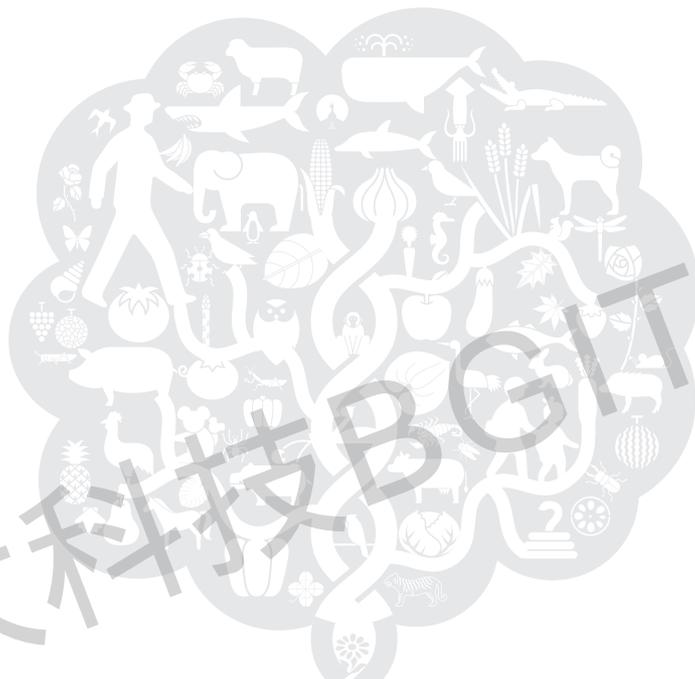




华大科技服务指南

●● 首选科技服务商 ●●

华大科技 BigITec



华大科技BGITech

目录

华大简介

- 003 公司简介
- 004 平台介绍
- 008 质量管理与资质

基因组学

- 011 动植物de novo测序
- 017 动植物全基因组重测序
- 021 人全基因组重测序
- 025 外显子组测序
- 029 细菌de novo测序
- 032 真菌de novo测序
- 035 细菌/真菌重测序

转录组学

- 039 转录组测序
- 044 全长转录组测序
- 049 RNA-Seq
- 053 Small RNA测序
- 057 长链非编码RNA测序

表观组学

- 063 全基因组甲基化测序 (WGBS)
- 067 ChIP-Seq

宏基因组学

- 073 16S/18S/ITS扩增子测序
- 080 宏基因组测序

蛋白质组学

- 087 定性蛋白质组学
- 090 定量蛋白质组学
- 097 修饰蛋白质组学
- 105 多肽组学
- 106 多组学关联分析
- 111 送样要求

代谢组学

- 115 非靶向代谢组学
- 117 脂质组学
- 119 靶向代谢组学
- 123 多组学关联分析
- 128 送样要求

合成生物学

- 131 普通基因合成
- 134 PCR克隆
- 135 定点突变
- 136 普通引物合成
- 137 修饰引物合成
- 139 Oligo Pools高通量芯片合成
- 140 代谢通路设计合成一站式服务

特色产品、服务

- 143 Dr. Tom
- 146 单管单细胞DNA测序
- 149 单管单细胞RNA测序
- 153 10X Genomics单细胞RNA-Seq
- 157 10X Genomics单细胞ATAC-Seq
- 161 10X Genomics单细胞免疫组库

常规分子生物学

- 167 Sanger测序
- 169 基因分型
- 173 甲基化验证
- 176 基因表达

华大科技BGITech

关于
我们

About
Us

| 基因科技 造福人类
Omics for All

华大科技BGITec

华大简介

公司简介	003
平台介绍	004
质量管理与资质	008

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“基因科技造福人类”的愿景，以推动生命科学研究进展、生命大数据应用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精准检测技术在民生健康方面的应用，致力于加速科技创新，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，全面助力精准医学。

华大基因主营业务为通过基因检测、质谱检测、生物信息分析等多组学大数据技术手段，为科研机构、企事业单位、医疗机构、社会卫生组织等提供研究服务和精准医学检测综合解决方案。公司总部位于中国深圳，在北京、天津、武汉、上海、广州、香港等国内主要城市设有分支机构和医学检验所，并在欧洲、美洲、亚太等地区设有海外中心及核心实验室，已形成“覆盖全国、辐射全球”的网络布局。

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科技服务商，为从事生命科学研究的机构和企业提供高质量、行业领先的基因测序、质谱、合成生物学、生物数据库、云计算等标准化的技术服务和综合的全流程解决方案。目前，公司服务已经覆盖了全球100多个国家和地区，拥有5,000多家合作单位，为20,000多位合作伙伴提供了杰出技术服务，更通过深度合作完成了一系列大型基因组科研计划和国际多边合作项目。科研积累上，截至2019年底，公司累计参与发表1,300+篇文章，其中CNNS 110+。华大自主测序平台DNBSEQ™2016年助力发表第一篇文章，截至2020年4月，已助力发表近900篇高质量文章。

平台介绍

核酸测序平台

拥有DNBSEQ™、Nanopore、PacBio、3730和HiSeq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



MGISEQ-200



BGISEQ-50



Nanopore平台 (PromethION)



Nanopore平台 (MinION)



PacBio平台 (Sequel I、II)



HiSeq 2500/4000

备注:

MGISEQ-2000、MGISEQ-200在部分海外国家和地区已更名为DNBSEQ-G400、DNBSEQ-G50

▶ 质谱平台

华大基因质谱平台始建于2007年，经过十余年的建设和发展，已形成基于高通量质谱检测技术为基础的蛋白质组学、多肽组学和代谢组学科学研究及高质量临床检测服务平台。华大基因质谱平台通过与基因组学、转录组学等高通量测序研究平台的优势整合，形成多组学关联分析的全贯穿研究模式，大力推动生命科学和系统生物学的快速发展。目前质谱平台拥有多台高精度、高分辨率质谱仪，能够开展大规模的蛋白质组学、代谢组学、以及靶标分子的检测和研究，并已经获得多项国家和省市的质量管理体系认证。

轨道离子阱质谱仪



Orbitrap Fusion Lumos



Q-Exacte/ Q-Exacte HF/
Q-Exacte HF-X



LTQ Orbitrap velos

三重四级杆质谱仪



QTRAP 6500+



TQS/TQD



TSQ Altis

飞行时间质谱仪 (TOF)



TripleTOF 5600+



Xevo-G2-XS



MicroTyper MS



Ultrafle Xtreme

▶ 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

▶ 合成平台

拥有192、768等多种型号的市场主流合成仪，及由多种行业领先的HPLC品牌机型组成的纯化平台。



Dr. Oligo 192合成仪



Waters 2695e系列高效液相色谱仪

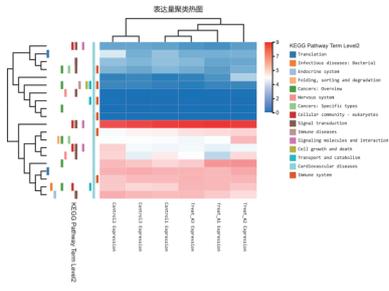


Agilent 1260系列高效液相色谱仪

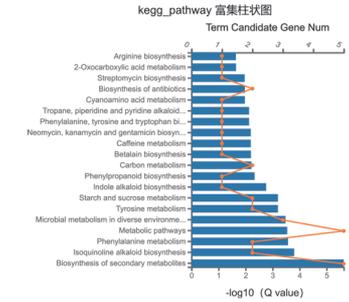
华天科技 BGITech

▶▶ Dr. Tom 系统

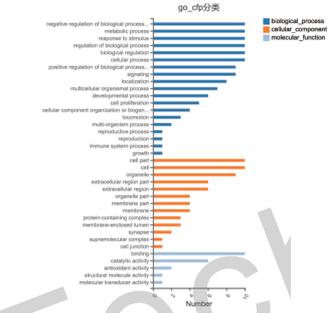
华大基因自主研发的多组学数据挖掘系统Dr. Tom是一款可以快速、深度、可交互式的用于挖掘基因和蛋白数据价值的系统，能够在复杂的多组学数据中快速找到目标基因和相关基因调控通路。该系统整合了多个国际常用的基因数据库和生物信息分析工具，并结合了云计算、人工智能和机器学习技术，还优化了数据挖掘的方式和展现形式。通过该系统，科研用户可以方便地进行自助式多组学数据挖掘。目前，该系统在全球已具有极强的竞争力，并且还在不断地升级和优化。



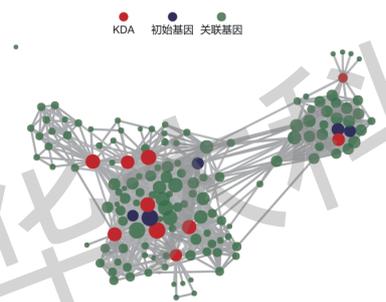
多维度表达量聚类热图



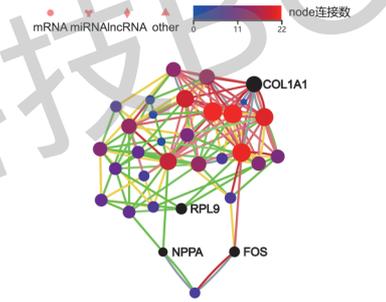
可切换气泡/柱状富集图



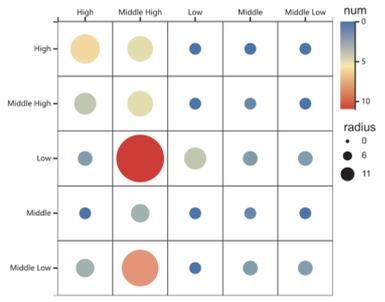
多数据库可选, 分类图



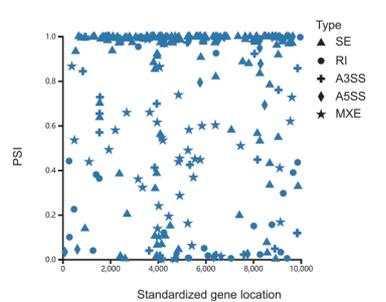
关键驱动基因分析, KDA图



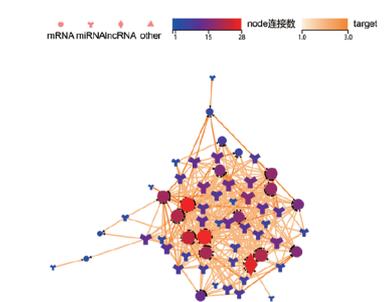
多种互动关系一键展示, 网络互动图



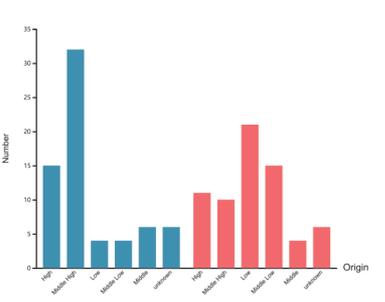
卡方检验



可变速切分析



关联网络图

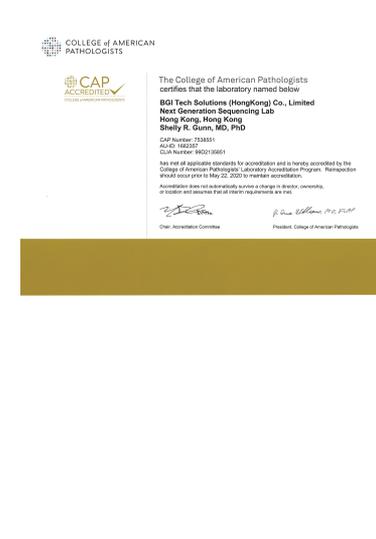
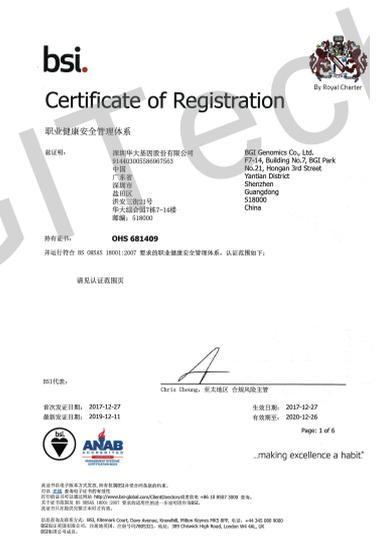
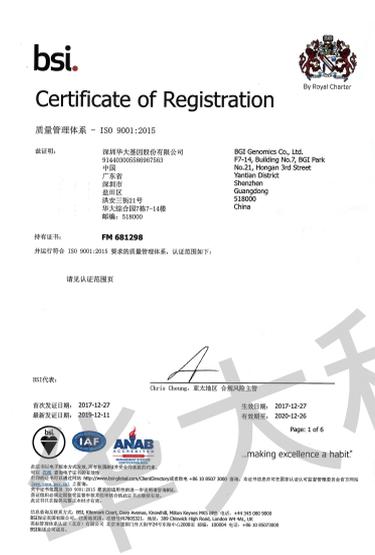


外部数据库-多组学关联分析

质量管理与资质

华大科技高度重视质量、环境、职业健康安全和信息化管理，2010年通过ISO 9001质量管理体系认证；2011年12月通过ISO 14001环境管理体系认证和OHSAS 18001职业健康安全管理体系认证；2013年6月通过ISO/IEC 27001信息安全管理体系认证。2014年5月，华大基因香港高通量测序实验室获美国病理学家协会（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基因测序领域通过CAP认可的实验室。2015年3月，武汉高通量测序实验室通过中国合格评定国家认可委员会（CNAS）审核，获得ISO/IEC 17025认可证书。

华大基因质谱平台从建立初期就重视质量的标准化建设，目前已获得ISO 9001、ISO 14001、OHSAS 18001三标一体认证，以及ISO/IEC 27001信息安全质量管理体系认证，处于同行业领先水平。此外，华大基因质谱平台已于2011年获批成立了深圳蛋白质组学工程实验室和广东省蛋白质组学工程实验室。



华大科技BGITech

产品
简介

Product
Profile

华大科技BGITec

基因组学

动植物de novo测序	011
动植物全基因组重测序	017
人全基因组重测序	021
外显子组测序	025
细菌de novo测序	029
真菌de novo测序	032
细菌/真菌重测序	035

动植物de novo测序

▶▶ 产品概述

从头测序即de novo测序，不需要任何参考序列资料即可对某个物种进行测序，用生物信息学分析方法进行拼接、组装，从而获得该物种的基因组序列图谱。利用全基因组从头测序技术，可以获得动植物的全基因组序列，带动这个物种下游一系列研究的开展，从而推进该物种的研究。全基因组序列图谱完成后，可以构建该物种的基因组数据库，为该物种的后基因组学研究搭建一个高效的平台，为后续的基因挖掘、功能验证提供DNA序列信息。华大基因结合短读长和长读长测序技术，可以高效、低成本地完成所有物种的基因组序列图谱。

▶▶ 产品优势



提取质量好

丰富的长片段提取经验，提取DNA质量高，提取成功率高达75%



准确度高

短读长和长读长测序联合组装、纠错，组装结果准确度高



分析经验多

分析经验丰富，已成功组装多类物种，为项目顺利交付保驾护航



平台多样

多平台联合应用保证超高测序通量，可实现快速交付



品牌底蕴足

华大基因已经完成1,000多个物种的全基因组从头测序，合作发表顶级期刊文章180+，其中26篇为封面文章

提取质量好

对于长读长测序来说，得到一份完整度高、纯度高的DNA是测序成功最关键的一步。华大基因专门针对长片段提取进行研发优化，提取成功率高达75%，针对很多疑难物种的优化提取获得了成功。

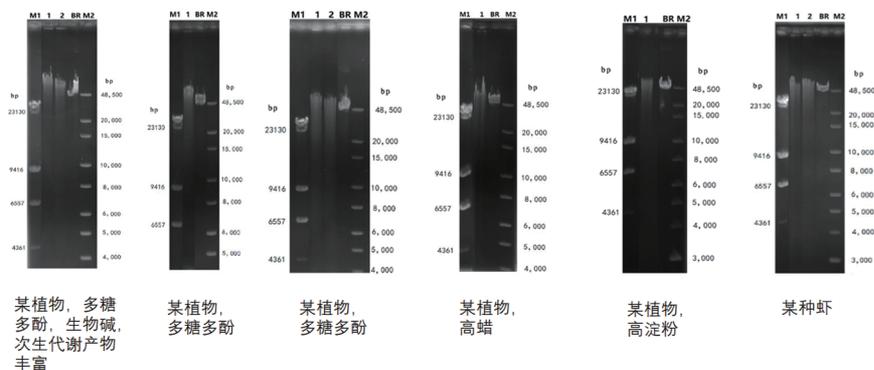


图1 疑难物种成功提取举例



图2 铁皮石斛提取成果

品牌底蕴足

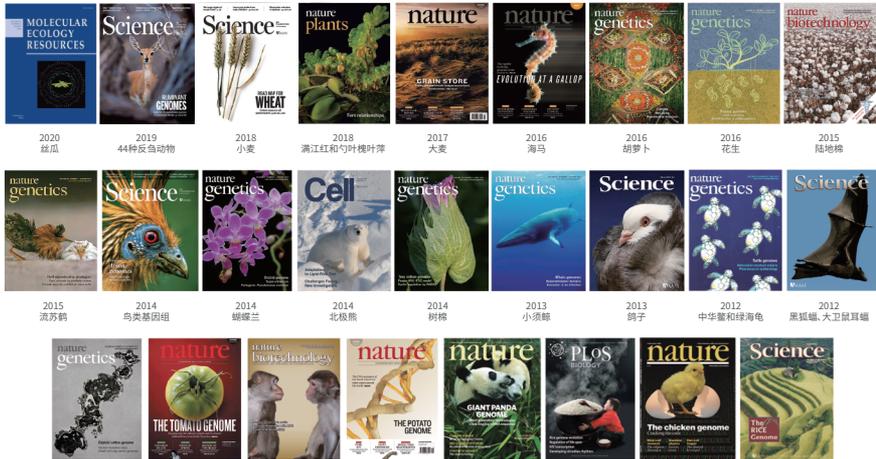


图3 华大参与发表的动植物基因组封面文章集锦

▶ 产品应用



▶ 技术流程

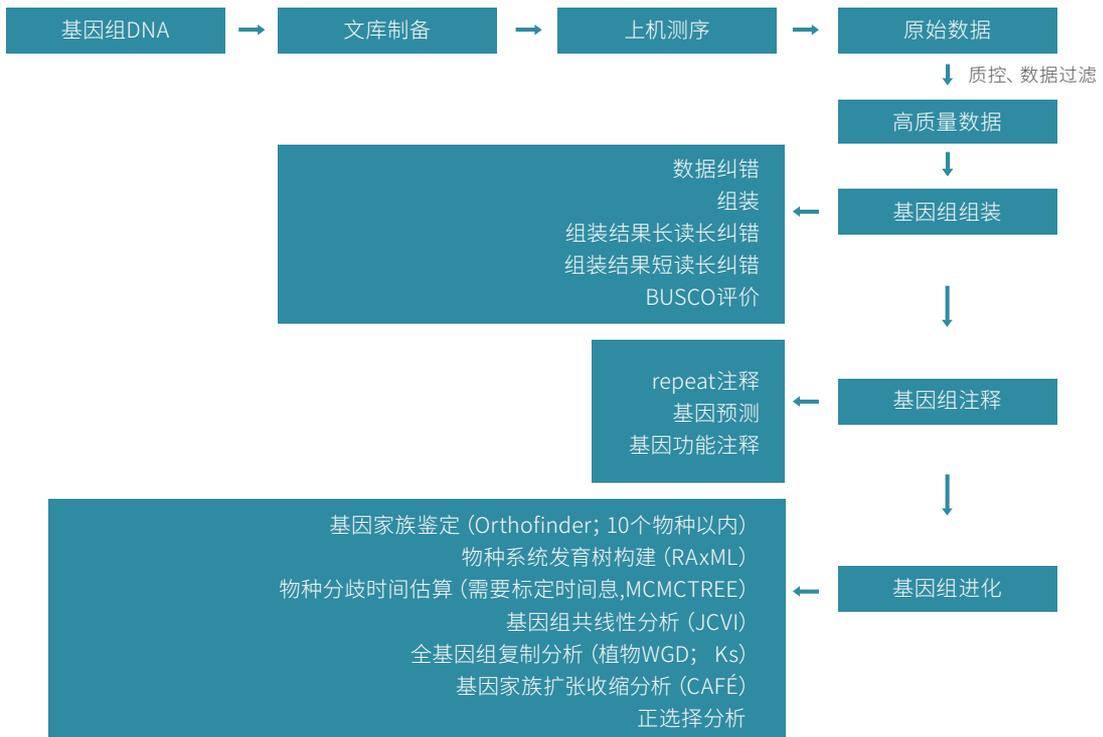


图4 动植物基因组de novo项目流程图

技术参数

样本要求:

表1 各个平台DNA送样要求

测序平台	文库类型	样品类型	样品浓度	样本量
Nanopore	20Kb	基因组DNA	20ng/ul	10μg
PacBio	20Kb		20ng/ul	10μg
DNBSEQ	350bp		20ng/ul	1.5μg
Hi-C	—	完成甲醛交联的DNA/新鲜血液/活体组织	—	—

注: 大片段文库不建议客户送DNA样本, 建议直接送组织: 组织样本需求量依据不同物种及组织样本类型而不同, 如有需要请咨询当地销售。

测序策略:

表2 动植物de novo各个平台的测序策略

测序平台	文库大小	测序读长	推荐测序深度	主要用途
Nanopore	20Kb及以上	10Kb以上	120X	组装
PacBio	20Kb及以上	10Kb以上	100X	组装
DNBSEQ	350bp	PE150	50-100X	Survey/纠错
Hi-C	350-450bp	PE100/PE150	100X	Hi-C辅助组装定位

注: 上文提及的辅助组装手段可全选, 也可根据物种特性及技术特点进行选择。

项目周期:

动植物: 简单基因组4个月, 复杂基因组6个月, 超大基因组周期另议。

案例分析

案例: 睡莲基因组和早期开花植物进化

发表期刊: Nature 测序策略: PacBio+Hi-C

★ 研究结果

1. 本项目通过122X的PacBio测序数据, 组装得到蓝星睡莲的高质量基因组, 组装基因组大小为409Mb, 组装指标为Contig N50=2.1Mb。通过系统发育组分析显示无油樟是最早分化出的被子植物类群, 其次是睡莲。

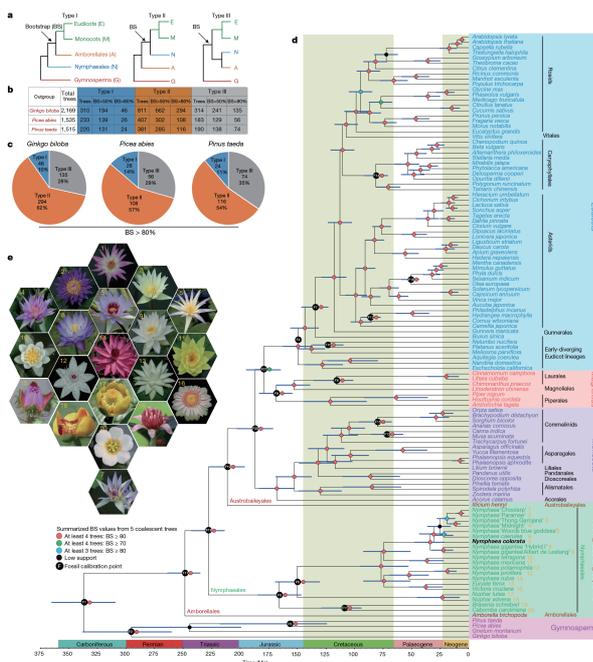


图5 睡莲基因组进化关系

2.通过基因组和转录组分析,睡莲科祖先发生了一次基因组加倍事件,这个事件可能在睡莲目祖先时期发生的。

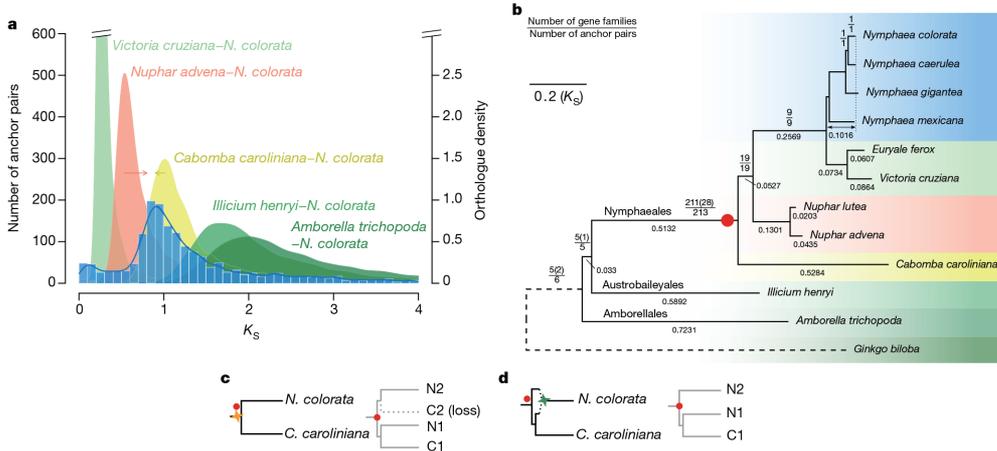


图6 睡莲基因组全基因组复制事件

3.确定了睡莲ABCE模型,揭示早期被子植物花发育特征。

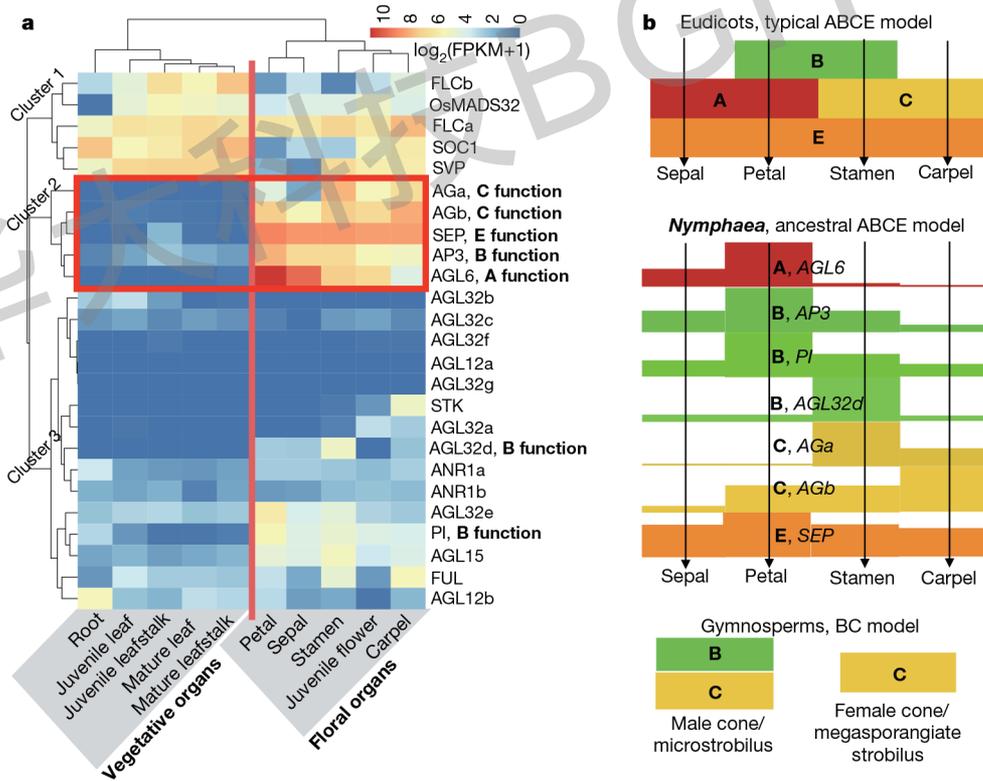


图7 睡莲花发育模型

4.本研究通过比较自然变异的白色花瓣蓝星睡莲与蓝色花瓣的转录组,发现两个重要基因可能编码蓝色花瓣合成途径关键酶。睡莲的花香有11种成分主要是萜类和脂肪酸等,其中合成倍半萜基因与单双子叶中已知基因不一样。

★ 参考文献

Zhang L, Chen F, Zhang X, et al. The water lily genome and the early evolution of flowering plants[J]. Nature, 2020, 577(7788):79-84.

► 常见问题

Q: 如何查询基因组的大小?

A: 查询植物基因组大小网站: C-values database
 查询动物基因组大小网站: Animal genome size database
 换算关系: 1pg=978Mb

Q: 如何判断简单基因组和复杂基因组?

A: 通常可以通过已测序的近缘种来判断物种复杂程度, 如一般鸟类和哺乳类动物都是简单基因组, 如果没有近缘物种供参考, 可以先做基因组survey评估。

简单基因组及复杂基因组定义		
基因组分类	普通基因组	复杂基因组
基因组大小	≤3.5Gb	>3.5Gb
染色体倍性	单倍体或纯合二倍体	杂合二倍体或多倍体
杂合率	<0.5%	≥0.5%
重复序列含量	<50%	≥50%
GC含量	35-65%	<35%或>65%
备注	满足所有条件为普通基因组	符合任一条件即为复杂基因组

Q: 基因组从头测序的组装结果好坏如何判断?

A: 一般用Contig N50和Scaffold N50来衡量基因组组装结果的好坏。N50是指把组装出的contigs或scaffolds从大到小排列, 当其累计长度刚刚超过全部组装序列总长度50%时, 最后一个contig或scaffold的大小即为N50的大小, N50对评价组装序列的连续性、完整性有重要意义; N70和N90的计算方法与N50类似, 只是百分数变为70%或90%。

Q: PacBio测序的优势是什么?

A: 最大的优势是测序读长长, 平均读长在12K以上, 且无GC偏向性; 对基因组的组装、大的结构变异检测、转录组全长测序结果均有极大提升。

► 华大合作发表文章(2017-2020, 部分)

华大基因自1999年成立以来, 发表文章总数超过1,000篇。截至2019年12月, 已经发表动植物de novo文章180+。

物种	发表时间	发表期刊	文献名称
植物研究			
丝瓜 <i>Luffa cylindrica</i> (L.)	2020	<i>Molecular Ecology Resources</i>	Long - read sequencing and de novo assembly of the <i>Luffa cylindrica</i> (L.) Roem. Genome
六倍体小麦-中国春 <i>Triticum aestivum</i> L	2018	<i>Science</i>	Shifting the limits in wheat research and breeding using a fully annotated reference genome
细叶满江红 <i>Azolla filiculoides</i> 勺叶槐叶萍 <i>Salvinia cucullata</i>	2018	<i>Nature Plants</i>	Fern genomes elucidate land plant evolution and cyanobacterial symbioses
大麦 <i>Hordeum vulgare</i>	2017	<i>Nature</i>	A chromosome conformation capture ordered sequence of the barley genome
胡萝卜 <i>Daucus carota</i>	2016	<i>Nature Genetics</i>	A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome

物种	发表时间	发表期刊	文献名称
花生A、花生B <i>Arachis duranensis</i> ; <i>Arachis ipaensis</i>	2016	Nature Genetics	The genome sequences of <i>Arachis duranensis</i> and <i>Arachis ipaensis</i> , the diploid ancestors of cultivated peanut
陆地棉 <i>Gossypium hirsutum</i> TM-1	2015	Nature Biotechnology	Genome sequence of cultivated Upland cotton (<i>Gossypium hirsutum</i> TM-1) provides insights into genome evolution
蝴蝶兰 <i>Phalaenopsis equestris</i>	2014	Nature Genetics	The genome sequence of the orchid <i>Phalaenopsis equestris</i>
树棉 (棉花A) <i>Gossypium arboreum</i>	2014	Nature Genetics	Genome sequence of the cultivated cotton <i>Gossypium arboreum</i>
雷蒙德氏棉 <i>Gossypium raimondii</i>	2012	Nature Genetics	The draft genome of a diploid cotton <i>Gossypium raimondii</i>
番茄 <i>Solanum lycopersicum</i>	2012	Nature	The tomato genome sequence provides insights into fleshy fruit evolution
马铃薯 <i>Solanum tuberosum</i>	2011	Nature	Genome sequence and analysis of the tuber crop potato
水稻 (9311精细图) <i>Oryza sativa ssp.indica</i>	2005	PLOS Biology	The genomes of <i>Oryza sativa</i> : A history of duplications
水稻 (9311框架图) <i>Oryza sativa ssp.indica</i>	2002	Science	A draft sequence of the rice (<i>Oryza sativa ssp. indica</i>) genome
动物研究			
44个反刍动物 ruminants	2019	Science	Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits
海马 <i>Hippocampus comes</i>	2016	Nature	The seahorse genome and the evolution of its specialized morphology
流苏鹬 <i>Philomachus pugnax</i>	2015	Nature Genetics	Structural genomic changes underlie alternative reproductive strategies in the ruff (<i>Philomachus pugnax</i>)
48种鸟类 (45种为新测) <i>Acanthisitta chloris</i> 等	2014	Science	Whole Genome Analyses Resolve the Early Branches to the Tree of Life of Modern Birds
北极熊 <i>Ursus maritimus</i>	2014	Cell	Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears
小须鲸 <i>Balaenoptera acutorostrata</i>	2013	Nature Genetics	Minke whale genome and aquatic adaptation in cetaceans
中华鳖和绿海龟 <i>Pelodiscus sinensis</i> ; <i>Chelonia mydas</i>	2013	Nature Genetics	The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan
鸽子 <i>Columba livia</i>	2013	Science	Genomic diversity and evolution of the head crest in the rock pigeon
黑狐蝠和大卫鼠耳蝠 <i>Pteropus alecto</i> ; <i>Myotis davidii</i>	2012	Science	Comparative analysis of bat genomes provides insight into the evolution of flight and immunity
食蟹猴和中国恒河猴 <i>Macaca fascicularis</i> ; <i>Macaca mulatta lasiota</i>	2011	Nature Biotechnology	Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques
熊猫 <i>Ailuropoda melanoleura</i>	2009	Nature	The sequence and de novo assembly of the giant panda genome
家鸡 <i>Gallus sonneratii</i>	2004	Nature	Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution



动植物全基因组重测序

▶▶ 产品概述

动植物全基因组重测序是对已知基因组序列的物种进行DNA测序，并在此基础上完成个体或群体分析。通过序列比对，可以检测到大量变异信息，包括单核苷酸多态性位点 (SNP)，插入缺失位点 (Insertion/Deletion, InDel)、结构变异 (Structure Variation, SV) 位点，拷贝数变异 (Copy Number Variation, CNV) 位点等，获得同一物种不同个体的遗传变异图谱。利用全基因组重测序技术有助于快速发现与动植物重要性状相关的遗传变异，应用于分子育种中，缩短育种周期。

▶▶ 产品优势



经验丰富、数据精准

至今完成10万+的动植物重测序样本，合作挂名发表文章100余篇
累计影响因子1,000+，其中一作或通讯作者文章50+



自主测序平台、成本可控

滚环扩增构建DNB测序文库，PCR-free重测序检测InDel更精准，无index hopping之忧，
低dup rate无需人为干预

▶▶ 产品应用

目标性状
基因挖掘

群体遗传学
研究

变异图谱构建

分子标记
开发及辅助
选择育种

物种/品种鉴定

动植物核心
资源普查

▶▶ 技术流程

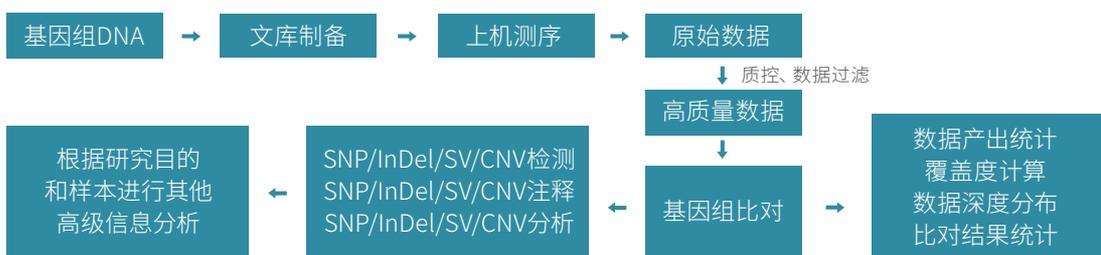


图1 动植物重测序技术流程图

▶▶ 技术参数

样品要求: 基因组DNA, 无降解或轻微降解

样品需求量 (单次): $\geq 1 \mu\text{g}$

样品浓度: $\geq 12.5 \text{ ng}/\mu\text{L}$

推荐数据量: 个体重测序推荐30X以上, 群体重测序推荐10X以上

交付周期: 具体项目的交付周期与样本数和信息分析内容有关

► 研究内容

信息分析条款		信息分析内容
标准信息分析		测序评估(碱基含量分布、测序质量分布及统计) 比对 比对统计(统计比对率、基因组覆盖度、测序深度等) SNP检测、注释及统计
		InDel 检测、注释及统计(群体项目不含) SV检测、注释及统计(群体项目不含) CNV检测、注释及统计(群体项目不含)
高级信息分析	群体进化	连锁不平衡分析(LD) 群体进化树分析(Phylogeny tree) 群体结构分析(Structure) 群体主成分分析(PCA) 多态性分析 选择分析(进化/驯化相关位点) (群体可分为野生种、驯化种/栽培种/地方种/商业化品种才可以做) 选择分析(环境适应性相关位点) (群体样本间根据明显的环境因素差异可以划分亚群才可以做) 选择区域基因注释(完成选择分析获得选择位点后做) 群体历史研究(最好群体有化石或史料信息,可以相互验证)
	BSA分析	子代混合样本SNP-index值计算 子代混合样本ED值计算 点突变位点检测、注释和统计 筛选两个DNA pool中等位基因频率差异显著的位点(两种方法交集) 候选基因检测、注释和统计
	遗传图谱构建及QTL定位	基因分型 遗传图谱/BinMap构建 QTL定位(需提供表型数据) 遗传图谱整合(需提供同一作图群体原图谱数据)
	全基因组关联分析(GWAS)	连锁不平衡分析(LD) 群体进化树分析(Phylogeny tree) 群体结构分析(Structure) 群体主成分分析(PCA) 基因型缺失数据处理(Imputation) GWAS 分析 筛选与性状相关的候选基因 候选基因功能注释
个性化信息分析		可结合客户的需求,协商确定个性化信息分析内容

► 案例分析

案例一: 3K水稻重测序&泛基因组研究(华大参与)

★ 研究背景

3,010份水稻(来自全球89个国家和地区)代表了全球78万份水稻种质约95%多样性的核心种质。通过全基因组重测序,每个样本平均测序深度14X,利用重测序数据共检测到32Mb的高质量SNPs和InDels。对亚洲栽培稻群体的结构和分化进行了更为细致和准确的描述和划分,由传统的5个群体增加到9个,分别是东亚(中国)的籼稻、南亚的籼稻、东南亚的籼稻和现代籼稻品种等4个籼稻群体,东南亚的温带粳稻、热带粳稻、亚热带粳稻等3个粳稻群体、以及来自印度和孟加拉的Aus和香稻。

研究首次揭示了亚洲栽培稻品种间存在的大量微细结构(>100bp)变异(SVs,包括易位、缺失、倒位和重复)。着重研究453个测序深度>20X的品系的SVs,利用SVs构建的进化树与SNP构建的进化树类似。大量的SVs可能是不同程度杂种不育和XI与GJ杂种衰退的遗传基础。同时构建了亚洲栽培稻的泛基因组,包括12,770个(62.1%)核心(core)基因家族和9,050个(37.9%)分散式(distributed)基因家族。发现了1.2万个全长新基因和数千个不完整的新基因。核心基因比较古老,大多数的新基因表现更年轻和长度偏短。

★ 研究策略

最初测序3,024份水稻样本，后来进行质控过滤掉14份，最终保留3,010份水稻样本进行深入研究。3K RG测序数据比对到参考基因组日本晴Nipponbare上检测SNPs、InDels。合并Nipponbare基因组序列和无冗余的新组装的基因组序列构建泛基因组。利用测序深度>20X，比对深度>15X的453个水稻材料进行SVs和PAVs分析。

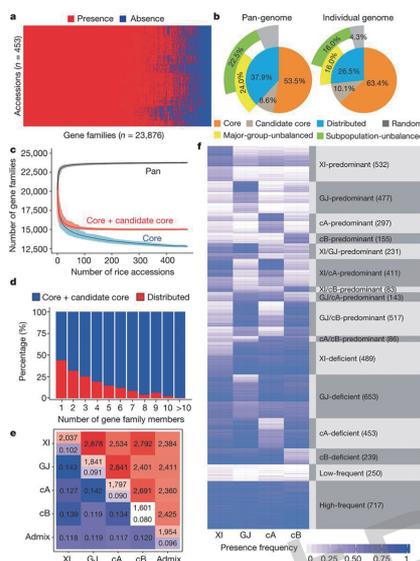


图2 水稻泛基因组

- 基因家族PAVs
- 泛基因组和一个单独的基因组的组成成份
- 基于500个随机筛选的水稻基因组模拟泛基因组和核心基因组
- 核心和分散式基因家族比例
- 两个品系间基因家族平均数量差异
- 5733主要群组不平衡基因家族特性

★ 参考文献

Wang W, Mauleon R, Hu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice[J]. Nature, 2018, 557(7703): 43.

案例二：鹰嘴豆群体进化和GWAS研究（华大参与）

★ 研究背景

研究人员应用429个鹰嘴豆品种4.97Mb的SNPs构建了进化树，7个野生种作为外群，栽培种形成4个亚群。进化树的拓扑结构与品种生物学地位和种子的2种主要类型均无关，这在作物群体研究当中比较少见，表明鹰嘴豆驯化路径较为复杂。

通过对来自于不同地理环境的鹰嘴豆样本的群体差异指数和群体多样性进行分析，揭示鹰嘴豆最初起源于地中海东部，经地中海/新月沃地到达中亚，之后可能从中亚同时到达东非（埃塞俄比亚）和南亚（印度）。分析数据也支持埃塞俄比亚是鹰嘴豆的第二起源中心，并且在鹰嘴豆的驯化过程中，有近80%的遗传多样性丢失。利用多样性衰减（ROD）和群体差异指数分析，得到基因组上受选择的候选区间122个，其中包含受选择的基因204个，后经注释发现，受选择的基因主要和胁迫应答、DNA修复、蛋白酶激活、种子发育、发芽、花发育等生理过程相关。研究同时针对272份鹰嘴豆样本，应用GLM/CMLM、FarmCPU和EMMAX模型，对3.65Mb的SNPs与来自多年多点（1-6个季节，1-6个地点）的20个与抗旱、耐热相关性状进行全基因组关联分析（GWAS），定位得到与13个重要性状相关的262个标记和候选基因，并发现地方种ICC14778具有稳定的耐热、抗旱性状，能用于培育抗性品种。

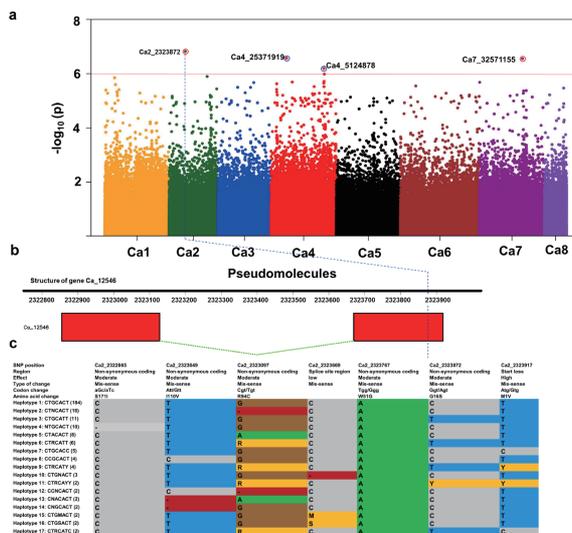


图3 高温胁迫下产量的GWAS分析结果

- SNP位点Ca2_2323872与早作条件下的产量有关联，候选基因Ca_12546预测和AB抗性相关；
- 基因Ca_12546结构信息；c. 基因Ca_12546的17个单倍型

★ 参考文献

Varshney RK, Thudi M, Roorkiwal M, et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits[J]. Nature Genetics, 2019, 51(5): 857.

▶▶ 常见问题

Q: 进行全基因组重测序数据推荐?

A: 每个样本推荐的数据量与样本类型和要做的信息分析内容相关。例如关注个体样本的SNP, 对SNP的准确度和覆盖度要求比较高, 一般推荐测序深度>30X, 对于稀有变异测序深度还要进一步提高; 用于研究群体结构的样本, 测序深度推荐10X以上, 纯合样本混样检测等位基因频率, 推荐平均每个样本的测序深度在1X以上, 混合样本测序深度不低于30X。

Q: 样本量选择多大合适?

A: 样本量大小与样本类型和研究目的相关。例如进行群体进化研究推荐30个样本以上, 因为在统计学上30个以上才属于大样本; 对于进行基因挖掘的项目来说, 无论是利用自然群体进行GWAS分析或是用家系群体进行连锁分析, 都是群体越大越好, 一般的情况下进行GWAS分析的样本推荐300个样本以上, 对于家系群体推荐200个以上。

Q: 连锁图谱构建适用于什么样的群体?

A: 连锁图谱的构建适用于作图群体, 它是由性状差异显著的亲本杂交衍生的群体。亲本选择的要求: 要考虑亲本间的遗传多态性、目标性状差异、亲本的纯合度和杂交后代的可育性。构建分离群体类型, 根据遗传稳定性可将分离群体分成两大类: 暂时性分离群体如 F_1 、 F_2 、BC等, 永久性分离群体如RIL、DH等。

Q: 重测序reads与参考基因组比对率低, 可能的原因是什么?

A: 重测序reads比对率低原因可能是: 1) 因为测序样本与参考基因组亲缘关系比较远。因为动植物品种多样, 但是目前已完成基因组组装的往往只是其中的一个品种, 同一个物种野生种与驯化种差异还是很大的; 2) 可能因为DNA不纯, 存在其他物种的污染; 3) 参考基因组序列组装质量较差, 引起比对率低; 4) 比对参数设置严格等。

▶▶ 华大合作发表文章 (2017-2019, 部分)

物种	发表日期	发表期刊	研究目的	研究内容
鹰嘴豆	2019	Nature Genetics	群体进化&GWAS	全基因组重测序鹰嘴豆揭示基因组多态性、进化和后进化候选基因
青稞	2018	Nature Communications	进化	青稞的起源与进化
梨	2018	Genome Biology	群体进化, 地理起源	亚洲梨和欧洲梨的多样性与独立进化
鹌鹑	2018	GigaScience	基因组+进化	群体基因组数据揭示与鹌鹑重要性状相关的基因
江豚	2018	Nature Communications	基因组+群体进化	鲸类对淡水环境的适应性进化机制
水稻	2018	Nature	群体进化&泛基因组	揭示亚洲栽培稻的起源和群体基因组变异结构, 剖析了水稻核心种质资源的基因组遗传多样性
梅花	2018	Nature Communications	群体进化+泛基因组+GWAS	木本植物梅花中花性状的遗传结构
古老大型犬科动物	2017	GigaScience	平台比较	DNBSEQ与HiSeq 2500平台比较



人全基因组重测序

产品概述

人全基因组重测序 (WGS) 是对人类不同个体或群体进行全基因组测序, 通过序列比对, 找到大量的单核苷酸多态性位点 (SNP), 插入缺失位点 (Insertion/Deletion, InDel)、结构变异位点 (Structure Variation, SV) 和拷贝数变异位点 (Copy Number Variation, CNV), 可全面挖掘DNA水平的遗传变异, 为筛选疾病的致病及易感基因, 研究发病及遗传机制提供重要信息。

全基因组测序是全基因组范围的突变, 包括外显子区、内含子区、调控区, 对检测拷贝数变化和基因重排具有无可替代的作用。

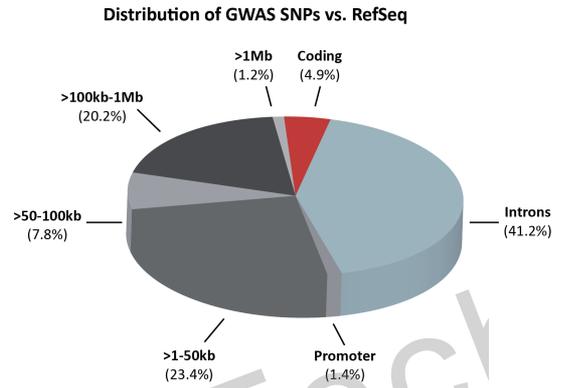


图1 许多常见疾病的变异位于非CDS区
Maurano MT, et al. Science, 2012

产品优势



自主测序平台、成本可控

滚环扩增构建DNB测序文库, PCR-free重测序检测InDel更精准, 无index hopping之忧, 低dup rate无需人为干预



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: 对原始数据进行去除接头污染序列及低质量reads的处理
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. 家系de novo突变研究高级信息分析

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:

DNA样本送样建议

样本类型	文库类型	总量	浓度	完整性(胶图)	纯度
Genomic DNA	PCR/PCR-Free	≥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

组织样本送样建议

组织类型	需求量
新鲜培养细胞(细胞数)	≥5×10 ⁶ cell
新鲜动物组织干重	≥50mg
全血(哺乳动物)	≥1ml
FFPE	≥10片、未染色、100mm ² 、5~10μm厚度

项目周期:

DNBSEQ+BGI Online实现人全基因组快速交付, 周期仅15个工作日, 让每个碱基都加速。

产品分类	交付周期	包含
人重快速交付	15个工作日	建库+测序+标准分析
人重常规交付	30个工作日	建库+测序+标准分析

交付指标:

建库策略	测序策略	推荐数据量
PCR/PCR-Free	PE100、PE150	产生不低于合同规定的clean data数据量

案例分析

案例一: 父母生育子女的年龄对后代的影响

★ 研究背景

研究者采集了14,688名冰岛人的血液或口腔细胞样品, 并对样品中的遗传物质进行了全基因组测序分析。该研究表明, 从基因组水平上来讲, 父母的年龄越大, 则其子女的新生突变频率就越高。相比目前来说, 父亲的生育年龄很大程度上决定了后代的新生突变, 而且后代中的这种新生突变数的增加与精神分裂症以及自闭症的发生概率具有一定的相关性。

随父母生育年龄变化, 源自父系的新生突变年增长数量是源自母系的四倍

测序识别出了包括SNP、InDel在内的共108,788个新生突变, 即受试的1,548个家族的“基因平均突变率”为70.3个/家族。从新生突变数量角度, 研究者发现, 以225组祖孙三代样本作为研究群体时, 源自母亲的新生突变数量随母亲年龄的增长每年增加0.37个, 而源自父亲的新生突变数量是前者的四倍, 随父亲年龄的增长每年增加1.51个。

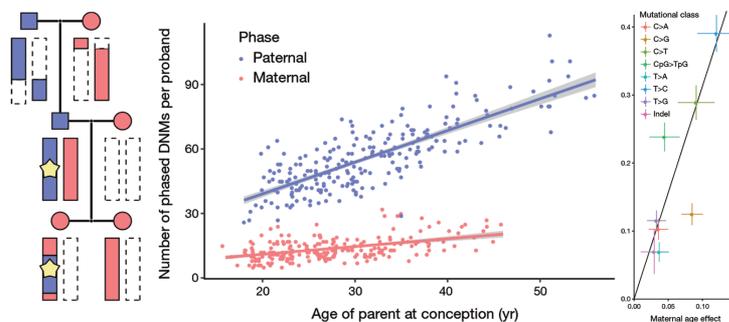


图2 左图: 判断某突变为新生突变的依据

中图: 后代基因组中新生突变数量与父/母生子时年龄的函数关系横轴

右图: 不同突变类型各自的父方(纵轴)/母方(横轴)遗传随年龄增长的变化率的对比

染色体上的聚类突变随母方生育年龄变化产生了更大的变化幅度

与父亲相比，母亲年龄的增长会让聚类新生突变产生更大的增幅。人类8号染色体短臂起始端一个大小为20Mb碱基的区域中，来自母亲的新生突变的出现密度是全部染色体中出现新生突变的平均密度的4.5倍。在C>G突变密集的染色体区域内，有56.5%的C>G新生突变簇出现来自母亲，相应来自父亲的仅为8.8%，印证了此种突变通常源于母亲一方遗传的说法。来自母亲的变异生成的基因簇片段的长度整体长于来自父亲的变异生成的基因簇。

★ 参考文献

Jónsson H, Sulem P, Kehr B, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland[J]. Nature, 2017, 549(7673):519.

案例二: DNBSEQ基因组测序揭示了肺鳞癌的潜在治疗靶标

★ 研究背景

文章通过外显子组重测序 (WES)，人全基因组重测序 (WGS)、靶区域捕获测序 (TS) 和CRISPR-Cas9基因组编辑技术，利用鳞状细胞肺癌手术肿瘤和对应的源自患者的异种移植瘤 (PDX) 样本探索和验证肺鳞癌的潜在治疗靶标。

文章亮点:

- Illumina HiSeq X Ten平台WES测序+ DNBSEQ平台WGS测序+ DNBSEQ平台TS测序，多平台数据联合分析
- LUSC PDX模型可广泛应用于潜在治疗目标和策略的验证
- 使用CRISPR系统对PDX肿瘤细胞中的驱动基因进行功能验证

通过基因组测序和CRISPR-Cas9基因组编辑的综合分析，在手术和PDX肿瘤上整合鉴定并验证了GRM8对LUSC肿瘤的促进功能。cAMP活化剂和MEK抑制剂可显著阻断具有GRM8突变的LUSC肿瘤细胞的增殖和存活。因此，GRM8信号传导通路的组成分子可能成为携带GRM8激活突变的鳞状细胞肺癌的治疗靶标。

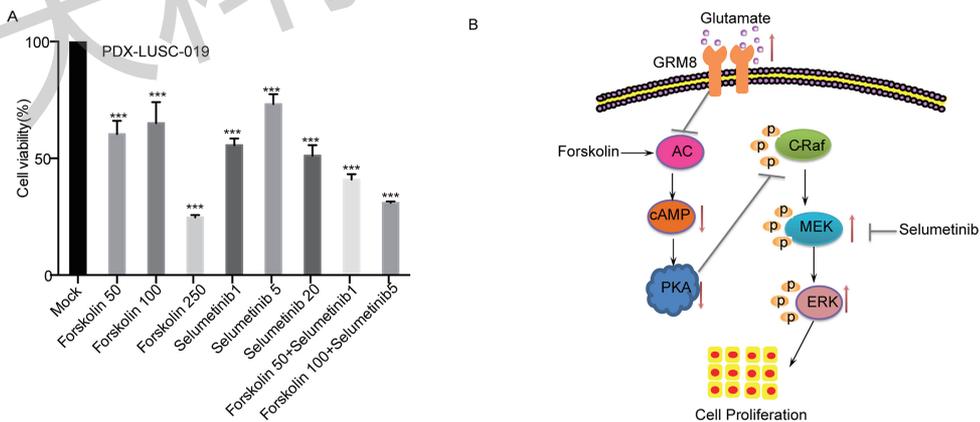


图3 cAMP活化剂和MEK抑制剂作为治疗策略

图A: Forskolin和Selumetinib的联合治疗对PDX肿瘤细胞活力具有抑制作用

图B: GRM8活化通过抑制cAMP通路和激活MAPK通路促进LUSC细胞的增殖

★ 参考文献

Zhang P, Kang B, Xie G, et al. Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer.[J]. Cancer Letters, 2019, 442:53-67.

▶▶ 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PCR-Free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PCR-Free测序则无法避免此类错误。因此，DNBSEQ PCR-Free 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PCR-Free数据。

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	文献名称
胰岛素瘤	2019	Gut	Whole-genome sequencing reveals distinct genetic bases for insulinomas and non-functional pancreatic neuroendocrine tumours: leading to a new classification system
stLFR方法	2019	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
肺鳞癌	2019	Cancer Research	Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer
参考人类基因组数据集	2017	GigaScience	A reference human genome dataset of the BGISEQ-500 sequencer
重度抑郁症	2015	Nature	Sparse whole-genome sequencing identifies two loci for major depressive disorder
癌症研究	2015	Human Genetics	Identification of cancer predisposition variants in apparently healthy individuals using a next-generation sequencing-based family genomics approach
神经母细胞瘤	2015	Nature	Telomerase activation by genomic rearrangements in high-risk neuroblastoma
宫颈癌	2015	Nature Genetics	Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism
老年性黄斑病变	2015	Nature Communications	New loci and coding variants confer risk for age-related macular degeneration in east Asians
单基因病	2013	Nature Genetics	Truncating mutations of MAGEL2 cause prader-willi phenotypes and autism
膀胱癌	2013	Nature Genetics	Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation
千人	2015	Nature	An integrated map of structural variation in 2,504 human genomes
	2015	Nature	A global reference for human genetic variation
	2012	Nature Methods	The 1,000 genomes project: data management and community access
澳洲土著人	2011	Science	An aboriginal australian genome reveals separate human dispersals into asia
千人	2010	Nature	A map of human genome variation from population-scale sequencing
古人	2010	Nature	Ancient human genome sequence of an extinct palaeo-eskimo
炎黄一号	2008	Nature	The diploid genome sequence of an asian individual
人类基因组计划	2001	Nature	Initial sequencing and analysis of the human genome

外显子组测序

产品概述

全外显子组测序 (Whole Exome Sequencing, WES) 是利用序列捕获技术, 将全基因组全部的外显子区域的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15万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
基于WES的GWAS可将关联变异位点锁定在编码区内, 为致病机制的阐明、疾病的治疗提供更丰富的线索; 是肿瘤精准免疫治疗必检项目, 可作为一线诊断工具



研究广泛
涉及60多种肿瘤类型, 100余种复杂疾病; 罕见病研究性文章超过150多篇

产品应用



技术流程

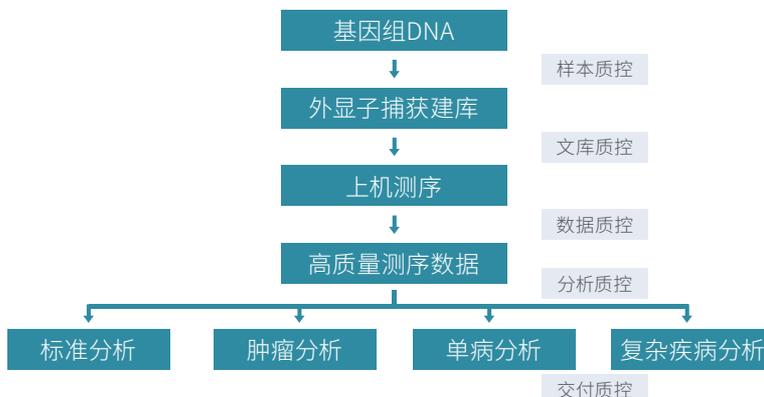


图1 外显子组测序技术流程

► 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	≥0.5g
全血	≥2ml
福尔马林固定石蜡包埋组织(FFPE)	>10个玻片或50mm ² , 5-10μm厚的切片10片
常规基因组DNA	m≥1μg, c≥12.5ng/μL
微量基因组DNA	>500ng, c≥2.5ng/μL
FFPE基因组DNA	m≥1μg, c≥10ng/μL
	m≥0.5μg, c≥2.5ng/μL, 风险送样

测序	推荐数据量	项目周期
PE100、PE150等	100X、200X、500X以上 (cfDNA)	短至11天

► 案例分析

案例: 发现与糖尿病相关的罕见DNA突变

★ 研究背景

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。

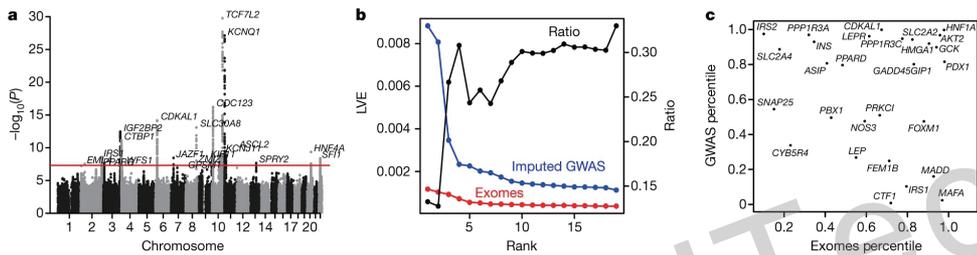


图2 外显子组测序与基于阵列的GWAS的比较

★ 参考文献

Flannick J, Mercader JM, Fuchsberger C, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls[J]. Nature, 2019, 570(7759):71-76.

► 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200×以上。如果只看经典SNP、非低频突变, 测序深度也至少应该在50×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:

$$\text{测序深度} = \frac{\text{比对到基因组上的有效数据 (去除DUP)} \times 10G \times \text{捕获效率} 60\%}{\text{目标区域大小} 60\text{Mb}} = 100X$$

► 华大合作发表文章 (部分)

(1) 自主平台DNBSEQ发表文章

测序平台	研究内容	年份	发表期刊	文献名称
DNBSEQ™	鳞状细胞肺癌	2019	Cancer Letters	Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer
	肿瘤细胞	2019	Cell Chemical Biology	Engineering forward genetics into cultured cancer cells for chemical target Identification
	非小细胞肺癌	2019	Journal for ImmunoTherapy of Cancer	Epigenetic alterations are associated with tumor mutation burden in non-small cell lung cancer
	骨髓增生性肿瘤	2019	Blood	Acquired resistance to JAK inhibitors in calr-mutated myeloproliferative neoplasms

(2) 华大历史经典文章

种类	研究内容	发表时间	发表期刊	文献名称
癌症	食管小细胞癌	2018	Cell Research	The genomic landscape of small cell carcinoma of the esophagus
	前列腺癌	2017	European Urology	Stromal gene expression is predictive for metastatic primary prostate cancer
	肺癌	2016	Nature Communications	Genomic heterogeneity of multiple synchronous lung cancer
	食管癌	2014	Nature	Identification of genomic alterations in oesophageal squamous cell cancer
	食管鳞状细胞癌	2014	Nature	Identification of genomic alterations in oesophageal squamous cell cancer
	肾上腺库欣综合征	2014	Science	Activating hotspot L205R mutation in PRKACA and adrenal cushing's syndrome
罕见病 / 遗传性疾病	早发性肌张力障碍	2016	Nature Genetics	Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia
	肢端点状角化病、唇炎、指节垫	2015	American Journal of Human Genetics	Loss-of-Function mutations in CAST cause peeling skin, leukonychia, acral punctate keratoses, cheilitis, and knuckle pads
	汗孔角化	2012	Nature Genetics	Exome sequencing identifies MVK mutations in disseminated superficial actinic porokeratosis
	先天性黑朦	2012	Nature Genetics	Exome sequencing identifies NMNAT1 mutations as a cause of Leber congenital amaurosis
	脊髓小脑共济失调	2010	Brain	TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing
复杂疾病	乙型肝炎病毒	2018	Blood	Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma
	银屑病	2014	Nature Genetics	A large-scale screen for coding variants predisposing to psoriasis
	中丹糖尿病三	2013	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
	糖尿病	2010	Nature Genetics	Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants

细菌de novo测序

▶▶ 产品概述

细菌基因组de novo测序是对细菌基因组测序后从头组装，并在组装的基础上，进行基因组组分、功能注释等分析；最终的组装水平根据研究的需要和细菌本身的特点而定，依据组装水平分为初级组装、高级组装和完成图。其中最高指标是组装成一条Contig，即基因组的完整序列。

▶▶ 产品优势



提取质量好
有丰富的长片段提取经验，提取DNA质量高



平台联合组装，提高准确度
短读长测序和单分子测序联合组装、纠错，打造“极致完成图”



组装指标高
极致完成图 ‘0’ gap, 1 Contig, 准确还原基因组原始状态（是否成环）



附加值高
结题报告文章化，深入挖掘基因组信息，并提供内容详细的结题报告解读视频



经验丰富
已完成上万个细菌de novo项目，发表细菌基因组文章170多篇

▶▶ 产品应用



▶▶ 技术流程

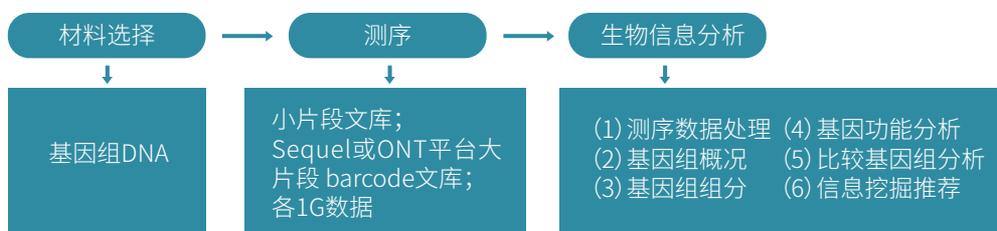


图1 细菌de novo测序流程图

信息分析内容

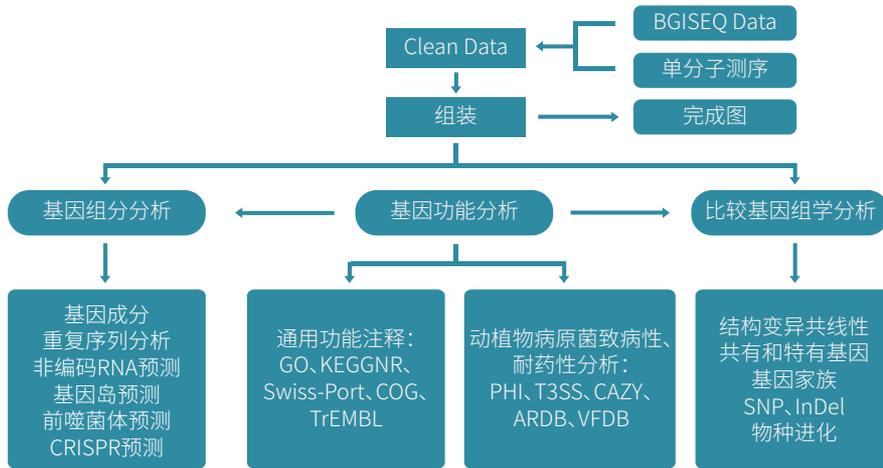


图2 细菌de novo测序信息分析内容

技术参数

样品要求: 基因组DNA, 无降解或轻微降解

小片段文库: 样品量 $m \geq 1\mu\text{g}$; 浓度 $c \geq 12.5\text{ng}/\mu\text{L}$

大片段文库 (Sequel/ONT): 样品量 $m \geq 5\mu\text{g}$; 浓度 $c \geq 60\text{ng}/\mu\text{L}$

OD260/280: 1.6-2.2; OD260/230: 1.6-2.2

推荐数据量:

测序策略	初级组装	完成图
小片段文库 (PE150)	1G	1G
大片段文库 (Sequel/ONT)	—	1G

项目周期: 初级组装35个工作日; 完成图45个工作日

案例解析

案例: 采用长读长Nanopore技术对食源性病原体进行快速、高通量的基因组和质粒测序

★ 研究背景

本研究使用Nanopore的MinION测序仪对沙门氏菌 (*Salmonella enterica* subsp. *enterica* serovar Bareilly) 和大肠杆菌 (*Escherichia coli* O157:H7) 进行测序, MinION平均读长达到8K以上, 平均质量值达到将近Q20的水平; 测序4小时即可组装得到完整的细菌基因组, 沙门氏菌和大肠杆菌组装结果与MiSeq平台的一致性分别为99.87%和99.89%。这些长读取组装结果提供了关于血清型, 毒力因子和抗微生物抗性基因的信息。在分析流程中增加polish步骤可进一步提高组装的结果的准确度。

Table 2: Assembly Data for MinION sequencing of *Salmonella*

Duration (min)	Reads	Subsampled Reads	Assembly Size	Circular Contigs ^a	Linear Contigs	Longest Contig	Longest Circular Contig	NG50 ^b	Average Identity in %	Reference Coverage
15	7229	7229	1135723	0	19	163786	0	0	99.13	24.36
30	14888	14888	4577215	0	18	841969	0	471499	99.55	95.38
60	29132	29132	4722179	1	0	4722179	4722179	4722179	99.79	98.4
120	51226	51226	4805334	2	0	4723663	4723663	4723663	99.84	100
240	84156	28492	4806150	2	0	4724389	4724389	4724389	99.87	100
480	132137	20193	4806518	2	0	4724724	4724724	4724724	99.87	100
960	248910	16221	4806892	2	0	4725103	4725103	4725103	99.89	100
1500	333298	15249	4806995	2	0	4725191	4725191	4725191	99.89	100

a. Two circular contigs indicates both the chromosome and the plasmid
b. NG50 - 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value

Table 3: Assembly Data for MinION sequencing of *E. coli*

Duration (min)	Reads	Subsampled Reads	Assembly Size	Circular Contigs ^a	Linear Contigs	Longest Contig	Longest Circular Contig	NG50 ^b	Average Identity in %
15	8731	8731	1352560	0	19	154626	0	0	99.18
30	18053	18053	5141583	0	14	155572	0	518318	99.63
60	35335	35335	5481126	1	0	5481126	5481126	5481126	99.82
120	62415	60362	5570410	1	1	5481662	5481662	5481662	99.87
240	103681	19589	5577045	2	0	5482542	5482542	5482542	99.89
480	164641	15265	5577346	2	0	5482831	5482831	5482831	99.90
960	317698	12941	5577818	2	0	5483284	5483284	5483284	99.91
1500	429909	12403	5577934	2	0	5483397	5483397	5483397	99.91

a. Two circular contigs indicates both the chromosome and the plasmid
b. NG50 - 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value

图3 MinION不同测序时间得到的数据组装结果评估。

★ 参考文献

Taylor T L, Volkening J D, DeJesus E, et al. Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology[J]. BioRxiv, 2019: 558718.

► 常见问题

Q: 单分子测序对纯度和样本完整性要求较高, 如果样本无法完全满足要求, 应该怎么处理?

A: 对于无法满足要求的样本处理建议如下:

①对于基因组已经降解的样本建议重新提取, 若样本珍贵, 可依据不同的降解程度做相应处理, 但仍可能存在文库及测序数据质量不够好的问题; 若普通跑胶后主带大于10Kb也可建库, 但建库及测序质量风险需客户自行承担; 降解的样本可用低倍数的磁珠选择大片段或是用blue pippin进行size selection, 并依据不同的降解程度相应提高样品起始量, 但由于样本的质量情况不同, 不能保证所有的小片段均会去除, 因此建库及测序质量存在一定风险需客户知悉。②对于260/280和260/230 比值不在正常范围内的及Qubit-HS/Nanodrop<0.5的样本, 可用磁珠或柱纯化后再检测。

Q: 细菌完成图可否只选择单分子测序?

A: 细菌完成图如果单从组装完整度上考虑, 一般只用单分子测序即可组装到完成图水平。为了保证组装结果的准确性, 我们推荐联合短读长测序进行纠错。

Q: 细菌完成图测序能否得到完整的质粒结果?

A: 细菌完成图组装不能保证质粒的完整性, 如果特别关注质粒信息的话, 可单独提取质粒进行测序。

Q: ONT细菌完成图和PB细菌完成图应该怎么选择?

A: 华大基因ONT细菌完成图采用DNBSEQ和ONT数据联合组装的策略, 可得到高质量的细菌完成图, 能满足绝大多数细菌完成图研究需求。建议将ONT细菌完成图作为大样本量细菌完成图以及非极端高GC细菌完成图首选策略。对于基因组较大的极端高GC细菌(基因组大约6Mb, GC含量大于65%)可根据具体需求选择: ①选择BGISEQ+ONT联合组装的策略, 准确度高, 高GC区域组装可能受影响, 有得不到完成图的风险; ②用ONT数据组装, 短读长数据纠错, 可以得到完成图, 但得到的基因组准确度可能偏低; ③选择PB细菌完成图, 可得到高质量的细菌完成图, 成本相对较高。

► 华大合作发表文章(部分)

物种	发表时间	发表期刊	文献名称
绿藻	2019	Scientific Reports	Sequencing and comparative analysis of three chlorella genomes provide insights into strain-specific adaptation to wastewater
镰刀菌	2019	Microorganisms	Fugacium spliced leader genes Identified from stranded RNA-seq datasets
产丁酸菌	2019	Microorganisms	Characterization a novel butyric acid-producing bacterium collinsella aerofaciens subsp. shenzhenensis subsp. nov.
紫色硫细菌	2018	Marine Drugs	First evidence of dehydroabiatic acid production by a marine phototrophic gammaproteobacterium, the purple sulfur bacterium allochromatium vinosum MT86
乳杆菌科/ 胆汁盐 水解酶 (BSH)	2018	Molecules	A comprehensive genome survey provides novel insights into bile salt hydrolase (BSH) in lactobacillaceae
放线菌	2018	International Journal of Systematic and Evolutionary Microbiology	Amnibacterium endophyticum sp. nov., an endophytic actinobacterium isolated from aegiceras corniculatum



真菌de novo测序

▶▶ 产品概述

新一代高通量测序技术，以其高通量，高准确率，周期短的特点，为真菌的研究提供强有力的支撑，可用于预测真菌的重要基因和蛋白，以了解其功能和可能机制。在研究植物致病性真菌方面，可以鉴定致病真菌与农作物互作及致病的相关基因，并可用于研究与其近缘物种的进化关系等；在食用真菌和药用真菌的研究方面，可以用于发现真菌复杂的代谢途径，鉴定其对人有利的代谢产物，及物种进化关系的研究等；另外还可用于生物防治真菌的研究，工业酵母菌的研究等。

▶▶ 产品优势



多平台

多种测序平台可选，满足客户多样化需求



准确度高

短读长数据纠错，组装结果准确度高



提取质量好

有丰富的大片段提取经验，提取DNA质量高



项目经验丰富

已完成数千个真菌de novo项目，发表真菌基因组文章50多篇

▶▶ 产品应用

食用真菌营养
分析及改良

致病真菌毒性
及机理研究

生物防治真菌
功能挖掘

▶▶ 技术参数

样品要求: 基因组DNA，无降解或轻微降解

小片段文库: 样品量 $m \geq 1\mu\text{g}$; 浓度 $c \geq 12.5\text{ng}/\mu\text{L}$

大片段文库 (Sequel/ONT): 大片段文库，样品量 $m \geq 5\mu\text{g}$; 浓度 $c \geq 60\text{ng}/\mu\text{L}$;

OD260/280: 1.6-2.2; OD260/230: 1.6-2.2

推荐数据量:

	初级组装	高级组装
小片段文库 (PE150)	$\geq 50\text{X}$	$\geq 50\text{X}$
大片段文库 (Sequel/ONT)	—	150X-200X

▶ 案例分析

案例: 大丽轮枝菌棉花致病机理研究

★ 研究背景

大丽轮枝菌 *Verticillium dahliae* 是一种广泛危害农作物的半知菌亚门轮枝菌属真菌, 寄主多达660种, 已成为一个世界性难题。不同的大丽轮枝菌有不同的原始宿主, 如 *V. dahliae* Vd991 以棉花为原始寄主, 它的两个姐妹菌株 *V. dahliae* JR2 和 *V. dahliae* VdLs.17 原始寄主分别是番茄和莴苣。

一般来说, 大丽轮枝菌对原始寄主的适应性和毒力最高, 对其他物种适应性和毒力相对较弱。不同菌株对宿主的适应性及毒力差异主要受染色体重组和种系特异性 (lineage-specific, LS) 基因区域影响。通过真菌 *de novo* 测序和比较基因组分析, 发现 *V. dahliae* Vd991 的可变基因区存在棉花特异性的毒力因子, 而一些棉花毒力相关基因很可能是从镰刀菌属 (*Fusarium*) 水平基因转移获得的。对目标菌株进行基因改造并进行毒力检测, 结果发现, Vd991 对棉花的毒力与其特有的7个 G-LSR2 基因密切相关。

★ 研究方法

1. Vd991 真菌 *de novo* 测序及相关分析

测序策略: PacBio RSII + MiSeq PE250+10Kb mate-paired 文库测序

基因组分析: 基因预测、重复序列、非编码 RNA

基因功能注释: nr、eggNOGs、INTERPROSCAN (incorporating InterPro, GO and KEGG pathway annotation)

毒力因子预测: CAZy、PHI、蛋白激酶 (PKs) 分析

比较基因组分析: 基因重排分析及 PCR 验证、基因家族分析、进化分析

2. 三株 *V. dahliae* 真菌 RNA-Seq 测序及相关分析

建库: TruSeq™ RNA Sample Preparation Kit

测序: HiSeq 2000

分析: 基因表达差异分析、KEGG pathway 分析

3. 真菌基因敲除、遗传转化实验及菌株毒力实验

★ 主要结果

1. LS 基因可能是大丽轮枝菌对寄主适应性和致病性差异产生的原因。
2. Vd991 的某些特异性基因区可能是从棉花枯萎病菌 *Fusarium* 基因水平转移获得的。

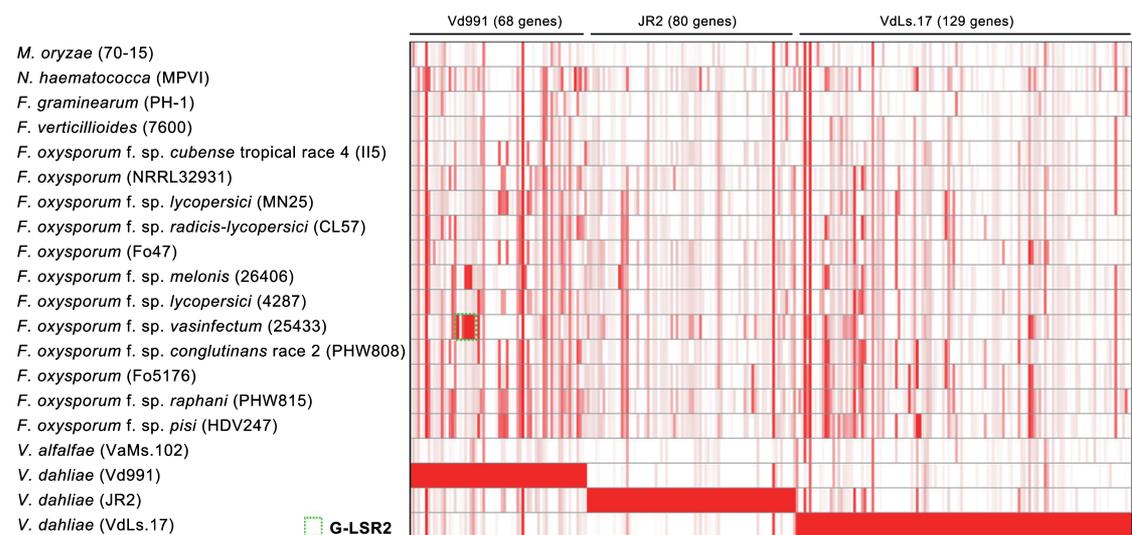


图 *Verticillium dahliae* 的 LS 基因和镰刀菌属基因同源性分析。
Vd991 的 7 个 G-LSR2 基因与 *F. oxysporum* f. sp. *vasinfectum* NRRL 25433 高度同源

3. 7 个 Vd991 菌株特异性 G-LSR2 基因与 *V. dahliae* 在棉花中的毒力高度相关。

★ 参考文献

Chen JY, Liu C, Gui YJ, et al. Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in *Verticillium dahliae* and evidence of horizontal gene transfer from *Fusarium*[J]. *New Phytologist*, 2018, 217(2): 756-770.

▶▶ 常见问题

Q: 真菌样品有什么要求? 客户提取真菌样品之后, 是否有必要做一下16S检测?

A: 建议送单倍体, 并建议送样之前做一下16S检测, 或者TA克隆, 这样可以避免细菌污染。

Q: 对于一些寄生、不能单独培养的真菌, 能做de novo测序吗?

A: 如果寄主的物种序列已知, 数据分析过程中可以将属于寄主的序列过滤掉, 再进行组装; 但在寄主的物种序列未知的情况下, 难度较高。

Q: 真菌的DNA样品准备有什么建议?

A: 选取无性生殖阶段、营养体的菌丝, 避免在生成高等复杂结构的时期提取样品。如果实验条件允许, 最好挑取单个孢子培养萌发大量菌丝, 然后利用这些由同一孢子萌发的菌丝进行样品制备。

Q: 真菌初级组装, 高级组装的区别是什么?

A: 初级组装一般用于对真菌基因组大小、GC含量高低、重复序列多少、杂合度大小以及是否受到污染做一个初步的评估。测序中推荐构建小片段文库, 测序数据量50X以上; 高级组装采用大片段测序和小片段测序联合的策略, 可以组装得到比较完整的基因组序列, 用于真菌基因和功能的挖掘。高级组装推荐同时构建大片段文库 (PacBio或ONT平台) 和小片段文库, 大片段文库测序150X-200X, 小片段文库测序50X以上。

▶▶ 华大合作发表文章 (部分)

物种	发表时间	发表期刊	文献名称
担子菌	2019	Genes (Basel)	Genomic evidence of recombination in the basidiomycete <i>walleimia mellicola</i>
刺柏 (寄生真菌)	2019	Genes (Basel)	Genome sequencing of <i>cladobotryum protrusum</i> provides insights into the evolution and pathogenic mechanisms of the cobweb disease pathogen on cultivated mushroom
酵母	2018	Nature Communications	Identifying and characterizing SCRaMbLEd synthetic yeast using ReSCuES
酵母	2018	Nature Communications	Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods
大丽轮枝菌	2017	New Phytologist	Comparative genomics reveals cotton-specific virulence factors in flexible genomic regions in <i>Verticillium dahliae</i> and evidence of horizontal gene transfer from <i>Fusarium</i>
蚂蚁、真菌	2016	Nature Communications	Reciprocal genomic evolution in the ant-fungus agricultural symbiosis
杏鲍菇亚种 <i>tuoliensis</i> (白灵菇)	2016	Molecules (Basel, Switzerland)	De novo sequencing and transcriptome analysis of <i>pleurotus eryngii</i> subsp. <i>tuoliensis</i> (<i>bailinggu</i>) mycelia in response to cold stimulation



细菌/真菌重测序

产品概述

基因组重测序是对已知的基因组进行测序，并对个体或者群体样品进行分析的研究。细菌/真菌重测序能够分析近源菌种之间的单核苷酸多态性 (SNP)、插入 (Insertion)、缺失 (Deletion) 等基因组变异类型，可以为筛查菌株的优良性状，菌株抗药性等研究提供指导与依据。

产品优势



性价比高
自主测序平台，成本可控



测序准确性高
DNBSEQ平台滚环扩增构建DNB测序文库，PCR扩增错误不会累积，高保真序列信息



Duplication率低
DNBSEQ平台Duplication率低，同样的数据量有效数据多出3%-17%



无index hopping担忧
DNBSEQ平台无index hopping担忧，结果更可靠



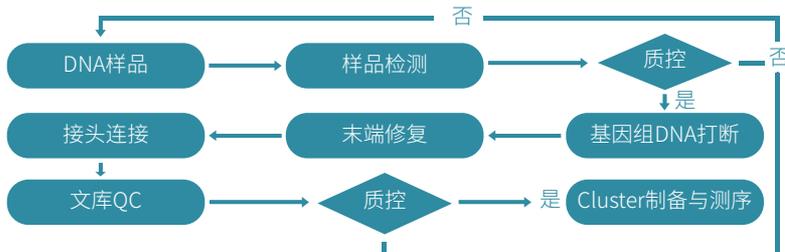
附加值高
数据挖掘推荐，数据交付6个月内，提供免费的项目咨询服务

技术流程

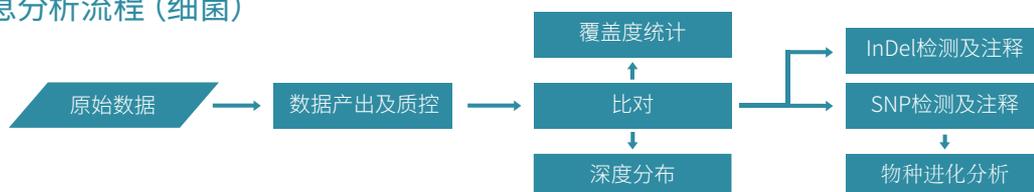
研究流程



实验流程



信息分析流程 (细菌)



► 研究内容 (细菌)

数据处理

1. 去除接头污染和低质量数据
2. 产出数据及质控统计

标准信息分析

1. 细菌初级组装及评价 (K-mer分析、GC-Depth分析)
2. 比较基因组分析 (参考基因组同源性比较、SNP/InDel、物种进化分析)

定制化分析

可结合客户的需求, 协商确定信息分析内容

► 技术参数

样品要求:

样品类型: 基因组DNA, 无降解或轻微降解

样品总量: $m \geq 1 \mu\text{g}$

样品浓度: $c \geq 12.5 \text{ ng}/\mu\text{L}$

测序策略: PE150

推荐数据量: 细菌100X以上 (或1Gb); 真菌50X以上

项目周期: 30个工作日

► 案例解析

案例: 中国地区161株结核杆菌分离株的群体进化和耐药分析

★ 研究背景

全球范围内多药耐药结核菌及交叉耐药结核菌的出现为结核病的治疗带来了严重的挑战, 结核菌耐药机制尚有待进一步研究。本研究对中国10省市161株广泛耐药性结核分枝杆菌样品进行全基因组测序, 其中44株为药物敏感株, 94株为多重耐药株, 23株为极端耐药株; 并加入22株已发表菌株序列。筛选检测到18,970个SNP位点, 这些位点在95%的样品中存在; 通过筛选研究发现72个新的基因, 11个非同义SNP及10个间区SNP与耐药性显著相关; 通过dn/ds选择压力分析表明, 耐药基因对耐药菌群具有很高的选择性。

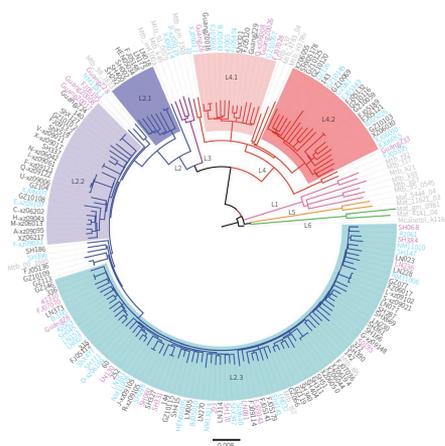
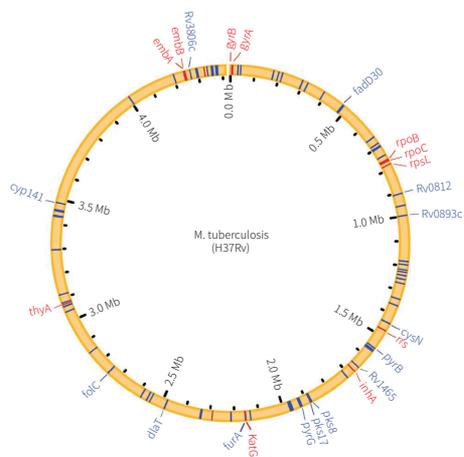


图1 系统进化树分析



华大科技BGITech

产品
简介

Product
Profile

| 基因科技 造福人类
Omics for All

华大科技BGITech

转录组学

转录组测序	039
全长转录组测序	044
RNA-Seq	049
Small RNA测序	053
长链非编码RNA测序	057



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种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，已广泛应用于物种遗传发育、改良，分子育种研究，以及癌症、复杂疾病及感染类疾病研究和药物研发等领域。

▶▶ 产品优势



全面丰富的样品类型

常规建库仅需200ng起，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，Science、Nature、Cell等1,000+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，即可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

样本提取只需2天，从样本检测、建库测序到Dr. Tom交付只需14个工作日，节约您的宝贵时间

▶▶ 产品应用



▶▶ 技术参数

样品要求：

样品类型：完整且无污染的总RNA

样品需求量（单次）：鼠200ng，其他物种1μg

样品浓度：人、鼠10-1,000ng/μL，其他物种40-2,500ng/μL

样品纯度：28S/18S≥1.0；人和其他动物RIN≥7.0，植物RIN≥6.0，真菌RIN≥6.5；昆虫无28S/18S及RIN值要求

测序策略：PE100/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）

► 技术流程

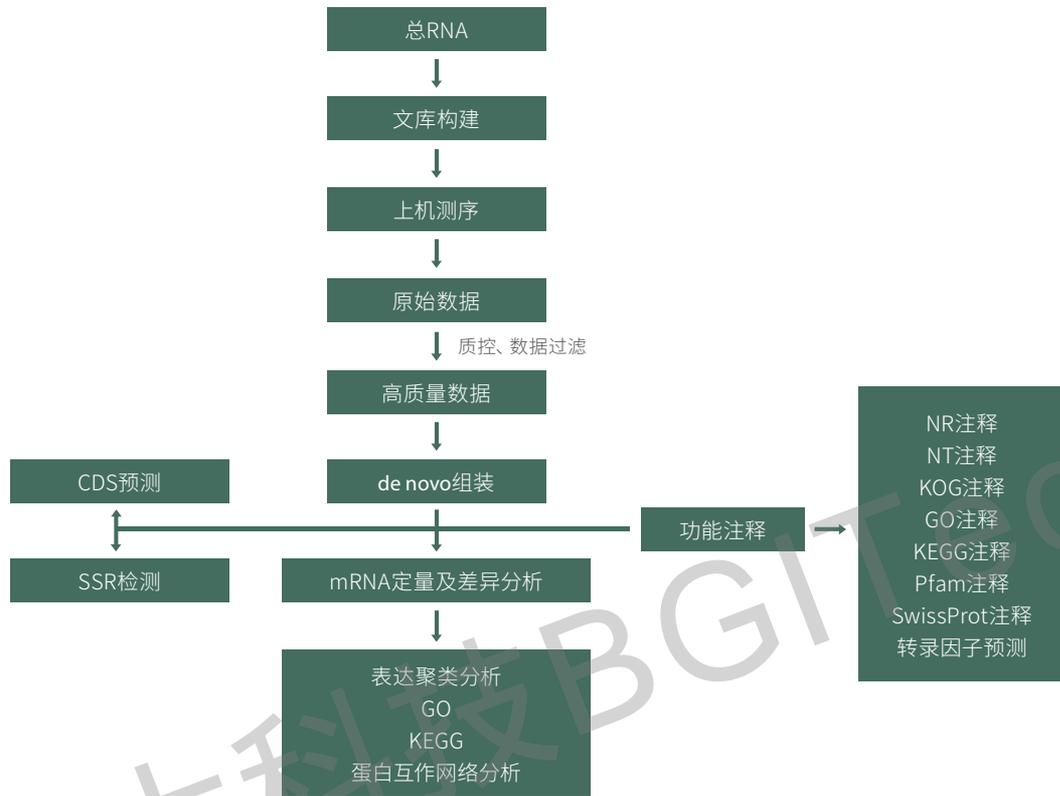


图1 无参考基因组序列的转录组测序技术流程

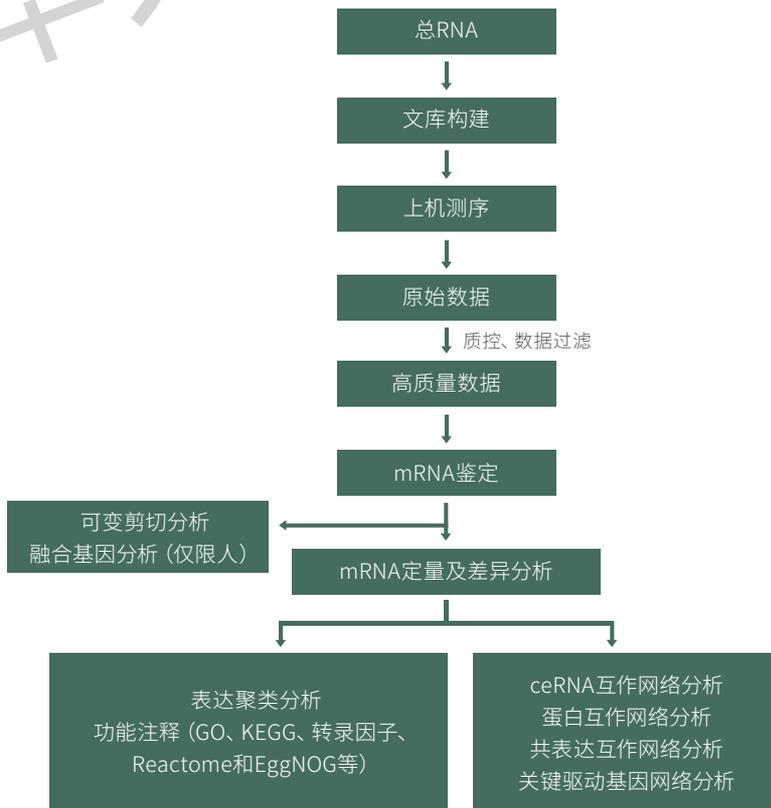


图2 有参考基因组序列的转录组测序技术流程

►► 研究内容

无参考基因组序列的转录组

标准信息分析

1. 测序数据过滤
2. 转录组 *de novo* 组装
3. Unigene 七大功能数据库注释
4. Unigene 的 CDS 预测
5. Unigene 的 TF 编码能力预测
6. Unigene 的 SSR 检测
7. Unigene 表达量计算 (基因表达水平、PCA 分析、基因表达水平箱线图和密度图、样品间的表达量韦恩图)
8. 时间序列分析
9. 差异表达基因检测
10. 差异表达基因聚类分析
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. 差异基因蛋白互作分析
14. 真菌致病基因预测 (真菌样本)
15. 植物抗病基因预测 (植物样本)

有参考基因组序列的转录组

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. Rfam、Pfam、Reactome、COG、EggNOG 和 InterPro 数据库注释
3. GSEA 分析

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 (仅限人)、ARCS4 (仅限鼠)
3. 卡方检验
4. 关键驱动基因网络分析
5. 时间序列分析

►► 案例分析

案例一: 水稻大规模转录组测序和蛋白质组数据集研究

本研究目的是通过转录组测序和蛋白质组数据, 提高水稻基因组注释。建立了一个全面的转录本数据库, 来自于 29 个公开的 RNA 测序数据集, 也重新分析了 9 个公开的水稻蛋白质组数据集, 共计鉴定到 420K 多肽谱。经过严格的筛选, 发现了 1,584 个高置信度的新多肽。这些新的多肽被聚类到 692 个基因组位点, 使注释得到改进。80% 的新肽在至少一种其他植物物种的蛋白质序列中具有同源性。对于在基因间区域聚集的多肽 (可能是潜在的新基因), 共鉴定了 101 个位点, 其中 43 个位点有高置信度的蛋白域。这些结果都可以获取, 帮助其他研究者重新注释水稻基因组。

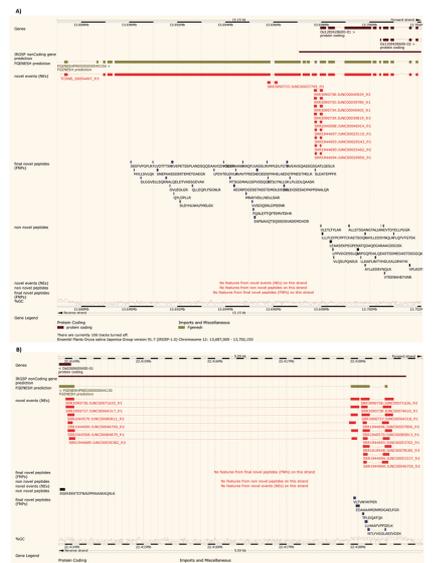


图3 发现的两个典型可变剪切事件 (NEs) 的基因组图谱

★ 参考文献

Zhe R, Da Q, Pugh N, et al. Improvements to the rice genome annotation through large-scale analysis of RNA-Seq and proteomics datasets[J]. *Molecular and Cellular Proteomics*, 2019, 18 (1), 86-98.

案例二：转录组研究乙型肝炎病毒感染增加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。

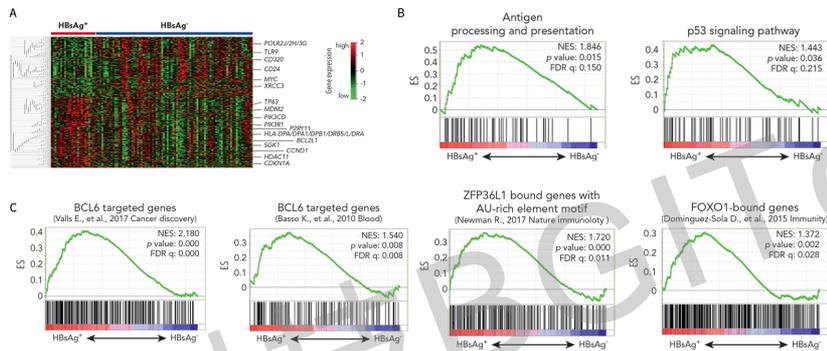


图4 转录组测序显示感染者具有独特的基因表达图谱

★ 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十万倍, 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2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DNBSEQ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	文献名称
医学研究			
小鼠	2019	Molecular Cancer	Hypoxic BMSC-derived exosomal miRNAs promote metastasis of lung cancer cells via STAT3-induced EMT
人细胞系	2019	Nucleic Acids Research	Diminished OPA1 expression and impaired mitochondrial morphology and homeostasis in Aprataxin-deficient cells
人细胞系	2019	Nature Communications	Precisely controlling endogenous protein dosage in hPSCs and derivatives to model FOXG1 syndrome
斑马鱼	2019	Chemosphere	Microcystin-LR exposure induced nephrotoxicity by triggering apoptosis in female zebrafish
人	2019	Genomics & Informatics	Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing
人胃肠道间质瘤	2019	PNAS	Mutational inactivation of mTORC1 repressor gene DEPDC5 in human gastrointestinal stromal tumors
人B细胞淋巴瘤	2018	Blood	Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma
小鼠	2018	Annals of the Rheumatic Diseases	Glucocorticoid receptor in stromal cells is essential for glucocorticoid-mediated suppression of inflammation in arthritis
动植物研究			
雨生红球藻	2019	Biotechnol Biofuels	Comparative transcriptome analysis of a long-time span two-step culture process reveals a potential mechanism for astaxanthin and biomass hyper-accumulation in Haematococcus pluvialis JNU35
拟南芥	2019	Environmental Science & Technology	Comparative physiological and transcriptomic analyses reveal the toxic effects of ZnO nanoparticles on plant growth
攀鲈	2019	Science of the Total Environment	Transcriptomic analysis of Anabas testudineus and its defensive mechanisms in response to persistent organic pollutants exposure
反刍动物	2019	Science	Genetic basis of ruminant headgear and rapid antler regeneration
千种植物	2019	Nature	One thousand plant transcriptomes and the phylogenomics of green plants
虎蛾	2018	Molecular Phylogenetics and Evolution	A phylogenomic analysis of lichen-feeding tiger moths uncovers evolutionary origins of host chemical sequestration
昆虫病毒	2018	Scientific Reports	Characterization of viral RNA splicing using whole-transcriptome datasets from host species



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/完善参考物种基因集。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	构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研究 基因/转录本差异研究	同样的数据量，可得到5倍的有效数据 检测到的基因/转录本数目多 基因/转录本定量结果准 其他优势同“常规全长转录组”
PolyA 全长转录组	PolyA相关研究 构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研究	可检测PolyA长度分布 可检测PolyA尾中非A碱基类型及出现频率 研究揭示PolyA在表达调控中的作用 其他优势同“常规全长转录组”
常规全长 转录组	构建参考基因集 基因功能研究 辅助基因组注释 转录本结构变异研究	不需要组装，就可以准确的得到全长转录本的序列信息 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 辅助基因组注释

技术流程

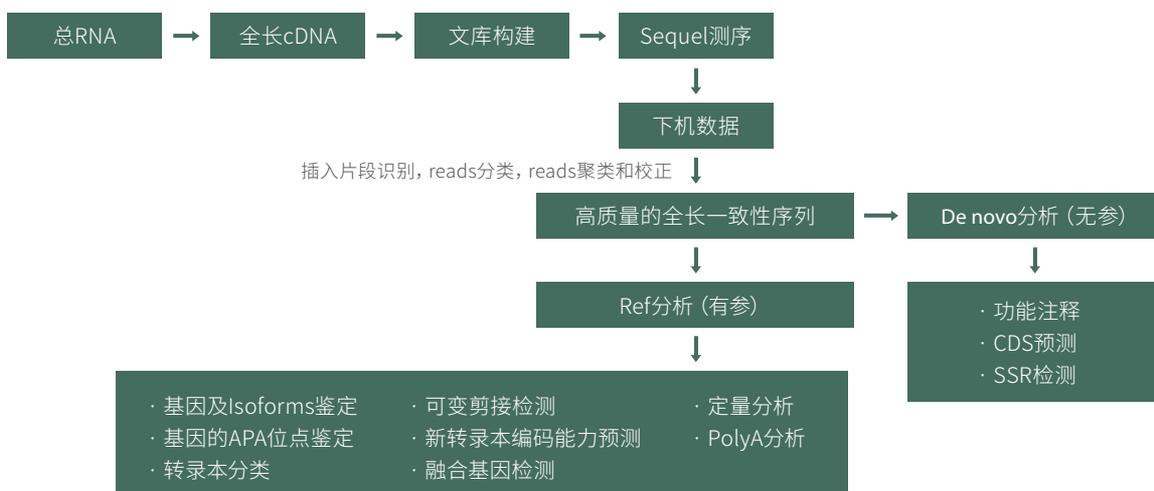


图1 全长转录组转录组测序技术流程

►► 信息分析内容

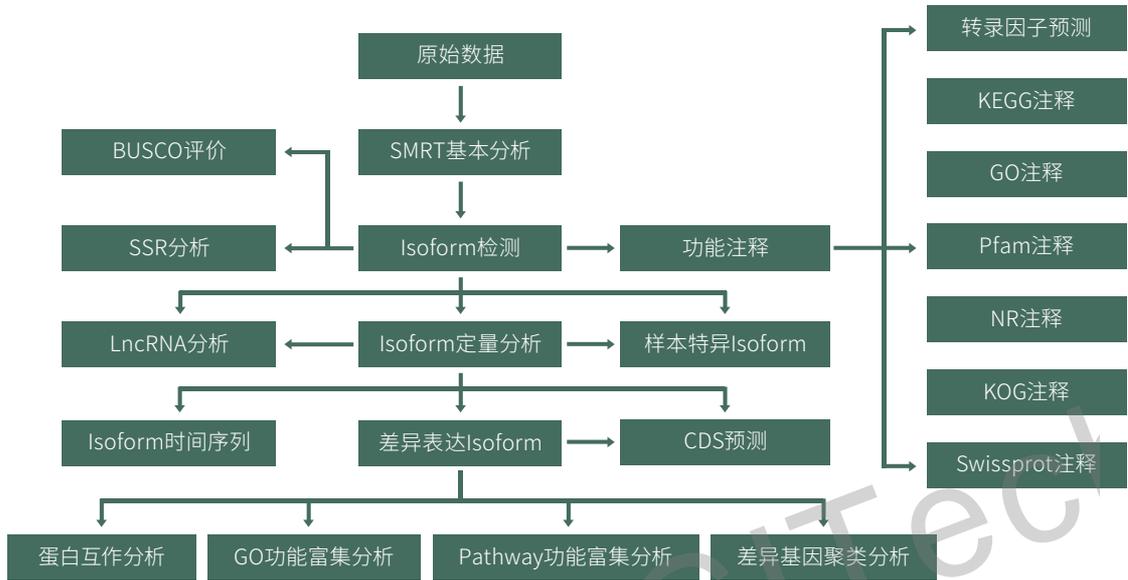


图2 Iso-Seq de novo标准分析条款

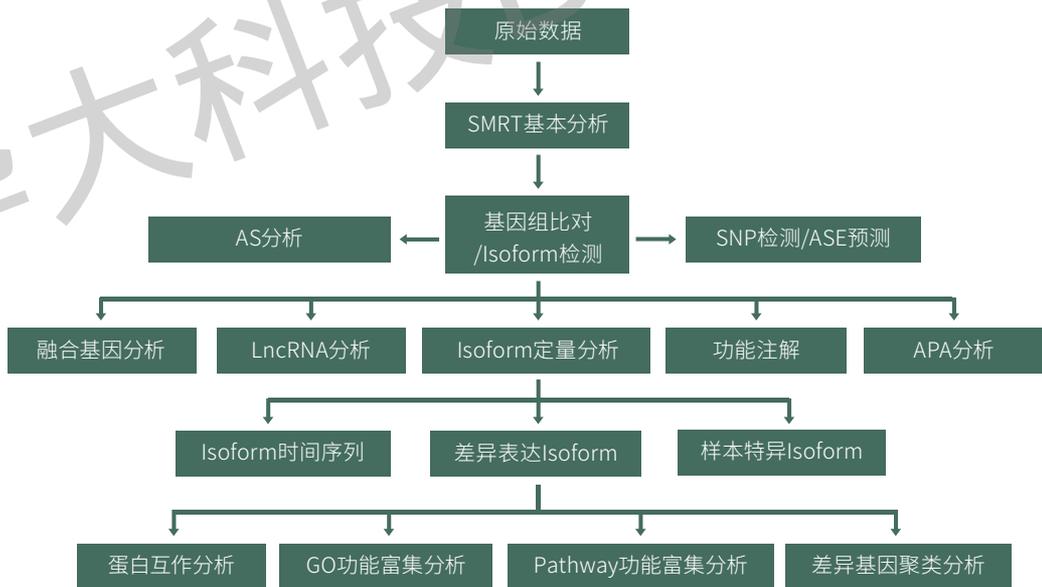


图3 ISO-Seq ref标准分析条款

►► 技术参数

样品要求:

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: $\text{RIN} \geq 8.0$, $28\text{S}/18\text{S} \geq 1.4$

推荐数据量: 测序1个cell或多个cell (sequel平台); 如果用sequel II平台, 可能需要多个样本pooling

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。

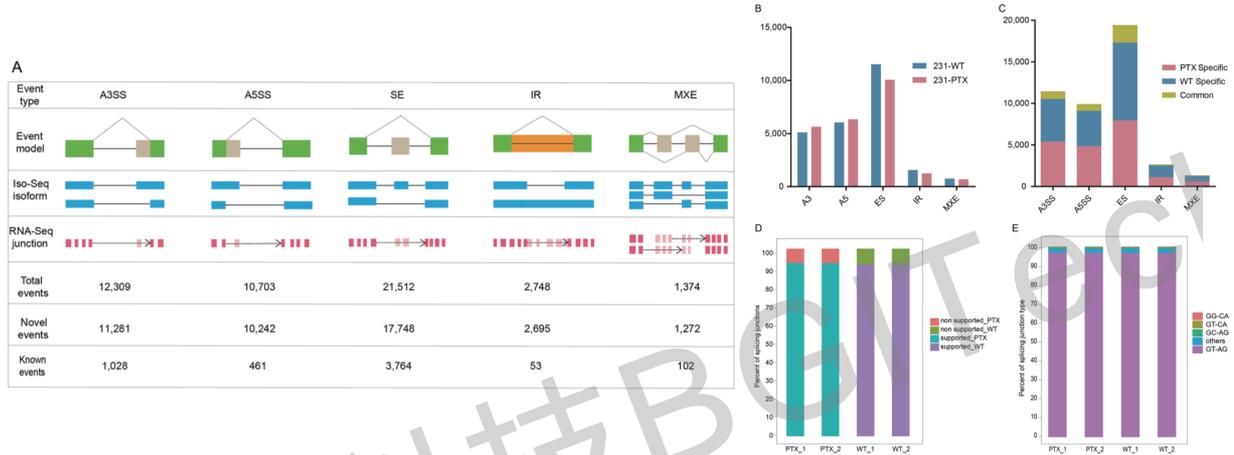


图4 乳腺癌细胞野生型 (231-WT) 和紫杉醇抗性型 (231-PTX) 可变剪接比较

★ 参考文献

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

案例二：玉米和高粱全长转录组测序

★ 方案设计

1. 材料选择

· 玉米 (B73)：根、茎、叶、幼苗 (14日龄)，雌穗 (ears, v8期)，雄穗 (tassels, v7期)，花粉 (r1期)、胚、胚乳和果皮 (授粉后20天从种子获得)，须 (silk, r1期)，苞片 (最内层)。

· 高粱 (BTx623)：根、茎、叶、幼苗 (组织萌芽后14天)，胚，胚乳和果皮 (授粉后20天)，花粉 (9-10周阶段)，花序-1 (1-5mm)，花序-2 (5-10mm)，花序-3 (1-2cm)

2. 测序策略

· PacBio测序

建库：Barcode建库；文库大小：<1 Kb、1-2 Kb、2-3 Kb、3-5 Kb、>5 Kb 文库

测序：PacBio RS II: 130 SMRT Cells; PacBio Sequel: 5 SMRT Cells;

· Illumina测序

HiSeq 2500 PE125

3. 研究内容

· Isoform 统计和分类

· 伴随无义介导的mRNA衰变的AS事件

· 转录因子Isoform分析

· LncRNA分析

· 基因进化分析

· Isoform 和可变剪接的组织特异性

· 可变剪接的保守性

· 多聚腺苷酸化分析 (APA)

· 基因表达聚类分析

★ 主要结论

1. 测序得到总数据量 (玉米+高粱) : 6,893,280 reads; 长度分布: 256–6643 bp; 玉米有1,570,093(96.7%)条高质量转录本能比对到玉米参考基因组, 非冗余序列136,745; 高粱有979,305(89.5%)条高质量转录本能比对到参考基因组, 非冗余序列95,380。

2. 高粱不同组织间基因表达相关性跟组织间的进化关系相关, 进化上越相近的组织, 相关性越高; 玉米不同组织间基因表达相关性跟组织进化无明显的相关性; 玉米和高粱同一组织基因表达相关性比同一物种不同组织相关性高。基因进化分析发现, 玉米比高粱有更多的“年轻”基因; 玉米的蛋白编码基因中有59.2%是“古老”基因, 3%是“年轻”基因; 高粱的蛋白编码基因中有66%是“古老”基因, 6.7%是“年轻”基因; 和进化“年轻”的基因相比, “古老”基因基因长度和ORF长度更长, Isoform种类更多。生殖组织比营养组织的进化年龄更高; 一般情况, 进化年龄指数 (TAI) 越高的组织, 转录组多样性指数 (TDI) 越高; 玉米和高粱同一组织的TAI和TDI指数有细微差别。

★ 参考文献

Wang B, Regulski M, Tseng E, et al. A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing[J]. Genome Research, 2018, 28(6): 921-932.

案例三: Poly(A)尾中散在分布的鸟苷酸 (G) 可抑制poly(A)结合蛋白 (PAB) 的结合从而降低mRNA的翻译效率

RNA poly(A)尾巴是成熟的mRNA和lncRNA的重要组成部分, 对RNA稳定性和翻译起着重要的调控作用。然而, 目前的poly(A)尾巴检测技术仍然非常有限。本研究基于Sequel平台建立了一种高准确度的RNA poly(A)尾巴检测技术。该方法可以实现转录组范围的全长转录本变体及其poly(A)尾巴的序列测定, 并揭示了在poly(A)尾巴的内部存在U、G和C碱基掺入, 为进一步研究RNA poly(A)尾巴中碱基组成的功能与调控奠定了基础。

★ 研究方法

PolyA尾高通量测序、CLIP-Seq、Ribo-Seq、mRNA稳定性检测。

★ 主要结论

1. AtPAB广泛结合植物体内mRNA的poly(A)尾, 但对不同mRNA的结合效率存在明显差异;
2. 拟南芥中10%的poly(A)尾含有至少一个G, 其含量在不同mRNA中的分布范围为0.8-28%;
3. Poly(A)尾中A的“纯度”越高 (G的含量越低), 与AtPAB的结合越强;
4. AtPAB对mRNA的结合可以提高mRNA的翻译效率;
5. 相应的, 在atpab突变体中, 带有不含G的“纯poly(A)尾”的mRNA翻译效率降低更为明显。
6. AtPAB2、AtPAB4、AtPAB8三突变体纯合致死, AtPABs的双突变体呈现多种发育异常表型, 暗示了由poly(A)尾与其结合蛋白介导的翻译调节机制对植物的正常生长发育具有至关重要的调控作用。

★ 参考文献

Liu Y, Nie H, Liu H, et al. Poly (A) inclusive RNA isoform sequencing (PALso-seq) reveals wide-spread non-adenosine residues within RNA poly (A) tails[J]. Nature Communications, 2019, 10(1): 1-13.

▶▶ 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。在常规建库测序中, 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)。

华大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。

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

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	~500,000	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

▶▶ 典型案例汇总

物种	发表时间	发表期刊	文献名称
拟南芥	2019	Genome Biology	Impact of poly(A)-tail G-content on Arabidopsis PAB binding and their role in enhancing translational efficiency
小鼠	2019	Nature Communications	Poly(A) inclusive RNA isoform sequencing (PAIso-seq) reveals wide-spread non-adenosine residues within RNA poly(A) tails
苜蓿	2019	BMC Plant Biology	Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (<i>Medicago sativa</i> L.)
玉米、高粱	2018	Genome Research	A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing
玉米	2017	Nature	Improved maize reference genome with single-molecule technologies

RNA-Seq

产品概述

RNA-Seq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，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。

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，Science、Nature、Cell等1,000+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，即可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

样本提取只需2天，从样本检测、建库测序到Dr. Tom交付只需11个工作日

产品应用



技术流程

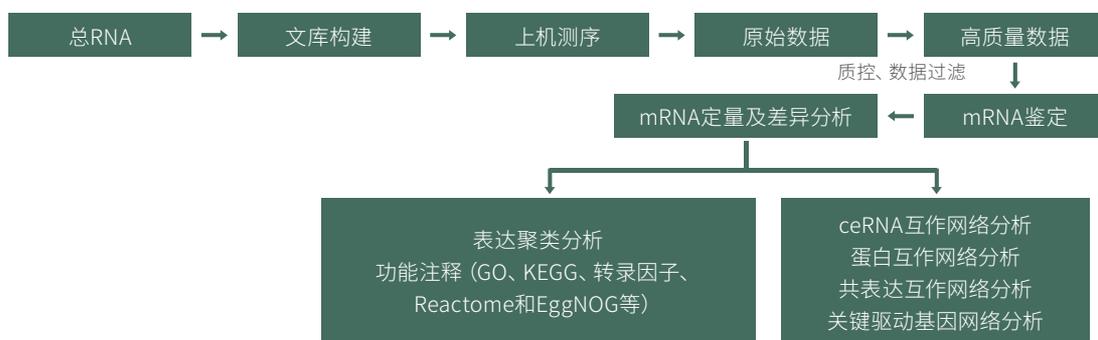


图1 RNA-Seq技术流程

►► 研究内容

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
- ④RNA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

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):人、鼠200ng,其他物种1 μ g

样品浓度:人、鼠10-1,000 ng/ μ L,其他物种40-1,000 ng/ μ L

样品纯度:28S/18S \geq 1.0;人鼠及动物RIN \geq 7.0,植物RIN \geq 6.0,和真菌RIN \geq 6.5;昆虫无28S/18S及RIN值要求

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等技术,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。

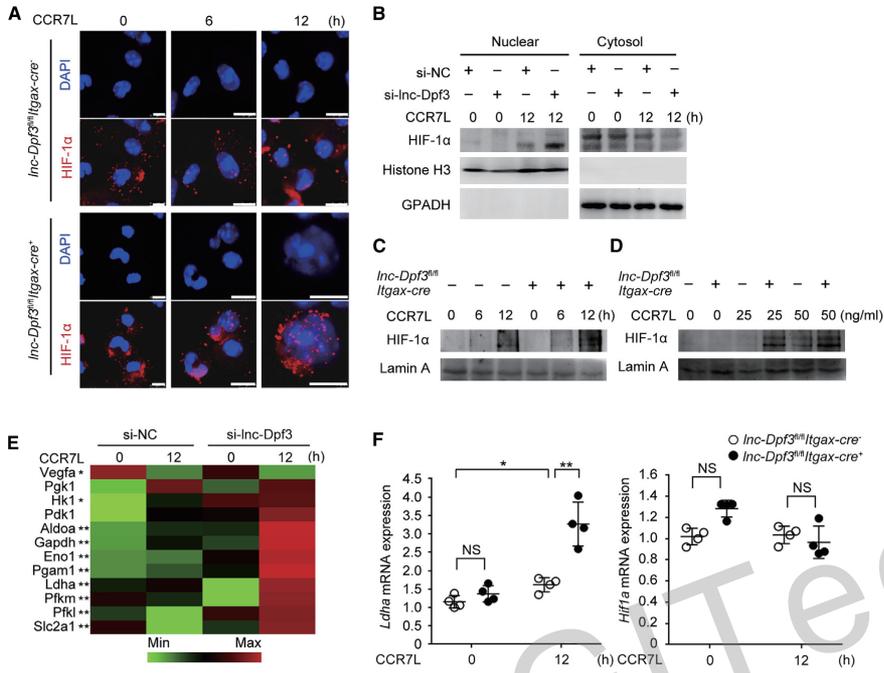


图2 Inc-Dpf3通过抑制HIF-1 α 诱导的Ldha转录, 抑制DC迁移

★ 参考文献

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

案例二: RNA-Seq在作物高效利用氮肥调控机制研究中的应用

通过选育半矮化作物, 极大地提高了绿色革命品种谷物的产量, 但因使用过多无机氮肥而导致环境恶化日益加剧。为加强全球粮食安全及可持续发展, 提高作物氮利用效率的需求日益迫切。因此, 亟需深入了解作物生长, 氮同化、碳固定的共同调节机制。材料为水稻半矮秆品系NJ6-sd1、对应野生型品系NJ6 (NJ6-sd1的氮素利用效率比NJ6低)。通过RNA-Seq和ChIP-Seq两项技术, 解析GRF4作用的分子机制。GRF4表达增加改善氮素吸收效率的同时, 不影响半矮化性状, 能进一步提高作物产量。

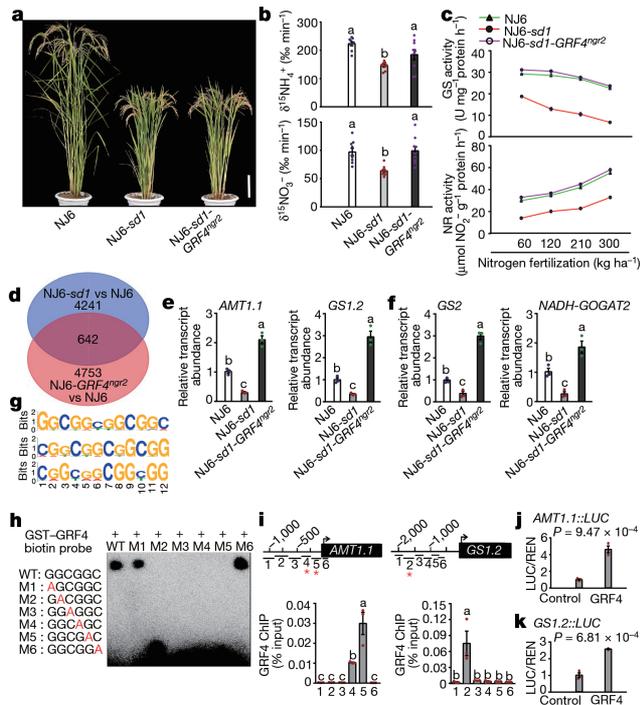


图3 GRF4调控多种氮代谢基因的表达

★ 参考文献

Li S, Tian Y, Wu K, et al. Modulating plant growth-metabolism coordination for sustainable agriculture[J]. Nature, 2018, 560(7720):595-600.

▶▶ 常见问题

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10-20 M 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: DNBSEQ 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	文献名称
医学研究			
人	2019	Immunity	CCR7 chemokine receptor-inducible Inc-Dpf3 restrains dendritic cell migration by inhibiting HIF-1 α -Mediated glycolysis
人	2019	Immunity	Specific decrease in B-Cell-Derived extracellular vesicles enhances post-chemotherapeutic CD8 ⁺ T Cell responses
小鼠	2019	Cell Research	Pioneering function of Isl1 in the epigenetic control of cardiomyocyte cell fate
人细胞系	2019	Cell Host & Microbe	Nutrient sensing by the intestinal epithelium orchestrates mucosal antimicrobial defense via translational control of Hes1
动植物研究			
秀丽隐杆线虫	2019	Cell Reports	A persistence detector for metabolic network rewiring in an animal
拟南芥	2019	Plant Physiology	The R2R3-MYB transcription factor MYB49 regulates cadmium accumulation
水稻	2019	Plant Biotechnology Journal	The basic helix - loop - helix transcription factor, OsPIL15 regulates grain size via directly targeting a purine permease gene OsPUP7 in rice



Small RNA测序

产品概述

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1ng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

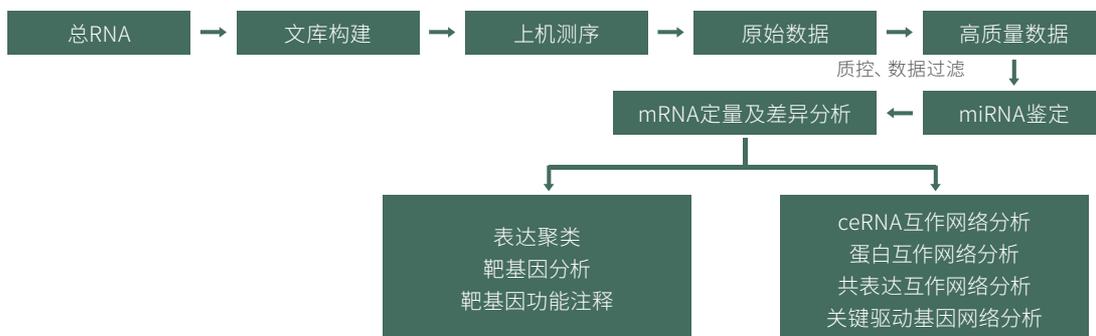


图1 Small RNA测序技术流程图

► 研究内容

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

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

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
 - ①miRNA-mRNA靶向关系分析
 - ②lncRNA-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

► 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、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: 标准流程的执行周期为30个工作日(样品数小于10个); 医学研究为25个工作日, 动植物研究为30个工作日

► 案例分析

案例一: miRNA表达谱研究——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。结果表明,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可能用于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。

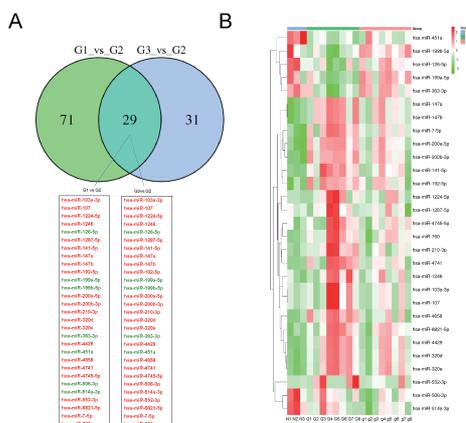


图2 腹膜化疗前后的GC患者和对照的外泌体miRNA差异

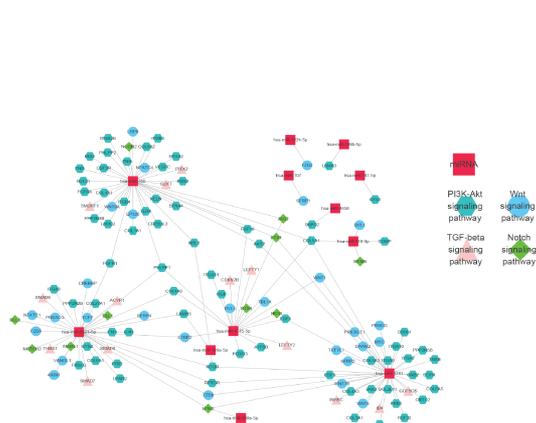


图3 腹膜扩散的GC患者中竞争RNA网络

★ 参考文献

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

案例二: miRNA机制研究——寄生昆虫衍生的miRNA调节宿主发育

寄生蜂会产生毒液、多DNA病毒 (PDV)、畸形细胞等, 改变宿主的生理状态有益于后代存活, 但改变的潜在分子机制仍不清楚。本研究发现寄生蜂 (*C. vestalis*) 的畸形细胞、多DNA病毒 (嵌入到寄生蜂基因组中特有的共生病毒) CvBV可以产生miRNA并通过不同方式将其传递到宿主中。寄生宿主小菜蛾 (*P. xylostella*) 中某些miRNA主要由畸形细胞产生, 而寄生宿主中由CvBV编码的miRNA的表达水平比非寄生宿主高100倍。一种畸形细胞产生的miRNA (Cve-miR-281-3p) 和一种CvBV产生的miRNA (Cve-miR-novel22-5p-1) 通过调节宿主蜕皮激素受体 (EcR) 的表达来阻止宿主生长。研究结果显示miRNA在动物寄生中可以进行跨物种调控, 及寄生期间宿主生理改变中miRNA的功能。

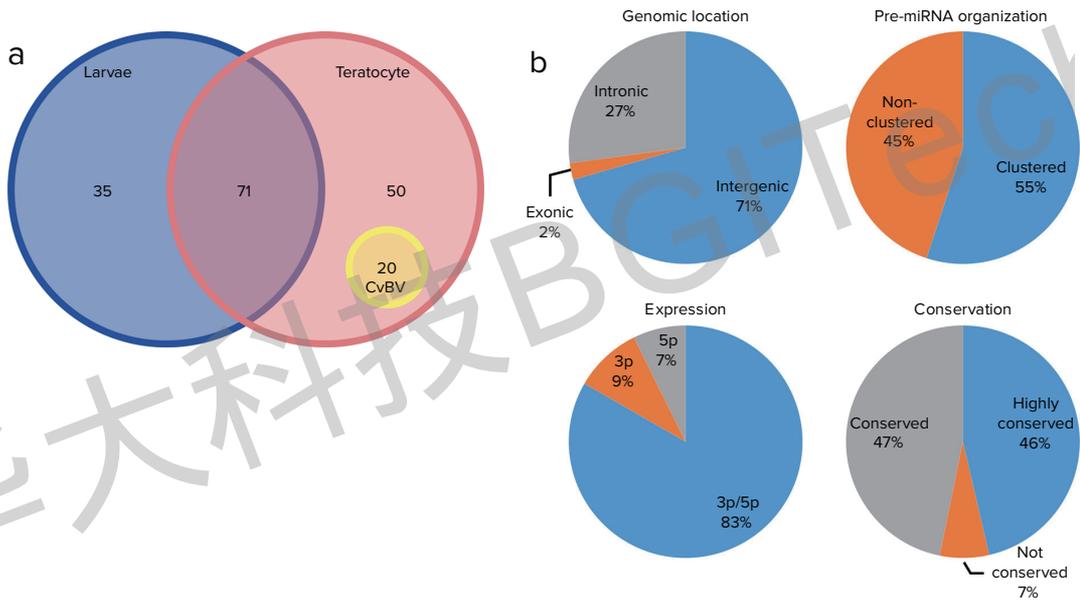


图4 寄生蜂幼虫和畸形细胞中的miRNA特征

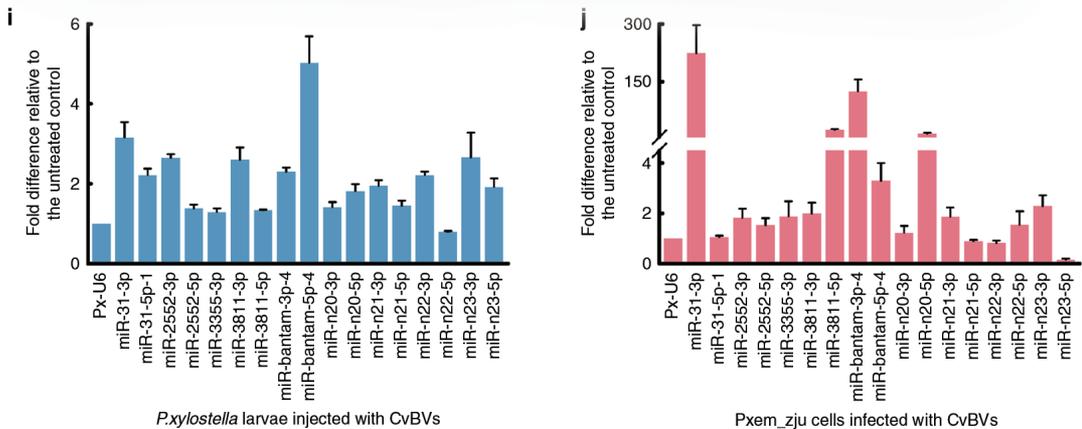


图5 CvBV编码的miRNA在CvBV感染的宿主小菜蛾中的表达

★ 参考文献

Wang Z, Ye X, Shi M, et al. Parasitic insect-derived miRNAs modulate host development[J]. *Nature Communications*, 2018, 9(1): 2205.

►► 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: 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, 只要满足送样要求, 均可用于小RNA建库测序。

Q: 进行small RNA测序对组织样品提取总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)、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/intron、repeat信息以及基因编码序列。

►► 华大合作发表文章 (部分)

研究方向	发表日期	发表期刊	文献名称
医学研究			
先天性免疫	2019	Nature Immunology	Inducible degradation of lncRNA Sros1 promotes IFN- γ -mediated activation of innate immune responses by stabilizing Stat1 mRNA
癌症	2019	Molecular Cell	Oncogenic biogenesis of pri-miR-17~92 reveals hierarchy and competition among polycistronic microRNAs
脂肪肝	2019	Cell Research	Complement C3 activation regulates the production of tRNA-derived fragments Gly-tRFs and promotes alcohol-induced liver injury and steatosis
人	2018	Molecular Cancer	The chromosome 11q13.3 amplification associated lymph node metastasis is driven by miR-548k through modulating tumor microenvironment
动植物研究			
芸薹属植物	2019	Molecular Plant	A Brassica miRNA Regulates Plant Growth and Immunity through Distinct Modes of Action
秀丽隐杆线虫	2019	Nature Communications	A secreted microRNA disrupts autophagy in distinct tissues of <i>Caenorhabditis elegans</i> upon ageing
蚊子	2019	Proceedings of the National Academy of Sciences of the United States of America	Subgenomic flavivirus RNA binds the mosquito DEAD/H-box helicase ME31B and determines Zika virus transmission by <i>Aedes aegypti</i>
小菜蛾	2018	Nature Communications	Parasitic insect-derived miRNAs modulate host development
小麦	2017	New Phytologist	<i>Puccinia striiformis</i> f. sp. <i>tritici</i> microRNA-like RNA 1 (Pst-miR1), an important pathogenicity factor of Pst, impairs wheat resistance to Pst by suppressing the wheat pathogenesis-related 2 gene



长链非编码RNA测序

▶▶ 产品概述

长链非编码RNA (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



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

细胞分化和发育研究

调控机理研究

疾病标志物寻找

疾病分子诊断

基因药物研发

▶▶ 技术参数

样本要求:

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: 人鼠为35个工作日, 非人鼠40个工作日

技术流程

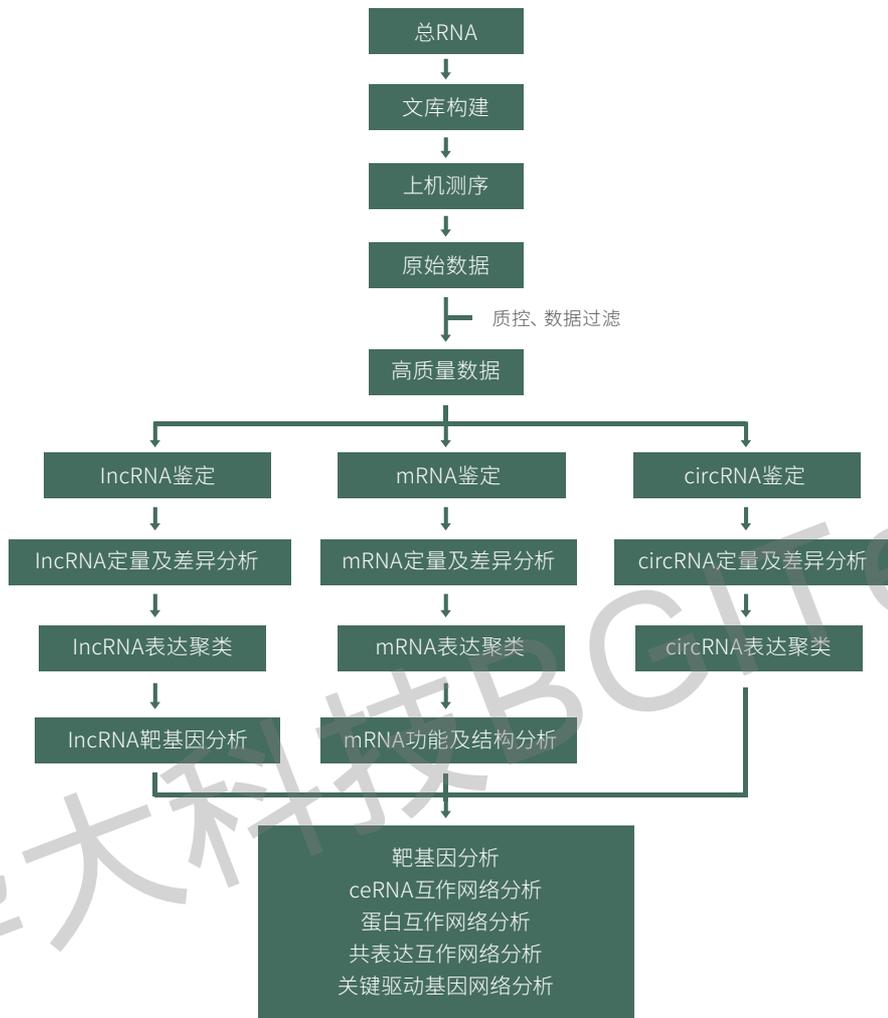


图1 长链非编码RNA测序技术流程图

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
 - ④RNA分类
- 2.参考基因组比对
- 3.IncRNA、mRNA、circRNA鉴定
- 4.IncRNA、mRNA、circRNA定量分析
- 5.IncRNA、mRNA、circRNA差异表达分析 (样本间、组间)
- 6.IncRNA、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.Rfam、Pfam、Reactome、COG、EggNOG和InterPro数据库注释
- 3.GSEA分析

互作网络分析

- 1.靶基因分析
 - ①miRNA-mRNA靶向关系分析
 - ②IncRNA-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

►► 案例分析

案例一：癌症lncRNA机制—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RNA 1) 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。

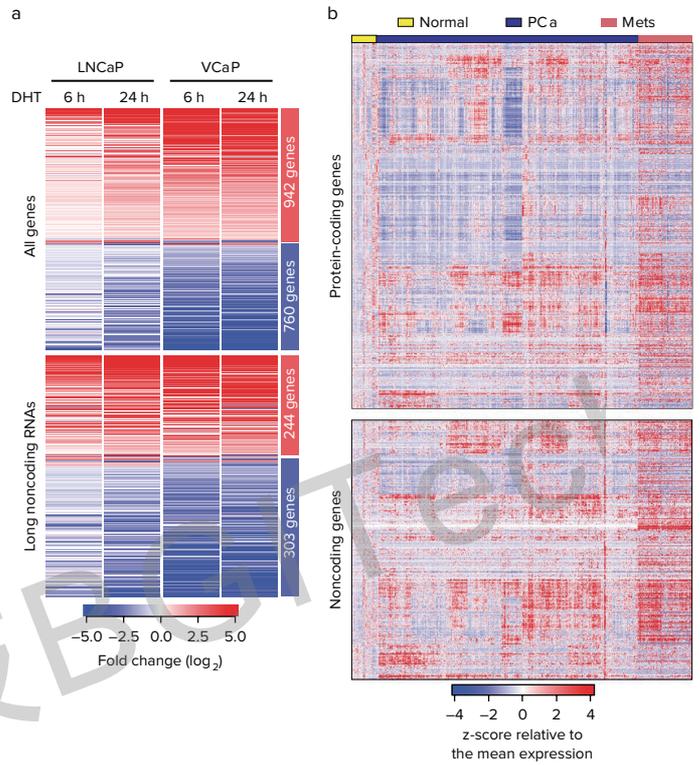


图2 前列腺癌中雄激素调节基因的鉴定热图

★ 参考文献

Zhang Y, Pitchiaya S, Cieslik 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.

案例二：棉花表型调控lncRNA—研究棉花衣分率和绒毛纤维

棉纤维是纺织工业重要的天然纤维。纤维起始数决定棉花衣分率, 是棉纤维产量的重要指标。虽然已通过转录组方法对棉纤维发育进行了分析, 但长链非编码RNA (lncRNA) 在棉绒和绒毛纤维起始调控方面的分子机制仍不甚清楚。本研究将Xu142与无纤维突变体Xu142 fl杂交, 获得了三个具有不同衣分率的品系。研究人员从开花后 (DPA) 0天和5天的Xu142、无纤维突变体Xu142 fl以及三个不同衣分率品系的胚珠 (附着纤维) 中收集表皮细胞, 以进行深度转录组测序。结果共鉴定了2,641个新基因、35,802个lncRNA和2,262个环状RNA (circRNA), 其中645个lncRNA优先在无纤维突变体Xu142 fl中表达, 651个lncRNA优先在纤维附着品系中表达。研究人员通过病毒诱导的基因沉默 (VIGS) 实验证明三种lncRNA参与棉纤维发育过程。本研究结果表明, 沉默X142 fl中的XLOC_545639与XLOC_039050增加了胚珠的纤维起始数, 但沉默Xu142中的XLOC_079089则导致纤维变短。

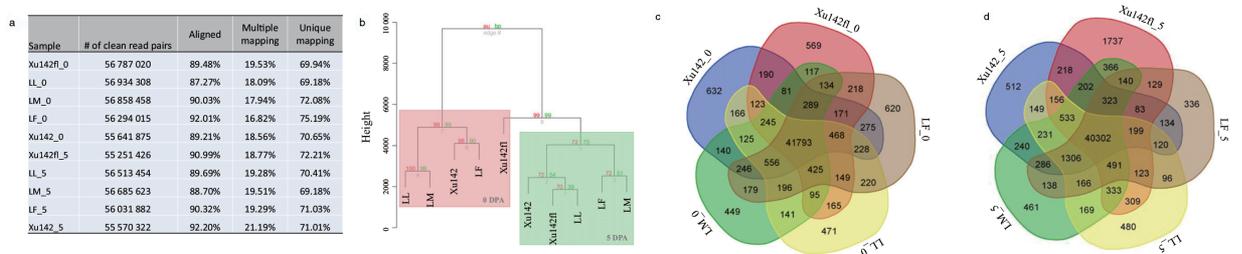


图3 测序数据分析

a. 10个样本的clean reads数及基因组比对率 b. 10个样本聚类分析
c-d. 0 DPA和5 DPA时期父母本及子代表达基因维恩图

★ 参考文献

Hu H, Wang M, Ding Y, et al. Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum* L.)[J]. Plant Biotechnology Journal, 2018, 16(5):1002-1012.

▶▶ 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: 长链非编码RNA的建库方案是什么?

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

Q: 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。

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。

▶▶ 华大合作发表文章 (部分)

研究内容	发表日期	发表期刊	文献名称
医学研究			
鼠-皮层神经元分化	2020	Protein Cell	Long non-coding RNA LncKdm2b regulates cortical neuronal differentiation by cis-activating Kdm2b
鼠-树突细胞迁移机制	2019	Immunity	CCR7 chemokine receptor-inducible lnc-Dpf3 restrains dendritic cell migration by inhibiting HIF-1 α -Mediated glycolysis
胚胎干细胞-核RNA功能	2019	Cell Reports	A functional link between nuclear RNA decay and transcriptional control mediated by the polycomb repressive complex 2
急性髓细胞样白血病	2017	Haematologica	Long noncoding RNA expression profile in cytogenetically normal acute myeloid leukemia identifies a distinct signature and a new biomarker in NPM1-mutated patients
动植物研究			
秀丽隐杆线虫	2019	Genome Biology	Systematic evaluation of <i>C. elegans</i> lincRNAs with CRISPR knockout mutants
棉纤维	2018	Plant Biotechnology Journal	Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (<i>Gossypium hirsutum</i> L.)
恒河猴	2016	Nucleic Acids Research	Extensive ceRNA-ceRNA interaction networks mediated by miRNAs regulate development in multiple rhesus tissues

华大科技BGITech

产品
简介

Product
Profile

| 基因科技 造福人类
Omics for All

华大科技BGITec

表观组学

全基因组甲基化测序 (WGBS)	063
ChIP-Seq	067



全基因组甲基化测序 (WGBS)

▶▶ 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 (Whole Genome Bisulfite Sequencing, WGBS) 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。

▶▶ 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

疾病发生
机制研究

免疫应答
干细胞研究

生命调节
机理研究

疾病标志物
的寻找

药物靶点
研究

▶▶ 技术参数

样本要求

样品总量: $\geq 2\mu\text{g}$

样品纯度: $\text{OD}_{260/280} = 1.8\sim 2.0$, $\text{A}_{260}/\text{A}_{230} \geq 1.6$; 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_2O ; 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: PE100

推荐数据量: $\geq 30\times$

项目执行周期: 45个工作日

► 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和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

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

► 案例分析

案例一：牛配子和体内早期胚胎的甲基化动态变化研究

★ 研究背景

哺乳动物配子和胚胎发育时期是甲基化活动最剧烈的时期；牛胚胎更像人胚胎，可以作为理解人类胚胎发育的良好模型；甲基化-基因不表达，去甲基化-基因表达。

★ 研究思路



图2 文章研究思路

★ 参考文献

Duan JE, Jiang ZC, Alqahtani F, et al. Methylome dynamics of bovine gametes and in vivo early embryos[J]. Frontiers in Genetics, 2019, 10: 512.

案例二: 孕妇烟草烟雾暴露与脐带血CD4⁺ 细胞DNA甲基化之间的关系

★ 研究背景

母体的烟草暴露会带来DNA甲基化修饰的改变;
以往测序均基于甲基化芯片, 只能检测有限的CpG, 且是测全血;
本文旨在研究母体暴露于烟草环境下脐带血CD4⁺ DNA甲基化。

★ 研究思路

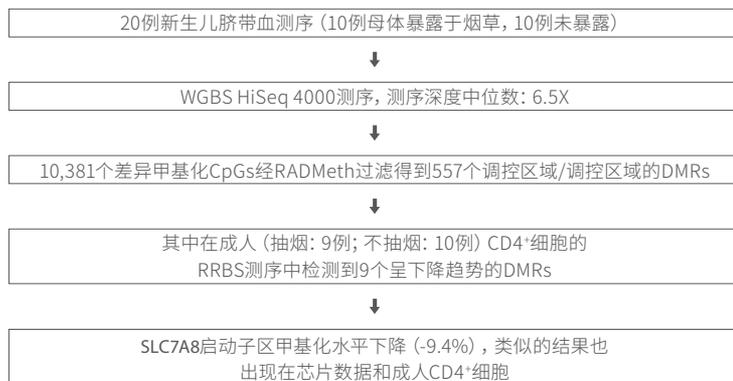


图3 文章研究思路

★ 参考文献

Howe CG, Zhou M, Wang X, et al. Associations between Maternal Tobacco Smoke Exposure and the Cord Blood CD4⁺ DNA Methylome[J]. Environmental Health Perspectives, 2019, 127(4), 047009.

►► 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没有不发生甲基化的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	文献名称
全基因组重亚硫酸盐测序揭示藏鸡对高海拔环境的独特适应性	2018	PLoS One	Whole genome bisulfite sequencing reveals unique adaptations to high-altitude environments in Tibetan chickens
RNA通过dCas9 甲基转移酶介导的DNA甲基化	2018	GigaScience	Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases
方法学, 利用Bisulfite-Seq分析病毒整合位点	2017	GigaScience	BS-virus-finder virus integration calling using bisulfite sequencing data
家蚕病毒感染DNA甲基化表观调控	2017	Scientific Reports	DNA methylation in silkworm genome may provide insights into epigenetic regulation of response to Bombyx mori cytopovirus infection
早发性肌张力障碍	2016	Nature Genetics	Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia
疾病研究	2016	Genome Biology	Novel regional age-associated DNA methylation changes within human common disease-associated loci
甲基化而在犬体细胞核转移性发育	2016	Scientific Reports	Stochastic anomaly of methylome but persistent SRY hypermethylation in disorder of sex development in canine somatic cell nuclear transfer
方法学	2015	Bioinformatics	BS-SNPer: SNP calling in Bisulfite-Seq data

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——染色质免疫共沉淀测序 (Chromatin Immunoprecipitation Sequencing, ChIP-Seq)，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，并通过高通量测序与数据分析，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。

▶▶ 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
DNBSEQ平台最低起始量仅需5ng，5ng以下亦可做定制化

▶▶ 产品应用

疾病发生
机制研究

免疫应答
干细胞研究

生命调节
机理研究

疾病标志物
的寻找

药物靶点
研究

▶▶ 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和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

►► 技术流程

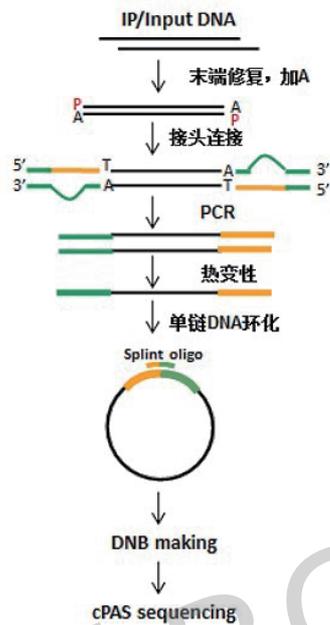


图1 DNBSEQ平台ChIP-Seq技术流程

►► 信息分析内容

标准信息分析

- 1.数据基本处理与质控
- 2.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

►► 案例分析

案例一: DNBSEQ ChIP-Seq在乳腺癌转移研究中的应用

★ 研究背景

乳腺癌是人类常见的一种恶性肿瘤。据统计,全球每年新发乳腺癌高达120万人,是女性第一高发的肿瘤。近年来我国的乳腺癌发病率明显增高,而乳腺癌转移是导致患者死亡的主要原因,乳腺癌转移的分子机理目前还不清楚。在以往的研究中观察到FOXN3在其他恶性肿瘤中表达失调的现象,但FOXN3在恶性肿瘤(包括乳腺癌)发生中的作用及分子机理仍有待研究,它的功能又需要哪些分子共同作用也需要探讨。

★ 研究思路及结果展示

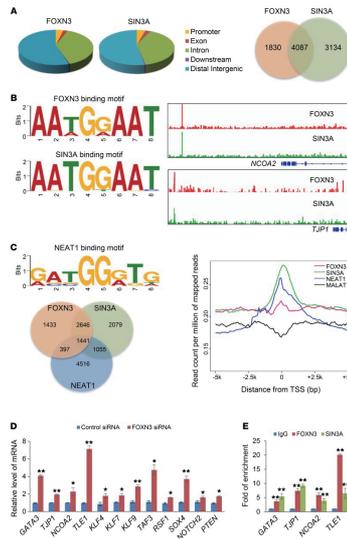


图2 FOXN3-NEAT1-SIN3A复合物转录靶点的全基因组分析

★ 参考文献

Li W, Zhang Z, Liu X, et al. The FOXN3-NEAT1-SIN3A repressor complex promotes progression of Hormonally responsive breast cancer[J]. Journal of Clinical Investigation, 2017, 127(9):3421-3440.

案例二: DNBSEQ ChIP-Seq与RNA-Seq在作物高效利用氮肥调控机制研究中的应用

★ 研究背景

通过选育半矮化作物,极大地提高了绿色革命品种谷物的产量,但因使用过多无机氮肥而导致环境恶化日益加剧。为加强全球粮食安全及可持续发展,提高作物氮利用效率的需求日益迫切。因此,亟需深入了解作物生长、氮同化、碳固定的共同调节机制。

★ 研究思路及结果展示

基于DNBSEQ平台的RNA-Seq和ChIP-Seq两项技术,解析GRF4作用的分子机制,证明GRF4是激活下游参与氮素吸收和氮素同化的相关基因。ChIP-Seq揭示了潜在的GRF4靶识别位点,其中主要是多个氮代谢基因启动子共有的GGCGG-motif。

GRF4与GIF1 (GRF-interacting factor 1) 复合体结合包含GGCG-motif的启动子,促进氮素同化吸收。SLR1 (属于DELLA家族) 在该途径中可抑制GRF4与GIF1的相互作用,导致氮素利用率降低。同时,GRF4还可以促进碳同化吸收相关基因的表达。

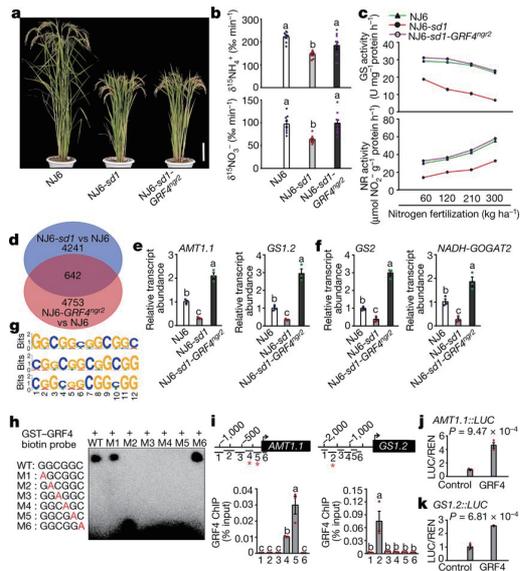


图3 GRF4调控多种氮代谢基因的表达

★ 参考文献

Li S, Tian Y, Wu K, et al. Modulating plant growth-metabolism coordination for sustainable agriculture[J]. Nature, 2018, 560(7720):595-600.

► 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。

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	文献名称
水稻高效利用氮肥	2018	Nature	Modulating plant growth-metabolism coordination for sustainable agriculture
dCas9甲基转移酶机制	2018	GigaScience	Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases
Wiskott-Aldrich综合征	2017	Genome Medicine	Nuclear Wiskott-Aldrich syndrome protein co-regulates T cell factor 1-mediated transcription in T cells
雌性激素受体结合位点研究	2017	Gene	Genome-wide mapping of estrogen receptor α binding sites by ChIP-seq to identify genes related to sexual maturity in hens
肿瘤研究	2017	Nature Communications	Menin enhances c-Myc-mediated transcription to promote cancer progression
胞内菌感染免疫学	2017	PNAS	Nuclear carbonic anhydrase 6B associates with PRMT5 to epigenetically promote IL-12 expression in innate response
乳腺癌研究	2017	The Journal of Clinical Investigation	The FOXN3-NEAT1-SIN3A repressor complex promotes progression of hormonally responsive breast cancer

华大科技BGITech

产品
简介

Product
Profile

华大科技BGITec

宏基因组学

16S/18S/ITS扩增子测序	073
宏基因组测序	080



16S/18S/ITS扩增子测序

▶▶ 产品概述

16S rDNA是细菌分类学研究中最常用的“分子钟”，其序列包含9个可变区 (Variable Region) 和10个保守区 (Constant Region)。可变区因细菌而异，且变异程度与细菌的系统发育密切相关。通过检测16S rDNA的序列变异和丰度，可以了解环境样品中群落多样性信息。基于16S rDNA的分析在微生物分类鉴定、微生态研究等方面起到重要作用。

18S rDNA或ITS (Internal Transcribed Spacer) 被广泛应用在真菌分类鉴定中。18S rDNA在系统发育研究中较适用于种及以上阶元的分类；ITS属于中度保守区域，利用它可研究种及种以下的分类阶元。

▶▶ 产品优势



策略多样

不同来源样本采用不同提取方法和建库测序策略，满足多种环境研究需求



平台多样

搭配短读长和长读长测序平台，可满足16S/18S/ITS不同高变区域或全长测序的需求



经验丰富

已测序样品类型涉及粪便、土壤、水体、唾液、牙菌斑、体腔、胃液、白带、空气、血液、皮屑等



样本需求量低

华大基因16S产品推荐DNA样本量50ng以上；对于样本获取困难的样本，只要样本量高于0ng也有可能建库成功

▶▶ 产品应用



▶▶ 技术流程

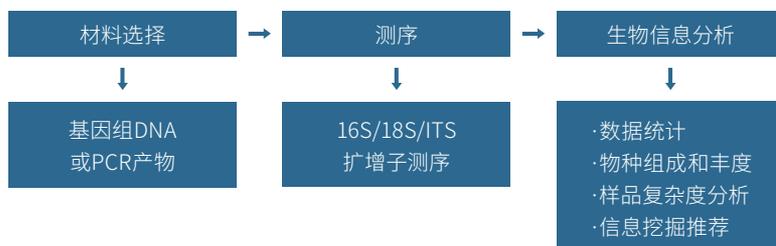


图1 16S/18S/ITS扩增子测序技术流程图

信息分析流程

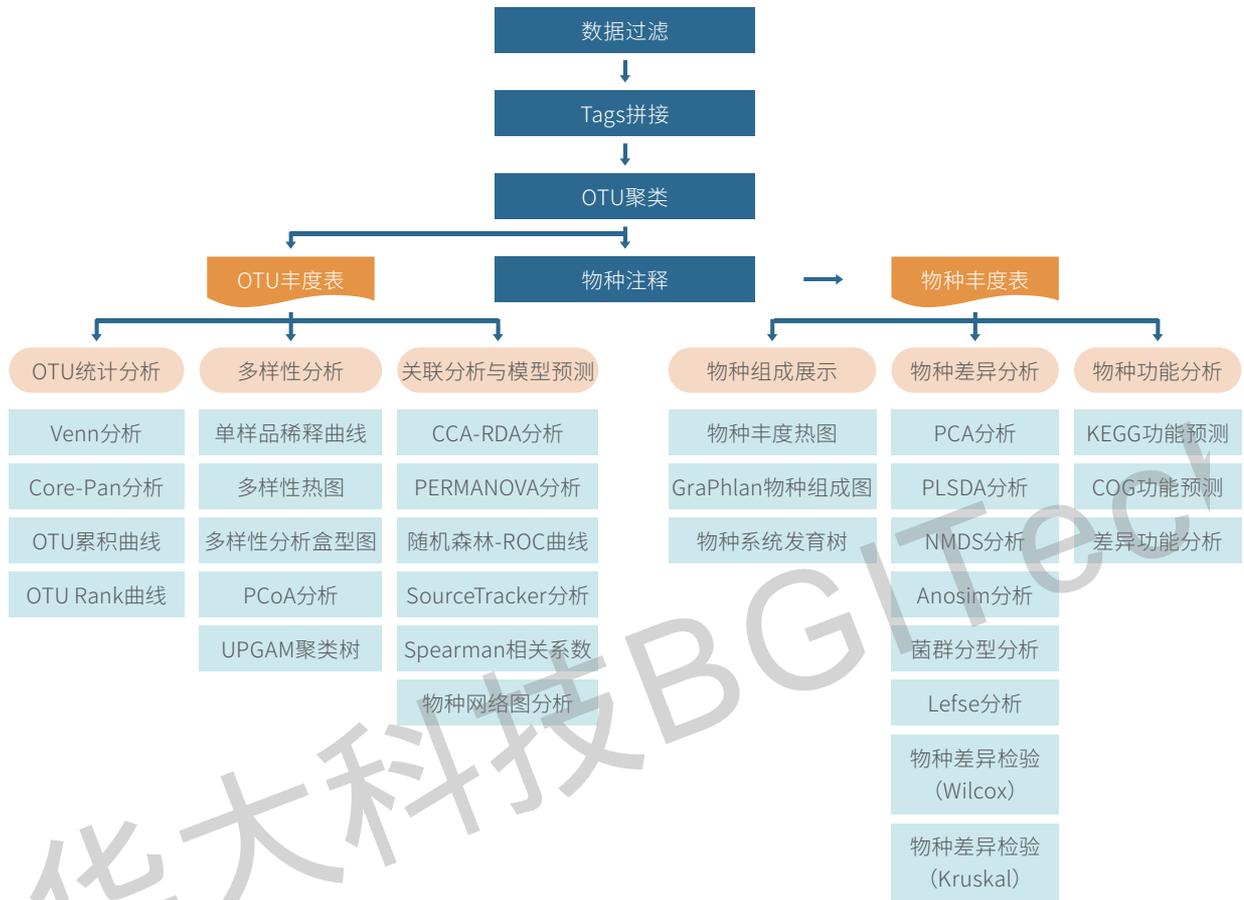


图2 16S/18S/ITS扩增子测序信息分析流程图

技术参数

样品要求: 基因组DNA, 无降解或轻微降解

样品总量: $m \geq 50\text{ng}$ (推荐)

样品浓度: $c \geq 0\text{ng}/\mu\text{L}$

测序策略:

	扩增区域	引物名称	引物对	测序类型*
细菌	V4	515F	GTGCCAGCMGCCGCGGTAA	PE250
		806R	GGACTACHVGGGTWTCTAAT	
	V1-V3	8F	AGAGTTTGATYMTGGCTCAG	PE300
		518R	ATTACCGCGGCTGCTGG	
	V3-V4	341F	ACTCCTACGGGAGGCAGCAG	PE300
		806R	GGACTACHVGGGTWTCTAAT	
	V4-V5	515F	GTGCCAGCMGCCGCGG	PE300
		907R	CCGTCAATTCMTTTRAGT	
真菌	ITS1	its1	CTTGGTCATTTAGAGGAAGTAA	PE250
		its2	GCTGCGTTCTTCATCGATGC	
	ITS2	its3	GCATCGATGAAGAACGCAGC	PE300
		its4	TCCTCCGCTTATTGATATGC	

*注: 只有两端完全测通的Reads (Tags)才能用于进一步的分析, 因此不同的扩增区域请严格遵循对应的测序类型

推荐数据量: 简单环境单个样品数据量不低于50,000tags; 复杂环境 (如土壤和海水) 样本建议测序数据量高于100,000tags

交付周期: 30个工作日

▶▶ 案例分析

案例一：女性生殖道微环境及其与生殖健康的关联

该项研究是大规模的育龄女性生殖道微生态研究项目，采用16S检测技术对女性盆腔与上下生殖道各部位菌群分布及与生殖系统疾病的关联进行研究。其研究成果打破了“盆腔和上生殖道为无菌环境”的传统认知，发现正常女性盆腔与上生殖道亦存在微生物。首次揭示了生殖道从阴道到宫颈管、宫腔、输卵管，直至盆腔的菌群结构具有一定连续性，并提出盆腔与女性生殖道微生态环境和生殖系统健康及相关疾病具有重要的关联性。

★ 样本来源

采集了110名因良性疾病接受手术治疗的育龄女性生殖道不同部位的样本，包括阴道下1/3 (CL)、阴道后穹窿 (CU)、宫颈管 (CV)、子宫腔 (ET)、输卵管 (FL) 及盆腔液 (PF)。

★ 主要方法

16S rRNA基因扩增测序分析、实时定量PCR和微生物传统培养法分析。

★ 主要结果

1. 上生殖道并非无菌环境，从阴道到输卵管及盆腔，各部位微生物物种组成呈解剖结构上的渐进性变化。

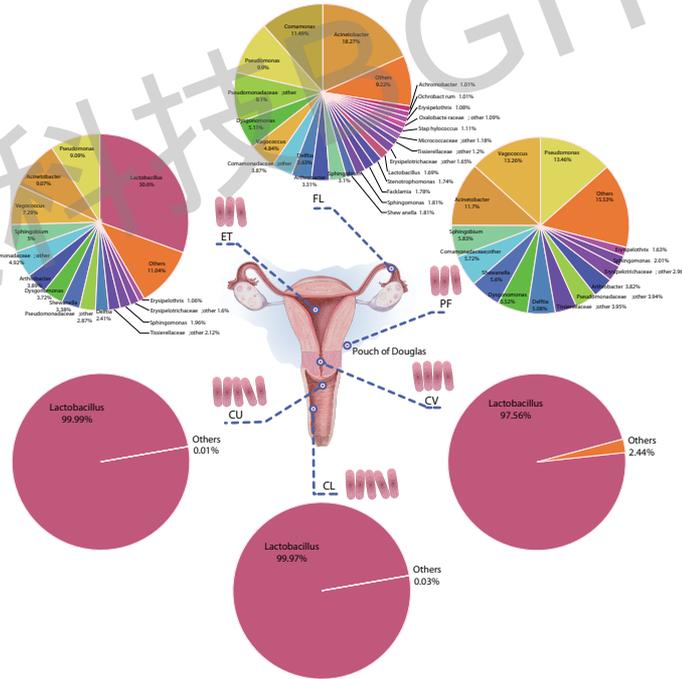


图3 生殖道不同部位微生物群落结构

2. 同一个体不同部位的样本间具有很高的相关性，不同个体间菌群组成变化明显。此外，同一个体的宫颈管样本与腹腔液样本具有显著的相关性，表明在普通人群中可以通过分析取得的宫颈粘液样本来评价宫腔和腹腔的菌群分布状况。

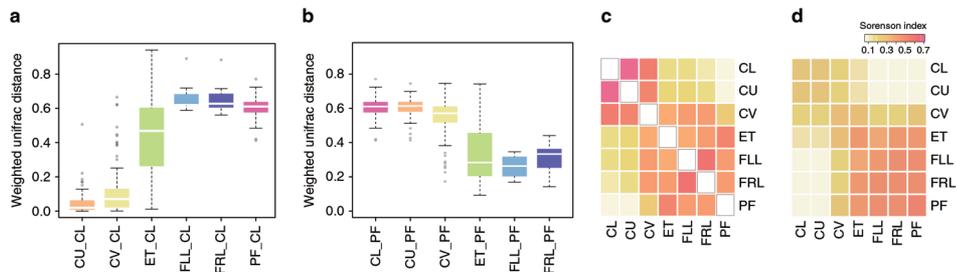


图4 不同个体/同一个体不同部位微生物群落结构一致性

3.对表型信息进行关联分析发现月经周期、孕产次数以及子宫内膜异位症、子宫肌瘤症等多种疾病相关的不孕等都与生殖道菌群的变化有关。

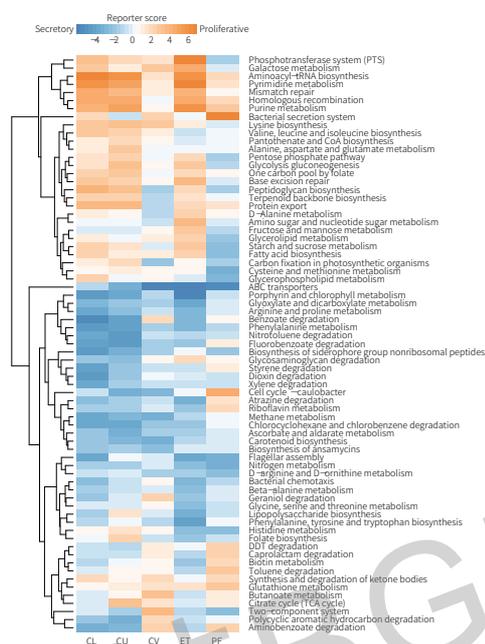


图5 生殖道微生物功能 (预测的) 与女性生理周期的关联分析

★ 参考文献

Chen C, Song X, Wei W, et al. The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases[J]. Nature Communications, 2017, 8(1): 875.

案例二：猪肠道菌群衍生细菌素可增强早期断奶仔猪腹泻抵抗力

在猪养殖过程中，早期断奶可以缩短猪的屠宰周期并改善母猪的繁殖性能。然而早期断奶容易导致应激性腹泻，仔猪死亡率上升，生长性能降低。使用抗生素可以预防仔猪断奶腹泻，降低饲养成本，但是由于病原菌抗生素抗性和抗生素残留问题，欧盟已完全禁止在动物饲养中使用抗生素。因此，寻找抗生素替代品以预防早期断奶仔猪的腹泻对于畜牧业和粮食安全至关重要。哺乳动物肠道菌群与宿主健康密切相关，通过粪菌移植或益生菌/益生元调控肠道菌群已成为有前景的胃肠道疾病治疗策略。

与商业杂交LY仔猪相比，CM仔猪（中国本土品种）对早期断奶应激诱导的腹泻抵抗力更强。本研究在早期断奶之前给LY仔猪口服CM仔猪粪便微生物群，LY仔猪的腹泻抗性增强。通过比较粪菌移植组和对照组LY仔猪肠道微生物群的相对丰度，鉴定到两个可能跟腹泻抗性相关的菌种加氏乳杆菌LA39 (*Lactobacillus gasseri* LA39) 和乳酸杆菌 (*Lactobacillus frumenti*)，并通过qPCR进行验证。腹泻抵抗力依赖于细菌素Gassericin A, Gassericin A与角蛋白19 (KRT19) 在肠上皮细胞质膜上的结合对于增强液体吸收和减少分泌至关重要。本研究结果表明L.gasseri LA39和L.frumenti可能是预防哺乳动物腹泻的有效抗生素替代品。

★ 方案设计

对LY仔猪和CM仔猪按不同处理进行分组如下：

- ①LY: LY仔猪，不经任何处理，n=3；
- ②LY (saline) : LY仔猪，day10-day18隔日口服生理盐水，n=3；
- ③LY (high dose) : LY仔猪，day10-day18隔日口服高浓度CM仔猪粪菌悬液，n=3；
- ④LY (low dose) : LY仔猪，day10-day18隔日口服低浓度CM仔猪粪菌悬液，n=3；
- ⑤LY (oxytetracycline) : LY仔猪，断奶日 (day21) 肌肉注射长效土霉素，n=3；
- ⑥CM: CM仔猪，不经任何处理，n=3；

以上各组仔猪在断奶后第3, 5, 6, 8, 11天收集粪便样本，进行16S V4和ITS2测序。

★ 主要结果

1. 粪菌移植仔猪腹泻症状缓解, 肠道菌群结构和功能发生改变

① 接受CM粪菌移植的LY仔猪早期断奶症状缓解

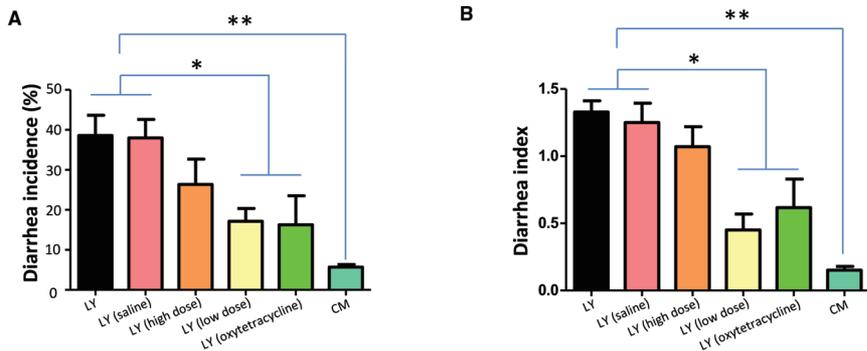


图6 各组仔猪腹泻发病率 (A) 和腹泻指数 (B)

② LY(saline)组, LY(low dose)组和CM组粪便细菌有明显差异; 真菌无明显分群

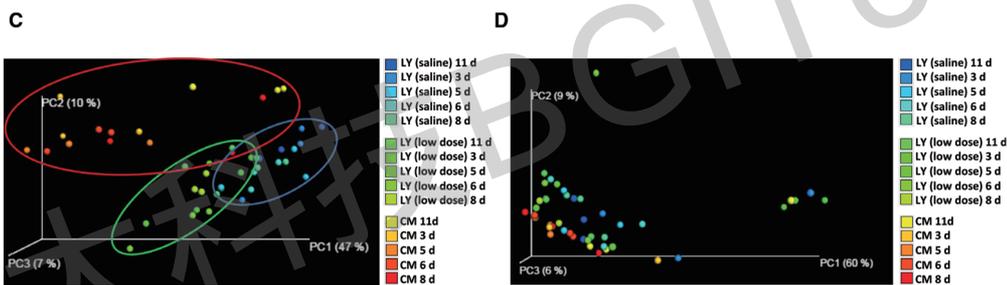


图7 细菌和真菌群落PCoA结果

③ 接受CM粪菌移植的LY仔猪细菌/真菌群落多样性发生了改变

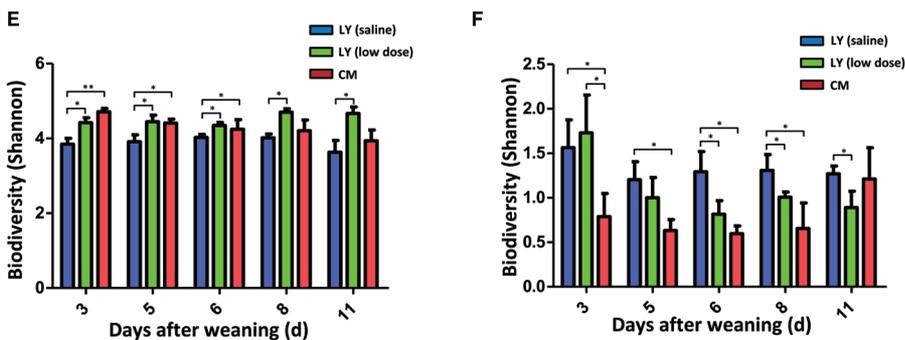


图8 细菌和真菌alpha diversity结果

④ LY(saline)组, LY(low dose)组和CM组粪便细菌功能结构有明显差异

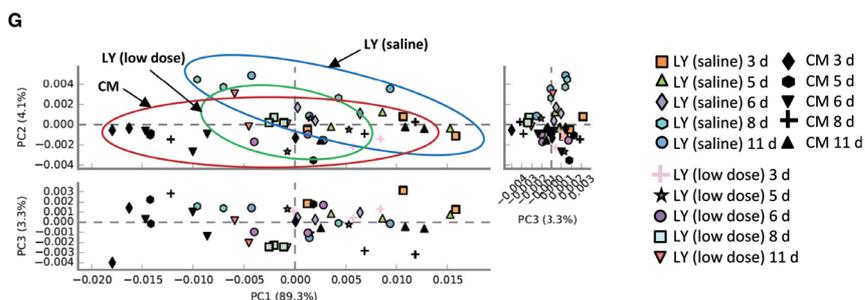


图9 仔猪肠道菌群功能pCoA图 (基于PICRUSt2软件 KEGG pathway分析)

2.口服腹泻抗性相关的肠道微生物可预防仔猪早期断奶应激引起的腹泻。鉴定得到5个跟腹泻抗性相关的菌种，其中加氏乳杆菌LA39 (*Lactobacillus gasseri* LA39) 和乳酸杆菌 (*Lactobacillus frumenti*) 单独服用即可增强仔猪腹泻抗性。

3.肠道微生物介导的抗腹泻功能依赖于分泌型Gassericin A。混合菌群移植和单独服用LG/LF菌株的LY仔猪肠道菌群Gassericin A编码基因GaaA丰度升高；口服GaaA缺陷LG菌株不能改善LY仔猪腹泻症状。

4.Gassericin A与肠上皮细胞中的细胞质膜结合，增强肠液吸收，减少肠液分泌。

5.KRT19介导的Gassericin A与细胞质膜结合，对于Gassericin A介导的肠液吸收增强和液体分泌减少至关重要。

6.Gassericin A通过激活由雷帕霉素机制靶标介导的磷酸二酯酶活性降低细胞环核苷酸水平，增加肠液吸收并减少肠液分泌。

★ 参考文献

Hu J, Ma L, Nie Y, et al. A microbiota-derived bacteriocin targets the host to confer diarrhea resistance in early-weaned piglets[J]. *Cell Host & Microbe*, 2018, 24(6): 817-832. e8.

► 常见问题

Q: 16S测序能不能进行功能分析?

A: 16S测序主要是基于16S rDNA序列相似性进行OTU聚类进而进行物种注释及相关多样性研究。因为16S测序并没有测到对应物种的基因组信息，不能直接基于测序结果进行功能注释。

利用软件PICRUSt可以进行16S功能预测，该软件的原理是：对由16S测序分析得到的OTU丰度进行拷贝数均一化，得到样品中可能出现的细菌及数目，从细菌的基因组信息得到对应的基因信息及注释信息，再结合均一化的OTU丰度来预测样品中可能存在的各级KEGG通路及丰度值以及COG功能信息及丰度值。

基于16S的功能预测可以为后续功能研究提供参考，但该分析不能反映群落中因基因表达差异导致的功能差异。如果主要关注功能差异，最好选择宏基因组测序来进行功能研究。

Q: 16S相关文章中选择的测序区域各不相同（如V4、V3-V4、V1-V3、V3等），选择依据是什么，选择哪个区域比较好?

A: 不同物种不同区域多样性不同，选择不同区域测序结果会有不同，可能会造成物种多样性的低估或高估。在非全长16S测序的情况下，测序区域也并非越长越好，不同样本需要的测序区域可能也不一样，跟全长16S结果最相近的测序区域即是最优选择。根据我们大量的项目经验，目前测序项目较多的区域为V4、V1-V3、V3-V4、V4-V5、V5-V6，具体项目测序区域请参考相关文献进行选择。

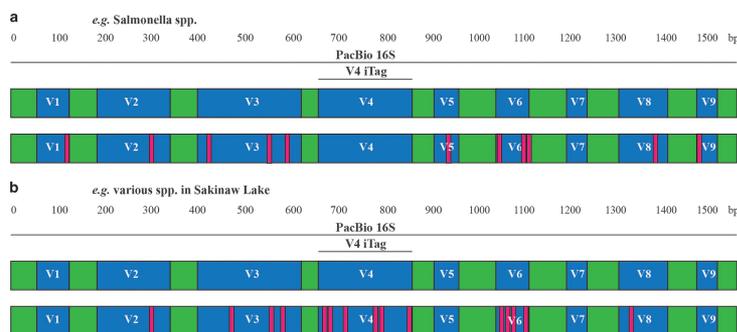


图10 不同物种16S各可变区变异程度不同

Q: 我的样本检测合格了，样本量也达到了推荐样本量要求，却建库失败了，可能是什么原因造成的，有没有什么解决方案?

A: 这种情况常见于宿主微生物样本，该类样本通常还有大量的宿主DNA，由于样本检测中不能区分宿主和微生物DNA，实际检测到的DNA其实绝大多数都是宿主DNA，导致检测到的样本量很高，却建库失败的情况。对于这类样本，推荐在样本制备过程中进行特殊处理，尽量减少宿主DNA含量。目前市面上有可以去宿主DNA的试剂盒（如QIAamp DNA Microbiome Kit），对于宿主DNA含量较高的样本如唾液、粘膜样本等，可以选择对应的试剂盒处理。

Q: 16S测序如果样本含有较多宿主DNA, 是否需要增加测序数据量?

A: 16S扩增子测序是对微生物16S rDNA 特定区域进行扩增并测序分析的产品, 扩增产物不包含宿主DNA的信息。所以样本中宿主DNA并不会对测序结果造成影响(宿主DNA含量过高可能会导致建库失败), 测序过程中也无需增加数据量。宏基因组测序是对环境中所有物种进行全基因组测序的产品, 样本中宿主DNA含量较高会导致宏基因组测序宿主比例偏高, 微生物有效数据量相对较少, 这种情况下可以通过增加测序数据量来增加有效数据。

▶▶ 华大合作发表文章 (部分)

样本来源	发表时间	发表期刊	文献名称
人肠道	2019	Food & Function	Probiotic strains improve high-fat diet-induced hypercholesterolemia through modulating gut microbiota in ways different from atorvastatin
茅台发酵剂	2019	Frontiers in Microbiology	Deciphering the composition and functional profile of the microbial communities in Chinese Moutai liquor starters
根系	2019	Nature Biotechnology	NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice
唾液、肺泡灌洗液	2019	Medical Science Monitor	A preliminary study of microbiota diversity in saliva and bronchoalveolar lavage fluid from patients with primary bronchogenic carcinoma
根际	2018	Nature Communications	The structure and function of the global citrus rhizosphere microbiome
小鼠肠道	2018	Molecules	Effects of rich-polyphenols extract of dendrobium loddigesii on anti-diabetic, anti-inflammatory, anti-oxidant, and gut microbiota modulation in db/db mice
猪肠道	2018	Cell Host & Microbe	A microbiota-derived bacteriocin targets the host to confer diarrhea resistance in early-weaned piglets
大鼠肠道	2018	Microbiome	Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats
生殖道	2017	Nature Communications	The microbiota continuum along the female reproductive tract and its relation to uterine-related diseases

宏基因组测序

产品概述

宏基因组学 (Metagenomics) 是将环境样品中的微生物群落作为整体进行研究的学科。宏基因组测序与其他研究环境群落的方法相比, 具有通量大、产出数据多、更为高效的优势, 不仅能够获得环境物种组成和丰度信息, 还能进行功能基因分析, 比较样品间基因差异, 研究物种间的代谢网络。通过深度挖掘具有应用价值的基因资源, 为研究和开发新的微生物活性物质提供有力支持。

产品优势



测序准确性高

DNBSEQ平台滚环扩增构建DNB测序文库, PCR扩增错误不累积, 高保真序列信息无index hopping担忧, 结果更可靠



成本可控, 数据利用率高

自主测序平台, 成本可控; Duplication率低, 同样的数据量有效数据多出3%-17%



样本需求量低

常规宏基因组建库建议样本量在1 μ g以上。对于难以获取的样本, 也可以选择微量建库, 样本量可低至几ng



经验丰富

有丰富的宏基因组项目经验, 特别是在人体微生物研究方面处于领先地位, 已发表文章100+, 其中CNS系列文章26篇; 有专业的meta大项目团队, 提供切实可行的项目方案, 兼顾商业合作、科研合作优势

产品应用



技术流程



图1 宏基因组测序技术流程图

研究内容

分析模块	分析条款	描述
项目概览	项目信息	
数据过滤	数据过滤统计	
组装	De novo组装	
基因预测及基因集构建	基因预测及基因集构建	
基因分析	①基因Venn图(2-5个样本或组别)	展示多样品共有和特有基因数目, 直观展示样品或者分组间基因的重叠情况
	②基因差异分析(分组 ≥ 2 , 每组样本数 ≥ 3)	寻找组间差异基因
基因集功能注释	Eggnog、KEGG、CARD、COG、CAZy、Swiss-prot, NR等	其他功能数据库可个性化分析
物种分析	①物种累积曲线图分析	抽样个数对物种多样性的影响, 判断采样量是否足够
	②物种组成分析	门纲目科属种水平进行物种分析
	③物种差异分析(分组 ≥ 2 , 每组样本数 ≥ 3)	寻找组间差异物种
	④差异物种丰度热图(分组 ≥ 2 , 每组样本数 ≥ 3)	根据物种或样品丰度相似性进行聚类, 并加上分组信息; 直观展示同一组样品的聚类情况及样品群落组成相似性和差异性
功能分析	①差异KO分析(分组 ≥ 2 , 每组样本数 ≥ 3)	寻找组间差异KO
	②差异KO丰度热图分析(分组 ≥ 2 , 每组样本数 ≥ 3)	根据KO或样品丰度相似性进行聚类, 并加上分组信息; 直观展示同一组样品的聚类情况及样品群落组成相似性和差异性
	③差异pathway分析(分组 ≥ 2 , 每组样本数 ≥ 8)	柱状图直观展示差异的pathway以及差异是否显著
多样性分析	①物种水平Alpha多样性分析	Shannon指数
	②KO水平Alpha多样性分析	Shannon指数
	③物种水平Beta多样性分析	Bray、jsd距离
	④KO水平Beta多样性分析	Bray、jsd距离
网络互作	①基于属水平的network分析	网络互作图, 直观展示群落中核心物种及物种之间的关联
相似性分析	①PCA分析	研究样本群落组成的相似性或相异性, 根据离散程度判断样品组成是否具有相似性
	②PCoA分析	研究样本群落组成的相似性或相异性
	③NMDS分析	研究样本群落组成的相似性或相异性, 将多维空间的研究对象简化到低维空间进行定位、分析和归类, 同时又保留对象间原始关系的数据分析方法
	④基于距离热图分析(物种、KO)	样品组成的形似性和差异性
关联分析与模型预测	①CCA分析(需提供环境因子数据)	分析物种或功能与环境因子的关系
	②PERMANOVA/Adonis置换多元方差分析(默认分组因素对样品差异的影响, 如关注表型对样品差异的影响需提供表型信息)	分析不同分组因素或环境因子(如临床表型数据、土壤理化指标等)对样品差异的影响大小, 并进行显著性分析
	③物种与表型Spearman相关系数分析(需提供表型信息)	绘制优势物种与表型之间Spearman相关性热图, 可以发现优势物种与表型之间的重要模式与关系
	④MaAsLin分析(需提供表型信息)	剔除已知影响变量, 关注潜在表型变量是否影响菌群
	⑤随机森林-ROC曲线(分组=2, 每组样本数 ≥ 30)	构建分类模型

信息分析流程

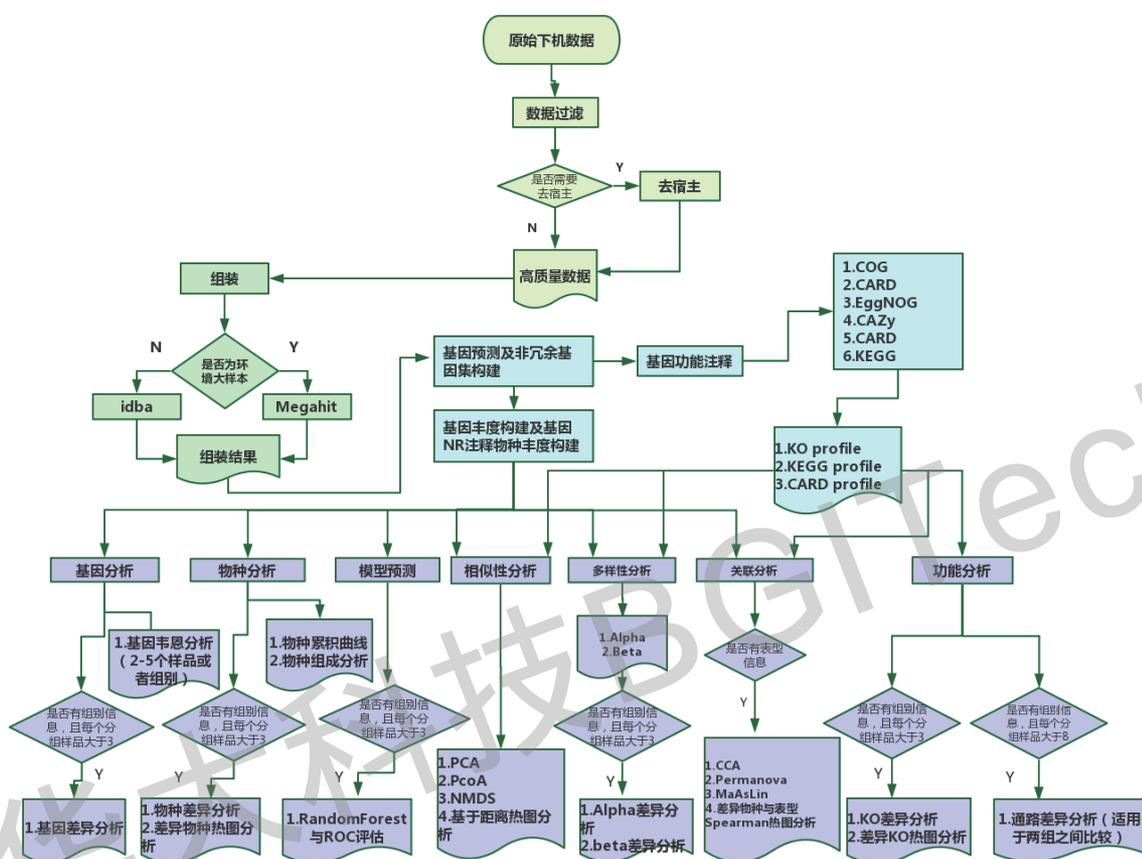


图2 宏基因组测序信息分析流程图

技术参数

样品要求: 基因组DNA, 无降解或轻微降解

样品总量: $m \geq 1\mu\text{g}$

样品浓度: $c \geq 12.5\text{ng}/\mu\text{L}$

测序策略: 小片段文库; PE150测序

推荐数据量: 简单环境 (哺乳动物肠道等) 建议不低于5Gb/样本; 复杂环境 (土壤、海洋等) 建议不低于10Gb/样本

交付周期: 40个工作日

案例分析

案例: 肠道微生物代谢的长链饱和脂肪酸可增强大鼠的结肠运动

胃肠功能紊乱是患有功能性胃肠病和其他肠道疾病的患者中常见的主要症状, 慢性或反复发作的胃肠运动障碍严重影响患者的生活质量。目前对胃肠功能紊乱的发病机制研究并不完整, 这限制了胃肠功能紊乱的个体化和精准有效药物的开发。前期研究表明肠道菌群在胃肠运动紊乱中起关键作用, 但肠道菌群与宿主胃肠道运动相互作用及其作用机制尚不明确。

新生儿母亲分离 (NMS) 的啮齿动物模型是一个成熟的长期结肠功能障碍模型, 主要表现是结肠动力增强。有研究表明小鼠和大鼠在受NMS处理后会表现出肠道微生物紊乱; 而NMS模型中肠道菌群变化与胃肠动力增强的关系仍有待研究。本研究通过大鼠NMS模型和无菌模型, 研究肠道菌群在胃肠功能紊乱中的作用, 并深入挖掘其作用机制。

★ 研究方案

1.NMS模型大鼠肠道代谢水平的改变

分组: 对照组VS NMS组

研究方法: 代谢组学

通过代谢组学检测NMS大鼠肠道代谢水平的变化, 得到与NMS模型大鼠胃肠动力改变相关的代谢产物 (SLCFA)

2.实验验证代谢水平的改变对大鼠胃肠的影响

体外实验: 通过器官浴系统研究不同剂量C17:0, C18:0以及游离LCFA受体拮抗剂对体外结肠环状肌收缩的影响

体内实验: 不同剂量C17:0和C18:0对正常大鼠 (n=8/group) 胃肠动力的影响

3.NMS模型大鼠肠道菌群及功能的改变

分组: 对照组VS NMS组

研究方法: 宏基因组学

通过宏基因组学检测NMS大鼠肠道菌群及基因变化, 得到与对应代谢产物 (SLCFA) 相关的菌群及基因

4.验证肠道菌群改变对胃肠运动及代谢产物的影响

伪无菌大鼠模型: 将对照组和NMS组大鼠粪便分别移植到伪无菌大鼠体内, 并检测移植后大鼠的粪便菌群、基因及胃肠运动和对代谢物水平的改变

抗生素治疗: 新霉素调节NMS大鼠肠道菌群, 以及治疗后胃肠动力、肠道菌群、代谢水平的变化

★ 主要结果

1.代谢组分析发现结肠中过量的SLCFAs与NMS组肠动力增强有关, 体内体外验证结果表明SLCFAs刺激大鼠结肠收缩并增强排便频率 (如图3)。

2.宏基因组分析发现NMS组中差异的粪便微生物与SLCFAs生成相关。粪便移植验证结果表明NMS供体的粪便微生物使无菌小鼠排便频率及结肠内SLCFAs含量增高。用新霉素改善NMS组肠道菌群失调可有效降低肠动力及SLCFAs产生 (如图4)。

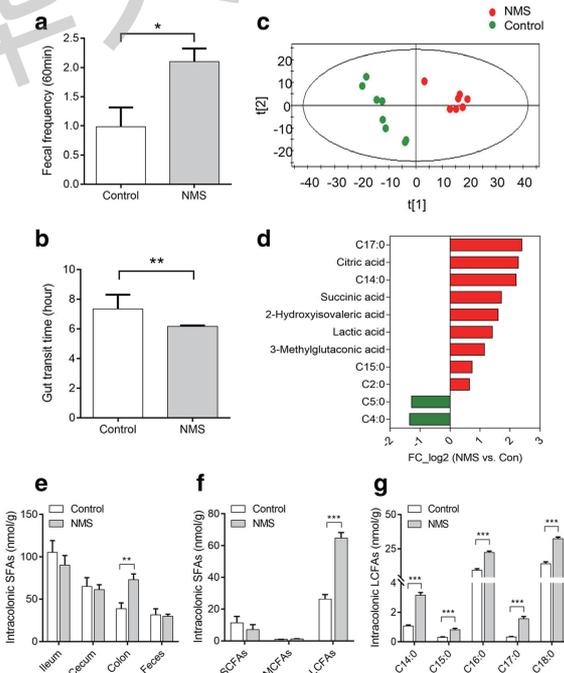


图3 与对照组相比, NMS大鼠的肠道运动性增强, 结肠内饱和和长链脂肪酸 (SLCFAs) 含量增加

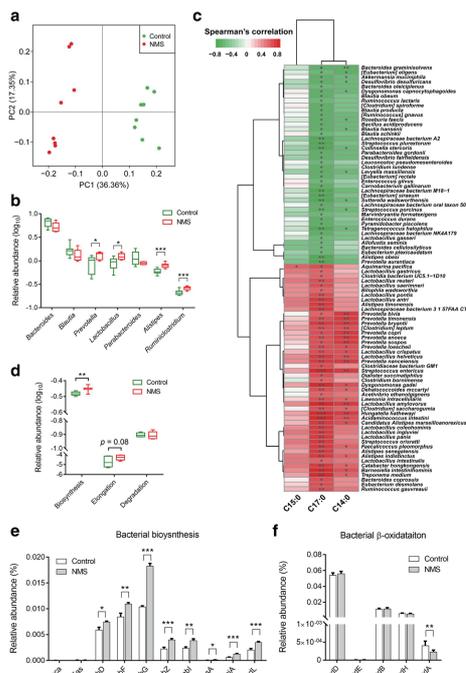


图4 粪便微生物组的改变与NMS大鼠中的 SLCFAs生成密切相关

★ 参考文献

Zhao L, Huang Y, Lu L, et al. Saturated long-chain fatty acid-producing bacteria contribute to enhanced colonic motility in rats[J]. Microbiome, 2018, 6(1): 107.

► 常见问题

Q: 如果存在较大的宿主污染, 且没有宿主基因组的参考序列可以进行宏基因组测序吗?

A: 不可以, 如果宿主的基因组序列在环境DNA中的量比较多, 测序之后, 我们没有办法通过已知的宿主基因组的序列来去污染, 会对最后的分析结果造成很大影响, 而且可用的数据量会很少; 但是如果在提取的过程, 宿主基因组的污染的量很少, 后期的数据分析还是可用的, 但是会存在一定的风险。

Q: 进行多样品比较分析的条件?

A: 样品数量超过2个, 且有比较意义的, 均可进行多样品比较分析。但若要对明显分组的样品进行比较分析, 建议至少2个组 (每组至少10个样品以平衡个体差异)。

Q: 宏基因组研究推荐多少样本量?

A: 宏基因组样本量需求跟研究目的直接相关。如果侧重于功能挖掘, 一个或几个样本即可; 如果侧重于组间差异分析, 如疾病相关的肠道菌群研究, 一般需要较大的样本量, 每组样本在50个左右, 可参考对应的经典研究案例。

Q: 对于有参考基因集的微生物群落研究 (如人肠道微生物、小鼠肠道、猪肠道), 是否可以直接比对已有的参考基因集进行物种、基因注释, 这种操作有什么优势/劣势?

A: 可以。

优势: 已经公开发表的参考基因集一般是基于大样本量数据构建而成, 基因集覆盖范围广, 质量高; 直接比对已有的数据库不受单次测序数据质量的影响, 结果准确, 且信息分析周期较短。既适用于一般的商业项目, 也适用于大型合作项目初筛。

劣势: 公开发表的参考基因集虽然覆盖范围较广, 来源特殊的样本可能和参考基因集差别较大; 在大型研究中, 可能也会忽略群落特异的 (新的) 基因或物种产生的影响; 因此在大型项目中, 可以先基于已发表的参考基因集进行初筛, 同时基于新的项目数据进一步完善已发表的参考基因集。

► 华大合作发表文章 (部分)

样本来源	发表时间	发表期刊	文章名称
小鼠模型	2020	Nature Communications	Gut microbiota mediates intermittent-fasting alleviation of diabetes-induced cognitive impairment
小鼠肠道	2019	Molecular Psychiatry	Transplantation of microbiota from drug-free patients with schizophrenia causes schizophrenia-like abnormal behaviors and dysregulated kynurenine metabolism in mice
根系	2019	Nature Biotechnology	NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice
人肠道	2019	Nature Biotechnology	1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses
小学生肠道	2019	Microbiome	Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children
人肠道	2018	Brain, Behavior, and Immunity	Alteration of gut microbiota-associated epitopes in children with autism spectrum disorders
根际	2018	Nature Communications	The structure and function of the global citrus rhizosphere microbiome
猪肠道	2018	Cell Host & Microbe	A microbiota-derived bacteriocin targets the host to confer diarrhea resistance in early-weaned piglets

华大科技BGITech

产品
简介

Product
Profile

华大科技BGITech

蛋白质组学

定性蛋白质组学	087
定量蛋白质组学	090
修饰蛋白质组学	097
多肽组学	105
多组学关联分析	106
送样要求	111

定性蛋白质组学

蛋白质全谱分析

蛋白质全谱分析是指蛋白质组分分析，是以完整的组织、血液或体液等复杂样本为研究对象，目的在于识别出样品中尽可能多的肽段和蛋白质。基于质谱技术的全谱分析，可为蛋白质高通量的定量和修饰分析提供参考信息。此外，将全谱分析数据和转录组数据结合，可相互补充并共同验证基因注释。

技术流程

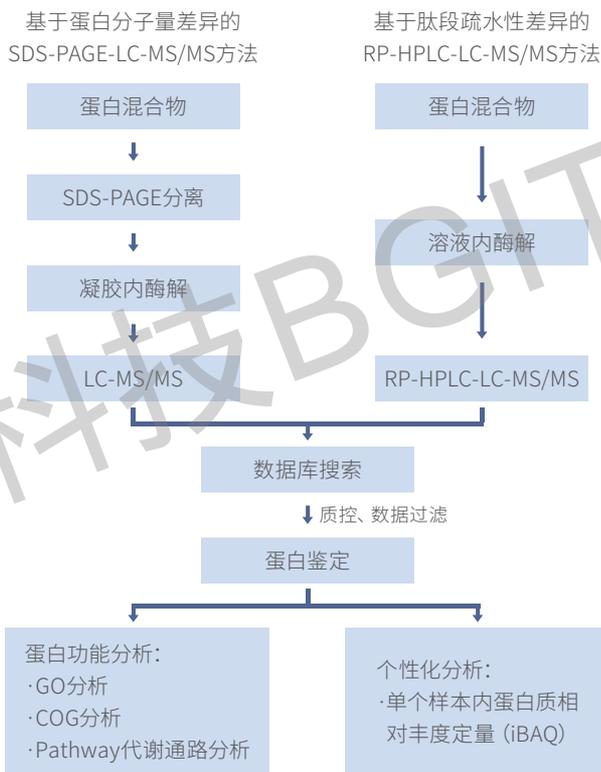


图1 蛋白全谱分析技术流程

蛋白质胶点和胶条鉴定

胶点鉴定是针对双向电泳胶中指定的胶点进行质谱鉴定。首先对胶点样品进行脱色 (decolorization)、酶解 (enzymolysis)，然后用质谱 (LC-MS/MS) 进行检测，得到的质谱图通过数据库搜索比对的方法鉴定蛋白质。

胶条鉴定是针对SDS-PAGE胶中的指定条带或者是蛋白种类比较少 (几十到一百左右) 的蛋白液进行质谱鉴定。先对胶条或蛋白样品进行酶解，然后用质谱 (LC-MS/MS) 进行检测，利用得到的质谱图，通过数据库搜索比对的方法鉴定目标条带中的蛋白质。

技术流程

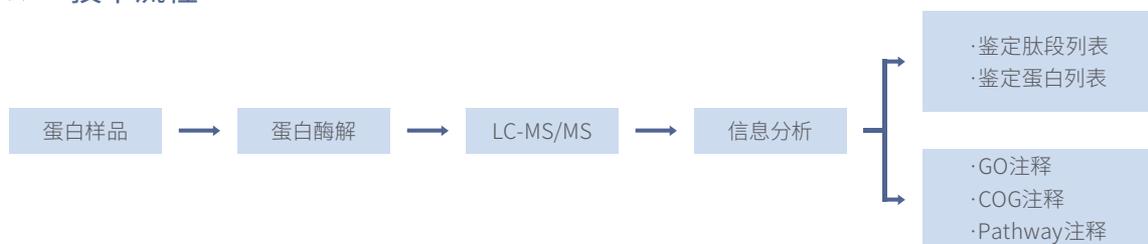


图2 胶点胶条技术流程

► 信息分析流程

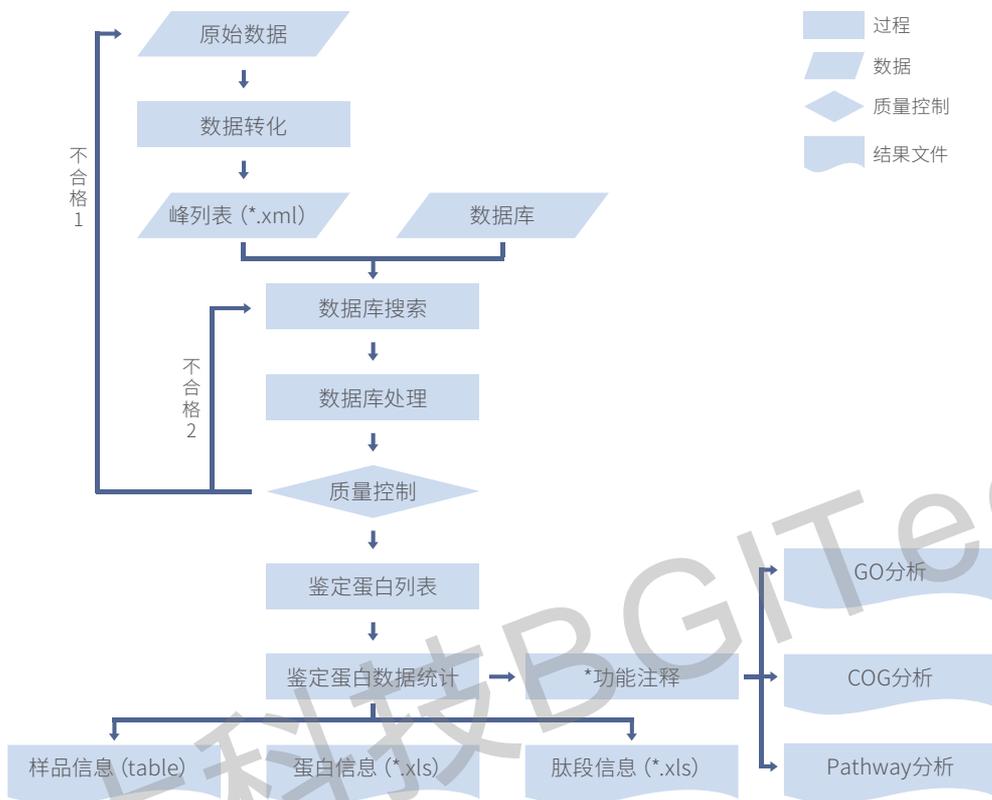


图3 蛋白鉴定信息分析流程, *表示胶条产品特有

●● 技术参数 ●●

产品类型		仪器平台	项目周期
定性蛋白质组学	蛋白全谱分析	Q Exactive HF或TripleTOF 5600	30个自然日 (≤2样本)
	胶点胶条鉴定	Q Exactive HF或TripleTOF 5600	14个自然日 (≤20个样本)

●● 案例分析 ●●

案例: 基于质谱技术对大枣的花进行蛋白质组全谱分析

研究内容: 大枣花蛋白质组学研究 发表期刊: PeerJ 研究策略: 蛋白全谱分析

★ 研究背景

目前,关于大枣花中表达的蛋白质数量和特性尚未研究清楚。本研究通过过滤器辅助样本制备,经过胰蛋白酶消化后,使用LC-MS/MS对枣花进行蛋白质组学分析,对大枣花进行蛋白质组学全谱分析。通过整合蛋白质组学和转录组学分析的方法,可以用于表征植物花卉生物学。

★ 研究思路

实验材料: 9年生枣树上收集的萌芽花

实验流程:



图4 SCX-LC-MS/MS进行枣花蛋白质全谱分析实验流程

★ 研究结果

1.从80,394张谱图、42,935种特异性肽段中, 鉴定到7,853种蛋白质, 约占枣花转录组全谱模型的30%。

2.蛋白组&转录组比较分析发现, 质谱鉴定到的94%的蛋白可以在mRNA水平检测到; 蛋白质组学分析鉴定的7,494个基因中, RPKM值小于1和5的基因分别有129 (1.6%) 和473 (6.0%) 个。

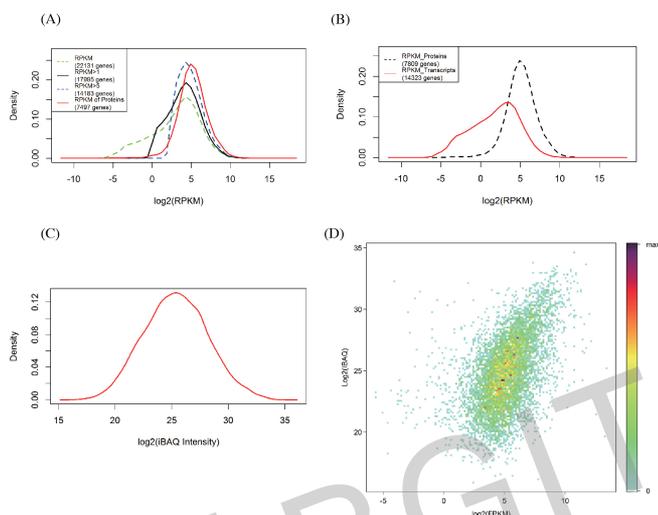


图5 蛋白质组学和RNA-Seq数据比较分析

A. 不同级别RPKM值分布; B. 蛋白质组中鉴定和未鉴定基因的RPKM值分布; C. 蛋白丰度分布; D. iBAQ与RPKM值密度散点图

3.蛋白丰度情况: 丰度最高的224种蛋白质 (2.85%) 占枣花蛋白质组丰度的50%, 691种最丰富的蛋白质占总蛋白丰度的80%, 41种最丰富的蛋白质占总蛋白的20%。

★ 研究结论

在枣花蛋白质组中, 最丰富的蛋白质是核糖体蛋白, 占蛋白质总质量的17%, 反映了花器官蛋白质组成中的不同生物学功能。蛋白质组学和转录组学分析显示, 具有较高表达水平的基因很可能被翻译成活性蛋白质。

★ 参考文献

Chen R, Chen G, Huang J. Shot-gun proteome and transcriptome mapping of the jujube floral organ and identification of a pollen-specific S-locus F-box gene[J]. PeerJ, 2017, 5:e3588.

典型案例汇总

发表时间	发表期刊	文献名称	研究策略
2019	Molecular & Cellular Proteomics	Improvements to the rice genome annotation through large-scale analysis of RNA-Seq and proteomics data sets	蛋白全谱
2019	Experimental Eye Research	Proteomic evidence that ABCA4 is vital for traumatic proliferative vitreoretinopathy formation and development	胶条
2017	PeerJ	Shot-gun proteome and transcriptome mapping of the jujube floral organ and identification of a pollen-specific S-locus F-box gene	蛋白全谱

定量蛋白质组学

定量蛋白质组学 (Quantitative Proteomics) 是对一个基因组表达的全部蛋白质或一个复杂混合体系内所有蛋白质进行精确鉴定和定量。蛋白质组定量可用于筛选和寻找任何因素引起的样本之间的差异表达蛋白, 结合生物信息学揭示生命体生理病理功能, 同时也可对某些关键蛋白进行定性和定量分析。

基于质谱检测技术的蛋白质组定量技术可以分为两大类: 标记定量技术 (Labeling Quantitation), 包括iTRAQ、IBT、TMT等; 非标记定量技术 (Label-free quantitation), 包括Label-free、DIA、MRM等。

蛋白标记定量分析-iTRAQ/IBT/TMT

iTRAQ (Isobaric Tags for Relative and Absolute Quantitation)、IBT (Isobaric Tags) 和TMT (Tandem Mass Tags) 分别是由AB SCIEX公司、华大基因和Thermo Fisher Scientific公司研发的体外同种同位素标记的相对与绝对定量的体外标记技术。iTRAQ、IBT、TMT这三种肽段体外标记试剂, 通过与氨基酸N端或赖氨酸侧链基团的氨基结合, 实现肽段的标记, 通过高精度质谱仪串联分析, 实现多个样本之间蛋白组表达量差异比较。

技术路线



图1 蛋白标记定量分析技术流程

信息分析流程

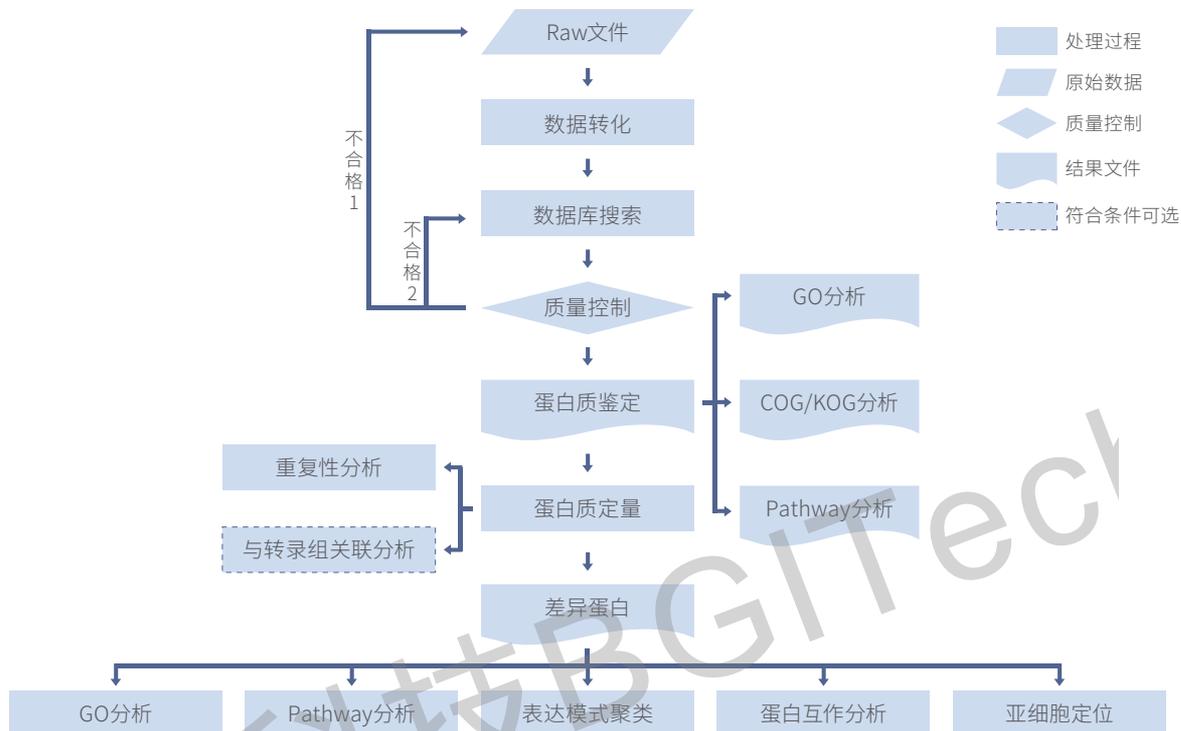


图2 蛋白标记定量信息分析流程

蛋白DIA定量分析

蛋白DIA定量分析主要采用数据非依赖性采集模式 (Data Independent Acquisition, DIA), 对特定质量范围内的所有母离子进行碎裂, 采集所有母离子的碎片离子信息进行定性和定量蛋白质组学分析。相比于iTRAQ、IBT、TMT等采用数据依赖性扫描模式 (Data Dependent Acquisition, DDA) 的蛋白定量技术, DIA能获得窗口范围内所有的肽段的碎片离子信息, 具有更高的数据挖掘性和检测通量, 在定量准确性和技术重复性方面也有一定优势。

技术原理

基于数据非依赖性采集 (DIA) 的蛋白定量技术, 首先需对样品进行传统的较高深度的DDA数据采集, 构建一个尽可能齐全的高质量肽段的谱图库 (Spectral Library); 之后是DIA模式采集, 在DIA采集中指定连续质荷比 (m/z) 窗口内的所有肽段都进行碎裂, 并采集相应高分辨率的二级质谱图; 在肽段谱图库的参照下, 对复杂的二级谱图进行匹配提取解析, 得到肽段和蛋白质的定性和定量结果。

技术路线

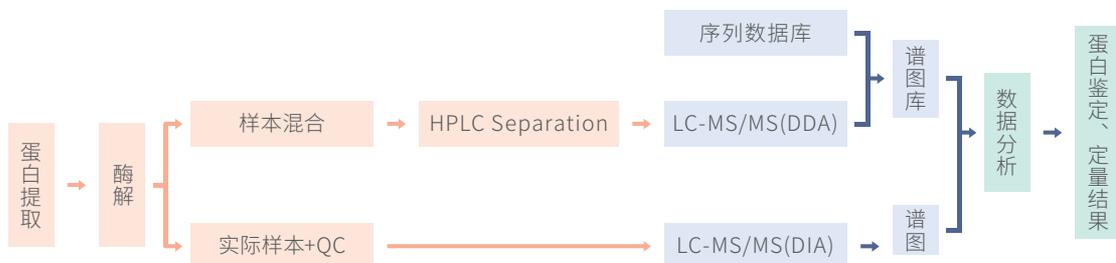


图3 蛋白DIA定量分析技术流程

●● 目标蛋白MRM定量分析 ●●

目标蛋白MRM定量分析是指基于质谱的多反应检测技术 (Multiple Reaction Monitoring, MRM), 有目标地分析检测特定的关键蛋白质在多样本中的含量变化情况的高通量检测技术产品。通过MRM检测分析可以得到目标蛋白质的上、下调情况, 推测这些蛋白质的生物学功能或潜在分子标志物信息。

多反应监测技术, 是一种用于靶向蛋白非标记量化的研究策略。与基于免疫学技术的目标蛋白方法相比, MRM目标蛋白组学验证的方法解决了蛋白标志物验证速度慢的难题。相比于一次只能研究一个蛋白的抗体技术, MRM技术能够同时分析20个及以上数量的目标蛋白, 从而实现大样本目标蛋白的验证。

► 技术原理

MRM技术基于三重四极杆质谱仪的应用, 如下图所示, 根据理论肽段母离子 m/z 和碎片离子 m/z , 选择特异性的母离子-子离子对 (Transition), 使得在Q1只筛选指定的母离子通过, 进入到Q2碰撞室进行母离子碎裂, 然后Q3只筛选指定的碎片离子通过, 从而实现Transition的筛选。通过母离子和子离子的双重选择, 既提高了选择特异性, 又去除了干扰离子, 降低化学背景噪音, 提高了灵敏度。MRM技术可以同时进行上百个Transition的筛选检测, 实现高通量分析。

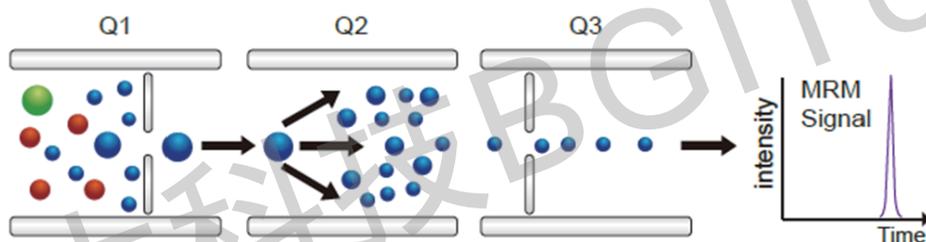


图4 目标蛋白MRM技术原理介绍

表1 目标蛋白MRM与抗体技术对比

产品类型	抗体技术	MRM
可行性	能否获得抗体	能否实现质谱方法建立
时间	抗体制备需数月或者更长	方法建立仅需2-4周
通量	一个抗体仅能检测一种蛋白	可同时对上百个蛋白进行检测
准确度	抗体质量参差不齐, 图像信号影响因素较多, 不同实验室之间的结果差异大, 缺乏统一标准	基于稳定统一的质谱物理信号, 定量结果更加准确
灵敏度	检测灵敏度较高	灵敏度中等
重复性	重复性较低	自动化检测, 重复性高

技术流程

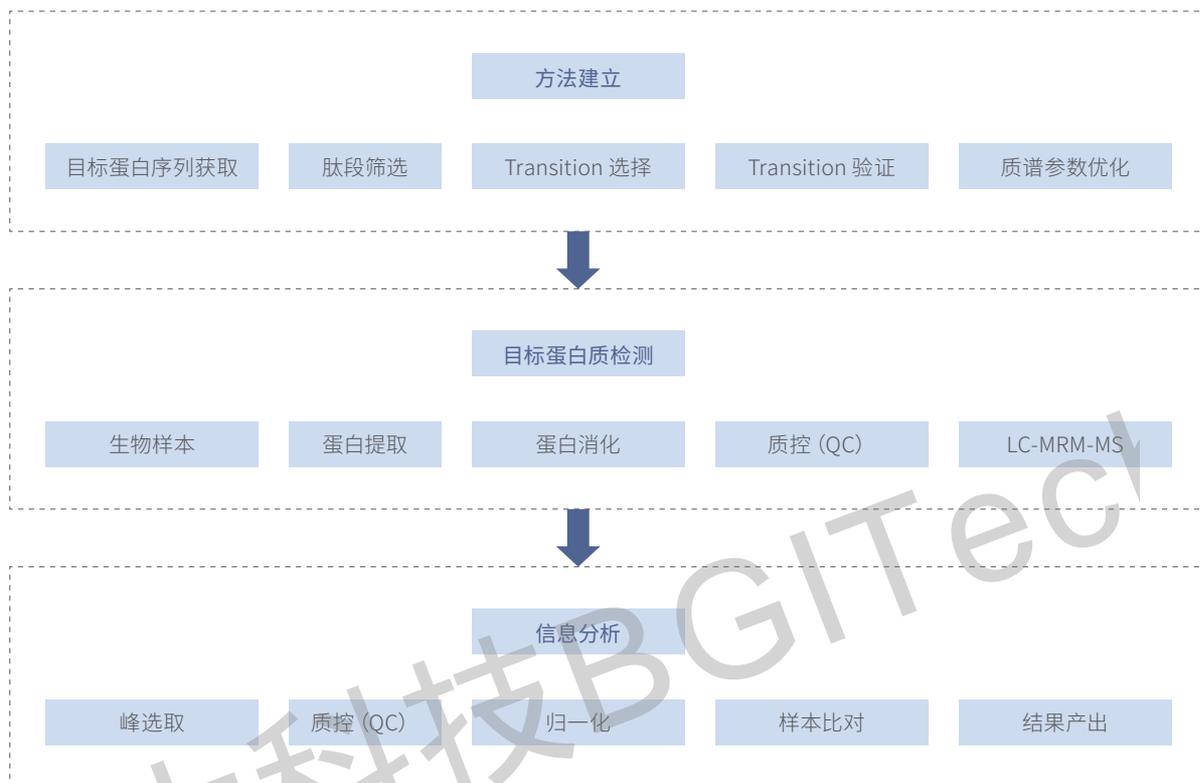


图5 目标蛋白MRM技术流程

蛋白组定量技术对比

表2 蛋白定量分析产品技术参数

产品类型		仪器平台	项目周期
定量蛋白质组学	iTRAQ	Thermo Q Exactive HF-X/HF	26个自然日 (≤1批次)
	IBT	Thermo Q Exactive HF-X/HF	26个自然日 (≤1批次)
	TMT	Thermo Q Exactive HF-X/HF	26个自然日 (≤1批次)
	DIA	Thermo Q Exactive HF-X/HF	42个自然日 (≤25个样本)
	MRM	QTRAP 6500+/TripleTOF 5600	48个自然日 (≤8个样本)

案例分析

案例一：现代玉米育种中蛋白质组的差异表征

研究内容：玉米育种中蛋白质组 发表期刊：Molecular & Cellular Proteomics 研究策略：iTRAQ+转录组

★ 研究背景

现代玉米育种的成功已经在上个世纪农业生产力的巨大提高中得到了证明。尽管科学家已经广泛研究了从热带到温带地区的玉米在适应过程中的潜在遗传变化，但仍缺乏对蛋白质组和mRNA表达水平的一致性的研究。

★ 研究思路

1.使用高通量高分辨率质谱 (MS) 在98个玉米自交系 (包括TST、NSS、SS三个基因组亚群) 的关联组中进行定量蛋白质组研究。

2.将定量蛋白质组学与RNA测序数据相结合,以揭示玉米从热带到温带地区适应过程中mRNA和蛋白质水平的一致性。

3.RNA-Seq分析中使用的所有近交系包括两个独立的生物学重复,而iTRAQ蛋白质组学分析中使用了7个代表性近交系包括两个独立的生物学重复,另外91个具有一个生物学重复的近交系 (总共15个8-plex iTRAQ)。

★ 研究结果

1.总共2,750种蛋白质被定义为具有唯一的ensembl基因ID并用于进一步分析。

2.蛋白质组亚型可能类似于基因组亚型,然而转录组亚型和基因组亚型存在比较大的不一致性。

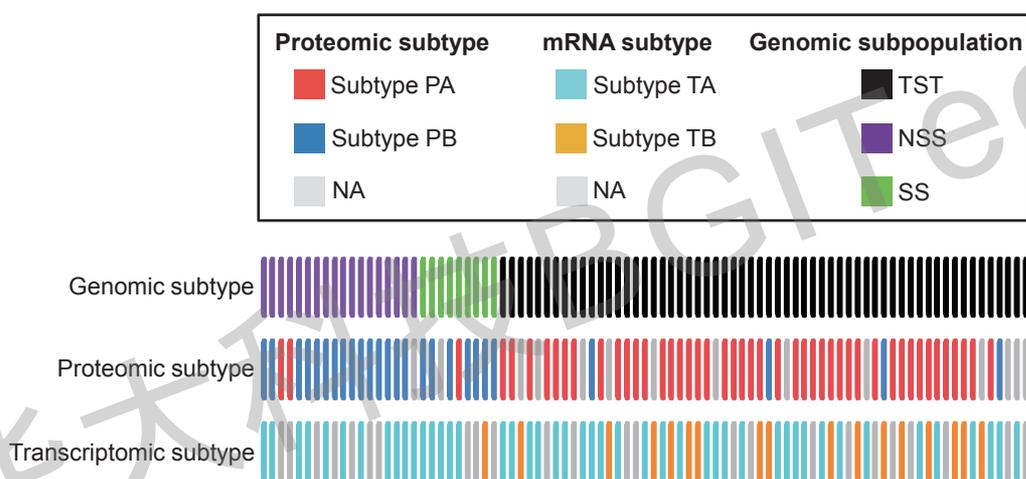


图6 蛋白质组或转录组亚型与基因组亚型的关联

3.在基因共表达网络分析 (WGCNA) 中,蛋白质和mRNA网络显示出适度的相关性 (仅有少量模块存在一致性)。虽然一些模块在mRNA和蛋白质网络之间存在重叠,但是仍有许多特定的蛋白质网络的模块。

4.在同一基因簇中,mRNA表达水平未随着基因组亚群的不同而产生变化。相反,在一些基因簇中,蛋白质表达水平在不同基因组亚群之间存在差异。

5.通过对5个基因簇中的其中3个进行GO富集分析发现,4号基因簇主要富集了编码核糖体蛋白或翻译相关的基因。1号基因簇中,mRNA在不同基因组亚群中表达水平相似,但在蛋白质层面,基因在TST亚群中有更高水平的表达。

★ 研究结论

上述研究结果表明,玉米通过有意调节涉及特定生物过程的某些基因的mRNA和蛋白的不同表达水平,适应了从热带到温带地区的气候地理变化。

★ 参考文献

Jiang LG, Li B, Liu SX. Characterization of proteome variation during modern maize breeding[J]. Molecular & Cell Proteomics, 2019,18(2):263-276.

案例二：蛋白定量DIA用于研究海马体不同区域应激反应机制

研究内容：海马体不同区域应激反应机制

发表期刊：Biological Psychiatry

研究策略：DIA

注：SWATH技术与DIA技术原理相同，同为二级谱采集荷质比范围内全部母离子的碎片离子信息，但不同仪器公司开发此技术时使用的名称有别，故可以理解为同一种技术。

★ 研究背景

作者为了研究海马体不同区域压力应答的分子效应，在已有的转录组研究基础之上，结合正在迅速发展的蛋白质组学方法，希望揭示生物学和临床相关的生物标志物。

★ 实验设计

对比四种压力下小鼠的海马组织背侧区域 (dHC) 和腹侧区域 (vHC) 蛋白及转录组差异。

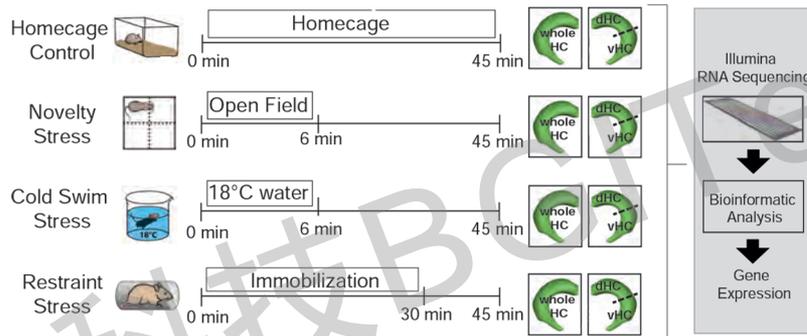


图7 实验设计

压力条件：家庭控制 (homeage control)、新颖性应力 (novelty stress)、冷游压力 (cold swim stress)、约束应力 (restraint stress)
海马组织不同区域：背侧区域 (dHC)、腹侧区域 (vHC)

★ 主要发现

1. 选择C57BL/6J雄性小鼠的海马组织不同区域：背侧区域 (dHC)、腹侧区域 (vHC) 进行测序，通过相互作用网络分析，dHC主要富集参与谷氨酸能突触传递、长时增强效应、内吞作用以及细胞骨架的调控通路，vHC主要富集神经配体-受体激活、含血清素的神经突触和氨基丁酸能突触的通路。

2. 使用SWATH技术，发现513个蛋白质在dHC和vHC之间存在差异。转录组和蛋白质组联合分析发现，在高富集的mRNA/蛋白质对中，Wfs1和Epha7被证实富含于海马CA1神经元中，说明蛋白质组学分析能够解决异质性脑切除中亚区域特异性表达差异。此外，作者选择在dHC中高富集的MOG (Myelin Oligodendrocyte Glycoprotein) 进行WB验证实验，实验结果和组学结果一致。

3. 通过蛋白互作网络分析，发现与dHC相比，在vHC观察到更大的相互作用网络，其特有表观遗传学簇Kdm6b响应vHC中的所有压力因素。WB检测冷水游泳应激后H3K27me3的水平，发现dHC中H3K27me3表达水平不变，vHC中H3K27me3表达明显下调。

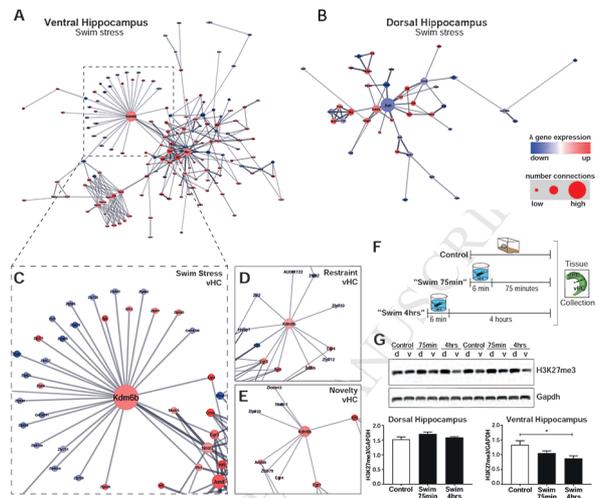


图8 腹侧海马中应激反应的表观遗传网络

★ 参考文献

Floriou-Servou A, von Ziegler L, Stalder L, et al. Distinct proteomic, transcriptomic, and epigenetic stress responses in dorsal and ventral hippocampus[J]. *Biological Psychiatry*, 2018, 84(7):531-541.

●●● 典型案例汇总 ●●●

发表时间	发表期刊	文献名称	研究策略
2019	Molecular & Cellular Proteomics	Characterization of proteome variation during modern maize breeding	iTRAQ定量+转录组
2019	Biotechnology for Biofuels	Proteomic profiling and integrated analysis with transcriptomic data bring new insights in the stress responses of <i>Kluyveromyces marxianus</i> after an arrest during high-temperature ethanol fermentation	iTRAQ定量+MRM定量
2019	Nature Ecology & Evolution	Rapid evolution of protein diversity by de novo origination in <i>Oryza</i>	MRM定量
2018	Analytical Chemistry	Reagents for isobaric labeling peptides in quantitative proteomics	IBT定量
2018	Frontiers in Plant Science	Targeted proteomics approach toward understanding the role of the mitochondrial protease FTSH4 in the biogenesis of OXPHOS during arabidopsis seed germination	MRM定量
2018	Biological Psychiatry	Distinct proteomic, transcriptomic, and epigenetic stress responses in dorsal and ventral hippocampus	DIA定量
2018	Molecular & Cellular Proteomics	Development of a gill assay library for ecological proteomics of threespine sticklebacks (<i>Gasterosteus aculeatus</i>)	DIA定量
2017	Nature Communications	Proteogenomic integration reveals therapeutic targets in breast cancer xenografts	iTRAQ定量+基因组
2017	Nature Communications	Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry	DIA定量

修饰蛋白质组学

蛋白质翻译后修饰 (Post-translational Modification, PTM) 指蛋白质经生物合成后的化学基团共价修饰及酶学修饰, 其修饰位点可发生在蛋白质氨基酸的侧链基团或其N端、C端。各种翻译后修饰在细胞生物过程中起了重要的调节调控作用, 如信号通路转导、基因转录编码活性调控、细胞通讯和免疫识别、黏附、蛋白质降解等。

由于翻译后修饰蛋白质在样本中含量低且动态范围广, 关于翻译后修饰的蛋白质组学研究极具挑战性, 需针对不同修饰类型采用不同的富集和质谱检测方法。亲和富集、多维分离等技术与生物质谱的结合为翻译后修饰蛋白质组学的发展提供了契机。目前, 已进行规模化研究的蛋白质修饰类型主要有磷酸化 (Phosphorylation)、乙酰化 (Acetylation)、糖基化 (Glycosylation)、泛素化 (Ubiquitination) 等修饰。

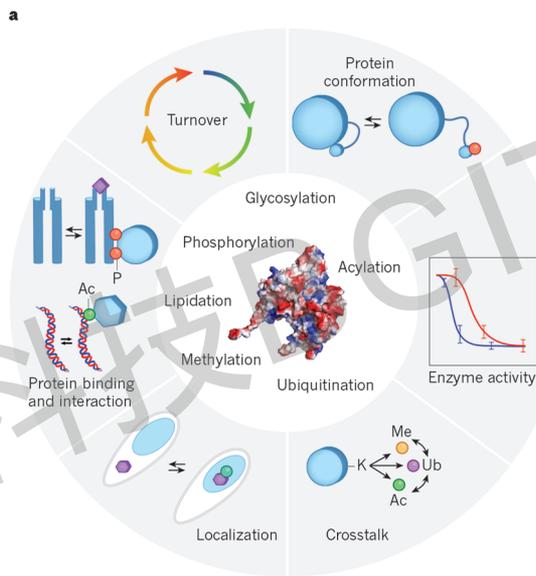


图1 细胞内翻译后修饰类型
Aebersold R, Mann M. *Nature*, 2016.

磷酸化修饰

蛋白质磷酸化 (Phosphorylation) 是最常见、最重要的一种蛋白翻译后修饰方式, 发生在蛋白质丝氨酸、苏氨酸以及酪氨酸上, 它由两个作用相反的酶系——蛋白激酶和磷酸酶进行调控。蛋白磷酸化和去磷酸化几乎调节生命活动的整个过程, 包括细胞的增殖、发育和分化、信号转导等。

通过质谱方法进行蛋白质磷酸化分析, 不仅可以检测磷酸化蛋白质, 还可以进一步确认蛋白质内发生磷酸化的肽段以及位点。磷酸化修饰产品主要包括磷酸化蛋白鉴定、磷酸化蛋白定量及目标磷酸化蛋白PRM定量。

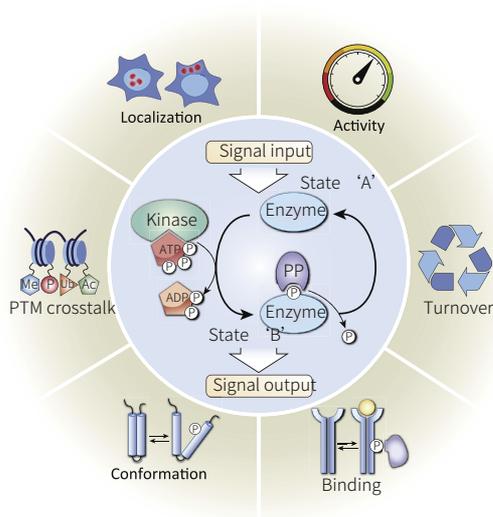


图2 蛋白磷酸化修饰生理功能
Humphrey SJ, et al. *Trends Endocrinol Metab*, 2015.

► 技术流程

纯化蛋白磷酸化鉴定/磷酸化蛋白全谱鉴定

纯化蛋白磷酸化鉴定是指对经过分离纯化或IP等实验富集的目标蛋白质（目标蛋白质纯度>80%）进行质谱鉴定，检测目标蛋白质的磷酸化位点。

磷酸化蛋白全谱鉴定是以组织、细胞为研究对象，目的在于鉴定样品中磷酸化蛋白质以及相应的磷酸化位点，阐述磷酸化蛋白质的生理意义及生物意义。

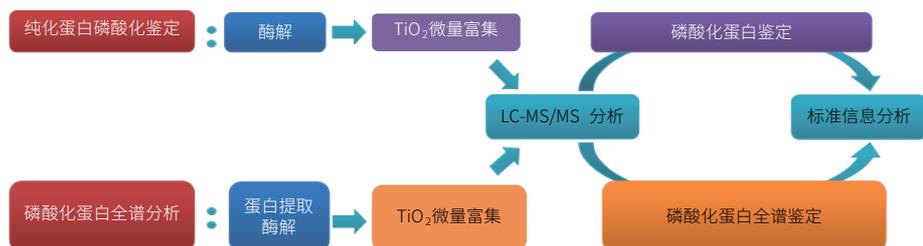


图3 纯化蛋白磷酸化鉴定/磷酸化蛋白全谱鉴定技术流程

磷酸化蛋白定量

磷酸化蛋白定量是以整个生物样本为研究对象，基于高精度、高灵敏度质谱仪，对生物体的磷酸化蛋白进行鉴定并相对定量，从而达到研究生物体内磷酸化蛋白质组动态变化与机体表型改变之间关系的目的。

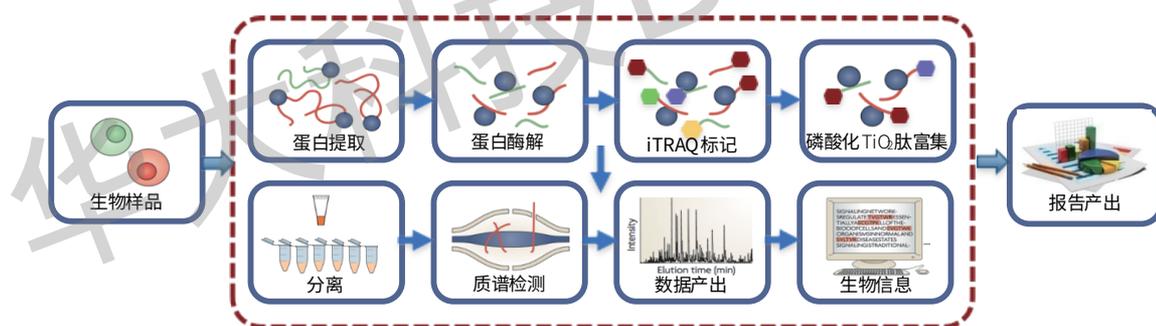


图4 磷酸化蛋白iTRAQ定量技术流程

目标磷酸化蛋白PRM定量

PRM (Parallel Reaction Monitoring)，平行反应监测技术是一种基于高分辨质谱仪对目标蛋白、目标肽段进行选择检测，从而实现目标蛋白质/肽段进行相对定量的技术。PRM技术首先利用一级质谱的选择性将目标蛋白的肽段筛选出来，接着在碰撞池中对母离子进行碎裂，最后利用高分辨的二级质谱对目标肽段的所有碎裂离子信号进行采集，即可对目标肽段进行定量。

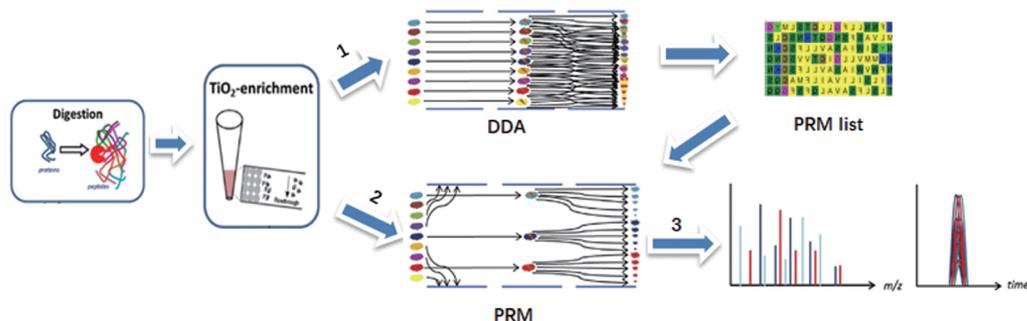


图5 目标磷酸化蛋白PRM定量技术流程

- 1.使用Orbitrap 系列质谱仪进行DDA（数据依赖性采集）模式采集数据，筛选出可检测的目标磷酸化肽段PRM list；
- 2.对靶标磷酸化肽段离子进行PRM模式检测，获得子离子的信号；
- 3.通过分析软件（如Skyline）进行特征子离子的筛选和峰面积积分，得到高分辨高特异性的磷酸化肽段相对定量值。

乙酰化修饰

蛋白质乙酰化修饰是在乙酰转移酶的作用下，在赖氨酸残基上添加乙酰基的过程，是一种高度保守的、可逆的翻译后修饰。从组蛋白与DNA的相互作用，到细胞周期的调控、神经退行性、癌症的发生发展等过程，都有赖氨酸乙酰化过程的广泛调控。

蛋白乙酰化分析包括乙酰化蛋白全谱鉴定及定量。乙酰化蛋白全谱鉴定是以组织、细胞为研究对象，通过蛋白提取富集，结合质谱分析，从而鉴定出样品中乙酰化蛋白质以及相应的乙酰化位点，阐述其生理意义及生物意义。乙酰化蛋白定量是基于高精度、高灵敏度质谱仪，将Label-free定量技术与特异性抗体富集技术相结合，对生物体的乙酰化蛋白进行鉴定及相对定量，从而达到研究生物体内乙酰化蛋白质组动态变化与机体表型改变之间关系的目的。

技术原理

与蛋白磷酸化修饰可通过原位磷酸化标记、灵敏的磷酸化特异性抗体等稳定有效的手段进行研究不同，对于蛋白质乙酰化研究的手段和方法十分有限。目前主要研究方法是先用赖氨酸乙酰化特异性抗体富集乙酰化肽段，再通过液相色谱质谱联用 (LC-MS) 的方法进行检测。

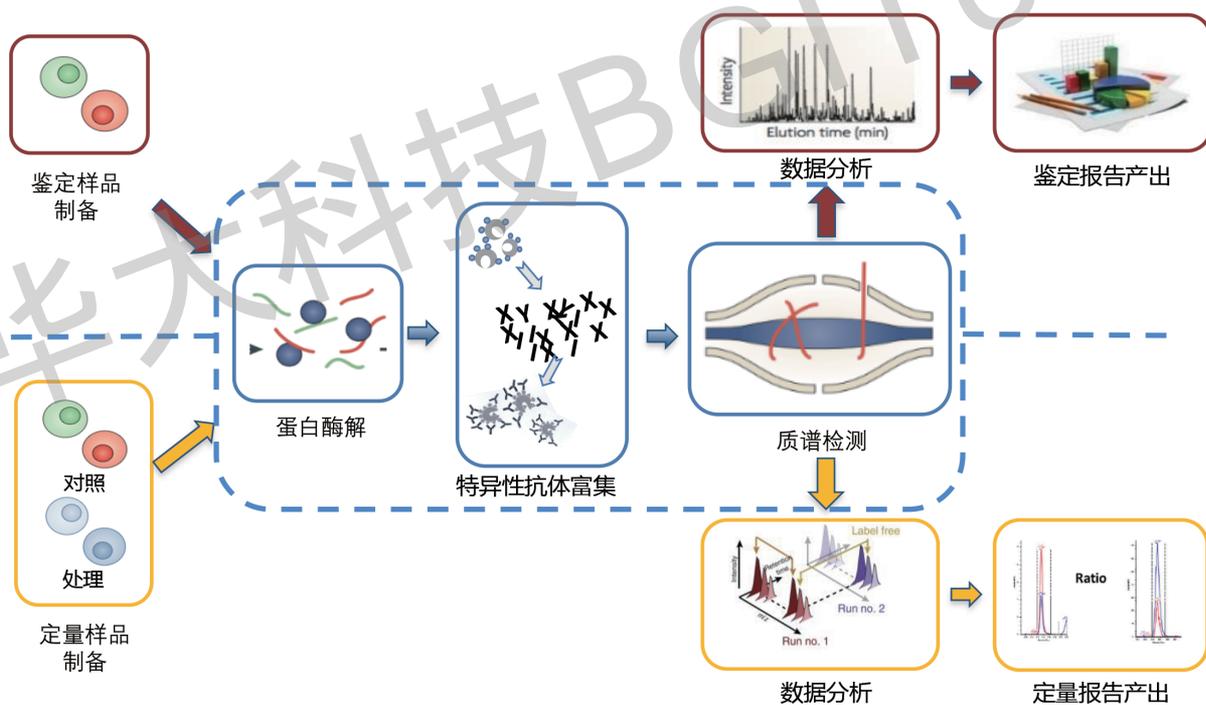


图6 乙酰化修饰分析技术路线

首先提取样本中的蛋白质，随后使用胰蛋白酶将蛋白质酶解成肽段，除盐后使用特异性抗体将乙酰化肽段富集，最后进行LC-MS/MS分析。数据经信息分析得到鉴定及定量结果。

N-糖基化修饰

糖基化修饰对蛋白质功能和结构的形成具有重要作用，在机体内细胞粘附、分子识别以及信号转导等过程均涉及糖基化蛋白质的参与。根据糖蛋白糖侧链的不同连接方式，较为常见的蛋白质糖基化修饰有O-连接糖基化(O-linked glycosylation)、N-连接糖基化(N-linked glycosylation)等。

当前国际上糖基化蛋白质组主要研究策略是利用现有的技术体系，分离富集糖蛋白或糖肽，再进行糖基化位点质量标记，从而实现大规模高通量糖蛋白及糖基化位点的鉴定。目前常用的糖蛋白分离富集技术有：凝集素亲和技术、胍化学富集法、亲水相互作用色谱法 (HILIC)、硼酸法、免疫法，五种分离富集方法比较如下：

表1 五种糖基化分离富集方法对比

方法	优点	缺点
亲水相互作用色谱法 (HILIC)	对糖基化覆盖率高、方法重现性好、易于与 MS 联用	非特异性吸附
凝集素亲和技术	选择性好	不具备普适性且作用较弱，富集效率较低
硼酸法	条件温和、反应可逆和不破坏糖链结构	易与其他二级非化学键作用有干扰
胍化学法	反应专一性高、无位点偏向性、共价键稳定	反应步骤多，条件不易控制，且操作较复杂，还有耗时较长，容易破坏糖链的结构
免疫法	专一性	缺乏通用性

华大基因质谱平台N-糖基化修饰主要采用亲水相互作用色谱法进行分离富集，主要包括N-糖基化蛋白全谱鉴定及定量两种产品。

N-糖基化蛋白全谱鉴定是以组织、细胞为研究对象，通过蛋白提取，亲水相互作用色谱法 (HILIC) 富集，结合 LC-MS/MS分析，从而大规模地鉴定出样品中N-糖基化蛋白质及其相应的糖基化位点，阐述其生理意义及生物意义。N-糖基化蛋白定量是基于高精度、高灵敏度质谱仪，将Label-free定量技术与HILIC富集技术相结合，对生物体的N-糖基化蛋白进行鉴定及相对定量，从而达到研究生物体内N-糖基化蛋白质组动态变化与机体表型改变之间关系的目的。

N-糖基化蛋白全谱鉴定和定量技术流程 (如图7)：首先从蛋白质的提取开始，经过酶解、HILIC亲水柱富集、PNGase F进行脱糖等步骤获得脱去N糖链的肽段，然后该肽段进行LC-MS/MS检测，最后获取数据进行鉴定及定量分析。其中，糖肽酶PNGaseF将N糖链切除后，发生糖基化的位点——天冬酰胺脱氨基生成天冬氨酸，造成相对分子量增加0.98Da，这种改变在数据库搜索比对时发现，从而可以确定糖基化位点。

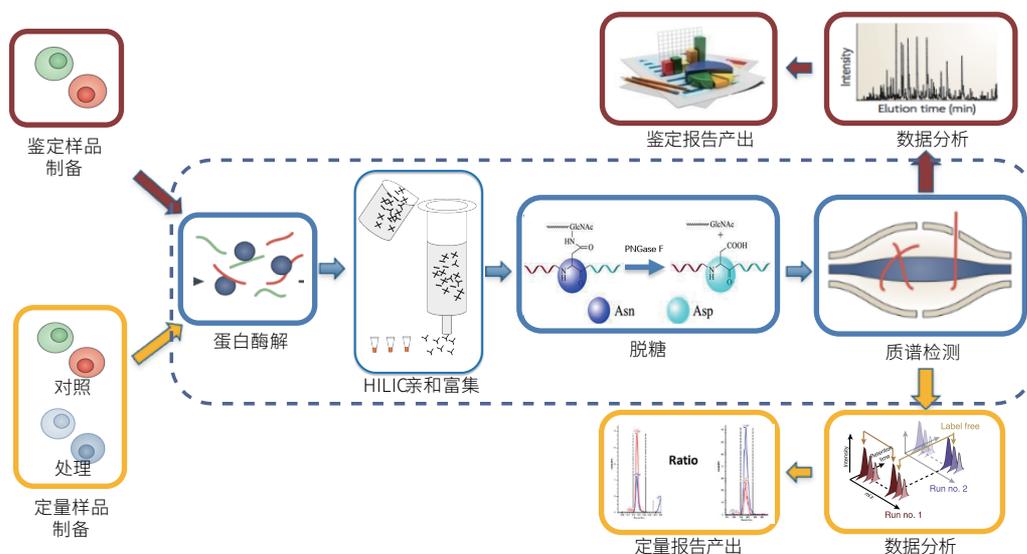


图7 N-糖基化修饰分析技术流程

技术参数

产品类型		仪器平台	项目周期	项目周期
磷酸化蛋白修饰	单个蛋白磷酸化鉴定	Orbitrap Fusion Lumos/ Q Exactive HF	TiO ₂ 富集	26个自然日 (2个样本以内)
	磷酸化蛋白全谱鉴定		TiO ₂ 富集	26个自然日 (2个样本以内)
	磷酸化蛋白iTRAQ定量		TiO ₂ 富集	30个自然日 (1批次)
	目标磷酸化蛋白PRM定量		TiO ₂ 富集	45个自然日 (≤8个样本)
乙酰化蛋白修饰	乙酰化蛋白全谱鉴定		抗体富集	28个自然日 (≤10个样本)
	乙酰化蛋白定量		抗体富集	34个自然日 (≤10个样本)
N-糖基化蛋白修饰	N-糖基化蛋白修饰全谱鉴定		HILIC亲和富集	28个自然日 (2个样本以内)
	N-糖基化蛋白修饰定量		HILIC亲和富集	34个自然日 (2个样本以内)

案例分析

案例一：利用蛋白质组与磷酸化蛋白质组研究生育能力低下的机制

研究内容: 小鼠睾丸的蛋白质组学和磷酸化蛋白质组学 发表期刊: Molecular & Cellular Proteomics
研究策略: iTRAQ+磷酸化蛋白质组

★ 研究背景

缺乏野生型p53诱导的磷酸酶1 (Wip1) 的小鼠显示出雄性生殖能力缺陷, 包括较小的睾丸, 生育能力低下和精子发生能力缺陷等等。然而, 对于这些异常的分子机制仍然不清楚。

★ 研究思路

将C57BL/6背景的10周龄雄性WT和Wip1-KO小鼠处死后解剖睾丸, 用液氮快速冷冻, 并储存在-80°C直至使用。在蛋白质组分析中, 采用了两组8-plex iTRAQ实验作为技术重复。每一个8-plex iTRAQ实验包含Wip1-WT的三个独立的生物学重复, 三个Wip1-KO的独立生物学重复, 以及合并后样品的两个技术重复。

★ 研究结果

1.在Wip1敲除小鼠中, 一共有23种显著差异蛋白和54种蛋白上的60个磷酸化位点上调, 35种显著差异蛋白和90种蛋白上的99个磷酸化位点下调。

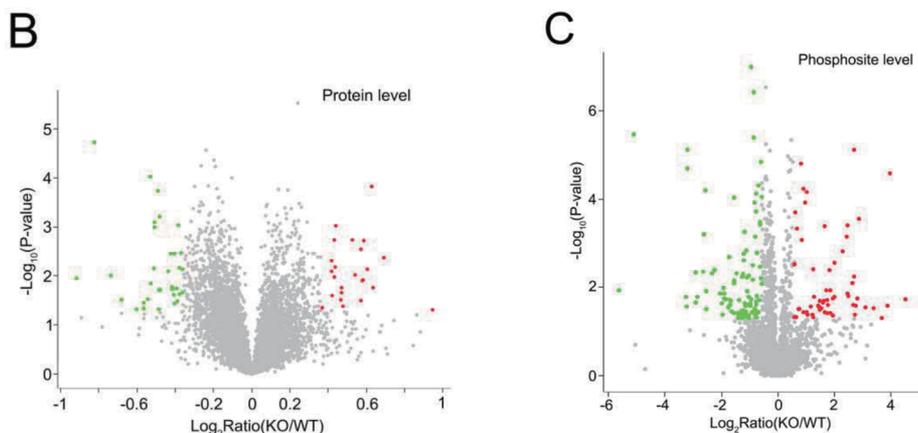


图8 B. 所有生物重复中的定量蛋白质的火山图 C. TiO₂富集后所有生物重复中定量磷酸化蛋白位点的火山图

2.对于显著差异蛋白GO富集分析显示,涵盖大多数差异表达的三个类别主要与粘附/紧密连接、精子发生、凋亡过程、免疫反应和微管/细胞骨架组织有关。进一步发现Wip1缺乏会损害连接相关蛋白表达和血睾屏障(BTB)完整性,并且BTB完整性破坏是导致生育能力降低的部分原因。

3.Wip1的敲除增加了睾丸中的细胞因子水平和生殖细胞凋亡。这些结果表明BTB完整性的破坏诱导睾丸中的生殖细胞凋亡和精子细胞损失,最终导致生育力降低。

★ 参考文献

Wei Y, Gao Q, Niu P, et al. Integrative proteomic and phosphoproteomic profiling of testis from Wip1 phosphatase-knockout mice: Insights into mechanisms of reduced fertility[J]. Molecular & Cellular Proteomics, 2019, 18 (2):216-230.

案例二: 通过乙酰化蛋白质组对比屠宰前不同处理方式对肉质的影响

研究内容: 影响肉质的乙酰化蛋白质组 发表期刊: Food Chemistry 研究策略: 乙酰化蛋白定量

★ 研究背景

为了探索蛋白质赖氨酸乙酰化在肌肉转化为肉类中的作用,对比屠宰前不同应激处理的猪肌肉的乙酰化差异。

★ 研究思路

实验样品: 18只杂交猪,约110kg,被随机分成3个处理组:①对照——屠宰前12小时送到屠宰场;②运输(压力)——屠宰前经过1小时的运输压力;③休息(恢复)——运输1小时后休息2小时再进行屠宰。每组6个样本,然后随机分成3个生物学重复。分别取屠宰后0h、0.5h、1h、4h、24h的肌肉样本进行试验分析。

实验策略: 蛋白乙酰化label-free定量

分析策略: GO、KEGG、蛋白互作分析

★ 研究结果

1.总共鉴定了681个乙酰化位点,771个乙酰化肽段,176个乙酰化蛋白。乙酰化蛋白集中于肌肉收缩、碳水化合物代谢、细胞凋亡和钙信号转导。

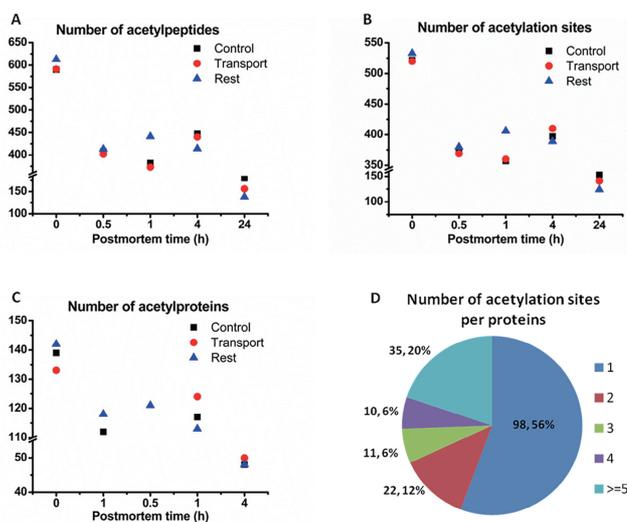


图9 乙酰化蛋白鉴定情况

2.生物信息学分析表明,预处理可能与屠宰后肌肉糖酵解及肉质形成有关,通过糖原分解或糖酵解的多种酶的乙酰化、收缩、ATP产生和钙信号相关蛋白的乙酰化,来调节僵硬、压力响应。

★ 参考文献

Zhou B, Shen Z, Liu Y, et al. Proteomic analysis reveals that lysine acetylation mediates the effect of antemortem stress on postmortem meat quality development[J]. Food Chemistry, 2019, 293:396-407.

案例三：鸡蛋黄的N-糖蛋白组学分析

研究内容：鸡蛋黄的N-糖蛋白质组学 发表期刊：Journal of Agricultural and Food Chemistry
 研究策略：N-糖蛋白质组

★ 研究背景

N-糖基化翻译后修饰在食物蛋白的结构和功能中发挥着重要作用。然而，鸡蛋黄 (Chicken Egg Yolk, CEY) 的N糖基化组尚未研究。

★ 研究思路

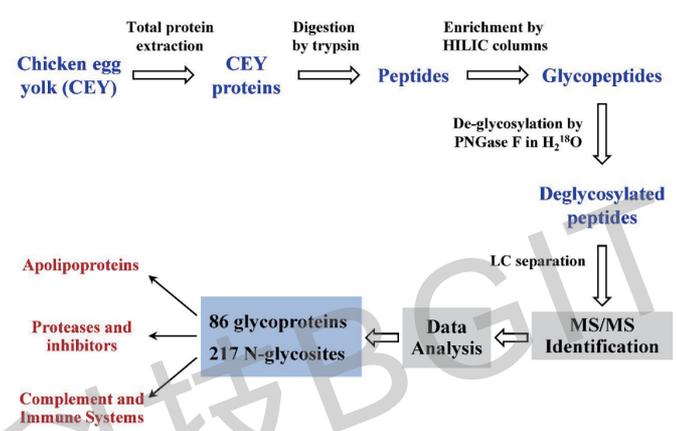


图10 实验设计
 鸡蛋黄、富集糖肽，用PNGase F去糖基化，使用鸟枪法进行鉴定

★ 研究结果

1. 鉴定217个糖基化位点和86个糖蛋白。这些糖蛋白主要参与结合、生物调节、催化活性和代谢过程。

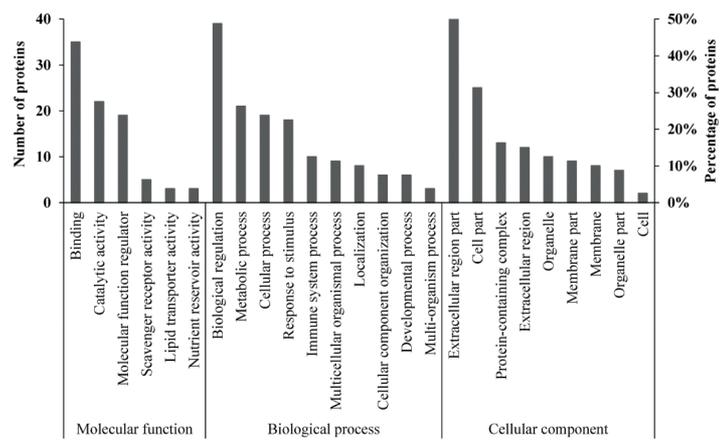


图11 糖蛋白功能注释

2. 在鉴定到糖蛋白中，有22种属于蛋白酶或蛋白酶抑制剂，说明蛋白酶或抑制剂调节系统的存在。此外，有15个糖蛋白属于免疫系统成员，为孵化期间提供保护，防止外界的威胁。

3. 本研究结果有助于理解鸡蛋黄在胚胎发育阶段改变机制以及储存过程的生理体征。

★ 参考文献

Geng F, Xie Y, Wang J, et al. N-glycoproteomic analysis of chicken egg yolk.[J]. Journal of Agricultural and Food Chemistry, 2018, 66(43):11510-11516.

典型案例汇总

发表时间	发表期刊	文献名称	研究策略
2019	Molecular & Cellular Proteomics	Integrative proteomic and phosphoproteomic profiling of testis from Wip1 phosphatase-knockout mice: Insights into mechanisms of reduced fertility	定量蛋白质组+磷酸化蛋白定量
2019	Journal of Proteome Research	Targeted quantitative kinome analysis identifies PRPS2 as a promoter for colorectal cancer metastasis	目标磷酸化蛋白PRM定量
2019	Food Chemistry	Proteomic analysis reveals that lysine acetylation mediates the effect of antemortem stress on postmortem meat quantity	乙酰化蛋白定量
2018	Analytical Chemistry	Pseudotargeted MS method for the sensitive analysis of protein phosphorylation in protein complexes	目标磷酸化蛋白PRM定量
2018	Nature	Quantitative phosphoproteome analysis of the molecular substrates of sleep need	定量蛋白质组+磷酸化蛋白定量
2018	Nature Communications	CoA synthase regulates mitotic fidelity via CBP-mediated acetylation	乙酰化蛋白定量
2018	Journal of Agricultural and Food Chemistry	N-glycoproteomic analysis of chicken egg yolks	N-糖基化蛋白定量
2016	New Phytologist	Endoplasmic reticulum-associated N-glycan degradation of cold-upregulated glycoproteins in response to chilling stress in Arabidopsis	N-糖基化蛋白定量

多肽组学

多肽组学研究对象为生物体内源性的小肽分子。内源性多肽通常分子量范围在1-10kDa范围内，具有特定功能活性或作为生理病理特异性的潜在分子标志物。内源性多肽来源于前体蛋白质的特定剪切或酶催化水解，可以看作是一种蛋白质的结构性修饰，其结构相对于常规大分子蛋白质而言，具有结构简单、稳定、穿透性强、免疫原性小等特点。

▶▶ 产品应用

- 1.发现新型疾病biomarker: 除了传统DNA、RNA、蛋白质分子和代谢小分子外, 内源性多肽分子可作为一种新型的分子标志物类型 (图1)。
- 2.发现新的功能多肽分子: 如激素肽、抗菌肽、毒素肽、神经肽、抗氧化/抗衰老肽等。
- 3.开发新型多肽药物前体: 多肽药物具有低免疫原性, 高药物活性、高特异性、高穿透性, 以及易人工合成等优势。
- 4.探索体内未知的酶催化反应: 从多肽产物和蛋白底物信息反推酶催化反应。

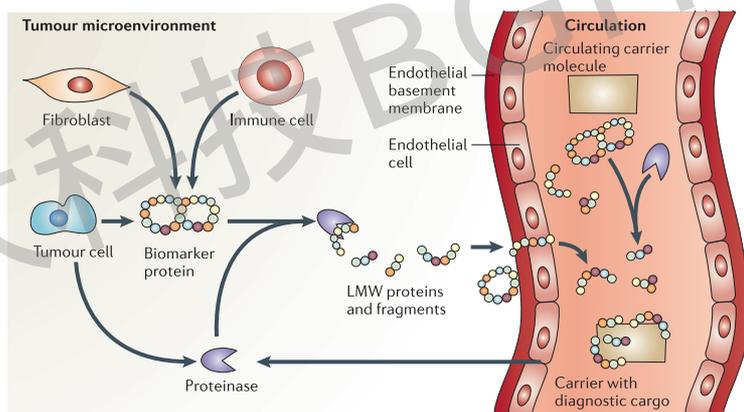


图1 疾病 (肿瘤等) 多肽在血浆中的特定存在形式
Petricoin EF, et al. Nat Rev Cancer, 2006

▶▶ 技术路线

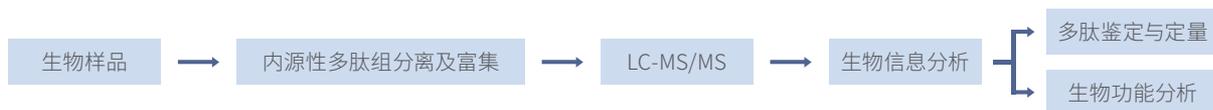


图2 胶点胶条技术流程

▶▶ 技术优势



高通量
可同时对样品中的大量多肽进行综合研究



高特异性
与传统免疫反应相比, 不存在抗体的交叉反应, 有较高的特异性



高灵敏度
利用高性能质谱平台LC-MS/MS进行检测, 较传统免疫反应和基于MALDI-TOF的检测具有更高的灵敏度和准确度

多组学关联分析

蛋白质组与转录组关联分析

产品概述

蛋白质是生命功能的执行者，其含量的变化在生物体的生长发育、环境应激、疾病发生发展等过程中发挥着重要的作用，因此对蛋白质的表达量进行深入研究显得至关重要。通过蛋白质组学研究可以得到基因组和转录组无法体现的功能性相关信息：基因实际表达的时空特性和丰度高低的特征；蛋白质的翻译后修饰形式具备特定的生物学功能；大多数蛋白质可形成具有功能性的复合物，诸如蛋白质与蛋白质复合体、蛋白质与核酸复合体、蛋白质与脂类复合体等。

mRNA的表达量是影响蛋白表达最为重要和直接的因素，通过分析蛋白组与转录组的关联性，可以系统全面地了解生物体内基因表达调控途径。

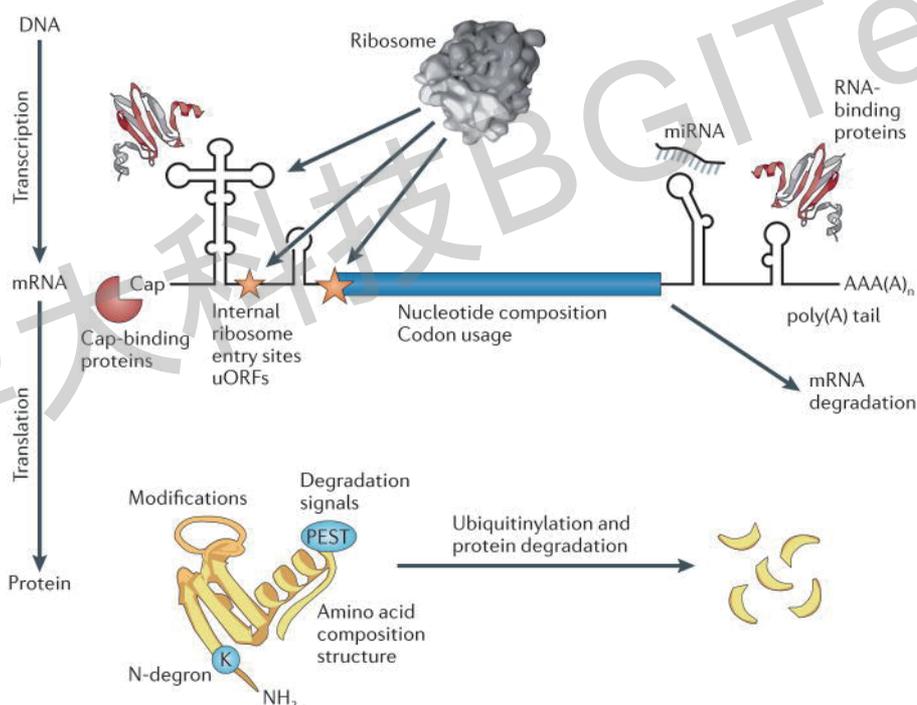


图1 转录调控和蛋白翻译后修饰模式

该图展示了蛋白质的丰度取决于RNA和蛋白产生与降解之间的平衡，图中上半部分主要解释转录和翻译过程的合成及稳定状态，下半部分主要阐述蛋白降解。因此，要全面探究生物体生长发育、环境应激机制、疾病发生发展规律，精准描绘关键基因的表达模式，同步检测mRNA和蛋白质的表达量并进行联合分析已成为当前研究的必然趋势。

拟解决的关键科学问题

1. 动植物生长发育过程中生理机制的研究，为动植物遗传品质改良，培育新品种提供理论数据；
2. 动植物环境应激性研究，为动植物生理机制提供理论数据，提高农牧业动植物环境适应性；
3. 动植物致病机理研究，提高农牧业动植物的抗病能力，同时也为人类疾病研究提供模式基础。

分析思路

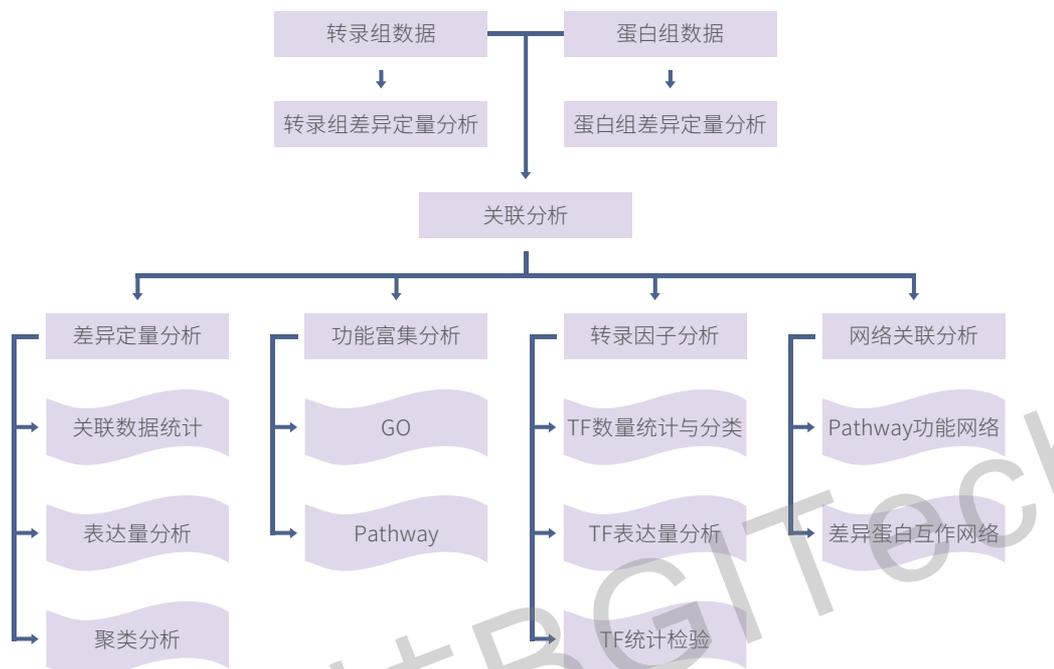


图2 蛋白组与转录组关联分析方案整体思路

分析内容

1. 定量关联分析:

对于转录组和蛋白组数据在表达量的层面上进行关联, 获得两个组学层面表达量趋势和相关性进行分析, 对表达情况进行聚类, 并对于表达趋势一致或者不同的基因进行深入分析。

首先, 按照蛋白和mRNA表达量的变化将所有关联到的基因分成5类, 以便细致描绘基因表达调控模式:

- A. 蛋白和mRNA表达趋势相同—DEPs_DEGs_Same Trend
- B. 蛋白和mRNA表达趋势相反—DEPs_DEGs_Opposite
- C. 蛋白表达有差异, mRNA表达无差异—DEPs_NDEGs
- D. 蛋白表达无差异, mRNA表达有差异—NDEPs_DEGs
- E. 蛋白表达和mRNA表达均无差异—NDEPs_NDEGs

对细分的5类表达关联类型, 有助于验证表达一致性 (A-正相关), 补充 (C/D/E-仅蛋白或RNA差异或无差异)、揭示特殊 (B-负相关) 的生物调控和代谢机制。

然后, 为了深入了解各种表达情况的相关性趋势和相关性系数, 需要对5类情况分别进行分析:

Correlation of proteome and transcriptome

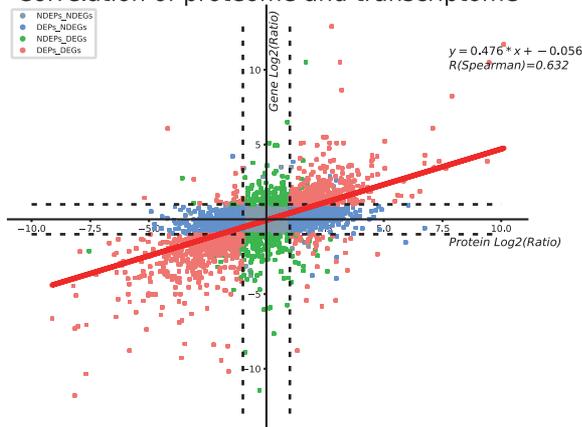


图3 所有定量蛋白质和基因的表达量关联图
横坐标为蛋白质的表达量, 纵坐标为基因的表达量。
黑点表示mRNA和蛋白都无显著差异;
红点表示mRNA无显著差异, 而蛋白显著差异;
绿点表示mRNA显著差异, 而蛋白无显著差异;
蓝点表示mRNA和蛋白都显著差异。

在动植物生理机制的研究中，最为关注的是转录组和蛋白组表达趋势一致的情况，表达情况正相关，有利于说明关键基因的表达情况在两个组学层面都得到了验证；对于转录组和蛋白组表达趋势相反的情况，一般是为了说明一些特殊的抑制调控方式，可根据具体的蛋白功能进行分析。另外，单一组学表达发生变化的情况，可通过后续的功能关联分析，寻找基因上下调的关系进行调控关键基因的深入挖掘。

2.功能关联分析:

通过两个组学的鉴定、定量层面的关联，可以分析基因表达产物mRNA和蛋白一对一的关联方式，但对于某一类基因或者具有上下游调控关系的基因，仅通过一对一的关联方式无法进一步分析，需要通过功能和代谢通路的分析进行解释和分析。

对差异表达基因和差异表达蛋白在GO条目/Pathway上的注释及富集情况进行分析，并将相同的GO功能条目/Pathway中注释到的基因和蛋白进行整合分析，同一GO条目/Pathway上的基因和蛋白在功能上相似，对环境因素的反馈可能存在共调控或是共表达等情况，有利于从基因集的层面研究基因表达调控。

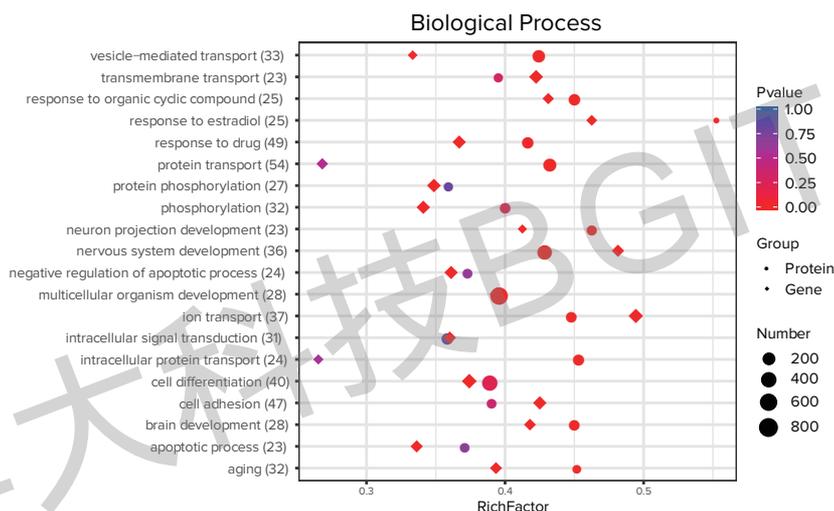


图4 GO关联数量及富集因子分析四维气泡图

图中X轴为富集因子，Y轴为各个GO条目，且在括号中展示了该条目所关联的蛋白数量；

不同的气泡形状代表不同组学数据，圆形代表蛋白质组的富集结果，菱形代表mRNA维度的富集结果。

气泡大小代表富集到该term的差异蛋白或者差异基因的数量；颜色代表显著富集Pvalue值。

富集因子越大，说明差异蛋白或差异基因在该类功能中所占的比例越大；Pvalue越小，说明富集到该类功能越显著。

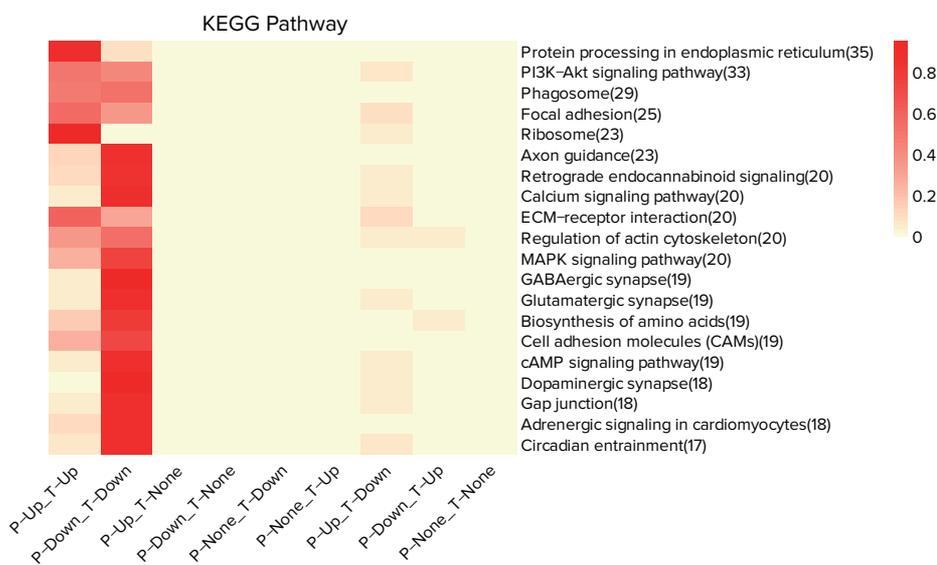


图5 GO条目关联分类概率热图

Y轴为各个GO条目，括号为该条目所关联的蛋白数量；X轴为不同差异类型分布情况；

热图颜色代表该情况在该GO条目关联上的蛋白中所占的比例。

根据蛋白组和转录组差异表达基因的分析结果，对其进行KEGG生物通路分类以及富集分析，最终将蛋白组和转录组差异表达基因的信息汇总在一张通路图中。分别显示mRNA上调，蛋白上调，mRNA和蛋白同时改变或者一方改变，mRNA和蛋白同时不变等情况，直观展示一条通路中的所有转录组和蛋白组鉴定和定量到的数据，更加方便地展示关键调控基因。

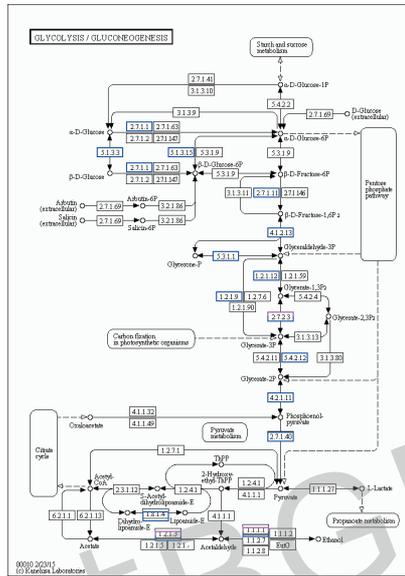


图6 差异蛋白和差异基因Pathway整合图
(红框代表差异蛋白，蓝框代表差异基因)

3.多组学网络关联分析:

针对上图中的所有差异蛋白，通过与STRING蛋白互作数据库比对，绘制网络图。

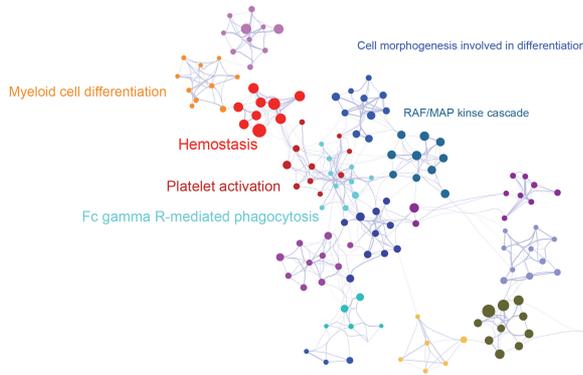


图7 差异蛋白和差异基因网络互作图

4.转录因子关联分析:

转录因子在动植物的生长发育及其对外界环境的反应中起着重要的调控作用，对转录因子的数量及表达量分布进行统计分析。

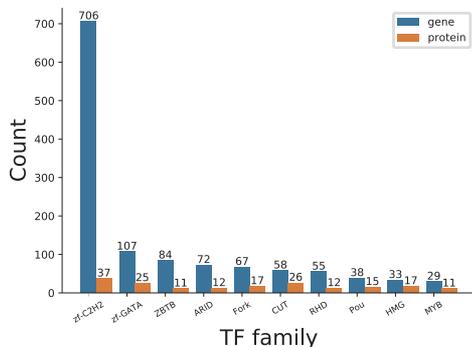


图8 各转录因子家族鉴定到的转录因子个数

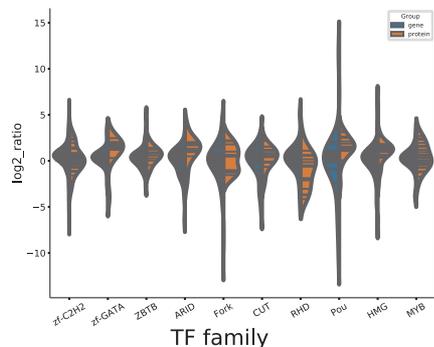


图9 各转录因子家族在基因及蛋白层面的表达量分布图

► 案例分析

案例一：黑腹果蝇发育过程中转录后调控定量分析

研究内容：黑腹果蝇发育机制 发表期刊：Nature Communications 研究策略：RNA-Seq、蛋白定量

★ 研究背景

复杂的转录后调控，使得mRNA水平和蛋白质丰度之间的相关性降低。为了研究转录组和蛋白组的相关性，本研究以果蝇胚胎发育为研究对象，生成了14个时间点的成对转录组-蛋白组时间进程数据集。结果显示，mRNA-蛋白质相关性有限 ($\rho=0.54$)，但在没有复杂的转录后调控的前提下，蛋白质翻译和降解的数学模型能描述达84%的基于mRNA动力学的蛋白质时间序列。本研究假设RNA结合蛋白Hrb98DE参与糖的转录后控制早期胚胎发育中的代谢，并使用Hrb98DE敲除验证了这一假设。本研究提出了一个可用于从大规模、时间序列的转录组和蛋白组数据中鉴定转录后基因调控的系统生物学框架。



图10 果蝇胚胎发育的转录组&蛋白组分析结果

a. 运用RNA-Seq和质谱检测的方法在果蝇胚胎发育期间测定配对的mRNA和蛋白质的时间点
b. mRNA和蛋白质时间序列的热图

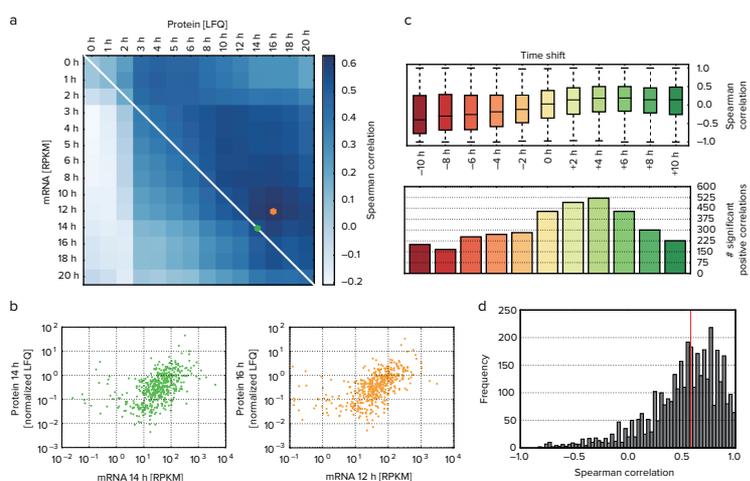


图11 mRNA和蛋白质相关性分析结果

a. 所有样品的RNA-蛋白质全谱相关性分析； b. 最大的RNA-蛋白质全谱相关性分析；
c. 个体基因mRNA和蛋白质时间序列局部相关性； d. 时间序列下，最大的斯皮尔曼相关系数分布

★ 参考文献

Becker K, Bluhm A, Casas-Vila N, et al. Quantifying post-transcriptional regulation in the development of *Drosophila melanogaster*[J]. Nature Communications, 2018, 9(1):4970.

送样要求

▶▶ 蛋白产品送样要求

样本类型	非修饰类产品送样量	修饰类产品送样量		备注	
		S/T磷酸化、N糖基化	乙酰化、泛素化		
动物类	常见动物组织：动物内脏（心、肝、脾、肺、肾）、皮肤、肌肉、脑等	≥10mg	≥20mg	≥200mg	非修饰DIA定量产品需要5mg
	软体动物类（弓形虫、血吸虫、果蝇、蠕虫、小菜蛾、灰飞虱、绦虫、蝉、涡鞭虫等）	≥10mg	≥10mg	≥100mg	
细胞类	悬浮细胞、贴壁细胞	≥5×10 ⁶	≥2×10 ⁷	≥2×10 ⁸	细胞解冻后部分细胞破碎使蛋白溶于液体中，无法对细胞再次进行洗涤，杂质会干扰后续实验，因此细胞收集后请务必用PBS清洗2-3次后再送样。送清洗细胞后残留的液体≤200μL即可
液体类	血浆、血清	≥200μL	-	-	若不需要去高丰度，只需要送20μl即可满足实验需求
	羊水、脑脊液、精液等	≥1mL	-	-	若不需要去高丰度，只需要送200μl即可满足实验需求
	唾液、乳汁	≥500μL	-	-	
	尿液	≥20mL	-	-	正常人的尿液蛋白含量较低，建议翻倍增加送样量
	泪液	≥10μL	-	-	
植物类	植物的嫩枝部（叶芽、嫩叶片）、藻类	≥500mg	≥2g	≥10g	
	植物的老叶、根、茎，树皮	≥1g	≥4g	≥20g	
	植物的花蕾、花粉	≥100mg	≥500mg	≥1g	花粉较容易沾在纸上，若储存于包装纸中会造成一定的损失，建议将样本保存于1.5mL离心管中
	植物的种子（水稻/小麦种子等）、果实（苹果、水蜜桃、梨）	≥100mg	≥500mg	≥1g	植物的种子和果实分别含淀粉和糖分较高，为避免提取的蛋白总量不达标需再次送样浪费周期，送样时尽量送2g以上

样本类型	非修饰类产品送样量	修饰类产品送样量		备注	
		S/T磷酸化、N糖基化	乙酰化、泛素化		
微生物类	原核细菌(大肠杆菌、沃氏葡萄球菌等), 真菌(酵母等)	菌体 $\geq 50\text{mg}$ 或 细胞数 $\geq 5 \times 10^6$	菌体 $\geq 200\text{mg}$ 或 细胞数 $\geq 2 \times 10^7$	菌体 $\geq 1\text{g}$ 或 细胞数 $\geq 1 \times 10^8$	原核细菌磷酸化个数很少, 原则上不接受该样本做磷酸化检测
胶条	考染或银染蛋白胶条	肉眼可见条带	内含 $\geq 10\mu\text{g}$ 蛋白	内含 $\geq 10\mu\text{g}$ 蛋白	
胶点	考染或银染胶点	肉眼可见斑点	内含 $\geq 1\mu\text{g}$ 蛋白	内含 $\geq 1\mu\text{g}$ 蛋白	仅适合不经过富集直接质谱检测的S/T磷酸化和乙酰化修饰鉴定
蛋白样品	复杂样本蛋白液、蛋白粉末	$\geq 100\mu\text{g}$, 浓度 $\geq 0.5\mu\text{g}/\mu\text{L}$	$\geq 1\text{mg}$, 浓度 $\geq 0.5\mu\text{g}/\mu\text{L}$	$\geq 10\text{mg}$, 浓度 $\geq 0.5\mu\text{g}/\mu\text{L}$	样品中不含较高浓度的盐分和去污剂, 蛋白液澄清; 非修饰DIA定量单次实验样本可低至 $20\mu\text{g}$
	单个蛋白(针对IP或CO-IP纯化下来的蛋白液)	肉眼可见斑点	内含 $\geq 1\mu\text{g}$ 蛋白	内含 $\geq 1\mu\text{g}$ 蛋白	仅适合不经过富集直接质谱检测的S/T磷酸化和乙酰化修饰鉴定
其他	粪便	$\geq 100\text{mg}$	-	-	

华大科技BGITech

产品
简介

Product
Profile

华大科技BGITech

代谢组学

非靶向代谢组学	115
脂质组学	117
靶向代谢组学	119
多组学关联分析	123
送样要求	128

非靶向代谢组学

产品概述

华大基因质谱平台采用的非靶向代谢组学检测技术为液质联用 (LC-MS) 技术, 通过LC-MS检测生物体内受到刺激或扰动前后的所有小分子代谢物, 并通过自主研发的代谢组学软件包metaX进行生物信息学分析, 筛选出实验组和对照组之间的差异代谢物, 对差异代谢物进行代谢通路分析, 进而研究代谢物与生理病理变化的关系。

技术流程

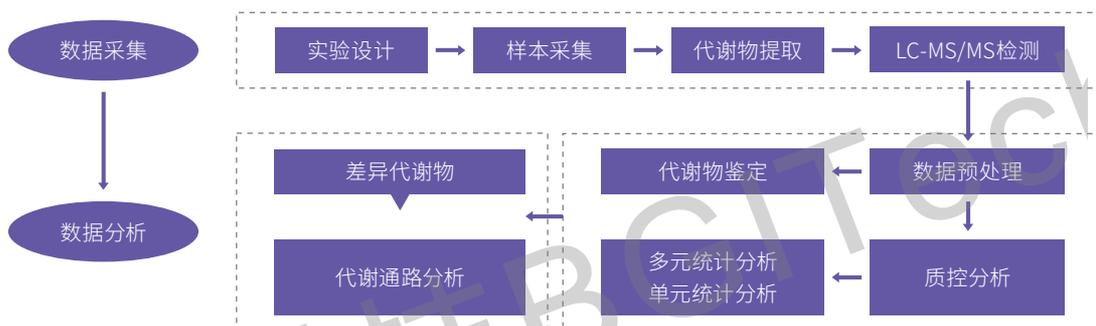


图1 技术流程图

信息分析流程

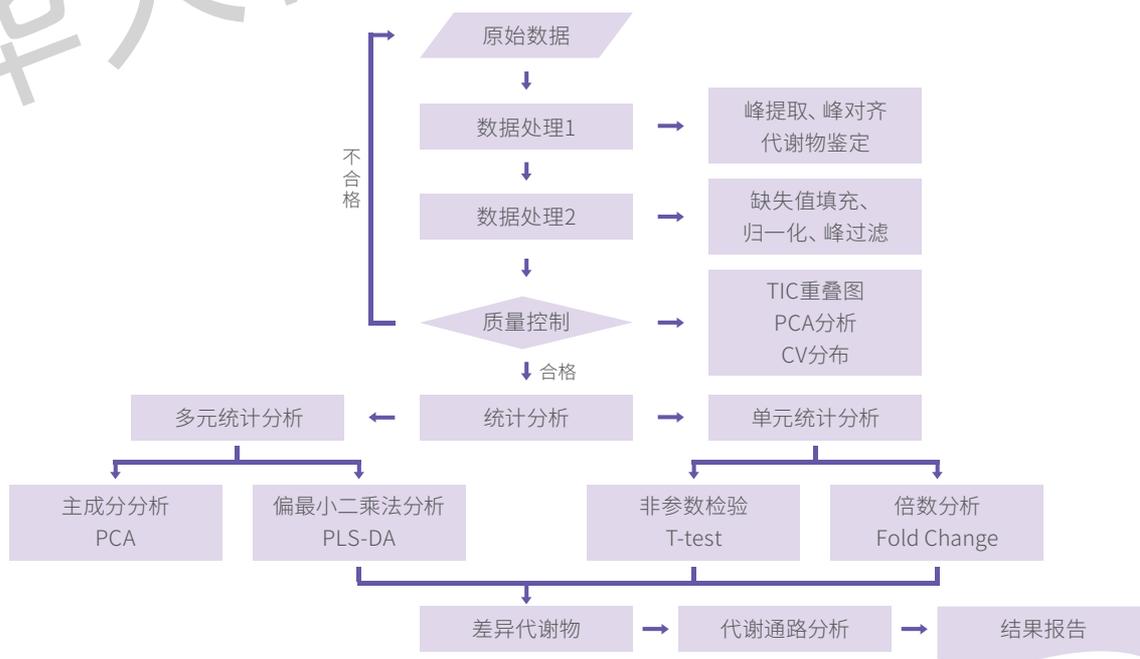


图2 信息分析流程图

技术参数

检测平台	仪器型号	软件	项目周期
LC-MS	Thermo Q Exactive 系列	Compound Discoverer	28个自然日 (≤50个样本)
	Waters Xevo G2-XS Q-ToF	Progenesis Q1	28个自然日 (≤50个样本)

► 生物学重复次数要求

样本类型	重复次数要求
植物和微生物样本	≥ 6个
动物样本	≥ 10个
临床样本	≥ 30个

► 案例分析

案例：结核病生物标志物的发现

研究内容：结核病生物标志物研究 发表期刊：Nature Communications 研究策略：非靶向代谢组学

★ 研究背景

结核病 (TB) 已成为全球最致命的传染病, 预估世界上四分之一的人口有潜在感染结核分枝杆菌 (Mtb) 的风险, 其中约10%的人发展成为结核病。控制结核病的手段不仅包括药物和疫苗, 同时还需要快速且敏感诊断结核病的检测方法。目前结核病风险预测主要使用TST和IGRA检测, 特异性较差, 存在假阳性风险。因而寻找针对结核病特异性强和准确度高的风险预估生物标志物具有重大意义。

★ 研究结论

本研究基于UPLC-MS/MS技术和GC-MS技术, 采用非靶向代谢组学的方法, 研究非洲4个不同地点的4,462例HIV阴性结核病家庭系, 发现活性TB前12个月皮质醇和大尿氨酸在结核病和对照组中代谢发生变化, 可作为结核病风险预估的生物标志物。

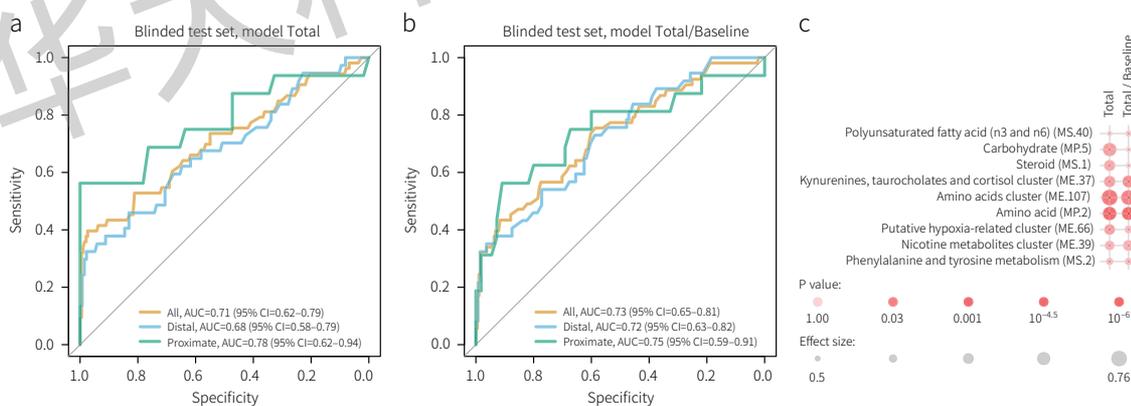


图3 通过模型确定不同组别中代谢物的差异

★ 参考文献

Weiner J 3rd, Maertzdorf J, Sutherland JS, et al. Metabolite changes in blood predict the onset of tuberculosis [J]. Nature Communications, 2018, 9(1):5208.

► 典型案例汇总

发表时间	发表期刊	文献名称	研究策略
2018	Nature Communications	Metabolite changes in blood predict the onset of tuberculosis	非靶向代谢组学
2017	Nature Medicine	Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention	非靶向代谢组学+靶向代谢组学+宏基因组学
2016	Scientific Reports	Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease	非靶向代谢组学

脂质组学

产品概述

脂质组学是代谢组学的一个重要分支，是对整体脂质进行系统分析的一门新兴学科。通过比较不同生理状态下脂代谢网络的变化，进而识别代谢调控中关键的脂质生物标志物，最终揭示脂质在各种生命活动中的作用机制。

美国国立卫生研究院 (NIH) 2003年所资助的“脂质代谢途径研究计划” (Lipid Metabolites and Pathways Strategy, LIPIDMAPS) 将脂质分为八大类, 包括: 脂肪酰类(Fatty Acyls)、甘油酯类(Glycerolipids)、甘油磷脂类(Glycerophospholipids)、鞘脂类(Sphingolipids)、固醇脂类 (Sterol Lipids)、糖脂类(Saccharolipids)、孕烯醇酮脂类(Prenol Lipids) 和聚酮类(Polyketides)。

华大基因质谱平台采用液质联用 (LC-MS) 技术进行脂质组学检测分析, 并通过自主研发的代谢组学软件包metaX进行生物信息学分析, 筛选出实验组和对照组之间的差异脂质分子, 进而研究脂质代谢调控在各种生命现象中的作用机制。

技术流程

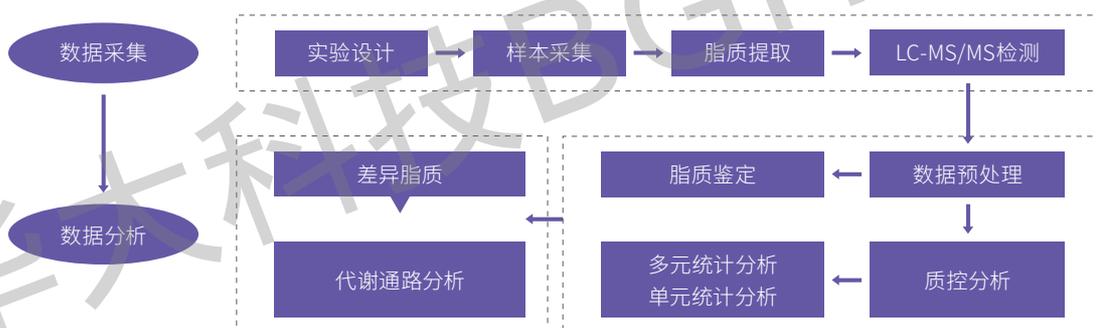


图1 技术流程图

技术参数

检测平台	仪器型号	软件	项目周期
LC-MS	Thermo Q Exactive 系列	LipidSearch	28个自然日 (≤50个样本)
	Waters Xevo G2-XS Q-Tof	Progenesis Q1	28个自然日 (≤50个样本)

生物学重复次数要求

样本类型	重复次数要求
植物和微生物样本	≥ 6 个
动物样本	≥ 10 个
临床样本	≥ 30 个

案例分析

案例: 首个中国人群糖尿病相关血浆脂质组学研究

研究内容: 糖尿病血浆脂质组 发表期刊: GigaScience 研究策略: 脂质组学

★ 研究背景

华大基因与苏州疾控中心合作，研究中国II型糖尿病患者、早期糖尿病患者与糖耐受正常人群的血浆脂质代谢物组成及差异，并为亚洲人群的脂代谢研究提供参考。

★ 研究思路

样品: 114位II型糖尿病患者 (T2D), 81位早期糖尿病患者 (prediabetic), 98位血糖正常的人 (NGT)
脂质组学: LC-MS/MS
数据分析: QI

★ 研究结果

- 1.脂质组学数据和几个血浆生化指标以及生理情况密切相关，从NGT到prediabetes再到T2D，代谢物持续发生变化；
- 2.成功构建28个特征峰模型，用于指示T2D的发生发展；
- 3.AUC=90.23%；
- 4.随机森林分析筛选出28个疾病有关离子，对其和疾病表型进行相关性分析。

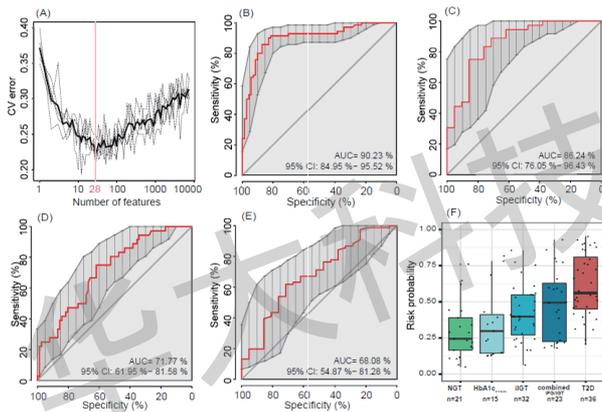


图2 代谢物持续变化情况图

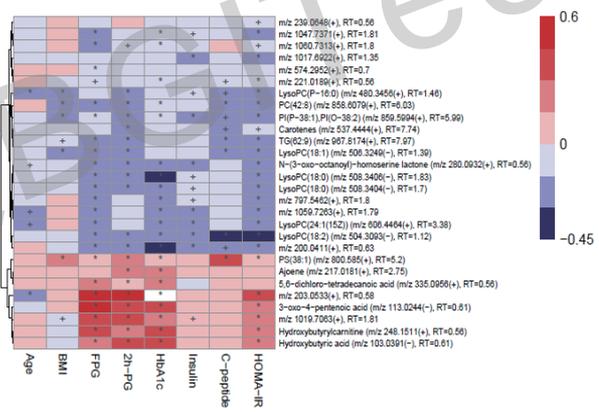


图3 一般线性模式对选择出的28个离子和表型进行相关性分析

★ 研究结论

脂质分子和糖尿病的临床指标之间具有相关性，溶血磷脂酰胆碱 (Lysophosphatidylcholine, lysoPC) 和酰基肉毒碱(Acylcarnitine)可作为II型糖尿病早期诊断的生物标志物。

★ 参考文献

Zhong H, Fang C, Fan Y, et al. Lipidomic profiling reveals distinct differences in plasma lipid composition in healthy, prediabetic, and type 2 diabetic individuals[J]. GigaScience, 2017, 6(7):1-12.

▶▶ 典型案例汇总

发表时间	发表期刊	文献名称	研究策略
2017	GigaScience	Lipidomic profiling reveals distinct differences in plasma lipid composition in healthy, prediabetic and type 2 diabetic individuals	脂质组学
2017	GigaScience	Lipidomics profiling reveals the role of glycerophospholipid metabolism in psoriasis	脂质组学
2016	Circulation	Plasma lipidomic profiles improve upon traditional risk factors for the prediction of cardiovascular events in type 2 diabetes	脂质组学

靶向代谢组学

产品概述

靶向代谢组学不同于非靶向代谢组学全面地检测代谢物，它针对特定的代谢物进行检测分析，通过标准品进行绝对定量，可以更准确地监测代谢的动态变化过程，阐明相关的代谢机制，为代谢生物标志物的鉴定和验证提供了有力的支持。靶向代谢组学具有特异性强、灵敏度高、准确度高、重现性好、线性动态范围宽等突出优点。

技术原理

靶向代谢组学常用的分析方法为多反应监测(Multiple Reaction Monitoring, MRM)技术。MRM 技术是基于目标分子已知或假定的信息设定质谱检测参数，对目标离子进行信号记录，去除大量非目标离子信号的干扰，从而得到质谱信息的一种仪器扫描模式。具体地说，就是根据目标分子母离子质荷比和碎片离子质荷比，选择母离子-子离子对 (Transition)，允许符合设定的母离子进入碰撞室，碰撞完成后只记录设定的子离子信号。通过母离子和子离子的两次选择，去除干扰离子，降低化学背景，提高方法的灵敏度和专属性。

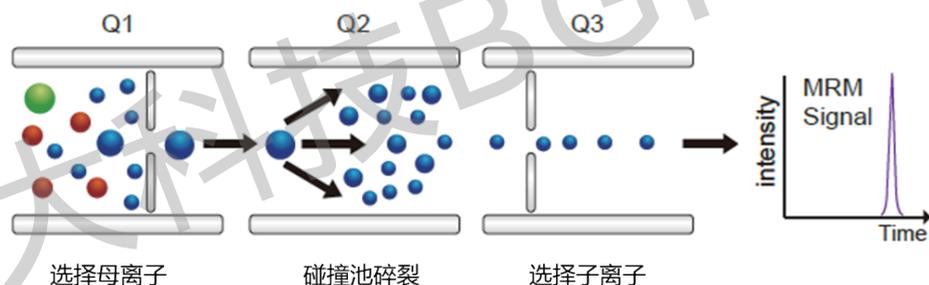


图1 MRM质谱技术的工作原理示意图

技术优势



高通量

可同时检测几十种同类型代谢物，节省时间



精确定量

灵敏度可达pg级，可以对低丰度的代谢物进行准确定量



高准确性

每种代谢物的标准品建立方法，保证结果的准确性



方法建立长期性

无需反复建立方法，每个代谢物的检测方法一次性建立，长期适用



重现性好

多重保障 (每个代谢物有多个特征离子对，每个信号多次检测)，在不同实验室也可获得高重现性

技术流程

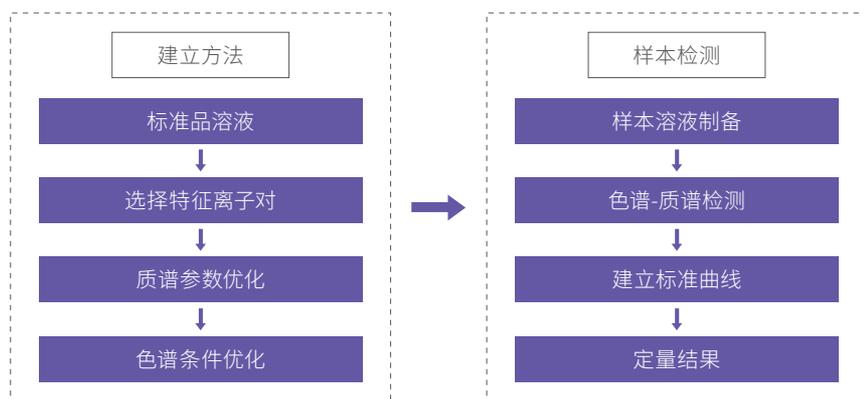


图2 技术流程图

技术参数

检测平台	仪器型号	项目周期
LC-MS	SCIEX QTRAP 4500/5500/6500系列, Waters Xevo TQ-S	35个自然日 (≤100个样本)
GC-MS	Thermo TSQ 9000, Agilent 7890B-5977A	35个自然日 (≤100个样本)

生物学重复次数要求

样本类型	建议重复次数	最少重复次数
植物和微生物样本	≥ 6个	3个
动物样本	≥ 10个	6个
临床样本	≥ 30个	10个

靶向代谢组学产品

类别	代表性化合物	代表性化合物
氨基酸 (20种)	精氨酸、组氨酸、异亮氨酸	血浆、血清、尿液、组织、粪便及肠道内容物、细胞、微生物及发酵液
氨基酸 (39种)	α -氨基己二酸、牛磺酸、同型胱氨酸	血浆、血清、尿液、组织、粪便及肠道内容物、细胞、微生物及发酵液
水溶性维生素 (7种)	维生素B1、维生素B2、维生素B9	血浆、血清、尿液、组织、粪便及肠道内容物、细胞、微生物及发酵液
脂溶性维生素 (5种)	维生素A、 α -生育酚、维生素K	血浆、血清、尿液、组织、粪便及肠道内容物、细胞、微生物及发酵液
类固醇激素 (13种)	睾酮、雌酮、孕酮、醛固酮	血浆、血清、尿液、组织、粪便及肠道内容物、细胞
胆汁酸 (15种)	胆酸、甘氨酸胆酸、牛磺胆酸	血浆、血清、尿液、组织、粪便及肠道内容物、细胞、胆汁
花青素相对定量 (45种)	芍药花青素 O-己糖苷、原花青素A1	鲜样、冻干样本
植物激素 (6类, 25种)	吲哚-3-乙酸、异戊烯腺嘌呤	植物鲜样
类胡萝卜素 (绝对定量/相对定量, 17种)	α -胡萝卜素、 β -胡萝卜素、叶黄素	鲜样、冻干样本
胆汁酸 (50种)	23-脱甲脱氧胆酸、3-氧代脱氧胆酸	血浆、血清、尿液、组织、粪便及肠道内容物
氧化脂质组 (绝对定量/相对定量, 64种)	PGJ2、LXA5、RvD1、1-Mar、9-oxoODE	血浆、血清、组织、粪便及肠道内容物、细胞

► 案例分析

案例：肠道微生物和血浆胆汁酸的分析对患者进行抗糖尿病分型治疗

研究内容：II型糖尿病分型

发表期刊：Nature Communications

研究策略：靶向代谢组学+宏基因组学

★ 研究背景

在中国和其他亚洲国家，阿卡波糖被推荐作为替代二甲双胍的II型糖尿病的治疗药物。阐明降糖药物对肠道微生物的影响和额外代谢获益机制及基于临床试验队列临床结果的关系，对治疗前II型糖尿病患者的分型具有重要意义。

★ 研究思路及结果

1.通过临床队列研究，对106例II型糖尿病初发患者随机采用阿卡波糖和格列吡嗪两种降糖药治疗。结果发现阿卡波糖给药组不仅具有降糖功效，而且具有显著减重、降脂、改善胰岛素抵抗的功能。

2.针对阿卡波糖降糖功能外代谢获益机制，发现其治疗后能够显著降低肠道共生微生物的基因丰度和生物多样性。通过代谢组学研究发现，胆汁酸的代谢过程则是导致肠道菌群微环境代谢变化的作用靶点。经过阿卡波糖治疗后，由肠道微生物合成的肝毒性的疏水性次级胆汁酸（脱氧胆酸和石胆酸）生成减少，血浆及粪便中的次级胆汁酸浓度降低。与此同时，有益的亲水性次级胆汁酸和熊脱氧胆酸生成得到促进，它们能够作用在肝脏及小肠等器官的FXR受体（胆汁酸受体）上，从而调节糖脂代谢，达到降糖外的获益。

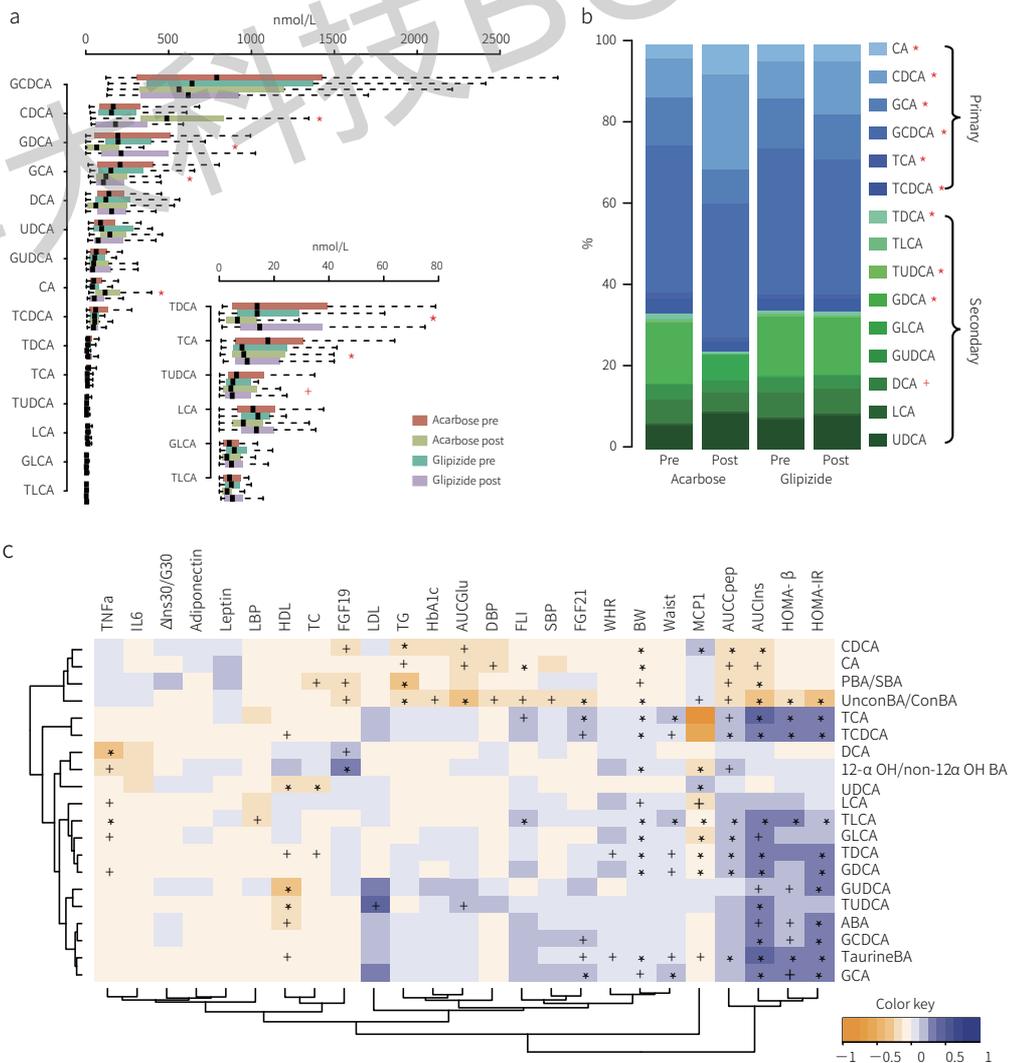


图3 阿卡波糖治疗后胆汁酸组成的变化与临床结果相关

3.通过宏基因组测序技术对受试者用药前肠道菌群进行分型,分为两大类:富含拟杆菌的B肠型以及富含普氏菌的P肠型;而在研究过程中发现在两组基线特征无明显差异的前提下,阿卡波糖治疗后B肠型患者的疗效更显著,空腹血糖改善、胰岛素抵抗改善程度更明显。

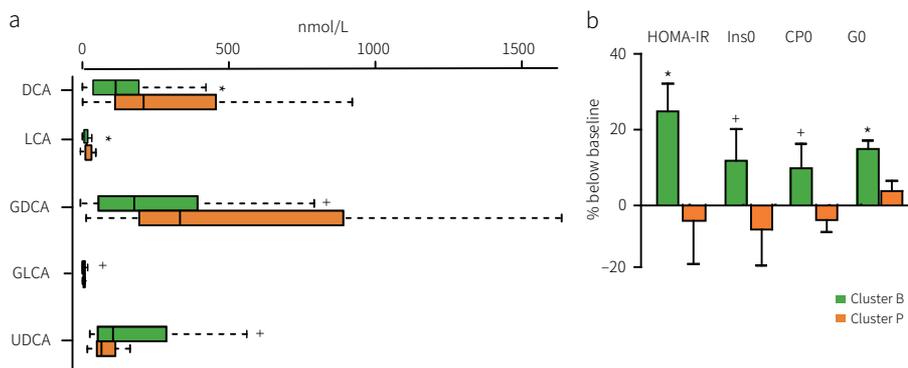


图4 不同基线菌群组成的患者血浆胆汁酸组成和不同的阿卡波糖治疗反应

★ 研究结论

该研究不仅破解了阿卡波糖众多降糖外代谢获益的机制,同时也为设计靶向胆汁酸信号的新型降糖药物提供了新的思路。同时更有意义的是,研究发现不同的肠道菌群组合即肠型是决定药物疗效的重要因素,对未来降糖疗效的判定有重要意义。

★ 参考文献

Gu Y, Wang X, Li J, et al. Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment[J]. Nature Communications, 2017, 8:1785 .

▶▶ 典型案例汇总

发表时间	发表期刊	文献名称	研究策略
2017	Nature Medicine	Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention	非靶向代谢组学+靶向代谢组学+宏基因组学
2017	Nature Communications	Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment	靶向代谢组学+宏基因组学
2016	Cell	Functional metabolomics describes the yeast biosynthetic regulome	靶向代谢组学+基因组学



多组学关联分析

代谢组与宏基因组关联分析

产品概述

近年来与肠道菌群相关的疾病研究迅速发展,许多研究结果表明近90%的疾病都有可能与肠道菌群相关。其中和肠道菌群相关研究较多的包括以下六大类:消化道类疾病;代谢性疾病,包括肥胖、II型糖尿病(T2D)等;肝脏类疾病;免疫类疾病;肺部疾病;精神类疾病等。肠道菌群参与了人体生命活动的各个途径,包括脂肪储存、血管生成、免疫系统的发育和成熟、某些维生素和必需氨基酸的合成、药物代谢、神经调节、食物消化和营养吸收、病原菌抵抗、上皮损伤修复、骨骼生成及骨代谢等。这些功能及其对应的pathway可以为肠道菌群相关机制的研究提供依据。

宏基因组学研究以环境中所有微生物基因组为研究对象,通过对环境样品中的全基因组DNA进行高通量测序,获得单个样品的饱和数据量、基因组组成及功能、特定环境相关的代谢通路等分析,从而进一步发掘和研究具有应用价值的基因及环境中微生物群落内部、微生物与环境间的相互关系,可为环境中微生物的研究、开发和利用提供基因资源库。

代谢组学(metabonomics/metabolomics)是继基因组学和蛋白质组学之后新发展起来的学科,通过对生物体内相对分子质量1,000Da以内的小分子代谢物进行定性定量分析,寻找代谢物与生理病理变化的关系。由于能够更为真实和直观地反映机体对于刺激的综合响应,代谢组学已经在疾病研究、药效检测等多个领域得到了广泛应用。

肠道菌群的研究方法也从单一组学,逐步发展到多组学联合分析。从最初的“关联”到探究微生物与疾病的“因果”关系,并深入挖掘微生物相关的发病机制。针对肠道样本进行宏基因组学和代谢组学关联分析研究,可分析不同疾病程度样本间基因、功能、物种、代谢物间的差异;将疾病发生、进展或治疗过程与肠道微生物进行关联分析,结合实验验证,深入研究复杂疾病发生机理。为更好的诠释肠道微生物与人类健康之间的关系提供科学依据,同时也推动肠道微生物研究在临床上的应用,有助于疾病的预防和治疗。

研究思路

由于宏基因组和代谢组数据具有高维度和复杂性等特征,因此我们设计的关联分析方案基于降维的思想,将代谢组数据和微生物组物种通过数据驱动聚类(Clustering of co-abundant metabolites或 Binning co-abundant genes)法降维,微生物组功能组成基于知识驱动如KEGG层级功能模块分类法降维,筛选与表型显著相关的数据特征进行跨组学关联分析(cross-omics association analysis),最后鉴定与表型显著相关的KEGG功能模块的驱动物种。

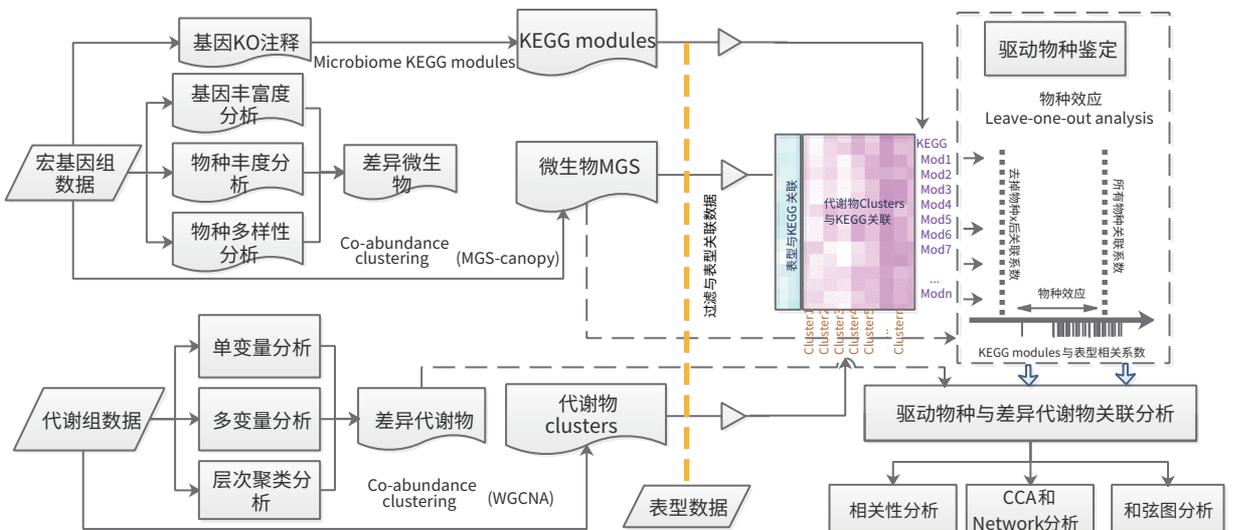


图1 宏基因组代谢组关联分析流程图

► 案例分析

案例一：肥胖干预前后肠道菌群和血清代谢物的变化

研究内容：肥胖干预前后肠道和代谢的变化 发表期刊：Nature Medicine 研究策略：粪便宏基因组、血清代谢组

★ 研究背景

肥胖已经成为世界性流行疾病，是II型糖尿病、心血管疾病和癌症的主要风险因子之一。而越来越多的证据表明肠道微生物和肥胖之间存在密切联系：

1. 肠道微生物可以改变宿主对能量的摄取和贮存，成为肥胖的一个重要环境因子；
2. 肥胖个体肠道菌群的物种多样性和基因丰富度比正常人群偏低，拟杆菌门和厚壁菌门的比值也偏低；
3. 研究表明肥胖个体通过饮食干预或者胃旁路手术（RYGB）降低体重后，拟杆菌门和厚壁菌门的比值升高；
4. 肠道微生物移植实验表明肠道微生物和肥胖的发生发展之间存在因果关系。

因此，研究肠道微生物和肥胖的关系对肥胖的预防和治疗尤为重要。

★ 研究思路

1. 疾病组VS对照组（肥胖组n=95，对照组n=105）肠道菌群和血清代谢物检测，研究出肥胖组肠道菌群的物种/基因变化，肠道菌群变化与血清代谢物的关系
2. 特定菌株移植小鼠，并检测移植后小鼠肥胖趋势及代谢的影响
3. 肥胖患者袖状胃切除术后肠道菌群和代谢的变化

★ 研究结果

1. 肥胖志愿者的肠道微生物多样性和基因数量都相对较低。发现350,524个与肥胖相关的基因，并鉴定出它们所对应的微生物。对照组以多形拟杆菌（*Bacteroides thetaiotaomicron*，简称BT菌）为代表的拟杆菌属细菌丰度很高；而肥胖人群中瘤胃球菌属、梭杆菌属等的细菌丰度较高。

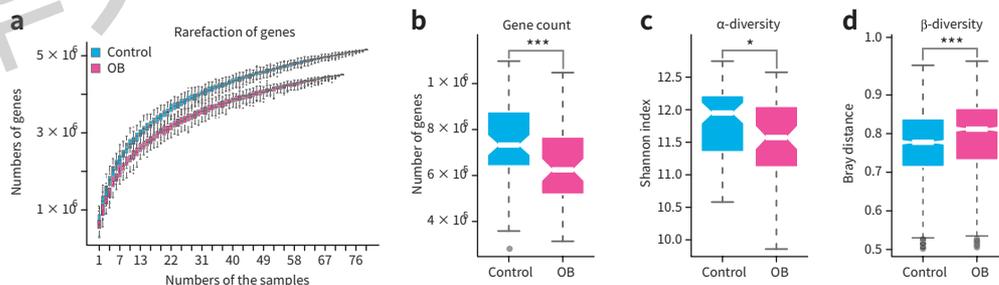


图2 肥胖组 (OB) 与对照组 (Control) 的基因稀疏性曲线、基因数量和丰富度的区别

2. 肥胖者的拟杆菌属细菌大量减少与代谢产物芳香族氨基酸 (AAA) 以及支链氨基酸 (BCAA) 的增加相关，这两类氨基酸在过去的研究中已经被证明与II型糖尿病的发生发展有关。对148种代谢物分析发现，肥胖人群中谷氨酸的含量非常高，与对照组差异也最大；肥胖人群中，谷氨酸的含量与多形拟杆菌的数量呈反比。

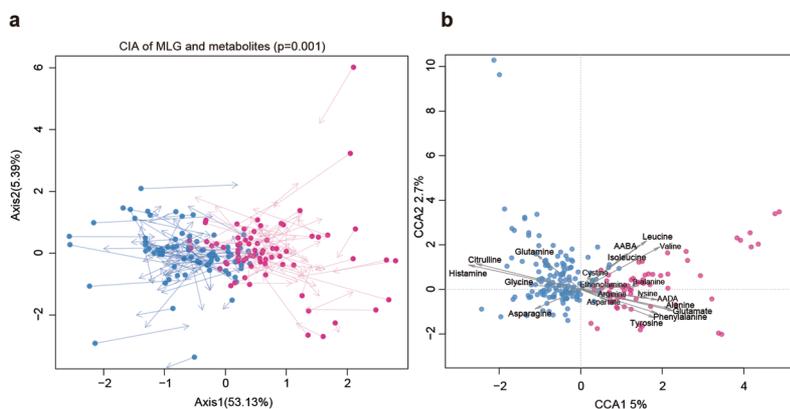


图3 微生物与代谢物相关性分析

3.正常饮食的小鼠移植了多形拟杆菌,总脂肪量减少、肌肉量增加;高脂饮食的小鼠移植多形拟杆菌后,体重增长的速度减慢,不易变肥胖。移植了活性拟杆菌组小鼠脂肪细胞体积明显减小,小鼠体内参与脂肪合成、炎症因子产生的基因的表达量明显下降,谷氨酸含量下降。

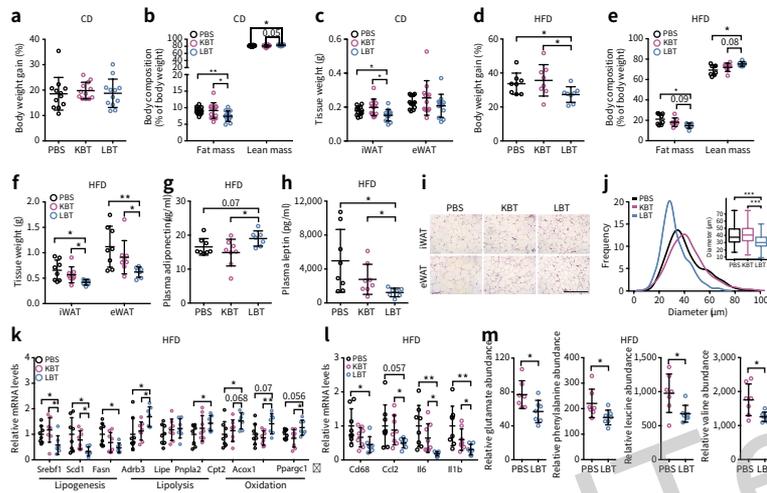


图4 BT菌对小鼠肥胖和代谢的影响

4.接受袖状胃切除术的肥胖患者体重迅速减少,高血糖、胰岛素抵抗和炎症反应显著减轻;肠道中的多形拟杆菌随着体重的下降慢慢增加,在3个月时,几乎恢复到了与对照组同样的水平;代谢循环中的谷氨酸含量也大大减少。

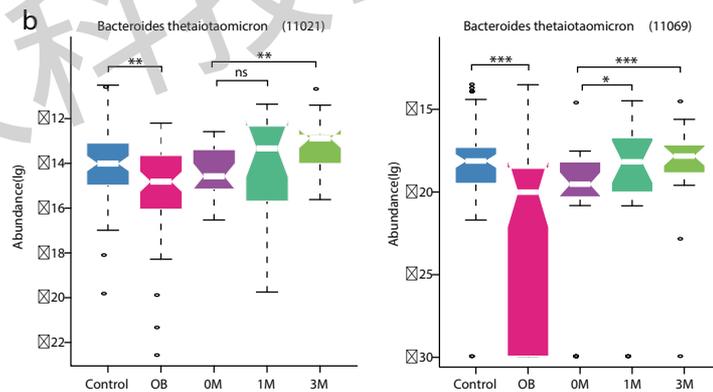


图5 减重手术患者术后3个月的BT菌基因关联群组丰度变化

★ 参考文献

Liu R, Hong J, Xu X, et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention[J]. Nature Medicine, 2017, 23(7): 859-868.

案例二：肠道与心血管疾病

研究内容：心血管病中肠道微生物功能和饮食微生物、代谢、免疫之间的相互作用
发表期刊：Circulation Research 研究策略：粪便宏基因组、血清代谢组

★ 研究背景

虽然已有研究报道了心血管病 (CVD) 肠道微生物的研究, 但是 CVD 中肠道微生物功能和饮食 - 微生物 - 代谢 - 免疫相互作用的研究依旧不多, 因此需要通过多组学基于系统生物学的方法去探究 CVD 中饮食 - 微生物 - 代谢 - 免疫相互作用。

★ 研究方法

LifeLines-DEEP队列 (LLD) : 本研究中 LLD 队列是荷兰 LifeLines 队列中的子队列, 剔除了 57 名服用抗生素或降脂药物和11名非禁食受试者的参与者, 剩下的978名受试者 (411名男性和567名女性) 的粪便宏基因组、血清代谢组和详细的饮食、12种炎症标志物和5种粪便的短链脂肪酸 (SCFAs) 用于分析。

300肥胖队列 (300-OB): 纳入了体重指数 (BMI) > 27 kg / m², 年龄在 55 至 80 岁之间的受试者, 进行了详细的心脏代谢表型分析, 包括评估颈动脉斑块和测量皮下和内脏脂肪组织和肝脏脂肪含量。队列纳入要求比较严格, 排除各种手术史、用药史和近期心脏事件。

★ 研究结果

1. 经过质控和估计~2.2%缺失值后, 确定了188个微生物物种、562个细菌代谢途径和231个代谢性状进行后续的关联分析。在校正年龄、性别和BMI后, 代谢物分别与物种和代谢途径在两个队列的关联情况如下:

	LLD	300-OB
微生物代谢物	210个代谢物与12个物种存在64个关联关系	42个代谢物与微生物物种存在一种关联
微生物pathway	210个代谢物与308个pathway存在4,135个关联关系	42个代谢物与19个pathway存在105个关联关系

其中, 300-OB关联数较少可能原因是受到样品数的影响。但结果显示LLD中鉴定的微生物因子通常对300-OB的代谢变异具有较低的预测值, 进一步比较了两个队列的关联强度和方向, 虽然300-OB样品数少, 但是存在一些LLD没有相关关联, 特别是脂蛋白亚类, 同时有些关联在两个队列相关方向是相反的。这些结论表明基于人群和肥胖群体之间的微生物关联存在显著差异。

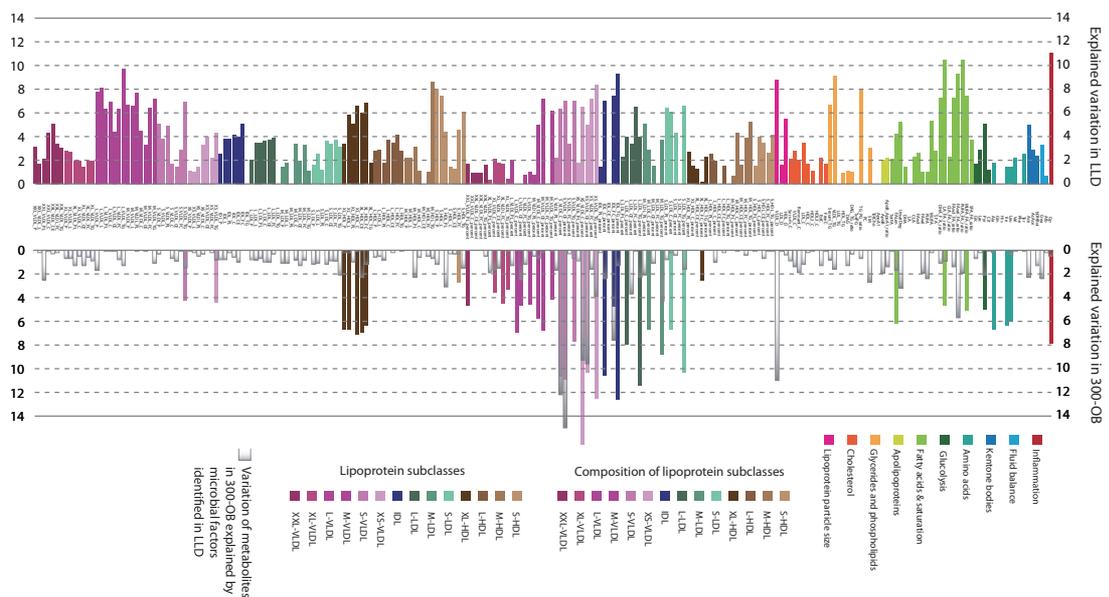


图6 肠道微生物组解释的血浆代谢物变化

2.为了探索血浆代谢物相关的微生物因子是否与心血管疾病临床表现相关，作者重点关注在2个队列中相关系数排前的物种和pathway，并计算它们与26个心脏临床表型的相关性。结果显示，在300-OB队列中，高丰度的细菌-L甲硫氨酸生物合成与斑块存在和最大狭窄程度显著相关；瘤胃球菌sp_5_1_39BFAA与肝脏脂肪含量 (Liver-fat) 呈正相关。虽然已有文献报道肠道微生物组衍生的TMAO会增加心血管疾病的风险，然而作者观察到本研究中，TMAO的血浆水平与内脏脂肪呈正相关，但与动脉粥样硬化表型和肝脂肪不相关，也与代谢物相关物种和细菌通路不相关。

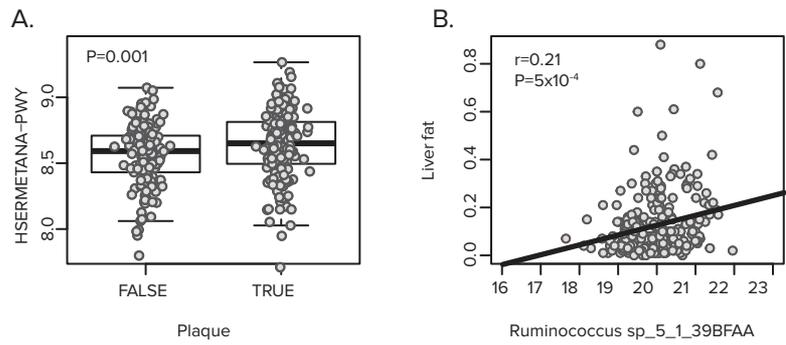


图7 300-OB队列中肠道微生物通路和物种与临床结果的关联

3.评估人群队列中CVD病情发展的个体代谢风险，作者使用33个确定与未来CVD事件相关但又不和已知风险因素（年龄、性别、BMI和吸烟状况）相关的代谢物biomarkers，基于加权风险模型构建个体的CVD代谢风险评分（MRS）。在LLD队列中，发现了48个微生物pathway与MRS关联。另外，还鉴定了与MRS关联的pathway的驱动物种（drive species），结果表明顶级分类群对MRS相关的微生物pathway贡献差异很大，相关系数介于0.26到0.89之间，平均值为0.60。这表明一些pathway是由一种占主导地位的微生物驱动，而其他pathway可能由许多不同的微生物驱动。

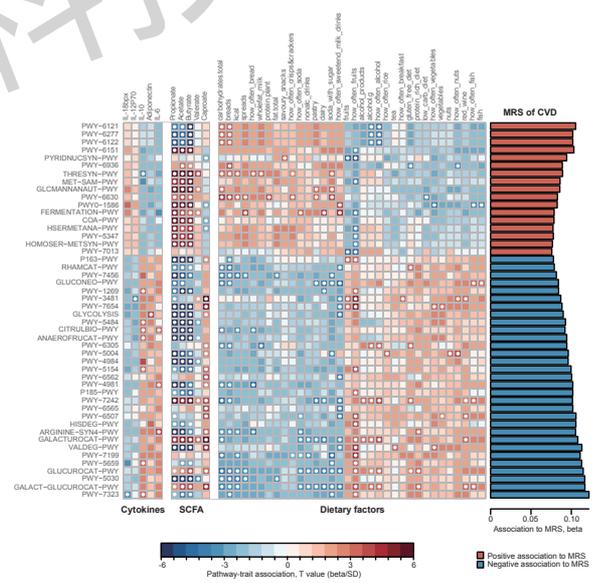


图8 48个生物通路与CVD MRS、粪便SCFAs、细胞因子和饮食因素的关联

4.作者还深入探讨了宿主-微生物-饮食相互作用对代谢和炎症的影响，将与MRS相关的48个细菌pathway与血浆水平的12种细胞因子和78个饮食因子进行系统整合分析，结果发现12个pathway和5种细胞因子存在14种显著关联。已知饮食会影响代谢，CVD风险和肠道微生物，结果发现78个饮食因素中有34个与MRS相关的微生物通路有关。MRS与饮食关联发现，水果和蔬菜的摄入量与48个MRS相关的pathway有关，对于这个现象可能的解释机制是细菌通过发酵膳食纤维产生SCFAs以影响CVD。于是作者测定了粪便中5种SCFAs浓度，结果表明这5种SCFAs与大部分的MRS相关的pathway显著关联，进一步分析了血浆中的SCFA浓度最高的乙酸盐，正如预期一样，血浆乙酸盐的浓度与低MRS和29个MRS相关的pathway相关。

★ 参考文献

Liu R, Hong J, Xu X, et al. Gut microbial associations to plasma metabolites linked to cardiovascular phenotypes and risk[J]. Circulation Research, 2019, 124(12):1808-1820.



送样要求

▶▶ 代谢产品送样要求

样本类型	建议送样量/例	样本类型	建议送样量/例
血清、血浆	$\geq 250 \mu\text{L}$	微生物、细胞	$\geq 10^7$ 个
尿液	$\geq 500 \mu\text{L}$	培养液	$\geq 1 \text{ mL}$
动物或临床组织	$\geq 200 \text{ mg}$	植物组织	$\geq 1\text{g}$
粪便、肠道内容物	$\geq 200 \text{ mg}$		

华大科技BGITech

华大科技BGITech

产品
简介

Product
Profile

合成生物学

普通基因合成	131
PCR克隆	134
定点突变	135
普通引物合成	136
修饰引物合成	137
Oligo Pools高通量芯片合成	139
代谢通路设计合成一站式服务	140

普通基因合成

▶▶ 产品概述

通过高通量、大规模、全自动的基因合成平台，采用自主研发的iBEST难度基因解决方案，利用先进的合成生物学技术可以解决以往从自然界获取DNA的复杂过程，提供任意复杂结构的长难度基因，包括高GC、高AT、单碱基及多碱基重复、长片段DNA的合成服务，同时提供针对不同物种合成基因的密码子优化服务，能够更好地满足科研工作者对基因、基因组水平DNA合成与组装更加高效、低成本、大规模的科研需求。

▶▶ iBEST技术原理介绍

我们基于iBEST™等温双向延伸合成技术Isotherma Bidirectional Extension Synthesis Technology基于特殊设计的茎环单链引物合成，配合其等温延伸组装酶系统的一体化开发。该技术可实现任意定制化基因合成，不依赖于PCR，具有易操作、高精度的特点，尤其是对各种复杂结构基因序列的合成具有明显技术优势。



→ 扫码观看iBEST技术原理

▶▶ 产品优势



任意复杂结构



任意载体克隆



自主知识产权



7*24h在线订购



最快4个工作日交付

▶▶ 技术流程



序列分析



片段拼接



克隆



筛选



核对测序结果



冻干粉质粒

▶▶ 交付周期

长度范围	交付周期
<1Kb	极速产品4个工作日
<1.5Kb	5-8工作日
1.5Kb-5Kb	8-13工作日
5Kb-8Kb	13-18工作日
8Kb-10Kb	18-23工作日
>10Kb	咨询

*仅限于10Kb以内常规基因序列

▶▶ 订购指南

- 1.填写《全基因合成订购单》，发送订单至syn-service@genomics.cn
- 2.通过客户自助系统下单<http://customer.genomics.cn>
- 3.联系当地销售

▶▶ 交付结果

产品	报告
>2ug/样，含有目的片段的高纯度质粒干粉DNA和含有重组质粒的穿刺菌	包括标准序列、测序结果、拼接文件、质粒图谱、发货样品 (COA) 报告、产品说明书

PCR克隆

产品概述

PCR克隆，即将一个基因从一个载体转移到另一个载体的过程。PCR克隆可以为客户量身打造DNA克隆解决方案，可以在克隆流程中的任意环节满足客户的各种需求，其中包括模板的获得、PCR产物的纯化、克隆方法的选择、载体的选择、后续的转化和质粒提取。

产品优势



高效无痕



操作规范



提供一站式改造方案



准确快速

技术流程



订购指南

1. 填写《PCR 克隆订购单》，发送订单至 syn-service@genomics.cn
2. 通过客户自助系统下单 <http://customer.genomics.cn>
3. 联系当地销售
4. 长度范围不限，价格参看评估结果

交付结果

产品	报告
>2ug/样, 含有目的片段的高纯度质粒干粉 DNA和含有重组质粒的穿刺菌	包括标准序列、测序结果、拼接文件、质粒图谱、 发货样品 (COA) 报告、产品说明书

产品应用

1. 农作物基因育种
2. 工业微生物菌株基因改良
3. 酶制剂体外进化改造
4. 抗体工程等领域中关键基因的体外克隆和大量制备

定点突变

产品概述

根据客户对突变需求的不同，在保留原始序列完整性的基础上，实现原基因序列中的任意位置突变，包括单碱基、多碱基的突变、缺失部分碱基以及插入部分碱基等，同时对原始模板基因突变体产物进行测序验证及克隆制备。

产品优势



经验丰富



性价比高

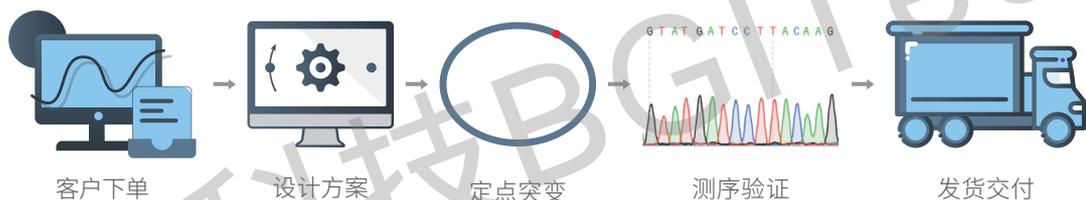


完善的质粒改造与设计系统



提供定制化服务

技术流程



订购指南

- 1.填写《定点突变订购单》，发送订单至 syn-service@genomics.cn
- 2.通过客户自助系统下单<http://customer.genomics.cn>
- 3.联系当地销售
- 4.长度范围不限，价格参看评估结果

交付结果

产品	报告
>2ug/样，含有目的片段的高纯度质粒干粉 DNA和含有重组质粒的穿刺菌	包括标准序列、测序结果、拼接文件、质粒图谱、 发货样品 (COA) 报告、产品说明书

产品应用

应用于抗体工程、酶工程、农作物性状改良等科研及工业应用领域中的功能基因序列修改或编辑再造

普通引物合成

产品概述

提供DSL、PAGE plus、PAGE、HPLC等多种纯化方式及合成规格，兼顾为客户提供个性化的寡核苷酸合成解决方案，可以满足客户的不同实验需求。

产品优势



纯化方式多样



提供个性化服务



设备先进



通量大速度快

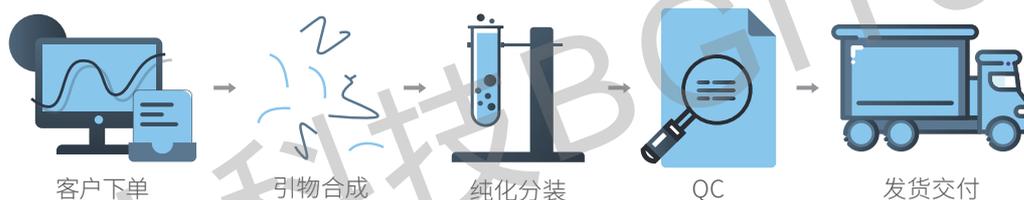


7*24h在线订购



图 Dr. oligo 192合成仪

技术流程



订购指南

1. 填写《引物合成订购单》，发送订单至order@service.genomics.cn
2. 通过客户自助系统下单<http://customer.genomics.cn>
3. 联系当地销售
4. 纯化方式选择指南

纯化方法	纯度	纯化原理	纯化效果	适用长度	使用范围
DSL	>70%	C18反向吸附柱纯化	可除去盐离子，无法去除合成失败的片段	<30nt	仅限普通PCR
				60-80nt 定制产品	定制产品可满足基因合成所需
HPLC	>90%	利用疏水性不同，通过高效液相色谱区分合成失败片段，收集正确引物	除盐，去除大部分合成失败片段	15-50nt	用于修饰、探针或纯度要求高的引物
PAGE plus	>85%	基于高品质纯化柱对DMT基团的高特异性吸附，分离合成产物；纯度优于DSL且效率高于PAGE	除盐，去除大部分合成失败片段	15-50nt	适用于普通、多重及定量等各种PCR反应，以及PCR产物用于克隆表达或基因重组、基因构建、全基因合成、亚克隆、点突变等
PAGE	>85%	利用聚丙烯酰胺凝胶电泳区分合成失败片段，特别适用于长片段	除盐，去除大部分合成失败片段	15-100nt	适用于普通、多重及定量等各种PCR反应，以及PCR产物用于克隆表达或基因重组、基因构建、全基因合成、亚克隆、点突变等

交付结果

产品	报告
对应合成量的引物干粉	引物报告单及使用说明书

产品应用

PCR扩增、基因克隆、定点突变、基因合成、DNA测序，为寡核苷酸药物、DNA新材料研究及应用提供关键的寡核苷酸原材料

修饰引物合成

产品概述

华大基因提供多种类型的修饰/标记引物，包括磷酸化、氨基、巯基及各种Spacers修饰、生物素、地高辛、荧光及碱基修饰、双标荧光探针、分子信标等，同时通过HPLC、MS等技术可实现序列纯度和准确度的验证。

产品优势



设备先进



通量大、速度快

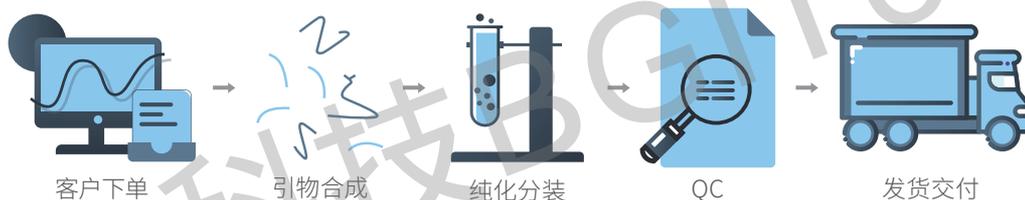


多种修饰类型



提供个性化服务

技术流程



订购指南

1. 填写《修饰引物合成订购单》，发送订单至 bgi-mp@service.genomics.cn
2. 联系当地销售
3. 修饰类型

5'修饰	5'磷酸化 (Phosphorylation)、5'氨基 (Aminolinker C6)、5'巯基 (SH C6)、5'生物素 (Biotin)、5'FAM、5'TET、5'HEX、5'JOE、5'ROX、5'TAMRA、5'CY3、5'CY5、5'地高辛 (Digoxin)
3'修饰	3'磷酸化 (Phosphorylation)、3'氨基 (Aminolinker C7)、3'巯基 (SH C6)、3'生物素 (Biotin)、3'Biotin TEG、3'地高辛 (Digoxin)、3'FAM、3'BHQ1、3'BHQ2、3'JOE、3'ROX、3'CY3、3'CY5、3'Dabcyl、3'TAMRA
中间修饰	硫代 (Phosphorothioate)、甲基化 (5-Methyl dC)、C3/C9/C12 Spacer、dSpacer、dU (deoxyUridine)、dI (deoxyInosine)、dT-Aminolinker C6、dT-Biotin、dT-Dabcyl、dT-Digoxin、dT-CY5、dT-CY3、dT-TAMRA、dT-FAM、dT-ROX、LNA
双标引物	5'-FAM and 3'-TAMRA、5'-FAM and 3'-BHQ-1、5'-CY3 and 3'-BHQ-1、5'-FAM and 3'-BHQ-2、5'-FAM and 3'-Dabcyl、5'-FAM and 3'-MGB、5'-FAM and 3'-Eclipse、5'-ROX and 3'-BHQ-2、5'-ROX and 3'-Eclipse、5'-Rox and 3'-MGB、5'-CY5 and 3'-MGB、5'-CY5 and 3'-BHQ-2、5'-HEX and 3'-BHQ-1、5'-HEX and 3'-Eclipse、5'-TET and 3'-BHQ-1、5'-TAMRA and 3'-Phos、5'-VIC and 3'-BHQ-1、5'-NED and 3'-BHQ-1、5'-VIC and 3'-TAMRA、5'-VIC and 3'-MGB、5'-VIC and 3'-BHQ-2、5'-TAMRA and 3'-BHQ-2、5'-HEX and 3'-MGB、5'-CY5 and 3'-Dabcyl、5'-Phos and 3'-FAM、5'-VIC and 3'-Dabcyl、5'-TET and 3'-TAMRA

*如需其他特殊碱基或修饰类型请咨询当地销售或拨打客服热线 400-706-6615

▶ 交付结果

产品	报告
合成的引物干粉	引物报告单及使用说明书

▶ 产品应用

应用于体外诊断、法医检测、高通量测序、生物标记等修饰探针及标记寡核苷酸合成，为医学检测及生物诊断等领域的科研及应用提供关键技术材料。

华大科技BGITech

Oligo Pools高通量芯片合成

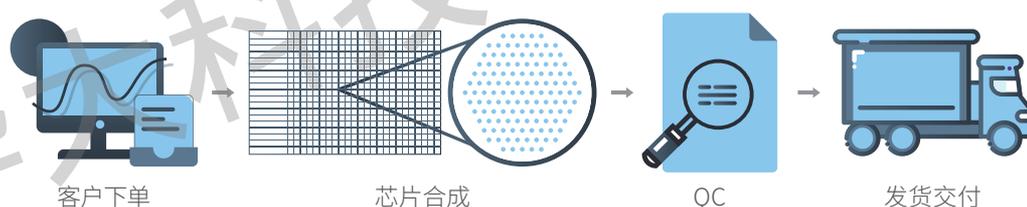
▶▶ 产品概述

在数万个半导体阵列上单独进行特定DNA序列的合成，合成后将引物从芯片上洗脱下来，即可得到含有数万条引物的混合溶液，即引物池 (Oligo Pools)。Oligo Pools经扩增和转录制备的探针，可用于数万次捕获，极大降低捕获成本；转录生成的RNA探针，均一性好，确保了较高的目标区域覆盖度和捕获效率；Oligo Pools转录生成的RNA探针可与常用NGS文库构建试剂盒兼容。华大基因可根据客户的应用需求，针对性地设计寡核苷酸序列文库以及包含不同寡核苷酸分子的混合文库。

▶▶ 产品优势



▶▶ 技术流程



▶▶ 订购指南

1. 发送订单至 bgi-mp@service.genomics.cn
2. 联系当地销售
3. 订购规格参考

合成类型	单链 DNA
长度范围	20-300nt
容量 (条)	2k-696k
引物量	>0.25fmol/条
合成周期	20个工作日

▶▶ 交付结果

成品	报告
Oligo Pools冻干粉	合成报告单

▶▶ 产品应用

应用于大规模的微生物、动物、植物细胞基因编辑改造；高通量基因组测序捕获探针合成；以DNA为介质的数据存储；为生命科学基础及应用研究提供关键的寡核苷酸库上游原料。

代谢通路设计合成一站式服务

产品概述

高通量合成与筛选手段为代谢工程与合成生物学研究提供了越来越多的可能，也为科研工作者提供了更多的可操作性，相对于传统代谢通路优化的方法，仅关注于代谢途径中的某几个基因，通过质粒过表达或者敲除敲入来调整表达强度，效率较低且容易引起中间产物积累，经费较高。华大基因特别推出代谢通路合成组装服务，显著降低了代谢通路整体设计的难度和成本。

产品优势



强大的基因合成能力



独有的难度基因解决方案



超过10年合成服务经验



从方案设计到后续验证全方位服务



专业的生物信息学分析系统

技术流程

前期沟通

项目调研

方案设计

项目实施

结果交付

服务内容

华大智能设计与基因合成平台可以为客户提供：

- 1.代谢挖掘、数字建模、合成基因通路、实验验证、菌株构建、启动子、RBS、基因的高通量模块化合成及客户定制的个性化服务；
- 2.可根据客户要求交付合成代谢通路的质粒或片段；
- 3.基于酵母项目实战，强大的片段整合能力，可根据客户需求特别定制代谢通路文库组装服务。

华大科技BGITech

产品
简介

Product
Profile

| 基因科技 造福人类
Omics for All

华大科技BGITech

特色产品、服务

Dr. Tom	143
单管单细胞DNA测序	146
单管单细胞RNA测序	149
10X Genomics单细胞RNA-Seq	153
10X Genomics 单细胞ATAC-Seq	157
10X Genomics单细胞免疫组库	161

▶▶ 产品概述

Dr. Tom是华大基因自主开发的一款多组学、交互式数据挖掘系统。它能够帮助科研人员快速的对基因组、转录组、蛋白质组和代谢组的数据信息进行深度挖掘，能够在复杂的多组学数据中快速找到目标基因和相关基因的调控通路。该系统整合了多个国际常用的基因数据库和生物信息分析工具，并优化了数据挖掘的方式和展现形式，还结合了云计算、人工智能和机器学习技术。通过该系统，科研用户可以方便的进行自助式多组学数据挖掘。

自2018年上线以来，该系统围绕“交互式”、“智能”的核心优势，给多组学大数据的分析挖掘带来了极大的便利性，分析效率提高50倍以上，获得了市场的高度认可。2019年5月，Dr. Tom 2.0上线，加强了“多组学”的数据挖掘内容——增加了非编码RNA模块，还陆续开放了蛋白组和代谢组的分析模块，同时扩展了50个可应用物种数据库。目前，该系统在全球已具有极强的竞争力，并且还在不断的升级和优化。

▶▶ 产品优势



测“一”得“多”

选中的目的基因，可通过系统中关联转换小工具，即可根据互作关系(PPI、Target、Coexpression、ceRNA、GGI、RNAplex)或者位置关系(上下游位置范围)查看所有与所选基因相关联的其他类型RNA，一键可视化互作关系，轻松解决大小RNA关联分析，ceRNA分析等分析热点和难点



多数据库联合分析，定向数据调取

除了常见的NCBI数据库之外，Dr. Tom增加了TCGA、ARCS4等多数据库联合分析，深度挖掘数据，拉取目的基因在肿瘤或常规组织中的表达情况，为后续分析方向提供更多可能。在任意基因表格中通过扩展列-外部数据库-即可在TCGA、ARCS4数据库中调取感兴趣的研究方向，选取后即可在表格中看到基因在外部数据库中的表达情况



告别查阅，一键读取基因&文献详情

系统界面中直接点击感兴趣的基因即可进入基因详情页，可以直接获取基因的注释链接、序列、表达情况，关联已发表的文献列表，文献直接点击即可进入阅读界面。同时，在绘制好的基因互作图中，直接点击两个基因之间的连接线，即可弹出支持互作关系的文献列表，一键点击即可进入阅读界面



多维度数据展示，低分到高分思维转变

在高分文章中，多维度的图片展示已经成为不可或缺的一环，Dr. Tom可以实现这一诉求。如表达量聚类图中，可添加多个内容作为纵向/横向分类，在常规的热图基础上，增加第三维度的分类信息，系统中多个作图环境都可以实现更高维度作图需求，助力高分文章发表



自定义模块，更个性化的数据处理

历史项目数据也可以上传至系统进行分析，可自定义上传同一参考基因组版本的其他定量数据，使用丰富的小工具模块自己分析。除此之外，还可以针对目的基因的特征，重新定义基因分类，更加个性化地整合单项目研究重点

系统工具展示

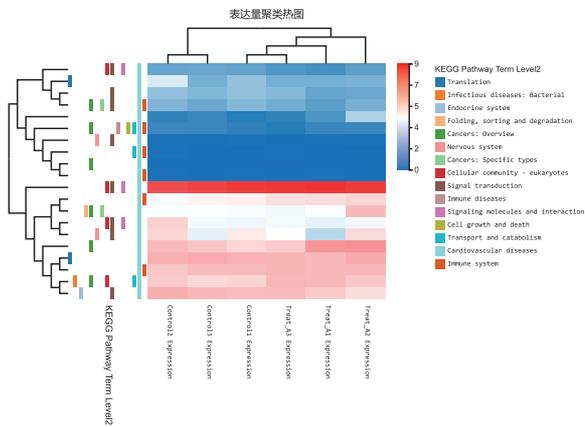


图1 多维度表达量聚类热图

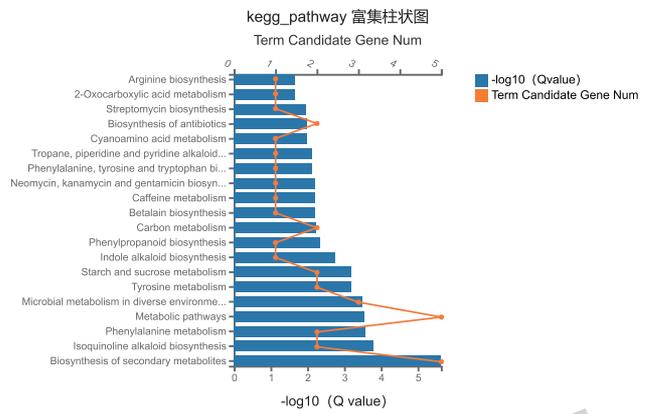


图2 可切换气泡/柱状富集图

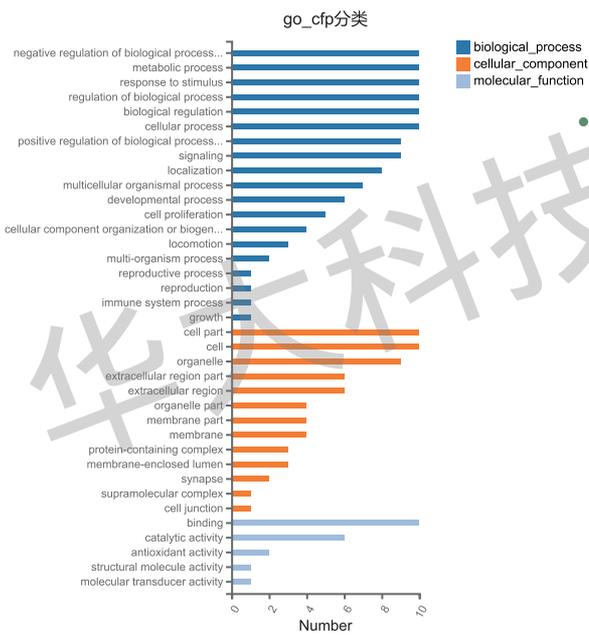


图3 多数据库可选, 分类图

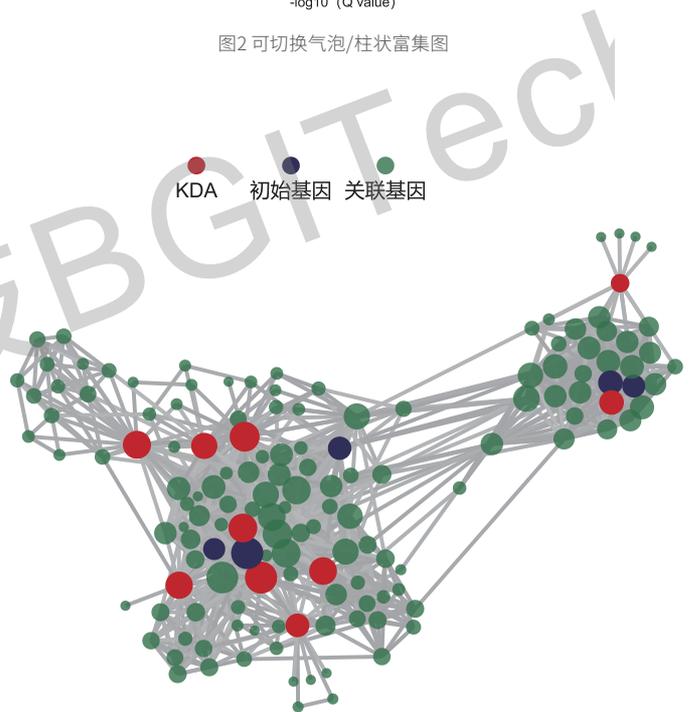


图4 关键驱动基因分析, KDA图

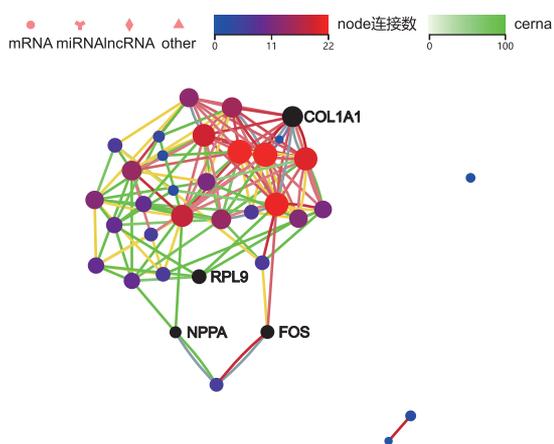


图5 多种互动关系一键展示, 网络互作图

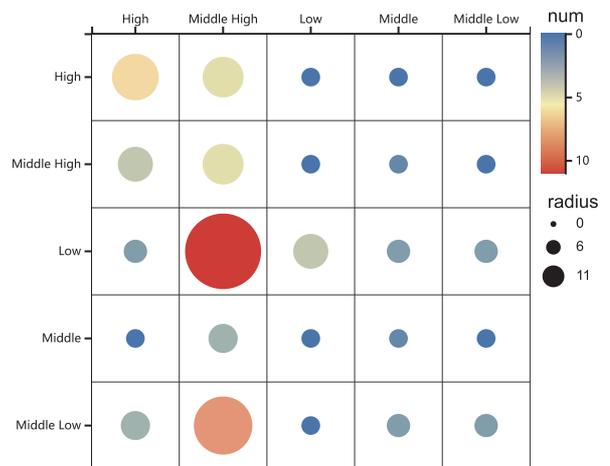


图6 卡方检验

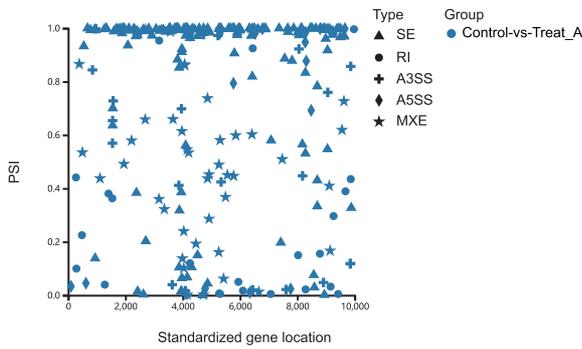


图7 可变剪切分析

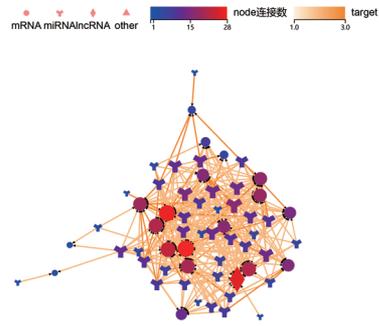


图8 关联网络图

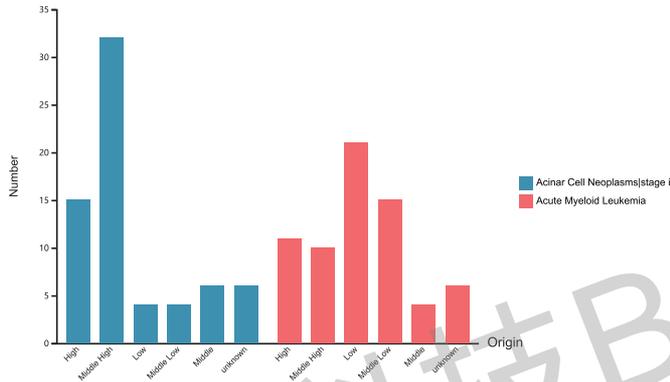


图9 外部数据库-多组学关联分析

常见问题

Q: 非华大测序的数据是否可以用Dr. Tom进行分析?

A: 可以, 需要信息分析人员重新处理后即可上传Dr. Tom分析, 如有此项需求可以联系驻地销售。

Q: 是否可以试用Dr. Tom系统?

A: 可以, 您可以联系驻地销售获取试用账号进行试用分析。

Dr. Tom 质谱云 (用于蛋白组和代谢组数据) 可在网站上自行申请账号密码注册登陆, 或仅以“游客身份”即可登陆试用。

Q: Dr. Tom 目前可以交付的项目类型是?

A: Dr. Tom RNA方向可以交付转录组, 小RNA和lncRNA分析, 覆盖物种50余种;

Dr. Tom质谱云可以交付蛋白iTRAQ定量分析、蛋白DIA定量分析、非靶向代谢组等; 如需评估项目是否适用, 请联系驻地销售。

Q: Dr. Tom是如何收费的?

A: 华大RNA和质谱方向的产品直接搭配Dr. Tom交付, 使用期限为一年, 这部分无额外费用。若是其他公司的测序数据, 需要支付信息分析的费用即可使用Dr. Tom, 具体资费情况可以联系驻地销售。

华大合作发表文章 (部分)

研究内容	发表日期	发表期刊	文献名称
非小细胞肺癌	2019	Cell Death & Disease	STYK1 promotes tumor growth and metastasis by reducing SPINT2/HAI-2 expression in non-small cell lung cancer
混合系白血病	2019	Clinical Epigenetics	Co-inhibition of HDAC and MLL-menin interaction targets MLL-rearranged acute myeloid leukemia cells via disruption of DNA damage checkpoint and DNA repair



单管单细胞DNA测序

产品概述

单管单细胞DNA测序首先使用口吸管分离的方法对细胞悬液进行细胞挑选，针对分离的单细胞进行高保真高覆盖度全基因组扩增，然后对扩增产物进行全基因组、外显子或者目标区域测序，再对高通量测序产出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，用于揭示细胞群体差异和细胞进化关系。

产品优势



技术稳定性好、重复性高
平均可检测>95%的基因组位点



高质量的扩增和极低的扩增错误率(10^{-6} - 10^{-7})
扩增产物长度可达20Kb, 有利于更多类型的变异检测



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



DNBSEQ自主平台, 丰富的单细胞分离经验

产品应用



技术流程



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: PE100或PE150

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: 单细胞测序揭示三阴性乳腺癌进化历史

采用外显子组测序, 对新辅助化疗 (NAC) 治疗前、治疗中、治疗后纵向检测20个三阴性乳腺癌 (TNBC) 患者的组织样本, 发现10例患者在NAC治疗后导致突变消失, 而剩下的10例患者在治疗后突变仍然存在。研究者对其中8例患者进行了更详细的研究, 使用单细胞DNA测序分析900个细胞, 探索8个病人的拷贝数进化, 进行深度为0.1X单细胞的WGS测序。对于新辅助化疗 (NAC) 治疗敏感的患者, DNA测序的CNV结果可以体现治疗前后明显的差异。为了说明四个克隆消亡患者的克隆亚结构, 研究者鉴定了群体中所有细胞的共同染色体截断点, 并进行了最佳聚类 and T-SNE预测。这一分析确定了每个患者的2个到3个非整倍体肿瘤细胞群和1个正常的二倍体细胞群。

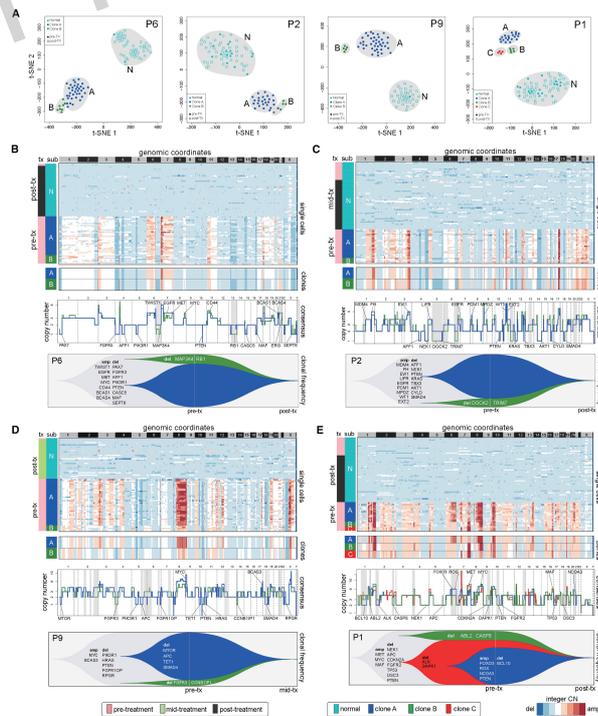


图 T-SNE确定了治疗前的三名患者 (P2, P6, P9) 中两个主要克隆, 另一名患者 (P1) 中有三个主要克隆, 以及全基因组范围拷贝数变化热图

★ 参考文献

Kim C, Gao R, Sei E, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing[J]. Cell, 2018,173(4):879-893.e13.

► 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%。

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3-5 μ g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扩增的特点是什么?

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	文献名称
多发性骨髓瘤	2018	Blood	Single cell exomes in an index case of amp1q21 multiple myeloma reveals more diverse mutanomes than the whole population
乳腺癌	2018	Journal of Clinical Investigation	Coexisting genomic aberrations associated with lymph node metastasis in breast cancer
CTC	2017	Cell Biology and Toxicology	Progress and challenges of sequencing and analyzing circulating tumor cells
PGD检测染色体异常	2017	European Journal of Medical Genetics	Massively parallel sequencing on human cleavage-stage embryos to detect chromosomal abnormality
CLL	2016	Nature Communications	Evolution of multiple cell clones over a 29-year period of a cll patient
膀胱癌干细胞	2016	European Urology	Single-cell sequencing reveals variants in arid1a, gprc5a and mll2 driving self-renewal of human bladder cancer stem cells



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。该技术可以解决传统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

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)

▶▶ 技术参数

样品要求:

单细胞: 细胞直径大于10 μ 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 1.0, RIN \geq 7.0, 浓度>50pg/ μ 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 μ L裂解液)
微量细胞		2 \leq X \leq 200 (4 μ 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: PE100建议6-10Gb clean data

技术流程



信息分析内容

PE100	
(需要提供参考基因组序列) 标准及高级信息分析	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, ④RNA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) 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
	Dr. Tom信息分析 数据库注释 1.转录因子注释 (AnimalTFDB/PlantTFDB) 2.Rfam、Pfam、Reactome、COG、EggNOG和 InterPro数据库注释 3.GSEA分析 互作网络分析 1.靶基因分析 ①miRNA-mRNA靶向关系分析 ②lncRNA-mRNA靶向关系分析 2.ceRNA互作网络分析 ①lncRNA-mRNA联合分析 ②circRNA-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
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。

案例分析

案例一: DNBSEQ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DNBSEQ平台单细胞转录组测序数据各个指标。由图1中可以看出DNBSEQ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。

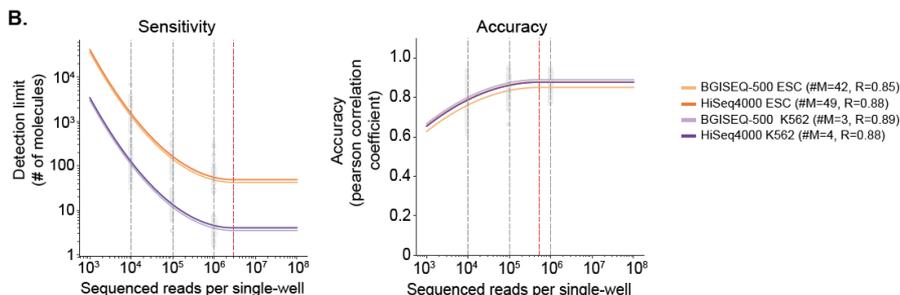


图1 DNBSEQ平台和Illumina平台单细胞转录组数据灵敏度和准确度统计

参考文献

Natarajan K, Miao Z, Jiang M, et al. Comparative analysis of sequencing technologies platforms for single-cell transcriptomics[J]. Genome Biology, 2019, 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）相关疾病激发的疼痛感可明显的联系到与特定的一种痛感神经元相关。单细胞RNA-Seq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。

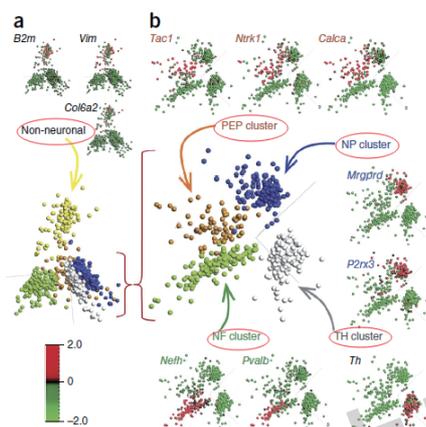


图2 细胞表达量PCA聚类图
图中每一个点为一个细胞，颜色反馈的是表达量的高低，根据表达谱的差异，把细胞聚成了5个类群

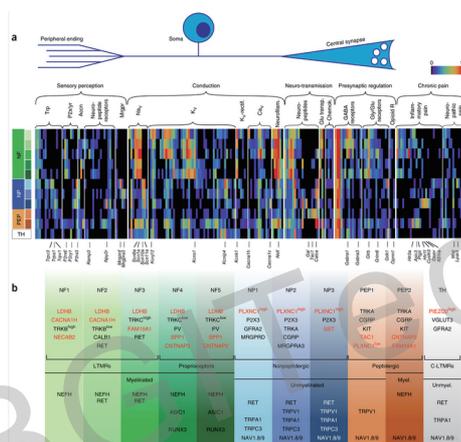


图3 通过对不同类群神经元进行表达量热图分析，找到亚群特征性的基因

★ 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.

▶▶ 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 μ 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，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	文献名称
mESCs和K562两种细胞	2019	Genome Biology	Comparative analysis of sequencing technologies for single-cell transcriptomics
人多能干细胞	2018	GigaScience	Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation
小神经胶质	2018	Nature Neuroscience	Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion
Hela细胞系	2016	Biotechnology Letters	Single-cell RNA-seq reveals lincRNA expression differences in HeLa-S3 cells
早期人类胚胎形成	2016	BMC Genomics	Single-cell RNA sequencing reveals dynamic changes in A-to-I RNA editome during early human embryogenesis
膀胱鳞状细胞	2016	Oncotarget	Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder
骨间质干细胞	2016	PLoS One	RNA-seq reveals the angiogenesis diversity between the fetal and adults bone mesenchyme stem cell
HeLa细胞系	2015	GigaScience	Full-length single-cell RNA-seq applied to a viral human cancer: applications to HPV expression and splicing analysis in HeLa S3 cells
人鼠胚胎发育	2013	Nature	Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing



10X Genomics单细胞RNA-Seq

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



丰富的项目经验
合作伙伴遍布全球，多项科研成果在顶尖杂志上发表



前沿热点信息分析内容
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



“0”单细胞数据拆分错误
搭载DNBSEQ™测序系统，避免样本间单个细胞数据弄混



一站式服务呵护您的样品
独家流式分选服务，一站式实现从组织到数据，避免细胞反复冻融



样本多样化
除常规活细胞悬液外，珍稀液氮速冻样本皆可实现单细胞水平研究

产品应用

免疫	神经元	肿瘤	植物
血液免疫细胞亚群 稀有细胞亚群监测 血液免疫疾病研究	神经元起源研究 神经元的异质性研究	肿瘤异质性研究 肿瘤的发生发展研究 肿瘤愈后检测研究	细胞亚群鉴定 稀有细胞类型鉴定

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 ⁵ /ml
	细胞悬液背景干净，无大量结团、碎片及杂质，不含Ca ²⁺ 和Mg ²⁺

测序策略：DNBSEQ™平台PE100

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（外周血单核细胞）样本，建议每个细胞平均测：50,000条Raw Reads。

技术流程



图1 10X Genomics平台单细胞RNA-Seq技术流程图

信息分析内容

信息分析条款	信息分析内容
<p>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)</p>	<ol style="list-style-type: none"> 1.测序结果统计 2.数据质控统计 3.比对结果统计 4.基因表达定量分析 5.Marker基因鉴定 6.细胞类型注释 7.单样本/多样本细胞聚类分析 8.特异marker基因差异分析 9.样本间相同cluster基因差异分析 10.Cluster差异表达基因GO功能分析 11.Cluster差异表达基因Pathway功能分析 12.Cluster差异表达基因TF编码能力预测 13.Cluster差异表达基因蛋白互作分析 14.Cluster基因相关性网络分析
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

案例分析

案例一: DNBSEQ™平台脐带血细胞特异性研究

研究人员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。

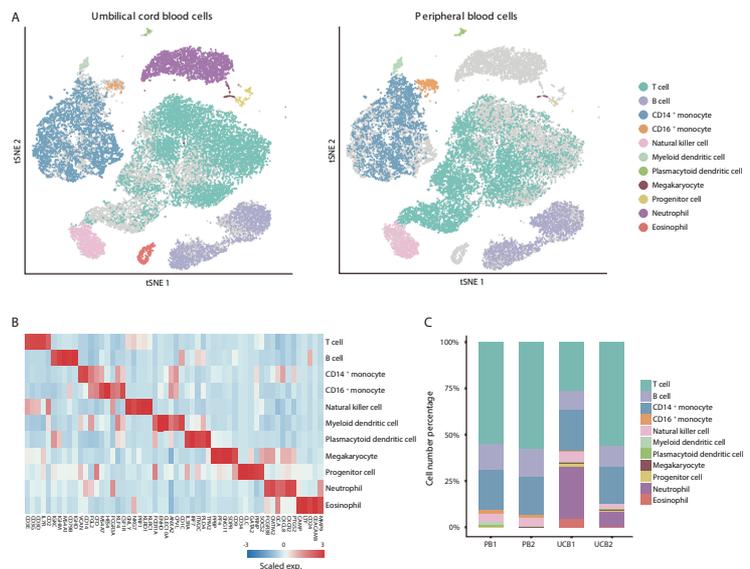


图1 外周血和脐带血不同细胞类型和比例

A. 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

B. 脐带血和外周血中主要的已知标记的基因表达热图

C. 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

★ 参考文献

Zhao Y, Li X, Zhao W, et al. Single-cell transcriptomic landscape of nucleated cells in umbilical cord blood[J]. GigaScience, 2019, 8(5): giz047.

案例二：10X单细胞转录组研究植物单细胞图谱

通过对10,000多个拟南芥 (*Arabidopsis thaliana*) 根细胞的原生质体进行单细胞转录组研究,发现所有主要组织和发育阶段都在这个单细胞转录组群中表现出来。此外,还发现了不同的亚群和罕见的细胞类型,对根表皮细胞转录本的分析确定了个体细胞从分生组织向根毛和非毛细胞分化的成熟阶段发展的轨迹。本研究证明了高通量单细胞转录组研究在植物中的可行性和实用性,并提供了拟南芥根单细胞分离的第一代单细胞基因表达图谱。

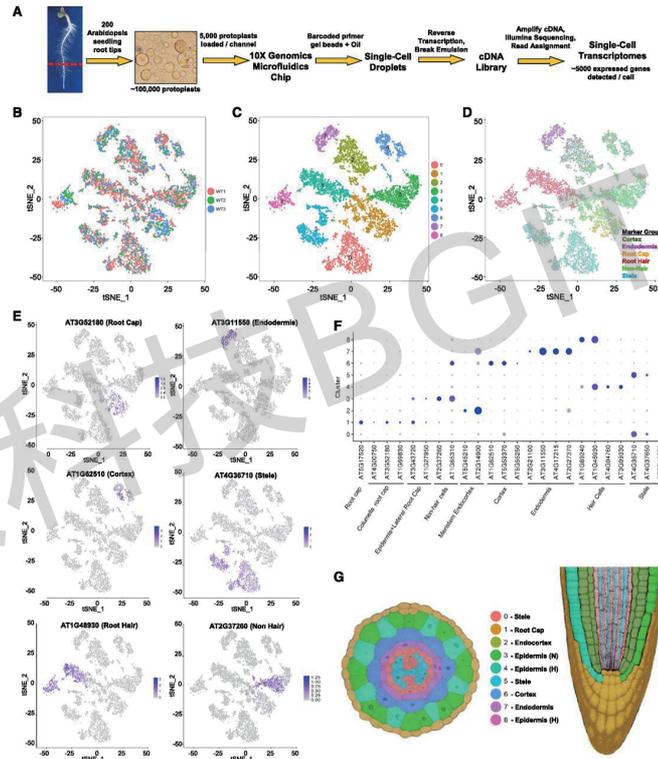


图2 野生型拟南芥根单细胞转录组的分离及聚类分析

A. 单个拟南芥根细胞转录组研究工作流程。B. t-sne结果图从三个生物重复的7522个个体野生型细胞转录本的空间分布。从每个生物学重复中获得的细胞转录本用不同的颜色表示(红色表示重复样本1;绿色表示重复样本2;蓝色表示重复样本3)。C. t-sne结果显示了7522个个体野生型根细胞转录子的9个主要簇。D. t-sne结果图显示了7522个野生型转录本上所有检测的标记基因的转录本积累,细胞类型分别是皮质、内皮层、根冠、根毛等。E. t-sne结果图显示特定组织/细胞标记基因的转录本积累。颜色强度表示每个细胞中指示基因的相对转录水平。F. 各细胞类型marker基因转录表达聚类图,横坐标表示9个细胞类型的24个marker基因,纵坐标表示各基因表达情况。G. 将细胞簇分配给根组织。拟南芥主根横切面(左)和纵切面(右)的描绘, H. 根毛细胞九个细胞类型图与相对应的组织示意图。

★ 参考文献

Ryu K, Huang L, Kang H, et al. Single-cell RNA sequencing resolves molecular relationships among individual plant cells[J]. Plant Physiology, 2019, 179(4): 1444-1456.

▶▶ 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: 细胞大小会有一定的限制,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<40μm。

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: 如果客户分离了1,000个细胞, 并检测出细胞活性很高, 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 5×10^5 个细胞, 且活率建议在80%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)

发表时间	发表期刊	文献名称
2020	Cancer Cell	An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates
2020	Cell	Single-cell transcriptome atlas of murine endothelial cells
2019	Functional & Integrative Genomics	Single-cell transcriptome provides novel insights into antler stem cells, a cell type capable of mammalian organ regeneration
2019	GigaScience	Single-cell transcriptomic landscape of nucleated cells in umbilical cord blood
2018	Nature Neuroscience	Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion



10X Genomics单细胞ATAC-Seq

▶▶ 产品概述

基于10X Genomics平台的高通量单细胞ATAC-Seq技术是利用液滴法的原理，使用Tn5酶进入核膜对细胞核进行预处理后，使用GemCode技术，通过控制微流体的进入，将带有barcode、引物及酶的凝胶珠 (GelBeads) 与处理后的细胞核混合，进行文库构建，从而实现大规模的单细胞级别的ATAC-Seq研究技术。可用于检测转录起始位点 (TSS) 和调控区域中富含信号的开放染色质，研究和比较感兴趣区域中的开放染色质，基于开放染色质区域和富集转录因子 (TF) 基序聚类和鉴定细胞等。

▶▶ 产品优势



低污染

线粒体DNA的reads数<2%



高通量, 高捕获效率

每个微流孔通道一次性可捕获500-10,000个单细胞核; 平台每个细胞核的捕获效率达到了65%



周期短

Tn5转座酶切割染色质的开放区域, 同时在DNA片段的末端添加测序引物, 省时省力



多数据量

获得染色质所有开放染色质区域, 不局限于某个转录因子位点或蛋白保护区域



低成本

相较于其他单细胞核捕获平台或传统酶切方法, 价格大幅降低

▶▶ 产品应用



▶▶ 技术流程



图1 10X Genomics平台单细胞ATAC-Seq技术流程图

► 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)	标准分析 1.测序结果统计 2.数据质控统计 3.比对结果统计 4.peak统计及鉴定 5.细胞类型鉴定及差异分析 ①样本间差异peak分析 ②样本间相同cluster 差异peak分析 6.功能分析 ①peak关联基因活性分析 ②顺式调控相互作用分析 ③差异peak关联基因GO功能分析 ④差异peak关联基因Pathway功能分析 ⑤差异peak关联基因蛋白互作分析 ⑥Motif鉴定和预测分析
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。 个性化分析 细胞轨迹分析(仅选取时间点样本可选做)

► 技术参数

样本要求:

细胞类型:

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

细胞活性: >70%

细胞数目: $\geq 1 \times 10^6$ cells/sample

测序平台: MGISEQ-2000

测序读长: PE50

数据量推荐: 1条lane

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

► 案例分析

案例一: 单细胞测序技术绘制哺乳动物染色质可接近性图谱

在此项研究中, 来自美国华盛顿大学的研究人员利用单细胞ATAC-Seq的测定方法, 分析了来自13种成年雄性小鼠组织的近10万个细胞。这些组织分别是骨髓、大肠、心脏、肾脏、肝脏、肺部、小肠、脾脏、睾丸、胸腺、全脑、小脑和大脑前额皮层。此次研究一共鉴定到85种不同的染色质可接近性模式, 并且能够将大部分染色质可接近性模式聚类, 鉴定相应的细胞类型, 还对40万个潜在的调节元件进行统计, 为深入了解哺乳动物细胞类型染色质调控环境及人类全基因组关联研究提供了一定的资源。

★ 主要结果

1. 17个8周龄雄性C57BL/6J小鼠的13个组织样本，进行单细胞ATAC测序，获得约100,000个细胞的染色质可接近性数据，共鉴定出436,206个可接近性位点。

2. 将通过质控的81,173个细胞的ATAC-Seq数据进行聚类分析，共鉴定出30个主要细胞亚群，其中大部分来源于同一组织的细胞都聚为同一亚群（比如91%肝脏来源细胞都在cluster 3）。基于30个主要细胞亚群的异质性，进一步细化分析，共鉴定出85种不同的染色体可接近性模式。

3. 将单细胞ATAC及转录组数据关联分析，发现大多数情况下两种方法分群注释数据表现出较高的相关性。利用基于KNN的分类方法将单细胞转录组数据中常见标签移植到单细胞ATAC数据中，发现两者细胞分群结果基本一致。

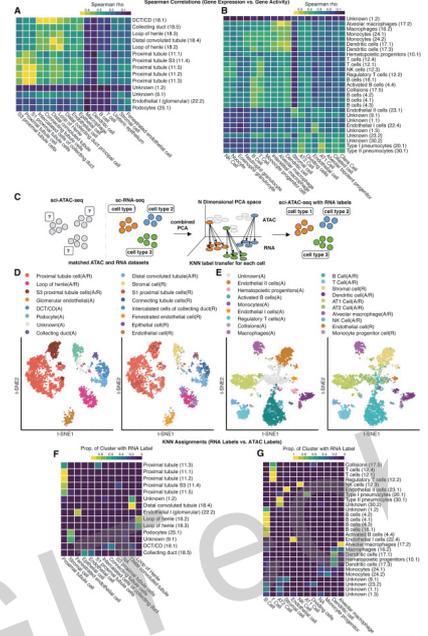


图2 单细胞ATAC及转录组数据联合分析结果展示

★ 参考文献

Cusanovich D, Hill A, Aghamirzaie D, et al. A single-cell atlas of in vivo mammalian chromatin accessibility[J]. Cell, 2018, 174(5): 1309-1324. e18.

案例二：单细胞染色质分析乳腺发育揭示细胞状态的转录调控和谱系关系

该研究首次在染色质可接近性方面，观察到胚胎乳腺原基在胚胎发育晚期（E18）已经开始获得成年鼠分化乳腺细胞特异的染色质可接近性特征。通过转录因子结合位点在染色质可接近区域的富集程度，推测了转录因子在不同细胞的活跃性。利用机器学习的方法预测了细胞谱系特异性的转录因子，其中包括多个已经实验验证的乳腺细胞转录因子。根据推测的每个细胞中转录因子的活跃性所构建的拟时间发育轨迹，高度吻合单细胞转录组数据。

★ 主要结果

1. 通过ATAC-Seq，一共鉴定到7,846个细胞核，胚胎晚期有2,577个，成年鼠5,269个；相比启动子-近端区域，使用启动子-远端区域也能更好地将E18细胞分组为亚群（图C）。

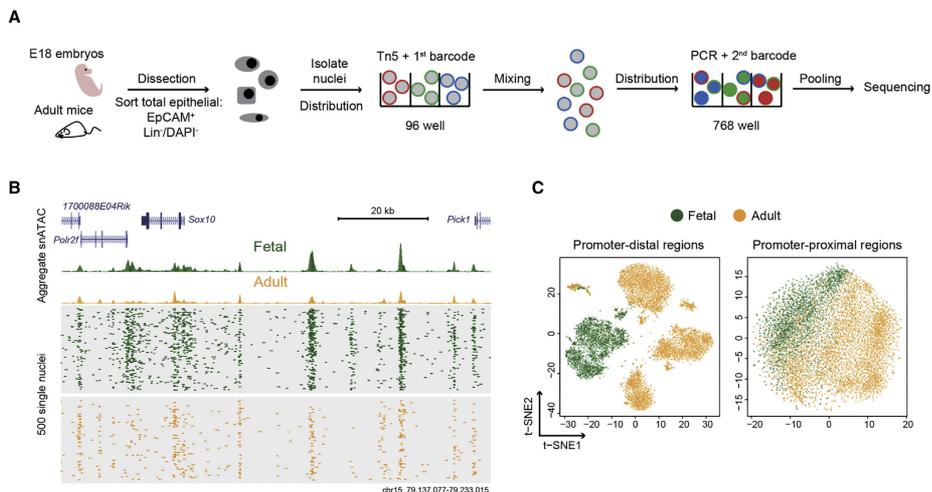


图3 A. 实验策略概述；

B. 聚集乳腺细胞的ATAC-seq剖面(上)和单核ATAC-seq剖面(下)，从500个随机选择的细胞中读取的数据被绘制成单核剖面；

C. 来自远端区域(>±1.5 kb转录起始位点[TSS])或启动子区域(<±1.5 kb TSS)的ATAC谱的t-SNE表达，所示的图是由两个生物复制组合而成的。

2.发现许多增强子与Sox10启动子具有很强的推定共可达性，并已经在斑马鱼等功能研究上证实它们是Sox10表达的主要驱动因素。除了Sox10位点外，还在重要的乳腺细胞状态指示基因中鉴定了共可及性位点。例如，发现了一个大型增强子簇，该簇富含与Krt8和Krt18基因连接的H3K27ac激活标记。

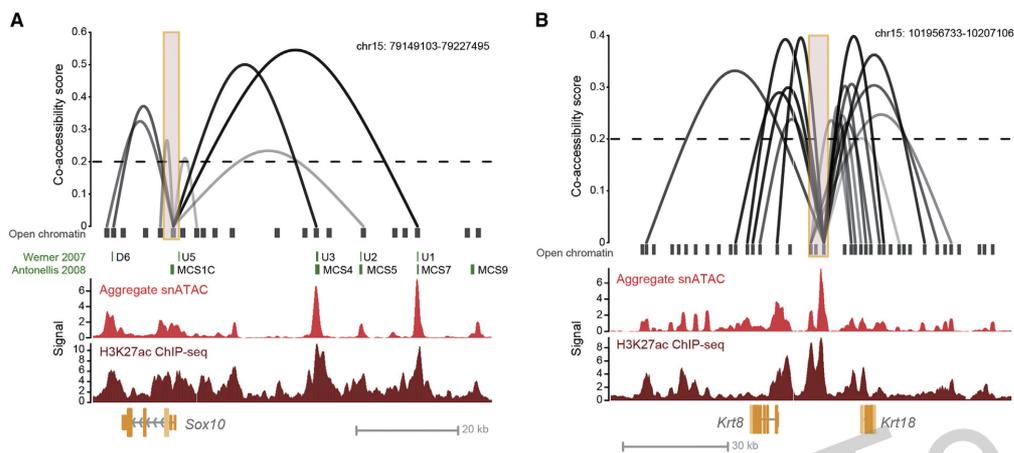


图4 peaks顺式调控关系图
cicero预测了Sox10 (A)、Krt8和Krt18 (B)位点的顺式调控相互作用的连锁图
(橙色框表示Sox10启动子区域或假定增强子区域为Krt8和Krt18)。之前描述的Sox10增强子用绿色表示。

★ 参考文献

Chung C, Ma Z, Dravis C, et al. Single-cell chromatin analysis of mammary gland development reveals cell-state transcriptional regulators and lineage relationships[J]. Cell Reports, 2019, 29(2):495-510. e6.

▶▶ 常见问题

Q: Tn5酶对细胞核进行预处理后，再形成微滴，是否会出现交叉污染？

A: 不会，因为每个细胞核有核膜包裹，相当于独立个体，等同于单细胞处理方式，形成油包水小液滴。

Q: 为什么要提前进行细胞核提取？

A: 因为Tn5酶不能进入细胞膜，所以需要单独提取细胞核，这样Tn5酶可以通过细胞核的核膜对染色质开放区域进行切割，从而进行后续实验。

▶▶ 典型案例汇总

物种	发表时间	发表期刊	文献名称
人、小鼠	2019	Nature Biotechnology	Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion
人、小鼠	2019	Nature Biotechnology	Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia
小鼠	2019	Cell	Rapid generation of somatic mouse mosaics with locus-specific, stably integrated transgenic elements
人	2018	Nature Biotechnology	Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion
小鼠	2018	Cell	A single-cell atlas of in vivo mammalian chromatin accessibility



10X Genomics 单细胞免疫组库

▶▶ 产品概述

10X Genomics单细胞免疫组库技术，即基于免疫组库技术基础上，进一步实现单个细胞水平上的免疫组库研究。基于10X Genomics平台，一次性能够分离500~10,000个单细胞，并同时获取5' 基因表达的数据以及TCR/BCR的V(D)J全长序列，是一种高效的同时检测表达和免疫组库信息的技术，在肿瘤微环境、感染性疾病、器官移植后排斥、免疫治疗等研究领域中有广泛的应用。

▶▶ 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，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）

▶▶ 产品应用

感染类疾病

病毒感染biomarker
感染过程中免疫动态变化
药物作用机制

抗体评估

抗体库测序
群体疫苗反应评估

移植

受体免疫重构研究
排异反应研究

免疫病

诊断biomarker
免疫病病程研究
药物作用机制

肿瘤

肿瘤免疫治疗
肿瘤异质性研究
肿瘤病程研究
诊断和预后biomarker

► 技术流程

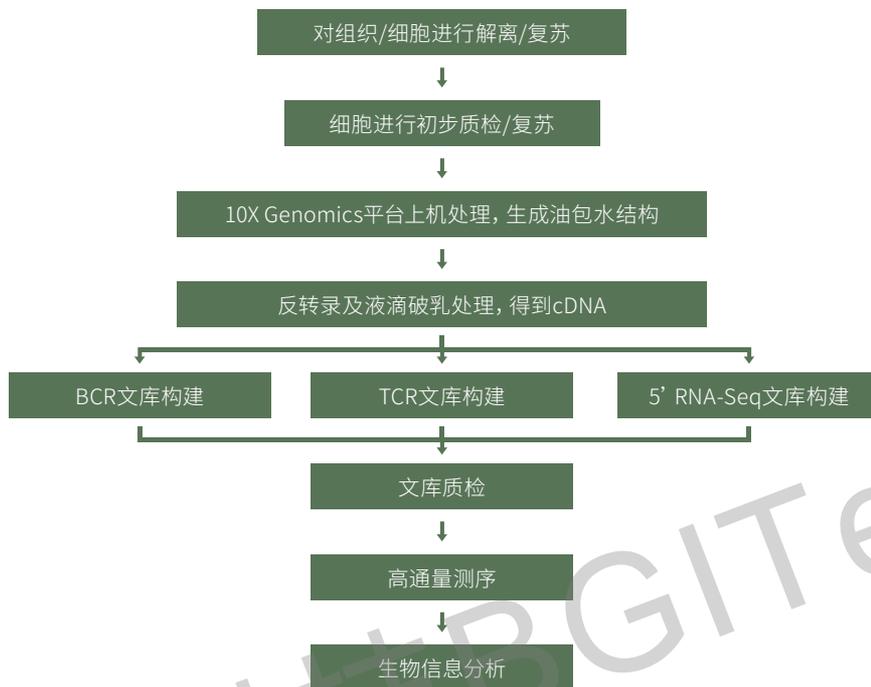


图1 10X免疫组库工作流程图

► 信息分析内容

表1 5' RACE mRNA-Seq文库信息分析内容列表

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)	1.测序结果统计 2.数据质控统计 3.比对结果统计 4.基因表达定量分析
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。

表2 TCR/BCR文库信息分析内容列表

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)	1.数据处理与质控分析 2.免疫组库细胞克隆型分析
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

样品要求:

基于10X Genomics平台的高通量单细胞RNA-Seq产品,接收的样本类型如下:

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 μ m; 细胞活性>80%; 细胞浓度>5 \times 10 ⁵ /ml
	细胞悬液背景干净, 无大量结团、碎片及杂质, 不含Ca ²⁺ 和Mg ²⁺

测序策略:

MGISEQ-2000平台

5' RACE mRNA-Seq文库: 建议PE100/PE150;

TCR/BCR文库: 建议PE150;

推荐数据量:

5' RACE mRNA-Seq文库:

细胞系样本, 建议每个细胞平均测: 50,000条Raw Reads;

组织样本, 建议每个细胞平均测: 500,000条Raw Reads;

PBMC样本, 建议每个细胞平均测: 50,000条Raw Reads。

TCR/BCR文库:

建议每个细胞至少测5,000条Raw Reads起。

项目执行周期:

从样品复苏后, 镜检符合上机要求、预付款到位开始 (不包括由于样品等问题停滞时间以及试剂订购周期), 细胞类型及生物信息分析整体完成周期视不同的类型而有所变化。

样品数 \leq 8个, 48个自然日;

样品数>8个, 请单独咨询周期。

售后服务:

数据交付3个月内, 提供免费的项目咨询服务;

最终数据提供后, 可继续保留数据1个月。

►► 案例分析

案例: 利用高通量单细胞免疫组库测序方法, 揭示了登革热四价疫苗诱导的T细胞免疫机制

样本/细胞类型: 全血分离出来的PBMC, 经过处理以及流式筛选后, 分离出登革热病毒激活的记忆性CD8⁺T细胞。
杂志: Nature Communications

登革热病毒有四种不同的血清型, 每年在世界各地感染多达2.8亿至5亿人, TAK-003是一种登革热四价减毒活疫苗。尽管很多感染登革热的病人很快从登革热感染中康复, 但据报道仍有约50万人罹患严重的登革热, 死亡率约为2.5%。如果病人感染一种血清型登革热, 然后再感染另一种不同的登革热血清型, 则发生严重疾病的可能性也会显著增加。这使登革热疫苗的研制更加复杂化, 该疫苗需要诱导对所有四种血清型的保护性反应, 并确保疫苗不会导致严重的登革热。T细胞介导的特异性细胞免疫反应, 是防治登革热和其他病毒性疾病的强有力且非常重要自身免疫反应; 尽管疫苗可以诱导和扩增T细胞(机体适应性免疫系统的关键部分), 但在决定T细胞反应的大小、多样性和持久性方面, 目前的大部分的科学研究对其中的细节还不清楚。

本项研究结合了单细胞免疫组库分析与标准的免疫监测手段, 发现了一种标记物, 转铁蛋白受体1(TfR1)可以鉴定CD8⁺细胞的分化潜能, 这表明T细胞免疫依赖于特定的代谢物。基于发现的这系列转录特征, 可定义一组标志物以判断最持久的疫苗反应性CD8⁺T细胞, 以评估疫苗诱导的细胞免疫的程度、多样性和持久性等。

★ 研究展示

1. 对接种登革热疫苗14天后的临床样本的外周血单核细胞进行单细胞测序，发现了TCR克隆型存在高度多样性，以及接种疫苗获得的免疫记忆性CD8⁺T细胞亚群。

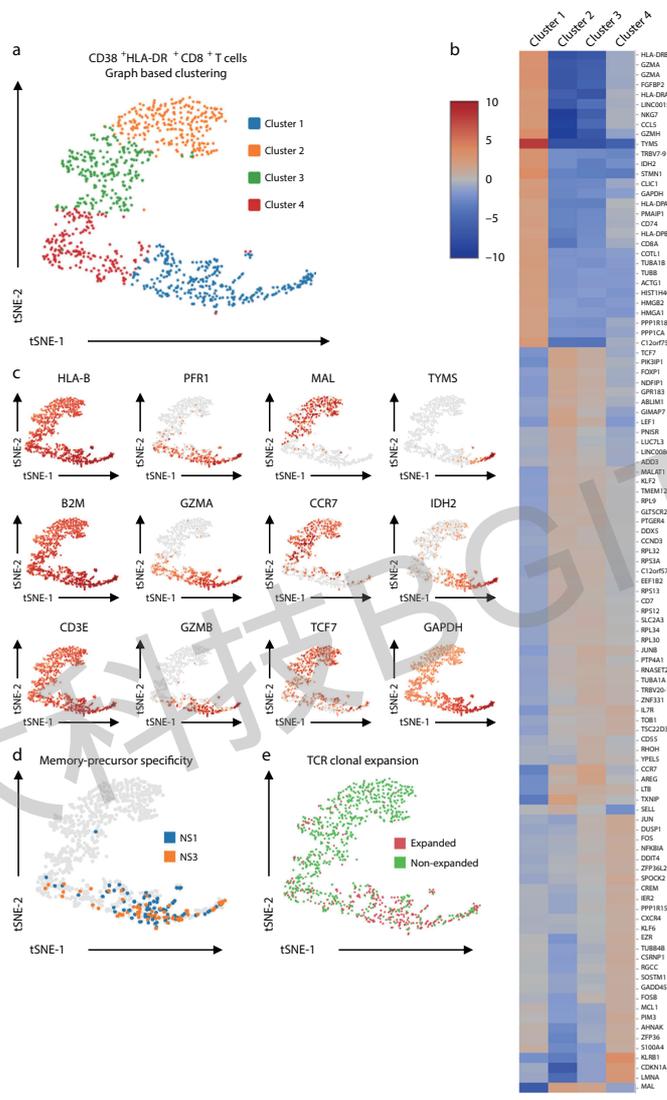


图2 TAK-003疫苗诱导的CD38⁺HLA-DR⁺ CD8⁺T细胞是高度多克隆型和持续有长期活性的记忆细胞

2. 通过对临床接种了四联登革热疫苗第14天后的样本进行处理，分离其外周血得到CD8⁺T细胞后，再进行单细胞RNA-Seq表达量测序分析，发现了TfR1⁺HLA-DR⁺或CD38⁺HLA-DR⁺CD8⁺的T细胞亚群，以及NS1或NS3激发的免疫CD8⁺T细胞亚群。同时，对同一样本进行流式细胞分选后，也得到了相应一致的细胞亚群。

★ 参考文献

Waickman A, Victor K, Li T, et al. Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling. *Nature Communications*, 2019, 10:3666.

▶▶ 典型案例汇总

发表时间	发表期刊	文献名称	信息分析条款
2019	Nature Communications	Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and cellular metabolic profiling	登革热疫苗引起的T细胞免疫机制

华大科技BGITech

产品
简介

Product
Profile

| 基因科技 造福人类
Omics for All

华大科技BGITec

常规分子生物学

Sanger测序	167
基因分型	169
甲基化验证	173
基因表达	176

Sanger测序

产品概述

Sanger测序是基于双脱氧链末端终止法的原理下，通过毛细管电泳技术，获得高准确度DNA序列数据，是DNA测序的金标准。我们可以为客户提供质粒提取、PCR扩增、PCR产物纯化、测序引物设计、特殊结构优化测序、DNA序列的基本数据分析，以及长片段walking测序后的序列拼接等生物信息学分析服务。

产品优势



通量高，满足批量测序需求
20+ABI3730测序仪



采用滚环复制技术，菌样交付速度提升一倍
200+专业技术团队



八大测序基地，满足本地化交付需求
北京、上海、广州、武汉、重庆、青岛、杭州、香港



下单便捷
自助系统可7*24小时在线下单；进入“六合华大”微信公众号可预约取样体验上门服务



IVD个性化交付
公司质量体系、PCR扩增等资质齐全；实验记录齐全，结果真实可靠；实验结果保密，可满足存储需求；有成熟解决方案，协助多家知名IVD企业顺利申报；可提供打包解决方案等个性化服务

服务流程

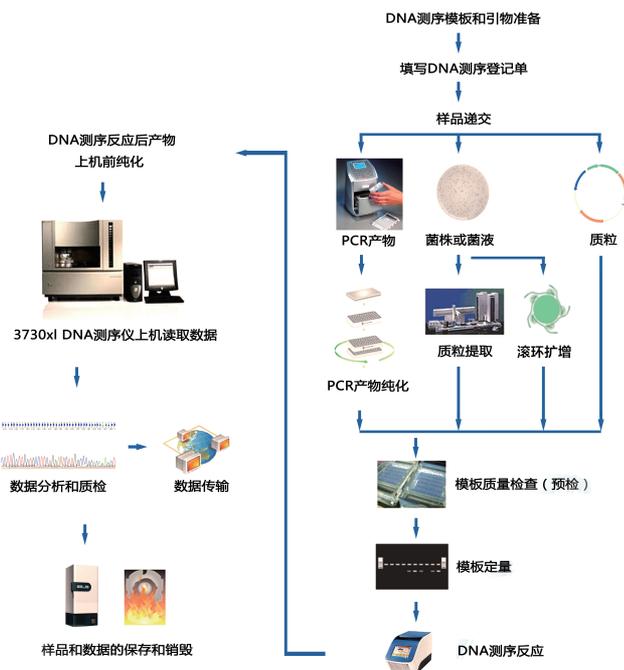


图1 Sanger测序服务流程图

技术参数

样本要求

样品类型	收样标准	样品类型	收样标准
菌液	体积 $\geq 50\mu\text{l}$	质粒	浓度 $\geq 100\text{ng}/\mu\text{l}$, 体积 $\geq 20\mu\text{l}$
PCR产物	浓度 $\geq 20\text{ng}/\mu\text{l}$, 体积 $\geq 20\mu\text{l}$	自带引物	浓度 $\geq 3.2\text{pmol}/\mu\text{l}$, 体积 $\geq 10\mu\text{l}$

*仅限于10kb以内常规基因序列

案例分析

案例一：菌液RCA（滚环扩增法）快速交付

滚环式复制是噬菌体中常见的DNA复制方式，采用多对随机引物对环状DNA进行扩增，具有操作简单效率高等特点。

★ 主要优势

次晨交付、50 μl 微量菌液送样。

大肠杆菌需经过夜培养、质粒提取后测序，时间成本较高。华大基因采用自主知识产权的滚环扩增技术，无需摇提就可快速富集高保真DNA，次日上午即可交付，既省时又可获得高质量测序结果。

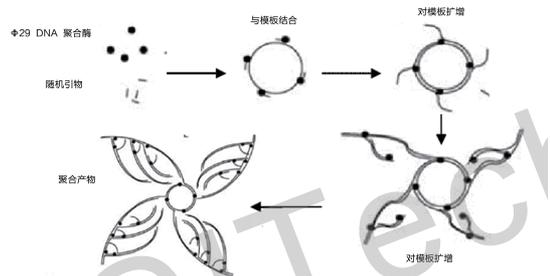
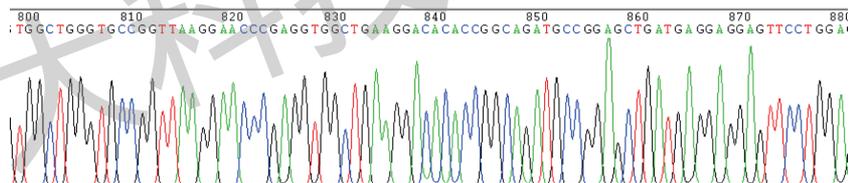


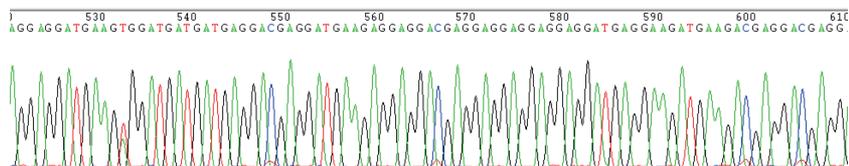
图2 滚环扩增法流程示意图



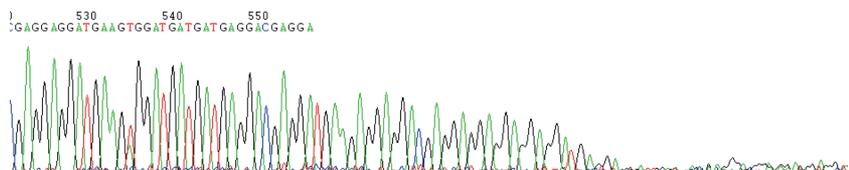
案例二：复杂结构、高GC含量样品

由于DNA自身GC含量高或者存在复杂结构等因素，常规测序方案中容易出现测序信号提前中断或者衰减过快现象，直接影响测序效果。华大基因拥有多套自主研发针对复杂结构、GC含量高的解决方案，客户认可度高。

★ 复杂结构-优化测序方案



★ 复杂结构-常规测序方案



案例三：工业客户服务

华大基因具备丰富的经验和先进的设备，并拥有专业的实验和数据分析人员，可以满足工业客户对于批量样品快速、高质量交付的需求。例如：单次送样7,000个反应，可实现24h极限交付。

基因分型

●● 基因芯片 ●●

▶▶ 产品概述

芯片基因分型主要包括Illumina SNP芯片分型平台和Affymetrix基因分型平台,可以在短时间内对大量的SNP进行分型,适用于大样本不同标记密度 (10^3 - 10^6) 的基因分型。实现了超大样本的GWAS检测,与其它技术NGS (外显子组、全基因组) 的搭配,可以更加有效的发现显著效应的基因变异。在全基因组层面上,为开展多中心、大样本、反复验证的基因与疾病的关联研究,揭示疾病发生、发展与治疗相关的遗传基因提供了技术支持。

▶▶ 产品优势



单次反应可以检测 10^3 - 10^6 位点



超高通量,每天可完成上千样品的验证



灵活多样,针对用户关注的位点灵活选择设计芯片



价格低,适用于已知高通量位点验证



准确性好,孟德尔遗传错误率小于0.1%,重复性99.9%以上

▶▶ 技术参数

样本要求:

- 1.DNA样本总量 $\geq 1\mu\text{g}$,浓度 $\geq 50\text{ng}/\mu\text{L}$,体积 $\geq 20\mu\text{L}$ 且 $\text{OD}_{260}/280=1.7\sim 2.0$, $\text{OD}_{260}/230\geq 1.4$,琼脂糖电泳条带 $>10\text{Kb}$ 且条带明亮单一,无RNA无蛋白污染
- 2.全血(推荐EDTA抗凝,避免使用肝素抗凝)体积 $\geq 1\text{mL}$
- 3.组织总重量 $\geq 100\text{mg}$ 且单个组织块重 $<30\text{mg}$,液氮或 -80°C 保存
- 4.细胞样本需细胞个数在 $1\times 10^6\sim 1\times 10^7$ 之间

结果交付:

- 1.完整的DNA质检报告
- 2.实验数据及报告
- 3.根据需求,可提供个性化的数据分析

项目周期:

样本 $<1,000$ 份质检合格后25-30个工作日
样本 $\geq 1,000$ 份需协商

●● 质谱分型 ●●

▶▶ 产品概述

MassARRAY系统提供的基因分型属于中等通量的SNP基因分型技术。该技术结合了引物延伸、基质辅助激光解吸电离飞行时间质谱的优点,能够快速、经济、高重现性地进行SNP分型验证。实验设计非常灵活(多重,一个well可以检测25~30个位点),分型结果准确性高,在对数十个位点进行数百至数千份样本检测时,性价比较高。适用于对全基因组范围内GWAS研究发现的位点以及其它可能存在的SNP位点进行验证。

▶ 产品优势



一个反应可检测25-30个SNP位点



超高通量，每天可检测10万个位点



针对20-200位点性价比较高



准确度大于99%，有效分型率90%以上

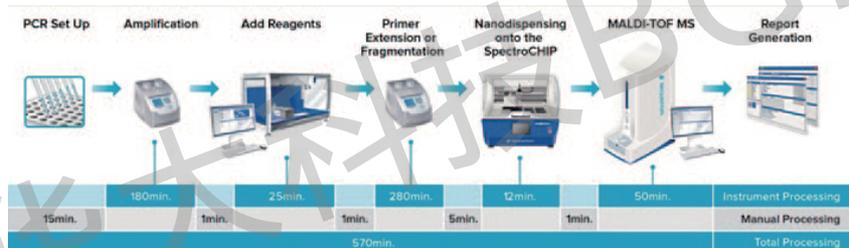
▶ 产品应用

肿瘤研究

遗传分析

分子诊断

▶ 技术流程



▶ 技术参数

样本要求:

1. DNA浓度 $\geq 25\text{ng}/\mu\text{L}$ ，体积 $\geq 10\mu\text{L}$ 且 $\text{OD}_{260}/\text{OD}_{280}=1.7\sim 2.0$ ，琼脂糖电泳条带 $>10\text{Kb}$ 且条带明亮单一，无RNA无蛋白污染

2. 全血（推荐EDTA抗凝，避免使用肝素抗凝）体积 $\geq 0.5\text{mL}$

结果交付:

1. 完整的DNA质检报告

2. 实验报告

3. 根据需求，可提供数据分析及报告

项目周期:

样品 $<1,000$ 份质检合格后25个工作日

样品 $\geq 1,000$ 份或需样品提取双方可协商

●● Taqman探针分型 ●●

▶ 产品概述

Taqman探针法是适用于少量位点进行大量样本验证。该方法基于ABI7900系统，采用384模块或96模块，可灵活、快速提供单个SNP位点分型服务，也可作为MassARRAY技术无法检测或检测效果不佳位点的检测技术补充。对于高度复杂的基因区域中SNP位点检测，分型数据更准确，比如药物基因检测中代谢酶的相关SNP位点等。因此，该方法在医学及健康检测服务领域中有广泛应用。

▶▶ 产品优势



准确度大于99%



设计灵活, 周期短



适用于上千样品, 每个样品1-5个位点

▶▶ 技术参数

样本要求:

1. DNA浓度 $\geq 25\text{ng}/\mu\text{L}$, 体积 $\geq 10\mu\text{L}$; OD260/280=1.7~2.0, 琼脂糖电泳条带 $>10\text{Kb}$ 且条带明亮单一, 无RNA无蛋白污染
2. 全血 (推荐EDTA抗凝, 避免使用肝素抗凝) 体积 $\geq 0.5\text{mL}$

结果交付:

1. 完整的DNA质检报告
2. 实验报告

项目周期:

- 样品 $<1,000$ 份质检合格后20个工作日
样品 $\geq 1,000$ 份或需样品提取双方可协商

●● SNaPshot分型 ●●

▶▶ 产品概述

SNaPshot技术是ABI公司推出了专为检测SNP设计的, 可对多个SNP位点同时进行基因分型。该方法在SNaPshot反应后, 产物通过电泳分离、五色荧光检测、Gene mapper分析, 可一次电泳检测多个SNP位点, 这是建立在3730, 3130等测序仪上的技术。

▶▶ 产品优势



一个反应可检测5-10个位点, 性价比较高



准确度99%以上



实验设计灵活



通量高

▶▶ 技术参数

样本要求:

1. DNA浓度 $\geq 25\text{ng}/\mu\text{L}$, 体积 $\geq 25\mu\text{L}$ (检测位点为1-5个, 如位点数增加, 则样品量需相应增加) 且 OD260/280=1.8~2.0, 溶液中不含PCR反应抑制剂
2. 全血 (推荐EDTA抗凝, 避免使用肝素抗凝) 体积 $\geq 1\text{mL}$

结果交付:

1. 原始检测结果
2. 位点分型结果
3. 多重检测结果峰图

项目周期:

一般15个工作日左右, 如样品数量较大, 需另行协商

PCR重测序

产品概述

PCR重测序是通过已知序列设计引物对待测样本进行扩增测序，此技术多用于同物种间不同个体的差异检测，如突变、缺失、插入等核酸水平上变化与表型的关联性研究。

产品优势



实验设计灵活



结果准确度高



周期短



适用于少量样品多位点

产品应用



技术参数

样本要求：

1. DNA浓度 $\geq 25\text{ng}/\mu\text{L}$ ，体积 $\geq 25\mu\text{L}$ （检测位点为1-5个，如位点数增加，则样品量需相应增加）；
OD_{260/280}=1.8~2.0，溶液中不含PCR反应抑制剂
2. 全血（推荐EDTA抗凝剂，避免使用肝素抗凝）体积 $\geq 0.5\text{mL}$
3. 组织总重量 $\geq 30\text{mg}$ 且单个组织块重 $< 30\text{mg}$ ，液氮或 -80°C 保存
4. 菌类样品培养好后寄送 $\geq 1.5\text{mL}$ ，避免再次转接

结果交付：

1. 测序数据质量达到Q20
2. 测序峰图、初步分析及结题报告

项目周期：

根据样本量及需检测的位点数双方协商

甲基化验证

●● 甲基化芯片 ●●

▶▶ 产品概述

采用Illumina Infinium MethylationEPIC BeadChip芯片 (850K芯片) 提供的人源全基因组DNA甲基化检测服务。850K芯片可检测人全基因组约85万个CpG位点的甲基化状态, 其中包含了原450K芯片91%的位点, 并增加了413,745个位点。850K芯片对基因组的功能元件提供了全面覆盖: CpG岛、RefSeq基因、ENCODE开放染色质、ENCODE转录因子结合区以及FANTOM5增强子。

▶▶ 产品优势



重复性好, >98%



流程简单, 无PCR过程



可用于FFPE样品



性价比高

▶▶ 产品应用

干细胞
研究

复杂疾
病研究

肿瘤

▶▶ 技术参数

样本要求:

1. DNA样本总量 $\geq 2\mu\text{g}$, 浓度 $\geq 50\text{ng}/\mu\text{L}$, 体积 $\geq 30\mu\text{L}$; OD260/280=1.7~2.0, 琼脂糖电泳条带 $>10\text{Kb}$ 且条带明亮单一, 无RNA无蛋白污染
2. 全血 (推荐EDTA抗凝, 避免使用肝素抗凝) 体积 $\geq 1\text{mL}$
3. 组织总重量 $\geq 100\text{mg}$ 且单个组织块重 $< 30\text{mg}$, 液氮或 -80°C 保存
4. 细胞样本需细胞个数在 $1 \times 10^6 \sim 1 \times 10^7$ 之间

结果交付:

1. 完整的DNA质检报告
2. 实验报告
3. 如进行后期验证, 可提供芯片数据整理及筛选

项目周期:

样本 < 100 份质检合格后20-25个工作日可完成检测
样本 ≥ 100 份需订货芯片双方协商实验周期

●● 质谱甲基化验证 ●●

▶▶ 产品概述

亚硫酸盐能够将 DNA 中未发生甲基化的 C 转化经扩增后变为 T，而甲基化的 C 不会变化，由于碱基分子之间的质量差异，利用 MassARRAY 系统能够快速地对 CpG 位点的甲基化进行定量。该方法单个反应可检测 300-500bp，适用于全基因组甲基化测序、EWAS 芯片筛选得到的靶向易感区域，以及候选基因筛选的重要调控区域的验证。

▶▶ 产品优势



灵敏度高，可检测 5% 的甲基化修饰



重复性好，CV ≤ 5%



适用于长片段检测

▶▶ 技术参数

样本要求：

1. DNA 样本总量 ≥ 1μg (检测片段为 1~2 个)，浓度 ≥ 50ng/μL，体积 ≥ 20μL；OD_{260/280} = 1.7~2.0，OD_{260/230} ≥ 1.4 以上，琼脂糖电泳条带 > 10Kb 且条带明亮单一，无 RNA 无蛋白污染
2. 全血 (推荐 EDTA 抗凝，避免使用肝素抗凝) 体积 ≥ 0.5mL
3. 组织总重量 ≥ 100mg 且单个组织块重 < 30mg，液氮或 -80°C 保存
4. 细胞样本需细胞个数在 $1 \times 10^6 \sim 1 \times 10^7$ 之间

结果交付：

1. 信息评估，提供详细的方案报告
2. 完整的 DNA 质检报告
3. 实验报告

项目周期：

- 样本 < 100 份质检合格后 25 个工作日内完成
样本 ≥ 100 需根据样品数及检测方案数确定

●● 焦磷酸测序 ●●

▶▶ 产品概述

目前采用 QIAGEN 的 PyroMark Q96 ID 及 PyroMark Q48 Autoprep 平台，反应原理是四种酶催化同一反应体系中的酶级联反应，依赖于 DNA 聚合过程中释放焦磷酸 (PPi)，PPi 在含有四种酶荧光素酶的反应体系作用下发出相应的荧光，从而能检测出目的 CpG 位点甲基化程度。该方法单个反应检测 45-100bp，适合位点及明确短区域的精确验证。

▶▶ 产品优势



准确度高，检测单碱基甲基化水平



可检测测序引物区域甲基化水平

▶▶ 技术流程



图 焦磷酸测序技术流程图

▶▶ 技术参数

样本要求:

1. DNA样本总量 $\geq 2 \mu\text{g}$ (检测1~5个片段), 体积 $\leq 30 \mu\text{L}$ 且 $\text{OD}_{260}/\text{OD}_{280} = 1.7\sim 2.0$, $\text{OD}_{260}/\text{OD}_{230} \geq 1.4$, 琼脂糖电泳条带 $> 10\text{Kb}$ 且条带明亮单一, 无RNA无蛋白污染
2. 全血 (推荐EDTA抗凝, 避免使用肝素抗凝) 体积 $\geq 1\text{mL}$
3. 组织总重量 $\geq 100\text{mg}$ 且单个组织块重 $< 30\text{mg}$ 方便称量提取, 液氮或 -80°C 保存
4. 细胞样本需细胞个数在 $1 \times 10^6 \sim 1 \times 10^7$ 之间

结果交付:

1. 信息评估, 提供引物设计报告
2. 实验报告

项目周期:

- 样本 < 100 份质检合格后25个工作日内完成
样本 ≥ 100 周期需根据样品数及检测反应数确定

基因表达

●● 荧光定量PCR ●●

▶ 产品概述

荧光定量PCR技术与常规PCR相比，具有特异性更强、更能有效的解决PCR污染问题、自动化程度高等特点，目前已成为国际公认的核酸分子定量的标准方法。采用ABI ViiA7平台384模块进行高通量测序后期样品验证，提供染料法及探针法两种检测方法，可以完成多种类型RNA (mRNA, miRNA, lncRNA, circRNA)、DNA的定量检测。

▶ 产品优势



准确性高、实验周期短



实验设计灵活、自动化程度更高

▶ 技术流程



图 荧光定量PCR技术流程图

▶ 技术参数

样品要求

提供目的基因的相关资料：基因名称、序列等信息

提供实验所需的足够量的实验样本：

- 1.Total RNA浓度 $\geq 100\text{ng}/\mu\text{l}$ ，体积 $\geq 15\mu\text{l}$ ；电泳检测28S rRNA、18S rRNA紫外灯下清晰可见，同时可见一条由tRNA、5.8S rRNA、5S rRNA组成迁移较快的条带，点样孔附近无条带（无DNA残留）
- 2.细胞数 $\geq 5 \times 10^6$ ，加入 Trizol reagent将细胞裂解后 -80°C 保存
- 3.新鲜组织 $\geq 100\text{mg}$ ，液氮或 -80°C 保存
- 4.血清/血浆体积 $\geq 600\mu\text{l}$ ，采血管采2-3 mL血液后颠倒混匀，以离心方式进行分离，分离后的血清/血浆 $-20^\circ\text{C}/-80^\circ\text{C}$ 保存

结果交付

- 1.完整的RNA质检报告及引物序列信息
- 2.实验报告，包括原始导出数据、完整的实验步骤报告、所有样本基因扩增曲线及溶解曲线

项目周期

一般时间周期为15-20工作日，具体时间根据样品数和指标数不同可另行协商



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

华大基因 BGITech

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

华大基因·澳洲 (布里斯班)

地址: L6, CBCRC, 300 Herston Road, Herston, Brisbane, Queensland, Australia
电话: +61 (07) 3362 0475

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: 2020年5月版