molecular systems biology

Landscape of nuclear transport receptor cargo specificity

Marie-Therese Mackmull, Bernd Klaus, Ivonne Heinze, Manopriya Chokkalingam, Andreas Beyer, Robert B Russell, Alessandro Ori, Martin Beck

Review timeline:

Submission date:	3 March 2017
Editorial Decision:	6 April 2017
Revision received:	18 August 2017
Editorial Decision:	5 October 2017
Revision received:	23 October 2017
Accepted:	10 November 2017

Editor: Thomas Lemberger

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 6 April 2017

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the referees who agreed to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, several important points on your work, which should be convincingly addressed in a major revision of the present work.

The reviewers find the data of high quality and the approach interesting. They are however suggesting several further analyses to deepen the insights gained from the study. While we do NOT think that it would be realistic to ask repeating the entire analysis in a second cell line, some additional verifications of the reported interactions and effects on transport and further insights into the redundancy among beta-importing and exportins, as suggested by the reviewers, would considerably improve the study.

REVIEWER REPORTS

Reviewer #1:

Review of:
MSB-17-7608
Landscape of nuclear transport receptor cargo selectivity
Mackmull et al.

Summary:
The authors use BioID to identify proximity interactors of 16 nuclear transport receptors (NTR) in a

human cell line.

Review
This is a high quality mass spectrometry-based study conducted by the Beck laboratory. Using BioID to identify proximity interactions for 16 different nuclear transport receptors, the authors report >1000 new putative human NTR-cargo interactions.

I appreciate the fact that the authors used both N- and C-terminal BirA tagging, characterized the intracellular locations of the biotinylated protein partners for each of the baits used in the study, and utilized several different types of relevant BioID controls, to generate a high-confidence set of NTR interactors. The follow-up data analysis is also impressive, conducting comparisons to previously published datasets, and using a number of different analytical tools to evaluate the dataset in several different ways.

Major Concerns
1. I do, however, feel that there is something very important missing in this work - validation of novel biological insight.

A truly transformative manuscript of this type will highlight new biological insights provided by the dataset, then validate these insights, to show that the dataset is useful to the field, and to demonstrate how the dataset can be mined.

While I would not presume to direct the authors research, it could be very interesting here to see: (i) some examples of knock down/knock out of individual (or multiple) NTRs and the accompanying effects on transport of specific cargo proteins, based on the new predictions in this dataset. (ii) The authors suggest that their dataset reveals that only specific members of a given complex establish direct interactions with NTRs. They could validate this statement by demonstrating direct transporter-cargo binding in vitro using recombinant proteins, and/or using mutational analysis to disrupt interactions discovered in the dataset.

2. Do the data reveal any novel NTR binding sequences/structural motifs? Any biological surprises that could be further explored (quickly)?

3. While this is admittedly a significant amount of additional work, it would also dramatically increase the impact of the manuscript if the same analysis was conducted in an additional cell line.

Minor concerns:
1. While the terminology used here has been widely propagated in the literature, this reviewer suggests that it is incorrect to call the R118G BirA mutant protein "promiscuous". This word implies that the protein itself is biotinylating surrounding lysines, as opposed to the "non-promiscuous" WT protein, which recognizes a linear amino acid sequence in substrate polypeptides. It is not. The mutant enzyme is actually "abortive", prematurely releasing activated biotin (biotinoyl-AMP) which then diffuses away and chemically reacts with nearby amine groups.

2. Are the BirA-tagged NTRs properly localized? While the authors did a nice job of showing where the biotinylated proximity interactors are for each bait protein, I would also like to see either IF or live cell imaging for the endogenous versions of each NTR, along with the bait proteins themselves, in the same cells.

Reviewer #2:

Mackmull et al presents results of proteomic studies to identity protein cargos and other interacting partners of beta-Importin nuclear transport receptors (NTRs). Several proteomic studies of NTRs were reported in the last few years, but the current manuscript uses a different biochemical strategy of the BioID proximity ligation coupled to mass spectrometry. The Mackmull et al study is therefore a useful comparison with the previously published proteomic analyses. Their results also provide hints of direct interactors through identification of biotinylated peptides. Furthermore, the current

study avoids potential pitfalls of previous studies such as the use of digitonin-permeabilized cells and depletion of NTRs.

The authors analyzed proteomic results to provide useful and interesting new information and to provide support for surprising findings from previously reported studies. They show that > 1/3 of the proteome is involved in active nucleocytoplasmic transport. Such large fraction supports previous suggestions that a large fraction of the proteome (Kirli et al 2015) enters the nucleus at some time during the life of a cell.

The authors show functional redundancy of NTRs, especially amongst Importin-alphas and between homologous beta-Importins such as the IPO4/IPO5, TNPO1/TNPO2 pairs and between Importin-alphas and KPNB1. This is nice to see, but not unexpected. Perhaps the authors could provide more information with analysis of redundancy between different beta-Importins and also between the Exportins? The question of redundancy between very different NTRs is one that is often posed and the lack of knowledge causes confusion.

Finally, there is significant interest from biologists of various disciplines - eg. those studying signal transduction, cancer biologists, those studying nuclear processes, those studying biochemical interactions of NTRs and others, who will find lists of proteins that the authors find interacting with the different NTRs, useful. It may be useful for the authors to publish Tables of cargos (top scoring ones) for individual NTRs in the main part of the manuscript rather than have them be obscured in Supplementary Information.

1st Revision - authors' response 18 August 2017

We want to thank the reviewers for their very constructive criticism and suggestions that were very helpful to revise and improve our manuscript. Our detailed point by point response follows below.

Reviewer #1:
Review of: MSB177608: Landscape of nuclear transport receptor cargo selectivity. Mackmull et al.

Summary:
The authors use BioID to identify proximity interactors of 16 nuclear transport receptors (NTR) in a human cell line.

Review
This is a high quality mass spectrometry based study conducted by the Beck laboratory. Using BioID to identify proximity interactions for 16 different nuclear transport receptors, the authors report >1000 new putative human NTR cargo interactions. I appreciate the fact that the authors used both N and C terminal BirA tagging, characterized the intracellular locations of the biotinylated protein partners for each of the baits used in the study, and utilized several different types of relevant BioID controls, to generate a high confidence set of NTR interactors. The follow up data analysis is also impressive, conducting comparisons to previously published datasets, and using a number of different analytical tools to evaluate the dataset in several different ways.

Major Concerns

1. I do, however, feel that there is something very important missing in this work validation of novel biological insight. A truly transformative manuscript of this type will highlight new biological insights provided by the dataset, then validate these insights, to show that the dataset is useful to the field, and to demonstrate how the dataset can be mined.
While I would not presume to direct the authors research, it could be very interesting here to see: (i) some examples of knock down/knock out of individual (or multiple) NTRs and the accompanying effects on transport of specific cargo proteins, based on the new predictions in this dataset. (ii) The authors suggest that their dataset reveals that only specific members of a given complex establish direct interactions with NTRs. They could validate this statement by demonstrating direct transporter cargo binding in vitro using recombinant proteins, and/or using mutational analysis to disrupt interactions discovered in the dataset.

We agree with this critique and included the following additional data to demonstrate that our large scale analysis is a useful resource:

We have generated BirA fusion proteins for a selected subset of cargos that to the best of our knowledge have not been previously characterized. We reciprocally validated the respective transport pathways with a good success rate (described in lines 338 - 359 of the revised manuscript; Figures 5 and S4 and Tables S4, S5 and S8).*

As suggested by the reviewer, we quantified the nucleocytoplasmic distribution of two cargo proteins in response to gene silencing experiments for various NTRs (described in lines 359 - 366 of the revised manuscript; and Figure 5C, 5D, S4D, S4E, S4H). Also this data validates the specificity of the respective transport pathways.

In this context, we also looked on potential motifs (signal sequences) predicted for specific subunits of complexes (discussed in our response to the reviewer's point 2 right below).

2. Do the data reveal any novel NTR binding sequences/structural motifs? Any biological surprises that could be further explored (quickly)?

NLS and NES prediction generally suffers from low accuracy, possibly because these motifs are bipartite or at least to some extent act as 3D folds. Most likely because the available training data sets are rather small for the respective algorithms. Previous studies such as e.g. Kirli et al. have not done such analysis because it is considered challenging.

We have teamed up with Rob Russell from the University of Heidelberg and used his Dillimot algorithm to predict potential short linear motifs in significant enriched cargos of individual NTRs. As expected, this analysis recovered the PY-NLSs and cNLSs for transportin and importin alphas, respectively, to quite some extent, although Dillimot is not designed to identify bipartite motifs (Table S6). Mutational analysis of potential cNLSs yielded the expected cytoplasmic enrichment (Table S5, S6 and Figure 5E and 5F). Mass spectrometry data showed that the targeted subunits were integrated into the respective protein complexes (Table S4 and Figure 5A and S4B). We also identified DE-rich motifs in cargos of e.g. importin betas, which was unexpected (Table S6). We included some experimental validation of this motif but further analysis is required to understand if it is directly relevant for nuclear transport (Figure S4F and S4G). The results are summarized in Table S5 and described in lines 367 – 391 of the revised version of the manuscript.

3. While this is admittedly a significant amount of additional work, it would also dramatically increase the impact of the manuscript if the same analysis was conducted in an additional cell line.

Although we agree that such analysis would likely recover more cargos, we have to stress that this would be at least a year of work, including the molecular cloning and literally hundreds of MS runs. We thus hope that the reviewer agrees that such analysis is beyond the scope of the present manuscript.

Minor concerns:

1. While the terminology used here has been widely propagated in the literature, this reviewer suggests that it is incorrect to call the R118G BirA mutant protein "promiscuous". This word implies that the protein itself is biotinylating surrounding lysines, as opposed to the "nonpromiscuous" WT protein, which recognizes a linear amino acid sequence in substrate polypeptides. It is not. The mutant enzyme is actually "abortive", prematurely releasing activated biotin (biotinoylAMP) which then diffuses away and chemically reacts with nearby amine groups.

Agreed and corrected.

2. Are the BirA tagged NTRs properly localized? While the authors did a nice job of showing where the biotinylated proximity interactors are for each bait protein, I would also like to see either IF or live cell imaging for the endogenous versions of each NTR, along with the bait proteins themselves, in the same cells.

We have investigated this issue experimentally using the FLAG epitope that we had included in all our fusion proteins and compared the results to the Human Protein Atlas and literature. This analysis has been included into Figure 2B and indicates appropriate localization.

Reviewer #2:

Mackmull et al presents results of proteomic studies to identity protein cargos and other interacting partners of beta Importin nuclear transport receptors (NTRs). Several proteomic studies of NTRs were reported in the last few years, but the current manuscript uses a different biochemical strategy of the BioID proximity ligation coupled to mass spectrometry. The Mackmull et al study is therefore a useful comparison with the previously published proteomic analyses. Their results also provide hints of direct interactors through identification of biotinylated peptides. Furthermore, the current study avoids potential pitfalls of previous studies such as the use of digitonin permeabilized cells and depletion of NTRs.
The authors analyzed proteomic results to provide useful and interesting new information and to provide support for surprising findings from previously reported studies. They show that > 1/3 of the proteome is involved in active nucleocytoplasmic transport. Such large fraction supports previous suggestions that a large fraction of the proteome (Kirli et al 2015) enters the nucleus at some time during the life of a cell. The authors show functional redundancy of NTRs, especially amongst Importin alphas and between homologous beta Importins such as the IPO4/IPO5, TNPO1/TNPO2 pairs and between Importin alphas and KPNB1. This is nice to see, but not unexpected. Perhaps the authors could provide more information with analysis of redundancy between different beta Importins and also between the Exportins? The question of redundancy between very different NTRs is one that is often posed and the lack of knowledge causes confusion.

We have expanded the respective part of the main text (lines 290 to 304, see also Figure 4C and 4D) and included an additional display item that visualizes a global quantification of this aspect (Figure S3A). The overlap among exportins is rather low as compared to e.g. importin alphas. Careful inspection of our data revealed that there is an overlap of cargos between IPO4, IPO5 and XPO7.

Finally, there is significant interest from biologists of various disciplines eg. Those studying signal transduction, cancer biologists, those studying nuclear processes, those studying biochemical interactions of NTRs and others, who will find lists of proteins that the authors find interacting with the different NTRs, useful. It may be useful for the authors to publish Tables of cargos (top scoring ones) for individual NTRs in the main part of the manuscript rather than have them be obscured in Supplementary Information.

This was a good suggestion. We have included an additional display item into the revised version (Table 1). We also made an effort to improve the usability of the respective supplementary table containing large-scale data, e.g. by adding a color code (Table 4).

2nd Editorial Decision 5 October 2017

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who accepted to evaluate the revised study. As you will see, the referees are now fully supportive and I am pleased to inform you that we will be able to accept you paper for publication in Molecular Systems Biology, pending the following minor modifications suggested by our editorial assistant at QC

REVIEWER REPORTS

Reviewer #1:

The authors have (more than) satisfactorily dealt with all of my previous issues / critiques.

Reviewer #2:

The reviewer is satisfied with the revisions.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name:	Martin Beck
Journal Submitted to:	Molecular Systems Biology
Manuscript Number:	MSB-17-7608R

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., *P* values = x but not *P* values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com	Antibodypedia
http://1degreebio.org	1DegreeBio
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor	ARRIVE Guidelines
http://grants.nih.gov/grants/olaw/olaw.htm	NIH Guidelines in animal use
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm	MRC Guidelines on animal use
http://ClinicalTrials.gov	Clinical Trial registration
http://www.consort-statement.org	CONSORT Flow Diagram
http://www.consort-statement.org/checklists/view/32-consort/66-title	CONSORT Check List
http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tum	REMARK Reporting Guidelines (marker prognostic studies)
http://datadryad.org	Dryad
http://figshare.com	Figshare
http://www.ncbi.nlm.nih.gov/gap	dbGAP
http://www.ebi.ac.uk/ega	EGA
http://biomodels.net/	Biomodels Database
http://biomodels.net/miriam/	MIRIAM Guidelines
http://jij.biochem.sun.ac.za	JWS Online
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html	Biosecurity Documents from NIH
http://www.selectagents.gov/	List of Select Agents

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	According to common good practice in the field of MS-based proteomics we chose to measure 4 biological replicates of all conditions.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Yes. Distributions were visually inspected.
Is there an estimate of variation within each group of data?	Yes.
Is the variance similar between the groups that are being statistically compared?	Yes.

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Catalogue numbers are given in methods section.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Cell lines were regularily tested for mycoplasma contamination bot not authenticated.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Data are available via ProteomeXchange with identifier PXD007976.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NA
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA