European Journal of Human Genetics (2008)16, 1155–1158
© 2008 Macmillan Publishers Limited All rights reserved 1018-4813/08 $30.00
www.nature.com/ejhg

SHORT REPORT

HuGE Watch: tracking trends and patterns of published studies of genetic association and human genome epidemiology in near-real time

Wei Yu*[,1], Anja Wulf[1], Ajay Yesupriya[1], Melinda Clyne[1], Muin Joseph Khoury[1] and Marta Gwinn[1]

[1]*National Office of Public Health Genomics, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA*

HuGE Watch is a web-based application for tracking the evolution of published studies on genetic association and human genome epidemiology in near-real time. The application allows users to display temporal trends and spatial distributions as line charts and google maps, providing a quick overview of progress in the field. http://www.hugenavigator.net/HuGENavigator/startPageWatch.do

European Journal of Human Genetics (2008) **16,** 1155–1158; doi:10.1038/ejhg.2008.95; published online 14 May 2008

Keywords: genetic association; human genome epidemiology; trend

Introduction

Completion of the Human Genome Project and advances in genomic technology has stimulated the emergence of new multidisciplinary research areas. One such area is human genome epidemiology,[1] which uses population-based, epidemiological methods to assess the relationship of human genetic variation to health and disease. Human genome epidemiology (HuGE) focuses on the study of gene–disease associations, gene–gene and gene–environment interactions, and the evaluation of genetic tests.[2] The Human Genome Epidemiology Network (HuGENet) (http://www.cdc.gov/genomics/hugenet/default.htm) is a global collaboration of individuals and organizations committed to developing and disseminating population-based human genome epidemiological information. HuGENet promotes quality reporting of genetic associations, as well as the systematic and quantitative synthesis of rapidly evolving information on gene-disease associations.

HuGENet maintains a knowledge base called HuGE Navigator,[3] which contains published data on genetic associations and human genome epidemiology. Since 2001, the data have been extracted weekly from PubMed, curated, and deposited in the knowledge base.[4] Recently, an automatic literature screening tool (GAPscreener) based on machine-learning techniques (Support Vector Machine) has been used for routine literature screening.[5] This new method has significantly increased the sensitivity of the screening process to an estimated 97.5%. We also have developed and implemented a novel data-mining method[6] to extract author profiles and geographical information from PubMed records.

We developed a Web-based application called HuGE Watch as one component of HuGE Navigator. HuGE Watch can be used to track the evolution of published studies in near-real time. HuGE Watch provides researchers, health care practitioners, and funding agencies a way to easily and quickly assess the current status of research in this field.

Implementation

The HuGE Navigator knowledge base is based on an open source infrastructure[7] developed by HuGENet. The Web-based HuGE Watch application was built on J2EE technology (http://java.sun.com/javaee/) and on other Java open source frameworks such as Hibernate (http://www.hibernate.org/) and Strut (http://struts.apache.org/). We used JChart open

*Correspondence: Dr W Yu, National Office of Public Health Genomics, Centers for Disease Control and Prevention, 4770 Buford Highway, MS K-89, Atlanta, GA 30341, USA.
Tel: +1 770 488 8435; Fax: + 1 770 488 8355;
E-mail: wby0@cdc.gov

Received 1 February 2008; revised 16 April 2008; accepted 17 April 2008; published online 14 May 2008

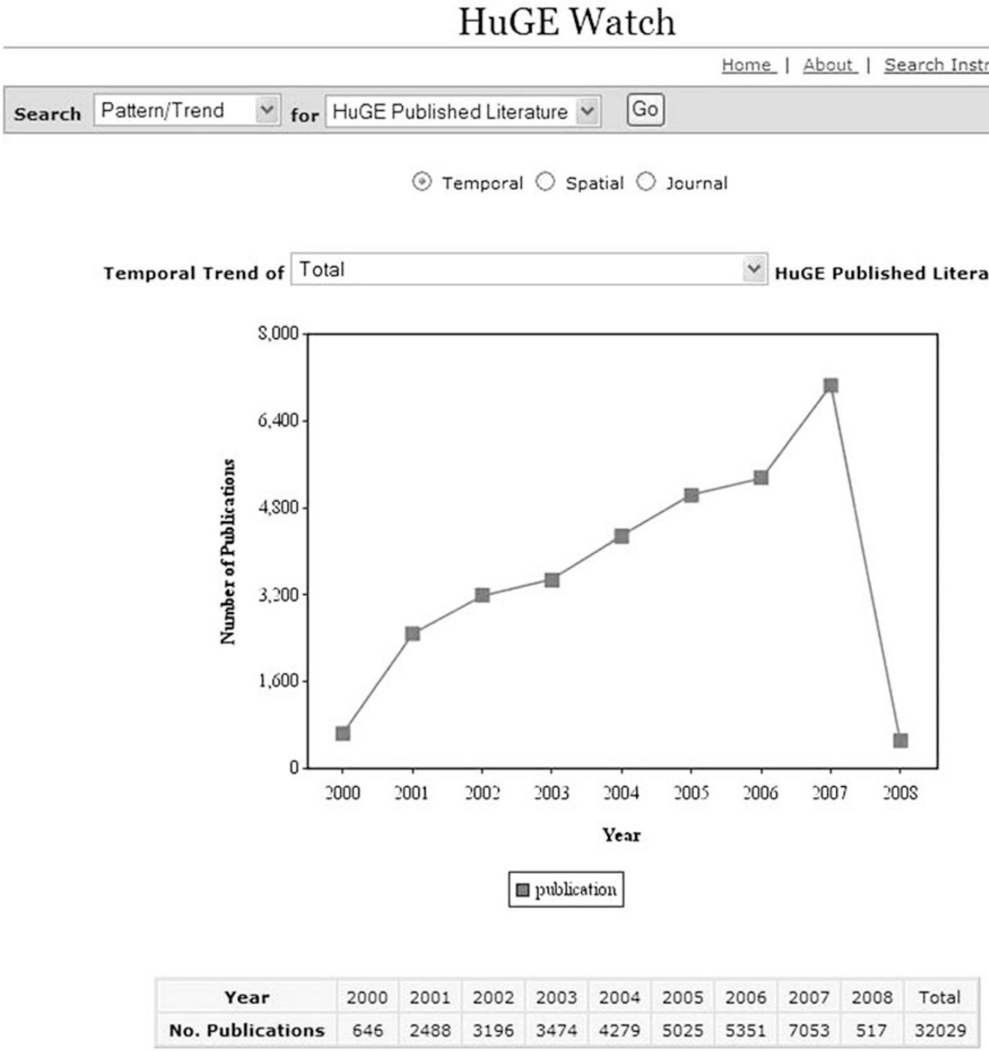
source software (http://jcharts.krysalis.org/) to generate dynamic charts and Google MAP API (http://www.google.com/apis/maps/documentation/) to build geographic maps.

HuGE Navigator records are indexed using several automatic and manual processes. A utility retrieves and parses the author information (including institute and country) from the affiliation string in PubMed records[6] and assigns it to all articles by that author. Relevant concept terms are indexed using Concept Unique Identifiers from the Unified Medical Language System (UMLS),[8] based on MeSH indexing of PubMed records. Concepts positioned under the disease category in the MeSH tree structure (http://www.nlm.nih.gov/bsd/disted/mesh/tree.html) are used for indexing by disease. The database curator indexes the gene symbol using the Entrez GeneID (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB = gene) and assigns a knowledge category (ie, genotype prevalence, gene–disease association, gene–gene interaction, gene–environment interaction, pharmacogenomics/toxicogenomics, and genetic testing) and study type (ie, observational study, meta-analysis, HuGE review, genome-wide association study, and clinical trial) to each record. Although the database is updated weekly, it lags PubMed by 1 week because of the curation process.

Features

Overall temporal trends and spatial distributions

HuGE Watch users can view publication patterns by year or by country. Patterns can be displayed for each of four parameters – number of relevant publications, number of relevant investigators, number of diseases, and number of genes. Patterns for each parameter can be specified



Search Pattern/Trend for HuGE Published Literature

● Temporal ○ Spatial ○ Journal

Temporal Trend of Total HuGE Published Literature

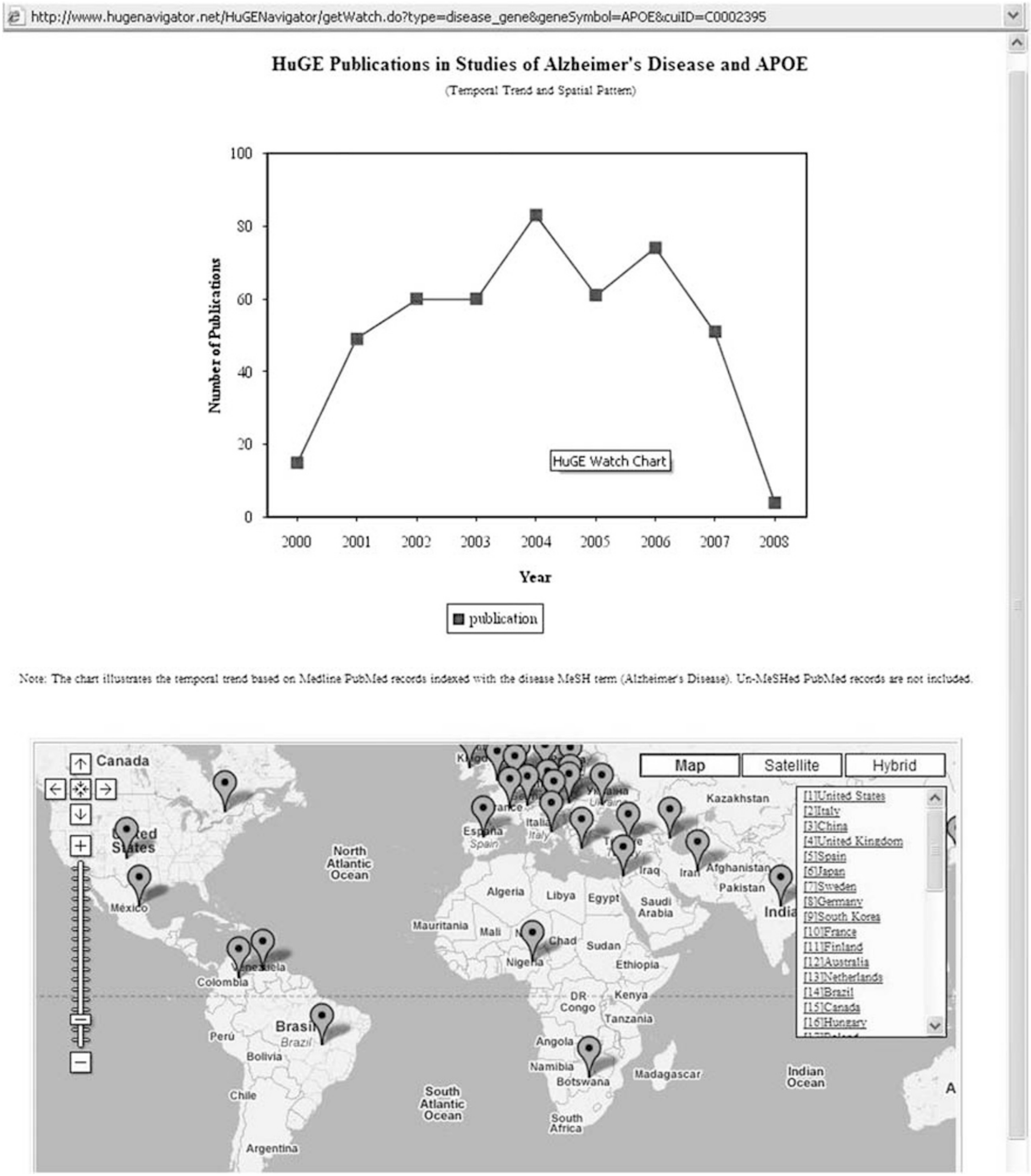
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	Total
No. Publications	646	2488	3196	3474	4279	5025	5351	7053	517	32029

Figure 1 One view in HuGE Watch.

further by category (genotype prevalence, gene–disease association, gene–gene interaction, gene–environment interaction, pharmacogenomics, genetic testing) or study type (observational study, genome-wide association, meta-analysis, HuGE review). Temporal trends are displayed as line charts, dynamically generated according to the weekly updated literature data in the database and the user's selection. Geographic distributions are presented by Google Map, with features that can be manipulated by the user (eg, zoom in or out) (Figure 1).

Temporal trends and spatial distributions for specific genes or diseases, alone or in combination

Publication patterns for specific genes or diseases, alone or in combination, can be viewed by using the two integrated components (Genopedia and Phenopedia) in HuGE Navigator. When searching any specific gene or disease term in Genopedia or Phenopedia, the publication patterns for a given query can be found by clicking HuGE Watch icon on the summary page (Figure 2).



http://www.hugenavigator.net/HuGENavigator/getWatch.do?type=disease_gene&geneSymbol=APOE&cuiID=C0002395

HuGE Publications in Studies of Alzheimer's Disease and APOE

(Temporal Trend and Spatial Pattern)

Note: The chart illustrates the temporal trend based on Medline PubMed records indexed with the disease MeSH term (Alzheimer's Disease). Un-MeSHed PubMed records are not included.

Figure 2 Temporal trends and spatial distributions of literature in a gene–disease combination view.

Journal rankings by the number of publications

In the HuGE Published Literature section, the Journal option allows users to view journals ranked by the number of publications in human genome epidemiology. Journal ranks also can be viewed for different itemized categories and study types.

Most studied genes and diseases

In the Genes Studied and Diseases Studied sections, the publication option allows users to view genes and diseases ranked by numbers of publications. The same information can be viewed for different itemized categories and study types. Publication patterns by year and by country also can be viewed for each gene or each disease term.

Examples of scenarios for HuGE Watch users

1. A quick citation analysis can provide an overview of human genome epidemiology research regionally or globally. For example, Adany and Pocsai[9] published a review of genetic epidemiology literature in Europe; a more comprehensive, updated review can now be done via HuGE Watch with just a few clicks.
2. The recent boom in genome-wide association research depends on collaboration to achieve the necessary large sample sizes. HuGE Watch combined with other components of HuGE Navigator could be used to identify potential collaborators in other countries.
3. Investigators can use HuGE Watch to help decide on the target journal for their manuscripts to increase the possibility of acceptance. For example, the *American Journal of Epidemiology* (*AJE*) has published the most HuGE review articles (47 of 71). Users might have a better chance of publishing such papers in *AJE*.

Conclusion

As the study of genetic associations and human genome epidemiology continues to grow throughout the world, research collaboration and synthesis are becoming increasingly important.[10] The HuGE Watch application is a valuable tool for monitoring and supporting these efforts. By documenting the rapid expansion of this field in near-real time, HuGE Watch provides a point of reference not only for investigators but also for funding agencies, publishers, and other stakeholders in genomics research.

References

1 Khoury MJ, Millikan R, Little J, Gwinn M: The emergence of epidemiology in the genomics age. *Int J Epidemiol* 2004; **33**: 936–944.
2 Khoury MJ, Little J, Burke W: *Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease*. New York, NY: Oxford University Press, 2004, p549.
3 Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ: A navigator for human genome epidemiology. *Nat Genet* 2008; **40**: 124–125.
4 Lin BK, Clyne M, Walsh M *et al*: Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol* 2006; **164**: 1–4.
5 Yu W, Clyne M, Dolan S *et al*: GAPscreener: An automatic tool for screening human genetic association literature in PubMed using the support vector machine technique. *BMC Bioinformatics* 2008; **9**: 205.
6 Yu W, Yesupriya A, Wulf A, Qu J, Gwinn M, Khoury MJ: An automatic method to generate domain-specific investigator networks using PubMed abstracts. *BMC Med Inform Decis Mak* 2007; **7**: 17.
7 Yu W, Yesupriya A, Wulf A, Qu J, Khoury MJ, Gwinn M: An open source infrastructure for managing knowledge and finding potential collaborators in a domain-specific subset of PubMed, with an example from human genome epidemiology. *BMC Bioinformatics* 2007; **8**: 436.
8 Lindberg DA, Humphreys BL, McCray AT: The Unified Medical Language System. *Methods Inf Med* 1993; **32**: 281–291.
9 Adany R, Pocsai Z: Genetic epidemiology literature in Europe – an overview. *Eur J Public Health* 2007; 30–32.
10 Ioannidis JP, Gwinn M, Little J *et al*: A network of investigator networks in human genome epidemiology. *Am J Epidemiol* 2005; **162**(4): 302–304.