Молекулярная генетика и геномика

Лекция 7

SAGE

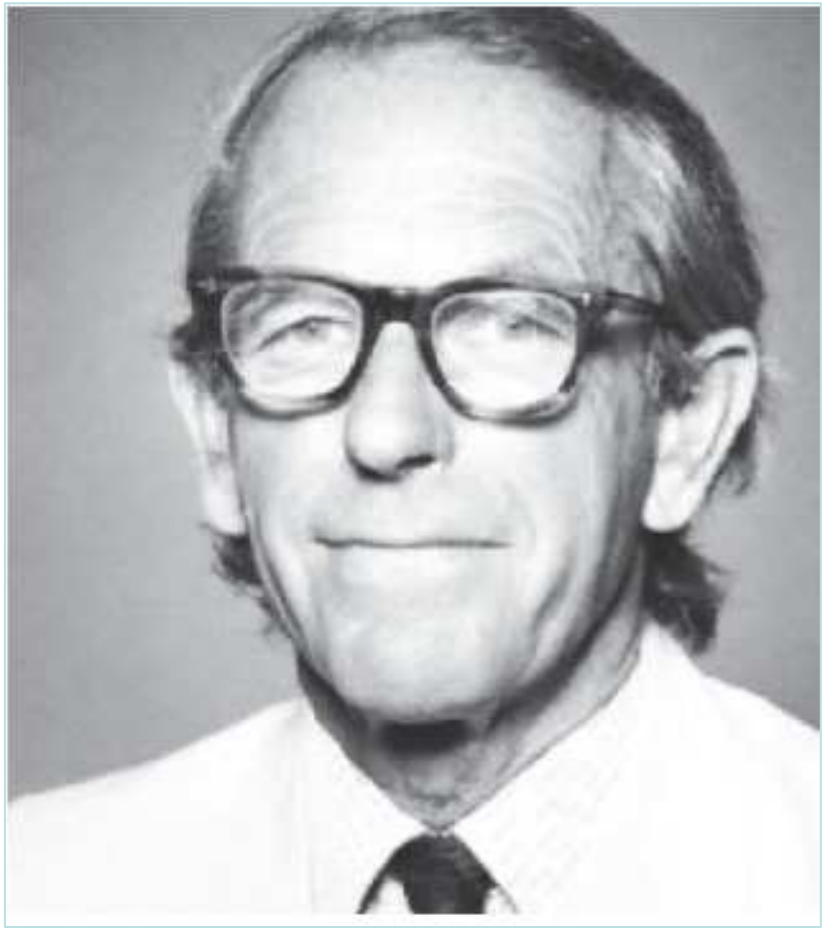
Современные платформы секвенирования

Сергей Лукьянов

Институт биоорганической химии имени академиков М.М. Шемякина и Ю.А. Овчинникова РАН

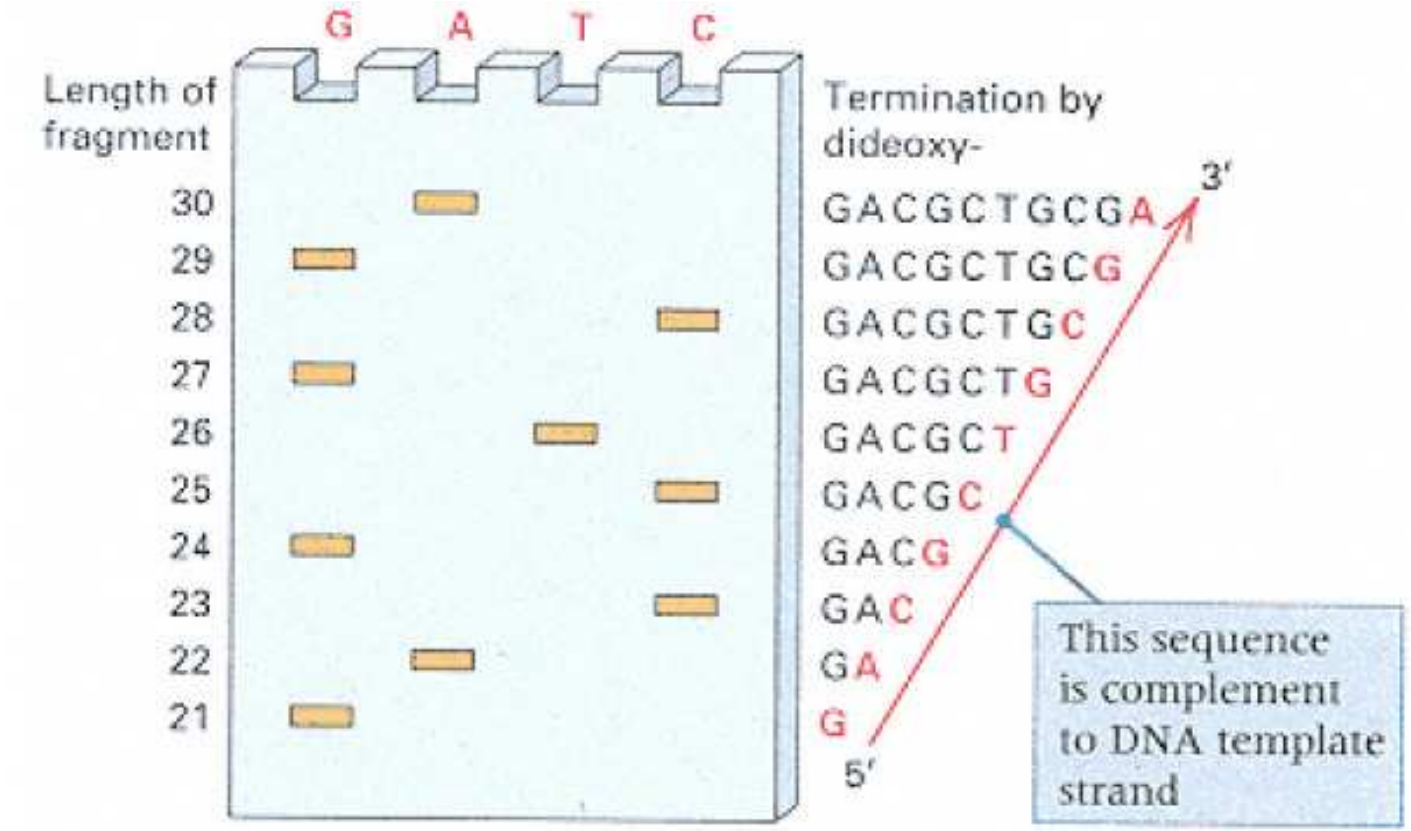
СОВРЕМЕННЫЕ ПЛАТФОРМЫ СЕКВЕНИРОВАНИЯ ДНК

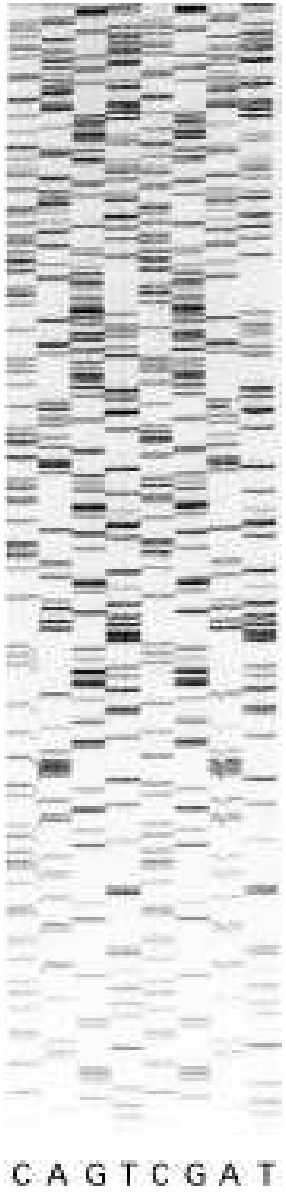
Секвенирование ДНК по методу Сэнгера



Frederick Sanger

Гель-электрофорез в ПААГ полученных фрагментов





Стратегия секвенирования больших геномов методом shotgun

Геномная ДНК

BAC-библиотека

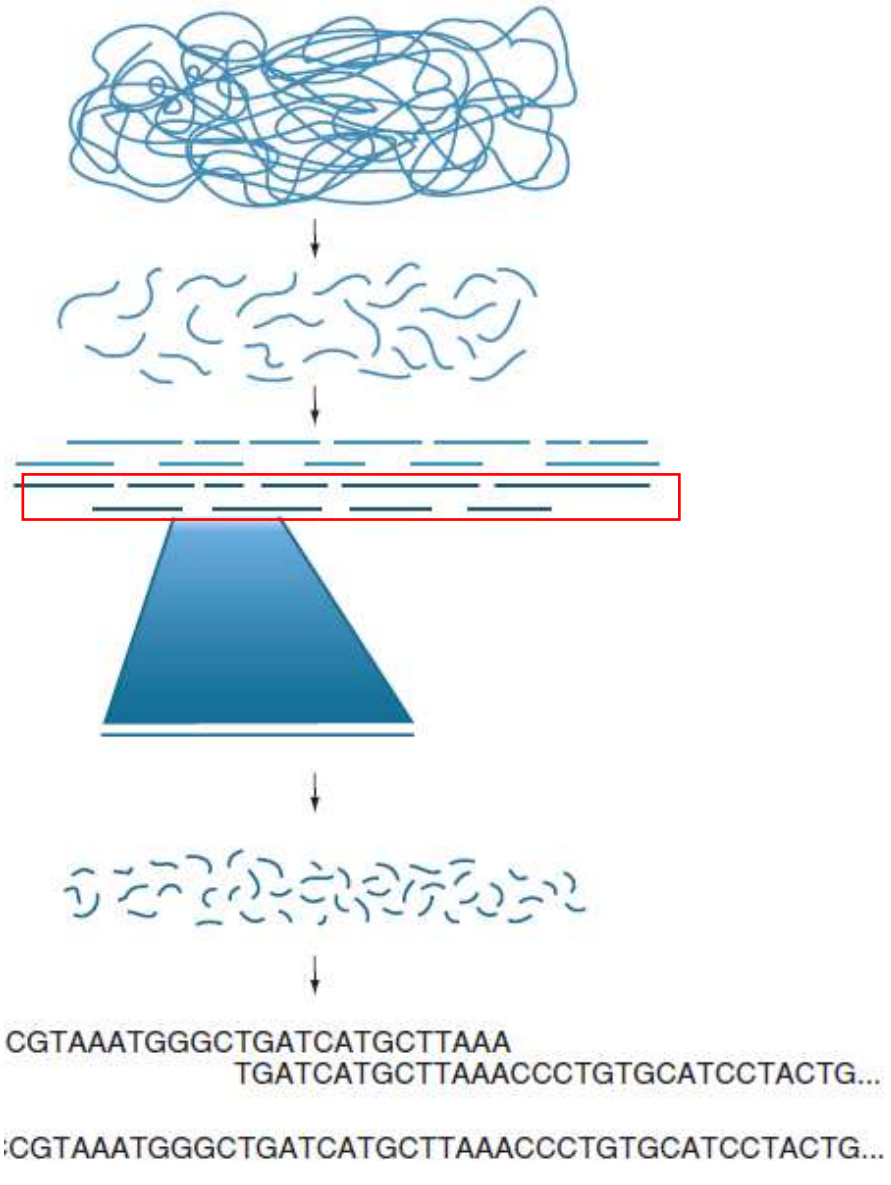
Упорядоченные клоны, объединенные в протяженные контиги: выбирается наиболее короткий путь секвенирования

BAC-клон, выбранный для секвенирования

Клоны, фрагменты BAC, секвенируемые методом shortgun

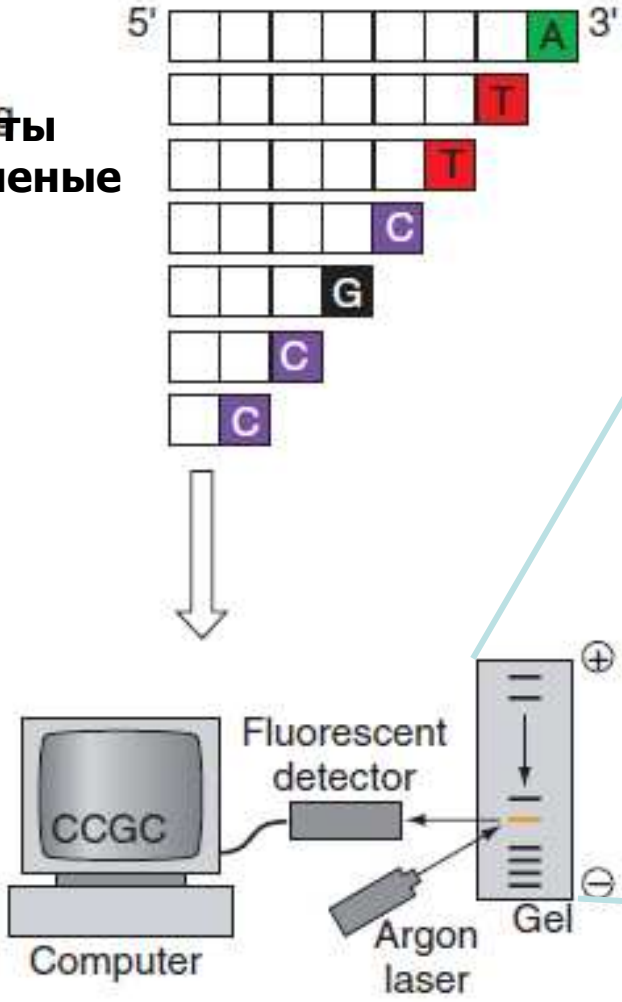
Объединение клонов по перекрыванию

Собранная последовательность



Секвенирование ДНК по методу Сэнгера на автоматическом секвенаторе

Фрагменты ДНК, меченые ddNTPs



Электрофорез в ПААГ

В одной пробирке присутствуют все 4 ddNTPs, меченые флуоресцентными красками

Центр по секвенированию геномов животных по методу Сэнгера

Одна машина может анализировать 96 образцов одновременно (96 капилляров), 750 п.н. за один прогон, 6 прогонов в день.
Итого:
одна машина может определять ~345 600 п.н. в один день

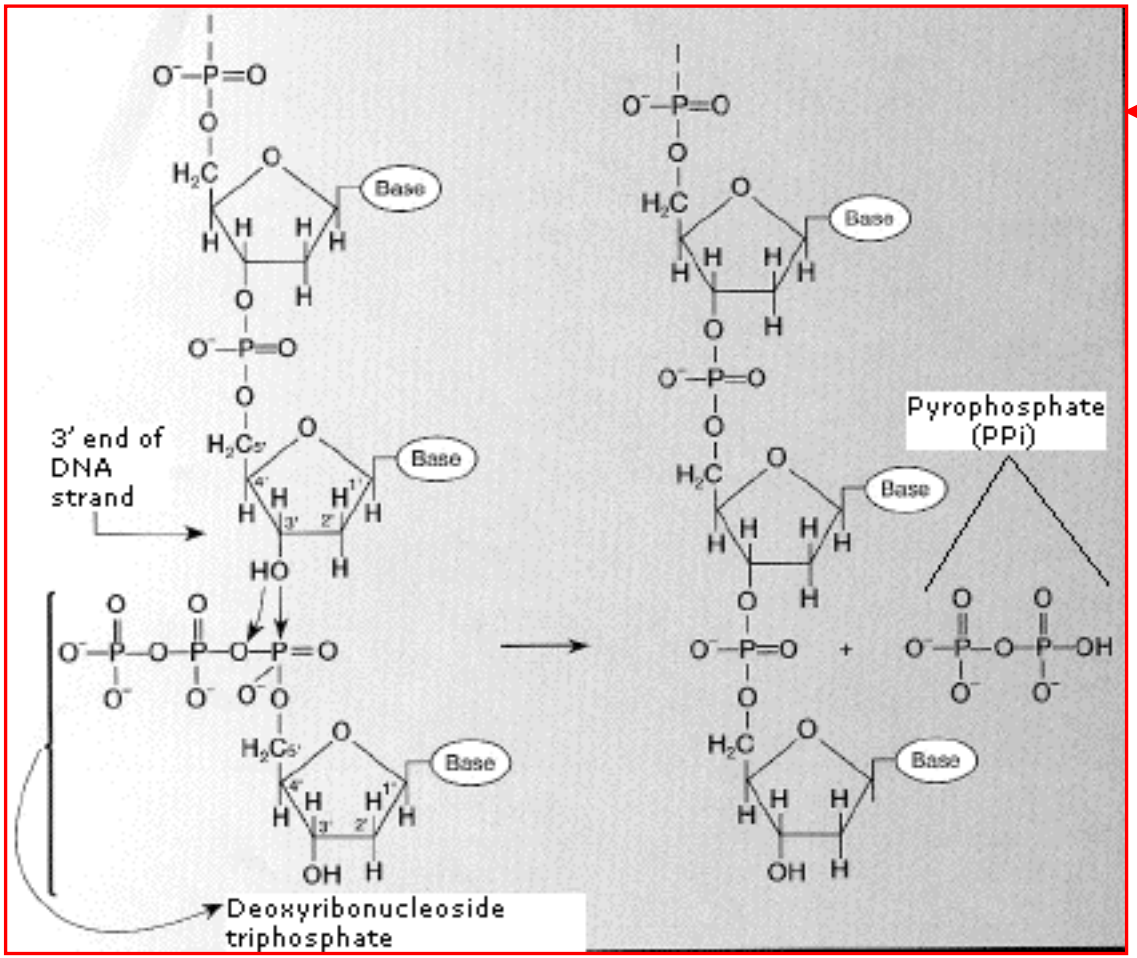
Секвенатор нового поколения

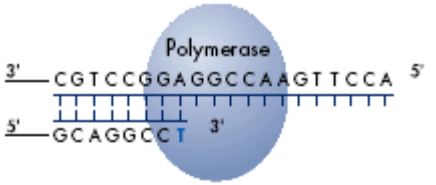
Applied Biosystem

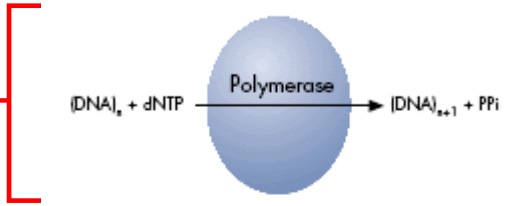
По производительности заменяет несколько сотен секвенаторов , основанных на методе Сэнгера

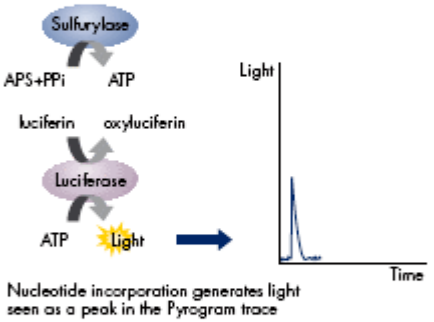
Principle of pyrosequencing

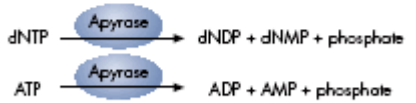
1. The incorporation of a nucleotide is accompanied by release of pyrophosphate (PPi).

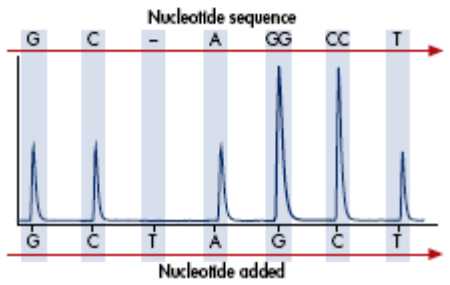






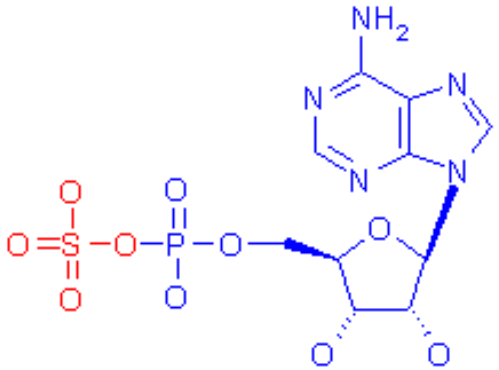




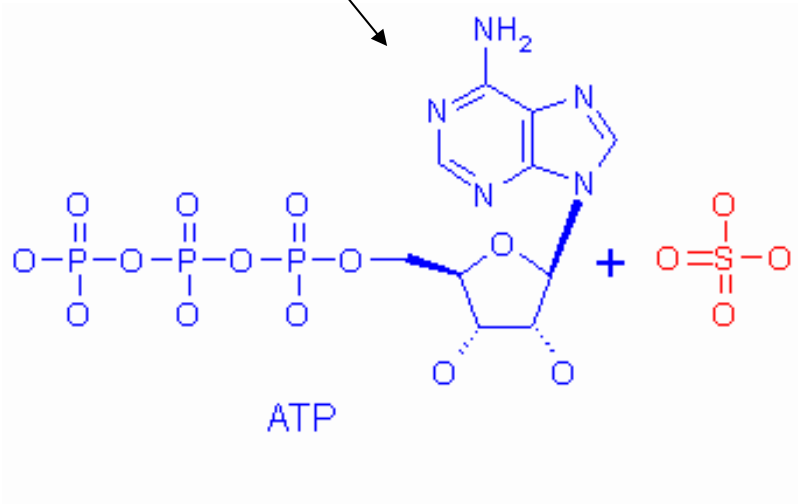


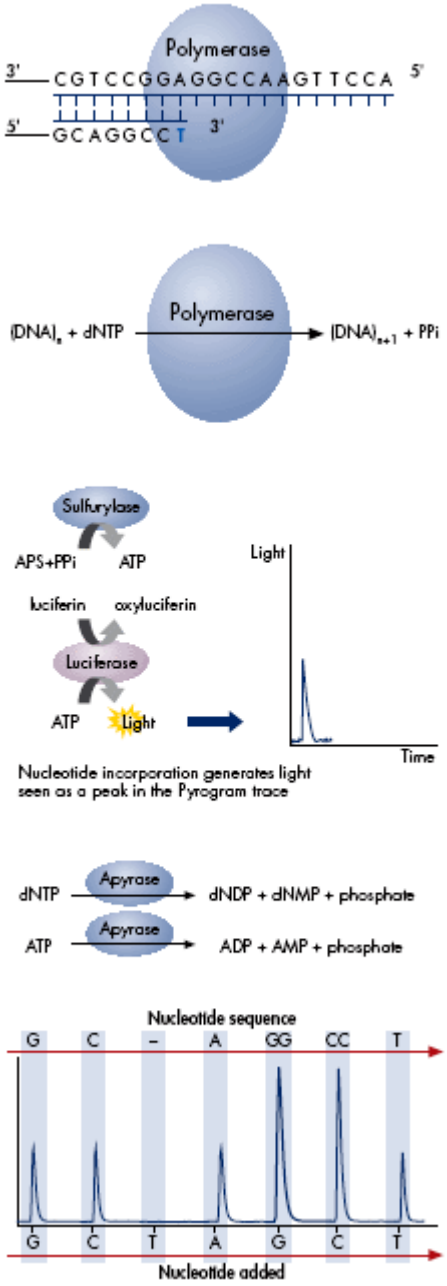
Principle of pyrosequencing

1. The incorporation of a nucleotide is accompanied by release of pyrophosphate (PPi).
2. The ATP sulfurylase quantitatively converts PPi to ATP.



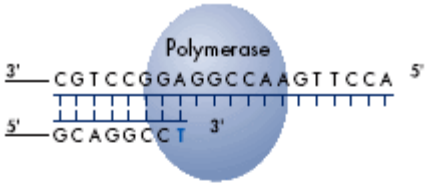
+ PPi

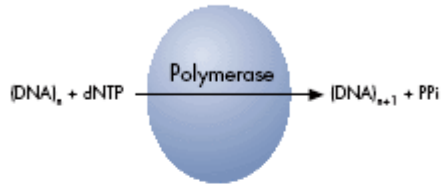


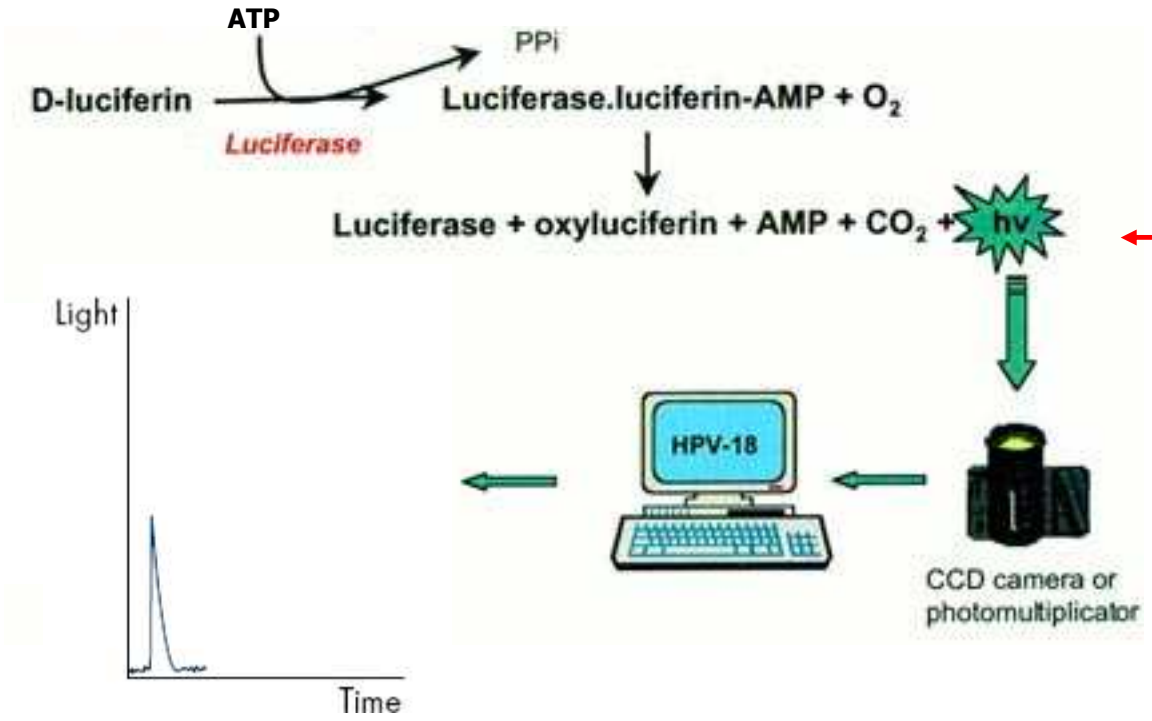


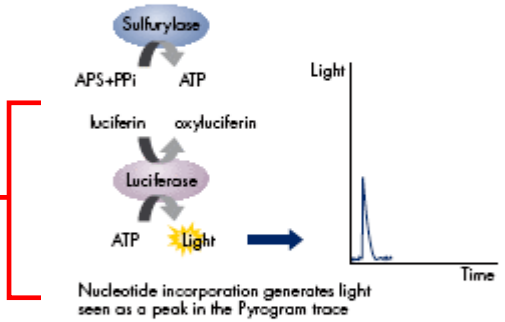
Principle of pyrosequencing

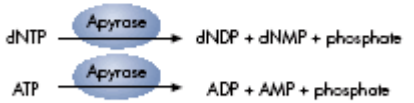
1. The incorporation of a nucleotide is accompanied by release of pyrophosphate (PPi).
2. The ATP sulfurylase quantitatively converts PPi to ATP.
3. The signal light produced by the luciferase-catalyzed reaction in presence of ATP is detected by a charge coupled device (CCD) camera and integrated as a peak in a Pyrogram.

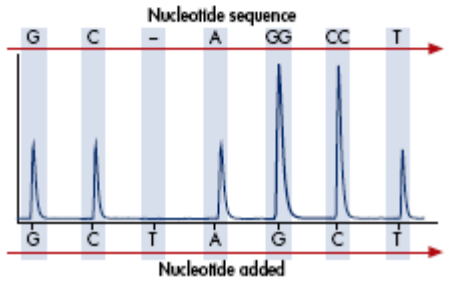






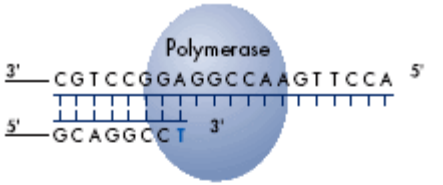


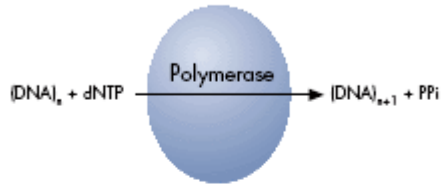


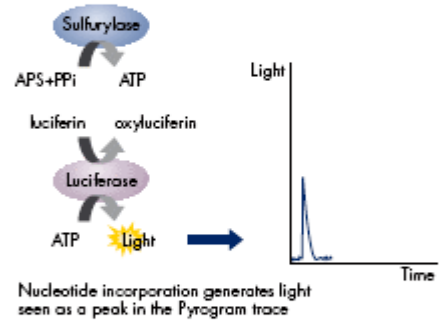


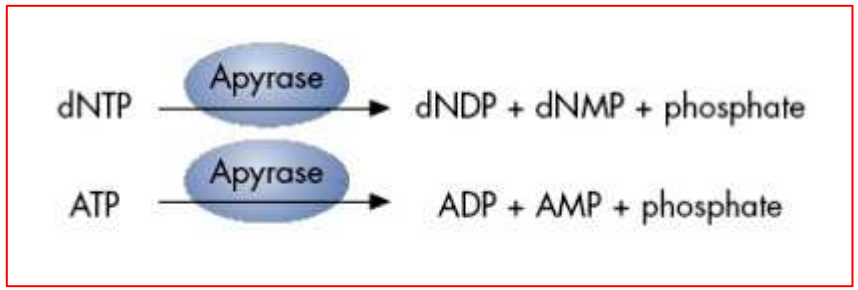
Principle of pyrosequencing

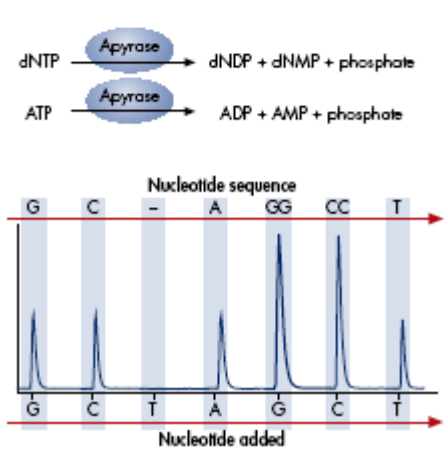
1. The incorporation of a nucleotide is accompanied by release of pyrophosphate (PPi).
2. The ATP sulfurylase quantitatively converts PPi to ATP.
3. The signal light produced by the luciferase-catalyzed reaction in presence of ATP is detected by a charge coupled device (CCD) camera and integrated as a peak in a Pyrogram.
4. The nucleotide degrading Apyrase enzyme continuously degrades ATP excess and unincorporated dNTPs.
5. The process continues with addition of the next dNTP and the nucleotide sequence of the complementary DNA strand is inferred from the signal peaks of the pyrogram.











In vitro cloning

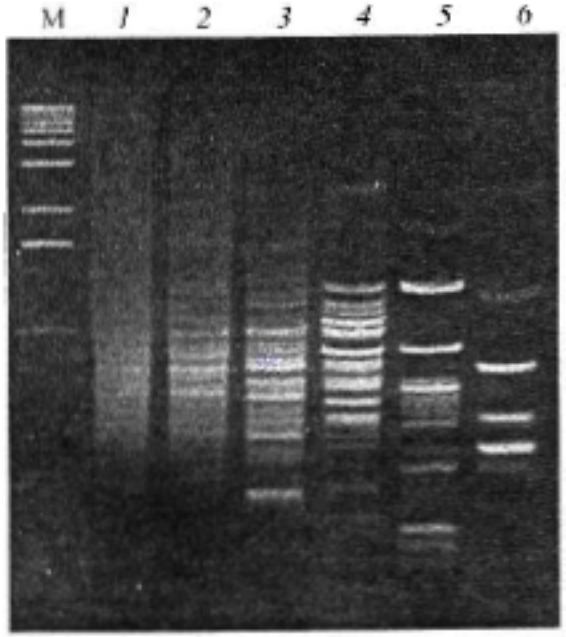


Рис. 3. Электрофорез в 2% агарозном геле продуктов ПЦР последовательных 3-кратных разведений неамплифицированной вычтенной кДНК. Дорожка М – маркер длин (1 kb ladder, Gibco BRL). Дорожки *1–6* – ПЦР с 100, 33, 11, 3.7, 1.2 и 0.4 нг исходной кДНК трейсера соответственно.

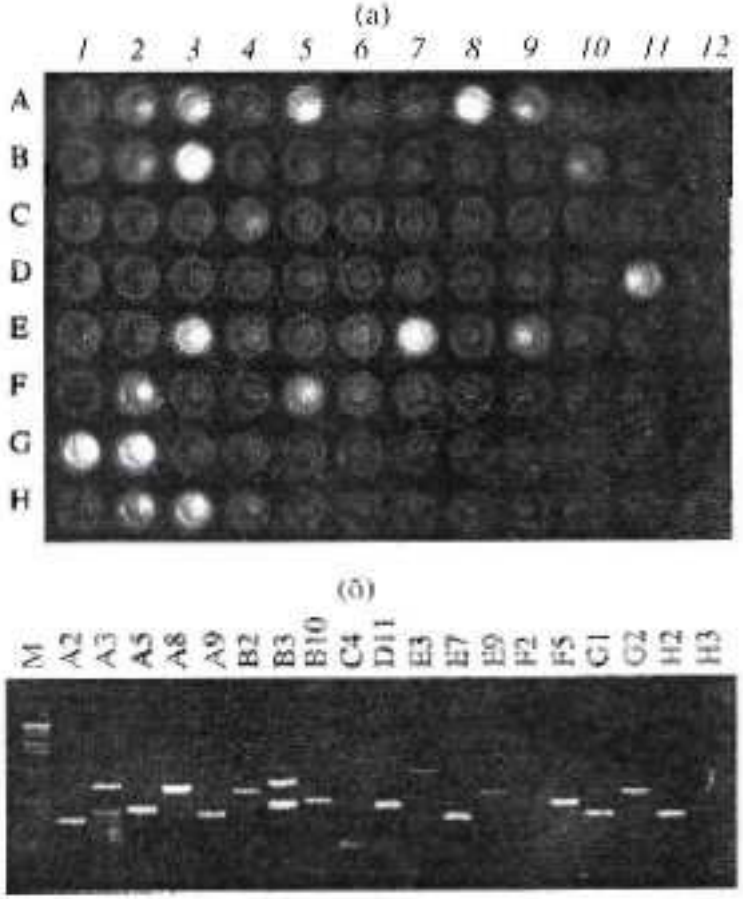
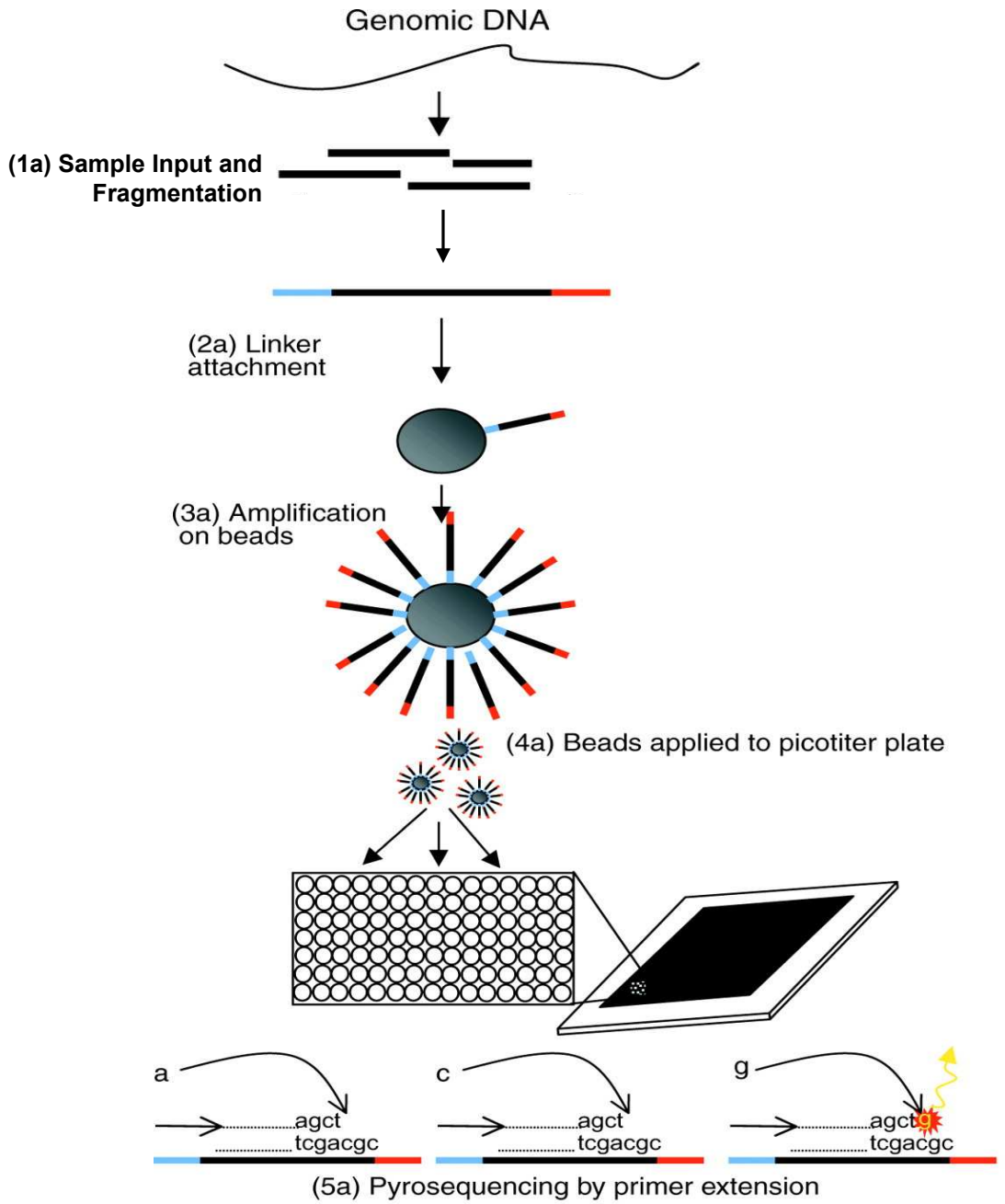
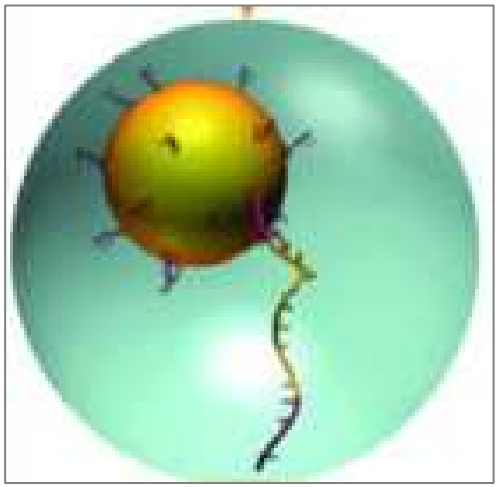


Рис. 4. Типичный результат клонирования in vitro в 96-луночном планшете в присутствии бромистого этидия: (а) – планшет в проходящем ультрафиолетовом свете после ПЦР; (б) – электрофорез в 2% агарозном геле продуктов ПЦР из окрашенных бромистым этидием проб (2 мкл из каждой пробы). Дорожка М – маркер длин (1 kb ladder, Gibco BRL). Над остальными дорожками указаны номера соответствующих лунок планшета.

454 technology
(Roche company)



454 technology

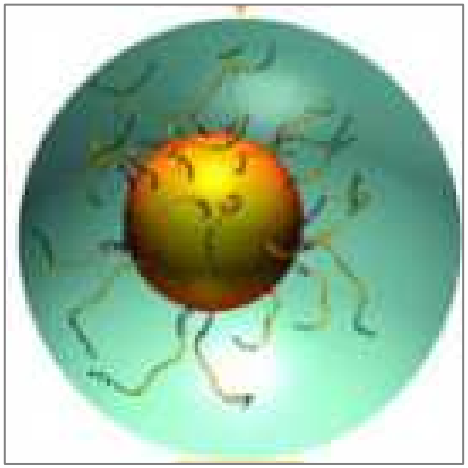


One Fragment = One Bead

- The single-stranded DNA library is immobilized onto specifically designed DNA Capture Beads.

- Each bead carries a unique single-stranded DNA library fragment.

- The bead-bound library is emulsified with amplification reagents in a water-in-oil mixture resulting in microreactors containing just one bead with one unique sample-library fragment.



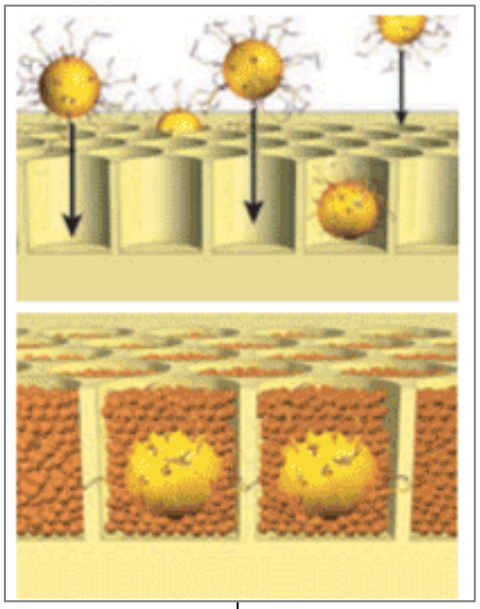
emPCR (Emulsion PCR) Amplification

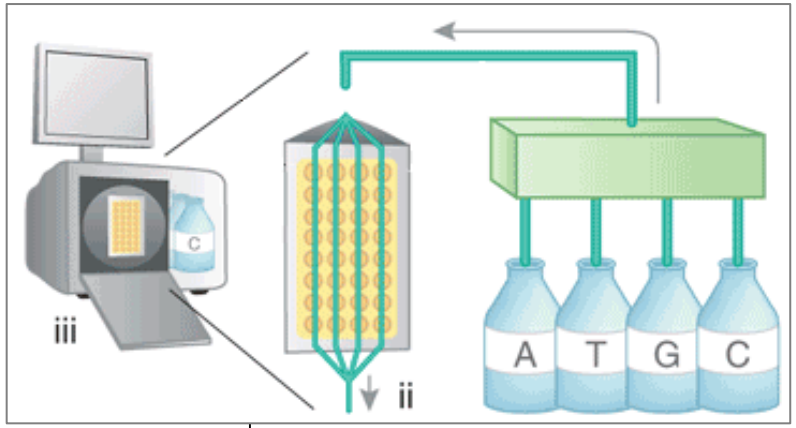
Each sample library fragment is amplified within its own microreactor, excluding competing or contaminating sequences.

Amplification of the entire fragment collection is done in parallel; for each fragment, this results in a copy number of several million per bead.

Subsequently, the emulsion PCR is broken while the amplified fragments remain bound to their specific beads.

454 technology

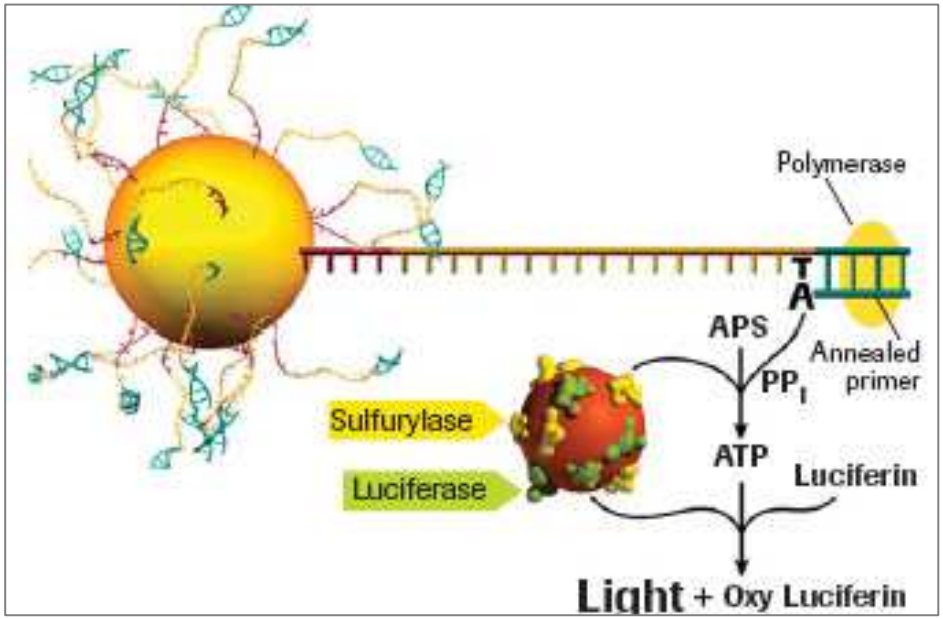




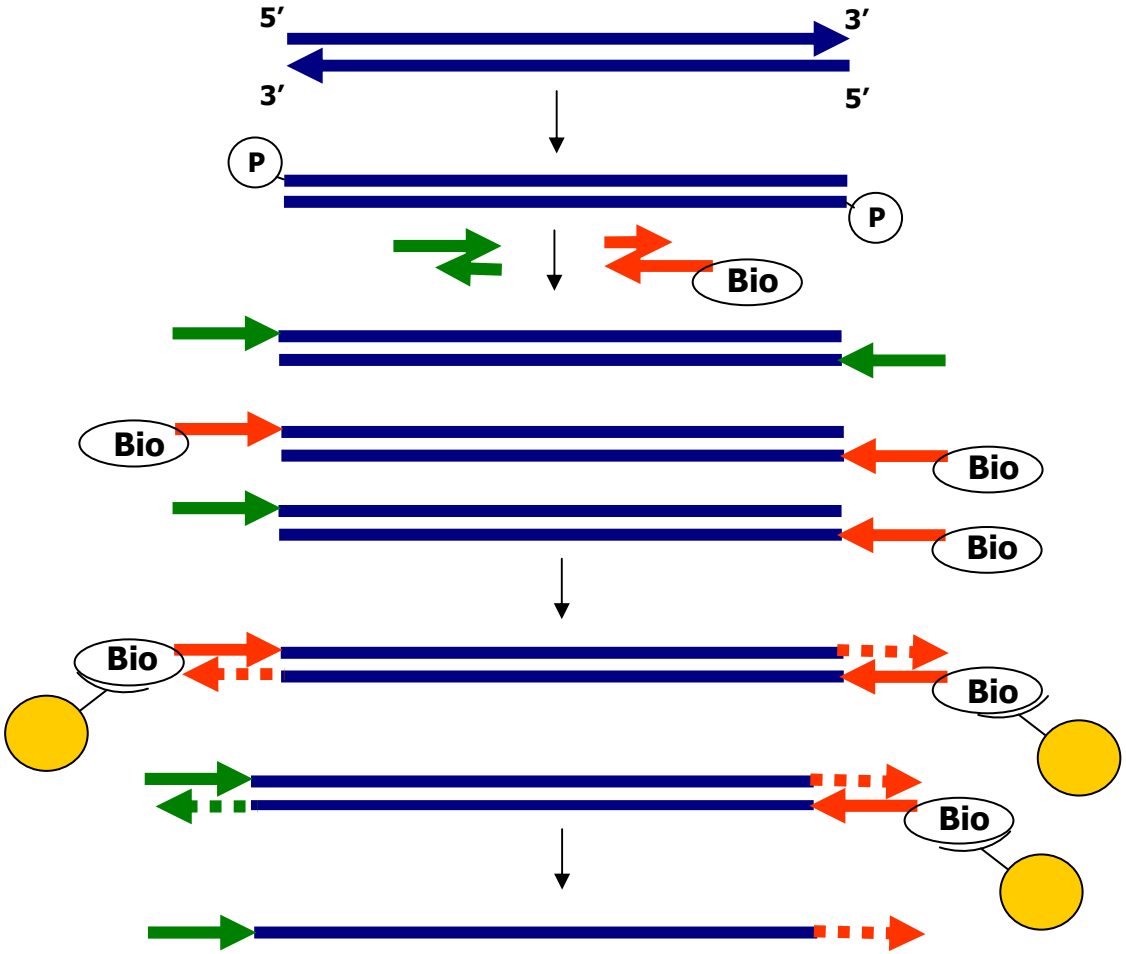
One Bead = One Read

The clonally amplified fragments are enriched and loaded onto a PicoTiterPlate device for sequencing. The diameter of the PicoTiterPlate wells allows for only one bead per well.

After addition of sequencing enzymes, the fluidics subsystem of the Genome Sequencer FLX Instrument flows individual nucleotides in a fixed order across the hundreds of thousands of wells containing one bead each. Addition of one (or more) nucleotide(s) complementary to the template strand results in a chemiluminescent signal recorded by the CCD camera.



454 library preparation process



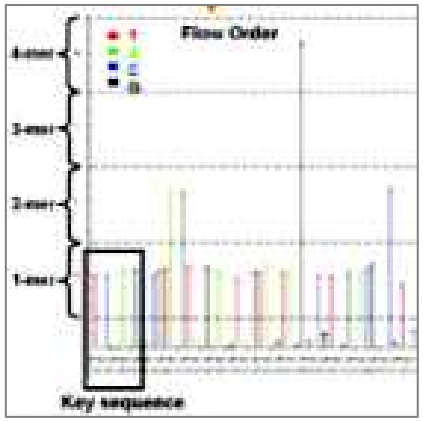
Double-stranded DNA molecules (blue) are made blunt-ended by T4 DNA polymerase, 5′-phosphorylated (stars) by T4 polynucleotide kinase

and ligated to one strand of nonphosphorylated double-stranded adaptors A (green) and B (red).

Ligation products carrying the biotinylated B adaptor are captured on Streptavidin beads (yellow), and the strand-displacing Bst DNA polymerase is used to extend the nicks between adaptors and template.

The DNA strands are then denatured, releasing the A-to-B strands, which are isolated and used as templates for emulsion PCR.

454 technology

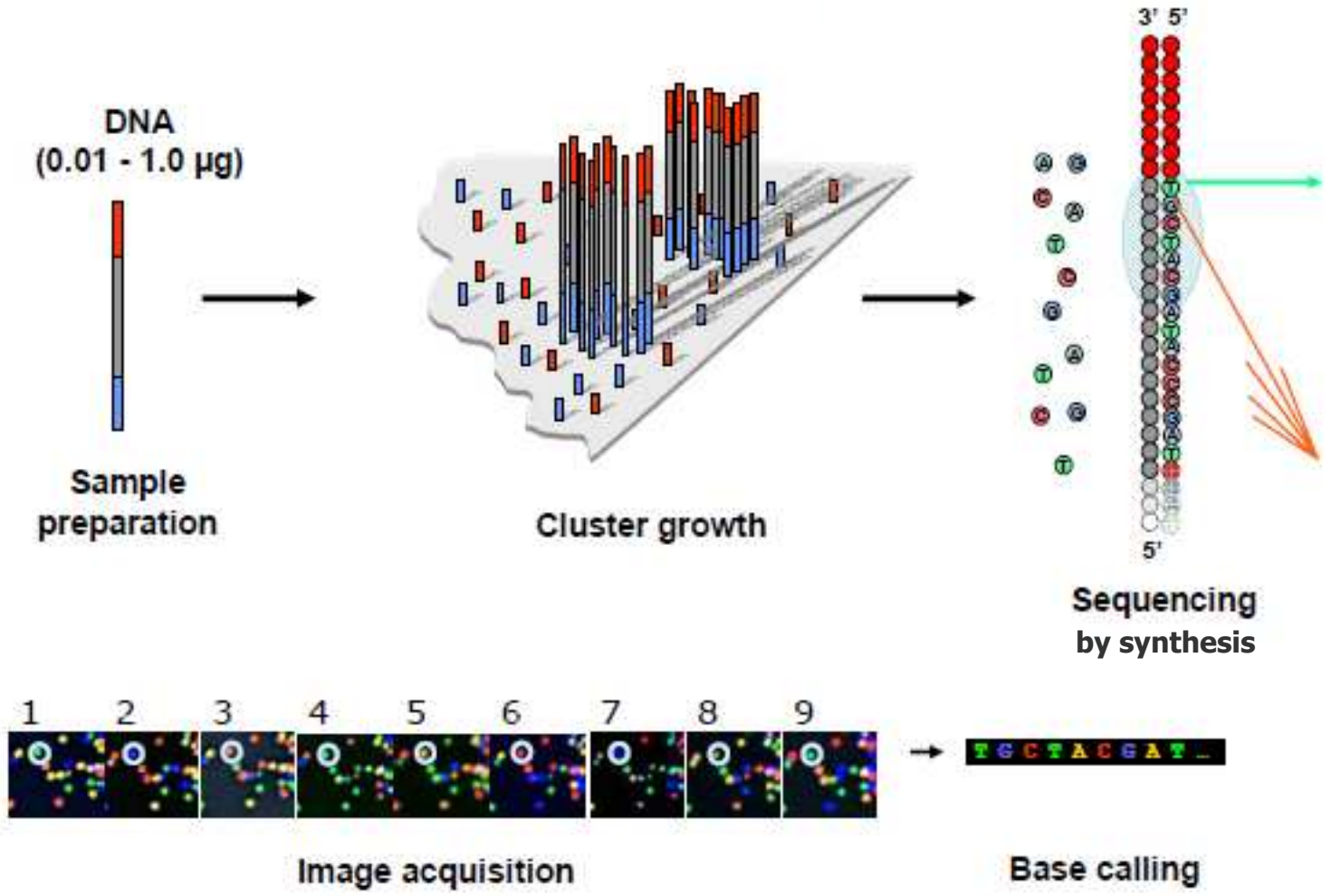


Data Analysis

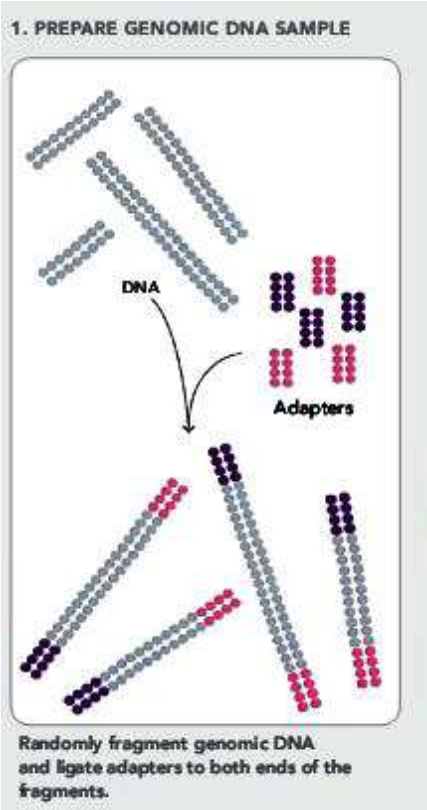
The combination of signal intensity and positional information generated across the PicoTiterPlate device allows the software to determine the sequence of more than 1,000,000 individual reads per 10-hour instrument run simultaneously.

For sequencing-data analysis, three different bioinformatics tools are available supporting the following applications: de novo assembly up to 400 megabases; resequencing genomes of any size; and amplicon variant detection by comparison with a known reference sequence.

Solexa sequencing (Illumina Inc)



Solexa sequencing

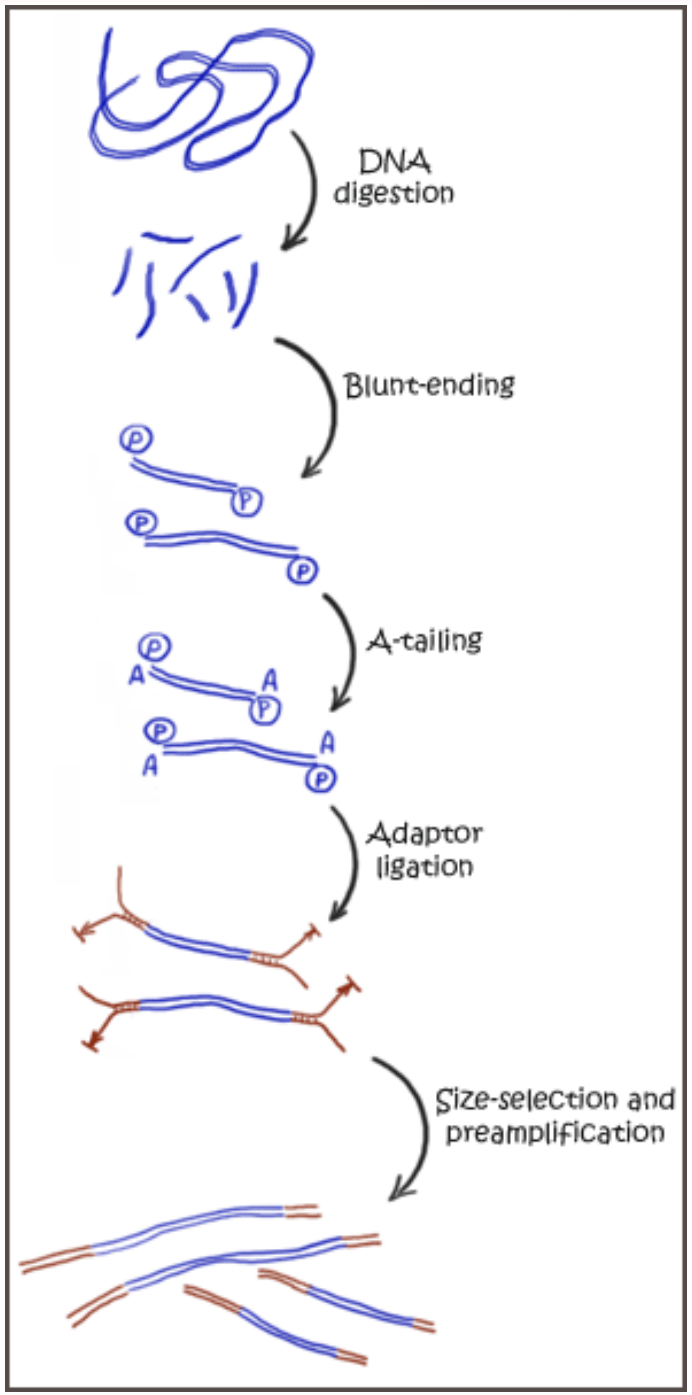


Step 1: Sample Preparation

-The DNA sample is sheared to appropriate size (average ~800bp) using a compressed air device known as a nebulizer.

-The ends of the DNA are polished, and two unique adapters are ligated to the fragments.

-Ligated fragments of the size range of 150-200bp are isolated via gel extraction and amplified using limited cycles of PCR.



Solexa sequencing

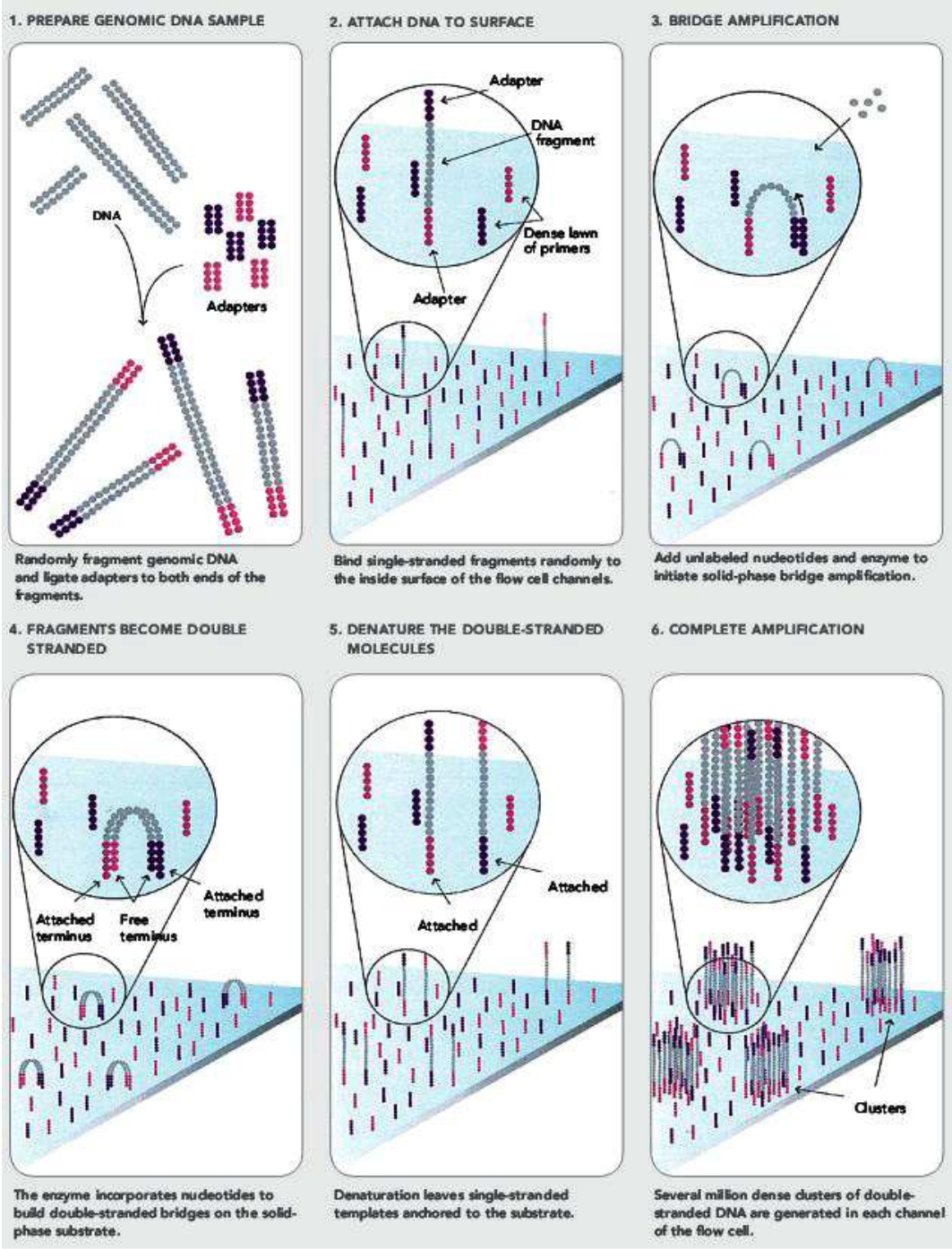
Steps 2-6: Cluster Generation by Bridge Amplification

- The flow cell surface is coated with single stranded oligonucleotides that correspond to the sequences of the adapters ligated during the sample preparation stage.

- Single-stranded, adapter-ligated fragments are bound to the surface of the flow cell exposed to reagents for polyermase-based extension. Priming occurs as the free/distal end of a ligated fragment "bridges" to a complementary oligo on the surface.

- Repeated denaturation and extension results in localized amplification of single molecules in millions of unique locations across the flow cell surface.

-This process occurs in what is referred to as Illumina's "cluster station", an automated flow cell processor.



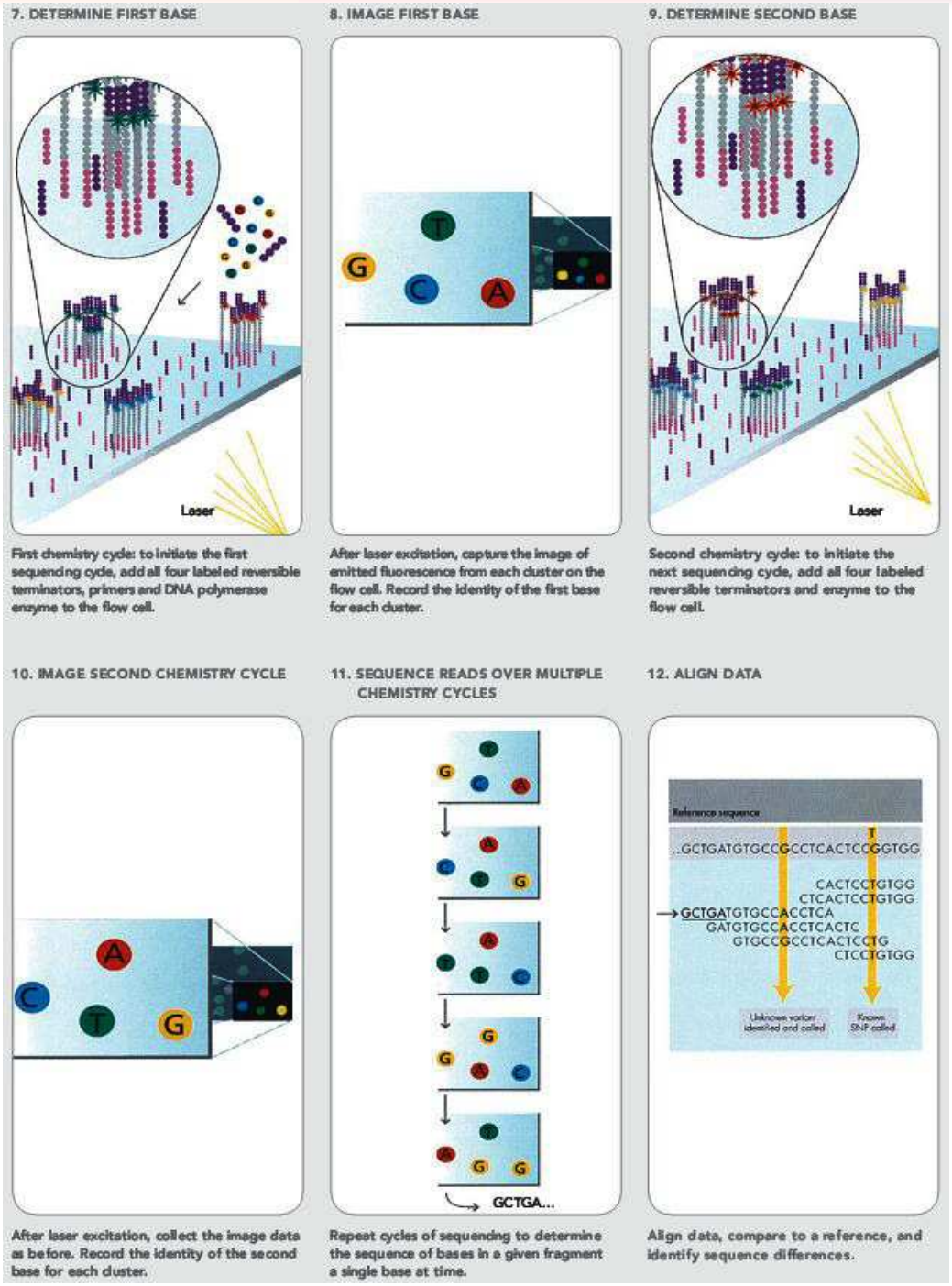
Solexa sequencing

Steps 7-12: Sequencing by Synthesis

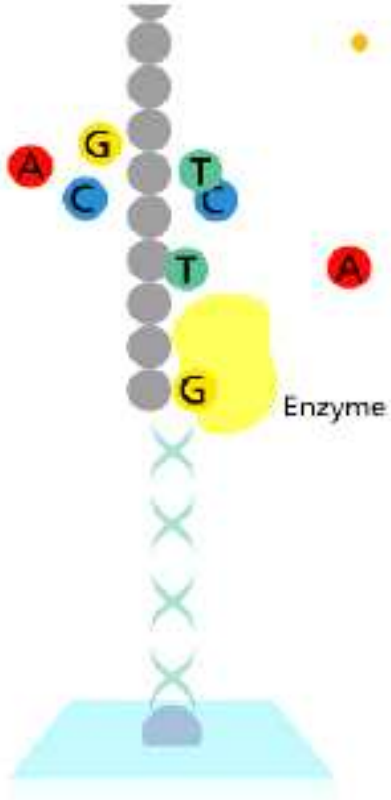
-A flow cell containing millions of unique clusters is now loaded into the 1G sequencer for automated cycles of extension and imaging.

- The first cycle of sequencing consists first of the incorporation of a single fluorescent nucleotide, followed by high resolution imaging of the entire flow cell. These images represent the data collected for the first base. Any signal above background identifies the physical location of a cluster (or polony), and the fluorescent emission identifies which of the four bases was incorporated at that position.

- This cycle is repeated, one base at a time, generating a series of images each representing a single base extension at a specific cluster. Base calls are derived with an algorithm that identifies the emission color over time. At this time reports of useful Illumina reads range from 26-50 bases.

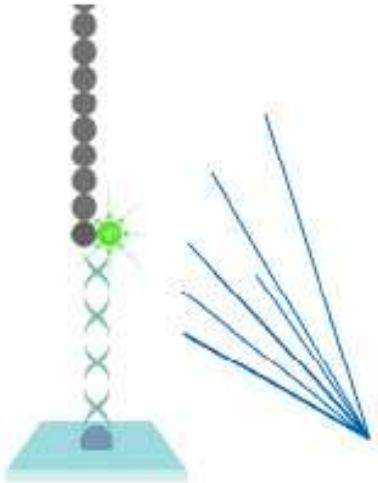


Solexa sequencing



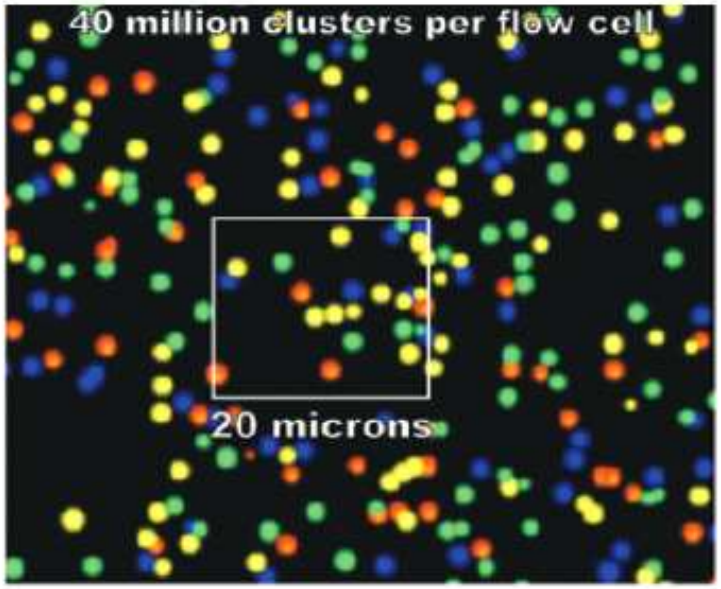
- Proprietary enzymes ensure nucleotide-specific incorporation

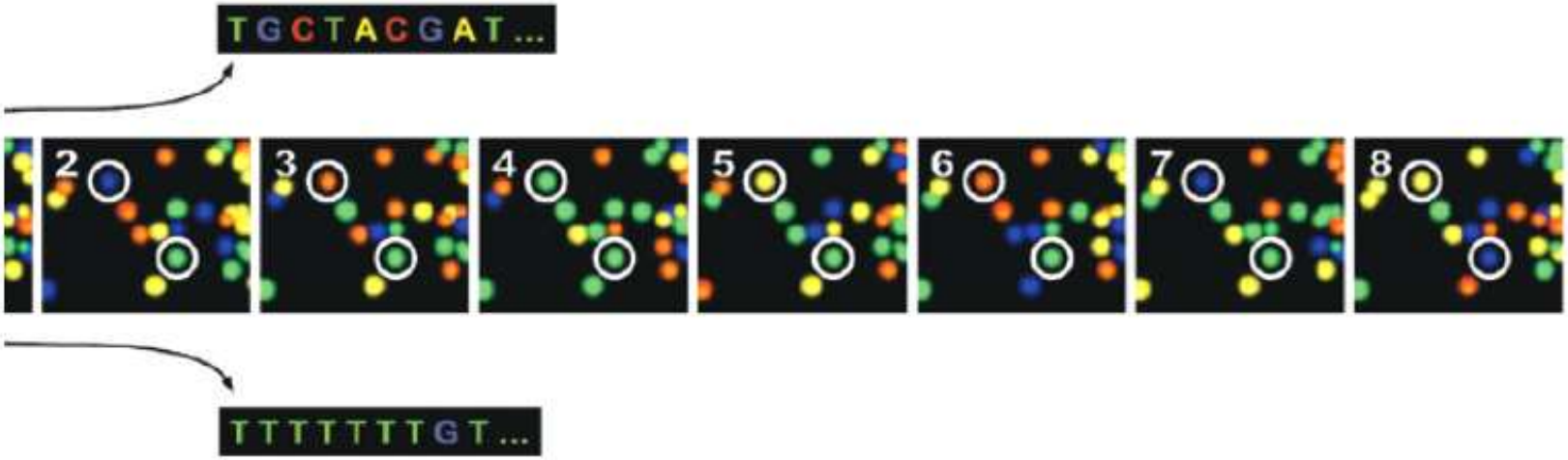
Cycle 1



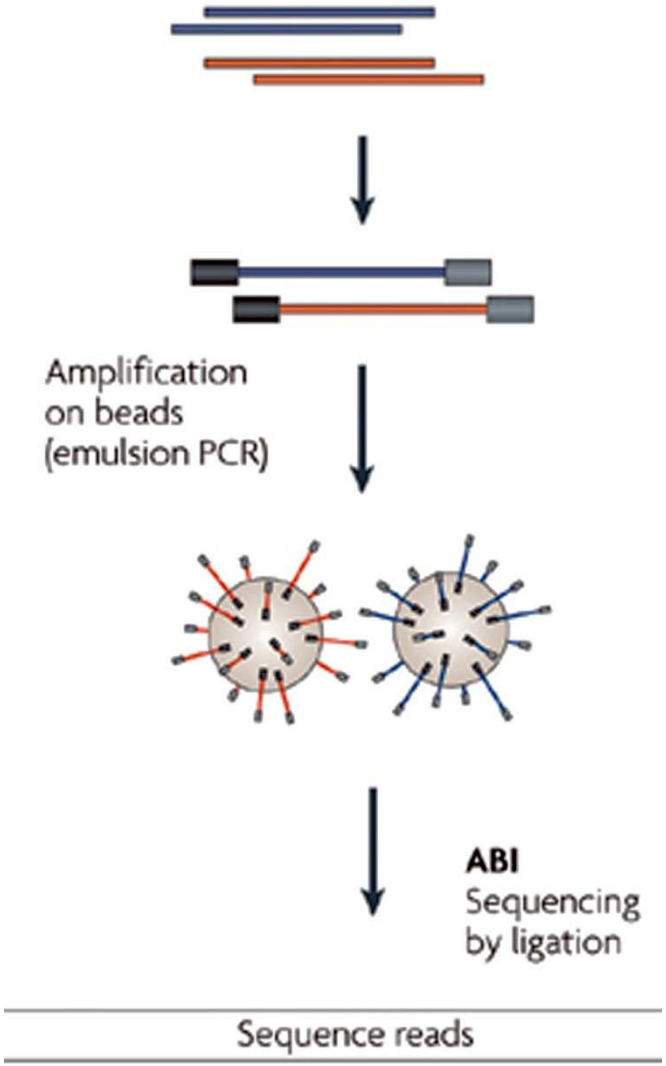
- Incorporated nucleotides detected by laser-excited fluorescence
- Detected by a CCD camera that rapidly scans the entire array
- Fluorescence is then removed

Solexa sequencing



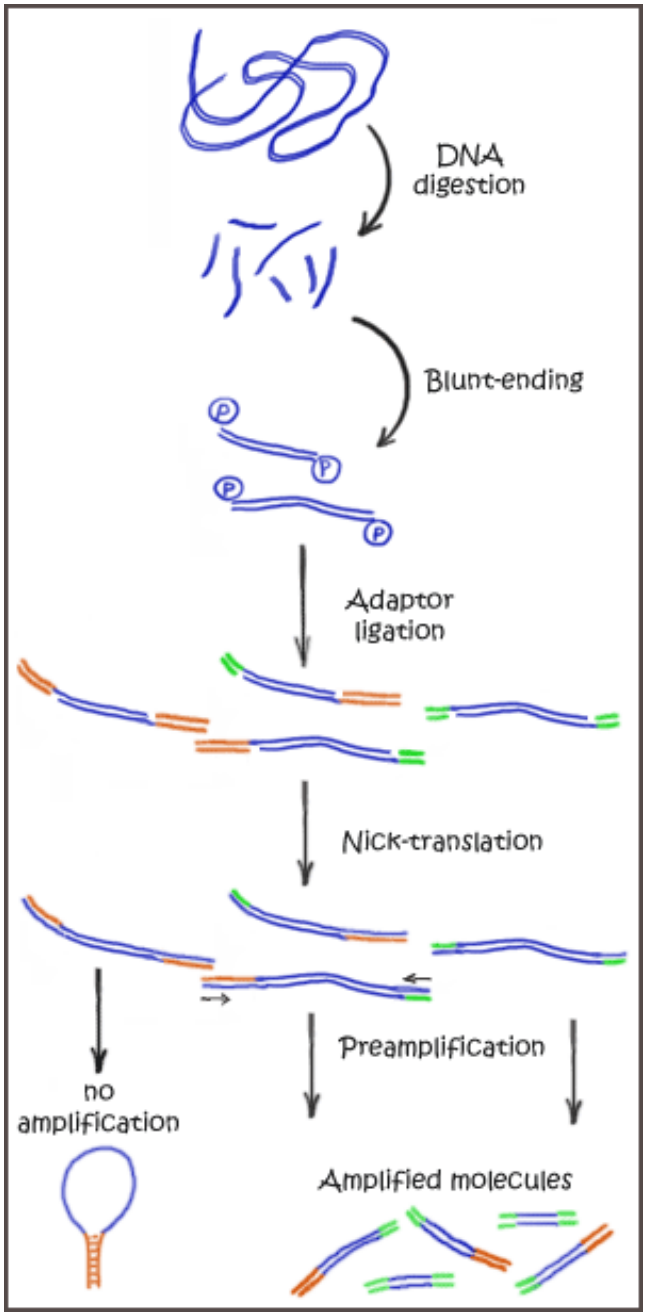


Solid sequencing (Applied Biosystems)



Solid sequencing

Step 1: Sample Preparation
The DNA sample is sheared to appropriate size (average ~800bp) using a compressed air device known as a nebulizer.
The ends of the DNA are polished, and two unique adapters are ligated to the fragments.

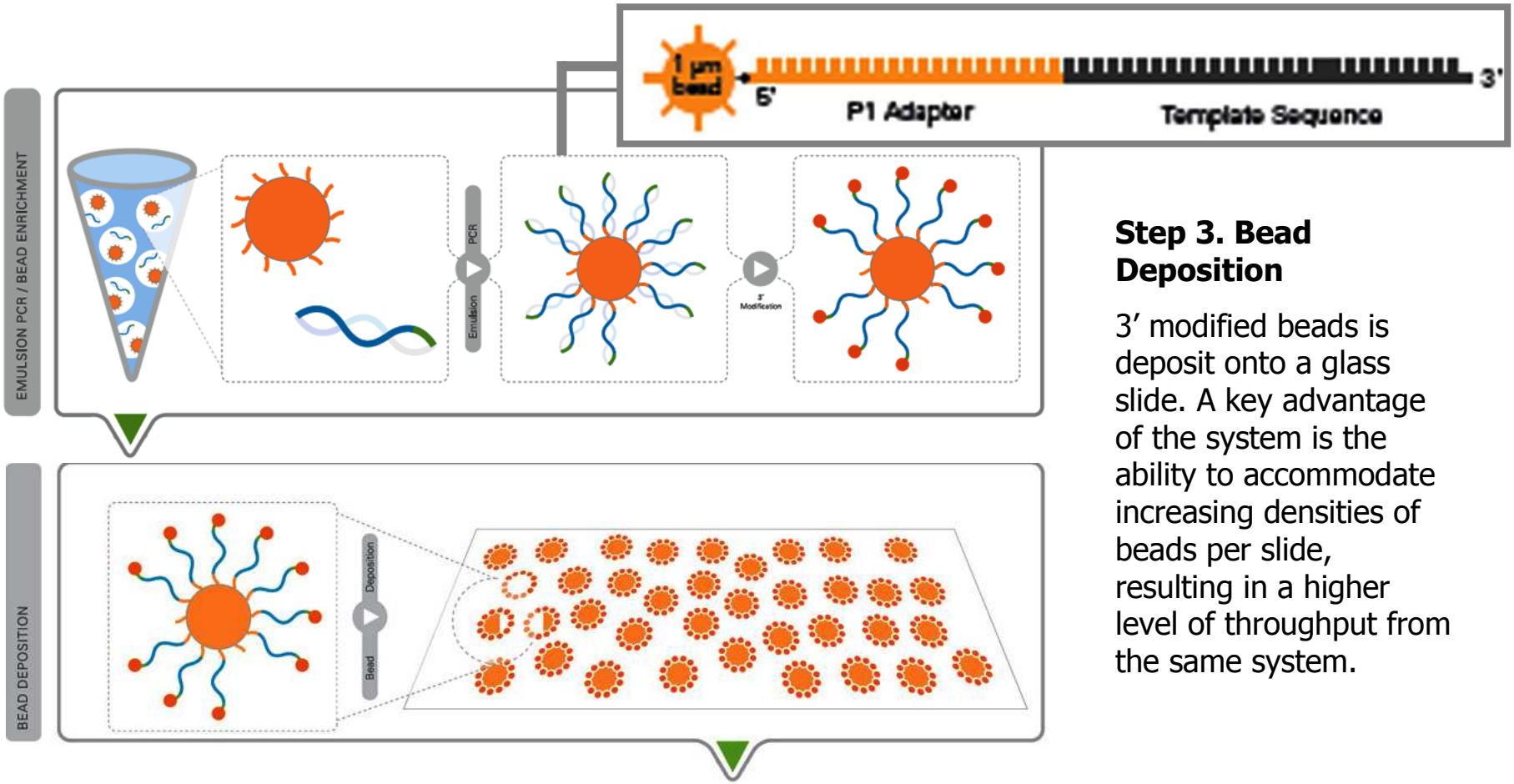


Solid sequencing

Step 2. Emulsion PCR

Clonal bead populations in microreactors containing template, PCR reaction components, beads, and primers are prepared.

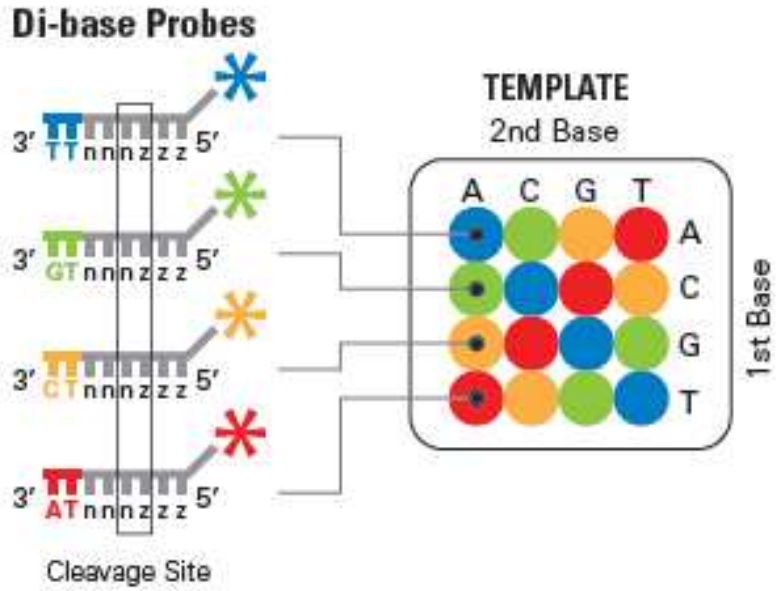
After PCR, the templates are denatured and bead population is enriched with beads with extended templates. The template on the selected beads undergoes a 3’ modification to allow covalent attachment to the slide.



Step 3. Bead Deposition

3’ modified beads is deposit onto a glass slide. A key advantage of the system is the ability to accommodate increasing densities of beads per slide, resulting in a higher level of throughput from the same system.

Solid sequencing

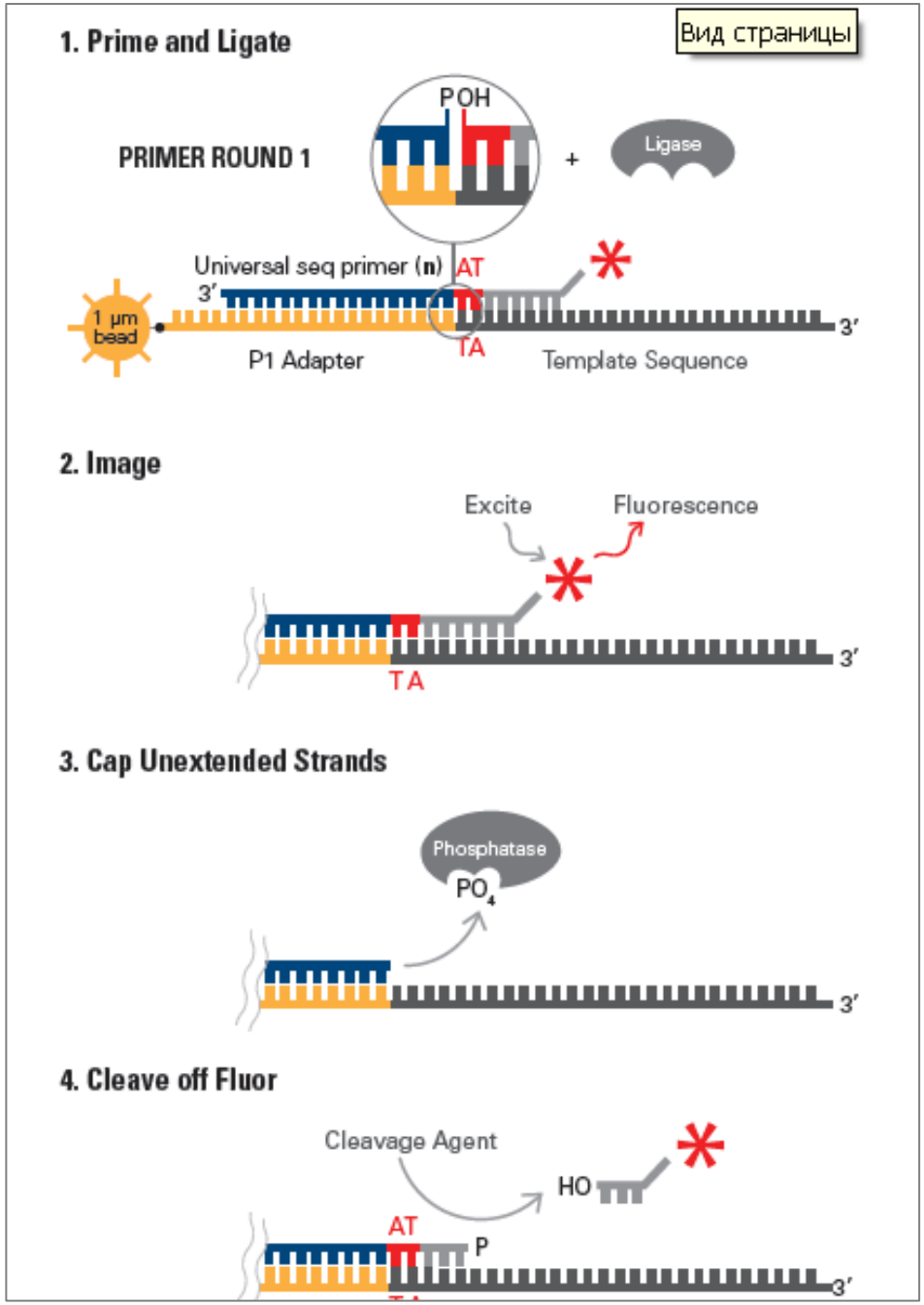


Step 4. Sequencing by Ligation

- Primers hybridize to the P1 adapter sequence on the templated beads.

- A set of four fluorescently labeled di-base probes compete for ligation to the sequencing primer. Specificity of the di-base probe is achieved by interrogating every 1st and 2nd base in each ligation reaction.

- Multiple cycles of ligation, detection and cleavage are performed with the number of cycles determining the eventual read length.

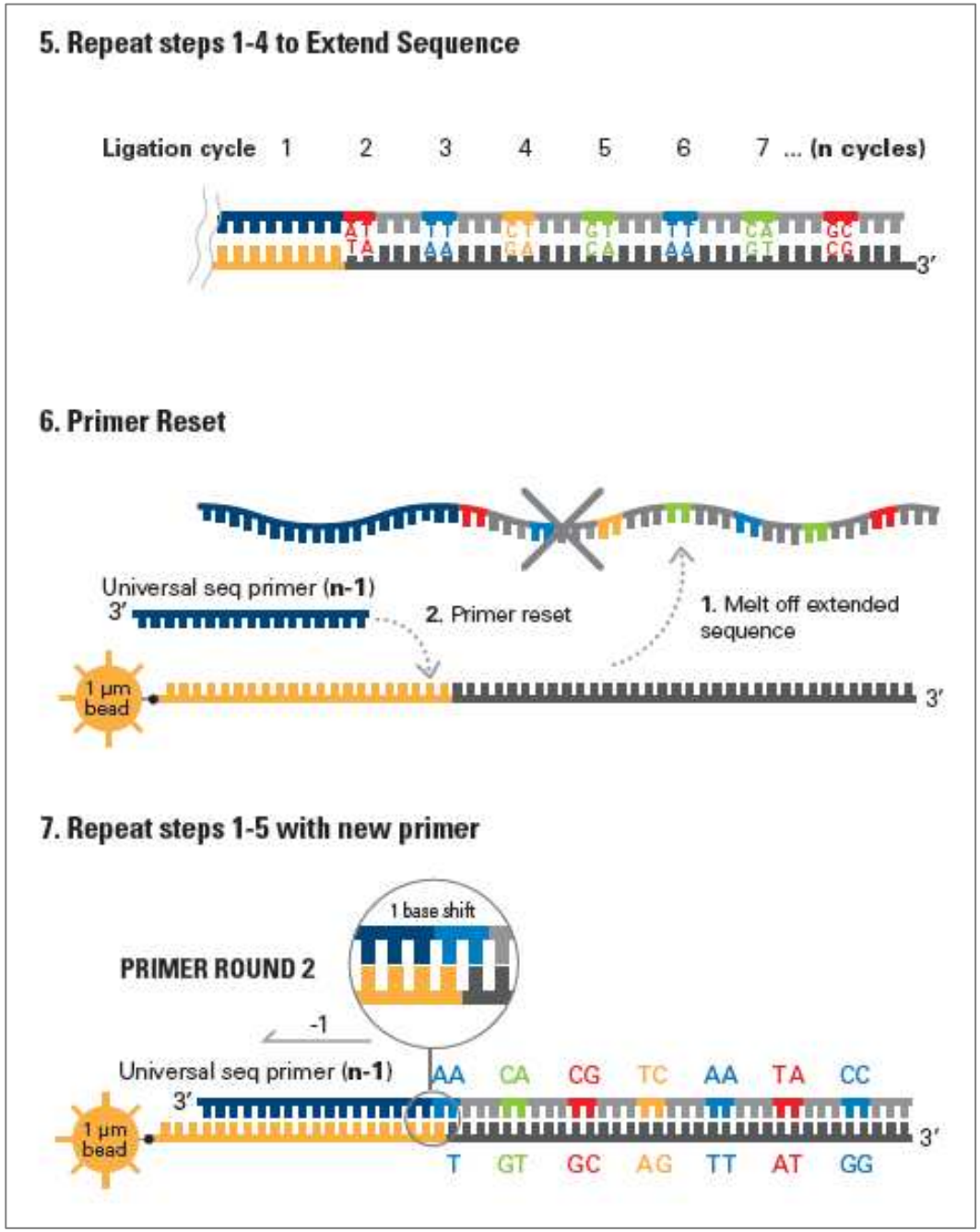


Solid Sequencing

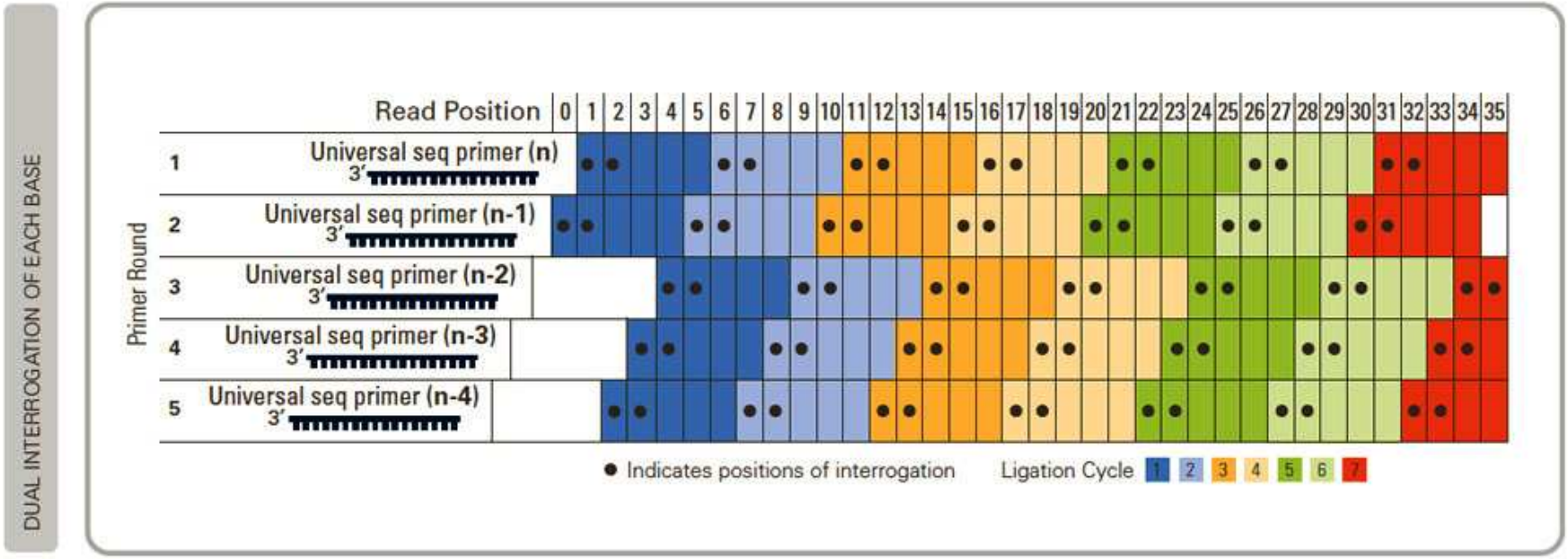
Step 5. Primer Reset

Following a series of ligation cycles, the extension product is removed and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles.

Five rounds of primer reset are completed for each sequence tag.



Solid sequencing

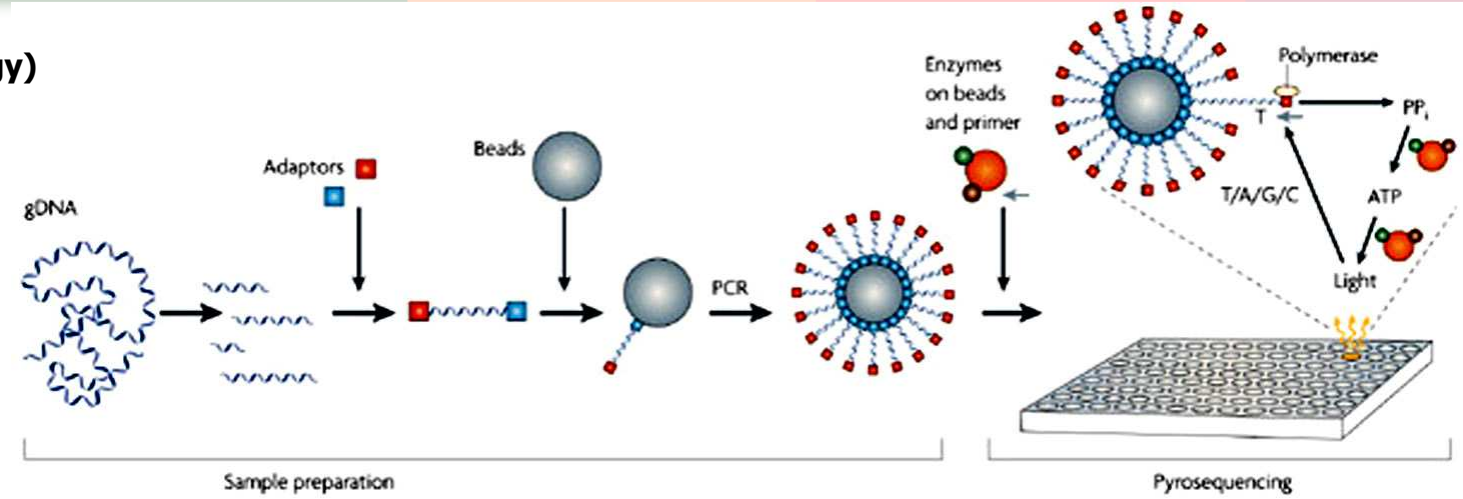


Primer Reset

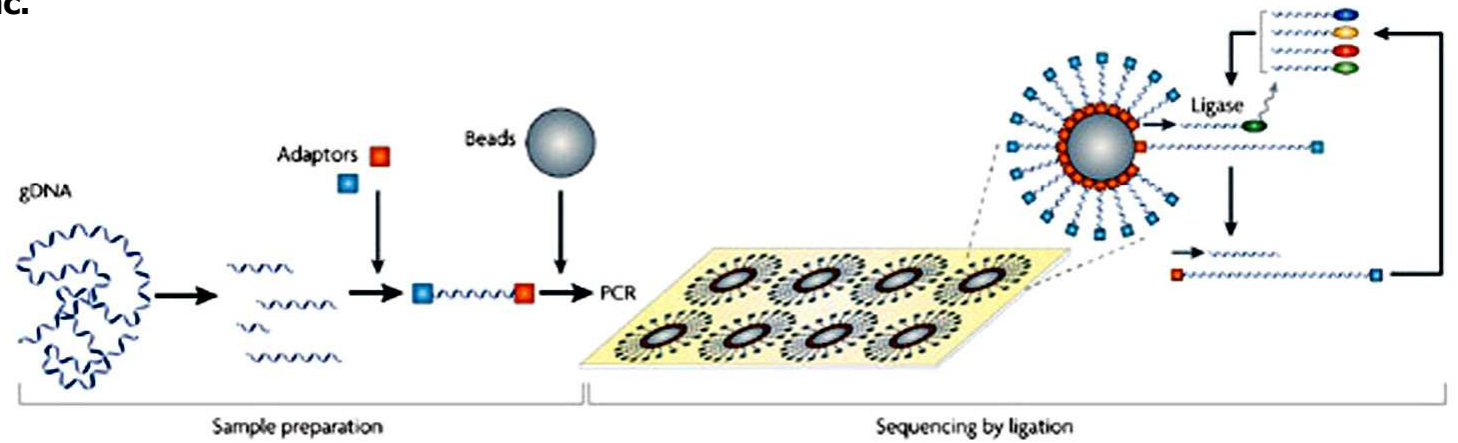
Five rounds of primer reset are completed for each sequence tag. Through the primer reset process, virtually every base is interrogated in two independent ligation reactions by two different primers.

For example, the base at read position 5 is assayed by primer number 2 in ligation cycle 2 and by primer number 3 in ligation cycle 1.

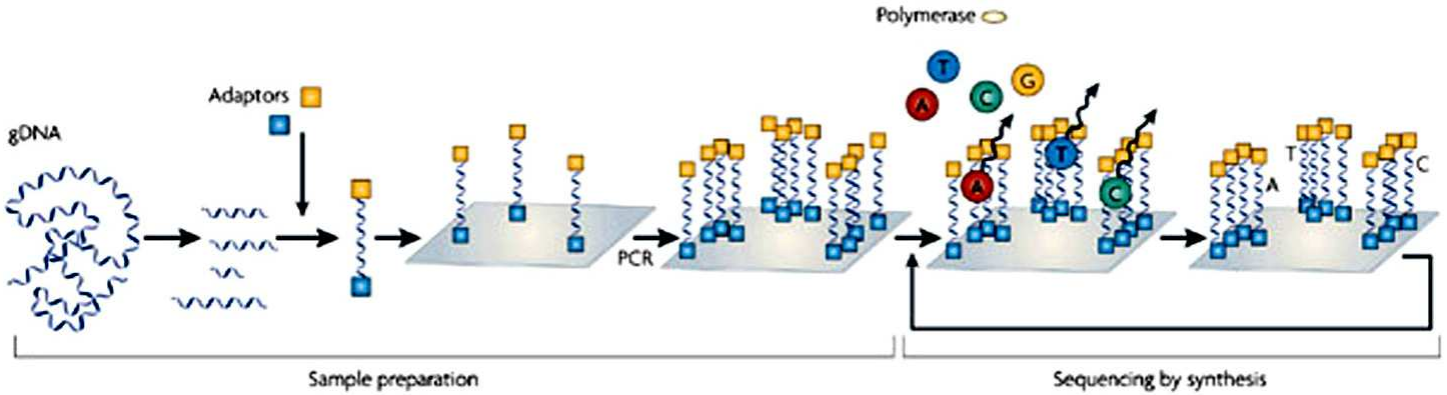
Roche (454 technology)



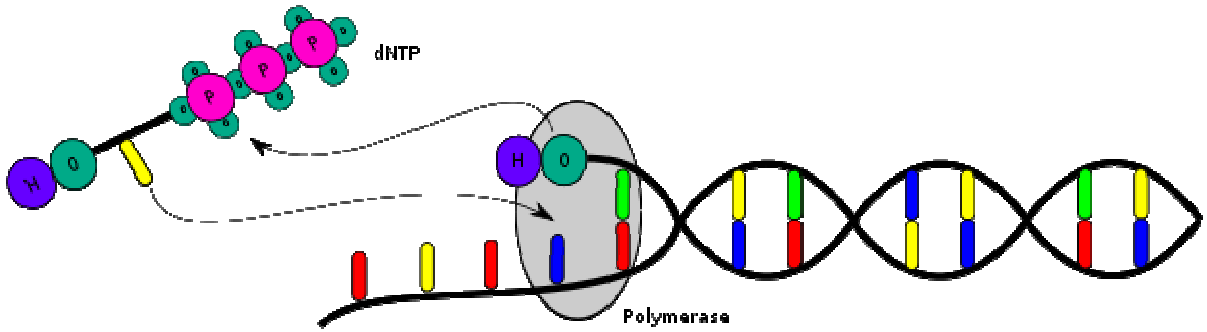
Applied Biosystems Inc. (SOLiDsystem)



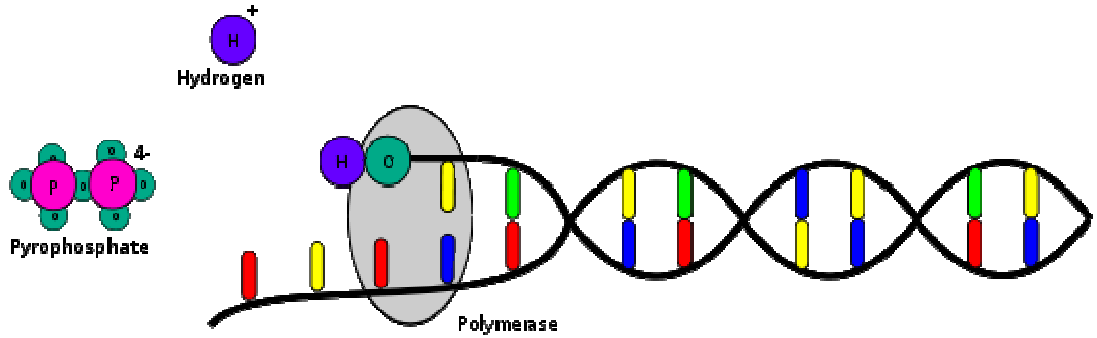
Illumina (Solexa)



Ion semiconductor sequencing



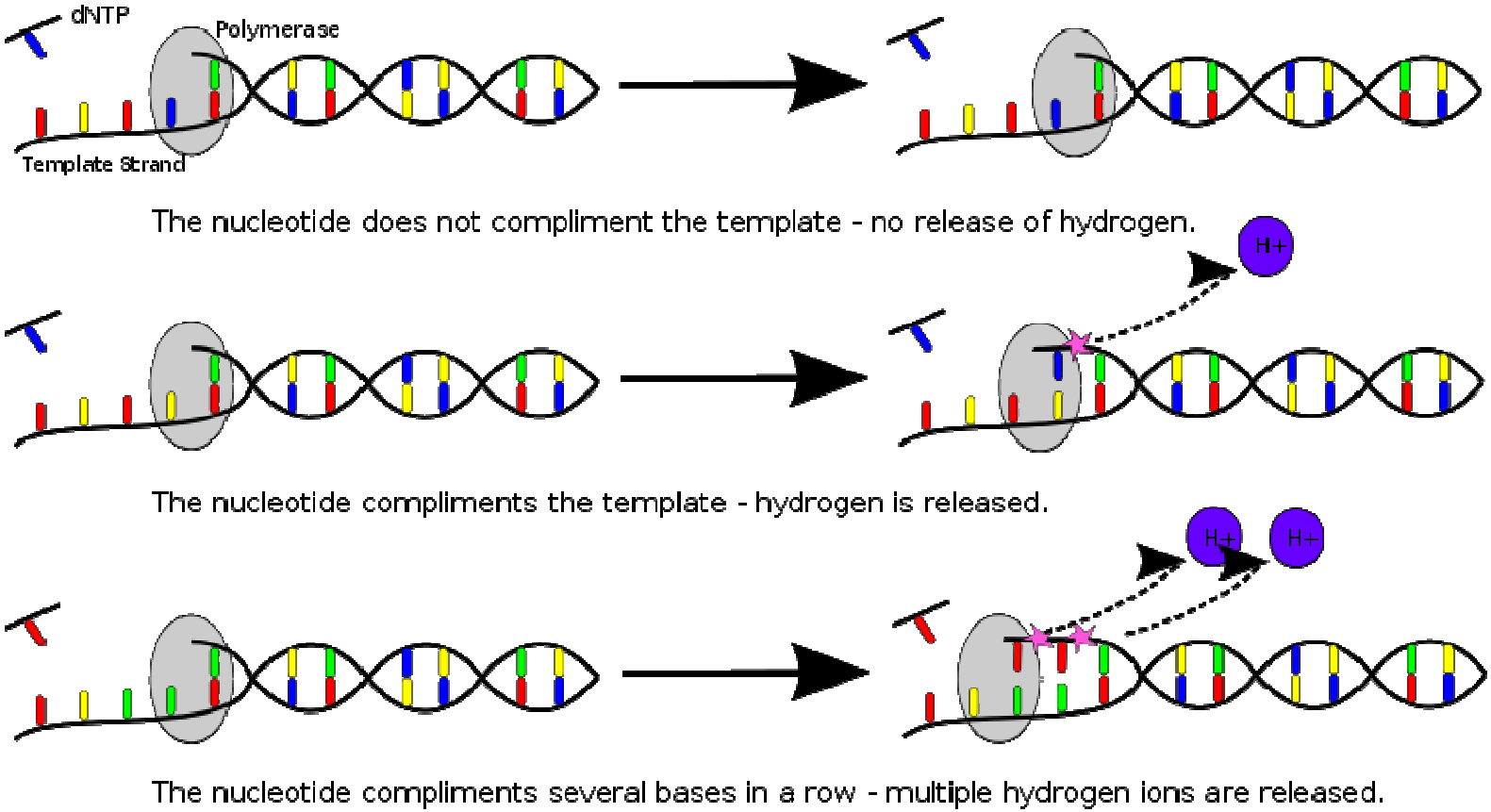
Polymerase integrates a nucleotide.



Hydrogen and pyrophosphate are released.

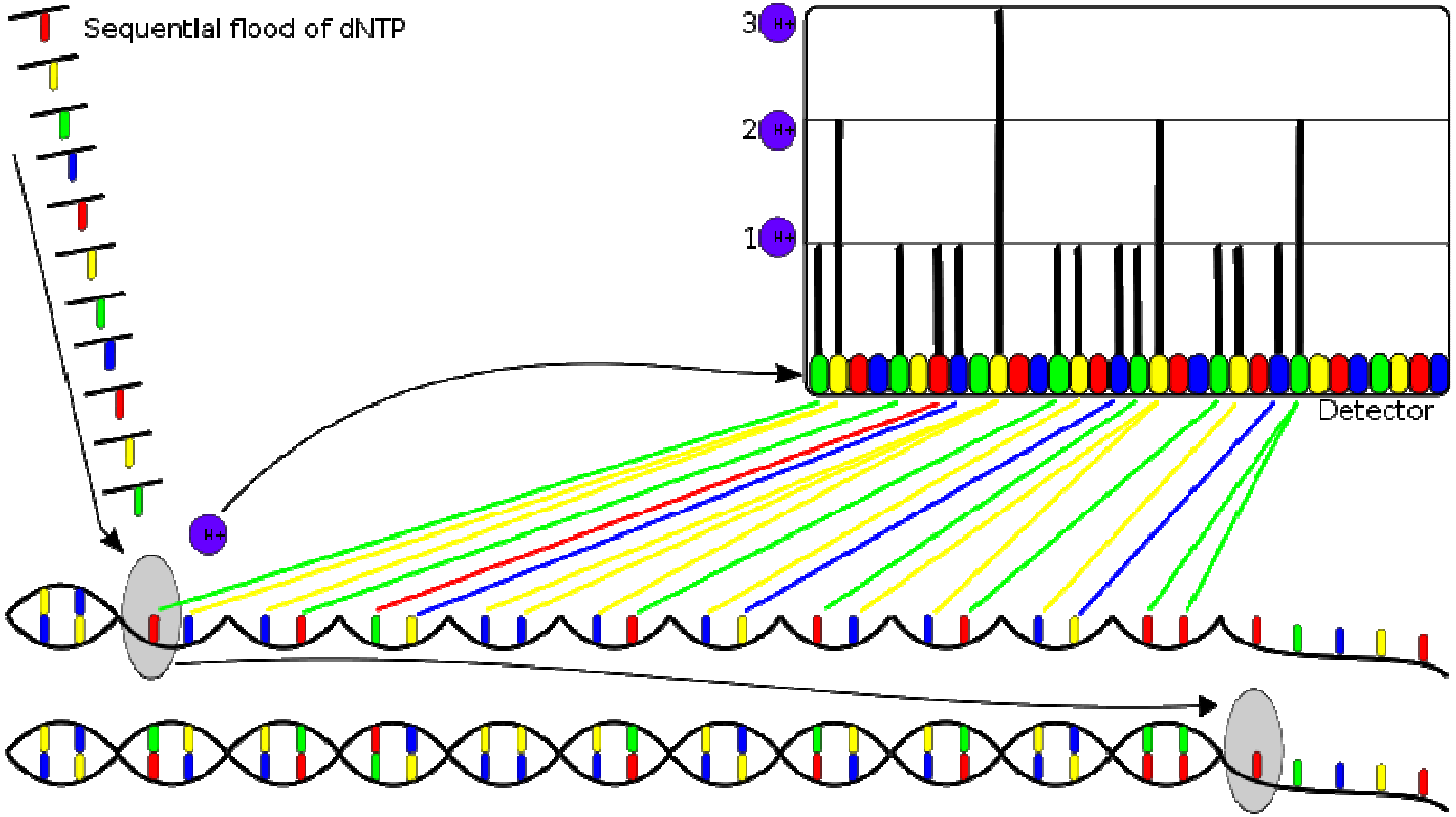
The incorporation of deoxyribonucleotide into a growing DNA strand causes the release of hydrogen and pyrophosphate

Ion semiconductor sequencing



The release of hydrogen ions indicate if zero, one or more nucleotides were incorporated.

Ion semiconductor sequencing



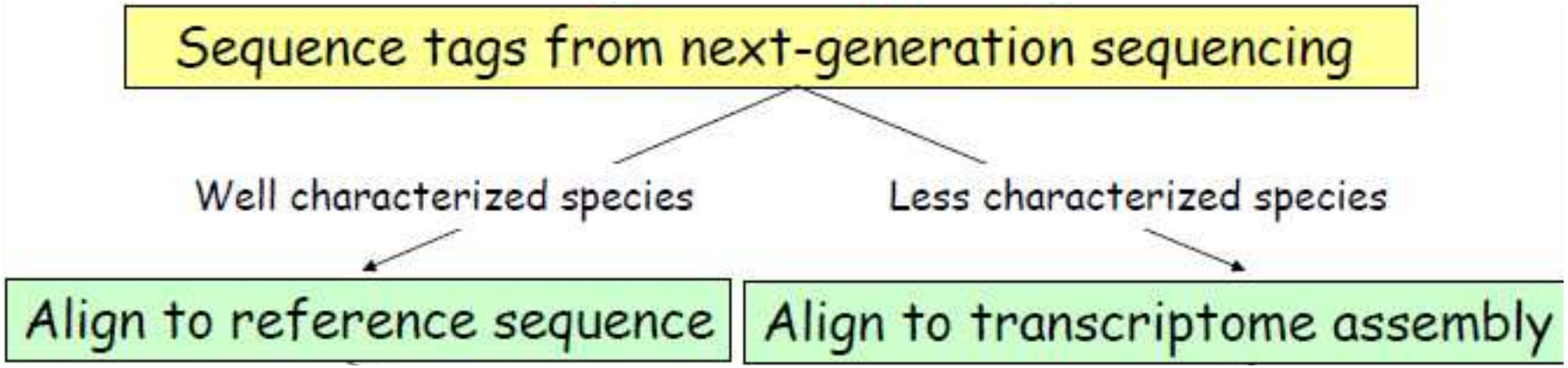
Released hydrogens ions are detected by an ion sensor. Multiple incorporations lead to a corresponding number of released hydrogens and intensity of signal.

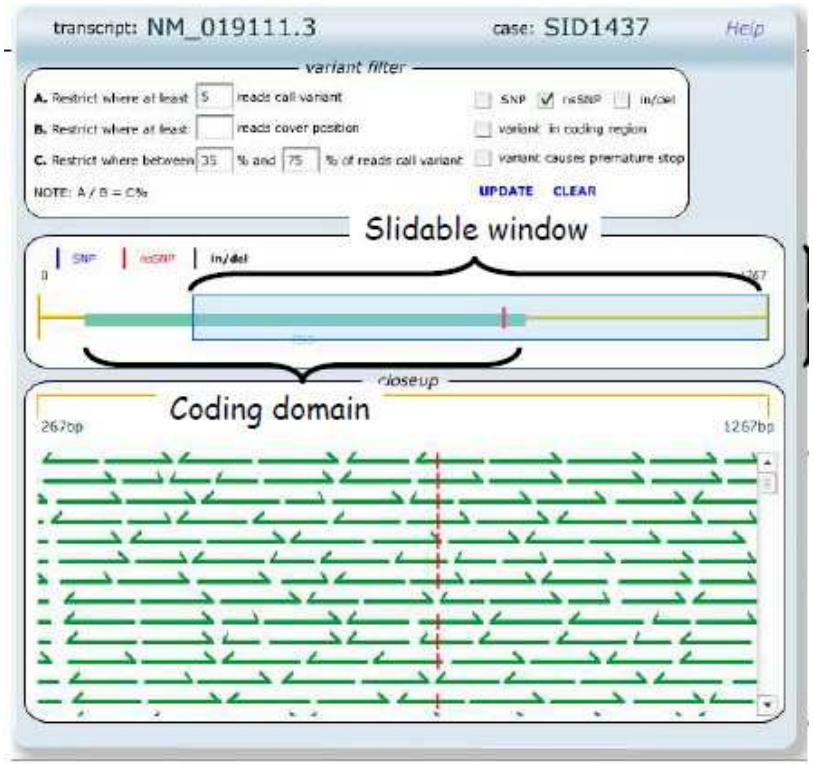
Comparison of sequencing methods

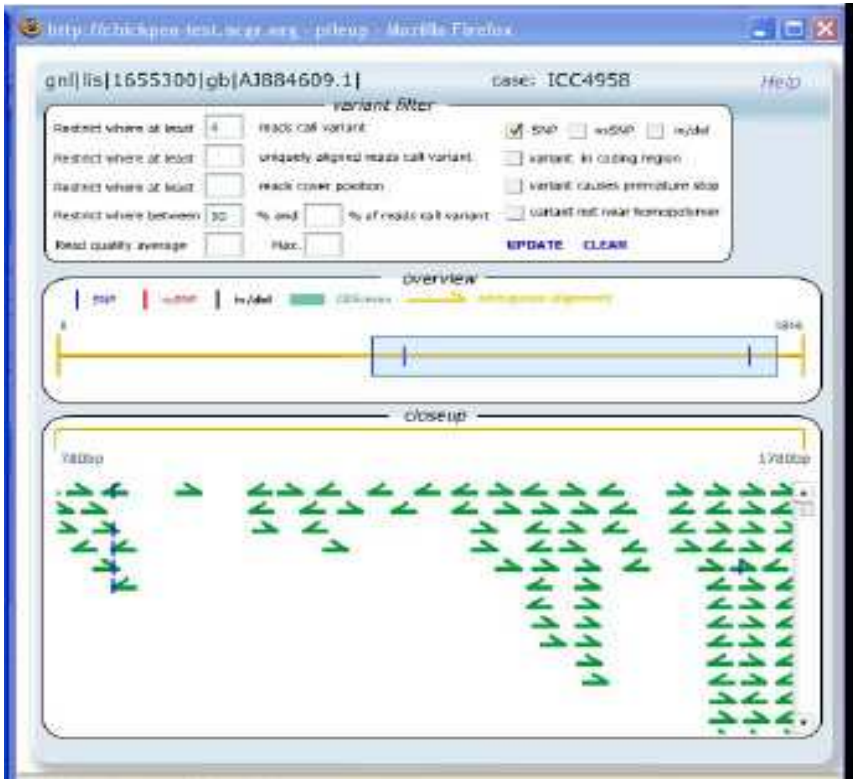
	Ion Torrent	454 Sequencing	Illumina	SOLiD
Sequencing Chemistry	Ion semiconductor sequencing	Pyrosequencing	Polymerase-based sequence-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Emulsion PCR	Bridge amplification	Emulsion PCR
Mb per run	100	100	600,000	170,000
Time per run	1.5 hours	7 hours	9 days	9 days
Read length	200 bp	400 bp	2x100 bp	35x75 bp
Cost per run	$ 350 USD	$ 8,438 USD	$ 20,000 USD	$ 4,000 USD
Cost per Mb	$ 5.00 USD	$ 84.39 USD	$ 0.03 USD	$ 0.04 USD
Cost per instrument	$ 50,000 USD	$ 500,000 USD	$ 600,000 USD	$ 595,000 USD


Oh, my God!
What should I do now?
FASTA
DELIVERY SERVICE
NGS machnies
Massive amount
of sequence data

Sequence processing







Высокопроизводительные платформы сиквенирования - революционные перспективы

Секвенирование геномов

- Человек – индивидуальная геномика
- Выявление новых ассоциированых с заболеваниями генов
- Новые виды животных- новые геномы
- Метагеномы природных биотопов

Секвенирование транскриптомов

- Описание транскриптома
- Анализ дифференциальной экспрессии
- Поиск мутаций в небольших популяциях клеток

Секвенирование генома человека

- 3 млрд bp = 6 млрд bp диплоидный набор
- Хорошее покрытие 20x = 120 млрд

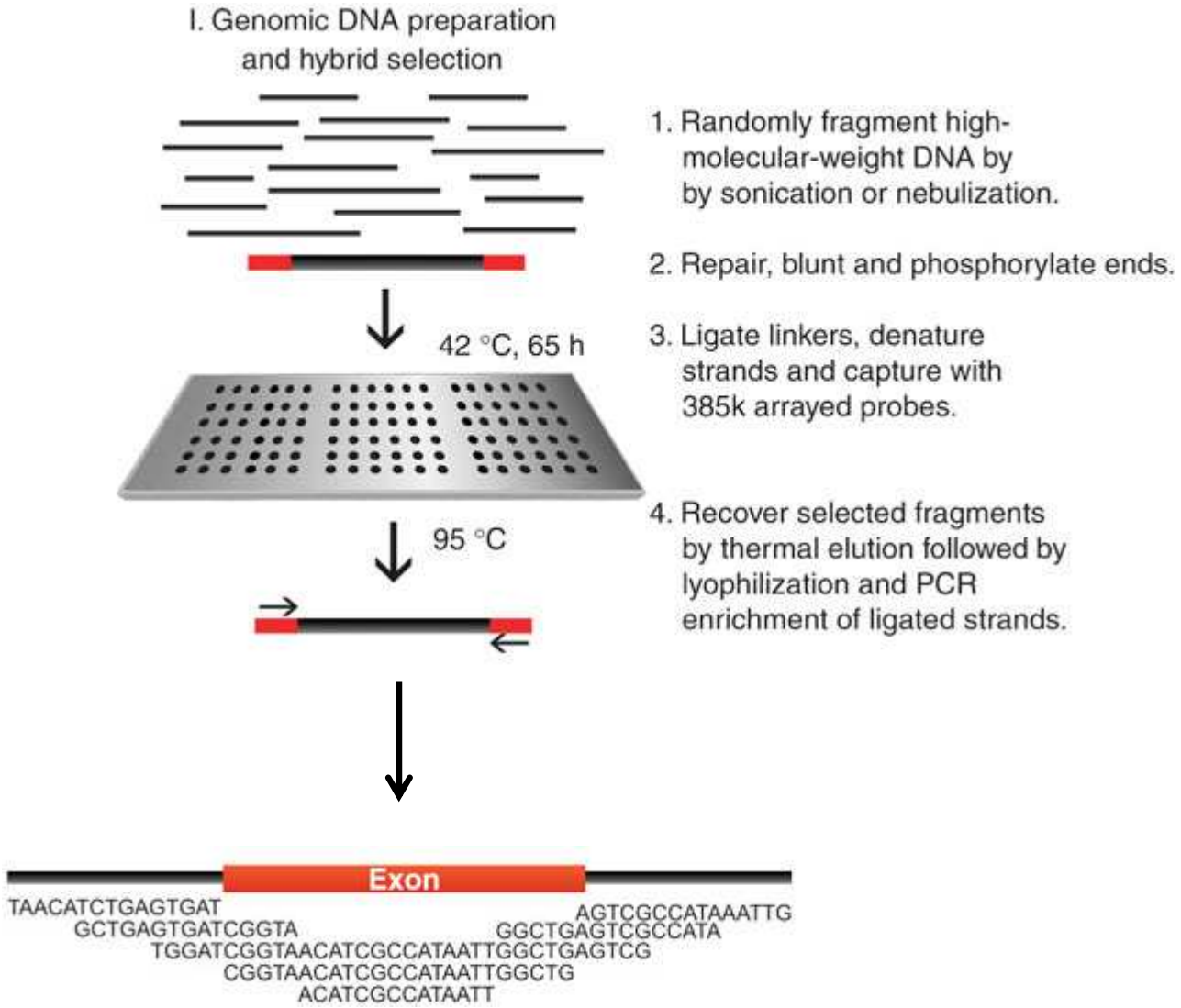
- Solexsa – Анонсирует 90 млрд за один прогон $10 000

Реально это 20-25 млрд т.е. Геном это $50 000 только сиквенс + подготовка образцов+обработка данных $100-150 0000 за 1-3 месяца

Пока слишком дорого и долго, нужно достичь уровня

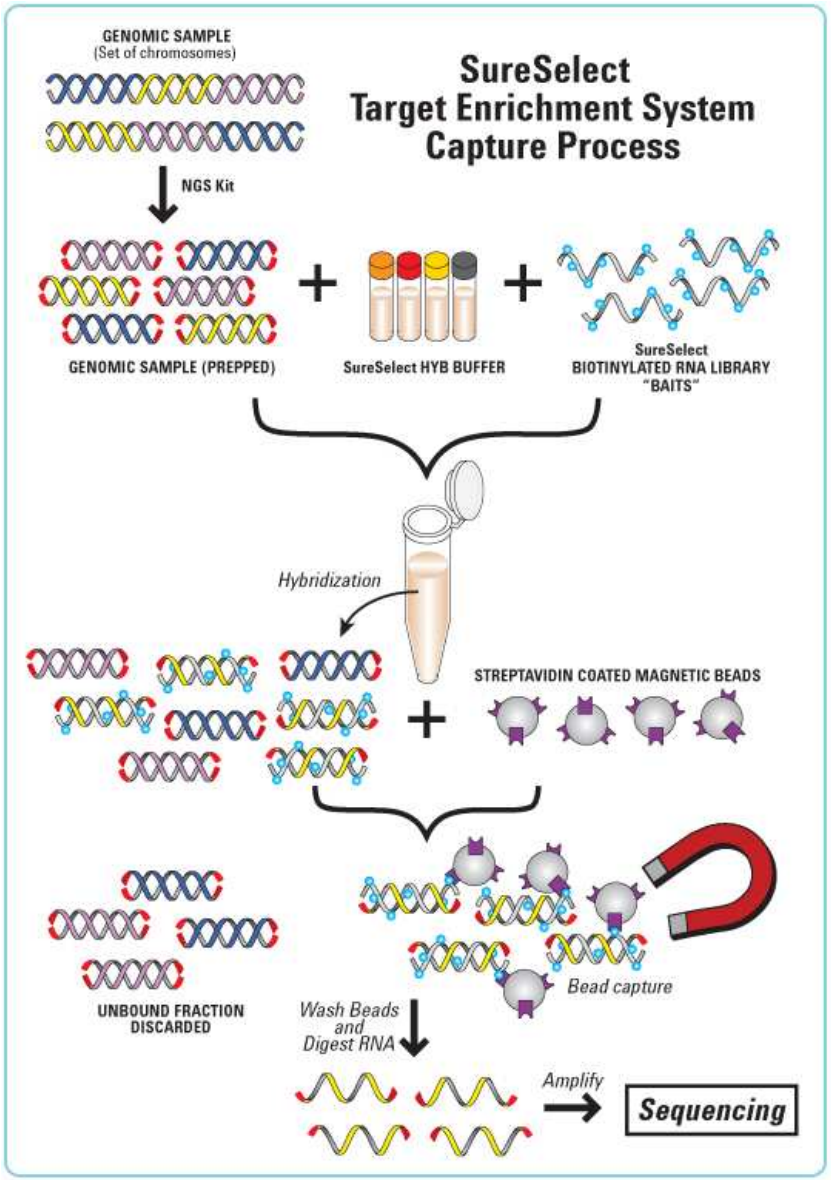
$5 000 за неделю

Genome-wide in situ exon capture for selective resequencing



Sequencing result

Sure Select


GENOMIC SAMPLE
(Set of chromosomes)
SureSelect
Target Enrichment System
Capture Process
NGS Kit
GENOMIC SAMPLE (PREPPED)
SureSelect HYB BUFFER
SureSelect
BIOTINYLATED RNA LIBRARY
"BAITS"
Hybridization
STREPTAVIDIN COATED MAGNETIC BEADS
Bead capture
UNBOUND FRACTION
DISCARDED
Wash Beads
and
Digest RNA
Amplify
Sequencing