



**BGI** Product Catalog  
Human Research  
产品手册——人类研究



华大基因  
**BGI**

**华大基因（深圳总部）**

地址：深圳市盐田区北山工业区综合楼（518083）  
电话：400-706-6615  
邮箱：info@bgi.com  
网站：www.bgi.com

**美洲华大基因（波士顿）**

地址：One Broadway, 14th Floor, Cambridge, MA 02142, USA  
电话：+1-617-500-2741  
邮箱：info@bgi.com  
网站：www.bgi.com

**欧洲华大基因（哥本哈根）**

地址：Ole Maalpes Vej 3, DK-2200 Copenhagen N, Denmark  
电话：+45-7026-0806  
邮箱：info@bgi.com  
网站：www.bgi.com

**日本华大基因（神户）**

地址：Kobe KIMEC Center BLDG.8F, 1-5-2 Minatojima-min-amimachi, Chuo-ku, Kobe City, Hyogo-pref. 650-0047, JAPAN  
电话：+81-785-996-108  
邮箱：info@bgi.com  
网站：www.bgi.com

**亚太、港澳台华大基因（香港）**

地址：香港新界大埔工业村大富街16号  
电话：+852-3610-3510  
邮箱：info@bgi.com  
网站：www.bgi.com

# ABOUT BGI

## 关于华大基因

深圳华大基因股份有限公司（简称华大基因）是华大基因集团下属子公司，华大基因秉承集团“基因科技造福人类”的愿景，以推动生物研究进展和提高全球医疗健康水平为出发点，基于基因领域研究成果及生物技术在民生健康方面的应用，进行科研和产业布局，致力于助力和加速科学创新，减少出生缺陷，加强肿瘤防控，抑制重大疾病对人类的危害，实现精准治愈感染，助力精准医学。

华大基因依托世界领先的生物信息研发、转化和应用平台，上百台高性能测序仪，质谱仪和强大的服务器存储为数据的输出、存储、分析提供有力保障。目前华大基因的主营业务为通过基因检测等手段，为医疗机构、科研机构、企业、事业单位等提供基因组学类的诊断和研究服务。

华大基因总部位于中国深圳，在京、津、汉、沪、穗等大陆主要城市设有分支机构和医学检验所，并在香港、欧洲、美洲、亚太等地区设有海外中心和核心实验室，已形成“覆盖全国、辐射全球”的网络布局。

公司主要服务于国内外的科研院校、研究所、独立实验室、制药公司等机构，以及各级医院、体检中心等医疗卫生机构、公司客户和大众客户。

目前公司服务已经覆盖了全球100多个国家和地区。包括国内31个省、市、自治区的1500多家科研机构和800多家医疗机构，其中三甲医院100多家；欧洲、美洲、大洋洲等地区合作的海外医疗和科研机构超过2000家。华大基因已经主导及参与发表了CNNS文章超过100篇，全部科研论文600余篇。



# Introduction

## 平台介绍

### 信息技术平台

拥有深圳、香港、北京、武汉、杭州等数个大型生物信息学超级计算中心，总峰值计算能力达到288.5T flops，总内存容量达到67.22TB，总存储能力达到35.09PB。其中位于深圳和香港的集群的峰值计算能力分列国内生物信息领域第一和第二位，有能力为海量生物信息学数据的存储、处理和分析提供稳定而高效的资源保障。下表所示为华大基因在中国各地所部署的超级计算集群。



深圳平台

天津平台

杭州平台

武汉平台

### 核酸测序平台

拥有 3730 和 Illumina HiSeq, Illumina MiSeq, BioNano Irys System, BGISEQ-500平台及Life Ion Proton、PacBio新一代测序平台，服务产品涵盖动植物、微生物、人及药物研发等多领域的核酸水平研究。



BGISEQ-500



MiSeq



HiSeq 2500/4000



BioNano Irys System



Sequel



PacBio RS II

# Introduction 平台介绍

## 质谱平台

拥有蛋白质组学和代谢组学领域强大的分析工具，如基于液相色谱的无胶分离系统，基于高灵敏度、高分辨度、高精度质谱的蛋白质和代谢物鉴定定量系统，和各种强大的生物信息分析软件等，可以对人类、动植物、微生物等生命科学领域的各类样本进行大规模的蛋白质和代谢组学研究。



G2-XS



LTQ-Orbitrap Velos



Q Exactive



TripleTOP 5600



Q Exactive HF



ultrafleXtreme



QTRAP 5500



## 分型平台

拥有芯片分型和质谱分型平台，可应用在疾病研究、药物筛选、动植物群体研究、分子育种等方面。



Illumina iScan



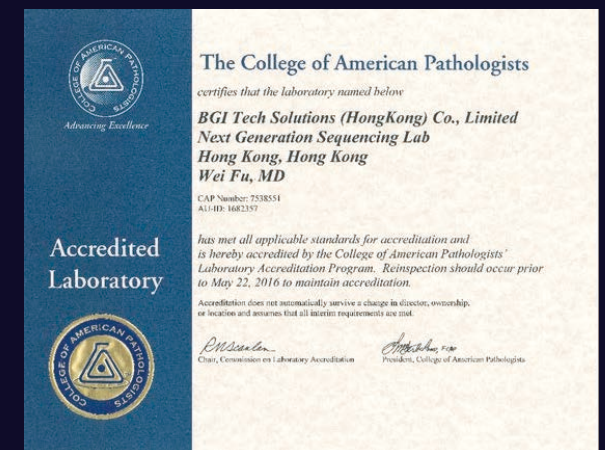
Affymetrix GeneTitan



MassARRAY

# Quality Assurance 质量管理

华大基因于2010年通过了ISO9001质量管理体系认证，于2011年12月通过了ISO14001环境管理体系认证和OHSAS18001职业健康安全管理体系认证，2012年6月通过了ISO/IEC 27001信息安全管理体认证，2015年7月华大基因香港高通量测序实验室获得美国病理学会（College of American Pathologists，以下简称“CAP”）颁发的CAP认可证书，成为中国首家在高通量测序服务和基因检测行业具有CAP认可资质的医学实验室。



# 目录

## CONTENTS

### 基因组学

- 07 人全基因组测序
- 15 全外显子测序
- 21 目标区域测序
- 25 芯片分型
- 28 质谱分型

### 表观组学

- 51 全基因组甲基化测序
- 55 目标区域捕获甲基化测序
- 59 ChIP-Seq

### 代谢组学

- 80 非靶向代谢组学
- 85 靶向代谢组学

### 转录组学

- 31 转录组测序
- 35 全长转录组测序
- 38 RNA-Seq
- 42 Small RNA测序
- 46 长链非编码RNA

### 蛋白组学

- 64 蛋白鉴定分析
- 69 蛋白定量分析
- 75 蛋白修饰分析

### 新技术

- 91 单细胞DNA测序
- 95 单细胞RNA测序
- 99 免疫组库测序

# 1 基因组学

- ◆ 人全基因组测序
- ◆ 全外显子测序
- ◆ 目标区域测序
- ◆ 芯片分型
- ◆ 质谱分型

# 人全基因组测序

## 基因组学

### ◆ 产品概述

人全基因组测序 (Human Whole genome sequencing, WGS) 是对人核酸样品进行全基因组范围的测序, 并在个体或群体水平进行差异性分析的方法。相比芯片检测或者外显子测序, 全基因组测序可以全面的挖掘基因序列差异和结构变异。随着测序成本降低, 云计算平台的推广使得科研人员对大数据的分析能力有所提高, 全基因组重测序已经成为人类遗传学、转化医学和群体进化领域最为迅速且有效的方法之一。

### ◆ 产品优势

**通量高:** 每个样本90Gb以上的数据产出量, 海量数据用于分析

**覆盖度高:** 能检测到99%的SNP

**分辨率高:** 单碱基分辨率

**检测范围广:** 可检测基因组内结构性变异 (SV) 及基因拷贝数变化 (CNV) 等, 可全面发掘罕见遗传变异和新遗传变异

**数据精准:** 数据偏差小、高均一性, 真实反映样本基因组信息

**经验丰富:** 华大基因主导或参与完成了多个基因组计划, 包括人类基因组计划、人类单体型计划、炎黄一号、千人基因组、荷兰人基因组、英国万人基因组

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

- **基本数据统计**
  - 1、对原始数据去除接头污染序列和低质量reads
  - 2、与人基因组参考序列比对
  - 3、统计测序深度和覆盖度
- **变异信息检测**
  - 1、SNP变异信息检测和注释
  - 2、InDel变异信息检测和注释
  - 3、SV变异信息检测和注释
  - 4、CNV变异信息检测和注释
- **高级信息分析**
  - 1、肿瘤研究高级信息分析
  - 2、人群体遗传学研究高级信息分析
  - 3、复杂疾病研究高级信息分析
  - 4、单基因病研究高级信息分析

### ◆ 技术参数

#### 样品要求

样品类型	最低送样量	建议送样量
gDNA	0.5μg	3μg
全血	0.5mL	2mL
细胞	$2 \times 10^5$ 个	$10^6$ 个

**推荐数据量:** 群体进化 (大样品量低深度, 1-3X)  
复杂疾病 (大样品量低深度, 1-3X)  
单基因病 (家系样品, 30-50X)  
肿瘤 (癌组织: 50X; 癌旁组织/血液样本: 30X)

**项目周期:** 标准流程的项目周期为40个工作日

## ◆ 案例解析

### ● 案例1 复杂疾病研究——大规模低深度测序发现重度抑郁症易感基因

抑郁症 (MDD major depressive disorder) 是最常发生的精神类疾病, 且病因复杂。此文对5303例中国女性周期性复发的抑郁症患者和5337例非抑郁症志愿者进行低深度全基因组测序, 发现10号染色体上的两个位点与抑郁症相关: 一个与 *SIRT1* 基因临近, 另一个位于 *LHPP* 基因的内含子。再分析4509例具有严重抑郁症患者, 得到 *SIRT1* 有一个增强的遗传信号, 再次得到验证。此次能够有重大发现应该归功于选取了症状相对一致的病例。

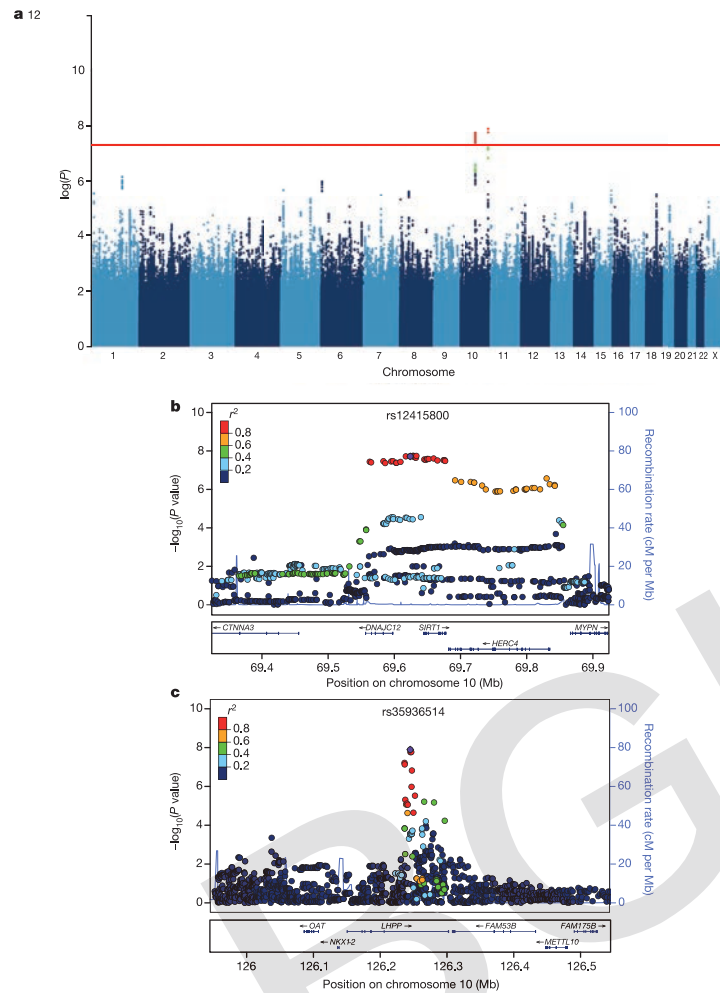


图1. 此次研究样品中发现10号染色体的两个位点与重度抑郁症相关。

- (A) 重度抑郁症全基因组相关的曼哈顿图
- (B) 10号染色体69.6Mb位置的 *SIRT1* 相关信息
- (C) 10号染色体126.2Mb位置的 *LHPP* 相关信息

参考文献:

Cai N, Bigdeli T B, Kretschmar W, et al. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*. 2015.

### ● 案例2 单基因病研究——单倍体测序方法验证 *MAGEL2* 截断突变引起 Prader-Willi 综合征和自闭症

单倍体测序方法是华大基因自主研发的新方法, 可以做到基因组单倍体分型, 有利于复合杂合异质性的疾病病因检测。Prader-Willi 综合征 (PWS) 是由于个体15q11-q13区域的基因, 父系和母系的基因均表达沉默引起的。这类基因表现为母系印记, 65-75% 的案例是由于父系的15q11-q13区域的基因缺失。此文报道了4个均是父系 *MAGEL2* 基因截断突变的个体, *MAGEL2* 是 PWS 区域的一个基因。第一个个体是通过全基因组测序确定, 其他三个是通过外显子测序确定的。四个病例均有自闭症 (ASD)、智力障碍以及不同程度的 Prader-Willi 综合征的临床特征。该研究发现 *MAGEL2* 是一个新的引起复杂自闭症的基因, 由于 *MAGEL2* 的功能缺失导致了 Prader-Willi 综合征的多个方面的表型。

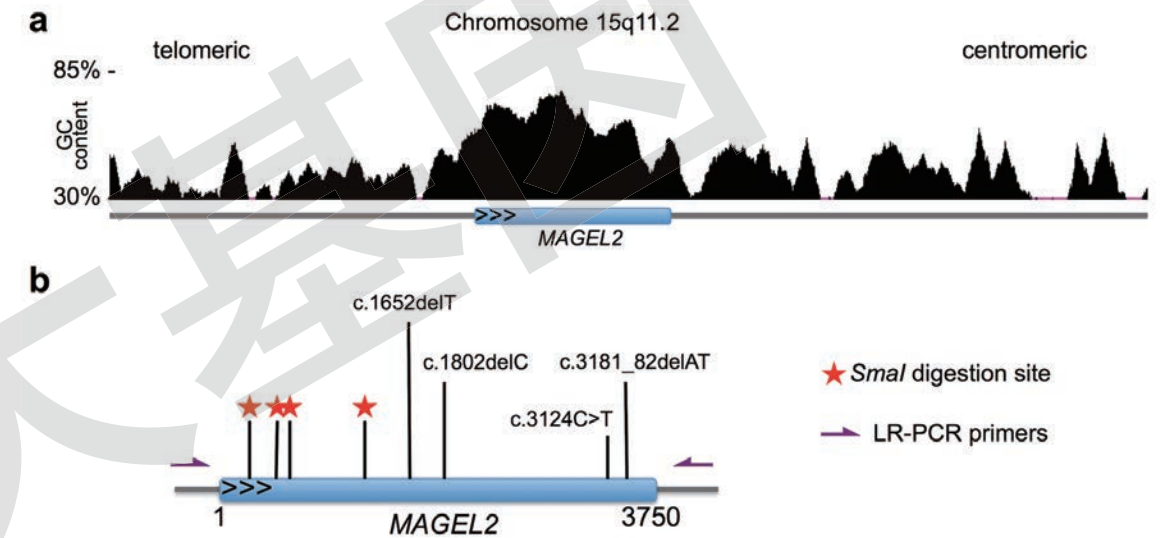


图2. *MAGEL2* 父本等位基因的缺失突变。

- (a) 15q11.2 染色体上 *MAGEL2* 以及侧翼序列的 GC 含量
- (b) 本文报道的 *MAGEL2* 缺失突变是发生在它唯一的外显子区域。

参考文献:

Schaaf C P, Gonzalez-Garay M L, Xia F, et al. Truncating mutations of *MAGEL2* cause Prader-Willi phenotypes and autism. *Nature Genetics*. 2013.

### ● 案例3 肿瘤研究——基因组重排导致高风险神经母细胞瘤端粒酶激活

神经母细胞瘤是交感神经系统的恶性儿科肿瘤, 通常一般的神经母细胞瘤会自愈或通过手术治疗。但是高风险的神经母细胞瘤, 即使采用多种治疗方法, 仍会恶化, 且该病的分子机制仍捉摸不透。本文对56个神经母细胞瘤样品进行全基因组测序 (39个高风险、17个低风险), 发现5p15.33附近的端粒酶反转录酶基因 (*TERT*) 发生了周期性的基因组重排。这些重排只发生在高风险的神经母细胞瘤 (12/39, 31%), 且与 *MYCN* 的扩增和 *ATRX* 突变相关。扩大样品 (217个样品) 验证发现, 染色体重排、*MYCN* 扩增都会引起 *TERT* 的转录表达量上调和病情恶化。然而神经母细胞瘤的端粒长度改变与 *TERT* 或 *MYCN* 的改变无关。进一步分析发现, 5p15.33重排和 *TERT* 编码序列共同使增强子作用, 导致大量的染色体重构, 并影响区域DNA甲基化。功能验证发现, 具有基因重排和 *MYCN* 扩增的神经母细胞瘤细胞系内, *TERT* 表达上调和端粒酶活性增强。本文发现高风险神经母细胞瘤通过基因组重排, 启动 *TERT* 的转录活性, 发挥了端粒酶的激活作用。

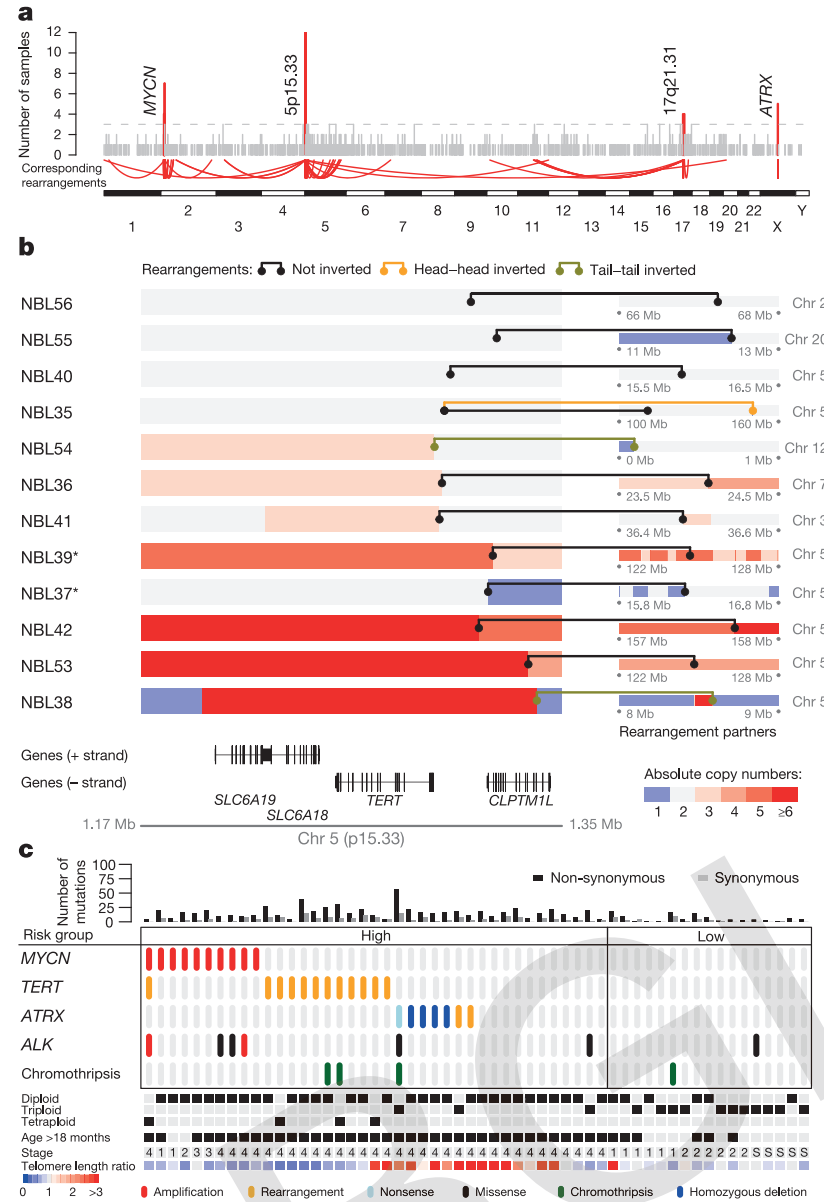


图3.基因组重排都集中在高风险神经母细胞瘤的5号染色体5p15.33区域

(a)56个原发性神经母细胞瘤内发生的100kb内区域基因组重排分布图。三个以上肿瘤发生重排的位置用红色标记

(b)染色体5p15.33发生染色体易位的细节图，右边是它们重排的搭档

(c) 56个原发性神经母细胞瘤各种基因变化的频率图谱，MYCN扩增、TERT重排、ATRX和ALK基因突变。

## ◆ 常见问题

Q: 人基因组部分区域GC含量过高对重测序有什么影响?

A: 由于化学本质的问题, GC含量高, 扩增困难, 偏差大, 从而引起测序偏差。因此, 一般高GC的地方, 平均深度较样品整个基因组的平均深度要低。另一方面, 高GC也影响酶扩增时候的准确性, 即高GC可能导致扩增错误率升高, 二代测序本身也是一种合成/扩增, 因此也会增加测序错误率, 最终引起整条序列的错误率升高。错误率越高, reads就有可能比对错误或无法比对。

Q: 人基因组中重复序列高的区域对重测序有什么影响?

A: 理论上, 一条序列, 它有可能来自于基因组的一个位置, 如果它在参考基因组上是重复序列, 能比对到5个不同的位置, 当我们随机选择一个位置的时候, 错误率为80%。当然, 如果他在所测的基因组上仍然是重复序列, 那随机取一个位置错误造成的影响会很低。即真正的重复序列对重测序的结果不会造成太大的错误, 但是由于测序错误造成的比对重复序列影响就较大(即序列本身属于uniq序列, 但是由于有几个碱基是测序错误, 造成能够同时比对到其他的一些地方, 那么, 序列比对到其他位置给出的序列信息就完全是错误的)。

Q: 人类基因组中常见变异类型有哪些?

表4 人类基因组中常见变异类型

突变类型	定义	基因组中的频率
SNP	在基因组上单个核苷酸的变异, 包括置换、颠换、缺失和插入	约每1000bp中有一个
插入/缺失	DNA片段的插入/缺失, 包括小规模的多态性变化以及大的染色体畸变, 大于1Kb的InDel又称为CNV	> 100百万个InDel(> 1bp)
微卫星	序列中小于200bp的1-6bp的重复序列	> 100百万个
微卫星及串联重复序列	多态性序列包括6-100bp重复的20-50个拷贝	约15000
多点突变(MSV)	由于CNV和基因转换引起的复杂单大碱基变化	数目不详
中等大小的结构变异ISV	获得或丢失大于8Kb的DNA序列, 包括倒位的Breakpoint	297个ISV已经被鉴定
CNV、CNP、LCVS	大于1Kb的InDel又称为CNV, 如果CNV频率大于1%, 则叫CNP, 当插入缺失的片段大于50Kb时, 则叫LCVS	数目不详
倒位	染色体结构变异的一种。由于重排引起的染色体上两个断裂点间的断片, 倒转180°后又重新连接	数目不详
易位	由重排引起的将DNA片段将连接到不同染色体	数目不详

Q: 一般用什么方法来验证call SNP 准确率?

A: 华大炎黄计划是用Sanger测序的方法和芯片分型两种方法来验证SNP的准确性的, 因为Sanger测序被认为是测序中的“金标准”。另外, 质谱也是常用的验证SNP准确性的方法。

Q: 如何在众多假阳性中把致病的SNP过滤出来?

A: 通过遗传变异数据库和对照数据去除常见突变; 通过蛋白序列预测去除不改变蛋白功能的突变; 在不同的病例突变数据库中取共同的突变, 最终得到少量的突变候选突变。

◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
人类基因组计划	2001.02	Nature	38.14	Initial sequencing and analysis of the human genome
单体型图谱	2003.12	Nature	38.14	The International HapMap Project
	2005.10	Nature	38.14	A haplotype map of the human genome
炎黄一号	2008.11	Nature	38.14	The diploid genome sequence of an Asian individual
古人	2010.02	Nature	38.14	Ancient human genome sequence of an extinct Palaeo-Eskimo
HapMap	2010.09	Nature	38.14	Integrating common and rare genetic variation in diverse human populations
澳洲土著人	2011.10	Science	34.66	An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia
荷兰人	2013.05	European Journal of Human Genetics	4.58	The Genome of the Netherlands: design, and project goals
	2014.06	Nature Genetics	31.62	Whole-genome sequence variation, population structure and demographic history of the Dutch population
	2015.03	Nature Communications	11.33	Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels
	2015.05	Nature Genetics	31.62	Genome-wide patterns and properties of de novo mutations in humans
千人	2010.10	Nature	38.14	A map of human genome variation from population-scale sequencing
	2012.04	Nature Methods	25.33	The 1000 Genomes Project: data management and community access
	2015.10	Nature	38.14	An integrated map of structural variation in 2,504 human genomes
	2015.10	Nature	38.14	A global reference for human genetic variation

研究内容	发表时间	发表期刊	影响因子	文献标题
UK10K	2015.09	Nature Communications	11.33	Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel
	2015.10	Nature	38.14	The UK10K project identifies rare variants in health and disease
	2015.10	Nature	38.14	Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture
	2015.09	Nature Genetics	31.62	Height-reducing variants and selection for short stature in Sardinia
膀胱癌	2013.12	Nature Genetics	31.62	Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation
智障	2014.07	Nature	38.14	Genome sequencing identifies major causes of severe intellectual disability
肺癌	2014.09	Cancer Research	8.56	Whole-genome sequencing of asian lung cancers: second-hand smoke unlikely to be responsible for higher incidence of lung cancer among Asian never-smokers
老年性黄斑病变	2015.01	Nature Communications	11.33	New loci and coding variants confer risk for age-related macular degeneration in East Asians
宫颈癌	2015.01	Nature Genetics	31.62	Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism
重度抑郁症	2015.07	Nature	38.14	Sparse whole-genome sequencing identifies two loci for major depressive disorder
神经母细胞瘤	2015.10	Nature	38.14	Telomerase activation by genomic rearrangements in high-risk neuroblastoma
LFR方法	2012.07	Nature	38.14	Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells
LFR+单基因病	2013.11	Nature Genetics	31.62	Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism
LFR方法	2015.01	Genetics	4.64	Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing
LFR+PGD	2015.01	Genome Research	11.35	Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing



# 全外显子测序

## 基因组学

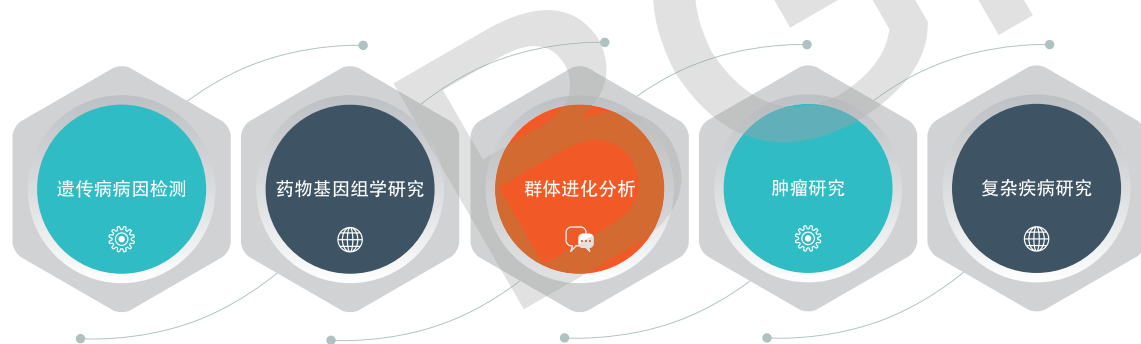
### ◆ 产品概述

外显子组仅占全基因组序列的1~2%左右，却包括大多数与疾病相关的变异。与全基因组测序相比，外显子组测序不仅经济高效，数据阐释也更简单。与全基因组关联分析相比，外显子组测序可鉴定低频和罕见变异，结果分析更准确。近年来外显子组测序已广泛应用于孟德尔遗传病致病基因的挖掘和癌症等复杂疾病易感基因的研究。华大基因作为外显子测序领导者，已经超过10万份例外显子样品的稳定数据为基础，为您提供最有保障的服务。

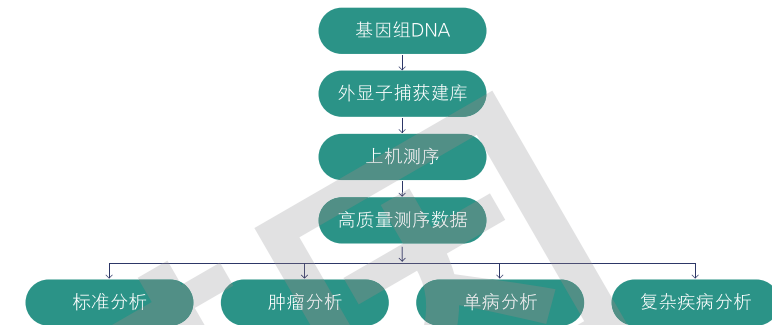
### ◆ 产品优势

- 高性价比：**仅测序1%的基因组序列，可以得到大约85%的人类致病突变位点
- 高准确性：**高深度测序，可发现常见变异和罕见变异，并发掘新变异
- 质量超群：**规范完整的项目流程和完善的质量体系，为项目质量保驾护航
- 经验丰富：**完成超过10万例样品建库测序，发表相关文章超过150篇
- 研究广泛：**涉及19种肿瘤类型，22种复杂疾病；罕见病研究性文章超过70篇

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

标准信息分析		
1、数据质控：去除接头污染和低质量数据		
2、与参考基因组比对、统计测序深度和覆盖度		
3、SNPs/InDels检测、多种数据库注释、保守性预测、致病性分析、统计		
癌症分析	复杂疾病分析	单基因病分析
1、样品成对初步鉴定	1、群体SNP检测、注释与统计、质控	1、多种软件数据比对、变异检测
2、Somatic SNV/InDel检测、注释、统计	2、基于遗传数据的样本质控	2、变异效应预测工具注释
3、CNV检测及注释	3、配对分析	3、群体等位基因频率注释
4、基因突变频率、样品突变情况统计	4、单个SNP和Gene-based的关联分析	4、有害性或保守性预测工具打分、通路相关注释
5、显著性基因和信号通路分析	5、候选基因优化，候选SNP的条件分析	5、OMIM相关注释、正常组织蛋白表达注释
6、潜在肿瘤相关通路分析	6、候选位点附近的LD分析及基于单体型域的单体型关联分析	6、标签性家系共分离筛选
7、克隆结构分析	7、GO功能分类，基于代谢通路的交互分析	
8、生存分析	8、ROC曲线与遗传方差解释	
	9、eQTL分析	
	10、群体CNV检测，注释与统计、关联分析	
	11、基于家系样本的 <i>de novo</i> mutation分析	

## ◆ 技术参数

标准信息分析	标准信息分析
组织	≥0.5g
全血	≥2mL
福尔马林固定石蜡包埋组织(FPE)	>10个玻片或50mm2,5-10μm厚的切片10片
常规基因组DNA	m≥2.5μg, c≥12.5ng/μL
微量基因组DNA	>200ng, c≥2.5ng/μL
FFPE基因组DNA	m≥1μg, c≥10ng/μL, 要求样品量
	m≥0.2μg, c≥2.5ng/μL, 风险送样

测序	推荐数据量	项目周期
PE50、PE101、PE151等	50X、100X、200X	25-40个工作日

## ◆ 案例解析

### ● 案例1 癌症研究——外显子测序发现8个和食管鳞癌相关的突变基因

食管癌是最具侵袭性的癌症类型之一，分鳞状细胞癌和腺癌两种，是全球第六位的主要癌症死亡原因。近70%的全球食管癌病例发生在中国，绝大多数病例(>90%)的组织病理形式是食管鳞状细胞癌(ESCC)。文章共选取了158个ESCC的患者通过对17个ESCC患者大于30X深度的全基因组测序和71个ESCC患者大于100X的全外显子测序以及在71个ESCC患者中的53名患者，加上没有进行上述分析的70个ESCC患者的比较基因组杂交分析发现了8个与食管鳞癌发生相关的重要基因突变，其中6个是已知的和肿瘤相关的基因：*TP53*、*RB1*、*CDKN2A*、*PIK3CA*、*NOTCH1*、*NFE2L2*，另外两个基因(*ADAM29*和*FAM135B*)在之前这种类型的癌症中没有被报道过。其中*FAM135B*是首次发现的肿瘤相关基因。

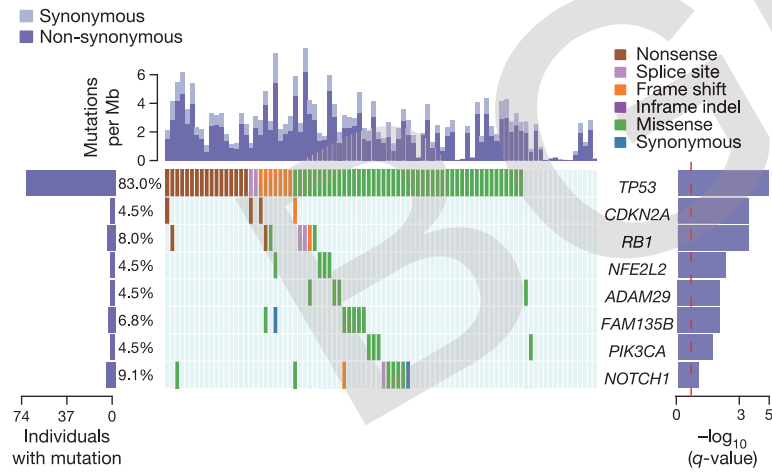


图1. 食管鳞状细胞癌中的显著突变基因

参考文献:

Song Y, Li L, Ou Y, Gao Z, et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature. 2014.

### ● 案例2 单基因病研究——外显子测序解析卵巢早衰的遗传因素

卵巢早衰通常指女性40岁之前闭经，1%的妇女患有此病，病因复杂，被认为受到遗传因素的影响。本研究利用外显子测序技术首次在中东家系1(MO1DA)的卵巢早衰病人样本进行分析，DNA由这个家族中的两个姊妹提供，姊妹俩其中一个健康的，另一个是不育的。研究发现减数分裂基因中的*STAG3*基因突变可以导致隐性遗传卵巢早衰，也在小鼠动物模型和卵巢早衰病患中得到了证实。为探索卵巢早衰或卵巢功能不全的发生机理，以及阐明该病的临床高度异质性和遗传病因复杂性开辟了一个新的研究路径。

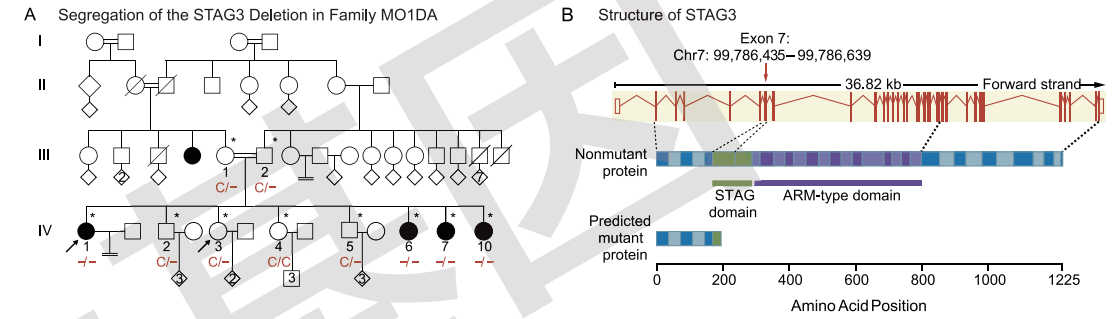


图2. MO1DA家系图谱

图3. STAG3基因结构图

参考文献:

Caburet S, Arboleda VA, Llano E, et al. Mutant cohesin in premature ovarian failure. N Engl J Med. 2014.

### ● 案例3 复杂疾病研究——外显子测序鉴定一种遗传变异与II型糖尿病相关

格陵兰岛人历史上是一个小型而孤立的人群，因此具有“奠基者效应”。“奠基者效应”是遗传漂变的一种形式，指的是由很少一群人繁衍出的种群，这个群体后来数量虽然会增加，但因未与其他生物群体交配繁殖，彼此之间基因的差异性甚小。丹麦哥本哈根大学托本·汉森和他的研究团队分析了5000份血样，利用芯片分型，找到250000个II型糖尿病相关性状的遗传变异，并对9个三人组共27个样本进行外显子测序，结果表明外显子测序SNP与SNP芯片结果没有任何不一致。他们发现，一个名叫*TBC1D4*基因中的突变，会让研究对象在口服相当于一顿饭的葡萄糖两个小时，血液中的葡萄糖和胰岛素含量更高，从而增加II型糖尿病的患病风险。此次研究和过往的II型糖尿病相关性状的规模全基因组关联分析相比，新观察到的遗传变异带来的影响大了好几倍。

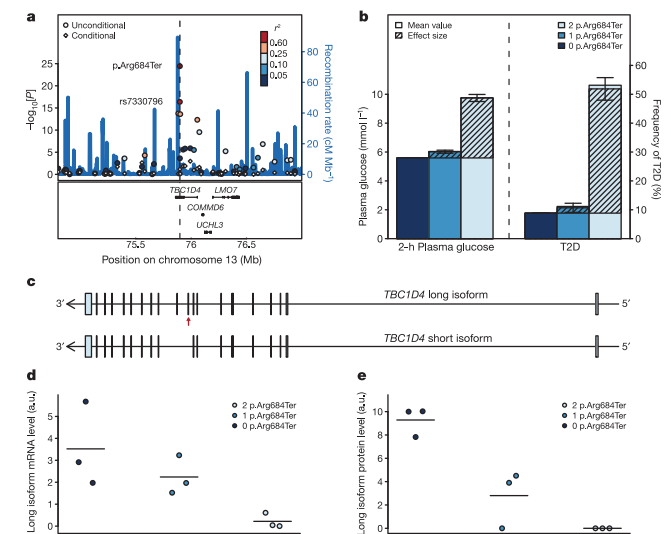


图4. *TBC1D4*基因p.Arg684Ter无义多态性的影响

参考文献:

Caburet S, Arboleda VA, Llano E, et al. Mutant cohesin in premature ovarian failure. N Engl J Med. 2014.

## ◆ 常见问题

Q: 石蜡样本提取的DNA可以用吗? 样本放多长时间就提不出高质量DNA了?

A: 石蜡样本的重测序和外显子测序我们都可以做。至于提取DNA, 我们可以通过小样本量的预实验来判断DNA质量, 如果能达到要求, 就可进行大样本测序。样本质量和石蜡包埋过程, 存放时间和存放温度都有关系, 在低温保存, 三年之内的石蜡样本质量较好, 但最终还是要看预实验结果。

Q: 外显子捕获测序的效率如何?

A: 外显子组覆盖度达到杂交芯片可捕获的不低于90%的目标区域, 发现杂交芯片可捕获区域中99%的SNPs(单核苷酸多态性)。外显子芯片区域的捕获效率一般有60%左右。建议外显子组测序的测序深度大于50X, 对于肿瘤或临床研究, 建议高精度测序。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
癌症	2015.04	Gut	14.92	Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer
	2014.06	Cell Research	14.81	Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing
	2014.03	Nature	38.14	Identification of genomic alterations in oesophageal squamous cell cancer
	2014.03	Science	34.66	Activating Hotspot L205R Mutation in PRKACA and Adrenal Cushing's Syndrome
	2013.12	Nature Genetics	31.62	Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation.
	2013.06	Advanced Materials	18.96	High-Purity Prostate Circulating Tumor Cell Isolation by a Polymer Nanofiber-Embedded Microchip for Whole Exome Sequencing
	2013.02	Nature Genetics	31.62	Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia.
	2012.03	Cell	28.71	Single-Cell Exome Sequencing and Monoclonal Evolution of a JAK2-Negative Myeloproliferative Neoplasm
	2012.03	Cell	28.71	Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor
	2012.02	Leukemia	12.10	Low frequency of DNMT3A mutations in pediatric AML, and the identification of the OCI-AML3 cell line as an in vitro model
	2011.12	Nature Genetics	31.62	Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma
	2011.08	Nature Genetics	31.62	Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder
	罕见病/遗传性疾病	2015.07	American Journal of Human Genetics	10.79

研究内容	发表时间	发表期刊	影响因子	文献标题
罕见病/遗传性疾病	2015.02	American Journal of Human Genetics	10.79	Keppen-Lubinsky Syndrome Is Caused by Mutations in the Inwardly Rectifying K+ Channel Encoded by KCNJ6
	2014.11	Cell Research	14.81	De novo mutation in ATP6V1B2 impairs lysosome acidification and causes dominant deafness-only-chodystrophy syndrome
	2014.01	Blood	11.84	Inherited bone marrow failure associated with germline mutation of ACD, the gene encoding telomere protein TPP1
	2012.09	Nature Genetics	31.62	Exome sequencing identifies NMNAT1 mutations as a cause of Leber congenital amaurosis
	2012.03	American Journal of Human Genetics	10.79	Exome Sequencing Reveals Mutations in TRPV3 as a Cause of Olmsted Syndrome
	2012.01	Nature Genetics	31.62	Exome sequencing identifies MVK mutations in disseminated superficial actinic porokeratosis
	2010.12	Brain	10.10	TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing
复杂疾病	2015.09	Journal of Allergy and Clinical Immunology	12.49	Novel mutations in TNFRSF7/CD27: Clinical, immunologic, and genetic characterization of human CD27 deficiency
	2015.04	Hepatology	11.71	The p.Ser267Phe variant in SLC10A1 is associated with resistance to chronic hepatitis B
	2014.08	Nature	38.14	A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes
	2014.07	Nature Communications	11.33	Sequencing-based approach identified three new susceptibility loci for psoriasis
	2014.01	Nature Genetics	31.62	A large-scale screen for coding variants predisposing to psoriasis
	2013.08	Gastroenterology	18.19	Association Between Variants of PRDM1 and NDP52 and Crohn's Disease, Based on Exome Sequencing and Functional Studies
	2012.11	Hepatology	11.71	Rare Inborn Errors Associated With Chronic Hepatitis B Virus Infection
	2010	Nature Genetics	31.62	Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants
	2015.03	Molecular Neurobiology	5.40	Two Novel Susceptibility SNPs for Ischemic Stroke Using Exome Sequencing in Chinese Han Population
	2015.03	Genome Biology and Evolution	4.10	Inference of purifying and positive selection in three subspecies of chimpanzees (Pan troglodytes) from exome sequencing
群体	2015.03	Molecular Neurobiology	5.40	An effort to use human-based exome capture methods to analyze chimpanzee and macaque exomes
	2012.02	PNAS	9.42	Extensive X-linked adaptive evolution in central chimpanzees
	2010.07	Science	34.66	Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude

# 目标区域捕获测序

## 基因组学

### ◆ 产品概述

目标区域测序是指利用定制的探针对感兴趣的基因组区域进行捕获，富集后进行高通量测序的基因组分析方法。该方法能够获得目标区域的遗传信息，极大地提高了基因组中特定区域的研究效率，降低研究成本，非常适合大样本量研究。通过目标区域测序，可以对候选基因或候选位点进行验证，也可以进一步找到其易感位点，结合大量的公共数据库信息更好地解释变异之间的关联和疾病的致病机理。

### ◆ 产品优势

**精准研究：**只需对感兴趣的区域及相关基因进行研究

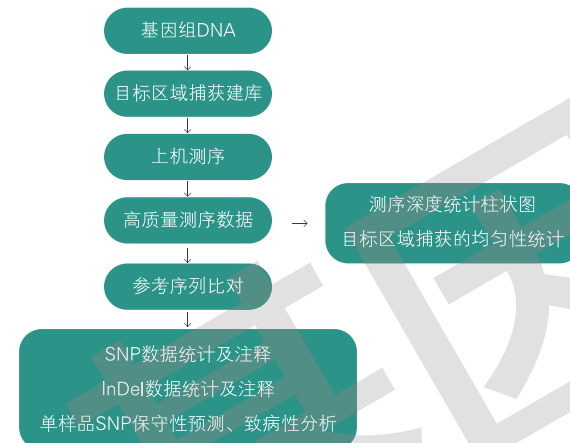
**高性价比：**适用于大样本量的候选基因筛查、小区域的高深度研究

**经验丰富：**累计已完成上万个样品的检测和分析

### ◆ 产品应用



### ◆ 技术流程



#### 定制化捕获试剂盒：

Agilent SureSelect XT Custom Kit  
NimbleGen SeqCap EZ Choice Enrichment Kit  
NimbleGen SeqCap EZ Choice XL Enrichment Kit

#### 设计好的捕获试剂盒：

TumorCare Cancer Panel (癌症热点突变基因，信息分析内容含SNV、InDel、CNV、Gene fusion)  
MHC Capture Kit (4.97Mb，技术参数及分析内容略有不同)  
病毒捕获芯片 (HBV/HPV)

### ◆ 信息分析内容

#### 标准信息分析

- 1、对原始数据进行去除接头、污染序列及低质量reads的处理
- 2、测序评估 (数据比对统计、测序饱和度分析、测序随机性的统计分析、reads在参考基因上的分布分析)
- 3、SNP、InDel检测，注释和统计
- 4、SNP保守性预测、致病性分析
- 5、SNP、InDel在各基因功能元件上的分布统计

#### 高级信息分析

- 1、非编码区SNP的注释分析 (基于ENCODE数据库进行注释分析)
- 2、癌症高级信息分析
- 3、复杂疾病高级信息分析
- 4、群体高级信息分析

### ◆ 技术参数

**样品总量：**≥ 1μg DNA，微量建库样品量低至200ng

**样品浓度：**≥ 12.5ng/μL

**项目周期：**标准流程的项目周期为40个工作日

## ◆ 案例解析

### ● 案例1 外显子和目标区域大规模筛查银屑病易感基因

对21309名中国银屑病患者样品进行目标区域测序，发现罕见突变（MAF<1%）对银屑病的影响可能很有限。本文采用第一阶段全外显子测序（样本量781例银屑病人/676例正常人），第二阶段目标区域测序（区域来自第一阶段发现的p值显著的位点或基因，以及前人研究已发现的免疫病相关的基因，共计1326个基因，样本量9946病例/9906正常人）。结果发现了*IL23R*和*GJB2*基因的两个低频错义突变（1%<MAF<5%），*LCE3D*、*ERAP1*、*CARD14*、*ZNF816A*中的5个常见错义突变（MAF>5%）与银屑病的发生显著关联，罕见SNP（MAF<1%）仅发现*FUT2*和*TARBP1*中两个错义突变可能与银屑病发病相关，但是并没有显著关联。

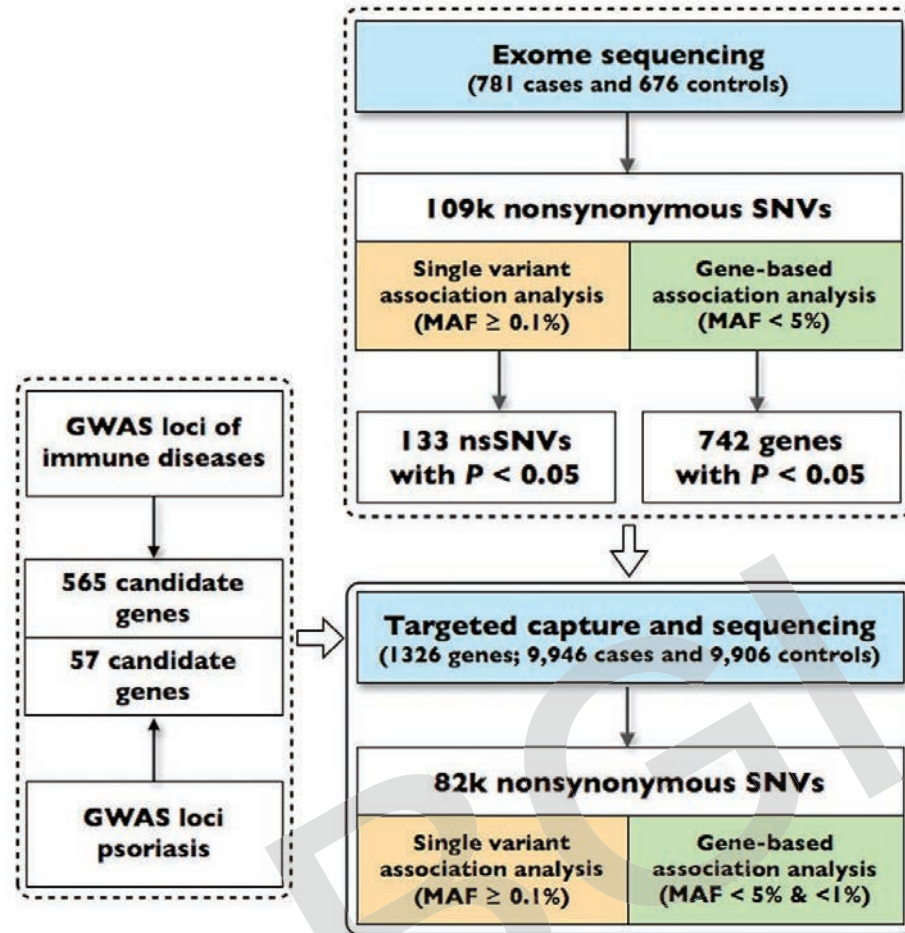


图1. 文章研究思路图

参考文献:

Tang H, Jin X, et al. A large-scale screen for coding variants predisposing to psoriasis. Nat Genet. 2014.

## ◆ 常见问题

Q: 捕获平台，一般推荐几杂？

A: 根据华大经验，推荐使用2杂或者4杂，具体看区域大小、样品类型和合同指标而定。

Q: 芯片的订货周期一般是多长？

A: 一般为45-60天。

Q: 捕获效率有哪些因素影响？推荐多少数据量？

A: 样本质量，区域复杂度等均可影响捕获效率。由于捕获效率未知，我们无法承诺有效深度，可以根据华大的经验捕获效率做个预估的深度。推荐测序量至少100倍，一般癌症样品测序深度需要更高，客户也可根据研究需要确定测序深度或数据量。

Q: 目标区域评估需要提供哪些信息？

A: 客户只需提供要捕获的物种名称，参考的基因组版本和目标区域所在的染色体号、起点位置、终点位置。该区域位置选择需要根据研究目的选择合适的区域如基因的exons、upstream、downstream或者连续的一段区域等。

Q: 目标区域有什么要求？

A: 可以是连续的DNA片段，也可以是分布在同一染色体不同区域或不同染色体上的片段。长度不定，原则上没有限制。但太小（几十Kb以下）且样本量很少时，建议散样测序；太大且目标区域都位于外显子情况下，成本很高，建议做外显子测序。难点区域：复杂区域，如重复序列较多，GC含量过高或过低，N区等存在探针设计困难。

# 芯片分型

## 基因组学

### ◆ 产品概述

华大基因拥有Illumina SNP芯片分型平台（Illumina Infinium®平台）和Affymetrix基因分型平台（Affymetrix GeneTitan®平台），这两大公司的基因分型平台是目前应用最为广泛的基因分型平台，适用于大样本不同标记密度的快速基因分型。

### ◆ 产品优势

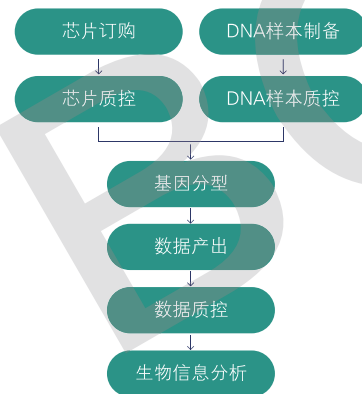
**基因分型服务通量高：**每周可完成上千例样品

**起始量低：**低至80ng，不同芯片要求不同

**精准研究：**成功率高，准确

**产品丰富：**标记密度高，芯片产品多样化、标准化

### ◆ 技术流程



### ◆ 基因分型芯片

Illumina基因分型产品目录	Affymetrix基因分型产品目录
人类疾病研究方向标准化芯片类型	
Human Omni zhonghua-8	Affymetrix GeneTitan® MC平台
Infinium Multi-Ethnic AMR/AFR-8	Axiom® Genome-Wide CHB1 And CHB2 1.2M
Infinium Multi-Ethnic EUR/EAS/SAS-8	Axiom® Genome-Wide EUR1 674K
Infinium Multi-Ethnic Global-8	Axiom® Genome-Wide EAS1 712K
HumanExome-12 BeadChip	Axiom® Genome-Wide AFR1 893K
Human OmniExpress-24	Axiom® Genome-Wide LAT1 817K
Human Omni 2.5M-8	Axiom® Genome-Wide CEU1 587K
HumanOmni 5-4	Axiom® Genome-Wide ASI1 600K
HumanOmni 5 Exome	Axiom® Genome-Wide PanAFR Genotyping Bundle(3 array set) 2.2M
Human Omni 2.5Exome -8	Affymetrix GeneChip® Scanner 3000 7G平台
Human OmniExpressExome-8	Axiom® Biobank Genotyping 675K
Infinium Human Core-24	Genome-Wide Human SNP Array 6.0 芯片
Infinium CoreExome-24 BeadChip Kit	OncoScan™ FFPE芯片
Infinium PsychArray-24	OncoScan™ CNV FFPE芯片
Infinium ImmunoArray-24 v2 BeadChip	CytoScan® HD 芯片
Infinium OncoArray-500K Beadchip	CytoScan® 750k 芯片
Infinium ONcoArray-500K+ Beadchip	CytoScan® Optima Array
Infinium HumanMethylation450 Beadchip	
Infinium MethylationEPIC(850K) Beadchip	
Infinium CytoSNP-850K BeadChip	
Infinium HumanCytoSNP -12	
Infinium HumanCytoSNP FFPE-12	
Cancer SNP Panel	
定制化芯片	
Infinium iSelect HD&HTS Custom Genotyping BeadChips	Axiom® myDesign™ Human Genotyping Arrays

## ◆ 信息分析内容

### 标准信息分析

- 1、数据产出与SNP calling
- 2、样品水平质控
- 3、位点水平质控

### 高级信息分析

- 1、缺失数据处理
- 2、群体分析
- 3、关联分析
- 4、注释分析

## ◆ 技术参数

样品总量：≥500ng DNA/次

样品浓度：≥50ng/μL

样品质量：无降解或轻微降解

项目周期：样本数低于500时，标准流程的项目周期为20个工作日（不包括芯片订购、高级信息分析，具体时间需根据具体芯片种类确定）

项目合格指标：芯片分型承诺平均位点检出率≥95%

# 质谱分型

## 基因组学

## ◆ 产品概述

华大基因采用世界领先的MassARRAY质谱系统进行SNP分型，该项技术曾被用于“人类单倍体型图计划（HapMap）”等多项基因组关联性研究项目。质谱SNP分型产品将多重PCR技术、单碱基延伸技术和基质辅助激光解吸电离飞行时间质谱（MALDI-TOF MS）技术完美结合，能够快速、经济、准确、高效地完成SNP分型。华大基因拥有十几年的技术服务经验、卓越的引物设计能力、严格的实验质控体系和优质的客户服务意识，从技术和服务上为您的课题保驾护航。

## ◆ 产品优势

**多重反应：**能够在在一个反应孔中实现多达30个SNP的检测

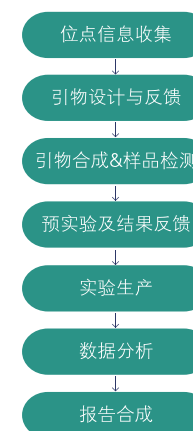
**通量高：**384孔板平行反应，一天最高检测通量可达10万位点

**设计灵活：**样品数和位点数的随意组合

**实验高效：**可设计95%的SNP位点，预实验采纳位点的检出率超过95%

**高准确性：**准确度高达99.7%

## ◆ 技术流程



## ◆ 技术参数

样品浓度：>10ng/μL

样品体积：>10μL/反应

样品纯度：OD260/280=1.8-2.0

结果交付：单个合格样品所有位点的检测结果单个位点的检出率代表性位点的峰图和聚类图

项目周期：1000个样品的项目周期约为20个工作日，实际项目周期根据样品数和位点数确定

## ◆ 常见问题

Q：位点数和样品数多少的情况下选择质谱分型比较合适？

A：位点数10-200 SNPs，样品数一百到数千不等的情况下适合选择质谱分型的方法。

Q：与其它SNP分型方法相比，质谱分型的优势在哪里？

A：质谱分型的准确性比芯片分型高，而相比于Taqman探针法和一代测序，质谱分型的多位点同时检测具有明显的通量优势，样品数较多的情况下建议选择质谱分型。

Q：位点评估一般需要多久？

A：收到有效位点信息后会在1-2个工作日内进行引物设计方案反馈。

Q：是否可以检测肿瘤组织的体细胞变异？

A：可以，为了尽可能地增加检测的灵敏度，引物设计和数据分析需要特殊的流程，在提供位点进行引物设计评估时需要特别注明。

Q：是否可以用于其它种属样品的SNP分型？

A：质谱SNP分型对种属无限制，只需要在位点评估时提供待检测位点的侧翼序列和变异类型即可。

Q：有没有运用该技术发表过的优质论文可供学习交流？

A：截止目前，运用质谱技术发表的SCI论文已有2500多篇，可通过登陆下面网址选择您感兴趣领域的论文进行学习参考：<http://agenabio.com/resources/publication-library/>。

# 2 转录组学

◆ 转录组测序

◆ 全长转录组测序

◆ RNA-Seq

◆ Small RNA测序

◆ 长链非编码RNA



# 转录组测序

## 转录组学

### ◆ 产品概述

转录组测序的研究对象为特定细胞在某一功能状态下所能转录出来的所有RNA的总和，目前该测序技术主要针对mRNA。转录组测序技术在检测基因表达水平变化的同时，还能发现未知转录本和稀有转录本，精确地识别可变剪切位点、基因融合、SNP以及RNA编辑位点，提供最全面的转录组信息。通过新一代高通量测序，能够全面快速地获得某一物种特定组织或器官在某一状态下的几乎所有转录本序列信息，已广泛应用于基础研究、临床研究和药物研发等领域。

### ◆ 产品优势

**样品多样：**提供单细胞、FFPE、微量样品等特殊样品服务，并且微量样本建库低至pg级

**结果精准：**独创的插入片段文库，完美匹配HiSeq4000的长读长，转录本组装长度更长，可变剪接、基因融合鉴定更灵敏

**内容前沿：**分析采用主流的分析软件，KEGG分析采用最新版数据库；包含主成分分析、RNA编辑分析等特有的分析

**报告简洁：**结果图片更加直观，可用于文章撰写；结题报告结构优化，有效提升阅读体验

**经验丰富：**专业的信息分析团队，目前已经完成100,000+例转录组测序，发表多篇高水平转录组文章

### ◆ 产品应用



### ◆ 技术流程



### ◆ 分析内容

- 1、测序数据过滤
- 2、参考基因组比对
- 3、新转录本预测
- 4、SNP和InDel检测
- 5、RNA编辑检测（基于第4条，且必须提供样品DNA数据检测出来的SNP结果）
- 6、差异剪接基因检测
- 7、融合基因检测（仅适用于人的癌症融合基因研究，每个样本8G以上数据量）
- 8、基因表达量计算
- 9、差异表达基因检测
- 10、差异表达基因层次聚类分析
- 11、差异表达基因GO功能分析
- 12、差异表达基因Pathway功能分析
- 13、差异表达基因蛋白互作分析

## ◆ 技术参数

样品类型：完整且无污染的总RNA

样品总量：200ng/次

样品浓度： $\geq 20\text{ng}/\mu\text{L}$

样品纯度：28S:18S $\geq 1.0$ ；RIN $\geq 7.0$

测序策略：PE101/PE151

推荐数据量：8G以上

项目周期：标准流程的项目周期为40个工作日

## ◆ 案例解析

### ● 案例1 转录组测序研究精神分裂症和躁郁症

通过对26个对照样本的大脑组织、31个精神分裂症（SCZ）大脑组织、26个躁郁症（BPD）大脑组织样本进行转录组测序，研究精神分裂症和躁郁症在基因表达层面的关联。在SCZ和BPD样本中找到105个和153个差异表达的基因；研究发现SCZ和BPD中的很多差异表达的基因在表达水平是一致的；选择 $q < 0.05$ 的差异表达基因（213个）进行代谢通路富集分析，发现了18个Pathway被显著富集，其中最显著的Pathway为溶酶体功能及肌动蛋白细胞骨架的调控；

结果展示：

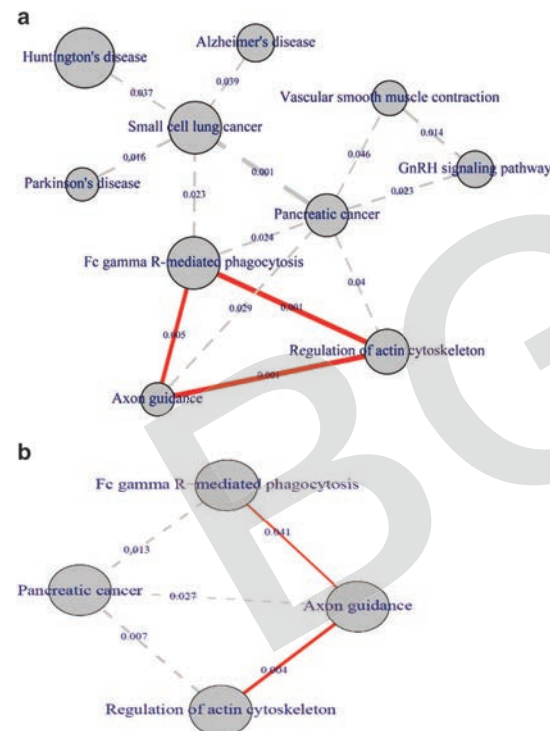


图1 SCZ和BPD样本的DCEG的通路互作网络分析图

参考文献：

Zhao Z, Xu J. et al. Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. Mol Psychiatry. 2015.

## ◆ 常见问题

Q: 转录组研究一般有哪些方法？

A: 基于测序的有cDNA文库、EST文库、SAGE；基于杂交的有cDNA芯片。

Q: 与基因芯片相比，转录组高通量测序有什么优势？

A: 基因芯片是用已知序列的探针和样本mRNA进行杂交来获得mRNA的序列信息的，如果样本mRNA以前从来没有报道过，那么基因芯片上面就不会有相应的探针序列，也就检测不到新的mRNA；另外，核酸杂交的背景噪音很高，存在交叉杂交现象。转录组是直接对样本mRNA进行测序，能够发现很多新的mRNA；转录组测序对基因表达上调或下降的检测范围能够达到几万倍，远比基因芯片的百倍左右要灵敏。而且在有参考基因组的情况下，通过转录组测序您还可以分析可变剪切、基因结构变异、全基因组水平基因表达丰度等情况。

Q: 是否需要生物学重复？重复几次？

A: 是的，至少需要两次生物学重复，3次以上的生物学重复更好。2011年7月Hansen发表的文章表明生物学差异是基因自身表达的特性，与检测技术的选择以及数据处理的方式无关。如果不设生物学重复，高影响因子的杂志可能会因此而拒稿。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
肝癌	2015.02	Cancer Letters	5.02	RNA over-editing of BLCAP contributes to hepatocarcinogenesis identified by whole-genome and transcriptome sequencing
HIV	2014.05	Journal of Virology	5.08	Deep sequencing of HIV infected cells: insights into nascent transcription and host-directed therapy
发育	2014.01	Cell Death and Differentiation	8.39	ADAR1 deaminase contributes to scheduled skeletal myogenesis progression via stage-specific functions
精神类疾病	2014.08	Molecular Psychiatry	15.15	Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder
方法学	2012.02	Nature Biotechnology	39.08	Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome
前列腺癌	2012.02	Cell Research	11.98	RNA-Seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicings

# 全长转录组测序

## 转录组学

### ◆ 产品概述

近年来，随着高通量测序技术的发展，转录组测序已经成为研究基因表达调控的主要手段。但是由于二代测序技术读长的限制，其拼接出来的转录本并不完整，准确性也大打折扣。华大基因基于PacBio单分子实时测序技术，可构建不同插入片段的文库，并借助其超长读长优势，所测即所得、不再需要组装，就可以直接得到高质量的全长转录本的信息。

### ◆ 产品优势

**超长读长：**不需组装就可以准确的得到全长转录本的序列信息

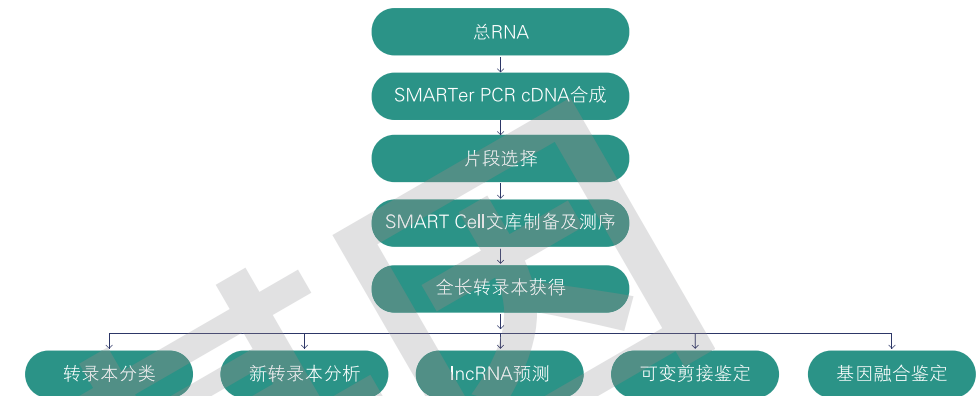
**优化方案：**通过Isoform的解决方案，可以改善基因表达定量的结果

**研究全面：**可以发现新的基因和转录异构体，鉴定可变剪接和基因融合

### ◆ 产品应用



### ◆ 技术流程



### ◆ 分析内容

- 1、过滤接头、低质量及过短的序列
- 2、根据reads读到接头与polyA的情况，区分全长与非全长转录本
- 3、对全长转录本进行聚类，得到一致性序列
- 4、用所有reads对一致性序列进行比对纠错，得到高质量的全长转录本
- 5、与参考基因组序列进行比对
- 6、对多个文库的数据合并去冗，最终得到高质量全长转录本
- 7、转录本分类
- 8、新转录本分析
- 9、lncRNA预测
- 10、可变剪接鉴定
- 11、基因融合鉴定

### ◆ 技术参数

**样品类型：**完整的RNA

**样品总量：**7 $\mu$ g

**样品浓度：** $\geq 300$ ng/ $\mu$ L

**样品纯度：**RIN $\geq 8$ , 28S/18S $\geq 1.4$

**推荐数据量：**3个文库，8个cell；或者4个文库，10个cell

**项目周期：**标准流程的项目周期为60个工作日

## ◆ 案例解析

### ● 案例1 三代单分子测序研究人的转录本结构

开展RNA水平的研究，对于我们了解生物学过程十分重要，但是芯片和基于二代短reads的高通量测序，很难得到完整的5'-3'RNA序列信息。本篇文章通过对20个人体组织和器官的混合样本进行三代单分子测序，得到了476,000个CCS序列，平均长度为1Kb。将这些序列与GENCODE的mRNA序列进行比对，最后鉴定得到了约14,000个转录异构体，其中有10%是前所未有的，研究结果显示基于三代测序的全长转录组对研究转录本结构很有帮助。

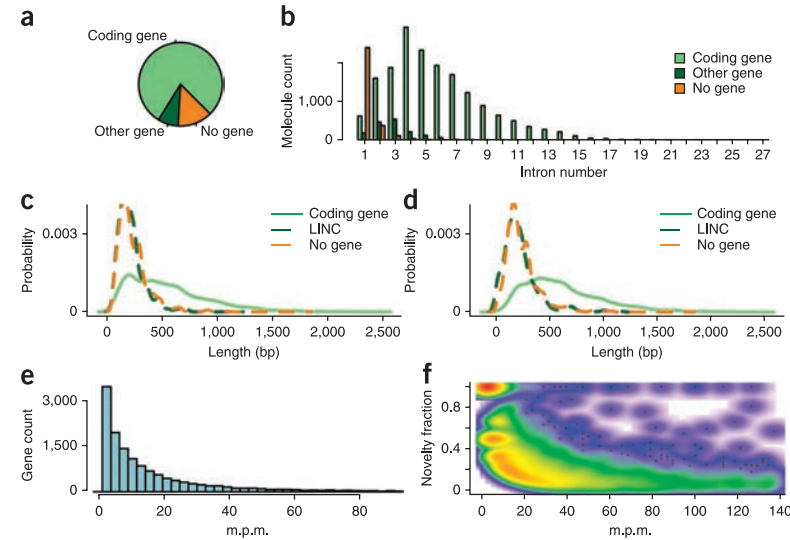


图1. 转录本的结构分析结果展示

参考文献:

Sharon D, Tilgner H, Grubert F, et al. A single-molecule long-read survey of the human transcriptome. Nature biotechnology. 2013.

## ◆ 常见问题

Q: PacBio平台全长转录组产品的优势?

A: 相比于HiSeq平台的转录组产品，PacBio全长转录组不需要组装，就可以得到全长转录本的信息。而基于二代HiSeq平台的转录组测序由于读长的限制，得到的全长转录本的数量、准确度及完整性均大幅下降。

Q: PacBio全长转录组为什么要建3-4个文库?

A: 如果构建1个文库，文库的片段范围是1-10K的话，上机数据大部分都是小片段，大片的转录本所占的比例会非常少，因为小片段更利于掉入ZMW孔中。

Q: PacBio全长转录组推荐的测序的方案?

A: 方案一：3个文库，8个cell；方案二：4个文库，10个cell。

方案1		方案2	
文库大小	数据量	文库大小	数据量
1k-2k	3个cell	1k-2k	3个cell
2k-3k	3个cell	2k-3k	3个cell
3k-6k	2个cell	3k-6k	2个cell
		5k-10k	2个cell
合计：3个文库	合计：8个cell	合计：4文库	合计：10个cell

# RNA-Seq

## 转录组学

### ◆ 产品概述

RNA-Seq是直接对某一物种或特定细胞在某一功能状态下产生的mRNA进行高通量测序，用来研究基因的表达差异情况，已经广泛应用于基础研究、临床研究和药物研发等领域。相比转录组，更加侧重基因定量研究，整体性价比更高。相比表达谱芯片，数字化信号、无背景噪音、无交叉杂交；没有物种限制；能发现未知基因和低丰度基因和新基因。

### ◆ 产品优势

**样品多样：**提供单细胞、极微量、FFPE样品等珍稀或特殊样品的技术服务，满足客户的特殊需求

**平台丰富：**已推出基于HiSeq4000、BGISEQ-500平台的RNA-Seq产品，满足客户的多样化需求

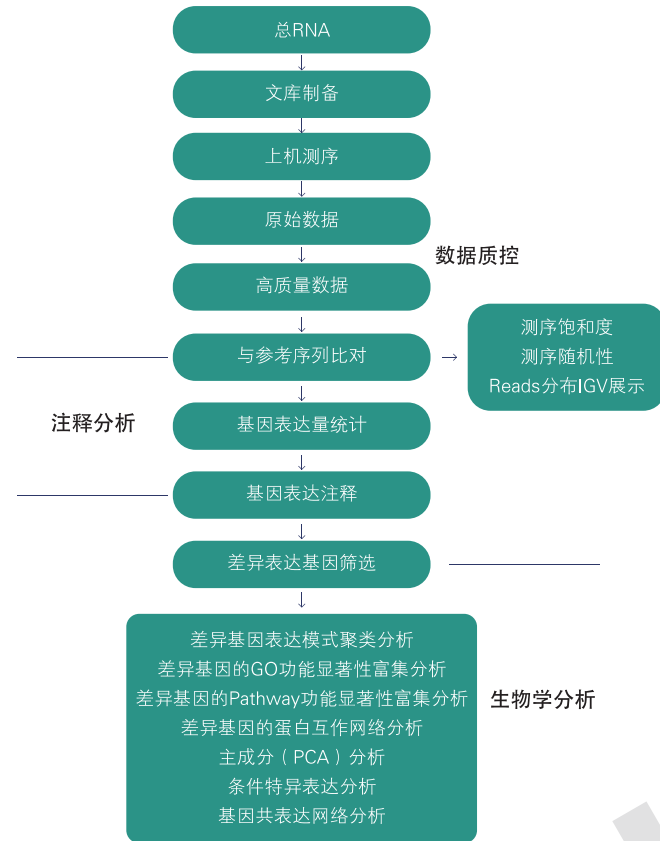
**质检严格：**服务过程中全面实施质量管理，保证所有客户均能获得优质服务

**报告简洁：**结题报告多次升级，做到结果文章化，可在最短时间内拿到极易使用的结果

### ◆ 产品应用



## ◆ 技术流程



## ◆ 分析内容

- 1、去除接头、污染序列及低质量reads
- 2、测序评估（包括：数据比对统计、测序质量评估、测序饱和度分析、测序随机性分析、Reads在参考基因/基因组上的分布）
- 3、基因表达注释（包括：基因覆盖度统计、结果文件列表、结果文件示例）
- 4、差异表达基因分析
- 5、组间差异表达基因分析
- 6、差异基因表达模式聚类分析
- 7、GO功能显著性富集分析
- 8、GO功能分类分析（WEGO分析）
- 9、Pathway显著性富集分析
- 10、蛋白互作网络分析（需要蛋白-蛋白相互作用数据库中有该物种的注释信息）
- 11、主成分（PCA）分析（5个或五个以上样品）
- 12、条件特异表达分析（5个或五个以上样品）
- 13、基因共表达网络分析（15个或以上样品）

## ◆ 技术参数

样品类型：完整且无污染的总RNA  
 样品总量：200ng/次  
 样品纯度：28S:18S $\geq$ 1.0；RIN $\geq$ 7.0  
 测序策略：SE50  
 推荐数据量：HiSeq平台推荐10–20M clean reads，BGISEQ–500平台，推荐20M clean reads  
 项目周期：HiSeq平台为30个工作日，BGISEQ–500平台25个工作日

## ◆ 案例解析

### ● 案例1 最完整的大鼠转录组图谱

研究采用11种器官、4个发育阶段、2种不同性别的320个大鼠样本，利用RNA-seq方法绘制了一张最完整的大鼠转录组图谱。发现大量的转录物显示器官特异性、年龄依赖性或性别特异性的差异表达模式。并构建一个与其他广泛应用的数据库互联、可通过网络开放获取的大鼠BodyMap数据库。

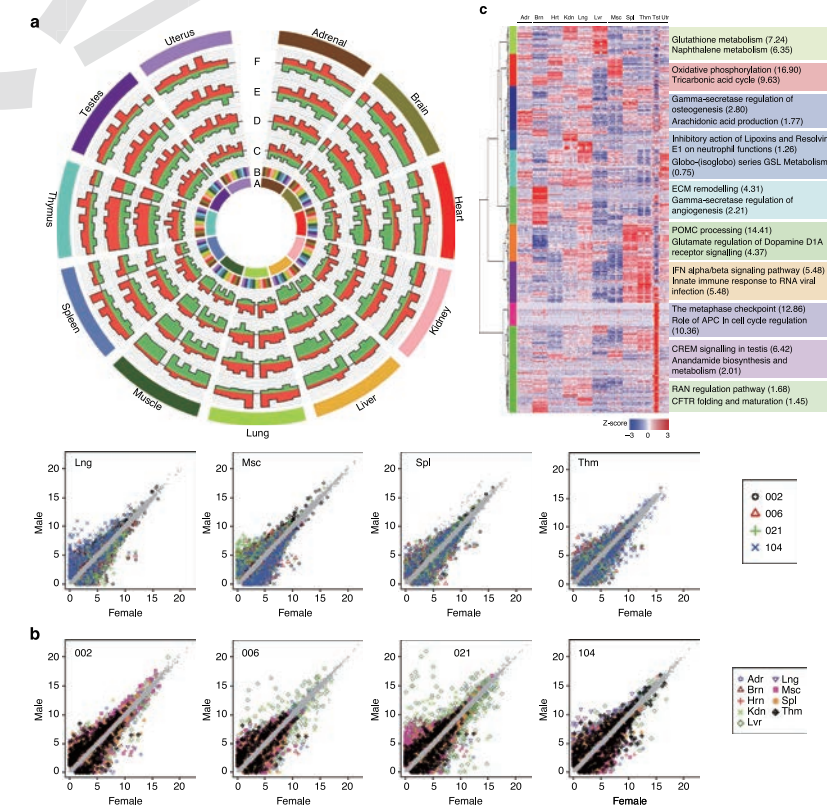


图1. 器官特异性表达基因、年龄特异性表达基因和性别特异性表达基因

参考文献：

Ying Yu, James C. Fuscoe, Chen Zhao, et al. A rat RNA-seq transcriptomic BodyMap across 11 organs and 4 developmental stages. Nature Communications. 2014.

## ◆ 常见问题

Q: RNA-Seq必须要有参考基因组序列吗?

A: 需要有参考序列而非参考基因组序列, unigene序列、mRNA序列、CDS序列等可以作为参考序列。

Q: 如何对差异表达基因进行功能分析?

A: 差异表达基因可以进行以下4种功能分析:

- 1、表达模式聚类分析, 即将具有相似表达模式的差异基因聚到一起, 从而筛选出感兴趣的基因。
- 2、GO分析, 该分析能确定差异基因行使的主要生物学功能分类, 发掘与基因差异表达现象关联的多个特征功能类。
- 3、Pathway 分析, 确定差异表达基因参与的最主要的生化代谢途径和信号转导途径。
- 4、蛋白-蛋白相互作用网络分析, 即找出跟差异表达基因编码蛋白发生直接相互作用的蛋白。

Q: RNA-Seq推荐测序数据量与基因组大小有关吗?

A: RNA-Seq 推荐测序数据量, 主要与基因的数量有关, 不同物种基因组大小相差比较大, 但是编码基因的数量相差并不大, 一般物种在3万左右, 所以, 对于一般物种RNA-Seq推荐10M clean reads数据量。

# Small RNA测序

转录组学

## ◆ 产品概述

Small RNA是生物体内一类重要的功能分子, 主要包括miRNA、siRNA和piRNA。它们通过各种序列特异性的基因沉默作用, 包括RNA干扰(RNAi)、翻译抑制、异染色质形成等, 诱导基因沉默, 调控诸如细胞生长发育、应激反应、沉默转座子等各种各样的细胞生理过程。Small RNA测序技术采用胶分离技术, 收集样品中18~30nt的RNA片段, 利用高通量测序技术, 一次性获得单碱基分辨率的数百万条Small RNA序列信息, 依托强大的生物信息分析平台, 鉴定已知Small RNA, 并预测新的Small RNA及其靶标基因。

## ◆ 产品优势

**样本多样:** 真核生物、原核生物、血清血浆、FFPE样品

**结果精准:** 3款可选靶基因和miRNA预测软件, 提高结果可靠性和准确性; 3种权威数据库用于Small RNA分类注释

**分析内容丰富:** 组间样品差异分析、生物学重复分析、piRNA、siRNA分析及靶基因预测等

**报告简洁:** 结果图片更加直观, 可用于文章撰写; 结题报告结构优化, 有效提升阅读体验

**平台多样:** 可选择HiSeq平台和BGISEQ-500平台

## ◆ 产品应用



## ◆ 技术流程



## ◆ 信息分析内容

- 1、数据产出统计，对原始信息采集数据去接头污染，去低质量reads
- 2、18–30 nt Small RNA信息采集结果的长度分布
- 3、Small RNA在选定的参考基因组上的分布（只能选定一个参考基因组）
- 4、Small RNA分类注释
  - ① Small RNA与rRNA、tRNA、snRNA、snoRNA的比对信息
  - ② Small RNA与重复序列的比对信息
  - ③ Small RNA与exon/intron的比对信息
  - ④ Small RNA与miRBase中指定范围的已知的miRNA的比对
  - ⑤ miRNA家族分析
  - ⑥ 动物保守piRNA注释分析（仅针对数据库中已有的物种：人、小鼠、斑马鱼、线虫、果蝇、大鼠、鸡、家蚕）
- 5、Small RNA预测（miRNA/siRNA/piRNA）
- 6、已知和新Small RNA定量分析（miRNA/siRNA/piRNA）
- 7、已知和新Small RNA差异表达分析（miRNA/siRNA/piRNA）
- 8、已知和新miRNA表达/差异表达聚类分析
- 9、Small RNA靶基因预测
- 10、差异Small RNA靶基因GO注释和KEGG通路分析

## ◆ 技术参数

**样品类型：**总RNA样品、200nt以内的小片段RNA样品、经胶回收的small RNA样品；组织培养与细胞系样品、血清/血浆样品、组织液样品，FFPE样品、RIP样品等

**样品总量：**≥100ng/次

**测序策略：**SE50

**推荐数据量：**10M clean reads

**项目周期：**标准流程的项目周期为30个工作日

## ◆ 案例解析

### ● 案例1 Small RNA测序鉴定结核病的血清microRNA分子标记物

通过高深度测序，对15个活动性肺结核患者、82个健康的志愿者共97人的血清共4组样本，检测到了904种miRNA，其中162种miRNA在活动性结核病人血清中差异表达显著，hsa-miR-516b、hsa-miR-196b和hsa-miR-376c显著上调，hsa-miR-486-5p显著下调，其中hsa-miR-196b和hsa-miR-376c的RT-PCR验证结果显示，与测序结果一致，表明这2个miRNA很可能是潜在的诊断标记。

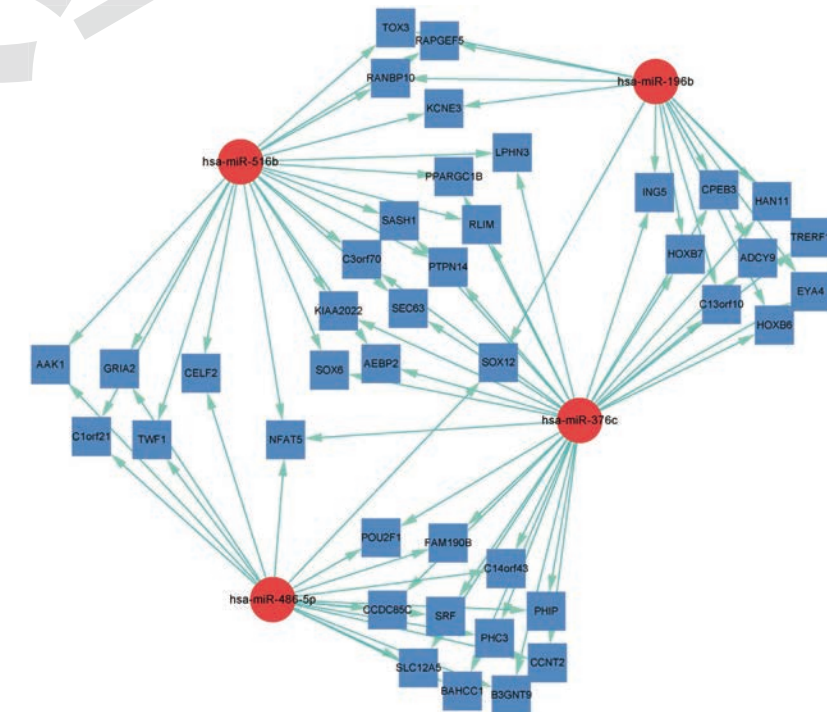


图1. 已鉴定的候选生物标记microRNA (hsa-miR-196b, hsa-miR-516b, hsa-miR-376c和hsa-miR-486-5p) 和靶基因的共表达网络图

参考文献：

Zhang H, Sun Z, et al. Identification of Serum microRNA Biomarkers for Tuberculosis Using RNA-seq. PLOS ONE. 2014.

## ◆ 常见问题

Q: Small RNA可以操作什么类型的样品建库测序?

A: 人、鼠总RNA样品、微生物总RNA样品、200nt以内的小片段RNA样品、经切胶回收的small RNA样品、组织培养与细胞系样品、血清/血浆样品、组织液样品, FFPE样品、RIP样品等, 只要满足华大送样要求, 均可在华大用于Small RNA建库测序。

Q: 进行Small RNA测序对于组织样品提取总RNA有什么特殊要求?

A: 提取总RNA时建议不要使用Qiagen等公司的过柱试剂盒, 也不要使用LiCl沉淀, 以免丢失小片段RNA。如果直接提供Small RNA样品, 可以使用Small RNA提取专用试剂盒来进行提取。

Q: miRNA验证方法?

A: 茎环实时定量PCR-Stem-loop quantitative real-time PCR (qRT-PCR)、Quantitative PCR (qPCR)、qRT-PCR等

Q: Small RNA分析需要合作伙伴提供参考序列么?

A: 需要, 最好是全基因组序列, 如果没有, 也可以提供近源物种的全基因组序列或者本物种的EST等序列作为参考序列, 另外还需要提供相关的exon/intro, repeat信息以及基因编码序列。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
人Hela细胞	2016.05	BMC Genomics	3.87	Fluorescence activated cell sorting followed by small RNA sequencing reveals stable microRNA expression during cell cycle progression
人鼻咽癌	2016.01	Journal of experimental & Clinical Cancer Research	1.23	Integrated analysis of microRNA regulatory network in nasopharyngeal carcinoma with deep sequencing
免疫性血小板减少症	2015.08	Biomedicine & Pharmacotherapy	2.02	Profiling of miRNA expression in immune thrombocytopenia patients before and after Qishunbaolier (QSBLE) treatment
人结肠直肠癌	2014.11	PLOS ONE	3.53	Deep sequencing of RNA from three different extracellular vesicle (EV) subtypes released from the human LIM1863 colon cancer cell line uncovers distinct miRNA-enrichment signatures.
大鼠肝纤维化	2014.07	Annals of Hepatology	2.07	Deep sequencing analysis of microRNA expression in porcine serum-induced hepatic fibrosis rats
人结核病	2014.02	PLOS ONE	3.53	Identification of Serum microRNA Biomarkers for Tuberculosis Using RNA-seq

# 长链非编码RNA测序

## 转录组学

## ◆ 产品概述

长链非编码RNAs (long non-coding RNAs, lncRNAs) 是一类长度大于200nt且不编码蛋白质的RNA, ENCODE计划发现大约75%的人类基因组能被转录成RNAs, 而74%则是非蛋白编码RNA (ncRNAs), 随着高通量测序技术在基因组学研究中广泛应用, 越来越多的非编码RNA作用被深入挖掘, 包括: 参与了X染色体沉默, 基因组印记以及染色质修饰, 转录激活, 转录干扰, 核内运输等多种重要的调控过程, 涉及到表观遗传调控、转录调控以及转录后调控等多个层面。lncRNA高通量测序是指使用特定方法对样本中大于200nt的非编码RNA进行测序研究, 从而快速全面准确地获得与特定生物学过程 (例如发育、疾病等) 相关lncRNA信息的研究方法。

## ◆ 产品优势

**技术稳定:** 同一样本重复建库相关性Pearson值大于0.993

**信息全面:** 基于高通量测序技术, 一次性获得样本中几乎全部的lncRNA信息

**高性价比:** 不局限于对已知lncRNA的研究, 还可进行novel lncRNA预测

**经验丰富:** 已完成万例lncRNA项目, 涉及人鼠及动植物物种

**高准确性:** 采用多重分析手段, 准确度高

## ◆ 产品应用





## ◆ 技术流程



## ◆ 分析内容

### 标准信息分析

#### 1、基本数据统计:

- ① 去除接头序列、低质量序列得到clean data
- ② 与核糖体数据库比对, 去除核糖体RNA数据

#### 2、lncRNA & mRNA鉴定

- ① 转录本组装
- ② 鉴定已知转录本 (包括已知lncRNA, mRNA)
- ③ 预测新转录本 (包括新的lncRNA, mRNA)

#### 3、lncRNA & mRNA定量分析

- ① 已知和新的lncRNA & mRNA定量分析
- ② 样本间差异分析 (至少2个样本)
- ③ 组间差异分析 (至少2组样本, 每组至少3个生物学重复)
- ④ lncRNA & mRNA 表达/差异表达聚类分析

- ⑤ lncRNA & mRNA 表达时间序列分析 (至少3个时间点的样本, 或3种不同处理的样本)

#### 4、lncRNA功能预测

- ① lncRNA靶基因预测
  - ② miRNA前体预测
  - ③ lncRNA家族预测
- #### 5、mRNA功能注释
- ① 数据库注释
  - ② 差异表达mRNA的GO和Pathway富集分析
- #### 6、mRNA结构分析
- ① 可变剪切分析
  - ② SNP/InDel分析

### 高级信息分析

- 1、lncRNA靶基因GO和Pathway富集分析
- 2、lncRNA及靶基因互作网络分析

## ◆ 技术参数

样品类型: 总RNA样品; 组织培养与细胞系样品、FFPE样品等  
 样品总量:  $\geq 200\text{ng}$ /次  
 测序策略: PE101  
 推荐数据量: 8–10G clean data  
 项目周期: 标准流程的项目周期为50个工作日

## ◆ 案例解析

### ● 案例1 研究对高低风险神经母细胞瘤lncRNA差异表达

通过高深度测序对3例 (1个亚型) 低风险神经母细胞瘤、12例 (2个亚型) 高风险神经母细胞瘤进行lncRNA建库和测序, 鉴定到了一个神经母细胞瘤相关的lncRNA转录本NBAT-1, 可以作为biomarker来预测神经母细胞瘤的临床结果。

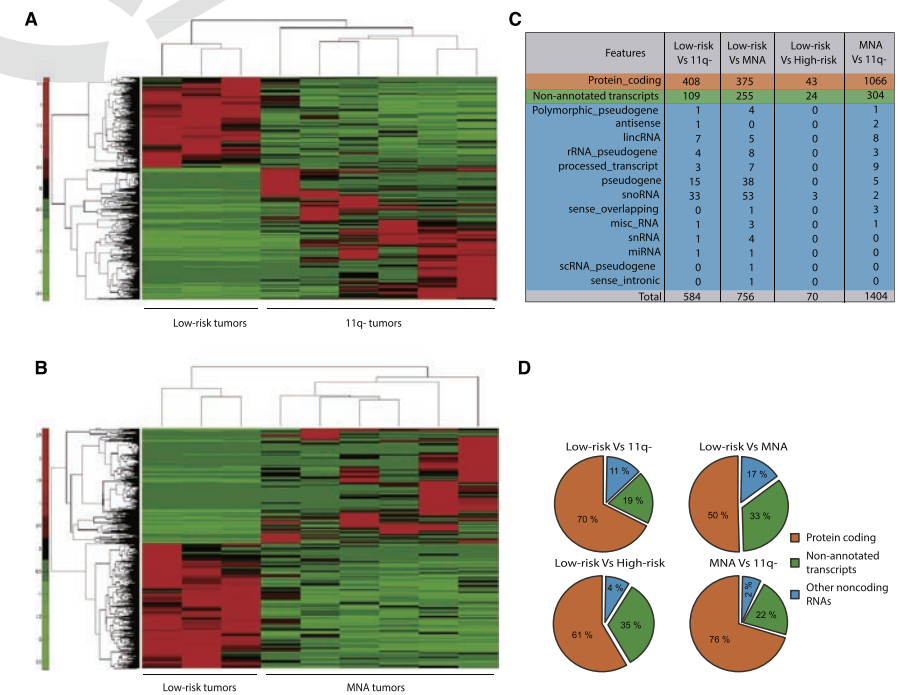


图1. 神经母细胞瘤不同亚型的特异的转录本差异表达  
 (A)(B)不同亚型肿瘤的表达模式聚类分析;  
 (C)不同亚型肿瘤编码和非编码转录本差异表达统计表;  
 (D)低风险和高风险的肿瘤编码和非编码转录本表达比例饼状图

参考文献:

Pandey GK, Mitra S, et al. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. Cancer Cell. 2014.

# 3 表观组学

## ◆ 常见问题

Q: 什么是长链非编码RNA? 以及它们的作用?

A: 哺乳动物基因组序列的约5%~10%被稳定转录, 蛋白质编码基因仅约占1%, 其余4%~9%的序列都转录为非编码RNA。而非编码RNA (non-coding RNA) 是指不能翻译为蛋白的功能性RNA分子。长链非编码RNA为这4%~9%中长度大于200nt的非编码RNA。它们的作用主要体现在四个方面: 第一, 影响周边基因的表达; 第二, 调控蛋白质活动及定位; 第三, 产生小分子RNA; 第四, 对其他RNA的调控作用。

Q: 长非编码RNA与芯片方法对比增加了哪些内容?

A: 与芯片方法相比增加了novel lncRNA预测; 同时可选择特异富集lncRNA。

Q: 长非编码RNA的建库方案是什么?

A: 长非编码RNA建库主要采用ribozero kit去除rRNA, 建库特异性文库, 目前建库比较稳定。

Q: lncRNA测序的数据中是否也含有 mRNA?

A: 是的。所以我们推荐在做lncRNA测序时, 也可利用同一套测序数据进行mRNA的分析。

Q: 怎么用分子生物学实验方法来验证分析结果?

A: 因为转录本组装的复杂性, 我们推荐使用传统的克隆或PCR的方法验证分析结果, 差异表达的lncRNA也可以用RT-PCR的方式进行验证。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
恒河猴	2015.11	Molecular BioSystems	3.2	Spatiotemporal-specific lncRNAs in the brain, colon, liver and lung of macaque during development
生殖细胞	2015.04	Database (Oxford)	3.37	Germ lncRNA: a unique catalogue of long non-coding RNAs and associated regulations in male germ cell development
神经卵母细胞瘤	2014.11	Cancer cell	23.90	The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation
前列腺癌	2012.01	Cell research	9.42	RNA-Seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicing

- ◆ 全基因组甲基化测序
- ◆ 目标区域捕获甲基化测序
- ◆ ChIP-Seq

# 全基因组甲基化测序

## 表观组学

### ◆ 产品概述

DNA甲基化是重要的表观遗传学标记，获得全基因组范围内所有C位点的甲基化水平数据，对于表观遗传学的时空特异性研究具有重要意义。全基因组甲基化测序将Bisulfite处理与高通量测序技术相结合，能够高效准确地绘制全基因组DNA甲基化图谱，实现高精度甲基化修饰模式的分析，在表观基因组学研究中具有里程碑式的意义。该方法能够广泛应用于细胞分化、组织发育等基础机制研究，以及动植物育种、人类健康与疾病等应用性研究。

### ◆ 产品优势

- 检测精准：单碱基分辨率，精确分析每一个C碱基的甲基化状态
- 高性价比：相对于传统的PCR+Sanger测序方法，费用较低
- 高效率：借助新一代高通量测序平台对全基因组5mC信息进行高效扫描

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

#### 标准信息分析

- 1、数据基本处理与质控
- 2、全基因组甲基化水平分析
- 3、甲基化C碱基中CG, CHG与CHH的分布比例
- 4、甲基化CG、CHG和CHH的甲基化水平分布
- 5、甲基化的CG、CHG、CHH附近碱基的序列特征分析
- 6、染色体水平的甲基化C碱基密度分布
- 7、基因组的不同区域的甲基化分布特征
- 8、基因组不同转录元件中的DNA平均甲基化水平
- 9、DMR的检测
- 10、DMR相关基因的GO和Pathway分析

### ◆ 技术参数

- 样品类型：DNA无降解，需提供凝胶电泳检测胶图
- 样品总量： $\geq 2\mu\text{g}$
- 样品纯度： $\text{OD}_{260/280} = 1.8 - 2.0$ ,  $\text{A}_{260}/\text{A}_{230} \geq 1.6$ ；没有蛋白、多糖和RNA污染
- 样品溶剂：建议样品用高质量试剂盒提取，溶于ddH<sub>2</sub>O。确保溶剂内不含有影响酶活的酒精，苯酚，氯仿或其它有机溶剂
- 测序策略：PE101/PE151
- 推荐数据量： $\geq 30\text{X}$
- 项目周期：标准流程的项目周期为35个工作日

## ◆ 案例解析

### ● 案例1 全基因组甲基化测序分析年龄与DNA甲基化的关系

人类的寿命不仅仅由遗传决定，环境因素可以通过改变表观遗传修饰对衰老的过程产生不可忽视的影响。本文通过WGBS技术比较了一位103岁老年男子和一名新生男婴的DNA样本，并在大量样本中进行了验证，最终发现与同样细胞类型的新生儿相比，整体上百岁老人DNA中有更多的非甲基化。

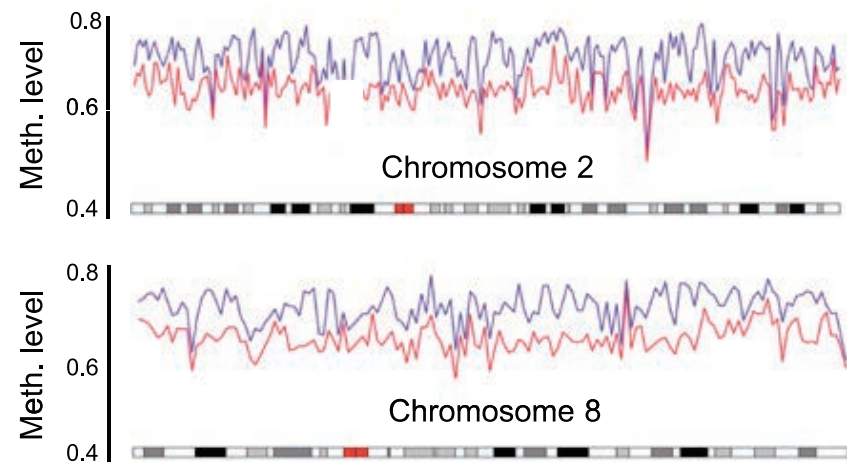


图1.百岁老人（红色）与新生儿（蓝色）的甲基化分布

参考文献:

Heyn H, Li N, Ferreira HJ, et al. Distinct DNA methylomes of newborns and centenarians. Proc Natl Acad Sci USA. 2012.

## ◆ 常见问题

Q: Bisulfite-Seq在项目开始之前需要考虑哪些因素?

A: 是否为低甲基化率的物种; 该物种的基因组完成情况如何(影响Bisulfite-Seq的比对); 基因组是否存在复杂因素, 如GC含量偏高、杂合度偏高、转座子、重复区域等。

Q: 可以对无参考基因组的物种进行Bisulfite研究吗?

A: Bisulfite-Seq强烈依赖基因组的完成程度, 基因质量的好坏直接影响后续的分析结果, 因此更适合有完整基因组信息的物种。

Q: 如果合作伙伴自己建库, 能否提供上机测序? 如何进行文库质量检测并保证测序质量?

A: 合作伙伴自己建的文库, 可以上机测序。如果用标准的Illumina kit, 要注明kit的类型及货号。如果不是使用Illumina kit建库, 那么合作伙伴提供信息单的同时必须说明建库方法、使用的试剂盒及品牌, 以及提供建库所用的接头引物序列和预期文库片段大小。如果是index文库, 要注明index的位置以及index的序列。若只完成部分建库过程请务必注明清楚; 如果文库是PCR产物建库或者插入片段中有特定序列, 请在提供的样品信息单中说明。否则会极大地影响数据质量。对合作伙伴文库的检测, 首先用Agilent 2100 Bioanalyzer 确定片段大小是否符合; 然后通过Q-PCR精确定量确定上机体积。

Q: Bisulfite的转化率是多少?

A: Bisulfite转化率达到99%以上; 如果样品DNA不含有未发生甲基化的DNA作为对照, 都会在样品中混入control DNA来验证Bisulfite的转化率。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
年龄	2012.06	PNAS	9.74	Distinct DNA methylomes of newborns and centenarians
方法学	2010.11	PLOS Biology	12.47	The DNA Methylome of Human Peripheral Blood-Mononuclear Cells

# 目标区域捕获甲基化测序

## 表观组学

### ◆ 产品概述

采用Agilent公司/Nimblegen公司的目标区域捕获kit，对全基因组范围内特定区域进行捕获，并结合Bisulfite转化测序的方法，对捕获区域内的C位点进行甲基化水平进行检测的一种技术，结合目标区域捕获和Bisulfite测序的方法，不仅能够在很大程度上降低研究的成本，还能够提高目标区域的测序深度，从而增加检测准确性，获得检测区域单碱基分辨率的甲基化信息，该技术使得对全基因组内感兴趣的目标区域进行高通量高精度的甲基化检测成为现实，在甲基化组学研究方面提供了一个全方位，稳定高效，高性价比的技术方法。

### ◆ 产品优势

**覆盖位点更多：**利用成品kit能获得比甲基化芯片更多的CpG位点

**检测范围更广：**除了覆盖已知的DMR区域之外，还能发现新的DMR区域

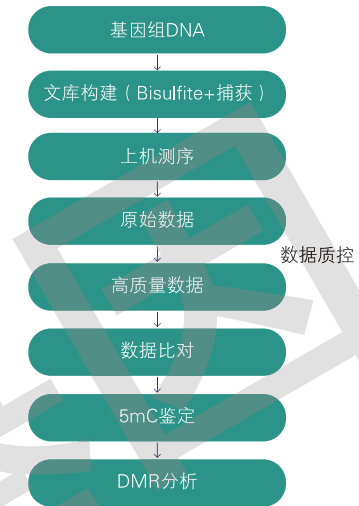
**测序分辨率高：**精确到单碱基分辨率，与已发表的全基因组甲基化测序数据有高度一致性

**技术稳定：**目标区域富集、测序深度分布、覆盖度的重复性高

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

#### 标准信息分析

- 1、数据基本处理与质控
- 2、C碱基的甲基化水平分析
- 3、mCG、mCHG和mCHH的比例
- 4、mCG、mCHG和mCHH的甲基化水平分布
- 5、不同基因组区域的甲基化分布特征
- 6、基因甲基化分析
- 7、DMR的计算统计
- 8、DMR相关基因的GO和Pathway分析

### ◆ 技术参数

**样品类型：**DNA无降解，需提供凝胶电泳检测胶图

**样品总量：**≥3μg

**样品纯度：**OD260/280 = 1.8 – 2.0，A260/A230 ≥ 1.6；没有蛋白、多糖和RNA污染

**样品溶剂：**建议样品用高质量试剂盒提取，溶于ddH<sub>2</sub>O。确保溶剂内不含有影响酶活的酒精，苯酚，氯仿或其它有机溶剂

**测序策略：**PE101/PE151

**推荐数据量：**≥100X

**项目周期：**标准流程的项目周期为35个工作日（不含kit采购周期，采购周期约2个月）

## ◆ 案例解析

### ● 案例1 目标区域捕获甲基化测序研究肝细胞癌的发生机理

DNA甲基化已被验证与癌症的发生有着密切的关系。而在肝细胞癌中全基因组的甲基化分析刚刚开始，为了更深入的研究肝细胞癌的发生机制，本文利用华大设计的启动子捕获kit对8对肝细胞癌及其癌旁组织进行了甲基化水平的分析，并寻找与标量异常基因相关的差异甲基化区域，发现12个差异甲基化相关的基因为差异表达基因；然后用BSP技术在另外的78对样品中验证这些差异甲基化区域。BSP验证了7个基因，其中5个基因在肝细胞癌中表现为高甲基化，2个表现为低甲基化。

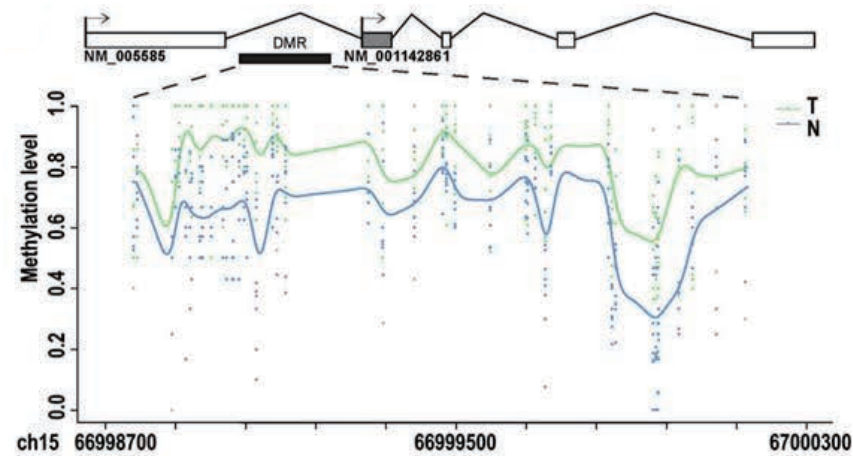


图1. 癌组织 (T) 与癌旁组织 (N) 甲基化水平分布

参考文献:

Gao F, Liang H, et al. Global analysis of DNA methylation in hepatocellular carcinoma by a liquid hybridization capture-based bisulfite sequencing approach. Clin Epigenetics. 2015.

## ◆ 常见问题

Q: 目标区域捕获甲基化在项目开始之前需要考虑哪些因素?

A: 首先要考虑选用哪种捕获kit, 目前成品Kit有三种, 华大自主设计的Promoter捕获kit (BGI Promoter Kit), 以及另外两种商业成品捕获kit ( Nimblegen SeaCap CpGiant Enrichment Kit /Agilent SureSelectXT Human Methyl-seq), 客户也可完全定制捕获Kit对感兴趣的目标区域进行研究, 产品选择多, 可满足客户多样的研究需求。

Q: 可以对无参考基因组的物种进行研究吗?

A: 甲基化分析强烈依赖基因组的完整程度, 基因质量的好坏直接影响后续的分析结果, 因此更适合有完整基因组信息的物种。

Q: 推荐的数据量?

A: 建议为50X以上深度 (约100X区域大小)。

Q: 目标区域捕获kit订购需要多长时间?

A: 一般成品kit订购大约需要2个月, 全定制的捕获kit大约需要3个月。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
癌症	2015.08	Clin Epigenetics	4.54	Global analysis of DNA methylation in hepatocellular carcinoma by a liquid hybridization capture-based bisulfite sequencing approach.
癌症	2014.03	Technol Cancer Res Treat	1.73	Clustering of Cancer Cell Lines Using A Promoter-Targeted Liquid Hybridization Capture-Based Bisulfite Sequencing Approach
方法学	2011.12	BMC Genomics	4.07	High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing

# ChIP-Seq

## 表观组学

### ◆ 产品概述

染色质免疫共沉淀 (ChIP) 是在体内环境中研究蛋白质与DNA相互作用的经典实验方法, 广泛应用于组蛋白修饰、特定转录因子的基因调控作用等相关领域。随着新一代测序技术的发展和成熟, 染色质免疫沉淀实验与高通量测序的整合——Chromatin Immunoprecipitation Sequencing (简称ChIP-Seq), 可在全基因组范围对蛋白结合位点进行高效而准确的筛选与鉴定, 同时也为进一步的深入研究打下基础。ChIP-Seq技术采用特异性抗体对目标蛋白进行免疫沉淀, 分离出与蛋白结合的基因组DNA片段, 通过高通量测序与数据分析, 在全基因组范围内寻找与目标蛋白结合的DNA位点, 并可基于多个样品进行差异比较。

### ◆ 产品优势

平台丰富: 提供BGISEQ-500、HiSeq4000等多种测序平台选择

高性价比: 基于抗体富集目标区域, 降低测序数据量及测序成本

检测范围广: 可对全基因组范围内的DNA与蛋白相互作用区域进行测序分析

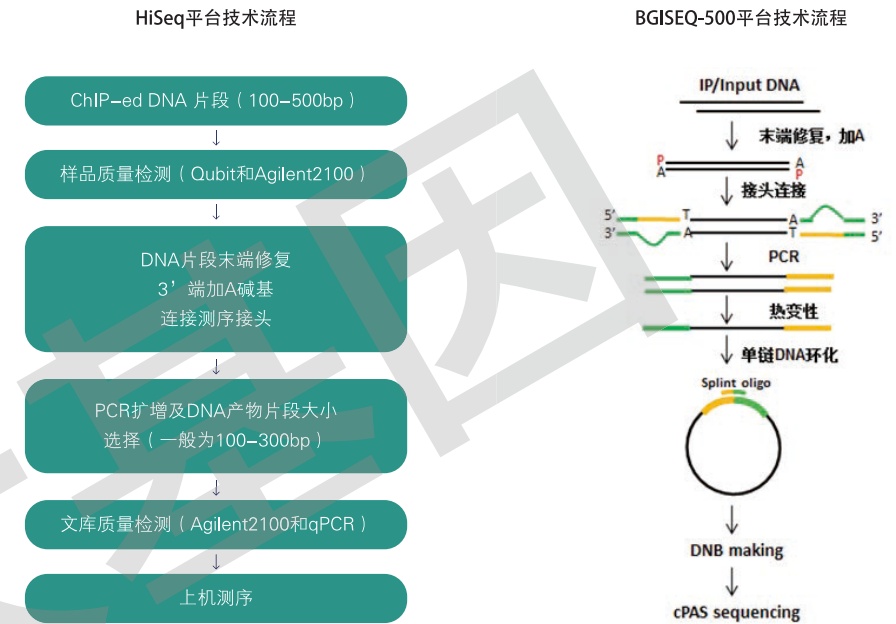
微量样品: BGISEQ-500平台最低起始量仅需5ng, 5ng以下亦可做定制化

关联分析: 可定制ChIP-Seq与RNA-Seq差异表达基因 (DEGs) 关联分析

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

#### 标准分析

- 1、数据基本处理与质控
- 2、基因组测序深度累积分布
- 3、Peak分析 (Peak扫描, Peak长度分布, Peak深度分布)
- 4、Peak在基因功能元件上的分布
- 5、Peak相关基因
- 6、Peak相关基因的GO功能显著性富集分析
- 7、Peak相关基因的pathway富集分析
- 8、样品间差异peak
- 9、样品间差异peak在基因功能元件的分布
- 10、样品间差异peak相关基因
- 11、样品间差异peak相关基因的GO和KEGG富集分析

#### 高级分析

Motif分析

#### 定制分析

RNA-Seq差异表达基因 (DEGs) 与ChIP-Seq peaks关联分析

## ◆ 技术参数

样品类型: ChIP-ed DNA样品 (未经PCR扩增)

样品总量:  $\geq 10\text{ng}$

样品浓度:  $\geq 1\text{ng}/\mu\text{L}$

样品纯度:  $\text{OD}_{260/280} = 1.8 - 2.0$

DNA片段大小: 分布在100~500bp范围, 且主带明显。请提供DNA打断后的检测胶图以确定DNA片段大小是否符合要求;

并请附加一份详细的样品信息单, 并提供ChIP后的q-PCR验证结果。

测序策略: SE50

推荐数据量: 20M或40M高质量reads

项目周期: 标准流程的执行周期为30个工作日

## ◆ 案例解析

### ● 案例1 索拉菲尼抑制上皮-间充质转化的表现机理

长期以来, 科学家发现上皮-间充质转化在肿瘤发生转移过程是必不可少的步骤, 而DNA修饰和组蛋白修饰的重编程在这个过程中基因表达的调控起到了十分重要的作用。索拉菲尼是第一个口服治疗癌症的药物, 它能阻断上皮-间充质转化从而抑制肿瘤的生长。本文以A549肺癌细胞系为研究对象, 通过ChIP-Seq的方法, 研究了癌细胞在索拉菲尼药物的作用下, 其组蛋白甲基化、乙酰化等修饰改变对基因表达调控的作用, 从而阐述了该药物在阻断上皮-间充质转化过程中的功能。

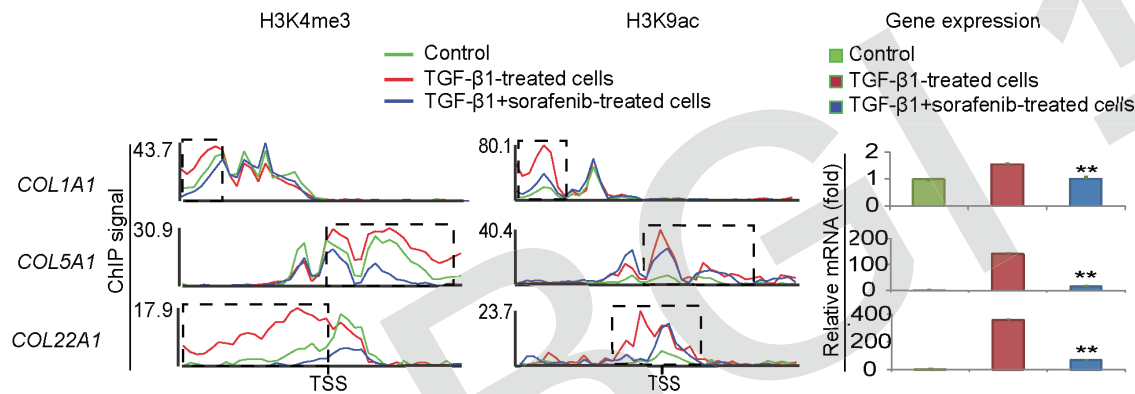


图1. 索拉菲尼能够通过影响组蛋白修饰从而调节上皮-间充质转化相关基因的表达

参考文献:

Zhang J, Chen YL, et al. Sorafenib inhibits epithelial-mesenchymal transition through an epigenetic-based mechanism in human lung epithelial cells. PLOS ONE. 2013.

## ◆ 常见问题

Q: N-ChIP和X-ChIP的区别是什么?

A: N-ChIP基于内切酶micrococcal nuclease (MNase) 酶切, 切割核小体, 适用于组蛋白的研究; X-ChIP基于化学交联, 适用于大多数研究情况。

Q: ChIP-Seq是否需要做阴性对照测序?

A: 一般情况下, 建议选择Input作为对照进行测序。

Q: 样品制备过程中是否需要PCR扩增? PCR扩增后是否会影响最终结果?

A: 由于ChIP下来的DNA样品量通常非常少, 所以在样品建库制备过程中需要经过一步PCR扩增, 主要是为了获得足够上机反应的DNA量。如果提供的DNA样品足量, 则减少PCR循环数或不进行PCR扩增。PCR扩增可能会增加结果的偏向性。

Q: 影响ChIP-Seq结果的因素有哪些?

A: 抗体的质量与特异性、目标区域在基因组上的比例、ChIP的实验操作、DNA片段长度范围等都会影响ChIP-Seq的结果。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
肺癌	2013.05	PLOS ONE	3.53	Sorafenib inhibits epithelial-mesenchymal transition through an epigenetic-based mechanism in human lung epithelial cells
组蛋白修饰图谱	2011.11	PLOS ONE	4.10	Global mapping of H3K4me1 and H3K4me3 reveals the chromatin state-based cell type-specific gene regulation in human Treg cells.
方法学	2011.06	Nature Methods	19.28	Single-tube linear DNA amplification (LinDA) for robust ChIP-seq



# 4 蛋白质组学

- ◆ 蛋白质鉴定分析
- ◆ 蛋白质定量分析
- ◆ 蛋白质修饰分析

## 蛋白质鉴定分析

### 蛋白质组学

#### ◆ 产品概述

蛋白质鉴定分析 (Protein Identification)，即利用质谱的方法对细胞、组织、体液，或者经过分离、纯化、富集后的蛋白胶/液等类型样本中的全部或特定的蛋白进行蛋白种类鉴定，包括蛋白全谱鉴定，胶点胶条和蛋白混合物鉴定等。充分了解特定环境下蛋白质种类情况是进行蛋白质组学研究的基础，为研究疾病发生发展过程以及动植物生长发育状态提供直观的蛋白质种类鉴定结果。

#### ◆ 产品优势

**仪器精准：**选用高分辨率（大于 $10^5$ ），高质量精度（小于1 ppm）质谱仪，确保鉴定结果准确性。

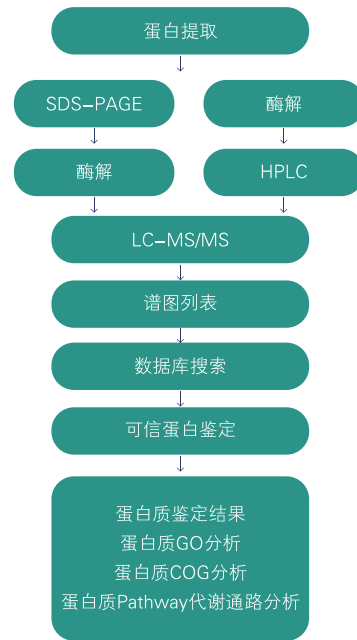
**检测广泛：**无物种特异性限制，理论上可适用于所有物种。

**经验丰富：**根据具体的研究对象，制定个性化的实验策略，帮助提高蛋白鉴定率，节省成本。

#### ◆ 产品应用



## ◆ 技术流程



## ◆ 分析内容

- 1、数据产出统计
- 2、蛋白质鉴定结果
- 3、蛋白质GO分析
- 4、蛋白质COG分析
- 5、蛋白质Pathway代谢通路分析

(注：双向电泳胶点或纯化后的唯一蛋白鉴定信息分析仅限前两条分析内容)

## ◆ 技术参数

样品要求：

服务类型	蛋白量	其他类型样本
蛋白全谱	≥500μg, 浓度≥1μg/μL	组织200mg; 细胞3-5 × 10 <sup>6</sup> (沉淀体积约为30-50μL左右); 血液(限血清、血浆) 500μL; 其他体液根据实际情况评估, 一般要求大于5mL
胶条鉴定	≥30μg, 浓度≥1μg/μL	胶条样本考染或银染肉眼可见
混合物鉴定	≥30μg, 浓度≥1μg/μL	样本跑胶后考染或银染肉眼可见
胶点鉴定		胶点样本考染或银染肉眼可见

注：蛋白全谱鉴定富含杂质或蛋白质含量低的样品量干重要求≥5g; 银染胶不能含有戊二醛, 且须在电泳后一周内切胶送样。

服务类型	研究对象	实验策略	质谱仪	项目周期
蛋白全谱	组织、细胞、体液样本中的全部蛋白种类	将整个组织、细胞、体液样本中的蛋白质组提取, 利用胶分离或液相色谱分离技术将蛋白质组分分成多个组分后, 每组单独进行质谱(LC-MS/MS)鉴定	Q-Exactive/5600/ LTQ-OrbitrapVelos	40
胶条鉴定	仅包括胶条内的蛋白, 一般包含几种到几百种蛋白	胶条酶解后质谱(LC-MS/MS)鉴定	Q-Exactive/5600/ LTQ-OrbitrapVelos	20
混合物鉴定	仅包括混合物内的蛋白, 一般包含几种到几百种蛋白	混合物酶解后质谱(LC-MS/MS)鉴定	Q-Exactive/5600/ LTQ-OrbitrapVelos	20
胶点鉴定	仅包括双向电泳胶点中的唯一蛋白	胶点酶解后质谱(LC-MS/MS或者MALDI-TOF/TOF)鉴定	LTQ-Orbitrap/ ultrafleXtreme	15

## ◆ 案例解析

### ● 案例1 利用蛋白鉴定技术建立了人泪液蛋白质组数据库

泪液是一种很重要的体液, 含有多种蛋白/肽段, 脂质, 小分子代谢物以及电解质等。已有研究表明, 基于人类泪液蛋白质组研究发现相关疾病的biomarkers。在本篇报道以健康人(3名女性和1名男性, 平均年龄36 ± 14岁)泪液为研究对象, 利用TripleTOF 5600质谱系统分析了人类泪液蛋白质组。共鉴定出1543个蛋白, 假阳性率(FDR) < 1%, 是迄今为止最大的人类泪液蛋白质组数据。之后, 又进行了GO分析, 并分别和人类血浆蛋白质组数据、NEIBank中泪腺和角膜基因组数据进行比较。本研究所得到的蛋白质组可以用于研究眼睛疾病biomarkers或者建立MRM分析的基础。

Table 1 -Summary of three LC-MS/MS runs. For each run, data from six SCX fractions were combined and processed using ProteinPilot software.

# Analysis	# Spectra used	# Unique peptides	# Protein IDs (<1% FDR)	# Protein IDs (<1% FDR, with ≥2 peptides)	# Protein IDs (<1% FDR, with 1 peptide)
1	39,433	9127	1262	922	340
2	39,781	8977	1273	927	346
3	41,887	8982	1233	913	320

图1. 蛋白鉴定统计数据

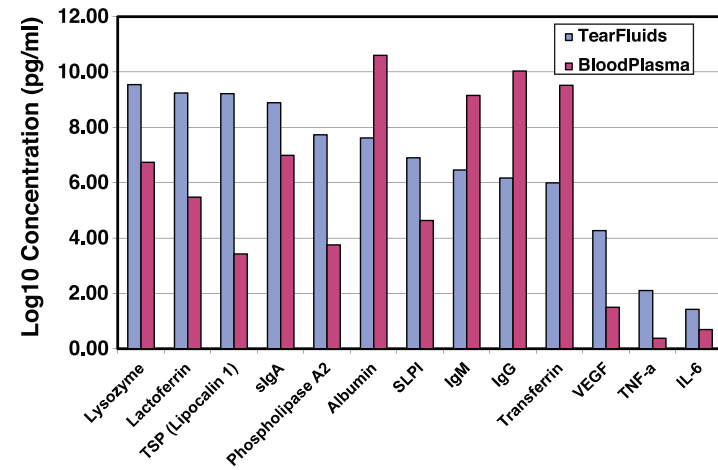


图2. 根据文献报道, 比较人类泪液和血浆蛋白质组浓度

参考文献:

Lei Zhou, Shao Zhen Zhao, Siew Kwan Koh, et al. In-depth analysis of the human tear proteome. J Proteomics. 2012.

### ● 案例2 利用蛋白鉴定技术发现更多的睾丸组织蛋白质种类 (华大发表)

泪液是一种很重要的体液, 含有多种蛋白/肽段, 脂质, 小分子代谢物以及电解质等。已有研究表明, 基于人类泪液蛋白质组研究发现相关疾病的biomarkers。在本篇报道以健康人(3名女性和1名男性, 平均年龄 $36 \pm 14$ 岁)泪液为研究对象, 利用TripleTOF 5600质谱系统分析了人类泪液蛋白质组。共鉴定出1543个蛋白, 假阳性率(FDR) < 1%, 是迄今为止最大的人类泪液蛋白质组数据。之后, 又进行了GO分析, 并分别和人类血浆蛋白质组数据、NEIBank中泪腺和角膜基因组数据进行比较。本研究所得到的蛋白质组可以用于研究眼睛疾病biomarkers或者建立MRM分析的基础。

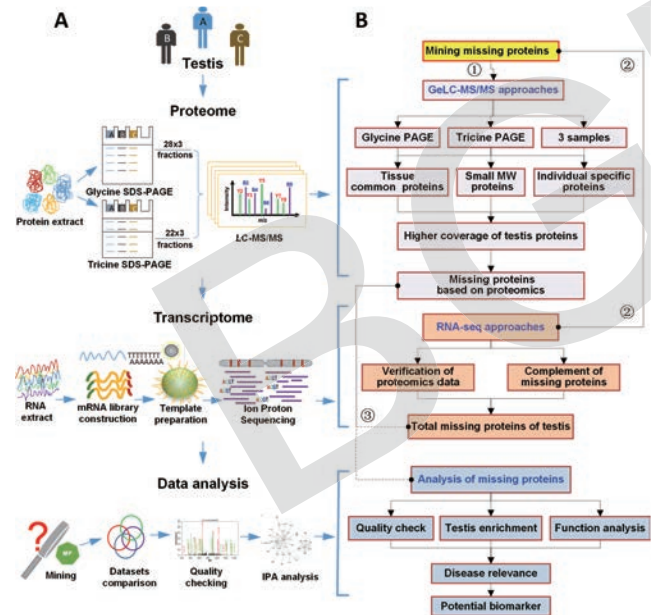


图1. 研究整体思路

参考文献:

Yao Zhang, Qidan Li, Feilin Wu, et al. Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. J Proteome Res. 2015.

### ◆ 常见问题

Q: 蛋白全谱分析能鉴定到多少蛋白?

A: 蛋白全谱分析鉴定到的蛋白数, 与物种、样本中蛋白质含量及复杂程度、蛋白组分分离程度、数据库的完整度都有很大的关系。如植物样本中蛋白质种类通常比动物样本多, 动物组织样本中蛋白质种类通常比其血液样本多。一般情况下, 一次全谱分析可以鉴定到3000-5000种蛋白。

Q: 目前可进行蛋白全谱分析的物种有哪些?

A: 只要数据库中该物种或近缘物种的蛋白质组数据趋于完整, 基因组已知或转录组数据足够丰富, 均可进行蛋白全谱分析。对于以上条件均不满足的物种, 可在进行蛋白质组全谱分析的同时, 做一个基因组或转录组测序, 以便为蛋白全谱分析提供支持。因此, 理论上所有的物种均可进行蛋白全谱分析。

Q: 双向凝胶电泳与质谱联用鉴定蛋白质的过程是什么?

A: 通过对蛋白质双向电泳的图谱扫描, 进行图谱差异分析, 找到差异点切出, 进行脱色、酶解、质谱检测, 得到质谱原始数据文件, 通过数据库搜索可以得到差异点蛋白质的详细信息。

### ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
人体睾丸组织	2015.09	Journal of Proteome Research	4.245	Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins
牡蛎	2010.10	Nature	41.456	The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation

# 蛋白定量分析

## 蛋白组学

### ◆ 产品概述

蛋白定量分析 (Quantitative Proteomics)，即利用质谱的方法对细胞、组织、体液等样本类型中的全部或特定的蛋白质进行表达量检测和种类鉴定，包括全部蛋白定量分析和目标蛋白定量分析两种类型。在研究的发现阶段，研究者获得全面的差异表达蛋白，需要对不同样本中尽可能多的蛋白质表达量进行比较，通常采用全部蛋白定量分析 (iTRAQ技术)；在研究的验证阶段，为了达到结果验证、假阳性排除、以及寻找潜在生物标志物的目的，还需要进行更为精确的目标蛋白定量分析 (多反应监测MRM技术)。蛋白定量分析可用于筛选和寻找任何因素引起的样本之间的蛋白差异表达，对某些关键蛋白进行定性定量分析，展现外界刺激或生长发育过程中生物体内的蛋白质组的变化，解释生理病理机制。

### ◆ 产品优势

**高通量：**iTRAQ技术可同时对8种不同样本进行定量分析；MRM技术可同时对一个样本中数十种蛋白进行精确定量

**周期短：**采用iTRAQ发现目标蛋白，再通过MRM技术进一步验证，避免制备抗体的过程，缩短科研项目周期

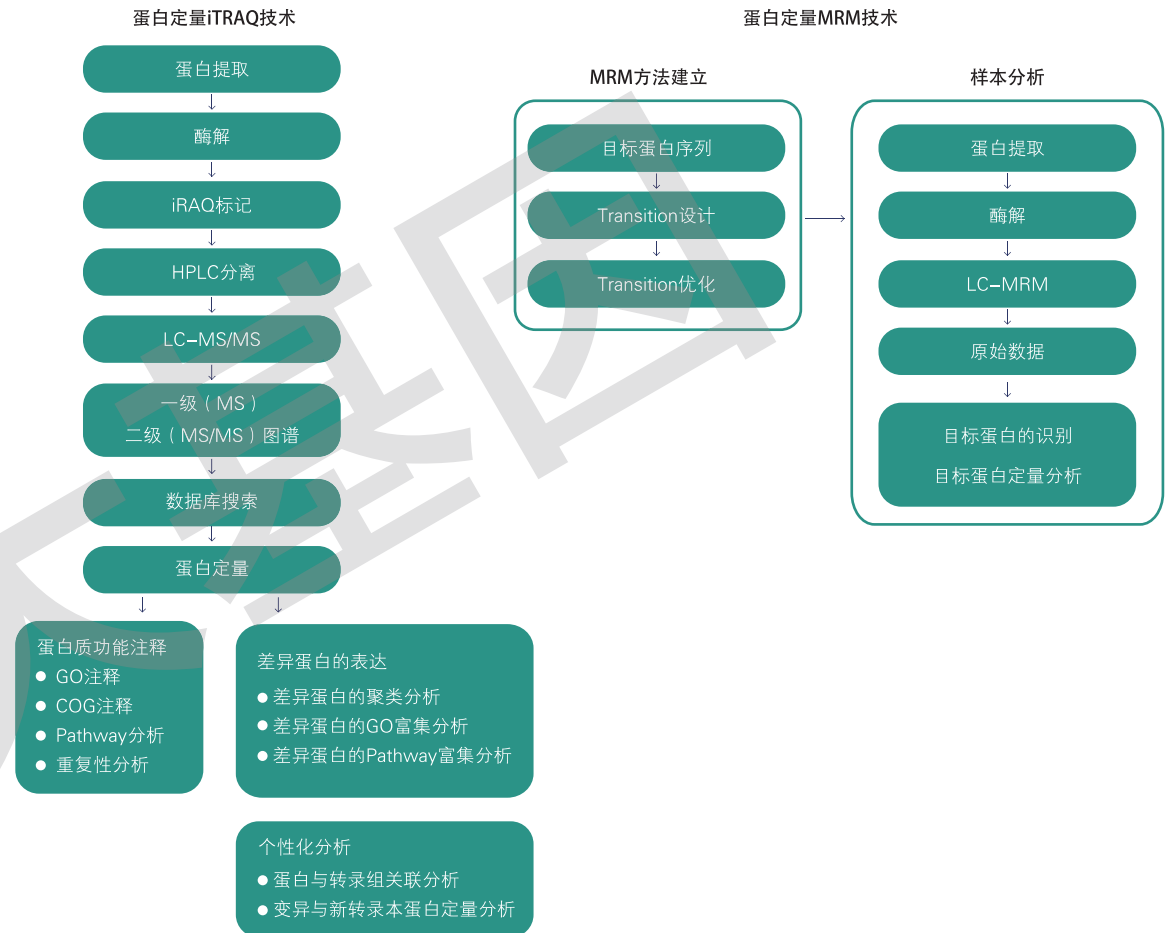
**高重复性：**iTRAQ技术，采用多个标记样本同时上机的方法，一次性比较多个样本，避免样本分别上机检测和传统制胶实验重复性差的问题

**实验高效：**MRM技术，同样采用液质联用技术，高效验证iTRAQ技术的结果

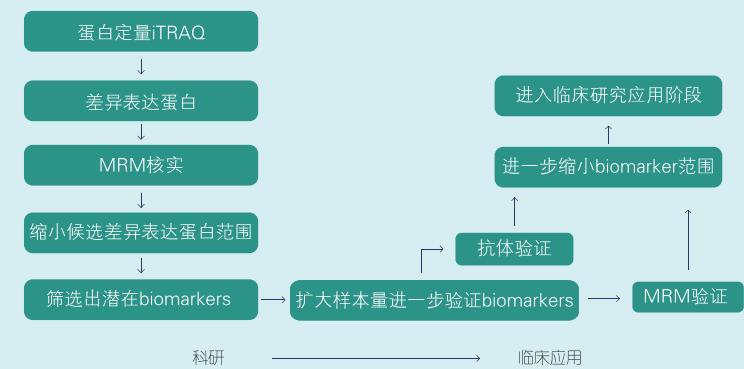
### ◆ 产品应用



### ◆ 技术流程



### iTRAQ联合MRM技术一站式解决方案技术路线



## ◆ 信息分析内容

### iTRAQ标准信息分析

- 1、数据产出统计
- 2、蛋白质定性，定量结果
- 3、蛋白质GO、COG分析
- 4、蛋白质Pathway代谢通路分析
- 5、差异蛋白的GO、Pathway富集分析
- 6、重复性分析（仅针对设计有重复的实验）
- 7、多样品间表达模式聚类（三个或三个以上样品可提供）

### iTRAQ定制化信息分析

- 1、蛋白质组与转录组/RNA-Seq（二选一）表达量关联分析

#### 表达量关联分析：

- ① 从鉴定、定量、差异三个层面进行关联统计
- ② 对定量层面关联结果细分为五类，并分别对其做GO、COG、Pathway注释分析
- ③ 蛋白组与转录聚类分析

#### 功能富集关联分析：

- ① GO、Pathway富集关联分析
- ② 转录组和蛋白的Pathway整合图

- 2、变异与新转录本检测（需要转录组resequencing/外显子/重测序数据（三选一））

- ① 在蛋白层面对SNP/SNV、InDel进行鉴定与定量
- ② 在蛋白层面对可变剪切、新转录本进行鉴定与定量（此项仅适用于转录本resequencing）

### MRM标准信息分析

- 1、数据产出统计
- 2、Transition list选择
- 3、目标蛋白质定量分析

## ◆ 技术参数

### 样品要求：

服务类型	蛋白量	其他类型样本
蛋白定量iTRAQ 蛋白定量MRM	≥500μg，浓度≥1μg/μL	组织200mg；细胞 $3-5 \times 10^6$ （沉淀体积约为30-50μL左右）；血液（限血清、血浆）500μL；其他体液根据实际情况评估，一般要求大于5mL

服务类型	研究对象	实验策略	质谱仪	项目周期
蛋白定量 iTRAQ	组织、细胞、体液样本中的全部蛋白种类和表达量	将整个组织、细胞、体液样本中的蛋白质组提取、酶解、标记并混合，利用液相色谱分离技术将蛋白质组分分成多个组分后，再进行质谱（LC-MS/MS）鉴定和定量	Q-Exactive/ 5600	40
蛋白定量 MRM	组织、细胞、体液样本中指定的几种蛋白质类型	MRM方法建立，选择最优的transition筛选合适的目标蛋白，再对样本分别进行处理并上质谱仪检测	Q-Exactive/ 5600/5500	40

## ◆ 案例解析

### ● 案例1 通过蛋白定量iTRAQ技术发现脓毒症预后不良的分子标志物

脓毒症是导致危重病人死亡的主要原因。本研究目的是寻找脓毒症预后的潜在分子标志物。在发现阶段，选取了30个脓毒症患者的尿液用iTRAQ技术进行定量分析，总共鉴定了232个蛋白质，对差异蛋白质做更深入的生理学和生物学分析，找到7个与预后相关蛋白，其中5个上调，2个下调。在验证阶段，选取另外54个脓毒症患者（其中27个预后正常，27个预后不良）采用Western Blot(WB)对差异蛋白进行验证，证明LAMP-1可作为早期预后评估的biomarker。

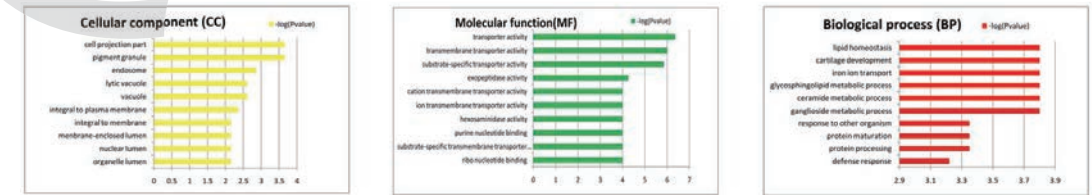


图1. GO注释图，CC是细胞位置分类、MF是分子功能、BP是生物过程

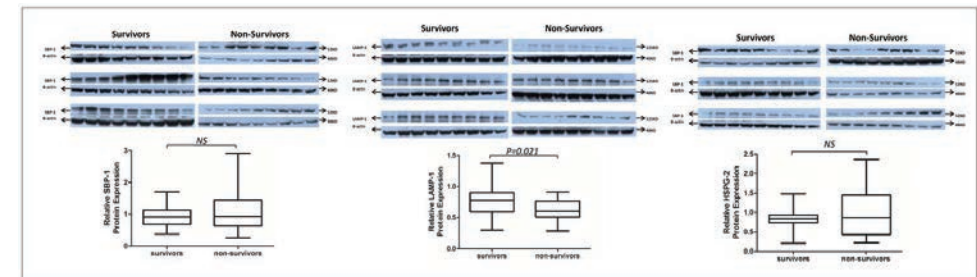


图2. WB验证图，第一个为SBP-1，第二个为LAMP-1，第三个为HSPG-2

### 参考文献：

Longxiang Su, Lichao Cao, Ruo Zhou, et al. Identification of Novel Biomarkers for Sepsis Prognosis via Urinary Proteomic Analysis Using iTRAQ Labeling and 2D-LC-MS/MS. PLOS ONE. 2013.

## ● 案例2 通过iTRAQ和MRM联合技术寻找并验证疾病的分子标志物

研究选用33份结肠癌组织，分别提取息肉和癌症组织（转移和非转移）中的膜蛋白，发现阶段，通过蛋白定量iTRAQ技术共鉴定到5566个蛋白，找到三种不同组织间的差异表达蛋白。验证阶段，通过SRM/MRM技术，105个候选的分子标志物中的69个得到了验证，另外选取三种组织各10个样本进行确认，69个分子标志物中的47个得到了进一步的确认。最终通过western blotting (Western Blot)，免疫组化和组织芯片证实C8orf55可能是诊断结肠癌、胃癌和乳腺癌潜在的分子标志物。

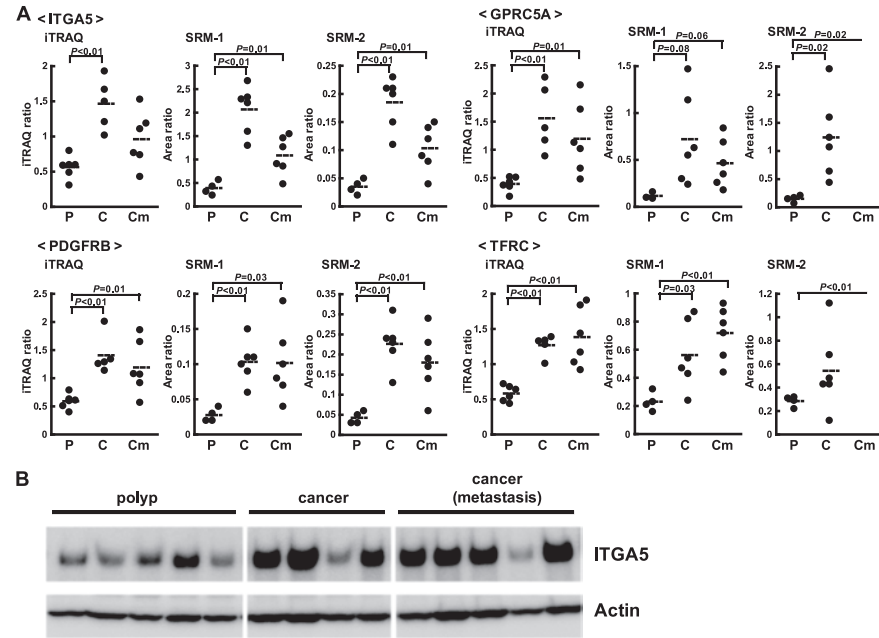


图1. 发现阶段iTRAQ技术及验证阶段MRM技术的典型结果展示

参考文献：

Hideaki Kume, Satoshi Muraoka, Takahisa Kuga, et al. Discovery of Colorectal Cancer Biomarker Candidates by Membrane Proteomic Analysis and Subsequent Verification using Selected Reaction Monitoring (SRM) and Tissue Microarray (TMA) Analysis. Mol Cell Proteomics. 2014.

## ◆ 常见问题

Q: iTRAQ怎样设置重复？

A: 推荐至少进行2次及以上生物学重复，否则iTRAQ实验结果一般不能被杂志接收。

Q: 不同试验的样本做iTRAQ定量可否放在一组上机？

A: 不能。因为不同试验的研究对象一般有差异，即便是同样的物种，处理方式也会有不同，这种情况下，不同来源的样本混在一起上机，会大大提高样本中蛋白种类的复杂程度，影响质谱的分辨能力，降低蛋白鉴定数目和定量效果。所以不同来源的样本，蛋白表达差异过大者，必须分别上机进行检测。

Q: 为什么有些差异表达基因在蛋白层面没有相应的差异蛋白？

A: 从生物学角度看，由于RNA调控、蛋白降解、蛋白分泌、转录和翻译效率不一致等原因，导致蛋白水平和转录水平未必呈现一样的趋势，这是正常现象。

Q: 只通过蛋白定量iTRAQ技术找到的差异蛋白，可否直接认定为biomarkers？

A: 从生物实验的角度来讲，通过一种实验得到的结果是需要再进行验证的。在发现阶段，用蛋白定量iTRAQ技术找到的差异蛋白，一般还需要通过其他技术进行验证后，再得出最终的结论，推荐采用目标蛋白定量MRM技术，或者传统的western blotting (Western Blot) 技术进行验证。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
结肠癌	2016.01	Journal of Proteomics	3.89	discovery of potential colorectal cancer serum biomarkers through quantitative proteomics on the colonic tissue interstitial fluids from the AOM-DSS mouse model

# 蛋白修饰分析

## 蛋白组学

### ◆ 产品概述

蛋白质磷酸化是最常见、最重要的一种蛋白翻译后修饰方式。磷酸化蛋白分析 (Phosphoproteome Analysis)，采用相对定量方法 (如iTRAQ, Label-free)，可以对发生磷酸化修饰的蛋白进行鉴定、磷酸化位点识别以及表达差异分析，并提供功能分析结果。这种研究方法可以用于高效筛选关键的磷酸化调控蛋白因子，揭示机体在特定生理病理条件下信号通路的开关机制。

### ◆ 产品优势

**高分辨率：**精确鉴定蛋白磷酸化位点，可定位到单个氨基酸位点的磷酸化修饰

**高鉴定率：**可提供CID、ETD、HCD等多种碎裂模式，提高磷酸化蛋白鉴定率

**高覆盖度：**一次实验可鉴定上千个磷酸化位点，快速构建蛋白磷酸化修饰谱

**实验高效：**利用iTRAQ标记的定量方法可同时分析8个样品的差异磷酸化蛋白

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

#### 标准信息分析

- 1、数据产出统计
  - 2、磷酸化位点鉴定、磷酸化肽段和磷酸化蛋白质鉴定
  - 3、磷酸化蛋白质GO、COG注释
  - 4、磷酸化蛋白质Pathway代谢通路注释
  - 5、差异磷酸化蛋白GO、Pathway富集分析
  - 6、差异磷酸化蛋白功能聚类分析
- (注：单个蛋白磷酸化鉴定分析仅限前两条信息分析内容)

#### 定制化信息分析

- 1、磷酸化蛋白功能分析-激酶与底物分析
- 2、磷酸化蛋白motif位点注释及分析
- 3、磷酸化蛋白底物与激酶互作关系

### ◆ 技术参数

#### 样品要求：

服务类型	样本类型	蛋白量	其他类型样本
磷酸化蛋白分析		≥10mg	细胞 $3-5 \times 10^7$ (沉淀体积约为300-500μL左右)；其他类型样本需进行评估后确定
单个蛋白磷酸化鉴定		>30μg, 目标蛋白纯度 >80%	-

服务类型	研究对象	实验策略	质谱仪	项目周期
磷酸化蛋白分析	组织、细胞、体液样本中的全部磷酸化蛋白	将整个组织、细胞中的蛋白质组提取、酶解、标记、TiO <sub>2</sub> 富集并混合，利用液相色谱分离技术将蛋白质组分分成多个组分后，再进行质谱 (LC-MS/MS) 鉴定和定量	Q-Exactive	60
单个蛋白磷酸化鉴定	分离纯化得到的单一磷酸化蛋白	对单一蛋白进行TiO <sub>2</sub> 富集，再进行质谱 (LC-MS/MS) 鉴定	Q-Exactive	25

## ◆ 案例解析

### ● 案例1 通过定量磷酸化组学分析揭示人细胞骨架重组调控机制

蛋白磷酸化和去磷酸化过程在细胞功能的众多方面起到调控作用。本研究将anti-CD3处理的人类T细胞与对照组T细胞进行磷酸化蛋白定量分析，鉴定差异表达的磷酸化蛋白。一共对334个T细胞受体应答蛋白进行分子功能分类和细胞定位，并构建T细胞受体应答蛋白互作网络。

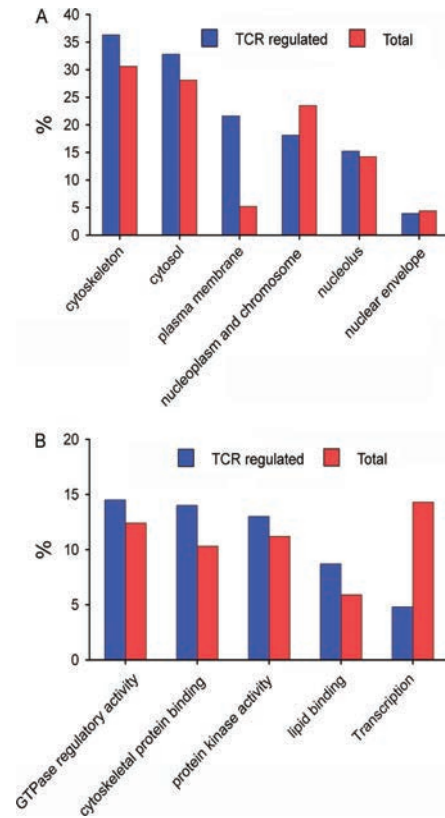


图1. 功能分类与细胞定位

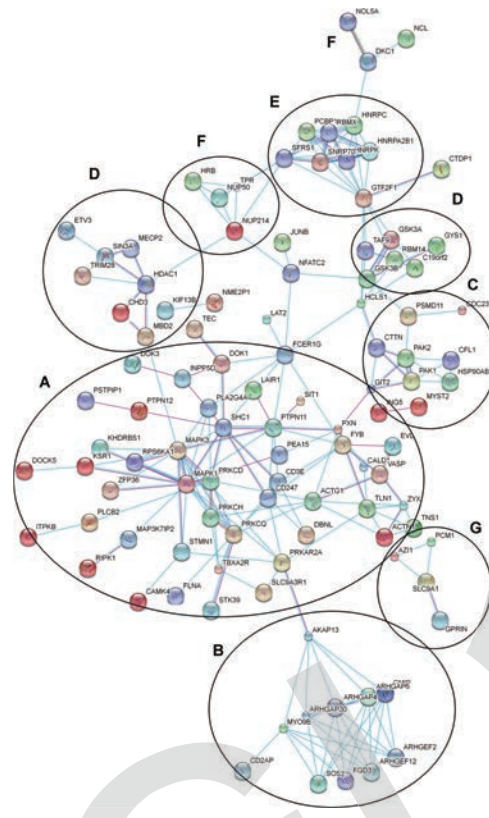


图2. T细胞受体应答蛋白的互作网络

参考文献:

Ruperez P, Gago-Martinez A, Burlingame AL, et al. Quantitative phosphoproteomic analysis reveals a role for serine and threonine kinases in the cytoskeletal reorganization in early T cell receptor activation in human primary T cells. *Mol Cell Proteomics*. 2012.

## ◆ 常见问题

Q: 如何确定磷酸化位点?

A: 一种特定的翻译后修饰通常会作用于一定的氨基酸，经修饰后，这一类氨基酸会增加相同的分子量，如磷酸化肽段因加入磷酸基团而产生+80的质量偏移。基于质量偏移的理论，利用生物信息学方法可以分析质谱数据鉴定磷酸化翻译后修饰。

Q: 目前用于磷酸化富集的方法是什么?

A: 我们选用的是TiO<sub>2</sub>富集法，它是目前最为成熟的金属氧化物磷酸化肽富集法 (Plant Sci. 2015 Dec;241:138–50.)。TiO<sub>2</sub>在不同的pH值下，可以表现为路易斯酸或路易斯碱。在酸性条件下，钛原子带正电表现为路易斯酸，可以与阴离子结合；在碱性条件下，则表现为路易斯碱，可以与阳离子结合。因此，磷酸化肽段的磷酸基在酸性条件下与之结合，碱性条件下洗脱，达到富集磷酸化肽段的目的。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
HeLa细胞系	2015.12	MCP	6.56	Lowering endogenous cathepsin D abundance results in ROS accumulation and cell senescence



# 5 代谢组学

- ◆ 非靶向代谢组学
- ◆ 靶向代谢组学

## 非靶向代谢组

### 代谢组学

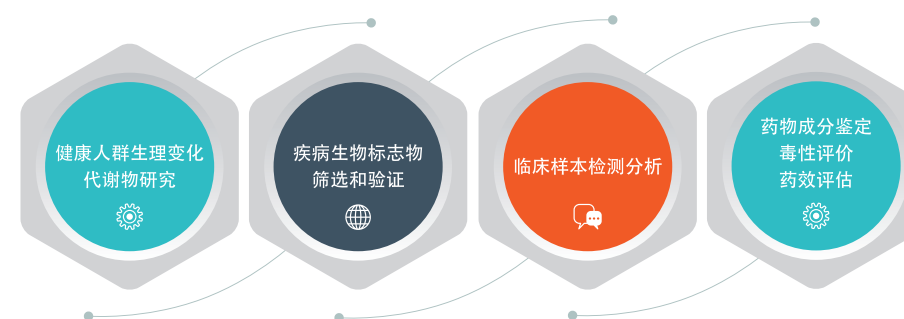
#### ◆ 产品概述

代谢组学 (metabonomics/metabolomics) 是继基因组学和蛋白质组学之后新兴发展起来的学科, 对生物体内所有代谢物进行定量分析, 并寻找代谢物与生理病理变化的相对关系的研究方式, 是系统生物学的组成部分。其研究对象是相对分子质量在1000Da以内的小分子物质, 如脂类、酮类、有机酸等。代谢组学根据其研究目的又可以分为靶向代谢组学和非靶向代谢组学, 其中非靶向代谢组学又可根据研究的物质分为常规代谢组学和脂质组学。非靶向代谢组学属于研究的发现阶段其目的是尽可能获得样本中全面的代谢谱图, 比对不同处理或不同生理状态下代谢谱图的差异, 测定其生物扰动诱导的生物指纹, 为后续验证阶段提供靶标。

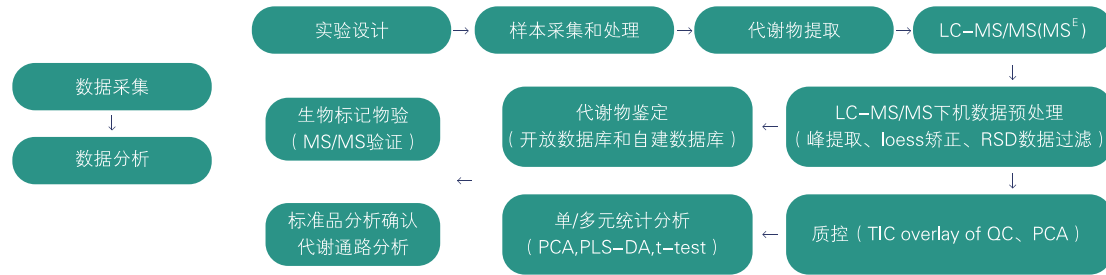
#### ◆ 产品优势

- 经验丰富: 丰富的代谢物提取经验
- 平台先进: 采用沃特世公司Xevo G2-XS Qtof质谱仪的MS<sup>E</sup>技术
- 专业分析: 采用Progenesis QI商业软件进行峰提取和鉴定
- 质量超群: 完善严格的统计分析和质控

#### ◆ 产品应用



## ◆ 技术流程



## ◆ 信息分析内容

- 1、数据预处理及归一化
- 2、数据质控：QC样本TIC重叠图；QC样本PCA分析
- 3、统计分析：单变量分析（差异倍数及T检验）；多变量分析（PCA, PLS-DA）
- 4、差异离子筛选及鉴定（注意此处鉴定为质谱数据与开源数据库匹配得到候选代谢物，最终确认代谢物还需要与标准品进行匹配）
- 5、聚类分析
- 6、候选差异代谢物Pathway注释

## ◆ 技术参数

样品要求：

服务类型	样本类型	蛋白量	组织	细胞
常规代谢组学/脂质组学		血浆、血清≥500μL；尿液≥2mL	≥200mg	个数≥10 <sup>7</sup>

（注：样本采集后应立即置于-80℃冻存，干冰寄送）

服务类型	研究对象	实验策略	质谱仪	项目周期
常规代谢组学/脂质组学	组织、细胞、体液样本中的全部代谢物或者脂质部分	提取整个体液、组织、细胞中的代谢组或脂质组后，进行质谱（LC-MS/MS）鉴定和定量	Xevo G2-XS	60

## ◆ 案例解析

### ● 案例1 利用代谢组学研究手段寻找帕金森症生物标志物（华大发表）

研究表明，帕金森综合症与体液代谢物变化密切相关。本研究拟找到帕金森症与尿液代谢组变化的关系，从而做到疾病早诊。实验通过采集104例健康人和106例帕金森病人（发病后0周、16周、32周）的尿液代谢组数据，找到健康组和疾病组间45个差异代谢物，这些差异代谢物集中于类固醇、脂肪酸β氧化、组氨酸代谢、苯丙氨酸代谢、色氨酸代谢、核酸代谢等代谢途径。帕金森果蝇模型研究中发现，过表达α-突触核蛋白会导致犬尿酸通路中色氨酸代谢变化，并用MRM方法验证成功。

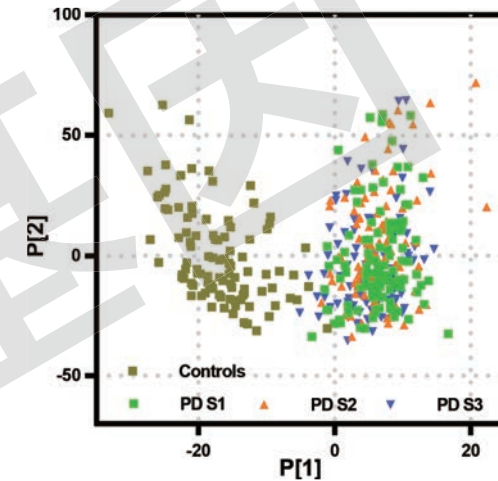


图1. 健康组与帕金森疾病组的OPLS-DA模型

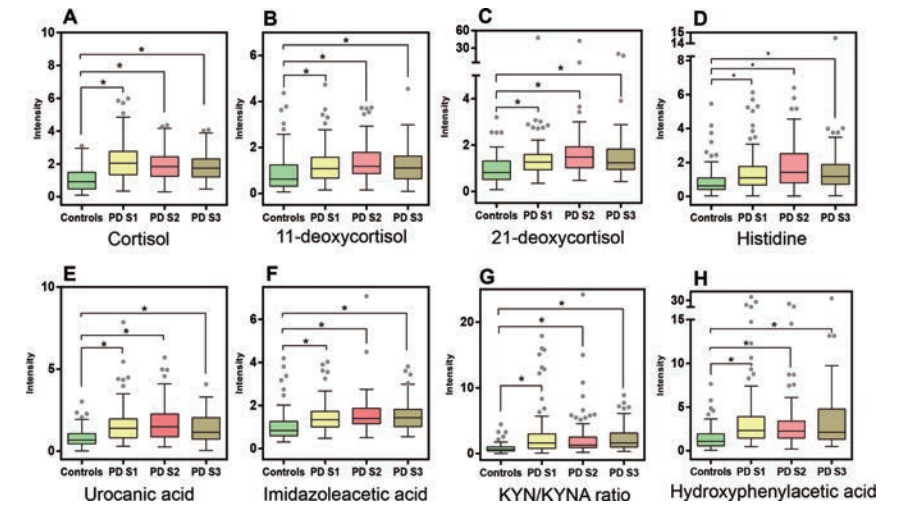


图2. 关键代谢物在健康组与帕金森疾病组的表达水平

参考文献：

Hemi Luan, Liang-Feng Liu, Nan Meng, et al. LC-MS-Based Urinary Metabolite Signatures in Idiopathic Parkinson's disease. J Proteome Res. 2015.

● 案例2 利用代谢组学方法研究孕妇周期性代谢物变化（华大发表）

妊娠期间，孕妇的代谢产物会发生非常大的变化，而这些变化可能与孕妇及胎儿的健康有密切关系。此前已有一些专家开始关注孕期疾病如妊娠糖尿病、先兆子痫等，而代谢物的变化也被证明和这些疾病有一定相关性。本实验召集180名（30\*6）健康孕妇作为志愿者，按照孕期分为6个组，分别提取血浆中的代谢物，通过非靶向代谢组学研究，发现不同孕期变化较大的代谢物有6个，涉及生物蝶呤代谢、磷脂代谢、氨基酸代谢和脂肪酸氧化等。其中色氨酸代谢途径发生显著变化，而且多种代谢物都和孕妇心理反应相关。

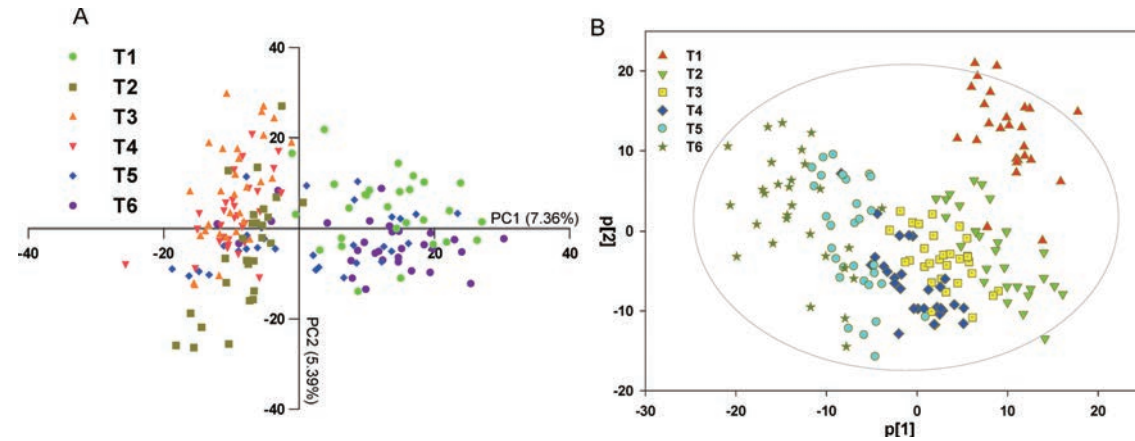


图3. 六个不同时期孕妇血浆代谢组学数据的PCA模型及PLS-DA模型

参考文献:

Hemi Luan, Nan Meng, Ping Liu, et al. Pregnancy-Induced Metabolic Phenotype Variations in Maternal Plasma. J Proteome Res. 2014.

◆ 常见问题

Q: 研究代谢组的三种平台GC-MS, LC-MS, NMR有什么区别呢?

A: LC-MS的覆盖范围很广泛，可用于绝大部分化合物的检测分析，样品前处理步骤简单。GC-MS主要用于易挥发、热稳定、中低极性物质的检测，比如含羟基、羧基、氨基和亚氨基等基团的极性强的物质，样品前处理相对繁琐，但谱图数据较为健全；NMR重复性很好，对样本制备要求少，快速，没有偏好性。但是灵敏度不及质谱技术。

Q: 代谢物鉴定搜索的数据库有哪些?

A: 人和哺乳动物样本会选择HMDB。

Q: 为什么代谢物鉴定与数据库匹配后还需要与标准品确认?

A: 由于化合物在不同仪器或不同实验条件下得到的谱图差异很大，因此通过数据库匹配的结果只供参考作用。而根据2007年MSI (Metabolomics Standards Initiative代谢组标准计划，由代谢组学协会提出)所提出的化合物鉴定原则，要真正确认化合物还是需要比对样品与标准品在相同实验条件相同仪器下得到谱图和出峰时间。

Q: 代谢物确认需要客户提供什么?

A: 只需要提供代谢物的标准品即可。标准品纯度>95%，可购于Sigma。

Q: 代谢组对于样本数量有要求吗?

A: 代谢组研究的目的在于寻找组间差异代谢物，因为代谢物的个体差异非常大，样本量少，找到的差异很有可能是个体差异，而不是群体差异，不具有统计学意义。因此临床样本数量每组建议不低于30个。

◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
冠心病	2016.03	Scientific Reports	5.58	Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease
小鼠弓形虫感染	2016.01	Scientific Reports	5.58	Metabolomic Profiling of Mice Serum during Toxoplasmosis Progression Using Liquid Chromatography-Mass Spectrometry
帕金森综合症	2014.08	Journal of Proteome Research	4.25	LC-MS-Based Urinary Metabolite Signatures in Idiopathic Parkinson's Disease
辣椒	2013.12	PNAS	9.42	Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization
孕妇血浆	2013.01	Journal of Proteome Research	4.25	Pregnancy-induced metabolic phenotype variations in maternal plasma

# 靶向代谢组

## 代谢组学

### ◆ 产品概述

靶向代谢组学是基于一定的研究基础，选取大量样本，对每个样本中特定的几种代谢物含量进行测定，从而进一步验证关键代谢物在生物体内的表达变化情况与表型的关系。多反应监测（multiple reaction monitoring, MRM）技术作为一种质谱检测的分析方法，具有特异性强、灵敏度高、准确度高、重现性好、线性动态范围宽、自动化高通量的突出优点，这些特质能够满足今天很多研究领域的迫切需要；传统方法多采用酶联免疫诊断技术、化学发光标记免疫分析或紫外分光光度法等技术，通量低，特异性不高，相比于传统方法的低通量、难以区分同分异构体等问题，MRM可作为靶向代谢组学研究的一种新的解决方法。

### ◆ 产品优势

**高准确性：**采用MRM技术，通过两级离子选择，排除干扰离子，降低化学背景，提高灵敏度

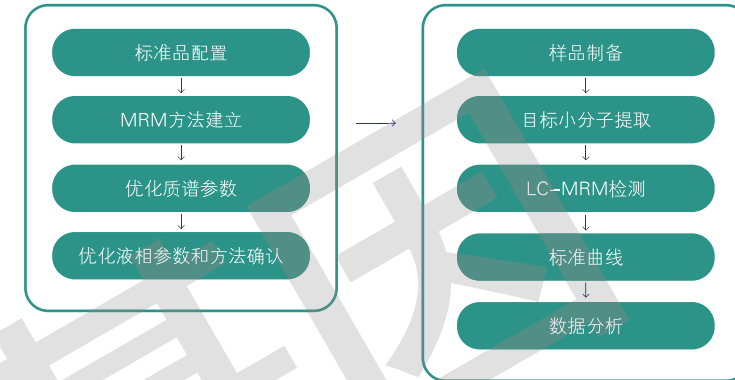
**特异性强：**通过质谱三重四级杆的筛选，特异性的选择目标代谢物，可避免传统实验方法特异性差的劣势

**高验证率：**前期通过液质联用的非靶向代谢组学方法获得的差异代谢物，再通过靶向代谢组MRM方法进行验证，一致性高

### ◆ 产品应用



### ◆ 技术流程



### ◆ 信息分析内容

- 1、母子离子对（Transition）选择
- 2、标准曲线绘制（绝对定量）
- 3、目标小分子定量分析

### ◆ 技术参数

样品要求：

服务类型	样本类型	蛋白量	组织	细胞
靶向代谢组学		≥1mL	≥1g	个数≥10 <sup>8</sup>
维生素		血清≥300μL	-	-
氨基酸		血浆≥100μL	-	-
激素		血清≥550μL	-	-

（注：样本采集后应立即置于-80℃冻存，干冰寄送；维生素、氨基酸、激素检测的样本类型必须是上表中标注的类型，其他样本类型需要评估。）

服务类型	研究对象	实验策略	质谱仪	项目周期
靶向代谢组学	组织、细胞、体液样本中的特定几种代谢物（客户提供检测物质列表）	质谱方法建立成功后，质谱上机检测样本中特定代谢物含量	5500/4500/TQS/TQD	60
维生素、氨基酸、激素检测	人体血液中特定的维生素、氨基酸、激素类物质	标准曲线制备，样本前处理后直接质谱上机检测	5500/4500/TQS/TQD	10

## ◆ 案例解析

### ● 案例1 靶向代谢组学研究II型糖尿病分子机制

通过多组学研究方法（RNA芯片、蛋白组SILAC定量、靶向代谢组MRM），研究者获得由高脂诱导的胰岛素抗性小鼠模型的生理调控整体情况，通过比较模型鼠两种组织（白色脂肪组织和肝脏组织）间的数据，并分析糖尿病药物罗格列酮（RGZ）的效用，寻找新的组织特异性药物靶点。靶向代谢组MRM技术在处理组和对照组中，分别定量检测153个代谢物和142个代谢物，整合转录组、蛋白组和代谢组的数据，得到了II型糖尿病分子机制的整体情况。通过网络和通路分析，研究证实了白色脂肪组织中的关键蛋白和相关代谢物紊乱，RGZ药物主要作用的分子机制，并发现肝脏特异性的药物靶点Sph和SP1。

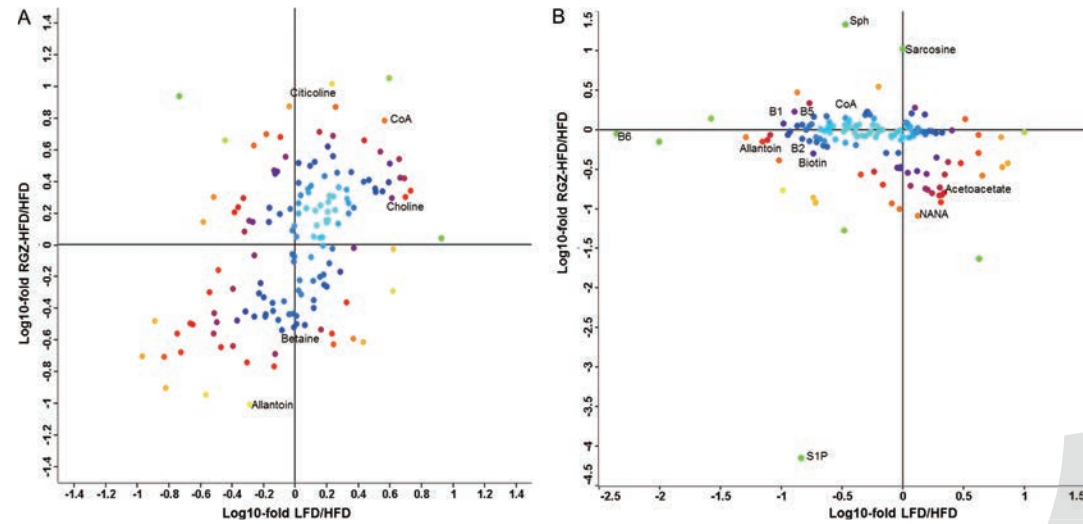


图1. 靶向代谢组学技术检测到两种组织中的组间差异代谢物的表达比值，白色脂肪组织（左）和肝脏组织（右）

参考文献:

David Meierhofer, Christopher Weidner, and Sascha Sauer. Integrative analysis of transcriptomics, proteomics, and metabolomics data of white adipose and liver tissue of high-fat diet and rosiglitazone-treated insulin-resistant mice identified pathway alterations and molecular hubs. J Proteome Res. 2014.

### ● 案例2 利用代谢组学方法寻找血液中老年痴呆早期生物标志物

阿兹海默症是一种世界范围内的疾病，为了避免有创的早期检测，本研究从早期认知损伤的患者血液中寻找生物标志物，帮助临床医生进行疾病早诊。研究选取了525位志愿者进行为期5年的随访。实验发现阶段，研究者选取了53例健康人和53例帕金森病人（其中18例为3年随访期间由健康转为发病者）的血浆进行非靶向代谢组学研究；验证阶段，研究者选取了20例健康人和21例帕金森病人（其中18例为3年随访期间由健康转为发病者）的血浆，进行靶向代谢组学验证，最终发现并验证了外周血中10个脂质代谢物，可作为指示2-3年间发病的候选生物标志物，准确度达到90%。

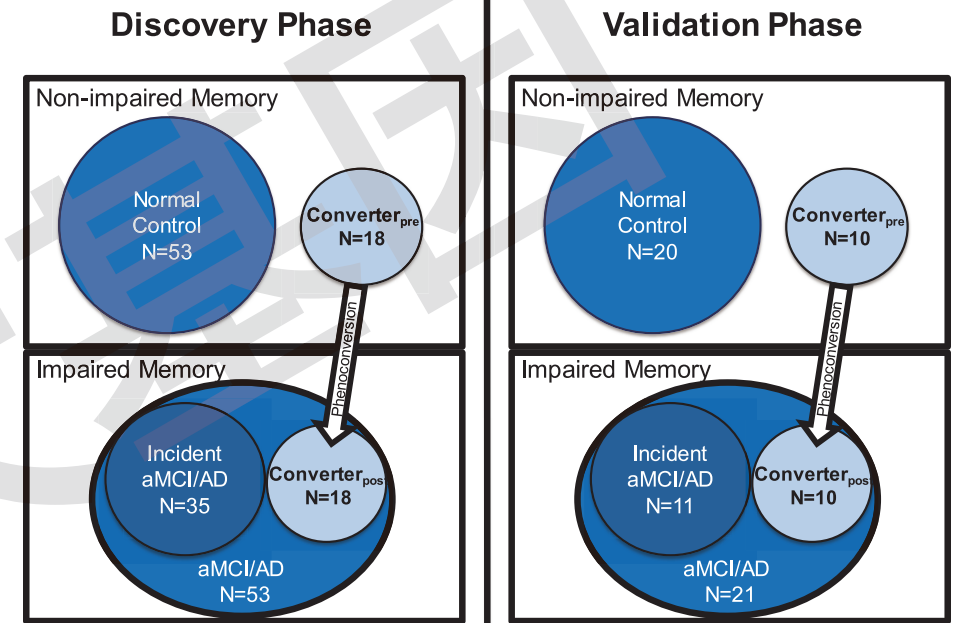


图2. 发现和验证阶段的取样

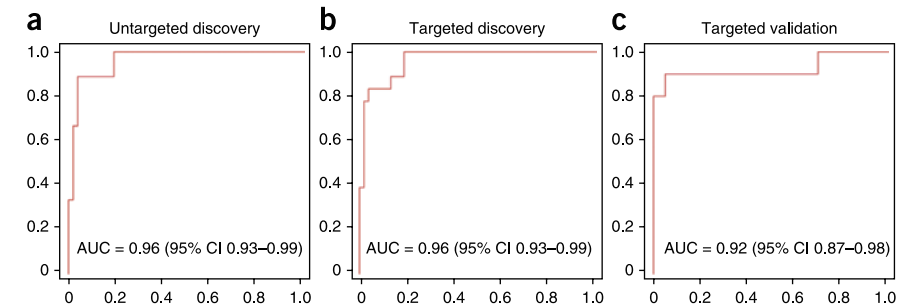


图3. 脂质代谢物的ROC分析

参考文献:

Mark Mapstone, Amrita K Cheema, Massimo S Fiandaca, et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med. 2014.

## ◆ 常见问题

Q: 如何选择母子离子对?

A: 一个母离子可能对应多个子离子, 在实验方法建立过程中, 每个目标化合物一般选择多个母子离子对 ( $\geq 2$ ), 最终会挑选两对最佳母子离子对, 其中一对用于定量, 另外一对用于辅助定性。

Q: 标准曲线如何制作?

A: 选取6个不同浓度的标准品溶液, 根据标准品浓度和标准品MRM信号响应峰面积拟合直线方程得到标准曲线。

Q: MRM技术进行质谱时, 可以一次监测多少个母子离子对? 可否进行多个特定分子同时监测?

A: MRM检测可实现多个目标分子的同时检测, 最大限度节省时间, 提高检测效率, 每个工作循环最多能处理300对母子离子对, 使用Schedule MRM可达700对母子离子对。

Q: 如何减少实验误差, 保证数据的质量?

A: 实验过程中会加入内标校正实验误差, 并为客户提供严格的QC质控确保数据的真实可靠。

Q: 是否需要提供同位素内部的标准品?

A: 同位素内标能校正实验误差, 提供准确的定量结果, 是目前绝对定量普遍采用的实验策略; 也可根据客户需求合理设置其它内标, 减少实验误差。

# 6 新技术

- ◆ 单细胞DNA测序
- ◆ 单细胞RNA测序
- ◆ 免疫组库测序

# 单细胞DNA测序

新技术

## ◆ 产品概述

单细胞DNA测序是对已分离好的单细胞进行高保真高覆盖度的全基因组扩增，然后对扩增产物进行全基因组、外显子组或目标区域测序，再对高通量测序产生的数据进行信息分析，从而快速全面的获取单细胞DNA水平上所有变异信息。单细胞测序技术的产生，及其在人类疾病健康研究的应用将更好地揭示疾病的发生、发展过程和规律,有助于人类对疾病的认识、预防、诊断和治疗。

## ◆ 产品优势

低样品量：单细胞级别的DNA起始量

精准度高：高质量的扩增和极低的扩增错误率（ $10^{-7}$ ， $10^{-6}$ ）

检测高效：平均可检测>95%的基因组位点，扩增产物长度达20kb，有利于更多类型的变异检测

技术稳定：技术稳定性较好、可重复性高

## ◆ 产品应用



## ◆ 技术流程



## ◆ 产品分类

单细胞全基因组重测序  
单细胞外显子组测序  
单细胞目标区域测序

## ◆ 信息分析内容

- 1、去除接头、污染序列及低质量数据
- 2、比对，产出数据统计
- 3、SNP、InDel检测、注释和统计
- 4、CNV、SV检测、注释和统计（仅全基因组测序）
- 5、结合客户的需求，协商确定定制化信息分析内容，如肿瘤的亚克隆演化，群体结构分析，主成分分析，系统发育树构建等（定制化分析内容）

## ◆ 技术参数

样品要求：

样本类型	WGS	WES/目标区域测序
单细胞	1-2个（4μL PBS悬液）	1-2个（4μL PBS悬液）
单细胞	1-2个（4μL PBS悬液）	1-2个（4μL PBS悬液）

- 测序策略：PE101或PE151
- 推荐数据量：大批量细胞测序，可采用低深度浅覆盖测序；有较大亲缘关系的细胞，可适量增加测序深度；小规模细胞测序，研究功能性变化，可参考常规全基因组重测序/外显子组测序/目标区域测序的测序深度；个性化研究，不同研究目的可适度调整测序深度。
- 项目周期：低于100个样品时（30X），标准流程的项目周期为50个工作日

## ◆ 案例解析

### ● 案例1 癌症单细胞测序——多重置换扩增用于原发性血小板增多症治疗方法研究

对取自一例典型的JAK2阴性ET（原发性血小板增多症）病人的90个单细胞进行了全外显子测序，分析发现该ET病人为单克隆进化，进一步分析发现了几个关键的ET候选基因，如SESN2和NTRK1，可能参与ET的发展。

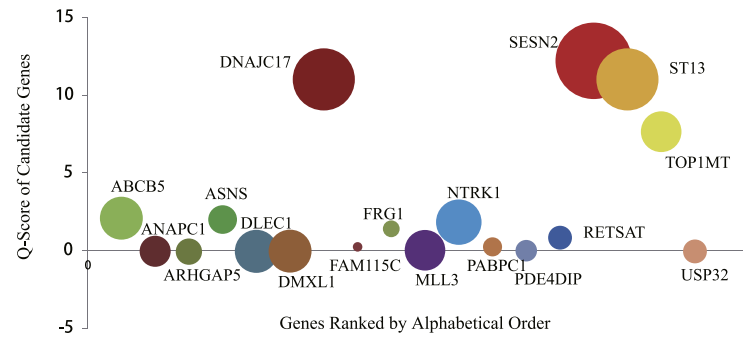


图1. ET病人的关键基因鉴定

通过Q score值挑选出18个ET病人的候选基因的dirver基因预测分析。

纵坐标是Q score值，圆圈的大小（直径）表示细胞突变的频率。

参考文献：

Hou, Y., Song, L., et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell. 2012.

### ● 案例2 单细胞外显子测序研究肾透明细胞癌的SNP特征

从一例肾癌样本中随机选取20个癌细胞和5个癌旁细胞，进行单细胞测序。聚类分析和进化分析均发现该例样本为单克隆，他们发现此例肾癌并非由常见的两个突变基因VHL和PBRM1导致，这说明在病人群体中所鉴定的频发突变（recurrent mutation）可能与肿瘤个体无关，同时也强调了在癌症分析和诊断过程中进行个性化治疗的重要性。

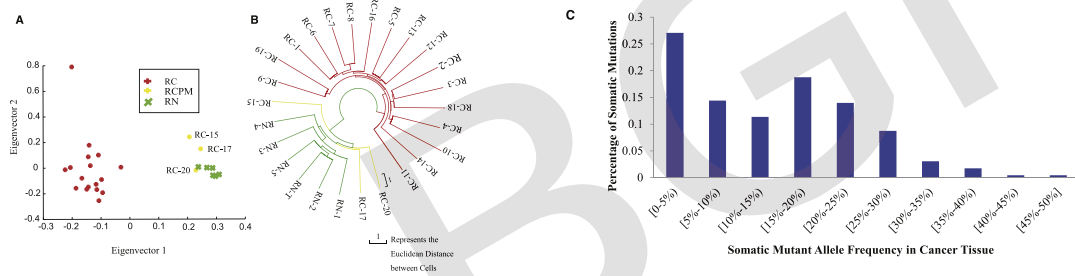


图2. PCA分析 (A) 和进化树分析

(B) 17个癌细胞中的突变频率分布

(C)。运用PCA分析，将癌细胞 (RC) 和正常细胞 (RN) 进行区分。

RCPM在细胞分离时被当作癌细胞，后期通过PCA分析发现实际为正常细胞。

参考文献：

Xu, X., Hou, Y., et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell. 2012.

## ◆ 常见问题

Q: 单细胞肿瘤研究如何避免癌细胞取样的假阳性?

A: 未经染色的单细胞，在显微镜下无法识别细胞是肿瘤或正常细胞，建议客户取样前通过病理学方法估计肿瘤组织的纯度，一般要求纯度高于80%。虽然后期信息分析能排除肿瘤组织中正常细胞的干扰，但是肿瘤细胞含量较低需要选取的细胞总量将会增多，增加项目投入。

Q: 经过全基因组扩增之后可以得到多少基因组 DNA ?

A: 全基因组扩增后可以得到3-5μg的DNA量，达到一般建库测序的DNA量要求，可用于各种常规的重测序。

Q: 扩增产物质控指标如何?

A: 全基因组扩增中扩增不均会导致测序覆盖度低，为确保数据利用率，全基因组扩增实验后须对扩增产物进行质控。我们通过PCR方法对管家基因进行检测。检测结果必须保证8个管家基因有6个以上阳性，反之则属不合格样品不能用于建库测序。现阶段管家基因质控的指标仅限于人，其他物种的还需要开发。此外我们还参考一般的DNA重测序样要求对扩增产物进行质控。

Q: MDA扩增的特点如何?

A: MDA具有扩增产物长、覆盖度高、扩增错误率低的优点，此外还存在两个主要的不足：扩增偏向性和等位基因丢失。基因组上GC含量越高扩增效率越高，导致高GC含量的区域测序深度远高于其他地方，呈现出明显的扩增偏向性。除此之外，在扩增过程中还存在等位基因丢失的问题（Allele dropout (ADO)），单细胞基因组上杂合的等位基因在扩增过程中会只扩增出其中一个等位基因，导致另外一个等位基因丢失。针对这些问题我们升级了扩增方法，使得基因组均一性比之前有了很大的提高；并且ADO也从之前Cell文章外显子数据的平均43%降低到5%以下（内部测试数据），大大降低了偏向性和ADO对下游信息分析的影响。

Q: 单细胞DNA测序可以做哪些变异检测?

A: 单细胞经过全基因组扩增之后，合格的扩增产物可以进行全基因组测序、外显子测序、目标区域捕获测序，可分别检测不同的变异类型。外显子测序和目标区域捕获测序只能进行SNP和InDel检测；全基因组测序则能进行所有变异检测，如SNP、InDel、CNV、SV，最全面地反馈单细胞水平上的所有变异类型。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
方法学	2015.08	GigaScience	7.46	Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing
结直肠癌	2014.06	Cell Research	14.81	Discovery of biclonal origin and a novel oncogene CLC12A5 in colon cancer by single-cell sequencing
膀胱癌	2012.08	GigaScience	7.46	Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer
骨髓增殖性肿瘤	2012.03	Cell	28.71	Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm
肾癌	2012.03	Cell	28.71	Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor



# 单细胞RNA测序

新技术

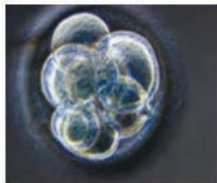
## ◆ 产品概述

单细胞RNA测序是通过对单个细胞中带PolyA尾的mRNA进行Smart-Seq2扩增，再结合高通量测序手段对其进行基因表达定量、功能富集、代谢通路、可变剪切等分析的一项新技术。Smart-Seq2扩增技术能改善转录后cDNA包含不完整的5'端的问题。通过单细胞RNA测序，可以解决传统RNA测序技术在早期胚胎发育、肝细胞、癌症、免疫等研究领域存在的样品量极低或细胞异质性问题，极大地拓展了RNA测序的应用范围。

## ◆ 产品优势

极低样品起始量，10pg  
 探测到90%以上的基因表达  
 正负链相关性大于98%  
 技术重复性高

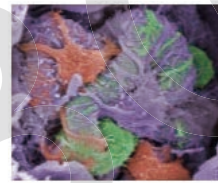
## ◆ 产品应用



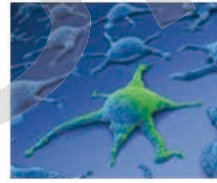
胚胎早期发育  
(时序表达)



iPSC  
(重编程细胞异质性)



免疫细胞群  
(组织异质性)



肿瘤干细胞  
(分化异质性)

## ◆ 技术流程



## ◆ 产品分类

单细胞转录组  
单细胞RNA-Seq定量

## ◆ 信息分析内容

单细胞转录组	单细胞RNA-Seq定量
测序数据过滤	对原始数据进行去除接头、污染序列及低质量reads的处理
参考基因组比对	测序评估(包括:数据比对统计、测序质量评估、测序饱和度和分析、测序随机性分析、Reads在参考基因/基因组上的分布)
新转录本预测	基因表达注释(包括:基因覆盖度统计、结果文件列表、结果文件示例)
SNP和InDel检测	差异表达基因分析
RNA编辑检测(基于第4条,且必须提供样品DNA数据检测出来的SNP结果)	组间差异表达基因分析(第4条和第5条只能二选一来做)
差异剪接基因检测	差异基因表达模式聚类分析
融合基因检测(仅适用于人的癌症融合基因研究,每个样本8G以上数据量)	GO功能显著性富集分析
基因表达量计算	GO功能分类分析(WEGO分析)
差异表达基因检测	Pathway显著性富集分析
差异表达基因层次聚类分析	蛋白互作网络分析(需要蛋白-蛋白相互作用数据库中有该物种的注释信息)
差异表达基因GO功能分析	转录因子分析(仅适用于植物)
差异表达基因Pathway功能分析	主成分(PCA)分析(五个或五个以上样品)
	条件特异表达分析(五个或五个以上样品)

## ◆ 技术参数

### 样品类型:

暂不提供单细胞分离服务，客户需自行分离得到单个细胞，置于我们提供的裂解液中送样。

目前可接收哺乳动物的生殖细胞、早期胚胎细胞、体细胞等。

### 注意事项:

- 1、分离单细胞时尽量减少细胞损伤，保证细胞活性。推荐口吸管分离。
- 2、确保操作环境无外源污染。设置阴性对照。
- 3、避免反复冻融，置入细胞后的裂解液干冰运送，运输时间不要超过72小时，确保样品送达时有足量干冰留存。
- 4、请客户自觉遵守道德伦理相关法律规定。

### 测序策略:

50SE (RNA-Seq) 或101PE (转录组)

### 推荐数据量:

50SE建议30M以上clean reads

101PE建议5-10G clean data

## ◆ 案例解析

### ● 案例1 单细胞RNA测序揭示人和小鼠早期胚胎发育进程

利用单细胞RNA测序技术，研究了人鼠的33个早期胚胎细胞，全面分析了人和小鼠从卵细胞到桑椹胚发育期，基因表达模式的动态变化。

### 结果展示:

1. 对人胚胎植入前7个阶段的单个分裂球的表达模式进行主成分分析，每个阶段的样品数目及主成分模式。

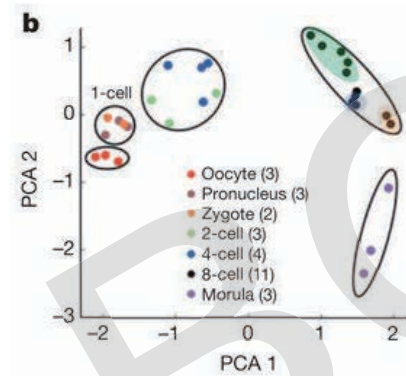


图1. 胚胎植入前7个阶段的单个分裂球表达图谱的主成分分析

2. 热图显示，7313个共表达基因在人早期胚胎发育的7个阶段，有阶段特有的表达模式，以及最相关的前3个GO条目。

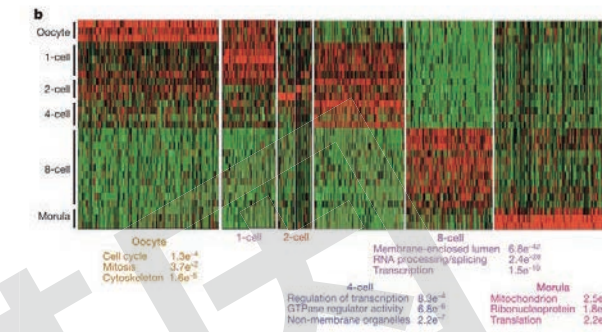


图2. 胚胎发育的7个阶段共表达基因热图

### 参考文献:

Xue Z, Huang K, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature. 2013.

## ◆ 常见问题

Q: 单细胞RNA-Seq可以用于哪些类型的细胞?

A: 单细胞RNA-Seq目前可以用于哺乳动物的生殖细胞、早期胚胎细胞、部分体细胞、部分细胞系。不同类型的细胞，成功率和风险也有所不同。总体来说，生殖细胞和早期胚胎细胞的成功率比较高，可达到70~90%。其次是细胞直径大于10 $\mu$ m的游离单细胞（如CTC）、细胞系等。有特殊结构、较小的细胞，风险比较大，不太适合目前的单细胞RNA测序技术。

Q: 单细胞RNA-Seq对低丰度转录本检测效率如何?

A: 单个细胞中的mRNA含量极少，只有10pg左右，在扩增过程中会产生一定偏向性，无法避免的随机性丢失一部分低丰度转录本的信息。根据经验，30M SE50 reads能够检测到90%以上表达的基因。目前的单细胞RNA-Seq可以检测到大多数的低丰度表达基因和绝大多数的高丰度表达基因。

Q: 客户是否需要设立阳性对照和阴性对照?

A: 客户在分离单细胞时，需要设置阴性对照，来检验操作过程和环境是否有污染。我们在扩增时会设立空白对照，用常规组织或者一般标准品作为阳性对照，与客户送过来的细胞平行进行实验操作。

## ◆ 发表文章

研究内容	发表时间	发表期刊	影响因子	文献标题
癌症	2015.11	GigaScience	7.46	Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells

# 免疫组库测序

## 新技术

### ◆ 产品概述

免疫组库测序 (Immune Repertoire sequencing, IR-Seq) 以 T/B 淋巴细胞为研究目标, 用多重PCR技术扩增决定B细胞受体 (BCR) 或T细胞受体 (TCR) 多样性的互补决定区, 再结合高通量测序技术, 全面评估免疫系统的多样性, 深入挖掘免疫组库与疾病的关系。

### ◆ 产品优势

扩增偏好性低: 采用两步法建库, 并优化引物配比, 降低扩增偏好性

重复性高: 同一样本建库两次克隆一致性高

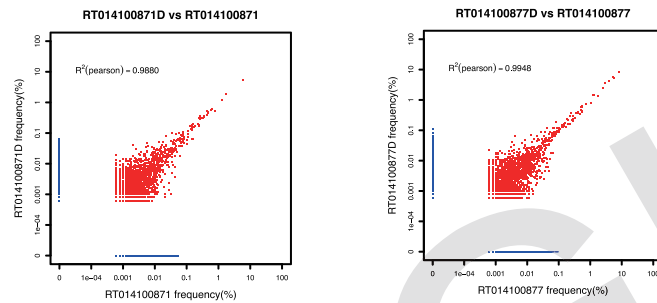


图1. 同一样本实验重复性评估 (HiSeq)

产品多元化: 可提供人T/B细胞不同链CDR3的检测服务

研究内容	目标区域	样品类型	建库方法	测序策略	推荐数据量
TCR $\alpha$ , TCR $\beta$	CDR3	DNA、RNA	M-PCR	PE101	2-3M raw reads
BCRH, BCRL	CDR3	DNA、RNA	M-PCR	PE151	2-3M raw reads
BCRH (鉴定抗体亚型)	CDR3	RNA	M-PCR (VC)	PE151	2-3M raw reads

精确的信息分析: 自主开发软件IMonitor可校正扩增及测序错误, 提供高级和个性化分析服务

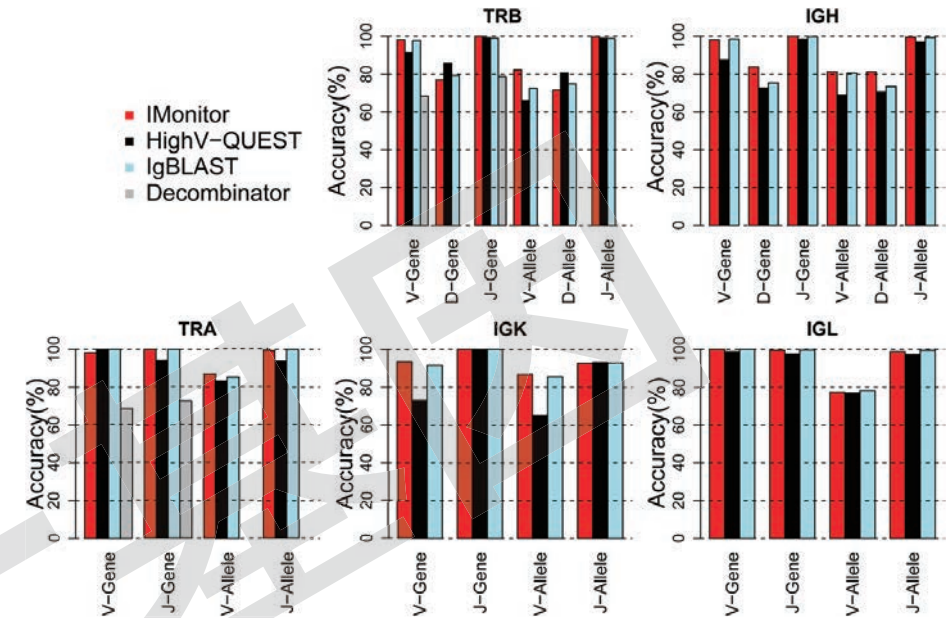


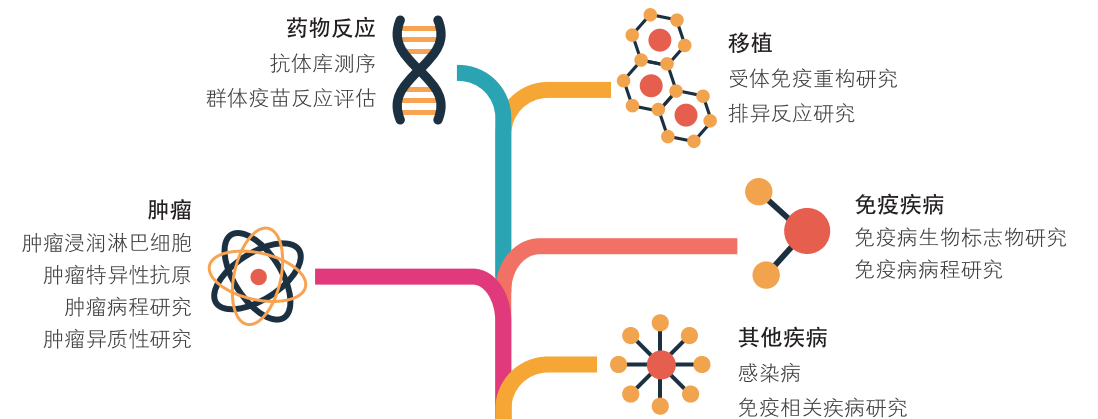
图2. 0.5%测序错误率下不同软件检测V/D/J gene & allele准确性评估

采用105个序列进行模拟, 评估四种软件检测V/D/J基因的准确性。

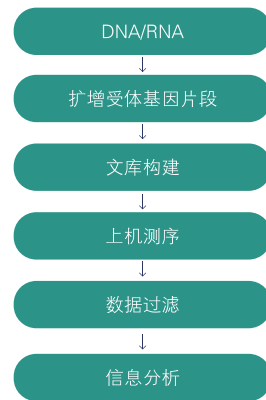
IMonitor和IgBLAST明显优于HighV-QUEST和Decombinator两款软件, 而IMonitor在检测IGH的D基因时, 表现优于IgBLAST。TRA/TRB代表TCR  $\alpha/\beta$  链, IGH/IGK/IGL代表BCR H/L链。

(Genetics. 2015 Oct;201(2):459-72.)

### ◆ 产品应用



## ◆ 技术流程



## ◆ 信息分析内容

### 数据统计

- 1、去除接头污染序列及低质量reads
- 2、数据拼接，消除测序背景及有效数据构建
- 3、数据产出统计及测序数据成分和质量评估

### 比对分析

- 1、与数据库 (IMGT) V/D/J 基因片段比对
- 2、寻找最佳 V/D/J 比对结果

### 序列结构分析

- 1、分析CDR序列组成及序列碱基成分
- 2、分析CDR序列的碱基插入和缺失
- 3、编码CDR序列翻译成氨基酸和肽链

### 免疫组库构建

- 1、构建免疫组库表达谱：统计多样性抗体库克隆表达情况
- 2、免疫组库多样性呈现：绘制 V/J 基因表达的 2D、3D图

### 免疫组库差异分析

- 1、样品间多样性差异分析 (辛普森系数、香农威纳系数)
- 2、样品间克隆表达差异分析 (CDR3, V-D-J)
- 3、分组样品间的克隆表达差异分析 (CDR3, V-D-J)

## ◆ 技术参数

样品要求:

样品类型	最低送样量	建议送样量
PBMC或T/B cells DNA/RNA	500ng	1μg
gDNA	1.5μg	3μg
RNA	1.5μg	3μg
全血	2mL	5mL
分选细胞		1 × 10 <sup>6</sup> 个

测序策略: 推荐 TCR 101PE/BCR 151PE

推荐数据量: 至少2-3M raw reads

项目周期: 标准流程的项目周期为40个工作日

## ◆ 案例解析

### ● 案例1 TCRβ CDR3表达特征可作为区分乙肝、肝癌亚型的生物标志物

研究发现, T细胞与肿瘤的预后和发展有密切的关系。文章选取160个RNA样品 (分别取自健康人、乙肝病人、三种亚型肝癌病人、结肠癌病人), 多重PCR扩增TCR β CDR3区域, 分析TCR β 在不同群体中的表达特征。研究发现, 癌症病人的TCR β 很相似, 与健康人完全不同。而HCC、ICC、MHC病人的TCR β 有差异, 可用于区分健康人、肝癌、乙肝病人。

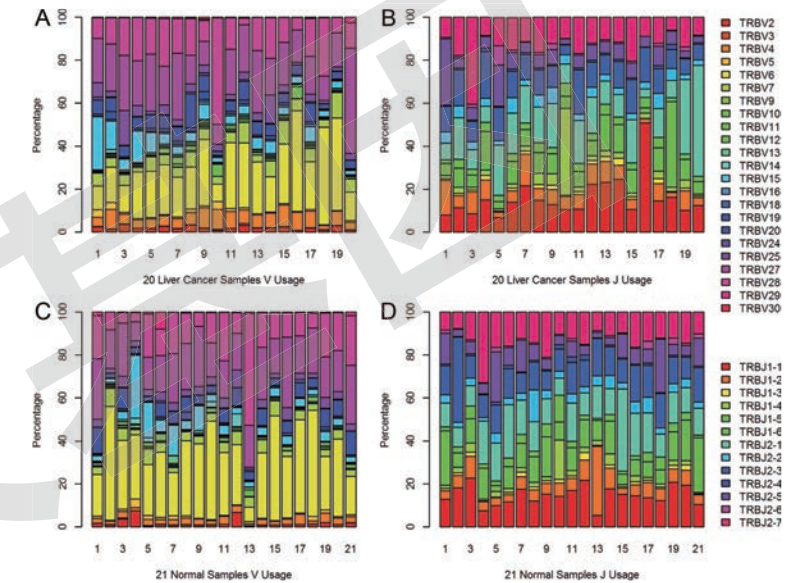


图3. 肝癌病人和健康人血液样品的V-J基因 (V-J usage) 半定量结果  
(A)20个肝癌病人血液样品中V基因用量  
(B)20个肝癌病人血液样品中J基因用量  
(C)21个健康人血液样品中V基因用量  
(D)21个健康人血液样品中J基因用量

参考文献:

Yingxin Han, Xing Liu, et al. Identification of characteristic TRB V usage in HBV-associated HCC by using differential expression profiling analysis. Oncoimmunology. 2015.

## ◆ 常见问题

Q: 华大可以提供什么物种的免疫组库测序?

A: 目前仅提供人TCR/BCR CDR3区域测序。

Q: 分选细胞需要提供多少细胞量?

A: 至少需要1 × 10<sup>6</sup>个细胞, 细胞量越高, 免疫细胞的多样性越高。

Q: 免疫组库同一个TCR或BCR的两条链要分开收费吗?

A: 是的。TCR的α或β, BCR的H或L链需要分开建库、测序、分析并收取费用, 因为每条链都要分别作扩增、建库和测序。

## ◆ 发表文章

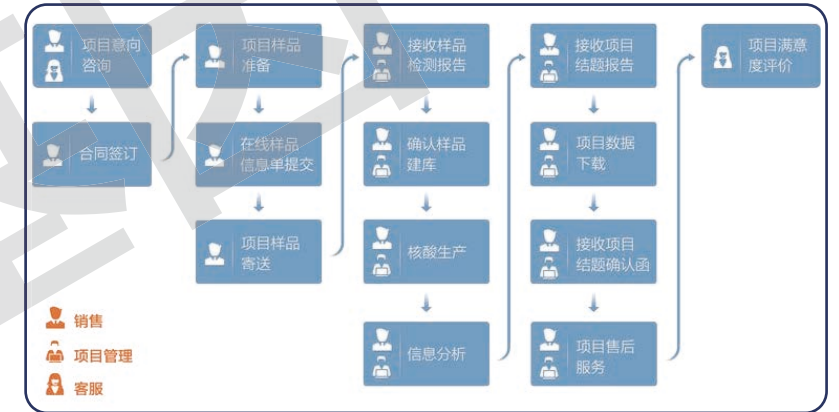
研究内容	发表时间	发表期刊	影响因子	文献标题
方法学	2016.03	PLOS ONE	3.06	Systematic Comparative Evaluation of Methods for Investigating the TCR $\beta$ Repertoire
方法学	2015.10	Genetics	4.39	IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis
综述	2015.07	Cancer	5.62	Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma
肝癌、乙肝	2015.04	Oncolmunology	6.27	Identification of characteristic TRB V usage in HBV-associated HCC by using differential expression profiling analysis

# 华大基因客户服务指南

您好! 非常感谢您选择华大基因作为您的合作伙伴!

为了让您全面了解华大基因的项目执行流程, 我们特拟定《华大基因客户服务指南》对项目流程及其各阶段进行详细说明。如果以下内容与您所选择的产品信息有任何不一致, 请以合同中双方约定的内容为准。

在项目执行过程中, 如果您有任何疑问或者问题, 请您及时联系我们。



华大基因项目执行流程

## 项目各节点应尽事项

节点	客户须知事项	细则/常见问题
合同签订	在合同起草阶段, 请按照标准合同模板操作; 对于非标准合同, 华大基因负责人需对合同进行审核。	标准合同是指按照华大基因现有合同模板签订的合同, 除标准合同外, 均为非标准合同。
	合同中所涉及金额、技术条款、法务条款、付款方式等内容如有变更, 华大基因需对变更后的合同重新审核。	合同生效, 要求须同时具备合同各方公章与法定代表人签字。
	合同签字盖章后, 客户需将双方签字盖章的合同于5个工作日内邮寄至华大基因。	邮寄地址: 广东省深圳市盐田区洪安三街21号华大综合园三办1308办公室 联系电话: 0755-36307088, 邮寄信息请告知相关销售。
	请确认发票中的付款单位名称与签订合同的客户单位名称一致。	如需开具增值税专用发票, 需提供客户单位相关资料(营业执照、税务登记证)的电子扫描件(PDF或图片格式)或复印件一份。按照审计规定, 不同的项目不能申请开同一张发票。
	请确认发票内容与华大基因营业范围相符, 发票内容须与合同中的服务或产品相关。	例如: 生物技术研发与服务, 技术转让与技术咨询等。
发票开出情况、发票真伪查询, 可请登陆深圳国税局官方网站查询。	深圳国税局官网: <a href="http://www.szgs.gov.cn">http://www.szgs.gov.cn</a>	

节点	客户须知事项	细则/常见问题
合同签订	客户发票如需作废，并重新开具新的发票，需要作废的发票邮寄至华大基因，待收件人确认收件后，可联系销售重新申请新的发票。	邮寄地址：广东省深圳市盐田区洪安三街21号华大综合园三办1308办公室 联系电话：0755-36307088
付款	银行汇款	1、请仔细查看发票信息后，按照发票金额汇款至合同指定账户，并注明款项相关信息（如：发票号码、项目编号、服务类型）。
	客户单位汇款成功后，客户可在财务部门查询汇款信息，并请及时告知销售。	客户如需华大基因帮助核实汇款是否已到账，可以联系销售。收款回执是由银行提供，用以证明华大基因已收到项目款项。客户需在汇款备注中注明发票信息（发票号码、付款单位、付款时间等）。
样品准备	客户准备好要寄送的样品，检查样品容器上的标识是否清晰和牢固（易脱落/在低温（如干冰）下易脱落的，可用油性笔再在管盖上写上样品信息，信息需与管壁保持一致），样品状态是否正常，做好可能出现异常情况的预防措施。	常见问题： 1、样品容器上的样品名称无法识别。 2、样品容器漏液导致样品量减少或导致样品污染。
	请客户认真核实样品信息，对易混淆的样品分袋包装，并在密封袋上标识。	常见问题： 1、华大基因样品处理组收到样品后，无法根据信息单中信息进行清点和识别。 2、要求返样，但未明确说明返样相关信息。
在线样品信息单提交	请客户登陆华大基因样品信息系统填写在线样品信息单。	在线样品信息系统网址： <a href="http://www.bgisample.com">http://www.bgisample.com</a> 如填写过程中有任何疑问请联系销售（若为系统异常问题，可截图反馈）。
	对于须在特殊温度下保存的样品，请在信息单备注中说明保存温度。	常见问题： 某些须特殊温度保存的样品，如菌体样品、培养基样品，没有说明保存温度，导致样品无法使用。
	对于有特殊实验要求的样品，请在信息单备注中说明。	常见问题： 1、多个样品合并检测，实验前需先离心，但是客户没有在备注中说明。 2、组织样品提取，客户没有注明提取要求。 3、样品处理，客户没有说明样品提取时是否需要混合，混合方法，物种本身特殊性等。
在线样品信息单提交	请检查所填写的信息是否完整，核对实际寄送样品信息与样品信息系统内提交信息单填写的内容（包括样品名称和数量等）是否一致。	常见问题： 为确保信息无误，和信息单不一致的样品需要进行确认，可能导致实验延期。如果没有签订正式的合同，在送样时需要与销售做好沟通，让销售在信息单中补充项目管理同事的信息，方便及时跟进。
样品寄送	请检查所寄送的样品是否是对应项目的所需样品。	常见问题： 客户样品寄送错误，导致样品无法进入生产环节。
	样品寄送前，建议向物流公司了解寄送所需的时间，干冰消耗量大约5kg/天，根据样品在途时间，计算干冰需求量，放置足量干冰。	常见问题： 客户放置干冰较少，导致华大基因收到样品时干冰已耗尽，样品所处常温状态，严重影响RNA样品、组织样品的质量。
	封装样品时，请准备较结实的外包装，包装内放置防震泡沫或气泡袋。	常见问题： 1、华大基因收到样品时，样品外包装破损，导致样品污染或样品暴露。 2、样品在运输过程中与干冰碰撞导致密封袋破损，样品散落在干冰中，或出现样品容器破裂等情况。

节点	客户须知事项	细则/常见问题
样品寄送	请按照在线样品信息单所指示的邮寄地址寄送，也可向销售或项目管理确认项目样品的邮寄地址。	常见问题： 样品寄送至错误目的地，导致样品无法进入生产环节。
	请准确填写寄件人信息。	常见问题： 华大基因收到的样品快递单上没有寄件人信息、寄件人信息错误或字迹无法辨认，导致找不到样品相应的联系人确认样品信息。
	寄送样品时，需随包裹寄送样品信息单或样品信息单批次号（建议打印信息单和编号）。	常见问题： 1、收到样品时没有样品信息单或样品批次号，样品不能及时被检测。 2、手写信息单字迹不清晰，样品信息无法确认。
	样品寄出后，请及时通知销售或项目管理，并跟踪样品的寄送情况。	常见问题： 销售或项目管理不知道此次样品的寄送信息，不能及时确认样品信息。
样品接收及信息确认	请及时处理由样品处理组、销售、项目管理发出的样品邮件，根据邮件提示进行样品信息确认。	常见问题： 1、收到样品时没有样品信息单或样品批次号，样品不能及时被检测。 2、手写信息单字迹不清晰，样品信息无法确认。
样品提取及检测	请接收来自销售或项目管理的样品检测报告。	如对样品检测结果有任何疑问，请直接联系项目管理或销售。
样品保存	华大基因对所有客户原始样本提供有期限的免费保存管理，请注意及时告知销售相应样品处理方式。	1、样品到达1个月内，需下达提取任务单或做其他指示（一般为返样、销毁、续存等），若无1个月后将默认销毁处理。 2、检测报告发出3个月内，需启动建库或做其他指示（一般为返样、销毁、续存等），若无3个月后将默认销毁处理。 3、总项目结题6个月内，需做其他指示（一般为返样、销毁、续存等），若无6个月后将默认销毁处理。
核酸生产	请接收来自销售或项目管理的项目进度反馈。	如需要销毁样品或返还样品，请提前联系项目管理或销售明确返样信息。如对核酸生产过程或结果有任何疑问，请直接联系项目管理或销售。
样品保存	请与项目管理或销售沟通信息分析方案，对于分析方案中的参考序列、分析名称、比对和聚类方案的设计要符合的实验设计，获取下载子项目数据的下载方法。	常见问题： 1、项目信息分析工作的按期执行需要在数据下机前，客户与项目管理或销售就信息分析方案达成一致。 2、如果客户没有就分析使用的参考数据库版本、样品名称或比对方案，提前向销售或项目管理反馈，信息分析一般默认参数执行。
	项目个性化分析，需提供个性化具体要求及参考资料。	个性化分析的具体内容及费用结算，需联系销售。
项目结题	项目完成后，项目管理或销售通过邮件提供项目结题报告及数据下载信息（华大基因数据交付系统ftp账号信息）。客户登陆华大基因数据交付系统下载项目结题报告及数据。如果通过硬盘传输，项目结束时，客户需提供硬盘接收地址、接收人信息等。	数据交付系统网址： 1、深圳： <a href="http://cdts-sz.genomics.cn/">http://cdts-sz.genomics.cn/</a> 2、香港： <a href="http://cdts-hk.genomics.cn/">http://cdts-hk.genomics.cn/</a> 3、武汉： <a href="http://cdts-wh.genomics.cn/">http://cdts-wh.genomics.cn/</a> 4、天津： <a href="http://cdts-tj.genomics.cn/">http://cdts-tj.genomics.cn/</a> 项目结题后，项目数据免费保留3个月，请在约定时间内下载数据。

节点	客户须知事项	细则/常见问题
项目 结题	项目完成后，项目管理或销售通过邮件提供项目结题报告及数据下载信息（华大基因数据交付系统ftp账号信息）。客户登陆华大基因数据交付系统下载项目结题报告及数据。如果通过硬盘传输，项目结束时，客户需提供硬盘接收地址、接收人信息等。	系统无法登陆或打不开ftp，数据下载困难或数据下载中出现问题，请直接联系销售或项目管理。
		数据交付形式有网络交付和移动设备交付两种。移动设备（硬盘）交付需在合同签订前协商是否采用硬盘交付。
	成功下载项目数据和结题报告后，回复项目结题确认函。	数据缺失或格式不符合使用需求，请直接联系项目管理、销售或华大科技客服。
项目结 题后服 务	客户帮助文档及信息分析结果使用	忘记回复或逾期未回复项目结题确认函，默认客户同意项目结题。
	结题报告解读及使用	使用数据有困难、对信息分析结果有疑问或对项目有新的想法和建议请直接联系销售、项目负责人或客服人员。
	项目满意度评价	对结题报告有疑问，请直接联系项目负责人、区域销售或客服人员。
	投诉及建议	项目结题后，会收到项目满意度调查问卷或满意度回访电话，请您对我们的服务及产品做出评价。 电话：400-706-6615 邮箱：info@bgi.com

注：以上信息是对项目执行流程及各节点提供说明，仅供参考。如以上内容与适用于您产品的信息有任何不一致，请以适用于您产品的合同内容为准。

联系我们：

电话：400-706-6615（请根据语音提示选择）

邮箱：info@bgi.com

官方网站：www.bgi.com

样品信息系统：www.bgisample.com