

TRANS PROF DB: A new resource for sharing translational profiles

Ruth V Spriggs* and Anne E Willis

MRC Toxicology Unit; Leicester, UK

Keywords: archive, database, hand-curated, meta-analyses, resource, polysome profile, translational profile

Abbreviations and acronyms: TRANS PROF DB, Translational Profiles Database

The translational efficiency of individual mRNAs can be measured on a genome-wide scale using translational profiling techniques. Data from such experiments are an enormously important resource in the quest to understand the impact of cellular state on gene expression. To improve our understanding of these data, we have created TRANS PROF DB, a manually curated resource containing the translational status of human mRNAs under defined conditions. Results are provided at the level of an annotated conclusion for each gene, e.g. "Translation up-regulated", and also, where available, in a rawer form such as the normalized analyzed microarray output. TRANS PROF DB aims to provide a central resource for the sharing of translational profiles to facilitate reuse of published data and to enable meta-analyses across data sets. As the database expands, it will provide an easily searchable archive of publicly available translational profiling data sets. We encourage all researchers to deposit their translational profiling data into TRANS PROF DB to enable us to create a truly comprehensive resource. TRANS PROF DB is available without restriction at mrctools.mrctox.le.ac.uk/TRANS_PROF_DB.

Introduction

Two techniques have been developed to monitor changes in the translational efficiency of mRNAs on a genome-wide scale, polysome profiling¹ and ribosome profiling.² Both of these techniques are based on the premise that mRNAs that are highly translated are associated with many ribosomes (polysomes), whereas those that are not, or are poorly translated, are found associated with subpolysomes or mRNPs. Thus polysomes are separated from subpolysomes using sucrose density gradient centrifugation and the distribution of mRNAs between these fractions is analyzed using microarray technology,¹ or RNA-seq.² These profiling techniques have been used to examine the translatome under many different conditions including genotoxic stress,³ viral infection⁴ and during the cell cycle.⁵

Regardless of the cell conditions examined, translational profiling ultimately produces lists of genes whose translation increases, decreases, or stays the same under the condition being tested. Unfortunately, the gene lists created by these analyses are rarely compared to other lists generated in the same way from the same or different experimental conditions. Consequently, there is much potential co-regulatory information that is not being mined from data that already exist, and the many translational profiling data sets that have been created are underexploited. There is an urgent need for a central resource for translational

profiling data which will allow the gene lists from experiments performed by different research groups to be downloaded and compared against each other in meta-analyses.

To address this requirement, we have created TRANS PROF DB, a database and online interface designed for the organization, storage, interrogation, and dissemination of translational profiling data. TRANS PROF DB also aims to act as an archive of experimental results, indicating which translational profiles have already been created for your gene, cell line, or condition, of interest, and will reduce the amount of time researchers spend creating translational profiles when the required data are already available. Every gene in each data set stored in TRANS PROF DB is tagged with a 'conclusion', such as "translation up-regulated", to enable instant comparisons to be made; there is no requirement for the user to interrogate any raw data from the profile.

TRANS PROF DB (TRANS^lational PROF^{il}es Data^Base) is freely available with no login or password requirements at mrctools.mrctox.le.ac.uk/TRANS_PROF_DB.

Results

Database implementation

The central unit in the database is a data set, defined as one experimental condition on one chip; an experiment involving

*Correspondence to: Ruth V Spriggs; Email: rvs3@leicester.ac.uk
Submitted: 05/30/2014; Revised: 07/07/2014; Accepted: 09/11/2014
<http://dx.doi.org/10.4161/2169074X.2014.965615>

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. The moral rights of the named author(s) have been asserted.

multiple conditions or multiple chips would be entered into the database as multiple data sets. A condition is defined as the experimental condition under investigation, e.g. UVB irradiation. Attached to each data set are 2 major forms of data, Results and Conclusions. The conclusions are given per gene and clearly describe whether the translation of that gene is up-regulated, down-regulated, or unchanged. The results contain the underlying data used to decide the conclusion for each gene, which is usually given for each clone/probe on the microarray chip. Examples of result types are the ratio of subpolysomal intensity to polysomal intensity, or the fold change in this ratio between control group and 'treated' group.

The database currently contains 7 polysome profiling data sets (from 4 independent research groups), from 3 different cell lines, under 6 different conditions, covering 7175 genes. The Home page lists the data sets that are currently coming through the data submission pipeline, but cannot be released yet. This may be because a submitted paper using the data is still under review or because the owner of the data has not yet provided all necessary information. We intend to expand the database rapidly using previously published papers (where full data sets have been made available as supplementary material) and, as researchers become aware of the database, with data sets provided to us directly. The full potential of the database can only be realized as increasing numbers of data sets are included.

The interface to the database provides 2 ways to access the data, a Search tool and a Browse tool. The interface allows the user to find data from the perspective of gene, cell line or condition, or multiples thereof. This allows, for example, one gene to be interrogated across all cell lines and conditions at once, or the state of every gene under a particular condition to be viewed. Every output screen is designed to be downloaded to allow analysis and comparisons to continue using other tools.

Search tool

As the user begins typing a gene symbol (e.g. GAPDH), a cell line (e.g. HeLa) or a condition (e.g. apoptosis) into the Search box, available matching terms in the database appear in a drop down menu (see Fig. 1A). A term is then selected and the Search button pressed; all available results and conclusions for the search term will be displayed. Example search terms are provided; clicking on an example will pre-fill the Search box. The search is case-insensitive.

Browse tool

The browse tool enables the user to see at a glance what data are available. Initially columns show all conditions, all cell lines, all genes and all result types available in the database. Greyed out columns are inactive until at least one selection has been made in the column to the left. Selections made alter the available options in subsequent columns to avoid allowing a combination of selections which will yield no results.

To begin to browse, the user selects one or more options under the Conditions title. The cell lines that are relevant for the conditions chosen will then become available to select. Once cell lines are chosen, the relevant genes will become

available to select, and once genes have been selected, the result types column will become active. When at least one selection has been made in each column, the Display button will be activated (see Fig. 1B). When the Display button is pressed, all available results and/or conclusions for the selected options will be displayed. The search used to produce the displayed output uses the Boolean operator AND to combine search terms between each column, and OR to combine terms within a column. Please note, therefore, that if multiple selections are made in each column, there may not be results available for every combination of terms, e.g. when 2 conditions are selected, all genes with results in either condition will be made available to select, but specific genes may only appear in one of the selected data sets.

Output from tools

The Search and Browse tools produce an output page with the same format; a table of Conclusions and/or a table of Results, depending on your selections (see Fig. 1C); followed by tables that give more detail about the data sets and the methods used. If a Results table is displayed, the units used are defined below the table. Every output page also includes the names and affiliations of the scientists responsible for creating the data as well as the details of any publications arising from it; a clickable link is also provided to the relevant PubMed entry.

If a search term that appears in many data sets is chosen, e.g. the cell line HeLa, or if very many terms are combined when browsing, the output page will not load instantly; large amounts of data from several data sets are being retrieved and combined to produce the display.

The output from searching or browsing can be downloaded as a tab-separated text file by clicking on the button labeled "Download the tables above in a tab-separated text file" at the bottom of the page; a pop-up window will invite you to open or save the file. This file can be opened by, or loaded into, any spreadsheet-based software, such as Microsoft Excel, for further analysis. Complete data sets can be downloaded by highlighting all genes, and if required, all result types, in the Browse tool, or by searching for a condition in the Search tool.

Discussion

TRANS PROF DB aims to provide an essential resource to researchers who are creating and using translational profiles, providing a place to archive projects and an easy platform for the sharing and comparing of results. The simple search and browse tools provided enable any user to easily find out what data are available and to retrieve those data of interest to them. The value of the profile data in TRANS PROF DB is increased by the addition of a 'conclusion' tag on every gene. Currently, the range of data in TRANS PROF DB is limited – used to initially demonstrate the tools and the potential of the resource. Data sets will be continually incorporated into the database and made available through the interface; from our group, from published papers,

A

Home Search Browse Help

Enter a gene symbol, cell line or experimental condition:
(e.g. BRCA2, HeLa or UVB irradiation)

brc

BRCA2 | Gene symbol
BRCC3 | Gene symbol

Search

B

Home Search Browse Help

Select at least one option from each menu in turn to display the relevant database content:

Conditions: Apoptosis, Endoplasmic Reticulum stress, Hypoxia, Mitosis, Poliovirus infection, UVB irradiation

Cell lines: HeLa

Genes: BRAF, BRCA2, BRD2, BRF1, BSG, BST2, BTF3, BTG1, BTN2A1, RTN2A2

Result types: Conclusion, Log ratio (subs over polys), Rank, Rank difference

AND AND AND

Display Reset

C

Home Search Browse Help

Translational profiling data retrieved through browsing options:

Condition(s) UVB irradiation
Cell line(s) HeLa
Gene(s) BRCA2
Result type(s) Conclusion, Log ratio (subs over polys), Rank, Rank difference

CONCLUSIONS

Dataset ID	Condition	Cell line	Gene symbol	Conclusion
D00002	UVB irradiation	HeLa	BRCA2	No change in translation

RESULTS

Dataset ID	Condition	Cell line	Clone ID	Result	Units	Gene symbol
D00002	UVB irradiation	HeLa	Hs005358	-0.0821	4h_LogRatio	BRCA2
D00002	UVB irradiation	HeLa	Hs005358	0.2269	Control_LogRatio	BRCA2
D00002	UVB irradiation	HeLa	Hs005358	953	4h_rank	BRCA2
D00002	UVB irradiation	HeLa	Hs005358	1259	Control_rank	BRCA2
D00002	UVB irradiation	HeLa	Hs005358	306	Rank_Difference	BRCA2

Definitions of Units:
4h_LogRatio = "normalised data from all 'treated' microarrays, combined to give log₂[SubPolysome/Polysome] ratio; 4 hours after 'treatment'"
Control LogRatio = "normalised data from all control microarrays, combined to give log₂[SubPolysome/Polysome] ratio"

Figure 1. Search, Browse and Results pages. (A) Screen shot of the Search page; the user has started typing brca2, and has selected BRCA2 from the options that appeared; the Search button can now be pressed to retrieve all BRCA2 results from the database. (B) Screen shot of the Browse page; the user has selected UVB irradiation from the available Conditions, then HeLa from the available Cell lines, then BRCA2 from the available Genes, and finally all the available Results types; Display is now available to press. (C) Screen shot of the top portion of the results page after pressing Display on the Browse page in (B) above. The conclusions and results tables are shown, together with the first few lines of the definitions of the units used. Further down the results page there are more details about the data set, including the names of the scientists responsible and links to connected papers.

and from other researchers as knowledge and use of TRANS PROF DB increases. As more groups deposit their data with us, the more useful the resource will become.

Reuse of gene expression data is a growing trend,⁶ but the tools available are designed for comparing transcript levels rather than the translational efficiency information that comes from translational profiles. TRANS PROF DB differs from resources such as the Gene Expression Omnibus⁷ and Expression Atlas⁸ in that it collates gene changes at the level of translation. Gene Expression Omnibus (GEO) provides a tool called GEO2R for identifying differentially expressed genes from deposited data, but only within one GEO Series. The tool requires that you place the samples into groups, such as control and treated, across which the comparisons are made. Polysomal, subpolysomal or total RNA samples are treated like any other sample, allowing the user to see any changes between, e.g., the polysomal samples, but not to see any changes in the ratio of polysomal levels to subpolysomal levels directly. For example, series GSE20277⁹ has subpolysomal and polysomal samples, but these are treated as another experimental factor to compare. GEO also provides GEO profiles, that allows the user to search for all the expression profiles for a particular gene, and to filter these by keyword, etc., but again, any subpolysomal/polysomal samples are treated as separate conditions.

GEO also has data in curated “DataSets”, where biologically and statistically similar samples are grouped together, and a tool is provided for picking out genes that are up or down regulated. There are some translational profiling DataSets, e.g. GDS2408¹⁰ has total RNA and polysomal RNA samples which are highlighted as different protocols in the expression profiles, but no direct comparison of the ratio of total RNA to polysomal RNA is made. The Differential Atlas within Expression Atlas allows users to search for whether genes are up or down regulated under different conditions. It is based on an in-house reanalysis of RNA-seq and microarray experiments from ArrayExpress,¹¹ but only gene expression microarray data are used. In TRANS PROF DB our use of translational conclusions attached to each gene makes it simple to compare the translational state of a gene across different studies, cell lines and conditions. The processed array data which is also provided then enables more subtle comparisons to be made if necessary.

Planned extensions coming soon to TRANS PROF DB include to incorporate non-human data, and to add translational change annotations that come from the relatively new technique of ribosome profiling,² where ribosome-loaded mRNA are identified through sequencing. Both these additions would increase the power of meta-analyses as they would allow comparison across species and across complementary techniques. We will also very shortly be able to incorporate data extracted from tissues, rather than purely from cell lines.

The results provided in TRANS PROF DB will not only be of interest to researchers within the field of translation, but also to researchers in many other fields where translational profiling results can give an insight into a particular cellular state, for example, during apoptosis or in tumor-derived cell lines.

Materials and Methods

Data collection

For a detailed description of the polysome profiling and microarray analysis techniques used to generate translational profiles, please see Melamed & Arava.¹² For a laboratory protocol for performing polysome profiling please see the TRANS PROF DB website. Briefly (see Fig. 2); cycloheximide is used to arrest the elongation step of translation and fix the ribosomes in position on the mRNA, a sucrose gradient is used to separate the mRNA according to the number of ribosomes attached, RNA is purified from fractions of the gradient, and labelled cDNA is generated and hybridized to a microarray slide. Subsequent analysis of the slides from control and treated samples provides a measure of how the relative number of ribosomes attached to specific mRNAs has been affected by the treatment, and therefore, with assumptions, how the translation rate of specific mRNAs has changed. For a detailed protocol for the ribosome profiling

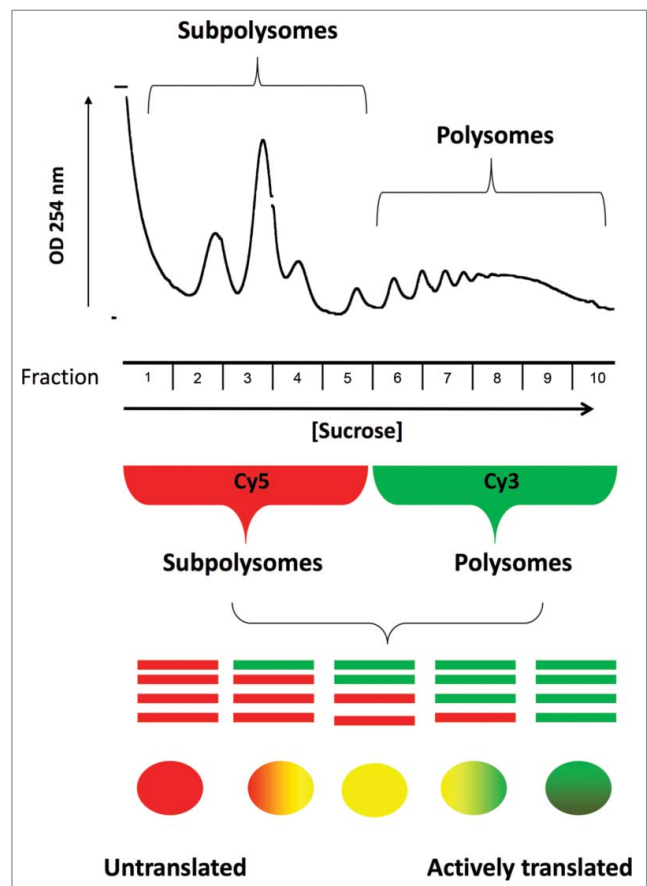


Figure 2. Polysome profiling. After arresting the elongation step of translation, cell extracts are applied to sucrose gradients to separate the mRNA according to the number of ribosomes attached. Fractions are pooled to create polysomal and subpolysomal samples. RNA is purified, then labelled cDNA is generated and hybridized to a microarray slide. Slides are scanned, and analysed to determine the abundance of each mRNA in the polysomal and subpolysomal samples. The ratio of polysomal abundance to subpolysomal abundance gives an indication of the translational status of each mRNA.

technique please see Ingolia.¹³ In brief; cells are treated with cycloheximide to arrest the ribosomes on the mRNAs, a post nuclear lysate is obtained and digested with RNase I before ribosome protected fragments (RPFs) are recovered by size selection and used to generate sequencing libraries. Total mRNA samples are extracted and sequenced at the same time. The ratio of RPFs to total mRNA can then be used to detect changes in the translational efficiency of individual genes in the condition under investigation, compared to the control.

Data curation

Data sets are provided by researchers, or are extracted from publications and any provided supplementary material. Data sets extracted from papers are generally smaller, and may only contain conclusions if no further processed data has been made available; the type of analysis performed and the type of results may still be described in the database, if they are described in the paper, to enable the values to be easily added if they become available to us. In all cases where data is extracted from papers, an attempt is made to contact the corresponding author to retrieve a complete data set.

Data input is a semi-automated process. To encourage the donation of data, the curator will accept data in any form. This is then hand-curated to a minimum form that can be accepted by an automated data-formatting program, which does all necessary integrity checks on the data before producing text files for import into the database. The data are stored in the database as they are provided, from the researcher or from the publication. No checks are made on the correctness of analyses, etc., that have been performed on the data. The only exception to this is the addition of official HUGO Gene Nomenclature Committee gene symbols¹⁴ to the annotations if these are not provided. Additional information for each data set, such as descriptions of the methods and of

the reasoning used to make conclusions about the translation of each gene, are manually curated. Usually this information is paraphrased from the associated paper or from knowledge provided by the depositor.

All research groups carrying out translational profiling are encouraged to deposit all their data, post-exploitation, into the database. Please contact the curator (transprofdb@le.ac.uk) or the authors if you would like to add your data to TRANS PROF DB. All contributions are fully acknowledged on the site whenever data are retrieved. TRANS PROF DB is hosted on a dedicated server within the MRC Toxicology Unit and will be maintained there in the long term. The curator will continue to add data sets to the database, and make them available online, for the foreseeable future.

Coding

The online interface is coded using HTML, PHP, and JavaScript, including the jQuery library (jquery.com), and makes use of CSS. The underlying relational database is coded in MySQL.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors wish to thank T. Beck for technical advice, and all the researchers whose data appear in TRANS PROF DB.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council [grant number BB/F02326X/2].

References

1. Spriggs KA, Stoneley M, Bushell M, Willis AE. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol Cell* 2008; 100:27-38; PMID:18072942; <http://dx.doi.org/10.1042/BC20070098>
2. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009; 324:218-23; PMID:19213877; <http://dx.doi.org/10.1126/science.1168978>
3. Powley IR, Kondrashov A, Young LA, Dobbyn HC, Hill K, Cannell IG, Stoneley M, Kong YW, Cotes JA, Smith GC, et al. Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes Dev* 2009; 23:1207-20; PMID:19451221; <http://dx.doi.org/10.1101/gad.516509>
4. Johannes G, Carter MS, Eisen MB, Brown PO, Sarnow P. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc Natl Acad Sci U S A* 1999; 96:13118-23; PMID:10557283; <http://dx.doi.org/10.1073/pnas.96.23.13118>
5. Qin XL, Sarnow P. Preferential translation of internal ribosome entry site-containing mRNAs during the mitotic cycle in mammalian cells. *J Biol Chem* 2004; 279:13721-8; PMID:14739278; <http://dx.doi.org/10.1074/jbc.M312854200>
6. Baker M. Gene data to hit milestone. *Nature* 2012; 487:282-3; PMID:22810669; <http://dx.doi.org/10.1038/487282a>
7. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013; 41:D991-5; PMID:23193258; <http://dx.doi.org/10.1093/nar/gks1193>
8. Petryszak R, Burdett T, Fiorelli B, Fonseca NA, Gonzalez-Porta M, Hastings E, Huber W, Jupp S, Keays M, Kryvych N, et al. Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res* 2014; 42:D926-32; PMID:24304889; <http://dx.doi.org/10.1093/nar/gkt1270>
9. Tebaldi T, Re A, Viero G, Pegoretti I, Passerini A, Blanzieri E, Quattrone A. Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC genomics* 2012; 13:220; PMID:22672192; <http://dx.doi.org/10.1186/1471-2164-13-220>
10. Patke A, Mecklenbrauker I, Erdjument-Bromage H, Tempst P, Tarakhovskiy A. BAF controls B cell metabolic fitness through a PKC beta- and Akt-dependent mechanism. *J Exp Med* 2006; 203:2551-62; PMID:17060474; <http://dx.doi.org/10.1084/jem.20060990>
11. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, et al. ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res* 2013; 41:D987-90; PMID:23193272; <http://dx.doi.org/10.1093/nar/gks1174>
12. Melamed D, Arava Y. Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. *Methods Enzymol* 2007; 431:177-201; PMID:17923236; [http://dx.doi.org/10.1016/S0076-6879\(07\)31010-0](http://dx.doi.org/10.1016/S0076-6879(07)31010-0)
13. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 2012; 7:1534-50; PMID:22836135; <http://dx.doi.org/10.1038/nprot.2012.086>
14. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res* 2013; 41:D545-52; PMID:23161694; <http://dx.doi.org/10.1093/nar/gks1066>