PlotXpress, a webtool for normalization and visualization of reporter expression data

Elias Brandorff^{1,*}, Marc Galland^{2,*}, Joachim Goedhart^{3,*}

¹ Swammerdam Institute for Life Sciences, Evolutionary Neurogenomics Department, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands.

² Swammerdam Institute for Life Sciences, Plant Physiology Department, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands.

³ Swammerdam Institute for Life Sciences, Section of Molecular Cytology, van Leeuwenhoek Centre for Advanced Microscopy, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands.

* Correspondence:

Email: J.E.Brandorff@uva.nl | ORCID: [0000-0002-5871-2567](https://orcid.org/0000-0002-5871-2567)

Email: M.D.Galland@uva.nl | ORCID: [0000-0003-2161-8689](https://orcid.org/0000-0003-2161-8689) | www: <https://www.mgalland.info>

Email: j.goedhart@uva.nl | ORCID: [0000-0002-0630-3825](https://orcid.org/0000-0002-0630-3825) | Twitter: [@joachimgoedhart](https://twitter.com/joachimgoedhart)

Abstract

In molecular cell biology, reporter assays are frequently used to investigate gene expression levels. Reporter assays employ a gene that encodes a light-emitting protein, of which the luminescence is quantified as a proxy of gene expression. Commercial parties provide reporter assay kits that include protocols and specialized detection machinery. However, downstream analysis of the output data and their presentation are not standardized. We have developed plotXpress to fill this gap, providing a free, open-source platform for the semi-automated analysis and standardized visualisation of experimental gene reporter data. Users can upload raw luminescence data acquired from a reporter gene assay with an internal control. In plotXpress, the data is corrected for sample variation with the internal control and the average for each condition is calculated. When a reference condition is selected the fold change is calculated for all other conditions, based on the selected reference. The results are shown as dot plots with a statistical summary, which can be adjusted to create publication-grade plots without requiring coding skills. Altogether, plotXpress is an open-source, low-threshold, web-based tool, that promotes a standardized and reproducible analysis while providing an appealing visualization of reporter data. The webtool can be accessed at: <https://huygens.science.uva.nl/PlotXpress/>

Introduction

Reporter gene assays are popular tools to investigate gene expression dynamics in cell biology (Barriscale et al., 2014; Liu et al., 2009; Schenborn and Groskreutz, 1999). In most systems, the reporter is a gene coding for a protein that emits light when it binds a substrate. An example is the firefly (*Photinus pyralis*) luciferase that emits light when it binds the substrate luciferin (Himes and Shannon, 2000). The luminescent light can be detected with dedicated equipment and reflects the expression of the reporter gene.

The expression level of the reporter gene is used as a proxy of the capacity of a DNA sequence to regulate gene expression under various experimental conditions. To this end, the DNA sequence of interest is cloned upstream of the reporter gene in a vector which is subsequently transfected into cells. Reporter gene activity can be compared between variations of a regulatory DNA sequence, for example by removing a transcription factor binding site or by modulating tandem repeat expansions in a promoter sequence (Rodriguez et al., 2020; Moparthi and Koch, 2020). Another option is to measure the effects on reporter activity of treatments such as pharmaceutical compounds or overexpression of transcription factors that interact with the sequence of interest (Asamitsu et al., 2021). In addition, reporter assays have been used to study the relation between microRNAs and cancer progression (Wang et al., 2021; Chengling et al., 2021). In the so-called “Dual-Luciferase® Reporter Assay System” (Promega: <https://nld.promega.com/resources/pubhub/in-vivo-evaluation-of-regulatory-sequences-by-analysis-of-luciferase-expression/>), an internal control is used to correct for variations in cell density and transfection efficiency. This internal control is co-transfected and consists of a vector with a constitutively active promoter that drives expression of another luciferase. This second luciferase emits a different color of light and is, in our example, derived from *Renilla reniformis* (sea pansy). Transfection of the internal control is kept the same in each well and reflects transfection efficiency and cellular protein production.

A given sequence of interest is cloned into a luciferase expressing plasmid in order to investigate its activity as a gene promoter. The transfection of an empty vector control, without this sequence, is an important reference condition and required to correct for

unintended effects of the plasmid alone. For example, in some vectors, luciferase is expressed under control of an SV40 promoter element. A transcription factor binding site (TFBS), cloned upstream of this promoter will influence its activity and the expression of luciferase. The measured effect is a composite of promoter activity *plus* the TFBS. The empty vector control allows normalization for baseline promoter activity and isolation of the influence of the TFBS alone. Here, we will call the empty vector condition the *reference* condition.

In reporter assays with multiple reporter constructs, cell types, treatments and reference conditions, experiments may increase rapidly in size and complexity, leading to challenging downstream analyses (Schagat et al. 2007). There is currently neither standardized method nor computational tool for the processing of reporter assay data and its visualization. PlotXpress was designed to simplify the analysis process of complex reporter assays by providing an online tool with a standard for data processing and visualization. PlotXpress was built following the philosophy of transparent and state-of-the-art data visualization implemented in the data visualization app PlotsOfData (Postma & Goedhart, 2019). In addition to a streamlined analysis, plotXpress enables transparent communication of the data. Instead of only providing averaged data and error bars to summarize gene reporter data in bar graphs, plotXpress produces dot plots maintaining individual data points. Data in both wide and tidy format (Wickham, 2014) can be provided. As coding skills are not required, plotXpress is a readily available low entry-level application that democratizes the processing and visualization of dual reporter expression data.

Figure 1. Graphic representation of a typical reporter assay. Firefly and renilla expressing plasmids are transfected into cells grown in 24-well plates. After incubation the cells are harvested and lysates pipetted into 96-wells optiplates. These are loaded into a plate reader where the luciferase substrates are added. Once the reactions are initiated, luminescence is detected, and readings are stored digitally.

Availability & code

The plotXpress web tool can be accessed at <https://huygens.science.uva.nl/PlotXpress/> (and a mirror is available at: <https://goedhart.shinyapps.io/PlotXpress>). Example data is provided to explore the functions of the app. The plotXpress code is written in the R programming language (<https://www.r-project.org>) using the following packages: ggplot2 (Wickham, 2016), tidyr, magrittr, readr, stringr, dplyr and readxl which are all part of the tidyverse suite of packages version 1.3.1 (Wickham et al., 2019), Shiny (<https://CRAN.R-project.org/package=shiny>), and DT (<https://CRAN.R-project.org/package=DT>).

This manuscript documents version 1.0.0 of the webtool which is archived (together with the example data) at Zenodo: <https://doi.org/10.5281/zenodo.4980068>

Background information, updates of the code and version releases will be published on GitHub: <https://github.com/ScienceParkStudyGroup/PlotXpress>. GitHub is the preferred channel for communication regarding issues and feature requests.

Example Data

The plotXpress app comes with an example dataset and a file that describes the experimental conditions for each well 'Tidy_design.csv' (**Fig 1**). Both files, 'DualLuc_example_data.xlsx' and 'Tidy_design.csv', are also available online at <https://github.com/ScienceParkStudyGroup/PlotXpress>.

Data upload

Users can upload their data in two different ways. The first option is for data acquired with the Promega GloMax plate reader. The alternative is a general-purpose option that accepts data in a tidy format.

GloMax data upload

The output of the GloMax plate reader (Promega) is a spreadsheet XLSX format with two tables containing firefly and renilla luminescence signals stored in a 96-well lay-out. PlotXpress reads the cells with the firefly and renilla readings and provides a graphic overview of the experiment by showing a 96-well plate where signal intensity is colour coded (**Fig 2**).

Figure 2. PlotXpress data upload. Screenshot of plotXpress app showing a 96-well format with signal intensities represented by false colour.

A separate table in CSV format with the experimental conditions per well is required for the data processing. A template is available for download within the app or online: https://github.com/ScienceParkStudyGroup/PlotXpress/blob/master/Tidy_design.csv

The firefly and renilla reads from the GloMax data are converted into a tidy format and merged with experimental conditions that are taken from the uploaded design (**Fig 3**). The resulting tidy dataframe is used for data processing.

firefly

	1	2	3	4	5	6
A	2010.00	1639000.00	1498000.00	1637.00	720340.00	1464540.00
B	3210.00	1639099.00	1498099.00	1736.00	720439.00	1464639.00
C	1965.00	1638451.00	1497451.00	1088.00	719791.00	1463991.00
D	2381.00	1639234.00	1498234.00	1871.00	720574.00	1464774.00

renilla

	1	2	3	4	5	6
A	2391540.00	2219540.00	2313540.00	1580540.00	2217540.00	1843540.00
B	2391639.00	2219639.00	2313639.00	1580639.00	2217639.00	1843639.00
C	2390991.00	2218991.00	2312991.00	1579991.00	2216991.00	1842991.00
D	2391774.00	2219774.00	2313774.00	1580774.00	2217774.00	1843774.00

→

tidy

Wells	row	column	condition	treatment1	treatment2	firefly	renilla
A01	A	1	control	HEK	-	2010	2391540
B01	B	1	control	HEK	-	3210	2391639
C01	C	1	control	HEK	-	1965	2390991
D01	D	1	control	HEK	-	2381	2391774
E01	E	1	control	neuron	-	1292	269021
F01	F	1	control	neuron	-	991	268918

Figure 3. GloMax data conversion to tidy format. PlotXpress allows uploading of GloMax spreadsheet input. An additional experimental design file in tidy format should be provided to connect the reads with the conditions. After supplying the design file, the GloMax data is converted to a table in tidy format, where each row contains a single observation and the experimental conditions. The table can be downloaded and used as input for plotXpress.

Tidy data upload

Instead of uploading a GloMax spreadsheet in 96-well format and an additional design file, users can choose to upload a single tidy dataset (.CSV) containing both experimental conditions and firefly and renilla signals. An example file is available: https://github.com/ScienceParkStudyGroup/PlotXpress/blob/master/plotXpress_Tidy-3.csv

All the relevant data from one experiment (design and measurements) should be present in one tidy table. Users are free in the size of the experiment: any number of cell types, sequences of interest, or treatments may be added. The minimal information

that is required is a column with wells (in the format A01, B01, ..), a column with intensity data and a column with the conditions. An additional column with reference data is optional and is used to normalize the data if supplied.

Data handling

In dual luciferase assays, each replicate consists of a firefly luciferase and renilla luciferase intensity reading. To normalize the reporter gene expression to the internal control, plotXpress calculates the ratio of firefly signal to renilla signal, resulting in the firefly/renilla ratio. An average firefly/renilla ratio is then calculated for each group of readings that share identical names in the “condition” column. Finally, the fold change is calculated by dividing the firefly/renilla ratio by the firefly/renilla ratio of a selected reference condition:

$$Fold\ Change = \frac{(firefly/renilla)}{(firefly/renilla)_{reference}} \quad (\text{equation 1})$$

PlotXpress offers flexibility in the selection of conditions to filter the data before plotting. Individual wells can be removed from the analysis or specific conditions can be filtered.

The combined dataset including information about the wells, condition, treatments, firefly and renilla signals and the calculated ratios can be downloaded in tidy format (CSV) for further processing and statistical testing.

Data visualization

In the plotting tab called “Plot”, a reference condition can be selected. Without selecting a reference, the firefly/renilla ratio is plotted on the y-axis. If a reference condition is selected, normalized firefly/renilla values from the other conditions are represented as fold-change relative to the reference (**Fig 4A**).

Figure 4. Plots based on example data. (A) Standard output plot produced by plotXpress after selecting a reference condition. The plots show the fold change in expression induced by gene A and gene B relative to a reference, for each of 8 different conditions. Individual datapoints are shown and the median of each condition is indicated with a horizontal line. (B) A 90° rotated panel for one of the conditions (drug1 in neurons) shown in panel A. Rotation of the panel improves readability of the fold change and allows for the display of a cartoon of the reporter gene next to it.

In plotXpress a categorical condition can be compared by selecting an option in the “compare” section. This condition is shown on the x-axis. Each datapoint of the reporter assay is plotted as a dot and represents the normalized firefly/renilla ratio. Dot size and transparency can be adjusted as well as other features such as axis scales, plot title, font size and the use of grid lines. The mean or median can be shown per condition as a line (Fig 4A). When presenting reporter expression data, it can be useful to display a rotated plot. Diagrams of reporter constructs can be included to visually support the experimental set up and show the corresponding measurements in line (**Fig 4B**). As an optional feature, the reference can be shown or hidden, depending on user preference.

Outputs

The combined data from an experiment can be downloaded as a CSV file in tidy format. This data format can be used for further processing, such as statistical testing and plotting. The tidy format is well handled by statistical software (R) and other web tools that we have developed (available at <https://huygens.science.uva.nl/>). For instance, when the data consists of a mix of technical and biological replicates, this can be visualized as a superplot (Lord et al., 2020) by uploading the data in the SuperPlotsOfData app (Goedhart, 2020). Final plots can be downloaded from PlotXpress in PNG or PDF format, the latter being ideal for downstream processing with vector-based graphic software.

PlotXpress also produces a summary table in multiple formats such as CSV or PDF that can be included as supplemental data in publications (**Fig 5**)

PlotXpress

condition	treatment1	treatment2	n	mean	sd	sem	CI_lo	CI_hi	median
control	HEK	-	4	1	0.2	0.1	0.6	1.4	0.9
control	neuron	-	4	1	0.1	0.1	0.8	1.2	1
geneA	HEK	-	4	738.4	0	0	738.4	738.5	738.5
geneA	neuron	-	4	112	3.7	2.1	105.3	118.8	112.2
geneB	HEK	-	4	647.5	0.1	0	647.4	647.6	647.5
geneB	neuron	-	4	248.9	18.1	10.5	215.6	282.2	249.4
control	HEK	drug1	4	1	0.2	0.1	0.6	1.4	1.1
control	neuron	drug1	4	1	0.1	0.1	0.8	1.2	1
geneA	HEK	drug1	4	324.3	0.1	0.1	324.1	324.5	324.3
geneA	neuron	drug1	4	454.6	24	13.9	410.4	498.7	465.1
geneB	HEK	drug1	4	793.2	0	0	793.1	793.2	793.2
geneB	neuron	drug1	4	949.8	55.4	32	848.1	1051.5	973.6

Figure 5. Data summary table. Table containing summary statistics of the data shown in figure 4A.

Perspectives

The current version of plotXpress was written to analyze expression data in tidy format as well as for GloMax output data. The import of data produced by plate readers from other manufacturers is not supported at this moment, but we welcome suggestions and example data to implement import functions for other data formats. Since plotXpress is open-source, users can modify the source code by making a fork to the GitHub repository and create a pull request to have their changes reviewed and possibly integrated. Although plotXpress was developed for the analysis and visualization of dual-luciferase experiments, it can be used for other types of data that require normalization or comparison with a reference condition, such as qPCR data.

Conclusion

Gene reporter assays are an invaluable tool for molecular biology, enabling the study of regulatory DNA sequences, transcription factors, pharmaceutical compounds, oligonucleotides and other factors that interact with DNA sequences. With each treatment or cellular context, the assay increases in complexity. PlotXpress simplifies the data analysis by bridging the gap between wet-lab standards (96-well format) and dry-lab conventions (tidy data format) that are typically used in downstream analysis of experimental biological data. The open-source web tool enables anyone to benefit from standardized data processing and state-of-the-art data visualization.

References

- Asamitsu, S., Y. Yabuki, S. Ikenoshita, K. Kawakubo, M. Kawasaki, S. Usuki, Y. Nakayama, K. Adachi, H. Kugoh, K. Ishii, T. Matsuura, E. Nanba, H. Sugiyama, K. Fukunaga, and N. Shioda. 2021. CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci. Adv.* 7:eabd9440. doi:10.1126/sciadv.abd9440.
- Barriscale, K.A., S.A. O'Sullivan, and T.V. McCarthy. 2014. A single secreted luciferase-based gene reporter assay. *Anal Biochem.* 453:44–49. doi:10.1016/j.ab.2014.02.019.
- Chengling, L., Z. Yulin, X. Xiaoyu, L. Xingchen, Z. Sen, W. Ziming, and C. Xianming. 2021. miR-325-3p, a novel regulator of osteoclastogenesis in osteolysis of colorectal cancer through targeting S100A4. *Mol Med.* 27:23. doi:10.1186/s10020-021-00282-7.
- Goedhart, J. 2020. SuperPlotsOfData – a web app for the transparent display and quantitative comparison of continuous data from discrete conditions. *bioRxiv.* 2020.09.01.276881. doi:10.1101/2020.09.01.276881.
- Himes, S.R., and M.F. Shannon. 2000. Assays for transcriptional activity based on the luciferase reporter gene. *Methods Mol Biol.* 130:165–174. doi:10.1385/1-59259-686-x:165.
- Liu, A.M.F., D.C. New, R.K.H. Lo, and Y.H. Wong. 2009. Reporter gene assays. *Methods Mol Biol.* 486:109–123. doi:10.1007/978-1-60327-545-3_8.
- Lord, S.J., K.B. Velle, R.D. Mullins, and L.K. Fritz-Laylin. 2020. SuperPlots: Communicating reproducibility and variability in cell biology. *Journal of Cell Biology.* 219. doi:10.1083/jcb.202001064.
- Moparthy, L., and S. Koch. 2020. A uniform expression library for the exploration of FOX transcription factor biology. *Differentiation.* 115:30–36. doi:10.1016/j.diff.2020.08.002.
- Rodriguez, C.M., S.E. Wright, M.G. Kearse, J.M. Haenfler, B.N. Flores, Y. Liu, M.F. Ifrim, M.R. Glineburg, A. Krans, P. Jafar-Nejad, M.A. Sutton, G.J. Bassell, J.M. Parent, F. Rigo, S.J. Barmada, and P.K. Todd. 2020. A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat Neurosci.* 23:386–397. doi:10.1038/s41593-020-0590-1.
- Schenborn, E., and D. Groskreutz. 1999. Reporter gene vectors and assays. *Mol Biotechnol.* 13:29–44. doi:10.1385/MB:13:1:29.
- Wang, D., M. Feng, X. Ma, K. Tao, and G. Wang. 2021. Transcription factor SP1-induced microRNA-146b-3p facilitates the progression and metastasis of colorectal cancer via regulating FAM107A. *Life Sci.* 277:119398. doi:10.1016/j.lfs.2021.119398.
- Wickham, H. 2014. Tidy Data. *J. Stat. Soft.* 59. doi:10.18637/jss.v059.i10.
- Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Springer International Publishing : Imprint: Springer, Cham. 1 pp.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemond, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the Tidyverse. *JOSS.* 4:1686. doi:10.21105/joss.01686.

Data availability

The data and code are deposited at zenodo: <https://doi.org/10.5281/zenodo.4980068>

Competing Interest

The authors declare no competing or financial interests.

Author contributions

E.B. initiated the project, contributed experimental data, wrote code and wrote the manuscript. M.G. supervised the project, wrote code, reviewed code and edited the manuscript. J.G. wrote code and edited the manuscript.

Acknowledgements

We thank our colleagues at the Swammerdam Institute for Life Sciences, University of Amsterdam, for their interest and support. We thank dr. F.M.J. Jacobs for providing resources and accommodation for this project.

Funding

None