



## Identification of Novel "Pathologs" (Human Disease-Related Gene Candidates) From the RIKEN Full-Length Mouse cDNA Data Set

Diego G. Silva, Christian Schönbach, Vladimir Brusic, et al.

*Genome Res.* 2003 13: 1559

Access the most recent version at doi:[10.1101/gr.1461303](https://doi.org/10.1101/gr.1461303)

---

### License

#### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



The NEW Vortex Mixer



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

## Abstract

# Identification of Novel “Pathologs” (Human Disease-Related Gene Candidates) From the RIKEN Full-Length Mouse cDNA Data Set

Diego G. Silva,<sup>1,2</sup> Christian Schönbach,<sup>3</sup> Vladimir Brusic,<sup>4</sup> Luis A. Socha,<sup>1,2</sup> Takeshi Nagashima,<sup>3,5</sup> RIKEN GER Group<sup>6</sup> and GSL Members,<sup>7,8</sup> and Nikolai Petrovsky<sup>1,2,9</sup>

<sup>1</sup>Medical Informatics Centre, University of Canberra, ACT 2601 Australia; <sup>2</sup>John Curtin School of Medical Research, Australian National University, Canberra ACT 2601, Australia; <sup>3</sup>Biomedical Knowledge Discovery Team, Bioinformatics Group, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan; <sup>4</sup>Laboratories for Information Technology, Singapore 11961; <sup>5</sup>Department of Knowledge Systems Science, Japan Institute of Science and Technology, Ishikawa, 923-1292, Japan; <sup>6</sup>Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; <sup>7</sup>Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan

The majority of common diseases such as cancer, allergy, diabetes, or heart disease are characterized by complex genetic traits, in which genetic and environmental components contribute to disease susceptibility. Our knowledge of the genetic factors underlying most of such diseases is limited. A major goal in the post-genomic era is to identify and characterize disease susceptibility genes and to use this knowledge for disease treatment and prevention. More than 500 genes are conserved across the invertebrate and vertebrate genomes. Because of gene conservation, various organisms including yeast, fruitfly, zebrafish, rat, and mouse have been used as genetic models for the study of human disease. The basic housekeeping genes such as those involved in metabolism, intracellular signalling, transcription/translation, DNA replication, and repair are highly conserved in eukaryotes, and yeast and fruitfly are useful, therefore, for the study of basic cellular processes and related diseases. However, these organisms do not share with humans large groups of genes, such as those involved in homeostasis, immunity, and cellular interactions. Rodents and humans have remarkably similar genomes and share closely related biochemical, physiological, and pathological pathways. The comparison of the human genome with the FANTOM1 mouse cDNA clone set showed that ~80% of mouse cDNA clones have matches in the human genome. In this work, we define the term patholog to mean a mouse gene with sequence similarity to a known human disease-related gene. Previous genome-wide studies of pathologs have largely focused on diseases exhibiting Mendelian inheritance patterns. In this work, we have expanded the analysis to all potential pathologs regardless of their pattern of inheritance. A bioinformatic analysis and human curation of 60,770 RIKEN full-length mouse cDNA clones produced 2622 sequences that showed similarity (70%–85% identity) to known human-disease genes or proteins. Using automated computational tools in parallel with human expert analysis of 33207 MEDLINE scientific abstracts, we identified 184 novel mRNA transcripts (targets) with sequence similarity to genes encoding proteins reported as disease-related in humans (reference proteins). Of these targets, 36 were identified by computational tools only, 49 by a human expert analysis only, and 99 by both methods. The reference proteins related to cancer (53%), hereditary (23%), immunological (5%), cardio-vascular (4%), or other (15%), disorders. The role of these candidate pathologs in disease pathogenesis will require further characterization. It is likely that at least some of these potential pathologs will not be confirmed experimentally because, for example, they represent nonfunctional transcripts or gene products with sequence similarity, but different function. Those pathologs that are experimentally validated as functionally relevant will be used as targets for genetic manipulation and development of mouse models of human disease. The similarity between mouse and human genomes and their closely related biochemical, physiological, and pathological pathways makes the mouse an invaluable model organism for the study of human disease.

<sup>8</sup>Takahiro Arakawa,<sup>6</sup> Piero Carninci,<sup>6,7</sup> Jun Kawai,<sup>6,7</sup> and Yoshihide Hayashizaki<sup>6,7</sup>

<sup>9</sup>Corresponding author.

E-MAIL: [nikolai.petrovsky@anu.edu.au](mailto:nikolai.petrovsky@anu.edu.au); FAX +61-2-62603372.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1461303>.