

## 2015 年中国医学遗传学研究领域若干重要进展

李元丰<sup>1</sup>, 韩玉波<sup>2</sup>, 曹鹏博<sup>1</sup>, 孟金凤<sup>1</sup>, 李海北<sup>1</sup>, 秦庚<sup>1</sup>, 张锋<sup>3</sup>, 靳光付<sup>4</sup>, 杨勇<sup>5</sup>, 郭玲仟<sup>6</sup>, 平杰<sup>1</sup>, 周钢桥<sup>1</sup>

1. 中国人民解放军军事医学科学院放射与辐射医学研究所, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 北京 100850;
2. 中国科学院遗传与发育生物学研究所, 北京 100101;
3. 复旦大学生命科学学院遗传工程国家重点实验室, 上海 200438;
4. 南京医科大学生殖医学国家重点实验室, 南京 210029;
5. 北京大学第一医院皮肤科, 北京 100034;
6. 中南大学医学遗传学国家重点实验室, 长沙 410008

**摘要:** 2015 年中国医学遗传学稳步发展, 众多具有原创性的研究论文在国际顶级杂志上发表。中国科学家在医学遗传学的诸多领域, 如罕见疾病的致病基因、复杂疾病的易感基因、癌症的体细胞突变、遗传学新方法新技术、疾病相关微小 RNA(microRNA, miRNA)、疾病相关长链非编码 RNA(Long non-coding RNA, lncRNA)、疾病相关竞争性内源 RNA(Competing endogenous RNA, ceRNA)、疾病相关可变剪接和分子进化等研究领域均取得了突破性的进展。中国科学家在医学遗传学研究中逐步从常见变异延伸到罕见变异, 从遗传学现象的描述到功能机制的验证, 从单组学分析扩展至多组学数据整合, 从基础研究走向临床应用。同时, 中国科学家的研究成果引起了国际同行的高度关注。本文概括性综述了 2015 年中国科学家在医学遗传学领域取得的若干重要研究进展, 旨在追踪当前中国医学遗传学领域发展的前沿, 与国内读者分享我国科学家在该领域取得的重要成果以及研究思路。

**关键词:** 中国; 医学遗传学; 表观遗传学; 研究进展; 2015 年

## Research advances on medical genetics in China in 2015

Yuanfeng Li<sup>1</sup>, Yubo Han<sup>2</sup>, Pengbo Cao<sup>1</sup>, Jinfeng Meng<sup>1</sup>, Haibei Li<sup>1</sup>, Geng Qin<sup>1</sup>, Feng Zhang<sup>3</sup>, Guangfu Jin<sup>4</sup>, Yong Yang<sup>5</sup>, Lingqian Wu<sup>6</sup>, Jie Ping<sup>1</sup>, Gangqiao Zhou<sup>1</sup>

1. State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 100850, China;
2. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China;
3. State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China;
4. State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 210029, China;
5. Department of Dermatology, Peking University First Hospital, Beijing 100034, China;
6. State Key Laboratory of Medical Genetics, Central South University, Changsha 410008, China

**Abstract:** Steady progress has been achieved in the medical genetics in China in 2015, as numerous original researches were published in the world's leading journals. Chinese scientists have made significant contributions to various fields of medical genetics, such as pathogenicity of rare diseases, predisposition of common diseases, somatic mutations of cancer,

收稿日期: 2016-04-27; 修回日期: 2016-05-13

通讯作者: 周钢桥, 研究员, 研究方向: 医学遗传学与基因组学。E-mail: zhougq114@126.com

DOI: 10.16288/j.ycz.16-174

网络出版时间:

URL:

new technologies and methods, disease-related microRNAs (miRNAs), disease-related long non-coding RNAs (lncRNAs), disease-related competing endogenous RNAs (ceRNAs), disease-related RNA splicing and molecular evolution. In these fields, Chinese scientists have gradually formed the tendency, from common variants to rare variants, from single omic analyses to multipleomics integration analyses, from genetic discovery to functional confirmation, from basic research to clinical application. Meanwhile, the findings of Chinese scientists have been drawn great attentions of international peers. This review aims to provide an overall picture of the front in Chinese medical genetics, and highlights the important findings and their research strategy.

**Keywords:** China; medical genetics; epigenetics; research advances; 2015

近年来, 中国科学家在医学遗传学研究领域取得了丰硕的成果, 获得了国际同行的广泛赞誉。2015年, 我国科学家在 *Nature*、*New England Journal of Medicine*、*Cell*、*Nature Genetics*、*Nature Medicine*、*Cancer Cell* 等国际著名的综合性学术期刊以及医学遗传学领域主流期刊上发表了众多具有原创意义的研究性论文, 涉及医学遗传学的各个重要领域。在诸如罕见疾病的致病基因、复杂疾病的易感基因、癌症的体细胞突变、遗传学新方法新技术、DNA 甲基化、疾病相关微小 RNA (microRNA, miRNA)、疾病相关长链非编码 RNA (Long non-coding RNA, lncRNA)、疾病相关竞争性内源 RNA (Competing endogenous RNA, ceRNA)、疾病相关可变剪接和分子进化研究等领域均取得了突破性的进展。例如, 在复杂疾病的易感基因研究中, 我国科学家为寻找复杂疾病“缺失的遗传度”开展了大样本量的全基因组关联研究、基于外显子芯片技术以增大编码区变异覆盖率的全基因组关联研究、基因-基因相互作用研究以及罕见变异研究。这些研究成果从不同角度揭示了复杂疾病“缺失的遗传度”, 是对复杂疾病遗传易感性的重要补充。在癌症的体细胞突变研究中, 我国科学家首次发现了 *PRPS1* 基因突变与急性成淋巴细胞性白血病耐药和复发密切相关, 为急性成淋巴细胞性白血病耐药和复发提供了治疗策略<sup>[1]</sup>, 该成果得到了国际同行的认可, *Nature Medicine* 同期发表评论, 认为该工作意义非常重要<sup>[2]</sup>。在 DNA 甲基化研究中, 我国科学家揭示了人类原始生殖细胞基因表达与表观遗传调控特征, 该成果为研究人类生殖细胞的表观遗传调控、早期胚胎全能性的建立以及 DNA 甲基化的隔代遗传等问题提供了理论基础<sup>[3]</sup>。该成果因其重大的科学意义, 入选了 2015

年度中国科学十大进展。在疾病相关 lncRNA 的研究中, 我国科学家揭示了 lncRNA NKILA 能够抑制肿瘤相关炎症的发生<sup>[4]</sup>。该成果作为 *Cancer Cell* 杂志的封面文章被报道, 充分体现中国科学家在 lncRNA 研究领域已处于国际前沿。在疾病相关 miRNA 的研究中, 我国科学家利用 miRNA 建立了肝癌早期预警的分类器<sup>[5]</sup>。该成果有重要的临床意义, *Lancet Oncology* 同期发表评论, 对该工作给予了高度评价<sup>[6]</sup>。

纵览过去一年医学遗传学领域的研究成果, 可喜地发现我国科学家在医学遗传学的研究正紧跟国际科学的前沿, 逐步从常见变异延伸到罕见变异, 从遗传学现象的描述扩展至功能机制的确证, 从单组学分析扩展至多组学数据整合, 从基础研究走向临床应用。相信我国科学家在这些研究领域内的成果将会极大推动相关科研工作的进展, 同时也会为我国的医疗卫生事业提供科学理论基础。

本文基于 2015 年我国科学家发表在国际主流期刊上的研究工作, 但不排除发表在其他期刊上的研究工作, 选取其中部分具有代表性的研究成果进行概括性介绍, 与读者分享。一方面, 回顾盘点过去一年我国科学家在医学遗传学领域取得的成果; 另一方面, 也希望借此客观、全面地展现我国医学遗传学发展的前沿和热点。由于资料收集体量较大以及篇幅所限, 难免会挂一漏万。如有纰漏, 亦请读者谅解。

## 1 疾病发生和发展的遗传病因研究

### 1.1 罕见疾病的致病基因研究

罕见疾病是指人群发病率极低的各类疾病。根据世界卫生组织的定义, 罕见疾病是指患病人数占

总人口的 0.65‰~1‰的疾病。罕见疾病一般由单个基因功能的改变所导致。虽然单种罕见疾病的发病率很低,但罕见疾病种类繁多,如果将所有罕见疾病纳入统计,全球范围内的罕见疾病患病人数高达数百万<sup>[7]</sup>。发现罕见疾病的致病基因,将为罕见病患者的诊断和治疗提供理论依据,具有重要的医学和社会意义。

高通量测序技术和高密度芯片技术的发展,为罕见疾病致病基因的研究提供了新的方法,使得罕见疾病致病基因的鉴定速度大幅加快。我国科学家采用高通量测序或高密度芯片技术,新发现了多种罕见疾病的致病基因。

发育不良性椎弓峡部裂是一种椎弓峡部发育缺损现象,是引起脊椎滑脱的潜在因素,多发生在腰骶之间,其遗传学机制尚不明确。温州医科大学蔡涛研究组和第四军医大学的罗卓荆研究组合作,应用全外显子组测序技术,在一个包含 5 例发育不良性椎弓峡部裂患者的家系中发现了 *SLC26A2* 基因的杂合突变 c.2286A>T。在另外 30 例无亲缘关系的患者中对该基因进行进一步筛查,发现了该基因上的另外两个突变(c.1922A>G 和 c.-93T>C)。进一步的功能研究显示,*SLC26A2* 基因在 14.5 天的小鼠胚胎腰骶椎中表达活跃。与野生型细胞比较,转染了突变型 *SLC26A2*(c.2286A>T 和 c.1922A>G)的 CHO 细胞(中国仓鼠卵巢细胞)硫酸盐吸收活性显著下降,从而证实了上述两个错义突变的致病性。总之,该研究揭示并证明了发育不良性椎弓峡部裂是由 *SLC26A2* 基因突变所导致<sup>[8]</sup>。

卵巢早衰是育龄期妇女最常见的内分泌性疾病之一。患者不仅表现为闭经、不孕等生殖障碍,而且由于雌激素水平持续较低,患者罹患心血管、神经系统及骨质疏松等疾病的风险显著增加。已有研究显示,卵巢早衰的遗传异质性较高。山东大学的陈子江研究组对一个包含 4 例卵巢早衰患者的中国汉族非近亲婚配家系的部分成员进行了全外显子组测序,发现 *PGBD3* 基因存在一个杂合错义突变 c.2237G>A。鉴于 *CSB-PGBD3* 融合基因(包含 *CSB* 基因的第 1~5 号外显子)以往曾被报道与 DNA 修复和科克因综合征(Cockayne syndrome)相关,他们进一步对 *CSB-PGBD3* 融合基因在 432 例散发性卵巢

早衰患者和 400 例健康对照个体中进行筛查,结果在患者中发现 *CSB-PGBD3* 融合基因上存在新的杂合错义突变 c.3166G>A 和杂合无义突变 c.643G>T。进一步的功能学研究结果表明,*CSB-PGBD3* 融合蛋白是卵母细胞中重要的 DNA 损伤修复因子,可能参与卵巢功能的维持;*CSB-PGBD3* 融合蛋白上的 3 个突变均不同程度地影响了 *CSB-PGBD3* 的 DNA 损伤修复功能。该项研究初步证明了 *CSB-PGBD3* 基因突变可能导致卵巢早衰的发生<sup>[9]</sup>。

除单碱基突变外,基因组拷贝数变异(Copy number variation, CNV)也是导致疾病发生的重要遗传变异形式。开展全基因组范围的 CNV 分析,可以揭示致病性 CNV 位点,为解释人类罕见疾病的发生提供新的视角。先天性脊柱侧凸是一种由于脊柱胚胎期发育异常导致的疾病。已有研究提示, DNA 变异可能是该疾病的重要病因,但其主要的致病基因一直未被发现。复旦大学的张锋研究组与北京协和医院的邱贵兴研究组合作,采用高密度比较基因组杂交芯片技术,解析了先天性脊柱侧凸患者的全基因组 CNV。研究发现,高达 7.5%的散发性先天性脊柱侧凸患者在人类基因组 16p11.2 区域内存在约 600 kb 的大片段 DNA 缺失。进一步的分析表明,缺失区域内的 *TBX6* 基因是该疾病的致病基因。在机制探寻中,他们发现 *TBX6* 基因的 CNV 等罕见变异不足以导致先天性脊柱侧凸,通常需要联合一个常见的 *TBX6* 等位基因单倍型共同致病。细胞功能实验显示,该单倍型可以下调 *TBX6* 基因的表达。这些发现不仅表明 *TBX6* 是迄今发现的最重要的先天性脊柱侧凸致病基因,而且揭示了 *TBX6* 基因致病的复合遗传机理,为先天性脊柱侧凸早期诊断及遗传咨询提供了理论依据。该项研究突破了在疾病遗传学研究中占主导地位的“常见疾病-常见变异”的理论框架,揭示了常见变异与罕见变异共同作用导致疾病发生的新机理<sup>[10]</sup>。

开展单碱基突变和 CNV 的联合分析,有望更全面地揭示罕见疾病的致病基因。汗孔角化症是多种角化性皮肤病的统称,该疾病具有显著的遗传性。但是,目前仅明确 *MVK* 基因是其中两种亚型的致病病因。复旦大学附属华山医院的张正华研究组通过对 134 例汗孔角化症先证者进行高通量测序和外显

子区域 CNV 的筛查,新鉴定了 3 个汗孔角化症致病基因(*PMVK*、*MVD* 和 *FDPS*)。这 3 个基因与已知的致病基因 *MVK* 均处在甲羟戊酸通路中。由此推测,甲羟戊酸通路产生的类异戊间二烯化合物很可能在汗孔角化症中起关键作用。同时他们还发现,携带相同致病基因的汗孔角化症病人,其临床表型往往相似。由此,该项研究不仅发现了汗孔角化症新的致病基因,而且提供了一种新的汗孔角化症分类体系<sup>[11]</sup>。

罕见疾病种类繁多,研究人员很难估计出其准确数量。目前,人类孟德尔遗传数据库(Online Mendelian Inheritance in Man, OMIM)和 Orphanet 数据库已收录超过 6000 种罕见疾病,更多的罕见疾病有待被定义。北京大学第一医院皮肤性病科的杨勇研究与英国伦敦大学玛丽皇后学院的 Kelsell 研究组合作,通过对一个近亲婚配的家族进行详细临床调查及分析后,确定了一种全新的疾病。该疾病是以皮肤脆性增加为主要特征的常染色体隐性遗传性皮肤病,患者主要表现为泛发性皮肤剥脱、白甲、肢端点状角化、唇炎及指节垫。在此之前,该疾病并不为人们所认识。他们取“皮肤剥脱(Peeling skin)、白甲(Leukonychia)、肢端点状角化(Acral punctate keratosis)、唇炎(Cheilitis)及指节垫(Kunckle pad)”的首字母,首次在国际上将其命名为 PLACK 综合征。通过对 3 个患病家系的全外显子组测序分析,发现编码钙蛋白酶抑制蛋白的 *CAST* 基因在 3 个患病家系中均发生了纯合突变,且所有 *CAST* 基因的致病性突变均可导致 *CAST* 蛋白功能的丧失。体外研究进一步证实 *CAST* 基因与角化细胞粘附相关。该项研究首次证实了 *CAST* 基因在皮肤复杂的角化生理过程中发挥重要作用,有望为大疱性皮肤病、鱼鳞病、异位性皮炎等疾病的治疗提供全新的药物靶点<sup>[12]</sup>。

严格的遗传学和流行病学分析手段,可有效发现并确证罕见疾病的致病基因。而基于病人病变组织或业内认可的细胞和动物模型,开展深入的基因功能机制研究,则可为致病基因的最终确证提供强有力的证据<sup>[13]</sup>。良性软骨肿瘤是人群中最常见的原发性骨组织肿瘤,该肿瘤以内生软骨瘤和骨软骨瘤为典型代表,目前其发病机制尚不明确。第三军医大学的陈林研究组在小鼠的软骨细胞中特异性敲除

*Fgfr3* 后,发现小鼠发生了多发性软骨瘤样病变。针对病变组织的研究显示,在 *Fgfr3* 特异性敲除的小鼠中,MAPK 信号通路活性降低;同时,与软骨生长密切相关的 IHH 蛋白表达量显著升高。他们进一步证实,抑制 MAPK 信号通路的活性,能够提高 *Ihh* 的转录水平。更为重要的是,他们还发现,如果给予 IHH 的抑制剂, *Fgfr3* 敲除小鼠的多发性软骨瘤的症状会减轻,进一步证实过多的 IHH 表达是多发性软骨瘤样病变的必要条件。由此,该项研究首次清晰地揭示了 *Fgfr3* 对于软骨瘤的抑制作用及其潜在的抑癌机制<sup>[14]</sup>。

值得说明的是,并不是所有被报道的候选突变最终都能够被证明是致病性的。致病突变的最终确认,需要遵循严格的标准<sup>[13]</sup>。首先,罕见疾病的诊断需要严格遵守临床诊断标准;其次,DNA 突变和基因致病性的评估需要符合目前关于遗传诊断的指南<sup>[13]</sup>。2015 年 2 月日本 Tatsuro Jo 研究组在 *New England Journal of Medicine* 发表文章,报告了两例日本急性发热性嗜中性皮病的致病性突变。然而,江苏省宿迁第一人民医院的 Han Fei 质疑他们的结论是草率的,甚至是有误导性的。他认为这两个病人所具有的皮肤病变不足以支持急性发热性嗜中性皮病的诊断,因为急性发热性嗜中性皮病的诊断需要满足两个主要标准,即有疼痛性红色斑块或结节、中性粒细胞广泛浸润真皮浅层且无白细胞破裂性脉管炎。然而,1 号病人根本不满足上述条件,2 号病人更像是坏疽性脓皮症<sup>[15]</sup>。此外,其他研究人员也对该报道提出了质疑,认为该报道中不仅临床诊断存在问题,发现的变异也不符合目前关于遗传诊断的指南<sup>[16]</sup>。

对确证的致病基因进行分子机制的探索,有望为罕见疾病的治疗提供潜在的药物靶点。含缬酪肽蛋白又称 p97,在泛素依赖的蛋白质水解、细胞自噬、胞膜重构等重要的生理过程中发挥关键作用。p97 单碱基突变能够导致一种被称为“包涵体肌病伴 Paget 骨病和额颞叶痴呆”的罕见疾病。p97 单碱基突变还被发现能够引发肌萎缩侧索硬化和帕金森症。但目前 p97 突变导致疾病的分子机制尚不明确。来自加利福尼亚洛杉矶生物医药研究所和首都医科大学药学院的 Xiaoyi Zhang 和 Lin Gui 研究组开展了

p97 辅因子 p37、p47 对野生型及突变型 p97 ATP 酶活性影响的研究。他们通过多项酶稳态动力学分析,证实 p97 辅因子 p37 能够增强 p97 的催化效率( $k_{cat}/K_m$ )。敲除 p47 中与 p37 不同源的氨基酸(第 69~92 位氨基酸)后, p47 由抑制辅因子变为与 p37 类似的促进辅因子。最为重要的是,他们发现 p97 上的突变将导致辅因子 p37 和 p47 无法正常激活 p97,由此导致了 p97 相关疾病的发生。该项研究证实了辅因子 p37 和 p47 在调控 p97 ATP 酶活性中的重要作用,为将来针对 p97 致病突变研发治疗药物提供了帮助<sup>[17]</sup>。

## 1.2 复杂疾病的易感基因研究

复杂疾病的发生发展是遗传因素和环境因素共同作用的结果。基于全基因组关联研究(Genome-wide association study, GWAS)策略,研究人员在近 10 年的研究中,成功发现了数以千计的复杂疾病易感基因区域,这些研究成果为复杂疾病的早期预警和个体化治疗提供了理论基础。尽管全基因组关联研究获得了巨大成功,通过全基因组关联研究发现的易感基因位点仅能解释复杂疾病的部分遗传度,而且这些易感基因位点的致病效应多数表现并不明显。因此,进一步开展更大规模的全基因组关联研究,可以有效发现复杂疾病中“缺失的遗传度”。

2015 年,我国科学家通过开展大规模的全基因组关联研究,新发现了多种复杂疾病的易感基因。其中,部分研究是国际上首次针对该疾病的全基因组关联研究,如垂体腺瘤的全基因组关联研究。垂体腺瘤是最常见的颅内神经内分泌肿瘤。流行病学研究显示,垂体腺瘤存在一定的遗传易感性,但目前对其遗传基础知之甚少。复旦大学附属华山医院的赵曜研究组联合国内多家研究单位,基于中国汉族人群,开展了国际上首项垂体腺瘤的全基因组关联研究。在发掘阶段,他们使用了 771 例垂体腺瘤病例和 2788 例对照个体;在两个独立验证人群中,他们共使用了 2542 例垂体腺瘤病例和 3620 例对照个体。最终,他们成功鉴定出 3 个新的垂体腺瘤易感基因区域,分别为:10p12.31(rs2359536),10q21.1(rs10763170)和 13q12.13(rs17083838)。对这 3 个区域进一步分析显示,10p12.31 区域的易感基因可能是 *NEBL* 或

*MALRDI*,10q21.1 区域的易感基因可能是 *PCDH15*,13q12.13 区域的易感基因可能是 *CDK8*。该项研究为垂体腺瘤的遗传基础提供了新的见解<sup>[18]</sup>。

先天性心脏病是目前世界上最常见的出生缺陷,多基因遗传缺陷是先天性心脏病的主要遗传病因。2013 年,南京医科大学的胡志斌研究组基于中国人开展了全基因组关联研究,并发现 2 个先天性心脏病的易感基因区域(1p12 和 4q31.1)<sup>[19]</sup>。为进一步探索先天性心脏病的遗传易感性,他们在前期全基因组关联研究的基础上,针对发掘阶段关联 *P* 值在  $10^{-4}$  到  $10^{-5}$  的 SNP 位点,在 6053 例病例和 7410 例对照中进行了验证。由此,新发现染色体 4q31.22、9p24.2、12q24.13 和 20q12 区域与先天性心脏病显著相关。这 4 个区域的标签 SNP 位点分别为:4q31.22 区域的 rs1400558(位于 *EDNRA* 基因上游)、9p24.2 区域的 rs7863990(*SMARCA2* 基因附近)、12q24.13 区域的 rs2433752(*TBX3* 和 *TBX5* 基因上游)和 20q12 区域的 rs490514(*PTPRT* 基因第 2 号内含子)。多项研究提示,上述基因均可能与先天性心脏病的发生有关。有趣的是,在欧洲人群中 rs490514 也与先天性心脏病显著相关。该项研究在中国人群中发现了新的先天性心脏病易感基因区域,为揭示先天性心脏病的发生机制提供了新的线索<sup>[20]</sup>。

痛风是男性最常见的关节炎之一。流行病学研究表明,肾功能受损和高尿酸血症是痛风的重要危险因素。前期全基因组关联研究已识别出一系列与血清尿酸盐浓度相关的遗传变异。然而,高尿酸血症并非痛风性关节炎发生的唯一影响因素。为拓宽对痛风遗传易感机制的理解,青岛大学医学院附属医院的李长贵和上海交通大学的师咏勇研究组开展了全基因组关联研究(包括 4275 例男性痛风病例和 6272 例男性健康对照),鉴别出 3 个新的痛风易感基因区域(17q23.2、9p24.2 和 11p15.5)。进一步的研究发现,上述易感基因区域与高尿酸血症并不相关,提示它们并非通过调节血尿酸盐浓度影响痛风的发生。染色体 17q23.2 区域内的 *BCAS3* 和 *TBX2*,以及 9p24.2 区域的 *RFX3*,均与痛风或血清尿酸盐浓度密切相关,提示以上 3 个基因可能是痛风的易感基因。该项研究发现了新的痛风易感基因位点,为阐明痛风的遗传和分子机制提供了理论基础<sup>[21]</sup>。

非综合征性唇腭裂是人类最常见的先天性出生缺陷之一。流行病学与遗传学研究表明,非综合征性唇腭裂是一种复杂疾病,其发生是环境因素和遗传因素共同作用的结果。目前,国外的研究人员基于全基因组关联研究,已发现 12 个非综合征性唇腭裂的遗传易感区域。为发现中国人群非综合征性唇腭裂的遗传易感区域,南京医科大学的沈洪兵、王林研究组开展了全基因组关联研究,共纳入 2521 例病例和 3122 例对照。他们发现了 1 个新的非综合征性唇腭裂易感基因区域(16p13.3),并验证出 4 个已知的易感基因区域(1q32.2、10q25.3、17p13.1 和 20q12)。染色体 16p13.3 区域的关联 SNP 位点 rs8049367 位于 *CREBBP* 和 *ADCY9* 基因间区。他们进一步证明,*ADCY9* 在非综合征性唇腭裂患者牙髓干细胞的表达水平显著上调,提示 *ADCY9* 可能是该区域的易感基因。与 rs8049367 存在强连锁不平衡的 SNP 位点 rs2262251 位于长链非编码 RNA *RP11-462G12.2* 外显子区域,*RP11-462G12.2* 与相邻的基因(*CREBBP* 和 *ADCY9*)均存在潜在的交互作用,提示 *RP11-462G12.2* 也可能是该区域的易感基因。该项研究是中国首次、亚洲最大规模的非综合征性唇腭裂全基因组关联研究,为揭示非综合征性唇腭裂遗传易感机制提供了新的证据<sup>[22]</sup>。

青少年特发性脊柱侧凸通常被认为是一种由多基因变异引起的复杂疾病。以往的全基因组关联研究已成功发现了 4 个脊柱侧凸的易感基因,但尚未有基于中国人群的研究结果。为发现中国人群脊柱侧凸的易感基因,香港中文大学的郑振耀研究组开展了全基因组关联研究,纳入来自中国人群的 4317 例病例和 6016 例对照。他们除验证出以往报道的易感基因 *LBXI*(10q24.32)外,还发现了 3 个新的脊柱侧凸易感基因区域(18q21.33、2q36.1 和 1p36.32)。染色体 10q24.32 区域的 SNP 位点 rs678741 位于 *LBX1AS1* 基因的内含子区域,该 SNP 位点位于增强子和 DNase I 超敏感区,提示该 SNP 位点可能参与调控 *LBX1AS1* 基因转录。染色体 18q21.33 区域的 SNP 位点 rs4940576 位于 *BCL-2* 基因内含子区域,该基因编码的产物对成骨细胞活性和软骨内骨化过程有重要调节作用。染色体 2q36.1 区域的 SNP 位点 rs13398147 位于 *EPHA4* 和 *PAX3* 基因间区,上述两

个基因均与神经管发育密切相关。染色体 1p36.32 区域的 SNP 位点 rs241215 位于 *AJAPI* 基因上游。*AJAPI* 能够影响细胞粘附、迁移和侵袭,而细胞粘附与骨骼生长和成骨细胞分化密切相关。该项研究发现了 3 个新的中国人群脊柱侧凸易感性位点,为揭示脊柱侧凸遗传易感机制提供了重要证据<sup>[23]</sup>。

IgA 肾病是最常见的原发性肾小球疾病。IgA 肾病的分布存在较大的地域差异,其中亚洲为高发地区。IgA 肾病存在明显的家族聚集性,表明遗传因素可能与该疾病的发生密切相关。以往的全基因组关联研究已发现 IgA 肾病的 5 个遗传易感区域(1q32、6p21、8p23、17p13.1 和 22q12)。为寻找新的 IgA 肾病遗传易感位点,中山大学附属第一医院的余学清和新加坡基因组研究院的刘建军研究组开展了全基因组关联研究,共纳入 8313 例病例和 19 680 例对照,由此发现了 3 个新的 IgA 肾病易感基因区域(3q27.3、11p11.2 和 8q22.3),并识别出已知的 8p23 区域存在的 3 个独立的关联信号。其中,3q27.3 区域的关联位点与区域内 *ST6GAL1* 基因的 mRNA 表达显著相关;11p11.2 区域的关联位点与区域内 *ACCS* 和 *EXT2* mRNA 表达显著相关。由此,提示这些基因可能是 IgA 肾病的易感基因。该项研究是中国汉族人群中迄今为止最大规模的 IgA 肾病全基因组关联研究,为阐明 IgA 肾病遗传易感机制提供了新的证据<sup>[24]</sup>。

乙型肝炎病毒(HBV)感染是全球重大的公共卫生问题。全球有超过 20 亿人口感染 HBV,其中 3.5 亿人口发展成为 HBV 慢性感染者。HBV 慢性感染将导致肝功能衰竭、肝硬化和肝癌等恶性肝脏疾病。然而,目前为止 HBV 感染慢性化的遗传机制仍不甚清楚。为发现新的 HBV 感染慢性化易感基因,复旦大学的余龙研究组联合多家研究单位,基于中国人群开展了 HBV 感染慢性化的全基因组关联研究。在发掘阶段,他们招募了来自江苏启东地区的 2514 例 HBV 慢性感染者和 1130 例健康个体。在验证阶段,使用了 4 个独立的病例对照人群(共计 6600 例 HBV 慢性感染者和 8127 例对照)。由此,重复出 7 个以往报道的与 HBV 感染慢性化相关的区域,同时还发现了 5 个新的区域。在新发现的 5 个区域中,有 4 个区域位于人类白细胞抗原(HLA)区,其标签 SNP 位点分别为 *CFB* 基因上的错义变异 rs12614、*NOTCH4*

基因上的错义变异 rs422951、*HLA-DOA* 基因上的同义变异 rs378352 和 *HLA-C* 基因上的非编码变异 rs2853953。另外一个新发现的区域位于人类基因组 20q13.1, 其标签 SNP 位点(*CD40* 基因 Kozak 序列上的变异 rs1883832)与 CD40 的表达显著相关。该项研究不仅能够完善人们对 HBV 感染慢性化的病因学理解, 而且为 HBV 感染慢性化的预防和治疗提供候选靶点<sup>[25]</sup>。

前列腺癌是全球男性发病率最高的肿瘤, 每年约有 90 万新发病例。前列腺癌的发生率在各个地区差异很大, 其中西方发达国家发病率最高, 而亚洲国家发病率最低。以往的全基因组关联研究已发现约 100 个与前列腺癌发生风险相关的位点。然而, 这些研究绝大多数是基于欧洲人群, 仅有 10 个位点是基于亚洲人群发现的。为了在亚洲人群中发现新的前列腺癌相关位点, 复旦大学的徐剑锋研究组对已有的两个亚洲人群全基因组关联研究数据进行了 Meta 分析(日本人群有 1583 例病例和 3386 例对照, 中国人群有 1417 例病例和 1008 例对照), 并对显著关联的位点进一步在 3 个独立的亚洲人群中进行了验证。结果发现 11p15.4 和 14q23.2 区域为前列腺癌新的易感基因区域, 11p15.4 区域内的 *PPFIBP2* 基因和 14q23.2 区域内的 *ESR2* 基因在前列腺癌中显著差异表达。该项研究在亚洲人群中新发现的 2 个易感基因区域, 为阐明前列腺癌潜在的分子机制提供了重要参考<sup>[26]</sup>。

已有研究报道, 近 5% 的遗传易感位点和近 17% 的遗传易感基因在多种不同性状中均存在显著关联<sup>[27]</sup>。这一现象在近期我国科学家发表的麻风病全基因组关联研究中得到了进一步的佐证。麻风病是一种由麻风分枝杆菌感染引起的慢性皮肤和神经系统疾病, 其发生发展与病人的免疫系统密切相关。全基因组关联研究已发现 11 个麻风病易感基因区域。但是, 这些已发现的位点只能解释部分的疾病易感性, 更多的遗传风险因子还有待发掘。山东大学、山东省医学科学院的张福仁研究组联合多家研究单位, 在中国人群中开展了新一轮的麻风病全基因组关联研究, 共包含 8313 例麻风病患者和 16 017 例对照个体。研究结果不仅证实了所有已报道的在中国人群中发现的显著的关联位点, 而且还发现了 6 个

新的易感位点。通过连锁不平衡分析、表达数量性状基因座/甲基化数量性状基因座分析以及蛋白-蛋白相互作用/信号通路分析等功能学研究, 他们确证了 3 个麻风病易感基因 *BATF3*、*CCDC88B* 和 *CHITA-SOCS1*。此外, 他们还系统评估了麻风病易感位点的多效性, 发现麻风病易感位点倾向于与自身免疫性疾病和炎症性疾病存在关联<sup>[28]</sup>。

全基因组关联研究对基因编码区的遗传变异位点的覆盖度有限, 是造成复杂疾病遗传度缺失的主要原因之一。银屑病是一种慢性炎症性皮肤病, 由机体免疫系统和上皮组织动态交互作用所致, 影响约 3% 的人口。目前, 全基因组关联研究已识别出 40 多个银屑病易感基因位点。然而, 这些位点对疾病的致病效应相对较弱, 仅能解释一小部分的遗传度。同时, 已报道的位点大多位于基因组非编码区, 其影响疾病发生的分子机制不明。多项基于高通量测序的研究表明, 编码区遗传变异更易导致银屑病的发生。为系统筛查与银屑病相关的编码区遗传变异, 安徽医科大学的张学军研究组开展了首项基于外显子芯片的银屑病全基因组关联研究, 该研究共纳入 17 614 例病例和 25 146 例对照。由此, 他们不仅重复出 4 个已知的易感基因区域(*TNIP1*、*NFKBIA*、*IL12B* 和 *LCE3D-LCE3E*), 还发现了 15 个新的易感基因区域(*C1orf141*、*ZNF683*、*TMC6*、*AIM2*、*IL1RL1*、*CASR*、*SON*、*ZFYVE16*、*MTHFR*、*CCDC129*、*ZNF143*、*AP5B1*、*SYNE2*、*IFNGR2* 和 3q26.2-q27)。该研究发现的 19 个易感基因区域能够解释银屑病遗传度的 1.9%。该项研究通过大样本、多阶段的病例对照设计, 发现了新的银屑病遗传易感区域, 突显了编码区常见变异对银屑病发生的遗传效应<sup>[29]</sup>。

复杂疾病中“缺失的遗传度”还可能与基因-基因或基因-环境相互作用有关。我国科学家成功发现了基因-基因相互作用可解释部分心房颤动的“缺失的遗传度”。心房颤动是临床上最常见的心率失常现象。以往的全基因组关联研究已发现了若干与心房颤动相关的遗传变异。在中国人群中, 仅有 *ZFHX3* 基因上的 rs2106261、*PITX2C* 附近的 rs2200733 和 *CAVI* 基因上的 rs3807989 三个 SNP 位点被重复验证与心房颤动密切相关。华中科技大学的王擎研究组基于以上 3 个位点, 在 3 个独立中国汉族人群(共

2020 例病例和 5315 例对照)中进行基因-基因互作分析, 结果发现 rs2106261 和 rs2200733 存在显著的交互作用。与两个位点的非风险基因型 GGCC 相比, 其风险基因型 AATT 患病的比值比最高(为 5.36), 明显高于 GGTT 和 AACC 的比值比(合并比值比为 3.31), 提示了这两个位点存在协同作用。进一步分析显示, 这两个位点在加性 × 加性模型下存在交互作用, 同时 rs2200733 指向的基因 *PITX2C* 能够负调控 miR-1 表达, 而 miR-1 又能够负调控 rs2106261 所在基因 *ZFH3* 的表达, 即 *PITX2C* 能够正向调控 *ZFH3* 的表达。有意思的是, *ZFH3* 也能够正向调控 *PITX2C* 的表达, 由此 *ZFH3* 和 *PITX2C* 形成了循环的交叉调控关系。此外, *ZFH3* 和 *PITX2C* 均能够调控 *NPPA*、*TBX5* 和 *NKX2.5* 基因的表达。以上证据提示, 这种循环的交叉调控基因表达模式是复杂疾病遗传学中基因-基因互作的分子基础<sup>[30]</sup>。

全基因组关联研究基于“常见疾病-常见变异”学说, 仅关注等位频率较高的常见遗传变异, 这可能是导致复杂疾病遗传度缺失的又一重要因素。越来越多的证据显示, 低频变异或罕见变异也是常见疾病发生的重要因素, 即“常见疾病-罕见变异”学说。最近, 我国科学家成功发现了与慢性乙肝、肺癌显著相关的低频变异, 进一步证明了低频变异或罕见变异在复杂疾病中的重要作用。乙型肝炎病毒 (HBV) 是一种具有包膜的 DNA 病毒, 其有非常严格的宿主种属和细胞类型特异性。在过去的 50 年中, 研究人员为寻找 HBV 受体付出了相当多的努力。近期的体外实验证实, NTCP(也被称为 SLC10A1) 是 HBV 的受体。值得关注的是, 体外证据还显示, SLC10A1 的 p.Ser267Phe 变异将导致 HBV 受体功能丧失。中山大学的王一鸣研究组采用了遗传关联的方法, 在 1899 例慢性乙肝患者和 1828 例对照中对 p.Ser267Phe 变异进行了检测。研究发现, 该变异在对照个体中富集。此外, 携带该变异的慢性乙肝患者, 发生急性肝功能衰竭的概率显著降低。通过计算他们发现, 该变异能够解释大约 3.2% 的 HBV 感染慢性化的遗传度。通过结构建模进一步发现, p.Ser267Phe 变异可能干扰了与配体的结合能力, 从而阻止了 HBV 进入细胞。该项研究不仅从遗传学角度证明了 SLC10A1 是 HBV 感染人类细胞的受体,

而且还发现了 p.Ser267Phe 变异可显著降低 HBV 感染慢性化的发生<sup>[31]</sup>。

肺癌是全球最常见的恶性肿瘤之一, 其致死率在所有恶性肿瘤中居于首位。吸烟是肺癌发生的主要因素, 但不同的个体对肺癌存在不同程度的遗传易感性。目前, 全基因组关联研究已经成功发现了一些与肺癌发生密切相关的常见变异, 但是这些变异仅能解释部分的肺癌易感性。为评估低频或罕见变异在肺癌中的作用, 南京医科大学的沈洪兵研究组对 1348 名肺癌患者和 1998 名对照个体进行了外显子组芯片分型, 采用基于单个位点和基于基因的两类方法进行了遗传关联分析, 并在 4 个独立人群(包含 4699 例肺癌患者以及 4915 例对照)中进行了验证, 结果发现了 3 个低频错义变异(6p21.33 区域 *BAT2* 基因上的 rs9469031、6p21.33 区域 *FKBPL* 基因上的 rs200847762 和 20q11.21 区域 *BPIFB1* 基因上的 rs6141383)与肺癌的发生风险显著相关。其中, rs9469031 和 rs6141383 与肺癌的发生年龄密切相关。上述 3 个基因在肺癌组织和癌旁组织中均存在显著的差异表达。基于基因的分析方法提示 *FKBPL* 基因与肺癌的发生风险相关。其中, 该基因内的两个罕见变异(rs200847762 与 rs117160266)是独立的关联信号, 由此推测 *FKBPL* 可能是 6p21.33 区域内的易感基因。该项研究证实了低频变异在肺癌易感性中发挥重要作用<sup>[32]</sup>。

随着高通量测序技术的发展, 基于全外显子组测序或全基因组测序技术筛查疾病相关的遗传变异成为现实。高通量测序技术能够全面系统地发现基因组的所有变异, 因此, 与 SNP 芯片技术相比, 更易于发现功能性位点或与疾病密切相关的罕见变异。老年黄斑变性是导致老年人出现不可逆中枢性失明的首要原因, 新生血管性老年黄斑变性是其中重要的一种亚型。为探索外显子区的遗传变异与新生血管性老年黄斑变性间的关联, 北京大学人民医院的黎晓新、赵明威和复旦大学的周鹏研究组合作, 对老年黄斑变性开展了全外显子组测序研究。他们对 216 例病例和 1553 例对照进行了全外显子测序, 筛选出其中显著关联的 SNP 位点后在 5 个独立人群中进行了验证(分别来自中国、日本和新加坡总计 3772 例病例与 6942 例对照), 最终鉴别出 1 个新的老年



黄斑变性的易感位点 (*UBE3D* 基因错义变异 rs7739323)。SKAT-O 分析结果显示, 新生血管性老年黄斑变性患者中 *UBE3D* 基因区域的遗传变异存在富集现象, 进一步提示 *UBE3D* 可能是该疾病的易感基因。功能学研究发现, 与野生型小鼠相比, *UBE3D*<sup>+/-</sup> 杂合子小鼠视网膜色素上皮细胞的绒毛更为稀疏, 有更多色素颗粒沉着, 且视网膜电图异常。该项研究证明了全外显子组测序技术可有效发现复杂疾病的易感基因位点<sup>[33]</sup>。

乳腺癌是女性最常见的恶性肿瘤, 约 10% 的患者存在乳腺癌家族史。然而, 仅 10%~15% 的家族性乳腺癌可归因于高外显率易感基因 *BRCA1* 或 *BRCA2* 的胚系突变; 5% 的家族性乳腺癌可归因于中外显率易感基因 (如 *PALB2*、*ATM*、*CHEK2*、*BRIP1* 和 *RAD51C*) 的突变。为寻找新的乳腺癌易感基因, 北京大学临床肿瘤学院乳腺中心的解云涛与中国农业大学生物学院的楼慧强研究组对 9 例不携带 *BRCA1/2* 基因胚系突变的早发家族性乳腺癌患者进行了全外显子组测序, 成功筛选出 3 个潜在的乳腺癌易感基因 *RECQL*、*TLL2* 和 *VSIG2*。其中, *RECQL* 包含两个重要的结构域 (解螺旋酶结构域和 RecQ 羟基末端结构域), 对维持基因组稳定性发挥重要作用。在 *RECQL* 基因敲除的小鼠模型中, 表现为基因组不稳定、对电离辐射敏感性增加、且 DNA 损伤程度加重。进一步对 *RECQL* 外显子区在 439 例不携带 *BRCA1/2* 基因突变的家族性乳腺癌病例中进行 Sanger 测序, 新发现了 9 个潜在的致病性胚系突变。该项研究首次报道了 *RECQL* 是乳腺癌潜在的易感基因, 提示了对 *BRCA1/2* 突变阴性的乳腺癌患者进行 *RECQL* 突变筛检具有重要的临床意义<sup>[34]</sup>。

### 1.3 癌症体细胞突变研究

癌症是一类特殊的复杂疾病。尽管遗传因素和环境因素均可导致癌症的发生风险增加, 但癌症的最终发生是由体细胞突变的累积直接导致的。因此, 体细胞突变在癌症发生过程中处于中心地位<sup>[35]</sup>。遗传因素和环境因素通过增加体细胞突变发生频率等方式, 增加癌症的发生风险<sup>[35]</sup>。体细胞突变是随机产生的, 所以绝大多数体细胞突变均不致癌, 仅有少数体细胞突变与肿瘤的发生密切相关, 被称之为

“驱动突变” (Driver mutation)。癌症驱动突变或驱动基因的发现, 有助于对肿瘤病人的早期诊断、个体化治疗和预后评估。

高通量测序技术的发展, 为肿瘤基因组的研究提供了便利。通过高通量测序技术, 我国科学家陆续发现了 T 细胞淋巴瘤、食管鳞癌、胃癌和骨髓增生异常综合征等肿瘤的驱动基因。自然杀伤/T 细胞淋巴瘤 (NKTCL) 是一种常见的非霍奇金淋巴瘤。该肿瘤在男性高发, 且与 EB 病毒感染密切相关。NKTCL 病程发展迅速, 在亚洲和南美发病率较高。然而, 目前对 NKTCL 发病的分子机理尚不清楚。因此, 对 NKTCL 发病机制的研究意义重大, 可为 NKTCL 靶向治疗提供理论基础。上海交通大学系统生物医学协同创新中心的陈赛娟、陈竺和赵维莅研究组对 25 例 NKTCL 进行了全外显子组测序, 并进一步对另外的 80 例 NKTCL 进行了靶向测序。研究发现, 在 NKTCL 中, RNA 解旋酶 (*DDX3X*)、肿瘤抑制因子 (*TP53* 和 *MGA*)、JAK-STAT 信号通路分子 (*STAT3* 和 *STAT5B*) 和表观遗传修饰相关基因 (*MLL2*、*ARID1A*、*EP300* 和 *ASXL3*) 发生了高频突变。他们进一步证明 *DDX3X* 突变能够降低其 RNA 解旋酶活性, 并使该基因抑制细胞周期进程的能力丧失; *DDX3X* 突变能够引起 NF- $\kappa$ B 和 MAPK 通路活性显著提高, 并且能够引起该基因抑制 ERK 磷酸化和降低细胞核内 RelB 的功能丧失。临床相关性分析发现, 携带 *DDX3X* 突变的病人其预后显著较差。该项研究为人们进一步了解 NKTCL 的发病机制提供了帮助<sup>[36]</sup>。

食管鳞状细胞癌是全球最常见的恶性肿瘤之一。我国的食管鳞状细胞癌的发病率很高, 全球约 70% 的病例发生在中国, 而太行山一带更是我国食管鳞癌发病率最高的地区。在我国, 食管鳞状细胞癌的致死率在所有恶性肿瘤中排名第四, 5 年生存率仅为 10%~25%。为进一步研究食管鳞状细胞癌的分子机理, 山西医科大学的崔永萍研究组、中国医学科学院詹启敏研究组等多家机构研究人员对 104 例来自太行山地区的食管鳞状细胞癌开展了高通量测序研究。在合并以往 88 例食管鳞状细胞癌高通量测序数据的基础上, 他们发现 47% 的食管鳞状细胞癌呈现出 APOBEC 介导的突变谱, 说明在食管鳞状细胞癌中 APOBEC 催化的脱氨基作用是引起 DNA 损伤

的重要因素。他们还发现, *CBX4* 和 *CBX8* 基因在食管鳞状细胞癌中存在局部扩增; *AJUBA*、*ZNF750*、*PTCH1* 和染色质重构相关基因(*CREBBP* 和 *BAP1*) 存在高频的失活型突变。后续的功能实验进一步证实 *CBX4*、*CBX8*、*AJUBA* 和 *ZNF570* 基因或基因上的突变与肿瘤的发生密切相关。此外, Hedgehog 和 PI3K 信号通路在 60% 的食管癌中被激活, 提示针对这些通路的靶向治疗有望提高食管癌的治疗效果。该项研究不仅全面系统分析了食管癌的突变谱, 而且还发现了若干新的潜在的早诊标志物和治疗靶点<sup>[37]</sup>。

胃癌也是世界最常见的恶性肿瘤之一。多种环境风险因素可能导致胃癌的发生, 如高盐饮食、吸烟和幽门螺旋杆菌感染事件。与其复杂的病因类似, 胃癌在临床和病理学上表现出非常明显的异质性, 各种亚型的 5 年生存率差异很大。目前标准的治疗方法忽视了不同亚型胃癌的异质性。发现潜在的、临床可行的胃癌治疗靶点, 可为胃癌的个体化治疗提供直接的帮助。天津医科大学的郝希山研究组对 78 例组织分级和解剖位置相异的胃癌进行了全外显子组测序, 同时还对另 2 例胃癌病人的 3 个原发灶和 2 个转移淋巴结进行了全基因组测序。他们使用 SciClone 软件预测每例胃癌的内部克隆数, 由此将胃癌分为高克隆和低克隆两组。高克隆组胃癌病人年龄更大, 携带 *TP53* 突变, 突变集中表现为 C-G 转换, 生存期明显较短; 低克隆组胃癌病人年龄更小, 携带 *ARID1A* 突变, 生存期明显较长。通过对 2 例胃癌病人的多个样本进行系统进化树分析, 证实了胃癌转移是一个克隆进化过程, 提示基因缺陷存在累积过程, 可能需要采取组合治疗的方法。在 216 例胃癌病人中对高频突变基因进一步检测发现, 同源重组 DNA 修复系统、Wnt 和 PI3K-ERBB 信号通路的成员发生突变频率最高。其中, 药物作用靶点 *NRG1* 和 *ERBB4* 在 10% 的胃癌病人中发生突变。此外, *BRCA2* 在 8% 的胃癌病人中发生突变, 且突变个体的生存时间显著延长。这一结果与 TCGA(The Cancer Genome Atlas)数据集中胃癌病人的结果类似。该项研究结果为我国胃癌人群提供了新的候选标志物, 同时也为胃癌的个体化治疗提供了理论基础<sup>[38]</sup>。

骨髓增生异常综合征(Myelodysplastic syndrome,

MDS)是一类异质性很强的造血系统肿瘤疾病。其典型特征表现为造血功能失效和外周血血球减少, 多达 30% 的 MDS 会发展成为急性髓性白血病。目前, MDS 患者的病程进展机制仍然未知。为揭示 MDS 病人病程进展的分子机理, 上海交通大学附属第六人民医院的李晓研究组对 3 例 MDS 患者样本(疾病进展前后的样本及其配对的正常口腔黏膜上皮细胞样本)进行了全外显子组测序, 同时还对 13 例 MDS 患者的样本进行了靶区域测序, 结果发现 *ROBO1* 和 *ROBO2* 在 37.5%(6/16)的进展性 MDS 中发生突变。在 193 个 MDS 病人中的验证结果也表明, 有 10.4% 的 MDS 病人携带 *ROBO1* 或 *ROBO2* 突变。此外, *ROBO1* 和 *ROBO2* 基因座上的拷贝数缺失和杂合缺失经常发生在进展性 MDS 病人或携带 *ROBO* 基因突变的病人样本中。体外实验表明, 过表达 *ROBO1* 或 *ROBO2* 将导致白血病细胞生长和迁移能力减弱。这一效应在 *ROBO* 突变或 *ROBO-SLIT2* 信号通路受损的情况下将会丧失。该项研究证明了 *ROBO* 突变能够影响 MDS 病人的预后, 并提示 *ROBO-SLIT2* 信号通路在 MDS 病程进展中发挥关键作用<sup>[39]</sup>。

病毒 DNA 整合进入人类基因组是肿瘤体细胞突变中的一种特殊形式。宫颈癌是女性常见的恶性肿瘤之一。HPV DNA 与宿主 DNA 的整合是导致宫颈癌形成的最重要的诱导因素。发现人类基因组上的 HPV 整合热点, 将有助于加深对 HPV 引发宫颈癌作用机理的理解。华中科技大学同济医学院的马丁研究组通过全基因组测序和自主开发的高通量病毒整合筛查技术(High-throughput viral integration detection, HIVID), 在 26 个宫颈上皮内瘤、104 个宫颈鳞状细胞癌和 5 株宫颈癌细胞系中发现了 3667 个 HPV 整合位点。他们不仅重复出以往报道的存在高频整合的基因, 而且还重新评估了这些基因上 HPV 的整合频率。此外, 他们还发现了一些全新的存在高频整合的基因, 如 *HMGA2*、*DLG2* 和 *SEMA3D*。HPV 整合的致癌机制可能是通过激活或抑制相关蛋白, 导致细胞恶性转变的几率增高。研究人员还提出了一种新的病毒整合模型: HPV 感染人体后造成人类基因组不稳定, 从而激活微同源序列介导的 DNA 修复通路, HPV DNA 抢在微同源序列前将自身的 DNA 融合进断裂的人类基因组中。该项研究深

入探索了 HPV 整合导致宫颈癌的作用机理,将有助于宫颈癌的早期个体化治疗和预后评估<sup>[40]</sup>。

携带体细胞突变的肿瘤细胞 DNA,有可能会被释放进入循环系统,从而成为肿瘤循环游离 DNA。对肿瘤循环游离 DNA 的检测和分析使得实体瘤的液体活检成为可能。尽管人们认识到其拥有巨大的临床应用潜力,肿瘤循环游离 DNA 的许多生物学特性仍旧未知,如血浆 DNA 分子大小等问题。许多研究显示肿瘤病人血浆中的 DNA 分子完整性更好,然而其他一些研究表明肿瘤病人血浆中的 DNA 分子更短。来自香港中文大学的卢煜明研究组对 90 例肝癌患者、67 例慢性乙肝患者、36 例乙肝相关肝硬化患者和 32 例健康对照个体的血浆 DNA 分子大小图谱进行了详细分析。他们使用高通量测序技术对血浆 DNA 的分子大小进行了测定。对于肿瘤来源的血浆 DNA 分子,进一步使用染色体臂水平 z-score 分析方法(CAZA)鉴定其分子大小。通过比较发现,在肝癌患者的血浆中存在异常长的和异常短的 DNA 分子,其中短的 DNA 分子倾向于携带肿瘤相关的拷贝数变异。此外还发现,在肝癌患者的血浆中线粒体 DNA 含量增高,而且这些 DNA 分子要比血浆细胞核中的 DNA 分子短得多。该项研究使人们对肿瘤循环游离 DNA 分子大小图谱有了更进一步的理解,同时进一步提高了研究人员利用血浆 DNA 作为分子诊断工具的能力<sup>[41]</sup>。

癌症的体细胞突变不仅与癌症的发生密切相关,也可能与癌症的复发和耐药相关。对复发和耐药相关体细胞突变的研究,将为癌症病人的个体化治疗和预后判断提供理论基础。急性成淋巴细胞性白血病(Acute lymphoblastic leukemia, ALL)的复发是导致患儿死亡的最重要因素。在化疗药物中,巯嘌呤类药物(6-巯基嘌呤和 6-巯代鸟嘌呤)是 ALL 综合治疗的核心药物。这些药物均为无生物活性的药物前体,需要通过体内嘌呤补救合成途径转变为有细胞毒性的巯鸟嘌呤核苷酸。因此,嘌呤稳态发生变化会对治疗效果产生影响。为研究儿童 ALL 耐药和复发的分子机制,上海交通大学的周斌兵研究组对 16 例 ALL 患者诊断时、缓解期和复发期的样本进行了全外显子组测序,结果发现编码一种限速嘌呤合成酶的 *PRPS1* 基因上的突变仅在复发样本中出现。进一

步在更多的复发样本中对该基因进行检测,证实该基因发生突变的频率为 6.7%(24/358)。携带 *PRPS1* 突变的个体在治疗早期复发,突变的 ALL 克隆在临床复发前便已经以指数方式增长。功能学研究证实 *PRPS1* 基因突变具有耐药性,能够激活 PRPS1,降低对嘌呤核苷酸从头合成的反馈抑制,同时竞争性抑制巯嘌呤的激活。尤其值得注意的是,嘌呤核苷酸从头合成抑制剂洛美曲索(Lometrexol)可以有效消除 *PRPS1* 突变引起的耐药性。该项研究结果突显了激活嘌呤核苷酸从头合成途径在巯嘌呤类药物耐药中的重要作用,为 ALL 耐药和复发提供了治疗策略<sup>[1]</sup>。

#### 1.4 遗传学研究新技术

以高通量测序技术和基因组编辑技术为代表的遗传学研究新技术,极大地推进了遗传学研究的进程。我国科学家在遗传学研究新技术的革新和应用中也发挥了重要作用。

##### 1.4.1 基因组编辑技术

以 CRISPR/Cas9 技术为代表的基因组编辑技术是近年来生物技术发展的热点,这不仅使在小鼠、大鼠、斑马鱼等各种模式生物中针对特定基因变异形式快速建立动物模型成为可能,而且还极大地推进了对生命科学各个领域的理论探索。例如,上海交通大学的吴强研究组通过 CRISPR 遗传编辑技术反转 CTCF 位点改变基因组拓扑结构和增强子与启动子的功能,阐明了基因组三维拓扑结构的形成及其调控基因表达模式的分子机制;清华大学的朱昕研究组、以色列特拉维夫大学的 Ebenstein 研究组及中国科学院微生物研究所的娄春波研究组合作开发的 CATCH 技术,实现了大型基因簇一步法靶向克隆,首次在体外应用 CRISPR/Cas9 技术实现对上百 kb 的基因组片段的靶向克隆。

线性的 DNA 分子在细胞核中以立体的三维结构呈现,且这种有规律的组织形式被认为是基因组进行表达调控的结构基础。例如,人类基因组包含十几个绝缘子结合蛋白 CTCF 结合位点(CBS),CTCF 蛋白通过结合 CBS 参与建立基因组 DNA 的远程交互网络。为阐明 CTCF 介导特异性染色质相互作用的分子机制,上海交通大学的吴强研究组利用 CRISPR/Cas9 介导的 DNA 片段编辑技术实现

CTCF 位点的原位反转,进而发现在原钙粘蛋白基因簇中 CTCF 识别其靶向 DNA 序列是具有方向性的,这种方向性决定了染色质环化的方向性。该研究还在珠蛋白基因簇中通过进一步的基因组编辑,发现位于上游正向的 CTCF 位点与位于下游反向的 CTCF 位点的 DNA 序列能够进行特异性长距离染色质环化,形成特异的染色质高级拓扑结构域,并通过增强子与基因启动子的特异性相互作用等形式调控基因的表达。计算生物学研究揭示这一规律在全基因组中具有普适性。该项研究提出了与前人基于报告基因检测方法认为增强子不具有方向性相反的新概念和观点,为进一步认识三维基因组的结构和功能以及疾病的发生发展奠定了理论基础<sup>[42]</sup>。

大片段基因簇的克隆是微生物功能基因组研究中最为关键的步骤。基于 PCR 等传统技术的克隆方法不仅受到克隆片段长度的限制,而且还容易产生突变,且效率不高。因此,亟需克服上述技术限制,实现快速简单的大尺度基因簇克隆。清华大学的朱昕研究与多家单位合作利用体外转录制备的 sgRNA 和表达纯化的 CRISPR-Cas9 核酸内切酶,精确地切割目标基因簇片段的两侧,并且通过 Gibson 组装方法将切割下来的大片段与目标载体进行连接,达到进一步克隆的目的。该方法成功实现了 50~150 kb 的大肠杆菌基因组片段的克隆,并在枯草芽孢杆菌、链霉菌中成功克隆了目标基因簇。该技术的克隆效率可与普通短片段克隆相当,可极大简化对能够表达具有高经济附加值生物大分子(比如抗菌素和抗肿瘤药物)的基因簇的克隆步骤,节省时间并降低成本,推动合成生物学等多个领域的发展<sup>[43]</sup>。

#### 1.4.2 基因组测序技术

开展单细胞研究是当前生命科学研究的重要趋势。一些关键的生命科学问题需要在单细胞水平上来分析和解答。在单细胞的基因组学研究中,由于可以获得的基因组 DNA 极少,因此必须通过全基因组的扩增技术将基因组 DNA 进行扩增。但是,扩增的不均匀性和扩增错误可导致扩增后的 DNA 样本发生较大的改变,从而导致对单细胞中 CNV 的结果和 SNV 位点的错误判读。目前,兼并寡核苷酸引物 PCR(Degenerate oligonucleotide-primed PCR, DOP-

PCR)、多重置换扩增(Multiple displacement amplification, MDA)和多次退火环状循环扩增(Multiple annealing and looping-based amplification cycles, MALBAC)等技术实现了单个细胞水平的全基因组分析。在这些技术的基础上,北京大学的黄岩谊研究组和谢晓亮研究组合作开发了一种用于单细胞测序的全基因组扩增新技术——乳液全基因组扩增(emulsion Whole-Genome Amplification, eWGA),该方法大幅度提高了全基因组扩增的均匀性和准确性,使得通过单细胞全基因组测序实现同时检测出小片段 CNV 和高精度的单核苷酸变异(SNV)成为可能。该方法利用微流控技术原理,将一个细胞的基因组片段分散在包含有几十万个皮升级微型反应液滴的乳液中,每个液滴中大致含有一个待扩增的 DNA 片段和用于等温多重置换扩增(MDA)的其他原料分子。当对整个乳液体系进行等温扩增时,可以在每个分隔的液滴内进行独立的扩增反应,并在这一受限体系内相继达到饱和。这样可以显著改善扩增的均匀性。该方法应用于单个肿瘤细胞的 CNV 检测,最高可以达到 250 kb 的分辨率<sup>[44]</sup>。

5-甲基胞嘧啶修饰(5mC)是真核细胞基因组 DNA 的重要表观遗传修饰形式。在 TET 家族酶的作用下,5mC 可以逐步氧化生成不同的 DNA 修饰形式,包括 5-羟甲基胞嘧啶(5hmC)、5-醛基胞嘧啶(5fC)和 5-羧基胞嘧啶(5caC)。对于这些新型的 DNA 修饰形式及其功能的研究是目前生命科学研究的前沿领域,但如何在基因组中实现 5fC 的精准定位是其中主要的难题之一。此前,通过抗体、亚硫酸氢盐处理等方法已经能够实现 5fC 的测序工作,但是这些传统的方法均具有明显的局限性,如依赖抗体的 5fC 测序技术分辨率较低,较难实现单碱基分辨率的 5fC 检测,而依赖亚硫酸氢盐处理的 5fC 测序技术则会导致严重的 DNA 降解、且成本昂贵。北京大学生命科学学院的伊成器研究组和美国芝加哥大学的何川研究组合作,通过化学生物学的研究手段,利用小分子化合物对 5fC 的特异性化学标记,开发出了不依赖于亚硫酸氢盐处理的 5fC 单碱基分辨率测序新技术——fC-CET(Cyclization-enabled C-to-T transition of 5fC)。利用 fC-CET 技术,他们成功实现了对小鼠胚胎干细胞基因组中 5fC 位点的鉴定,并发

现 5fC 富集的基因组区域比 5hmC 富集的区域其转录更为活跃。需要特别指出的是,由于 fC-CET 技术不会导致明显的基因组 DNA 降解,因而只需小量核酸样本即可实现全基因组 5fC 修饰图谱的测序分析。该项研究摒弃了传统的亚硫酸氢盐处理的方法,实现了在全基因组水平对 5fC 进行单碱基分辨率的测序分析<sup>[45]</sup>。

测序技术的发展,离不开针对测序数据的分析方法开发。华大基因研究院的研究团队组装完成了迄今最完整的单倍体水平的人类二倍体基因组序列。虽然通过全基因组测序技术可以获得近乎整个基因组的 DNA 序列信息,但是这些序列信息在二倍体基因组中哪个单倍型上分布的信息是不完整的。完整解读每个染色单体的 DNA 序列及其遗传变异信息具有重要的生物学意义。该研究在不使用人类基因组参考序列的前提下,仅凭借高通量测序数据,通过全基因组鸟枪测序法结合 Fosmid-pooling 策略的分级组装方法,组装出了单倍体水平的二倍体基因组序列。通过对第一个亚洲人基因组 YH 1 号进行重新测序,组装出 5.15 Gb 的二倍体基因组序列,其单体型 N50 长度达到 484 kb,全面分析和展示了人类二倍体基因组完整的序列和变异信息。这些重要数据和信息可以被用于同其他组学数据的综合分析,帮助揭示基因表达调控等多方面的生物学机理。此外,该研究也为解决其他杂合度高或多倍体物种的基因组序列组装问题提供了新的技术途径<sup>[46]</sup>。

#### 1.4.3 其他新技术和新方法

RNAi 是由 siRNA 介导的转录后基因沉默现象,目前已被广泛应用于基因功能的研究,且有可能成为小核酸药物用于疾病的治疗。siRNA 可由化学方法合成后转入细胞,或者通过 DNA 载体在细胞内表达 shRNA(Short hairpin RNA)后由 Dicer 酶加工生成 siRNA。但 shRNA 的脱靶效应(Off-target effect)以及对细胞内源 miRNA 的竞争性抑制所带来的毒副作用限制了 RNAi 技术的应用。中科院上海生科院生物化学与细胞生物学研究所的吴立刚研究组对 siRNA 前体的加工机制进行了深入研究,并在此基础上发明了一种比传统 shRNA 效率更高、脱靶作用更少的 saiRNA(Single-stranded, Ago2-processed in-

terfering RNA)。saiRNA 的双链区较短,为 16~18 个碱基,且靶向区延伸入顶端环序列,其加工不依赖于 Dicer,而由 RNAi 的核心蛋白 Ago2 直接在双链区 3'侧第 10 位碱基进行切割,并在细胞内被进一步加工生成长度为 24~27 个碱基的单链 siRNA。在 saiRNA 的 3'末端融合丁型肝炎病毒(HDV)核酶后,利用核酶的高效自切割活性准确地在 saiRNA 的 3'端产生了两个碱基的悬垂,大大增强了与 Ago2 蛋白的结合和对靶基因的沉默效率。由于核酶增强的 saiRNA 仅产生一条单链 siRNA,并特异性与 Ago2 结合,因此具有更小的脱靶作用和对细胞内源 miRNA 的竞争性抑制作用。这一新型的 RNAi 载体或将成为生命科学基础研究和疾病治疗应用的重要手段和工具<sup>[47]</sup>。

CNV 是人类个体间遗传多样性的另一来源,可为解释人类个体间生理和体质特征的多样性提供新的视角。中科院上海生命科学研究院计算生物学研究所的徐书华研究组设计了一种搜寻人群特异性 CNV 的新方法——WinXPCNVer。他们使用该方法探索了与藏族人群适应高原低氧环境相关的 CNV 位点。在临近 *EPAS1* 基因的非编码区域检测到一段藏族人群特异的约 3.4 kb 长度的缺失。该缺失的频率在藏族中高达 90%,其中纯合缺失个体的频率达到了 50%,远远高于世界其他人群中该缺失区域的频率。研究发现该缺失区域包含有组蛋白增强子,并且与之前报道的血红蛋白浓度相关联的 SNP 呈强连锁不平衡状态。一系列的证据都支持该 *EPAS1* 基因调控区的 CNV 在藏族人群低氧适应中发挥重要功能,为后续开展藏族人群高原适应性的分子机制研究指明了方向。更为有趣的是,这个 CNV 不存在于丹尼索瓦人的基因组中,提示现代人祖先与非现代人祖先之间的基因交流格局和适应性进化机制比学界目前所理解的要更加复杂<sup>[48]</sup>。

现代人类祖先走出非洲逐渐迁徙扩散到世界各地的过程中,曾经跨越不同的自然地理区域、经历多样的气候、接触不同的传染性疾病。因此,现存的不同人群曾受到了不同的自然选择压力的洗礼,进而在各自的基因组上留下不同的遗传适应性标记。对人类群体间此类存在自然选择差异的遗传位点进行鉴定和分析,将有助于人们认识人类的适应性进

化历史及其生物学功能。虽然科学家们已经开发了多种方法来鉴定受到自然选择的遗传位点,但是现有的手段并不能对人群间的选择性差异进行很好的度量。中科院上海生命科学研究院计算生物学研究所何云刚和复旦大学的金力研究组合作开发了一种基于统计模型的用于检测和估算人群间经受不同程度的自然选择的新方法。基于遗传漂变和自然选择的概率模型,通过人群间等位基因频率比值的对数值,可以对人群间选择系数的差异进行估算。利用全基因组遗传变异数据,可以很好地实现对人群间选择系数差异的度量并给出置信区间。这一新方法被用于对汉族和藏族人群的全基因组数据分析,有效验证了已知的汉族和藏族间存在自然选择差异的 *EPAS1* 和 *EGLN1* 等基因位点。可以预期,该方法将会被广泛用于检测和量化人群间存在的自然选择差异性位点,揭示新的适应性进化位点及其生物学机理<sup>[49]</sup>。

## 2 疾病发生和发展的表观遗传学研究

### 2.1 DNA 甲基化研究

DNA 甲基化是一种重要的表观遗传修饰方式,是调控基因表达的重要机制。DNA 甲基化已被证明在多种生理、病理过程中发挥重要作用。生殖细胞对于遗传信息的世代传递和维持物种的延续至关重要。原始生殖细胞(Primordial germ cell, PGC)产生于胚胎发育早期,是发育成为成熟的精子和卵母细胞的前体细胞。目前,研究人员对小鼠 PGC 已有相对全面的了解。但是,对人类 PGC 的表达调控机制尚不了解。北京大学第三医院的乔杰研究组采用单细胞 RNA 测序技术,对不同发育阶段的人类 PGC 进行了分析,发现人类 PGC 具有独特的基因转录模式。他们还对 PGC 的 DNA 甲基化谱进行了分析,发现了全局层面的脱甲基化现象。怀孕早期,PGC 中的 DNA 甲基化水平较高,大部分基因的表达受到抑制;孕期 10~11 周时,PGC 中的大部分 DNA 甲基化会被擦除,甲基化水平降至最低点。这也是人类已知的所有细胞类型中,DNA 甲基化水平最低的细胞,说明 PGC 的甲基化组具有明显的独特性。被保留下来的甲基化位点大部分位于重复序列元件上,尤其是微卫星及一些进化上年代较近的元件,为人类隔代遗传现象的表观遗传学研究提供了线索。该

项研究为研究人类生殖细胞的表观遗传调控、早期胚胎全能性的建立以及 DNA 甲基化的隔代遗传等科学问题提供了理论基础<sup>[3]</sup>。

### 2.2 lncRNA 研究

lncRNA 是一类核苷酸长度大于 200 个核苷酸(nt)、不编码蛋白质的 RNA 转录物。最新的研究揭示,lncRNA 可在表观遗传水平、转录水平和转录后水平调控基因表达。lncRNA 的表达异常可引起基因组印迹、基因转录调控等重要生物学过程的紊乱,从而可能导致细胞恶性转化及肿瘤发生。我国科学家陆续发现,lncRNA 能够通过调控组蛋白甲基化或者与蛋白质相互作用,从而调控基因的表达或影响蛋白的活性。这些研究进一步证明了 lncRNA 在肿瘤发生过程中的重要作用。

上海交通大学医学院的张何朋、范先群和葛盛芳研究组发现,lncRNA ROR 可通过与组蛋白 G9A 甲基转移酶竞争结合靶基因 *TESC* 的启动子区,从而抑制该区域中的组蛋白 H3K9 甲基化水平,实现对 *TESC* 基因表达的调控。ROR 在 DNA 损伤和胚胎干细胞自我更新中发挥作用,但其在肿瘤中的作用未见有相关报道。在该项研究中,他们首先通过表型实验,证实了 ROR 具有致癌功能,然后通过表达谱分析,发现 *TESC* 基因的表达水平受到 ROR 调控。研究还发现 ROR 的异常表达能够引起 *TESC* 启动子区中的 H3K9 甲基化状态的改变,进而影响 *TESC* 的表达。有趣的是,ROR 并不直接与 G9A 发生直接相互作用,而是通过与 G9A 竞争结合靶基因 *TESC* 的启动子区,从而引起 H3K9 甲基化状态的改变。该项研究不仅揭示了 ROR 在肿瘤发生过程中发挥的重要作用,而且还揭示了一种全新的调控组蛋白修饰的方式<sup>[50]</sup>。

lncRNA 已被报道能够调控多种肿瘤相关的进程,如细胞增殖、凋亡、转移、衰老和耐药等,然而其在肿瘤相关炎症中扮演的角色却一直未知。NF- $\kappa$ B 家族是一类在炎症和免疫反应中发挥重要作用的转录因子。中山大学的宋尔卫研究组首次发现了 lncRNA NKILA 能够通过结合 NF- $\kappa$ B:IkB 复合体而抑制 NF- $\kappa$ B 活性,从而抑制肿瘤相关炎症的发生。他们以乳腺癌为研究模型,发现 NKILA 在各种炎症刺激因子的诱导下均显著上调表达,并且其上调表

达是由 NF- $\kappa$ B 信号通路引起的。进一步研究发现, NKILA 能够负调控 NF- $\kappa$ B 信号通路。NKILA 通过结合 NF- $\kappa$ B:I $\kappa$ B 复合物, 掩盖 I $\kappa$ B 的磷酸化模体, 进而抑制了 IKK 诱导的 I $\kappa$ B 的磷酸化和 NF- $\kappa$ B 的激活。临床相关性分析显示, NKILA 的低表达与乳腺癌的转移以及病人不良的预后显著相关。该项研究证实了 lncRNA 能与蛋白相互作用, 从而调控肿瘤相关炎症, 引起肿瘤的发生<sup>[4]</sup>。

lncRNA 在肿瘤相关的细胞衰老中扮演的角色一直未知。北京大学医学部的毛泽斌研究组发现, lncRNA SALNR 能够通过结合 NF90, 进而抑制细胞的衰老。他们首先通过对人成纤维细胞进行 lncRNA 测序, 发现了衰老细胞中 968 个显著下调和 899 个显著上调的 lncRNA。其中, 存在 1 个 lncRNA 在细胞衰老过程中以及结肠癌癌变前均显著下调, 他们将其命名为 SALNR。研究发现, 在人成纤维细胞中过表达 SALNR, 可以延缓细胞的衰老。体外结合实验证实, SALNR 可与 RNA 结合蛋白 NF90 直接相互作用。抑制 NF90 将导致衰老相关的 miRNA 显著上调表达。由此, 该项研究揭示了 lncRNA 在肿瘤相关的细胞衰老过程中的作用机理<sup>[51]</sup>。

Hotair 是位于 Hoxc 基因簇上的 lncRNA。以往报道显示, Hotair 能够促进多种肿瘤的发生。然而, Hotair 的致癌机制仍未明确。香港中文大学的李刚和张锦芳研究组开展了对 Hotair 致癌机制的研究。他们首先发现, 在肝癌组织和细胞系中, Hotair 与 miR-218 的表达均呈现负调控关系。进一步的研究发现, 沉默 Hotair, 可通过 EZH2 促进 miR-218 的表达, 进而下调 miR-218 的功能性靶基因 Bmi-1, 并激活下游抑癌基因 P16 和 P14, 最终在体外和在体水平抑制肿瘤的生长。临床相关性分析进一步证实了以上发现。由此, 该研究揭示了 Hotair 的致癌机制<sup>[52]</sup>。

### 2.3 microRNA 研究

microRNA(miRNA)是一类长约 20~24nt 的内源性单链非编码小 RNA。miRNA 能够通过与其靶 mRNA 3' UTR 的特异性结合, 抑制靶 mRNA 的翻译或引起序列特异性的降解。以往的研究已经表明 miRNA 与多种生理、病理事件密切相关。我国科学家陆续证

明了 miRNA 参与细胞存活与死亡、外周血调节性 T 细胞(Regulatory T, Treg)的生成和 Treg 调节的免疫稳态等过程, 此外还发现 miRNA 的异常会引起动脉粥样硬化、银屑病、肿瘤等各类疾病, 并探索了 miRNA 在疾病治疗过程中的临床应用价值。

#### 2.3.1 miRNA 的生物学功能

经典的 miRNA 生物合成过程包括: miRNA 基因被 RNA 聚合酶 转录为 pri-miRNA, pri-miRNA 由 RNase 酶 Drosha 和辅因子 DGCR8 加工成为 65~80 bp 的 pre-miRNA, pre-miRNA 经过 RNase 酶 Dicer 再次加工剪接成约 22 bp 的成熟的 miRNA。其中, Drosha 将 pri-miRNA 加工成 pre-miRNA 是 miRNA 合成的关键步骤。第四军医大学唐都医院的杨倩与亚特兰大埃默里大学的 Zixu Mao 研究组联合开展了对 Drosha 调控 miRNA 作用机理的研究。他们发现, 在高温或 H<sub>2</sub>O<sub>2</sub> 刺激下, p38 MAPK 可促进 Drosha 的磷酸化水平, 导致 Drosha 与 DGCR8 的结合能力降低, 从而促进 DGCR8 被钙蛋白酶转运和降解。进一步研究发现, 敲低 Drosha 可使细胞变得更加敏感并导致细胞死亡增加。相反, 过表达 Drosha 可导致细胞的死亡减少。这些研究结果证明了 Drosha 蛋白水平的变化与细胞存活或死亡密切相关。该项研究不仅首次明确了 Drosha 作为新的应激反应调节蛋白的重要性, 同时也揭示了抑制 miRNA 的产生对调节细胞存活/死亡的关键作用和意义<sup>[53]</sup>。

miRNA 介导的基因沉默是基因转录后水平调控的重要方式, 但是其作用效率的调控机制尚不清楚。上海交通大学医学院的余健秀研究组发现, TARBP2 的 SUMO 化修饰能够使 miRNA/siRNA 有效发挥基因沉默的功能。以往报道显示, TARBP2 蛋白通过直接与 DICER 和 PACT 相互作用来稳定 RLC 复合物, 用于 miRNA 的加工和基因沉默。他们研究发现, TARBP2 蛋白的磷酸化修饰能够增强其第 52 位赖氨酸的 SUMO 化修饰。同时, TARBP2 的 SUMO 化修饰通过抑制 TARBP2 蛋白第 48 位上赖氨酸的泛素化, 起到稳定 TARBP2 蛋白的作用。TARBP2 蛋白的 SUMO 化修饰对 miRNA 的加工和生成没有影响, 但却能够调控 miRNA/siRNA 的有效性。SUMO 化的 TARBP2 蛋白招募 Ago2 蛋白形成



RLC 复合物,同时促进更多 pre-miRNA 装载入 RLC。由此,Ago2 蛋白更加稳定,并促使与 TARBP2/Dicer 结合的 miRNAs/siRNA 更加有效地传递到 Ago2 蛋白上。这些反应将促进用于 RNA 干扰的 RISC 复合物的有效形成。该项研究有助于人们进一步理解 miRNA/siRNA 介导的基因沉默的分子机制<sup>[54]</sup>。

T 细胞在获得性免疫中具有重要作用,它们的激活依赖于来自 T 细胞受体、共刺激受体及多种细胞因子的信号。一旦被某种抗原激活,初始 CD4<sup>+</sup> 的 T 细胞会增殖和分化为多种辅助性 T(T helper, TH)细胞,包括 TH1、TH2、TH17 和 Treg。这些辅助细胞能释放出不同的细胞因子,从而发挥不同的功能效应。其中,Treg 细胞具有免疫抑制作用,在多发硬化病等多种自身免疫疾病中有重要的临床应用价值。然而,目前 miRNA 在 T 细胞向 Treg 细胞分化过程中的分子机制却并不清楚。上海交通大学的王宏林研究组研究发现 miRNA 参与了调控外周血 Treg 细胞的生成过程。在小鼠 CD4<sup>+</sup> 的 T 细胞中,条件性敲除 miR-31 可以增强小鼠外周血 Treg 细胞的产生,减缓多发硬化病的发生与发展。进一步研究发现,miR-31 可以通过靶向 G 蛋白偶联受体家族成员 *Gprc5a* 的 3' UTR 区域发挥作用。基于 *Gprc5a* 基因敲除的小鼠模型发现,*Gprc5a* 的缺失能显著抑制初始 T 细胞向 Treg 细胞的分化,从而显著增加多发硬化病的严重程度。该项研究揭示了 miR-31 通过直接靶向 *Gprc5a* 负向调控外周血 Treg 细胞的分化,为包括多发硬化病在内的多种自身免疫性疾病提供了新的候选治疗靶标<sup>[55]</sup>。

此外,上海交通大学的沈南研究组还发现,miRNA 与 Treg 细胞调节的免疫稳态密切相关。miR-125a 的表达在多种人类自身免疫性疾病患者的 CD4<sup>+</sup> T 细胞中显著下调,这一结论在相关的自身免疫性小鼠模型中也得到了证实。基于 miR-125a 敲除的小鼠模型,他们发现,miR-125a 的缺失会导致 Treg 分化缺陷,原先的免疫稳态被打破,从免疫抑制状态转变为炎症状态,从而导致小鼠发生更为严重的结肠炎和自身免疫性脑脊髓炎。在全基因组层面对 miRNA 的靶基因分析发现,miR-125a 可抑制 Stat3、Il13 和 Ifng 等免疫反应相关因子。利用化学合成的 miR-125a 类似物或拮抗剂,在自身免疫性脑脊髓炎

小鼠模型中进行了尾静脉注射,结果发现这种化学合成的类似物或拮抗剂可以越过血脑屏障,在大部分的组织中实现 miR-125a 的过表达或抑制。miR-125 类似物不仅可以阻止脑脊髓炎的发生,而且可以减轻已发生的脑脊髓炎的严重程度。该项研究为开发基于 miRNA 的自身免疫性疾病的治疗提供了新的视角<sup>[56]</sup>。

### 2.3.2 miRNA 与复杂疾病的发生发展

miRNA 的异常会引起各类复杂疾病的发生。银屑病是一种复杂的慢性炎症性疾病,在人群中的感染比例大约为 2%~3%。其主要特征表现为皮肤颗粒层缺失导致的表皮增生、角质化增厚、角质细胞异常分化以及主要炎症免疫细胞渗透进入真皮和表皮。NF- $\kappa$ B 在银屑病表皮中持续活化,然而其促进角质细胞增殖的机理尚不清楚。上海交通大学医学院的王宏林研究组发现,miR-31 在银屑病患者和 T 细胞调节的牛皮癣小鼠模型的表皮中均显著高表达。在表皮角质化细胞中,miR-31 的激活需要 NF- $\kappa$ B 信号通路的持续活化。在小鼠表皮基底角质细胞中条件性敲除 miR-31 后,小鼠表皮的增生和真皮细胞的浸润受到抑制。机制研究发现,miR-31 可靶向细胞周期负向调控元件 *Ppp6c*,进而发挥促进角质细胞增殖的功能。该项研究揭示了 miR-31 可能是银屑病治疗的潜在靶点<sup>[57]</sup>。

组织纤维化导致许多慢性炎症疾病过程中的器官衰竭,巨噬细胞在组织纤维化中发挥重要作用。多项研究表明,miRNA 是免疫细胞发挥功能的关键调控因子,但其在巨噬细胞介导的组织纤维化过程中是否发挥重要作用,仍没有相关研究被报道。中山大学的宋尔卫、江山平和苏士成研究组发现,在巨噬细胞中,IL-4 和 IL-3 可通过促进 miR-142-5p 和抑制 miR-130a-3p,维持巨噬细胞的促纤维化效应。体外研究证明,miR-142-5p 可靶向 SOCS1,进而抑制 STAT6 的磷酸化水平,miR-130a-3p 可靶向 PPAR $\gamma$ ,进而调控 STAT6 信号通路。体内研究结果显示,抑制 miR-142-5p 或过表达 miR-130a-3p,可抑制 CCL4 诱导的肝纤维化和博来霉素诱导的肺纤维化。他们进一步从肝硬化和肺纤维化病人的样本组织中获取巨噬细胞来检测 miR-142-5p 和 miR-130a-3p 的表达,



结果发现在获取的巨噬细胞中, miR-142-5p 表达上调, 而 miR-130a-3p 表达下调。该项研究揭示了 miR-142-5p 和 miR-130-3p 在巨噬细胞介导的组织纤维化过程中发挥重要的调控作用<sup>[58]</sup>。

雌激素具有抑制肝脏脂肪累积的功效。为揭示雌激素通过 miRNA 影响非酒精性脂肪肝的机制, 复旦大学上海医学院的李希研究组开展了相关研究。他们通过体外和在体实验发现, 雌激素受体 $\alpha$ (ER $\alpha$ ) 可显著促进 miR-125b 的表达, 而 miR-125b 可抑制脂肪的累积。报告基因实验和染色质免疫共沉淀实验结果显示, ER $\alpha$  可直接转录激活 miR-125b 的表达。当切除卵巢或肝脏特异性敲除 ER $\alpha$  的小鼠接受 miR-125b 过表达的治疗, 其对高脂饮食引起的脂肪肝有抵抗能力, 相反抑制 miR-125b 将轻微促进高脂饮食小鼠的脂肪肝。该项研究首次揭示了雌激素通过上调 miR-125b 抑制肝脏脂肪病变的作用机制, 有望为肝脏脂肪变性提供新的治疗策略<sup>[59]</sup>。

异常的胆管增生是人类胆汁淤积型肝炎的典型特征。以往的体外和在体研究显示, 胆汁淤积型肝炎与 miRNAs 的异常表达显著相关, 然而, miRNAs 是否参与胆汁淤积引起的胆管上皮细胞增殖仍然未知。上海交通大学的蔡威研究组发现, miR-124 和 miR-200 的表达失调可以促进淤胆型肝炎中胆管上皮细胞的增殖。通过对胆管闭锁患者和胆管结扎小鼠的肝脏样本进行检测, 他们发现 IL-6 的表达水平显著上调, 而 miR-124 的表达水平则显著下调。同时, miR-124 与 *STAT3* 和 *IL-6R* 的 mRNA 表达水平呈负相关。通过体外和体内研究进一步证实, miR-124 可靶向 *STAT3* 和 *IL-6R* 的 3'UTR 区, 抑制 *STAT3* 和 *IL-6R* 的蛋白表达, 从而抑制了 IL-6 介导的体外胆管上皮细胞的增殖和体内胆管上皮细胞的增生。此外他们还发现, miR-200 家族成员 miR-200a、miR-200b 和 miR-200c 在胆道闭锁患者和胆管结扎小鼠的肝脏组织中表达水平显著升高, 且 miR-200 家族成员与 *FOXA2* 的表达呈负相关。miR-200 家族成员可能通过结合 *FOXA2* 的 3'UTR 区, 抑制 *FOXA2* 的表达, 进而促进 IL-6 的表达。该项研究首次揭示了 miR-124 和 miR-200 参与胆管增生的作用机制<sup>[60]</sup>。

在肿瘤研究领域, 我国科学家还陆续证明了 miRNA 的异常可能引起肺癌、结肠癌、鼻咽癌和肝

癌在内的多种恶性肿瘤的发生发展。肺癌是全球致死率最高的癌症, 非小细胞肺癌占有肺癌的 80%。阐明非小细胞肺癌的致病和复发机理对非小细胞肺癌的治疗具有重要意义。Wnt 信号通路在癌症干细胞的生物学过程中具有重要作用, 但非小细胞肺癌中  $\beta$ -catenin 和 *APC* 基因的突变并不常见, 因此非小细胞肺癌中 Wnt 信号通路的激活机制有待被进一步阐明。中山大学的黎孟枫研究组发现, miR-582-3p 在非小细胞肺癌的癌组织和癌细胞株中均显著高表达, 且其表达水平与非小细胞肺癌病人的总体生存期和无病生存期显著相关。进一步的机制研究显示, miR-582-3p 可以同时靶向多个 Wnt/ $\beta$ -catenin 信号通路的负调控因子(包括 *AXIN2*、*DKK3* 和 *SFRP1*), 激活 Wnt/ $\beta$ -catenin 信号通路, 从而维持肺癌干细胞样特性。该项研究证明 miR-582-3p 可以有效调节 Wnt/ $\beta$ -catenin 信号通路的活化程度, 有望成为非小细胞肺癌潜在的治疗靶点<sup>[61]</sup>。

结直肠癌是致死率较高的恶性肿瘤之一。已有研究表明, miRNA 与结直肠癌的发生发展密切相关。但是, miRNA 是否参与结直肠癌中糖代谢的调控尚不清楚。上海复旦大学的何祥火研究组基于小 RNA 深度测序和 miRNA 表达谱, 筛选获得 7 个与肿瘤代谢相关的 miRNA。其中, 3 个 miRNA 在 3 株结直肠癌细胞系中均可显著降低细胞中乳酸的产生。他们随后对效应最强的 miR-124 进行了深入研究, 并成功发现磷酸核糖焦磷酸合成酶 1(*PRPS1*)和核糖-5-磷酸异构酶-A(*RPIA*)是 miR-124 的直接作用靶点。与过表达 miR-124 的结果一致, 敲低 *PRPS1* 和 *RPIA* 可降低结直肠癌细胞中的糖消耗。相反, 过表达 *PRPS1* 或 *RPIA* 可恢复结直肠癌细胞的糖代谢。他们还发现, *PRPS1* 和 *RPIA* 可以促进核酸的代谢, 为 DNA 和 RNA 的生物合成提供前体。当过表达 miR-124, 抑制 *PRPS1* 或 *RPIA* 的表达后, 结直肠癌细胞的 DNA 合成和细胞生长受阻。在小鼠移植瘤模型中, 敲低 miR-124 可显著降低小鼠肿瘤的体积和重量。临床相关性研究进一步证实了上述发现。该项研究揭示了 miR-124 抑制结直肠癌增殖的作用机制<sup>[62]</sup>。

神经母细胞瘤是儿童最常见的颅外肿瘤。乙酰肝素酶(Heparanase, HPSE)是一种降解硫酸乙酰肝

素蛋白聚糖的内源性糖类内切酶,可以促进肿瘤的生长、侵袭、转移和血管形成。HPSE 在神经母细胞瘤中高表达,然而其异常表达的调控机制并不清楚。华中科技大学同济医学院的童强松研究组证实,miR-558 可以激活 HPSE 的表达,从而促进神经母细胞瘤的发生发展。在传统的认识中,miRNA 通常结合到靶基因的 3'UTR 区域,抑制其靶基因的翻译。然而他们的研究却发现,miR-558 可以通过结合到 HPSE 的启动子区,以 AGO1 依赖的形式诱导 HPSE 的启动子活性,上调 HPSE 的转录表达。在神经母细胞瘤细胞系中,miR-558 具有类似 HPSE 的功能,能促进肿瘤细胞的侵袭、转移和血管形成。拯救实验也进一步证实,过表达 HPSE 可以恢复因抑制 miR-558 表达而导致的细胞表型变化。裸鼠体内成瘤实验发现,miR-558 可以显著促进肿瘤的生长、转移和血管形成。临床相关性研究证实了上述发现。该项研究为 miR-558 在神经母细胞瘤发生中的作用机制提供了证据<sup>[63]</sup>。

肝细胞肝癌是世界常见的恶性肿瘤之一,然而目前肝癌发生发展的分子机制尚未完全阐明。中南大学湘雅医学院的杨连粤研究组发现,miR-188-5p 可以通过抑制 FGF5,进而促进肝癌的发生发展。研究发现,miR-188-5p 在肝癌组织中显著下调表达,并且在转移灶中表达更低。miR-188-5p 的低表达与肝癌患者的多结节、血管浸润能力以及患者的总体生存期和无病生存期呈显著负相关。体外实验结果显示外源过表达 miR-188-5p 可以抑制肝癌细胞系的增殖、侵袭及迁移能力。裸鼠肝脏原位成瘤实验表明,过表达 miR-188-5p 可以显著抑制肝癌的生长及转移。进一步的分子机制研究发现,miR-188-5p 通过结合 FGF5 的 3'UTR 抑制其表达,且两者表达的负相关性在肝癌临床组织样本中也得到了验证。体内及体外的靶基因拯救及依赖性实验证实,miR-188-5p 的抑癌功能需要通过抑制 FGF5 的表达来实现。该项研究揭示肝癌相关的抑癌 miRNA miR-188-5p 的作用机制,为肝癌的治疗和预后提供了一个新的候选药物靶标和标志物<sup>[64]</sup>。

第二军医大学东方肝胆外科医院的周伟平和林川研究组揭示,miR-422a 能够抑制肝癌的生长和转移。他们发现,miR-422a 在肝癌组织和细胞系中显

著下调表达,并且 miR-422a 的表达水平与肝癌的病理分级、复发和转移呈负相关。进一步的研究证实,miR-422a 可显著抑制肝癌细胞系的增殖、平板克隆形成、迁移和侵袭能力,同时 miR-422a 还可显著抑制体内肝癌的生长和转移。分子机制研究显示,miR-422a 可靶向 FOX 家族中 FOXG1、FOXQ1 和 FOXE1 基因的 3'UTR。体外和在体的实验结果表明,3 个靶基因敲低后均能显著抑制肿瘤的生长和转移。值得注意的是,他们还发现 miR-422a 和靶基因 FOXG1/Q1/E1 之间存在双向负调控的反馈现象。在 miR-422a 上游 1~2 kb 基因组区域内存在 miR-422a 潜在的启动子,该启动子活性与 miR-422a 的表达均能够被 miR-422a 的 3 个靶基因负向调控。在 DEN 诱导的肝癌小鼠模型中,通过过表达 miR-422a 治疗,发现肝癌得到抑制。该项研究提示 miR-422a 有可能是肝癌潜在的治疗靶标<sup>[65]</sup>。

以往的研究表明趋化因子受体 4(CXCR4)在肝癌的进展中具有多种功能,然而 CXCR4 引起肝癌的机制及其在肝癌中的治疗价值一直未被阐释。复旦大学上海医学院的徐洁杰研究组揭示,miR-622 能够通过调控 CXCR4 的表达影响肝癌的发生和发展。他们首先在肝癌癌组织中发现 CXCR4 存在异常的上调表达,且其上调表达与肝癌的较差预后和进展相关。敲低 CXCR4 可抑制体内和体外肝癌的生长,而过表达 CXCR4 则促进肝癌的生长。进一步研究发现 miR-622 可以直接结合 CXCR4 的 3'UTR 区,从而负向调控 CXCR4 的表达。研究还发现,miR-622 的表达受到 EZH2 的调控。EZH2 可诱导 H3K27 三甲基化和 CXCR4 启动子区的甲基化,从而抑制 miR-622 的表达。临床相关性分析进一步证实,miR-622 与 CXCR4 的蛋白表达呈负相关,CXCR4 与 EZH2 的蛋白表达水平呈正相关;同时,CXCR4 高表达且 miR-622 低表达的肝癌患者,无病生存期显著降低。该项研究揭示了 EZH2/miR-622/CXCR4 通路 with 肝癌的不良预后显著相关,为肝癌的治疗提供了潜在靶点<sup>[66]</sup>。

军事医学科学院输血研究所的裴雪涛和岳文研究组证实,miR-125b 可通过靶向 SMAD2 和 SMAD4,抑制上皮-间质转化(EMT),从而抑制肝癌肿瘤干细胞的产生与增殖。以往的研究报道 miRNA 与肝癌的

发生发展密切相关,但是其在 EMT 过程中的机制以及潜在的治疗价值并未被广泛的研究。他们通过在肝癌细胞中构建 EMT 模型,发现了一批显著下调表达的 miRNA,其中 miR-125b 表现最为显著。随后研究发现,miR-125b 不仅在绝大多数肝癌细胞系和肝癌组织中下调表达,同时其表达还与肝癌患者的分化程度相关。体外和体内实验进一步证实 miR-125b 对小鼠的 EMT 有明显的抑制作用。此外,miR-125b 还能够抑制 EMT 相关的一些性状,并且与 EMT 和癌症干细胞标记分子表达呈负相关。分子机制研究显示 miR-125b 可通过靶向 SMAD2 和 SMAD4,从而抑制 EMT。最为重要的是,他们将人工合成的 miR-125b 导入小鼠肝癌组织后发现,肿瘤干细胞和转移被显著抑制。该项研究揭示 miR-125b 可能是治疗肝癌的靶点<sup>[67]</sup>。

南方医科大学的刘莉和吴德华研究组发现,miR-34a 能够通过调控 lncRNA UFC1 而影响肝癌的发生发展。他们首先发现,UFC1 在肝癌组织中显著高表达,并且其表达与肿瘤的大小、分级以及病人的预后密切相关。通过体外细胞实验发现 UFC1 能促进肝癌细胞的增殖、抑制其凋亡,并且调节肝癌细胞的细胞周期。裸鼠成瘤实验表明,过表达 UFC1 可以显著促进肿瘤的生长,敲低 UFC1 则对肿瘤的生长产生明显的抑制作用。分子机制研究显示,UFC1 通过与 mRNA 稳定因子 HuR 蛋白相互作用,上调  $\beta$ -catenin 的表达,从而参与肿瘤的发生。他们还发现,miR-34a 可以直接与 UFC1 结合,从而调节其表达水平,进而影响肝癌的发生发展。该项研究揭示 miR-34a 通过与 UFC1 结合,影响癌基因 UFC1 的表达,从而参与肝癌的发生发展<sup>[68]</sup>。

病毒感染是人类肿瘤发生的重要病因。宿主 miRNA 的异常能够导致肿瘤的发生。事实上,病毒也能编码 miRNA,而且其编码的 miRNA 也可能与人类肿瘤的发生密切相关。鼻咽癌是我国南方以及东南沿海地区常见的恶性肿瘤之一,其在广东地区发病率最高。EB(Epstein-Barr)病毒在全球多于 90% 的人口中潜伏感染,是多种人类恶性肿瘤的病因。EB 病毒是首个被发现能编码 miRNA 的人类病毒。目前,EB 病毒编码的 miRNA 在鼻咽癌中的功能尚不清楚。南方医科大学的李欣和李纪良研究组研究

发现,在鼻咽癌中,EB 病毒编码的 miRNA BART1 通过激活 PTEN 依赖的信号通路,诱导鼻咽癌的转移。他们首先发现,BART1 在鼻咽癌中高表达,且与临床分期显著相关。BART1 可以促进鼻咽癌细胞系的迁移和侵袭,并促进体内肿瘤转移。分子机制研究发现,BART1 可以直接靶向抑癌基因 PTEN,激活 PTEN 依赖的 PI3K-Akt、FAK-p130<sup>Cas</sup> 和 Shc-MAPK/ERK1/2 信号通路,驱动上皮间质转化,从而促进鼻咽癌细胞的迁移和侵袭。该研究首次发现 EB 病毒 miRNA 对鼻咽癌发生发展的影响,为鼻咽癌的干预治疗提供了新的策略<sup>[69]</sup>。

miRNA 还被证明在多种致癌途径中发挥重要功能。我国科学家陆续发现,miRNA 在肿瘤相关的葡萄糖代谢、乳酸代谢和外泌体途径中发挥重要的桥梁作用。

激素及其受体在代谢调控中发挥重要作用。然而,激素及其受体在肿瘤代谢过程中的作用尚未完全被揭示。上海交通大学的张志刚和夏强研究组首次发现了盐皮质激素受体 MR 通过调控 miR-338-3p 促进乳酸的产生,进而抑制肝癌的发展。他们首先通过 siRNA 筛选的方法,分别沉默 20 个著名的激素受体,并检测其对肝癌细胞的 Warburg 效应。结果发现,多个激素受体沉默后均能对乳酸代谢产物产生影响。其中 MR 被沉默后,可使所有肝癌细胞系的乳酸显著增加。进一步通过体外和在体实验证明,MR 功能的增强或缺失能够显著影响肝癌细胞系的增殖、细胞周期和凋亡。分子机制研究显示,MR 作为转录因子可直接调控 miR-338-3p 的表达,进而由 miR-338-3p 靶向醣酵解关键酶 PKLR,从而抑制肝癌细胞系的 Warburg 效应。此外,他们还发现 MR 由于染色体缺失和组蛋白去乙酰化,在肝癌组织中显著下调表达,且 MR 的低表达与肝癌的不良预后显著相关。该研究为人们认识激素受体 MR 在肝癌发生发展中的作用机制提供了新的证据<sup>[70]</sup>。

葡萄糖代谢的重编程是肿瘤细胞的一个重要生物学事件。癌细胞可以迅速调整能量来源,从氧化磷酸化途径转换为糖酵解代谢途径,以满足其在缺氧环境中高效增殖的需要,但这个转变开关的分子机制却并不完全清楚。中国科学院上海生命科学研究院生物化学与细胞生物学研究所的刘默芳研究组

首次发现,在肝癌细胞中,缺氧可以诱导 RNA 结合蛋白 HuR 特异性结合初始的 miR-199a 转录体,抑制 miR-199a 的成熟,从而在肝癌的糖酵解中起到转换开关的作用。他们首先在低氧诱导的肝癌细胞中,发现 miR-199a 的表达显著下调,它可以通过靶向糖酵解基因己糖激酶 Hk2 和丙酮酸激酶 Pkm2,从而抑制它们的转录水平和转录后翻译。Hk2 和 Pkm2 在调节葡萄糖代谢和肿瘤发生中起着重要的作用。基于小鼠肝癌模型,他们进一步发现,miR-199a 过表达能显著抑制小鼠的葡萄糖摄取能力,减弱肿瘤的生长。该研究不仅揭示了肿瘤能量代谢重编程的新机制,同时也为肝癌的治疗提供了一种可能的途径<sup>[71]</sup>。

外泌体是直径大约 40~100 nm,来源于细胞的胞外膜泡,最初形成的成熟多泡体为管腔内的囊泡 (Intraluminal vesicles, ILVs)。这些 ILVs 通过限制膜和质膜的融合被释放到胞外基质中,成为人们所知的外泌体。肿瘤细胞可以通过自分泌和旁分泌的形式分泌外泌体而改变肿瘤微环境。越来越多的证据表明,外泌体通过将蛋白质和核酸传递给受体细胞,从而影响受体细胞的生物学过程,在细胞间的通讯中发挥重要作用。然而,外泌体是否通过调控 miRNA 表达谱影响肝癌的发生发展尚不清楚。中山大学的刘培庆和闵军研究组证实,外泌体可作为中介体,调控肝癌细胞中的 miRNA 表达谱,从而影响肝癌的发生发展。他们首先通过体外实验发现,肝癌细胞系自分泌的外泌体能促进肝癌细胞系自身的生长、侵袭和迁移。同时,他们发现负调控外泌体生物合成相关的基因 Vps4A 在肝癌组织中下调表达,并且与肝癌的进展及复发转移显著相关。进一步的体外和体内实验证实,Vps4A 具有抑制肝癌细胞系生长及迁移的能力。分子机制研究显示,Vps4A 能显著抑制原癌 miRNA 通过外泌体分泌出细胞,同时还能显著促进抑癌 miRNA 的分泌,并最终通过抑制 PI3K/AKT 信号通路的活性而发挥抑癌功能。该研究为寻找新的肝癌治疗靶标和发展新的治疗策略提供了线索<sup>[72]</sup>。

中国科学院上海生命科学研究院生物化学与细胞生物研究所的惠静毅研究组首次发现,mRNA 包装蛋白 YB-1 在 miR-29 的加工过程中具有重要的调控作用。他们采用 iCLIP-seq 技术在体内发现,

YB-1 蛋白可以与大部分的编码基因的 mRNA 前体及成熟体相互结合,结合的 motif 主要为 UYAUC 回文序列。此外他们还发现,YB-1 还与 pri-/pre-miR-29b-2 的终端环区域相互结合。二者的结合阻止了与 miRNA 生物合成相关的 DGCR8 及 Dicer 酶被招募到 miR-29b-2 前体的过程。在多形性胶质母细胞瘤中,miR-29 的表达显著下调,与 YB-1 的表达显著负相关。体外细胞表型实验证实,miR-29 的表达异常能显著影响多形性胶质母细胞瘤细胞系的生长。该研究首次发现了 YB-1 在 miRNA 生物合成过程中的调控作用,为揭示 miRNA 异常表达的分子机制提供了证据<sup>[73]</sup>。

### 2.3.3 miRNA 的临床相关性研究

我国科学家还完成了多项基于 miRNA 的临床相关性研究,如肿瘤的早期诊断、肿瘤的分型以及各类疾病的治疗。

肝癌的早期诊断包括血清学检测和影像学检查。目前常用的早期诊断标志物主要有甲胎蛋白( $\alpha$ -feto-protein, AFP)、脱- $\gamma$ -羧基凝血酶原(Des-gamma-carboxyprothrombin, DCP)和 AFP-L3 等,但是它们的临床诊断敏感度均较低。为寻找新的肝癌早期诊断标志物,中山大学第三附属医院的庄诗美研究组基于血清 miRNA 组学,开展了相关研究。在对 6 例肝癌患者和 18 例慢性乙肝患者的血清分析中,他们发现了 19 个在肝癌患者中显著上调表达且丰度较高的 miRNA。进而,他们在 108 例肝癌患者、47 例肝硬化患者、51 例慢性乙肝患者及 51 例健康对照的血清样本中进行了验证,并采用逻辑回归构建了诊断模型。最终确认由 miR-29a、miR-29c、miR-133a、miR-143、miR-145、miR-192 和 miR-505 等 7 个 miRNA 分子构成的诊断模型效果最佳(AUC=0.826;灵敏度=80.6%;特异度=84.6%)。在另外 153 例肝癌患者、71 例肝硬化患者、68 例慢乙肝患者及 60 例健康对照的血清样本中进行验证的结果也显示,此预测模型的 AUC 可以达到 0.817,灵敏度为 74.5%,特异度为 88.9%;另外一组独立样本集(49 例肝癌患者、42 例非活化 HBV 携带者及 48 例健康对照的血清样本)的验证结果也显示,该预测模型的 AUC 可以达到 0.884,灵敏度为 85.7%,特异度为 91.1%;在

巢式病例-对照人群(27 个通过超声、组织病理学及 CT 或 MRI 确诊的肝癌患者和 135 例肝硬化或慢乙肝患者)验证中,他们发现在确诊前 12 个月、9 个月、6 个月、3 个月及确诊时由这 7 个 miRNA 分子构成的诊断模型的表现均显著高于现有的检测金标准 AFP<sub>20</sub> 及 AFP<sub>400</sub>。该项研究建立的诊断模型有望应用于肝癌的早期筛查和诊断<sup>[5]</sup>。

天津医科大学肿瘤医院的陈可欣研究组利用 100 例肝癌组织样本的 mRNA 表达谱和 miRNA 表达谱数据开展了肝癌的分子分型研究。他们将肝癌分为两个亚型,建立的分类方法在独立样本集中得到了进一步验证。此外他们还发现,干细胞相关基因在肝癌亚型 1 中显著富集,并且该亚型肝癌病人的总体生存期和无病生存期均显著低于肝癌亚型 2。进一步分析表明,miR-148a 在肝癌亚型 1 中显著下调表达。已有的报道显示,miR-148a 能够靶向结合 *ACVR1* 的 3'UTR,从而抑制 *ACVR1* 的表达,而 *ACVR1* 是 BMP 通路中重要的受体,参与调控了多种干细胞标志物以及 IL-8 的表达。临床相关性分析显示,*ACVR1* 及其下游基因 CD24、CD90 和 IL-8 的高表达与肝癌患者较差的预后显著相关。他们进一步研究证实 miR-148a 能够抑制肝癌的发生。该项研究不仅建立了肝癌分子分型的分类方法,而且还揭示了一种可能用于肝癌预后检测的标志物和潜在的治疗靶点<sup>[74]</sup>。

miRNA 还被证明在间充质干细胞移植治疗疾病的过程中具有重要作用。间充质干细胞移植已成功用于治疗多种疾病,但其作用机制尚未完全被解析。第四军医大学的金岩研究组联合宾夕法尼亚大学和南加州大学的施松涛研究组发现,在 Fas 缺失的 MRL/lpr 狼疮小鼠中,间充质干细胞移植能够挽救骨髓间充质干细胞的功能并改善骨质缺乏。分子机制研究发现,在 MRL/lpr 狼疮小鼠的骨髓间充质干细胞中,Fas 缺失将造成 miR-29b 无法被释放,使得细胞内的 miR-29b 的水平上升,进而导致 DNA 甲基转移酶 1(Dnmt1)的表达下调以及 *NOTCH1* 启动子区域的低甲基化,从而激活 Notch 信号通路,造成骨分化受损。他们还发现,经移植的间充质干细胞分泌的外泌体能够将 Fas 传递到 MRL/lpr 狼疮小鼠的骨髓间充质干细胞中,从而降低细胞内 miR-29b

的水平,由此改善 MRL/lpr 狼疮小鼠中骨髓间充质干细胞的功能。该研究揭示了 miRNA 在临床治疗中的重要意义<sup>[75]</sup>。

#### 2.4 竞争性内源 RNA(ceRNA)研究

近年来,随着 miRNA 研究的不断深入,竞争性内源 RNA(Competing endogenous RNA, ceRNA)假说正逐渐被学界接受和认可。该假说认为,miRNA 能够通过与其靶 mRNA 的 3'UTR 的特异性结合,抑制靶 mRNA 的翻译或引起目的基因的降解;各类转录物可通过 miRNA 反应元件,与靶 mRNA 竞争结合相同的 miRNA,从而调节靶基因的表达。ceRNA 假说为人们解析疾病的发生发展机制提供了新的思路和线索,我国科学家也陆续证实 ceRNA 与人类多种复杂疾病的发生发展相关。

细胞自噬是在进化上高度保守的应激反应,损伤的细胞器和异常的蛋白复合体在细胞自噬过程中被溶酶体降解。异常的细胞自噬通常与各类心血管疾病密切相关。lncRNA 被报道参与基因的表达调控,然而其在心血管疾病发生发展过程中如何发挥调控作用却仍旧未知。青岛大学转化医学研究院的李培峰研究组发现,lncRNA APF 可以与 miR-188-3p 竞争性结合,发挥 ceRNA 的功能调节 miR-188-3p 的靶基因 *ATG7* 的表达水平,进而调控自噬和心肌梗死。该项研究既揭示了 APF-miR-188-3p-*ATG7* 这一调控细胞自噬和心肌梗塞的新机制,也为心肌梗塞和心力衰竭的诊断和治疗提供了新的潜在的治疗靶点<sup>[76]</sup>。

病理性血管新生是眼部病变、肿瘤和动脉硬化等疾病的关键组成部分,它通常由异常的细胞增殖、细胞迁移、免疫、炎症反应等生物学过程导致。lncRNA 是上述生物学过程的重要调控分子,然而其在糖尿病导致的微血管功能障碍过程中的作用机制仍然不清楚。南京医科大学附属眼科医院的蒋沁研究组发现,lncRNA MIAT 在糖尿病鼠的视网膜中以及高糖处理下的血管内皮细胞中均显著上调表达。进一步的在体和体外实验证实,敲低 MIAT 能够在体内明显缓解糖尿病导致的视觉功能障碍和视网膜损伤,在体外则减弱细胞的活力、加速细胞凋亡以及抑制细胞的增殖。分子机制研究显示,MIAT 具有 ceRNA 的功能,可与 miR-150-5p 竞争性结合,从

而调节 miR-150-5p 的靶基因 VEGF 的表达水平。该研究证明了 lncRNA MIAT 在血管新生过程中可发挥重要作用<sup>[77]</sup>。

肝纤维化是各种慢性肝病向肝硬化发展的必经阶段。其特点是细胞外基质过多的沉积以及正常肝结构的变形。其中,肝星形细胞的活化被认为是肝纤维化发生和发展的关键环节。因此,抑制肝星形细胞的活化被认为是治疗肝纤维化的理想策略。复旦大学金山医院的樊晓明研究组揭示,lncRNA GAS5 能够通过 ceRNA 机制竞争结合 miR-222,从而有效抑制肝星形细胞的活化。他们首先发现,GAS5 在肝纤维化组织以及激活的肝星形细胞中存在下调表达。过表达 GAS5 能够抑制体外肝星形细胞的活化,并降低体内胶原的积累。生物信息学分析预测和功能实验发现,GAS5 上存在 miR-222 的结合位点,且 miR-222 能够抑制 GAS5 的表达。有趣的是,GAS5 也能够抑制 miR-222 的表达。进一步研究发现,GAS5 通过 ceRNA 机制竞争结合 miR-222,进而影响 miR-222 的靶基因 p27 的表达,而 P27 是调控细胞周期的关键性抑制因子。该项研究不仅阐明了 GAS5 通过调控 miR-222 和 p27 来抑制肝星形细胞的活化和增殖的作用机制,同时也有望为肝纤维化提供潜在的治疗靶标<sup>[78]</sup>。

ceRNA 还被报道与肿瘤的发生密切相关。卵巢癌是一种致死率极高的妇科恶性肿瘤。lncRNA HOST2 被发现在人卵巢癌中特异性高表达,然而其在卵巢癌发生发展过程中究竟发挥什么样的作用仍不清楚。第二军医大学的辉宁研究组通过体外实验发现,HOST2 在卵巢上皮癌中能促进肿瘤细胞的增殖、迁移和侵袭等生物学过程。裸鼠成瘤实验表明,敲低 HOST2 可以显著抑制肿瘤的生长。生物信息学预测并结合相关的功能实验发现,HOST2 通过与 miR-let-7b 竞争结合,调节 miR-let-7b 的功能,进而促进肿瘤细胞的增殖、迁移和侵袭能力。该研究揭示了 lncRNA HOST2 在卵巢癌中通过与 miR-let-7b 竞争性结合致癌的机制,为卵巢癌的诊断和治疗提供了一定的理论基础<sup>[79]</sup>。

ceRNA 还被报道与肿瘤的转移密切相关。上海第二军医大学的孙树汉研究组于 2014 年揭示,lncRNA ATB 作为 ceRNA,竞争性结合 miR-200 家

族,上调 ZEB1 和 ZEB2 的表达水平,从而促进 EMT 和肝癌细胞的侵袭<sup>[80]</sup>。同时,ATB 还能够结合 IL-11 mRNA 并诱导 IL-11 的表达,从而激活 STAT3 信号通路,导致肝癌细胞的转移。香港中文大学的 Sun Tingting 和 Wong Nathalie 在高度评价此项研究的同时,进一步提出了该研究尚未解决的两个问题:既然 ATB 能够同时结合 miR-200 和 IL-11,那么,ATB 结合二者的比例是如何被调控的?此外,在促进肝癌转移的过程中,经典的 TGF- $\beta$ -Smad-EMT 信号通路是否与上述报道的 TGF- $\beta$ -lncRNA-EMT 信号通路存在某种联系<sup>[81]</sup>?

## 2.5 mRNA 可变剪接研究

在真核生物中,剪接是形成成熟 mRNA 的必经过程,可在转录后水平调控基因表达。可变剪接机制的存在,使得生物体的复杂性大大提高。可变剪接在生理过程中发挥着重要的调控功能。越来越多的证据表明,异常的可变剪接是导致人类疾病的重要因素。多囊卵巢综合征是育龄妇女最常见的内分泌代谢疾病之一,发病率约为 6%~8%,患者通常表现为持续不排卵、高雄激素血症及胰岛素抵抗、卵巢多囊样改变等,是不孕症的常见原因。目前多囊卵巢综合征(PCOS)的病因不明。高雄激素血症及卵泡发育障碍是 PCOS 患者最常见的症状,因此众多研究人员从雄激素代谢相关基因入手研究 PCOS 的遗传学基础。浙江大学的黄荷凤研究组对 68 名 PCOS 患者卵巢颗粒细胞上的雄激素受体(AR)进行研究,发现 62%的患者存在两种不同的剪接变体,即 AR 基因 cDNA 的 2 号、3 号外显子之间插入了一段 69 bp 的内含子序列或者 3 号外显子缺失。但是这两种剪接变体在 120 名正常女性中均未被检测到。通过比较不同剪接变体的 PCOS 患者体内的激素水平,发现表达 AR 剪接变体的患者表现出更加明显的全身及卵巢局部的雄激素增高、卵泡发育及排卵障碍等症状。通过染色质免疫共沉淀测序(ChIP-seq)及转录组测序(RNA-seq)进一步发现,与野生型 AR 相比,AR 剪接变体显著改变了下游基因的募集状态及下游基因的表达,而且发生改变的基因与卵泡发育及雄激素代谢密切相关。该项研究首次揭示了 AR 剪接变体是 PCOS 患者雄性激素过多和卵泡异常的主要原因<sup>[82]</sup>。

肝细胞肝癌(HCC)具有着较高的致死率。发生肿瘤门静脉血栓(PVTT)的HCC病人更是极少被治愈,同时这些病人预后极差。Merlin蛋白由抑癌基因 *Nf2* 编码,该蛋白在肿瘤发生和转移过程中发挥重要作用。但是目前为止,人们还尚未认识到 Merlin 剪接体的重要功能意义。第二军医大学的王红阳和河南科技大学的李忠研究组发现,在HCC中,尤其在转移灶中,Merlin表达量很低,并且Merlin低表达与较差预后相关。有趣的是,Merlin的一个剪接变体(缺失第2~4号外显子,  $\Delta^{2-4}$ Merlin)在HCC和PVTT中高表达,而且该变体也在PVTT来源的肝癌细胞系CSQT2中高表达。进一步研究发现,  $\Delta^{2-4}$ Merlin在细胞质而非细胞表面表达,能够干扰野生型Merlin结合  $\beta$ -catenin和ERM。过表达  $\Delta^{2-4}$ Merlin能够增加  $\beta$ -catenin和多种相关基因的表达水平,进而促进肿瘤的转移。该项研究揭示了一种新的肝癌转移和复发的机制<sup>[83]</sup>。

### 3 分子进化研究

高通量测序和高密度SNP芯片技术也为人类的进化研究提供了便利。前面已经提及,中国科学院上海生命科学研究院计算生物研究所的徐书华研究组利用新开发的算法WinXPCNVer,在临近*EPAS1*基因的非编码区域检测到一段藏族人群特异的约3.4 kb长度基因组片段的缺失。该研究为后续开展藏族人群高原适应性的分子机制研究指明了方向。同时,该研究还有另外一个重要发现,即该CNV在代表非现代人的丹尼索瓦人的基因组中并不存在。这一发现提示现代人祖先与非现代人祖先之间的基因交流格局和适应性进化机制比学界目前所理解的更加复杂<sup>[48]</sup>。

中国科学院古脊椎动物与古人类研究所的付巧梅博士与多国学者开展合作,发现罗马尼亚的Peștera cu Oase洞穴中的1个现代人与尼安德特人存在基因交流。据估计,尼安德特人在大约3.9~4.1万年前消失。尽管如此,现代亚欧大陆人种含有约1%~3%的尼安德特人DNA。他们首先通过目的区域核DNA富集实验,发现这个约4万年前的欧洲现代人含有6%~9%左右的尼安德特人基因,超出了目前任何已知的早期现代人基因组和现存欧亚大陆人基因组含有的尼安德特人基因组的含量(1%~4%)。然

后,他们通过数学模型估算出了尼安德特人的基因组的长度。运用平均基因遗传图谱,推算出该个体的前4~6代的祖先中存在尼安德特人,从而揭示了该个体的遗传物质中确实存在与尼安德特人的基因交流,时间大概在距今5万年到6万年前。然而,在此之前其他学者推测的现代人祖先与尼安德特人的基因交流发生在距今8.6万年至3.7万年间。该项研究不仅发现了1个早期现代人个体的基因组与尼安德特古人类存在密切联系,而且还将现代人祖先与尼安德特人的基因交流的时间范围缩小了2~3万年<sup>[84]</sup>。

### 参考文献(References):

- [1] Li BS, Li H, Bai Y, Kirschner-Schwabe R, Yang JJ, Chen Y, Lu G, Tzoneva G, Ma XT, Wu TM, Li WJ, Lu HS, Ding LX, Liang HH, Huang XH, Yang MJ, Jin L, Kang H, Chen ST, Du A, Shen SH, Ding JP, Chen HZ, Chen J, von Stackelberg A, Gu LJ, Zhang JH, Ferrando A, Tang JY, Wang SY, Zhou BB. Negative feedback-defective PRPS1 mutants drive thiopurine resistance in relapsed childhood ALL. *Nat Med*, 2015, 21(6): 563-571.
- [2] Mullighan CG. Mutant PRPS1: a new therapeutic target in relapsed acute lymphoblastic leukemia. *Nat Med*, 2015, 21(6): 553-554.
- [3] Guo F, Yan LY, Guo HS, Li L, Hu BQ, Zhao YY, Yong J, Hu YQ, Wang XY, Wei Y, Wang W, Li R, Yan J, Zhi X, Zhang Y, Jin HY, Zhang WX, Hou Y, Zhu P, Li JY, Zhang L, Liu SR, Ren YX, Zhu XH, Wen L, Gao YQ, Tang FC, Qiao J. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell*, 2015, 161(6): 1437-1452.
- [4] Liu BD, Sun LJ, Liu Q, Gong C, Yao YD, Lv XB, Lin L, Yao HR, Su FX, Li DS, Zeng MS, Song EW. A cytoplasmic NF- $\kappa$ B interacting long noncoding RNA blocks I $\kappa$ B phosphorylation and suppresses breast cancer metastasis. *Cancer Cell*, 2015, 27(3): 370-381.
- [5] Lin XJ, Chong YT, Guo ZW, Xie C, Yang XJ, Zhang Q, Li SP, Xiong YJ, Yuan YF, Min J, Jia WH, Jie YS, Chen MS, Chen MX, Fang JH, Zeng CX, Zhang YJ, Guo RP, Wu YK, Lin GL, Zheng LM, Zhuang SM. A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncol*, 2015, 16(7): 804-815.
- [6] Forner A. Hepatocellular carcinoma surveillance with miRNAs. *Lancet Oncol*, 2015, 16(7): 743-745.

- [7] Carter CO. Monogenic disorders. *J Med Genet*, 1977, 14(5): 316–320.
- [8] Cai T, Yang L, Cai WS, Guo S, Yu P, Li JC, Hu XY, Yan M, Shao QZ, Jin Y, Sun ZS, Luo ZJ. Dysplastic spondylolysis is caused by mutations in the diastrophic dysplasia sulfate transporter gene. *Proc Natl Acad Sci USA*, 2015, 112(26): 8064–8069.
- [9] Qin YY, Guo T, Li GY, Tang TS, Zhao SD, Jiao X, Gong JJ, Gao F, Guo CX, Simpson JL, Chen ZJ. CSB-PGBD3 mutations cause premature ovarian failure. *PLoS Genet*, 2015, 11(7): e1005419.
- [10] Wu N, Ming X, Xiao JQ, Wu ZH, Chen XL, Shinawi M, Shen YP, Yu GJ, Liu JQ, Xie H, Gucev ZS, Liu S, Yang N, Al-Kateb H, Chen J, Zhang J, Hauser N, Zhang T, Tasic V, Liu PF, Su XL, Pan XD, Liu CY, Wang LW, Shen J, Shen JX, Chen YL, Zhang T, Zhang JG, Choy KW, Wang J, Wang QQ, Li SG, Zhou WC, Guo J, Wang YP, Zhang C, Zhao H, An Y, Zhao Y, Wang JC, Liu ZL, Zuo YZ, Tian Y, Weng XS, Sutton VR, Wang HY, Ming Y, Kulkarni S, Zhong TP, Giampietro PF, Dunwoodie SL, Cheung SW, Zhang X, Jin L, Lupski JR, Qiu GX, Zhang F. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med*, 2015, 372(4): 341–350.
- [11] Zhang ZH, Li CH, Wu F, Ma RX, Luan J, Yang F, Liu WD, Wang L, Zhang SM, Liu Y, Gu J, Hua WL, Fan M, Peng H, Meng XM, Song NJ, Bi XL, Gu CY, Zhang Z, Huang Q, Chen LJ, Xiang LH, Xu JH, Zheng ZZ, Jiang ZW. Genomic variations of the mevalonate pathway in porokeratosis. *eLife*, 2015, 4: e06322.
- [12] Lin ZM, Zhao JH, Nitoiu D, Scott CA, Plagnol V, Smith FJD, Wilson NJ, Cole C, Schwartz ME, McLean WH, Wang HJ, Feng C, Duo LN, Zhou EY, Ren YL, Dai LL, Chen YL, Zhang JG, Xu X, O'Toole EA, Kelsell DP, Yang Y. Loss-of-function mutations in CAST cause peeling skin, leukonychia, acral punctate keratoses, cheilitis, and knuckle pads. *Am J Hum Genet*, 2015, 96(3): 440–447.
- [13] MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. *Nature*, 2014, 508(7497): 469–476.
- [14] Zhou SR, Xie YL, Tang JZ, Huang JL, Huang QZ, Xu W, Wang ZQ, Luo FT, Wang Q, Chen HG, Du XL, Shen Y, Chen D, Chen L. FGFR3 deficiency causes multiple chondroma-like lesions by upregulating Hedgehog signaling. *PLoS Genet*, 2015, 11(6): e1005214.
- [15] Han F. More on Sweet's syndrome in patients with MDS and MEFV mutations. *N Engl J Med*, 2015, 372(20): 1970–1971.
- [16] Kone-Paut I, Hentgen V, Touitou I. More on Sweet's syndrome in patients with MDS and MEFV mutations. *N Engl J Med*, 2015, 372(20): 1970.
- [17] Zhang XY, Gui L, Zhang XY, Bulfer SL, Sanghez V, Wong DE, Lee Y, Lehmann L, Lee JS, Shih PY, Lin HJ, Iacovino M, Wehl CC, Arkin MR, Wang YZ, Chou TF. Altered cofactor regulation with disease-associated p97/VCP mutations. *Proc Natl Acad Sci USA*, 2015, 112(14): E1705–E1714.
- [18] Ye Z, Li ZQ, Wang YF, Mao Y, Shen M, Zhang QL, Li SQ, Zhou LF, Shou XF, Chen JH, Song ZJ, Ma ZY, Zhang ZY, Li YM, Ye HY, Huang CX, Wang T, He WQ, Zhang YC, Xie R, Qiao ND, Qiu HJ, Huang S, Wang M, Shen JW, Wen ZJ, Li WJ, Liu K, Zhou J, Wang L, Ji J, Chen H, Cheng HX, Shi ZF, Zhu YQ, Geng DY, Yao ZW, Tang WJ, Lu B, Pan L, Zhang Y, Bao WM, Wu JS, Zheng K, Shi YY, Zhao Y. Common variants at 10p12.31, 10q21.1 and 13q12.13 are associated with sporadic pituitary adenoma. *Nat Genet*, 2015, 47(7): 793–797.
- [19] Hu ZB, Shi YY, Mo XM, Xu J, Zhao BJ, Lin Y, Yang SW, Xu ZF, Dai JC, Pan SD, Da M, Wang XW, Qian B, Wen Y, Wen J, Xing JL, Guo XJ, Xia YK, Ma HX, Jin GF, Yu SQ, Liu JY, Zhou ZM, Wang XR, Chen YJ, Sha JH, Shen HB. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat Genet*, 2013, 45(7): 818–821.
- [20] Lin Y, Guo XJ, Zhao BJ, Liu JJ, Da M, Wen Y, Hu YL, Ni BX, Zhang K, Yang SW, Xu J, Dai JC, Wang XW, Xia YK, Ma HX, Jin GF, Yu SQ, Liu JY, Keavney BD, Goodship JA, Cordell HJ, Wang XR, Shen HB, Sha JH, Zhou ZM, Chen YJ, Mo XM, Luo LF, Hu ZB. Association analysis identifies new risk loci for congenital heart disease in Chinese populations. *Nat Commun*, 2015, 6: 8082.
- [21] Li CG, Li ZQ, Liu SG, Wang C, Han L, Cui LL, Zhou JG, Zou HJ, Liu Z, Chen JH, Cheng XY, Zhou ZW, Ding CC, Wang M, Chen T, Cui Y, He HM, Zhang KK, Yin CC, Wang YL, Xing SC, Li BJ, Ji J, Jia ZT, Ma LD, Niu JP, Xin Y, Liu T, Chu N, Yu Q, Ren W, Wang XF, Zhang AQ, Sun YP, Wang HL, Lu J, Li YY, Qing YF, Chen G, Wang YG, Zhou L, Niu HT, Liang J, Dong Q, Li XD, Mi QS, Shi YY. Genome-wide association analysis identifies three



- new risk loci for gout arthritis in Han Chinese. *Nat Commun*, 2015, 6: 7041.
- [22] Sun YM, Huang YQ, Yin AH, Pan YC, Wang YR, Wang C, Du Y, Wang ML, Lan FF, Hu ZB, Wang GQ, Jiang M, Ma JQ, Zhang XZ, Ma HX, Ma J, Zhang WB, Huang Q, Zhou ZW, Ma L, Li YD, Jiang HB, Xie L, Jiang YY, Shi B, Cheng J, Shen HB, Wang L, Yang YX. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat Commun*, 2015, 6: 6414.
- [23] Zhu ZZ, Tang NL, Xu LL, Qin XD, Mao SH, Song YM, Liu LM, Li FC, Liu P, Yi L, Chang J, Jiang L, Ng BK, Shi BL, Zhang W, Qiao J, Sun X, Qiu XS, Wang Z, Wang F, Xie DD, Chen L, Chen ZH, Jin MR, Han X, Hu ZS, Zhang Z, Liu Z, Zhu F, Qian BP, Yu Y, Wang B, Lee KM, Lee WYW, Lam TP, Qiu Y, Cheng JC. Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls. *Nat Commun*, 2015, 6: 8355.
- [24] Li M, Foo JN, Wang JQ, Low HQ, Tang XQ, Toh KY, Yin PR, Khor CC, Goh YF, Irwan ID, Xu RC, Andiappan AK, Bei JX, Rotzschke O, Chen MH, Cheng CY, Sun LD, Jiang GR, Wong TY, Lin HL, Aung T, Liao YH, Saw SM, Ye K, Ebstein RP, Chen QK, Shi W, Chew SH, Chen J, Zhang FR, Li SP, Xu G, Tai ES, Wang L, Chen N, Zhang XJ, Zeng YX, Zhang H, Liu ZH, Yu XQ, Liu JJ. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat Commun*, 2015, 6: 7270.
- [25] Jiang DK, Ma XP, Yu HJ, Cao GW, Ding DL, Chen HT, Huang HX, Gao YZ, Wu XP, Long XD, Zhang HX, Zhang YJ, Gao Y, Chen TY, Ren WH, Zhang PY, Shi ZQ, Jiang W, Wan B, Saiyin H, Yin JH, Zhou YF, Zhai Y, Lu PX, Zhang HW, Gu XL, Tan AH, Wang JB, Zuo XB, Sun LD, Liu JO, Yi Q, Mo ZN, Zhou GQ, Liu Y, Sun JL, Shugart YY, Zheng SL, Zhang XJ, Xu JF, Yu L. Genetic variants in five novel loci including CFB and CD40 predispose to chronic hepatitis B. *Hepatology*, 2015, 62(1): 118–128.
- [26] Wang ML, Takahashi A, Liu F, Ye DW, Ding Q, Qin C, Yin CJ, Zhang ZD, Matsuda K, Kubo M, Na R, Lin XL, Jiang HW, Ren SC, Sun JL, Zheng SL, Le Marchand L, Isaacs WB, Mo ZN, Haiman CA, Sun YH, Nakagawa H, Xu JF. Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer. *Nat Commun*, 2015, 6: 8469.
- [27] Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet*, 2011, 89(5): 607–618.
- [28] Liu H, Irwanto A, Fu XA, Yu GQ, Yu YX, Sun YH, Wang C, Wang ZZ, Okada Y, Low HQ, Li Y, Liany H, Chen MF, Bao FF, Li JH, You JB, Zhang QL, Liu J, Chu TS, Andiappan AK, Wang N, Niu GY, Liu DC, Yu XL, Zhang L, Tian HQ, Zhou GZ, Rotzschke O, Chen SM, Zhang XJ, Liu JJ, Zhang FR. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat Genet*, 2015, 47(3): 267–271.
- [29] Zuo XB, Sun LD, Yin XY, Gao JP, Sheng YJ, Xu JH, Zhang JZ, He CD, Qiu Y, Wen GD, Tian HQ, Zheng XD, Liu SX, Wang WJ, Li WR, Cheng YY, Liu LD, Chang Y, Wang ZX, Li ZG, Li LN, Wu JP, Fang L, Shen CB, Zhou FS, Liang B, Chen G, Li H, Cui Y, Xu AE, Yang XQ, Hao F, Xu LM, Fan X, Li YZ, Wu RN, Wang XL, Liu XM, Zheng M, Song SP, Ji BH, Fang H, Yu JB, Sun YX, Hui Y, Zhang FR, Yang RY, Yang S, Zhang XJ. Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. *Nat Commun*, 2015, 6: 6793.
- [30] Huang YF, Wang CC, Yao YF, Zuo XY, Chen SS, Xu CQ, Zhang HF, Lu QL, Chang L, Wang F, Wang PX, Zhang RF, Hu ZK, Song QX, Yang XW, Li C, Li SS, Zhao YY, Yang Q, Yin D, Wang XJ, Si WX, Li XC, Xiong X, Wang D, Huang Y, Luo CY, Li J, Wang JJ, Chen J, Wang LF, Wang L, Han M, Ye J, Chen FF, Liu JQ, Liu Y, Wu G, Yang B, Cheng X, Liao YH, Wu YX, Ke T, Chen QY, Tu X, Elston R, Rao SQ, Yang YZ, Xia YL, Wang QK, Williams SM. Molecular basis of gene-gene interaction: cyclic cross-regulation of gene expression and post-GWAS gene-gene interaction involved in atrial fibrillation. *PLoS Genet*, 2015, 11(8): e1005393.
- [31] Peng L, Zhao Q, Li QB, Li MX, Li CX, Xu TT, Jing XY, Zhu X, Wang Y, Li FC, Liu RH, Zhong C, Pan QH, Zeng BH, Liao QJ, Hu B, Hu ZX, Huang YS, Sham P, Liu JS, Xu SH, Wang J, Gao ZL, Wang YM. The p.Ser267Phe variant in SLC10A1 is associated with resistance to chronic hepatitis B. *Hepatology*, 2015, 61(4): 1251–1260.
- [32] Jin GF, Zhu M, Yin R, Shen W, Liu J, Sun J, Wang C, Dai JC, Ma HX, Wu C, Yin ZH, Huang JQ, Higgs BW, Xu L, Yao YH, Christiani DC, Amos CI, Hu ZB, Zhou BS, Shi YY, Lin DX, Shen HB. Low-frequency coding variants at 6p21.33 and 20q11.21 are associated with lung cancer risk in Chinese populations. *Am J Hum Genet*, 2015, 96(5): 832–840.
- [33] Huang LZ, Li YJ, Xie XF, Zhang JJ, Cheng CY, Yamashiro K, Chen LJ, Ma XY, Cheung CM, Wang YS, Zhang CF, Bai YJ, Hou J, Chen XL, Qi Y, Li SS, Sun YY, Mei JP, Cheng Y, Yu WZ, Hu XB, Zhuang FF, Fan L, Lu Y, Sun XH, Zhu XJ, Shen DF, Chan CC, Zhao MW, Yoshimura N,

- Pang CP, Wong TY, Khor CC, Zhang K, Zhou P, Li XX. Whole-exome sequencing implicates UBE3D in age-related macular degeneration in East Asian populations. *Nat Commun*, 2015, 6: 6687.
- [34] Sun J, Wang YX, Xia YS, Xu Y, Ouyang T, Li JF, Wang TF, Fan ZQ, Fan T, Lin BY, Lou HQ, Xie YT, Narod SA. Mutations in RECQL gene are associated with predisposition to breast cancer. *PLoS Genet*, 2015, 11(5): e1005228.
- [35] Luzzatto L. Somatic mutations in cancer development. *Environ Health*, 2011, 10(Suppl. 1): S12.
- [36] Jiang L, Gu ZH, Yan ZX, Zhao X, Xie YY, Zhang ZG, Pan CM, Hu Y, Cai CP, Dong Y, Huang JY, Wang L, Shen Y, Meng GY, Zhou JF, Hu JD, Wang JF, Liu YH, Yang LH, Zhang F, Wang JM, Wang Z, Peng ZG, Chen FY, Sun ZM, Ding H, Shi JM, Hou J, Yan JS, Shi JY, Xu L, Li Y, Lu J, Zheng Z, Xue W, Zhao WL, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat Genet*, 2015, 47(9): 1061–1066.
- [37] Zhang L, Zhou Y, Cheng CX, Cui HY, Cheng L, Kong PZ, Wang JQ, Li Y, Chen WL, Song B, Wang F, Jia ZW, Li L, Li YP, Yang B, Liu J, Shi RY, Bi YH, Zhang YY, Wang J, Zhao ZX, Hu XL, Yang J, Li HY, Gao ZB, Chen G, Huang XL, Yang XK, Wan SQ, Chen C, Li B, Tan YK, Chen LY, He MH, Xie S, Li XC, Zhuang XH, Wang MY, Xia Z, Luo LH, Ma J, Dong B, Zhao JZ, Song YM, Ou YW, Li EM, Xu LY, Wang JF, Xi YF, Li GD, Xu EW, Liang JF, Yang XF, Guo JS, Chen X, Zhang YB, Li QS, Liu LX, Li YR, Zhang XQ, Yang HM, Lin DX, Cheng XL, Guo YJ, Wang J, Zhan QM, Cui YP. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet*, 2015, 96(4): 597–611.
- [38] Chen KX, Yang D, Li XC, Sun BC, Song FJ, Cao WF, Brat DJ, Gao ZB, Li HX, Liang H, Zhao YR, Zheng H, Li M, Buckner J, Patterson SD, Ye X, Reinhard C, Bhatena A, Joshi D, Mischel PS, Croce CM, Wang YM, Raghavakaimal S, Li H, Lu X, Pan Y, Chang H, Ba SJ, Luo LH, Cavenee WK, Zhang W, Hao XS. Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci USA*, 2015, 112(4): 1107–1112.
- [39] Xu F, Wu LY, Chang CK, He Q, Zhang Z, Liu L, Shi WH, Guo J, Zhu Y, Zhao YS, Gu SC, Fei CM, Wu D, Zhou LY, Su JY, Song LX, Xiao C, Li X. Whole-exome and targeted sequencing identify ROBO1 and ROBO2 mutations as progression-related drivers in myelodysplastic syndromes. *Nat Commun*, 2015, 6: 8806.
- [40] Hu Z, Zhu D, Wang W, Li WY, Jia WL, Zeng X, Ding WC, Yu L, Wang XL, Wang LM, Shen H, Zhang CL, Liu HJ, Liu X, Zhao Y, Fang XD, Li SC, Chen W, Tang T, Fu AS, Wang Z, Chen G, Gao QL, Li S, Xi L, Wang CY, Liao SJ, Ma XY, Wu P, Li KZ, Wang SX, Zhou JF, Wang J, Xu X, Wang H, Ma D. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet*, 2015, 47(2): 158–163.
- [41] Jiang PY, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong VW, Wong GLH, Chan SL, Mok TSK, Chan HLY, Lai PBS, Chiu RWK, Lo YMD. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci USA*, 2015, 112(11): E1317–E1325.
- [42] Guo Y, Xu Q, Canzio D, Shou J, Li JH, Gorkin DU, Jung I, Wu HY, Zhai Y, Tang YX, Lu YC, Wu YH, Jia ZL, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, 2015, 162(4): 900–910.
- [43] Jiang WJ, Zhao XJ, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun*, 2015, 6: 8101.
- [44] Fu YS, Li CM, Lu SJ, Zhou WX, Tang FC, Xie XS, Huang YY. Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc Natl Acad Sci USA*, 2015, 112(38): 11923–11928.
- [45] Xia B, Han DL, Lu XY, Sun ZZ, Zhou AK, Yin QZ, Zeng H, Liu MH, Jiang X, Xie W, He C, Yi CQ. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat Methods*, 2015, 12(11): 1047–1050.
- [46] Cao HZ, Wu HL, Luo RB, Huang SJ, Sun YH, Tong X, Xie YL, Liu BH, Yang HL, Zheng HC, Li J, Li B, Wang Y, Yang F, Sun P, Liu SY, Gao P, Huang HD, Sun J, Chen D, He GZ, Huang WH, Huang Z, Li Y, Tellier LCAM, Liu X, Feng Q, Xu X, Zhang XQ, Bolund L, Krogh A, Kristiansen K, Drmanac R, Drmanac S, Nielsen R, Li SG, Wang J, Yang HM, Li YR, Wong GK, Wang J. De novo assembly of a haplotype-resolved human genome. *Nat Biotechnol*, 2015, 33(6): 617–622.
- [47] Shang RF, Zhang FJ, Xu BY, Xi HR, Zhang X, Wang WH, Wu LG. Ribozyme-enhanced single-stranded Ago2-processed interfering RNA triggers efficient gene silencing with fewer off-target effects. *Nat Commun*, 2015, 6: 8430.
- [48] Lou HY, Lu Y, Lu DS, Fu RQ, Wang XJ, Feng QD, Wu SJ, Yang YJ, Li SL, Kang LL, Guan YQ, Hoh BP, Chung YJ,

- Jin L, Su B, Xu SH. A 3.4-kb copy-number deletion near EPAS1 is significantly enriched in high-altitude tibetans but absent from the denisovan sequence. *Am J Hum Genet*, 2015, 97(1): 54–66.
- [49] He YG, Wang MX, Huang X, Li R, Xu HY, Xu SH, Jin L. A probabilistic method for testing and estimating selection differences between populations. *Genome Res*, 2015, 25(12): 1903–1909.
- [50] Fan JY, Xing Y, Wen XY, Jia RB, Ni HY, He J, Ding X, Pan H, Qian GX, Ge SF, Hoffman AR, Zhang H, Fan XQ. Long non-coding RNA ROR decoys gene-specific histone methylation to promote tumorigenesis. *Genome Biol*, 2015, 16: 139.
- [51] Wu CL, Wang Y, Jin B, Chen H, Xie BS, Mao ZB. Senescence-associated long non-coding RNA (SALNR) delays oncogene-induced senescence through NF90 regulation. *J Biol Chem*, 2015, 290(50): 30175–30192.
- [52] Fu WM, Zhu X, Wang WM, Lu YF, Hu BG, Wang H, Liang WC, Wang SS, Ko CH, Wayne MM, Kung HF, Li G, Zhang JF. Hotair mediates hepatocarcinogenesis through suppressing miRNA-218 expression and activating P14 and P16 signaling. *J Hepatol*, 2015, 63(4): 886–895.
- [53] Yang Q, Li WM, She H, Dou J, Duong DM, Du YH, Yang SH, Seyfried NT, Fu HA, Gao GD, Mao ZX. Stress induces p38 MAPK-mediated phosphorylation and inhibition of Drosha-dependent cell survival. *Mol Cell*, 2015, 57(4): 721–734.
- [54] Chen C, Zhu CH, Huang J, Zhao X, Deng R, Zhang HL, Dou JZ, Chen Q, Xu M, Yuan HH, Wang YL, Yu JX. SUMOylation of TARBP2 regulates miRNA/siRNA efficiency. *Nat Commun*, 2015, 6: 8899.
- [55] Zhang LY, Ke F, Liu ZY, Bai J, Liu JL, Yan S, Xu ZY, Lou FZ, Wang H, Zhu HY, Sun Y, Cai W, Gao YY, Li Q, Yu XZ, Qian YC, Hua ZC, Deng J, Li QJ, Wang HL. MicroRNA-31 negatively regulates peripherally derived regulatory T-cell generation by repressing retinoic acid-inducible protein 3. *Nat Commun*, 2015, 6: 7639.
- [56] Pan W, Zhu S, Dai D, Liu Z, Li D, Li B, Gagliani N, Zheng YJ, Tang YJ, Weirauch MT, Chen XT, Zhu W, Wang Y, Chen B, Qian YC, Chen YX, Fang JY, Herbst R, Richman L, Jallal B, Harley JB, Flavell RA, Yao YH, Shen N. MiR-125a targets effector programs to stabilize Treg-mediated immune homeostasis. *Nat Commun*, 2015, 6: 7096.
- [57] Yan S, Xu ZY, Lou FZ, Zhang LY, Ke F, Bai J, Liu ZY, Liu JL, Wang H, Zhu HY, Sun Y, Cai W, Gao YY, Su B, Li Q, Yang X, Yu JX, Lai YP, Yu XZ, Zheng Y, Shen N, Chin YE, Wang HL. NF- $\kappa$ B-induced microRNA-31 promotes epidermal hyperplasia by repressing protein phosphatase 6 in psoriasis. *Nat Commun*, 2015, 6: 7652.
- [58] Su SC, Zhao QY, He CH, Huang D, Liu J, Chen F, Chen JN, Liao JY, Cui XY, Zeng YJ, Yao HR, Su FX, Liu Q, Jiang SP, Song EW. miR-142-5p and miR-130a-3p are regulated by IL-4 and IL-13 and control profibrogenic macrophage program. *Nat Commun*, 2015, 6: 8523.
- [59] Zhang ZC, Liu Y, Xiao LL, Li SF, Jiang JH, Zhao Y, Qian SW, Tang QQ, Li X. Upregulation of miR-125b by estrogen protects against non-alcoholic fatty liver in female mice. *J Hepatol*, 2015, 63(6): 1466–1475.
- [60] Xiao YT, Wang J, Yan WH, Zhou Y, Chen YW, Zhou KJ, Wen J, Wang Y, Cai W. Dysregulated miR-124 and miR-200 expression contribute to cholangiocyte proliferation in the cholestatic liver by targeting IL-6/STAT3 signalling. *J Hepatol*, 2015, 62(4): 889–896.
- [61] Fang LS, Cai JC, Chen BX, Wu SS, Li R, Xu XN, Yang Y, Guan HY, Zhu X, Zhang L, Yuan J, Wu JH, Li MF. Aberrantly expressed miR-582-3p maintains lung cancer stem cell-like traits by activating Wnt/ $\beta$ -catenin signalling. *Nat Commun*, 2015, 6: 8640.
- [62] Qiu ZP, Guo WJ, Wang QF, Chen ZA, Huang SL, Zhao FY, Yao M, Zhao YJ, He XH. MicroRNA-124 reduces the pentose phosphate pathway and proliferation by targeting PRPS1 and RPIA mRNAs in human colorectal cancer cells. *Gastroenterology*, 2015, 149(6): 1587–1598 e1511.
- [63] Qu HX, Zheng LD, Pu JR, Mei H, Xiang X, Zhao X, Li D, Li SW, Mao L, Huang K, Tong QS. miRNA-558 promotes tumorigenesis and aggressiveness of neuroblastoma cells through activating the transcription of heparanase. *Hum Mol Genet*, 2015, 24(9): 2539–2551.
- [64] Fang F, Chang RM, Yu L, Lei X, Xiao S, Yang H, Yang LY. MicroRNA-188-5p suppresses tumor cell proliferation and metastasis by directly targeting FGF5 in hepatocellular carcinoma. *J Hepatol*, 2015, 63(4): 874–885.
- [65] Zhang J, Yang Y, Yang T, Yuan SX, Wang RY, Pan ZY, Yang Y, Huang G, Gu FM, Jiang BG, Lin C, Zhou WP. Double-negative feedback loop between microRNA-422a and forkhead box (FOX)G1/Q1/E1 regulates hepatocellular carcinoma tumor growth and metastasis. *Hepatology*, 2015, 61(2): 561–573.
- [66] Liu HO, Liu YD, Liu WS, Zhang WJ, Xu JJ. EZH2-mediated loss of miR-622 determines CXCR4 activation in hepatocellular carcinoma. *Nat Commun*, 2015, 6: 8494.
- [67] Zhou JN, Zeng Q, Wang HY, Zhang B, Li ST, Nan X, Cao N, Fu CJ, Yan XL, Jia YL, Wang JX, Zhao AH, Li ZW, Li YH, Xie XY, Zhang XM, Dong Y, Xu YC, He LJ, Yue W, Pei XT. MicroRNA-125b attenuates epithelial-mesenchymal transitions and targets stem-like liver cancer cells

- through small mothers against decapentaplegic 2 and 4. *Hepatology*, 2015, 62(3): 801–815.
- [68] Cao CH, Sun JY, Zhang DY, Guo XJ, Xie LW, Li X, Wu DH, Liu L. The long intergenic noncoding RNA UFC1, a target of MicroRNA 34a, interacts with the mRNA stabilizing protein HuR to increase levels of  $\beta$ -catenin in HCC cells. *Gastroenterology*, 2015, 148(2): 415–426.
- [69] Cai LM, Ye YF, Jiang Q, Chen YX, Lyu X, Li JB, Wang S, Liu TF, Cai HB, Yao KT, Li JL, Li X. Epstein-Barr virus-encoded microRNA BART1 induces tumour metastasis by regulating PTEN-dependent pathways in nasopharyngeal carcinoma. *Nat Commun*, 2015, 6: 7353.
- [70] Nie HZ, Li J, Yang XM, Cao QZ, Feng MX, Xue F, Wei L, Qin WX, Gu JR, Xia Q, Zhang ZG. Mineralocorticoid receptor suppresses cancer progression and the Warburg effect by modulating the miR-338-3p-PKLR axis in hepatocellular carcinoma. *Hepatology*, 2015, 62(4): 1145–1159.
- [71] Zhang LF, Lou JT, Lu MH, Gao CF, Zhao S, Li B, Liang S, Li Y, Li DS, Liu MF. Suppression of miR-199a maturation by HuR is crucial for hypoxia-induced glycolytic switch in hepatocellular carcinoma. *EMBO J*, 2015, 34(21): 2671–2685.
- [72] Wei JX, Lv LH, Wan YL, Cao Y, Li GL, Lin HM, Zhou R, Shang CZ, Cao J, He H, Han QF, Liu PQ, Zhou G, Min J. Vps4A functions as a tumor suppressor by regulating the secretion and uptake of exosomal microRNAs in human hepatoma cells. *Hepatology*, 2015, 61(4): 1284–1294.
- [73] Wu SL, Fu X, Huang JY, Jia TT, Zong FY, Mu SR, Zhu H, Yan Y, Qiu SW, Wu Q, Yan W, Peng Y, Chen JX, Hui JY. Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastoma multiforme. *Nucleic Acids Res*, 2015, 43(17): 8516–8528.
- [74] Li L, Liu YX, Guo Y, Liu B, Zhao YR, Li P, Song FJ, Zheng H, Yu JP, Song TQ, Niu RF, Li Q, Wang XW, Zhang W, Chen KX. Regulatory MiR-148a-ACVR1/BMP circuit defines a cancer stem cell-like aggressive subtype of hepatocellular carcinoma. *Hepatology*, 2015, 61(2): 574–584.
- [75] Liu SY, Liu DW, Chen C, Hamamura K, Moshaverinia A, Yang RL, Liu Y, Jin Y, Shi ST. MSC transplantation improves osteopenia via epigenetic regulation of notch signaling in lupus. *Cell Metab*, 2015, 22(4): 606–618.
- [76] Wang K, Liu CY, Zhou LY, Wang JX, Wang M, Zhao B, Zhao WK, Xu SJ, Fan LH, Zhang XJ, Feng C, Wang CQ, Zhao YF, Li PF. APF lncRNA regulates autophagy and myocardial infarction by targeting miR-188-3p. *Nat Commun*, 2015, 6: 6779.
- [77] Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, Tao ZF, Song YC, Chen Q, Jiang Q. lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. *Circ Res*, 2015, 116(7): 1143–1156.
- [78] Yu FJ, Zheng JJ, Mao YQ, Dong PH, Lu ZQ, Li GJ, Guo CY, Liu ZJ, Fan XM. Long Non-coding RNA growth arrest-specific transcript 5 (GAS5) inhibits liver fibrogenesis through a mechanism of competing endogenous RNA. *J Biol Chem*, 2015, 290(47): 28286–28298.
- [79] Gao Y, Meng H, Liu SP, Hu JJ, Zhang YM, Jiao TT, Liu YJ, Ou J, Wang D, Yao L, Liu SR, Hui N. lncRNA-HOST2 regulates cell biological behaviors in epithelial ovarian cancer through a mechanism involving microRNA let-7b. *Hum Mol Genet*, 2015, 24(3): 841–852.
- [80] Yuan JH, Yang F, Wang F, Ma JZ, Guo YJ, Tao QF, Liu F, Pan W, Wang TT, Zhou CC, Wang SB, Wang YZ, Yang Y, Yang N, Zhou WP, Yang GS, Sun SH. A long noncoding RNA activated by TGF- $\beta$  promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell*, 2014, 25(5): 666–681.
- [81] Sun TT, Wong N. Transforming growth factor- $\beta$ -induced long noncoding RNA promotes liver cancer metastasis via RNA-RNA crosstalk. *Hepatology*, 2015, 61(2): 722–724.
- [82] Wang FF, Pan JX, Liu Y, Meng Q, Lv PP, Qu F, Ding GL, Klausen C, Leung PCK, Chan HC, Yao WM, Zhou CY, Shi BW, Zhang JY, Sheng JZ, Huang HF. Alternative splicing of the androgen receptor in polycystic ovary syndrome. *Proc Natl Acad Sci USA*, 2015, 112(15): 4743–4748.
- [83] Luo ZL, Cheng SQ, Shi J, Zhang HL, Zhang CZ, Chen HY, Qiu BJ, Tang L, Hu CL, Wang HY, Li Z. A splicing variant of Merlin promotes metastasis in hepatocellular carcinoma. *Nat Commun*, 2015, 6: 8457.
- [84] Fu QM, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, Viola B, Prufer K, Meyer M, Kelso J, Reich D, Paabo S. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, 2015, 524(7564): 216–219.