

神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤及肾母细胞瘤中全基因组关联分析计量和热点研究

林彩侠¹,许春伟²

(1. 江山市人民医院,浙江 江山 324100;2.中国人民解放军北京军区总医院,北京 100700)

摘要:[目的] 分析神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤及肾母细胞瘤全基因组关联研究(GWAS)英文文献的数据构成特征。[方法] 对国际最新GWAS研究数据库之一——美国国立人类基因组研究所(National Human Genome Research Institute, NHGRI)数据库 2008 年 11 月 25 日至 2013 年 8 月 26 日收录的神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤和肾母细胞瘤相关文献基本特征、第一阶段样本量、二阶段样本量、染色体区域相关单核苷酸多态性(SNP)、对照组相关风险等位基因频率、P 值文献中最强 SNP、优势比(odds ratio, OR)或 β 相关系数和实验平台及检测 SNP 数量 8 项数据进行分析。[结果] 共检索 7 篇 4 种肿瘤 GWAS 研究文献(神经母细胞瘤 3 篇,多形性胶质母细胞瘤 1 篇,尤文肉瘤 1 篇和肾母细胞瘤 1 篇),涉及 27 个 SNP(神经母细胞瘤 10 个,多形性胶质母细胞瘤 1 个,尤文肉瘤 3 个,肾母细胞瘤 13 个),以在 *Nat Genet* 发表的文献最多。文献第一阶段病例组样本量从 315 例增加到 2101 例,对照组从 1879 例增加到 3851 例;第二阶段病例组样本量从 434 例增加到 1488 例,对照组从 199 例增加到 3851 例。染色体区域相关 SNP 以 2,5,6 和 11 号居多。对照组相关风险等位基因频率以 0.40~ 组最多,未发现稀有变异。最强 SNP 相关风险等位基因 P 值以 1×10^{-10} ~ 为主,最小 P 值为 2×10^{-26} 。OR 或 β 相关系数范围为 1.23~2.64。研究平台以 Illumina 为主,占 85.71%。检测 SNP 数量从 286 966 个到 599 255 个。[结论] 2,5,6 和 11 号染色体区域 SNP 多态性可能是神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤及肾母细胞瘤遗传易感区域,深入此易感区域研究可能为阐述肿瘤发生、治疗及预后评估提供一定线索。

主题词:全基因组关联研究(GWAS);文献计量学;神经母细胞瘤;胶质母细胞瘤;尤文肉瘤;肾母细胞瘤

中图分类号:R73 **文献标识码:**A **文章编号:**1671-170X(2014)10-0833-05

doi:10.11735/j.issn.1671-170X.2014.10.B011

Bibliometrics and Hot Research of Genome-wide Association Study in Neuroblastoma, Glioblastoma, Ewing Sarcoma and Wilms Tumor

LIN Cai-xia¹, XU Chun-wei²

(1. People's Hospital of Jiangshan City, Jiangshan 324100, China; 2. General Hospital of Beijing Military Command, Beijing 100700, China)

Abstract: [Purpose] To analyze the data constitution features of genome-wide association studies (GWAS) in English literatures of neuroblastoma, glioblastoma, Ewing sarcoma and Wilms tumor. [Methods] National Human Genome Institute GWAS database, one of the latest international research database of GWAS, was meticulously data mined. Eight characteristics of NHGRI literatures in neuroblastoma, glioblastoma, Ewing sarcoma and Wilms tumor from November 25, 2008 to August 26, 2013 were analyzed bibliometrically. These eight characteristics were the basic characteristic of literature, the sample sizes of first and second stages, chromosomal regions associated with single nucleotide polymorphism (SNP), related risk allele frequency in the control groups, P value with the strongest SNP in the literatures, odds ratio(OR) or beta coefficient of correlation, and experimental platform including testing the SNP number. [Results] Total 7 articles about 4 types of rare tumors from GWAS studies were found (3 for neuroblastoma, 1 for glioblastoma, 1 for Ewing sarcoma and 1 for Wilms tumor). These papers involved a total of 27 SNP(10 for neuroblastoma, 1 for glioblastoma, 3 for Ewing sarcoma, 13 for Wilms tumor). Most of the literatures came from *Nat Genet*. The first stage group sample sizes of these literatures were from 315 to 2101 cases in case groups and from 1879 to 3851 cases in control groups; the second stage group sample sizes were from 434 to 1488 cases in case groups and 199 to 3851 cases in control group. Chromosome regions associated SNP were mainly in number 2,5,6 and 11. Control group related risk allele frequency were commonly in 0.40~ group, and no rare mutation was found. P value in the strongest risk associated with SNP alleles was dominated by 1×10^{-10} ~, the smallest P value was 2×10^{-26} . OR or beta coefficient ranged from 1.23 to 2.64. Majority of research platform was Illumina (85.71%). Detective number of SNP was from 286 966 to 599 255. [Conclusion] Chromosome 2,5,6 and 11 regional SNP may be genetic susceptibility regions of neuroblastoma, glioblastoma, Ewing sarcoma and Wilms tumor, and further study on the susceptible areas can provide some clues in the treatment and prognosis assessment of cancer.

Subject words: genome-wide association study(GWAS); bibliometrics; neuroblastoma; glioblastoma; Ewing sarcoma; nephroblastoma

通讯作者: 许春伟,助理研究员,硕士;中国人民解放军北京军区总医院病理科,
北京市东城区南门仓 5 号(100700),E-mail:xuchunweibbb@163.com

收稿日期:2014-03-11;修回日期:2014-05-08

全基因组关联研究 (genome-wide association study, GWAS) 是利用高通量基因芯片技术, 对人类全基因组范围内常见遗传变异——单核苷酸多态性 (single nucleotide polymorphism, SNP) 和拷贝数变异 (copy number variation, CNV) 进行总体关联分析的研究方法^[1]。它是基于 DNA 是可遗传的且和邻近的等位基因从上一代传递给下一代这一理论。国际人类基因组测序完成, 单体型图谱工程的完成和经济高效的高通量基因分型技术的开展, 使得全基因组范围内筛选与疾病相关的序列变异成为可能, 它可以在病例和对照中比较全基因组范围内所有变异的等位基因频率, 从中发现与疾病相关联的序列变异^[2,3]。GWAS 研究设计类型有单个阶段研究、两个阶段研究和多阶段研究设计^[2], 其统计分析的基本原则是减少系统偏倚和加大研究强度, 譬如进行强有力地 SNP 设定分析^[4]。2007 年 4 月 Duggan 等^[5]和 Murbito 等^[6]利用 Illumina 及 Affymetrix 的基因芯片对前列腺癌进行了 GWAS, 其结果同时首次在 *Science* 杂志上发表, 自此以后越来越多肿瘤方面的 GWAS 研究结果陆续发表。肿瘤方面 GWAS 研究处于生物医学科技前沿, 所有原始研究文选均以第一时间在高质量的英文科技刊物发表。现通过对最新 GWAS 数据库之一——美国国立人类基因组研究所 (NHGRI) 数据库, 进行神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤和肾母细胞瘤文献特征数据分析挖掘, 所获得的少见肿瘤 GWAS 数据信息或许能对国内 GWAS 有参考作用。

1 资料与方法

1.1 资料

来源于 NHGRI 数据库 2008 年 11 月 25 日至 2013 年 8 月 26 日收录的神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤和肾母细胞瘤相关的文献。登入页面 <http://www.genome.gov/page.cfm?pageid=26525384#searchForm>。在页面下 Disease/Trait 一栏中输入“Neuroblastoma”, 点击 Search 后, 再点击“Download Spreadsheet of Search Result”保存“MyGWASSearch.xls”, 并改名为“Neuroblastoma.xls”, 并用同样的步骤获得“Glioblastoma.xls”、“Ewing sarcoma.xls”、“Wilms tumor.xls”。

1.2 方法

运用文献计量学方法分析文献基本特征、第一、二阶段样本量、染色体区域相关 SNP、对照组相关风险等位基因频率、P 值文献中最强 SNP、优势比(odds ratio, OR) 或 β 相关系数和实验平台及检测 SNP 数量 8 项数据, 获得少见肿瘤 GWAS 的共性特征。

2 结果

2.1 文献基本特征

数据库原记录包含 7 篇论文^[7-13], 其中神经母细胞瘤 3 篇, 多形性胶质母细胞瘤 1 篇, 尤文肉瘤 1 篇, 肾母细胞瘤 2 篇, 涉及 27 个 SNP, 其中神经母细胞瘤 10 个 SNP, 多形性胶质母细胞瘤 1 个 SNP, 尤文肉瘤 3 个 SNP, 肾母细胞瘤 13 个 SNP。论文分别发表在 *N Engl J Med* 1 篇, *Nature* 1 篇, *Nat Genet* 4 篇, *Clin Cancer Res* 1 篇。

2.2 第一(筛选)阶段样本量

在 7 篇少见肿瘤 GWAS 研究文献中第一阶段病例组样本量从 315 例到 2101 例不等; 对照组例数从 1879 例到 4947 例。14.29% 文献没有明确报告对照组样本量 (Table 1)。

Table 1 First-stage sample size distribution of rare cancer GWAS literatures

| Sample | Literature of case group (%) | Literature of control group (%) |
|--------|------------------------------|---------------------------------|
| 100~ | 3(42.86) | 0 |
| 1000~ | 4(57.14) | 6(85.71) |
| NR | 0 | 1(14.29) |
| Total | 7(100.00) | 7(100.00) |

Note: NR: not report.

2.3 第二(验证)阶段样本量

第二阶段病例组样本量从 434 例到 1488 例, 对照组例数从 199 例到 3851 例。14.29% 文献没有报告对照组样本量 (Table 2)。

Table 2 Second-stage sample size distribution of rare cancer GWAS literatures

| Sample | Literature of case group (%) | Literature of control group (%) |
|--------|------------------------------|---------------------------------|
| 100~ | 6(85.71) | 0 |
| 1000~ | 1(14.29) | 6(85.71) |
| NR | 0 | 1(14.29) |
| Total | 7(100.00) | 7(100.00) |

2.4 染色体区域相关SNP(NCBI 登记号)

所定位的染色体区域分布大致与染色体长度近似。其中2、5、6和11号染色体明显多于邻近的染色体(Table 3)。

Table 3 NCBI registration number in chromosome region SNP of rare cancer GWAS literatures

| Chromosome | Number | Ratio(%) |
|------------|--------|----------|
| 1 | 2 | 7.41 |
| 2 | 4 | 14.81 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 5 | 18.53 |
| 6 | 5 | 18.53 |
| 7 | 0 | 0 |
| 8 | 0 | 0 |
| 9 | 1 | 3.70 |
| 10 | 1 | 3.70 |
| 11 | 4 | 14.81 |
| 12 | 2 | 7.41 |
| 13 | 0 | 0 |
| 14 | 0 | 0 |
| 15 | 1 | 3.70 |
| 16 | 0 | 0 |
| 17 | 0 | 0 |
| 18 | 0 | 0 |
| 19 | 0 | 0 |
| 20 | 0 | 0 |
| 21 | 0 | 0 |
| 22 | 1 | 3.70 |
| X/Y | 1 | 3.70 |
| NR | 0 | 0 |
| Total | 27 | 100.00 |

2.5 对照组相关风险等位基因频率

对照组中与最强 SNP 相关风险等位基因频率在少见肿瘤 GWAS 文献中的分布,未发现稀有变异(即在对照人群中的发生频率<0.01) 的少见肿瘤 GWAS 文献(Table 4)。

2.6 相关风险等位基因 P 值文献中最强 SNP

相关风险等位基因 P 值选择 $<1\times10^{-6}$ 为录入标准,最小 P 值 2×10^{-26} (Table 5)。

2.7 OR 或 β 相关系数

文献中最强 SNP 相关风险等位基因优势比(或 β 相关系数)从 1.23 到 2.64(Table 6)。

2.8 实验平台及检测 SNP 数量

6 篇文献(85.71%) 的少见肿瘤 GWAS 研究独

Table 4 The strongest risks associated with SNP allele frequency distribution in control groups in rare cancer GWAS literatures

| Risk allele frequency | Number | Ratio(%) |
|-----------------------|--------|----------|
| 0.01~ | 4 | 14.81 |
| 0.10~ | 3 | 11.11 |
| 0.20~ | 3 | 11.11 |
| 0.30~ | 1 | 3.70 |
| 0.40~ | 6 | 22.24 |
| 0.50~ | 4 | 14.81 |
| 0.60~ | 3 | 11.11 |
| 0.70~ | 2 | 7.41 |
| 0.80~ | 1 | 3.70 |
| 0.90~ | 0 | 0 |
| NR | 0 | 0 |
| Total | 27 | 100.00 |

Table 5 Risks associated with allele P-value distribution

| OR group | Number | Ratio(%) |
|----------|--------|----------|
| (-30)~ | 0 | 0 |
| (-20)~ | 4 | 14.81 |
| (-10)~ | 17 | 62.97 |
| (-9)~ | 2 | 7.41 |
| (-8)~ | 2 | 7.41 |
| (-7)~ | 1 | 3.70 |
| (-6)~ | 1 | 3.70 |
| NS | 0 | 0 |
| Total | 27 | 100.00 |

Table 6 The strongest risks associated with SNP allele ratio distribution

| OR group | Number | Ratio(%) |
|----------|--------|----------|
| 0.1~ | 0 | 0 |
| 1.00~ | 23 | 85.19 |
| 2.00~ | 4 | 14.81 |
| NR | 0 | 0 |
| Total | 27 | 100.00 |

立使用了 Illumina 研究平台,1 篇研究平台未报道。检测 SNP 数量从 286 966 个到 599 255 个(Table 7)

Table 7 The number of SNP detection by using test platforms in rare tumor GWAS literatures

| Number of SNP detection | Literature | Ratio(%) |
|-------------------------|------------|----------|
| 20000~ | 2 | 28.57 |
| 30000~ | 1 | 14.29 |
| 40000~ | 3 | 42.86 |
| 50000~ | 1 | 14.29 |
| Total | 7 | 100.00 |

3 讨 论

目前少见肿瘤 GWAS 方面仅报道了神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤和肾母细胞瘤, 其中在神经母细胞瘤 GWAS 方面, Maris 等^[7]在北美人群中研究发现染色体 6p22.3 区域 *FLJ22536* 和 *FLJ44180* 基因中 rs6939340 是与神经母细胞瘤转移相关。随后 Wang 等^[8] 在欧洲人群中验证了 6p22.3 区域 *FLJ22536* 和 *FLJ44180* 基因中 rs6939340 与神经母细胞瘤的遗传易感性, 并发现 11p15.4 区域 *LMO1* 基因中 rs110419 与神经母细胞瘤疾病进展相关。Diskin 等^[9] 在证实 6p22.3 区域 *FLJ22536* 和 *FLJ44180* 基因中 rs6939340 与神经母细胞瘤的遗传易感性相关外, 在 6p16.3 上发现 *HACE1* 基因中 rs4336470 和 *LIN28B* 基因中 rs17065417 共同变异影响神经母细胞瘤易感性并在疾病的发展中发挥作用; 在多形性胶质母细胞瘤 GWAS 方面, Xiao 等^[10] 发现在 5q14 区域 *SSBP2* 基因中 rs7732320 与多形性胶质母细胞瘤预后相关; 在尤文肉瘤 GWAS 方面, Postel-Vinay 等^[11] 在欧洲法国人群中发现 1p36.2 区域 *TARDBP* 基因 rs9430161, 10q21.3 区域 *EGR2*、*ADO* 和 *ZNF365* 3 个基因中 rs224278 以及 15q15 区域 *EIF2AK4*、*SRP14* 和 *BMF3* 3 个基因中 rs4924410 与尤文肉瘤的遗传易感性相关; 在肾母细胞瘤 GWAS 方面 Turnbull 等^[12,13] 研究发现 2p24 区域 *DDX1* 和 *MYC* 基因两基因中 rs3755132 和 rs807624 以及 11q14.1 区域 *DLG2* 基因中 rs790356 与肾母细胞瘤的发生有关, 另外 5p15.3 区域 (*TERT* 基因)、5p15.3 区域 (*CLPTM1L* 基因)、9p24 区域 (*DMRT1* 基因)、12p13.1 区域 (*ATF7IP* 基因)、5q31.3 基因 (*SPRY4* 基因) 和 12q21.3 区域 (*KITLG* 基因) 中相对应的 rs2736100、rs4635969、rs755383、rs2900333、rs4624820 和 rs3782181 与睾丸生殖细胞肿瘤遗传易感性相关, 其中 rs755383 也与小鼠畸胎瘤遗传易感性相关。

通过对 NHGRI 数据库中 4 种少见肿瘤 8 项数据进行不完全统计后, 我们发现少见肿瘤有如下特点: ①在文献基本特征方面: 少见肿瘤在 *Nat Genet* 发表的文献最多, 占 57.14%, 并在影响因子位居第 2 的国际顶级期刊 *N Engl J Med* 发表 1 篇, 这也是目前为止肿瘤 GWAS 方面发表的影响因子最高的 GWAS。笔者推测 *Nat Genet* 以其影响力将是优质少

见肿瘤的 GWAS 的主要阵地。目前罕见胚胎性肿瘤肝母细胞瘤、胰母细胞瘤、视网膜母细胞瘤等罕见肿瘤仍然缺乏预防及预后的指标的数据。②少见肿瘤 GWAS 第一、二阶段的病例组样本量方面: 一般从几百例到几千例不等, 且以百例为主, 最大样本量均为超过 2500 例; 对照组样本量也从几百例到几千例不等, 但以千例以上为主, 最大样本量均为超过 5000 例。③染色体区域相关 SNP 方面: 以 2、5、6 和 11 号染色体区域 SNP 居多, 其中 5 号染色体区域 SNP 在脑肿瘤、头颈部肿瘤、皮肤肿瘤和血液系统肿瘤中也居多, 6 号染色体区域 SNP 在脑肿瘤、头颈部肿瘤、皮肤肿瘤和肝胆胰消化系统肿瘤中也居多。④对照组相关风险等位基因频率方面: 少见肿瘤 GWAS 中风险等位基因以风险等位基因频率 0.40 左右最多, 其次是 0.01 和 0.50, 但未发现稀有变异。⑤P 值文献中最强 SNP 方面: 罕少见肿瘤 P 值以 10^{-10} 为主, 其中最强 P 值 2×10^{-26} 。⑥OR 或 β 相关系数方面: 少见肿瘤 OR 主要集中在 1.00 左右的弱相关, 2.00 以上的强相关所占的比例很少。⑦实验平台及检测 SNP 数量方面: 少见肿瘤实验平台方面以 Illumina 研究平台占大多数, 未见 Affymetrix、Perlegen 或 Invader 研究平台的报道。检测 SNP 数量 286 966~599 255 个, 相比常见消化系统、血液系统、泌尿生殖系统、甲状腺、乳腺、皮肤和妇科肿瘤相对过少。

肿瘤方面 GWAS 的最终目标是应用于临床, 主要体现在对正常人群进行肿瘤的危险度预测评估和指导肿瘤的个体化治疗。个体化治疗推崇在肿瘤遗传谱指导下的个体化医学, 为肿瘤发生和治疗疗效提供早期预测, 促进肿瘤预防和治疗。少见肿瘤 GWAS 经过 6 年的发展, 已在神经母细胞瘤、多形性胶质母细胞瘤、尤文肉瘤和肾母细胞瘤遗传易感性研究方面取得了很大成功, 为个体化治疗打下了基础。本文通过运用文献计量学方法客观分析了 NHGRI 的文献基本特征、第一、二阶段样本量、染色体区域相关 SNP、对照组相关风险等位基因频率、P 值文献中最强 SNP、OR 或 β 相关系数, 实验平台及检测 SNP 数量 8 项数据, 找到了少见肿瘤 GWAS 的一些共性特征, 可为国内少见肿瘤 GWAS 作一定参考。

参 考 文 献:

- [1] Zhang XJ. Genetics implications of genome-wide association studies for psoriasis [J]. Journal of Zhejiang University

- (Medical Sciences), 2009, 38 (4):333–337.[张学军. 全基因组关联分析对银屑病遗传学研究的启示 [J]. 浙江大学学报(医学版), 2009, 38(4):333–337.]
- [2] Yan WL. Advances in genome-wide association of complex diseases—study design and genetic markers [J]. Genetic, 2008, 30(4):400–406.[严卫丽. 复杂疾病全基因组关联研究进展——研究设计和遗传标记 [J]. 遗传, 2008, 30(4):400–406.]
- [3] Yan WL. Advances in genome-wide association of complex diseases—statistical analysis of the genetic [J]. Genetic, 2008, 30(5):543–549.[严卫丽. 复杂疾病全基因组关联研究进展——遗传统计分析[J]. 遗传, 2008, 30(5): 543–549.]
- [4] Kraft P, Cox DG. Study designs for genome-wide association studies[J]. Adv Genet, 2008, 60(17):465–504.
- [5] Duggan D, Zheng SL, Knowlton M, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP[J]. J Natl Cancer Inst, 2007, 99(24):1836–1844.
- [6] Murabito JM, Rosenberg CL, Finger D, et al. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study [J]. BMC Med Genet, 2007, 8(Suppl 1):S6.
- [7] Maris JM, Mosse YP, Bradfield JP, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma[J]. N Engl J Med, 2008, 358(24):2585–2593.
- [8] Wang K, Diskin SJ, Zhang H, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene[J]. Nature, 2011, 469(7329):216–220.
- [9] Diskin SJ, Capasso M, Schnepf RW, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma [J]. Nat Genet, 2012, 44(10): 1126–1130.
- [10] Xiao Y, Decker PA, Rice T, et al. SSBP2 variants are associated with survival in glioblastoma patients [J]. Clin Cancer Res, 2012, 18(11):3154–3162.
- [11] Postel-Vinay S, Veron AS, Tirode F, et al. Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma[J]. Nat Genet, 2012, 44(3):323–327.
- [12] Turnbull C, Perdeaux ER, Pernet D, et al. A genome-wide association study identifies susceptibility loci for Wilms tumor[J]. Nat Genet, 2012, 44(6):681–684.
- [13] Turnbull C, Rapley EA, Seal S, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer[J]. Nat Genet, 2010, 42(7):604–607.

关于启用稿件远程处理系统的通知

本刊已启用稿件远程处理系统,该系统包括作者在线投稿/查询、主编办公、专家审稿、编辑办公等功能,通过网上投稿、网上查稿、网上审稿,实现作者、编辑、审稿专家的一体化在线协作处理,从而构建一个协作化、网络化、角色化的编辑稿件业务处理平台。对于广大作者而言,该系统最大的优点是支持在线投稿,方便作者及时了解稿件处理进程,缩短稿件处理时滞。使用过程中具体注意事项如下:

(1)第1次使用本系统投稿的作者,必须在“作者登录”中先注册,才能投稿。注册时各项信息请填写完整。作者自己设定用户名和密码,该用户名密码长期有效。

(2)已注册过的作者,请不要重复注册,否则将导致查询稿件信息不完整。如果遗忘密码,可以致电编辑部查询。

(3)作者投稿请点击“作者登录”,登录后按照提示操作即可。投稿成功后,系统自动发送回执邮件,作者投稿后请随时关注邮箱提示,也可随时点击“作者登录”,获知该稿件的审理情况、处理进展、审稿意见等。

稿件远程处理系统启用后,我刊只接受网上投稿,不再接收电子邮件投稿和纸质稿,《肿瘤学杂志》网址:<http://www.chinaoncology.cn>

如有任何问题,请与编辑部联系! 联系电话:0571-88122280。

《肿瘤学杂志》编辑部