

**DESENVOLVIMENTO DE UM BANCO DE DADOS DE FATOR DE
TRANSCRIÇÃO HUMANO**

DEVELOPMENT OF HUMAN TRANSCRIPTION FACTOR DATABASE

**DESARROLLO DE UNA BASE DE DATOS DE FACTOR DE TRANSCRIPCIÓN
HUMANO**

OSVALDO CESAR PINHEIRO DE ALMEIDA¹

LUIZ AUGUSTO BOVOLENTA²

MARCIO LUIS ACENCIO³

NEY LEMKE⁴

Recebido em setembro de 2009. Aprovado em setembro de 2009.

¹ Professor da Faculdade de Tecnologia de Botucatu. Graduado em Ciências da Computação pela Universidade de São Paulo. Mestre em Ciências da Computação e Matemática Computacional pela Universidade de São Paulo. Endereço eletrônico: cesar@fatecbt.edu.br.

² Graduado em Informática para Gestão de Negócios pela Faculdade de Tecnologia de Botucatu. Endereço eletrônico: labovolenta@gmail.com.

³ Graduado em Ciências Biológicas pela Universidade Federal de São Paulo. Mestre em Biotecnologia pela Universidade de São Paulo. Doutorando em Genética pela Universidade Estadual Paulista. Endereço eletrônico: mlacencio@gmail.com.

⁴ Professor da Universidade Estadual Paulista. Graduado em Física pela Universidade Federal do Rio Grande do Sul. Mestre em Física pela Universidade Federal do Rio Grande do Sul. Doutor pela Universidade Federal do Rio Grande do Sul. Endereço eletrônico: ney.nlemke@gmail.com.

Agradecimentos aos órgãos de fomento FAPESP e CNPq pelo apoio financeiro ao projeto.

Tékhnē e Lógos, Botucatu, SP, v.1, n. 1, out. 2009

DESENVOLVIMENTO DE UM BANCO DE DADOS DE FATOR DE TRANSCRIÇÃO HUMANO

RESUMO

Há uma dificuldade no acesso às informações sobre fatores de transcrição humanos, pois estas se encontram dispersas na literatura biomédica. A centralização destas informações em bancos de dados facilita o acesso às informações sobre os fatores de transcrição. Atualmente, existem dois principais bancos de dados responsáveis pelo armazenamento dessas informações: TRANSFAC e TRED. Este trabalho tem como objetivo o desenvolvimento de uma nova base de dados para o armazenamento de informações relacionadas aos fatores de transcrição, aproveitando as vantagens dos dois principais bancos de dados existentes e suprindo as suas necessidades e desvantagens. A base de dados gerada concentra informações sobre os fatores de transcrição humanos, incluindo os seus respectivos genes regulados, bem como as referências contendo as provas experimentais da regulação dos genes por esses fatores de transcrição. Para armazenar as informações sobre fatores de transcrição, foi utilizado o Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL. Foi implementada uma *interface web*, usando linguagem de programação Java e a tecnologia *Servlet* e JSP, para acessar as informações do banco de dados. O desenvolvimento deste trabalho resultou em uma base de dados gratuita sobre fatores de transcrição humanos. O sistema *web* para acessar estas informações é de fácil utilização e, por isso, possibilita a ampla disseminação das informações contidas nesta base.

PALAVRAS-CHAVE: Banco de Dados. Fator de Transcrição. Regulação Gênica.

DEVELOPMENT OF HUMAN TRANSCRIPTION FACTOR DATABASE

ABSTRACT

There is a difficulty in accessing human transcription factors information, because they are scattered in biomedical literature. The centralization of this information in databases make it easy to access transcription factors information. Currently, there are two main databases responsible for storing this information: TRANSFAC and TRED. This project aims to develop a new database for storing information related to transcription factors, taking advantage of the two main existing databases and meeting their needs and disadvantages. The generated database brings together information about the human transcription factors, including their regulated genes as well the references containing the experimental evidence of gene regulation by these transcription factors. To store information about transcription factors was used Management System Database (DBMS) PostgreSQL. We implemented a web interface using the Java programming language and Servlet and JSP technology to access information from the database. The development of this work resulted in a free database of human transcription factors. The web system to access this information is easy to use and, therefore, enables the wide dissemination of the information contained on this basis.

KEYWORDS: Data Base. Transcription Factor. Gene Regulation.

DESARROLLO DE UNA BASE DE DATOS DE FACTOR DE TRANSCRIPCIÓN HUMANO

RESUMEN

Hay una dificultad en el acceso a la información sobre los factores de transcripción humanos, porque están dispersos en la literatura biomédica. La centralización de la información en base de datos facilita el acceso a la información sobre los factores de transcripción. Actualmente, hay dos bases de datos principales responsables por almacenar esta información: TRANSFAC y TRED. Este estudio tiene como objetivo desarrollar una nueva base de datos para almacenar información relacionada a los factores de transcripción, tomando ventaja de las dos principales bases de datos existentes y satisfacer sus necesidades y desventajas. La base de datos reúne la información generada acerca de los factores de transcripción humanos, incluidos sus genes regulados, así como las referencias con las pruebas experimentales de la regulación génica de estos factores de transcripción. Para almacenar información acerca de los factores de transcripción se utilizó sistema de gestión de bases de datos (DBMS) PostgreSQL. Hemos implementado una interfaz web utilizando el lenguaje de programación Java y la tecnología Servlets y JSP para acceder a la información de la base de datos. El desarrollo de este trabajo ha resultado en una base de datos libre de factores de transcripción humanos. El sistema web para acceder a esta información es fácil de usar y, por tanto, permite a la amplia difusión de la información contenida en esta base.

PALABRAS CLAVE: Base de Datos. Factor de Transcripción. Regulación Génica.

1 INTRODUÇÃO

Fatores de transcrição (FTs) são proteínas reguladoras que se ligam a sequências nucleotídicas específicas localizadas a montante das regiões promotoras de genes e, através dessa ligação, podem ativar ou reprimir a transcrição dos genes. A regulação gênica, através da ligação dos FTs, é o mecanismo primário pelo qual processos biológicos complexos, como proliferação e diferenciação celulares, são controlados (DAVIDSON, 2006). Devido a essa função reguladora, a identificação e a caracterização dos FTs são importantes para a compreensão desses processos e para o desenvolvimento de terapias dirigidas a distúrbios originados pelo funcionamento anormal dos FTs, como distúrbios de desenvolvimento de tumores (DAVIDSON, 2006).

Há uma profusão de dados sobre os fatores de transcrição humanos, incluindo os genes que tais fatores regulam. Porém, esses dados encontram-se dispersos na literatura biomédica. A centralização dessas informações em bancos de dados é uma solução apropriada para a rápida obtenção de informações sobre um determinado fator de transcrição. Há, no momento, dois principais bancos de dados de fatores de transcrição humanos disponíveis que possuem informações sobre FTs e sobre

quais genes esses fatores regulam: TRANSFAC e TRED. Esses dois bancos de dados, porém, são limitados. O TRANSFAC é o maior repositório de informações sobre fatores de transcrição de diversas espécies, porém, o acesso às informações contidas neste banco é restrito: é necessário o pagamento de uma taxa para um acesso completo. O TRED, embora possa ser acessado livremente, é incompleto: estão disponíveis informações sobre 113 FTs, menos de 10% da quantidade de FTs conhecidos em seres humanos, que é cerca de 1400 (MESSINA et al. 2004).

O objetivo deste trabalho foi o desenvolvimento de uma nova base de dados sobre FTs, a fim de ampliar e facilitar o acesso às informações apresentadas anteriormente. Esta nova base de dados contém interações regulatórias validadas experimentalmente entre FTs humanos e seus respectivos genes alvos. Além de informações sobre os FTs, possui os genes regulados pelos FTs e as referências contendo as provas experimentais da regulação dos genes pelos FTs. Esse banco de dados foi implementado para ser disponibilizado gratuitamente para a comunidade, por meio de uma *interface web*, implementada em linguagem de programação Java. A criação desta base possibilitará às pessoas e às organizações interessadas, e aos que necessitam obter

informações relacionadas aos FTs, a utilização gratuita de recursos de buscas para uma obtenção agilizada e precisa dessas informações.

Para isso essa nova base de dados será completa com as seguintes informações:

- Símbolos oficiais dos genes humanos que codificam os fatores de transcrição de acordo com o banco de dados *Human Genome Organization Gene Nomenclature Committee* (HGNC);

- GeneIDs, números identificadores únicos dos genes humanos que codificam fatores de transcrição, de acordo com o banco de dados EntrezGene do *National Center for Biotechnology Information* (NCBI);

- Símbolos oficiais dos genes regulados pelos fatores de transcrição de acordo com o HGNC;

- GeneIDs dos genes regulados pelos fatores de transcrição de acordo com o banco de dados Entrez Gene do NCBI;

- *Links* para resumos de artigos da literatura no Pubmed, contendo as provas experimentais das regulações dos genes pelos fatores de transcrição.

O desenvolvimento de um novo banco de dados que disponibilize gratuitamente acesso às informações sobre FTs facilitará a procura e obtenção destas informações. Isto deve auxiliar na disseminação e entendimento dos dados

Tékhnē e Lógos, Botucatu, SP, v.1, n. 1, out. 2009

sobre FTs, facilitando a gestão destas informações para pesquisadores que têm como objetivo a compreensão do funcionamento dos processos regulatórios. Apoiando, assim, o desenvolvimento de terapias dirigidas aos distúrbios originados pelo funcionamento anormal dos fatores de transcrição, como distúrbios de desenvolvimento de tumores e outras doenças.

2 TRANSCRIÇÃO GÊNICA

Todos os processos biológicos que ocorrem em um ser vivo dependem de uma série de instruções herdadas e passadas de geração para geração. Essas instruções são armazenadas de maneira codificada no ácido desoxirribonucléico (DNA) (ALBERTS et al. 1999).

O DNA é formado por uma sequência de nucleotídeos que, por sua vez, são estruturas formadas por um grupo fosfato, um açúcar (desoxirribose) e uma das bases nitrogenadas. As bases nitrogenadas são a adenina (A), a citosina (C), a guanina (G) ou a timina (T) e é a ordem ou sequência dessas bases em certos trechos do DNA, chamadas de genes, que codifica as informações necessárias para o funcionamento correto dos processos biológicos na célula (BERG et al. 2002).

Os genes são sequências de nucleotídeos no DNA que contêm as

informações para a produção e o controle da produção das proteínas. As proteínas são as moléculas que efetivamente atuam nos processos biológicos e são formadas por uma sequência de aminoácidos que é determinada por certa sequência de nucleotídeos contida no gene. A decodificação das informações contidas nessa parte do gene para a produção da sequência correta de aminoácidos ocorre da seguinte forma: as informações contidas no gene são transcritas em outra molécula, chamada de RNA mensageiro (RNAm), e as informações transcritas nesse RNAm são, por sua vez, traduzidas em uma sequência de aminoácidos.

O controle da produção de proteínas é feito principalmente através do controle da transcrição do gene, isto é, o processo de formação do RNAm. As informações para esse controle estão localizadas em sequências de nucleotídeos nos genes, próximas às sequências que codificam as proteínas, conhecidas como sequências regulatórias. O processo biológico relacionado a este controle da transcrição do gene é denominado de regulação gênica (ALBERTS et al. 1999).

Segundo Alberts et al. (1999) na regulação gênica, proteínas denominadas fatores de transcrição (FTs) se ligam às sequências regulatórias presentes nos genes e regulam a transcrição destes. Centenas de sequências regulatórias de genes foram

identificadas e cada uma dessas sequências é reconhecida por um ou mais fatores de transcrição que, dependendo da situação, podem aumentar ou diminuir a transcrição do gene e, conseqüentemente, aumentar ou diminuir a produção da proteína codificada por esse gene.

Os fatores de transcrição são decisivos nos organismos vivos. O desenvolvimento de um órgão inteiro pode ser disparado por apenas um fator de transcrição, como demonstrado, por exemplo, em estudos do desenvolvimento do olho na mosca-da-fruta *Drosophila melanogaster*. Neste organismo, um gene chamado *Ey* foi transcrito artificialmente nas etapas iniciais do desenvolvimento em células que normalmente iriam formar as patas. Essa transcrição gênica anormal resultou no desenvolvimento de um olho no meio das patas (HALDER et al. 1995).

Em humanos, fatores de transcrição estão relacionados também a vários distúrbios, sendo que dos 1400 dos fatores de transcrição conhecidos em seres humanos, cerca de 300 estão envolvidos com doenças humanas, como por exemplo, distúrbio de crescimento, no qual foram encontradas mutações em fatores de transcrição relacionados à formação da hipófise, glândula que produz numerosos e importantes hormônios, ocasionando a baixa estatura (MARUI et al. 2002).

2.1 FTs e a indústria farmacêutica

Como descrito anteriormente, dos 1400 fatores de transcrição conhecidos em seres humanos, cerca de 300 estão envolvidos com doenças humanas de acordo com o banco de dados *Morbid Map do Online Mendelian Inheritance in Man* (OMIM) (BAXEVANIS, 2003).

Segundo Wishart (2006) desses 300 FTs, há alguns exemplos que já são alvos de drogas para o tratamento de diversas doenças. O fator de transcrição RARA, por exemplo, é alvo da droga alitretinoína para o tratamento de acne e de outras doenças e, segundo Caetano (1999), essa droga está presente no medicamento Panretin produzido pelo laboratório *Ligand Pharmaceuticals UK Ltd.* O fator de transcrição PPARA, por sua vez, é alvo da droga clofibrato para o tratamento de hiperlipoproteinemia tipo III e hipertrigliceridemia grave. Segundo Caetano (1999), esta droga está presente nos medicamentos Claripex AL, Lipofacton e Sinteroid produzido pelo laboratório Sanofi. Há também exemplos de drogas utilizadas no tratamento do câncer, como o fator de transcrição AR, que é alvo da droga nilutamida no tratamento de câncer de próstata. Segundo Caetano (1999), a nilutamida está no medicamento Anadron produzido pelo laboratório Hoechst. Outro exemplo é o fator de transcrição ER que é

alvo da droga tamoxifeno no tratamento de câncer de mama. O tamoxifeno presente nos medicamentos Novadex e Novaldex D, ambos produzidos pelo laboratório Zeneca, está presente também no medicamento Tecnotax produzido pelo laboratório Zodiac (CAETANO, 1999).

Segundo Karamouzis et al. (2002), drogas desenvolvidas para controlar os FTs no tratamento do câncer são mais eficientes do que as drogas convencionais. As drogas convencionais têm como alvos proteínas localizadas na membrana das células doentes, que funcionam como receptores e transmissores de informações do meio extracelular para o núcleo, para que a célula tenha uma resposta adequada ao seu meio extracelular. Essas proteínas de membrana, portanto, afetam todo comportamento celular e, justamente por estarem na superfície celular, são mais fáceis de serem alvos para drogas. Porém, como o processo de transmissão das informações até o núcleo pode seguir por mais de uma via, essas drogas convencionais podem elicitar comportamentos celulares indesejáveis e, portanto, podem não ter o efeito esperado, com consequente perda de eficiência. Já as drogas que têm como alvos os FTs atuam diretamente no núcleo e, portanto, a probabilidade de ocorrência de comportamentos celulares indesejáveis é menor (KARAMOUZIS et al. 2002).

2.2 Bancos de dados sobre FTs

Há, no momento, dois principais bancos de dados que armazenam os dados das interações regulatórias entre fatores de transcrição e seus genes alvos: o TRANSFAC e o TRED.

O TRANSFAC é o maior repositório de informações sobre fatores de transcrição de diversas espécies, porém, o acesso às informações contidas neste banco é restrito, sendo necessário o pagamento de uma taxa anual (MATYS et al. 2006). Para os Estados Unidos, Canadá e Japão, essa taxa é de 750 dólares para direito de acesso e 1200 dólares para direito de acesso e *download*. Para países da Europa e outros países ao redor do mundo, exceto os citados acima, essa taxa é equivalente a 600 euros para direito de somente acesso ao banco de dados e de 1000 euros para direito de acesso e *download* dos dados (TYSIAK e WINGENDER, 1999).

O TRED, embora possa ser acessado livremente, é incompleto. Estão disponíveis informações sobre 113 fatores de transcrição, menos de 10% da quantidade de fatores de transcrição conhecidos em seres humanos, que é cerca de 1400 (MESSINA et al. 2004).

2.3 Pesquisas sobre FTs

Segundo Sayers et al. (2008), o Pubmed, base de dados onde as publicações dos artigos científicos biomédicos são divulgadas, os trabalhos em regulação gênica têm ocupado um espaço cada vez maior. No período de 1980 a 1989, dos cerca de 3 milhões de artigos científicos publicados, cerca de 12 mil (0,4%) deles eram sobre regulação gênica. No período de 1990 a 1999, dos 4,3 milhões de artigos publicados, aproximadamente 57 mil (1,3%) eram sobre regulação gênica: isso representa um aumento de aproximadamente 200% no número de artigos sobre regulação gênica em relação à década de 80. De 2000 a 2007 foram publicados cerca de 90 mil artigos científicos sobre regulação gênica, o que representa aproximadamente 1,8% do total de artigos publicados nesse período. Somente nesses sete anos, houve um aumento de aproximadamente 46% no número de artigos sobre regulação gênica em relação à década de 90 e 225 % em relação à década de 80. Esses resultados sugerem, portanto, que o interesse sobre regulação gênica vem aumentando.

3 DESENVOLVIMENTO WEB

No início, a internet oferecia apenas páginas estáticas, mas agora com as

diversas tecnologias, são comuns as aplicações que disponibilizam páginas com conteúdo dinâmico e personalizado. Aplicações que possibilitam consultas a banco de dados, integração com sistemas corporativos entre outras funcionalidades (TEMPLE et al. 2004).

Tecnologias como *Servlets* e páginas JSP, que são componentes Java que permitem disponibilizar informações em páginas na internet, possibilitam o desenvolvimento dessas aplicações. Elas oferecem vantagens sobre as tecnologias existentes, herdadas da própria linguagem de programação Java: portabilidade, aplicação pode ser implantada em diversas plataformas, facilidade de programação, flexibilidade, e outros benefícios como escalabilidade e eficiência (TEMPLE et al. 2004).

3.1 Java

Conforme Figueira (2005), a linguagem de programação Java foi apresentada ao público por sua desenvolvedora Sun Microsystems, no ano de 1994. Projetada inicialmente para ser utilizada em dispositivos eletrônicos inteligentes (agendas eletrônicas, televisores, etc.), ganhou aceitação e popularidade ao ser usada em pequenos programas (*Applets*) executados em navegadores da Internet.

Java apresenta várias vantagens em relação às outras linguagens de programação, dentre as quais se destaca a independência de plataforma, uma das principais características, e que permite programas escritos nesta linguagem serem executados em qualquer máquina ou sistema operacional. Além disso, apresenta outras características, tais como a simplicidade e uma estrutura mais limpa que torna a programação mais simples em relação a outras linguagens. Por ser uma linguagem completamente orientada a objetos, permite a reutilização do código, aumentando assim a produtividade. A linguagem também é elogiada, pelo conjunto bem documentado de bibliotecas (FIGUEIRA, 2005).

Programas escritos em linguagem orientada a objetos, como Java, são constituídos por várias partes, denominadas classes, e estas, por sua vez, são constituídas de métodos, que são responsáveis por realizarem tarefas e retornarem informações ao completarem as tarefas (DEITEL e DEITEL, 2001).

3.2 Modelo MVC

Segundo Temple et al. (2004), o desenvolvimento *web* envolve diferentes tipos de equipes: equipes responsáveis pelo desenvolvimento da lógica de programação das aplicações, os programadores, e equipes

responsáveis pela formatação e *layout* das páginas das aplicações, os *web-designers*. Existe um ponto de intersecção entre essas equipes. A inicialização do trabalho de uma necessita a finalização do trabalho de outra, e alterações efetuadas por uma equipe, envolverá necessariamente o trabalho da outra, situação que pode prejudicar tanto o cronograma quanto o custo de um projeto.

O modelo MVC ou *Model View Controller* é utilizado com o intuito de tornar o trabalho dessas equipes o mais independente possível (TEMPLE et al. 2004).

A arquitetura básica do modelo MVC, se vale do uso de *Servlets*, *JavaBeans* e páginas *JSP*, onde os *Servlets* controlam as requisições recebidas (*Controller*), os *JavaBeans* implementam a lógica da aplicação (*Model*), e as páginas *JSP* se encarregam da apresentação do resultado (*View*) (TEMPLE et al. 2004).

3.3 SGBD - PostgreSQL

Sistemas Gerenciadores de Banco de Dados (SGBD) são constituídos por um conjunto de programas para acessar um conjunto de dados associados, denominados banco de dados, podendo ser armazenados diferentes tipos de dados a esse conjunto. Os SGBDs têm com principal objetivo proporcionar um ambiente conveniente e eficiente para a recuperação e

Tékhnē e Lógos, Botucatu, SP, v.1, n. 1, out. 2009

armazenamento das informações do banco de dados (SILBERSCHATZ et al. 1999).

Segundo Date (2000), Sistema de Gerenciamento de Banco de Dados (SGBD) é um conjunto de *softwares* que tratam todo o acesso ao banco de dados, ou seja, todas as solicitações de acesso ao banco de dados são processadas pelo SGBD. Portanto, uma das funcionalidades do SGBD é isolar os usuários do banco de dados dos detalhes do nível de *hardware*, por meio de operações em um nível mais alto, via linguagem SQL.

PostgreSQL é um SGBD de livre distribuição, podendo ser comparado aos melhores sistemas de gerenciamento de banco de dados existentes. O PostgreSQL possui várias vantagens, como a utilização de baixos recursos de memória e processamento. Além de proporcionar uma grande economia em licenças de *software*, possui também uma grande capacidade de gerenciar grandes volumes de dados (LIMA, 2007).

4 METODOLOGIA

Para o desenvolvimento deste trabalho foi realizado um levantamento dos requisitos necessários para a implementação de uma base de dados de FTs. Em seguida, foram elaborados os modelos utilizados para a estruturação e desenvolvimento do banco de dados.

O desenvolvimento do banco de dados iniciou com a elaboração do Modelo Entidade Relacionamento (MER), modelo utilizado para representar a estrutura do banco de dados de uma forma conceitual, e que pode ser representado de uma maneira gráfica por meio do Diagrama Entidade Relacionamento (DER).

Após a elaboração do Diagrama Entidade Relacionamento (DER), iniciou-se o processo de identificação das tabelas a serem criadas no banco de dados. Para isso, o DER resultante foi interpretado em um modelo de dados, equivalente as tabelas que serão geradas no SGBD. A partir deste modelo foram gerados os *scripts* de criação das tabelas baseado em linguagem SQL. Os *scripts* foram aplicados no SGBD PostgreSQL para a criação das tabelas sobre FTs.

No desenvolvimento da *interface web*, baseado na linguagem de programação Java, foi necessário a instalação e configuração de um ambiente de desenvolvimento de aplicações para *web*. Esse ambiente de desenvolvimento foi composto de uma versão Java SE *Development Kit* (JDK), que é utilizada na compilação de aplicações Java. Também foi instalado um servidor de aplicações Java chamado JBOSS. Este servidor permite o executar aplicações Java baseadas em *interface web* baseada nas tecnologias *Servlets* e JSP.

As páginas desenvolvidas com recursos dinâmicos foram codificadas baseadas no modelo MVC. Neste caso, as páginas JSP são responsáveis pelas entradas de requisições e da visualização das respostas das consultas. Os *Servlets* são responsáveis pelo controle dos processos e o retorno das respostas para as páginas JSP. E o modelo dos dados requisitados são responsabilidade dos JavaBeans. Com este modelo, a lógica de programação e a visualização das informações se mantêm separadas.

Em seguida, foi implementada a conexão com a base de dados. Utilizando o recurso do Java de JDBC, foram interligados os *Servlets*, de controle, com todas as tabelas da base de dados de FTs. A implementação foi realizada de modo que os dados, consultados no banco de dados, são transportados até as páginas JSP usando JavaBeans.

Os dados relacionados aos fatores de transcrição que serão inseridos no banco de dados foram selecionados manualmente a partir das seguintes fontes:

- Nomes oficiais dos genes que codificam os fatores de transcrição 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: Entrez Gene do

National Center for Biotechnology Information (NCBI);

- Genes regulados pelos fatores de transcrição: literatura biomédica no Pubmed e bancos de dados TRED;

Com as informações cadastradas no banco de dados de FTs, e as páginas do sistema *web* implementadas, é possível realizar as consultas sobre os FTs. A Figura 1 ilustra como serão realizadas as consultas, e como as respostas serão retornadas, permitindo que os usuários do sistema pesquisem os dados sobre FTs.

5 RESULTADOS E DISCUSSÃO

O resultado dos estudos e desenvolvimento deste trabalho foi uma base de dados de fatores de transcrição humanos. Base disponibilizada gratuitamente por meio de uma *interface web*, desenvolvida com base na tecnologia Java, que oferece como características uma maior portabilidade, segurança e facilidade ao acesso aos dados armazenados nessa base de dados. O acesso a esta base

disponibiliza organizações e instituições de ensino às pessoas que necessitam de informações relacionadas ao processo de Transcrição Gênica, uma ferramenta de pesquisa que oferece um recurso de busca rápido, fácil, preciso e gratuito de informações relacionadas aos fatores de transcrição de seres humanos.

Com a centralização dos dados relacionados a fatores de transcrição de seres humanos, introduzida por esta nova base de dados, os especialistas biomédicos e outros pesquisadores da área (como por exemplo: pesquisadores da área de P&D de laboratórios farmacêuticos) ganham uma nova ferramenta de pesquisa, que facilita a obtenção de informações e o entendimento do processo de transcrição gênica. Isto contribui para o aumento das possibilidades de descoberta de novos medicamentos, que podem agir de maneira mais eficaz contra doenças relacionadas aos processos de transcrição gênica anormais, e contribui, também, na descoberta de possíveis curas para essas doenças.

Essa base de dados, acessada por

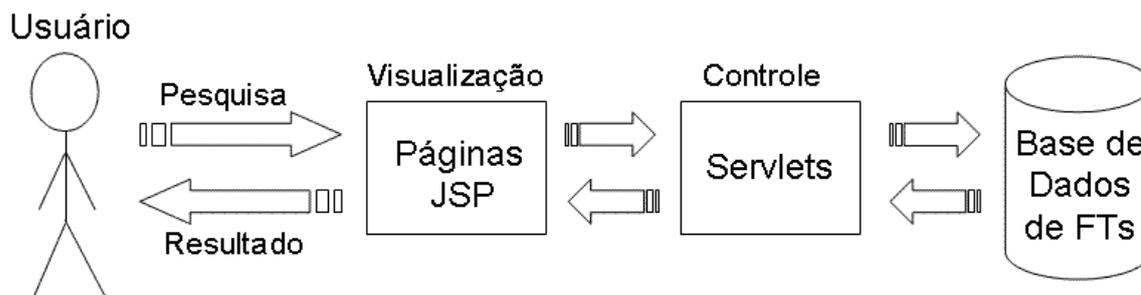


Figura 1. Esquema ilustrativo sobre o funcionamento do sistema *web* para pesquisa sobre FTs.

meio de uma *interface web*, foi desenvolvida para ser de fácil manuseio e acesso aos pesquisadores. Com isso, deve proporcionar rapidez e facilidade nas buscas sobre as informações relacionadas aos fatores de transcrição.

A página de introdução da *interface web* do banco de dados *Human Transcriptional Regulation Interaction Database* - HTRIdb (denominação do banco de dados desenvolvido), (ilustrada na Figura 2), descreve uma breve história sobre a importância do estudo dos fatores de transcrição para a formação de uma rede de fatores de transcrição de seres humanos.

Além disso, contém informações sobre quais dados estão inseridos no banco de dados, como: os números identificadores dos fatores de transcrição; os nomes oficiais dos fatores de transcrição; as informações adicionais dos fatores de transcrição; os números identificadores dos genes alvos; os nomes oficiais dos genes alvos; as informações adicionais dos genes alvos; as referências bibliográficas que comprovam as interações entre esses fatores de transcrição e os genes alvos; e a técnica utilizada para comprovar essa interação.

A página de pesquisa do HTRIdb oferece aos pesquisadores dois recursos de

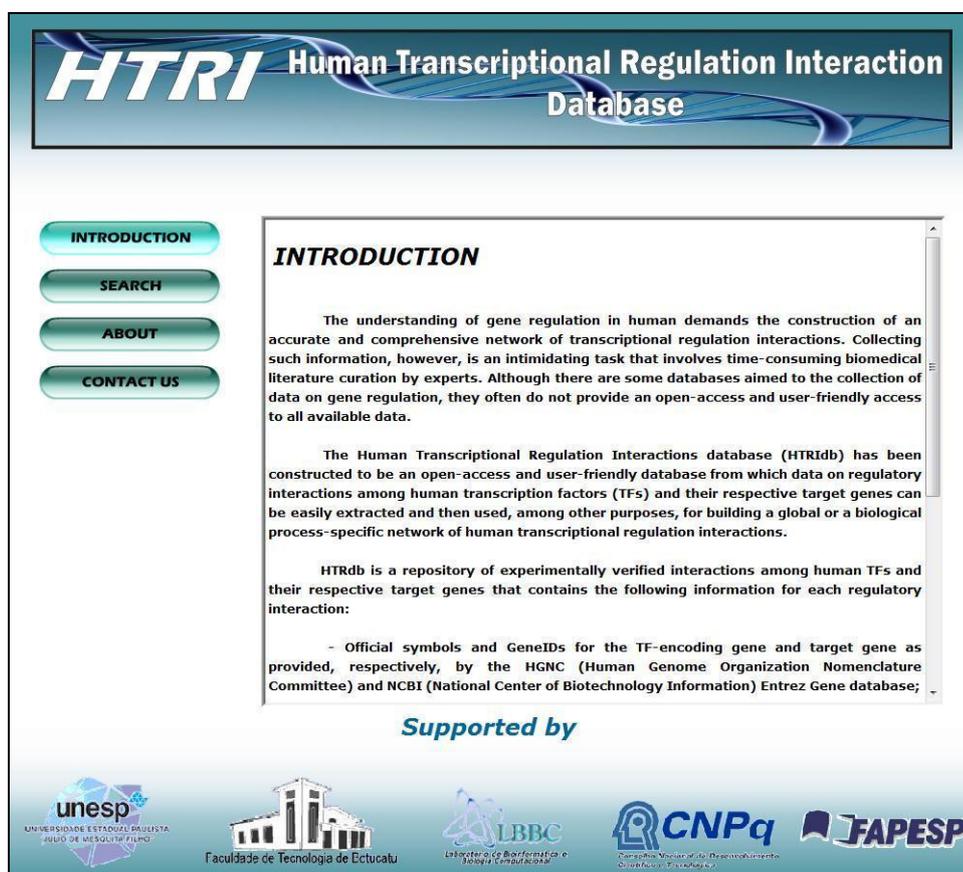


Figura 2. Página de introdução do sistema *web* para disponibilização de uma base de dados sobre os fatores de Transcrição Gênica.

pesquisas: 1) pesquisas realizadas usando os dados conhecidos de fatores de transcrição; 2) pesquisas usando os dados conhecidos de genes alvos. A pesquisa por fatores de transcrição possui recursos de procura pelo número identificador do fator de transcrição e pelo nome oficial do fator de transcrição. Já a pesquisa por genes alvos possui os recursos de procura pelo número identificador do gene alvo e pelo nome oficial do gene alvo. As páginas de pesquisa oferecem, ainda, dois tipos de procura e listagem das informações: 1) aplicado apenas a um fator de transcrição, ou um gene alvo; 2) aplicados a todos os fatores de transcrição ou a todos os genes alvo, presentes no banco de dados. A Figura

3 ilustra a página de pesquisa por genes alvos (similar a página de pesquisa por fatores de transcrição).

As páginas de resposta das consultas por genes alvos e fatores de transcrição são similares, contudo, cada tipo de procura acaba gerando um resultado com tabulação distinta. Na consulta por genes alvos os dados são tabulados por: número identificador do gene alvo; nome do gene alvo; mais informações sobre o gene alvo na base de dados do NCBI; número identificador do fator de transcrição; nome do fator de transcrição; informações sobre o fator de transcrição na base de dados do NCBI; técnicas utilizadas para a comprovação da interação entre os fatores

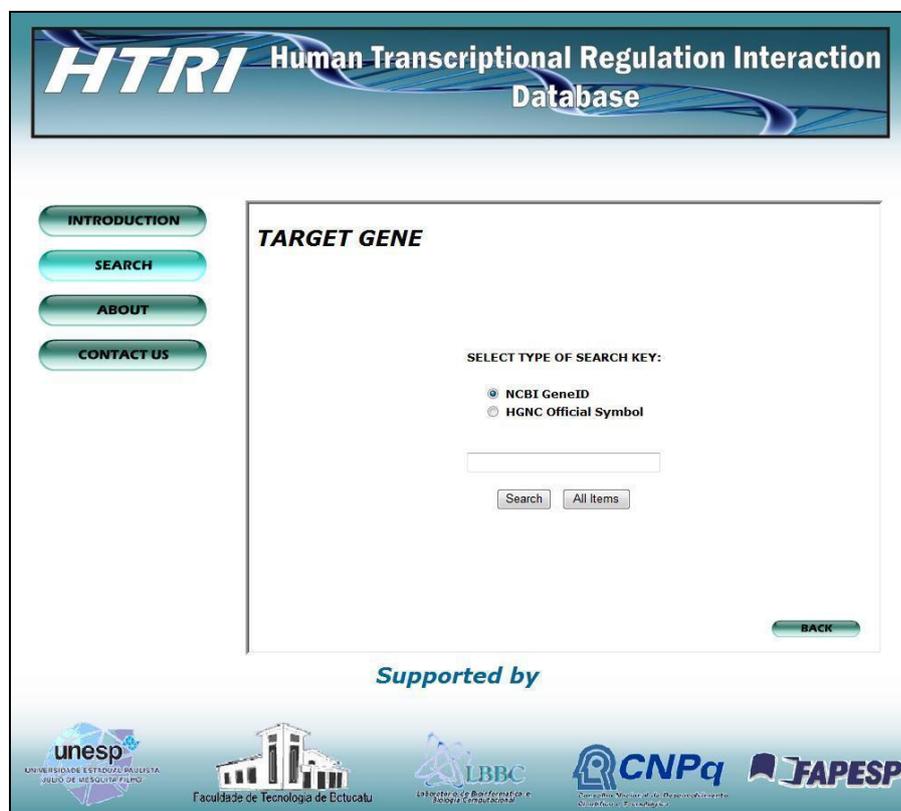


Figura 3. Página de pesquisa do HTRIdb baseado nos genes alvos.

de transcrição e os genes alvos; e as referências. Já na consulta por fatores de transcrição os dados são tabulados por: identificador do fator de transcrição; nome do fator de transcrição; informações sobre o fator de transcrição na base de dados do NCBI; número identificador do gene alvo; nome do gene alvo; mais informações sobre o gene alvo na base de dados do NCBI; técnicas utilizadas para a comprovação da interação entre os fatores de transcrição e os genes alvos; e as referências. A Figura 4 ilustra o resultado de uma consulta por fatores de transcrição.

O banco de dados desenvolvido neste trabalho oferece acesso simplificado aos dados armazenados, pois utiliza uma interface amigável e simples. Isto o diferencia dos demais bancos de dados existentes, como o TRANSFAC, que possui uma interface gráfica caracterizada pela grande poluição visual, o que dificulta o seu manuseio. Além disso, o TRANSFAC impõe a obrigatoriedade da geração de um cadastro, para que se possa ter direito ao acesso ao banco de dados, pois este direito de acesso é cobrado.

Em relação ao banco de dados

GeneID TF	Transcription Factor TF	More Info	GeneID TG	Target Gene TG	More Info	Techniques	Ref
1874	E2F4	NCBI	4613	MYCN	NCBI	in vivo footprinting	9
1874	E2F4	NCBI	4613	MYCN	NCBI	reporter assay	9
1874	E2F4	NCBI	7157	TP53	NCBI	chromatin immunoprecipitation	11
1958	EGR1	NCBI	7157	TP53	NCBI	Immunohistochemical analysis	11
2353	FOS	NCBI	969	CD69	NCBI	Electrophoretic mobility shift assay	12
2353	FOS	NCBI	4922	NTS	NCBI	gel shift	7
2353	FOS	NCBI	4922	NTS	NCBI	DNase I footprinting	7
2353	FOS	NCBI	6352	CCL5	NCBI	Electrophoretic mobility shift assay	15
2353	FOS	NCBI	7157	TP53	NCBI	gel shift competition	10
2353	FOS	NCBI	7157	TP53	NCBI	functional analysis	10
2353	FOS	NCBI	7157	TP53	NCBI	direct gel shift	10
2353	FOS	NCBI	7157	TP53	NCBI	supershift (antibody binding)	10
2354	FOSB	NCBI	6352	CCL5	NCBI	Electrophoretic mobility shift assay	15
3091	HIF1A	NCBI	7157	TP53	NCBI	Electrophoretic mobility shift assay	12
3725	JUN	NCBI	4852	NPY	NCBI	gel shift	2
3725	JUN	NCBI	4852	NPY	NCBI	DNase I footprinting	2
3725	JUN	NCBI	4852	NPY	NCBI	southwestern blotting	2
3725	JUN	NCBI	7031	TFF1	NCBI	Western blot	12
3725	JUN	NCBI	7031	TFF1	NCBI	chromatin immunoprecipitation	12

Figura 4. Página de resultado de uma pesquisa baseada no fator de transcrição.

TRED, o HTRIdb possui como vantagem uma maior quantidade de dados que serão inseridos sobre fatores de transcrição de seres humanos. Além disso, há também a facilidade de acesso a esses dados, pois o HTRIdb é direcionado apenas aos dados de fatores de transcrição em humanos. Ainda, os resultados das consultas são exibidos de maneira mais clara e simplificada ao usuário no banco de dados HTRIdb do que no banco de dados TRED.

6 CONCLUSÃO

Este trabalho permitiu o desenvolvimento de um novo banco de dados sobre transcrição gênica, que centraliza uma grande quantidade de informações e as disponibilizam de maneira gratuita, por meio de uma *interface web* simples de utilizar. Toda esta base 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, como os departamentos de P&D de Laboratórios Farmacêuticos. Permitindo que estes pesquisadores encontrem informações sobre fatores de transcrição em humanos de maneira mais eficiente. Promovendo, com isso, o aumento da produtividade em suas pesquisas e o desenvolvimento de novos trabalhos sobre os processos de transcrição gênica.

Tékhnē e Lógos, Botucatu, SP, v.1, n. 1, out. 2009

Possibilitando, também, um maior entendimento sobre a rede de fatores de transcrição 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 de transcrição gênica anormais.

7 REFERÊNCIAS

- ALBERTS, B. et al. *Fundamentos de biologia celular: uma introdução à biologia molecular da célula*. Porto Alegre, RS: Editora Artemed, 1999. 757p.
- BAXEVANIS, A.D. *Searching Online Mendelian Inheritance in Man (OMIM) for information for genetic loci involved in human disease*. *Curr Protoc Hum Genet.*; Chapter 9:Unit9.13, feb. 2003.
- BERG, J.M.; TYMOCZKO, J.L.; STRYER, L. *Biochemistry*. 5.ed. Nova Iorque, NY: W. H. Freeman and Company, 2002. 1120p.
- CAETANO, N. *Guia de remédios*. 4.ed. São Paulo, SP: Editora Escola, 1999. 400p.
- DATE, C.J. *Introdução a sistemas de banco de dados*. 7.ed. Rio de Janeiro, RJ: Editora Campus 2000. 803p.
- DAVIDSON, E.H. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. 1. ed. Nova Iorque, NY: Academic Press, 2006. 304p.

DEITEL, H.M.; DEITEL, P.J. *Java, como programar*. Trad. Furnankiewicz, E. 3.ed. Porto Alegre, RS: Editora Bookman, 2001. 1201p.

FIGUEIRA, J.S. Easy Java simulations - Modelagem computacional para o ensino de Física. *Revista Brasileira de Ensino de Física*, Pato Branco, PR, v.27, n.4, p.613-618, ago. 2005.

HALDER, G. et al. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science*. v.267, p.1788-1792, 1995.

KARAMOUZIS, M.V. et al. Transcription Factors and Neoplasia: Vistas in Novel Drug Design. *Clinical Cancer Research*. Greece, v.8, p.949-961, may 2002.

LIMA, V.R. *Desenvolvimento de um sistema baseado na Web para suporte ao programa de monitoração individual interna do ipen*. 2007. 97f. Dissertação (Mestrado)- Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo, São Paulo, SP, 2007.

MARUI, S. et al. Bases Genéticas dos Distúrbios de Crescimento. *Arq Bras Endocrinol Metab*. São Paulo, SP, v.46, n.4, 2002. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27302002000400016&lng=en&nrm=iso>. Acesso em: 4 set 2008.

MATYS, V. et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. Wolfenbüttel, Germany, v.34, *Database issue*, p.D108-D110, out. 2006.

MESSINA, D.N. et al. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res*. St. Louis, MO, v.14, n.10B, p.2041-2047, 2004.

TEMPLE, A. et al. *Programação Web com JSP, Servlets e J2EE*. Stanford, CA: Creative, 2004. 381p.

TYSIAK, M.; WINGENDER, E. *Biobase Biological Databases: Transfac*. Wolfenbuettel, Germany, 1999. Disponível em <<http://www.biobase-international.com/index.php?id=316>>. Acesso em: 2 set 2008.

SAYERS, E.W. *Database resources of the National Center for Biotechnology Information*. *Nucleic Acids Res*. oct. 2008.

SILBERSCHATZ, A. et al. *Sistema de Banco de Dados*. 3.ed. São Paulo, SP: Makron books, 1999. 808p.

WISHART, D.S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. v.34, *Database issue*, p.D668-D672, jan. 2006.