OXFORD

Genome analysis

EWAS: epigenome-wide association study software 2.0

Jing Xu[1,2,†], **Linna Zhao**[1,2,†], **Di Liu**[1,2,†], **Simeng Hu**[1,2,†], **Xiuling Song**[1,2], **Jin Li**[1], **Hongchao Lv**[1], **Lian Duan**[1], **Mingming Zhang**[1], **Qinghua Jiang**[3], **Guiyou Liu**[3], **Shuilin Jin**[4], **Mingzhi Liao**[5], **Meng Zhang**[2,6], **Rennan Feng**[2,6], **Fanwu Kong**[7,*], **Liangde Xu**[1,2,*] and **Yongshuai Jiang**[1,2,*]

[1]College of Bioinformatics Science and Technology, [2]Training Center for Students Innovation and Entrepreneurship Education, Harbin Medical University, Harbin 150086, China, [3]Center for Bioinformatics, School of Life Science and Technology, [4]Department of Mathematics, Harbin Institute of Technology, Harbin 150001, China, [5]College of Life Science, Northwest A&F University, Yangling 712100, Shaanxi, China, [6]Department of Nutrition and Food Hygiene, Public Health College and [7]Department of Nephrology, The Second Affiliated Hospital, Harbin Medical University, Harbin 150086, China

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

Abstract

Motivation: With the development of biotechnology, DNA methylation data showed exponential growth. Epigenome-wide association study (EWAS) provide a systematic approach to uncovering epigenetic variants underlying common diseases/phenotypes. But the EWAS software has lagged behind compared with genome-wide association study (GWAS). To meet the requirements of users, we developed a convenient and useful software, EWAS2.0.

Results: EWAS2.0 can analyze EWAS data and identify the association between epigenetic variations and disease/phenotype. On the basis of EWAS1.0, we have added more distinctive features. EWAS2.0 software was developed based on our 'population epigenetic framework' and can perform: (i) epigenome-wide single marker association study; (ii) epigenome-wide methylation haplotype (meplotype) association study and (iii) epigenome-wide association meta-analysis. Users can use EWAS2.0 to execute chi-square test, t-test, linear regression analysis, logistic regression analysis, identify the association between epi-alleles, identify the methylation disequilibrium (MD) blocks, calculate the MD coefficient, the frequency of meplotype and Pearson's correlation coefficients and carry out meta-analysis and so on. Finally, we expect EWAS2.0 to become a popular software and be widely used in epigenome-wide associated studies in the future.

Availability and implementation: The EWAS software is freely available at http://www.ewas.org.cn or http://www.bioapp.org/ewas.

Contact: jiangyongshuai@gmail.com or xuliangde@hrbmu.edu.cn or kidney1979@163.com

1 Introduction

Epigenome-wide association study (EWAS) is an effective tool to identify the association between epigenetic variation and common disease/phenotype (Rakyan *et al.*, 2011; Wahl *et al.*, 2017).

Compared with genome-wide association study (GWAS), the analysis tools of EWAS have lagged behind. To fill this gap, we developed novel and unique features, and improved upon the previous version EWAS1.0 (Xu *et al.*, 2016).

EWAS1.0 was originally designed only for identifying the association between combinations of methylation levels (beta-value) and diseases. EWAS2.0 (http://www.ewas.org.cn) is a fully functional software.

2 Features

EWAS2.0 software can perform: (i) epigenome-wide single marker association study; (ii) epigenome-wide methylation haplotype (meplotype) association study and (iii) epigenome-wide association meta-analysis. The methylation data should be cleaned and normalized.

For each DNA methylation loci, EWAS2.0 can carry out t-test or logistic regression analysis to identify the significant associations with case/control or binomial phenotype, perform linear regression analysis to identify the significant results associated with continuous phenotype, and calculate the Pearson's correlation coefficients between beta-value and continuous phenotype.

According to our 'population epigenetic framework' (Zhao *et al.*, 2016), EWAS2.0 can analyze the methylation genotypes (menotypes: MM, MU and UU, where M is methylation epi-allele and U is unmethylation epi-allele) data, calculate the epi-allele frequency and identify risk epi-allele (calculate Chi-square, *P*-value, odd ratio and 95% confidential interval). EWAS2.0 can also analyze the association between two epi-alleles (M and U) in the same locus, and label the type of the relationships: synergic (two members of homologous chromosomes tend to be methylated simultaneously) or exclusive (one member of homologous chromosomes is methylated, the other member of homologous chromosomes tends to be unmethylated) (Zhao *et al.*, 2016).

For multiple DNA methylation loci that are physically close to each other, there are non-random associations of epi-alleles between these loci, which we call methylation disequilibrium (MD) (Zhao *et al.*, 2016). EWAS2.0 can calculate the MD coefficients (Zhao *et al.*, 2016), identify the MD blocks using Gabriel *et al.*'s algorithm (Barrett *et al.*, 2005; Gabriel *et al.*, 2002) and estimate the frequency of meplotype (a group of specific epi-alleles on a chromosome) using Excoffier *et al.*'s Maximum Likelihood Estimate method (Excoffier and Slatkin, 1995). For case/control data, EWAS2.0 can scan the whole epigenome and identify the disease-related meplotype (calculate Chi-square, *P*-value, odd ratio and 95% confidential interval). We suggest that users perform meplotype analysis to identify the combinations of some SMP loci related to diseases/phenotypes after performing the single SMP analysis.

Since the results of the similar EWAS studies are often inconsistent, we developed an epigenome-wide meta-analysis module. At first, EWAS2.0 test the heterogeneity between individual studies using Cochran's Q-statistics. Then, the fixed effects model (all studies share a common effect size) and a random effects model (each study has a specific effect size) were used to evaluate the association between marker and disease/phenotype. We suggest that users select fixed effects model for low heterogeneity, and random effects model for high heterogeneity.

EWAS2.0 software is a JAVA application based on JAVA 1.7 and is freely available at: http://www.ewas.org.cn. The current status of EWAS2.0 is depicted in Table 1. More functions will be added in the future version (such as EWAS for gene region, KEGG

Table 1. Overview of novel functions in EWAS2.0

Category	Description
-t.test	T-test for case/control or binomial phenotype
-linear	Linear regression analysis for continuous phenotype
-logistic	Logistic regression analysis for case/control or binomial phenotype
-cor	The Pearson's correlation coefficients for continuous phenotype
-SMP.allele_chisq	Chisq-square test for epi-allele: 2 (phenotype)*2 (M vs. U) table
-SMP.aa	Identify the type of epi-allele association: synergic or exclusive
-meplotype	Epigenome-wide meplotype association analysis
-MD	Calculate the MD coefficient
-block	Identify the MD blocks and calculate the frequency of meplotype
-meta	Epigenome-wide association meta-analysis

pathway, GO categories, network, interacting with genetic marker, regulation of gene expression, RNA modification, histone modification). Some comparisons between different methods can be found in a supplement (http://www.bioapp.org/ewas/supplement.html). We expect it to become a useful tool.

References

Barrett,J.C. *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Excoffier,L. and Slatkin,M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, **12**, 921–927.

Gabriel,S.B. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

Rakyan,V.K. *et al.* (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, **12**, 529–541.

Wahl,S. *et al.* (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*, **541**, 81–86.

Xu,J. *et al.* (2016) EWAS: epigenome-wide association studies software 1.0 - identifying the association between combinations of methylation levels and diseases. *Sci. Rep.*, **6**, 37951.

Zhao,L. *et al.* (2016) The framework for population epigenetic study. *Brief. Bioinf.*, **19**, 89–100.