

Microarray Technology

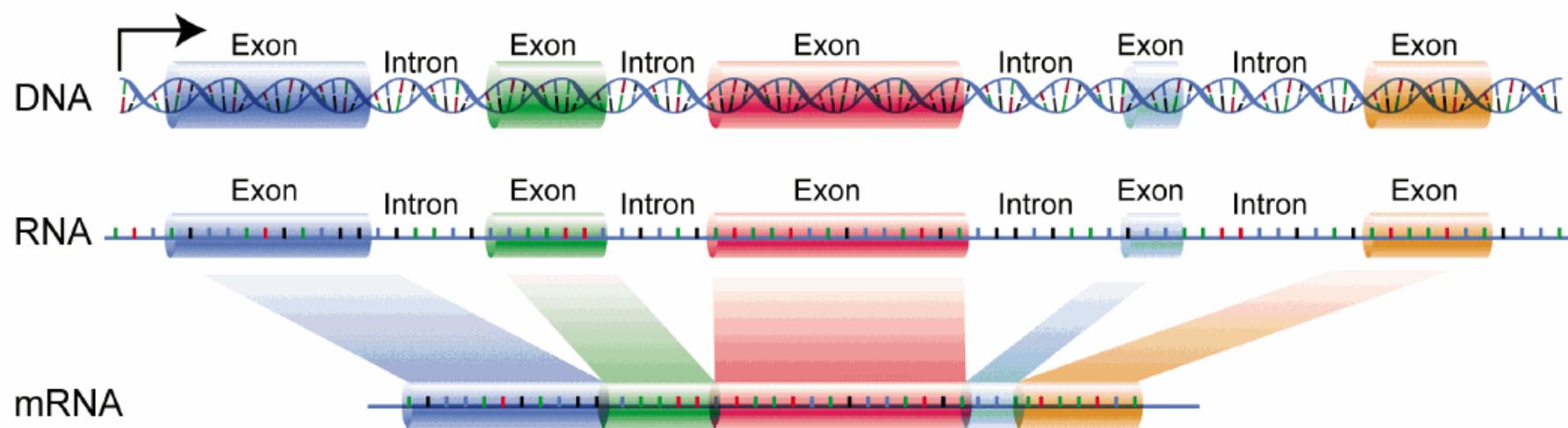
Mikhail Dozmorov, Ph.D.,
Arthritis & Clinical Immunology Department
Oklahoma Medical Research Foundation
Mikhail-Dozmorov@omrf.org
January 16, 2013

Agenda

- ❖ General microarray principles
- ❖ Single-channel Affymetrix and Illumina microarrays
- ❖ Exon arrays
- ❖ RNA-sequencing

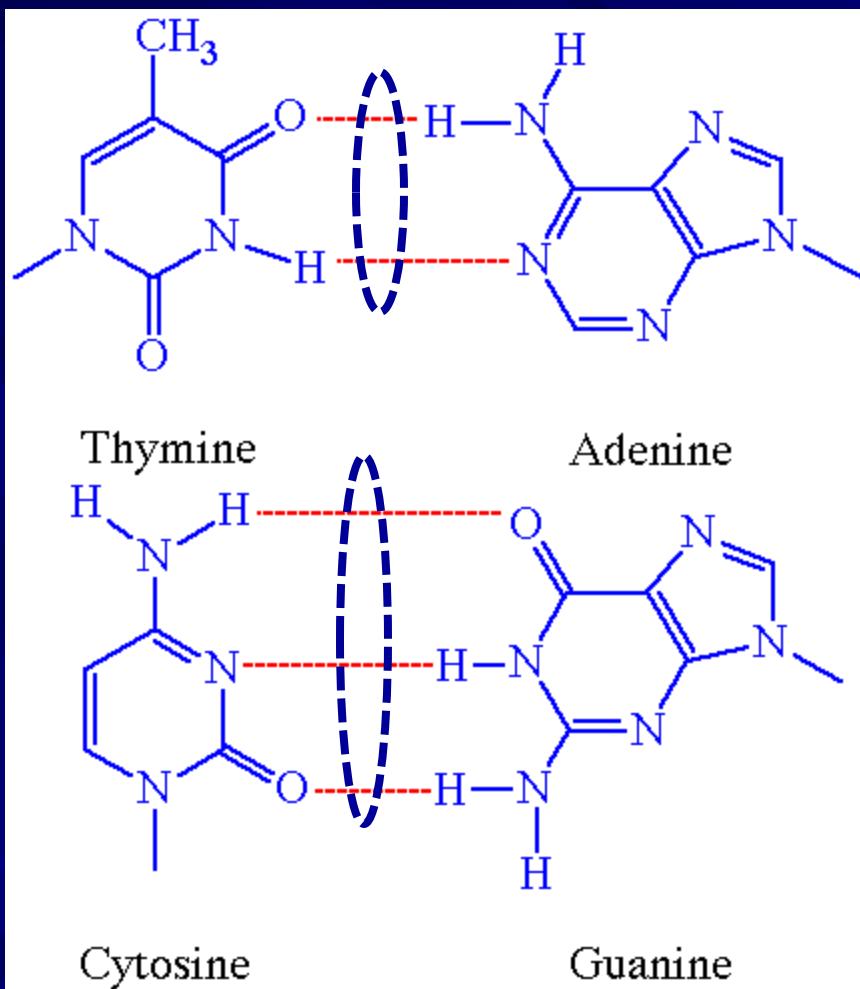
General Microarray Principles

Gene structure in eukaryotes



- ❖ Exon – Expressed regiON, coding part
- ❖ Intron – INT^{er}genic regiON, non-coding part
- ❖ mRNA – messenger RNA, carrier of genetic information.

Rules of base pairing



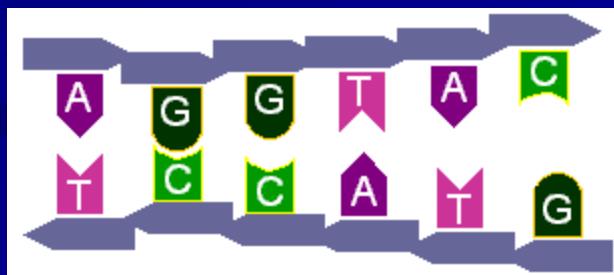
- All genetic code is spelled out with just four chemical letters, or bases: adenine (A), thymine (T), cytosine (C) and guanine (G).
- These pair up, A with T and C with G.
- The human genome has between 2.8 and 3.5 billion base pairs.

Complementary hybridization

- A-T, C-G

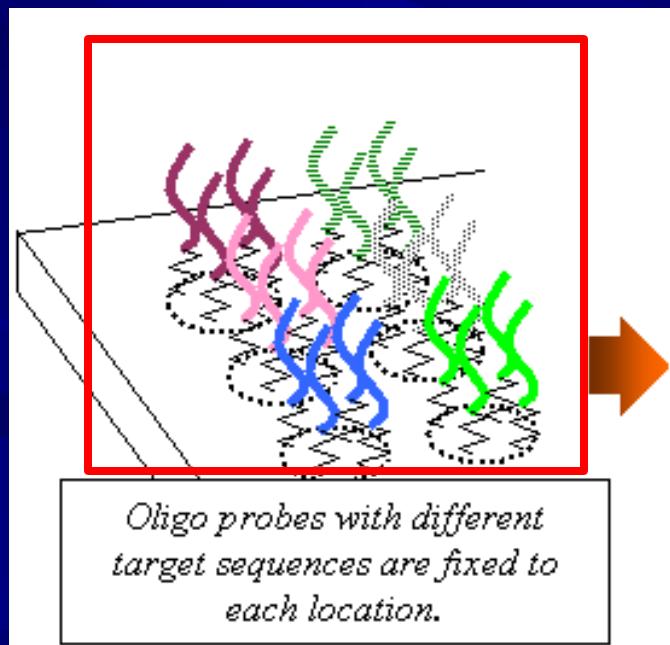


- Sequence fully complementary to a target will hybridize with much higher efficiency than partially complementary.



Expression of different genes

- An oligonucleotide (from Greek prefix oligo-, "having few, having little") is a short nucleic acid polymer.

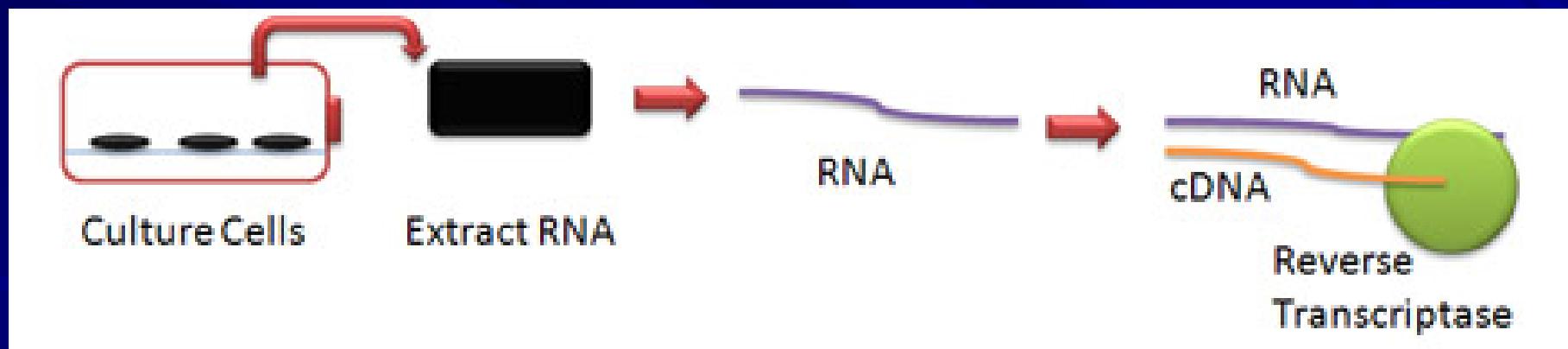


Type of “bait” on the array determines the fish caught

- Long (70) vs. short (25mer) oligos.
 - Specificity higher for long oligos.
 - Sophisticated algorithms should be used to design short oligos.

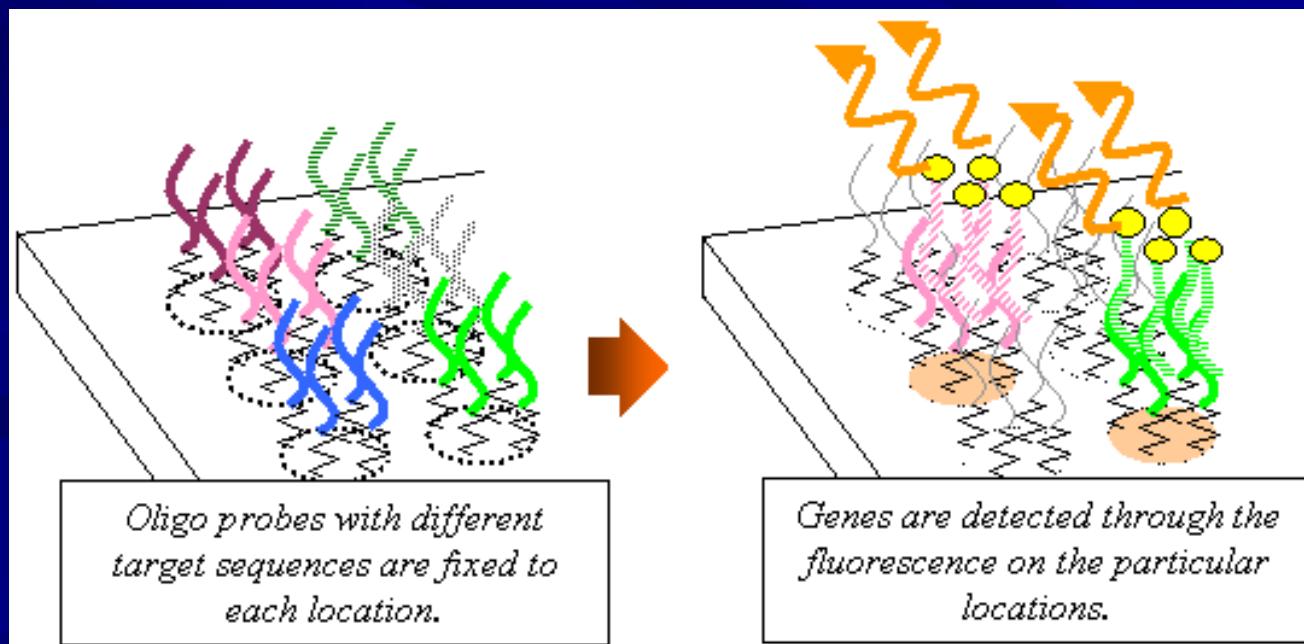
Expression of different genes

- mRNA converted to cDNA (complementary DNA) and labeled with fluorescent dye



Expression of different genes

- The locations in which fluorescence occurs can be used to determine the presence of specific kinds of genes in the sample.



Principles of all microarrays

1. Negotiate expression of thousands of genes.
2. Complementary hybridization of fluorescently labeled cDNA to a target spot.
3. Amount of fluorescence reflects relative amount of RNA expression.

Kinds of Arrays

- Nylon Membrane

- Low density, cheap.

- Glass slide

- Long oligos printed on a glass slide.

- Affymetrix

- Printed by photolithography. Short oligos.

- Illumina Bead Array

- Random beads with decode sequence.

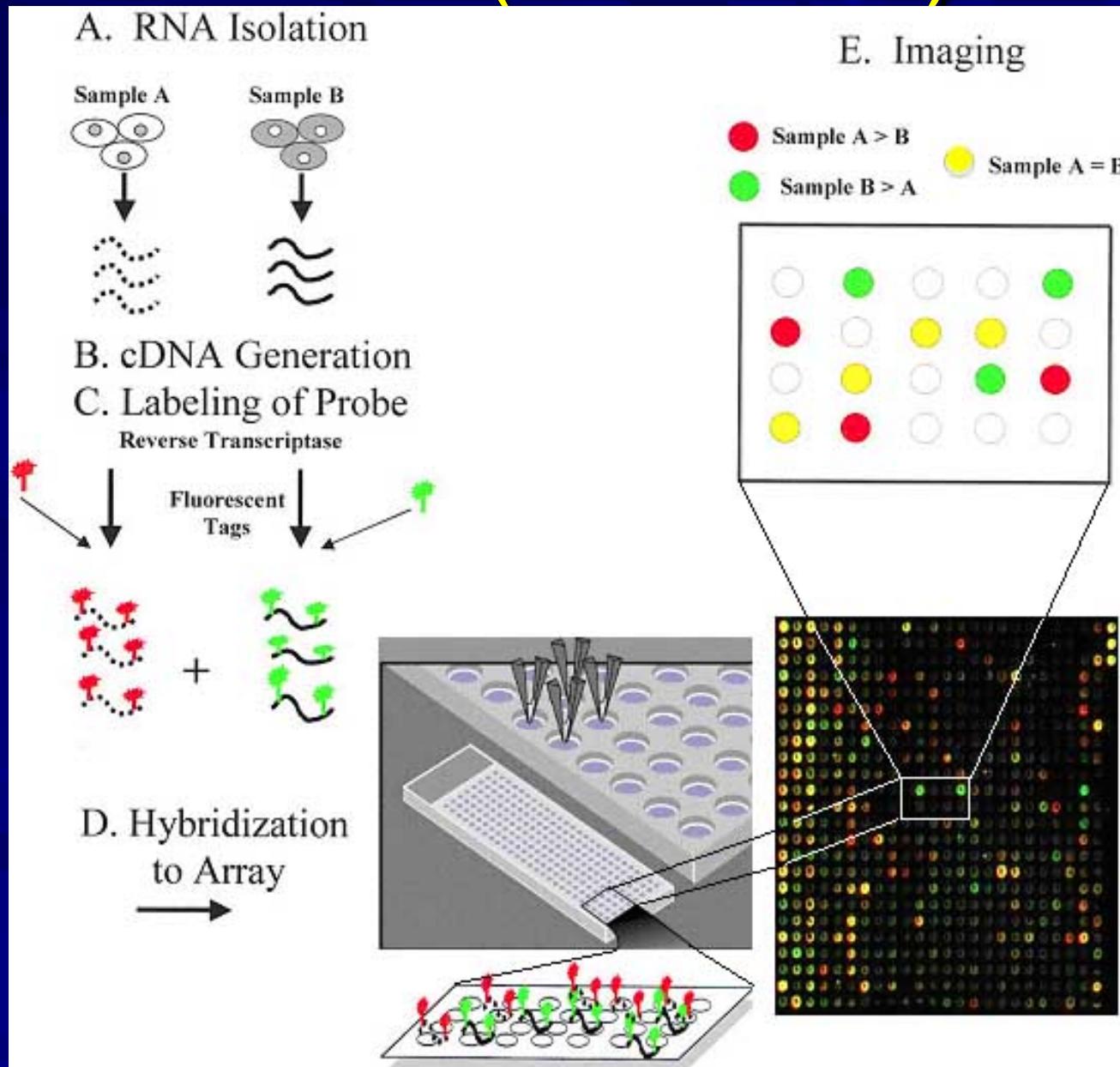
Glass Slide Arrays

- Same principle, but glass slides are used.
- Can be one-channel or two-channel.

Use of two-channel arrays

- ❖ Use *control* (reference) and a *test* samples, e.g. exposed and unexposed. Take identical amounts of mRNA of each and convert to cDNA
- ❖ Incorporate a **GREEN** fluorescent dye into one cDNA (e.g. *control*)
Incorporate a **RED** fluorescent dye into the other (e.g. *test*)
- ❖ **GREEN** spot indicates mRNA expression only in *control* sample, **RED** – in *test* sample, **ORANGE** – in both

Two-channel (two-color) arrays



Two-channel arrays

Advantages

- ❖ Assessment of gene expression in two samples on a single array
- ❖ Two samples have the same background variability on an array

Two-channel arrays

Disadvantages

- ❖ More laborious
- ❖ Each channel may behave differently
- ❖ Normalization of microarray data
WITHIN and BETWEEN the arrays is still needed

Single-channel arrays: Affymetrix and Illumina

Single-channel arrays (Affymetrix)

- Short oligos.
- Several probes per gene.
- Well developed methods for data processing.

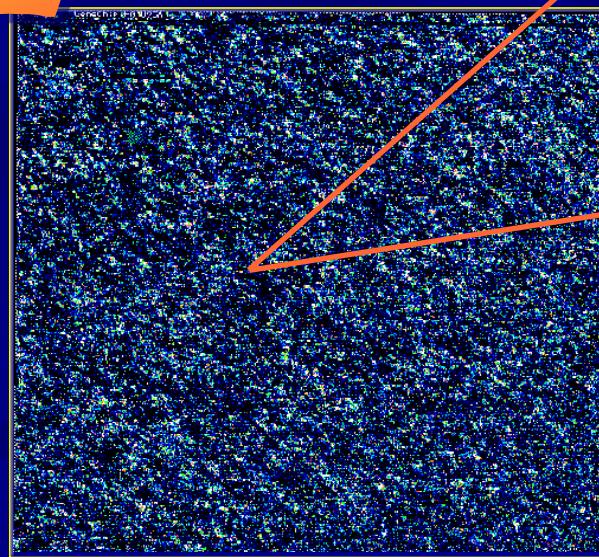
Single-channel arrays (Affymetrix)

UV light

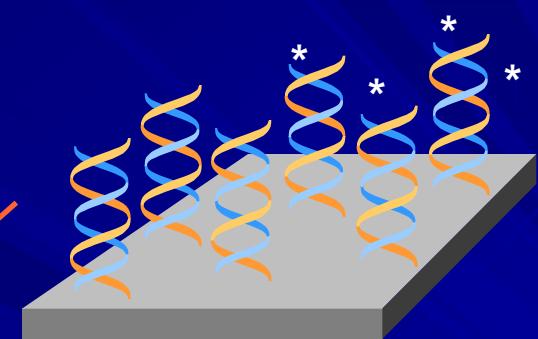
Photolithography



Substrate

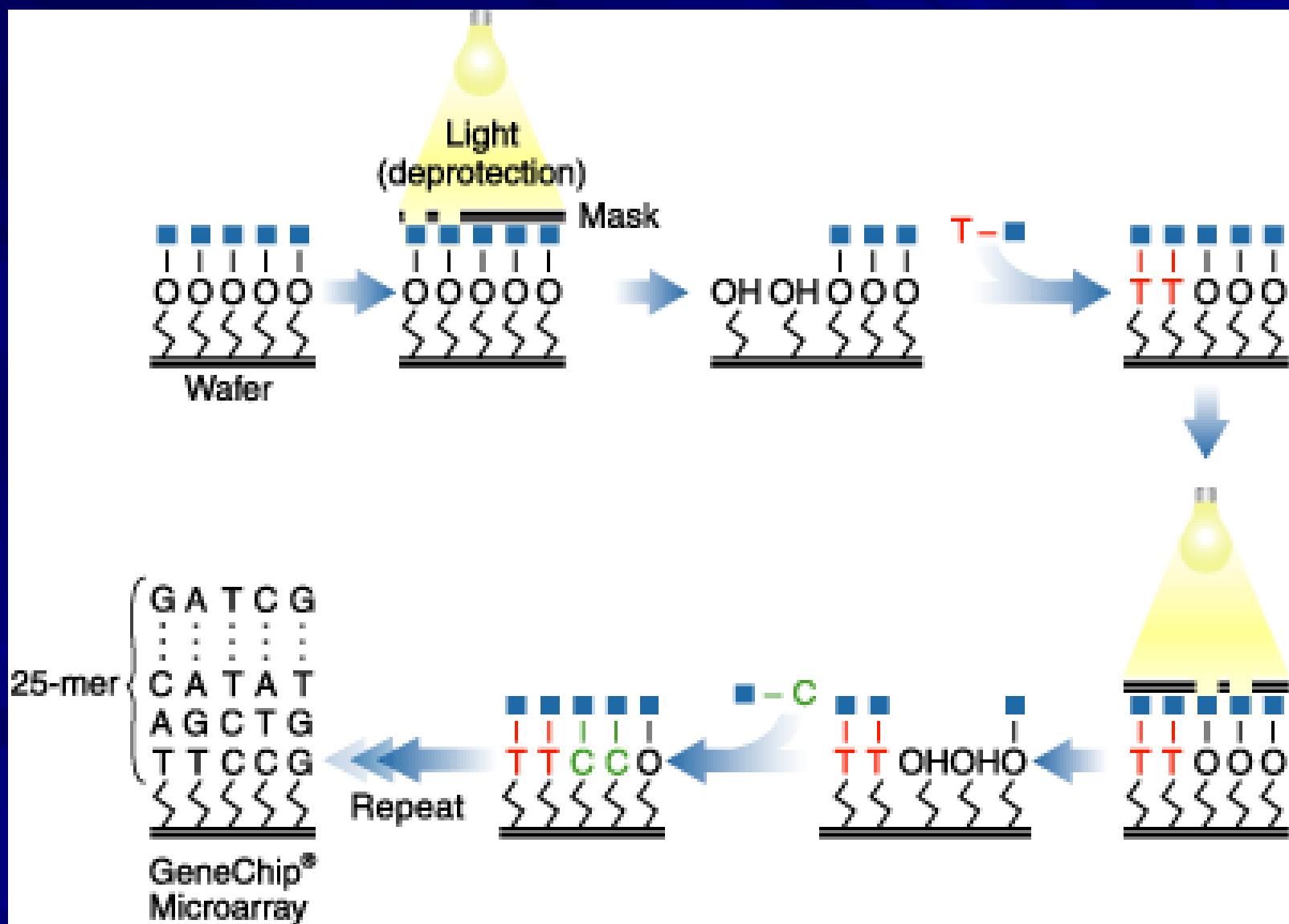


1.28cm

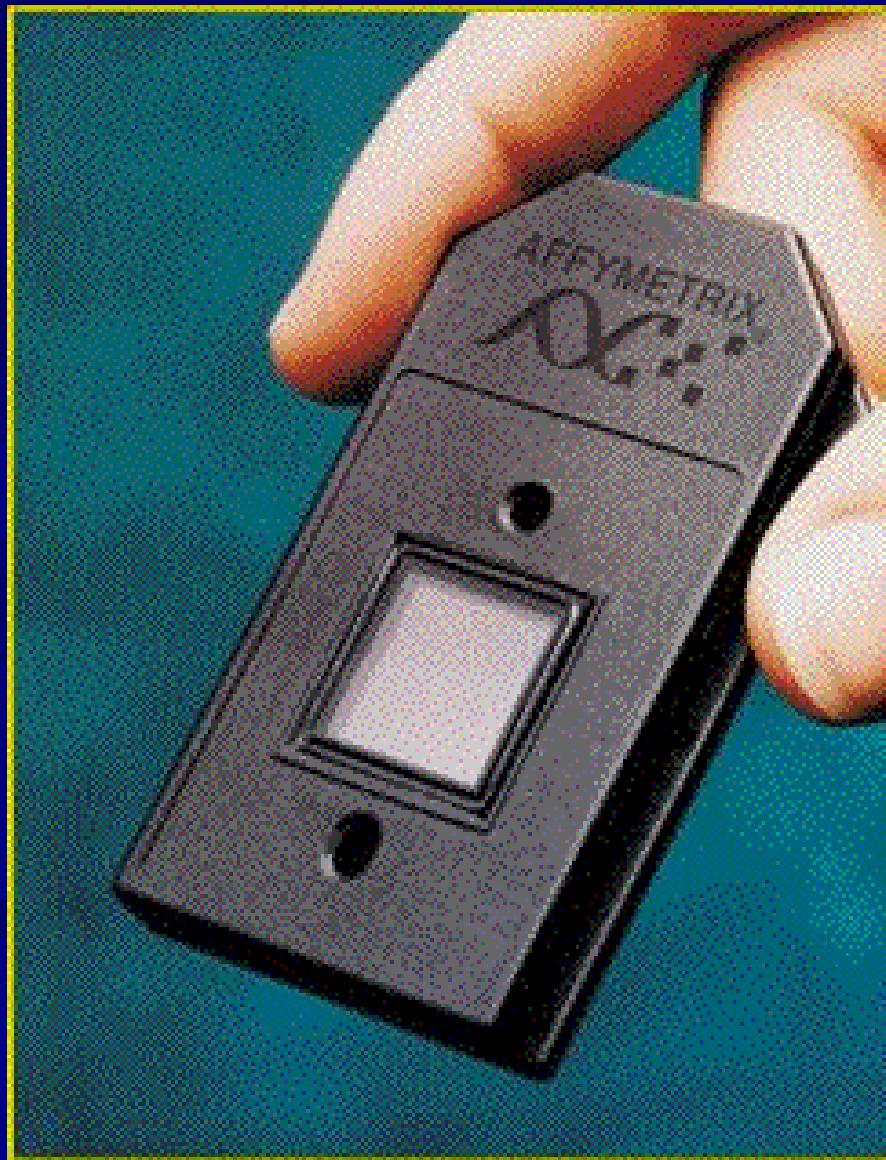


Millions of
probes
recognizing
unique
sequences

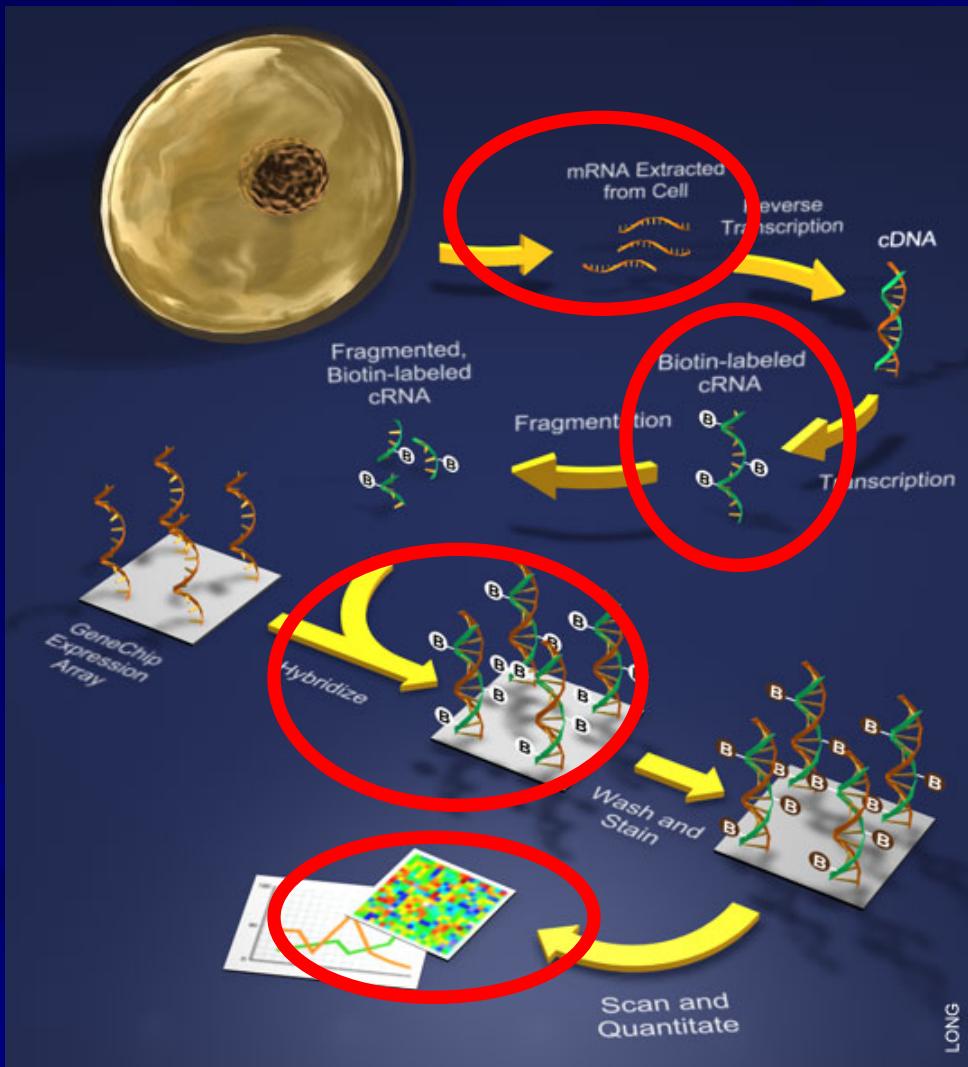
Making Affymetrix arrays



Affymetrix chip

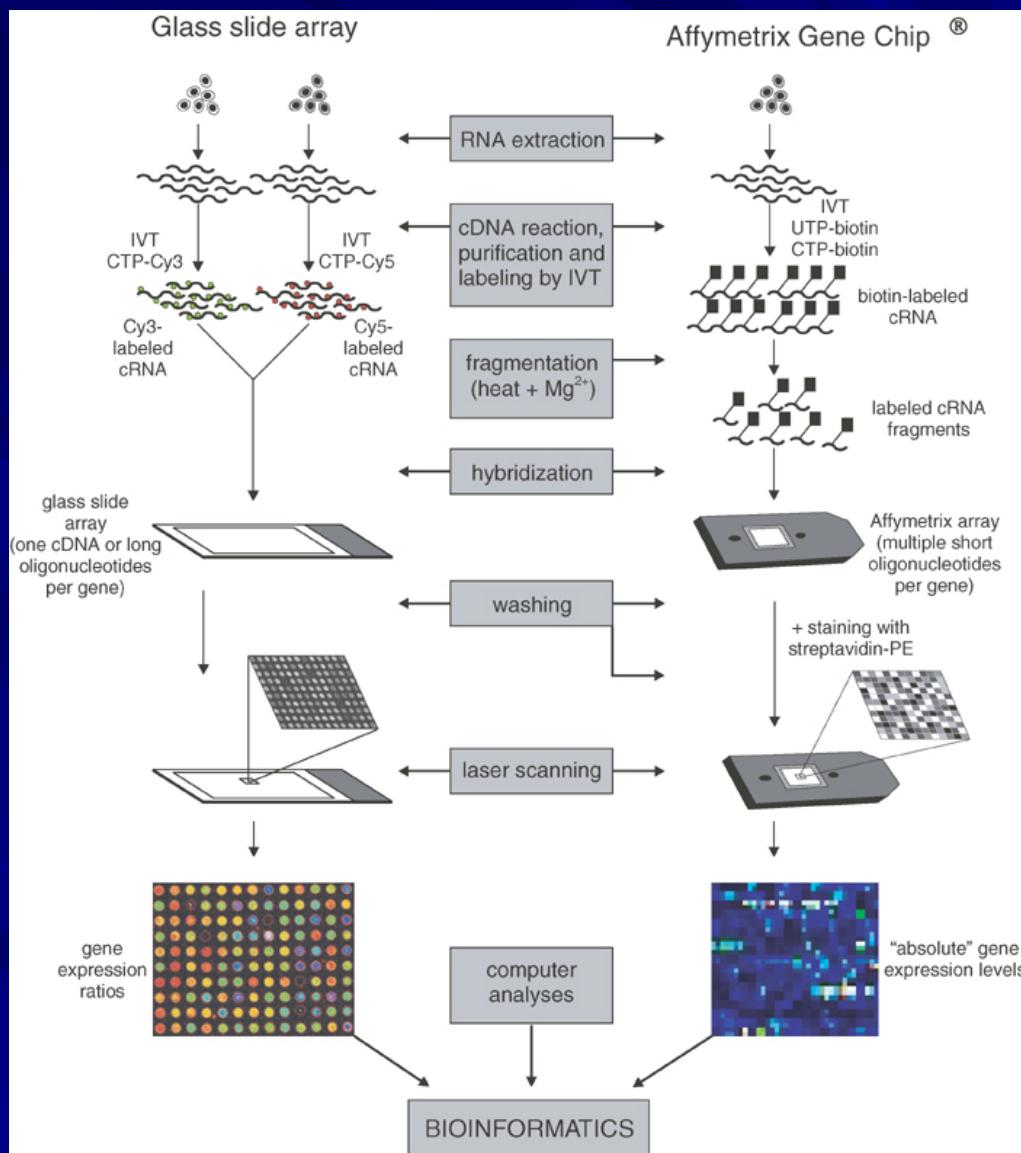


The use of single-channel arrays



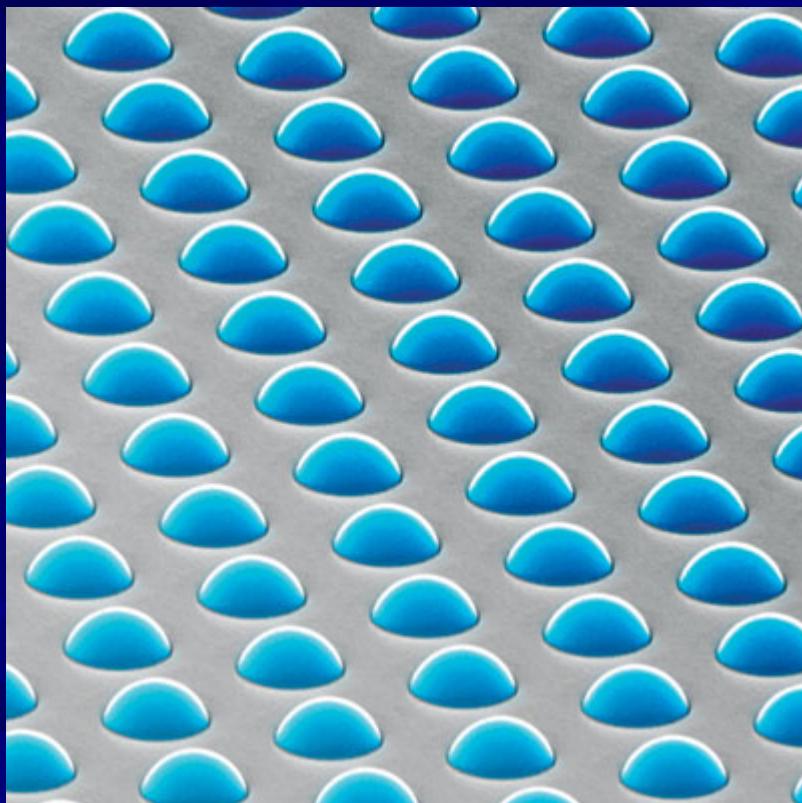
- ❖ mRNA extraction from one sample
- ❖ cDNA synthesis and fluorescent dye-labeling
- ❖ cDNA hybridization onto array
- ❖ Scanning and quantification of fluorescence of each spot

Glass vs. Affy arrays

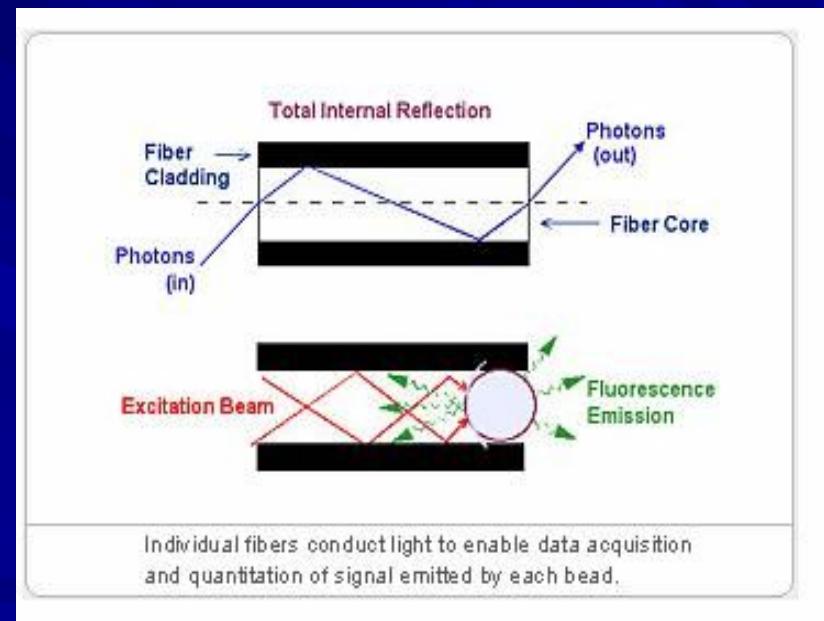


Illumina Bead Arrays

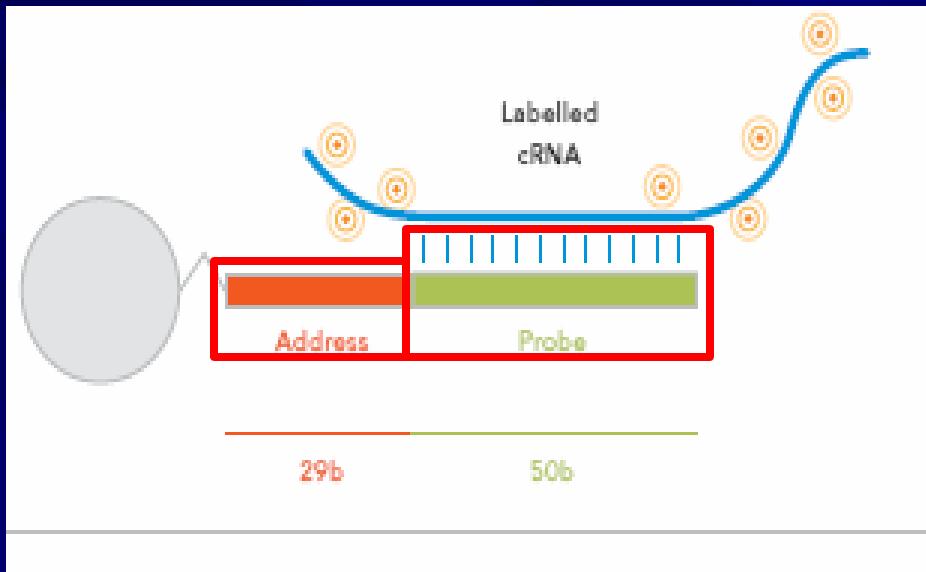
Beads form array on light fibers.



Illumination from below excites fluorescence—quantifies probe bound.



Illumina Bead Arrays



- Address allows decoding (“barcode”) with decoder oligo.
- At least 30 beads per gene
- Each bead contains ~100,000 copies of a single sequence.

The use of single-channel arrays

Advantages

- ❖ Analysis of ONE sample per array
- ❖ Straightforward approach – more fluorescence = more RNA

The use of single-channel arrays

Disadvantages

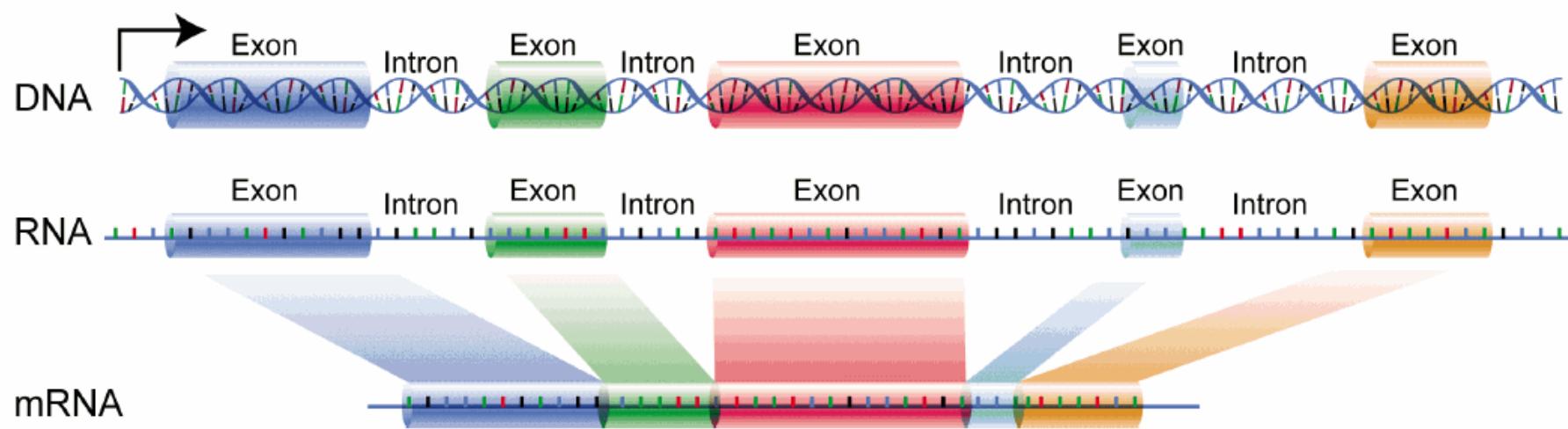
- ❖ Need to use another array(s) for comparative analysis
- ❖ Careful normalization of one microarray data to the other is a must

Affymetrix exon arrays

Exon array principles

- Gene-level and exon-level detection of expression.
- Allow detection of alternative splicing mRNA transcripts.

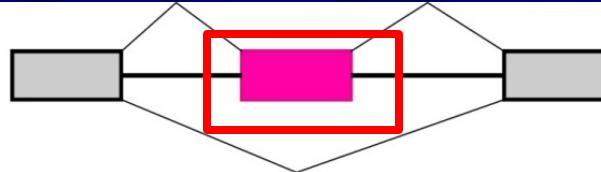
Gene structure in eukaryotes



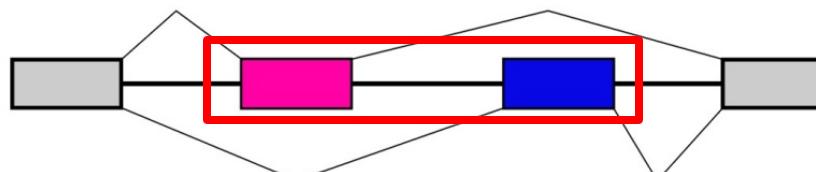
- ❖ Exon – Expressed regiON, coding part
- ❖ Intron – INTergenic regiON, non-coding part
- ❖ mRNA – messenger RNA, carrier of genetic information.

Alternative splicing

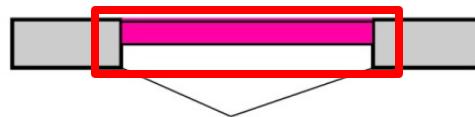
Cassette Exon



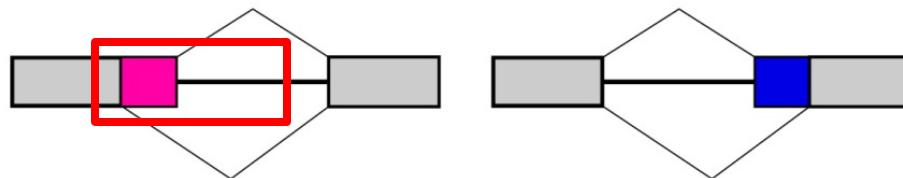
Mutually Exclusive Exons



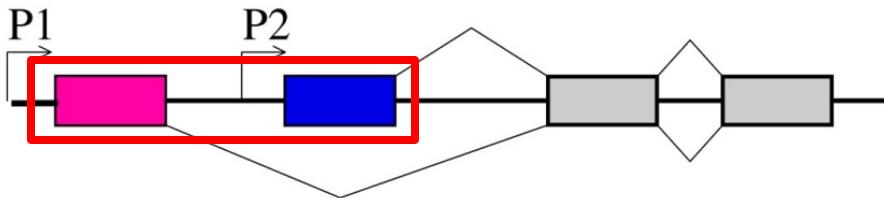
Intron Retention



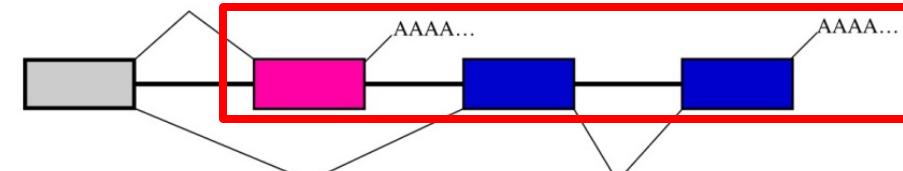
**Alternative 5'
or 3' Splice Sites**



Alternative Promoters

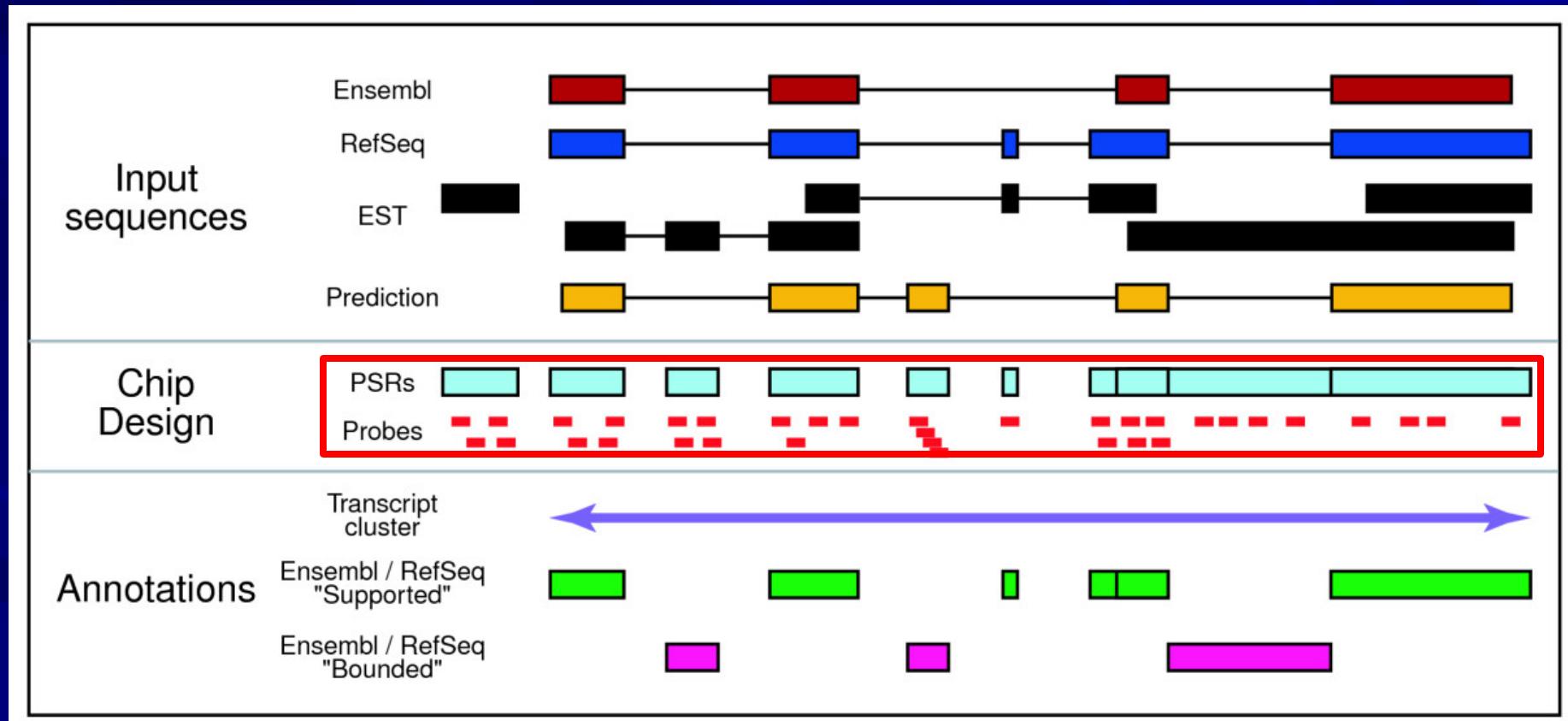


**Alternative Splicing
and Polyadenylation**



Exon array design

■ PSR – Probe Selection Region



Affymetrix exon arrays

- ❑ Affymetrix GeneChip Exon 1.0 ST
 - Wide coverage
 - Well annotated genes plus gene prediction sets
 - Over 1.4 million probe sets

The use of exon array

Advantages

- Allow detection of alternative splicing.
- Cost is about the same as for regular microarrays.

The use of exon array

Disadvantages

- Careful probe design is imperative.
- Methods for analysis are not well developed.

Why we love microarrays

- ❖ The technology is global, tracks expression of the whole genome
- ❖ Repeatability is not bad
- ❖ Many diseases have pronounced gene expression profiles

Why we hate microarrays

- ❖ Proteins are more important.
- ❖ mRNA changes may not reflect protein changed.
- ❖ The most interesting genes are really unstable.

Additional Microarray Platforms

- Gene Expression Arrays
- Exon Arrays
- SNP/CNV Arrays
 - Whole Genome Association Studies
- ChIP-chip arrays
- Micro-RNA Arrays
- Protein arrays

What about RNA-sequencing?

Sequencing-based methods are very powerful but have typically been prohibitively expensive. With recent advances in low-cost, high-throughput **next generation sequencing**, these methods—referred to as “**RNA-seq**”—are becoming more common and will soon be dominant.

RNA-seq

Although details of the methods vary, the concept behind RNA-seq is simple:

- isolate all mRNA
- convert to cDNA using reverse transcriptase
- sequence the cDNA
- map sequences to the genome

The more times a given sequence is detected, the more abundantly transcribed it is. If enough sequences are generated, *a comprehensive and quantitative view* of the entire transcriptome of an organism or tissue can be obtained.

RNA-seq vs. Microarray

Hypothesis neutral, unbiased coverage, high sensitivity

Figure 1. Correlation of RNA-Seq and Exon Array Ratios

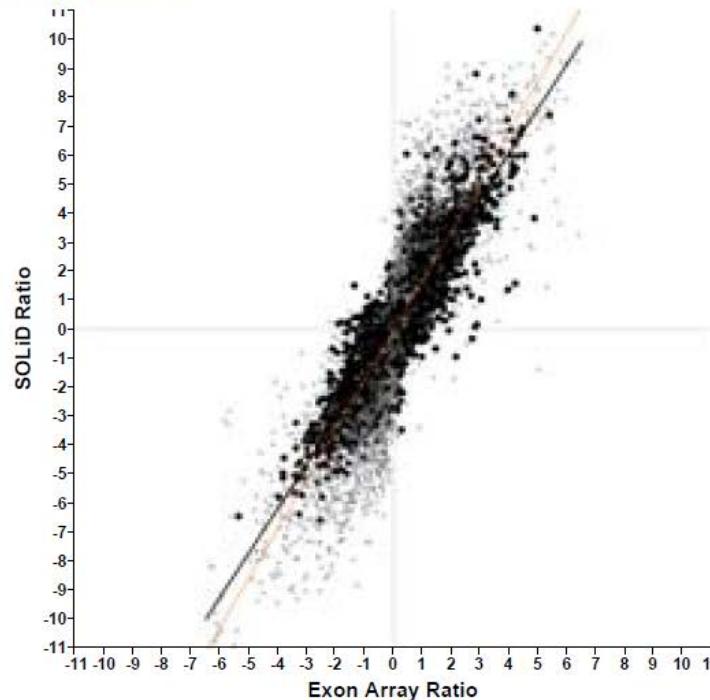
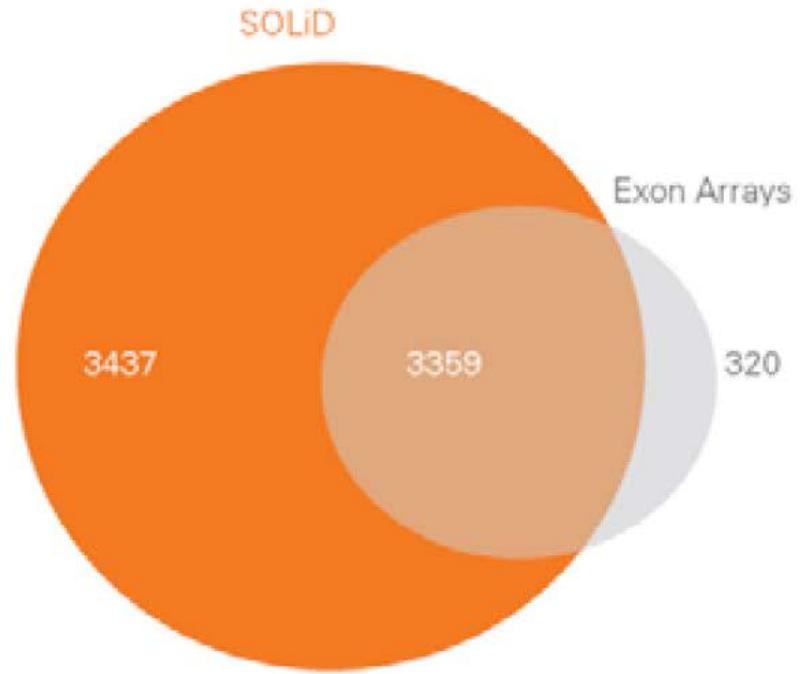


Figure 2. Concordance of Differentially Expressed Transcripts

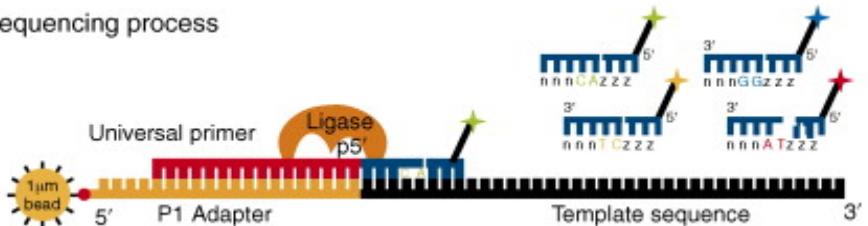


Comparing Sequencers

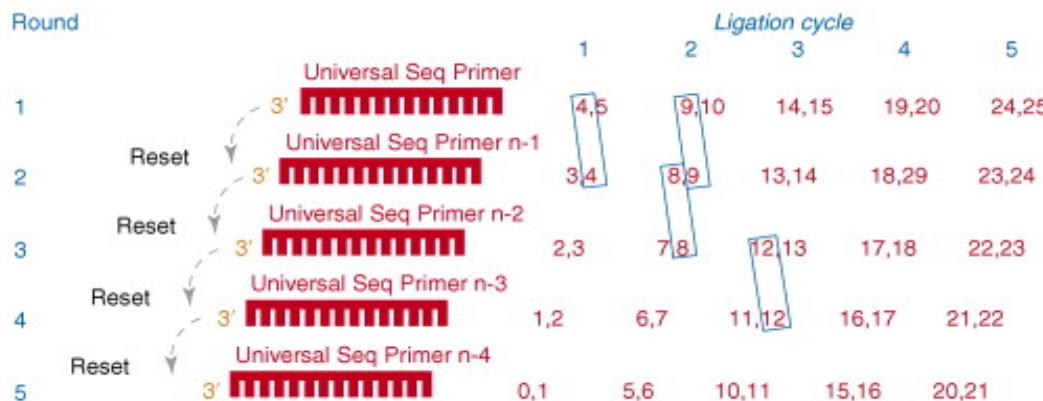
	Roche (454)	Illumina	SOLiD
Chemistry	Pyrosequencing	Polymerase-based	Ligation-based
Amplification	Emulsion PCR	Bridge Amp	Emulsion PCR
Paired ends/sep	Yes/3kb	Yes/200 bp	Yes/3 kb
Mb/run	100 Mb	1300 Mb	3000 Mb
Time/run	7 h	4 days	5 days
Read length	250 bp	32-40 bp	35 bp
Cost per run (total)	\$8439	\$8950	\$17447
Cost per Mb	\$84.39	\$5.97	\$5.81

ABI SOLiD Workflow

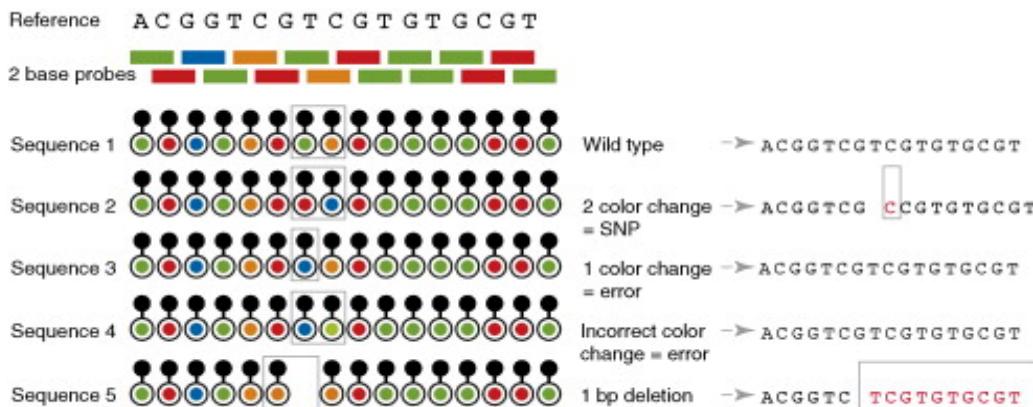
(a) Solid sequencing process



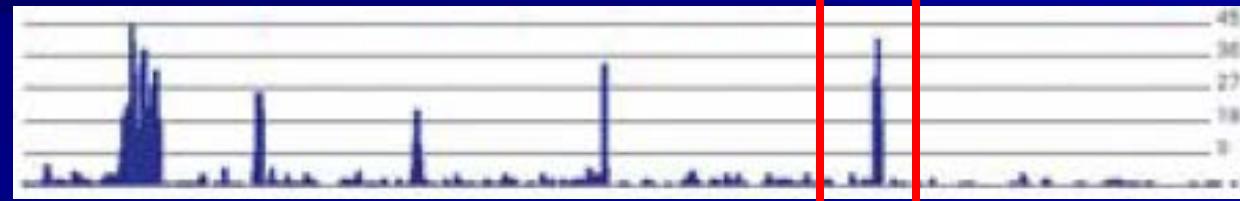
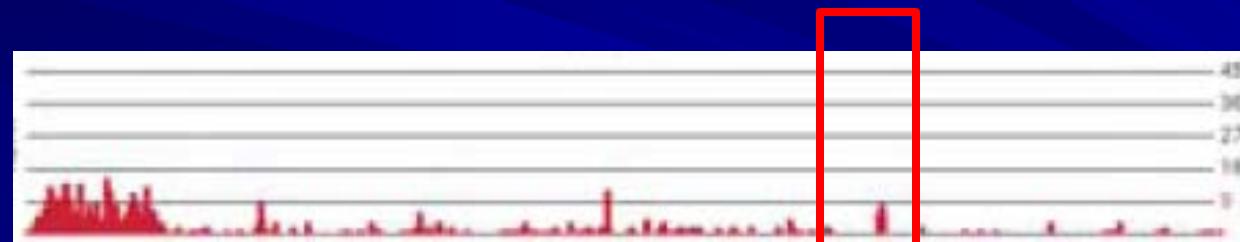
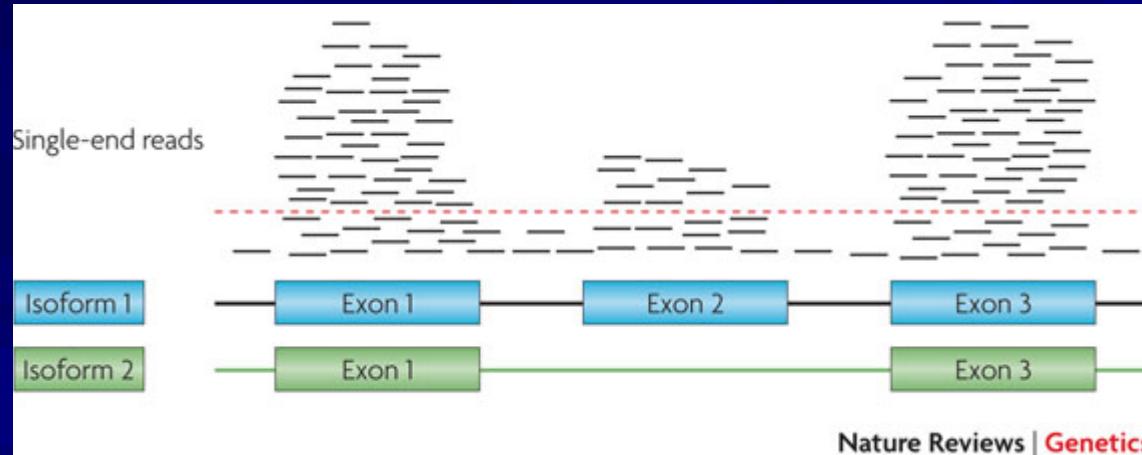
Round



(b) Principles of two base encoding



RNA-seq



High Throughput Sequencing

Prepare gDNA Library

3 hours hands-on, 6 hours
total time

Generate Clusters

<1 hour hands-on time, 5 hours
total time

Sequence Clusters

2.5 days single read
(36 bases)



RNA sequencing: good and bad

Advantages: wide applications

- Expression of different gene alleles
- Post-transcription mutations
- Gene fusions
- Splice variants

RNA sequencing: good and bad

Disadvantages

- High cost
- Methods for sequencing analysis are not comprehensive

Conclusion

- ❖ All microarrays are based on hybridization principle.
- ❖ Microarrays offer us a "peek under the hood" of the cell.
- ❖ Careful data processing is imperative.



Human explained

6. Body

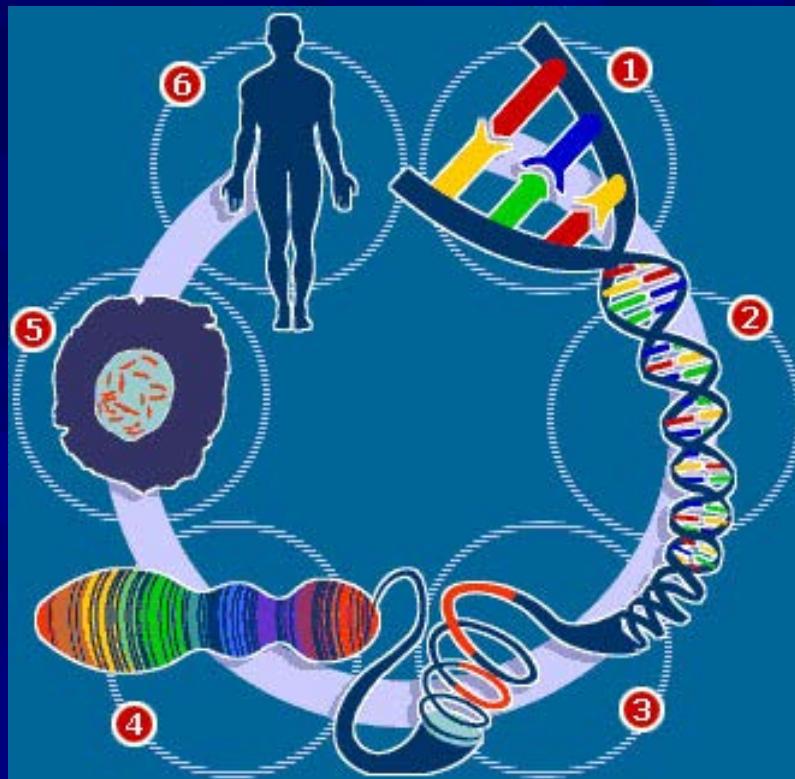
Each of the cells becomes specialized by obeying just some of the instructions in the DNA. Blood, muscle, bone, organs and so on result. The body is built from 100 trillion of these cells.

5. Nucleus and Cell

The 46 chromosomes are held in the nucleus found in most cells in the human body. Nearly every cell in the body contains the full DNA code for producing a human.

4. Chromosomes

The total number of genes is not known - estimates range from 30,000 to 120,000. However many there are, they, and all the junk DNA, are wrapped up into bundles called chromosomes. Every human has 23 pairs of chromosomes, one set from each parent.



1. The four letters

All genetic code is spelled out with just four chemical letters, or bases: adenine (A), thymine (T), cytosine (C) and guanine (G). These pair up, A with T and C with G. The human genome has between 2.8 and 3.5 billion base pairs.

2. DNA double helix

The base pairs form the rungs of the ladder-like DNA double helix. Running up and down the ladder are the long sequences of bases which are the code for life. Each cell in the human body contains two meters (six feet) of DNA.

3. Genes

As little as 3% of the total genome is made of genes - the rest is meaningless "junk". Genes are special sequences of hundreds or thousands of base pairs that provide the templates for all the proteins which the body needs to produce.

Thank you!

Microarray Technology

Mikhail Dozmorov, Ph.D.,
Arthritis & Clinical Immunology Department
Oklahoma Medical Research Foundation
Mikhail-Dozmorov@omrf.org
January 16, 2013