Special Topics on Genetics

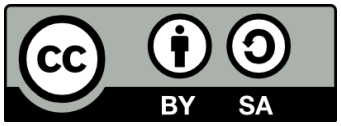
Section 6: Functional Genomics Platforms

Triantafyllidis A.
School of Biology

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

License

- The offered educational material is subject to Creative Commons licensing.
- For educational material, such as images, that is subject to other form of licensing, the license is explicitly referred to within the presentation.



Funding

- The offered educational material has been developed as part of the educational work of the Instructor.
- The project "Open Academic Courses at Aristotle University of Thessaloniki" has financially supported only the reorganization of the educational material.
- The project is implemented under the Operational Program "Education and Lifelong Learning" and is co-funded by the European Union (European Social Fund) and national resources.

Licensing of figures

We warmly thank the Pearson Education Inc for granting the right to use the following figure of this presentation:
Figure: 1

This figure comes from the book Peter Russell, iGenetics: A Mendelian approach, 2006, Pearson Education Inc, publishing as Benjamin Cummings.

Section Contents

- Functional Genomics – An Introduction
- First-generation sequencing platforms
- Second-generation sequencing platforms
- Third-generation sequencing platforms
- Comparison of sequencing platforms
- DNA Arrays
- Mass spectrometer machines

Functional Genomics (1/2)

The need for comprehensive analyses has led to the development of **high-throughput platforms**.

One genomic high-throughput platform includes:

- an automated system for the production of DNA fragments for sequencing,
- an automated sequencer, and
- many computational methods to collect and process data.

Functional Genomics (2/2)

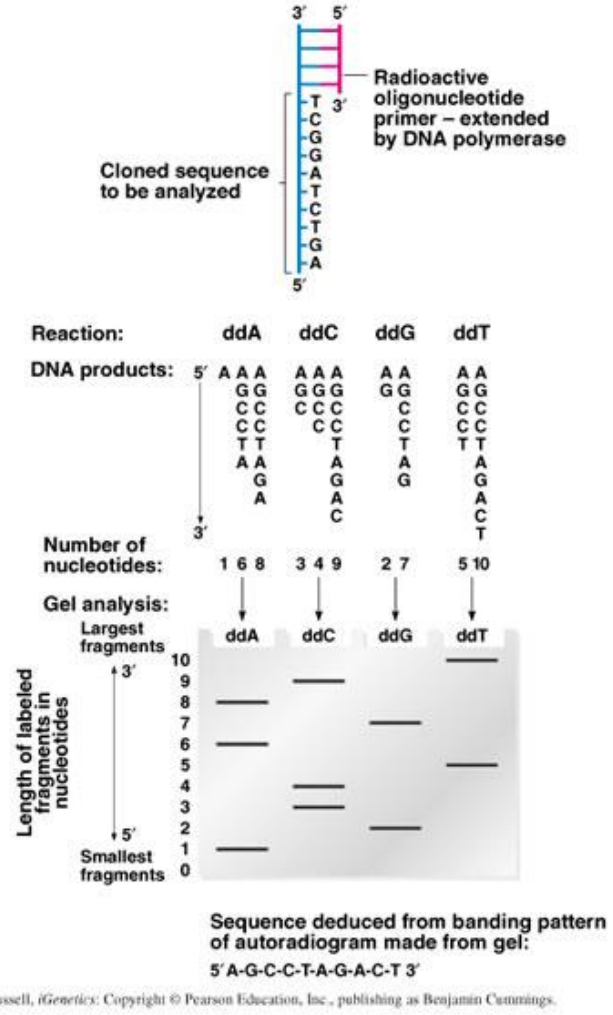
Such machines that actually produce BIG data are the following:

- **Sequencers**
- **DNA arrays**
- **Mass spectrometers**

First-generation sequencing platforms (1/5)

Figure 1: Sanger sequencing method

A method that uses dideoxynucleotides.



http://www.youtube.com/watch?v=oYpllbI0qF8

First-generation sequencing platforms (2/5)

Stage 1

DNA polymerase synthesizes a complementary strand starting from a primer.

4 different reactions, each including a dideoxy analog of the four nucleotides G, A, C, and T are mixed with normal nucleotides.

Approximately 0.5% of synthesized sequences are terminated.

Labelling of the 4 reactions is done with **4 fluorescent dyes** attached to the primers or to the dideoxy-nucleotides

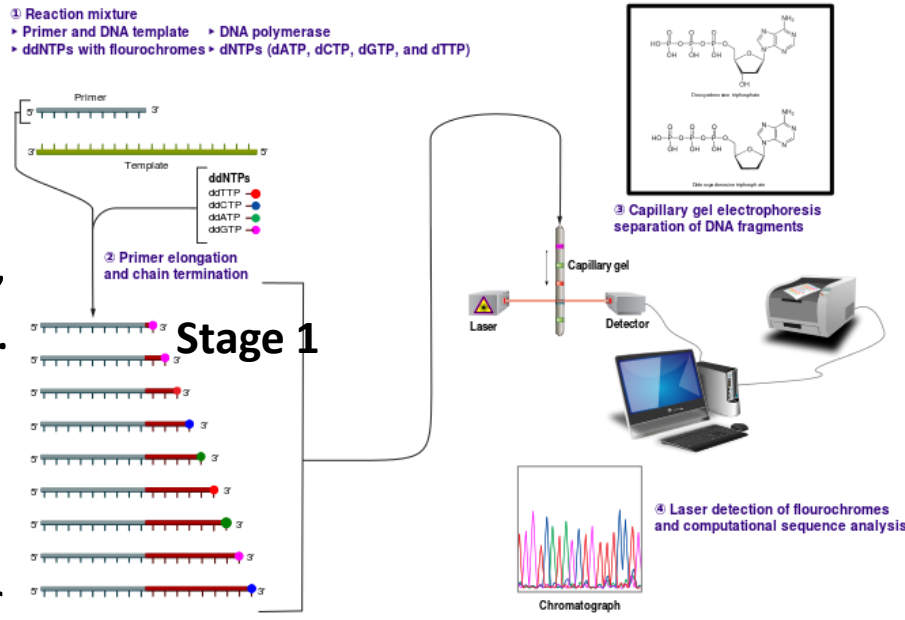


Figure 2: Sanger sequencing.

http://www.youtube.com/watch?v=dUjMf2ZezIw

☞Radioactive labeling is not used anymore

First-generation sequencing platforms (3/5)

Stage 2

The four different reactions with fragments of different size are electrophoresed in one column or capillary.

The electrophoretic conditions allow the separation of DNA fragments that differ in size of a nucleotide.

http://www.yourgenome.org/downloads/animations.shtml

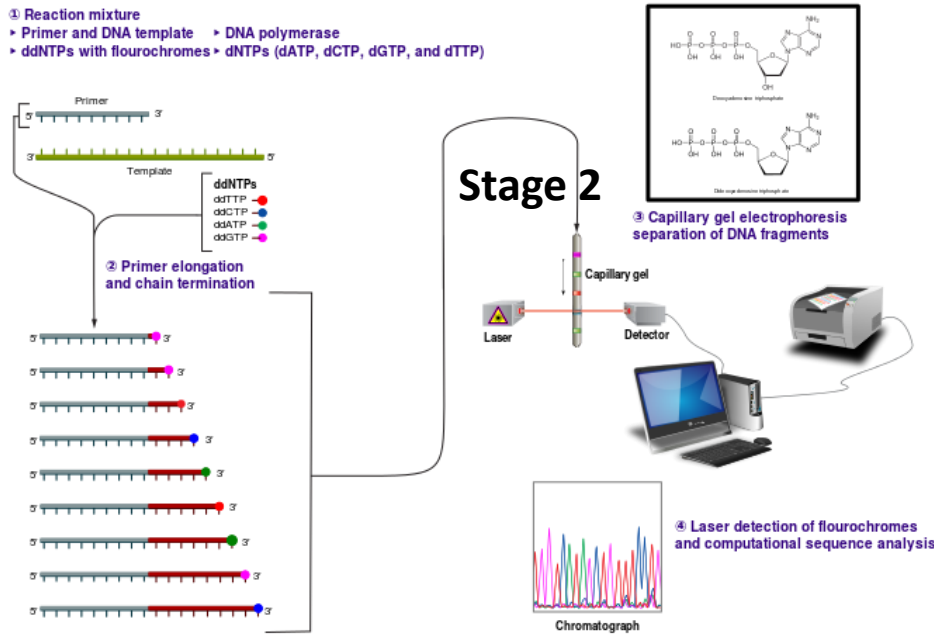


Figure 2: Sanger sequencing.

First-generation sequencing platforms (4/5)

Stage 3

When DNA fragments reach the end of the electrophoresis, they **are scanned with the help of a laser.**
With the help of the photocell, light emission is detected and the signal is transferred to a computer that automatically analyzes and stores data.

It has been established that the presence of nucleotide A is labeled and fluoresces with green color, G black, C blue and T red.

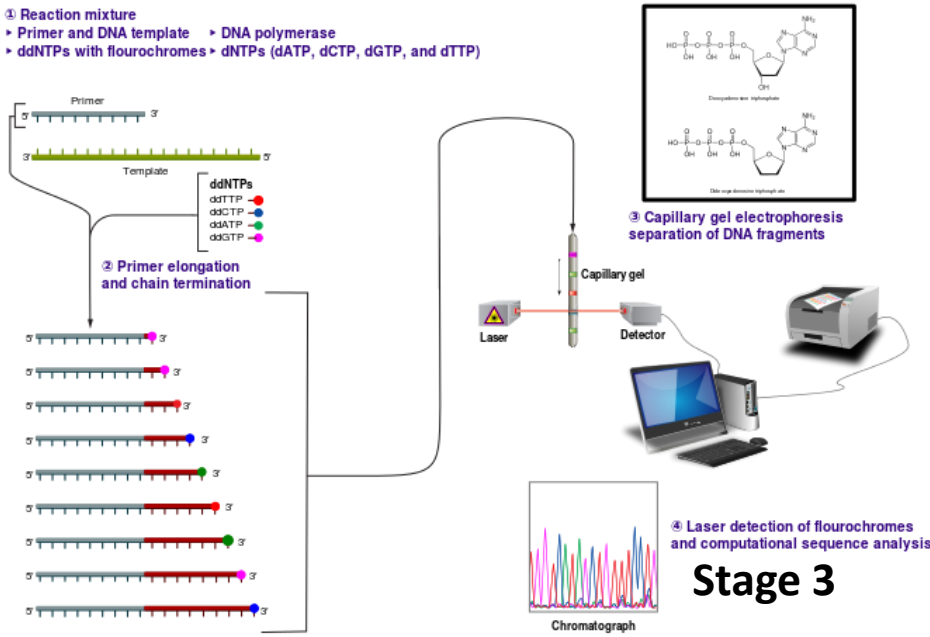


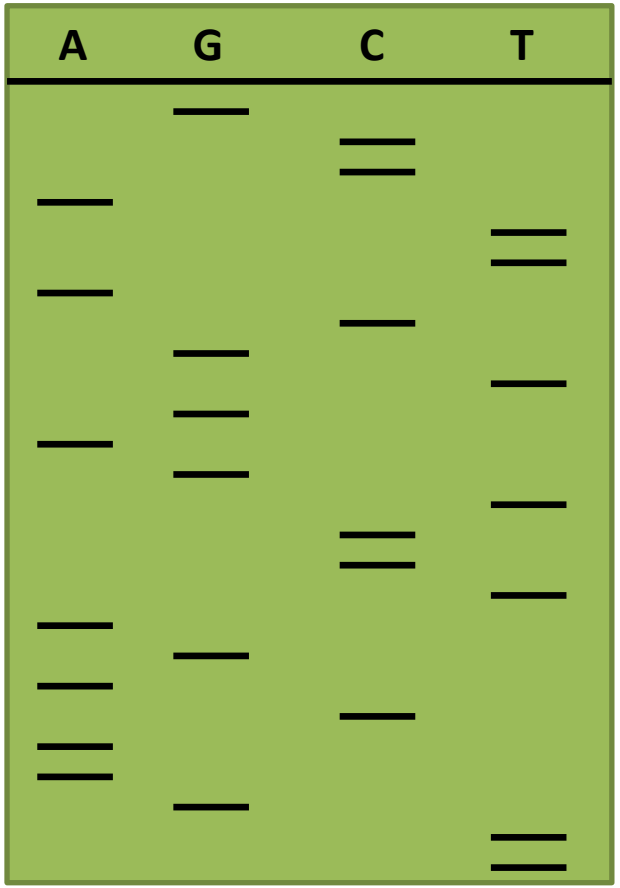
Figure 2: Sanger sequencing.

First-generation sequencing platforms (5/5)

- Automated sequencers can easily simultaneously analyze 384 DNA fragments.
- Production of more than 10 Mb of data per day is possible.

Exercise 1

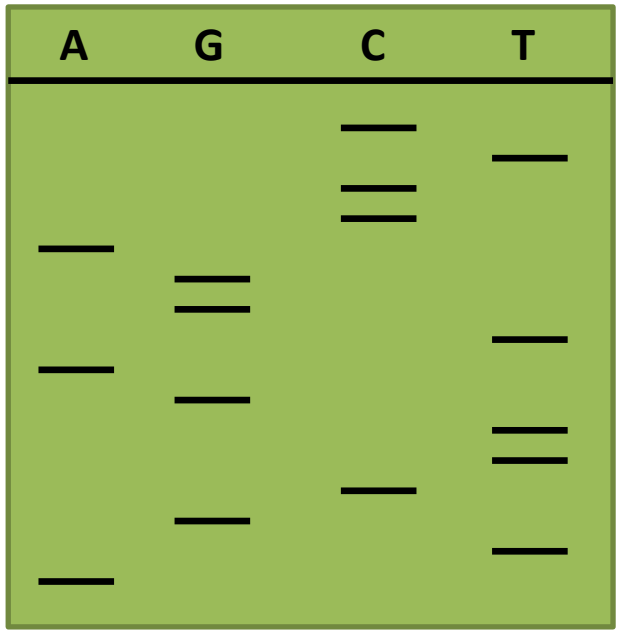
Which is the sequence of the initial strand template?



Solution:
5′ TTGAACAGATCCTGAGTGCATTACCG 3′
5′ CGGTAATGCACTCAGGATCTGTTCAA 3′

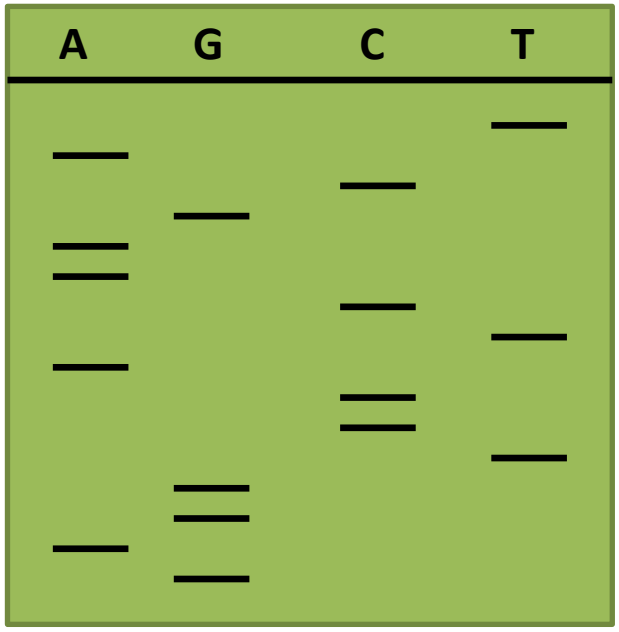
Exercise 2

Draw the sequencing pattern in the complementary strand



Exercise 2 - Solution

ATGCTTGATGGACCTC → GAGGTCCATCAAGCAT



… and rates constantly increase

…next generation sequencers (NGS)

…2nd generation sequencers

…3rd generation sequencers

Second-generation sequencing platforms (1/8)

- In 1987 an automated ABI machine produced 4800 bases/day!!!
- Illumina machines - produce 25 billion bases/day
- Life Technologies - 100 billion bases (soon 300 billion)
- Costs may be less than $ 10000

We can easily produce a human genome within a week.

Second-generation sequencing platforms (2/8)

454 sequencing machines

- Nature, Rothberg *et al*. 2005

25.000.000 bases in 4 hours!

- Producing a whole human sequence within a few days.
- Sequencing the genome of *Mycoplasma genitalium* in 4 hours with 99.96% accuracy.
- Its size is similar to a microwave.

The revolution is based on the sample preparation and analysis platform = 1.6 million reactors in slides 6.4 cm^2.

Massively parallel machines!

http://www.454.com/ http://454.com/products/technology.asp

Second-generation sequencing platforms (3/8)

454 sequencing machines

- Fragmentation of DNA
- Attachment of single stranded DNA sequences to adapters, that allow the attachment to a unique ball-bead 28 μm in diameter
- They are incorporated in a drop of oil, where necessary reagents for emPCR are present
- The oil drops do not fuse → no contamination between different fragment PCRs is allowed

Second-generation sequencing platforms (4/8)

454 sequencing machines

- The beads with the DNA substrate are placed into slides of photosensitive optical fibers, in which the necessary enzymes for sequencing are present.
- With each addition of a nucleotide, a photosensitive sensorreads the nucleotide addition (or not) in each one of the 1.6 million reactors!

Second-generation sequencing platforms (5/8)

454 sequencing – Problems in 2005

- Ideal for re-sequencing of small genomes or for sequencing of small genomes with unique DNA ... BUT
- Its reads were small (fragments of 80-120 bases only)
- With reduced accuracy
- accurate assembly was not possible
- The procedures were not fully automated

- Read length nowadays: At least >500 bases

Hutchison CA 2007 Nucleic Acids Research 1-11
Metzker 2010 Nature Reviews Genetics 31-46

Second-generation sequencing platforms (6/8)

Applied Bioplatforms SOLiD

- 2006
- Performs 100 millions reactions
- Read length: 35 bases

http://www.appliedbioplatforms.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-platforms/solid-sequencing-chemistry.html

https://www.youtube.com/watch?v=nlvyF8bFDwM

Second-generation sequencing platforms (7/8)

Illumina (in 2007)

- Performs 80 million reactions
- Read length: 45 bases

http://res.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf
http://www.illumina.com/platforms/miseq.ilmn
https://www.youtube.com/watch?v=womKfikWlxM

In the 2nd quarter of 2014, Illumina reported adjusted earnings of 57% per share – most probably the biggest increase in the companies history.

Second-generation sequencing platforms (8/8)

Ion Torrent-Life tech-ABI

1.000 $ genome candidate!

- It costs only $50000
- It needs ONLY an ion microchip which costs $99 to read a bacterial genome (25 Mb) in 2 hours
- There are 1.2 million sites for reactions (similar to 454)
- It is based on the ability of the machine to recognize a pH change due to the ions released by the addition of bases during DNA synthesis
- Average fragment size 100bp
- Already successful

http://www.iontorrent.com/

http://www.youtube.com/watch?v=yVf2295JqUg&feature=BFa&list=PL171604F04548D64E

http://www.youtube.com/watch?v=fxCY_f0QaZQ

Third-generation sequencing platforms (1/3)

Pacific Biosciences, Oxford nanopore, Starlight

- Sequencing of a SINGLE MOLECULE without AMPLIFICATION
- Reads the sequence of the existing clone
- We can overcome the target of one genome per person and
- We can take "photos" from the same person, from different tissues / cells such as germ, stem cells, cancer cells

See:
http://www.genome.gov/27543255

Third-generation sequencing platforms (2/3)

Single-molecule, real-time DNA SMRT sequencing

http://www.pacificbiosciences.com/

- Its method is based on fluorescence
- Real-time observation of the DNA polymerization
- Four phosphorus-linked nucleotides of different colors are used
- Fluorescence is released simultaneously with nucleotide incorporation

J. Eid et al., Science 323, 133 -138 (2009)

http://www.dnatube.com/video/3003/SMRT-DNA-Sequencing

Third-generation sequencing platforms (3/3)

Oxford Nanopore

Completely different approaches: these machines use nanotechnology and proteins with nanopores. They detect each different nucleotide as it passes through the pores.
The success of these methods, is based on the sensitivity of detection of very small differences in the physicochemical properties of the four basic nucleotides and also on the achievement of a constant rate of transit of the nucleotide sequence into the pores (by the inclusion of exonuclease, which cuts one by one the nucleotides and reads them afterwards).
Such a machine with size similar to a USB is now available!

https://www.nanoporetech.com/ https://www.youtube.com/watch?v=Sx6FbYoFGmM
https://nanoporetech.com/technology/the-minion-device-a-miniaturised-sensing-system/the-minion-device-a-miniaturised-sensing-system

Comparison of sequencing platforms (1/15)

Read lengths and data production have increased due to the upgrades of NGS machines.

Field guide to next-generation DNA sequencers –Glenn 2011
http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2011.03024.x/pdf

http://flxlexblog.wordpress.com/2013/10/01/developments-in-next-generation-sequencing-october-2013-edition/

Comparison of sequencing platforms (2/15)

Sequencing costs

Machines can be divided into two groups:

- Those that produce a moderate number of reads with large size, with relatively high cost per Mb of sequence (454, Ion Torrent, PacBio, Starlight).
- On the other hand, those that produce huge number of reads with small size, with low cost per Mb of sequence (Illumina, SOLiD).

- New third generation machines (e.g. Oxford Nanopore) may be able to produce several reads, with large size with relatively low cost.
- For de novo sequencing, however, the combination of Illumina machines and 454 is suggested.

Field guide to next-generation DNA sequencers – Glenn 2011
http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2011.03024.x/pdf

Comparison of sequencing platforms (3/15)

• We have to remember that none of the machines, that have been developed so far, possess such a technology, that could allow for full continuous sequencing of a chromosome even of the smallest genome.

• Therefore, intelligent bioinformatics approaches for the assembly of the produced sequences into a whole genome are essential for the success of genome sequencing approaches.

• Its also evident that older machines will not become obsolete. Each one, has its own characteristics and each researcher depending on his budget and the program objectives should choose after careful consideration.

Comparison of sequencing platforms (4/15)

Sequencing cost

Gene sequencing leaves the laboratory

Nature 2013, 494, 290-291

http://www.nature.com/news/gene-sequencing-leaves-the-laboratory-1.12454

- Cost of human genome: ~ $5000
- The cost and the capabilities of available technologies, have greatly enabled medical research and related applications
- Some machines can sequence with a cost of less than 1 cent/Mb

In October 2014, PacBio released an updated chemistry with an average read length 10000-15000 bases and even larger reads up to 40000 bases.

Comparison of sequencing platforms (5/15)

Some compalnies have lost at that game!
The 1000$ genome

Roche will stop the production of 454 machines after 2015.

It will invest in the biodiagnostic potential of Pacific Biosciences machines ($ PACB), which costs $ 75 million.

Comparison of sequencing platforms (6/15)

Some will win!
The 1000$ genome

15/1/2014
Production of 16 Human Genomes / 3 days = 18000 HG / year, 30X coverage, thanks to better packaging of the reactions within microreactors, better recording of the signal through improved camera techniques, cheaper polymerase.
But … you have to purchase 10 such machines = 10 M $
http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530

Comparison of sequencing platforms (7/15)

Platform	Company	Method of Sequencing	Method of Substrate Amplification	Reason of success
454	Roche	Synthesis (clone)	emPCR	1st Next generation Machine, Large read size
Illumina	Illumina	Synthesis	BridgePCR	1st Machine of short reads, less mistakes, larger data production
SOLiD	Life Technologies	Hybridization	emPCR	2nd Machine of short reads, very low error rate
Heliscope	Helicos	Synthesis	No	First single DNA molecule sequencing machine
Ion Torrent	Life Technologies	Synthesis (H+ detection)	emPCR	1st sequencing machine that is not based on light emission, cost < $100.000
PacBio	Pacific Biosciences	Synthesis	No	1st single DNA molecule sequencing machine in real time
Starlight*	Life Technologies	Synthesis	No	single DNA molecule Sequencing machine with quantum dots

Comparison of sequencing platforms (8/15)

Platform	Company	Basic Applications
454	Roche	1,2,3,4,7,8
Illumina	Illumina	1,2,3,4,5,6,7,8
SOLiD	Life Technologies	3,5,6,8
Heliscope	Helicos	5,8
Ion Torrent	Life Technologies	1,2,3,4,8
PacBio	Pacific Biosciences	1,2,3,7,8
Starlight*	Life Technologies	1,2,7,8

1 = de novo sequencing in BACs and bacterial genomes
2 = transcriptome characterization
3 = targeted resequencing
4 = de novo sequencing in eukaryotic genomes,
5 = resequencing and transcript counting
6 = detection of mutations,
7 = metagenomics analyses,
8 = others (ChIP-Seq, µRNA-Seq, Methyl-Seq)

Field guide to next-generation DNA sequencers –Glenn 2011
http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2011.03024.x/pdf

Comparison of sequencing platforms (9/15)

Reducing assembly complexity of microbial genomes with single-molecule sequencing

Koren *et al*. 2013 Genome Biology

http://genomebiology.com/2013/14/9/R101

- First and second generation sequencing technologies cannot always reconstruct a microbial genome.
- During sequencing and assembly of six bacterial genomes of varying complexity, it was revealed that the majority of the genomes of bacteria and archaea, can be assembled without gaps, using a Pacific Biosciences sequencing approach.
- This type of assembly is more precise than typical assemblies with small read length.

Comparison of sequencing platforms (10/15)

• In the following figure we see that low complexity genomes (Class I) may be assembled using any sequencing machine.
• Genomes of medium complexity (Class II) present satisfying assembly results, only when using Pacific Biosciences machines (notice the complete outer circle, that indicates a complete assembly).
• High complexity genomes (Class III), possessing many repeats can not be perfectly assembled with any sequencing machine (even with Pacific Biosciences machines) even though its read length may exceed 7000 bases.

Comparison of sequencing platforms (11/15)

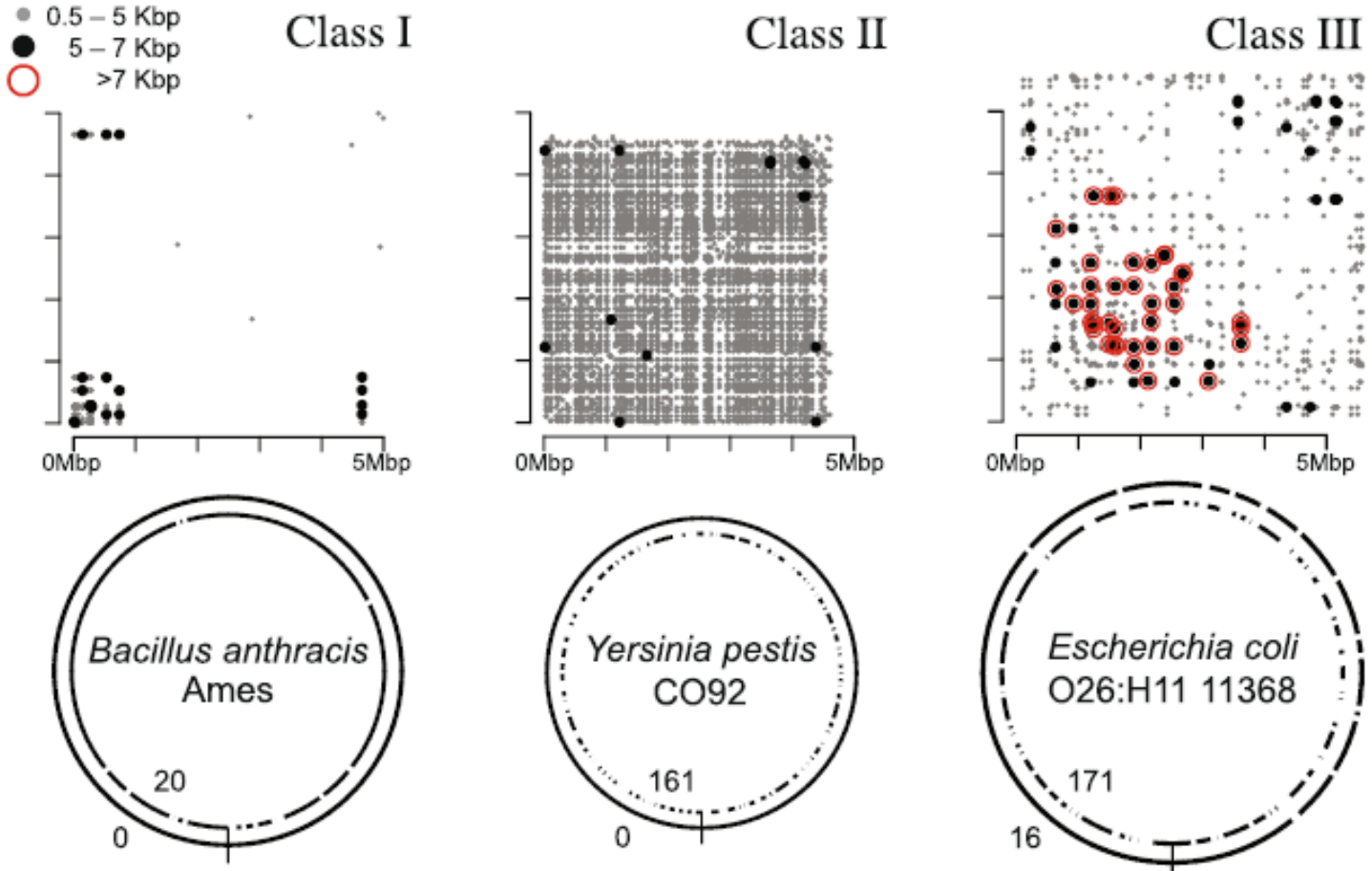


Figure 3: Three classes of assembly complexity of microbial genomes.

Koren *et al*. 2013 Genome Biology, http://creativecommons.org/licenses/by/2.0

Comparison of sequencing platforms (12/15)

…data assembly?

There are many programs
Each one with its own characteristics, algorithms and weaknesses

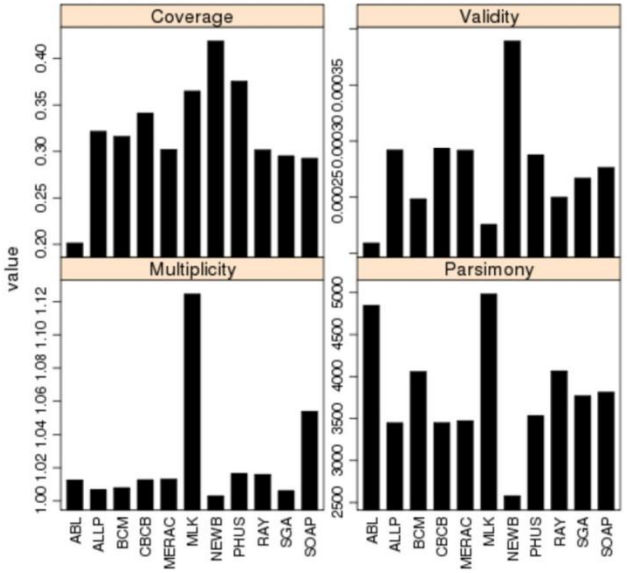
The assembly program is determined based on the characteristics of the studied genome but as well as on the sequencing machine

- SOAP: Short Oligonucleotide Analysis Package (soap.genomics.org.cn/)
- ABySS: Assembly By Short Sequences

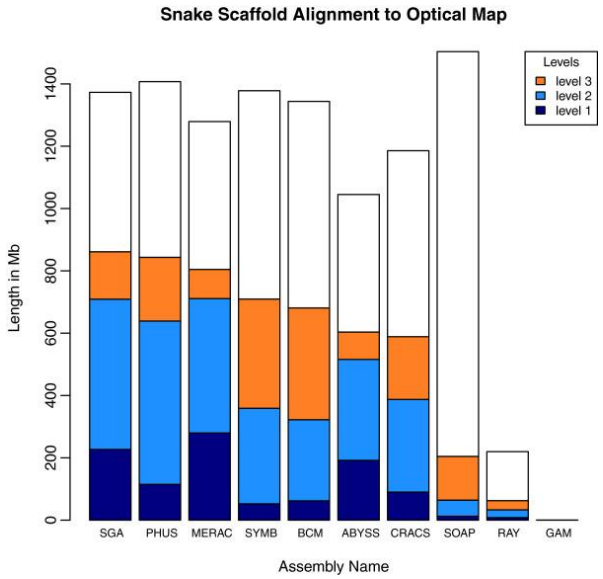
So do you want to be a computational biologist?
http://www.nature.com/nbt/journal/v31/n11/full/nbt.2740.html

Comparison of sequencing platforms (13/15)

Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species: Bradnam *et al*. 2013, Gigascience (http://www.gigasciencejournal.com/content/2/1/10)



Figures 4 (left) and **5** (right) show the results of data assembly in birds (left) and snakes (right).



It seems that there is still much room for improvement in assembly programs.

Comparison of sequencing platforms (14/15)

Insights into the phylogeny and coding potential of microbial dark matter Nature (2013)

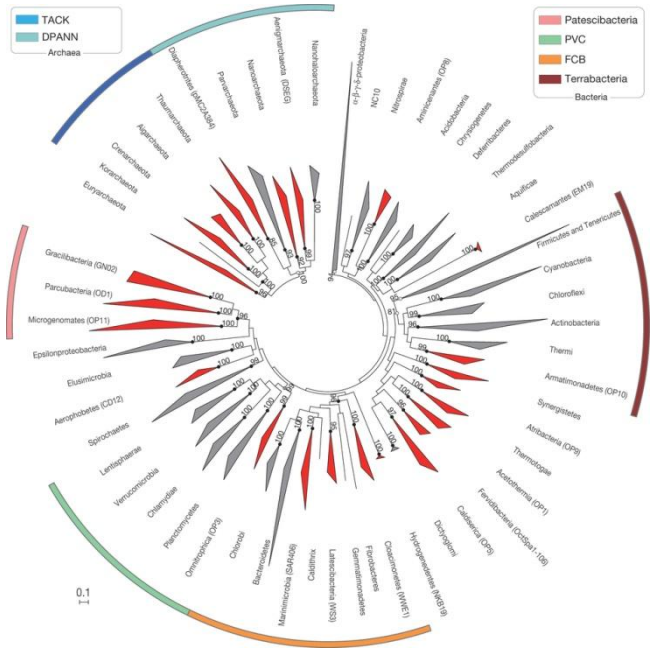


Figure 6: Maximum-likelihood phylogenetic inference of Archaea and Bacteria. http://www.nature.com/nature/journal/v499/n7459/full/nature12352.html, CC-BY-NC-SA-3.0, http://creativecommons.org/licenses/by-nc-sa/3.0/

http://genome.jgi.doe.gov/MDM/

Comparison of sequencing platforms (15/15)

The knoSYS100 is an integrated system for full production, analysis and data storage of human genomes. At a cost of $ 125000, the company's aim is to enable hospital services to give specific information that can interest anyone, and to explain results quickly and accurately.

www.knome.com

Exercise 3

You work in a large genomic center.

You want to sequence a bacterial genome.

You want to sequence the genome of a primate mammal in order to compare its evolution to the human and the chimpanzee genomes.

You want to sequence the genome of the mosquito *Anopheles gambiae* in order to find ways to combat malaria.

Which sequencing technique would you use in each case and why;

Exercise 3 - Solution

Bacterial genomes: You can use any NGS machine. In case that the genome is of high complexity, you may have to use Pacific Biosciences machines.

Primate genome: Since you want to study evolution you need great accuracy. Therefore you need a hierarchical approach.

The genome of the mosquito *Anopheles gambiae.* In this case you are interested in the genes (mainly unique regions) and you are not interested in repetitive DNA. A WGS approach can easily produce the results you need.

Exercise 4

If you had to sequence the genome of the 12ploid beet or the 8ploid commercial strawberry what technique would you use?

Exercise 4 - Solution

You would certainly require a machine, that can produce large reads due to the complexity of the genome and the large areas of repeated DNA (eg perhaps Pacific Biosciences) You might also need to use a hierarchical approach to aid you, with the creation of physical and genetic maps to further help the assembly.

DNA of a single cell

Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. Science 2012
http://bernstein.harvard.edu/research/MALBAC.html

High-performance sequencing technologies rapidly change the ways in which biological questions are answered. The sequencing of DNA of single cells is now possible.

DNA of cancer cells

Clonal evolution in breast cancer revealed by single nucleus genome sequencing
Wang et al 2014 Nature 512, 155-160
The Nuc-Seq approach sequences cells with duplicated genetic material (just before their division), starting from more genetic material and reducing error rates when sequencing.

A defining decade in DNA sequencing
John D McPherson, Nature Methods 11, 1003–1005 (2014)

DNA Arrays

- They consist of a large number of DNA fragments placed on a solid substrate.
- DNA fragments are synthesized, which consist of fragments of genes and are placed at specific locations on a solid substrate.
- A mixture of DNA or RNA fragments is hybridized (labeled with fluorescent dyes or radioactivity) on the DNA array.
- Thus, the expression levels of the genes placed on the array are qualitatively and quantitatively measured.

http://bme240.eng.uci.edu/students/08s/jentel/Diagnose-hereditary-disease.htm

DNA Arrays

DNA macroarrays

- A cDNA library is created, which is transferred to agarose and then onto microtiter wells. Using robotics, up to 20000 clones are collected and fixed on **nylon membranes,** creating an array.
- The DNA mixture, which will be hybridized, is labeled with radioactivity, and the result of hybridization is checked using autoradiography.
- They are relatively inexpensive, and were used in the preliminary stages of control of expression.

http://openi.nlm.nih.gov/detailedresult.php?img=3110109_1471-2229-11-70-9&req=4

DNA Arrays

DNA microarrays

- DNA fragments amplified by PCR are fixed on glass slides / chip.
- A slide accepts 5000-40000 PCR products (e.g. all 6200 yeast genes).
- The mixture of cDNAs and mRNAs which will be analyzed, is labeled with fluorescent dyes and hybridized.
- The level of fluorescence is proportional to gene expression levels.

DNA Arrays

DNA microarrays

- Simultaneous analysis of two different samples of cDNA (or RNA) from one cell type at two different states (e.g., normal / cancer cell) or two different times or in two different organisms (mouse / human) or in two different cell types (brain / gut).
- Each is labeled with a different fluorescent dye. The hybridization levels for each position are measured. The level of fluorescence is proportional to gene expression levels. Conclusions are reached regarding protein expression.

DNA Arrays

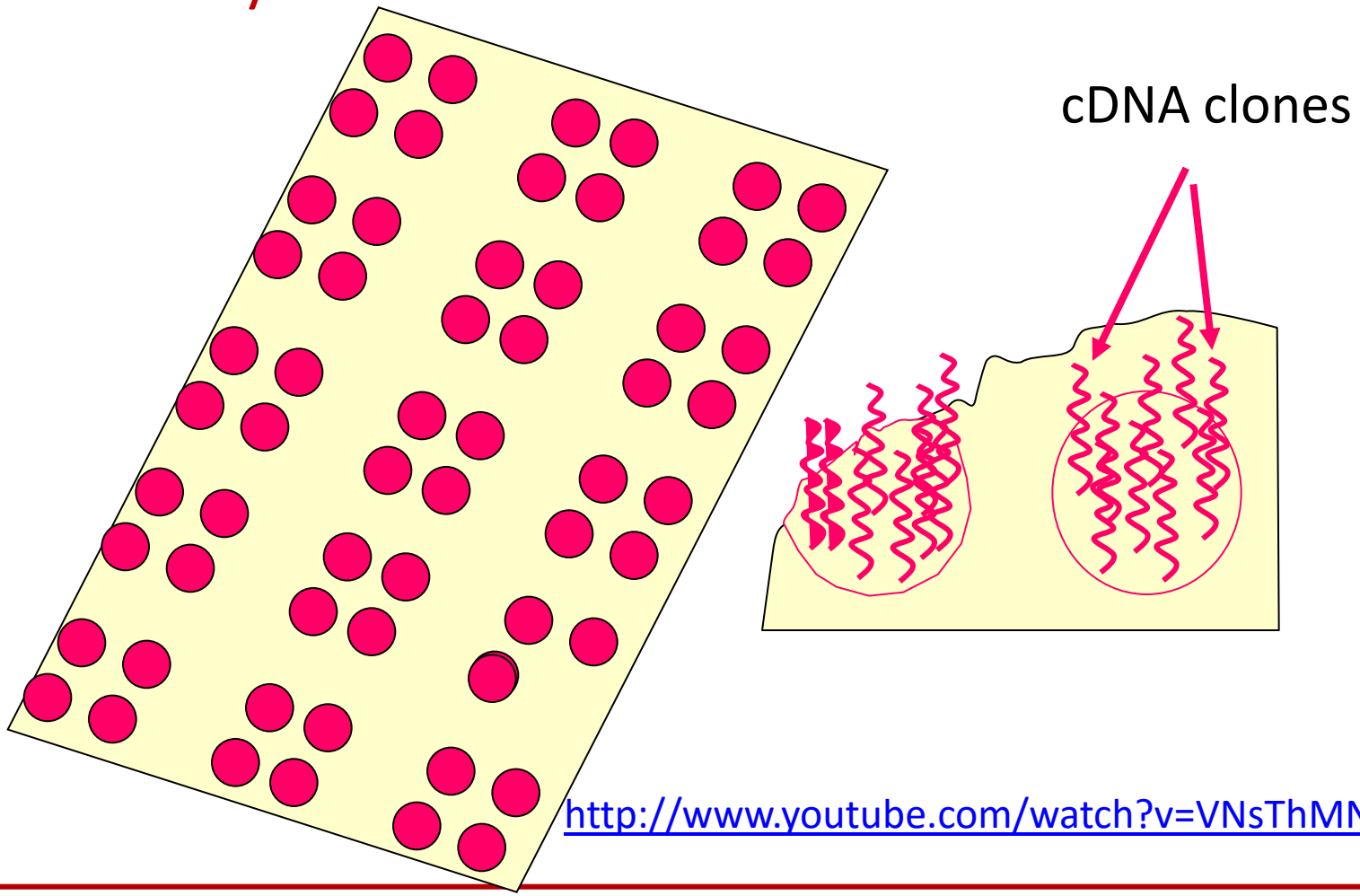
DNA microarrays

Suppose that cDNA samples are labeled with red (for cancer cells) and green (for normal cells) colors respectively, and hybridized on a DNA microarray. When both samples are hybridized at a position, a yellow color is produced. Red indicates mRNA (or cDNA) overexpressed in a cancer cell and green mRNA (or cDNA) overexpressed in normal cells.

http://www.web-books.com/MoBio/Free/Ch9F.htm

DNA Arrays

cDNA microarrays



http://www.youtube.com/watch?v=VNsThMNjKhM

DNA Arrays

cDNA microarrays

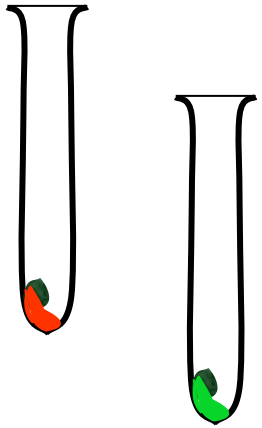
Comparison of expression of two samples

PRINT

cDNA Gene in each position

Samples

cDNA red/green



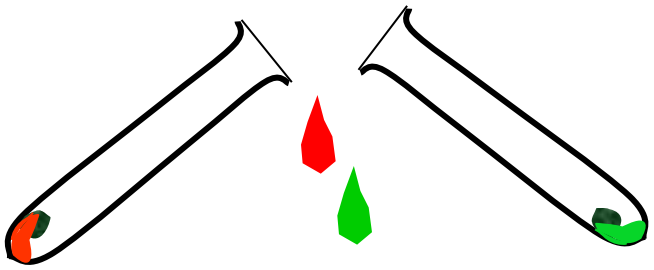
e.g. reference Rna / target Rna

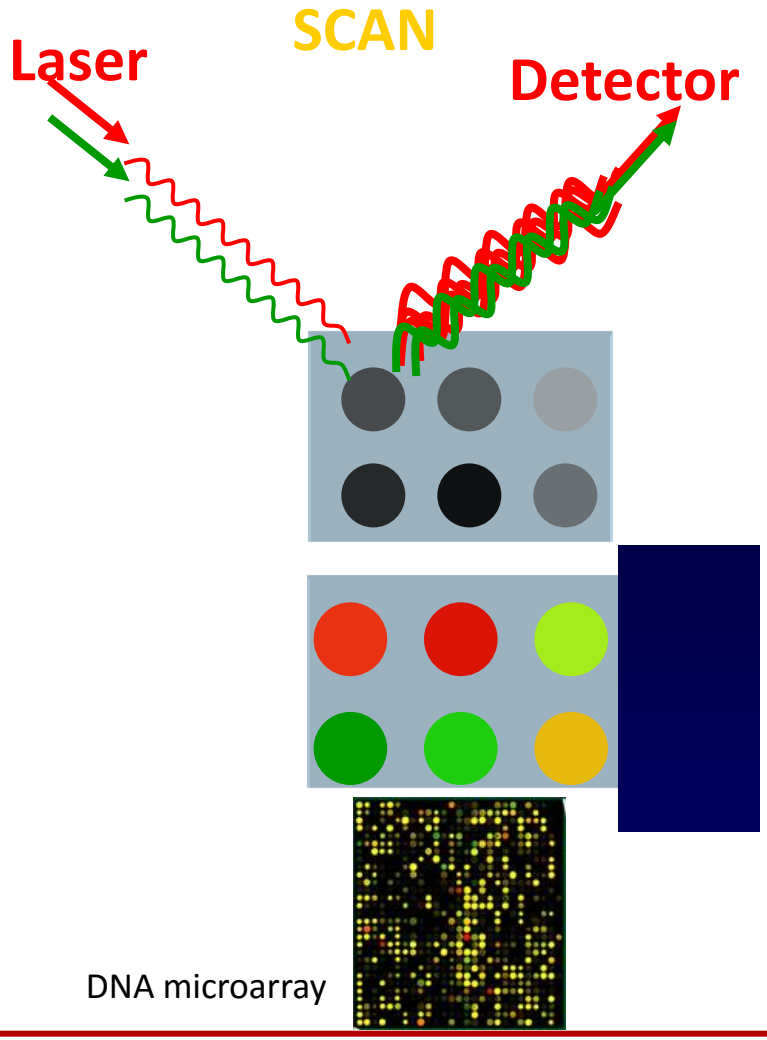
DNA Arrays

cDNA microarrays

Hybridization

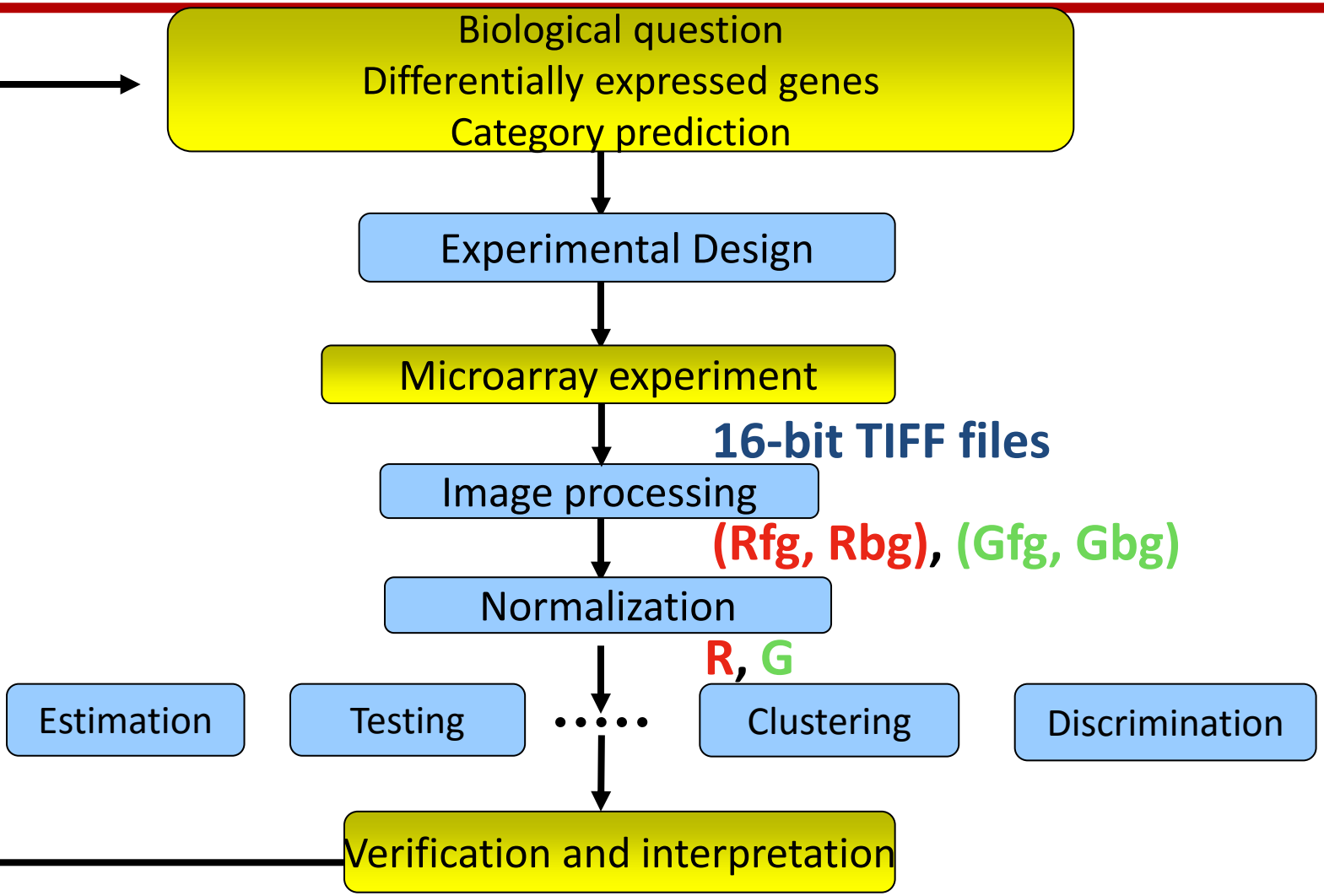
Put equal amounts categorized cDNA samples in microarray





DNA microarray

DNA Arrays


Biological question
Differentially expressed genes
Category prediction
Experimental Design
Microarray experiment
16-bit TIFF files
Image processing
(Rfg, Rbg), (Gfg, Gbg)
Normalization
R, G
Estimation
Testing
Clustering
Discrimination
Verification and interpretation

DNA Arrays

DNA microarrays

A number of known and unknown genes are identified.
e.g. detection of expression level changes for genes related to the conversion of a normal cell into a cancer, or when switching from the primary cancer stages into metastatic stages.
By obtaining the gene expression profile of an individual using DNA microarray, it might become possible to make a diagnosis of a disease stage.
Nevertheless we are still far away from using these machines to find biomarkers (Nature March 2011) –> 90% sensitivity and> 90% specificity are still required and difficult to achieve

DNA Arrays

DNA microarrays

- Cancer diagnosis

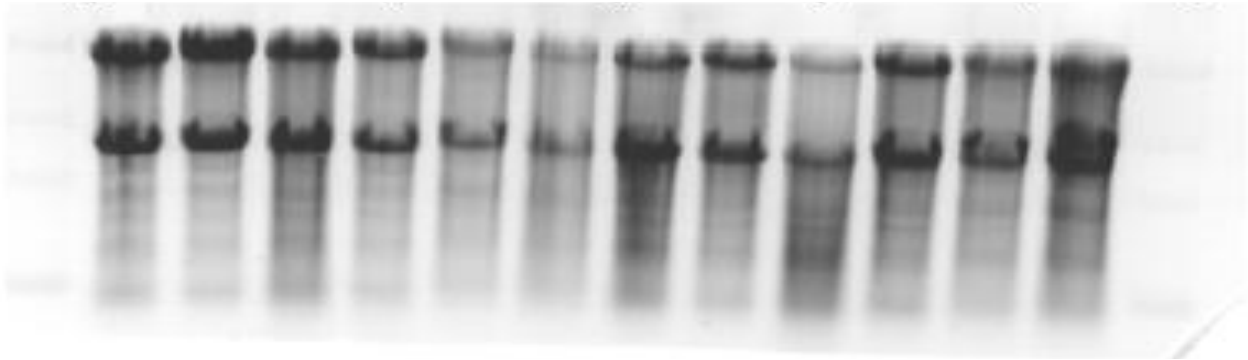
- When studying the response of patients to chemotherapy using microarrays…
- Patients appear to have different gene expression profiles

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0036383

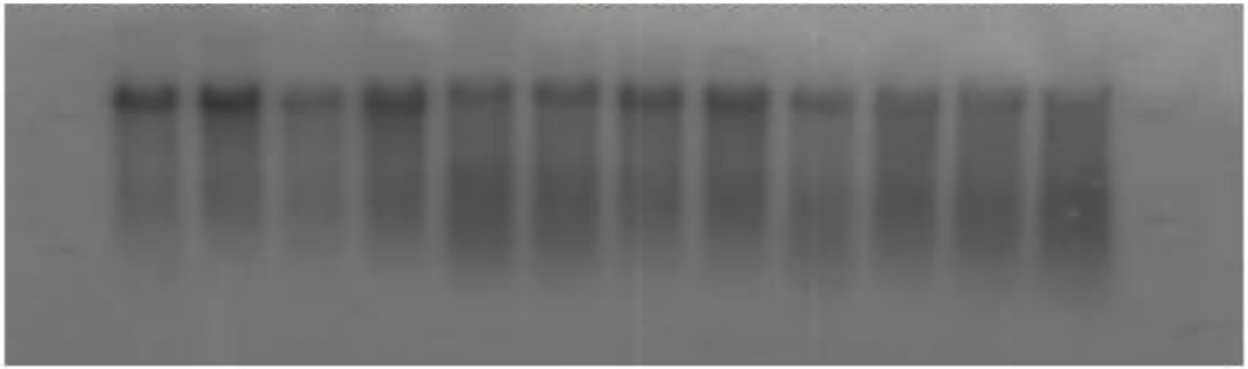
DNA Arrays

In Northern Blot analysis the expression of a single gene only was studied

Figure 7 a): Separation of RNA on agarose gel



b): Detection of specific mRNA by hybridization with labeled probe.



Northern Blot

Measurement of mRNA levels by dot blot and hybridization:
http://ajpgi.physiology.org/content/274/6/G1077 (Fig 11)

Comparison of Arrays/Northern Blot

	Northern	Array
Target	A mixture of RNA molecules separated through gel electrophoresis by size, mounted on a membrane	cDNA molecules (resulting from mRNA using reverse transcriptase) labeled with fluorescent dyes
Probe	Radioactive labeled cDNA in solution	Tens of thousands of cDNA molecules fixed on a chip
Quality	High, Size of studied mRNA is determined	Not so good, size of activated genes is not known
Gene number	One / control	Up to 35,000 genes / control

DNA Arrays

Oligonucleotide arrays

Composition and fixation of at least 100000 oligonucleotides on glass slides of some cm^2 surface.
The oligonucleotides are synthesized artificially and are about 20-60 bp.

Advantages

1) They detect SNP polymorphisms & separate gene products which differ few bps.
2) They do not require the use of libraries or large amounts of PCR products. An oligonucleotide array requires 50-70 bp fragments for each gene.

http://www.cs.tau.ac.il/~rshamir/algmb/00/scribe00/html/lec11/node10.html

DNA Arrays

DNA microarrays

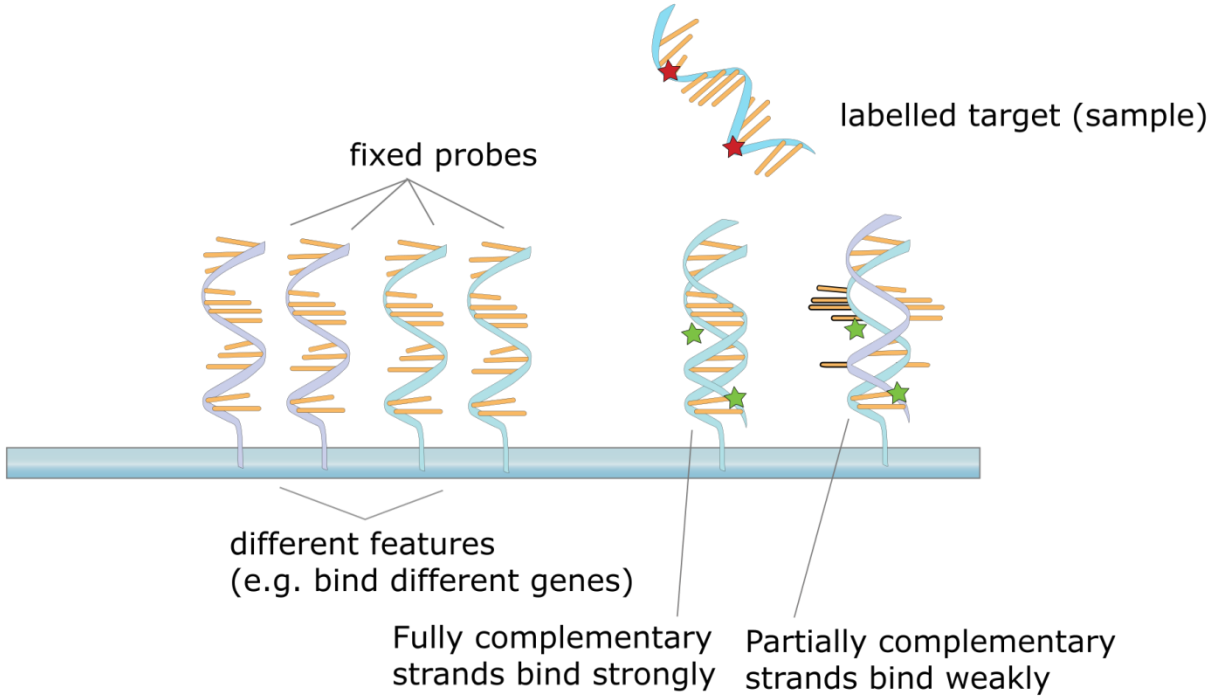


Figure 8: DNA microarray, by Squidonius, https://commons.wikimedia.org/wiki/File:NA_hybrid.svg

DNA Arrays

Oligonucleotide arrays

New techniques: optical fibers (of 1mm diameter) 96 optical fibers per machine.

Up to 50,000 wells / fiber. 50,000 beads are placed there on which oligonucleotides (25 bp in size) are fixed.

The individual analyses that can be executed for SNP polymorphism detection or at the level of protein expression increase dramatically.

50.000 wells X 96 fibers = 4.700.000 analyses!!

DNA Arrays

DNA microarrays

MassARRAY Nanodispenser (Sequenom)

http://bgiamericas.com/service-solutions/genotyping/massarray-genotyping/

384 positions / 8 min

- **6-30 nl**

Advantages:

- High sensitivity
- Cost effective
- Flexible

DNA Arrays

Micro/Oligoarrays DNA

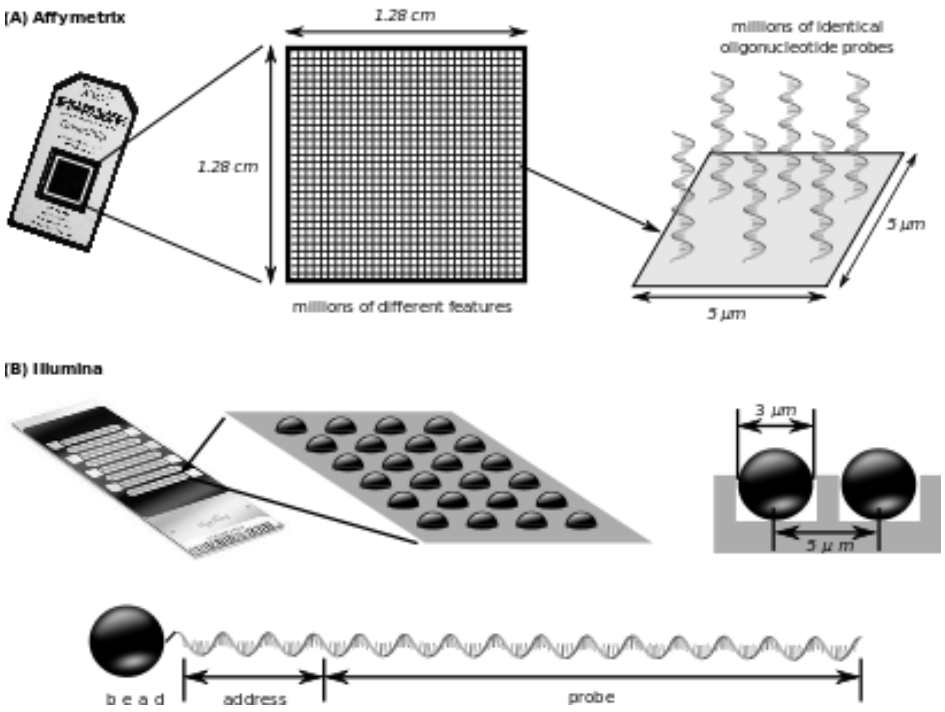


Figure 9: Affymetrix GeneChip and Illumina BeadChip designs.

GeneChip Zebrafish Genome Array: It provides comprehensive coverage of the zebrafish genome and it is an important tool for developmental studies.

www.affymetrix.com

www.agilent.com

www.illumina.com

www.roche.com

DNA Arrays

List of Affymetrix Arrays

Arabidopsis Genome Arrays
B. subtilis Genome Array (Antisense)
Barley Genome Array
Bovine Genome Array
C. elegans Genome Array
Canine Genome Array
Drosophila Genome Arrays
E. coli Genome Arrays
Human Genome Arrays
Mouse Genome Arrays
P. aeruginosa Genome Array
Plasmodium/Anopheles Genome Array
Rat Genome Arrays
S. aureus Genome Array
Soybean Genome Array
Vitis vinifera (Grape) Array
Xenopus laevis Genome Array
Yeast Genome Arrays
Zebrafish Genome Array

DNA Arrays

DNA microarrays - Cost

$100 /chip, 20 chips/experiment= $2000

+ 2000 consumables

+1000 for qPCR

Total $5000

For a paper in a moderately good scientific journal.

Websites with extra info for Microarrays

Microarray methods and reagents:
http://www.genisphere.com/pdf/array50v2_10_19_04.pdf, Array 50 Kit Manual (Genisphere)
DNA sequence databases and related resources:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy, NCBI Taxonomy Browser
http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=r_trout, TIGR rainbow trout gene index
http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=salmon, TIGR Atlantic salmon gene index
http://www.molecularcloning.com, Molecular Cloning, a Laboratory Manual on the Web
http://www.ncbi.nlm.nih.gov, National Center for Biotechnology Information (NCBI)
http://www.phrap.org, PHRAP program for DNA sequence assembly
Minimum Information About a Microarray Experiment guidelines:
http://www.mged.org/Workgroups/MIAME/miame_checklist.html, MIAME Checklist (Brazma et al. 2001)
Microarray data extraction and analysis software:
http://www.biodiscovery.com/index/imagene, ImaGene microarray image processing software (BioDiscovery)
http://www.silicongenetics.com/cgi/SiG.cgi/Products/GeneSpring/features.smf, GeneSpring microarray data analysis software (Agilent Technologies)
http://www.moleculardevices.com/pages/software/gn_genepix_pro.html, GenePix Pro 6.0 microarray image analysis software (Molecular Devices)
http://www.bioconductor.org, Bioconductor 1.8 (open source software for genomic data analysis)
Functional annotation of nucleic acid or protein sequences, and identification of molecular pathways:
http://www.geneontology.org, Gene Ontology Consortium home page
http://ca.expasy.org, Expert Protein Analysis System (ExPASy) proteomics server
http://www.genmapp.org/introduction.asp, GenMAPP Gene Map Annotator and Pathway Profile
Submission of microarray platform, sample, and series information to a public data repository:
http://www.ncbi.nlm.nih.gov/geo, NCBI GEO (Gene Expression Omnibus)

Mass Spectrometer (1/4)

Estimates the mass of molecules (proteins, peptides, oligonucleotides) based on the speed of movement through an electric field (small masses move faster than large).

The main elements are:

- The ion source,
- The mass analyser
- The detector

Mass Spectrometer (2/4)

- The ion source ionizes the molecules.
- The mass analyzer separates ionised sections, according to the mass charge ratio (m/z).
- The detector calculates the separation time.

- Complex computer programs search within protein databases, which protein would give the specific peptides with the specific masses.

Mass Spectrometer (3/4)

- The MS can sequence peptides, breaking them and then calculating the mass of the novel peptides.
- Such a machine is called **Tandem mass spectrometer (MS/MS)**.
- It's a tool for the identification and sequencing of peptides from protein complexes and also for SNP genotyping.

http://www.nature.com/nprot/journal/v1/n4/fig_tab/nprot.2006.257_F4.html

Mass Spectrometer (4/4)

Protein Analysis

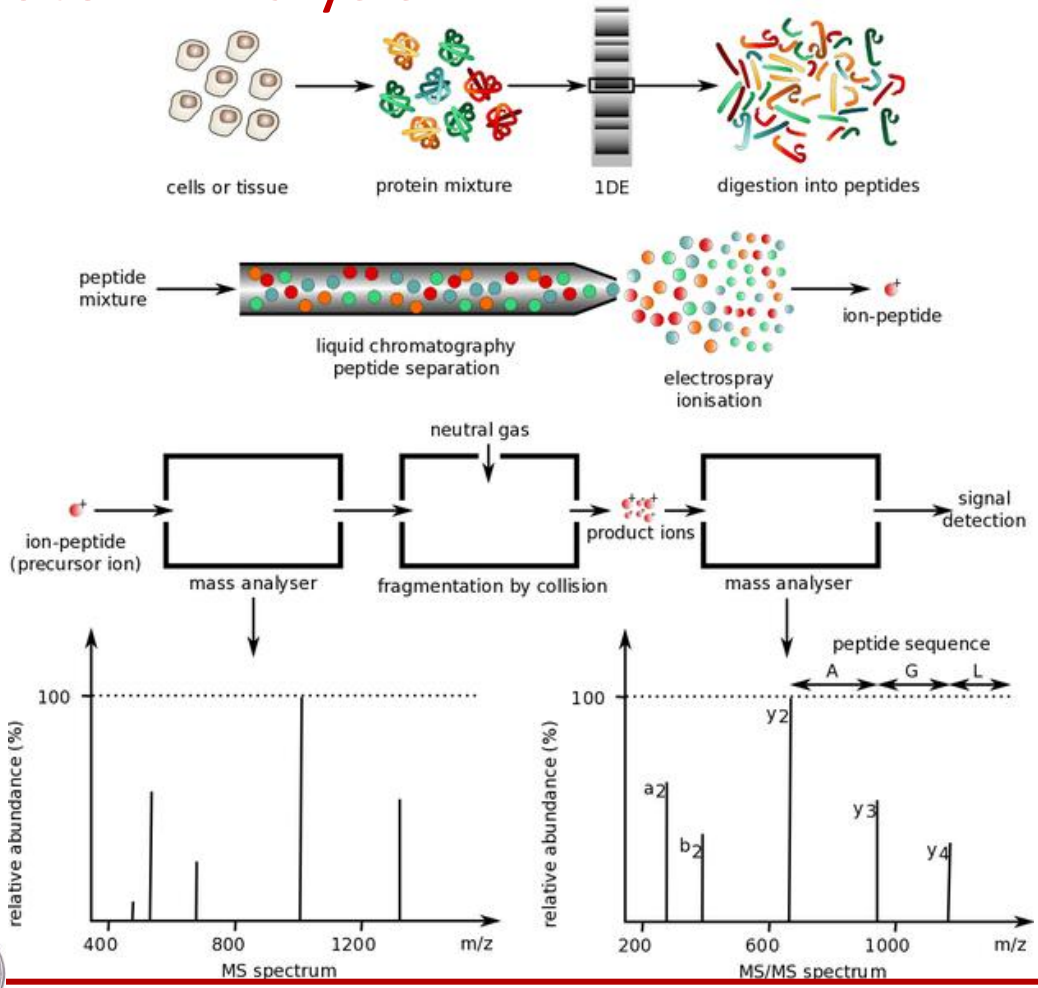


Figure 10: Mass spectrometry protocol.

Note of use of third party works

Sanger sequencing, http://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg, by Estevezj, CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/deed.en)
DNA microarray, http://www.flickr.com/photos/ajc1/2034113679/sizes/o/, by AJC1, CC-BY-NC-2.0 (http://creativecommons.org/licenses/by-nc/2.0/)
Affymetrix GeneChip and Illumina BeadChip designs, http://commons.wikimedia.org/wiki/File:Affymetrix_GeneChip_and_Illumina_BeadChip_designs.svg , by Philippe Hupe, CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/deed.en)
Mass spectrometry protocol, http://en.wikipedia.org/wiki/File:Mass_spectrometry_protocol.png, by Philippe Hupe, CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/deed.en)

Reference note

Licensing note

End of Section

Processing: Minoudi Styliani
Thessaloniki, Winter Semester 2014-2015

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Notes Preservation

Any reproduction or adaptation of the material should include:

- the Reference Note
- the Licence Note
- the Notes Preservation
- Note of use of third party works

accompanied with their hyperlinks.