

## 结肠直肠癌全基因组关联的研究进展

吴叶娇 易洪刚

**【关键词】** 结肠直肠癌; 全基因组关联研究; 单核苷酸多态性

**Progress of colorectal cancer by study the association of genome-wide** WU Ye-jiao<sup>1</sup>, YI Hong-gang<sup>2</sup>. 1 The First School of Clinical Medicine, 2 Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 210029, China

Corresponding author: YI Hong-gang, Email: ohcepf@163.com  
This work was supported by grants from the National Natural Science Foundation of China (No. 81202283, 81072389), the Natural Science Foundation of Higher Education Institutions of Jiangsu Province (No. 10KJA330034), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20113234110002) and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

**【Key words】** Colorectal cancer; Genome-wide association study; Single nucleotide polymorphism

结肠直肠癌(CRC)在全球的发病率和死亡率呈上升趋势。2000年CRC新增病例94.5万,2008年达到120万,平均年增加3.37%;2000年死亡病例数为49.2万,2008年达到60.9万,年均增加2.97%<sup>[1,2]</sup>。2008年全球范围内CRC列男性常见肿瘤第三位,而在女性中居第二位<sup>[3]</sup>。同时我国CRC发病率也呈快速上升趋势<sup>[4]</sup>。2000年CRC在我国恶性肿瘤发病率中居第五位,于2007年上升至第三位。

CRC作为一种复杂性疾病,其确切病因至今尚未明确,通常认为环境因素和遗传因素共同作用导致了CRC的发生和发展。目前可能的环境因素包括吸烟、缺乏体育锻炼、超重或者肥胖、过度食用红肉或者熟食和过量饮酒等<sup>[5]</sup>。有研究显示,在CRC的发生发展中,遗传因素约占35%(95%CI: 10%~48%)<sup>[5]</sup>。可见针对CRC遗传因素的研究在其发病机制的探讨中扮演重要角色。随着全基因组关联研究(genome-wide association study, GWAS)的出现和发展,对CRC遗传机制的研究已有很大进展,发现了许多与CRC有关的遗传易感性位点,为CRC发病机制研究提供了重要的新线索。本文系统介绍GWAS应用于CRC的研究进展。

DOI: 10.3760/cma.j.issn.0254-6450.2013.10.021

基金项目:国家自然科学基金(81202283,81072389);江苏省高校自然科学研究重大项目(10KJA330034);高等学校博士学科点专项科研基金(20113234110002);江苏高校优势学科建设工程项目

作者单位:210029 南京医科大学第一临床医学院(吴叶娇),公共卫生学院流行病与卫生统计学系(易洪刚)

通信作者:易洪刚,Email: ohcepf@163.com

1. GWAS: GWAS是利用人类全基因组高通量测序和分型技术,对研究对象的整个基因组中单核苷酸多态性(SNP)(通常是超过数十万个SNP)分型,并利用生物统计学(biostatistics)和生物信息学(bioinformatics)方法,检验SNP或基因与复杂疾病或可测性状的关联性,全面揭示疾病发生、发展与治疗相关的遗传特征。

GWAS的理论是基于人类常见复杂性疾病的遗传基础主要由常见的等位基因遗传变异(the common disease common variant hypothesis, CDCV)所致<sup>[6]</sup>。GWAS通常利用覆盖人类基因组中的高密度SNP来寻找病例(有特定疾病或特定性状者)和对照个体间等位基因频率的差异,如差异有统计学意义,即提示在基因组相应区域可能具有潜在功能意义、与所研究的疾病或性状有关联的DNA序列变异。随后在新样本中将不断验证这种有统计学意义关联的结果<sup>[7]</sup>。

GWAS主要有4个步骤<sup>[8]</sup>。①研究设计和准备阶段。包括查阅文献、提出科研假说、进行研究设计、招募合格的研究对象、问卷调查和采集生物学标本。②实验室基因分型阶段。包括DNA制备、芯片制备、基因分型和实验室质量控制等。③数据分析阶段。包括调查资料的清理、基因分型资料清理[如缺失值的控制、最小等位基因频率(minor allele frequency, MAF)控制、Hardy-Weinberg连锁不平衡检验(H-W LD test)、亲属关系鉴别等<sup>[9]</sup>、人群分层分析(鉴别种族和调整人群混杂效应)、各位点与疾病的效应分析,进而分析每个基因、每条通路(pathway)与疾病的关联性,以及基因-环境、基因-基因的交互作用]。④重复验证阶段。对新发现的疾病易感性位点、基因等,在同类人群中和其他种族人群中进行验证,评价研究结果的可重复性,减少假发现率(false discovery rate)<sup>[10]</sup>。包括内部验证(internal validation)和外部验证(external validation)。

1996年由Risch和Merikangas<sup>[11]</sup>首次提出GWAS。自2005年Science发表第一篇GWAS论文后<sup>[12]</sup>,与GWAS相关的文献量逐年增加,至2012年11月9日公开发表的GWAS文献有1418篇<sup>[13]</sup>。其中心血管疾病<sup>[14,15]</sup>、乳腺癌<sup>[16,17]</sup>、肺癌<sup>[18,19]</sup>和2型糖尿病<sup>[20,21]</sup>等许多复杂性疾病均有GWAS,并发现许多SNP与复杂性疾病有统计学关联,并在不同样本中得到验证。由此可见,GWAS目前已经成为人类复杂性疾病遗传易感性的主要研究策略。

2. GWAS与CRC:目前GWAS结果显示,可能与CRC有关的遗传易感性位点已达到31个,并在对部分SNP做大样本验证时发现了新的遗传易感性位点<sup>[22-31]</sup>。其中对部分位点已开展了生物学功能研究,结果提示部分易感性位点在CRC的

发生发展过程中可能具有重要作用。

(1) rs6983267: 2007 年 Tomlinson 等<sup>[32]</sup>采用 Illumina Hap550 Bead 芯片在 550 000 个标记 SNP 中,对 930 例欧洲裔 CRC 患者和 960 例对照进行初筛,随后在 4361 例 CRC 患者和 3752 例对照、1901 例 CRC 患者和 1079 例对照、1072 例 CRC 患者和 415 例对照中分别进行 3 次重复验证,结果均发现该位点与 CRC 发病有统计学关联(3 次重复验证合并  $P=1.27 \times 10^{-14}$ ,  $OR=1.21$ ,  $95\% CI: 1.15 \sim 1.27$ , 位于 8q24)。自此,rs6983267 与 CRC 的关系在多个大样本且不同人群的研究中得到重复验证<sup>[24-29,33]</sup>。

基于以上研究,Tuupanen 等<sup>[34]</sup>对 rs6983267 与 CRC 的关系进一步做了功能学方面的探索。结果显示,rs6983267 是 Wnt 信号通路上转录因子 TCF4 的结合位点。而几乎在所有 CRC 发生的早期,均能观察到 Wnt 信号通路的激活,包括 Wnt 信号通路上的转录因子 TCF4 的激活<sup>[35]</sup>。Wnt 信号通路的开启和关闭直接控制着大量与生长和代谢相关基因的表达水平,参与了多种生物学过程的调控,包括胚胎生长和形态发育、稳定组织、平衡能量代谢和维护干细胞。

经典的 Wnt 通路是由 Wnt 蛋白与细胞表面 Frizzled 受体结合后引发的一系列反应,其中涉及胞浆内  $\beta$ -catenin 的降解复合物。在结肠腺瘤的形成过程中,该复合物失活,胞浆内的  $\beta$ -catenin 得以稳定存在,部分  $\beta$ -catenin 进入细胞核与转录因子 TCF 作用,结合于 rs6983267,促进了靶基因的转录和表达,其中包括 MYC 基因<sup>[36,37]</sup>,即距离 rs6983267 最近的编码蛋白质癌基因(图 1)。

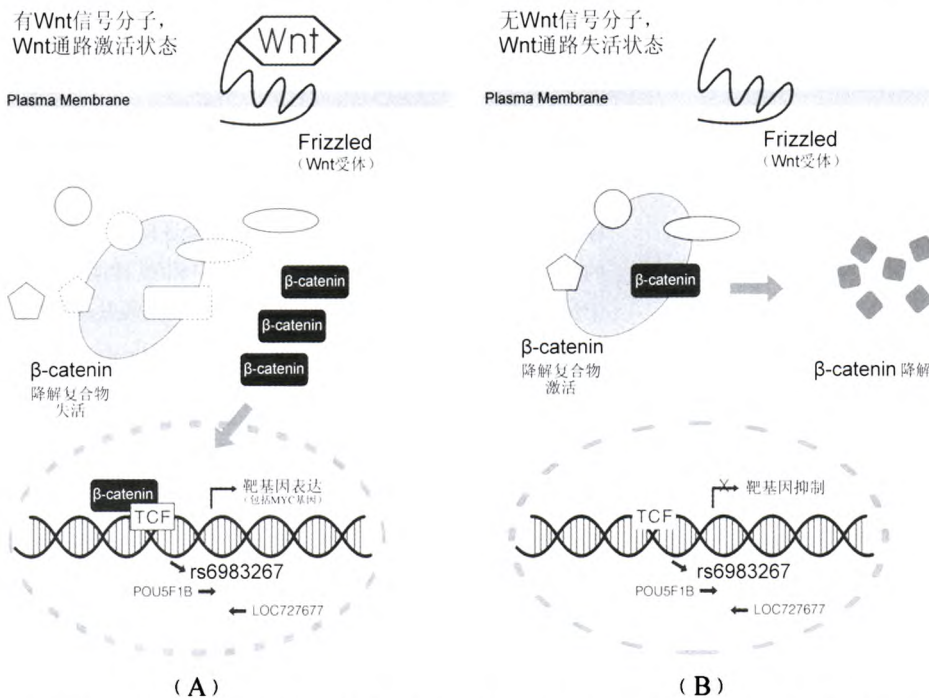
Pomerantz 等<sup>[38]</sup>还发现,rs6983267 位于一个转录增强子

(transcriptional enhancer)内,且该危险区域与原癌基因 MYC 有密切关联。故该位点很可能与某个未知的控制 MYC 基因表达元件(element)存在连锁不平衡(linkage disequilibrium, LD)<sup>[34]</sup>。这为研究非编码区易感位点的分子生物机制提供了一个很好的思路。

(2) rs4939827: 2007 年 Broderick 等<sup>[39]</sup>采用 Illumina Hap550 Bead 芯片在 547 647 个 SNP 中,对 930 例英国白种人 CRC 患者和 960 例对照进行初筛,随后在欧洲裔 4422 例 CRC 患者和 3844 例对照、1992 例 CRC 患者和 1680 例对照、963 例 CRC 患者和 343 例对照中分别进行 3 次重复验证,提示 rs4939827 与 CRC 有统计学关联( $P=1.0 \times 10^{-12}$ ,  $OR=0.85$ ,  $95\% CI: 0.81 \sim 0.89$ , 位于 18q21.1, SMAD7)。此后,rs4939827 与 CRC 的关联性还在其他大样本且不同人群的研究中得到验证<sup>[25,28,40-41]</sup>。

已有报道显示,SMAD7 是 TGF- $\beta$  信号通路的胞内拮抗剂<sup>[42]</sup>。TGF- $\beta$  是一种多功能的细胞因子,对几乎所有类型细胞均有作用,在维持胚胎发育和组织稳定中起到重要作用。TGF- $\beta$  在正常肠上皮细胞中抑制细胞的增殖和诱导凋亡,从而抑制癌症发生。在对负反馈循环和拮抗剂信号做出应答时,SMAD7 抑制 TGF- $\beta$  信号通路<sup>[43,44]</sup>。癌细胞可逃逸 TGF- $\beta$  的抑癌效应,抵制其诱导的增殖抑制而不断自我更新<sup>[45]</sup>。但在 CRC 的形成晚期,TGF- $\beta$  高表达,促进低分化和转移的 CRC 细胞不断增殖并向外侵袭,加快癌症发展进程<sup>[46]</sup>。

(3) rs4444235、rs961253 和 rs4925386: 2008 年 Houlston 等<sup>[47]</sup>采用 Illumina HumanHap 芯片在 555 000 余个 SNP 中,对伦敦 940 例 CRC 患者、965 例对照和爱丁堡 1012 例 CRC 患者、1012 例对照进行初筛,之后又在 38 000 余个 SNP 中对伦敦 2873 例 CRC 患者、2871 例对照和爱丁堡 2057 例 CRC 患者、2111 例对照进行二次筛选,随后在 8 组欧裔,共计 13 406 例 CRC 患者和 14 012 例对照中进行重复验证。结果提示 rs4444235 和 rs961253 与 CRC 存在统计学关联(合并  $P=8.1 \times 10^{-10}$ ,  $OR=1.11$ ,  $95\% CI: 1.08 \sim 1.15$ ; 合并  $P=2.0 \times 10^{-10}$ ,  $OR=1.12$ ,  $95\% CI: 1.08 \sim 1.16$ )。两者与 CRC 的关联性在其他大样本的人群研究中也得到验证<sup>[30,31]</sup>。rs4444235 位于染色体 14q22.2 上,在 BMP4 基因区域内。rs961253 位于染色体 20p12.3 上,该区域内无基因,但 BMP2 基因的端粒距离该位点 342 kb。



注:(A)有 Wnt 信号分子存在时,Wnt 信号分子与受体结合, $\beta$ -catenin 降解复合物失活解体, $\beta$ -catenin 不能降解而不断累积,进入核内与 TCF 结合于 rs6983267,靶基因表达;(B)无 Wnt 信号分子存在时, $\beta$ -catenin 在胞浆内  $\beta$ -catenin 降解复合物作用下进入降解途径,因此缺少  $\beta$ -catenin 与 TCF 结合于 rs6983267,靶基因的表达受抑制

图 1 Wnt 信号通路的传导过程

2010 年 Houlston 等<sup>[48]</sup>采用 Illumina HumanHap 芯片在 555 000 个 SNP 中对 5 组欧裔, 共计 8212 例 CRC 患者和 9542 例对照进行筛选, 之后对 6 组欧裔, 共计 9973 例 CRC 患者和 10 655 例对照重复验证。结果提示 rs4925386 与 CRC 存在统计学关联 (合并  $P=1.89 \times 10^{-10}$ ,  $OR=0.93$ , 95%  $CI: 0.91 \sim 0.95$ )。rs4925386 在 LAMA5 基因区域内, 此基因编码大型层粘连蛋白 A5, 并参与 BMP 细胞外拮抗剂 Noggin 的合成, 从而起到拮抗 BMP 的作用。

BMP 是一类分泌型糖蛋白, 通过调节一系列下游基因的活性, 参与中胚层形成、神经系统分化以及牙齿和骨骼发育等重要生物学过程。近年研究表明, BMP 属于 TGF- $\beta$  超家族中的重要成员, 在 CSC 的形成过程中起重要的作用。BMP 信号通路调控良好有助于 Wnt 通路的稳态维持, 从而抑制肠内基底细胞的自我更新<sup>[49]</sup>。而在腺瘤晚期至肿瘤早期

的过程中, 上皮内 BMP 通路常常丢失<sup>[50]</sup>。但是进入肿瘤进展期, 完整的 BMP 通路却与肿瘤的恶性程度相关<sup>[50]</sup>。

(4)rs7758229: 2010 年 Cui 等<sup>[33]</sup>采用 Illumina HumanHap 芯片对 1583 例 CRC 患者和 1898 例对照进行筛选, 之后利用 Invader Assay 对 4809 例 CRC 患者和 2973 例对照进行 3 次重复验证。结果提示 rs7758229 与 CRC 存在统计学关联 (合并  $P=7.92 \times 10^{-9}$ ,  $OR=1.28$ , 95%  $CI: 1.18 \sim 1.39$ )。rs7758229 位于 SLC22A3 基因区域内。

SLC22A3 属于有机阳离子转运家族, 在肝脏、肾脏、肠道和脑组织中广泛表达<sup>[51]</sup>。该家族成员在转运带阳离子药物、毒物和内源性代谢产物中起到重要作用<sup>[52]</sup>。有研究表明, 某些毒物或内源性代谢产物, 例如脂多糖和亚油酸的代谢物, 可引发肿瘤的形成。鉴于此, 可能是由于 SLC22A3 基因参与清除某致癌物的通路发生异常而诱发 CRC<sup>[33]</sup>。见表 1。

表 1 结肠直肠癌的 GWAS 中报道的遗传易感位点

第一作者(年份)	初筛样本	重复样本	SNP	基因座	基因	P 值	OR 值(95%CI)
Dunlop(2012) <sup>[33]</sup>	8682/9649	21 096/19 555	rs1321311	6p21	intergenic	$1.1 \times 10^{-10}$	1.10(1.07 ~ 1.13)
			rs3824999	11q13.4	POLD3	$3.7 \times 10^{-10}$	1.08(1.05 ~ 1.10)
			rs5934683	Xp22.2	SHROOM2	$7.3 \times 10^{-10}$	1.07(1.04 ~ 1.10)
Peters(2012) <sup>[40]</sup>	2906/3416 European descent	8161/9101 European descent	rs4779584	15q13.3	Intergenic	$1.8 \times 10^{-8}$	1.18(1.11 ~ 1.24)
			rs4939827	18q21.1	SMAD7	$1.1 \times 10^{-7}$	0.88(0.85 ~ 0.93)
			rs16892766	8q23.3	Intergenic	$4.0 \times 10^{-7}$	1.24(1.14 ~ 1.34)
			rs3802842	11q23.1	C11orf93	$3.8 \times 10^{-7}$	1.14(1.08 ~ 1.20)
Cui(2011) <sup>[33]</sup>	1583/1898 Japanese	4809/2973 Asian	rs7758229	6q25.3	SLC22A3	$7.9 \times 10^{-9}$	1.28(1.18 ~ 1.39)
			rs6983267	8q24.2	intergenic	$1.5 \times 10^{-8}$	1.18(1.11 ~ 1.25)
			rs7837328	8q24.2	intergenic	$7.4 \times 10^{-8}$	1.17(1.10 ~ 1.24)
Houlston(2010) <sup>[48]</sup>	3334/4628 UK	18 095/20 197 European ancestry	rs6691170	1q41	intergenic	$9.6 \times 10^{-10}$	1.06(1.03 ~ 1.09)
			rs11169552	12q13.1	intergenic	$1.9 \times 10^{-10}$	0.92(0.90 ~ 0.95)
			rs4925386	20q13.3	LAMA5	$1.9 \times 10^{-10}$	0.93(0.91 ~ 0.95)
			rs6687758	1q41	intergenic	$2.3 \times 10^{-9}$	1.09(1.06 ~ 1.12)
			rs10936599	3q26.2	MYNN	$3.4 \times 10^{-8}$	0.93(0.91 ~ 0.96)
			rs7136702	12q13.1	intergenic	$4.0 \times 10^{-8}$	1.06(1.04 ~ 1.08)
			rs961253	20p12	intergenic	$2.0 \times 10^{-10}$	1.12(1.08 ~ 1.16)
Houlston(2008) <sup>[47]</sup>	6780/6843 London and Edinburgh	13 406/14 012 European ancestry	rs355527	20p12	intergenic	$2.1 \times 10^{-10}$	1.12(1.08 ~ 1.17)
			rs4444235	14q22	intergenic	$8.1 \times 10^{-10}$	1.11(1.08 ~ 1.15)
			rs10411210	19q13.1	RHPN2	$4.6 \times 10^{-9}$	0.87(0.83 ~ 0.91)
			rs9929218	16q22	CDH1	$1.2 \times 10^{-8}$	0.91(0.89 ~ 0.94)
			rs1862748	16q22	CNH1	$2.9 \times 10^{-8}$	0.91(0.88 ~ 0.94)
			rs7259371	19q13.1	RHPN2	$2.2 \times 10^{-7}$	0.89(0.85 ~ 0.93)
			rs4939827	18q21.1	SMAD7	$7.8 \times 10^{-28}$	1.20(1.16 ~ 1.24)
Tenesa(2008) <sup>[41]</sup>	3069/3123 Scotland	14 500/13 294 All population	rs7014346	8q24.2	intergenic	$8.6 \times 10^{-26}$	1.19(1.15 ~ 1.23)
			rs12953717	18q21.1	intergenic	$3.1 \times 10^{-11}$	1.18(1.12 ~ 1.23)
			rs3802842	11q23.1	C11orf93	$5.8 \times 10^{-10}$	1.11(1.08 ~ 1.15)
			rs11213809	11q23.1	C11orf53	$7.9 \times 10^{-8}$	1.11(1.07 ~ 1.16)
			rs16892766	8q23	intergenic	$3.3 \times 10^{-18}$	1.25(1.19 ~ 1.32)
Tomlinson(2008) <sup>[54]</sup>	3813/3836 European ancestry	15 018/14 704 European ancestry	rs10795668	10p14	intergenic	$2.5 \times 10^{-13}$	0.89(0.83 ~ 0.91)
			rs4939827	18q21.1	SMAD7	$1.0 \times 10^{-12}$	0.85(0.81 ~ 0.89)
Broderick(2007) <sup>[39]</sup>	940/965 White UK	7 473/5 984 European ancestry	rs12953717	18q21.2	SMAD7	$9.1 \times 10^{-12}$	1.17(1.12 ~ 1.22)
			rs4464148	18q21.1	SMAD7	$6.7 \times 10^{-8}$	1.15(1.09 ~ 1.21)
Tomlinson(2007) <sup>[32]</sup>	930/960 UK residents	7 334/5 246 European ancestry	rs6983267	8q24.2	intergenic	$1.3 \times 10^{-14}$	1.21(1.15 ~ 1.27)
Zanke(2007) <sup>[55]</sup>	1257/1336 Ontario	6 223/6 443 European ancestry	rs10505477	8q24.2	intergenic	$3.6 \times 10^{-11}$	1.17(1.12 ~ 1.23)

3. 总结与展望:针对 CRC 的 GWAS 发现了大量易感位点,为深入研究其生物学发病机制提供了依据。与其他研究方法相比,GWAS 的最大优势在于研究前不需要提出致病位点所在基因组位置的生物学假设,而是从基因组所有 SNP 中筛选。采用高通量、高效率技术,辅以大样本量和重复验证,大大提高了 GWAS 从海量基因组信息中检测出与 CRC 相关的遗传变异的效能<sup>[56]</sup>。

但 GWAS 并非毫无瑕疵。GWAS 提示的易感位点大部分位于基因组的非编码区或“沙漠基因”中,目前对其生物学功能尚缺乏深入认识。如 GWAS 对影响 CRC 发生和发展的罕见变异和结构变异不敏感;LD 导致 GWAS 无法识别真正与 CRC 相关的遗传变异;GWAS 可能存在假阳性和分型错误。因此,有学者在经典 GWAS 的基础上提出了相关策略和方法,以弥补其不足。例如“多种疾病共同易感基因研究”、“基于通路的 GWAS”、“基因-基因、基因-环境交互作用研究”等<sup>[57]</sup>。

“后 GWAS”(Post-GWAS)时期应致力于描述人群、肿瘤分子亚型与临床相关变异三者之间相互联系的流行病学特征;探索基因-环境交互作用,增强其对人群患病风险的解释能力;确定最佳的候选致病变异用于后续的功能性研究,以阐明其潜在的生物学病因机制<sup>[58]</sup>。随着下一代测序技术的出现和发展,GWAS 将不断吸收和利用新技术,并将其应用到复杂性疾病、性状和药物效应的研究中。

### 参 考 文 献

- [1] Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncol*, 2001, 2(9): 533-543.
- [2] Jemal A, Center MM, DeSantis C, et al. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev*, 2010, 19(8): 1893-1907.
- [3] Jemal A, Bray F, Center MM, et al. Global cancer statistics. *CA Cancer J Clin*, 2011, 61(2): 69-90.
- [4] Wu F, Lin GZ, Zhang JX. An overview of cancer incidence and trend in China. *Chin Cancer*, 2012, 21(2): 81-85. (in Chinese)  
吴菲, 林国桢, 张晋昕. 我国恶性肿瘤发病现状及趋势. *中国肿瘤*, 2012, 21(2): 81-85.
- [5] Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*, 2000, 343(2): 78-85.
- [6] Iles MM. What can genome-wide association studies tell us about the genetics of common disease? *PLoS Genet*, 2008, 4(2): e33.
- [7] McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 2008, 9(5): 356-369.
- [8] Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med*, 2009, 360(17): 1759-1768.
- [9] Teo YY. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr Opin Lipidol*, 2008, 19(2): 133-143.
- [10] Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature*, 2007, 447(7145): 655-660.
- [11] Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1996, 273(5281): 1516-1517.
- [12] Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*, 2005, 308(5720): 385-389.
- [13] URL: <http://www.genome.gov/gwastudies/index.cfm>.
- [14] Reilly MP, Li M, He J, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet*, 2011, 377(9763): 383-392.
- [15] Smith EN, Chen W, Kahonen M, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet*, 2010, 6(9): e1001094.
- [16] Gaudet MM, Kirchoff T, Green T, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet*, 2010, 6(10): e1001183.
- [17] Fletcher O, Johnson N, Orr N, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst*, 2011, 103(5): 425-435.
- [18] Sato Y, Yamamoto N, Kunitoh H, et al. Genome-wide association study on overall survival of advanced non-small cell lung cancer patients treated with carboplatin and paclitaxel. *J Thorac Oncol*, 2011, 6(1): 132-138.
- [19] Yoon KA, Park JH, Han J, et al. A genome-wide association study reveals susceptibility variants for non-small cell lung cancer in the Korean population. *Hum Mol Genet*, 2010, 19(24): 4948-4954.
- [20] Shu XO, Long J, Cai Q, et al. Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet*, 2010, 6(9): e1001127.
- [21] Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet*, 2010, 42(10): 864-868.
- [22] Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet*, 2008, 40(1): 26-28.
- [23] Jia WH, Zhang B, Matsuo K, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet*, 2013, 45(2): 191-196.
- [24] Haiman CA, Le Marchand L, Yamamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet*, 2007, 39(8): 954-956.
- [25] Curtin K, Lin WY, George R, et al. Meta association of colorectal cancer confirms risk alleles at 8q24 and 18q21. *Cancer Epidemiol Biomarkers Prev*, 2009, 18(2): 616-621.
- [26] Schafmayer C, Buch S, Volzke H, et al. Investigation of the colorectal cancer susceptibility region on chromosome 8q24.21 in a large German case-control sample. *Int J Cancer*, 2009, 124(1): 75-80.
- [27] Wijnen JT, Brohet RM, van Eijk R, et al. Chromosome 8q23.3

- and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. *Gastroenterology*, 2009, 136(1):131-137.
- [28] Xiong F, Wu C, Bi X, et al. Risk of genome-wide association study-identified genetic variants for colorectal cancer in a Chinese population. *Cancer Epidemiol Biomarkers Prev*, 2010, 19(7):1855-1861.
- [29] He J, Wilkens LR, Stram DO, et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. *Cancer Epidemiol Biomarkers Prev*, 2011, 20(1):70-81.
- [30] Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet*, 2011, 7(6):e1002105.
- [31] Fernandez-Rozadilla C, Palles C, Carvajal-Carmona L, et al. BMP2/BMP4 colorectal cancer susceptibility loci in northern and southern European populations. *Carcinogenesis*, 2013, 34(2):314-318.
- [32] Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*, 2007, 39(8):984-988.
- [33] Cui R, Okada Y, Jang SG, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut*, 2011, 60(6):799-805.
- [34] Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet*, 2009, 41(8):885-890.
- [35] Bienz M, Clevers H. Linking colorectal cancer to Wnt signaling. *Cell*, 2000, 103(2):311-320.
- [36] He T-C, Sparks AB, Rago C, et al. Identification of c-MYC as a Target of the APC Pathway. *Science*, 1998, 281(5382):1509-1512.
- [37] Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science*, 1997, 275(5307):1784-1787.
- [38] Pomerantz MM, Ahmadiyah N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet*, 2009, 41(8):882-884.
- [39] Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet*, 2007, 39(11):1315-1317.
- [40] Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet*, 2012, 131(2):217-234.
- [41] Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet*, 2008, 40(5):631-637.
- [42] ten Dijke P, Hill CS. New insights into TGF-beta-Smad signalling. *Trends Biochem Sci*, 2004, 29(5):265-273.
- [43] Heldin CH, Miyazono K, ten Dijke P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature*, 1997, 390(6659):465-471.
- [44] Derynck R, Feng XH. TGF-beta receptor signaling. *Biochim Biophys Acta*, 1997, 1333(2):F105-150.
- [45] Engle SJ, Hoying JB, Boivin GP, et al. Transforming growth factor beta 1 suppresses nonmetastatic colon cancer at an early stage of tumorigenesis. *Cancer Res*, 1999, 59(14):3379-3386.
- [46] Xu Y, Pasche B. TGF-beta signaling alterations and susceptibility to colorectal cancer. *Hum Mol Genet*, 2007, 16 Spec 1:R14-20.
- [47] Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*, 2008, 40(12):1426-1435.
- [48] Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet*, 2010, 42(11):973-977.
- [49] Kosinski C, Li VS, Chan AS, et al. Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc Natl Acad Sci USA*, 2007, 104(39):15418-15423.
- [50] Kodach LL, Bleuming SA, Musler AR, et al. The bone morphogenetic protein pathway is active in human colon adenomas and inactivated in colorectal cancer. *Cancer*, 2008, 112(2):300-306.
- [51] Kekuda R, Prasad PD, Wu X, et al. Cloning and functional characterization of a potential-sensitive, polyspecific organic cation transporter (OCT3) most abundantly expressed in placenta. *J Biol Chem*, 1998, 273(26):15971-15979.
- [52] Inui KI, Masuda S, Saito H. Cellular and molecular aspects of drug transport in the kidney. *Kidney Int*, 2000, 58(3):944-958.
- [53] Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet*, 2012, 44(7):770-776.
- [54] Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet*, 2008, 40(5):623-630.
- [55] Zanke BW, Greenwood CMT, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*, 2007, 39(8):989-994.
- [56] Yan WL. Genome-wide association study on complex diseases: study design and genetic markers. *Hereditas*, 2008, 30(5):543-549. (in Chinese)  
严卫丽. 复杂疾病全基因组关联研究进展——遗传统计分析. *遗传*, 2008, 30(5):543-549.
- [57] Quan C, Zhang XJ. Research strategies for the next step of genome-wide association study. *Hereditas*, 2011, 33(2):100-108. (in Chinese)  
权晟, 张学军. 全基因组关联研究的深度分析策略. *遗传*, 2011, 33(2):100-108.
- [58] Le Marchand L. Genome-wide association studies and colorectal cancer. *Surg Oncol Clin N Am*, 2009, 18(4):663-668.

(收稿日期:2013-05-12)

(本文编辑:张林东)