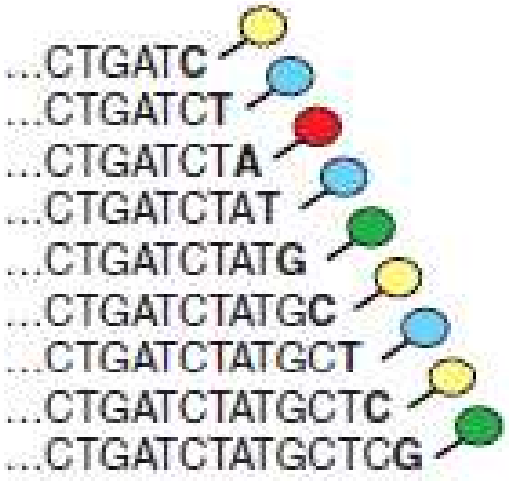
Chapter 7. DNA Microarrays

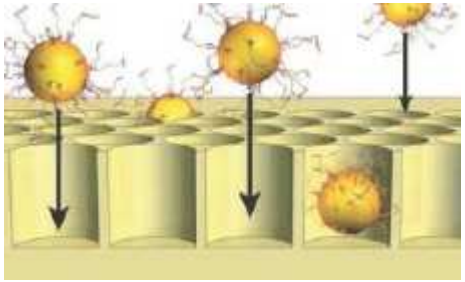
7.9 Next Generation Sequencing

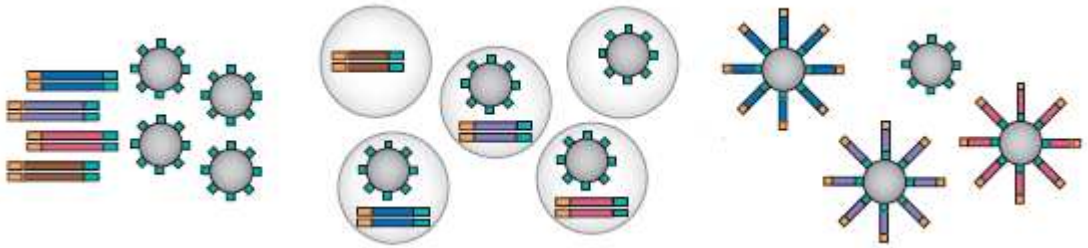
454 Sequencing

Solexa Illumina

Solid ™ System







7.9 Next Generation Sequencing

Sequencing

"Process of determining the nucleotide order of a given DNA fragment"

Wikipedia

To identify, diagnose and potentially develop treatments for genetic diseases.

Treatments for contagious diseases in pathogenesis

Chain termination method: Frederick Sanger

Sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates (**di**-deoxynucleotide)

Extension is initiated at a specific site on the template DNA by 'primer'

Pyrosequencing (Sequencing by synthesis): Pål Nyhren and Mostafa Ronaghi

454 Life Sciences array-based method Stockholm in 1996

7.9 Next Generation Sequencing

Pyrosequencing

- Detection of pyrophosphate release on nucleotide incorporation, rather than chain termination with dideoxynucleotides
- Components: ssDNA template+ primer + enzymes (DNA polymerase, ATP sulfurylase, luciferase and apyrase) , and adenosine 5´ phosphosulfate (APS) and luciferin
- DNA nucleotides are added sequentially in a fixed order → light signal recorded by the camera in the instrument.
- The signal strength is proportional to the number of nucleotides

7.9 Next Generation Sequencing

1. The addition of one (dNTPs) DNA polymerase incorporates the correct, complementary dNTPs onto the template.

   → Releases pyrophosphate (PPi)

2. ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5´ phosphosulfate.

   → Generation of visible light in proportional amounts to the amount of ATP (ATP luciferase-catalyzed reaction).

   → Light produced in the is detected by a camera and analyzed in a program

   → Homopolymer stretches incorporated in a single nucleotide flow generate a greater signal than single nucleotides. Signal strength is linear only up to eight consecutive nucleotides after which the signal falls-off rapidly

3. Unincorporated nucleotides and ATP are degraded by the apyrase, and
4. Reaction is restarted with another nucleotide

7.9 Next Generation Sequencing

Next Generation Sequencing

Massive parallel sequencing

Community of genomics and transcriptomics

Sequences of the GENOME

DNA Suffix and prefix

BIOINFORMATICS CHALLENGE

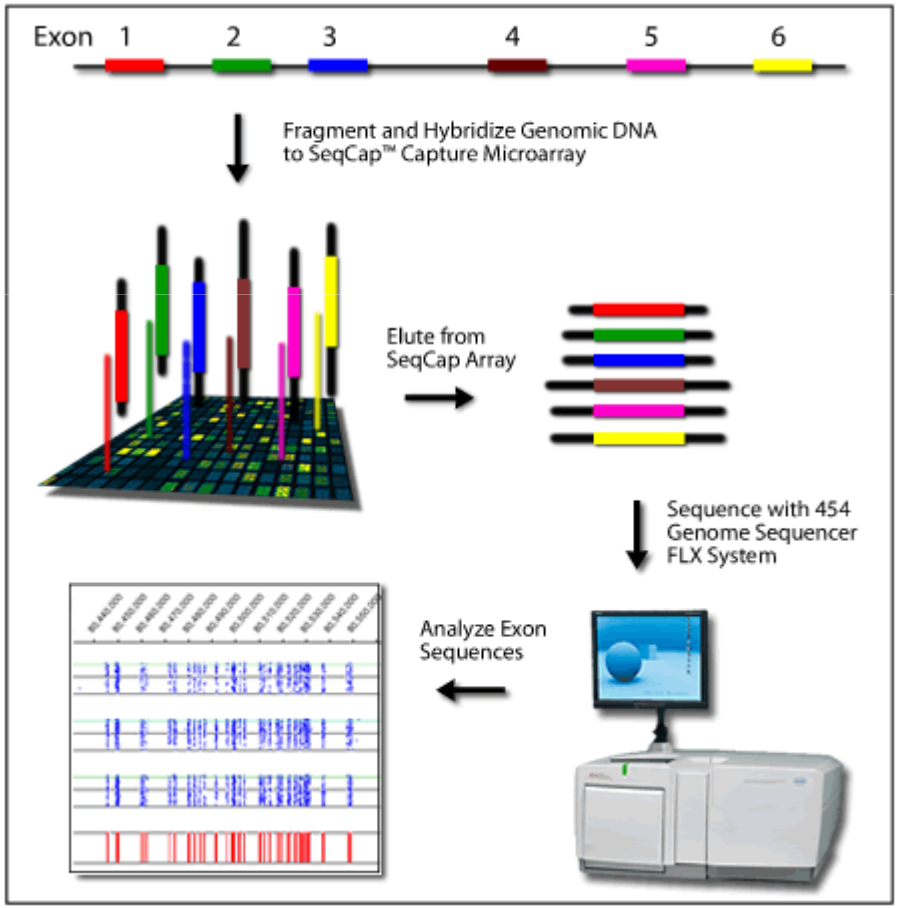
Data Infrastructure

+

Data Analysis

Parallel and GRID computing

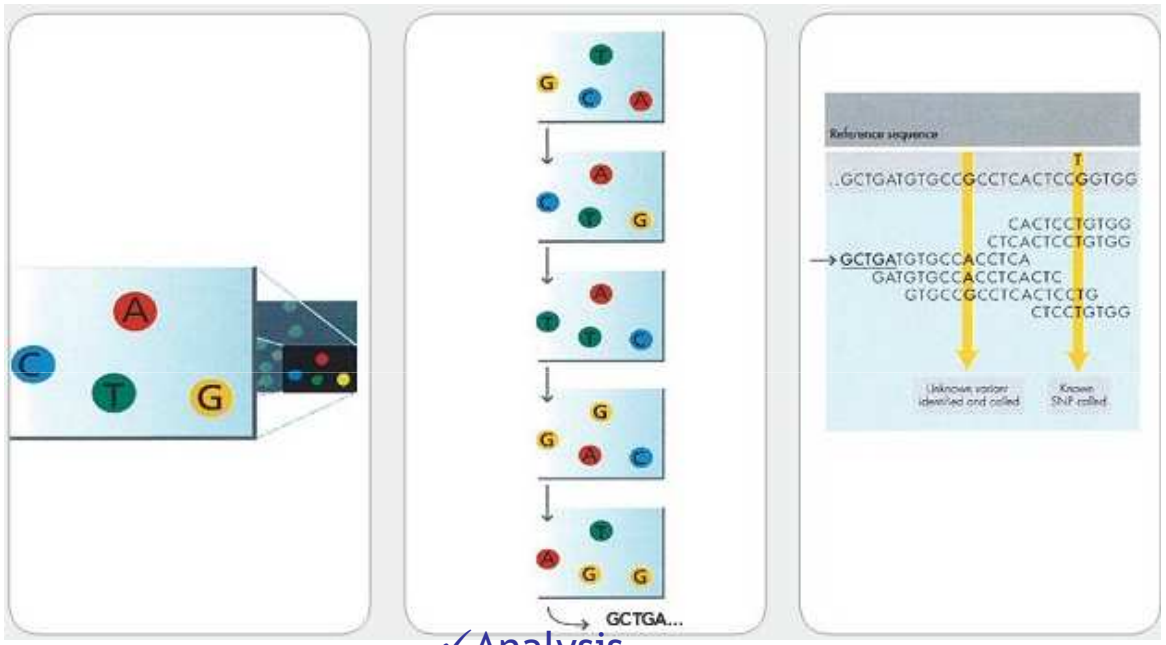
http://www.austriangrid.at/



7.9 Next Generation Sequencing

ILLUMINA Sequencing

- ✓ Reads
- ✓ Storing



✓Back-mapping

Algorithms main memory and Need of parallel on multiprocessors and/or Run on computers GRID

✓Analysis

Assemble a genome
Transcripts determination
Transcripts concentration
Nucleosome position detection
SNPs detection, CNV estimation

7.9 Next Generation Sequencing

ILLUMINA Sequencing

Produces more than 50 million reads

- One read: 30 – 72 long prefix or suffix sequences of DNA fragments with length 100 to 500 base pairs
- Lane: 150 Gb image data per run/experiment
- First: Reads divided into 8 lanes
  - 1 Experiment 1.2 Tb of image data→ 100 Experiments 120 Tb
- Second: reads mapped back to the reference genome
- Third: analysis on the reference genome

7.9 Next Generation Sequencing

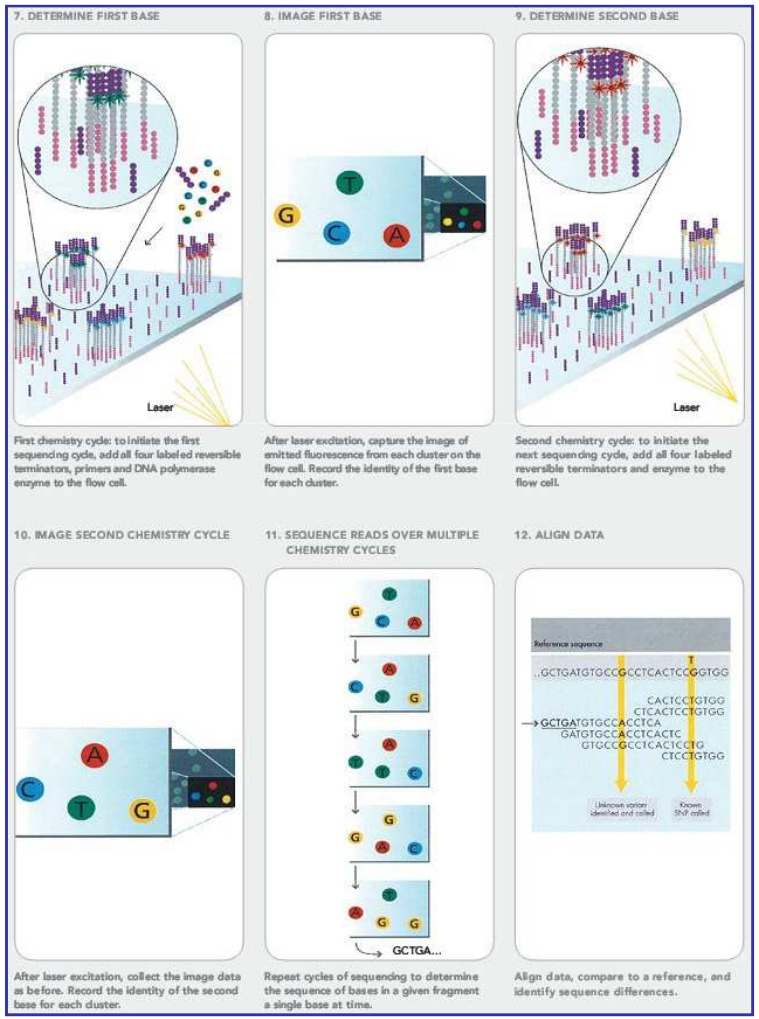
Cluster Generation by Bridge Amplification

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



7.9 Next Generation Sequencing

Illumina Summary

Massive parallel sequencing of millions of fragments by reversible terminator-based chemistry

Templates sequenced using four color DNA sequencing-by-synthesis

Removable terminators with removable fluorescent dyes

Randomly fragmented DNA attached to a planar and transparent surface

Attached DNA fragments are extended and bridge amplified

Cell flow created with > 50 million clusters (each approx. 1000 copies same template)

High sensitivity fluorescence detection by laser excitation

“Paired ends”

Both prefix and suffix (algorithm has to include the constraint of matching pairs of reads)

Second > 36bp reads from the opposite end of the fragment and a cluster formation by a bridge (Total > 3Gb paired-end data)

7.9 Next Generation Sequencing

454 Roche

Reads: 400.000 per run BUT 400-800 bases length ?????

Better suited for genome assembly by larger overlaps

Coverage of the genome 100 times smaller than Solexa

SOLID™ AppliedBiosystem

Sequencing of clonally amplified DNA fragments linked to beads

The sequencing methodology is based on sequential ligation with dye-labeled oligonucleotides

Comparable to Solexa

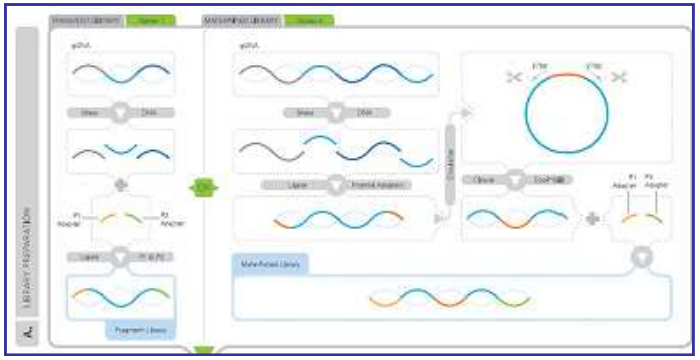
Reads: 10-20 millions per run

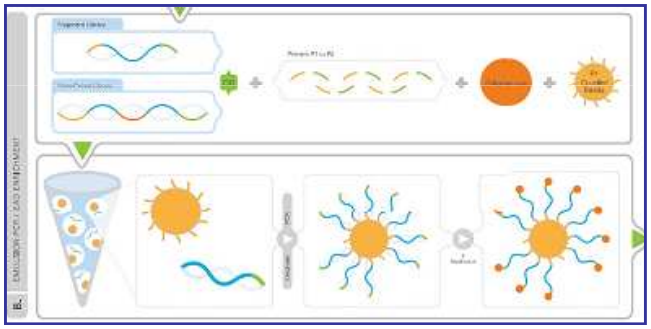
7.9 Next Generation Sequencing

SOLID™ AppliedBiosystem

1. Library preparation

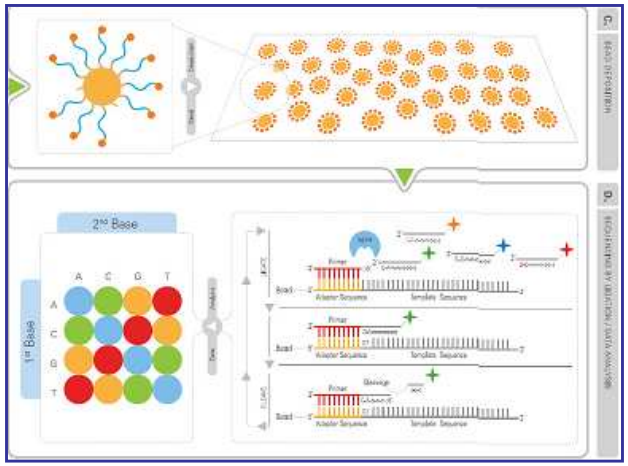
   Two types of libraries can be used





2. Emulsion PCR/ Beads enrichment

   Cloning fragments are prepared



3. Bead deposition and sequencing by ligation data analysis

7.9 Next Generation Sequencing

Solexa Illumina
http://www.illumina.com

s454 Sequencing
http://www.roche.com/

SOLID™ System
http://www3.appliedbiosystems.com/AB_Home/index.htm

Papers

1. **Next-generation DNA sequencing**. Jay Shendure & Hanlee Ji

2. **Next-generation sequencing (**Short communication**).**Jorge S Reis-Filho

3. **CNV-seq, a new method to detect copy number variation using high-throughput sequencing** Chao Xie and Martti T Tammi