



网址:www.bgitechslutions.com  
邮箱:info@genomics.cn

华大基因  
BGI

华大基因  
BGI

华大基因·总部(深圳)  
地址:深圳市盐田区北山工业区综合楼(518083)  
电话:400-706-6615

华大基因·美洲(波士顿)  
地址:One Broadway, 14th Floor, Cambridge, MA02142, USA  
电话:+1-617-500-2741

华大基因·欧洲(哥本哈根)  
地址:Ole Maaloes Vej 3, DK-2200 Copenhagen N, Denmark  
电话:+45-7026-0806

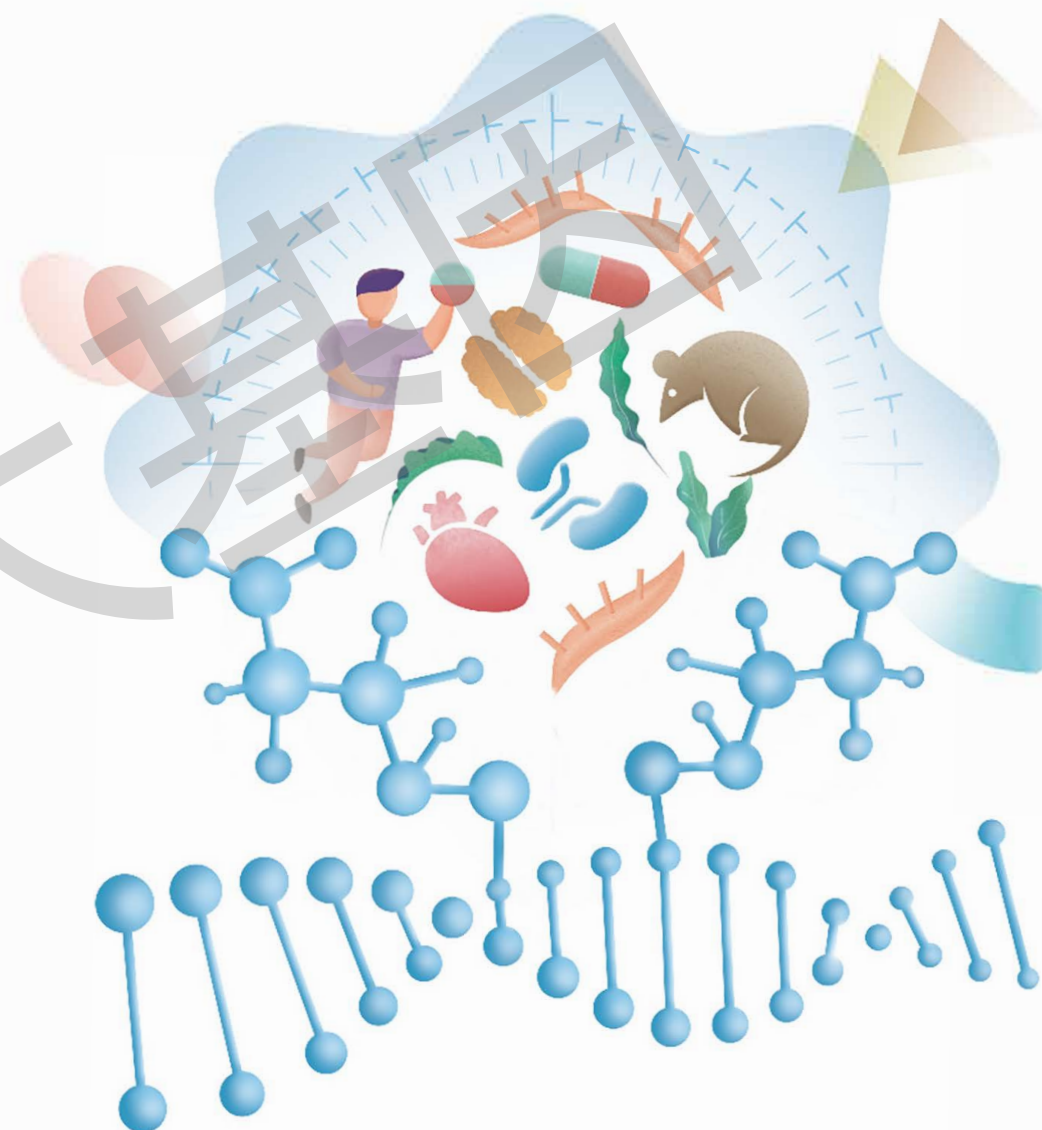
华大基因·日本(神户)  
地址:Kobe KIMEC Center BLDG.8F, 1-5-2 Minatojima-min-amimachi, Chuo-ku,  
Kobe City, Hyogo-pref.650-0047, JAPAN  
电话:+81-785-996-108

华大基因·亚太(中国香港)  
地址:香港新界大埔工业村大富街16号  
电话:+852-3610-3510

本手册仅供客户学习、交流和研究使用,请勿用于商业用途,违者必究。

版权声明:本手册版权属于深圳华大基因股份有限公司所有,未经本公司书面许可,任何其他个人或组织均不得以任何形式将本手册中的各项内容进行复制、拷贝、编辑或翻译为其他语言。本手册中所有商标或标志均属于深圳华大基因股份有限公司及其提供者所有。

版次:2019年9月版



产品手册 医学研究

# 目录

## 02

### 华大简介

公司简介	03
平台介绍	04
质量管理	06

## 07

### 基因组学

人全基因组测序	09
外显子测序	17
目标区域测序	25

## 29

### 转录组学

转录组测序	31
全长转录组测序	35
RNA-Seq	39
Small RNA测序	43
长链非编码RNA测序	47

## 53

### 表观组学

全基因组甲基化测序	55
目标区域捕获甲基化测序	58
ChIP-Seq	62

## 65

### 新产品

单管单细胞RNA测序	67
10X Genomics 单细胞	
RNA-Seq测序	72
单细胞DNA测序	76
免疫组库测序	81



## 华大简介

- 01 | 公司简介
- 02 | 平台介绍
- 03 | 质量管理

关于  
我们

About  
Us

## 公司简介



深圳华大基因股份有限公司（简称华大基因）是华大集团下属子公司，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，数百台高性能测序仪、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、穗等主要城市设有分支机构和临床检验中心，并在中国香港、欧洲、美洲、亚太等地区设有核心实验室，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，以及国内外的各级医院、体检机构等医疗卫生机构、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及科研机构超过2000家。

华大科技于2012年完成整合，致力于成为全球生命科学研究机构的首选合作伙伴。为从事生命科学研究的机构和企业提供高质量、行业领先的新一代测序、信息分析、基因分型、蛋白和代谢组质谱检测、Sanger测序、Oligo合成、生物云计算等标准化的生物技术服务，也可依据用户的个性化需求提供定制产品服务。

目前，华大科技已经主导及参与发表了CNNS文章超过100篇，全部科研论文1000余篇。

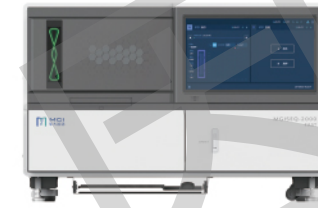
## 平台介绍

### 核酸测序平台

拥有BGISEQ、Nanopore、PacBio、3730和Illumina HiSeq/MiSeq 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。



DNBSEQ-T7



MGISEQ-2000



BGISEQ-500



BGISEQ-200



BGISEQ-50



Nanopore平台 (PromethION)



Nanopore平台 (MiniION)



PacBio平台 (Sequel I、II)



HiSeq 2500/4000



MiSeq



## ▶▶ 质谱平台

拥有包括Orbitrap Fusion Lumos、Q Exactive/HF/HF-X、QTRAP 6500+、Triple TOF 5600+、Xevo G2-XS、TQS等涵盖轨道阱、飞行时间、三重四级杆以及各类离子源和碎裂模式的40余套质谱仪系统，配套高效样本制备和分子分离平台，能实现工业规模和科研层次的蛋白质组学、多肽组学、代谢组学研究及目标分子检测。



Orbitrap Fusion Lumos



Q-Exactive/ Q-Exactive HF/  
Q-Exactive HF-X



QTRAP 6500+



Xevo-G2-XS



TQS/TQD



TSQ Altis

## ▶▶ 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 820.1028 T flops，总内存容量达到 243.144 TB，总存储能力达到 104.3152 PB (截至 2018 年 3 月)。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、处理和提供稳定而高效的资源保障。下图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



武汉平台



杭州平台



香港平台

## ▶▶ 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。



芯片分型平台-iScan



芯片分型平台-Affymetrix Genetian



质谱分型平台-MassARRAY

## 质量管理

华大科技于2010年通过了ISO9001质量管理体系标准认证；于2011年12月通过了ISO14001环境管理体系认证和ISO18001职业健康安全管理体系认证；2012年6月通过了ISO/IEC 27001信息安全管理标准认证；2015年7月华大基因香港高通量测序实验室获美国病理学会 (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。





## 基因组学

- 01 | 人全基因组测序
- 02 | 外显子测序
- 03 | 目标区域测序

产品  
简介

Product  
Profile

# 人全基因组测序

## 产品概述

人全基因组重测序 (WGS) 基于人类参考基因组序列进行全基因组测序, 在个体或群体水平更全面地挖掘基因序列差异和结构变异, 在全基因组水平上扫描并检测与表型差异、疾病、群体进化等相关的突变位点。

随着测序成本降低、云计算平台的推广及科研人员对于大数据分析能力的提高, 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而有效的方法之一。

## 产品优势

- 测序准确性高**  
采用独特的DNB核心测序技术, 在扩增过程避免错误累积的发生, 有效提高测序准确度;
- Dup低, 可用数据多**  
测相同的数据量, BGISEQ平台有效测序数据要高出至少10%;
- 无Index hopping担忧**  
DNA纳米球技术的高index保真度, 可有效避免因标签错配带来的样本混淆;
- 从建库到测序的真正0 PCR**  
基于PCR-free的滚环扩增技术, 全程“0”PCR, 可多检测4万个真InDel;
- 严格的质控保驾护航**  
每一环节均有严格的质量控制流程, 保证每一文库每一样本的高效、快速、准确的执行;
- 经验丰富**  
主导及参与完成了多个国际大型基因组计划, 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、英国万人基因组等。

## 产品应用



## 技术流程



## 信息分析内容

- 基本数据统计**
  1. 数据过滤: 去除原始数据接头污染序列及低质量reads
  2. 数据比对: 通过比对软件与人参考序列对比
  3. 数据统计: 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

### 样本要求

样本类型	文库类型	总量	浓度	完整性(胶图)	纯度
Genomic DNA	PCR/0 PCR	≥1μg	≥12.5ng/μl	主峰>20kb	无蛋白, RNA/盐离子等污染, 样本无色、透明、不粘稠

组织类型	需求量
新鲜培养细胞(细胞数)	≥5×10 <sup>6</sup> cell
新鲜动物组织干重	≥50mg
新鲜植物组织干重	≥200mg
全血(哺乳动物)	≥1ml
菌体(细胞数或干重)	≥5×10 <sup>6</sup> cell or ≥200mg
FFPE	≥10片、未染色、100mm <sup>2</sup> 、5~10μm厚度

### 推荐数据量

- 群体进化 (大样品量低深度)
- 复杂疾病 (大样品量低深度)
- 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 30 个工作日



## ► 案例分析

### 案例一: 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基因的内含子区。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, 实验结果再次得到验证。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。

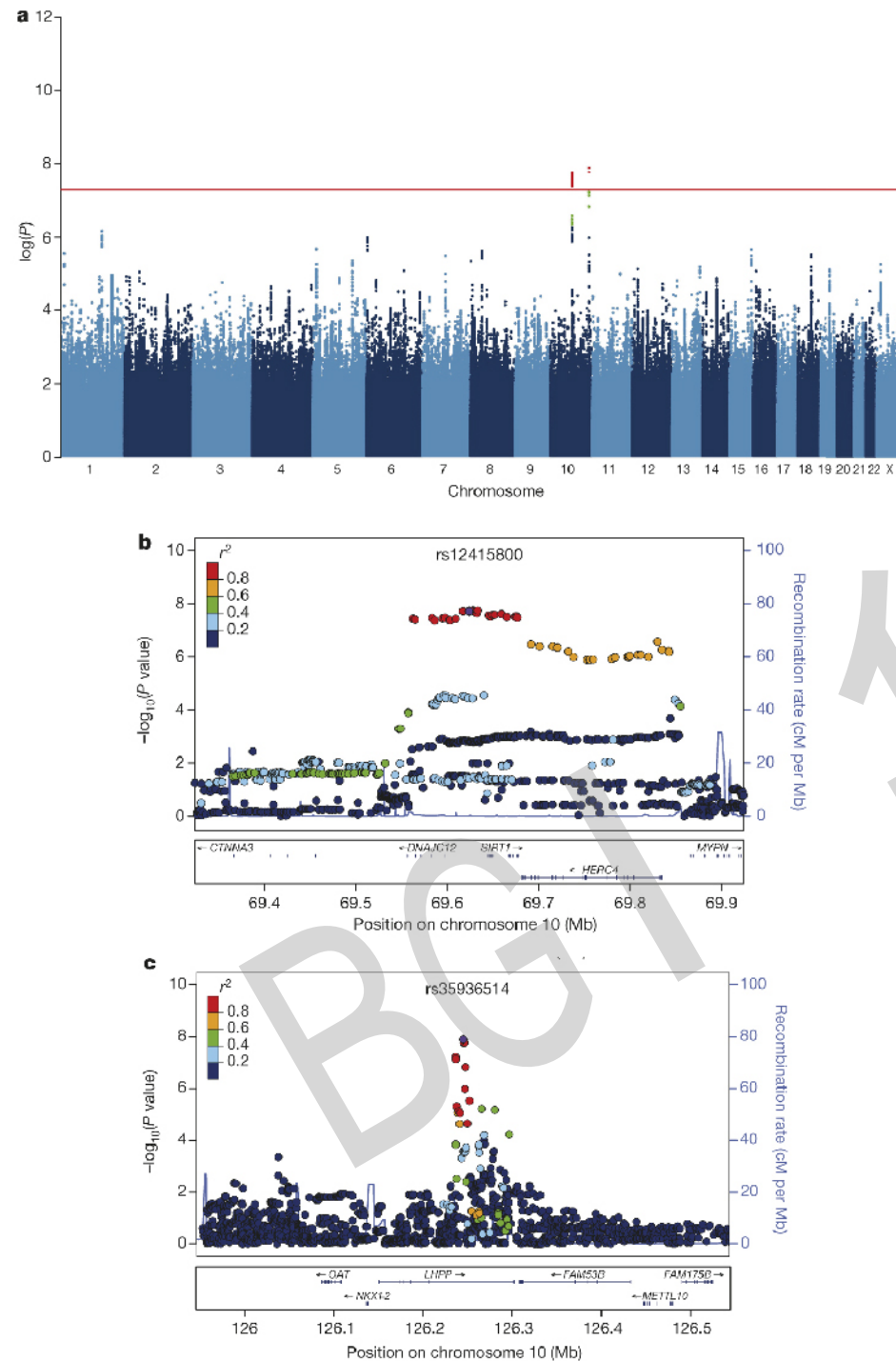


图1 此次研究样品中发现10号染色体的两个位点与重度抑郁症相关  
a. 重度抑郁症全基因组相关的曼哈顿图 b. 10号染色体69.6Mb位置的SIRT1相关信息  
c. 10号染色体126.2Mb位置的LHPP相关信息

### ★ 参考文献

Sparse whole-genome sequencing identifies two loci for major depressive disorder[J]. Nature, 2015, 523(7562):588-591.

### 案例二: 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, 其他三个均通过外显子测序确定。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。

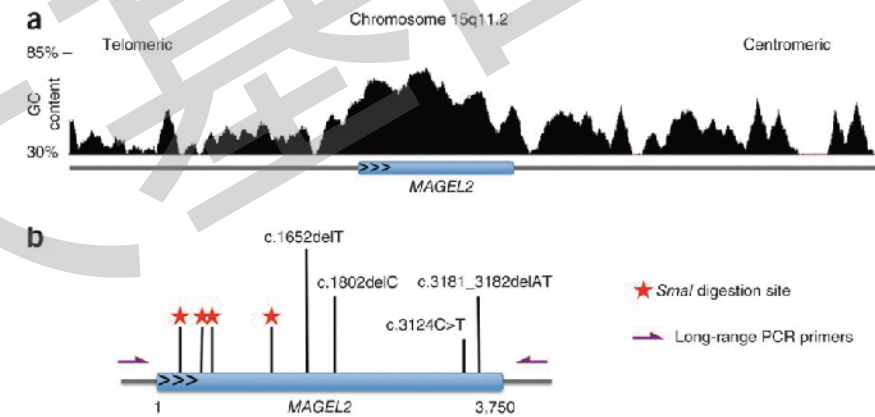


图2 MAGEL2父本等位基因的缺失突变  
a. 15q11.2染色体上MAGEL2以及侧翼序列的GC含量  
b. 本文报道的MAGEL2缺失突变是发生在它唯一的外显子区域

### ★ 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.[J]. Nature Genetics, 2013, 45(11):1405.

### 案例三: 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。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, 具有基因组重排和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。综上, 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。

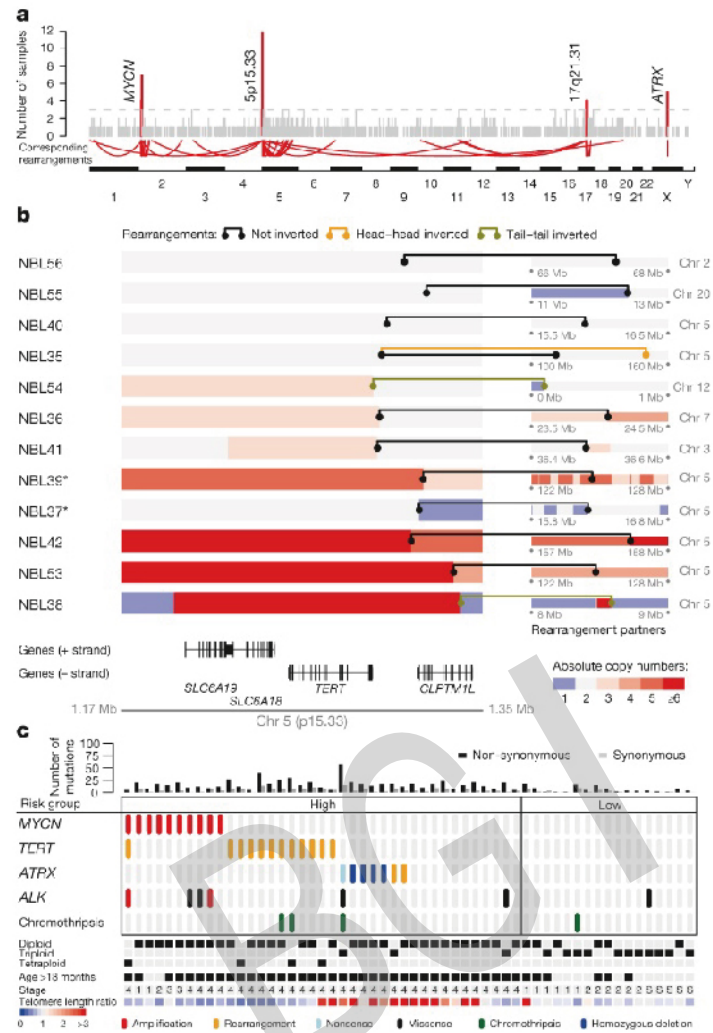


图3 基因组重排都集中在高风险神经母细胞瘤的5号染色体5p15.33区域

- 56个原发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
- 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
- 56个原发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基因组变化

#### ★ 参考文献

Telomerase activation by genomic rearrangements in high-risk neuroblastoma[J]. Nature, 2015.

### 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区域的平均深度较整个基因组的平均深度要低。

另一方面, 高GC也影响酶扩增时的准确性, 即高GC可能提高扩增时的错误率, 短读长测序本身也是一种合成/扩增, 因此也会提高测序错误率, 导致整条序列的错误率提高。错误率越高, reads比对时就有可能比对到错误的位置, 或者比对不上。

基于DNB技术的0 PCR测序不会在测序前信号放大阶段累积复制错误, 而基于指数扩增Cluster技术的0 PCR测序则无法避免此类错误。因此, BGISEQ 0 PCR WGS实现了从文库构建到测序过程中始终保持原始模板DNA的“原貌”, 有效避免PCR扩增偏向性, 获得真正的0 PCR数据。

#### Q: 人类基因组中常见变异类型有哪些?

A: 详情见表1

表1 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1,000bp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。由于重排引导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, 最终得到少量的候选突变。



► 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
人类基因组计划	2001.02	Nature	Initial sequencing and analysis of the human genome
单体型图谱	2003.12	Nature	The International HapMap Project
	2005.01	Nature	A haplotype map of the human genome
炎黄一号	2008.11	Nature	The diploid genome sequence of an Asian individual
古人	2010.01	Nature	Ancient human genome sequence of an extinct Palaeo-Eskimo
千人	2010.02	Nature	A map of human genome variation from population-scale sequencing
HapMap	2010.09	Nature	Integrating common and rare genetic variation in diverse human populations
澳洲土著人	2011.01	Science	An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia
LFR方法	2012.04	Nature	Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells
千人	2012.07	Nature Methods	The 1000 Genomes Project: data management and community access
	2015.01	Nature	A global reference for human genetic variation
	2015.01	Nature	An integrated map of structural variation in 2,504 human genomes
膀胱癌	2013.12	Nature Genetic	Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation
单基因病	2013.11	Nature Genetics	Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism
荷兰人	2014.02	European Journal of Human Genetics	The Genome of the Netherlands: design, and project goals
	2014.08	Nature Genetics	Whole-genome sequence variation, population structure and demographic history of the Dutch population
	2015.03	Nature Communications	Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels
	2015.07	Nature Genetics	Genome-wide patterns and properties of de novo mutations in humans
智力障碍	2014.07	Nature	Genome sequencing identifies major causes of severe intellectual disability

研究内容	发表时间	发表期刊	文献标题
UK10K	2015.09	Nature Communications	Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel
	2015.10	Nature	The UK10K project identifies rare variants in health and disease
	2015.10	Nature	Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture
	2015.11	Nature Genetics	Height-reducing variants and selection for short stature in Sardinia
老年性黄斑病变	2015.01	Nature Communications	New loci and coding variants confer risk for age-related macular degeneration in East Asians
宫颈癌	2015.02	Nature Genetics	Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism
神经母细胞瘤	2015.10	Nature	Telomerase activation by genomic rearrangements in high-risk neuroblastoma
癌症研究	2015.06	Human Genetics	Identification of cancer predisposition variants in apparently healthy individuals using a next-generation sequencing-based family genomics approach
肺癌	2019.02	Cancer Research	Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer
LFR方法	2015.01	Genetics	Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing
重度抑郁症	2015.07	Nature	Sparse whole-genome sequencing identifies two loci for major depressive disorder
PGD	2015.03	Genome Research	Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing
StLFR方法	2019.05	Genome Research	Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly

# 外显子测序

## 产品概述

人类基因组全部外显子区域的集合称为外显子组，是基因中重要的编码蛋白的部分，并涵盖了与个体表型相关的大部分功能性变异。全外显子组测序 (whole exome sequencing, WES) 是利用序列捕获技术，将全基因组外显子区域的DNA捕捉并富集后，进行高通量测序的基因组分析方法。

与全基因组测序相比，WES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，WES可鉴定低频和罕见变异，结果分析更准确。因此，近年来WES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。

华大基因作为工业级的外显子测序领导者，以完成超过12万例外显子样品数据的稳定性为基础，为您提供最有保障的服务。

## 产品优势



**高效精准**  
直接对蛋白编码序列测序，更适合高深度测序；



**质量卓越**  
DNBSEQ™对InDel检测有更好的灵敏度；



**省力好用**  
1. 基于WES的GWAS可将关联变异位点锁定在编码区内，能够为致病机制的阐明、疾病的治疗提供更丰富的线索；  
2. 应用于表型复杂的疑似遗传病诊断，发现新基因以及重分析；  
3. 是肿瘤精准免疫治疗必检项目，能够预测肿瘤新抗原结构及功能，可作为一线诊断工具；



**质控完整**  
严格规范的项目流程和质量体系，为项目质量保驾护航；



**经验丰富**  
完成超过12万例样品，发表相关文章超过200篇；



**研究广泛**  
涉及21种肿瘤类型，28种复杂疾病；罕见病研究性文章超过96篇。

## 产品应用

### 癌症

- 食管癌
- 肾上腺肿瘤
- 肾癌
- 膀胱癌
- ...

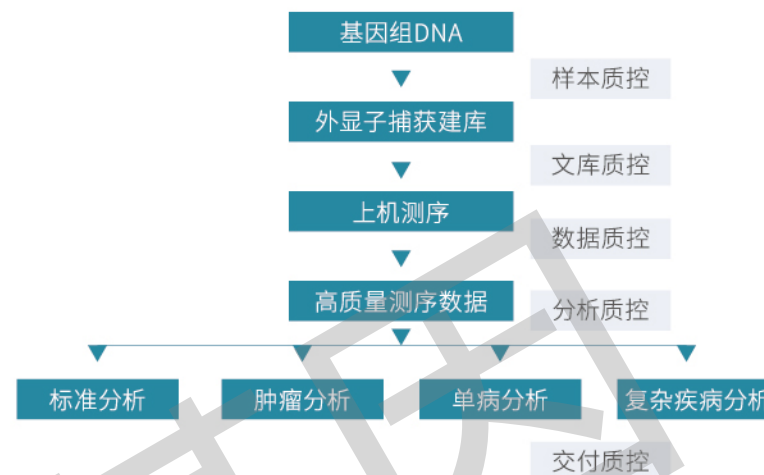
### 复杂疾病

- 糖尿病
- 银屑病
- 克罗恩病
- 慢性肝炎
- ...

### 罕见病

- 汗孔角化症
- 先天性黑蒙
- 遗传性耳聋
- 小脑共济失调
- ...

## 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. SNP/InDel检测、多种数据库注释、保守性预测、致病性分析、统计

### 癌症研究

1. 体细胞SNV/InDel检测和注释
2. 体细胞CNV检测和注释
3. 易感基因筛查
4. 配对样品同源性验证
5. 肿瘤突变特征分析
6. 高频突变基因统计和富集分析
7. 高频突变基因相关性分析
8. 突变可视化展示 (突变全景图)
9. 驱动基因的预测
10. 肿瘤克隆演化分析
11. 肿瘤进化树分析
12. 肿瘤免疫治疗
13. 生存分析

### 复杂疾病研究 (散发)

1. 基于遗传数据的样本质控
2. 基于单个SNP的关联分析
3. 基于基因和pathway的关联分析
4. 候选变异、基因优化
5. 候选SNP的独立性分析
6. 多组织多数据库的eQTL分析
7. 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
8. GO功能分类, 以及GO/KEGG功能富集分析
9. 基于通路或蛋白交互网络数据库的交互分析
10. ROC曲线与遗传方差解释
11. 线粒体异质性和位点关联分析
12. CNV关联分析和区域解读

### 单基因病研究

1. 多种软件数据比对、变异检测
2. 变异效应预测工具(VEP)注释
3. 群体等位基因频率注释
4. 有害性或保守性预测工具打分、通路相关注释
5. OMIM相关注释、正常组织蛋白表达注释
6. 增加家系共分离筛选标签
7. 基因功能影响、变异可靠性分析
8. 核心家系de novo突变、近亲家系纯合子区分析



► 技术参数

样品要求/样品类型	体积/总量
组织	≥0.5 g
全血	≥2 ml
福尔马林固定石蜡包埋组织(FFPE)	>10个玻片或50mm <sup>2</sup> , 5-10μm厚的切片10片
常规基因组DNA	m≥1μg, c≥12.5 ng/μL
微量基因组DNA	>200ng, c≥2.5ng/μL
FFPE基因组DNA	m≥1μg, c≥10 ng/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
PE101、PE151等	100X, 200X, 500X以上 (cfDNA)	短至11天

► 案例分析

案例一: 全外显子测序在单基因遗传性疾病中的应用

心血管疾病 (CVD) 是世界范围内的一类重要的致死疾病, 其中血液中高浓度的低密度脂蛋白胆固醇 (low-density lipoprotein cholesterol, LDL-C) 是它的一个主要危险因素之一。LDL-C如果浓度过高, 会沉积在动脉壁中, 形成斑块, 造成血管堵塞, 从而引发心血管疾病。该文章选用的是家系样本。对三个LDL-C低浓度个体和一个正常个体进行全外显子测序, 过滤后, 只剩LDL-C个体的突变位点, 接着通过过滤同义突变、dbSNP数据库、SIFT分值高而POLYPHEN V2和Mutation Taster分值低的突变位点。经全外显子测序以及Sanger测序验证最终发现了一个未知突变LIMA1-K306fs, 后对1000多个哈萨克族个体全基因组的LIMA1基因片段进行靶向测序, 发现另外3个家系的LIMA1基因中含有L25I突变 (LIMA1-L25I (Leu→Ile))。

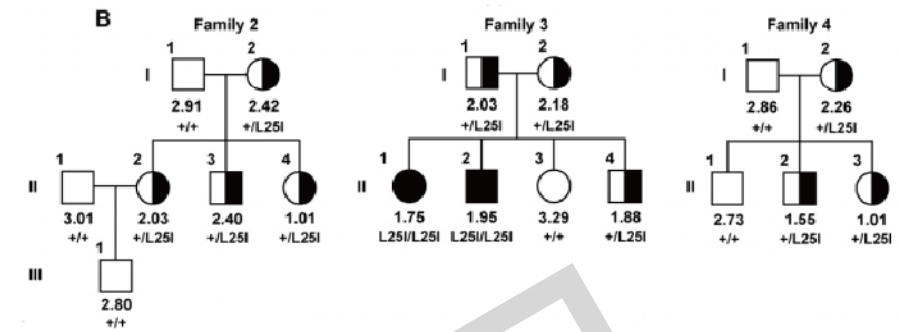
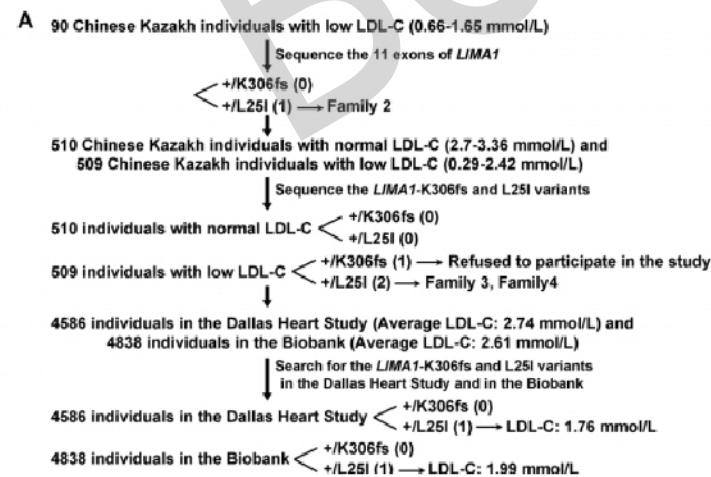


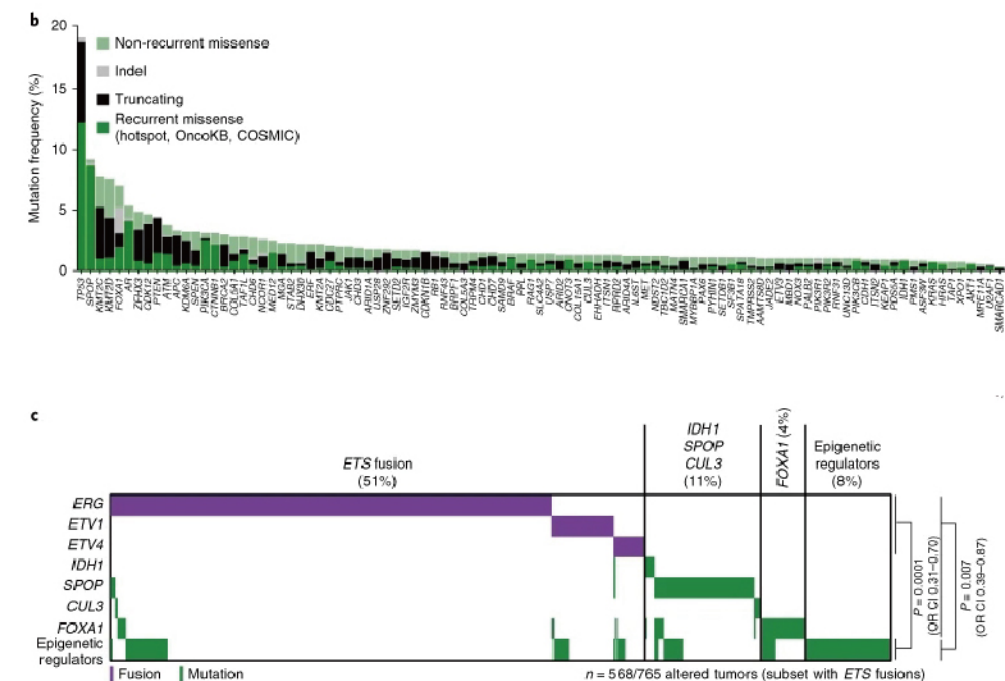
图1 3个中国哈萨克族低水平LDL-C家族LIMA1-L25I突变鉴定示意图

★ 参考文献

Zhang Y Y, Fu Z Y, Wei J, et al. A LIMA1 variant promotes low plasma LDL cholesterol and decreases intestinal cholesterol absorption[J]. Science, 2018, 360(6393): 1087-1092.

案例二: 前列腺癌致癌基因的长尾效应分析

前列腺癌的深度基因组学分析已经鉴定出一些复发性相关的变异基因, 这些基因参与雄激素信号传导, 比如DNA修复和PI3K信号传导等基因。然而, 更大规模的基因组学分析可以鉴定出一些其他低频的反复突变基因。在这篇文章中, 研究人员汇总并统一分析来自1,013个前列腺癌的外显子组测序数据。鉴定和验证了一类新的由表观遗传调节因子中的突变所定义的E26转化特异性 (E26 transformation-specific, ETS) -融合阴性肿瘤, 以及在之前的前列腺癌研究中未涉及到的途径中的变异, 如剪接体途径。同时作者还发现显著突变基因 (significantly mutated genes, SMG) 的突变率遵循长尾分布, 许多基因的突变率不到3%。作者总共确定了97个SMG, 包括70个在之前的研究中未报道的前列腺癌SMG, 例如泛素连接酶CUL3和转录因子SPEN。最后, 通过比较原发性和转移性前列腺癌中的突变位点信息, 鉴定出一组可以预测前列腺癌危险分层的基因组标记。





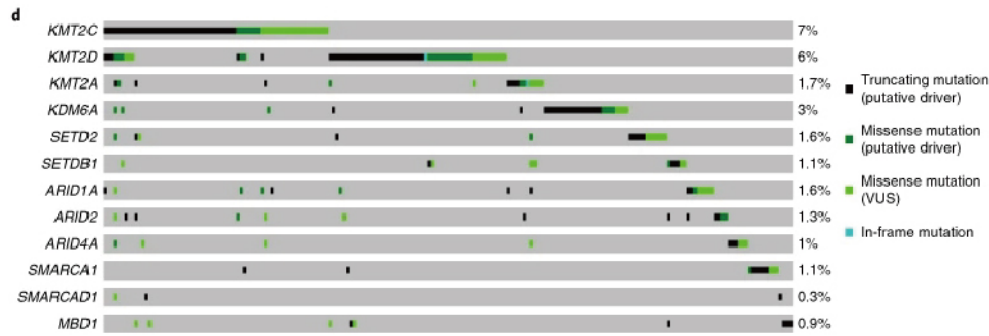


图2 1,013个前列腺癌的突变显著基因

## ★ 参考文献

Armenia J, Wankowicz SAM, Liu D, et al. The long tail of oncogenic drivers in prostate cancer[J]. Nature genetics, 2018, 50(5): 645.

## 案例三: 发现与糖尿病相关的罕见DNA突变

目前,全基因组关联分析(GWAS)是寻找疾病相关变异非常流行的一种方法。这种方法可以非常有效地在整个基因组中发现常见的疾病变异,但缺点是可能会漏掉不太常见的外显子变异。这项研究以外显子测序为手段,分析了近5万人(40X)的蛋白质编码基因,鉴定出与2型糖尿病相关的新型罕见变异。这一发现或有助于改进对2型糖尿病的特征鉴别和治疗。通过外显子组关联分析找出7个位点上15个变体表现出显著关联,其中2个是过去GWAS没有发现的新变异。在基因级别上,有3个基因达到显著关联。

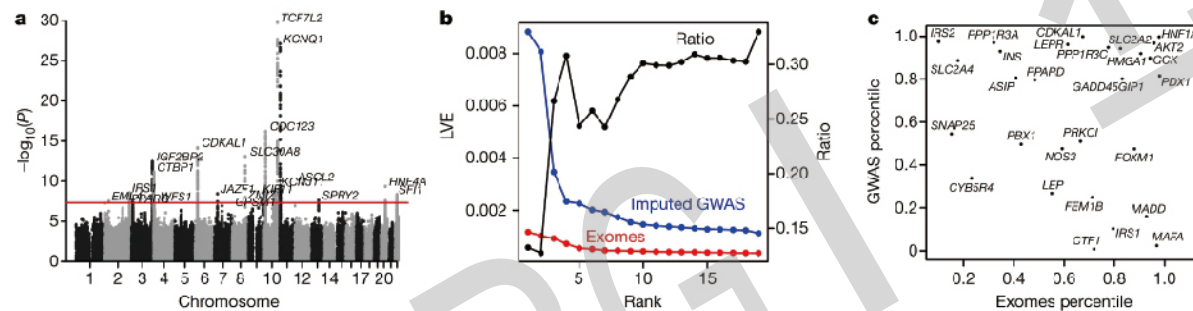


图3 外显子组测序与基于阵列的GWAS的比较

## ★ 参考文献

Jason Flannick et al., (2019) Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature. DOI: 10.1038/s41586-019-1231-2

## ► 常见问题

## Q: 何时选择全外显子捕获? 何时选择定制探针捕获?

A: 人类和一些常见物种的全外显子捕获都有预设的探针产品,其中人类全外显子探针已优化过多个版本。如果目标区域很多且均为外显子(几十Mb以上),建议直接选择全外显子捕获,因为全外显子有设计更成熟的探针组和更低的成本,捕获效果更好。对于较小的目的片段,或者也关注外显子以外的区段,建议采用定制探针。

## Q: 全外显子组测序深度的意义是什么? 测序深度如何换算?

A: 测序深度代表了序列被探针组覆盖的次数,次数越高,测序结果的识别就越精确,后续的统计分析也就越准确。如果做肿瘤、低频突变研究,建议测序深度至少应达到200X以上。如果只看经典SNP、非低频突变,测序深度也至少应该在50X以上。测序深度换算方法:一般目标区域的捕获效率在60-70%,安捷伦和罗氏等外显子捕获试剂盒的目标区域大小在60Mb左右,即测序深度=  $\frac{\text{比对到基因组上的有效数据(去除DUP)} \times 10G \times \text{捕获效率} 60\%}{\text{目标区域大小} 60Mb} = 100X$

## ► 华大合作发表文章(部分)

种类	研究内容	发表时间	发表期刊	文献标题
癌症	膀胱癌	2011.08	Nature Genetics	Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder
	肾癌	2011.11	Nature Genetics	Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma
	白血病	2013.02	Nature Genetics	Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia
	前列腺癌	2013.06	Adv. Mater	High-Purity Prostate Circulating Tumor Cell Isolation by a Polymer Nanofiber-Embedded Microchip for Whole Exome Sequencing
	移行细胞癌	2013.06	European Urology	Somatic Mutation of the Androgen Receptor Gene Is Not Associated with Transitional Cell Carcinoma: A "Negative" Study by Whole-exome Sequencing Analysis
	膀胱癌	2013.11	Nature Genetics	Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation
	胰岛素瘤	2013.11	Nature Communications	Whole exome sequencing of insulinoma reveals recurrent T372R mutations in YY1
	食管癌	2014.03	Nature	Identification of genomic alterations in oesophageal squamous cell cancer
	食管鳞状细胞癌	2014.05	Nature	Identification of genomic alterations in oesophageal squamous cell cancer
	肾上腺库欣综合征	2014.05	Science	Activating Hotspot L205R Mutation in PRKACA and Adrenal Cushing's Syndrome
	结直肠癌	2014.06	Gut	Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer

种类	研究内容	发表时间	发表期刊	文献标题
癌症	大B细胞淋巴瘤	2014.08	Blood	Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients
	食管癌	2015.04	The American Journal of Human Genetics	Genomic Analyses Reveal Mutational Signatures and Frequently Altered Genes in Esophageal Squamous Cell Carcinoma
	食管鳞状细胞癌	2015.04	The American Journal of Human Genetics	Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma
	多同步肺癌	2016.01	Nature Communications	Genomic heterogeneity of multiple synchronous lung cancer
	肺癌	2016.01	Nature Communications	Genomic heterogeneity of multiple synchronous lung cancer
	肝癌	2017.03	Journal of Hepatology	Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma
	前列腺癌	2017.03	European Urology	Stromal Gene Expression is Predictive for Metastatic Primary Prostate Cancer
	食管小细胞癌	2018.05	Cell Research	The genomic landscape of small cell carcinoma of the esophagus
罕见病 / 遗传性疾病	脊髓小脑共济失调	2010.12	Brain	TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing
	发作性运动障碍	2011.12	Brain	Identification of PRRT2 as the causative gene of paroxysmal kinesigenic dyskinesias
	汗孔角化	2012.10	Nature Genetics	Exome sequencing identifies MVK mutations in disseminated superficial actinic porokeratosis
	先天性黑蒙	2012.09	Nature Genetics	Exome sequencing identifies NMNAT1 mutations as a cause of Leber congenital amaurosis
	Olmsted 综合征	2012.03	American Journal of Human Genetics	Exome Sequencing Reveals Mutations in TRPV3 as a Cause of Olmsted Syndrome
	遗传性耳聋	2014.06	Cell Research	De novo mutation in ATP6V1B2 impairs lysosome acidification and causes dominant deafness-onychodystrophy syndrome
	遗传性骨髓衰竭	2014.09	Blood	Inherited bone marrow failure associated with germline mutation of ACD, the gene encoding telomere protein TPP1
	Keppen-Lubinsky 综合征	2015.01	American Journal of Human Genetics	Keppen-Lubinsky Syndrome Is Caused by Mutations in the Inwardly Rectifying K <sup>+</sup> Channel Encoded by KCNJ6

种类	研究内容	发表时间	发表期刊	文献标题
罕见病 / 遗传性疾病	肢端点状角化病、唇炎、指节垫	2015.02	American Journal of Human Genetics	Loss-of-Function Mutations in CAST Cause Peeling Skin, Leukonychia, Acral Punctate Keratoses, Cheilitis, and Knuckle Pads
	线粒体脑肌病	2015.06	American Journal of Human Genetics	RNASEH1 Mutations Impair mtDNA Replication and Cause Adult-Onset Mitochondrial Encephalomyopathy
	皮肤疾病	2016.01	Nature Genetics	Stabilizing mutations of KLHL24 ubiquitin ligase cause loss of keratin 14 and human skin fragility
	遗传性脑病	2016.06	Brain	CSF1R mosaicism in a family with hereditary diffuse leukoencephalopathy with spheroids
	早发性肌张力障碍	2016.12	Nature Genetics	Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia
复杂疾病	糖尿病	2010.01	Nature Genetics	Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants
	乙肝	2012.11	Hepatology	Rare inborn errors associated with chronic hepatitis B virus infection
	克罗恩病	2013.08	Gastroenterology	Association Between Variants of PRDM1 and NDP52 and Crohn's Disease, Based on Exome Sequencing and Functional Studies
	中丹糖尿病三	2013.11	The American Journal of Human Genetics	Whole-Exome Sequencing of 2,000 Danish Individuals and the Role of Rare Coding Variants in Type 2 Diabetes
	银屑病	2014.06	Nature Genetics	A large-scale screen for coding variants predisposing to psoriasis
	免疫失调综合征	2015.04	Journal of Allergy and Clinical Immunology	Novel mutations in TNFRSF7/CD27: Clinical, immunologic, and genetic characterization of human CD27 deficiency
	老年性黄斑变性	2015.04	Nature Communications	Whole-exome sequencing implicates UBE3D in age-related macular degeneration in East Asian populations
	慢性肝炎	2015.04	Hepatology	The p.Ser267Phe variant in SLC10A1 is associated with resistance to chronic hepatitis B
	主要抗体不足	2017.08	Journal of Allergy and Clinical Immunology	Plasma cell deficiency in humans with heterozygous mutations in SEC61A1
	乙型肝炎病毒	2018.03	Blood	Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma



# 目标区域捕获测序

## 产品概述

目标序列捕获测序是将感兴趣的基因组区域定制成特异性探针与基因组DNA在序列捕获芯片或溶液中进行杂交，将目标基因组区域的DNA片段进行富集后再利用高通量测序技术进行测序的方法。研究人员只需对基因组中感兴趣的候选区域进行测序就可满足研究需求，大幅缩小了测序区域，极大地降低了成本，非常适合大样本量研究。

通过目标区域测序，可以对候选位点或候选基因进行验证，也可以进一步找到候选区域或候选基因内的易感位点，并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



**研究区域明确**  
只需对感兴趣的区域及相关基因进行研究；



**性价比高**  
适用于大样本量的候选基因筛查、小区域的高深度研究；



**项目经验丰富**  
累计已完成数万个样品的检测和分析。

## 产品应用

对于肿瘤临床诊断来说，靶向测序是一种便宜、好用、易于解释的好工具。



## 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
- TargetSeq enrichment kit ; MultipSeq enrichment kit

## ★ 目录产品捕获试剂盒

- 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不同)

## 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. SNP/InDel检测、多种数据库注释、保守性预测、致病性分析、统计

### • 癌症研究

1. 体细胞SNV/InDel检测和注释
2. 体细胞CNV检测和注释
3. 易感基因筛查
4. 配对样品同源性验证
5. 肿瘤突变特征分析
6. 高频突变基因统计和富集分析
7. 高频突变基因相关性分析
8. 突变可视化展示 (突变全景图)
9. 驱动基因的预测
10. 肿瘤克隆演化分析
11. 肿瘤进化树分析
12. 肿瘤免疫治疗
13. 生存分析

### • 复杂疾病研究 (散发)

1. 基于遗传数据的样本质控
2. 基于单个SNP的关联分析
3. 基于基因和pathway的关联分析
4. 候选变异、基因优化
5. 候选SNP的独立性分析
6. 多组织多数据库的eQTL分析
7. 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
8. GO功能分类, 以及GO/KEGG功能富集分析
9. 基于通路或蛋白交互网络数据库的交互分析
10. ROC曲线与遗传方差解释
11. 线粒体异质性和位点关联分析
12. CNV关联分析和区域解读

### • 单基因病研究

1. 多种软件数据比对、变异检测
2. 变异效应预测工具(VEP)注释
3. 群体等位基因频率注释
4. 有害性或保守性预测工具打分、通路相关注释
5. OMIM相关注释、正常组织蛋白表达注释
6. 增加家系共分离筛选标签
7. 基因功能影响、变异可靠性分析
8. 核心家系de novo突变、近亲家系纯合子区分析

## 技术参数

- 样品总量  $\geq 1\mu\text{g}$  DNA, 微量建库样品量低至200ng
- 样品浓度  $\geq 12.5\text{ng}/\mu\text{L}$

## ► 案例分析

### 案例一: 中国人群MHC区域的深度测序有助于复杂疾病的研究

研究者对20,635名汉族血统(10,689名对照和9,946名银屑病患者)中的整个5Mb MHC区域进行了测序,并构建了一个Han-MHC数据库,其中包括两种变体和高精度的HLA基因分型结果。进一步确定了HLA-C, HLA-B, HLA-DPB1和BTNL2中的多个独立的新易感位点与牛皮癣相关的基因间变体rs118179173,证实了已确立的风险等位基因HLA-C\*06:02。研究者预计,通过对大量样品进行深度测序而构建的Han-MHC参考组将作为研究MHC区域在多种疾病中的作用的有效工具,从而促进对这些疾病发病机理的理解。

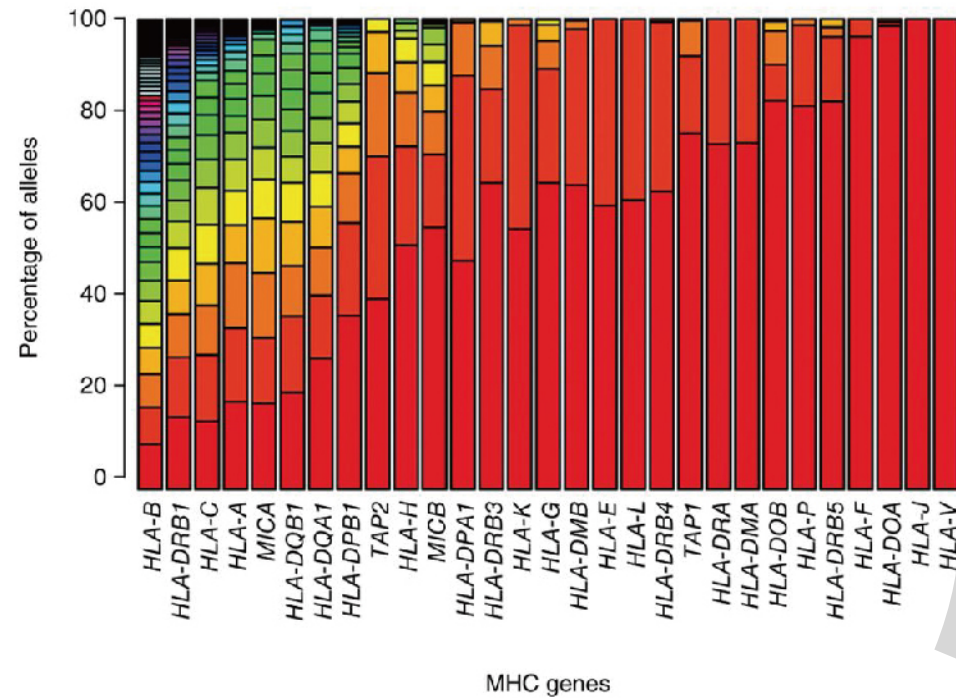


图1 汉族人群中29种MHC基因的多样性

### ★ 参考文献

Zhou F, Cao H, Zuo X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease[J]. Nature genetics, 2016, 48(7): 740.

## ► 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1.5-2个月。

### Q: 捕获效率有哪些因素影响? 推荐多少数据量?

A: 捕获效率影响因素有样本质量, 区域复杂度等。由于捕获效率未知, 我们无法承诺有效深度, 但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来进行, 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K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。



## 转录组学

- 01 | 转录组测序
- 02 | 全长转录组测序
- 03 | RNA-Seq
- 04 | Small RNA测序
- 05 | 长链非编码RNA测序

产品  
简介

Product  
Profile



# 转录组测序

## 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，提供最全面的转录组信息。通过高通量测序，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、临床诊断和药物研发等领域。

## 产品优势

- 全面丰富的样品类型**  
小鼠样本常规建库仅需200ng，提供单细胞、FFPE、微量样品的特殊样品服务，微量建库低至pg级；
- 高质量的自主测序平台**  
滚环扩增构建DNB测序文库，PCR扩增错误不累积，无index hopping之忧，低Dup rate无需人为干预，《Nature》、《Cell》、《Immunity》等1000+累积影响因子；
- Dr. Tom系统解决个性化难题**  
只需任意一种RNA测序数据，就获得多组学关联信息，多数据库联合分析，多维度数据展示，循环挖掘数据，轻松锁定解释生物学问题的核心基因；
- 极速交付**  
从提取和检测开始，到Dr.Tom个性化分析，每个环节都极速。

## 产品应用



## 技术流程

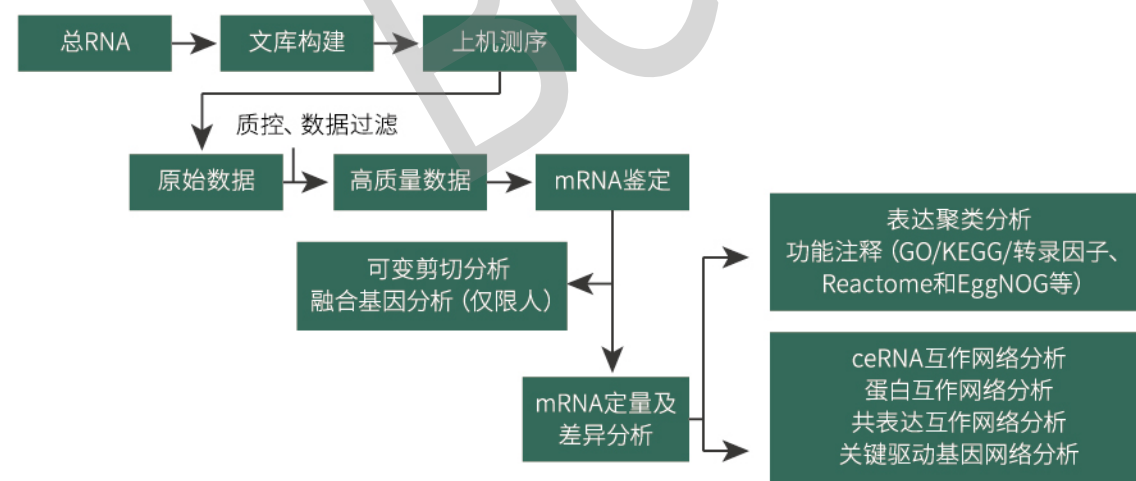


图1 转录组测序技术流程

## 信息分析内容

- 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reads信息, ② 样品相关性, ③ 表达量分布, ④ RNA分类;
  2. 参考基因组比对;
  3. mRNA鉴定;
  4. mRNA定量分析;
  5. mRNA差异表达分析 (样本间、组间);
  6. mRNA表达/差异基因聚类;
  7. mRNA差异基因GO分类、富集;
  8. mRNA差异基因KEGG分类、富集;
  9. mRNA结构分析: ① 可变剪切分析, ② 融合基因分析 (仅限人)。
- Dr.Tom信息分析**
  - 数据库注释
    1. 转录因子注释 (AnimalTFDB/PlantTFDB);
    2. GSEA分析 (仅限人);
    3. Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释。
  - 互作网络分析
    1. 靶基因分析① miRNA-mRNA靶向关系分析, ② lncRNA-mRNA靶向关系分析;
    2. ceRNA互作网络分析① lncRNA-mRNA联合分析, ② circRNA-mRNA联合分析 (仅限人、小鼠);
    3. 蛋白互作网络分析;
    4. 共表达互作网络分析。
  - 特色分析
    1. 自定义标签和自有数据上传;
    2. 外部数据库信息 (TCGA、ARCHS4);
    3. 卡方检验;
    4. 关键驱动基因网络图分析;
    5. 时间序列分析。

## 技术参数

- 样本要求**

样品类型: 完整且无污染的总RNA, 样品需求量 (单次): 200ng,  
样品浓度: 10-1,000 ng/μL,  
样品纯度: 28S/18S≥1.0; RIN≥7.0,
- 测序策略** PE100或PE150
- 推荐数据量** 6G或10G
- 项目执行周期**

标准执行周期为24个工作日, 极致交付周期为14个工作日 (含样品检测)。

## ► 案例分析

### 案例一：转录组研究乙肝病毒感染增加B细胞淋巴瘤患病风险

乙型肝炎病毒 (HBV) 感染还可能会增加几种B细胞淋巴瘤的发病率。对108个中国人B细胞淋巴瘤患者进行转录组测序, 包括24个同时感染HBV的患者。与未感染HBV的人相比, 感染者具有独特的基因表达图谱, 有377个上调基因和324个下调基因。进一步研究显示感染HBV的B细胞淋巴瘤患者, HBV相关的基因表达特征由BCL6、FOXO1和ZFP36L1调节的基因富集构成, 这与这些转录因子在感染者中倾向突变是一致的。研究首次揭示了此类病人特有的分子学特征, 并证明了HBV感染与恶性B细胞淋巴瘤发生有着直接的联系。

#### • 结果展示

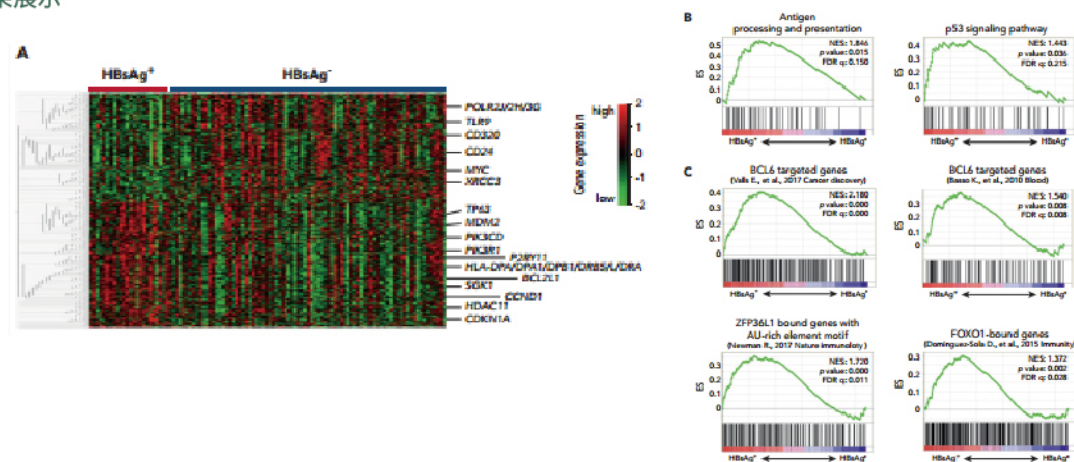


图2 转录组测序显示感染者具有独特的基因表达图谱

#### ★ 参考文献

Ren W, Ye X, Su H, et al. Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma[J]. Blood, 2018, 131(24): 2670-2681.

## ► 常见问题

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。

Q: 可否提供BGISEQ转录组标准品测试数据供客户测评?

A: 可以, 我们1个UHRR标准品, 构建了3个文库, 数据都已经上传到EBI。



←扫描二维码, 即可获取标准品数据

## ► 华大合作发表文章(部分)

物种	发表时间	发表期刊	文献标题
人	2017. 01	Cell Reports	The Anti-Warburg Effect Elicited by the cAMPPGC1 $\alpha$ Pathway Drives Differentiation of Glioblastoma Cells into Astrocytes
人B细胞淋巴瘤	2018. 03	Blood	Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma
小鼠	2018. 11	Annals of the Rheumatic Diseases	Glucocorticoid receptor in stromal cells is essential for glucocorticoid-mediated suppression of inflammation in arthritis
斑马鱼	2019. 01	Chemosphere	Microcystin-LR exposure induced nephrotoxicity by triggering apoptosis in female zebrafish
人	2019. 01	Parasitology Research	Toxoplasma gondii ROP17 inhibits the innate immune response of HEK293T cells to promote its survival
人细胞系	2019. 02	Nucleic Acids Research	Diminished OPA1 expression and impaired mitochondrial morphology and homeostasis in Aprataxin-deficient cells
小鼠	2019. 03	Neuropsychiatric Disease and Treatment	Identification of key genes, pathways, and miRNA/mRNA regulatory networks of CUMS-induced depression in nucleus accumbens by integrated bioinformatics analysis
小鼠	2019. 03	Molecular Cancer	Hypoxic BMSC-derived exosomal miRNAs promote metastasis of lung cancer cells via STAT3-induced EMT



# 全长转录组测序

## 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。PacBio的Sequel测序平台测转录组，因其长读长优势，可以直接得到全长转录本信息，已被越来越多地用于构建/完善参考物种基因组。华大基因基于PacBio平台，在常规全长转录组的基础上开发了独特的UMI全长转录组产品、5X全长转录组产品以及polyA+全长转录组产品，以更低的成本为大家提供定性、定量的全方位调控研究方案。

## 产品应用&产品优势

产品名称	产品应用	产品优势
5X全长转录组	<ul style="list-style-type: none"> <li>构建参考基因组</li> <li>基因功能研究</li> <li>辅助基因组注释</li> <li>转录本结构变异研究</li> <li>基因/转录本差异研究</li> </ul>	<ul style="list-style-type: none"> <li>同样的数据量，可得到5倍的有效数据</li> <li>检测到的基因/转录本数目多</li> <li>基因/转录本定量结果准</li> <li>其他优势同“常规全长转录组”</li> </ul>
PolyA+全长转录组	<ul style="list-style-type: none"> <li>PolyA相关研究</li> <li>构建参考基因组</li> <li>基因功能研究</li> <li>辅助基因组注释</li> <li>转录本结构变异研究</li> </ul>	<ul style="list-style-type: none"> <li>可检测PolyA长度分布</li> <li>可检测PolyA尾中非A碱基类型及出现频率</li> <li>研究揭示polyA在表达调控中的作用</li> <li>其他优势同“常规全长转录组”</li> </ul>
常规全长转录组	<ul style="list-style-type: none"> <li>构建参考基因组</li> <li>基因功能研究</li> <li>辅助基因组注释</li> <li>转录本结构变异研究</li> </ul>	<ul style="list-style-type: none"> <li>不需要组装，就可以准确的得到全长转录本的序列信息</li> <li>可以发现新的基因和转录异构体</li> <li>鉴定可变剪接和基因融合</li> <li>辅助基因组注释</li> </ul>

## 技术流程

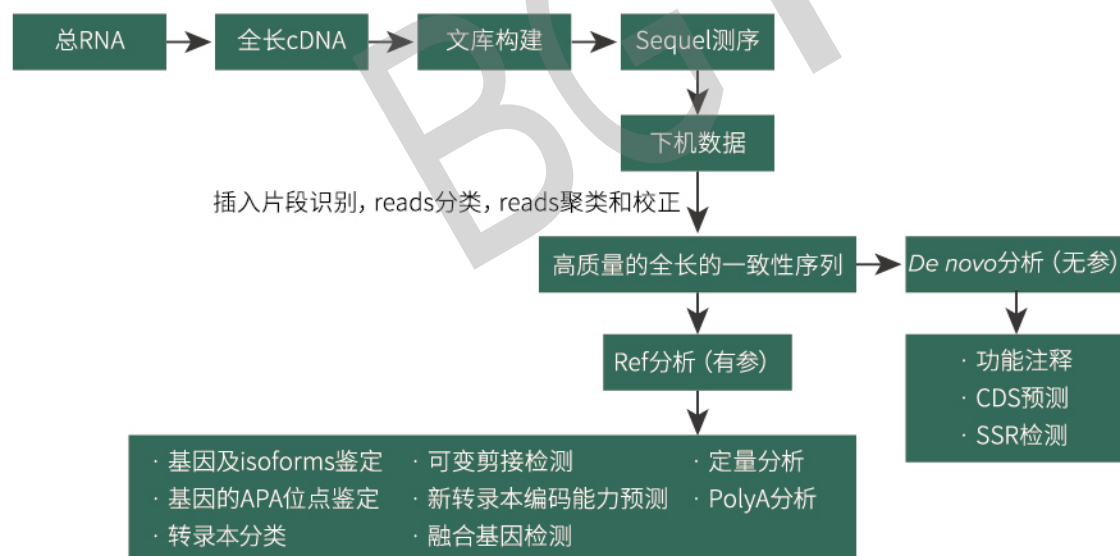


图1 全长转录组转录组测序技术流程

## 信息分析内容

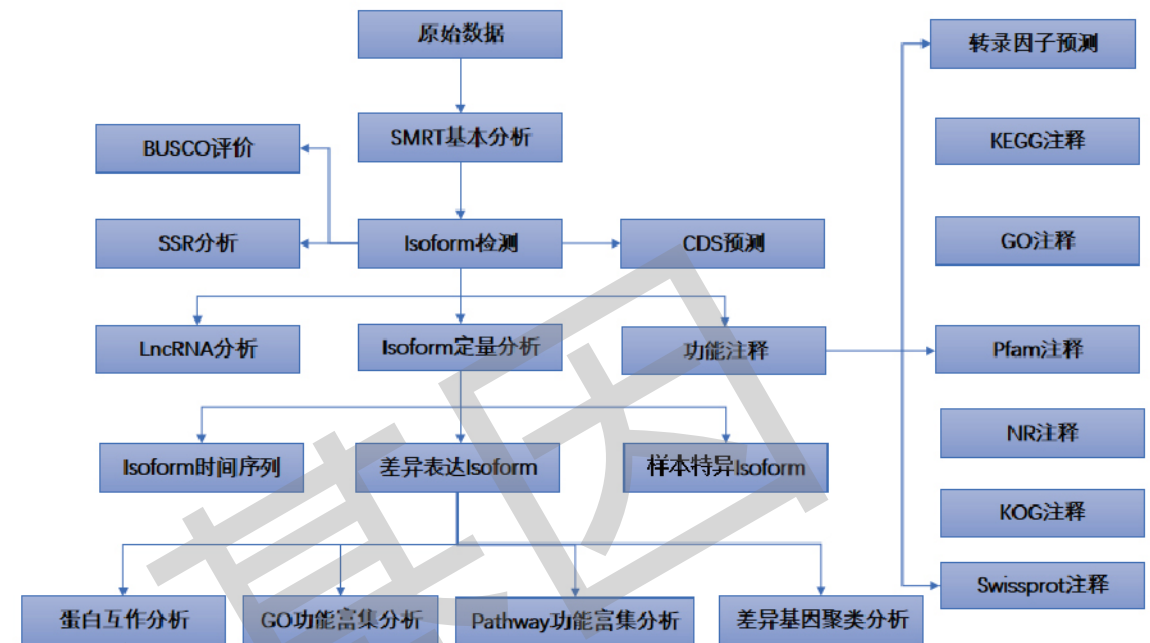


图2 Iso-Seq de novo标准分析条款

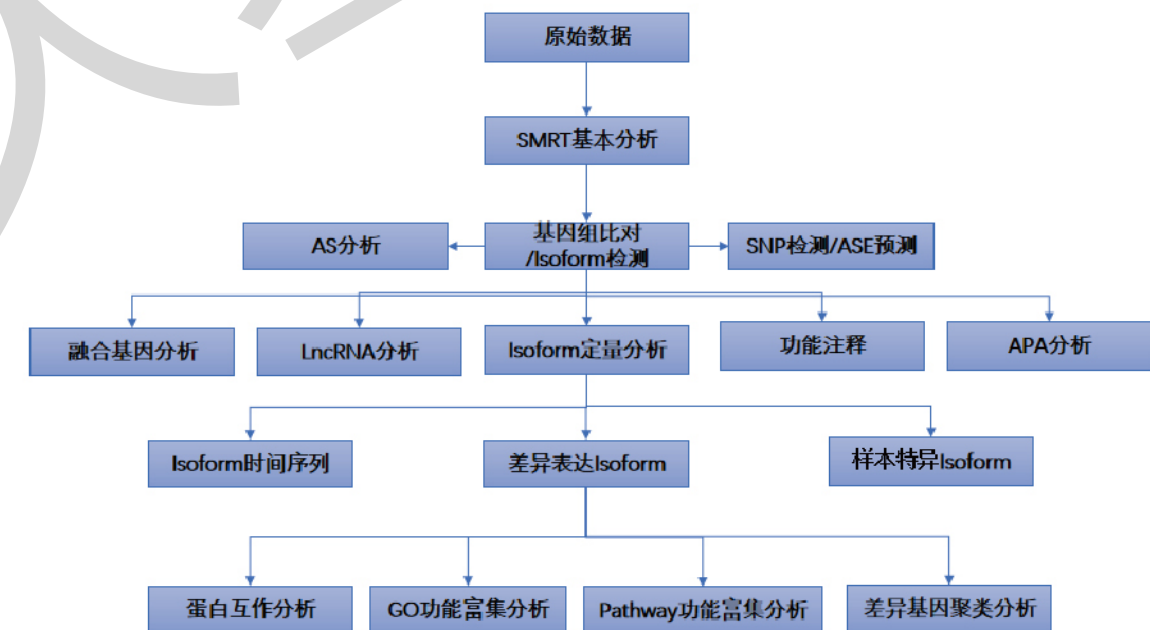


图3 ISO-Seq ref标准分析条款

## 技术参数

- 样品类型 完整且无污染的总RNA
- 样品需求量  $\geq 1\mu\text{g}$
- 样品浓度  $c \geq 285\text{ng}/\mu\text{L}$
- 样品纯度  $\text{RIN} \geq 8, 28\text{S}/18\text{S} \geq 1.4$
- 推荐数据量 根据需求选择5X全长转录组文库、polyA+全长转录组文库或常规全长转录组文库；根据数据量需求Sequel平台一个样本可测序1个cell或多个cell，Sequel II平台1个cell可pooling多个样本。
- 项目执行周期 40个工作日

## ► 案例解析

### 案例一：通过乳腺癌中的单分子长读长RNA测序揭示紫杉醇抗性的新靶点

RNA测序已成为研究癌症中转录组的最常用技术之一，而短读长测序长度较短限制了其在可变剪接 (AS) 和发现新isoform中的应用。本研究选择人乳腺癌细胞MDA-MBA-231野生型 (231-WT) 和紫杉醇抗性型 (231-PTX)，同时采用单分子RNA测序 (Iso-seq) 和短读长RNA测序 (RNA-seq) *de novo* 组装。这两种测序技术均提供了准确的转录本序列和高深度转录本覆盖率。然后结合short-read和long-reading RNA-seq来分析可变剪接事件和新isoform。最终选择BAK1作为我们的候选目标来验证我们的分析。本研究结果表明，应用单分子长读长RNA测序可以获得更深入和更准确的转录水平信息，从而推进癌症基因组功能特征的研究。

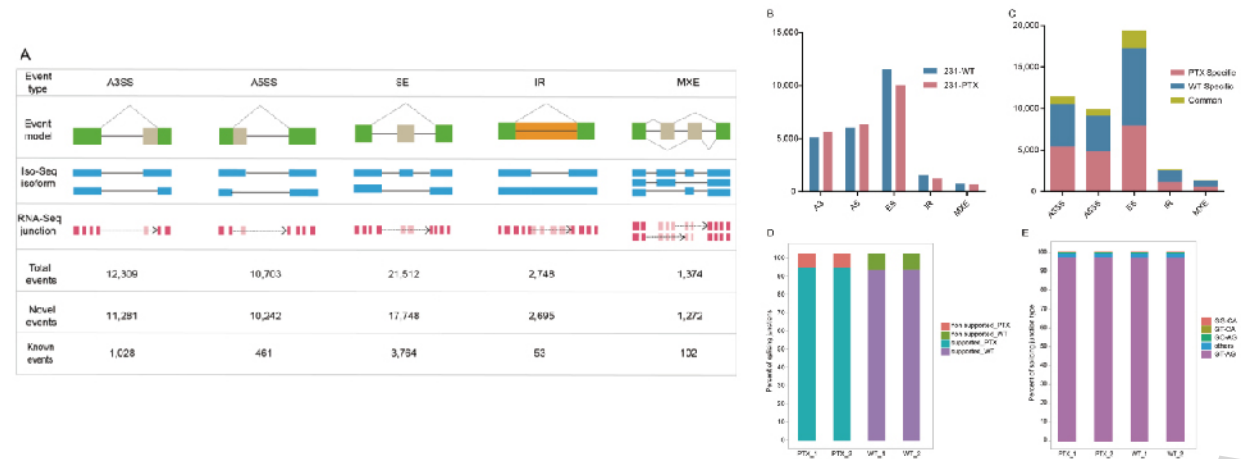


图4 乳腺癌细胞野生型 (231-WT) 和紫杉醇抗性型 (231-PTX) 可变剪接比较

### ★ 参考文献

Lian B, Hu X, Shao Z. Unveiling novel targets of paclitaxel resistance by single molecule long-read RNA sequencing in breast cancer[J]. Scientific reports, 2019, 9(1): 6032.

## ► 常见问题

Q: 为什么我测的数据量增加了好几倍，得到全长转录本数基本没什么变化？

A: Sequel 平台测序数据量主要受读长和reads数两个因素影响。Reads数的多少主要受限于测序芯片上ZMW孔的多少，sequel一张测序芯片上有1M个孔，一般测序loading率在30-70%之间波动（即每次测序会有300K-700K的有效reads），这个数量不会受测序芯片和试剂的升级影响。在常规文库测序中，测序得到的每一条reads代表一条转录本序列（全长比例大约占50%左右），即一般测序一个cell会得到300K-700K的转录本（其中全长转录本有150K-350K条）。

随着sequel测序试剂的升级，测序读长越来越长，在reads数不变的情况下，测序产量也得到了大幅提升，下表是sequel不同试剂版本对应的产量、读长和reads数，从表中可以看出，3.0试剂相比2.0试剂测序产量提升了8倍，平均reads数只提升了不到50%。所以在全长转录组测序中不能只关注测序产量，得到的转录本数量是一个更重要的关注指标。

	Sequel (2.0试剂)	Sequel (2.1试剂)	Sequel (3.0 LR试剂)	提升率 (3.0 LR试剂/2.0试剂)
平均产量	4.5G	10G	40G	8.88
平均读长	13K	23K	81K	6.23
平均Reads数	346k	434k	494K	1.42

5X全长转录组技术是将多条转录本连成一条reads进行建库测序，测得的一条reads代表多条转录本，从而提高了全长转录本的得率。相同的测序数据量，可得到常规文库测序5倍的全长转录本。

Q: 5X全长转录组产品的优势？

A: 相比于常规全长转录组测序，5X全长转录组产品有以下优势：

1. 有效转录本reads数可提升3-5倍
2. 基因/转录本检测数目接近Illumina平台
3. Isoform准确度高
4. 测序长度分布无偏向性
5. UMI和5X全长转录组完美结合，基因定量结果更准确
6. 采用5X全长转录组技术，大部分物种测1个sequel cell数据就可满足研究需求

Q: 用5X全长转录组技术，一个样本需要测多少数据？

A: 采用5X全长转录组技术，大部分物种测1个sequel cell数据就可满足研究需求。

以下是部分已发表物种全长转录组测序数据量，请选择转录本复杂程度相近的物种作参考：

	测序Reads数	转录本数目	推荐测序Cell数 (sequel)	推荐测序Cell数 (5X全长转录组)
海岛棉	2,542,318	176,849	~6cell	2cell
玉米	3,716,604	111,151	~6cell	2cell
拟南芥	844,003	78,010	~4cell	1cell
菠萝	1,068,545	45,876	~3cell	1cell
毛竹	288,312	42,280	~3cell	1cell
盐生草	433,420	54,835	~3cell	1cell
高粱	1,838,330	27,860	~2cell	1cell
苜蓿	448,454	52,787	~2cell	1cell
黄芪	494,408	27,975	~2cell	1cell
鸡	1,849,786	64,277	~3cell	1cell
真菌	~500K	22,956	~2cell	1cell

Q: PolyA+全长转录组可以做哪些分析？

A: 目前通过PolyA+全长转录组可以研究以下内容：

- PolyA长度分布
- PolyA尾中非A碱基类型及出现频率
- 转录本表达量、结构及功能与polyA的关联
- 结合翻译组及RNA半衰期检测，还可以研究：PolyA长度及其他碱基修饰与mRNA稳定性或翻译效率的关系



# RNA-Seq

## 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,能发现低丰度基因。

## 产品优势

- 全面丰富的样品类型**  
人鼠样本常规建库仅需200ng,提供单细胞、FFPE、微量样品的特殊样品服务,微量建库低至pg级;
- 高质量的自主测序平台**  
滚环扩增构建DNB测序文库,PCR扩增错误不会累积,无index hopping之忧,低Dup rate无需人为干预,《Nature》、《Cell》、《Immunity》等1000+累积影响因子;
- Dr. Tom系统解决个性化难题**  
只需任意一种RNA测序数据,就获得多组学关联信息,多数据库联合分析,多维度数据展示,循环挖掘数据,轻松锁定解释生物学问题的核心基因;
- 极速交付**  
从提取和检测开始,到Dr.Tom个性化分析,每个环节都极速。

## 产品应用



## 技术流程

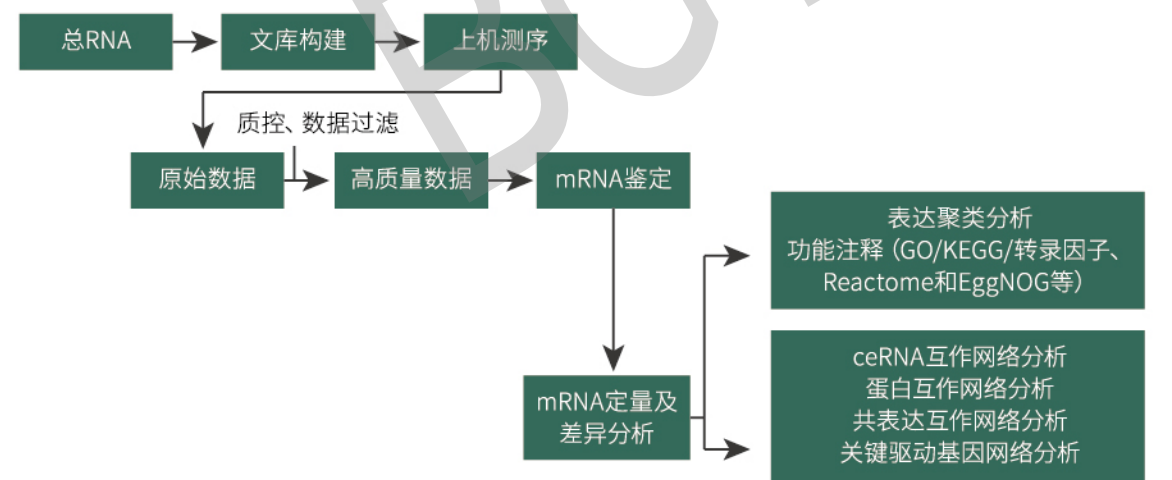


图1 RNA-Seq技术流程

## 信息分析内容

- 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reads信息,② 样品相关性,③ 表达量分布,④ RNA分类;
  - 2、参考基因组比对;
  - 3、mRNA鉴定;
  - 4、mRNA定量分析;
  - 5、mRNA差异表达分析(样本间、组间);
  - 6、mRNA表达/差异基因聚类;
  - 7、mRNA差异基因GO分类、富集;
  - 8、mRNA差异基因KEGG分类、富集

## Dr.Tom信息分析

### 数据库注释

- 1、转录因子注释(AnimalTFDB/PlantTFDB);
- 2、GSEA分析(仅限人);
- 3、Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释;

### 互作网络分析

- 1、靶基因分析① miRNA-mRNA靶向关系分析,② lncRNA-mRNA靶向关系分析;
- 2、ceRNA互作网络分析① lncRNA-mRNA联合分析,② circRNA-mRNA联合分析(仅限人、小鼠);
- 3、蛋白互作网络分析;
- 4、共表达互作网络分析;

### 特色分析

- 1、自定义标签和自有数据上传;
- 2、外部数据库信息(TCGA、ARCHS4);
- 3、卡方检验;
- 4、关键驱动基因网络图分析;
- 5、时间序列分析。

## 技术参数

- 样本要求**  
样品类型:完整且无污染的总RNA,样品需求量(单次):200 ng  
样品浓度:10-1,000 ng/ $\mu$ L  
样品纯度:28S/18S $\geq$ 1.0; RIN $\geq$ 7.0
- 测序策略** SE50
- 推荐数据量** 20M clean reads
- 项目执行周期**  
标准执行周期为15个工作日,极致交付周期为11个工作日(含样品检测)。



## ► 案例分析

### 案例一：RNA-Seq发现天然免疫及炎症调控新靶标

树突状细胞 (DC) 是一类重要的天然免疫细胞, DC迁移紊乱可能导致DC在炎症部位的过度聚集及活化, 导致组织过度炎症, 甚至引发炎症性疾病。此前对于DC迁移的表观调控机制尚缺乏了解, 特别是长链非编码RNA在DC迁移及炎症性疾病中的作用未见报道。研究团队通过华大提供的lncRNA测序和RNA-Seq等, 发现天然免疫及炎症调控新靶标——长链非编码RNA lnc-Dpf3, 通过抑制树突状细胞糖酵解过程, 从而抑制树突状细胞体内迁移及炎症性疾病的发生发展。该研究有助于解释与树突状细胞过度迁移相关的炎症性疾病的发生机制, 为设计和探索疾病新型治疗方法提供了新的思路与依据。

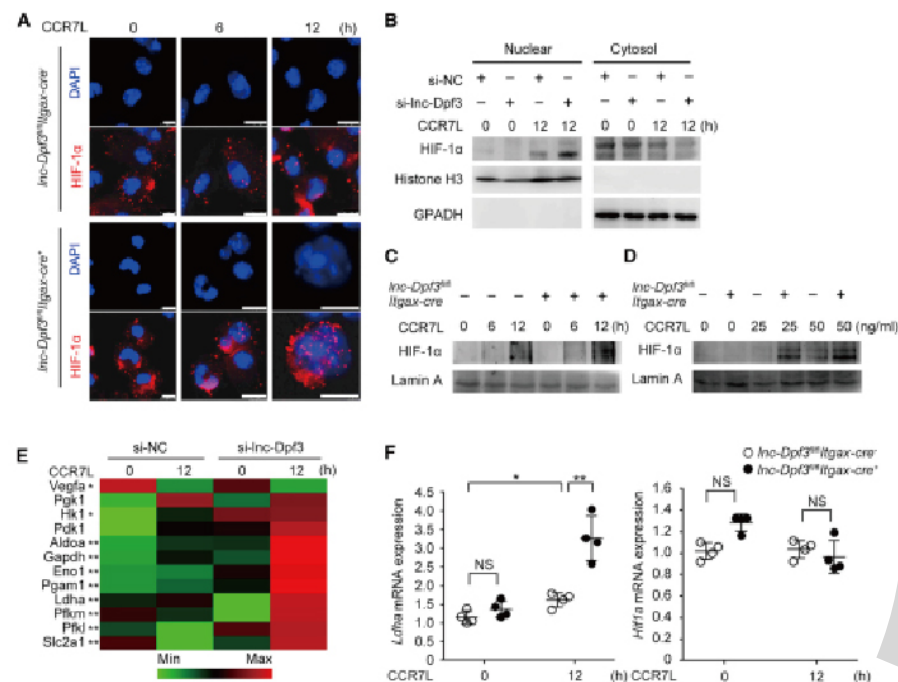


图2 lnc-Dpf3通过抑制HIF-1α诱导的Ldha转录, 抑制DC迁移

### ★ 参考文献

Liu J, Zhang X, Chen K, et al. CCR7 Chemokine Receptor-Inducible lnc-Dpf3 Restrains Dendritic Cell Migration by Inhibiting HIF-1α-Mediated Glycolysis[J]. Immunity, 2019, 50(3): 600-615. e15.

## ► 常见问题

### Q: RNA-Seq必须要有参考基因组序列吗?

A: 不是必须要有参考基因组序列, 但一定要有参考序列, 如unigene序列、mRNA序列、CDS序列等可以作为参考序列。

### Q: 若无参考序列, 能够进行哪些生物信息分析?

A: 在无参考序列情况下, 可通过参考有基因组序列的近缘物种(需预先指定)的基因注释信息来完成相应的标准信息分析。

### 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。

### Q: 测序后有何验证方法?

A: 实验验证的方法最常见的是通过实时荧光定量 PCR (qRT-PCR) 技术来验证测序结果。还有FISH (原位荧光杂交)、微阵列芯片技术、Northern blot等。功能验证一般是基因敲除、敲低或过表达, 转基因等。

### Q: BGISEQ RNA-Seq有没有发布demo数据?

A: 已发布一个UHRR人标准品下机数据并共享了数据结果和分析方法。



←扫描二维码, 即可获取数据

## ► 华大合作发表文章(部分)





物种	发表时间	发表期刊	文献标题
人	2017.01	Cell	Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity
人银屑病	2018.04	Immunity	An Interleukin-25-Mediated Autoregulatory Circuit in Keratinocytes Plays a Pivotal Role in Psoriatic Skin Inflammation
人	2019.02	Immunity	CCR7 Chemokine Receptor-Inducible lnc-Dpf3 Restrains Dendritic Cell Migration by Inhibiting HIF-1α-Mediated Glycolysis
小鼠	2019.02	Hepatology	Hepatic Interferon Regulatory Factor 6 Alleviates Liver Steatosis and Metabolic Disorder by Transcriptionally Suppressing Peroxisome Proliferator-Activated Receptor $\gamma$ in Mice
人	2019.02	Immunity	Specific Decrease in B-Cell-Derived Extracellular Vesicles Enhances Post-Chemotherapeutic CD8 <sup>+</sup> T Cell Responses
人	2019.03	Neuron	Nucleoporin Seh1 Interacts with Olig2/Brd7 to Promote Oligodendrocyte Differentiation and Myelination
小鼠	2019.04	Cell Research	Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate
人细胞系	2019.04	Genome Biology	Translation of the circular RNA circ $\beta$ -catenin promotes liver cancer cell growth through activation of the Wnt pathway
人细胞系	2019.04	Cell Host & Microbe	Nutrient Sensing by the Intestinal Epithelium Orchestrates Mucosal Antimicrobial Defense via Translational Control of Hes1
小鼠	2019.05	Hepatology	Integrated Omics Reveals Tollip as an Aggravator and Therapeutic Target for Hepatic Ischemia - Reperfusion Injury in Mice
人	2019.05	Journal of Clinical Investigation	Breast milk alkylglycerols sustain beige adipocytes through adipose tissue macrophages

# Small RNA测序

## 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UMI建库技术、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，鉴定小RNA，并预测其靶标基因。

## 产品优势

-  全面的样品类型  
真核、原核生物、血清血浆、FFPE样品、外泌体、RIP富集RNA等多种类型；
-  更低的样品总量  
小RNA样本实现最低1 ng起始量，低起始量建库成功率达95%；
-  更准确的测序结果  
建库引入UMI技术，实现无偏向的精准定量；
-  更丰富的分析内容  
采用Dr.Tom多组学数据挖掘系统，10大数据库注释，多维度结果图片展示，数据图表循环挖掘。

## 产品应用



## 技术流程

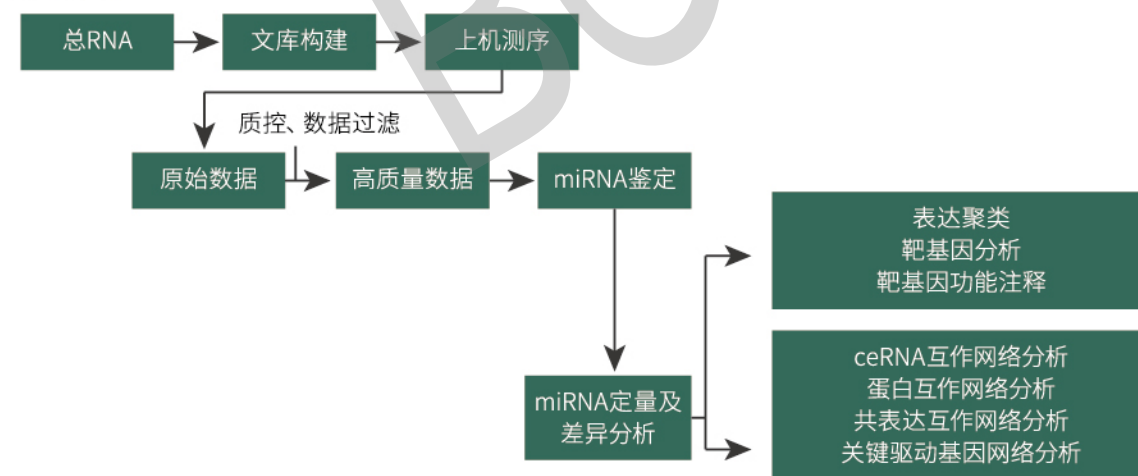


图1 Small RNA技术流程

## 信息分析内容

- 标准信息分析
  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. small RNA 信息采集结果的长度分布
  3. small RNA在选定参考基因组上的分布
  4. small RNA分类统计
  5. miRNA定量分析
  6. miRNA差异表达分析
  7. miRNA表达/差异表达聚类分析
  8. miRNA靶基因分析
  9. 差异miRNA 靶基因GO 注释和KEGG 通路分析

## 高级信息分析

- 数据库注释
  1. 转录因子注释 (AnimalTFDB/PlantTFDB)
  2. GSEA分析
  3. Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释
- 互作网络分析
  1. 靶基因分析 ① miRNA-mRNA靶向关系分析 ② lncRNA-mRNA靶向关系分析
  2. ceRNA互作网络分析
  3. 蛋白互作网络分析
  4. 共表达互作网络分析
- 特色分析
  1. 外部数据库关联分析 (TCGA、ARCHS4)
  2. 关键驱动基因网络图分析
  3. 时间序列分析

## 技术参数

- 样本要求  
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、外泌体、RIP样品等
- 样品最低需求量 (单次)  $\geq 1\text{ng}$
- 测序策略 SE50
- 推荐数据量 20M clean reads
- 项目执行周期 25个工作日 (样品数小于10个)

## 案例分析

### 案例一：Small RNA测序鉴定晚期胃癌外泌体miRNA特征谱

腹膜扩散 (PD) 是晚期胃癌 (GC) 患者中预后不良的最常见转移。本研究分析来自GC患者的恶性腹水 (MA) 来源的外泌体对肿瘤细胞的影响，并阐明其潜在机制。体外、体内分析显示，与外泌体耗尽的上清液相比，来自GC患者的MA的外泌体通过上调上皮-间质转化 (EMT) 信号传导促进AGS细胞的侵袭。在小鼠腹部异种移植模型中，MA衍生的外泌体处理后的中值存活期比对照组短 (35.5天对67天， $p = 0.0005$ )。此外，在腹膜化疗前后8名配对GC患者和3名非恶性疾病患者中，通过高通量测序鉴定了29个来自腹水的外泌体miRNA。总之，来自GC患者的MA衍生的外泌体促进GC细胞和小鼠腹膜肿瘤模型中的EMT信号传导。差异外泌体miRNA可能在治疗上用于抑制腹膜转移，这为GC中PD的分子机制提供了新的见解。



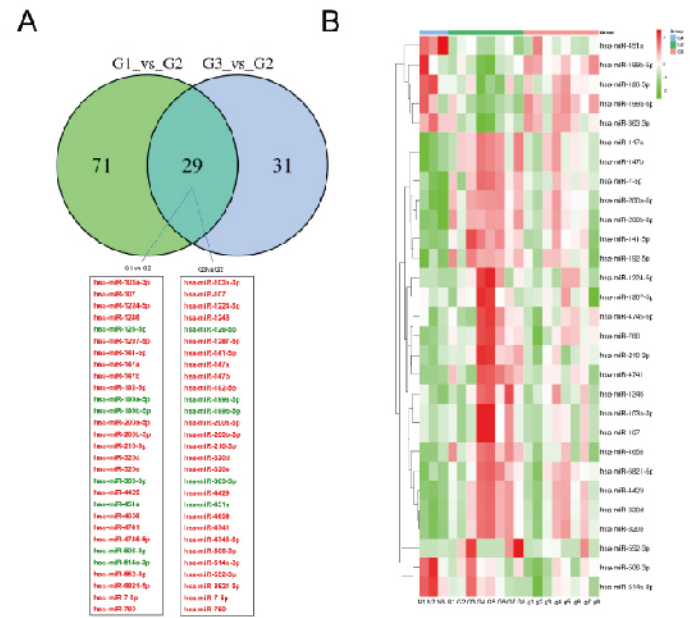


图2 腹膜化疗前后的GC患者和对照的外泌体miRNA差异

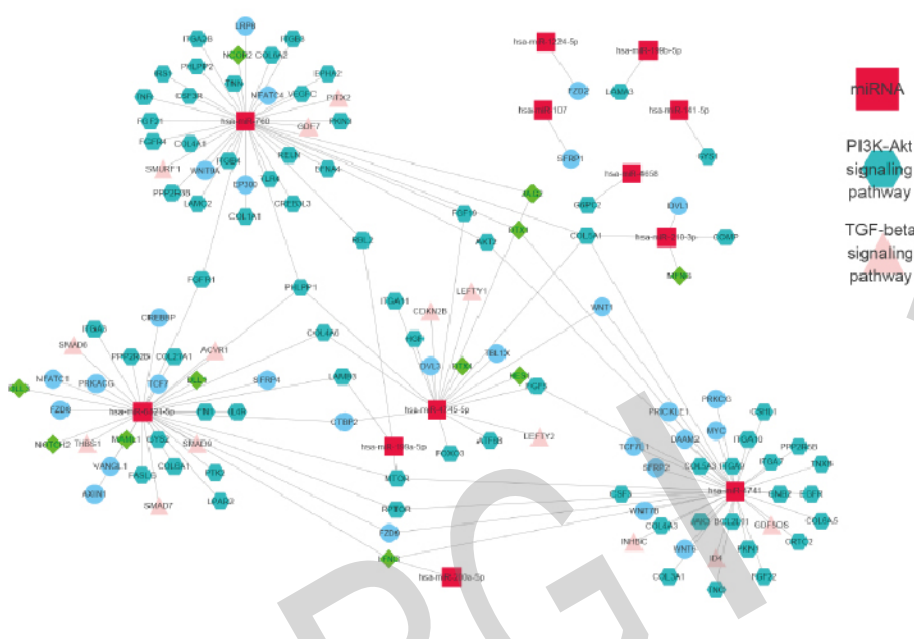


图3 腹膜扩散的GC患者中竞争RNA网络

★ 参考文献

Hu Yanting, Qi Changsong, Liu Xiang et al. Malignant ascites-derived exosomes promote peritoneal tumor cell dissemination and reveal a distinct miRNA signature in advanced gastric cancer. [J]. Cancer Lett., 2019, 457: 142-150.

► 常见问题

Q: 华大Small RNA可以操作什么类型的样品建库测序?

A: 动植物总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、外泌体、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时不建议使用过柱试剂盒, 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是本物种的EST等序列作为参考序列, 另外还需要老师提供相关的exon/intro, repeat信息以及基因编码序列。

► 华大合作发表文章(部分)

物种	发表时间	发表期刊	文献标题
血清	2008. 11	Cell Research	Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases
肝癌	2011. 02	Cancer Cell	Identification of miRNomes in Human Liver and Hepatocellular Carcinoma Reveals miR-199a/b-3p as Therapeutic Target for Hepatocellular Carcinoma
小鼠	2014. 08	Journal of Autoimmunity	DNA methylation and mRNA and microRNA expression of SLE CD4+ T cells correlate with disease phenotype
鞭毛虫	2014. 09	PNAS	Both endo-siRNAs and tRNA-derived small RNAs are involved in the differentiation of primitive eukaryote Giardia lamblia
人	2015. 12	Nucleic Acids Research	A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes
小鼠	2017. 02	Cell Death and Disease	MicroRNA-144 is regulated by CP2 and decreases COX-2 expression and PGE2 production in mouse ovarian granulosa cells
心脏	2017. 09	European journal of heart failure	Serially measured circulating microRNAs and adverse clinical outcomes in patients with acute heart failure
骨髓	2017. 11	Molecular Therapy	Global microRNA profiling in human bone marrow skeletal (stromal or mesenchymal) stem cells identified candidates for bone regeneration
人	2018. 08	Molecular Cancer	The chromosome 11q13.3 amplification associated lymph node metastasis is driven by miR-548k through modulating tumor micro-environment
人	2018. 09	Analytical Chemistry	Next Generation Sequencing Analysis of Total Small Noncoding RNAs from Low Input RNA from Dried Blood Sampling.



# 长链非编码RNA测序

## 产品概述

长链非编码 RNAs (long non-coding RNAs, lncRNAs) 是一类长度大于 200 nt 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 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;



**性价比高**  
一次测序可以得到 mRNA、lncRNA、circRNA 信息;



**经验丰富**  
已完成万例 lncRNA 项目, 涉及人鼠及动植物物种;

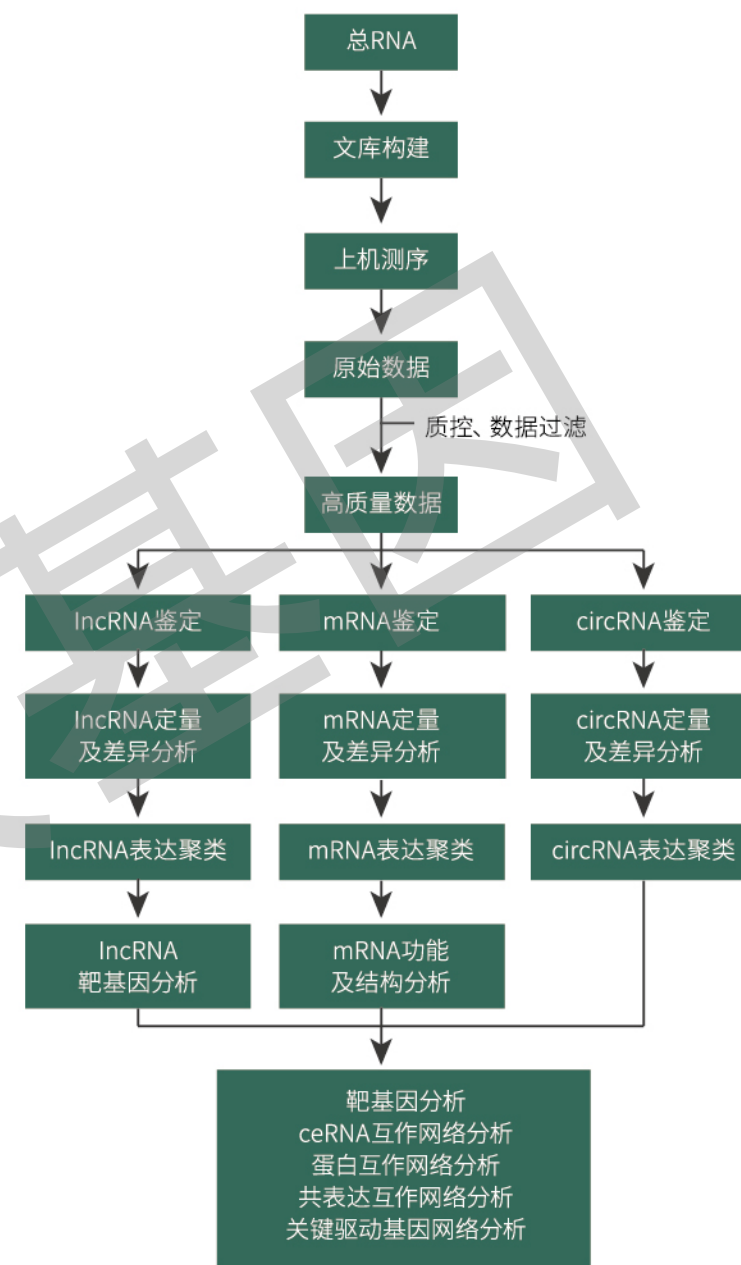


**分析交互**  
采用 Dr.Tom 多组学数据挖掘系统, 10 大数据库注释, 多维度结果图片展示, 数据图表循环挖掘。

## 产品应用



## 技术流程



## 信息分析内容

- 标准信息分析 (circRNA 仅限人、小鼠)
  1. 基本数据统计
    - ① 去除接头序列、低质量序列得到 reads 信息
    - ② 样品相关性
    - ③ 表达量分布
    - ④ RNA 分类
  2. 参考基因组比对
  3. lncRNA、mRNA、circRNA 鉴定
  4. lncRNA、mRNA、circRNA 定量分析
  5. lncRNA、mRNA、circRNA 差异表达分析 (样本间、组间)
  6. lncRNA、mRNA、circRNA 表达/差异基因聚类



7. mRNA差异基因GO分类、富集
8. mRNA差异基因KEGG分类、富集
9. mRNA结构分析
  - ① 可变剪切分析
  - ② 融合基因分析 (仅限人)

## ► 高级信息分析

- 数据库注释
  1. 转录因子注释 (AnimalTFDB/PlantTFDB)
  2. GSEA分析
  3. Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释
- 互作网络分析
  1. 靶基因分析
    - ① miRNA-mRNA靶向关系分析
    - ② lncRNA-mRNA靶向关系分析
  2. ceRNA互作网络分析
  3. 蛋白互作网络分析
  4. 共表达互作网络分析
- 特色分析
  1. 外部数据库关联分析 (TCGA、ARCHS4)
  2. 关键驱动基因网络图分析
  3. 时间序列分析

## ► 技术参数

- 样本类型 总RNA样品; 组织培养与细胞系样品、外泌体、FFPE样品等
- 样品最低需求量 (单次)  $\geq 200\text{ng}$
- 测序策略 PE100
- 推荐数据量 10G clean data
- 项目执行周期 35个工作日 (小于10个样本)

## ► 案例分析

### 案例一: lncRNA在前列腺癌进展中的作用

本研究利用雄激素受体(AR)配体二氢睾酮(DHT)刺激前列腺癌细胞系,通过RNA-Seq结果结合先前发表的ChIP-Seq数据,再对比分析TCGA中正常前列腺组织、临床前列腺癌的RNA-Seq数据,发现ARLNC1(雄激素受体调节的长链非编码RNA1)与前列腺癌进展中雄激素受体信号传导密切相关。ARLNC1不仅由雄激素受体蛋白诱导,而且可通过RNA-RNA相互作用稳定雄激素受体转录物。体外体内功能验证发现,敲低ARLNC1会抑制雄激素受体表达、信号传导和前列腺癌生长。因此,ARLNC1不仅具有成为生物标志物的潜力,更有望成为新的癌症治疗靶标。

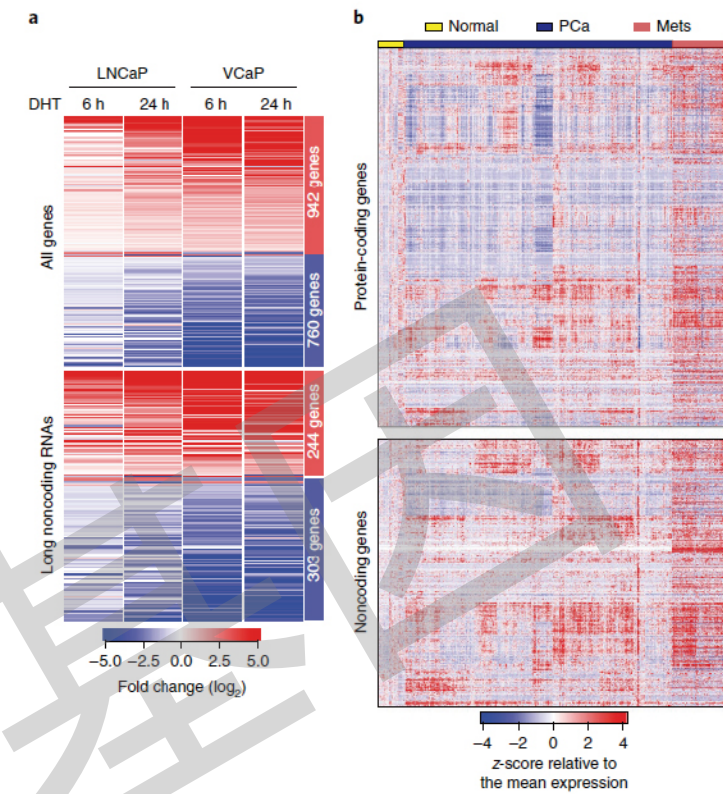


图1 前列腺癌中雄激素调节基因的鉴定热图

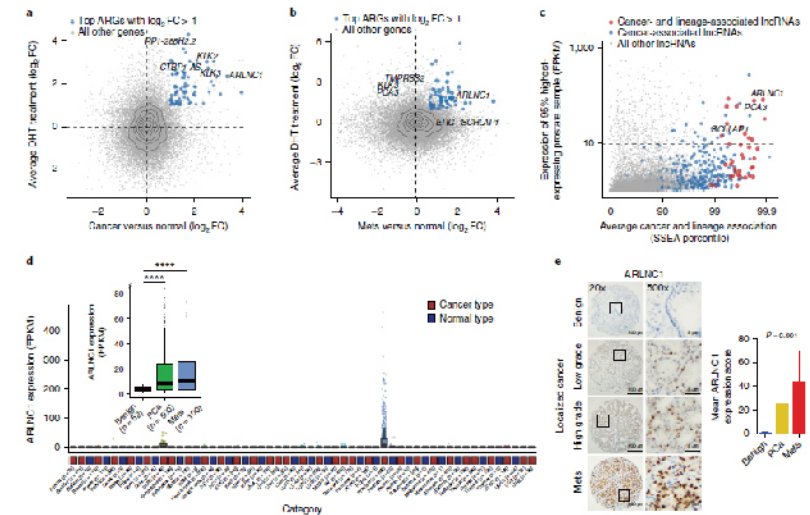


图2 前列腺癌中ARLNC1的鉴定和原位表征

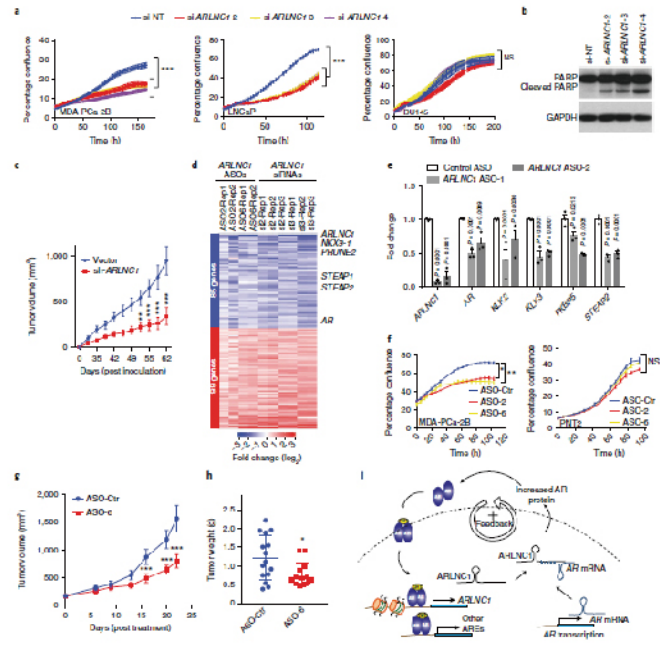


图3 ARLNC1敲低的体外体内验证

★ 参考文献

Zhang Y, Pitchiaya S, Cieřlik M, et al. Analysis of the androgen receptor-regulated lncRNA landscape identifies a role for ARLNC1 in prostate cancer progression[J]. Nature genetics, 2018, 50(6): 814-824.

► 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: 长链编码RNA的建库方案是什么?

A: 长链编码RNA建库主要采用去除rRNA, 构建链特异性文库, 建库稳定。

Q: LncRNA测序的数据中是否也含有 mRNA、circRNA?

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、circRNA 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。

► 华大合作发表文章(部分)

物种	发表时间	发表期刊	文献标题
人-前列腺癌	2012.01	Cell Research	RNA-Seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions, cancer-associated long noncoding RNAs and aberrant alternative splicing
人-神经母细胞瘤	2014.11	Cancer Cell	The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation
人-生殖细胞	2015.04	Database (Oxford)	GermlncRNA: a unique catalogue of long non-coding RNAs and associated regulations in male germ cell development
癌症	2016.03	Cell Death and Disease	MiR-125a-5p decreases after long non-coding RNA HOTAIR knock-down to promote cancer cell apoptosis by releasing caspase 2
小鼠精原干细胞	2016.03	Cell Death and Disease	A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells
人-各组织表达谱	2016.06	Scientific Reports	Identification of Tissue-Specific Protein-Coding and Noncoding Transcripts across 14 Human Tissues Using RNA-seq
人-结肠癌	2016.06	Scientific Reports	Transcriptome and long noncoding RNA sequencing of three extracellular vesicle subtypes released from the human colon cancer LIM1863 cell line
急性髓细胞样白血病	2017.07	Haematologica	Long noncoding RNA expression profile in cytogenetically normal acute myeloid leukemia identifies a distinct signature and a new biomarker in NPM1-mutated patients
胚胎着床失败	2018.01	Reprod Sci	Genome-Wide Profiling of Long Noncoding RNA Expression Patterns in Women With Repeated Implantation Failure by RNA Sequencing
乳腺癌	2018.12	Scientific Reports	lncRNA profile study reveals the mRNAs and lncRNAs associated with docetaxel resistance in breast cancer cells



产品介绍

Product Profile

## 表观组学

- 01 | 全基因组甲基化测序
- 02 | 目标区域捕获甲基化测序
- 03 | ChIP-Seq

# 全基因组甲基化测序

## 产品概述

DNA甲基化是重要的表观遗传学标记信息，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



**检测精度高**  
单碱基分辨率，精确分析每一个C碱基的甲基化状态；



**性价比高**  
相对于传统的PCR+sanger测序法费用少；

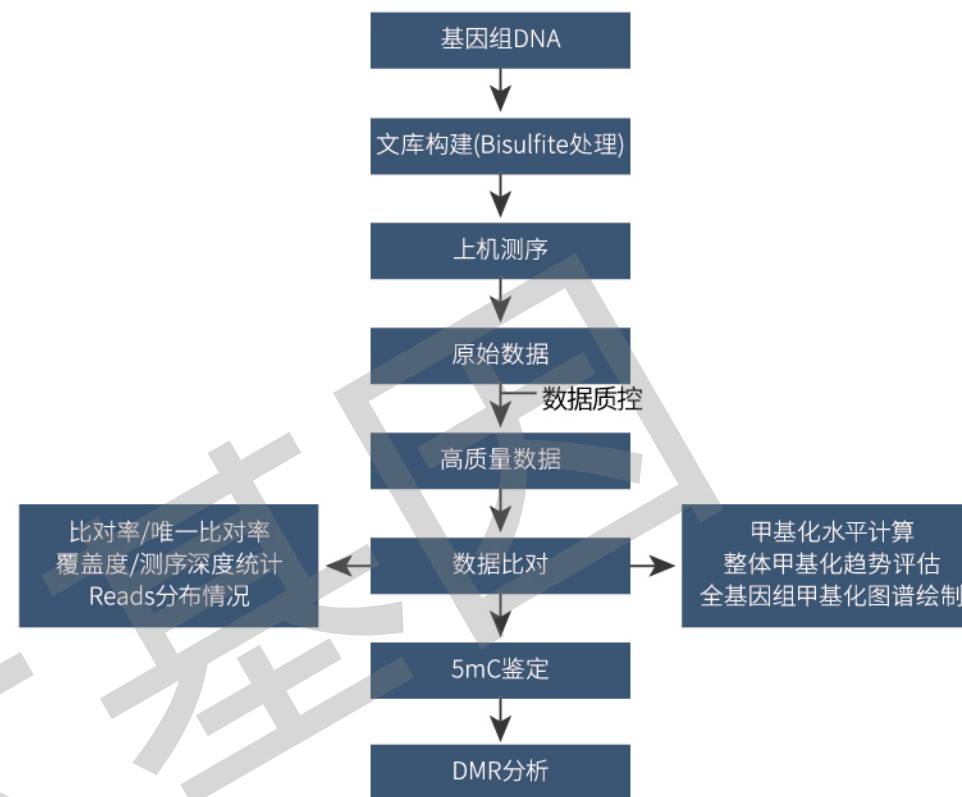


**效率高、耗时少**  
借助新一代高通量测序平台对全基因组5mC信息进行高效扫描双链环化建库方式，GC偏向性更低，全基因组覆盖度更平均，甲基化率更真实。

## 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. 基因组不同转录元件中的DNA平均甲基化水平
  8. 全基因组差异甲基化区域DMR的检测
  9. DMR相关基因的GO和Pathway分析
  10. 其他定制化信息分析（可结合客户的需求，协商确定定制化信息分析内容）

## 技术参数

- 样本要求
  - 样品总量:  $\geq 1 \mu\text{g}$
  - 样品纯度:  $\text{OD}_{260}/\text{OD}_{280} = 1.8\sim 2.0$ ,  $\text{A}_{260}/\text{A}_{230} \geq 1.6$ ; 没有蛋白、多糖和RNA污染
  - 样品完整性: DNA 无明显降解, 需提供凝胶电泳检测胶图
  - 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
- 推荐数据量  $\geq 30\times$
- 项目执行周期 40个工作日



## ► 案例分析

### 案例一：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，并在更大样本量中进行了验证，最终发现与同样的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。

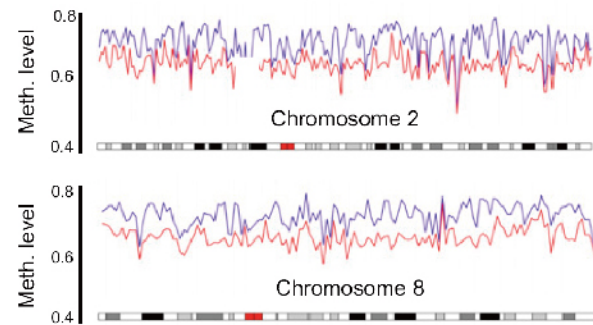


图1 百岁老人(红色)与新生儿(蓝色)的甲基化分布

### ★ 参考文献

Heyn H, Li N, Ferreira H J, et al. Distinct DNA methylomes of newborns and centenarians[J]. Proceedings of the National Academy of Sciences, 2012, 109(26): 10522-10527.

## ► 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: 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，基因组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: Bisulfite的转化率是多少?

A: Bisulfite转化率达到99%以上; 如果样品的DNA不存在不发生甲基化的DNA作为对照，都会在样品中混入control DNA来验证Bisulfite的转化率。

## ► 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
方法学	2010.09	Plos Biology	The DNA Methylome of Human Peripheral Blood Mononuclear Cells
年龄	2012.06	PNAS	Distinct DNA methylomes of newborns and centenarians

# ChIP-Seq

## ► 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的蛋白进行免疫沉淀，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的蛋白结合的DNA位点，并可基于多个样品进行差异比较。

## ► 产品优势



数据利用率高  
SE50最优读长，没有数据浪费；



性价比高  
基于抗体富集目标区域，有效降低测序数据量；



检测范围广  
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平台最低起始量仅需5ng，5ng以下亦可做定制化。

## ► 产品应用

### 癌症

肝癌

乳腺癌

膀胱癌

淋巴瘤等

### 复杂疾病

自身免疫性疾病

心血管疾病

精神类疾病

基因印记病等

### 其他

细胞分化

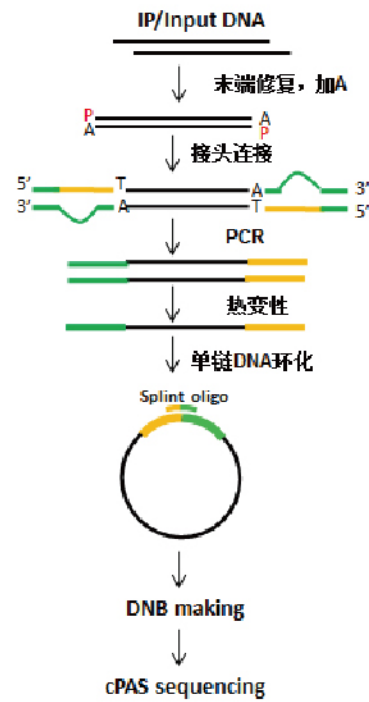
胚胎发育

年龄、长寿

甲基化图谱等

## ► 技术流程

BGISEQ平台技术流程



## ► 技术参数

### • 样本要求

样品类型: ChIPed 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:  $\text{OD}_{260}/\text{OD}_{280} = 1.8\sim 2.0$ ;

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;

• 项目执行周期 标准流程的执行周期为30个工作日。

## ► 案例分析

### 案例一: 索拉菲尼抑制上皮-间充质转化的表观机理

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癌症治疗的药物, 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。

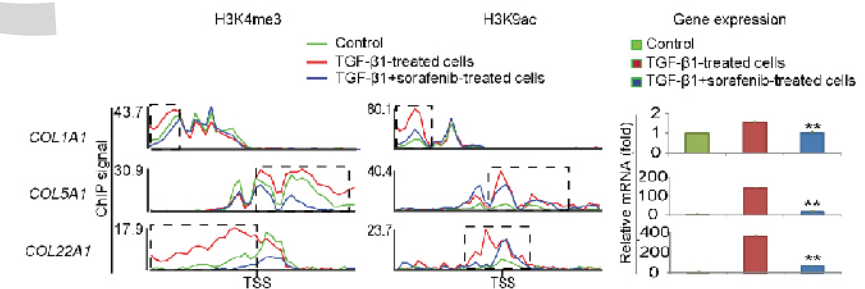


图1 索拉菲尼能够通过影响组蛋白修饰从而调节上皮-间充质转化相关基因的表达

### ★ 参考文献

Zhang J, Chen Y L, Ji G, et al. Sorafenib inhibits epithelial-mesenchymal transition through an epigenetic-based mechanism in human lung epithelial cells[J]. PLOS One, 2013, 8(5): e64954.

## ► 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进行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, 则可减少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。

## ► 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扫描
4. Peak长度分布
5. Peak深度分布
6. Peak在基因功能元件上的分布
7. Peak相关基因
8. Peak相关基因的GO功能显著性富集分析
9. Peak相关基因的pathway富集分析
10. 样品间差异peak
11. 样品间差异peak在基因功能元件的分布
12. 样品间差异peak相关基因
13. 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



Q: 影响ChIP-Seq 结果的因素有哪些?

A: 抗体的质量与特异性、需要富集的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。

► 华大合作发表文章(部分)







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

# ATAC-Seq

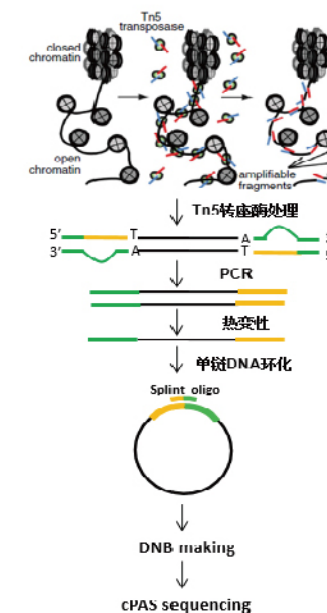
► 产品概述

ATAC-Seq利用转座酶可以攻击开放的染色质并催化转座子整合到基因组的特性,使用改造后的活性极高的Tn5转座酶,并将转座子DNA设计为测序接头,这样ATAC-Seq 可以通过“剪切”和“粘贴”的方式,直接将测序接头引入开放染色质中,进一步实现对开放染色质进行捕获和测序。最终,通过ATAC-Seq 我们将获得全基因组开放性调控元件图谱。ATAC-Seq作为一种高性价比、获取信息全面、起始量低的研究方法,在大规模临床样本的研究中具有广泛的应用前景。

► 产品优势

-  **化繁为简**  
实验流程精简,样本适应范围更广;
-  **周期更短**  
相对传统技术交付更快;
-  **准确性高**  
与同类技术(如Dnase-Seq)一致性高;
-  **起始量低**  
同类技术所需细胞材料最少、适用于临床样本(100个-5万细胞);
-  **性价比高**  
依托BGISEQ平台测序成本优势,更具竞争力;
-  **更多信息**  
获得信息量最大的表观基因组测序技术(全基因组调控活性图谱、全转录因子结合图谱)。

► 技术流程



BGI

华大基因

## ► 信息分析内容

1. 数据基本处理与质控
2. 文库插入片段长度分布
3. 基因组测序深度累积分布
4. 各样品基因promoter (up2K) 区测序深度分布
5. Peak检测 (Peak长度分布、Peak深度分布)
6. 样本生物学重复 IDR 分析
7. Peak在基因功能元件上的分布
8. Peak相关基因
9. Peak相关基因的GO功能显著性富集分析
10. Peak相关基因的pathway富集分析
11. 样本间差异peak检测
12. 样品间差异peak在基因功能元件的分布
13. 样品间差异peak相关基因
14. 样品间差异peak相关基因的GO和KEGG富集分析
15. Motif分析

## ► 技术参数

### • 样本要求 细胞样品送样建议

### • 要求 收集活性高(大于70%)

### • 细胞活性检测方法参考

细胞使用台盼蓝进行染色, 4%台盼蓝母液使用时用PBS稀释至0.4%, 并用0.2 $\mu$ m的滤膜进行过滤处理。细胞悬液与0.4%台盼蓝溶液以9:1混合混匀。(终浓度0.04%), 在三分钟内, 分别计数活细胞和死细胞。细胞活性在70%以上较为理想。若细胞活性率偏低, 建议进行活细胞筛选。

### • 细胞数量

生长状态良好的细胞, 加入适量配制好的冻存培养液, 用吸管轻轻吹打使细胞均匀, 计数, 调节冻存液中细胞的最终密度为 $1 \times 10^6$ /ml~ $1 \times 10^7$ /ml, 送样体积在1ml以上, 转移至冻存管中, 把冻存管放到程序降温盒内, 再把程序降温盒放到-80 $^{\circ}$ C的冰箱里, 盒内温度会成线性下降。保护细胞不被损伤。

### • 动物组织样品送样建议

推荐客户制备成细胞悬液送样, 若只能提供组织样本, 取新鲜组织浸在1XPBS (pH7.4) 或HBSS缓冲液中, 移至超净台, 清洗一下组织表面, 去除表面的血迹; 将组织置于60mm培养皿中(或1.5ml离心管中), 加入适量培养基(DMEM), 用无菌手术剪去除掉脂肪或结缔组织, 剪碎组织(2mm<sup>3</sup>), 加入冻存液并转移至冻存管, 体积在1ml以上, 组织重量大于100mg, 放入程序降温盒进行-80 $^{\circ}$ C冷冻保存。

## ► 案例分析

### 案例一：哺乳动物胚胎细胞染色质结构开放图谱研究

由于技术限制, 很难对少量细胞进行染色质状态研究, ATAC-Seq只需要少量细胞即可对全基因组调控活性序列进行分析, 文章用ATAC-Seq检测了早期胚胎发育过程中各时期染色质结构开放区域的动态变化, 发现不同时期的细胞开放区域的动态变化情况, 并得到胚胎发育时期关键转录因子。

样本选取: 小鼠早期胚胎各时期, 2个重复

## • 研究成果

得到早期胚胎各时期细胞染色质开放序列信息。

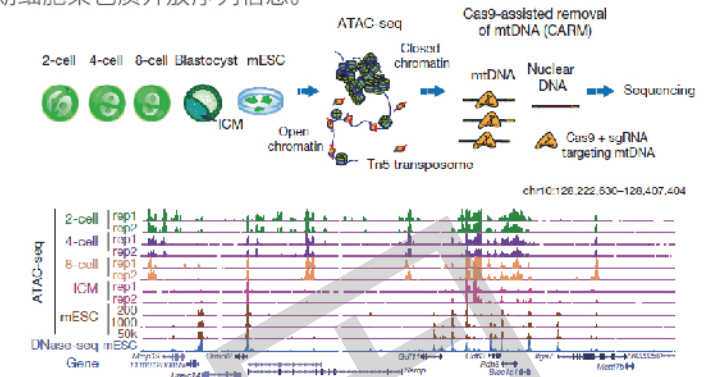


图1 小鼠着床前胚胎细胞染色质开放性可视图

找到各时期调控元件分析及转录因子, 结合RNA-Seq数据找到关键调控因子。

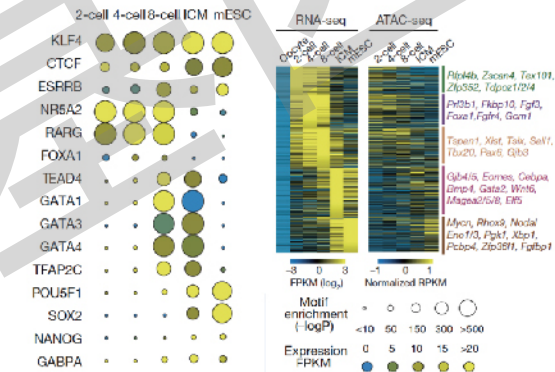


图2 各时期关键调控元件及转录因子预测结果

## ★ 参考文献

Wu J, Huang B, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. Nature. 2016 Jun.

## ► 常见问题

### Q: 客户送样前有哪些需要注意?

A: 送样前最好检测下细胞活性, 尽量送活性(大于70%)较高的新鲜细胞, 组织样本需先处理成细胞悬液。

### Q: ATAC-Seq没有参考基因组能否做?

A: ATAC-Seq需要有完整的参考基因组或者参考序列用作比对, 基因组组装结果的好坏直接影响信息分析结果, 如果没有参考基因组, 可以选择近源物种或者转录组拼接的结果来作为参考基因组, 但准确性不做保证。

### Q: ATAC-Seq如果样本细胞数量达不到送样要求, 是不是一定不能做?

A: 不是, ATAC-Seq建库一般细胞起始数量范围为: 500-50,000个细胞, 具体起始量与细胞活性及细胞类型相关, 甚至可做到单细胞水平, 所以即使客户样本的细胞数量达不到送样要求, 并不代表一定不能建库, 只是相对来说, 细胞数量足够的样本, 成功率更高。

### Q: 除了PE50测序策略, 是否可以其他测序策略?

A: 可以, 但根据文库特征, 大部分插入片段分布在100bp左右, PE50为最佳策略, 如客户选择其他测序策略, 影响到测序质量或信息分析结果, 后果由客户自己承担。也可送细胞样品, 如果客户送的是细胞样品, 细节参照《合作伙伴送样建议》, 且富集实验走定制化评估流程。

### Q: 推荐多少数据量? 能否接不同数据量的项目?

A: BGISEQ ATAC-Seq推荐50M clean reads/样本, 非常规样本需重新评估, 可根据项目执行情况进行增减。



BGI 华大基因

产品  
介绍

Product  
Profile

## 新产品

- 01 | 单管单细胞RNA测序
- 02 | 10X Genomics 单细胞RNA-Seq测序
- 03 | 单细胞DNA测序
- 04 | 免疫组库测序

# 单管单细胞RNA测序

## 产品概述

单管单细胞RNA测序，采用的是基于Smart-Seq2的扩增方法，获得全长转录本，随后Tn5转座酶进行建库测序，该扩增方法不仅可以用于转录本的定量，还可以分析结构变异和功能方面的信息。可以解决传统RNA测序技术在早期胚胎发育、干细胞、癌症、免疫等研究领域中的样品量极低或细胞异质性的问题，是单细胞水平研究基因表达强有力的工具，极大地拓展了RNA测序的应用范围。

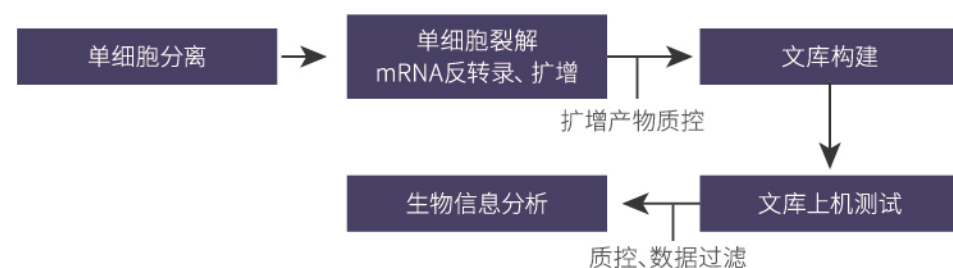
## 产品优势

-  起始量低至10pg;
-  单细胞百万倍扩增;
-  技术重复性高;
-  正负链相关性>98.0%;
-  探测到90%以上的gene表达。

## 产品应用



## 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

	PE100	SE50
(需要提供参考基因组序列)	<p>标准信息分析:</p> <ol style="list-style-type: none"> <li>1. 测序数据过滤;</li> <li>2. 参考基因组比对;</li> <li>3. 新转录本预测;</li> <li>4. SNP和INDEL检测;</li> <li>5. 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);</li> <li>6. 差异剪接基因检测;</li> <li>7. 基因表达量计算(基因比对率及表达数目统计、Reads对转录本的覆盖度及分布、样品间的相关性、PCA分析、样品中基因表达量的分布、样品间及组间的基因表达情况);</li> <li>8. SNP、INDEL、基因表达量和基因融合Circos图(仅限人的样本);</li> <li>9. 时间序列分析;</li> <li>10. 差异表达基因检测;</li> <li>11. 差异表达基因维恩图分析;</li> <li>12. 差异表达基因层次聚类分析;</li> <li>13. 差异表达基因GO功能分析;</li> <li>14. 差异表达基因Pathway功能分析;</li> <li>15. 差异表达基因TF编码能力预测;</li> <li>16. 差异表达基因蛋白互作分析;</li> <li>17. 差异表达基因中的真菌致病基因预测(真菌样本);</li> <li>18. 差异表达基因中的植物抗病基因预测(植物样本)。</li> </ol> <p>高级信息分析:</p> <ol style="list-style-type: none"> <li>19. 基因共表达网络分析(15个及以上样品, 不包含在标准执行周期内)。</li> </ol>	<p>标准信息分析:</p> <ol style="list-style-type: none"> <li>1. 测序数据过滤;</li> <li>2. 参考基因组比对;</li> <li>3. 基因表达量分析;</li> <li>4. 基因表达量聚类分析;</li> <li>5. 时间序列分析(3个或3组及以上样品);</li> <li>6. 差异表达基因检测;</li> <li>7. 差异表达基因维恩分析;</li> <li>8. 差异表达基因层次聚类分析;</li> <li>9. 差异表达基因GO功能分析;</li> <li>10. 差异表达基因Pathway功能分析;</li> <li>11. 差异表达基因转录因子编码能力预测;</li> <li>12. 差异表达基因蛋白互作分析;</li> <li>13. 差异表达基因中的真菌致病基因预测(真菌样品);</li> <li>14. 差异表达基因中的植物抗病基因预测(植物样品)。</li> </ol> <p>高级信息分析:</p> <ol style="list-style-type: none"> <li>15. 基因共表达网络分析(15个及以上样品, 不包含在标准执行周期内)。</li> </ol>
定制化信息分析	可结合客户的需求, 协商确定定制化分析内容。	



## ► 技术参数

### • 样本要求

单细胞: 细胞直径大于10 $\mu$ m, 如卵母细胞、动物桑葚期至囊胚期单个细胞, 细胞系分离的单细胞等。

微量细胞: 细胞数目在200个以下, 如动物桑葚胚、囊胚。

微量total RNA: 一般要求RNA 28S/18S $\geq$ 1, RIN $\geq$ 7, 浓度>50pg/ $\mu$ L。

表1 单细胞RNA测序送样要求

样品类型	测序类型	转录组/RNA-Seq
单细胞		1-2个 (4 $\mu$ L裂解液)
微量细胞		2 $\leq$ X $\leq$ 200 (4 $\mu$ L裂解液)

注: 表中X为细胞数目

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 SE50 (RNA-Seq) 或PE100 (转录组)

• 推荐数据量 SE50建议30M以上clean reads  
PE100建议5-10G clean data

## ► 案例分析

### 案例一: BGISEQ单细胞转录组数据与其他平台比较

文章选用468个单细胞来做测试, 有mESCs和K562两种细胞类型, Fluidigm C1和plate-based两个捕获平台组合, 对比了Illumina平台和BGISEQ平台单细胞转录组测序数据各个指标。由图1中可以看出BGISEQ平台在基因定量的灵敏度和准确度上表现优秀, 与Illumina平台高度一致。

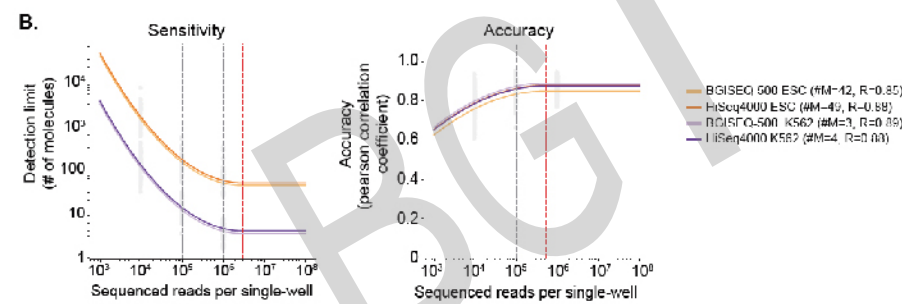


图1 BGISEQ平台和Illumina平台单细胞转录组数据灵敏度和准确度统计

### ★ 参考文献

Natarajan K N, Miao Z, Jiang M, et al. Comparative analysis of sequencing technologies platforms for single-cell transcriptomics[J]. Genome Biology, 2019 Apr 9;20(1):70.

### 案例二: 单细胞RNA-Seq对感觉传递神经元进行精确细胞亚群区分

文章通过大规模的单细胞RNA测序, 对799个细胞表达量PCA聚类分析, 鉴定离散的感觉神经元细胞群。鉴定结果显示神经元细胞个数为622个, 非神经元细胞个数109个, 68个细胞没有找到明确的定义。把这些细胞分成了5个类群, 其中一个类群属于非神经元, 其余4个类群是神经元。从分子水平验证了这些不同的神经元细胞类型, 并发现了一些新的细胞亚型和找到这些新的细胞亚型对应的marker; 发现了皮肤炎症 (例如过敏性皮炎) 相关的疾病激发的疼痛感是明显的联系到特定的一种痛感神经元相关。以及发现了单细胞RNA-seq的方法可以很好的分析和研究感官反应细胞与特定类型的神经元之间的对应关系。文章的结果说明了感官细胞的多样性及其对体内感觉的复杂性等。

### • 结果展示

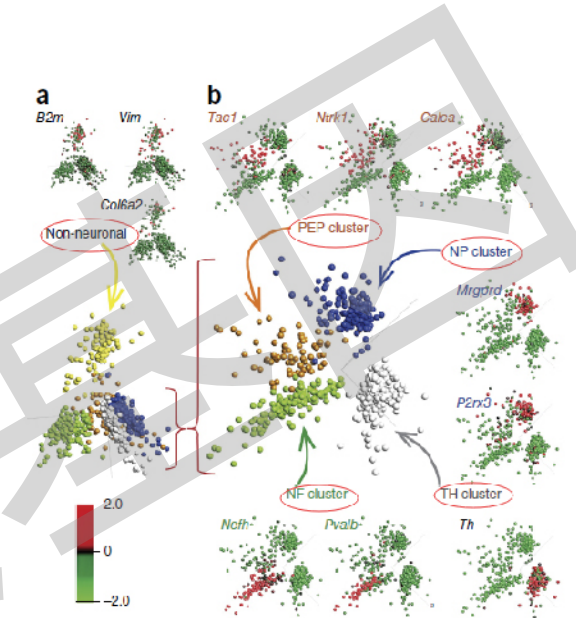


图2 细胞表达量PCA聚类图

图中每一个点为一个细胞, 颜色反馈的是表达量的高低, 根据表达谱的差异, 把细胞聚成了5个类群

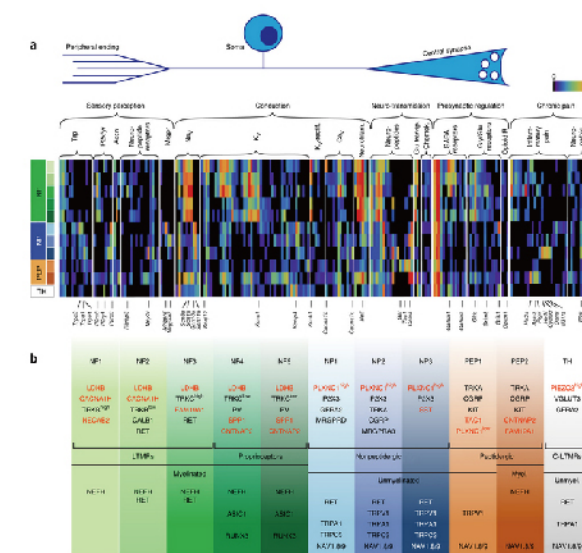


图3 通过对不同类群神经元进行表达量热图分析, 找到亚群特征性的基因

### ★ 参考文献

Usoskin D, Furlan A, Islam S, et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing[J]. Nature neuroscience, 2015, 18(1): 145.

# 10X Genomics 单细胞RNA-Seq测序

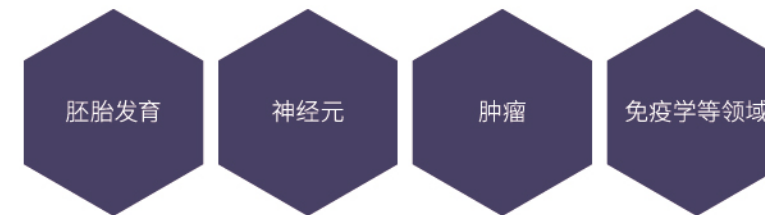
## 产品概述

基于10X Genomics平台，利用微流控液滴法等技术来实现高通量的细胞捕获及文库构建，结合高通量测序平台，获得单个细胞的3'端的转录组数据，进而实现区分混合细胞类型、鉴定单一类型细胞分化周期，区分癌细胞的异质性、鉴定罕见细胞亚群等。目前主要应用在发育、神经元、肿瘤、免疫学等领域。

## 产品优势

- 更专业的单细胞研究团队**  
单细胞Droplet技术、微流控方向资深研发技术总监率领的研发团队；
- 更丰富的项目经验**  
华大基因成立至今，单细胞测序技术服务时间长达10余年，期间，合作伙伴遍布全球，多项科研成果在顶尖杂志上发表；
- 更多样的样本类型**  
可接受常规的细胞悬液样本、细胞系样本以外，还可以接收组织样本，提供组织解离服务（肿瘤组织）；
- 更先进的10X Genomics分选建库仪器**  
一次建库完成10,000个细胞的分离和文库制备，独立的GEMs (Gelbead in Emulsion) 反应体系，避免细胞间的交叉污染；无需繁琐的细胞分离实验，高度自动化集成系统；
- 更接近前沿热点信息分析内容**  
除标准分析以外，还可提供模拟时空变化表达聚类分析等，多种可视化分析软件可供选择，让数据易懂性更强；
- 自主平台**  
特有的DNB纳米球滚环扩增技术，保证测序的准确性。

## 产品应用



## 技术流程

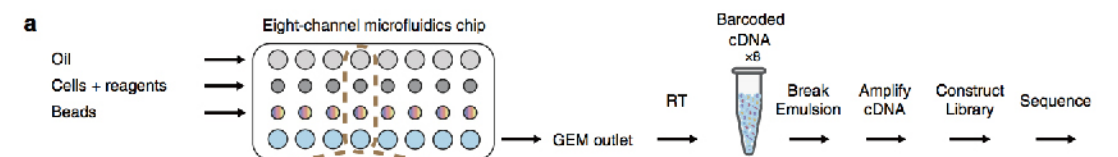


图1 10X Genomics平台单细胞RNA-Seq工作流程图

## 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10μM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，这个问题是无法避免的，会随机丢失一部分低丰度转录本的信息。不过根据我们的经验，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: 客户是否有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。

## 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
人鼠胚胎发育	2013.08	Nature	Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing
HeLa细胞系	2015.11	GigaScience	Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells.
骨间质干细胞	2016.02	PLoS One	RNA-Seq Reveals the Angiogenesis Diversity between the Fetal and Adults Bone Mesenchyme Stem Cell
膀胱鳞状细胞	2016.09	Oncotarget	Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder
早期人类胚胎形成	2016.09	BMC Genomics	Single-cell RNA sequencing reveals dynamic changes in A-to-I RNA editome during early human embryogenesis.
Hela细胞系	2016.11	Biotechnology Letters	Single-cell RNA-seq reveals lincRNA expression differences in HeLa-S3 cells
小神经胶质	2018.02	Nature Neuroscience	Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion
人多能干细胞	2018.11	GigaScience	Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation.
mESCs和K562两种细胞	2019.04	Genome Biology	Comparative analysis of sequencing technologies for single-cell transcriptomics



## ► 信息分析内容

信息分析条款	信息分析内容
标准信息分析 (需要基于良好的参考序列,且每个样本拆分出的细胞数目上下浮动在20%或以内的误差皆属于正常范围)	<ol style="list-style-type: none"> <li>1. 测序结果统计</li> <li>2. 数据质控统计</li> <li>3. 比对结果统计</li> <li>4. 基因表达定量分析</li> <li>5. Marker基因鉴定</li> <li>6. 细胞类型注释</li> <li>7. 单样本/多样本细胞聚类分析</li> <li>8. 特异marker基因差异分析</li> <li>9. 样本间相同cluster基因差异分析</li> <li>10. Cluster差异表达基因GO功能分析</li> <li>11. Cluster差异表达基因Pathway功能分析</li> <li>12. Cluster差异表达基因TF编码能力预测</li> <li>13. Cluster差异表达基因蛋白互作分析</li> <li>14. Cluster基因相关性网络分析</li> </ol>
高级信息分析	细胞轨迹分析(仅选取时间点样本可选做)
定制化信息分析	可结合客户的需求,协商确定定制化信息分析内容

## ► 技术参数

### • 样本要求

类型	组织: 新鲜/冻存的肿瘤组织, 新鲜人全血 细胞: 新鲜/冻存肿瘤细胞、生殖细胞、胚胎细胞、免疫细胞、PBMC、其他原代细胞、细胞系等
细胞状态	细胞直径<40 μm; 细胞活性>80%; 细胞浓度>5×10 <sup>5</sup> /ml 细胞悬液背景干净, 无大量结团、碎片及杂质, 不含Ca <sup>2+</sup> 和Mg <sup>2+</sup>

### • 测序策略 BGISEQ平台 PE100/PE150

### • 推荐数据量

细胞系样本, 建议每个细胞平均测: 50,000条Raw Reads; 10,000个细胞/lane  
组织样本, 建议每个细胞平均测: 500,000条Raw Reads; 10,000个细胞/lane  
PBMC (外周血单个核细胞) 样本, 建议每个细胞平均测: 50,000条Raw Reads

## ► 案例分析

### 案例一: 胚胎发育单细胞研究

本文中的研究人员采用了大规模胚胎细胞的单细胞测序技术, 检测在小鼠胚胎发育最初48小时的细胞分化(细胞亚群)和基因表达的变化。

样本取自411个完整的受精后6.5-8.5天的小鼠原始胚胎。每间隔6小时取样一次, 一共分离得到了116,312个单细胞, 并进行单细胞转录组测序。使用了10× Genomics平台分批对各个时期的单细胞进行RNA-Seq以及细胞聚类和注释等分析, 最后得到了37个重要的细胞亚群。同时, 发现这些亚群细胞类型与样本采集时间点密切相关; 并且在一些重要的时期, 比如器官开始发生分化时, 每个胚层细胞多样性相应地发生了变化。据此, 本文构建了一个小鼠原肠胚形

成和早期器官发生的单细胞分辨率图谱, 它描绘了细胞发育轨迹。本研究构建的哺乳动物早期胚胎发育图谱模型, 在研究多种发育突变体的分子机制和细胞发育轨迹上都具有重要作用, 不仅能促进我们对脊椎动物胚胎发育的了解, 还能为各种发育相关的疾病研究、机制研究带来极大的便利, 该图谱是研究界的一个重要且有用的资源。

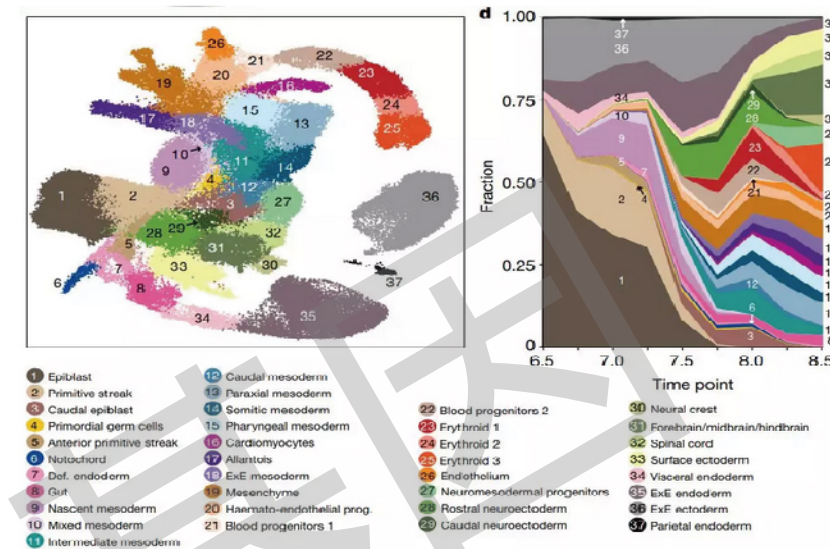


图2 胚胎发育细胞亚群和随时间推移发生的分化谱系图

### 案例二: BGISEQ平台脐带血细胞特异性研究

脐带血 (UCB) 移植, 是治疗小儿和成人多种血液系统疾病类型的最佳选择, 例如: 血癌、骨髓增生性疾病、遗传性疾病和代谢障碍等。然而, 细胞异质性水平和有核细胞的多样性在UCB尚未进行评估。所以, 此次研究对脐带血核细胞进行单细胞RNA测序, 绘制单细胞RNA表达图谱, 为后续研究提供重要的数据资源。此次构建了19,052个脐带血细胞的转录组图谱, 分成11个主要细胞类型。这些细胞类型由不同的亚群组成, 包括UCB中NK和NKT细胞类型。有核红细胞 (NRBC) 的假定量分析发现不同的激活和转录调节因子的抑制机制, 导致细胞状态极化。UCB中的祖细胞还包括两个亚群, 因激活了不同的转录, 导致特异性细胞分化差异。总的来说, 此次研究的单细胞转录组研究表明, 它可以揭示未知的细胞类型, 代谢途径和基因表达调控, 这些调控可能有助于UCB移植的疗效和结果, 拓宽了研究范围和临床创新。

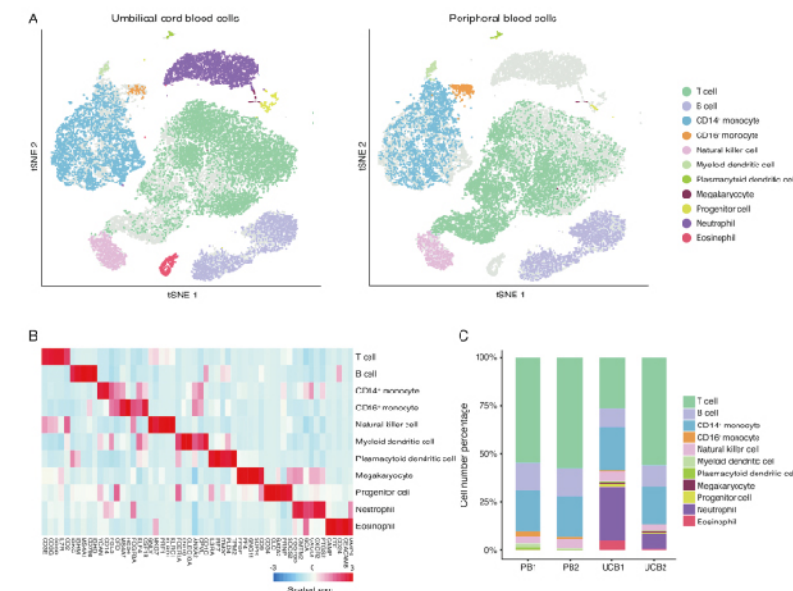


图3 外周血和脐带血不同细胞类型和比例

- 二维对外周血和脐带血细胞根据已知标记进行分类, 左侧图为脐带血细胞, 右侧图为外周血细胞, 对相同类型的细胞标记相同的颜色
- 脐带血和外周血中主要的已知标记的基因表达热图
- 外周血样本1/2和脐带血1/2细胞类型比例, 使用相同的颜色(同分类A)标记











# 单细胞DNA测序

## 产品概述

单细胞重测序首先使用口吸管分离的方法对组织或者细胞悬液的细胞进行分析，针对分离的单细胞进行高保真高覆盖度的全基因组扩增，然后对扩增产品进行全基因组、外显子或者目标区域测序，再对高通量测序产生的数据进行信息分析，从而快速全面获取单细胞水平上所有变异信息的产品，用于揭示细胞群体差异和细胞进化关系。它解决了用组织样本测序或样本少无法解决的细胞异质性难题，为从单核苷酸水平深入研究癌症发生、发展机制及其诊断、治疗提供了新的研究思路并开辟了新的研究方向。

## 产品优势

-  丰富的单细胞分离经验；
-  单细胞级别的DNA起始量；
-  单个样品可被扩增上百万倍，满足测序及后续分析的需求；
-  扩增产物长度可达20kb，有利于更多类型的变异检测；
-  高质量的扩增和极低的扩增错误率( $10^{-6}$ - $10^{-7}$ )；
-  平均可检测>95%的基因组位点；
-  技术稳定性较好、重复性高；
-  BGISEQ自主平台性价比高。

## 产品应用



## 常见问题

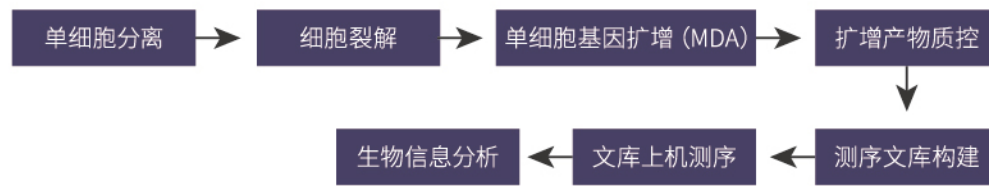
- Q: 10X Genomics平台分离的单细胞会有细胞大小的限制吗?**  
 A: 细胞大小会有一些限制，但限制条件相对固体芯片没有那么严格。同时，生产实验过程中，为了避免细胞出现长团情况，在进行10X平台细胞分离处理前，会做一个粗略的过筛处理。建议细胞大小<40 $\mu$ m。
- Q: 请问10X Genomics单细胞RNA-Seq可以进行可变剪切等结构变化分析吗?**  
 A: 10X Genomics单细胞RNA-Seq产品主要定位为基因表达定量分析为主，数据不建议进行可变剪切等基因结构变化分析。如需研究结构变异信息，可进行单管单细胞RNA测序。
- Q: 10X Genomics单细胞RNA-Seq采用的3'端扩增技术与Smart-seq2的扩增技术有什么区别?**  
 A: Smart-seq2扩增技术针对的是3'-5'端全部的mRNA；3'端扩增技术主要是扩增3'端，是否能延伸至5'端，取决于酶的活性等系列实验因素决定。因此，扩增产物的长度上会有一些区别和差异。
- Q: 如果客户现场分离了1,000个细胞，并检测出细胞活度很高，问：可否直接使用1,000个细胞进行10X Genomics平台上机?**  
 A: 不建议。因为10X Genomics单细胞RNA-Seq最低建议细胞送样量为 $10^6$ 个细胞，且活率建议在90%以上。
- Q: 单个细胞的测序reads数目应该在多少合适?**  
 A: 根据研究目的，细胞类型，以及分析内容的差异，对测序数据量会有不同的要求。针对不同的细胞类型及其相应的测序饱和度经验值，建议：细胞系样本，平均每个细胞测50,000条Raw Reads；组织样本，平均每个细胞测500,000条Raw Reads；PBMC样本，平均每个细胞测50,000条Raw Reads。同时，不同研究目的，以及不同分析内容，视具体情况而定。

## 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
鹿茸生茸组织	2013.08	BioRxiv	Single-cell transcriptome of antler stem cells from antlerogenic periosteum
小鼠胚胎发育	2013.08	Nature	Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing
HeLa细胞系	2015.11	GigaScience	Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells.
骨间质干细胞	2016.02	PLoS One	RNA-Seq Reveals the Angiogenesis Diversity between the Fetal and Adults Bone Mesenchyme Stem Cell
膀胱鳞状细胞	2016.09	Oncotarget	Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder
早期人类胚胎形成	2016.09	BMC Genomics	Single-cell RNA sequencing reveals dynamic changes in A-to-I RNA editome during early human embryogenesis.
Hela细胞系	2016.11	Biotechnology Letters	Single-cell RNA-seq reveals lincRNA expression differences in HeLa-S3 cells
小神经胶质	2018.02	Nature Neuroscience	Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion
脐带血细胞	2018.06	GigaScience	Single-cell Transcriptomic Landscape of Nucleated Cells in Umbilical Cord Blood
人多能干细胞	2018.11	GigaScience	Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation.
mESCs和K562两种细胞	2019.04	Genome Biology	Comparative analysis of sequencing technologies for single-cell transcriptomics



## 技术流程



## 产品分类

- 单细胞全基因组重测序
- 单细胞外显子测序
- 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检测、注释和统计
4. InDel检测、注释和统计
5. CNV检测、注释和统计 (仅全基因组测序)
6. SV检测、注释和统计 (仅全基因组测序)
7. 可结合客户的需求, 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

### • 样本要求

样品类型	测序类型	WGS	转录组/RNA-Seq
单细胞		1-2个 (4μL PBS悬液)	1-2个 (4μL PBS悬液)
微量细胞		2≤X≤1000 (4μL PBS悬液)	2≤X≤1000 (4μL PBS悬液)

### • 测序策略 PE101或PE151

### • 推荐数据量

针对大批量细胞的测序, 可以进行低深度浅覆盖测序; 对于细胞之间亲缘关系较大的, 可以适当增加测序深度; 针对小规模细胞测序, 研究功能性变化的, 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; 其余个性化研究目的, 测序深度视不同的研究目的可适度调整。

### • 项目周期 <100个样本 (30X) 运转周期约为50个工作日

## 案例分析

### 案例一: 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。

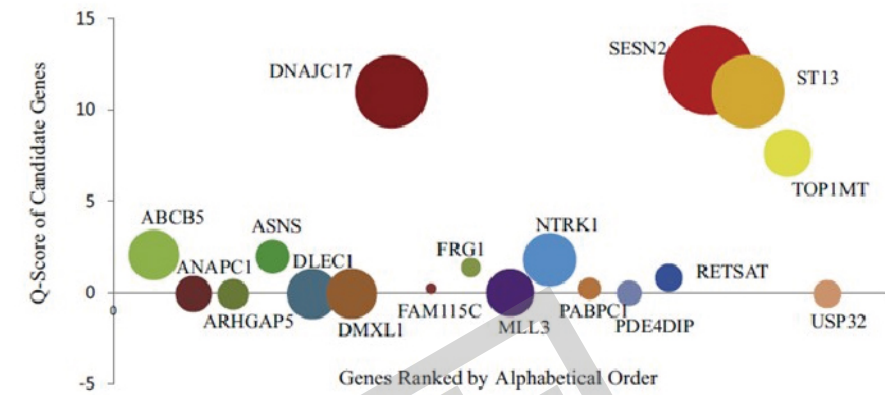


图1 ET病人的关键基因鉴定

通过Q score值挑选出18个ET病人的候选基因的driver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, Zhu P, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm[J]. Cell, 2012, 148(5): 873-885.

### 案例二: 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, 研究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。

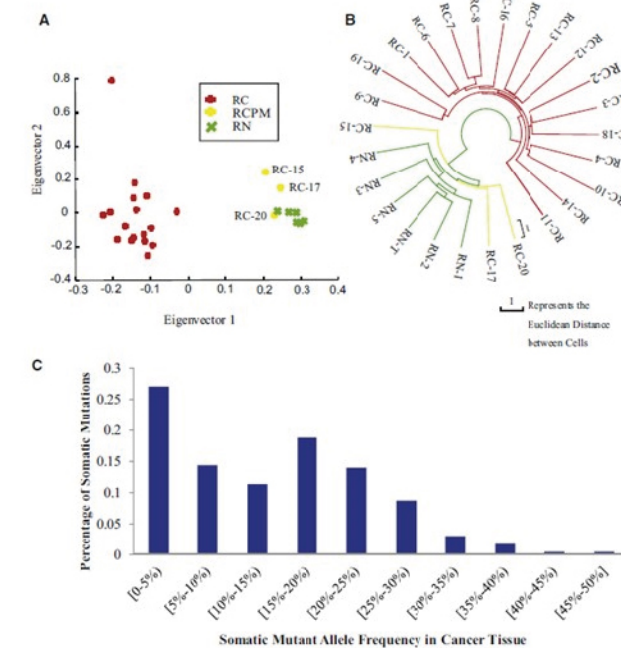


图2 PCA分析

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, Yin X, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor[J]. Cell, 2012, 148(5): 886-895.

## ► 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: 华大对单细胞不经过任何染色, 在显微镜下无法识别细胞是肿瘤细胞还是正常细胞, 建议客户取样前通过病理学的方法估计肿瘤组织的纯度, 我们一般要求纯度高于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, 这会加大项目投入。

### 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: 华大对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重测序送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: 扩增均匀性和等位基因丢失。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。除了存在扩增偏向性之外, 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的现象)。针对这些问题华大升级了扩增的方法, 使得基因组上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: 单细胞重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。不同的测序类型可以进行不同类型的变异检测, 外显子测序和目标区域捕获测序只能进行SNP和Indel的检测; 全基因组测序则能进行所有类型的变异检测, 如SNP、Indel、CNV、SV, 最全面地反馈单细胞水平上的所有变异类型。

## ► 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
骨髓增生性肿瘤	2012.03	Cell	Single-Cell Exome Sequencing and Monoclonal Evolution of a JAK2-Negative Myeloproliferative Neoplasm
ccRCC	2012.03	Cell	Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor
膀胱癌	2012.08	GigaScience	Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer
胚胎发育 CNV	2013.01	PLoS one	A Single Cell Level Based Method for Copy Number Variation Analysis by Low Coverage Massively Parallel Sequencing
双克隆起源结肠癌	2014.04	Cell Research	Discovery of biclonal origin and a novel oncogene SLC12A5 in colon cancer by single-cell sequencing
植入前β地贫	2015.02	Clinical Chemistry	Embryo Genome Profiling by Single-Cell Sequencing for Preimplantation Genetic Diagnosis in a β-Thalassemia Family

研究内容	发表时间	发表期刊	文献标题
CTC	2016.06	Oncotarget	Tumor-selective replication herpes simplex virus-based technology significantly improves clinical detection and prognostication of viable circulating tumor cells
膀胱癌干细胞	2016.07	European Urology	Single-cell sequencing reveals variants in ARID1A, GPRC5A and MLL2 driving self-renewal of human bladder cancer stem cells
CLL	2016.12	Nature Communication	Evolution of multiple cell clones over a 29-year period of a CLL patient
PGD检测染色体异常	2017.10	European Journal of Medical Genetics	Massively parallel sequencing on human cleavage-stage embryos to detect chromosomal abnormality
CTC	2017.11	Cell Biology and Toxicology	Progress and challenges of sequencing and analyzing circulating tumor cells
乳腺癌	2018.04	Journal of Clinical Investigation	Coexisting genomic aberrations associated with lymph node metastasis in breast cancer.
多发性骨髓瘤	2018.05	Blood	Single cell exomes in an index case of amp1q21 multiple myeloma reveals more diverse mutanomes than the whole population



# 免疫组库测序

## 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;



**可重复性高**  
同一样本建库两次克隆一致性高;



**产品多元化**  
可提供人 T/B 细胞不同链 CDR3 的检测服务。

研究对象	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。

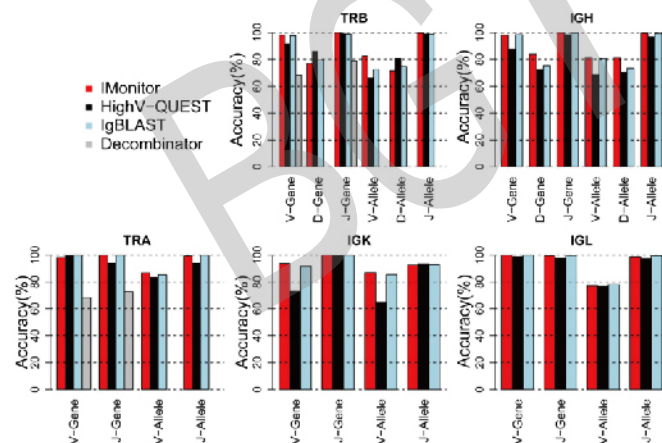


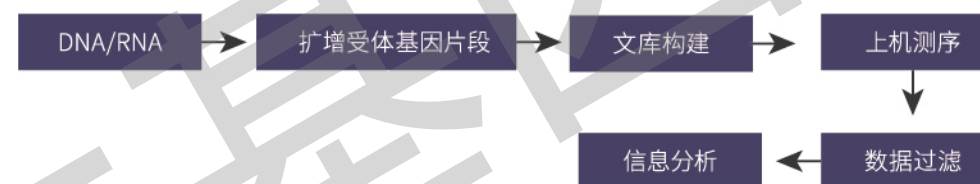
图1 0.5%测序错误率下不同软件检测V/D/J gene & allele准确性评估  
(Genetics. 2015 Oct;201(2):459-72.) 采用105个序列进行模拟, 评估四种软件检测V/D/J基因的准确性。

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。

## 产品应用

- 肿瘤**
  - 肿瘤浸润淋巴细胞
  - 肿瘤特异性抗原
  - 肿瘤病程研究
  - 肿瘤异质性研究
- 免疫疾病**
  - 免疫病生物标志物研究
  - 免疫病病程研究
- 移植**
  - 受体免疫重构研究
  - 排斥反应研究
- 药物反应**
  - 抗体库测序
  - 群体疫苗反应评估
- 其他疾病**
  - 感染病、免疫相关疾病研究

## 技术流程



## 信息分析内容

- 基本数据统计**
  - 数据过滤: 去除接头污染序列及低质量reads
  - 数据搭建: 数据拼接, 消除测序背景及有效数据构建
  - 数据统计: 数据产出统计及测序数据的成分和质量评估
- 数据比对分析**
  - 比对分析: 与数据库 (IMGT) V/D/J 基因片段比对
  - 重新比对: 寻找最佳 V/D/J 比对结果
- 序列结构分析**
  - 分析 CDR 序列组成及序列碱基成分
  - 编码 CDR 序列翻译成氨基酸和肽链
- 免疫组库构建**
  - 构建免疫组库表达谱: 统计多样性抗体库克隆表达情况
  - 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**
  - 样品间多样性差异分析 (辛普森系数、香农威纳系数)
  - 样品间克隆表达差异分析 (CDR3, V-D-J)
  - 分组样品间的克隆表达差异分析 (CDR3, V-D-J)

## ▶▶ 技术参数

### • 样本要求

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>	

- 测序策略 推荐 TCR 101PE/BCR 151PE
- 推荐数据量 至少1G raw reads
- 项目周期 40 个工作日

## ▶▶ 案例分析

### 案例一: 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TCRβ CDR3区域, 分析TCRβ在不同群体中的表达特征。研究发现, 癌症病人的TCRβ很相似, 与健康人完全不同。而HCC、ICC、MHC病人的TCRβ有差异, 可用于区分健康人、肝癌病人、乙肝病人。

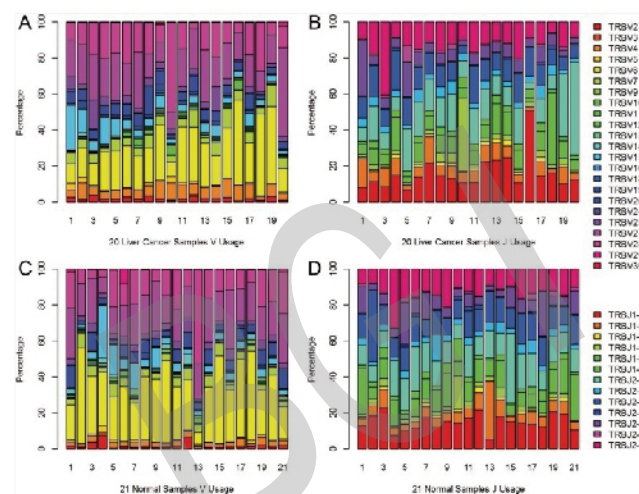


图2 肝癌病人和正常样品的V-J基因(V-J usage)半定量结果

- 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

Y Han Y, Liu X, Wang Y, et al. Identification of characteristic TRB V usage in HBV-associated HCC by using differential expression profiling analysis[J]. Oncoimmunology, 2015, 4(8): e1021537

## ▶▶ 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alpha或beta, BCR的H或L链需要分开收建库、测序、分析费用, 因为每条链都要分别作扩增建库和测序。

## ▶▶ 华大合作发表文章(部分)

研究内容	发表时间	发表期刊	文献标题
肝癌、乙肝	2015.04	Oncoimmunology	Identification of characteristic TRB V usage in HBV-associated HCC by using differential expression profiling analysis
综述	2015.07	Cancer Letters	Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma
方法学	2015.10	Genetics	IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis
方法学	2016.03	PLoS One	Systematic Comparative Evaluation of Methods for Investigating the TCRβ Repertoire