

暴露组与暴露组学

任爱国

【导读】 人类复杂疾病主要是由环境因素或者环境与遗传因素相互作用所致。自人类基因组计划开展以来,有关遗传因素与复杂疾病发病关系的研究进展非常迅速,相比之下,环境暴露与复杂疾病关系的研究却没有得到应有的重视和发展。暴露组与暴露组学的提出,为推动环境因素与人类健康之间关系的研究提供了新思路。

【关键词】 暴露组;暴露组学;生物标志物;全暴露组关联研究

On exposome and exposomics REN Ai-guo, Institute of Reproductive and Child Health/Ministry of Health Key Laboratory of Reproductive Health, Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

Corresponding author: REN Ai-guo, Email: renag@bjmu.edu.cn

【Key words】 Exposome; Exposomics; Biomarkers; Exposome-wide association study

流行病学研究的主要任务之一是分析暴露与疾病(或其他健康问题)之间的关系。疾病的发生可能是环境因素、遗传因素或者是环境和遗传因素相互作用的结果^[1]。一般认为,慢性非传染性疾病(复杂疾病)主要由环境因素导致^[2],或者是由遗传与环境因素交互作用所导致。随着人类基因组(genome)计划及人类基因组单倍体型组计划的完成,对复杂疾病遗传因素的研究进展迅速。相比之下,对环境因素的研究进展缓慢。为了使复杂疾病遗传和环境因素两方面的研究得到平衡发展,有学者提出了暴露组(exposome)和暴露组学(exposomics)的概念^[3-5]。

1. 环境因素在复杂疾病发病中的作用:疾病的分布和干预试验表明,环境因素在癌症、心血管疾病、先天性缺陷等复杂疾病发生中起主要作用。

日本人曾是胃癌高发人群,但日本人移居至美国后胃癌死亡率有所下降,而在美国出生的日本后裔其胃癌死亡率下降更为明显^[6]。表明环境因素在胃癌发病中起主要作用。对瑞典、丹麦和芬兰共计44 788对双生子28种癌症发病情况调查发现,遗传因素对前列腺癌的发病贡献最大(42%),其次为结肠直肠癌(35%)、乳腺癌(27%);而遗传因素对宫颈癌和子宫体癌的发病贡献为0,胃癌为28%,肺癌为26%,卵巢癌为22%,白血病为21%^[7]。提示遗传因素在大多数癌症发病中不起主要作用。

有些疾病在同一种族(民族)不同地区间发病率(或死亡率)差别很大。如我国1986—1987年调查发现,山西省胎儿神经管畸形出生患病率为10.5%,而宁夏回族自治区只有2.5%^[8]。相邻地区间4倍的差异很难用遗传因素的作用来解释。对1972—1989年上海市区居民癌症发病率监测发现,18年间食管癌的发病率男女性均下降了50%以上^[9]。表明食管癌的发病因素主要是环境因素,如饮食的改善等。

一般认为,遗传因素在先天性疾病发病中的作用可能更大。但对于胎儿神经管缺陷(NTDs),如母亲孕前至孕后1个月每日服用0.4 mg叶酸,可降低80%以上胎儿的NTDs风险^[10,11]。曾经生育过NTDs患儿的妇女,从孕前开始每日服用4 mg叶酸,可降低72%再次生育NTDs患儿的风险,如全程服药,则可降低83%的风险^[12]。表明叶酸缺乏是胎儿NTDs发生的主要原因。

有学者对2010年英国18种癌症归因于14种生活方式和环境因素的分值进行估计,其中宫颈癌(100.0%)、喉癌(92.5%)、口腔癌(90.6%)、肺癌(89.2%)、食管癌(89.0%)、胃癌(74.9%)等常见癌症的分值均较高^[13]。表明生活方式和环境因素在大多数癌症发病中起主要作用。

2. 遗传变异增加复杂疾病发生的易感性:同一人群相似的生活环境,却只有少数人发生某种疾病。当然,即使生活在相似的环境中,个体间的暴露水平或者暴露“窗口”可能还是有差异。不过一般认

DOI: 10.3760/cma.j.issn.0254-6450.2012.09.022

作者单位:100191 北京大学学生健康研究所 卫生部学生健康重点实验室/北京大学公共卫生学院流行病学与卫生统计学系

通信作者:任爱国, Email: renag@bjmu.edu.cn

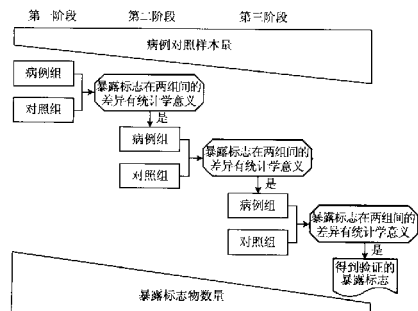
为,这是因为某些个体更易发生这种疾病,也就是具有易感性。疾病易感性主要由遗传因素决定,即遗传易感性(genetic susceptibility)。易感性也受非遗传因素影响,如营养不良会使机体抵抗力下降,从而对某些疾病的发病风险增加;或者某些环境暴露物质进入机体后,对内分泌系统或其他生理或生化过程产生影响,从而更容易发生某些疾病^[4]。在人类基因组计划出现之前,流行病学对遗传易感性的研究主要通过家族史、连锁分析等传统手段。随着人类基因组计划及单体型图计划的完成,通过基因多态性(SNPs)对遗传易感性的研究成为流行病学研究的热点,其中以全基因组关联研究(GWAS)为代表。

3. 暴露组和暴露组学:与飞速发展的基因组或相关技术相比,对环境暴露的研究进展缓慢,尽管环境因素在复杂疾病发病中起主要作用。Wild^[5]形象地将这种不平衡现象比喻为雄性大西洋泥沼潮蟹(*Uca pugnax*) (该蟹一只螯巨大,重量可达整体的65%;而另一只螯极小,欠发育)。为此Wild于2005年提出了暴露组的概念,作为对基因组的补充。按此定义,暴露组是从受精卵开始,系终生的环境暴露,包括其生活方式^[5]。随后,Rappaport和Smith^[15]对暴露组的概念进行了发展,认为暴露组涵盖所有来源、进入人体内环境暴露的全部。这些暴露不仅包括来自于空气、水、食物等外环境的化学物质,也包括机体因炎症、氧化应激、脂质过氧化、感染、肠道菌群等产生的化学物质。美国疾病预防控制中心将暴露组学定义为一个人一生中的所有暴露以及这些暴露如何与疾病发生联系^[16]。随着暴露组概念的提出,相应地出现了暴露组学的概念。Rappaport^[17]将暴露组学定义为研究暴露组以及暴露组对人类疾病过程影响的学问。

4. 暴露组学的研究思路:环境暴露与健康之间的关系研究可从测定病例组和对照组外环境中的空气、水、食物内的污染物入手,检验两组间是否存在差异,进而通过对主要物质的摄入量、代谢等进行考量,估计暴露剂量,这称之为自上而下的方法(bottom-up)。另一种方法是从测量病例组和对照组血液或其他体液内的物质入手,检验各类物质与疾病之间的统计学联系,最终确定导致疾病的物质及其暴露来源,称之为自上而下的方法(top-town)^{5,15]}。Rappaport^[17]认为后种方法优于前种。

与传统意义上的流行病学或暴露科学研究不同,暴露组学研究多是无目标设计(untargeted design)^[17]。这与GWAS类似。研究中,并不针对单一或少数几种暴露物质,而是检测所有可能的暴露

标志,通过统计学分析发现病例组和对照组差异最大的几种暴露标志,再在更大或多个独立样本中重复测定,最终获得验证后的暴露标志(图1)。当然,对得到验证的暴露标志物,可进一步采用动物实验或其他研究方法验证或进行致病机制研究。与GWAS相对应,这类研究可称为全暴露组关联研究(exposome-wide association study, EWAS)。可见GWAS和EWAS都是数据驱动(data-driven),无研究假设(hypothesis free)或无明确研究假设^[18],是从众多的比较中发现暴露与疾病之间的统计学联系,进而产生病因假设。



注:第一阶段为发现阶段,对于第一阶段统计学显著性最高的几个暴露标志物,在重复阶段用另外一个或多个独立的病例对照研究进行检验,得到验证的暴露标志物即认为与疾病的发病有关

图1 全暴露组关联研究

暴露组的研究结果还可用来确定具体的暴露,开发用于高通量筛查的特异标志物,或者用于确定外源性或内源性暴露的来源^[17,19]。

美国疾病预防控制中心每2年进行一次全国健康与营养调查(National Health and Nutrition Examination Survey, NHANES)。对采集的血液标本测定各种临床(如空腹血糖)或环境暴露标志物,每次测定的标志物类别和数量不尽相同。Patel等^[20]利用1999—2000年(111个环境暴露标志物)、2001—2002年(146个环境暴露标志物)、2003—2004年(211个环境暴露标志物)、2005—2006年(75个环境暴露标志物)4次调查数据进行2型糖尿病全环境关联分析(environment-wide association study)。分析时控制假发现率(false discovery rate, FDR),并调整年龄、性别、BMI、种族以及社会经济状况。首先在各次调查样本内分析环境暴露标志物与2型糖尿病发病风险之间的关系。结果分别发现4~23种,共计37种暴露标志与2型糖尿病风险之

间存在统计学联系;然后对这37种暴露标志再次在4次调查样本中进行“验证”,如果某种标志在2个或以上调查样本中均存在统计学关联,则认为该标志物得到了验证。最终发现环氧七氯(heptachlor epoxide)、维生素E(γ -tocopherol)和多氯联苯-170为2型糖尿病的危险因素,而 β -胡萝卜素为2型糖尿病的保护因素。虽然这4次均为横断面调查,并无法得到这些危险因素与2型糖尿病的因果关系,但可作为全环境或全暴露组关联研究的雏形。

5. 提出暴露组学的必要性:如前所述,人类复杂疾病主要由环境因素,或者由遗传与环境因素相互作用而导致。但对遗传因素和环境因素的研究处于极度不平衡状态,这不利于开发针对疾病环境病因的理论和方法研究,亦阻碍了对疾病的预防和治疗。

以往已有很多利用暴露标志物进行的疾病病因研究,但往往只针对单一化学物质或少数几类化学物质。这一点与针对单个或少数几个基因多态性位点的遗传学研究非常相似。这种研究不但效率低,更重要的是不能了解暴露的全貌,无法分析众多暴露标志物之间的复杂关系。暴露组概念的提出,为研究疾病的环境病因提供了一种新的思维方式和研究思路。

暴露组研究可以补充基因组学在人类复杂疾病病因研究中的不足。基因组学研究尽管投入巨大,但产出并未像最初期望的那样,很快就可以确定众多复杂疾病的遗传易感性,并据此设计出高危人群的筛查方法,以及针对个体遗传背景制定的治疗或保健方案^[21-23]。以人体身高研究为例。通常认为,身高的遗传度有80%。但通过对数千人的研究所发现的40多个基因多态性位点中能解释身高变异只有5%左右^[24]。由于围受孕期妇女增补叶酸可预防胎儿NTDs,因此叶酸代谢相关基因多态性与胎儿NTDs风险之间关系的研究是基因变异与疾病遗传易感性研究的热点之一。其中又以叶酸代谢酶——5, 10-亚甲基四氢叶酸还原酶基因(MTHFR)677C>T多态性的研究最多。但对32项相关研究进行Meta分析发现,MTHFR 677TT基因型使胎儿NTDs风险增加主要表现在非拉丁裔欧洲人(风险增加62%),而在拉丁裔欧洲人中,TT基因型不增加NTDs风险。即使在非拉丁裔欧洲人,TT基因型能够解释的人群NTDs百分比只有5.7%^[25]。在近期一项研究中,对239名脊柱裂病例和231名健康对照叶酸代谢通路上31条基因共402个外显子进行测序,虽然发现了1441个多态性位点,但如根据多重检验调整P值,无任何一个多态性位点与脊柱裂的发病

风险存在统计学联系^[26]。如果以受试者工作特征曲线(ROC)下面积(AUC)作为风险预测指标,根据GWAS结果对18种复杂疾病计算的预测能力与根据传统因素得到的预测能力非常相似^[27]。虽然已有种原因来解释GWAS“丢失”的遗传度(missing heritability),并提出改进途径^[22],将基因与环境暴露结合分析,开展全基因组全环境交互作用研究(gene-environment-wide interaction study, GEWIS)^[28]或全基因组全环境关联分析(gene-environment wide association study, GEWAS)^[29],以更全面揭示疾病的病因和发病机制,有助于高危个体的筛查,提高对疾病的预防和治疗能力^[25]。

6. 暴露组学与其他组学间的关系:暴露组学与现有的多种组学具有密切的关系。首先,暴露组学借用基因组学研究思路 and 统计方法。另外,暴露物在体内的吸收和代谢,受基因变异的影响,GWAS与EWAS相结合,有助于发现疾病的遗传和环境病因,进而提高对个体发病风险的预测能力。环境暴露物进入体内与生物大分子共价结合,形成加合物(adduct),因而,暴露组学与加合物组学密切相关。环境暴露物可以影响基因的表达,因此同时分析暴露组和转录组有助于揭示环境暴露物的致病机制。如果某一环境暴露物影响基因的表达,则可能进一步影响蛋白质的翻译,因而在蛋白质层面留下印迹。环境暴露物进入体内后,大多需要经过代谢,因而也会在代谢组内有所反映。

7. 暴露组学面临的挑战:由于人类基因从受精卵开始就已确定且一生固定不变(体细胞突变除外),因此,针对基因型与疾病发病关系的研究可在疾病的任何阶段进行。但暴露组是动态的,人体对环境的暴露会随时间,甚至是机体生理生化状态的变化而变化。因此,暴露组学研究的关键在于能够准确测定各种暴露以及暴露的健康效应^[15]。尤其重要的是,对暴露内部标志物的测定要与疾病发生的“时间窗”相对应,否则,研究结果将可能因为暴露和疾病发生时间上的先后关系不清而无法得出可靠结论。例如,人类的神经管在受精后第25至28天之间闭合^[30],要研究影响神经管闭合的环境暴露,就需要采集紧邻该时间段的血液或其他生物标本,测定其中环境暴露标志。由此可见,暴露组学研究宜采用队列研究(或巢式病例对照研究)设计,尤其在生命过程的多个关键阶段采集生物标本的队列研究设计。

由于暴露组研究依赖于暴露的内部标志或生物学有效剂量标志,也就面临许多技术困难^[31]:①某些

暴露尚未发现内部标志物,如噪声、热应激、电磁场暴露;②有些暴露缺乏特异的标志物,如复杂混合暴露(复杂混合物)、环境烟草暴露、颗粒性空气污染物、光化学烟雾等;③有些标志物具有双重身份,既可作为内暴露标志,又是对暴露的早期生理反应标志;④个体的暴露变异很大且随时间(年龄)变化而不同;⑤对一些暴露标志的测定可能在困难,因为有些化学物质很快被排除体外,而在体内停留时间非常短暂;⑥某些暴露标志尚无灵敏而精确的测定方法。

GWAS具有设备依赖性,而EWAS除依赖设备之外,还具有技术依赖性。目前,对体内暴露标志分析多使用基于质谱的分析仪器(如GC-MS、GC-TOF-MS、GCxGC-TOF-MS、HPLC-MS、UPLC-MS、CE-MS等);虽然很多机构拥有这些设备,但并不是每个机构都具备分析众多化学标志物的能力。

最后,暴露组学是否能够像其他组学那样发展成为一种真正的“组学”,还有待开发出有别于其他组学特定的实验室方法和数据分析方法。

8. 最新进展:美国科学院已组织了两次有关暴露组的学术研讨会。2010年2月25—26日召开了名为“The Exposure: A Powerful Approach for Evaluating Environmental Exposures and Their Influences on Human Diseases”的学术研讨会;2011年12月8—9日召开了名为“Emerging Technologies for Measuring Individual Exposomes”的学术研讨会^[19],大大推进了暴露组学的发展。

为了旨在最终发起全环境关联研究或全暴露组关联研究以及人类暴露组计划(Human Exposome Project, HEP),由美国加州大学 Berkeley 分校(UC Berkeley)、英国伦敦帝国学院(Imperial and Kings Colleges London)、国际癌症研究机构(International Agency for Research on Cancer, IARC)以及美国国家环境健康科学研究所(National Institute of Environmental Health Sciences, NIEHS)等单位共同发起成立了暴露组联盟(Exposome Alliance Project, <http://exposomealliance.org>)。人类暴露组计划的目的是阐明居住在不同环境下、具有各种疾病危险因素的多种族个体的暴露组。如此以来,就能像人类基因组单倍体计划那样,清晰地了解人类暴露的多样性和类型。虽然目前高通量的技术尚不成熟,但如人类基因组计划一样,只要方向明确,相信不久设想即可变为现实。

参 考 文 献

- Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*, 2005, 6: 287-298.
- Willett WC. Balancing life-style and genomics research for disease prevention. *Science*, 2002, 296: 695-698.
- Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*, 2005, 14: 1847-1850.
- Adler T, Sawyer K, Shelton-Davenport M. The exposome: a powerful approach for evaluating environmental exposures and their influences on human disease. <http://nas-sites.org/emergingscience/files/2011/05/03-exposome-newsletter-508.pdf>.
- Rappaport SM. Implications of the exposome for exposure science. *J Expo Sci Environ Epidemiol*, 2011, 21: 5-9.
- Demon JE. Cancer epidemiology in populations of the United States—with emphasis on Hawaii and California—and Japan. *Cancer Res*, 1975, 35: 3240-3245.
- 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: 78-85.
- Chinese Birth Defects Monitoring Collaborative Group. Atlas of birth defects in China. Chengdu Map Publishing House, 1992. (in Chinese)
- 中国出生缺陷监测协作组. 中国出生缺陷地图集. 成都: 成都地图出版社, 1992.
- Zheng W, Jin F, Devesa SS, et al. Declining incidence is greater for esophageal than gastric cancer in Shanghai, People's Republic of China. *Br J Cancer*, 1993, 68: 978-982.
- Czeizel AE, Dudas I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med*, 1992, 327: 1832-1835.
- Berry RJ, Li Z, Erickson JD, et al. Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. *N Engl J Med*, 1999, 341: 1485-1490.
- MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet*, 1991, 338: 131-137.
- Parkin DM, Boyd L, Walker LC. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br J Cancer*, 2011, 105: S77-81.
- Schug TT, Janesick A, Blumberg B, et al. Endocrine disrupting chemicals and disease susceptibility. *J Steroid Biochem Mol Biol*, 2011, 127: 204-215.
- Rappaport SM, Smith MT. Environment and Disease Risks. *Science*, 2010, 330: 460-461.
- Center for Disease Control and Prevention. Exposome and exposomics. <http://www.cdc.gov/niosh/topics/exposome/>.
- Rappaport SM. Discovering environmental causes of disease. *J Epidemiol Community Health*, 2012, 66: 99-102.
- Kitsios GD, Zintzaras E. Genome-wide association studies: hypothesis-free or “engaged”? *Transl Res*, 2009, 154: 161-164.
- Lioy PJ, Rappaport SM. Exposure science and the exposome: an opportunity for coherence in the environmental health sciences. *Environ Health Perspect*, 2011, 119: A466-467.
- Patel CJ, Bhattacharya J, Butte AJ. An environment-wide association study (EWAS) on type 2 diabetes mellitus. *PLoS One*, 2010, 5: e10746.
- Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA*, 2008, 299: 1335-1344.
- Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*, 2009, 461: 747-753.
- Bookman EB, McAllister K, Gillanders E, et al. Gene-environment interplay in common complex diseases: forging an integrative model—recommendations from an NIH workshop. *Genet Epidemiol*, 2011, 35: 217-225.
- Visscher PM. Sizing up human height variation. *Nat Genet*, 2008, 40: 489-490.
- Amorim MR, Lima MAC, Castilla EE, et al. Non-Latin European descent could be a requirement for association of NTR1 and MTHFR variant 677C>T: a meta-analysis. *Am J Med Genet Part A*, 2007, 143A: 1726-1732.
- Marini NJ, Hoffmann TJ, Lammert EJ, et al. A genetic signature of spina bifida risk from pathway-informed comprehensive gene-variant analysis. *PLoS One*, 2011, 6: e28408.
- Justins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet*, 2011, 20: R182-R188.
- Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies—challenges and opportunities. *Am J Epidemiol*, 2009, 169: 227-230.
- Thomas D. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*, 2010, 11(4): 259-272.
- Sadler TW. Embryology of neural tube development. *Am J Med Genet C Semin Med Genet*, 2005, 135C: 2-8.
- Peters A, Hoek G, Katsouyanni K. Understanding the link between environmental exposures and health: does the exposome promise too much? *J Epidemiol Community Health*, 2012, 66: 103-105. (收稿日期: 2012-05-23)

(本文编辑: 张林东)