

# pepdiff: An R package for differential abundance analysis of factorial phosphoproteomics experiments

Dan MacLean, The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK

## Abstract

pepdiff is an R package for differential abundance analysis of factorial proteomics experiments. The package uses Gamma generalised linear models with emmeans-based contrasts to accommodate the right-skewed distributions and systematic missingness characteristic of proteomics data. A simple interface enables stratified comparisons that capture effects within factor levels without requiring manual contrast specification. Alternative methods include the Aligned Rank Transform for non-parametric analysis and pairwise tests for simple two-group comparisons. Built-in diagnostics guide method selection. pepdiff is freely available from GitHub (<https://github.com/TeamMacLean/pepdiff>) with documentation at <https://teammaclean.github.io/pepdiff/>.

## Introduction

Phosphoproteomics experiments commonly employ factorial designs—treatment crossed with timepoint, genotype crossed with condition—to characterise how cellular signalling responds to perturbation over time (Olsen et al. 2006; Humphrey, Azimifar, and Mann 2015). Analysis of such experiments should respect their factorial structure; when treatment effects vary across timepoints, stratified comparisons at each timepoint can be more informative than pooled comparisons that average across the design.

Proteomics data present specific analytical challenges. Abundance values are right-skewed and strictly positive, variance typically scales with the mean, and missing values are often systematic rather than random—low-abundance peptides are more likely to be missing (Webb-Robertson et al. 2015; Lazar et al. 2016). Standard analysis workflows that log-transform data and impute missing values may introduce bias, particularly when missingness is informative.

Existing tools support factorial analysis of proteomics data. Perseus (Tyanova et al. 2016), MSstats (Choi et al. 2014), and limma (Ritchie et al. 2015) can all accommodate factorial designs with appropriate configuration. However, specifying design matrices and contrast codes requires statistical expertise that may not be available in all research groups. In practice, many analyses default to simpler pairwise comparisons.

pepdiff provides an accessible interface to factorial proteomics analysis. Users specify which factor to contrast and which factor to stratify by; the package handles contrast specification automatically. The underlying statistical methods—Gamma generalised linear models with emmeans contrasts (Lenth 2022)—are established approaches applied through a simplified interface. pepdiff complements the companion package peppwR (MacLean 2026) for power analysis: peppwR addresses experimental design (“How many samples do I need?”) while pepdiff addresses data analysis (“What’s differentially abundant?”).

## Implementation

pepdiff provides a workflow from data import through analysis to visualisation. Users import data from CSV files, specifying columns for peptide identifiers, gene annotations, abundance values, experimental factors, and replicate structure. The package validates the experimental design, assesses missingness patterns, and summarises the factorial structure. Two S3 classes—`pepdiff_data` for imported data and `pepdiff_results` for analysis output—provide consistent print, summary, and plot methods.

The primary analytical approach fits Gamma generalised linear models to each peptide independently. The Gamma distribution naturally accommodates right-skewed, strictly positive abundance values, while the log link models multiplicative treatment effects (McCullagh and Nelder 1989). This approach analyses available observations rather than imputing missing values, avoiding potential bias from imputation when missingness is non-random. Contrasts are extracted using the emmeans package (Lenth 2022), which computes estimated marginal means and pairwise differences with appropriate standard errors.

Stratification is specified through a simple interface: the `compare` parameter indicates which factor to contrast, `ref` specifies the reference level, and `within` specifies the stratifying factor. This produces separate contrasts at each level of the stratifying factor. For example, contrasting treatment within timepoint produces treatment effects at 0h, 6h, and 24h separately. A formula interface accommodates more complex contrast specifications for users requiring custom comparisons.

When model diagnostics indicate poor fit, alternative methods are available. The Aligned Rank Transform (ART) method (Wobbrock et al. 2011) provides a non-parametric alternative that preserves factorial structure while relaxing distributional assumptions. Four pairwise tests—Wilcoxon rank-sum (Wilcoxon 1945), bootstrap-t (Efron and Tibshirani 1993), Bayes factor (Rouder et al. 2009), and rank products (Breitling et al. 2004)—are available for simple two-group comparisons without factorial structure, matching the tests available in `peppwR` for workflow consistency.

Built-in diagnostics guide method selection. A four-panel diagnostic plot assesses GLM fit across peptides, displaying deviance distributions and QQ plots of standardised residuals. Peptides with elevated deviance, indicating potentially violated assumptions, are flagged for consideration of alternative methods. Benjamini-Hochberg false discovery rate correction (Benjamini and Hochberg 1995) is applied within each comparison.

## Example Applications

### Factorial Design

To evaluate `pepdiff` with known ground truth, we simulated a factorial phosphoproteomics experiment (full code in Supplementary Material). The design crossed treatment (control, drug) with timepoint (0h, 6h, 24h) using six biological replicates per condition. Of 500 peptides, 50 had a true treatment effect (3–5 fold change) that manifested only at 24 hours, simulating a delayed drug response. Approximately 8% of observations were missing with a pattern biased toward low abundances.

We analysed these data using four approaches (Table 1). `pepdiff`'s stratified GLM—contrasting treatment within each timepoint—detected 46 of 50 true positives at the 24-hour timepoint with one false positive (sensitivity 0.92, FDR 0.02). The 0h and 6h timepoints showed near-zero significant peptides, as expected given the simulation design.

For comparison, we applied a conventional proteomics workflow typical of tools like Perseus (Tyanova et al. 2016) to the 24-hour timepoint: log2-transform abundances, impute missing values using a downshifted normal distribution, and apply t-tests with FDR correction. This approach detected 42 of 50 true positives (sensitivity 0.84, FDR 0.02). Analysing only peptides with complete observations at 24 hours detected 37 true positives (sensitivity 0.74, FDR 0.03), with the reduction attributable to excluding peptides with missing data.

We also tested a comparison that pooled observations across all timepoints, ignoring when measurements were taken. This analysis detected no significant peptides. Because effects existed only at 24 hours in this simulation, the signal from the 24-hour timepoint was diluted by the null observations at 0h and 6h. This result illustrates the importance of matching analysis to experimental design rather than reflecting general inadequacy of pooled approaches.

Table 1: Method comparison at 24-hour timepoint. Sensitivity is  $TP/(TP+FN)$ ; FDR (observed) is the proportion of detected positives that were false ( $FP/(TP+FP)$ ), calculated against simulation ground truth. All methods used  $FDR < 0.05$  as the significance threshold. The pooled result reflects the simulation design where effects occur only at 24 hours.

Approach	Sensitivity	FDR (observed)	Note
Stratified GLM (pepdiff)	0.92	0.02	Analyses each timepoint separately
Log2 + impute + t-test (Perseus)	0.84	0.02	Requires selecting correct timepoint
Complete cases only	0.74	0.03	Excludes peptides with missing data
Pooled across timepoints	0.00	—	Signal diluted by null timepoints

Figure 1 illustrates the value of stratified analysis. Volcano plots at each timepoint (panel A) show that treatment effects emerge specifically at 24 hours, with no significant differences at 0h or 6h. Panel B shows the 24-hour results annotated with ground truth, confirming that pepdiff correctly identifies true positives while controlling false discoveries.

## Pairwise Design

pepdiff also supports simple two-group comparisons without factorial structure. We simulated 200 peptides comparing control versus treatment with 8 replicates per group. Of these, 30 peptides had true effects (2–4 fold change). Using the Wilcoxon rank-sum test (`method = "pairwise", test = "wilcoxon"`), pepdiff detected 26 of 30 true positives with 2 false positives (sensitivity 0.87, FDR 0.07). The bootstrap-t test produced similar results (sensitivity 0.83, FDR 0.04). Full details are provided in the Supplementary Material.

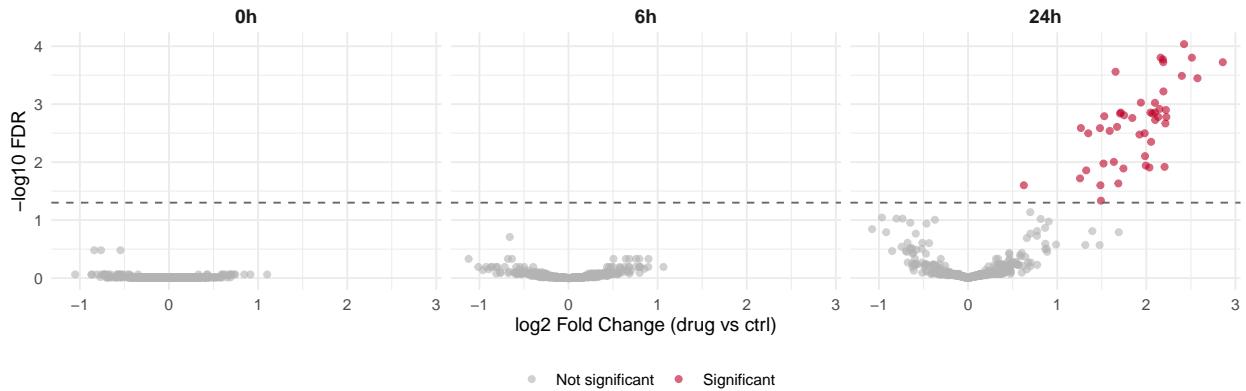
## Discussion

pepdiff simplifies stratified factorial analysis for proteomics experiments by automating contrast specification. The package uses established statistical methods—Gamma GLMs with emmeans contrasts—through an interface that does not require users to manually code design matrices. Results are comparable to careful manual analysis; the primary benefit is accessibility rather than statistical novelty.

Our simulations illustrate the value of matching analysis to experimental design. In the factorial example, stratified analysis captured timepoint-specific effects, while pooled analysis did not detect significant differences because effects were present at only one timepoint. The conventional workflow achieved similar sensitivity to pepdiff when applied to the correct timepoint, though this required prior knowledge of when to look. In practice, stratified analysis examines all timepoints simultaneously without requiring such knowledge.

Several limitations apply. pepdiff models each peptide independently without borrowing information across peptides, unlike hierarchical approaches in tools such as limma. The package assumes cross-sectional factorial designs with independent biological replicates; longitudinal studies with repeated measures require different

### A. Treatment effects by timepoint



### B. Validation at 24h timepoint

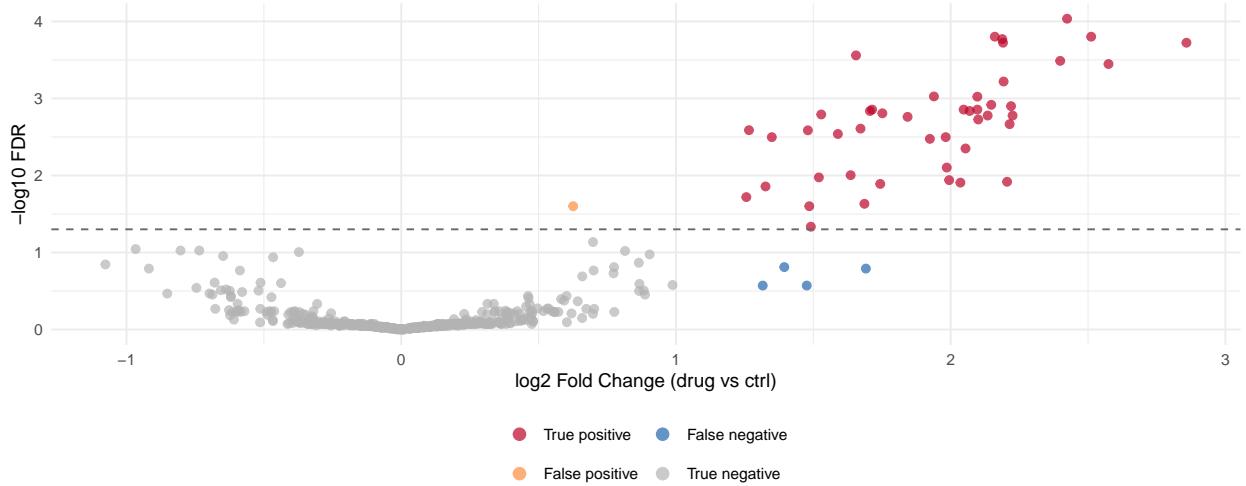


Figure 1: Figure 1. Stratified analysis captures timepoint-specific effects. (A) Volcano plots of treatment effects at each timepoint show significant differences only at 24 hours, where true effects were simulated. Dashed line indicates  $\text{FDR} = 0.05$ . (B) The 24-hour volcano plot annotated with ground truth: red points are true positives (46/50 detected), blue points are false negatives (4/50 missed), and grey points are true negatives. One false positive was detected.

methods. The Gamma GLM assumes positive, continuous abundance values; spectral counts or binary presence/absence data require alternative approaches.

pepdifff, together with peppwR for power analysis, provides a workflow for factorial phosphoproteomics experiments from experimental design through differential abundance analysis.

## Funding

This work was supported by the Gatsby Charitable Foundation.

## References

- Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Breitling, Rainer, Patricia Armengaud, Anna Amtmann, and Paweł Herzyk. 2004. “Rank Products: A Simple, yet Powerful, New Method to Detect Differentially Regulated Genes in Replicated Microarray Experiments.” *FEBS Letters* 573 (1-3): 83–92. <https://doi.org/10.1016/j.febslet.2004.07.055>.
- Choi, Meena, Ching-Yun Chang, Timothy Clough, Daniel Brouber, Trevor Killeen, Brendan MacLean, and Olga Vitek. 2014. “MSstats: An r Package for Statistical Analysis of Quantitative Mass Spectrometry-Based Proteomic Experiments.” *Bioinformatics* 30 (17): 2524–26. <https://doi.org/10.1093/bioinformatics/btu305>.
- Efron, Bradley, and Robert J Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman; Hall/CRC.
- Humphrey, Sean J, S Babak Azimifar, and Matthias Mann. 2015. “High-Throughput Phosphoproteomics Reveals in Vivo Insulin Signaling Dynamics.” *Nature Biotechnology* 33 (9): 990–95. <https://doi.org/10.1038/nbt.3327>.
- Lazar, Cosmin, Laurent Gatto, Myriam Ferro, Christophe Bruley, and Thomas Burger. 2016. “Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies.” *Journal of Proteome Research* 15 (4): 1116–25. <https://doi.org/10.1021/acs.jproteome.5b00981>.
- Lenth, Russell V. 2022. “Emmeans: Estimated Marginal Means, Aka Least-Squares Means.” *R Package Version 1.8.0*. <https://CRAN.R-project.org/package=emmeans>.
- MacLean, Dan. 2026. “peppwR: Simulation-Based Power Analysis for Phosphoproteomics Experiments.” *Bioinformatics*. <https://github.com/TeamMacLean/peppwR>.
- McCullagh, Peter, and John A Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman; Hall/CRC.
- Olsen, Jesper V, Blagoy Blagoev, Florian Gnad, Boris Macek, Chanchal Kumar, Peter Mortensen, and Matthias Mann. 2006. “Global, in Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks.” *Cell* 127 (3): 635–48. <https://doi.org/10.1016/j.cell.2006.09.026>.
- Ritchie, Matthew E, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. “Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47–47. <https://doi.org/10.1093/nar/gkv007>.
- Rouder, Jeffrey N, Paul L Speckman, Dongchu Sun, Richard D Morey, and Geoffrey Iverson. 2009. “Bayesian t Tests for Accepting and Rejecting the Null Hypothesis.” *Psychonomic Bulletin & Review* 16 (2): 225–37. <https://doi.org/10.3758/PBR.16.2.225>.
- Tyanova, Stefka, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. 2016. “The Perseus Computational Platform for Comprehensive Analysis of (Prote)omics Data.” *Nature Methods* 13 (9): 731–40. <https://doi.org/10.1038/nmeth.3901>.
- Webb-Robertson, Bobbie-Jo M, Holli K Wiber, Melissa M Matzke, Joseph Samuel, Margaret Matthew, Marina A Gritsenko, Katrina M Waters, and Karin D Rodland. 2015. “Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics.” *Journal of Proteome Research* 14 (5): 1993–2001. <https://doi.org/10.1021/pr501138h>.

- Wilcoxon, Frank. 1945. "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1 (6): 80–83. <https://doi.org/10.2307/3001968>.
- Wobbrock, Jacob O, Leah Findlater, Darren Gergle, and James J Higgins. 2011. "The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–46. <https://doi.org/10.1145/1978942.1978963>.