulation of genetic errors during cell division.
- **In cancer:** Loss of function leads to a "mutator phenotype," where cells acquire mutations at a high rate, accelerating tumor evolution.
- **Examples:**
  - **Mismatch repair genes (MLH1, MSH2, MSH6, PMS2):** Defects cause microsatellite instability, common in colorectal cancer.
  - **DNA repair genes:**
    - *Nucleotide excision repair:* Defects → xeroderma pigmentosum, leading to skin cancer susceptibility.
    - *Homologous recombination repair:* BRCA1/2 (also tumor suppressors).
  - **Telomere maintenance genes:** Dysfunction leads to chromosomal instability.

Summary

- **Proto-oncogenes:** Growth accelerators → mutations turn them into "stuck accelerators."
- **Tumor suppressors:** Growth brakes → mutations disable them.
- **Cell cycle regulators:** Gatekeepers → mutations remove checkpoints.
- **Apoptosis genes:** Suicide program → mutations block self-destruction.
- **Genomic stability genes:** DNA caretakers → mutations let damage pile up.
- Together, mutations in these categories allow cells to grow uncontrollably, avoid death, and accumulate more mutations — the hallmarks of cancer.

Significance of Molecular Biological Testing in Breast Cancer

- **Screening high-risk groups:** About 4%–5% of breast cancers are hereditary. Among breast cancers occurring at age ≤30, **20%–30% have BRCA1, BRCA2, and TP53 mutations**, and the prognosis is generally poor.
- **Reflects the biological behavior of the tumor.**
- **Predicts the risk of early breast cancer recurrence and metastasis.**
- **Helps select patients suitable for endocrine therapy, chemotherapy, and targeted therapy.**

Significance of Molecular Biological Testing in Leukemia:

- Assists in **MIC subtyping (**morphology, immunology, and cytogenetics);
- **Identifies new subtypes**: About 25% of precursor B-cell ALL (acute lymphoblastic leukemia) patients have unknown genetic subtypes, meaning they lack obvious chromosomal abnormalities or fusion genes. These patients are prone to relapse;
- **Monitors minimal residual disease (MRD)**;
- **Helps study the pathogenesis of leukemia.**

Molecular Testing Strategies for Viral Infections

- **Two Main Strategies**

- **General Detection Strategy**
  - Detects whether viral DNA or RNA exists (yes/no).
  - Useful for diagnosing viral infections quickly.
  - Often based on PCR primers targeting conserved viral sequences.
- **Complete Detection Strategy**
  - Provides more detailed insights:
    - Confirms viral presence/absence.
    - Identifies carriers and latent infections.
    - Quantifies viral load.
    - Determines genotypes, subtypes, and drug resistance.

Complete Detection – Clinical Value

- **Accurate diagnosis** of infection and carrier status.
- **Viral load measurement** → monitors infection progression and treatment response.
- **Genotyping & subtyping** → guides therapy choices.
- **Drug resistance testing** → ensures effective treatment selection.

Clinical Significance of HBV Molecular Testing

- **Viral Load Testing**
  - HBV DNA quantification reflects replication level and infectivity.
  - Used to guide treatment decisions and monitor therapy.
- **Genotyping & Subtyping**
  - Certain HBV genotypes/subtypes correlate with disease progression and treatment response.
  - Helps predict prognosis and personalize therapy.
- **Drug Resistance Testing**
  - Detects HBV DNA mutations causing resistance to antiviral drugs.
  - Provides evidence for adjusting treatment regimens.

Techniques

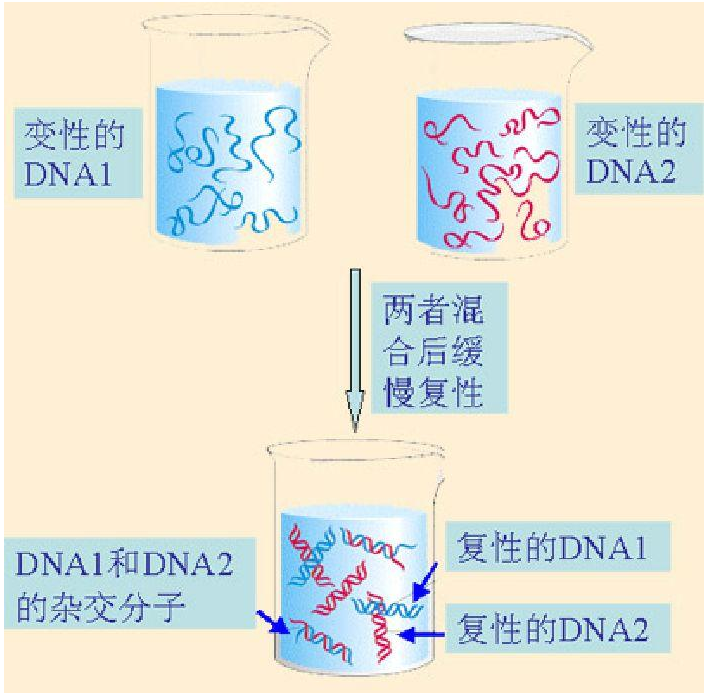
- Hybridization
  - Biochip
- DNA sequencing
- Polymerase chain reaction

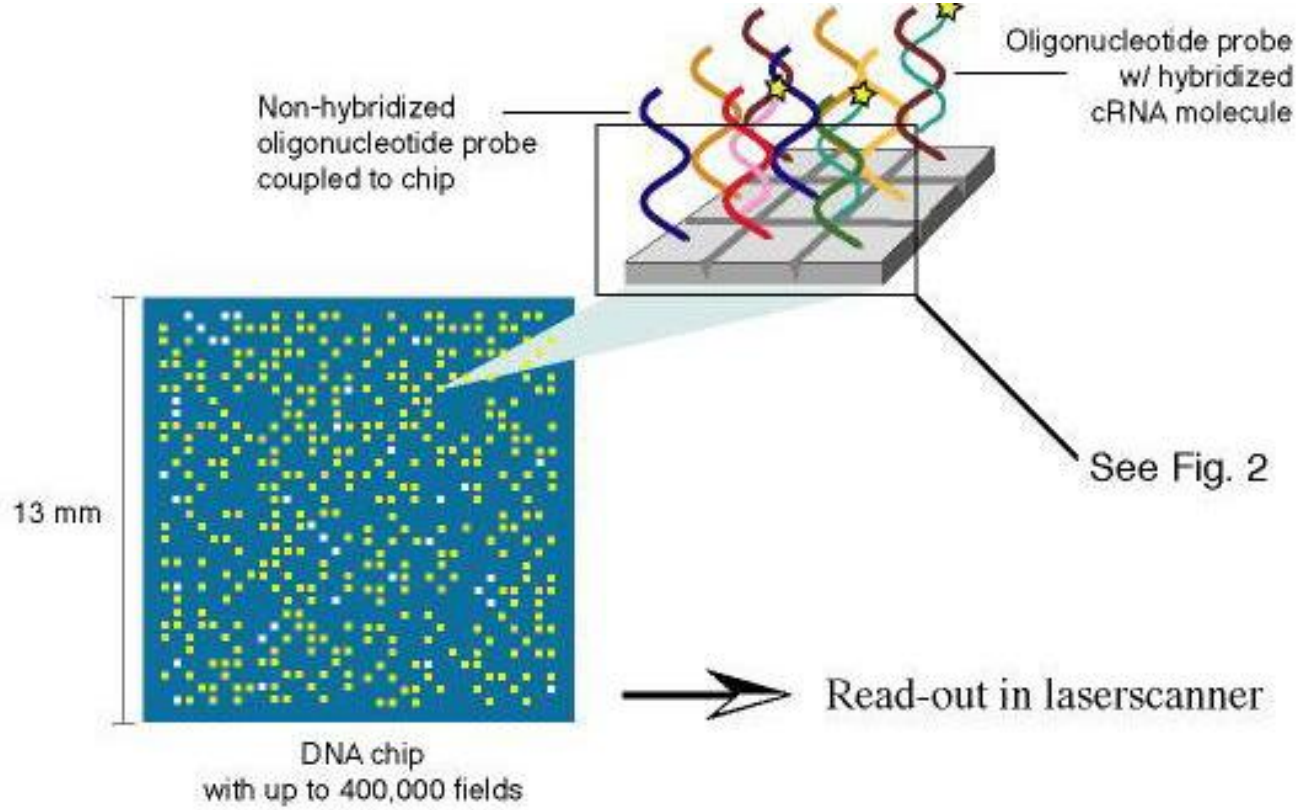
Hybridization

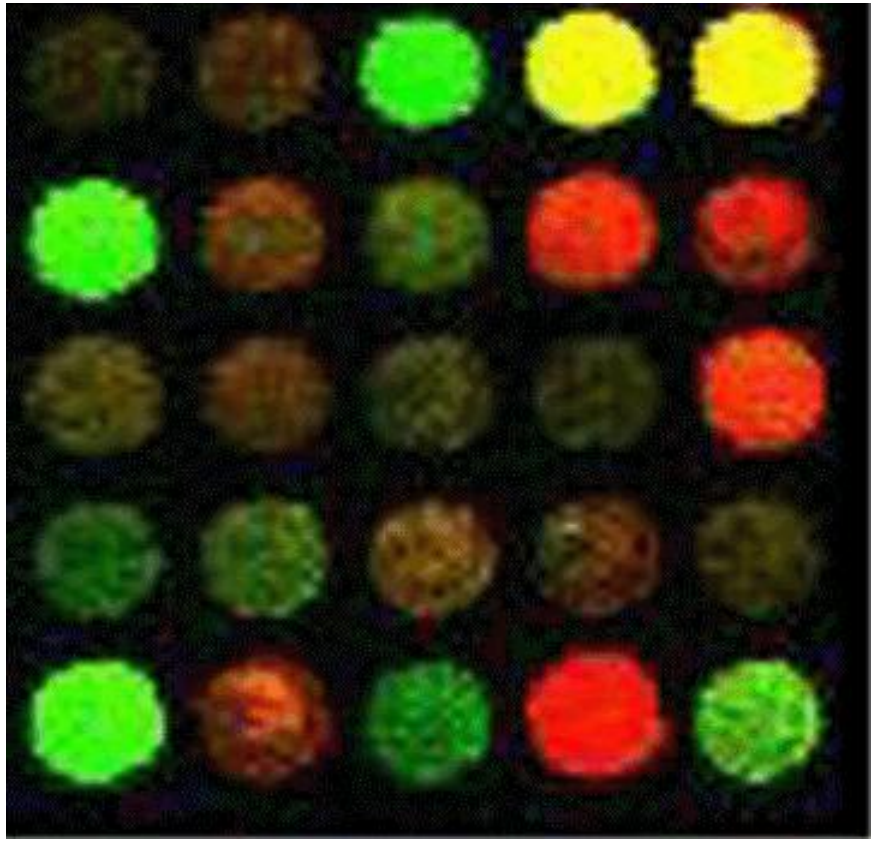
DNA-DNA

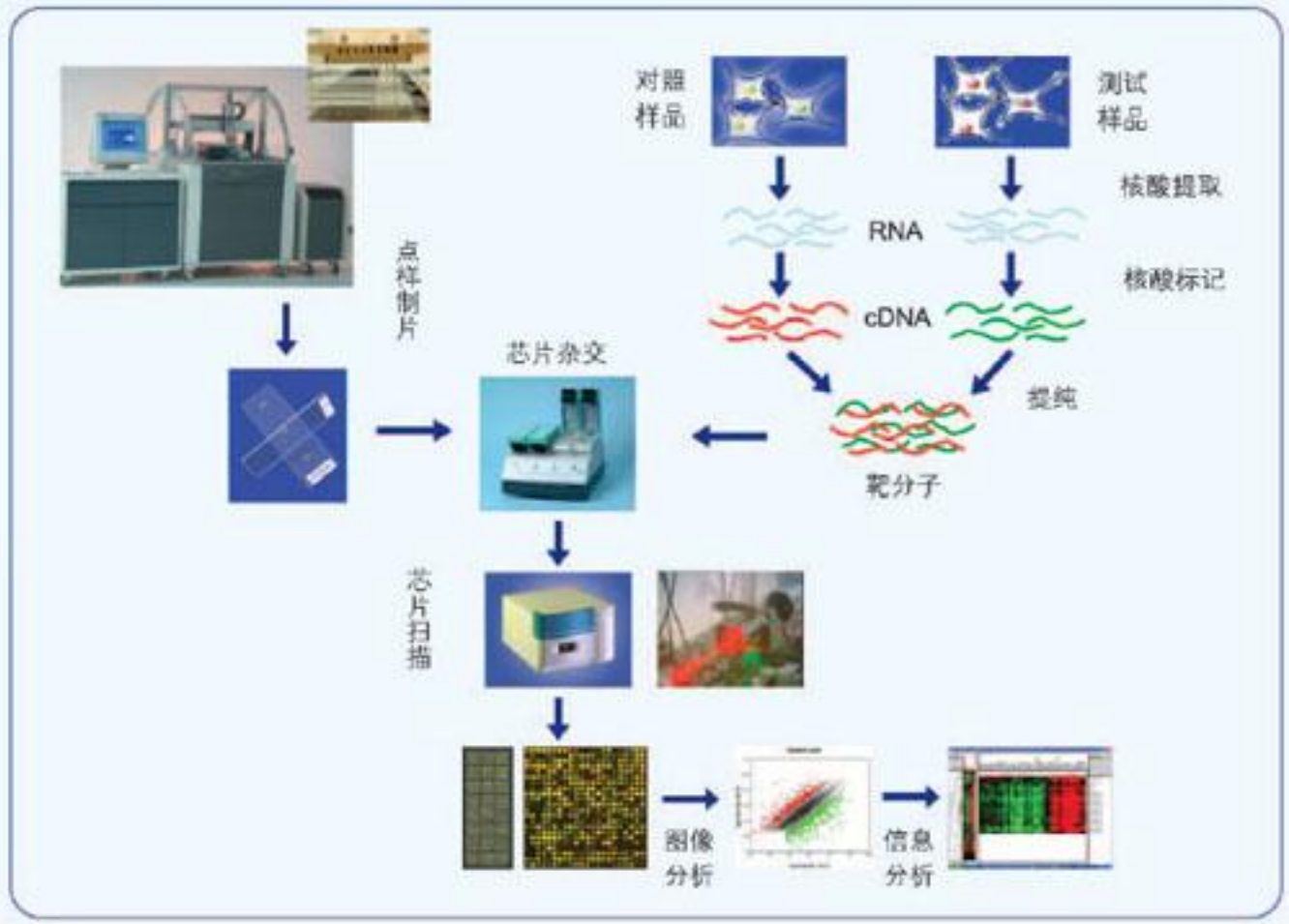
DNA-RNA

RNA-RNA




Oligonucleotide probe
w/ hybridized
cRNA molecule
Non-hybridized
oligonucleotide probe
coupled to chip
See Fig. 2
13 mm
Read-out in laserscanner
DNA chip
with up to 400,000 fields




对照
样品
测试
样品
核酸提取
RNA
核酸标记
cDNA
点样制片
芯片杂交
提纯
靶分子
芯片扫描
图像
分析
信息
分析

DNA sequencing

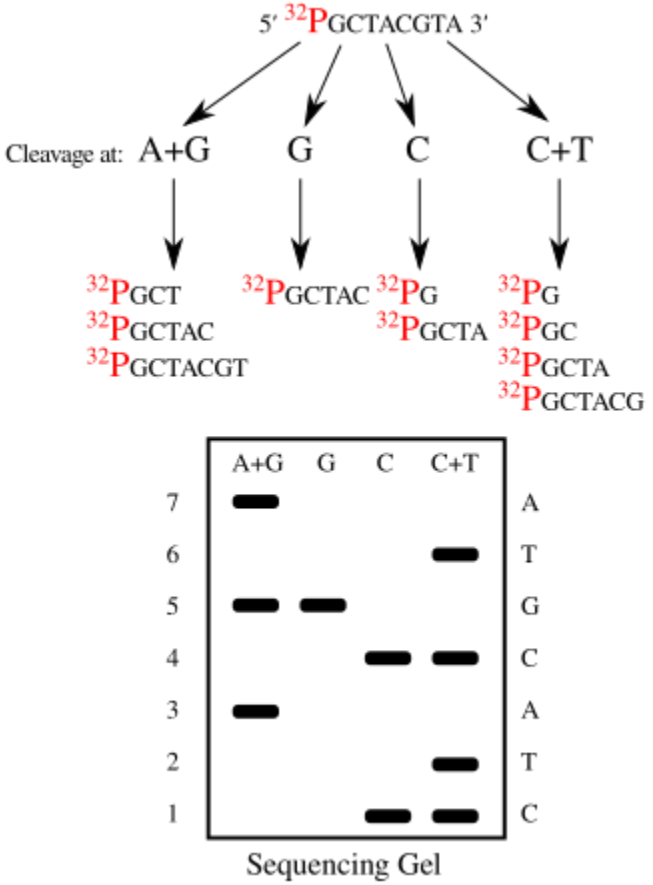
- First generation
  - Maxam–Gilbert sequencing
  - Sanger's method
- Second: NGS
- Third: Single molecule

Maxam–Gilbert sequencing

- The **chemical degradation method** was developed by **Allan Maxam and Walter Gilbert in 1977**, and it was one of the **first DNA sequencing technologies**.
  It determines the sequence of nucleotides (A, T, G, C) in DNA by **chemically modifying and cleaving** DNA strands at specific bases, then analyzing the resulting fragments.

Principle

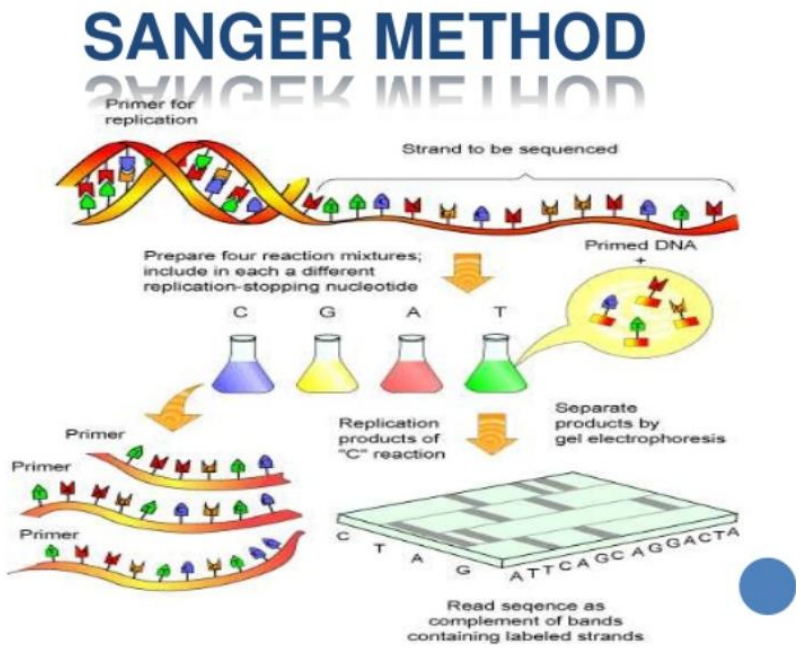
- The method works by:
- **Labeling** one end of a DNA molecule with a radioactive marker (usually at the 5′ end);
- **Chemically treating** the DNA to break it at specific bases (G, A+G, C, or C+T);
- **Separating** the resulting fragments by **polyacrylamide gel electrophoresis (PAGE)**;
- **Visualizing** the bands using autoradiography (X-ray film exposure);
- **Reading** the sequence from the band pattern — the order of fragments corresponds to the nucleotide sequence.


5′ 32PGCTACGTA 3′
Cleavage at: A+G G C C+T
32PGCT
32PGCTAC
32PGCTACGT
32PGCTAC
32PG
32PGCTA
32PG
32PGC
32PGCTA
32PGCTACG
A+G G C C+T
7 A
6 T
5 G
4 C
3 A
2 T
1 C
Sequencing Gel

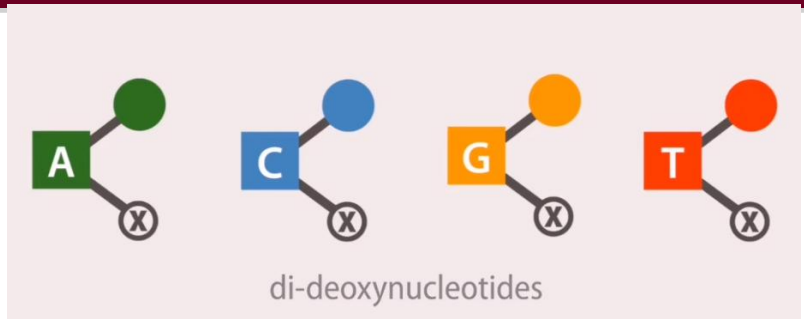
- https://www.youtube.com/watch?v=V9a3OcjN0lM

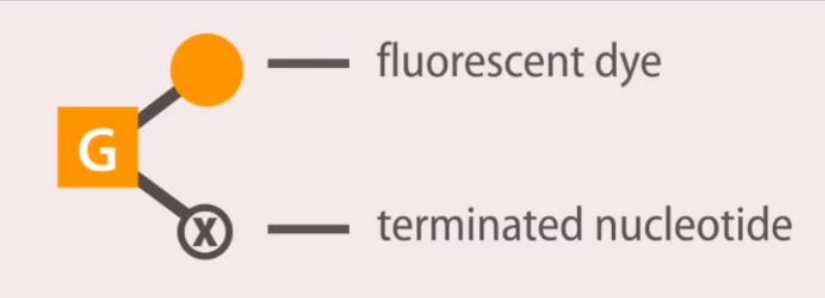
Sanger sequencing

- 1977，the Sanger Method, or Chain Termination Method
- 视频：0:40-2:25

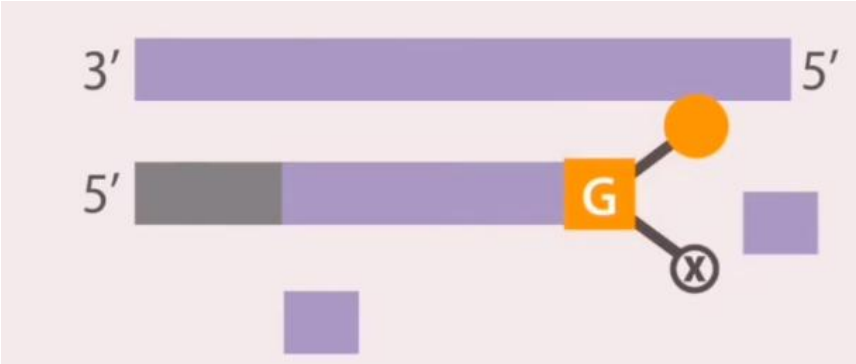


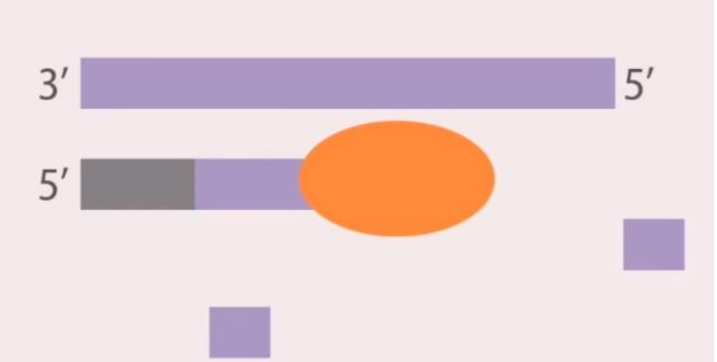
https://www.bilibili.com/video/BV1bE411c77b?from=search&seid=17190867935327337723

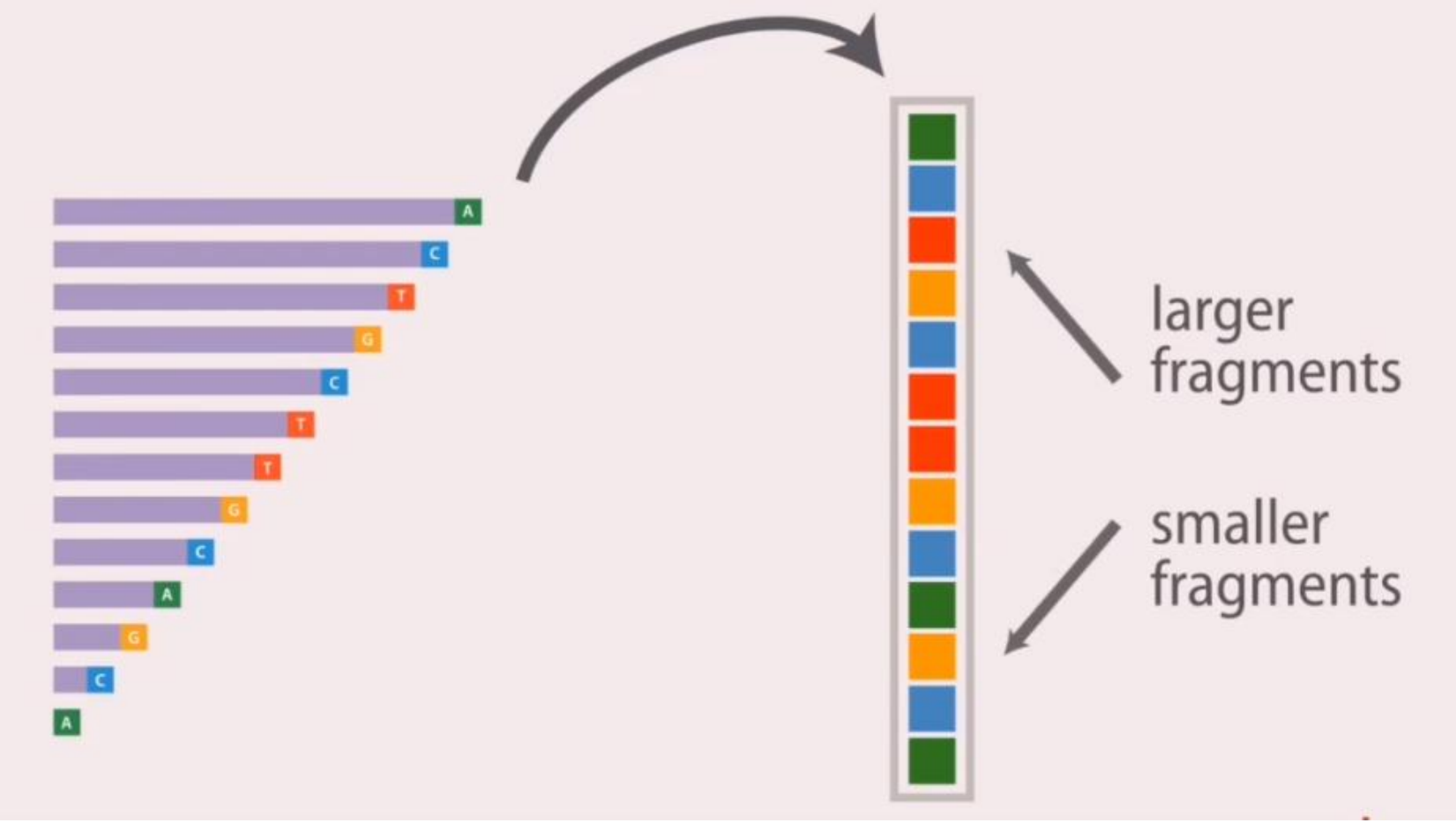




ddNTP vs dNTP






A
C
T
G
C
T
T
G
C
A
G
C
A
larger fragments
smaller fragments

或：凝胶电泳

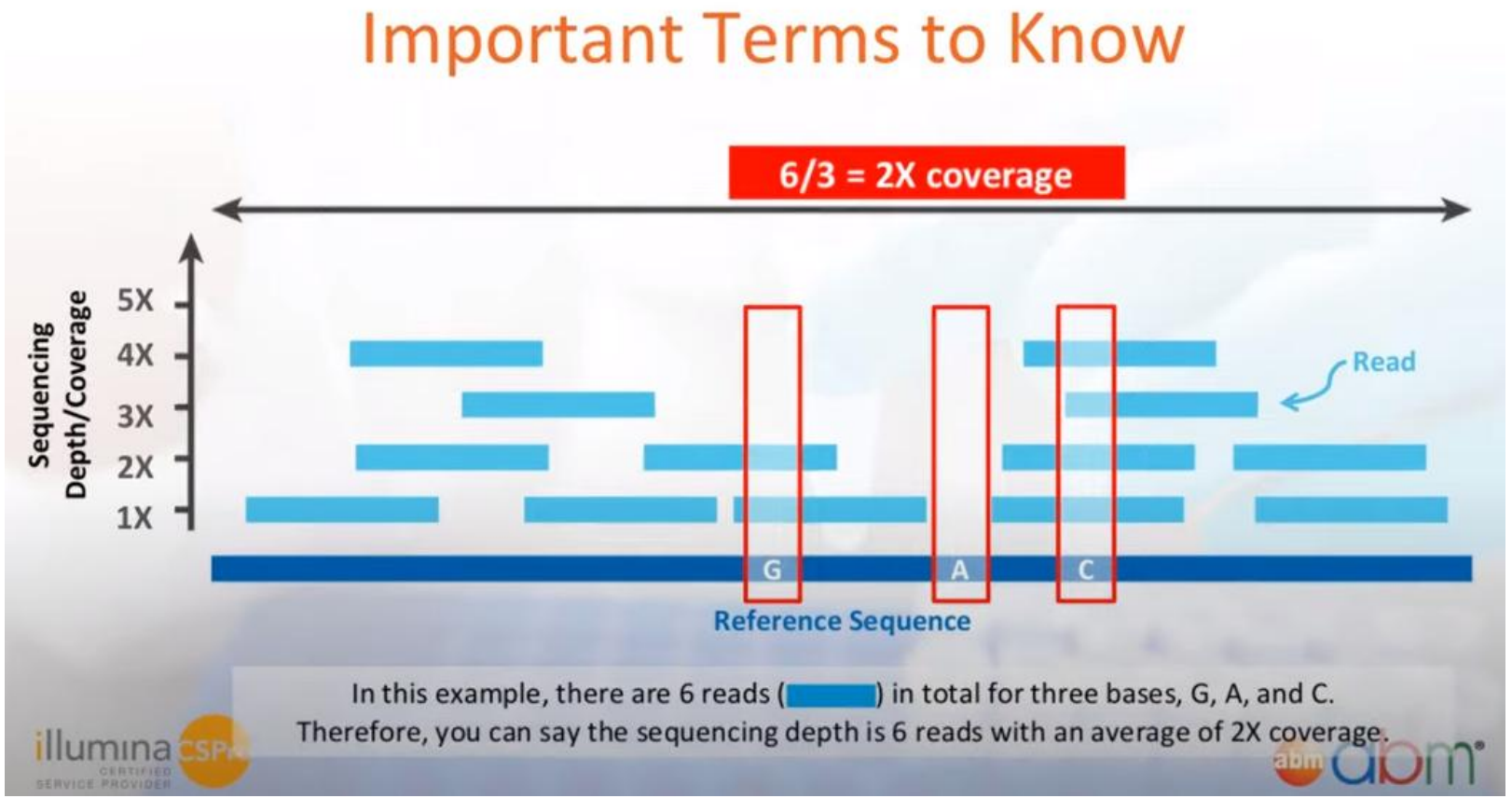
Gel electrophoresis: running gels； 跑胶

NGS

- 特点
  - 通量高
  - 读长较短
    - < 300 nt

测序的深度和覆盖率

- Depth and coverage



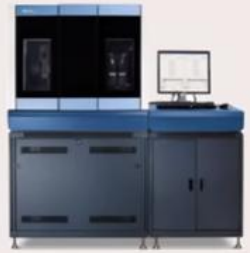
https://www.youtube.com/watch?v=8lAVfKbRK3I

第二代测序 (NGS)

Illumina MySeq

Illumina HiSeq

454 Sequencer



SOLiD System

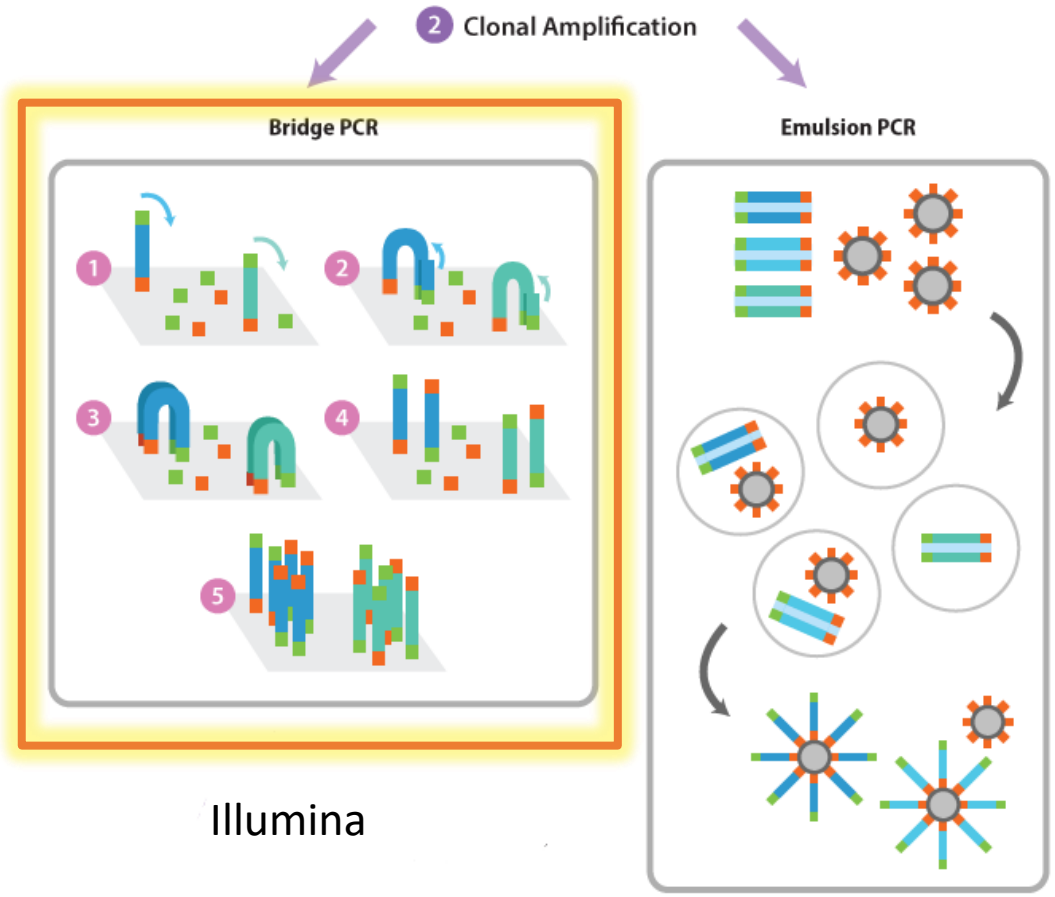
Ion Proton

Examples of NGS systems

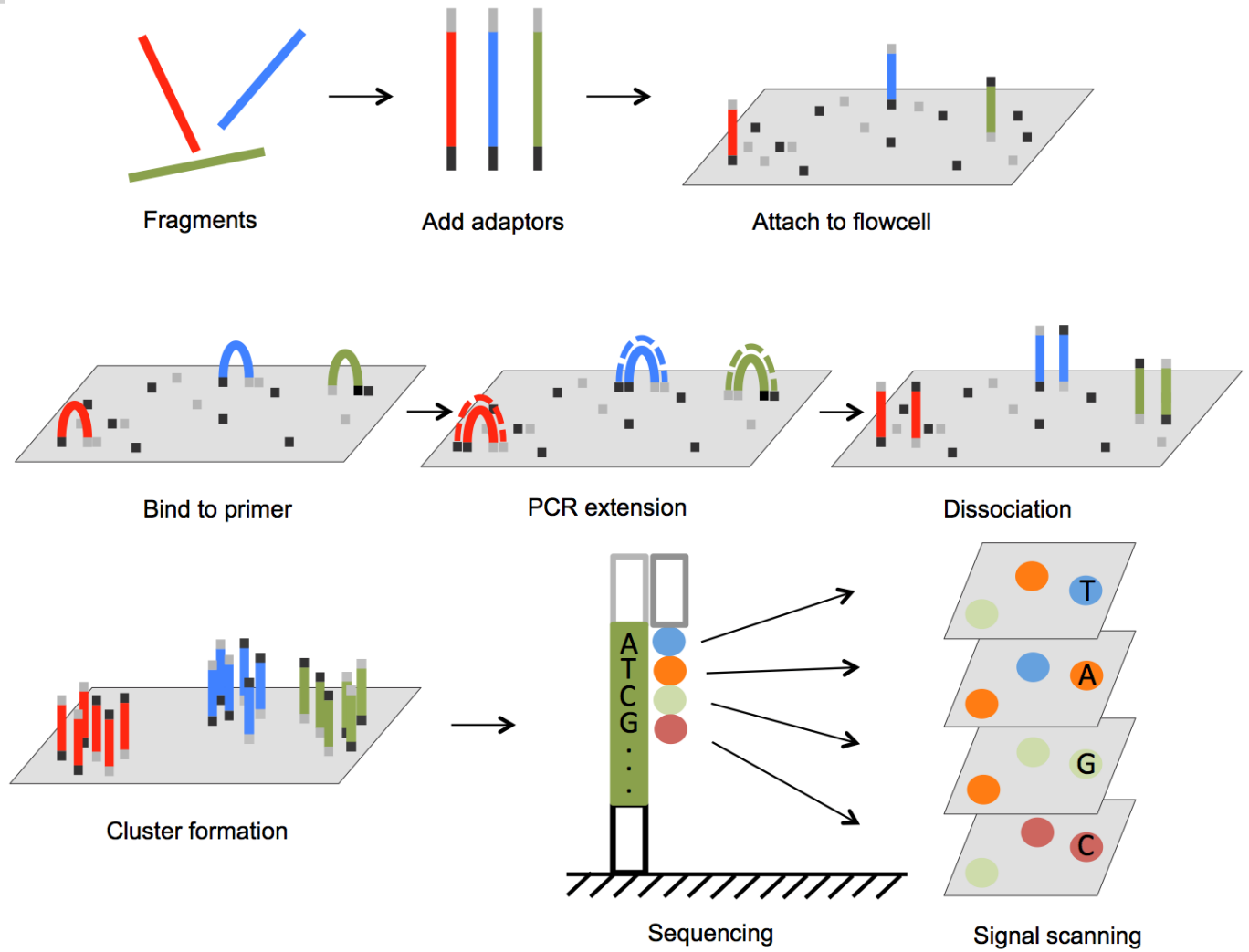
Next Generation Sequencing

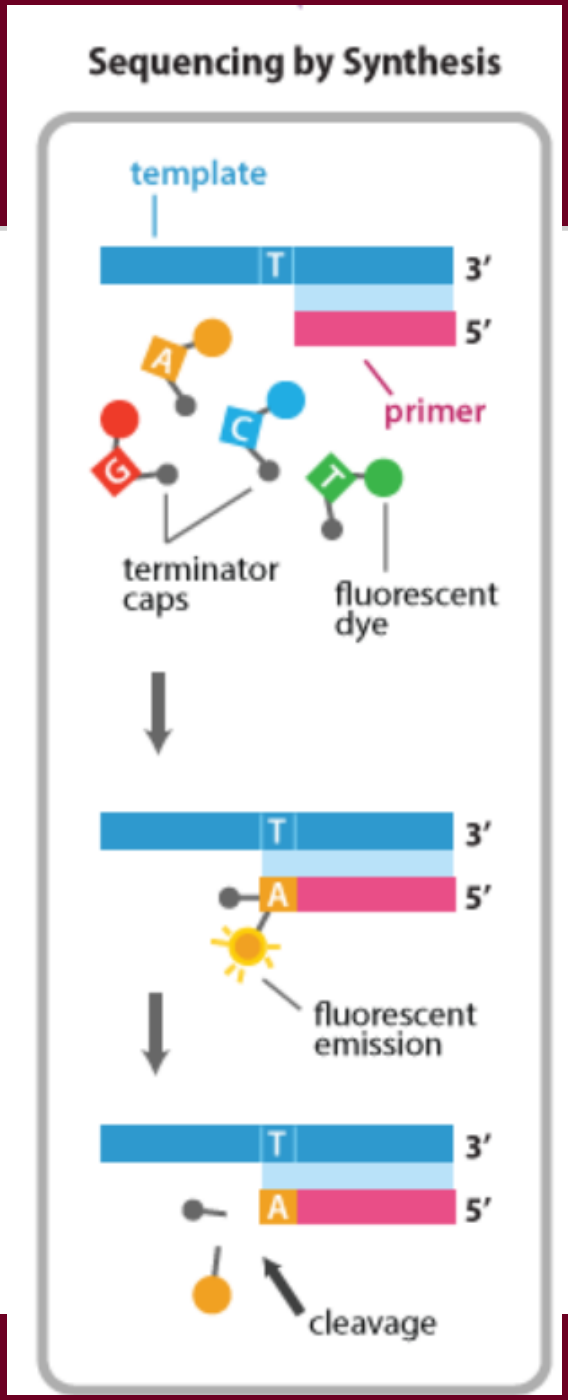
1 Library Preparation

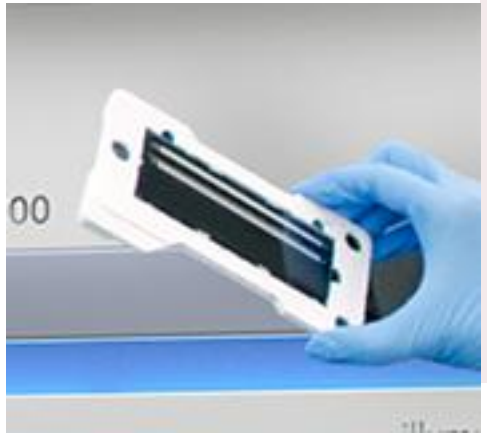
Input DNA → Fragmentation → Adaptor ligation



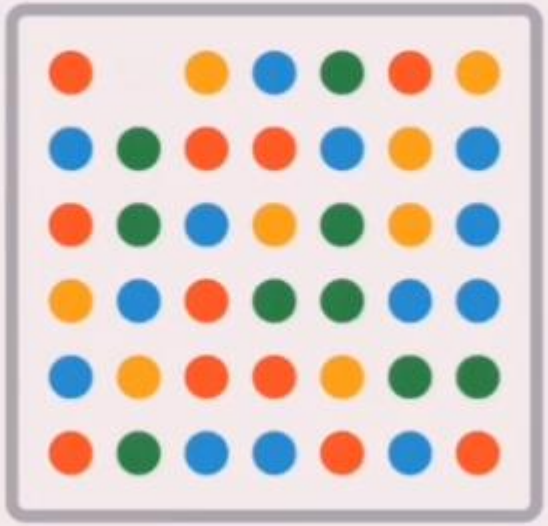
Illumina


Fragments
Add adaptors
Attach to flowcell
Bind to primer
PCR extension
Dissociation
Cluster formation
ATCG . . .
Sequencing
T
A
G
C
Signal scanning

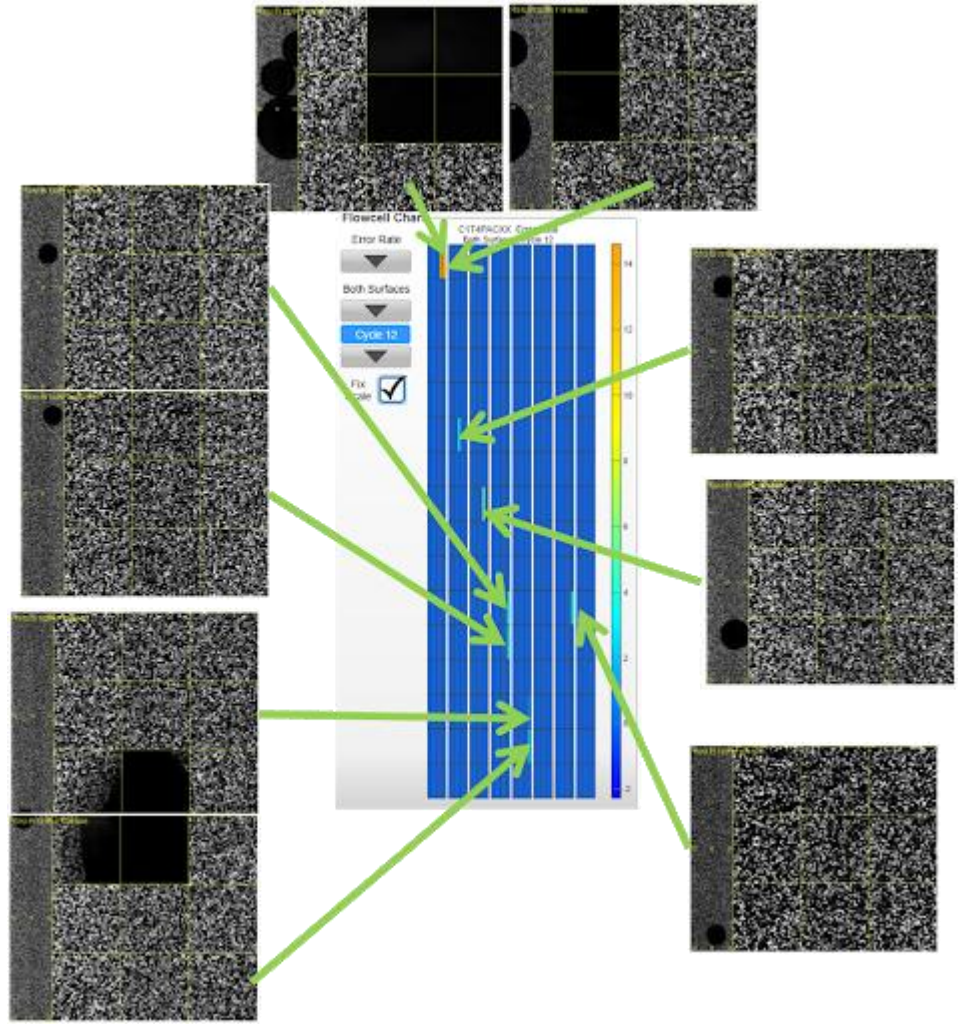

Sequencing by Synthesis
template
T
3′
5′
primer
A
C
G
T
terminator caps
fluorescent dye
T
3′
A
5′
fluorescent emission
T
3′
A
5′
cleavage



3. Data output



Raw data presented on DNA chips

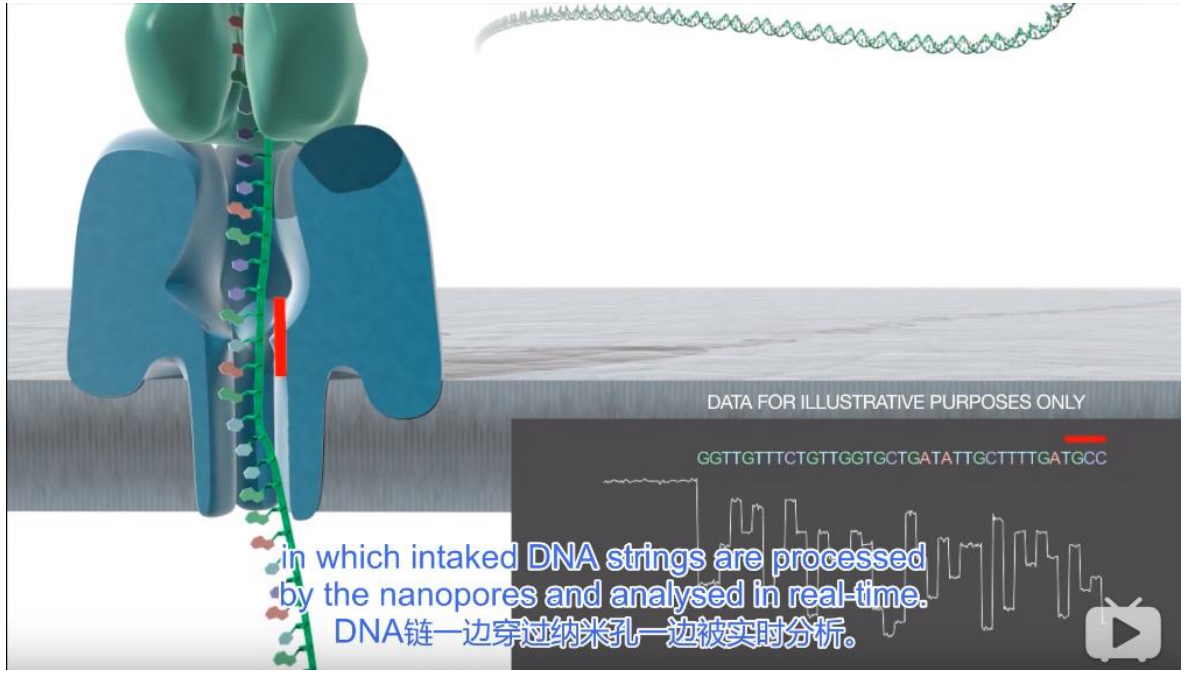

Flowcell Char
Error Rate
Both Surfaces
Cycle 12
Fix Scale

- bilibili.com/video/BV1EJ411n75i/

Third generation

- 特点
  - Single molecule
  - Long read
- Oxford Nanopore
  - Notably, the mean flow cell **read lengths** ranged from ~13 kb to ~ 20 kb, with a maximum **read length** of ~134 kb
- PacBio
  - Average 10–14 kb

Nanopore sequecing



- https://www.bilibili.com/video/BV1E4411h738/?spm_id_from=333.788.videocard.0

PacBio

- Single Molecule, Real-Time (**SMRT**) **Sequencing**
- https://www.bilibili.com/video/BV1Pt4y127dt?from=search&seid=14347789591565060087

PacBio

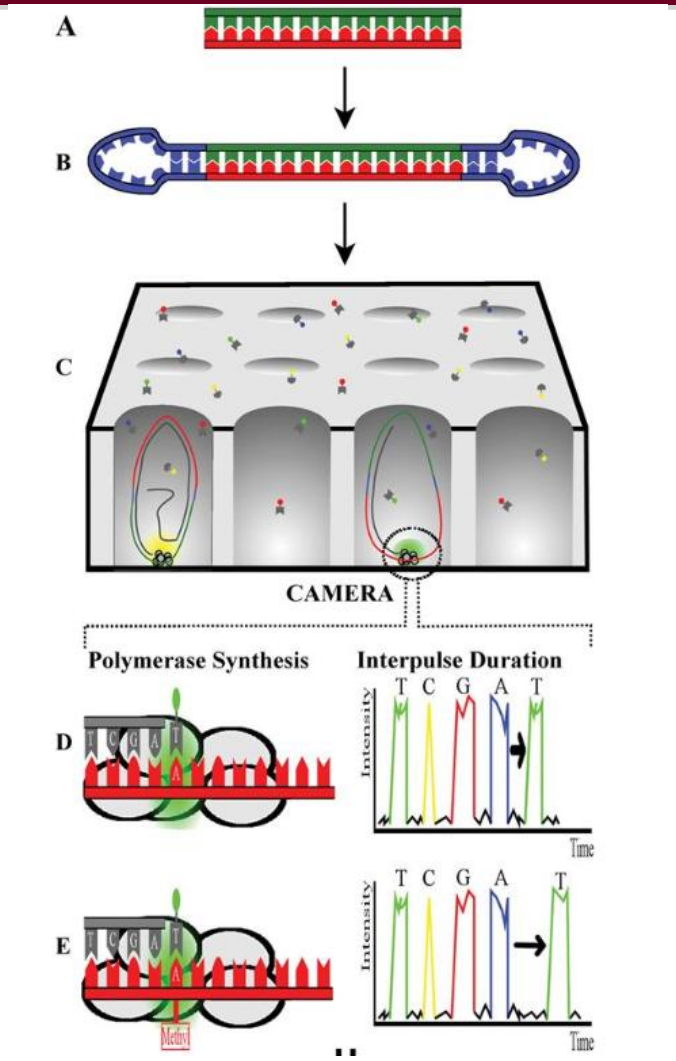

A
B
C
CAMERA
Polymerase Synthesis
Interpulse Duration
D
E
T C G A T
Intensity
Time
T C G A T
Intensity
Time

Table 1.

Comparison of PacBio sequencing platforms to two current industry standards

Platform	Read length	Number reads	Error rate	Run rime
PacBio RSII (per SMRT cell)	Average 10–16 kb	~55 000	13–15%	0.5–6 hours
PacBio Sequel (per SMRT cell)	Average 10–14 kb	~365 000	13–15%	0.5–10 hours
Illumina HiSeq 4000	2 × 150 bp	5 billion	~0.1%	<1–3.5 days
Illumina MiSeq	2 × 300 bp	25 million	~0.1%	4–55 hours

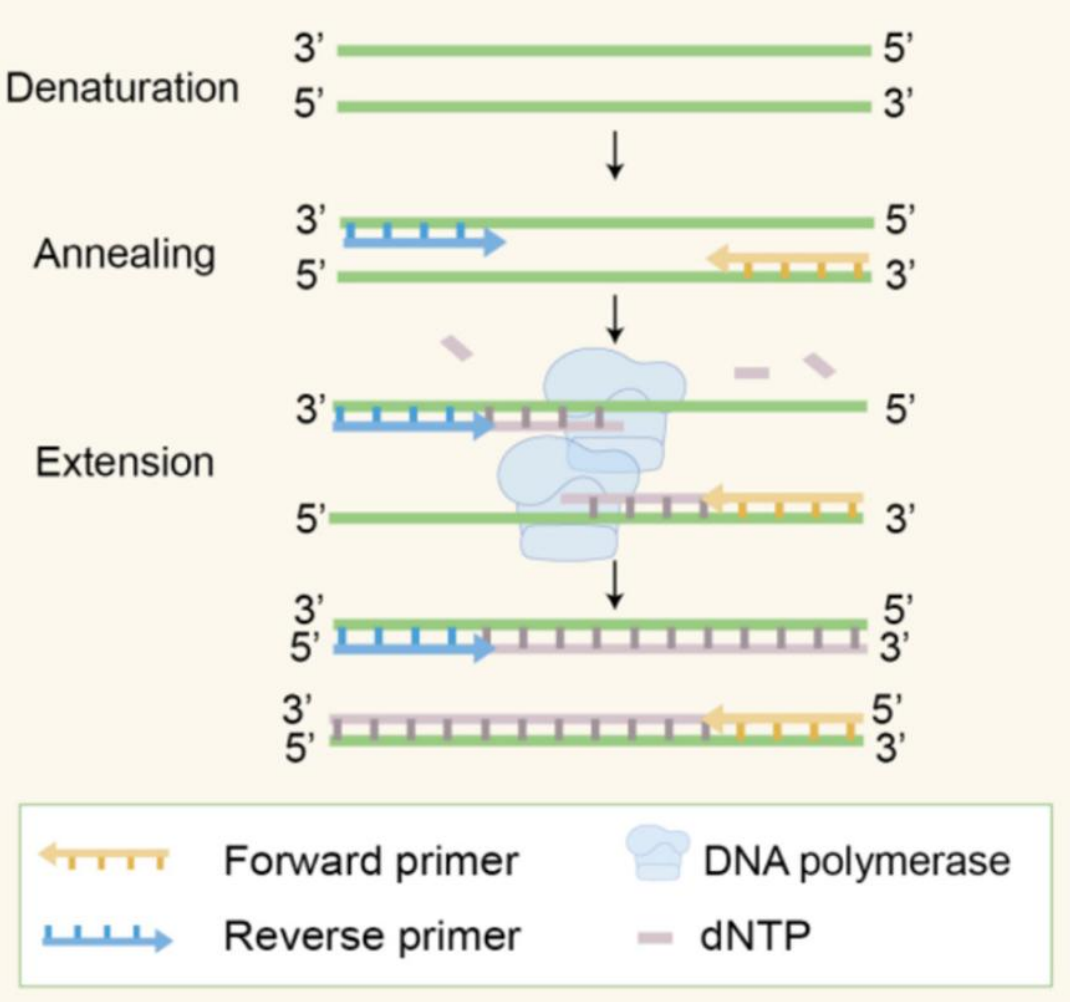
Numbers from personal experience and company website (www.pacb.com and www.illumina.com) queries on 14 November 2017.

Simon Ardui, *Nucleic Acids Research*, Volume 46, Issue 5, 16 March 2018

Nucleic acid amplification

- Polymerase chain reaction
- LAMP
- RPA
- RCA

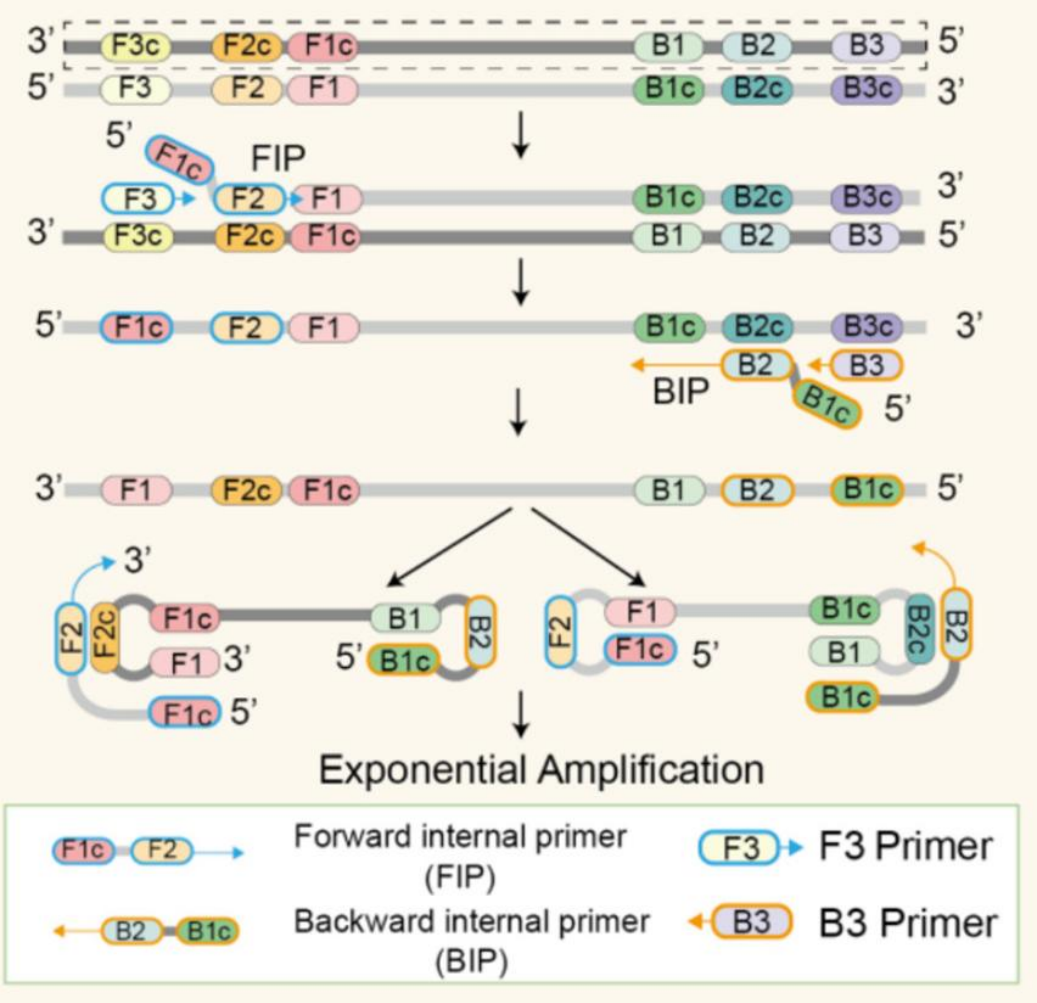
PCR



1.Denaturation (94–95 °C):
DNA double helix separates into single strands.
2.Annealing (50–65 °C):
Short synthetic primers bind (anneal) to target sequences.
3.Extension (72 °C):
Taq DNA polymerase extends the primers, synthesizing new DNA strands.

- One cycle doubles the DNA → repeated **~25–35 times**
- After *n* cycles → roughly **2^n copies** of target sequence

LAMP



Overview

- Developed by **Notomi et al., 2000**
- DNA amplification method performed at a **constant temperature** (≈60–65 °C) — **no thermal cycler needed**
- Uses **strand-displacing DNA polymerase** and **4–6 specially designed primers** recognizing **6–8 regions** of the target DNA

Main Steps

1.Primer Binding:

Four primers (F3, B3, FIP, BIP) recognize six regions of the target DNA.

2.Strand Displacement & Loop Formation:

The **Bst DNA polymerase** synthesizes new strands while displacing existing ones, forming **looped structures** at the ends.

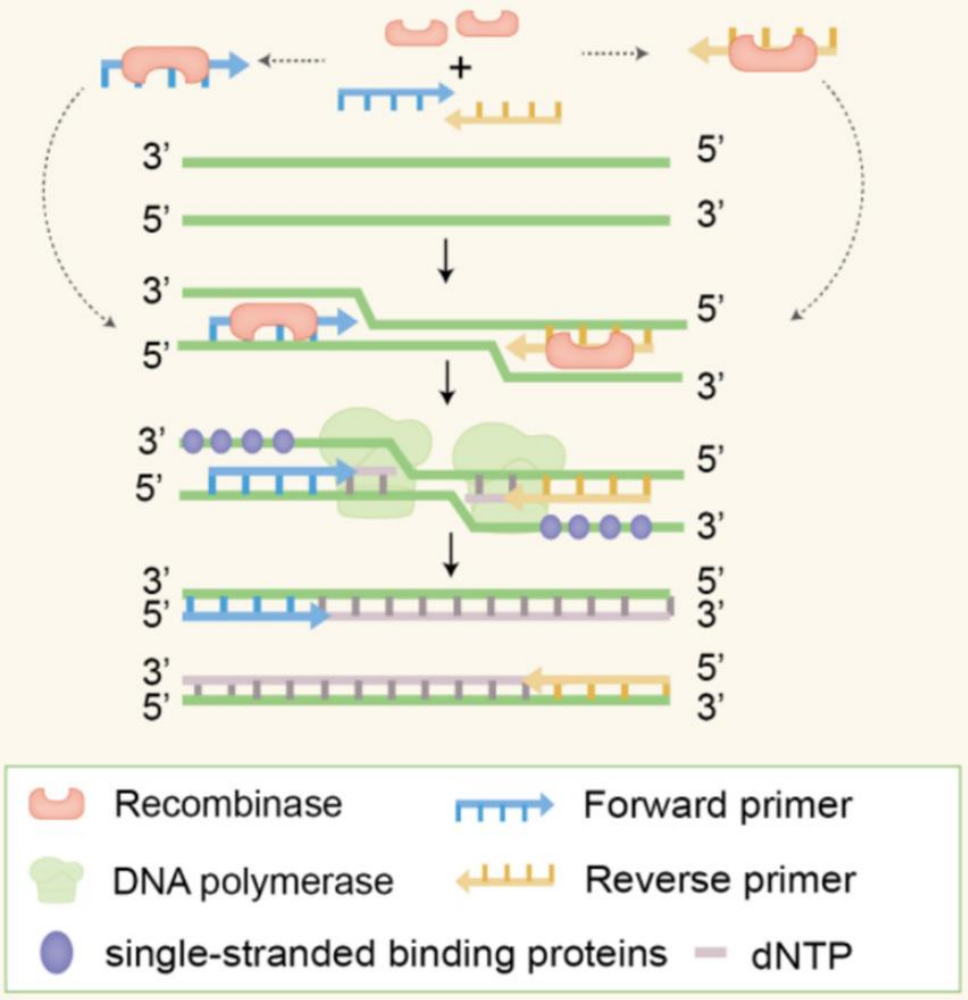
3.Cycling Amplification:

The looped DNA serves as a template for **continuous self-primed amplification**, generating large amounts of DNA.

4.Product Detection:

Amplification produces **turbidity**, **fluorescence**, or **color change** (visible even to the naked eye).

RPA



🧬 **Overview**

•Developed in the early 2000s by **TwistDx Ltd.**

•**Isothermal amplification** method — works efficiently at **37–42 °C**

•Requires **no thermal cycling** → suitable for portable, point-of-care diagnostics

•Amplifies DNA rapidly (often **within 10–20 minutes**)

⚙️ **Main Steps**

1.Recombinase–Primer Complex Formation:

Recombinase proteins bind to primers, forming complexes that search for homologous sequences on double-stranded DNA.

2.Strand Invasion:

The complex opens the DNA duplex and inserts primers at complementary sites.

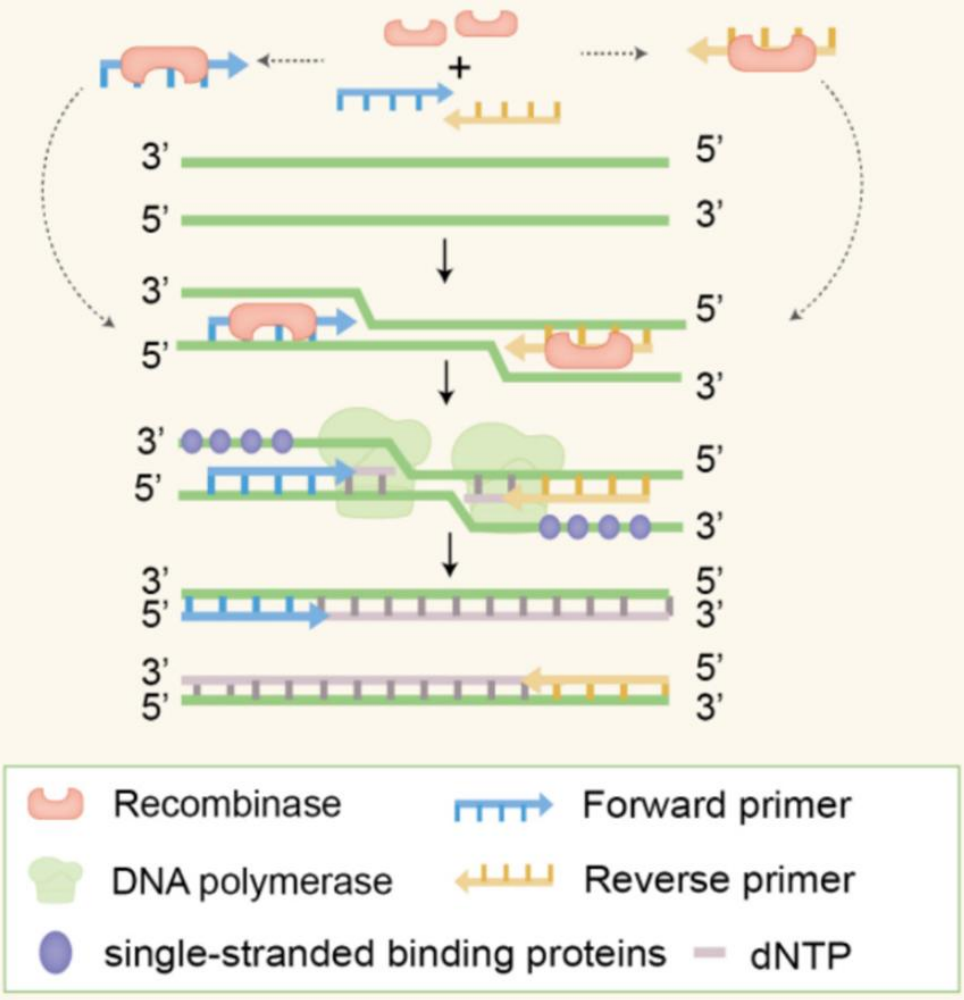
3.DNA Synthesis:

A **strand-displacing DNA polymerase** extends the primers, synthesizing new DNA strands.

4.Exponential Amplification:

The displaced strands serve as templates for additional reactions — rapid, continuous amplification.

RCA



Overview

•Developed in the **1990s** as an **isothermal amplification** technique.

•Uses a **circular DNA template** and a **strand-displacing DNA polymerase** (e.g., Φ29 polymerase).

•Produces **long single-stranded DNA concatemers** containing multiple repeats of the circular sequence.

Main Steps

1.Circular Template Preparation:

The target DNA is circularized using a **ligase** (often after primer binding or padlock probe formation).

2.Primer Annealing:

A short primer binds to the circular template.

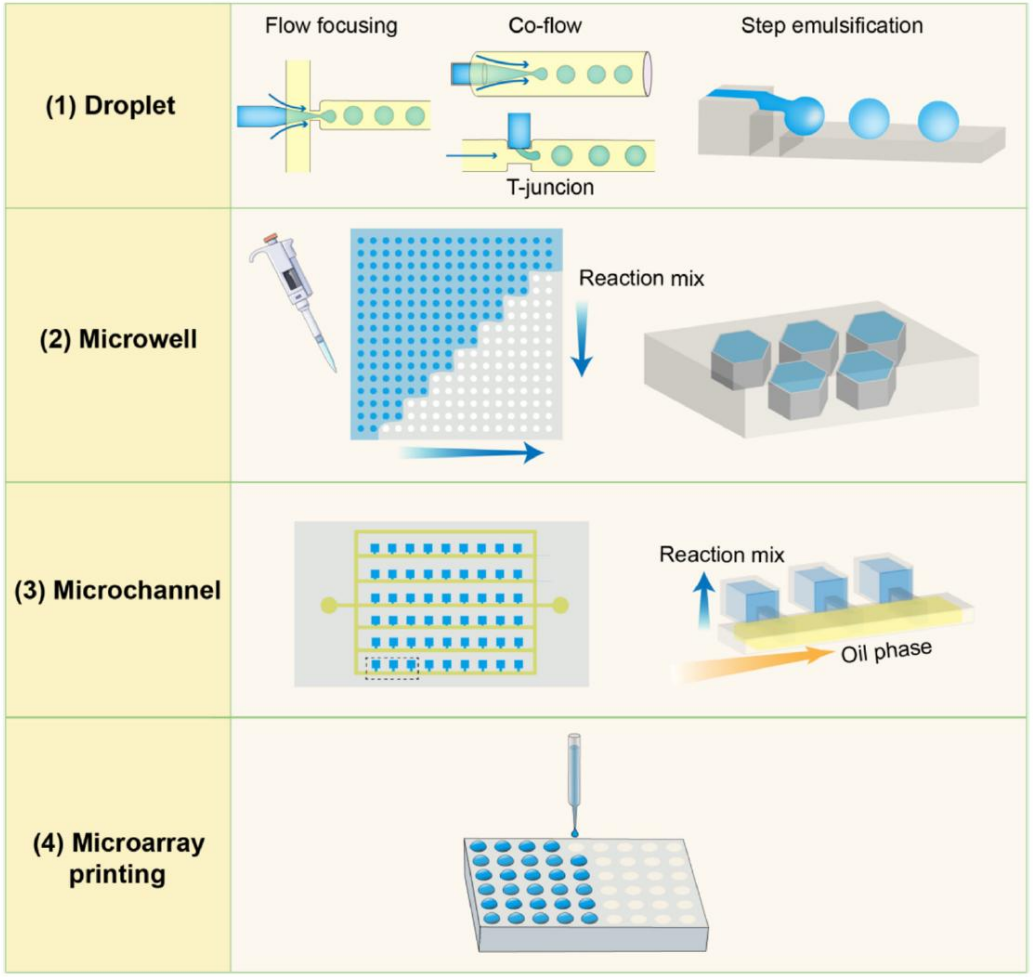
3.Rolling Circle Synthesis:

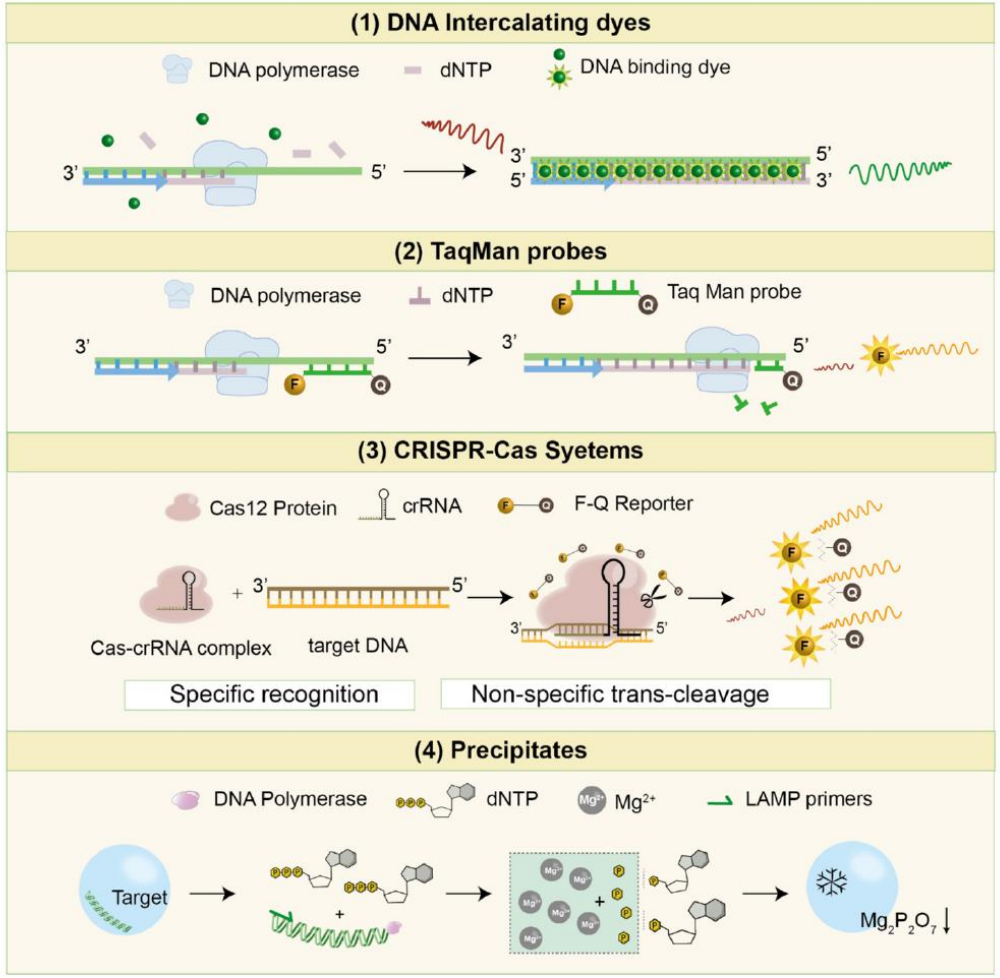
DNA polymerase extends the primer continuously around the circle, **displacing the previous strand** and producing a long linear DNA strand with tandem repeats.

4.(Optional) Hyperbranched RCA:

Additional primers can initiate secondary amplification from the product, increasing yield and sensitivity.

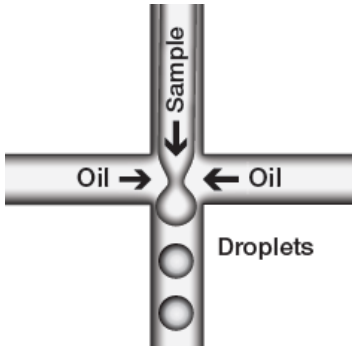
- 🔬 **Key Features**
- **Isothermal** (typically 30–37 °C)
- **Linear amplification** (or exponential in branched RCA)
- **High fidelity** using Φ29 polymerase
- **Applications:**
  - DNA nanotechnology (e.g., DNA origami)
  - Biosensors and microarrays
  - Pathogen and microRNA detection

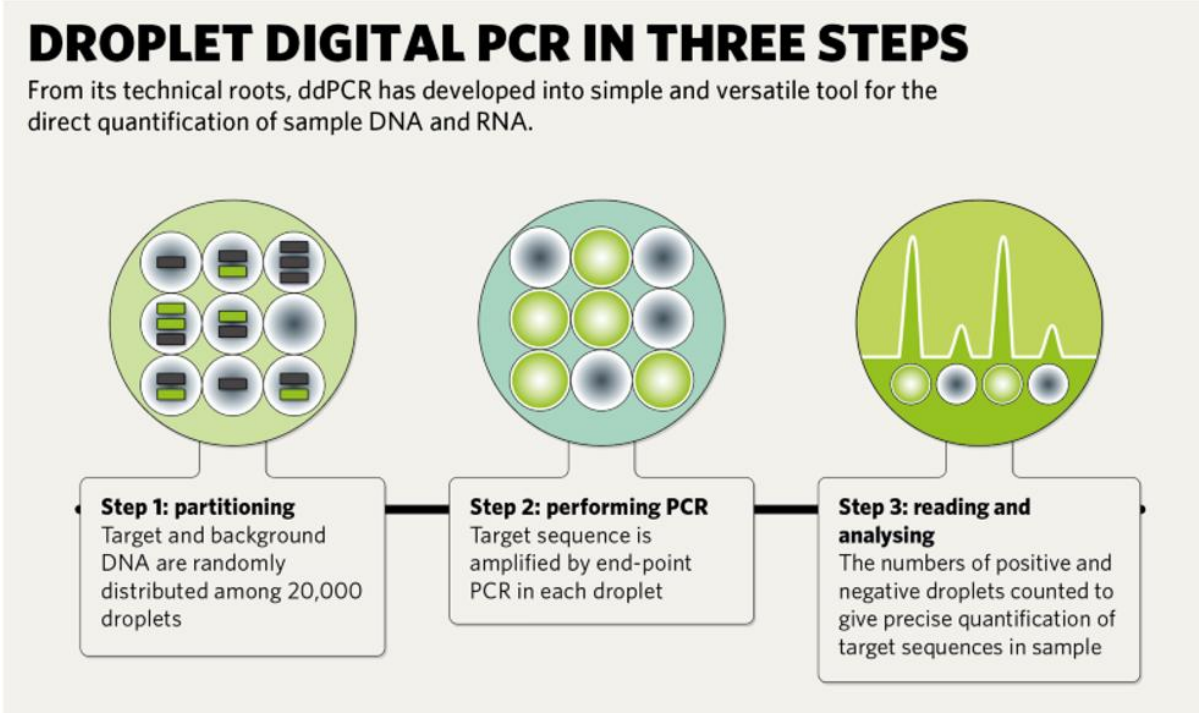

(1) Droplet
Flow focusing
Co-flow
Step emulsification
T-juncion
(2) Microwell
Reaction mix
(3) Microchannel
Reaction mix
Oil phase
(4) Microarray printing


(1) DNA Intercalating dyes
DNA polymerase
dNTP
DNA binding dye
3'
5'
5'
3'
3'
5'
(2) TaqMan probes
DNA polymerase
dNTP
Taq Man probe
F
Q
3'
5'
3'
5'
(3) CRISPR-Cas Syetems
Cas12 Protein
crRNA
F-Q Reporter
3'
5'
3'
5'
Cas-crRNA complex
target DNA
Specific recognition
Non-specific trans-cleavage
(4) Precipitates
DNA Polymerase
dNTP
Mg2+
LAMP primers
Target
Mg2P2O7

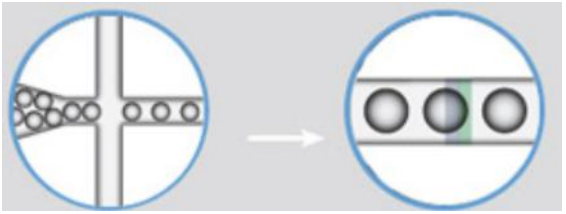
Part II:
Digital Nucleic Acid Detection

Background:
Digital PCR Requires Complex Equipment





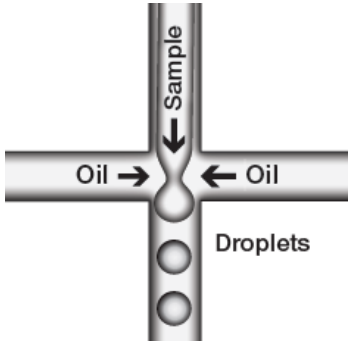
Bio-Rad Laboratories

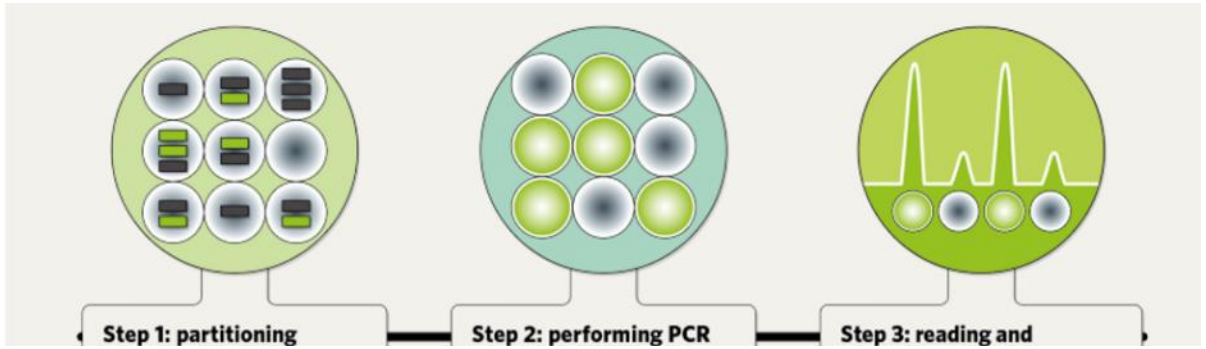


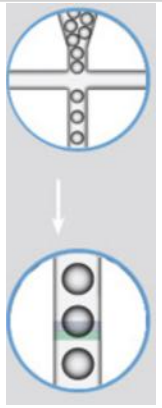
Droplet Generator

Droplet Reader

Background:
Uniform Droplets -> Non-uniform Droplets
Fluorescence -> Brightfield







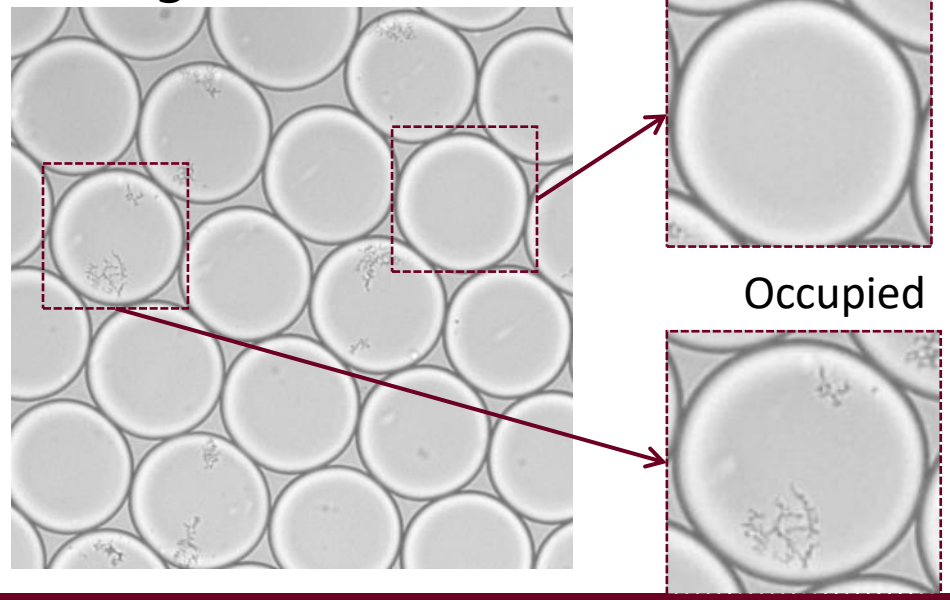
Algorithm for Uniform Droplets

$$C = -\ln P^{-} / V$$

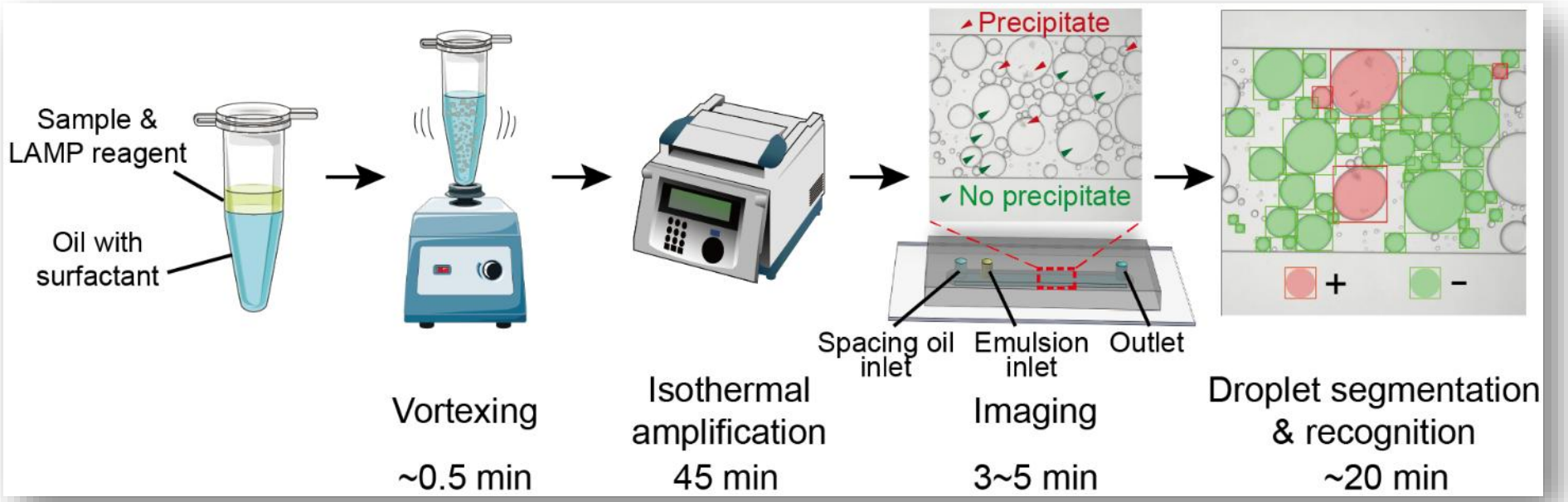
Non-uniform Droplets

$$\sum_{i=1}^{M} V_i^{-} + \sum_{j=1}^{N} V_j^{+} = \sum_{j=1}^{N} \frac{V_j^{+}}{[1 - \exp(-V_j^{+} \cdot C)]}$$

- Brightfield

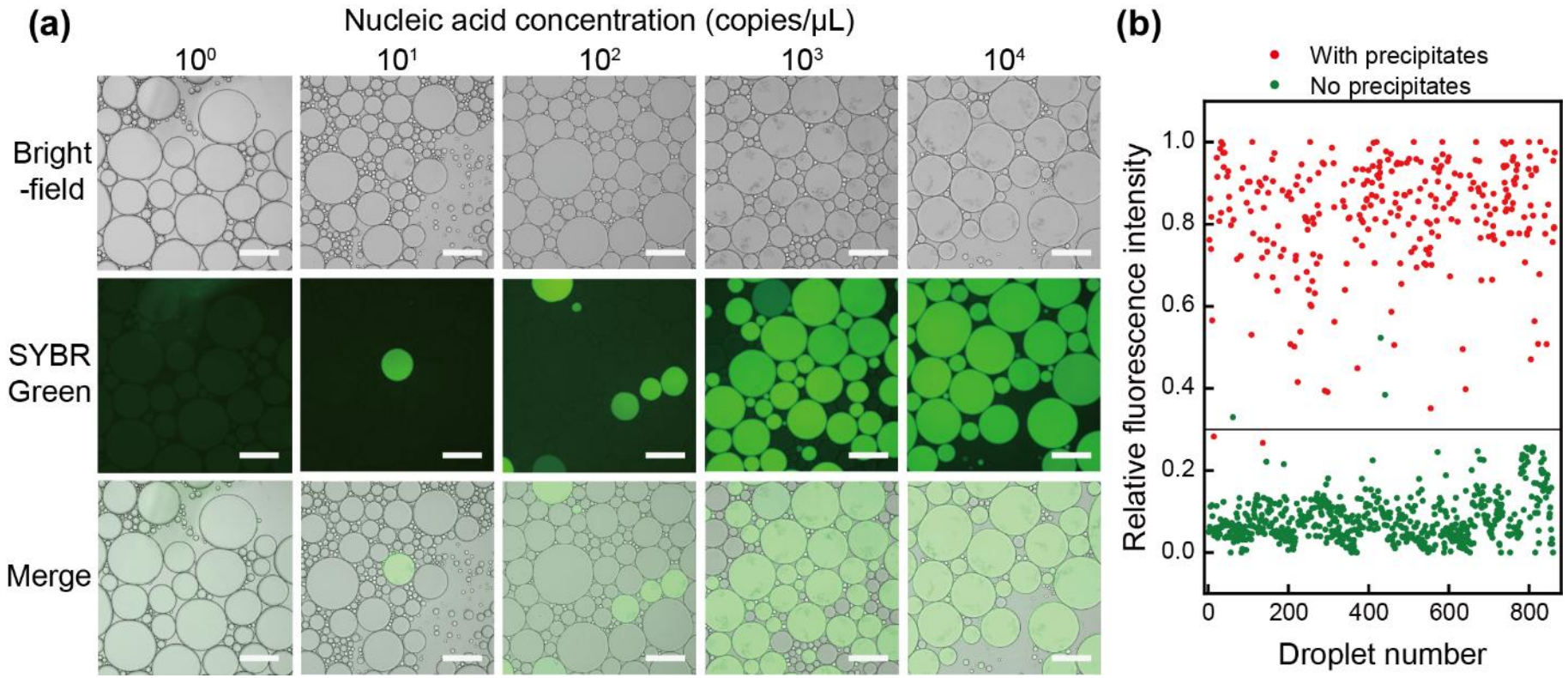


Work 2.1:
Non-uniform Droplets, Label-free Digital LAMP

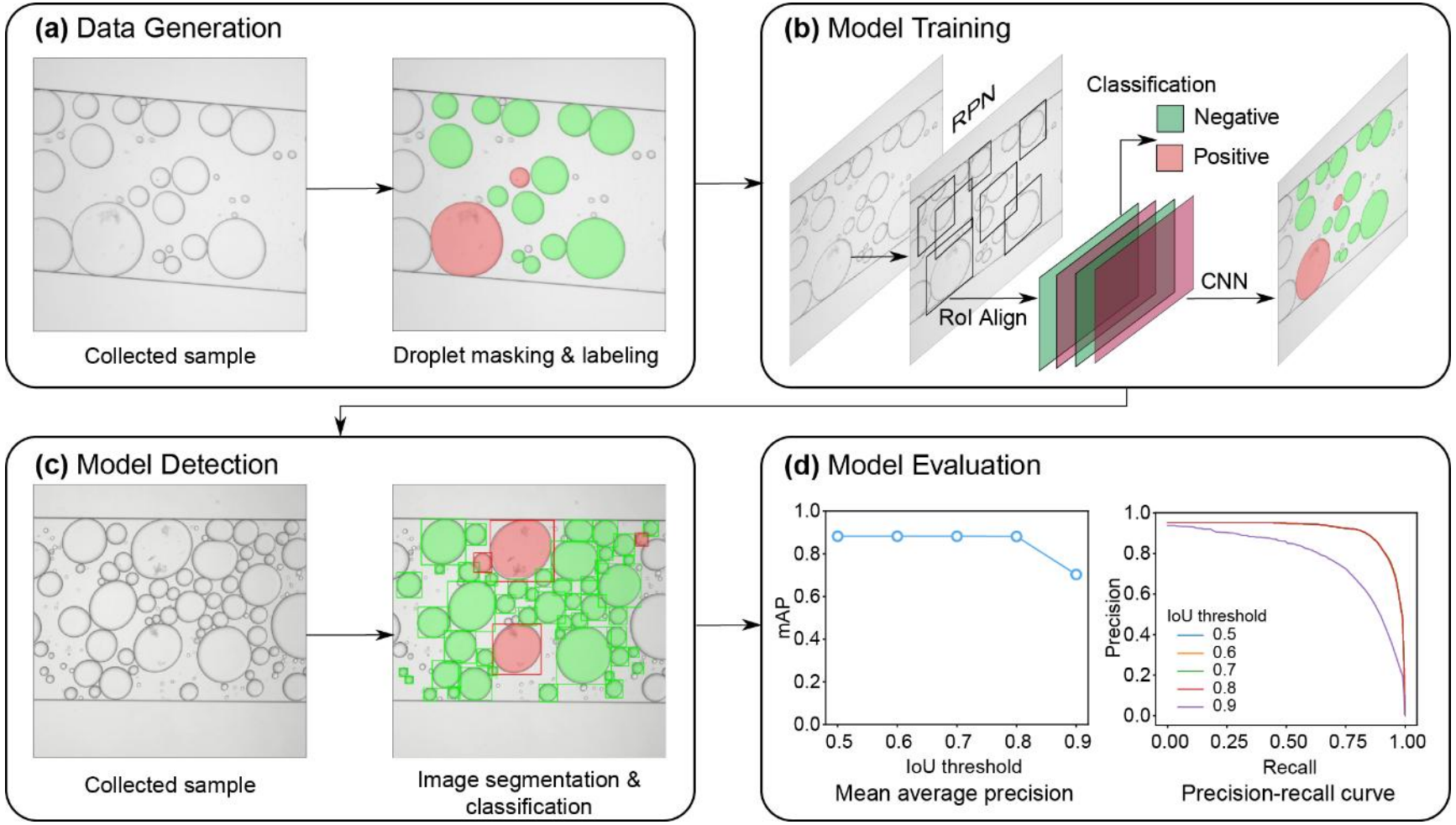


Chen, Ding, Li*, et al. *Adv. Sci.* 2022, 9, 9

Presence of precipitates is a reliable indicator of occupied emulsions.



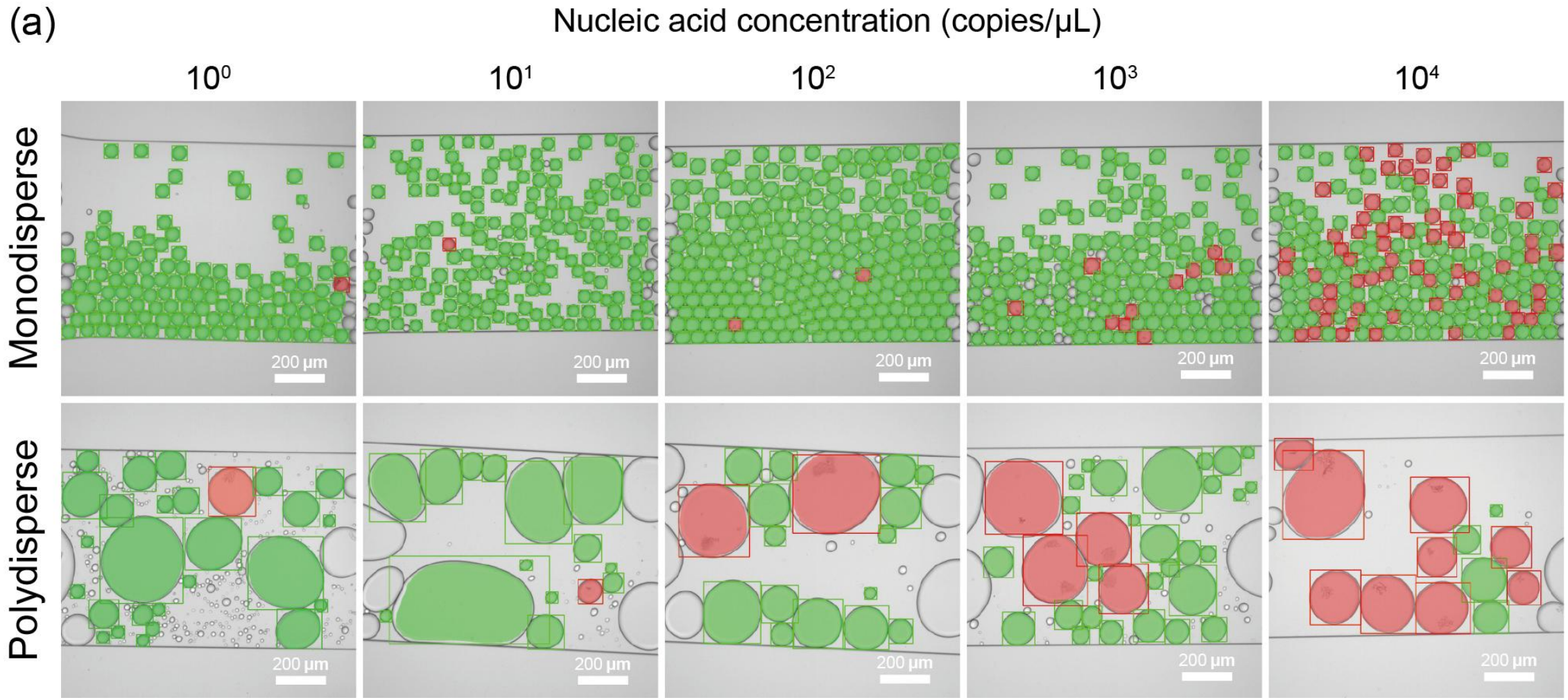
Deep learning image analysis



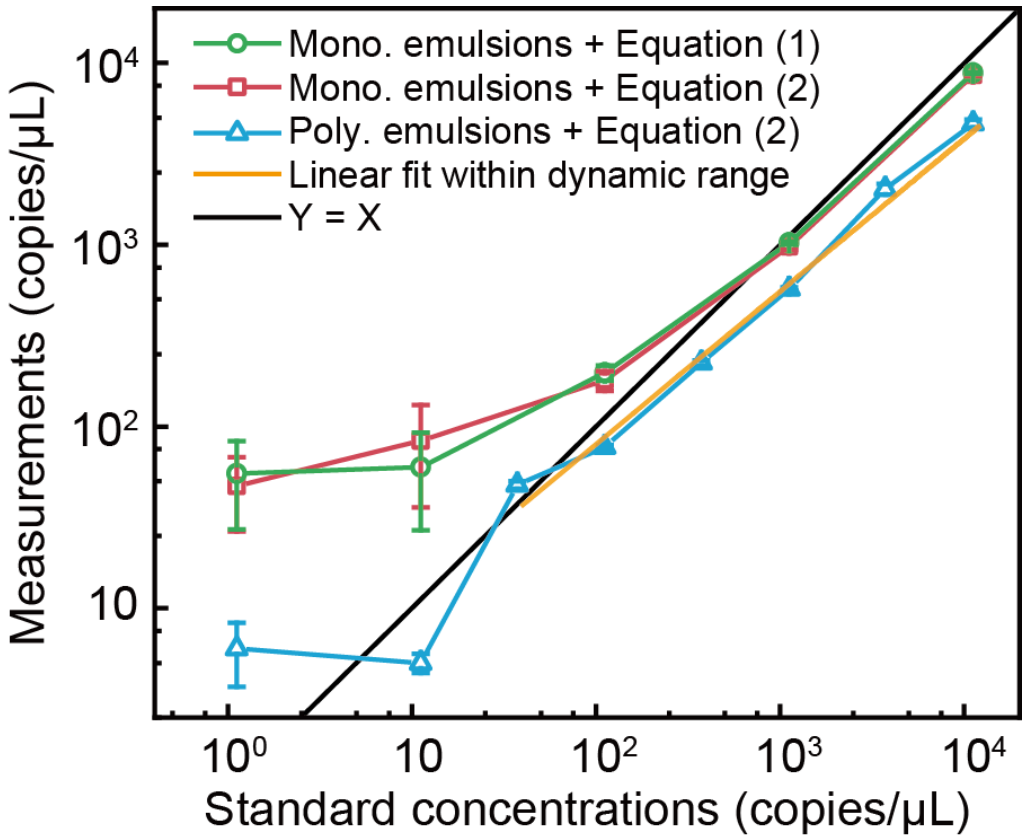
Validation

1. Image analysis results are accurate (in practice)
   - Use standard (monodisperse) digital nucleic acid test to validate
2. Nucleic acid quantification using polydisperse emulsion is valid
   - Use full version of the deep-dLAMP to validate

Representative micrographs showing the droplet segmentation and classification after the image analysis using monodisperse and polydisperse emulsions



Measurements of samples with different nucleic acid concentrations



$$C = -\ln P^{-} / V$$

$$\sum_{i=1}^{M} V_i^{-} + \sum_{j=1}^{N} V_j^{+} = \sum_{j=1}^{N} \frac{V_j^{+}}{[1 - \exp(-V_j^{+} \cdot C)]}$$

R^2=0.99

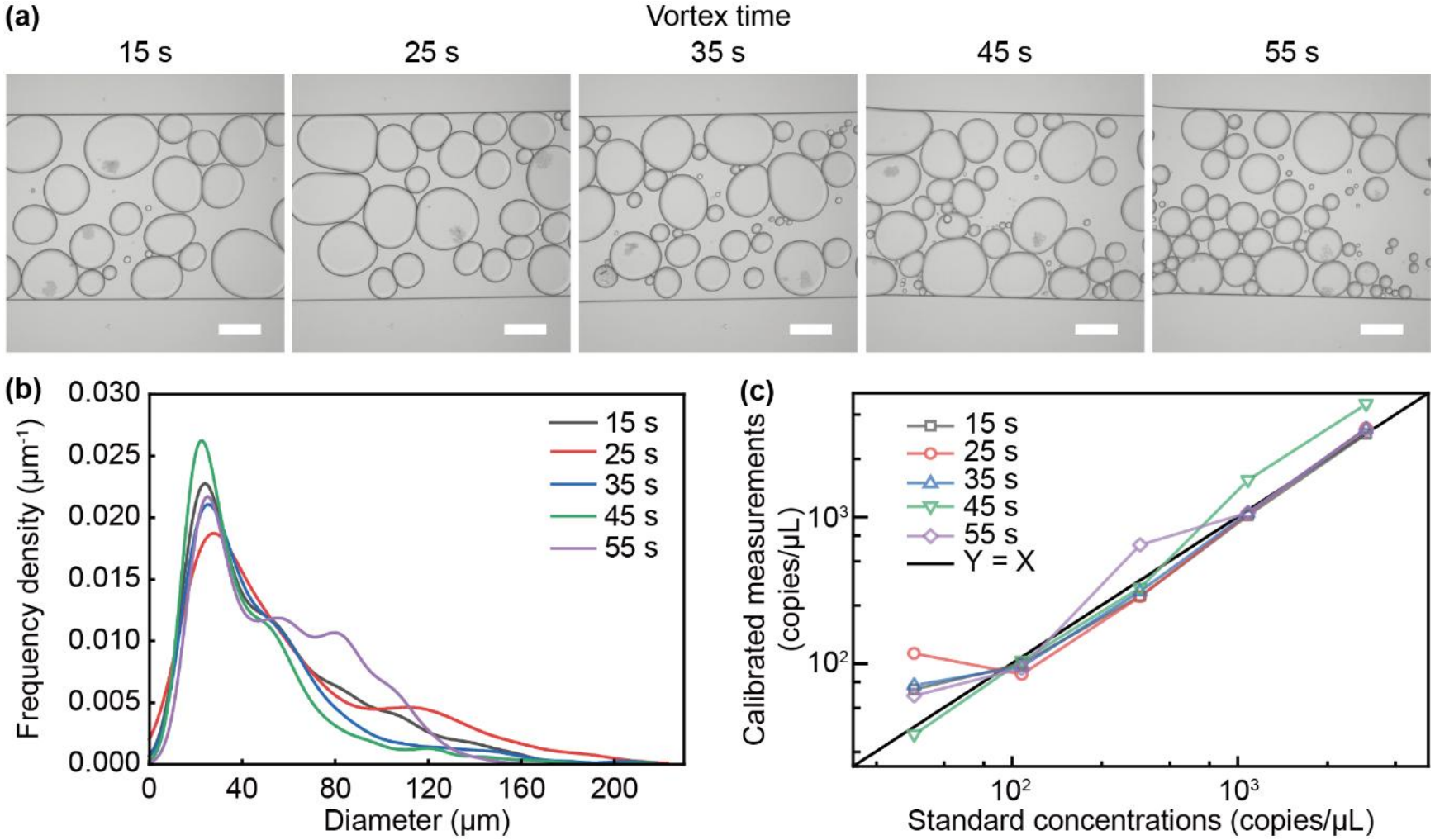
Limit of detection:

5.6 copies/μL

Dynamic range:

37.2-11,000 copies/μL

Effect of vortex time on deep-dLAMP measurements.



Effect of image qualities on deep-dLAMP measurements.