Overview of DNA Technology Application in Medicine

Ahmad Aizat Abdul Aziz Human Genome Centre School of Medical Sciences

Slide overview

DNA technology application in medicine

- I. Polymerase Chain Reaction (PCR)
- II. Microarray system
- III. DNA sequencing and Next Generation Sequencing (NGS)
- IV. Liquid biopsy









- Polymerase Chain Reaction (PCR) is a biochemical technology in molecular biology to amplify a single or a few copies of a piece of DNA, and generate thousands to millions of copies of DNA sequence.
- Method was developed by Kary Mullis in year 1983



Basic principles of PCR

PCR is a three-step reaction process referred to as a cycle

- *One PCR cycle consists of the following steps:
 - 1. Denaturation
 - 2. Annealing
 - 3. Extension

Usually carry out between 25 and 35 cycles



Types of PCR and Principles

Туре		Principles
PCR-RFLP		Technique that exploits variations in homologous DNA sequences , known as polymorphisms, in order to distinguish individuals, populations, or species or to pinpoint the locations of genes within a sequence using restriction enzyme
RT-PCR	qRT-PCR	Reverse transcription-PCR, or RT-PCR, allows the use of RNA as a template. An additional step allows the detection and amplification of RNA. The RNA is reverse transcribed into complementary DNA (cDNA) , using reverse transcriptase.
qPCR/real-time PCR		Quantitative polymerase chain reaction (Q-PCR) or real time PCR is a method by which the amount of the PCR product can be determined, in real-time. Useful for investigating gene expression .
Multiplex PCR		Technique that used for amplification of multiple targets in a single PCR experiment . In a multiplexing assay, more than one target sequence can be amplified by using multiple primer pairs in a reaction mixture
Allele Specific PCR		Allele-specific polymerase chain reaction (ASPCR) is an application of the polymerase chain reaction (PCR) that permits the direct detection of any point mutation in human DNA.
Nested PCR		Nested PCR is a modification of PCR that was designed to improve sensitivity and specificity . Nested PCR involves the use of two primer sets and two successive PCR reactions .
Touchdown PCR		TD-PCR is a modification of PCR in which the initial annealing temperature is higher than the optimal Tm of the primers and is gradually reduced over subsequent cycles until the Tm temperature or "touchdown temperature" is reached.

PCR Applications

 PCR is now a common and often indispensable technique used in medical and biological research labs for a variety of applications.

 Structural analysis 	•Mapping
 DNA typing 	 Site-directed mutagenesis
 Disease detection 	•Sequencing
•Cloning	 Forensic medicine
 Mutation analysis 	 Scientific research
 Detection of gene expression 	•Pre-natal diagnosis

PCR applications

• Diagnosis of genetic diseases

The use of PCR in diagnosing genetic disease, whether inherited genetic changes or as a result of a spontaneous genetic mutations, is becoming more common. Diseases can be diagnosed even before birth. Examples include: Genetic counseling-screening the parents for genetic diseases before deciding on having children

• Genetic fingerprints

One of the most famous uses for PCR is in the creation of a genetic fingerprint (also known as DNA profiling) from a sample of blood or semen, or from a hair root. Genetic fingerprints are profiles of specific stretches of DNA that vary from person to person. PCR also plays a role in mitochondrial DNA analysis, used for samples from hair shafts and bones when other samples are not available. Genetic analysis based on PCR is also used in paternity testing, and in tissue typing for organ transplantation.

• Detection and diagnosis of infectious diseases

PCR can detect infectious disease before standard serological laboratory tests (tests to detect the presence of antibodies), so allowing treatment to start much earlier. PCR is also useful for screening donated blood for infections, and is especially useful for infections that are difficult to culture in the laboratory, such as tuberculosis.

• Detection of infection in the environment

PCR is used to monitor and track the spread of infectious disease within an animal or human population. PCR can also be used to detect bacterial and viral DNA in the environment, for example looking at pathogens in water supplies.

• Personalized medicine

PCR is used in personalized medicine to select patients for certain treatments, for example in cancer when patients have a genetic change that makes a patient more or less likely to respond to a certain treatment.

• PCR in research

The Human Genome Project used PCR. PCR can be used in analysis of gene expression, for example looking at levels of expression and when genes are switched on and off in physiological processes, including in health and disease.



• What are the differences between PCR, RT-PCR, qPCR, and RT-qPCR?

http://www.enzolifesciences.com/science-center/technotes/2017/march/what-are-the-differences-between-pcr-rt-pcr-qpcr-and-rt-qpcr?/

DNA Microarray

- Genetic disorders can be screened using a DNA microarray - a collection DNA spots / probes (representing genes) embedded on a solid matrix
- A single microarray can contain 10,000 or more probes and can hold representative fragments from the entire genome



Microarray platforms

- Codelink
- Affymetrix
- Agilent
- NimbleGen
- ABI
- Febit

Based on:

- 1. Array format
- 2. Starting RNA requirement
- 3. Amount cDNA hybridized per array
- 4. Hybridized time
- 5. Hybridized temperature
- 6. Detection method
- 7. Sensitivity

Advantages and disadvantages

Table 1. Micro	array platforms co	mparison.					
				Platforms			
	CodeLink	Affymetrix	Agilent	NimbleGen	ABI	Febit	Spotted cDNA arrays
Array format	30-mer	25-mer	60-mer	24-mer	60-mer	25- to 30-mer	
Starting RNA requirement	200 ng–2 µg total RNA	5 µg total RNA	Fluorescent linear amplification: 5 μg total RNA, 200 ng polyA+ RNA Direct labeling: 10 μg total RNA, 200 ng polyA+ RNA Low RNA input: 50 ng total RNA	15 µg total RNA or 3 µg polyA⁺ RNA	0.5 µg for RT-IVT labeling; 40 µg for RT labeling	5 μg total RNA	Fluorescent linear amplification: 5 μg total RNA, 200 ng polyA+ RNA Direct labeling: 10 μg total RNA, 200 ng polyA+ RNA Indirect labeling: 10 μg total RNA, 200 ng polyA+ RNA
Amount cRNA/cDNA hybridized per array	10 μg cRNA	15 μg cRNA	0.75 µg Cy3 cRNA and 0.75 µg Cy5 cRNA per 22,000 array Combined Cy3 and Cy5 fluorescent cDNA derived from direct labeling Combined Cy3 and Cy5 fluorescent cRNA derived from low-input protocol	12 μg cRNA	10 µg cRNA or cRNA	15 µg cRNA	0.75 µg Cy3 cRNA and 0.75 µg Cy5 cRNA per 20,000 array Combined Cy3 and Cy5 fluorescent cDNA derived from direct and indirect labeling
Hybridization time	18 h	16 h	17 h	16 h	16 h	16 h	17 h
Hybridization temperature	37°C	45°C	60°C	45°C	55°C	45°C	60°C
Detection method	Streptavidin-Alexa Fluor® 647	Streptavidin- phycoerythrin	Cy3 and Cy5	Streptavidin-Cy3	Digoxygenin	Streptavidin- phycoerythrin	Cy3 and Cy5
Secondary detection method	None	Biotinylated anti-streptavidin	None	None	Anti-digoxygenin/ alkaline phosphatase chemiluminiscence	None	None
Scanner/ analyzer	GenePix [®] 4000B; Affymetrix 428™; ArrayWoRx [®] ; GeneTAC™; Agilent G2565BA	Agilent GeneArray® scanner	Agilent G2565BA	GenePix 4000B or alternative scanner with 5 µm scanning resolution	Applied Biosystems 1700 Chemiluminescent Microarray Analyzer	Geniom [®] one	GenePix 4000B; Affymetrix 428; ArrayWoRx; GeneTAC; Agilent G2565BA

3D: Three-dimensional; ABI: Applied Biosystems; Cy: Cyanine; IVT: In vitro transcription; RT: Reverse transcription.

Retrieved from Hardiman, 2004

continued

				Platforms			
	CodeLink	Affymetrix	Agilent	NimbleGen	ABI	Febit	Spotted cDNA arrays
lmage acquisition and data analysis	CodeLink [™] Scanning CodeLink Expression V4.0 analysis Raw and normalized data in Excel sheet Export to Rosetta Resolver [®] , GeneSpring [®] , Genesight [™] , Spotfire [®] , or dChip	GeneChip analysis suite Normalized average difference cells in Excel sheet	Agilent feature extraction software Rosetta Resolver gene expression data analysis system Rosetta luminator gene expression data analysis system	NimbleScan R statistical package and Spotfire decision site	Applied Biosystems 1700 chemiluminiscent microarray analyzer resident software Export to GeneSpring, Spotfire or as a text file	Geniom® one software	Axon GenePics, Biodiscovery Imagene, Agilent feature extraction software Export to Rosetta Resolver, GeneSpring, Genesight, Spotfire
Sensitivity	1:900,000	1:100,000	1:1,000,000			1:100,000	1:300,000
Advantages	Sensitivity; 3D surface; liquid hybridization kinetics; can be utilized with any microarray scanner; customization is possible	Reproducibility; content; mature platform; customization	Reproducibility; content; mature platform; sensitivity; customization	Customization	Sensitivity; less background with chemiluminescence	Customization	Inexpensive; can be utilized with any microarray scanner; customization is possible
Disadvantages	Non-contact printing – printing- related issue, such as poor spot morphology	Short oligonucleotides – less sensitive	Two-color dye bias and ozone-related degradation	Short oligonucleotides – less sensitive	Currently only available for human studies	Short oligonucleotides – less specific	cDNAs – poor specificity; two-color dye bias and ozone-related degradation; poor reproducibility; reporter feature identity errors

3D: Three-dimensional; ABI: Applied Biosystems; Cy: Cyanine; IVT: In vitro transcription; RT: Reverse transcription.

Retrieved from Hardiman, 2004

Type of microarray and its application

Application or technology	Synopsis
Gene expression profiling	In an mRNA or gene expression profiling experiment the expression levels of thousands of genes are simultaneously monitored to study the effects of certain treatments , diseases , and developmental stages on gene expression . For example, microarray-based gene expression profiling can be used to identify genes whose expression is changed in response to pathogens or other organisms by comparing gene expression in infected to that in uninfected cells or tissues.
Comparative genomic hybridization (Array CGH)	Comparative genomic hybridization is a molecular cytogenetic method for analysing copy number variations (CNVs) relative to ploidy level in the DNA of a test sample compared to a reference sample, without the need for culturing cells. The aim of this technique is to quickly and efficiently compare two genomic DNA samples arising from two sources, which are most often closely related, because it is suspected that they contain differences in terms of either gains or losses of either whole chromosomes or subchromosomal regions (a portion of a whole chromosome).
SNP detection	Identifying single nucleotide polymorphism among alleles within or between populations. Several applications of microarrays make use of SNP detection, including genotyping, forensic analysis, measuring predisposition to disease, identifying drug-candidates, evaluating germline mutations in individuals or somatic mutations in cancers, assessing loss of heterozygosity, or genetic linkage analysis.
Fusion genes microarray	A Fusion gene microarray can detect fusion transcripts , e.g. from cancer specimens. The principle behind this is building on the alternative splicing microarrays. The oligo design strategy enables combined measurements of chimeric transcript junctions with exon-wise measurements of individual fusion partners. Ex; BCR-ABL fusion gene

Applications

- Gene expression profiling
- Differential expression analysis
- Diagnostics (SNPs detection, deletion and duplication)
- Cellular profiling
- Chromosomal aberrations
- Genomic abnormalities in cancer
- Prenatal genetic diagnosis

ADVANTAGES

- Provides data for thousands of genes.
- One experiment instead of many.
- Fast and easy to obtain results.
- Different parts of DNA can be used to study gene expression.

Disadvantages:

- The biggest disadvantage of DNA chips is that they are **expensive** to create.
- The production of too many results at a time requires long time for analysis, which is quite complex in nature.
- The DNA chips **do not have very long shelf life**, which proves to be another major disadvantage of the technology.

•Identify gene expression of only those who already reported.

DNA Sequencing

- DNA sequencing is the method to determine the sequence/order of nucleic acid A (adenine),T (Thymine),G (Guanine) and C (Cytosine)
- Two basic methods for DNA sequencing:
 - Maxam-Gilbelt Sequencing

Based on chemical modification of DNA and subsequent cleavage at specific bases

Sanger Sequencing

Based on the selection incorporation of chain-terminating dideoxynucleosides by DNA polymerase during in vitro DNA replication

Basic principle of Maxam-Gilbert Sequencing

- Developed by Allan Maxam and Walter Gilbert.
- This method is based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified nucleotides.
 - Chemical modification of DNA; radioactive labelling at one 5' end of the DNA
 - Purification of the DNA
 - Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T).
 - Run on the gel



Basic principle of Sanger DNA Sequencing

Components: Primer, DNA template, DNA polymerase, dNTPs (A,T,C,G) and ddNTPs (ddA, ddT, ddC or ddG).

Sanger DNA sequencing steps

- DNA-primer mixture is put into separate tube 1. with DNA polymerase, dNTPs (A,T,G,C) and ddNTPs (ddA, ddT, ddC or ddG).
- Next, ddNTPs are into 4 difference tubes, resulting in different lengths of fragments
- Fragments produced form each of reactions 3. separated by gel electrophoresis
- The fragments are to be read from bottom to top and this represents the complementary sequence of the original strand of DNA

*TCAGTAAT *TCAGTAA

*TCAGTA *TCAGT *TCAG

*тса *тс

*т

https://www.youtube.com/watch?v=FvHRio1yyhQ



Method Comparisons

Sanger

- Enzymatic
- Requires DNA synthesis
- Termination of chain elongation
- Single stranded DNA

Maxam Gilbert

- Chemical
- Requires DNA
- Breaks DNA at different nucleotides
- Double stranded or single stranded DNA

Uses and limitations

- High quality sequence for relatively long stretches of DNA (up to 900 bp)
- Typically used to sequence individual pieces of DNA (bacteria plasmid or DNA copied in PCR) results validation
- Expensive and inefficient for large-scale project (whole genome sequencing)-New large-scale sequencing techniques are needed such as Next Generation Sequencing (NGS).

High throughput DNA sequencing method

- Next generation sequencing (NGS), massively parallel or deep sequencing are related terms that describe a DNA sequencing technology which has revolutionized genomic research.
- Using NGS an entire human genome can be sequenced within a single day.
- In contrast, the previous Sanger sequencing technology, used to decipher the human genome, required over a decade to deliver the final draft



Overview of selected Next generation Sequencing

- Illumina
- Life technologies/ThermoFisher/Ion Torrent
- Pacific Biosciences
- Oxford Nanopore Technologies

Advantages & Disadvantages

Table II. Comparing performance characteristics of various next generation sequencing (NGS) platforms						
Properties	Roche (454)	Illumina	ABI SOLiD	ION PGM	Heliscope	Pacific Biosciences
Sequencing chemistry	Pyrosequencing	Sequencing by synthesis	Ligation based sequencing	Semiconductor sequencing	Single molecule approach	Real time single molecule approach
Accuracy (%)	99	99.9	99.94	99	99.5	99.999
Millions of reads/run	1	3.4	>700	0.10	800	0.01
Run time ^a	10 h	26 h	$\sim 1-2 \text{ wk}$	~2 h	~1 day	~2 h
^a Run time depends on the length of the genome to be sequenced. Here, it is considered to be the length of the human genome. <i>Source</i> : Refs 74-78						

Retrieved from Desikan and Narayanan, 2015)

Table III. Advantages and disadvantages of the various next generation sequencing (NGS) platforms				
NGS platform	Primary advantages	Primary disadvantages		
Roche (454)	Produces maximum read length.	Very expensive – high cost per Mb		
Illumina	Versatile instrument and scalable in future.	Relatively few reads and higher cost / Mb		
ABI SOLID TM	Low cost of the instrument and high accuracy.	Takes a long time for sequencing and cost per read is high.		
ION PGM	Simple and a low cost instrument can be easily upgraded.	New platform with low accuracy.		
Heliscope™	Single molecule approach produces large number of reads.	Accuracy and longevity of this approach remain questionable.		
Pacific Biosciences	Real time sequencing through single molecule approach produces large number of reads; the run time is very low thereby enabling sequencing of large number of samples.	High capital cost.		
Mb, mega base pair <i>Source</i> : Refs 74,75,78	;			

Retrieved from Desikan and Narayanan, 2015)

Application of NGS in cancer research and clinical application



Retrieved from Shyr and Liu, 2013)

Active cancer studies using NGS as primary outcome measure

Study Title/ <i>Sponsor</i>	NCT#/# Enrolled/Start Date	Condition	Description	Sequencing Technologies
Tumor Specific Plasma DNA in Breast Cancer/Dartmouth- Hitchcock Medical Center	NCT01617915/6/October 2012	Breast Cancer	Analyze chromosomal rearrangements and genomic alterations	Whole genome sequencing
Whole Exon Sequencing of Down Syndrome Acute Myeloid Leukemia/ <i>Children's Oncology Group</i>	NCT01507441/10 /February 2012	Leukemia	Examine DNA samples of patients with Leukemia and Down Syndrome and identify DNA alterations	Whole exome Sequencing
Studying Genes in Samples From Younger Patients with Adrenocortical Tumor/ <i>Children's Oncology Group</i>	NCT01528956/10 /February 2012	Adrenocortical Carcinoma	Study genes from patients with adrenocortical tumor	Whole genome Sequencing
Feasibility Clinical Study of Targeted and Genome-Wide Sequencing/ <i>University Health Network, Toronto</i>	NCT01345513/150/March 2011	Solid Tumors	Identify gene mutations in cancer patients	Whole genome sequencing
An Ancillary Pilot Trial Using Whole Genome Sequencing in Patients with Advance Refractor Cancer/ <i>Scottsdale</i> <i>Healthcare</i>	NCT01443390/10 /September 2011	Advanced Cancer	Investigate patients with cancer that are using Phase I drugs and its effect on the patient	Whole genome Sequencing
Cancer Genome Analysis/ <i>Seoul National University</i> Hospital	NCT01458604/100 /August 2011	Malignant Tumor	Identify and analyze genetic alterations in tumors for therapeutic agents	Targeted Sequencing, whole exome sequencing and RNA- seq
RNA Biomarkers in Tissue Samples From Infants with Acute Meyloid Leukemia/ <i>Children's Oncology Group</i>	NCT01229124/20/October 2010	Leukemia	Analyze tissue samples and identify biomarkers from RNA	RNA-seq
Molecular Analysis of Solid Tumors/ <i>St. Jude Children's</i> Research Hospital	NCT01050296/360 /January 2010	Pediatric Solid Tumors	Analyze gene expression profiles of tumor and examine genetic alterations	Whole genome Sequencing
Deep Sequencing of the Breast Cancer Transcriptome/ <i>University of Arkansas</i>	NCT01141530/30/Sept 2009	Breast Cancer	Examine transcriptional regulation and triple negative breast cancer	RNA-seq

Table Z. Examples of consolita-based Floject	Table 2.	Examples of	Consortia-Based	Projects
--	----------	-------------	------------------------	-----------------

Initiative	Purpose	Website
1000 Genomes Project	Cataloging normal variation in diverse human populations.	www.1000genomes.org
The Encyclopedia of DNA Elements	Identifying functional genomic elements in the human genome.	www.encodeproject.org
Roadmap Epigenomics Project	Catalogue human epigenomic data with the goal of advancing basic biology and disease-oriented research.	www.roadmapepigenomics.org
Human Microbiome Project	Comprehensive characterization of the human microbiome and analysis of its role in human health and disease.	www.hmpdacc.org
Genotype-Tissue Expression Program	Characterizing gene expression and regulation in many human tissues and correlating with genetic variation and disease.	www.commonfund.nih.gov/GTEx/ index
Human Immunology Project Consortium	Characterizing the diverse states of the human immune system following infection, vaccination or treatment.	http://www.immuneprofiling.org
Grand Opportunity Exome Sequencing Project	Discovery of novel genes and mechanisms contributing to heart, lung and blood disorders.	https://esp.gs.washington.edu/drupal
The Cancer Genome Atlas	Understanding the molecular basis of cancer.	www.cancergenome.nih.gov
International Cancer Genome Consortium	Describing the genomic, transcriptomic and epigenomic changes in 50 different tumor types.	www.icgc.org
Clinical Sequencing Exploratory Research Program	Develop methods as well as the legal and ethical frameworks necessary to integrate sequencing into the clinic.	www.genome.gov/27546194
Centers for Mendelian Genomics	Discovering the genes and genetic variants underlying human Mendelian disorders.	www.mendelian.org
Undiagnosed Diseases Network	Promoting the use of genomic data to elucidate the mechanisms underlying the diseases of unknown etiology.	www.commonfund.nih.gov/Diseases/ index
Newborn Sequencing in Genomic Medicine and Public Health	Exploring the challenges and opportunities associated with using genomic sequence information in the newborn period.	www.genome.gov/27558493
The Pediatric Cardiac Genomics Consortium	Determining the genes responsible for congenital heart disease.	www.benchtobassinet.com
Alzheimer's Disease Sequencing Project	Identifying genes contributing to risk of developing Alzheimer's disease in multiethnic populations.	www.niagads.org/adsp

Retrieved from Reuter et al., 2015)

Liquid Biopsy

Fluid biopsy or fluid phase biopsy, is the sampling and analysis of nonsolid biological tissue, primarily blood

Types of liquid biopsy

- circulating tumor cells (CTCs)
- circulating tumor DNA (ctDNA)
- cell-free RNA (cfRNA)
- Exosomes (ex: miRNA)



Figure 1. Molecular applications of circulating tumor cells (CTCs), circulating tumor DNA (ctDNA) and exosomes as liquid biopsy for personalized medicine.



Time-Intensive Procedure Localized Sampling of Tissue Not Easily Obtained Some Pain/Risk Invasive

Liquid Biopsy



Quick Comprehensive Tissue Profile Easily Obtained Minimal Pain/Risk Minimally Invasive

How it works:

- ctDNA is found in serum and plasma fractions from blood.
- The mechanism of ctDNA release is unknown, though apoptosis, necrosis, and active secretion from tumor cells have been hypothesized.

As tumors increase in volume, the capacity of phagocytes to eliminate and clear apoptotic and necrotic fragments can be exceeded, leading to passive release of ctDNA into the bloodstream

- 1. Cancer mutations in ctDNA mirror those found in traditional tumor biopsies, which allows them to be used as molecular biomarkers to track the disease
- 2. Scientists can purify and then analyze ctDNA using next-generation sequencing (NGS) or PCR-based methods such as digital PCR

Table 1. Potential applications of ctDNA in cancer diagnostics and management.

Application	Summary	Potential for clinical utility	References
Prognosis	 Undetectable ctDNA in the bloodstream after surgery are correlated with improved prognosis and smaller chances of relapse Strength and type of chemotherapy can be informed by ctDNA analysis showing l ikelihood of relapse 	Excellent	13–31, 70, 88, 89
Monitoring treatment efficacy and early relapse detection	 Increased number of mutations or rising ctDNA concentration can indicate treatment failure, resistance, and relapse 	Excellent, maybe in combination with protein markers	13–20, 23–29
Selection of treatment	 Knowledge of mutations in the ctDNA informs choice of therapy (personalized treatments) ctDNA avoids tumor heterogeneity issues by providing overview of all the mutations in the tumor (assuming all tumor cells secrete DNA at the same rate) 	Excellent	17, 28, 30–33, 73
Tumor size/disease burden	 Larger amount of ctDNA in blood denotes advanced tumor stage and volume Blood testing does not carry the risk of radiation exposure or poor accuracy of imaging and can be repeated more often than traditional biopsies 	Excellent, especially in combination with imaging	15, 17, 18, 20–29, 88, 89
Detection of cancer in asymptomatic individuals/population screening	 Most studies show poor sensitivity, especially for early-stage tumors For small tumors, there is not enough ctDNA present to allow for an accurate test result Threshold appears to be 1-cm-diameter tumors 	Under intense investigation; could be used in combination with protein biomarkers	13, 30, 70, 88, 89

THANK YOU

HUMAN GENOME PROJECT

- · The biggest challenge for the life sciences
- 15 years project (NIH, DOE of USA)
- Primary goal ⇒ Sequence base pairs of human beings that form DNA
- Identifying & mapping approx. 20K-25K genes
- Significance → Physical & functional
 standpoint

ADVANTAGES

- · Improved diagnosis of disease
- Identify the genes causing genetic diseases
- Identifying crime suspects

DISADVANTAGES

- · Whole genome cannot be sequenced at once
- Very slow and time consuming

The project formally launched in 1990 and was declared complete on April 14, 2003