深圳华大基因股份有限公司

电话：400-706-6615

网址：www.bgitechsolutions.com

邮箱：info@genomics.cn

地址：深圳市盐田区洪安三街21号（518083）

华大科技公众号　BGI技术支持-质谱类

华大科技质谱

产 | 品 | 手 | 册

华大基因 BGI

目录 Contents

01 华大科技质谱平台

简介	02
专业资质	02
仪器配置	03
生物信息分析平台	05
领军人物	06

02 蛋白质组学产品

定性蛋白质组学 09

蛋白质全谱分析	09
蛋白质胶点和胶条鉴定	10
典型结果展示	11
应用案例解读	12

定量蛋白质组学 16

蛋白标记定量分析–iTRAQ/IBT/TMT	16
蛋白DIA定量分析	18
目标蛋白MRM定量分析	20
蛋白组定量技术对比	21
典型结果展示	22
应用案例解读	27

修饰蛋白质组学 37

磷酸化修饰	38
乙酰化修饰	41
N–糖基化修饰	43
典型结果展示	46
应用案例解读	49

多肽组学 55

应用方向	55
技术路线	56
产品优势	56
结果示例	57

03 代谢组学产品

非靶向代谢组学 61

产品介绍	61
典型结果展示	62
应用案例解读	64

脂质组学 66

产品介绍	66
典型结果展示	68
应用案例解读	69

靶向代谢组学 72

产品介绍	72
典型结果展示	75
应用案例解读	76

04 多组学关联分析

蛋白质组与转录组关联分析 83

代谢组与宏基因组关联分析 92

05 送样要求

送样要求 104

BGI 华大基因

01

华大科技质谱平台

简介

专业资质

仪器配置

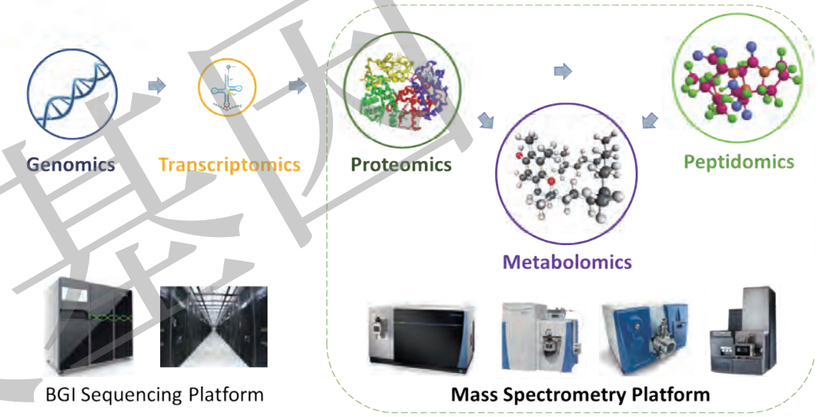
生物信息分析平台

领军人物

简 介

华大成立于1999年，是全球领先的生命科学前沿机构。华大秉承“基因科技造福人类”的使命，以“产学研”一体化的发展模式引领基因组学的创新发展，通过遍布全球100多个国家和地区的分支机构，与产业链各方建立广泛的合作。华大科技于2012年完成整合，致力于成为全球生命科学研究机构的首选合作伙伴，为其提供高质量、行业领先的新一代测序、高通量质谱检测、生物云计算等标准化的生物技术服务，也可依据用户的个性化需求提供定制产品服务。

华大科技质谱平台始建于2007年，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科技服务平台。华大科技质谱平台通过与华大的基因组学、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。



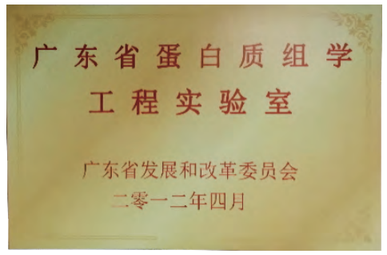
截止到2019年2月，华大科技质谱平台在蛋白质组学和代谢组学领域已发表85篇研究论文，总计影响因子达432。十余年风雨兼程，华大科技质谱平台始终聚焦于组学研究成果和技术的应用，开发了一系列的医学检验技术，如新生儿遗传代谢病检测、母乳营养检测等；始终积极从事蛋白质组学和代谢组学的科研服务，打造了一系列涵盖蛋白质、多肽、代谢小分子以及核酸的高质量质谱技术综合服务产品，并形成了与科研人员高度合作的销售、市场、客服的专业服务团队。

专业资质

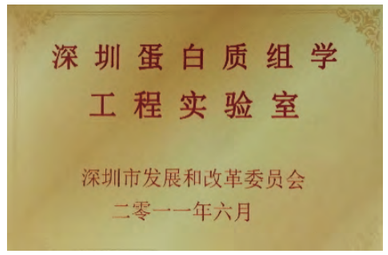
华大科技质谱平台从建立初期就重视质量的标准化管理，目前已获得ISO9001、ISO14001、OHSAS18001三标一体认证，通过了CNAS 17025质量管理体系及ISO27001信息安全质量管理体系认可，标准化管理能力处于行业领先水平。

此外，华大科技质谱平台已于2011和2012年被提升为“深圳市蛋白质组学工程实验室”和“广东省蛋白质组学工程实验室”。

BGI 华大基因



省级蛋白质组学工程实验室



市级蛋白质组学工程实验室

ISO17025实验室管理体系认

ISO9001质量保证体系认证

ISO14001环境管理体系认证

OHSAS18001职业健康安全管理体系认证

ISO27001信息安全质量管理体系认证

仪器配置

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

部分仪器信息列表如下：

仪器类型	仪器名称	厂家	应用方向
色谱仪	I-CLASS	Waters	与质谱联用的代谢小分子液相分离
	二维ACQUTTY UPLC		
	2777C		
	UPLC 20AD XR	SHIMADZU	
	Transcend II	Thermo	
	HPLC 20AD	SHIMADZU	肽段预分离
	ExionLC AD	SCIEX	蛋白、肽段预分离
	GC7890B	Agilent	与质谱联用的代谢小分子气相分离
	Nano Acquity	Waters	与质谱联用的肽段分离
	nanoLC-20AD	SHIMADZU	
	UltiMate 3000 RSLCnano	Thermo	
	Eksigent NanoLC 415	SCIEX	
	nanoLC Ultra 2D		
轨道阱质谱仪 (Orbitrap)	Orbitrap Fusion Lumos	Thermo	蛋白质组学、多肽组学、修饰蛋白质组学、互作蛋白质组学、代谢组学
	Q Exactive HF-X		
	Q Exactive HF		
	Q Exactive		
	LTQ Orbitrap velos ETD		
三重四极杆质谱仪 (Triple Quad)	Q TRAP 6500+	SCIEX	靶标蛋白定量，靶标小分子定量，如氨基酸、维生素、激素、新生儿遗传代谢物检测、胆汁酸、抗生素等
	Q TRAP 5500		
	Q TRAP 4500		
	Triple Quad 5500		
	TQS	Waters	
	TQD		
飞行时间质谱仪 (TOF)	Triple TOF 5600+	SCIEX	蛋白质组学、代谢组学
	Triple TOF 5600		
	Xevo-G2-XS	Waters	代谢组学
	MicroTyper MS	Skyray	微生物、核酸检测
	UltrafleXtreme	Bruker	蛋白质组学
其他类型质谱仪	ICP-MS	Agilent	微量元素与重金属检测
	5977A-GC/MSD		
	TSQ9000 GC-MS	Thermo	有机酸、代谢物、VOC检测

BGI 华大基因

生物信息分析平台

华大科技质谱平台拥有30多名具有多年项目与开发经验的专业信息技术人员组成的生物信息分析团队，致力于大规模蛋白质组学、代谢组学数据的处理、分析与深度挖掘，以及多组学数据的关联整合分析。依托华大基因的高性能计算平台，应用蛋白质组和代谢组专业的Mascot、MaxQuant、ProteinPilot、Proteome Discoverer、Spectronaut X™、Skyline、Compound Discoverer、Progenesis QI等分析软件，从高通量高精度质谱数据中通过与UniProt、Swiss-Prot、TrEMBL、NCBI、KEGG、HMDB、LIPID MAPS、mzCloud等数据库进行搜索匹配得到大量蛋白质的序列鉴定和定量信息，代谢物的结构鉴定和定量信息。再联合华大科技质谱平台自主开发的多种数据分析软件IQuant、IPeak、MetaX等，进一步统计分析得到蛋白质和代谢物功能注释，代谢通路以及差异蛋白质和代谢物等分析。另外，从上游基因组到转录组、蛋白质组和下游代谢组，各组学数据之间还可以进一步做关联整合分析，提供多角度更全面的分析结果。大型计算机资源、多套软件处理系统、自动化分析流程以及个性化分析方法为高通量质谱数据以及多组学数据整合分析提供了重要的分析平台和分析方向。

部分质谱数据分析软件及数据库展示如下：

软件分类	分析软件/平台	数据库
质谱数据查看软件	Xcalibur（Thermo）、PeakView（Sciex）、MassLynx（Waters）、FlexAnalysis（Bruker）、Analyst（Sciex）等	UniProt、NCBI、Ensembl、KEGG、RNA ref等
蛋白质组搜索引擎	Mascot、Sequest HT、Andromeda、X!tandem、MS-GF+等	
蛋白质组学分析软件	IQuant（标记定量）、IPeak（综合鉴定）、Proteome Discoverer（鉴定、标记/非标记定量）、Maxquant（DDA非标记定量）、Spectronaut（DIA非标记定量）、Proteinpilot（鉴定、标记定量）、Skyline（靶向定量）等	
代谢组学分析软件	Progenesis QI、metaX、Compound Discover等	KEGG、HMDB、LIPID、m/zCloud、自建标准品库
靶向定量分析软件	Skyline、MRMPilot、TraceFinder等	

华大基因自主开发的蛋白质组和代谢组数据分析软件（有软件著作权）包括：

❶ IQuant-蛋白标记定量软件

标记定量（iTRAQ/IBT/TMT）是目前蛋白质定量领域广泛采用的定量技术，质谱平台针对iTRAQ/IBT/TMT类试剂标记数据，开发了一个高通量精度高的定量软件IQuant。该软件基于Mascot搜索鉴定结果进行定量及统计分析，利用先进的机器学习算法对鉴定进行重打分并进行严格的质量控制，同时利用最新的统计算法对肽段的定量信息进行缺失值处理，归一化以及权重计算，然后利用非参数检验方法进行蛋白质表达差异统计检验分析，与常用类似软件相比，可获得更高质量的定量结果。

参考文献：Bo W, et al. IQuant: an automated pipeline for quantitative proteomics based upon isobaric tags. Proteomics. 2014, 14(20):2280-2285

❷ metaX代谢组学数据处理软件

metaX是基于R语言开发的代谢组学数据分析软件包，它能够通过一组可互换的模块进行端到端代谢组学数据分析。具体来说，metaX提供了多种功能，如峰值拾取和注释、数据质量评估、缺失值插补、数据归一化、单变量和多变量统计，功率分析和样本量估计，接收器操作特征分析，生物标记选择，路径注释，相关网络分析和代谢物鉴定。

参考文献：Bo W, Mei Z, ZengC. et al. metaX: a flexible and comprehensive software for processing metabolomics data. BMC bioinformatic. 2017, 18(1):183

❸ IPeak 蛋白质组综合鉴定软件

IPeak是一款综合了Percolator算法和几款主流蛋白质组鉴定搜索引擎的软件，既有图形用户界面，也有命令行使用界面。该软件在不降低准确度的前提下，提升了肽段鉴定的灵敏度，从而提高了蛋白质组鉴定效果。

参考文献：Wen B, Du C, Li G, et al. IPeak: An open source tool to combine results from multiple MS/MS search engines. Proteomics. 2015 Sep;15(17):2916-20.

领军人物

刘斯奇

华大基因董事会董事，华大基因蛋白质组首席科学家。兼任中国科学院北京基因组研究所研究员，中国科学院研究生院教授，中国农业大学兼职教授，路易斯维尔大学医学院客座教授。

刘博士于1999年参与北京华大基因研究中心和中科院北京基因所的创建工作，主要负责蛋白质组学平台的建设，质谱技术在蛋白质分析中的应用和提高生物信息学在蛋白质结构和功能上的应用。作为首席科学家主持了国家重大科学研究计划“蛋白质组海量质谱数据的解析及其在人类基因组注释中的作用”，国家自然科学基金重大研究计划“小鼠AOM结直肠癌模型的微进化研究”。近年来带领团队主持参与了10余项国家863、973计划项目课题，作为通讯作者发表研究论文60余篇。

王融

深圳华大基因质谱首席技术官，美国纽约西奈山伊坎医学院遗传学和基因组学兼职教授，生物科学研究员（广东省自然科学研究研究员/副研究员/高级实验师资格评审委员会认证）。

自1982年留校建设北京医科大学中心实验室开始，王融博士一直从事质谱和蛋白质分析的技术和方法的开发及应用。1994年获美国氰胺物理与分析研究奖，1994年和2005年两次获得美国阿尔茨海默病协会研究奖。自1997年任美国健康卫生院（NIH）、美国阿尔茨海默病协会、香港研究资助委员会（RGC）和中国科技部（973项目）的专家评审，发表SCI检索论文140余篇，包括*Science, Nature, Nature Medicine, Nature Neuroscience, Nature Genetics, Neuron, JBC, PNAS*等，被引用18,000余次（h-index 64），拥有已授专利4项，均已产业化。王融博士自2016年12月辞去美国西奈山医学院终身教职回国以来，着力于大健康质谱和跨组学的研究和转化工作，已领导和开展了西藏地区小学生营养代谢研究，西藏包虫病跨组学研究，和百万人基因组与代谢组学研究等多个惠及民生健康的项目。

BGI 华大基因

02 蛋白质组学产品

定性蛋白质组学

蛋白质全谱分析

蛋白质胶点和胶条鉴定

典型结果展示

应用案例解读

定量蛋白质组学

蛋白标记定量分析-iTRAQ/IBT/TMT

蛋白DIA定量分析

目标蛋白MRM定量分析

蛋白组定量技术对比

典型结果展示

应用案例解读

修饰蛋白质组学

磷酸化修饰

乙酰化修饰

N-糖基化修饰

典型结果展示

应用案例解读

多肽组学产品

应用方向

技术路线

产品优势

结果示例

蛋白质组学（Proteomics）是后基因组时代的一门重要学科，其研究对象为生物体内所有翻译表达并执行生命功能的蛋白质，蛋白质之间的相互关系，以及蛋白质与DNA、RNA和代谢产物之间的关联关系等。基于液相色谱和串联质谱联用（LC-MS/MS）的高通量高深度蛋白质组学研究可以获得蛋白质表达谱、蛋白质翻译后修饰、定量差异蛋白表达谱、多肽谱、定量差异多肽谱，以及蛋白质相互作用信息，可应用于基因表达研究、辅助基因注释和修正、蛋白标志物和多肽标志物的发现、功能蛋白和多肽的挖掘、信号通路和分子机理机制等研究。

截至2019年2月，华大科技质谱平台在蛋白质组学和代谢组学领域发表85篇研究论文，累计影响因子达432。

蛋白质组学研究工作流程

（含严格质控体系 - 数据质量保证）

BGI 华大基因

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

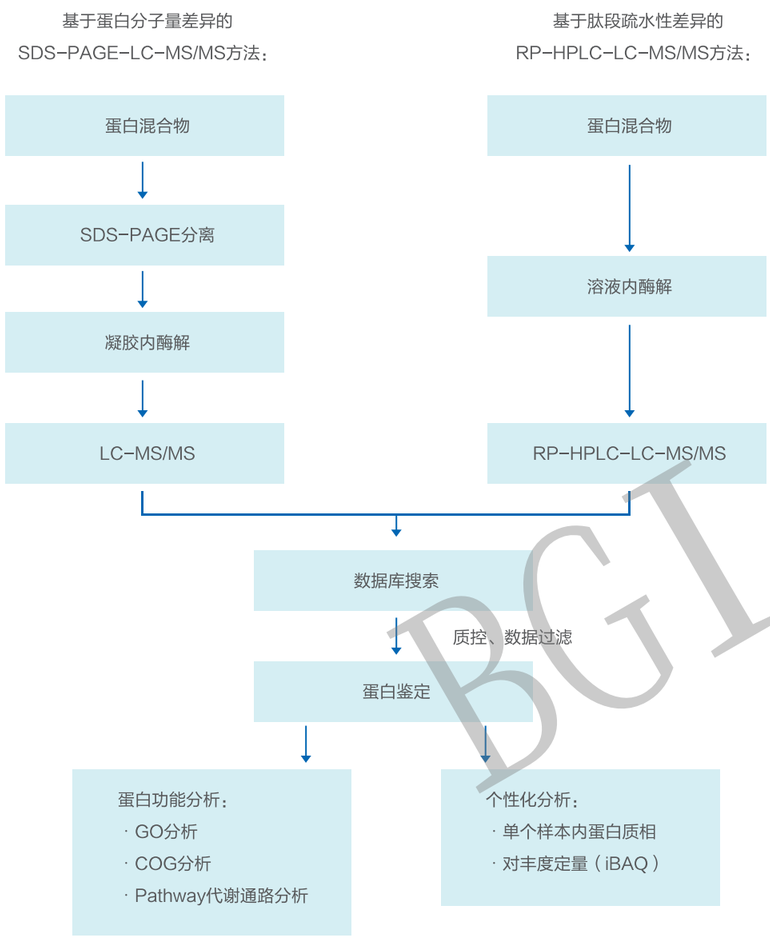


图1 蛋白全谱分析技术流程

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

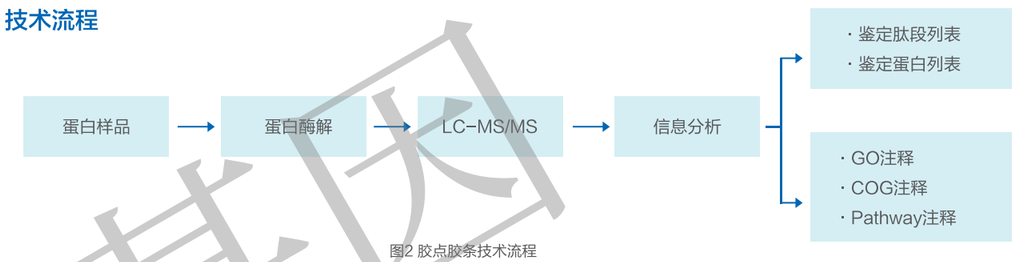


图2 胶点胶条技术流程

• 信息分析流程

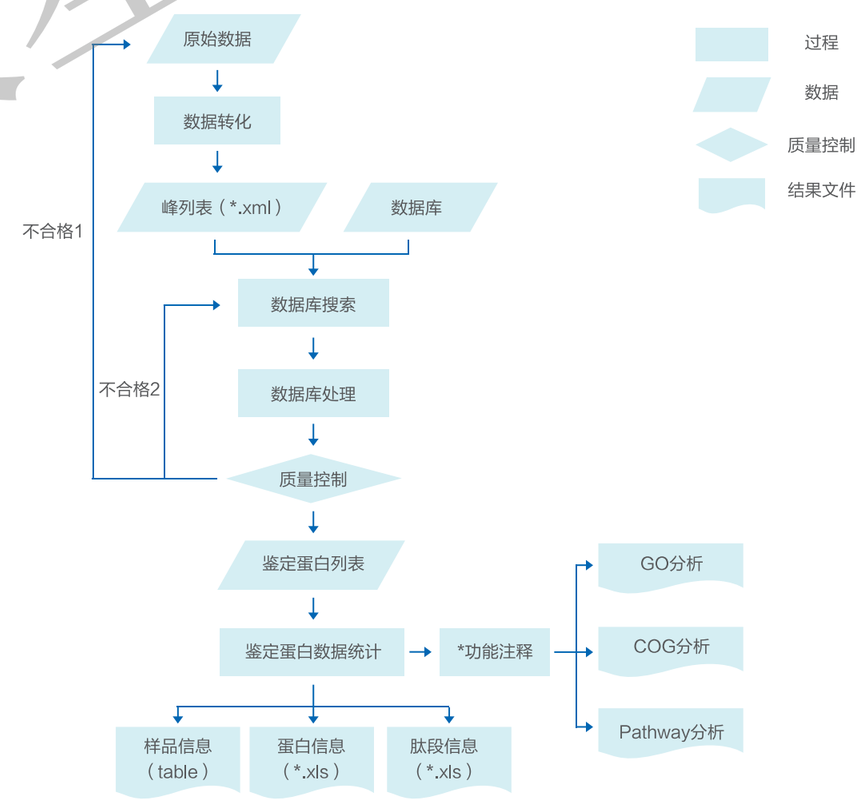


图3 蛋白鉴定信息分析流程，*表示胶条产品特有

BGI 华大基因

- **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	22个工作日
	胶点胶条鉴定	Q Exactive HF或TripleTOF 5600	10个工作日

典型结果展示

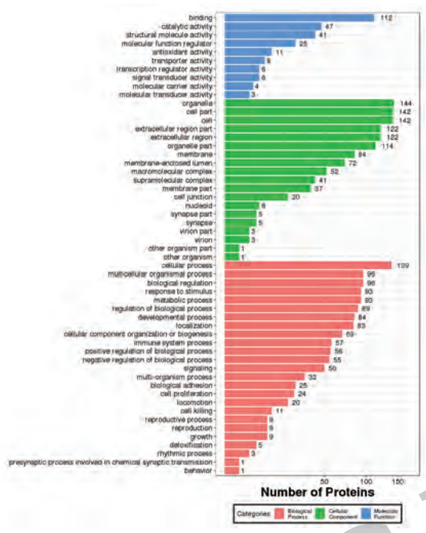


图4 蛋白质GO功能注释

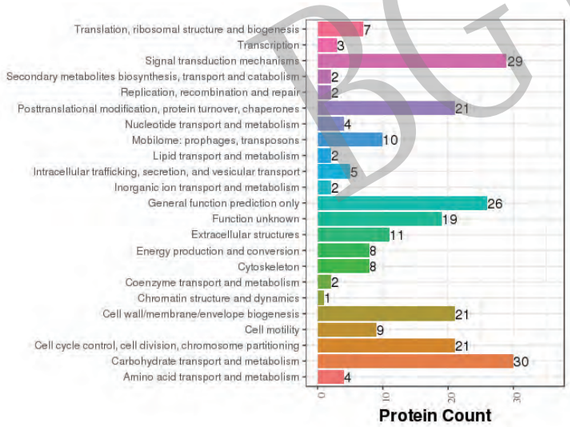


图5 COG注释结果统计

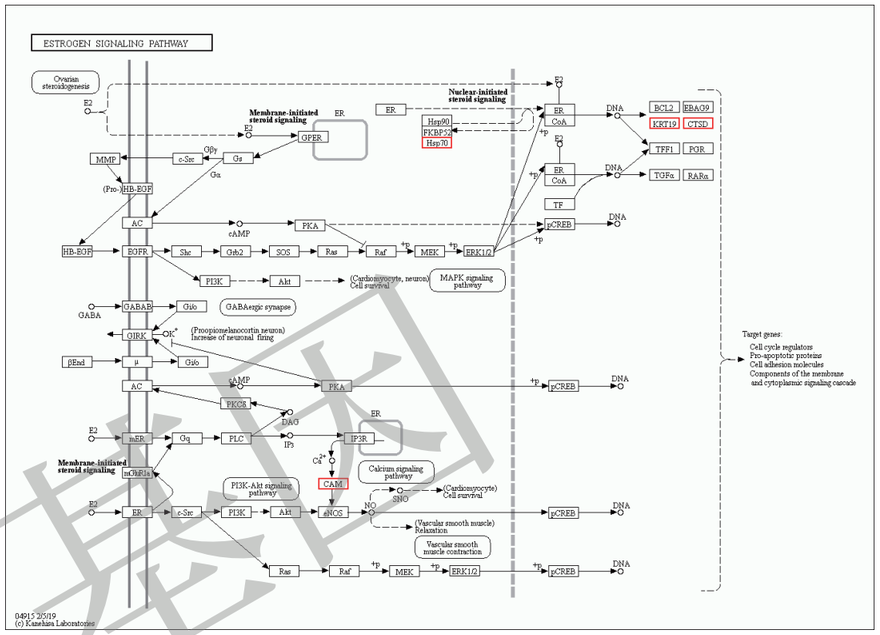


图6 KEGG数据库详细的通路图

应用案例解读

- **典型案例节选**

文章名	研究策略	发表时间	发表期刊
Improvements to the Rice Genome Annotation Through Large-Scale Analysis of RNA-Seq and Proteomics Data Sets	蛋白全谱	2019	Mol Cell Proteomics
Proteomic evidence that ABCA4 is vital for traumatic proliferative vitreoretinopathy formation and development	胶条	2019	Experimental Eye Research
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	2017	PeerJ

BGI 华大基因

案例一

基于质谱技术对大枣的花进行蛋白质组全谱分析

Shot-gun proteome and transcriptome mapping of the jujube floral organ and identification of a pollen-specific S-locus F-box gene

研究内容：大枣花蛋白组学研究　　研究策略：蛋白全谱分析　　发表时间：2017年　　发表期刊：PeerJ

- **研究背景**

花是植物的生殖器官，随着花期的结束形成植物果实的一部分。目前关于大枣花中表达的蛋白质的数量和特性尚未研究清楚。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。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对大枣花进行蛋白质全谱学分析。

- **研究思路**

实验材料：9年生枣树上收集的萌芽花

实验流程：

蛋白提取及质量测定 → 溶液内蛋白消化 → SCX色谱分离肽段 → LC-MS/MS → 信息分析 → 后续转录组分析

图1 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%的蛋白基因可以在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。

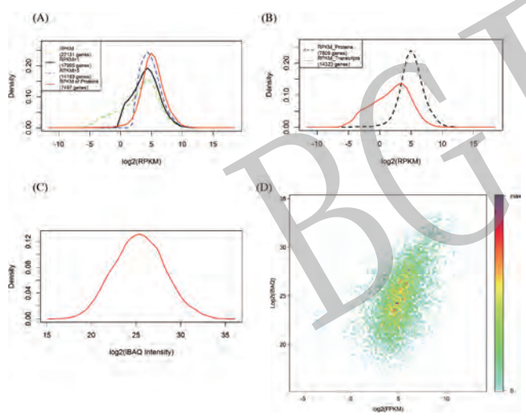


图2 蛋白质组学和RNA-Seq数据比较分析

（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%。

图3 前10个蛋白质组和转录组GO类别

（A. 生物过程；B. 细胞组成；C. 分子功能）

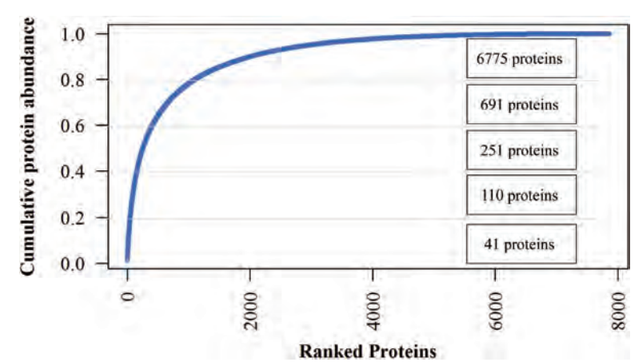


图4 枣花蛋白全谱中蛋白丰度分布情况

- **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。

BGI 华大基因

案例二

高通量转录组测序和蛋白质组数据对水稻基因组注释的提升作用

Improvements to the Rice Genome Annotation Through Large-Scale Analysis of RNA-Seq and Proteomics Data Sets

研究内容：辅助水稻基因组注释研究　研究策略：转录组测序+蛋白质组检测　发表时间：2019年　发表期刊：Mol Cell Proteomics

- **研究背景**

基因测序作为研究动植物遗传品质的重要工具，在水稻物种上已经得到了研究应用。然而，基因组测序所获得的所有基因注释仍面临着很大的挑战。本研究通过将大量水稻蛋白质相关的公共数据进行重新分析，得到了基因序列注释的校正，并发现了新的基因。

- **研究结果**

1、通过公共蛋白质组数据与水稻基因组和转录组数据搜索鉴定；
2、发现1,584个在之前基因模型中没有得到解释的新肽段；
3、101个新的基因位点通过新肽段得到匹配。

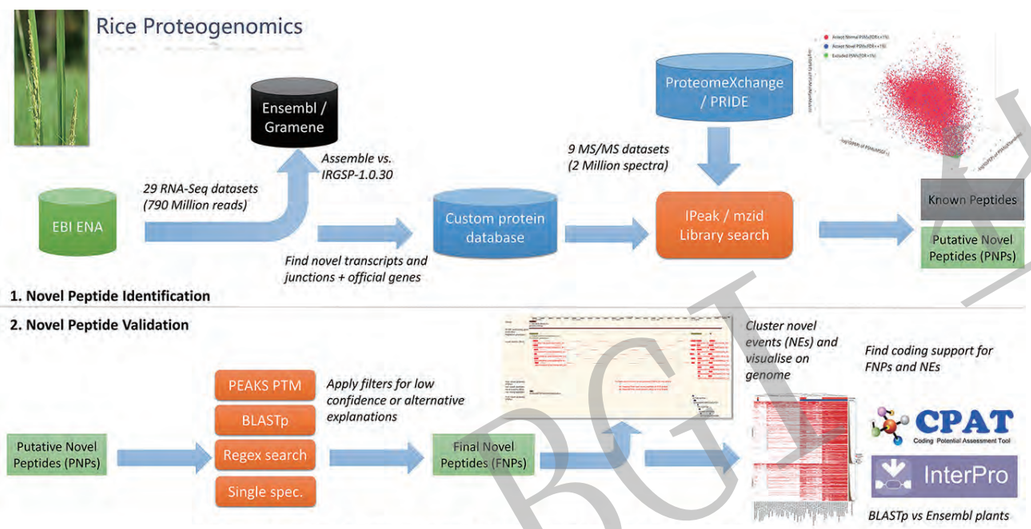


图1 研究内容和技术路线图

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。

- **技术原理**

由于iTRAQ、IBT、TMT三种标记定量技术原理类似，接下来我们以iTRAQ试剂为例，介绍蛋白质标记定量技术。

从结构上看，iTRAQ试剂包括3个化学基因：报告基团、平衡基团和反应基团（图1）。不同的报告基团与其相对应的平衡基团的分子量之和是相同的；反应基团与肽段的氨基发生特异性的共价结合反应，将iTRAQ试剂连接到肽段上。在二级质谱中，平衡基团发生中性丢失，而报告基团则会产生相应的报告离子，这些报告离子的强度就代表了相应样本蛋白质/肽段的强度，从而实现蛋白的相对和绝对定量。

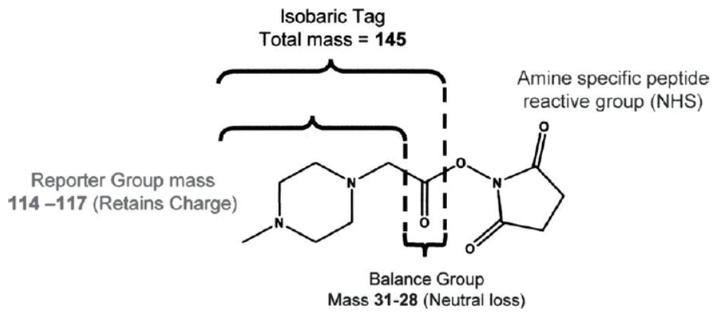


图1 iTRAQ试剂结构

iTRAQ试剂结构包括三个部分：报告基团、平衡基团、反应基团，其中同位素标签部分是指报告基团和平衡基团两部分。

BGI 华大基因

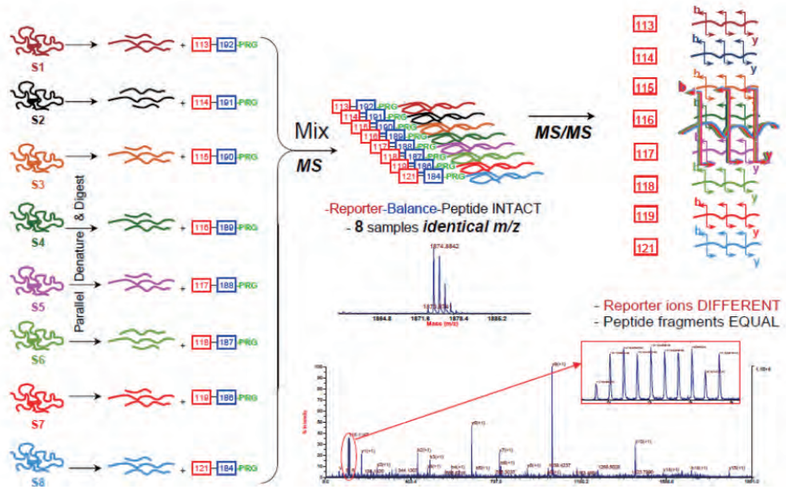


图2 蛋白定量iTRAQ原理示意图

蛋白酶解后的肽段通过与iTRAQ试剂标记反应而带上了不同的标签，样本混合后进入质谱检测。一级质谱中，由于不同标签的分子量相同，所以不同样本中相同肽段的表现无差异；二级质谱中，由于不同标签在平衡基团和报告基团间不同配比的碎裂，产生不同的m/z大小的报告离子，其强度代表了不同样本间多肽的相对量的差异。

• 技术流程

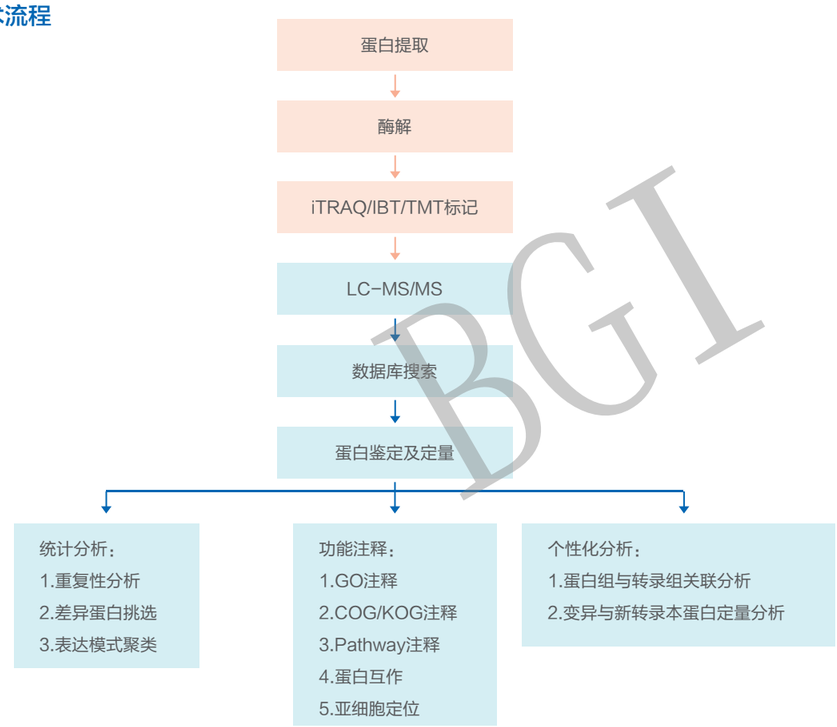


图3 蛋白标记定量分析技术流程

• 信息分析流程

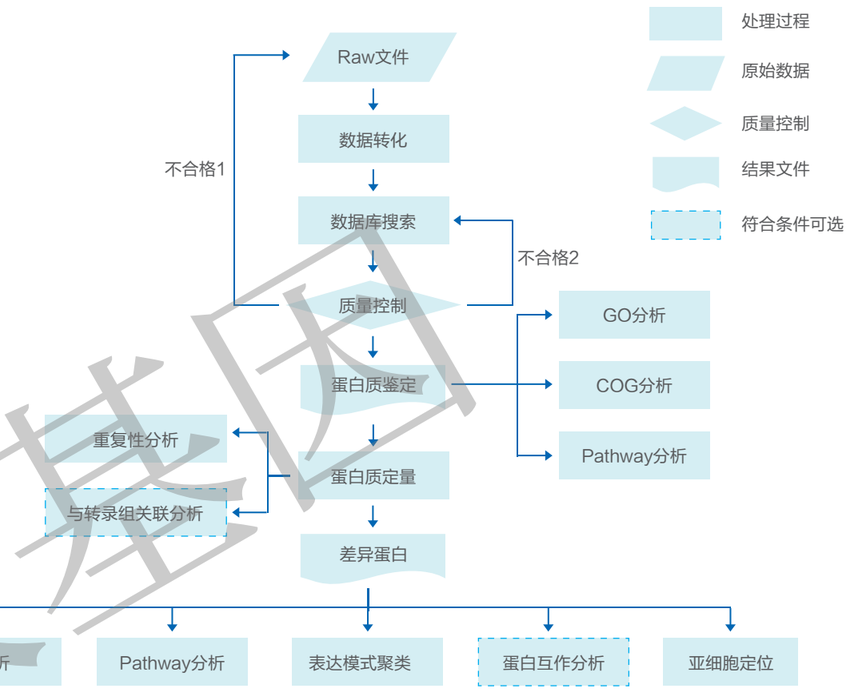


图4 蛋白标记定量信息分析流程

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能获取窗口范围内所有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谱图；通过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。

BGI 华大基因

- **技术流程**

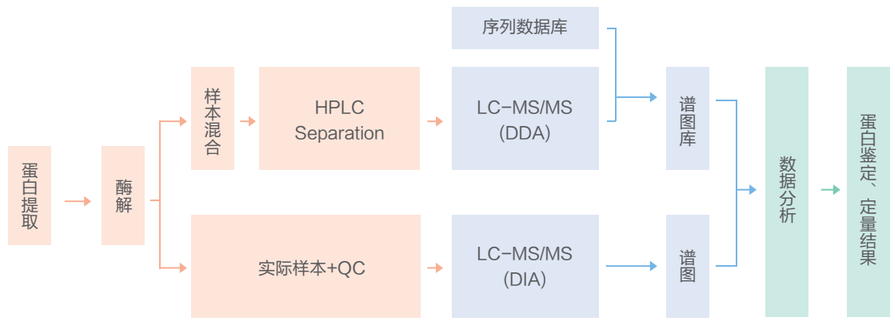


图5 蛋白DIA定量分析技术流程

- **信息分析流程**

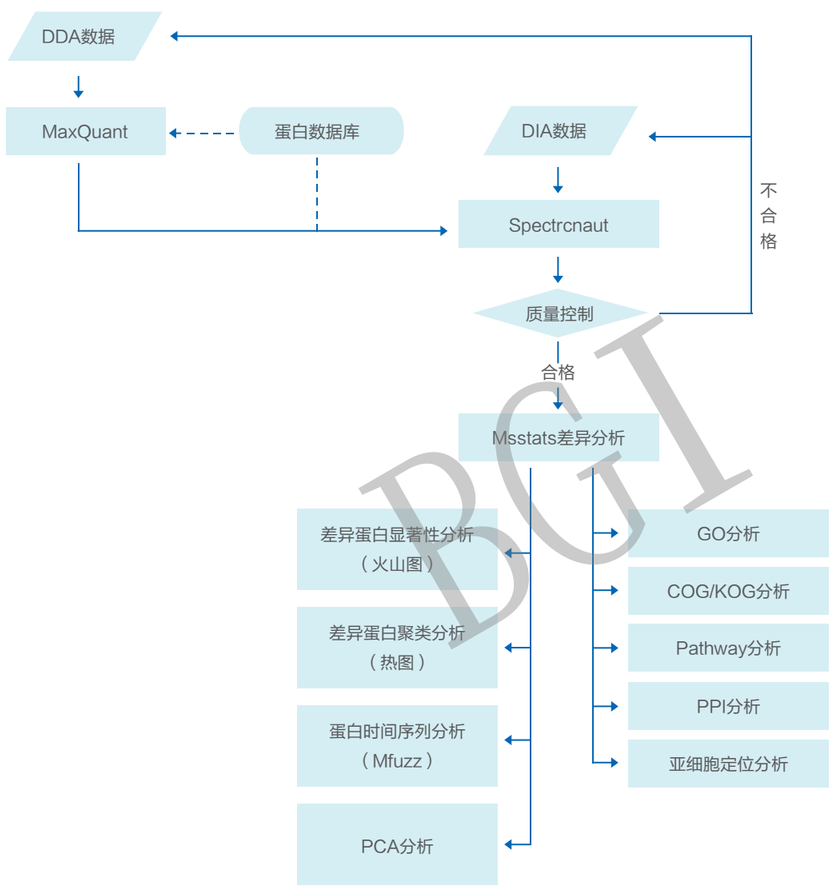


图6 蛋白DIA定量信息分析流程

目标蛋白MRM定量分析

目标蛋白MRM定量分析产品是指基于质谱的多反应检测技术（Multiple Reaction Monitoring，MRM），有目标地检测分析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定量的研究策略。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及以上目标蛋白，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，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。

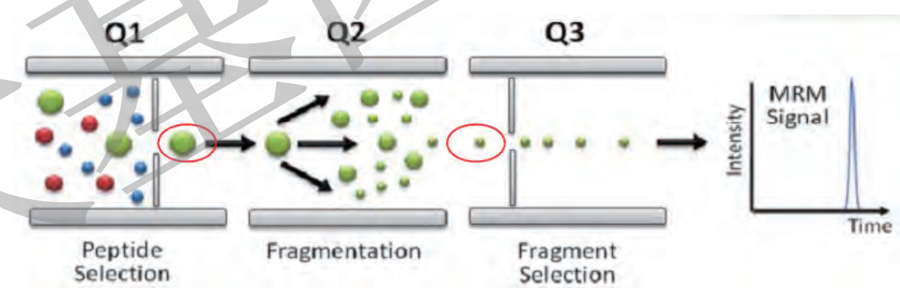


图7 目标蛋白MRM技术原理介绍

表1 目标蛋白MRM与抗体技术对比

	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

BGI 华大基因

- **技术流程**

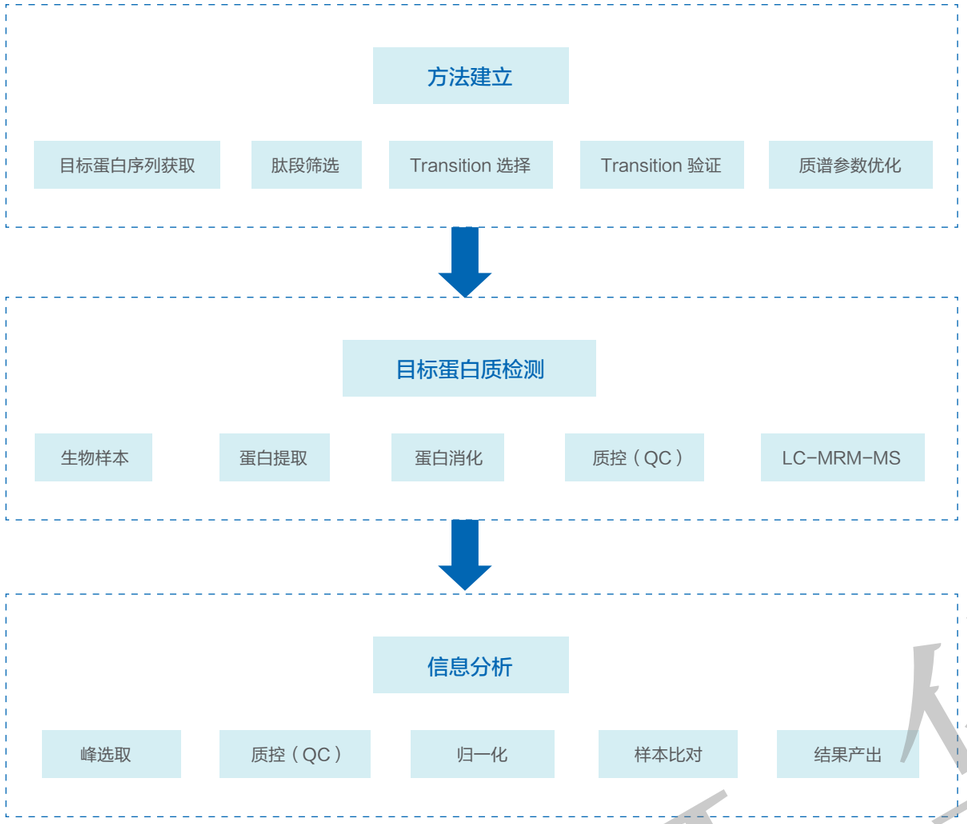


图8 目标蛋白MRM技术流程

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	10-20个工作日
	IBT	Thermo Q Exactive HF-X/HF	10-20个工作日
	TMT	Thermo Q Exactive HF-X/HF	10-20个工作日
	DIA	Thermo Q Exactive HF/HF	10-30个工作日
	MRM	QTRAP 6500+/TripleTOF 5600	方法建立10-20个工作日 样本定量10-15个工作日

表3 蛋白组定量技术对比

方法 / 比较项目	2-DE/DIGE	质谱检测技术							
		标记定量技术						非标记定量技术	
		体内标记法	体外标记法						
		^{15}N标记	SILAC	ICAT	iTRAQ	TMT	IBT	Label-free（DDA）	Label-free（DIA）
样本类型	丰度较高蛋白	可培养的细胞	可培养的细胞	所有样本	所有样本	所有样本	所有样本	所有样本	所有样本
通量	1-2	2	2-3	2	2-8	2-10	2-10	可>10	可>10
蛋白鉴定数	+	++	++	++	++	+++	+++	+++	+++
动态范围	+	++	++	++	++	++	++	+++	+++
定量准确性	+	++	++++	++	+++	+++	+++	++	+++
实验经费	+	+++	+++	++	++	+++	+	+++	++

典型结果展示

iTRAQ/IBT/TMT、DIA定量蛋白质组学结果展示

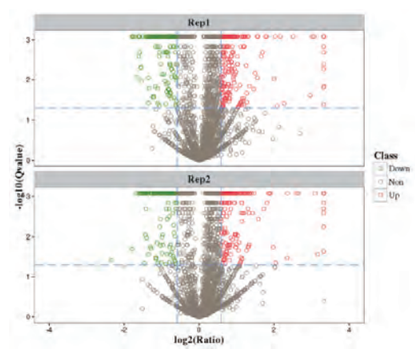


图9 差异蛋白火山图

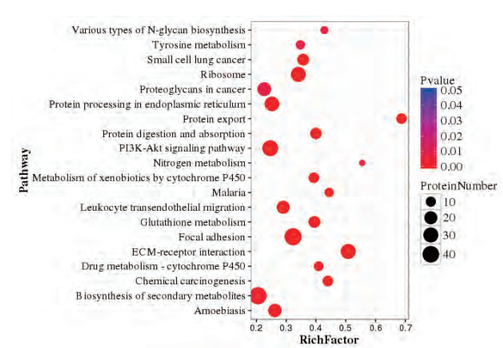


图10 显著差异的pathway统计图

BGI 华大基因

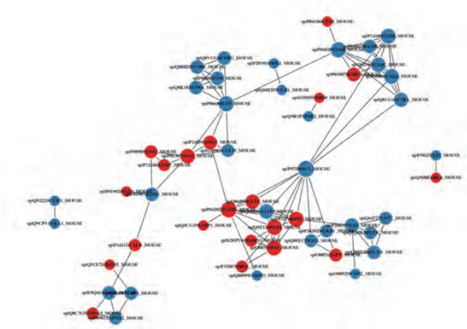


图11 差异蛋白互作关系网络图

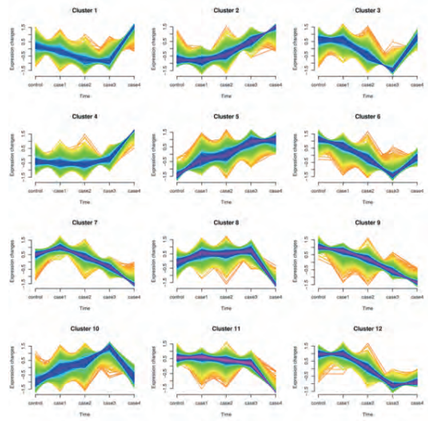


图12 时间序列分析图

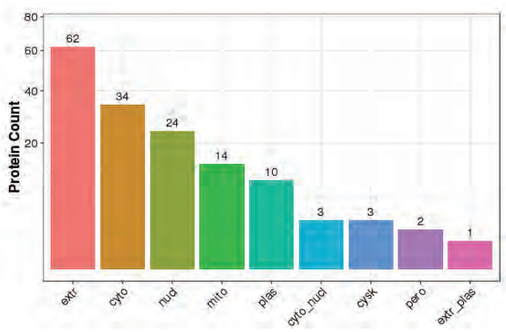


图13 差异蛋白亚细胞定位柱状图

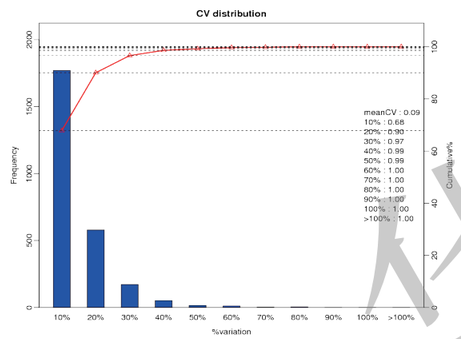


图14 重复性实验CV分布图

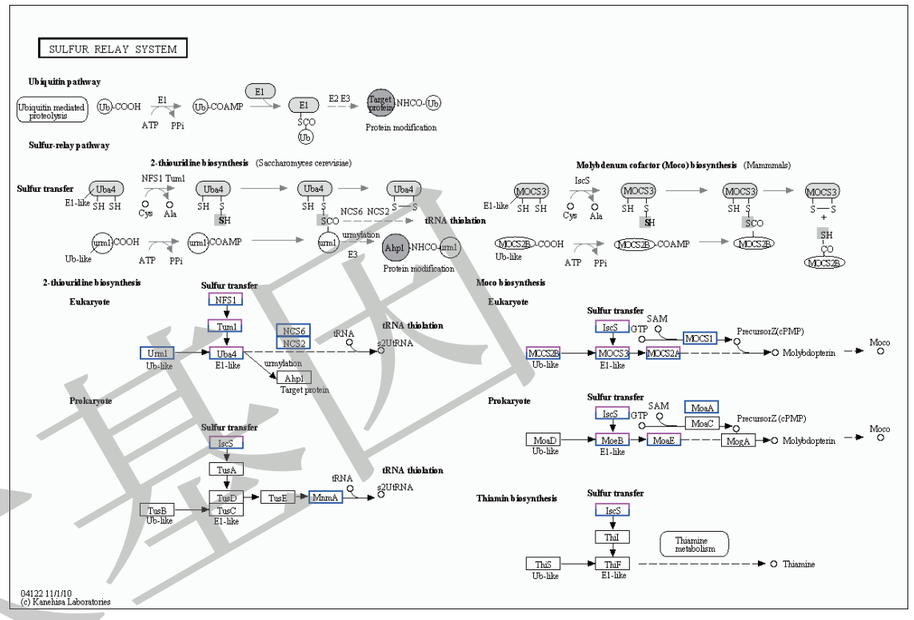


图17 转录组和蛋白组代谢通路整合

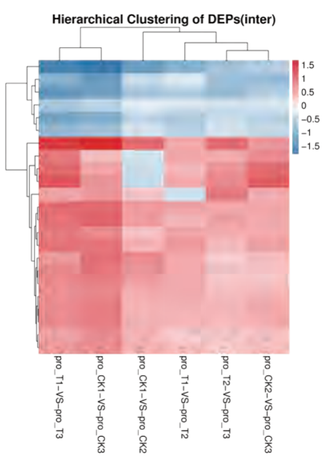


图15 差异蛋白的表达聚类图

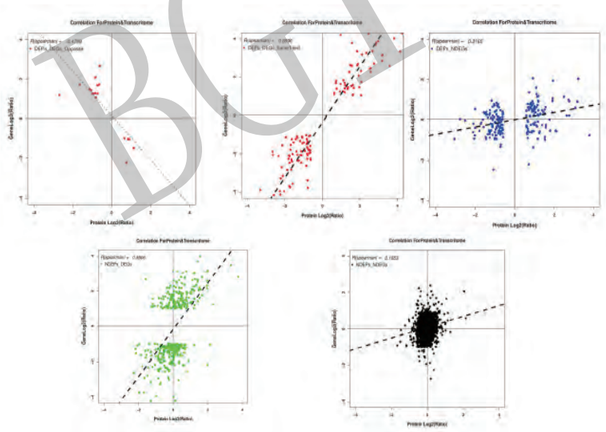


图16 蛋白质组与转录组/RNA-seq表达量关联分析

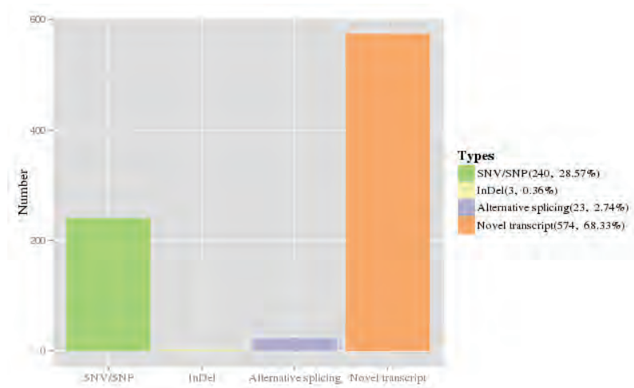


图18 变异与新转录本检测

BGI 华大基因

MRM典型结果展示

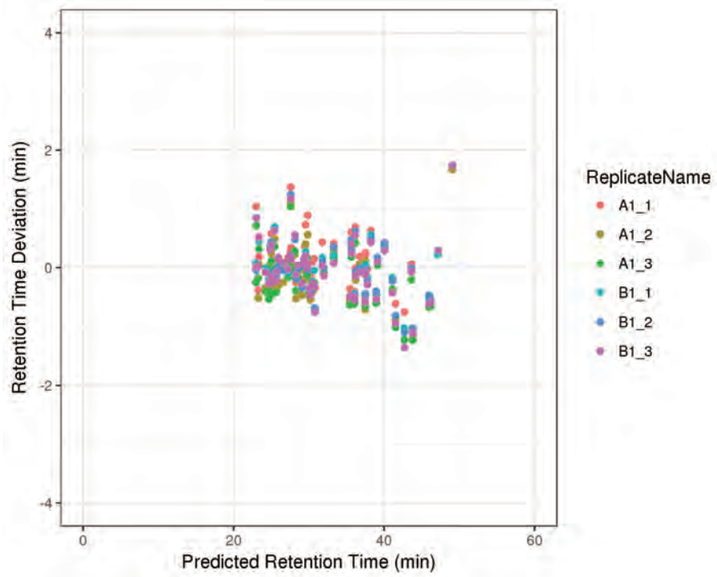


图19 目标蛋白每个肽段的色谱出峰时间误差

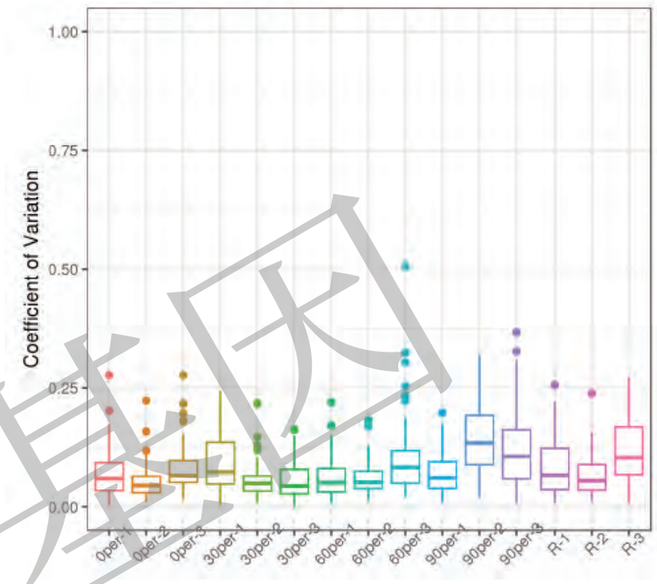


图21 目标蛋白所有肽段质定量信息变异系数分布

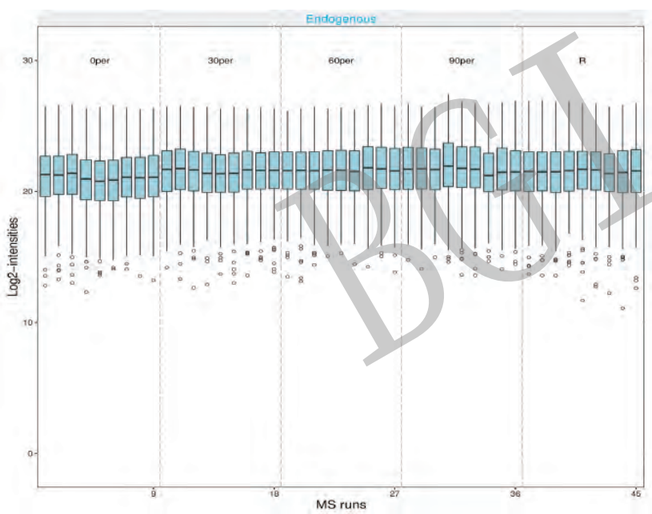


图20 目标蛋白所有transition归一化强度分布

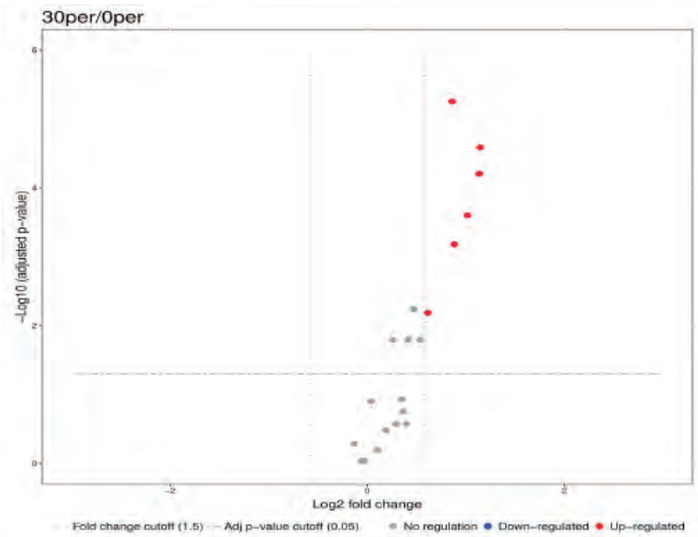


图22 目标蛋白在所有比较组的相对定量分布

BGI 华大基因

应用案例解读

• 典型案例汇总

文献名称	研究策略	发表时间	发表期刊
Characterization of proteome variation during modern maize breeding	iTRAQ定量+转录组	2019	Molecular & Cellular Proteomics
Proteomic profiling and integrated analysis with transcriptomic data bring new insights in the stress responses of Kluyveromyces marxianus after an arrest during high-temperature ethanol fermentation	iTRAQ定量+MRM定量	2019	Biotechnology for Biofuels
Rapid evolution of protein diversity by de novo origination in Oryza	MRM定量	2019	Nature Ecology & Evolution
The alternative sigma factor SigF is a key player in the control of secretion mechanisms in Synechocystis sp. PCC 6803	iTRAQ定量	2019	Environ Microbiol
Proteomic Analysis and Biochemical Correlates of Mitochondrial Dysfunction after Low-Intensity Primary Blast Exposure	TMT定量	2019	Journal of Neurotrauma
Reagents for Isobaric Labeling Peptides in Quantitative Proteomics	IBT定量	2018	Anal. Chem
Complementary iTRAQ-based proteomic and RNA sequencing based transcriptomic analyses reveal a complex network regulating pomegranate (Punica granatum L.) fruit peel colour	iTRAQ定量+转录组	2018	Scientific Reports
A comprehensive proteomic investigation of Ebf1 heterozygosity in pro-B lymphocytes utilizing data independent acquisition (DIA)	iTRAQ定量+PRM定量	2018	J. Proteome Res
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	Frontiers in Plant Science
Quantitative evaluation of serum protein uncovers a protein signature related to maturity-onset diabetes of the young	iTRAQ定量+MRM定量	2018	J. Proteome Res
The role of PTRF/Cavin1 as a biomarker in both glioma and serum exosomes	TMT定量+DIA定量	2018	Theranostics
Benchmarking common quantification strategies for large-scale phosphoproteomics.	非标定量+TMT定量	2018	Nature Communications
Distinct Proteomic, Transcriptomic, and Epigenetic Stress Responses in Dorsal and Ventral Hippocampus	DIA定量	2018	Biol. Psychiatry
Development of a Gill Assay Library for Ecological Proteomics of Threespine Sticklebacks (Gasterosteus aculeatus)	DIA定量	2018	Molecular & Cellular Proteomics
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	2017	Nature Communications

案例一

现代玉米育种中蛋白质组的差异表征

Characterization of Proteome Variation During Modern Maize Breeding

研究内容：玉米育种中蛋白质组　研究策略：iTRAQ+转录组　发表时间：2019年　发表期刊：Molecular & Cellular Proteomics

• 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较大的不一致性。

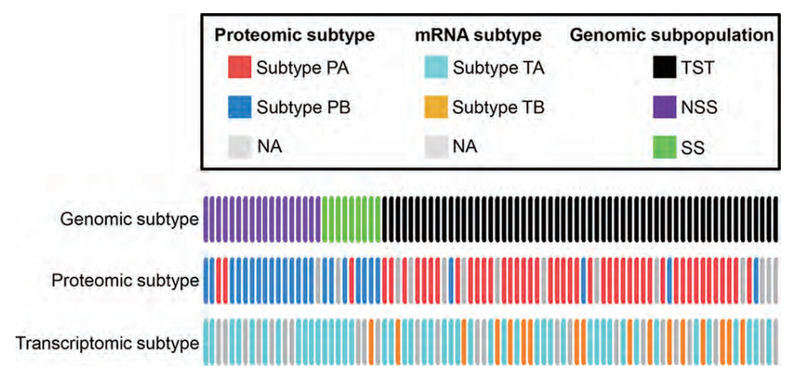


图1 蛋白质组或转录组亚型与基因组亚型的关联

BGI 华大基因

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。

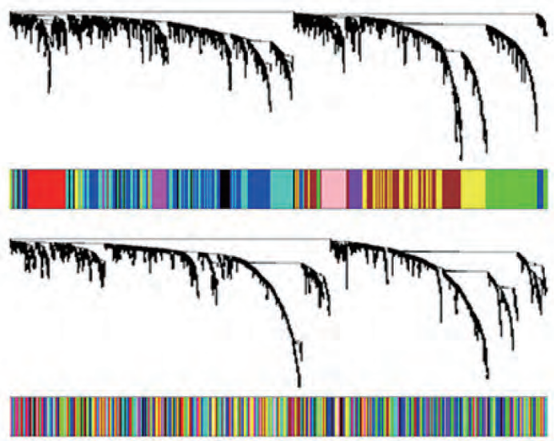


图2 WGCNA衍生的共表达树状图和相应的模块，蛋白质网络（上）和mRNA网络（下）

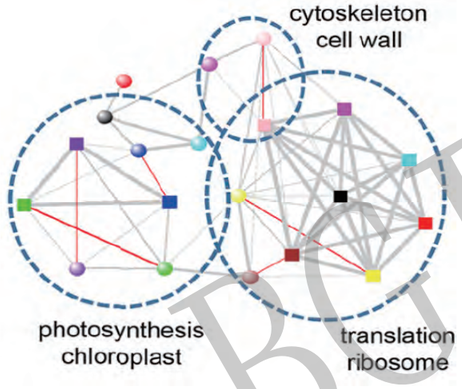


图3 蛋白质和mRNA网络之间的模块重叠

（模块点的颜色与WGCNA中的模块颜色对应。球形模块代表蛋白质数据，矩形模块代表mRNA数据，线宽取决于相应模块之间的重叠显著性，红线表示显著相关的mRNA-蛋白质网络。）

为了寻找能在蛋白质和mRNA水平上区分基因组亚群的基因，研究者采用一种基于机器学习的SOM（self-organizing maps）分析，来发现多维度下的样本关联性。

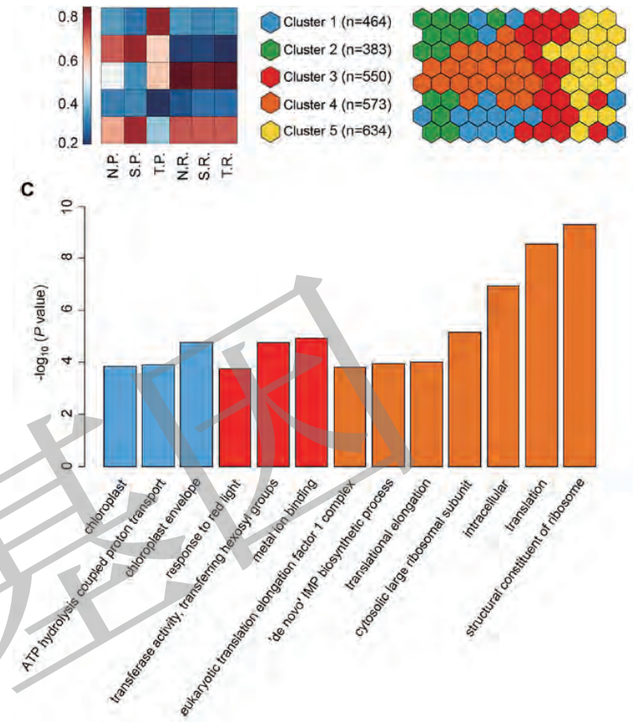


图4 神经元按相似性传播聚类后的聚类图以及基因簇GO分析柱形图

（柱的颜色与基因簇的颜色保持一致）

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个与叶绿体结构相关的基因簇的其中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。

BGI 华大基因

案例二

蛋白质组和基因组在异种移植模型中联合揭示乳腺癌治疗靶点

Proteogenomic integration reveals therapeutic targets in breast cancer xenografts

研究内容：乳腺癌治疗靶点　研究策略：iTRAQ+基因组　发表时间：2017年　发表期刊：Nature Communications

• 研究背景

在许多病例中使用NGS技术难以检测到潜在的成药基因改变。全面的蛋白质组学分析为验证基因水平变异和发现新的药物靶点提供了一种潜在的可能性。

• 研究思路

从原发性或转移性乳腺肿瘤中选择了24个异种移植（PDX）模型，用于全面的蛋白质组学表征分析。然后分别对23个PDX模型进行 DNA 和 RNA 测序，并在最后1个模型中对热点突变进行Sanger DNA测序。用iTRAQ（4-plex）对所有24个PDX模型的蛋白质组进行定量，然后用非标记定量（LFQ）对其中的18个PDX模型进行验证。

• 研究结果

1、从这24个模型的RNA-Seq数据中定量了16,209个人类特异性基因的mRNA表达。应用了两种不同的MS数据采集方法iTRAQ和LFQ，分别定量了12,794种人类蛋白质（11,879种基因）和8,648种蛋白质（8,035种基因）的表达。此外使用iTRAQ还鉴定了总共56,874个磷酸化修饰位点。

2、首先评估了24个PDX模型中的mRNA表达和蛋白质表达之间的相关性：83.6%的基因具有正相关性（皮尔逊系数R=0.536）

图1 mRNA和蛋白表达水平之间的相关性揭示了一些mRNA和蛋白质共调控或不相关的信号通路

3、多数关键基因在CNV、mRNA、蛋白质和磷酸化修饰位点分析中的表达模式与临床亚型的结果一致。

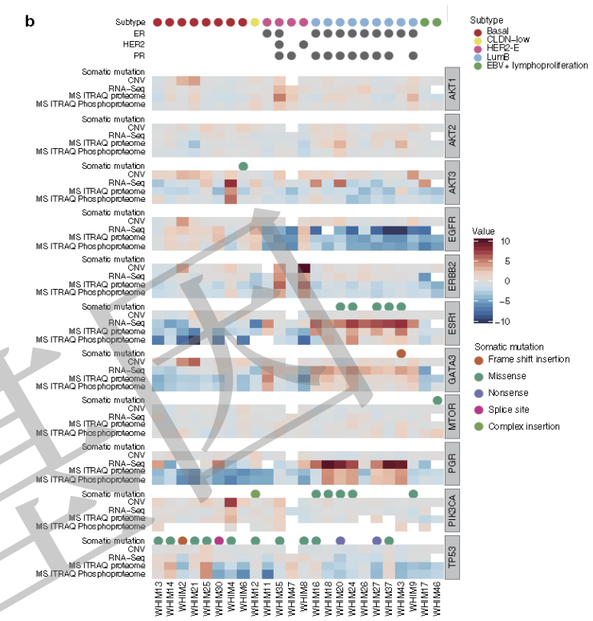


图2 异种移植模型中的基因组和蛋白质组表达水平显示CNV、mRNA、蛋白质组和磷酸化修饰位点之间的变异相关性。所涉及的基因包括4种临床亚型。

4、基于PAM50基因的RNA-Seq数据显示转录组表达模式很大程度上与乳腺癌的内在亚型一致。与之相比蛋白质组和磷酸蛋白组与内在亚型的关联性则要稍弱一些。总的来说，虽然基底和管腔亚型的分布保持一致，但在mRNA、蛋白质和磷酸蛋白质水平上，PDX样本之间的层级距离是不同的。蛋白质组和磷酸化蛋白质组与mRNA表达定义的亚型之间的偏离表明了蛋白质组层面的分子异质性。

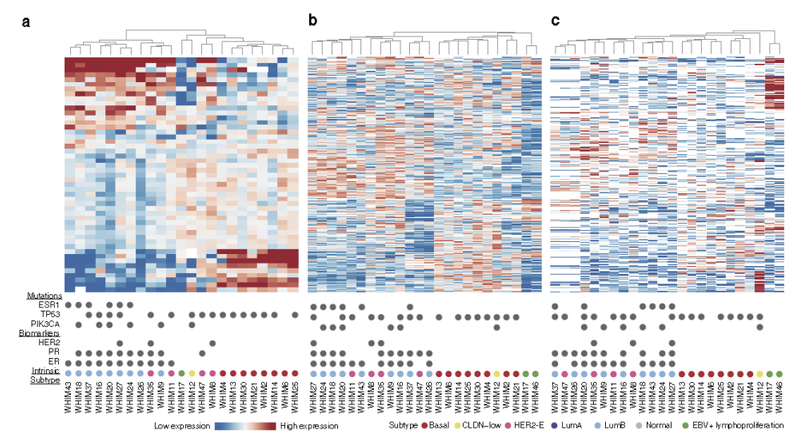


图3 a. 基于PAM50基因表达的PDX乳腺癌的转录组聚类； b. 基于最显著差异的436个蛋白质的PDX乳腺癌的蛋白质组聚类；c. 基于最显著差异的1737个磷酸化修饰位点的PDX乳腺癌的磷酸蛋白质组聚类

BGI 华大基因

5、蛋白质表达上调与大量CNV或mRNA上调情况保持一致。例如，在WHIM8和WHIM35样本的ERBB2基因中同样观察到了HER2蛋白和磷酸化修饰的异常表达，印证了使用拉帕替尼进行治疗的有效性。此外，在PDX和人体样本中，有些过表达的蛋白质在基因组水平上并没有明显的异常表达，包括EGFR的蛋白异常表达，以及ARAF、BRAF、HSP90AB1、PTPN11和TOP2A的异常磷酸化修饰等，表明了乳腺癌的潜在治疗机会。

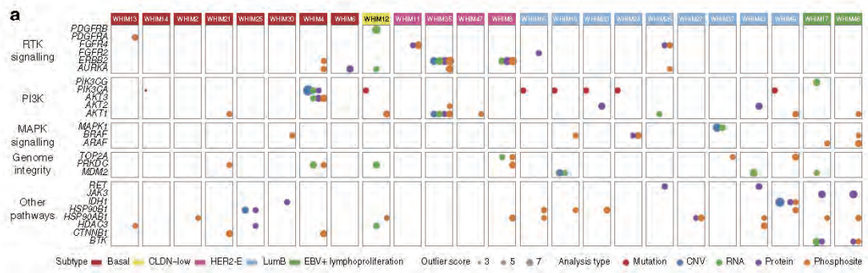


图4 异常表达分析揭示了在不同维度下4种乳腺癌亚型中RTK、PI3K、MAPK信号传导中潜在的药物靶点

• 研究结论

蛋白质组和磷酸化蛋白质组与mRNA表达定义的亚型之间的偏离表明了蛋白质组层面的分子异质性。差异蛋白表达和异常磷酸化修饰的研究为乳腺癌的治疗提供了潜在的机会。

案例三

蛋白定量DIA用于研究海马体不同区域应激反应机制

Distinct Proteomic, Transcriptomic, and Epigenetic Stress Responses in Dorsal and Ventral Hippocampus

研究内容：海马体不同区域应激反应机制　　研究策略：DIA　　发表时间：2018年　　发表期刊：Biol. Psychiatry

*注：SWATH技术与DIA技术原理相同，同为二级质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

生活中的压力会导致焦虑、创伤后应激障碍和人类认知障碍。急性压力通过改变海马体不同区域的功能引发不同的应答反应，因此海马体是参与急性应激反应的一个关键组织。海马体的内在组织其长背腹侧轴高度保守，背侧（前部）海马体（dHC）参与介导空间导航和记忆形成，而腹侧（后部）海马体（vHC）则参与情绪调节。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。

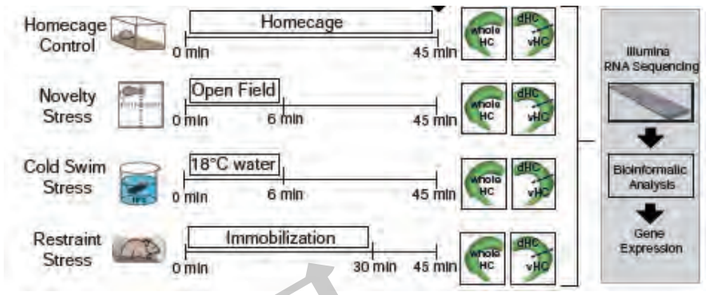


图1 实验设计

压力条件：家庭控制（homecage 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，作者发现了Wfs1和Epha7这两个基因，它们已被证实富含于海马CA1神经元中，这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。

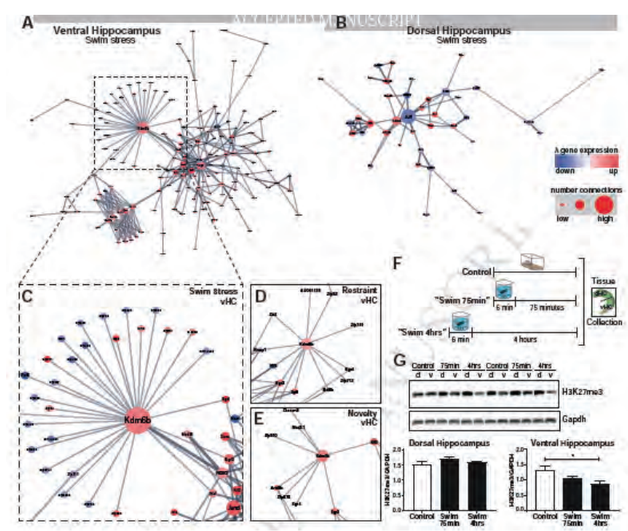


图2 腹侧海马中应激反应的表观遗传网络

BGI 华大基因

案例四

血清定量蛋白质组学揭示成熟型糖尿病的蛋白特征

Quantitative Evaluation of Serum Proteins Uncovers a Protein Signature Related to Maturity-Onset Diabetes of the Young (MODY)

研究内容：成熟型糖尿病血清蛋白质组学　研究策略：iTRAQ+MRM　发表时间：2018年
发表期刊：Journal of Proteome Research

• 研究背景

成熟型糖尿病（MODY）是一种单基因遗传性疾病。MODY中的基因突变经常引起非同义变化，直接导致蛋白质的功能失真和病理结果。研究人员提出，MODY家族中发现的遗传突变可能导致蛋白质丰度的紊乱，这种现象在血清中尤为突出。从三代维吾尔族MODY家族收集的血清样本，去除处理后的血清蛋白以表征MODY相关血清蛋白，然后使用靶向定量蛋白质组学进行验证。

• 研究思路

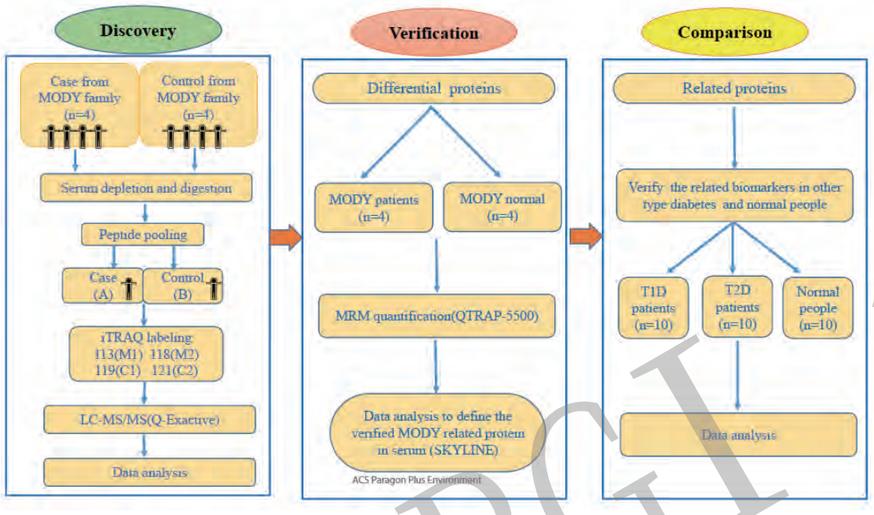


图1 研究思路

• 研究结果

1、初步鉴定出32种MODY相关的血清蛋白。
2、针对个体样品进一步测试，证明了MODY患者中具有显著不同丰度的12个候选物。
3、比较了Ⅰ型糖尿病、Ⅱ型糖尿病、MODY、健康受试者血清中的12种蛋白质，揭示了与MODY相关的血清蛋白质如SERPINA7、APOC4、LPA、C6和F5的蛋白特征。

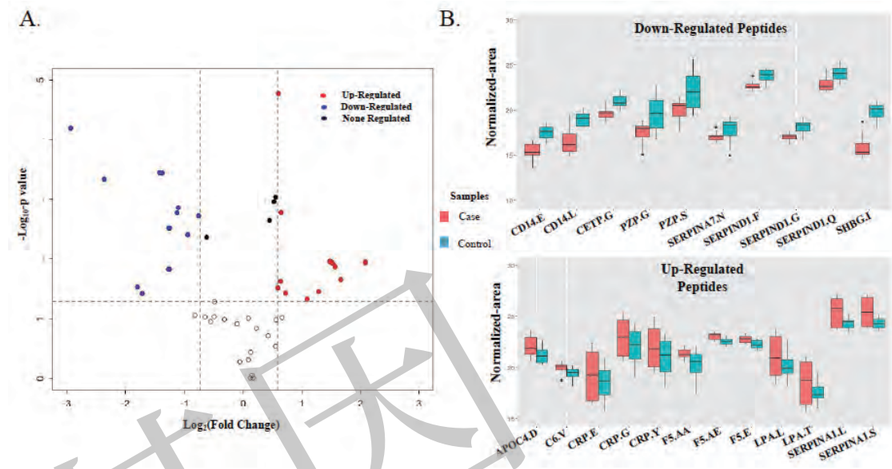


图2 MRM验证MODY相关血清蛋白结果

A. MODY患者和健康对照之间血清蛋白的丰度比的火山图，x轴表示两组蛋白质的对数倍数变化，y轴表示倍数变化的p值取对数；
B. 个体MODY和健康对照样品之间MODY相关蛋白丰度的对比分析，分布图用于比较两组之间12种血清蛋白的22种肽的丰度差异。

案例五

硬骨鱼肌间骨和肋骨的比较蛋白质组学研究

Comparative proteomics analysis of teleost intermuscular bones and ribs provides insight into their development

研究内容：硬骨鱼肌间骨和肋骨的比较蛋白质组学　研究策略：iTRAQ+MRM　发表时间：2017年　发表期刊：BMC Genomics

• 研究背景和思路

肌间骨（IBs）和肋骨都是硬骨鱼骨骼系统的一部分，但具有不同的发育过程。目前尚未研究鱼类IBs和肋骨的化学成分以及它们发育的潜在机制。本研究中，组织学结构显示一个含有破骨细胞的骨腔存在于团头鲂的肋骨中，但不存在于肌间骨中。我们构建了第一个包含肌间骨和肋骨的鱼骨蛋白质组学图谱，并通过iTRAQ LC-MS/MS蛋白质组学分析鉴定了IBs和肋骨之间差异表达的蛋白质。

在6只1岁、6只2岁的团头鲂中，取肌间骨和肋骨中的组织样本进行SDS-PAGE蛋白分离后，通过标记定量蛋白质组学技术（iTRAQ）进行蛋白定量，通过MRM验证iTRAQ蛋白定量结果中差异表达的蛋白质。

• 研究结果

1、2,342种蛋白质被定量，1,451种蛋白质在GO注释中被成功注释。
2、在1-IB-vs-1-Rib（1岁肌间骨vs1岁肋骨）和2-IB-vs-2-Rib（2岁肌间骨vs2岁肋骨）的对照组中，共鉴定出93种和154种差异表达的蛋白质，表明这两种骨组织之间的化学组成明显不同。
3、与肌间骨相比，与破骨细胞分化相关的两种蛋白质（玻连蛋白b前体和基质金属蛋白酶-2）显著上调，这与组织结构的结果一致。
4、在1-IB-vs-2-IB和1-Rib-vs-2-Rib的比较组中，鉴定到33和51种差异表达的蛋白质，功能注释结果表明这些蛋白质参与调节骨骼发育和分化。
5、通过MRM测定验证了与骨发育相关的1-IB-vs-1-Rib和1-IB-vs-2-IB对照组中的11和13种候选蛋白。

BGI 华大基因

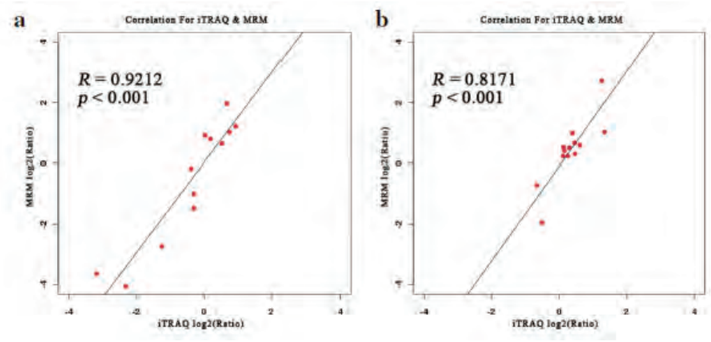


图1 两个对照组iTRAQ定量和MRM相关性定量相关性

注：a表示1-IB-vs-1-Rib，b表示1-IB-vs-2-IB

- **研究结论**

目前的研究表明，鱼肋骨和IBs的组成以及它们的生长发育涉及不同的关键蛋白质。

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。

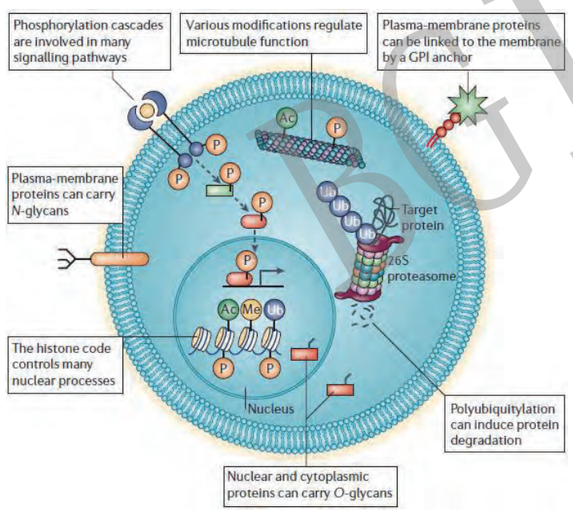


图1 细胞内翻译后修饰类型

Nature Reviews Molecular Cell Biology 7, 391–403

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着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。

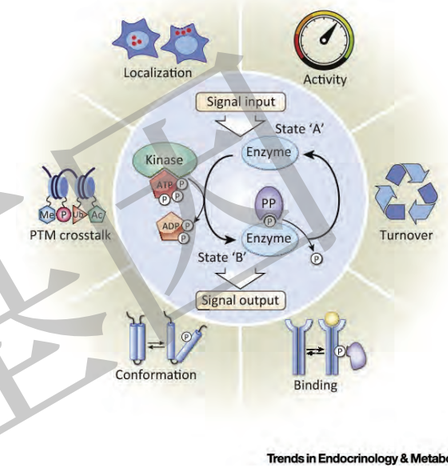


图2 蛋白磷酸化修饰生理功能

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。

- **技术原理**

蛋白磷酸化分析流程为：先对样品进行酶解、富集磷酸化肽段，然后用质谱（LC-MS/MS）进行检测，利用得到的质谱图与数据库搜索比较，从而鉴定磷酸化蛋白及磷酸化位点并对磷酸化肽段进行相对定量。

纯化蛋白磷酸化鉴定/磷酸化蛋白全谱鉴定技术流程

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。

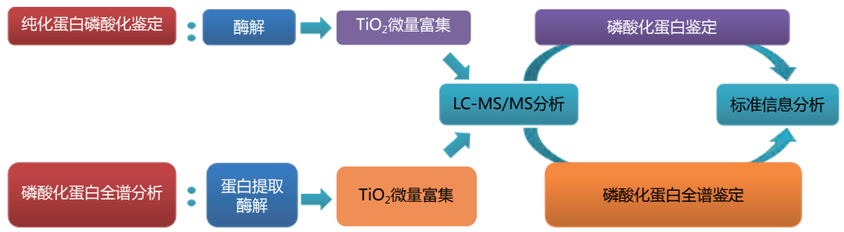


图3 纯化蛋白磷酸化鉴定/磷酸化蛋白全谱鉴定技术流程

BGI 华大基因

磷酸化蛋白定量技术流程

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，包括磷酸化蛋白iTRAQ定量及酪氨酸磷酸化蛋白Label-free定量。

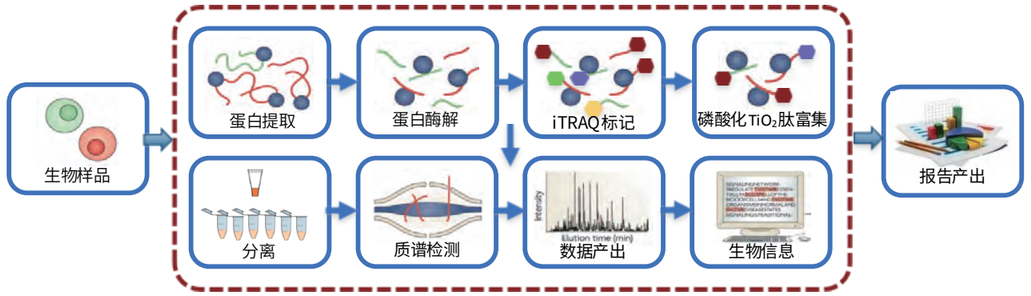


图4 磷酸化蛋白iTRAQ定量技术流程

目标磷酸化蛋白PRM定量技术流程

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、目标肽段进行选择性检测，从而实现对目标蛋白质/肽段进行相对定量的技术。它与MRM多反应监测技术（multiple reaction monitoring，又称选择反应监测技术SRM，selected reaction monitoring）类似，都可作为靶向蛋白质组研究策略。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。

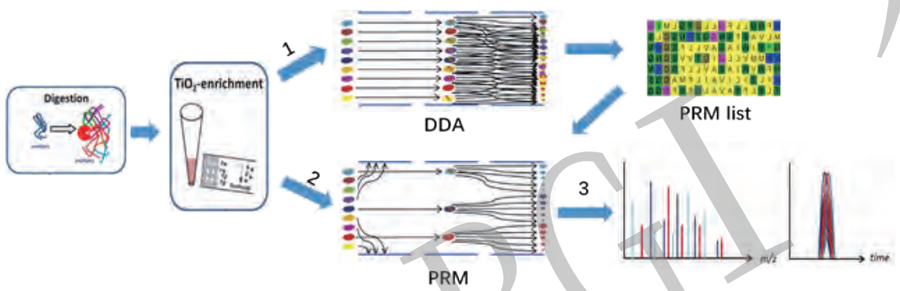


图5 目标磷酸化蛋白PRM定量技术流程

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。

• 信息分析流程

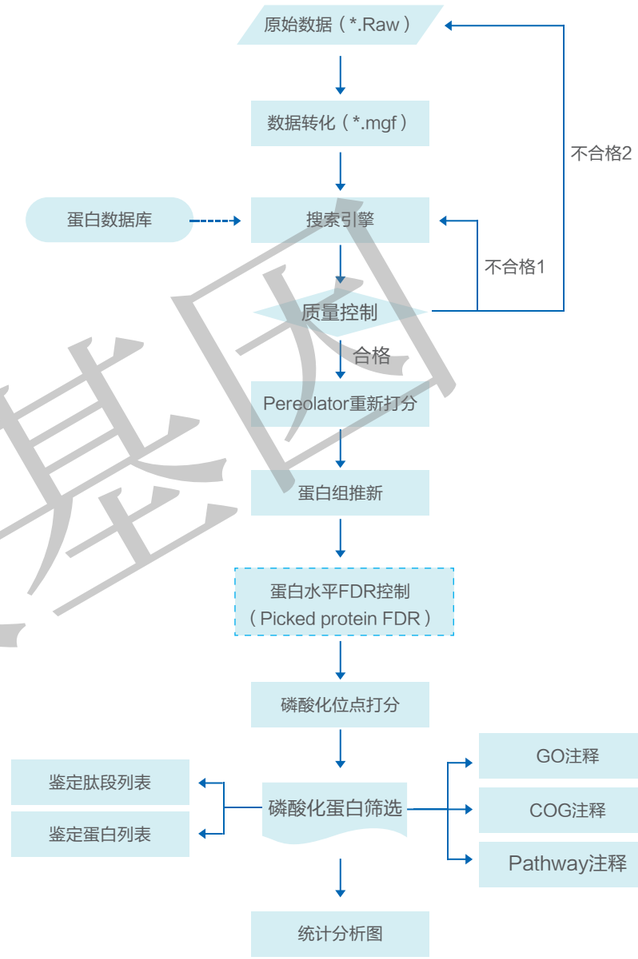


图6 磷酸化蛋白质鉴定信息分析流程图

BGI 华大基因

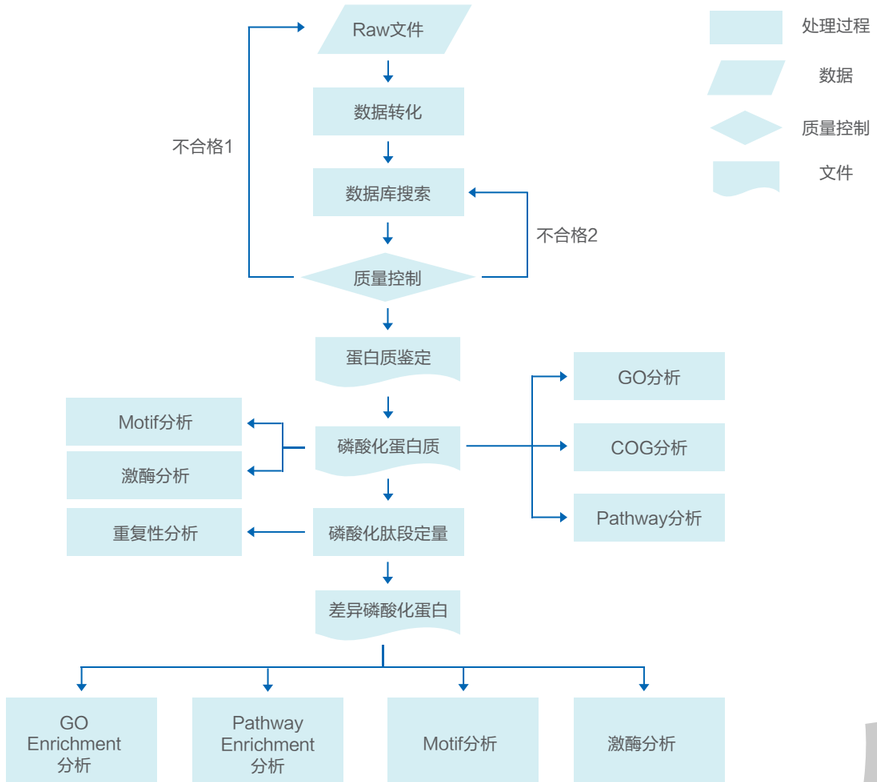


图7 磷酸化蛋白质定量信息分析流程图

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。

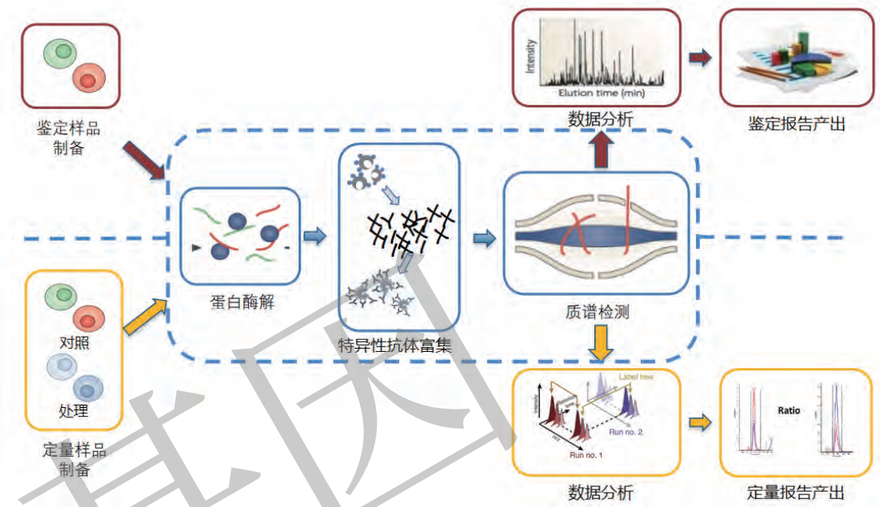


图8 乙酰化修饰分析技术路线

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。

• 信息分析流程

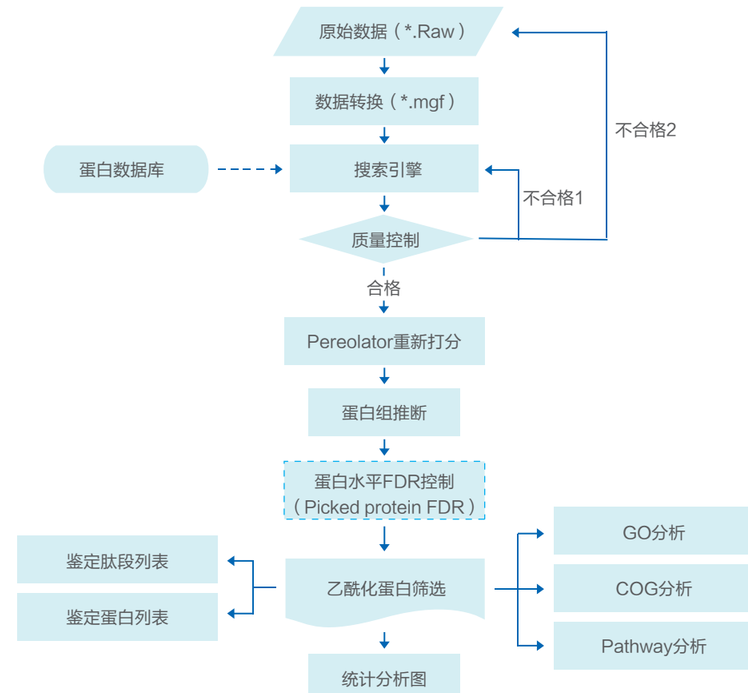


图9 乙酰化修饰鉴定产品信息分析流程

BGI 华大基因

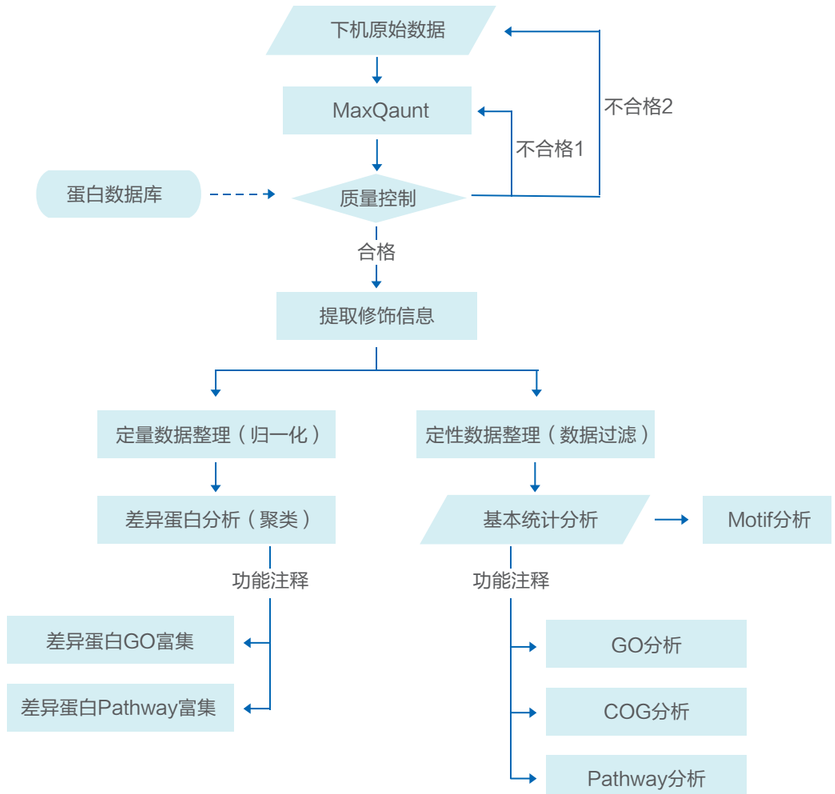


图10 乙酰化修饰定量产品信息分析流程

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、肼化学富集法、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
肼化学法	反应专一性高、无位点偏向性、共价腙键稳定	反应步骤多，条件不易控制，且操作较复杂，还有耗时较长，容易破坏了糖链的结构
免疫法	专一性	缺乏通用性

华大科技质谱平台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11）：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糖基化肽段，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PNGase F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。

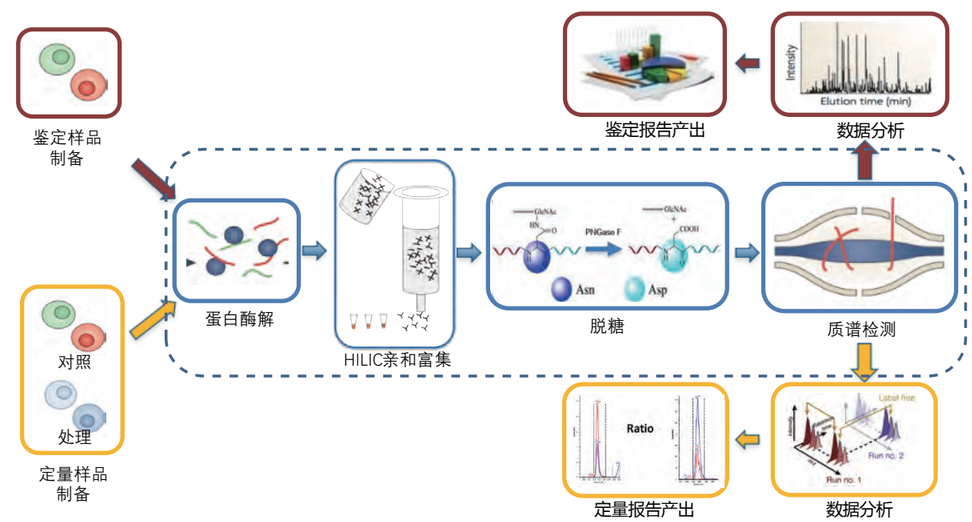


图11 N-糖基化修饰分析技术流程

BGI 华大基因

• 信息分析流程

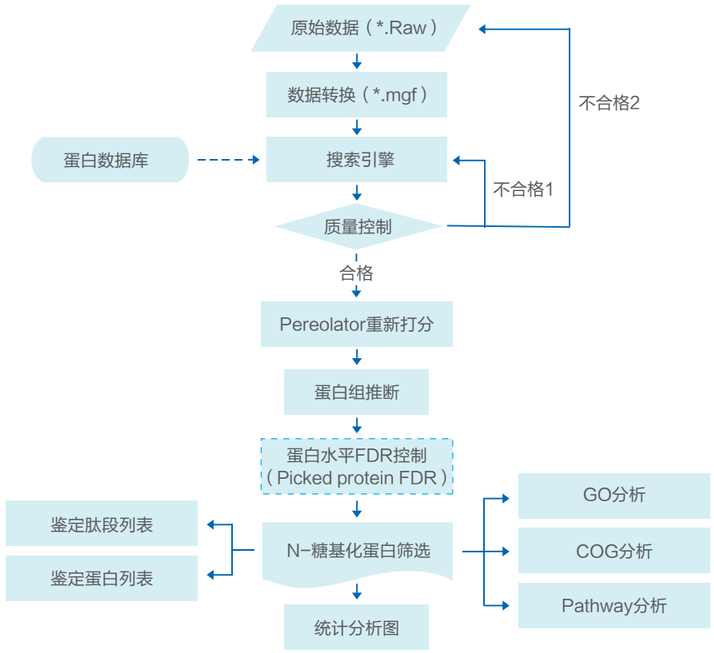


图12 N-糖基化修饰鉴定信息分析流程

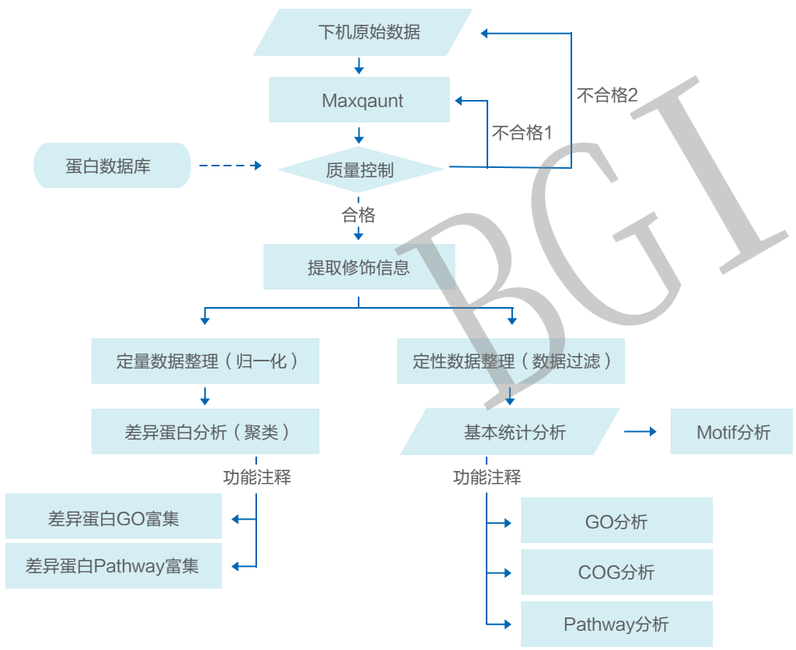


图13 N-糖基化修饰定量信息分析流程

• 技术参数

产品类型		仪器平台	富集方法	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	20个工作日
	磷酸化蛋白全谱鉴定		TiO_2富集	20个工作日
	磷酸化蛋白iTRAQ定量		TiO_2富集	23个工作日
	目标磷酸化蛋白PRM定量		TiO_2富集	33个工作日
乙酰化蛋白修饰	乙酰化蛋白全谱鉴定		抗体富集	26个工作日
	乙酰化蛋白定量		抗体富集	30个工作日
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	25个工作日
	N-糖基化蛋白修饰定量		HILIC亲和富集	29个工作日

典型结果展示

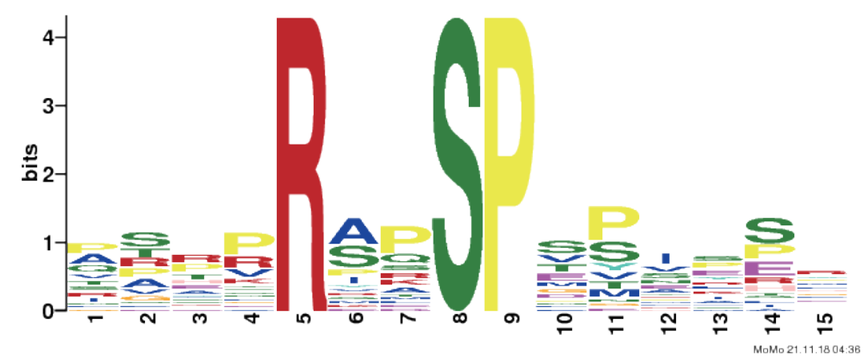


图14 修饰位点motif分布

BGI 华大基因

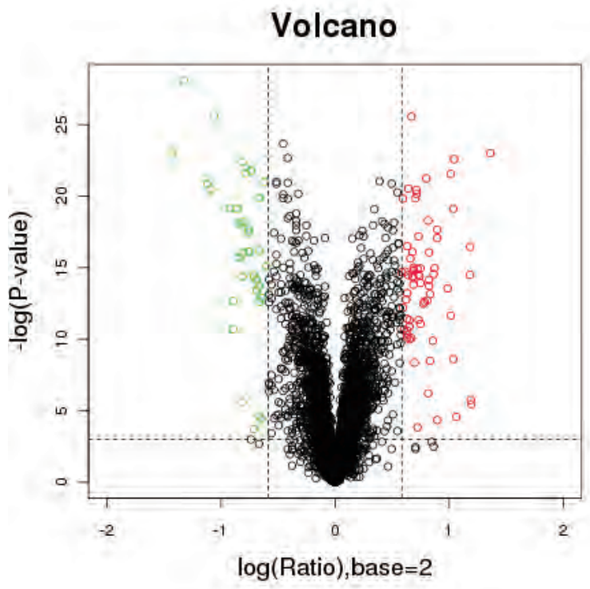


图15 修饰肽段定量火山图

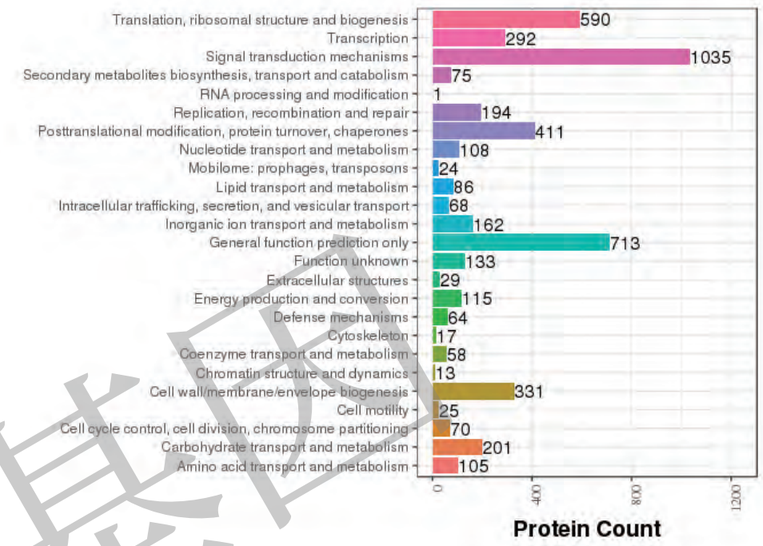


图17 COG注释柱图

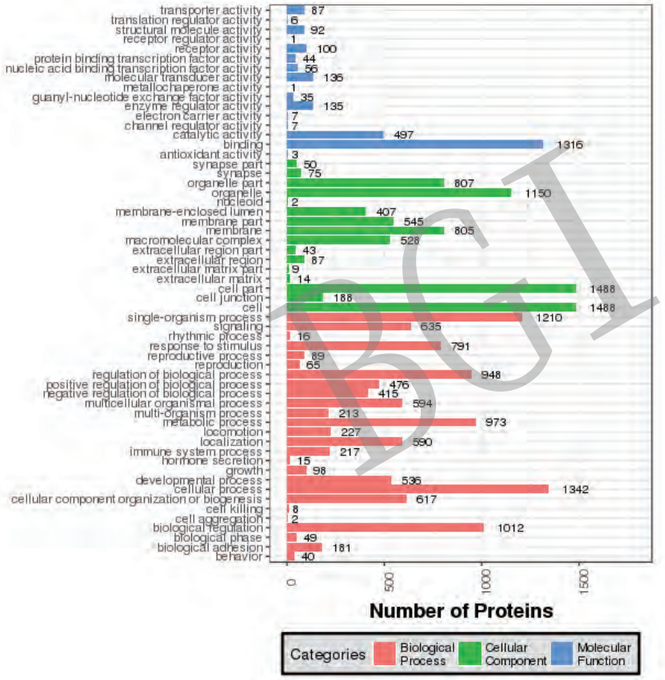


图16 GO功能注释图

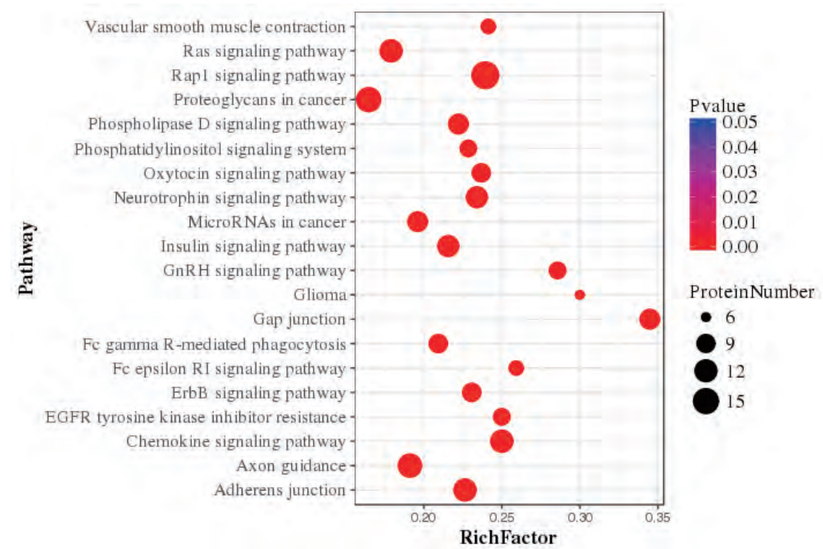


图18 显著富集的Pathway统计图

BGI 华大基因

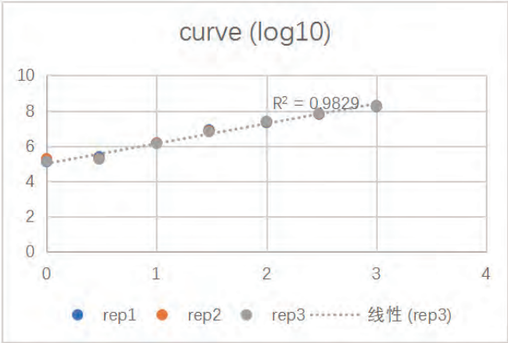


图19 磷酸化PRM定量线性展示图

应用案例解读

- **典型案例节选**

文献名称	研究策略	发表时间	发表期刊
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	Molecular & Cellular Proteomics
Targeted quantitative kinome analysis identifies PRPS2 as a promoter for colorectal cancer metastasis	目标磷酸化蛋白PRM定量	2019	J Proteome Res
Proteomic analysis reveals that lysine acetylation mediates the effect of antemortem stresee on postmortem meat quantility	乙酰化蛋白定量	2019	Food 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	Analytical Chemistry
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
CoA synthase regulates mitotic fidelity via CBP-mediated acetylation	乙酰化蛋白定量	2018	Nat Commun
N-glycoproteomic analysis of chicken egg yolks	N-糖基化蛋白定量	2018	J Agric Food Chem
Endoplasmic reticulum-associated N-glycan degradation of cold-upregulated glycoproteinsin respons to chilling stress in Arabidopsis	N-糖基化蛋白定量	2016	New Phytol

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

Studies on mechanisms of low fertility using proteome and phosphorylated proteome

研究内容：小鼠睾丸的蛋白质组学和磷酸化蛋白质组学　研究策略：iTRAQ+磷酸化蛋白质组　发表时间：2019年
发表期刊：Molecular & Cellular Proteomics

- **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，其表现包括睾丸较小、生育能力低下和精子发生能力缺陷等等。然而，目前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℃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。

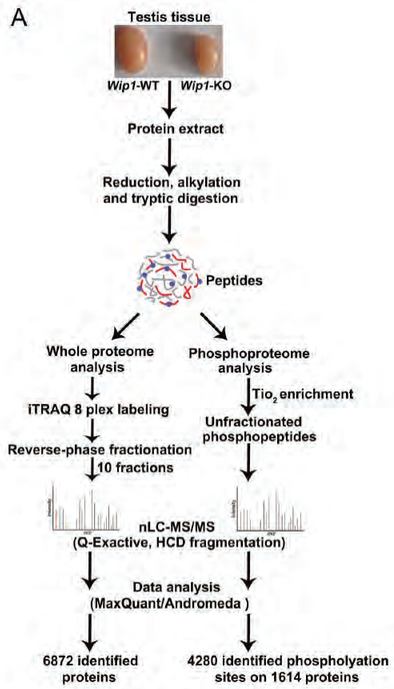


图1 主要实验流程

BGI 华大基因

• 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。

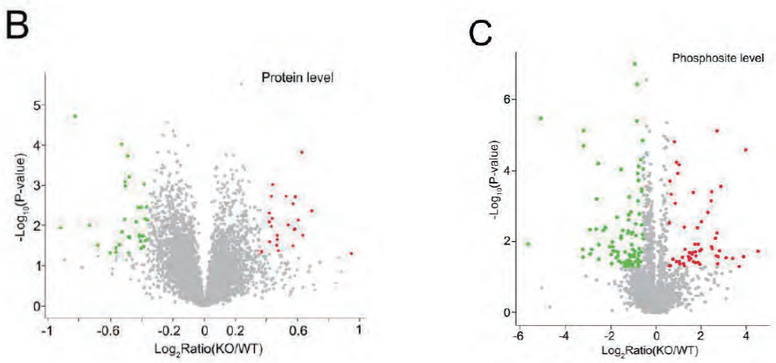


图2 B. 所有生物重复中的定量蛋白质的火山图

C. TiO_2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。

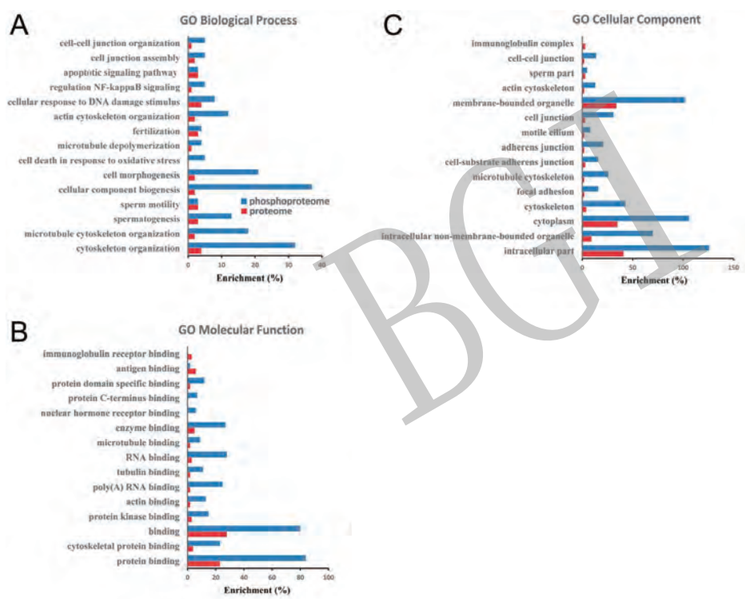


图3 Wip1缺陷的蛋白质组和磷酸化蛋白质组的整合分析鉴定细胞粘连/紧密连接通路

A-C. 显著差异蛋白和显著变化的磷酸化水平的GO功能富集A. 生物过程分析;

B. 分子功能分析; C. 细胞成分分析；红色和蓝色列分别代表蛋白质组和磷酸化蛋白质组。

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。

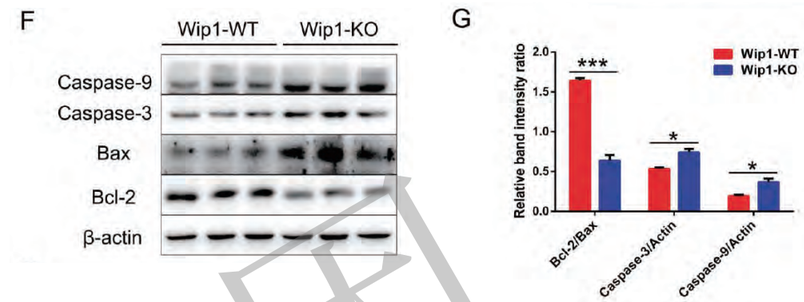


图4 免疫印迹验证Wip1缺陷小鼠在睾丸中升高的生殖细胞凋亡水平

F. WT和Wip1-KO小鼠睾丸中凋亡相关组分的相对蛋白质表达水平，β-肌动蛋白作为上样对照。

G. 通过条带强度的光密度扫描计算Bcl-2 / Bax，半胱天冬酶-3 /肌动蛋白和半胱天冬酶-9 /肌动蛋白的相对表达水平。

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

Proteomic analysis reveals that lysine acetylation mediates the effect of antemortem stress on postmortem meat quantility development

研究内容：影响肉质的乙酰化蛋白质组　研究策略：乙酰化蛋白定量　发表时间：2019年　发表期刊：Food Chemistry

• 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。

BGI 华大基因

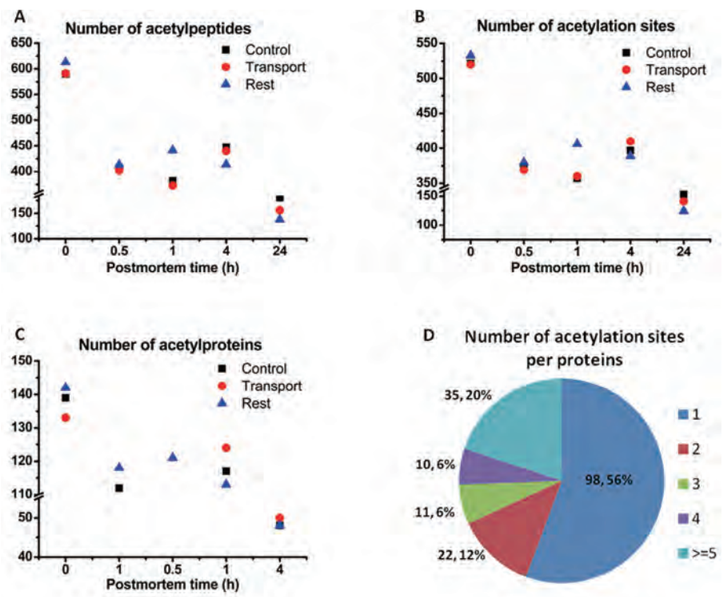


图1 乙酰化蛋白鉴定情况

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来调节肌肉僵硬，和通过改变热休克蛋白和通透性转换孔复合物的功能来调控压力应激、细胞凋亡和肉嫩化。本研究首次报道了屠宰后肌肉乙酰化受到预处理的影响，并拓宽了生物化学调节肉质发展的知识。

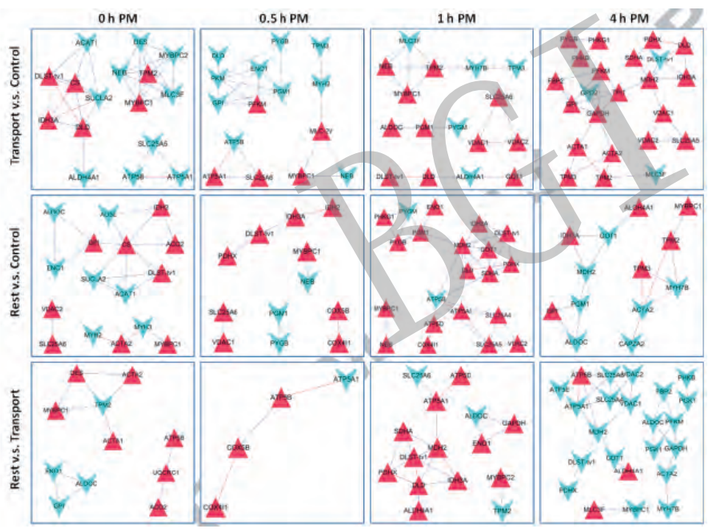


图2 差异乙酰化蛋白互作分析

案例三

鸡蛋黄的N-糖蛋白组学分析

N-glycoproteomic analysis of chicken egg yolks

研究内容：鸡蛋黄的N-糖蛋白质组学　研究策略：N-糖蛋白质组　发表时间：2018年　发表期刊：J Agric Food Chem

• 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

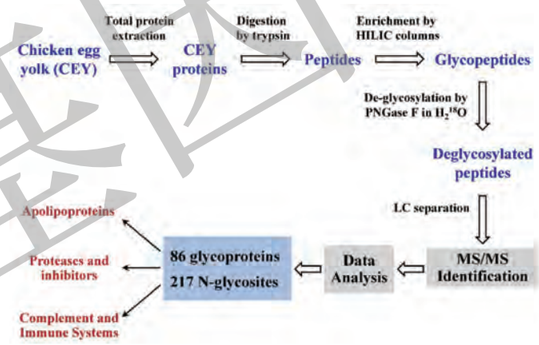


图1 实验设计

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。

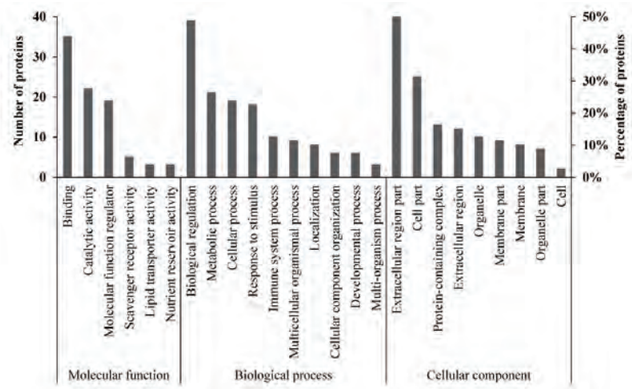


图2 糖蛋白功能注释

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。

BGI 华大基因

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内，是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。

应用方向

1、新型疾病biomarker发现，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。如图1所示。
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。

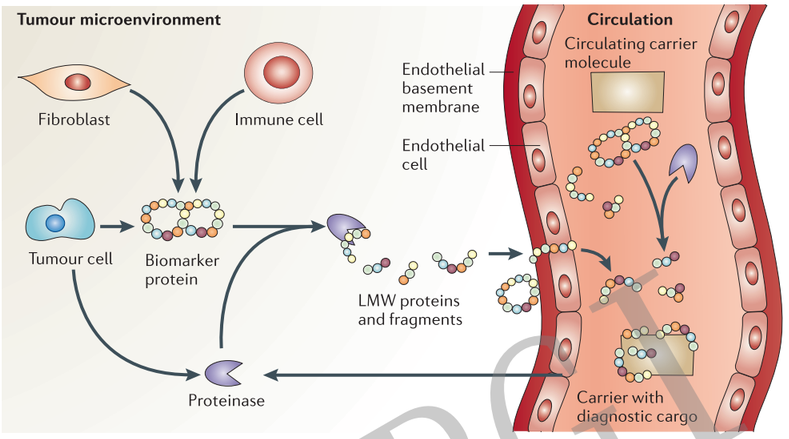


图1 疾病（肿瘤等）多肽在血浆中的特定存在形式
Nature Reviews Cancer 6, 961-967(December 2006) | doi:1038/nrc2011

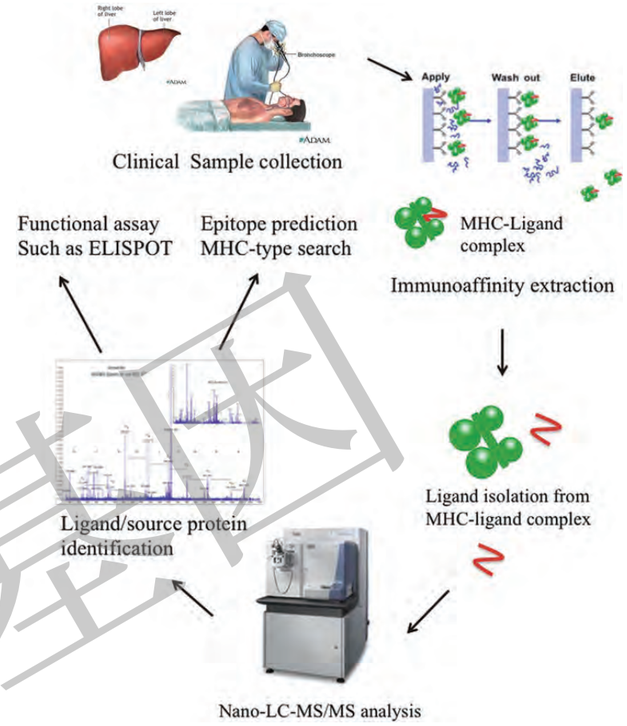


图2 肿瘤特异性抗原（MHC结合多肽）的鉴定与肿瘤免疫多肽治疗的应用
Methods Mol Biol.2013;203-18.doi:10.1007/978-1-4614-7209-4_14

技术路线

生物样品 → 内源性多肽组分离及富集 → LC-MS/MS → 生物信息分析 < 多肽鉴定与定量 / 生物功能分析

产品优势

1. 高通量

可同时对样品中的大量多肽进行综合研究。

2. 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。

3. 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。

BGI 华大基因

结果示例

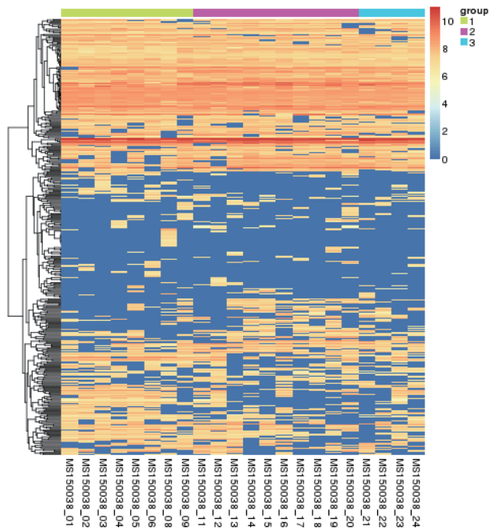


图3 多肽组定量聚类热图

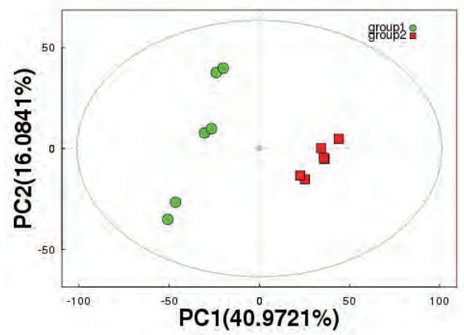


图4 PCA聚类分析

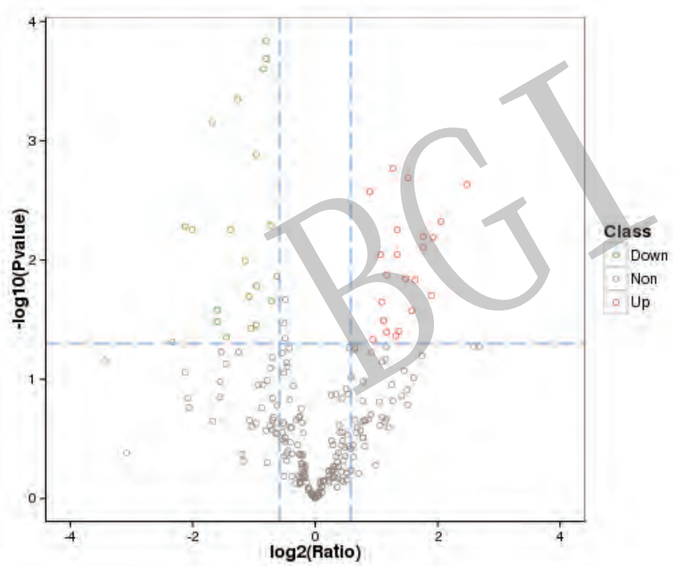


图5 定量差异多肽分布火山图

03 代谢组学产品

非靶向代谢组学

产品介绍

典型结果展示

应用案例解读

脂质组学

产品介绍

典型结果展示

应用案例解读

靶向代谢组学

产品介绍

典型结果展示

应用案例解读

华大基因 BGI

代谢组学

代谢组学(Metabonomics/Metabolomics)是继基因组学和蛋白质组学之后发展起来的新兴学科。其主要目标是定性定量地研究生命体对外界刺激、生理病理变化以及基因突变而产生的体内代谢物水平的多元动态反应，研究对象大都是相对分子质量1000 Da以内的小分子物质，如有机酸、氨基酸、核苷酸、糖、脂质、维生素等。代谢组学研究主要基于核磁共振技术、色谱-质谱联用技术采集样品的代谢物信息，比较不同组别样品中代谢产物的含量，寻找差异代谢物，并探索差异代谢物参与的代谢通路，广泛应用于生物标志物的发现、中医药研究、健康评估、疾病诊断、环境监测、分子机制的揭示等研究。根据研究目的不同，代谢组学主要分为非靶向代谢组学、脂质组学、靶向代谢组学等。

依托在基因组、转录组和蛋白质组研究领域取得的众多成果和顶尖仪器平台、高端专业人才，华大在代谢组学研究方面具有充分的优势，也取得了瞩目的成就。我们致力于帮助合作伙伴在整个生命科学和系统生物学研究中实现更好的贯穿、整合与互补。

代谢组学研究工作流程
（含严格质控体系 - 数据质量保证）

图1 代谢组学工作流程图

代谢组学研究采用的检测技术主要有NMR、GC-MS、LC-MS三种，三种不同检测技术的差异比较如下：

表1 三种代谢组学检测技术的比较

仪器类型	色谱质谱联用		核磁共振NMR
	GC-MS	LC-MS	
适用范围	易挥发、热稳定、中低极性物质	涵盖大部分化合物，特别适用于热不稳定，不易挥发，不易衍生化等物质	没有偏向性，适用于体液分析研究，活体和原位研究
样品前处理	需要衍生化，比较繁琐	较简单	简单
适用范围	高	高	较低，动态范围有限
化合物定性	好，有参考比对的标准谱图库	数据库不完善	好，有完整数据库重现性高

代谢组研究中，一般会用到两种分析策略：非靶向代谢组学（即代谢全谱分析）和靶向代谢组学（即目标小分子定量分析）。首先采用非靶向代谢组学无偏向性地对样本中所有的代谢物进行检测分析，筛选出差异代谢物，进行差异代谢物的鉴定，使用标准品对鉴定结果进行确认，完成生物学意义的阐释。再通过靶向代谢组学对有意义的代谢物进行验证分析，为生物标志物的深入研究和开发利用提供有力支持。

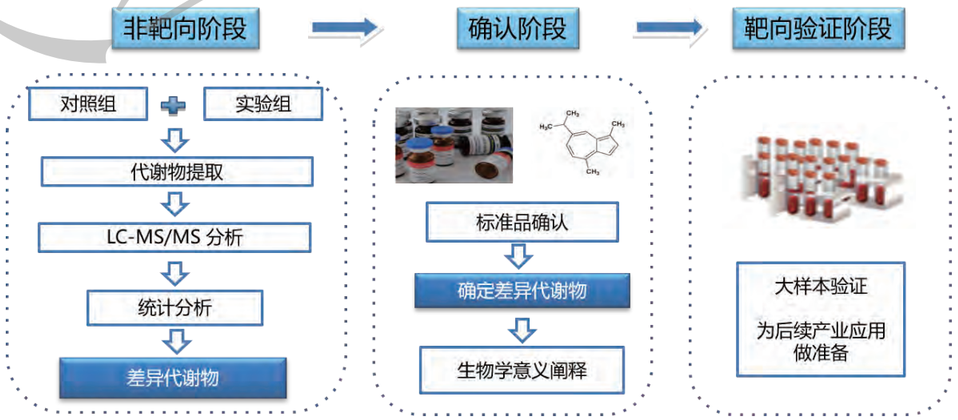


图2 代谢组学整体研究策略

BGI 华大基因

非靶向代谢组学

产品介绍

华大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相关关系。

• 技术流程

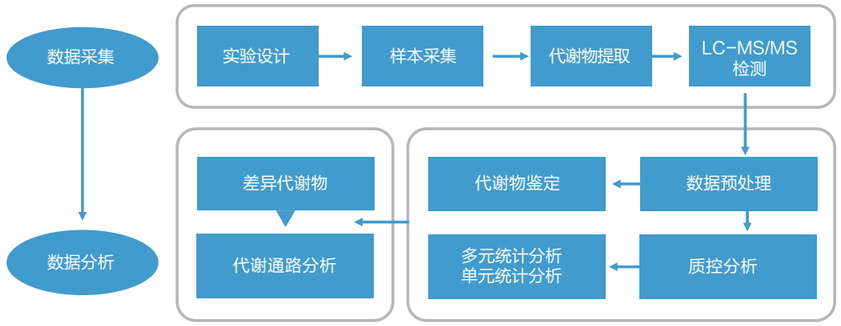


图1 技术流程图

• 信息分析流程

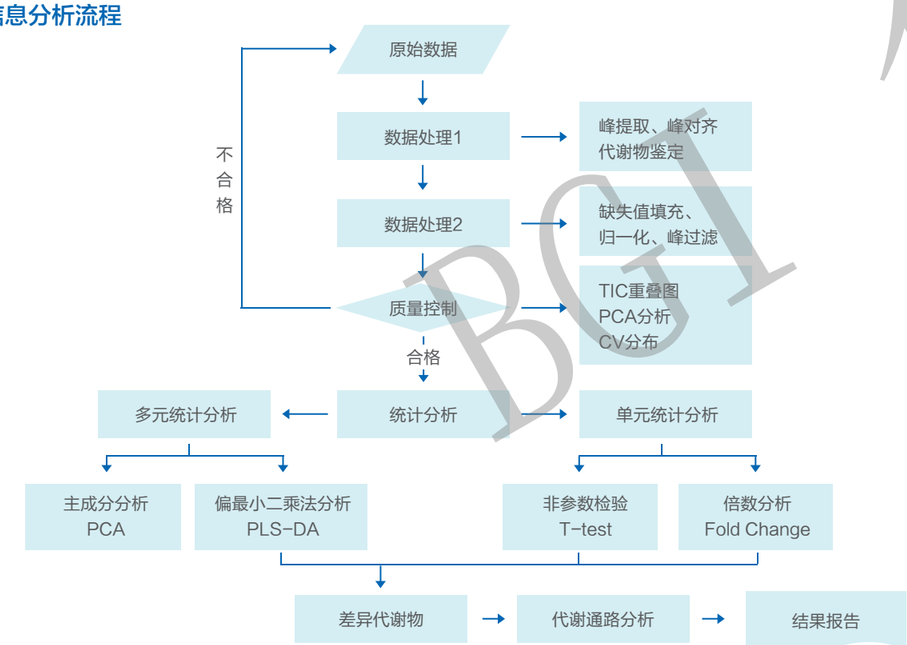


图2 信息分析流程图

• 技术参数

表1 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	25个工作日
	Waters Xevo G2-XS Q-Tof	Progenesis QI	25个工作日

• 生物学重复次数要求

表2 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

典型结果展示

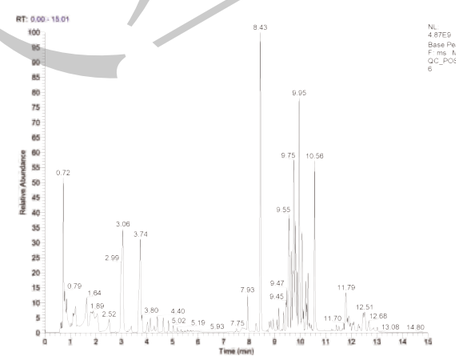


图3 正离子模式的基峰色谱图

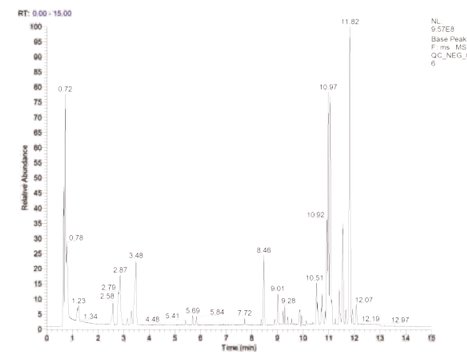


图4 负离子模式的基峰色谱图

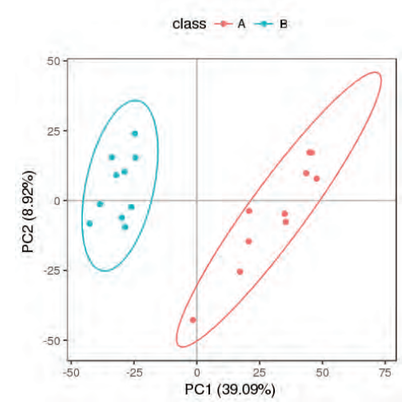


图5 PCA分析得分图

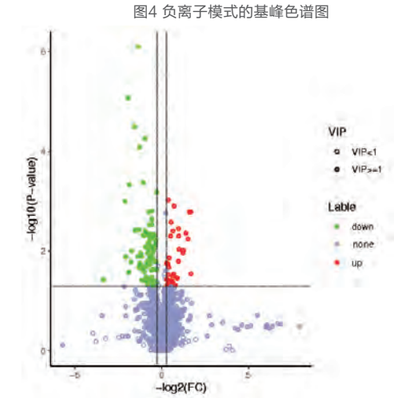


图6 火山图

BGI 华大基因

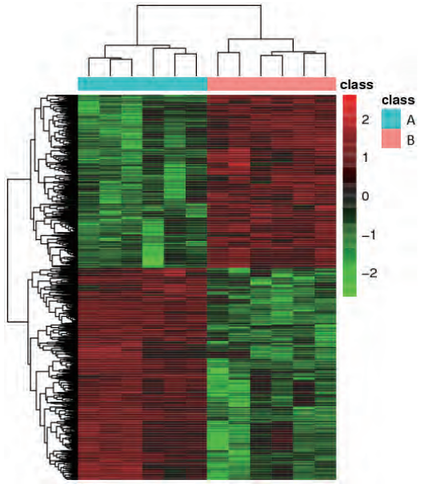


图7 层次聚类分析图

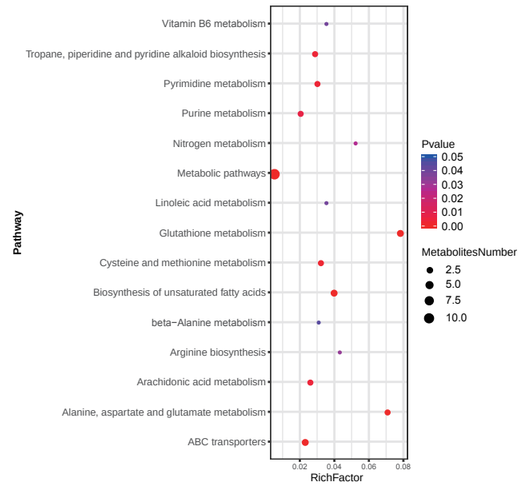


图8 代谢通路富集分析气泡图

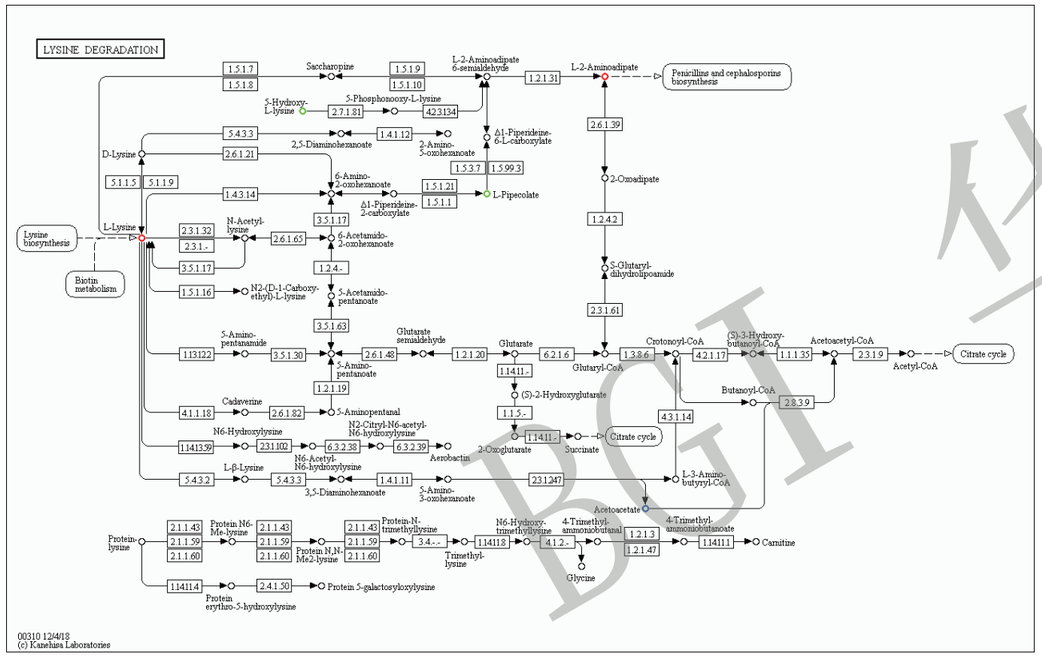


图9 代谢通路信号图

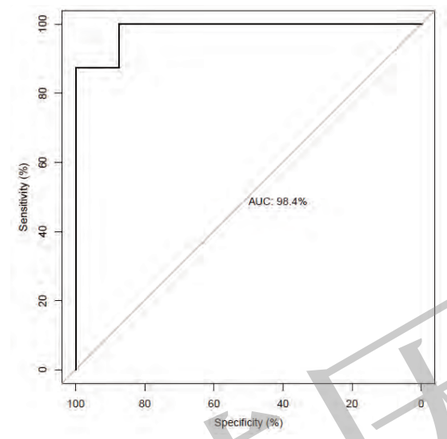


图10 ROC曲线图

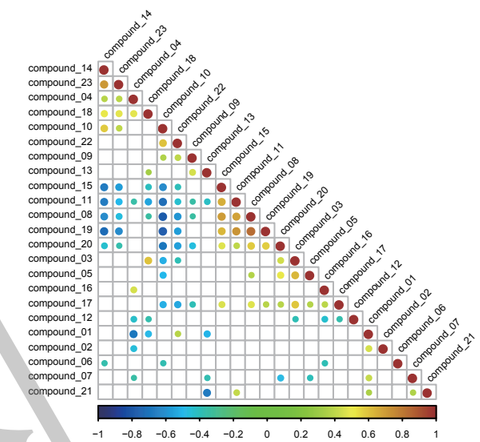


图11 相关性分析图

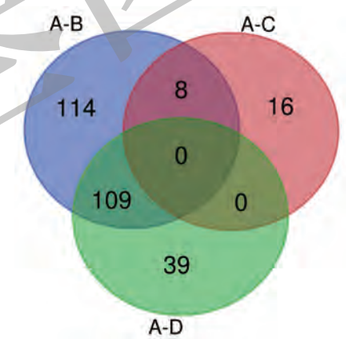


图12 Venn图

应用案例解读

- **典型案例汇总**

文献名称	研究策略	发表时间	发表期刊
Metabolite changes in blood predict the onset of tuberculosis	非靶向代谢组学	2018	Nature Communications
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 Medicine
Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease	非靶向代谢组学	2016	Scientific Reports

BGI 华大基因

案例

结核病生物标志物的发现

Metabolite changes in blood predict the onset of tuberculosis

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　发表时间：2018年

- **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TB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，容易使一些健康个体处于不必要的风险中。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。

- **技术路线**

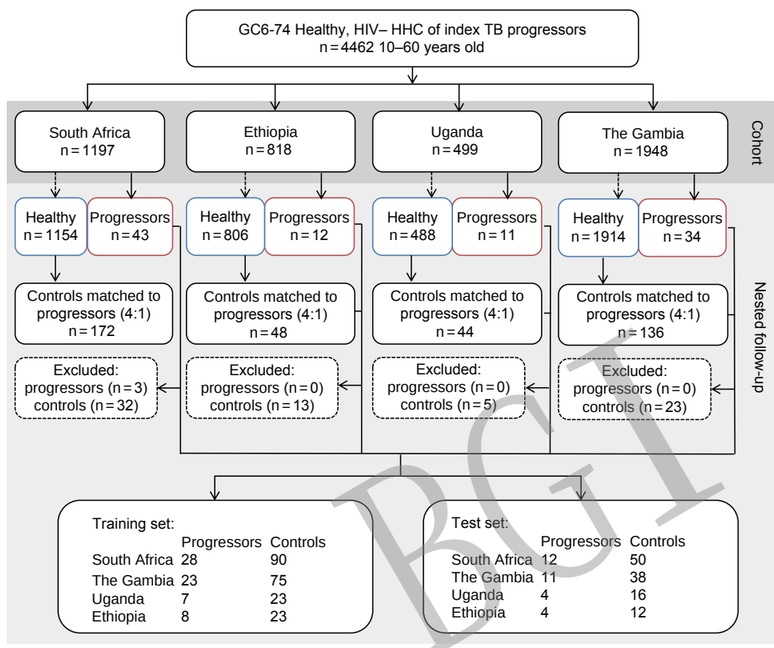


图1 技术路线

- **研究结果**

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4462例HIV阴性结核病家庭系，发现活性TB前12个月皮质醇和犬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记物。

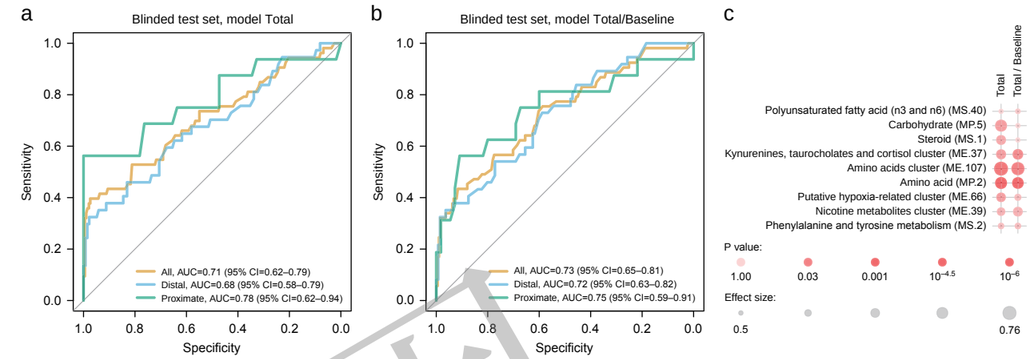


图2 通过模型确定不同组别中代谢物的差异

脂质组学

产品介绍

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，LIPID MAPS）将脂质分为八大类，包括：脂肪酰类(Fatty Acyls)、甘油脂类(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**

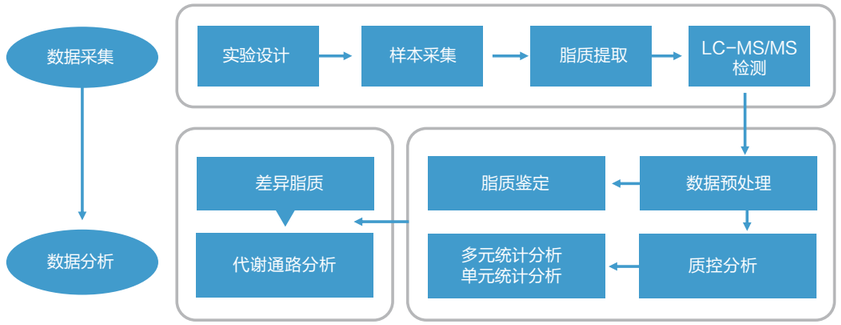


图1 技术流程图

BGI 华大基因

• 信息分析流程

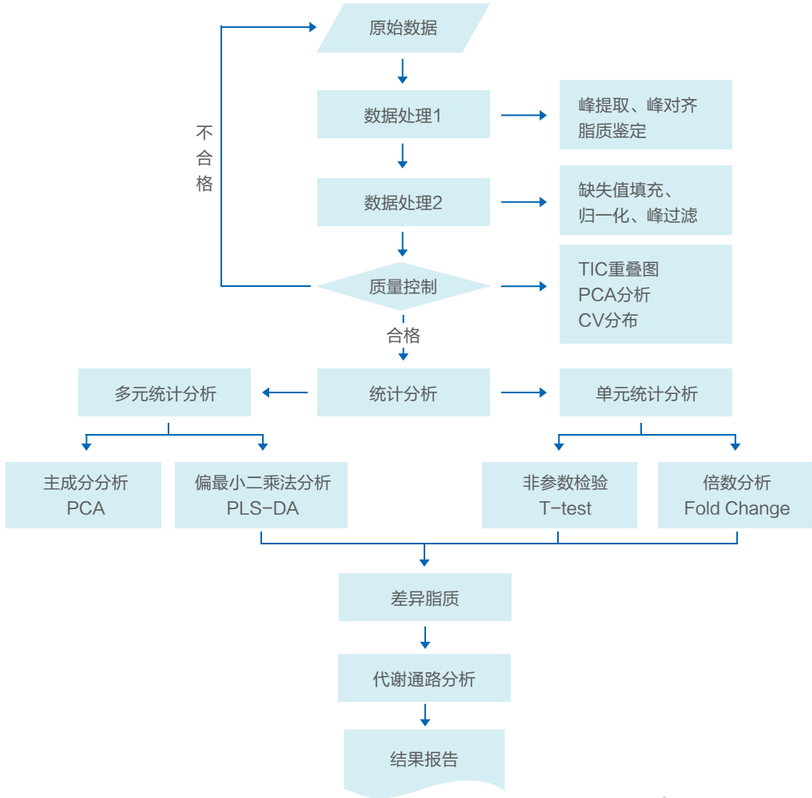


图2 信息分析流程图

• 技术参数

表1 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	25个工作日
	Waters Xevo G2-XS Q-Tof	Progenesis QI	25个工作日

• 生物学重复次数要求

表2 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

典型结果展示

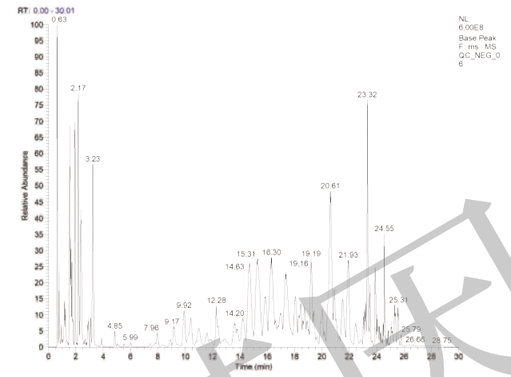


图3 正离子模式的基峰色谱图

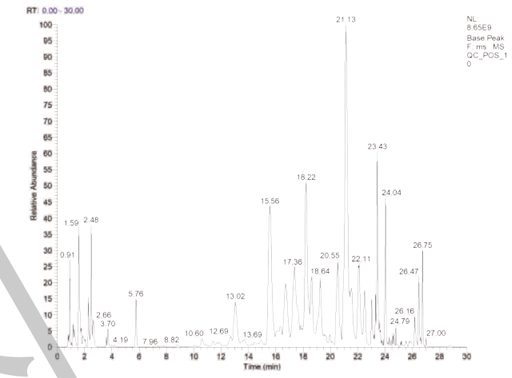


图4 负离子模式的基峰色谱图

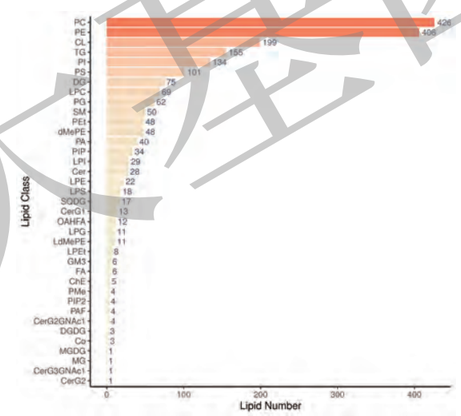


图5 脂质亚类和对应的分子数量统计图

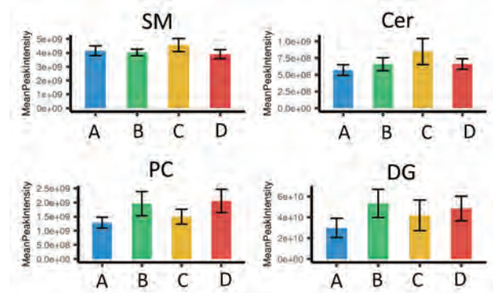


图6 脂质亚类(部分)的相对强度分析

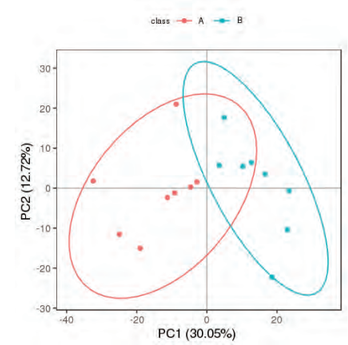


图7 PCA分析得分图

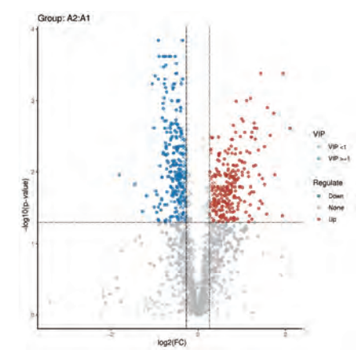


图8 火山图

BGI 华大基因

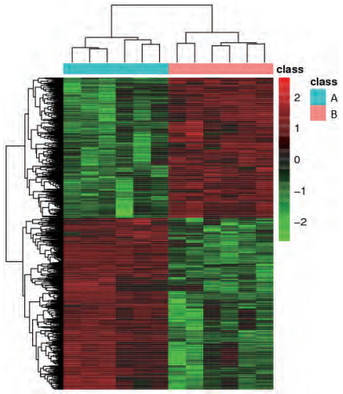


图9 层次聚类分析图

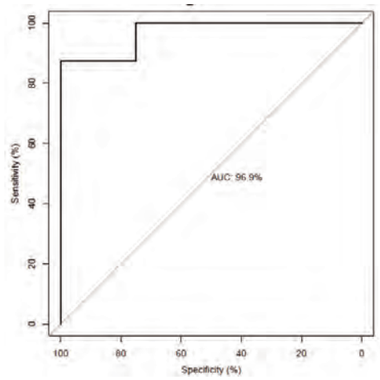


图10 ROC曲线图

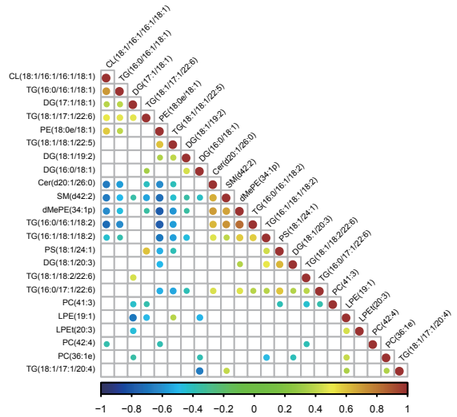


图11 相关性分析图

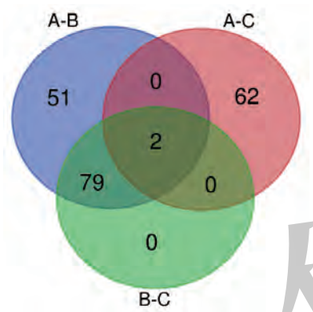


图12 Venn图

应用案例解读

- **典型案例汇总**

文献名称	研究策略	发表时间	发表期刊
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	2017	GigaScience
Plasma lipidomic profiles improve upon traditional risk factors for the prediction of cardiovascular events in type 2 diabetes	脂质组学	2016	Circulation

案例

首个中国人群糖尿病相关血浆脂质组学研究

Lipidomic profiling reveals distinct differences in plasma lipid composition in healthy, prediabetic and type 2 diabetic individuals

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　　发表时间：2017年

- **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、糖尿病前期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了参考。

- **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

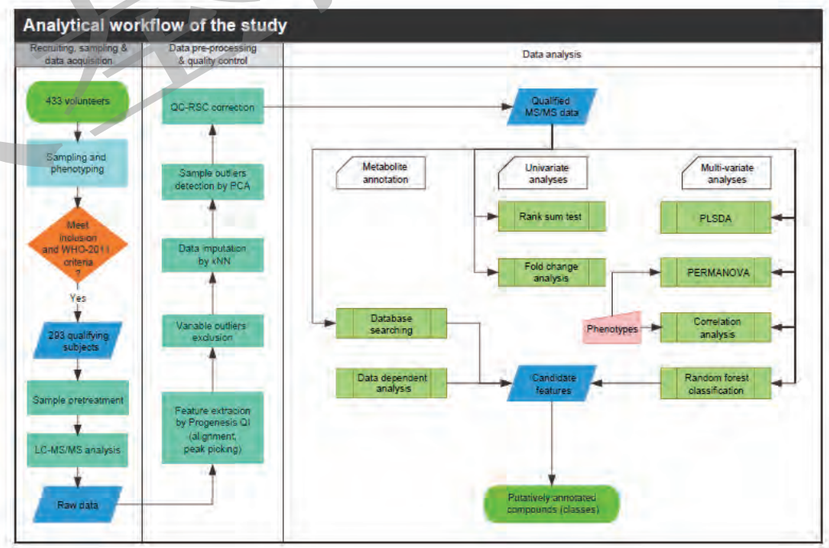


图1 技术路线

- **研究结果**

1、脂质组学数据和几个血浆参数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

BGI 华大基因

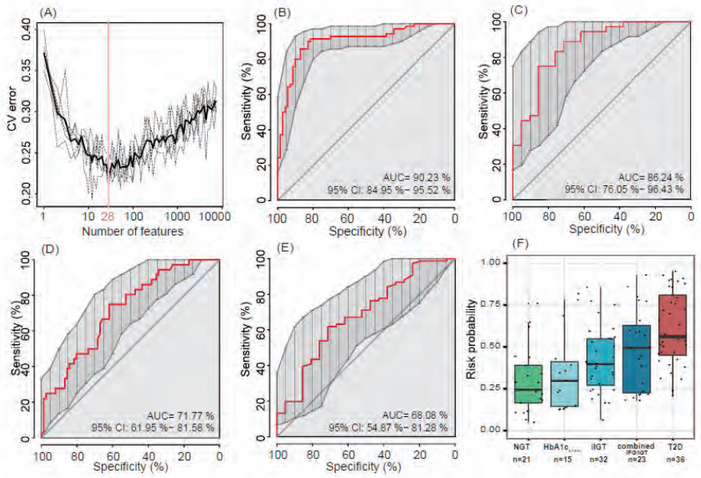


图2 代谢物持续变化情况图

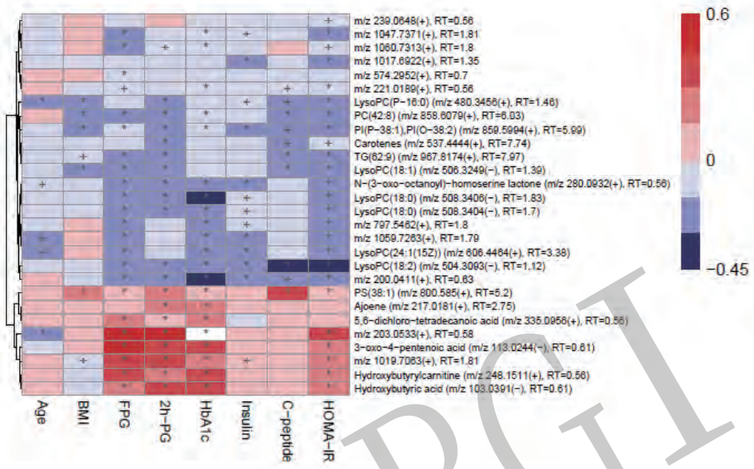


图3 一般线性模式对选择出的28个离子和表型进行相关性分析

• 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biomarker。

靶向代谢组学

产品介绍

靶向代谢组学不同于非靶向代谢组学全面地尽可能多地检测代谢物，靶向代谢组学是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，可以更准确的监测代谢的动态变化过程，阐明相关的代谢机制，更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、自动化高通量等突出优点。

• 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数据获取方式。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，提高灵敏度和方法的专属性。

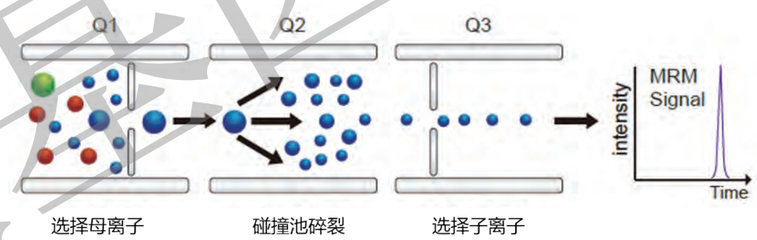


图1 MRM质谱技术的工作原理示意图

• 技术特点

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；

标准品：通过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，每个代谢物的特征离子对一次性建立，长期适用；

重现性好：多重保障（每个代谢物多个特征离子对，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

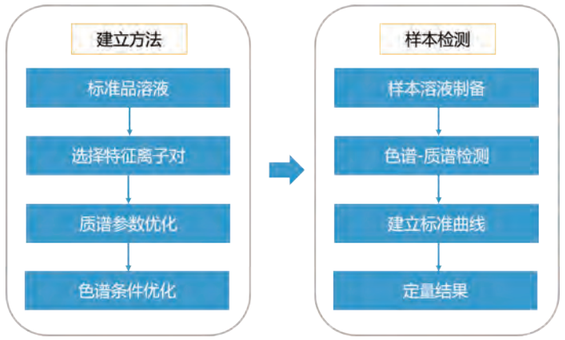


图2 技术流程图

BGI 华大基因

• 信息分析流程

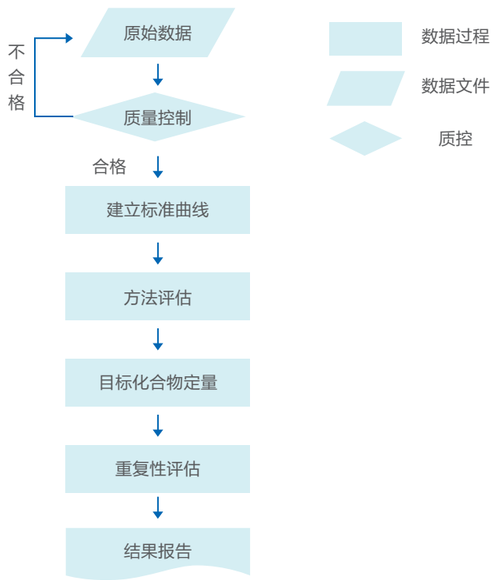


图3 信息分析流程图

• 技术参数

表1 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	25-35个工作日
GC-MS	Thermo TSQ 9000，Agilent 7890B-5977A	25-35个工作日

• 生物学重复次数要求

表2 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

• 已建立方法的靶向代谢组学检测类别

表3 靶向代谢组学检测类别

类别	检测仪器	化合物数目	代表性化合物
短链脂肪酸	GC-MS	7	乙酸、丙酸、丁酸、戊酸、异丁酸、异戊酸、己酸
有机酸	GC-MS	9	丙酮酸、乳酸、马来酸、琥珀酸、戊二酸、苹果酸、柠檬酸等
尿液代谢物相对定量	GC-MS	132	2-丙基-3-羟基戊酸、2-己烯酸、2-酮己二酸等
氨基酸	LC-MS	39	精氨酸、组氨酸、异亮氨酸、亮氨酸、赖氨酸、甲硫氨酸等
维生素	LC-MS	12	维生素B1、吡哆酸、5-甲基四氢叶酸；维生素A（视黄醇）、α-生育酚（维生素E）等
类固醇激素	LC-MS	13	睾酮、雄烯二酮、雌二醇、孕酮、11-脱氧皮质醇、可的松等
儿茶酚胺及其代谢物	LC-MS	7	多巴胺、去甲肾上腺素、肾上腺素、高香草酸、香草扁桃酸等
胆汁酸	LC-MS	50	胆酸、鹅脱氧胆酸、甘氨胆酸、甘氨鹅脱氧胆酸、牛磺胆酸等
氧化三甲胺及其相关代谢物	LC-MS	5	三甲胺、氧化三甲胺、甜菜碱、胆碱、左旋肉碱
磷脂酰胆碱及鞘脂的相对定量	LC-MS	226	LPC(12:0)、LPC(14:0)、PC(24:0)、PC(25:0)、SM(30:1)等
酰基肉碱相对定量	LC-MS	40	游离肉碱、乙酰肉碱、丙酰基肉碱、丙二酰肉碱、丁酰基肉碱等
神经递质	LC-MS	26	5-羟色胺、变肾上腺素、高香草酸、γ-氨基丁酸、左旋多巴、褪黑素等
花青素相对定量	LC-MS	45	芍药花青素 O-己糖苷、矢车菊属 3-O-丙二酰己糖苷、花翠素 O-丙二酰己糖苷、锦葵色素等
植物激素	LC-MS	25	吲哚-3-乙酸（IAA）、反式玉米素（tZ）、茉莉酸甲酯（MEJA）、茉莉酸（JA）、水杨酸（SA）、脱落酸（ABA）等
类胡萝卜素（绝对定量/相对定量）	LC-MS	17	α-胡萝卜素、叶黄素、α-隐黄质、番茄红素、虾青素、辣椒红素等
氧化脂质	LC-MS	64	花生四烯酸（AA）、亚油酸（LA）、α-亚麻酸（ALA）、DHA、EPA、DGLA等多不饱和脂肪酸的氧化代谢产物。

• 其他已建立方法的靶向定量检测类别

表4 靶向定量检测类别

类别	检测仪器	化合物数目	代表性化合物
核糖核苷酸（含甲基化）	LC-MS	16	dA、dG、T、dC、5-hmdc、m6dA、A、C、G、U、Ψ、m1Ψ、m3U、m5U、m5Um、m5C
免疫抑制剂	LC-MS	4	他克莫司、环孢霉素A、依维莫司、西罗莫司
抗生素	LC-MS	40	万古霉素、阿莫西林、氨苄西林、奥比沙星、二甲硝唑、环丙沙星等
微量元素与重金属	ICP-MS	15	镁Mg、锶Sr、铜Cu、锌Zn、铬Cr、锰Mn、硒Se、碘I、铅Pb、砷As、镉Cd、铁Fe、汞Hg、铊Tl、钴Co

BGI 华大基因

典型结果展示

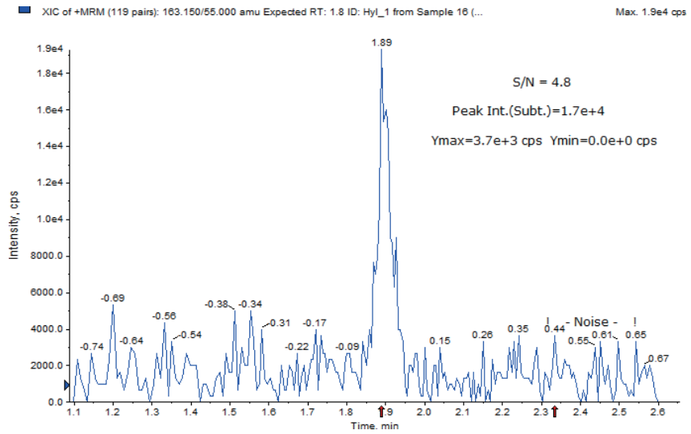


图4 LOD（检出限）图

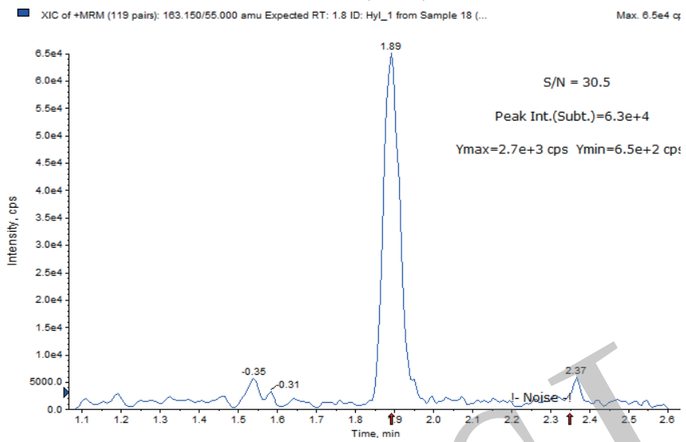


图5 LOQ（定量限）图

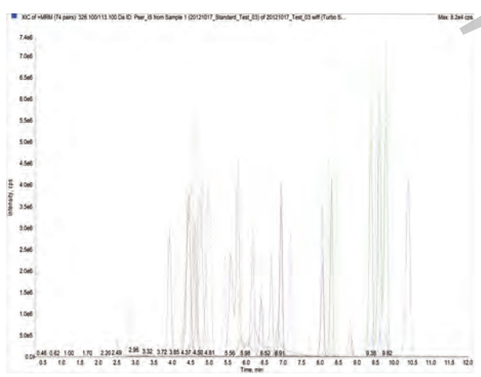


图6 代谢物色谱图

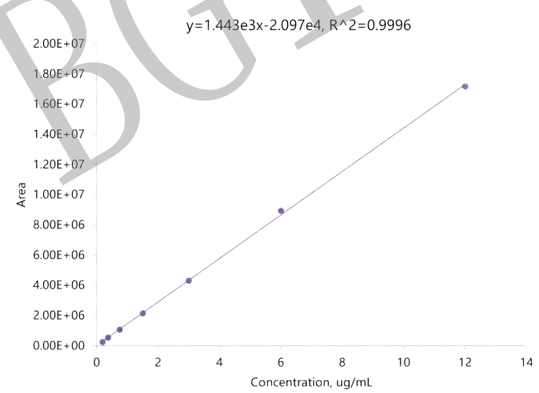


图7 标准曲线图

应用案例解读

- **典型案例汇总**

文献名称	研究策略	发表时间	发表期刊
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 Medicine
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	2017	Nature Communications
Functional metabolomics describes the yeast biosynthetic regulome	靶向代谢组学+基因组学	2016	Cell

案例一

功能代谢组学描述酵母的生物合成规律

Functional Metabolomics Describes the Yeast Biosynthetic Regulome

研究内容：酵母的生物合成过程　发表期刊：Cell　研究策略：靶向代谢组学+基因组学　发表时间：2016年

- **研究背景**

氨基酸作为新陈代谢的关键组成部分，但仍缺乏其调控背后遗传网络的全面描述。而且氨基酸的生物合成、稳态和分解代谢在代谢通量中占很大比例，在极性代谢物的总质量中占四分之三。因此，通过对酿酒酵母靶向氨基酸代谢组学及基因敲除后反应的系统性研究，可以绘制出基因组-代谢组相互作用图谱，从而了解细胞生长期间代谢的调节和内稳态生理机制。通过与其代谢表征的关联，提高对未知功能基因的注释潜力。

- **研究思路**

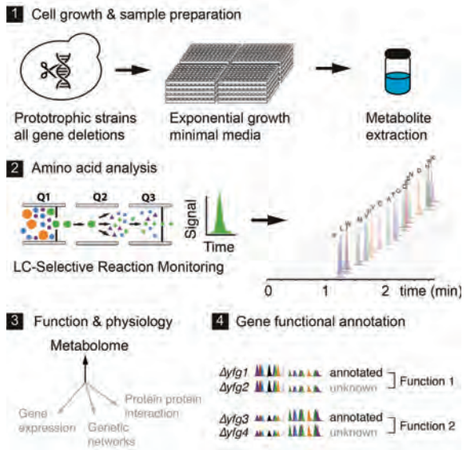


图1 技术路线

BGI 华大基因

• 研究结果

1、研究发现至少1,519个基因对酵母的氨基酸代谢有特异性的影响，即三分之一的编码基因显著影响酵母的生物合成代谢，而且每个氨基酸的浓度独立地受基因缺失地影响。

2、氨基酸代谢与染色质和转运蛋白最为相关。在大多数情况下，由于相关基因的缺失会产生高度特异性的浓度信号，而染色质水平的基因调控、蛋白质翻译和蛋白质运输机制主要影响氨基酸代谢调节网络。

3、TORC1通过囊泡介导转运影响生物合成氨基酸代谢。研究发现以TORC1复合物为中心的激酶网络对氨基酸代谢有强烈的影响。删除TORC1上游激活剂和使用抑制剂以获得下游的转录和代谢数据，在基因组水平上阐明囊泡介导运输机制，确定了在氨基酸代谢过程中，TORC1具有维持处于指数生长期细胞内膜转运功能。

4、功能代谢组学具有较高的功能基因注释潜力。与物理和遗传相互作用网络相比，代谢组学特征提供了与基因相关的正交信息，富集了目前未被鉴定的一半以上的酵母基因，并为其提供功能注释。而编码基因的一个主要功能是参与基因–代谢的相互作用，因此可以阐明整个代谢调控网络，解释未知基因功能。

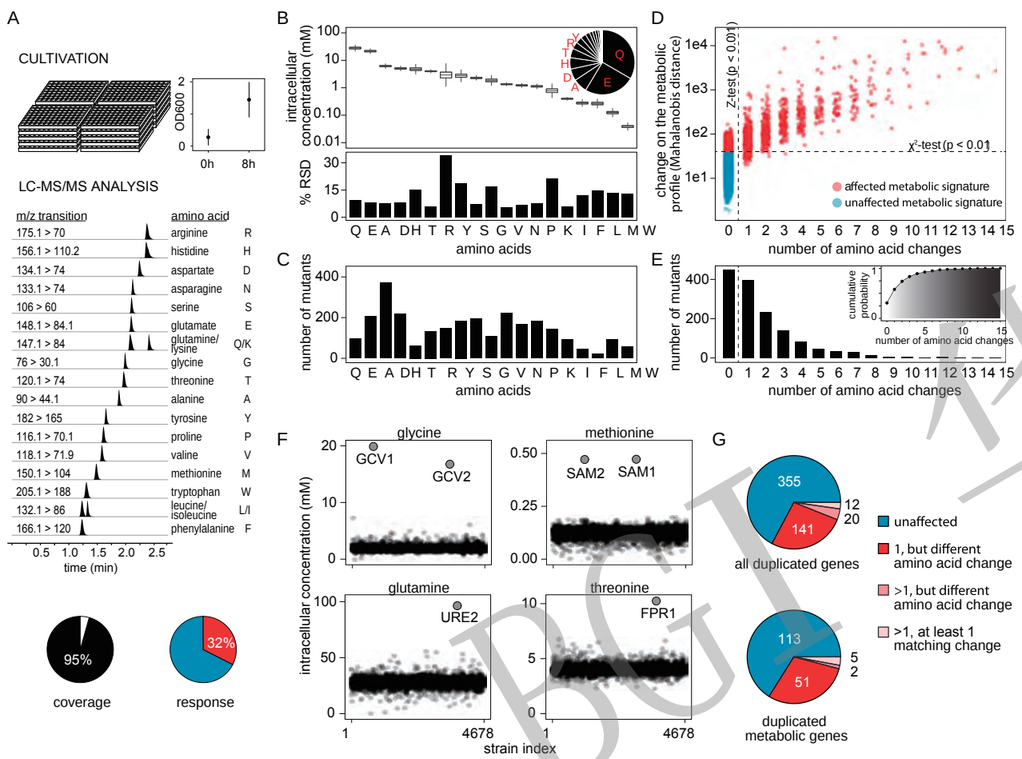


图2 影响氨基酸生物合成代谢的基因组–代谢组相互作用

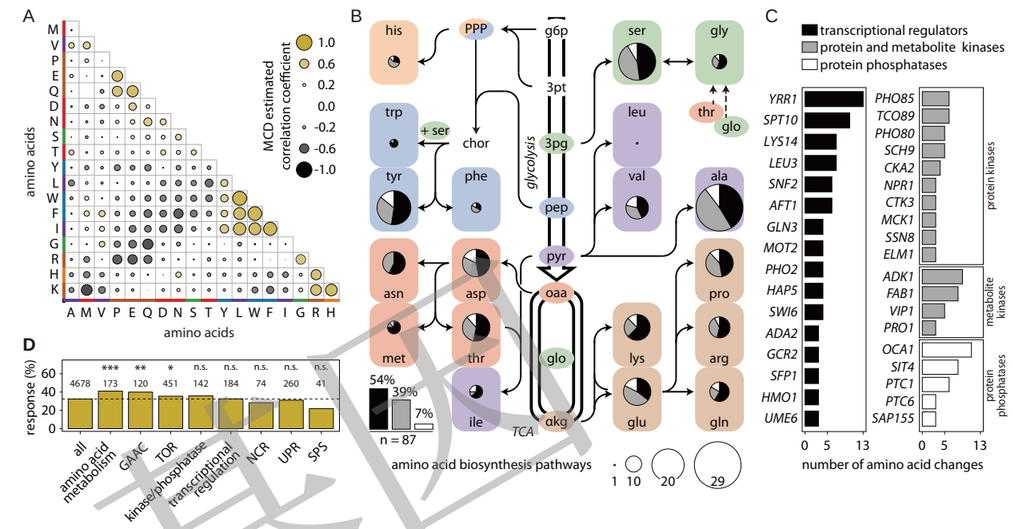


图3 氨基酸浓度与基因组规模相关，并取决于转录和转录后调控机制的遗传成分

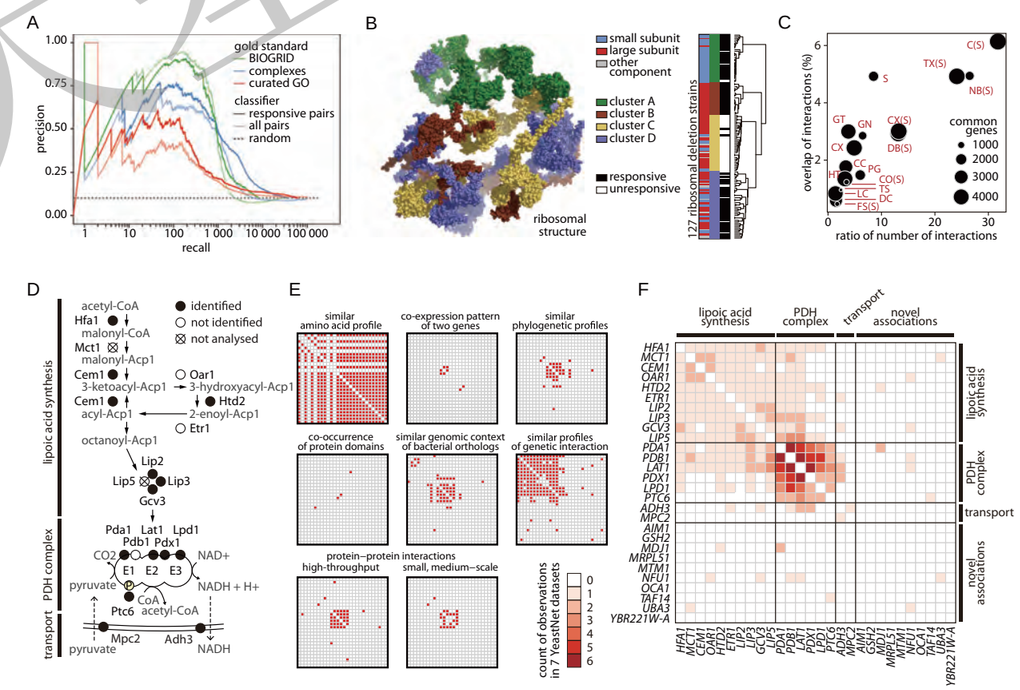


图4 氨基酸特征与基因功能的关系以及其他分子网络的相关性体现

BGI 华大基因

案例二

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

Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment

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　发表时间：2017年

- **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型糖尿病患者进行分型具有重要意义。

- **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。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。

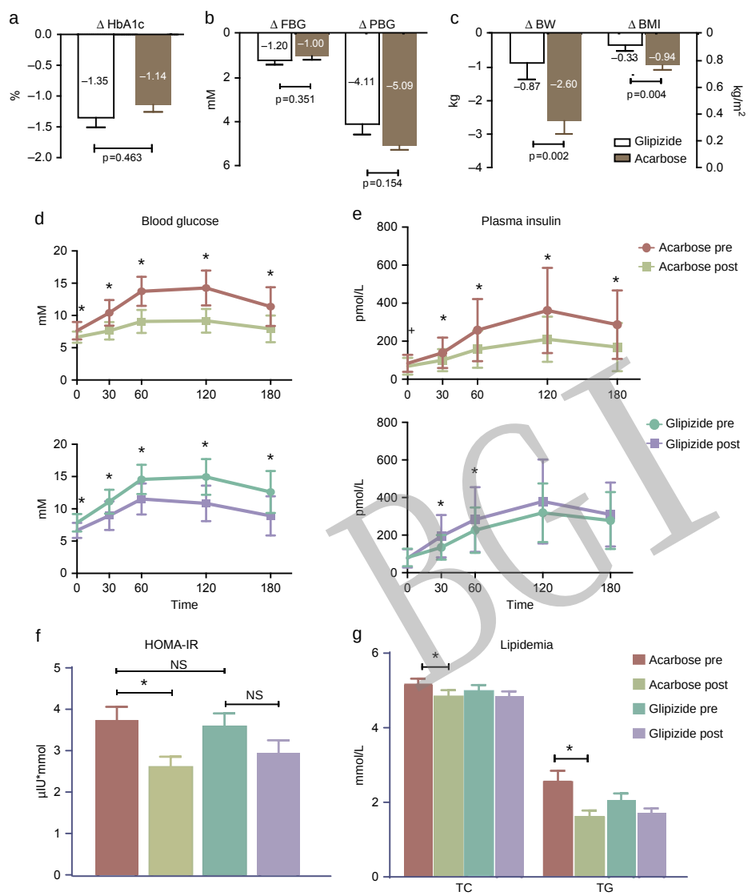


图1 阿卡波糖和格列吡嗪治疗3个月后患者的主要临床结果

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汁酸生成得到促进，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。

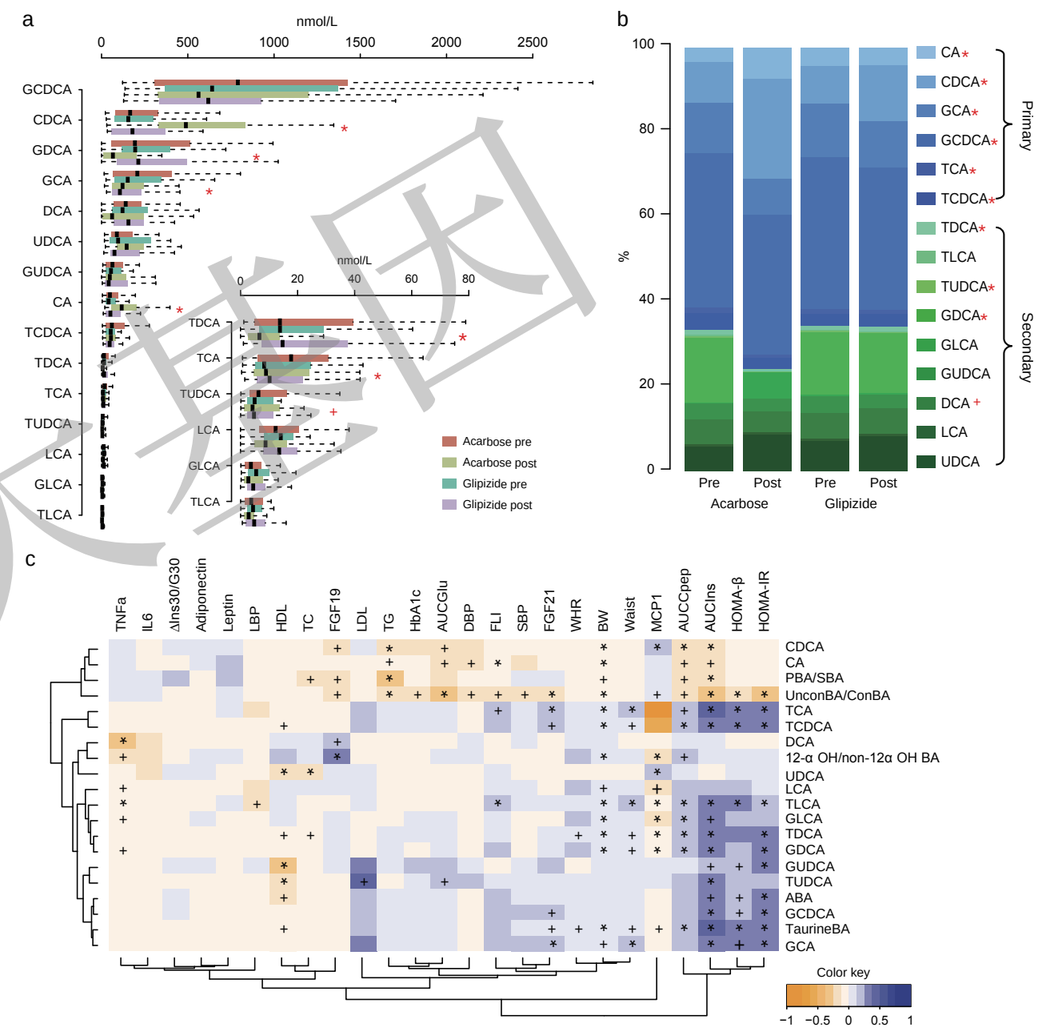
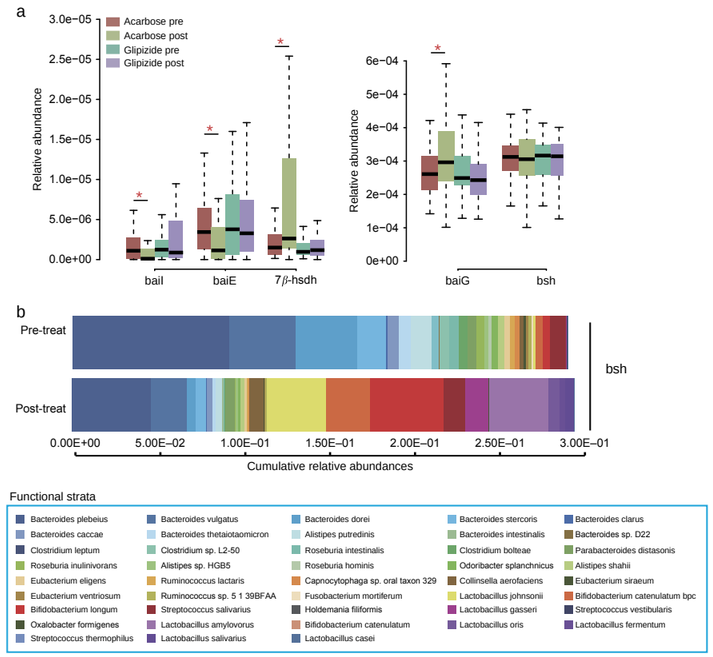


图2 阿卡波糖治疗后胆汁酸组成的变化与临床结果相关

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。

BGI 华大基因



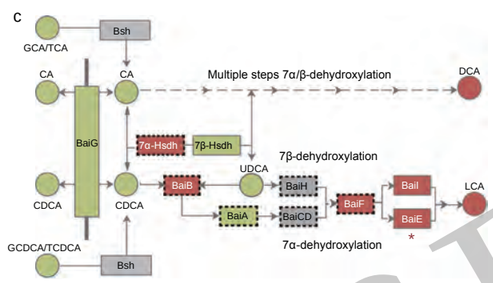


图3 阿卡波糖治疗潜在影响次级胆汁酸代谢

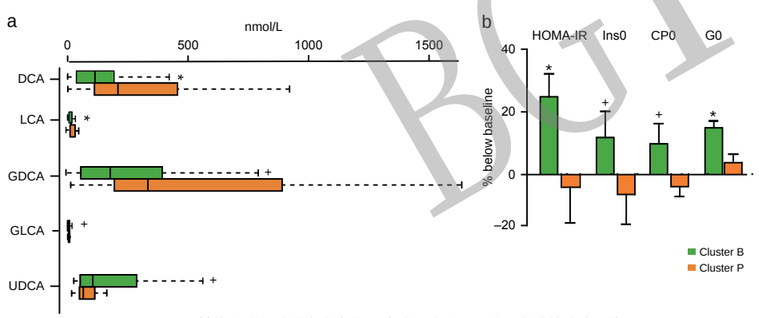


图4 不同基线菌群组成的患者血浆胆汁酸组成和不同的阿卡波糖治疗反应

• 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。

04 多组学关联分析

蛋白质组与转录组关联分析

代谢组与宏基因组关联分析

BGI 华大基因

蛋白质组与转录组关联分析

• 研究背景

生命体是一个多层次，多功能的复杂结构体系，从DNA、RNA、蛋白质到代谢物的过程中涉及到一整套精细的表达调控机制，如转录调控、转录后调控、翻译调控、翻译后调控等。高通量技术的发展积累了大量的组学数据，这使得由精细的分解研究转向系统的整体研究成为可能。整合多组学数据能够实现对生物系统的全面了解，建立有效指示表型的模型，揭示重要的生物标志物。

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。

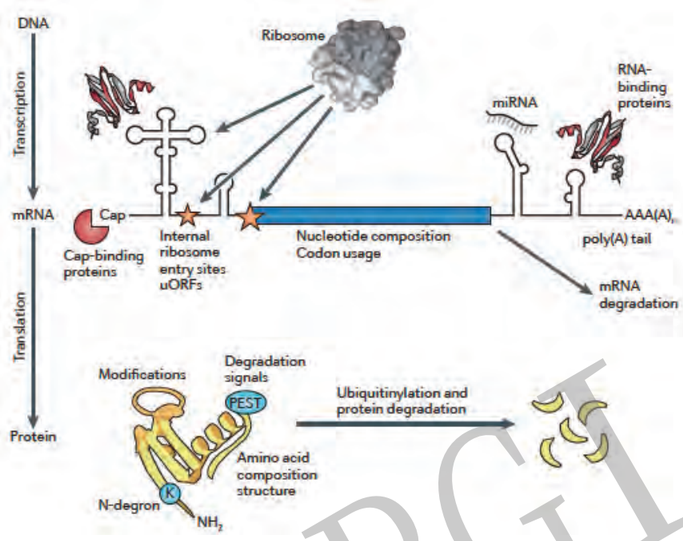


图1 转录调控和蛋白翻译后修饰模式

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

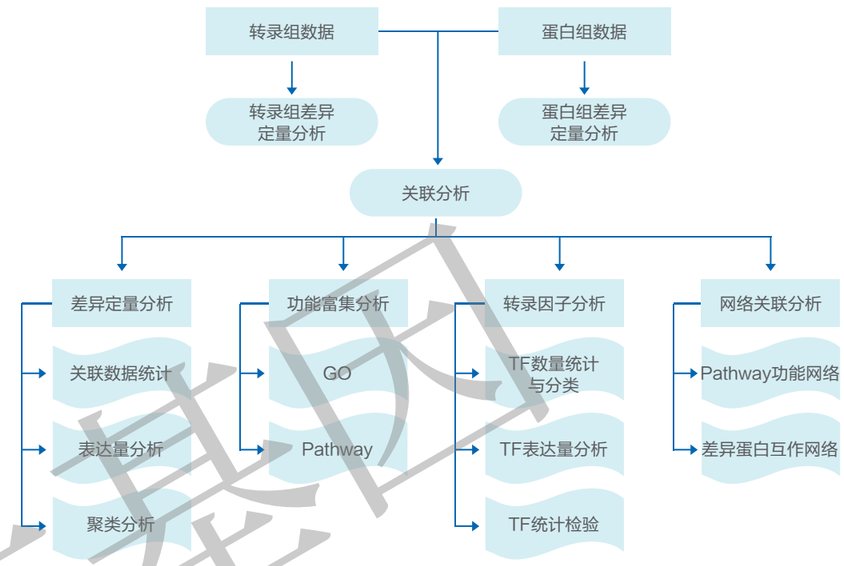


图2 蛋白组与转录组关联分析方案整体思路

• 分析内容

对转录组和蛋白组分别获得的表达量情况和功能分析结果进行关联分析，获得两个组学上的定性、定量和功能层面上的关联分析结果。

1、鉴定关联分析：利用转录组数据库构建蛋白质数据库，提高蛋白鉴定数。

图3中绿色代表转录组鉴定到的基因个数，红色代表蛋白质组鉴定到的蛋白个数，两个圆圈重叠的部分即为转录组与蛋白组共同鉴定到的基因个数。

通过鉴定关联分析韦恩图，可从整体上分析鉴定到的mRNA和蛋白质的情况。

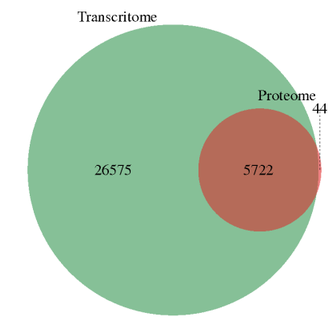


图3 定性关联韦恩图

2、关联与非关联表达量分布：基于转录组数据，对基因分为关联与非关联，展示两种分类的表达量分布，从而指导后续对比较组mRNA转录后调控机制分析。

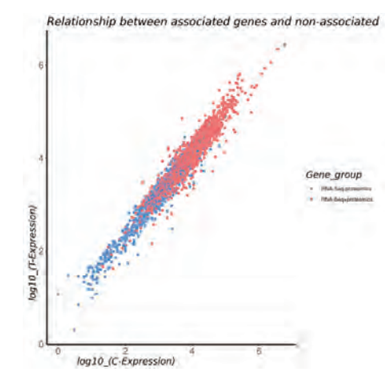


图4 关联与非关联表达量散点图

横坐标为对照组的表达量，纵坐标为实验组的表达量。其中红点为关联上的基因，蓝点为没有关联上的基因

BGI 华大基因

3、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：

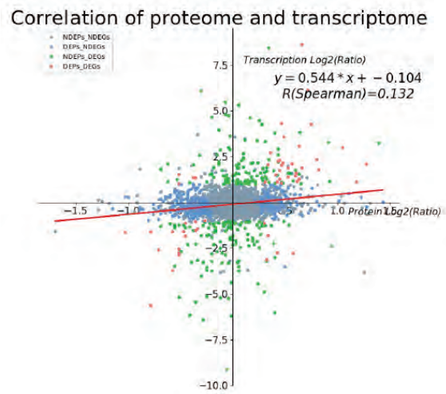


图5 所有定量蛋白质和基因的表达量关联图

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。

4、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。

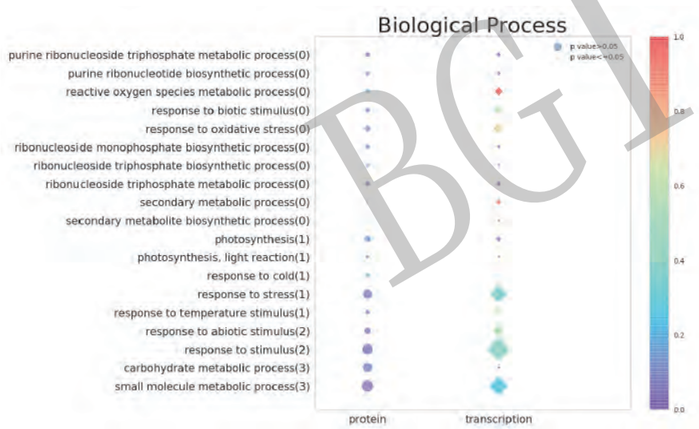


图6 GO关联数量及富集因子分析四维气泡图

Y轴为各GO条目，括号中统计了该条目所关联的蛋白数量，气泡大小代表富集数量，

颜色代表富集因子，圆圈代谢转录组（蛋白质组）P值大于等于0.05，钻石形代表转录组（蛋白质组）P值小于0.05

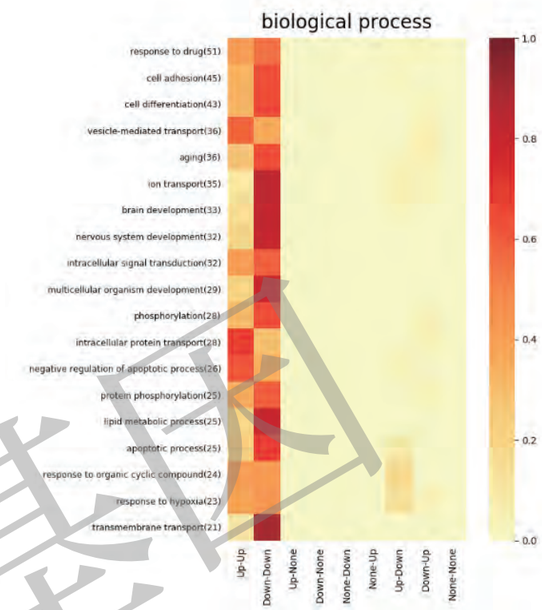


图7 GO条目关联分类概率热图

Y轴为各GO条目，括号为该条目的关联数量；X轴为不同差异类型分布情况

热图颜色代表该类型在该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上下调，蛋白上下调，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。KEGG (Kyoto Encyclopedia of Genes and Genomes)是有关Pathway的主要公共数据库，该数据库整合了基因组、化学以及系统功能信息，特别是测序得到的基因集与细胞、生物体以及生态环境的系统性功能相关联。

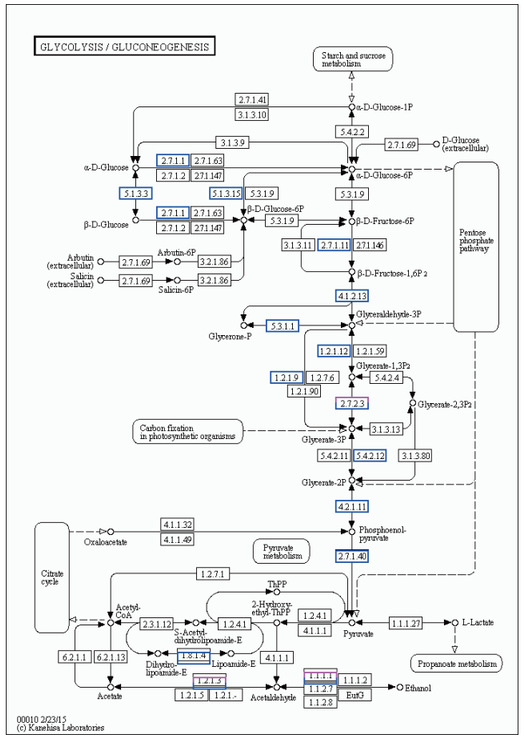


图8 差异蛋白和差异基因Pathway整合图（红框代表差异蛋白，蓝框代表差异基因）

BGI 华大基因

5、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。

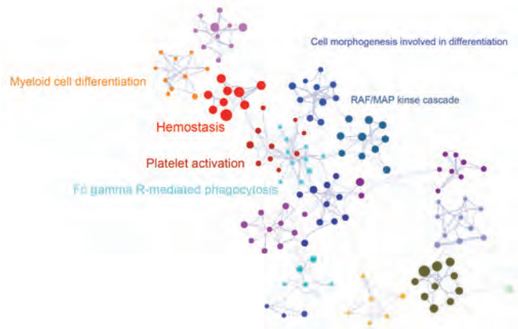


图9 差异蛋白和差异基因网络互作图

6、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。

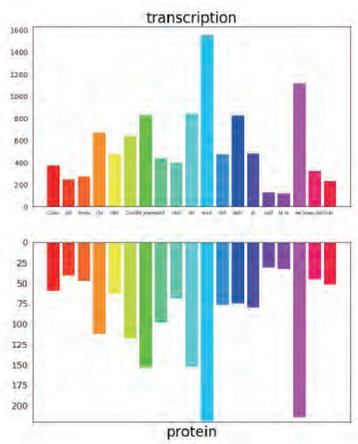


图10 各转录因子家族鉴定到的转录因子个数

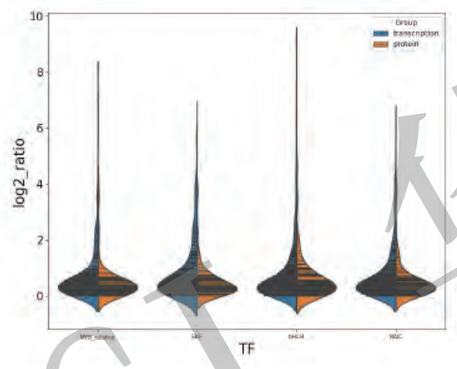


图11 各转录因子家族在基因及蛋白层面的表达量分布图

应用案例

案例一

黑腹果蝇发育过程中转录后调控定量分析

Quantifying post-transcriptional regulation in the development of Drosophila melanogaster

研究内容：黑腹果蝇发育机制　研究策略：RNA-Seq、蛋白定量　发表时间：2018年　发表期刊：Nature Communications

复杂的转录后调控的存在，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生为研究对象，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（ρ=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敲出炎症了这一假设。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。

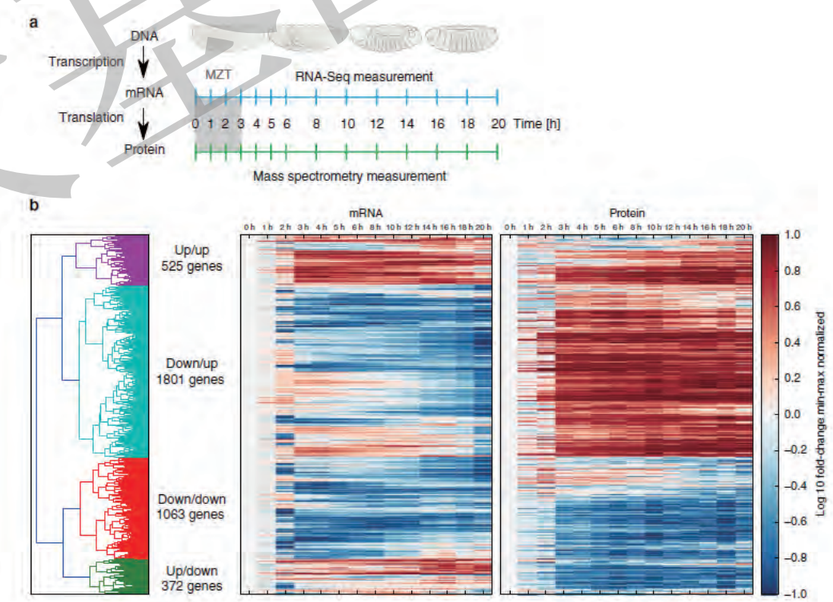


图1 果蝇胚胎发生的转录组&蛋白组分析结果

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

BGI 华大基因

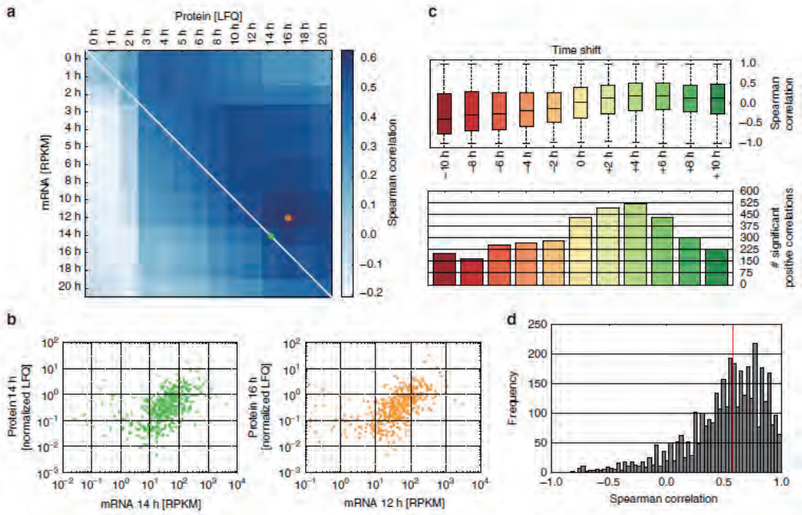


图2 mRNA和蛋白质相关性分析结果

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

案例二

通过多组学数据展示不同发育阶段小鼠胃的分子数据

A time-resolved multi-omic atlas of the developing mouse stomach

研究内容：小鼠胃发育机制　　研究策略：转录组、蛋白质组　　发表时间：2018年　　发表期刊：Nature Communications

哺乳动物的胃在结构上呈现高度多样化，其功能很大程度上依赖于胚胎发育的情况。虽然前人已经报道对胃发育期间形态变化的研究，但缺乏对其潜在分子变化进行系统化分析。本研究展示了小鼠胃在多个发育阶段的全面转录组和蛋白组图谱。研究基于蛋白质和RNA的变化水平，对三个不同阶段12,108个基因产物进行了定量分析，获得纵向时间尺度上胃功能分子标记。转录组分析发现了与发育相关的重要亚型，并在肽段水平上功能验证了未注释的新可变剪切转录本。蛋白组分析发现了胃发育中差异表达的蛋白质在弥漫型胃癌中也显著表达。最后，研究提出了胃的发育与胃癌发生的信号通路密切相关。

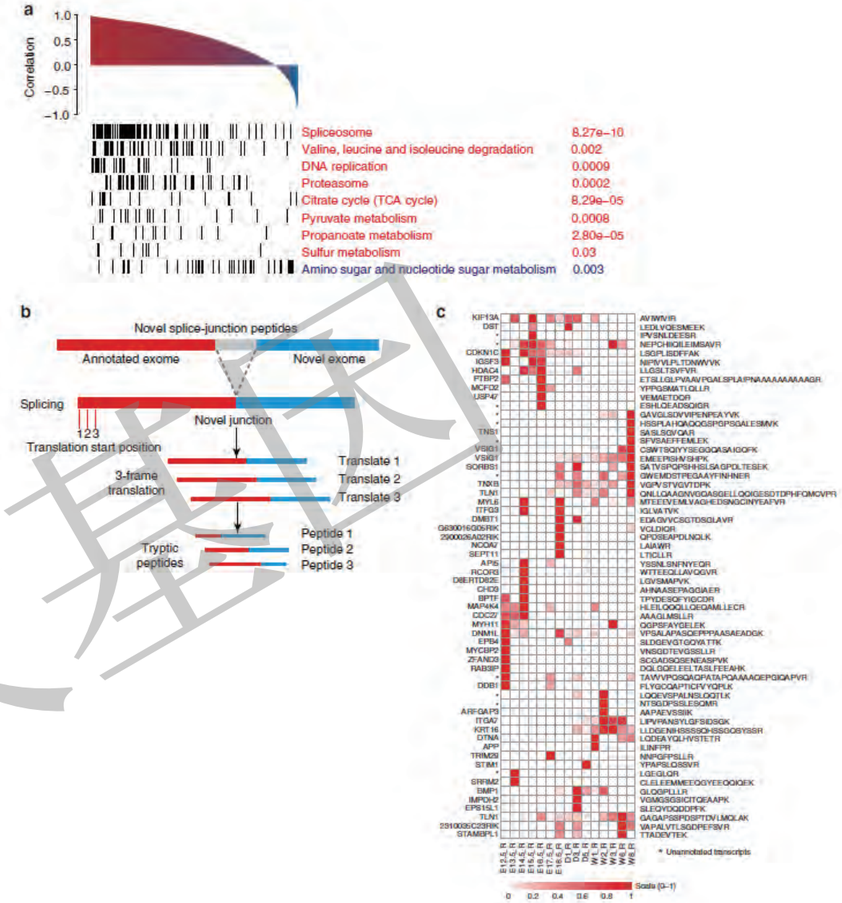


图1 小鼠胃蛋白质组和转录组整合分析

a.根据所有时间点下每个基因产物的标准化分数（z-score），通过Pearson相关系数，计算蛋白质-RNA相关性

红色表示正相关基因及其富集途径，蓝色表示负相关基因及其富集途径；

b.发现新型剪接点多肽的途径；c.所有15个时间点上，60种新型剪接体多肽的相对丰度分布

BGI 华大基因

案例三

通过转录组和蛋白质组研究蜱虫感染巴贝虫病的分子机制

Transcriptome and proteome response of Rhipicephalus annulatus tick vector to Babesia bigemina infection

研究内容：蜱虫感染巴贝虫病机制　研究策略：转录组、蛋白质组　发表时间：2019年　发表期刊：Frontiers in Physiology

用系统生物学方法深入了解蜱虫载体与病原体之间的相互作用。蜱虫是巴贝虫病的主要载体之一，对动物健康有着巨大影响，获取该物种的感染信息，更好地了解蜱和病原体生物学，为病原体传播动态提供新的控制方法。在蜱虫中，唾液腺（SGs）在病原体感染和传播过程中起着关键作用。对未感染和已感染巴贝虫病的雌性蜱虫唾液腺进行转录组测序，获得6,823和6,475个unigenes，360个差异基因（$p<0.05$），总共鉴定了3,679个蜱虫蛋白，406个差异蛋白。巴贝虫感染导致mRNA和蛋白质水平降低，综合转录组和蛋白质组学数据表明应激反应和凋亡途径对感染起关键作用。对6个编码基因进行RNAi发现，敲除gp80、dapk-1和bi-1导致巴贝虫感染水平显著增加，而hsp和QtRibosyl敲除感染水平并未显著降低。基因敲除后并不影响蜱虫的存活率，但与对照相比，gp80、dapk-1和QtRibosyl敲除组的雌性个体体重和产蛋量受到影响，这些结果促进我们对蜱虫-巴贝虫分子相互作用的理解，并为新的蜱虫抗原作为疫苗接种提供新的靶标，以控制蜱虫感染和病原体感染/传播。

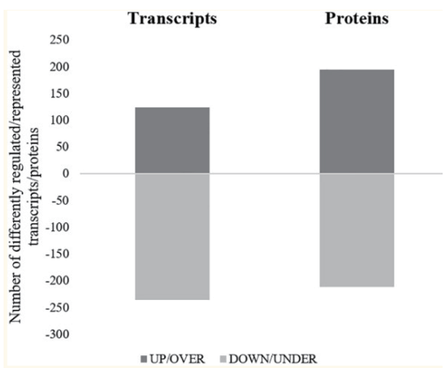


图1 感染巴贝虫病后的差异蛋白质及差异基因

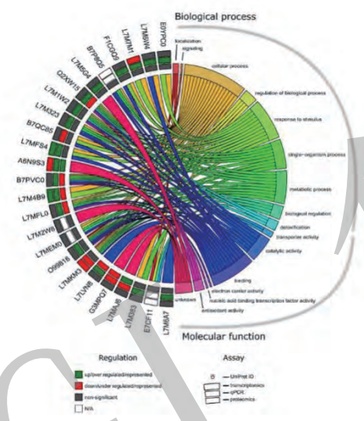


图2 对差异蛋白及差异基因的功能进行和弦图分析

代谢组与宏基因组关联分析

• 研究背景

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：1、消化道类疾病；2、代谢性疾病，包括肥胖、II型糖尿病T2D等；3、肝脏类疾病；4、免疫类疾病；5、肺部疾病；6、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、某些维生素和非必需氨基酸的合成、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1000Da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应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。

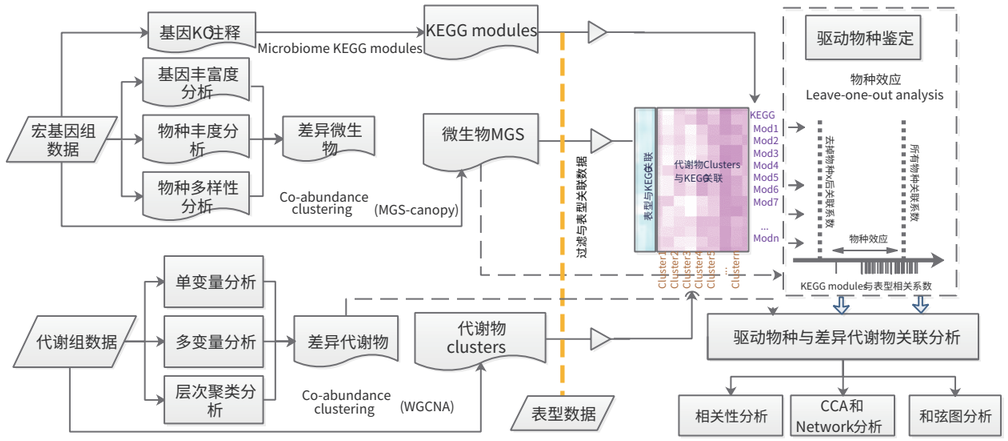


图1 宏基因组代谢组关联分析流程图

BGI华大基因

分析结果展示

• 物种丰度分析

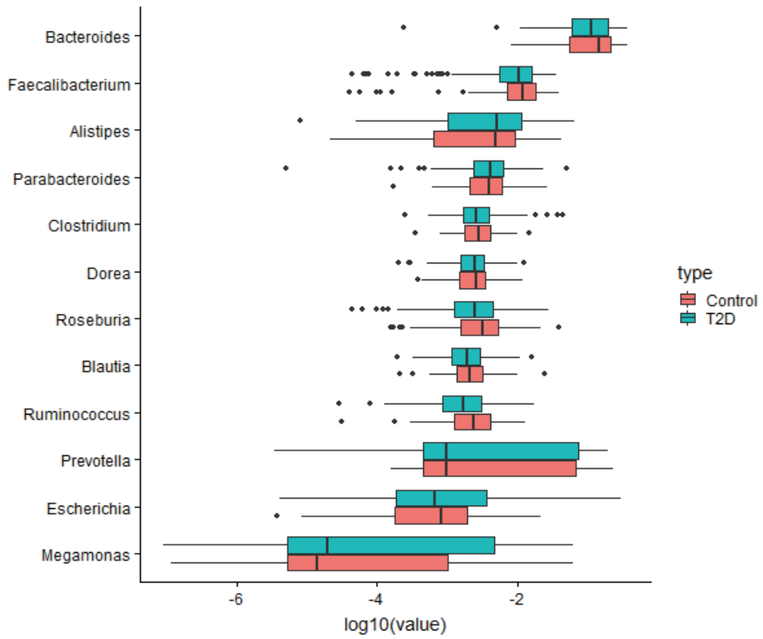


图2 Case组和Control组中物种丰度的箱线图

• 物种多样性分析

使用α-diversity和β-diversity对Case组和Control组的物种多样性进行评估，如下示例图：

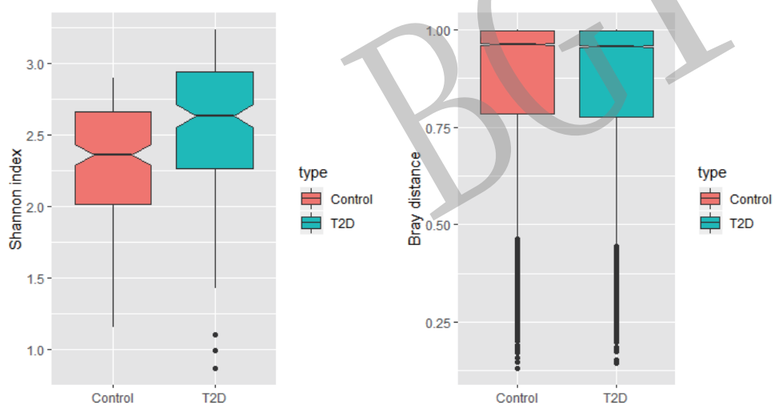


图3 Case组和Control组中α-diversity和β-diversity多样性箱线图

• 物种关系网络分析

通过网络图直观展示疾病组和对照组物种组成的差别以及物种之间的相互作用。

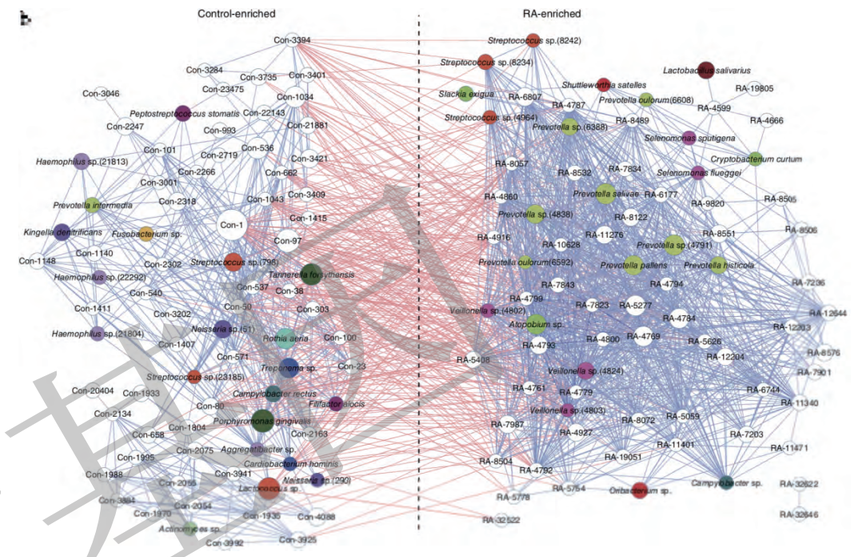


图4 物种关系网络分析图

• 代谢组单变量分析

代谢组学数据分析中常用统计检验的方法（T检验、Wilcoxon秩和检验、方差分析等）和差异倍数分析等单变量分析方法对数据所反映的数量变动进行分析。本产品采用的是T检验和变异倍数分析（Fold change analysis，FC analysis）。在统计分析过程中，进一步对统计检验产生的p-value进行校正得到q-value。最终结果以火山图（Volcano plot）形式呈现差异倍数（Fold change）和q-value两个指标，通常以差异倍数≥1.2或≤0.8333，q-value值小于0.05作为筛选差异代谢物的条件。筛选的结果如图所示：

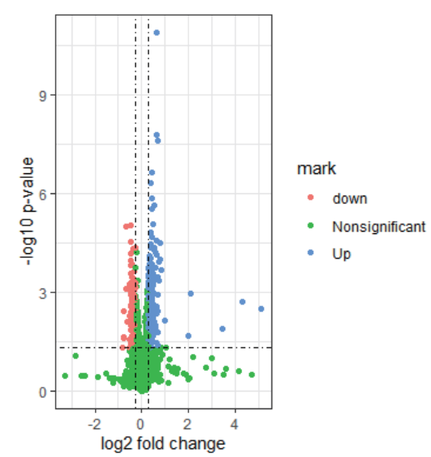


图5 Case组和Control组中代谢物的差异表达火山图

BGI 华大基因

• 多变量分析

要从海量数据中发现代谢组学潜在标志物，除了单变量分析还需借助多变量统计分析方法，常用的多变量分析方法有主成分分析（PCA）和偏最小二乘法-判别分析（PLS-DA）。本项目分别采用PCA和PLS-DA方法进行多变量分析，如图所示：

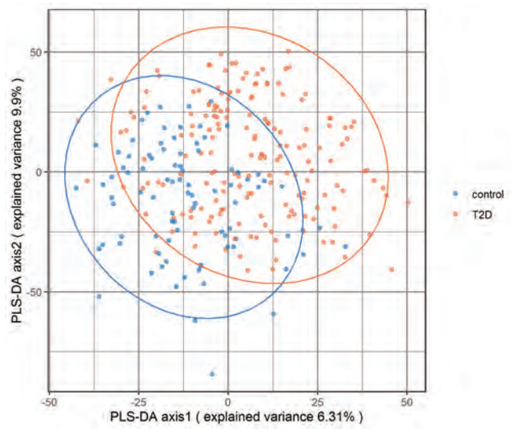


图6 Case组和Control组血清代谢物和PLS-DA图

• 差异代谢物层次聚类分析

本产品采用多变量分析PLS-DA模型前两个主成分的VIP（Variable Important for the Projection）值结合单变量分析差异倍数（Fold change）和q-value值来筛选差异表达的代谢物。 筛选条件: 1、VIP≥1；2、fold change≥1.2 或者≤ 0.8333；3、q-value < 0.05，三者取交集，得到共有的离子即差异代谢物。如图7所示：

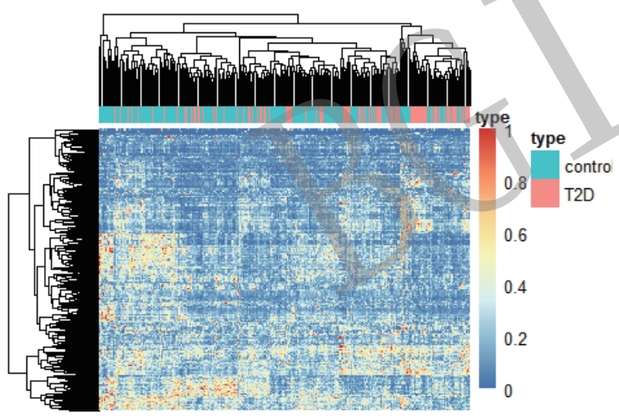


图7 差异代谢物层次聚类热图

宏基因组和代谢组关联分析

• 代谢物Clusters与表型关联

将代谢物通过数据驱动聚类（WGCNA）进行降维处理后生成不同的clusters，每一个cluster中的特征向量值与表型数据进行关联分析（Spearman correlation），寻找最显著关联的clusters，从而选择显著相关的代谢物cluster进行后续跨组学关联分析，过滤掉与表型数据不相关的clusters。如图所示：

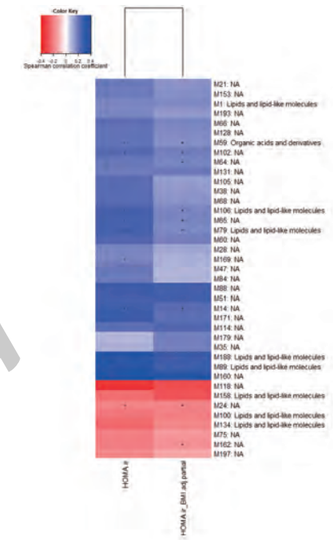


图8 代谢物clusters与表型特征相关性热图

• 微生物MGS与表型关联

将微生物组物种通过数据驱动分箱（MGS）法进行降维处理后生成不同的MGSs，对每一个微生物MGS所有基因求中位值（Median），然后与表型数据进行关联分析(Spearman correlation)，寻找最显著关联的MGS，从而选择显著相关的微生物MGS进行后续的跨组学关联分析。如图所示：

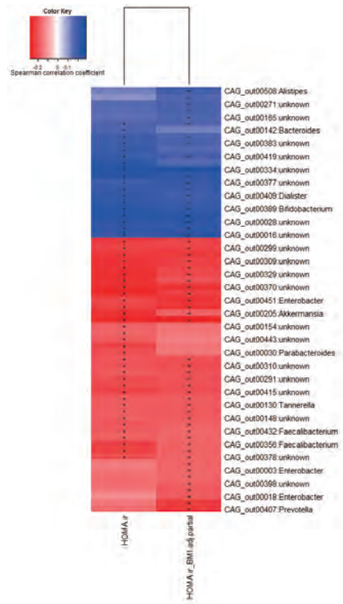


图9 微生物MGS与表型特征相关性热图

深圳华大基因

• 微生物KEGG功能模块与表型关联

将微生物组功能组成通过知识驱动如KEGG层级功能模块分类法进行降维处理后，生成不同的KEGG功能模块，对每一个module所有基因求中位值（Median），然后与表型数据进行关联（Spearman correlation），寻找最显著关联的module，用于后续的跨组学关联分析。如图所示：

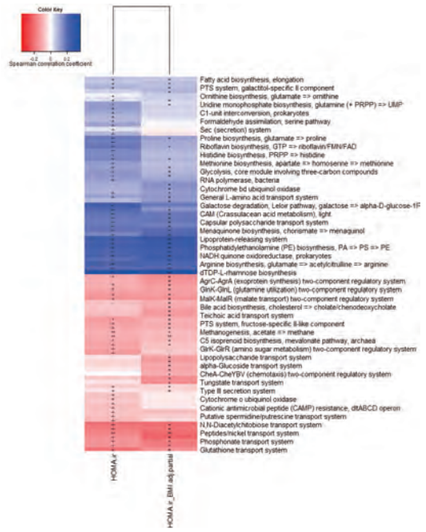


图10 微生物KEGG功能模块与表型特征相关性热图

• 代谢物Cluster与微生物KEGG模块跨组学关联

将与表型显著关联的代谢物clusters和与表型显著关联的KEGG功能模块进行跨组学关联，揭示肠道微生物与血清代谢物之间的相互作用关系。如图所示：

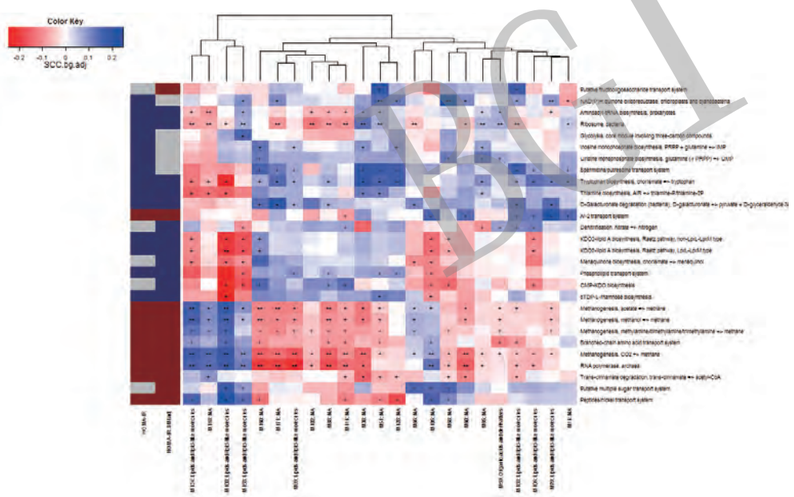


图11 微生物KEGG功能模块与表型特征、代谢物相关性热图

• 驱动物种（driver species）鉴定

为了探究与表型显著相关的KEGG功能模块中是哪个或哪些物种起主要的驱动作用，采用leave-one-MGS-out方法进行驱动物种的鉴定。由于篇幅有限，其中以BCAA生物合成KEGG模块进行驱动物种鉴定的为例进行说明，如图所示：

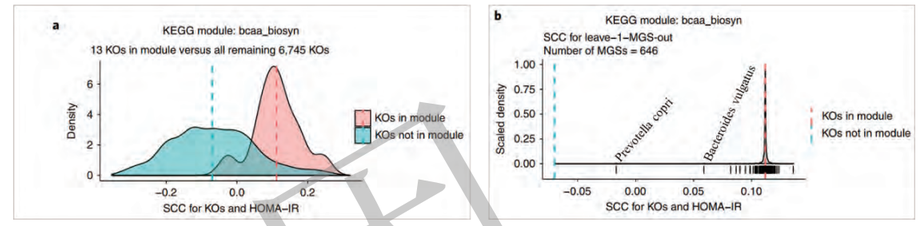


图12 支链氨基酸合成KEGG功能模块

（BCAA biosynthesis：M00019, M00570, M00535 和M00432，共13个KOs）与表型IR的相关性SCC密度图。

a.合成BCAA生物功能模块的KOs与其他所有KOs的SCC密度分布图。红色区域代表BCAA biosynthesis模块中13个KOs与表型的SCC分布情况。绿色区域代表去掉13个KOs后剩余KOs与表型SCC分布情况。

b. 去掉BCAA biosynthesis功能模块中每个MGS的基因后，剩余KO与HOMA-IR相关性SCC分布图。

• 驱动物种和显著代谢物相关性分析

为了进一步研究驱动物种与表型显著相关代谢物之间的内在相互作用关系，首先每个KEGG模块挑选最重要的5个驱动物种，如图所示：

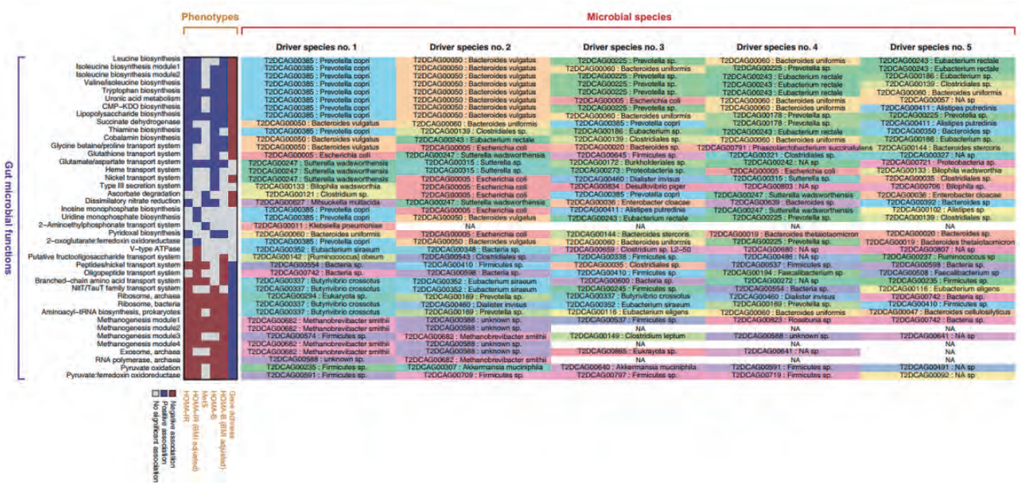


图13 与表型显著相关的每个KEGG模块的最重要的5个驱动物种

华大基因

每种物种用不同颜色代表。然后将每个KEGG模块的最重要的5个驱动物种与表型显著相关差异代谢物(FDR<0.001)进行spearman相关性分析，如图所示：

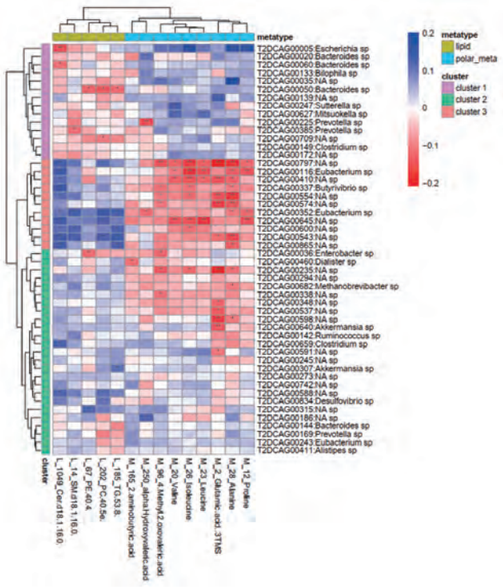


图14 驱动物种与显著代谢物spearman相关性热图

• 驱动物种和显著代谢物CCA分析

基于上一步驱动物种与显著代谢物的spearman相关性分析，筛选出相关性显著（P<0.01）驱动物种和差异代谢物进一步做CCA分析，如下图所示：

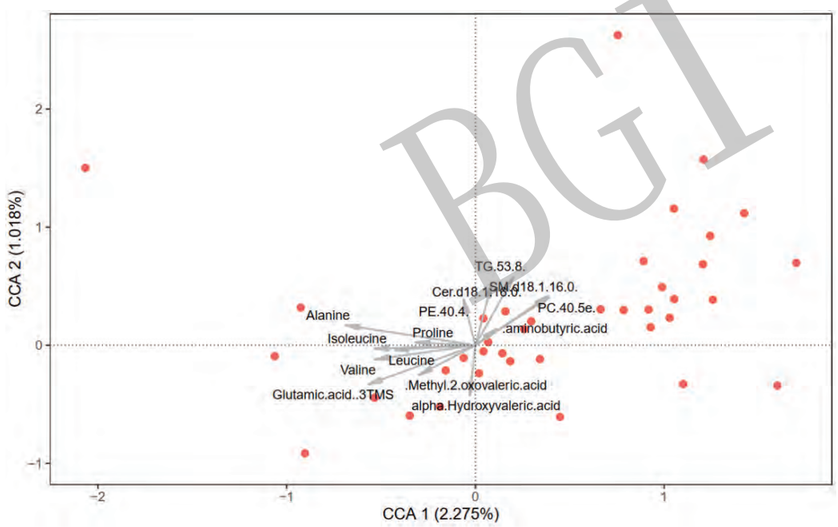


图15 驱动物种和显著代谢物CCA图

• 驱动物种和显著代谢物和弦图分析

基于驱动物种与显著代谢物的spearman相关性分析，筛选出正相关的驱动物种和差异代谢物（P<0.01）进一步做和弦图分析，如图所示：

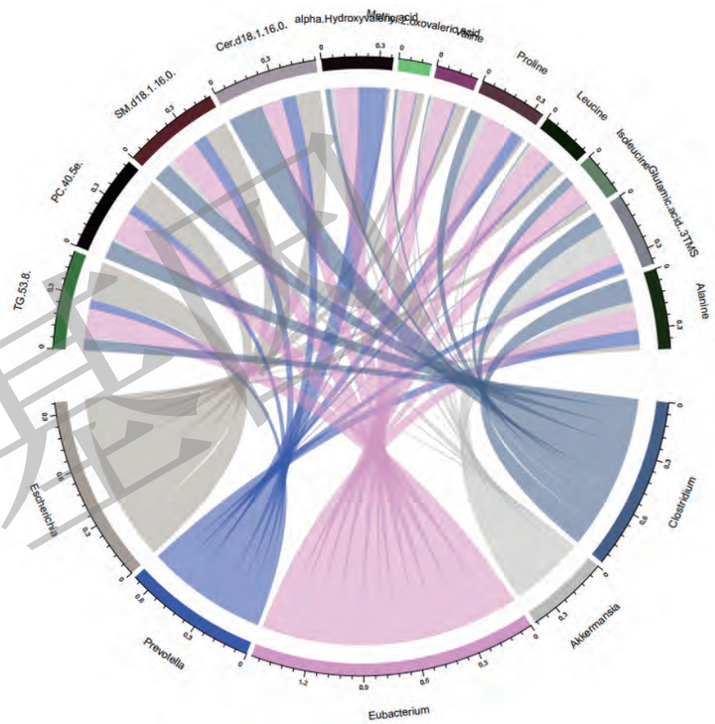


图16 驱动物种和显著代谢物正相关的和弦图

BGI 华大基因

应用案例

案例一

肥胖干预前后肠道菌群和血清代谢物的变化

Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention

研究内容：肥胖干预前后肠道和代谢的变化　研究策略：粪便宏基因组、血清代谢组

发表期刊：Nature Medicine　发表时间：2017年

• 研究背景

肥胖已经成为世界性流行疾病，是2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。

• 样本来源

1、疾病组 VS对照组（肥胖组n=95，对照组n=105）粪便、血清

2、特定菌株移植小鼠，小鼠粪便、血清

3、肥胖患者袖状胃切除术前后及对照组粪便、血清（对照组n=26；肥胖组n=23(术前OM，n=23；术后1月1M，n=17；术后3月3M，n=17)

• 方法与策略

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

• 结果与分析

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。

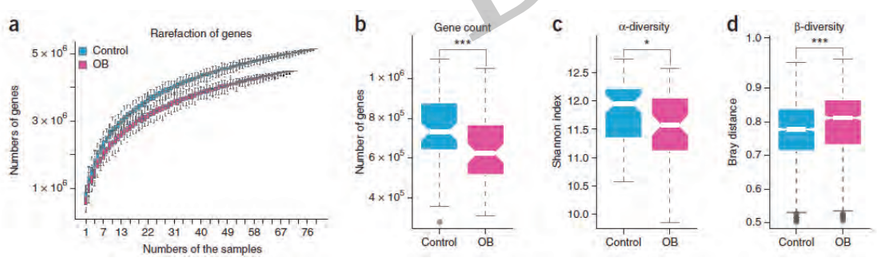


图1 肥胖组（OD）与对照组（Control）的基因稀疏性曲线、基因数量和丰富度的区别

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。

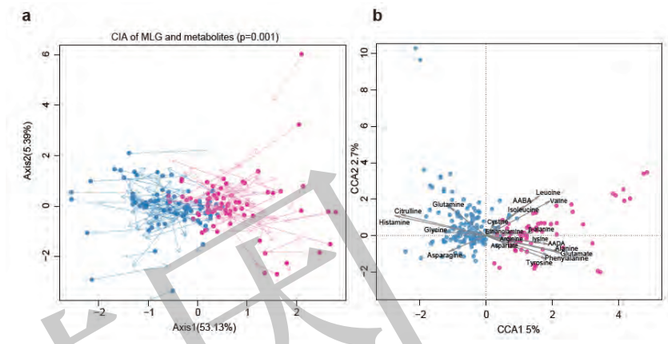


图2 微生物与代谢物相关性分析

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。

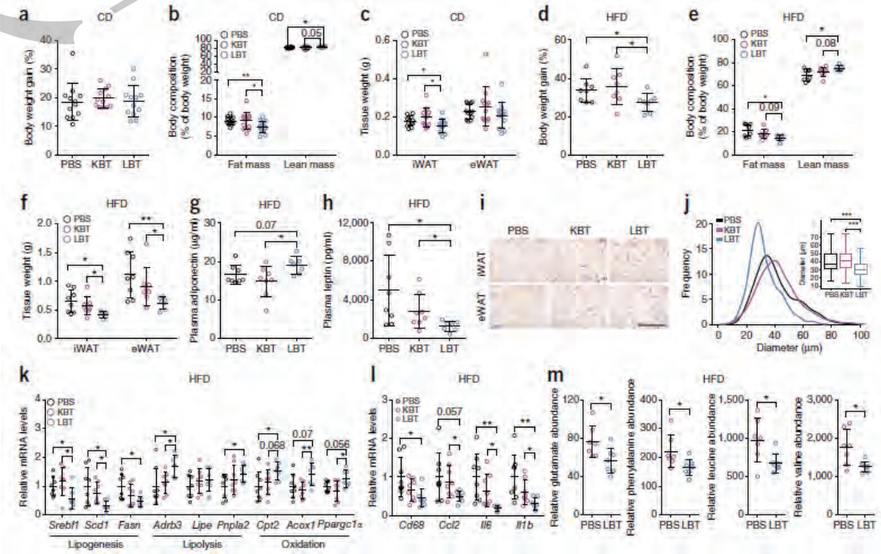


图3 BT菌对小鼠肥胖和代谢的影响

BGI 华大基因

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的同样的水平；代谢循环中的谷氨酸含量也大大减少。

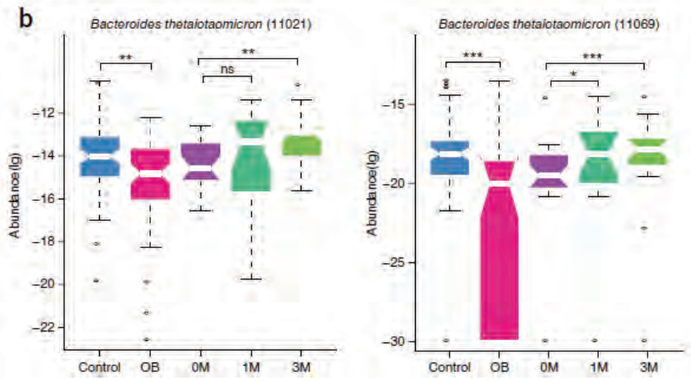


图4 减重手术患者术后3个月的BT菌基因关联群组丰度变化

案例二

肠道与心血管疾病

Gut microbial associations to plasma metabolites linked to cardiovascular phenotypes and risk: a cross-sectional study

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　研究策略：粪便宏基因组、血清代谢组

发表期刊：Circulation Research　发表时间：2019年

- **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知之甚少，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。

- **样本来源及方法**

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 / m^2，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。

- **结果与分析**

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个物种存有64个关联关系	42个代谢物与微生物物种有一种关联
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存有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排前发关联强度和方向，即便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一些强有力的差异。

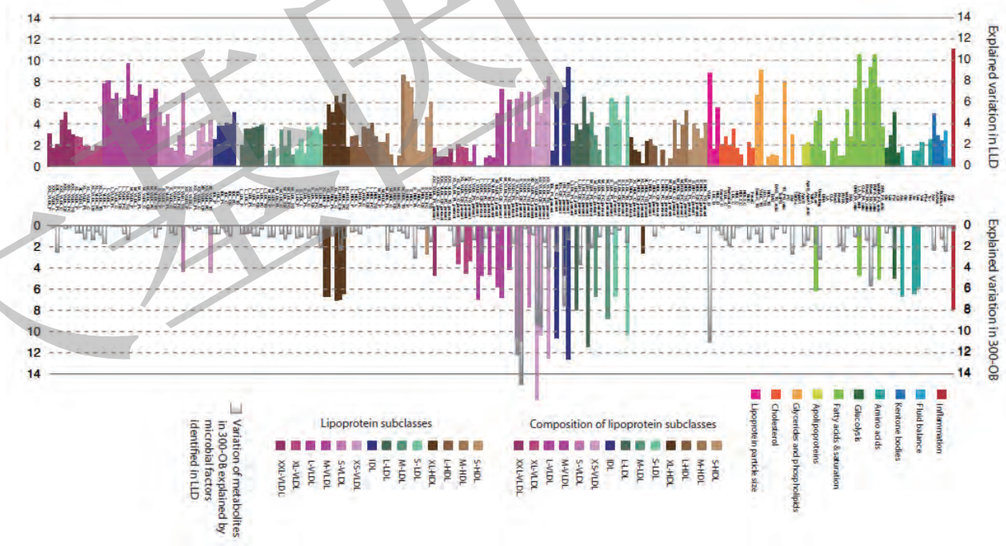


图1 肠道微生物组解释的血浆代谢物变化

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。

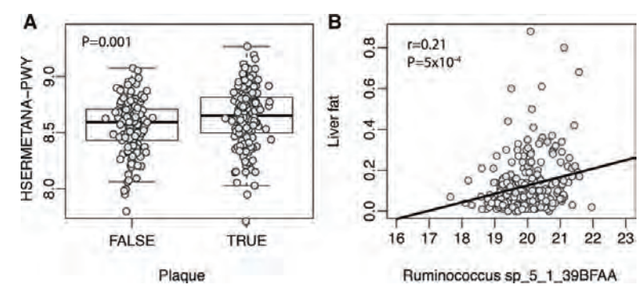


图2 300-OB队列中肠道微生物通路和物种与临床结果的关联

BGI 华大基因

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。

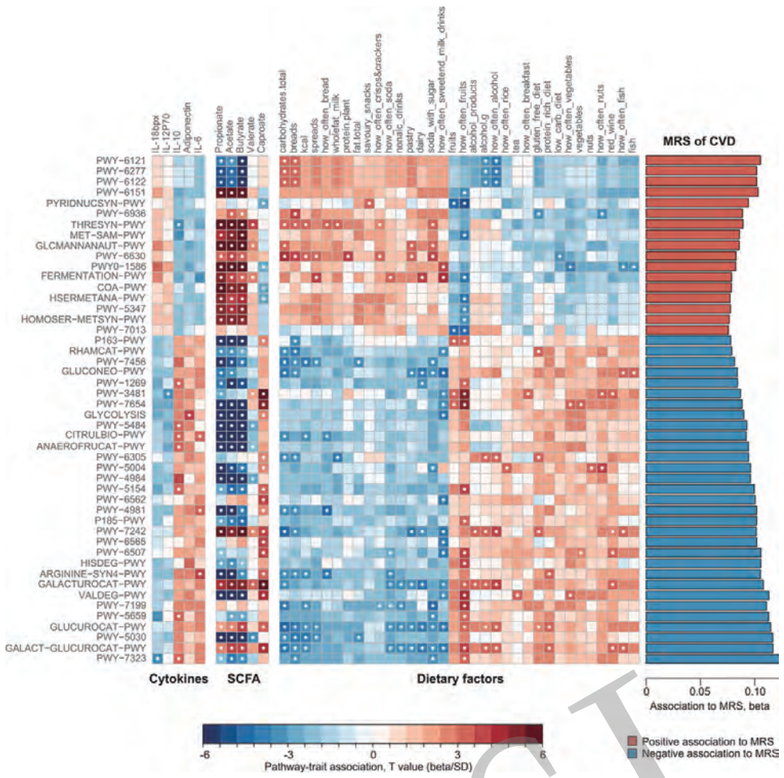


图3 48个生物通路与CVD MRS、粪便SCFAs、细胞因子和饮食因素的关联

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种细胞因子存有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。

BGI 华大基因

05 送样要求

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、酪氨酸磷酸化	
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定量产品单次实验样本可低至1mg。
	软体动物类（弓形虫、血吸虫、果蝇、螨虫、小菜蛾、灰飞虱、绦虫、蝉、涡鞭虫等）	≥5mg	≥10mg	≥100mg	
细胞类	悬浮细胞、贴壁细胞	$\geq 5\times10^5$	$\geq 5\times10^6$	$\geq 5\times10^7$	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≤200mL即可。
液体类	血浆、血清	≥200μL	≥200μL	≥1mL	若不需要去高峰度，只需要送20μl即可满足实验需求。
	羊水、脑脊液、精液等	≥100μL	≥200μL	≥1mL	若不需要去高丰度，只需要送50μl即可满足实验需求。
	唾液、乳汁	≥500μL	≥2mL	≥10mL	
	尿液	≥15mL	≥50mL	–	正常人的尿液蛋白含量较低，建议翻倍增加送样量；“–”表示样本量很难满足，不适合做。
	泪液	≥10μL	≥200μL	–	样本难收集，不适合做乙酰化和酪氨酸磷酸化。
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	≥2g	≥4g	≥20g	
	植物的花蕾、花粉	≥100mg	≥500mg	≥2g	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。

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、酪氨酸磷酸化	
植物类	植物的种子（水稻/小麦种子等）、果实（苹果、水蜜桃、梨）	≥100mg	≥500mg	≥2g	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
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≥50mg 或 细胞数$\geq 5\times10^6$	菌体≥200mg 或 细胞数$\geq 2\times10^7$	菌体≥1g 或 细胞数$\geq 1\times10^8$	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≥10μg蛋白	内含≥10μg蛋白	对于修饰位点检测，若需要进行富集操作，则需将送样量提高3倍。
胶点	考染或银染胶点	肉眼可见斑点	内含≥1μg蛋白	内含≥1μg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	≥100μg，浓度≥0.5μg/μL	≥1mg，浓度≥0.5μg/μL	≥5mg，浓度≥0.5μg/μL	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20μg。
	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≥1μg蛋白	内含≥1ug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	≥100mg	≥500mg	–	“–”表示不建议做乙酰化和酪氨酸磷酸化。

代谢产品送样要求

样品类型	建议送样量/例
血清、血浆	≥ 250 μL
尿液	≥ 500 μL
动物或临床组织	≥ 200 mg
粪便、肠道内容物	≥ 200 mg
微生物、细胞	≥ 107 个
培养液	≥ 1 mL
植物组织	1-3g，最少0.6g

BGI 华大基因