ECRONICON OPEN ACCESS

EC CLINICAL AND MEDICAL CASE REPORTS

Research Article

Correlation Analysis on Transcriptomes from Published Human Skin Studies Show Variations between Control Samples

Rebecca SY Teng[1], Jasmine CY Kwang[1], Angelena SQ Chin[1], Clara J Sander[1], Irwin ZL Ang[1], Jun Hang Foong[1], Kar Chi Cheong[1], Raphael YH Hon[1] and Maurice HT Ling[1,2]*

[1]*School of Applied Sciences, Temasek Polytechnic, Singapore*

[2]*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, HOHY PTE LTD, Singapore.

Received: May 11, 2020; **Published:** May 26, 2020

Abstract

Reproducibility has been shown to be a problem in many areas of science, leading to a "reproducibility crisis". Many studies had examined factors limiting experimental reproducibility and one of the factors suggested is the stability of control samples underpinning all experimental findings. This study examines the transcriptomes of the control samples from three published human skin studies using correlation analysis to evaluate the stability of clinical control samples. Our results show significant differences (t-test p-value < 5.4E-5, Mann-Whitney U p-value < 0.00001) between within data set correlations and between data set correlations, suggesting significant differences between control samples from different data sets. This may have potential implications on the interpretation of clinically important results.

Keywords: *Transcriptomes; Human Skin Studies; Reproducibility*

Introduction

Reproducibility is a hallmark of scientific and failures to reproduce scientific results in many disciplines had led to a reproducibility or replication crisis [1]. A survey by Nature magazine on 1576 scientists found that more than 70% of the scientists reported having failed to reproduce the results from another scientist while more than 50% had even failed to reproduce their own experiments, and 52% agreed on the presence of a significant reproducibility crisis [2]. A study in 2015 [3] estimated that approximately USD 28 billion per year was spent on irreproducible preclinical research in the United States alone, leading to a call for acknowledgement of the situation and improvement of research practices [4].

Failure to reproduce existing results is multifactorial, with selective reporting and pressure to publish being the most cited factors [2]. Other important factors are fraud, uncontrolled factors in the experiments and poor experimental design and/or statistical analysis [5]. Wilful data fabrication and falsification, which are interlinked to the pressure to publish [6], has been plaguing the research community for centuries [7, 8]. Uncontrolled factors are factors that may be out of the researcher's control; such as variability of reagents [9,10] and errors in source samples [11]. An aspect highlighted by Eisner [5] is the suitability and stability of control samples, which falls under experimental design and/or statistical analysis. Indeed, many experimental findings are predicated on the suitability and stability of control samples. Barrows., *et al.* [12] repeated the same experiment after 5 months using the same materials and reported significant differences in their transfection results despite their best efforts to ensure reproducibility. This clearly suggests variability in reagents and biological samples.

Given that control samples from medical studies usually originate as patient samples, variations in control samples are plausible. This is supported by a study [13] demonstrating high cell variability leading to poor reproducibility of primary insect explant cultures. There has been no study examining the stability of clinical controls to date. In this study, correlation analysis was performed on transcriptomes of the control samples from three published human skin studies to evaluate the stability of clinical control samples. Our results suggest significant differences (t-test p-value < 5.4E-5, Mann-Whitney U p-value < 0.00001) between control samples from different data sets.

Materials and Methods

Transcriptome data

Three data sets from published transcriptomic studies (GDS2935 [14], GDS4460 [15] and GDS4491 [16]) on human skin and assayed on Affymetrix Human Genome U133 Plus 2.0 Array (GPL570) were identified from NCBI Gene Expression Omnibus. The control samples were identified from each study for analysis. GDS2935 has 3 control samples (GSM144362, GSM144371 and GSM144376). GDS4460 has 10 control samples (GSM803586, GSM803589, GSM803592, GSM803595, GSM803598, GSM803601, GSM803604, GSM803607, GSM803610 and GSM803613). GDS4491 has 8 control samples (GSM815451, GSM815452, GSM815453, GSM815454, GSM815455, GSM815456, GSM815457 and GSM 815458).

Correlation analysis

Spearman's rank correlation coefficients within and between each data set were calculated. Correlations from between data sets were analyzed for mean differences compared to correlations within data sets using 2-samples t-test assuming unequal variances and Mann-Whitney U test. Spearman's rank correlation and Mann-Whitney U test were used as they do not assume normality but 2-samples t-test was used as Mann-Whitney U test requires more than 5 data points per sample.

Results and Discussion

The mean correlation coefficients of control samples within the same data set (n = 76) ranged from 0.915 to 0.986 while the mean correlation coefficients of control samples across data sets (n = 135) ranged from 0.845 to 0.926 (Figure 1 and table 1). Taking all the correlation coefficients from within data set comparison (Figure 1), the mean and median correlation coefficient are 0.944 and 0.920, respectively. For between data set comparisons, the mean and median correlation coefficient are 0.899 and 0.925, respectively. There is significant difference (t-test p-value = 1.5E-14, Mann-Whitney U p-value < 0.00001) between within data set correlations and between data set correlations. Considering each data set separately (Table 2), there are also significant differences (t-test p-value < 5.4E-5, Mann-Whitney U p-value < 0.00001) between within data set correlations and between data set correlations.

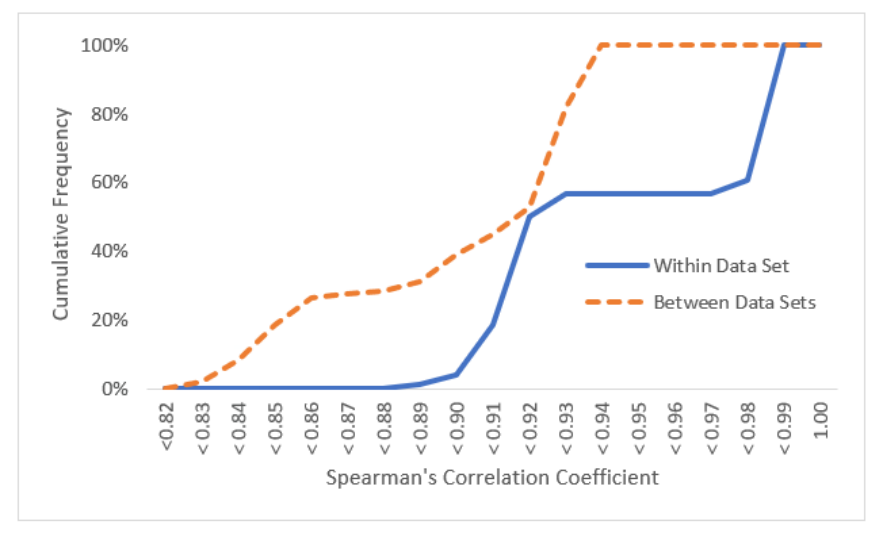

Figure 1: *Distribution of Spearman's correlation coefficients.*

Data Set	Count	Spearman's Rank Correlation Coefficient			
		Minimum	Maximum	Average	Standard Deviation
Within GDS2935	3	0.978	0.983	0.981	0.0029
Within GDS4460	45	0.887	0.988	0.915	0.0168
Within GDS4491	28	0.977	0.990	0.986	0.0032
GDS2935 vs GDS4491	24	0.879	0.911	0.898	0.0080
GDS2935 vs GDS4460	30	0.821	0.863	0.845	0.0103
GDS4460 vs GDS4491	81	0.902	0.937	0.926	0.0070

Table 1: *Spearman's rank correlations within and between data sets.*

Between Data Sets	Within Data Sets	t-test p-value	Mann-Whitney U p-value
GDS2935 vs GDS4491	GDS2935	3.7E-09	Not calculated as GDS2935 has less than 5 correlations
GDS2935 vs GDS4491	GDS4491	4.5E-30	< 0.00001
GDS2935 vs GDS4460	GDS2935	7.5E-11	Not calculated as GDS2935 has less than 5 correlations
GDS2935 vs GDS4460	GDS4460	7.2E-36	< 0.00001
GDS4460 vs GDS4491	GDS4460	5.4E-05	< 0.00001
GDS4460 vs GDS4491	GDS4491	4.5E-80	< 0.00001

Table 2*: Statistical test of correlations.*

Our null hypothesis can be stated as, the average correlation coefficients from pairwise comparisons of control samples within the same data set (n = 76) is statistically equal to the average correlation coefficients from pairwise comparisons of control samples across data sets (n = 135). However, our results suggest statistically significant differences (t-test p-value < 5.4E-5, Mann-Whitney U p-value < 0.00001) in the average correlation coefficients from within the same data set compared to that from across data sets, leading to the rejection of null hypothesis and the acceptance of the alternate hypothesis. Therefore, variations between the control samples across data sets is larger than that the variations between control samples within the same data set. The differences are significant (p-value < 0.00001) despite using a non-parametric statistical test, Mann-Whitney U test. Our results are supported by a current studies showing significant differences between collected samples [13] and when the same samples are test at different timings [12]. A meta-analysis of swaps self-collected by patient and clinician-collected samples showed assay-specific differences [17]. Moreover, a recent study suggested that the same sample underwent different experimental protocols may yield different results [18]. This suggest the potential for differences in control samples across difference studies.

Conclusion

As control samples are the basis on which experimental interpretations are constructed, statistically significant differences in control samples across multiple studies may have an impact on the interpretation of clinically important results. To the best of our knowledge, this study is the first that examined potential stability of control samples published by multiple studies. The main limitation of this study is the small number of data sets used and only one type of human sample, skin, was examined. Future work can expand on the both volume of data sets and the types of samples used.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

Rebecca SY Teng, Jasmine CY Kwang and Angelena SQ Chin have been contributed equally for the article.

Bibliography

1. S Laraway., *et al*. "An Overview of Scientific Reproducibility: Consideration of Relevant Issues for Behavior Science/Analysis". *Perspectives on Behavior Science* 42.1 (2019): 33-57.
2. M Baker. "1,500 Scientists Lift the Lid on Reproducibility". *Nature* 533.7604 (2016): 452-454.
3. LP Freedman., *et al*. "The Economics of Reproducibility in Preclinical Research". *PLOS Biology* 13.6 (2015): e1002165.
4. RA Poldrack. "The Costs of Reproducibility". *Neuron* 101.1 (2019): 11-14.
5. DA Eisner. "Reproducibility of Science: Fraud, Impact Factors And Carelessness". *Journal of Molecular and Cellular Cardiology* 114 (2018): 364-368.
6. MS Anderson., *et al*. "The Perverse Effects of Competition on Scientists' Work and Relationships". *Science and Engineering Ethics* 13.4 (2007): 437-461.
7. D Fanelli. "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data". *Plos One* 4.5 (2009): e5738.
8. MH Ling. "Science/Education Portraits III: Perceived Prevalence of Data Fabrication and/or Falsification in Research". *Advances in Biotechnology and Microbiology* 11.5 (2018): 555824.
9. S Thompson and D Chesher. "Lot-to-Lot Variation". *The Clinical Biochemist Reviews* 39.2 (2018): 51-60.
10. M Baker. "Reproducibility crisis: Blame it on the antibodies". *Nature* 521.7552 (2015): 274-276.
11. JR Lorsch., *et al*. "Fixing Problems with Cell Lines". *Science* 346.6216 (2014): 1452-1453.
12. NJ Barrows., *et al*. "Factors Affecting Reproducibility Between Genome-Scale Sirna-Based Screens". *Journal of Biomolecular Screening* 15.7 (2010): 735-747.
13. N Ogata and K Iwabuchi. "Relevant Principal Factors Affecting the Reproducibility of Insect Primary Culture". *In Vitro Cell and Developmental Biology - Animal* 53.6 (2017): 532-537.
14. MB Pedersen., *et al*. "Gene Expression Time Course in the Human Skin During Elicitation of Allergic Contact Dermatitis". *Journal of Investigative Dermatology* 127.11 (2007): 2585-2595.
15. JM Jensen., *et al*. "Gene Expression Is Differently Affected by Pimecrolimus and Betamethasone in Lesional Skin of Atopic Dermatitis". *Allergy* 67.3 (2012): 413-423.
16. M Suárez-Fariñas., *et al*. "Nonlesional Atopic Dermatitis Skin is Characterized by Broad Terminal Differentiation Defects and Variable Immune Abnormalities". *The Journal of Allergy and Clinical Immunology* 127.4 (2011): 954-964.
17. M Arbyn., *et al*. "Accuracy of Human Papillomavirus Testing on Self-Collected Versus Clinician-Collected Samples: A Meta-Analysis". *The Lancet Oncology* 15.2 (2014): 172-183.
18. J Fitzpatrick., *et al*. "Analysis of Platelet-Rich Plasma Extraction: Variations in Platelet and Blood Components between 4 Common Commercial Kits". *The Orthopaedic Journal of Sports Medicine* 5.1 (2017): 2325967116675272.

Volume 3 Issue 6 June 2020