

多组学技术助力构建肺癌诊断模型

华大智造DNBSEQ测序技术辅助肺癌诊断多组学联合模型的开发

以北京大学医学院为主的研究团队基于华大智造MGISEQ-2000RS基因测序仪进行高通量测序并结合多组学技术对肺癌及良性肺结节的液体活检标志物进行探究，揭示多组学液体活检标志物在肺癌辅助诊断及预后评估上的应用价值。

相关成果于2021年以题为“Non-invasive lung cancer diagnosis and prognosis based on multi-analyte liquid biopsy”发表于*Molecular Cancer*¹。

推荐应用：肿瘤组学(肺癌)

推荐机型：MGISEQ-2000RS

• 测序数据质量高

DNBSEQ测序技术具有高准确性、低重复序列率、低标签跳跃等的重要特性，可为肿瘤研究提供准确的测序数据。

• 多组学模型诊断效果更佳

相比单一指标或不含CEA的模型，cfDNA的基因学和表观遗传学特征，与血清蛋白标记物CEA的联合模型对恶性与良性病例的分类能力最好。



背景介绍

肺癌是我国发病率和致死率最高的恶性肿瘤²，导致肺癌发生的风险因素包括吸烟、遗传、不良饮食、空气污染和职业接触，如石棉、金属和混合职业接触、二氧化硅、多环芳烃和柴油废气。早诊断³、早治疗⁴、早监测⁵是提高病人生存率、降低社会家庭经济负担的关键。目前肺癌筛查应用最广泛的方法是低剂量计算机断层扫描 (Computed tomography, CT)，但其仍有一定的局限性，不仅会带来辐射暴露，且在临床上区分良恶性结节的作用有限，容易出现假阳性。而肺癌的预后指标临床上主要依赖肺癌病理 TNM 分期，但同一分期内的患者预后往往差异较大，仍有必要寻找额外的预后参数。

肿瘤驱动基因突变状态和表达特征等分子改变与肺癌 (Lung cancer, LC) 预后密切相关；同时，已有新的证据支持表观遗传改变的预后价值，这仍有待阐明。因此，适用于肺癌诊断及预后评估的新型液体活检标志物具有重要的临床价值，已成为近年来的研究热点。与使用组织活检的传统癌症诊断相比，液体活检更可行、侵入性更小，并且比组织活检更全面地评估肿瘤异质性⁶，因为所有肿瘤部位都会将循环肿瘤 DNA (Circulating tumor DNA ,ctDNA) 释放到血液中。

研究描述

癌症患者血浆中的 ctDNA 为癌症基因组提供了有价值的信息，也为非侵入性癌症检测提供了很大的希望^{7,8}。然而，由于 ctDNA 被大量非癌源的循环细胞游离 DNA (cell-free DNA, cfDNA) 稀释，其检测十分具有挑战，特别是在肿瘤肿块很小的早期癌症阶段。在本研究中，研究者开发了一套实验和计算工具，通过高通量测序，测量 LC 患者和良性肺结节 (Benign lung nodule, BLN) 患者血浆 cfDNA 的遗传和表观遗传信号，结合血清蛋白标记物 CEA 建立联合模型，旨在探索基于血液的生物标志物在 LC 诊断和预后中的潜在应用价值⁹。

实验方法和数据分析

样本采集

本研究收集了128例肺癌患者和94例良性肺结节患者的术前血液样本及术中肺癌组织及癌旁正常组织样本，并进行DNA提取。同时考虑白细胞基因组 DNA (WBC gDNA)污染，对白细胞基因组DNA进行提取。

文库制备及测序

a.探针panel设计

基于TCGA数据库和COSMIC数据库设计一个包含139个基因的全癌panel以用于后续的靶向高深度测序。

b.WGS文库的靶向测序

为减少PCR或测序产生的影响，制备获得的DNA样本采用双端唯一分子标识符 (UMI) 策略构建目标超高深靶向测序文库(NGS)。

cfDNA、肺癌组织、癌旁正常组织以及 WBC gDNA 均采用此策略进行文库构建。构建好的WGS文库利用IDT公司的xGen® Lockdown® 试剂进行靶向捕获实验，具体操作可参考相关说明书。所获文库在MGISEQ-2000测序仪上完成双端100个碱基的测序 (PE100)。

c. WGBS文库的靶向测序

同时为检测肺癌特异性表观遗传学的变化，采用单链DNA文库制备策略对cfDNA、肺癌组织、癌旁正常组织以及 WBC gDNA 构建全基因组亚硫酸氢盐测序文库(WGBS)。构建好的WGBS文库利用罗氏的SeqCap Epi CpGiant探针进行靶向捕获实验，具体操作可参考相关说明书。所获文库在MGISEQ-2000测序仪上完成双端100个碱基的测序 (PE100)。

数据分析

对MGISEQ-2000下机数据进行处理。使用FASTQ对Reads进行UMI裁剪及低质量Reads进行过滤后，与人类基因组Hg19进行比对并识别热点突变，对基于TCGA和COSMIC数据库选择的139个癌症驱动基因的外显子突变情况进行分析，通过突变评分确定良恶性肺结节的突变差异。此外通过WGBS靶向测序，识别并检测肺癌组织及癌旁组织的315个肺癌特异性甲基化位点(GpC)。随后，对cfDNA的全基因组甲基化测序结果分析覆盖560万个GpC位点，计算每个cfDNA的区域甲基化比以确定甲基化标志物的有效性。最后根据机器学习方法构建分类模型，将血清蛋白标志物与甲基化特性等结合，筛选获得可用于良恶性肺结节诊断的甲基化位点和血清蛋白标志物。

样品收集与饲养	文库制备和测序	生信分析	结果分析
128例肺癌患者和94例良性肺结节患者的术前血液样本，术中肺癌组织，癌旁正常组织样本和白细胞基因组DNA	<div> WGS和WGBS靶向测序文库的构建</div> <div> MGISEQ-2000 测序仪</div>	FASTQ SOAPnuke-2.0.3 BWA-MEM BitMapperBS MethylDackel	良恶性肿瘤的突变差异 肺癌特异性甲基化位点 甲基化标志物的有效性 机器学习方法构建分类模型

结果

靶向超深测序检测血浆 cfDNA 和 WBC gDNA 的不同突变谱

128 例 LC 患者代表了自然的肿瘤分期分布 (66% 的病例为 0 期或 I 期), 94 例 BLN 患者入选本研究 (图 1a), 对来自同一个体的 cfDNA 及白细胞基因组 DNA 进行了基于双端分子标签 (UMI) 的超高深度肿瘤驱动基因大 panel 靶向测序 (平均去重深度达 5000X)。结果表明, 在 cfDNA 及白细胞同时测到的突变具有高度的相关性 (图 1b), 提示这些突变来源于白细胞基因组 DNA, 因而应该在后续的分析中进行过滤。

过滤后结果显示, 在 111 个 LC 样本中检测到 67 个 (60.36%) 样本中共 153 个突变 (图 1c)。在 78 个 BLN 样本中检测到 23 个 (29.49%) 样本中共 28 个突变 (图 1e)。在良性肺结节病人的 cfDNA 中也检测到了来自肿瘤驱动基因的突变。虽然这些突变的频率相对较低, 且突变图谱和肺癌病人存在一定差异 (图 1d)。

基于体细胞突变的分类模型区分 LC 和 BLN

建立基于检测到的突变来区分 LC 和 BLN 的预测模型, SUMAF 模型的 AUC 为 0.67, 敏感性为 55.9%, 特异性为 76.9%; weighted_SUMAF 模型的 AUC 为 0.68, 敏感性 59.5%, 特异性 71.8% (图 1f)。这些结果表明, 仅基于突变评分建立的分类模型对 LC 和 BLN 血浆的分类能力有限。BLN cfDNA 携带的癌症驱动基因的基因组序列改变可能比以前认为的更普遍, 因此限制了基于突变的诊断分析的效用。多分析方法更有可能提高癌症信号的检测。

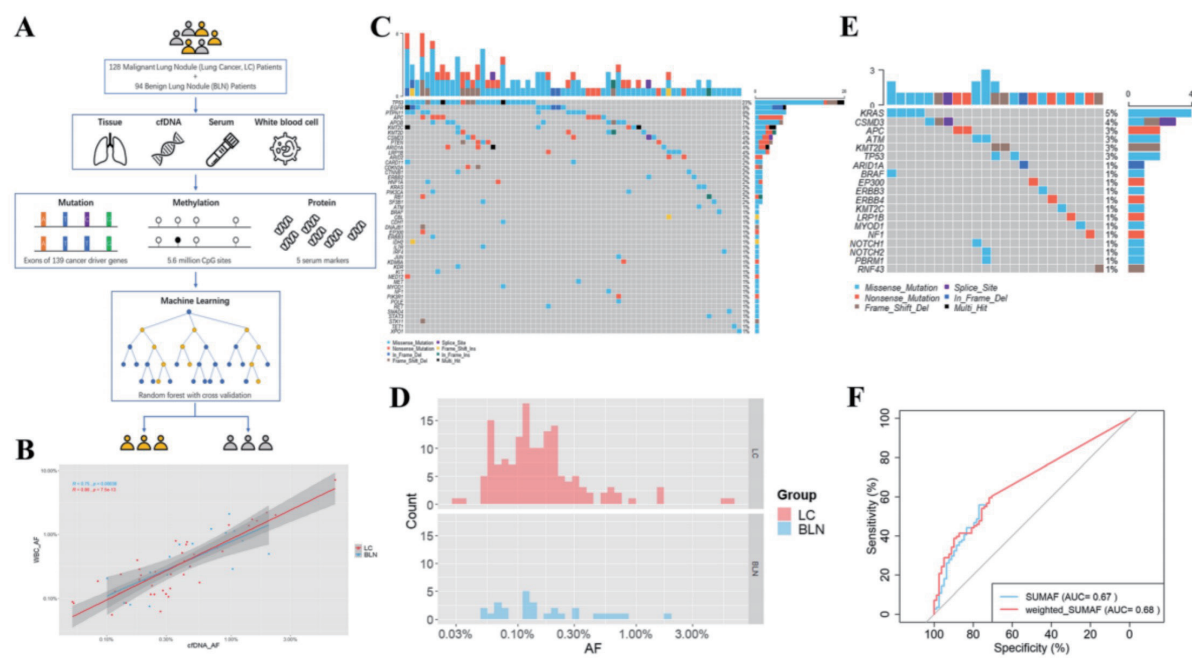


图1.血浆cfDNA中变异的研究设计和分类模型, 通过匹配的WBC样本筛选共享变异

基于 cfDNA 甲基化数据对 LC 和 BLN 血浆进行分类

为了识别肺癌特有的表观组学变异，研究者首先对肺癌及癌旁组织进行全基因组甲基化测序（图 2a），共识别出了 315 个肺癌特异性的差异甲基化区域（图 2b），其中包含 293 个高甲基化区域及 22 个低甲基化区域。GO（Gene Ontology）分析表明这 293 个高甲基化区域多位于转录调控区域（图 2c）。随后，研究者对来自肺癌病人及肺良性结节病人的 cfDNA 进行了高深度甲基化靶向测序，并用机器学习方法筛选出了可用于肺部良恶性结节辨别的甲基化标志。

多组学分析鉴别 LC 和 BLN 血浆

同时还检测了受检者血浆中 CEA，CYFRA21-1，NSE，CA19-9 和 CA125 这五种蛋白标志物的含量，LC 患者中只有 CEA 水平明显高于 BLN 患者 ($p = 0.04$, Student's t 检验)，分级 AUC 为 0.66（图 3、4），故将其纳入诊断模型中。结果表明，在同一组样本中，与不含 CEA 的模型相比，基于 wSUMAF 突变评分、54 个选定 DMRs 的区域甲基化率和血清 CEA 水平的多组预测模型可以更加有效区分肺癌病人及肺良性结节病人 ($AUC = 0.78$ ，图 2d； $AUC = 0.74$ ，图 2e)，基于 cfDNA 突变及甲基化标志物构建的肺癌预后模型效能也优于单一组学模型（图 2f、g； $AUC = 0.74$ ，图 2h）。

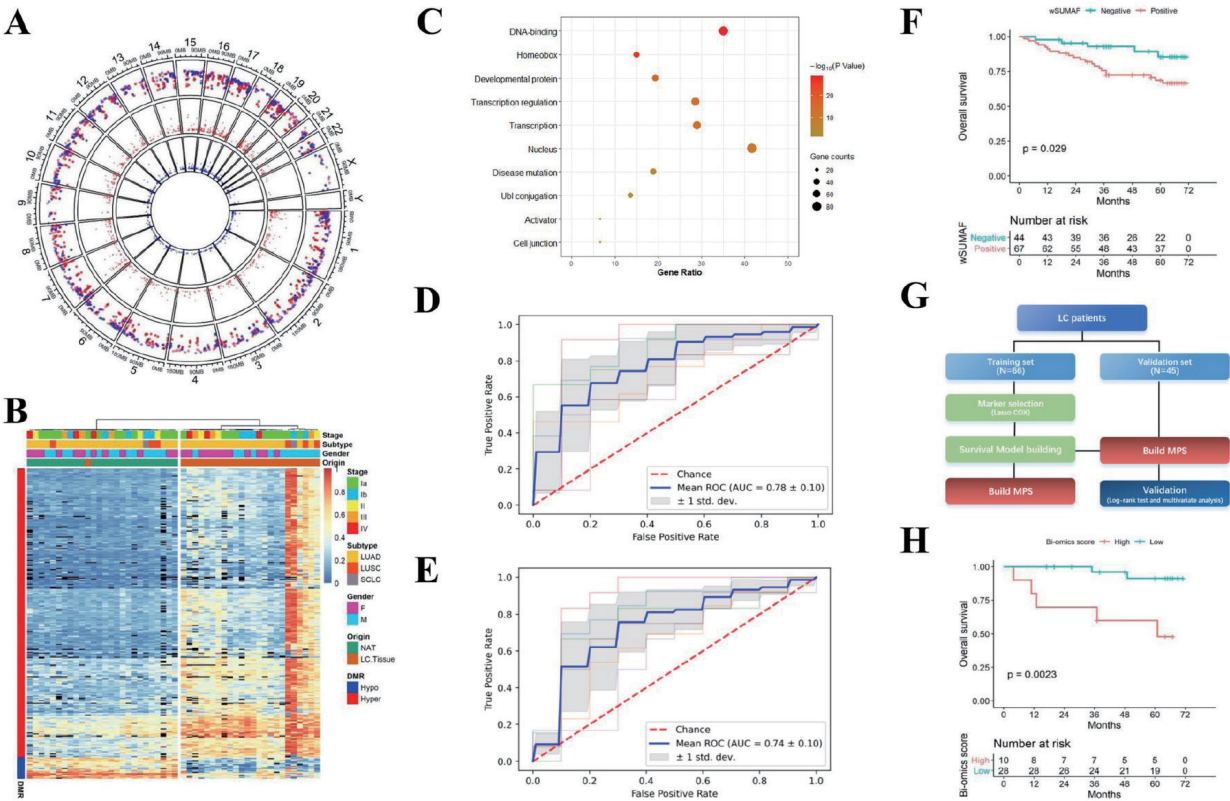


图2.DNA甲基化和多组学分析

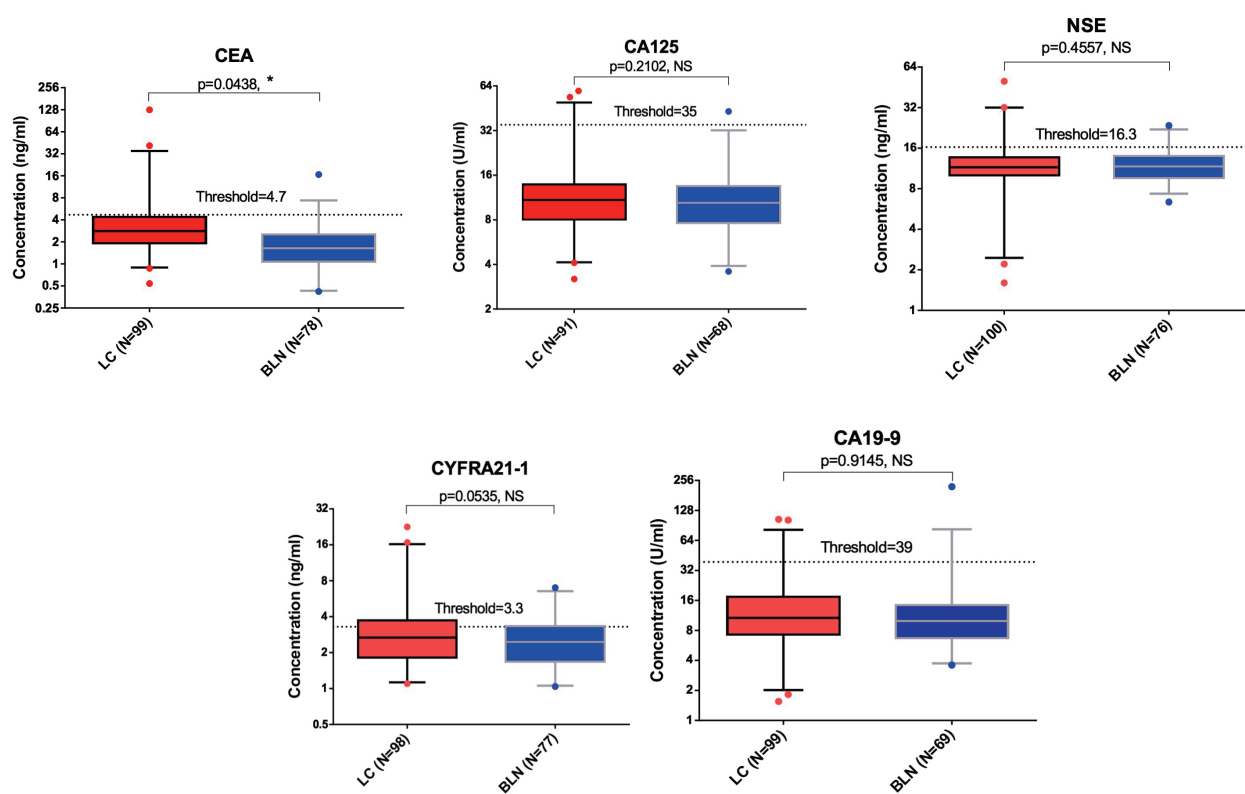


图3. BLN患者5种血清蛋白标志物水平的比较(T检验)。虚线表示临床常用的每个标记的阈值

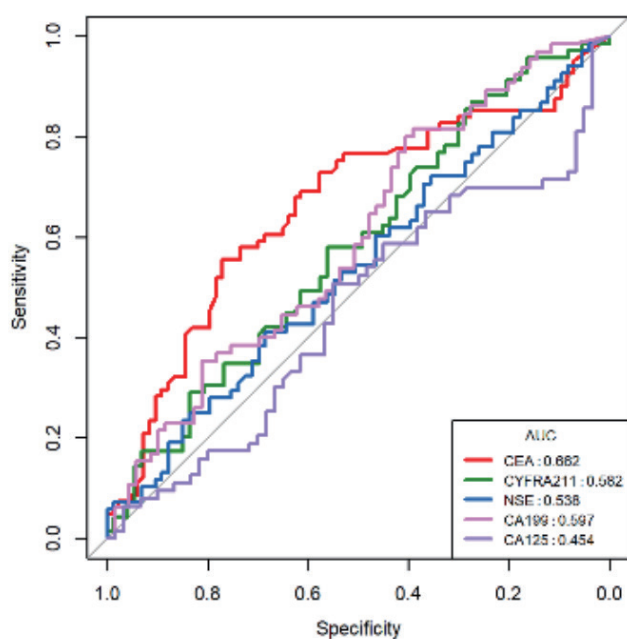


图4. 基于单一血清蛋白标志物区分LC和BLN血浆cfDNA的预测模型

总结

cfDNA 的基因学和表观遗传学特征，与血清蛋白标记物 CEA 的联合模型对恶性与良性病例的分类能力最好。此外，结合 cfDNA 突变状态和基于甲基化的预后标志物的集成模型有可能提高肺癌生存率的预测。

该结果是通过使用华大智造基因测序仪 MGISEQ-2000RS 实现的，研究者构建目标超高深靶向测序文库，对 cfDNA 和 WBC gDNA 的基因组序列改变进行检测。同时对肺癌组织和癌旁正常组织进行 WGBS，确定肺癌特异性表观遗传学改变。制备获得的文库进行扩增后在 MGISEQ-2000 测序仪上完成 PE100 测序，经数据分析后得出结论。



基因测序仪 MGISEQ-2000RS

参考文献

1. Chen, K. et al. Non-invasive lung cancer diagnosis and prognosis based on multi-analyte liquid biopsy. *Mol Cancer* 20, 23, doi:10.1186/s12943-021-01323-9 (2021).
2. Zhang, S., Sun, K., Zheng, R., Zeng, H. & He, J. Cancer incidence and mortality in China, 2015. *Journal of the National Cancer Center* (2020).
3. Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C. & Boffetta, P. Risk factors for lung cancer worldwide. *European Respiratory Journal*, 889 (2016).
4. Nooreldeen, R. & Bach, H. Current and future development in lung cancer diagnosis. *International Journal of Molecular Sciences* 22, 8661 (2021).
5. Cipriano, L. E. et al. Lung cancer treatment costs, including patient responsibility, by disease stage and treatment modality, 1992 to 2003. *Value in Health* 14, 41-52 (2011).
6. Siena, S. et al. Dynamic Molecular Analysis and Clinical Correlates of Tumor Evolution Within a Phase 2 Trial of Panitumumab-Based Therapy in Metastatic Colorectal Cancer. *Annals of Oncology Official Journal of the European Society for Medical Oncology* (2017).
7. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 359, eaar3247 (2018).
8. Lennon, A. M., Buchanan, A. H., Kinde, I., Warren, A. & Papadopoulos, N. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* 369, eabb9601 (2020).
9. Phallen, J. et al. Direct detection of early-stage cancers using circulating tumor DNA. *American Association for the Advancement of Science* (2017).

推荐订购信息

产品类型	产品名称	产品货号
仪器	基因测序仪 MGISEQ-2000RS	900-000035-00
	MGISP-100RS 自动化样本制备系统	900-000070-00
	MGISP-960RS 自动化样本制备系统	900-000100-00
软件	MegaBOLT 生信分析加速器(工作站式服务器)	970-000085-00
建库试剂	MGIEasy 全基因组甲基化文库制备试剂盒(16 RXN)	1000005251
测序试剂	MGISEQ-2000RS 高通量测序试剂套装(FCL PE100)	1000012554

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋



4000-688-114



www.mgi-tech.com



MGI-service@mgi-tech.com

股票简称：华大智造

股票代码：688114



仅供研究使用

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

版本：2023年2月版

撰稿：方嘉宜 杨骞

责任编辑：王其伟

审稿：江遥