

华大基因
BGI

高分文章锦集

BGISEQ助力科技服务

【文章218篇】



深圳华大基因股份有限公司

▲ 华大基因(深圳总部)

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

▲ 欧洲华大基因(哥本哈根)

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

▲ 亚太、港澳台华大基因(香港)

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

▲ 美洲华大基因(波士顿)

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

▲ 日本华大基因(神户)

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



BGISEQ 平台简介

BGISEQ-500 桌面型、高通量、一站式、开放测序平台



精准

- DNA 纳米球 • Patterned array

采用优化的联合探针锚定聚合技术（cPAS）和独有的DNA纳米球（DNB）核心测序技术，结合芯片位点阵列排布（Pattern Array）技术和高分辨率成像系统，其中DNB通过线性扩增增强信号，降低单拷贝的错误率，使测序准确度达到前所未有的新高度。BGISEQ测序准确度达到同行业领先水平。

简易

- 一键测序 • 自动化应用分析

BGISEQ采用智能化图标触控式操作系统，测序操作简单便捷。可实现自动化样本管理，通过对样品、芯片、试剂等标签信息的自动化读取，内置专业的高性能应用分析软件，可提供一键测序功能，自动完成信息分析。

灵活

- 双芯片独立运行 • 多种通量模式可选

配备双芯片平台，可适配FCL和FCS两种芯片类型，通过不同设置可实现16种测序模式，数据产出通量在8-200Gb之间自由选择，可满足实验室和临床不同应用情景的测序需求。目前主流的测序仪如果要产生这十多种模式至少需要3台仪器才能完成，这大大节省了整个实验室仪器采购和管理成本。

快速

- 一站式测序操作流程 • 最快完成周期24小时

提供一站式测序操作流程，测序流程清晰明确。配合可选的全自动文库制备系统和全自动样品加载系统，从DNA到数据分析结果，最快完成周期可缩短在24小时内。

MGISEQ-2000

MGISEQ-2000采用全新的芯片系统，能够灵活支持多种不同的测序模式，并采用优化设计的光学及生化系统，能够在较短时间内完成完整的测序流程，带给使用者更加精简流畅的测序体验。

单次运行数据产出18.75-1080Gb；满负荷PE100仅需48小时；支持多种读长；支持多领域应用。三大核心技术：DNB、Patterned array、优化的联合探针锚定聚合技术（cPAS）。



BGISEQ 平台历程

2015年6月27日

华大基因在中国展出一款桌面化的测序仪系统BGISEQ-500。

2016年3月19日

BGISEQ-500亮相中国发展高层论坛创新前沿展。

2016年6月1日

亮相国家“十二五”科技创新成就展。

2016年7月23日

BGISEQ-500助力完成中国首例输入性裂谷热病毒基因组高通量测序。

2018年10月25日

在第十三届国际基因组学大会（ICG-13）上，震撼发布超高通量基因测序仪MGISEQ-T7。

2019年1月7日

MGISEQ-T7亮相第37届J.P.摩根健康产业大会。

2015年10月24日

在第十届国际基因组学大会（ICG-10）上，全球最大的基因组学研发机构华大基因发布了其自主研发的新型桌面化测序系统BGISEQ-500。

2016年5月27日

华大基因初夏发布会发布了测序应用整体解决方案BGISEQ-500n，这是全球测序技术领域首次大批量列装“中国造”标志着中国在高端测序技术的源头创新和上下游协同发展的突破及引领。

2016年7月5日

BGISEQ-500系列基因测序仪规模化生产。

2017年10月28日

在第十二届国际基因组学大会（ICG-12）上，发布了MGISEQ-2000测序仪。

2018年11月15日

MGISEQ-2000亮相世界公认最大医疗设备展览会MEDICA

2019年2月11日

MGISEQ-2000亮相韩国基因组学机构召开的KOGO冬季研讨会。



BGISEQ 平台科技应用发展历程



BGISEQ 平台测序原理

样品通过超声波高性能样品处理系统 (Covaris) 随机打断, 经过片段选择后得到适合长度的片段。随后进行 DNA 片段末端修复, 3' 端加上 "A" 碱基, 两端加上文库接头。接头连接后的文库进行线性扩增 (LM-PCR)。取适量扩增产物进行单链分离和环化处理, 环化文库经过滚环复制 (rolling circle amplification, RCA) 生成 DNA 纳米球 (DNA Nano Ball, DNB), 质控合格后即可上机测序。

BGISEQ 采用优化的联合探针锚定聚合技术 (cPAS) 和改进的 DNA 纳米球 (DNB) 核心测序技术, 是行业领先的高通量测序平台之一。具体而言, 首先 DNA 分子锚和荧光探针在纳米球上进行聚合, 随后高分辨率成像系统对光信号进行采集, 光信号经过数字化处理后即可获得待测序列。其中, DNB 通过线性扩增增强信号, 降低单拷贝的错误率。而且, DNB 大小与芯片上活性位点的大小相匹配, 每个位点结合一个 DNA 纳米球, 在保证测序精度的情况下提高了测序芯片的利用效率。采用先进的半导体精密加工工艺, 在硅片表面形成结合位点阵列, 实现 DNA 纳米球的规则排列吸附。

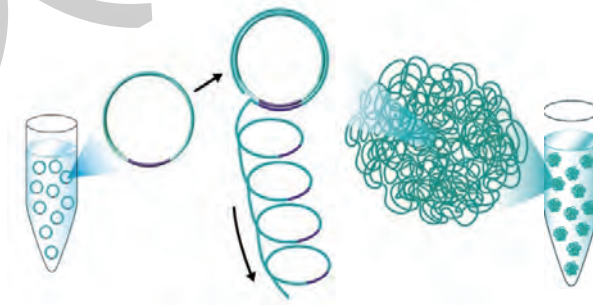


图1 DNA纳米球 (DNB)

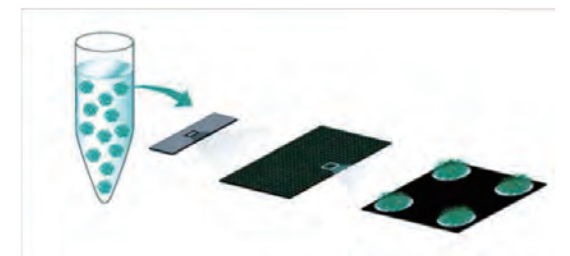


图2 独创的硅基纳米阵列

测序得到的原始图像数据, 经碱基识别软件 (Base Calling) 转化为原始序列数据 (raw reads), 数据以 FASTQ 文件格式存储。

BGISEQ平台优势

1、DUP低，测序准确性高

重复序列Duplicate reads (Dup)，指通过测序得到的两对或两对以上的Pair-End Reads同时比对到参考基因组上相同的起始和结束位置的序列。这些重复序列在总测序序列中占比简称为Dup rate。

解决duplicates最根本的方法是在源头尽量减少duplicates的引入。与I平台排他性扩增不同，BGISEQ平台采用DNB核心测序技术，具有独特的线性扩增模式。

DNB技术是目前全球少有的能在溶液中完成模板扩增的技术，在扩增过程避免错误累积，有效提高测序准确度。原理：将DNA进行片段化处理，加接头序列，环化，形成单链环状DNA；通过滚环扩增(Rolling circle amplification, RCA)将单链环状DNA扩增2-3个数量级，最终产生的扩增产物即DNB。因基于滚环扩增，DNB技术有效增加了待测DNA的拷贝数，大大增强了信号强度；即使复制过程中引入单个碱基的复制错误，这个错误也不会像PCR那样把信号放大。

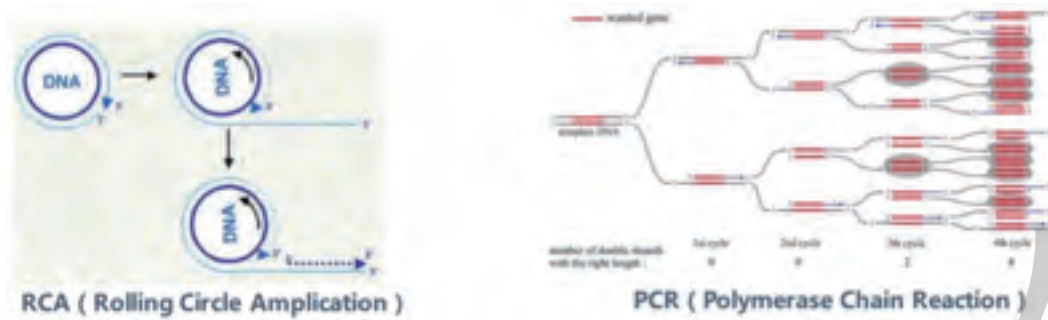


图3 RCA与PCR技术差异展示

完成模版扩增后，将DNB转载到Patterned Array (规则阵列)上。Patterned Array采用先进的纳米硅半导体精密加工工艺，实现DNB的规则排列、信号均匀、密度高，且彼此互不干扰，提供了最高的成像效率和最优的试剂用量，不仅保证了测序准确度，而且提高了测序芯片的利用效率。因DNB在溶液里提前完成扩增，loading过程中没有聚合酶、引物和dNTP等PCR条件，所以该技术的duplicate rate更低。

2、无Index Hopping担忧

样品混合测序，将独特的index加到每个样品上，并将多个样品合并一起在同一run中测序，降低测序成本，使基因测序得到广泛应用。

研究发现，基于ExAmp(排他性扩增)的测序平台，例如H、N平台，混样上机测序会出现index错误分配(index hopping)问题，样本错误分配率超过1%，且对于PCR-free文库，index hopping比例可高达6%。即使采用繁琐的non-combinatorial dual index方案，index污染率也只能降低到0.08%。



图4 Illumina index hopping白皮书

BGISEQ平台极低的index hopping比例可以有效避免样本的“张冠李戴”，同时可带来：

- 1、高的检测准确度，保证体细胞低频突变、HPV检测等基因检测的准确性；
- 2、Single index避免了繁琐的non-combinatorial dual index带来的额外成本和劳动力浪费；
- 3、避免高通量测序中样本数据完整性的丢失。

3、严格质控保驾护航

华大高通量实验室，每年度均参与并通过《实验室室间质量测评》，严格质量控制流程保证结果准确度。武汉实验室的资质有ISO 9001质量管理体系认证、ISO 14001环境管理体系认证、OHSAS 18001职业健康安全管理体系认证、ISO 17025检测实验室认可、ISO 27001信息安全管理体系认证。深圳实验室有ISO 9001、ISO 14001、OHSAS 18001、ISO 15189医学实验室认可、ISO 27001。

临床级别的华大实验室保证每一样本测序结果的准确性，对于样本检测、文库构建、上机测序、数据下机及信息分析每一环节均有严格的质量控制流程，保证每一文库每一样本的高效、快速、准确的执行。



图5 BGISEQ平台质控环节

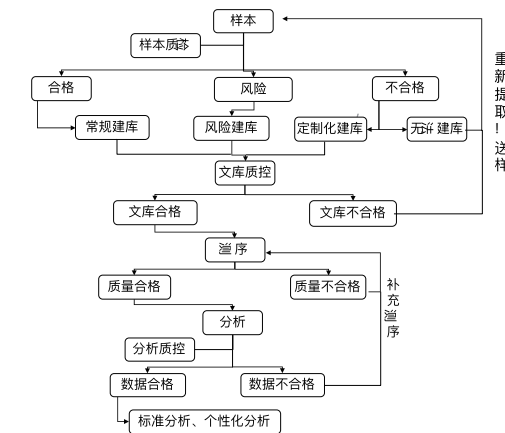


图6 BGISEQ平台质控详细流程图

4、发文218篇，影响因子1073

截止2019年2月，BGISEQ平台共发表文章218篇，累计影响因子1073。已发表文章涵盖人方向文章72篇，动物方向68篇，植物方向51篇，微生物19篇，涉及105种物种。

2018年8月，BGISEQ助力解析高效利用氮肥调控机制，推动作物绿色变革新方向，发表于《Nature》杂志。影响因子大于5的文章共79篇，其中影响因子大于10的文章共20篇。

目录

真核转录组&RNA-Seq / 01

- 02 | 文章1: 植物的生长代谢调控促进可持续农业
- 06 | 文章2: 甲基转移酶SETD2介导的STAT1甲基化修饰在IFN的抗病毒免疫中起关键作用
- 09 | 文章3: 代谢通路的协同作用: 通过耦合光学纯的乳酸生物合成提高1,3-丙二醇生产中的碳平衡
- 13 | 文章4: 桑黄多糖增强高脂高糖饮食小鼠的胰岛素敏感性
- 18 | 文章5: 拟南芥AtACS8基因在铜离子激活下, 在乙烯早期生物合成中起关键作用
- 21 | 文章6: 两个叶甲姐妹种化学感应基因的比较转录组学分析, 揭示化学感应物种形成
- 23 | 文章7: 药物引起果蝇类似抑郁症的表型和基因表达谱改变
- 26 | 文章8: 过表达OsHSP18.0-Cl提高水稻对细菌性条斑病的抗性
- 29 | 文章9: 转录组水平分析发现酿酒酵母FKS1突变体中MAPK通路和碳代谢在细胞壁维持中的作用
- 33 | 附表: 真核转录组&RNA-Seq

Small RNA测序 / 37

- 38 | 文章1: 基于cPAS的BGISEQ-500平台非编码小RNA测序研究
- 40 | 文章2: 甘蓝、芥蓝嵌合体的无性后代嫁接诱导表观遗传变异的维持
- 43 | 附表: Small RNA文章列表

ChIP-Seq / 44

- 45 | 文章1: FOXN3-NEAT1-SIN3A阻遏物复合物促进激素反应性乳腺癌的进展
- 50 | 附表: ChIP-Seq文章列表

全基因组重测序 (WGS) / 51

- 52 | 文章1: BGISEQ-500测序仪的参考人类基因组数据集
- 56 | 文章2: BGISEQ-500与Illumina HiSeq 2500测序平台对古基因组测序性能的比较
- 60 | 文章3: BGISEQ-500和HiSeq X Ten全基因组测序鉴定生殖细胞和体细胞变异
- 64 | 文章4: 基于DNB的测序平台可有效避免index hopping

全外显子测序 (WES) / 67

- 68 | 文章1: 首篇BGISEQ外显子应用文章, 助力恶性卵巢癌多组学研究
- 72 | 附表: WES文章列表

宏基因组学 (Metagenomics) / 73

- 74 | 文章1: 基于cPAS的BGISEQ-500平台对宏基因组测序的评估
- 77 | 文章2: 基于BGISEQ平台完成全球首个大鼠肠道微生物参考基因组构建
- 79 | 附表: Meta文章列表

单细胞测序 / 80

- 81 | 文章1: 单细胞多组学研究揭示了细胞调节异质性
- 84 | 附表: 单细胞测序文章列表

华大基因介绍 / 85

真核转录组&RNA-Seq

新一代高通量基因组测序仪的迅速发展 (Solexa, 454 GS-FLX, SOLiD, tSMS) 不仅给基因组领域带来革命性的突破, 同时也给转录组检测方法带来重大革新。采用类似SAGE技术和MPSS技术的理念, 新一代高通量基因组测序仪可以通过测定细胞全部转录产物序列, 通过序列比对得到最后的转录组, 标志着一个新的测定转录组的“RNA测序”法出现, 该技术称为RNA测序技术 (RNA Sequencing, RNA-Seq)。

基因转录水平的研究是功能基因组学和医学研究的基础。RNA-Seq是用来研究某一生物对象在特定生物过程中基因表达差异的技术。该技术结合了转录组测序建库的实验方法与数字基因表达谱 (Digital Gene Expression Tag Profiling, DGE) 的信息分析手段, 可广泛应用于生理调控、农业性状、生物标记、环境改造、疾病机制和药物筛选等领域。

文章1

植物的生长代谢调控促进可持续农业

Modulating plant growth - metabolism coordination for sustainable agriculture



发表期刊: *Nature*

影响因子: 41.58

研究对象: 水稻

合作单位: 中国科学院遗传与发育生物学研究所、
中国科学院大学生命科学学院

研究背景

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

研究方法

取 材: 水稻半矮秆品系NJ6-*sd1*, 对应野生型品系NJ6, NJ6-*sd1*的氮素利用效率比NJ6低

研究方法: 以NJ6 (*SD1*)为对照, 对36个含有*sd1*的籼稻品种分析NH₄⁺的利用率。再以NJ6为轮回亲本与一个高NH₄⁺吸收率品系NM73杂交创建BC1F2群体, 通过QTL定位、图位克隆等技术获得氮肥高效利用的关键基因。分析关键基因启动子的三种SNP分型, 筛选不同型别影响作物产量的潜在联系

测 序: RNA-Seq BGISEQ; ChIP-Seq BGISEQ

研究概要

研究深入解析作物生长、氮素利用相关GRF4-DELLA的平衡调控机制, 不影响优良性状作物半矮化的同时提高产量。GRF4的遗传变异将成为提高作物产量和养分利用效率的主要目标, 新的育种策略将引领未来的绿色革命, 增加产量可持续化的同时, 减少因使用农业氮而导致的环境恶化。

主要结果

1、GRF4促进“绿色革命”作物水稻对NH₄⁺的摄取

在水稻和小麦中，DELLA的积累抑制作物生长和氮吸收。利用RNA干涉将GRF4沉默或降低表达，证明GRF4可以促进对NH₄⁺的吸收。

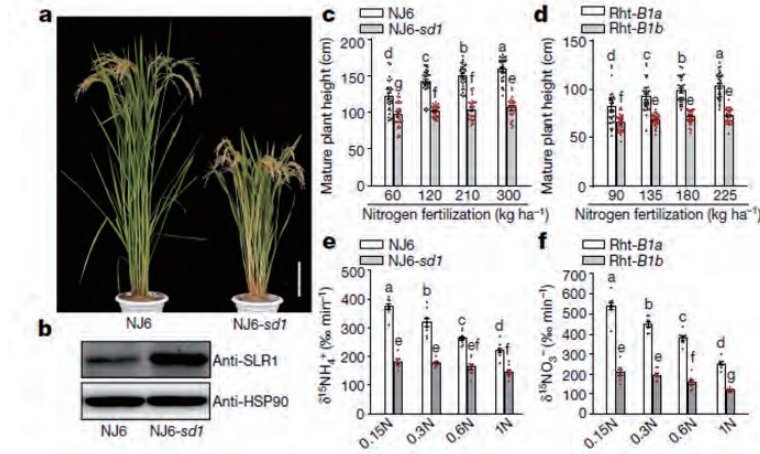


图1 DELLA的积累抑制水稻和小麦的生长及氮素吸收

a 水稻半矮杆品系NJ6-*sd1*，对应野生型品系NJ6。
b SLR1的积累。HSP90为对照，印迹结果。
c 水稻高度。
d 小麦高度。
e NH₄⁺不同氮浓度下的摄取速率。
f NO₃⁻不同氮浓度下的摄取速率。

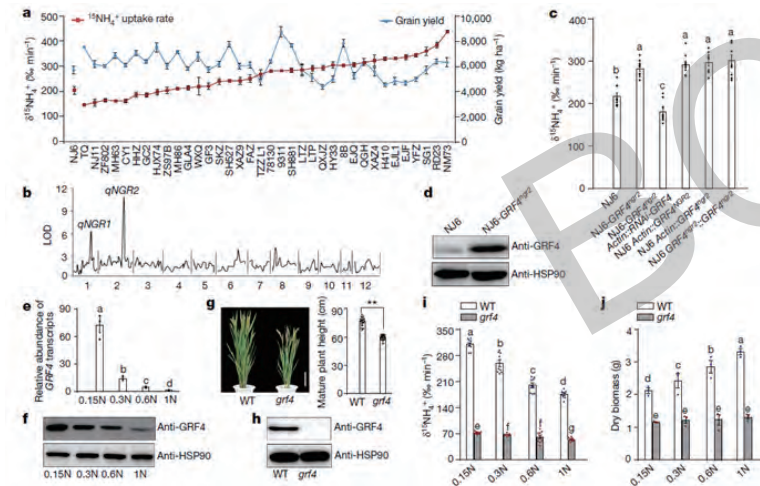


图2. GRF4调节水稻吸收NH₄⁺，以及对氮素利用率的生长反应

a NH₄⁺吸收和产量变化。
b QTL分析。
c NH₄⁺摄取率。
d GRF4的累积。
e 增加氮素供应，NJ6根中GRF4转录物丰度。
f NJ6中GRF4的累积。
g 成熟株高。
h NH₄⁺摄取率。
i 四周龄植物的干重。

2、GRF4-SLR1相互作用调节氮代谢

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

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

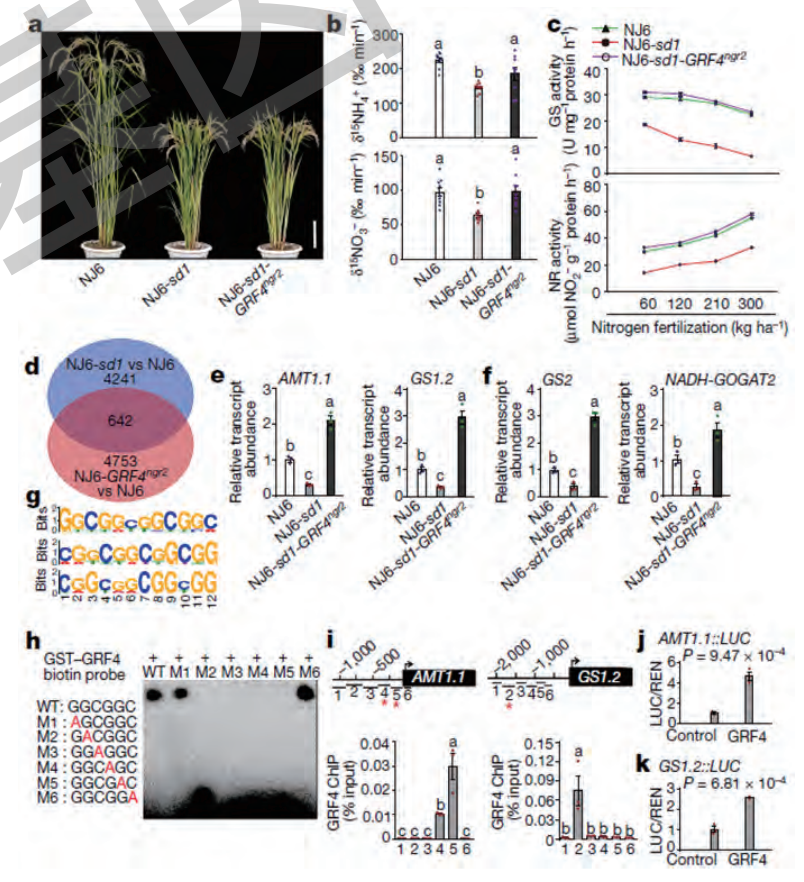


图3 GRF4调控多种氮代谢基因的表达

研究结论

(1) GRF4-DELLA 相互平衡作用调节植物的生长和代谢。GRF4 (GROWTH REGULATING FACTOR 4) 是生长调节因子, 调节促进植物氮代谢和稳态, 也协调促进植物碳代谢和生长, 而 DELLA 生长阻遏物抑制这些过程。GRF4 与 DELLA 是拮抗调节关系来维持平衡。

(2) 半矮化作物中, 在适度的氮素水平下增加 GRF4 的表达, GRF4-DELLA 平衡机制偏向于 GRF4 的促进作用, 从而利于氮、碳的吸收, 增加叶、茎的宽度, 但对植物的高度影响很小。因此, 改善氮素吸收效率的同时, 不影响半矮化性状, 进一步提高作物产量。

文章2

甲基转移酶SETD2介导的STAT1甲基化修饰在IFN的抗病毒免疫中起关键作用

Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity



发表期刊: *Cell*

影响因子: 31.40

研究对象: 人

合作单位: 浙江大学医学院免疫学研究所、
中国工程院院士曹雪涛研究团队

研究概要

全球由乙型肝炎病毒 (HBV) 感染而导致的乙肝患者超过3.5亿, 其中约1/3在中国。除了接种疫苗, 目前广泛使用抗病毒天然免疫细胞因子IFN α (I型干扰素) 治疗, 但效率低。为揭示其分子机制, 研究团队通过高通量小RNA干扰筛选体系筛选了711个已知的表观遗传修饰基因, 发现组蛋白甲基转移酶SETD2表达对于干扰素抵抗HBV至关重要。通过肝细胞特异性敲除SETD2基因的小鼠模型, 确定SETD2增强干扰素抑制HBV作用。发现SETD2分子通过其SET结构域发挥甲基转移酶活性, 直接催化信号蛋白STAT1的第525位赖氨酸发生单甲基化修饰, 从而促进干扰素效应信号的活化。此外, SETD2选择性地催化干扰素诱导性基因远端启动子区发生组蛋白H3K36me3的修饰, 从而直接促进这些具有抗病毒作用基因的转录活化和表达。该研究揭示了SETD2分子直接催化信号蛋白STAT1甲基化修饰的新机制, 为病毒感染性疾病的治疗提供了新靶点。

研究方法

取 材: 4组不同的人细胞样品, 每组2个生物学重复, 共8个样品

研究技术: siRNA库高通量筛选、赖氨酸甲基化质谱分析、RNA干扰、qRT-PCR分析、流式细胞分析、ELISA、免疫沉淀反应、Western杂交、Pull-down试验、体外甲基化分析、免疫荧光、ChIP、BGISEQ RNA-Seq等。

主要结果

1、SETD2经SET域 (包含片段) 促进STAT1活化, 从而提高干扰素诱导的抗病毒功能

HepG2细胞中敲除SETD2，产生出3个SETD2-KO细胞系，呈现出SETD2表达缺失和总的H3K36me3水平下降（图1F），与对照HepG2相比，都表现出干扰素诱导的STAT1磷酸化作用受损（图1G）。观察干扰素处理的SETD2-KO HepG2细胞，*IFNAR1/R2*或STAT1上游信号分子的表达没有显著差异。RNA-Seq分析发现一系列ISGs响应干扰素刺激，在SETD2-KO细胞比HepG2 WT细胞表达更低（图1H）。qRT-PCR分析确定干扰素刺激后在SETD2-KO细胞中两个ISG表达降低，包括ISG15和MX2（图1I）。所以，在干扰素刺激下，SETD2提高STAT1磷酸化作用和ISG表达。

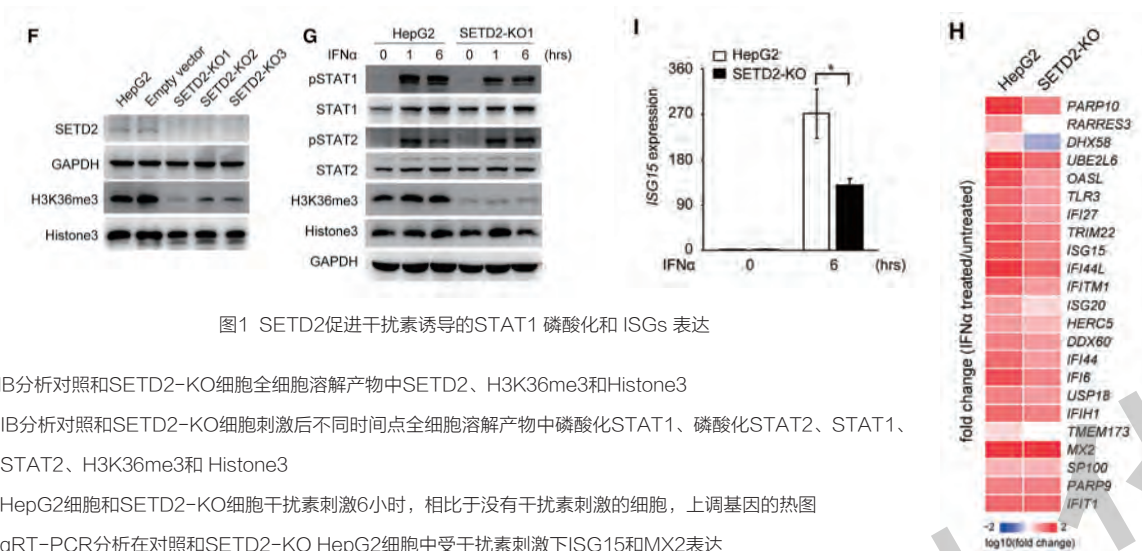


图1 SETD2促进干扰素诱导的STAT1 磷酸化和 ISGs 表达

F IB分析对照和SETD2-KO细胞全细胞溶解产物中SETD2、H3K36me3和Histone3

G IB分析对照和SETD2-KO细胞刺激后不同时间点全细胞溶解产物中磷酸化STAT1、磷酸化STAT2、STAT1、STAT2、H3K36me3和 Histone3

H HepG2细胞和SETD2-KO细胞干扰素刺激6小时，相比于没有干扰素刺激的细胞，上调基因的热图

I qRT-PCR分析在对照和SETD2-KO HepG2细胞中受干扰素刺激下ISG15和MX2表达

2、SETD2选择性促进H3K36me3修饰的ISGs表达

利用RNA-Seq、SETD2-KO细胞、STAT1-K525A-Re细胞、STAT1-KO细胞以及对照的HepG2 细胞在干扰素刺激下6小时（图2A）。干扰素刺激后对照HepG2细胞中222个基因被诱导，SETD2-KO细胞、STAT1-K525A-Re细胞和STAT1-KO 细胞，与对照HepG2细胞相比，分别有89、128和180个基因显著性下调（图2B）。GO富集分析显示SETD2-KO细胞和STAT1-K525A-Re细胞中表达更低的基因被归类于病毒的防御反应或I型干扰素通路相关基因（图2C）。STAT1-KO细胞在干扰素刺激下下调的180个基因中有40%的基因在SETD2-KO细胞在干扰素刺激下也表达更低，包括SETD2促进一系列的STAT1依赖的ISG（图2D）。另外，干扰素刺激后的SETD2-KO细胞和STAT1-K525A-Re细胞中，75个基因表达更低，包括有抗病毒功能的ISGs，如ISG15、ISG20、MX2等（图2D和2E）。进一步确定SETD2介导的ISGs调控是在很大程度上依赖于在介导STAT1 K525单甲基化作用中它起到的作用。

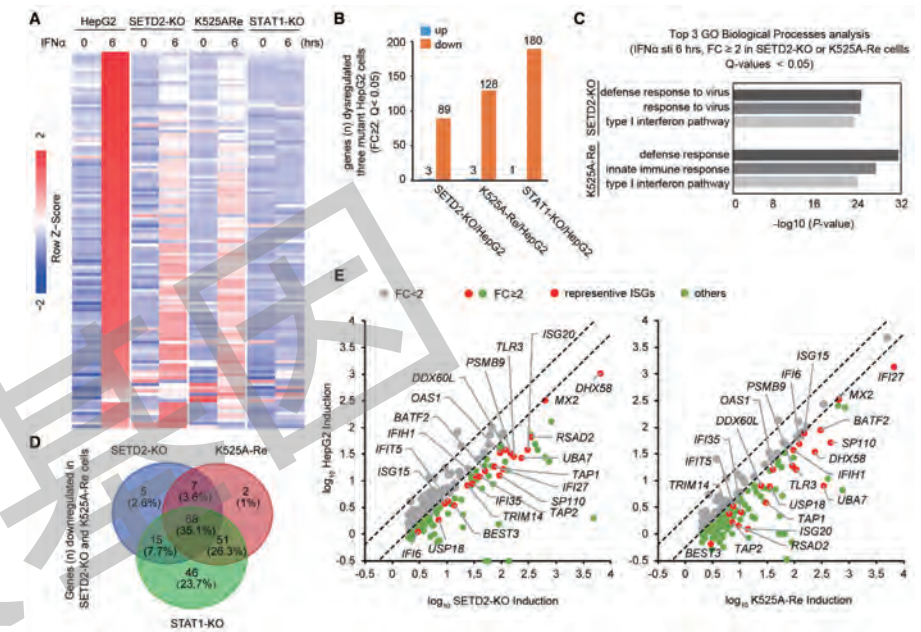


图2 SETD2选择性催化一些ISGs的H3K36me3 修饰

A 干扰素刺激6小时，HepG2、SETD2-KO、K525A-Re和STAT1-KO细胞中ISGs mRNA表达热图；

B SETD2-KO、K525A-Re和STAT1-KO HepG2细胞相对于对照HepG2细胞的差异诱导基因统计；

C GO分析干扰素刺激6小时，SETD2-KO、K525A-Re与对照HepG2细胞相比的差异表达基因；

D 干扰素刺激6小时，SETD2-KO、K525A-Re和STAT1-KO的差异表达基因维恩图；

E 干扰素刺激下SETD2-KO（左）、K525A-Re（右）细胞相对于对照HepG2细胞抑制表达基因散点图。灰点代表两种细胞中表达没变化，红点代表选择的ISGs在两种细胞中都表达更低，绿点代表其他基因。

文章3

代谢通路的协同作用:通过耦合光学纯的乳酸生物合成提高1,3-丙二醇生产中的碳平衡

Coordination of metabolic pathways: Enhanced carbon conservation in 1,3-propanediol production by coupling with optically pure lactate biosynthesis



Metabolic Engineering
Volume 41, May 2017, Pages 102-114



Coordination of metabolic pathways: Enhanced carbon conservation in 1,3-propanediol production by coupling with optically pure lactate biosynthesis

Bo Xin^{a, b, 1}, Fei Tao^{a, 1}, Yu Wang^a, Hongyu Liu^a, Cuiqing Ma^b, Ping Xu^{a, c, *}

发表期刊: *Metabolic Engineering*

影响因子: 7.67

研究对象: 克雷伯氏菌

合作单位: 上海交通大学、
山东大学和华东理工大学

研究背景

代谢工程是生物合成大量优质化学制品的一个强有力工具。自然条件下各种代谢途径的协同作用会带来复杂的代谢变化, 阻碍生物炼制的发展。

1,3-丙二醇 (PD) 是一种被广泛用于多聚物、化妆品、润滑剂及药物生产的化学原料, PD的全球需求量正快速增长。丙三醇是PD生产中极好的发酵底物, 利用微生物代谢从丙三醇中合成PD涉及氧化和还原两条平行分支。在氧化分支中, 丙三醇为细胞生长提供能量并形成一系列副产物; 在还原分支中, 通过消耗氧化反应分支中产生的NADH进行PD的生物合成。因此, 丙三醇的氧化反应对于PD的生物合成是必需的, 但同时也带来相当大量的副产物。

许多微生物都可从丙三醇中合成PD, 其中兼性厌氧型克雷伯氏菌 (*K. oxytoca*) 因其良好的底物耐受性、产量、生产效率以及易于遗传改造, 得到很多关注。至今几乎没有报道关注过丙三醇氧化还原反应产生PD的过程中, 在氧化分支最大程度的改变底物并形成新的平衡。本研究选取产酸克雷伯菌为宿主菌株, 选择乳酸作为副产物, 运用系统的代谢工程方法, 改造产酸克雷伯菌, 使其成为利用丙三醇生产PD和乳酸的高效细胞工厂。

研究策略

取材: *K. oxytoca* strain PDL-0(WT) 筛选自上海各地的土壤样本, 挑选其中产PD和乳酸能力最佳的菌株用于后续研究, 命名为PDL-0。经16S rDNA克隆和测序, 确定为*K. oxytoca*菌株。使用基因工程的方法对*K. oxytoca*菌株进行改造(图1)。

建库: 指数增长晚中期的*K. oxytoca*细胞用于RNA植被及文库构建

测序: BGISEQ RNA-Seq

分析: 原始数据过滤得到FASTQ格式的clean reads, 通过HISAT将clean reads比对到PDL-0基因组, 用RSEM软件包进行基因表达定量, 使用NOISeq方法筛选差异表达基因。

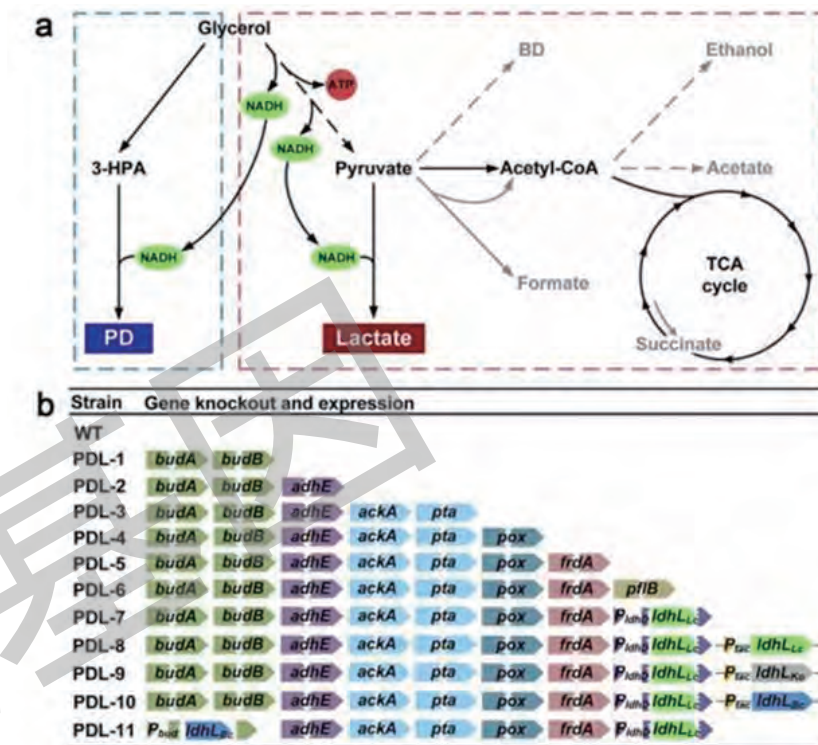


图1 *K. oxytoca* PDL-0联产PD和光学纯乳酸的代谢工程改造

(a) 从丙三醇中联产PD和乳酸的技术路线图。实线表示一步反应, 虚线表示多步反应, 黑线表示在代谢工程菌中被激活的通路, 灰线表示在代谢工程菌中被阻断的通路。红色框代表丙三醇代谢中的氧化分支, 蓝色框代表丙三醇代谢中的还原分支。

(b) 本研究使用的重组菌株。箭头代表基因表达, 断裂的箭头代表基因被敲除。

研究概要

代谢工程是利用基因重组技术有目的的对细胞代谢途径进行改造, 用以构建新的代谢途径来生产目的产物, 在改造菌株时往往采用生化方式对代谢产物进行测定, 无法从基因层面了解代谢通路的变化。本研究将BGISEQ测序平台的RNA-Seq分析与常规测定结果相结合, 一方面从基因水平验证了二者之间的一致性, 另一方面也能通过基因表达的变化或差异解释一些生理表象, 为后续研究奠定基础。

研究结果

- 1、选择乳酸作为*K. oxytoca*利用丙三醇产生PD过程中有利的副产物。
- 2、分析*K. oxytoca*PDL-0发酵丙三醇得到的代谢产物。
- 3、为了在*K. oxytoca*发酵丙三醇过程中得到更多PD和光学纯的D-乳酸, 通过阻断BD、乙醇、乙酸盐、琥珀酸

盐合成通路来减少其副产物的形成。

- ◇ 在PDL-1菌株中，BD合成通路阻断带来的丙酮酸累积；
- ◇ 在PDL-2菌株中，乙醇合成通路阻断使得PD和D-乳酸产量达到86%；
- ◇ 在PDL-3菌株中，ackA-pta的敲除减少了乙酸盐的形成
- ◇ 在PDL-4菌株中，poxB的敲除使乙酸盐浓度轻微下降；
- ◇ 在PDL-5菌株中，琥珀酸盐合成通路被阻断，琥珀酸盐产量下降，且对PD和乳酸产量无明显影响，使其成为良好的候选菌株。

4、阻断甲酸盐合成通路

- ◇ 在PDL-6菌株中，尽管甲酸盐副产物减少，但PD和D-乳酸产量显著下降。PDL-5菌株被用于后续实验。

5、细胞内NADH/NAD⁺分析

- ◇ PDL-5菌株的NADH/NAD⁺比率相当，说明其代谢通路的阻断，对于其细胞内的氧化还原反应无明显影响，而PDL-6菌株的NADH/NAD⁺比率则出现显著升高，是抑制其生长的首要因素。

6、RNA-Seq分析

- ◇ 为全面研究副产物生物合成通路被阻断带来的影响，PDL-1、PDL-2、PDL-4、PDL-5和PDL-6菌株突变体被用于转录组学分析，野生型菌株PDL-0作为对照。
- ◇ PDL-1、PDL-2、PDL-4、PDL-5和PDL-6菌株中分别发现125、172、592、867和733个差异表达基因。产量表现最佳的PDL-5突变菌株中共有404个上调基因，463个下调基因以及5611个非调节基因。（图2a）

◇ 主成分分析：研究不同突变体之间的关系，来自同一突变体的重复样本被聚类到一起，说明实验方法的准确性及重复性良好。PDL-6突变菌株明显离群，显示其与其他菌株具有很大的差异。这与该突变体与其他菌株相比细胞内NADH水平更高及生长状况更差的结论一致。（图2b）

◇ 差异表达基因聚类：挑选1054个基因，分成20个功能组。PDL-1到PDL-6在差异表达上呈现积累的趋势，说明代谢网络在逐步的遗传操作中可能确实是逐渐改变的。分析结果还显示突变菌株PDL-1的改变与其他菌株不太相同。

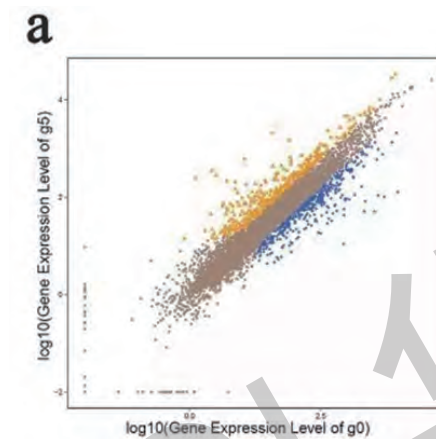


图2a PDL-5 (g5) 与PDL-0 (g0) 所有表达基因的比较；蓝色代表下调基因，橙色代表上调基因，棕色代表非调节基因

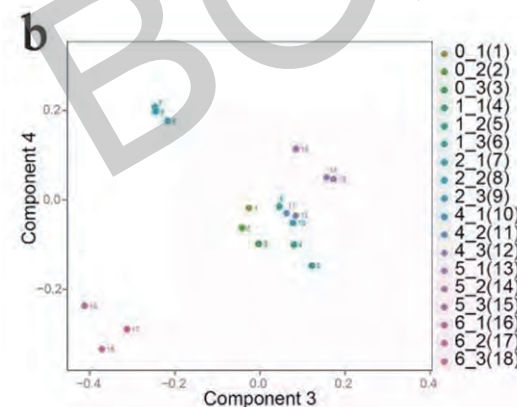


图2b 突变菌株的主成分分析；0_1,0_2,0_3代表PDL-0的三个重复，其他突变菌株也遵循此原则

◇ 图2c还显示在如氨基酸转运和代谢、碳水化合物运输和代谢、无机离子运输和代谢等功能组之间的一些显著变化。PDL-1菌株中，氨基酸转运和代谢增强，同时其他突变体减弱，尤其在PDL-5和PDL-6菌株中，这与此两株突变体生长稍弱是一致的。碳水化合物运输和代谢从PDL-1到PDL-6菌株中是逐渐增强的，正好可以解释逐渐增加的PD产量。在PDL-1菌株的无机离子运输和代谢途径中，上游基因先上调，然后逐步下调。表明在遗传改造后，细胞营养需求也改变了。（图2c）

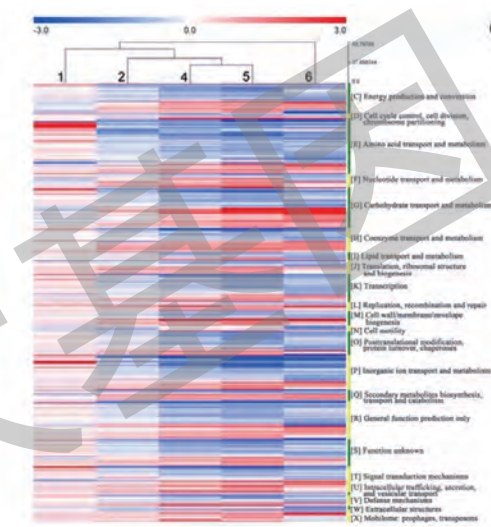


图2c 不同突变菌株与野生型比较的转录水平变化热图；1、2、4、5、6分别代表PDL-1、PDL-2、PDL-4、PDL-5、PDL-6

◇ 对与PD生物合成直接相关的差异表达基因更深入的研究发现，B12辅酶相关基因在PDL-5菌株中通常上调，与该菌株具有更好的表现是一致的。

7、最优化的PD和D-乳酸联合产物

◇ PDL-5作为最佳联产候选菌株，采用分批加料发酵法优化发酵条件，PD和D-乳酸转化率分别达到42.6%和52.8%，总转化率达95.4%，且未检测到L-乳酸。

8、PD和L-乳酸副产物

◇ 通过在PDL-7突变菌株染色体的*budAB*位点插入*ldhLbc*得到PLL菌株，其PD和L-乳酸转化率分别达到42.5%和53.4%，且副产物浓度很低或未检测到。

研究意义

研究亮点在于通过基因工程的方式对菌株进行改造，优化代谢通路，筛选出能从丙三醇中高效联产PD和光学纯的乳酸的*K. oxytoca*突变株。结合BGISEQ平台的RNA-Seq方法，对不同突变体及其生物学重复的基因表达进行分析，发现不同突变体的生物学重复被聚类到一起。同时，差异表达基因聚类也显示了代谢通路中的一些变化规律，从基因层面解释遗传改造后的生理变化。BGISEQ测序平台良好的重复性与准确性，与传统的生理生化指标测定相互验证补充，为高质量的代谢工程研究提供了基因水平的依据。

文章4

桑黄多糖增强高脂高糖饮食小鼠的胰岛素敏感性

Polysaccharides extracted from Phellinus linteus ameliorate high-fat high-fructose diet induced insulin resistance in mice.



Carbohydrate Polymers
Volume 200, 15 November 2018, Pages 144-153



发表期刊: Carbohydrate Polymers

影响因子: 5.158

研究对象: 小鼠

合作单位: 武汉轻工大学、香港大学

Polysaccharides extracted from *Phellinus linteus* ameliorate high-fat high-fructose diet induced insulin resistance in mice

Han Feng^a, Shijie Zhang^a, Jennifer Man-Fan Wan^b, Lifeng Gui^a, Mengcheng Ruan^a, Na Li^a, Hongyu Zhang^a, Zhiguo Liu^a✉, Hualin Wang^a✉

研究概要

桑黄多糖治疗糖尿病效果显著，但其机制尚不清楚。研究者基于高糖高脂饮食的小鼠模型，在BGISEQ平台完成转录组测序，通过层层深入的数据分析挖掘，构建出关键代谢通路及代谢网络，最终锁定一种肠道细菌，阐明了桑黄多糖增强胰岛素敏感性的机制。

研究策略

取材: 3组小鼠肝脏样本，每组3个重复；对照组: 常规饮食16周；HFD组: 高糖高脂饮食12周服用盐水4周；

PLP组: 高糖高脂饮食12周服用桑黄多糖4周；

测序: BGISEQ RNA-Seq；

分析: 基于转录组测序数据，应用基础GO、Pathway代谢通路富集分析，结合常见生化实验。

研究结果

1、RNA测序研究

处理3组小鼠肝脏样本，进行生化指标数据采集，其中HFD和PLP两组（每组3个重复样本）基于BGISEQ平台完成6个样本的RNA-seq测序。

分析转录组数据，相比于HFD组，在PLP组中11110个基因表达量上调，221个基因表达量下调。通过KEGG代谢通路富集分析，发现差异表达基因主要富集在信号转导通路，而其中的FOXO通路、钙代谢通路等正好与研究者一直关注的性状相关（图1）。

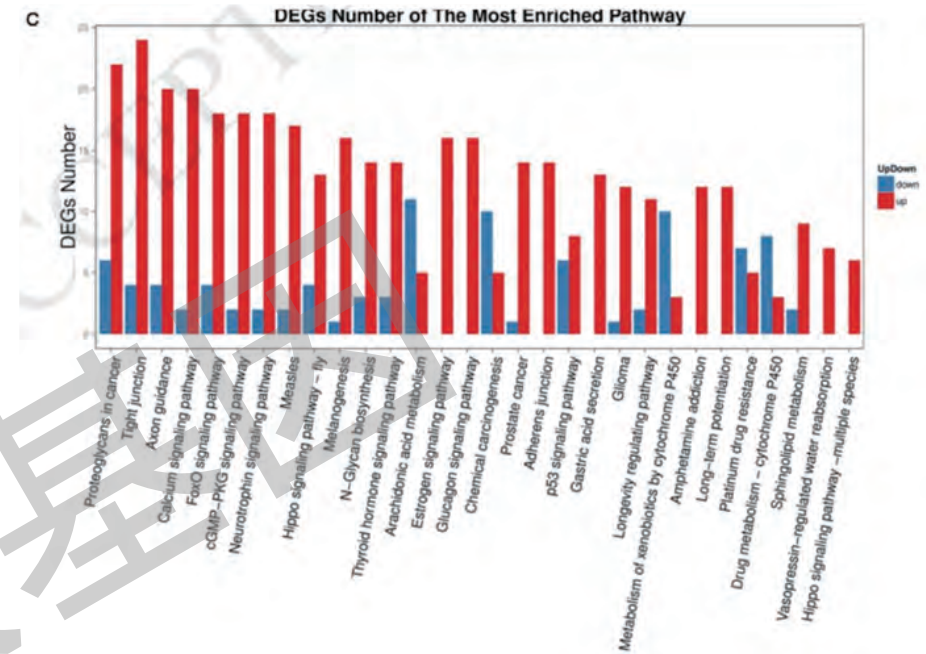


图1 差异表达基因KEGG分析，主要富集在FOXO通路、钙通路（图中红色方框）等

2、桑黄多糖作用于肝脏FOXO代谢通路

聚焦差异表达基因富集分析捕获的FOXO通路。FOXO蛋白作为胰岛素代谢通路上的重要分子，对调控机体血糖的平衡有重要作用。但该通路当中各基因上下调关系未知，研究者通过分析KEGG通路图，发现该通路上的关键基因总体上都是上调（图2），后续qPCR验证与测序结果一致（图3）。

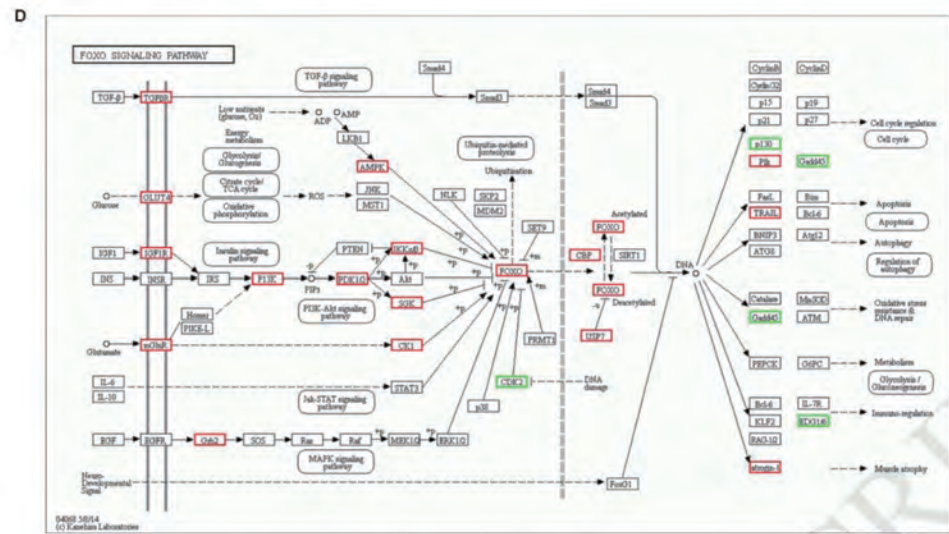


图2 小鼠肝脏转录组测序，高脂高糖+桑黄多糖组，和高糖高脂组，FOXO通路当中的基因总体都上调

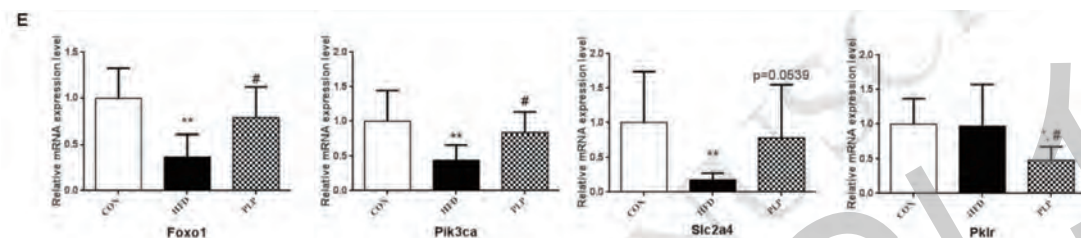


图3 qPCR验证实验与测序结果一致

3、钙信号通路

研究者还关注差异表达基因富集通路当中所包含的钙信号通路。肝脏细胞的钙平衡，受磷脂酰胆碱与磷脂酰乙醇胺的比例（PC/PE）调控研究。发现，PC/PE比例在高糖高脂组降低，服用桑黄多糖后恢复（图4）。

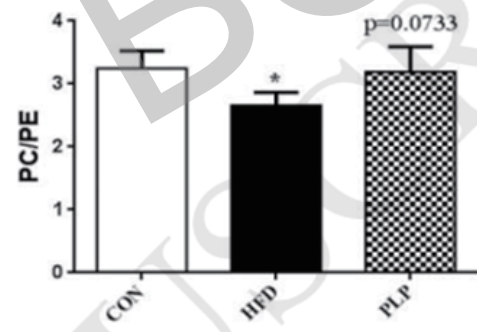


图4 高糖高脂饮食降低了PC/PE 比例，而桑黄多糖能使比例恢复正常水平

4、PC/PE比例的调控机制分析

研究者首先分析PEMT相关基因表达。qPCR实验显示，*Pemt*（表达PEMT的基因）在高糖高脂组中表达量高于常规饮食组，服用桑黄多糖后，*Pemt*表达量下降。

随后分析SAM基因表达量变化情况。回归转录组数据，发现有两个差异表达基因参与SAM合成：*Mat1a*和*Mat2b*。高糖高脂+桑黄多糖组*Mat1a*表达量上调，但*Mat2b*表达量下调。SAM的浓度在组间没有显著差异，但其反式甲基反应副产物s-腺苷高半胱氨酸（SAH）在高糖高脂组升高，在桑黄多糖组恢复，从SAM/SAH比例看，高糖高脂组比常规组低，桑黄多糖服用组与常规组持平。SAH经代谢转化为高半胱氨酸（HCY），但HCY在三组间并无显著差异（图5）。

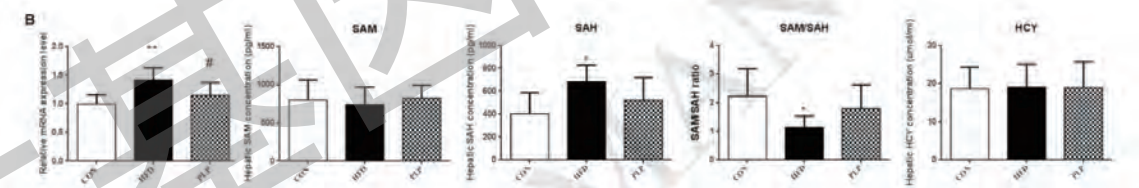


图5 SAH在高糖高脂组升高，服用桑黄多糖后恢复，从SAM/SAH比率看，高糖高脂组比常规组低，在桑黄多糖组与常规组持平。HCY在三组间并无显著差异。

5、桑黄多糖对VB12的影响机制

进一步挖掘转录组测序数据，发现桑黄多糖摄入改变了3个参与维生素B12运输和代谢相关基因的表达（从KEGG和GO代谢路径分析可得）；再查看维生素B12的表型数据，发现高糖高脂组血浆维生素B12水平相比于常规组显著降低，服用桑黄多糖后改善（图6）。

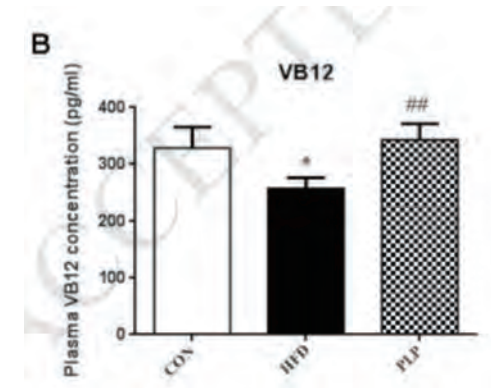


图6 高糖高脂组血浆维生素B12水平相比于常规组显著降低，服用桑黄多糖后改善

6、桑黄多糖对肠道菌群的影响机制

维生素B12参与HCY合成甲硫氨酸的代谢路径，是甲硫氨酸合成酶的辅因子。回归转录组数据，发现桑黄多糖处理提高了甲硫氨酸合成还原酶基因 (*Mtrr*) 表达，MTRR能恢复甲硫氨酸合成酶活性，促进氧化态维生素B12还原再生。

动物自身不能合成维生素B12，但肠道菌群可以。为了弄清桑黄多糖是否通过改变肠道菌群来影响维生素B12代谢，研究者采用PCR-DGGE实验，在高糖高脂饮食+桑黄多糖服用组，发现了一个特别的细菌——卟啉单胞菌丰度的改变。

这个菌是不是真的影响了维生素B12水平？钴胺酸a,c-二酰胺合酶 (*cbiA*) 是参与维生素B12合成的一个关键酶，检测编码该酶的DNA在肠道菌群中的相对水平，发现服用桑黄多糖后，该DNA含量更高，这促进了维生素B12血浆浓度的恢复 (图7)。

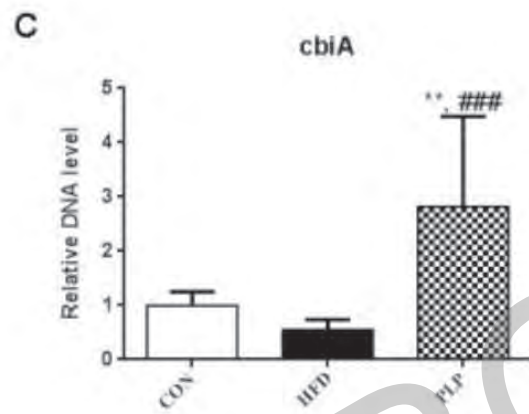


图7 服用桑黄多糖显著提升编码*cbiA*的DNA相对水平

研究结论

致此，两条代谢通路分析结果交汇到了一起，研究者据此推断，桑黄多糖的摄入影响了肠道菌群构成，特异性地促进了肠道细菌——卟啉单胞菌的增殖，促进了维生素B12合成，进而调节肝脏的甲硫氨酸代谢，使得磷脂酰胆碱与磷脂酰乙醇胺的比例恢复平衡。

拟南芥*AtACS8*基因在铜离子激活下，在乙烯早期生物合成中起关键作用

文章5

AtACS8 plays a critical role in the early biosynthesis of ethylene elicited by copper ions in Arabidopsis

© 2017. Published by The Company of Biologists Ltd.

AtACS8 Plays a Critical Role in the Early Biosynthesis of Ethylene Elicited by Copper Ions in Arabidopsis

发表期刊: *Journal of Cell Science*

影响因子: 4.431

研究对象: 拟南芥

合作单位: 山东农业大学

研究概要

铜离子是所有生物体所必需的，拟南芥中铜离子能引起防御反应，依赖于乙烯信号通路。但铜离子引起乙烯生物合成的机制仍不清楚。CuSO₄处理后可以快速提高乙烯合成。另外，它能上调一些防御相关基因的表达，如一些乙烯合成基因，其中*AtACS8*是铜离子诱导下，防御反应和早期乙烯合成中必不可少的。而且，铜离子诱导的*AtACS8*表达依赖于*AtACS8*启动子区的铜响应顺式元件 (CuRE)。铜离子特异性激活拟南芥*AtACS8*基因表达，来促进乙烯早期生物合成，进而引起依赖于CuRE调控元件的拟南芥防御反应。

研究策略

取 材: 两周大野生型拟南芥植株样品，CuSO₄处理后0、2、24h，2个生物学重复，共6个样品

研究技术: BGISEQ RNA-Seq、SE50

验证方法: qRT-PCR、功能验证等

研究结果

1、拟南芥中铜离子促进乙烯合成

处理24小时后，CuSO₄处理比MgSO₄合成了更多乙烯 (图1A)，而且在转录水平上乙烯响应因子1，*ERF1*基因 (乙烯合成信号通路的一个标记) 表达更多。说明铜离子可促进拟南芥乙烯合成和激活乙烯信号通路。

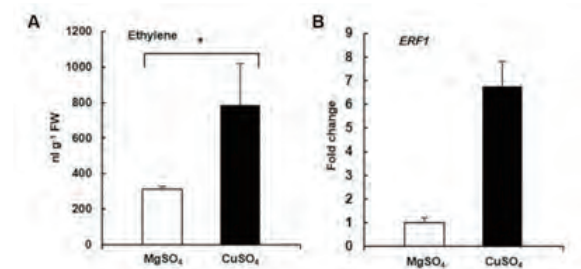


图1 铜离子提高拟南芥乙烯水平和上调*ERF1*表达

2、拟南芥CuSO₄处理后基因表达差异

两周大野生型拟南芥植株样品，CuSO₄处理后0、2、24h，2个生物学重复，共6个样品BGISEQ-500、RNA-Seq、SE50、24M clean reads、Clean Data Rate高达99.95%。

表1 测序数据统计

	Sample	Sequencing Strategy	Raw Data Size (bp)	Raw Reads Number	Clean Data Size (bp)	Clean Reads Number	Clean Data Rate (%)
Replicate 1	R1-0h	SE50	1206855350	24137107	1205184250	24103685	99.86
	R1-2h	SE50	1206843900	24136878	1206354400	24127088	99.95
	R1-24h	SE50	1206821350	24136427	1206413250	24128265	99.96
Replicate 2	R2-0h	SE50	1206874850	24137497	1206698800	24133976	99.98
	R2-2h	SE50	1206879900	24137598	1206647550	24132951	99.98
	R2-24h	SE50	1206868050	24137361	1206622350	24132447	99.97

CuSO₄处理后，2小时后2206个上调，1009个下调基因，24小时后2423个上调，590个下调基因（图2A）。快速（2小时）上调基因中，370个基因是被鞭毛蛋白和延伸因子上调（图2B）。qRT-PCR验证上调基因，发现所有挑选基因验证结果与测序相似（图2C）。GO分析发现79.30%的上调基因可以归类于代谢进程，细胞进程和刺激响应。

3、铜离子激活拟南芥乙烯生物合成通路

通过测序获得CuSO₄处理2小时后快速地受诱导的基因中，有一些是参与乙烯生物合成的（图3A），其中包含*AtACS8*等基因。qRT-PCR验证获得的这些乙烯合成基因，所有挑选的基因的验证结果与测序相似（图3B）说明铜离子诱导了拟南芥乙烯合成基因的表达。

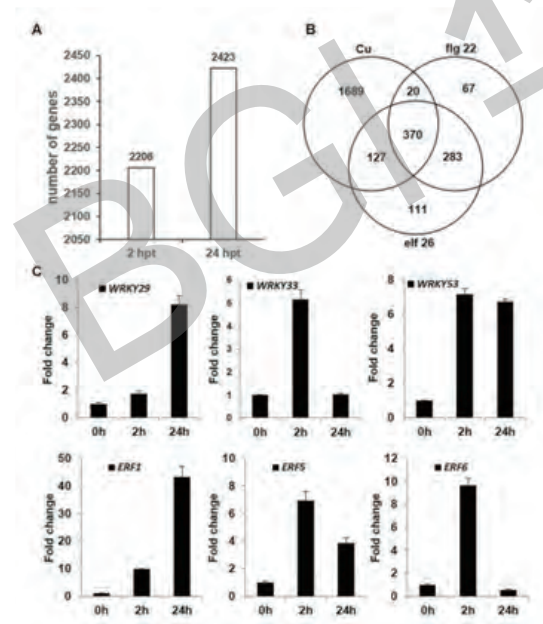


图2 铜离子调控下拟南芥基因表达情况

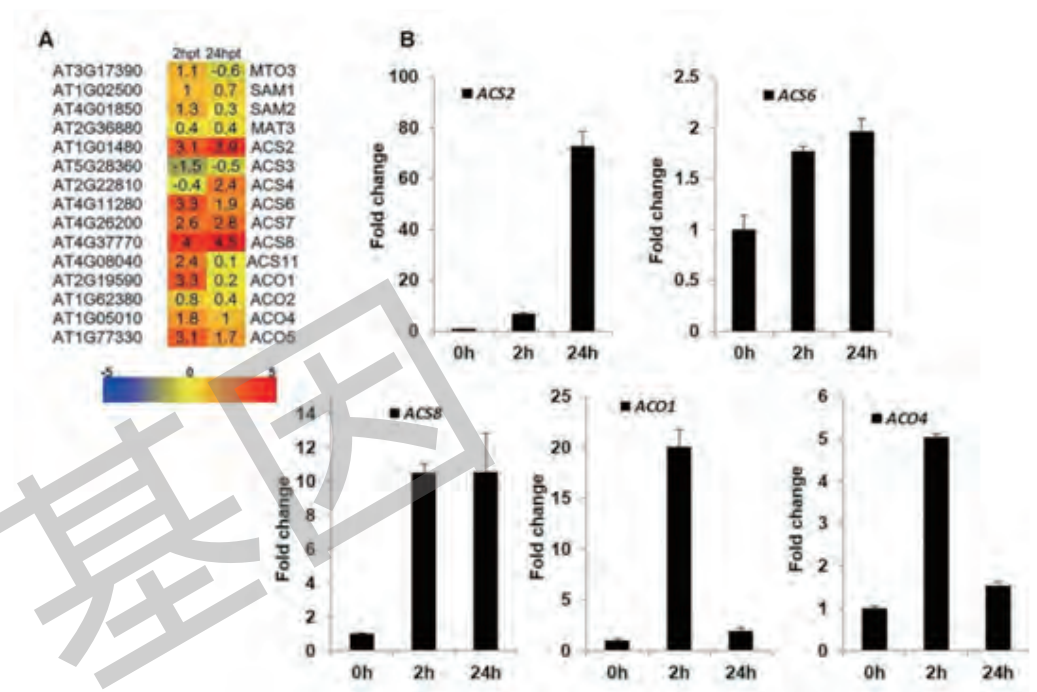


图3 CuSO₄处理后乙烯生物合成通路基因的表达情况

4、铜离子可以迅速地激活乙烯生物合成

分析不同时间点的乙烯产率，证明铜离子可以迅速地激活乙烯生物合成。

5、铜离子诱导的防御反应和乙烯合成依赖于*AtACS8*基因

通过测序和qRT-PCR验证，说明铜离子快速诱导*AtACS8*表达（图3）。检测野生型和敲出突变植株的MgSO₄或CuSO₄处理后乙烯水平和乙烯合成信号通路的标记基因表达水平，以及更多验证工作，证明*AtACS8*基因在铜离子诱导的乙烯早期生物合成中起关键作用。

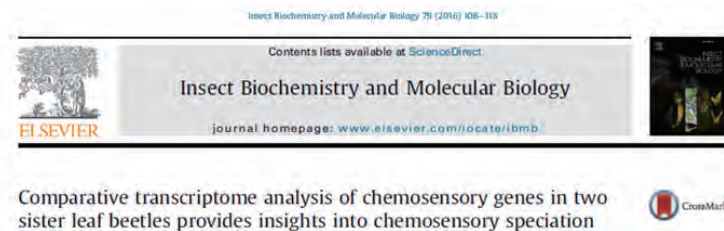
6、鉴定CuRE在铜离子诱导表达*AtACS8*中的作用

通过构建突变体、定点敲除等方法，证明铜离子诱导的*AtACS8*表达，是依赖于*AtACS8*启动子区的铜响应顺式元件（CuRE）。

文章6

两个叶甲姐妹种化学感应基因的比较转录组学分析, 揭示化学感应物种形成

Comparative transcriptome analysis of chemosensory genes in two sister leaf beetles provides insights into chemosensory speciation



发表期刊: *Insect Biochemistry and Molecular Biology*

影响因子: 3.562

研究对象: 叶甲

合作单位: 中国科学院动物研究所等

研究背景

化学感受对于昆虫非常重要, 昆虫利用自己的化学感受器识别环境中的化学信号, 并产生相应的行为反应。因此昆虫近缘种间化学感应基因的分化产生行为分化, 比如食物、产卵地点和交配的偏好, 行为分化进而产生交配前生殖隔离, 此即为化学感应物种形成。要查明化学感应基因在物种形成中的作用, 一个方法就是在近缘种之间比较这些基因, 来发现分歧进化的遗传特征。本文研究的是一种植食性的昆虫——叶甲。挑选了叶甲两个姐妹种 *P. aenescens* 和 *P. maculicollis*。这两个物种具有生殖隔离和宿主植物的偏好, 他们都以同样的宿主植物为食, 但只食用特定株龄。

研究策略

取材: 两个种成虫混样做转录组 *de novo*; RNA-Seq 定量, 4个 *P. aenescens* 和 3个 *P. maculicollis* 样品做生物重复

测序: HiSeq 2500, PE100, 转录组 *de novo*; BGISEQ, RNA-Seq 定量, SE50

分析: 转录组 *de novo* 组装和功能注释; 构建ML树, 鉴定化学感应基因直系同源对; 差异表达基因分析

验证: qRT-PCR、RNA-Seq

研究概要

叶甲姐妹种 *Paenescens* 吃成树叶子, *Pmaculicollis* 吃籽苗叶子, 做转录组 *de novo*+BGISEQ RNA-Seq, 构建进化树, 筛选正选择基因和差异表达基因, 从而寻找化学感应物种形成的关键基因。使用 BGISEQ RNA-Seq, 研究叶甲姐妹种中化学感应直系同源基因表达水平, 随机挑了 12 个, 用 qRT-PCR 验证 RNA-Seq 结果。

表1 显著性差异表达的化学感应物种形成的候选直系同源基因

Orthologs	Different expression ^a	
	log ₂ FC	P _{adj}
OBP1	-4.4144	5.36E-48
OBP10	-0.3831	0.3772
OBP12	2.6033	6.31E-14
OBP25	2.8873	0.0026
CSP1	-3.7225	1.38E-17
OR12	-1.9408	6.32E-10
OR14	-1.7672	0.0101
OR15	-3.3870	7.58E-35
GR4	2.3611	7.77E-10

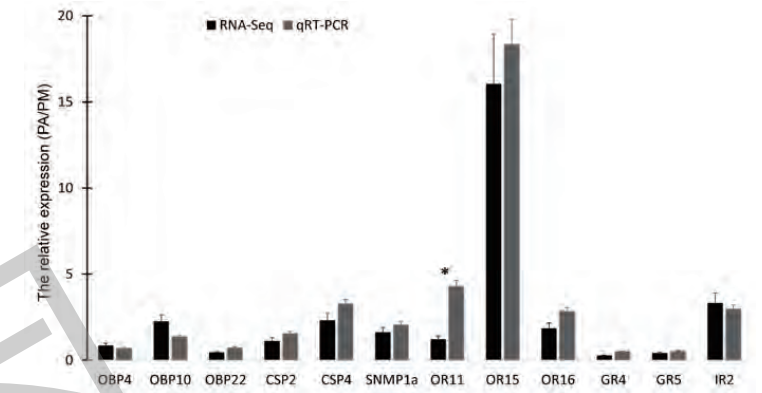


图1 随机挑选12个直系同源基因, 两个种基因表达量倍数比值 (*P. aenescens*/*P. maculicollis*) 的RNA-Seq和qRT-PCR比较

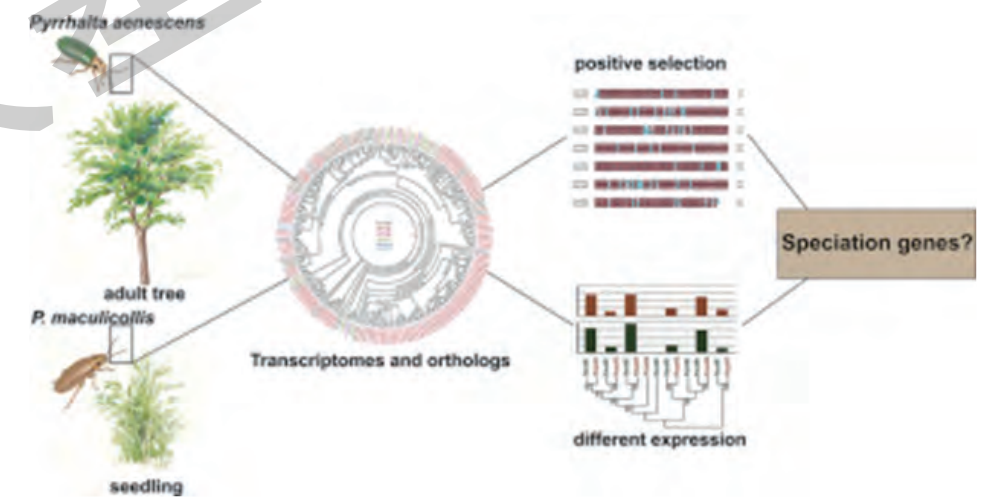


图2 研究思路

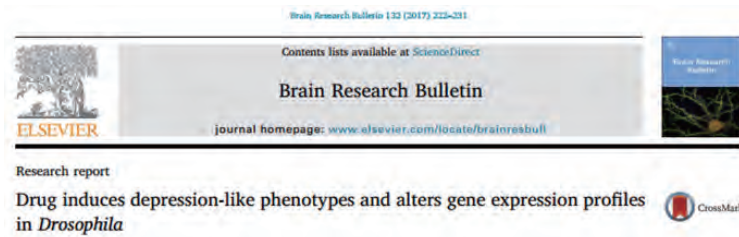
研究结果

- 1、功能注释发现两个姐妹种的转录组高度相似, 整体上是低分歧的, 这与预期相符。
- 2、注释化学感应基因, 并进行进化分析, 构建maximum likelihood (ML) 树, 鉴定两个种的化学感应基因直系同源对及了解和其他昆虫的关系。
- 3、发现一些正选择下高表达和差异表达的化学感应基因在化学引导的生殖隔离和生态隔离中可能起了关键作用。

文章7

药物引起果蝇类似抑郁症的表型和基因表达谱改变

Drug induces depression-like phenotypes and alters gene expression profiles in Drosophila



发表期刊: *Brain Research Bulletin*

影响因子: 3.44

研究对象: 黑腹果蝇

合作单位: 华中师范大学

研究概要

重度抑郁症 (MDD) 是一种严重的精神疾病, 分子机制目前还不清楚。该研究选取与人类在药物作用下有相似的类似抑郁症状表现的黑腹果蝇为动物模型。发现喂养黑腹果蝇左旋多巴或者氯丙嗪能引起类似抑郁症表型。通过RNA-Seq, 发现氯丙嗪处理 (CPZ) 组与对照组雄性果蝇相比, 有467个基因差异表达, 其中312个基因上调表达, 155个基因下调表达。这些基因参与的代谢通路有帕金森病、亨廷顿病、阿尔茨海默病和溶解酵素通路。qRT-PCR确定了19个CPZ组与对照组雄性果蝇中差异表达的基因。

研究策略

取材: 6个雄性果蝇RNA-Seq, 包含2个组, 即3个生物学重复。3个对照组样品, 3个CPZ组样品, 每个样品包含30个果蝇

测序: BGISEQ RNA-Seq

分析: GO分析、代谢通路分析

验证: qRT-PCR

研究结果

1、表型

通过抑郁症类似行为、生物标记和遗传毒性等研究, 发现氯丙嗪2000 mg/L, 48小时处理引起雄性果蝇更加明显的抑郁症类似的特征, 包括显著性降低的食物摄入量、交配频率、血清素 (5-HT) 浓度和增加的丙二醛 (MDA) 浓度以及减少的超氧化物歧化酶 (SOD) 活性, 雄性比雌性对药物更敏感。氯丙嗪处理的果蝇, 其后代也有这些类似抑郁症的特征。

2、RNA-Seq

(1) 为探究抑郁症分子机制, 测序6个雄性果蝇RNA-Seq, 包含2个组, 即3个生物学重复。3个对照组样品, 3个CPZ组样品, 每个样品包含30个果蝇。

(2) 平均24,137,120 raw reads, 过滤后24,100,281 clean reads, clean reads比率高达99.85%。

(3) 每个样品鉴定到的基因数量。数据库中一共有基因13907个, 每个样品鉴定的基因有12,702-12,898个, 鉴定出的基因数占数据库基因总数比例很高, 91.34%-92.74% (图1)。

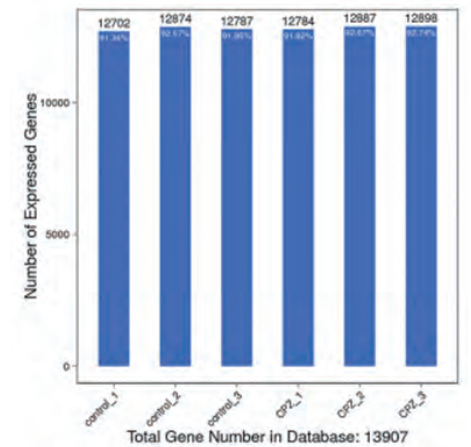


图1 鉴定到的基因数量

(4) CPZ与对照组的差异表达基因。一共312个显著上调基因和155个显著下调基因 (图2)。

(5) GO分析。GO分为分子功能 (MF)、细胞组分 (CC) 和生物过程 (BP) 三大功能类, 分别分析了三类参与基因最多的功能 (图3)。

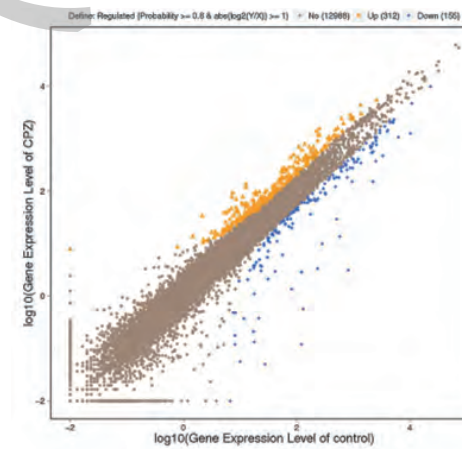


图2 所有表达基因Scatter-plot分布图

(蓝色代表下调基因、橘色代表上调基因、棕色代表非差异表达基因)

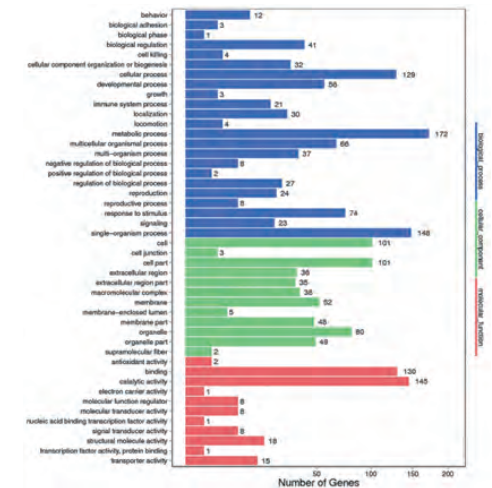


图3 差异基因GO功能分类

(6) 代谢通路分析。在Top20差异基因KEGG富集通路中, 鉴定出一些神经性的疾病相关通路, 包括帕金森病、亨廷顿病、阿尔茨海默病 (图4)。一些其他人类疾病相关通路也展现在图中, 如百日咳、胰液分泌、非酒精性脂肪肝病、刺激神经组织的配体-受体互作、溶解酵素、雌激素信号通路、心肌收缩和抗原加工提呈。

3、qRT-PCR验证

挑选19个基因进行qRT-PCR，挑选原则是差异大的差异表达基因，以及与神经病相关的基因（图5）。19个基因qRT-PCR结果全部都与RNA-Seq相似。

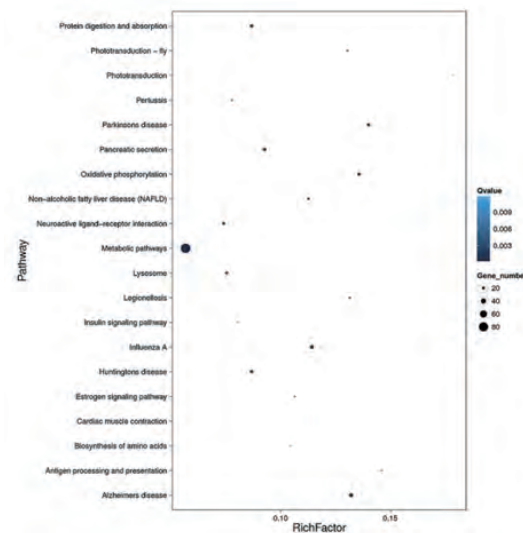


图4 Top20差异基因KEGG富集结果

研究意义

药物引起黑腹果蝇类似抑郁症的表型。通过RNA-Seq发现467个差异表达基因，涉及到新陈代谢、人类疾病和溶解酵素通路，提供解释人类抑郁症分子机制的补充观点，可能有助于临床治疗。

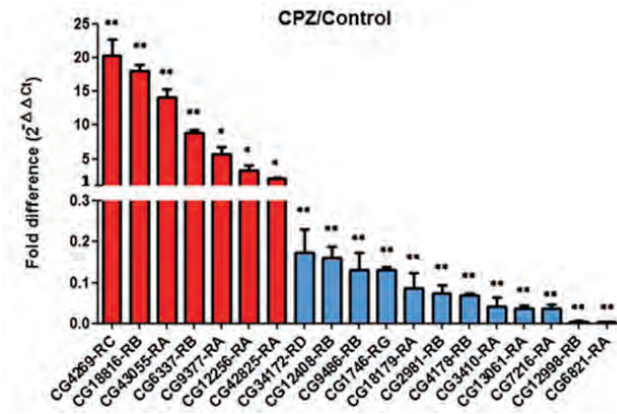


图5 qRT-PCR验证RNA-Seq差异表达基因

文章8

过表达*OsHSP18.0-CI*提高水稻对细菌性条斑病的抗性

Overexpression of OsHSP18.0-CI Enhances Resistance to Bacterial Leaf Streak in Rice

Ju et al. *Rice* (2017) 10:12
DOI 10.1186/s12284-017-0153-6

Rice

ORIGINAL ARTICLE

Open Access

Overexpression of *OsHSP18.0-CI* Enhances Resistance to Bacterial Leaf Streak in Rice

Yanhui Ju¹, Hongjuan Tian^{1,3}, Ruihua Zhang², Liping Zuo¹, Guixiu Jin², Qian Xu¹, Xinhua Ding^{1,4}, Xiangkui Li² and Zhaohui Chu^{1,5*}

研究背景

小热激蛋白是一个蛋白大家族，响应各种非生物与生物胁迫。水稻中的小热激蛋白基因*OsHSP18.0-CI*有报道发现响应盐和镉等非生物胁迫，但对病原菌的直接功能并不清楚。

研究策略

取材：野生型材料，水稻品种Shengdao 806，细菌性条斑病敏感品种；过表达和抑制表达转基因材料，从806叶片DNA中克隆*OsHSP18.0-CI*的片段，构建载体转入到806中；接种条斑病菌RS105

研究方法：克隆基因、构建载体、转基因、表型观察、PCR、qRT-PCR、RNA-Seq等

测序：BGISEQ RNA-Seq, SE50

分析：从整体和重点基因两个角度，进行差异表达基因与GO富集、KEGG和转录因子等

验证：qRT-PCR

研究概要

发现*OsHSP18.0-CI*，能被水稻病原细菌性条斑病菌诱导表达，超量表达能显著增强水稻对条斑病菌RS105的抗性，抑制表达则能显著提高敏感性。

为探究该基因提高抗性的作用机理，对接种条斑病菌RS105前后的野生型材料WT和过表达的转基因材料OE进行了RNA-Seq，对所有表达基因进行分析。

差异基因的GO分析结果显示，OE接种后的差异表达基因与WT表现出类似的GO分类结果，但OE中富集到每个Go term的基因数量都明显高于WT。OE和WT的共同差异基因主要富集在与免疫反应相关的PR基因、水杨酸Sa、茉莉酸JA合成相关基因等，并且这些基因在OE中表现出更显著的差异表达。

这些结果说明*OsHSP18.0-CI*是通过增强水稻的基础防卫反应，提高水稻对条斑病的抗性的。

研究思路

构建好转基因材料，在当代和后代中通过表型观察和qRT-PCR方法确定了*OsHSP18.0-C*抗病作用。但是深入的作用机制还不明确，该基因是否也改变了其他基因表达水平。因而通过RNA-Seq对整体上所有表达基因进行分析，并验证。最后通过表型观察来评估过表达对生长发育影响。

研究结果

1、*OsHSP18.0-CI*与水稻病原细菌性条斑病抗性关系

*OsHSP18.0-CI*能被水稻病原细菌性条斑病菌诱导表达。过表达提高水稻对条斑病菌RS105的抗性，抑制表达则提高敏感性。而且对多个条斑病菌株如HGA4、HGA2等都具有抗性。

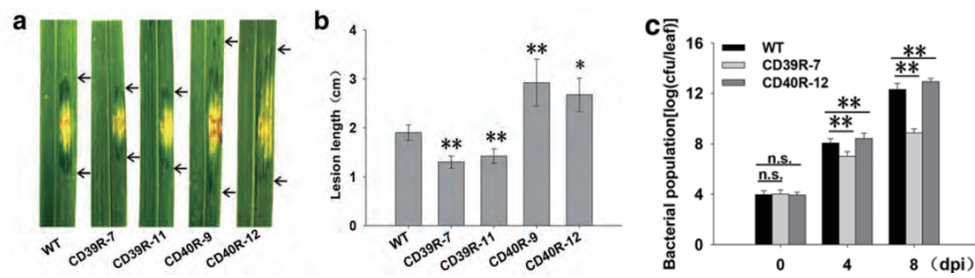


图1 转基因植株后代的抗性情况

a和b叶片损害长度的观察图和测量图，c叶片中细菌含量；野生型WT，过量表达*OsHSP18.0-CI*的转基因材料CD39R-7和CD39R-11，抑制表达转基因材料CD40R-9和CD40R-12

2、BGISEQ RNA-Seq鉴定过表达转基因材料和野生型材料的差异表达基因

对接种条斑病菌RS105前后的野生型材料WT和过表达*OsHSP18.0-CI*的转基因材料OE进行了RNA-Seq分析。4个样品，每个样品平均24Mreads，基因组比对率87%，28,872个基因，clean/raw reads比率极高99.98%。

表1 RNA-Seq数据统计

Sample	Raw Data Size (bp)	Raw Reads Number	Clean Data Size (bp)	Clean Reads Number	Total Mapped Reads (%)	Unique Match(%)	Number of expressed genes
WT	1,206,861,900	24,137,238	1,206,636,750	24,132,735	87.55	80.32	28,037
OE	1,206,839,800	24,136,796	1,206,614,350	24,132,287	87.00	79.59	28,771
WT-24	1,206,846,900	24,136,938	1,206,567,200	24,131,344	86.85	79.97	28,978
OE-24	1,206,841,800	24,136,836	1,206,272,600	24,125,452	86.86	79.46	29,701

在病菌RS105刺激下，野生型材料和过表达材料分别产生了898个和1844个上调表达基因，738个共同差异基因（a）。这738个基因GO分析发现，差异表达基因在34个不同功能类别中的分布是相似的。

同样，在病菌接种后，野生型材料和过表达材料分别产生了128和154个下调表达基因，29个共同差异基因，在野生型和过表达材料中都被抑制表达（b）。

相比于野生型，738个共同上调表达基因中，RS105诱导694个基因在过表达材料中表达更高（c）。29个共同下调表达基因中，16个基因在过表达材料中被RS105抑制更强烈（c）。

综上所述，大多数共同的差异表达基因，在接种后过表达材料中变化更强烈。说明*OsHSP18.0-CI*是通过增强水稻的基础防卫反应提高水稻对条斑病的抗性。

3、BGISEQ RNA-Seq，水稻细菌性条斑病菌抗性有关基因的差异表达分析

发病机理相关基因的诱导表达和水杨酸茉莉酸含量提高，对提高水稻细菌性条斑病菌抗性十分重要。在738个和29个共同的上调和下调基因中，鉴定出28个基因，包含13个发病机理相关基因PR和8个水杨酸Sa生物合成、7个茉莉酸JA生物合成有关基因（a）。这些基因都是在接种后过表达材料，表达水平更高（a）。与基因表达一致，水杨酸和茉莉酸的含量在过表达材料中也提高了（c和d）。此外，也发现一些乙烯或植物生长激素信号通路差异表达基因。

10个RNA-Seq上调表达基因全部通过qRT-PCR验证。10个基因在4个样品中基因表达的倍数变化趋势一致，即接种病菌后刺激了过表达和野生型表达量上调，而且过表达材料比野生型表达高（b）。

4、过表达*OsHSP18.0-CI*基因对植株生长影响

结果发现负面影响极小。

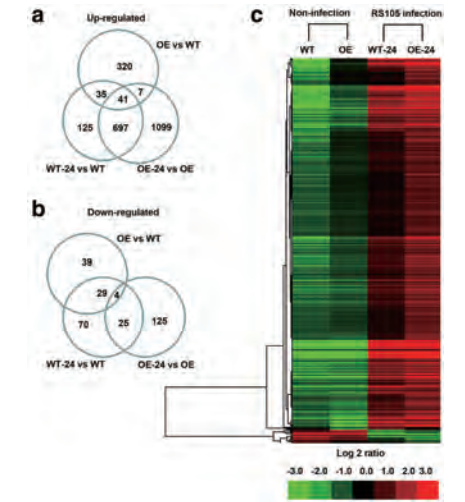


图2 RNA-Seq研究基因表达谱

a 过表达vs野生型，野生型接种vs未接种，过表达接种vs未接种的上调基因统计，b下调统计，c差异表达基因层次聚类

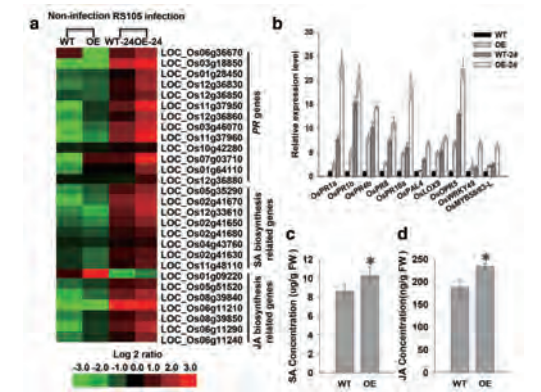


图3 发病机理相关基因和水杨酸茉莉酸生物合成相关基因的表达

a RNA-Seq层次聚类分析，b qRT-PCR验证RNA-Seq，c 水杨酸含量，d 茉莉酸含量

文章9

转录组水平分析发现酿酒酵母 *FKS1* 突变体中MAPK通路和碳代谢在细胞壁维持中的作用

RNA-Seq analysis of global transcriptomic changes suggests a roles for the MAPK pathway and carbon metabolism in cell wall maintenance in a *Saccharomyces cerevisiae FKS1* mutant



发表期刊: *Biochemical and Biophysical Research Communications*

影响因子: 2.559

研究对象: 酿酒酵母

合作单位: 华南理工大学等

RNA-Seq analysis of global transcriptomic changes suggests a roles for the MAPK pathway and carbon metabolism in cell wall maintenance in a *Saccharomyces cerevisiae FKS1* mutant

Cong Huang, Fengguang Zhao, Ying Lin, Suiping Zheng, Shuli Liang, Shuangyan Han*

研究概要

细胞壁有维持细胞自身形态构造、抵抗外界机械压力和维持渗透压等作用。很难想象有细胞壁缺陷的酵母细胞如何生存。*FKS1*编码的 β -1,3-葡聚糖合成酶(GS),负责合成酿酒酵母细胞壁的主要组分 β -1,3-葡聚糖。如果*FKS1*遭到破坏,会导致合成酶活性显著下降,抑制 β -1,3-葡聚糖的合成,细胞壁出现 β -1,3-葡聚糖缺陷。所以*FKS1*突变型,是细胞壁缺陷型。

研究者通过转录组测序展示了*FKS1*突变型中转录组全貌变化,发现了*FKS1*突变引起的细胞壁表型变化与响应变化的分子机制之间的相互关系,详细描述了MAPK通路和碳代谢通路如何响应*FKS1*突变,加强细胞壁重塑和细胞壁组分的合成,所以*FKS1*突变型就是靠这样调节生存下来。

研究策略

取材: 酿酒酵母*FKS1*突变型构建,分2组突变型和对照,3个生物学重复,共6个样品

测序: BGISEQ真核转录组测序, PE100

验证: qRT-PCR

表型: 离子交换色谱分析法定量细胞壁的多糖,3个生物学重复

研究结果

1、转录组测序

通过转录组测序,发现*FKS1*突变型相比于对照,有1151个差异表达基因,显著性地富集到MAPK通路和碳代谢的7个通路。利用qRT-PCR验证转录组测序结果。

2、MAPK通路

MAPK通路为*FKS1*突变型生存而被激活。*FKS1*突变导致产生的各种信号通过MAPK通路传递,加强那些维持细胞壁完整的代谢过程。在酿酒酵母中MAPK通路能响应环境信号和各种各样的压力,比如细胞壁压力等(图1),在突变体中这些通路被显著性富集到,其中参与细胞系压力信号传递的许多正调控的基因上调表达,负调控的基因则下调表达(图1中cell wall stress通路)。

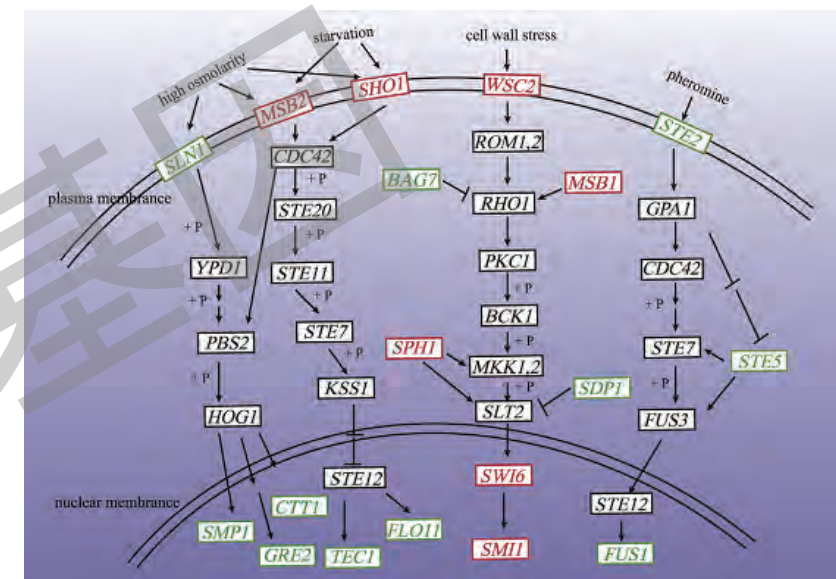


图1 *FKS1*突变引起的MAPK通路表达变化(红色上调,绿色下调,+P代表磷酸化过程)

3、碳代谢通路

分析*FKS1*突变体中细胞壁组分的碳代谢通路变化,发现细胞壁组分合成有关的葡聚糖、甘露聚糖和几丁质的合成被增强。与重构细胞壁有关的 β -1,3-葡聚糖合成的碳代谢过程如图2A所示。在*FKS1*突变型细胞中,将UDP-葡萄糖转化其他糖的过程减少到最低(图2中第一三四条通路),从而增强了UDP-葡萄糖转化成 β -1,3-葡聚糖合成(图2第二条通路),并改变细胞壁结构补偿 β -1,3-葡聚糖含量减少。同样的,与细胞壁合成有关的甘露聚糖合成、细胞壁组分几丁质合成被增强,合成有关的碳代谢(图2B和C)。

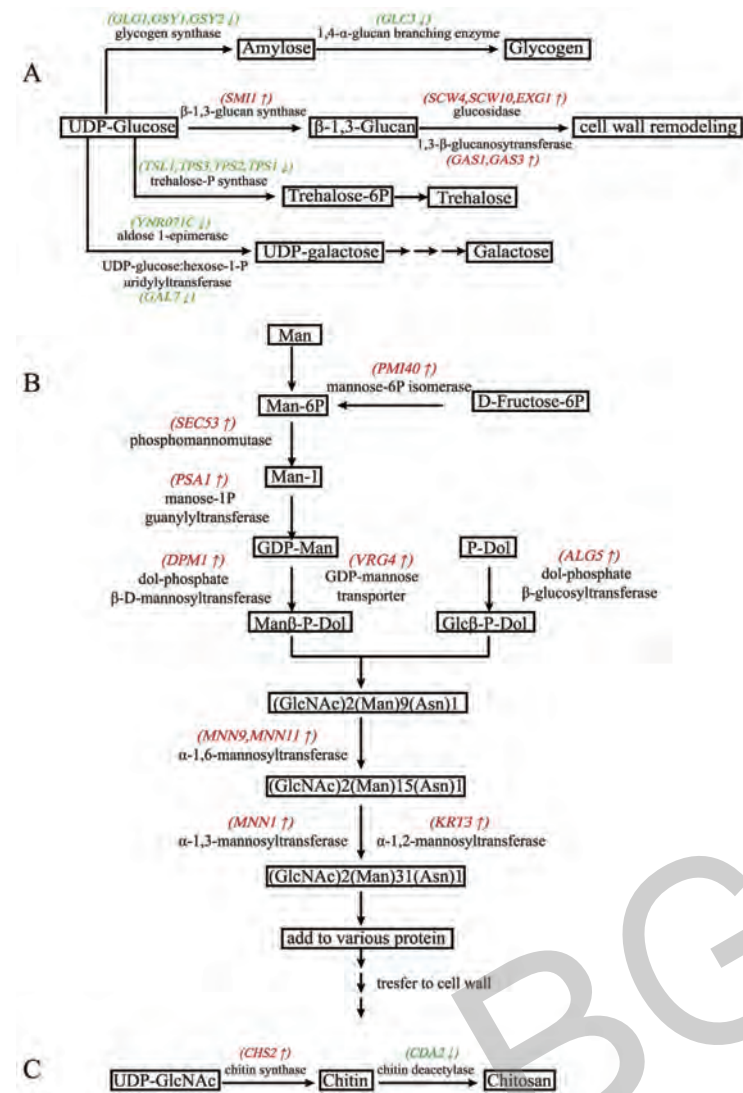


图2 *FKS1*突变型中细胞壁组分合成通路和参与这些通路的基因表达变化，
A β -1,3-葡聚糖合成，B 甘露聚糖合成，C 几丁质合成（红色上调，绿色下调）

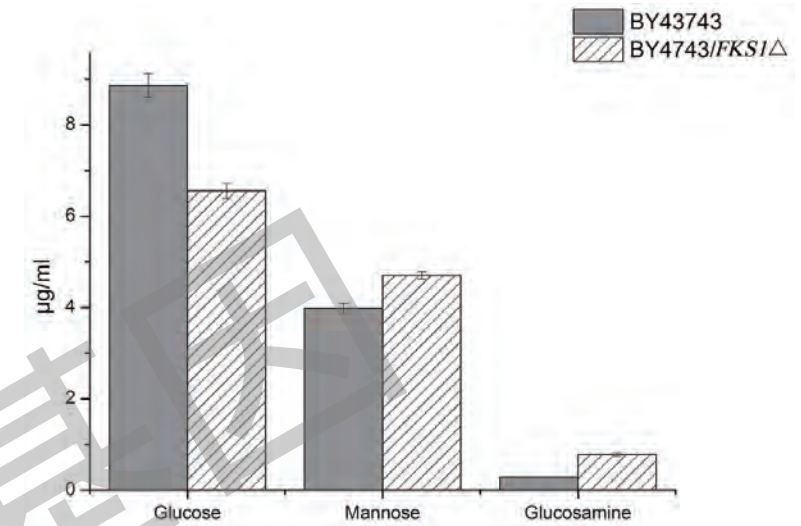


图3 *FKS1*突变型和对照细胞的细胞壁组分水解后的葡萄糖、甘露糖和葡萄糖胺的定量分析（灰色为对照，斜线为突变型）

4、检测细胞壁表型变化，与转录组测序相互印证

提取细胞壁，葡聚糖、甘露聚糖和几丁质水解产物，葡萄糖、甘露糖和葡萄糖胺。*FKS1*突变型和对照细胞相比，甘露糖和葡萄糖胺分别提高18.09%和169.72%，而葡萄糖却降低25.99%（图3）。原因是，尽管如上所述中有一种补偿机制增强葡萄糖相关酶的表达，但突变型有 β -1,3-葡聚糖缺陷，同时突变型又通过增强甘露聚糖和几丁质水平来补偿葡聚糖的缺陷。

附表：真核转录组&RNA-Seq文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	An Interleukin-25-Mediated Autoregulatory Circuit in Keratinocytes Plays a Pivotal Role in Psoriatic Skin Inflammation	<i>Immunity</i>	22.845	人	清华大学
2	Repopulated microglia are solely derived from the proliferation of residual microglia after acute depletion	<i>Nature Neuroscience</i>	17.839	小鼠	中科院深圳先进技术研究院等
3	KRAB-type zinc-finger proteins PITA and PISA specifically regulate p53-dependent glycolysis and mitochondrial respiration	<i>Cell research</i>	15.606	小鼠	北京生命组学研究所
4	Menin Deficiency Leads to Depressive-like Behaviors in Mice by Modulating Astrocyte-Mediated Neuroinflammation	<i>Nature</i>	14.318	小鼠	厦门大学
5	Glucocorticoid receptor in stromal cells is essential for glucocorticoid-mediated suppression of inflammation in arthritis	<i>Annals of the Rheumatic Diseases</i>	12.380	小鼠	乌尔姆大学
6	STAG2 deficiency induces interferon responses via cGAS-STING pathway and restricts virus infection	<i>Nature Communications</i>	12.124	人	斯坦福大学
7	Mutually exclusive acetylation and ubiquitylation of the splicing factor SRSF5 control tumor growth	<i>Nature Communications</i>	12.124	人	北京生命组学研究所
8	Roles of the CSE1L-mediated nuclear import pathway in epigenetic silencing	<i>PNAS</i>	9.661	人	北京协和医学院和中国医学科学院
9	Nuclear carbonic anhydrase 6B associates with PRMT5 to epigenetically promote IL-12 expression in innate response	<i>PANS</i>	9.661	小鼠	中国医学科学院、中国工程院院士曹雪涛研究团队

10	Phytochrome A Negatively Regulates the Shade Avoidance Response by Increasing Auxin/Indole Acetic Acid Protein Stability	<i>Developmental Cell</i>	9.174	拟南芥	复旦大学生命科学学院遗传工程国家重点实验室和植物生物学研究所
11	PAF-Myc-Controlled Cell Stemness Is Required for Intestinal Regeneration and Tumorigenesis	<i>Developmental Cell</i>	9.174	小鼠	德克萨斯大学
12	Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy	<i>The American Journal of Human Genetics</i>	9.025	人	纽卡斯尔大学
13	Overcoming erlotinib resistance in EGFR mutation-positive lung adenocarcinomas through repression of phosphoglycerate dehydrogenase	<i>Theranostics</i>	8.766	人	上海交通大学医学院、上海大学转化医学协作创新中心
14	The LIS1/NDE1 Complex Is Essential for FGF Signaling by Regulating FGF Receptor Intracellular	<i>Cell Reports</i>	8.282	小鼠	浙江大学
15	Leukemia cell-derived microvesicles induce T cell exhaustion via miRNA delivery	<i>Oncolmmunology</i>	7.719	人	华中科技大学
16	TBX3 promotes proliferation of papillary thyroid carcinoma cells through facilitating PRC2-mediated p57KIP2 repression	<i>Oncogene</i>	7.519	人	天津医科大学
17	A novel HD-ZIP IV/MIXTA complex promotes glandular trichome initiation and cuticle development in <i>Artemisia annua</i>	<i>New Phytologist</i>	7.330	青蒿	上海交通大学
18	Draft genome sequence of the Tibetan medicinal herb, <i>Rhodiola crenulata</i>	<i>Giga Science</i>	7.267	大花红景天	东南大学、华大基因
19	PBRM1 restricts the basal activity of innate immune system through repressing RIG-I-like receptor signaling and is a potential prognostic biomarker for colon cancer	<i>The Journal of Pathology</i>	6.894	人	深圳大学、四川大学华西第二医院

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
20	An integrated biomanufacturing platform for the large-scale expansion and neuronal differentiation of human pluripotent stem cell-derived neural progenitor cells	<i>Acta Biomaterialia</i>	6.319	人	亚利桑那州立大学
21	The epigenetic modifier PBRM1 restricts the basal activity of the innate immune system by repressing retinoic acid-inducible gene-1-like receptor	<i>Journal of Pathology</i>	6.253	人	深圳大学
22	The lncRNA MACC1-AS1 promotes gastric cancer cell metabolic plasticity via AMPK/Lin28 mediated mRNA stability of MACC1	<i>Molecular Cancer</i>	6.204	人	南方医科大学
23	Effects of In Utero PFOS Exposure on Transcriptome, Lipidome and Function of Mouse Testis	<i>Environmental Science & Technology</i>	6.198	小鼠	香港大学、香港中文大学
24	Preinduction with bone morphogenetic protein-2 enhances cardiomyogenic differentiation of c-kit+ mesenchymal stem cells and repair of infarcted myocardium	<i>International Journal of Cardiology</i>	6.189	大鼠	复旦大学上海医学院
25	FBXL10 contributes to the development of diffuse large B-cell lymphoma by epigenetically enhancing ERK1/2 signaling pathway	<i>Cell death & disease</i>	5.965	人	天津医科大学
26	Autophagy mediates glucose starvation-induced glioblastoma cell quiescence and chemoresistance through coordinating cell metabolism, cell cycle, and survival	<i>Cell death & disease</i>	5.965	人	同济大学医学院
27	In Situ Lignin Bioconversion Promotes Complete Carbohydrate Conversion of Rice Straw by <i>Cupriavidus basilensis</i> B-8	<i>ACS Sustainable Chemistry & Engineering</i>	5.950	细菌	中南大学
28	Genetic screening and multipotency in rhesus monkey haploid neural progenitor cells	<i>Development</i>	5.843	恒河猴	南开大学

29	The GARP/MYB-related grape transcription factor AQUILO improves cold tolerance and promotes the accumulation of raffinose family oligosaccharides	<i>Journal of Experimental Botany</i>	5.830	拟南芥	中国科学院植物学研究所
30	Selenite and methylseleninic acid epigenetically affects distinct gene sets in myeloid leukemia: A genome wide epigenetic analysis	<i>Free Radical Biology and Medicine</i>	5.606	人	卡罗林斯卡学院
31	Cambogin suppresses dextran sulphate sodium - induced colitis by enhancing Treg cell stability and function	<i>British journal of pharmacology</i>	5.490	小鼠	上海中医药大学
32	Dual inhibition of DNMTs and EZH2 can overcome both intrinsic and acquired resistance of myeloma cells to IMiDs in a cereblon - independent manner	<i>Molecular Oncology</i>	5.314	人	哥本哈根大学医院
33	Direct inhibition of ACTN4 by ellagic acid limits breast cancer metastasis via regulation of β -catenin stabilization in cancer stem cells	<i>Journal of Experimental & Clinical Cancer Research</i>	5.189	人	中山大学、广东省中医院等
34	Establishment of a normal-derived estrogen receptor-positive cell line comparable to the prevailing human breast cancer subtype	<i>Oncotarget</i>	5.168	人	哥本哈根大学

Small RNA测序

Small RNA 是生物体内一类重要的功能分子，主要包括 miRNA、siRNA 和 piRNA。它们通过各种序列特异性的基因沉默作用，包括 RNA 干扰 (RNAi)、翻译抑制、异染色质形成等，诱导基因沉默，调控诸如细胞生长发育、应激反应、沉默转座子等各种各样的细胞生理过程。

Small RNA 测序技术则是借助第二代高通量测序技术，对某物种某组织在特定状态下的 18-30nt (或 18-40 nt) 的 small RNA 进行高通量测序。继而通过数据库比对，对获得的 small RNA 序列进行分析、鉴定，将数百万条 small RNA 序列分类成 rRNA、tRNA、snRNA、snoRNA 和 miRNA、siRNA、piRNA (部分物种)，进行定量、差异分析等；另外，还进行预测新的 miRNA、siRNA、piRNA，并对鉴定出的相应已知 miRNA 和新 miRNA 进行靶基因预测。通过对 miRNA 靶基因的相关分析，如 GO 分析和 pathway 分析，可将鉴定出的 miRNA 与相应表型差异关联起来，从而为解释相关生理过程的调控机制提供证据，如开花时间，花的颜色，种子发育，植株抗病，病理发生调控机制等。

文章1

基于cPAS的BGISEQ-500平台非编码小RNA测序研究

cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs

Clinical Epigenetics

Home About Articles Submission Guidelines

Research Open Access

cPAS-based sequencing on the BGISEQ-500 to explore small non-coding RNAs

Tobias Fehlmann, Stefanie Reinheimer, Chunyu Geng, Xiaoshan Su, Snezana Drmanac, Andrei Alexeev, Chunyan Zhang, Christina Backes, Nicole Ludwig, Martin Hart, Dan An, Zhenzhen Zhu, Chongjun Xu, Ao Chen, Ming Ni, Jian Liu, Yuxiang Li, Matthew Poulter, Yongping Li, Cord Stähler, Radoje Drmanac, Xun Xu, Eckart Meese and Andreas Keller

Clinical Epigenetics. The official journal of the Clinical Epigenetics Society 2016 8:123
<https://doi.org/10.1186/s13148-016-0287-1> | © The Author(s). 2016
 Received: 6 October 2016 | Accepted: 4 November 2016 | Published: 21 November 2016

发表期刊: *Clinical Epigenetics*
 影响因子: 6.091
 研究对象: 人
 合作单位: 德国萨尔兰大学、
 华大基因

研究概要

此次研究首次利用基于联合探针锚定聚合技术 (cPAS) 的BGISEQ-500平台对人样品 (脑、心脏、血液) 进行Small RNA高通量测序，并与HiSeq平台、芯片比较。优化的联合探针锚定聚合技术 (cPAS) 使得测序数据具有很高的技术重复性 (相关性系数高达0.99)，BGISEQ-500平台是对现有测序平台的补充，是Small RNA测序研究的优选平台。

研究方法

取材: 选取6个脑的重复样品，2个心脏的重复样品，2个正常人血液样品

测序: BGISEQ-500 Small RNA SE50、Illumina HiSeq

分析: BGISEQ-500技术重复分析，平台间技术分析，血液特有miRNA分析，靶基因互作分析

主要结果

1、BGISEQ-500平台技术重复性高

测序技术重复平均相关性系数达0.98，部分样本间可至0.99。

2、BGISEQ-500平台与HiSeq平台及芯片平台三者比较相关性基本一致

BGISEQ-500平台与HiSeq平台测序检测已知miRNA定量相关性达到0.75。

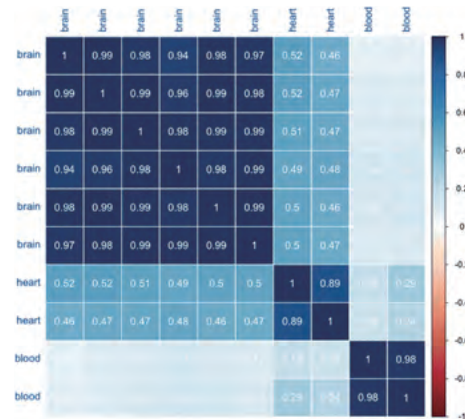


图1 BGISEQ平台技术重复相关性分析

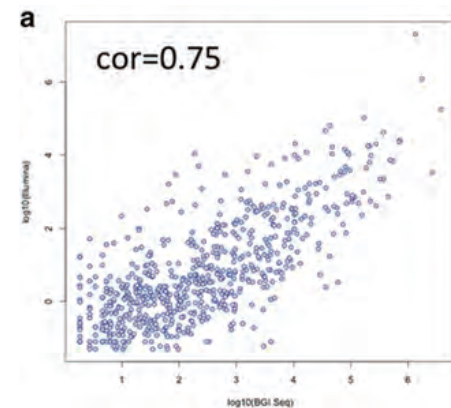


图2 a BGISEQ与H平台miRNA表达量相关性分析

3、血液样品miRNA表达分布，BGISEQ-500平台与芯片分析结果基本一致

HiSeq平台测序数据比对到miR-451a仅有0.8%的reads, BGISEQ-500平台测序数据比对到miR-451a高达45.9%的reads, 而在芯片数据中发现miR-451a有37.2%的reads比对。

文章主要比较三种miRNA研究平台的技术, BGISEQ-500平台测序检测到miRNA技术重复性更好, 与芯片已知miRNA定量结果更一致。HiSeq平台采用桥式PCR, DNA指数扩增测序; BGISEQ-500平台采用DNA纳米球(DNB), 线性DNA扩增测序, 此种扩增技术不会累积扩增错误而产生测序偏向性。

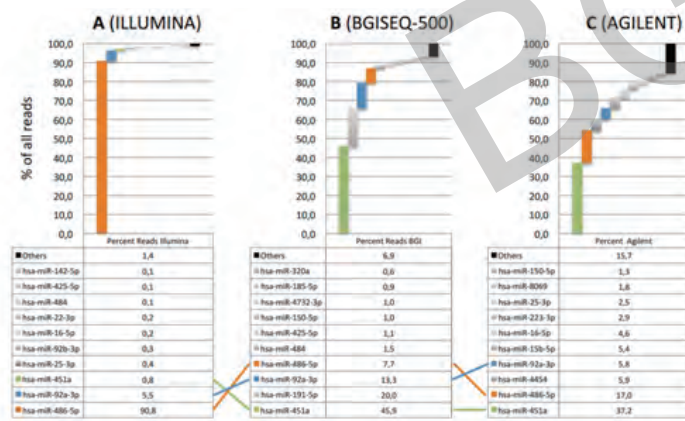


图3 HiSeq平台、BGISEQ-500平台和芯片平台数据中表达量排名前10 miRNA表达量占比

文章2

甘蓝、芥蓝嵌合体的无性后代嫁接诱导表观遗传变异的维持

Maintenance of grafting-induced epigenetic variations in the asexual progeny of *Brassica oleracea* and *B. juncea* chimera.

the plant journal



Original Article

Maintenance of grafting-induced epigenetic variations in the asexual progeny of *Brassica oleracea* and *B. juncea* chimera

Ningning Yu, Liwen Cao, Lu Yuan, Xiao Zhi, Yiqian Chen, Susheng Gan, Liping Chen

发表期刊: *The Plant Journal*

影响因子: 5.775

研究对象: 紫甘蓝、芥蓝

合作单位: 浙江大学、康奈尔大学

研究概要

嫁接会发生遗传变异, 并且能在细胞增殖和再生的过程中维持, 但其内在的机制仍是未知的。理论上, 嫁接诱导的变异主要有两种机制: (1) 异种中的遗传物质或大分子通过近距离的细胞接触传输, 远距离通过韧皮部传输; (2) 嫁接衍生的sRNA和内源sRNA通过嫁接处转移。本次研究的主要目的是确定嫁接获得的甲基化变化是否由sRNA的变化引起, 及其机制。

研究策略

取材: 芥蓝表示为TTT (代表顶端分生组织的三个细胞层, 从最外层到最内层), 紫甘蓝表示为CCC

方法: 通过嫁接方式, 无性繁殖得到芥蓝和紫甘蓝的嵌合体TCC。取无腋芽的TCC嵌合体的茎段及测序段进行培养, 诱导出不定芽, 再生得到无性子代r-CCC。相同的方法无性繁殖多代, 关注不同个体中甲基化修饰和sRNA的变化。

测序: BGISEQ Small RNA SE50, Illumina HiSeq4000 WGBS

研究结果

1、嵌合体TCC来源不定芽的鉴定

通过分子标记鉴定无性子代的遗传信息与亲本紫甘蓝相同, 但形态学表现有差异。这种遗传变异在无性繁殖的多代中都存在。通过DNA甲基化图谱发现, 无性子代中CHH胞嘧啶甲基化的数量和占比多, 且重复元件中CHH甲基化得变化最明显。r-CCC4中转座子表现更高的CHH甲基化水平。

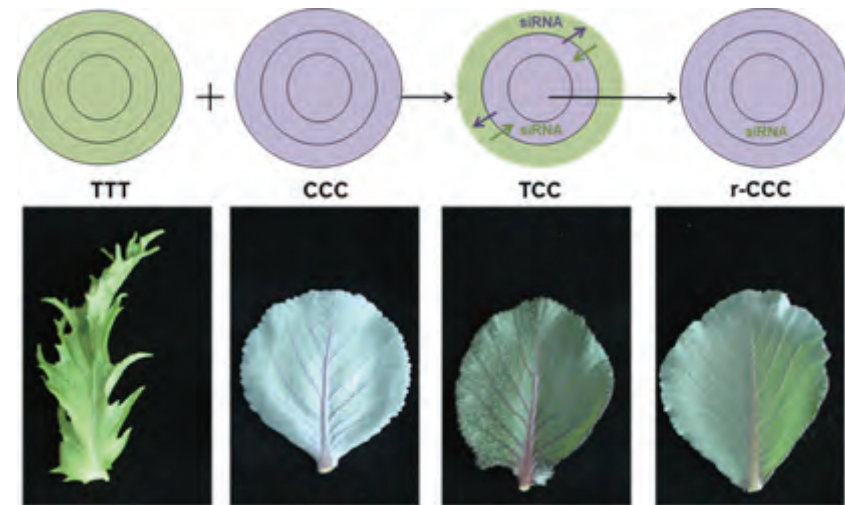
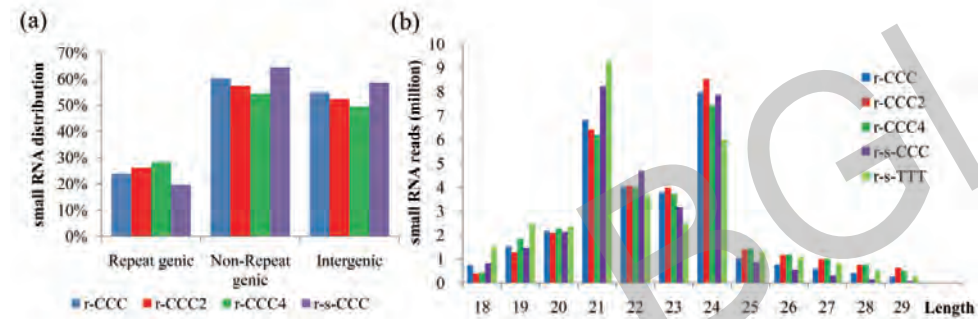


图1 嫁接嵌合体、无性繁殖后代及亲本的形态学观察与分子鉴定

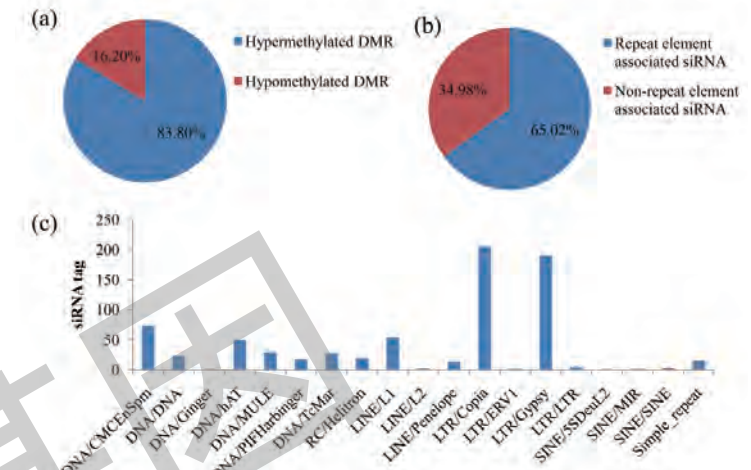
2、sRNA图谱比较

BGISEQ Small RNA高通量测序结果显示，嵌合体无性后代中大多数siRNA的表达水平明显升高，并且在子代中靶向重复序列区域的sRNA比例逐渐升高。

图2 无性后代和相应亲本中sRNA的基因组分布
(a) sRNA的分布；(b) 无性后代和相应亲本中sRNA的长度分布。

3、siRNA和甲基化的关联分析

嫁接诱导上调的siRNA中，1135个在子代中特异表达。其中65.02%的嫁接特异siRNA与重复元件相关，且在无性后代中这些siRNA靶向区域的CHH胞嘧啶表现为高甲基化修饰。因此，推测维持CHH甲基化修饰与特定的siRNAs在细胞间传递有关。

图3 特异性siRNA靶向转座子中的DNA甲基化
(a) 分析r-CCC4中特异性siRNA重叠的差异甲基化区域 (DMR)；
(b) 重复元件相关siRNA和非重复元件相关siRNA的比例；
(c) 与重复元件相关的特定siRNA的分布。

研究结论

- (1) 嫁接诱导变异出现的重要原因：嵌合体异源细胞间的交流和互作导致受体细胞中siRNA种类和表达量的改变，并通过影响DNA甲基化，特别是转座子的甲基化修饰，从而影响相关基因的表达水平。
- (2) 嫁接变异在后代中维持的实现：嫁接过程诱导的转座子相关siRNA能够在细胞再生和无性繁殖多代中稳定存在，这些siRNA可能参与维持基因组的稳定性。

附表：Small RNA文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	DDM1 represses noncoding RNA expression and RNA-directed DNA methylation in heterochromatin	<i>Plant Physiology</i>	6.456	水稻	华中农业大学
2	Transcriptome-wide analysis of immune-responsive microRNAs against poly (I:C) challenge in <i>Branchiostoma belcheri</i> by deep sequencing and bioinformatics	<i>Oncotarget</i>	5.168	文昌鱼	南京大学
3	Comparative genome-wide analysis of extracellular small RNAs from the mucor-mycosis pathogen <i>Rhizopus delemar</i>	<i>Scientific Reports</i>	4.259	根菌	南京农业大学
4	Exosomes derived from preadipocytes improve osteogenic differentiation, potentially via reduced miR-223 expression	<i>Mol Med Rep.</i>	1.922	鼠	浙江大学
5	Conserved and ubiquitous expression of piRNAs and PIWI genes in mollusks antedates the origin of somatic PIWI/piRNA expression to the root of bilaterians	<i>bioRxiv</i>	—	螺和牡蛎	德国美因兹大学有机与分子进化研究所

ChIP-Seq

染色质免疫共沉淀技术（Chromatin immunoprecipitation, ChIP）是一种重要的在体内环境中研究蛋白质与DNA相互作用的经典实验方法，广泛应用于组蛋白修饰、特定转录因子的基因调控作用等相关领域。

随着新一代测序技术的发展和成熟，依托高通量测序并整合染色质免疫沉淀技术，形成了新的技术，即ChIP sequencing（Chromatin Immunoprecipitation Sequencing, ChIP-Seq）。可以在全基因组范围内对蛋白质结合位点进行高效而准确地筛选与鉴定，广泛应用于组蛋白修饰、转录因子调控等相关领域的研究。

采用特异性抗体对目的蛋白进行免疫沉淀后，分离与其结合的基因组DNA片段，再通过高通量测序与数据分析，在全基因组范围内寻找目的蛋白的DNA结合位点，并且可以基于多个样品进行差异比较。

文章1

FOXN3-NEAT1-SIN3A阻遏物复合物促进激素反应性乳腺癌的进展

The FOXN3-NEAT1-SIN3A repressor complex promotes progression of hormonally responsive breast cancer



发表期刊: *The Journal of Clinical*

Investigation

影响因子: 12.784

研究对象: 人

合作单位: 北京大学医学部

研究背景

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

研究方法

取材: MCF-7细胞的总蛋白、乳腺癌组织

研究技术: 染色质免疫共沉淀测序 (BGISEQ CHIP-Seq)、RNA免疫共沉淀测序 (iRIP-Seq)、lncRNA表达谱芯片、CHART-Seq、快速蛋白质液相色谱法 (FPLC) 等

验证方法: qPCR、Western blotting

研究概要

使用ChIP-Seq和iRIP-Seq技术, 揭示了FOXN3-NEAT1-SIN3A阻遏物复合体的存在方式FOXN3-NEAT1-SIN3A阻遏物复合物通过抑制GATA3基因表达, 改变了上皮-间质转化 (EMT), 鉴定了FOXN3-SIN3A复合物靶向基因, 验证FOXN3-NEAT1-SIN3A阻遏物复合物在体外促进了乳腺癌细胞的EMT转化侵袭以及体内乳腺癌的转移。

研究结果

1、揭示了FOXN3-NEAT1-SIN3A阻遏物复合体的存在模式

FOXN3是人类细胞中的转录抑制因子, 为证实FOXN3与SIN3A阻遏物相互作用关系, 提取MCF-7细胞的总蛋白, 使用FOXN3的抗体进行免疫沉淀, 免疫印迹法显示SIN3A蛋白同时被FOXN3抗体免疫共沉淀, 相反, 使用SIN3A阻遏蛋白抗体富集, 同时检测到FOXN3蛋白被SIN3A抗体免疫共沉淀。在RNase存在的情况下, FOXN3-SIN3A复合物不再被检测到, 认为是某种RNA分子参与了阻遏物复合物, 为验证, 针对MCF-7细胞使用FOXN3抗体进行RNA免疫沉淀偶联高通量测序 (iRIP-Seq), 分析鉴定出2个显著富集lncRNA (图1A), NEAT1和MALAT1, 为确定具体是哪一种lncRNA参与了阻遏复合体的作用, 使用lncRNA表达谱芯片对MCF-7和MDA-MB-231细胞中lncRNAs进行比较分析, 结果表明MCF-7细胞中NEAT1表达水平明显高于MDA-MB-231细胞, 而MALAT1的表达水平在这两种细胞株中相当 (图1C)。通过qPCR分析MCF-7、T-47D和MDA-MB-231细胞中NEAT1和MALAT1的表达情况, NEAT1在MCF-7和T-47D细胞中高表达, 进一步验证了NEAT1参与了阻遏复合体的作用。文章用了多种方法共同揭示了FOXN3-NEAT1-SIN3A阻遏物复合体的存在模式。

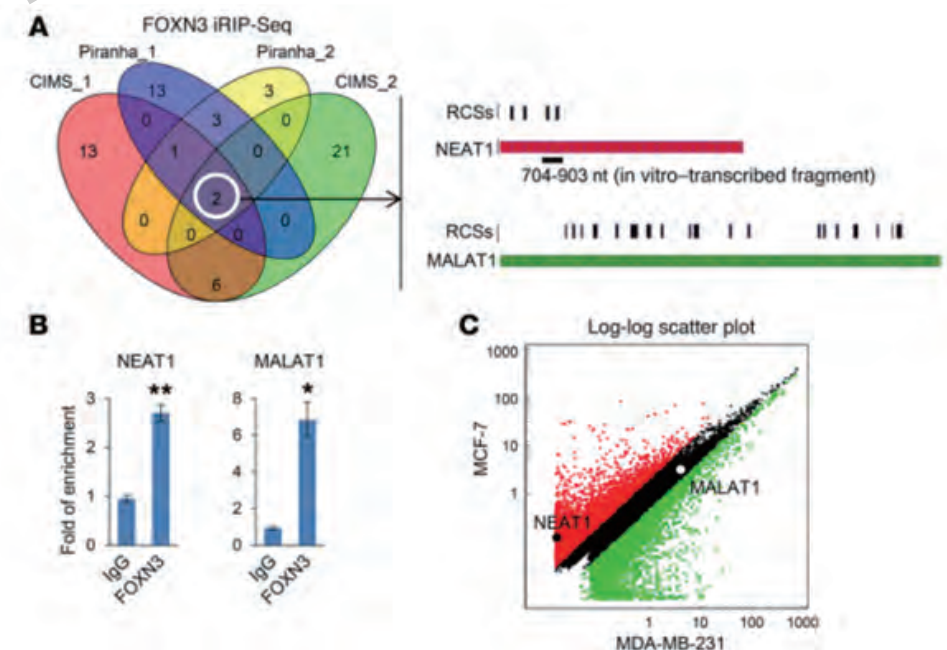


图1 A.FOXN3 iRIP-Seq显著富集lncRNA分析结果

B. iRIP-Seq结果qPCR验证结果

C.lncRNA表达芯片检测散点图

2、全基因组范围内鉴定了FOXN3-NEAT1-SIN3A阻遏物复合体的目标基因

为研究FOXN3-NEAT1-SIN3A阻遏物复合体的功能，使用BGISEQ-500 ChIP-Seq对MCF-7细胞中的FOXN3蛋白和SIN3A蛋白分别进行靶向基因鉴定，得到FOXN3-SIN3A复合物靶向基因4087个(图2A)，FOXN3和SIN3A具有非常相似的motif结构，在一些代表性的靶向基因上表现出高度相似的peak峰图(图2B)。将ChIP-Seq结果与公认的NEAT1的CHART-Seq数据进行关联分析，共鉴定到1441个FOXN3-NEAT1-SIN3A阻遏复合体基因，这些基因在上皮-间质转化过程中调节乳腺上皮细胞分化和乳腺癌进展期细胞侵袭中的作用被广泛认可。

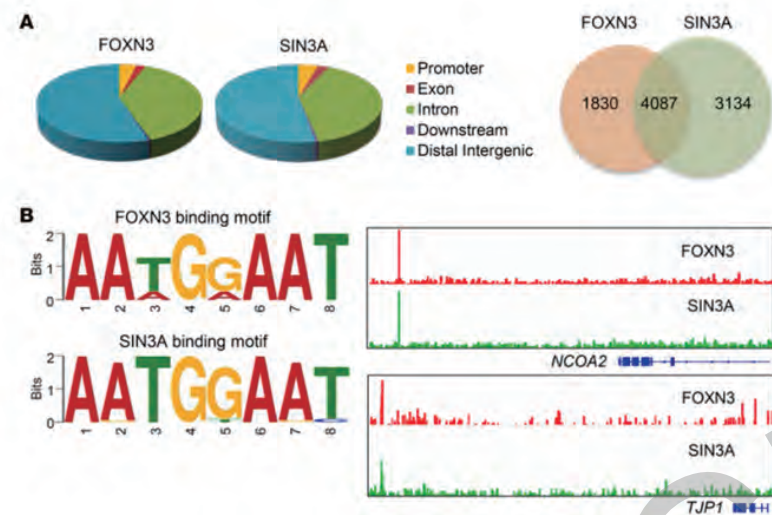


图2 A. 基于FOXN3和SIN3A抗体的BGISEQ-500 ChIP-Seq结果维恩图

B. 显著性富集的peak可视化对比和binding motif分析

3、FOXN3-NEAT1-SIN3A阻遏复合体促进乳腺癌细胞的EMT转化和侵袭

用FOXN3或NEAT1转染MCF-7细胞，通过相差显微镜做形态学检查，上皮（E-钙粘蛋白和TJP1）和间质（波形蛋白和纤连蛋白）标记物通过共聚焦显微镜进行免疫荧光染色和分析，观察到FOXN3或NEAT1过表达细胞，细胞与细胞间出现接触质膜的损失，细胞变得分散，并且它们鹅卵石形状的被成纤维细胞形状所代替，表明EMT的特征性形态学发生了变化（图3），说明FOXN3-NEAT1-SIN3A阻遏复合体促进了EMT转化。进一步用Transwell细胞侵袭测定，证明FOXN3-NEAT1-SIN3A阻遏复合体通过抑制下游靶基因（包括GATA3）的转录促进激素反应性乳腺癌中的细胞侵袭。

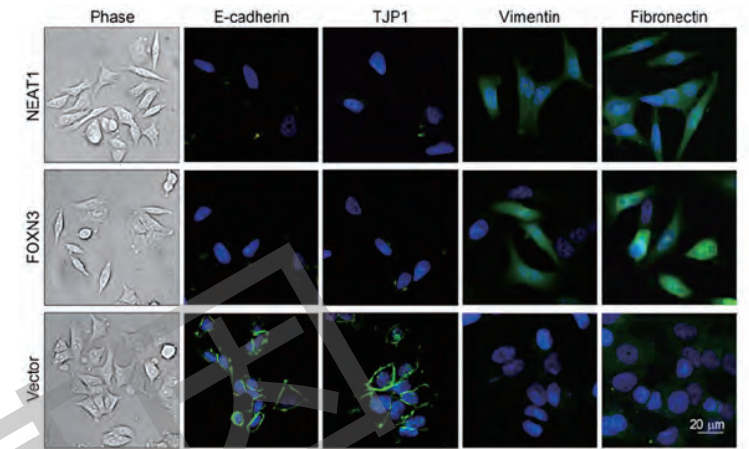


图3 相差显微镜形态学检查结果

4、FOXN3-NEAT1-SIN3A阻遏物复合体促进乳腺癌的转移

设计用于稳定表达萤火虫荧光素酶（MCF-7RR-Luc, PerkinElmer）的MCF-7细胞，使用携带空载质粒或对照shRNA, FOXN3, NEAT1, FOXN3 + shSIN3A, NEAT1 + shSIN3A或FOXN3 + GATA3慢病毒转染，原位植入6周龄雌性免疫缺陷小鼠（n=8）的腹部乳腺脂肪垫中，10周后使用IVIS成像系统进行生物光学成像分析，结果表明FOXN3或NEAT1的过度表达与SIN3A阻遏物有效地促进了MCF-7RR-Luc肿瘤的肺转移。

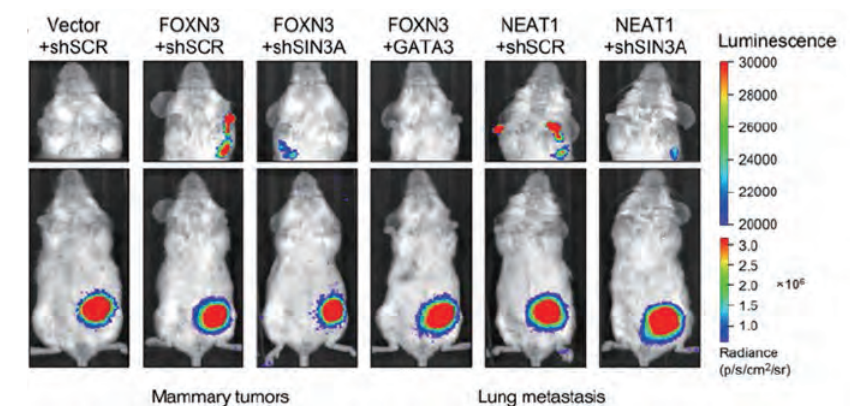


图4 慢病毒转染后IVIS成像系统分析结果

5、FOXN3和NEAT1在乳腺癌中上调，其高水平与较高的肿瘤分级和较差的存活率相关

收集与乳腺癌患者的相邻正常乳腺组织配对的24个乳腺癌样品，通过qPCR分析FOXN3和NEAT1的表达。结果表明，FOXN3和NEAT1均在乳腺癌中表达上调，GATA3的表达水平与FOXN3或NEAT1的表达水平存在显著的负相关。进一步对正常（15）、I级（26）、II级（82）、III级（52）共175例乳腺癌样本的FOXN3、NEAT1表达水平进行分析，同时结合Oncomine中的Bittner的乳腺癌数据，显示FOXN3和NEAT1的表达水平与ER+乳腺癌的组织学分级呈正相关。进一步得出FOXN3和NEAT1的高水平表达，乳腺癌患者的总体存活率更差。

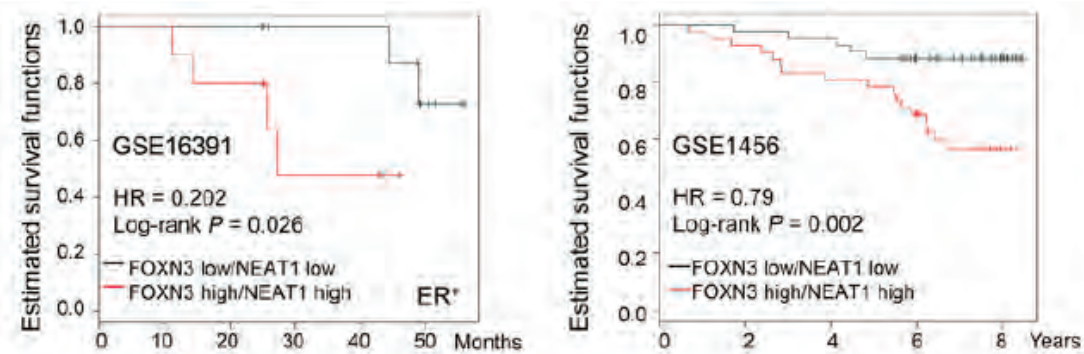


图5 FOXN3和NEAT1表达水平与乳腺癌患者的总体存活率相关性分析

6、展望

近年来，研究发现FOXN3和NEAT1广泛涉及到不同组织恶性肿瘤的发生，但关于FOXN3和NEAT1在乳腺癌发展中的作用却知之甚少，在本研究中，FOXN3-NEAT1-SIN3A阻遏物复合体通过抑制下游靶基因的转录促进体外乳腺癌细胞的EMT和侵袭，并在体内促进乳腺癌的传播和转移，为理解FOXN3的病理及生理功能提供了分子基础，这些分子可作为乳腺癌治疗的潜在靶标，复合物对GATA3的转录抑制表明FOXN3在乳腺发育中的重要作用。需要进一步确定动物发育中FOXN3和FOXN3-NEAT1-SIN3A阻遏物复合体的大小和数量，以及为什么只在脊椎动物中FOXN3-SIN3A复合物作用需要NEAT1参与。

研究意义

研究人员将免疫共沉淀技术发挥到极致，同时分析特定蛋白与DNA互作关系 (ChIP-Seq) 和RNA互作关系 (iRIP-Seq)，揭示FOXN3-NEAT1-SIN3A阻遏物复合体的存在模式，结合lncRNA表达谱芯片、CHART-Seq等技术，对阻遏复合体的生物学功能进行进一步挖掘并验证。

附表：ChIP-Seq文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	Nuclear carbonic anhydrase 6B associates with PRMT5 to epigenetically promote IL-12 expression in innate response	<i>PNAS</i>	9.661	小鼠	北京协和医学院

全基因组重测序 (WGS)

全基因组重测序 (WGS) 是对已知基因组序列的物种进行不同个体的基因组测序, 并在此基础上对个体或群体进行差异性分析。全基因组重测序的个体, 通过序列比对, 可以找到大量的单核苷酸多态性位点 (SNP), 插入缺失位点 (InDel, Insertion/Deletion)、结构变异位点 (SV, Structure Variation) 位点, 通过生物信息手段, 分析不同个体基因组间的结构差异, 可以获得同一物种不同个体之间的遗传变异图谱。

人全基因组重测序是基于人基因组参考序列对个体或群体进行全基因组测序, 并在个体或群体水平进行生物信息分析的方法。基于全基因组重测序的人类遗传学和群体进化学的研究, 能够快速筛选出基因组范围内的遗传变异, 实现基因型多样性分析、遗传进化分析以及致病和易感性基因等的筛选。随着二代测序技术的发展与普及, 全基因组重测序已成为人类遗传学、转化医学和群体进化领域最为迅速而有效的方法之一, 可更全面地挖掘基因序列差异和结构变异, 包括单碱基突变、插入缺失变异、拷贝数变异和结构变异, 在全基因组水平上扫描并检测与表型差异、疾病、进化等相关的突变位点, 具有重大的科研价值。

随着测序成本降低和已知基因组序列物种增多, 全基因组重测序已经成为动植物基因挖掘和研究群体进化的最为迅速、有效的方法之一。全基因组重测序对基因组序列已知物种的样品进行测序, 并在个体或群体水平进行全基因组水平的差异性分析, 可以全面的挖掘基因序列差异和结构变异。同时, 在全基因组水平上扫描并检测与生物体重要性状相关的基因位点, 具有重大的科研价值和产业价值。

文章1

BGISEQ-500测序仪的参考人类基因组数据集

A reference human genome dataset of the BGISEQ-500 sequencer



发表期刊: *GigaScience*
 影响因子: 7.267
 研究对象: 人
 合作单位: 华大基因、国家食品药品监督管理总局 (NIFDC) 和国家食品药品监督管理局湖北医疗器械质量监督检验中心

研究概要

与公开发布的HiSeq 2500人全基因组重测序 (WGS) 数据分析比较, BGISEQ-500测序平台在比对与变异检测方面与HiSeq 2500表现出较高一致性。

研究策略

取材: “瓶中基因组”联盟发布的人类细胞系HG001(NA12878)

测序: BGISEQ-500 PE100、PE50; Illumina HiSeq 2500 PE150

分析: 采用相同分析、统计方法, 对两平台的数据质量、比对结果、变异检测等进行比较

研究结果

1、两平台通量水平相当

BGISEQ-500测序仪产出的PE50及PE100数据分别约135.5 Gbp、153.6 Gbp, 共两张芯片, 耗时约72小时, 与HiSeq 2500测序仪快速模式v1 (per single flow cell 40小时产出80Gbp) 通量相似。

2、两平台都具有较高的碱基质量分布和较相似的GC含量分布

对原始Fastq数据使用相同条件进行数据过滤, 过滤后的数据显示BGISEQ-500和HiSeq 2500两平台都具有较高的碱基质量分布和较相似的GC含量分布 (图1)。

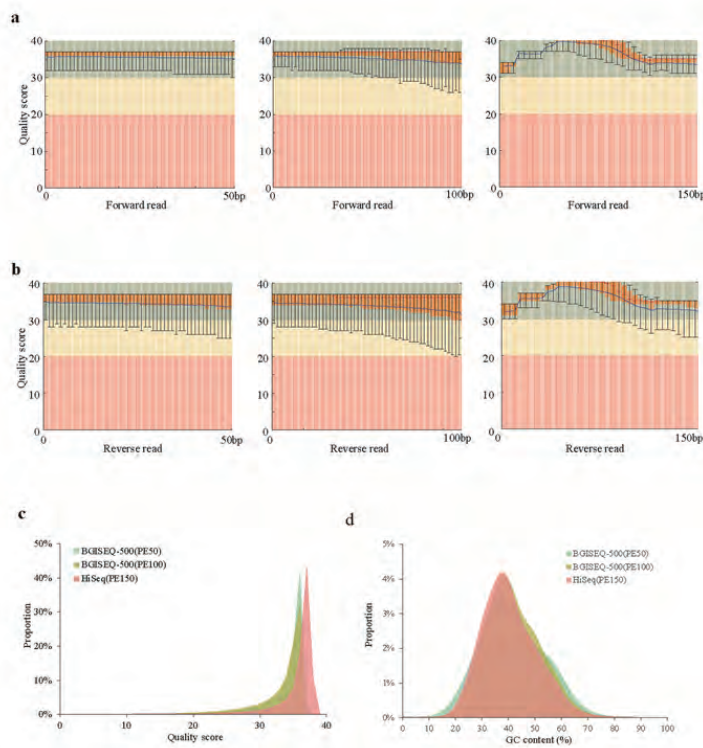


图1 数据质量比较

- a. read1碱基质量分布;
- b. read2 碱基质量分布;
- c. PE50、PE100、HiSeq总体碱基质量分布;
- d. PE50、PE100、HiSeq GC含量分布

3、两平台具有较高的比对率和相似的覆盖度和错配率，较好的基因组覆盖度

使用BWA + GATK流程 (图2) 对BGISEQ-500 PE100、PE50以及HiSeq 2500 PE 150数据进行比对和变异检测，与HiSeq 2500平台相比，两组数据都具有较高的比对率和相似的覆盖度和错配率，此外，两平台也显示出较好的基因组覆盖度 (表1)。

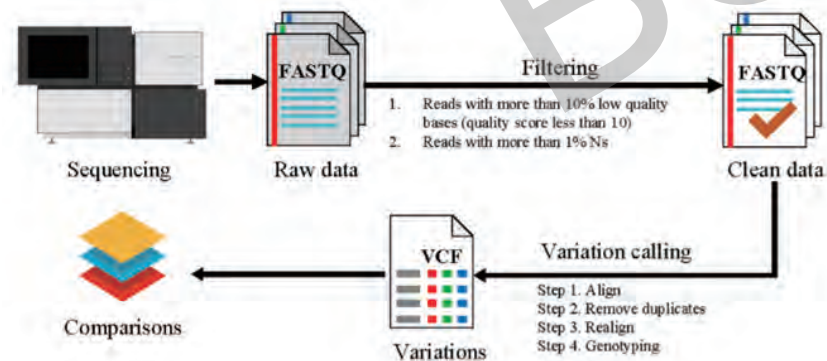


图2 BWA + GATK 比对变异检测流程

表1 比对统计

Metrics	BGISEQ-500 PE50	BGISEQ-500 PE100	HiSeq 2500 PE150
Clean reads	2,378,725,921	1,136,008,901	708,941,148
Clean bases (bp)	118,936,296,050	113,600,890,100	104,923,289,904
Mapping rate	97.87%	99.22%	99.05%
Unique rate	93.17%	96.47%	97.06%
Duplicate rate	6.26%	2.47%	1.52%
Mismatch rate	0.34%	0.58%	0.56%
Average sequencing depth	37.57	37.44	34.52
Coverage	99.28%	99.12%	99.06%
Coverage at least 4x	98.90%	98.69%	98.60%
Coverage at least 10x	97.97%	97.81%	97.83%
Coverage at least 20x	95.78%	96.06%	94.81%

4、两平台SNP变异检测具有相似的灵敏度，阳性检出率，假阳性率和假阴性率

在变异检测方面，BGISEQ-500 PE50、PE100 SNP检出数略低于HiSeq 2500 PE150，但BGISEQ-500 PE100与HiSeq 2500 PE150具有相似的灵敏度、阳性检出率、假阳性率和假阴性率 (表2、表3)。

表2 变异检测统计

Metrics	BGISEQ-500 PE50	BGISEQ-500 PE100	HiSeq 2500 PE150
SNPs	3,451,124	3,477,642	3,609,606
1000genome and dbSNP	3,242,083	3,288,653	3,347,441
1000genome specific	1,260	420	693
dbSNP specific	180,935	179,967	243,256
dbSNP rate	99.19%	99.74%	99.48%
Novel	26,846	8,602	18,216
Homozygous	1,426,328	1,433,490	1,472,063
Heterzygous	2,024,796	2,044,152	2,137,543
Synonymous	19,880	20,012	20,860
Ti/Tv	2.0462	2.065	2.0427
dbSNP Ti/Tv	2.0608	2.0693	2.0503
Novel Ti/Tv	0.8948	0.9775	1.0544

表3 变异评估统计

Variant type	Metrics	BGISEQ-500 PE50	BGISEQ-500 PE100	HiSeq 2500 PE150
SNPs	True Positive	3,006,132	3,071,579	3,084,449
	False Positive	15,203	6,907	4,318
	False Negative	186,825	121,379	108,508
	Precision	99.50%	99.78%	99.86%
	Sensitivity	94.15%	96.20%	96.60%
	FPR	0.00060%	0.00020%	0.00017%
	FNR	5.85%	3.80%	3.40%

研究意义

从变异检测结果看, BGISEQ-500平台数据与HiSeq 2500平台相差不大, 证明该测序仪可以用于各种科研与临床应用。随着新技术的研发, BGISEQ-500的测序质量、速度以及通量仍有较大的提升空间, 测序读长、不同插入片段以及分析软件的研发必将提升数据的整体表现。同时, BGISEQ-500测序仪可用于转录组、表观基因组、宏基因组等测序。

本次测序数据是BGISEQ-500测序仪的第一个参考数据集, 展示了华大基因国产新测序平台的整体面貌及基本特点, 希望能够促进新技术以及软件的发展, 能够更加快速、准确地分析BGISEQ-500的测序数据。

文章2

BGISEQ-500与Illumina HiSeq 2500测序平台对古基因组测序性能的比较

Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing



发表期刊: *Giga Science*

影响因子: 7.267

研究对象: 狗和狼

合作单位: 华大基因、丹麦哥本哈根大学、丹麦技术大学等

研究概要

分别采用BGISEQ-500和Illumina HiSeq 2500平台, 对91-14000年前的8个古老大型犬科动物的DNA 样本进行测序(表1), 对测序性能和数据质量进行比较。研究结果表明, 数据表现在两个测序平台上基本相当(表2)。BGISEQ-500是可用于古基因组测序的有效并有价值的潜在替代平台, 值得用此平台进一步探索降解DNA。

研究策略

取材: 91-14000年前8个古老大型犬科动物DNA

测序: BGISEQ-500; Illumina HiSeq 2500

分析: 采用相同分析、统计方法, 对两平台的数据质量、比对结果、变异检测等进行比较

表1 8个历史古老的狗和狼样本信息

Sample	Original ID	Material	Species	Locality	Age	Extraction
214	CN 214	Hide	Wolf	Uummannaq, Greenland	Before 1869 AD	A
1921	CN 1921	Hide	Wolf	Rosenvinge Bugt, Greenland	1925 AD	A
P84	MGUH VP 3332	Humerus	Wolf	Vøvedal, Greenland	ca. 7,620 cal YBP	B
P83	NKA 1950x2906	Canine tooth	Dog	GUS, Greenland	ca. 600-1,000 YBP	B
P79	ZMK 350/1982	Tibia	Dog	Qajã, Greenland	ca. 3,6-2,700 YBP	B
FRC	FRC	Cartilage	Large canid	Tumat, Siberia	ca. 14,122 cal YBP	C
L	L	Liver	Large canid	Tumat, Siberia	ca. 14,122 cal YBP	C
M1	M1	Muscle	Large canid	Tumat, Siberia	ca. 14,122 cal YBP	C

表2 数据统计总结

Sample	Platform	Total reads	Normalised % reads retained after adapter removal	Normalized clonality	Normalized endogenous DNA (%)	Normalised length of uniquely mapped reads	θ	δD	δS	GC Content (%)	mtDNA (%)
1921	Illumina	3.08E+07	94.69	0.11	58.73	40.77	0.008	0.008	0.154	51.58	4.51E-03
	BGISEQ-500	5.32E+07	83.97	0.15	59.37	42.14	0.009	0.008	0.132	50.42	2.57E-03
214	Illumina	1.35E+07	99.13	0.07	74.25	49.37	0.008	0.011	0.084	48.60	4.15E-03
	BGISEQ-500	1.98E+08	99.55	0.07	75.51	53.08	0.009	0.012	0.061	47.75	3.11E-04
FRC	Illumina	1.64E+07	99.54	0.03	11.58	73.05	0.008	0.012	0.399	44.01	4.55E-03
	BGISEQ-500	3.39E+08	99.79	0.02	10.22	75.63	0.012	0.012	0.325	43.64	1.98E-04
L	Illumina	2.91E+07	99.63	0.09	1.03	64.65	0.013	0.010	0.415	43.24	6.04E-03
	BGISEQ-500	2.44E+08	99.77	0.08	0.85	66.72	0.013	0.009	0.262	45.99	7.09E-04
M1	Illumina	5.10E+07	99.38	0.06	64.09	72.95	0.007	0.010	0.395	44.27	8.02E-03
	BGISEQ-500	1.79E+08	99.74	0.06	54.80	76.76	0.012	0.010	0.258	43.23	2.31E-03
P79	Illumina	4.18E+07	98.48	0.38	0.07	52.45	0.030	0.012	0.880	43.36	4.65E-06
	BGISEQ-500	8.55E+07	98.08	0.10	0.06	45.77	0.039	0.011	0.550	44.21	6.40E-07
P83	Illumina	2.77E+07	84.67	0.58	0.64	65.78	0.014	0.040	0.842	42.32	4.85E-04
	BGISEQ-500	2.32E+07	86.84	0.32	0.47	66.55	0.017	0.040	0.773	44.30	3.87E-04
P84	Illumina	5.94E+07	98.70	0.31	0.12	54.79	0.015	0.030	0.355	44.42	2.71E-06
	BGISEQ-500	1.57E+08	92.45	0.08	0.10	51.13	0.022	0.020	0.154	47.99	5.15E-07

* δD : 双 (Double) 链损伤率, 即双链胞嘧啶脱氢比率, (C->T);
 δS : 单 (Single) 链损伤率, 即单链胞嘧啶脱氢比率, (C->T);
 θ : 考虑DNA损伤后参考序列和样本之间差异的背景率;
 Clonality: 克隆率, 在去DUP之后保留的reads数与去adapter和trim后的reads的比率。

研究结果

1、内源DNA水平

决定古基因组测序成本的最大因素是内源核DNA含量, 以及内源核DNA平均序列长度, 两个指标越高测序成本则越低。这两个指标在两平台间并没有显著差异 (表3)。

2、测序准确性

双链DNA损伤率 (δD) 没有显著差异。BGISEQ-500平台的单链DNA损伤率 (δS) 略低。与参考基因组差异的背景值 θ , BGISEQ-500平台略高。

BGISEQ-500平台产生的8个文库中, 有7个文库复杂度显著高于HiSeq平台 (图1)。其中差异最大的, 是距今14,000年前的最古老样本之一。

表3 数据统计分析结果

Test	Paired t-test t	P-value
% Reads retained	-1.131308	0.295
Clonality levels	-1.942886	0.093
% Endogenous DNA	-0.956158	0.371
Endogenous DNA average read length	0.0375544	0.718
θ	3.366145	0.012*
δD	-1.09765	0.309
δS	-3.425669	0.011*
% GC	1.091076	0.311
% mtDNA	-3.073585	0.018*

* Significant at P<0.05

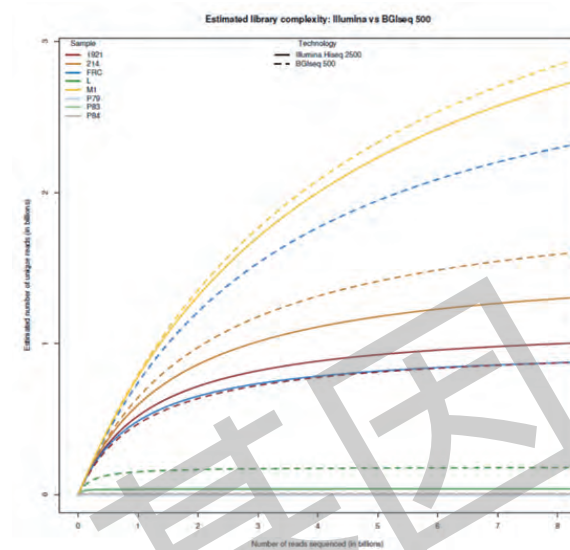


图1 实线是在Illumina HiSeq 2500平台上文库复杂性的估计; 虚线是对BGISEQ-500平台测序的文库复杂性的估计; 8个样本中的每一个由不同的颜色表示

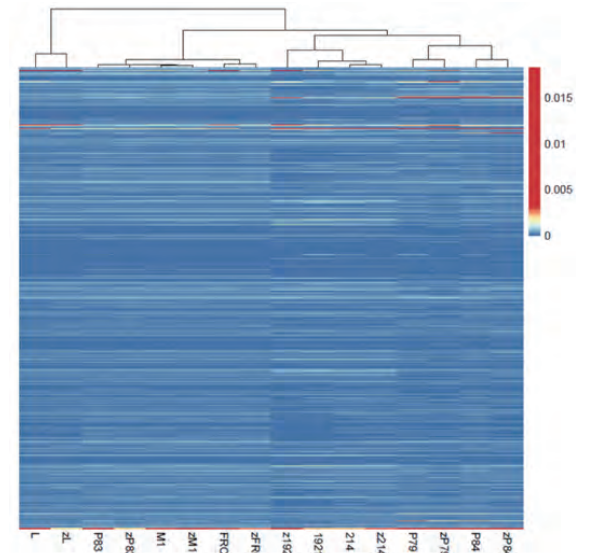


图2 文库间k-mer计数的热图

3、基因组覆盖均一性

使用k-mer频率和GC含量讨论测序在基因组上的覆盖区域是否有平台特异性。多数样品的两个平台测序方法之间的k-mer模式是一致的, 样品配对聚集在一起。只有2个样本 (P83和1921) 的差异略微明显 (图2)

平均GC含量没有统计学差异 (表2, 表3), 进一步分析中, 相同窗口的片段数在BGISEQ-500和Illumina之间有很好的相关性 (图3, 图4)。

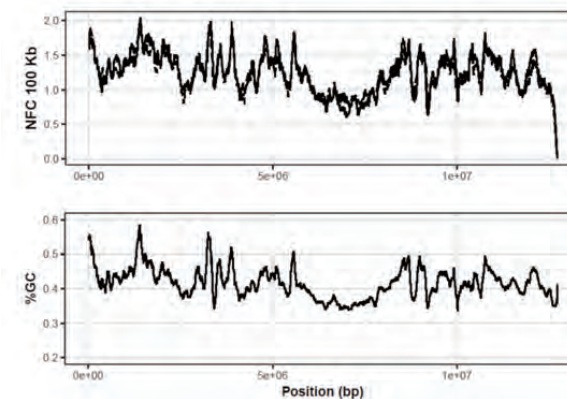


图3 上图: 每100Kb窗口的中值归一化片段数 (normalised fragment count, NFC), 实线表示Illumina数据, 虚线表示BGISEQ-500数据; 下图: 与上图相同的窗口中计算GC的百分比。

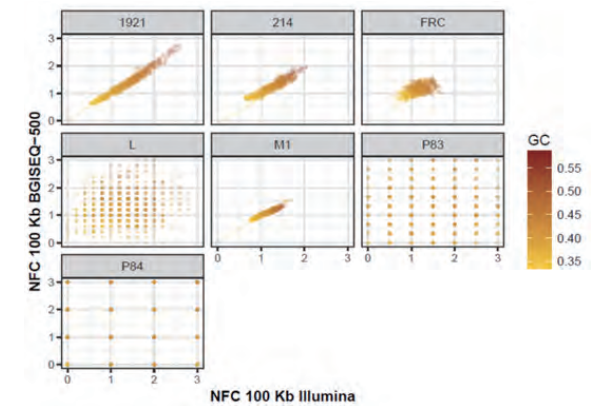


图4 100Kb的窗口的Illumina vs BGISEQ-500的中值归一化片段数 (NFC)。每个点的颜色对应于窗口的GC含量。对于高质量样本 (1921, 214, FRC, M1), 可以观察到两个平台之间的NFC有很好的相关性, 片段数似乎与GC含量相关。

4、CNV

通过质控的相关系数 r^2 值的范围在0.35-0.96之间(表4)。结果说明两测序平台对高质量CN calls都可行,但很难说明CN calls差异造成的原因。

表4 平台之间的相同基因组窗口中拷贝数的相关系数

Sample	CW Size					
	1000Kbp	100Kbp	50Kbp	10Kbp	5Kbp	1Kbp
214	0.905 ^a	0.331 ^a	0.354 ^a	0.506 ^a	0.519 ^b	0.433 ^c
1921	0.963 ^a	0.384 ^a	0.392 ^a	0.428 ^a	0.432 ^b	0.393 ^c
FRC	0.582 ^a	0.847 ^a	0.870 ^a	0.873 ^b	0.870 ^c	0.783 ^c
L	0.941 ^b	0.957 ^c	0.964 ^c	0.958 ^c	0.955 ^c	ND
M1	0.665 ^a	0.943 ^a	0.952 ^a	0.953 ^a	0.950 ^a	0.910 ^b
P79	0.672 ^b	ND	ND	ND	ND	ND
P83	0.203 ^b	0.003 ^c	0.004 ^c	0.003 ^c	0.002 ^c	ND
P84	0.919 ^b	0.001 ^c	0.001 ^c	ND	ND	ND

a 质量控制通过; b 质控结果良好; c 至少有一个平台的质控失败

研究意义

虽然本研究仅基于8个标本,但其内源DNA含量(<1-75%) and 标准化平均内源性DNA序列长度(约42-76 bp)都是其他许多古代样品的典型值。研究结果除了错误率有非常细微的高之外, BGISEQ-500的性能在准确性、基因组覆盖均一性、CNV calling上都是可比的。

文章3

BGISEQ-500和HiSeq X Ten全基因组测序鉴定生殖细胞和体细胞变异

Germline and somatic variant identification using BGISEQ-500 and HiSeq X Ten whole genome sequencing



RESEARCH ARTICLE

Germline and somatic variant identification using BGISEQ-500 and HiSeq X Ten whole genome sequencing

Ann-Marie Patch^{1*}, Katia Nones^{1*}, Stephen H. Kazakoff^{1*}, Felicity Newell¹, Scott Wood¹, Conrad Leonard¹, Oliver Holmes¹, Qinying Xu¹, Venkateswar Addala¹, Jenette Creaney^{2,3}, Bruce W. Robinson^{2,3}, Shujin Fu⁴, Chunyu Geng⁴, Tong Li⁴, Wenwei Zhang⁴, Xinming Liang⁴, Junhua Rao⁴, Jiahao Wang⁴, Mingyu Tian⁴, Yonggang Zhao⁴, Fei Teng⁴, Honglan Gou⁴, Bicheng Yang⁴, Hui Jiang⁴, Feng Mu⁴, John V. Pearson^{1*}, Nicola Waddell^{1*}

¹ Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia, ² National Centre for Asbestos Related Disease, School of Medicine and Pharmacology, University of Western Australia, Nedlands, Western Australia, Australia, ³ Department of Respiratory Medicine, Sir Charles Gardner Hospital, Nedlands, Western Australia, Australia, ⁴ BGI, BGI-Sherzhen, Shenzhen, China



发表期刊: PLOS ONE

影响因子: 3.057

研究对象: 人

合作单位: 华大基因、西澳大利亚大学医学和药学院、澳大利亚医院呼吸科等

研究概要

使用BGISEQ-500平台对三种恶性胸膜间皮瘤及其对照的正常样本进行全基因组测序,并与Illumina HiSeq X Ten平台测序结果进行评估。两平台数据均使用相同的分析流程,分别比较生殖细胞和体细胞单核苷酸变异(SNP)、小插入或缺失(InDel)。结果表明BGISEQ-500平台通过全基因组测序来鉴定肿瘤样本的体细胞和生殖细胞突变是有潜力的可适用性的平台,这也是该平台首次公开可用的癌症基因组数据。

研究策略

取材: 三种恶性胸膜间皮瘤及其对照的正常样本

测序: BGISEQ-500; Illumina HiSeq X Ten

分析: 比较生殖细胞和体细胞单核苷酸变异(SNP)、小插入或缺失(InDel)

研究结果

1、生殖细胞突变

BGISEQ-500平台和HiSeq X Ten平台识别SNP的能力与SNP分型芯片(Infinium Omni2.5-8, Illumina)高度一致(>99%)。两个测序平台中鉴定的生殖细胞SNV和indels也高度一致(分别为86%和81.5%)。

表1 SNP芯片数据分别与BGISEQ-500和HiSeq X Ten数据比较, 种系基因型一致性的百分比

Table 1. The percent concordance of germline genotypes ascertained by SNP arrays compared to the BGISEQ-500 and HiSeq X Ten data.

Patient	SNP array vs BGISEQ-500	SNP array vs HiSeq X Ten
9869	99.797	99.789
11202	99.794	99.794
11398	99.797	99.795

在3个样本数据总10,686,962个生殖细胞SNV中, BGISEQ-500和HiSeq X Ten平台特有的SNV突变比例分别为0.36%和1.85%; 699,321个生殖细胞indels中, BGISEQ-500和HiSeq X Ten平台特有的indels比例分别为0.19%和3.23%, 具体比例如图1所示。

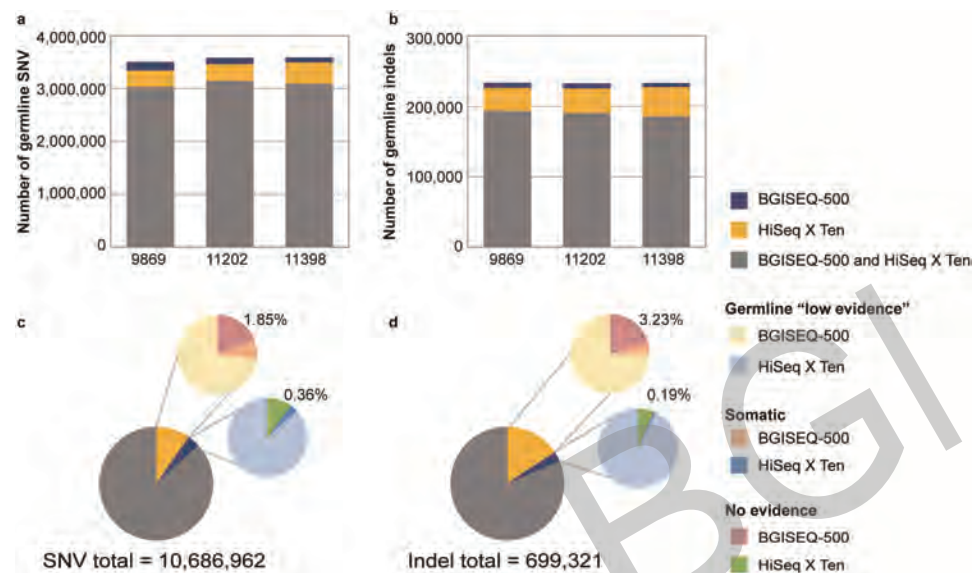


Fig 2. Germline variants identified in three mesothelioma samples (patients: 9869, 11202 and 11398) using BGISEQ-500 and HiSeq X Ten data. The number of germline SNV (a) and indels (b) identified in each patient using the BGISEQ-500 and HiSeq X Ten platforms. We investigated germline SNV (c) and indels (d) which were only called in one platform and that fall into three categories: i) identified as germline in the other platform but with low evidence; ii) identified in the other platform but predicted as a somatic variant; or iii) not identified in the other platform. Across the 3 patients only 197,434 (1.85%) SNVs were truly unique to the HiSeq X Ten and not identified in the BGISEQ-500 (c). Similarly in the BGISEQ-500 platform only 38,236 SNVs (0.36% of the total) were truly unique to the BGISEQ-500, not called in the HiSeq X Ten data (c). The same pattern was observed for indels (d), only 3.23% were unique to HiSeq X Ten and 0.19% to BGISEQ-500.

图1 利用BGISEQ-500和HiSeq X Ten的数据鉴定3个胸膜间皮瘤的生殖细胞突变

2、体细胞突变

三名患者中总共10,890个体细胞SNV, 大部分体细胞SNV (72%) 在两个平台中被识别, 小部分为BGISEQ-500和HiSeq X Ten两平台特有的 (分别为14%, 14%)。只有38%的indels可以同时被两个测序平台识别, BGISEQ-500和HiSeq X Ten分别特有的为38%和14%。和SNV分析相比, 体细胞indels突变有较高的不一致性, 具体比例如图2所示。

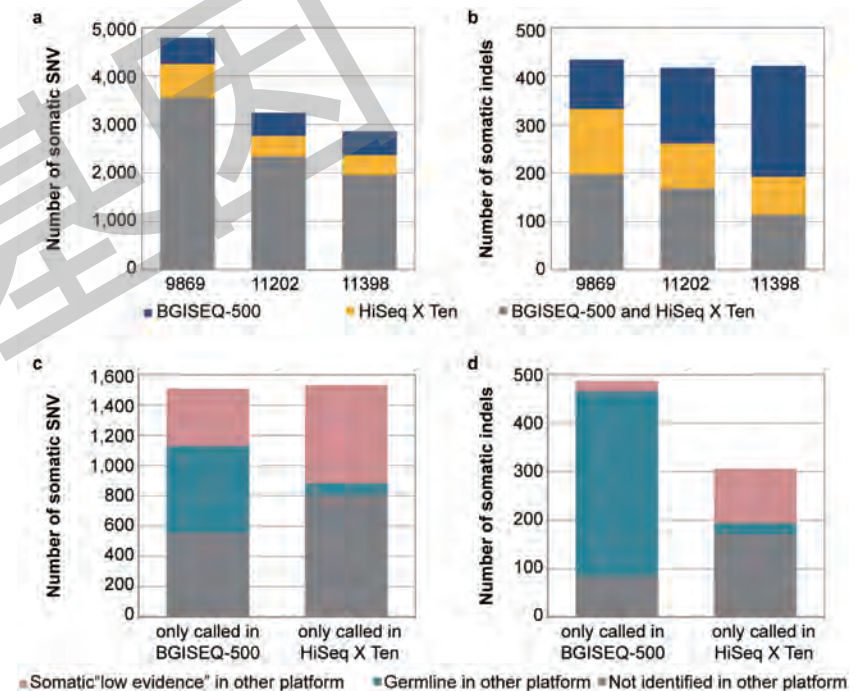


Fig 4. Somatic variants in mesothelioma patients identified using BGISEQ-500 and HiSeq X Ten data. A summary of the somatic variants identified in 3 mesothelioma patient samples (patient ID: 9869, 11202 and 11398) using different sequencing platforms. The number of somatic SNV (a) and indels (b) identified using the BGISEQ-500 and HiSeq X Ten platforms in each patient. The somatic SNV (c) and indels (d) which were only called in one platform fall into three categories: i) identified as somatic in the other platform but with low evidence; ii) identified in the other platform but predicted as a germline variant; or iii) not identified in the other platform.

图2 利用BGISEQ-500和HiSeq X Ten的数据鉴定3个胸膜间皮瘤的体细胞突变

共156个体细胞突变 (141个SNV和15个indels) 位于基因编码区。其中, 在两个测序平台中共同鉴定出109个编码突变 (70%), 包括已知的间皮瘤驱动基因BAP1, 仅在 BGISEQ-500和HiSeq X Ten分别独有的突变为20和27个。

3、讨论与结论

使用BGISEQ-500和HiSeq X Ten测序平台对三个间皮瘤患者的癌症和对照的正常DNA进行了测序, 结果显示大部分突变均可以通过两个测序平台进行鉴定。

两个平台分别特有一小部分变异的差异是由多重因素导致的。一个关键因素是两个平台 (BGISEQ-500平台50 bp和HiSeq X Ten平台150 bp) 之间读长的差异。比对分析使用BWA-MEM, 是基于Burrows-Wheeler Aligner算法开发的, 此算法专门针对读长大于70bp而设计。另外一个可能导致突变结果较小不一致的因素是, 两个平台采用了相同的分析流程。分析流程要求至少35个碱基连续匹配, 且少于3个错配, 这对BGISEQ-500 PE50读长来说是不公平的。另外, 尽管为所有三个样品对提供了来自相同核酸提取的高分子量DNA, 等分试样到每个测序中心, 最大限度地减少样品质量差异导致不一致的可能性, 但在文库制备和测序过程中随机抽取DNA分子可能是数据中不一致的来源。

研究意义

这篇文章的BGISEQ-500数据是平台进行的首次公开可用的癌症基因组测序。尽管BGISEQ-500平台是早期PE50的数据, 但两平台生殖细胞突变基因型与SNP芯片分型结果高度一致 (>99%), 且在检测生殖细胞和体细胞SNV和插入缺失的能力方面都表现出很高的一致性。

文章4

基于DNB的测序平台可有效避免index hopping

Reliable Multiplex Sequencing with Rare Index Mis-Assignment on DNB-Based NGS Platform



研究对象: 人

合作单位: 华大基因、国家基因库

研究概要

使用三种主要的文库制备方法研究了DNB测序平台的Index hopping问题。BGISEQ测序仪利用独特的DNA纳米球 (DNB) 技术, 基于滚环复制 (RCR) 进行文库扩增, 这种线性扩增可以避免常规PCR带来的错误累积。基于DNB的NGS应用仅使用单个index就实现了前所未有的0.0001%至0.0004%低样本错误分配率。此外, 用水代替DNA, 加入index增加空白对照, DNB测序平台发生错误匹配的概率为36 million reads分之一, 即0.0000028%。

研究策略

取材: 8个基因的400bp片段构建WGS文库, HPV文库, 癌症文库

测序: BGISEQ-500 PE100、PE50

研究结果

1、DNA纳米球技术的高index保真度

BGISEQ平台将DNB加载到规则阵列 (patterned arrays) 上, 并利用组合引物锚定测序技术 (cPAS) 进行测序。独特的DNB技术采用具有强链置换活性的Phi29聚合酶和能够进行线性扩增的RCR工艺, 每个扩增循环都以原始的单链环状DNA文库为模板, 保持每个拷贝子的独立性 (图1a)。因此, 即使出现寡核苷酸的index hopping等错误, 也不会累积错误拷贝, 正确的序列总是会在后面的DNA拷贝中复制, 保证最高的扩增保真度。

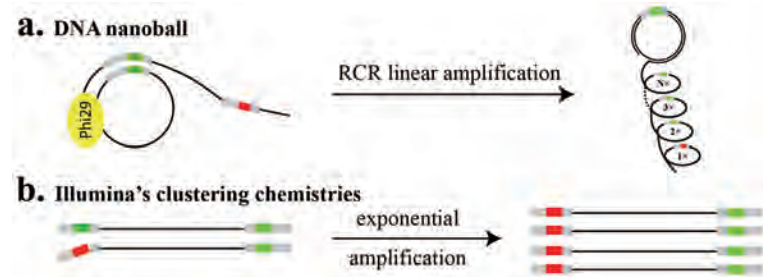


Figure 1: Mechanisms of index hopping on different sequencing platforms. (a) Sequencing using DNA nanoball technology is accomplished through Phi29 and RCR linear amplification; each copy is amplified independently using the same template ssCir. In this case, error reads from index hopping cannot accumulate, and most of the signal originates from correct indexes. (b) Bridge PCR or ExAmp chemistry utilizes exponential amplification, and index hopping can accumulate as amplification proceeds through each cycle, resulting in mis-assigned samples. Green, correct index; red, wrong index.

图1 Index hopping在不同的测序平台产生的机制

2、信号外溢造成极少的index hopping

Index污染可以通过实验操作、PCR错误、测序错误、寡核苷酸合成错误引入。为了检测BGISEQ平台极少index hopping的原因，我们首先找到那些出现错配的DNB，调查其测序质量。发现这些不匹配的DNB在index区域的平均Q30仅为36.66%，显著低于正确匹配的DNB的index区域（平均Q30=91.19%）。同时，统计了DNB在芯片上的位置，并计算了与其周围DNB共享相同index序列的概率。平均而言，无index hopping的DNB与相邻DNB共享相同index序列的概率为20.21%，存在index hopping的DNB共享相同index的概率则为57.04%。该结果表明相邻DNB的信号外溢造成了index hopping。而对于这种情况，index的测序质量通常偏低，我们可以通过Q30>60%来降低因此产生的index hopping。

3、Oligo合成污染为另一主要原因

还发现EFEMP2 / LOX被错误分配到index 7的 reads的index区域的测序质量值很高（平均Q30= 85.03%和82.38%），如表1所示。index 2和index 7之间的汉明距离是8，索引3和7之间的汉明距离是9，因此，EFEMP2 / LOX的index hopping也不太可能由随机测序错误引起的。

很可能的一个原因是，Index 7 oligo在合成或寡核苷酸处理过程中污染了其他寡核苷酸，导致在测序时候在其他的里面发现了index7。寡核苷酸合成污染可能是index错误分配的另一个主要原因，可通过使用IDT的TruGrade的oligo，有效避免此类index hopping的产生。

此外，对于PCR-free的文库，由于起始量高，需要加入更多浓度的index，因此相比PCR文库更容易造成去除不干净的情况，如图2所示：

除了常规 PCR 文库外，文中还对 PCR-free 文库在 BGISEQ 平台的 index hopping 情况进行调查，未经过任何 Q30 过滤的 99.9998%精度再次证实了 DNB 可以在很大程度上最小化 index 污染。与上面的常规 PCR 文库类似，污染率平均约为 0.0004%。

表1 BGISEQ平台的index hopping结果

Barcode	No Q30 Filter							
	NO TCH10P (Chr1)	EFEMP2P (Chr1)	LOX7P (Chr5)	V7P (Chr7)	S2F3-4P (Chr8)	S2F3-7P (Chr7)	MR104-2 (Chr8)	S2F3-25P (Chr2)
1	8281070	268	168	80	73	359	80	158
2	887	7981566	150	113	86	95	102	178
3	890	307	7481812	115	105	181	30	182
4	35	543	208	7257788	25	49	30	85
5	85	300	81	46	7211188	51	254	55
6	581	245	210	150	244	7811795	84	181
7	486	1746	906	307	311	451	8728404	567
8	175	252	178	105	105	311	104	6731644
Percent of read length	11.5184%	0.0004%	0.0003%	0.0001%	0.0001%	0.0002%	0.0002%	0.0003%
1	0.0001%	0.0005%	12.4021%	0.0003%	0.0002%	0.0002%	0.0001%	0.0001%
2	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%
3	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%
4	0.0001%	0.0001%	0.0001%	11.8408%	0.0001%	0.0001%	0.0001%	0.0001%
5	0.0001%	0.0001%	0.0001%	0.0001%	11.7914%	0.0001%	0.0001%	0.0001%
6	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	12.1899%	0.0001%	0.0001%
7	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	14.2402%	0.0001%
8	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	0.0001%	35.9641%

表2 PCR-free 文库index污染比率

Library Index	HBB	HPV11	HPV18	HPV31	HPV33	HPV32	HPV45
1	2,594,608	1,348,826	83	96	14	23	33
2	2,722,311	75	2,310,955	31	17	24	31
3	1,891,540	53	65	1,566,954	10	8	18
4	2,936,888	54	90	80	940,365	18	25
5	2,289,158	61	52	24	14	1,237,126	22
6	1,747,834	251,390	53	17	9	18	1,497,716
7	27	5	0	3	0	16	5
Percent of read length		14.7309%	0.0009%	0.0004%	0.0002%	0.0003%	0.0004%
2		0.0008%	25.2386%	0.0003%	0.0002%	0.0003%	0.0003%
3		0.0006%	0.0007%	17.1132%	0.0001%	0.0001%	0.0002%
4		0.0006%	0.0010%	0.0009%	10.2700%	0.0002%	0.0003%
5		0.0007%	0.0006%	0.0003%	0.0002%	13.5110%	0.0002%
6		2.7673%	0.0006%	0.0002%	0.0001%	0.0002%	16.3570%
7		0.0000%	0.0000%	0.0000%	0.0000%	0.0002%	0.0001%

最后，对目标区域捕获文库进行测试，BGI 肺癌试剂盒的数据验证了 DNB 测序平台上的 single index 不易发生 index hopping，可用于准确检测癌症等疾病的低频变异。结果显示在不同库类型中 BGISEQ index hopping 比率都非常低，表明 DNB 测序技术的数据准确性不受文库构建方法的影响。

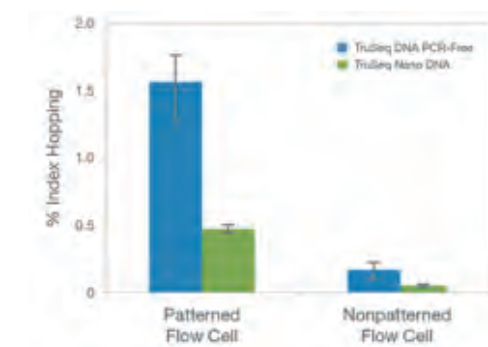


图2 PCR-free和PCR在两种测序芯片上index hopping比例的比较

全外显子测序 (WES)

外显子组测序是指利用序列捕获或者靶向技术将全基因组外显子区域DNA富集后再进行高通量测序的基因组分析方法。

在人类基因中大约有180,000外显子, 占人类基因组的1%, 约30MB。人类基因组的蛋白编码区大约包含85%的致病突变。外显子组测序主要用于识别和研究与疾病、种群进化相关的编码区及UTR区域内的结构变异。结合大量的公共数据库提供的外显子数据, 有利于更好地解释所得变异结构之间的关联和致病机理。

斯坦福大学医学院遗传学系的研究人员对市场上主流的外显子组测序平台进行了比较。该研究结果发表在《Nature Biotechnology》上。研究人员还比较了同一样品的外显子组测序和全基因组测序 (WGS), 显示外显子组测序能够检测到全基因组测序错过的小变异。

除了人的外显子测序, 华大还提供小鼠外显子测序服务。小鼠是人类疾病研究, 药物研发的重要模式生物。现在有了小鼠全外显子捕获试剂盒, 使得小鼠基因组研究更加深入, 成本大为降低。

猴子与人类具有十分重要的亲缘关系, 二者在基因水平和生理水平上较为相近, 所以灵长类常用于研究人类疾病和药物研发。它们主要用于药物临床前期研究, 包括发现新型疫苗、药物、行为研究等等。华大基因完成了恒河猴(Macaca mulatta)和猕猴(Macaca fascicularis)的全基因组的组装和注释。结合两种猴子基因组序列, 全面设计外显子芯片。猴外显子捕获芯片与高通量测序平台的高效结合, 不但能简化对猴基因组的研究流程, 还可使研究人员轻松捕获猴外显子, 靶向性地捕获所需基因组片段, 快速获得完整外显子区域信息, 从而加速科研新发现。

文章1

首篇BGISEQ外显子应用文章, 助力恶性卵巢癌多组学研究

C/EBP β enhances platinum resistance of ovarian cancer cells by reprogramming H3K79 methylation



发表期刊: *Nature Communications*
影响因子: 12.35
研究对象: 人
合作单位: 华中科技大学同济医学院

研究概要

化疗耐药是卵巢癌治疗的主要临床障碍。表观遗传学在调节恶性表型中起着关键性的作用, 并有潜力开发治疗有价值的靶点, 以改善这种疾病的不良后果。文章表明, 一系列的转录因子, 包括C/EBP β , GCM1和GATA1, 可以作为肿瘤细胞中组蛋白甲基化的潜在调节剂。值得注意的是, C/EBP β 是卵巢癌患者的独立预后因子, 它介导了一种重要的机制, 通过这种机制, 表观遗传学酶以相关的方式修饰功能相关基因组。通过甲基转移酶DOT1L, C/EBP β 可以通过多个耐药基因的H3K79甲基化维持开放的染色质状态, 从而增强肿瘤细胞的化疗耐药性。该论文提出了一种抗癌表观遗传学的新途径, 其中鉴定和靶向表观遗传学关键调节因子如C/EBP β 可能为卵巢癌提供更精确的治疗选择。

研究策略

取材: 20例高级别浆液卵巢癌 (HG-SOC), 20例良性疾病患者 (10例子宫肌瘤和10例子宫腺瘤)

测序: BGISEQ-500 WES PE100; Hiseq 2000/2500 ChIP-Seq SE50

分析: 通过BWA比对软件和GATK的HaplotypeCaller分别进行序列比对和变异检测

研究结果

1、组蛋白甲基化, 发现关键基因C/EBP β

在HG-SOC组织中, 研究人员应用ChIP-seq技术对H3K4、H3K36、H3K79 (转录激活相关) 和H3K9、H3K27、H4K20 (转录抑制相关) 六个甲基化位点进行基因层面的分析, 结果显示HG-SOC中许多基因的H3K79甲基化上调, 并在IHC实验中得到证实。RNA-Seq结果表明HG-SOC中上调的基因与H3K79高甲基化相关, 说明H3K79甲基化对于调控基因表达发挥重要作用。使用癌症基因组图谱 (TCGA) 数据集分析了转录因子的下游基因, 证明CEBPB不仅调节许多基因的表达, 而且调节HG-SOC中超过三分之一的H3K79相关上调基因的表达。接着, 利用生信软件HOMER, 研究人员找到与组蛋白甲基化相关的35个转录因子(TF), 其中CEBPB (编码C/EBP β 蛋白) 与H3K79甲基化上调成最显著相关性, 见图1。

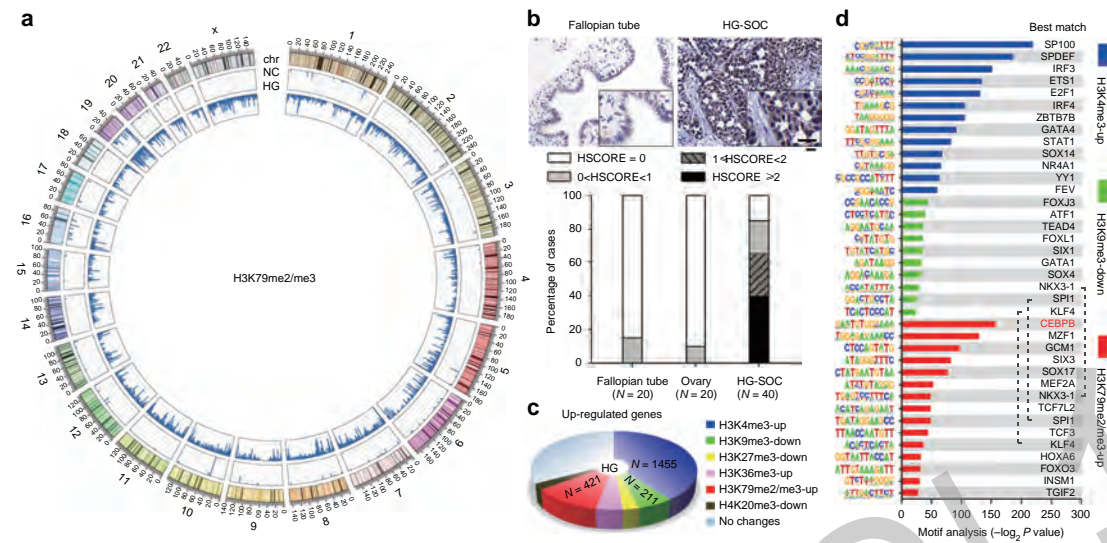


图1 HG-SOC表观遗传重编程转录因子

2、C/EBP β 高表达促进卵巢癌对顺铂耐药

卵巢癌患者在临床治疗中遇到的主要障碍是耐药性, 导致患者预后较差。而C/EBP β 在不良预后病人中高表达, 研究人员推测C/EBP β 与卵巢癌对顺铂耐药相关。

利用TCGA数据库和IHC实验, 对C/EBP β 和卵巢癌病人的顺铂耐药可能性进行评估, 结果显示高表达的C/EBP β 与顺铂耐药正相关。应用C/EBP β 基因过表达和敲降技术, 在细胞实验和动物实验同步证实: C/EBP β 过表达促进顺铂耐药, 而敲降C/EBP β 可以降低顺铂耐药。此外, 临床卵巢癌患者在顺铂治疗后, 肿瘤组织中的C/EBP β 蛋白表达水平也有明显上调, 如图2。

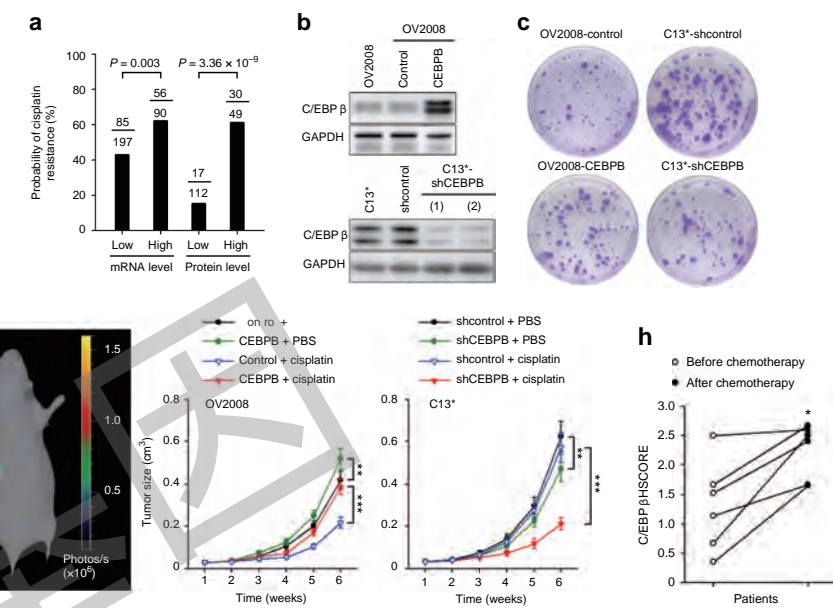


图2 C/EBP β 与顺铂抗性分析

3、C/EBP β 重新编程H3K79甲基化

ChIP-seq实验结果显示, 在两种不同的卵巢癌细胞系, 敲降C/EBP β 可以导致H3K79甲基化水平下调, 进而可致H3K79调控基因表达下调。RNA-Seq分析发现C/EBP β 敲降后, 1238个基因发生上调, 504个基因发生下调, 其中9个与顺铂抗性相关基因的表达水平伴随着H3K79甲基化的下调而降低。IPA分析显示这些基因主要参与到顺铂抗性相关的DNA损伤修复 (DDR) 信号和ERK5信号中, 如图3。

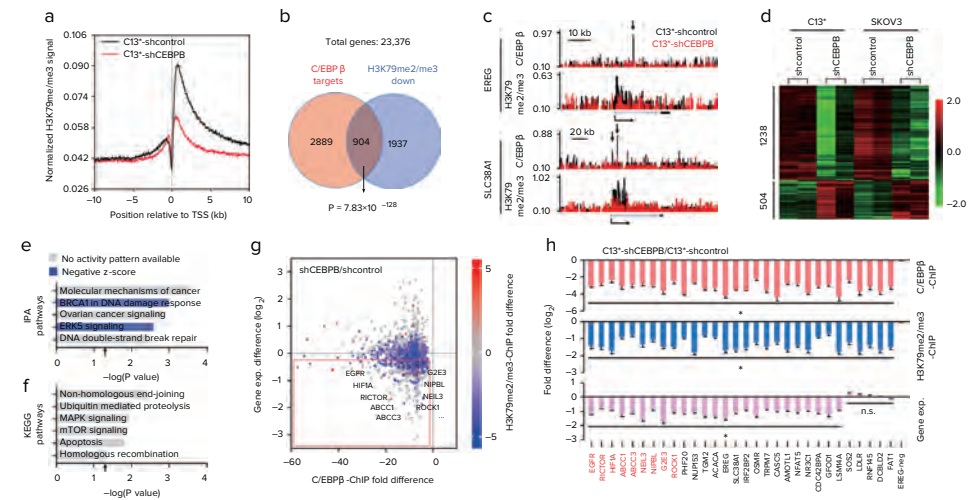
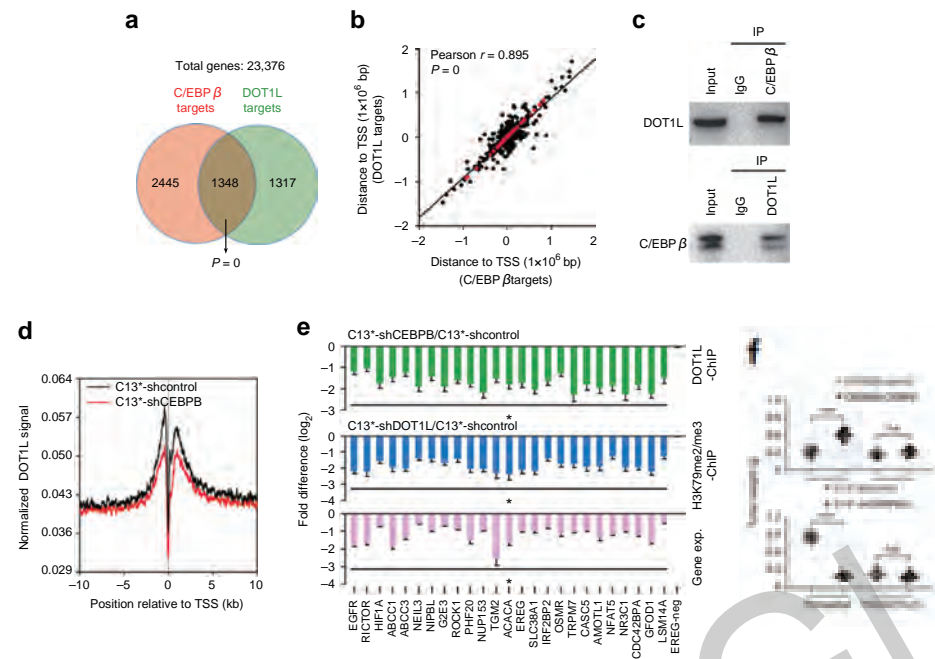


图3 C/EBP β 调节H3K79甲基化重编程基因表达

4、发现C/EBP β 调控H3K79甲基转移酶DOT1L

ChIP-seq实验结果显示, C/EBP β 靶向基因与DOT1L靶向基因具有显著相关性, 并且C/EBP β 结合位点与DOT1L结合位点距离接近, 两者有物理相互作用。经过ChIP-qPCR验证结果进一步证实, C/EBP β 通过增强DOT1L结合能力调节H3K79甲基化。应用DOT1L抑制剂或shRNA进行体内外抑制实验, 发现DOT1L抑制后, C/EBP β 高表达引起的顺铂耐药性也随之降低, 见图4。

图4 C/EBP β 通过DOT1L调节H3K79甲基化

研究总结

该研究采用了经典的甲基化研究思路, 找到了甲基转移酶复合物C/EBP β -DOT1L在卵巢癌顺铂耐药中发挥作用, 通过DNA测序、RNA测序和表观学结合的方法揭开了表观调控与肿瘤耐药之间关系, 找到了明确的耐药机制, 为甲基化调控深入研究提供了典范。

附表: WES文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth	<i>PLOS Genetics</i>	5.540	人	奥卢大学医院
2	A novel PTCH1 mutation underlies non-syndromic cleft lip and/or palate in a Han Chinese family	<i>oral diseases</i>	2.010	人	北京大学口腔医学院
3	Identification of a mutL-homolog 1 mutation via whole-exome sequencing in a Chinese family with Gardner syndrome	<i>Molecular Medicine Reports</i>	1.692	人	重庆大学肿瘤医院
4	A novel mutation in PRPF31, causative of autosomal dominant retinitis pigmentosa, using the BGISEQ-500 sequencer	<i>International journal of ophthalmology</i>	1.100	人	中国科学院大学
5	A new massively parallel nanoball sequencing platform for whole exome research	<i>bioRxiv</i>	--	人	James D. Watson 基因组科学研究所

宏基因组学 (Metagenomics)

宏基因组学 (Metagenomics) 又称为环境基因组学, 其研究对象是直接从中环境中获得的全部微小生物的遗传物质的总和。宏基因组鸟枪法测序 (Metagenomic shotgun sequencing) 作为微生物学领域继显微镜发明之后又一重大技术突破, 使得直接研究微生物群落成为可能, 无须分离培养即能发现大量微生物“暗物质”, 大大推动了包括人体肠道菌群在内的新兴微生物学领域的研究。华大自 2008 年作为唯一一个非欧盟单位参与欧盟 Meta-HIT 项目以来, 始终致力于以宏基因组为核心的多组学技术开发以及其在人体疾病等方面的应用, 取得了一系列标志性成果, 发表包括《Nature》封面文章在内的CNS论文19篇, 提交专利申请70余件, 开发产品婴幼儿肠道菌群检测服务1项。

宏基因组研究所依托深圳华大生命科学研究院高通量测序平台和大数据分析平台, 致力于人体共生微生物对人体健康影响的研究, 包括: (1)人体各个部位的共生微生物的群落结构以及与健康之间的关系, 如口腔、生殖道、皮肤、尤其是消化道; (2)人体肠道微生物与慢性复杂疾病发生发展相关的内容, 从“饮食-人体共生微生物-宿主代谢”的角度尝试寻找新的健康管理及疾病干预手段。

基于cPAS的BGISEQ-500平台对宏基因组测序的评估

文章1

Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing



华大基因
BGI

发表期刊: *GigaScience*

影响因子: 7.267

研究对象: 人

合作单位: 深圳华大生命科学研究院
宏基因组研究所转化应用实验团队

研究概要

首次公开使用BGISEQ-500对健康人群肠道样本进行深度宏基因组测序的数据, 通过一系列性能评估, 展示BGISEQ-500宏基因组测序应用性能的优越性、准确性、稳定性及跨平台一致性。

研究策略

取材: 20例健康成年人粪便DNA

测序: BGISEQ-500 SE100、Illumina HiSeq 2000、HiSeq 4000

研究结果

随机选择8例DNA样本作为技术重复, 产出8对文库构建重复与8对测序重复数据集以评估该测序平台的稳定性, 并完成全部20例DNA样本在Illumina HiSeq 2000 与 HiSeq 4000双平台的测序以评估跨平台数据的一致性 (图1)。

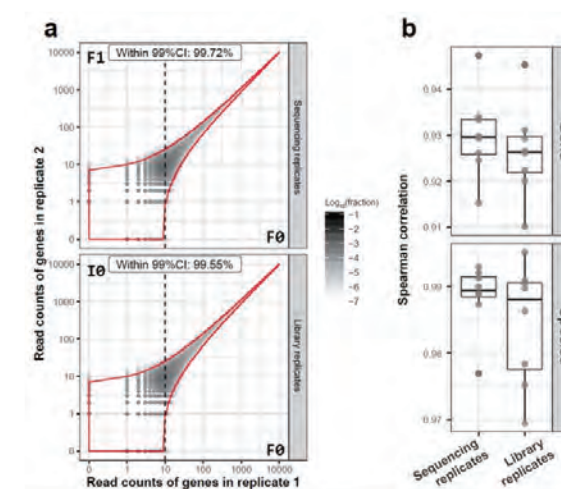


图1 BGISEQ-500平台内技术重复一致性

1、BGISEQ-500平台的高质量性

针对BGISEQ-500数据特征, 开发宏基因组数据质控流程, 结果显示, 质控后高质量数据占原始数据的96.06%, 其中Q30超过90.56%。

2、BGISEQ-500平台内高稳定性

基于文库重复数据集与测序重复数据集的评估显示: 定性层面, BGISEQ-500平台内的两组技术重复内可稳定检测的基因数高达99.5%以上; 定量层面, 两组数据间的基因相对丰度相关性可高达0.92以上, 物种相对丰度相关性可高达0.99(图2)。

3、BGISEQ-500与HiSeq 2000&4000的跨平台高一致性

BGISEQ-500与HiSeq 2000比较: 跨平台数据集可稳定检测的基因一致性高达91.8%, 对平台间基因丰度的统计检验仅检出3.25%显著差异基因, 两平台在物种丰度的相关性仍达94.8%(图3)。这些差异基因在GC含量上存在一定的平台特异性, 表现为高GC含量的基因在HiSeq 2000平台相对丰度偏高, 该发现在BGISEQ-500与HiSeq 4000的数据集比较中也得到了验证。

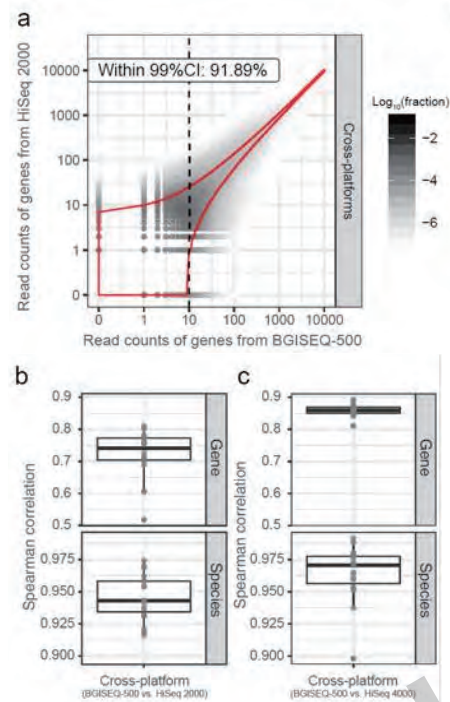


图2 BGISEQ-500与HiSeq 2000/HiSeq 4000平台间的一致性评价

来自Joslin diabetes center的Dr. Kostic及其团队也受邀参与了该数据集的评估工作, 证实了物种GC含量在BGISEQ-500与HiSeq 2000平台间、以及物种丰度之间的独立性, 表明极少数差异基因的GC偏向性对物种的定量并无显著影响。

研究意义

通过对BGISEQ-500在平台内、平台间的宏基因组定性和定量性能全面评估, 全面证实了BGISEQ-500平台对宏基因组应用的高度可行性, 其高通量的测序能力与更经济的测序成本也将为未来大规模的宏基因组研究提供可行的技术与平台支撑。

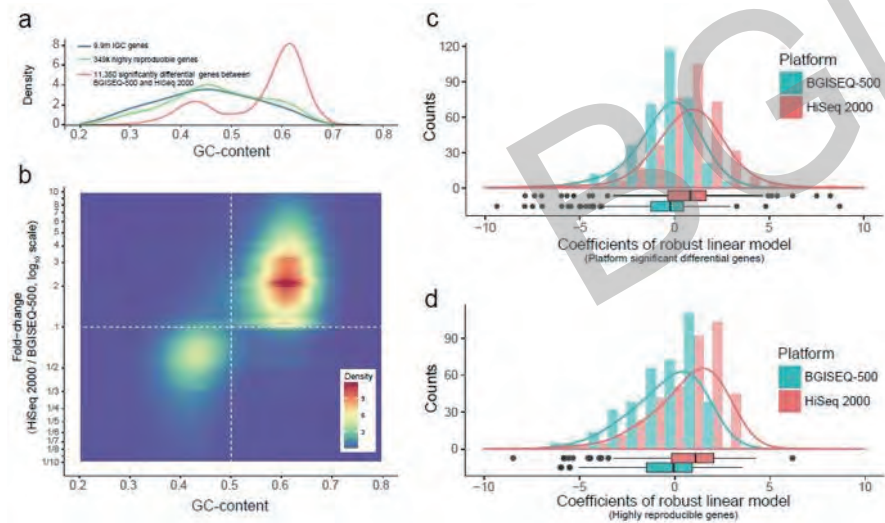


图3 BGISEQ-500与HiSeq 2000平台间差异基因的比较

文章2

基于BGISEQ平台完成全球首个大鼠肠道微生物参考基因集构建

A gene catalogue of the Sprague-Dawley rat gut metagenome



华大基因
BGI

发表期刊: *GigaScience*

影响因子: 7.267

研究对象: 大鼠

合作单位: 澳门科技大学

研究概要

全球首个大鼠肠道微生物参考基因集的构建工作完成, BGISEQ平台的宏基因组测序数据, 展示出良好的组装表现; 同时说明了大鼠适合作为研究人类疾病与肠道菌群之间关系的动物实验模型, 能够为将来的研究提供重要的参考数据。

研究策略

取材: 49只SD大鼠, 用完全弗氏佐剂诱导关节炎模型, 进行为期30天、5个时间点和7个实验组的药物干预实验, 取粪便样本

测序: BGISEQ SE100, PE50

研究结果

1、基于BGISEQ平台, 采用宏基因组鸟枪法测序技术, 对5个时间点的大鼠粪便样本进行单末端测序 (Single-end, SE100) 测序, 并挑选首尾两个时间点的大鼠粪便样本进行双末端测序 (Paired-end, PE50) 测序, 共得到245个单端测序数据和98个双端测序数据, 总计测序量2,512Gb; 并进行数据组装、基因预测, 去除冗余基因序列, 最终得到包含约510万个非冗余的SD大鼠肠道微生物参考基因集, 约为大鼠自身基因的200倍, 并且趋近于饱和。其中, 64.6%的基因可注释到物种门水平, 主要归属于厚壁菌门 (75.90%)、拟杆菌门 (10.83%) 和变形杆菌门 (6.77%); 26.7%的基因可注释到物种属水平, 主要为梭菌属 (8.74%)、拟杆菌属 (6.25%)、罗斯氏菌属 (4.75%)、瘤胃球菌属 (4.44%) 和毛螺菌属 (2.58%)。同时, 在基因功能水平上, 53.1%的基因可以注释到KEGG数据库。

2、该研究还将人类、小鼠和大鼠肠道微生物参考基因集进行了比较分析, 发现在大鼠肠道基因集与人肠道基因集的共有基因占比 (2.47%) 高于小鼠肠道与人肠道基的共有基因占比 (1.19%); 相对于基因, 三者肠道菌群在功能上更为相似; 同时, 人肠道菌群KOs (93.65%) 能够在大鼠肠道菌群中发现, 同样高于小鼠肠道菌群 (80.03%)。

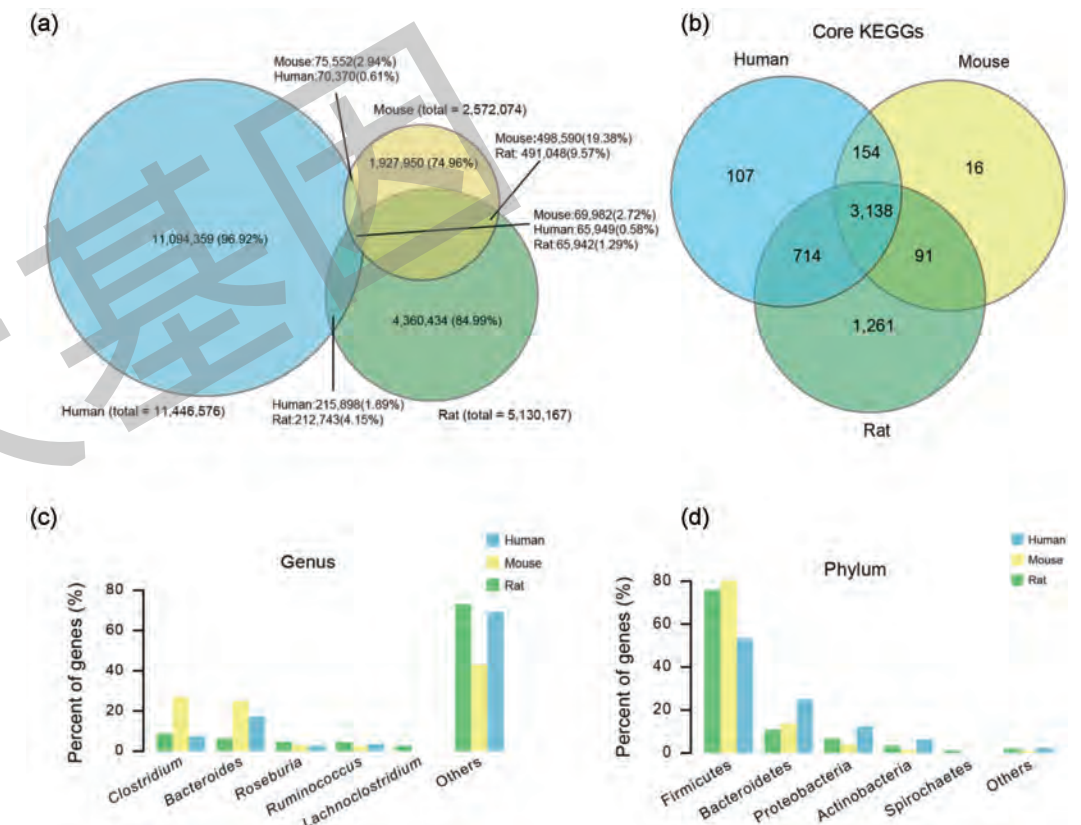


图1. 人、小鼠、大鼠肠道菌群基因集比较。

a, 非冗余基因 Venn 图; b, KO 功能 Venn 图; c, 三个基因集中各个属基因占比; d, 三个基因集中各个门基因占比。

附表: Meta文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	A novel affordable reagent for room temperature storage and transport of fecal samples for metagenomic analyses	<i>Microbiome</i>	8.496	肠道微生物	BGI、哥本哈根大学
2	The metagenome of the female upper reproductive tract.	<i>Gigascience.</i>	7.267	生殖道微生物	中国科学院大学
3	Efficacy and safety of faecal microbiota transplantation in patients with psoriatic arthritis: protocol for a 6-month, double-blind, randomised, placebo-controlled trial	<i>BMJ OPEN</i>	2.369	肠道微生物	欧登塞大学医院
4	Identification of pathogens in culture-negative infective endocarditis with metagenomic analysis	<i>bioRxiv</i>	--	人-感染性心内膜炎	国家心血管病中心

单细胞测序

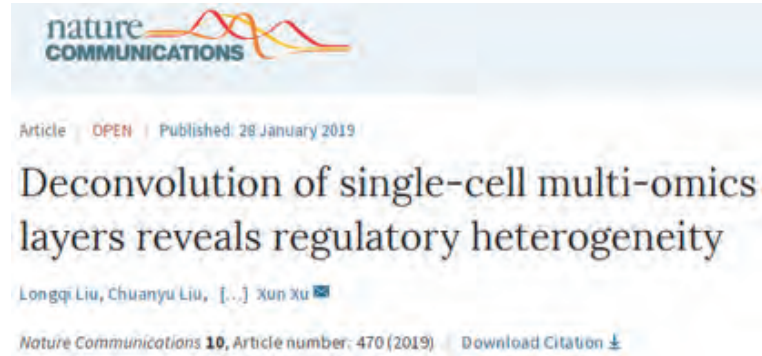
单细胞测序是对单独一个细胞或少数细胞的DNA或RNA进行测序,区别于常规群体测序,单细胞测序对样本需求量极少,能体现细胞的异质性。随着生命科学和医学基础研究的深入发展,人们发现越来越多的特殊标本和特定的生物学现象,如法医鉴定的微量标本、肿瘤内部异质性等无法用常规组织测序的方法进行研究。单细胞测序技术的出现,给这类样本的研究带来了极大的便利。

单细胞测序包括单管单细胞RNA测序、高通量单细胞RNA-Seq;单细胞DNA测序产品、高通量单细胞免疫组库测序、单细胞重测序、单细胞全基因组甲基化测序等。单细胞测序有助于研究基因调控网络和基因表达的异质性与随机性,结合系统生物学的方法,能应用到肿瘤微环境研究、疾病发展与进程监控、持续追踪肿瘤相关基因的动态表达,还可以研究胚胎不同发育阶段的组织器官图谱等。

文章1

单细胞多组学研究揭示了细胞调节异质性

Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity



发表期刊: *Nature Communications*

影响因子: 12.353

研究对象: 人

合作单位: 深圳国家基因库、深圳华大生命科学研究院、哈尔滨工业大学深圳研究院、德国肿瘤研究中心等

研究概要

单细胞转录组、全基因组或者外显子组等, 是研究单细胞核酸序列层面的信息; 单细胞ATAC、Chip-seq或者Hi-C等, 则是研究表观调控层面的信息。单细胞CAT-Seq则是一个把单细胞ATAC测序和单细胞转录组测序结合起来, 能够联合分析单个细胞染色质可接近区域与基因转录表达的新技术。

研究策略

取材: 三种细胞模型, K562慢性粒细胞白血病细胞、HCT116大肠癌细胞、Hela-S3宫颈癌细胞; 构造图谱时使用了植入前的胚胎细胞

测序: BGISEQ PE100 scATAC-Seq、单细胞转录组

研究结果

1、单细胞染色质可接近区域和基因表达的联合分析

对192个K562慢性粒细胞白血病细胞进行scCAT-Seq测序, 得到CA和GE两部分信息。通过对比发现此技术得到有效CA和GE数据量与已公布的scATAC-Seq或scRNA-Seq数据量相当。同时, 研究发现, CA集中于组蛋白修饰较活跃的顺势作用因子 (CREs) 区域, 在修饰少或不修饰区域片段少 (图1.d)。在组蛋白修饰较活跃的区域, 作者联合CA和GE的分析发现, 与H3K36me3关联的CA低水平区域有较高的GE; 而在转录因子结合位点研究GE与CA之间的关联时, 发现与EZH2结合的区域基因表达水平较低, 因此通过scCAT-Seq能够证实一些已知的生物学过程, 评估不同染色质区域的基因表达活性。

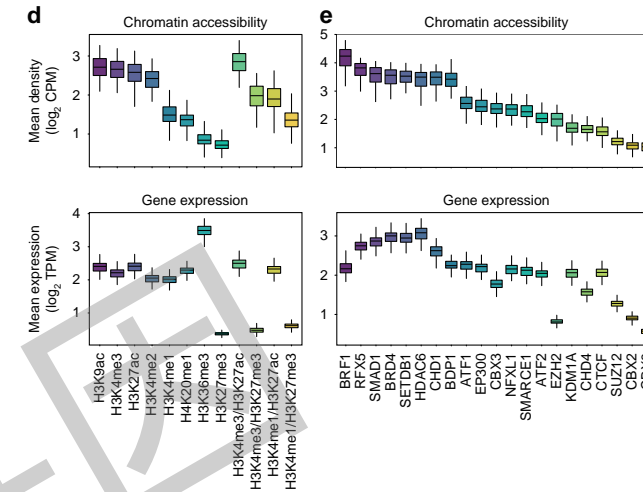


图1. 特定区域的染色质可接近区域与平均基因表达水平图

2、建立单个细胞CREs与基因表达调控的新模型

对三种细胞类型 (K562、HeLa-S3、HCT116) 的两个组学层面进行动态关联分析。发现CA和GE关联分析比两个CREs之间的关联分析更好地反应染色体高级结构特点; 通过构建以单个细胞为纵坐标, 染色质可接近区域和基因表达的调控关系为横坐标的矩阵图, 而矩形表示每个细胞调控关系的进行或者沉默状态, 然后运用NMF方法在矩阵中构建了单个细胞身份识别信号, 能够不同细胞类型 (K562、HeLa-S3、HCT116) 中同时反应出多个单细胞的表达调控特征, 以及组织之间的表达调节异质性。

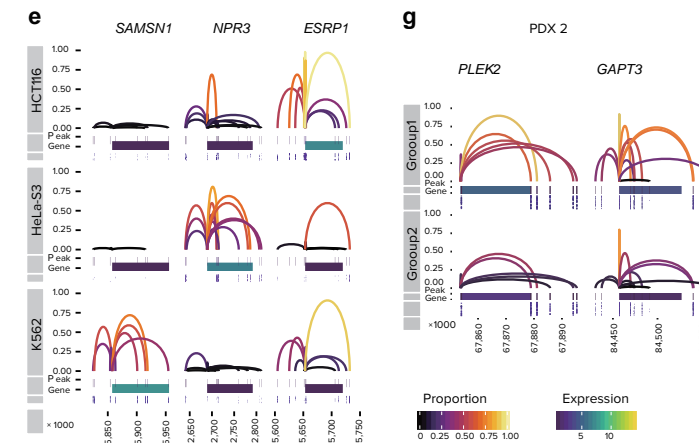


图2. 不同细胞类型中单个细胞特定基因的表达调控关系

3、早期胚胎单个细胞表观组学和转录组联合图谱的绘制

对早期胚胎中110个不同发育阶段的细胞进行scCAT-Seq测序，通过将染色质可接近区域TF和不同细胞CA水平构建一个ChromVAR，计算所有单细胞中已知TF的偏差。通过这种方式细胞被分成了明显的两个亚群，并发现了转录因子GATA3在决定滋养层细胞分化中的重要意义以及其他转录因子在决定细胞发育中的作用。这些结论说明了研究所构建的分析方法有助于在单细胞分辨率下发现一些TF的重要功能，同时发现scCAT-Seq对于在单细胞分辨率下发现复杂细胞类群中一些特殊细胞有突出作用。

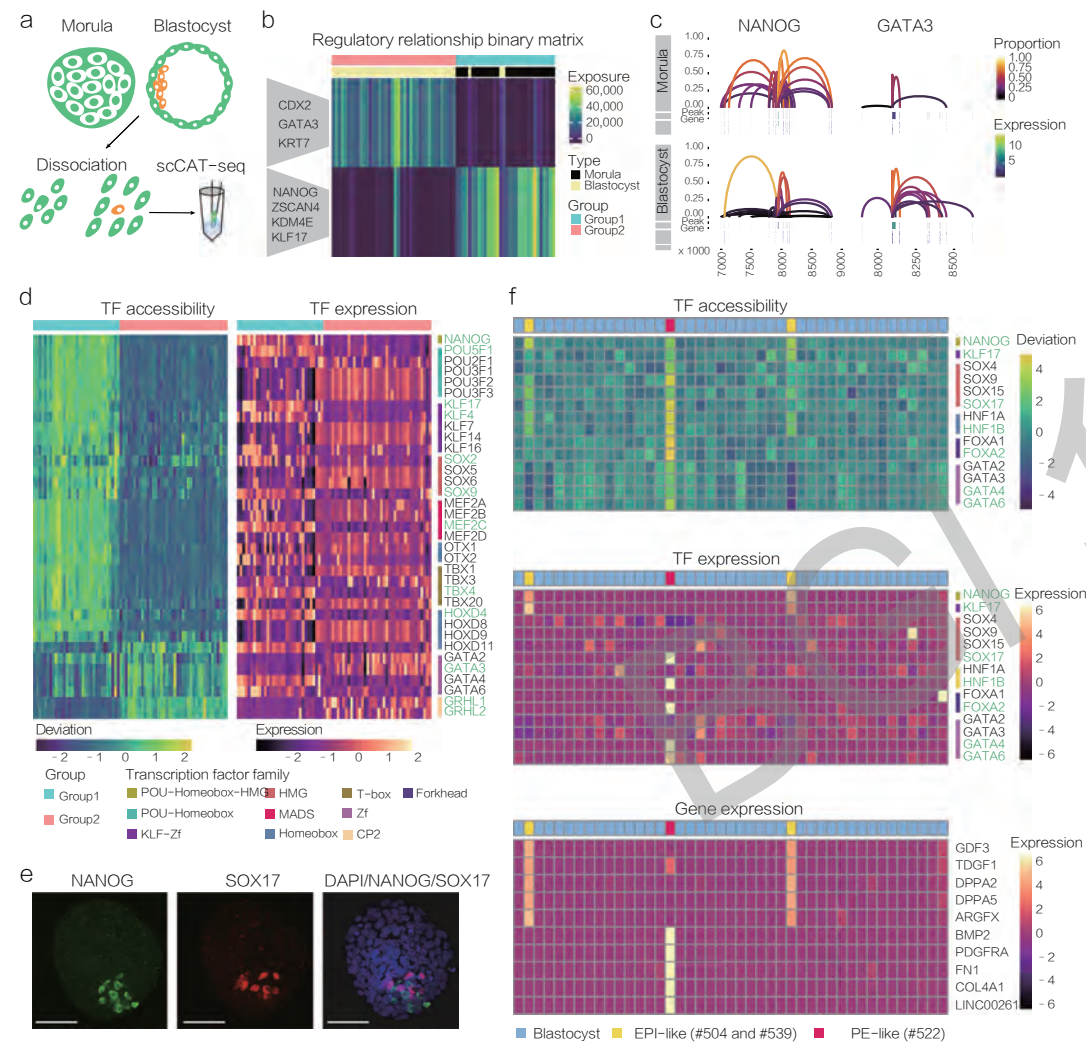


图3. 基于不同发育阶段的胚胎细胞scCAT-Seq数据构建图谱

文章总结

单细胞基因活性分析技术被《科学》杂志评为2018年十大科学突破之首，单细胞RNA-Seq被科学家认为会在未来10年改变基础生物学和医学的研究格局，并且单细胞多组学联合研究是2018年单细胞研究的重大进展之一。

本文介绍的scCAT-Seq能够对于单个细胞的表达调控特质进行综合表观组学和转录组的解决方法。scCAT-Seq的主要特点是，能同时测定单个细胞的染色质开放区域与转录谱，并且通过分析二者之间的关联性重塑了单细胞中顺势调控元件与靶基因之间的关系，这无疑给研究细胞之间的调控模式差异提供了一个新的思路。同时，本文中通过对早期胚胎scCAT-Seq测序的实践，绘制出了第一个早期胚胎单细胞染色质开放区域与转录谱的关联图谱，这进一步证实了scCAT-Seq在精确分析细胞中特定区域转录因子的重要作用。这两个层面的信息也给了“单个细胞”更确切的定义，同时使得细胞间调控异质性在复杂的细胞类群中被呈现出来。

附表：单细胞测序文章列表

序号	文献名称	发表期刊	影响因子	研究对象	合作单位
1	Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity	<i>Nature Communications</i>	12.353	人	深圳国家基因库、深圳华大生命科学研究院、哈尔滨工业大学深圳研究院、德国肿瘤研究中心等
2	Comparative analysis of sequencing technologies platforms for single-cell transcriptomics	<i>bioRxiv</i>	-	老鼠胚胎干细胞和人类K562细胞	西北大学

华大基因介绍

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

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

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

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

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

技术平台

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



深圳平台



天津平台



杭州平台



武汉平台

质量管理

华大基因于2010年通过了ISO 9001质量管理体系认证, 2011年12月通过了ISO 14001环境管理体系认证和OHSAS18001职业健康安全管理体系认证, 2012年6月通过了ISO/IEC 27001信息管理体系认证, 2015年7月华大基因香港高通量测序实验室获得美国病理学会 (College of American Pathologists, CAP) 颁发的CAP认可证书, 成为中国首家在高通量测序服务和基因检测行业具有CAP认可资质的医学实验室。

