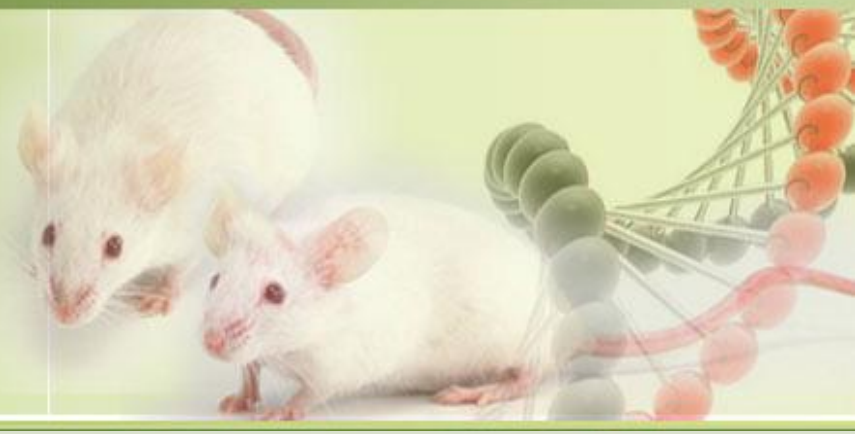
基迪奥生物
GENE DENOVO



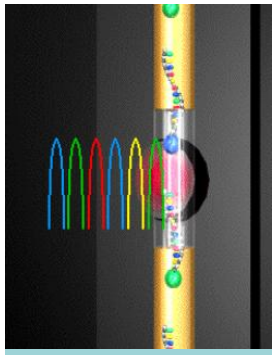
从测序大数据到“小目标”

——项目设计与数据挖掘

周煌凯　2014/3
hkzhou@genedenovo.com

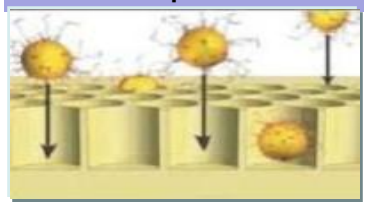
广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com　电话:020-39340225 传真:020-39354111

测序技术的发展

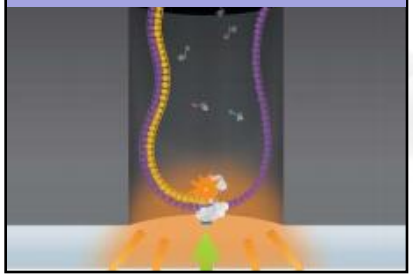


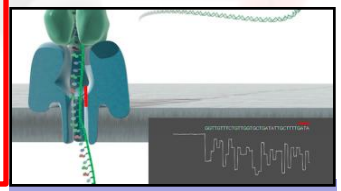
毛细管电泳
(1996)

Heliscope(2008)



Pacific SMRT
(2011)

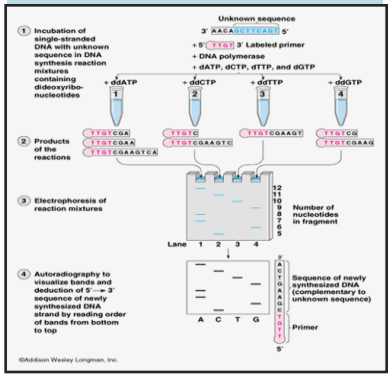




Nanopore(2011)

1950 1960 1970 1980 1990 2000 2010

Sanger技术 (1977)



454 (2005)

Roche

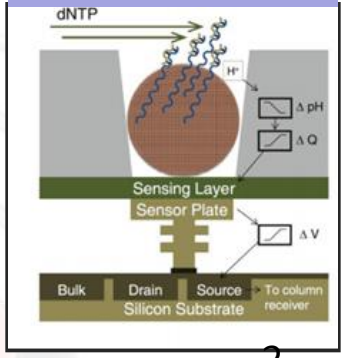
Solexa(2006)

illumina

Solid(2007)

applied biosystems
by life technologies

Iron Torrent(2010)



基迪奥生物
GENE DENOVO

大基因组，大数据

- 基因组：

  哺乳动物基因 3Gb， 2万 编码gene

  玉米基因组 2.3Gb， 5万 编码gene

- 每个基因都有不同的可变剪切
- noncoding RNA（人类 93%）
- 表观修饰
- 所有以上信息，都相互作用，形成调控网络

……

基迪奥生物
GENE DENOVO

高通量技术应用的问题

- 新技术→研究目的（方案）
- 海量数据→**亮点**

实验数据 → **如何挖掘？**
公共数据 → **价值？**
大样本、多组学 → **如何整合、存储？**

核心问题：挖掘实现数据的价值？

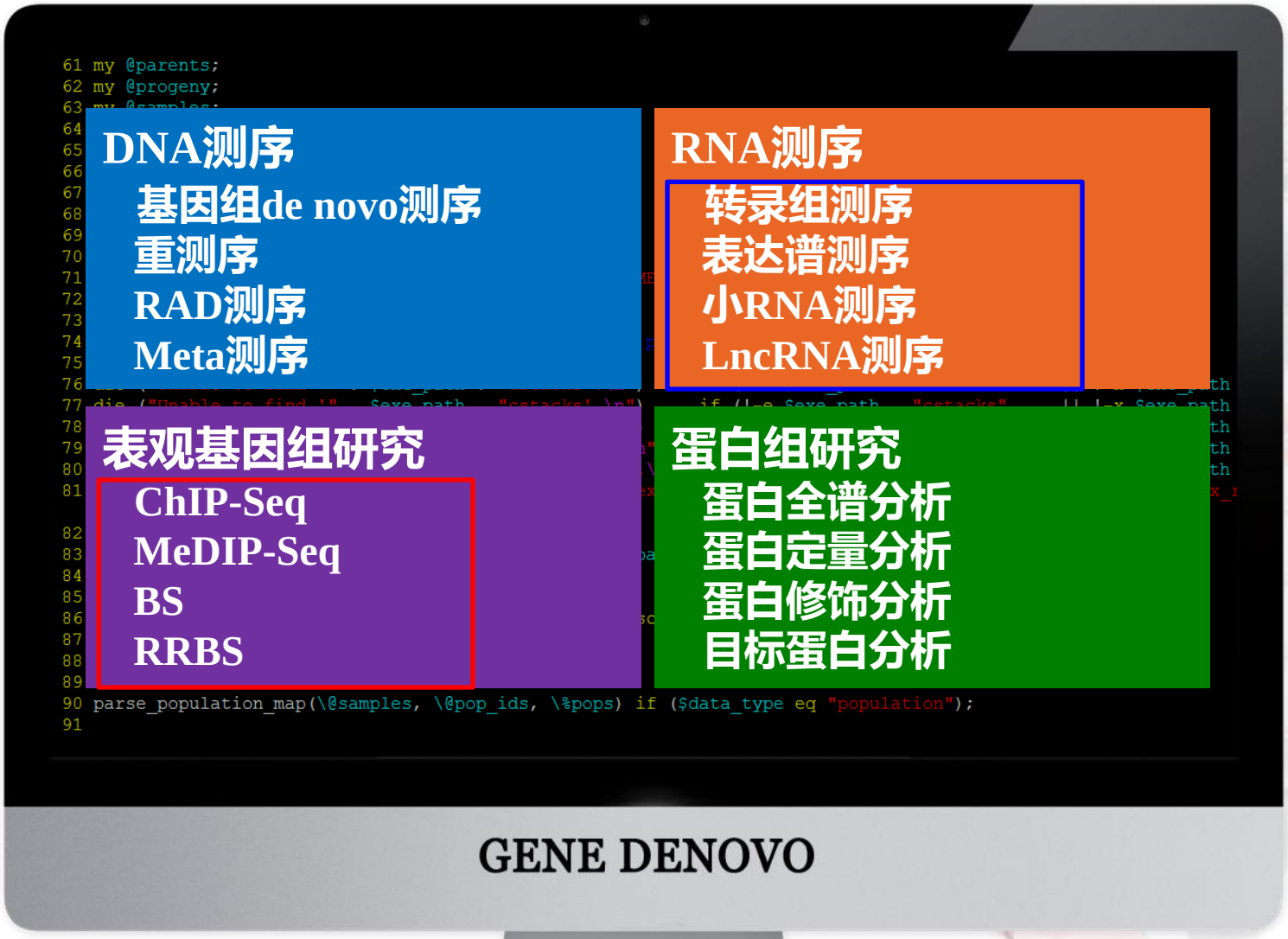
基迪奥生物
GENE DENOVO

提 纲

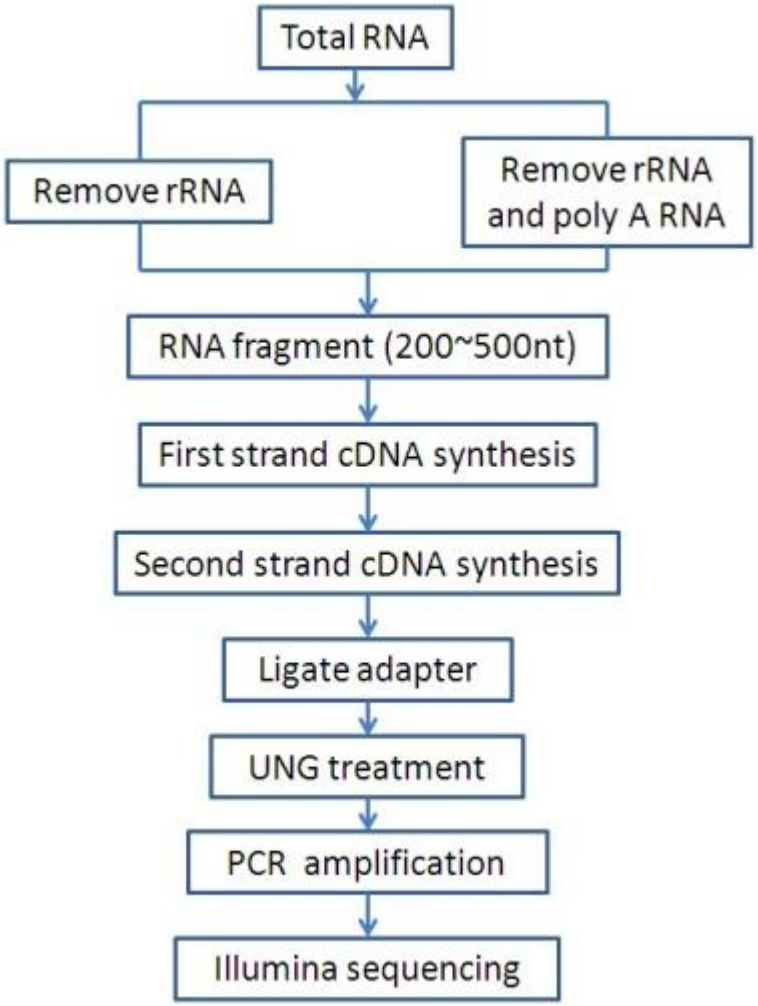
1. **基础——选择适合的NGS技术**
2. 框架——定制化方案设计
3. 亮点——精细挖掘

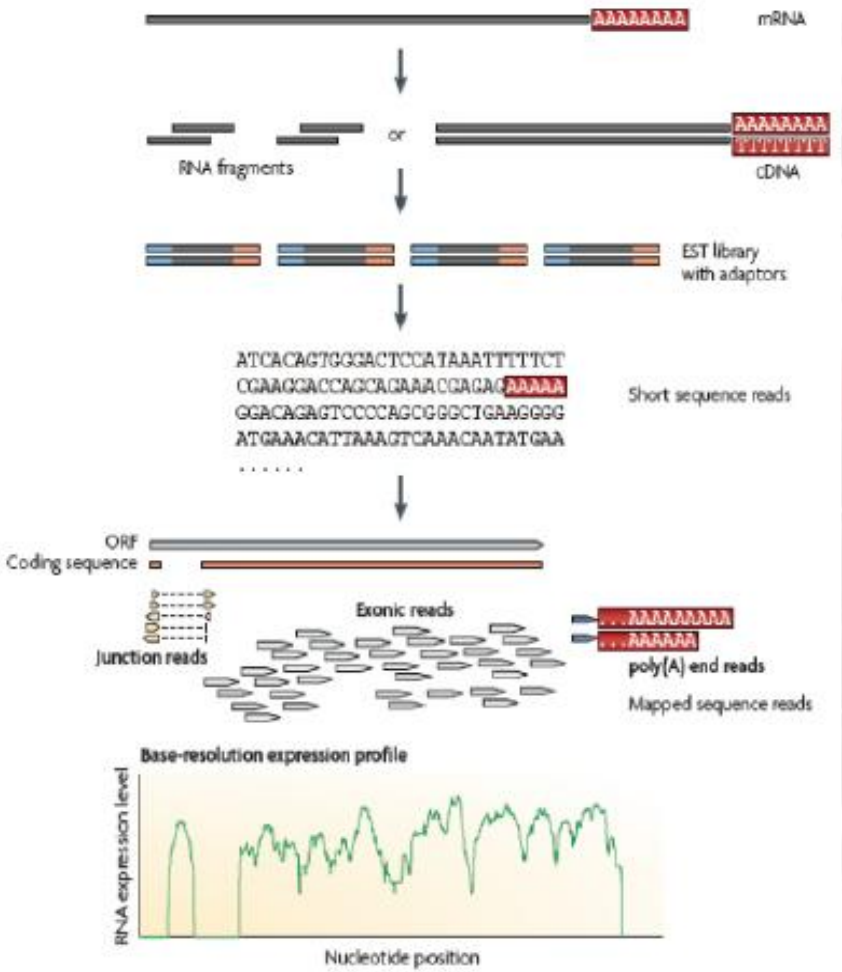
基迪奥生物
GENE DENOVO

不同组学水平的应用



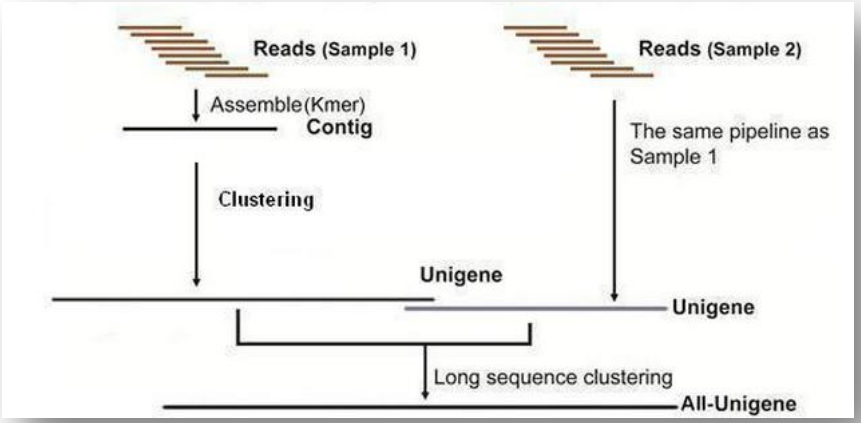
转录组与表达谱建库测序流程



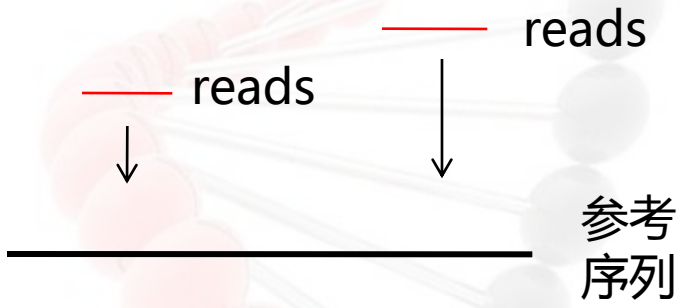


RNA-seq:转录组 or 表达谱

	转录组（transcriptom）	表达谱 (Qualification)
推荐数据量	4~6G	2G
测序读长	100PE	50 SE 或 100PE



Vs.

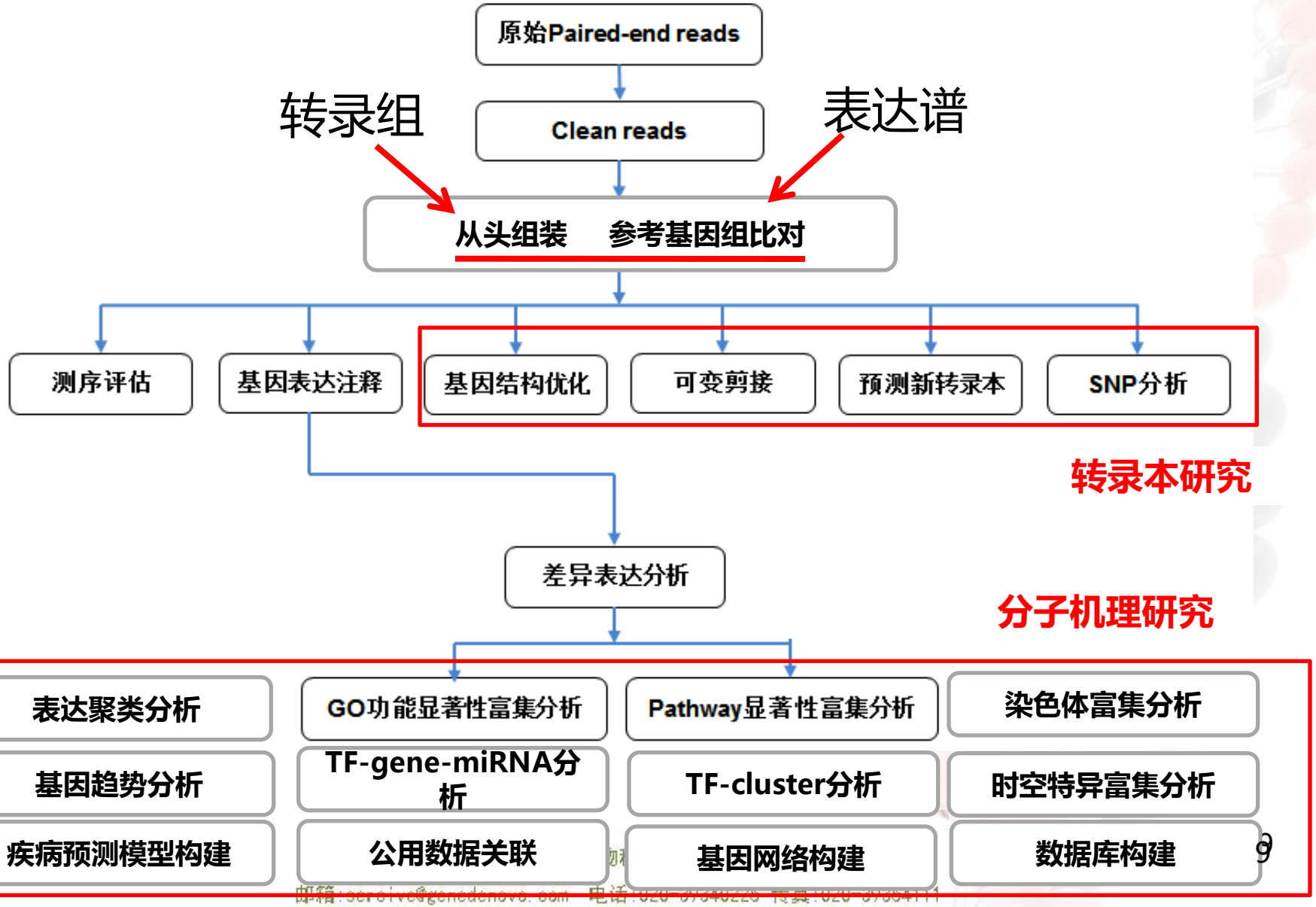


转录组：组装

表达谱：比对

基迪奥生物
GENE DENOVO

RNA-seq分析流程



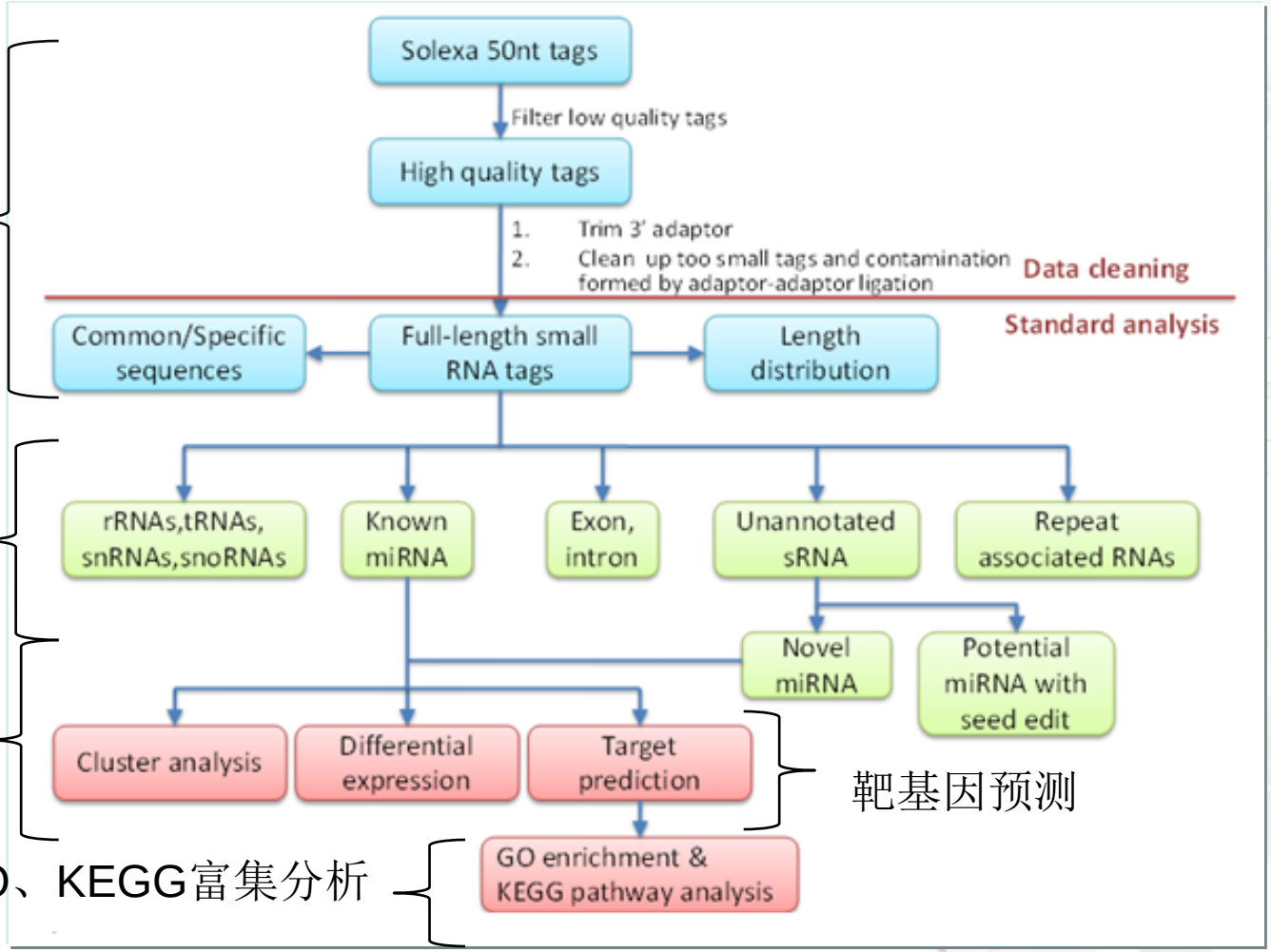
小RNA信息分析内容

测序及序列分类、统计

比对注释

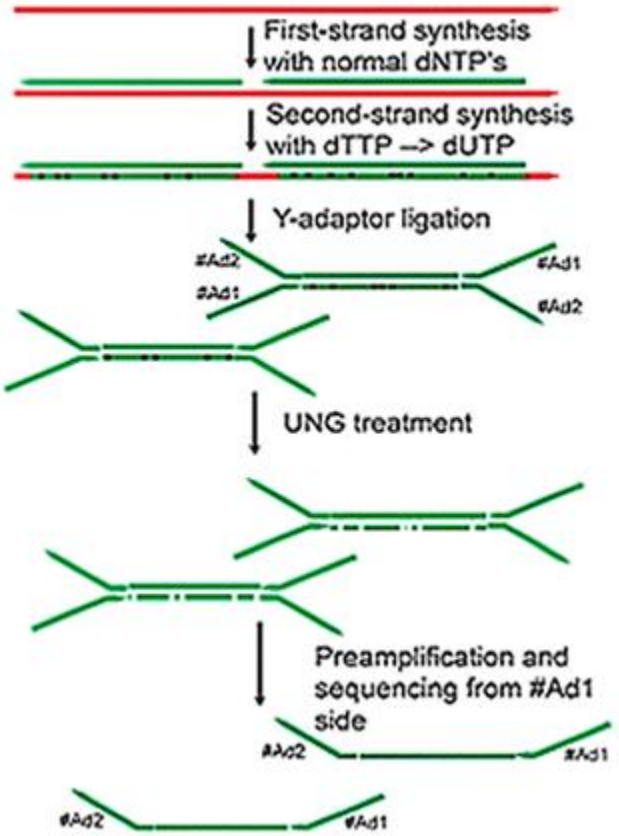
差异分析

GO、KEGG富集分析

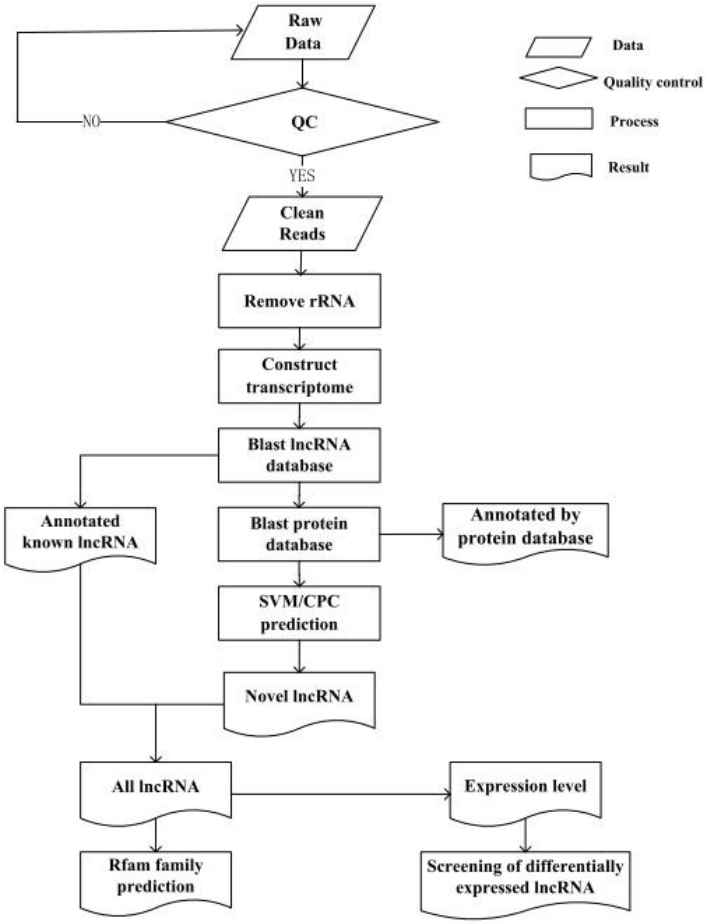


基迪奥生物
GENE DENOVO

Lnc RNA测序与分析

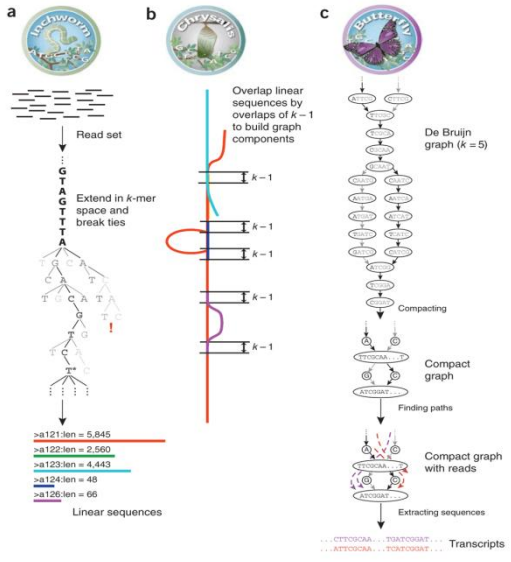


链特异性文库

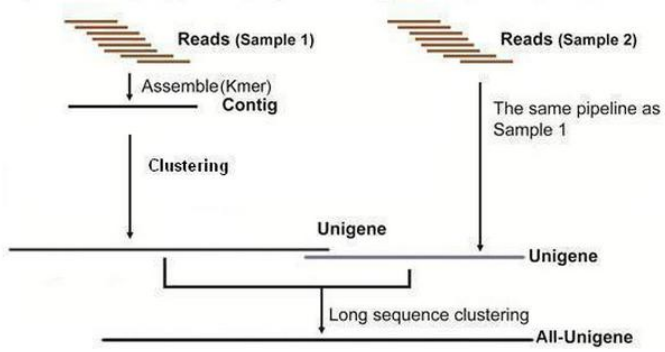


lncRNA分析流程

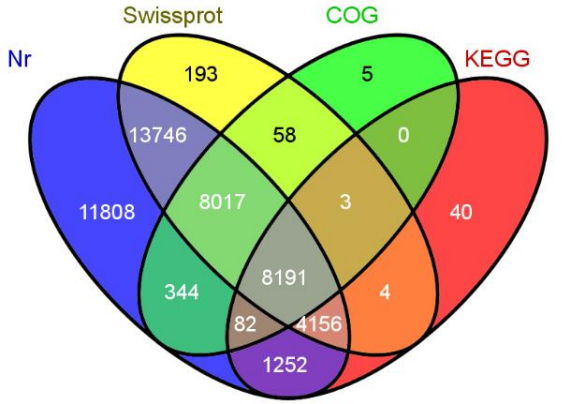
组装与注释



Trinity软件

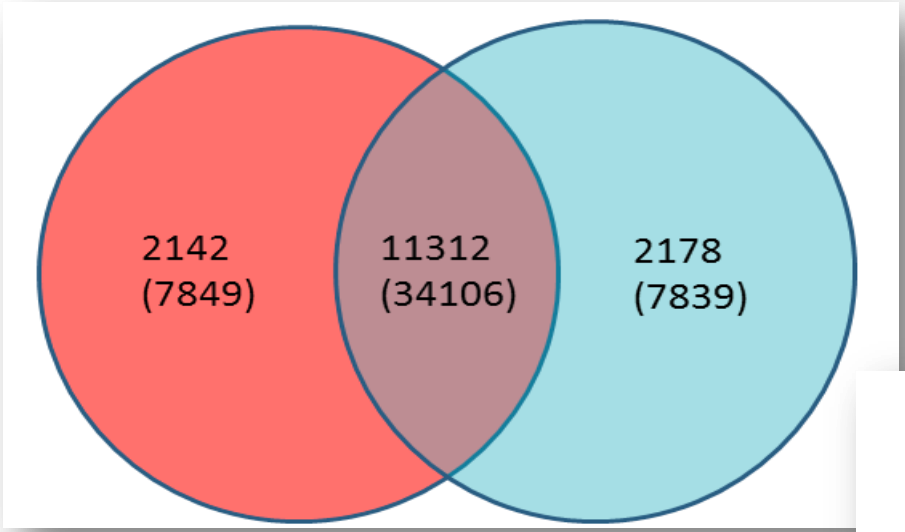


Kmer深度的设置可以改变组装的结果

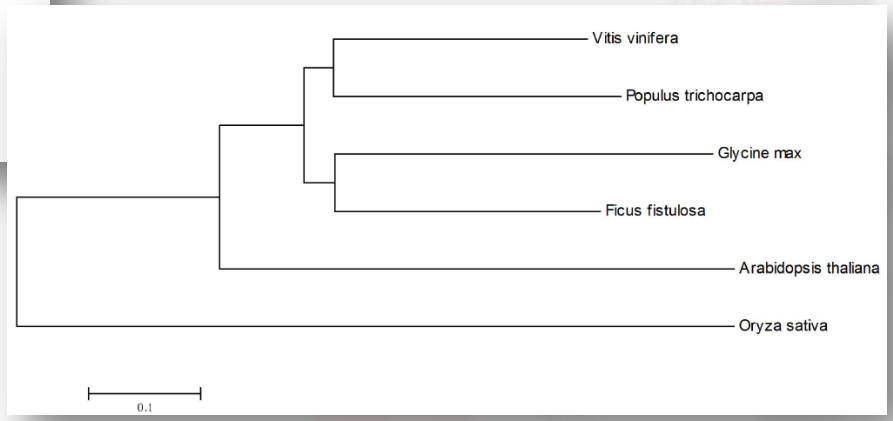


多个数据库注释

蛋白基因家族分析与进化树



跨物种的比较——同源基因与物种特有基因

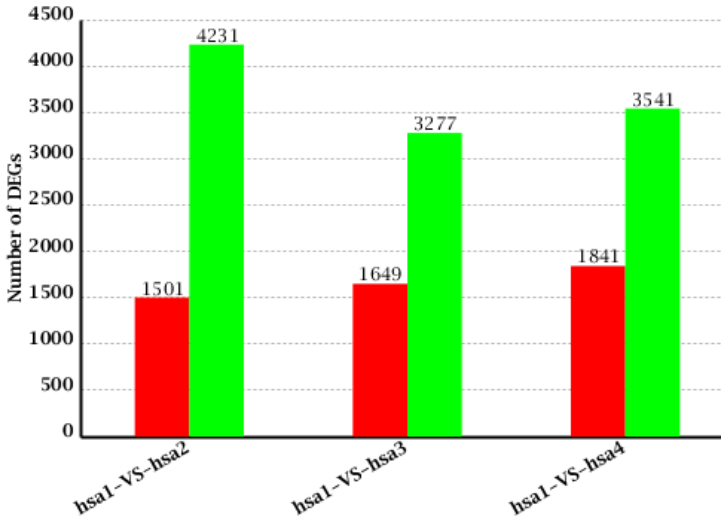


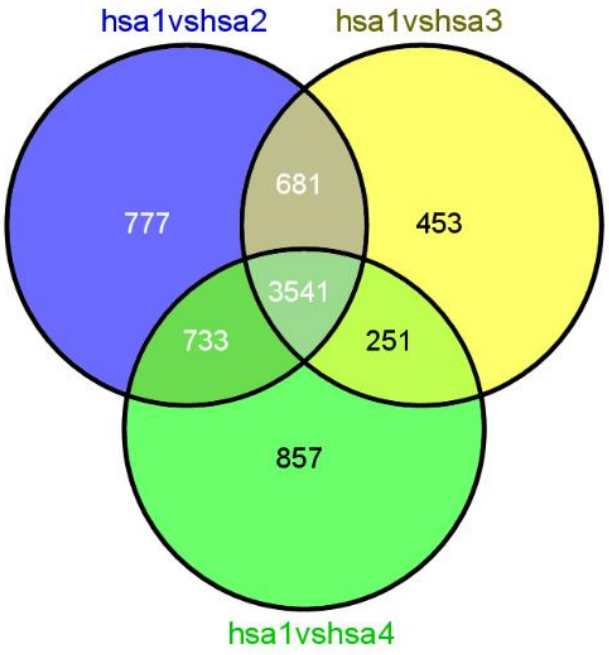
进化树

基迪奥生物
GENE DENOVO

差异比较——高级分析的基础

差异基因统计图



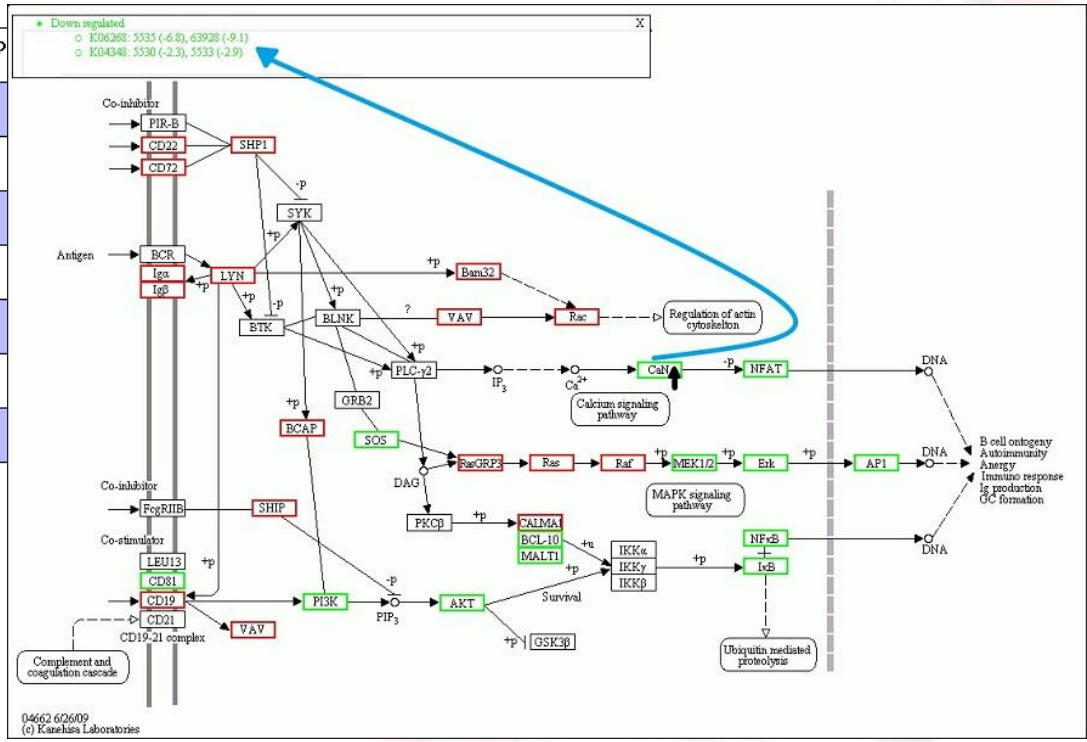


- 差异两倍
- Adjusted P-value < 0.001

基础分类：交集与并集

差异基因分类——KEGG富集分析

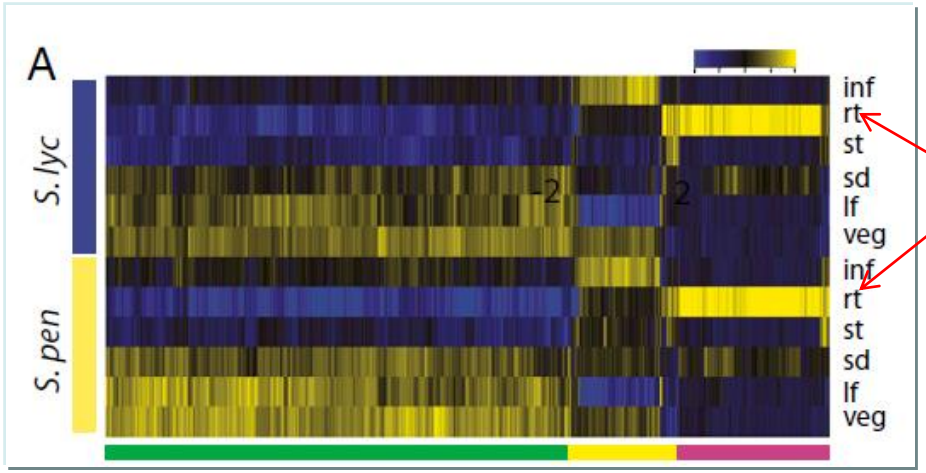
#	Pathway	DEGs with p
1	Systemic lupus erythematosus	34 (1.49%)
2	DNA replication	45 (1.98%)
3	Cell cycle	82 (3.6%)
4	Gap junction	27 (1.19%)
5	Citrate cycle (TCA cycle)	31 (1.36%)
6	Photosynthesis	28 (1.23%)
7	Pathogenic Escherichia coli infection	29 (1.27%)
8	Ribosome	119 (5.23%)



差异表达的基因趋势可能偏向于集中在某些功能或代谢通路

关注点： 基因(零散） → 功能类型（集中）

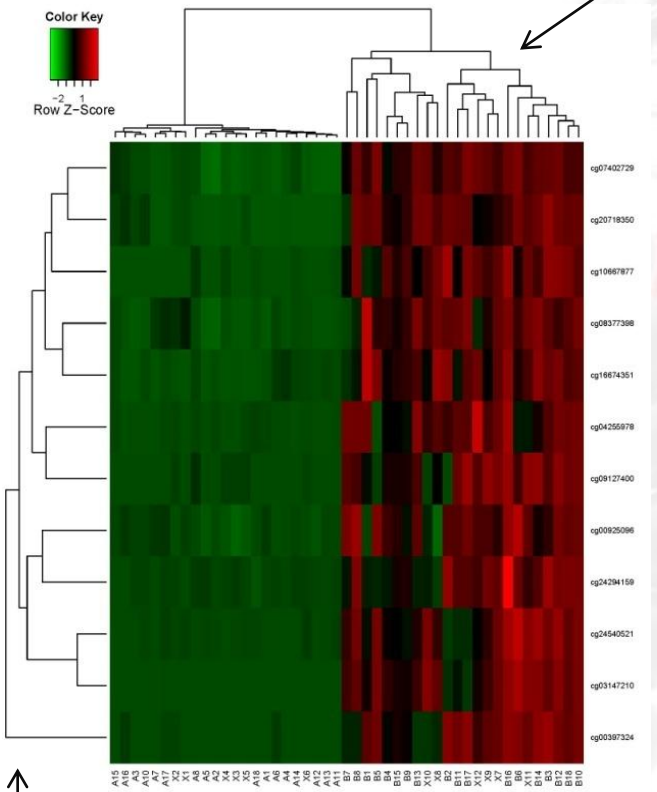
热图分析——展示+聚类



直观展示某类基因的整体表达模式特点

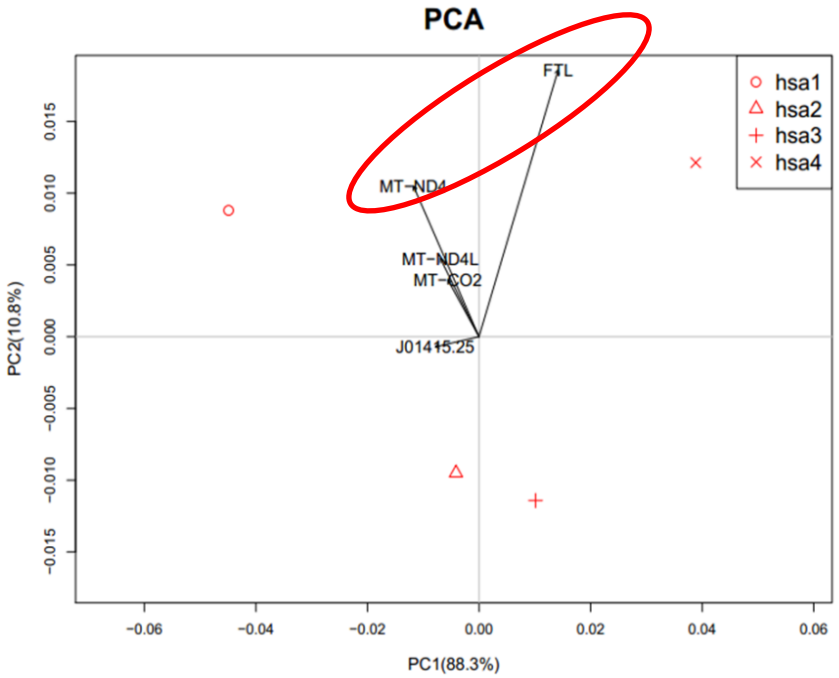
根

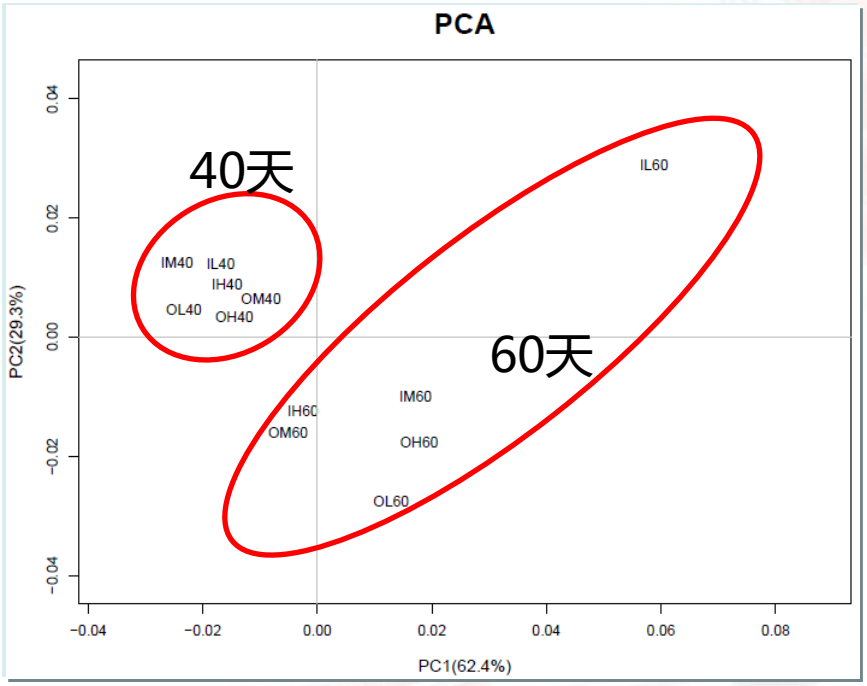
样本关系



基因关系

多样本聚类——主成分(PCA)分析





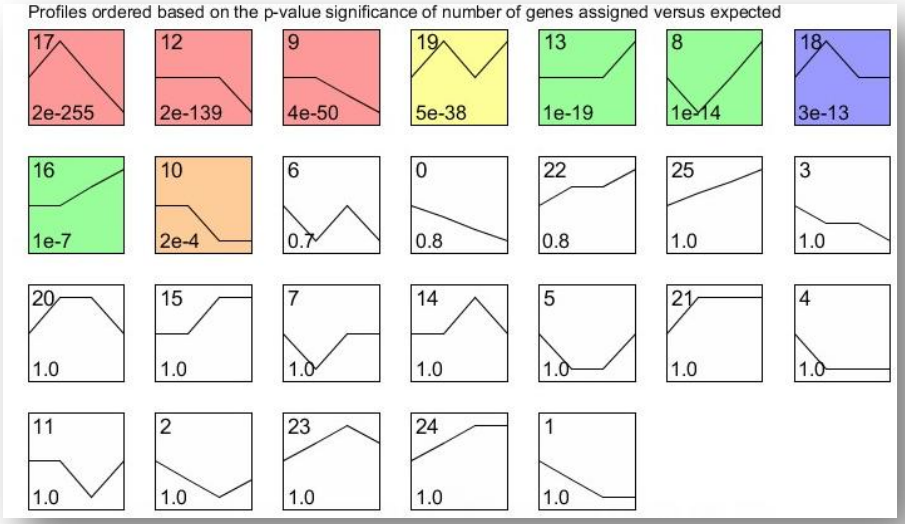
- 样本聚类（样本间差异,主要实验因素，异常样本...）
- 主要基因鉴定

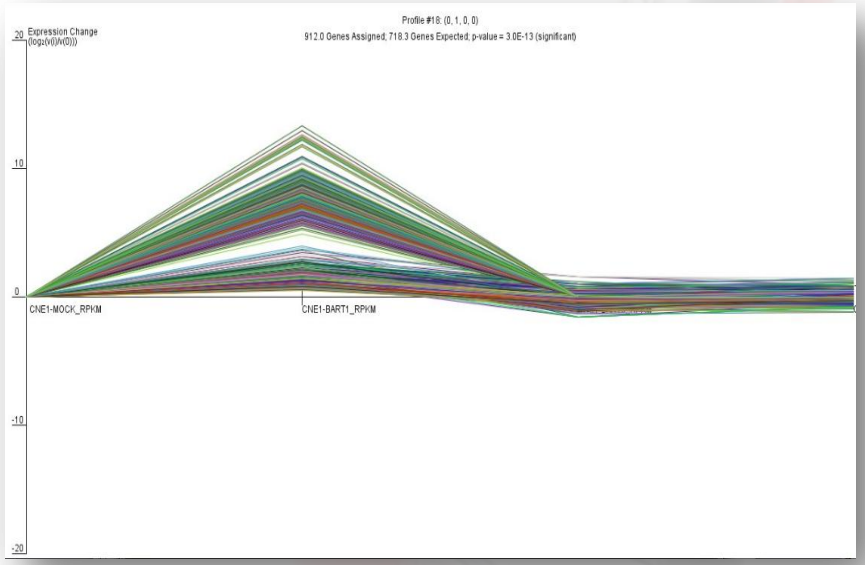
基迪奥生物
GENE DENOVO

多样本聚类——基因表达趋势分析

样本1　样本2　样本3　样本4

生命周期



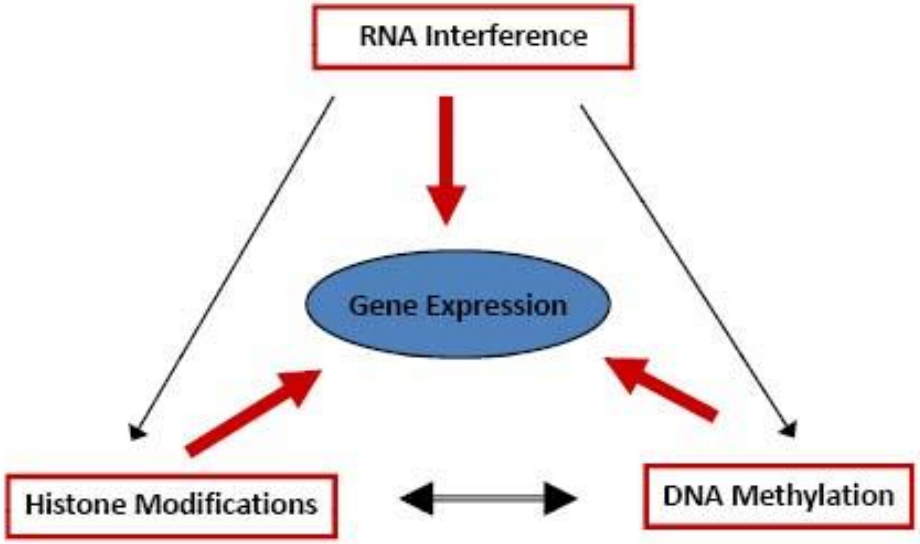


利用表达模式（趋势）对基因进行分类。

基迪奥生物
GENE DENOVO

表观遗传调控

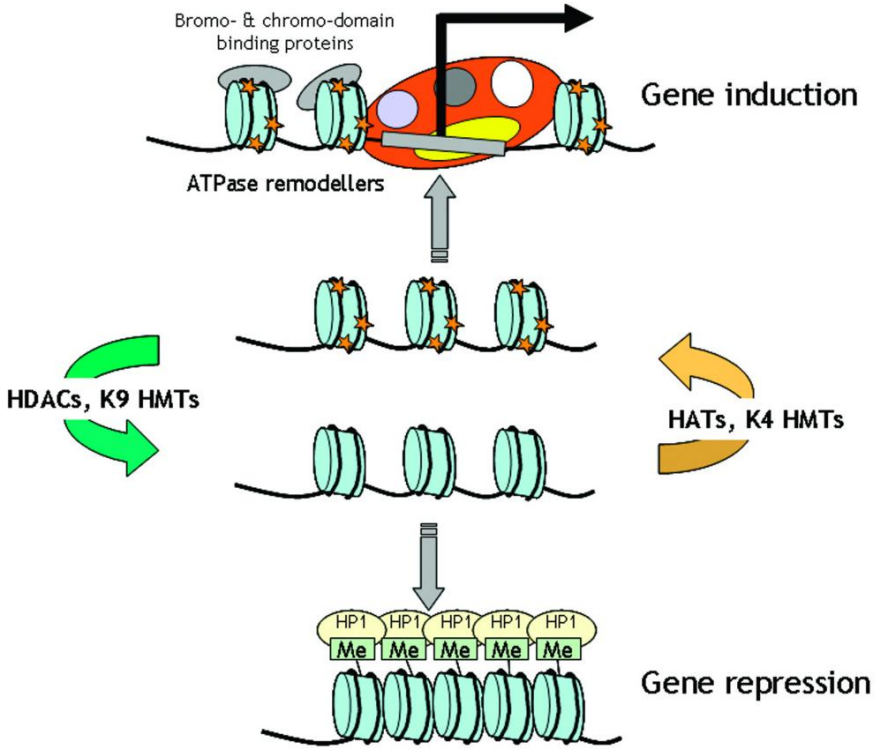
- RNA 干扰；
- DNA 甲基化；
- 组蛋白修饰以及蛋白DNA互作；



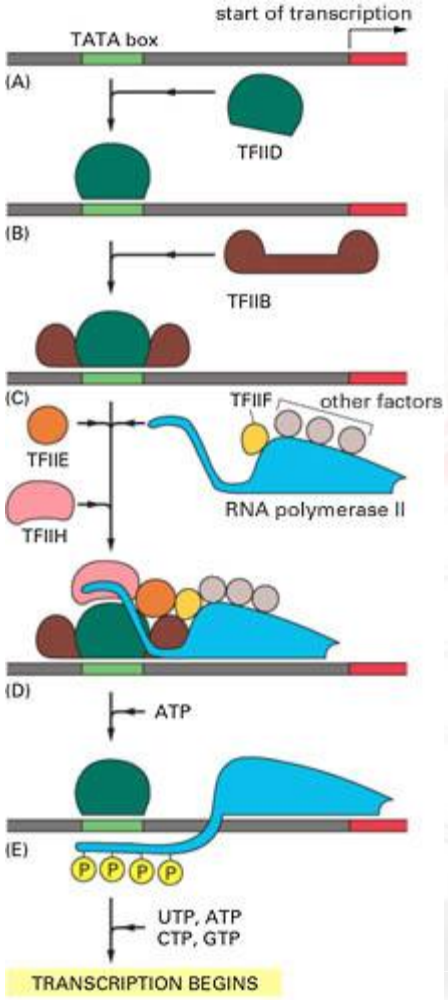
基迪奥生物
GENE DENOVO

蛋白DNA互作

组蛋白修饰



转录因子



广州基迪奥生物科技有限公司
邮箱:sercive@genedenovo.com 电话:020-39340225 传真:020-39354111

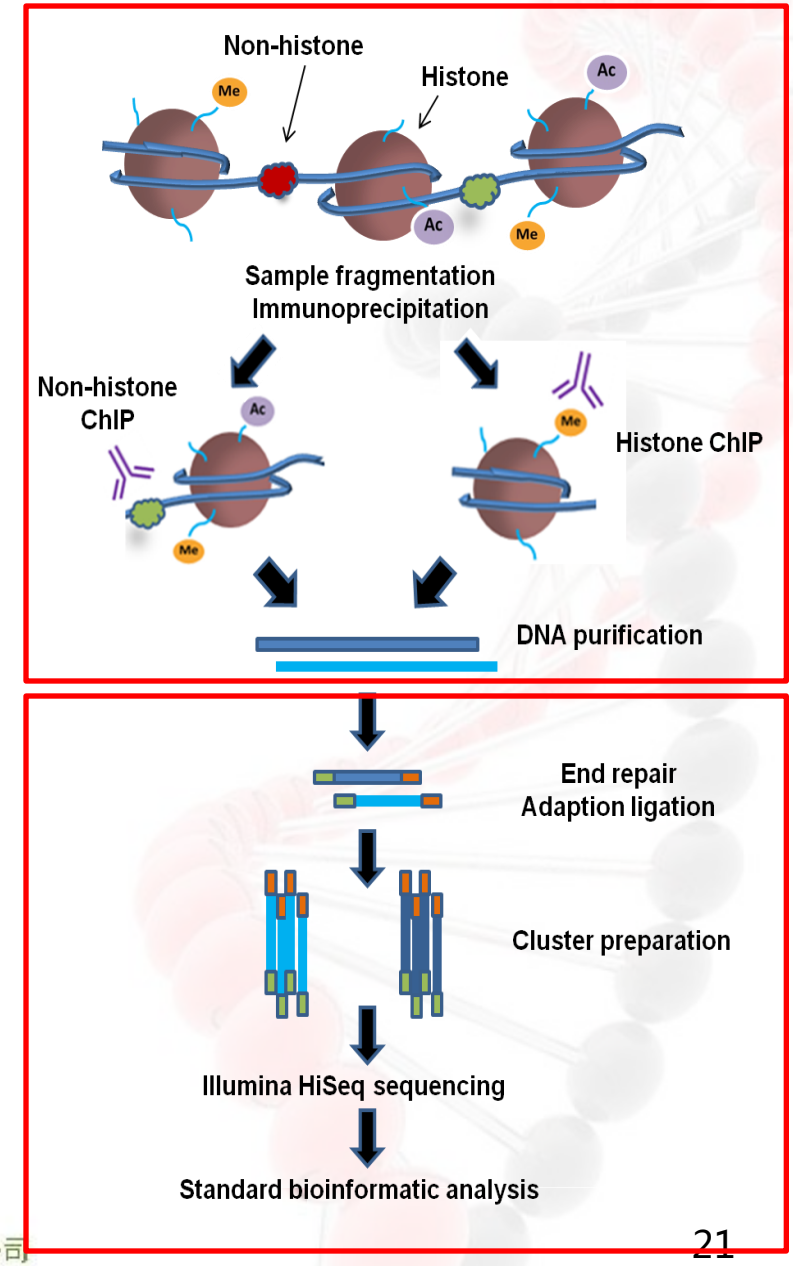
基迪奥生物
GENE DENOVO

ChIP-Seq

◆ **CHIP + Seq**
Chromatin immunoprecipitation和高通量测序的结合

◆全基因组范围内研究蛋白与**DNA**的相互作用机制

◆高性价比，抗体富集后仅需较少的数据量规模即可鉴定全基因组范围内的结合位点



广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

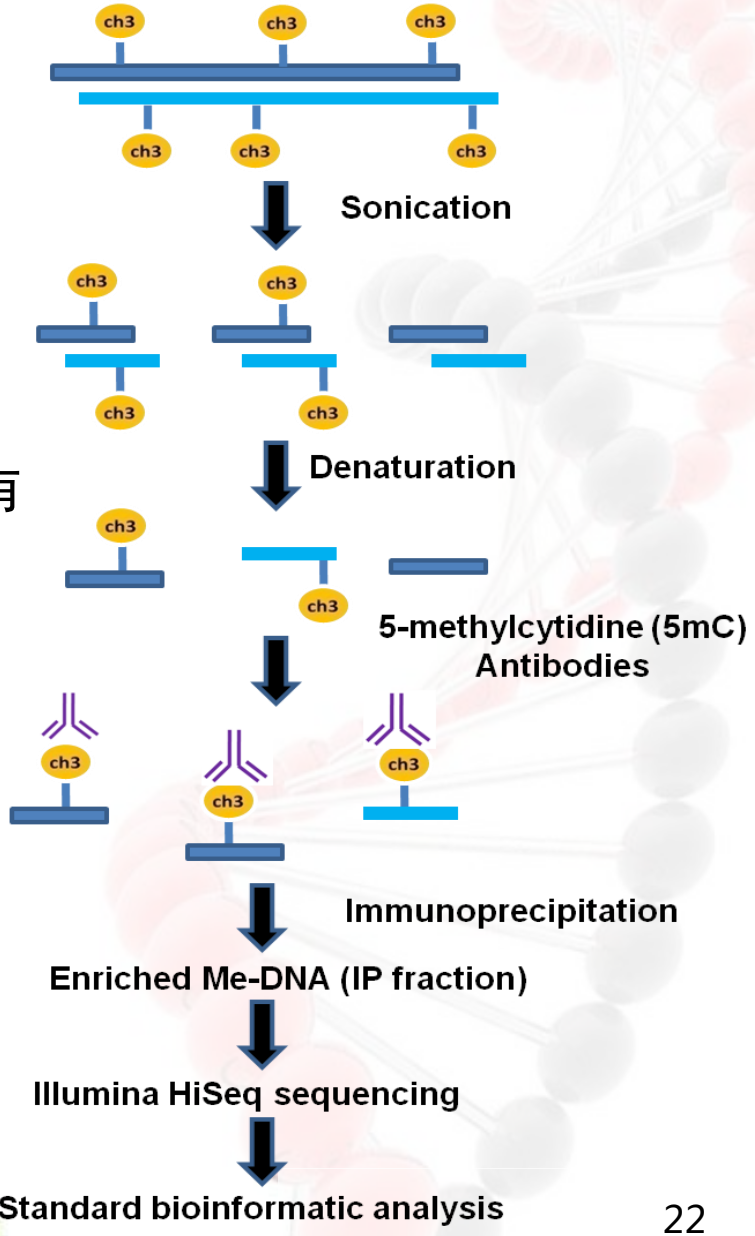
基迪奥生物
GENE DENOVO

DNA methylation
-- MeDIP-Seq

◆基于富集原理，高性价比

◆在高CpG密度，高DNA甲基化水平区域具有很好的抗体亲合性

◆尤适用于大样本量的表观研究



MeDIp/CHIP-Seq分析流程

原始数据

数据清理

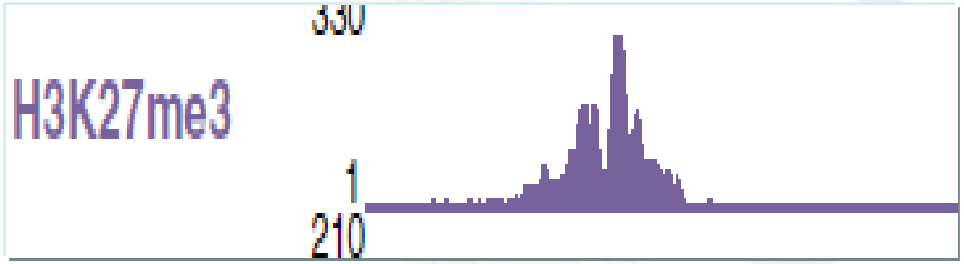
序列比对

Peak 分布 ← Peak 扫描 → Genome Browser可视化

Peak相关基因

GO功能分析

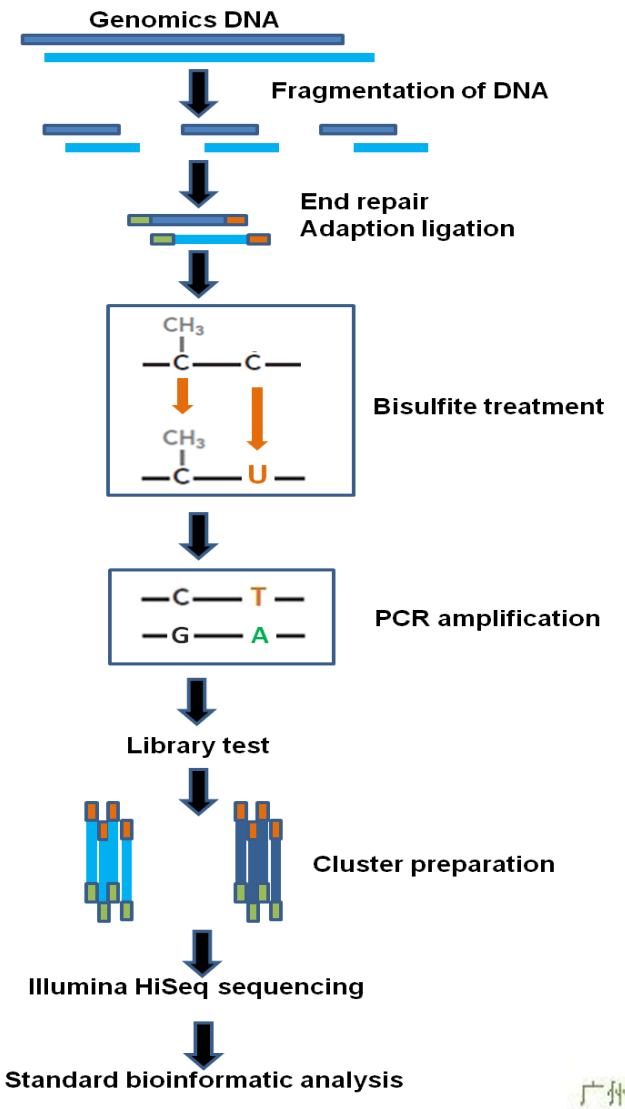
多个样品的差异分析

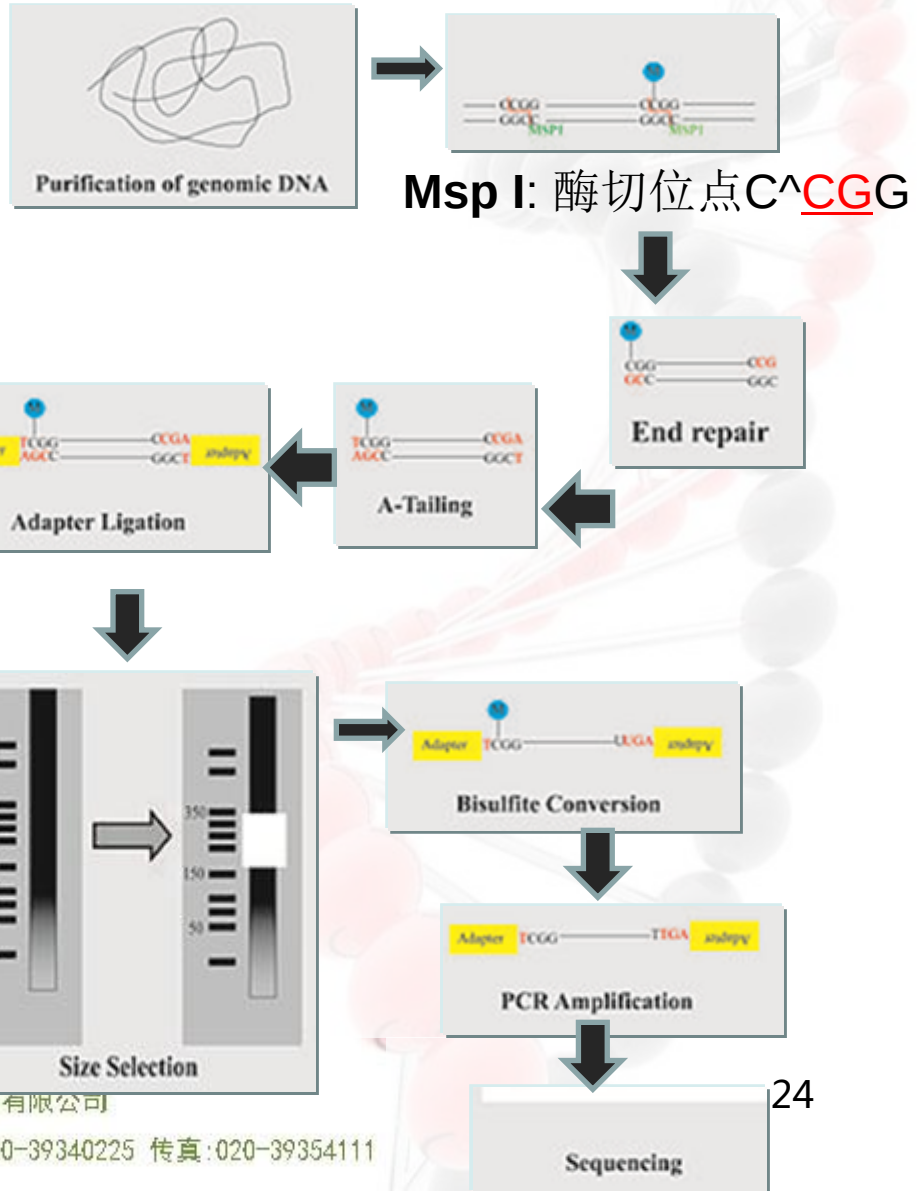


Peak 区域

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

Whole genome BS and RRBS（亚硫酸盐处理）





基迪奥生物
GENE DENOVO

四种技术

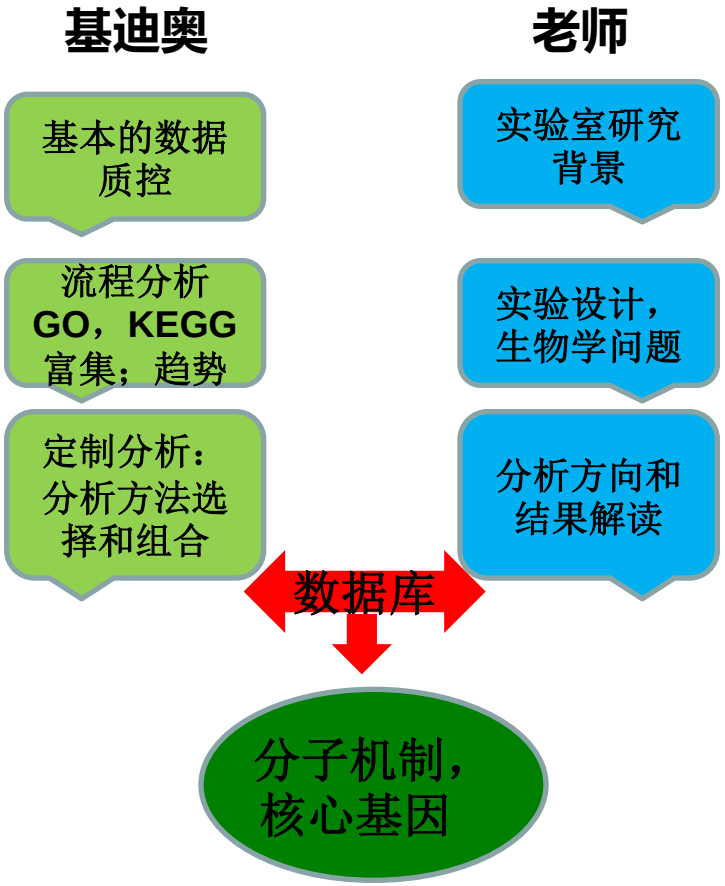
	Whole-genome BS	RRBS	MeDIP-seq	CHIP-seq
覆盖范围	全基因组范围	酶切位点的区域	高甲基化、高CpG密度区域	对应蛋白修饰的区域
精度	单碱基	单碱基	约100bp	约100bp
推荐数据量	基因组大小 30x， 至少20x	4~8G	1-4G，视具体物种不同	1-4G，视具体物种不同
适用范围	获得全基因组高分辨率甲基化图谱	相对甲基化差异的比较，适于大样本量研究	相对甲基化差异的比较，适于大样本量研究	蛋白DNA互作（或多样本差异比较）

基迪奥生物
GENE DENOVO

提 纲

1. 基础——选择适合的NGS技术
2. **框架——定制化方案设计**
3. 亮点——精细挖掘

框架——定制化方案



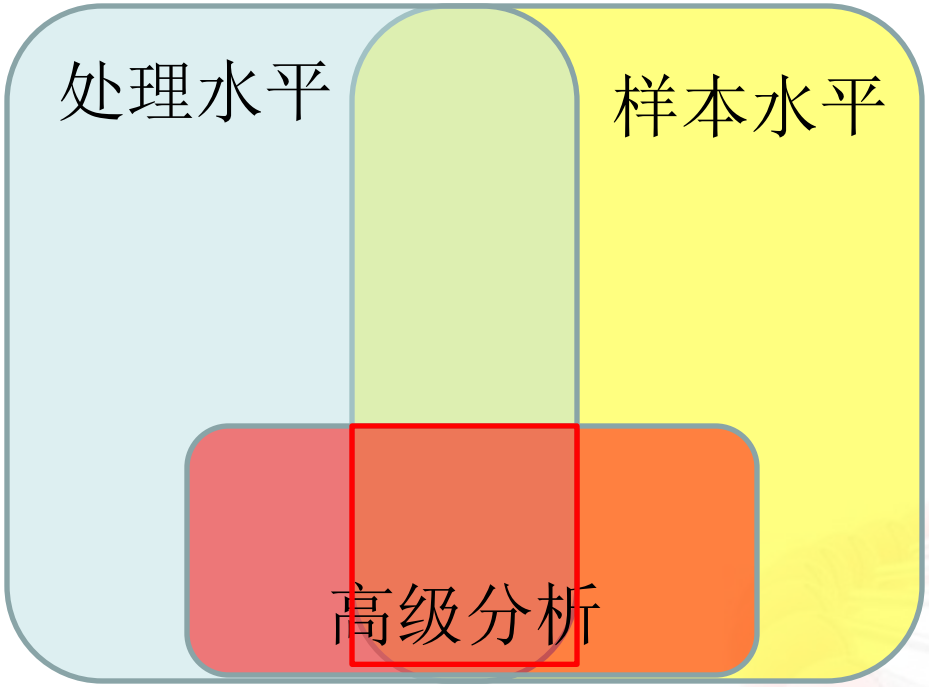
方案：

- 材料特点
- 实验组合
- 分析策略：多组学贯穿、大样本

方案——根据项目特点定制化设计

基迪奥生物
GENE DENOVO

方案组合的意义
——减少目标（候选基因）的数量



思路：过滤 + 分类

基迪奥生物
GENE DENOVO

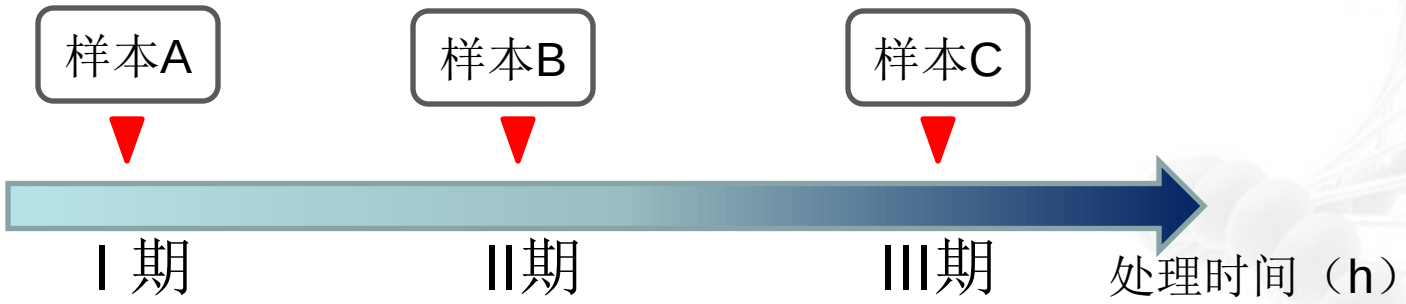
常见方案组合

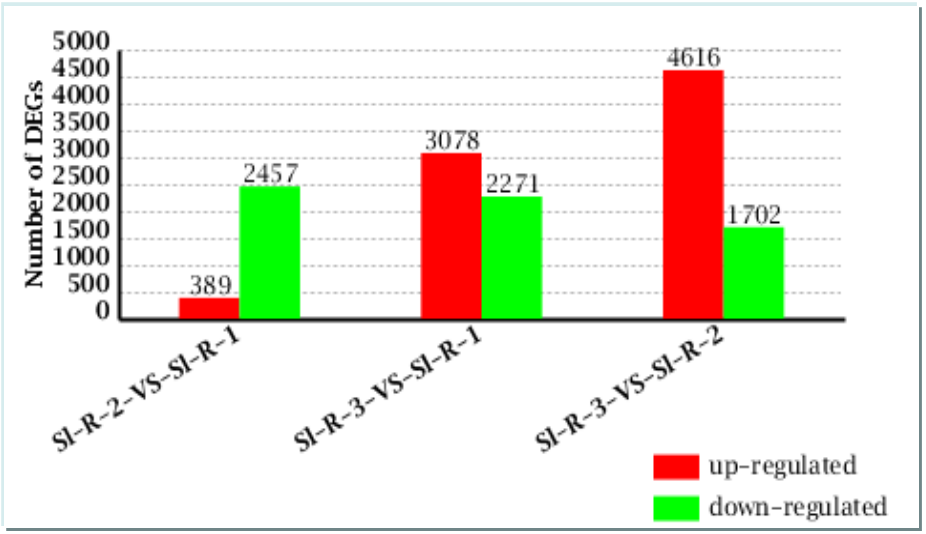
处理	样本关系	高级分析
对照处理	设置重复	贯穿分析
梯度处理	不同品系	基因网络分析
	不同种	特征基因选择
	病原宿主	……

样本、处理组合 + 分析策略 = 核心信息的筛选过滤

邮箱:sercive@genedenovo.com 电话:020-39340225 传真:020-39354111

对照 vs 处理





两两差异分析的不足：

●片面：基因动态变化过程；

●差异基因太多：背景干扰；

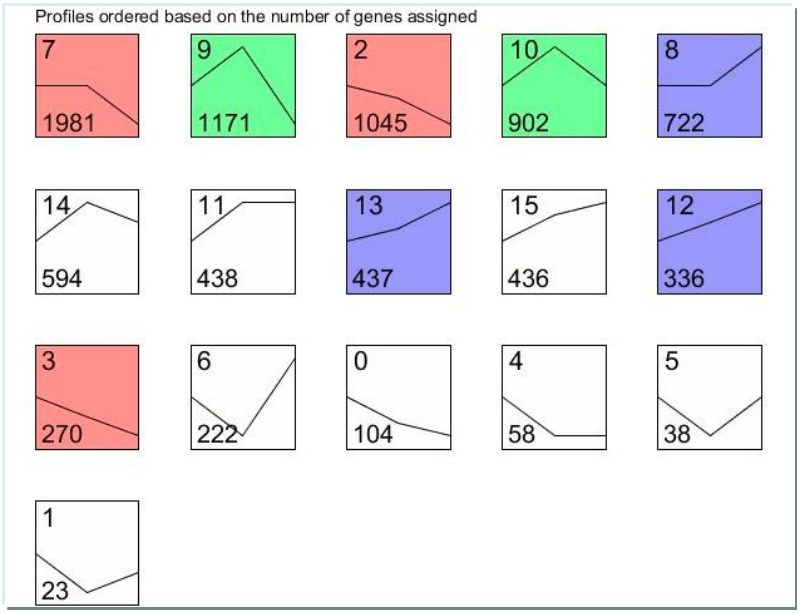
不良影响：

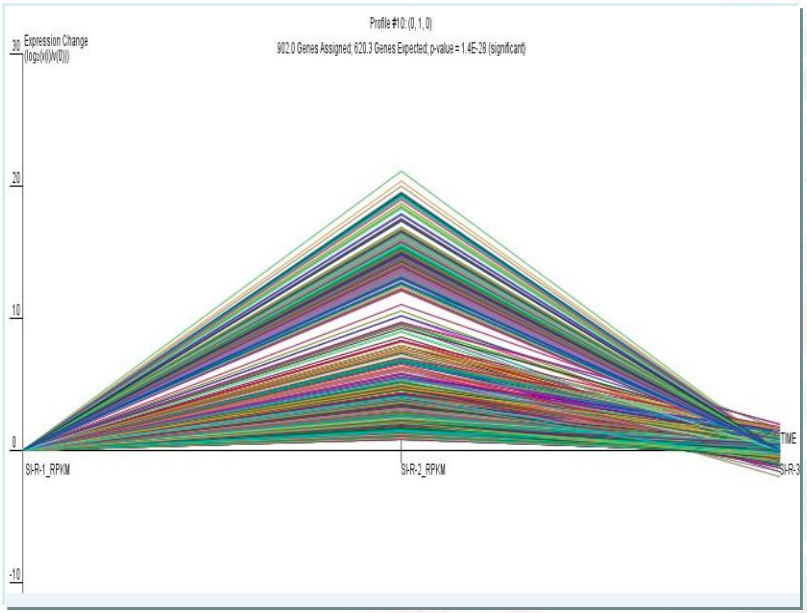
真正的目标淹没在“噪音”中

策略：进一步分类去噪

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

表达趋势分析





●基因数：

8000（总差异基因集）

vs

902（趋势10，潜在生物学意义）

趋势分析筛选关键Pathway

老师预先提出的关注的Pathway:

cytokine-cytokine receptor interaction; ECM-receptor interaction; focal adhesion; gap junction; Jak-STAT signaling; hedgehog signaling; Wnt signaling; renin angiotensin system; VEGF signaling…

#	Pathway	DEGs genes with pathway annotation (309)	All genes with pathway annotation (3125)
1	Focal adhesion	25 (8.09%)	83 (2.66%)
2	Adherens junction	12 (3.88%)	33 (1.06%)
3	ECM-receptor interaction	16 (5.18%)	57 (1.82%)
4	Cytokine-cytokine receptor interaction	11 (3.56%)	31 (0.99%)
5	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	9 (2.91%)	23 (0.74%)
6	Leukocyte transendothelial migration	10 (3.24%)	32 (1.02%)
7	Autoimmune thyroid disease	7 (2.27%)	18 (0.58%)
8	Cytosolic DNA-sensing pathway	6 (1.94%)	14 (0.45%)
9	Axon guidance	15 (4.85%)	66 (2.11%)
10	RNA polymerase	9 (2.91%)	30 (0.96%)
11	Jak-STAT signaling pathway	8 (2.59%)	25 (0.8%)
12	Amino sugar and nucleotide sugar metabolism	12 (3.88%)	51 (1.63%)

邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

处理 + 不同样本 + 重复(案例)

PNAS

Genomic basis for coral resilience to climate change

Daniel J. Barshis[1,2], Jason T. Ladner, Thomas A. Oliver, François O. Seneca, Nikki Traylor-Knowles, and Stephen R. Palumbi

Department of Biology, Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950

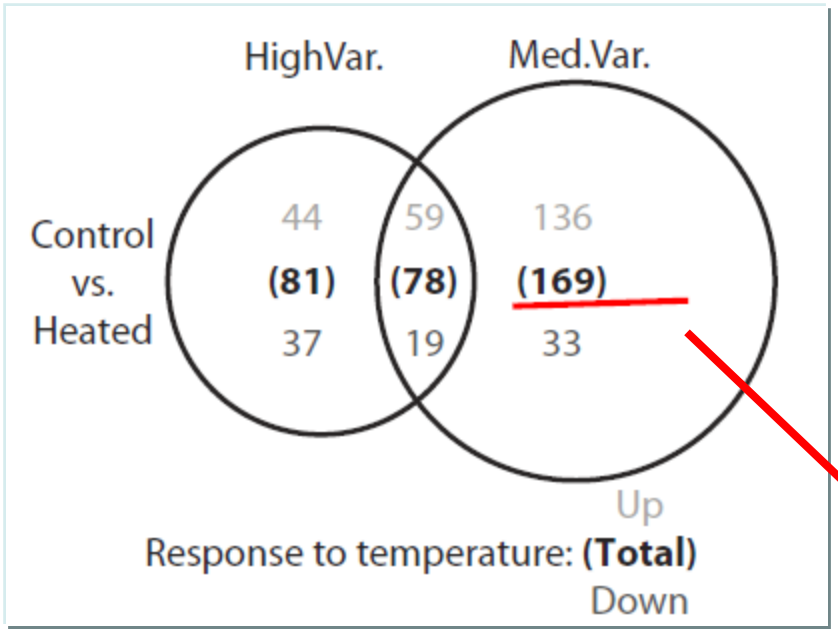
- 研究目的：珊瑚的高温适应性分子机理
- 材料处理

	普通品种（A1）	耐高温品种（A2）
常温处理（B1）	A1B1	A2B1
高温处理（B2）	A1B2	A2B2

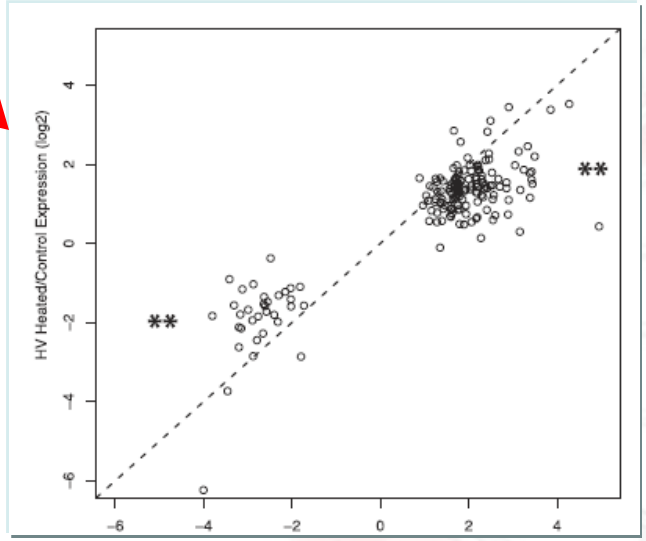
备注：每个组都设置了生物学重复

- 研究方法：RNA-seq

处理 + 品系 + 重复

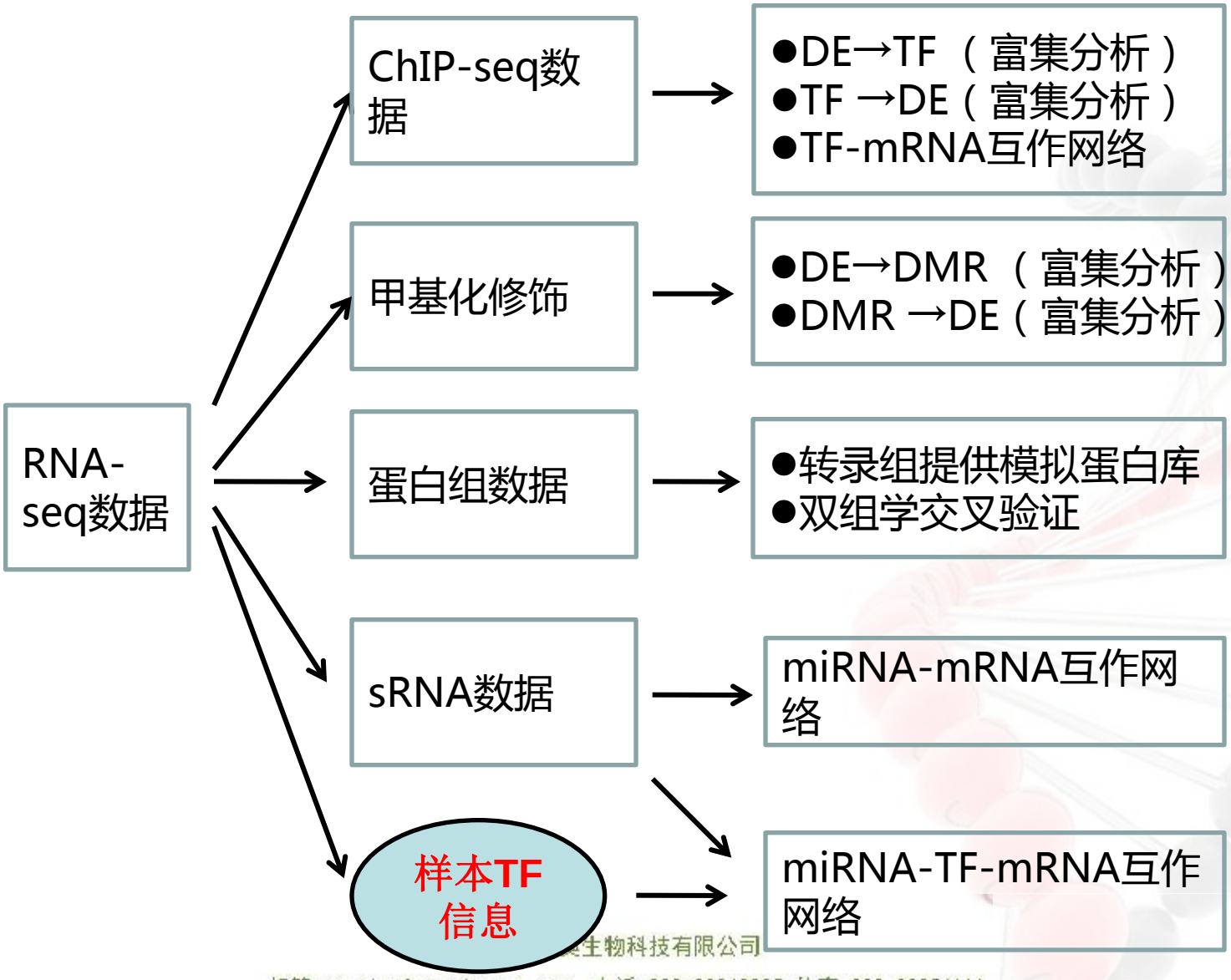


- 对照-高温 差异分析
- 两个品种差异基因对比（维恩图）
- 高温品种应激反应更小



结论：珊瑚中的Hsp70, TNF, peroxidasin和zinc metalloproteases 与其适应热环境能力有关

RNA-seq数据的后续贯穿分析



基迪奥生物
GENE DENOVO

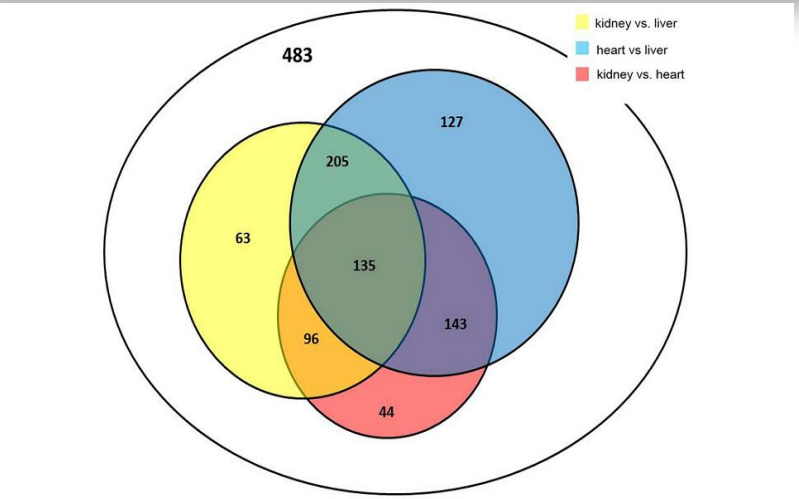
表达差异与甲基化差异的关系

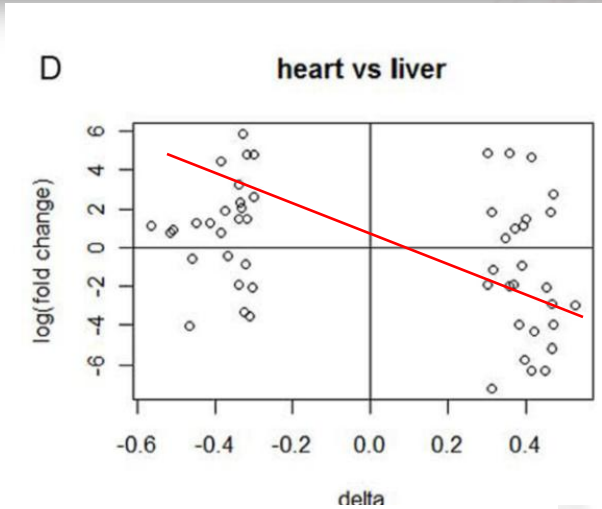
Xie *et al. BMC Systems Biology* 2011, **5**(Suppl 3):S4
http://www.biomedcentral.com/1752-0509/5/S3/S4

BMC Systems Biology

RESEARCH Open Access

An integrative analysis of DNA methylation and RNA-Seq data for human heart, kidney and liver





- 总共有1296个差异表达基因
- 其中有813个基因（2/3）对应的甲基化差异
- 甲基化差异与基因表达差异呈负相关

基迪奥生物
GENE DENOVO

不同组学的贯穿
——表达谱与miRNA贯穿分析

- **客户材料及分析需求：**

材料：两个品种10个生长发育时期的表达谱和小RNA 数据；

研究目的：两个品种生长发育差异的调控机制（转录因子、小RNA、基因表达三者间的关系）；

问题：

（1）多个样本数据如何比较？（两两比较=> 局部 ≠ 整体）

（2）如果构建简洁有效的网络 （简洁 *vs* 信息完整）

	生长阶段									
	1期	2期	3期	4期	5期	6期	7期	8期	9期	10期
品种A	sample	sample	sample	sample	sample	sample	sample	sample	sample	sample
品种B	sample	sample	sample	sample	sample	sample	sample	sample	sample	sample

项目总体方案确认

基迪奥生物
GENE DENOVO

老师

科研思路

- 材料：两个品种10个生长发育时期的表达谱和小RNA 数据；
- 研究目的：两个品种骨骼肌生长发育差异的调控机制（转录因子、小RNA、基因表达三者间的关系）；

整体解决方案

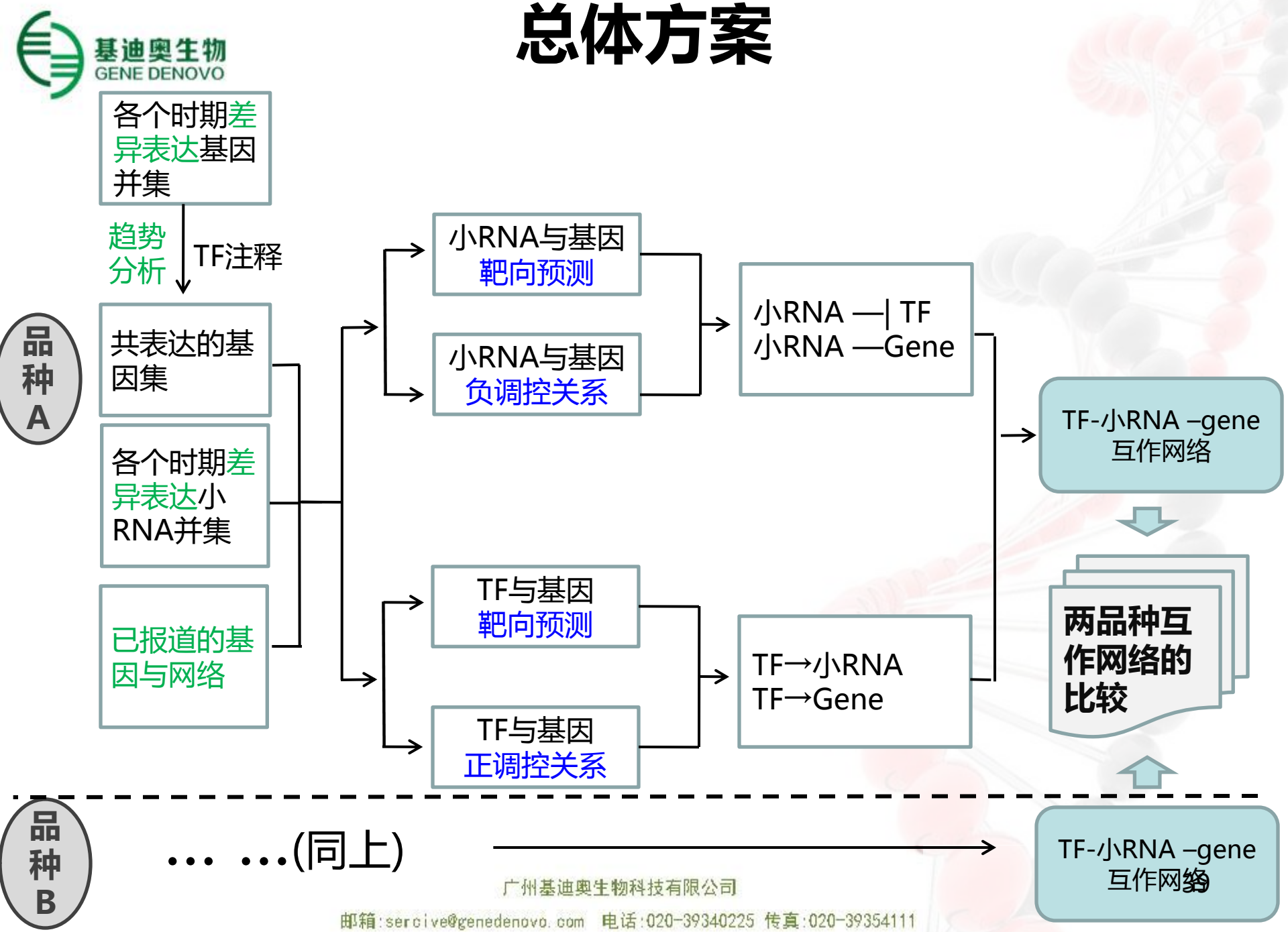
基迪奥

分析策略

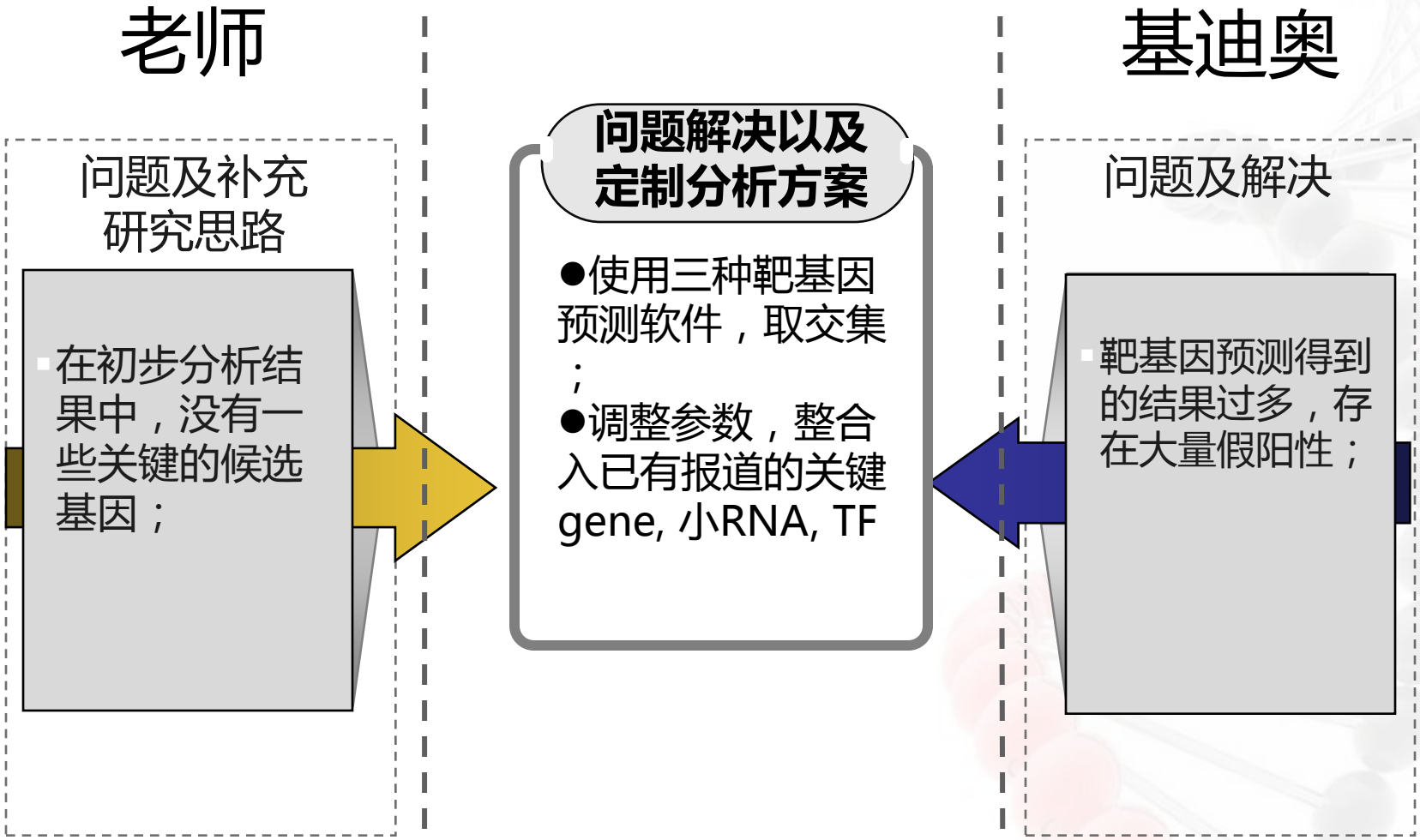
- 差异表达分析；
- 靶基因预测；
- 趋势分析；
- 相关分析；
- 基因互作网络分析；

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

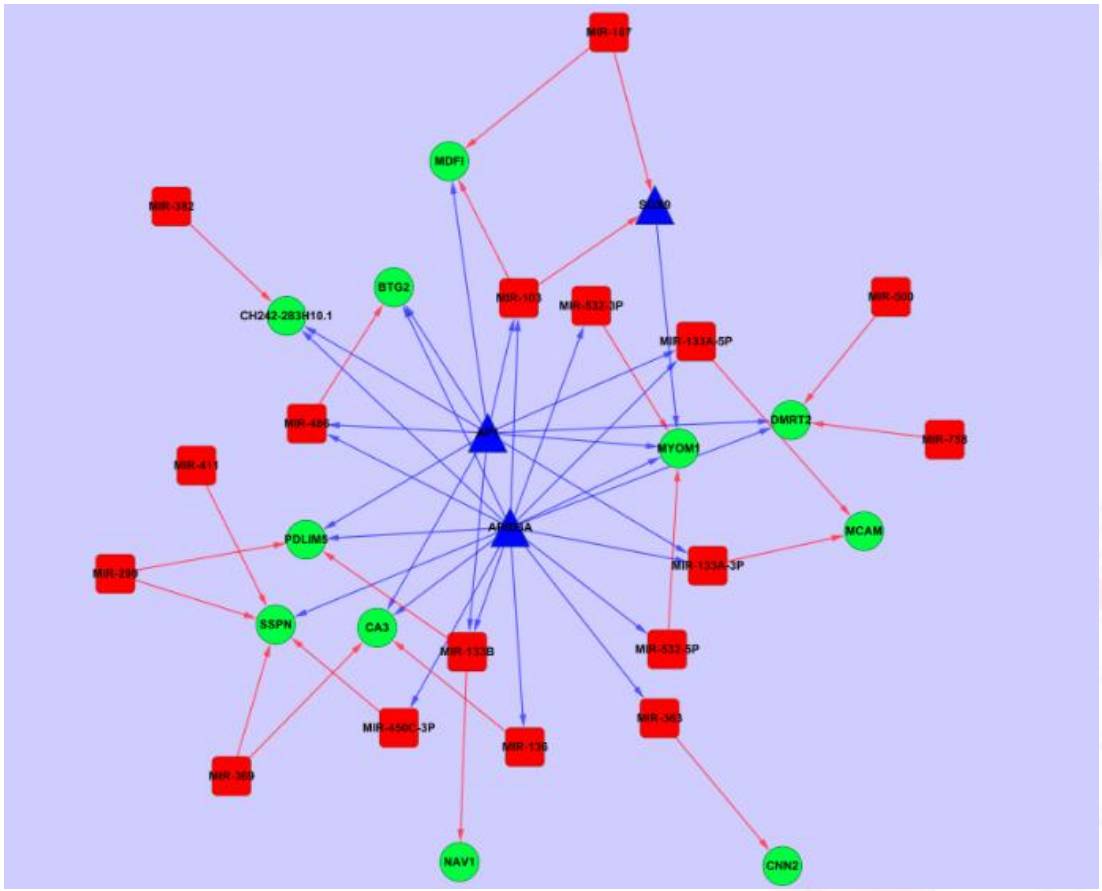
总体方案


基迪奥生物
GENE DENOVO
各个时期差异表达基因并集
趋势分析
TF注释
品种A
共表达的基因集
各个时期差异表达小RNA并集
已报道的基因与网络
小RNA与基因靶向预测
小RNA与基因负调控关系
TF与基因靶向预测
TF与基因正调控关系
小RNA —| TF
小RNA —Gene
TF→小RNA
TF→Gene
TF-小RNA –gene互作网络
两品种互作网络的比较
品种B
… …(同上)
TF-小RNA –gene互作网络
广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

分析优化


老师
基迪奥
问题及补充
研究思路
在初步分析结果中，没有一些关键的候选基因；
问题解决以及
定制分析方案
●使用三种靶基因预测软件，取交集；
●调整参数，整合入已有报道的关键gene, 小RNA, TF
问题及解决
靶基因预测得到的结果过多，存在大量假阳性；

结果——互作网络



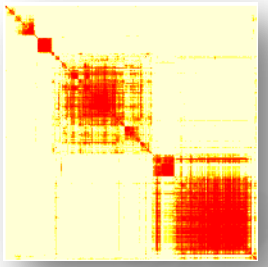
TF-miRNA –gene 互作网络

广州基迪奥生物科技有限公司
邮箱:sercive@genedenovo.com 电话:020-39340225 传真:020-39354111

WGCNA（权重共表达网络）

关系矩阵构建

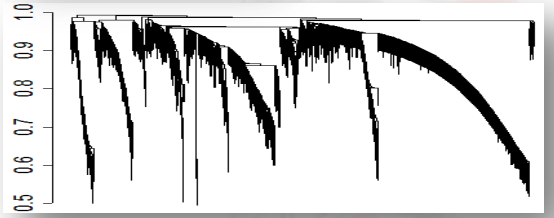
基本原理：利用基因间表达量的相关系数。



↓

模块识别

基本原理：利用拓扑树结构区分基因模块。

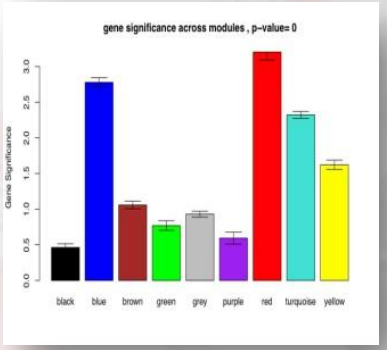


↓

核心模块挑选

基本原理：分析模块内基因的特性，进一步寻找有生物学意义的模块。

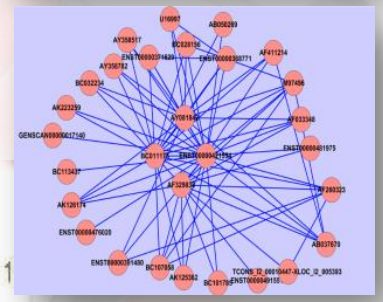
分析策略：模块特征值与表型间的相关性，模块内基因的KO、GO分析。



↓

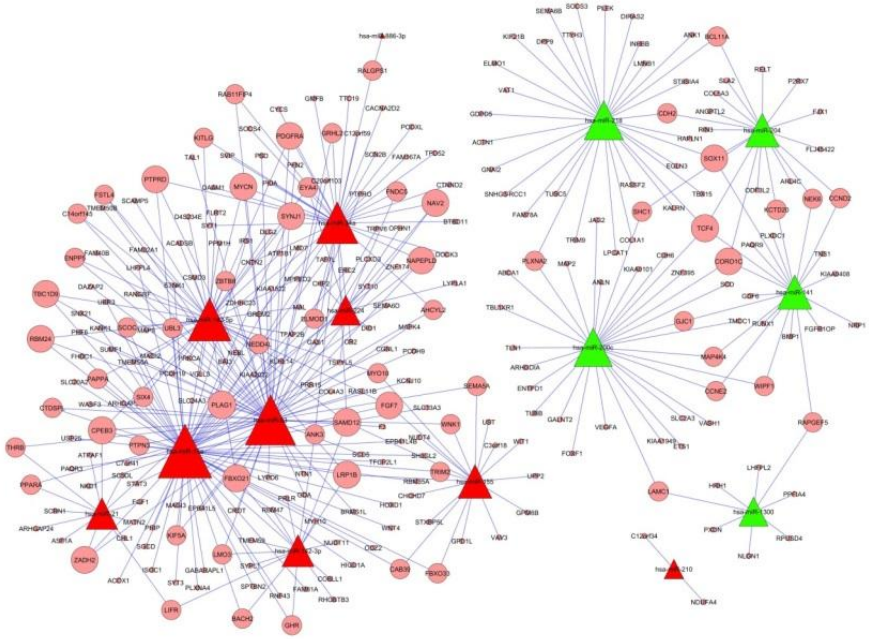
核心基因的挑选，并构建网络

基本原理：利用基因连通性信息挑选核心基因，并围绕其构建网络



基迪奥生物
GENE DENOVO

数据库来源



- 实验数据
- 公用数据

MaizeGDB
Maize Genetics and Genomics Database

DDBJ
DNA Data Bank of Japan

OMIM®

MGI

NCBI

e!Ensembl ASIA

FlyBase

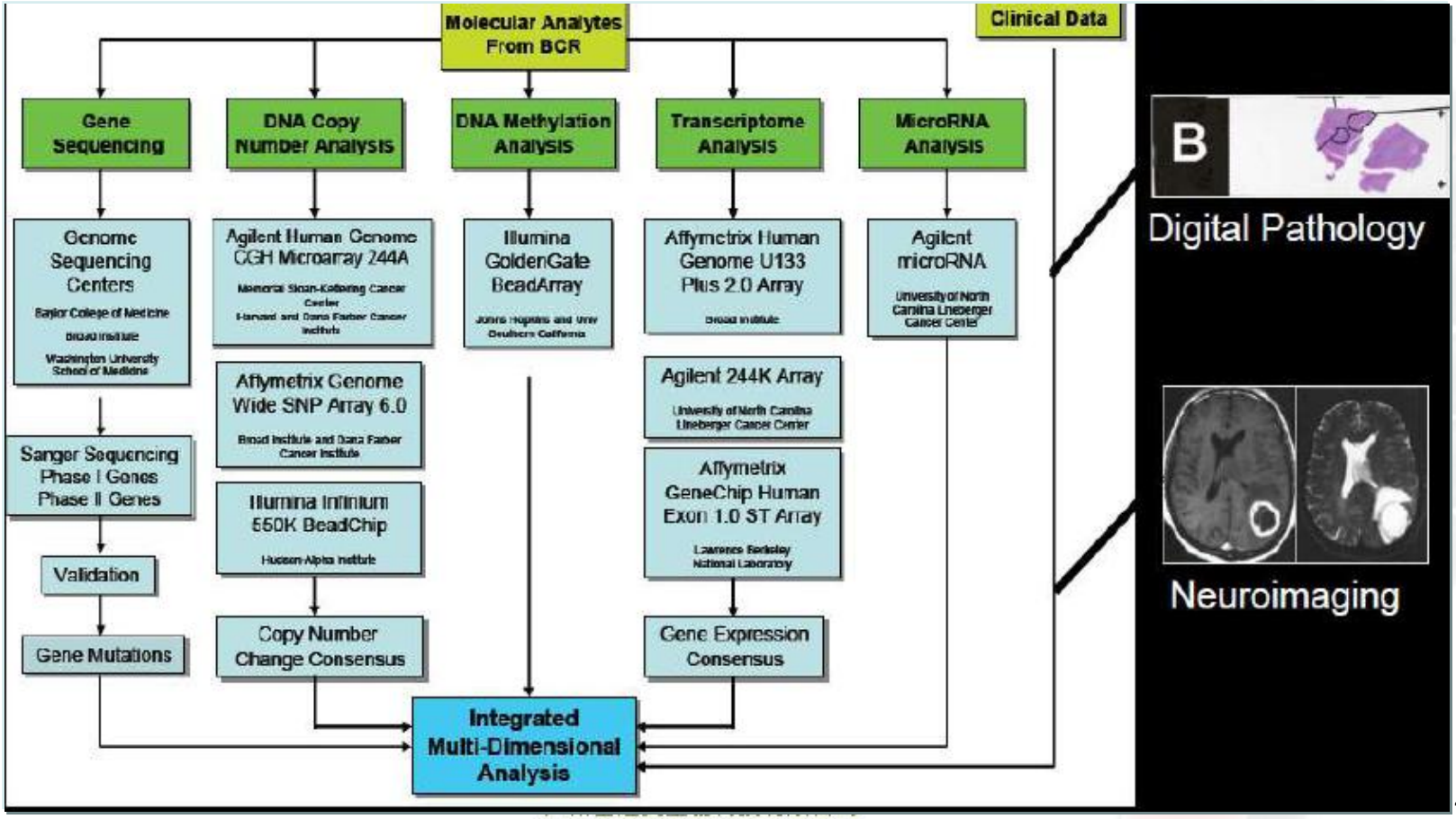
GeneCards®
The Human Gene Compendium

PlantGDB

AnimalTFDB
Animal Transcription

TCGA(The Cancer Genome Atlas, NIH)数据库

- 一共整合了六个维度的癌症数据

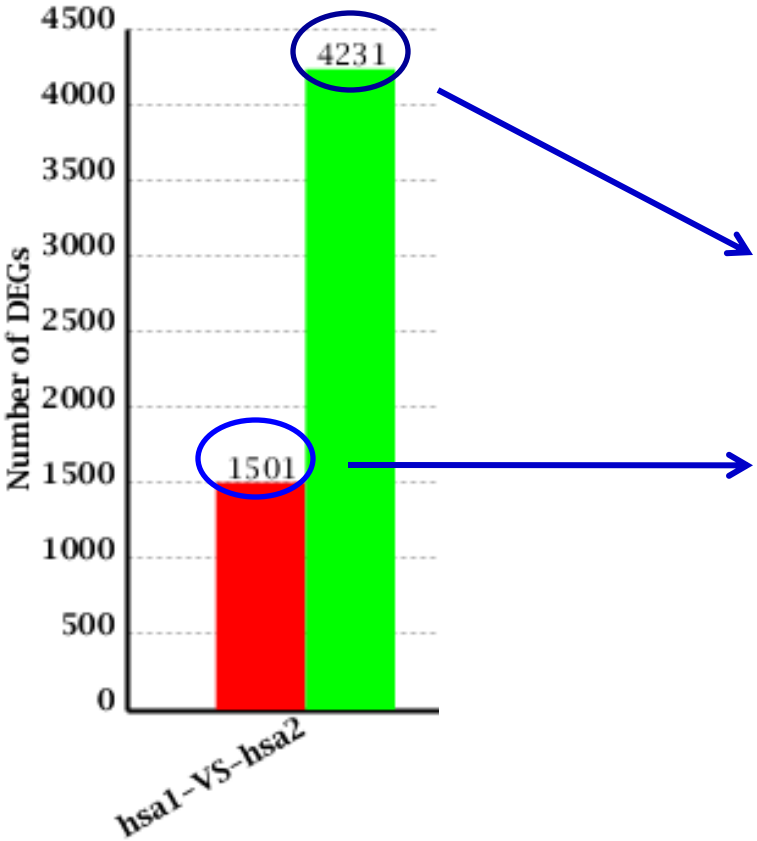


邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

基迪奥生物
GENE DENOVO

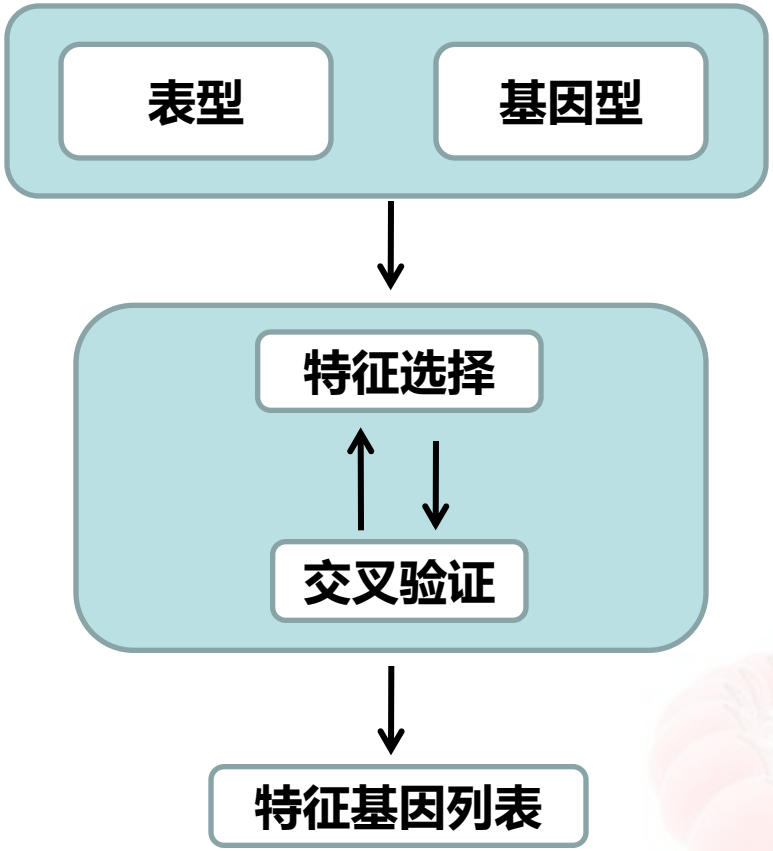
二分类样本——常用策略的瓶颈

差异基因统计图



1）差异基因往往数量巨大；
2）但实际上决定样本性状的，往往只是很少一部分的关键基因；
3）我们如何从中挑选与疾病或性状最相关的基因？

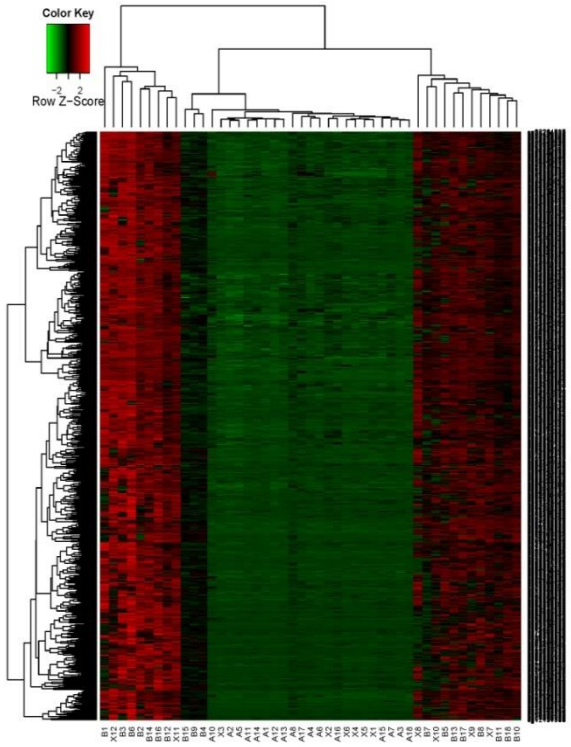
特征基因选择——机器学习



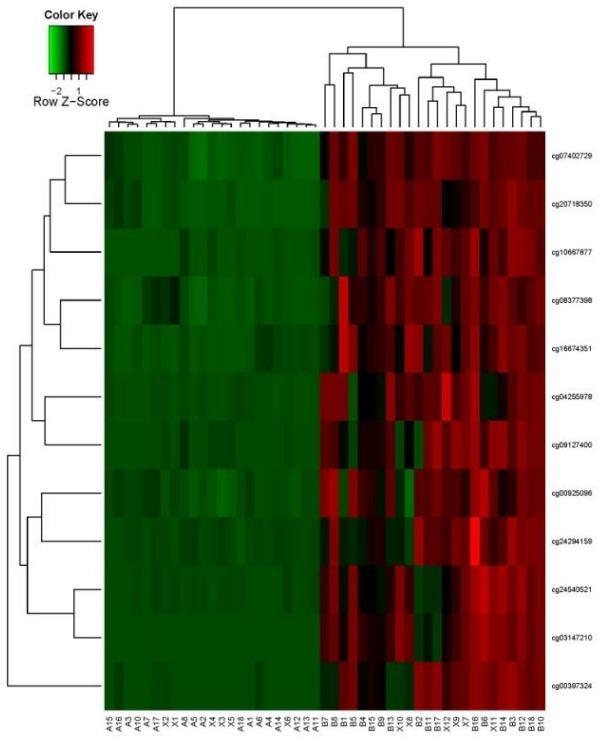
- 特征基因列表：按照表达量与表型的相关性排列。
- 实际分析案例表明，排名靠前的基因往往具有生物学意义。

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

正常人vs 病人分类预测



所有基因的分类结果不明确



12个特征基因的分类清晰明确

➢**原始数据：**

48个样本，1030个基因

➢**方法：**

SVM-FRE

➢**结果：**

•分类准确率100%。

•特征基因：7个为已经报道，预测5个基因新基因

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

基迪奥生物
GENE DENOVO

疾病相关分子标记物筛选的思路

- 自有数据 + 同类的公共数据
  - 例1：采用基因芯片对少数（如20例）肿瘤组织和正常组织进行测量，获得marker，采用公共库（已发表）的大型数据（如>400例）进行独立验证
  - 例2：有cell lines实验中筛选到某重要基因（集），通过已发表数据集观察该基因（或基因集合组成的signature）与临床指标的相关性，反映其临床重要性

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

特征基因与个体化医疗

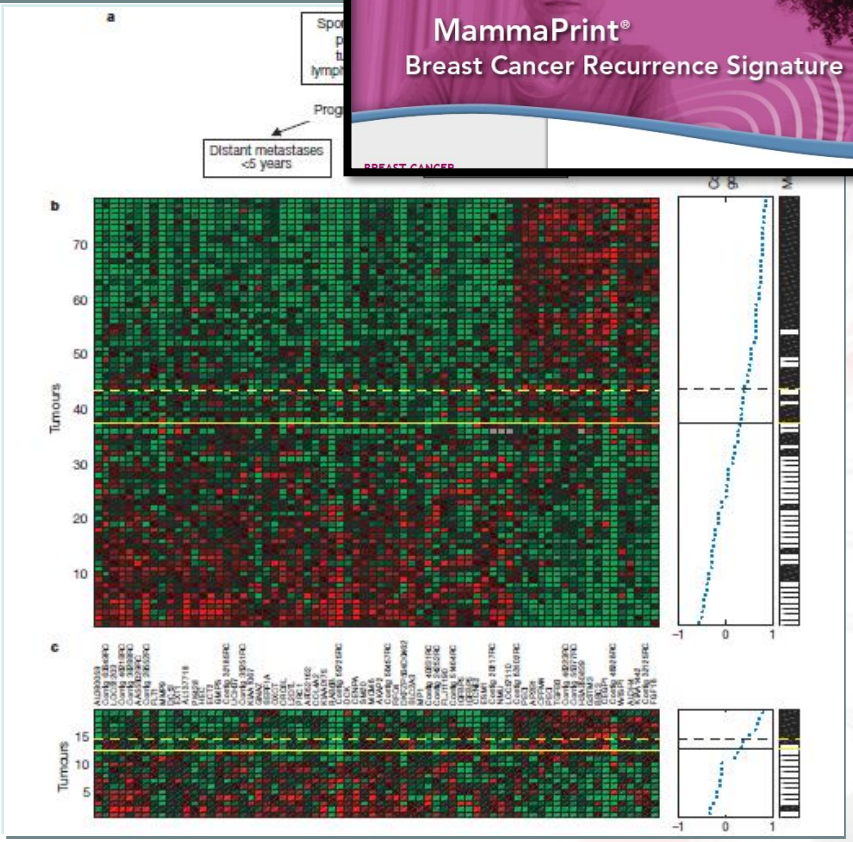
Gene expression profiling predicts clinical outcome of breast cance
LJ van't Veer, H Dai, MJ Van De Vijver, YD He... - nature, 2002 - nature.com
Abstract Breast cancer patients with the same stage of disease can have markedly
treatment responses and overall outcome. The strongest predictors for metastases
example, lymph node status and histological grade) fail to classify accurately breas
被引用次数：6866 相关文章 所有 88 个版本 引用 保存

Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer*†, Hongyue Dai†‡, Marc J. van de Vijver*†, Yudong D. He‡, Augustinus A. M. Hart*, Mao Mao‡, Hans L. Peterse*, Karin van der Kooy*, Matthew J. Marton‡, Anke T. Witteveen*, George J. Schreiber‡, Ron M. Kerkhoven*, Chris Roberts‡, Peter S. Linsley‡, René Bernards* & Stephen H. Friend‡

* *Divisions of Diagnostic Oncology, Radiotherapy and Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands*
‡ *Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA*
† *These authors contributed equally to this work*

Breast cancer patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour[1–3]. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70–80% of patients receiving this treatment would have survived without it[4,5]. None of the signatures of breast cancer gene expression reported to date[6–12] allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of *BRCA1* carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and

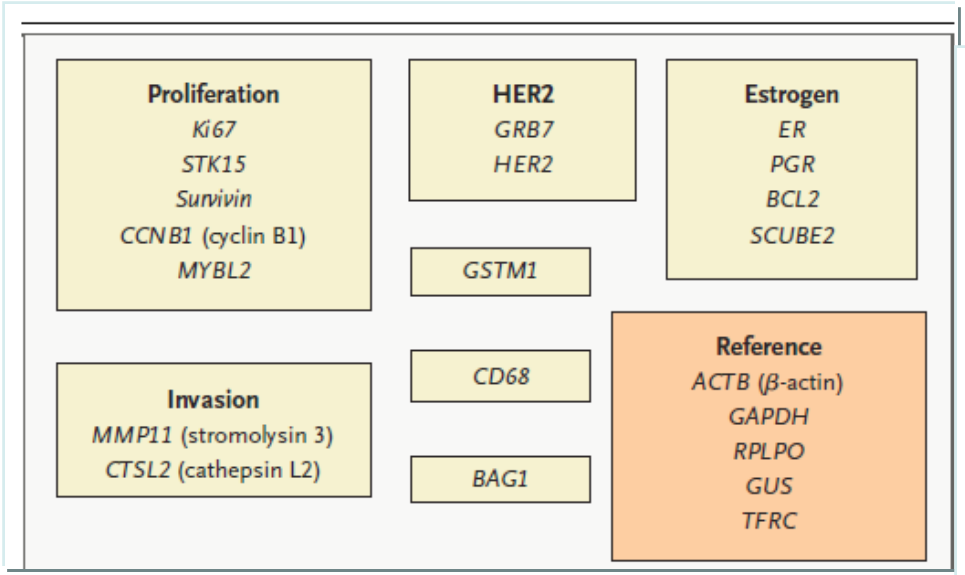


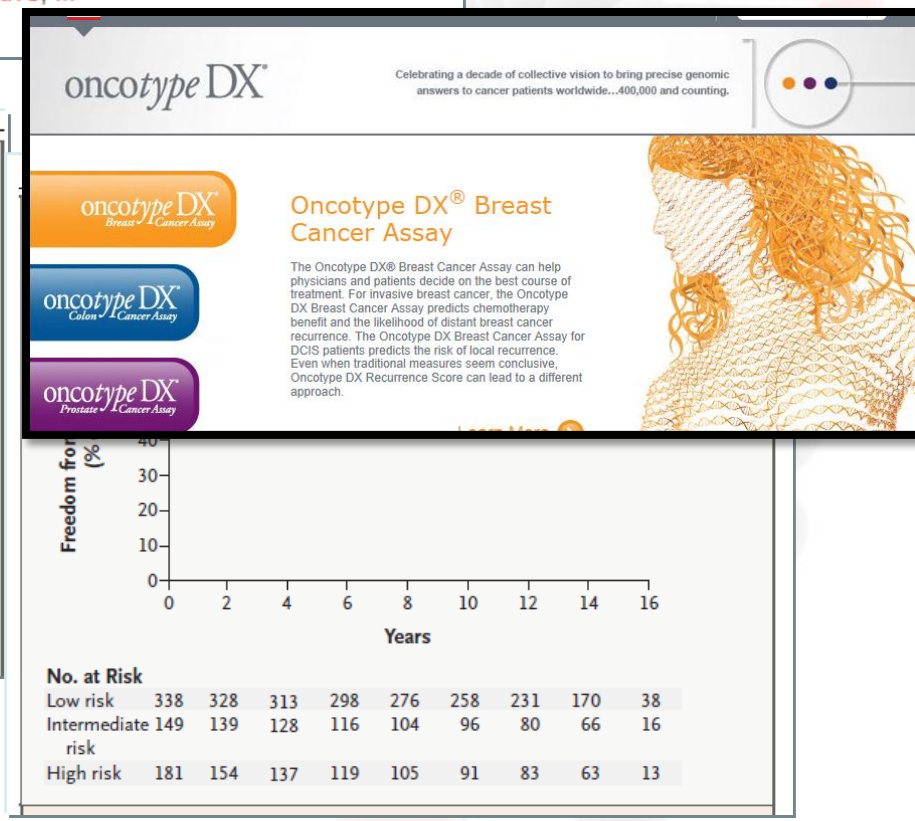
2002年，nature文章：与乳腺癌转移相关的70个特征基因，已经被引用6866次

基迪奥生物
GENE DENOVO

特征基因与个体化医疗

A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer
S Paik, S Shak, G Tang, C Kim, J Baker... - ... England Journal of ..., 2004 - Mass Medical Soc
We tested whether the results of a reverse-transcriptase–polymerase-chain-reaction (RT-PCR) assay of 21 prospectively selected genes in paraffin-embedded tumor tissue would correlate with the likelihood of distant recurrence in patients with node-negative, ...
被引用次数：3036　相关文章　所有 18 个版本　引用　保存





2004年，《新英格兰医学》文章：利用21个基因进行乳腺癌预后预测，已经被引用3036次

基迪奥生物
GENE DENOVO

提　纲

1. 基础——选择适合的NGS技术
2. 框架——定制化方案设计
3. **亮点——精细挖掘**

基迪奥生物
GENE DENOVO

亮点——精细挖掘

- “问题”项目解决
- 目标挖掘
- 结果展示
- 新方法开发

基迪奥生物
GENE DENOVO

问题项目与特殊材料

1. 微量建库样品的处理；
2. 特殊样本——例如肿瘤样本；
3. 污染样品的处理；

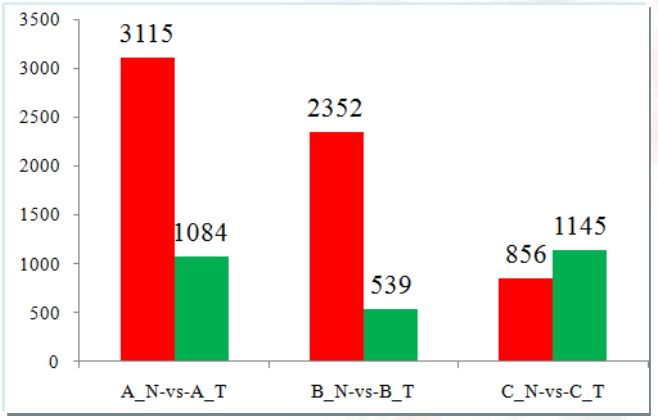
基迪奥生物
GENE DENOVO

肿瘤细胞样本的处理

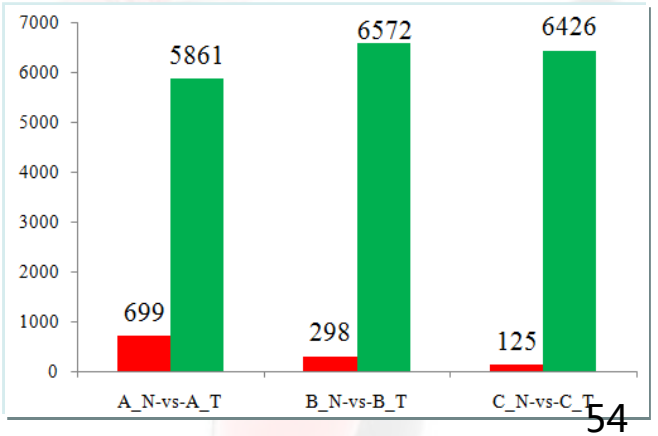
- 样本：肿瘤样本
- 算法：RPKM

$$RPKM = \frac{10^6 C}{NL/10^3}$$

参照系I：总mRNA量

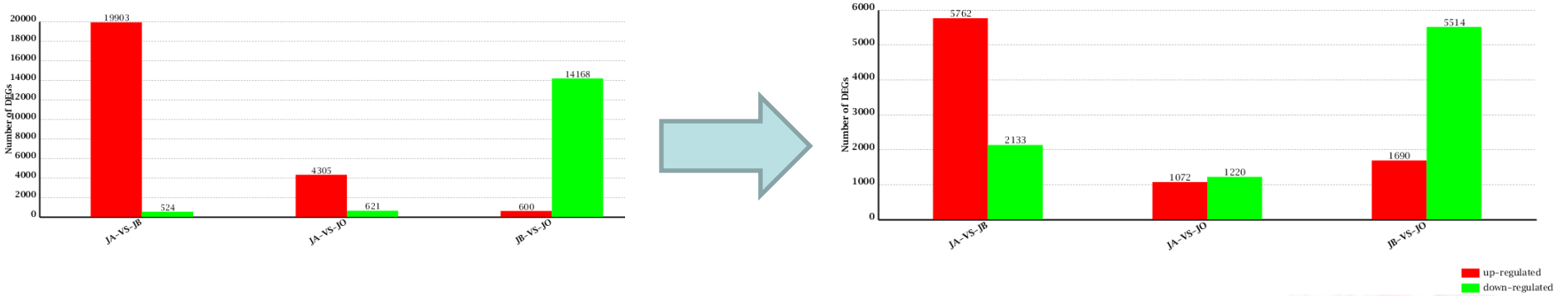


- 参照系II：3个内参基因加权



广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

污染样本

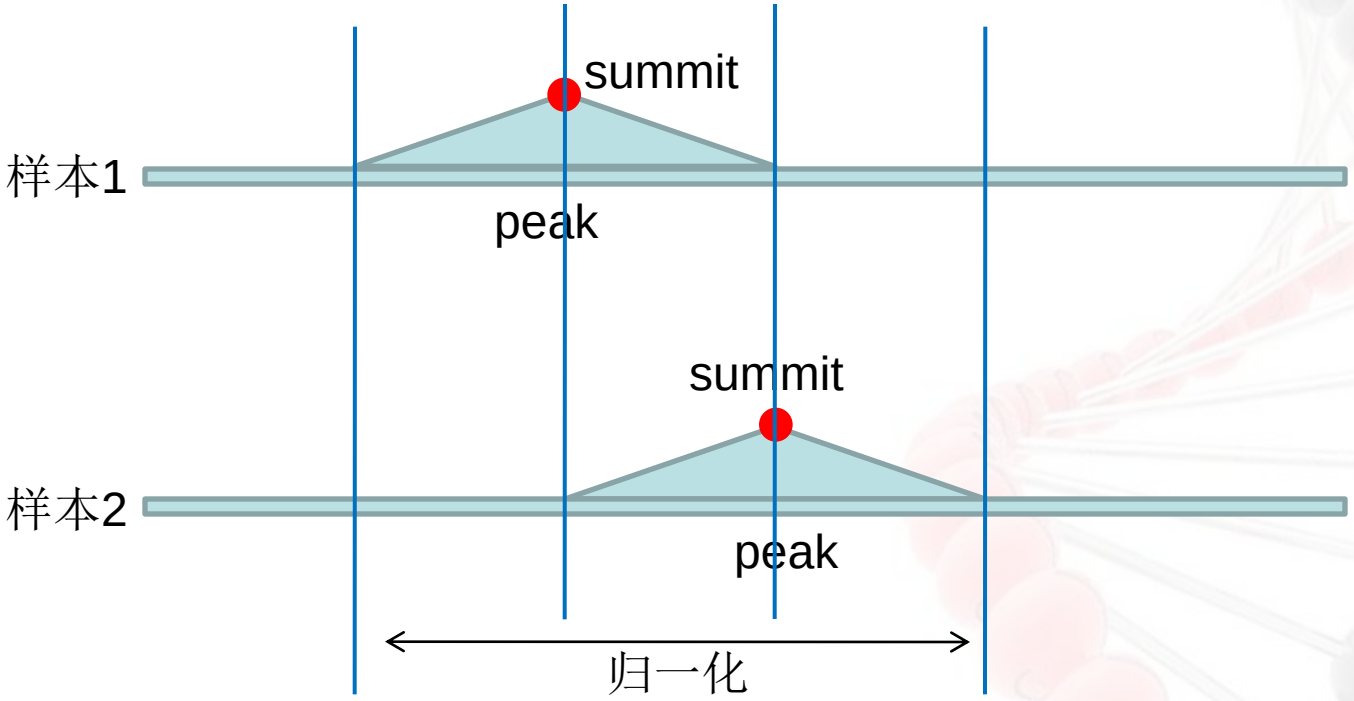


- 对照考察：内参基因
- 参考序列：NCBI → Ensemble
- 算法调整：RPKM → TMM
- 解决问题：有病毒序列污染

基迪奥生物
GENE DENOVO

分析方法可能导致的问题

MeDIP样本的差异甲基化分析



基迪奥生物
GENE DENOVO

目标挖掘——整合已报道信息

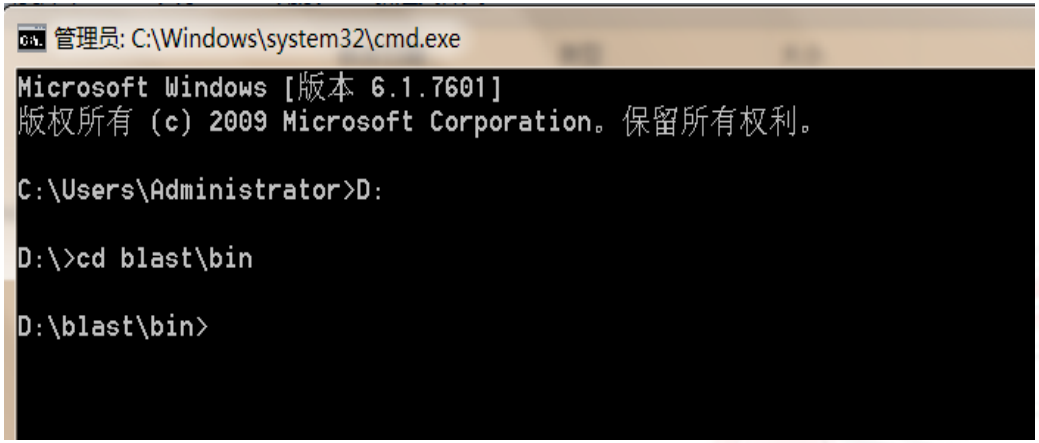
已经报道的基因（重要）

- 寻找：是否存在（blast）
- 比较：如表达模式是否相同
- 线索：相关调控网络以及在网络中的位置

基迪奥生物
GENE DENOVO

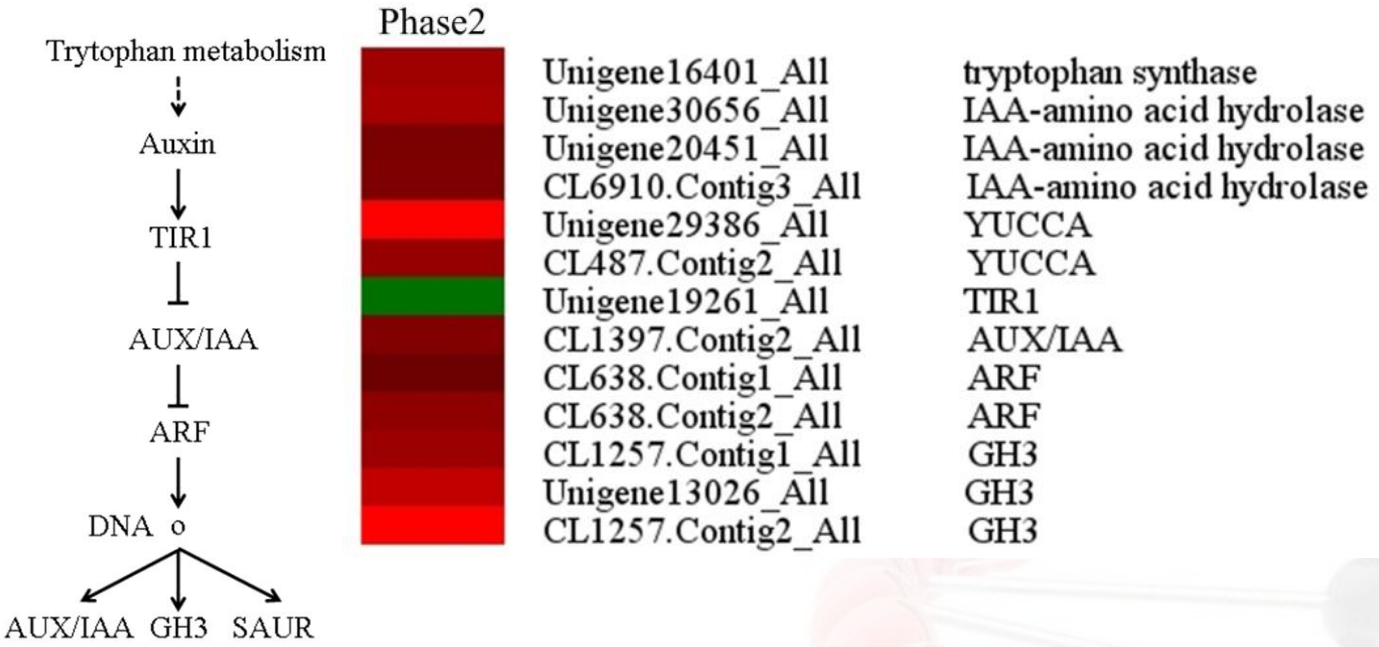
寻找和比较

本地化Blast



我的转录组数据中是否有某个小鼠同源基因？

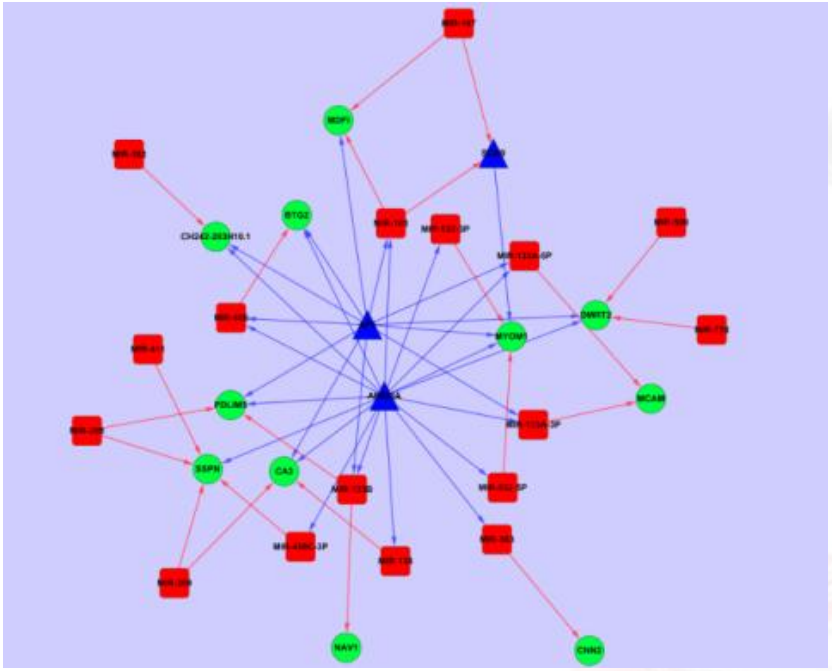
寻找和比较



观察整个Pathway 的表达量变化

基迪奥生物
GENE DENOVO

线索——重新构建网络



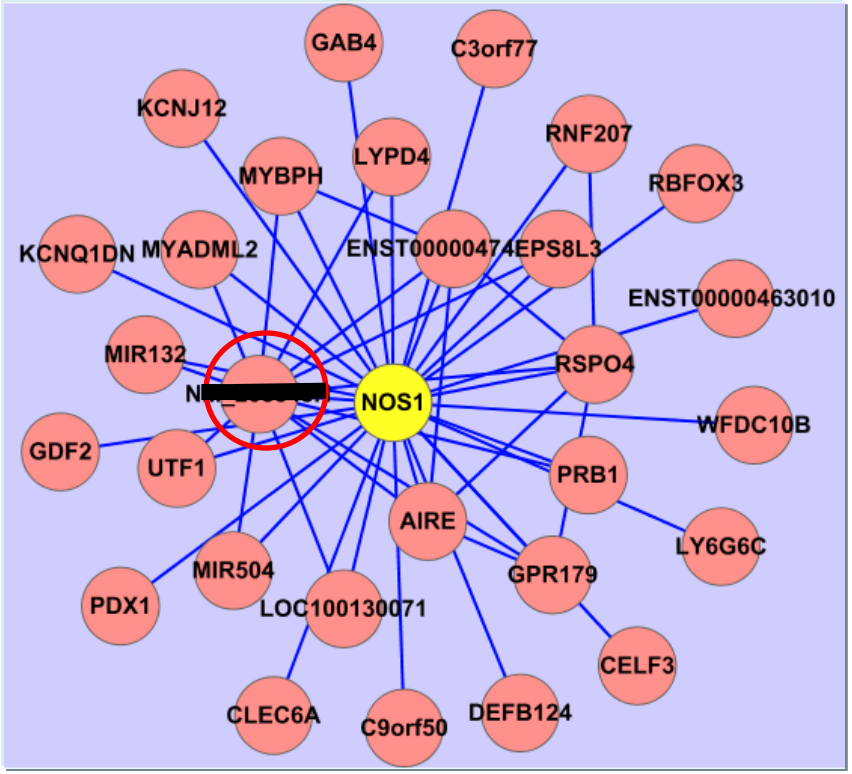
网络构建的生物学意义：

- ●可发表可读：拆分为局部网络，核心网络等
- ●可验证：部分基因是否在文献中验证
- ●预测性：已知基因与新基因关系，提供线索

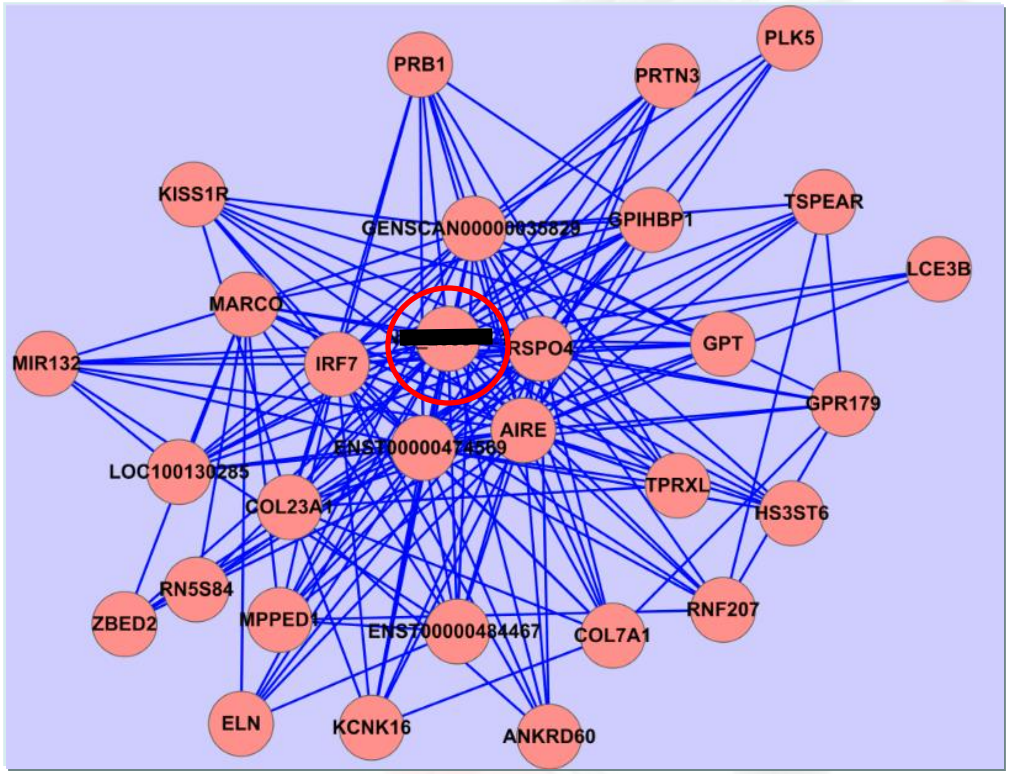
基迪奥生物
GENE DENOVO

WGCNA调控网络

目标基因NOS1相关基因的关系网络



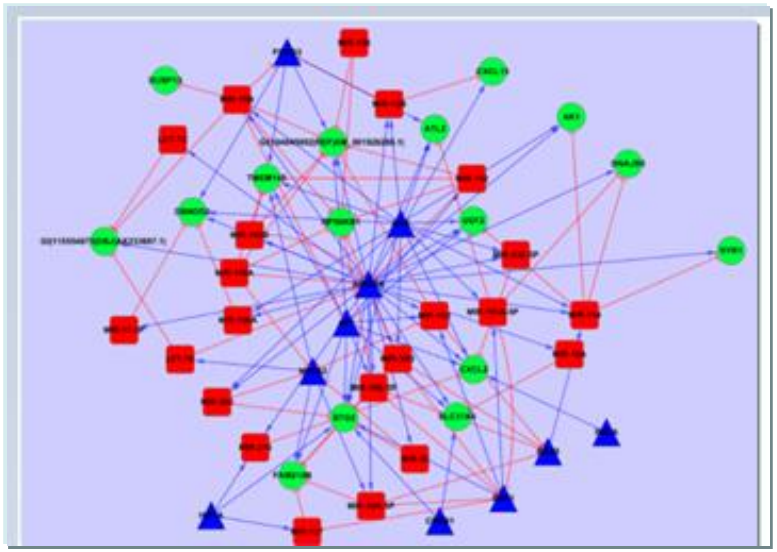
共表达模块中连通性排名前30的**核心基因**



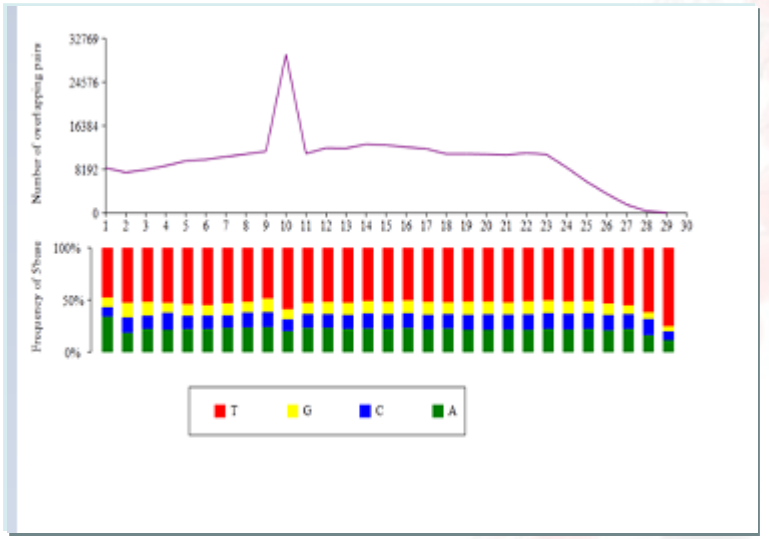
个性化图表

基迪奥生物
GENE DENOVO

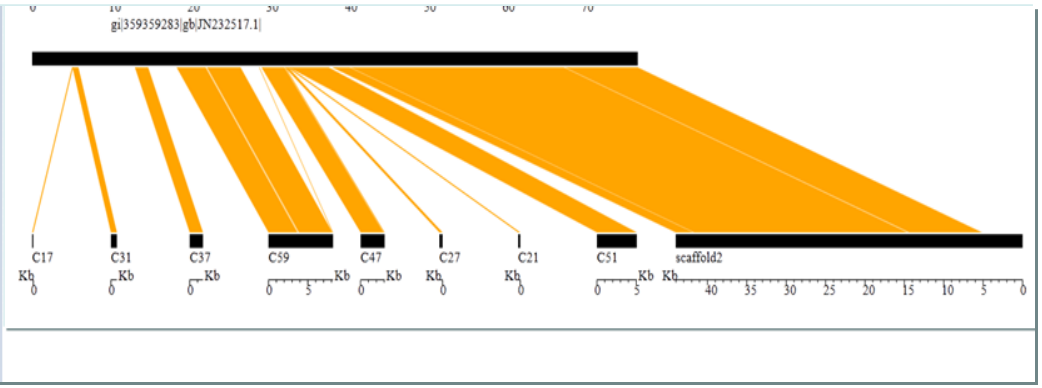
网络图



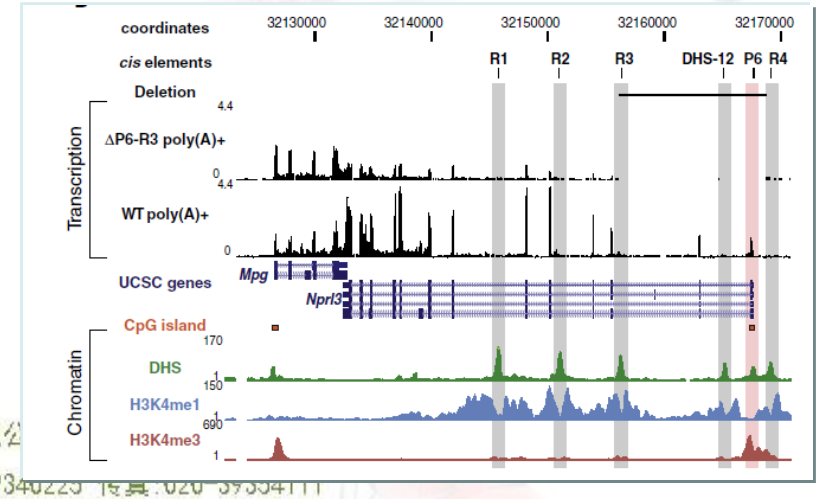
碱基统计



共线性图



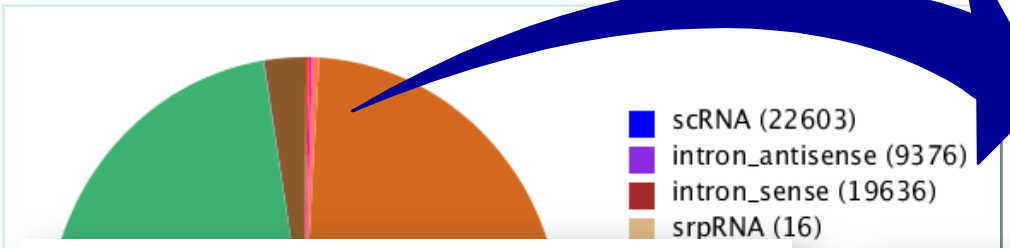
基因原件图



广州基迪奥生物科技有限公
邮箱:service@genedenovo.com 电话:020-39340223 传真:020-39334111

开发——小RNA数据的挖掘

- Novel miRNA
- Mirtron
- PiRNA



Arensburger et al. BMC Genomics 2011, 12:606
http://www.biomedcentral.com/1471-2164/12/606

BMC Genomics

RESEARCH ARTICLE Open Access

The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs

Peter Arensburger[1], Robert H Hice[1], Jennifer A Wright[1], Nancy L Craig[2] and Peter W Atkinson[1*]

Abstract

Background: The piRNA pathway has been shown in model organisms to be involved in silencing of transposons thereby providing genome stability. In *D. melanogaster* the majority of piRNAs map to these sequences. The medically important mosquito species *Aedes aegypti* has a large genome size, a high transposon load which includes Miniature Inverted repeat Transposable Elements (MITES) and an expansion of the piRNA biogenesis genes. Studies of transgenic lines of *Ae. aegypti* have indicated that introduced transposons are poorly remobilized and we sought to explore the basis of this. We wished to analyze the piRNA profile of *Ae. aegypti* and thereby determine if it is responsible for transposon silencing in this mosquito.

Results: Estimated piRNA sequence diversity was comparable between *Ae. aegypti* and *D. melanogaster*, but surprisingly only 19% of mosquito piRNAs mapped to transposons compared to 51% for *D. melanogaster*. *Ae. aegypti* piRNA clusters made up a larger percentage of the total genome than those of *D. melanogaster* but did not contain significantly higher percentages of transposon derived sequences than other regions of the genome. *Ae. aegypti* contains a number of protein coding genes that may be sources of piRNA biogenesis with two, *traffic jam* and *maelstrom*, implicated in this process in model organisms. Several genes of viral origin were also targeted by piRNAs. Examination of six mosquito libraries that had previously been transformed with transposon derived sequence revealed that new piRNA sequences had been generated to the transformed sequences, suggesting that they may have stimulated a transposon inactivation mechanism.

Conclusions: *Ae. aegypti* has a large piRNA complement that maps to transposons but primarily gene sequences, including many viral-derived sequences. This, together the more uniform distribution of piRNA clusters throughout its genome, suggest that some aspects of the piRNA system differ between *Ae. aegypti* and *D. melanogaster*.

Research

Discovery of hundreds of mirtrons in mouse and human small RNA data

Erik Ladewig,[1] Katsutomo Okamura,[1,2] Alex S. Flynt,[1] Jakub O. Westholm,[1] and Eric C. Lai[1,3]

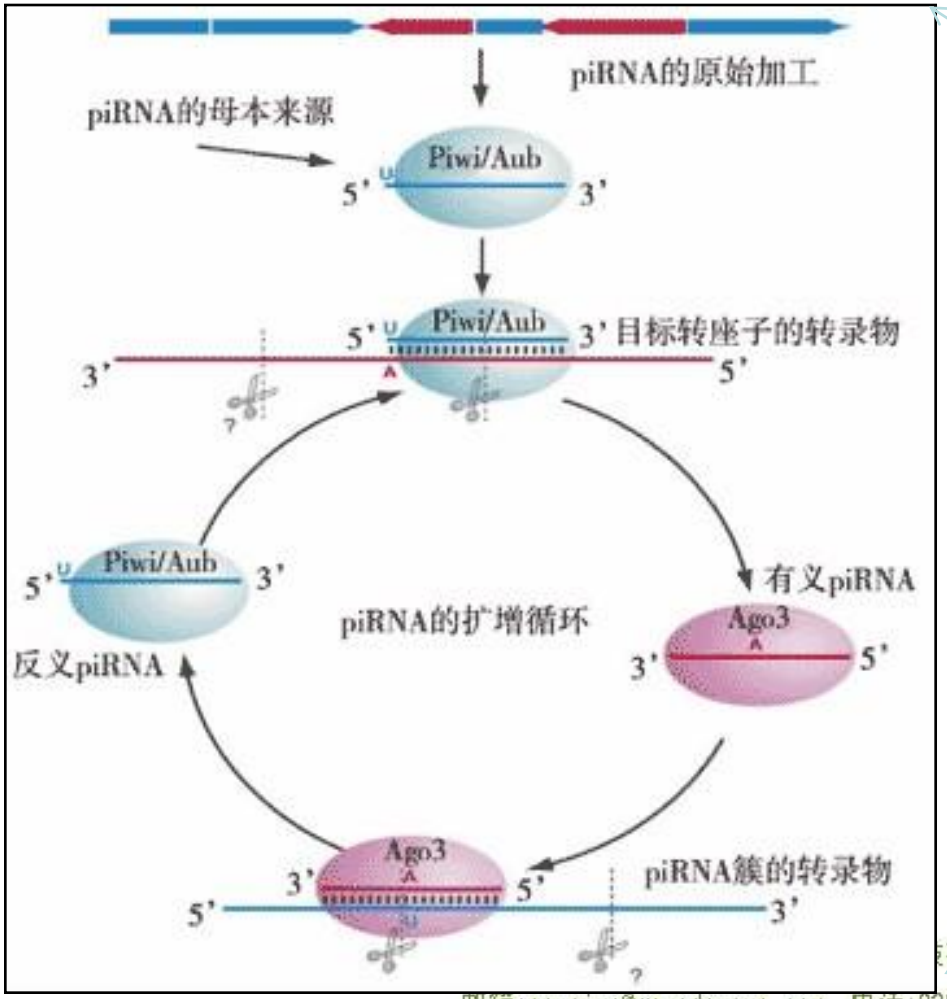
[1]*Department of Developmental Biology, Sloan-Kettering Institute, New York, New York 10065, USA;* [2]*Temasek Life Sciences Laboratory, National University of Singapore, Singapore 117604*

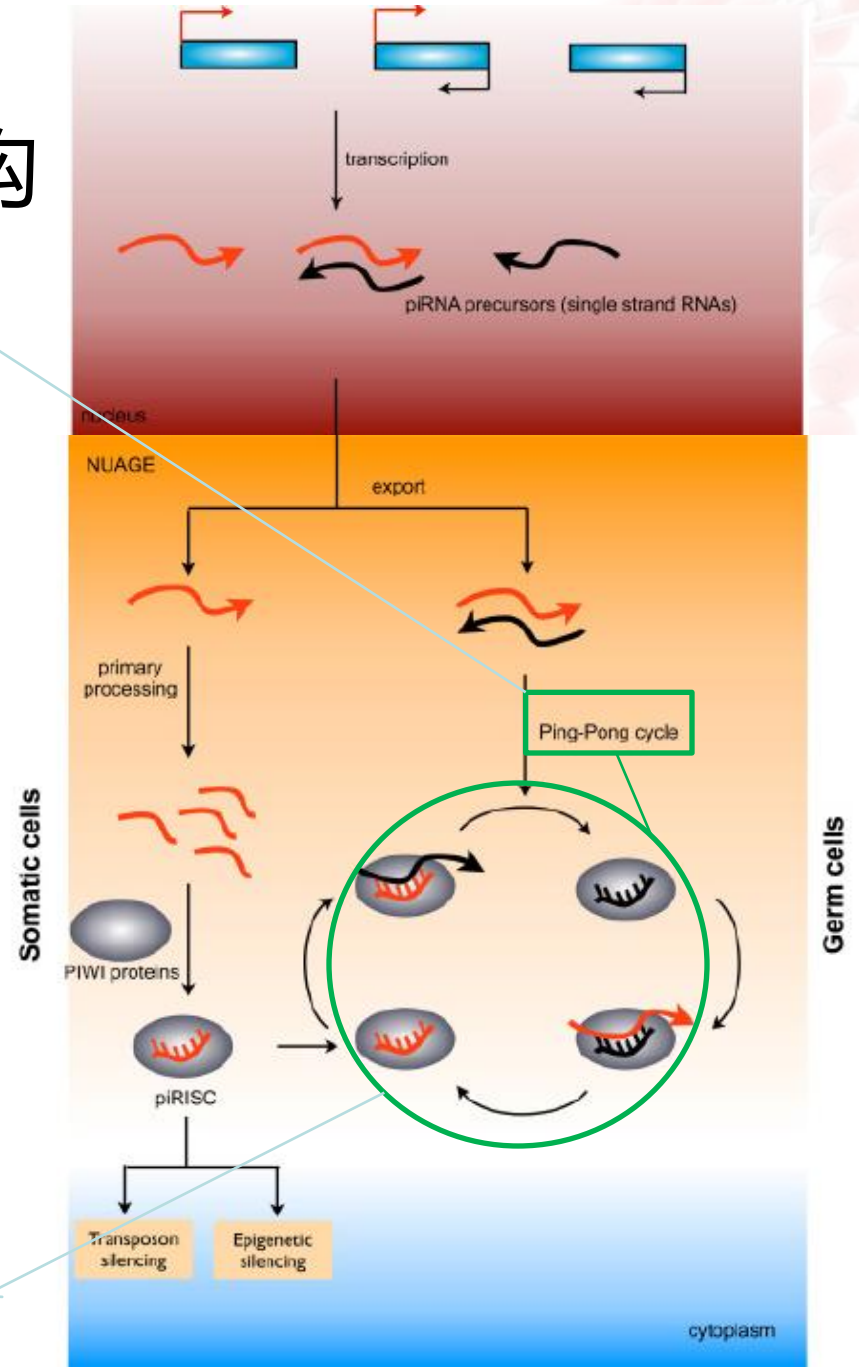
Atypical miRNA substrates do not fit criteria often used to annotate canonical miRNAs, and can escape the notice of miRNA genefinders. Recent analyses expanded the catalogs of invertebrate splicing-derived miRNAs ("mirtrons"), but only a few tens of mammalian mirtrons have been recognized to date. We performed meta-analysis of 737 mouse and human small RNA data sets comprising 2.83 billion raw reads. Using strict and conservative criteria, we provide confident annotation for 237 mouse and 240 human splicing-derived miRNAs, the vast majority of which are novel genes. These comprise three classes of splicing-derived miRNAs in mammals: conventional mirtrons, 5′-tailed mirtrons, and 3′-tailed mirtrons. In addition, we segregated several hundred additional human and mouse loci with candidate (and often compelling) evidence. Most of these loci arose relatively recently in their respective lineages. Nevertheless, some members in each of the three mirtron classes are conserved, indicating their incorporation into beneficial regulatory networks. We also provide the first Northern validation for mammalian mirtrons, and demonstrate Dicer-dependent association of mature miRNAs from all three classes of mirtrons with Ago2. The recognition of hundreds of mammalian mirtrons provides a new foundation for understanding the scope and evolutionary dynamics of Dicer substrates in mammals.

[Supplemental material is available for this article.]

基迪奥生物
GENE DENOVO

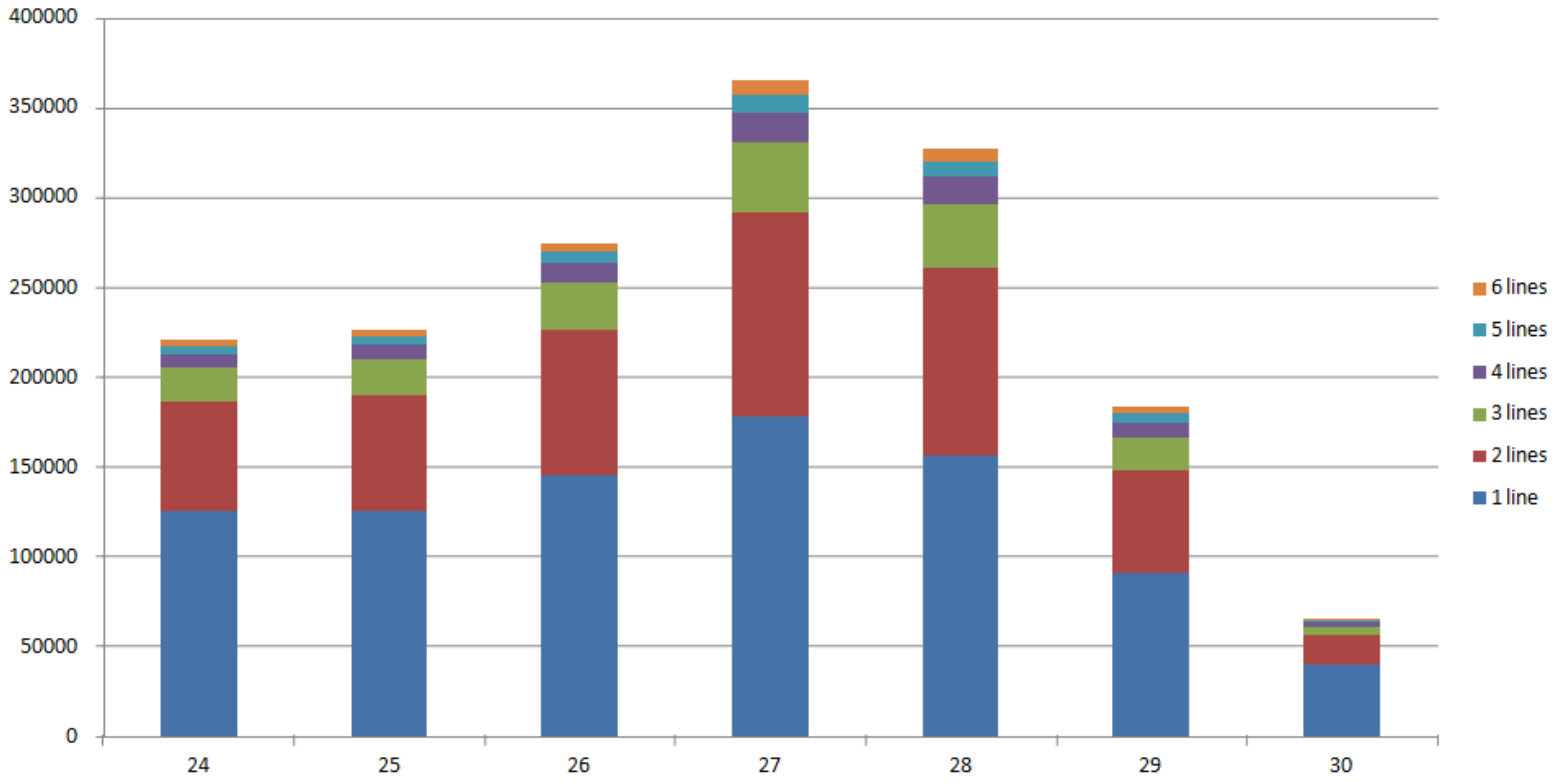
piRNA的ping-pong 结构





基迪奥生物
GENE DENOVO

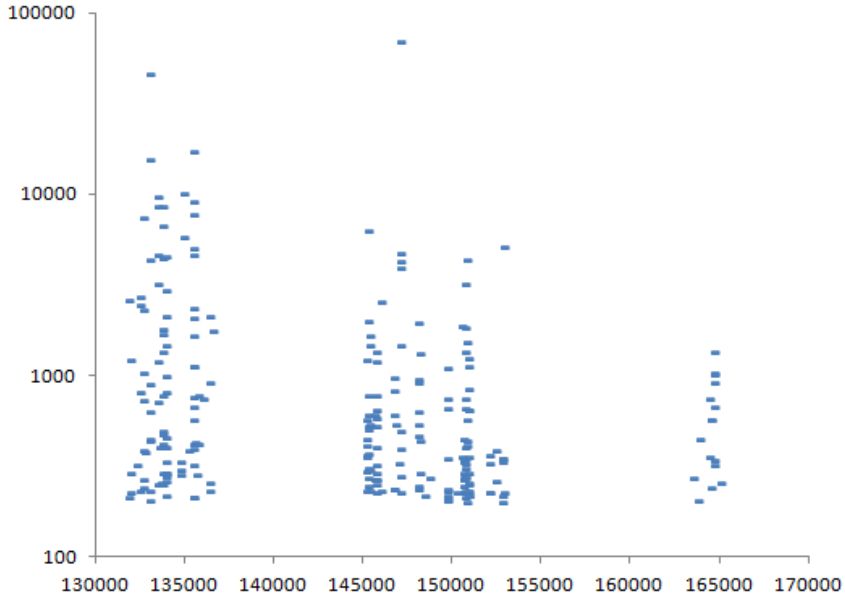
piRNA在6个样品中的分布图

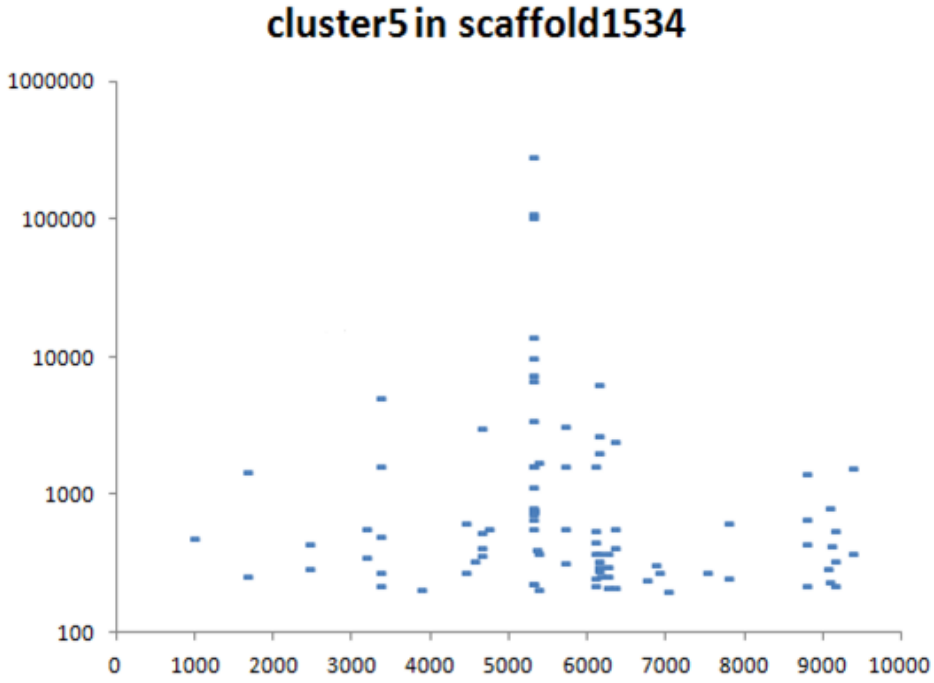


广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

基迪奥生物
GENE DENOVO

piRNA的成簇分布





广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

开发——实验室专有数据库

ADMEregulator

广东省医学科学院 Guangdong Academy of Medical Sciences　广东省人民医院 Guangdong General Hospital

Home | Search | Browse | Help | Search Result

Welcome to ADMEregulator

ADMEregulator now provides a Bayesian regulatory network for each ADME gene. For a given ADME gene, ADMEregulator able to:

- Search for the first level and the seconde level of causal transcript factors, miRNAs, and lncRNAs that regulate ADME gene expression.
- Visualize Bayesian regulatory network of an ADME gene.

ADME regulator count: 85 entries

Enter a ADME gene name or keyword:

CYP3A4

Search | Reset | Example

gene_name	eQTL_mark	first_type	first_relation	first_id	first_name	eQTL_mark1	first_cor	first_fdr	second_type	second_relation	second_id	second_name	eQTL_mark2	second_cor	second_fdr	second_gene_cor	second_gene_fdr
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000005889	ZFX	NA	0.753162393	0.008831 4	0.474871795	0.074797
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000068305	MEF2A	eQTL:rs11853151	0.815384615	0.008127	0.476923077	0.073248
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000069667	RORA	NA	0.635555556	0.040738	0.500854701	0.056406
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000134532	SOX5	NA	0.730598291	0.012351	0.405128205	0.14504
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000148516	ZEB1	NA	0.721709402	0.014076	0.483076923	0.068614
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000156273	BACH1	NA	0.797606838	0.006124 9	0.433162393	0.11282
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000164749	HNF4G	eQTL:rs1460249	0.692307692	0.020952	0.417435897	0.13025
CYP3A4	NA	lncRNA	relation	Unigene1886_All-lncRNA	NONHSAT122255	NA	0.699145299	0.017786	tf	tfbs	ENSG00000177932	ZNF354C	NA	0.792820513	0.005991 6	0.627350427	0.0099713
				Unigene1889	NONHSAT		0.60410	0.0537			ENSG0000			0.63487		0.4748717	

基迪奥生物
GENE DENOVO

项目成功的拼图

确立基石：

技术——选择合适的NGS技术

绘制蓝图

框架——定制化方案设计

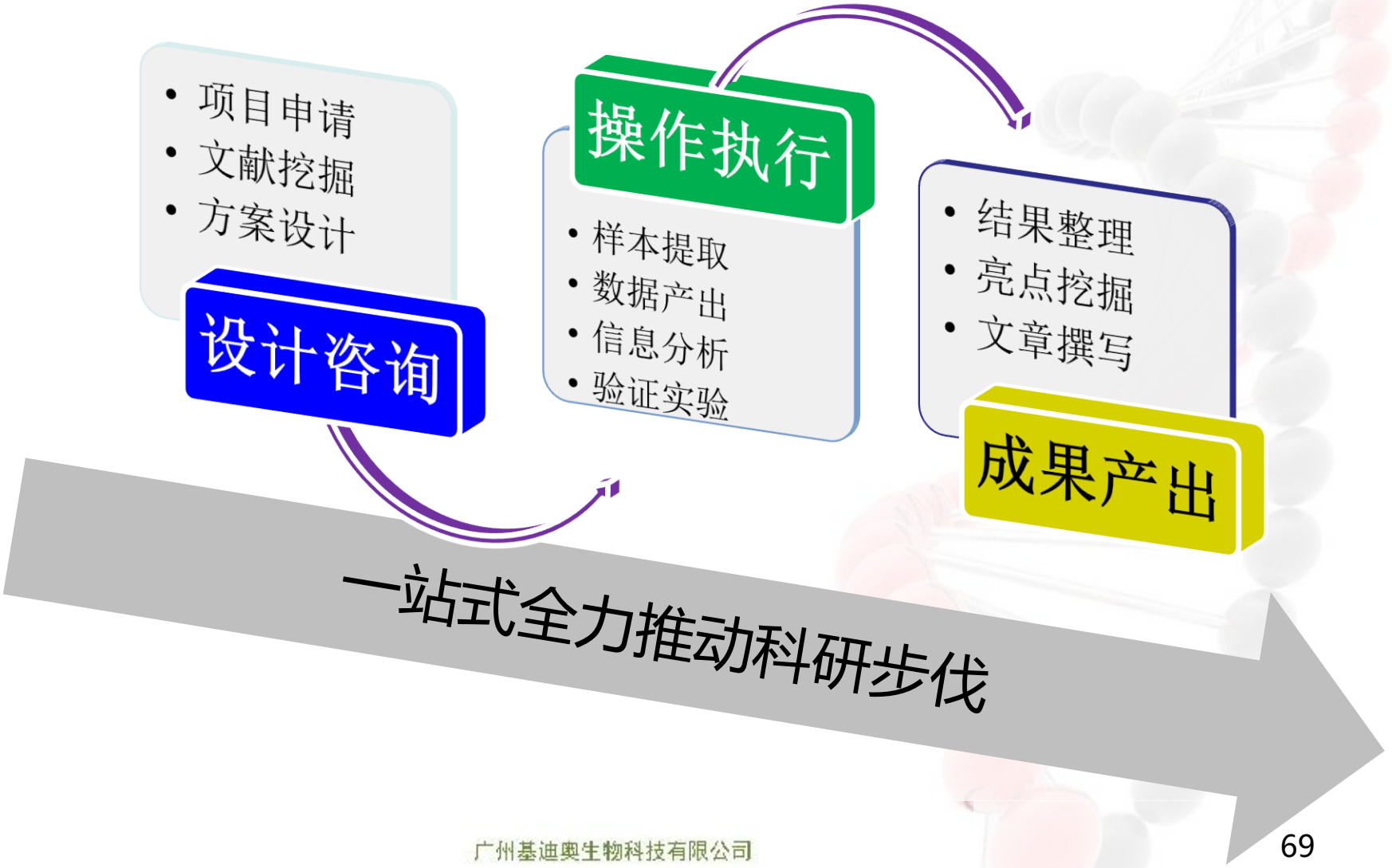
结果产出

细节——精细挖掘

广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

基迪奥生物
GENE DENOVO

基迪奥“一站式”服务



广州基迪奥生物科技有限公司
邮箱:service@genedenovo.com 电话:020-39340225 传真:020-39354111

基迪奥生物
GENE DENOVO

用我们的服务，加速您的研究！

谢谢！

service@genedenovo.com