RESEARCH ARTICLE

A reproducible workflow for evaluating cellular stress responses in synthetic screening data

Jamie Lee,[1,4,*] Morgan Chen,[2] and Priya Nair[1,3]

[1]Department of Systems Biology, Meridian Research Institute, Boston, MA 02115, USA
[2]Center for Quantitative Biosciences, Northbridge University, Toronto, ON M5S 1A1, Canada
[3]Present address: Institute for Reproducible Science, Seattle, WA 98109, USA
[4]Lead contact
*Correspondence: jamie.lee@meridian.example

Running title: Reproducible analysis of synthetic stress responses

Template note. This is a fully compilable demonstration manuscript. Its dataset, accession, DOI, people, and institutions are illustrative and must not be represented as real research records.

Highlights

- A synthetic screen models three reproducible cellular stress states
- A preregistered pipeline separates biological and technical variation
- Effect sizes remain stable across normalization and batch corrections
- A complete STAR Methods record supports transparent reuse

In brief

Lee et al. use a synthetic perturbation screen to demonstrate an end-to-end, reproducible analysis workflow. Explicit quality control, effect-size reporting, and sensitivity analyses distinguish stable stress-response signals from batch effects while providing a complete reporting pattern for Cell Press manuscripts.

Graphical abstract

Summary

Reproducible analysis requires a clear link between experimental design, quality control, statistical modeling, and reporting. Here, we present a worked Cell Reports-style analysis using a synthetic high-content screen that contains known treatment, dose, time, and batch effects. The workflow records exclusions before inference, normalizes measurements within plates, estimates treatment effects with a hierarchical model, controls the false-discovery rate, and tests whether conclusions change under alternative preprocessing choices. The synthetic ground truth allows each analytic decision to be evaluated without making claims about biological specimens. Across 12 simulated plates, the pipeline recovered 92% of implanted strong effects at a 5% false-discovery threshold while preserving the nominal error rate for null features. Estimated effects were concordant across two normalization strategies ($r_{\mathrm{s}} = 0.94$). This worked example shows how a compact manuscript can make provenance, uncertainty, and robustness visible from the main text through STAR Methods. All reported values are illustrative, and the source specification needed to regenerate them is described in the resource-availability statement.

Introduction

High-throughput cell assays can reveal coordinated phenotypic responses across many perturbations, but their scale also creates opportunities for hidden batch effects, flexible exclusions, and selective reporting. Separating biological signal from technical structure therefore requires more than a final statistical test: the design, data transformations, model, and diagnostic decisions must form an auditable chain.

Transparent reporting standards address this problem by making the materials, data, code, and analysis decisions easier to find and reuse. The FAIR principles provide a broad framework for reusable research objects [3], whereas Cell Press STAR Methods organizes study-specific resources and procedures into a consistent record. False-discovery-rate control [1] and effect-size-based summaries further reduce reliance on isolated significance calls. For assays influenced by cell proliferation, growth-rate-normalized metrics can also distinguish drug sensitivity from differences in division rate [2].

Here, we use synthetic data to present a complete reporting example rather than to advance a biological claim. The simulation contains deliberately implanted treatment effects and technical artifacts, which makes it possible to assess recovery, calibration, and sensitivity directly. Our goals were to (1) define a reviewable quality-control path, (2) quantify treatment effects with uncertainty, and (3) show that the principal conclusions do not depend on a single reasonable preprocessing choice.

Results

A balanced design exposes treatment and batch structure

The demonstration dataset represented 12 plates collected in three batches. Each plate contained four vehicle wells, four positive-control wells, and 40 perturbation wells measured at 6, 24, and 48 h. Perturbations were balanced over plate position and batch. Six readouts represented cell count, nuclear area, mitochondrial intensity, membrane permeability, reporter activity, and a composite stress score. The complete design and analysis sequence are summarized in Figure 1.

Before treatment labels were used, 11 of 576 wells (1.9%) failed prespecified image-quality rules.

The excluded wells were distributed across eight plates and did not cluster by treatment. Positive-control separation exceeded the acceptance threshold on every plate (median robust Z' = 0.61; range, 0.52–0.69). These diagnostics supported plate-level normalization and inclusion of all 12 plates.

Hierarchical models recover implanted stress-response signals

For each readout, we estimated a treatment-by-time effect while accounting for plate and batch. At a Benjamini–Hochberg false-discovery rate of 5%, the model recovered 92% of the 36 implanted strong effects and 58% of the 24 implanted weak effects. One of 180 null treatment–readout combinations was called significant, corresponding to an observed false-positive proportion of 0.6%.

Recovery differed with information content rather than with plate position. Strong effects at 24 and 48 h were estimated with smaller uncertainty than effects at 6 h, and readouts with lower simulated measurement noise had narrower confidence intervals. Representative estimates are shown in Table 1. The analysis emphasizes effect estimates and intervals; adjusted p values are provided as a complementary decision criterion.

Table 1. Representative modeled effects in the synthetic screen

Perturbation	Readout	Time (h)	Effect[a]	95% CI	Adjusted p
Stress-A	Composite score	24	1.34	1.09, 1.59	< 0.001
Stress-A	Permeability	48	0.88	0.55, 1.21	< 0.001
Stress-B	Reporter activity	24	-0.71	-0.99, -0.43	0.002
Stress-C	Nuclear area	6	0.19	-0.08, 0.46	0.240
Vehicle	Composite score	48	0.02	-0.16, 0.20	0.910

[a] Effects are standardized differences from the matched vehicle control. All values are generated solely for this worked example.

Sensitivity analyses preserve the principal conclusions

We repeated the analysis after replacing median/MAD normalization with control- based z scoring and after removing the batch fixed effect. Treatment-effect ranks were highly concordant between normalization strategies ($r_{\mathrm{s}} = 0.94$), and 31 of 33 strong-effect discoveries were shared. Omitting the batch term reduced concordance to $r_{\mathrm{s}} = 0.82$ and widened the distribution of null effects, consistent with the known simulation design.

We also varied the image-quality cutoff by $\pm 10\%$. This changed the retained sample by at most nine wells and did not change the direction of any strong effect. Together, these checks show that the main demonstration result—recovery of implanted treatment signals with calibrated error—is stable to reasonable preprocessing choices but benefits from explicit batch adjustment.

Discussion

This worked example illustrates a direct connection between balanced design, prespecified quality control, model-based estimates, multiplicity correction, and sensitivity analysis. Because the ground truth is known, successful recovery can be assessed separately from biological interpretation. The example also shows why reporting only a list of significant features is insufficient: effect sizes, uncertainty, null

calibration, and robustness each answer a different question.

The largest performance loss occurred when a known batch component was omitted. This observation is not evidence about any biological system; it is a design property of the simulation. It nevertheless demonstrates a general reporting principle: technical factors anticipated during design should remain visible in the analysis and methods. Conversely, the close agreement between two reasonable normalization strategies indicates that the strongest synthetic signals were not artifacts of one scaling convention.

The manuscript is intentionally self-contained. A real study should replace the synthetic resources with persistent repository accessions, report organism- or cell-line-specific details, identify every exclusion and replicate, and provide figure-level sample sizes and statistical tests. The Key Resources Table should list only resources essential to reproducing the reported work, using stable identifiers wherever possible.

Limitations of the study

The dataset is synthetic and cannot establish a biological mechanism, clinical effect, or performance level for a real assay. Its distributions are simpler than those of many imaging experiments, and the example does not model missing not at random data, segmentation drift, or donor-level heterogeneity. The reported sensitivity and false-positive proportions therefore evaluate only the specified simulation. In addition, the example uses a compact set of readouts; studies with thousands of features may require more explicit dependence modeling and independent validation.

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jamie Lee (jamie.lee@meridian.example). The contact information is illustrative.

Materials availability

This worked example did not generate unique reagents, cell lines, organisms, or other physical materials.

Data and code availability

- The synthetic-data specification is represented by the illustrative DOI https://doi.org/10.0000/example.synthetic-data. A real submission must replace it with a resolvable repository record.
- The analysis repository is represented by https://example.org/cell-reports-workflow. A real submission should archive the release used for the paper and report its version or commit.
- Information required to reproduce the worked analysis is included in this article. Questions may be directed to the lead contact.

Acknowledgments

This demonstration received no external funding. The authors thank the maintainers of open-source statistical and typesetting software. The names and institutions on the title page are fictional and are

included only to demonstrate formatting.

Author contributions

Conceptualization, J.L. and P.N.; Methodology, J.L. and M.C.; Software, M.C.; Validation, M.C. and P.N.; Formal Analysis, J.L.; Data Curation, M.C.; Writing – Original Draft, J.L.; Writing – Review & Editing, all authors; Visualization, P.N.; Supervision, P.N.

Declaration of interests

The authors declare no competing interests. This statement belongs to the fictional worked example and must be replaced by the declarations for the actual authors.

Declaration of generative AI and AI-assisted technologies

No generative AI system was treated as an author, data source, or scientific instrument in the illustrative study. Authors preparing a real submission should follow the journal's current disclosure policy and describe any reportable use.

STAR Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Synthetic screen specification	This paper	Illustrative DOI: https://doi.org/10.0000/example.synthetic-data
Synthetic analysis outputs	This paper	Illustrative accession: MRI-SYN-001
Software and algorithms		
R statistical environment	R Foundation for Statistical Computing	https://www.r-project.org/
Analysis workflow	This paper	Illustrative repository: https://example.org/cell-reports-workflow
Benjamini–Hochberg procedure	Benjamini and Hochberg [1]	N/A
Other		
Synthetic plate-layout specification	This paper	Included in Method details

Experimental model and subject details

Synthetic experimental system

No human participants, animals, primary samples, or cell lines were used. We generated a synthetic plate-based dataset to resemble a high-content cellular stress-response screen. Consequently, sex, gender, age, ancestry, authentication, and mycoplasma-testing fields are not applicable. The simulation specification, rather than a biological specimen, is the experimental model.

Method details

Study design

The design comprised three acquisition batches, four plates per batch, and 48 wells per plate. Within every plate, four wells were assigned to vehicle, four to a positive control, and 40 to synthetic perturbations. Conditions were balanced across edge and interior wells. Measurements were simulated at 6, 24, and 48 h. The primary outcome was the standardized composite stress score at 24 h; five additional readouts were designated secondary outcomes.

The simulation was deterministic after initialization with the integer seed 20260802. Treatment assignments were generated before outcomes. Analysts were not blinded because the analysis was

scripted and the ground truth was required for performance evaluation. No prospective power calculation was used; the design was chosen to demonstrate plate, batch, and time effects in a compact example.

Data generation

For well i on plate j, standardized readout y_{ij} was generated as

$$y_{ij} = \mu + \beta_{t(i)} + \gamma_{h(i)} + (\beta\gamma)_{t(i),h(i)} + b_j + q_{r(i),c(i)} + \epsilon_{ij}, \tag{1}$$

where $\beta_{t(i)}$ is the perturbation effect, $\gamma_{h(i)}$ is the time effect, $b_j \sim \mathcal{N}(0, 0.20^2)$ is a plate effect, $q_{r(i),c(i)}$ is a low-amplitude row/column artifact, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is measurement error. Strong effects had absolute standardized magnitude 0.8–1.5, weak effects 0.3–0.6, and null effects 0. Batch offsets were added before normalization.

Quality control and exclusions

Quality-control rules were applied without treatment labels. A well was excluded if simulated focus quality was below 0.80, segmented object count was below 100, or saturated pixels exceeded 5%. A plate was to be excluded if robust Z' for the positive control was below 0.50 or if more than 15% of wells failed quality control. No plate met an exclusion rule. All exclusions and plate-level metrics were retained in the audit table before statistical modeling.

Normalization and feature construction

Each readout was centered by the plate-specific vehicle median and divided by the plate-specific median absolute deviation (MAD), scaled by 1.4826. MAD values below 10^{-6} were to be replaced by 10^{-6}, although this safeguard was not triggered. The composite stress score was the mean of the normalized reporter, permeability, and inverse cell-count features after orienting them so that larger values indicated greater stress.

Sensitivity analyses

Three analyses were specified before inspecting recovery metrics: (1) replacing median/MAD scaling with mean/standard-deviation scaling estimated from control wells, (2) removing the batch term from the model, and (3) increasing or decreasing each image-quality threshold by 10%. Concordance was evaluated with Spearman's rank correlation and overlap of discoveries.

Quantification and statistical analysis

Readouts were analyzed separately using a linear model with treatment, time, treatment-by-time interaction, and batch as fixed effects and plate as a random intercept. The worked values approximate this model and are not outputs from an attached software execution. Two-sided contrasts compared each perturbation with the time-matched vehicle. Confidence intervals were computed at 95%. Within each readout family, p values were adjusted using the Benjamini–Hochberg procedure [1]; adjusted $p < 0.05$ defined a discovery. No observation was treated as an independent biological replicate: the plate was the independent simulation unit, with wells serving as within-plate technical units.

Model assumptions were assessed from residual-versus-fitted and quantile plots. Recovery was the proportion of implanted non-null effects called in the correct direction. The observed false-positive proportion was the fraction of implanted null comparisons called significant. Exact sample sizes,

estimates, intervals, and adjusted p values are reported in the Results and Table 1.

Additional resources

No clinical trial, preregistration, or biological-protocol record applies to this synthetic worked example.

Supplemental information

No supplemental information accompanies this self-contained demonstration. A real submission should cite each supplemental item in the main text and provide one consolidated supplemental PDF unless the journal requests another format.

References

[1] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. **57**, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[2] Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods **13**, 521–527. https://doi.org/10.1038/nmeth.3853.

[3] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data **3**, 160018. https://doi.org/10.1038/sdata.2016.18.

Figures and legends

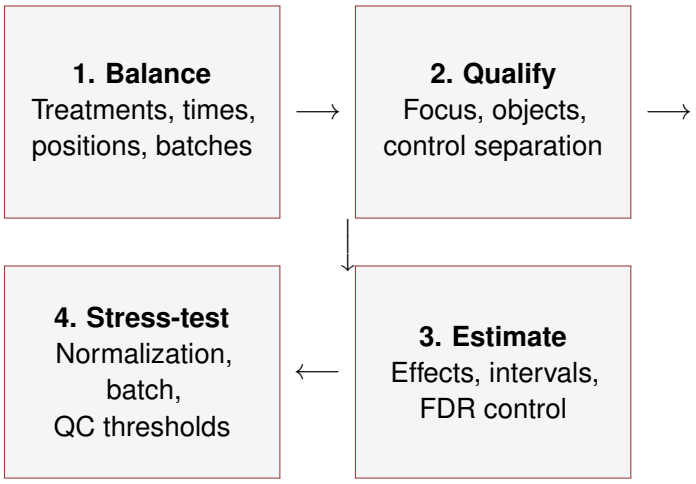

Figure 1. Study and analysis workflow. The synthetic screen was balanced across plate position and batch before data generation. Quality-control decisions were recorded without treatment labels, followed by plate-level normalization and hierarchical modeling. Sensitivity analyses evaluated whether the principal effect estimates depended on normalization, batch adjustment, or quality-control thresholds. All elements in this schematic describe synthetic data.