人类复杂疾病全基因组关联研究

张学军^{1,2}

1. 复旦大学皮肤病研究所(附属华山医院), 上海 200000;

2. 安徽医科大学皮肤病研究所, 安徽医科大学第一附属医院皮肤性病科, 合肥 230000

E-mail: ayzj@vip.sina.com

2019-10-05 收稿, 2019-11-23 修回, 2019-11-25 接受, 2019-11-27 网络版发表

摘要 复杂疾病是由环境因素与遗传因素共同作用所致, 具有明显的遗传异质性和表型复杂性的特征, 且在人群中的发病率高, 严重影响人们的身心健康。随着HGP, HapMap和千人基因组计划的顺利完成, 研究人员已获得人类基因组中常见变异位点的详细图谱, 全基因组关联研究(genome-wide association study, GWAS)得到迅速发展。从2005年至今, 全世界众多研究小组开展了大量复杂疾病易感基因的GWAS, 发现了1995个疾病/表型相关变异, 推动了人类对这些疾病/性状的认识, 为探讨复杂疾病/性状的遗传学发病机制提供新的思路和方法, 从而揭示疾病发生、发展与治疗相关的遗传基础, 为将来疾病预警、风险预测、临床诊断、药物开发及个体化用药指导等在内的精准医学奠定了坚实的理论基础。本文就人类复杂疾病全基因组关联研究的兴起、发展、扩展、深入分析、临床转化及未来发展进行评述。

关键词 全基因组关联研究, 复杂疾病, 易感基因, 精准医学

人类基因组计划的实施, 对生命科学的研究带来了深远的影响。随着人类基因组计划(HGP)、人类基因组单体型图计划(HapMap)和千人基因组计划的顺利完成, 基因分型技术的快速发展和成本的不断降低, 结合强大效能的统计学方法和软件的出现, 使得全基因组关联研究(genome-wide association study, GWAS)得到迅速发展。被广泛应用于复杂疾病的易感基因搜寻研究中, 并取得了一系列研究成果, 为探讨复杂疾病/性状遗传学发病机制的搜寻提供新的思路和方法, 从而揭示疾病发生、发展与治疗相关的遗传基础。在此基础上迅速发展起来的各类组学研究将GWAS与功能学及临床研究紧密结合, 为真正实现人类疾病诊疗个体化的精准医学研究奠定坚实的遗传学基础。本文就全基因组关联研究的兴起、发展、扩展、深入及未来发展进行评述。

1 GWAS研究的兴起

复杂疾病是由环境因素与基因共同作用所致, 具

有明显的遗传异质性和表型复杂化的特征, 且发病率高, 严重影响人们的身心健康。越来越多的研究提示, 遗传因素在复杂疾病的发病过程中起到重要作用。在关联研究方法提出以前, 研究人员主要利用连锁研究方法开展对复杂疾病/性状遗传易感性研究, 发现了一些疾病和性状的易感基因。然而由于方法自身的局限性以及复杂疾病的表型复杂性、遗传异质性和种族差异性等特点, 以家系为基础的“全基因组连锁研究”的应用遇到了瓶颈, 人们对复杂疾病的研究效率存在质疑^[1~3]。以银屑病为例, 2002年, 安徽医科大学遗传性皮肤病研究团队^[4]通过对中国汉族寻常型银屑病家系进行全基因组扫描, 发现4q31(PSORS9)上存在与寻常型银屑病相关的易感基因区域。随后, 与中国国家人类基因南方研究中心陈竺和黄薇团队以及新加坡国立研究中心刘建军^[5]合作, 再次对该区域进行精细定位, 缩小易感区域范围, 选择银屑病病例与正常对照开展候选基因SNPs筛查, 发现IL-15基因可能参与银屑病的发病

引用格式: 张学军. 人类复杂疾病全基因组关联研究. 科学通报, 2020, 65: 671-683

Zhang X J. Review of genome-wide association study (in Chinese). Chin Sci Bull, 2020, 65: 671-683, doi: 10.1360/TB-2019-0063

机制。但后续的研究发现,连锁分析在复杂疾病家系中发现的易感位点或基因,很难在复杂疾病散发病例中得以重复或验证,因此连锁分析在寻找复杂疾病易感基因上面临困境。2009~2010年,本团队^[6,7]先后对汉族人群银屑病开展GWAS系列研究,进一步证实了1996年Risch和Merikangas^[8]提出的关于常见复杂疾病的遗传学研究中关联分析较连锁研究有更高效力的理论。随着HGP, HapMap和千人基因组计划的顺利完成,人类基因组中常见变异位点的详细图谱已被获取^[9]。在此基础上选取覆盖全基因组的代表性的分子标记位点,研发出高通量的基因芯片,对数百万个常见的变异位点进行成批检测。研究成果能将遗传多态位点(高密度的遗传变异图谱)和特定疾病(性状)的风险联系起来,使得GWAS成为可能,从根本上改变了复杂疾病基因研究的“蓝图”。

2 GWAS的发展

GWAS基于“常见疾病, 常见变异”(common disease, common variant)原理, 应用基因组中数以百万计的单核苷酸多态性(single nucleotide polymorphism, SNP)为分子遗传标记, 进行总体的关联分析, 在全基因组范围内选择遗传变异进行基因分型, 比较患者与对照组之间遗传变异及其频率的差异, 统计分析遗传变异与疾病/性状之间的相关性大小; 筛选出最相关的遗传变异开展多中心、大样本、反复验证, 最终确认与疾病/性状相关的遗传变异^[10]。

2.1 国际GWAS研究

2005年, *Science*报道了一篇年龄相关性视网膜黄斑变性的GWAS研究, 是国际上第一个关于复杂疾病易感基因GWAS研究的成果^[11], 从此拉开全世界复杂疾病GWAS研究的序幕。此后一系列复杂疾病易感基因的GWAS报道不断涌现, 研究成果相继发表在*Science*, *Nature*, *Nat Genet*和*N Engl J Med*等世界权威学术杂志上。人类遗传变异相关研究入选*Science* 2007年“世界十大科学进展”; GWAS相关研究于2011年再次被*Nature*评为年度“重大发现和科学事件”第二位。据美国NIH全基因组关联研究权威数据库收录(<http://www.genome.gov/GWASStudies/>)。统计显示, 目前肿瘤、自身免疫性疾病、心血管疾病、神经退行性疾病、精神疾病、炎症性疾病、感染性疾病等, 以及肥胖、身高、肤色、骨密度、血脂等与人类生命密切相关的重要复

杂疾病/性状的易感基因被陆续揭示^[12]。截至2018年12月, 该数据库共收录了3675篇相关科技论文, 覆盖了1995个人类复杂疾病或性状, 包含了与这些疾病/性状相关的68289个SNPs, 推动了人类对这些疾病/性状的认识。该研究方法已成为搜寻易感基因最为有效的手段, 为探讨复杂疾病/性状遗传学发病机制的搜寻提供新的思路和方法, 从而揭示疾病发生、发展与治疗相关的遗传基因。

2.2 中国的GWAS研究

GWAS作为揭示复杂遗传性疾病易感基因的一种行之有效的方法而备受关注。中国拥有最大的医学资源市场和庞大的患者群体, 具有民族多样性, 且疾病结构谱正在发生改变, 慢性疾病逐渐增多, 有利于进行大规模的基因分型研究, 揭示复杂疾病的遗传学机制, 为复杂疾病的防治指明方向。此前中国已积极参与各项人类基因组计划, 包括HGP, HapMap以及千人基因组计划, 促进国际交流, GWAS研究基础扎实。中国在GWAS研究领域虽然起步较晚, 但是在相关复杂疾病及性状的研究领域也做出了突出的贡献并取得重大成果。上海交通大学Bio-X研究院的贺林、安徽医科大学张学军和孙良丹^[13], 在*Nature*发表了题为“The progress of GWAS in China”的点评性文章, 介绍了中国在复杂疾病GWAS研究领域的进展。中国研究人员已经对30余种不同的人类疾病进行GWAS研究, 占国际GWAS研究疾病的18.5%, 占国际GWAS发表同类论文的8.1%。国内GWAS疾病主要包括: 精神疾病(例如精神分裂症、自闭症等)、癌症(例如肝癌、肺癌、鼻咽癌、宫颈癌、食道癌等)、谢性疾病(例如2型糖尿病、多囊卵巢综合征和骨质疏松症)、自身免疫性疾病(例如IgA肾病和强直性脊柱炎)、心血管疾病(例如冠状动脉疾病、高血压)、皮肤复杂疾病(例如银屑病、系统性红斑狼疮、麻风、白癜风和特应性皮炎)等。上述这些GWAS研究报道了约120个与疾病/性状的易感基因座(<http://www.genome.gov/GWASStudies/>)。据不完全统计, 截止到2018年12月, GWAS权威数据库已收录中国科学家发表的GWAS研究共计179篇, 包括在*N Engl J Med*上发表2篇, *Nat Genet*上发表44篇; 研究的疾病/性状涵盖127种, 病例样本数量约30万, 对照样本数量近40万。

2009年, 安徽医科大学张学军团队与国家人类基因组南方研究中心陈竺、黄薇等17个团队^[6]合作发表

的中国第一篇GWAS成果。利用GWAS研究策略首次发现,位于染色体1q21上的*LCE*基因是银屑病的易感基因,该基因编码表皮终末分化角质外膜蛋白,与银屑病最基本的组织病理学改变——角质形成细胞过度增殖密切相关。*Nature China*评价“*LCE*基因有望成为未来银屑病治疗靶标”。2010年,该团队在前期银屑病GWAS研究基础上进行了多阶段研究,又发现6个银屑病易感基因^[7],增加了银屑病的遗传风险因子,其中一些已被证实与某些自身免疫疾病相关。随后与深圳华大基因研究院的杨焕明、汪建、王俊团队合作^[14],通过外显子测序和靶向测序发现了7个常见和低频的非同义风险编码SNPs与银屑病发病风险相关,不仅揭示了疾病免疫失衡的机制和关键调控靶点,也为新的免疫干预治疗策略提供理论基础和实验依据。此后,安徽医科大学皮肤病研究所团队^[15~20]又陆续完成系统性红斑狼疮(SLE)、麻风、白癜风、特应性皮炎的系列GWAS研究,相关成果入选2010年度“中国科学十大进展”。这些易感基因/位点的发现为阐明疾病的遗传学发病机制提供重要理论依据,也为国内其他团队研究复杂疾病提供了宝贵的经验。

中国科学家在肿瘤、精神/神经疾病、口腔疾病等疾病易感基因研究上也做出了非凡成就。2010年,北京蛋白质组研究中心贺福初团队^[21]利用GWAS方法发现1p36.22区域是肝癌新的易感位点,提示*KIF1B*, *UBE4B*或*PGD*基因相关通路可能参与肿瘤的发病机制。第二军医大学孙颖浩研究团队^[22]在中国人群中发现2个前列腺癌易感位点;中国科学院昆明动物研究所张亚平研究团队^[23]通过GWAS鉴别出了中国汉族人群重型痤疮的2个易感位点等,基因表达研究显示,痤疮患者皮损中*SELL*基因表达水平明显升高,从而促进痤疮患者炎症反应加重。上海交通大学Bio-X研究院贺林研究团队^[24]通过GWAS发现染色体8p12区域和1q24.2区域内的基因变异影响脑组织中*ASH2L*, *DDHD2*, *PPAPDC1B*, *LETM2*和*FGFR1*基因表达;在精神分裂症和重度抑郁症患者脑海马组织中, *FGFR1*表达水平明显升高从而影响儿茶酚胺能神经元的功能,从而可能与精神分裂症的发病机制相关。中国医学科学院和北京协和医学院林东昕研究团队^[25]在中国人胰腺癌的GWAS研究中分别发现5个肿瘤相关易感基因,为胰腺癌的治疗提供分子靶标。广州中山大学曾益新研究团队^[26]在鼻咽癌的GWAS研究中发现3个易感基因/位点,提示转化生长因子 β 和c-Jun N端激酶信号通路在鼻咽癌发病过程中

的重要作用。中国医学科学院阜外医院顾东风研究团队^[27]利用GWAS在汉族人群中发现了4个冠状动脉疾病易感位点。华中科技大学同济医学院马丁研究团队^[28]在宫颈癌的GWAS研究中成功发现2个易感位点,进一步强化了遗传免疫学及一些致癌因素通过影响持续感染及HPV的共同作用机制在宫颈癌的发病风险中发挥重要作用的理论,为宫颈癌的遗传学研究拓展了新的思路。南京医科大学沈洪兵研究团队^[29]利用GWAS成功鉴定出4个肺癌风险易感区域,其中2个位点(13q12.12和22q12.2)是首次报道的中国人群肺癌易感位点,13q12.12区域的*TNFRSF19*可能参与c-Jun氨基末端激酶通路,调节非caspase依赖性的细胞死亡。22q12.2区域的*LIF*可能通过活化Janus激酶/信号传导和转录激活因子3途径参与肺癌的发病机制。华中科技大学邬堂春研究团队^[30]发现位于6p21.3区域的*HSPA1B*基因启动子多态性与中国汉族人群肺癌的发生风险和预后有关。中南大学夏昆团队^[31]鉴定一个自闭症易感位点,该区域的*TRIM33*基因不仅在其他与自闭症有着许多相似特征的精神疾病有关,且在自闭症患者的大脑中表达有显著差异,因此为揭示自闭症的发病机制提供新的思路。北京大学精神卫生研究所张岱团队^[32]鉴定出一个与精神分裂症相关的易感位点,该位点的*NKAPL*基因可能在早期神经发育过程中参与调节神经元的迁移;另一个基因*ZKSCAN4*可能通过影响出生后的的大脑发育过程参与精神分裂症的发病机制。河南新乡医学院癌症研究中心王立东^[33]在人类第10号和20号染色体上首次发现两个食管癌易感基因, *PLCE1*基因可能参与调节细胞生长、分化、凋亡和血管生成。*C20orf54*基因负责运输核黄素作用,而核黄素缺乏已被证明是食道鳞状细胞癌和贲门腺癌的危险因素,随后中国医学科学院肿瘤医院肿瘤研究所林东昕研究团队与王立东团队^[34]合作,通过扩大样本量再次发现两个食管癌相关易感位点,其中*TMEM173*基因可通过产生1型干扰素,促进对病毒和细菌的先天免疫反应参与肿瘤的发生发展过程,为中国食管癌发病机制研究提供了重要的遗传学证据。武汉大学口腔医院边专团队^[35]发现14个新的非综合征型唇腭裂相关易感基因,大大增加*NSCLP*的遗传易感性位点的数量以及对颅面畸形遗传病因学的重要见解。安徽医科大学汪凯和北京大学第三医院邓敏研究团队^[36]通过GWAS发现两个肌萎缩侧索硬化易感位点,这两个位点可解释中国汉族人群该病12.48%疾病风险性,提示肌萎缩性脊髓侧索硬

化症新的发病机制以及遗传异质性。复旦大学余龙团队^[37]运用GWAS发现两个与肝癌(乙肝)相关的易感基因。山东省皮肤病与性病防治研究所张福仁^[17]通过GWAS发现*NOD2*, *RIPK2*, *TNFSF15*, *HLA-DRB1*和*PARK2*基因遗传变异,通过调节干扰素- γ 水平,影响NF- κ B和*NOD2*介导的生物学信号通路而参与其疾病可能的发病机制,随后再次利用GWAS深入分析,发现两个与麻风病相关的易感基因,揭示了*IL23R*在感染性疾病易感性中的作用,提示自吞噬作用参与麻风发病机制^[38]。第二军医大学附属长征医院徐沪济研究团队^[39]在中国汉族人群类风湿关节炎中发现了两个新的易感位点,揭示了不同人群类风湿性关节炎遗传背景的差异。北京协和医院张奉春团队^[40]通过GWAS鉴定出一个与原发性干燥综合征相关的易感位点,提示T细胞和B细胞免疫信号在该病中的作用。中国科学院上海生命科学研究院林旭团队^[41]发现两个中国汉族人群T2D特有的易感位点,揭示不同人群2型糖尿病遗传易感性的差异,并揭示该病 β 细胞受损和胰岛素敏感性降低的机制。中山大学附属第一医院余学清研究团队^[42]则从中国汉族人群中鉴定出了IgA肾病的多个易感位点,进一步证明了免疫和炎症中涉及的遗传多态性可以影响IgA肾病的易感性和临床表现。上海交通大学医学院附属仁济医院马雄团队、东南大学生命科学研究院刘向东课题组及苏州大学附属第一人民医院陈卫昌团队^[43]的合作研究,揭示了染色体6个区域内的基因与原发性胆汁性胆管炎(PBC)的发病机制相关,支持*IL21*信号通路和Tfh细胞参与原发性胆汁性胆管炎发病的假说。人类疾病基因研究四川省重点实验室杨正林研究团队^[44]鉴定出一个与高度近视眼相关的易感基因,表明不同种族人群存在遗传异质性,这为高度近视的发病机制提供了有价值的参考等。以上GWAS成果进一步促进了复杂疾病的遗传学发病机制研究,为疾病的风险预测、临床诊断、药物开发及个体化用药指导等提供了理论基础,相关成果均在国际顶级杂志*N Engl J Med*, *Nature*及其子刊等发表。多项成果入选贺林院士^[45]撰写的我国改革开放40年医学遗传学辉煌成就中(表S1)。

复旦大学金力和新加坡国立研究中心刘建军^[46,47]在近年来开展的GWAS研究中发现,不同人群、不同地域的群体存在明显的差异。分析汉族人群的GWAS数据,发现中国的南北人群间存在着明显的遗传异质性。在最近开展的SLE多人群GWAS Meta分析中,发现

了HLA区域和*IRF5*位点在不同人群中的显著遗传异质性^[48]。疾病在不同人群中的遗传异质性提示,不同人群中存在特异的疾病生物学标记,进一步证明开展中国人群特异性GWAS研究在揭示中国人群疾病特异性发生机制、发掘潜在关键性分子标志物,以及开发适应中国人群的特异性新药治疗靶点等方面有重要意义^[49]。

在中国科学院文献情报中心与汤森路透旗下的知识产权与科技事业部联合举办和发布的《2014研究前沿》中,“全基因组关联研究在人类疾病等复杂性状研究中的应用”被称为生物科学领域年均被引频次最高的热点研究前沿。2016年实施的“精准医学研究”是被优先启动的重点专项之一,其中涉及多类常见高发、危害重大疾病的基因组学研究。利用GWAS的研究方法,为引领遗传学、临床医学等各个相关学科的研究,以及生物医学发展带来新的契机。由此可见,对中国人群开展更多、更大规模的GWAS研究具有至关重要的意义。

2.3 国际疾病基因组学研究论坛

中国复杂疾病易感基因研究整体水平达到国际先进,受到国际学术界高度关注和认可。为进一步推动全基因组关联研究在中国的应用,促进国际合作,2011年,由国际知名的自然出版集团与安徽医科大学共同主办的第一届全基因组关联研究国际论坛在合肥隆重举行。为提升全基因组关联研究国际论坛学术影响力,自然出版集团随后与中国GWAS科研团队多次联手,先后于2012~2019年分别在杭州、上海、北京、长沙、深圳、广州、成都、南京又连续成功举办了共计九届全基因组关联研究国际论坛,搭建高端学术平台,增强中国疾病基因组研究领域的国内外学术交流与合作。该项国际盛会的多次召开,得到*Nat Genet*主编的高度认可,先后3次发表专论称赞会议的成功,具有即时的、广泛的战略意义。2011年,首届论坛成功召开,*Nat Genet*主编发表专论“Milestone in Anhui”,充分强调开展多国际中心的全基因组关联研究有助于深入挖掘复杂疾病/性状的生物学发病机制^[50]。2012年,第二届论坛成功召开,*Nat Genet*主编发表专论“ Asking for more”,指出随着全基因组关联研究的应用,需要对前期发现的易感位点进行精细定位研究,进一步发现与疾病相关的功能变异,从而鉴别疾病易感基因^[51]。2016年,第六届论坛成功召开,*Nat Genet*主编发表专论“Genome var-

iation in precision medicine”, 在全外显子组范围内搜寻与疾病相关的变异有利于推动精准医学的发展, 安徽医科大学张学军教授研究团队^[14]利用全基因组外显子测序和靶向测序, 鉴定MHC区域可能存在多个与疾病相关的编码变异, 该项工作的成功开展显示器官移植供体和受体遗传多样性和分布, 可能成为未来健康研究的核心^[52]。本论坛旨在围绕疾病遗传学和基因组学研究领域的前沿和热点, 通过时任*Nat Genet*主编Myles Axton等权威专家学者的主题报告和专题讨论的形式, 探讨疾病遗传学和基因组学发展方向和趋势, 推进疾病遗传学和基因组学研究成果向临床转化应用, 增进学术交流, 促进国际合作, 搭建国内外合作交流平台, 进而为如何利用发现的疾病潜在治疗靶点和生物学标记服务社会公共卫生和提升人类健康水平提供依据。

3 GWAS研究的扩展

随着GWAS在多种疾病中的广泛应用, 研究者们开始察觉GWAS发现的绝大部分变异并不能够完全解释疾病的发病风险(OR 1.1~1.5)^[53]。许多复杂疾病是由罕见变异造成的^[54], 而这种基于芯片的GWAS在实验的最开始就漏掉了这部分信息, 因而无法解释复杂的遗传特征。因此, GWAS后续研究亟须扩展新的方法和研究领域, 如高通量测序、表观基因组学研究、基因型填补研究、eQTL及药物基因组学研究等方法, 来捕获罕见的致病性变异。这些新方法发现的致病性变异将与GWAS研究成果形成互补, 更准确地揭示疾病复杂的致病机理。

3.1 高通量测序

2006年, 以Illumina/Solexa为代表的新一代测序仪的推出, 为基因组多态性的识别带来新的解决方案。人类基因组绝大部分由重复序列和非蛋白编码区域构成, 而外显子区域, 即蛋白编码区, 仅占人类基因组的1%, 是基因组中与功能最为相关的部分, 85%的功能变异都潜藏在外显子中。因此, 选择性地针对全基因组编码区(外显子组)进行测序, 将大大降低测序费用, 同时获取最有价值的信息^[55]。随着2008年Ng等人^[56]首次对一个体外显子组的变异进行全面分析, 多项运用新一代测序技术进行全基因组外显子测序的研究已经相继开展。Sarah等人^[57,58]利用Illumina Genome Analyzer II系统, 先后对4例弗里曼谢尔登综合征(FSS)和4例米勒综合征(Miller syndrome)进行外显子测序, 通过多步骤分

类过滤法, 准确鉴定出FSS的致病基因*MYH3*的突变, 并成功发现了米勒综合征的致病基因*DHODH*。外显子测序也同样成功地应用于各种肿瘤(如胰腺癌^[59]、类胶质母细胞瘤^[60]等多基因复杂疾病)。寻找易感基因的研究中, Jones等人^[61]通过对一例胰腺癌患者的全部外显子序列分析, 发现了一个新的易感基因*PALB2*, 该基因的突变可以解释约3%的家族性胰腺癌的发生。Huyghe等人^[62]利用基于全基因组测序技术的GWAS研究, 发现结直肠癌人群在*CHD1*上有一个保护型的强关联变异信号。同时通过对125478人的大数据进行整合, 开展Meta分析, 又发现了40个新的独立遗传信号($P < 5 \times 10^{-8}$), 从而将结直肠癌已发现的独立遗传信号的数量扩增到约100个。这些研究充分证实了使用全基因组外显子测序来寻找遗传性疾病的致病/易感基因策略的可行性和有效性。这种基于其高通量、低成本的优点, 通过全基因组水平的深度测序, 不仅能够识别在人群中发生概率较高的多态性位点, 对于那些从前无法识别的低频多态性位点也可以精确检出, 因此能够更为有效地发现疾病的易感或致病基因, 进而发现疾病新的发病机制。Sino-Danish利用新一代测序技术对数千名患者和对照进行外显子测序, 涉及内脏肥胖症、2型糖尿病、高血压等代谢病, 以期发现与代谢疾病相关的常见变异和低频变异。相信随着新一代测序技术的发展, 以数据为导向, 对个体进行基因组水平、大规模研究, 即在掌握基因组全局的情况下, 直接发现影响疾病的基因位点的方法将成为继GWAS后研究复杂疾病易感基因最为有效的方法。

3.2 表达数量性状位点(eQTL)分析方法

表达数量性状位点(eQTL)分析方法是2001年由Jansen和Nap^[63]提出。原理是通过将整个基因组中所有基因多态性与组织中基因的mRNA表达水平进行定量作图, 发现特定组织对特定的基因表达存在影响的基因多态性。eQTL是基因组内与基因表达水平连锁或关联的遗传变异位点, 主要通过mRNA稳定性、DNA甲基化程度、基因间相互作用和调控等来影响这些基因的转录水平, 改变关键的蛋白量, 从而影响表型。因此, 大多数GWAS发现的性状相关SNP位点可能具有eQTL效应^[64]。

应用GWAS发现的大量疾病相关易感位点位于非编码序列, 对其发病机制的解释多局限于位点对mRNA表达水平的影响。而eQTL分析主要研究mRNA

水平与基因组的关联性,进而探究遗传变异与基因表达关系、基因间相互作用和基因调控网络等.将GWAS与eQTL联合应用,可为阐明非编码区遗传变异影响疾病表型的作用机制提供有力的方法学支持.因此,将eQTL分析应用于疾病基因组学研究,为发现更多影响疾病发病机制的功能性遗传变异提供有效研究手段^[65],为疾病临床个体化诊疗的开展提供理论依据.

3.3 药物基因组学

药物基因组学是由遗传学衍生出来的一门新兴学科,旨在运用高通量技术搜寻所有与个体药物反应差异相关的变异位点、mRNA及其编码蛋白^[66].将GWAS分析方法更好地应用于临床,发现更多与药物疗效及毒性相关的生物标志物,为临床个体化用药的开展提供遗传学证据.

国内外银屑病的药物基因组学研究现已陆续开展并取得了一定成果.甲氨蝶呤、环孢素和阿维A是传统治疗中重度银屑病的一线系统用药,研究发现,*ABCC1*(rs35592, rs2238746, rs28364006), *ABCG2*(rs17731538, rs13120400), *FOXP3*(rs3761548)和HLA-Cw*0602等与甲氨蝶呤治疗银屑病的疗效有关^[67-69].环孢素治疗银屑病产生的个体差异与*CYP3A4*, *CYP3A5*及*ABCB1*的多态性有关^[69,70].阿维A与疗效间的关系则主要受到*VEGF*上SNPs变异的影响^[71].国内在该项领域中的研究较国外相对缺乏且大多仍属于遗传药理学范畴,国内团队利用候选基因法发现了*VEGF*和*EGF*与银屑病发病及阿维A疗效相关^[71];国外已开展了基于GWAS的药物基因组学和基于转录组的药物疗效研究^[72,73],发现*TNF- α* 拮抗剂疗效与*TLR10*等基因相关,并发现Th细胞相关基因(Th1, Th17和Th22)mRNA的表达水平在服用甲氨蝶呤后明显降低.目前药物基因组学仍处于起步阶段,其结果已被讨论纳入常规药物审查项目.未来该领域研究将会在预测个体药物疗效及不良反应方面起到重要作用,在临床实验中收集大量遗传学数据将是今后研究的一个主要方向^[74].

4 GWAS数据的深入分析

4.1 功能学研究

基因功能学研究既能研究基因功能,也可以对预测后的基因加以验证.目前的GWAS已经发现了大量

的变异位点,可以通过结合已发表的公共功能基因组学数据库,检测调控变异位点功能、明确变异位点靶基因等步骤进行功能学的探索^[75].主要的实验学方法包括基因转导、人工染色体的转导^[76]、基因编辑新技术、微阵列分析^[77,78]、单倍体胚胎干细胞研究^[79]等.

(i) 基因功能的预测.结构决定功能是众所周知的,对基因功能的研究可以从结构上进行合理地预测或运用生物信息学方法,再通过实验学方法进行验证加以确定.研究人员已用微阵列表达数据、蛋白质的三维结构、蛋白质结构域、蛋白-蛋白相互作用网络^[80]、系统发生谱来预测基因的功能^[81].

(ii) 基因的时空表达谱分析.基因的时空表达谱包括蛋白质和mRNA两个水平^[82],基因在个体的不同组织和细胞中以及不同的发育阶段表达均不相同,所以在研究基因功能之前,要分析基因的时空表达谱.常用的方法有原位杂交、免疫组织化学、蛋白印迹技术、实时PCR技术等.

4.2 新的分析方法的发展

近年来,各国研究者正在将一些新的策略和方法尝试应用到复杂疾病/性状GWAS的后续研究中,如Meta分析、基因型填补、Gene-based分析及基于通路的关联分析等.这些新的分析方法弥补了经典GWAS的一些不足,能够发现更多复杂疾病/性状的易感基因,进一步完善了人类对复杂疾病/性状遗传学机制的认识^[83].

(i) Meta分析.目前的GWAS样本量多为1000病例/1000对照左右,发现的疾病相关SNPs基本上都属于常见(MAF>0.2)且具有中度效力(OR $\geq$ 1.2)的变异;而对于频率低(MAF<0.05)、效力弱(OR值接近1)的SNPs,由于统计把握度较低很难被检出,因此增加GWAS样本量是提高检验效能最直接和有效的方式.但是鉴于GWAS巨大的工作量以及昂贵的实验费用,单个GWAS样本量可能难以达到研究需要.而Meta分析可以通过合并多个研究数据的方式增加样本含量,提高统计把握度,有利于发现新的易感位点^[80,84].GWAS的Meta分析不仅可以检出常见变异,同样可以发现一些强效的罕见变异,是发掘现有GWAS数据,鉴定更多疾病易感基因的一种经济和高效的方法.

(ii) Imputation. 基因型填补(genotype imputation)是一种基于观察对象已有基因型,对未进行基因分型(genotyping)的位点之等位基因(allele)进行预测的方法.

填补能增加GWAS中单核苷酸多态性(single-nucleotide polymorphism, SNP)的密度,使得在已经发现的关联性位点周围寻找疾病位点(disease locus)成为可能,同时也能提高对采用了不合适的标签(tagger)SNP进行标记的SNP的检验效能。基因型填补也被用于GWAS数据的Meta分析。

由于不同的GWAS所采用的芯片不同,SNP的密度和个数也不尽相同。对现有数据采用同一数据库作为模板进行填补,可以将各研究得到的数据放大到相同尺度和密度上,从而在同一个位点综合多个研究的信息^[85,86]。基因型填补还能估计现有SNP中基因分型失败的个体信息,增加有效样本量。此外,GWAS的验证步骤非常重要,在某个GWAS发现可能具有科学意义的位点后,往往需要在另一个独立研究人群中进行验证。若其他研究中未对此位点进行基因分型,可将该位点利用现有数据进行填补,再进行验证。

(iii) Gene-based分析。GWAS主要以单个SNP的统计分析作为核心研究手段。单个SNP的分析往往只挑选最具有统计学意义的SNP(其假阳性也最小),舍弃了其他统计学意义相对较弱的SNP,这样就遗漏了与疾病存在关联但关联效应较弱的基因,而这种基因可能占据了易感基因的大多数;另一方面,许多不是最具有统计学意义的基因上可能有多个相对独立的较具有统计学意义的SNP。越来越多研究者将目光放在对GWAS数据的深度挖掘和生物信息学分析上^[87-91]。生物网络分析是生物信息学处理GWAS数据的核心手段。基因是基本的分析单元,它反映了基因上所有或者部分最重要的SNP对疾病的综合遗传关联结果,因此在网络分析之前,需要通过统计学方法将基因上全部或者部分单个SNP的遗传关联结果综合起来,即基因水平的关联分析(gene-based分析)^[92]。

(iv) Pathway分析。将GWAS所有分型的SNPs按照不同的生物学通路排列,然后比较各个通路在病例/对照间的差异,这就是基于通路GWAS的研究思路^[93]。基于通路的GWAS和基于单个位点的GWAS是两种互补的分析方法,将两者有效地结合可以发现更多的疾病易感基因或者相关生物学通路。由于疾病相关的生物学通路能够更准确反映疾病的发病机制,对于新药的开发和临床诊疗技术的提高具有更大价值。

此外还包括,不同疾病共同易感基因^[18,94-97]、主成分分析和傅里叶分析^[98,99]、基于Permutation分析^[100]等研究方法。

5 GWAS的临床转化

如上所述,GWAS研究显示了其在揭示复杂疾病遗传学发生机制中的强大优势,并为多种复杂疾病积累了丰富的基因组变异位点。然而,这些位点在复杂疾病中往往发挥低到中等的效应,如何有效地利用这些基因组变异数据,更好地为临床工作服务是目前研究的热点。早在2011年,当GWAS研究发现复杂疾病/性状1000余个易感变异时,Hirschhorn和Gajdos^[101]即针对“GWAS早期几年成果在临床医学上的应用”进行了探讨,推测其在疾病个体化医疗的预测、治疗以及阐释人类生理和病理学方面的重要作用。而在2013年,当GWAS研究已经发现约2000个复杂疾病关联位点时,鉴于人们对GWAS发现在临床应用方面的怀疑,主要包括关联变异有限的效应尺度以及变异所致尚不明确的功能结局,Manolio^[102]在*Nat Rev Genet*撰文,举例和总结了GWAS成果在临床的应用。目前GWAS成果已应用于临床,主要表现在以下几个方面。

5.1 疾病风险预测

通过结合多个已发现的疾病相关易感基因,在早期就可预测未来患病的风险,可用于疾病早期筛查和干预。目前常用的预测模型是利用已知的疾病易感位点,计算个体多基因风险分数(PRS),预测疾病的发病风险^[103]。利用GWAS所发现位点构建模型进行疾病预测已见于多篇报道,包括类风湿性关节炎、2型糖尿病、结直肠癌^[104-107]等。以银屑病为例,我们的GWAS研究和广义GWAS研究已经发现了近90个银屑病易感基因^[80],深入阐释了银屑病的遗传学发生机制。本团队在2015年尝试地使用其中14个易感基因,所构建的遗传预测模型显示了很好的预测潜能,预测精度达到88%^[108],这为提前识别复杂疾病易感个体,尽早进行个性化干预提供了临床依据。

5.2 药物开发

复杂疾病的GWAS研究概括了多种疾病相关的已知生物学途径,也为识别新的生物学途径提供了很多线索,为临床新药的研发提供了理论基础。从而为疾病的治疗提供了更多的生物靶标。以骨质疏松为例,有报道目前批准的8类抗骨质疏松症治疗药物中有5种在临床试验中得到遗传数据的支持,是通过GWAS定位到的药物靶点^[109]。在其他多种疾病中也发现了

GWAS易感基因所对应的靶向药物^[110]。银屑病GWAS发现的易感基因*IL12B*, *IL23R*已被公司成功开发出优特克单抗、古赛库单抗、蒂尔他昔单抗等多种银屑病靶向性生物制剂^[6]。通过不同生物学通路易感基因的分析,能够发掘关键的生物学靶标,从而开发出作用更显著、特异性更强的新药,同时为不同疾病相同生物学通路中的老药新用提供依据。

5.3 药物临床指导

目前多种疾病的治疗指南中,对于如何选择药物治疗,参考的是循证医学的数据,然而即使是一线药物,在同一种疾病的不同个体中也会出现不同的疗效。最近一项甲氨蝶呤治疗银屑病的前瞻性研究中,伴有和不伴有关节病型银屑病的患者在甲氨蝶呤治疗12周后达到PASI75的分别为41.4%和50.5%^[111]。本团队基于前期GWAS所发现的*TNIP1*, *LCE3D*和*HLA-Cw6*等易感基因分析携带不同基因型的患者接受甲氨蝶呤治疗的疗效,结果发现携带*TNIP1*基因rs10036748位点的TT基因型能够将甲氨蝶呤治疗银屑病疗效从45%提高至70%,证明GWAS研究成果及技术(药物基因组学)在今后指导临床用药过程中将发挥越来越重要的作用。

5.4 药物不良反应预测

除了同种药物疗效的因人而异外,药物对不同个体的不良反应也有差异。正如我们团队上述研究中所报道的,甲氨蝶呤治疗伴有和不伴有关节病型银屑病的患者,出现头晕的分别为9.4%和0.9%,出现胃肠道症状的为25.0%和12.1%,出现肝脏毒性的为26.6%和15.0%。SLE药物治疗方面,因巯嘌呤甲基转移酶(TPMT)是免疫抑制剂巯嘌呤药物代谢过程中的关键酶,TPMT的5种主要突变(TPMT *1/*2/*3A/*3B/*3C)均可引起患者TPMT活性下降,导致患者使用标准剂量治疗SLE时发生严重的毒副作用,因此使用时需要根据基因型进行剂量调整^[112]。药物基因组学采用GWAS研究的原理,将同种药物治疗中不同疗效或不同不良反应进行分组,比较其中的基因组差异,从而获得与药物疗效或不良反应相关的基因组变异标记。在一项关于氨苯砜治疗麻风所引起的超敏反应综合征的研究中,*HLA-B*13:01*被确认是氨苯砜超敏反应综合征的风险因素,携带*HLA-B*13:01*对预测该综合征的敏感

性和特异性分别达到了85.5%和85.7%^[113]。

美国FDA(U.S. Food and Drug Administration)官网已经收录了多种药物基因组学的生物标记,囊括药物不良反应的风险以及基因型特异性剂量等合理用药的不同治疗方面(<https://www.fda.gov/Drugs/ScienceResearch>)。中日友好医院王辰研究团队^[114]牵头20余家国内三甲医院和机构共同成立“中国精准医学临床研究与应用联盟”,编写《精准医学:药物治疗纲要》专著,其中涵盖116种药物,为合理用药的精准治疗提供了坚实依据。药物在疾病治疗过程中的精准使用,将更大程度上提高疾病的治疗效果,具有巨大的应用价值。

6 GWAS与精准医学

2011年,美国科学院提出基因组学推进了生物医学信息学和临床医学的整合,迈进精准医学时代。2015年,时任美国总统奥巴马提出了“精准医学”计划,以提高医疗经费,推动个体化基因组学研究,为癌症及其他疾病患者制定个体医疗。我们团队与*Nat Genet*从2011年起每年共同举办国际疾病基因组学研究论坛,探讨和交流国际上基因组学研究的进展,自2013年开始,我们就将论坛的主题定为“Genome Variation in Precision Medicine”,并持续至今。精准医学是医学科技发展的前沿方向,已列入科技部“十三五”国家重点研发专项指南。基因组学信息和临床大数据是多种疾病研究的基石,同时基因组变异也是疾病发生机制研究的上游,寻找基因组变异是推动未来精准医学发展的一个重要环节。本团队2016年应邀在*J Genet Genomics*中发表“GWAS promotes precision medicine in China”,总结了我国GWAS的研究进展,并预测了GWAS研究对疾病精准医学的促进作用^[115]。

GWAS已被证实是复杂疾病遗传学机制高效的研究策略,目前已经成功运用于具有不同遗传背景的多个不同种群的复杂疾病,构建并不断丰富了疾病的遗传变异谱。随着GWAS研究的不断扩展和深入,必将发现疾病更多的遗传变异、易感基因和生物学通路。在未来的10年中,基因组“个体化”或“精准”医学将不断普及,并在常见疾病中不断开发出直接和间接的临床应用,为包括疾病预警、风险预测、临床诊断、药物开发及个体化用药指导等在内的精准医学奠定坚实的基础。

参考文献

- 1 Lander E, Kruglyak L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature*, 1995, 11: 241–247
- 2 Göring H H H, Terwilliger J D, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Human Genets*, 2001, 69: 1357–1369
- 3 Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. *Genetics*, 2003, 164: 829–833
- 4 Zhang X J, He P P, Wang Z X, et al. Evidence for a major psoriasis susceptibility locus at 6p21(PSORS1) and a novel candidate region at 4q31 by genome-wide scan in Chinese hans. *J Invest Dermatol*, 2002, 119: 1361–1366
- 5 Zhang X J, Yan K L, Wang Z M, et al. Polymorphisms in interleukin-15 gene on chromosome 4q31.2 are associated with psoriasis vulgaris in Chinese population. *J Invest Dermatol*, 2007, 127: 2544–2551
- 6 Zhang X J, Huang W, Yang S, et al. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat Genet*, 2009, 41: 205–210
- 7 Sun L D, Cheng H, Wang Z X, et al. Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat Genet*, 2010, 42: 1005–1009
- 8 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1996, 273: 1516–1517
- 9 International HapMap C. The International HapMap Project. *Nature*, 2003, 426: 789–796
- 10 Smith J G, Newton-Cheh C. Genome-wide association study in humans. *Methods Mol Biol*, 2009, 573: 231–258
- 11 Klein R J, Zeiss C, Chew E Y, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*, 2005, 308: 385–389
- 12 Han J W, Zhang X J. Current status of genome-wide association study (in Chinese). *Hereditas (Beijing)*, 2011, 33: 25–35 [韩建文, 张学军. 全基因组关联研究现状. *遗传*, 2011, 33: 25–35]
- 13 Zhang X J, He L, Sun L D. The progress of GWAS in China. *Nature*, 2012, 490: 316
- 14 Tang H, Jin X, Li Y, et al. A large-scale screen for coding variants predisposing to psoriasis. *Nat Genet*, 2014, 46: 45–50
- 15 Han J W, Zheng H F, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet*, 2009, 41: 1234–1237
- 16 Yang W, Tang H, Zhang Y, et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. *Am J Human Genets*, 2013, 92: 41–51
- 17 Zhang F R, Huang W, Chen S M, et al. Genomewide association study of leprosy. *N Engl J Med*, 2009, 361: 2609–2618
- 18 Quan C, Ren Y Q, Xiang L H, et al. Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. *Nat Genet*, 2010, 42: 614–618
- 19 Tang X F, Zhang Z, Hu D Y, et al. Association analyses identify three susceptibility Loci for vitiligo in the Chinese Han population. *J Invest Dermatol*, 2013, 133: 403–410
- 20 Sun L D, Xiao F L, Li Y, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat Genet*, 2011, 43: 690–694
- 21 Zhang H, Zhai Y, Hu Z, et al. Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat Genet*, 2010, 42: 755–758
- 22 Xu J, Mo Z, Ye D, et al. Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. *Nat Genet*, 2012, 44: 1231–1235
- 23 He L, Wu W J, Yang J K, et al. Two new susceptibility loci 1q24.2 and 11p11.2 confer risk to severe acne. *Nat Commun*, 2014, 5: 2870
- 24 Shi Y, Li Z, Xu Q, et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat Genet*, 2011, 43: 1224–1227
- 25 Wu C, Miao X, Huang L, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet*, 2011, 44: 62–66
- 26 Bei J X, Li Y, Jia W H, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet*, 2010, 42: 599–603
- 27 Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nat Genet*, 2012, 44: 890–894
- 28 Shi Y, Li L, Hu Z, et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat Genet*, 2013, 45: 918–922
- 29 Hu Z, Wu C, Shi Y, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. *Nat Genet*, 2011, 43: 792–796

- 30 Guo H, Deng Q, Wu C, et al. Variations in HSPA1B at 6p21.3 are associated with lung cancer risk and prognosis in Chinese populations. *Cancer Res*, 2011, 71: 7576–7586
- 31 Xia K, Guo H, Hu Z, et al. Common genetic variants on 1p13.2 associate with risk of autism. *Nature*, 2014, 19: 1212–1219
- 32 Yue W H, Wang H F, Sun L D, et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat Genet*, 2011, 43: 1228–1231
- 33 Wang L D, Zhou F Y, Li X M, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies a susceptibility locus at PLCE1. *Nat Genet*, 2010, 42: 759–763
- 34 Wu C, Wang Z, Song X, et al. Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nature*, 2014, 46: 1001–1006
- 35 Yu Y, Zuo X, He M, et al. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun*, 2017, 8: 14364
- 36 Deng M, Wei L, Zuo X, et al. Genome-wide association analyses in Han Chinese identify two new susceptibility loci for amyotrophic lateral sclerosis. *Nat Genet*, 2013, 45: 697–700
- 37 Jiang D K, Sun J, Cao G, et al. Genetic variants in *STAT4* and HLA-DQ genes confer risk of hepatitis B virus—Related hepatocellular carcinoma. *Nat Genet*, 2013, 45: 72–75
- 38 Zhang F, Liu H, Chen S, et al. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet*, 2011, 43: 1247–1251
- 39 Jiang L, Yin J, Ye L, et al. Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans. *Arthritis Rheumatol*, 2014, 66: 1121–1132
- 40 Li Y, Zhang K, Chen H, et al. A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjögren's syndrome at 7q11.23. *Nat Genet*, 2013, 45: 1361–1365
- 41 Li H, Gan W, Lu L, et al. A genome-wide association study identifies *GRK5* and *RASGRP1* as type 2 diabetes loci in Chinese hans. *Diabetes*, 2013, 62: 291–298
- 42 Yu X Q, Li M, Zhang H, et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet*, 2011, 44: 178–182
- 43 Qiu F, Tang R, Zuo X, et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. *Nature*, 2017, 8: 14828
- 44 Shi Y, Qu J, Zhang D, et al. Genetic variants at 13q12.12 are associated with high myopia in the Han Chinese population. *Am J Human Genet*, 2011, 88: 805–813
- 45 Cai L, Zheng L A, He L. The forty years of medical genetics in China. *J Genet Genomics*, 2018, 45: 569–582
- 46 Chen J, Zheng H, Bei J X, et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am J Human Genet*, 2009, 85: 775–785
- 47 Xu S, Yin X, Li S, et al. Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am J Human Genet*, 2009, 85: 762–774
- 48 Morris D L, Sheng Y, Zhang Y, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat Genet*, 2016, 48: 940–946
- 49 Cui Y, Sheng Y, Zhang X. Genetic susceptibility to SLE: Recent progress from GWAS. *J Autoimmun*, 2013, 41: 25–33
- 50 Editorial. Milestone in Anhui. *Nat Genet*, 2011, 43: 613
- 51 Editorial. Asking for more. *Nat Genet*, 2012, 44: 733
- 52 Editorial. Genome variation in precision medicine. *Nat Genet*, 2016, 48: 701
- 53 Manolio T A, Collins F S, Cox N J, et al. Finding the missing heritability of complex diseases. *Nature*, 2009, 461: 747–753
- 54 Ji W, Foo J N, O'Roak B J, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet*, 2008, 40: 592–599
- 55 Zhang X, Li M, Zhang X J. Exome sequencing and its application (in Chinese). *Hereditas (Beijing)*, 2011, 33: 847–856 [张鑫, 李敏, 张学军. 全基因组外显子测序及其应用. *遗传*, 2011, 33: 847–856]
- 56 Ng P C, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet*, 2008, 4: e1000160
- 57 Ng S B, Buckingham K J, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, 2010, 42: 30–35
- 58 Ng S B, Turner E H, Robertson P D, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 2009, 461: 272–276
- 59 Jones S, Zhang X, Parsons D W, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, 2008, 321: 1801–1806
- 60 Parsons D W, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 2008, 321: 1807–1812
- 61 Jones S, Hruban R H, Kamiyama M, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science*, 2009, 324:

- 62 Huyghe J R, Bien S A, Harrison T A, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*, 2019, 51: 76–87
- 63 Jansen R. Genetical genomics: The added value from segregation. *Trends Genets*, 2001, 17: 388–391
- 64 Gamazon E R, Huang R S, Cox N J, et al. Chemotherapeutic drug susceptibility associated SNPs are enriched in expression quantitative trait loci. *Proc Natl Acad Sci USA*, 2010, 107: 9287–9292
- 65 Hormozdiari F, van de Bunt M, Segrè A V, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Human Genets*, 2016, 99: 1245–1260
- 66 Ovejero-Benito M C, Muñoz-Aceituno E, Reolid A, et al. Pharmacogenetics and pharmacogenomics in moderate-to-severe psoriasis. *Am J Clin Dermatol*, 2018, 19: 209–222
- 67 Warren R B, Smith R L L, Campalani E, et al. Genetic variation in efflux transporters influences outcome to methotrexate therapy in patients with psoriasis. *J Invest Dermatol*, 2008, 128: 1925–1929
- 68 Indhumathi S, Rajappa M, Chandrashekar L, et al. Pharmacogenetic markers to predict the clinical response to methotrexate in south Indian Tamil patients with psoriasis. *Eur J Clin Pharmacol*, 2017, 73: 965–971
- 69 Vasilopoulos Y, Sarri C, Zafiriou E, et al. A pharmacogenetic study of ABCB1 polymorphisms and cyclosporine treatment response in patients with psoriasis in the Greek population. *Pharmacogenomics J*, 2014, 14: 523–525
- 70 Naesens M, Kuypers D R J, Sarwal M. Calcineurin inhibitor nephrotoxicity. *Clin J Am Soc Nephrol*, 2009, 4: 481–508
- 71 Young H S, Summers A M, Read I R, et al. Interaction between genetic control of vascular endothelial growth factor production and retinoid responsiveness in psoriasis. *J Invest Dermatol*, 2006, 126: 453–459
- 72 Nishikawa R, Nagai H, Bito T, et al. Genetic prediction of the effectiveness of biologics for psoriasis treatment. *J Dermatol*, 2016, 43: 1273–1277
- 73 Goldminz A M, Suárez-Fariñas M, Wang A C, et al. *CCL20* and *IL22* messenger RNA expression after adalimumab vs methotrexate treatment of psoriasis. *JAMA Dermatol*, 2015, 151: 837–846
- 74 Pincelli C, Pignatti M, Borroni R G. Pharmacogenomics in dermatology: From susceptibility genes to personalized therapy. *Exp Dermatol*, 2009, 18: 337–349
- 75 Gallagher M D, Chen-Plotkin A S. The post-GWAS era: From association to function. *Am J Human Genets*, 2018, 102: 717–730
- 76 Kim J H, Kononenko A, Erliandri I, et al. Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. *Proc Natl Acad Sci USA*, 2011, 108: 20048–20053
- 77 Liu Y J, Zhang J Y. Time-series microarray data simulation modeled with a case-control label. *Genet Mol Res*, 2016, 15: 15027287
- 78 Huggins C E, Domenighetti A A, Ritchie M E, et al. Functional and metabolic remodelling in GLUT4-deficient hearts confers hyper-responsiveness to substrate intervention. *J Mol Cell Cardiol*, 2008, 44: 270–280
- 79 Horii T, Morita S, Kimura M, et al. Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *PeerJ*, 2013, 1: e230
- 80 Tsoi L C, Stuart P E, Tian C, et al. Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat Commun*, 2017, 8: 15382
- 81 Phuong T, Nhung N. Predicting gene function using similarity learning. *BMC Genomics*, 2013, 14: S4
- 82 Nair R P, Duffin K C, Helms C, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF- κ B pathways. *Nat Genet*, 2009, 41: 199–204
- 83 Quan C, Zhang X J. Research strategies for the next step of genome-wide association study (in Chinese). *Hereditas (Beijing)*, 2011, 33: 100–108 [权晟, 张学军. 全基因组关联研究的深度分析策略. *遗传*, 2011, 33: 100–108]
- 84 Yin X, Low H Q, Wang L, et al. Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. *Nat Commun*, 2015, 6: 6916
- 85 Zeggini E, Scott L J, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*, 2008, 40: 638–645
- 86 Lessard C J, Adrianto I, Ice J A, et al. Identification of IRF8, TMEM39A, and IKZF3-ZBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. *Am J Human Genets*, 2012, 90: 648–660
- 87 Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Human Genets*, 2007, 81: 1278–1283
- 88 Peng G, Luo L, Siu H, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet*, 2010, 18: 111–117
- 89 Jia P, Zheng S, Long J, et al. dmGWAS: Dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*, 2011, 27: 95–102
- 90 Eleftherohorinou H, Wright V, Hoggart C, et al. Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. *PLoS One*, 2009, 4: e8068

- 91 Baranzini S E, Galwey N W, Wang J, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Human Mol Genets*, 2009, 18: 2078–2090
- 92 Salvi E, Wang Z, Rizzi F, et al. Genome-wide and gene-based meta-analyses identify novel loci influencing blood pressure response to hydrochlorothiazide. *Hypertension*, 2017, 69: 51–59
- 93 Christoforou A, Espeseth T, Davies G, et al. GWAS-based pathway analysis differentiates between fluid and crystallized intelligence. *Genes Brain Behav*, 2014, 13: 663–674
- 94 Stadler Z K, Vijai J, Thom P, et al. Genome-wide association studies of cancer predisposition. *Hematol/Oncol Clin North Am*, 2010, 24: 973–996
- 95 Kochi Y, Okada Y, Suzuki A, et al. A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat Genet*, 2010, 42: 515–519
- 96 Jin Y, Birlea S A, Fain P R, et al. Common variants in FOXP1 are associated with generalized vitiligo. *Nat Genet*, 2010, 42: 576–578
- 97 Rafnar T, Sulem P, Stacey S N, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet*, 2009, 41: 221–227
- 98 Peng Q, Zhao J, Xue F. PCA-based bootstrap confidence interval tests for gene-disease association involving multiple SNPs. *BMC Genet*, 2010, 11: 6
- 99 Gao Q, He Y, Yuan Z, et al. Gene- or region-based association study via kernel principal component analysis. *BMC Genet*, 2011, 12: 75
- 100 Liu J Z, McRae A F, Nyholt D R, et al. A versatile gene-based test for genome-wide association studies. *Am J Human Genets*, 2010, 87: 139–145
- 101 Hirschhorn J N, Gajdos Z K Z. Genome-wide association studies: Results from the first few years and potential implications for clinical medicine. *Annu Rev Med*, 2011, 62: 11–24
- 102 Manolio T A. Bringing genome-wide association findings into clinical use. *Nat Rev Genet*, 2013, 14: 549–558
- 103 Tada H, Melander O, Louie J Z, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J*, 2016, 37: 561–567
- 104 Bao W, Hu F B, Rong S, et al. Predicting risk of type 2 diabetes mellitus with genetic risk models on the basis of established genome-wide association markers: A systematic review. *Am J Epidemiol*, 2013, 178: 1197–1207
- 105 Yarwood A, Han B, Raychaudhuri S, et al. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. *Ann Rheum Dis*, 2015, 74: 170–176
- 106 Scott I C, Seegobin S D, Steer S, et al. Predicting the risk of rheumatoid arthritis and its age of onset through modelling genetic risk variants with smoking. *PLoS Genet*, 2013, 9: e1003808
- 107 Frampton M J E, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol*, 2016, 27: 429–434
- 108 Yin X, Cheng H, Lin Y, et al. A weighted polygenic risk score using 14 known susceptibility variants to estimate risk and age onset of psoriasis in Han Chinese. *PLoS One*, 2015, 10: e0125369
- 109 Sabik O L, Farber C R. Using GWAS to identify novel therapeutic targets for osteoporosis. *Transl Res*, 2017, 181: 15–26
- 110 Visscher P M, Wray N R, Zhang Q, et al. 10 Years of GWAS discovery: Biology, function, and translation. *Am J Human Genets*, 2017, 101: 5–22
- 111 Yan K, Zhang Y, Han L, et al. Safety and efficacy of methotrexate for Chinese adults with psoriasis with and without psoriatic arthritis. *JAMA Dermatol*, 2019, 155: 397
- 112 Dean L. Azathioprine therapy and TPMT genotype. In: Pratt V, McLeod H, Rubinstein W, et al., eds. Medical Genetics Summaries [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2012. 1–7
- 113 Zhang F R, Liu H, Irwanto A, et al. *HLA-B*13:01* and the dapsone hypersensitivity syndrome. *N Engl J Med*, 2013, 369: 1620–1628
- 114 Wang C, Yao S K. Precision Medicine: Outline of Drug Treatment (in Chinese). Beijing: People's Medical Publishing House, 2016 [王辰, 姚树坤. 精准医学: 药物治疗纲要. 北京: 人民卫生出版社, 2016]
- 115 Sun L, Zhang X, He L. GWAS promotes precision medicine in China. *J Genets Genomics*, 2016, 43: 477–479

补充材料

表S1 中国的GWAS研究

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “人类复杂疾病全基因组关联研究”

Review of genome-wide association study

Xuejun Zhang^{1,2}

¹ *Institute of Dermatology, Fudan University (Affiliated Huashan Hospital), Shanghai 200000, China;*

² *Department of Dermatology at No.1 Hospital Institute of Dermatology, Institute of Dermatology, Anhui Medical University, Hefei 230000, China*

E-mail: ayzxj@vip.sina.com

Common diseases are caused by the combined effects of genetic and environmental factors with obvious genetic heterogeneity and phenotypic complexity. These diseases have a high incidence and seriously affect the physical and mental health of patients. With the consecutive completions of the Human Genome Project (HGP), the International HapMap Project and the 1000 Genomes Project, the detailed map of human genomic variations has been gradually decoded. Along with the rapid development of the technology and the rapid reduction of the cost of genotyping, genome wide association studies (GWAS) have become the most efficient and popular method to identify the susceptibility genes of common diseases and traits. Based on the principle of “common disease, common variation”, GWAS can explore the differences in allele frequencies between different groups using association analysis at the whole genome level. Following the first successful application of GWAS on age-related macular degeneration in 2005, a large number of susceptibility genes have been identified for common diseases and traits. In China, there are 179 GWAS on 127 kinds of common human diseases, and traits have been determined in a total sample size of 300000 cases and 400000 controls. These studies have identified a large number of susceptibility genes and revealed plentiful potential biological pathways that have tremendously expanded the mechanism of these common diseases and traits. To expand the application fields, several novel technologies based on the results or principle have been performed to further identify the genetic pathogenesis of common diseases, including high-throughput sequencing, epigenomic study, imputation analysis, expression quantitative trait loci (eQTL) analysis and pharmacogenomic study. To adequately utilize the massive genome data acquired from GWAS, many downstream analyses have also been performed with these vast databases. First, to better understand the exact mechanism of a susceptibility gene in the pathogenesis of a disease, functional studies at different levels have been widely performed. In addition, new analysis methods were also developed to further elucidate the underlying results of these data, including meta-analyses with data from different platforms, different sample cohorts or different ethnic groups, gene-based analysis and pathway analysis. Another research field is how to translate these findings to clinical applications. To date, several explorations have been performed using the GWAS principle and data in predictions of disease risk and drug effects, while various susceptibility genes have been proven to be the targets of several biological agents. With the fast development of this technology, GWAS have provided novel ideas and methods to explore the genetic pathogenesis of common diseases, which revealed the basis of the occurrence, development and treatment of these diseases and laid a solid foundation for precision medicine. This paper reviews the rise, development, expansion, clinical transformation and future of GWAS.

genome-wide association study, common diseases, susceptibility genes, precision medicine

doi: [10.1360/TB-2019-0063](https://doi.org/10.1360/TB-2019-0063)