Published online 7 November 2011 *Nucleic Acids Research, 2012, Vol. 40, Database issue* ***D377–D380***
doi:10.1093/nar/gkr960

NucleaRDB: information system for nuclear receptors

Bas Vroling[1,2], David Thorne[3], Philip McDermott[3], Henk-Jan Joosten[2], Teresa K. Attwood[4], Steve Pettifer[3] and Gert Vriend[1,*]

[1]CMBI, NCMLS, Radboud University Nijmegen Medical Centre, Nijmegen, [2]Bio-Prodict, Dreijenplein 10, 6703 HB Wageningen, The Netherlands, [3]School of Computer Science and [4]Faculty of Life Sciences, University of Manchester, Manchester, M13 9PL, UK

Received September 6, 2011; Revised October 9, 2011; Accepted October 13, 2011

ABSTRACT

The NucleaRDB is a Molecular Class-Specific Information System that collects, combines, validates and disseminates large amounts of heterogeneous data on nuclear hormone receptors. It contains both experimental and computationally derived data. The data and knowledge present in the NucleaRDB can be accessed using a number of different interactive and programmatic methods and query systems. A nuclear hormone receptor-specific PDF reader interface is available that can integrate the contents of the NucleaRDB with full-text scientific articles. The NucleaRDB is freely available at http://www.receptors.org/nucleardb.

INTRODUCTION

Nuclear receptors (NRs) are ligand-inducible transcription factors that regulate processes, such as homeostasis, differentiation, embryonic development and organ physiology. A total of 49 human NRs have been identified (1). Their ligands are lipophilic compounds such as steroids, thyroid hormone, vitamin D3 and retinoids (2). The endogenous ligands are not yet known for 30% of the NRs (3). As nuclear receptors are involved in almost all aspects of human physiology and are implicated in many important diseases including cancer, diabetes and osteoporosis, understanding of these receptors has major implications for human biology and for the development of new drug treatments. Nuclear receptors are targets for pharmaceutical industries with similar importance (4), as the G protein-coupled receptors (GPCRs), ion channels and kinases.

Due to the increasing amounts of experimental and computational data buried in numerous databases and scientific articles, the task of extracting, combining and validating this data is becoming an increasingly large hurdle for the individual scientist. Databases that revolve around a single protein family can help researchers in using all data needed for their research, while relieving them of the onerous tasks related to the retrieval of many data from different sources (5).

The NucleaRDB is a data source that holds many different data types (Table 1) in a well organized and easily accessible form (6). The data are validated, internally consistent and updated regularly. The NucleaRDB provides access to the data via various interfaces, which depending on the users' needs, are suited either for automated access or interactive usage.

DATA CONTENTS

Primary data

The NucleaRDB contains three different primary data types: sequences, structures and mutations. Sequences and structures were updated as described previously (7). Mutation data was obtained from the Nuclear Receptor Mutation Database (8) and fully integrated in the NucleaRDB. In addition, a large body of mutations was extracted from literature by the software package MuteXt (9).

Computational data

A large and diverse collection of computationally generated data are present in the NucleaRDB. Multiple sequence alignments (MSAs) form the heart of the system and allow users to easily transfer information between different proteins. MSAs are available for all families and subfamilies, and can be viewed using JalView (10) or can be directly downloaded in a number of formats. MSAs were created as described previously (7).

Correlated mutation analyses (CMA) can be used to identify groups of residues that mutate in tandem. Residues that show correlated mutation behavior are

*To whom correspondence should be addressed. Tel: +31 (0)24 36 19390; Fax: +31 (0)24 36 19395; Email: vriend@cmbi.ru.nl

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

likely to be functionally related, and networks of those correlating residues indicate functional units (11). Correlation scores are available for all (sub-)families.

The entropy and variability for a position in a MSA can be an indicator of the evolutionary pressures exerted at that position (12). Entropy and variability scores are available in tabular form and via an interactive page displaying an integrated view via plots, tables and structure models.

In addition to the already large amount of structural information that is present in the NucleaRDB, homology models based on multiple template structures have been built for all NRs. All structure models were built using YASARA (13) and are available for download or can be viewed directly using Jmol (14).

INFORMATION RETRIEVAL

All data in the NucleaRDB web interface are extensively connected, allowing for easy navigation between different data types. The main way of accessing the NucleaRDB's contents is via the hierarchical family tree. For each family, users can access the individual receptors, multiple sequence alignments (and all derived data and analyses such as correlation scores and protein distance networks), mutations, structures and models (Figure 1). All pages contain links to all related data and information. Extensive search facilities are available, allowing the search for proteins, sequences, structures, families and mutations using various search criteria and filters. A BLAST service is available that allows users to run their own sequences against the NucleaRDB.

All data types and search facilities are accessible from the web pages as well as from the web service endpoints, allowing users to write workflows or in-house software that uses the NucleaRDB.

Annotating scientific literature

Utopia Documents (15,16) is a new PDF reader that offers unique opportunities to place information and knowledge in the context of scientific literature. We have integrated the NucleaRDB with the Utopia Documents PDF reader in such a way as to present to scientists, in a non-intrusive way, all NR-relevant data and information discussed in an

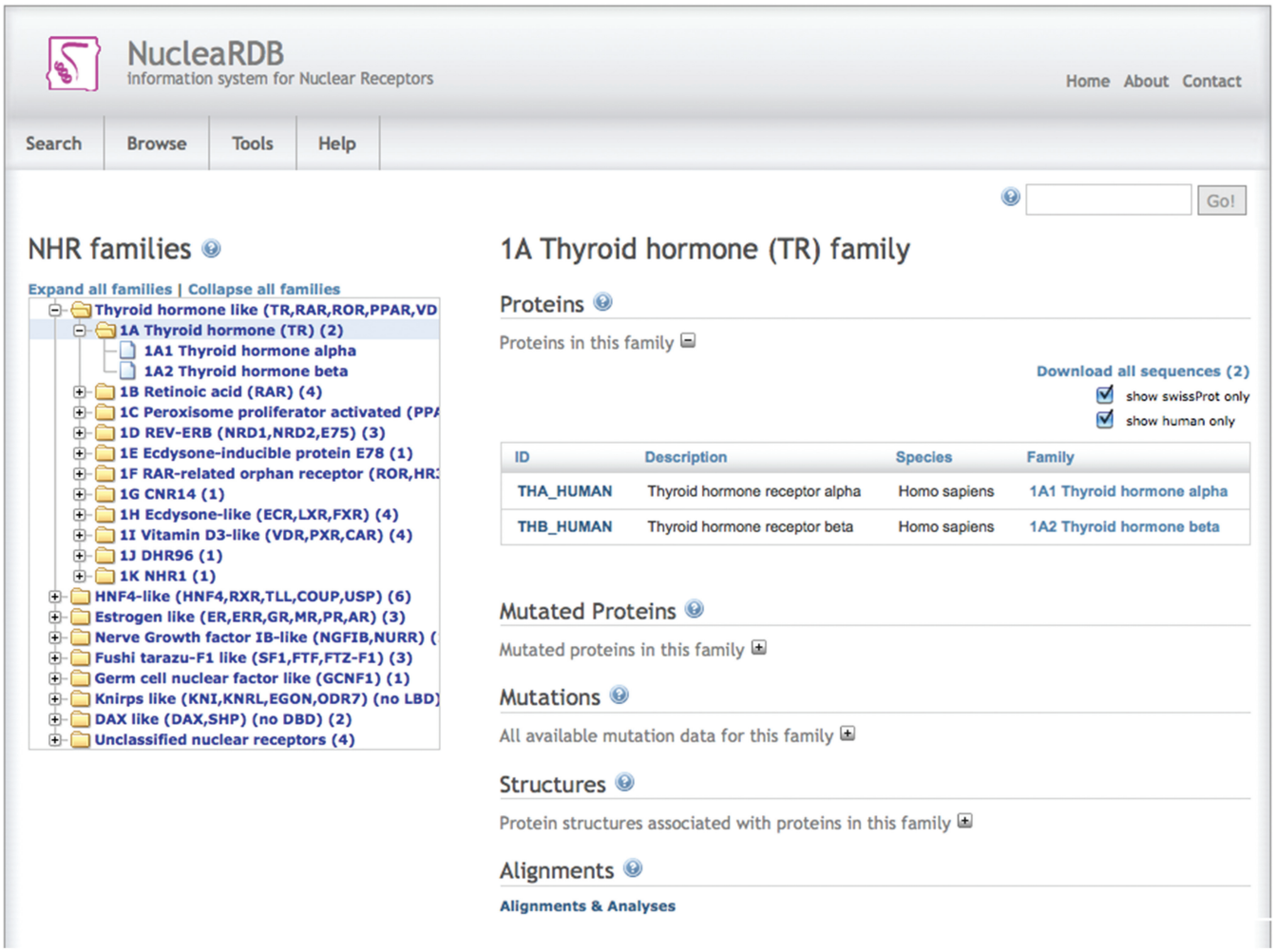

Figure 1. Screenshot of the NucleaRDB family page. The family tree is shown on the left with the thyroid hormone family expanded. On the right-hand side, the data for the selected family is shown.

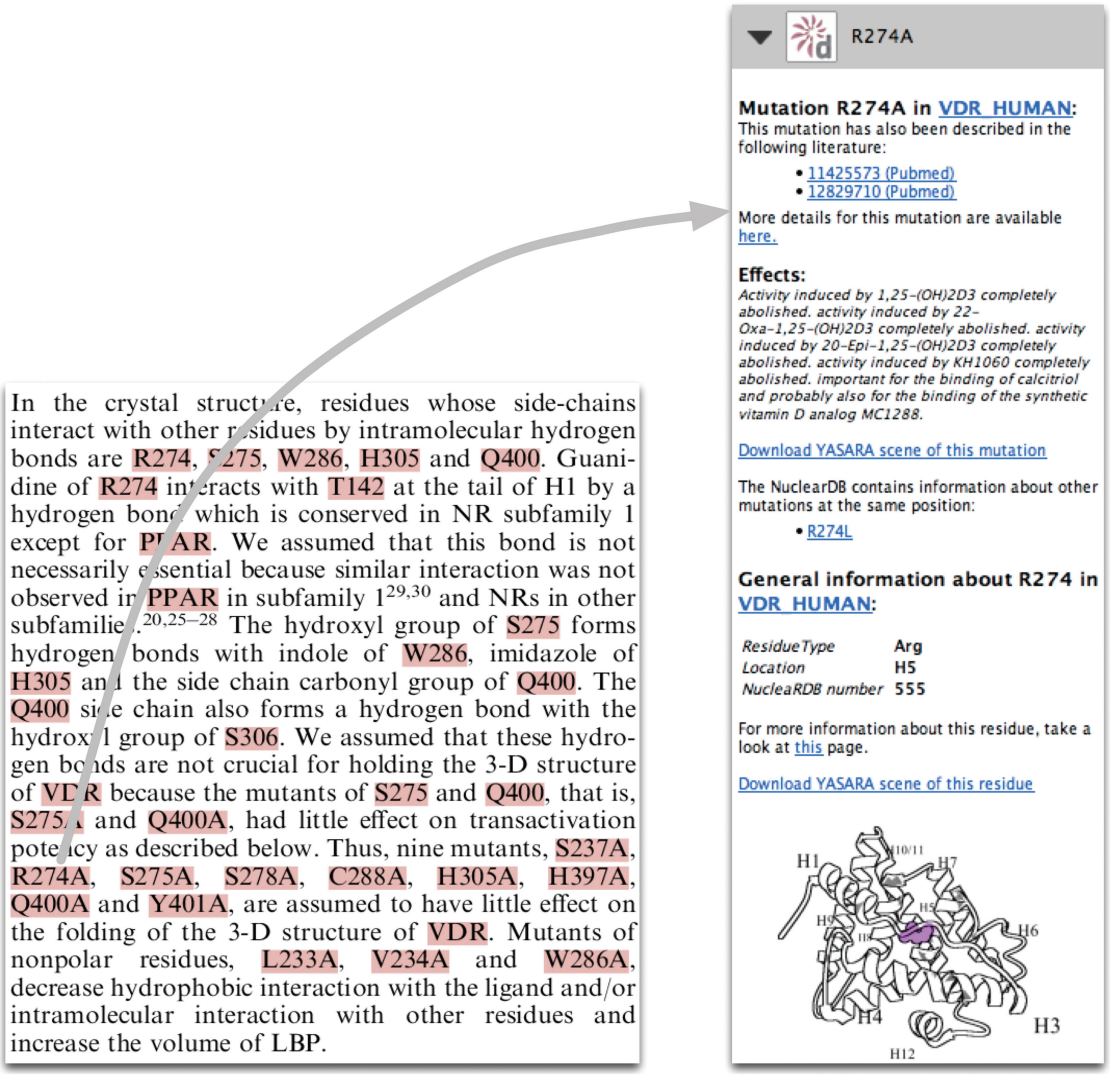

Figure 2. An impression of the Utopia Documents PDF reader interface to the NucleaRDB data. On the left-hand side a part of a scientific paper (17) is shown that is annotated by the NucleaRDB. Annotations are available for all the highlighted words. On the right-hand side an example of such an annotation (the mutation R274A) is displayed.

Table 1. Contents of the NucleaRDB

Proteins	3764
Families	123
Mutations	1543
Protein structures	613
Structure models	3764
Residues	2 012 651
Species	339

article at hand. Annotations are provided for proteins, residues and mutations mentioned in the PDF. For each of these concepts the annotations contain carefully selected information, as well as pointers to relevant web pages and related scientific literature. An example is shown in Figure 2. The PDF reader presents the scientist, in a non-intrusive way, all relevant data and information related to the topics discussed in the article. This alleviates the troubles associated with navigating the many links between existing data and information available from the many articles in this field. The scientist neither struggles to get access to information related to topics within an article, nor is swamped by unnecessary information that still needs disambiguation; only data and information relevant to the topic of the article is made available.

IMPLEMENTATION

The data in the NucleaRDB is stored in a PostgreSQL (www.postgresql.org) relational database. The web service interface is developed with the Apache CXF (cxf.apache.org) web services framework. We offer both Simple Object Access Protocol and Representational state transfer endpoints. The web interface is built using the Apache Wicket (wicket.apache.org) web application framework. The database is accessed via a Hibernate

(www.hibernate.org) object-relational mapping layer. The server is running within Sun's Glassfish (www.glassfish.org) application server.

CONCLUSION

The NucleaRDB provides researchers with a single point of access for nuclear receptor-related data. Not only does the NucleaRDB hold a large amount of information, it also provides a broad scope of tools and dissemination facilities, relieving scientist of many of the tasks that come with collecting, validating and integrating many diverse data.

ACKNOWLEDGEMENTS

We thank Maarten Hekkelman, Wilmar Teunissen and Tim teBeek for their support with computer science issues. We thank TIPharma for financial support.

FUNDING

BioRange program of the Netherlands Bioinformatics Centre (NBIC); BSIK grant through the Netherlands Genomics Initiative (NGI); EMBRACE project that is funded by the European Commission within its FP6 Programme, under the thematic area 'Life sciences, genomics and biotechnology for health' (contract number LHSG-CT-2004-512092); and TIPharma. Funding for open access charge: RUNMC.

Conflict of interest statement. None declared.

REFERENCES

1. Robinson-Rechavi,M., Carpentier,A.S., Duffraisse,M. and Laudet,V. (2001) How many nuclear hormone receptors are there in the human genome? *Trends Genet*., **17**, 554–556.
2. Mangelsdorf,D.J., Thummel,C., Beato,M., Herrlich,P., Schütz,G., Umesono,K., Blumberg,B., Kastner,P., Mark,M., Chambon,P. *et al*. (1995) The nuclear receptor superfamily: the second decade. *Cell*, **83**, 835–839.
3. Kliewer,S.A., Lehmann,J.M. and Willson,T.M. (1999) Orphan nuclear receptors: shifting endocrinology into reverse. *Science*, **284**, 757–760.
4. Hopkins,A.L. and Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov*., **1**, 727–730.
5. Folkertsma,S., van Noort,P., Van Durme,J., Joosten,H.-J., Bettler,E., Fleuren,W., Oliveira,L., Horn,F., de Vlieg,J. and Vriend,G. (2004) A family-based approach reveals the function of residues in the nuclear receptor ligand-binding domain. *J. Mol. Biol*., **341**, 321–335.
6. Horn,F., Vriend,G. and Cohen,F.E. (2001) Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic Acids Res*., **29**, 346–349.
7. Vroling,B., Sanders,M., Baakman,C., Borrmann,A., Verhoeven,S., Klomp,J., Oliveira,L., de Vlieg,J. and Vriend,G. (2011) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res*., **39**, D309–D319.
8. Van Durme,J.J.J., Bettler,E., Folkertsma,S., Horn,F. and Vriend,G. (2003) NRMD: Nuclear Receptor Mutation Database. *Nucleic Acids Res*., **31**, 331–333.
9. Horn,F., Lau,A.L. and Cohen,F.E. (2004) Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors. *Bioinformatics*, **20**, 557–568.
10. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
11. Oliveira,L., Paiva,A.C.M. and Vriend,G. (2002) Correlated mutation analyses on very large sequence families. *Chembiochem*, **3**, 1010–1017.
12. Ye,K., Lameijer,E.-W.M., Beukers,M.W. and Ijzerman,A.P. (2006) A two-entropies analysis to identify functional positions in the transmembrane region of class A G protein-coupled receptors. *Proteins*, **63**, 1018–1030.
13. Krieger,E., Joo,K., Lee,J., Lee,J., Raman,S., Thompson,J., Tyka,M., Baker,D. and Karplus,K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins*, **77(Suppl. 9)**, 114–122.
14. Herráez,A. (2002) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ*., **34**, 255–261.
15. Attwood,T.K., Kell,D.B., McDermott,P., Marsh,J., Pettifer,S.R. and Thorne,D. (2009) Calling international rescue: knowledge lost in literature and data landslide! *Biochem. J*., **424**, 317–333.
16. Attwood,T.K., Kell,D.B., McDermott,P., Marsh,J., Pettifer,S.R. and Thorne,D. (2010) Utopia documents: linking scholarly literature with research data. *Bioinformatics*, **26**, i568–i574.
17. Choi,M., Yamamoto,K., Masuno,H., Nakashima,K., Taga,T. and Yamada,S. (2001) Ligand recognition by the vitamin D receptor. *Bioorg. Med. Chem*., **9**, 1721–1730.