CHIP-chip

报告人：彭颖

2010.11.5.

CHIP--Chromatin Immunoprecipitation

染色体免疫共沉淀

保持蛋白质和DNA联合的同时，通过运用对应于一个特定蛋白标记的生物抗体，染色质被切成很小的片断，并沉淀下来

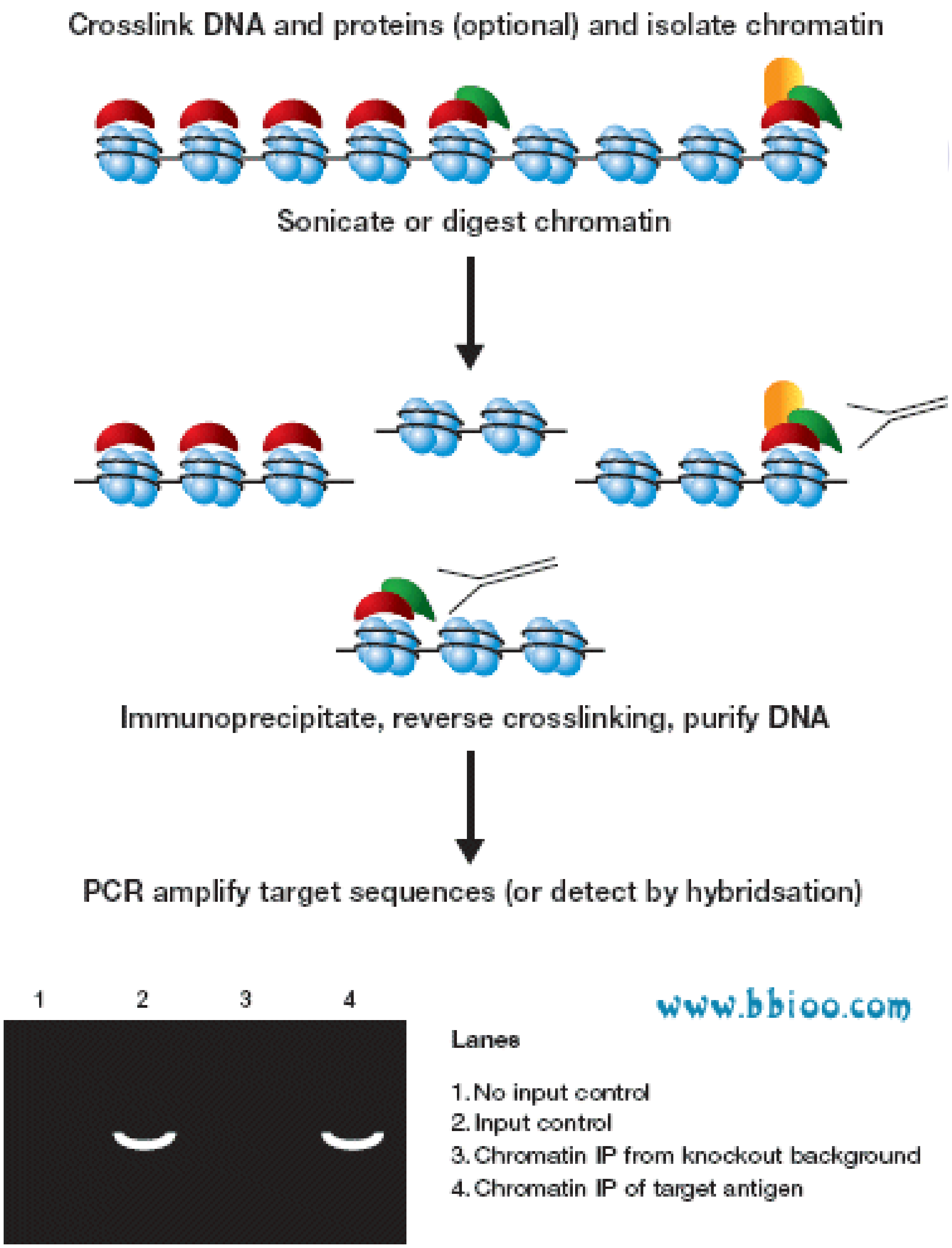
CHIP:

将全基因组内的

互作DNA

一网打尽！

In vivo



Crosslink DNA and proteins (optional) and isolate chromatin

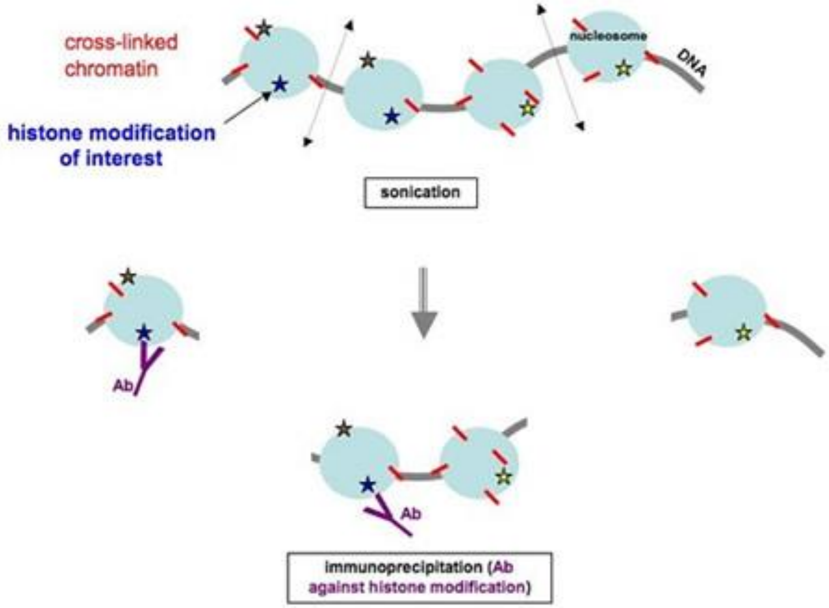
Sonicate or digest chromatin

Immunoprecipitate, reverse crosslinking, purify DNA

PCR amplify target sequences (or detect by hybridsation)

Lanes

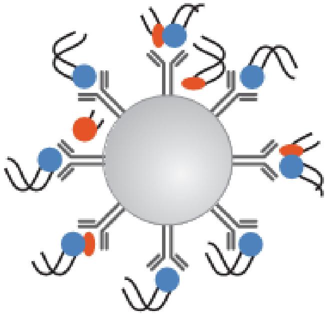
1. No input control
2. Input control
3. Chromatin IP from knockout background
4. Chromatin IP of target antigen



甲醛处理细胞

收集细胞，破碎

加入目的蛋白抗体，与复合物结合



加proteinA

清洗复合物

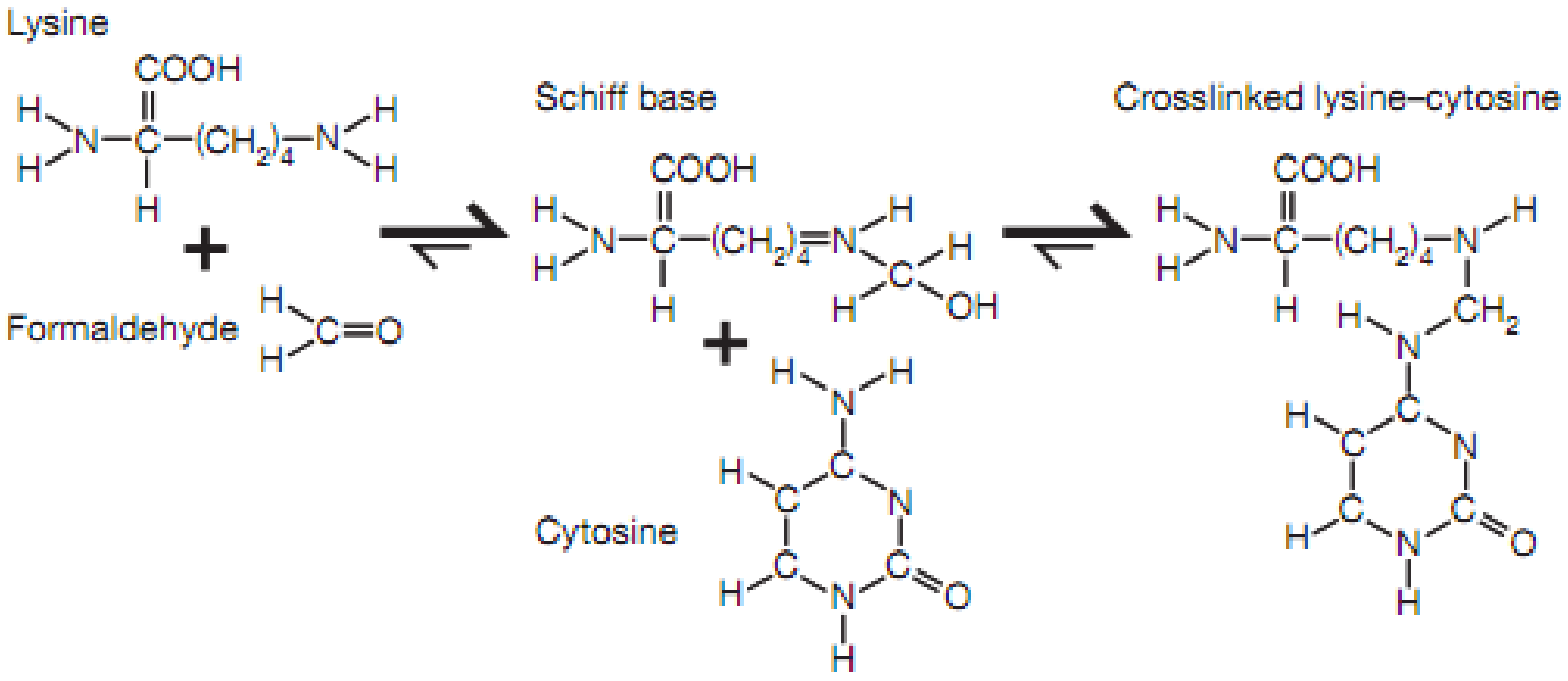
洗脱

解交联

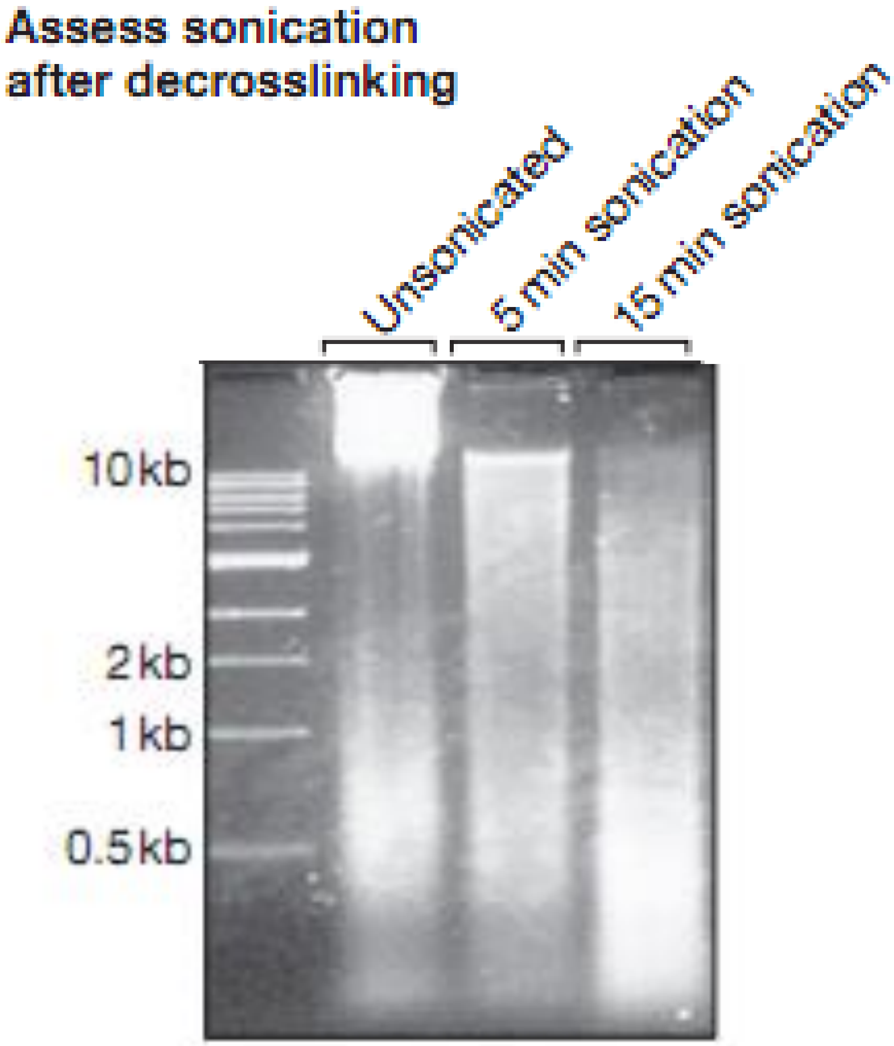
纯化富集DNA

交联

e.g.赖氨酸和胞嘧啶



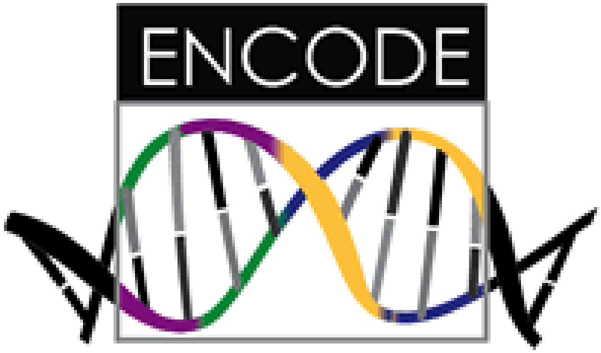
超声波破碎后对DNA进行凝胶检测



常用试剂盒

Thermo

Encyclopedia of DNA Elements



Funded by

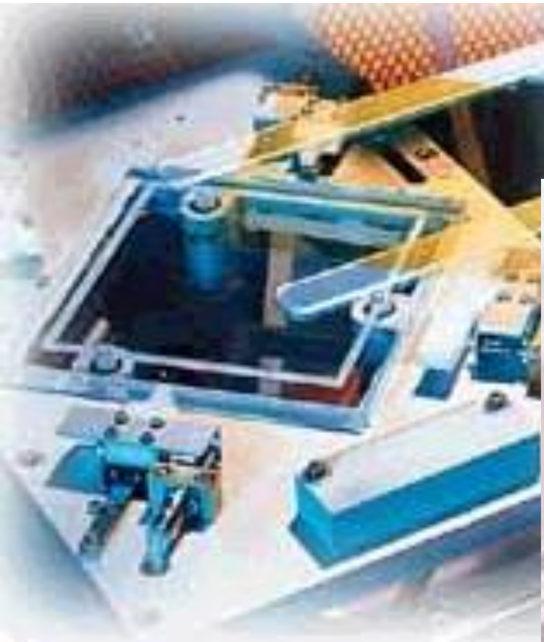
National Human Genome Research Institute (NHGRI)

Goal

build a comprehensive parts list of the functional elements of the human genome

Microarray

- 用经特殊处理的玻璃片（载玻片）作固体支持表面，将目标基因DNA排列在玻片表面并固定。
- 探针采用多色荧光标记。
- 其特点是，密度高、制作方便，是目前应用最广泛的基因芯片之一

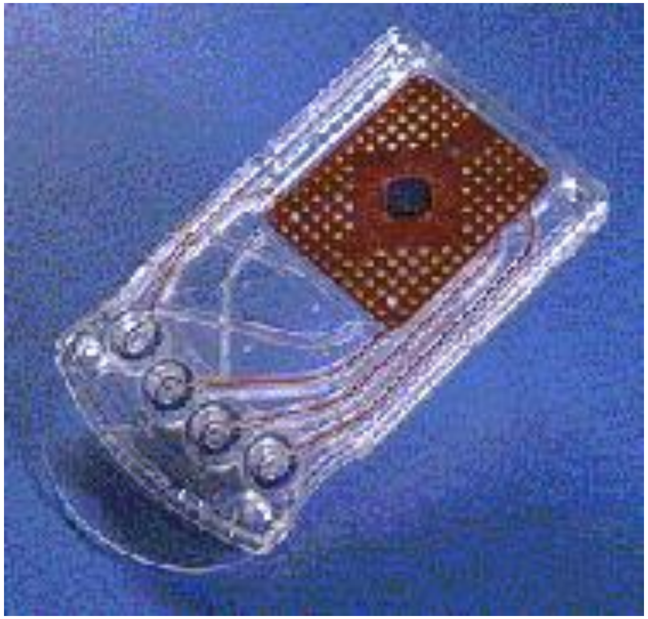


靶DNA原位合成

芯片组装

基因芯片

Lab on a chip



Nanogen公司

- 运用集成电路控制电极，使电极与目标DNA分子结合
- 与探针杂交
- 自动化程度高，具有极高的发展前景

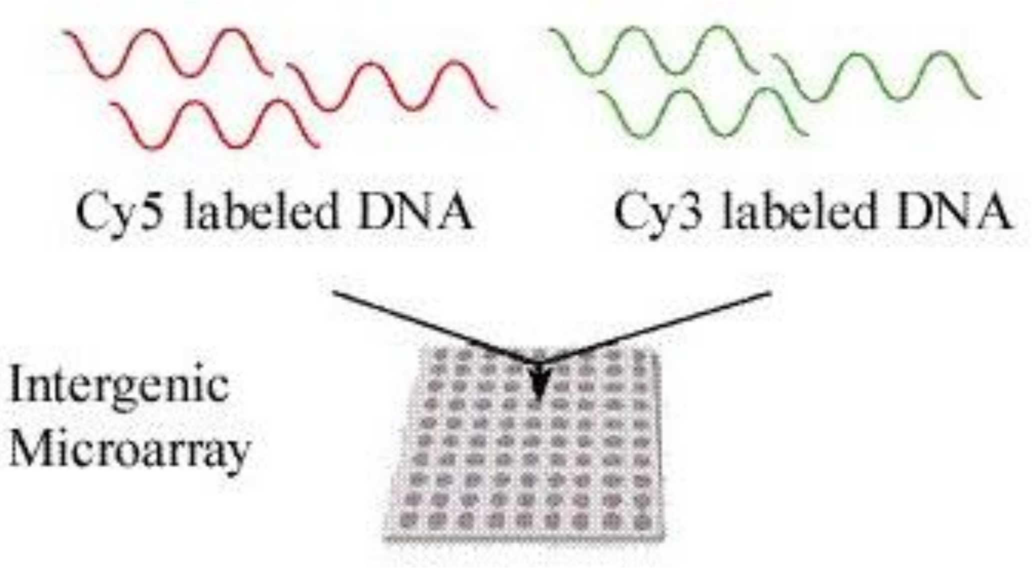
chip

探针

- before: 200bp
- now: 25-70bp

区域

- end-to-end
- spaced



Different kinds of arrays

Promoter array

Candidate region array

CpG island array

Encyclopedia of DNA Elements array

Non-repetitive sequence array (tilling array)

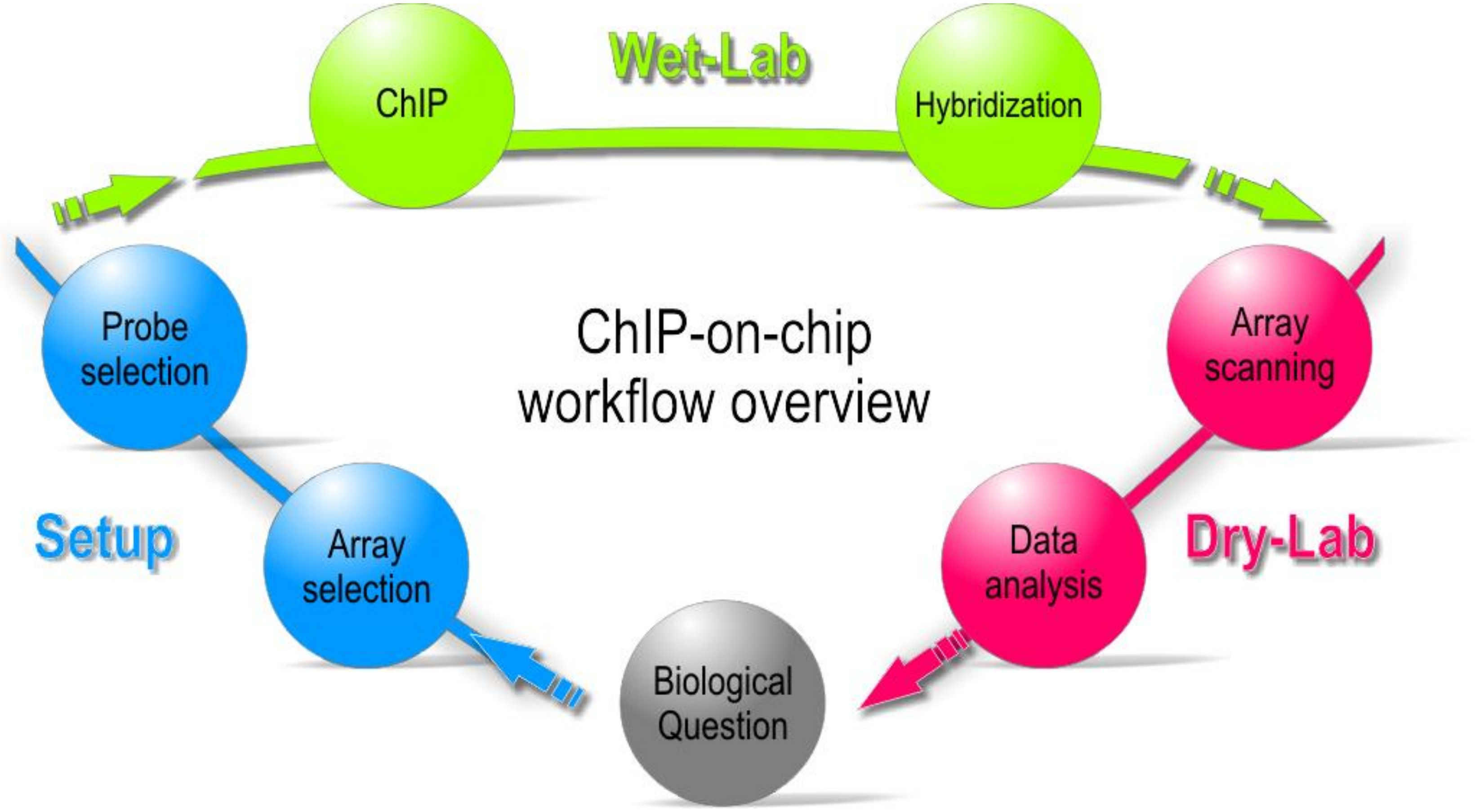
Affymetrix

the yeast genome： a resolution of 5bp (all in all 3.2 million probes).

the human genome： 90 million probes (Feb 2007)

芯片扫描仪器

生物信息与实验技术的结合


Wet-Lab

ChIP

Hybridization

Probe selection

ChIP-on-chip workflow overview

Array scanning

Setup

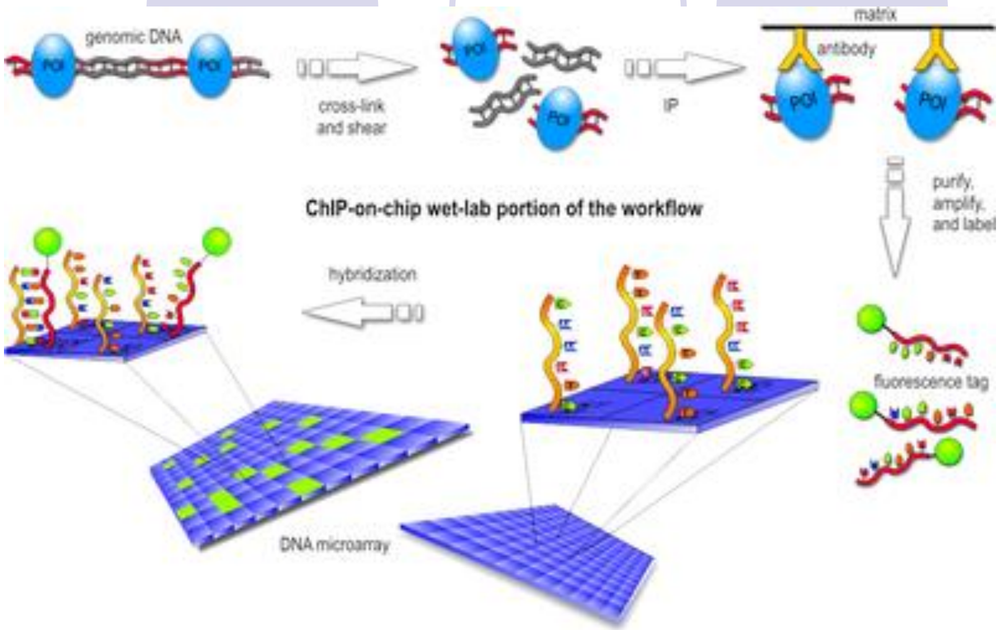
Array selection

Data analysis

Dry-Lab

Biological Question

CHIP-chip

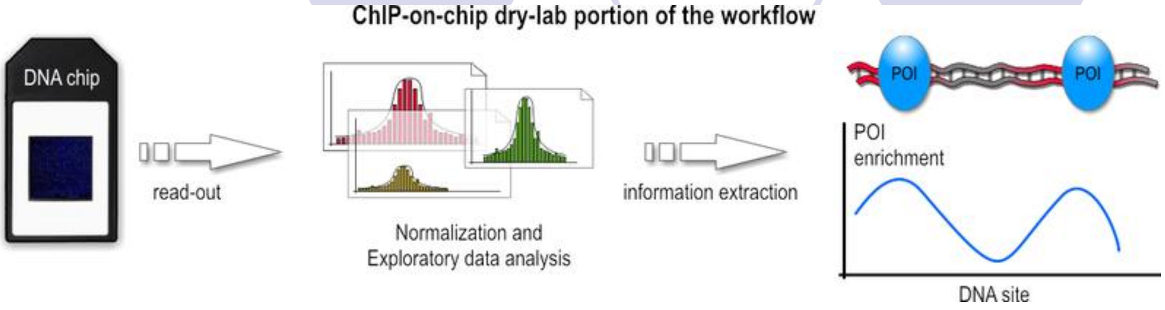


目前主要用于研究

与TF结合和条件特异性

组蛋白修饰、组蛋白修饰蛋白和染色体重建

信号处理



1.收集应该信号后，利用参照信号讲信息标准化，

扣除非特异性杂交

2.数值检测和统计测试，确认**POI**丰富区域

median percentil,single-array error,

sliding-window

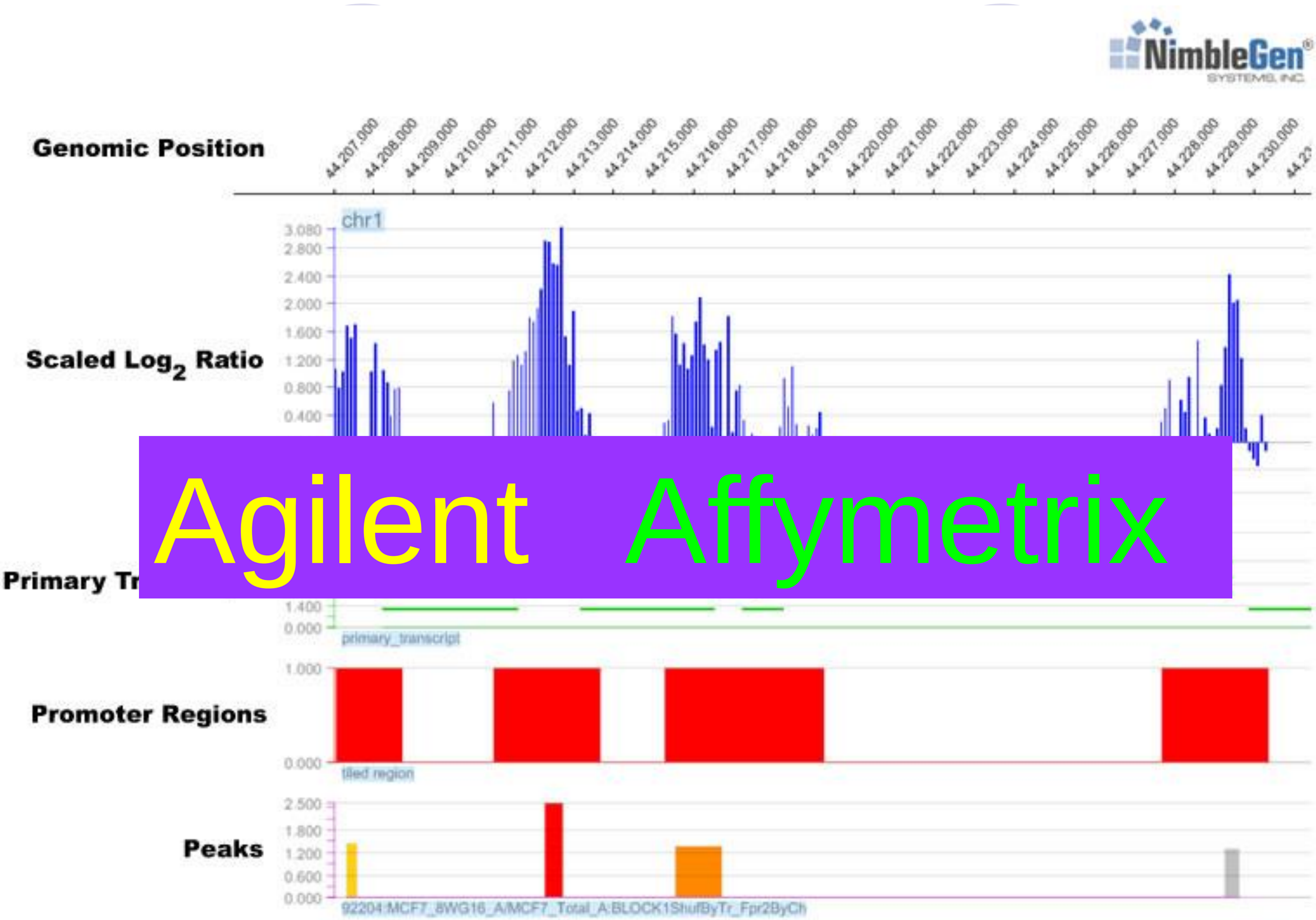
3.进一步分析

核酸结构、基因组中的功能解析

2-step paradigm

1.获得的数据中在结合部位抽取500-1000bp的序列信号

2.在这些序列中进行模式对比，获得粗糙的数据集，进行数据挖掘和分析



Agilent Affymetrix

芯片数据分析软件

Cluster 2.20	对大量微阵列数据组进行各种聚类（**Cluster)**分析与其它各种处理的软件。
TreeView 1.60	用图形来显示**Cluster**软件分析的结果，用来发表。
cluster 3.0 10.04.05	日本学者编写的**cluster**软件的增强版，增加了**K-Means**算法，以寻找最佳聚类结果。
ScanAlyze 2.51	进行微阵列荧光图像分析，包括半自动定义格栅与像素点分析。输出为分隔的文本格式，可很容易地转化为任何数据库。
数个芯片数据分析小程序	**standford**大学一位教授写的数个芯片数据分析小程序。包括**AACK : Add Annotation**；**CCACK : Constant Cutoff Analysis**；**FLICK : File Linker**；**FRICK : Filter/Retrieve IDs**；**GACK : Genomotyping Analysis**；**GODACK : Good Data**；**HIMACK : Histogram Maker**；**LACK : Lexical Analysis**；**ALACK: Automated LACK**；**NACK : Name Averaging**；**REDUCK : Remove Duplicates**；**Samster : SAM to Cluster**；**COP : Competition Plotter**；**LD50 Calculators**；**DRACK : Differential Restriction Analysis**；**FOCK : Frequency of Oligos**；**LOCK : Locater of Oligos**；**Motif Search**，还包括源程序。

芯片数据分析软件

ArrayMaker 2.6	美国stanford大学Brown实验室提供的基因芯片研究全套设备相配套的软件与文档。
J-Express pro 2009	是一个用JAVA语言写的应用程序，用来分析微阵列（ microarray ）实验获得的基因表达数据。
MAExplorer Version 0.96.34.01	是一个基于JAVA程序的微阵列数据库数据采集软件，用来帮助发现和癌症、 疾病相关的调节基因，可进行各种微阵列数据分析工作，直接链接到网上的各种基因組数据库。
FreeView和FreeOView	基于JAVA语言的系统树生成软件，接收Cluster生成的数据，比Treeview增强了某些功能。
dchip 08.07.12	DNA芯片分析软件，似乎是两个华人开发的软件，最好在WIN2000下运行。
SAM 3.09C	Significance Analysis of Microarrays 的缩写，微阵列显著性分析软件，EXCEL软件的插件，Stanford大学编制。
BRB-ArrayTools 4.1.0 b2	显示和统计分析DNA微阵列基因表达数据软件包。
AMADA 2.0.7	是Analysis of Microarray Data的缩写，是一个集成软件，用来組织、 研究、 显示、 分析微阵列(Microarray)数据。
GalFileGenerator v1.2	数据转换软件，将96孔板文件转换为384孔板文件，将384孔板文件转换为GAL文件。
Array-A-Lizer 1.03	DNA阵列实验结果快速分析软件。在对数据进行进一步分析前，用来快速确定DNA阵列实验数据的质量。

芯片数据分析软件

GenePattern 3.2.3	MIT学者用JAVA语言编写的表达分析高级生物医学分析软件平台。
GenMAPP 2.1	将基因表达微阵列数据进行分析并图形化显示软件。
MADAM v4.0	MicroArray DAta Manager的缩写，微阵列数据管理器，用来将数据输入为交互数据库格式，并对数据进行管理。
Spotfinder v3.2.1	快速分析微阵列图像工具软件，定量化基因表达结果并输出为TAV格式、EXCEL格式或数据库格式。
MIDAS 2.22	微阵列数据分析系统，Microarray Data Analysis System的缩写，对各种微阵列数据进行标准格式转换，输出格式为TAV格式。
MeV 4.6	MultiExperiment Viewer的缩写， 通用微阵列分析工具，运用各种算法对格式化好的微阵列数据进行聚类、统计、显示、分析。
TreeArrange	对微阵列表达数据进行重新排序和聚类分析软件。如果无法下载，请自行到原始网站注册下载。
Java Treeview 1.1.5r2	一个开放源代码的微阵列数据显示软件。Java语言写成。
EXPANDER 5.1	(EXpression Analyzer and DisplayER)的缩写。基因表达与显示软件包。用JAVA语言写成，可对微阵列数据进行聚类分析、显示及处理。
OligoArray 2.1	用于基因組规模微阵列的寡核苷酸探针设计软件。
BAGEL 3.62	Bayesian Analysis of Gene Expression Levels的缩写，微阵列数据统计分析软件。
CAGED 1.1	是Cluster Analysis of Gene Expression Dynamics的缩写，基因表达数据分析软件。
GAAS	是Gene Array Analyzer Software的缩写，大规模基因表达数据分析、显示与储存软件。
GLR	微阵列数据统计分析软件。
ANOVA 2.0	微阵列基因表达数据统计软件。
Rosetta	基因表达数据分析系统。

芯片数据分析软件

UCSF Spot	另一个微阵列扫描图像分析软件。
ArrayMiner 5.3简易版	微阵列数据分析软件的简易版，使用有限制。
Venn Mapper 1.01	比较非同源微阵列数据软件。
MAIA 2.75 Demo	微阵列图像分析软件。
MAGIC Tool 2.1	微阵列数据图像分析软件。
Array Tools 3.3	微阵列数据处理excel插件
GENOM 2005 3.40	分析微阵列实验数据软件。
Mayday 2.10	显示、分析、储存微阵列数据的软件平台
ChipInspector 2.1	芯片微阵列数据分析软件
OligoWiz 2.1.3	DNA微阵列寡核苷酸设计的客户端软件
ENIGMA 1.1	从扰动微阵列数据分析基因表达模块软件。JAVA语言写成，运行需要java环境。
PermutMatrix 1.9.3	基因表达微阵列数据分析软件。
KegArray 1.2.3	微阵列数据分析软件。
Correlate 1.00	Excel插件，用来进行两組基因組数据組稀疏典型相关性统计分析。
Picky 2.14 demo	Oligo Microarray设计和分析软件。
ParaSAM	并行微阵列数据显著性分析软件。
VIZARD 1.2	分析和显示Affymetrix基因芯片数据软件 。
COCAS V2.4	染色质免疫沉淀基因芯片分析ChIP-on-chip软件包 。

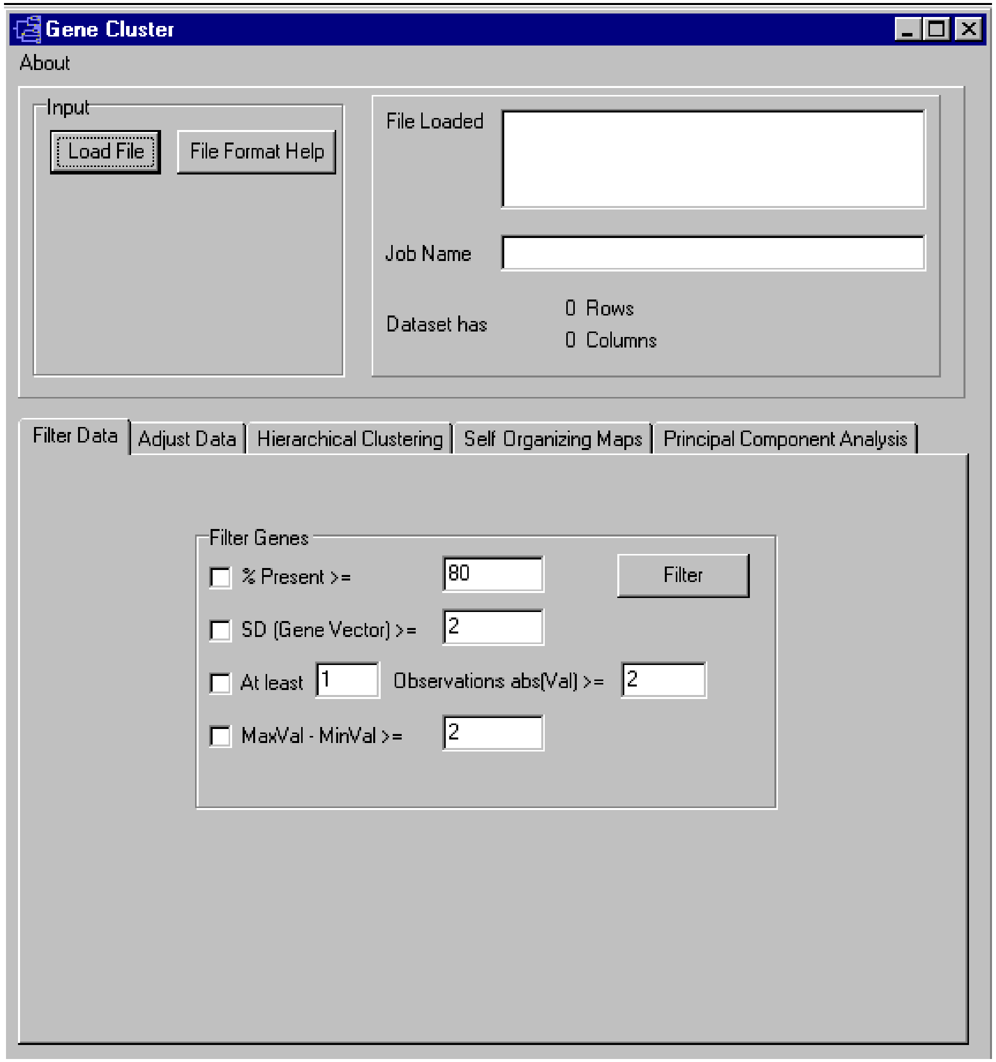
CHIP-chip数据处理软件举例

Cluster and TreeView

The program Cluster organizes and analyzes the data in a number of different ways. （cdt,gtr)

TreeView allows the organized data to be visualized and browsed.

Cluster操作窗口



File Loaded

Job Name

Dataset has
0 Rows
0 Columns

Filter Genes

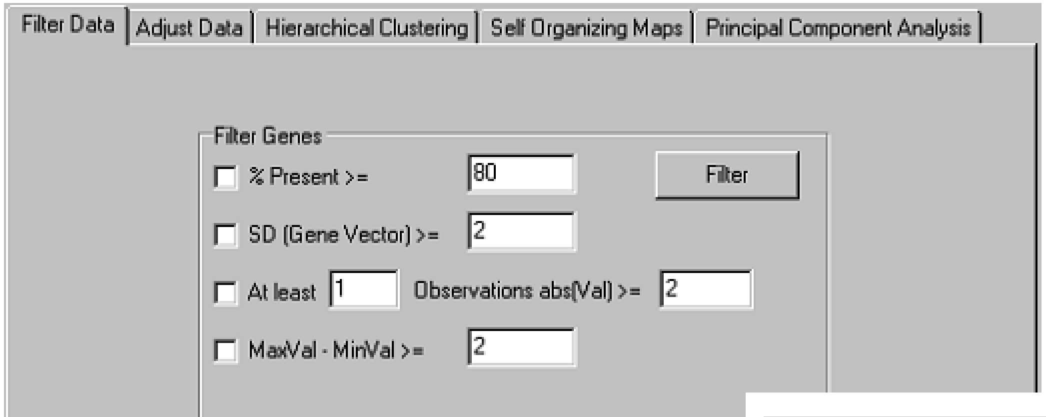
☐ % Present >= 80

☐ SD (Gene Vector) >= 2

☐ At least 1 Observations abs(Val) >= 2

☐ MaxVal - MinVal >= 2

Filtering Data



Filter Genes

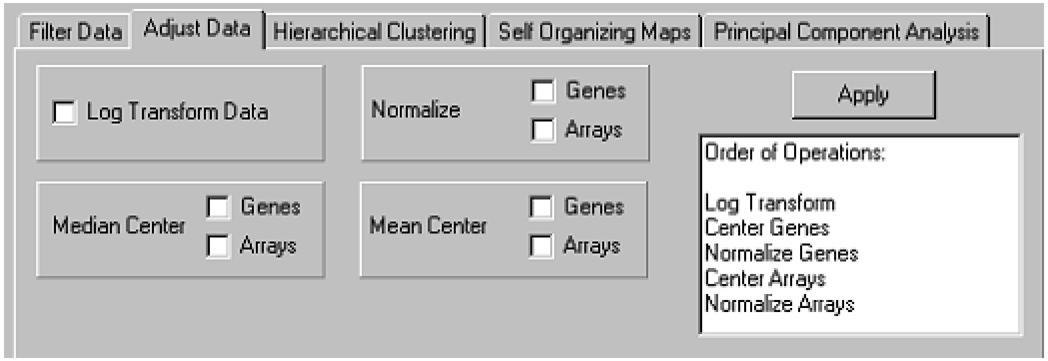
☐ % Present >= 80

☐ SD (Gene Vector) >= 2

☐ At least 1 Observations abs(Val) >= 2

☐ MaxVal - MinVal >= 2

Adjust Data



☐ Log Transform Data

Normalize ☐ Genes ☐ Arrays

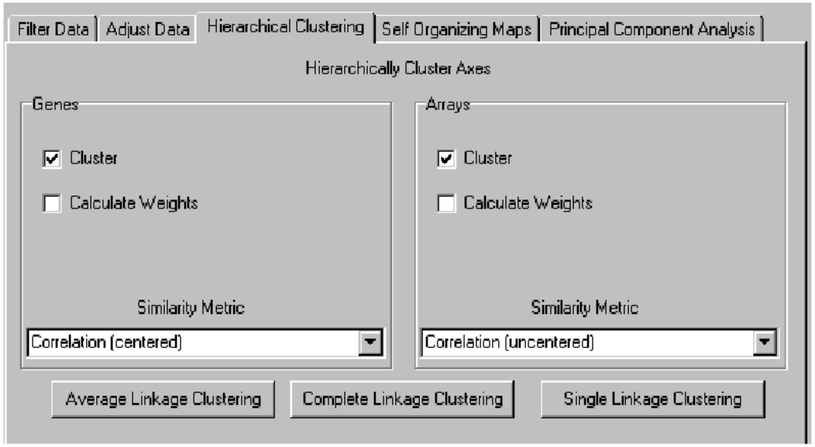
Median Center ☐ Genes ☐ Arrays

Mean Center ☐ Genes ☐ Arrays

Order of Operations:

Log Transform
Center Genes
Normalize Genes
Center Arrays
Normalize Arrays

Hierarchical Clustering



Hierarchically Cluster Axes

Genes

☑ Cluster

☐ Calculate Weights

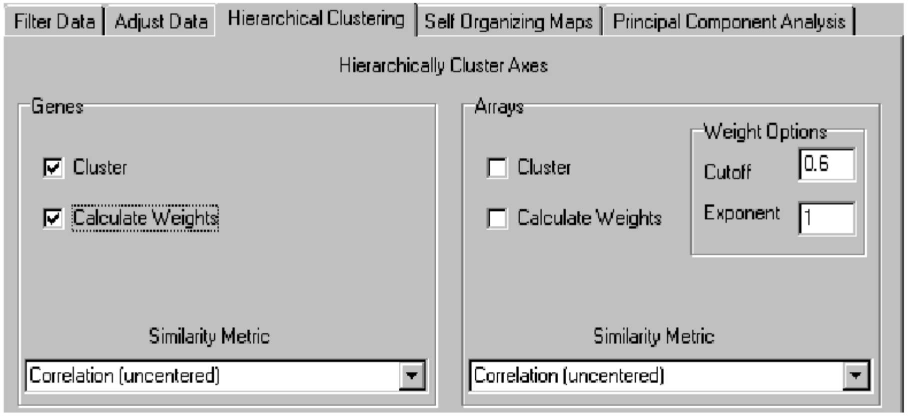
Similarity Metric: Correlation (centered)

Arrays

☑ Cluster

☐ Calculate Weights

Similarity Metric: Correlation (uncentered)



Hierarchically Cluster Axes

Genes

☑ Cluster

☑ Calculate Weights

Similarity Metric: Correlation (uncentered)

Arrays

☐ Cluster

☐ Calculate Weights

Weight Options

Cutoff 0.6

Exponent 1

Similarity Metric: Correlation (uncentered)

K-mean Clustering (new)

☑ Organize Genes

10 K (number of

100 Max Cycles

○ KMeans

● KMediods

● KMediods

各聚类本身尽可能紧凑，

各聚类之间尽量分开

从N个数据对象任意选择k

中心；

象则

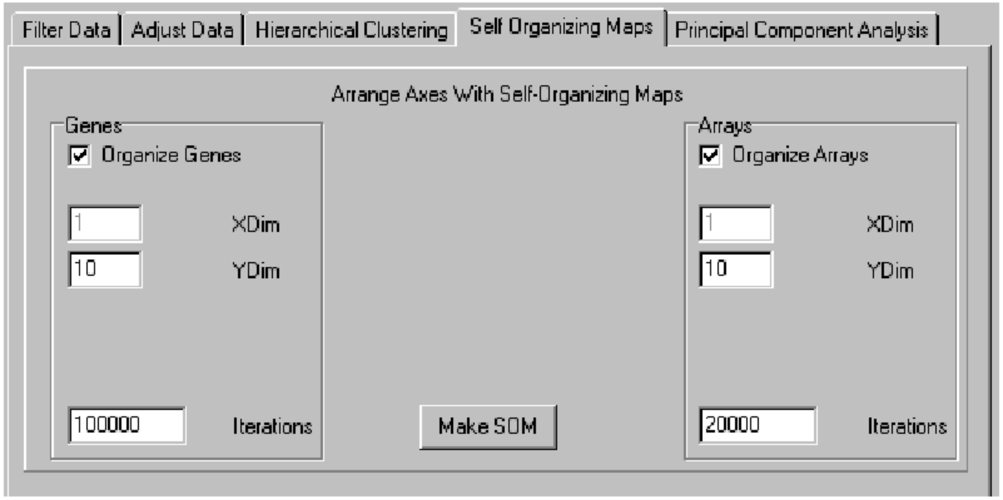
似度

相似

个聚

类中所有对象的均值，不断重复知道标准测度函数开始收敛

Self-Organizing Maps



Genes

☑ Organize Genes

1 XDim

10 YDim

100000 Iterations

Arrays

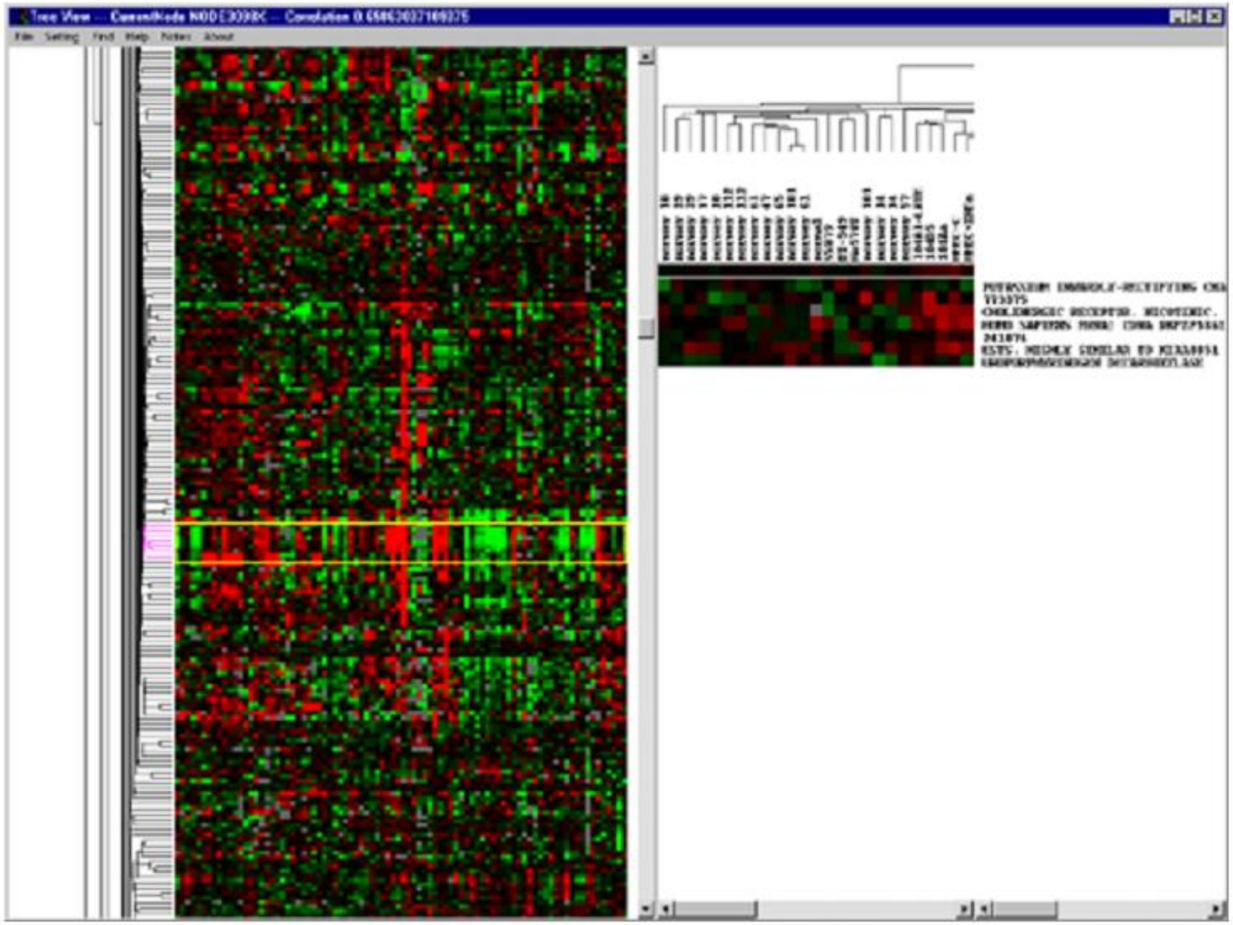
☑ Organize Arrays

1 XDim

10 YDim

20000 Iterations

TreeView



CHIP-chip数据处理软件举例

ChIP-on-chip Analysis Suite （CoCAS）

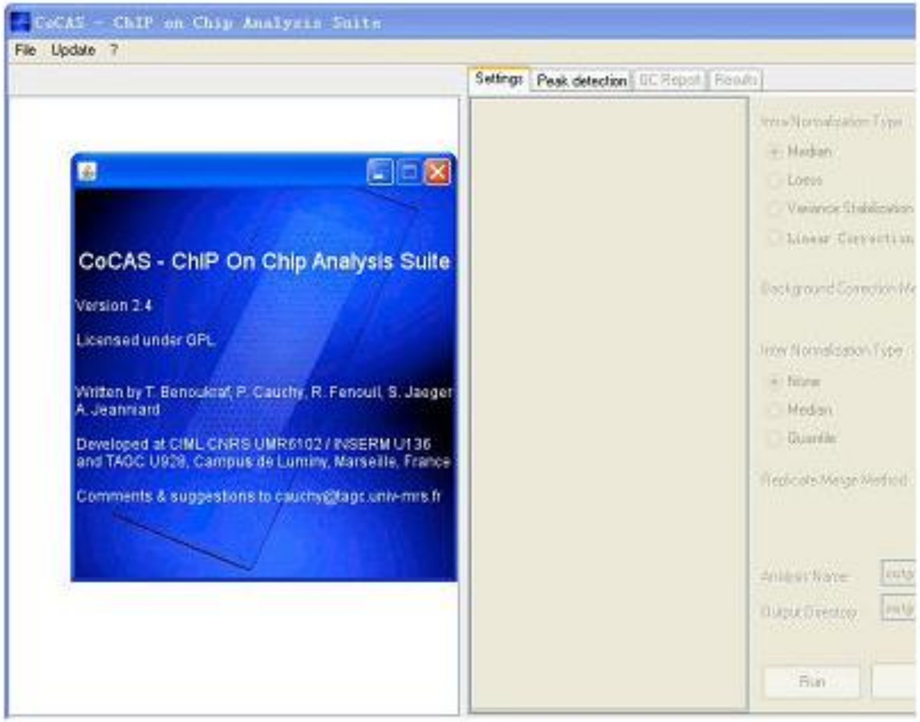
Data normalization

Peak detection

Quality control reports

Windows XP/Vista,

Mac OSX, Linux



CoCAS - ChIP On Chip Analysis Suite

Version 2.4

Licensed under GPL

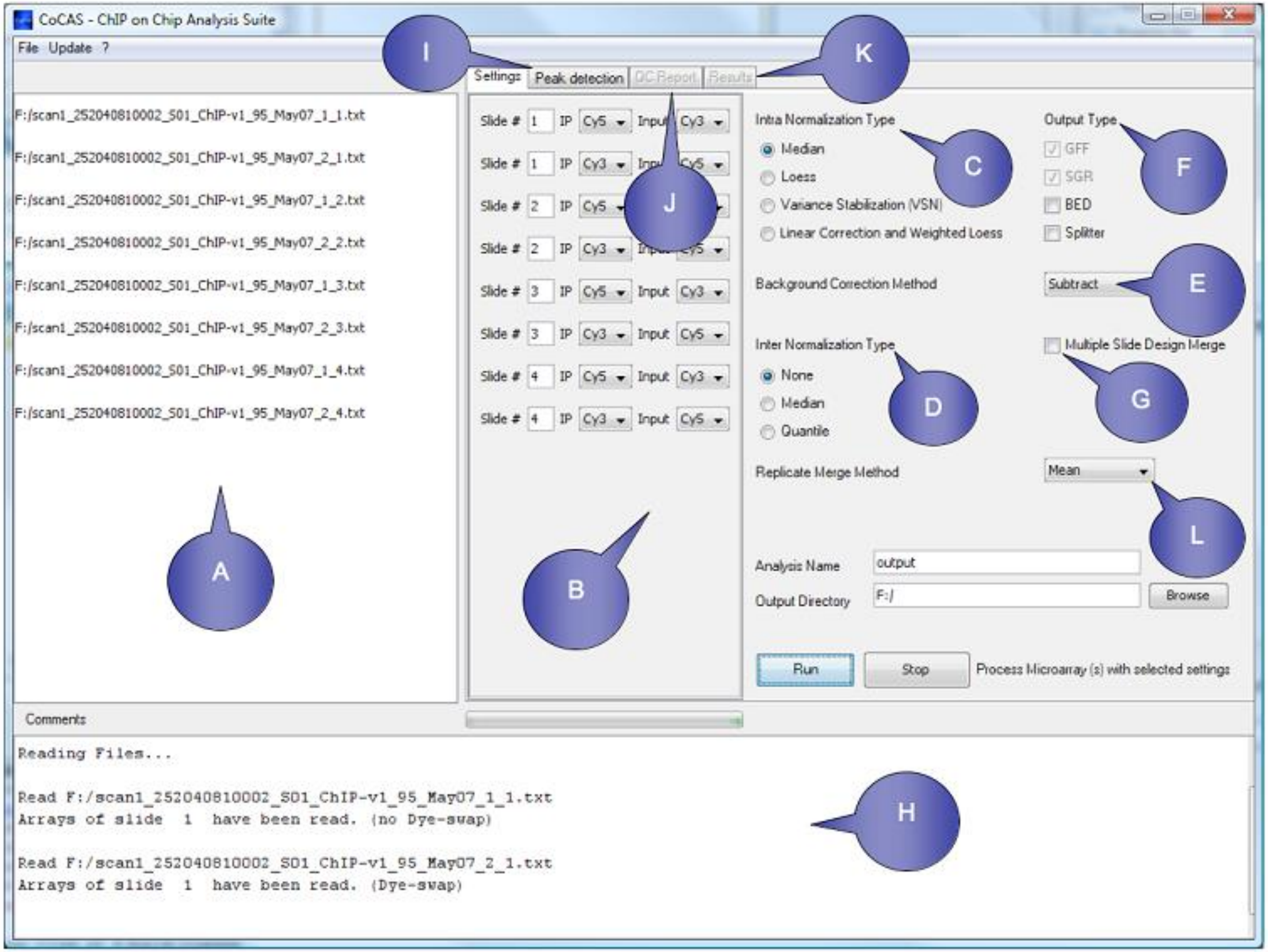
Written by T. Benoukraf, P. Cauchy, R. Fenouil, S. Jaeger
A. Jeanniard

Developed at CIML CNRS UMR6102 / INSERM U136
and TAGC U928, Campus de Luminy, Marseille, France

Comments & suggestions to cauchy@tagc.univ-mrs.fr

Java and R programming languages,notably BioConductor

CoCAS操作窗口



F:/scan1_252040810002_S01_ChIP-v1_95_May07_1_1.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_2_1.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_1_2.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_2_2.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_1_3.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_2_3.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_1_4.txt
F:/scan1_252040810002_S01_ChIP-v1_95_May07_2_4.txt

Slide # 1 IP Cy5 Input Cy3
Slide # 1 IP Cy3 Input Cy5
Slide # 2 IP Cy5
Slide # 2 IP Cy3 Input Cy5
Slide # 3 IP Cy5 Input Cy3
Slide # 3 IP Cy3 Input Cy5
Slide # 4 IP Cy5 Input Cy3
Slide # 4 IP Cy3 Input Cy5

Intra Normalization Type
● Median
○ Loess
○ Variance Stabilization (VSN)
○ Linear Correction and Weighted Loess

Output Type
☑ GFF
☑ SGR
☐ BED
☐ Splitter

Background Correction Method: Subtract

Inter Normalization Type
● None
○ Median
○ Quantile

☐ Multiple Slide Design Merge

Replicate Merge Method: Mean

Analysis Name: output
Output Directory: F:/

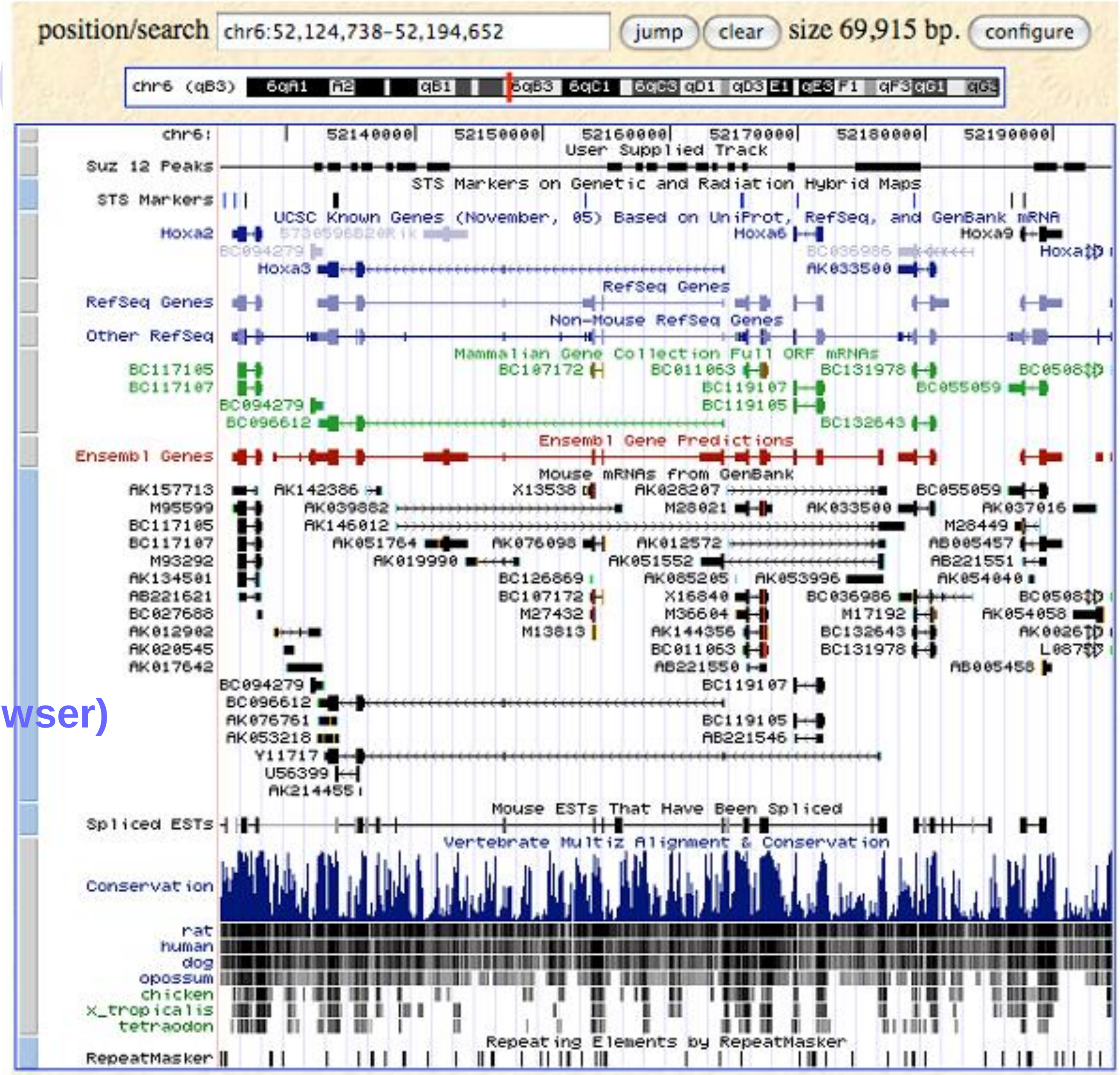
Process Microarray (s) with selected settings

```
Reading Files...

Read F:/scan1_252040810002_S01_ChIP-v1_95_May07_1_1.txt
Arrays of slide  1  have been read. (no Dye-swap)

Read F:/scan1_252040810002_S01_ChIP-v1_95_May07_2_1.txt
Arrays of slide  1  have been read. (Dye-swap)
```

CoCAS


position/search chr6:52,124,738-52,194,652 size 69,915 bp.

Connection to

Genome Browsers

(UCSC Genome Browser)

CoCAS分析结果

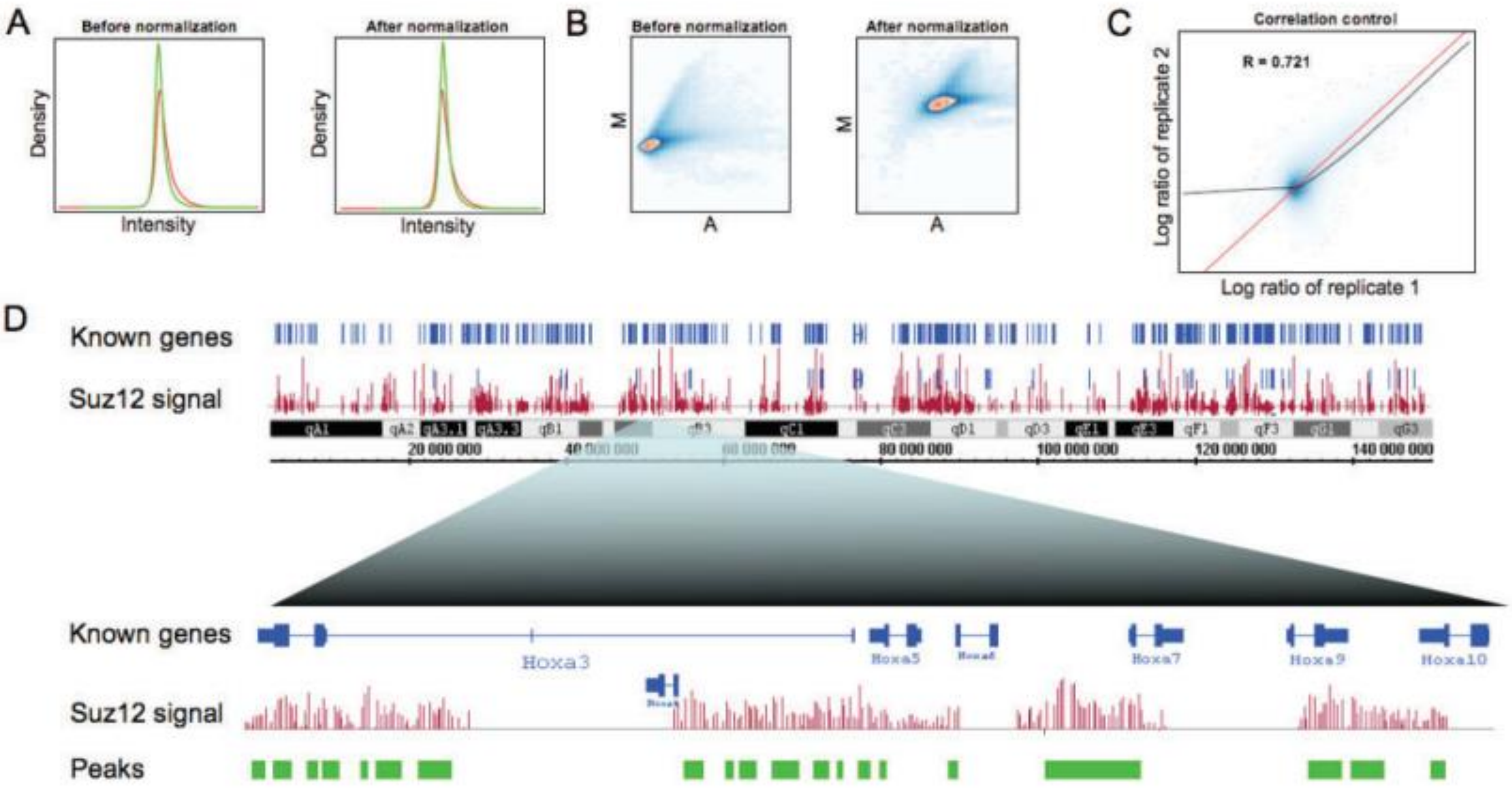


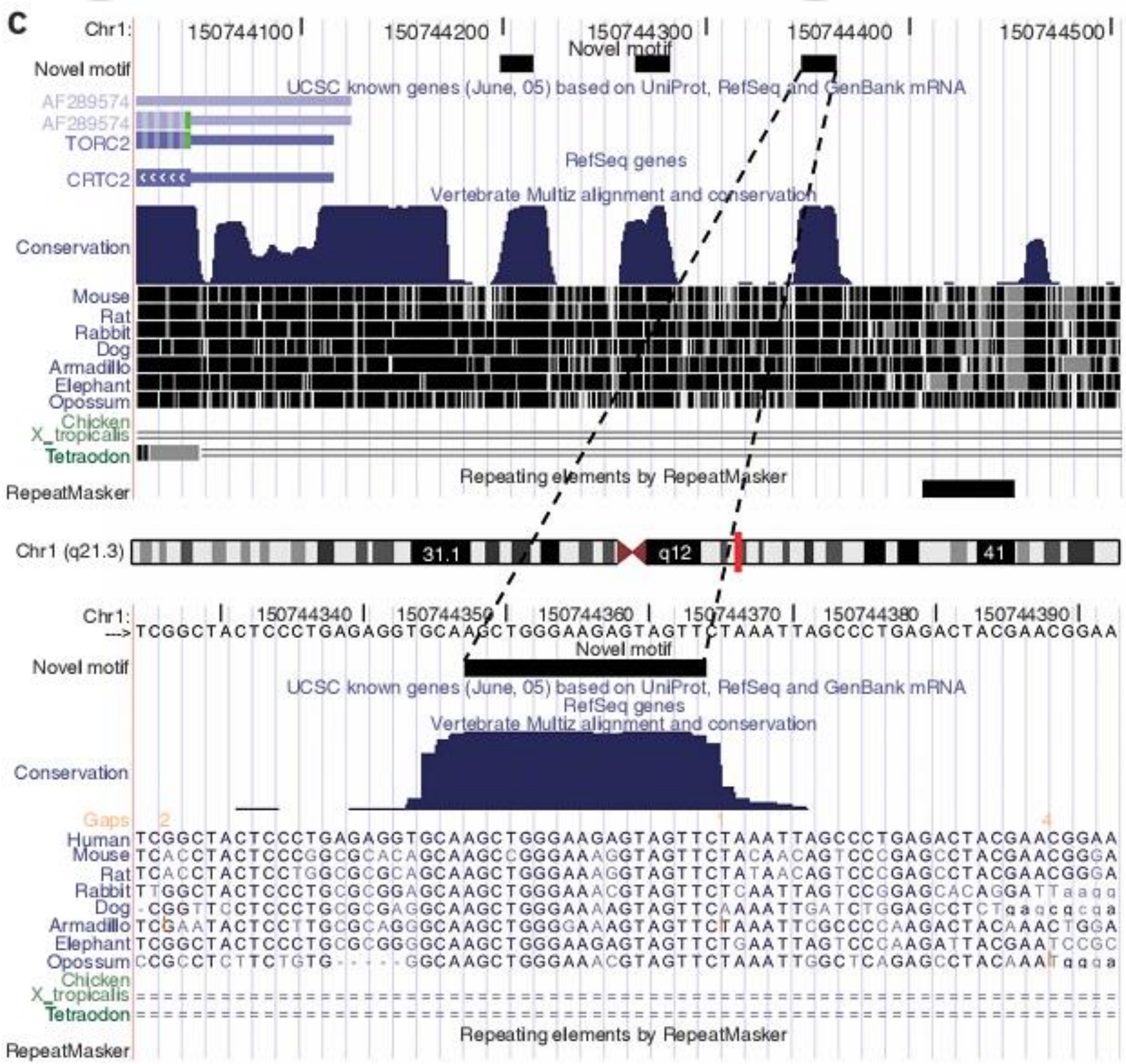
Fig. 1. Stepwise data analysis of Suz12 ChIP-on-chip in CoCAS. Quality control reports include (**A**) density plots of immunoprecipitated (IP) DNA, in red, and Input DNA, in green, so as to detect any dye bias; (**B**) MA plots which allow assessment of normalization quality and probe enrichment; (**C**) replicate correlation plots, which also help estimate background noise (which shows no correlation at low intensities). (**D**) Chromosomal view (chromosome 6) of Suz12 IP over input log ratios (in red) via IGB (top), followed by peak detection (green track) on a close up in the Hox cluster region (bottom).

MA强度

$$M = log_2 R - log_2 G$$
$$A = \frac{1}{2} \times (log_2 R + log_2 G)$$

The MA plot gives a quick overview of the distribution of the data. the general assumption is that most of the genes would not see any change in their expression. Therefore the majority of the points on the y axis (M) would be located at 0, since Log(1) is 0. If this is not the case, then a normalization method

GUI


c
Chr1: 150744100 150744200 150744300 150744400 150744500
Novel motif
UCSC known genes (June, 05) based on UniProt, RefSeq and GenBank mRNA
AF289574
AF289574
TORC2
CRTC2
RefSeq genes
Vertebrate Multiz alignment and conservation
Conservation
Mouse
Rat
Rabbit
Dog
Armadillo
Elephant
Opossum
Chicken
X_tropicalis
Tetraodon
Repeating elements by RepeatMasker
RepeatMasker
Chr1 (q21.3)
31.1
q12
41
Chr1: 150744340 150744350 150744360 150744370 150744380 150744390
---> TCGGCTACTCCCTGAGAGGTGCAAGCTGGGAAGAGTAGTTCTAAATTAGCCCTGAGACTACGAACGGAA
Novel motif
Gaps
Human
Mouse
Rat
Rabbit
Dog
Armadillo
Elephant
Opossum
Chicken
X_tropicalis
Tetraodon
RepeatMasker

实验中的障碍

哺乳动物中存在大量重复序列

确认相似的基因调控网络

芯片的缺陷

数据处理是巨大的挑战

实验产生大量数据，背景干扰大，标准化过程有待优化，没法确定错误率

Thanks for your attention!