主要的基因芯片技术平台

Affymetrix Gene Chip

Agilent DNA microarray

Illumina Sentrix™ Arrays

博奥的芯片平台

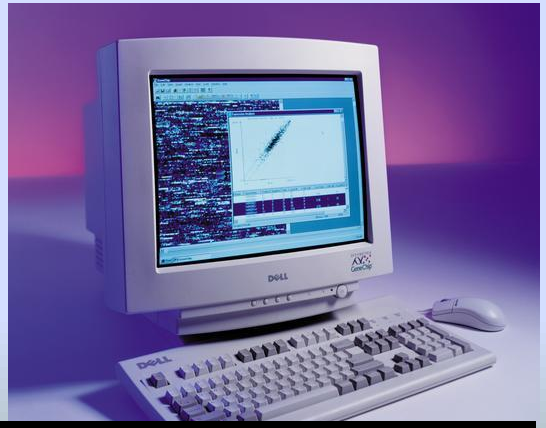
Affymetrix 基因芯片

预制的基因芯片

芯片滚动杂交仪

全自动芯片洗涤工作站

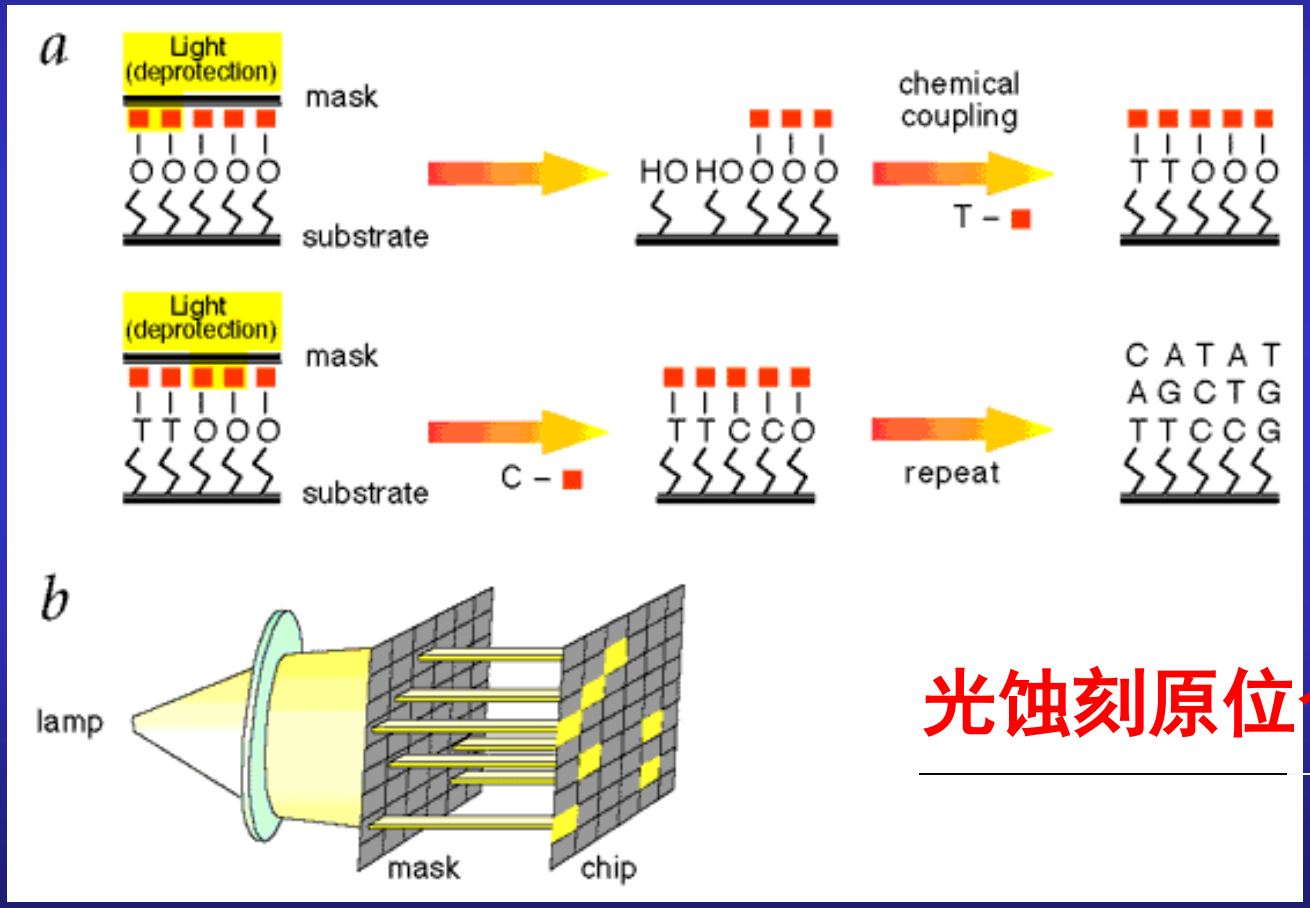
高分辨率的芯片扫描仪



智能化的分析软件

Affymetrix 基因芯片的特点（1）

光蚀刻原位合成

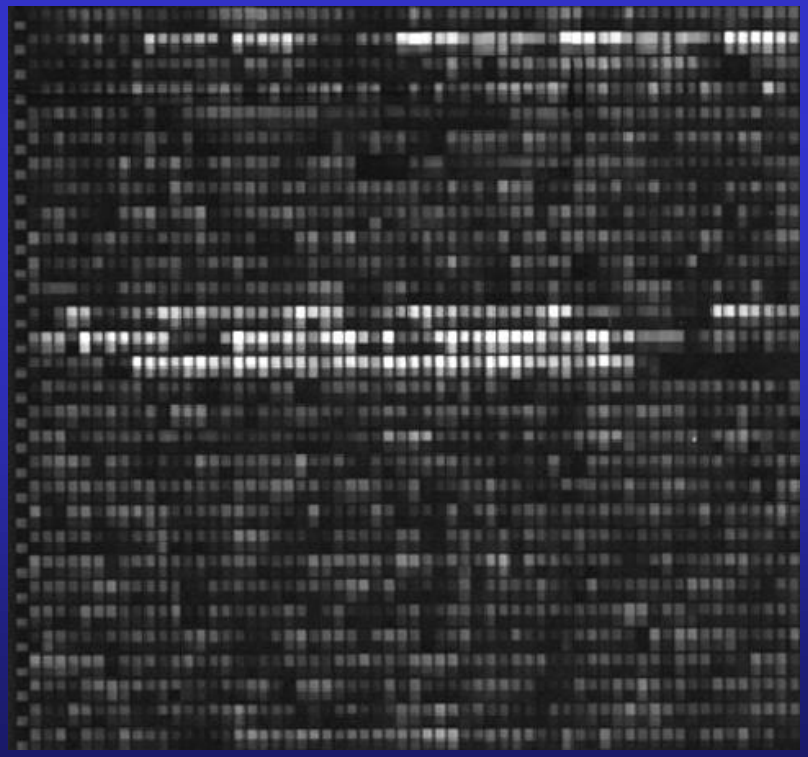


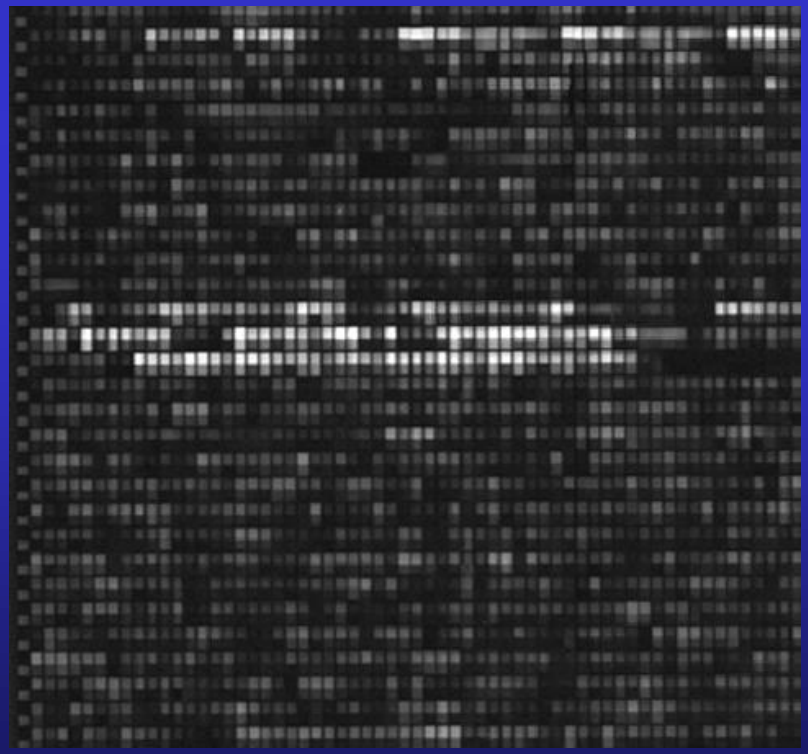
光蚀刻原位合成技术

Affymetrix 基因芯片的特点（2）

高密度的点阵技术

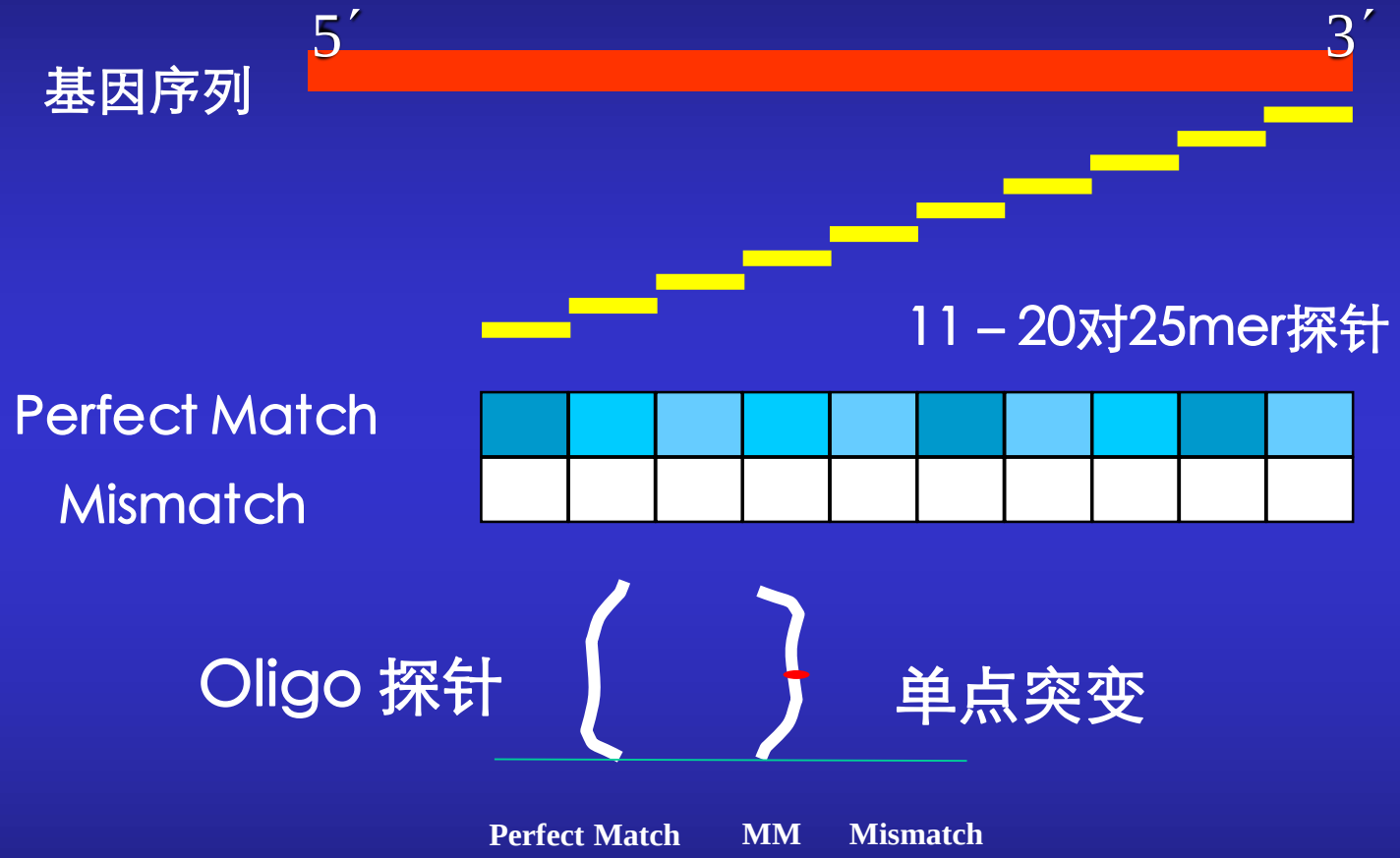
1 个平方厘米的面积至少可排列几十万到一百多万个探针合成区（“点”）





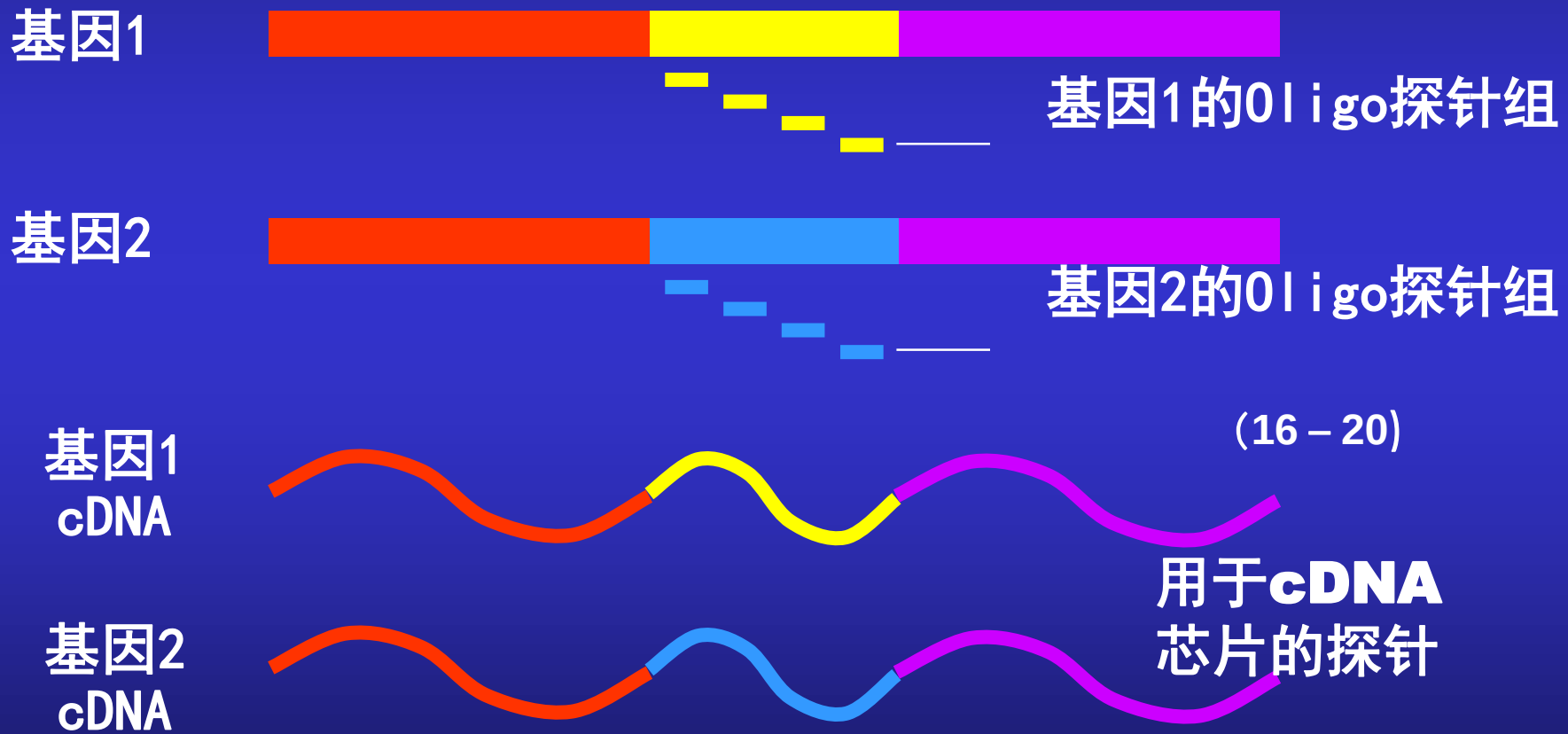
Affymetrix 基因芯片的特点（3）

Affymertix独特的PM – MM探针设计



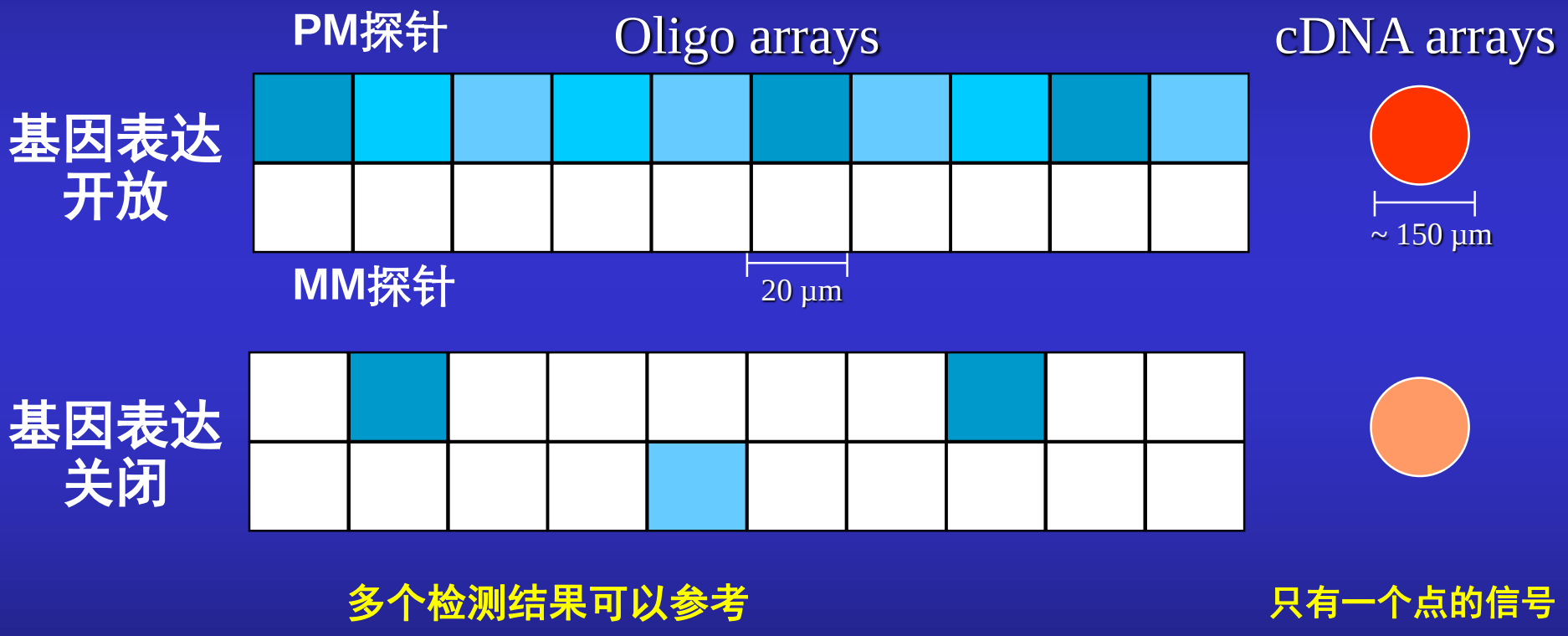
Affymetrix 基因芯片的特点（4）

特异性高



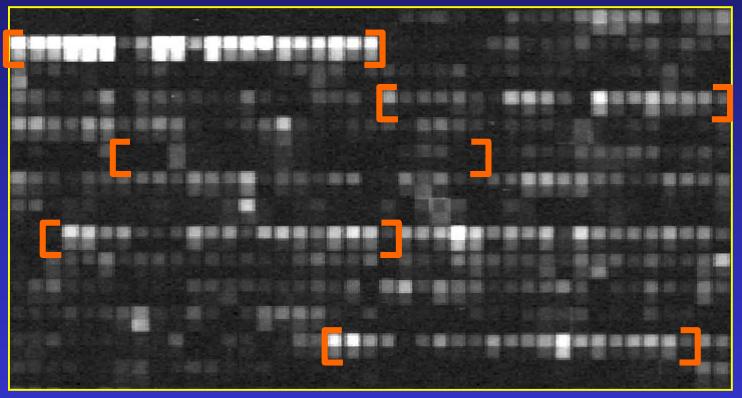
Affymetrix 基因芯片的特点（5）

芯片结果准确可靠

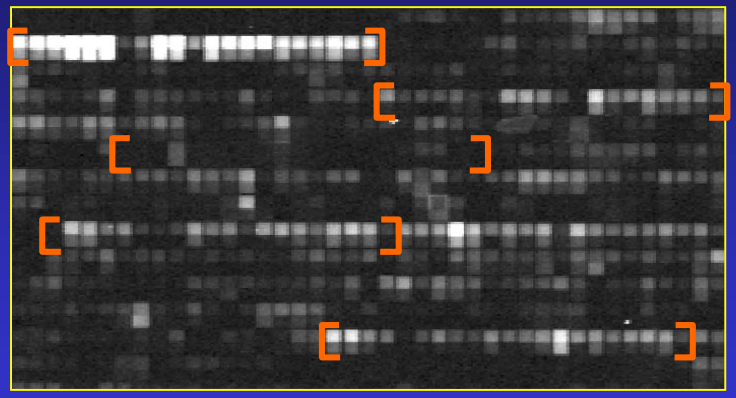


Affymetrix 基因芯片的特点（6）

重复性高



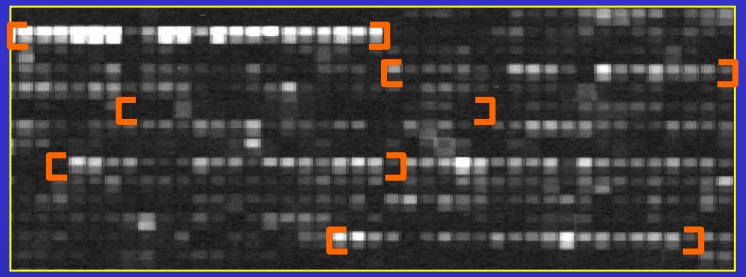
Signal Intensities: 13,400



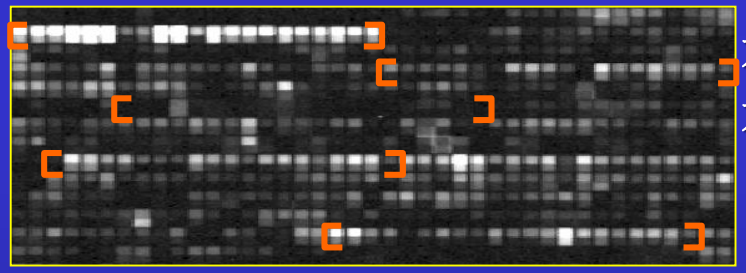
Signal Intensities: 13,090

同一样本，

不同批次的芯片



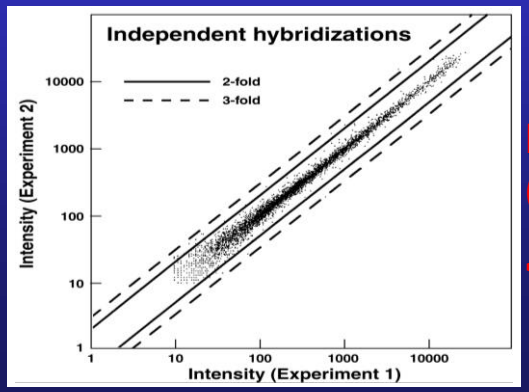
Signal Intensities:13,400



Signal Intensities: 11,670

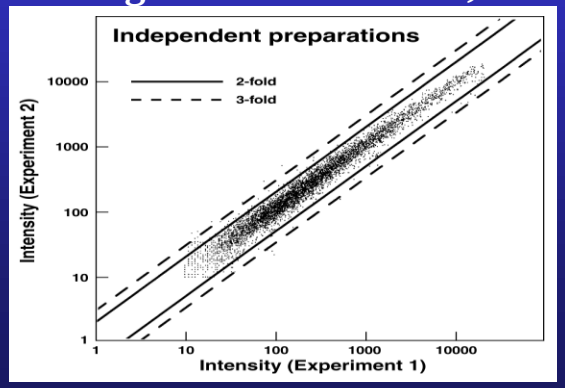
不同的样本，

不同批次的芯片



False I/D
(≥ 2 fold)

< 0.3%

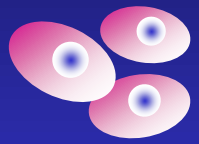


False I/D
(≥ 2 fold)

< 1%

GeneChip®的操作流程

以真核生物为例



总RNA的制备

起始用量5-10ug

AAAA

反转录

cDNA

体外转录（IVT）

生物素标记的cRNA

片段化处理

带标记的cRNA片断

35-200 bases

0.5-2 ug/ul

Control Oligo B2

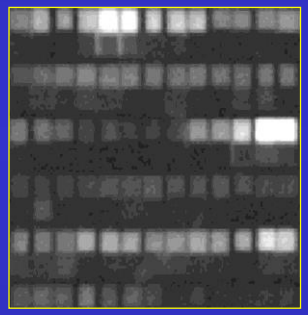
Eukaryotic Hyb.Control

杂交混合液的制备

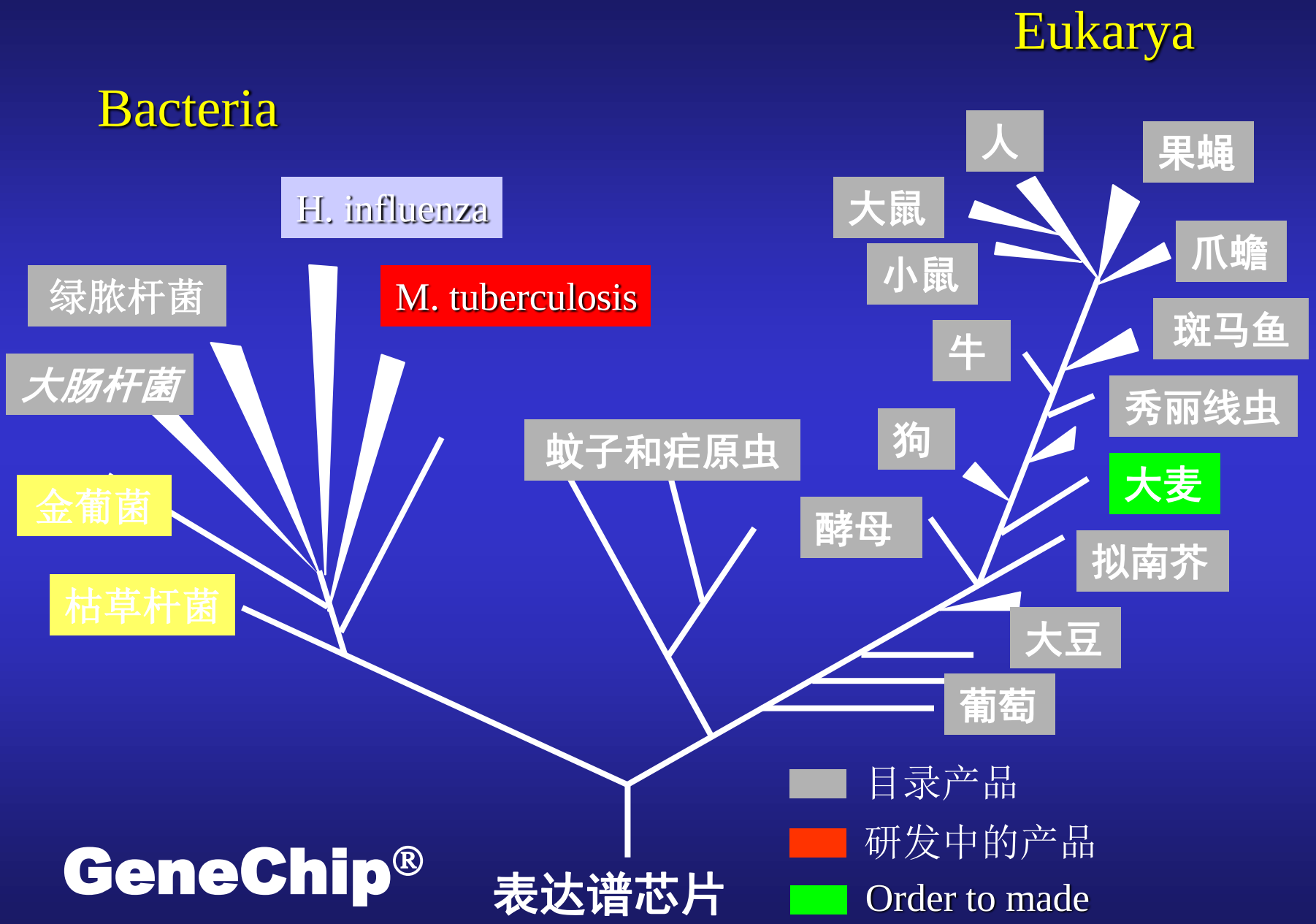
杂交

洗涤

扫描

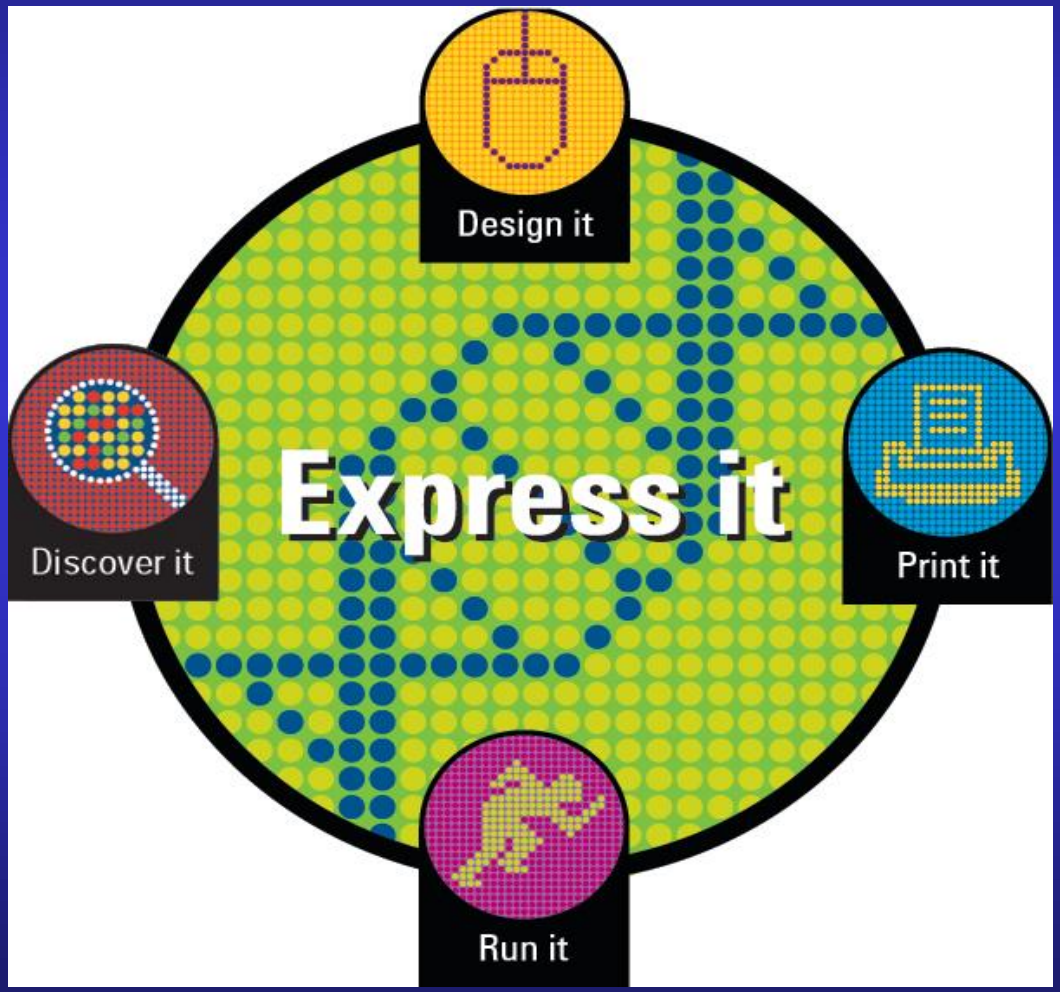


数据分析


Eukarya
Bacteria
人
果蝇
大鼠
爪蟾
小鼠
斑马鱼
牛
秀丽线虫
狗
大麦
酵母
拟南芥
大豆
葡萄
H. influenza
绿脓杆菌
M. tuberculosis
大肠杆菌
蚊子和疟原虫
金葡菌
枯草杆菌
目录产品
研发中的产品
Order to made
GeneChip®
表达谱芯片

主要的基因芯片技术平台（2）

Agilent Gene Expression Plateform



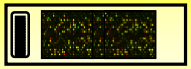
Total Solutions for Gene Expression

- **Total Solutions**
  - **Provide quality tools for gene expression analysis: Microarray, Labeling, Scanner, Analysis Software.**
- **Open Platform**
  - **Can be used for existent system.**
  - **Or, can be used as a whole system.**

Optimized Experiment Protocol

cDNA Array

Oligo Array

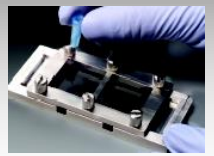


Custom Array

Sample Check

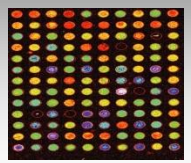
Labeling

Hybridization & Wash

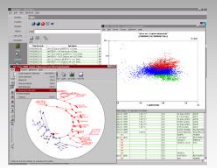


Scanner

Spot Analysis Software

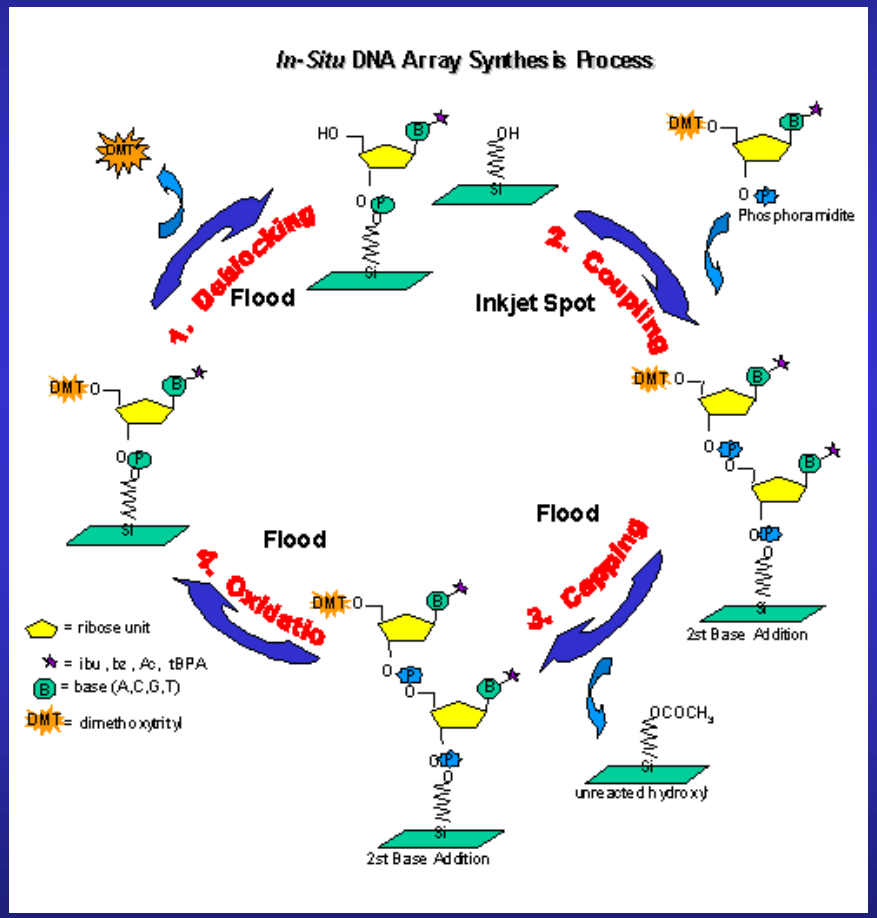


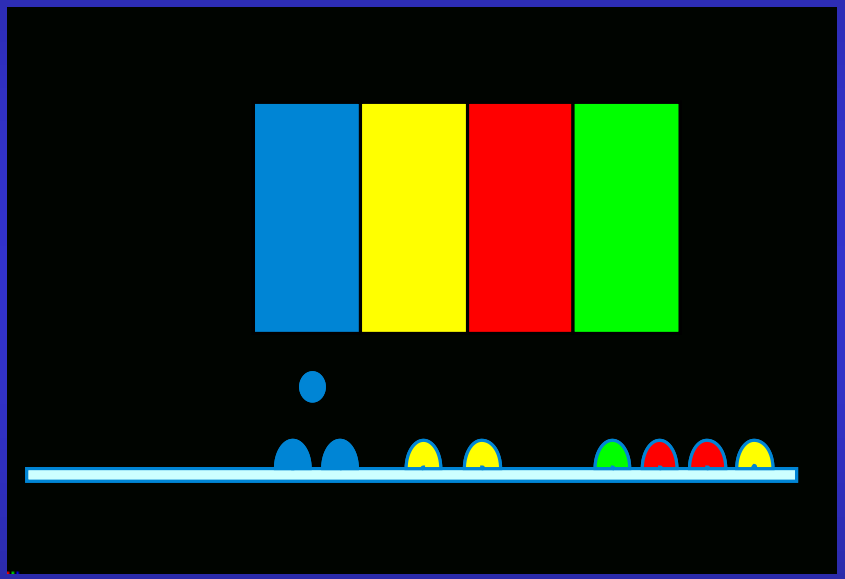
Data Analysis Data Base



In Situ Oligo Synthesis

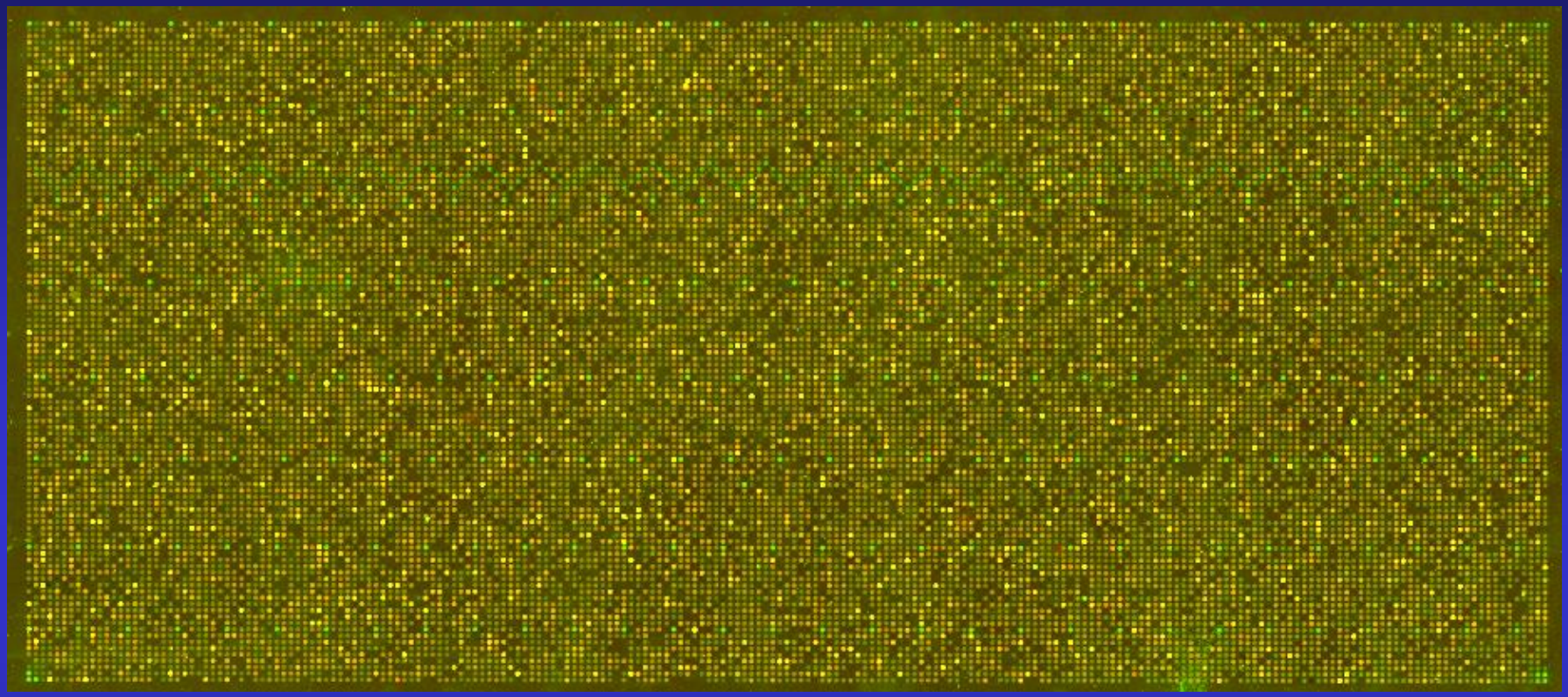
Phosphoamidite Reaction





Glass Substrate

In Situ Microarray: Microarrays & Formats

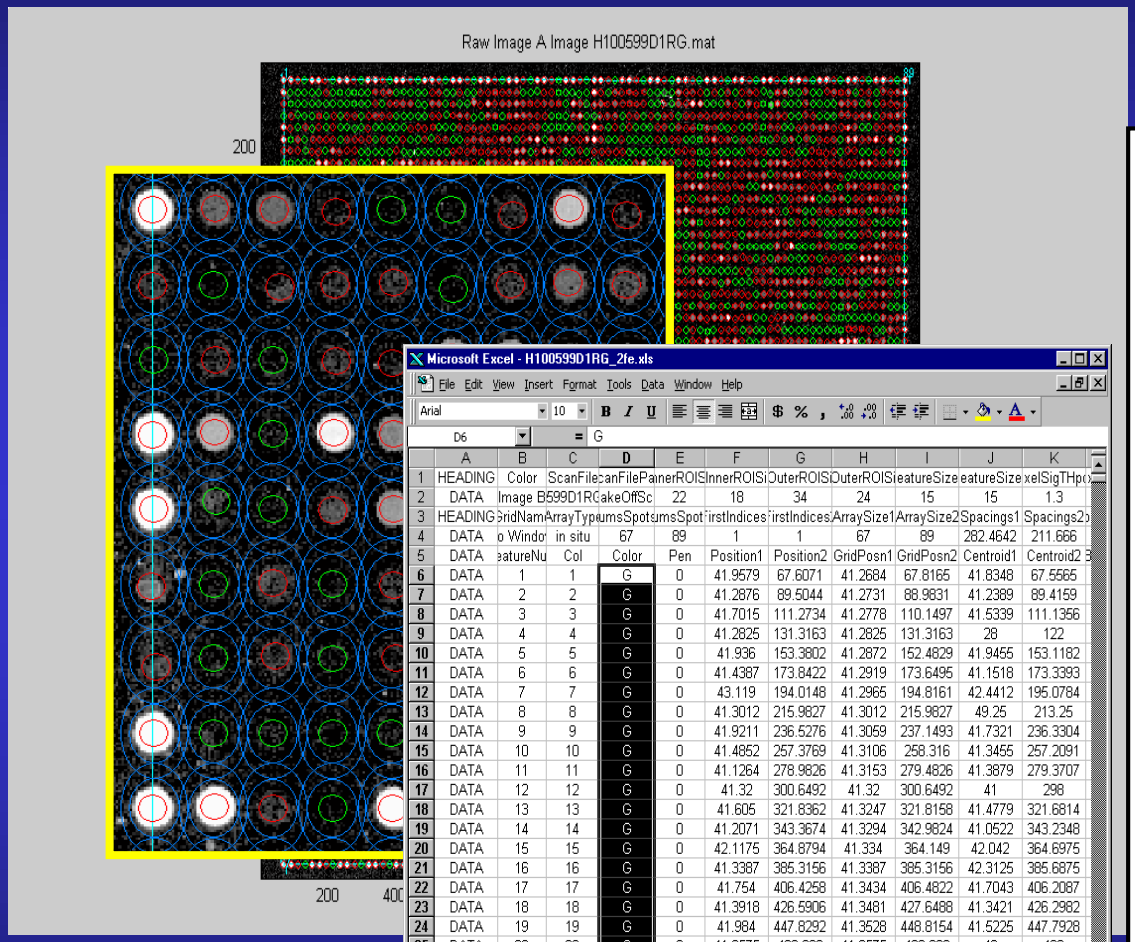


In situ Oligonucleotide
- 60 mer

formats:
- 15 k array
- 44k array
- 60k array
-

-180K array
-244K array
- 400K array
-1 M array

Automated Analysis and Feature Extraction

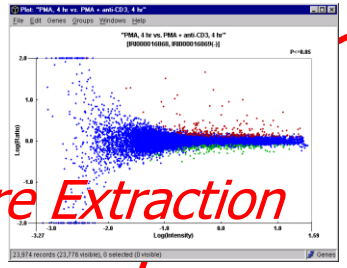


Feature Extraction Software

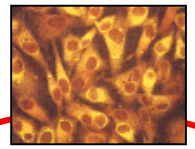
- Rapid
- Automated feature finding
- Intra-feature pixel statistics
- Inter-feature statistics
- Flexible dye normalization
- Statistical significance of signals and ratios

Agilent's Gene Expression Workflow

- Rosetta bioinformatics solutions both Resolver and Luminator for gene expression analysis



~1 Min for Feature Extraction



- QA/QC of RNA Integrity

30 min to run 12 RNA Samples

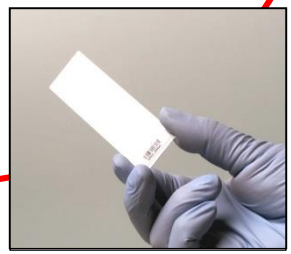
- Most high-throughput and sensitive Microarray scanner on the market

7 Min Scan

Total Solution for Gene Expression

- 50ng of total RNA for sensitive labeling solutions

6 hr labeling rxn



- Superior Sensitivity of 60mer oligo microarray

- Sure Hybe chamber for easy manipulation

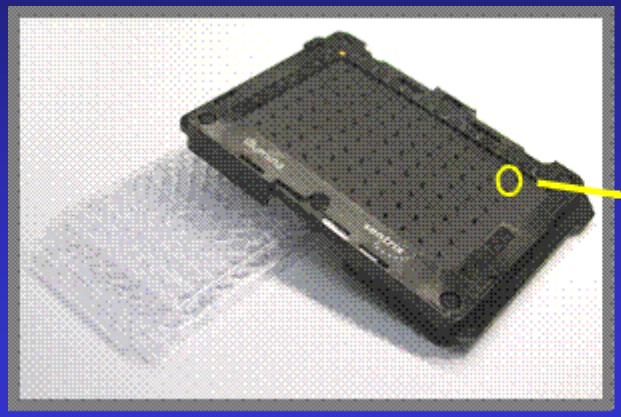
17 hr hybridization

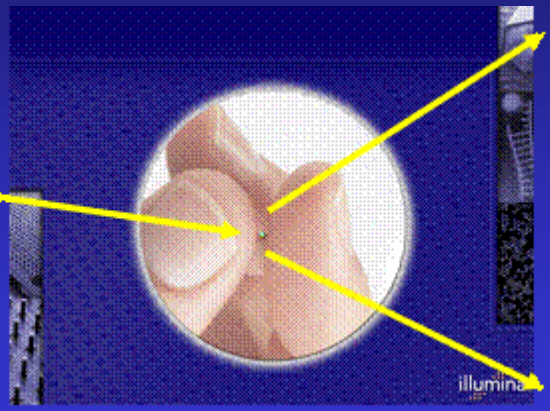
30 Min Frag RXN and Clean-up

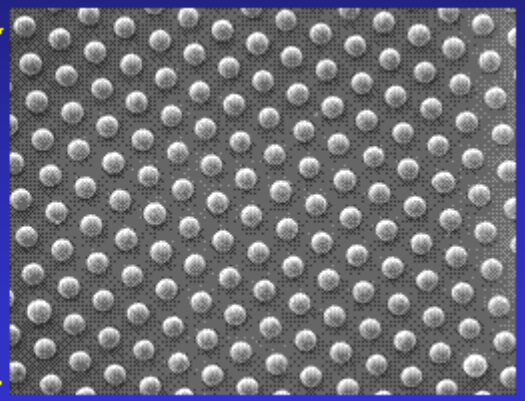
Agilent Technologies

主要的基因芯片技术平台（3）

Illumina Sentrix™ Arrays

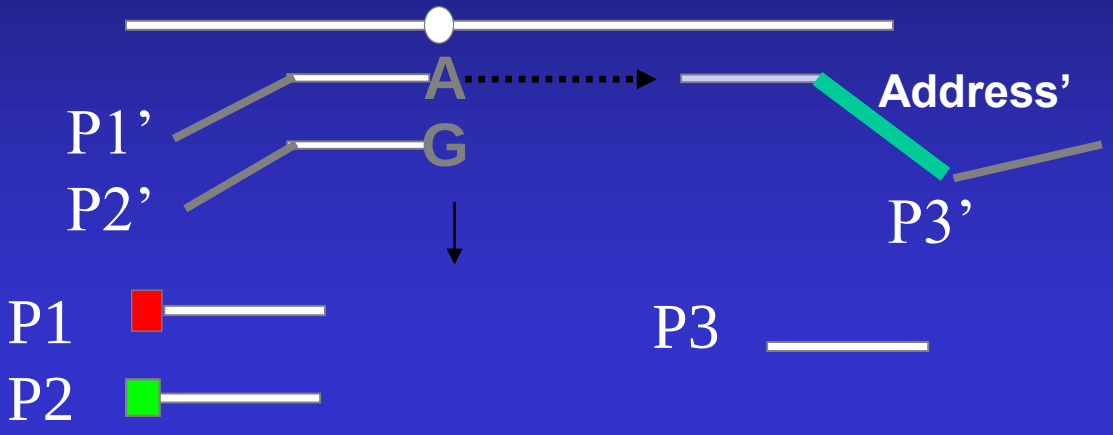






- BeadArray™
  - 96 arrays configured to standard format to enable high throughput
  - Parallel processing of 96 samples
  - Up to 150,000 assays per array matrix
- One array—an optical fiber bundle
  - 1.5 mm in diameter
  - Containing 50,000 individual optical fibers
- Beads on each bundle—an optical fiber bundle
  - Each bead is positioned to the end of an optical fiber
  - Bead diameter ~3 mm
  - Bead center-to-center distance ~5 mm—highest feature density
  - Bead surface is derivatized with DNA probes

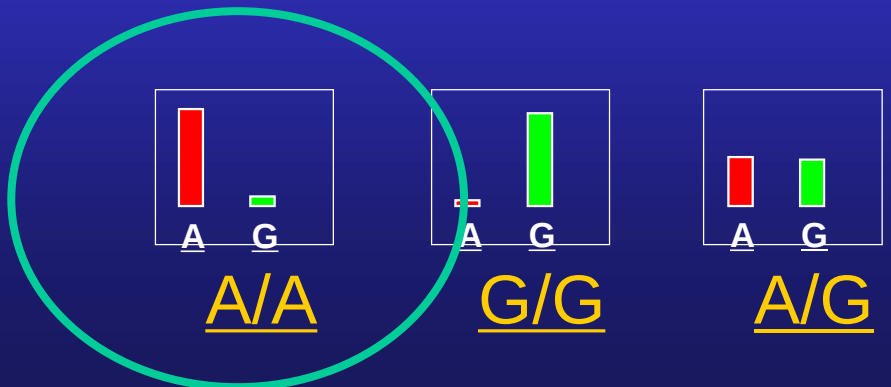
Allele Specific Extension Assay



Allele Specific Extension

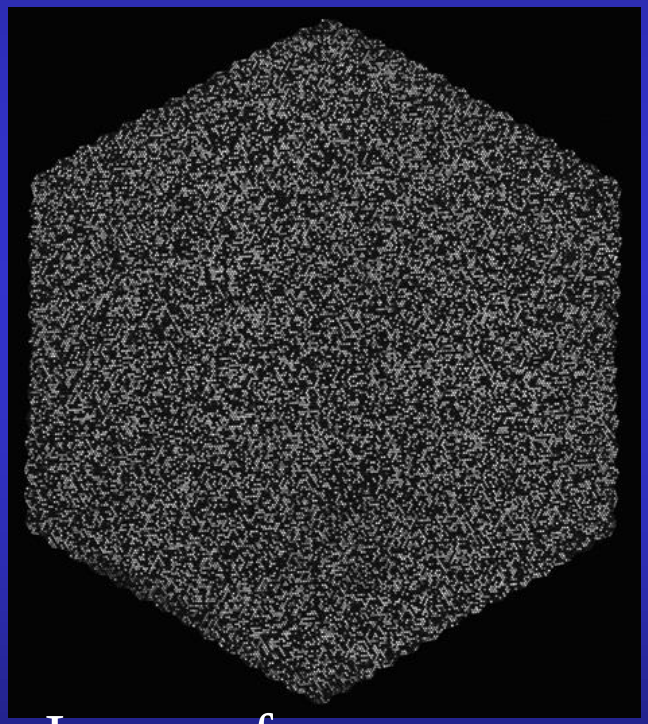
PCR with common primers

Product capture by hybridization to array

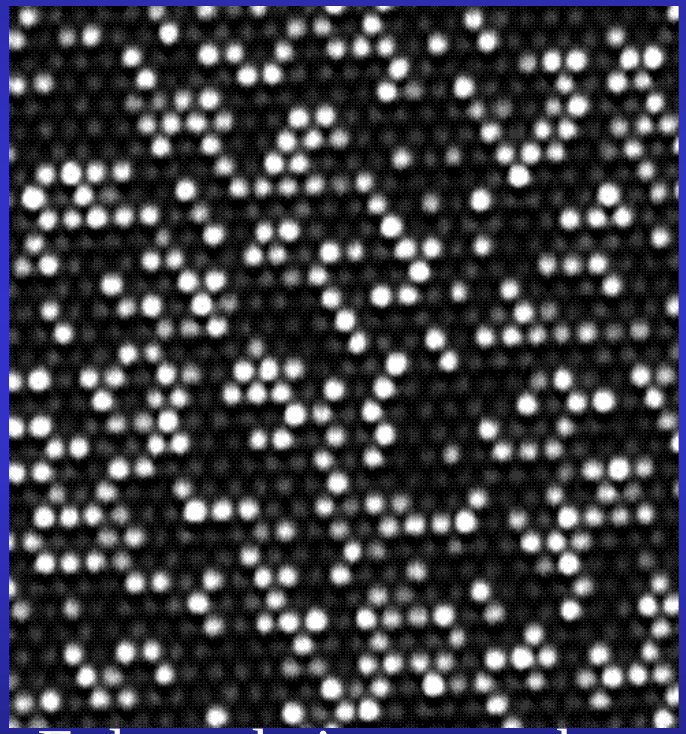


Readout

Confocal Scanning Produces Superior Image Quality

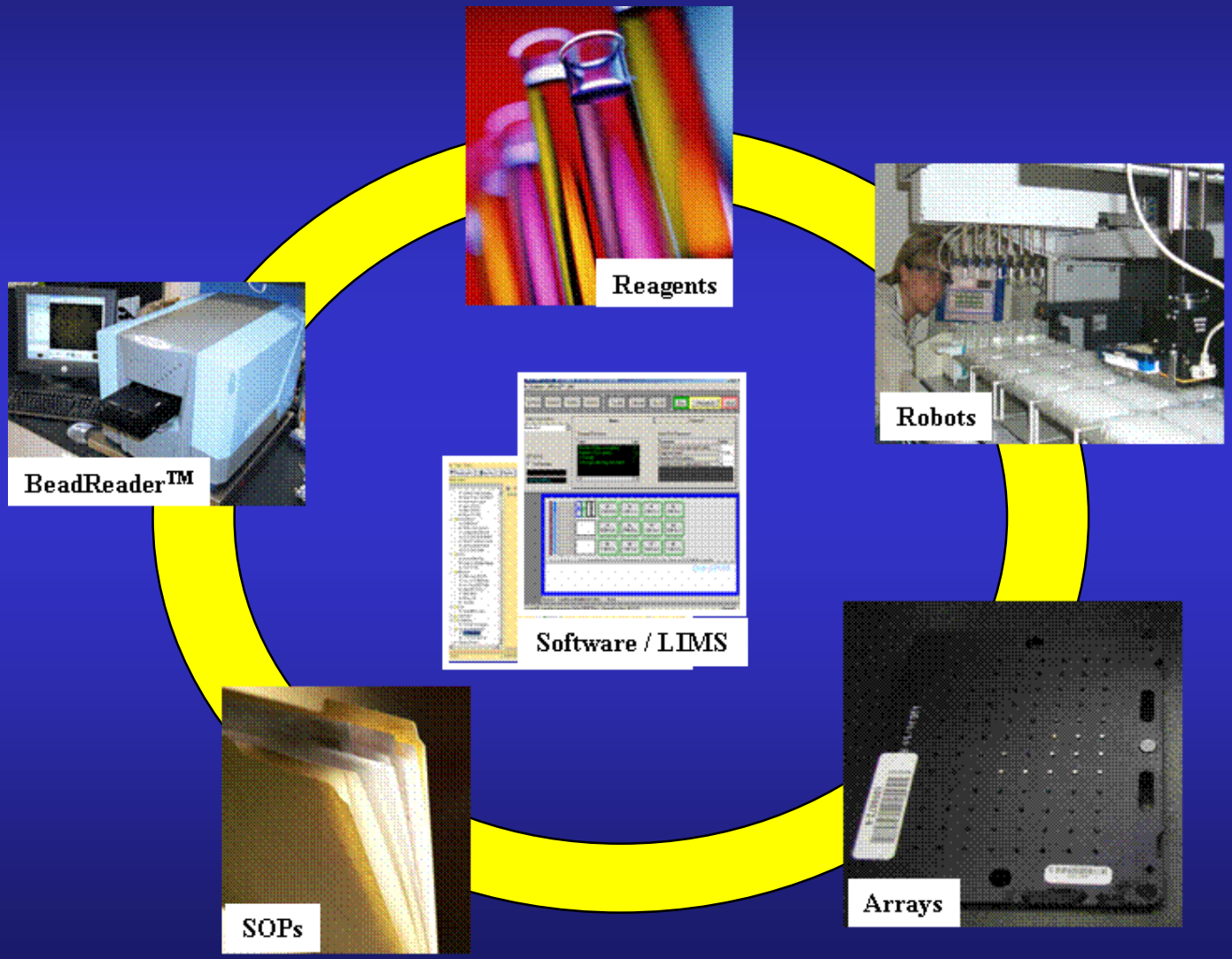


- Image of one BeadArray™ bundle



- Enlarged view—each bright spot is a bead with fluorescent signal

BeadLab™—Integrated Genotyping Laboratory

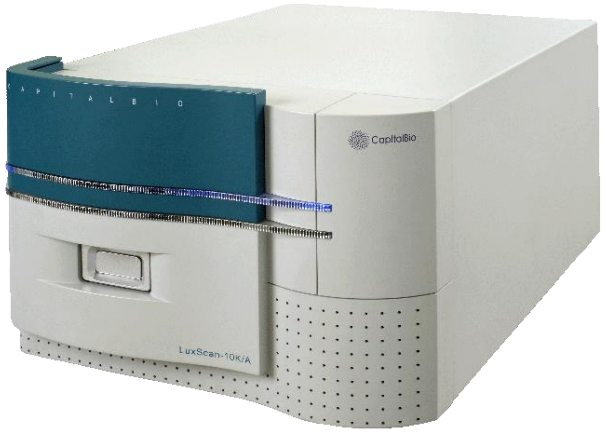


Lab Automation &
Lab Management

博奥的芯片平台

点样仪

杂交仪

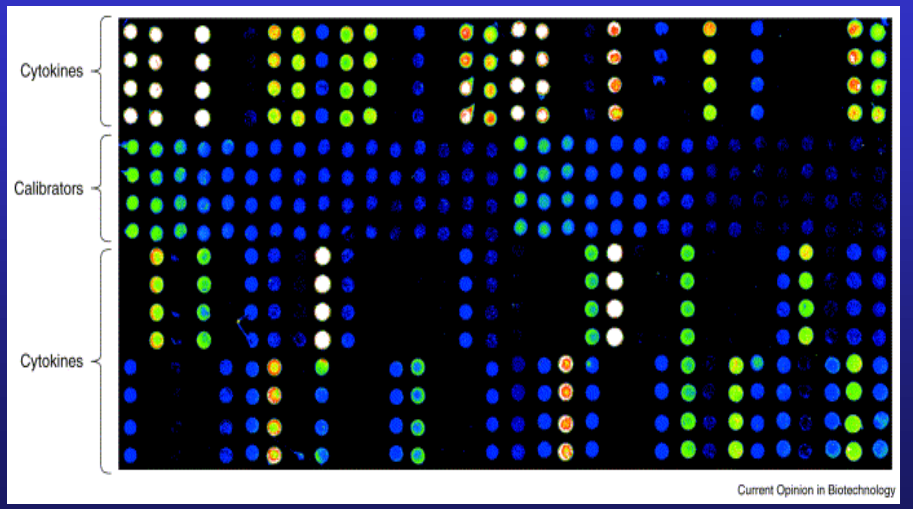


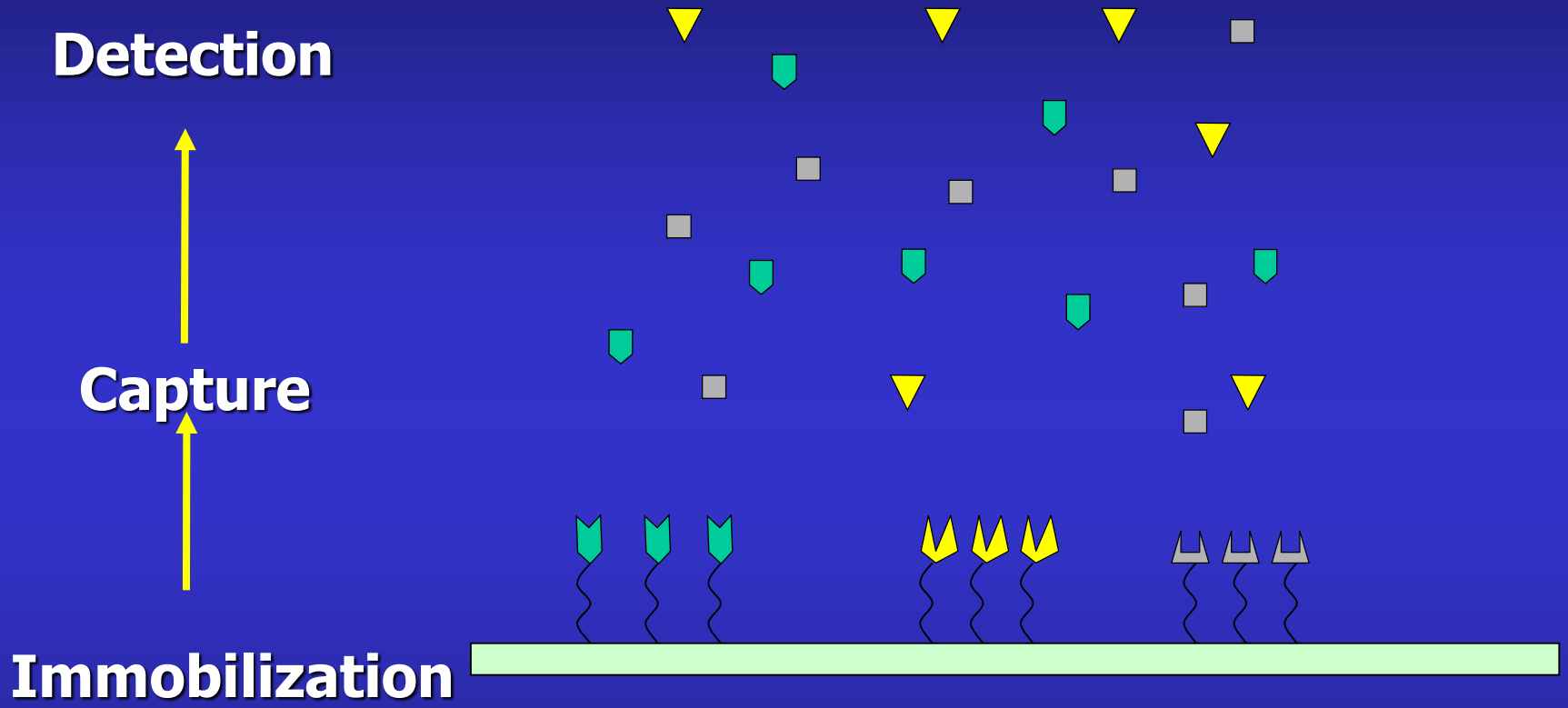
扫描仪

第三部分：蛋白质芯片

一、蛋白芯片的基本原理

蛋白质芯片是指在固相支持物上有序排列的各自独立的多肽、蛋白或相应配体，利用蛋白质与蛋白质、蛋白质与DNA、蛋白质与RNA，或与其它配体间的相互作用，同时平行分析大量蛋白质的生物化学性质，在蛋白质组学基础研究以及临床诊断、药物研究、环境监测、食品卫生等方面有广阔的应用前景。





蛋白质芯片原理示意图

二、蛋白质芯片（Protein Microarray)的分类

Interaction Arrays

—— Proteins on Chip- Labelled Sample

Caputure Arrays

—— Antibody on Chip-Labelled Sample

μELISA on Array

—— Miniaturised Sandwich Assays

Reversed Phased Arrays

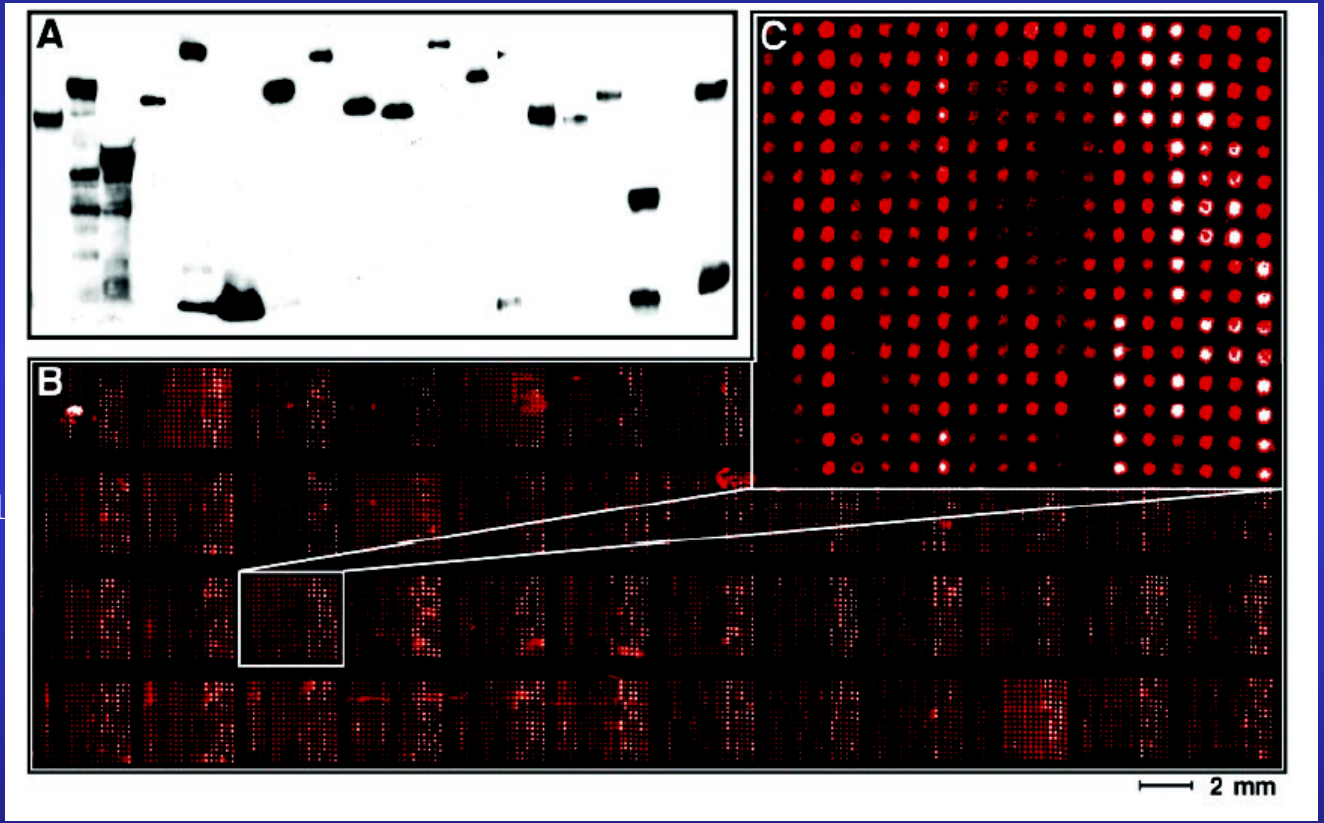
—— Multitude of arrayed Samples

Interaction Arrays

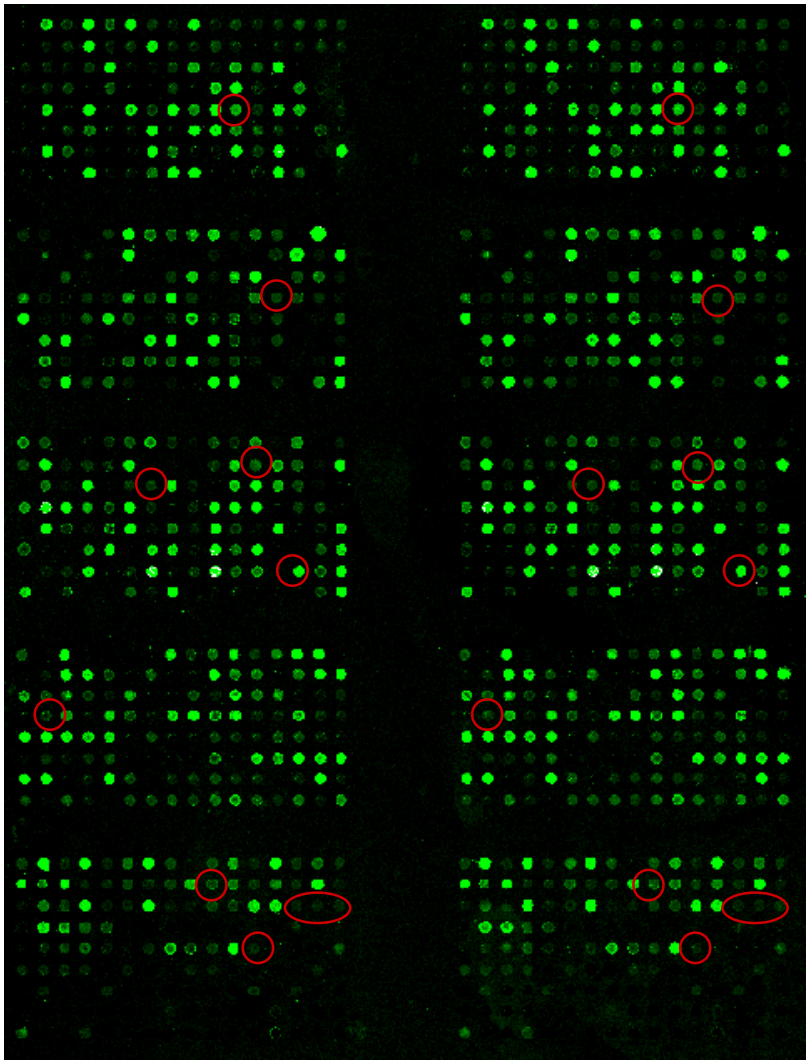
-Cloning and expression of 6600 yeast protein

-Complex arrays for protein-protein interaction studies

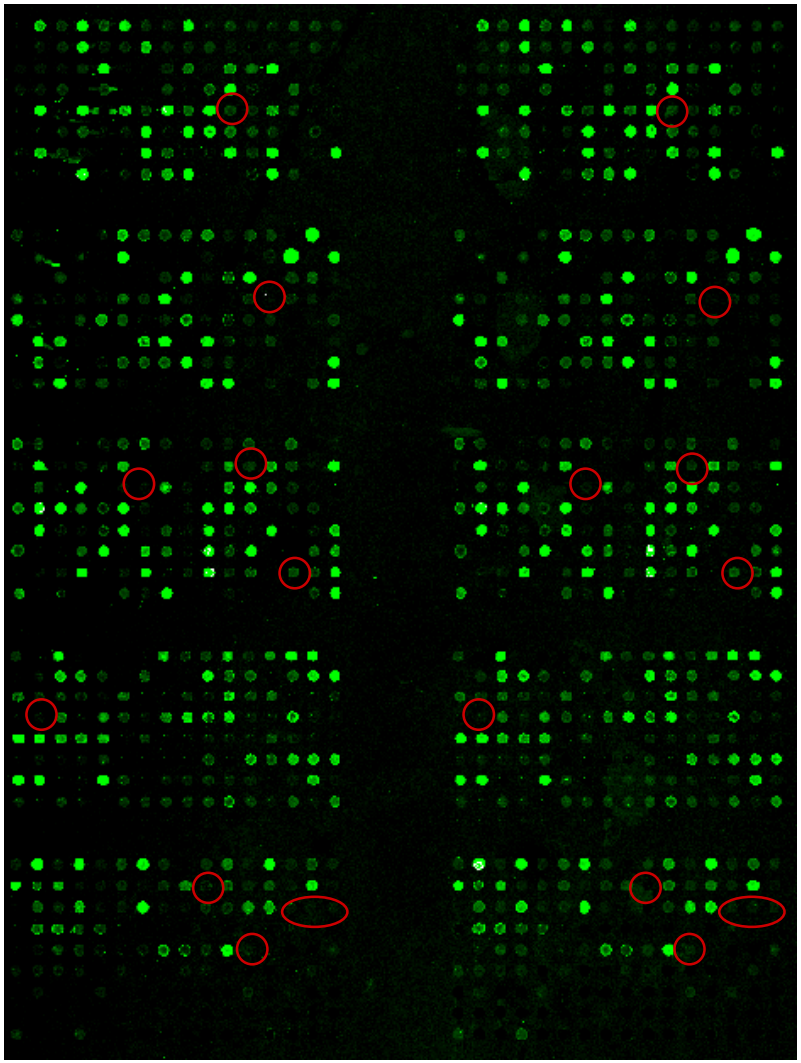
Zhu et al. (2001)Science.293,2101



Protein Array Assay Images

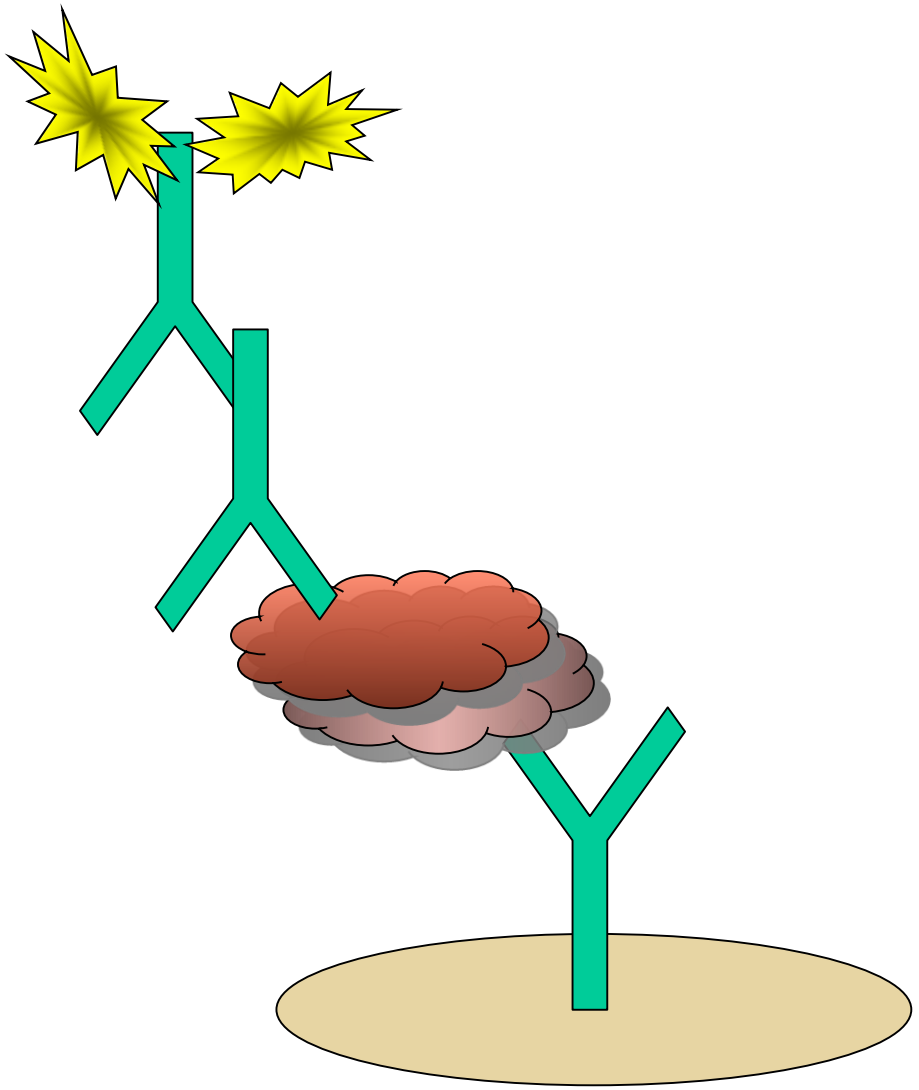


Control Sample



Treated Sample

Capture Arrays



ANTIBODY ARRAYS

Sample Labelling Approaches

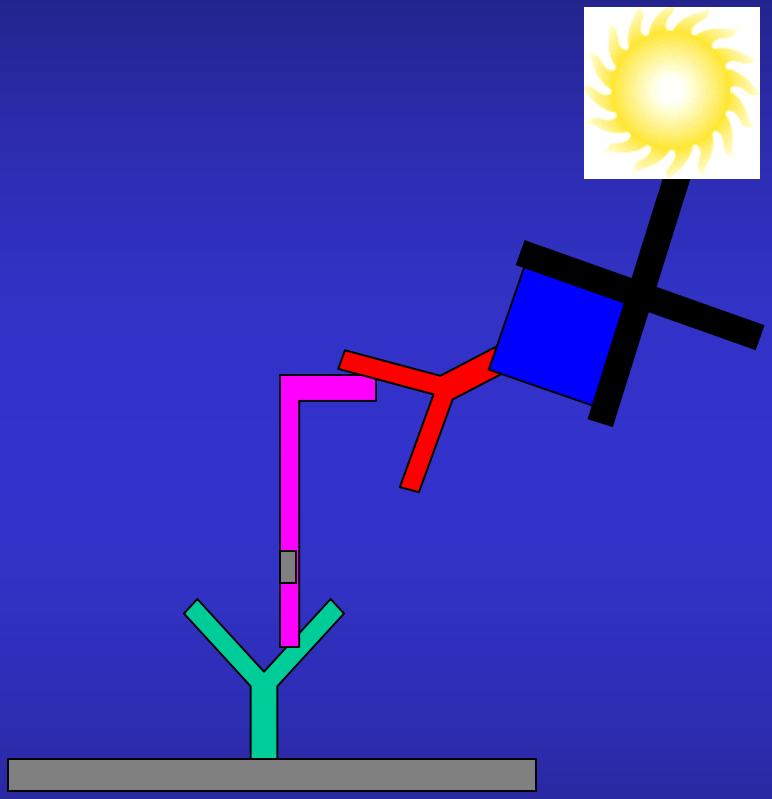
ANTIBODY ARRAYS

Mutiplexed Sandwich ELISAs

SAMPLE ARRAYS

Reverse Phase Screen

Sandwich Immunoassays



- highly parallel
- high sensitivity
- expensive equipment
- labour intense
- automation difficult

三、蛋白质芯片的制备

（一）、基质材料的选择和处理

- 聚丙烯酰胺凝胶
- 聚偏二氟乙烯膜（PVDF）
- 硝酸纤维素膜
- 载玻片
- 硅片
- 纸
- 金等
- 微珠

固定蛋白的方式包括吸附、共价结合、分子组装等

（二）、蛋白芯片的点样

制备好的蛋白通过使用DNA芯片中成熟的机械针点样系统或是微量液体分配系统实现。也可以通过原位合成的方法制备。

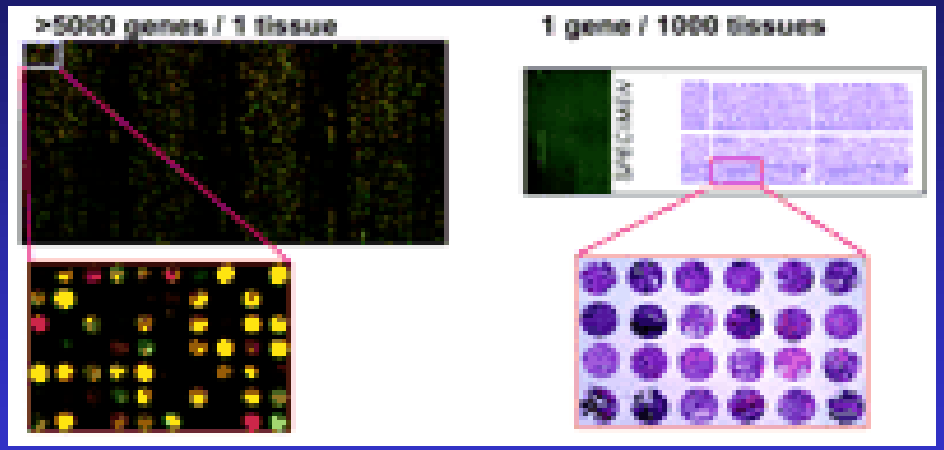
值得注意的是蛋白的粘稠度比较高，选择点样针的时候，需要针的孔径大一些。

（三）、蛋白芯片的检测

荧光，同位素标记方法，原子力显微技术，表面等离子体共振技术， 多光子检测技术，基体辅助激光解吸电离飞行时间质谱(MALDI-TOF/MS), 表面增强激光解吸附作用/离子化技术(SELDI)

第四部分，组织芯片

概



述

- ❖ 高通量
- ❖ 多组织标本
- ❖ 同时、平行分析
- ❖ 同时研究一个或多个基因或蛋白的研究工具
- ❖ 高质、高效，省材，操作便捷
- ❖ 应用广泛

→
- 成百甚至上千份
- 正常或处于疾病状态
- 疾病不同发展阶段

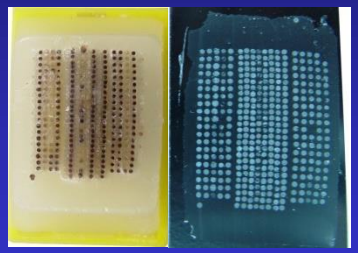
→
- **FISH**（荧光原位分子杂交）
- **mRNA ISH**（原位杂交）
- **IHC**（免疫组织化学）

→ 基因、转录、表达水平三方面

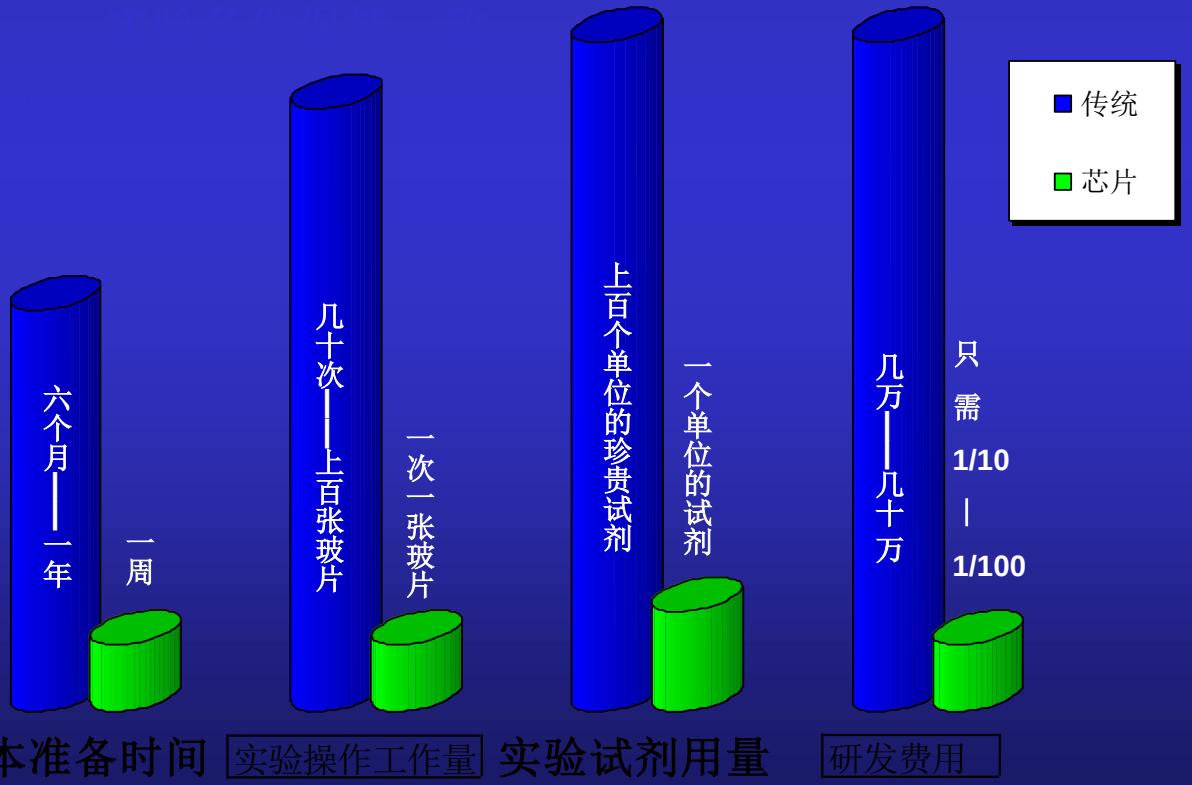
组织芯片的优点

- 高通量——短时（一次即获大量信息 ）、高效（成百成千倍地提高效率 ）

仅2CM²



含数十甚至数百例组织样本，
一次操作即可完成普通实验所
需的数十甚至数百次操作。



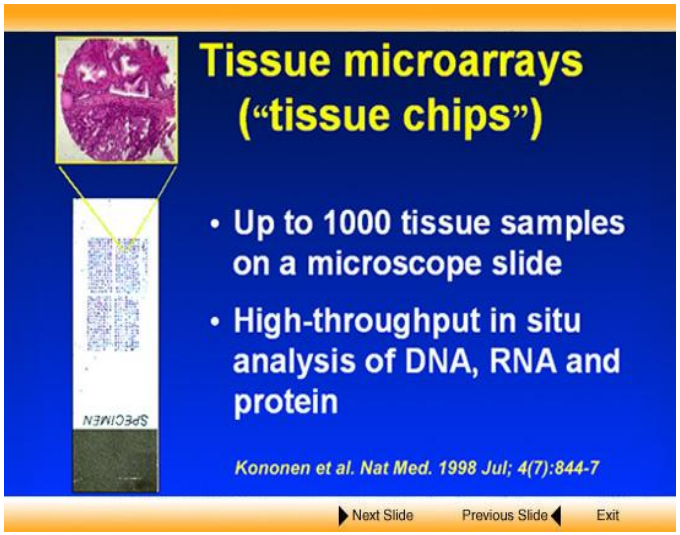
组织芯片（意义）

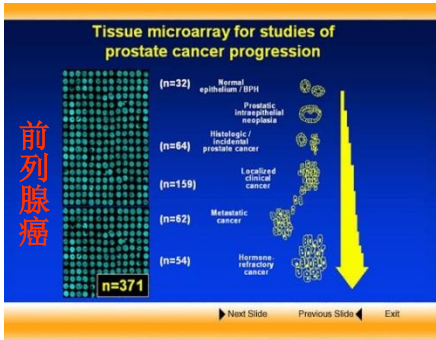
对人类基因组学、后基因组学的深入研究与发展

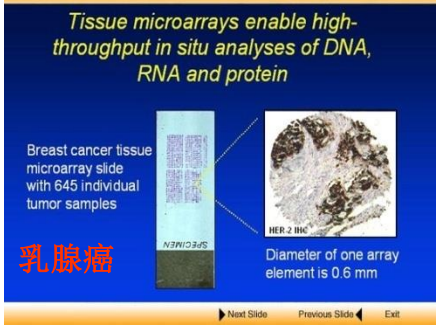
特别表现在如下几个方面：

- ❖ 研究特定基因及其所表达的蛋白质与疾病之间的相互关系
- ❖ 疾病的分子诊断
- ❖ 预后指标的确定
- ❖ 治疗靶点的定位
- ❖ 治疗过程的追踪与预测
- ❖ 抗体和新药物的开发和筛选
- ❖ 基因治疗的研发等方面

广 阔 的 应 用 前 景



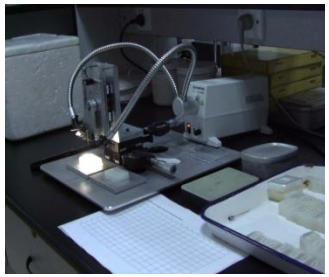


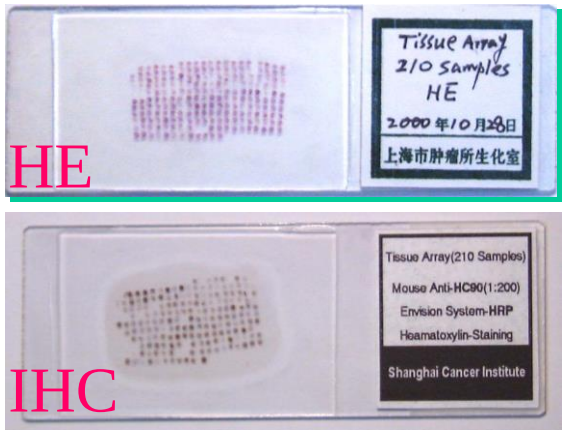


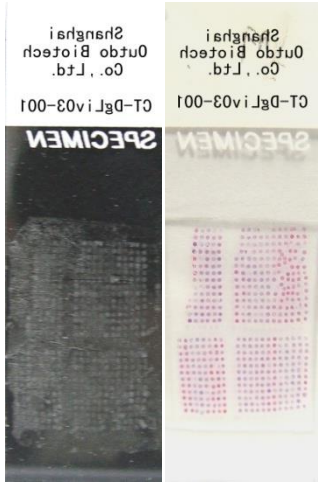
高通量研究DNA、RNA和PROTEIN

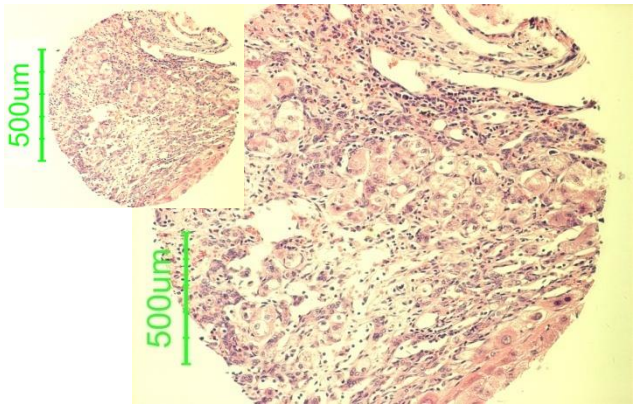
肝癌组织芯片（ OD-CT-DgLiv03-001 ）

- 高通量——478个样本
- 高质量——最新鲜的标本材料和最敏感的生物学活性
- 高价值——203例肝细胞癌 18例肝硬化 10例正常肝 14例其他组织对照









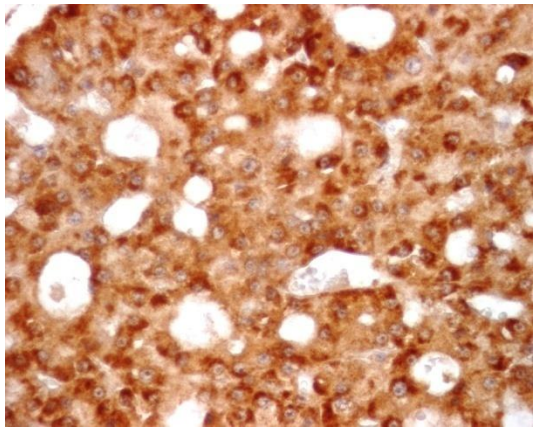
Tissue Microaray of HCC（2001/10/03）

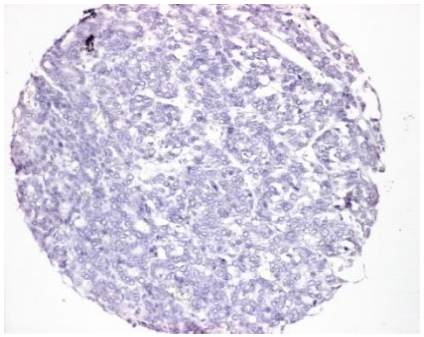
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	QD-01K	QD-01N	QD-02K	QD-02N	QD-03K	QD-03N	QD-04K	QD-04N	QD-05K	QD-05N	QD-06K	QD-06N	QD-07K	QD-07N	QD-08K	QD-08N	QD-09K	QD-09N
2	QD-10K	QD-10N	QD-11K	QD-11N	QD-12K	QD-12N	QD-13K	QD-13N	QD-14K	QD-14N	QD-15K	QD-15N	QD-16K	QD-16N	QD-17K	QD-17N	QD-18K	QD-18N
3	QD-19K	QD-19N	QD-20K	QD-20N	QD-21K	QD-21N	QD-22K	QD-22N	QD-23K	QD-23N	QD-24K	QD-24N	ZS-01K	ZS-01N	ZS-02K	ZS-02N	ZS-03K	ZS-03N
4	ZS-04N	ZS-04N	ZS-05N	ZS-05N	SY-01K	SY-01N	SY-02K	SY-02N	SY-03K	SY-03N	SY-04K	SY-04N	SY-05K	SY-05N	GX-01K	GX-01N	GX-02K	GX-02N
5	GX-03K	GX-03N	GX-04K	GX-04N	GX-05K	GX-05N	GX-06K	GX-06N	GX-07K	GX-07N	GX-08K	GX-08N	GX-09K	GX-09N	GX-10K	GX-10N	GX-11K	GX-11N
6	GX-12K	GX-12N	GX-13K	GX-13N	GX-14K	GX-14N	GX-15K	GX-15N	GX-16K	GX-16N	GX-17K	GX-17N	GX-18K1	GX-18N1	GX-18K2	GX-18N2	GX-19K	GX-19N
7	LIVER1	LIVER2	LIVER3	LIVER4	LIVER5	LIVER6	LIVER7	LIVER8	LIVER9			中山（M）	中山（H）	军 HCC	浙 1号 K	浙 1号 N	Q01-06K	Q01-06N
8	WM-L	BALB/c-L	SCID-L-T	7402-1	7402-2	7721	Hep3B-W	Hep3B-M	Hep3B-P									

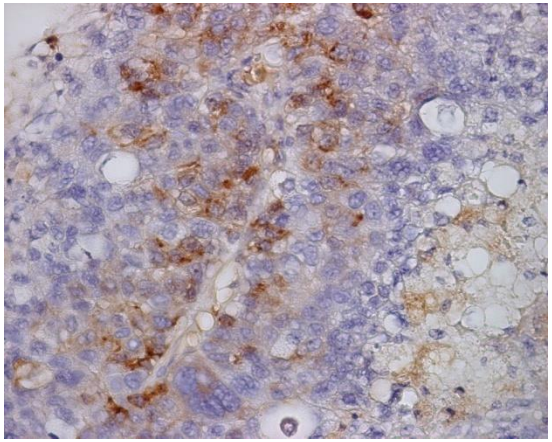
肝癌组织芯片的应用

p53、COX-2、AFP蛋白表达

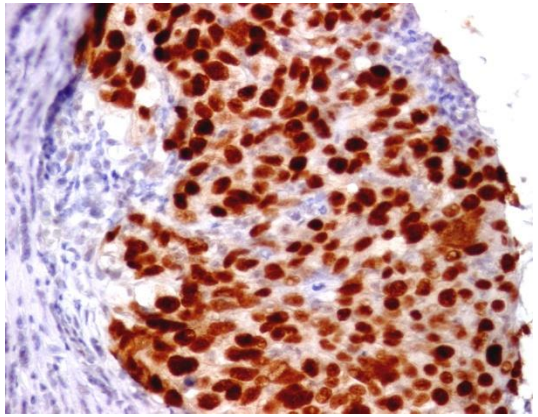
COX-2

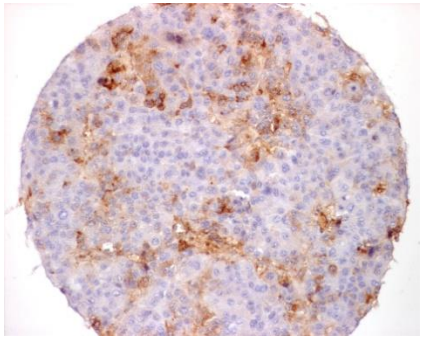




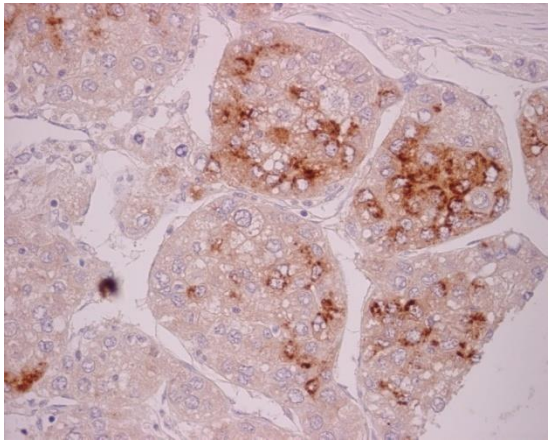


p53





AFP



人胰腺癌组织芯片

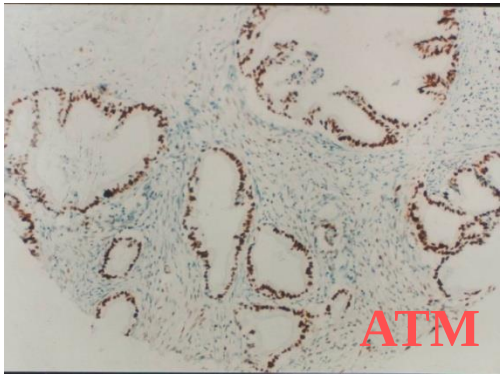
高通量——289个样本

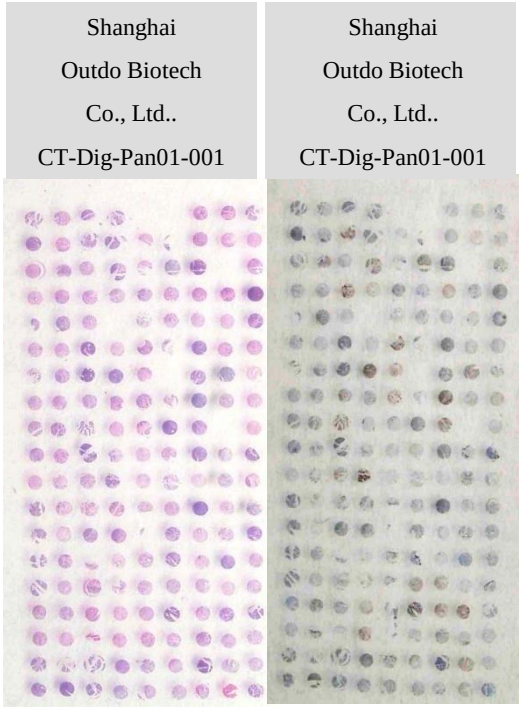
高质量——最新鲜的标本材料和最敏感的生物学活性

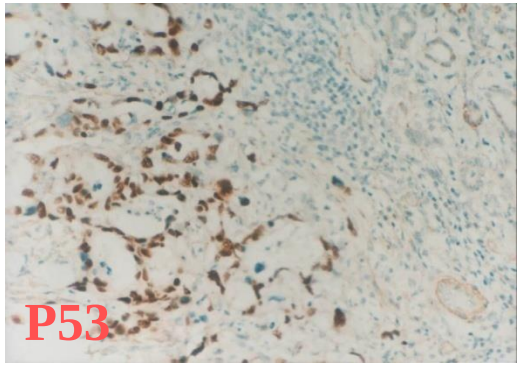
高价值——167 胰腺癌（胰腺导管腺癌133例，粘液性囊腺癌16例，腺鳞癌7例，其他11例）01 癌旁胰腺组织、7例粘液性囊腺瘤、胰腺炎组织4例与10正常胰腺

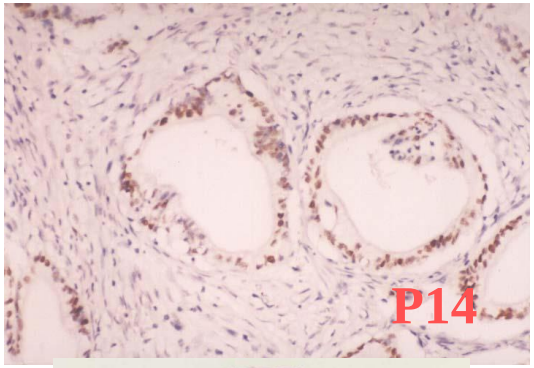
胰腺癌组织芯片的应用

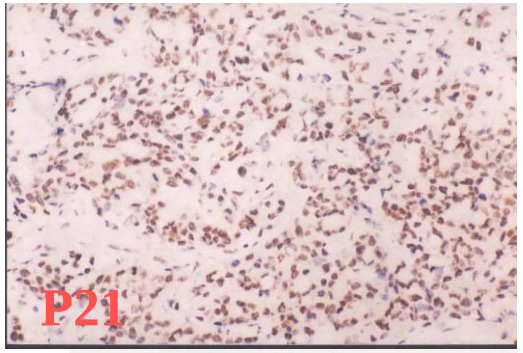
ATM、p53、 p21、 p14、 p33 AFP蛋白表达

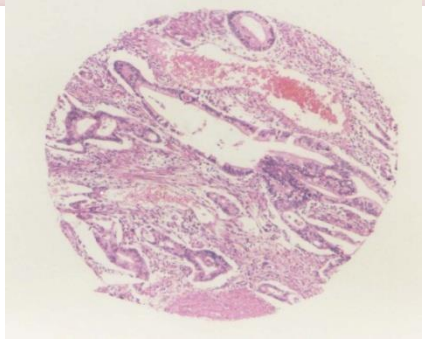


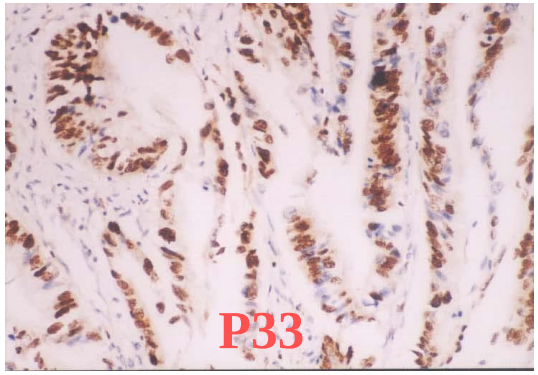


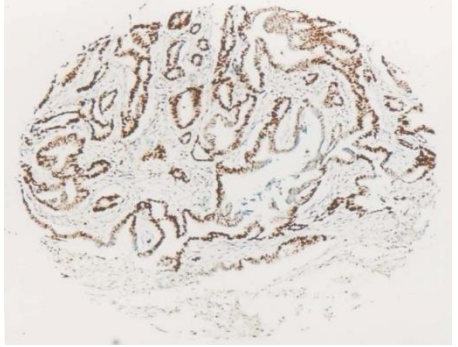






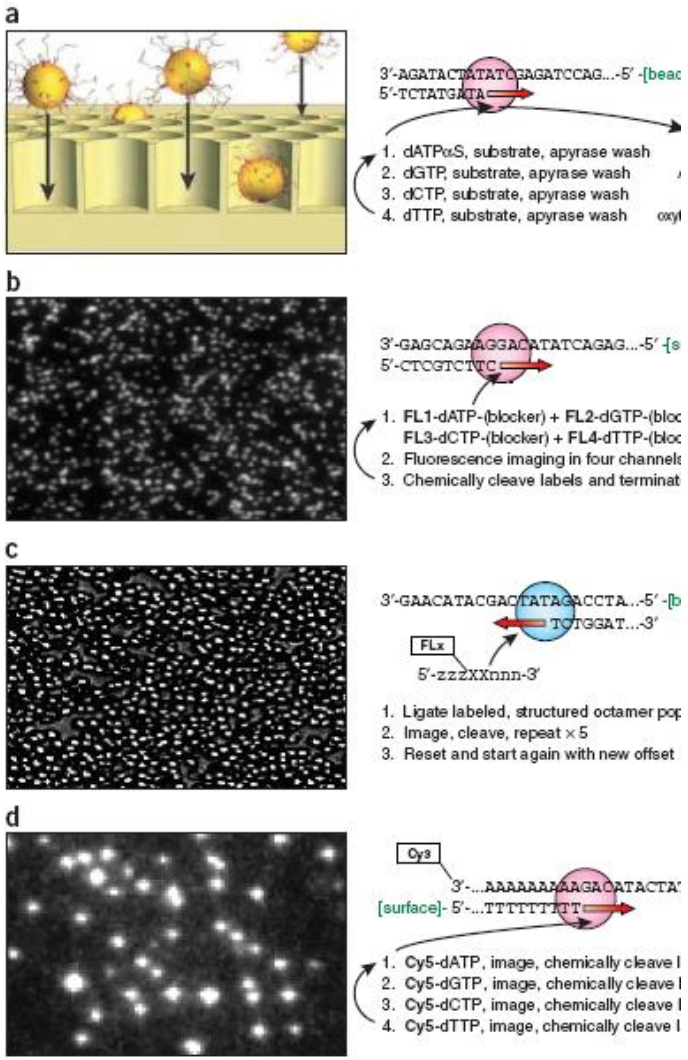






高通量测序技术正变革着当今的生物学研究

Nature biotechnology October 2008

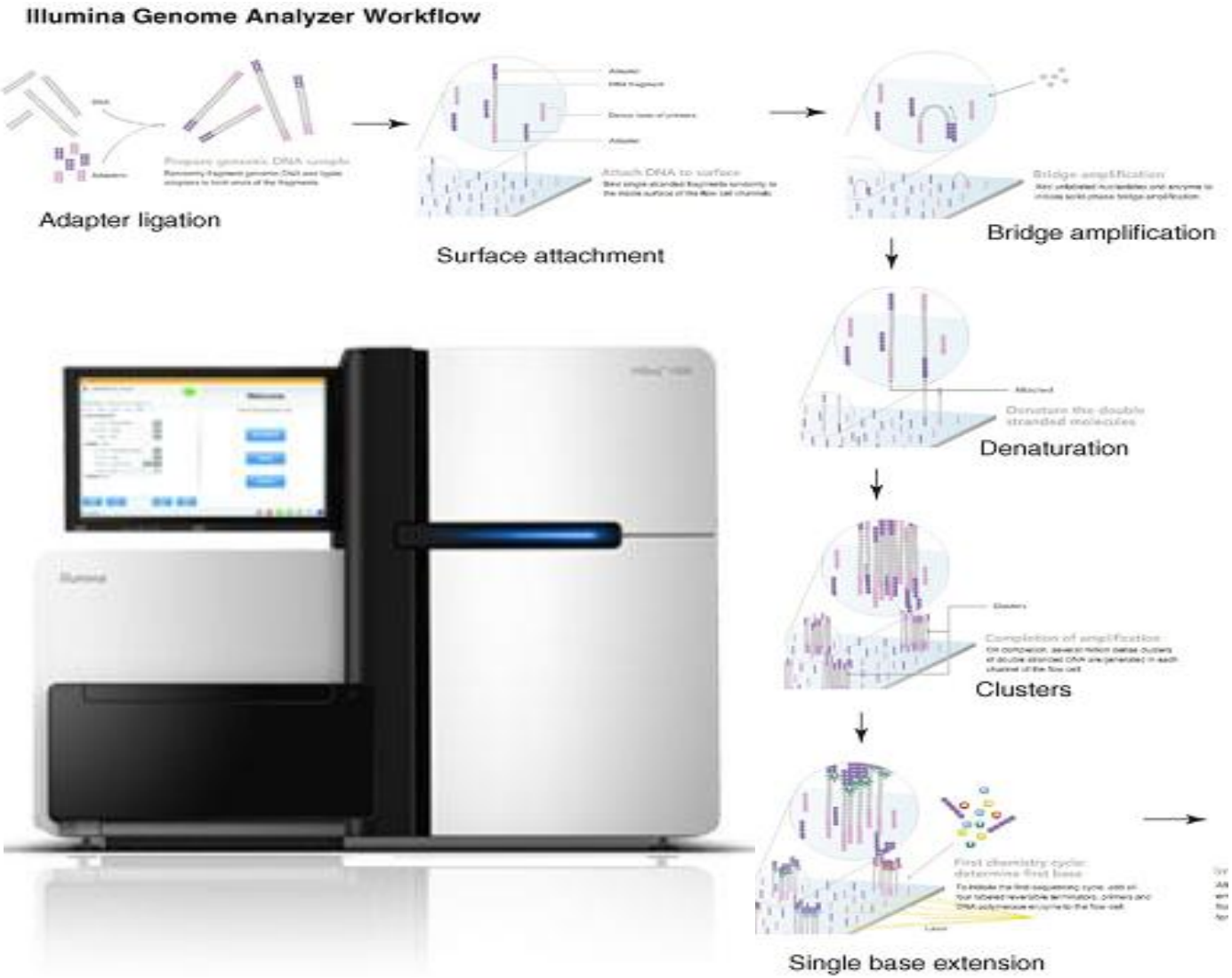


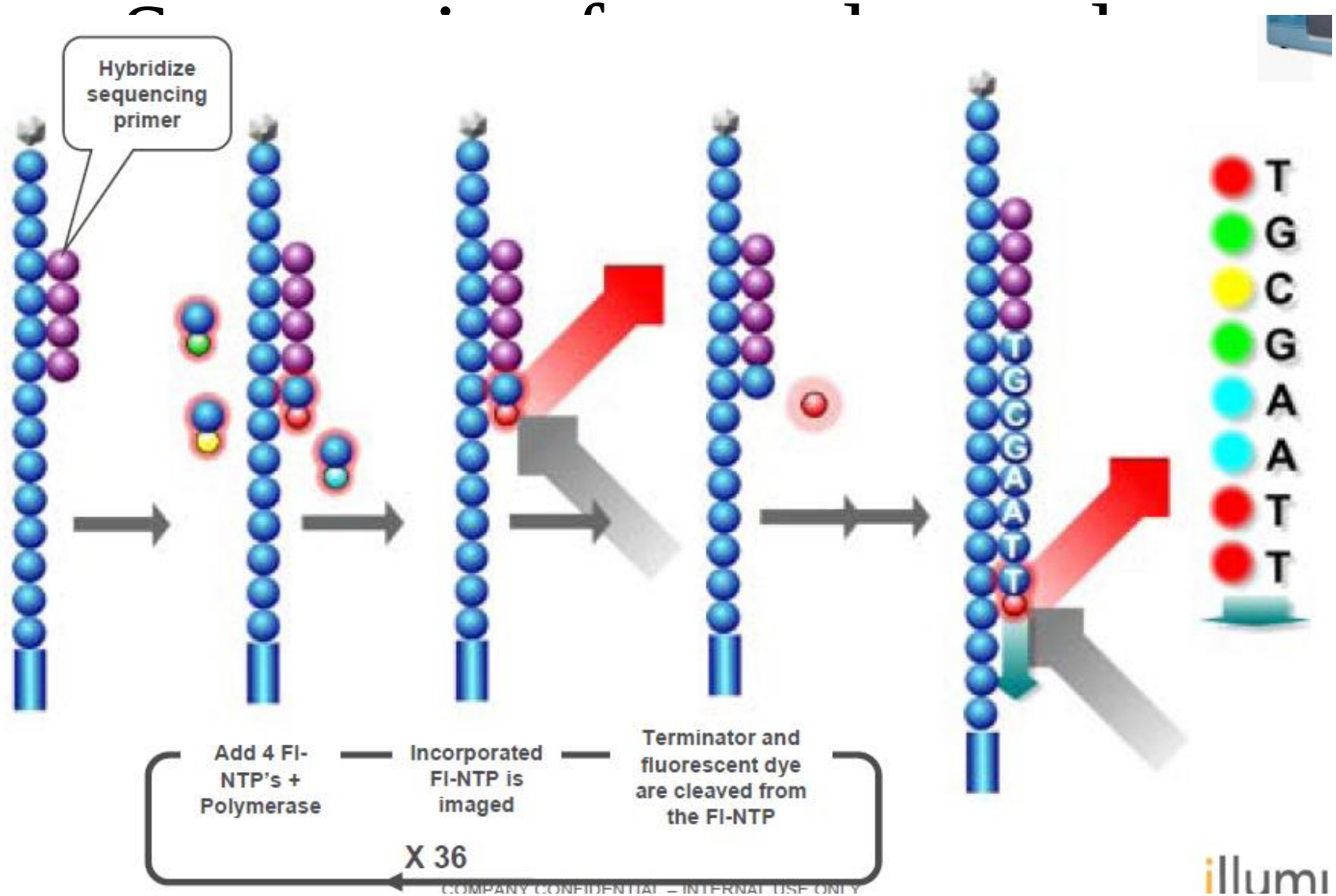
Roche 454
焦磷酸测序

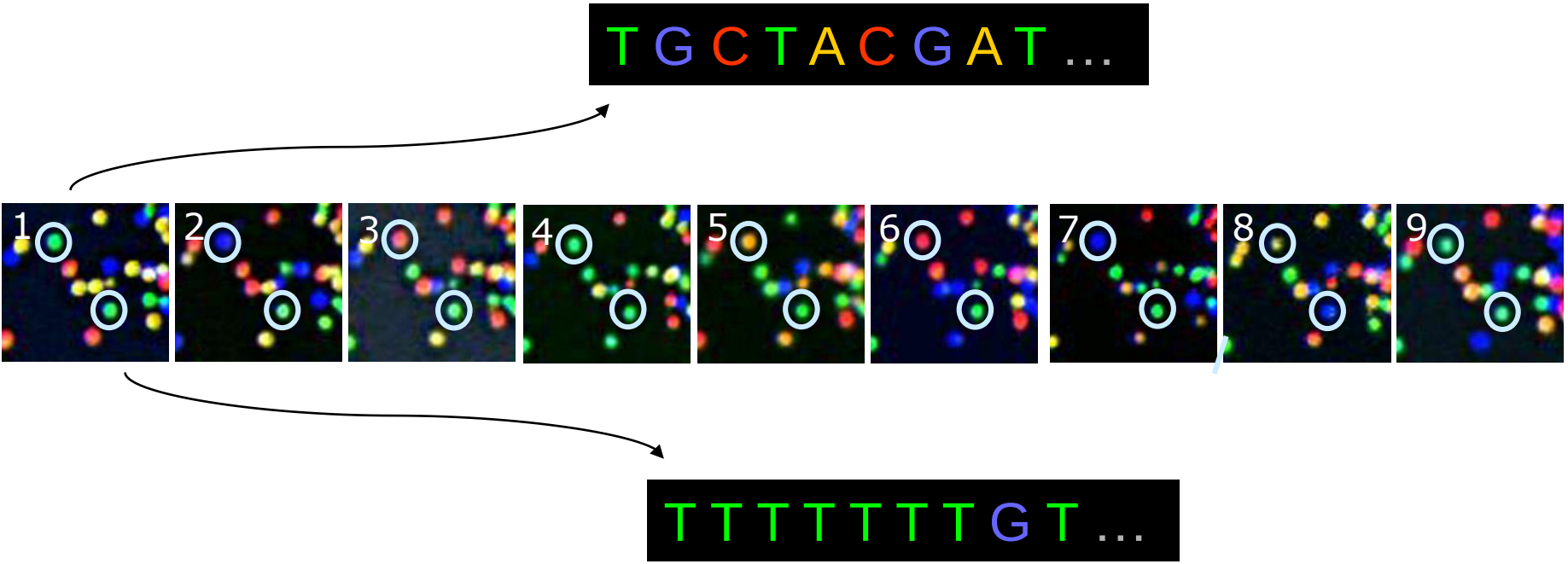
Illumina Sole
合成法测序

ABI SOLiD
连接法测序

Illumina 测序原理







The identity of each base of a cluster is read off from sequential images

根据每个点每轮反应读取的荧光信号序列，转换成相应的**DNA**序列

二代半测序（Ion Torrent， MiSeq）

	Illumina MiSeq		Ion Torrent	
Systerm	Reagent Kit v2	Reagent Kit v3	PGM	Proton
Read length (bp)	1×36 2×25 2×150 2×250	2×75 2×300	1×200 1×400	1×200
Output (Gb) / run	540-610 Mb 750-850 Mb 4.5-5.1 Gb 7.5-8.5 Gb	3.3-3.8 Gb 13.2-15 Gb	30 Mb~1 Gb 60 Mb~2 Gb	10 Gb
Run time (day)	4 h 5.5 h 24 h 39 h	24 h 65 h	2.3~4.4 h 3.7~7.3 h	2~4 h

通量更高、速度更快

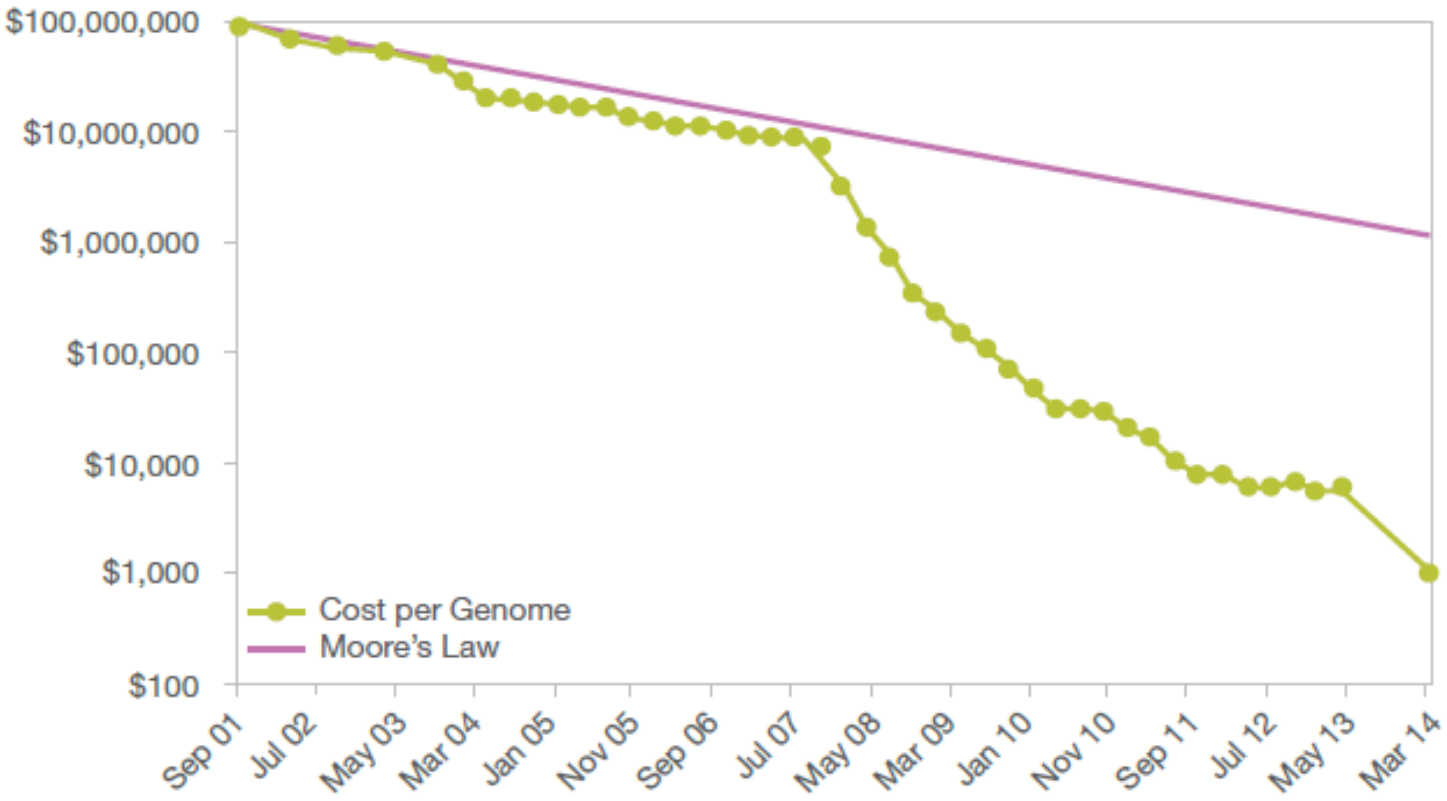
Population power. Extreme throughput. $1,000 human genome

The HiSeq X Ten is a set of ten ultra-high-throughput sequencers, purpose-built for large-scale human whole-genome sequencing.

	Dual Flow Cell	Single Flow Cell
Output/Run	1.6–1.8 Tb	800–900 Gb
Reads Passing Filter†	≤ 6 billion	≤ 3 billion
Supported Read Length	2 × 150	
Run Time	< 3 days	
Quality	≥ 75% of bases above Q30 at 2 × 150 bp	

Prep	Sequence	Analyze	Share
5.5 hours hands-on	10 minutes hands-on	5 minutes hands-on, on-site or in cloud	Secure, unlimited storage
9 HOURS	15-18 HOURS	<4 HOURS PER SAMPLE	INSTANTLY

测序成本：1000美元/每人类基因组



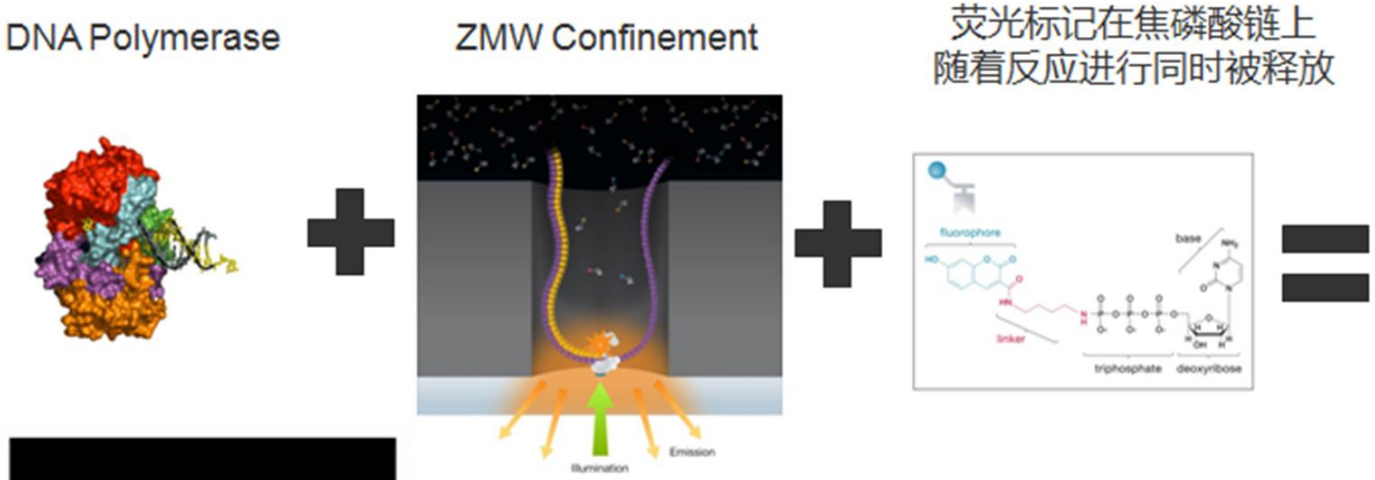
NovaSeq

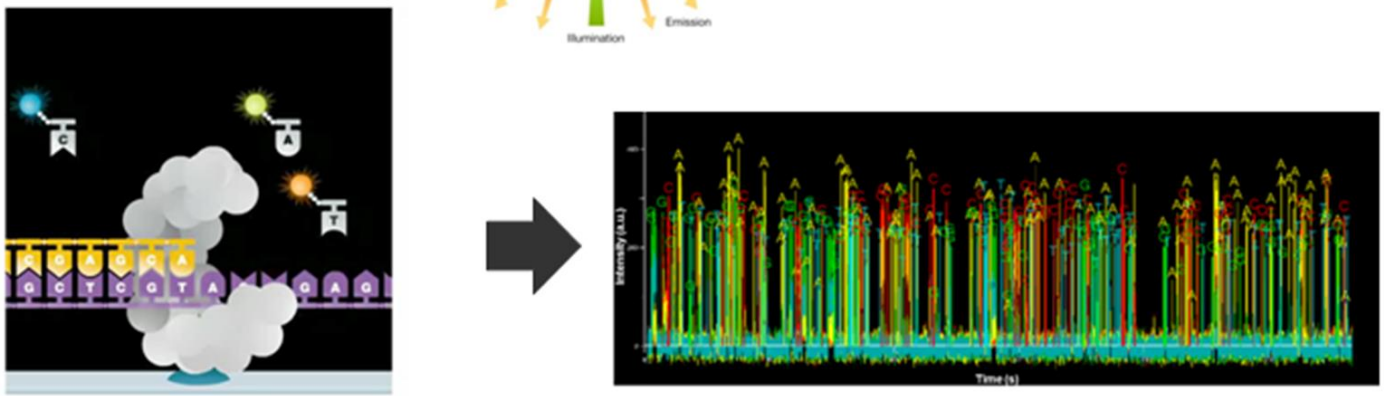

The next era in sequencing starts now
The NovaSeq Series of Sequencing Systems unleashes groundbreaking innovations that leverage our proven technology. Now you can get scalable throughput and flexibility for virtually any sequencing method, genome, and scale of project.
Contact an Illumina Representative

第三代测序：单分子测序

1、Pacific Biosciences

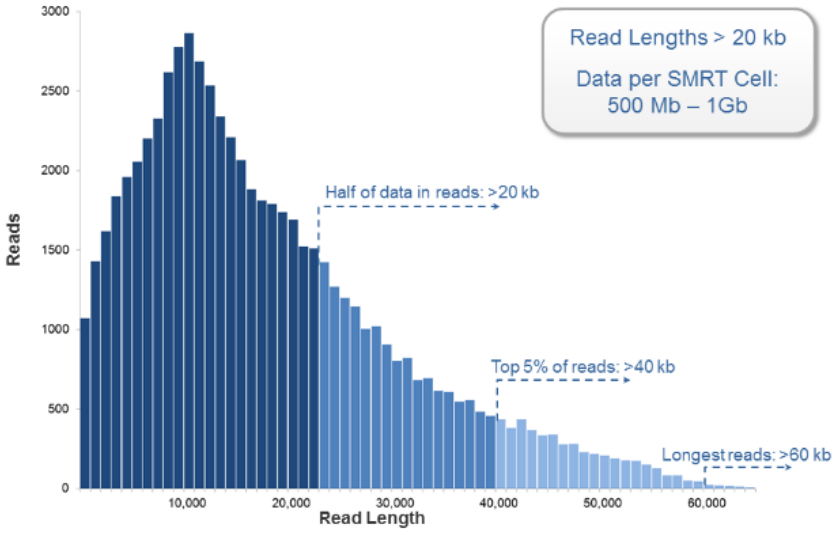
- 技术： Single Molecule Real Time (SMI
- 平台： **PACBIO RS II**

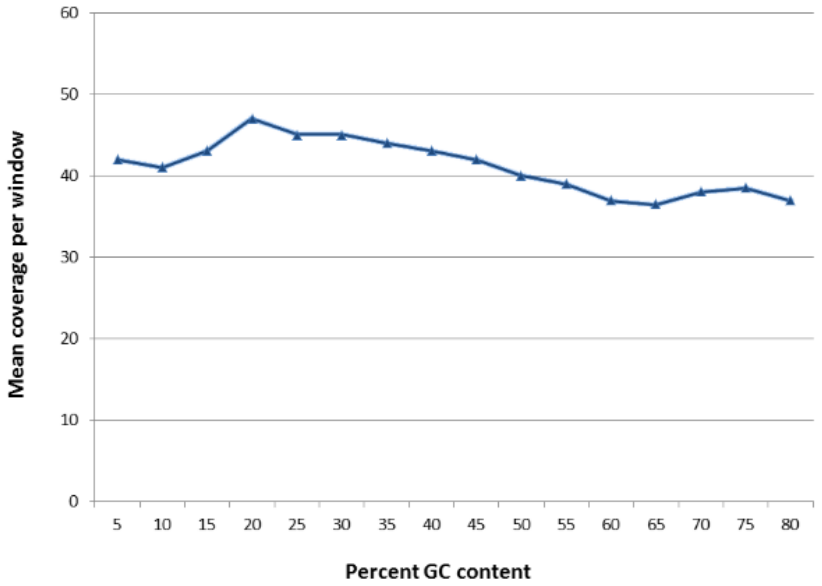




PacBio的技术优势

1. **Longest Read lengths**
   - Sequence reads >20kb
   - Some reads >60 kb
2. **Highest Accuracy**
   - Achieves >99.999% (QV50)
   - Lack of systematic sequencing errors
3. **Greatest Uniformity**
   - Lack of GC content or sequence complexity bias
4. **Can sequence native DNA**
   - No DNA amplification
   - Epigenome characterization





2、Oxford Nanopore Technologies

Oxford Nanopore Technologies

- 技术： nanopore-based electronic systems for analysis of single molecules

MinION

- 平

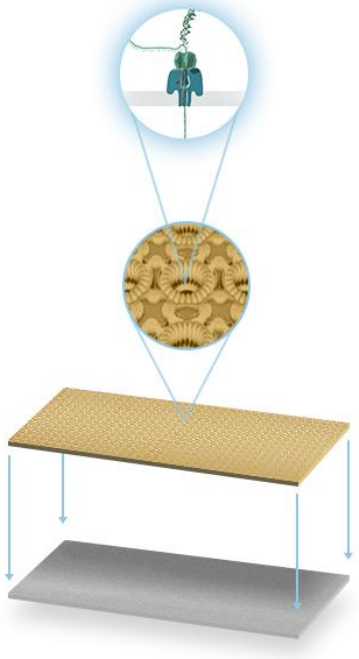
PromethION

GridION

MinION的基本结构和参数

MinION Mkl Flow cell

The MinION Mkl flow cell contains a sensor array of several hundred channels, to enable multiple nanopore experiments to be performed in parallel. The flow cell is compatible with the MinION Mkl device.



样品量（PCR Free）： 10pg - 1μg

通量：

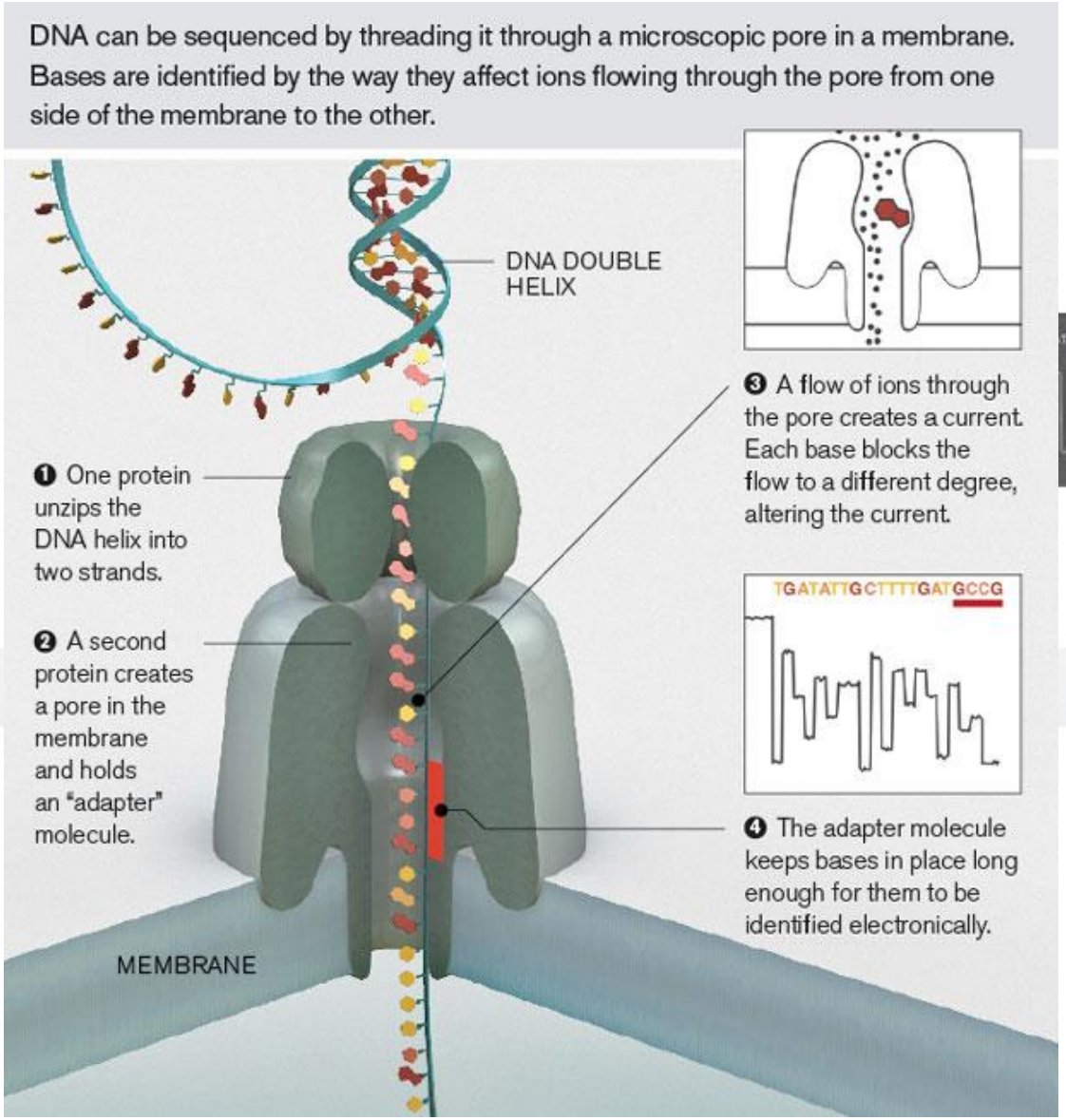
6G（10kb，标准模式）

42G （10kb，快速模式）

最长读长：230-300kb

运行时间：48 h

Nanopore的测序原理



测序市场份额

图 13：2013 年全球新一代测序仪市场格局

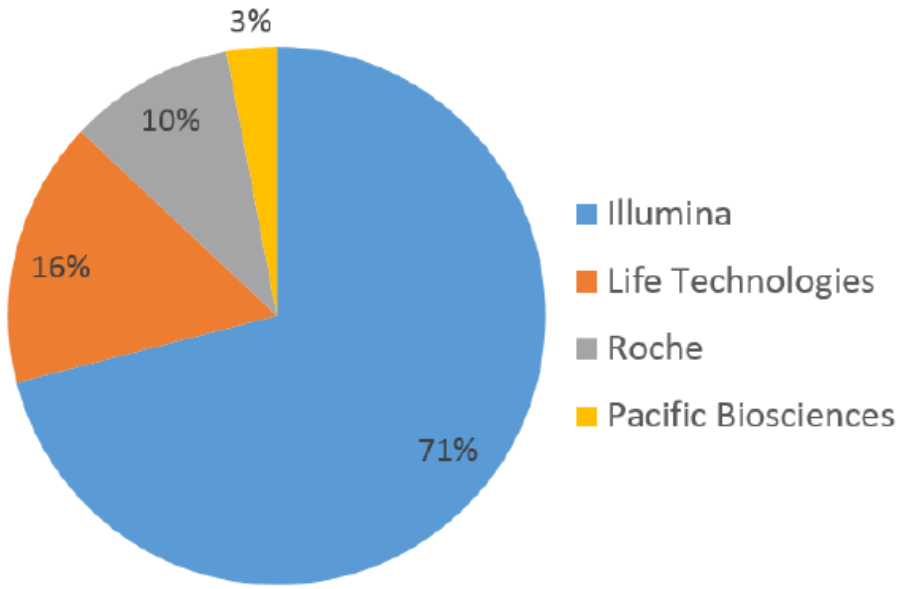
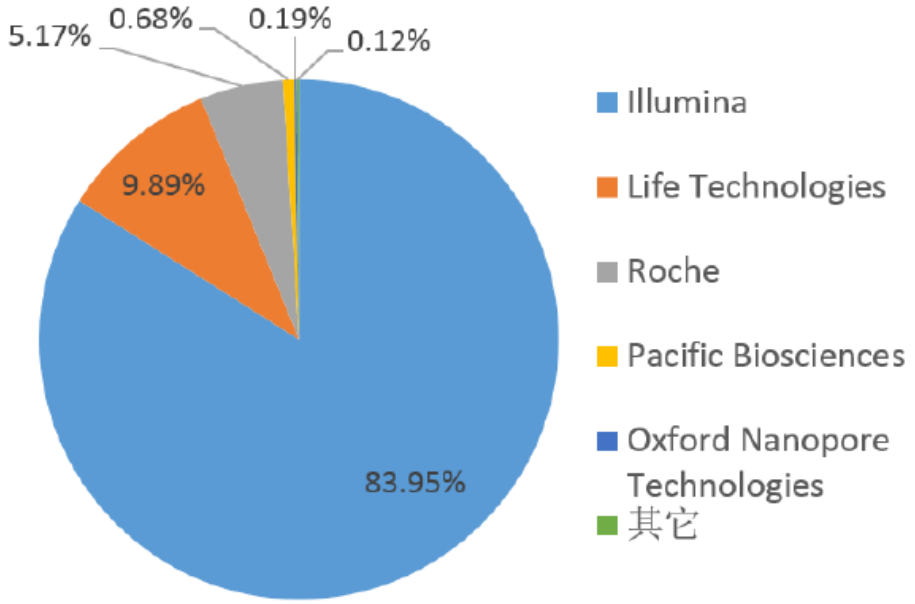
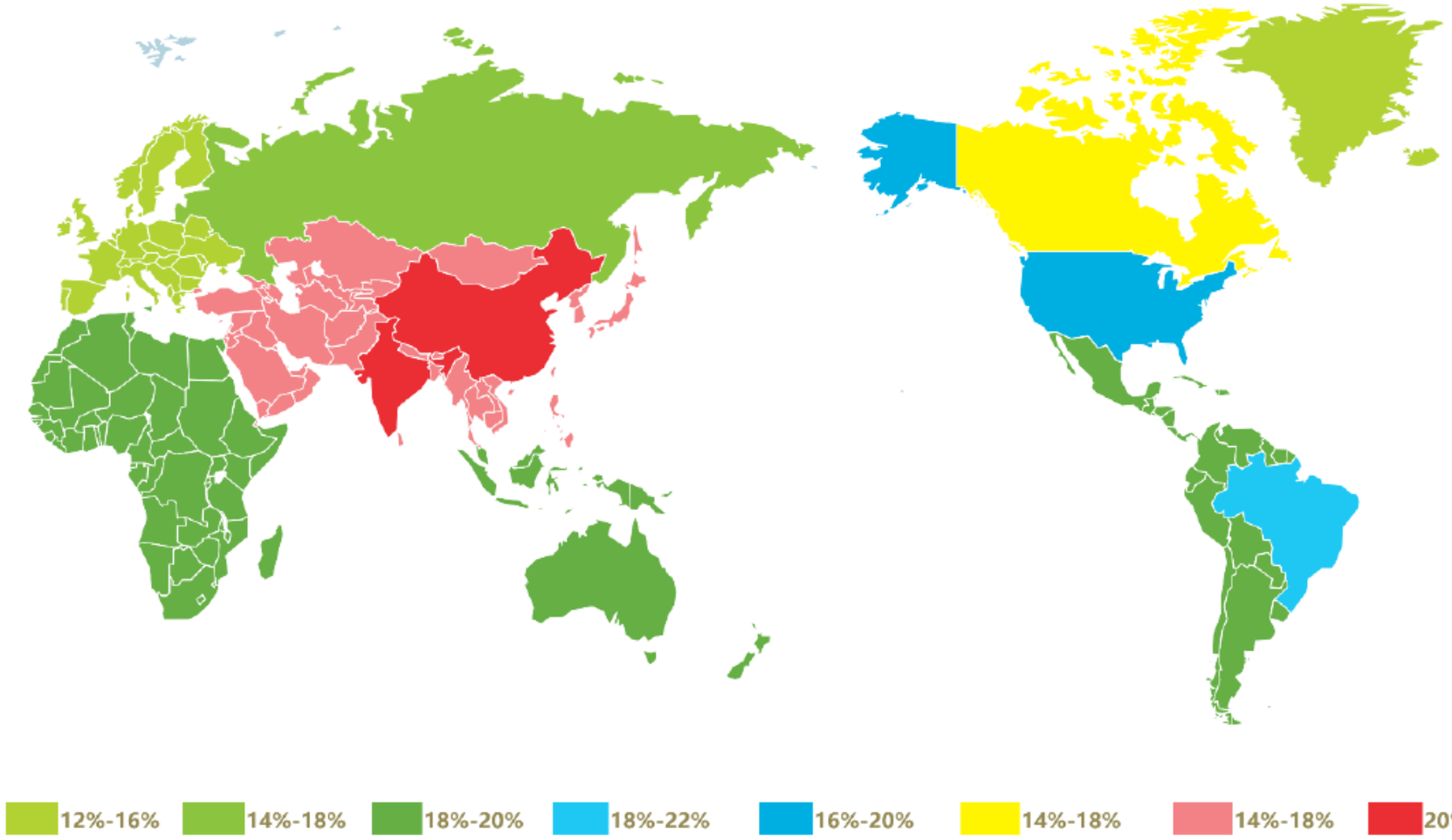


图 14：2016 年全球新一代测序仪市场格局

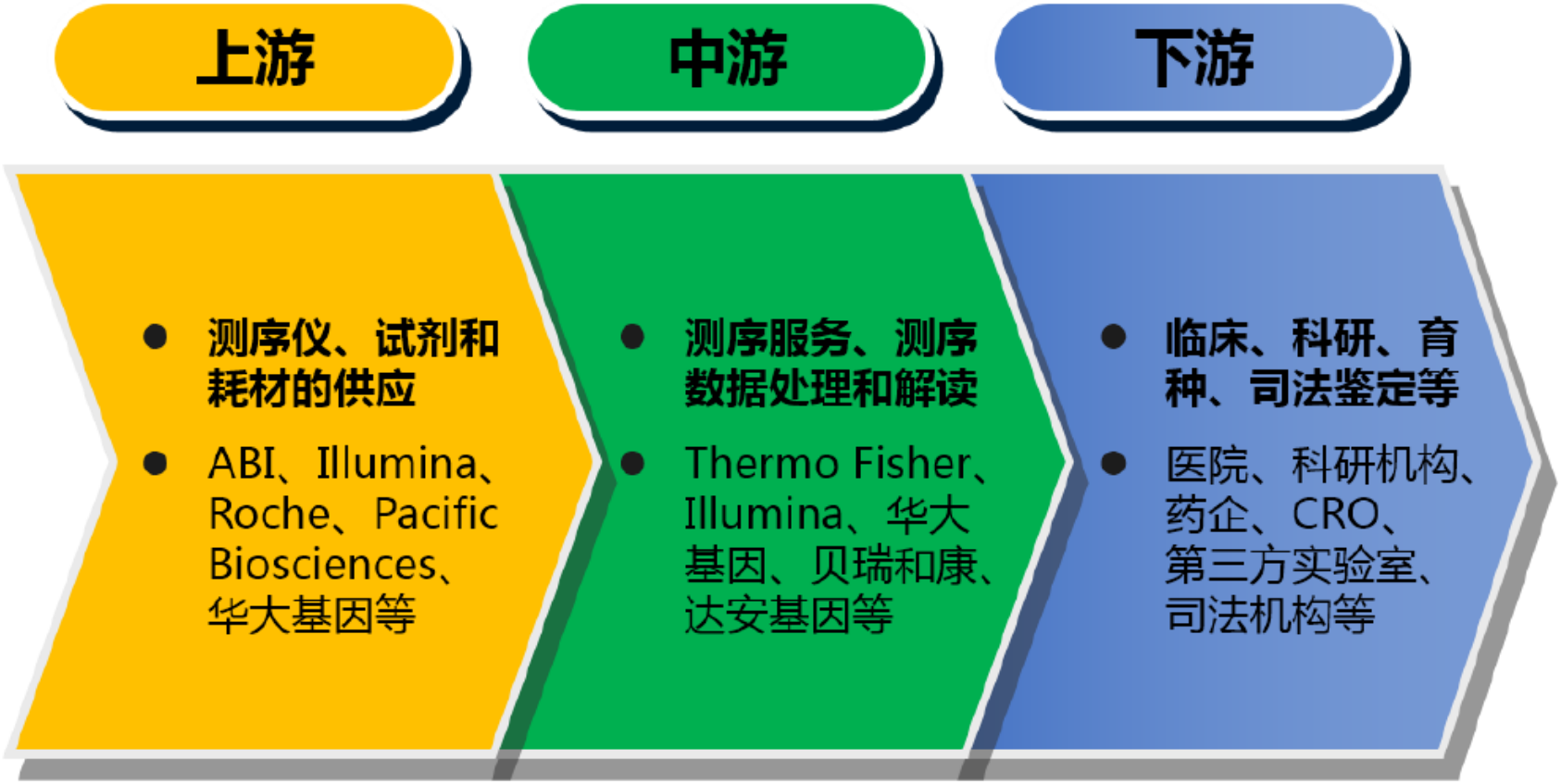


全球基因测序市场增速

图 11：2012-2017 年全球基因测序市场增速分布



基因测序行业产业链


上游
中游
下游
测序仪、试剂和耗材的供应
ABI、Illumina、Roche、Pacific Biosciences、华大基因等
测序服务、测序数据处理和解读
Thermo Fisher、Illumina、华大基因、贝瑞和康、达安基因等
临床、科研、育种、司法鉴定等
医院、科研机构、药企、CRO、第三方实验室、司法机构等

基因组/转录组 re-sequencing

Vol 461|8 October 2009|doi:10.1038/nature08489 nature

LETTERS

Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution

Sohrab P. Shah[1,2]*, Ryan D. Morin[3]*, Jaswinder Khattra[1], Leah Prentice[1], Trevor Pugh[3], Angela Burleigh[1], Allen Delaney[3], Karen Gelmon[4], Ryan Guliany[1], Janine Senz[2], Christian Steidl[2,5], Robert A. Holt[3], Steven Jones[3], Mark Sun[1], Gillian Leung[1], Richard Moore[3], Tesa Severson[3], Greg A. Taylor[3], Andrew E. Teschendorff[6], Kane Tse[1], Gulisa Turashvili[1], Richard Varhol[3], René L. Warren[3], Peter Watson[7], Yongjun Zhao[3], Carlos Caldas[6], David Huntsman[2,5], Martin Hirst[3], Marco A. Marra[3] & Samuel Aparicio[1,2,5]

Recent advances in next generation sequencing[1–4] have made it possible to precisely characterize all somatic coding mutations that occur during the development and progression of individual cancers. Here we used these approaches to sequence the genomes (>43-fold coverage) and transcriptomes of an oestrogen-receptor-α-positive metastatic lobular breast cancer at depth. We found 32 somatic non-synonymous coding mutations present in the metastasis, and measured the frequency of these somatic mutations in DNA from the primary tumour of the same patient, which arose 9 years earlier. Five of the 32 mutations (in *ABCB11*, *HAUS3*, *SLC24A4*, *SNX4* and *PALB2*) were prevalent in the DNA of the primary tumour removed at diagnosis 9 years earlier, six (in *KIF1C*, *USP28*, *MYH8*, *MORC1*, *KIAA1468* and *RNASEH2A*) were present at lower frequencies (1–13%), 19 were not detected in the primary tumour, and two were undetermined. The combined analysis of genome and transcriptome data revealed two new RNA-editing events that recode the amino acid sequence of *SRP9* and *COG3*. Taken together, our data show that single nucleotide mutational heterogeneity can be a property of low or intermediate grade primary breast cancers and that significant evolution can occur with disease progression.

Lobular breast cancer is an oestrogen-receptor-positive (ER^+, also known as $ESR1^+$) subtype of breast cancer (approximately 15% of all breast cancers). It is usually of low-intermediate histological grade and can recur many years after initial diagnosis. To interrogate the genomic landscape of this class of tumour, we re-sequenced[1–4] the DNA from a metastatic lobular breast cancer specimen (89% tumour cellularity; Supplementary Fig. 1) at approximately 43.1-fold aligned, haploid reference genome coverage (120.7 gigabases (Gb) aligned paired-end sequence; Supplementary Fig. 2, Table 1 and Supplementary Methods). Deep high-throughput transcriptome sequencing (RNA-seq)[5] performed on the same sample generated 160.9-million reads that could be aligned (Supplementary Table 1, see also Supplementary Fig. 2 and Supplementary Methods). The saturation of the genome (Table 1) and RNA-seq (Supplementary Table 1) libraries for single nucleotide variant (SNV) detection is discussed in Supplementary Information. The aligned (hg18) reads were used to identify (Supplementary Fig. 2) the presence of genomic aberrations, including SNVs (Supplementary Table 2), insertions/deletions (indels), gene fusions, translocations, inversions and copy number alterations (Supplementary Methods). We examined predicted coding indels and predicted inversions (coding or non-coding; Supplementary Methods); however, all of the events that were validated by Sanger re-sequencing were also present in the germ line (Supplementary Tables 3 and 4). None of the 12 predicted gene fusions revalidated. We also computed the segmental copy number (Supplementary Methods and Supplementary Table 5a) from aligned reads, and revalidated high level amplicons by fluorescence *in situ* hybridization (FISH) (Supplementary Table 5b), revealing the presence of a new low-level amplicon in the *INSR* locus (Supplementary Fig. 3).

We identified coding SNVs from aligned reads, using a Binomial mixture model, SNVMix (Supplementary Table 2, Methods and Supplementary Appendix 1). From the RNA-seq (WTSS-PE) and genome (WGSS-PE) libraries we predicted 1,456 new coding non-synonymous SNVMix variants (Supplementary Table 2). After the removal of pseudogene and HLA sequences (1,178 positions remaining) and after primer design, we re-sequenced (Sanger amplicons) 1,120 non-synonymous coding SNV positions in the tumour DNA and normal lymphocyte DNA. Some 437 positions (268 unique to WGSS-PE, 15 unique to WTSS-PE, and 154 in common) were confirmed as non-synonymous coding variants. Of these, 405 were new

Table 1 | Summary of sequence library coverage

	WGSS-PE	WTSS-PE
Total number of reads	2,922,713,774	182,532,650
Total nucleotides (Gb)	140.991	7.108
Number of aligned reads	2,502,465,226	160,919,484
Aligned nucleotides (Gb)	120.718	6.266
Estimated error rate	0.021	0.013
Estimated depth (non-gap regions)	43.114	NA
Canonically aligned reads	2,294,067,534	109,093,616
Exons covered	93.5 at >10 reads; 95.7 at >5 reads	82,200 at 10 reads (see also Supplementary Table 1)
Reads aligned canonically (%)	78.49	67.79
Unaligned reads	420,248,548	21,613,166
Mean read length (bp)	48.24	38.94

The WGSS-PE column shows the genome paired-end read coverage for DNA from the metastatic pleural effusion sample. The WTSS-PE column shows coverage for the complementary DNA reads from the matched transcriptome libraries of the metastatic pleural effusion. Coverage of exon bases in the reference genome (hg18) is shown at 5 or more reads per position, and 10 or more reads per position for the metastatic genome. bp, base pairs; NA, not applicable.

[1]Molecular Oncology, [2]Centre for Translational and Applied Genomics, [3]Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver V5Z 1L3, Canada. [4]Medical Oncology, BC Cancer Agency, 600 West 10th Avenue, Vancouver V5Z 1L3, Canada. [5]Department of Pathology, University of British Columbia, G227-2211 Wesbrook Mall, British Columbia, Vancouver V6T 2B5, Canada. [6]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [7]Deeley Research Centre, BC Cancer Agency, Victoria V8R 6V5, Canada.
*These authors contributed equally to this work.

Nature October 2009

利用二代测序技术对同一个样品进行基因组和转录组的重测序工作， 在人类乳腺小叶肿瘤中发现未报道过的碱基突变。

Table 1 | Summary of sequence library coverage

	WGSS-PE	WTSS-PE
Total number of reads	2,922,713,774	182,532,650
Total nucleotides (Gb)	140.991	7.108
Number of aligned reads	2,502,465,226	160,919,484
Aligned nucleotides (Gb)	120.718	6.266
Estimated error rate	0.021	0.013
Estimated depth (non-gap regions)	43.114	NA
Canonically aligned reads	2,294,067,534	109,093,616
Exons covered	93.5 at >10 reads; 95.7 at >5 reads	82,200 at 10 reads (see also Supplementary Table 1)
Reads aligned canonically (%)	78.49	67.79
Unaligned reads	420,248,548	21,613,166
Mean read length (bp)	48.24	38.94

The WGSS-PE column shows the genome paired-end read coverage for DNA from the metastatic pleural effusion sample. The WTSS-PE column shows coverage for the complementary DNA reads from the matched transcriptome libraries of the metastatic pleural effusion. Coverage of exon bases in the reference genome (hg18) is shown at 5 or more reads per position, and 10 or more reads per position for the metastatic genome. bp, base pairs; NA, not

WGSS，基因组重测序
WTSS，转录组重测序

基因组/转录组 re-sequencing

找到基因组和转录组同时存在新的位点突变
发现新拷贝数变异位点， **CNV**

Table 2 | Somatic coding sequence SNVs validated by Sanger sequencing

Gene	Description	Position	Source	Allele change	Amino acid change	Protein domain affected	Expression (sequenced bases per exonic base)	Allelic expression bias (R, NR allele)	Copy number classification (HMM state)
ABCB11	Bile salt export pump (ATP-binding cassette sub-family B member 11)	2:169497197	WGSS	C>T	R>H	Transmembrane helix 3	0.3	1, 1	Amplification (4)
HAUS3	HAUS3 coiled-coil protein (C4orf15)	4:2203607	WGSS, WTSS	C>T	V>M	Unknown	14.1	4, 23	Neutral (2)
CDC6	Cell division control protein 6 homologue	17:35701114	WGSS, WTSS	G>A	E>K	N-terminal, unknown	2.7	3, 3	Amplification (4)
CHD3	Chromodomain-helicase-DNA-binding protein 3	17:7751231	WGSS	G>A	E>K	Unknown, C-terminal	3.9	41, 11 ($Q < 0.01$)	Neutral (2)
DLG4	Disks large homologue 4	17:7052251	WGSS	G>A	P>L	Unknown, N-terminal	5.5	7, 1	Neutral (2)
ERBB2	Receptor tyrosine-protein kinase erb-b2	17:35133783	WGSS, WTSS	C>G	I>M	Kinase domain	67.1	62, 35	Amplification (4)
FGA	Fibrinogen alpha chain	4:155726802	WGSS	C>T	W>stop	Fibrinogen a/b/c domain	0.01	NA	Gain (3)
*GOLGA4**	Golgin subfamily A member 4	3:37267947	WGSS, WTSS	G>C	E>Q	Unknown, N-terminal	111.8	37, 12	Gain (3)
GSTCD	Glutathione *S*-transferase C-terminal domain-containing protein	4:106982671	WGSS, WTSS	G>C	E>Q	Unknown, C-terminal	23.2	23, 8	Neutral (2)
*KIAA1468**	LisH domain and HEAT repeat-containing protein	18:58076768	WGSS, WTSS	G>C	R>T	ARM type fold	36.1	23, 11	Neutral (2)
KIF1C	Kinesin-like protein KIF1C	17:4848025	WGSS, WTSS	G>C	K>N	Kinesin motor domain	28.5	16, 13	Neutral (2)
KLHL4	Kelch-like protein 4	X:86659878	WGSS	C>T	S>L	Unknown, N-terminal	1.7	1, 0	Neutral (2)
MYH8	Myosin 8 (myosin heavy chain 8)	17:10248420	WGSS	C>G	M>I	Actin-interacting protein domain	0	NA	Neutral (2)
PALB2	Partner and localizer	16:23559936	WGSS	T>G	E>A	N-terminal	13.0	NA	Amplification

序列捕获/Targeted sequencing

nature Vol 461|10 September 2009|doi:10.1038/nature08250

LETTERS

Targeted capture and massively parallel sequencing of 12 human exomes

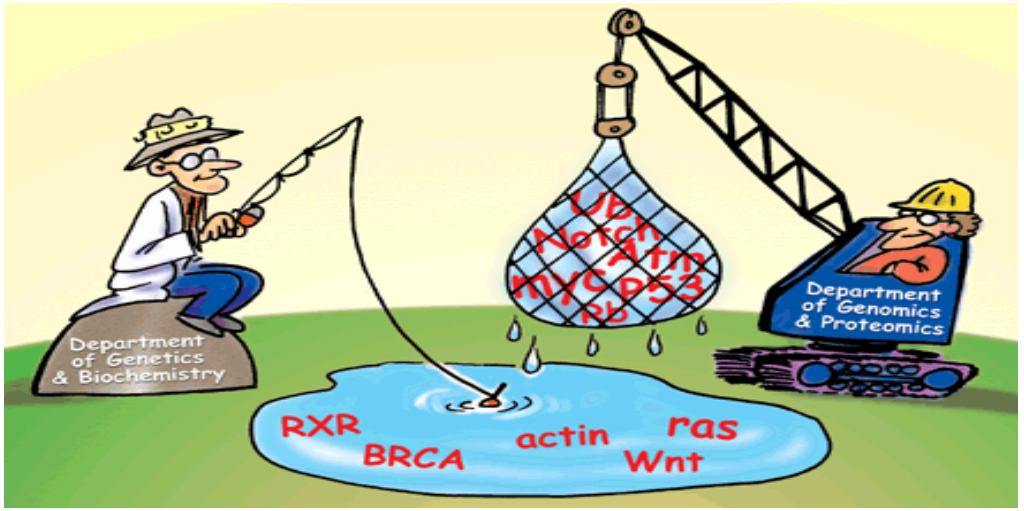
Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[4], Evan E. Eichler[1,3], Michael Bamshad[2], Deborah A. Nickerson[1] & Jay Shendure[1]

Genome-wide association studies suggest that common genetic variants explain only a modest fraction of heritable risk for common diseases, raising the question of whether rare variants account for a significant fraction of unexplained heritability[1,2]. Although DNA sequencing costs have fallen markedly[3], they remain far from what is necessary for rare and novel variants to be routinely identified at a genome-wide scale in large cohorts. We have therefore sought to develop second-generation methods for targeted sequencing of all protein-coding regions ('exomes'), to reduce costs while enriching for discovery of highly penetrant variants. Here we report on the targeted capture and massively parallel sequencing of the exomes of 12 humans. These include eight HapMap individuals representing three populations[4], and four unrelated individuals with a rare dominantly inherited disorder, Freeman-Sheldon syndrome (FSS)[5]. We demonstrate the sensitive and specific identification of ... coding ... candidate ... sequencing. This ... genetics ... non-synonymous ... or ... force ... is ... ration ... genome ... exome ... µg of ... To ... ped a ... start- ... target ... NCBI ... ids the ... rowder ... ion of ... length ... nome ... previ- ... ictural ... iduals ... delian ... arthro- ... ed by mutations in *MYH3* (ref. 5). Unpaired, 76 base-pair (bp) reads[12] from post-enrichment shotgun libraries were aligned to the reference genome[13]. On average, 6.4 gigabases (Gb) of mappable sequence was generated per individual (20-fold less than whole genome sequencing with the same platform[12]), and 69% of reads mapped to targets (Supplementary Table 1). After removing duplicate reads that represent potential polymerase chain reaction artefacts[14], the average fold-coverage of each exome was 51× (Supplementary Fig. 1). On average per exome, 99.7% of targeted bases were covered at least once, and 96.3% (25.6 Mb) were covered sufficiently for variant calling (≥8× coverage and Phred-like[15] consensus quality ≥30). This corresponded to 78% of genes having >95% of their coding bases called (Supplementary Fig. 2 and Supplementary Data 2). The average pairwise correlation coefficient between individuals for gene-by-gene coverage was 0.87, consistent with systematic bias in coverage between individual exomes.

False positives and false negatives are critical issues in genomic resequencing. We assessed the quality of our exome data in four ways. First, comparing sequence-based calls for the eight HapMap exomes to array-based genotyping, we observed a high concordance with both homozygous (99.94%; $n = 219,077$) and heterozygous (99.57%; $n = 43,070$) genotypes (Table 1). Second, we compared our coding single-nucleotide polymorphism (cSNP) catalogue to ~1 Mb of coding sequence determined in each of the eight HapMap individuals by molecular inversion probe (MIP) capture and direct resequencing[16]. At coordinates called in both data sets, 99.9% of all cSNPs ($n = 4,620$) and 100% of novel cSNPs ($n = 334$) identified here were concordant, consistent with a low false discovery rate. Third, we compared the NA18507 cSNPs identified here to those called by recent whole genome sequencing of this individual[11], and found substantial overlap (Supplementary Fig. 3). The relative numbers of cSNPs called by only one approach, and the proportions of these represented in dbSNP, indicate that exome sequencing has equivalent sensitivity for cSNP detection compared to whole genome sequencing. Fourth, we compared our data to cSNPs in high-quality Sanger sequence of single haplotype regions from fosmid clones of the same HapMap individuals[17]. Most fosmid-defined cSNPs (38 of 40) were at coordinates with sufficient coverage in our data for variant calling. Of these, 38 of 38 were correctly identified as variant.

A comparison of our data to past reports on exonic[18] or exomic[8] array-based capture revealed roughly equivalent capture specificity, but greater completeness in terms of coverage and variant calling (Supplementary Table 2). These improvements probably arise from a combination of greater sequencing depth and differences in array designs and in experimental conditions for capture. Within the set of

[1]Department of Genome Sciences, [2]Department of Pediatrics, University of Washington, [3]Howard Hughes Medical Institute, Seattle, Washington 98195, USA. [4]Agilent Technologies, Santa Clara, California 95051, USA.

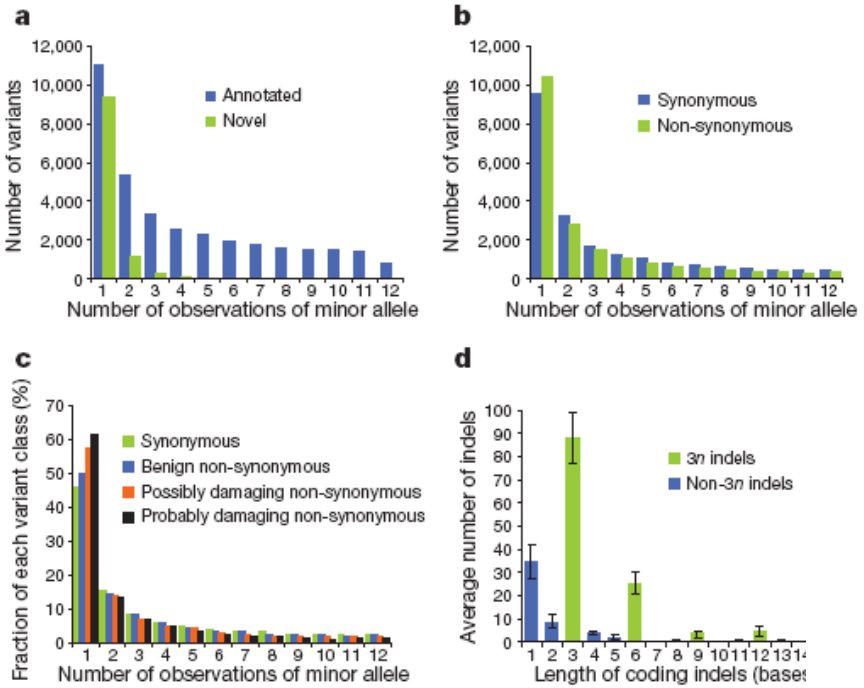


采用Target enrichment(Agilent)+ NGS(solexa)的技术，对一个已知疾病，*Freeman–Sheldon syndrome*（弗-谢二氏综合征, 颅腕跗营养不良）进行基因组突变扫描

探索利用外显子测序技术通过对小量样品（相对于全基因组扫描），进行孟德尔遗传疾病候选基因筛查的技术可行性

Nature September，2009

序列捕获/Targeted sequencing



在12个样品中经过数据过滤后发现的候选SNP位点分布情况

◆A:已知变化 vs 新发现变化

◆B：同义突变 vs 非同义突变

◆C：同义突变，良性突变，可疑突变和致病突变的的分布

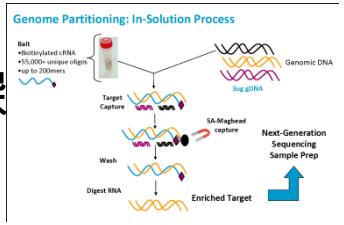
◆D：3碱基缺失突变 vs 非三碱基缺失突变

通过测序在多个样品中定位的致病基因

Number of genes in which each affected has at least one...	FSS24895	FSS24895 FSS10208	FSS24895 FSS10208 FSS10066	FSS24895 FSS10208 FSS10066 FSS22194	Any 3 of 4 FSS24895 FSS10208 FSS10066 FSS22194
Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,264	2,765	2,479	3,768
NS/SS/I not in dbSNP	513	128	71	53	119
NS/SS/I not in eight HapMap exomes	799	168	53	21	160
NS/SS/I neither in dbSNP nor eight HapMap exomes	360	38	8	1 (*MYH3*)	22
...And predicted to be damaging	160	10	2	1 (*MYH3*)	3

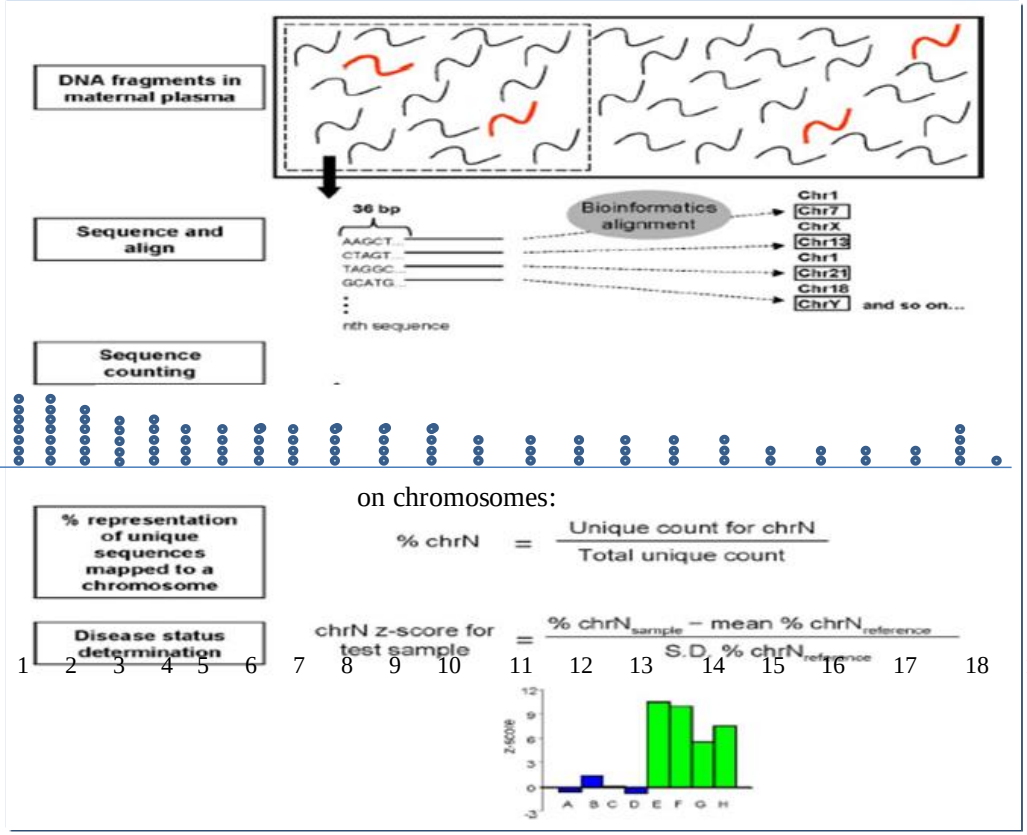
序列捕获的主要平台

- 对感兴趣的DNA片段进行测序
- 在不增加测序成本的前提下提高目的区段测序的覆盖深
- 采用的技术平台
  - ➢ 捕获
    - Agilent SureSelect（固相/液相）
    - Nimblegen Sequence Capture（固相）
  - ➢ 测序
    - SOLiD
    - Solexa



NimbleGen Sequence Capture
2.1M Human Exome Array
Capture the human exome on a single array

无创产前基因检测：母血浆DNA+高通量测序



母体和胎儿ＤＮＡ的混合物

Unique reads (单一序列读取）**: no mismatch, one location**

(Lo *et al*, PNAS, 2008)

孕妇外周血中存在胎儿DNA

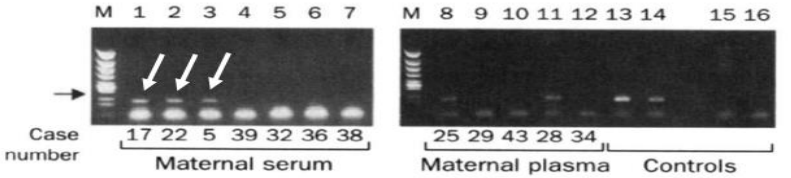
THE LANCET

Early report

Lancet 1997; **350:** 485–87

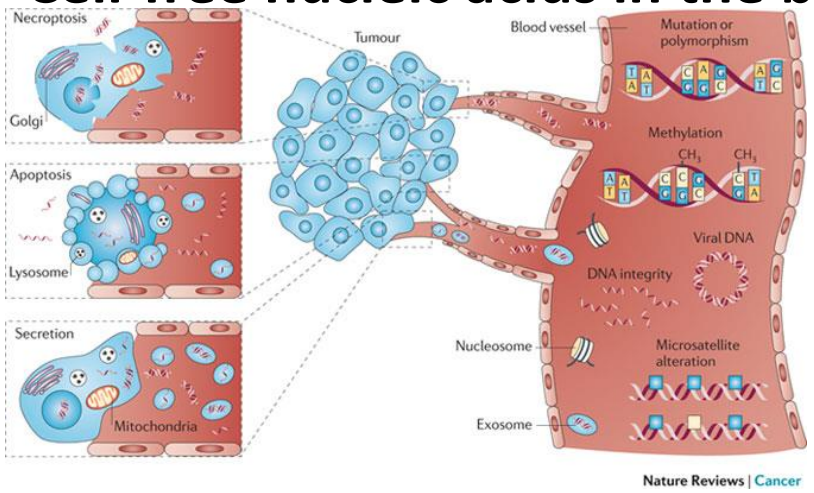
Presence of fetal DNA in maternal plasma and serum

Y M Dennis Lo, Noemi Corbetta, Paul F Chamberlain, Vik Rai, Ian L Sargent, Christopher W G Redman, James S Wainscoat



Amplification of fetal Y-chromosomal sequences from maternal plasma and serum

Cell-free nucleic acids in the blo



Cell-free fetal DNA

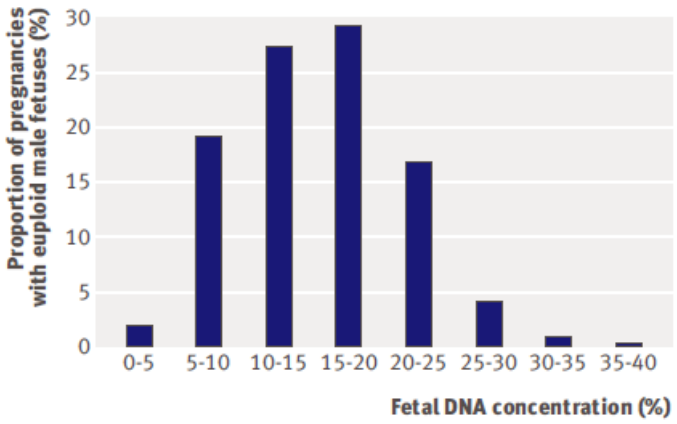
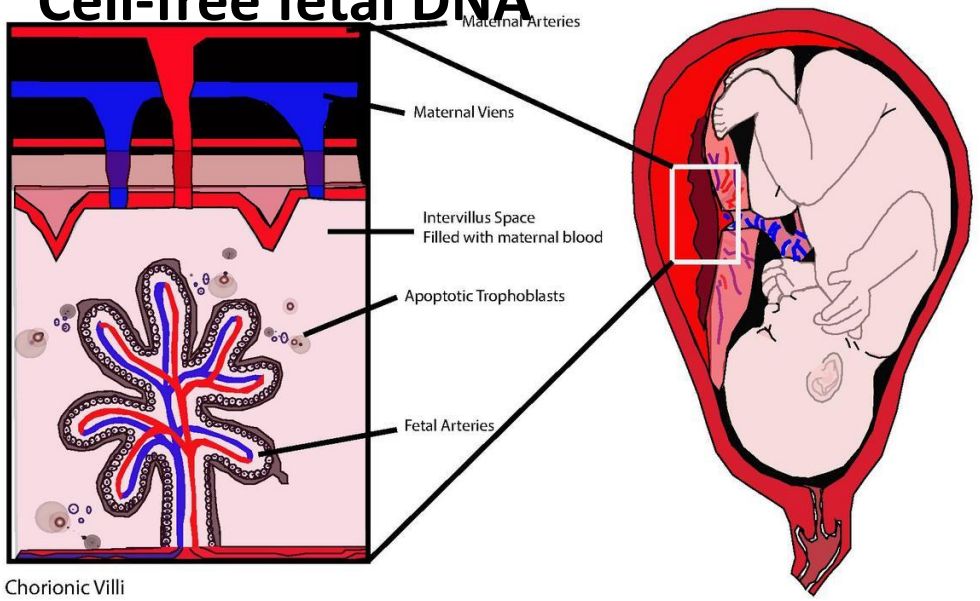


Fig 4| Fetal DNA concentration found in maternal plasma among pregnancies with euploid male fetuses

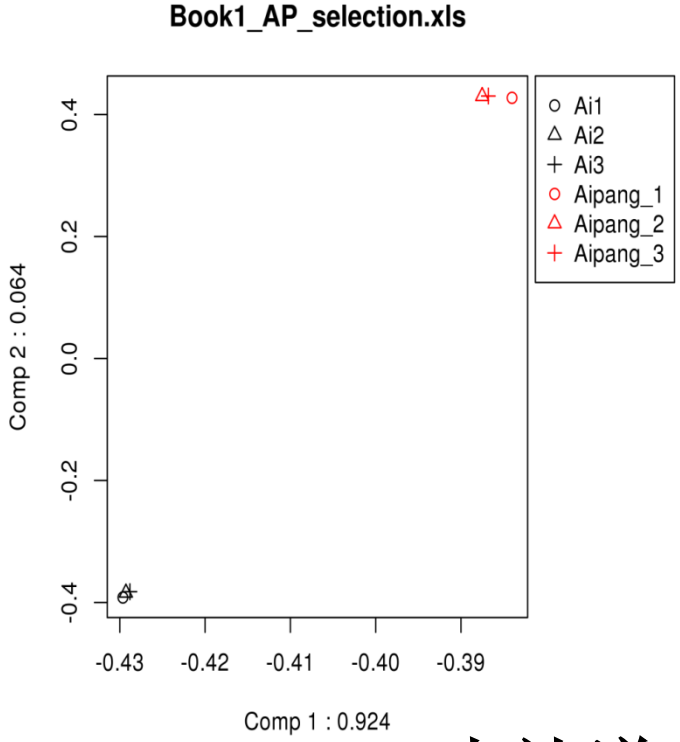


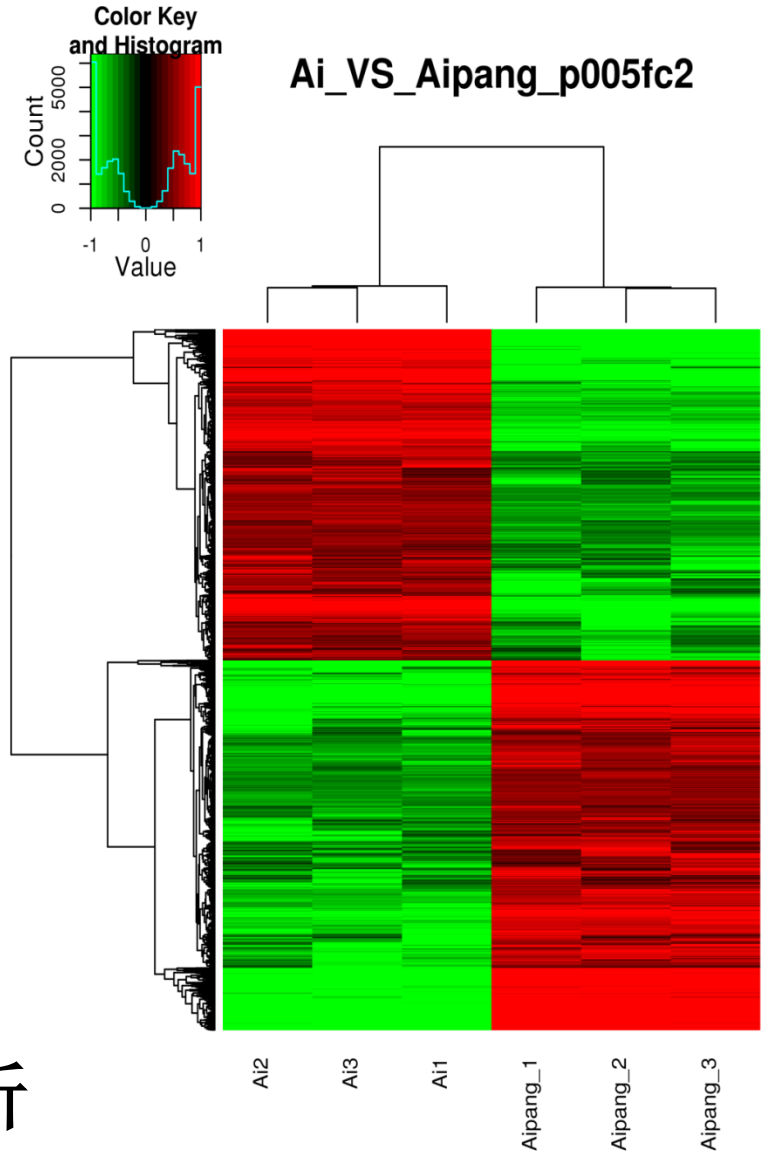
无创产前基因检测的优势：

- 准确性高于99%；
- 早期诊断，12孕周即可检测；减少妊娠晚期引产对孕妇造成的伤害；
- 非侵入性，减轻孕妇心理负担，无流产风险以及产前诊断引起的并发症；
- 可以同时检测其它染色体异常
- 所需样本量少，只需5毫升母体外周血。

实测案例1：结直肠癌的个体化治疗

病例：女，49岁，发病前无任何征兆，通过体检发现CEA和CA19.9高出正常值10倍，后确诊为结直肠癌肝转移。





表达谱分析

药靶筛选（伊立替康）

- 适用表型：TOP1高表达患者该药物疗效好；UGT1A1分型，ATG上游TATA box“TA”拷贝数分析，数量越少药物毒副作用越小。
- 表达谱芯片数据表明癌组织TOP1表达2倍与癌旁
- 测序数据表明UGT1A1的“TA”拷贝数5-6，属于低拷贝基因型
- 结论：推测适用伊立替康（化疗）
- 筛选出5组候选药物组合：
- 1.Irinotecan + Bevacizumab
- 2.Sorafenib + Bevacizumab
- 3.Erlotinib + Irinotecan
- 4.Irinotecan + Decitabine
- 5.5-FU + Cetuximab

后续的PDX的药物敏感性实验结果也验证了伊立替康敏感。

病人的治疗情况

- 2014年5月确诊为肠癌，肝转移，不能手术
- 先进行化疗方案
- 基因表达数据分析结果出来后，增加了伊立替康
- CEA和CA19.9明显下降。
- 2014年8月底，手术切除肠上的病灶。
- 继续化疗。
- 2014年12月底，手术切除肝上病灶。
- CEA和CA19.9恢复到正常水平。

病人的样本取材情况

- 确诊后，取材病人全血和血浆，肠镜新鲜组织样本（癌和癌旁）
- 肠上病灶的新鲜组织样本（癌和癌旁）
- 肝上病灶的新鲜组织样本（癌和癌旁）
- 不同治疗阶段的血浆样本（初诊，化疗后，第一次手术前，手术后，第二次手术前后，第二次手术后半年，一年。。。。。。

实测案例2：乳腺癌肝转移

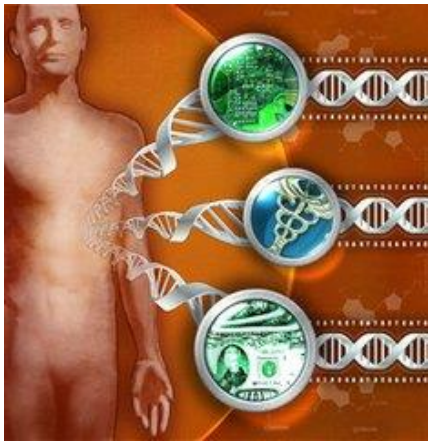
The Open Access Journal of Science and Technology
Vol. 3 (2015), Article ID 101114, 10 pages
doi:10.11131/2015/101114

AgiAl Publishing House
http://www.agialpress.com/

Case Report

A Therapeutic Targeting Identification from Microarray Data and Quantitative Network Analysis

Hongliang Hu[1], Qinghua Zhang[2], Shen Li[3], Xiaona Zhang[2], Junsong Han[2], Huasheng Xiao[2], David Yan[4], Jie Zheng[5], and Biaoru Li[6]



该研究通过对一例乳腺癌肝转移患者进行表达谱RNA分析研究，找到了肝转移特异性高表达基因和乳腺癌特异性高表达基因，在此基础上，筛选特异性“靶向”药物——MPA和doxorubicin，并在临床用药后取得明显成效，为个性化精准医疗提供了一个很好的思路和策略。

Affymetrix U133-A芯片在本研究中用于差异表达基因筛选

病人病史

患者：女，55岁，65kg。于2007年确诊为三阴性右侧浸润性乳腺癌，进行手术切除及6个疗程化疗（顺铂、紫杉醇）、1个疗程放疗。

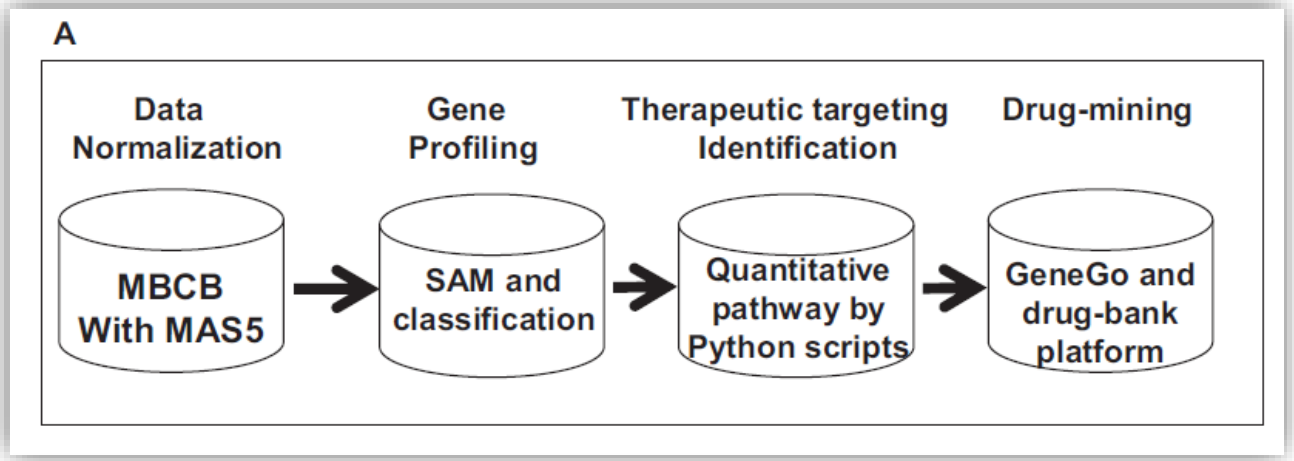
2009年，出现脊柱转移，进行6个疗程化疗、1个疗程放疗。

2011年，出现肝转移，进行1个疗程介入治疗。

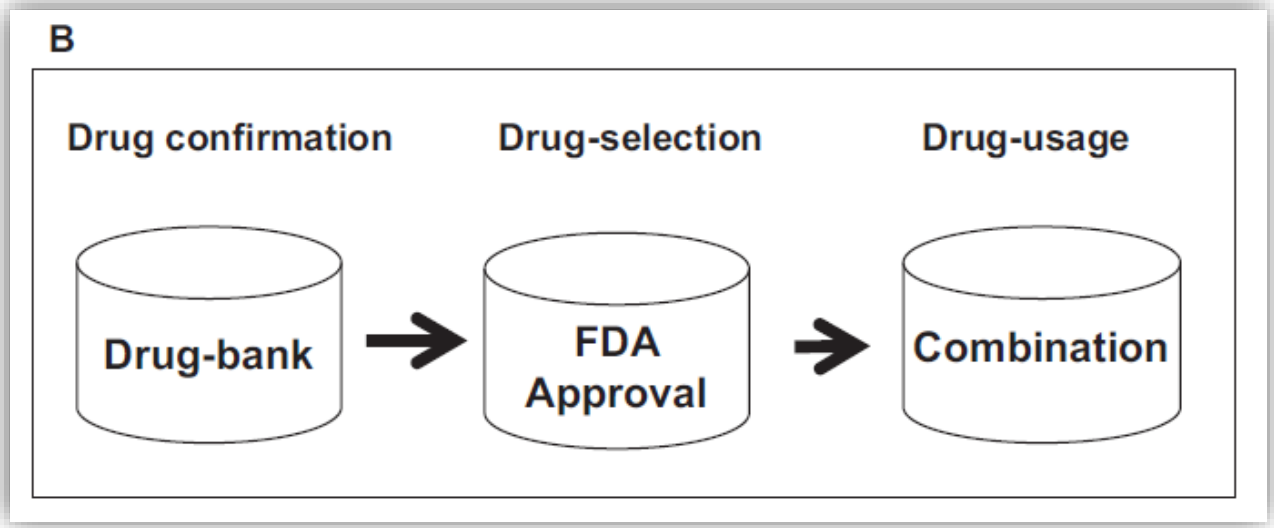
2012年，再次出现肝转移和多淋巴节转移。利用表达谱芯片分析筛选特异性“靶向”药物进行个性化治疗。

个性化治疗整体思路

第一步：遗传学数据整理与分析



第二步：药物发现与运用



乳腺癌相关特异性差异上调基因（set-2）

Table 1: Gene expression level from set-2 profile between breast tumor and normal breast duct.

Gene Symbol	Gene Expression fold-change (metastatic liver/normal liver)	Gene expression (log) from GEO of breast tumor			Gene expression (log) from GEO of normal breast duct		
		GSM134698	GSM134701	GSM134704	GSM134584	GSM134588	GSM134687
AGTPBP1	**3.43**	7.57	7.57	7.62	5.81	5.59	5.99
cct6a	**2.89**	9.09	8.91	8.95	6.98	7.4	7.95
CENPF	**12.48**	8.1	8	8.17	3.89	3.72	5.8
copb2	**2.36**	9.35	9.33	9.35	8.3	7.98	8.02
CUX1	**3.31**	7.59	7.42	7.71	5.7	5.44	6.29
DDX21	**3.45**	8.84	8.94	8.77	7.12	7.52	6.56
fam60a	**10.1**	9.15	9.08	9.06	5.32	6.01	5.91
ISG15	**12.25**	10.7	10.85	10.74	7.29	7.38	6.79
MXRA5	**6.41**	9.21	9.28	8.96	6.16	7.21	5.99
NCBP1	**4.28**	6.96	6.78	6.53	5.52	5.07	5.23
OSBPL10	**3.38**	6.2	6.4	6.41	3.8	5.01	4.94
PLP2	**6.39**	8.41	8.42	8.45	5.93	5.49	5.87
RBMS1	**3.59**	9.49	9.65	9.64	7.35	8.35	7.53
Sulf1	**47.15**	9.37	9.39	9.36	3.32	3.82	4.28
SYNCRIP	**6.25**	8.79	8.73	8.86	5.99	5.74	6.71
TAF1D	**6.73**	8.33	8.59	8.12	5.49	6.93	6.47

set-2： primary breast tumor

肝转移特异性高表达基因（Set-1）

Table 2: Results of genomic expression signature from set-1 profile.

Gene Symbol	Patients Gene Expression Fold-change (metastatic liver/normal liver)	Connectivity	Betweenness
HSPA1A	3.41	6	0.015
HIST1H2BG	7.45	3	0.050
EIF3H	3.86	6	0.012
SET	3.50	21	0.042
ST14	4.72	5	0.158
S100A6 (Calcyclin)	16.41	6	0.021
RACGAP1	5.95	17	0.052
ORC3L	4.08	12	0.095
AGTPBP1	3.46	6	0.015
CBX5	3.64	28	0.126

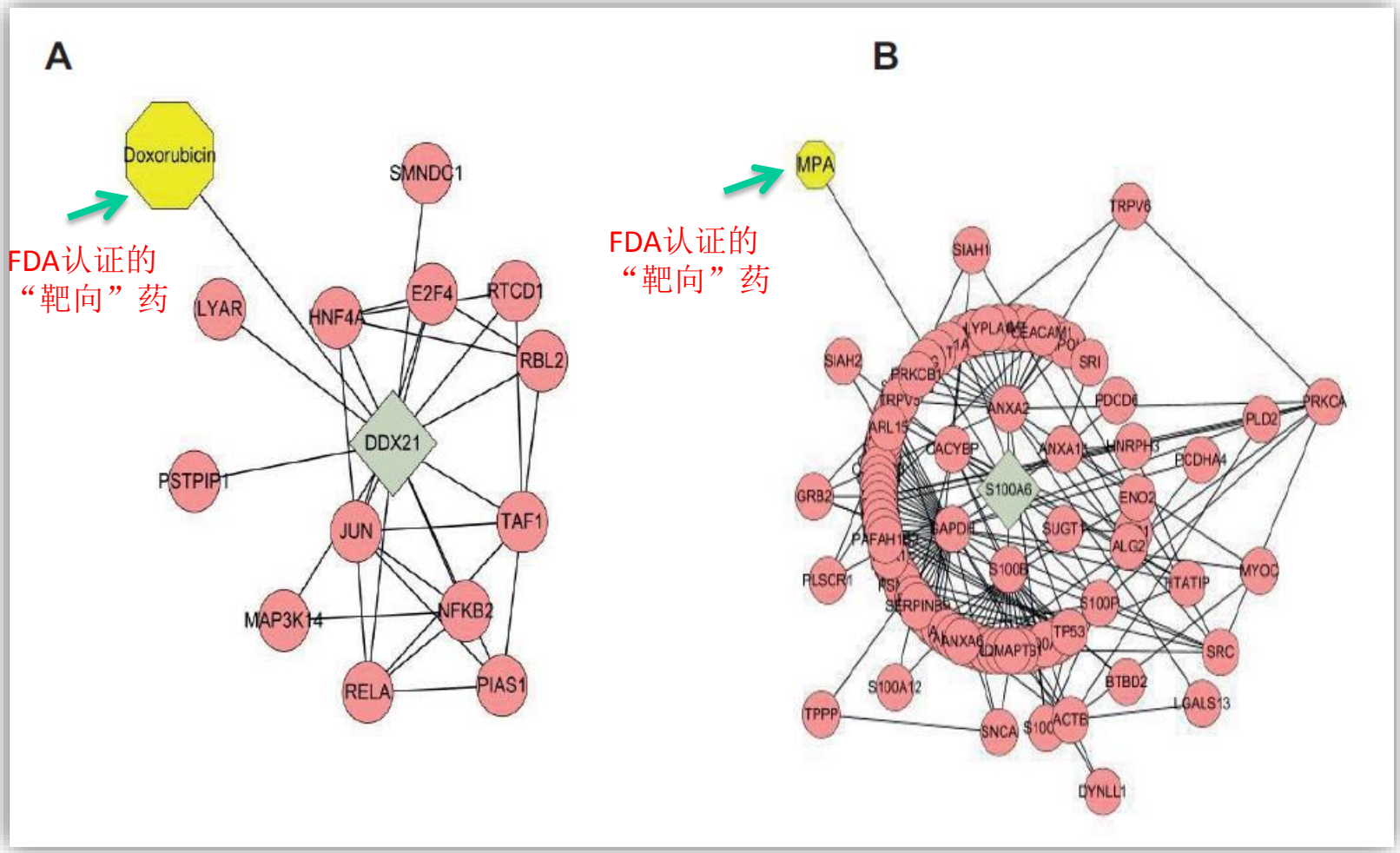
set-1：hepatic metastasis

药物比对分析

Table 3: Result of drug discovery by GeneGo and Genebank platform.

Drug targeting	Drug names	Drug targeting proteins	Drug targeting function
Primary breast ductal tumor	GSK923295	CENP-E	Inhibition
Primary breast ductal tumor	Doxorubicin	DDX21	Inhibition
Liver metastasis	Medroxyprogesterone (MPA)	Calcyclin (S100A6)	Inhibition
Liver metastasis	Secobarbital	RacGAP1	Inhibition
Liver metastasis	Pentobarbital	RacGAP1	Inhibition
Liver metastasis	Pentamidine	ST14 (Matriptase)	Inhibition

候选基因Pathway网络图



Set-2　　Set-1

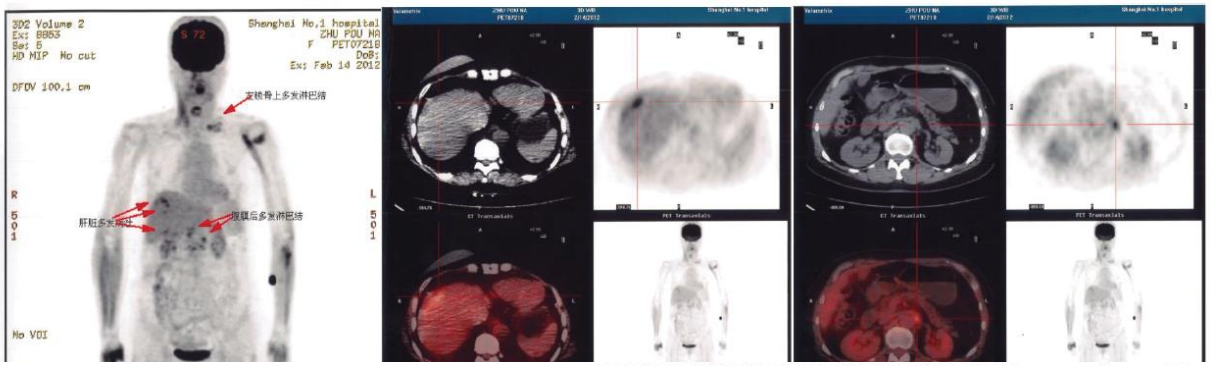
个性化治疗方案

Table 4: Drug treatment plan and application.

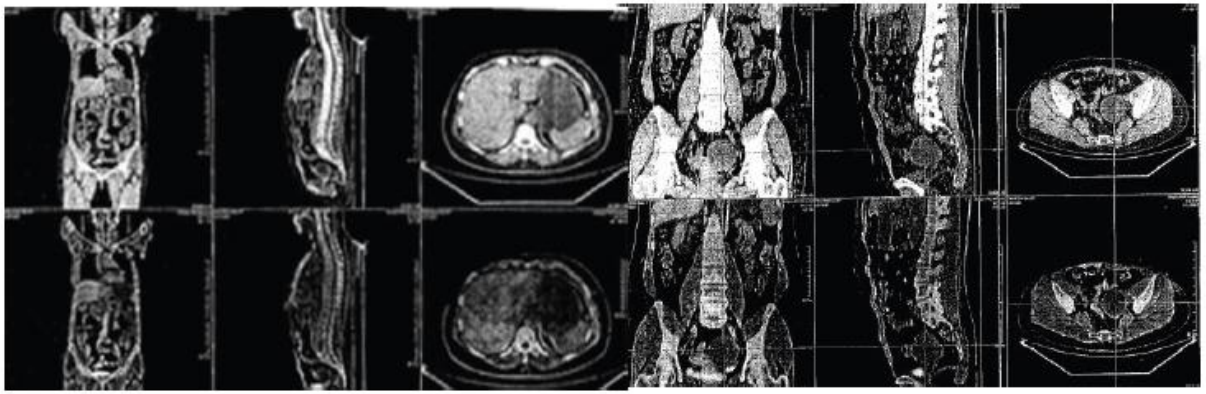
Drugs	**FDA approval**	**Dosage**	**Application methods**
MPA	FDA approved	4 x 250mg/daily/ 6 month	Oral
Doxorubicin	FDA approved	60 mg/m^2/21day/5 cycle	intravenous
Docetaxel	FDA approved	75 mg/m^2/21day/5 cycle	intravenous
Cyclophosphamide	FDA approved	500 mg/m^2/21 day/2 cycle	Intravenous

靶向治疗后效果良好

A



治疗前，患者出现明显多器官肿瘤转移。



Non increase uptake of FDG

个性化治疗3个月后，肝与淋巴管有良好药物反应。