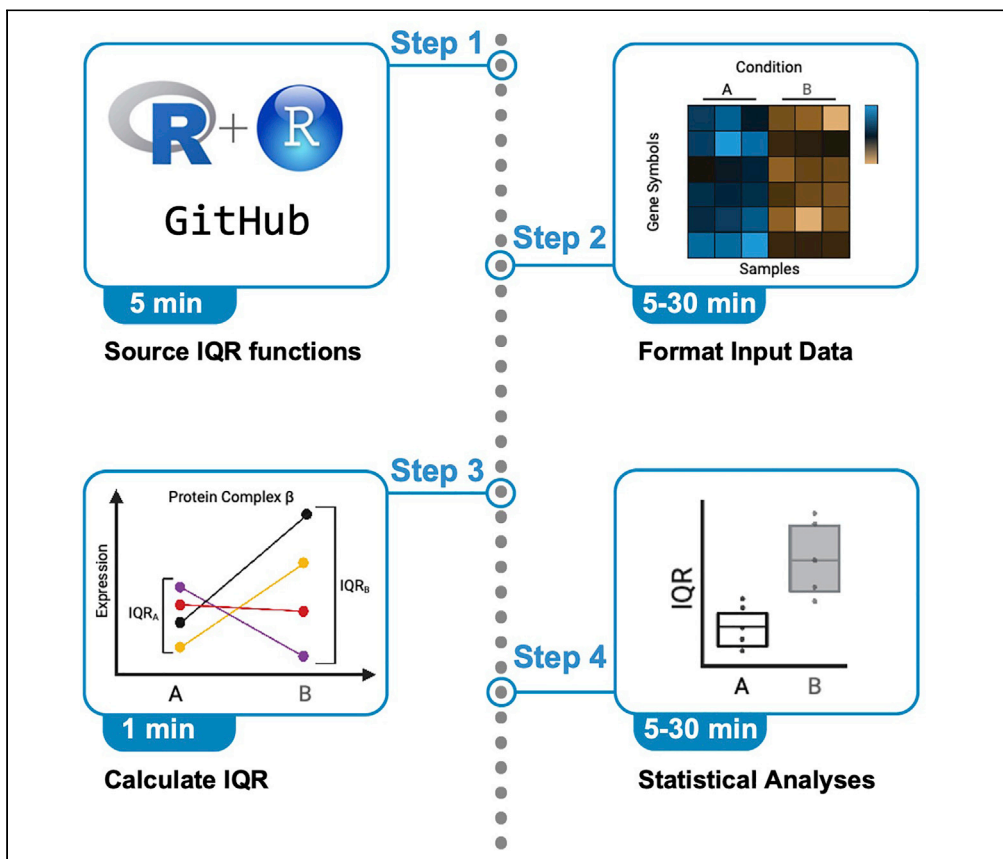


Protocol

Protocol for computationally evaluating the loss of stoichiometry and coordinated expression of proteins



Dysregulation of the transcriptional or translational machinery can alter the stoichiometry of multiprotein complexes and occurs in natural processes such as aging. Loss of stoichiometry has been shown to alter protein complex functions. We provide a protocol and associated code that use omics data to quantify these stoichiometric changes via statistical dispersion utilizing the interquartile range of expression values per grouping variable. This descriptive statistical approach enables the quantification of stoichiometry changes without additional data acquisition.

Stefan Hinz, Michael E. Todhunter, Mark A. LaBarge

shinz@coh.org (S.H.)
mlabarge@coh.org (M.A.L.)

Highlights

A protocol to quantify stoichiometry changes of protein complexes

Robust and versatile output based on interquartile range of expression

Lightweight R functions easily loaded into data pipeline via GitHub

Conveniently plot measured stoichiometry changes with ggplot2 wrapper function

Hinz et al., STAR Protocols 3, 101182

June 17, 2022 © 2022 The Author(s).

<https://doi.org/10.1016/j.xpro.2022.101182>



Protocol

Protocol for computationally evaluating the loss of stoichiometry and coordinated expression of proteins

Stefan Hinz,^{1,3,*} Michael E. Todhunter,^{1,3} and Mark A. LaBarge^{1,2,4,*}¹Department of Population Sciences, Beckman Research Institute, City of Hope, 1500 E. Duarte Road, Duarte, CA 91010, USA²Center for Cancer and Aging Research, City of Hope, 1500 E. Duarte Road, Duarte, CA 91010, USA³Technical contact⁴Lead contact*Correspondence: shinz@coh.org (S.H.), mlabarge@coh.org (M.A.L.)
<https://doi.org/10.1016/j.xpro.2022.101182>

SUMMARY

Dysregulation of the transcriptional or translational machinery can alter the stoichiometry of multiprotein complexes and occurs in natural processes such as aging. Loss of stoichiometry has been shown to alter protein complex functions. We provide a protocol and associated code that use omics data to quantify these stoichiometric changes via statistical dispersion utilizing the interquartile range of expression values per grouping variable. This descriptive statistical approach enables the quantification of stoichiometry changes without additional data acquisition. For complete details on the use and execution of this protocol, please refer to Hinz et al. (2021).

BEFORE YOU BEGIN

General considerations

This protocol uses numerical expression data to evaluate the changing stoichiometry of gene or protein complexes over a condition variable (e.g., age, time, or treatment). The script assumes that all genes or proteins are present at all tested conditions. The protocol assumes that the expression of gene/protein complexes must change concordantly to maintain stoichiometry (coordinated change of expression). A progressive reduction in the correlation between protein and mRNA causes a progressive loss of stoichiometry in several protein complexes, including ribosomes (Kelmer Sacramento et al., 2020), which is observed as an uncoordinated change of expression. The figure of merit is the interquartile range (IQR) of expression. IQR describes the difference between the 75th and 25th percentile ($x_{75} - x_{25}$) and, if the proteins complexed are unchanged between condition variables, the IQR stays unchanged, whereas the IQR changes given coordination changes (Figure 1). These analyses have been used to identify changes in proteostasis. (Hinz et al., 2021; Kelmer Sacramento et al., 2020)

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
R (V3.6.1)	CRAN	r-project.org
RStudio (V1.2.1335)	CRAN	rstudio.com
dplyr (V1.0.5)	CRAN	https://cran.r-project.org/web/packages/dplyr/index.html
ggrepel (V0.8.2)	CRAN	https://cran.r-project.org/web/packages/ggrepel/index.html
ggplot2 (V3.3.3)	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html

(Continued on next page)



Continued

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
ggsci (V2.9)	CRAN	https://cran.r-project.org/web/packages/ggsci/index.html
readr (V1.4.0)	CRAN	https://cran.r-project.org/web/packages/readr/index.html
Deposited data		
Numerical expression data	GitHub	https://raw.githubusercontent.com/LaBargeLab/IQR_test/main/example_gene_data.csv
Other		
Computer	NA	NA
R functions	GitHub/zenodo	https://github.com/LaBargeLab/IQR_test https://doi.org/10.5281/zenodo.5879559

STEP-BY-STEP METHOD DETAILS

Source IQR functions

⌚ Timing: 5 min

To run this protocol, a sourcing of provided convenience R functions from GitHub is required to calculate and visualize the IQR analyses. The functions, example data, and a tutorial are available at https://github.com/LaBargeLab/IQR_test.

1. Download IQR functions in R.

```
if(!require(devtools)){ install.packages("devtools") # If not already installed }
source_url("https://raw.githubusercontent.com/LaBargeLab/IQR_test/main/clean_functions.R")
```

Format input data

⌚ Timing: 5–30 min

The functions require input data with gene name, sample name, expression value, and grouping variable columns. Any parametric units are valid for expression values, such as counts-per-million, reads-per-kilobase transcript for RNA-seq, or protein abundance data. An example can be downloaded through the provided GitHub page.

2. Import expression matrix.

⚠ **CRITICAL:** Data must be in long format - i.e., every combination of gene name and sample name must have its own row.

```
urlfile <- ``https://raw.githubusercontent.com/LaBargeLab/IQR_test/main/example_gene_data.csv``
data <- read.csv(urlfile)
data
# Symbol name value variable
## <chr> <chr> <dbl> <chr>
# 1 gene1 sample1 99 A
```



```
# 2 gene1 sample2 131 A
# 3 gene1 sample3 74 A
# 4 gene1 sample4 145 A
#
#
#...
#10000 gene1000 sample10 0 B
.
```

Run analyses

⌚ Timing: 1 min

- Execute the stoichiometry function with the long format input data from step 2 to calculate the IQR for every sample by conditions and return a data frame with the results.

⚠ **CRITICAL:** The stoichiometry function input requires the following arguments:

symbol - character vector of gene symbols

expression - numeric vector of expression values

variable - character or factor vector with condition information

sample - character vector of sample IDs

geneset - character vector of interested genes (same nomenclature as symbol); if no geneset is supplied IQR analyses are performed based on all provided symbols

```
stoi <- stoichiometry( expression = data$value,
                      symbol = data$Symbol,
                      variable = data$variable,
                      geneset = c("gene1", "gene2", "gene3", "gene4", "gene5"),
                      sample = data$name)
```

- Use the output from the stoichiometry function as input for the plotting function for visualization.

```
stoi_plot(stoi)
```

EXPECTED OUTCOMES

The provided functions output a data frame with the IQR values on a sample level, and these can be conveniently plotted using boxplots (Figure 2).

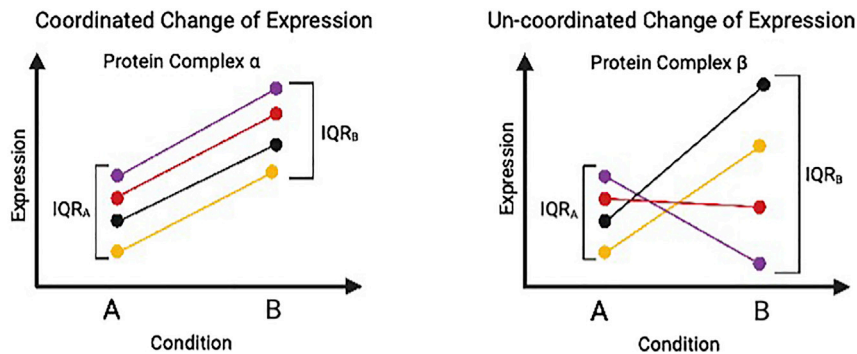


Figure 1. Examples of coordinated and uncoordinated change of expression

The provided function requires numerical expression data with n samples for m conditions (exemplar conditions A and B) to calculate the interquartile range.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical significance can be calculated using an appropriate statistical test, such as the Welch two-sample t-test for comparing two conditions or ANOVA for more than two conditions. The tests utilize the per-sample IQR data as input. Therefore, a power estimation is recommended to assess minimum sample size for meaningful analyses.

```
t.test(stoi$IQR ~ stoi$variable)
# Welch Two Sample t-test
# data : stoi$IQR by stoi$variable
# alternative hypothesis: true difference in means is not equal to 0
# -216.46803 -65.53197
# sample estimates:
# mean in group A mean in group B
#      36.6      177.6 95 percent confidence interval:
# -15.48451 22.68451
# sample estimates:
# mean in group A mean in group B
#      36.6      33.0
```

LIMITATIONS

The method described here is a statistical approach to assess the deregulation of protein complexes and does not replace confirmatory experiments.

TROUBLESHOOTING

Problem 1

The provided function does not load or work (step 3).

Potential solution

Confirm that all dependencies are installed (see [key resources table](#)). In case issues persist, an issue can be opened through the GitHub page.

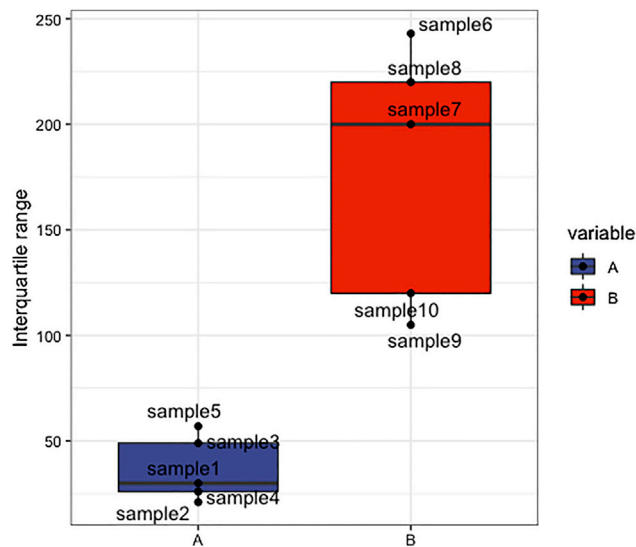


Figure 2. Results of example IQR analyses

Boxes of box whisker plot represent 25–75 percentile ranges, vertical lines represent $1.5 \times$ inter quartile range, and horizontal bars represent medians.

Problem 2

The expression data includes NA values (step 1).

Potential solution

Remove data with NA values or impute expression data if appropriate.

Problem 3

Expression data is in wide format not the required long format (step 2).

Potential solution

There are multiple tools to reshape data in R. The authors suggest the use of `pivot_longer()` function from the tidyR package.

Problem 4

Where to find curated genesets (step 3)?

Potential solution

There are multiple databases of curated genesets. The authors of this protocol recommend Molecular Signatures Database (MSigDB), Kyoto Encyclopedia of Genes and Genomes (KEGG), Drug Signatures Database (DSigDB), Gene Ontology Resource (GO), or HUGO Gene Nomenclature Committee as starting points.

Problem 5

The geneset does not match any symbols provided in the dataset (step 3).

Potential solution

Confirm that the symbol nomenclature matches between geneset and dataset symbol. In case of differing format, consider utilizing symbol conversion tools (e.g.,: `biomaRt`).

RESOURCE AVAILABILITY

Lead contact

Mark A. LaBarge, mlabarge@coh.org

Materials availability

This study did not generate new unique reagents.

Data and code availability

Code and data is available through the GitHub repository: https://github.com/LaBargeLab/IQR_test.

This repository has been archived at Zenodo: <https://doi.org/10.5281/zenodo.5879559>.

ACKNOWLEDGMENTS

This work was supported by awards from the Department of Defense/Army Breast Cancer Era of Hope Scholar Award (BC141351), City of Hope Center for Cancer and Aging to M.A.L.; National Institutes of Health/National Cancer Institute (NIH/NCI) grants R01CA237602, U01CA244109, R33AG059206, and R01EB024989 to M.A.L.; American Cancer Society – Fred Ross Desert Spirit Postdoctoral Fellowship (PF-21-184-01-CSM) to S.H. and American Cancer Society Postdoctoral Fellowship (131311-PF-18-188-01-TBG) to M.E.T. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Graphical abstract and [Figure 1](#) created in part with BioRender.com.

AUTHOR CONTRIBUTIONS

S.H. and M.A.L. conceived the protocol; S.H. and M.E.T. wrote code; and S.H. wrote the manuscript. All authors discussed and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Hinz, S., Manousopoulou, A., Miyano, M., Sayaman, R.W., Aguilera, K.Y., Todhunter, M.E., Lopez, J.C., Sohn, L.L., Wang, L.D., and LaBarge, M.A. (2021). Deep proteome profiling of human mammary epithelia at lineage and age resolution. *iScience* 24, 103026.
- Kelmer Sacramento, E., Kirkpatrick, J.M., Mazzetto, M., Baumgart, M., Bartolome, A., Di Sanzo, S., Caterino, C., Sanguanini, M., Papaevgeniou, N., Lefaki, M., et al. (2020). Reduced proteasome activity in the aging brain results in ribosome stoichiometry loss and aggregation. *Mol. Syst. Biol.* 16, e9596.