AUTHOR CORRECTION

Open Access

Author Correction: scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured

Tianyi Sun[1], Dongyuan Song[2], Wei Vivian Li[3*] and Jingyi Jessica Li[1,4,5,6*]

*Correspondence:
vivian.li@rutgers.edu; jli@stat.ucla.edu; lijy03@g.ucla.edu

[1] Department of Statistics, University of California, 90095-1554 Los Angeles, CA, USA
[2] Interdepartmental Program of Bioinformatics, University of California, 90095-7246 Los Angeles, CA, USA
[3] Department of Biostatistics and Epidemiology, Rutgers School of Public Health, 08854 Piscataway, NJ, USA
[4] Department of Human Genetics, University of California, 90095-7088 Los Angeles, CA, USA
[5] Department of Computational Medicine, University of California, 90095-1766 Los Angeles, CA, USA
[6] Department of Biostatistics, University of California, 90095-1772 Los Angeles, CA, USA

Following the publication of the original paper [1], the authors identified two errors in the notation and formulas in the Methods section.

identifiable [111, 112]. To resolve this unidentifiability issue, scDesign2 uses the technique of distributional transform [113]: first draw $V_{ij} \sim \text{Uniform}[0,1]$ independently for $i=1,...,p$ and $j=1,...,n$; second define U_{ij} as:

$$U_{ij} = V_{ij}F_i(X_{ij} - 1) + (1 - V_{ij})F_i(X_{ij}). \tag{1}$$

Equation (1) in the online version.

The first error is the U_{ij} expression. Specifically, Eq. (1) should be corrected as

$$U_{ij} = (1 - V_{ij})F_i(X_{ij} - 1) + V_{ij}F_i(X_{ij}). \tag{1}$$

Hence, $\tilde{X}_{ij} \sim \tilde{F}_i$; that is, the continuous random variable $\tilde{X}_{ij}$ constructed from X_{ij} and V_{ij} follows $\tilde{F}_{ij}$. Defining $U_{ij} = \tilde{F}_i(\tilde{X}_{ij})$, we have $U_{ij} \sim \text{Uniform}[0,1]$ and:

$$\begin{aligned} U_{ij} &= \tilde{F}_i(X_{ij} + V_{ij}) = F_i(X_{ij} - 1) + V_{ij}\left(F_i(X_{ij}) - F_i(X_{ij} - 1)\right) \\ &= V_{ij}F_i(X_{ij} - 1) + (1 - V_{ij})F_i(X_{ij}), \end{aligned}$$

The end of the derivation of Eq. (1) in the online version.

Accordingly, at the end of the derivation of

$$U_{ij} = \tilde{F}_i(X_{ij} + V_{ij}) = F_i(X_{ij} - 1) + V_{ij}\big(F_i(X_{ij}) - F_i(X_{ij} - 1)\big)$$
$$= V_{ij}F_i(X_{ij} - 1) + (1 - V_{ij})F_i(X_{ij}),$$

The last line should be changed to

$$U_{ij} = \cdots$$
$$= (1 - V_{ij})F_i(X_{ij} - 1) + V_{ij}F_i(X_{ij}),$$

The second error is one notation in the sentence just before the above-corrected expression. In "Hence, $\tilde{X}_{ij} \sim \tilde{F}_i$; that is, the continuous random variable $\tilde{X}_{ij}$ constructed from X_{ij} and V_{ij} follows $\tilde{F}_{ij}$", at the end, $\tilde{F}_{ij}$ should be corrected as $\tilde{F}_i$.

The authors confirm that the notational errors corrected here had no effect on the conclusions or the performance of the method scDesign2.

Published online: 22 February 2023

Reference
1.   Sun T, Song D, Li WV, et al. scDesign2: a transparent simulator that generates high-fidelity single-cell gene expression count data with gene correlations captured. Genome Biol. 2021;22:163. https://doi.org/10.1186/s13059-021-02367-2.