Journal of Infection and Public Health (2016) 9, 331—338

http://www.elsevier.com/locate/jiph

Human Disease Insight: An integrated knowledge-based platform for disease-gene-drug information

Munazzah Tasleem, Romana Ishrat, Asimul Islam, Faizan Ahmad, Md. Imtaiyaz Hassan*

Centre for Interdisciplinary Research In Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi 110025, India

Received 5 April 2015; received in revised form 21 May 2015; accepted 23 October 2015

KEYWORDS
Human disease database;
Human genome project;
Data integration;
Knowledge management system;
Relational database management system

Summary The scope of the Human Disease Insight (HDI) database is not limited to researchers or physicians as it also provides basic information to non-professionals and creates disease awareness, thereby reducing the chances of patient suffering due to ignorance. HDI is a knowledge-based resource providing information on human diseases to both scientists and the general public. Here, our mission is to provide a comprehensive human disease database containing most of the available useful information, with extensive cross-referencing. HDI is a knowledge management system that acts as a central hub to access information about human diseases and associated drugs and genes. In addition, HDI contains well-classified bioinformatics tools with helpful descriptions. These integrated bioinformatics tools enable researchers to annotate disease-specific genes and perform protein analysis, search for biomarkers and identify potential vaccine candidates. Eventually, these tools will facilitate the analysis of disease-associated data. The HDI provides two types of search capabilities and includes provisions for downloading, uploading and searching disease/gene/drug-related information. The logistical design of the HDI allows for regular updating. The database is designed to work best with Mozilla Firefox and Google Chrome and is freely accessible at http://humandiseaseinsight.com.

Introduction

Scientists have documented diseases within specific categories in various online databases. Due to advancements in science and technology, especially

* Corresponding author. Tel.: +91 11 2698 3409; fax: +91 11 2698 3409; mobile: +91 9990323217.
E-mail address: imtiyaz.hassan@gmail.com (Md.I. Hassan).

http://dx.doi.org/10.1016/j.jiph.2015.10.018

in genomics and information technology, we have entered in an exciting era of modern biology. The major challenge that the medical science community is presently facing is the integration of vast and rapidly growing amounts of information on various diseases into a holistic understanding. Recently, there has been considerable progress in disease genetics and genome-related medicine, leading to the generation of extensive data. The remarkable approach adopted by the Human Genome Project [1,2] making human genome, transcriptome and proteome data publicly available through online databases has facilitated in-depth investigations of disease genetics.

Currently, databases containing information about human diseases are focused predominantly on a particular disease category, such as all known Mendelian disorders described in the Online Mendelian Inheritance in Man (OMIM) [3]; infectious disease information, such as that found in the Infectious Disease Biomarker Database [4]; rare childhood diseases (http://www.madisonsfoundation.org/index.php); hereditary ocular diseases (https://disorders.eyes.arizona.edu); dermatological diseases (http://www.aocd.org/); and gastrointestinal diseases (http://www.gastro.net.au/). Other databases, such as the Malaysian National Cardiovascular Disease Database (NCVD), give an overview of cardiovascular disease and maintain records of patients suffering from these diseases [5]. In addition, the Indian Genetic Disease Database (IGDD) is a mutation data repository for genetic diseases in India; however, the information stored in the IGDD is helpful only to researchers [6]. Another database that provides information about autoimmune disorders is the Autoimmune Disease Database, which gives descriptions of autoimmune disorders and links these diseases to candidate genes, which is, again, a database that useful only for researchers [7]. The Comparative Toxicogenomic Database (CTD) is a rich resource for researchers to access information about the etiology of environmental diseases and explore chemical-gene and protein interactions [8]. Such attempts have contributed enormously to efforts related to the prevention, diagnosis and treatment of diseases and have resulted in the development of new approaches to alleviate the consequences of life-threatening illnesses. However, no disease database providing guidance related to bioinformatics tools and information available to members of the non-scientific public currently exists. Hence, the integration of information on all human diseases from different categories within a common place has become an important issue in the field of bioinformatics.

In recent years, the scientific community has been able to gain information through a number of useful internet-accessible resources, in addition to text books of biological and medical information. In the database development field, internet-accessible information retrieval systems have recently become popular due to the reduced costs of data storage and transfer. Vast amounts of biomedical information can be accessed through the World-Wide Web (WWW), although this information is scattered. Additionally, the heterogeneity and complexity of the available resources means that some information cannot be retrieved in a timely fashion. Furthermore, the rapidly growing fields related to disease information, genomics and proteomics databases and drug discovery and the expeditious development of computational tools to solve biological queries necessitate the integration of all of these information sources in a well-organized and concise database. Over a period of time, advancements in diagnostic evaluation and treatment have emerged. To provide the community with the most recent knowledge on human diseases and the discovery of genes involved in diseases, we have created a Knowledge Management System (KMS) that includes information on various categories of human diseases, the drugs used to cure these diseases, and the genes involved in causing these diseases, as well as bioinformatics tools for analyzing the genes in question. The HDI is therefore a comprehensive database of human diseases belonging to various categories that is cross-linked to other databases to retrieve detailed information on genes, drugs and tools. The HDI exhibits broad utility as it renders clinical information for physicians, genetic information and tool classification for researchers and disease descriptions for the general public. Thus, the HDI aims to provide a better understanding of human diseases, genes and drugs, as well as their relationships with one another, and allows easy retrieval of information through its user-friendly web-based applications.

Materials and methods

Overview of the database, HDI

The Human Disease Insight (HDI) database introduces an integrated knowledge base of diseases, genes and drugs and a list of bioinformatics tools with a user-friendly interface. The database was developed to allow simple retrieval of disease/gene/drug information and exploratory analysis of disease-specific genes within a single