



# 生物信息学

## 第3讲 转录组学与调控网络分析（1）

张高川

遗传学与生物信息学系  
基础医学与生物科学学院  
苏州大学医学部

Email: [zhanggaochuan@suda.edu.cn](mailto:zhanggaochuan@suda.edu.cn)

Tel: 18962111592

WX: [zhanggaochuan770609](#)

QQ: 2257916241



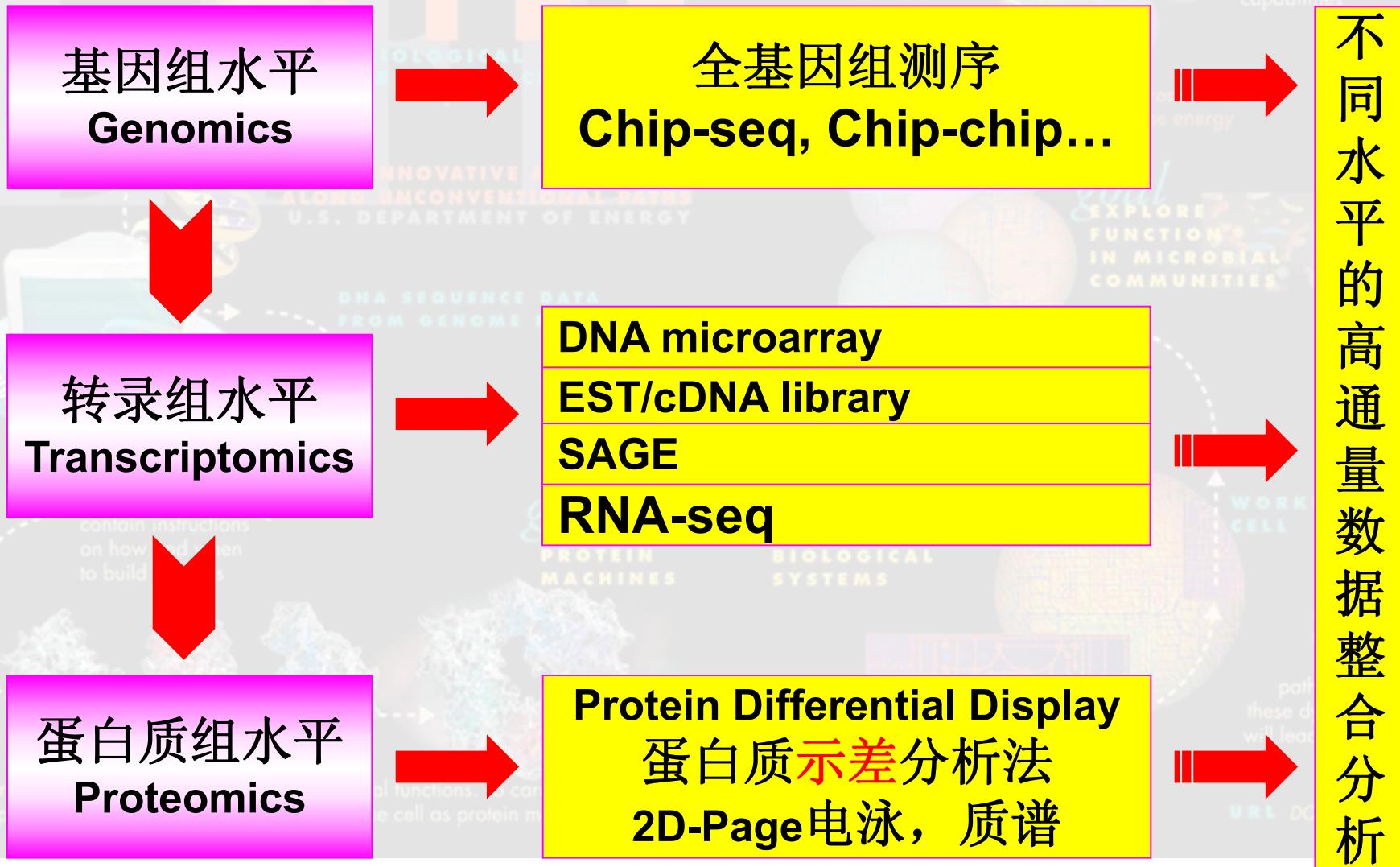
# 回顾 >> 上一讲内容

## NCBI-Genbank-Entrez-Gene

## NCBI-Genbank-Entrez-OMIM



# Introduction



# GENOMES to LIFE

BIOLOGICAL SOLUTIONS FOR ENERGY CHALLENGES

Protect workers and the public  
Clean up the environment  
Sequester excess carbon  
Produce and use energy  
Apply knowledge of microbial functional capabilities

## 什么是转录组和转录组学？

### What is transcriptome and transcriptomics?

Genes and other DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal  
IDENTIFY PROTEIN MACHINES

CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

goal  
CHARACTERIZE GENE REGULATORY NETWORKS

WORKING CELL

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02

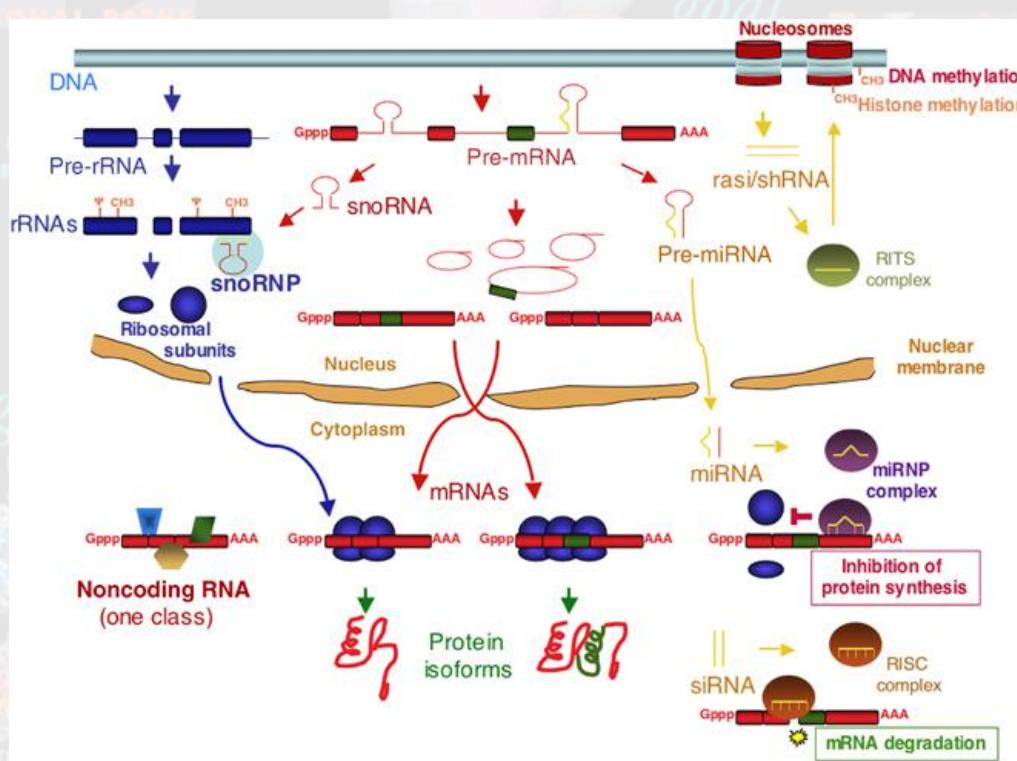




# What is transcriptome and transcriptomics?

# Transcriptome

The transcriptome is the set of *all RNA molecules*, including **mRNA**, **rRNA**, **tRNA**, and other **non-coding RNA** produced in one or a population of cells.





# What is transcriptome and transcriptomics?

## Transcriptome

The transcriptome is the set of ***all RNA molecules***, including **mRNA**, **rRNA**, **tRNA**, and other **non-coding RNA** produced in one or a population of cells.

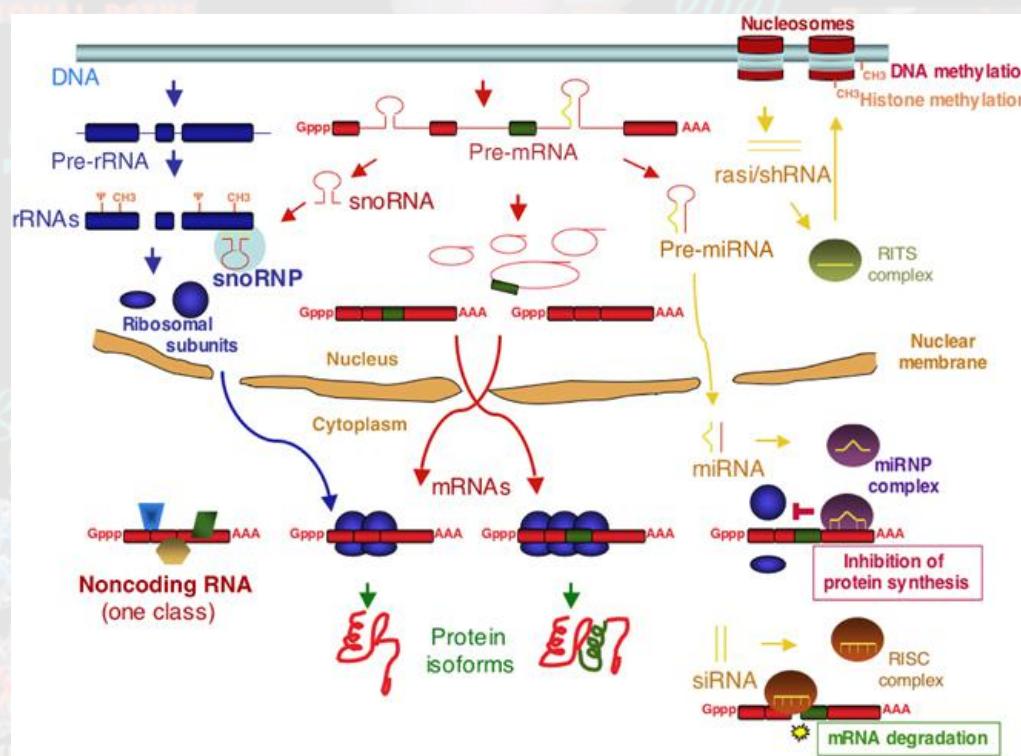


## Transcriptomics

to build proteins



Proteins perform many of life's most essential functions. To carry out specific roles, they often work together in the cell as protein machines.



REGULATORY NETWORKS

URL: DOEgenomes2life.org

10/02





# 转录组学的研究进展如何？

=> “transcriptom\*[tiab]” , 2018.3.18

NCBI Resources How To

Sign in to NCBI

PubMed.gov US National Library of Medicine National Institutes of Health

PubMed transcriptom\*[tiab] Search Create RSS Create alert Advanced Help

Article types Clinical Trial Review Customize ... Text availability Abstract Free full text Full text Publication dates 5 years 10 years Custom range... Species Humans Other Animals Clear all Show additional filters

Format: Summary Sort by: Most Recent Per page: 20 Send to Filters: Manage Filters

**Best matches for transcriptom\*[tiab]:**

[Human Pre-gastrulation Development.](#)  
Wamaitha SE et al. Curr Top Dev Biol. (2018)

[Omics Advances in Ecotoxicology.](#)  
Zhang X et al. Environ Sci Technol. (2018)

[The evolution of mollusc shells.](#)  
McDougall C et al. Wiley Interdiscip Rev Dev Biol. (2018)

Switch to our new best match sort order

**Sort by:** Best match Most recent

**Results by year**

Download CSV

**Search results**

Items: 1 to 20 of 51851 Page 1 of 2593 Next > Last >>

[Tn5Prime, a Tn5 based 5' capture method for single cell RNA-seq.](#)  
1. Cole C, Byrne A, Beaudin AE, Forsberg EC, Vollmers C.  
Nucleic Acids Res. 2018 Mar 14. doi: 10.1093/nar/gky182. [Epub ahead of print]  
PMID: 29548006  
[Similar articles](#)

[The impact of single-cell RNA sequencing on understanding the functional organization of the immune system.](#)  
2. Vegh P, Haniffa M.  
Brief Funct Genomics. 2018 Mar 14. doi: 10.1093/bfgp/ely003. [Epub ahead of print]  
PMID: 29547972  
[Similar articles](#)

[Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic](#)

**Find related data**

Database: Select

[Find Items](#)

**Search details**

transcriptom[tiab] OR transcriptoma[tiab] OR transcriptomal[tiab] OR transcriptomas[tiab] OR transcriptomatic[tiab] OR transcriptome[tiab] OR

Search

苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）

苏州大学校徽、基础医学与生物科学学院院徽

# 转录组研究方法简介 (1)

## >> 传统生化与分子生物学方法

核酸杂交- RNA印迹 (Northern blot)

原位杂交 (ISH, FISH)

逆转录PCR (RT-PCR)

Real-time quantitative PCR  
(RT-qPCR, QT-PCR)

RACE  
(rapid-amplification of cDNA ends)

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# 转录组研究方法简介 (2)

## >> 高通量技术和方法

Apply knowledge of  
microbial functional  
capabilities

### DNA microarray (DNA chip or gene chip)

### Serial analysis of gene expression (SAGE)

### RNA-Seq

### EST/cDNA library

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



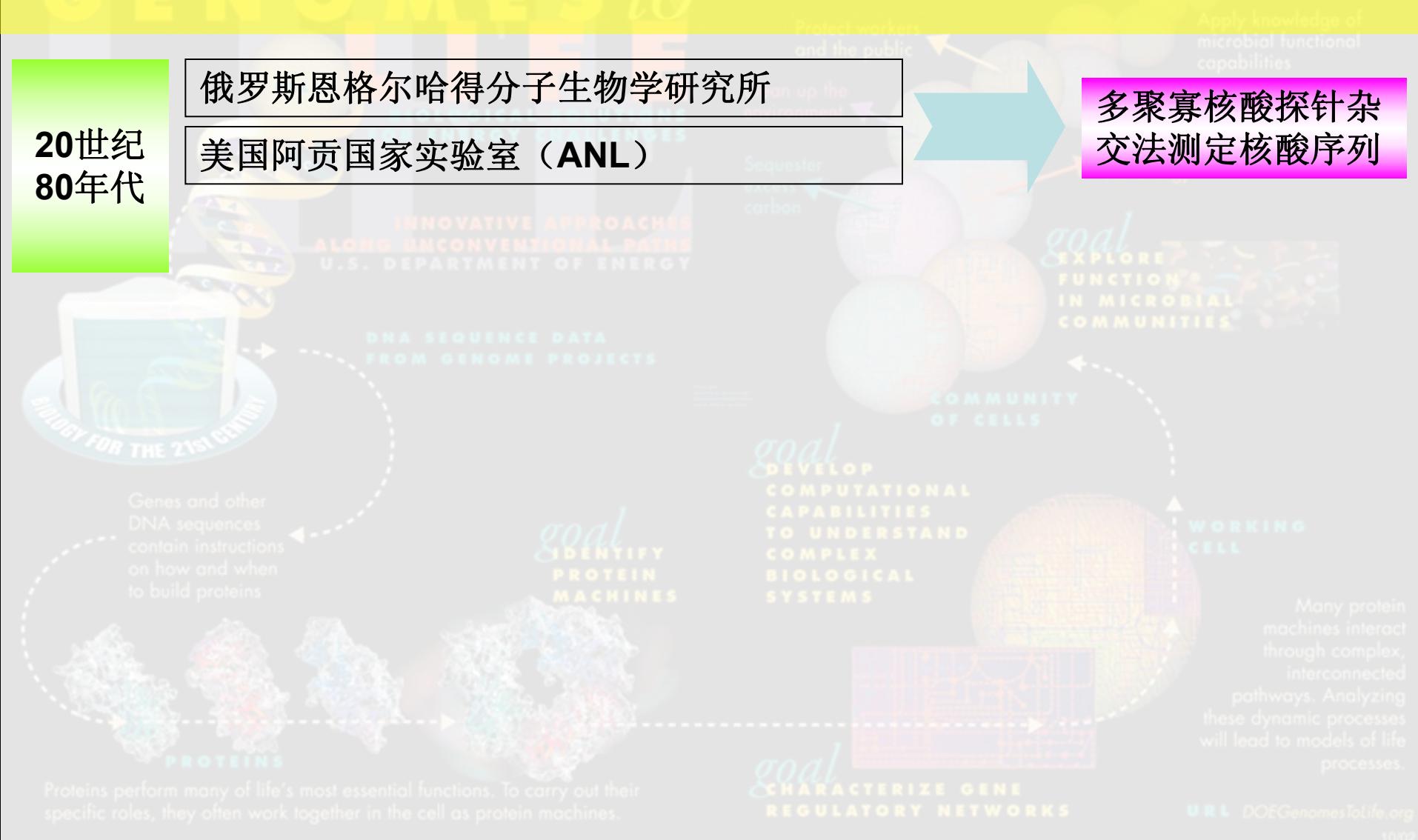
# 发展史简介

20世纪  
80年代

俄罗斯恩格尔哈得分子生物学研究所

美国阿贡国家实验室 (ANL)

多聚寡核酸探针杂交法测定核酸序列



# 发展史简介

20世纪  
80年代

俄罗斯恩格尔哈得分子生物学研究所

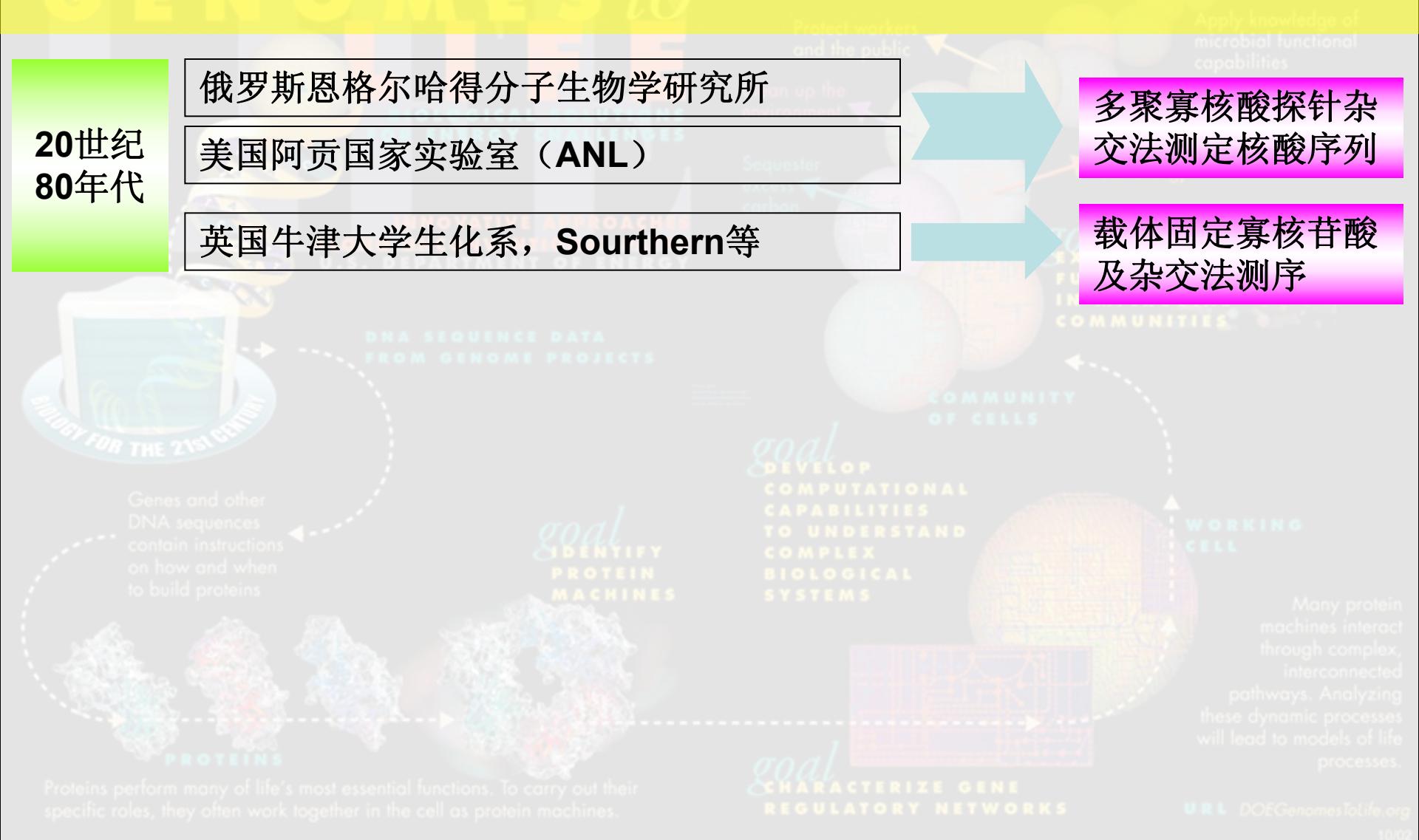
美国阿贡国家实验室（ANL）

英国牛津大学生化系，Sourthern等

Apply knowledge of microbial functional capabilities

多聚寡核酸探针杂交法测定核酸序列

载体固定寡核苷酸及杂交法测序



# 发展史简介

20世纪  
80年代

俄罗斯恩格尔哈得分子生物学研究所

美国阿贡国家实验室（ANL）

英国牛津大学生化系，Southern等

Apply knowledge of microbial functional capabilities

多聚寡核酸探针杂交法测定核酸序列

载体固定寡核苷酸及杂交法测序

1994

美国能源部防御研究计划署

俄罗斯科学院

俄罗斯人类基因组计划

1000多  
万美元

生物芯片

地中海贫血  
病人血样的  
基因突变



# 发展史简介

20世纪  
80年代

俄罗斯恩格尔哈得分子生物学研究所

美国阿贡国家实验室（ANL）

英国牛津大学生化系，Southern等

Apply knowledge of microbial functional capabilities

多聚寡核酸探针杂交法测定核酸序列

载体固定寡核苷酸及杂交法测序

1994

美国能源部防御研究计划署

俄罗斯科学院

俄罗斯人类基因组计划

1000多  
万美元

生物芯片

地中海贫血  
病人血样的  
基因突变

1995

使用小型化的微阵列基因表达分析的首次报道



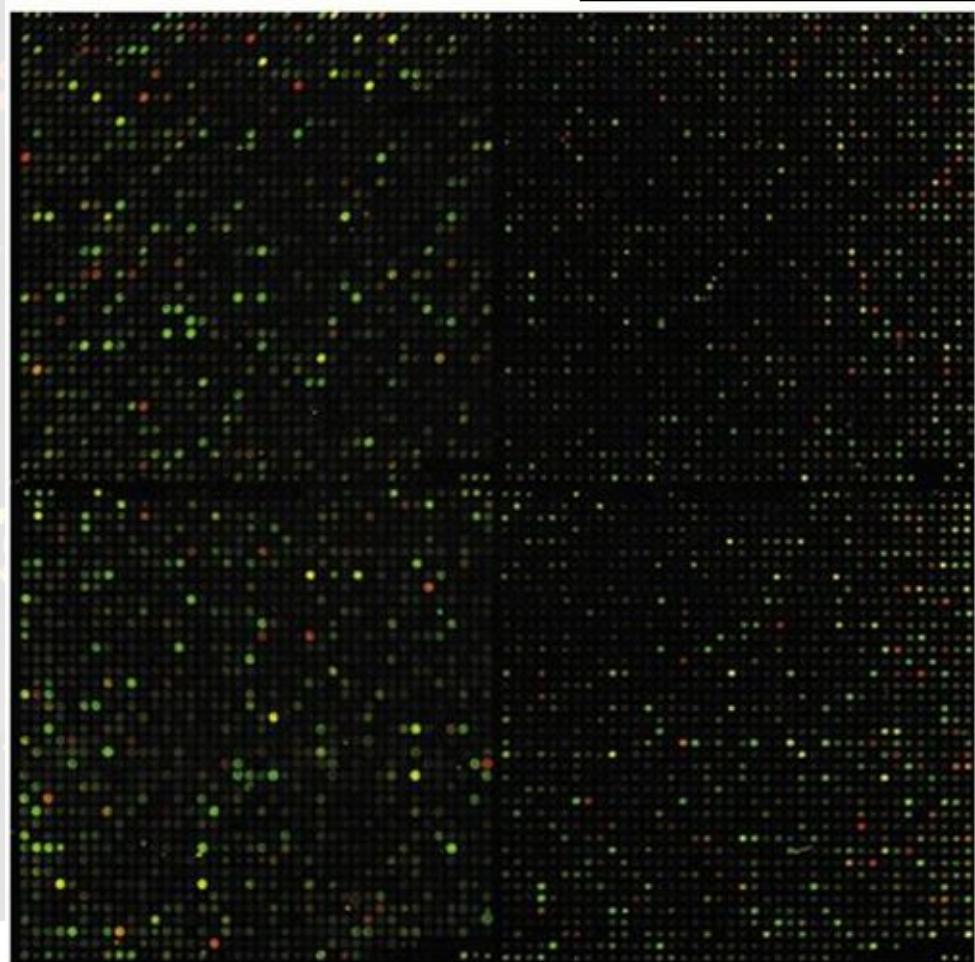
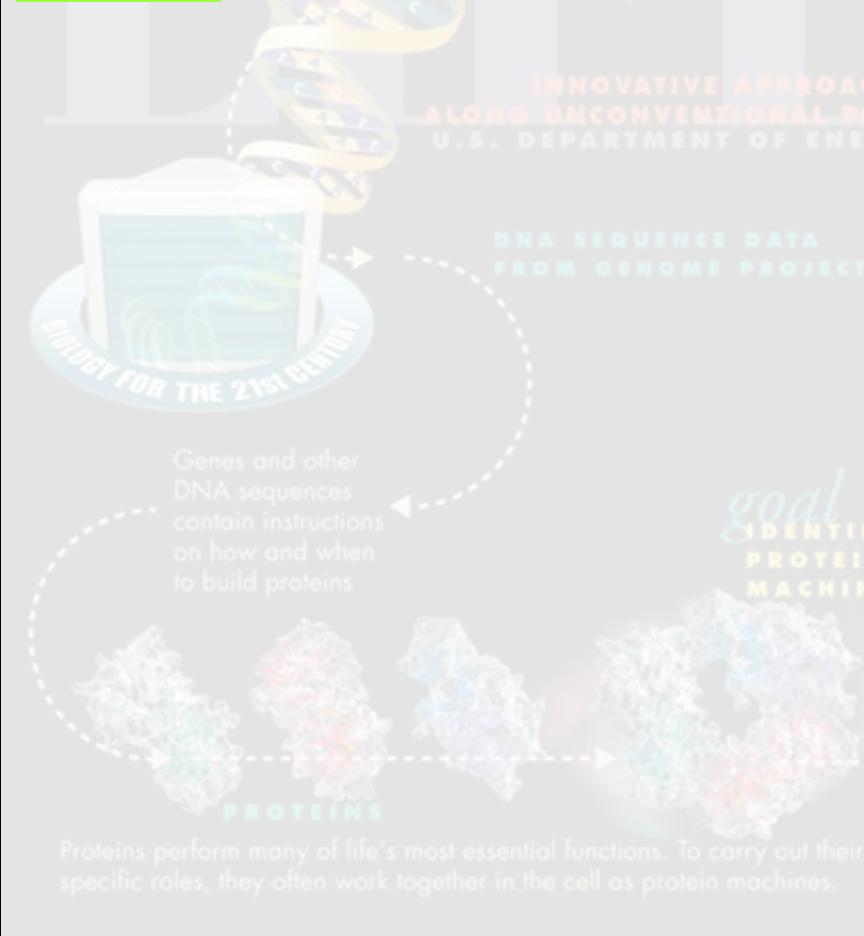
# 发展史简介

1997

斯坦福大学Brown实验室

世界上第一张  
全基因组芯片

含有6166个基因酶  
母全基因组芯片



# 发展史简介

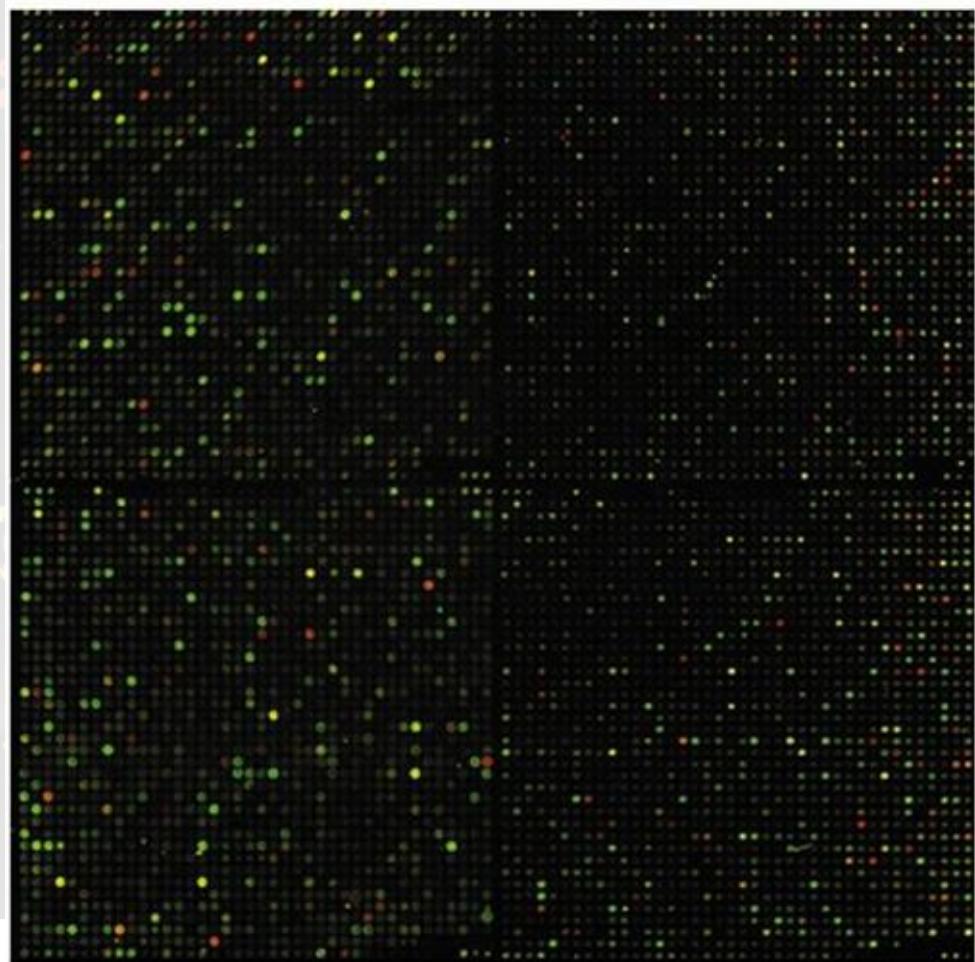
1997

斯坦福大学Brown实验室

世界上第一张  
全基因组芯片

含有6166个基因酶  
母全基因组芯片

如何确保点样准确性?  
如何解读检测结果?

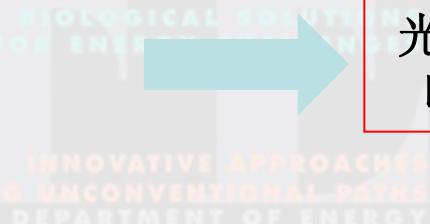


Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.



# 发展史简介（2）

Dr. Stephen Fodor in 1992



光导向平板  
印刷技术

直接在硅片上合  
成寡核苷酸点阵  
的高密度芯片



# 发展史简介（2）

Dr. Stephen Fodor in 1992



光导向平板  
印刷技术

直接在硅片上合  
成寡核苷酸点阵  
的高密度芯片

1994年，第一个产品：一个艾滋病毒基因分型基因芯片



# 发展史简介（2）

Dr. Stephen Fodor in 1992



光导向平板  
印刷技术

直接在硅片上合  
成寡核苷酸点阵  
的高密度芯片

1994年，第一个产品：一个艾滋病毒基因分型基因芯片

基因芯片扫描

流路工作站

计算机软件分析系统



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# 应用领域

## 基因序列分析

## 基因诊断

## 基因表达研究

## 基因组研究

## 新基因发现

## 各种病原体的诊断

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Apply knowledge of microbial functional capabilities

Produce and use energy

goal  
EXPLORE  
FUNCTION  
IN MICROBIAL  
COMMUNITIES

COMMUNITY  
SCIENCE

WORKING  
CELL

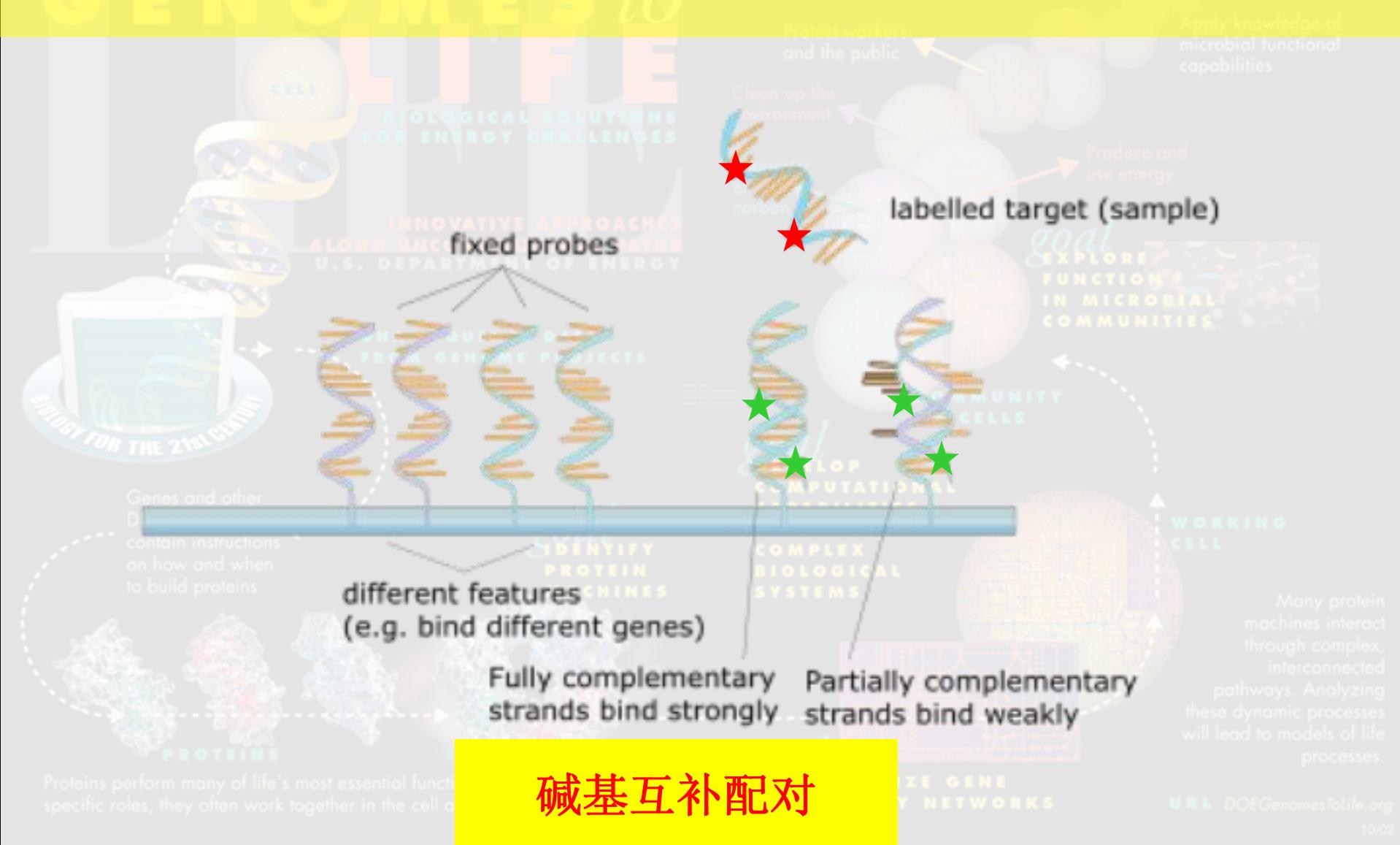
Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

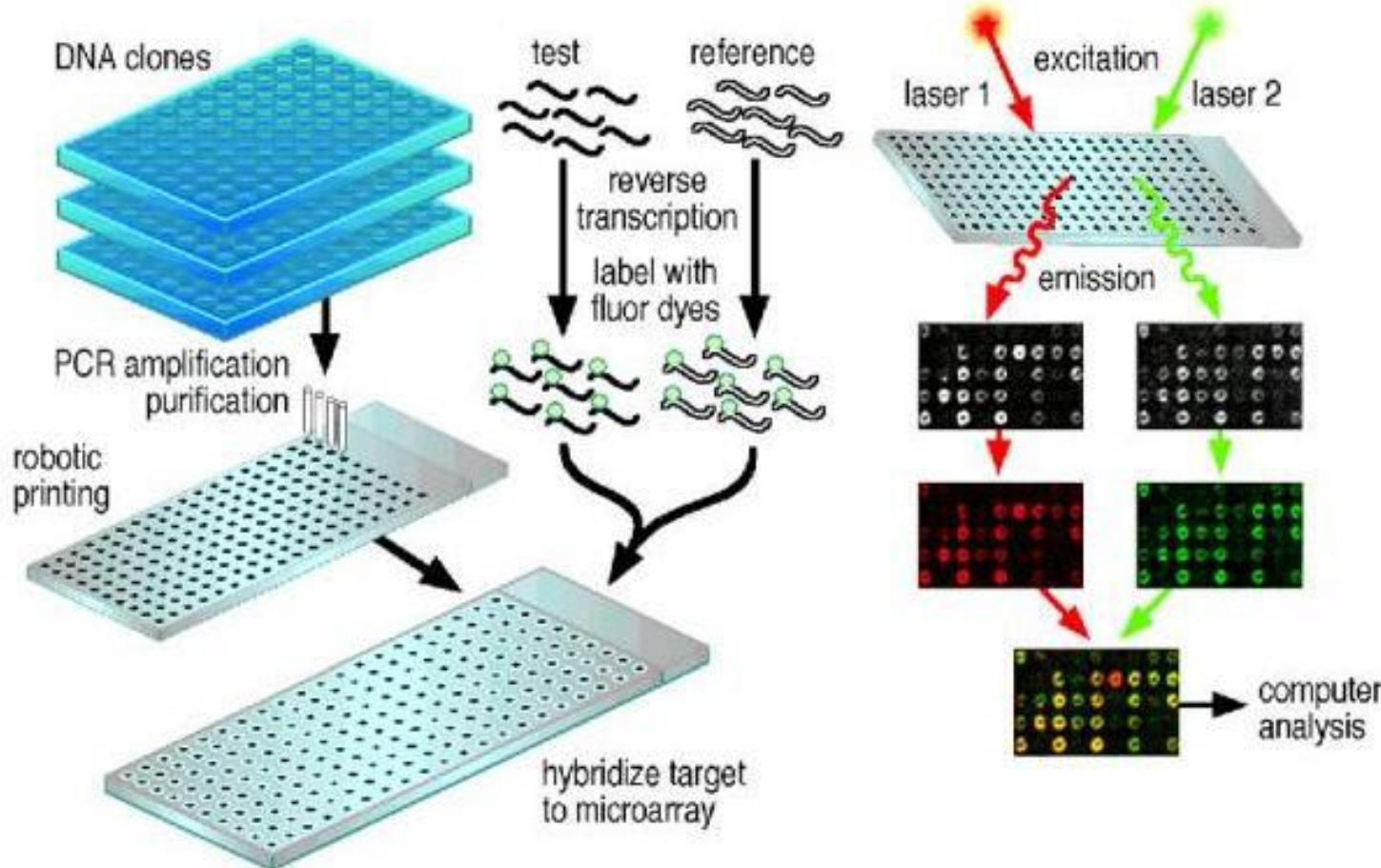
10/02



# 基本原理 (Principle)



# 基本工作流程 (Workflow)



## PROTEINS

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

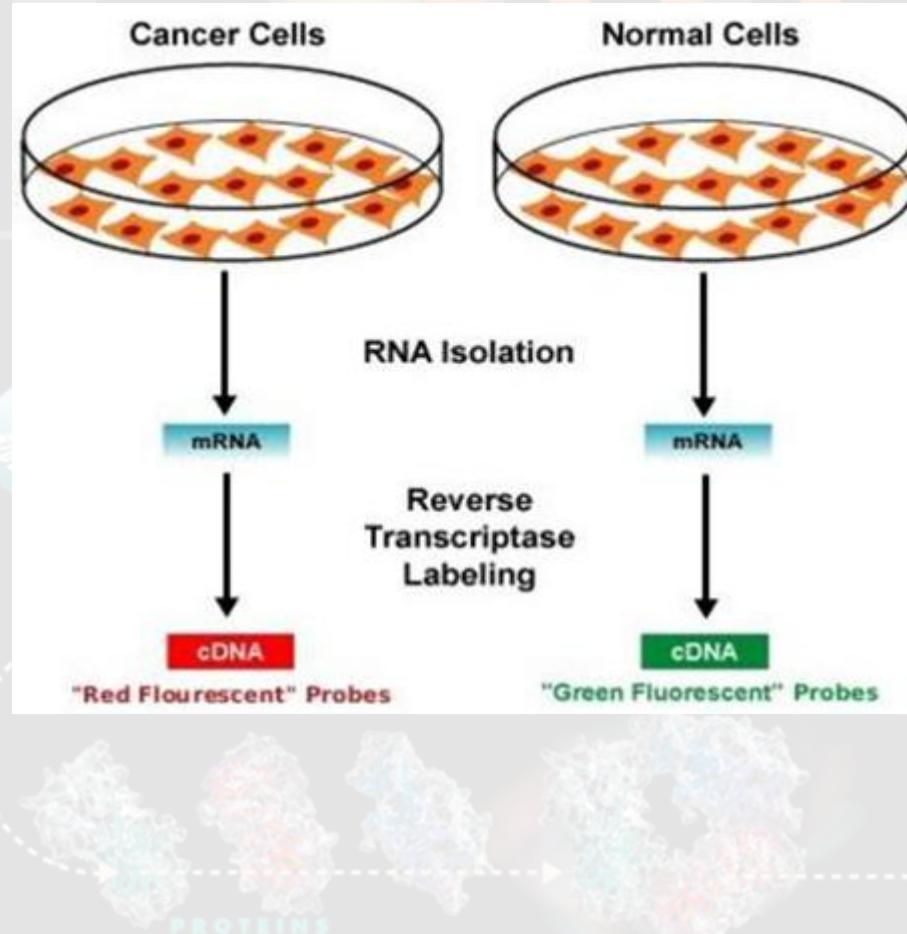
goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

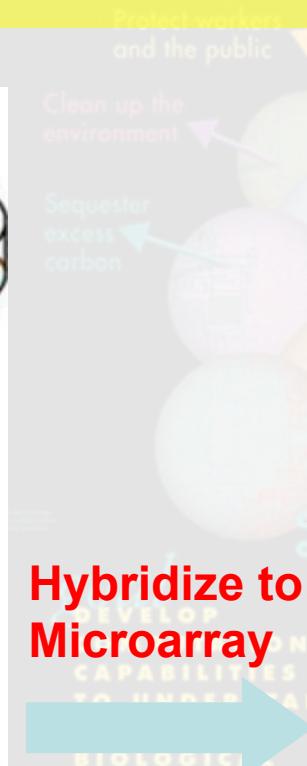
10/02



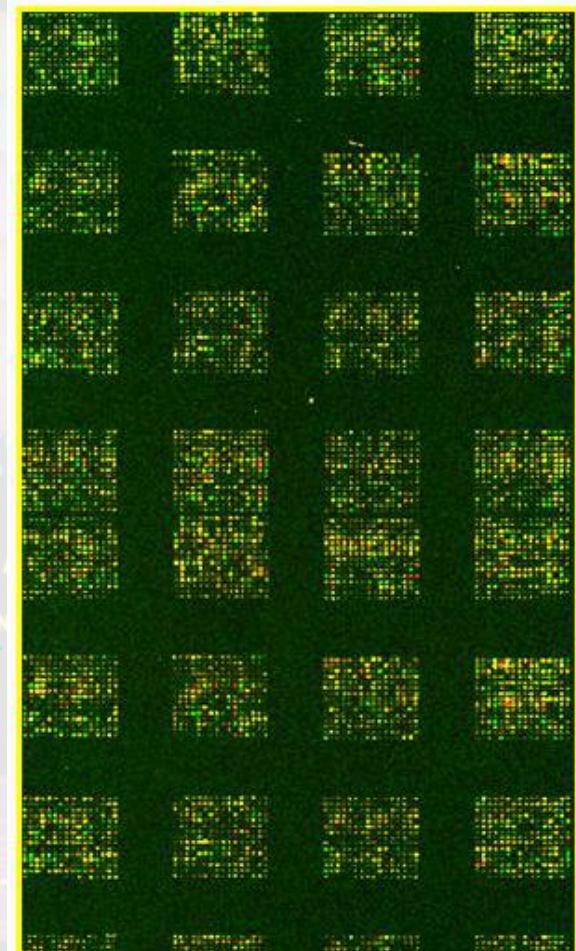
# Scheme of DNA microarray in cancer research



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.



*goal*  
CHARACTERIZE GENE  
REGULATORY NETWORKS

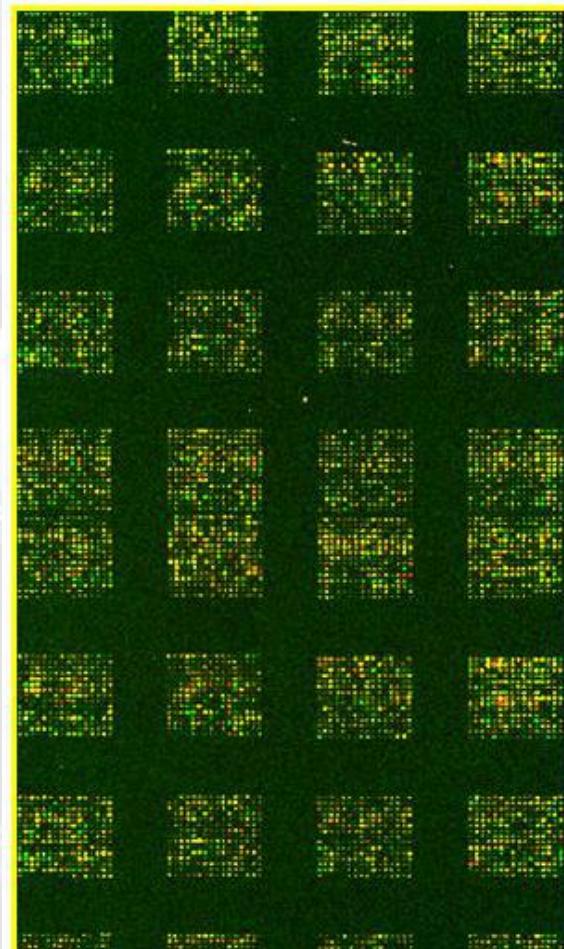


URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

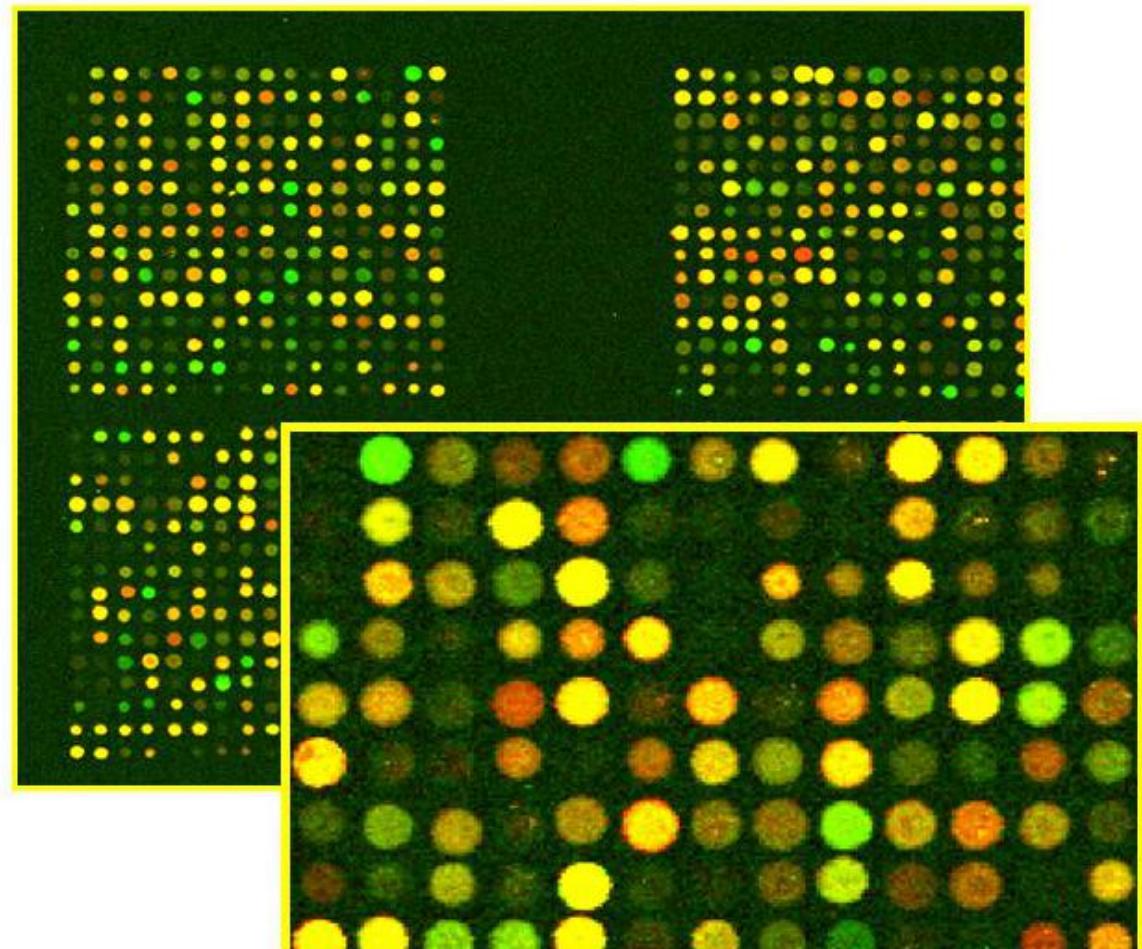
10/02



# Scheme of DNA microarray in cancer research



双通道



单通道





# DNA microarray研究进展如何？

=> “DNA microarray[tiab]” , 2018.3.18

? knowledge of

NCBI Resources How To

PubMed DNA microarray[tiab] Search

PubMed.gov US National Library of Medicine National Institutes of Health Help

**Article types**

Clinical Trial  
Review  
Customize ...

**Text availability**

Abstract  
Free full text  
Full text

**Publication dates**

5 years  
10 years  
Custom range...

**Species**

Humans  
Other Animals

[Clear all](#)

[Show additional filters](#)

Format: Summary Sort by: Most Recent Per page: 20

**Search results**

Items: 1 to 20 of 7156

<< First < Prev Page 1 of 358 Next > Last >>

- [Ensemble of rankers for efficient gene signature extraction in smoke exposure classification.](#)
  1. Giordano M, Tripathi KP, Guaraccino MR. *BMC Bioinformatics*. 2018 Mar 8;19(Suppl 2):48. doi: 10.1186/s12859-018-2035-3. PMID: 29536823 [Free Article](#) [Similar articles](#)
- [Visible DNA microarray system as an adjunctive molecular test in the identification of pathogenic fungi directly from a blood culture bottle.](#)
  2. Sturaro LL, Gonoi T, Busso-Lopes AF, Tararam CA, Levy CE, Lyra L, Trabasso P, Schreiber AZ, Kamei K, Moretti ML. *J Clin Microbiol*. 2018 Mar 7. pii: JCM.01908-17. doi: 10.1128/JCM.01908-17. [Epub ahead of print] PMID: 29514940 [Similar articles](#)
- [Post-translational modification of OCT4 in breast cancer tumorigenesis.](#)
  3. Cho Y, Kang HG, Kim SJ, Lee S, Jee S, Ahn SG, Kang MJ, Song JS, Chung JY, Yi EC, Chun KH. *Cell Death Differ*. 2018 Mar 6. doi: 10.1038/s41418-018-0079-6. [Epub ahead of print] PMID: 29511337 [Similar articles](#)
- [Down-regulation of senescence marker protein 30 by iron-specific chelator deferoxamine drives cell senescence.](#)
  4. Inoue H, Hanawa N, Katsumata-Tsuboi R, Katsumata SI, Takahashi N, Uehara M. *Biosci Biotechnol Biochem*. 2018 Mar 1:1-4. doi: 10.1080/09168451.2018.1440190. [Epub ahead of print] PMID: 29490582

Send to Filters: [Manage Filters](#)

Sort by: Best match Most recent

**Results by year**

Download CSV

**Titles with your search terms**

Profiling DNA methylation differences between inbred mouse strains on the IIII [PLOS One. 2018]

DNA microarray-based resonance light scattering assay for multiplexed c [Analyst. 2018]

Universal informative CpG sites for inferring tumor purity from [J Bioinform Comput Biol. 2017]

[See more...](#)

**Find related data**

Database: [Select](#)

[Find items](#)

苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）



# DNA microarray研究进展如何?





# DNA microarray研究进展如何?

Progress towards  
and the public

Apply knowledge of  
microbial functional  
capabilities



Genes and DNA sequences contain instructions on how to build proteins.

PROTEINS

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02

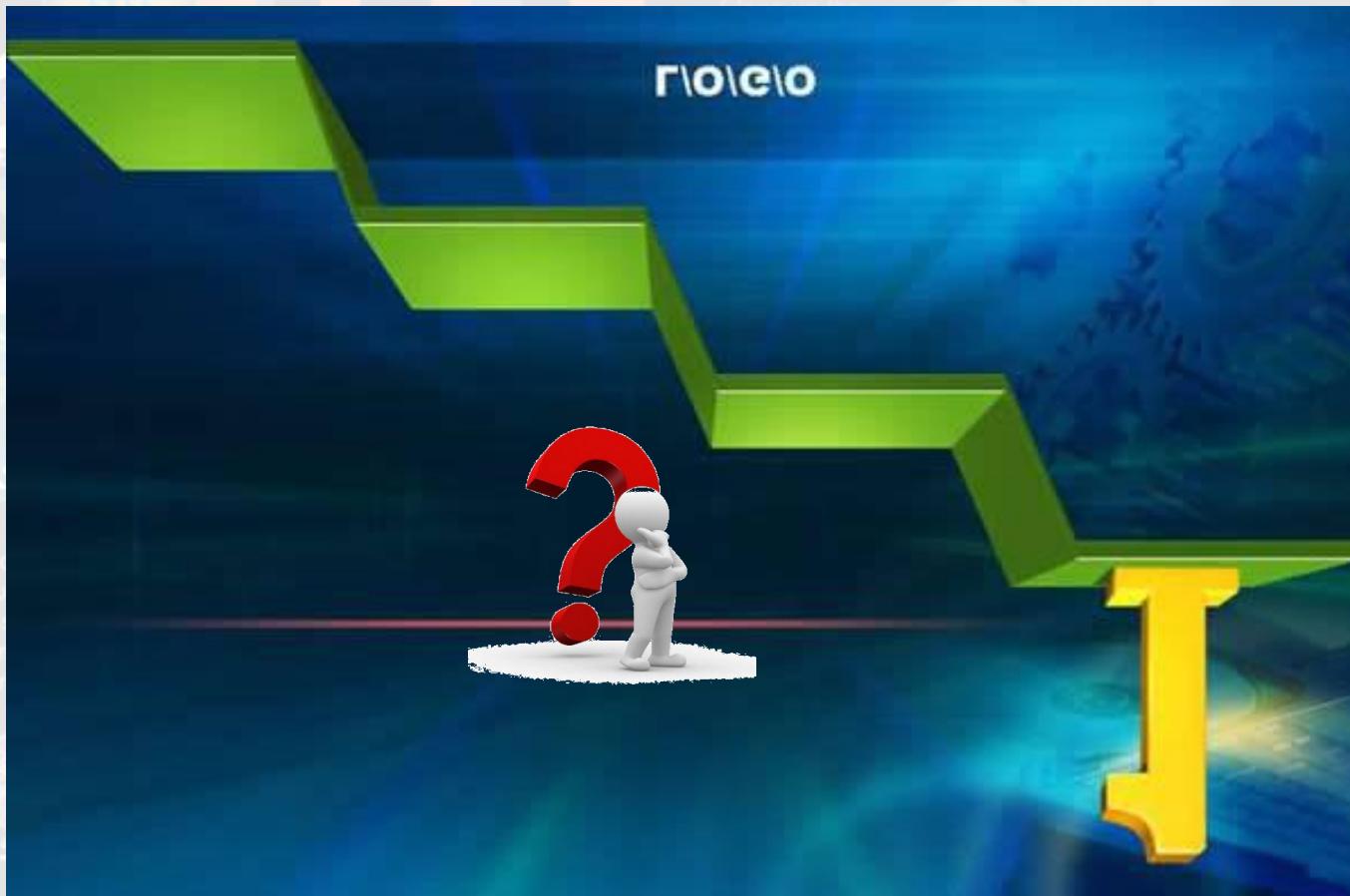




# DNA microarray研究进展如何?

Progress towards  
and the public

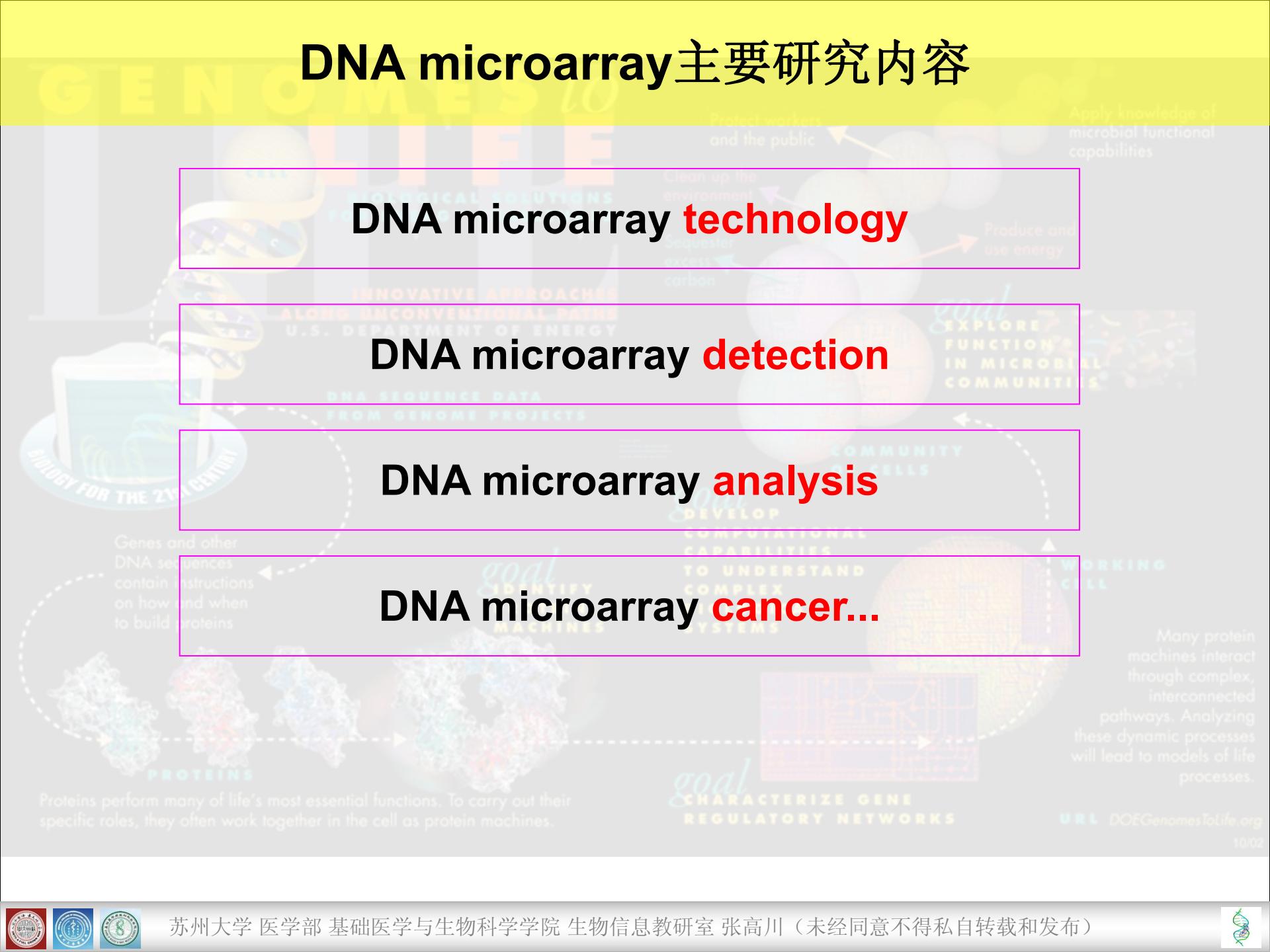
Apply knowledge of  
microbial functional  
capabilities



## 基因芯片技术的固有缺陷



# DNA microarray 主要研究内容



# DNA microarray 分析案例

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis.

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan.

### Abstract

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infectious states. To investigate how to represent LCAP efficacy, we have conducted gene expression analyses from the peripheral blood of RA patients treated with non-woven polyethylene terephthalate filters. Peripheral blood samples were collected immediately before and after treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R.basic package: if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25–370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# DNA microarray 分析案例

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis. 类风湿关节炎

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan.

### Abstract

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infectious states. To investigate how to represent LCAP efficacy, we have conducted gene expression analyses from the peripheral blood of RA patients treated with non-woven polyethylene terephthalate filters. Peripheral blood samples were collected immediately before and after treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R.basic package: if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25–370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# DNA microarray 分析案例

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

Apply knowledge of  
microarray technology

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis.

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan

Ab  
**Welch's t-test**

**R.basic package**

**P<0.05**

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infectious states. To investigate how to represent LCAP efficacy, we have conducted gene expression analyses from the peripheral blood of RA patients treated with non-woven polyethylene terephthalate filters. Peripheral blood samples were collected immediately before and after treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R basic package; if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25–370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# DNA microarray 分析案例

Apply knowledge of  
microarray technology

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis.

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan.

### Abstract

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infi-

**25370 genes**

**2110 up-regulated genes**

**1864 down-regulated genes**

gate how t patients treated with non-woven poly treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R.basic package: if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25 370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# DNA microarray 分析案例

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

Apply knowledge of  
microarray technology

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis.

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan.

### Abstract

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infectious states. To investigate how to represent LCAP efficacy, we have conducted gene expression analyses from the peripheral blood of RA patients treated with non-woven polyethylene terephthalate filters. Peripheral blood samples were collected immediately before and after treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R.basic package: if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25–370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

pathway analysis by MetaCore program

生物学知识



# DNA microarray 分析案例

Apply knowledge of  
microarray technology

Ther Apher Dial. 2012 Oct;16(5):456-66. doi: 10.1111/j.1744-9987.2012.01111.x.

## Gene expression analysis using a high-resolution DNA microarray of peripheral whole blood immediately before and after leukocytapheresis for rheumatoid arthritis. 类风湿关节炎

Kusaoi M, Yamaji K, Murayama G, Yasui M, Yamada R, Hishinuma R, Nemoto T, Hohtatsu K, Kageyama M, Kawamoto T, Sugimoto K, Sekiya F, Kon T, Ogasawara M, Kempe K, Tsuda H, Takasaki Y.

Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine Department of General Medicine, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan

Welch's t-test

R.basic package

P<0.05

Leukocytapheresis (LCAP) is a safe, unique therapy pertaining to intractable rheumatoid arthritis (RA) even in cases of drug allergy or infi-  
25370 genes nate how to 2110 up-regulated genes 1 gene e 1864 down-regulated genes

patients treated with non-woven poly treatment from eight RA patients who received LCAP. Among these patients, all of them achieved 20% improvement in the core set of the American College of Rheumatology (ACR20), and thus, they were confirmed as LCAP responders. Gene expression analysis was done with a high-resolution DNA microarray. The results of each of the two groups' gene expression values (immediately before and after LCAP) were calculated using Welch's t-test. Calculations were performed with a statistical software R.basic package: if the P-value was less than 0.05, this was seen as a significant change. In a comparison of 25 370 gene expressions, the number of genes showing a P-value < 0.05 in the upregulating group was 2110, and in the downregulating group it was 1864. The results of pathway analysis using the MetaCore program indicate that gene groups work for cytoskeletal remodeling are upregulated, and genes related to immune responses, such as antigens presenting via major histocompatibility complex class I and II, are downregulated just after LCAP. These findings may relate to LCAP efficacy for RA patients, but this needs further investigation.

© 2012 The Authors. Therapeutic Apheresis and Dialysis © 2012 International Society for Apheresis.

PMID: 23046371 [PubMed - in process]

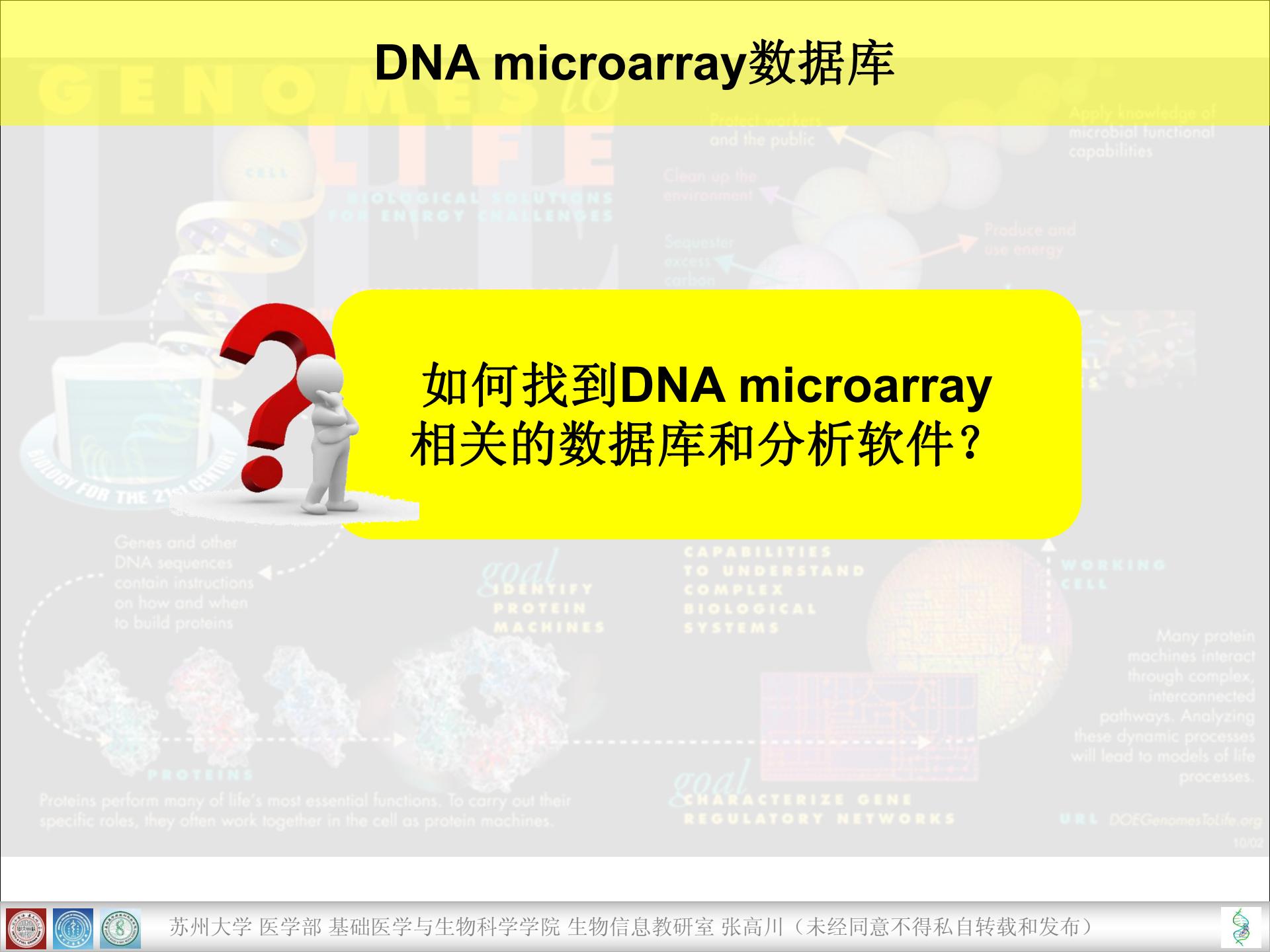
pathway analysis by MetaCore program

生物学知识



# DNA microarray数据库

## 如何找到DNA microarray 相关的数据库和分析软件？



# 转录组研究方法简介 (2)

## >> 高通量技术和方法

DNA microarray

(DNA chip or gene chip)

Serial analysis of gene expression  
(SAGE)

RNA-Seq

EST/cDNA library

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02





# What's SAGE?

A technique used by molecular biologists to produce a **snapshot** of the **mRNA population** in a sample of interest in the form of **small tags** that correspond to fragments of **those transcripts**.

**10bp => 4<sup>10</sup> different tags**



The Oncology Center of Johns Hopkins University

约翰·霍普金斯大学肿瘤中心

Dr. Victor Velculescu

Science. 1995 Oct 20;270(5235):484-7.



goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS



Science. 1995 Oct 20;270(5235):484-7.

## Serial analysis of gene expression.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW.

Source

Oncology Center, Johns Hopkins University, Baltimore, MD 21231, USA.

Apply knowledge of  
microbial functional  
capabilities

### Abstract

The characteristics of an organism are determined by the genes expressed within it.

A method was developed, called serial analysis of gene expression (SAGE), that allows the quantitative and simultaneous analysis of a large number of transcripts.

To demonstrate this strategy, short diagnostic sequence tags were isolated from pancreas, concatenated, and cloned.

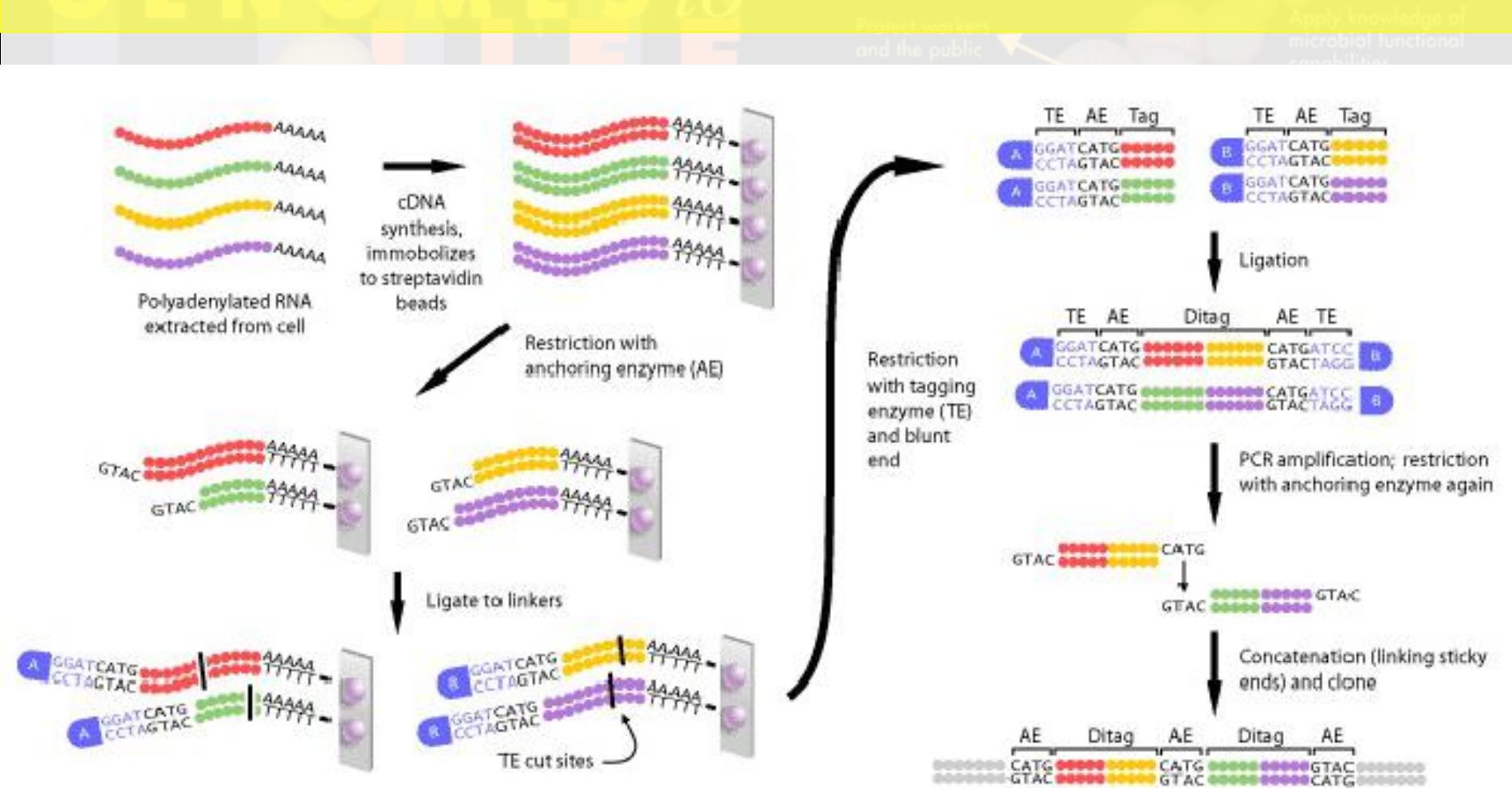
Manual sequencing of **1000 tags** revealed a gene expression pattern characteristic of **pancreatic function**. **New pancreatic transcripts corresponding to novel tags were identified.**

SAGE should provide a broadly applicable means for the quantitative cataloging and comparison of expressed genes in a variety of normal, developmental, and disease states.

PMID: 7570003



# Scheme of SAGE method



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

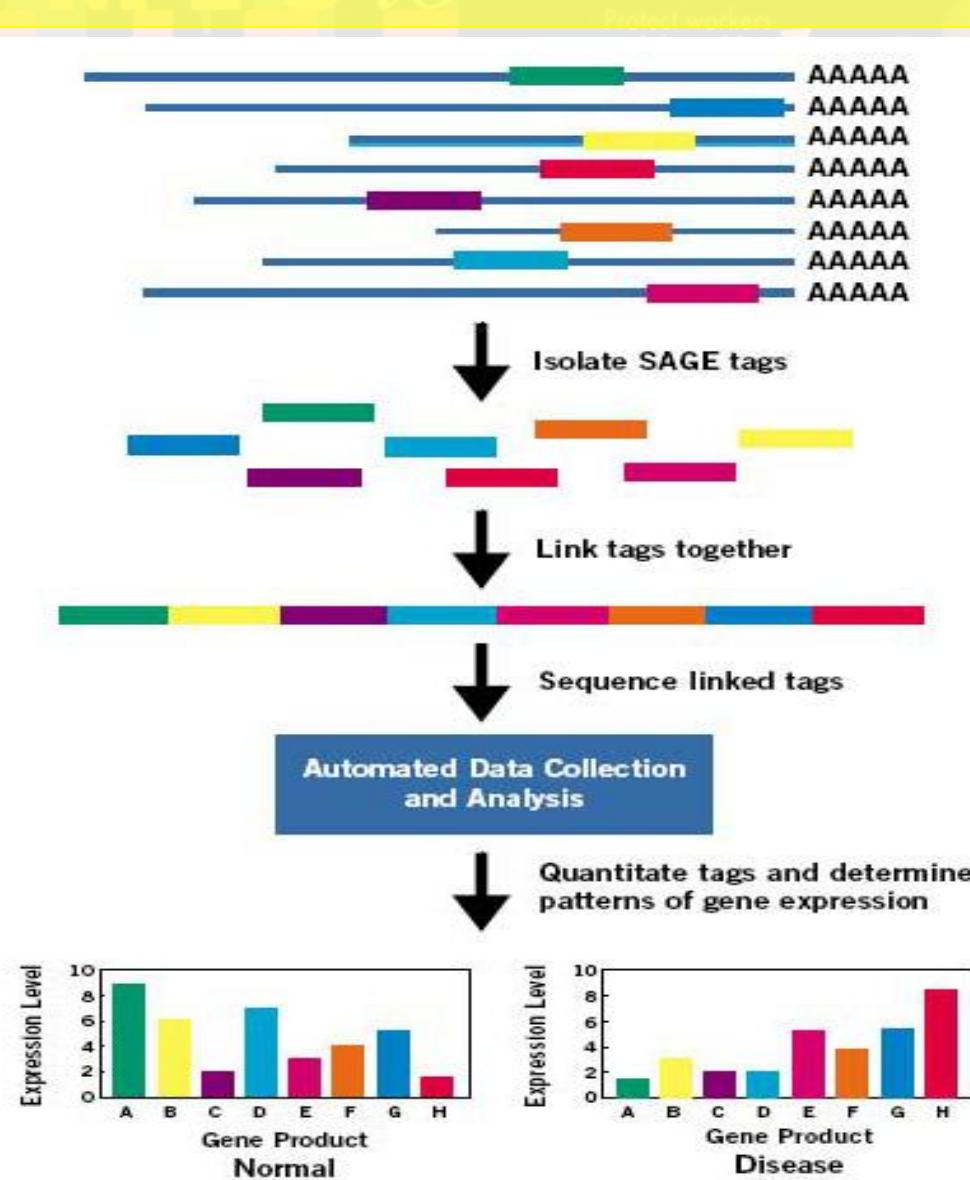
CHARACTERIZE GENE REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# Application of SAGE method



Apply knowledge of microbial functional capabilities

Produce and use energy

Core action of microbial communities

Working cell

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# PubMed search >> "serial analysis of gene expression"

NCBI Resources How To Sign in to NCBI Help

PubMed serial analysis of gene expression Search

Create RSS Create alert Advanced

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#) [Show additional filters](#)

Format: Summary Sort by: Most Recent Per page: 20 Send to Filters: [Manage Filters](#)

Sort by: Best match Most recent

Search results Items: 1 to 20 of 3205 << First < Prev Page 1 of 161 Next > Last >>

[Endogenous Selenoprotein P, a Liver-Derived Secretory Protein, Mediates Myocardial Ischemia/Reperfusion Injury in Mice.](#)  
1. Chadani H, Usui S, Inoue O, Kusayama T, Takashima SI, Kato T, Murai H, Furusho H, Nomura A, Misu H, Takamura T, Kaneko S, Takamura M.  
*Int J Mol Sci.* 2018 Mar 16;19(3). pii: E878. doi: 10.3390/ijms19030878.  
PMID: 29547524 [Similar articles](#)

[Homologous recombination mediates stable Fah gene integration and phenotypic correction in tyrosinaemia mouse-model.](#)  
2. Junge N, Yuan Q, Vu TH, Krooss S, Bednarski C, Balakrishnan A, Cathomen T, Manns MP, Baumann U, Sharma AD, Ott M.  
*World J Hepatol.* 2018 Feb 27;10(2):277-286. doi: 10.4254/wjh.v10.i2.277.  
PMID: 29527263 [Free PMC Article](#) [Similar articles](#)

[Pancreatic tumor microenvironment confers highly malignant properties on pancreatic cancer cells.](#)  
3. Takahashi K, Ehata S, Koinuma D, Morishita Y, Soda M, Mano H, Miyazono K.  
*Oncogene.* 2018 Mar 7. doi: 10.1038/s41388-018-0144-0. [Epub ahead of print]  
PMID: 29511349 [Similar articles](#)

[Loss of CDX2 gene expression is associated with DNA repair proteins and is a crucial member of the Wnt signaling pathway in liver metastasis of colorectal cancer.](#)  
4. Tóth C, Sükösd F, Valicsek E, Herpel E, Schirmacher P, Tiszlavicz L.  
*Oncol Lett.* 2018 Mar;15(3):3586-3593. doi: 10.3892/ol.2018.7756. Epub 2018 Jan 9.  
PMID: 29467879 [Free PMC Article](#)

Results by year Download CSV

Titles with your search terms Exploring the molecular pathogenesis and biomarkers of high risk [Eur J Cancer Prev. 2017]

Serial analysis of gene expression in adenocarcinoma samples [Genet Mol Res. 2015]

Screening of crucial long non-coding RNAs in oral epithelial dysplasia by [Genet Mol Res. 2015]

See more...

Find related data Database: Select Find Items

# PubMed search >> "serial analysis of gene expression"

NCBI Resources How To Sign in to NCBI Help

PubMed serial analysis of gene expression Search

Create RSS Create alert Advanced

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#) [Show additional filters](#)

Format: Summary Sort by: Most Recent Per page: 20 Send to Filters: [Manage Filters](#)

Sort by: Best match Most recent

Search results Items: 1 to 20 of 3205 << First < Prev Page 1 of 161 Next > Last >>

[Endogenous Selenoprotein P, a Liver-Derived Secretory Protein, Mediates Myocardial Ischemia/Reperfusion Injury in Mice.](#)  
1. Chadani H, Usui S, Inoue O, Kusayama T, Takashima SI, Kato T, Murai H, Furusho H, Nomura A, Misu H, Takamura T, Kaneko S, Takamura M.  
Int J Mol Sci. 2018 Mar 16;19(3). pii: E878. doi: 10.3390/ijms19030878.  
PMID: 29547524 [Similar articles](#)

[Homologous recombination mediates stable Fah gene integration and phenotypic correction in tyrosinaemia mouse-model.](#)  
2. Junge N, Yuan Q, Vu TH, Krooss S, Bednarski C, Balakrishnan A, Cathomen T, Manns MP, Baumann U, Sharma AD, Ott M.  
World J Hepatol. 2018 Feb 27;10(2):277-286. doi: 10.4254/wjh.v10.i2.277.  
PMID: 29527263 [Free PMC Article](#) [Similar articles](#)

[Pancreatic tumor microenvironment confers highly malignant properties on pancreatic cancer cells.](#)  
3. Takahashi K, Ehata S, Koinuma D, Morishita Y, Soda M, Mano H, Miyazono K.  
Oncogene. 2018 Mar 7. doi: 10.1038/s41388-018-0144-0. [Epub ahead of print]  
PMID: 29511349 [Similar articles](#)

[Loss of CDX2 gene expression is associated with DNA repair proteins and is a crucial member of the Wnt signaling pathway in liver metastasis of colorectal cancer.](#)  
4. Tóth C, Sükösd F, Valicsek E, Herpel E, Schirmacher P, Tiszlavicz L.  
Oncol Lett. 2018 Mar;15(3):3586-3593. doi: 10.3892/ol.2018.7756. Epub 2018 Jan 9.  
PMID: 29467879 [Free PMC Article](#)

Results by year Download CSV

Titles with your search terms Exploring the molecular pathogenesis and biomarkers of high... [Prev. 2017]

Serial analysis of gene expression in adenocarcinoma samples [Genet Mol Res. 2015]

Screening of crucial long non-coding RNAs in oral epithelial dysplasia by [Genet Mol Res. 2015]

See more...

Find related data Database: Select

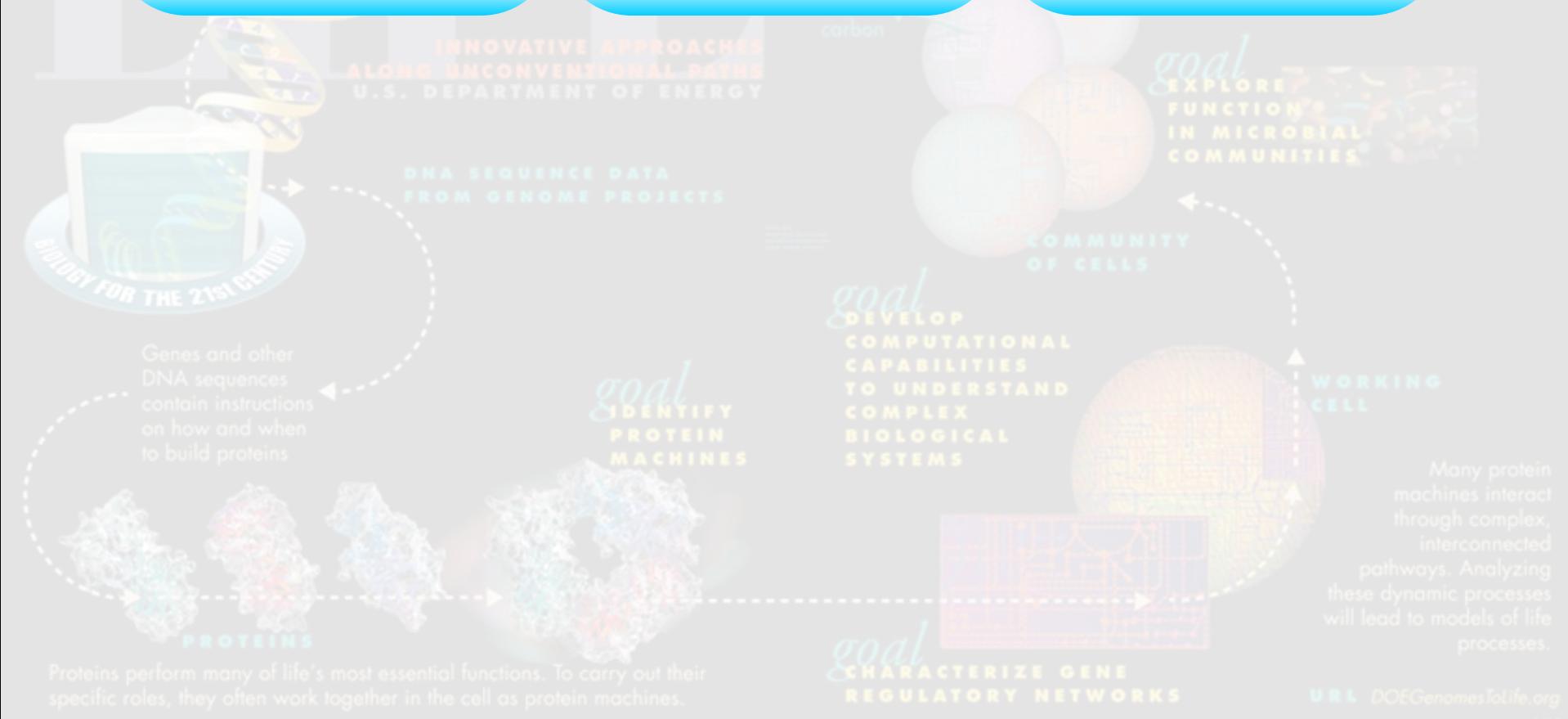
Find Items

# Several developed SAGE variants

LongSAGE (~20bp)

RL-SAGE

SuperSAGE (~26bp)



# Several developed SAGE variants

Protect workers  
and the public

Apply knowledge of  
microbial functional  
capabilities

## LongSAGE (~20bp)

## RL-SAGE

## SuperSAGE (~26bp)

Nat Biotechnol. 2002 May;20(5):508-12.

### Using the transcriptome to annotate the genome.

Saha S<sup>1</sup>, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE.

#### ⊕ Author information

#### Abstract

A remaining challenge for the human genome project involves the identification and annotation of expressed genes. The public and private sequencing efforts have identified approximately 15,000 sequences that meet stringent criteria for genes, such as correspondence with known genes from humans or other species, and have made another approximately 10,000-20,000 gene predictions of lower confidence, supported by various types of *in silico* evidence, including homology studies, domain searches, and *ab initio* gene predictions. These computational methods have limitations, both because they are unable to identify a significant fraction of genes and exons and because they are unable to provide definitive evidence about whether a hypothetical gene is actually expressed. As the *in silico* approaches identified a smaller number of genes than anticipated, we wondered whether high-throughput experimental analyses could be used to provide evidence for the expression of hypothetical genes and to reveal previously undiscovered genes. We describe here the development of such a method--called long serial analysis of gene expression (LongSAGE), an adaption of the original SAGE approach--that can be used to rapidly identify novel genes and exons.

PMID: 11981567 [PubMed - indexed for MEDLINE]

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)



# Several developed SAGE variants

Protect workers  
and the public

Apply knowledge of  
microbial functional  
capabilities

## LongSAGE (~20bp)

## RL-SAGE

## SuperSAGE (~26bp)

Plant Physiol. 2004 Mar; 134(3):890-7.

### **Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis.**

Gowda M<sup>1</sup>, Jantasuriyarat C, Dean RA, Wang GL.

#### **Author information**

<sup>1</sup>Department of Plant Pathology, Ohio State University, Columbus, Ohio 43210, USA.

#### **Abstract**

Serial analysis of gene expression (SAGE) is a widely used technique for large-scale transcriptome analysis in mammalian systems. Recently, a modified version called LongSAGE (S. Saha, A.B. Sparks, C. Rago, V. Akmaev, C.J. Wang, B. Vogelstein, K.W. Kinzler [2002] *Nat Biotechnol* 20: 508-512) was reported by increasing tag length up to 21 bp. Although the procedures for these two methods are similar, a detailed protocol for LongSAGE library construction has not been reported yet, and several technical difficulties associated with concatemer cloning and purification have not been solved. In this study, we report a substantially improved LongSAGE method called Robust-LongSAGE, which has four major improvements when compared with the previously reported protocols. First, a small amount of mRNA (50 ng) was enough for a library construction. Second, enhancement of cDNA adapter and ditag formation was achieved through an extended ligation period (overnight). Third, only 20 ditag polymerase chain reactions were needed to obtain a complete library (up to 90% reduction compared with the original protocols). Fourth, concatemers were partially digested with NlaIII before cloning into vector (pZErO-1), greatly improving cloning efficiency. The significant contribution of Robust-LongSAGE is that it solved the major technical difficulties, such as low cloning efficiency and small insert sizes associated with existing SAGE and LongSAGE protocols. Using this protocol, one can generate two to three libraries, each containing over 4.5 million tags, within a month. We recently have constructed five libraries from rice (*Oryza sativa*), one from maize (*Zea mays*), and one from the rice blast fungus (*Magnaporthe grisea*).

PMID: 15020752 [PubMed - indexed for MEDLINE] PMCID: PMC389912 [Free PMC Article](#)

Many protein machines interact through complex interconnected ways. Analyzing dynamic processes to models of life processes.

EGenomesToLife.org



# Several developed SAGE variants

Protect workers  
and the public

Apply knowledge of  
microbial functional  
capabilities

LongSAGE (~20bp)

RL-SAGE

SuperSAGE (~26bp)

Cell Microbiol. 2005 Jan;7(1):11-8.

## SuperSAGE.

Matsumura H<sup>1</sup>, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Krüger DH, Terauchi R.

### ⊕ Author information

#### Abstract

The application of transcriptomics to study host-pathogen interactions has already brought important insights into the mechanisms of pathogenesis, and is expanding further keeping pace with the accumulation of genomic sequences of host organisms (human and economically important organisms such as food crops) and their pathogens (viruses, bacteria, fungi and protozoa). In this review, we introduce SuperSAGE, a substantially improved variant of serial analysis of gene expression (SAGE), as a potent tool for the transcriptomics of host-pathogen interactions. Notably, the generation of 26 bp tags in the SuperSAGE procedure allows to decipher the 'interaction transcriptome', i.e. the simultaneous monitoring of quantitative gene expression, of both a host and one of its eukaryotic pathogens. The potential of SuperSAGE tags for a rapid functional analysis of target genes is also discussed.

PMID: 15617519 [PubMed - indexed for MEDLINE]

#### PROTEINS

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal  
CHARACTERIZE GENE  
REGULATORY NETWORKS

pathways. Analyzing  
these dynamic processes  
will lead to models of  
life processes.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)



# SAGE分析案例

Mol Biol Rep. 2012 Oct 12.

## Novel differential transcript expression identified by LongSAGE in the mouse endometrium during the implantation window.

Ding YB, He JL, Chen XM, Liu XQ, Wang YX.

### Source

Department of Public Health, Chongqing Medical University, No. 1 Yixueyuan Rd, Box 197, Chongqing, 400016, People's Republic of China, dingyb@gmail.com.

### Abstract

Full development of a receptive uterus is necessary for embryo implantation; however, many genes that are required for the endometrial modifications that occur during this process remain unidentified. To identify novel genes that control endometrial modifications during this period, we investigated the differential gene expression profile in the endometrium of mice on days 2 (D2) (pre-implantation) and 4 (D4) of pregnancy (i.e., the implantation window) using 17-bp long serial analysis of gene expression (LongSAGE).

One hundred fifty-six tags were annotated as unique transcripts. Of these, 101 tags were significantly upregulated, and 55 tags were downregulated in the D4 library relative to the D2 library. These differentially expressed genes should therefore be of increased importance in the establishment of uterine receptivity. The differential expressions of certain of the identified genes, namely, Hspa8, Tctp, Sparc, Ifitm1, Ik, serbp1 and Dnmt1, were validated by semi-quantitative RT-PCR and/or immunohistochemistry. Functional grouping analysis classified 86 of the mapped tags into 17 categories, which are closely associated with morphological modifications of the endometrium during pregnancy. Ingenuity pathways analysis revealed that the identified differentially expressed genes fell into six primary networks, which themselves contain numerous factors that are related to key modulators of signaling pathways that are vital for endometrial modifications. These findings will aid in the further understanding of the molecular events that underlie the implantation physiology in mice.

PMID: 23065227



# 转录组研究方法简介 (2)

## >> 高通量技术和方法

- DNA microarray  
(DNA chip or gene chip)**
- Serial analysis of gene expression  
(SAGE)**

**RNA-Seq**

**EST/cDNA library**

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

*goal*  
CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# RNA-seq简介

## Whole Shotgun Sequencing

WTSS

Transcriptome

next-generation sequencing (NGS)

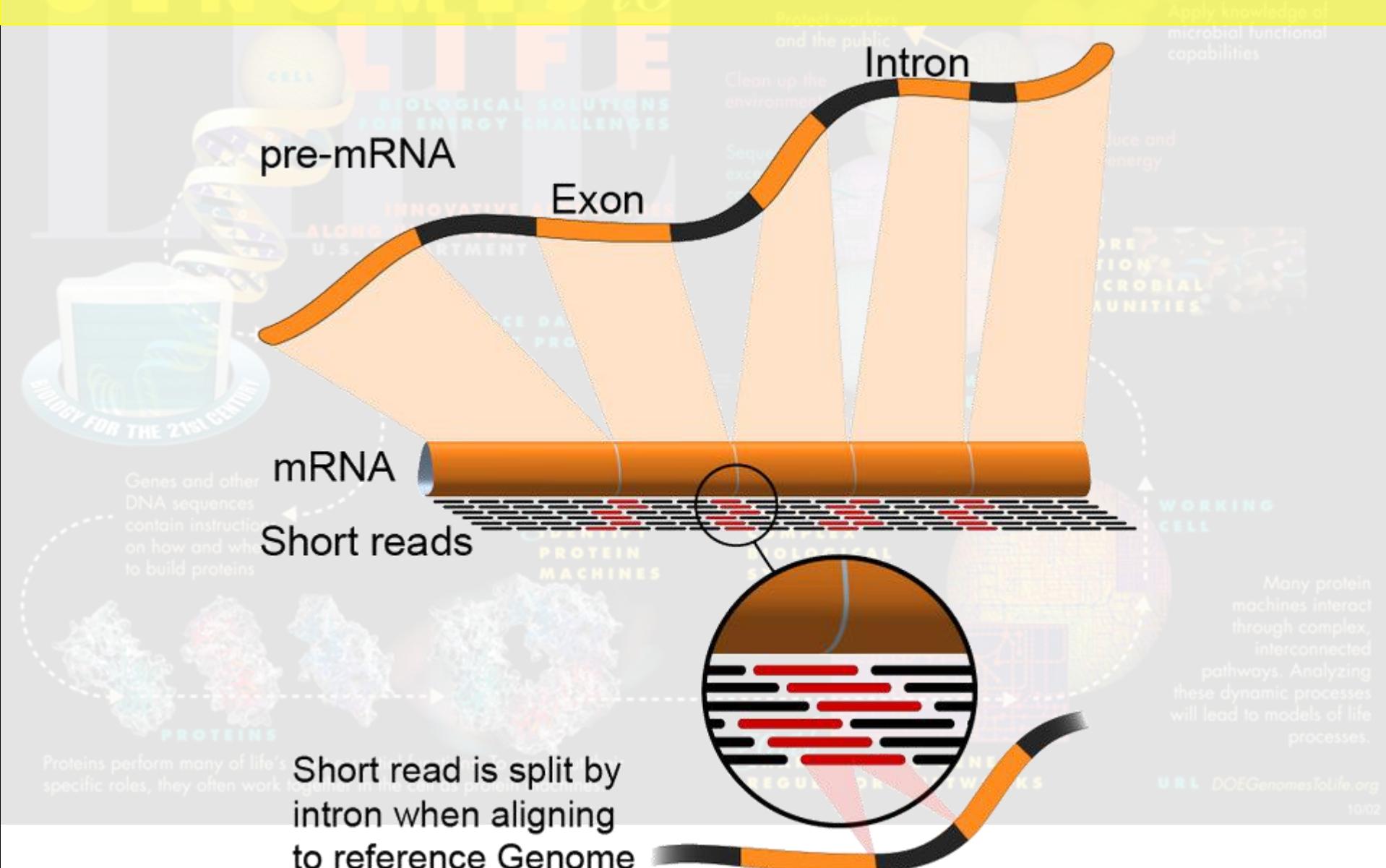
Sanger sequencing

RNA 'Poly(A)' Library

Proteins perform many of life's most essential specific roles; they often work together in the co-



# Transcriptome alignment



# 应用

## Gene expression (compare to Microarray approach)

## Single nucleotide variation (polymorphism) discovery – SNV, SNP

### Germline vs expressed alleles

### Fusion gene detection

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

Apply knowledge of microbial functional capabilities

Produce and use biofuels

FUNCTIONS IN MICROBIAL COMMUNITIES

COMMUNITY OF CELLS

goal  
DEVELOP COMPUTATIONAL CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS

WORKING CELL

Many protein machines interact through complex interconnected pathways. Analyzing these dynamic processes will lead to models of life processes.

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# 应用

## Gene expression (compare to Microarray approach)

## Single nucleotide variation (polymorphism) discovery – SNV, SNP

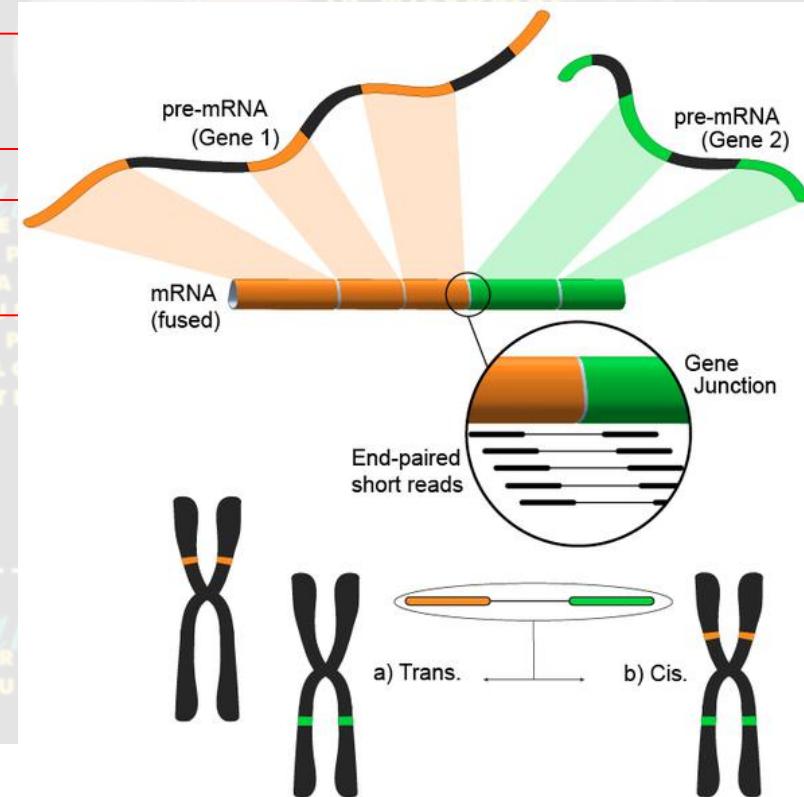
### Germline vs expressed alleles

### Fusion gene detection

DNA sequences contain instructions on how and when to build proteins



Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.



# PubMed search >> RNA-seq[tiab]

NCBI Resources How To

Sign in to NCBI

PubMed RNA-seq[tiab] Search

Create RSS Create alert Advanced Help

Article types Clinical Trial Review Customize ...

Text availability Abstract Free full text Full text

Publication dates 5 years 10 years Custom range...

Species Humans Other Animals

[Clear all](#)

[Show additional filters](#)

Format: Summary Sort by: Most Recent Per page: 20

Send to Filters: [Manage Filters](#)

Sort by: Best match Most recent

Search results Items: 1 to 20 of 12487

<< First < Prev Page 1 of 625 Next > Last >>

[The consistent differential expression of genetic pathways following exposure of an industrial Pseudomonas aeruginosa strain to preservatives and a laundry detergent formulation.](#)  
1. [Pseudomonas aeruginosa strain to preservatives and a laundry detergent formulation.](#)  
Green AE, Amézquita A, Le-Marc Y, Bull MJ, Connor TR, Mahenthiralingam E.  
FEMS Microbiol Lett. 2018 Mar 14. doi: 10.1093/femsle/fny062. [Epub ahead of print]  
PMID: 29548026  
[Similar articles](#)

[Tn5Prime, a Tn5 based 5' capture method for single cell RNA-seq.](#)  
2. [Tn5Prime, a Tn5 based 5' capture method for single cell RNA-seq.](#)  
Cole C, Byrne A, Beaudin AE, Forsberg EC, Vollmers C.  
Nucleic Acids Res. 2018 Mar 14. doi: 10.1093/nar/gky182. [Epub ahead of print]  
PMID: 29548006  
[Similar articles](#)

[DNA Methylome and Transcriptome Alterations and Cancer Prevention by Curcumin in Colitis-accelerated Colon Cancer in Mice.](#)  
3. [DNA Methylome and Transcriptome Alterations and Cancer Prevention by Curcumin in Colitis-accelerated Colon Cancer in Mice.](#)  
Guo Y, Wu R, Gaspar JM, Sargsyan D, Su ZY, Zhang C, Gao L, Cheng D, Li W, Wang C, Yin R, Fang M, Verzi MP, Hart RP, Kong AN.  
Carcinogenesis. 2018 Mar 14. doi: 10.1093/carcin/bgy043. [Epub ahead of print]  
PMID: 29547900  
[Similar articles](#)

[SEASTAR: systematic evaluation of alternative transcription start sites in RNA.](#)  
4. [SEASTAR: systematic evaluation of alternative transcription start sites in RNA.](#)  
Qin Z, Stoilov P, Zhang X, Xing Y.  
Nucleic Acids Res. 2018 Mar 13. doi: 10.1093/nar/gky053. [Epub ahead of print]  
PMID: 29546410

Results by year

Download CSV

Titles with your search terms

Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hyb [Genome Biol. 2018]

A single-cell RNA-seq survey of the developmental landscape of the hu [Nature. 2018]

iSeq: Web-Based RNA-seq Data Analysis and Visualization. [Methods Mol Biol. 2018]

See more...

Find related data

Database: Select

Find Items

苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）

# RNA-seq文献

Nucleic Acids Res. 2018 Mar 14. doi: 10.1093/nar/gky182. [Epub ahead of print]

Apply knowledge of  
microbial functional  
capabilities

## Tn5Prime, a Tn5 based 5' capture method for single cell RNA-seq.

Cole C1, Byrne A2, Beaudin AE3, Forsberg EC1,4, Vollmers C1.

### Author information

- 1 Department of Biomolecular Engineering, University of California Santa Cruz, CA, 95064 USA.
- 2 Department of Molecular, Cellular, Developmental Biology, University of California Santa Cruz, CA, 95064 USA.
- 3 Department of Molecular and Cell Biology, School of Natural Sciences, University of California Merced, CA, 95340 USA.
- 4 Institute for the Biology of Stem Cells, University of California Santa Cruz, CA, 95064 USA.

### Abstract

RNA-sequencing (RNA-seq) is a powerful technique to investigate and quantify entire transcriptomes. Recent advances in the field have made it possible to explore the transcriptomes of single cells. However, most widely used RNA-seq protocols fail to provide crucial information regarding transcription start sites. Here we present a protocol, Tn5Prime, that takes advantage of the Tn5 transposase-based Smart-seq2 protocol to create RNA-seq libraries that capture the 5' end of transcripts. The Tn5Prime method dramatically streamlines the 5' capture process and is both cost effective and reliable. By applying Tn5Prime to bulk RNA and single cell samples, we were able to define transcription start sites as well as quantify transcriptomes at high accuracy and reproducibility. Additionally, similar to 3' end-based high-throughput methods like Drop-seq and 10 $\times$  Genomics Chromium, the 5' capture Tn5Prime method allows the introduction of cellular identifiers during reverse transcription, simplifying the analysis of large numbers of single cells. In contrast to 3' end-based methods, Tn5Prime also enables the assembly of the variable 5' ends of the antibody sequences present in single B-cell data. Therefore, Tn5Prime presents a robust tool for both basic and applied research into the adaptive immune system and beyond.

PMID: 29548006





# 如何访问NCBI的GEO数据库? (Gene Expression Ominus)

Apply knowledge of  
microbiology

Sign in to NCBI

NCBI Resources How To

NCBI

National Center for  
Biotechnology Information

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

All Databases

Books

ClinVar

Clone

Conserved Domains

dbGaP

dbVar

Epigenomics

EST

Gene

Genome

GEO DataSets

GEO Profiles

GSS

HomoloGene

MedGen

MeSH

NCBI Web Site

NLM Catalog

Nucleotide

OMIM

NCBI

National Center for Biotechnology Information advances science and health by  
providing access to biomedical and genomic information.

Home | Mission | Organization | Research | NCBI News

Analyze data using NCBI software

Get NCBI data or software

Learn how to accomplish specific tasks at NCBI

Submit data to GenBank or other NCBI databases

## Genomic Structural Variation

dbVar archives large scale genomic  
variation data and associates defined  
variants with phenotypic information.



1 2 3 4 5 6 7 8

## Popular Resources

PubMed

Bookshelf

PubMed Central

PubMed Health

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

## NCBI Announcements

Virus Variation Resource pages for  
Ebola virus, MERS coronavirus give  
quick and easy access to related  
sequences and other data

Sep 19 2014

Simplified FASTA headers included on  
new NCBI Genomes FTP site

Sep 17, 2014

Last month, a major revision of the NCBI





# 如何访问NCBI的GEO数据库? (Gene Expression Ominus)

Apply knowledge of  
microbiology

Sign in to NCBI

NCBI Resources How To

NCBI

National Center for  
Biotechnology Information

GEO DataSets

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome

The National

Information advances science and health by

Popular Resources

PubMed

Se

Ent

Genome

SNP

Gene

Protein

PubChem

空搜索

直接跳转到所选数据库的主页

- [Tools](#): Analyze data using NCBI software
- [Downloads](#): Get NCBI data or software
- [How Tos](#): Learn how to accomplish specific tasks at NCBI
- [Submissions](#): Submit data to GenBank or other NCBI databases

## Education Resources

Central point of access for helpful documents, teaching materials, news, outlets, and other educational resources.



II 1 2 3 4 5 6 7 8

## NCBI Announcements

Virus Variation Resource pages for Ebolavirus, MERS coronoavirus give quick and easy access to related sequences and other data

Sep 19 2014

Simplified FASTA headers included on new NCBI Genomes FTP site

Sep 17, 2014

Last month, a major revision of the NCBI





# 如何访问NCBI的GEO数据库? => GEO DataSets 主页

Apply knowledge of  
microbiology

NCBI Resources How To

Sign in to NCBI

GEO DataSets

GEO DataSets

Advanced

Search

Help



## GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

### Getting Started

[GEO Documentation](#)

[GEO FAQ](#)

[About GEO DataSets](#)

[Construct a Query](#)

[Download Options](#)

### GEO Tools

[Submit to GEO](#)

[Advanced Search](#)

[DataSet Browser](#)

[Programmatic Access](#)

[GEO2R](#)

### More Resources

[GEO Home](#)

[GEO Profiles](#)

[Epigenomics](#)

[SRA](#)

### Example Searches

Keywords and species

`(smok* OR diet) AND (mammals[organism] NOT human[organism])`

Study type

`"expression profiling by high throughput sequencing"[DataSet Type]`

Studies with CEL files

`cel[Supplementary Files]`

DataSets that have 'age' as an experimental variable

`age[Subset Variable Type]`





# 如何访问NCBI的GEO数据库? => GEO Profiles 主页

Apply knowledge of  
microbiology

Sign in to NCBI

NCBI Resources How To

GEO Profiles

GEO Profiles

Advanced

Search

Help



## GEO Profiles

This database stores individual gene expression profiles from curated DataSets in the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

### Getting Started

[GEO Documentation](#)

[GEO FAQ](#)

[About GEO Profiles](#)

[Construct a Query](#)

[Download Options](#)

### GEO Tools

[Submit to GEO](#)

[Advanced Search](#)

[DataSet Browser](#)

[Programmatic Access](#)

### More Resources

[GEO Home](#)

[GEO DataSets](#)

[Epigenomics](#)

[SRA](#)

### Example Searches

Gene symbol

[CYP1A1\[Gene Symbol\]](#)

Gene symbols in DataSets that contain specific keywords

[\(CYP1A1\[Gene Symbol\] OR ME1\[Gene Symbol\]\) AND \(smok\\* OR diet\)](#)

Partial gene name in a specific DataSet

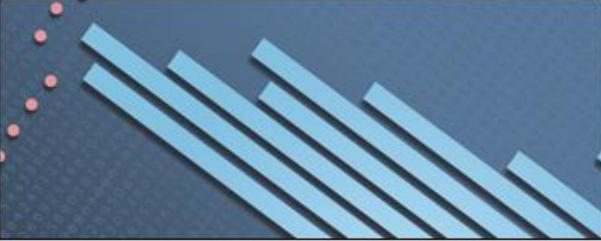
[kinase\[Gene Description\] AND GDS182](#)



# Genbank >> GEO Profiles

NCBI Resources How To Sign in to NCBI Apply knowledge of

GEO Profiles GEO Profiles Advanced Search Help



## GEO Profiles

This database stores individual gene expression profiles from curated DataSets in the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

### Getting Started

- [GEO Documentation](#)
- [GEO FAQ](#)
- [About GEO Profiles](#)
- [Construct a Query](#)
- [Download Options](#)

### GEO Tools

- [Submit to GEO](#)
- [Advanced Search](#)
- [DataSet Browser](#)
- [Programmatic Access](#)

### More Resources

- [GEO Home](#)
- [GEO DataSets](#)
- [Epigenomics](#)
- [SRA](#)

### Example Searches

Gene symbol  
Gene symbols in DataSets  
Partial gene name in a specific DataSet  
Gene Ontology(GO) term in a specific DataSet  
Chromosome region and species  
Genes that show subset effects in DataSets that examine

某个基因      某种疾病      某个生物学过程

CYP1A1[Gene Symbol] AND (CYP1A1[Gene Symbol] AND kinase[Gene Ontology] AND apoptosis[Gene Ontology]) AND GO:00102 [Chromosome] AND 10000:3000000[Base Position]) AND mouse[organism] AND "value subset effect"[Flag Type]

  苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川 (未经同意不得私自转载和发布)

# Genbank >> GEO Datasets

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets Advanced Search Help

## GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

### Getting Started

- [GEO Documentation](#)
- [GEO FAQ](#)
- [About GEO DataSets](#)
- [Construct a Query](#)
- [Download Options](#)

### GEO Tools

- [Submit to GEO](#)
- [Advanced Search](#)
- [DataSet Browser](#)
- [Programmatic Access](#)
- [GEO2R](#)

### More Resources

- [GEO Home](#)
- [GEO Profiles](#)
- [Epigenomics](#)
- [SRA](#)

### Example Searches

Keywords and species  
某个基因 (smok\* OR diet\*) AND (expression AND disease) NOT human[Organism]

Study type  
某种疾病 (expression AND disease)[Data Type]

Studies with CEL files  
某个生物学过程 (cell[Supplement] AND age[Subset Variable Type])

DataSets that have 'age' as an experimental variable  
100:500[Number of Samples]

Studies with between 100 and 500 samples  
smith a[Author]

Author



# Genbank >> GEO Datasets >> Search “leukemia”

Report a problem | Help | Apply knowledge of

NCBI Resources How To Sign in to NCBI

GEO DataSets  Advanced  Help

## GEO DataSets

This database stores curated gene expression DataSets, as well as original Series and Platform records in the Gene Expression Omnibus (GEO) repository. Enter search terms to locate experiments of interest. DataSet records contain additional resources including cluster tools and differential expression queries.

**Getting Started**

- [GEO Documentation](#)
- [GEO FAQ](#)
- [About GEO DataSets](#)
- [Construct a Query](#)
- [Download Options](#)

**GEO Tools**

- [Submit to GEO](#)
- [Advanced Search](#)
- [DataSet Browser](#)
- [Programmatic Access](#)
- [GEO2R](#)

**More Resources**

- [GEO Home](#)
- [GEO Profiles](#)
- [Epigenomics](#)
- [SRA](#)

**Example Searches**

Keywords and species	<a href="#">(smok* OR diet) AND (mammals[organism] NOT human[organism])</a>
Study type	<a href="#">"expression profiling by high throughput sequencing"[DataSet Type]</a>
Studies with CEL files	<a href="#">cell[Supplementary Files]</a>
DataSets that have 'age' as an experimental variable	<a href="#">age[Subset Variable Type]</a>
Studies with between 100 and 500 samples	<a href="#">100:500[Number of Samples]</a>
Author	<a href="#">smith a[Author]</a>



苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）

# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

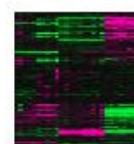
Study type Expression profiling by array Methylation profiling by array More ...

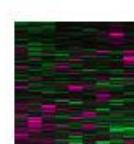
Author Select ...

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters

**Entry type**  
DataSets (175)  
Series (2639)  
Samples (46535)  
Platforms (30)

**Organism**  
Select ...

**Study type**  
Expression profiling by array  
Methylation profiling by array  
More ...

**Author**  
Select ...

**Attribute name**  
tissue  
strain  
More ...

**Publication dates**  
30 days  
1 year  
Custom range...

**Display Settings:** Summary, 20 per page, Sorted by Default order

**Results: 1 to 20 of 49379**

<< First < Prev Page 1 of 2469 Next > Last >>

**Send to:** Filters: Manage Filters

**Top Organisms [Tree]**

- Homo sapiens (42106)
- Mus musculus (7155)
- Rattus norvegicus (79)
- Macaca mulatta (70)
- synthetic construct (48)

**Find related data**

Database: Select

Find items

**Search details**

“leukemia”[MeSH Terms] OR leukemia[All Fields]

Search See more...

**Recent activity**

Turn Off Clear

leukemia (49379)

GEO DataSets

**1. Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course**

Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.

Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)

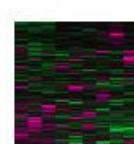
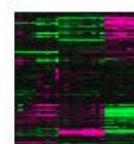
**2. Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells**

Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.

Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)

**3. NOD/SCID/huALL xenotransplant model of pediatric acute leukemia**

Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

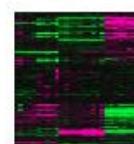
Study type Expression profiling by array Methylation profiling by array More ...

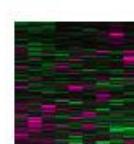
Author Select ...

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

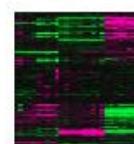
Study type Expression profiling by array Methylation profiling by array More ...

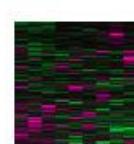
Author Select ...

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

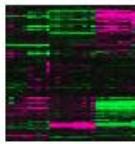
Study type Expression profiling by array Methylation profiling by array More ...

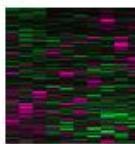
Author Select ... (highlighted with a red border)

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

▼ Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

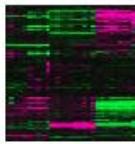
Study type Expression profiling by array Methylation profiling by array More ...

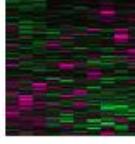
Author Select ...

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

▼ Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” Results

NCBI Resources How To Sign in to NCBI Help

GEO DataSets GEO DataSets leukemia Save search Advanced

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Entry type DataSets (175) Series (2639) Samples (46535) Platforms (30)

Organism Select ...

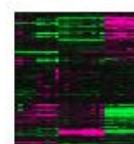
Study type Expression profiling by array Methylation profiling by array More ...

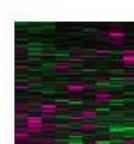
Author Select ...

Attribute name tissue strain More ...

Publication dates 30 days 1 year Custom range...

Results: 1 to 20 of 49379 << First < Prev Page 1 of 2469 Next > Last >>

[Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course](#)  
1. Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.  
Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[Induced myeloproliferative neoplasm effect on Scl-tTA::TRE-BCR/ABL double-transgenic model of chronic myeloid leukemia: osteoblastic lineage cells](#)  
2. Analysis of osteoblastic lineage cells (OBC) from Scl-tTA::TRE-BCR/ABL (BA) double-transgenics withdrawn from doxycycline at 5 weeks of age to induce myeloproliferative neoplasia (MPN). MPN development drives OBC expansion. Results provide insight into molecular features of MPN-expanded OBCs.  
Organism: Mus musculus  
Type: Expression profiling by array, transformed count, 2 genotype/variation sets  
Platform: GPL6246 Series: GSE48438 10 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5024 ID: 5024  
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)  


[NOD/SCID/huALL xenotransplant model of pediatric acute leukemia](#)  
3. Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with

Top Organisms [Tree]  
Homo sapiens (42106)  
Mus musculus (7155)  
Rattus norvegicus (79)  
Macaca mulatta (70)  
synthetic construct (48)  
More...

Find related data Database: Select Find items

Search details "leukemia" [MeSH Terms] OR leukemia [All Fields]

Search See more... Turn Off Clear

Recent activity leukemia (49379)

GEO DataSets



# Search “leukemia” >> Filtered Results

Project navigation  
Help

Apply knowledge of microbial functional

NCBI Resources How To Sign in to NCBI

GEO DataSets GEO DataSets leukemia Save search Advanced Help

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Default order

Send to: Filters: Manage Filters

Results: 1 to 20 of 126

<< First < Prev Page 1 of 7 Next > Last >>

Filters activated: DataSets, human, Expression profiling by array Clear all

Top Organisms [Tree]

Homo sapiens (126)  
Mus musculus (1)

Find related data

Database: Select

Find items

Search details

(“leukemia” [MeSH Terms] OR leukemia[All Fields]) AND (“gds” [Filter] AND “human” [Organism] AND “Expression profiling by array” [Filter])

Search See more...

Recent activity

Turn Off Clear

leukemia AND (“gds”[Filter] AND “human”[Organism] AND “I” GEO DataSets)

leukemia AND (“gds”[Filter] AND “Expression profiling by arr” GEO DataSets)

leukemia AND (“gds”[Filter]) (175)

Entry type DataSets (126) clear

Organism human clear

Study type Expression profiling by array clear

Attribute name tissue

Publication dates Custom range...

Clear all Show additional filters

Acute myeloid leukemia fusion oncogenes in vitro effect on primary CD34+ cells: time course

Analysis of CD34+ cells nucleofected (6h) or transduced (3d, 8d) with AML leukemia fusion oncogenes: AML1-ETO, MLL-AF9, PML-RARA or NUP98-HOXA9. Results provide insight into molecular mechanisms of fusion oncogene-driven leukemogenesis and early temporal patterns of gene deregulation in AML.

Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 10 protocol, 3 time sets  
Platform: GPL570 Series: GSE57194 30 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS5059 ID: 5059  
PubMed Full text in PMC Similar studies GEO Profiles Analyze DataSet

NOD/SCID/huALL xenotransplant model of pediatric acute leukemia

Analysis of xenograft leukemia samples from NOD/SCID mice transplanted with leukemia cells from patients with pediatric BCP-ALL. The xenograft samples displayed short or long “Time To Leukemia” (TTL). Results provide insight into molecular mechanisms underlying the distinct TTL phenotypes.

Organism: Homo sapiens  
Type: Expression profiling by array, transformed count, 2 disease state sets  
Platform: GPL570 Series: GSE13576 12 Samples  
Download data: GEO (CEL)  
DataSet Accession: GDS4779 ID: 4779  
PubMed Similar studies GEO Profiles Analyze DataSet

Bone marrow microenvironment effect on imatinib-treated chronic myeloid leukemia CD34+ cells

SUJU MEDICAL LIBRARY

苏州大学 医学部 基础医学与生物科学学院 生物信息教研室 张高川（未经同意不得私自转载和发布）

# GDS3057



## DATASET BROWSER



Search for GDS3057[ACCN]

### INNOVATIVE APPROACHES

DataSet Record GDS3057: [Expression Profiles](#) [Data Analysis Tools](#) [Sample Subsets](#)

Title: Acute myeloid leukemia

Summary: Comparison of leukemic blasts from 26 acute myeloid leukemia (AML) patients with normal hematopoietic cells at a variety of stages of differentiation from 38 healthy donors. Results provide insight into the molecular mechanisms of AML and the biological significance of those genes with AML-specific expression changes.

Organism: *Homo sapiens*

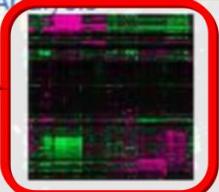
Platform: GPL96: [HG-U133A] Affymetrix

Citation: Stirewalt DL, Meshinchi S, Kornblith PL, et al. Abnormal gene expression changes in acute myeloid leukemia. *Cancer* 2008 Jan;113(1):8-20.

Reference Series: [GSE9476](#)

Value type: transformed count

### Cluster Analysis



**DataSet full SOFT file**  
**DataSet SOFT file**  
**Series family SOFT file**  
**Series family MINiML file**  
**Annotation SOFT file**

[DataSet full SOFT file](#)  
[DataSet SOFT file](#)  
[Series family SOFT file](#)  
[Series family MINiML file](#)  
[Annotation SOFT file](#)

### Data Analysis Tools

Find genes

PROTEINS

Compare 2 sets of samples

Cluster heatmaps

Experiment design and value distribution

Find gene name or symbol:

Find genes that are up/down for this condition(s):  
 disease state  
 cell type



# GDS3057



## DATASET BROWSER



Search for GDS3057[ACCN]

### INNOVATIVE APPROACHES

DataSet Record GDS3057: [Expression Profiles](#) [Data Analysis Tools](#) [Sample Subsets](#)

Title:	Acute myeloid leukemia
Summary:	Comparison of leukemic blasts from 26 acute myeloid leukemia (AML) patients with normal hematopoietic cells at a variety of different stages of maturation from 38 healthy donors. Results provide insight into the possible clinical significance of those genes with AML-specific expression changes.
Organism:	<i>Homo sapiens</i>
Platform:	GPL96: [HG-U133A] Affymetrix Human Genome U133A Array
Citation:	Stirewalt DL, Meshinchi S, Kopecky KJ, Fan W et al. Identification of genes with abnormal expression changes in acute myeloid leukemia. <i>Genes Chromosomes Cancer</i> 2008 Jan;47(1):8-20. PMID: 17910043
Reference Series:	GSE9476
Value type:	transformed count
Sample count:	64
Series published:	2007/11/01



### Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

### Data Analysis Tools

Find genes ?

PROTEINS

Compare 2 sets of samples

Cluster heatmaps

Experiment design and value distribution

Find gene name or symbol:

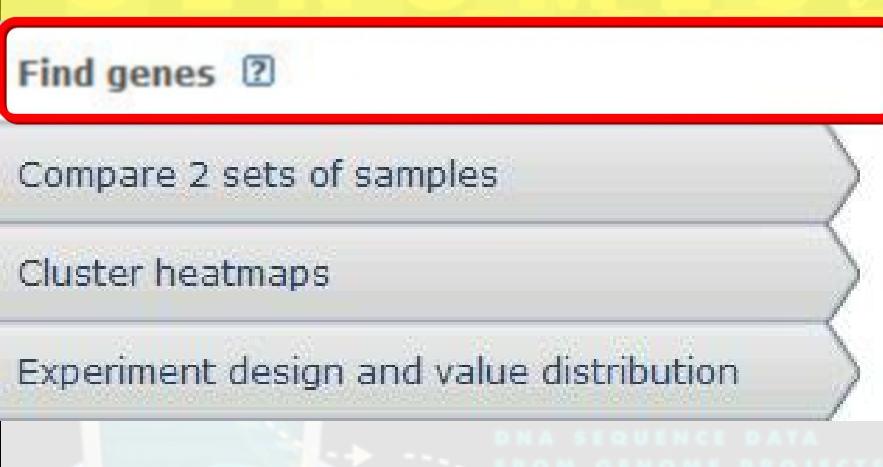
Find genes that are up/down for this condition(s):

disease state

cell type



# Data Analysis Tools >> Find Genes



Find gene name or symbol:

Go

Find genes that are up/down  
for this condition(s):



disease state



cell type

Go



# Find Genes >> Results

NCBI Resources How To

My NCBI Sign In

GEO Profiles

GEO Profiles

("disease state"[FINF] OR "cell type"[FINF]) AND GDS3057[ACCN]

Search

Save search Limits Advanced

Help

Display Settings:  Summary, 20 per page, Sorted by Subgroup effect

EXCESS  
carbon

Send to:

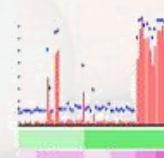
Filters: Manage Filters

Results: 1 to 20 of 96

Page 1 of 5

CRHBP - Acute myeloid leukemia

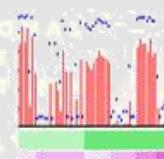
1. Annotation: CRHBP, corticotropin releasing hormone binding protein  
Organism: Homo sapiens  
Reporter: GPL96, 205984\_at (ID\_REF), GDS3057, NM\_001882  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47487511  
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Download profile data

ELANE - Acute myeloid leukemia

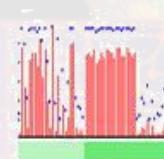
2. Annotation: ELANE, elastase, neutrophil expressed  
Organism: Homo sapiens  
Reporter: GPL96, 206871\_at (ID\_REF), GDS3057, NM\_001972  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47488397  
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Find pathways

VCAN - Acute myeloid leukemia

3. Annotation: VCAN, versican  
Organism: Homo sapiens  
Reporter: GPL96, 215646\_s\_at (ID\_REF), GDS3057, R94644  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47497019  
[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Find items

FGL2 - Acute myeloid leukemia

4. Annotation: FGL2, fibrinogen-like 2  
Organism: Homo sapiens



Search

Many protein machines interact

Search details

("disease state"[FINF]  
OR "cell type"[FINF]) AND  
GDS3057[ACCN]

See more...



# Find Genes >> Results

NCBI Resources How To

My NCBI Sign In

GEO Profiles

GEO Profiles

("disease state"[FINF] OR "cell type"[FINF]) AND GDS3057[ACCN]

Search

Save search Limits Advanced

Help

Display Settings:  Summary, 20 per page, Sorted by Subgroup effect

EXCESS  
carbon

Send to:

Filters: Manage Filters

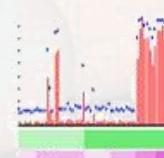
Results: 1 to 20 of 96

Page 1 of 5

CRHBP - Acute myeloid leukemia

- Annotation: CRHBP, corticotropin releasing hormone binding protein  
Organism: Homo sapiens  
Reporter: GPL96, 205984\_at (ID\_REF), GDS3057, NM\_001882  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47487511

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)

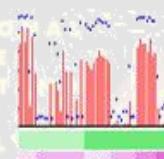


Download profile data

ELANE - Acute myeloid leukemia

- Annotation: ELANE, elastase, neutrophil expressed  
Organism: Homo sapiens  
Reporter: GPL96, 206871\_at (ID\_REF), GDS3057, NM\_001972  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47488397

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)

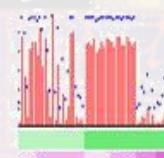


Find pathways

VCAN - Acute myeloid leukemia

- Annotation: VCAN, versican  
Organism: Homo sapiens  
Reporter: GPL96, 215646\_s\_at (ID\_REF), GDS3057, R94644  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47497019

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Find items

FGL2 - Acute myeloid leukemia

- Annotation: FGL2, fibrinogen-like 2  
Organism: Homo sapiens

[GEO DataSets](#) [Gene](#) [UniGene](#) [Profile neighbors](#) [Chromosome neighbors](#) [Sequence neighbors](#) [Homologene neighbors](#)



Search

Many protein machines interact

Search details

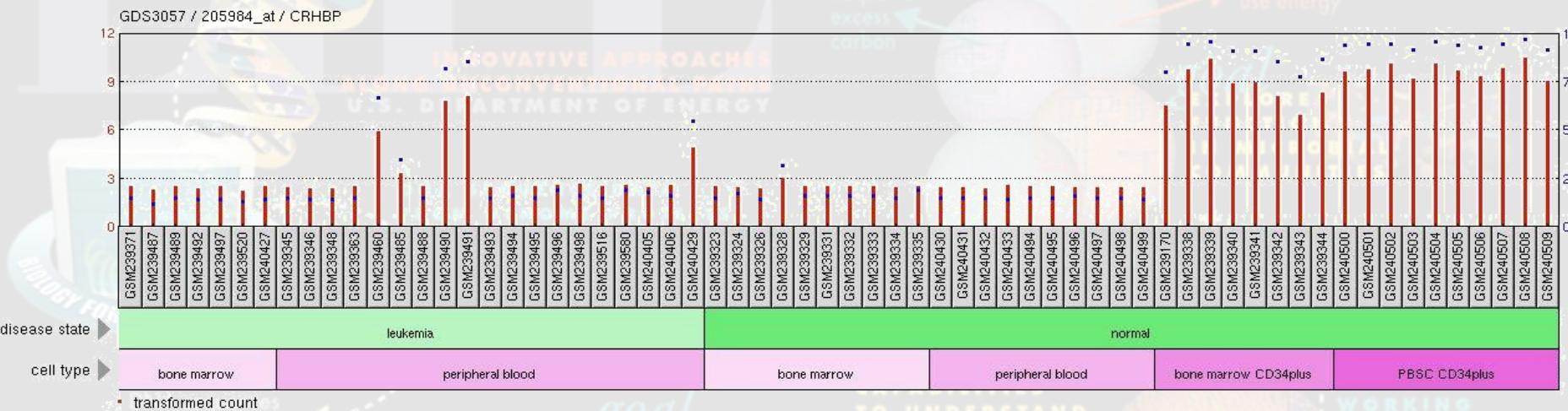
("disease state"[FINF]  
OR "cell type"[FINF]) AND  
GDS3057[ACCN]

See more...



# Find Genes >> Results (2)

Profile GDS3057 / 205984\_at / CRHBP  
 Title Acute myeloid leukemia  
 Organism Homo sapiens



## Red column chart - Value

## Blue dot - Rank

# Find Genes >> Results

Project workspace      Apply knowledge of microarray analysis

NCBI Resources How To My NCBI Sign In

GEO Profiles      GEO Profiles ("disease state"[FINF] OR "cell type"[FINF]) AND GDS3057[ACCN]      Search

Save search      Limits      Advanced      Help

Display Settings:  Summary, 20 per page, Sorted by Subgroup effect

Results: 1 to 20 of 96

CRHBP - Acute myeloid leukemia

1. Annotation: CRHBP, corticotropin releasing hormone receptor  
Organism: Homo sapiens  
Reporter: GPL96, 205984\_at (ID\_REF), GDS3057, NM\_001882  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47487511  
GEO DataSets Gene UniGene Profile neighbors

ELANE - Acute myeloid leukemia

2. Annotation: ELANE, elastase, neutrophil expressed  
Organism: Homo sapiens  
Reporter: GPL96, 206871\_at (ID\_REF), GDS3057, NM\_001972  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47488397  
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

VCAN - Acute myeloid leukemia

3. Annotation: VCAN, versican  
Organism: Homo sapiens  
Reporter: GPL96, 215646\_s\_at (ID\_REF), GDS3057, R94644  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47497019  
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Sequence neighbors Homologene neighbors

FGL2 - Acute myeloid leukemia

4. Annotation: FGL2, fibrinogen-like 2  
Organism: Homo sapiens

Download profile data      Find pathways

Profile data      Filters: Manage Filters

Profile pathways      Find pathways

Find related data      Database: Select

Find items

Search details

("disease state"[FINF]  
OR "cell type"[FINF]) AND  
GDS3057[ACCN]

Search

See more...



# Find Genes >> Find pathways



Promote workers  
and the public

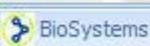
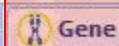
Apply knowledge of  
microbial functional  
capabilities

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains PubChem BioSystems

## FLink - Frequency weighted Links

ABOUT SEARCH HOW TO HELP FAQ NEWS PUBLICATIONS DISCOVER

Links from geoprofiles records to biosystems records weighted by frequency (click to see details)



Clear Selections

Show

Download CSV

Summary

?

<input type="checkbox"/> Frequency	Gene ID	Gene Symbol	Gene Name
5 Genes and other DNA sequences contain instructions on how and when to build proteins	1462	VCAN	versican
2	55353	LAPTM4B	lysosomal protein transmembrane 4 beta
2	5265	SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
2	4353	MPO	myeloperoxidase
2	3492	IGH@	immunoglobulin heavy locus
2	2993	GYPAA	glycophorin A (MNS blood group)
2	2273	FHL1	four and a half LIM domains 1
2	948	CD36	CD36 molecule (thrombospondin receptor)
1 Genes and other DNA sequences contain instructions on how and when to build proteins	100126583	LOC100126583	hypothetical LOC100126583
1	729230	CCR2	chemokine (C-C motif) receptor 2



# Find Genes >> Find pathways (2)



Protect workers  
and the public

Apply knowledge of  
microbial functional  
capabilities

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains PubChem BioSystems

## FLink - Frequency weighted Links

ABOUT SEARCH HOW TO HELP FAQ NEWS PUBLICATIONS DISCOVER

Links from geoprofiles records to biosystems records weighted by frequency (click to see details)

<input type="checkbox"/> Gene	<input checked="" type="checkbox"/> BioSystems
<input type="checkbox"/>	<a href="#">Clear Selections</a>
<input type="checkbox"/>	<a href="#">Show</a>
<input type="checkbox"/>	<a href="#">Download CSV</a>
<input type="checkbox"/>	<a href="#">Summary</a>
Frequency	BSID
18 Genes and other DNA sequences	106386
10 contain instructions on how and when to build proteins	366160
8	477135
8	106387
7	477114
6	530733
6	523016
6	522987
Source	Name
REACTOME	Immune System
REACTOME	Adaptive Immune System
REACTOME	Metabolism
REACTOME	Innate Immune System
REACTOME	Signal Transduction
REACTOME	Cell Cycle
KEGG	Transcriptional misregulation in cancer
KEGG	Transcriptional misregulation in cancer
Type	Organism
organism-specific biosystem	Escherichia coli, Homo sapiens
organism-specific biosystem	Homo sapiens
organism-specific biosystem	Homo sapiens
organism-specific biosystem	Escherichia coli, Homo sapiens
organism-specific biosystem	Homo sapiens
organism-specific biosystem	Homo sapiens
organism-specific biosystem	Homo sapiens
conserved biosystem	



# Data Analysis Tools

## >> Compare 2 sets of samples



**Step 1: Select test and significance level**

Two-tailed t-test (A vs B) ▾ Significance level: 0.100 ▾

**Step 2: Select which Samples to put in Group A and Group B**

PROTEINS  
Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

goal CAPABILITIES TO UNDERSTAND COMPLEX WORKING

Many protein machines interact through complex interconnected pathways. Analyzing processes: of life processes.

CHARACTERIZE GENE REGULATORY NETWORKS

www.DCGenomesToolkit.org

10/02

### Step 3: Query Group A vs. B



# Data Analysis Tools

## >> Compare 2 sets of samples

Find genes

Compare 2 sets of samples 

Cluster heatmaps

Experiment design and value distribution

Apply knowledge of microbial functional capabilities

Clean up the environment

Sequester excess carbon

Produce and use energy

goal  
EXPLORE FUNCTION IN MICROBIAL COMMUNITIES

### Step 1: Select test and significance level

Two-tailed t-test (A vs B) 

Two-tailed t-test (A vs B)

One-tailed t-test (A > B)

One-tailed t-test (A < B)

Value means difference

Rank means difference

Significance level: 0.100 

Many protein machines interact through complex interconnected pathways. Analyzing processes: of life processes.

es to put in Group A and Group B

goal  
CHARACTERIZE GENE REGULATORY NETWORKS

www.DNAgenomeslotIt.org

10/02



# Data Analysis Tools

## >> Compare 2 sets of samples

Find genes

Compare 2 sets of samples 

Cluster heatmaps

Experiment design and value distribution

Provide public  
and the public

Clean up the  
environment

Sequester  
excess  
carbon

Produce and  
use energy

goal  
EXPLORE  
FUNCTIONS  
IN MICROBIAL  
COMMUNITIES

### Step 1: Select test and significance level

Two-tailed t-test (A vs B) 

Significance level:

0.100 

0.100  
0.050  
0.010

### Step 2: Select which Samples to put in Group A and Group B

PROTEINS  
Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENES  
REGULATORY NETWORKS

http://DCCgenomes.iitit.org

10/02

### Step 3: Query Group A vs. B



# Data Analysis Tools

## >> Compare 2 sets of samples

Find genes

Compare 2 sets of samples 

Cluster heatmaps

Experiment design and value distribution

Provide data  
and the public

Clean up the  
environment

Sequester  
excess  
carbon

Produce and  
use energy

goal  
EXPLORE  
FUNCTIONS  
IN MICROBIAL  
COMMUNITIES

### Step 1: Select test and significance level

Two-tailed t-test (A vs B) 

Significance level: 0.100 

### Step 2: Select which Samples to put in Group A and Group B

### Step 3: Query Group A vs. B



# Data Analysis Tools

## >> Compare 2 sets of samples

Apply knowledge of



Click on accessions to select samples individually, click on colored blocks and then on blinking arrows to select groups of samples.

Samples, Group A	Factors		Samples, Group B
	disease state	cell type	
GSM239371			GSM239371
GSM239487			GSM239487
GSM239489			GSM239489
GSM239492			GSM239492
GSM239497			GSM239497
GSM239520			GSM239520
GSM240427			GSM240427
GSM239345			GSM239345
GSM239346			GSM239346
GSM239348			GSM239348
GSM239363			GSM239363
GSM239460			GSM239460
GSM239485			GSM239485
GSM239488			GSM239488
GSM239490			GSM239490
GSM239491			GSM239491
GSM239493			GSM239493
leukemia		bone marrow	
		peripheral blood	

Ok

Reset

Cancel

alysis Tools

Sample Subsets

AML) patients  
f maturation  
: clinical  
s.

Cluster Analysis



Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

/01

Find genes

Compare 2 sets of samples [?](#)

Cluster heatmaps

Experiment design and value distribution

Step 1: Select test and significance level

Two-tailed t-test (A vs B)  Significance level: 0.050

Step 2: Select which Samples to put in Group A and Group B

Step 3: Query Group A vs. B

# Data Analysis Tools

## >> Compare 2 sets of samples

Apply knowledge of



Gene Expression Omnibus

Click on accessions to select samples individually, click on colored blocks and then on blinking arrows to select groups of samples.

Ok

Reset

Cancel

alysis Tools

Sample Subsets

Cluster Analysis



Download

DataSet full SOFT file

DataSet SOFT file

Series family SOFT file

Series family MINiML file

Annotation SOFT file

/01

Samples, Group A	Factors		Samples, Group B
	disease state	cell type	
GSM239371			GSM239371
GSM239487			GSM239487
GSM239489			GSM239489
GSM239492		bone marrow	GSM239492
GSM239497			GSM239497
GSM239520			GSM239520
GSM240427			GSM240427
GSM239345			GSM239345
GSM239346			GSM239346
GSM239348			GSM239348
GSM239363			GSM239363
GSM239460			GSM239460
GSM239485	leukemia		GSM239485
GSM239488			GSM239488
GSM239490			GSM239490
GSM239491			GSM239491
GSM239493		peripheral blood	GSM239493

Find genes

Compare 2 sets of samples

Cluster heatmaps

Experiment design and value distribution

Step 1: Select test and significance level

Two-tailed t-test (A vs B)

Significance level: 0.050

Step 2: Select which Samples to put in Group A and Group B

Step 3: Query Group A vs. B

# Data Analysis Tools

## >> Compare 2 sets of samples

Apply knowledge of

### Step 1: Select test and significance level

Two-tailed t-test (A vs B)  Significance level: 0.050

### Step 2: Select which Samples to put in Group A and Group B

**Group A:** GSM239371, GSM239487, GSM239489, GSM239492, GSM239497, GSM239520, GSM240427, GSM239345, GSM239346, GSM239348, GSM239363, GSM239460, GSM239485, GSM239488, GSM239490, GSM239491, GSM239493, GSM239494, GSM239495, GSM239496, GSM239498, GSM239516, GSM239580, GSM240405, GSM240406, GSM240429

**Group B:** GSM239323, GSM239324, GSM239326, GSM239328, GSM239329, GSM239331, GSM239332, GSM239333, GSM239334, GSM239335, GSM240430, GSM240431, GSM240432, GSM240433, GSM240494, GSM240495, GSM240496, GSM240497, GSM240498, GSM240499, GSM239170, GSM239338, GSM239339, GSM239340, GSM239341, GSM239342, GSM239343, GSM239344, GSM240500, GSM240501, GSM240502, GSM240503, GSM240504, GSM240505, GSM240506, GSM240507, GSM240508, GSM240509

### Step 3: Query Group A vs. B



# Data Analysis Tools

>> Compare 2 sets of samples >> Results

NCBI Resources How To

GEO Profiles GEO Profiles Search Limits Advanced Help

Display Settings:  Summary, 20 per page, Sorted by Default order

Results: 1 to 20 of 6630

<< First < Prev Page 1 of 332 Next > Last >>

Acute myeloid leukemia  
1. Annotation: No annotation available  
Organism: Homo sapiens  
Reporter: GPL96, AFFX-ThrX-5\_at (ID\_REF), GDS3057, --Control (SPOT ID)  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47504279

Acute myeloid leukemia  
2. Annotation: No annotation available  
Organism: Homo sapiens  
Reporter: GPL96, AFFX-r2-Bs-thr-5\_s\_at (ID\_REF), GDS3057, --Control (SPOT ID)  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47504261

Acute myeloid leukemia  
3. Annotation: No annotation available  
Organism: Homo sapiens  
Reporter: GPL96, AFFX-r2-Bs-phe-M\_at (ID\_REF), GDS3057, --Control (SPOT ID)  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47504259

STAT1 - Acute myeloid leukemia  
4. Annotation: STAT1, signal transducer and activator of transcription 1, 91kDa  
Organism: Homo sapiens  
Reporter: GPL96, AFFX-HUMISGF3A/M97935\_MB\_at (ID\_REF), GDS3057, --Control (SPOT ID)  
DataSet type: Expression profiling by array, transformed count, 64 samples  
ID: 47504237

Send to:  Filters: [Manage Filters](#)

**Profile data**

[Download profile data](#)

**Profile pathways**

[Find pathways](#)

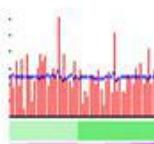
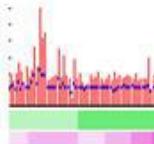
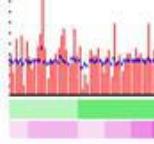
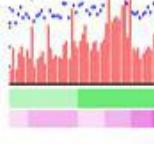
**Find related data**

Database:

[Find items](#)

**Recent activity**

[Turn Off](#) [Clear](#)

Myeloid leukemia cell response to thrombopoietin receptor agonist GDSBrowser

(GDS3606[ACCN]) AND GDS[filter] (1) GDSBrowser

leukemia AND ("human"[Organism] AND "expression profiling by" GEO DataSets)



# Downloaded Profile Data

	A	B	C	D	E	F	G
1	GDS3057						
2	ID_REF	GSM239371	GSM239487	GSM239489	GSM239492	GSM239497	GSM239520
3	disease state	leukemia	leukemia	leukemia	leukemia	leukemia	leukemia
4	cell type	bone marrow					
5	201393_s_at	2.68506	6.92861	6.28011	2.31923	6.44752	7.77314
6	201743_at	2.69449	9.2758	7.24947	2.75118	2.79617	9.74539
7	202833_s_at	6.38448	10.0681	7.64008	3.46602	5.75939	7.06087
8	202870_s_at	4.63858	8.72134	6.64636	5.67736	7.25293	7.224
9	203948_s_at	11.7501	6.95483	9.57061	9.76478	9.8773	9.24497
10	204834_at	3.11119	8.46838	8.78863	2.87765	8.66496	8.71096
11	204961_s_at	2.05988	5.77771	6.48172	3.95743	5.86574	8.18133
12	205237_at	2.31403	8.63986	9.30197	7.31201	6.14012	9.55777
13	205984_at	2.54583	2.3319	2.58222	2.44386	2.58714	2.26815
14	206660_at	8.30563	2.79302	5.63719	9.77477	3.74263	2.15869
15	206871_at	11.1581	12.8085	10.8801	12.7086	4.44054	11.6918
16	207341_at	9.49535	9.0176	6.63088	10.4557	4.11607	8.25436
17	209949_at	4.43437	8.57677	8.96456	4.88518	8.1161	9.29813
18	210895_s_at	5.40815	7.89175	6.26662	4.13634	6.18376	7.84984
19	211571_s_at	3.59926	11.596	7.70835	3.82008	10.9362	11.2489
20	211821_x_at	2.02952	1.88636	2.01706	1.95411	1.99606	1.85076
21	212148_at	2.18282	2.06802	2.18531	2.15039	2.19182	1.95605
22	212372_at	4.78092	3.16697	3.18689	3.89479	3.20336	3.00223
23	215646_s_at	2.75725	11.588	7.4926	3.30949	10.3628	11.0229
24	218454_at	3.41286	7.28606	4.94926	3.38578	4.93365	10.4836



# GEO DataSets >> GSE9476

Scope: Self Format: HTML Amount: Quick GEO accession: GSE9476

Apply knowledge of

## Series GSE9476

Query DataSets for GSE9476

Status	Public on Nov 01, 2007
Title	Abnormal Expression Changes in AML
Organism	<a href="#">Homo sapiens</a>
Experiment type	Expression profiling by array
Summary	Acute myeloid leukemia (AML) is one of the most common and deadly forms of hematopoietic malignancies. We hypothesized that microarray studies could identify previously unrecognized expression changes that only occur only in AML blasts. We were particularly interested in those genes with increased expression in AML, believing that these genes may be potential therapeutic targets. Keywords: AML vs Normal hematopoietic cells
Overall design	We compared gene expression profiles between normal hematopoietic cells from 38 healthy donors and leukemic blasts from 26 AML patients. Normal hematopoietic samples included CD34+ selected cells (N = 18), unselected bone marrows (N = 10), and unselected peripheral bloods (N = 10).
Contributor(s)	<a href="#">Stirewalt DL</a> , <a href="#">Meshinchi S</a> , <a href="#">Kopecky KJ</a> , <a href="#">Fan W</a> , <a href="#">Pogosova-Agadjanyan EL</a> , <a href="#">Engel JH</a> , <a href="#">Cronk MR</a> , <a href="#">Shannon Dorcy K</a> , <a href="#">McQuary AR</a> , <a href="#">Hockenberry D</a> , <a href="#">Wood B</a> , <a href="#">Heimfeld S</a> , <a href="#">Radich JP</a>
Citation(s)	Stirewalt DL, Meshinchi S, Kopecky KJ, Fan W et al. Identification of genes with abnormal expression changes in acute myeloid leukemia. <i>Genes Chromosomes Cancer</i> 2008 Jan;47(1):8-20. PMID: 17910043
Submission date	Oct 30, 2007
Last update date	Jun 08, 2012

# GEO DataSets >> GSE9476

Apply knowledge of

Platforms (1) GPL96 [HG-U133A] Affymetrix Human Genome U133A Array

Samples (64) GSM239170 BM\_CD34\_R000030

+ More...  
GSM239323 BM\_01000  
GSM239324 BM\_02300

## Relations

BioProject 103247

## Analyze with GEO2R

### Download family

SOFT formatted family file(s)

### Format

SOFT [?](#)

MINiML formatted family file(s)

MINiML [?](#)

Series Matrix File(s)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE9476_RAW.tar	215.2 Mb	(ftp)(http) (custom)	TAR (of CEL)

Raw data provided as supplementary file

Processed data included within Sample table

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# GEO DataSets >> GSE9476

Apply knowledge of

Platforms (1) GPL96 [HG-U133A] Affymetrix Human Genome U133A Array

Samples (64) GSM239170 BM\_CD34\_R000030

+ More...  
GSM239323 BM\_01000  
GSM239324 BM\_02300

## Relations

BioProject 103247

## Analyze with GEO2R

### Download family

SOFT formatted family file(s)

### Format

SOFT [?](#)

MJNiML formatted family file(s)

MJNiML [?](#)

Series Matrix File(s)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE9476_RAW.tar	215.2 Mb	(ftp)(http) (custom)	TAR (of CEL)

FTP 目录 /pub/geo/DATA/SeriesMatrix/GSE9476/ 位于 [ftp.ncbi.nlm.nih.gov](ftp://ftp.ncbi.nlm.nih.gov)

[转到高层目录](#)

06/09/2012 02:44下午

7,910,750 [GSE9476\\_series\\_matrix.txt.gz](#)



# GEO DataSets >> GSE9476

Apply knowledge of

Platforms (1) GPL96 [HG-U133A] Affymetrix Human Genome U133A Array

Samples (64) GSM239170 BM\_CD34\_R000030

+ More... GSM239323 BM\_01000

GSM239324 BM\_02300

## Relations

BioProject 103247

## Analyze with GEO2R

### Download family

SOFT formatted family file(s)

### Format

SOFT [?](#)

MINiML formatted family file(s)

MINiML [?](#)

Series Matrix File(s)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE9476_RAW.tar	215.2 Mb	(ftp)(http) (custom)	TAR (of CEL)

Raw data provided as supplementary file

Processed data included within Sample table

Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.

CHARACTERIZE GENE  
REGULATORY NETWORKS

URL: [DOEGenomesToLife.org](http://DOEGenomesToLife.org)

10/02



# GPL96

Apply knowledge of

Scope: Self Format: HTML Amount: Quick GEO accession:

## Platform GPL96

## Query DataSets for GPL96

Status Public on Mar 11, 2002  
Title [HG-U133A] Affymetrix Human Genome U133A Array  
Technology type in situ oligonucleotide  
Distribution commercial  
Organism Homo sapiens  
Manufacturer Affymetrix  
Manufacture protocol see manufacturer's web site

The U133 set includes 2 arrays with a total of 44928 entries and was indexed 29-Jan-2002.  
The set includes over 1,000,000 unique oligonucleotide features covering more than 39,000 transcript variants, which in turn represent greater than 33,000 of the best characterized human genes.  
Sequences were selected from GenBank, dbEST, and RefSeq. Sequence clusters were created from Build 133 of UniGene (April 20, 2001) and refined by analysis and comparison with a number of other publicly available databases including the Washington University EST trace repository and the University of California, Santa Cruz golden-path human genome database (April 2001 release). In addition, ESTs were analyzed for untrimmed low-quality sequence information, correct orientation, false priming, false clustering, alternative splicing and alternative polyadenylation.

Description Affymetrix submissions are typically submitted to GEO using the GEOarchive method described at [http://www.ncbi.nlm.nih.gov/projects/geo/info/geo\\_affy.html](http://www.ncbi.nlm.nih.gov/projects/geo/info/geo_affy.html)  
June 03, 2009: annotation table updated with netaffx build 28  
June 08, 2012: annotation table updated with netaffx build 32

Web link [http://www.affymetrix.com/support/technical/byproduct.affx?  
product=hgu133](http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133)  
<http://www.affymetrix.com/analysis/index.affx>

Submission date Feb 19, 2002  
Last update date Jun 08, 2012  
Organization Affymetrix, Inc.  
E-mail(s) [geo@ncbi.nlm.nih.gov](mailto:geo@ncbi.nlm.nih.gov), [support@affymetrix.com](mailto:support@affymetrix.com)  
Phone 888-362-2447  
URL <http://www.affymetrix.com/index.affx>  
Street address  
City Santa Clara  
State/province CA  
ZIP/Postal code 95051  
Country USA



# GPL96

Apply knowledge of

Samples (31609) [GSM3921](#), [GSM3923](#), [GSM3925](#), [GSM3927](#), [GSM3929](#), [GSM3931](#)

[+ More...](#)

Series (932) [GSE362](#) Normal human muscle - U133 arrays

[+ More...](#)

[GSE409](#) IMR-90 cells: Collagen Mesh Time Course

[GSE410](#) IMR-90 cells: Tissue Culture Polystyrene Time Course

## Relations

Alternative to [GPL4557](#) (Alternative CDF)

Alternative to [GPL7566](#) (Alternative CDF)

Alternative to [GPL9021](#) (Alternative CDF)

Alternative to [GPL9741](#) (Alternative CDF)

Alternative to [GPL10553](#) (Alternative CDF: ncilpg\_U133Agb)

Alternative to [GPL14663](#) (Alternative CDF)

## Data table header descriptions

**ID** Affymetrix Probe Set ID

**GB\_ACC** GenBank Accession Number

**SPOT\_ID** identifies controls

**Species Scientific Name** The genus and species of the organism represented by the probe set.

**Annotation Date** The date that the annotations for this probe array were last updated.  
It will generally be earlier than the date when the annotations were posted on the Affymetrix web site.

**Sequence Type**

**Sequence Source** The database from which the sequence used to design this probe set was taken.

**Target Description**

**Representative Public ID** The accession number of a representative sequence. Note that for consensus-based probe sets, the representative sequence is only one of several sequences (sequence sub-clusters) used to build the consensus sequence and it is not directly used to derive the probe sequences. The representative sequence is chosen during array design as a sequence that is best associated with the transcribed region being interrogated by the probe set. Refer to the "Sequence Source" field to determine the database used.

**Gene Title**

Title of Gene represented by the probe set.



# GPL96

Apply knowledge of

## Data table

ID	GB_ACC	SPOT_ID	Species Scientific Name	Annotation Date	Sequence Type	Sequence Source	Target Description
1007_s_at	U48705		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	U48705 /FEATURE=mRNA /DEFIN
1053_at	M87338		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	M87338 /FEATURE= /DEFINITION
117_at	X51757		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	X51757 /FEATURE=cds /DEFINITI
121_at	X69699		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	X69699 /FEATURE= /DEFINITION
1255_g_at	L36861		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	L36861 /FEATURE=expanded_cds
1294_at	L13852		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	L13852 /FEATURE= /DEFINITION
1316_at	X55005		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	X55005 /FEATURE=mRNA /DEFIN
1320_at	X79510		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	X79510 /FEATURE=cds /DEFINITI
1405_i_at	M21121		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	M21121 /FEATURE= /DEFINITION
1431_at	J02843		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	J02843 /FEATURE=cds /DEFINITI
1438_at	X75208		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	X75208 /FEATURE=cds /DEFINITI
1487_at	L38487		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	L38487 /FEATURE=mRNA /DEFIN
1494_f_at	M33318		Homo sapiens	Jun 9, 2011	Exemplar sequence	Affymetrix Proprietary Database	M33318 /FEATURE=mRNA /DEFIN
1598_g_at	L13720		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	L13720 /FEATURE= /DEFINITION
160020_at	Z48481		Homo sapiens	Jun 9, 2011	Consensus sequence	GenBank	Z48481 /FEATURE=cds /DEFINITI
1729_at	L41690		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	L41690 /FEATURE= /DEFINITION
1773_at	L00635		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	L00635 /FEATURE= /DEFINITION
177_at	U38545		Homo sapiens	Jun 9, 2011	Exemplar sequence	GenBank	U38545 /FEATURE= /DEFINITION



Total number of rows: 22283

Table truncated, full table size 33784 Kbytes.

[Download full table...](#)

[Annotation SOFT table...](#)

### Download family

[SOFT formatted family file\(s\)](#)  
[MTNiMI formatted family file\(s\)](#)

### Format

[SOFT](#)   
[MTNiMI](#)



# 拓展阅读和实践练习

**Transcriptomics, DNA Microarray, SAGE, RNA-seq**  
相关文献资料查阅

生化与分子生物学方法相关文献资料的阅读

搜索并了解更多转录组相关数据库和分析工具

熟悉GEO数据库【Profiles和DataSets】

熟悉GEO数据库中GDS、GSE和GPL的内容

练习GEO数据库中整合的Data Analysis Tools



法養  
天地正氣  
古今完人  
楊永清題



*Thanks for your attention !*

