OXFORD

Systems biology

PlateDesigner: a web-based application for the design of microplate experiments

Maria Suprun [1] and Mayte Suárez-Fariñas[2,3,*]

[1]Department of Pediatrics, Allergy and Immunology and [2]Department of Population Health Science and Policy and [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on May 24, 2018; revised on September 9, 2018; editorial decision on September 30, 2018; accepted on October 8, 2018

Abstract

Summary: In biological assays, systematic variability, known as a batch effect, can often confound the effects of true biological conditions and has been well documented for a variety of high-throughput technologies. In microplate-based multiplex experiments, such as Luminex or OLINK assays, researchers need to consider both position and plate effects. Those effects can be easily accounted for if the experiments are properly designed, which includes randomization of the samples across multiple experimental runs. However, doing the ad hoc randomization becomes challenging when handling multiple samples. *PlateDesigner* is the first web-based application that provides randomization for microplate experiments, ensuring that the main principles of the experimental design, such as grouping samples from the same biological units and balancing the distribution of experimental conditions, are applied. Creating randomizations with *PlateDesigner* is simple and the results can be exported in a variety of formats, and easily integrated with microplate readers and statistical analysis software.

Availability and implementation: *PlateDesigner* is written in R/Shiny and is hosted online by the Center of Biostatistics at the Icahn School of Medicine at Mount Sinai. This application is freely available at platedesigner.net.

Contact: mayte.suarezfarinas@mssm.edu

1 Introduction

High-throughput technologies are used in many fields of biological research and allow measurement of protein and gene expression, epigenetic modifications and cell populations. While facilitating great discoveries, these studies can be quite erroneous if systematic variability is not accounted for, a phenomenon known as a batch effect (Lambert and Black, 2012).

Batch effects can be due to a variety of factors, including equipment and experimental fluctuations or sample processing habits. These effects have been well documented for a wide range of technologies, from microarray chips (Benito *et al.*, 2004) and RNA-sequencing lanes (Auer and Doerge, 2010) to the microplates (Browne *et al.*, 2013) used in the ELISA or multiplex assays, such as Luminex and OLINK. An unidentified batch effect will lead to increased variability of the data, making it more difficult to identify true biological effects;

since oftentimes the experimental variability can be higher than the biological variability. For example, Leek *et al.* (2010) found that 17% of sequence variability in the 1000 Genome Project was associated with the biological differences, whereas an immense 32% could be explained by the date when the samples were run.

Identification of batch effect relies on a careful exploratory analysis of documented potential biases using principal component analysis and then quantifying the percent of explained variance by PVCA (Li *et al.*, 2009). More recently, a formal statistical test has been introduced to identify the presence of batch effect (Nyamundanda *et al.*, 2017). Once the batch is identified, the variability of the data can be reduced by modelling the batch effect simultaneously with the biological effect of interest using a variety of available techniques, including sva and ComBat (Leek *et al.*, 2012), or linear models (Clarke *et al.*, 2013; Johnson *et al.*, 2007).

However, if there are confounding factors between batch and the biological variable, the two sources of variability are intertwined and cannot be reliably estimated. This scenario renders the data unusable, or worse - published spurious findings (Dressman *et al.*, 2007; Sebastiani *et al.*, 2011).

Such scenarios can be avoided if experiments are properly designed. Even though the majority of labs have rigorously standardized sample processing to ensure consistent experimental conditions, the application of experimental design principles has been lagging behind, in part due to the overconfidence in the precision of the technology (Lambert and Black, 2012). Different assays have a number of potential confounders that can be explored. For example, in RNA-sequencing one should consider accounting for separate library preparations, lanes, and dates when the samples were run (Auer and Doerge, 2010). In microplate-based experiments, which are commonly used in protein and epitope studies, both well position and use of multiple plates have been identified as potential sources of bias (Browne *et al.*, 2013; Harrison and Hammock, 1988). Fortunately, most of those effects can be easily addressed if the experimental conditions, samples, and plates are properly randomized, ensuring that samples are randomly assigned positions within plates and experimental conditions are well represented among all the plates (Qu, 2011; Roselle *et al.*, 2016). Though this task might seem trivial for someone with quantitative and informatics background, for the researchers with no coding experience or statistical understanding of randomization methodologies, attempting a successful ad hoc randomization will be next to impossible, even for a moderate number of samples. Thus, sample randomization often becomes secondary in hopes of addressing batch and location effects post hoc at the analysis stage. However, the analysis of any scientific study should be considered at the design phase, as even the most sophisticated statistical methods cannot always 'rescue' a poorly designed experiment. This golden rule readily embraced in clinical trials is not yet fully adopted in the lab.

To facilitate experimental design in microplate experiments we have developed *PlateDesigner*—a web application written in R/Shiny (Chang, 2017)—to streamline proper experimental design, and to ensure that biological effects can be adjusted for systematic variability.

2 Materials and methods

PlateDesigner is straightforward to use and is highly customizable. After uploading initial information about their samples, researchers can design the experimental setup by specifying the randomization units (RU), the number and position of replicates, and the type of control samples, such as background, quality controls and standard curve dilutions (Fig. 1A). In some cases, it is desirable that a certain experimental condition (i.e. exposure, disease, or study site) is adequately represented in each plate. In *PlateDesigner* this can be achieved by specifying a 'balancing condition'. Sampling weights for the RU are equal or defined by the frequency distribution of the 'balancing condition' (eg. the proportion of patients in the drug and placebo groups).

Plates are populated sequentially: (i) the number (N) of RUs that fits the plate is determined. (ii) N RU are then drawn from the pool of available RU with sampling weights. (iii) Samples from those N RU, controls and their replicates are positioned in each plate-well. Placement can be completely randomized or keeping the replicates together (horizontally or vertically), as to ease the technician's labor. The process iterates until the available RU pool is empty. This process
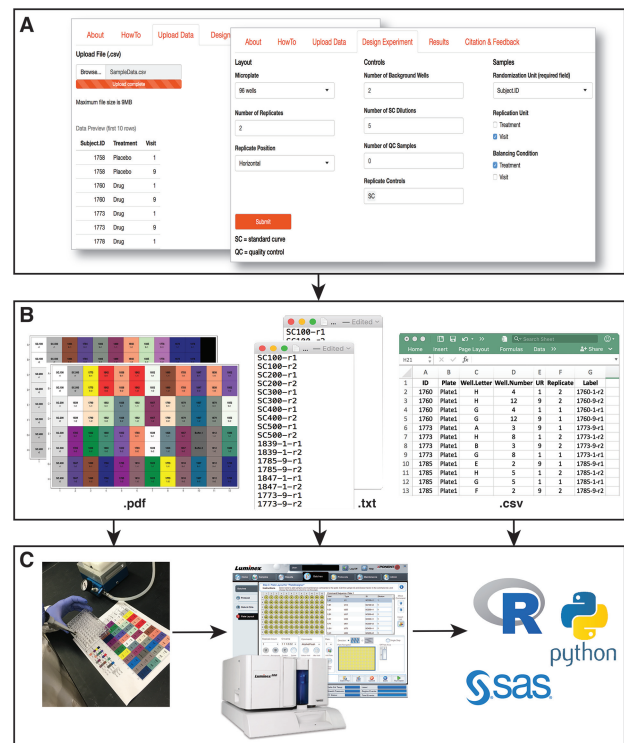

Fig. 1. *PlateDesigner* pipeline. (**A**) A snapshot of the application's 'Data Upload' and 'Design Experiment' web pages. (**B–C**) Outputs that are provided by the application and their utility: color-coded plate templates in a.pdf format for the researchers to add samples to a microplate; plain text (.txt) files for each randomized plate that can be uploaded to a plate-reader software; randomization information for each sample in a.csv format that can be used in the analysis

ensures that all samples from the same RU are placed in the same plate (eg. repeated visits from same patient) and experimental conditions are well-balanced. Furthermore, empty wells are kept together on the last plate, allowing for the re-use of the plate in future experiments.

The plate design can be exported in a variety of ways: (i) a pdf file with a convenient color-coded plate template that will aid the sample pipetting to the microplates; (ii) a csv file with sample identifiers and their assigned plate and well positions, easily read by statistical programs and languages, and (iii) a machine-readable text file to be uploaded to the microplate reader software, i.e. xPONENT for the Luminex-based assays (Luminex Corporation, Austin, Texas) (Fig. 1B and C). By doing so, researchers will be freed from the tedious and error prone manual entry of sample identifiers matched to well location into the assay's software.

3 Summary

PlateDesigner is easy to use and does not require background knowledge of randomization techniques or programming. The user just needs a web browser and is not required to install any programs or packages. Researchers will upload basic information about their samples, make menu selections for their experimental setup, and within seconds get an experiment plan that is properly designed, reducing data entry time and ensuring correct recording of experimental information. To our knowledge, *PlateDesigner* is the first and only web-based application currently available to researchers that generates randomization schemes for microplate experiments

and does not require researchers to have any coding background. *PlateDesigner* facilitates randomization, allowing seamless execution of well-designed, unbiased experiments.

Acknowledgements

Funding

References

Auer,P.L., and Doerge,R.W. (2010) Statistical design and analysis of RNA sequencing data. *Genetics*, **185**, 405–416.

Benito,M. *et al*. (2004) Adjustment of systematic microarray data biases. *Bioinformatics*, **20**, 105–114.

Browne,R.W. *et al*. (2013) Performance of multiplex cytokine assays in serum and saliva among community-dwelling postmenopausal women. *PLoS One*, **8**, e59498.

Chang,W. *et al*. (2017) shiny: Web Application Framework for R. Release R package version 1.0.5.

Clarke,D.C. *et al*. (2013) Normalization and statistical analysis of multiplexed bead-based immunoassay data using mixed-effects modeling. *Mol. Cell. Proteom*., **12**, 245–262.

Dressman,H.K. *et al*. (2007) An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J. Clin. Oncol*., **25**, 517–525.

Harrison,R.O., and Hammock,B.D. (1988) Location dependent biases in automatic 96-well microplate readers. *J. Assoc. Off. Anal. Chem*., **71**, 981–987.

Johnson,W.E. *et al*. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.

Lambert,C.G., and Black,L.J. (2012) Learning from our GWAS mistakes: from experimental design to scientific method. *Biostatistics*, **13**, 195–203.

Leek,J.T. *et al*. (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, **28**, 882–883.

Leek,J.T. *et al*. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet*., **11**, 733–739.

Li,J. *et al*. (2009) Principal variance components analysis: estimating batch effects in microarray gene expression data. In, *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*. John Wiley & Sons, Ltd. pp. 141-154.

Nyamundanda,G. *et al*. (2017) A novel statistical method to diagnose, quantify and correct batch effects in genomic studies. *Sci. Rep*., **7**, 10849.

Qu,X. (2011) Design and analysis of high-throughput screening experiments. *J. Syst. Sci. Complex*., **24**, 711–724.

Roselle,C. *et al*. (2016) Mitigation of microtiter plate positioning effects using a block randomization scheme. *Anal. Bioanal. Chem*., **408**, 3969–3979.

Sebastiani,P. *et al*. (2011) Genetic signatures of exceptional longevity in humans. *Science*, **333**, 404.