Differential gene expression in blastocysts following pronuclear transfer

Edward H. Morrow[1] and Fiona C. Ingleby[1]

[1] Evolution, Behaviour and Environment Group, School of Life Sciences, University of Sussex, John Maynard Smith Building, Brighton, BN1 9QG, UK

Corresponding author:

Edward H. Morrow[1]

Email address: ted.morrow@sussex.ac.uk

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

Abstract

Nuclear transfer techniques (a.k.a. mitochondrial replacement therapies) are currently under development to provide a route to eliminating particular instances of mitochondrial disease from the germline. Before these kinds of techniques are implemented clinically it is of primary concern that their safety and efficacy is established. In a recent paper, Hyslop et al (2016) utilized a specific version of pronuclear transfer (PNT) to investigate the consequences for gene expression in the developing embryo, which may indicate whether or not developmental pathways have been perturbed. However, the study was only able to include a small number of blastocysts within each treatment group, although a larger number of single cell expression profiles from each blastocyst were acquired. Using simulated datasets we show that the size and experimental design of this study cannot provide conclusive evidence that expression profiles of manipulated or control samples are indistinguishable from one another due to low power.

Introduction

Two main methods of mitochondrial replacement - pronuclear transfer (PNT) and maternal spindle transfer (MST) - are currently under development as potential germline therapies for eliminating some forms of mitochondrial disease. Hyslop et al. (2016) examined the consequences for early stage embryos following an ‘early’ version of PNT (termed ePNT), where zygotes had completed meiosis but not yet undergone mitosis. Gene expression profiles were obtained from single cell samples of blastocysts created using 4 different main methods: ePNT of oocytes from two different unrelated women (heterologous, n = 9), unmanipulated controls (n = 3), and two types of procedural

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

controls – ePNT of oocytes from the same donor (autologous, n = 1), and ePNT of oocytes from two related sisters (homologous, n = 1). Including autologous and heterologous controls potentially enables the authors to disentangle the effects of the ePNT procedure itself from any effects that may arise from switching the nuclear genomes between different mitochondrial genetic backgrounds. This mitonuclear mismatching is a potential safety concern for the clinical implementation of any of the various versions of mitochondrial replacement therapy (Reinhardt, Dowling & Morrow, 2013; Dunham-Snary & Ballinger, 2015). RNAseq data from blastocyst-derived single cells were explored via principle component analysis (PCA), t-distributed stochastic neighbour embedding, and unsupervised hierarchical clustering. On the basis of these exploratory analyses, the authors concluded that gene expression levels were indistinguishable between control and ePNT blastocysts.

However, there a number of shortcomings to the analytical approaches undertaken. First, the power to detect differences between treatment groups is low due to the small number of biologically independent samples, which is at the level of blastocyst and not single cell sample. For instance, a test of the mitonuclear mismatching hypothesis would compare 9 heterologous versus a maximum of 2 autologous/homologous blastocysts. Second, no statistical modeling of treatment effects was conducted, which obviously precludes the possibility of making any conclusions about whether or not there are statistical differences overall, or between specific treatment groups.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

We investigated these issues using simulated datasets and subsequent power analysis and conclude that based on the number of samples included and the magnitude of effect sizes that might reasonably be expected to be present, the study is unable to provide clear evidence that the manipulated samples are indistinguishable from controls.

Methods

The power to detect differential gene expression between treatments was examined via simulation, where simulated datasets based on the experimental design used here were analysed for differences between treatments using a mixed effects linear model. In order to resemble a transcriptomic analysis of differential gene expression, simulations were ran in batches of 100 (i.e. analogous to analysing 100 genes) and the power was calculated from each batch as the percentage of significant tests. These batches were repeated to produce 100 power estimates from simulated data. Two sets of simulations were ran: Set 1 tested a range of effect sizes, and Set 2 tested a range of sample sizes. All analyses used R v3.2.1 and the 'lmer' mixed modelling function in the 'lme4' package (Bates et al., 2014). Methods are described below, and annotated R code that also generates two plots is provided in a supplement.

Each simulated dataset was set up by initially specifying a small effect size for differences in gene expression between cell types, variance estimates both within blastocyst and for error variance (both based on the real data from Hyslop et al. (2016), and the effect size for treatment. In Set 1, the effect size for treatment was tested for all values between 1 and 10, whereas in Set 2, the effect size for treatment was fixed at 2.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

The effect sizes as shown are unstandardized, but when standardised using the error variance specified in the models, i.e. with a standard deviation = 10, an effect size of 1 is approximately d = 0.1 (very small; see Cohen (2013)) and an effect size of 10 is approximately d = 1 (very large). Note that the real effect size for differential gene expression will vary widely from one gene to another, but is likely to be low on average across all genes.

Next, the experimental design for each simulated dataset was set up as a balanced design, based on the numbers of samples in Hyslop et al. (2016) (although the actual study is unbalanced). The first set of simulations used 8 blastocysts with 4 samples from each blastocyst (by comparison, Hyslop et al. (2016) successfully sequenced RNA from 10 grade A-D blastocysts, with between 1 and 11 samples sequenced from each). In the simulated data, samples were split across a fully-factorial design between 4 different cell types (primitive endoderm, epiblast, trophectoderm and ambiguous) and 4 different treatments (control, autologous, homologous and heterologous). These factors represent the 4 cell types and 4 treatments in Hyslop et al. (2016), although samples were unbalanced across these factors. As in the study, all samples from the same blastocyst were under the same treatment. Set 2 of simulations varied the total number of blastocysts, but scaled the experiment to have the same fully-factorial design as the Set 1 simulations. Note that simulations were re-ran with an unbalanced design that more closely matched the variable levels of replication in Hyslop et al. (2016), and very similar power estimates were obtained.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

To simulate the data, gene expression values were generated as the sum of cell type and treatment effects (calculated using the effect sizes), as well as blastocyst and error variance (see R code for details). The data was analysed in a mixed linear model as follows:

$$Y \sim T + C + B + \varepsilon$$

where Y is the simulated expression data, T and C are 4-level fixed factors representing treatment and cell type, respectively, and B is a random factor representing blastocyst ID. P values for the treatment effect were obtained via model simplification (Crawley, 2007). This simulation process was ran separately for treatment effect sizes 1-10 (assuming 8 blastocysts; Set 1), and then separately for 48, 96, 144, 192 and 240 blastocysts (assuming a treatment effect size of 2; Set 2). Results are shown as the mean of 100 power estimates for each effect size (Set 1) and the mean of 100 power estimates for each blastocyst sample size (Set 2), with 95% confidence intervals.

Results

The simulated analysis of differential expression between treatments, based on this experimental design, clearly demonstrates that reasonable statistical power to detect treatment effects would only be possible if: (i) effect sizes were unusually strong (Figure 1); or (ii) a far higher number of blastocysts were sequenced (Figure 2).

Conclusions

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

On the basis of the low power and the descriptive nature of the methods employed by Hyslop et al. (2016) the conclusion that blastocysts created via ePNT versus controls, or between the different ePNT treatments are indistinguishable from one another is premature until sufficient data is available to carry out statistical modelling.

REFERENCES

Bates D., Mächler M., Bolker B., Walker S. 2014. Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]*.

Cohen J. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Crawley MJ. 2007. *The R Book*. Wiley Publishing.

Dunham-Snary KJ., Ballinger SW. 2015. Mitochondrial-nuclear DNA mismatch matters. *Science* 349:1449–1450. DOI: 10.1126/science.aac5271.

Hyslop LA., Blakeley P., Craven L., Richardson J., Fogarty NME., Fragouli E., Lamb M., Wamaitha SE., Prathalingam N., Zhang Q., O'Keefe H., Takeda Y., Arizzi L., Alfarawati S., Tuppen HA., Irving L., Kalleas D., Choudhary M., Wells D., Murdoch AP., Turnbull DM., Niakan KK., Herbert M. 2016. Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* advance online publication. DOI: 10.1038/nature18303.

Reinhardt K., Dowling DK., Morrow EH. 2013. Mitochondrial Replacement, Evolution, and the Clinic. *Science* 341:1345–1346. DOI: 10.1126/science.1237146.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016

149 **Figure legends**

150

151 **Figure 1**

152 **Simulated power based on unstandardized effect sizes**

153 Results shown are the mean power estimate (±95% confidence intervals) for 100

154 simulations at each effect size (ranging from 1-10). Effect size simulations use a similar

155 experimental design and size as the experiment described in Hyslop et al. (2016).

156

157 **Figure 2**

158 **Simulated power based on datasets with varying number of blastocysts**

159 Results shown are the mean power estimate (±95% confidence intervals) for 100

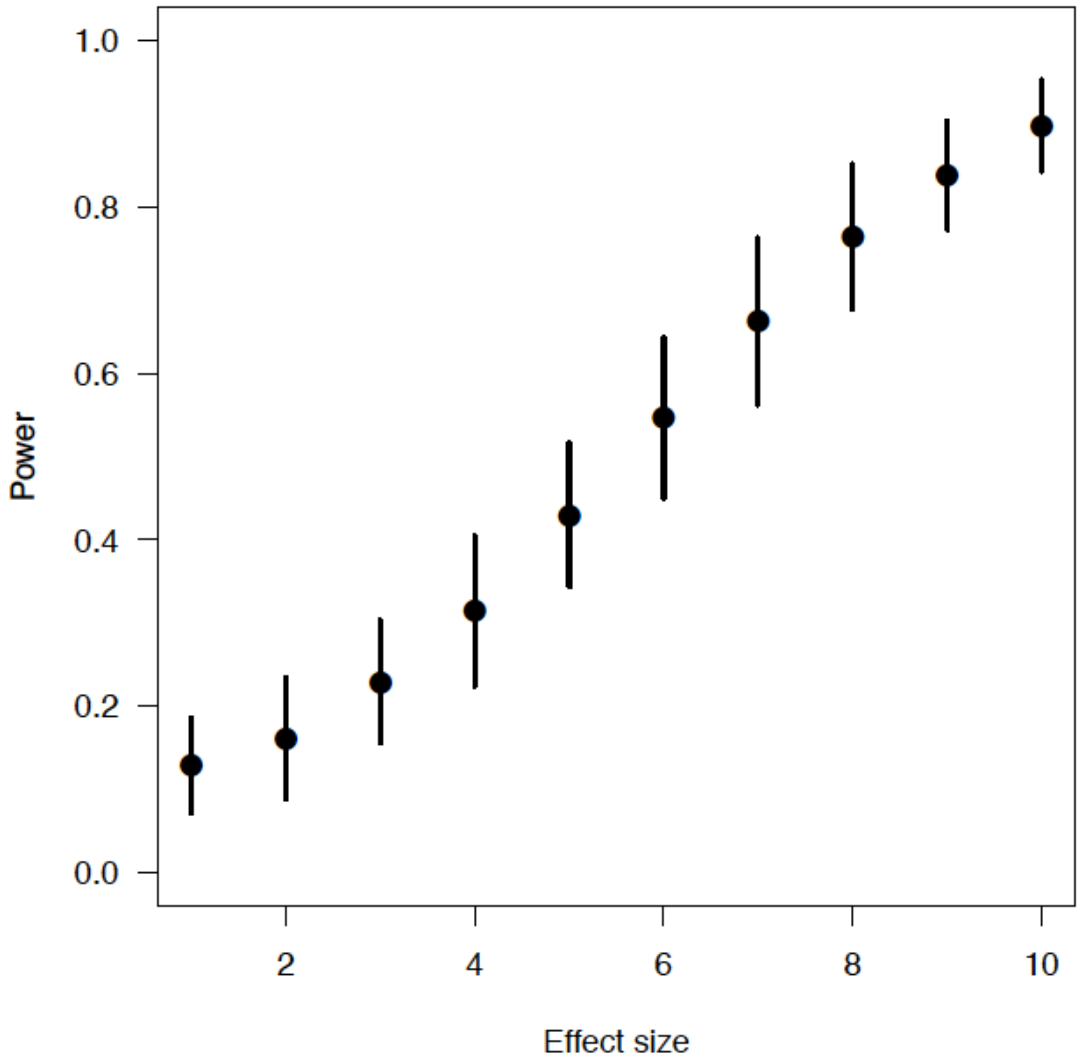
160 simulations for each number of blastocysts (ranging from 8-240). Blastocyst number

161 simulations use a similar experimental design as described in Hyslop et al. (2016), but

162 scale the design to increase the number of blastocyst samples.

163

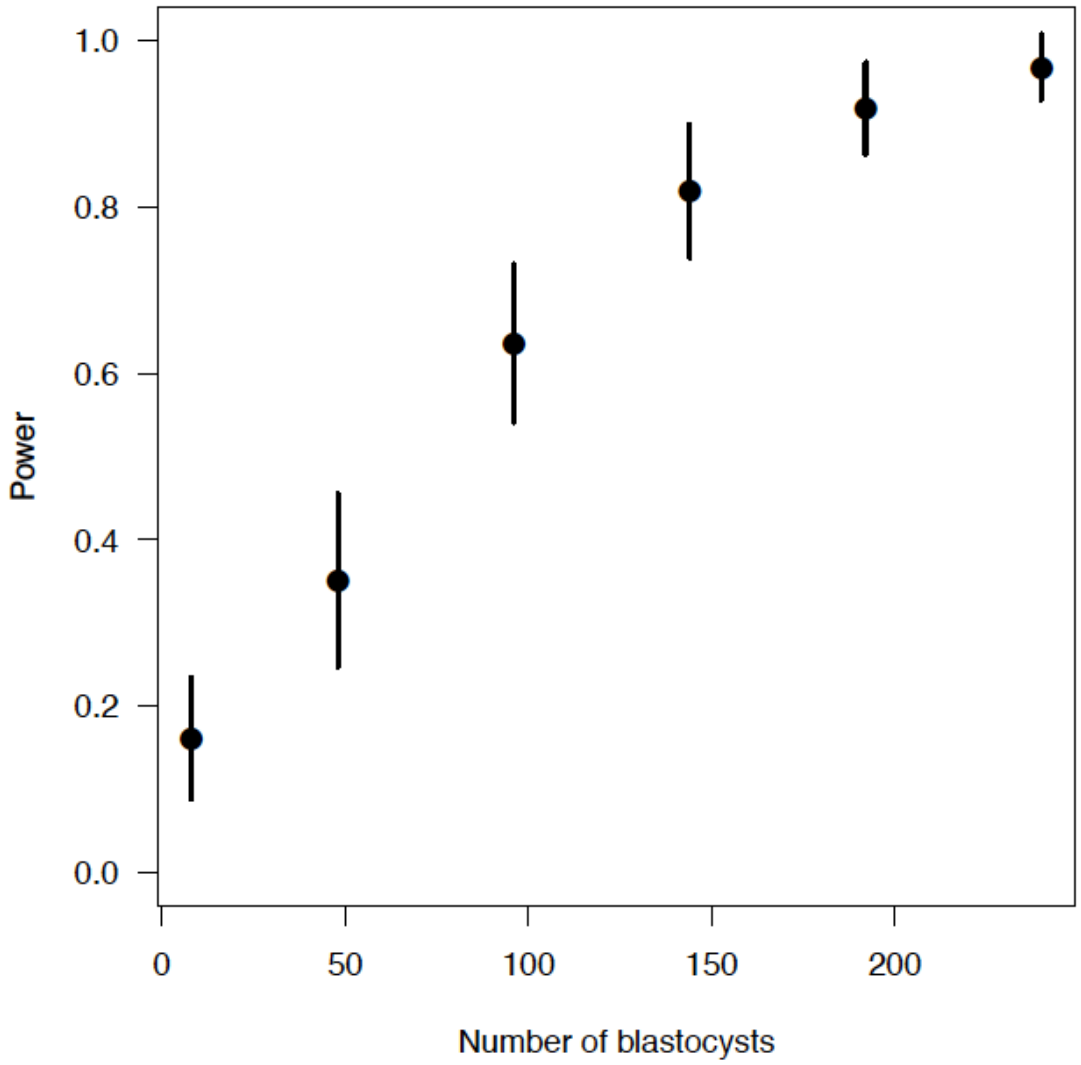
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016



164

165

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016



166

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2412v1 | CC BY 4.0 Open Access | rec: 2 Sep 2016, publ: 2 Sep 2016