Free shopping in a virtual bazaar of gene regulation data

October 4 2007

An international team has opened a virtual bazaar, called PAZAR, which allows biologists to share information about gene regulation through individually managed 'boutiques' (data collections). According to research published in the online open access journal, Genome Biology, customers can access data without any charge from any boutique or extract information from the 'superstores' that aggregate data of similar types.

In deciphering the human genome sequence, researchers hope to understand the when and where of gene expression because this could speed development of novel cancer therapies or stem cell treatments for degenerative disease, and explain complex diseases such as diabetes.

Much of the information gathered in costly studies of gene regulation is poorly accessible if available at all. Individual research teams often generate databases or post files on the internet, but these data are fragmented and can be lost over time. The research team, led by Wyeth Wasserman at the University of British Columbia and Child & Family Research Institute, Vancouver, Canada, and colleagues from Bulgaria, Canada, France, and the USA, describe a novel approach to managing this information by bringing it together for the first time using PAZAR.

PAZAR is, explain the researchers, "an open-access and open-source database of transcription factor and regulatory sequence annotation". As such, it fulfils a longstanding need for a large data collection of regulatory sequences unrestricted by commercial concerns. Its novel

shopping-mall-like structure (pazar is Bulgarian for shopping mall) will allow researchers to share data collections and computational predictions in an organized and accessible manner.

In order to demonstrate the advantages and features of PAZAR and its depth of annotation, the researchers used the Pleiades Promoter Project collection of brain-linked regulatory sequences as a show case. They have been working internationally with boutique operators and are currently expanding the data represented, and improving the curation tools. By bringing small data collections together Wasserman and colleagues are aiming to bring the data to international scientific customers and are encouraging other researchers to open new boutiques in this genomic mall.

Citation: Software PAZAR: a framework for collection and dissemination of cis-regulatory sequence annotation, Elodie Portales-Casamar, Stefan Kirov, Jonathan Lim, Stuart Lithwick, Magdalena I Swanson, Amy Ticoll, Jay Snoddy and Wyeth W Wasserman, Genome Biology (in press)

Source: BioMed Central