

Human Transcriptional Regulation Interaction database (HTRIdb): Atualização 2010.

Luiz A. Bovolenta¹, Marcio L. Acencio¹, Ney Lemke¹

¹Instituto de Biociências de Botucatu – Universidade Estadual Paulista (UNESP)
CEP 18618-970 – Botucatu – SP – Brasil

{labovolenta, lemke, mlacencio}@ibb.unesp.br

Abstract. *In this paper we described the Human Transcription Regulation Interaction database (HTRIdb), an open-source database of transcription factor (TF)-regulated gene (RG) interactions with Pubmed reference to the experimental evidence of regulation that has been created in response to need of a user-friendly open-access repository of TF-regulated gene interactions. Currently, most TF-RG data are either dispersed in literature or found in paid subscription databases or open-access, non-user friendly databases. The HTRIdb allows users to easily retrieve and download TF-RG interaction data in spreadsheet or text format as well as to graphically visualize the interactions.*

Resumo. *Neste trabalho, nós descrevemos o desenvolvimento de um banco de dados de interações transcricionais de seres humanos experimentalmente verificados, denominado de Human Transcription Regulation Interaction database (HTRIdb), para suprir a necessidade de uma base de dados de acesso livre e mais completa que centralize os dados que atualmente se encontram dispersos na literatura biomédica ou em bancos de dados pagos ou de uso complexo. O HTRIdb permite aos usuários recuperarem os dados de uma interação regulatória específica de interesse ou todas as interações regulatórias presentes no banco de dados, fazerem download dos dados em diferentes formatos de arquivo e visualizarem graficamente as interações regulatórias e com outras proteínas.*

1. Introdução

O entendimento da regulação gênica exige um conhecimento preciso e abrangente das interações de regulação transcricional. A fim de responder à necessidade crescente de repositórios integrados de fatores de transcrição (FT) e seus respectivos genes regulados, vários bancos de dados de fatores de transcrição de organismos específicos foram desenvolvidos, tais como o "C. elegans differential gene expression database" (EDGEDb) [Barrasa et al. 2007], o "Yeast Search for Transcriptional Regulators And Consensus Tracking" (YEASTRACT) [Teixeira et al. 2006], o "TRANSFAC® database" [Matys et al. 2006] e o "Transcriptional Regulatory Element Database" (TRED) [Jiang et al. 2007]. Estes dois últimos bancos de dados contêm dados de interações de regulação transcricional de seres humanos, mas ambos possuem limitações: TRANSFAC® [Matys et al. 2006] é o principal banco de dados de FTs, mas o acesso é restrito com apenas uma parte do seu conteúdo disponível na versão pública; TRED [Jiang et al. 2007] é um banco de dados de acesso livre, mas contém um número limitado de FTs humanos (113 FTs de um total de ≈1500 FTs conhecidos). Para superar essas

limitações, foi desenvolvido o *Human Transcriptional Regulation Interactions database (HTRIdb)*, um banco de dados relacional de interações de FTs com seus genes regulados com referência para a prova experimental da regulação no *Pubmed* baseado no Sistema de Gerenciamento de Banco de Dados PostgreSQL. Até agora, *HTRIdb* foi populado com ≈ 100 FTs que regulam ≈ 2000 genes. O *HTRIdb* pode ser acessado através de uma interface web (<http://tinyurl.com/lbbc-htridb>) onde os usuários são capazes de recuperar os dados de uma interação regulatória específica de interesse ou todas as interações regulatórias presentes no banco de dados, Além disso, usuários podem fazer *download* e recuperar os dados em formato de planilha eletrônica ou em formato de texto. O *HTRIdb* também disponibiliza uma ferramenta capaz de apresentar de forma gráfica as interações regulatórias.

2. Metodologia

2.1. Banco de Dados

A primeira fase do desenvolvimento foi a modelagem de dados, ou seja, a transformação do problema do mundo real em uma estrutura abstrata capaz de demonstrar a estrutura que irá armazenar os dados. Essa fase é realizada através de conceitos descritos no Modelo Entidade Relacionamento (MER) criado por Peter Chen [Chen 1976]. Esta estrutura abstrata pode ser representada graficamente utilizando figuras geométricas, onde cada figura representa de maneira distinta os conceitos descritos no MER, conhecido como Diagrama Entidade Relacionamento (DER).

A segunda fase foi a criação do modelo físico e a codificação da estrutura do *HTRIdb* em SQL, que foi realizada com o auxílio de uma ferramenta case, *Case Studio 2*, responsável por auxiliar na elaboração do modelo e gerar os *scripts* SQL automaticamente que foram implantados no Sistema de Gerenciamento de Banco de Dados Objeto-Relacional (SGBDOR) PostgreSQL.

2.2. Interface

As páginas da interface com conteúdo estático foram desenvolvidas em HTML, e as páginas com conteúdo dinâmico foram desenvolvidas através dos recursos da linguagem Java (JavaBeans, JSP e Servlets). A programação das páginas com conteúdo dinâmico seguiu o modelo *MVC (Model View Controller)*. Neste modelo, as páginas JSP são responsáveis pelas entradas de requisições e da visualização e formatação das respostas dos processamentos. Os Servlets são responsáveis pelo controle dos processos e o retorno das respostas para as páginas JSP. O modelo dos dados requisitados, por sua vez, são de responsabilidade do JavaBeans. Com este modelo, a lógica de programação e a visualização das informações ficam separadas.

Foram utilizadas ferramentas de tratamento de imagem para a criação das imagens presentes em cada uma das páginas da interface do banco de dados e recursos de CSS para a estruturação do *layout* da interface gráfica.

Para possibilitar a visualização gráfica das interações entre os FTs e seus genes alvos e outras proteínas, foi incorporada uma ferramenta denominada de Cytoscape Web, desenvolvida através de tecnologias Flash e JavaScript pelo laboratório de Gary Bader e Quaid Morris na Universidade de Toronto.

2.3. Coleta e inserção dos dados

Os dados das interações entre os FTs e seus genes alvos foram coletados manualmente a partir da literatura biomédica. Os dados coletados foram os seguintes:

- Nomes oficiais dos genes que codificam os FTs e dos genes regulados por esses fatores: *Human Genome Organization Gene Nomenclature Committe (HGNC)*.
- GeneIDs dos genes que codificam os fatores de transcrição e dos genes regulados por esses fatores: *National Center for Biotechnology Information (NCBI)*;
- Interações regulatórias entre FTs e seus genes alvos: literatura biomédica no *Pubmed*;
- Dados das proteínas que interagem fisicamente entre si [Costa et al. 2010].

3. Resultados

O resultado foi o desenvolvimento de um banco de dados de fatores de transcrição humanos disponibilizado gratuitamente por meio de uma interface web e desenvolvida com base na tecnologia Java.

A estrutura do banco de dados foi o primeiro resultado obtido durante o desenvolvimento. Este é demonstrado através do DER na Figura 1.

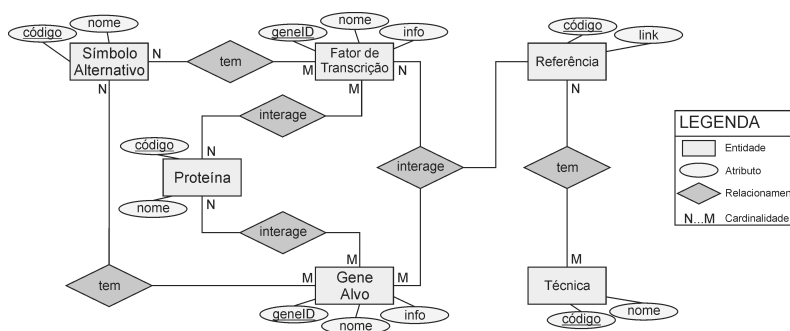
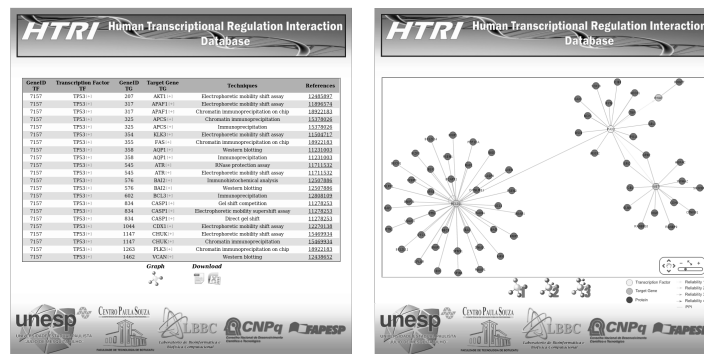


Figura 1. Diagrama Entidade Relacionamento do HTRIdb.

Outro resultado é a interface web, que possui recurso de busca que disponibiliza as seguintes informações relacionadas às interações entre os FTs e seus genes alvos em forma de uma tabela: os números identificadores e os símbolos oficiais dos genes que codificam os FTs e de seus genes alvos, os links direcionados para as páginas do *EntrezGene* contendo informações completas sobre os genes que codificam os FTs e seus genes alvos, as técnicas experimentais utilizadas para detectar as interações e os links direcionados para as referências no *Pubmed* que provam essas interações (Figura 2(a)).

O *HTRIdb* também oferece aos usuários recursos de exportação parcial ou integral dos dados presentes no banco de dados e a visualização gráfica das interações entre os FTs e seus genes alvos e com outras proteínas (Figura 2(b)).

Esse interface gráfica também disponibiliza uma página que permite especialistas a contribuir com novas descobertas relacionadas às interações regulatórias entre FTs e seus genes alvos.



(a) Página dos resultados de pesquisas realizadas. (b) Página de visualização gráfica das interações.

Figura 2. Interface gráfica do HTRIdb

4. Conclusão

O banco de dados e os meios de busca do sistema compõem uma ferramenta capaz de auxiliar na gestão de informação para pesquisadores biomédicos e outros pesquisadores da área, permitindo com que estes pesquisadores encontrem informações sobre as interações entre FTs e genes regulados em humanos de maneira mais eficiente. Possibilita, também, um maior entendimento sobre a rede de interações entre FTs e seus genes alvos que controlam todos os processos celulares, promovendo novas descobertas que podem resultar na criação de novos medicamentos capazes de melhorar tratamentos ou, até mesmo, possibilitar a cura de doenças relacionadas aos processos anormais de transcrição gênica.

Referências

Barrasa, M. I., Vaglio, P., Cavasino, F., Jacotot, L., and Walhout, A. (2007). Edgedb: a transcription factor-dna interaction database for the analysis of c. elegans differential gene expression. *BMC Genomics*, 8(1):21.

Chen, P. P.-S. (1976). The entity-relationship model toward a unified view of data. *ACM Trans. Database Syst.*, 1:9–36.

Costa, P. R., Acencio, M. L., and Lemke, N. (2010). A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. *BMC Genomics*, 11:S9.

Jiang, C., Xuan, Z., Zhao, F., and Zhang, M. Q. (2007). Tred: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res*, 35(Database issue):D137–40.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E., and Wingender, E. (2006). Transfac and its module transcompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*, 34(Database issue):D108–10.

Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., Alenquer, M., Freitas, A. T., Oliveira, A. L., and Sa-Correia, I. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *NUCLEIC ACIDS RESEARCH*, 34(Sp. Iss. SI):D446–D451.