中南大学学报(医学版) 2011, 36(5) 457- DOI: 10.3969/j.issn.1672-7347.2011. ISSN: 1672-7347 CN: 43-1427/R

综述

DNA混合分析技术的单体型频率估计方法

何柳， 唐迅， 胡永华

北京大学医学部公共卫生学院流行病与卫生统计学系，流行病学教育部重点实验室，北京 100191

摘要：

DNA混合分析技术的广泛使用，能够明显降低研究的工作量和成本，然而仍然有其自身局限，例如损失了大量个体遗传信息、测量误差较大等。为了尽量克服DNA混合分析技术的自身问题，其检测和分析方法在不断发展与完善。在单体型频率估计方面，基于最大期望（expectation-maximization，EM）算法的新方法不断涌现，如HaploPool算法和PoooL算法，其准确性、实用性均增强。

关键词： DNA混合分析技术 单体型 遗传流行病学

Estimation of haplotypes based on DNA pooling

HE Liu， TANG Xun， HU Yonghua

Department of Epidemiology & Biostatistics, Key Laboratory of Epidemiolog of Ministry of Education, Health Science Center, Peking University, Beijing 100191, China

Abstract:

DNA pooling, a fast and economic study strategy, is widely used in areas of scientific research. In spite of various limits, researchers are making their efforts to improve DNA pooling toward a more perfect direction, including allele frequency detection and estimation of haplotypes. In haplotype estimation, more and more analyzing methods originated from the expectation-maximization algorithm have appeared, with improved accuracy and practicality, such as HaploPool algorithm and PoooL algorithm.

Keywords: DNA pooling haplotypes genetic epidemiology

收稿日期 2010-04-22 修回日期 网络版发布日期

DOI: 10.3969/j.issn.1672-7347.2011.

基金项目:

国家自然科学基金(30671807，30872173)。

通讯作者: 胡永华,E-mail:yhhu@bjmu.edu.cn

作者简介: 何柳，博士研究生，主要从事遗传流行病学方面的研究。

作者Email: yhhu@bjmu.edu.cn

参考文献：

［1］Arnheim N C. Strange H E. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease:Studies of HLA class II loci［J］. Proceedings Nat Acad Sci, 1985, 82(20):6970-6974.

［2］Bharti A R, Letendre S L, Patra K P, et al. Malaria diagnosis by a polymerase chain reaction-based assay using a pooling strategy［J］. Am J Trop Med Hyg, 2009, 81(5):754-757.

［3］Bugeja M J, Booth D, Bennetts B, et al. An investigation of polymorphisms in the 17q11.2-12 CC chemokine gene cluster for association with multiple sclerosis in Australians［J］. BMC Med Genet, 2006, 7: 64.

［4］Matkovich S J, Van Booven D J, Hindes A, et al. Cardiac signaling genes exhibit unexpected sequence diversity in sporadic cardiomyopathy, revealing HSPB7 polymorphisms associated with disease［J］. J Clin Invest, 2010, 120(1): 280-289.

［5］Baro J A, Carleos C, Corral N, et al. Power analysis of QTL detection in half-sib families using selective DNA pooling［J］. Genet Sel Evol, 2001, 33(3):231-247.

［6］Sham P, Bader J S, Craig I, et al. DNA Pooling: a tool for large-scale association studies［J］. Nat Rev Genet, 2002,3(11):862-871.

DNA混合分析技术的单体型频率估计方法 何柳，等封三［7］Pearson J V, Huentelman M J, Halperin R F, et al. Identification of the genetic basis for complex disorders by use of pooling-based genomewide single-nucleotide-polymorphism association studies［J］. Am J Hum Genet, 2007,80(1):126-139.
［8］De Bakker P I, Graham R R, Altshuler D, et al. Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations［J］. Pac Symp Biocomput, 2006,11:478-486.
［9］International HapMap Consortium. The International HapMap Project［J］. Nature, 2003,426 (6968):789-796.
［10］Hoh J, Matsuda F, Peng X, et al. SNP haplotype tagging from DNA pools of two individuals［J］. BMC Bioinformatics, 2003, 4: 14.
［11］Ito T, Chiku S, Inoue E, et al. Estimation of haplotype frequencies, linkage-disequilibrium measures, and combination of haplotype copies in each pool by use of pooled DNA data［J］. Am J Hum Genet, 2003,72(2):384-398.
［12］Yang Y, Zhang J, Hoh J, et al. Efficiency of single-nucleotide polymorphism haplotype estimation from pooled DNA［J］. Proc Natl Acad Sci U S A, 2003,100(12):7225-7230.
［13］Barrett J C, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps［J］. Bioinformatics, 2005,21(2):263-265.
［14］Kirkpatrick B, Armendariz C S, Karp R M, et al. HaploPool: improving haplotype frequency estimation through DNA pools and phylogenetic modeling［J］. Bioinformatics, 2007,23(22):3048-3055.
［15］Philip E B. RECOMB’04: Proceedings of the Eighth annual international Conference on Resaerch in Computational Molecular Biology［M］. New York,USA:ACM Press,2004：10-19.
［16］Stephens M, Smith N J, Donnelly P. A new statistical method for haplotype reconstruction from population data［J］. Am J Hum Genet, 2001,68(4):978-989.
［17］Zhang H, Yang H C, Yang Y. PoooL: an efficient method for estimating haplotype frequencies from large DNA pools［J］. Bioinformatics, 2008,24(17):1942-1948.
［18］Zhang H, Zhang H, Li Z, et al. Statistical methods for haplotype-based matched case-control association studies［J］. Genet Epidemiol, 2007,31(4):316-326.
［19］Valle T, Tuomilehto J, Bergman R N, et al. Mapping genes for NIDDM. Design of the Finland-United States investigation of NIDDM Genetics (FUSION) study［J］. Diabetes Care, 1998,21(6):949-958.
［20］Lin D Y, Zeng D. Likelihood-based inference on haplotype effects in genetic association studies［J］. J Am Stat Assoc, 2006,101(473): 89-104.
［21］Yang Y, Hoh J, Xu F, et al. Efficiency of SNP haplotype estimation from pooled DNA［J］. Proceedings Nat Acad Sci, 2003,100(12):7225-7230.
［22］Jain S, Tang X, Narayanan C S, et al. Angiotensinogen gene polymorphism at -217 affects basal promoter activity and is associated with hypertension in African-Americans［J］. J Biol Chem, 2002, 277 (39): 36889-36896.
［23］Yang H C, Pan C C, Lin C Y, et al. PDA: Pooled DNA analyzer［J］. BMC Bioinformatics, 2006,7:233.
［24］Niu T, Qin Z S, Xu X, et al. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms［J］. Am J Hum Genet, 2002,70(1):157-169.
［25］Zuo Y, Zou G, Zhao H. Two-stage designs in case-control association analysis［J］. Genetics, 2006, 173(3):1747-1760.

本刊中的类似文章