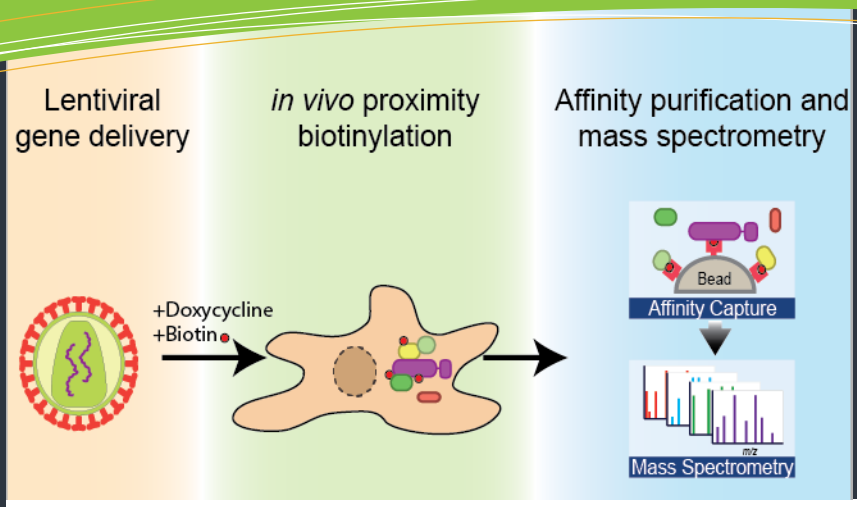
January 2019

Network Biology Collaborative Centre

Tech Note



This issue

BioID

BioID is a short form for proximity-dependent biotin identification. In BioID, a protein of interest (the bait) is tagged with an abortive biotin ligase that marks proteins (the preys) in close proximity to the bait with biotin. Biotinylated proteins are purified and identified by mass spectrometry.

Advantages

Proteins in close proximity are covalently marked so prey identification is not reliant on maintaining interactions with the bait. Ideally suited for proteins localized to insoluble cellular structures (e.g. plasma membranes or organelles) or for capturing low affinity interactions.

Lentiviral Delivery

Lentiviral delivery using vectors described in this Tech Note expand BioID to hard-to-transfect cells including immortalized and primary cells.

Expanding BioID to Diverse Cell Types

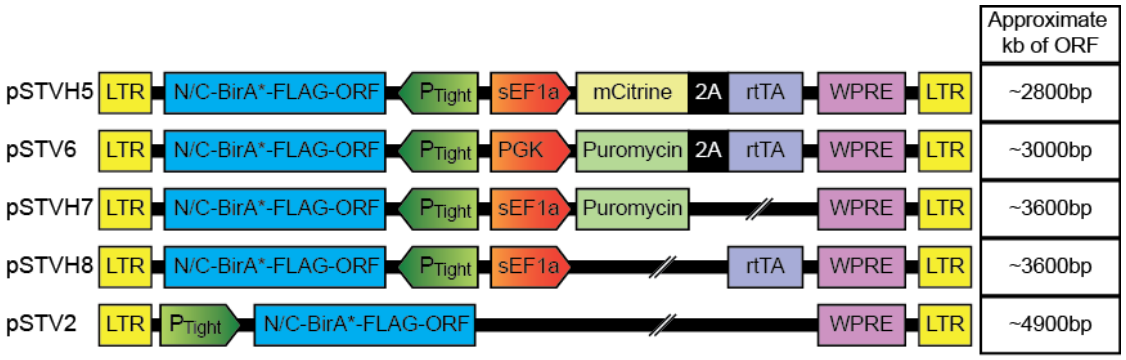
Deciphering a protein's function necessitates an understanding of the cellular environment in which it resides and the interactions it makes with neighbouring proteins. Since its first use over 6 years ago, the BioID approach is increasingly becoming the tool of choice to identify proteins in close proximity to a target protein. Here, we describe a toolkit of new lentiviral vectors for BioID that extend the application of this technique to a diverse range of cell types.

BioID was first described in 2012 where it was used to identify new components of nuclear lamina[1]. Since then it has been employed to define protein constituents of centrosomes and cilia[2,3], focal adhesions[4], stress granules and P-bodies[5] as well as to uncover new components of signaling pathways[6]. In BioID, the bait protein is fused to a mutated BirA ligase (R118G, denoted as BirA*). In cells, the BirA* tag converts biotin to the intermediate biotinoyl-5'-AMP. The biotin-AMP diffuses away from the tagged bait and covalently modifies epsilon amine groups of lysines on proteins within an approximate 10 nM radius[7]. Biotinylated proteins are then purified using streptavidin-conjugated beads and bound proteins are identified by mass spectrometry. Since the proteins are covalently labeled with biotin, the cells can be lysed under harsher conditions making this technique particularly suitable for cellular compartments that are often intractable to traditional immunoprecipitation experiments where gentle lysis is required to maintain interactions.

Because the BirA* fusion has to be introduced into cells, most BioID experiments have been performed in model cell lines such as HEK293 or HeLa that are easy to grow, transfect and generate stable cell lines (a strategy we have used succesfully with Flp-In T-REx versions of these cells that permit single integration of a tetracycline-regulatable construct[2,5,6]). To expand BioID to more diverse and biologically relevant cells including primary cells and tissue, post-doctoral fellow Payman Samavarchi-Tehrani from the Gingras lab at the Lunenfeld-Tanenbaum Research Institute developed a suite of lentiviral vectors to perform BioID in hard-to-transfect cells[8]. We have benchmarked this vector toolkit using 4 baits (3 nuclear, 1 cytoplasmic) in HeLa cells, primary mouse embryonic fibroblasts (MEFs) and human foreskin fibroblasts. This Tech Note describes these vectors in more detail, highlights key requirements for their effective use and provides ordering information and links to detailed protocols.

Figure 1.

Schematic of the 5 lentiviral backbone vectors and the size of open-reading frames that can be packaged.



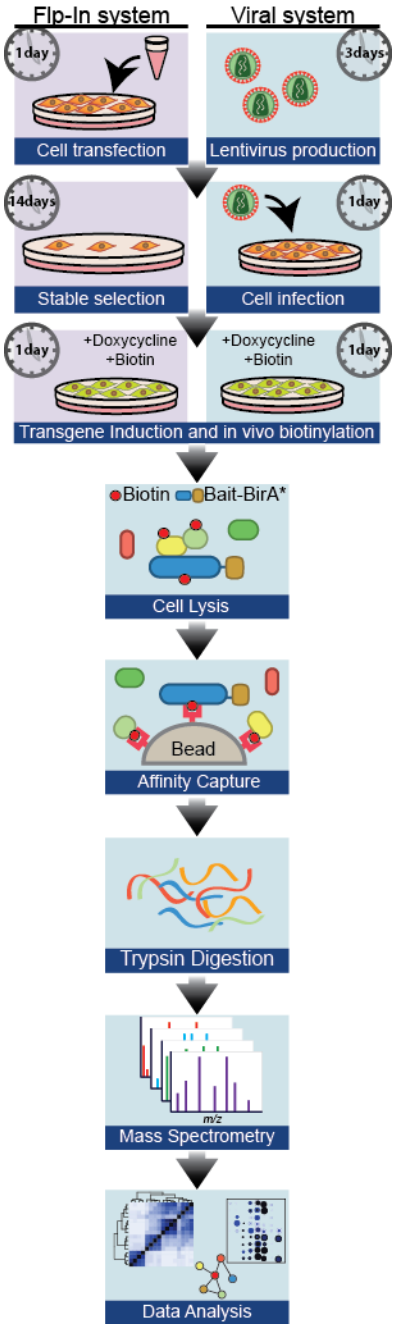


Figure 2.

BioID workflow: stable cells vs. lentiviral infection.

Features of the Lentiviral Vectors

- 5 backbone vectors with N- or C-terminal BirA-FLAG tagging options
- Gateway cloning sites for rapid cloning of open-reading frames (ORFs) from entry clones
- Tetracycline-inducible promoter to control gene expression (rtTA transactivator needed to induce expression is available in indicated vectors)
- Selectable marker options: puromycin for drug selection, mCitrine fluorescent protein to sort cells or to monitor infection
- Lentiviral packaging limits of ~10 KB[9,10] may restrict choice of vectors from all-in-one vectors (pSTVH5, pSTV6, pSTVH8) to the streamlined pSTVH8 or pSTV2 vectors. The approximate size of open-reading frames that can be packaged is shown in **Figure 1**.

Using the Lentiviral Vectors in BioID Experiments

The workflows for stable cell line generation (Flp-In T-Rex system from ThermoFisher Scientific) and viral infection are shown in **Figure 2**. In addition to expanding the cell types available, the lentiviral workflow shortens the time required to generate material for BioID.

Prior to performing a BioID experiment, it is important to optimize BioID for the cell line that is being used. Titration of virus is required to ensure that comparable proportions of cells are infected by the transgene. Doxycycline for induction of gene expression may also need to be adjusted to ensure similar expression across experimental conditions or cell types.

To start, we suggest infecting a 10 cm dish at 35-40% density with sufficient virus to infect 75-80% of cells and to scale these cells to one 15 cm dish for BioID. Experiments should be performed in at least biological duplicates. With careful attention to the protocol, correlation between replicates is high (R^2 >.93[8]).

Inclusion of proper controls is imperative to differentiate specific proximal interactors from background proteins identified by mass spectrometry. Performing BioID from cells not expressing the BirA* tag will identify endogenously modified proteins and proteins that bind non-specifically to the streptavidin support. BirA* expressed alone or fused to a fluorescent protein (e.g. EGFP) will help identify non-specific biotinylation. For nuclear baits, a fusion of BirA* to a nuclear localization signal (such as BirA*-EGFP-NLS) will help eliminate compartment-specific background proteins.

We recommend using SAINTexpress[11] (as a standalone tool, through our ProHits-LIMS management system[12] or more easily on the Contaminant Repository for Affinity Purification website (CRAPome.org[13]) to identify preys specific to your bait. Tools for data visualization and analysis are available on the ProHits-viz[14] website.

Q&A with Dr. Payman Samavarchi-Tehrani

Q: What motivated you to create this resource and how do you think it will be applied?

A: The advent of proximal biotinylation techniques such as BioID has significantly changed how we can study the proximal neighbourhood of a protein of interest.

To apply BioID to my own research question, I needed to apply this technique to stem cells and primary cells. This led me to generate a few lentiviral vectors to deliver the BirA*-tagged baits into these cells. However, after talking to a few colleagues about my project, I realized that these reagents could have broader utility. So I extended my efforts to generate a few additional vectors with various features, worked-out some of the experimental details and benchmarked the validity and reproducibilty of the workflow.

With my own focus now turned back to my original research question, I am most interested in implementation of BioID to study cell fate specification using stem cell differentiation and somatic cell reprogramming systems. I'm also really eager to see the various ways in which the scientific community utilizes these reagents. Although it would be technically challenging, I would love to see *in vivo* application of BioID using lentiviral delivery to study normal development or disease onset and progression.

Dr. Samavarchi-Tehrani is a post-doctoral fellow in the Gingras Lab and first author on the publication describing the lentiviral vector toolkit.

How to get started:

- **Read background publications.**

  The lentiviral toolkit – Samavarchi-Tehrani *et al.*, Mol Cell Proteomics, 2018

  Parallel exploration of interactions by BioID and AP-MS – Hesketh *et al.*, Methods Mol Bio, 2017

  Getting to know the neighbourhood – Gingras et al., Curr Open Chem Biol, 2018

- **Request lentiviral vectors from the NBCC.**

  For more detailed information including vector maps and sequences and to request vectors, please visit our website at nbcc.lunenfeld.ca/resources.

- **Download the protocols for sample preparation.**

  For lentiviral production and cell preparation – NBCC proteomics protocol 4

  For affinity enrichment and preparation of samples for mass spectrometry – NBCC proteomics protocol 5

- **Need Help?**

  The Network Biology Collaborative Centre offers service in sample preparation, mass spectrometry and data analysis. To find out more about our services, please visit our website at nbcc.lunenfeld.ca or contact us at nbcc@lunenfeld.ca.

Network Biology Collaborative Centre
Lunenfeld-Tanenbaum Research Institute
Sinai Health System
600 University Avenue, Room 970
Toronto, Ontario, M5G 1X5
416-586-4800 x3018 ph
https://nbcc.lunenfeld.ca

Lunenfeld-Tanenbaum Research Institute
Sinai Health System

Written by Karen Colwill, PhD
NBCC Manager

References

1 Roux, K. J. *et al. The Journal of cell biology* **196**, 801-810 (2012).
2 Gupta, G. D. *et al. Cell* **163**, 1484-1499 (2015).
3 Firat-Karalar, E. N. *et al. Curr Biol* **24**, 664-670 (2014).
4 Dong, J. M. *et al. Science signaling* **9**, rs4 (2016).
5 Youn, J. Y. *et al. Molecular cell* **69**, 517-532 e511 (2018).
6 Couzens, A. L. *et al. Science signaling* **6**, rs15 (2013).
7 Kim, D. I. *et al. Proceedings of the National Academy of Sciences of the United States of America* **111**, E2453-2461 (2014).
8 Samavarchi-Tehrani, P. *et al. Mol Cell Proteomics* **17, 2256-2269** (2018).
9 Kumar, M. *et al. Hum Gene Ther* **12**, 1893-1905 (2001).
10 al Yacoub, N. *et al. J Gene Med* **9**, 579-584 (2007).
11 Teo, G. *et al. J Proteomics* **100**, 37-43 (2014).
12 Liu, G. et al. Nature biotechnology 28, 1015-1017 (2010).
13 Mellacheruvu, D. et al. Nature methods 10, 730-736 (2013).
14 Knight, J. D. R. et al. Nature methods 14, 645-646 (2017).

About the NBCC

The Network Biology Collaborative Centre provides integrated solutions for functional genomics and proteomics. The Centre provides complementary services including mass spectrometry-based proteomics, high-throughput screening and drug discovery, next-generation sequencing and high-content to high-resolution imaging. Integrated data analysis enables data to be combined from these different services.