Data Descriptor Title (110 character maximum, inc. spaces)

Firstname Lastname[1], Firstname Lastname[2*]

May 24, 2017

1. An affiliation 2. A different affiliation *corresponding author(s): Firstname Lastname (email@address)

Abstract

This is a manuscript template for Data Descriptor submissions to *Scientific Data* (http://www.nature.com/scientificdata). The abstract must be no longer than 170 words, and should succinctly describe the study, the assay(s) performed, the resulting data, and the reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section.

Background & Summary

(700 words maximum) An overview of the study design, the assay(s) performed, and the created data, including any background information needed to put this study in the context of previous work and the literature. The section should also briefly outline the broader goals that motivated the creation of this dataset and the potential reuse value. We also encourage authors to include a figure that provides a schematic overview of the study and assay(s) design. This section and the other main body sections of the manuscript should include citations to the literature as needed [1, 2]. References should be included within the manuscript file itself as our system cannot accept BibTeX bibliography files. Authors who wish to use BibTeX to prepare their references should therefore copy the reference list from the .bbl file that BibTeX generates and paste it into the main manuscript .tex file (and delete the associated \bibliography and \bibliographystyle commands).

Methods

The Methods should include detailed text describing any steps or procedures used in producing the data, including full descriptions of the experimental design, data acquisition assays, and any computational processing (e.g. normalization, image feature extraction). Related methods should be grouped under

corresponding subheadings where possible, and methods should be described in enough detail to allow other researchers to interpret and repeat, if required, the full study. Specific data outputs should be explicitly referenced via data citation (see Data Records and Data Citations, below). Authors should cite previous descriptions of the methods under use, but ideally the method descriptions should be complete enough for others to understand and reproduce the methods and processing steps without referring to associated publications. There is no limit to the length of the Methods section.

Code availability

For all studies using custom code in the generation or processing of datasets, a statement must be included here, indicating whether and how the code can be accessed, including any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset.

Data Records

Please explain each data record associated with this work, including the repository where this information is stored, and an overview of the data files and their formats. Each external data record should be listed in Data Citation section at the end of this template, and records should be cited throughout the manuscript as, for example (Data Citation 1).

Tables should be used to support the data records, and should clearly indicate the samples and subjects, their provenance, and the experimental manipulations performed on each. They should also specify the data output resulting from each data-collection or analytical step, should these form part of the archived record. Please see the submission guidelines at the *Scientific Data* website, and our Word templates for more information on preparing such tables.

Technical Validation

This section presents any experiments or analyses that are needed to support the technical quality of the dataset. This section may be supported by up figures and tables, as needed. This is a required section; authors must present information justifying the reliability of their data.

Usage Notes

Brief instructions that may help other researchers reuse these dataset. This is an optional section, but strongly encouraged when helpful to readers. This may include discussion of software packages that are suitable for analyzing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or

tips for integrating or comparing this with other datasets. If needed, authors are encouraged to provide code, programs, or data processing workflows when they may help others analyse the data. We encourage authors to archive related code in a DOI-issuing archive when possible, but code may also be supplied as supplementary information files.

For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, and what criteria will be used to determine who may access the data, and any limitations on data use.

Acknowledgements

Text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged. Author contributions Please describe briefly the contributions of each author to this work on a separate line.

AK did this and that.

BG did this and that and the other.

Competing financial interests

A competing financial interests statement is required for all accepted papers published in *Scientific Data*. If none exist simply write, "The author(s) declare no competing financial interests".

Figures Legends

Figure should be referred to using a consistent numbering scheme through the entire Data Descriptor. For initial submissions, authors may choose to supply this document as a single PDF with embedded figures, but separate figure image files must be provided for revisions and accepted manuscripts. In most cases, a Data Descriptor should not contain more than three figures, but more may be allowed when needed. We discourage the inclusion of figures in the Supplementary Information – all key figures should be included here in the main Figure section.

Figure legends begin with a brief title sentence for the whole figure and continue with a short description of what is shown in each panel, as well as explaining any symbols used. Legend must total no more than 350 words, and may contain literature references.

Tables

Tables supporting the Data Descriptor. These can provide summary information (sample numbers, demographics, etc.), but they should generally not be used to present primary data (i.e. measurements). Tables containing primary data should be submitted to an appropriate data repository.

Tables may be provided within the LaTeX document or as separate files (tab-delimited text or Excel files). Legends, where needed, should be included here. Generally, a Data Descriptor should have fewer than ten Tables, but more may be allowed when needed. Tables may be of any size, but only Tables which fit onto a single printed page will be included in the PDF version of the article (up to a maximum of three).

References

[1] Califano, A., Butte, A. J., Friend, S., Ideker, T. & Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature Genetics* **44**, 841–847 (2012).

[2] Wang, R. *et al.* PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nature Biotechnology* **30**, 135–137 (2012).

Data Citations

Bibliographic information for the data records described in the manuscript.

1. Lastname1, Initial1., Lastname2, Initial2., ...& LastnameN, InitialN. *Repository name* Dataset accession number or DOI (YYYY).