SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

A transcriptomic study of Williams-Beuren syndrome associated genes in mouse embryonic stem cells

Rossella De Cegli[1*], Simona Iacobacci[1], Anthony Fedele[2], Andrea Ballabio[1,3,4,5] & Diego di Bernardo[1,6]

Williams-Beuren syndrome (WBS) is a relatively rare disease caused by the deletion of 1.5 to 1.8 Mb on chromosome 7 which contains approximately 28 genes. This multisystem disorder is mainly characterized by supravalvular aortic stenosis, mental retardation, and distinctive facial features. We generated mouse embryonic stem (ES) cells clones expressing each of the 4 human WBS genes (WBSCR1, GTF2I, GTF2IRD1 and GTF2IRD2) found in the specific delated region 7q11.23 causative of the WBS. We generated at least three stable clones for each gene with stable integration in the ROSA26 locus of a tetracycline-inducible upstream of the coding sequence of the genet tagged with a 3xFLAG epitope. Three clones for each gene were transcriptionally profiled in inducing versus non-inducing conditions for a total of 24 profiles. This small collection of human WBS-ES cell clones represents a resource to facilitate the study of the function of these genes during differentiation.

Background & Summary

Williams-Beuren Syndrome (WBS) is a neurodevelopmental disorder caused by a hemizygous deletion of 1.5 Mb segment occurring in approximately 95% of cases and a larger 1.84 Mb deletion observed in about 1 of 20 cases[1,2]. Clinical main features comprise, distinctive facial features (elfin face)[3,4], supravalvular aortic stenosis, connective tissue anomalies, hypertension, infantile hypercalcemia[5], dental, kidney and thyroid abnormalities, premature ageing of the skin[6], impaired glucose tolerance and silent diabetes[2,7]. The cognitive hallmark includes mental retardation, hypersensitivity to sound due to the absence of acoustic reflexes and hypersociability[8,9]. While the primary cause of WBS is well understood[10], we still know little about the molecular basis of the phenotype. The first genome-wide transcription study performed in primary fibroblasts from eight individuals with WBS resulted in set of candidate pathways mis-regulated in WBS possibly involved in associated phenotypes[2].

To facilitate the study of genes involved in WBS, we generated and transcriptionally profiled of mouse embryonic stem (ES) cells[11,12] with inducible expression of the three GTF-transcription factors (*GTF2IRD1*, *GTF2IRD2* and *GTF2I*) together with the translation initiation factor *Eif4h* (the human homolog is known as *WBSCR1*[13,14]. The ES properties to self-renew[15] and to differentiate in the three germ layers[16,17] have made these cells a unique *in vitro* system for studying the molecular mechanisms that regulate lineage specification. The three GTF-family members are all highly expressed in the brain. Mouse hemizygote models for *GTF2I* and *GTF2IRD1* present cognitive and behavioural phenotypes associated with WBS[3,4], moreover *GTF2I* deletion is known to be associated with increased sociability while the *GTF2I* duplication results in increased separation anxiety[18,19]. Targeted *Gtf2IRD1* knockout mouse is known to cause the up-regulation of growth factors and other genes involved in brain development and cellular proliferation which may be linked with the extreme thickening of the epidermis observed in the mouse model[20]. Moreover it has been reported that the transgenic expression of each of the three family members in skeletal muscle causes significant fiber type shifts[21]. Finally, *WBSCR1*, the human homolog of *Eif4h*, is known to contribute to neuroanatomical WBS deficits[22]: in *vivo* studies on knockout mice displayed

[1]Telethon Institute of Genetics and Medicine (TIGEM), via Campi Flegrei, 34, 80078, Pozzuoli, Naples, Italy. [2]Lysosomal Diseases Research Unit, South Australian Health and Medical Research Institute, Adelaide, SA, Australia. [3]Medical Genetics Unit, Department of Medical and Translational Science, Federico II University, Via Pansini 5, 80131, Naples, Italy. [4]Jan and Dan Duncan Neurological Research Institute, Houston, TX, 77030, USA. [5]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA. [6]Department of Chemical, Materials and Industrial Engineering, University of Naples 'Federico II', 80125, Naples, Italy. *email: decegli@tigem.it

growth retardation, a smaller brain volume, a reduction in both the number and complexity of neurons and severe impairments of fear-related associative learning and memory formation[22].

In a previous study on Down Syndrome, we generated a collection of mouse ES clones capable of the inducible expression of 32 mouse genes (orthologs of human chromosome 21 genes) under the control of the tetracycline-response element (tetO)[14]. Here we used the same approach exploiting the ROSA-TET system[23] to generate 12 mouse ES clones carrying the 4 Open Reading Frames (ORFs) of the GTF-transcription factors (*GTF2IRD1*, *GTF2IRD2* and *GTF2I*) together with the translation initiation factor *Eif4h* (Fig. 1). Three positive clones (Supplementary Fig. 1) for each gene were selected and grown in medium deprived of tetracycline (Tc) to perform an induction time course. RNA was extracted (Supplementary Fig. 2) from each clone at the time-point of maximal expression (24 hrs, Supplementary Fig. 3) and total RNA extracted from un-induced clones used as control. Total RNA was profiled by Affimetrix microarrays (the whole set of results is available in the GEO database [GSE96701[24]])[25,26]. This analysis was performed to detect differentially expressed genes (that is, in induced versus non-induced cells, Supplementary Fig. 4) in ES cells modeling the WBS.

Methods

Generation of recombinant WBS-ES clones. The generation of recombinant WBS-ES clones started with the modification of the cell line EBRTcH3 (EB3) as described in[14]. The cells were cultured in ES media supplemented with the leukemia inhibitory factor (LIF), at 37 °C in 5% CO_2. The ES media contained DMEM high glucose (Invitrogen, Catalog No. 11995–065) supplemented with 15% fetal bovine serum defined (hyClone, Catalog No. SH30070.03), 0.1 mM non-essential aminoacids (Gibco-Brl, Catalog No. 11140–050), 0.1 mM 2-mercaptoethanol (Sigma, Catalog No. M6250) and 1,000 U/ml ESGRO-LIF (Millipore, Catalog No. ESG1107). The basal expression of the transgenes in each stable clone was assured by the growth of the cells in this ES media +LIF supplemented with 1 μg/ml Tetracyclin (Tc) (Sigma, Catalog No. T7660). The selection of positive recombinant clones was assured by the growth of the cells the ES media (+LIF and +Tc) supplemented with 1.5 μg/ml of Puromycin (Puro, Sigma, Catalog No. P9620). After trypsinization (Trypsin-EDTA solution 10x, Sigma, Catalog No. T4174) the cells were plated 1 day before the nucleofection on a layer of 0.1% Gelatin (Gelatin Type I from porcine skin, Sigma) in 100-mm dishes (Nunc, Catalog No. 150350) in ES media (+LIF and +Tc). For Nucleofection protocol 2×10^6 cells were counted for each sample. Plasmids were prepared using Qiagen plasmid Midi-kit (Catalog No. 12145): 5–6 μg of pPthC vector in which the ORF of interest were cloned were incubated with 3 μg of pCAGGS-Cre vector[27] and 100 μl of Mouse ES Cell Nucleofector Kit (Amaxa, Catalog No. VPH-1001) was added to the plasmid mix as described in[14]. Cells were then incubated for 15 minutes at room temperature in complete medium and then plated. The day after, the cells were washed twice with PBS (Dulbecco Phosphate buffered Saline 1x, Gibco, Catalog No. 14190), and switched to selection media (+LIF +Tc +1.5 μg/ml Puro). The colonies were grown for one week before they were individually trypsinized and transferred to 96-well U-bottom plates (Nunc, Catalog No. 163320), then each clone was equally distributed among two gelatin-coated 48-well plates for selection in "selection media" (ES media +LIF and +150 μg/ml Hygromicin B in PBS, (Invitrogen, Catalog No. 10687–010)): the clones resistant to selection media and in parallel dead in selection media were isolated, replicated in 12-well plates (Nunc, Catalog No. 150628) and then in 6-well plates (Nunc, Catalog No. 140675) to extract the genomic DNA using standard conditions.

Cloning strategy. Each human coding sequence was cloned from the ATG to the stop codon without the 5′ and 3′ UTRs. For the 4 WBS ORFs, we cloned the longest annotated coding sequence (NM_001368300 for GTF2IRD2; NM_001199207 for GTF2IRD1; NM_032999 for GTF2I; NM_022170 for WBSCR1). The exchange vector pPthC-*Oct-3/4*was modified as described in[14] and the epitope 3xFLAG was designed to be in frame with the stop codon of each ORF. The cDNAs were amplified using the plasmids as templates by PCR in standard conditions: the forward and reverse primers were designed to include in the sequence the restriction sites recognized by the enzymes AscI and PacI at the 5′ and 3′ ends, respectively (Supplementary File 1). After digestion with specific restriction enzymes, the cDNA fragments were cloned into pTOPO-bluntII (Invitrogen, Catalog No. K2875J10), and then the cDNAs was cleaved by AscI-PacI. The fragments obtained by digestion were separated from pTOPO-bluntII as described in[14], the purified cDNA fragments were then inserted into the appropriately digested and purified pPthC vector[23]. The *Escherichia coli* positive clones were selected by enzymatic digestions and then sequenced by using the universal M13Fw primer and, for longer sequences, internal forward primers specific to the gene of interest.

Induction of transgene expression. The induction of the 4 transgenes' expression to Tc was verified on three positive clones for each WBS gene of interest. The complete removal of Tc results in sufficient induction of the Tet-off system as decribed in[28]. Cells to be induced were grown in medium deprived of Tc to perform a time course of induction (17, 24, 39 and 48 hours), by using the growth in Tc as control, time 0. Total RNA was extract at each time point of the time course and at time 0 and then 1 μg of each reverse-transcribed as described in[14]. The levels of each transcript was measured by Real-time RT-PCR experiments by using Light Cycler 480 Syber Green I Mastermix (Roche, Catalog No. 04887352001) for cDNA amplification and in LightCycler 480 II (Roche) for signal detection. RT-PCR results were analyzed using the comparative Ct method normalized against the housekeeping gene *Actin B* (refer to Supplementary File 2). All primer pair sequences used for RT-PCR are available in Supplementary File 2. For the time course of induction of the GTF2I clones (named C6, B3, A1), of the GTF2IRD1 clones (named D1, D2, D4), of the GTF2IRD2 clones (named A3, A4, A5) and of the WBSCR1 clones (named A3, A4, A1) refer to Supplementary File 3, Supplementary Fig. 3 and Online-only Table 1.

Microarray hybridization, data processing and statistical analysis. The preparation of the RNA' samples for the microarray hybridization on the Affymetrix GeneChip Mouse Genome 430_2 array was described

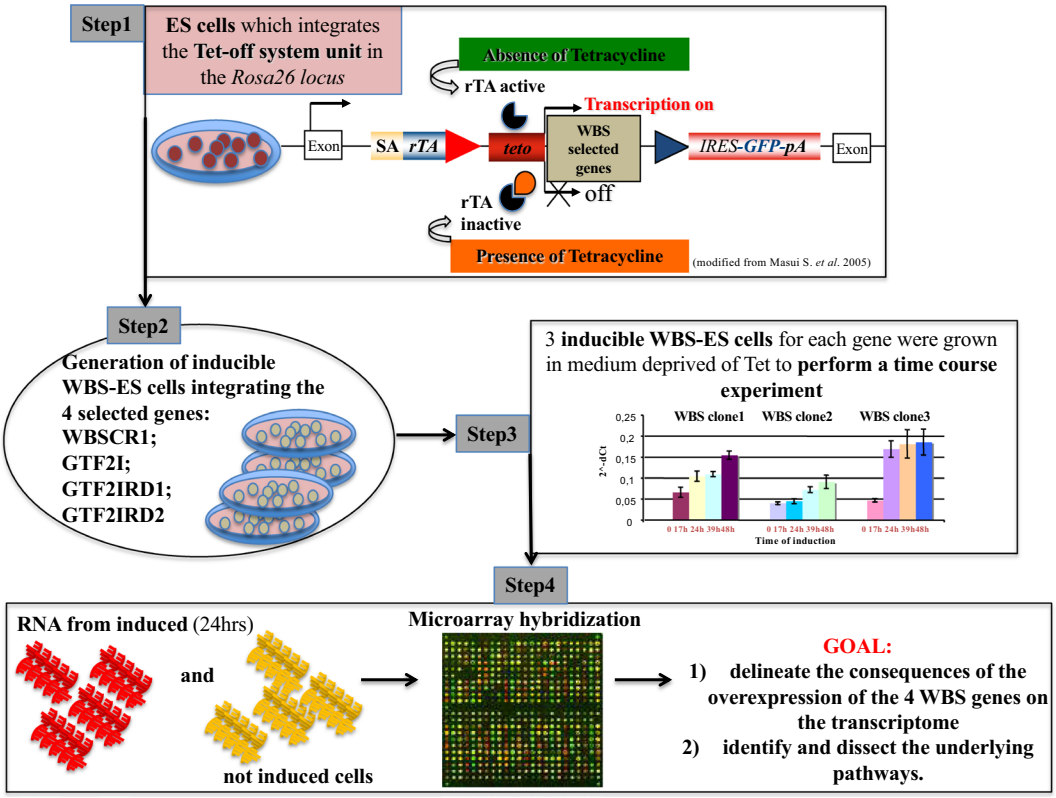

Fig. 1 Flowchart of experimental design of this study.

in[14]. Low-level analysis was performed by robust multiarray average (RMA) implemented using the RMA function of the Affymetrix package of the Bioconductor project[29,30] in the R programming language[31]. The low-level analysis for the BAMarray tool (v3.0) was performed using the MAS5 method as described in[14] and implemented using the corresponding function of the same Bioconductor package. For each gene, a *t*-test was used on RMA normalized data to determine the differentially expressed genes (induced versus uninduced). *P*-value adjustment for multiple comparisons was done with the FDR of Benjamini-Hochberg[32] (threshold FDR <0.05, refer to Supplementary File 4 and Supplementary Fig. 4).

Accession codes. The whole set of results is available in the GEO database[25,26] as "A transcriptomic study of Williams-Beuren syndrome associated genes in mouse embryonic stem", SuperSerie code GSE96701[24] (Supplementary File 4, Supplementary Fig. 4 and Online-only Table 1). The title of the SuprSeries is "Expression data from inducible ES stable cell line overexpressing the human GTF2IRD1, GTF2IRD2, WBSCR1, or GTF2I". In details: 1) GSE95267 refers to expression data from inducible ES stable cell line overexpressing specifically the human gene GTF2IRD1; 2) GSE95268 refers to expression data from inducible ES stable cell line overexpressing specifically the human gene GTF2IRD2; 3) GSE95269 refers to expression data from inducible ES stable cell line overexpressing specifically the human gene WBSCR1; 4) GSE95270 refers to expression data from inducible ES stable cell line overexpressing specifically the human gene GTF2I Fig. 1.

Data Records

The whole set of results is available in the GEO database[25,26] as "A transcriptomic study of Williams-Beuren syndrome associated genes in mouse embryonic stem", SuperSerie code GSE96701[24].

Technical Validation

The overexpression of the 4 selected WBS genes was based on the inducible expression by means of a tetracycline-repressible promoter (tet-off system). The first validation of the system was based on the cloning of the *luciferase* (*Luc*) into the exchange vector as described in[14], the second was the establishment of the expression of the *YFP* reporter gene, which is separated from the *Luc* gene in the recombinant locus by an *IRES* sequence, by detecting a comparable level of the *YFP* expression and protein accumulation following induction[14]. The study of the growth properties of our mES line (EB3) compared to the parental line (E14) (data not shown) and the ability of these cells to differentiate in the three main germ layers was also performed in[14]: in details the down-regulation of the pluripotens' marker Oct3/4 was also confermed in the EB3 as well as a farther induction of the mesodermal (Brachyury), ectodermal (Gfap) and endodermal (Afp) markers during mES differentiation. Collectively these data suggest that the system we chose allows the efficient and long-term overexpression of the transgene in a dose and time-dependent manner. It is therefore suitable for systematic expression of WBS cDNAs. The positive

clones overexpressing the 4 selected WBS genes were identified by PCR using the primer pair used in previous studies[13,14]: 5′-GCATCAAGTCGCTAAAGAAGAAAG-3′ and 5′-GAGTGCTGGGGCGTCGGTTTCC-3′ (Supplementary Fig. 1).

Code availability

Codes that were used for data processing are included in the Methods and available as supplementary material (Supplementary File 1 includes the sequences Asc1-Pac1 of the 4WBS ORFs; Supplementary File 2 the Primers used for RT-PCR_WBS). The whole set of results is available in the GEO database[25,26] as "A transcriptomic study of Williams-Beuren syndrome associated genes in mouse embryonic stem", SuperSerie code GSE96701[24] (Supplementary File 4).

Received: 24 July 2019; Accepted: 11 October 2019;
Published online: 06 November 2019

References

1. Ewart, A. K. *et al*. Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* **5**, 11–16, https://doi.org/10.1038/ng0993-11 (1993).
2. Henrichsen, C. N. *et al*. Using transcription modules to identify expression clusters perturbed in Williams-Beuren syndrome. *PLoS Comput Biol* **7**, e1001054, https://doi.org/10.1371/journal.pcbi.1001054 (2011).
3. Tassabehji, M. *et al*. GTF2IRD1 in craniofacial development of humans and mice. *Science* **310**, 1184–1187, https://doi.org/10.1126/science.1116142 (2005).
4. Antonell, A. *et al*. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. *J Med Genet* **47**, 312–320, https://doi.org/10.1136/jmg.2009.071712 (2010).
5. Sindhar, S. *et al*. Hypercalcemia in Patients with Williams-Beuren Syndrome. *J Pediatr* **178**, 254–260 e254, https://doi.org/10.1016/j.jpeds.2016.08.027 (2016).
6. Kozel, B. A. *et al*. Skin findings in Williams syndrome. *Am J Med Genet A* **164A**, 2217–2225, https://doi.org/10.1002/ajmg.a.36628 (2014).
7. Pober, B. R. Williams-Beuren syndrome. *N Engl J Med* **362**, 239–252, https://doi.org/10.1056/NEJMra0903074 (2010).
8. Jarvinen, A., Korenberg, J. R. & Bellugi, U. The social phenotype of Williams syndrome. *Curr Opin Neurobiol* **23**, 414–422, https://doi.org/10.1016/j.conb.2012.12.006 (2013).
9. Goldman, K. J., Shulman, C., Bar-Haim, Y., Abend, R. & Burack, J. A. Attention allocation to facial expressions of emotion among persons with Williams and Down syndromes. *Dev Psychopathol* **29**, 1189–1197, https://doi.org/10.1017/S0954579416001231 (2017).
10. Osborne, L. R. *et al*. Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. *Genomics* **36**, 328–336, https://doi.org/10.1006/geno.1996.0469 (1996).
11. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
12. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634–7638 (1981).
13. De Cegli, R. *et al*. Reverse engineering a mouse embryonic stem cell-specific transcriptional network reveals a new modulator of neuronal differentiation. *Nucleic Acids Res* **41**, 711–726, https://doi.org/10.1093/nar/gks1136 (2013).
14. De Cegli, R. *et al*. A mouse embryonic stem cell bank for inducible overexpression of human chromosome 21 genes. *Genome Biol* **11**, R64, https://doi.org/10.1186/gb-2010-11-6-r64 (2010).
15. Smith, A. G. Embryo-derived stem cells: of mice and men. *Annual review of cell and developmental biology* **17**, 435–462 (2001).
16. Suda, Y., Suzuki, M., Ikawa, Y. & Aizawa, S. Mouse embryonic stem cells exhibit indefinite proliferative potential. *Journal of cellular physiology* **133**, 197–201 (1987).
17. Palmqvist, L. *et al*. Correlation of murine embryonic stem cell gene expression profiles with functional measures of pluripotency. *Stem cells (Dayton, Ohio)* **23**, 663–680 (2005).
18. Walton, J. R., Martens, M. A. & Pober, B. R. The proceedings of the 15th professional conference on Williams Syndrome. *Am J Med Genet A* **173**, 1159–1171, https://doi.org/10.1002/ajmg.a.38156 (2017).
19. Mervis, C. B. *et al*. Duplication of GTF2I results in separation anxiety in mice and humans. *Am J Hum Genet* **90**, 1064–1070, https://doi.org/10.1016/j.ajhg.2012.04.012 (2012).
20. Corley, S. M. *et al*. RNA-Seq analysis of Gtf2ird1 knockout epidermal tissue provides potential insights into molecular mechanisms underpinning Williams-Beuren syndrome. *BMC Genomics* **17**, 450, https://doi.org/10.1186/s12864-016-2801-4 (2016).
21. Palmer, S. J. *et al*. GTF2IRD2 from the Williams-Beuren critical region encodes a mobile-element-derived fusion protein that antagonizes the action of its related family members. *J Cell Sci* **125**, 5040–5050, https://doi.org/10.1242/jcs.102798 (2012).
22. Capossela, S. *et al*. Growth defects and impaired cognitive-behavioral abilities in mice with knockout for Eif4h, a gene located in the mouse homolog of the Williams-Beuren syndrome critical region. *Am J Pathol* **180**, 1121–1135, https://doi.org/10.1016/j.ajpath.2011.12.008 (2012).
23. Masui, S. *et al*. An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res* **33**, e43 (2005).
24. *Gene Expression Omnibus*, https://identifiers.org/geo:GSE96701 (2019).
25. Barrett, T. & Edgar, R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. *Methods Mol Biol* **338**, 175–190, https://doi.org/10.1385/1-59745-097-9:175 (2006).
26. Chen, G. *et al*. Restructured GEO: restructuring Gene Expression Omnibus metadata for genome dynamics analysis. *Database* **2019**, bay145, https://doi.org/10.1093/database/bay145 (2019).
27. Araki, K., Imaizumi, T., Okuyama, K., Oike, Y. & Yamamura, K. Efficiency of recombination by Cre transient expression in embryonic stem cells: comparison of various promoters. *Journal of biochemistry* **122**, 977–982 (1997).
28. Rennel, E. & Gerwins, P. How to make tetracycline-regulated transgene expression go on and off. *Anal Biochem* **309**, 79–84 (2002).
29. Heller, D. S. *et al*. Demonstration of her-2 protein in cervical carcinomas. *J Low Genit Tract Dis* **7**, 47–50 (2003).
30. Gentleman, R. C. *et al*. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).
31. Piccolboni, D., Ciccone, F., Settembre, A. & Corcione, F. The role of echo-laparoscopy in abdominal surgery: five years' experience in a dedicated center. *Surg Endosc* **22**, 112–117, https://doi.org/10.1007/s00464-007-9382-x (2008).
32. Klipper-Aurbach, Y. *et al*. Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. *Med Hypotheses* **45**, 486–490 (1995).

Acknowledgements
We thank Gilda Cobellis and Nicoletta D'Alessio for technical assistance in the generation of WBS transgenic mouse ES cell lines. We thank Antonio Romito and Patrick Descombes for technical support. We thank Dr Hitoshi Niwa for providing the parental cell line EB3 and the exchange vector pPthC-*Oct-3/4*. We thank Dr Lucia Perone and the Cell Culture and Cytogenetics Core of Tigem for the karyotyping of mouse ES clones. This work was supported by the FP7 European Union grant 'Aneuploidy' (Contract Number 037627), the Swiss Science Foundation and the Italian Telethon Foundation.

Author contributions
Collection and assembly of data, data analysis and interpretation, conception and design, manuscript writing: R.D.C. Performed the experiments: R.D.C., S.I. and A.F. Analyzed the data: R.D.C. and D.d.B. Contributed reagents/materials/analysis tools: R.D.C., S.I. and A.B. A.O.F. provided technical input with respect to cloning. R.D.C. wrote the manuscript. Conception and financial support: A.B.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41597-019-0281-5.

Correspondence and requests for materials should be addressed to R.D.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.