NPInter v2.0: an updated database of ncRNA interactions

Jiao Yuan1,2, Wei Wu1,2, Chaoyong Xie2,3, Guoguang Zhao3, Yi Zhao3,* and Runsheng Chen1,*

1Laboratory of Noncoding RNA, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, 2University of Chinese Academy of Sciences, Beijing 100049, China and 3Bioinformatics Research Group, Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Received August 15, 2013; Revised September 29, 2013; Accepted October 12, 2013

ABSTRACT

NPInter (http://www.bioinfo.org/NPInter) is a database that integrates experimentally verified functional interactions between noncoding RNAs (excluding tRNAs and rRNAs) and other biomolecules (proteins, RNAs and genomic DNAs). Extensive studies on ncRNA interactions have shown that ncRNAs could act as part of enzymatic or structural complexes, gene regulators or other functional elements. With the development of high-throughput biotechnology, such as cross-linking immunoprecipitation and high-throughput sequencing (CLIP-seq), the number of known ncRNA interactions, especially those formed by protein binding, has grown rapidly in recent years. In this work, we updated NPInter to version 2.0 by collecting ncRNA interactions from recent literature and related databases, expanding the number of entries to 201,107 covering 18 species. In addition, NPInter v2.0 incorporated a service for the BLAST alignment search as well as visualization of interactions.

INTRODUCTION

Interactions of RNA with other biomolecules are fundamental in cellular processes. In particular, noncoding RNA–protein interactions play important roles in protein synthesis (1,2), gene expression (3,4), RNA processing (5–7), developmental regulation (8,9), etc. Interactions between ncRNAs and proteins can be either direct or indirect. Some IncRNAs function as transcriptional regulators through direct association with transcription factors (10–12), while others indirectly interact with proteins in cooperation with genomic DNAs (13). The functional importance of many ncRNA–protein interactions for correct transcriptional regulation has been demonstrated (14–19), suggesting wide-ranging effects of ncRNA–protein interaction. In addition to targeting or being targeted by proteins or protein-coding transcripts, ncRNAs could also potentially target other ncRNAs, resulting in a layer of regulatory interactions between noncoding RNA classes (11,20). Consequently, cataloguing interactions of ncRNAs and other biomolecules is significant for gaining insight into biological processes and understanding the mechanism by which ncRNAs carry out their regulatory function.

Given the importance of ncRNA interactions in various pathways, we embarked on a project to build a comprehensive catalogue of such data and established the NPInter database (21). A large amount of new research has led to deeper insight into ncRNA interactions at a variety of levels; thus, NPInter has been updated to v2.0 to accommodate for such expanding data resources.

Following data collection as described below, noncoding molecules in NPInter were automatically filtered and assigned identifiers from NONCODE (22) or miRBase (23), while protein-coding molecules were assigned identifiers from UniProt (24), RefSeq (25) or UniGene (26). With the exception of rRNAs and tRNAs, all reported interactions of ncRNAs were included. The aim of the database is to provide a platform that will facilitate both bioinformatic as well as experimental research. In addition to a user-friendly interface and a convenient search option that allow efficient recovery of related interactions and other information, NPInter v2.0 also provides a visualization platform for related interactions. Whole datasets can be

*To whom correspondence should be addressed. Tel: +86 10 6488 8543; Fax: +86 10 6487 7837; Email: crs@sun5.ibp.ac.cn
Correspondence may also be addressed to Yi Zhao. Tel: +86 106 260 1010; Fax: +86 106 260 1356; Email: biozy@ict.ac.cn

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© Crown copyright 2013.
downloaded and search results can be exported in text format.

DATA COLLECTION AND ANNOTATION

Based on the first version of NPInter database, new datasets from literature and other related databases were collected. In this update, data from literature accounts for the major source. We first retrieved literature published in the last 5 years from PubMed, employing key words including ‘CLIP interaction’, ‘non-coding RNA bind protein post-transcriptional’, ‘rRNA–rRNA interaction’, ‘lncRNA protein interaction’, ‘lncRNA bind’, ‘RNA protein cross linking’, ‘RIP non-coding RNA’, etc, and found 1270 relevant articles. Information on reported interactions, either verified by experiments or derived from binding sites generated by sequencing, was manually extracted. For the latter, we only extracted processed data by authors, rather than raw sequencing data. Binding sites were first compared with RefSeq coding genes to eliminate those located within coding regions and then screened against NONCODE, which serves as an ncRNA reference database. Binding sites that lie within NONCODE ncRNAs were retained and assigned NONCODE IDs, while others were discarded. As interaction partners, proteins, protein-coding RNAs and DNA were assigned UniProt IDs, RefSeq IDs and UniGene IDs, respectively, while other noncoding RNAs and DNA were still assigned NONCODE IDs or miRBase IDs. We also integrated data from external resources, mainly LncRNA Disease (27), which curated 478 experimentally supported lncRNA interactions at various levels, including binding, regulation and co-expression. Molecules from LncRNA Disease were subjected to the same annotation pipeline. By manual curation on data from literature and integration of data from LncRNA Disease, NPInter v2.0 provides a multilevel snapshot of the interactome. A process of redundancy elimination was then performed on the whole dataset, including previously existing data and newly collected data as mentioned above. Redundant interactions were aggregated into a single record. The workflow used for the generation of NPInter is schematically shown in Figure 1.

DATABASE CONTENT AND STRUCTURE

The purpose of the database is to serve as a knowledge base for experimentally-oriented studies and as a resource for bioinformatics applications. The first release of NPInter in 2006 contained 700 published functional

---

**Figure 1.** Overview of the NPInter v2.0 Database. NcRNA interactions were obtained from literature and specialized databases. RNA binding sites of individual proteins identified using genome-wide techniques were screened against NONCODE and only those correspond to NONCODE RNAs were extracted.
interactions from six model organisms. NPInter v2.0 presently contains 201,107 ncRNA interactions distributed on 18 organisms, collected from 529 published articles. The significant growth in the amount of data is primarily because of systematic identification of protein binding sites on the transcriptome through a combination of CLIP and RNA sequencing, whereas other interactions, including ncRNA–RNA interactions and transcript factor (TF)–ncRNA interactions, were obtained mainly from interaction studies on individual ncRNA.

The basic information on each interaction contains an interaction ID, name of ncRNA and its interaction partner, organism in which the interaction was identified, level and class of interaction manually defined in NPInter, as well as tags manually added. The interaction partners can be DNA, RNA or protein. For example, at the binding level, which accounts for most of the NPInter datasets, ncRNAs may interact with DNA (13,28), ncRNA (29), miRNA (30), mRNA (31) and protein (32). At the level of indirect interactions, ncRNAs may either regulate or be regulated by an interaction partner (33,34). The level of interaction defined in NPInter represents the types of interacting molecules and characteristics of the interaction, including ‘RNA–protein’, ‘RNA–RNA’ and ‘DNA–TF’. In general, NPInter classifies all interactions into three classes, which are ‘binding’, ‘regulatory’ and ‘co-expression’. Tags are added to give brief introductions to each interaction, suggesting in which biological process the interaction participates. Our tags are ‘expression correlation’, ‘genetic interaction’, ‘genomic location related’, ‘indirect’, ‘miRNA’, ‘miRNA target interaction’, ‘ncRNA affects synthesis or function of protein’, ‘ncRNA is regulated’, ‘ncRNA–protein binding’, ‘expression, processing, or function of ncRNA is affected’, ‘ncRNA targets mRNA’, ‘other linkages’, ‘promoter as action site’, ‘regulatory’ and ‘RNA–RNA interaction’. Each interaction can be labelled simultaneously with several tags. (See Supplementary Data: Supplementary I).

The NPInter database is composed of three linked tables:

- The Interaction table gives a description of details of the interaction between two molecules. For example, for the interaction between Kcnq1ot1 and Dnmt1 in mouse, information in this table states that this interaction occurs at the level of ‘RNA–protein’ and is classified into ‘binding’ class with a tag ‘ncRNA–protein binding’.
- The Molecule table describes information on interacting molecules, containing identifiers from UniGene,

![Figure 2](http://nar.oxfordjournals.org/)

**Figure 2.** Graph visualization of NPInter interactions. Node colour and size indicates molecule type, while edge colour indicates interaction source. The initial placement of the nodes is determined by a ‘force-directed’ layout algorithm that aims to keep the more similar nodes closer together, but the placement may later be adjusted by the user.
Noncoding RNAs have emerged as key molecular players in different biological processes (37). Characterizing ncRNA interactions will largely contribute towards the discovery of novel ncRNA functions. However, existing databases are not enough to provide a comprehensive resource for such a purpose. For example, starBase (38) lists all binding sites of several proteins in the transcriptome, but there is no specific annotation for noncoding transcripts. For another example, NPIDB (39) focuses on nucleic acid-protein complexes from PDB, thus, most RNAs included are either rRNAs or tRNAs.

In contrast, NPInter is more informative on ncRNA interactions than other databases, with aims on integrating interactions of the other types of ncRNAs that serve a variety of functions. In other words, NPInter is one of the most comprehensive databases of interactions between ncRNAs and other biomolecules.

Compared with the previous version of NPInter, the new version is a step towards a more integrated knowledge database. The total number of functional interactions of ncRNAs has been expanded. NPInter v2.0 is also developed to present graphical molecular interaction networks that will enable biological scientists to explore their data in a more systems-oriented manner. NPInter will continue to keep track of and promptly collect new interactions. The authors plan to update NPInter whenever there is an accumulation of novel ncRNA interactions reported in literature or other sources.

It is worth mentioning that NPInter is included in our systematic platform for noncoding RNAs. Our platform consists of ncRNA resources such as NONCODE and NPInter, as well as online tools and web servers (40–43) including CNCl (40) and ncFANs (43) for analysis of ncRNAs. As a member of the union of specific databases and tools for noncoding RNAs, NPInter is expected to remain an informative and valuable data source on the biological roles of ncRNAs for the scientific community.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank Andrew Plygawko and Jianjun Luo for carefully reading our manuscript.

FUNDING
Funding for open access charge: National High-tech Research and Development Projects 863 [2012AA020402]; Training Program of the Major Research plan of the National Natural Science Foundation of China [91229120]; National Natural Science Foundation of China [31000586]; Chinese Academy of Science Strategic Project of Leading Science and Technology [XDA01020402]; National Key Basic Research and Development Program 973 [2009CB825401]; the National Center for Mathematics and Interdisciplinary Sciences.

Conflict of interest statement. None declared.

REFERENCES


