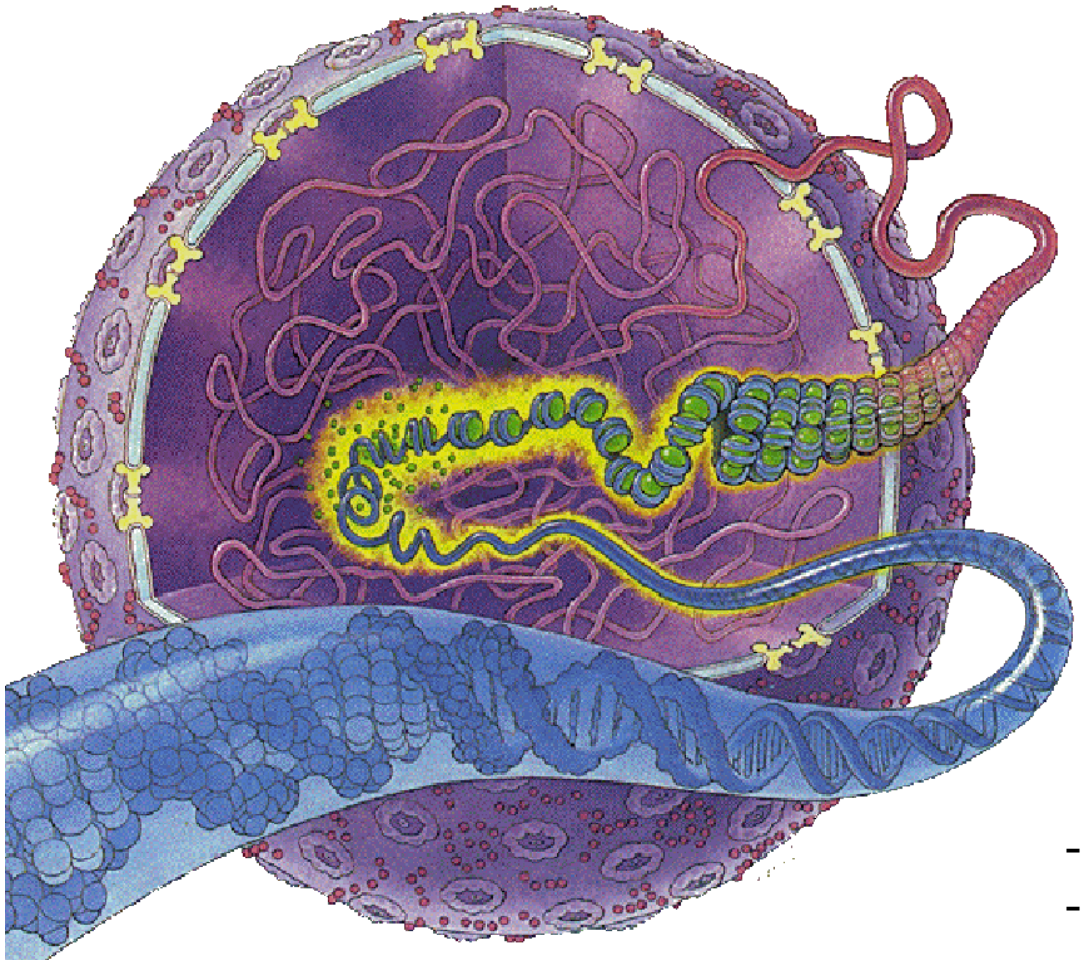
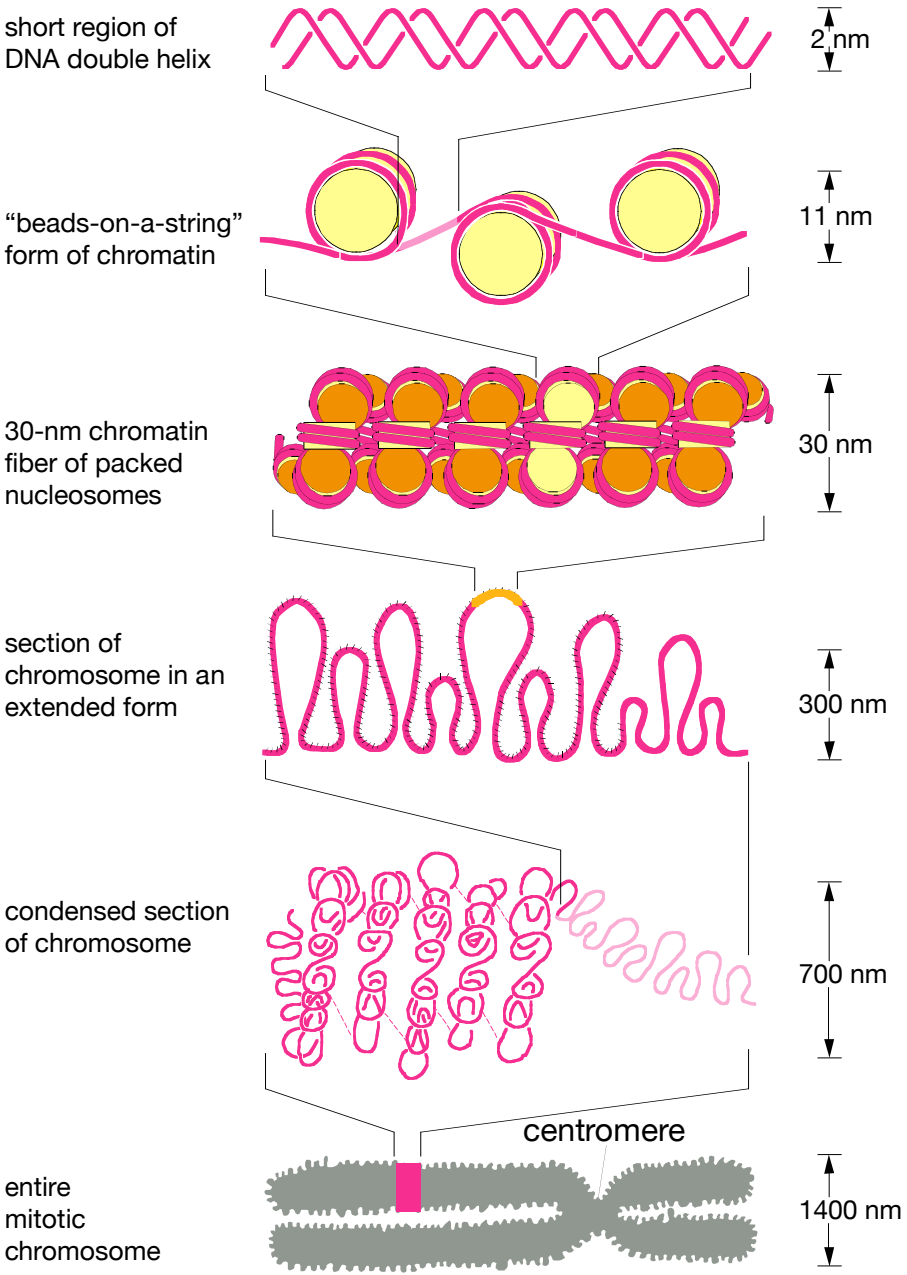
The human genome

A human cell nucleus



- 2 m DNA
- nucleus ≈ 10 µm diameter
- about 30 000 genes
-10 000 different nuclear proteins

from DNA to chromosomes



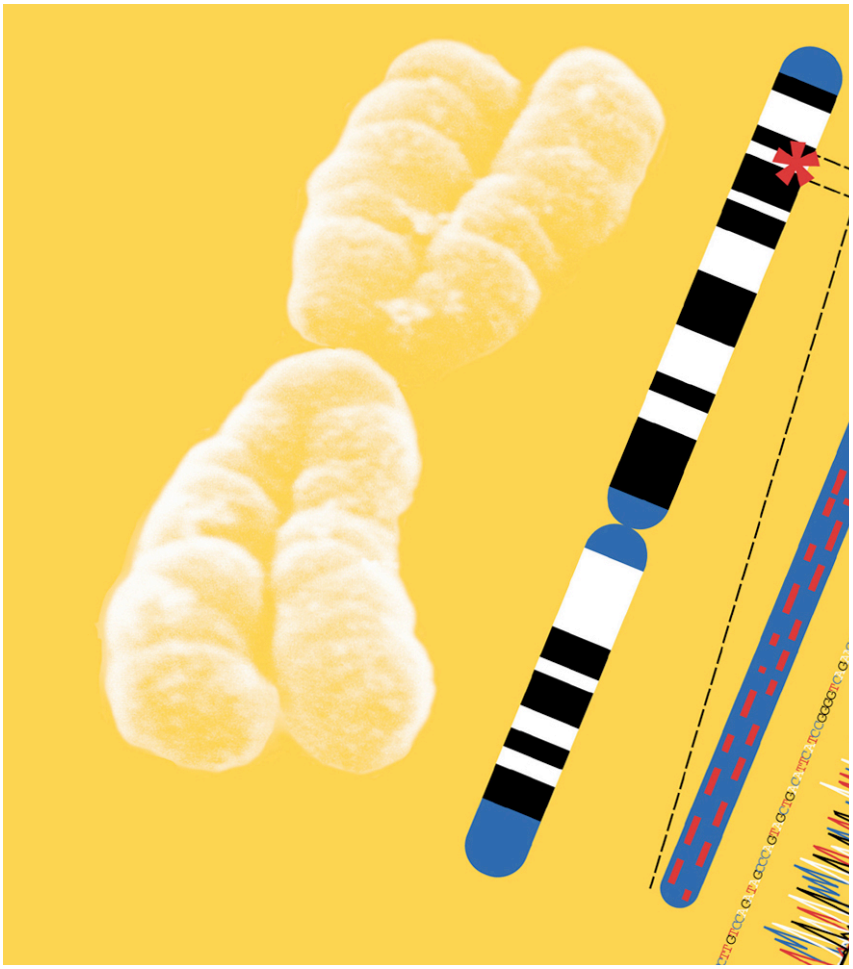
Telomeres

centromere

p arm

q arm

sister chromatids



How many genes are there in the human genome?

30,000 more? or less?

What percentage of the genome actually codes for genes?

only about 1-3% What is the rest for?

How large is the human genome?

3 billion base pairs

The largest gene: dystrophin (associated with Duchene's muscular dystrophy)

2 400 000 bases

The smallest genes: tRNA genes, about 100 bases

How did they estimate the number of genes?

1) genomic sequencing - extrapolation from sequencing large chromosome regions.

2) CpG island numbers - (short stretches of DNA--about 1-2 kilobases long) About 56% of genes are associated with CpG islands. The total number of CpG islands is about 45000.

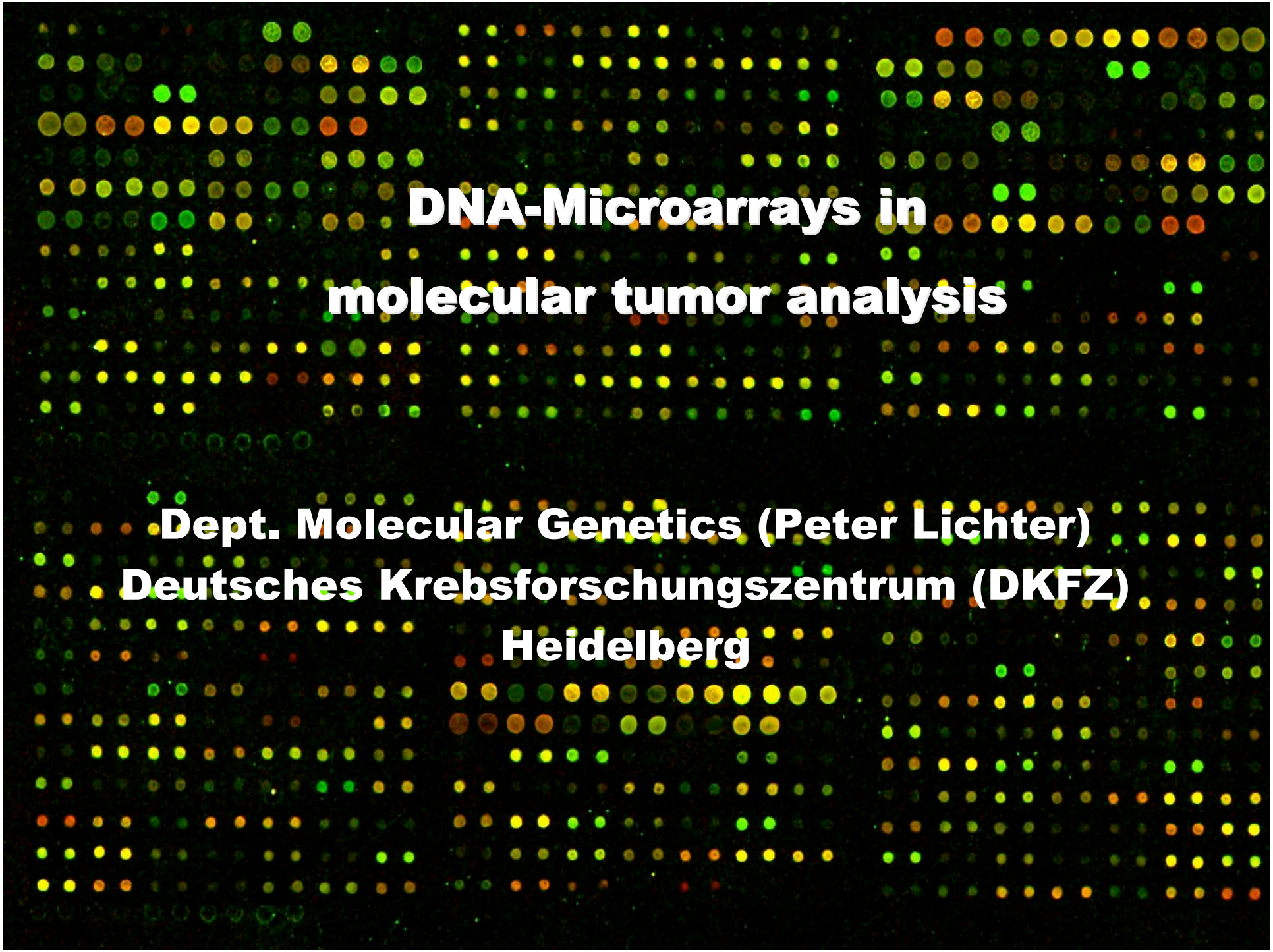
3) EST=expressed sequence tags

Nucleotide content of Human DNA

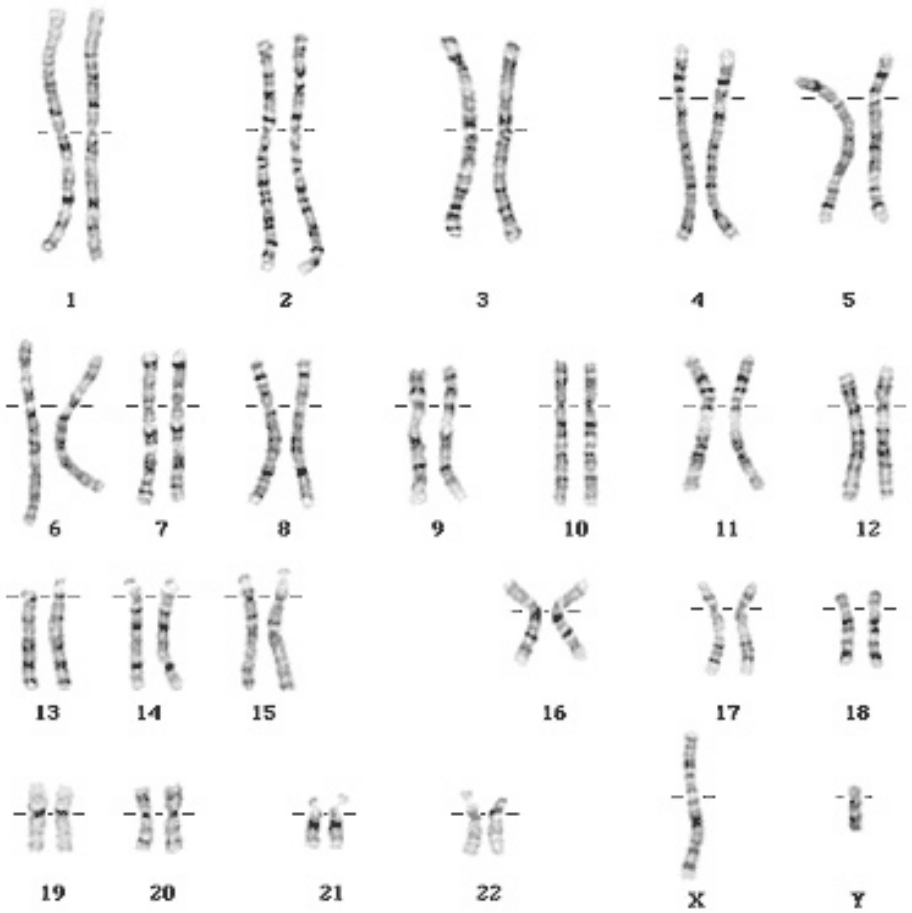
A	G	C	methyl C	T
29.9%	20.7%	19.9%	0.7%	30%

A - T

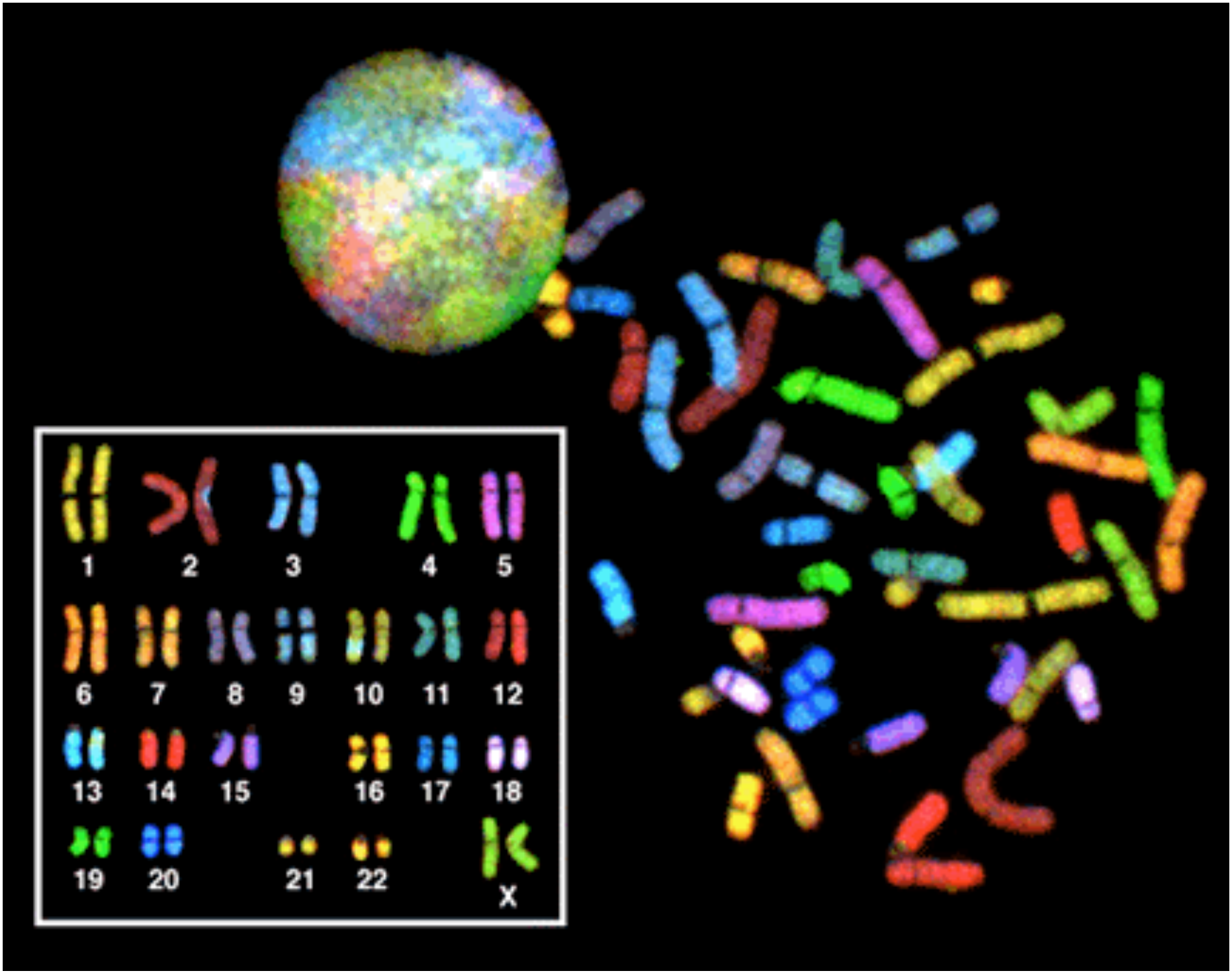
G - C or methyl C


DNA-Microarrays in
molecular tumor analysis
Dept. Molecular Genetics (Peter Lichter)
Deutsches Krebsforschungszentrum (DKFZ)
Heidelberg

Human chromosomes



24 chromosomes total
- 22 autosomal chromosomes
- X, Y

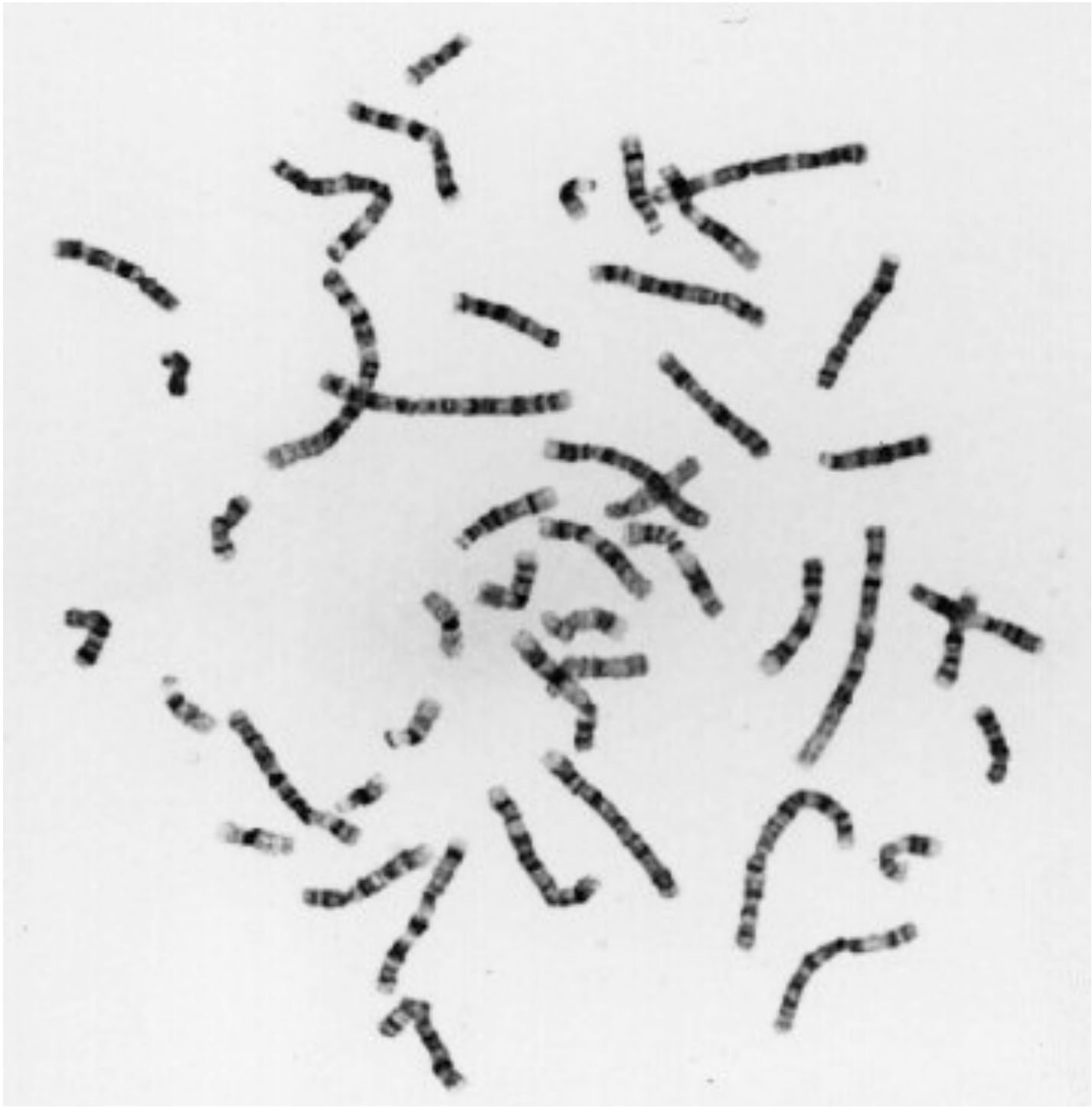

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X

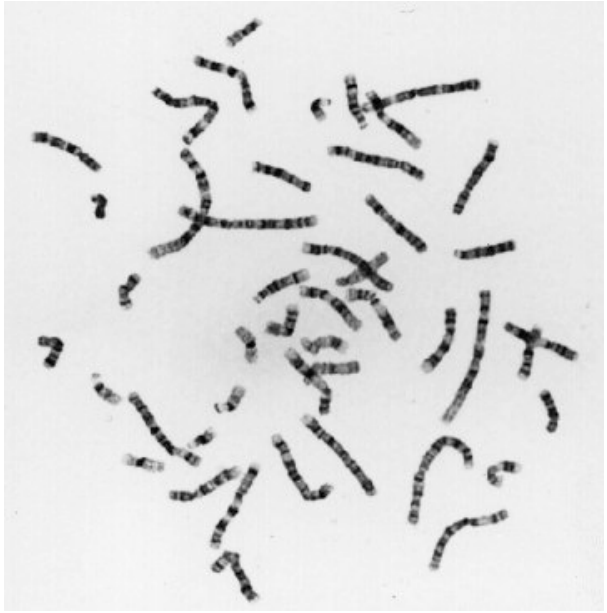
The 22 somatic chromosomes and the X and Y chromosomes are easily identified by

-size
-centromere location
-secondary constrictions (present on the long arms of 1,9, and 16)
-G band patterns

19 and 22 are gene rich

4 and 18 are gene poor





Properties of the Dark G Bands

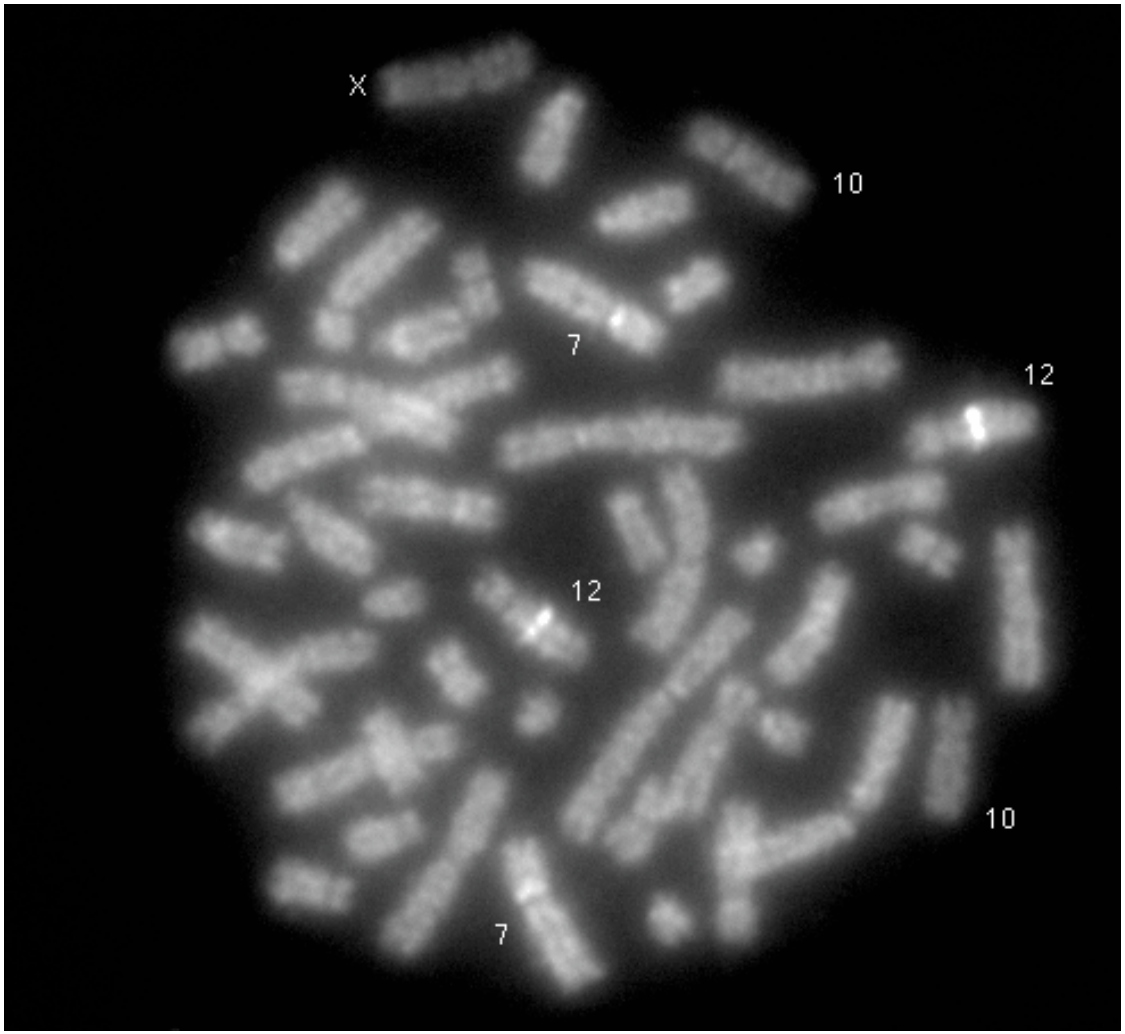
AT rich
Dnase insensitive
condense early in the cell cycle but, replicate late
gene poor
but, genes that are there, have large introns

More than 99.9% of human DNA sequences are the same across the population of all humans in the world.

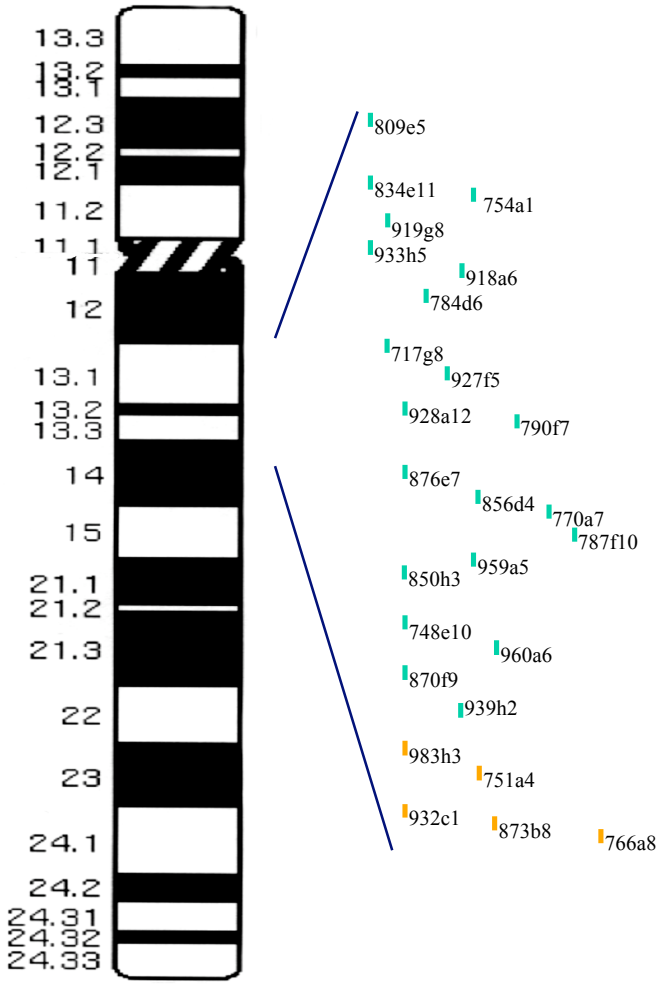
Single-nucleotide polymorphisms (SNPs) occur about once every 100 to 300 bases.

polymorphism = many forms

Amplification of oncogenes in a human tumour



gene card

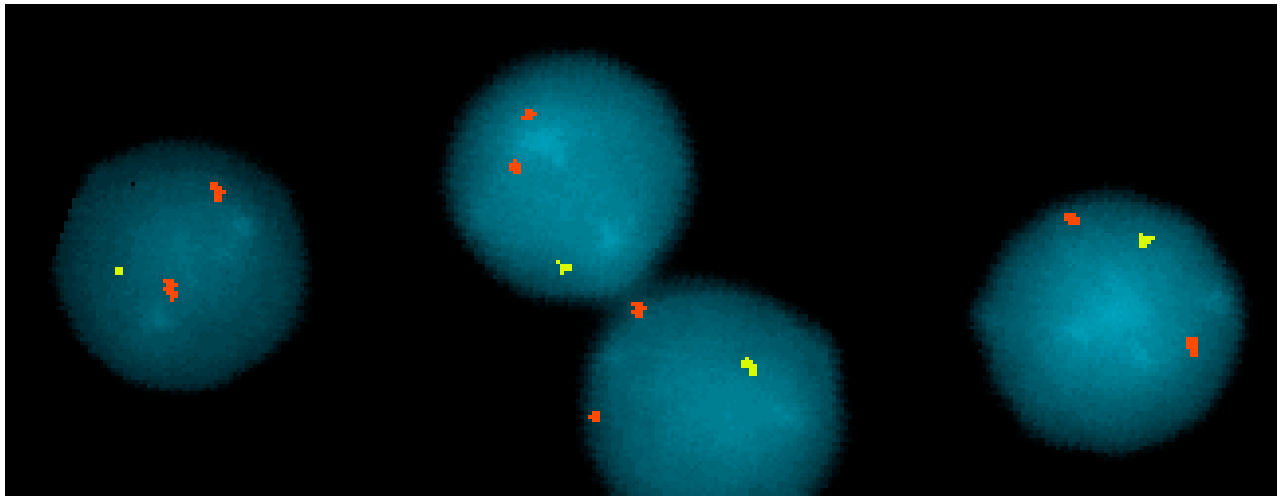


B-cell chronic lymphocytic leukemia (B-CLL)

mature but immuno-incompetent B-cells in peripheral blood

Impaired proliferation ?

Impaired induction of apoptosis



Profile of genomic alterations:

Loss: subregions within 6q, 10q, 11q (ATM), 13q (BCMS?), 17p (TP53)

Gain: subregions within 3q, 8q (MYC), 12q

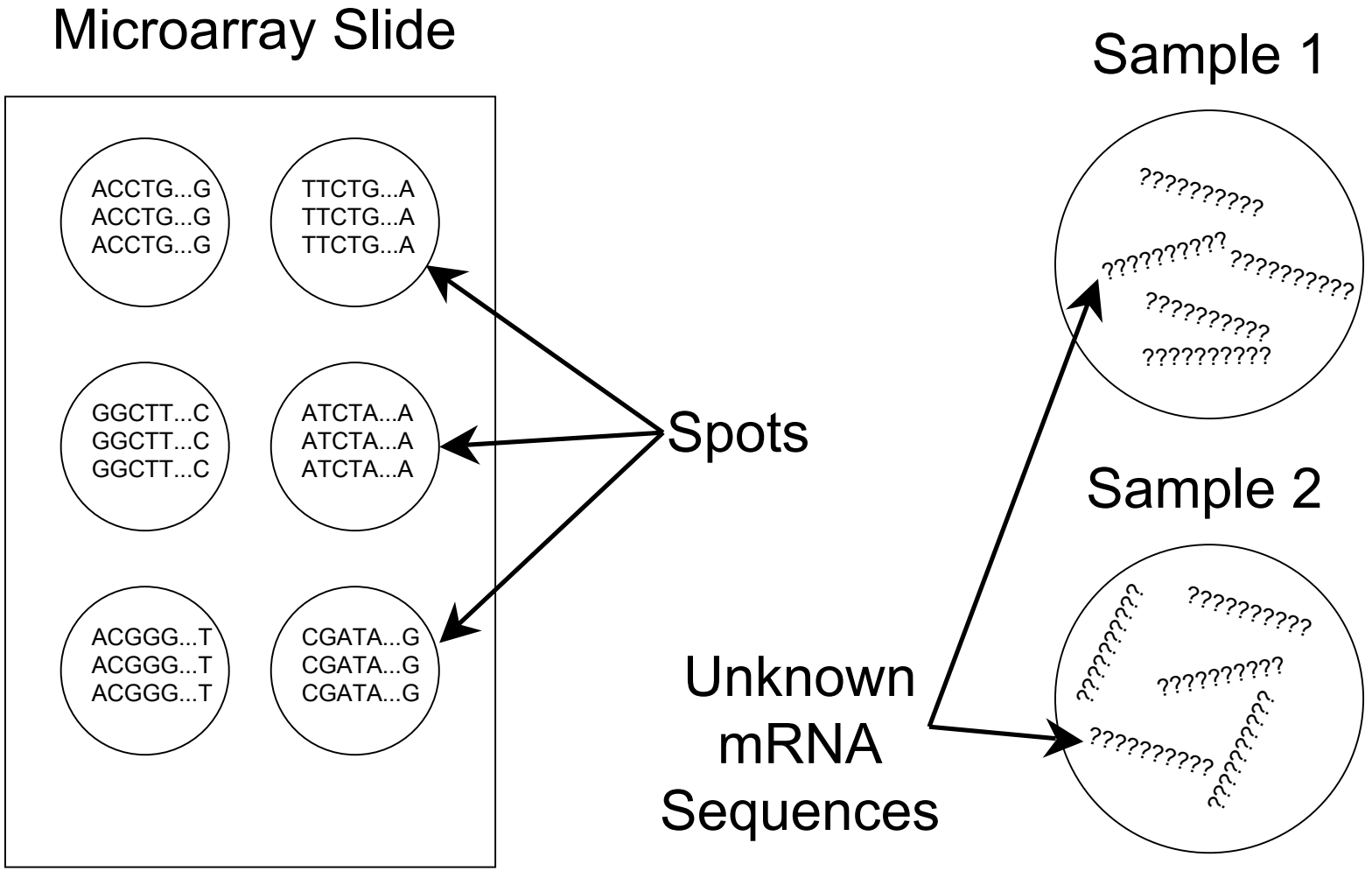
DNA microarrays

Microarray Technology

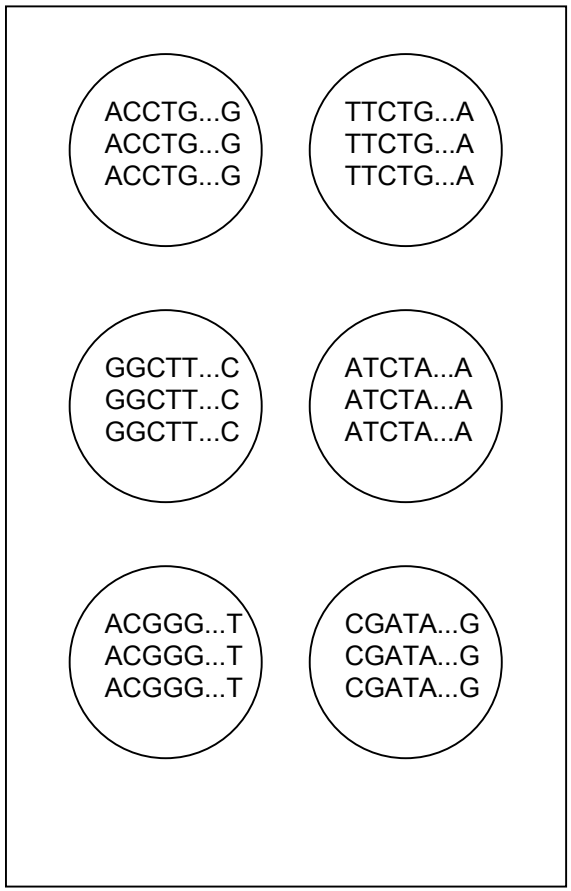
- Microarrays allow researchers to measure the abundance of thousands of mRNA transcripts in multiple biological samples.

- By understanding how transcript abundance changes across experimental conditions, researchers gain clues about gene function and learn how genes work together to carry out biological processes.

Two-Color Microarrays (cartoon version)

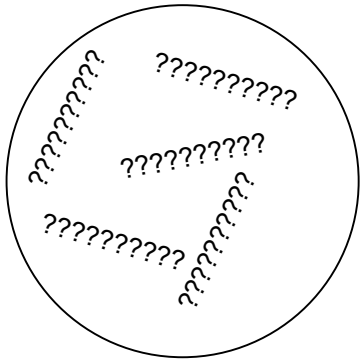


Extract mRNA

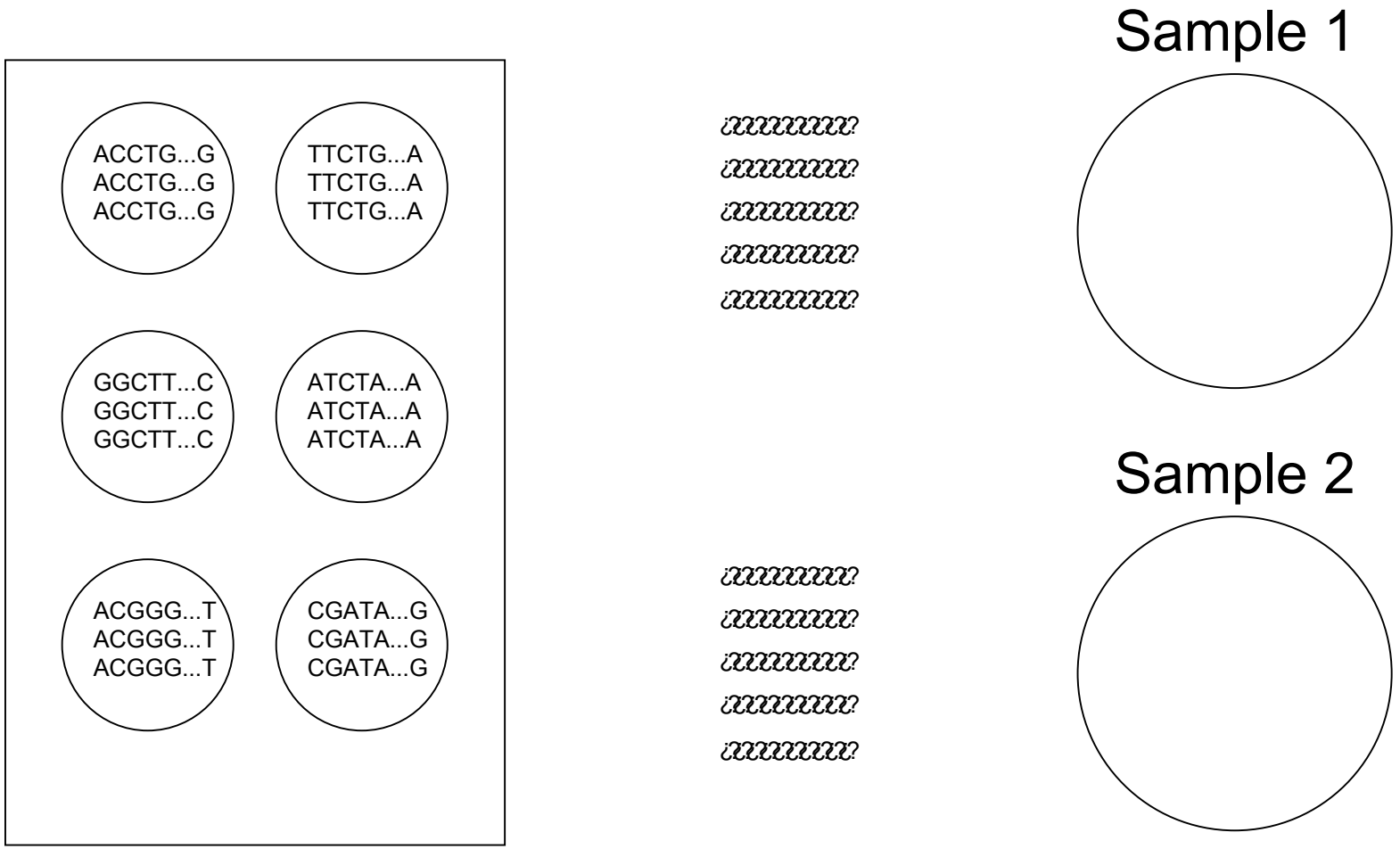


Sample 1

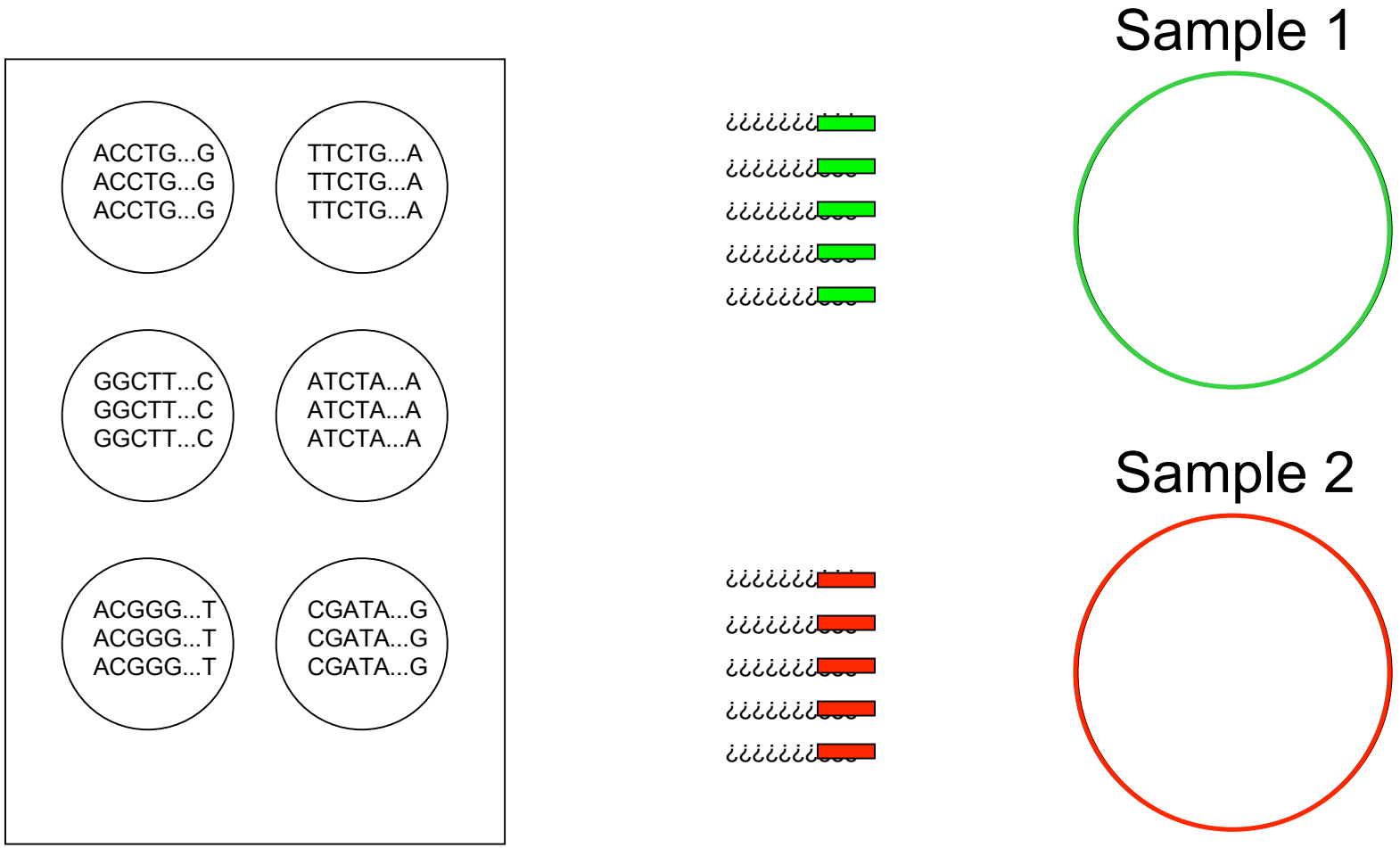
Sample 2



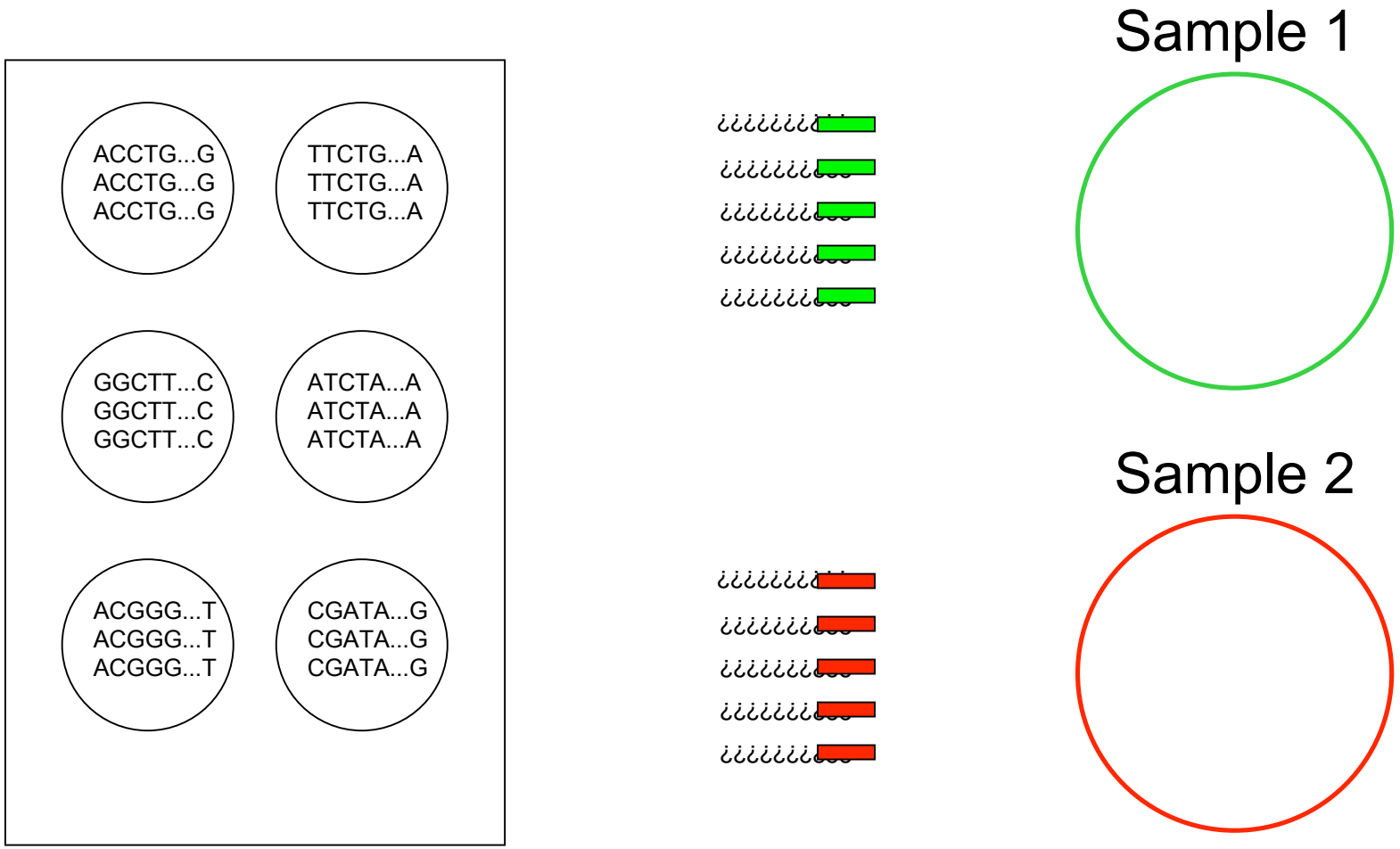
Convert to cDNA



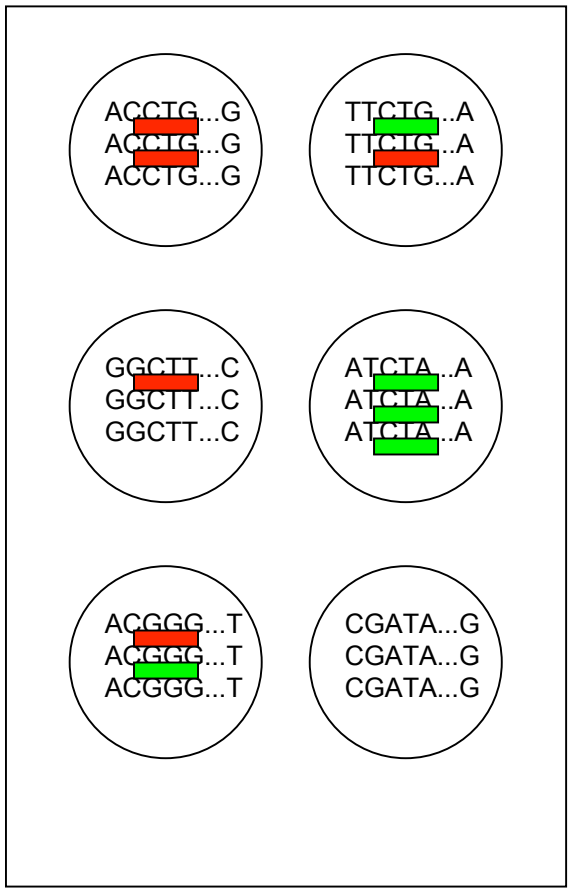
Tag cDNA from Different Samples with Different Fluorescent Dye



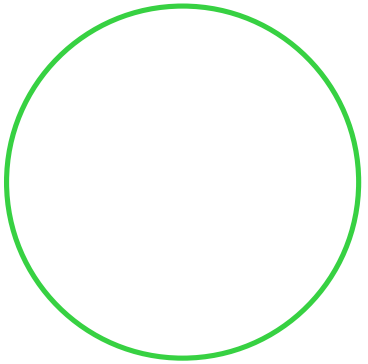
Hybridize cDNA to the Slide



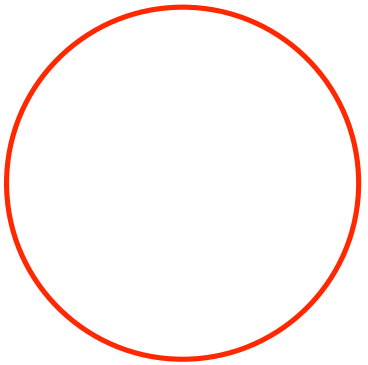
Hybridize cDNA to the Slide



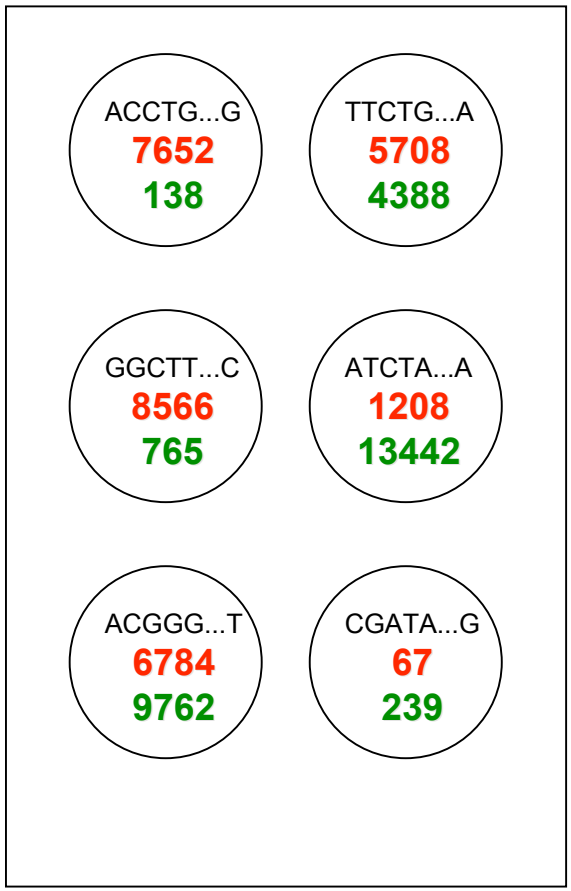
Sample 1

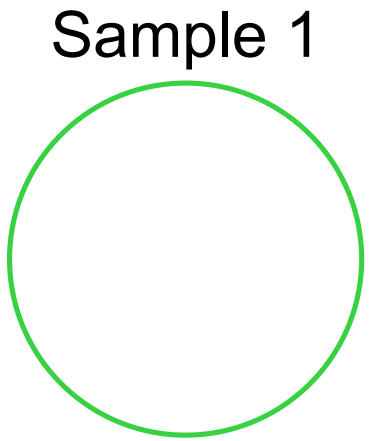


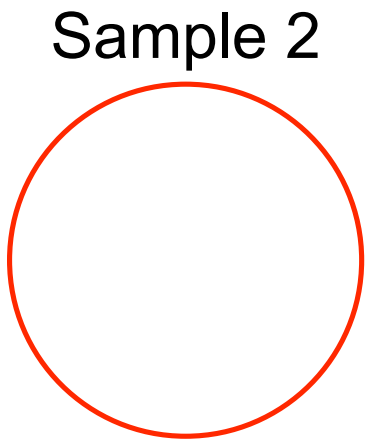
Sample 2



Scan and Quantify Signals



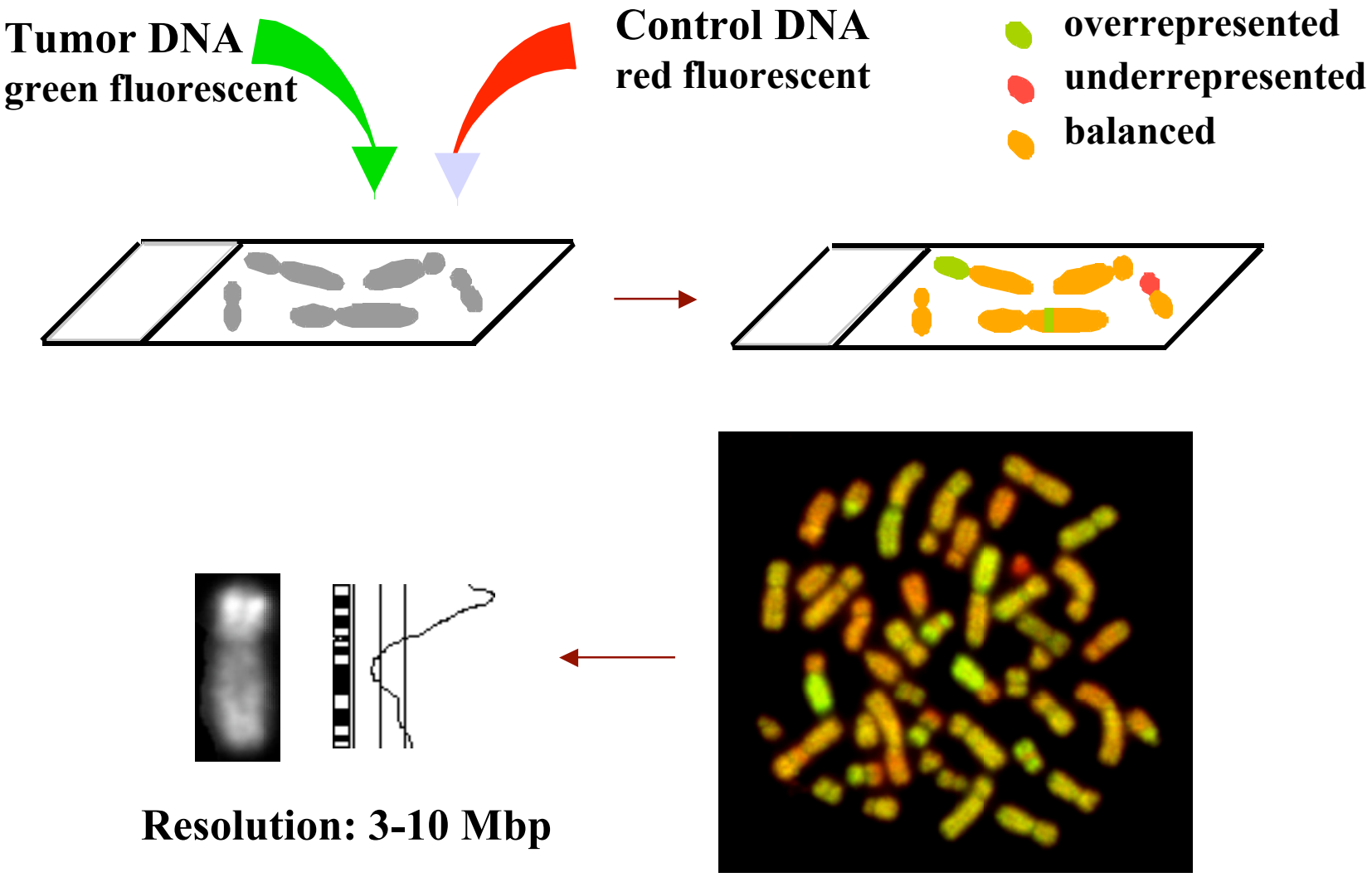




Matrix-CGH

Matrix-CGH is a chip-based, high-resolution method for the analysis of genomic-DNA copy number. It is used this important tool to characterize chromosomal aberrations in human tumors.

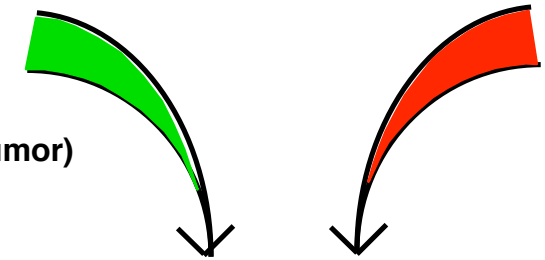
Comparative Genomic Hybridization (CGH)



Matrix-CGH

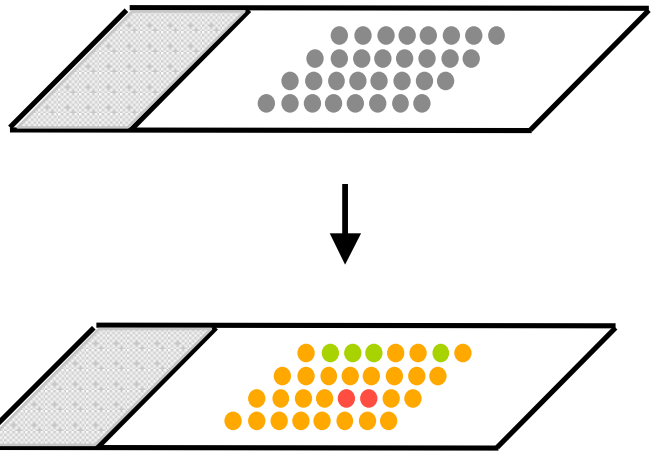
Test DNA

(e.g. from tumor)



Control DNA

Solinas-Toldo et al.
Genes Chromosom. & Cancer
20, 399-407, 1997

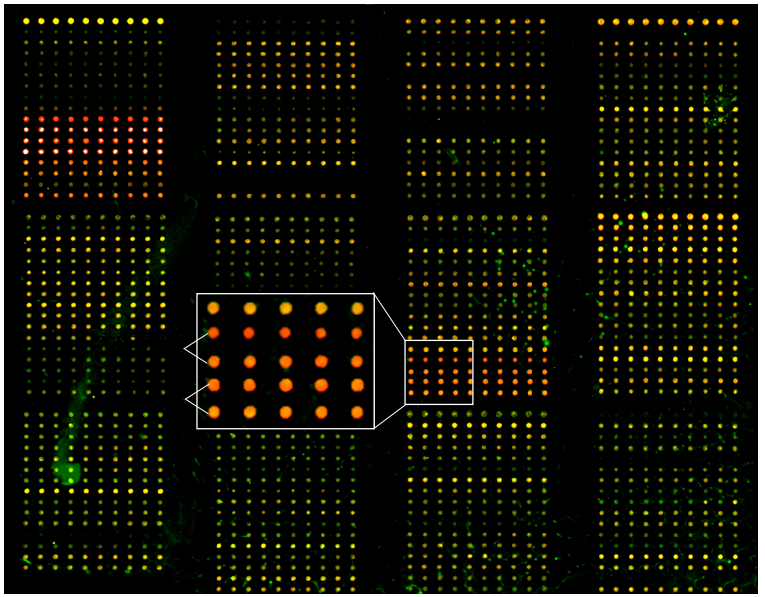


Matrix of microarrayed
genomic DNA fragments

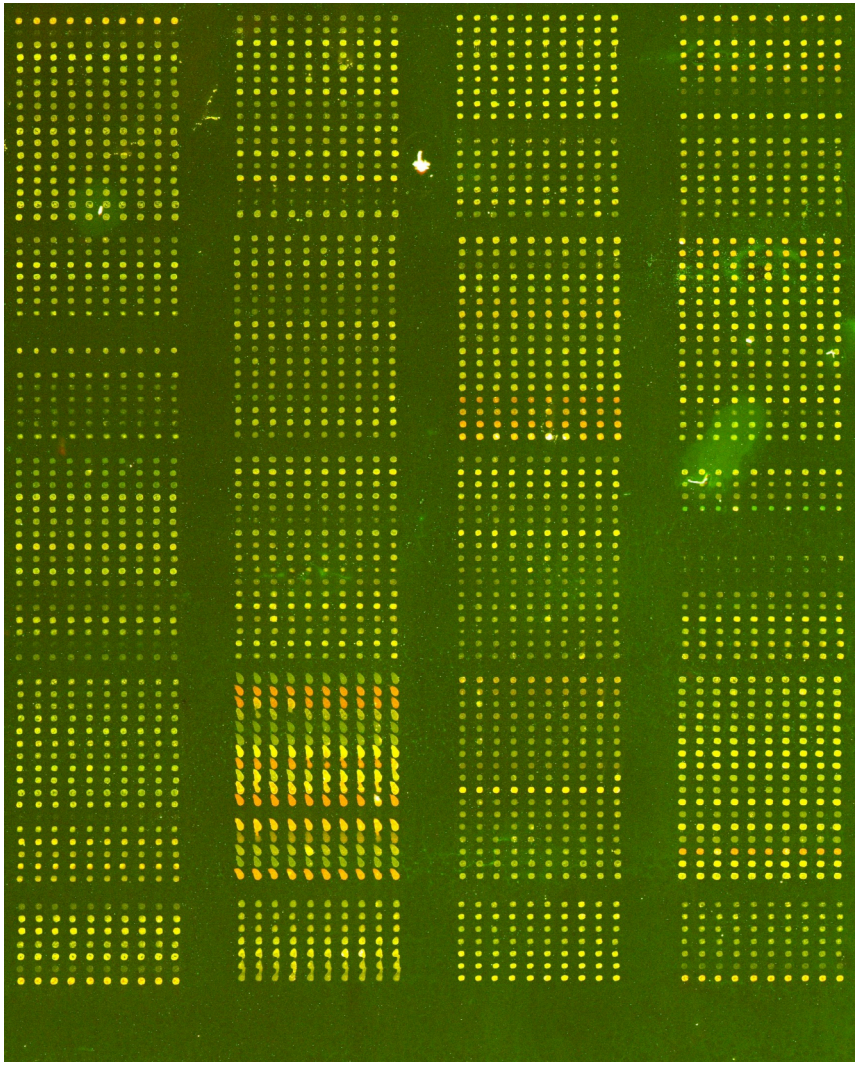
= material gained in test genome

= material lost in test genome

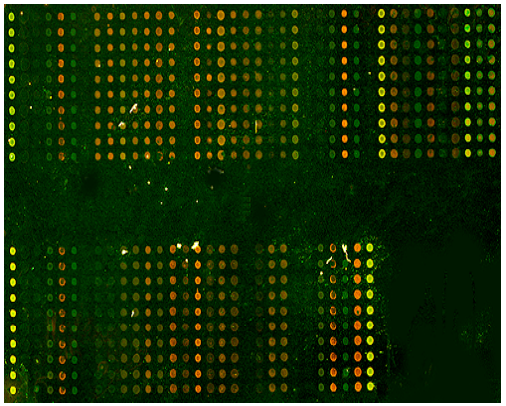
= material balanced in test genome



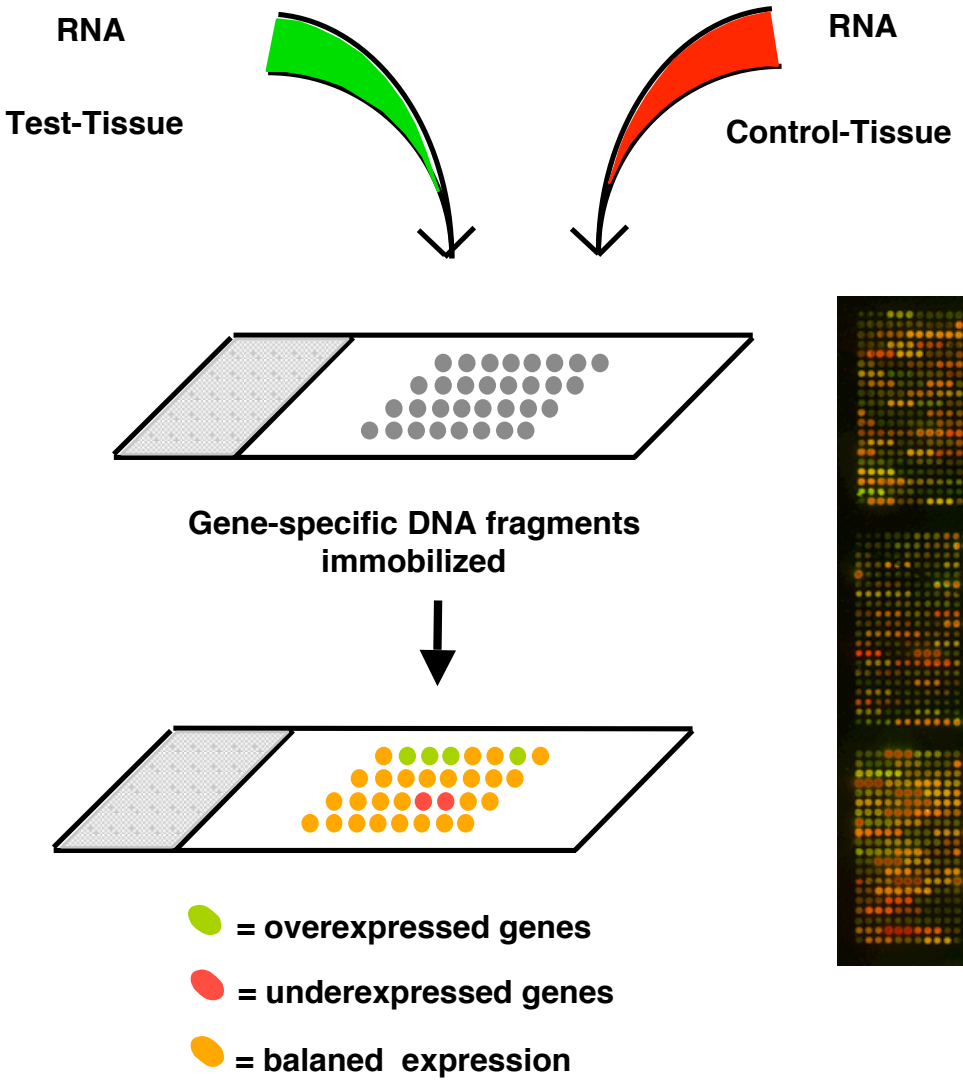
Diagnostic DNA-chips:

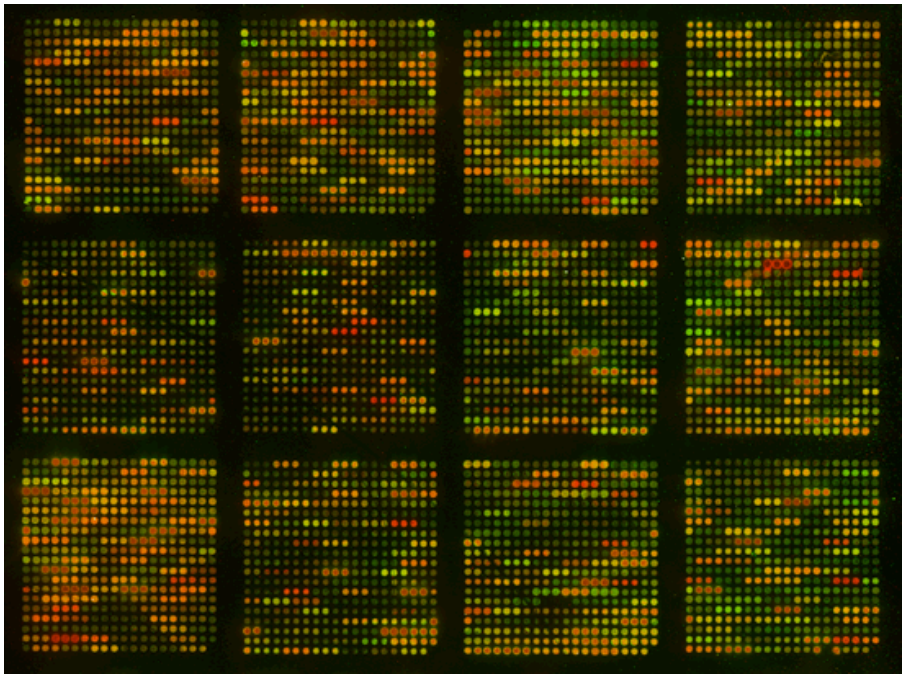


e.g. diagnosis of B-cell chronic lymphocytic leukemia

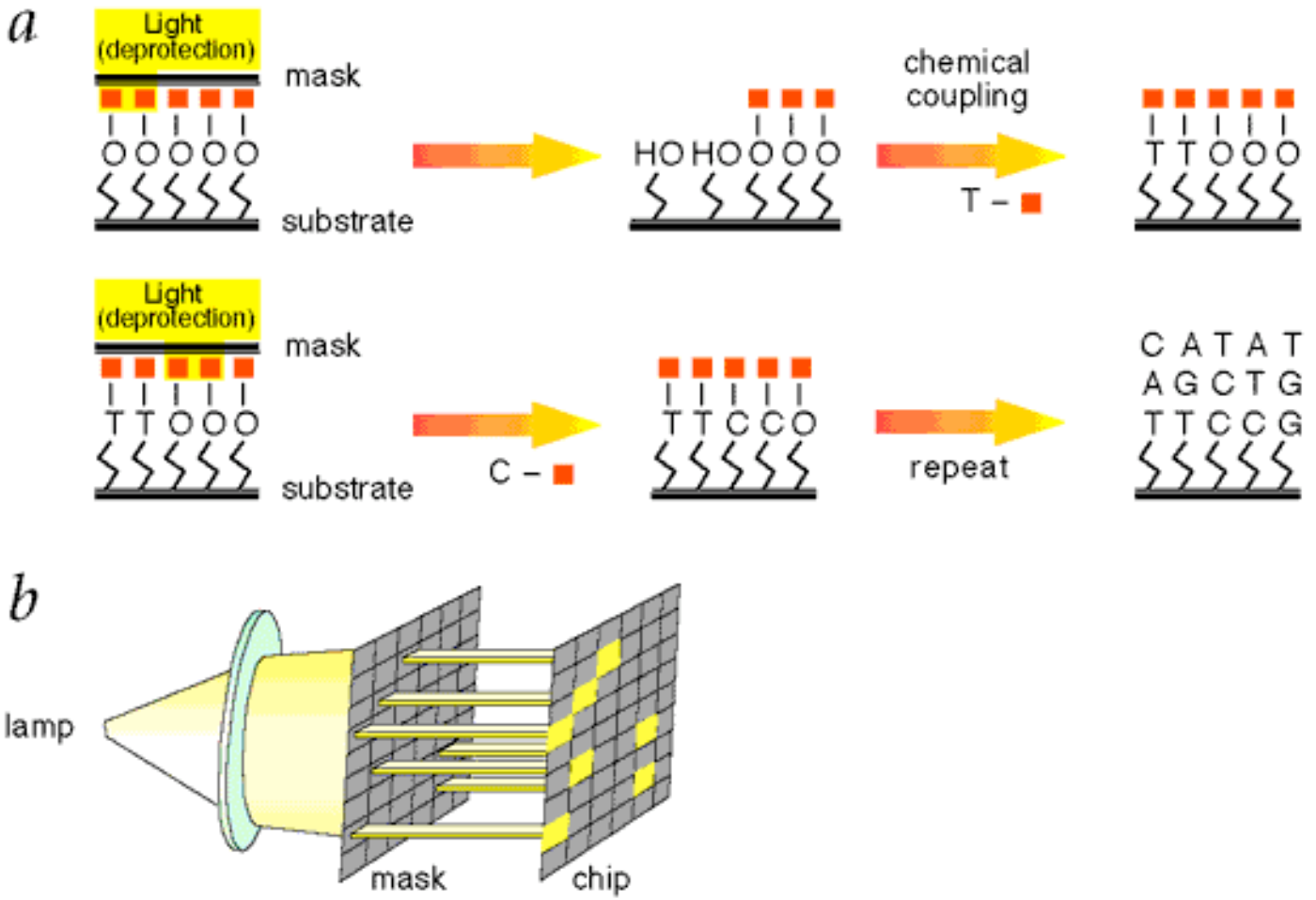


Gene Expression Profiling



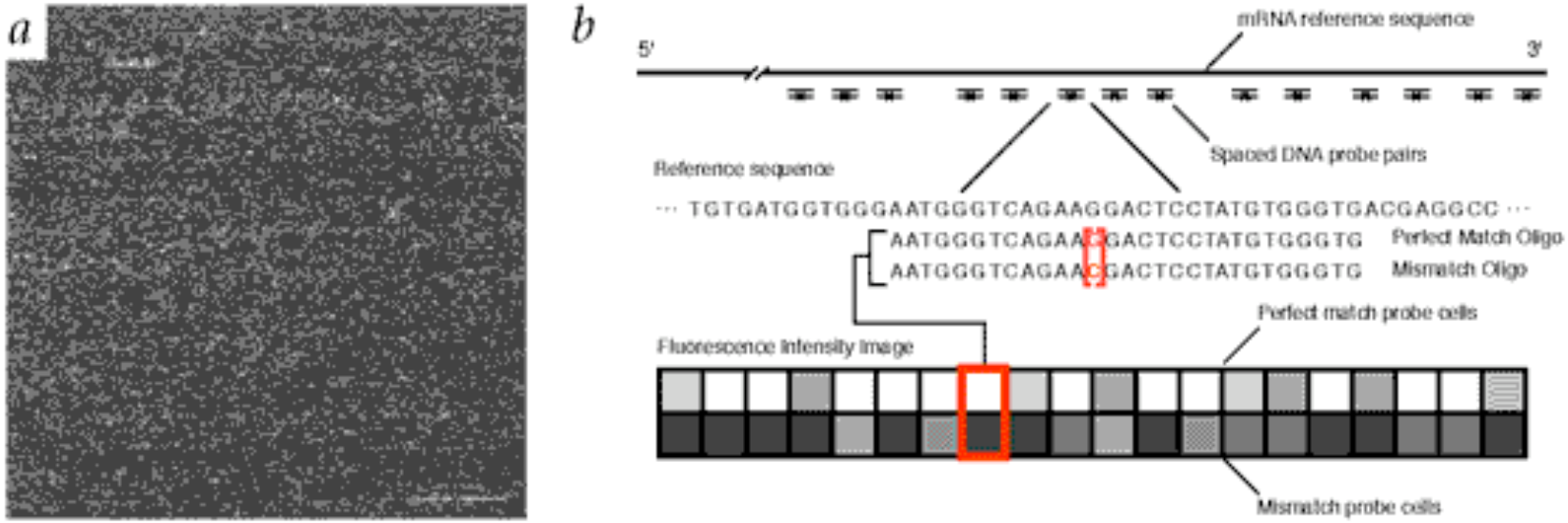


Affymetrix-Chips and CYP450

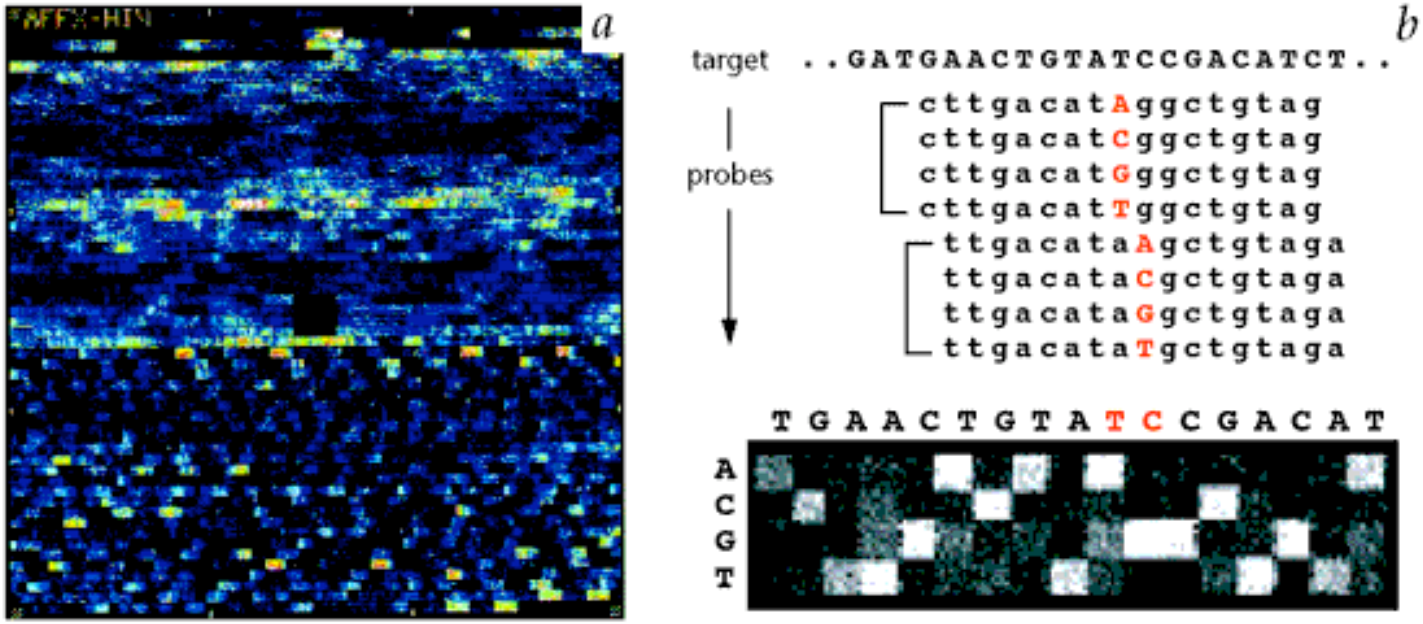


Light directed oligonucleotide synthesis.
A solid support is derivatized with a covalent linker molecule terminated with a photolabile protecting group. Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites. The process is repeated, activating different sets of sites and coupling different bases allowing arbitrary DNA probes to be constructed at each site. ***b***, Schematic representation of the lamp, mask and array.

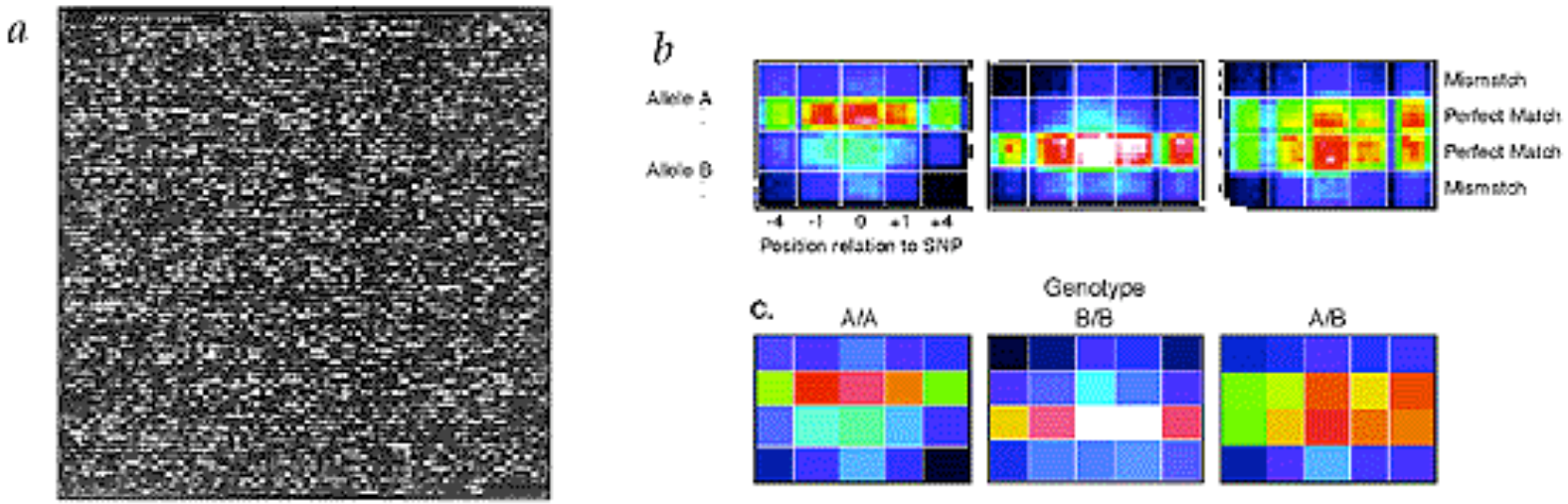
High density synthetic oligonucleotide arrays. Robert J. Lipshutz *et al.*, *Nature Genetics* **21**, 20 – 24 (1999)



Gene expression monitoring with oligonucleotide arrays.
a, A single 1.28 1.28 cm array containing probe sets for approximately 40,000 human genes and ESTs. This array contains features smaller than 22 22 m and only four probe pairs per gene or EST. ***b***, Expression probe and array design. Oligonucleotide probes are chosen based on uniqueness criteria and composition design rules. For eukaryotic organisms, probes are chosen typically from the 3′ end of the gene or transcript (nearer to the poly(A) tail) to reduce problems that may arise from the use of partially degraded mRNA. The use of the PM minus MM differences averaged across a set of probes greatly reduces the contribution of background and cross–hybridization and increases the quantitative accuracy and reproducibility of the measurements.



Sequence analysis arrays.
a, Image of an HIV–1 genotyping array (HIV PRT) hybridized to labelled PCR–generated DNA copies of the protease (PR) and reverse transcriptase (RT) genes (0.8 0.8 cm array). Each base on both strands covering a total of 1515 base pairs is interrogated in the HIV–1 PR (codons 1–99) and RT (codons 1–400) genes. Additionally, to ensure detection in the event of multiple mutations occurring in close proximity, most common drug–resistance conferring mutations are encoded on the array as specialized tilings. ***b***, General tiling strategy. Detection of mutations or polymorphisms in a sequence is accomplished by using a four–probe interrogation strategy. In this illustration, four 17–mer oligonucleotide probes are used to determine the identity of the base in the middle of the probe sequence. The probe that forms the most stable duplex will provide the highest fluorescent signal among the four probes assigned to interrogate the central base. The next nucleotide in the target sequence is interrogated in the same manner, using another set of four oligonucleotide probes. Probes with interrogation positions other than the central position, or probes of different lengths can also be used to query the targeted base. Analysis of both strands of a target can be carried out on the same array to increase the confidence of the base determination.



Genotyping arrays.
a, A single array with over 120,000 probes designed to determine the genotype of a sample at over 3,000 biallelic loci. ***b***, The fluorescence intensity pattern for a set of probes designed to interrogate a single locus showing the presence of an AA homozygote, a BB heterozygote, and a BB homozygote. The upper and lower halves of the probe blocks interrogate the A and B alleles, respectively. Each half consists of pairs of probes centered on the polymorphic position and offset one and four bases to either side. The probe pairs consist of a perfect match and single base mismatch to the reference sequence for the specific allele. For each locus, interrogation blocks are included for both the sense and anti–sense strands.

Table 2 • Commercial GeneChip® probe arrays[a]

Application	Species	Information
expression	human[b]	~42,000 genes/ESTs
	mouse[b]	~30,000 genes/ESTs
	rat[c]	>11,000 genes/ESTs
	yeast (*S. cerevisiae*)	whole genome (all ORFs)
	Drosophila melanogaster[c]	>12,000 genes/ESTs
	Arabidopsis thaliana[c]	>12,000 genes/ESTs
	C. elegans[c]	whole genome (all ORFs)
	E. coli[c]	whole genome (all ORFs)
	other bacteria[d]	whole genome (all ORFs)
	targeted[e]	functionally selected gene sets
	custom[f]	any eukaryotic organism
genotyping	human	~2,000 SNPs
polymorphism screening	human	screening service
variant analysis	human	*CYP450* (2D6, 2C19)
	human	p53 (exons 2–11)
	HIV-1, Clade B	HIV (protease, rev. transcriptase)

[a]The GeneChip® System is required to run the arrays. The complete system is priced at $175,000 in the U.S. and includes the GeneChip® Fluidics Station, Hewlett Packard GeneArray™ Scanner, GeneChip® workstation and GeneChip® 3.1 analysis software. [b]Array update including additional gene and EST information expected in 1999. [c]Expected to be available mid-1999. [d]Five bacterial species planned for late 1999. [e]Specific subsets of genes selected based on their relevance to a given biological, clinical or disease area. [f]Gene or EST sets (public or proprietary) chosen by external users.

Table 3 • Array capacity and feature size

Feature Size	Expression[a]	Sequence Analysis[b]	Genotyping[c]
50 μm	1600–6400 genes	8–16 kb	2,000–4,000 markers
20 μm	10,000–50,000 genes	50–100 kb	12,000–25,000 markers
2 μm	>1 million genes	500–1,000 kb	1.2–2.5 $\times 10^6$ markers

All numbers calculated for 1.28×1.28 cm arrays. [a]Assuming 4–20 probe pairs per gene. [b]Assuming 4–8 probes per basepair. [c]Assuming 6–32 probes per marker.

AmpliChip CYP450

komplexe und umfassende Analyse von

29 *CYP2D6*-Varianten sowie

2 *CYP2C19*-Varianten

in einem einzigen Assay

Roche Diagnostics: erste Firma, welche

die Affymetrix Mikroarray-Technologie

für eine diagnostische Fragestellung nutzt

***CYP2D6*-kodierte Enzyme metabolisieren:**

• zahlreiche Anti-Depresiva

• Anti-Psychotika

•Anti-Arrhytmika

•Schmerzmittel

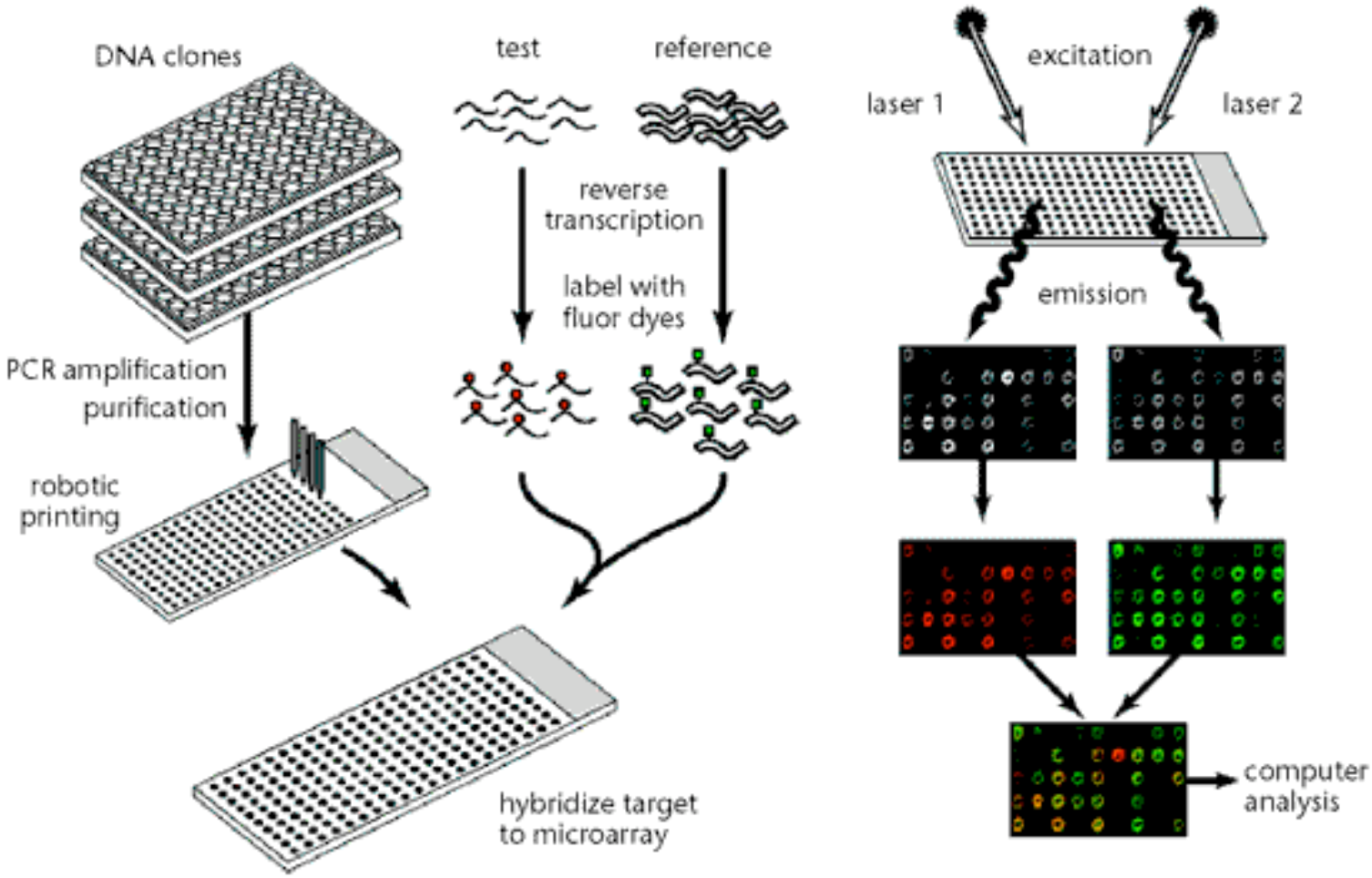
•beta-Blocker

***CYP2C19*-kodierte Enzyme:**

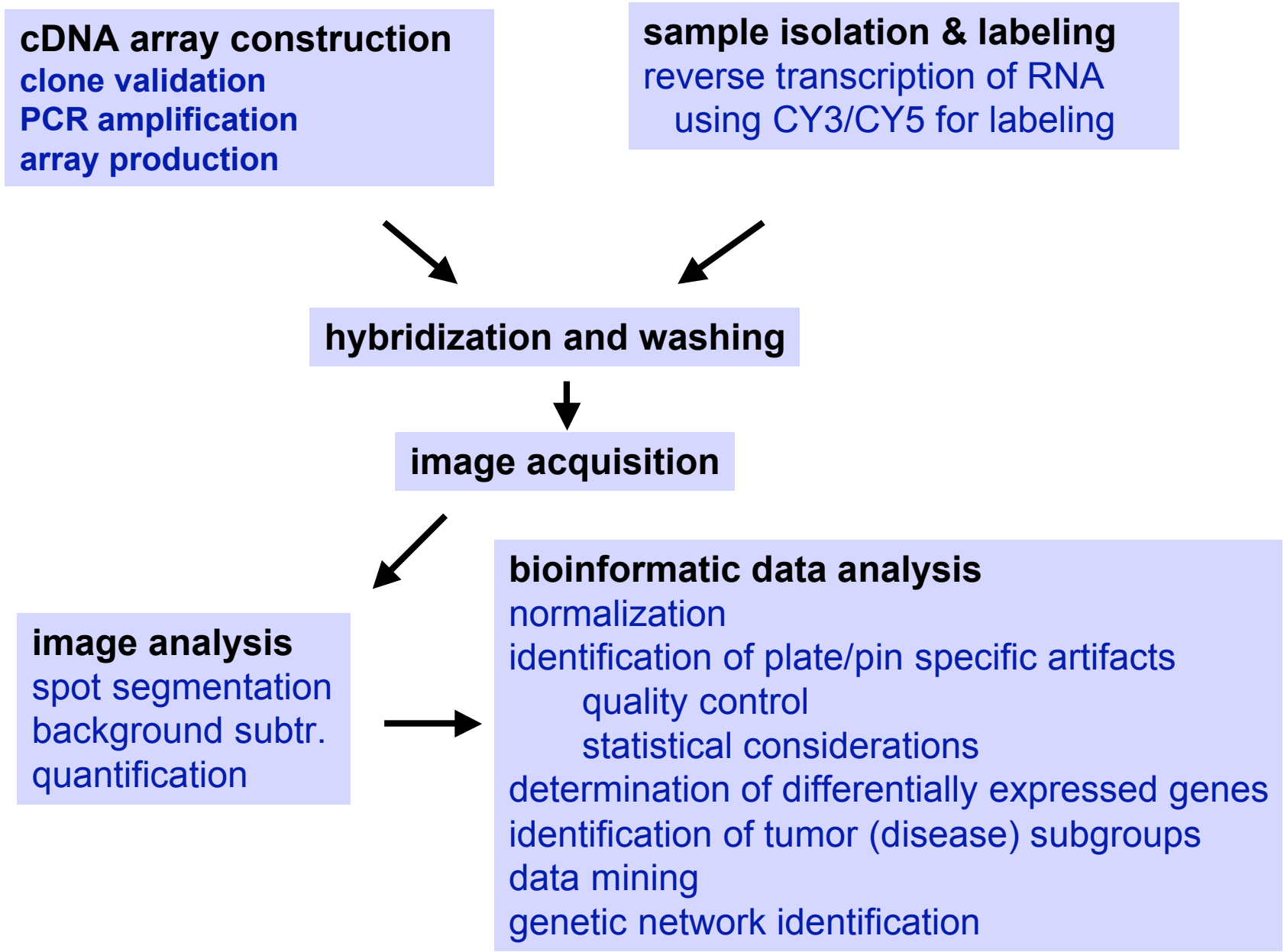
metabolisieren diverse Substanzen, u.a. Protonenpumpen-Inhibitoren, Anti-Koagulantien, Benzodiazepine, Anti-Malariamittel

Making microarrays

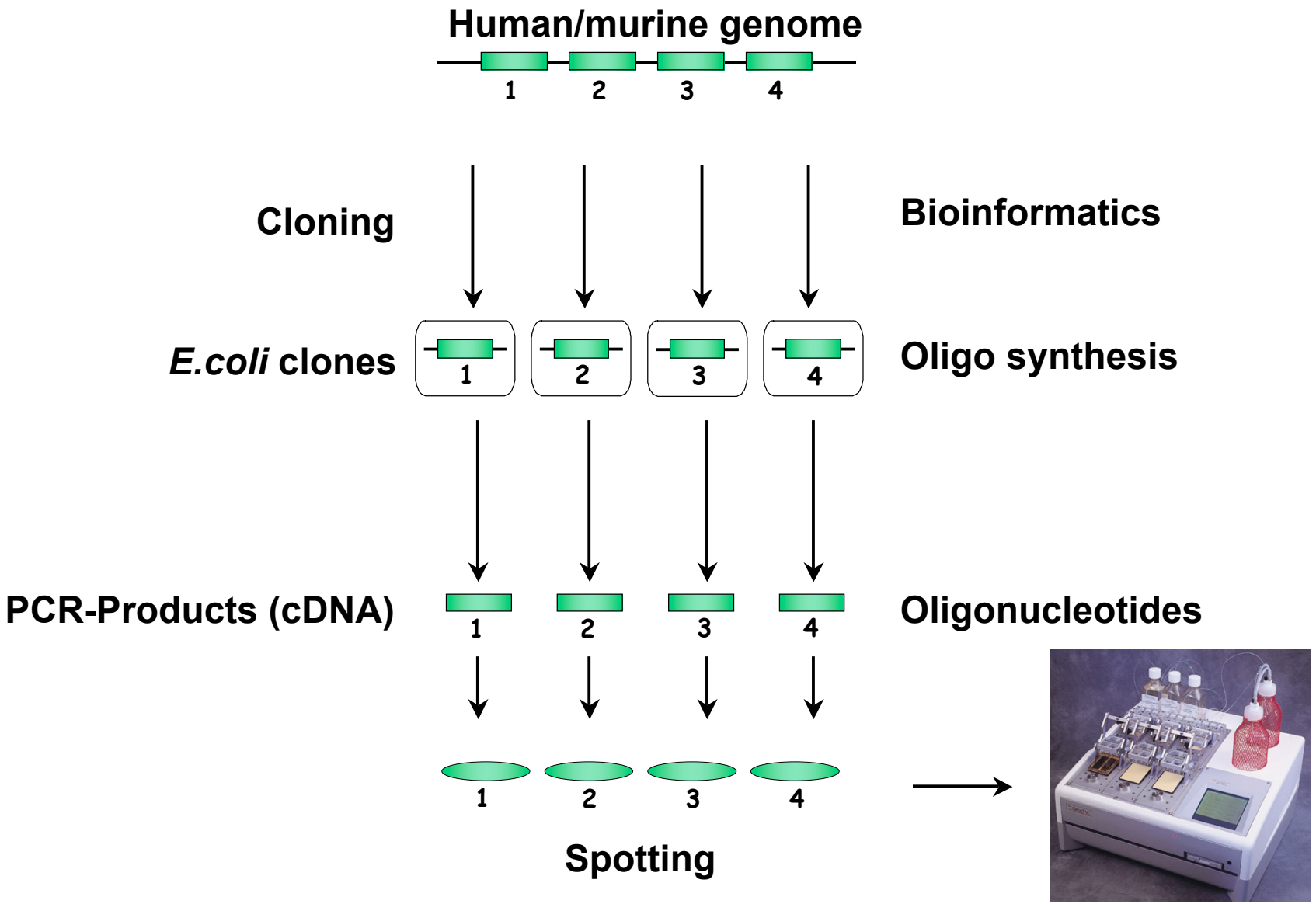
The Principle of cDNA Microarrays



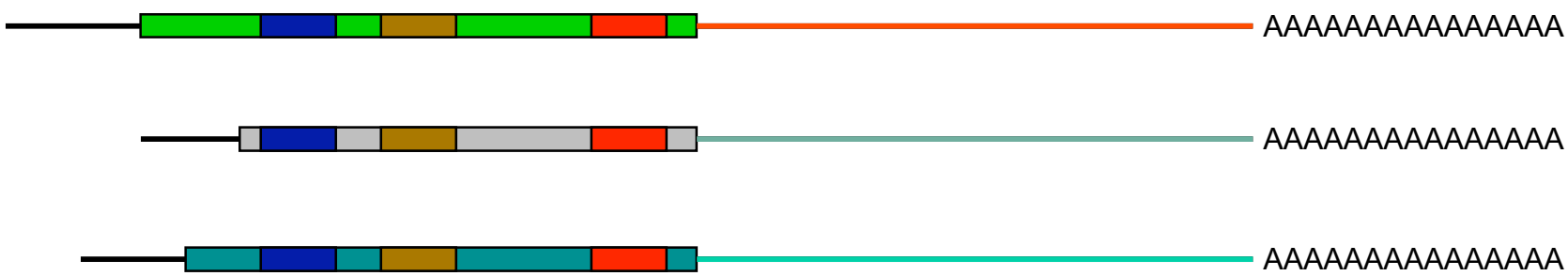
Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet*. 1999, 21(1 Suppl):10-14.


cDNA array construction
clone validation
PCR amplification
array production
sample isolation & labeling
reverse transcription of RNA
using CY3/CY5 for labeling
hybridization and washing
image acquisition
image analysis
spot segmentation
background subtr.
quantification
bioinformatic data analysis
normalization
identification of plate/pin specific artifacts
quality control
statistical considerations
determination of differentially expressed genes
identification of tumor (disease) subgroups
data mining
genetic network identification

Expression Profiling



arrayTAGs:
Optimized cDNA Fragments for Microarrays



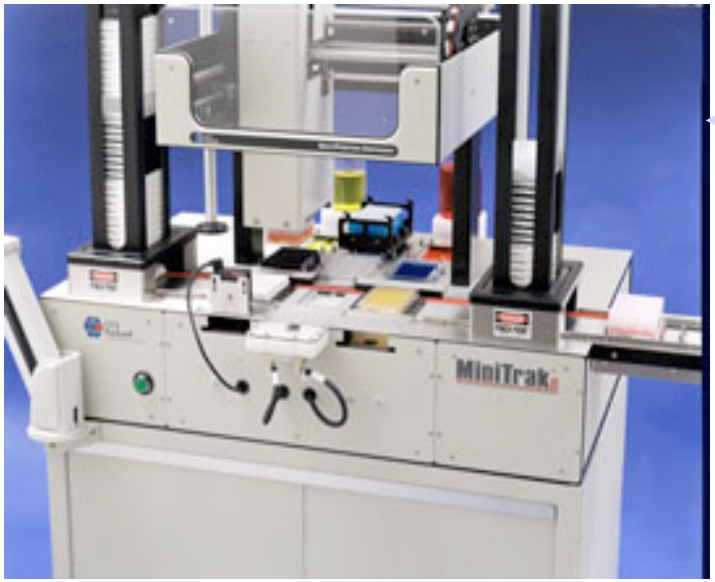
arrayTAG fragments

LION Bioscience

10,080 Clones

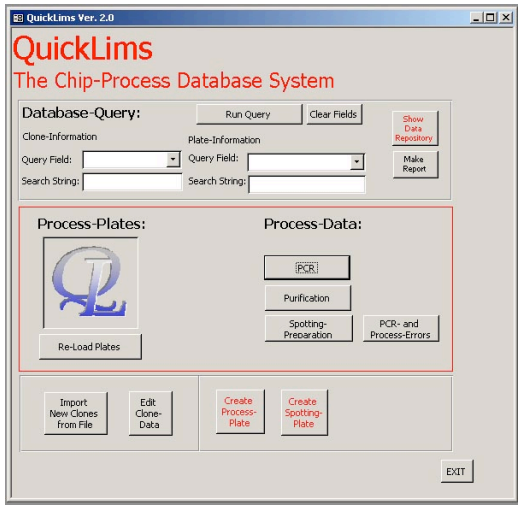
- **homogenous size (~ 500bp)**
- **no poly A - tail**
- **no repetitive elements**
- **3´UTR exclusively**

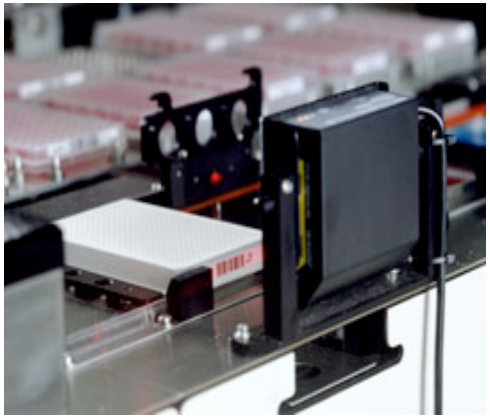
The Liquid Handling System I



Read Barcode

Request plate information from QL

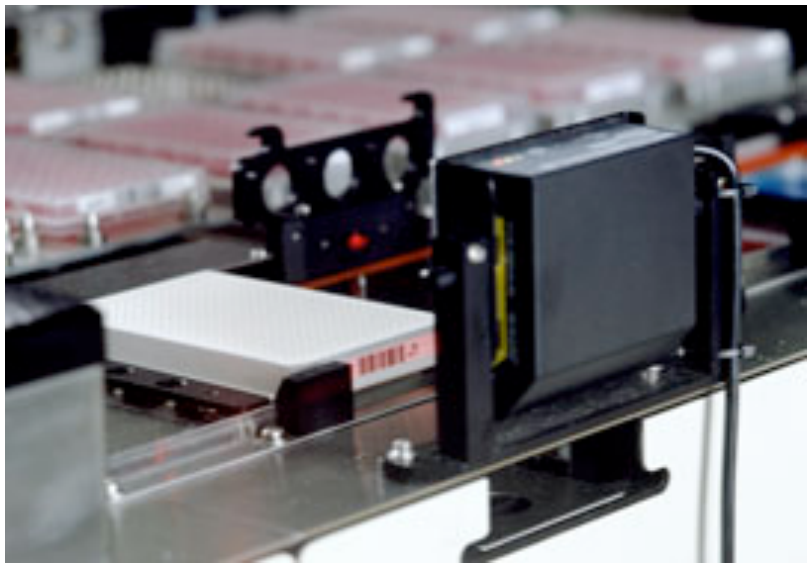




Pipet according to step information (provided by QL)

The Liquid Handling System II

Packard Bioscience MiniTrak conveyor-based robotic liquid handling system

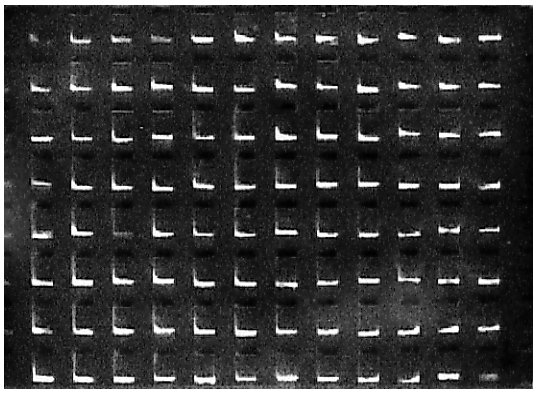


The robot uses a bar code reader to identify individual process plates. This enables the VB control script, which is connected to a custom database system, to monitor, automate and document the whole production process

PCR Process

MiniTrak Liquid Handling System

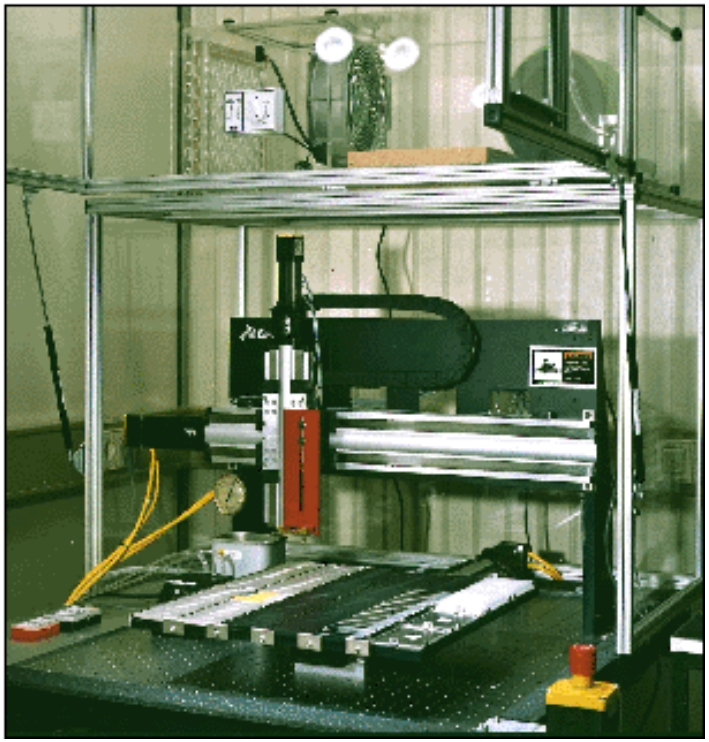
96 Well Steel Replicator



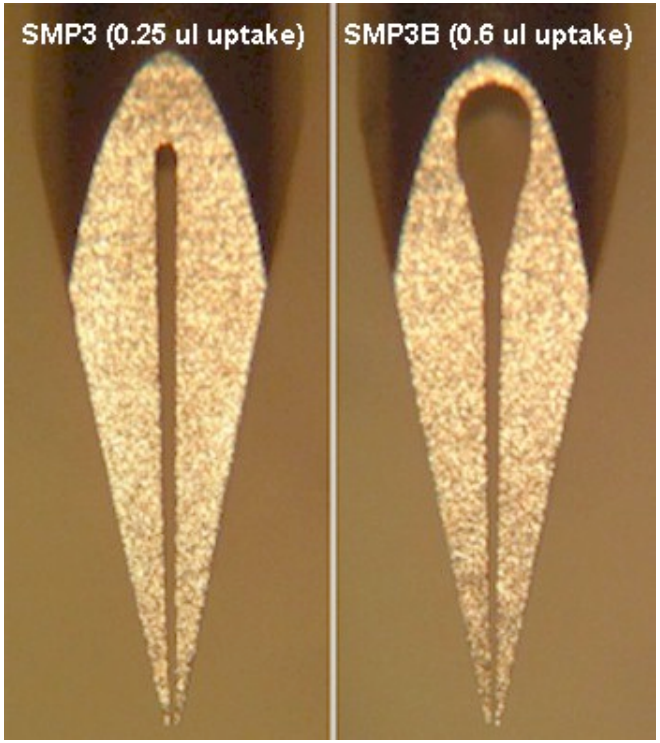
96-well Ready-to-Run Agarose Gel

4 x 4 Block PCR Machine (96 Wells)

Production of cDNA Microarrays I



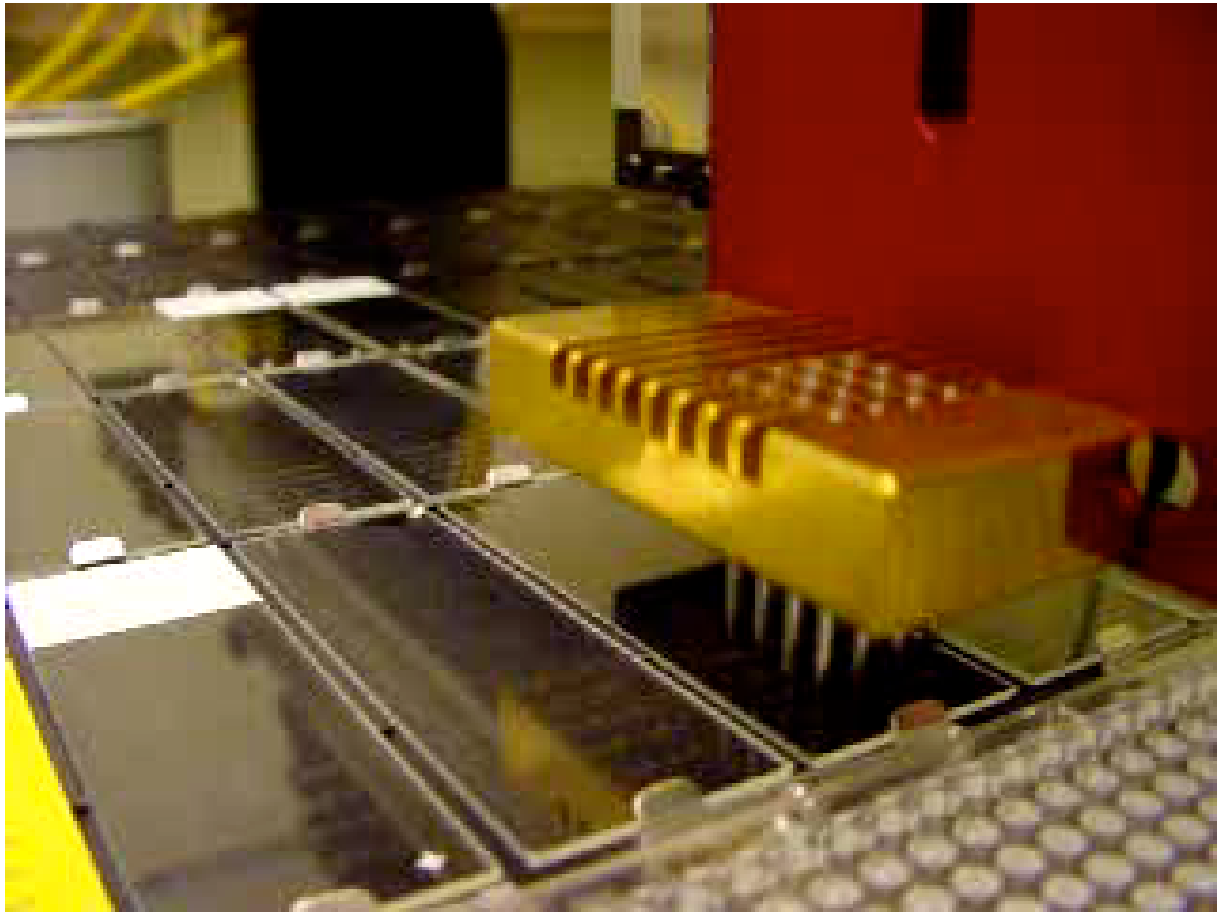
GeneMachines OmniGrid multi-axis microarrayer



Telechem Stealth Micro Spotting Pins SMP3 and SMP3B (split / capillary pins)

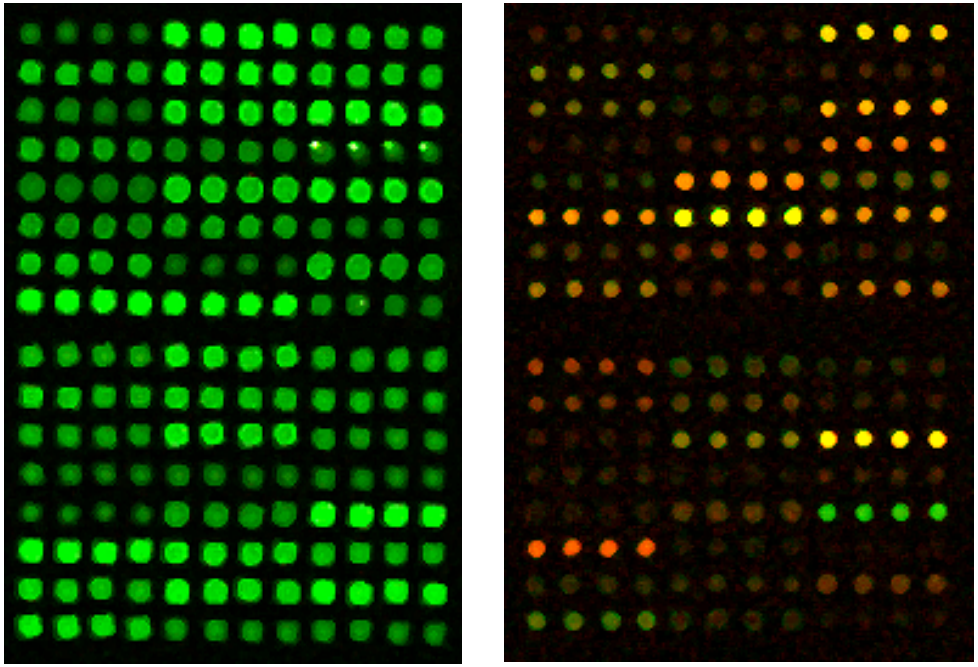
Production of cDNA Microarrays II

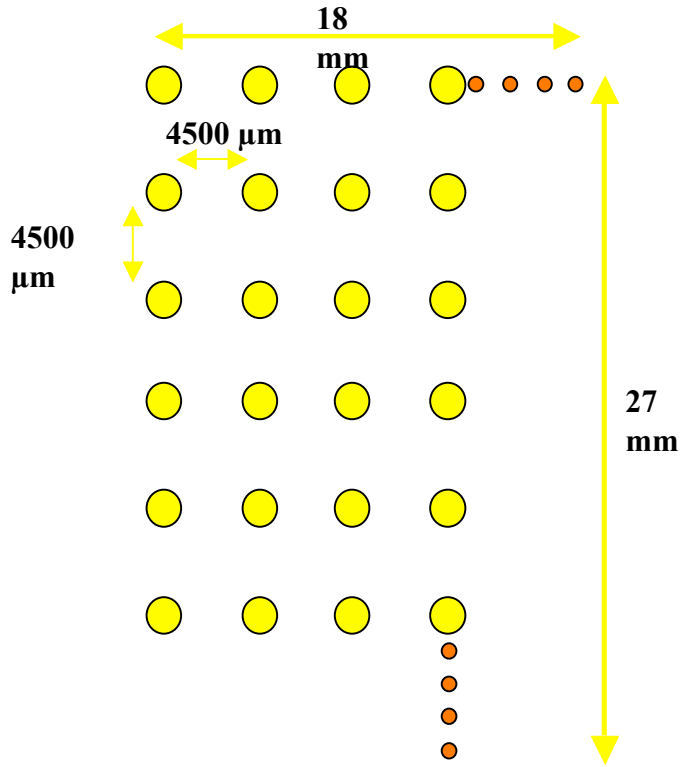
Production of cDNA Microarrays III



Spots at 140 µm Dot Spacing

- Spotting can be performed at 140 µm dot spacing
- Printable area (for HybStation): 18 x 54 mm
- $(4500 : 140)^2$ x 4 x 6 x 1 x 2 = 49,590 spots
- 24,900 clones as replicate array, 24 pins (Virtek Microarrayer)





Hybridization of Microarrays

TeleChem hybridization cassette

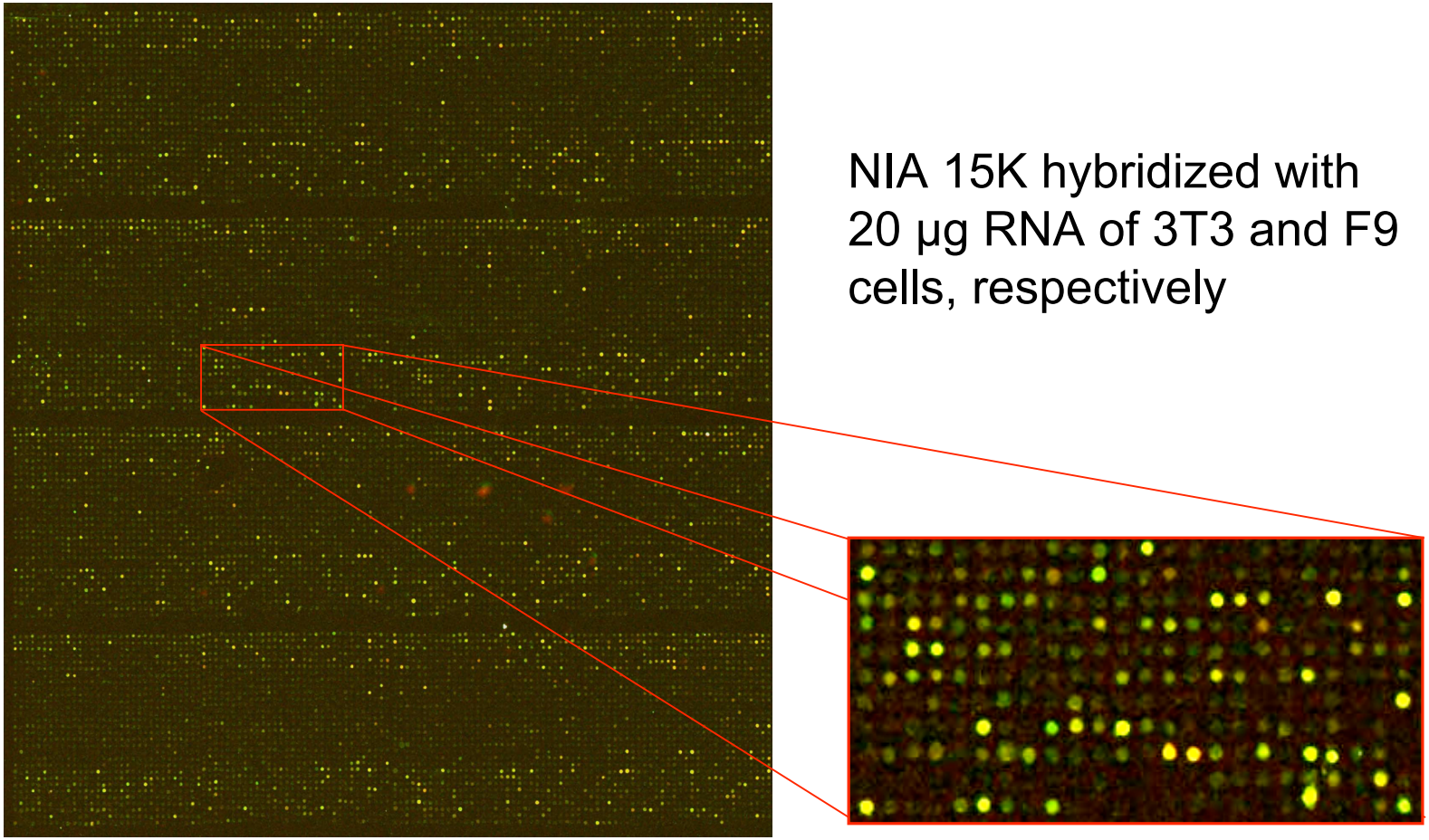
Genomic Solutions GeneTAC Hyb-Station (ASP)

Scanning of Microarrays

Axon GenePix 4000A Microarray Scanner

- **Resolution: 10 µm / pixel**
- **2 laser-PMT units**
- **Detection @ 532 nm (Cy3) and 635 nm (Cy5)**
- **Speed: 3-4 min /slide**

What it looks like...



NIA 15K hybridized with 20 µg RNA of 3T3 and F9 cells, respectively