D218–D220 *Nucleic Acids Research, 2006, Vol. 34, Database issue*
doi:10.1093/nar/gkj004

NOPdb: Nucleolar Proteome Database

Anthony Kar Lun Leung*, Laura Trinkle-Mulcahy, Yun Wah Lam, Jens S. Andersen[1], Matthias Mann[1] and Angus I. Lamond

Division of Gene Regulation and Expression, Wellcome Trust Biocentre, School of Life Sciences, University of Dundee, Dundee DD1 5EH, UK and [1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Received August 13, 2005; Accepted August 16, 2005

ABSTRACT

The Nucleolar Proteome Database (NOPdb) archives data on >700 proteins that were identified by multiple mass spectrometry (MS) analyses from highly purified preparations of human nucleoli, the most prominent nuclear organelle. Each protein entry is annotated with information about its corresponding gene, its domain structures and relevant protein homologues across species, as well as documenting its MS identification history including all the peptides sequenced by tandem MS/MS. Moreover, data showing the quantitative changes in the relative levels of ~500 nucleolar proteins are compared at different timepoints upon transcriptional inhibition. Correlating changes in protein abundance at multiple timepoints, highlighted by visualization means in the NOPdb, provides clues regarding the potential interactions and relationships between nucleolar proteins and thereby suggests putative functions for factors within the 30% of the proteome which comprises novel/uncharacterized proteins. The NOPdb (http://www.lamondlab.com/NOPdb) is searchable by either gene names, nucleotide or protein sequences, Gene Ontology terms or motifs, or by limiting the range for isoelectric points and/or molecular weights and links to other databases (e.g. LocusLink, OMIM and PubMed).

INTRODUCTION

The nucleolus is the most prominent structure within the eukaryotic nucleus and is known for its role in ribosomal RNA (rRNA) transcription, processing and the subsequent assembly of processed rRNA with ribosomal proteins to form ribosomal subunits (1–3). Recent studies suggested that the mammalian nucleolus may also play roles in tumourigenesis (4), viral replication (5) and cellular stress responses (6). However, the pathway and the identities of the molecular machineries involved in these mechanisms within this nuclear organelle remained largely unknown. Due to its inherent high density, nucleoli from cultured human cells can be isolated readily from sonicated nuclear extracts (7). Taking advantage of this, we and others have previously employed mass spectrometry (MS) techniques to identify the protein components from highly purified nucleolar preparations (8–10). Furthermore, fluorescent protein-tagging experiments and photobleaching analyses have vividly demonstrated the dynamic nature of the nucleolar proteome, where proteins only accumulate in the nucleolus either under specific metabolic conditions, or at specific cell cycle stages (11). Recently, we have extended our MS analyses to measure the dynamic behaviour of the nucleolar proteome by quantitating the relative level of individual nucleolar components upon transcriptional inhibition using a method known as stable isotope labelling with amino acids in cell culture (SILAC) (12).

DATABASE ACCESS AND CONTENT

To facilitate the analysis of these quantitative proteomic data, we have established the Nucleolar Proteome Database (NOPdb), a database aiming to archive all the human nucleolar proteins identified by MS analyses so far (13). The current version 2.0 of the database is available at http://www.lamondlab.com/NOPdb/ and is searchable by gene name/symbol, protein sequence, motif (14–16), Gene Ontology (GO) terms (17) or by setting the range of the predicted isoelectric point and/or molecular weight (Figure 1). To date, NOPdb archives 728 human nucleolar proteins (covering ~2.5% of the predicted human proteome) verified by multiple MS analyses and documents the quantitative changes in protein levels for 498 of these proteins at multiple timepoints after transcription is inhibited by treating cells with Actinomycin D.

*To whom correspondence should be addressed. Tel: +1 617 253 0265; Fax: +1 617 253 3867; Email: akleung@mit.edu
Present address:
Anthony K. L. Leung, Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

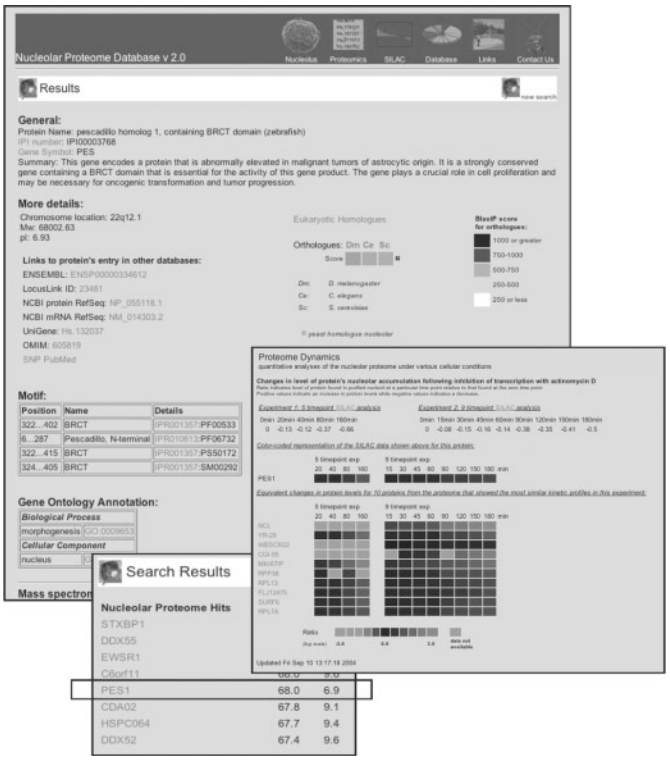

Figure 1. Snapshots of the NOPdb (http://www.lamondlab.com/NOPdb/). The database was searched against molecular weights between 65 and 70 kDa and here we show an overview page for the PES1 protein (pescadillo) documenting its motif distribution, its GO annotations, its identification history by multiple MS analysis and its quantitative data from SILAC analyses. Proteins of similar kinetic profiles based on correlation coefficient are identified for future investigation. The kinetic profiles are ranked according to the Pearson's correlation coefficients for the log value of the peak intensities of multiple peptides at a particular timepoint normalized to the respective peak intensities at zero timepoints.

The NOPdb provides (i) information on gene sequences and chromosomal localization, (ii) information on primary protein sequence (including protein sequence, predicted isoelectric point and molecular weight and motif structure) and (iii) information about putative nucleolar protein homologues in fruitfly, nematode and yeast, and also their localization data in these species, if available (18,19). A dedicated section for MS data has included the identification history of these nucleolar proteins in multiple MS analyses, peptide sequences deduced by tandem MS and the details of the MS experiments. Functions of these proteins are described using GO terms and detailed comments manually curated in the Entrez Gene database (20). In addition, the NOPdb also acts as a gateway to other databases, including NCBI LocusLink (20), OMIM (21), PubMed (9), UniGene (20) and ENSEMBL (22).

ACCESS TO PROTEOME DYNAMICS

A general problem experienced in proteome analyses is the abundance of novel/uncharacterized proteins (~30% in the case of the nucleolus) where limited information is available regarding their function (9,13). Therefore, the availability of quantitative information allows for the first time the ability to annotate/classify the proteome according to the changes in individual protein levels at multiple timepoints upon drug treatment. Analogous to the gene expression profiles generated for microarray data (23), we used SILAC data to generate a unique kinetic profile over time for each protein, where the relative abundance of each protein is compared with its respective level at the initial timepoint. Unlike microarray data, the quantitative measurements are made at the post-transcriptional level. The changes in the levels of protein in the nucleolus after drug treatment likely reflect their respective functional roles. Moreover, proteins with similar kinetic profiles based on Pearson's correlation coefficients can be identified, through the visualization means in the NOPdb, where available. This information makes direct predictions that can subsequently be tested both *in vivo* and *in vitro*.

PERSPECTIVES

Future versions of the NOPdb will include additional kinetic profiles for each protein, based on their responses to both different drug treatments and other metabolic and cell cycle variations. Clustering of such data may offer useful information for predicting the potential functions of these novel proteins (24). Apart from shedding light to the functions of novel proteins, clustered protein groups can be served as refined sets for motif search. Bioinformatic tools will also be developed to provide means to interact with the related microarray data deposited in the public domain. Comparison of these profiles with gene expression profiles from parallel microarray data may yield fresh understanding of the post-transcriptional regulation of the corresponding genes. Current analyses on the primary sequences deposited in the NOPdb determined a number of properties of the nucleolar proteome in terms of the distribution of amino acid/short peptide composition (13), domain structure and GO terms (Supplementary Tables 1 and 2), which are statistically different from the profiles of proteins accumulated within other cellular structures or organelles. In summary, the NOPdb provides a useful resource for the scientific community to explore the plurifunctionality of nucleolus, where further surprises are probably still in store.

SUPPLEMENTARY DATA

Supplementary Data is available at NAR Online.

ACKNOWLEDGEMENTS

A.K.L.L. was supported by a Croucher Foundation Scholarship. A.I.L. is a Wellcome Trust Principal Research Fellow. The Human Frontier Science Program is acknowledged for a research grant entitled 'Functional organization of the cell nucleus investigated through proteomics and molecular dynamics'. The work in the Lamond laboratory is supported by the Wellcome Trust and work in the Mann laboratory is funded by a Danish National Research Foundation grant to the Centre for Experimental Bioinformatics. Funding to pay the Open Access publication charges for this article was provided by Joint Information Systems Committee of the UK.

Conflict of interest statement. None declared.

REFERENCES

1. Leary,D.J. and Huang,S. (2001) Regulation of ribosome biogenesis within the nucleolus. *FEBS Lett*., **509**, 145–150.
2. Tschochner,H. and Hurt,E. (2003) Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends Cell Biol*., **13**, 255–263.
3. Pederson,T. (1998) The plurifunctional nucleolus. *Nucleic Acids Res*., **26**, 3871–3876.
4. Ruggero,D. and Pandolfi,P.P. (2003) Does the ribosome translate cancer? *Nature Rev. Cancer*, **3**, 179–192.
5. Hiscox,J.A. (2002) The nucleolus—a gateway to viral infection? *Arch. Virol*., **147**, 1077–1089.
6. Olson,M.O. (2004) Sensing cellular stress: another new function for the nucleolus? *Sci. STKE*, **2004**, pe10.
7. Busch,H., Muramatsu,M., Adams,H., Steele,W.J., Liau,M.C. and Smetana,K. (1963) Isolation of nucleoli. *Exp. Cell Res*., **24** (Suppl 9), 150–163.
8. Scherl,A., Coute,Y., Deon,C., Calle,A., Kindbeiter,K., Sanchez,J.C., Greco,A., Hochstrasser,D. and Diaz,J.J. (2002) Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell*, **13**, 4100–4109.
9. Andersen,J.S., Lyon,C.E., Fox,A.H., Leung,A.K., Lam,Y.W., Steen,H., Mann,M. and Lamond,A.I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol*., **12**, 1–11.
10. Andersen,J.S., Lam,Y.W., Leung,A.K., Ong,S.E., Lyon,C.E., Lamond,A.I. and Mann,M. (2005) Nucleolar proteome dynamics. *Nature*, **433**, 77–83.
11. Leung,A.K. and Lamond,A.I. (2003) The dynamics of the nucleolus. *Crit. Rev. Eukaryot. Gene Expr*., **13**, 39–54.
12. Ong,S.E., Blagoev,B., Kratchmarova,I., Kristensen,D.B., Steen,H., Pandey,A. and Mann,M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics*, **1**, 376–386.
13. Leung,A.K., Andersen,J.S., Mann,M. and Lamond,A.I. (2003) Bioinformatic analysis of the nucleolus. *Biochem. J*., **376**, 553–569.
14. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Barrell,D., Bateman,A., Binns,D., Biswas,M., Bradley,P., Bork,P. *et al*. (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res*., **31**, 315–318.
15. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L. *et al*. (2004) The Pfam protein families database. *Nucleic Acids Res*., **32**, D138–D141.
16. Letunic,I., Copley,R.R., Schmidt,S., Ciccarelli,F.D., Doerks,T., Schultz,J., Ponting,C.P. and Bork,P. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res*., **32**, D142–D144.
17. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet*., **25**, 25–29.
18. Huh,W.K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
19. Mewes,H.W., Amid,C., Arnold,R., Frishman,D., Guldener,U., Mannhaupt,G., Munsterkotter,M., Pagel,P., Strack,N., Stumpflen,V. *et al*. (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res*., **32**, D41–D44.
20. Wheeler,D.L., Church,D.M., Edgar,R., Federhen,S., Helmberg,W., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E. *et al*. (2004) Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res*., **32**, D35–D40.
21. Hamosh,A., Scott,A.F., Amberger,J., Bocchini,C., Valle,D. and McKusick,V.A. (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*., **30**, 52–55.
22. Birney,E., Andrews,D., Bevan,P., Caccamo,M., Cameron,G., Chen,Y., Clarke,L., Coates,G., Cox,T., Cuff,J. *et al*. (2004) Ensembl 2004. *Nucleic Acids Res*., **32**, D468–D470.
23. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
24. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.