

Production and Service

产品与服务 » 科技服务|探秘基因，助力科研

科技服务事业部基于国际先进的高通量测序平台和博士硕士为主的研发团队,开展了对转录调控、医学、群体、基因组、微生物等研究的主体业务,学术研究硕果累累,先后在《Nature Genetics》、《Nature Communications》、《PNAS》、《Plant Cell》等国际顶级学术刊物发布论文数千篇,累计影响因子 3000+, 此外还拥有 40 余项专利和 150 余项软件著作权, 多项经营指标连续多年保持 100%以上增长率。

遗传群体研究

遗传图谱 (重测序、SLAF、转录组)

BSA (重测序、SLAF、转录组)

KASP 基因分型

个体重测序

基因编辑

建库测序 (DNA、RNA)

医学科技

人重测序

全基因组甲基化测序

简化甲基化测序

区域捕获测序

全转录组测序

全外显子组测序

医学转录组测序

医学外泌体 RNA 测序

医学小 RNA 测序

医学 LncRNA 测序

医学 circRNA 测序

动植物基因组研究

基因组调研图

利用二代测序完成基因组精细图



利用二代&三代测序完成基因组精细图

利用三代测序完成基因组精细图

泛基因组

光学图谱辅助基因组组装

Hi-C 技术辅助构建染色体

利用 Nanopore 技术完成基因组精细图

HIC 互做

自然群体研究

遗传进化 (重测序、SLAF、转录组)

GWAS (重测序、SLAF、转录组)

全长比较转录组

2+3 重测序

转录调控研究

真核普通转录组

真核小 RNA 测序

真核小 LncRNA 测序

真核单细胞转录组测序

Circ RNA 测序

真核全长转录组 (PacBio 平台)

转录组与蛋白组联合分析

转录组与代谢组联合分析

全基因组甲基化测序、

真核单细胞转录组测序

微生物研究

细菌基因组完成图

细菌泛基因组研究

真菌基因组调研图

真菌基因组精细图

真菌基因组完成图

微生物多样性测序

宏基因组测序

宏转录组测序



生物云平台

单次项目研究套餐 (次账号)

包年分析研究套餐 (年账号)

微生物多样性研究套餐 (微云)

公共数据库及物种数据库的构建

物种综合数据库搭建

私有云/专用机

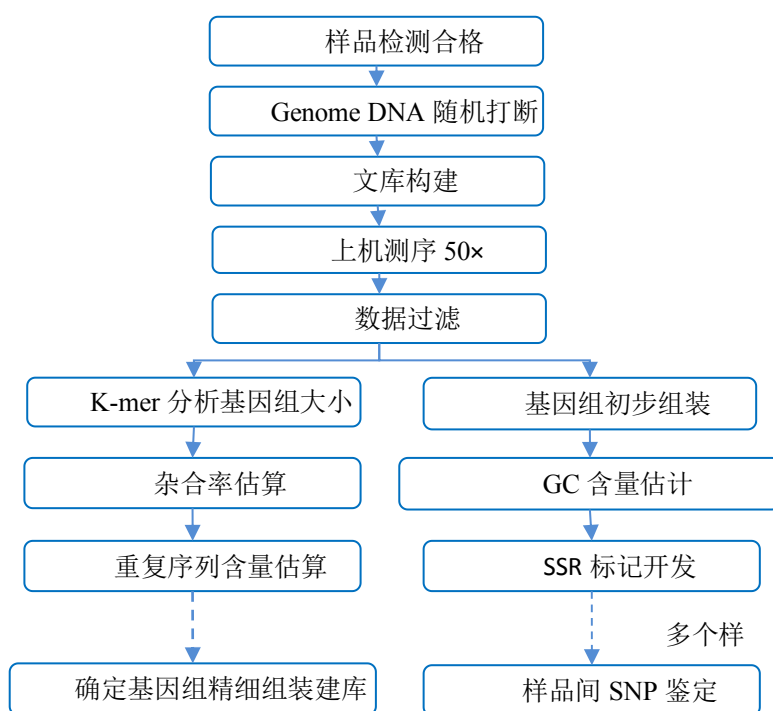
104 款工具包

基因组

基因组调研图 (survey)

快速获得基因组大小、杂合度、重复序列比例等基本信息,为制定该物种的全基因组 de novo 测序策略提供有效依据。

技术路线



survey 分析流程

百迈客调研图优势

免费的个性化分析

完成基本信息后,免费进行基因组组装注释,启动精细图还会获得更多优惠。

丰富的项目经验

对经济作物、水产、林木、昆虫、哺乳动物、鸟类等都有基因组项目经验。

高深度测序

50 倍的基因组覆盖度高深度测序,结果更准确,了解更全面。

基因组调研图的意义

启动全基因组测序的必要前提;
了解与近缘物种间的基因组差异状况;
得到该物种基因组的基本信息。

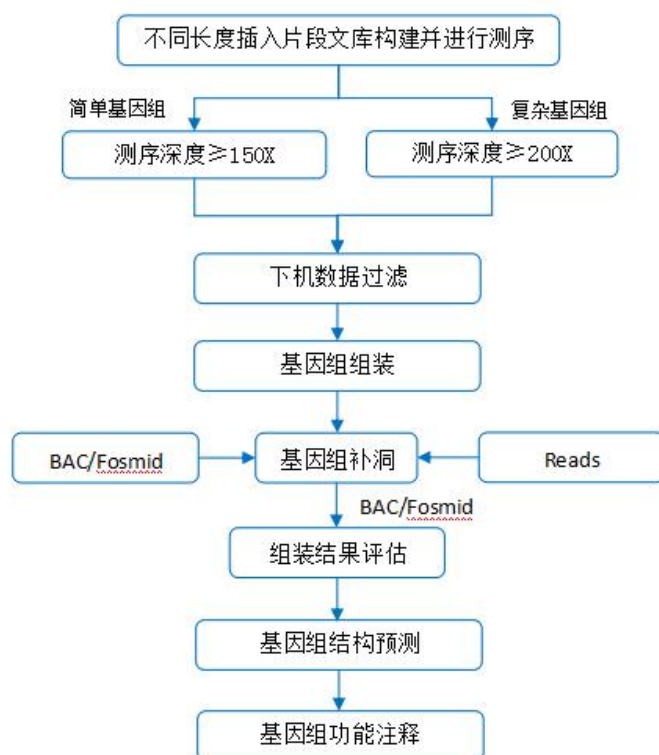
基因组调研图送样要求

物种	植物	动物
采样	植物组织; DNA	动物组织; 全血; DNA
总量	组织样总量 ≥ 0.5 g; DNA (Qubit 检测) 总量 ≥ 5 ug, 同时 ≥ 30 ng/ul。	组织样总量 ≥ 0.5 g, 血液 ≥ 2 mL。 DNA (Qubit 检测) 总量 ≥ 5 ug, 同时 ≥ 30 ng/ul。
运输条件	干冰运输, 确保 72 h 内送至百迈客实验室	

二代基因组精细图

在已知基因组的大小难度，或者有调研图结果的情况下，基于二代测序技术手段，进行基因组组装和基本分析，快速获得优质基因组序列，并完成标准的比较基因组和进化分析，为该物种的后基因组学研究奠定基础。

技术路线



二代基因组分析流程

基因组分析内容

1. 下机数据过滤

2. 完成基因组精细图

- a) 基因组组装；
- b) GC 含量分析；
- c) 测序深度分析；
- d) 单碱基错误率评估；
- e) 基因组覆盖度评估；
- f) 基因区覆盖度评估。

3. 基因组组分析

- a) 重复序列注释；
- b) 编码基因预测；
- c) Non-coding RNA 注释；
- d) 假基因 (Pseudogene) 注释。

4. 基因功能注释

- a) Nt 注释分析；
- b) Nr 注释分析；
- c) KEGG 代谢通路分析；
- d) COG 分析；
- e) GO 注释 (Gene Ontology)；
- f) TrEMBL 注释分析；
- g) 调控 Motif 预测。

5. 比较基因组和进化分析

- a) 基因家族的聚类、共有及特有的基因家族分析；
- b) 基因家族的收缩与扩张；
- c) 系统发育树的构建；
- d) 估算物种分化时间；
- e) 全基因组复制事件分析；
- f) 共线性分析。

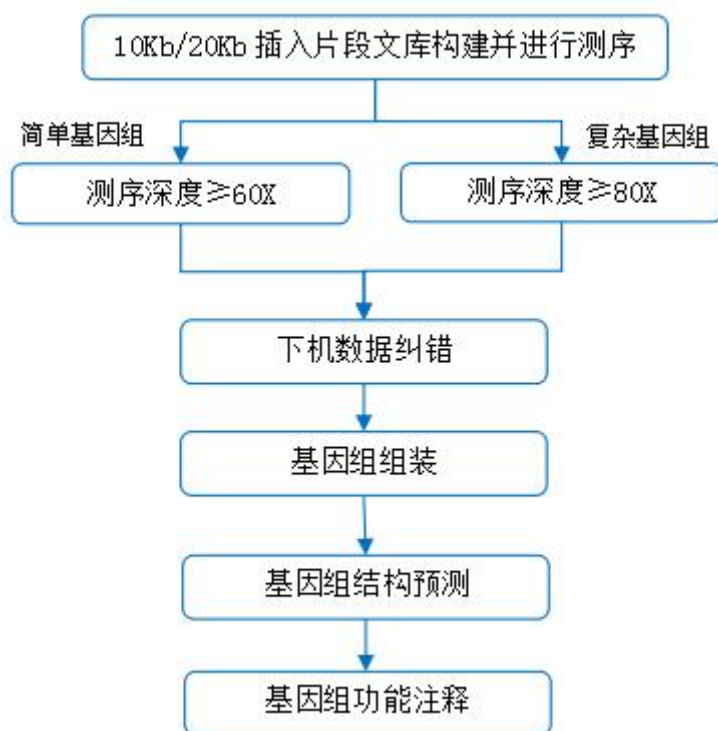
二代测序送样要求

物种	植物	动物
采样	植物组织；DNA	动物组织；全血；DNA
总量	组织样总量≥5 g； DNA (Qubit 检测) 总量≥50 ug，同时≥50 ng/ul。	组织样总量≥5 g；血液≥25 mL； DNA (Qubit 检测) 总量≥50 ug，同时≥50 ng/ul。
运输条件	干冰运输，确保 72h 内送至百迈客实验室	

三代基因组精细图

在已知基因组的大小难度，或者有调研图结果的情况下，基于 PacBio 三代测序技术手段，进行基因组组装和基本分析，获得高质量基因组序列，为该物种的后基因组学研究奠定基础。

技术路线



三代基因组分析流程

三代基因组组装的优势

超高的 contig N50 指标

三代的平均 reads 读长为 8-15Kb ,远高于二代测序,可以跨过基因组中较短的重复序列和杂合区域,高效解决组装难题,所以能够获得更高的 contig 指标。

组装正确率高

三代的长 reads 在基因组拼接时,不需要像二代测序一样借助复杂的算法,可以直接通过 overlap 进行组装。这样可以大幅减少在基因组组装上的拼接错误。

组装完整性好

三代的长 reads 可以有效解决重复序列区域的组装,减少重复区域组装的缩水现象,提高组装完整性,使组装的基因组大小更接近于物种真实的基因组大小。

二代测序送样要求

物种	植物	动物
采样	植物组织 ; DNA	动物组织 ; 全血 ; DNA
总量	组织样总量 ≥ 10 g ; DNA (Qubit 检测) 总量 ≥ 30 ug , 同时 ≥ 50 ng/ul。	组织样总量 ≥ 5 g ; 血液 ≥ 15 mL ; DNA (Qubit 检测) 总量 ≥ 30 ug , 同时 ≥ 50 ng/ul。
运输条件	干冰运输 , 确保 72h 内送至百迈客实验室	

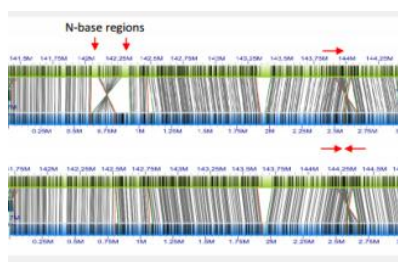
4 光学图谱辅助基因组组装

光学图谱的概念

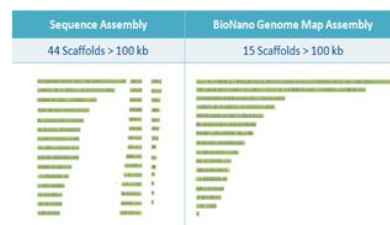
光学图谱技术是一个利用单个 DNA 分子基因组限制性内切酶图谱快速生成高分辨率、有序的全基因组限制性内切酶图谱的方法。现在最普遍是 BioNano Irys 系统，该系统利用内切酶对 DNA 进行识别酶切并标记荧光，再利用极细的毛细管电泳来把 DNA 分子拉直，进行超长单分子高分辨率荧光成像，通过酶切位点进行拼接即生成了一幅酶切位点分布图。

光学图谱的作用

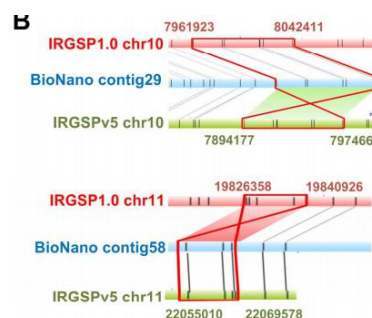
提升基因组 Scaffold 水平



纠正基因组的组装错误



多样本间的结构变异检测



光学图谱辅助基因组组装的优势

超高准确度

光学图谱构建所需要的 DNA 分子是最原始的，不经过扩增打断等处理，所以其反应的是物种 DNA 最真实状态；光学图谱的具有超长的读长，达到上百 Kb 或 Mb，远大于测序读长，可以跨过部分重复区域，所以在解决重复序列等组装难题时，具有很好的效果。。

提升 scaffold 水平

光学图谱可以在有基因组组装的情况下，构建或提高基因组 scaffold 水平。基因组组装质量越好，光学图谱的提升效果也就越好。

明确基因组 gap 大小

光学图谱在辅助基因组组装时，可以通过酶切标记固定基因组序列的位置，明确片段之间的 gap 大小。

光学图谱送样要求

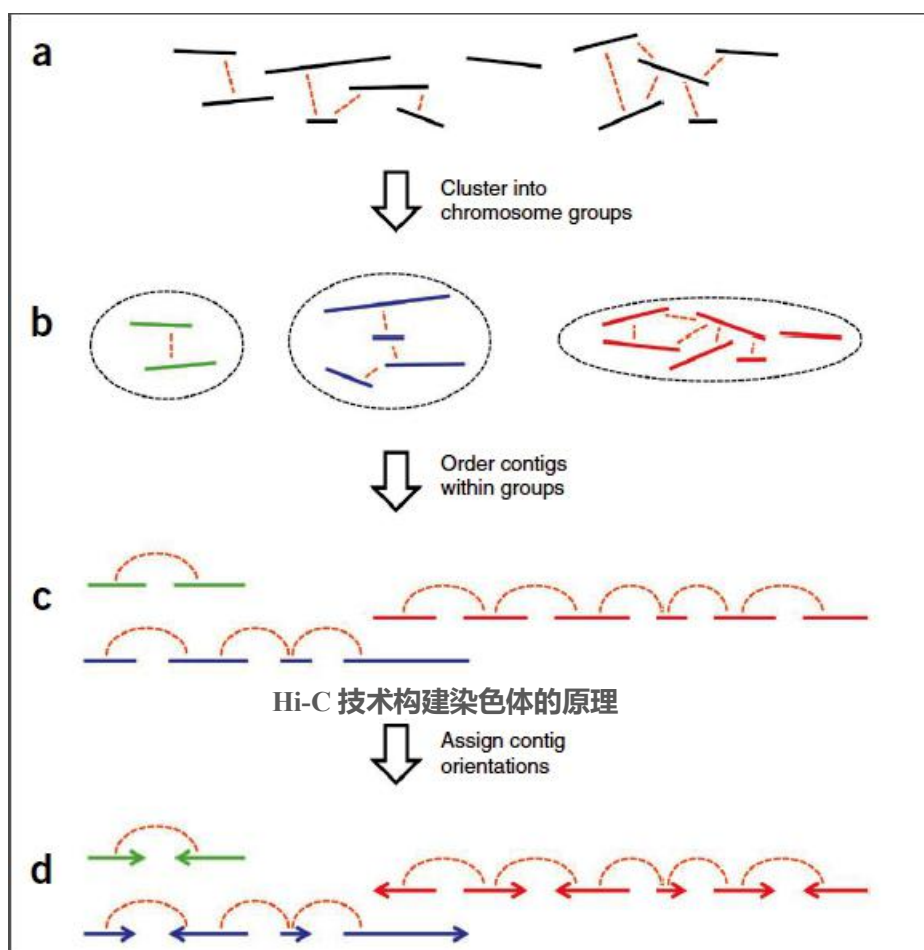
物种	植物	动物	节肢动物
采样	黄化饥饿、幼嫩组织	动物组织；全血	动物组织（需饥饿处理，置于离心管中，做好透气）
总量	组织样总量≥15 g	组织样总量≥15g；血液≥15ml	组织样总量≥500 mg
运输条件	干冰运输，确保 72h 内送至百迈客实验室		

5 Hi-C 染色体构建技术

Hi-C 技术的概念

Hi-C 技术是染色质构象捕获技术 (Chromosome conformation capture) 与高通量测序 (High-throughput sequencing) 结合衍生的一种技术。主要是利用全基因组范围内整个染色体 DNA 在空间位置上的关系, 对染色质内全部 DNA 相互作用模式进行捕获, 结合生物信息学方法, 来获得染色体水平的基因组序列。

Hi-C 技术的原理



Hi-C 技术构建染色体的优势

1. Hi-C 无需群体，单个体就能实现染色体定位

很多物种都无法构建遗传群体，包括大部分高等动物、野生动植物、林木、果树等等。Hi-C 是通过染色体上空间距离、线性距离的不同而导致的交互频率的不同来完成染色体的定位，所以不需要构建群体。

2. 标记密度更大，序列定位更完整

相比遗传图谱，染色质之间的交互频率具有更高的标记密度，如此高密度的图谱不仅可以挂载上长的 Scaffold，较短的 Scaffold 也可以被定位，所以通过 Hi-C 技术，一般可以将 90%以上基因组序列定位到染色体。

3. 可以对已组装的基因组进行纠错

通过 scaffold 间的交互频率大小，很容易确定 scaffold 在染色体上的排列顺序；同时，Hi-C 可以对已组装的基因组序列进行纠错。

4. 误差小，更可靠

直接通过单株系进行染色体定位，减少构建群体时所造成的误差，又不依赖于遗传群体的数量和质量，所以准确性更高，结果更可靠。

样品要求

Hi-C 采样前请联系百迈客技术人员，要求如下：



植物需要送活体幼苗，组织量在 10g 以上；

动物需要送全血样品，确保 48h 内送至百迈客实验室，需要 10ml 以上；



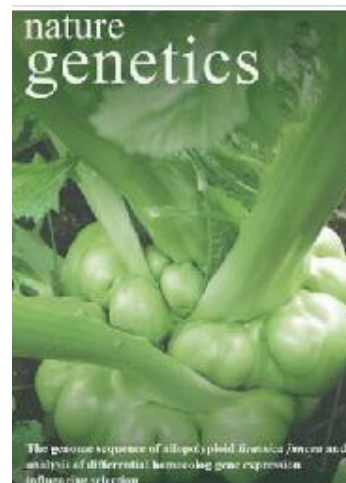
节肢动物需饥饿处理，置于离心管中，做好透气，24h 内送至百迈客，总量需在 500mg 以上。

百迈客基因组文章研究思路

芥菜基因组

期刊：Nature genetics

发表时间：2016. 09



研究背景

异源四倍体芥菜 (AABB) 属于十字花科芸薹属, 是重要的经济作物, 主要包括菜用和油用芥菜两大类群。芥菜是"禹氏三角"中重要的一员, 由白菜和黑芥杂交后加倍而来, 至少发生了三次古多倍化事件, 非常具有研究价值。本文成功组装出高质量的芥菜基因组图谱, 从多角度论证了芥菜 A 亚基因组起源问题, 为多倍体物种遗传育种提供了新的方向。

研究内容

1. 芥菜基因组组装

本文利用二代测序和三代测序相结合的方法进行芥菜基因组初步组装, 并利用光学图谱进行校正, contig N50 由 28 Kb 提升到 61 Kb, scaffold N50 由 710 Kb 提升到 1.5 Mb, 基因组完整性达到 85%。利用遗传图和光学图谱对 A、B 亚基因组进行区分, A 亚基因组挂载率为 91.48%, B 亚基因组挂载率为 72.32%。另外还利用二代测序技术组装了一版黑芥的基因组, 基因组大小为 591 Mb, 完整性为 68%。

2. 芥菜 A 亚基因组起源问题

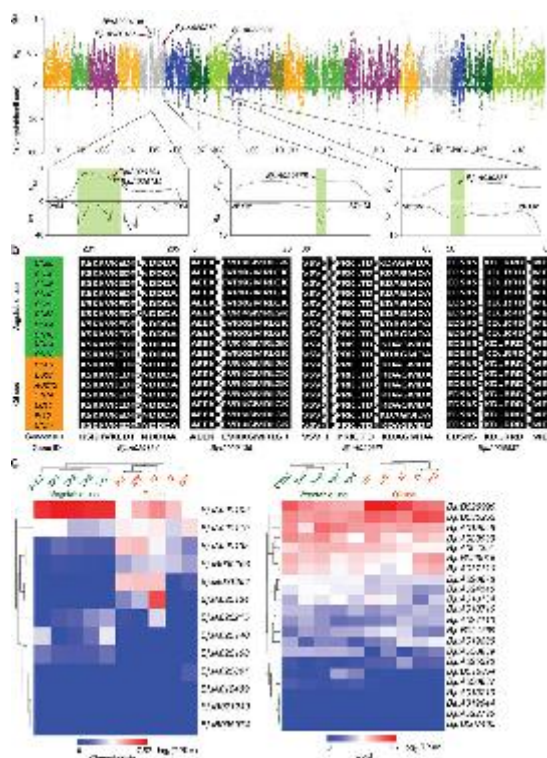
通过对 10 个菜用的芥菜、7 个油用的芥菜、5 个甘蓝型油菜和 27 个白菜 (多亚种) 进行重测序, 绘制进化树, 结果显示芥菜全部聚在一起, 没有出现分散的情况, 说明芥菜中 A 的基因组是来源于同一个亚种, 属于单系起源, 另外还从 PCA 聚类和 Fixed SNP 角度验证了单系起源的结论。

3. 基因表达的 dominance 现象

通过计算等位基因的表达量,在不同时期,不同组织之间,发现存在 dominance 基因,存在 dominance 的基因经受的选择压力大于 neutral 基因(不存在 dominance 现象,功能非常重要,纯化作用较强,不轻易突变),小于 subordinate 基因(作用不重要,纯化作用较小,易丢失)。

4. 油用芥菜和菜用芥菜的选择与分化

通过对菜用和油用芥菜群体进行选择清除分析,发现菜用和油用芥菜中 36% 的基因存在受选择区域 (HED)。dominance 的基因被筛选出来的比例较高,结合转录组数据,这部分基因在油用和菜用两个群体中差异表达,同时发现与硫苷,脂类代谢显著相关并且存在 dominance 的基因组,这些基因在油用菜用群体中有各自独特基因分型且受选择基因 (HED) 与基因分型和表达关联。



芥菜选择清除分析

参考文献:

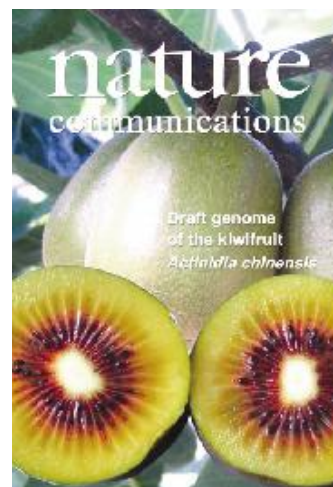
Jinghua Yang, Dongyuan Liu, Xiaowu Wang, et al. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection[J]. *Nature Genetics*, 2016.

百迈客基因组文章研究思路

猕猴桃基因组

期刊：Nature communications

发表时间：2013.05



研究背景

猕猴桃是杜鹃花目蔓生木本植物，果实具有较高的营养价值和良好口感。猕猴桃起源于中国，在 20 世纪初被作为一种优秀的经济作物广泛传播到世界各地。由于木本植物基因组杂合率远高于其他物种，猕猴桃基因组的杂合率高达 1.05% (绝大多数生物基因组杂合率<0.5%)，测序和组装存在很大难度；百迈客成功构建高质量的猕猴桃全基因组图谱，为猕猴桃的品种改良和遗传进化研究奠定了理论基础。

研究内容

1. 猕猴桃基因组组装

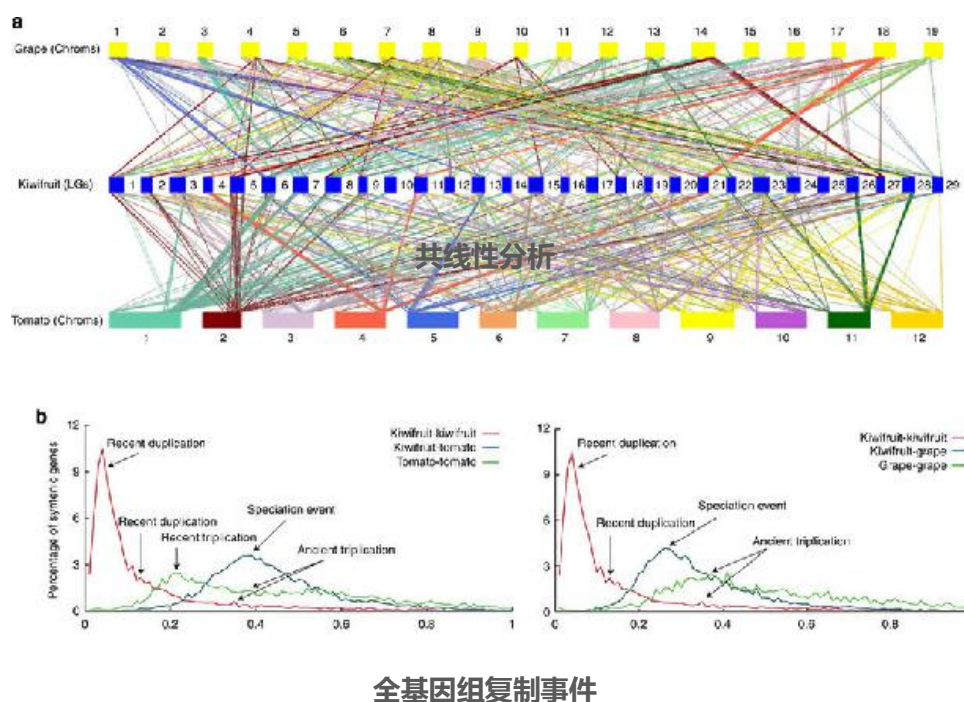
以中华猕猴桃 (*Actinidia chinensis*) 为材料，全面解析了首个猕猴桃全基因组序列信息，采用 Illumina Hiseq2000 测序平台，总数据量为 105.6Gb，覆盖度为 140 ×。并结合遗传图谱，猕猴桃基因组组装达到染色体水平，初步评估得到基因组大小为 758 Mb，组装得到基因组大小为 616.61Mb，contig N50 为 58.8 kb，scaffold N50 为 646.8 kb。

2. 比较基因组分析

研究同时比较分析了猕猴桃与西红柿、水稻、葡萄和拟南芥的基因组信息，揭示出大约 8000 种基因共同存在于这 5 个物种中，并描绘了包括与生长、成熟、营养代谢和抗病性相关的发育基因的进化关系。

3. 全基因组复制事件研究

基于全面完整的全基因组序列信息，科学家们进一步观察到猕猴桃 DNA 存在高比例的相似性，进而揭示出在大约 2700 万和 8000 万年前，猕猴桃在细胞分裂过程中发生了两次不同寻常的重大变异事件：在基因组发生加倍事件导致广泛的基因扩增之后，又发生了大规模的基因缺失。在基因复制过程中，额外的基因频繁发生突变，产生了猕猴桃全新的功能。参与调控猕猴桃重要经济特征，例如水果维生素 C、类黄酮和类胡萝卜素代谢的基因大量增加。



猕猴桃作为杜鹃花目首个被全面解码基因组信息的植物，为比较基因组学和进化研究提供了宝贵的理论基础，也为猕猴桃的种质资源保护和性状改良提供了必要的研究依据，具有很高的学术和应用价值。

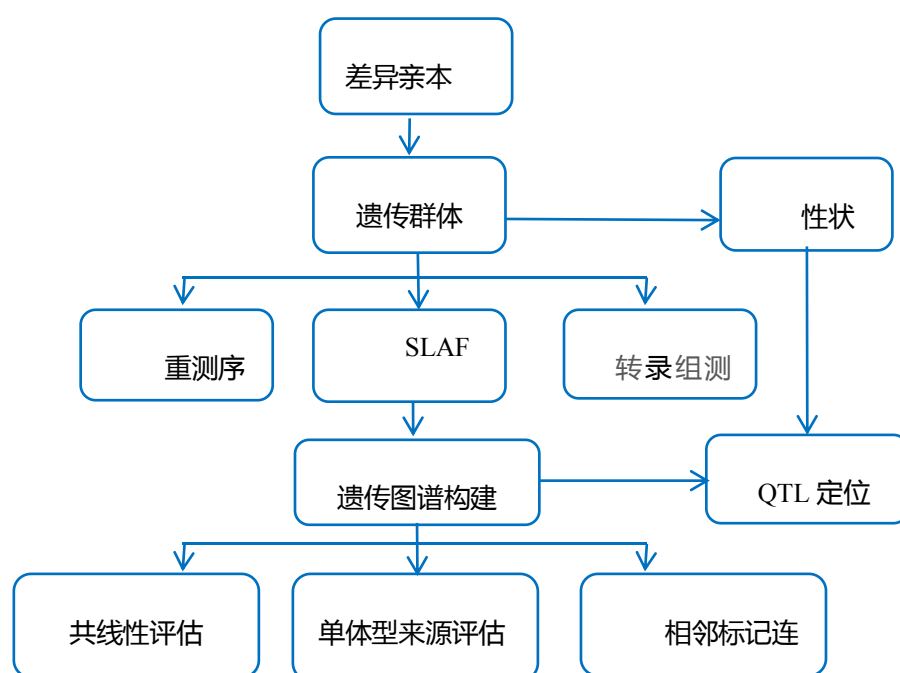
参考文献:

Huang S, Ding J, Deng D, et al. Draft genome of the kiwifruit *Actinidia chinensis*[J]. Nature Communications, 2013.

群体

遗传图谱简介

遗传图谱是指依据分子标记在染色体上的重组值(或交换值),将染色体上的各个标记之间的距离和顺序标志出来,绘制而成的图谱。采用类似单基因定位的方法将 QTL 定位在遗传图谱上,确定 QTL 与遗传标记间的距离。



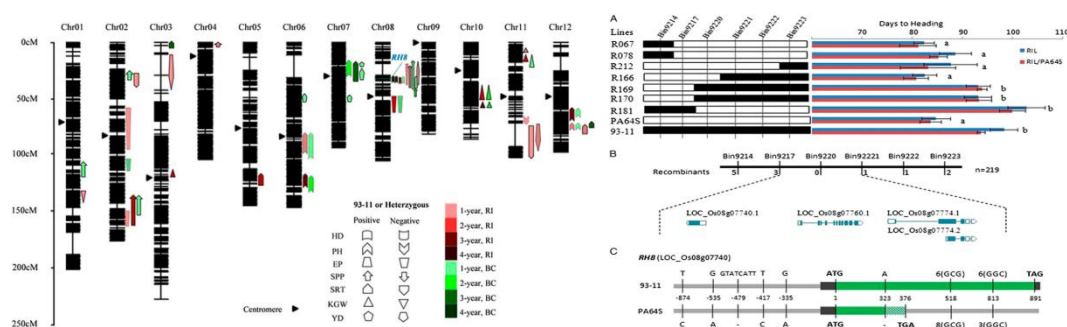
技术优势

方案灵活	根据物种情况和研究目的设计标记开发方案, 在最优性价比基础上完成研究
图谱质量高	百迈客自主研发的 HighMap 软件, 标记数量多, 容错能力更强, 图距更真实
项目经验丰富	完成图谱项目 300+, 个性化分析 1000+, 发表文章 80 余篇, 研究涉及作物、水产、林木花卉、蔬菜等 50 余个物种
周期短	项目从建库测序、图谱构建到性状定位仅需 70-90 天, 加速功能基因研究
云平台数据深入挖掘	客户通过百迈客云平台自主进行基因数据深入挖掘、QTL 定位、引物设计等
后期验证	百迈客推出 KASP 基因分型验证服务, 助您一步到位拿到候选基因

百迈客遗传图谱优秀案例分享

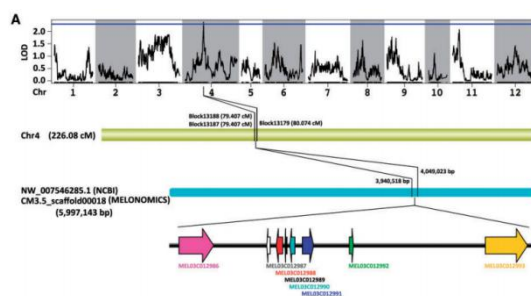
案例一 多组学联合分析定位多个水稻产量杂种优势位点

杂种优势指的是杂交子代在表型上超越其亲本的现象，在本研究中，作者利用整合分析的方法，发现了多个数量性状位点，通过调节两个产量组成性状（每穗小花数、有效穗数）造成两系杂交稻产量优势，同时，结合转录组分析以及 QTL 定位的结果，确定了多个与杂种优势相关候选基因。



案例二 重测序遗传图谱解析甜瓜抗蔓枯病机理

蔓枯病是葫芦科的一种由子囊菌引起的破坏性病害，对甜瓜生产造成严重影响。近期，《DNA Research》在线发表了由百迈客和浙江大学杨景华教授课题组协作完成的甜瓜重测序遗传图谱文章，该文章利用高密度遗传图谱，成功定位到了控制甜瓜抗蔓枯病的候选基因，并对该基因的功能进行了初步验证。



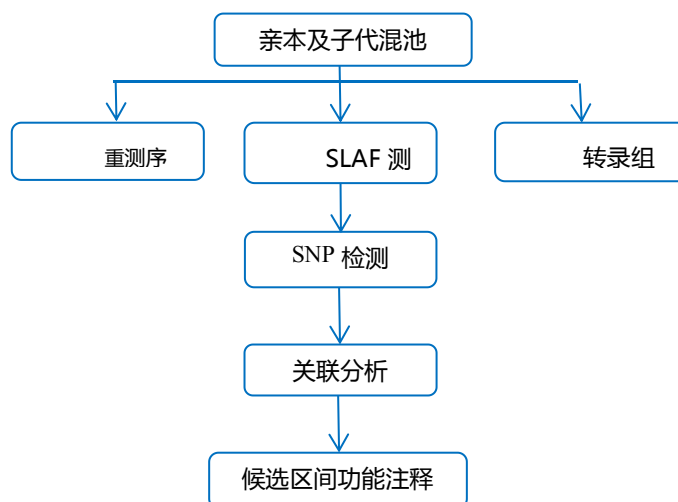
参考文献

1. Li D, Huang Z, Song S, et al. Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase.[J]. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(41):E6026.
2. Hu Z Y, Deng G C, Mou H P, et al. A re-sequencing-based ultra-dense genetic map reveals a gummy stem blight resistance-associated gene in Cucumis melo[J]. DNA Research, 2017.

BSA 简介

将高通量测序技术与集群分离分析方法相结合的新一代基因精细定位研究策略。通过对极端性状混池进行高通量测序，再筛选与目的性状紧密连锁的 SNP 和 InDel，进行目的基因的快速、精细定位。

BSA 技术路线



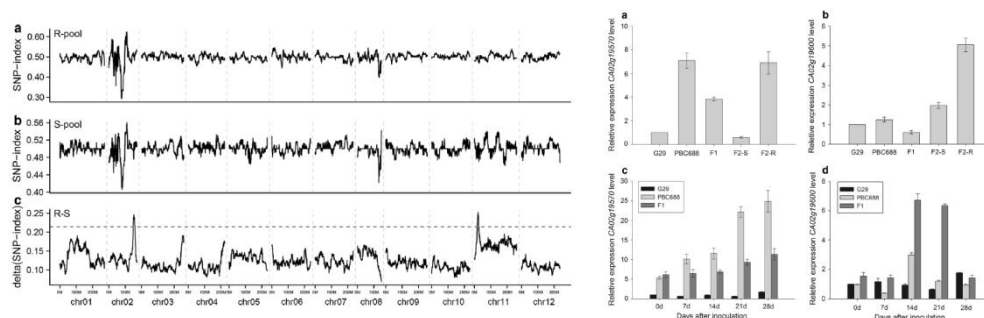
BSA 技术优势

适用范围广	有无参考基因组均可
混池规模大	从 30+30 到 200+200 不等，混池越大，背景干扰越小
项目经验丰富	BSA 发表文章 40+，累计影响因子 100+
方案设计灵活多变	从 DNA 到 RNA，从全基因组测序到简化测序均可
项目周期短	仅需 45 天，快速高效
云平台数据深入挖掘	百迈客云平台自主实现 BSA 分析、阈值调整、引物设计等
后期验证有保证	百迈客提供 KASP 基因分型验证服务，一步到位拿到候选基因

百迈客 BSA 优秀案例分享

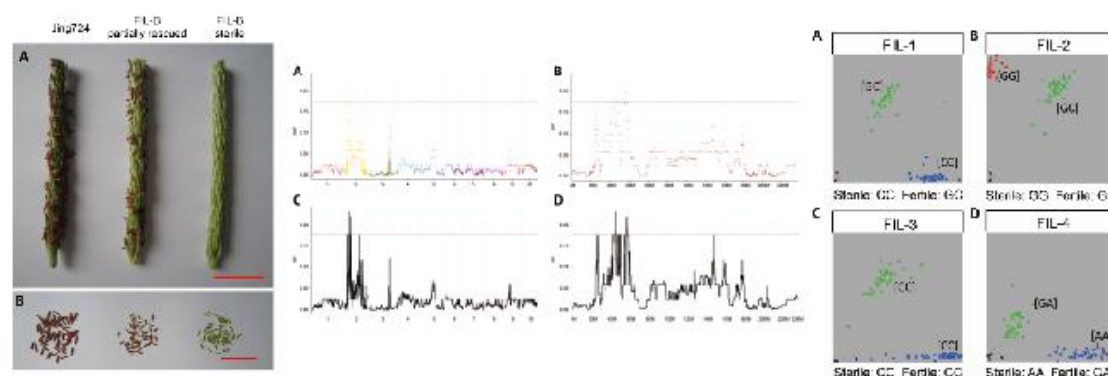
案例一 SLAF-BSA 定位辣椒抗黄瓜花叶病毒 (CMV) 基因

本研究以 PBC688 (高抗 CMV) 和 G29 (易感 CMV) 为亲本, 构建 BC1 和 F2 群体, 选取抗性极端的个体分别构建混池, 基于 SLAF 标签进行 BSA 关联分析, 在 2 号和 11 号染色体上分别定位到 4Mb 的区间, 结合局部遗传图谱分析, 最终把定位区间缩小到 330Kb 的区间, 在该区间内有多候选基因, 其中 2 个基因与抗 CMV 极为相关。



案例二 转录组 BSR 定位玉米雄性不育基因

Jing724 和 MD32 (CMS-S: S 型细胞质雄性不育系) 为亲本构建的 BC4 群体, 对完全雄性不育和育性部分恢复混池进行 RNA 测序, 获得混池间的差异基因数目 3672 个, 利用 BSA 分析共得到 5 个候选区域, 结合数据库注释结果及转录组差异基因的结果, 最终获得 12 个候选基因, 并利用 KASP 对候选基因进行了验证。



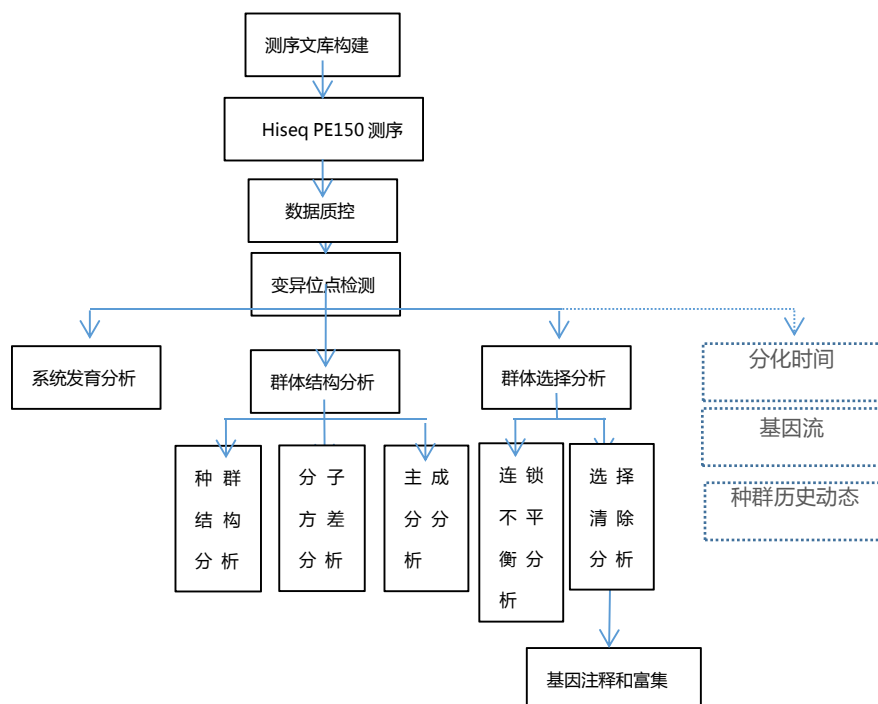
参考文献

- Guo G, Wang S, Liu J, et al. Rapid identification of QTLs underlying resistance to Cucumber mosaic virus in pepper (*Capsicum frutescens*) [J]. Theoretical & Applied Genetics, 2016:1-12.
- Su A, Song W, Xing J, et al. Identification of Genes Potentially Associated with the Fertility Instability of S-Type Cytoplasmic Male Sterility in Maize via Bulk Segregant RNA-Seq [J]. Plos One, 2016, 11(9):e0163489.

遗传进化简介

通过对代表性材料（建议最少两个亚群以上，每个亚群 ≥ 10 份个体或混池）进行全基因组重测序，检测 SNPs 和 Indels 等变异信息并进行群体遗传学分析，如连锁不平衡分析、系统发育分析、群体结构分析、有效群体大小分析、基因流分析等。

遗传进化技术路线



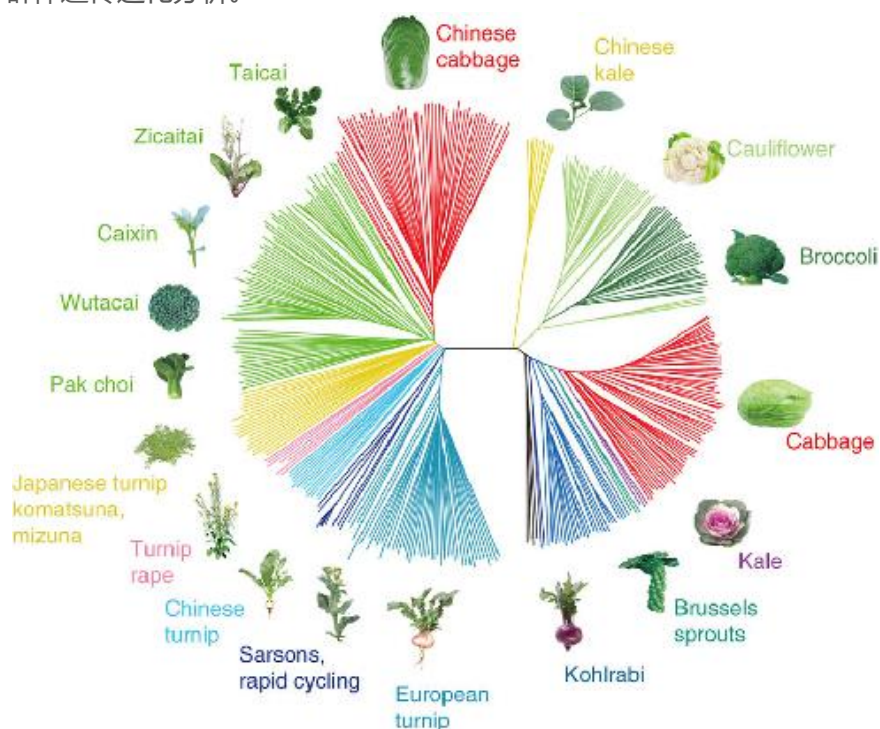
送样要求

样品	DNA	RNA
完整性	主带清晰，无降解或轻度降解，无 RNA 污染	28s 和 18s 条带清晰，且 28s 亮度大于 18s 的两倍，植物 RIN 值 ≥ 6.5 ，非植物 RIN 值 ≥ 7.0 ，28S/18S ≥ 1.0
总量	DNA 总量 $\geq 2\mu\text{g}$ ，同时 $\geq 15\text{ng}/\mu\text{l}$	RNA 总量 $\geq 3\mu\text{g}$ ，同时 $\geq 40\text{ng}/\mu\text{l}$
纯度	OD260/280 为 1.7~2.2	OD260/280 为 1.7~2.5，OD260/230 为 0.5~2.5

百迈客遗传进化案例分享

文章通过对不同形态类型的白菜和甘蓝类蔬菜的基因组重测序,构建了白菜类和甘蓝类蔬菜的群体基因组变异图谱,确定了许多与白菜和甘蓝叶球形成与膨大根(茎)驯化选择的基因组信号和相关基因。在结球的大白菜和甘蓝中,分别发现了 19 个和 16 个参与叶球形成的关键基因位点。叶片背面和腹面分别由决定背性和腹性两类不同极性的基因控制,且叶片背性和腹性两类极性控制基因参与了叶球的形成。而在根膨大的芜菁和茎膨大的茎蓝中,各确定了 7 个与根(茎)膨大有关的基因位点。膨大根(茎)形成过程中,除生长素相关基因受选择外,细胞快速膨大和糖转运蛋白相关基因也受到高度选择。研究首次发现,结球白菜和结球甘蓝,芜菁和茎蓝等不同物种产生类似产品器官的现象,与其共同祖先的在 1100 万年前发生的一次全基因组三倍化事件有关。本研究获得了白菜和甘蓝类蔬菜作物全基因组的大量变异,确定了一些与白菜甘蓝类蔬菜叶球形成和根(茎)膨大有关的重要基因,为加快白菜与甘蓝类蔬菜分子育种奠定了重要基础。

研究策略：全基因组重测序, 13 种形态型的白菜种 *B. rapa* 共 199 份, 9 种形态型的甘蓝种 *B. oleracea* 共 119 份, Hiseq2000 测序, 插入片段 350bp, 测序深度 $\geq 8\times$, 并基于 SNP 标记进行群体遗传进化分析。



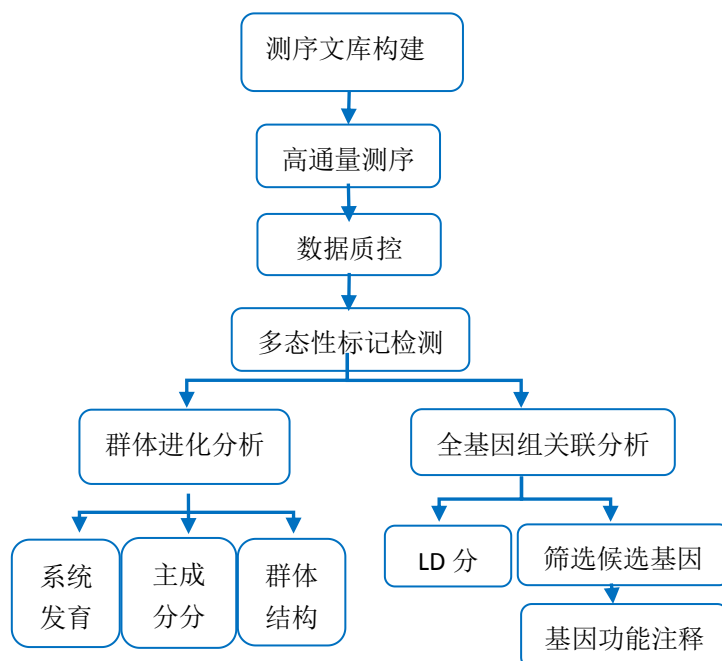
参考文献

Cheng F, Sun R, Hou X, et al. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea* [J]. *Nature Genetics*, 2016.

GWAS 简介

通过全基因组大样本重测序对动植物重要种质资源进行全基因组的基因型鉴定,并与关注的表型数据进行全基因组关联分析 (GWAS), 进而找到与关注表型相关的 SNP 位点,定位与性状相关基因,为后续动植物的育种提供科学理论依据。

GWAS 技术路线



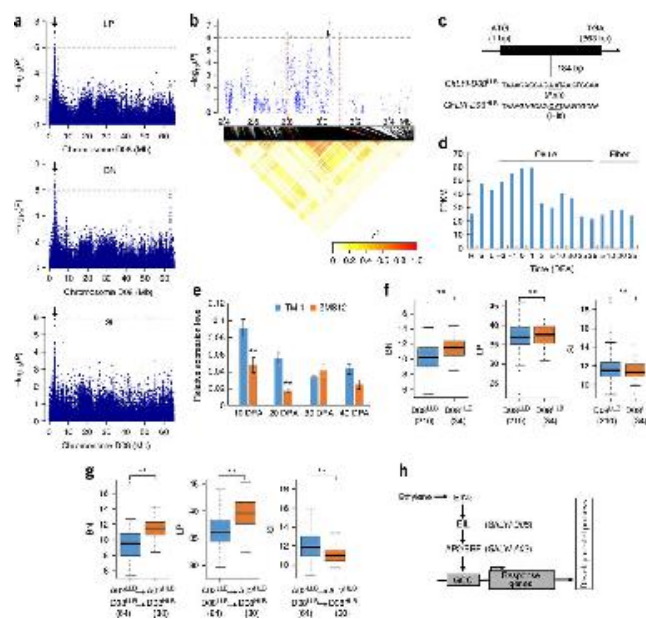
送样要求

样品	DNA	RNA
完整性	主带清晰,无降解或轻度降解,无 RNA 污染	28s 和 18s 条带清晰,且 28s 亮度大于 18s 的两倍,植物 RIN 值 ≥ 6.5 ,非植物 RIN 值 ≥ 7.0 , 28S/18S ≥ 1.0
总量	DNA 总量 $\geq 2\mu\text{g}$,同时 $\geq 15\text{ng}/\mu\text{l}$	RNA 总量 $\geq 3\mu\text{g}$,同时 $\geq 40\text{ng}/\mu\text{l}$
纯度	OD260/280 为 1.7~2.2	OD260/280 为 1.7~2.5, OD260/230 为 0.5~2.5

GWAS 案例分享

棉花纤维是重要的天然纺织纤维，占据着重要经济地位。中国陆地棉栽培始于 20 世纪初，主要从美国引种，但我国陆地棉遗传资源狭窄，同质性较高，导致遗传育种工作遇到了瓶颈。本研究对 318 份棉花（35 份地方种+258 份栽培种）进行全基因组重测序，对地方种和栽培种棉花的核苷酸多态性分析获得 25 个受选择区域，对 258 份栽培种棉花的关联分析得到 119 个与产量（71 个）、纤维品质（45 个）、黄萎病抗性（3 个）相关的位点，其中，陆地棉 A 亚基因组更多的是产量、纤维品质相关的位点，D 亚基因组更多的是与环境适应性相关的位点。把 GWAS 关联到的位点和受选择区域进行比较，发现有 4 个受选择区域存在与棉花产量、纤维长度和纤维硬度相关的位点。此外，研究人员鉴定出能同时提高 2-3 个产量性状或纤维品质性状的关联基因，为棉花“精准育种”提供了优异的基因资源和理论指导，有非常重要的利用价值。

研究策略：全基因组重测序，35 份地方种棉花，258 份栽培种棉花，Hiseq 2500 测序，测序深度 $\geq 5\times$ ，并基于 SNP 标记进行关联分析。陆地棉 TM-1 and ZMS 12，Hiseq 2000 测序，RNA-Seq，单个文库 ≥ 4.6 Gb。



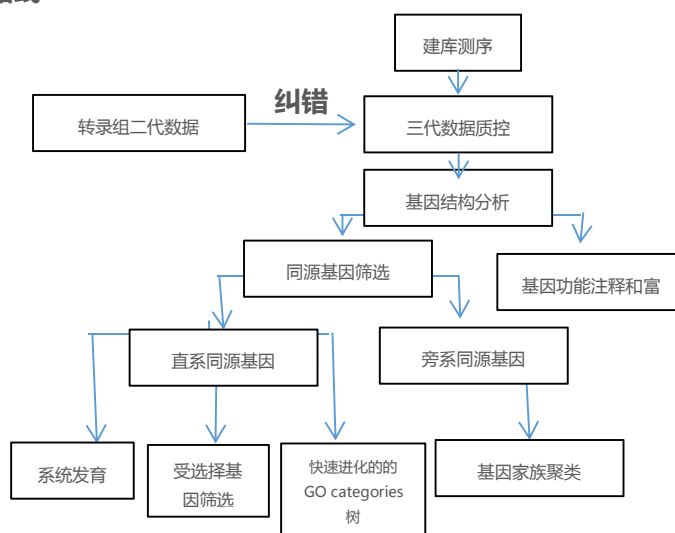
参考文献

Fang L, Wang q, Hu y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits[J]. Nature Genetics, 2017.

全长比较转录组简介

基于转录组测序的全长比较转录组是通过对同一科、属内不同物种，尤其是复杂的或者无参考基因组的物种，提取 mRNA 后进行 PacBio 三代测序，从基因序列水平研究起源于共同祖先的不同物种间的系统发育关系以及分化时间等进化事件，同时挖掘明显受到正向选择或者负向选择的基因，从而快速、全面地揭示选择和适应机制。通过二代测序对三代序列进行校正后，不需要组装即得到了高质量的全长转录本，从而避免了由于拼接引入的错误，分析更精准。

全长比较转录组技术路线



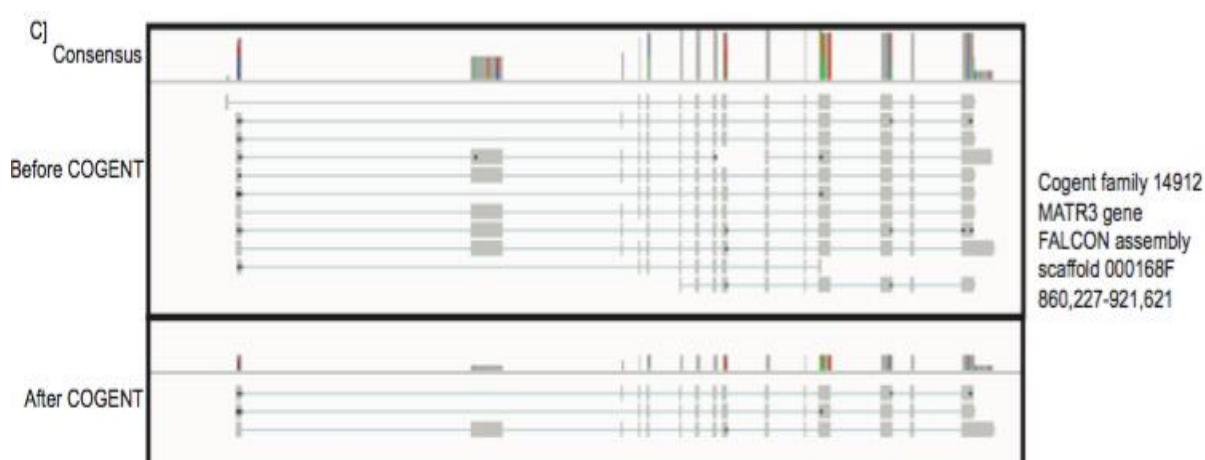
送样要求

测序策略	二代	三代
完整性	28s 和 18s 条带清晰，且 28s 亮度大于 18s 的两倍，植物 RIN 值 ≥ 6.5 ，非植物 RIN 值 ≥ 7.0 ，28S/18S ≥ 1.0	28s 和 18s 条带清晰，且 28s 亮度大于 18s 的两倍，植物 RIN 值 ≥ 7.5 ，非植物 RIN 值 ≥ 8.0 ，28S/18S ≥ 1.3
总量	DNA 总量 $\geq 3\mu\text{g}$ ，同时 $\geq 40\text{ng}/\mu\text{l}$	RNA 总量 $\geq 3\mu\text{g}$ ，同时 $\geq 300\text{ng}/\mu\text{l}$
纯度	OD260/280 为 1.7~2.5，OD260/230 为 0.5~2.5	OD260/280 为 1.7~2.5，OD260/230 为 0.5~2.5

全长比较转录组案例分享

文章运用 SMRT 测序技术进行蜂鸟全长转录组分析。文中作者提到“蜂鸟能够在短时间内对任意一种能源代谢进行切换，这一过程具有显著的灵活性。”并且“储存式脂肪产生的原则就是由日常糖前体的摄入到脂质的从头合成。有时在即将迁徙飞行的前几天，也能够很快的进行这一过程。该文章在没有参考基因组的情况下，“运用全长转录本数据，发现得到基因 isoform 数据的清晰全图已然没有必要进行比对。”作者还提到，“长读长使得转录本完全不需要进行组装，可以在没有基因组的情况下进行分析。”接近一半的读长在最终分析时覆盖了包括 5' 端和 3' 端的全长基因，例如分析到了 polyA 尾的信息。针对同一基因家族的转录本，并关注于那些独特的 isoform，作者们还运用 COGENT 方法对转录本序列进行分析。这一分析结果呈现了一系列高度多样性的 isoform，正如作者所说“这是蜂鸟肝脏的中，更接近完整转录组的分析结果。”在这样一系列的数据结果中，科学家们发现了一些蜂鸟独特的基因。作者指出，“在脂类代谢途径上，这些基因表现出特异的富集，说明蜂鸟通过这些基因的不同异构体，达到高水平的代谢效率”。

研究策略：PacBio RSII 平台，构建了 1-2kb，2-3kb，3-6kb，5-10kb 四种片段大小的文库，COGENT 分析。



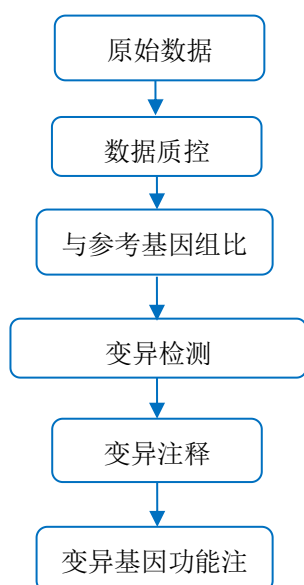
参考文献

Rachael E. W, Alexander M. M, Elizabeth T, G. et al. Single molecule, full-length transcript sequencing provides insight into the extreme metabolism of ruby-throated hummingbird *Archilochus colubris*[J]. BioRxiv, 2017.

个体重测序简介

基于全基因组重测序的变异图谱通过测序手段结合生物信息分析研究同一物种不同个体之间的变异情况，获得大量的变异信息，如 SNP、Indel、SV 等。主要可以快速地获得大量的分子标记以及不同个体在基因组水平上的差异。

个体重测序技术路线



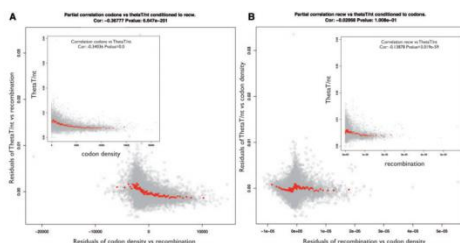
送样要求

样品	DNA
完整性	主带清晰，无降解或轻度降解，无 RNA 污染
总量	DNA 总量 $\geq 2\mu\text{g}$ ，同时 $\geq 15\text{ng}/\mu\text{l}$
纯度	OD260/280 为 1.7~2.2

个体重测序案例分享

本文对七个甜瓜品种进行个体重测序，鉴定与甜瓜重要性状有关的 SNP、SV、TE 等遗传变异。这七个甜瓜品种的遗传学变异影响了 902 个基因。研究人员预测了其中 53 个基因的功能，发现它们在抗病性、细胞壁代谢和糖代谢中起作用。该研究揭示了甜瓜基因组的高度可塑性，可以帮助人们更好地进行育种，提升这种水果的品质或者它的抗虫抗病能力

研究策略：重测序，测序深度 $\geq 19.45\times$ 。



参考文献

Rachael E. W, Alexander M. M, Elizabeth T, G. et al. Single molecule, full-length transcript sequencing provides insight into the extreme metabolism of ruby-throated hummingbird *Archilochus colubris*[J]. Molecular Biology and Evolution, 2015.

转录调控

百迈客云服务

百迈客云平台是由北京百迈客生物科技有限公司开发，集成生物信息分析软件、数据库以及云计算为一体的生物数据信息分析平台。

百迈客云服务是基于传统科技服务项目的基础上，利用云平台交互性、可视化、可实时查看的优势，提供标准分析及

数据深度挖掘方案以及配套的云平台数据挖掘工具。

百迈客云服务，为您提供从建库测序到文章发表的一站式服务，让您离高分文章更近一步！

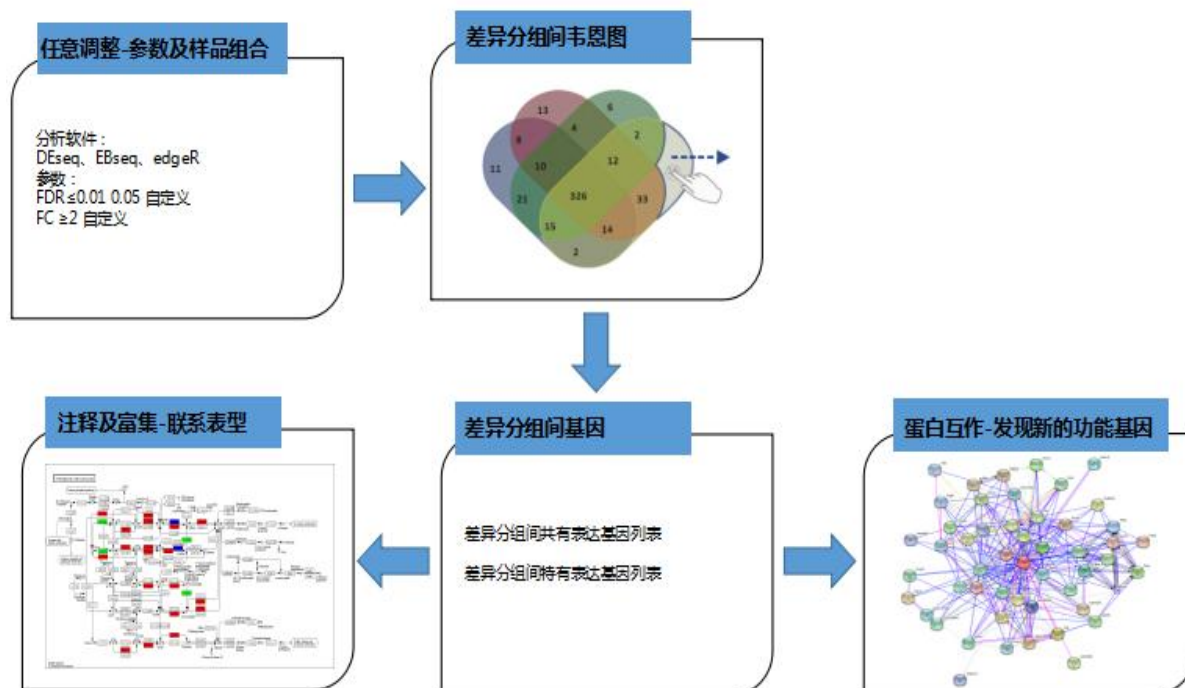


云端结题报告和数据可视化呈现，项目进度实时查看、沟通和评价及云定制化 8 小时交付

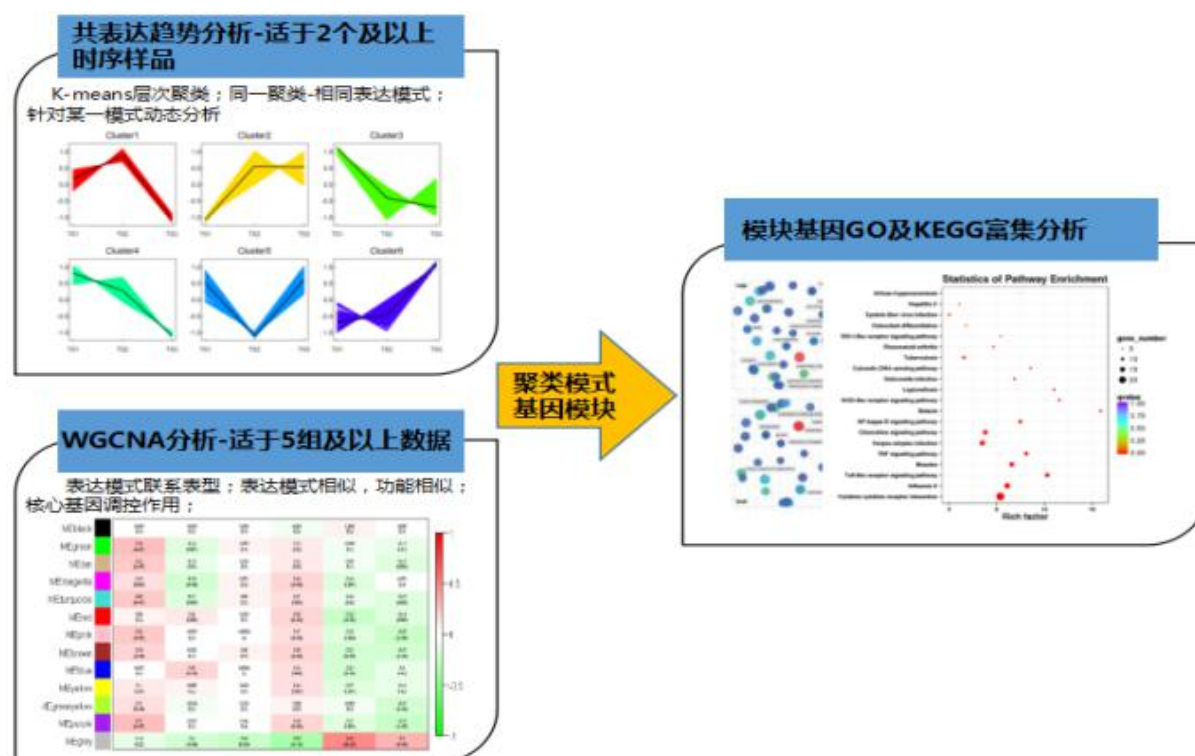


数据深度挖掘方案

方案 1：经典数据挖掘思路-基于差异表达基因组间的数据



方案 2：热点数据挖掘-基于共表达趋势分析 & WGCNA 分析



真核转录组 云次账号—SMART

转录组测序技术主要是针对转录产物 mRNA 进行高通量测序，可全面快速的获得某一

物种特定组织或器官在某一状态下的 RNA 的序列变化与表达差异信息，是基因功能及结构研究的基础，目前已广泛应用于生物学基础研究等多个领域。

云次账号依托百迈客转录调控 8 年项目经验及云服务，在常规有参转录组项目的基础上去除了 SNP/Indel 和新基

因的分析，开发出了一个极速的 smart 模式，小于 100G 的数据在 24h 内就能够分析完成！！

转录组+云次账号，助力文章的快速发表！



成功案例 通过转录组分析合浦珠母贝同种异体移植血细胞中免疫相关差异表达基因的鉴定

Wei J, et al. Differentially expressed immune-related genes in hemocytes of the pearl oyster *Pinctada fucata* against allograft identified by transcriptome analysis[J]. Fish & Shellfish Immunology, 2017, 62:247-256. (IF=3.05)

本文通过对异体移植后的珍珠贝血淋巴进行转录组测序，通过比较不同移植时期样品的基因差异表达情况，阐明了珍珠免疫的分子机制。短短 5 个月的时间，中国水产科学院南海水产研究所的老师利用云平台，第二篇对珠母贝异种异体移植后免疫相关研究成果再次被该杂志发表。转录组+云次账号助力文章快速发表！！

全转录组

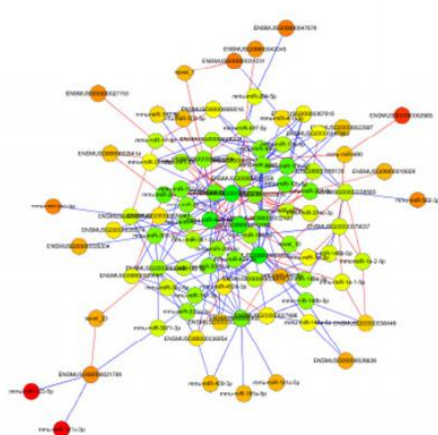
非编码 RNA 测序

小 RNA 是一类长度在 18~30nt，能够通过与靶 mRNA 特异性结合来调控其表达的内源性非编码 RNA。小 RNA 测序能够快速对小 RNA 进行鉴定、特征分析、表达分析及其靶基因分析。

长链非编码 RNA(lncRNA)是一类转录本长度>200nt 的功能分子，能够在转录及转录后水平上调节蛋白编码基因的表达。lncRNA 测序能够快速对 lncRNA 进行鉴定、特征分析、表达分析及其靶基因分析。

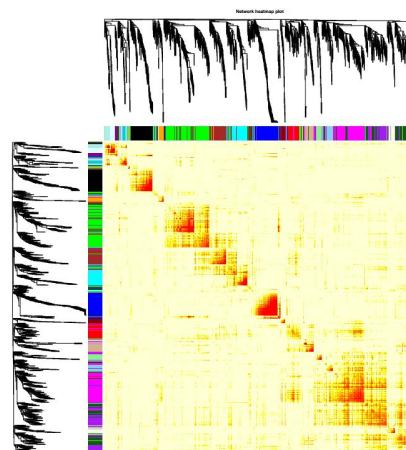
circRNA 是一类特殊的非编码环状 RNA，能够通过结合 miRNA 来解除 miRNA 对靶基因的抑制作用，还能够调控其它类型 RNA 的表达。circRNA 测序可以快速对 circRNA 进行鉴定、特征分析、表达分析及 miRNA 结合位点分析。

**联合测序： 转录组测序+小 RNA 测序联合分析
lncRNA 测序+小 RNA 测序联合分析**



miRNA-mRNA 调控网络分析

**转录组测序+lncRNA 测序联合分析
circRNA 测序+小 RNA 测序联合分析**



mRNA 与 lncRNA 共表达网络分析

全转录组测序

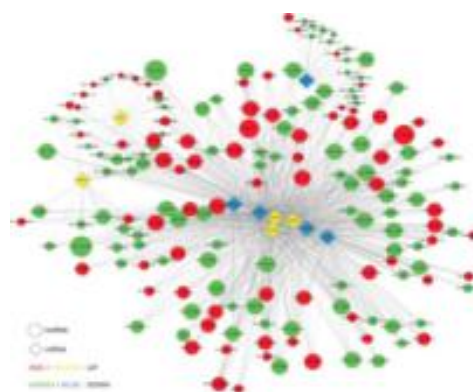
不同类型的 RNA 分子都是相互联系，相互作用的。单一的 mRNA 或 ncRNA 研究已无法满足科研需求，需要结合多种 RNA 信息进行整合分析，探索潜在的调控网络机制。而全转录组测序不仅可以对 mRNA 或 ncRNA 分别进行研究，更能够对多种 RNA 进行联合分析，研究调控网络，无疑是阐释生物学问题的利器！

联合分析内容

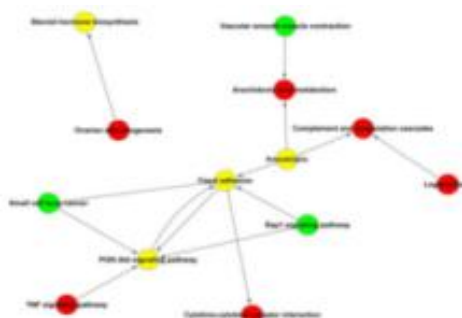
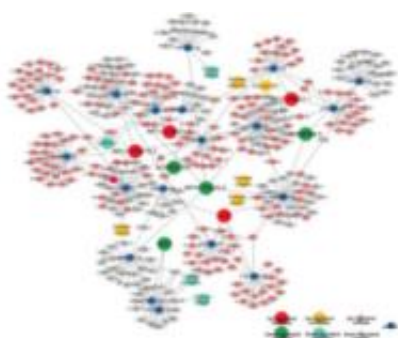
1. RNA 数目比例统计，各个样品的全局的表达量展示
2. 差异 RNA 的全局展示
3. 共表达网络
 - lncRNA_mRNA 共表达
 - miRNA_circRNA 共表达
4. KEGG 网络整合
5. ceRNA 网络背景构建，差异 RNA 子网提取整合
6. 共表达网络与 ceRNA 整合提取子网
7. 网络随机游走，提取关键 RNA，并提取子网
 - 对子网中 mRNA 进行功能注释
 - 对关键 lncRNA 绘制靶基 因关系图



从整体水平上观察不同样本 RNA 表达



miRNA-circRNA/mRNA-lncRNA 相关性网络



转录组与蛋白质组联合分析

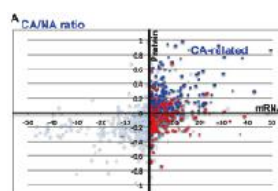
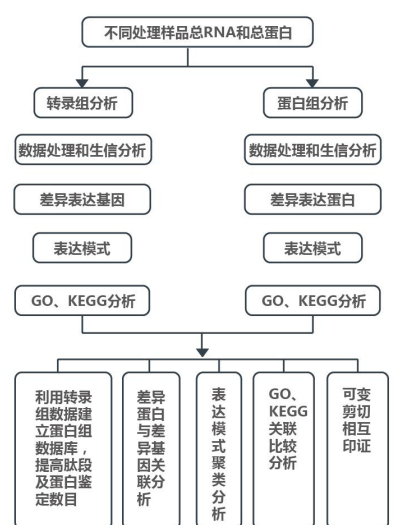
遗传物质从 DNA、RNA 到蛋白质的过程，受到不同水平的表达调控，如表观调控、转录调控、翻译调控等。mRNA 的研究只能在转录调控水平上研究，结合蛋白质组学研究可以比较分析基因在转录和翻译不同水平的调控机制。蛋白质组学研究，即把一个样品中表达的大多数蛋白质进行精确的定量和鉴定。

iTRAQ 和 label-free 是目前定量蛋白质组学常用的技术。iTRAQ 技术是由 AB Sciex 和 Thermo 公司研发的蛋白体外标记定量技术，利用多种同位素特异性标记蛋白的氨基集团，进行质谱串联分析，比较不同样品蛋白质的相对含量。非标定量法 (label-free) 是一种不依赖于同位素标记的新型蛋白质定量技术，通过比较质谱分析次数或质谱峰强度，分析不同样品蛋白的数量变化。

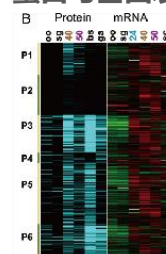
定量蛋白质组技术方法比较

技术方法	优点	缺点
iTRAQ	- 蛋白通量高，覆盖度高 - 定量准确，可信度高 - 分离能力强、分析范围广，定性结果可靠 - 自动化程度高、分析时间短，分离效果好	- 需同位素标记，成本较高 - 样本数受限制，只可平行比较 2-8 个样本 - 样本处理过程容易引入误差
Label free	- 实验成本低，无需同位素 - 样本数不受限制，适用大样本 - 对样本的操作较少，所需样本量低 - 检测肽段范围宽，有利于检测低丰度蛋白	- 数据处理复杂 - 对质谱结果稳定性依赖高，影响重复性和定量准确性

转录组和蛋白组联合分析路线



蛋白与基因表达水平比较分析



蛋白与基因表达模式聚类分析

转录组与代谢组联合分析

生物体的表型由 DNA、RNA、蛋白质及代谢物等多个层面调控决定，mRNA 研究可以获得样本之间差异基因、复杂调控网络，而基因与表型不能直接关联，难以鉴定控制关键途径的结构及调控基因，结合代谢组，可以从大量转录组信息中挖掘差异基因，快速鉴定核心调控网络，确定关键候选基因。代谢组学研究，是对生物体内所有小分子代谢物进行定性和定量。

非靶向代谢组和广泛靶向代谢组是目前代谢组学常用的技术。非靶向代谢组是将对照组和实验组样本的全部代谢物进行比对，以找出其代谢物的差异。广泛靶向代谢组是高通量、高灵敏度靶向代谢物检测技术，提供了定性、定量检测大批量、低丰度代谢物高效的方法，有助于全面解析差异代谢物与代谢途径。

代谢组技术方法比较

技术方法	优点	缺点
非靶向代谢组	■检测 ■通量高，操作简单	■难以准确鉴定代谢物 ■灵敏度低、难检测低丰度代谢物 ■依赖公共数据库，只能检测数据库已知代谢物，难以发现新代谢物
广泛靶向代谢组	■批量，准确鉴定代谢物 ■高灵敏度，发现低丰度代谢物 ■自建代谢物数据库，既检测已知代谢物 ■也发现新代谢物，同时实现相对定量	■需自建代谢物数据库 ■进样处理复杂，检测过程较繁琐

主要分析内容

- 1、数据质控
- 2、代谢物定性定量分析；
- 3、差异代谢物分析；
- 4、差异代谢物聚类分析；
- 5、差异代谢物 KEGG 富集分析；

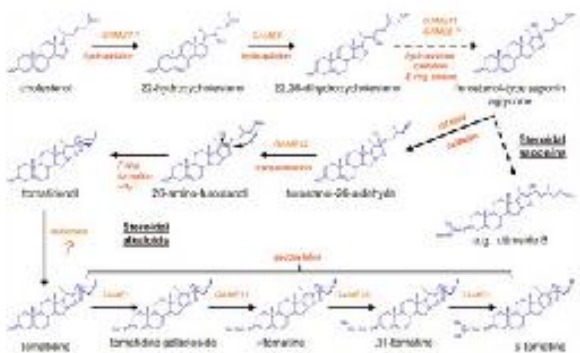
与转录组联合分析

- 1、差异表达基因与差异代谢共表达分析；
- 2、差异剪切、位点变异基因和差异代谢物分析；差异代谢物分析；
- 3、关键驱动基因与转录因子分析；
- 4、筛选核心调控网络，解析调控机理；

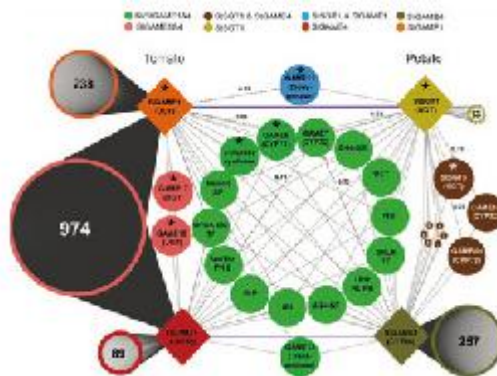
案例分享 转录组和代谢组解析茄科植物中生物碱合成途径

Itkin M, Heinig U, Tzfadia O, et al. Biosynthesis of Antinutritional Alkaloids in Solanaceous Crops Is Mediated by Clustered Genes. *Science* 2013.341:175-179. (IF=30.75)

茄科植物中的类固醇糖生物碱（SGAs）能引起胃肠道和神经系统疾病，而生物碱合成途径尚未完整解析。本研究通过转录组和代谢组全面解析土豆和番茄生物碱合成途径，鉴定了 10 个候选基因，修正了生物碱生物合成途径，为改良茄科植物品质及医用价值提供依据。



茄科植物中生物碱和皂苷生物合成途径



生物碱基因的共表达网络分析

大样本（全）转录组

转录调控高分文章发表现状

随着高通量测序在功能基因组学领域的广泛使用,转录组测序已经是常规的技术手段之一。通过调查近五年高通量测序转录调控相关文献发现,文献数目呈指数增长,但是绝大多数发表的文章水平普遍较低(影响因子 ≤ 5)。进一步分析 2016-2017 年转录调控高分文章(影响因子 ≥ 5),依据文章主要结构和研究手段可分为四类:A、小样本转录组数据分析+深入基因功能研究(分子实验);B、转录组数据分析+其他组学数据(代谢组或蛋白组);C、大样本转录组数据深入分析+基础功能验证;D、大样本转录组+非编码 RNA 数据深入挖掘+基础功能验证。



转录调控高分文章解决方案一大样本（全）转录组

根据转录调控高分文章分类分析可见，C类和D类是我们比较好的选择。大样本（全）转录组是实现这两类文章的最佳解决方案。

- 1、转录组研究是功能基因的研究利器，是后续实验设计、基金申请的基础数据，可长期使用；
- 2、转录组研究最大的特点“短平快”，材料获取和实验实施较容易进行，测序分析到文章发表周期快；
- 3、项目性价比高，随着样本数目的增加，结合百迈客的实验设计及配套的项目分析方案，不需要太多分子实验；
- 4、全转录组在转录组基础上加入了非编码RNA的内容，非编码RNA本身是一定的热点研究，有助于提高文章发表档次。



大样本（全）转录组

解决方案--实验设计、分析方案

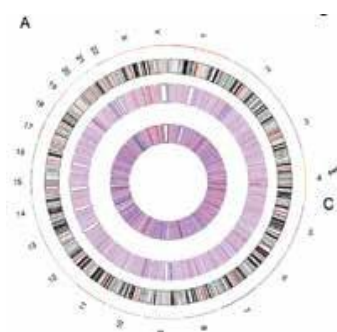
依据物种和研究方向、前期试验结果，结合大量文献调研，提供配套针对性的实验设计。

依据实验目的和实验设计，针对大样本的数据量，制定个性化项目分析内容和数据深入挖掘思路，按照写作文章思路分析数据。

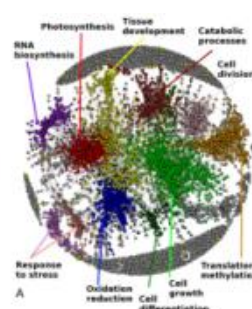
解决方案--发表文章的分析结果

高分文章一个很大的特点是对研究结果呈现的美观性，因此大样本（全）转录组项目可提交优化的直接用于文章

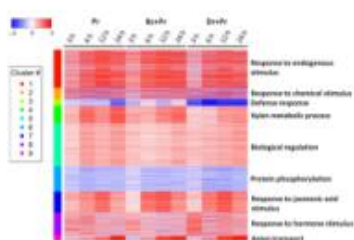
为了提高文章中数据结果呈现的多样化和个性化，需要自主细致地调整分析结果的图表。因此，通过云平台上数据分析、文章图表绘制等专门的培训班，更好地优化分析结果。



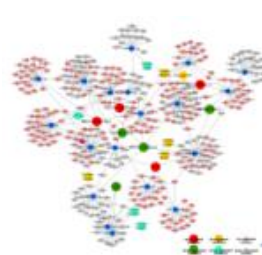
lncRNA 染色体分布 Circos 图



共表达基因模块调控网络



层次聚类分析



mRNA 与非编码 RNA 调控网络

解决方案--云平台助力数据挖掘

依托强大的生物信息分析平台—百迈客云平台，即时自主地进行数据分析，绘制美图。

依托百迈客云平台，整合已发表海量公共数据，结合自测数据，有效降低研究成本，提升数据分析深度。

APP 个性化分析内容

数据整理：检索、筛选、整合、调整参数等

美图绘制：热图、韦恩图、分类和富集图等

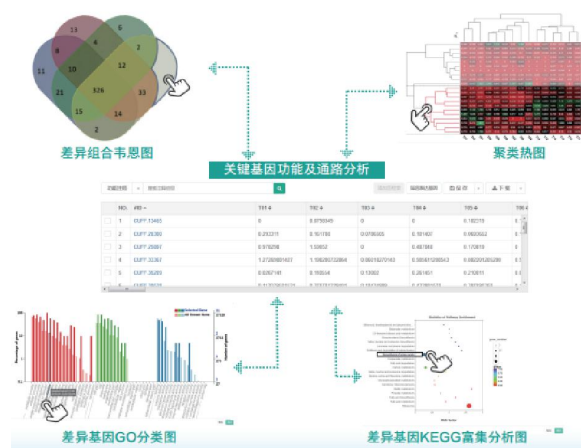
联合分析：mRNA 与非编码 RNA 联合分析

APP 个性化分析优势

随心所欲、即时分析

可视化、交互性操作

快速高效、分分钟呈现结果



依据大样本全转录组的数据信息，可搭建该物种研究方向的转录调控数据库，用于后续功能基因、转录调控机制等研究。

大样本转录组解决方案—实验设计和分析方案

实验设计

依据物种和研究方向、前期预实验结果，结合大量文献调研，提供配套针对性的实验设计。

案例：转录组解析连续生物和非生物胁迫下拟南芥的反应机制 Coolen S. Plant Jour, 2016, 86: 249-267.

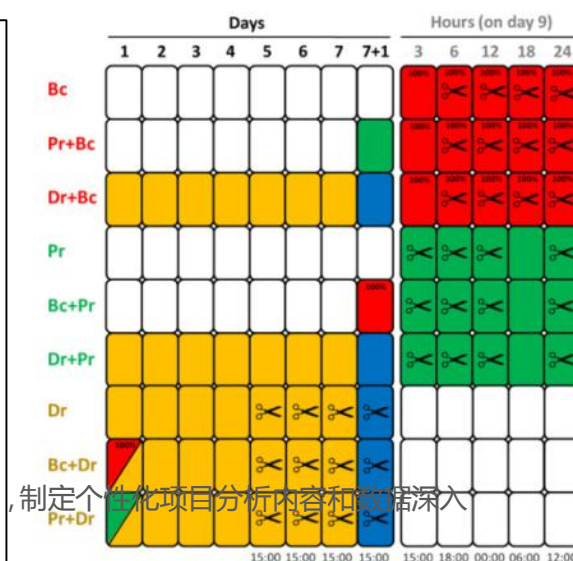
实验材料：4 周拟南芥小苗；

实验处理：

- 1、灰霉菌（*B. cinerea*）处理：用菌液侵染第八片叶片，在 6、12、18 和 24 小时取样；
- 2、菜青虫（*P. rapae*）处理：用幼虫处理第八片叶片，在 3、6、12 和 24 小时取样；
- 3、干旱（Dr）处理：控制供水后 5、6、7 天和恢复供水后第一天（7+1）取样；
- 4、连续两个处理：
 - Dr+Bc or Pr，在干旱处理 7+1 天时进行 Bc 或 Pr 处理；
 - Bc+Dr or Pr，湿度从 100% 调为 70%，进行 Dr 或 Pr 处理；
 - Pr+Dr or Bc，将幼虫从叶片上去除后，进行 Dr 或 Bc 处理；

取 样：每组处理均设置对照组，设置 4 个时间点，取叶片提取 RNA 用于测序，设置 3 个生物学重复，共 167 个样品。

测序策略：RNA-seq，Illumina Hi-seq 2000。



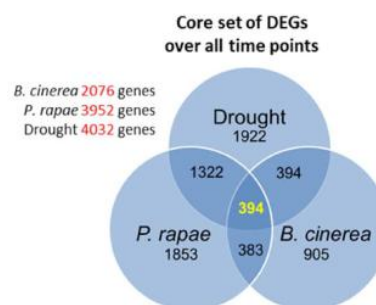
制定个性化项目分析内容和数据深入

文章思路 案例：转录组解析连续生物和非生物胁迫下拟南芥的反应机制

- 1、单个胁迫和连续两个胁迫处理的差异基因数目分析；
- 2、三组单个胁迫处理的核心差异基因集鉴定和功能分析；
- 3、比较 Bc 处理组和 Dr or Pr+Bc 处理组鉴定 Bc 核心基因集，表达量聚类、共表达和 GO 功能富集分析；
- 4、比较 Pr 处理组和 Dr or Bc+Pr 处理组鉴定 Pr 核心基因集，表达量聚类、共表达和 GO 功能富集分析；
- 5、比较 Dr 处理组和 Pr or Bc+Dr 处理组鉴定 Dr 核心基因集，表达量聚类、共表达和 GO 功能富集分析；
- 6、连续胁迫处理时，第一个胁迫对第二个胁迫响应的影响。

要点小结

- 1、从整体分析各处理组差异基因集数目和上下调情况；
- 2、按大组来进行差异基因比较分析，用韦恩图直观呈现；
- 3、按基因时间序列表达模式聚类分析，结合 GO 富集进行功能分析；
- 4、基因共表达网络分析有助于从复杂数据中提取关键信息。



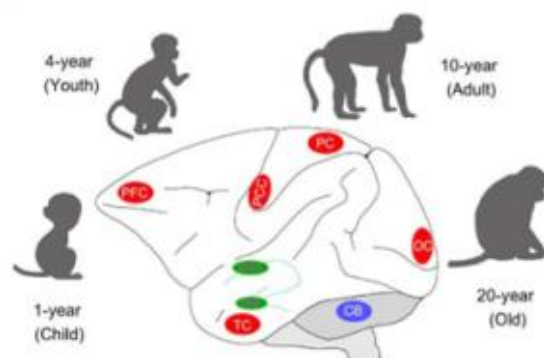
大样本全转录组 I 案例分享-LncRNA+mRNA

猕猴大脑发育和老化过程中长链非编码 RNAs 介导的表观遗传调控基础

Liu S.Wang Z.Chen D et al.Annotation and cluster analysis of spatiotemporal and sex-related lncRNA expression in rhesus macaque brain.Genome research 2017. doi: 10.1101/gr.217463.116.

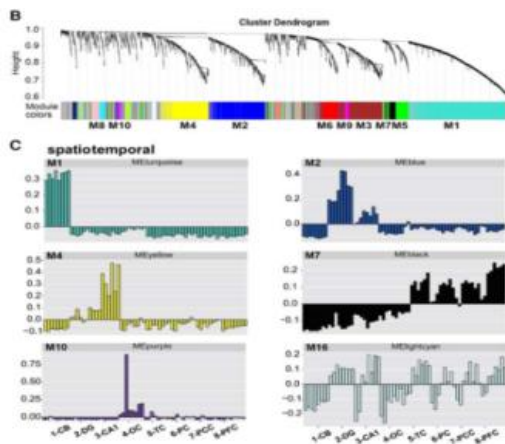
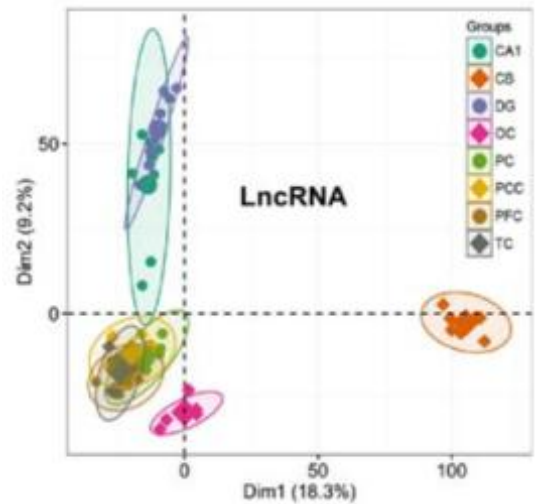
实验设计

- 1、物种：恒河猕猴；
- 2、取材：两只公猴和两只母猴，选取四个发育阶段（儿童-1岁青年-4岁、成年-10岁和老年-20岁），脑部的八个不同区域，分别是前额叶皮质（PFC）、后扣带皮质（PCC）、颞皮层（TC）、顶叶皮质（PC）、枕叶皮质（OC）、海马（CA1）、齿状回（DG）和小脑皮质（CB）。公猴和母猴的组织分别混样，共计 64 个样品。
- 3、测序策略：RNA-seq, Illumina HiSeq 2000；

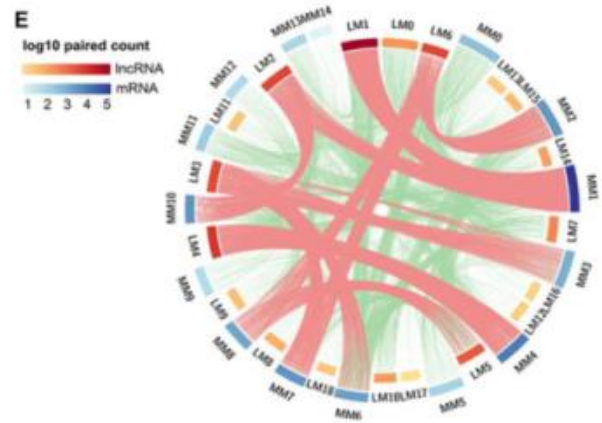


文章思路

- 1、比较猕猴脑中 lncRNA 序列与大猩猩、人、小鼠的相似性；
- 2、比较不同性别、时期及组织的 lncRNA 与 mRNA 的表达模式；
- 3、WGCNA 分析鉴定时空、性别特异性表达的 lncRNA；
- 4、CAGE-seq 分析 lncRNA 转录起始位点和 5'加帽动态变化；
- 5、lncRNA 与 mRNA 共表达分析，鉴定模块和共表达网络；
- 6、分析 mRNA-lncRNA 和 lncRNA-lncRNA 之间的调控机制。



WGCNA 分析 lncRNA 表达模式



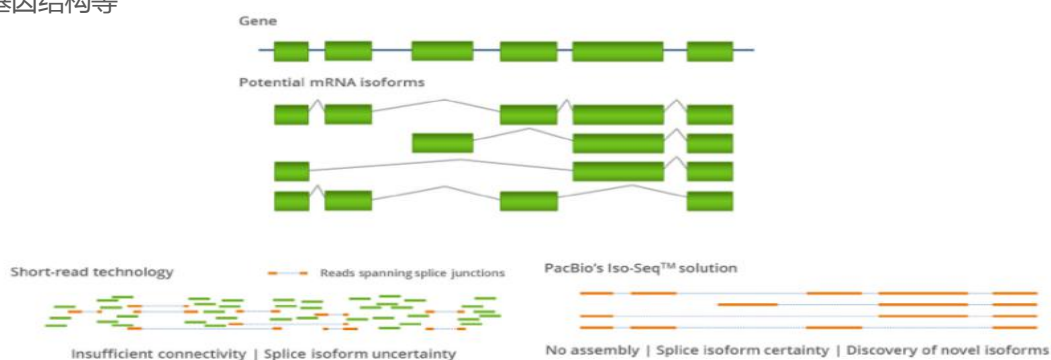
lncRNA 和 mRNA 共表达基因模块调控图

三代全长转录组概述

全长转录组测序是利用 PacBio 三代测序平台进行的测序，目前，PacBio 平台的平均读长已经达到 12kb，其长度大大超过一般转录组中典型基因的长度（人转录本平均长度 2.5kb，鸡为 1.2kb）。利用 PacBio 三代测序平台其

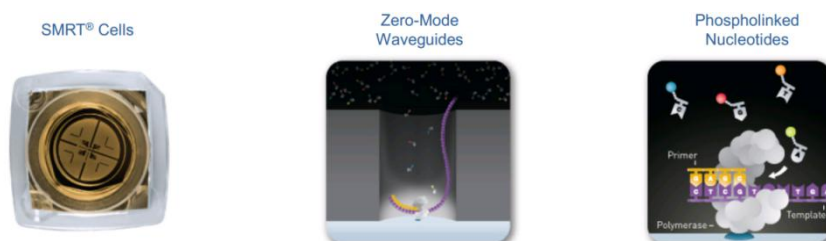
优势在于：

- 无需打断，可直接读取从 5'端到 3'端 polyA 尾的高质量的单个 RNA 分子的全长序列（包含 5' 3' UTR 区）
- 消除转录组组装带来的错误，无参定量更为准确
- 可准确鉴定可变剪接、APA（可选择性多聚腺苷酸化）和转录起始位点、融合基因和基因家族等信息
- 对于多倍体或杂合物种，由于亚基因组序列相似性较高，全长转录组可准确鉴定多倍体物种基因结构等



三代测序原理

Pacbio 测序核心技术原理是单分子实时测序技术 (SMRT)，SMRT 测序采用四色荧光标记的 dNTP 和 ZMW 孔对单个 DNA 分子进行测序。ZMW 孔直径为 50-100nm，每个 SMRT Cell 中含有大量这种圆形纳米小孔。单链环状的 DNA 分子模板与引物结合，然后结合固定在 ZMW 孔底部的某种特殊的 DNA 聚合酶后，开始边合成边测序。



全长转录组产品优势

- **短平快**：材料的获取和实验的实施较容易进行，基本不需要进行分子实验即可发表文章，因此测序分析到文章发表周期短；
- **性价比高**：所测即所得，无需做 Race 实验，节约大量经费、时间；无参物种获得物种的全长转录本序列，可作为后续转录组项目的参考序列；
- **基因结构研究利器**：可准确鉴定可变剪接和转录起始位点、融合基因和基因家族等信息
- **研究热点**：随着测序平台和分析技术的成熟，越来越多的研究采用三代全长转录组阐释科学问题，三代全长转录组已然成为热点研究对象。研究物种从模式动物向经济作物、药用植物、水产转移，研究物种更加丰富。



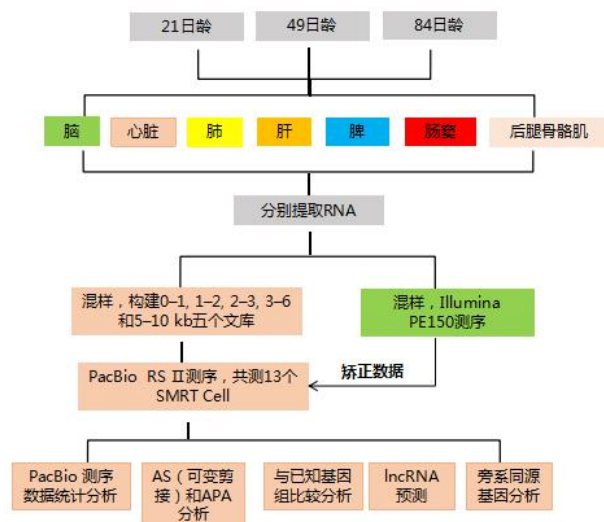
成功案例：全长转录组测序揭示兔转录本图谱

Chen S Y, Deng F, Jia X, et al. A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing[J]. Scientific Reports, 2017, 7.

案例介绍

兔 (*Oryctolagus cuniculus*) 在脂蛋白代谢方面与人类相似，因此被认为是研究人类高胆固醇的首选动物模型。

从 14,474 个基因座中获得了 36,186 个高可信度转录本，其中超过 23% 的基因和 66% 的转录本在目前的参考基因组中尚未被注释。此外，约 17% 的转录本显示为非编码 RNA。在此重新构建的转录本中检测到多达 24,797 个可变剪接 (AS) 和 11,184 个可选择性多聚腺苷酸化 (APA) 事件。



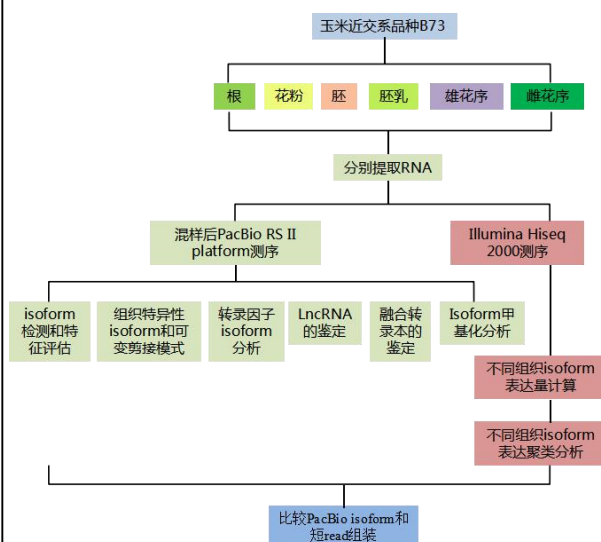
案例解析：全长转录组解析玉米转录组复杂性

Wang B, Tseng E, Regulski M, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing[J]. Nature Communications, 2016, 7.

案例介绍

玉米是世界上最重要的农作物之一，也是阐明植物转录组调控网络的模式植物。

利用 Pacbio 数据改善玉米基因组注释，纠正了之前的错误注释；共发现 110322 个 splice isoform，其中有至少 60% 的可变剪切形式是新发现的；转录组在不同组织中有较大的差异，无论是转录本表达量，还是可变剪切形式；鉴定出 1430 个融合转录本，玉米中的基因融合现象比之前的认知要普遍的多；鉴定出 878 个 lncRNA，lncRNA 的比例也比之前发现的高。



微生物多样性测序

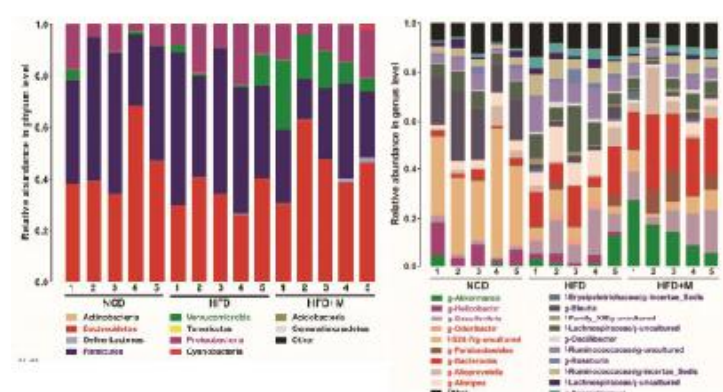
微生物多样性测序通过对环境样品中细菌 16S rDNA 和真菌 ITS 高变区进行测序,能同时对环境中优势物种、低丰度物种和难以分离培养的物种进行检测,解析环境中微生物在不同水平下的组成和丰度,比较不同环境中微生物的组成和丰度的差异。

微生物多样性测序分析内容

物种分类分析	单样品多样性分析	多样品多样性分析	高级分析内容
物种注释统计 物种分布柱状图 物种丰度聚类热图 物种系统进化树	稀释曲线 香农指数曲线 等级丰度曲线 物种累计曲线	样品 Heatmap PCA/PCoA 分析 NMDS 分析 样品层次聚类	RDA/CCA 分析 多元统计分析 Lefse 分析 16S 功能预测分析 相关性分析

经典案例：褪黑素通过调节小鼠肠道菌群预防肥胖

本研究通过分析不同膳食喂养 (NCD: 正常膳食、HFD: 高脂肪膳食、HFD+M: 高脂肪膳食+褪黑素) 的小鼠的肠道菌群,发现褪黑素能显著地控制小鼠体重增加,褪黑素能显著降低小鼠肠道微生物的多样性和厚壁菌门与拟杆菌门的比例,并且会显著提高疣杆菌门的相对丰度;通过 Lefse 分析,共找到 30 个与褪黑素相关的标志 OTU,这些 OTU 属于 Ruminococcaceae、Desulfovibrionaceae、Bacteroides、Porphyromonadaceae、Helicobacteraceae、Christensenellaceae,有研究表明这些微生物与肥胖相关。



不同膳食喂养小鼠肠道门、属水平微生物群落结构



差异 OTU 丰度热图

微云 v3.0--微生物多样性数据分析平台

微云是基于百迈客云 (BMKCloud) 开发的一种用于微生物多样性分析的 APP, 以专业、高效、丰富、简单的特点实现微生物多样性分析的私人定制, 通过全面丰富多角度的个性化分析, 深度解读微生物数据。

微云特征

内容专业

专业的微生物研发团队,
专业的分析流程, 专业的
个性化分析, 专业的微生物
多样性数据分析平台

分析高效

用户可直接和数据对接,
省略中间反复沟通环节,
高效分析可一天拿到分
析结果。

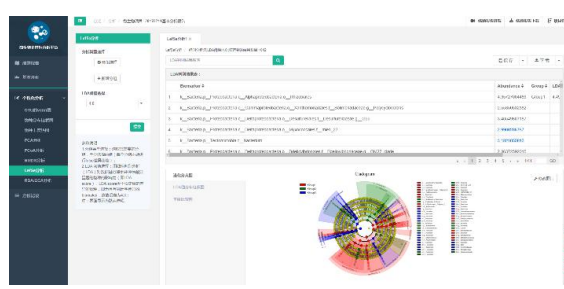
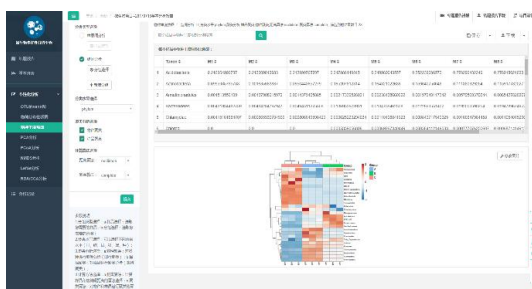
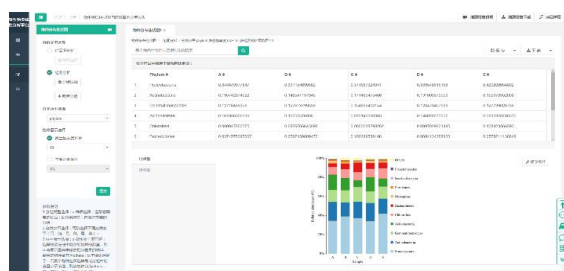
结果丰富

30+基本分析内容,
8+款个性化分析,
多角度多参数解析
微生物多样性数据。

操作简单

简单易学的操作流程,
可快速掌握微云主流
程以及个性化分析,
全面进行数据解析。

微云界面展示



宏基因组测序

宏基因组对环境微生物总 DNA 进行测序研究，关注环境微生物群体种群结构、进化分析、功能基因，解析微生物功能与环境之间的关系，挖掘和研究特定功能基因及代谢通路。

宏基因组测序的分析内容

物种分类分析

物种组成分析
物种丰度聚类
多样性比较分析
物种 PCA 分析

基因组注释

基因预测
基因丰度分析
基因归属分析
前噬菌体预测
转座元件预测

基因功能注释

通用数据注释
专有数据库注释
功能丰度分析
功能聚类分析

差异分析

差异基因筛选
差异基因富集分析
STAMP 差异分析
Anosim 相似性分析

应用方向

基因功能注释

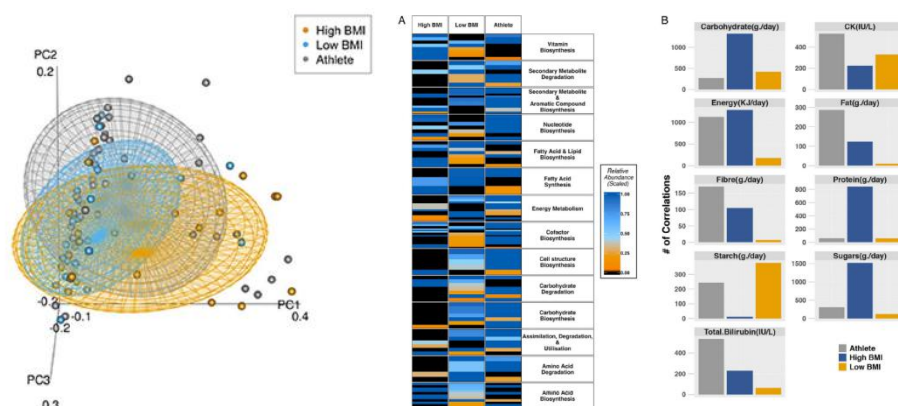
功能基因挖掘

人类疾病发生

养分循环利用

经典案例：运动员与正常人肠道宏基因组研究

本研究选取 40 个专业运动员和 46 个健康人（根据 BMI 指数分为高低两组）的粪便样品，通过宏基因组测序对运动员和健康人群肠道微生物功能进行研究，结果表明与正常人相比，运动员的肠道微生物具有更高的功能多样性，并且与肌肉运动和健康相关的代谢通路的丰度获得了显著地提升。



健康人群与运动员肠道微生物功能 PCoA 分析

健康人群与运动员肠道微生物功能变化和代谢差异

The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level. Gut. 2017 Mar 30.

宏转录组测序

宏转录组测序以环境样品总 RNA 入手，原位研究特定时空的活跃菌群组成与丰度、活性基因的差异与表达，结合理化生境的不同，解析差异条件下的基因表达与调控。

宏转录组测序的服务内容

物种分类学分析

物种组成分析
物种丰度聚类
物种进化分析

功能注释分析

通用数据库注释
CAZyme 数据库注释
ARDB 数据库注释

差异基因表达与富集

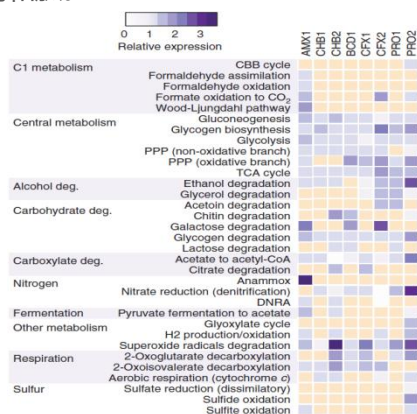
差异基因功能注释
差异基因 GO、KEGG 富集
调控网络分析

应用方向

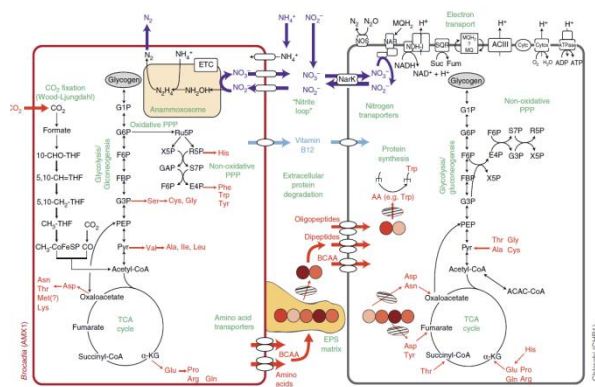
活跃菌群的组成 活性基因差异表达 养分循环利用 代谢途径与分子调控

经典案例：宏基因组和宏转录组研究揭示厌氧氨氧化颗粒中的微生物群落相互作用

本研究选取厌氧稳定运行的氨氧化反应器中高脱氮效率的生物样本，通过宏基因组和宏转录组测序研究异养和厌氧细菌之间的相互作用。结果表明，环境中厌氧菌 *Brocadia* 和异养菌 *Chlorobi* 是优势活跃物种，并通过功能注释发现 *Chlorobi* 是高度活跃的蛋白质降解剂，分解代谢细胞外肽的同时将硝酸盐还原成亚硝酸盐，并可能有助于清除由厌氧细菌产生的碎屑和肽。



MAG 相关的主要代谢途径的基因表达量热图



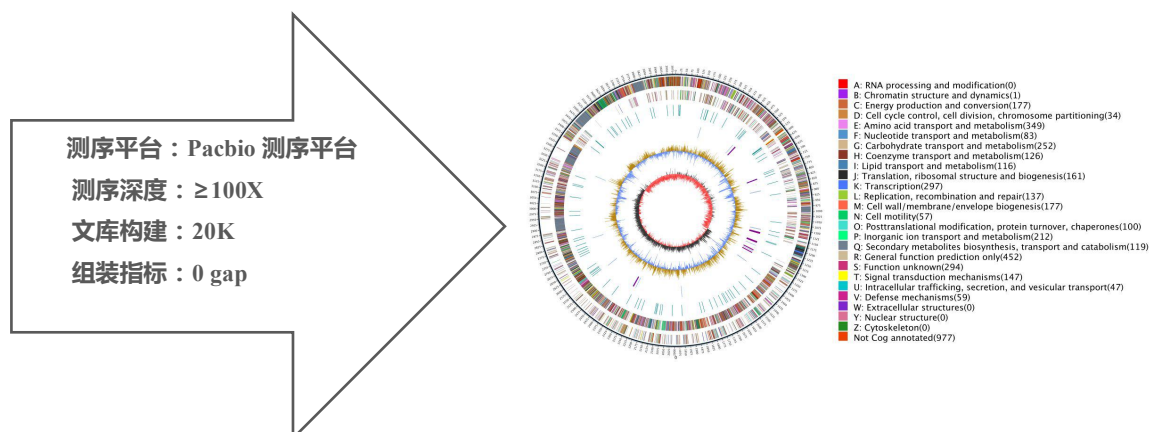
Brocadia 和 *Chlorobi* 之间的代谢互作图

Metabolic network analysis reveals microbial community interactions in anammox granules. Nat Commun.2017.

细菌基因组完成图

细菌基因组完成图基于 Illumina 二代测序以及 PacBio 三代测序平台，实现细菌基因组的完整组装，获得全面丰富的基因组信息，从而可在致病性、耐药性、环境适应性、代谢途径等多角度进行功能基因的深入挖掘，对细菌不同表型等特性进行深层次研究。

细菌基因组完成图研究策略



细菌基因组完成图研究内容

基因组组分研究

编码基因预测
重复序列注释
非编码 RNA 预测
基因岛预测
前噬菌体预测
CRISPR 预测

基因功能研究

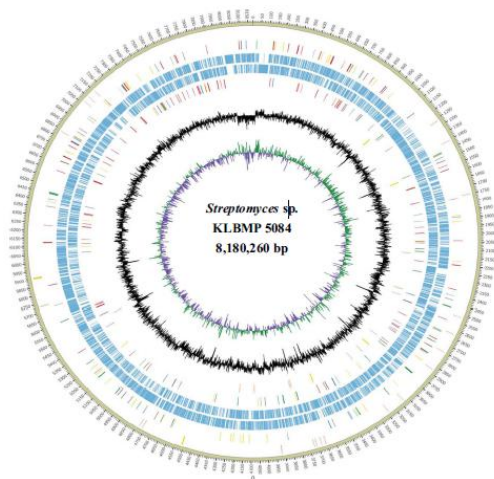
通用数据库注释
专用数据库注释
分泌蛋白分析
转运蛋白分析
信号肽预测
基因组表观修饰

比较基因组研究

基因家族分类
特有和共有基因检测
快速进化基因检测
物种间进化关系分析
共线性分析
水平基因组转移分析

细菌基因组完成图案例解析

该文章通过基因组测序内生链霉菌 KLBMP 5084 获得完成图，研究在高盐分失调环境下对其宿主补血草生长和生理的响应，并对菌株的基因组进行测序分析以了解其环境适应和增强代谢潜力的机制。该研究有助于了解内生菌如何帮助生长在沿海盐沼的植物调节耐盐度，并提供一个可能的方法提高濒危盐土植物和重要农作物的生长和耐盐性。完整基因组信息用于其致病机制以及进化关系的进化关系的研究。



链霉菌 KLBMP 5084 圈图

Features	Chromosome
No. of all scaffolds	1
Genome size (bp)	8,180,260
G + C content	72.41 %
Plasmid	0
miRNAs	3
rRNAs	18
tRNAs	66
Protein-coding genes (CDS)	6433
Secondary metabolites clusters	36

组装以及预测结果

Plant growth-promoting effect and genomic analysis of the beneficial endophyte *Streptomyces* sp. KLBMP 5084 isolated from halophyte *Limonium sinense*. Plant Soil.2017

真菌基因组测序

真菌基因组测序基于第二代、三代测序以及光学图谱技术相结合，组装高质量真菌基因组图谱，获得全面丰富基因信息，解析食药真菌重要代谢产物、病原性真菌致病性差异机理以及对重要的功能蛋白、遗传进化等全面分析，可根据研究不同需求选择真菌基因组精细图或真菌基因组准完成图。

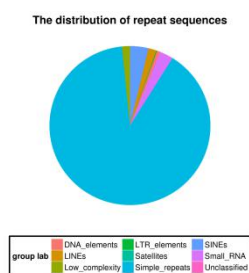
真菌基因组精细图

测序平台：Hiseq+Pacbio 测序
测序总深度： $\geq 130X$
组装指标：Contig N50 $\geq 1M$

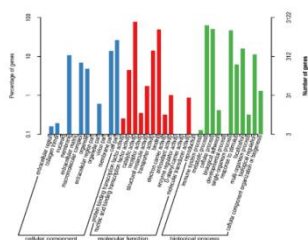
真菌基因组准完成图

测序平台：Hiseq+Pacbio+Irys 测序
测序总深度： $\geq 300X$
组装指标：接近染色体水平

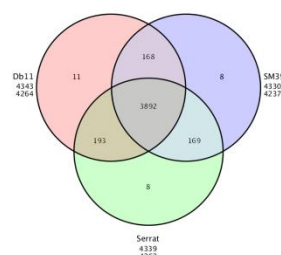
真菌基因组测序结果展示



重复序列分析



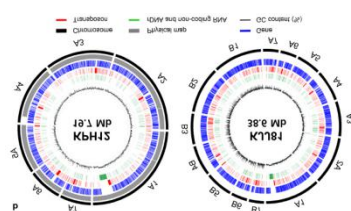
GO 功能注释分析



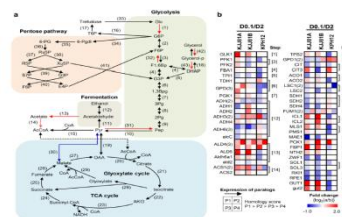
比较基因组分析

真菌基因组测序案例解析

本研究是对分离于韩国小麦的两株真菌 *S.fibuligera* KJJ81(38 Mb)和 KPH12(18 Mb)进行基因组测序，通过单分子和二代测序以及光学图谱，得到更高质量的组装水平（7 条和 14 条线性 scaffold），通过基因组注释、比较基因组分析以及转录组数据，对其碳代谢、硫代谢以及纤维素降解等进行重点研究，为其起源进化以及菌种改良起到关键的指导作用。



S.fibuligera KJJ81 和 KPH12 拼接结果



碳代谢注释研究结果图

Que Yet al. Whole-genome de novo sequencing, combined with RNA-Seq analysis, reveals unique genome and physiological features of the amylolytic yeast *Saccharomycopsis fibuligera* and its interspecies hybrid
Biotechnology for Biofuels 2016

医学科技

外泌体 RNA 测序

外泌体 (Exosome) 是一种能被大多数细胞分泌的微小膜泡, 广泛分布于外周血、尿液、唾液、腹水、羊水等体液中。其来源多样, 可来自各种类型细胞, 其内容物含蛋白、脂质和核酸等, 特别是在其中发现的 non-coding RNA (miRNA、lncRNA、circRNA 等) 使其更具有意义。外周血等体液中的分子标志物的检测具有无创伤、可重复特点, 同时可于同一时间点进行多项指标的检测。通过高通量技术对外泌体 RNA 进行测序, 可快速高效获得外泌体 RNA 的全面信息, 是疾病诊断和随访的理想手段。

技术优势

强大的科研实力: 国内首家推出低起始量建库策略 (低至 2ng), 在确保结果准确的同时可以解决体液样本存在的样品量极低问题, 是在外泌体水平研究基因表达调控强有力的方法, 极大地拓展了 RNA-Seq 的应用范围;

丰富的项目经验: 目前已拥有上百个项目经验, 涉及体液样本类型丰富, 包括血浆、血清、细胞上清、尿液、乳汁、灌洗液、精液、唾液等;

专业的科研团队: 国内唯一一家拥有独立的外泌体 RNA 测序研究团队; 根据研究需求, 量身定制, 免费提供最优的项目方案。

技术参数

类型	测序量	建库策略	获得RNA信息	生物学重复设置
miRNA	10M	miRNA 文库	miRNA	每个样本至少3个重复
lncRNA	16G	lncRNA 文库	lncRNA、mRNA、circRNA	
circRNA	≥10G	去线性文库	circRNA	

1. circRNA 也可采用 lncRNA 建库策略方式, 适用于互作关系研究;
2. 机理方面的研究: 建议根据不同患病时期选取不同阶段的梯度样本。

外泌体 RNA 研究思路

研究目的决定取样思路, 根据不同的研究方向, 外泌体 RNA 取样思路如下:

肿瘤机理研究 1. “两阶段”研究方案: 某肿瘤发生前后组织来源细胞上清液 2. “多阶段”研究方案: 某肿瘤发生、发展中不同阶段的组织来源细胞上清液 3. 肿瘤微环境方案: 不同组件来源上清液: 基质、上皮细胞、内皮细胞、脂肪组织、纤维组织等 4. 外泌体及其来源联合研究方案: 同时收集外泌体及其相应的肿瘤组织或细胞	分子标志物开发 1. 组织鉴定方案: 根据研究肿瘤的组织特异性, 针对性选择样本, 如肾-尿液、脑相关-脑脊液等 2. 疗效跟踪方案: 治疗 (肿瘤切除或药物治疗后) 预后的跟踪诊断, 选择预后不同时期鉴定体液样本 3. 不同患病时期体液样本进行分子标志物开发
--	---

案例解析

外泌体来源的 miR-1290 和 375 可作为雄激素非依赖性前列腺癌预后标志物
Exosomal miR-1290 and miR-375 as Prognostic Markers in Castration-resistant Prostate Cancer. European urology, 2015, IF=14.976'

研究背景：

细胞外的 miRNA 可以被包被在外泌体中，可以作为前列腺癌预后的分子标志物。本研究通过鉴定和评估血浆来源的

外泌体 miRNA，筛选可作为雄激素非依赖性前列腺癌预后标志物。

样本取材：

23 个 CRPC 患者血浆：用于测序

100 个 CRPC 患者血浆：用于 qPCR 验证

研究思路：

1. Discovery set 23 个样本通过高通量的方法进行初筛，鉴定真实表达的 miRNA 432 个。
2. 初筛 432 个分子经参数调整进一步，缩小到 267 个。
3. 二筛通过 COX 回归模型及 KM 生存率分析筛选出 3 个。
4. 扩大样本验证 100 个样本经 qPCR 及 KM 生存分析缩至 2 个。

研究结果：

miR-1290 和 miR-375 可以作为 CRPC 晚期的预后良好分子标志物。

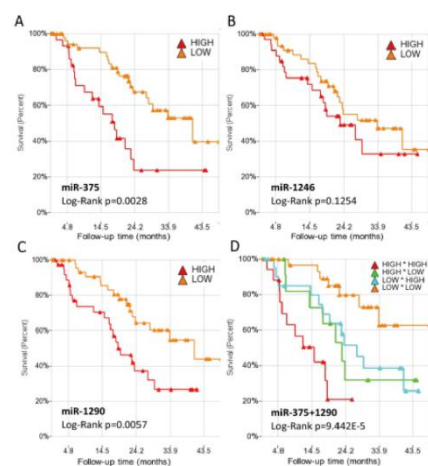


图 1. Kaplan-Meier 生存分析曲线

1. 本研究首先基于高通量测序技术筛选到 375 个已知 miRNA 和 57 个新的 miRNA，通过控制 log2 transformed read counts <5，最终筛选获得 248 个已知 miRNA 和 19 个新的 miRNA。
2. 通过 COX 回归模型分析，进一步筛选找到 3 个 FDR ≤ 0.05 与生存相关的 miRNA：miR-1290、miR-1246、miR-375。通过 KM 生存率分析，进一步确认 miR-1290、miR-1246 和 miR-375 与生存显著性相关 (p < 0.05)。
3. 扩大样本进行验证
在已有数据库中选取健康人 50 个、患病人 119 个，最终确定选取 miR-30a-5p、miR-30e-5p 作为标准化的内参，经过 qPCR 验证进一步进行 Kaplan-Meier 生存分析 miR-1290、miR-375，p 值分别为 0.0028 和 0.0057 有意义。miR-1246，p = 0.13 无意义

最终确定 miR-1290 和 miR-375 可以作为 CRPC 晚期的预后良好相关分子标志物。

全外显子组测序

全外显子组测序 (Whole Exome Sequencing, WES), 又称定向外显子捕获测序 (Targeted Exome Capture Sequencing), 指利用序列捕获技术选择性地富集占全基因组约 1-2% 的外显子区域 DNA 进行高通量测序, 从而直接发现与蛋白功能变异相关的变异, 经济高效且更易于分析, 从而广泛应用于单基因遗传病、复杂疾病和肿瘤变异研究。

遗传病致病基因研究

1. 单基因遗传病: 受一对等位基因控制的遗传病, 又称孟德尔遗传病, 其传递规律遵循孟德尔分离定律, 根据遗传方式分为常染色体显性、常染色体隐性、x 伴性显性、x 伴性隐性和 Y 伴性遗传病等。
2. 多基因遗传病: 遗传信息通过两对以上致病基因的累积效应所致的遗传病, 是遗传因素与环境因素共同起作用。一般有家族性倾向, 与患者血缘关系越近, 患病率越高。

项目流程

- 样本选取** → 按照家系遗传规律, 选取一个/几个相同疾病家系的核心成员 (患者 2-3 个/正常对照 1-2 个) 或散发样本等
- 建库测序** → 提取 DNA (通常为全血样本) 质控, 构建外显子捕获文库, 文库质控合格后上机测序, 数据量建议 10G/样本 (有效测序深度约 100X)
- 分析内容** → 基本分析: 数据质控>比对统计>变异检测>变异注释 (变异位点位置和功能区域信息、数据库频率、保守/有害性预测、基因型信息、疾病数据库以及变异基因功能注释)
高级分析: 有害变异过滤>结合遗传模式筛选 (显性遗传、隐性遗传、新生突变等)>个性化分析
- 致病证据** → 变异验证: 在本研究样本及更多病例中利用 Sanger 测序等方法验证候选致病基因/位点的致病比例
功能验证: 结合功能注释在细胞、动物模型、临床样本等对候选基因进行功能实验, 深入挖掘致病机制

肿瘤体细胞变异研究

体细胞变异 (somatic mutations), 发生在正常机体细胞中的突变, 比如发生在皮肤或器官中的突变, 这样的突变不会传给后代, 却可以引起当代某些细胞的遗传结构发生改变。多数肿瘤是个体对肿瘤的遗传易感性基础上, 致癌因子引起细胞遗传物质结构或功能异常 (体细胞突变), 导致其克隆性异常增生而形成的异常病变。

驱动突变 (driver mutation): 对肿瘤的增殖扩散有重要影响, 使肿瘤获得选择性的增长

优势，这样的突变称为 driver mutation；另外一种对肿瘤增殖扩散影响很小或者没有影响的突变称 passenger mutation。

体细胞变异研究策略

研究目的	研究策略	样本量&数据量
体细胞突变	实体肿瘤（肿瘤组织 VS 血液/癌旁/正常组织） 血液肿瘤（骨髓细胞/血液 VS 口腔黏膜细胞/皮肤）	推荐 30-50 对配对样本 肿瘤样本测序深度 150X 以上，即>15G 数据量/样本 配对的非肿瘤组织样本>10G 数据量/样本
驱动基因	大量样本寻找高频突变基因 分析已知驱动基因，预测驱动基因	
转移复发	不同转移灶 vs. 原发灶+配对正常对照 复发灶 vs. 原发灶+配对正常对照	
演进机制	时间：不同时期/分期样本 空间：同一时期不同部位/肿瘤组织不同分区	
药物/治疗应答	肿瘤药物治疗/化疗/放疗敏感性研究 敏感患者 vs. 耐受患者肿瘤样本+配对正常对照	

复杂疾病遗传易感性研究

复杂疾病是在众多因素共同作用下发生的，如多个基因、一个基因的多个突变、环境作用以及未知的随机因素，遗传模式复杂。复杂疾病在普通人群中发病率较高（一般不少于 1%），也称“常见疾病”。遗传基础是由多基因构成的，它部分决定了个体发病的风险，这种由遗传基础决定一个个体患病风险称为易感性。

复杂疾病遗传易感性研究策略

1. GWAS 研究方案

两阶段或多阶段设计 two/multiple-stage design	- 第一阶段小样本量(50-100 对以上)进行病例对照研究(NGS)，推荐低深度测序，外显子测序建议 20-50X（即>2-5G 数据量），统计分析筛选出较少数量的阳性 SNP--Discovery Stage - 第二阶段或随后的多阶段中采用更大样本的病例对照样本群进行基因分型--Replication Stage - 采用两阶段或多阶段进行关联研究，是一种经济高效的方法
两阶段设计 two stage design	- 患者共有有害突变基因筛选:筛查 2 个病例以上共有的罕见有害变异或变异所在基因，基因功能注释可能与疾病发生相关 - 若有对照，则剔除正常对照中的（或正常对照中频率>90%）变异 - 第一阶段为 NGS 阶段（候选基因发现阶段），病例组约为 10-50 例，推荐数据量>10G；可进行第二阶段扩大样本量验证

Production and Service

产品与服务 » 医学检测|解读基因，守护健康

百迈客医学检测专注于基因临床检测的服务和产品，主要方向为肿瘤和遗传病相关的个性化用药基因检测和早期发现与预防。包括基于 ctDNA 的无创肿瘤检测和实体瘤基因检测，单基因遗传病检测、疾病易感基因检测等产品。基因检测是精准医疗的核心组成部分，是实现个体化治疗的重要技术手段，对于提高治疗效果，降低病患痛苦有着重要意义。

百迈客医学检测将秉承严谨、专业的精神去践行我们的使命：解读基因，守护健康。

检测项目



肿瘤诊疗

肿瘤个体化靶向用药
肿瘤个体化化疗用药

爱检康
肿瘤个体化用药基因检测
利用先进的高通量测序技术，对患者实体瘤或静脉血样本进行基因检测，通过专业的生物信息学分析，结合 NCCN 指南及肿瘤研究进展前沿，分析靶向药物的疗效和副作用，为医生提供可信的治疗依据。



生殖生育

无创产前 DNA 基因检测

爱汝康
遗传性乳腺癌风险基因检测
利用高通量测序技术，检测 BRCA1、BRCA2、等 18 个遗传性乳腺癌易感基因，进行风险评估，提供健康管理建议。



健康管理

疾病易感基因检测
叶酸代谢能力基因检测
酒精代谢能力基因检测
用药安全基因检测

检益康
144 种疾病个体易感性检测
针对中国人群常见的 12 大类 144 种疾病进行风险预测和评估，并提供专业的遗传检测报告和健康指南



产品介绍

百贝康
无创产前 DNA 基因检测
利用第二代高通量测序技术对母体外周血中的游离 DNA 片段（包括胎儿游离 DNA）进行测序，并将测序结果进行生物信息分析，可以从中得到胎儿的遗传信息，从而评估胎儿是否患唐氏综合征（T21）、爱德华氏综合征（T18）、帕陶氏综合征（T13）疾病的风险。

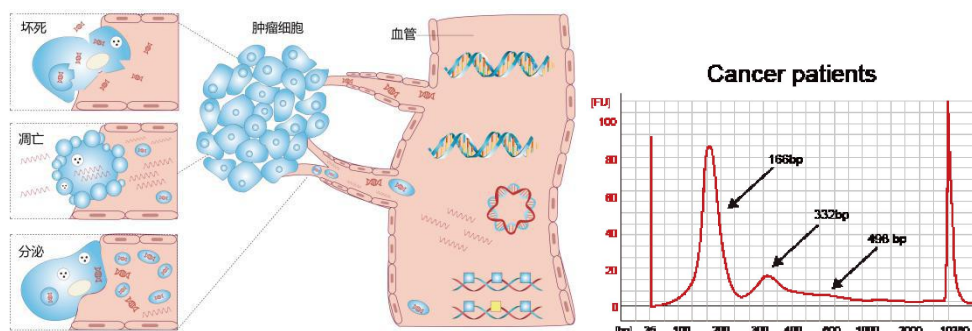
酒因真经
酒精代谢能力基因检测
结合基因检测与身体指标，计算健康饮酒量及饮酒相关疾病风险，提供个性化的健康饮酒策略和疾病预防措施

叶欣安
叶酸代谢能力基因检测
依据国际公认的研究成果，精确计算受检者的叶酸代谢能力，并定制精准的孕期叶酸补充方案，呵护母婴健康

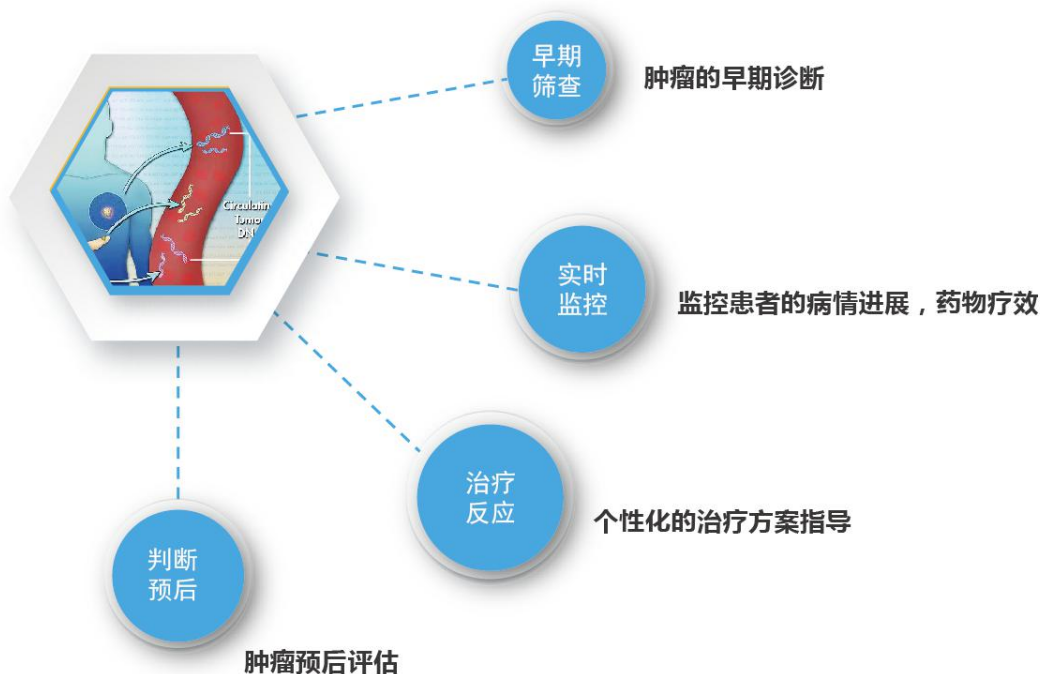
爱检康™

无创循环肿瘤 DNA (ctDNA) 检测

循环肿瘤 DNA (circulating tumor DNA , ctDNA) , 是一种存在于血浆或血清等体液中的细胞外 DNA , 主要来自于坏死、凋亡的肿瘤细胞和肿瘤细胞分泌的外排体。ctDNA 是游离 DNA (cell-free DNA , cfDNA) 中的一类, 所占比例较低 (0.1%-1%之间), 大小通常为 160-180 bp。由于 ctDNA 的无创性和易获得性, 在其中发现的肿瘤标志物, 被认为可以用于肿瘤早期诊断, 肿瘤进展监控, 预后判断及个性化用药指导等领域。



应用范围



ctDNA 检测的特点

1. 无创、安全

ctDNA 检测只需抽取血液即可完成对肿瘤细胞 DNA 的检测，避免穿刺、手术等有创方式采集肿瘤组织所带来的风险和痛苦。

2. 样品易获取，可实时多次采样进行检测

在恶性肿瘤发生早期，ctDNA 含量和基因的异常就可以被检测到，可实现实时、多阶段、个体化诊疗服务。

3. 灵敏度高，准确度高，假阳性率低

利用 NGS 检测 ctDNA，灵敏准确，可检测到低至 0.1% 的低频突变。ctDNA 具有更高的敏感性，可以更早的反应肿瘤的进展情况和治疗效果。

4. 特殊患者可受益

对于较难通过手术或穿刺取得肿瘤组织的肿瘤患者，也可通过 ctDNA 检测获益。

服务流程



我们的优势

- 自主研发的游离 DNA 建库方法，操作更便捷，浓度和纯度更高，建库 DNA 起始量降至 10ng；
- 检测内容全面，一次建库即可检测所有类型的突变，包括 SNP、Indel、Fusion、CNV 等多种突变类型；
- 采用目标区域的捕获测序法，测序深度高达 10000X，有效保证超低频突变的充分检测，可检测低至 0.1% 的突变；
- 专业的信息分析，优化突变过滤算法，有效降低背景噪音，保证超低频突变检测的准确性；

Production and Service

产品与服务 » 生物云平台|基因生态，共享未来

百迈客云平台是国内最大的生物大数据信息分析平台，是全球第一个针对生命科学家的“Turn-key BioCloud”，提供数十款集成式分析软件和百款工具，汇集八大数据库、文献库和云课堂，能让生物学家自主进行可视化、交互式、一键式基因大数据深度挖掘，加快发SCI论文速度，产出更多高水平科研成果。

百迈客云平台登录网址：www.biocloud.net。

平台特点

a) 图形化操作界面，无需用户掌握编程基础

数据导入、参数设定、任务投放、结果查看&整理均可在平台网页界面完成，区别于传统生信分析过程所使用的 linux 命令行界面，更加适用于国内无编程基础的基因组学研究用户。

b) 高度集成化的分析流程，快速完成数据深度挖掘

平台所部署分析流程集成化程度较高，一套流程可同时完成对数据的基础分析及个性化分析；所有流程的搭建、维护、更新均有后台有 8 年项目经验的生信团队完成，保证了流程的专业性以及分析结果的准确性。

c) 平台配套的公共数据库，有效降低研究成本、提升研究深度

整合分析已有公共数据可以有效的降低研究成本、提升数据分析深度，是基因组学研究中一项重要的分析手段，然而，传统的公共数据整合分析存在较高的技术门槛，要求研究者具备良好的生物信息基础。为解决这一问题，部署于 BMKCloud 的公共数据模块在数据规模同步了 SRA、GEO 两大世界权威测序数据库的基础上，又在数据检索、下载、分析等环节根据国内研究者的使用习惯进行了大量优化，目前已实现数据免下载，一键导入 BMKCloud APP 即可完成分析的生信 0 基础用户使用模式。

d) 云计算架构，对计算资源的利用更加高效

系统层采用云计算架构，实现计算资源按需分配、可弹性扩展，相较于传统的本地服务器使用方式，对计算资源的使用更加高效与成本。



科研服务

农业育种

医学检测

生信研发

平台特点



注册百迈客云平台账户



在确认电子邮件中验证您的账户



登录并开始探索您的数据

1.原始数据导入

2.综合选项

3.基因功能注释

4.差异表达分析

5.参数确认

lncRNA正式分析Cabbage项目

项目编码: lncRNA_Cabbage_20151113123955 项目完成时间: 2015/12/18

摘要

完成4个样品的长链非编码测序, 共获得42.87Gb Clean Data, 各样品Clean Data均达到10.18Gb, Q30碱基百分比在90.24%及以上。分别将各样品的Clean Reads与指定的参考基因组进行序列比对, 比对效率从49.48%到55.79%不等。基于比对结果, 进行可变剪接预测分析、基因结构优化分析以及新基因的发展, 发现新基因1,172个, 其中1,023个得到功能注释。基于比对结果, 进行基因表达量分析。根据基因在不同样品中的表达量识别差异表达基因, 并对其进行功能注释和富集分析。

图片预览



1 实验流程

- 1.1 RNA样品检测
- 1.2 RNA文库构建
- 1.3 文库质检
- 1.4 上机测序
- 2 生物信息学分析
- 2.1 测序数据及其质量控制
- 2.2 长链非编码数据与参考基因组序列比对
- 2.3 长链非编码文库质量评估
- 2.4 基因结构分析
- 2.5 新基因分析
- 2.6 基因表达量分析
- 2.7 差异表达分析
- 2.8 差异表达基因功能注释和富集分析
- 2.9 转录本拼接

三步注册

操作简易

详尽报告

个性化分析

基因挖掘

表达量挖掘

差异基因挖掘

转录因子预测

.....

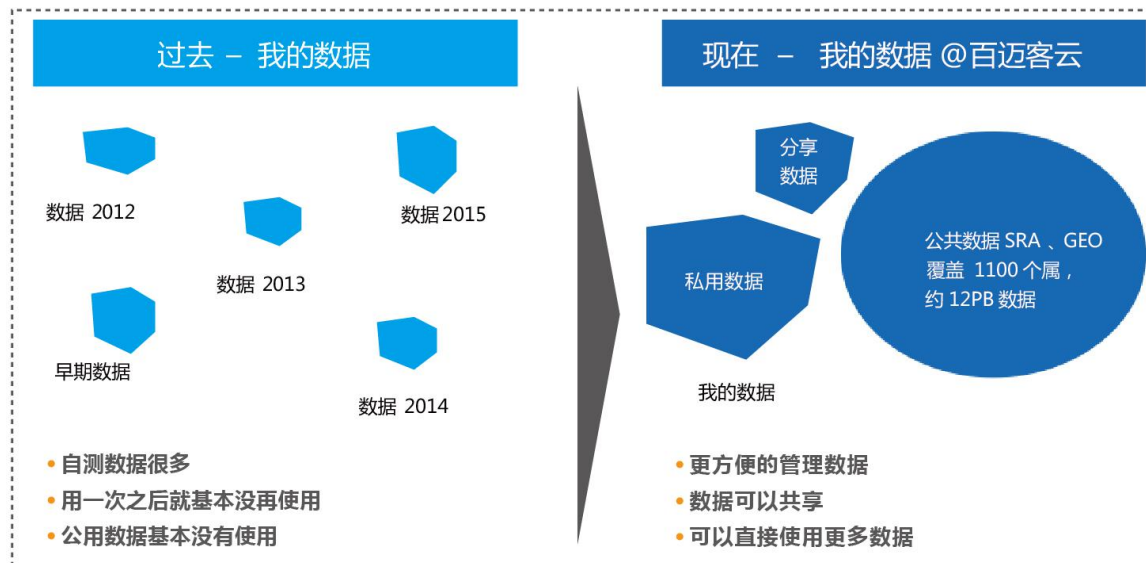


深入挖掘

可视化结题报

镶嵌式的数据深入挖掘模块, 快速实现从数据到文章

平台特点



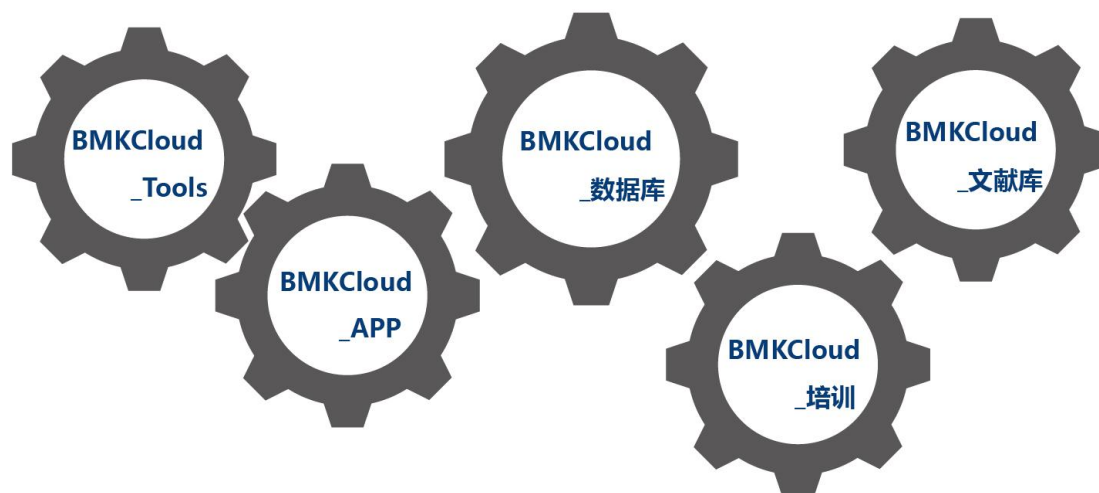
整合公共数据，让数据生产更多价值



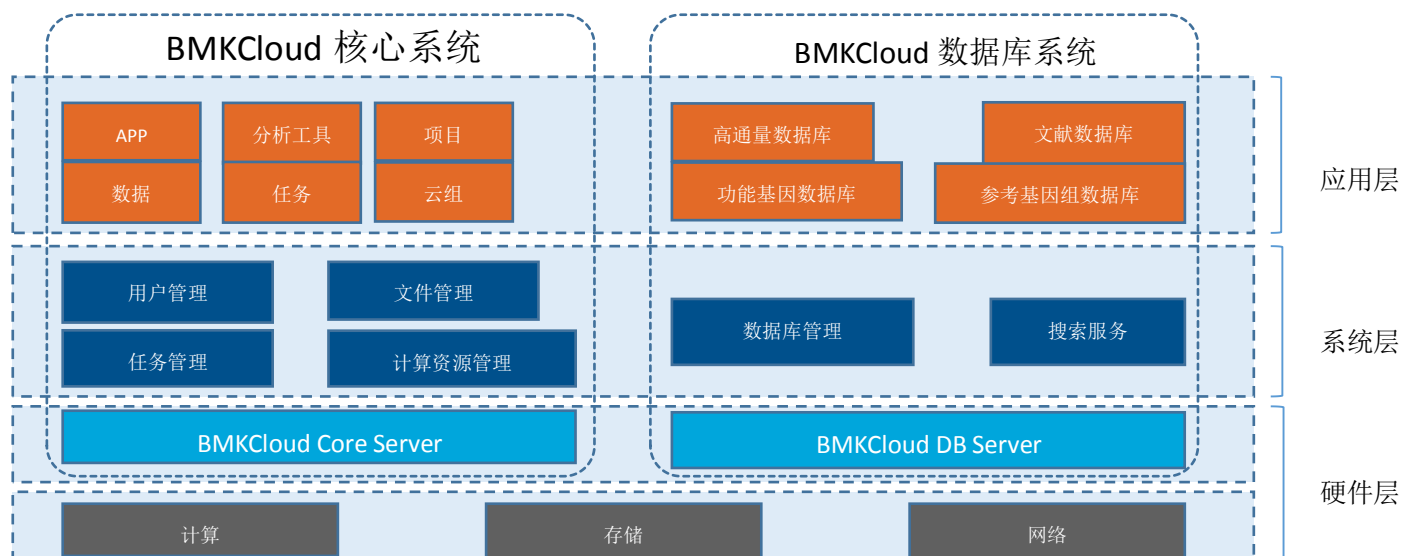
一个百迈客云账户=一个专业的生物信息分析团队

平台主要功能模块

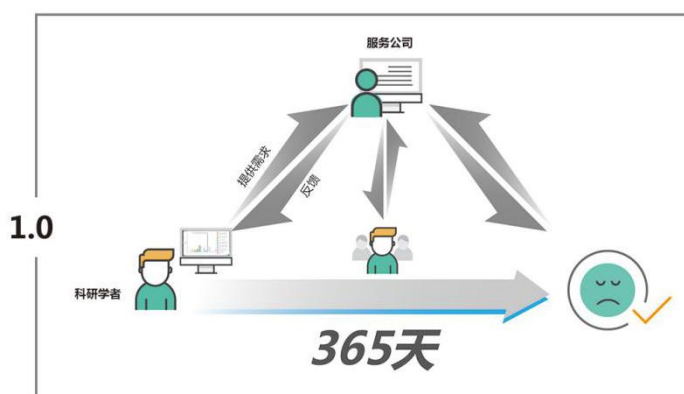
百迈客云平台 5 大功能模块



平台架构



百迈客云模式引领基因科技服务 2.0 时代

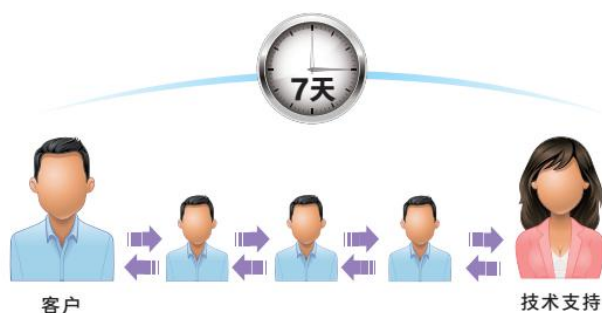


百迈客“云账号”主要针对做过或熟悉高通量测序,手里有一些没有完全挖掘的数据,还想充分利用公用数据的客户,经过简单的培训后,就可以在云平台上自主分析数据,多种分析软件都可以使用,深入挖掘自己以前的所有项目数据,也可以利用 NCBI 等公共数据进行挖掘研究,课题组的多个成员可以同时开展多个项目,提高研究的效率。



百迈客云平台 (BMKCloud) 自 2015 年 10 月正式商用以来已有 12000+ 注册客户。

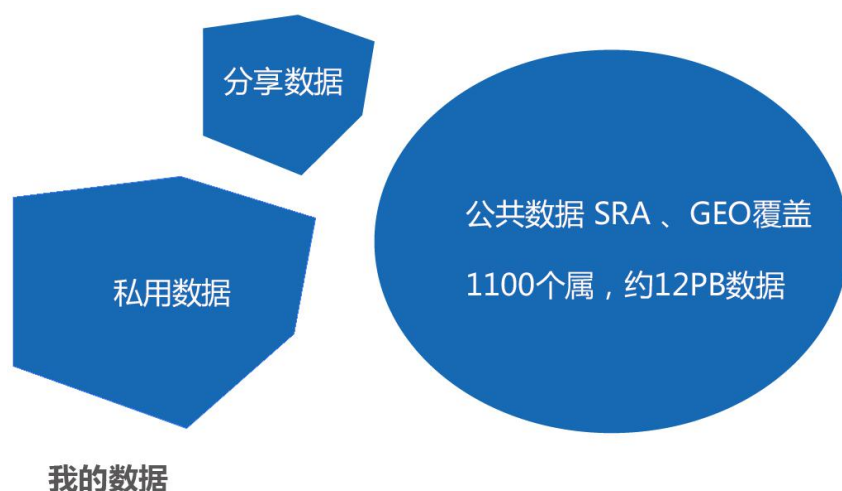
市场传统个性化解决方案



百迈客云个性化解决方案



公共数据整合分析场景篇



公共数据价值



物种	碱基数 (Tb)	物种	碱基数 (Tb)
模式植物	56	鸡	15
模式动物	679	猪	29
人	6627	小麦	33
玉米	39	番茄	17
杨树	38	大豆	21
牛	58	羊	22
水稻	74	棉花	17

NCBI SRA 数据库的数据已经达到 12PB 以上。

Nature Reviews Genetics 报道称,目前已发表的高通量测序数据利用率不足 20%,很多有价值的信息被研究者所忽略,基于公共数据挖掘和分析将成为科学研究发展的趋势。

最有代表性的是最大的癌症基因信息的数据库, Cancer Genome Atlas/TCGA 蕴藏着难以想象的宝贵信息,围绕 TCGA 已经有很多重量级文章出现, PubMed 收录的利用其进行数据挖掘发表的文章已经超过 1500 篇,而且对它的使用必然愈发重要。可以说,已经公开的高通量测序数据如同“金矿”一般,具有极高的学术价值。

可视化分析平台

无需编程、随想随做

五款 RNA 研究 APP



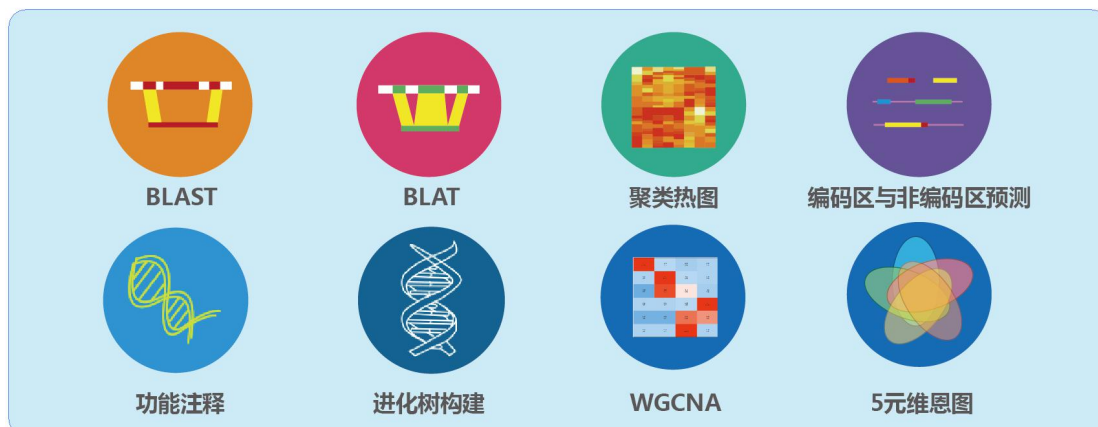
五款 DNA 研究 APP



标准化分析：可视化操作，参数自定义，无需编程即可做生信分析

个性化分析：深入解读，无需等待沟通结果，随想随做

100+ 分析工具



覆盖面全：数据处理、基因功能注释、绘图等功能一应俱全

灵活组合：小工具之间或与 APP 可灵活组合，适应特有分析流程

公共数据挖掘

探索数据背后的生物学价值

玉米公共测序数据库：一键下载、整合分析

48000G 玉米测序数据



9376 个样本数据

涉及研究方向：

发育调控类：玉米产量、叶发育、根发育、胚乳发育、胚芽发育、花药发育、

非生物胁迫类：干旱胁迫、淹水胁迫、寒冷胁迫、热胁迫、盐胁迫、低磷胁迫、氮胁迫、甘露醇胁迫、机械损伤应答、蔗糖处理胚乳、种子中 Zn 积累

生物胁迫类/免疫互动类：轮状镰刀霉菌感染、禾谷镰刀菌感染、立枯丝核菌感染、水稻黑条矮缩病毒感染、蠕孢菌感染、虫害应答 玉米叶斑病、吸胀冷害

其他：ncRNA 鉴定、转录因子鉴定、基因组组装、标记开发、遗传进化杂种优势、线粒体功能研究、表观遗传.....

玉米基因功能数据库：即时基因查询注释

52979 条玉米基因功能信息

功能信息涵盖

- 1) 功能注释、
- 2) 序列、
- 3) 相关文献、
- 4) 其他物种的直系同源基因、
- 5) 相关 ncRNA 信息、
- 6) 蛋白互作信息、
- 7) 公共项目中表达量信息、
- 8) 变异信息



玉米研究文献数据库：文献关联数据、快速验证思路

30018 篇

玉米研究相关文献

730 篇

影响因子 10 分以上文献

公共数据分析思路

分析思路 1

共表达分析中,整合大量相关公共样本测序数据,可构建出相较于常规样本量下可靠度更高的基因共表达网络,从而基于该网络进行更加准确的后续分析:a)预测目标转录因子的下游调控基因、目标调控网络中的关键转录因子;b)预测 ncRNA 与 mRNA 之间的靶向关系;c)基于网络中已知功能基因推测同网络中其他功能未知基因功能;e)将每个共表达模块分别作为一个整体,计算其与各组织或各发育时间点之间的相关性,建立各组织相关或各时期相关基因表达网络.....

支撑文献:

- [1] Wisecaver JH et al. A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. Plant Cell. 2017
- [2] Yu C et al. Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. Proc Natl Acad Sci . 2015

分析思路 2

研究某一类基因的转录水平在不同处理下(或不同组织部位间、或不同发育时间点等)的变化规律,可整合多个类似研究中的公共测序数据来共同揭示该变化规律,使结果更加可靠。

支撑文献:

- [1] Patil G. et al. Soybean (Glycine max) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. BMC Genomics. 2015
- [2] Matthijs M et al. Profiling of the Early Nitrogen Stress Response in the Diatom *Phaeodactylum tricornutum* Reveals a Novel Family of RING-Domain Transcription Factors. Plant Physiol. 2016

其他分析思路

1. 整合同一物种公共测序数据,构建物种完备转录本序列参考集,用于后续功能分析。

支撑文献:

- Iyer MK et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015
- 2. mRNA-ncRNA 联合分析中,利用公共数据补充其中一种类型的 RNA 测序数据。



扫码阅读更多详细内容

支撑文献:

Liu X et al. MicroRNA-mRNA regulatory networking fine-tunes the porcine muscle fiber type, muscular mitochondrial respiratory and metabolic enzyme activities. BMC Genomics. 2016

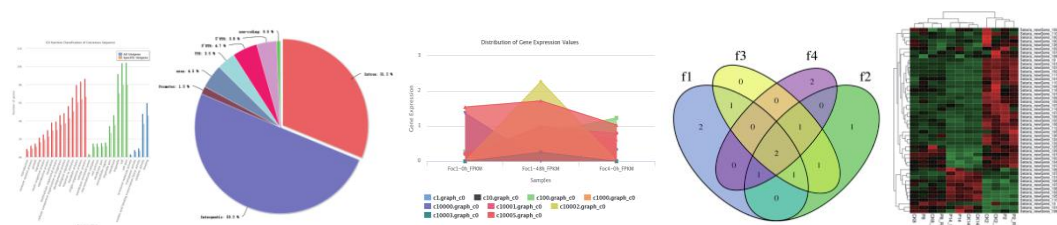
云平台 72 款小工具免费用

72 款小工具免费用

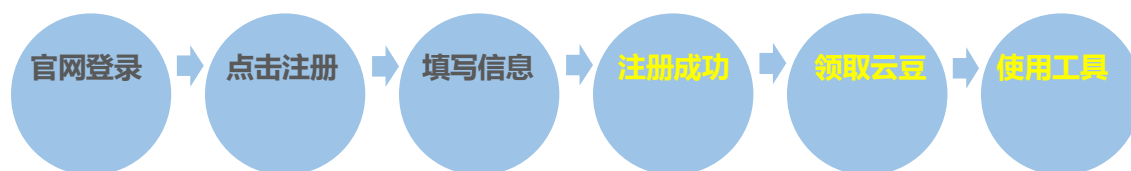
聚类热图、PCA 分析、GO/KEGG 分类富集图、lncRNA 分类及对应基因注释、遗传共线性分析、差异基因 GO 富集分析、柱状图、多平面(纵向)点线图绘制、miRNA 基因家族分类、参考基因组 reads 统计分布图、关联区域分析、单因素曼哈顿图、绘图工具等。

12 大功能类别

基因分析、遗传进化、ncRNA、质控、组装、比对、数据提取、突变、统计、表格处理、绘图、序列分析等。



操作指南



获得云豆的方法	获得云豆数量 (个)
注册	200
完善个人信息	200
邀请好友注册	2000
签到	50
每日首次登录	3
上传文献	500
项目评价	100