

# 泰山学者特聘专家工作计划书

## 一、背景概述

（主要包括所属研究领域、国内外发展水平、申报人选及团队在国内外本学科领域所处的地位或具有的影响）

### （一）所属研究领域及国内外发展现状

#### 1、研究领域

申请人研究领域为肿瘤的表观遗传学以及与肿瘤表观遗传学相关的早期诊断、监测和预后的生物标记物的研发。

#### 2、国内外发展现状及肿瘤表观遗传学研究领域的发展趋势

##### 2.1 胞嘧啶修饰与复杂疾病的关联及研究趋势：

表观遗传修饰（例如 DNA 序列中胞嘧啶的甲基化修饰、组蛋白修饰以及 microRNA 等）通过调控基因表达，对于包含癌症在内的人类复杂疾病的发生、发展具有重要的生物学作用。由于胞嘧啶修饰在基因组中广泛分布，调控基因的表达，比其他表观遗传标记如 microRNA 等更稳定，因此，本研究将以胞嘧啶的甲基化及羟甲基化修饰为主要研究对象。相对于 DNA 序列本身的改变（体突变）而言，胞嘧啶修饰在分子进化层面所受的选择压力较小，更有利于成为癌症发生、发展的早期标志物。因此，本计划的主要目的就是利用最新的检测技术开发基于胞嘧啶修饰的用于癌症早期诊断、监测以及预后的生物标记物。

过去十几年，国内外对表观遗传学的研究基本上集中在 5-mC (5-methylation-胞嘧啶)，并且广泛证明 CpG 位点上的胞嘧啶甲基化修饰水平可以影响基因表达并与包括癌症在内的复杂疾病（或其他复杂性状）的发生、发展以及治疗反应等临床表现相关。值得注意的是，近年来一些研究“再次”发现 5-hmC (5-hydroxymethylation-胞嘧啶羟甲基化) 作为第 6 种碱基存在于人类基因组内（A、C、T、G、5-mC 以外）。5-hmC 可以产生自 Tet 类酶 (Ten-eleven Translocation Enzymes) 对 5-mC 的氧化，因而长期被人们认为是一个没有特定功能的中间产物。最近的研究则显示 5-hmC 是哺乳类动物基因组中一类拥有独特功能的、相当稳定的重要表观遗传标记。例如，Branco 等研究发现 (Nat Rev Genet 2012; 13: 7-13)，在哺乳动物大脑中发现大约 40% 的被修饰胞嘧啶为 5-hmC。在干细胞中，Tet 类酶的变异会引起特异表型。在人类和小鼠全基因组的扫描中进一步发现，5-hmC 位点相当稳定，并且在胚胎干细胞中多达 50% 以上，说明 5-hmC 不可能仅为 5-mC 的氧化中间产物。同时，有些蛋白（如转录因子）也被发现可以优先结合在基因组中的 5-hmC 位点，值得注意的是，这些蛋白包括 MECP2 (methyl CpG binding protein)，而 MECP2 可以通过抑制 IL-6 在癌症中起重要作用。所以，作为一类具有重要生物学功能的、稳定的胞嘧啶修饰，5-hmC 极有可能在人类疾病（如恶性肿瘤）的

发生发展中起重要作用，从而具有成为一类新型生物标记的潜力。

## 2.2 癌症诊断和检测技术的突破——液体活检（即体液活检）：

随着对癌症研究的深入，发现在临床诊断和治疗过程中使用的组织活检技术存在一定的局限性。主要表现为：1）肿瘤内部具有异质性（Intra-tumoral heterogeneity），对于组织活检或癌细胞已经发生转移的患者而言，仅仅取某个部位的肿瘤组织，并不能反映病人的整体情况，而且在临床上对所有的肿瘤组织都取样检测不可能做到；2）某些患者自身的情况较差，不适合做组织活检；受到手术的扰动之后，有些肿瘤有加速转移的风险；3）组织活检的局限性对患者的治疗不利。因此对于癌症的诊断和检测技术有更高的要求。所以，利用血液等的液体活检（体液活检）有可能解决上述的问题，同时由于液体活检的便利性和微创性也更容易使病人接受检查。

**针对癌症检测，目前有下列几类液体活检的技术：**

1）循环肿瘤细胞（Circulating tumor cells, CTCs）：简言，CTCs 是指存在于外周血中的各类肿瘤细胞的统称；对 CTCs 检测包含细胞富集和细胞检测两部分，可以进行不同的组合，形成不同的技术。最常见的富集方法是抗体介导或物

理方法富集，接下来进行免疫组织化学计数或遗传学、组学分析。

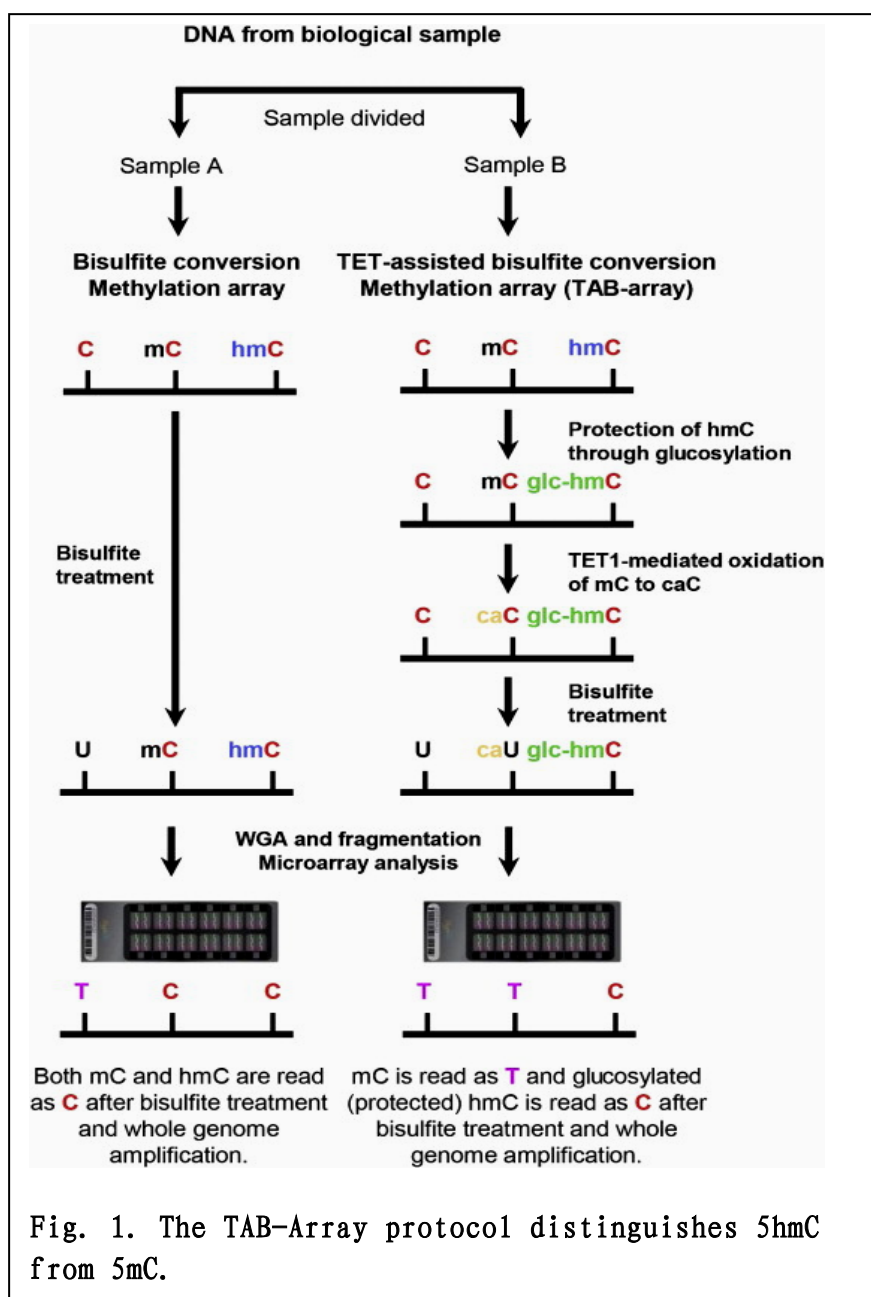
2) 血浆游离肿瘤 DNA (Cell-free circulating tumor DNA, ctDNA) 或血浆游离 DNA (Cell-free circulating DNA, cfDNA): ctDNA 即循环肿瘤 DNA 是人体血液循环系统中不断流动的携带一定特征 (包括突变, 甲基化等) 来自肿瘤基因的 DNA 片段。一般来源于坏死或凋亡的肿瘤细胞、循环肿瘤细胞以及肿瘤细胞分泌的外泌体。而 cfDNA 是血浆中游离存在的 DNA, 它们有的来自于正常细胞, 有的来自于异常细胞 (如肿瘤细胞), 还有部分来自于人体外部 (如病毒 DNA)。

目前国际上利用上述液体活检的技术进行癌症标志物开发的研究方兴未艾。例如最近 Nature Genetics 发表有中国鵬远基因和美国加州大学 San Diego 分校的一项利用 cfDNA 中的 5-mC 来区分定位人体肿瘤的研究 (Nat Genet. 2017 Apr; 49 (4): 635-642)。过去几年其他的一些研究也表明液体活检在癌症标志物开发中的巨大潜力和前景。例如, Science Translational Medicine 发表的一项研究报告中指出, 一种新的基因测序技术可能比传统的肿瘤活检技术检测癌性突变更安全、更廉价。这种叫做 "TAm-Seq" 的技术可通过收集血浆中循环 DNA 片段, 放大与癌症相关的 DNA 部分并深入挖掘基

因序列来发现那些难以辨认的突变 (Sci Transl Med. 2012 May 30; 4(136):136ra68)。

2.3 区分 5-hmC/5-mC 以及适合微量 DNA 的新技术:

5-mC 富集于启动子和 CpG 岛区域并且多为肿瘤抑制基因, 与 5-mC 不同, 5-hmC 主要富集于增强子和基因内部 (转录起始和转



录终止位置之间的区域), 并且可能更多地与活跃表达的基因关联, 增加或维持基因表达。尽管这两类胞嘧啶修饰存在不同的基因组分布和各自独特的基因调控功能, 目前主流的基于亚硫酸盐处理的检测胞嘧啶修饰的平台 (包括广泛使用

的 450K 微芯片以及 2016 年初发布的最新 Illumina EPIC 芯片) 从技术上不能区分 5-hmC 和 5-mC。为了区分 5-hmC 和 5-mC, 申请人的长期合作者, 美国芝加哥大学化学系的何川教授在 2011 年发明了 TAB-Seq (Tet-assisted Bisulfite Sequencing) 方法 (Cell, 2012. 149: 1368-80.)。就技术特点简言之, TAB-Seq 技术通过糖基化保护基因组内的 5-hmC 以在后续的亚硫酸盐测序时能够有效地区分 5-hmC 和 5-mC (图 1, Fig. 1)。

尽管 TAB-Seq 可以精确地区分出 5-hmC 和 5-mC 位点, 但依赖于全基因组测序的技术, 在临床应用方面存在费用和可行性的问题。针对这一限制, 一项利用化学标记 (Chemical Labeling) 的称作 nano-Seal 的技术被用来特异性地富集 5-hmC 区域 (Nat Biotechnol, 2011. 29: 68-72)。nano-Seal 已被验证可以用于少至 1-2 ng 的 DNA 量, 因而可能被用于诸如采自病人液体活检的 cfDNA (1-2ng 的 DNA 量大致相当于 1-2ml 血浆中的 cfDNA) (Mol Cell, 2016. 63: 1-9)。本项目将利用 nano-Seal 这一最新技术来研究癌症的表观遗传学, 从而寻找新型癌症相关标志物。nano-Seal 的详细步骤见“总体思路 and 计划”。

## 2.4 研究计划的目标:

综上所述, 液体活检 (即体液活检) 技术使得癌症早期阶段的筛查成为可能, 无论是肿瘤诊断、治疗还是监控领

域，液体活检的优势都使其成为最具发展潜力的肿瘤无创诊断手段，因此具有极高的临床应用价值和市场前景，也将带领巨大的经济效益和社会效益。当然，作为一个新兴领域，各种技术还需要进一步完善和验证。申请人团队将充分发挥在人类遗传学、肿瘤表观遗传学、生物信息学等研究领域的经验和技能，来探索在液体活检（特别是 cfDNA）中利用 5-hmC 和 5-mC 作为癌症标记物的可能性，研发更便利、创伤更小、费用更低、病人更易接受的生物标记物，为肿瘤的早期诊断、进展监测、预后及治疗反应的预测等提供科学依据。

在本计划中，我们将利用山东地域性流行病学的特殊情况，针对山东人群一个特别重要的癌症——食管癌进行攻关。山东省消化道肿瘤发病率和死亡率在所有恶性肿瘤中占四个，其中胃癌和食管癌分别在第二和第三位。在济宁地区，食管癌的发病率（1400 例/年）仅次于胃癌（1900 例/年）。本计划的研究内容将来可以扩展到其他癌症或复杂疾病，因而会对相关医学研究产生巨大的影响。目前，国内外相关产业也处于起步阶段，本计划的实施将加强山东省在这一新兴领域的研发实力和未来生物医药技术领域的发展潜力。

（二）申报人选及团队在国内外本学科领域所处的地位或具有的影响

申请人长期从事人类复杂疾病与性状的遗传/表观遗传学研究，在人类遗传学、肿瘤的表观遗传学、生物信息学、全

基因组大数据处理分析、高通量“组学”测量领域积累了大量的经验，已发表超过 125 篇 SCI 论文（其中第一作者或通讯作者超过 50 篇）、3 篇国外出版的教科书章节（通讯作者）和 100 多次国际学术会议的口头报告和摘要。申请人先后在美国伊利诺伊大学和美国西北大学担任教职，作为 PI 承担多项与本研究计划相关的美国卫生总署（NIH）和其他机构资助的科研项目，积累了丰富的追踪国际前沿研究及先进技术的经验。

现将申报人选及团队在本学科领域的研究工作和具有的影响简述如下：

### **1、胞嘧啶修饰在人群中的差异、遗传学构架及大数据库的建立：**

作为目前研究最深入的表观遗传学标记，胞嘧啶修饰已知广泛地参与各种生物学过程及复杂性状和疾病的发生（如癌症发生、基因表达、药物反应等）。例如，申请人曾经利用 Illumina 450K 人类甲基化微芯片对 133 个国际 HapMap 计划的类淋巴母细胞系（lymphoblastoid cell line, LCL）样品进行了全基因组胞嘧啶修饰（5-mC）的测定。在这 133 个样品中，60 个来源于欧洲裔白人（美国犹他州居民），73 个来源于非洲裔黑人（尼日利亚 Yoruba 部族）（Genetics, 2013; 194: 987-96）。申请人的这一研究是在世界上首次利用 450K 微芯片这一平台在全基因组水平上观察到胞嘧啶修饰存



在显著的个体和族群差异，并且族群（白种人和黑种人）特异 CpG 位点可能和某些族群特异的疾病有关。同时申请人分析显示了有相当部分的基因表达和它们邻近（比如 100kb 以内）的 CpG 位点有关联，从而验证了胞嘧啶修饰与人群基因表达这一复杂性状的差异有关。通过结合 DNA 基因型和胞嘧啶修饰，申请人揭示了遗传多态性（single nucleotide polymorphism，SNP）；通过 mQTL（methylation quantitative trait locus）控制一部分胞嘧啶修饰水平；同时胞嘧啶修饰也是一类近距离作用的基因调控因子。我们的分析结果显示 CpG 附近的 DNA 顺序（SNPs）的差异会导致个体间在同一个 CpG 位点不同的修饰水平，从而证明了遗传学和表观遗传学之间存在一定的联系，并且这种 CpG 和 SNP 的关系决定了相当一部分种族间特异的胞嘧啶修饰水平（Human Molecular Genetics, 2014; 23: 5893-905）。

值得一提的是申请人除了基础研究外，也拥有丰富的开发生物信息数据库的经验，迄今开发和维护了多个提供全基因组研究结果的数据库（Bioinformatics. 2010; 26(2): 259-62），特别是提供基于 HapMap LCL 样品的结果的数据库。例如 SCANdb（SNP and Copy Number Variation Annotation Database）是申请人课题组与美国芝加哥大学医学系 Nancy Cox 教授（现为美国 Vanderbilt 大学遗传学研究所主任）长期合作、共同开发的一个公共、免费的公共数据

库，提供基于 HapMap LCL 样品的全基因组分析结果，内容包括申请人发表在《美国人类遗传学杂志》（Am J Hum Genet）上的两大族裔（欧洲裔和非洲裔）的 eQTL (expression quantitative trait loci)（Am J Hum Genet. 2008; 82(3):631-40.）以及来自申请人发表在 Hum Mol Genet 的 mQTL 数据（Hum Mol Genet. 2014; 23(22):5893-905）。SCANdb 同时向用户提供相关的生物学信息，包括 SNP 位置、目标（如 CpG 位点）附近的基因、连锁不平衡（LD: linkage disequilibrium）等信息。SCANdb 自 2009 年开发以来已经有来自世界各地的网络访问超过 200 万次，分别来自近 60,000 个不同的互联网地址，充分证明了这类生物信息数据库的重要性及受欢迎程度。

## 2、胞嘧啶修饰与复杂性状及疾病的关联性：

在复杂性状或疾病的表观遗传学方面，申请人同样积累了相当多的研究成果。例如，利用上述的 HapMap 计划的类淋巴细胞系，申请人研究了胞嘧啶修饰和抗癌药物反应之间的关系。申请人发现相当大一部分个体之间对于某些抗癌化疗药物反应的差异可以通过某些药物代谢有关基因的表观遗传差异获得解释，从而加深了们对于药物反应这一复杂性状的认识，这些研究为将来进行癌症有关的精准医疗打下基础（Human Molecular Genetics 2013; 22: 4007-20）。申请人还涉及其他一些疾病的表观遗传学研究，包括急性肺损伤、

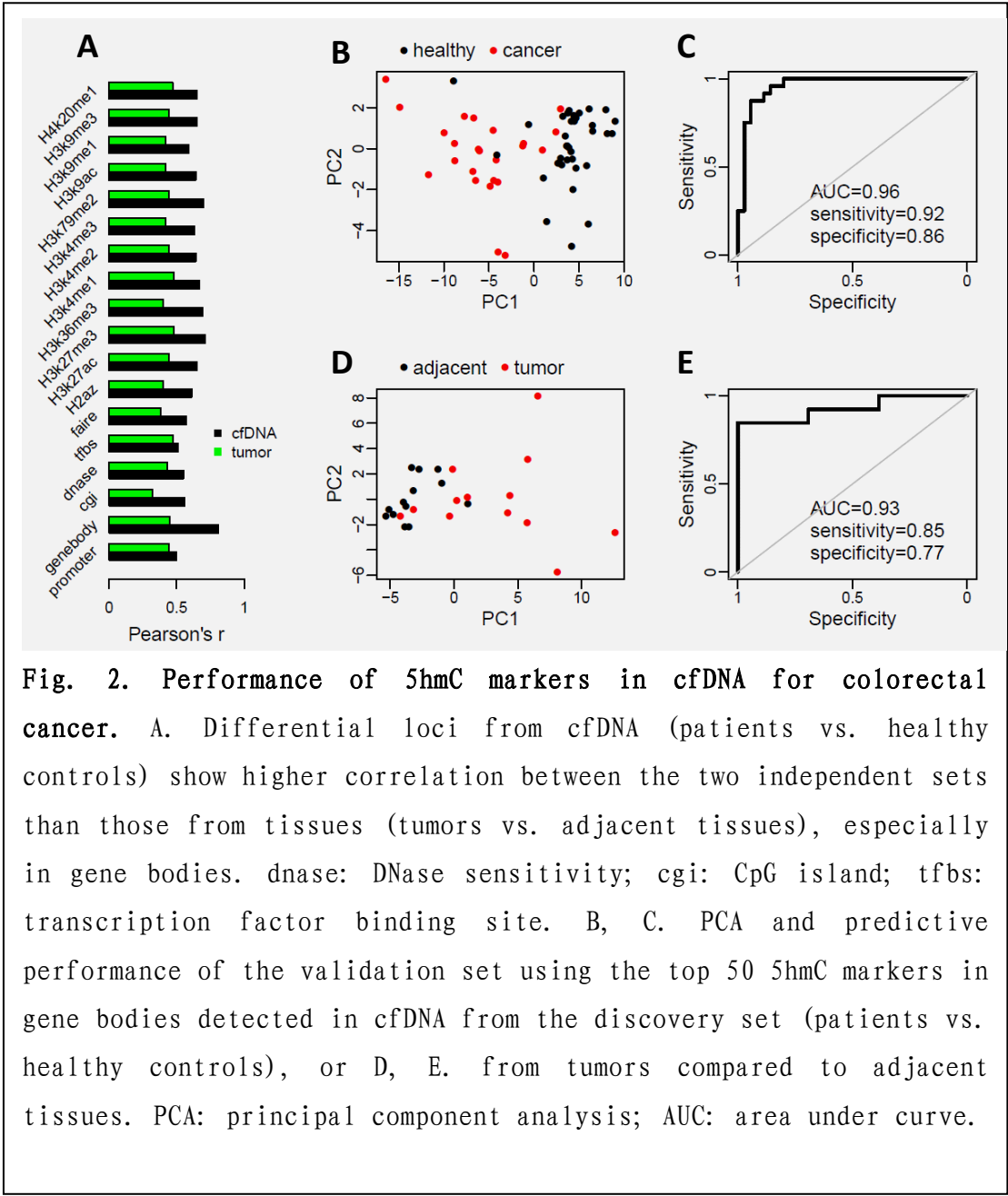
环境健康等 (Transl Res. 2017; 180:12-21; Genetics. 2016; 203(2):985-95; Environ Health. 2015; 14(1):65)

### 3、5-hmC 作为癌症标志物以及液体活检用于癌症检测:

5-hmC 作为癌症标志物属于前沿研究。申请人目前主持两项美国 NIH 和两项其他独立机构资助的癌症表观遗传研究项目, 其中与本研究计划相关的有“5-hmC 生物学标志物用于胰腺癌早期检测”(美国国家癌症研究所 NIH/NCI); “5-hmC 作为一类新型生物标志物用于预测脑癌临床表现”(美国西北大学脑癌研究所及 Phi Beta Psi Sorority 基金会)。申请人同时作为重要合作者正在参与其他多项相关研究(包括结直肠癌、脑胶质瘤等), 因此申请人已经积累相当多的与本研究计划直接有关的研究经验和思路。

有关在 cfDNA 中寻找基于胞嘧啶修饰, 特别是 5-hmC 癌症标志物的研究领域, 申请人及团队处于世界领先的位置。申请人和合作者目前已经在美国西北大学、芝加哥大学以及中国复旦大学华山医院用 nano-Seal 技术对于超过 500 例来自肠癌、胃癌、肝癌、胰腺癌, 以及其他复杂疾病(如 IBD、肺纤维化等)病人的血浆 cfDNA 成功地检测分析, 筛选到一系列与疾病相关联的基于 cfDNA 的 5-hmC 标志物, 目前正在组织临床验证及扩大试验。申请人作为共同通讯作者的相关论文正在 Nature 杂志审稿过程中(W Li, et al., Wei Zhang\*, Chuan He\*, and Jie Liu\*, DNA 5-

hydroxymethylcytosines from cell-free circulating DNA as diagnostic biomarkers for human cancers. Nature, 2017. revision submitted; \*共同通讯作者)。这篇论文中详细介绍了在本计划中将使用的主要技术并列示筛选肠癌、胃癌的 cfDNA 表观遗传学标志物作为例子 (图 2 , Fig. 2)。



### (一) 总体思路

Fig. 3 简要描述了本计划的总体流程和要实现的目标，详述如下。

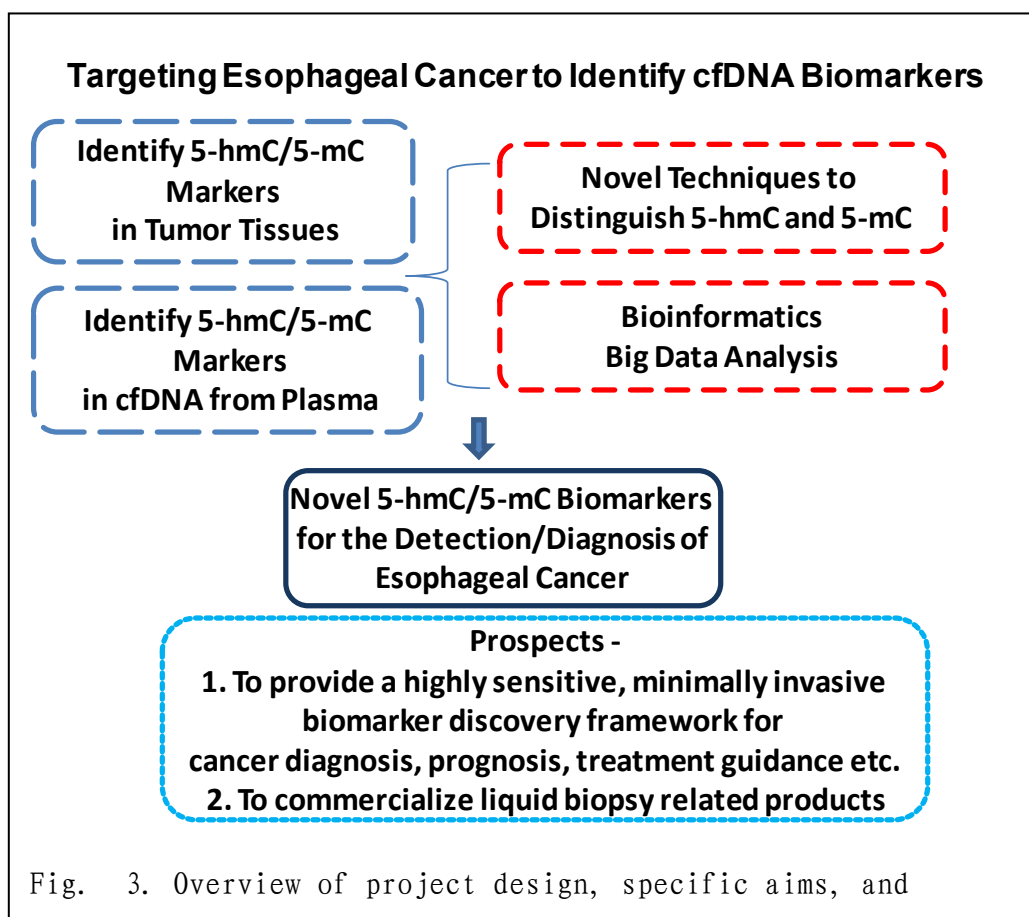


Fig. 3. Overview of project design, specific aims, and

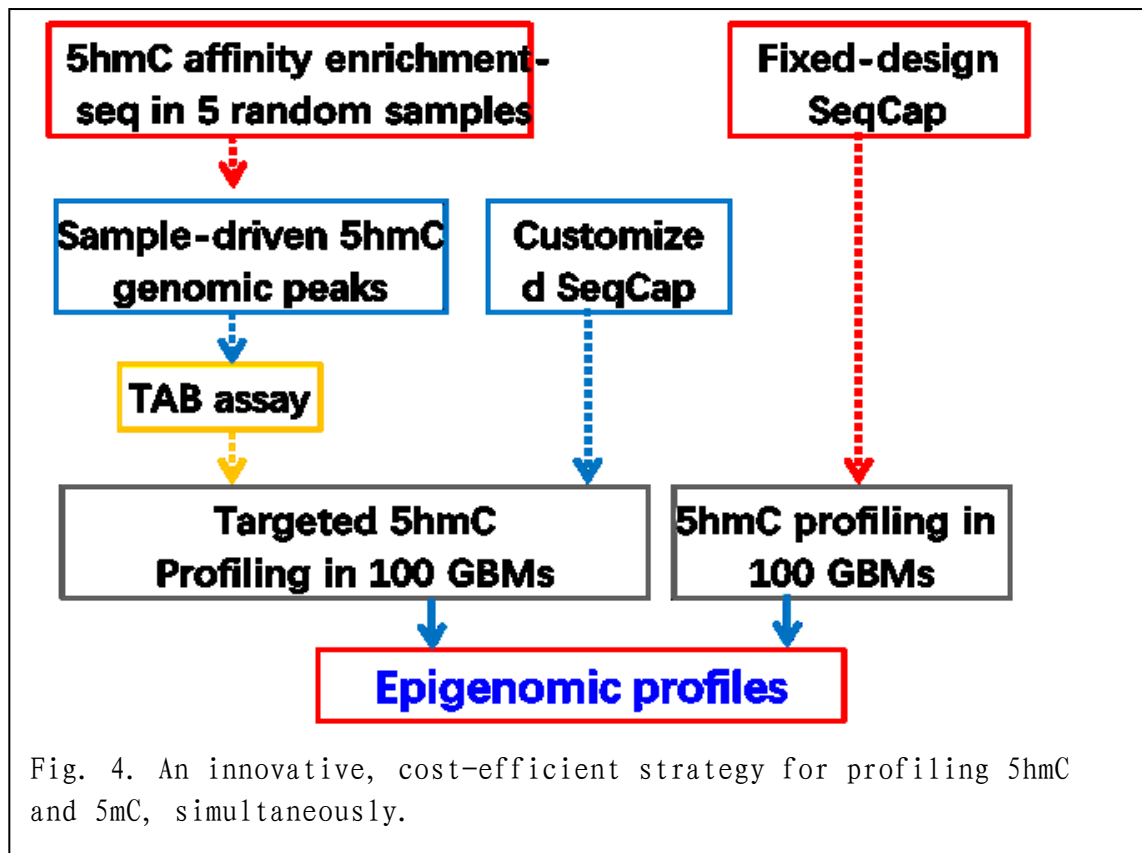
## 研究对象 - 以食管癌为攻关对象研究 5-hmC/5-mC 作为肿瘤标志物

我们将利用如上所述的最新技术，如 TAB-Seq、Nano-Sea1 分别在肿瘤病理组织和血液来源的 cfDNA 中检测和寻找 5-hmC/5-mC 标志物，并比较这两种来源 DNA 胞嘧啶修饰物用于区分癌症病人/肿瘤组织的准确性。我们的总体思路是利用山东省多发的食管癌为突破口，研究开发基于表观遗传学和液体活检的癌症早期诊断工具，为将来相关的大规模临床研究和产业开发打下基础。

以下将详细介绍有关的研究计划。

### 1、在肿瘤病理样品中寻找 5-hmC 和 5-mC 标记

### (1) 5-hmC/5-mC 的检测



目前商业化的表观遗传测量平台普遍基于亚硫酸盐处理 (Bisulfite conversion)，因而不能区分 5-hmC/5-mC 这两类胞嘧啶修饰。同时，由于基于全基因组测序的平台费用高昂，而其它基于微芯片的平台，如 Illumina 450K 以及 2016 年初推出的 EPIC 微芯片等并非针对 5-hmC 特异性设计，且只有在欧洲裔白人中优化设计，所以我们将要建立一个可根据样品来源，特别是中国人样品，灵活调整的、高通量的、相对价格低廉、高性价比（和全基因组测序对比）的 5-hmC 检测流程策略。特别强调的是，在本课题中，我们将结合 5-hmC 亲和富集测序的“粗”筛选和可定制的 CpG 甲基化测量平台（即高通量“细”测）来达到定向测量 5-hmC 的

目的。其后，在整合商业化的 5-mC 高通量测量平台的结果后，我们将可以区分 5-hmC 和 5-mC 用于后续的分析研究。这个测量策略（图 4， Fig. 4）包含以下几个步骤：

**第 1 步：**利用 5-hmC 亲合富集测序在一部分有代表性的样品中进行“粗”筛，目的是得到基于所研究中国病人群体的 5-hmC 富集区域。由于肿瘤组织和对照组织有可能存在有不同的 5-hmC 位点，并且我们预期个体之间存在差异，因此，在这一步我们将随机选择测试 15 个样品（5 个肿瘤组织，包括 1-2 个早期和 3-5 晚期、5 个相应的癌旁正常组织、5 个健康对照组织）。对于测序的结果我们将利用 Galaxy 平台（<http://galaxyproject.org/>）的二代测序流程来进行处理（利用开源的 Bowtie2 等分析软件）、质量控制（如去除低质量的测序结果）和探测 5-hmC 富集区域（利用 MACS 软件 – Model-based Analysis for ChIP-Seq 寻找富集的 5-hmC 峰域，<http://liulab.dfci.harvard.edu/MACS/>）等分析，申请人研究团队有丰富的分析二代测序和 5-hmC 亲合富集测序数据的经验。

**第 2 步：**得到 5-hmC 富集区域的信息（所包含的 CpG 位点）后，我们将和 NimbleGen 公司合作定制一个基于 NimbleGen Epi Choice 平台的 5-hmC 测量平台系统。我们将合并所有样品中找到的 5-hmC 富集区域，基于我们的前期预实验，我们预计大致可以获得约 2 万左右高质量的 5-hmC 基因组区域



（约 5 万 CpG 位点）。为了区分 5-hmC 位点我们将利用 TAB-Seq 技术（**Fig. 1**）来处理样品。TAB 通过特异性地糖基化基因组内的 5-hmC 来区分不被保护的 5-mC 和未修饰的胞嘧啶。

**第 3 步：** 利用 NimbleGen Epi Enrichment 平台系统来测量 5-mC。该平台的设计为全基因组覆盖，既包括启动子区域也包括有基因体，非翻译区以及基因间区域，共涵盖有约 5.5 百万的 CpG 位点，与表观遗传组学常用的 Illumina 微芯片（450K 芯片约 48 万 CpG 位点，或 EPIC 芯片 85 万 CpG 位点）比较，该系统有更广泛的基因组覆盖度，因而可以提供更为详尽的 CpG 甲基化信息。步骤 2 和 3 的原始数据也将利用如上所描述的 Galaxy 平台二代测序分析流程来处理、分析。

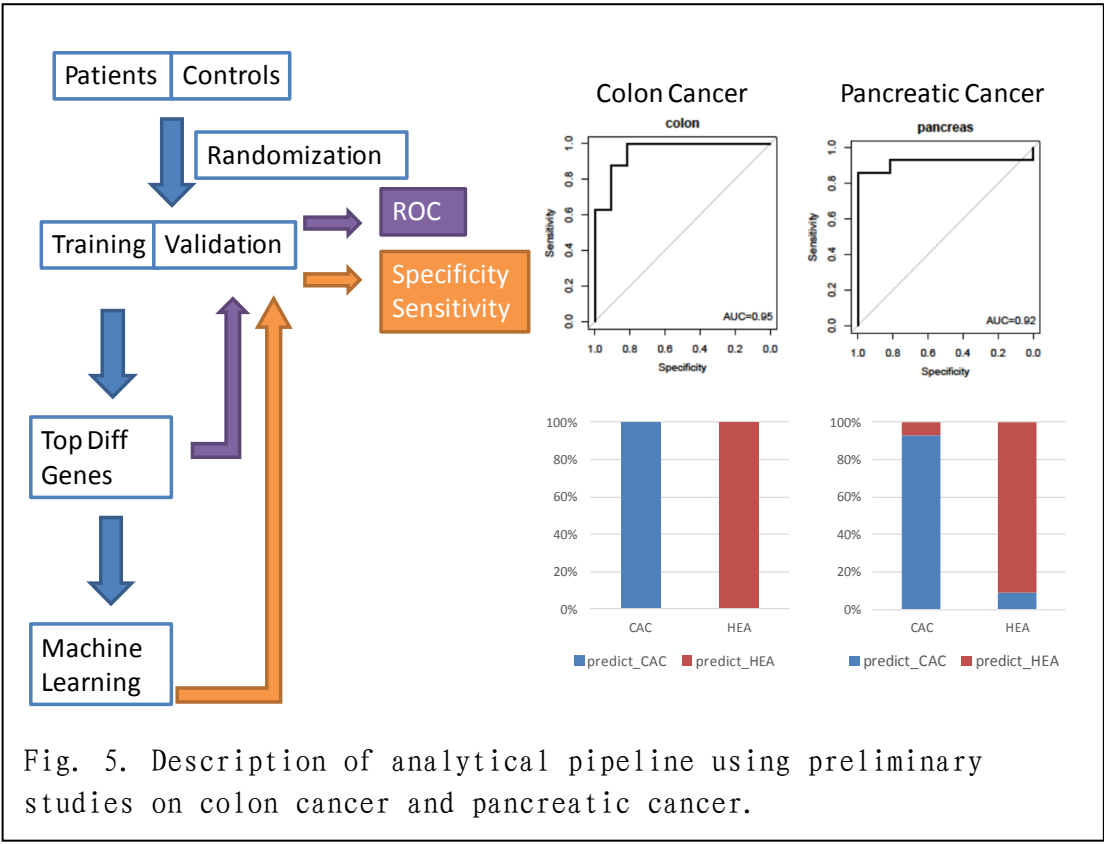
综合来自步骤 2（5-hmC）和步骤 3（5-mC）的测量结果后，我们将可以获得这两类胞嘧啶修饰在所测样品中的分布和修饰水平，在进行标准化（如 Quantile Normalization）和去除批次影响（如利用 COMBAT 工具来控制，<https://www.bu.edu/jlab/wp-ssets/ComBat/Abstract.html>）后的数据将被用于后续分析。

**（2）区分肿瘤中的胞嘧啶修饰-**我们将利用如上描述的检测策略（**Fig. 4**）对我们的“发现组”样品，即在济宁医学院精准医学研究院收集的 50 个经病理确认的食管癌肿瘤组织及癌旁正常组织（~1/3 为早期病人），以及 30 例年龄、性别、

BMI (Body Mass Index) 配对的健康对照组进行全表观基因组 5-hmC 和 5-mC 修饰水平的检测分析。这些食管癌病人都将经过严格的病理组织学检验 (例如免疫组化确定样品中癌细胞的含量达到 80%以上, 癌旁样品中则不含有癌细胞), 并且排除经过放化疗或凝血障碍的病人, 而健康对照组则来源于非癌症病人的食道、胃镜检查病人。所有涉及的样本收集都将得到参与者的知情同意及学校伦理委员会的许可。这些标本都从济宁医学院附属医院收集, 收集后立即放在液氮罐中冷冻保存, 在实验室-80℃冰箱长期存放。

**(3) 多因素回归方法及寻找食管癌 5-hmC 或 5-mC 标记** -对于肿瘤和癌旁对照 (或健康组织) 的数据利用由 (1) 得到的食管癌差异 5-hmC 或 5-mC 位点 (如基于 5%假阳性的在肿瘤和对照组间有差异的位点), 开展对食管癌风险的多因子回归分析 (如利用 Elastic Net) (Journal of the Royal Statistical Society. Series B: Statistical Methodology, 2005. 67: 301-320), 建立加权多因子回归模型。我们并可对单一和混合因子 (也就是 5-hmC 或 5-mC 单独作为因子, 或两者共同作为模型中的因子) 估算这些标记的专一性、灵敏度, 通过 ROC (Receiver Operating Characteristic) 曲线的线下面积 (AUROC) 判断模型总的准确度。我们将采用 95%CI 标准评估这两种胞嘧啶修饰是否对

模型的准确度有互相辅助的作用。Fig. 5 以我们预实验的结果为例简单描述了大致的统计分析流程。



我们最终的目标是建立一个由 30 – 40 个 5-hm C / 5-mC 位点所组成的有准确度的食管癌标记组。选择决定时我们将考虑多重标准，比如统计显著性（p-Value 或关联度），是否靠近已知基因（便于后续基因调控功能研究），是否为食管癌特异位点，是否和功能位点重合（如基于 ENCODE – Encyclopedia of DNA Elements 所发表的人类基因组调控因子的数据（Nature, 2012. 489: 57-74），例如对于 DNase I 的敏感区域，也即基因表达区域）。

（4）验证所寻找到的食管癌关联胞嘧啶修饰

① 收集独立的验证组样品：

我们将在本项目前 2 年里持续收集约 100 个年龄、性别、癌症级别及其它临床指标配对的一组新的食管癌病人样品。除了 100 对肿瘤组织及癌旁组织外，我们也将收集 50 例年龄、性别、BMI 配对的非食管癌病人组织作为正常的对照（临床标准和发现组样本相同）。基于精准医学研究院从附属医院收集标本的情况看，可以顺利地收集到“验证组”所需的样本数。

## **② 测试验证食管癌 5-hmC 或 5-mC 标记：**

在“验证组”样品中，我们将对于从前述（1）（2）分析所得的 30 - 40 个食管癌关联胞嘧啶修饰位点采用 TAB-Seq（Fig. 1）来测量 5-hmC，以及使用焦磷酸测序来测量 5-mC。然后我们将在“验证组”中利用如上所述我们所建立的食管癌加权多因子回归模型来估算每一个样品被检测出作为肿瘤的概率（或者也可以称为“预测分数”；分数 > 0.5 可以作为阳性 / 癌症，分数 < 0.5 则作为阴性 / 非癌症）。接下来，我们将利用这些标记在验证组样品中计算出该标记组的专一性、灵敏度，并通过 ROC 曲线的线下面积（AUROC）来判断模型总的准确度。

## **2、在血液 cfDNA 样品中寻找 5-hmC 和 5-mC 标记**

我们将利用 nano-Seal 技术在采自病人血浆的 cfDNA 中寻找食管癌 5-hmC 和 5-mC 标记，并和肿瘤组织中找到的 5-hmC 和 5-mC 标记比较总体准确度、敏感度和专一性。

(1) nano-Seal 技术

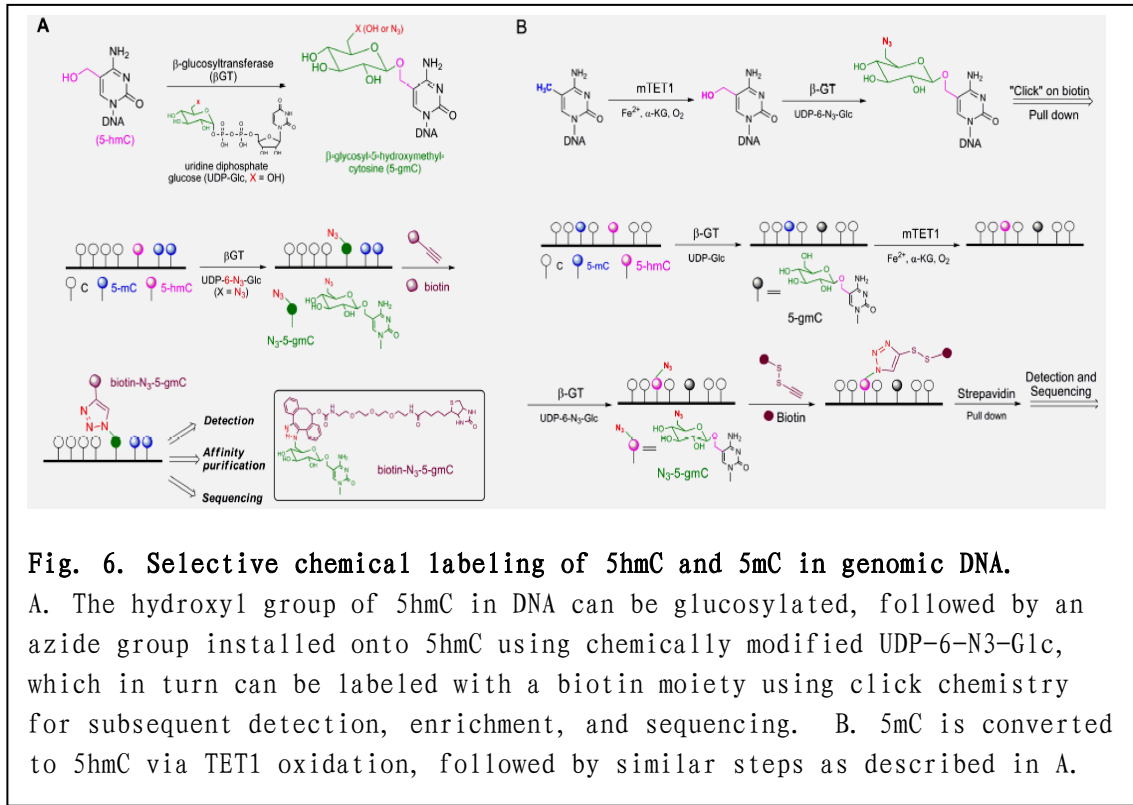


图 6 (Fig. 6) 简单介绍了这一技术步骤。申请人和合作者的前期工作证明了 nano-Seal 技术可以适用于少至 1-2ng 的 DNA 量，因而可以从约 2-3ml 的血浆中获得足够的 DNA。

我们将收集 50 个食管癌病人的血浆，这些血样将在病人诊断时获得，因而不会受到放化疗的影响。同时我们将收集 50 例健康人的血浆，健康人来自普通体检人群或其他非严重系统性疾病的人群，我们将对食管癌病人和健康样品进行

年龄段和性别的配对。CfDNA 将利用 QIAamp Circulating Nucleic Acid Kit (Qiagen) 提取。每个样品至少 1 ng 的 cfDNA 将被用来进行 nano-Seal 的建库及随后的二代测序。

## **(2) 数据分析和寻找 cfDNA 中的 5-hmC 和 5-mC 标记**

二代测序数据处理将大致如前所述，特别的地方在于我们将会按照 Ref-Seq 基因 (Nucleic Acids Res, 2005. 33: D501-4. PMC539979) 和 ENCODE 调控因子来标准化各个样品的数据，以利于分析比较。其后的数据处理和统计分析大致上与前面所述肿瘤组织内的分析一样，我们的目标是寻找 30 - 50 个基于 cfDNA 的食管癌 5-hmC 和 5-mC 标记。我们将通过比较 cfDNA 和组织内 DNA 分别所获得 5-hmC 和 5-mC 标记的总体准确度、敏感度和专一性来判断 cfDNA 是否可以替代组织活检技术，最终用于食管癌的检测。

## **(二) 工作计划**

(分年度列出 5 年工作计划)

### **1、 2018 年 1 月—2018 年 12 月**

1-6 月 准备发现组的肿瘤组织样品，提取 DNA，检测 DNA 质量，测试 5-hmC 亲合富集测序流程；

7-12 月 随机抽取 10 个样品 (5 个肿瘤，5 个正常组织) 进行 5-hmC 亲合富集测序，处理分析数据；开始收集 cfDNA 样品；

## **2、 2019 年 1 月—2019 年 12 月**

1-6 月 定制 NimbleGen Epi Choice 系统在发现组中（60 个肿瘤，60 个配对正常组织）中测试 5-hmC；

7-12 月 分析发现组 5-hmC 数据，利用 NimbleGen Epi Enrichment 系统测量 5-mC；挑选 20-30 个候选 CpG 位点；继续收集 cfDNA 样品；

## **3、 2020 年 1 月—2020 年 12 月**

1-6 月 继续完成所有样品的测量，完成发现组数据处理，对于选择的位点进行验证确认大规模测量数据的可靠性；完成收集 cfDNA 样品；

7-12 月 进行 cfDNA 中的 nano-Seal 测试；

## **4、 2021 年 1 月—2021 年 12 月**

1-6 月 完成 cfDNA 中的 nano-Seal 测试，开始数据处理；

7-12 月 完成 cfDNA 数据的分析；

## **5、 2022 年 1 月—2022 年 12 月**

1-6 月 在验证组中测试 cfDNA 的 5-hmC/5-mC 标记；

7-12 月 比较 cfDNA 和组织标记的总体准确度、敏感度和专一性。

## **三、目标任务**

### **（一）科研攻关目标**

（主要包括：拟申请立项的重大项目等）

1、主要研究目标是：探索食管癌的表观遗传标记，明确 5-hmC/5-mC 作为一类新型癌症标志物。在肿瘤表观遗传学及分子标志物等方面申请国家级课题 2-3 项，省部级课题 3-4 项。

2、攻关目标：

- (1) 食管癌的肿瘤组织的表观遗传标记；
- (2) 食管癌的液体活检的表观遗传标记。
- (3) 建立起适合临床大规模应用的相关生物信息分析流程

(二) 科技成果目标

(主要包括：科技奖励、高水平论文和专利等)

1、在国际知名期刊发表 SCI 论文 10 篇以上，单篇影响因子大于 5.0 的不少于 5 篇，其中影响因子大于 10.0 的 1-2 篇；

2、获批 1 项省部级以上科学技术奖励。

(三) 团队建设目标

(主要包括：人才培养、引进和创新团队建设等)

1、人才培养：

- 培养 2-3 名具有博士学位的青年教师成为研究团队的骨干：具有一定的肿瘤表观遗传学的理论基础；具有一定的独立分析问题解决问题的能力；掌握液体活检、表观遗传学检测分析等实验技能。



- 研究团队成员 1-2 名获批山东省泰山学者计划青年专家项目或入选省级及以上其他人才项目。

2、人才引进：引进 1-2 名学科带头人和 4-5 名优秀青年博士：具有坚实的 肿瘤表观遗传学 的理论基础；具有严谨的逻辑思考，具备提出问题、分析问题、解决问题的能力；具有良好的英文文献阅读和论文撰写能力，成为研究团队的项目带头人。

### 3、创新团队建设

- 打造一个研究基础雄厚、人才结构合理、发展前景良好的具有国际领先水平的肿瘤表观遗传学及液体活检、肿瘤标志物等研究团队；
- 构建表观遗传学及生物信息学的研究平台。

#### （四）学科平台建设目标

通过本项目的实施，造就一支由优秀中青年科研人员组成的，具有国际竞争力的肿瘤表观遗传学、生物信息学和肿瘤标志物的科技创新团队；培养一批从事高水平科学研究的后备力量；建立国际一流的实验平台；发表一批具有国际影响力的高水平研究论文和具有自主知识产权的新方法和新技术。

#### （五）经济社会效益目标

食管癌已经成为我国恶性肿瘤中发病率持续在高水平的一种肿瘤，在临床就诊的病人一般已经处于中晚期，花费巨

大，治疗效果不佳，5 年生存率极低。因此，对于食管癌的高风险人群筛查和早期诊断非常重要。我们将要利用液体活检、生物信息学等技术发现独特的表观遗传学标志物，用于食管癌筛查和早期诊断、精确诊断，将有助于为食管癌高危人群的早期诊断，使食管癌得以早发现、早治疗，减少食管癌对人民健康的危害，节省花费，提高生存生活质量，促进社会的可持续发展。因此，本项目具有广泛的应用前景，必将带来巨大的经济和社会效益，为满足国家“逐步达到促进国民健康、带动科学发展、加速社会进步的战略目标”作出贡献。

## **四、保障措施**

### **（一）实验室及仪器设备保障**

本项目将依托“十三五”山东省高等学校肿瘤精准医学重点实验室、山东省“十二五”病理学与病理生理学重点学科（强化）、济宁医学院精准医学研究院以及国家基因检测示范中心（济宁医学院）研究平台，能够在实验室空间、工作环境、仪器设备、人员团队配置等方面给予充分的保障与支持。

### **（二）配套资金保障**

学校将积极为引进人才提供配套资金保障，为引进人才提供 500 万元的科研启动经费，每年为引进人才提供不少于

15 万元的科研配套经费，每年为引进人才及其所带科研团队提供不少 5 万元的岗位津贴配套经费。

### （三）人员团队保障

团队现有人员的情况：

鲍永华，女，40 岁，博士，教授。鲍教授为临床医学专业本科毕业，分子生物学和生物化学博士（吉林大学）和博士后（中国农业大学）。主要从事消化道肿瘤的分子机制和生物标志物研究，主持国家自然科学基金项目 1 项，参与国家自然科学基金重大研究计划项目和面上项目各 1 项，发表 SCI 论文 20 多篇，获得国家授权发明专利 3 项。

郭永臣，男，40 岁，博士，副教授。郭永臣毕业于临床医学专业本科，分子生物学和生物化学硕士（吉林大学）和博士（中国农业大学）。美国威斯康辛医学院生物信息学博士后。主要从事肿瘤的生物标志物和生物信息学研究，主持山东省自然科学基金面上项目 1 项，发表 SCI 论文 10 多篇，获得国家授权发明专利 3 项。

艾志营，男，32 岁，博士，讲师。艾志营本科毕业于生物科学专业，西北农林科技大学生物化学与分子生物学硕士、细胞生物学专业博士。参与国家自然科学基金面上项目 1 项。主要从事肿瘤干细胞的自我更新和定向分化研究，发表 SCI 论文 13 篇。

张营，男，34 岁，博士，讲师。张营本科毕业于生物技术专业，生物化学与分子生物学硕士（四川大学），机械工程专业博士（美国西弗尼吉亚大学）。已发表 SCI 论文 12 篇，其中一作 6 篇。主持山东省自然科学基金项目 1 项。

刘孝然，男，31 岁，博士，讲师。刘孝然本科毕业于生物科学，华中农业大学动物遗传育种与繁殖专业博士）。目前主要从事肿瘤生物标志物的研究，，参与国家自然科学基金面上项目 1 项，发表 SCI 论文 8 篇，获得国家专利 1 项。

王姗姗，女，32 岁，博士，讲师。本科毕业于山东大学生物技术专业，博士期间就读细胞生物学专业，博士毕业于中山大学。主要从事 HCV 病毒侵染机制和 EMT 相关肺癌转移机制研究，参与山东省自然科学基金项目 2 项。在 Scientific Reports, Plos One, Nature Communication 杂志发表 SCI 论文 4 篇。

根据学校的实际，优先为设置“泰山学者”特聘专家岗位的学科安排研究生招生计划和工作人员招聘计划，支持“泰山学者”特聘专家培养更多的中青年学术带头人。

#### （四）生活保障

1、津贴：每年提供 20 万元人民币（税前）的津贴，按月支付，并按照《济宁医学院科研工作量津贴计算办法》有关规定，根据其研究成果支付相应的科研业绩津贴。

2、住房：为泰山学者特聘专家提供周转房 1 套。