

Towards Optimising the Production of and Expression from Polycistronic Vectors in Embryonic Stem Cells

Steven Y. Gao¹, Michelle M. Jack^{1,2}, Christopher O'Neill^{1*}

1 Centre for Developmental and Regenerative Medicine, Kolling Institute for Medical Research, Sydney Medical School, University of Sydney, Sydney, NSW, Australia, **2** Department of Endocrinology, Royal North Shore Hospital, Sydney, NSW, Australia

Abstract

Polycistronic vectors linked by self-processing 2A peptides have been successfully used in cellular reprogramming. The expression of these vectors has yet to be well documented in embryonic stem cells. In the present study, we generated expression cassettes containing combinatorial arrangements of 3 pancreatic transcription factors (*Pdx1*, *Nkx2.2* and *Ngn3*) together with an *eGFP* reporter, all linked by self-processing 2A peptides. The study tested the utility of constructing complex expression cassettes by ligating multiple components, each flanked by unique restriction sites. This approach allowed flexible and efficient design and construction of a combinatorial array of polycistronic constructs, which were expressed after transient transfection into embryonic stem cells. The inclusion of EGFP provided for a convenient proxy measure of expression and showed that expression was similar regardless of EGFP's position within a 2A polycistronic construct. Expression of terminal EGFP was 51% and 24% more efficient when linked by T2A compared to F2A or E2A peptides, respectively. The highest level of expression was achieved when all genes in a construct were linked exclusively by T2A peptides. This effect of T2A was independent of the type of promoter used, as a similar increase in terminal EGFP expression was observed when the polycistronic constructs were under the control of a CAG promoter compared to the CMV promoter, even though the GAG promoter was more efficient in this model than the CMV promoter. The study provides guidance on design strategies and methods for the efficient generation and expression of 2A polycistronic constructs in embryonic stem cells.

Citation: Gao SY, Jack MM, O'Neill C (2012) Towards Optimising the Production of and Expression from Polycistronic Vectors in Embryonic Stem Cells. PLoS ONE 7(11): e48668. doi:10.1371/journal.pone.0048668

Editor: Andre Van Wijnen, University of Massachusetts Medical, United States of America

Received: June 3, 2012; **Accepted:** September 28, 2012; **Published:** November 6, 2012

Copyright: © 2012 Gao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported in part by grants from the NHMRC to CO and by funds provided by the Favaloro bequest to the Department of Endocrinology. SG was supported by a Rotary Postgraduate Scholarship and a Lola Douglas Paediatric Research Scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: chris.oneill@sydney.edu.au

Introduction

Various approaches have been employed to co-express multiple proteins in cells. These include the use of internal ribosomal entry site (*IRES*) elements [1], dual promoter systems [2] or transfection of multiple vectors [3]. Each of these is associated with a number of limitations such as uneven or unreliable protein expression levels, silencing of some promoters [4,5], or increased toxicity to cells (with multiple transfections) [6]. An alternative strategy is the use of the self-processing viral 2A peptides, which are reported to allow the efficient expression of multiple genes under the control of a single promoter [7,8].

The 2A peptides are members of the *cis*-acting-hydrolase elements (*CHYSELs*) family. These elements were first identified in picornaviruses, and were subsequently found in a number of other viral systems [9]. The most studied 2A peptide sequence is from foot-and-mouth disease virus (F2A) [10]. Other commonly used 2A elements are from the equine rhinitis A virus (E2A) and *Thossea asigna* virus (T2A) [11].

All commonly used 2A peptides share a highly conserved 18 amino acid region responsible for cleavage. This occurs at the C-terminus of the sequence, between the 2A glycine and 2B proline [11,12]. The cleavage is achieved through a putative "ribosomal skip" mechanism [13]. The cleavage efficiency of the 2A peptide-

linked proteins are generally high, however, it has been demonstrated that the efficiency can be influenced by the identity of the protein at the N-terminus of the peptide [5]. The cleavage efficiency can be improved by placing a Gly-Ser-Gly linker between the N-terminal protein and the 2A peptide [11]. This "flexible" linker creates a space between the N-terminus protein and the peptide, favouring a conformation of the peptide which facilitates efficient cleavage [14].

Self-processing 2A peptides have been used to achieve co-expression of heterogeneous proteins and show cleavage efficiency in a range of differentiated cell lines and embryonic stem cells [4,7,8,11,15,16,17]. Expression of up to five genes linked by 2A sequences under the control of a single promoter has been demonstrated [18], but the efficiency of expression of individual proteins from a polycistronic construct may be lower than their expression from monocistronic constructs [8], although a definitive reason for this reduced efficiency has not been shown.

The cleavage of 2A peptides may be affected by the nature of the protein attached to its N-terminus [19,20], the order of the genes used in a specific expression cassette [21], the variant of the 2A peptide used (even if it was derived from the same peptide) [22], and the expression environment [20]. The P2A peptide is reported to have the highest cleavage efficiency in an *in vivo* environment [23], however, there remains relatively limited

Table 1. The sequences of primers used in this study, the restriction enzyme sites used in the study are underlined.

Symbols	Sequences
eGFP Full F	5'-GCCACCATGGTGAGCAAG-3'
eGFP Full R	5'-TACTTGTACAGCTCGTCCATGC-3'
SacI-GFP-F2A-SacI F	5'-GAGCTCGCGCCGCTGGATCCGCCACCATGGTGAGCAAG-3'
SacI-GFP-F2A-SacI R	5'-GGATCCAGCGCCGCGAGCTCGGGCCCTGGTTGGACTCCACGTCTCCCGCAACTTGAGAAGGTCAAATTCAAA GTCTGTTTACCAGGACCCTGCCCTTGTACAGCTCGTCCATGC-3'
BamHI/Hind-eGFP	5'-GGATCCAAGCTTATGGTGAGCAAGGGCGAG-3'
eGFP-Sall	5'-GTCGACTTACTTGTACAGCTCGTCCATGC-3'
NGN3 Full F	5'-ATGGCGCCTCATCCCTTG-3'
NGN3 Full R	5'-CAAGAAGTGTGAGAACCACAGTGCT-3'
BamHI-NGN3-F2A-HindIII F	5'-GGATCCGCCACCATGGCGCCTCATCCCTTG-3'
BamHI-NGN3-F2A-HindIII R	5'-AAGCTTGGGCGCTGGTTGGACTCCACGTCTCCCGCAACTTGAGAAGGTCAAATTCAAAGTCTGTTTACCAGGA CCGCTGCCAAGAAGTGTGAGAACCACAGTGCT-3'
BamHI-NGN3-T2A-HindIII R	5'-AAGCTTGGGCCAGGATTCTCTCGAGTCCACCGCATGTTAGCAGACTTCTCTGCCCTCACCGCTGCCAAGAA GTCTGAGAACCACAGTGCT-3'
PDX1 Full F	5'-ATGAACAGTGAGGAGCAGTACTACGCG-3'
PDX1 Full R	5'-CCGGGGTTCCTGCGGTGCG-3'
SacI-PDX1-E2A-NotI/BamHI F	5'-GAGCTCGCCACCATGAACAGTGGAGGAGCAGTACTACGCG-3'
SacI-PDX1-E2A-NotI/BamHI R	5'-GGATCCAGCGCCGCGAGGACCGGGTACTTTCAACATCGCCAGCGAGTTTCAACAAGCGTAGTTAGTACATT GGCCAGAAACCCCGGGTCTCTGCGGTGCG-3'
SacI-PDX1-T2A-NotI/BamHI R	5'-GGATCCAGCGCCGCTGGGCCAGGATTCTCTCGAGTCCACCGCATGTTAGCAGACTTCTCTGCCCTGCCA GAACCCCGGGTCTCTGCGGTGCG-3'
NKX22 Full F	5'-CCCCAGTCACAGCCTACATT-3'
NKX22 Full R	5'-AACCAACCGTGGTAAGGATCG-3'
NotI-NKX2-T2-BamHI F	5'-GCGGCCGCTGCCACCATGTCGCTGACCAACACAAAGACG-3'
NotI-NKX2-T2A-BamHI R	5'-GGATCCTGGGCCAGGATTCTCTCGAGTCCACCGCATGTTAGCAGACTTCTCTGCCCTGCCAGAACCCCAA GTCCACTGCTGGGCC-3'
ASEI-CAG F	5'-ATTAATAGTAATCAATTACGGGGTCA-3'
CAG-NheI R	5'-GCTAGCGAATTCTTCCAAAATGATGA-3'
β -Actin F	5'-GCAGGCCTAGTAACCGAGACA-3'
β -Actin R	5'-AGTTTTGGCGATGGGTGCT-3'

doi:10.1371/journal.pone.0048668.t001

systematic analysis of the effects of different 2A peptides on the protein expression levels within polycistronic expression cassette. To date, there is not a sufficient body of evidence to allow rational design of 2A-linked polycistronic constructs with optimal expression efficacy.

The most conventional method for generating polycistronic expression cassettes linked by 2A peptides involves the use of recombinant PCR [24]. This method is relatively inflexible because the use of the same 2A peptide sequence within a PCR

results in their self-annealing. Thus, designs are constrained by the need to use either different 2A peptides, or different variables between each linked proteins. Furthermore, any changes made to the structure of the cassette, such as the addition of a gene, requires the re-design of the large primers used in construction as well as the cost and time-consuming process of re-constructing the cassette. These matters limit the feasibility of constructing large numbers of constructs for screening and optimisation purposes. The present study assesses a more flexible approach for generating 2A polycistronic expression cassettes. Each gene and its assigned 2A peptide were designed to have flanking restriction enzyme sites. This allowed the generation of many different gene combinations by adding and removing gene-2A peptide units using standard restriction enzyme digestion and ligation methodologies. This approach allows the use of multiple copies of a given 2A peptide within an expression cassette.

The aim of this study was to systematically analyze the effects of design variables on the efficiency of production and expression of 2A polycistronic vectors in mouse embryonic stem cells. The number of genes, their order within a construct, the type of promoter and the type of 2A peptide used were examined. We found that the type of promoter and the 2A peptides used independently impacted on the efficiency of expression and

Table 2. Different glycine-serine-glycine (GSG) linkers (Forward) and their corresponding reverse primer sequences (Reverse) used for the generation of 2A polycistronic constructs.

Forward	Reverse
5'-GGTTCTGGC-3'	5'-GCCAGAACC-3'
5'-GGCAGCGGT-3'	5'-ACCGCTGCC-3'
5'-GGAAGAGGA-3'	5'-TCCTCTCC-3'

doi:10.1371/journal.pone.0048668.t002

A

pCMV-Script Multiple Cloning Site Region
(sequence shown 620-799)

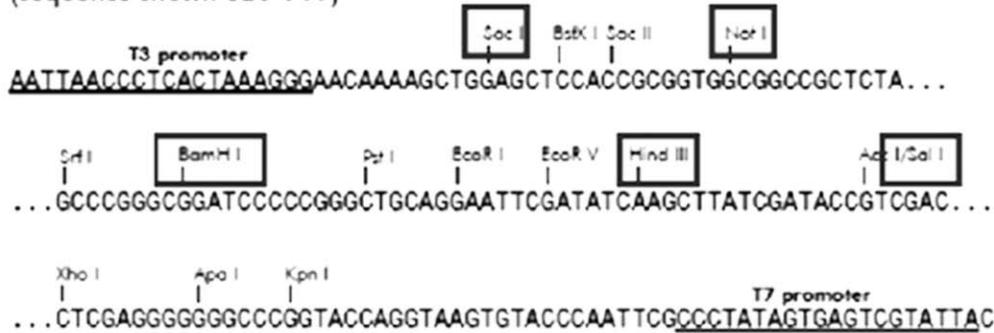
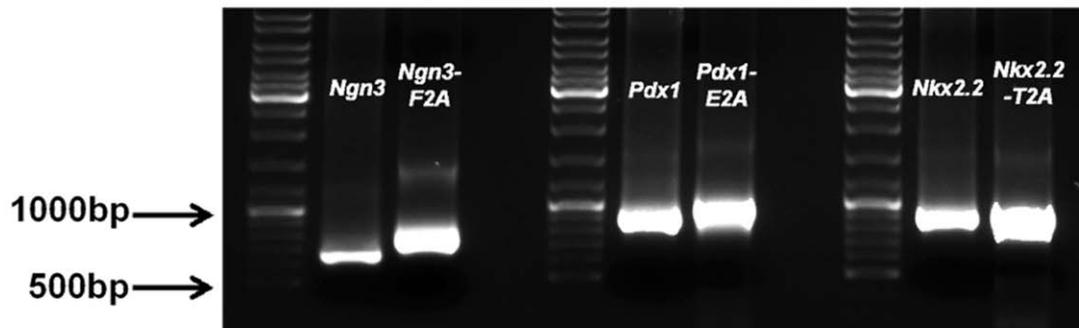
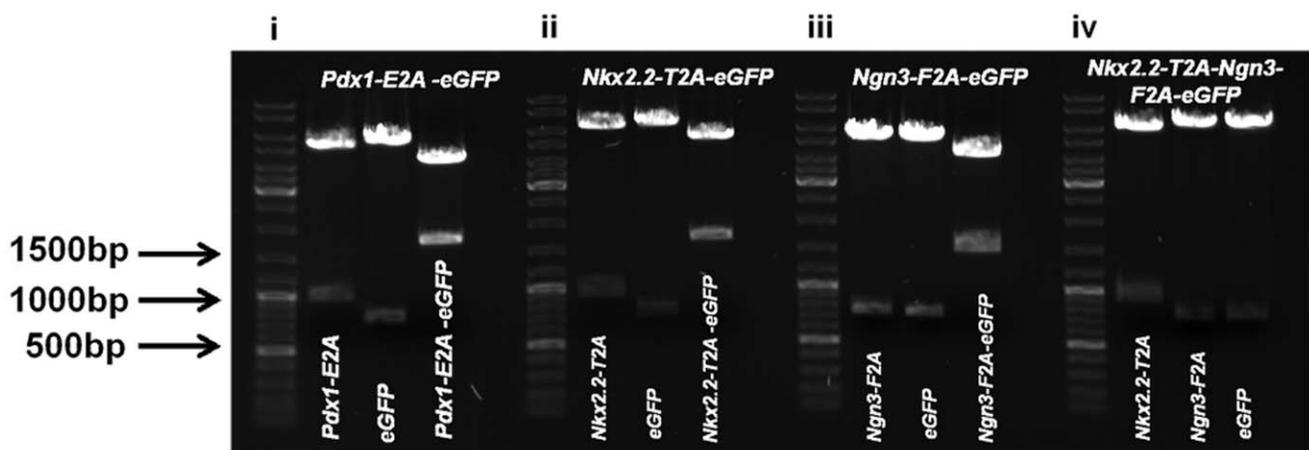
**B****C**

Figure 1. The construction of 2A polycistronic expression cassettes. (A) All 2A polycistronic expression cassettes were generated by restriction enzyme digestion and ligation utilising the restriction enzyme cutting sites: SacI, NotI, BamHI, HindIII and Sall (boxed, as described in Materials and methods) present in the multicloning site of the pCMV-Script plasmid (Stratagene). (B) The amplification of *Ngn3-F2A*, *Pdx1-E2A* and *Nkx2.2-T2A*, a size difference between the native full-length gene and the gene with a 2A-peptide tag can be observed using gel electrophoresis. Note that the native *Nkx2.2* amplified from MIN6 cell cDNA contained regions outside of the coding region of the mRNA, as the primers were optimised for maximum efficiency, and thus is larger in size compared to *Nkx2.2-T2A*. Each group of three lanes is (from left to right) molecular weight ladder, native sequence, and native sequence +2A. (C) Examples of the flexibility of the restriction enzyme digestion/ligation construction method. Each

component of a 2A polycistronic expression cassette can be removed by using the appropriate restriction enzymes: (i) For *Pdx1-E2A-eGFP*, *Pdx1-E2A* can be cut out using *SacI* and *NotI/BamHI* (left), *eGFP* can be cut out using *NotI/BamHI* and *Sall* (middle) and *Pdx1-E2A-eGFP* can be cut out using *SacI* and *Sall*; (ii) for *Nkx2.2-T2A-eGFP*, *Nkx2.2-T2A* can be cut out using *NotI* and *BamHI* (left), and *eGFP* can be cut out using *BamHI* and *Sall* (middle), and *Nkx2.2-T2A-eGFP* can be cut out using *NotI* and *Sall* (right); (iii) for *Ngn3-F2A-eGFP*, *Ngn3-F2A* can be cut out using *BamHI* and *HindIII* (left), *eGFP* can be cut out using *HindIII* and *Sall* (middle), and *Ngn3-F2A-eGFP* can be cut out using *BamHI* and *Sall* (right); (iv) for *Nkx2.2-T2A-Ngn3-F2A-eGFP*, *Nkx2.2-T2A* can be cut out using *NotI* and *BamHI* (left), *Ngn3-F2A* can be cut out using *BamHI* and *HindIII* (middle), and *eGFP* can be cut out using *HindIII* and *Sall*. The sizes of the DNA fragments were confirmed by comparison to the molecular weight ladder on the left of each data set. doi:10.1371/journal.pone.0048668.g001

demonstrated the utility and efficiency of this approach to testing polycistronic constructs.

Materials and Methods

Cell Cultures

The mouse ES cell line D3 (P22-P32) was maintained on 0.1% (v/v) gelatin (BDH Chemicals, Poole, UK) coated tissue culture dishes in ES cell media (DMEM (Thermo Fisher Scientific, Scoresby, VIC, Australia) supplemented with 10% (v/v) FBS (Invitrogen, Carlsbad, CA), 0.1 mM β -mercaptoethanol (Sigma, St. Louis, MO), 50 μ g/mL penicillin/streptomycin mix (Sigma, St. Louis, MO) and 1000 units of Leukaemia inhibitory factor (LIF) (GIBCO, Grand Island, NY)). ES cell media was changed every 1–2 days and the cells were passaged using 0.25% (w/v) trypsin-EDTA (GIBCO) upon reaching 70% confluence. Cells were cultured in 5% CO₂ in air at 37°C.

Amplification of the Full-length Sequences Using PCR

The primers used for generating constructs, and their restriction sites are shown in (Table 1). Full-length sequences of mouse *Ngn3* and *Pdx1* were amplified from *pAd-Ngn3-I-nGFP* and *pAd-Pdx1-I-nGFP* plasmids, respectively (plasmid number 19410 & 19411, kindly supplied by Dr. D. Melton through Addgene, Cambridge, MA) [25] using primers *Pdx1 Full F*, *Pdx1 Full R*, *Ngn3 Full F* and *Ngn3 Full R* (Table 1). The full-length sequence of *Nkx2.2* was amplified using the mouse insulinoma (MIN6) cell line [26] cDNA as a template with primers *Nkx22 Full F* and *Nkx22 Full R* (Table 1). The full-length sequence of *eGFP* was amplified from

the 613bp to 1329bp of the *pEGFP-C1* plasmid (Clontech, Mountain View, CA) using primers *eGFP Full F* and *eGFP Full R* (Table 1).

Vector Construction

The stop codons of *Pdx1*, *Nkx2.2* and *Ngn3* were removed from their respective reverse primers and a sequence encoding a glycine-serine-glycine (*GSG*) linker (Table 2) was added to the 3'-end of each gene. This linker is reported to facilitate efficient cleavage of the 2A peptides [27]. *PCMV-Script* (Stratagene, Cedar Creek, TX) was used as the backbone plasmid. Each gene was assigned a 2A peptide and flanked with appropriate restriction enzyme sites as found on the *pCMV-Script* plasmid. The following sequences were generated for the initial testing of the expression cassettes: *SacI-Pdx1-E2A-NotI/BamHI*, *NotI-Nkx2.2-T2A-BamHI*, *BamHI-Ngn3-F2A-HindIII*, *BamHI/HindIII-eGFP-Sall*. To move the *eGFP* from the final position to the first position, the *F2A* sequence attached to *Ngn3* was removed and a stop codon was added to the end of *Ngn3* to generate *BamHI-Ngn3-HindIII*. This replaced *BamHI-Ngn3-F2A-HindIII*. The stop codon of the *eGFP* was removed and the *F2A* sequence was added to the 3'-end of the *eGFP* to become *SacI-eGFP-F2A-SacI*. *T2A* sequences were used in place of the *F2A* and *E2A* sequences, while no changes to the restriction enzyme sites used were made.

The template sequences of the *F2A*, *E2A* and *T2A* peptides were obtained from a previously published study [11]. The primers used to generate the 2A constructs were: *BamHI/HindIII-eGFP*, *eGFP-Sall*, *BamHI-Ngn3-F2A-HindIII F*, *BamHI-Ngn3-F2A-HindIII R*, *SacI-Pdx1-E2A-NotI/BamHI F*, *SacI-Pdx1-E2A-NotI/BamHI R*, *NotI-Nkx2.2-T2A-BamHI F*, *NotI-Nkx2.2-T2A-BamHI R*, *SacI-eGFP-F2A-SacI F*, *SacI-eGFP-F2A-SacI R* (Table 1). The amplified sequences were then ligated into the multicloning site of *pCMV-Script* using T4 ligase (Promega, Madison, WI) to obtain the desired 2A polycistronic vectors. All restriction enzymes and buffers used were purchased from New England Biolabs (Ipswich, MA).

The primers used for the T2A experiments are: *BamHI-Ngn3-T2A-HindIII F*, *BamHI-Ngn3-T2A-HindIII R* and *SacI-Pdx1-E2A-NotI/BamHI F*, *SacI-Pdx1-T2A-NotI/BamHI R* (Table 1). The generated sequences were then ligated into the desired locations of the 2A polycistronic vectors to replace the original sequences.

Transient Transfection of Constructs

Mouse ES cells were trypsinised and plated at 100,000 cells/cm² on 48 well tissue culture plates (Nunc, Roskilde, Denmark). Transfection media was ES cell media without penicillin/streptomycin. Sterile vector DNA at a concentration of 1.2 μ g/100,000 cells was transfected using Lipofectamine 2000 at 3.6 μ L/1.2 μ g of DNA (Invitrogen). Samples were analysed 24 h post transfection. The backbone *pCMV-Script* plasmid (Stratagene) was used as a negative control.

Table 3. List of vectors generated for this study.

Vectors	
<i>pCMV-eGFP</i>	<i>CMV-G</i>
<i>pCMV-Ngn3-F2A-eGFP</i>	<i>CMV-N-F2A-G</i>
<i>pCMV-Pdx1-E2A-eGFP</i>	<i>CMV-P-E2A-G</i>
<i>pCMV-Nkx2.2-T2A-eGFP</i>	<i>CMV-N2-T2A-G</i>
<i>pCMV-Pdx1-E2A-Ngn3-F2A-eGFP</i>	<i>CMV-P-E2A-N-F2A-G</i>
<i>pCMV-Pdx1-E2A-Nkx2.2-T2A-eGFP</i>	<i>CMV-P-E2A-N2-T2A-G</i>
<i>pCMV-Nkx2.2-T2A-Ngn3-F2A-eGFP</i>	<i>CMV-N2-T2A-N-F2A-G</i>
<i>pCMV-Pdx1-E2A-Nkx2.2-T2A-Ngn3-F2A-eGFP</i>	<i>CMV-P-E2A-N2-T2A-N-F2A-G</i>
<i>pCMV-eGFP-F2A-Pdx1-E2A-Nkx2.2-T2A-Ngn3</i>	<i>CMV-G-F2A-P-E2A-N2-T2A-N</i>
<i>pCMV-Ngn3-T2A-eGFP</i>	<i>CMV-N-T2A-G</i>
<i>pCMV-Pdx1-T2A-eGFP</i>	<i>CMV-P-T2A-G</i>
<i>pCMV-Pdx1-T2A-Ngn3-F2A-eGFP</i>	<i>CMV-P-T2A-N-F2A-G</i>
<i>pCMV-Pdx1-E2A-Ngn3-T2A-eGFP</i>	<i>CMV-P-E2A-N-T2A-G</i>
<i>pCMV-Pdx1-T2A-Nkx2.2-T2A-Ngn3-T2A-eGFP</i>	<i>CMV-P-T2A-N2-T2A-N-T2A-G</i>
<i>pCAG-Pdx1-E2A-Nkx2.2-T2A-Ngn3-F2A-eGFP</i>	<i>CAG-P-E2A-N2-T2A-N-F2A-G</i>
<i>pCMV-Pdx1-T2A-Nkx2.2-T2A-Ngn3-T2A-eGFP</i>	<i>CMV-P-T2A-N2-T2A-N-T2A-G</i>

doi:10.1371/journal.pone.0048668.t003

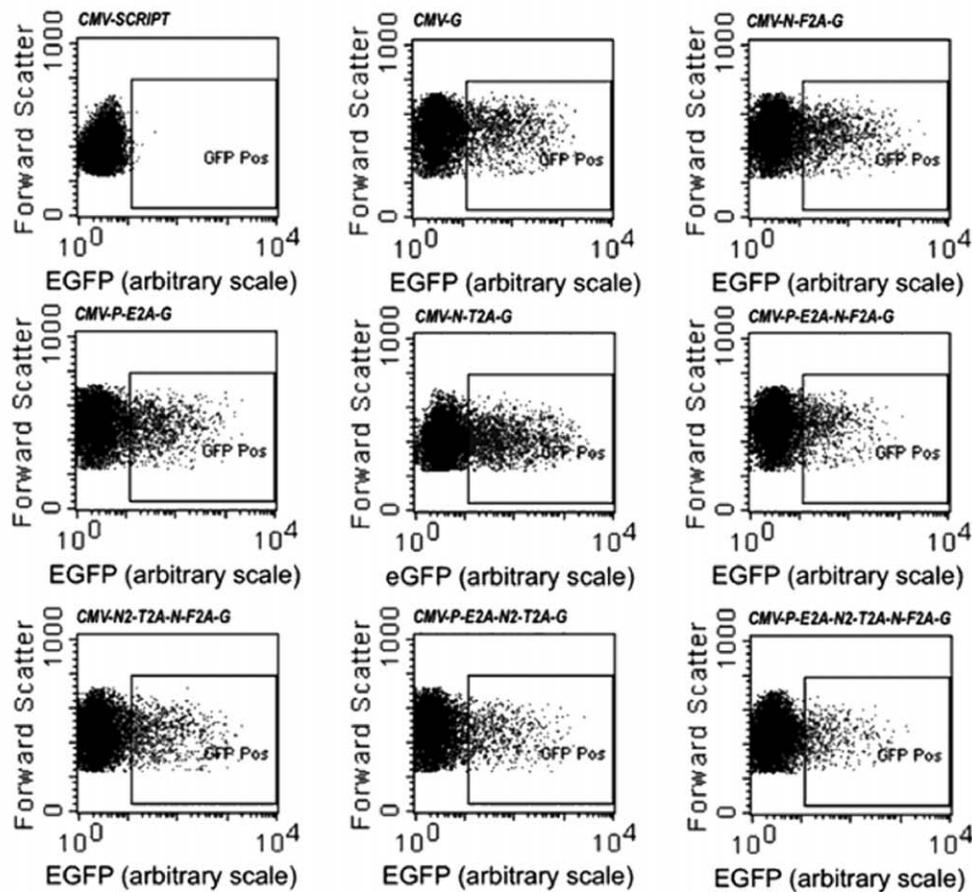
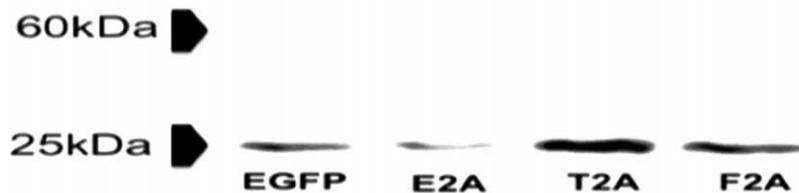
A**B**

Figure 2. Analysis of expression and cleavage of polycistronic vectors in mouse ES cells. (A) Each of the vectors shown in Table 3 were transiently transfected into mouse ES cells and EGFP expression assessed 24 h later. Representative analysis of at least 3 replicates is shown for each vector. The *pCMV-Script* backbone plasmid was used as a negative control and expression above these levels were taken as an evidence of positive EGFP expression. Each graph shows the relative level of EGFP expression in arbitrary units per cell against forward scatter. (B) The cleavage efficiency of E2A, T2A and F2A in mouse ES cells was determined by using western blot analysis of EGFP from protein samples extracted from cells transfected with two-gene vectors each linked by one of the three 2A peptides.
doi:10.1371/journal.pone.0048668.g002

Cloning of the CMV Early Enhancer/chicken β -actin (CAG) Promoter

In order to determine whether the choice of promoter influenced the efficiency of *T2A*, the *CMV* promoter was replaced by a *CAG* promoter/enhancer. The full *CAG* promoter sequence with a 5'-*NheI* restriction enzyme overhang was amplified with primers *ASEI-CAG F* and *CAG NheI R* (Table 1) from plasmid *pCAGEN* (kindly provided by Dr. C. Cepko through Addgene, plasmid number 11160) [28] using PCR. The *CMV* enhancer/

promoter from the *pCMV-Pdx1-Nkx2.2-Ngn3-eGFP* vectors was removed by digesting the vector with *AseI* and *NheI*. The *CAG-NheI* sequence was then digested and ligated into the *AseI* and *NheI* sites of the *Pdx1-Nkx2.2-Ngn3-eGFP* vectors.

Fluorescence-activated Cell Sorting

Transfected cells were analyzed 24 h post transfection using a BD DACS Calibur flow cytometer (BD Biosciences, Franklin Lakes, NJ). The cells were gated based on size and EGFP

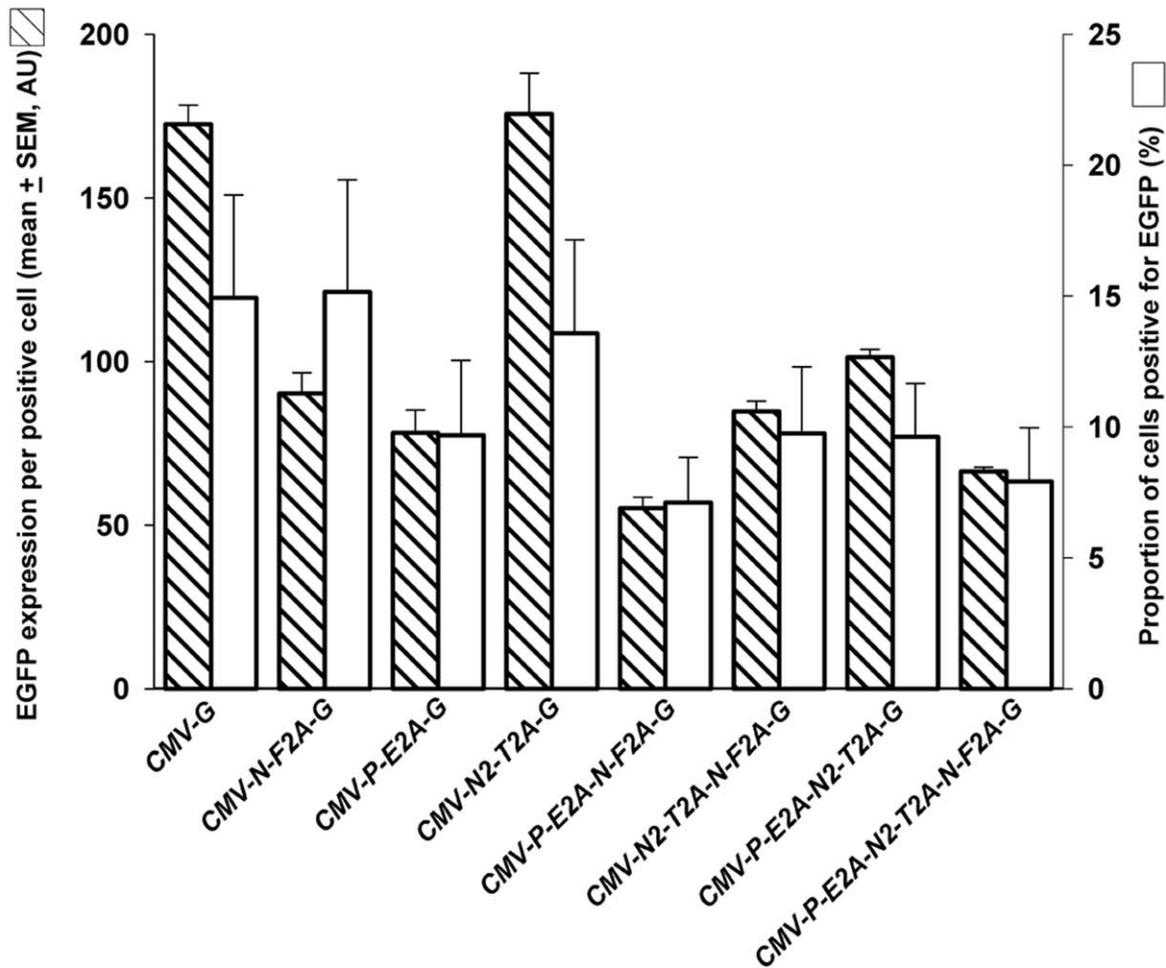
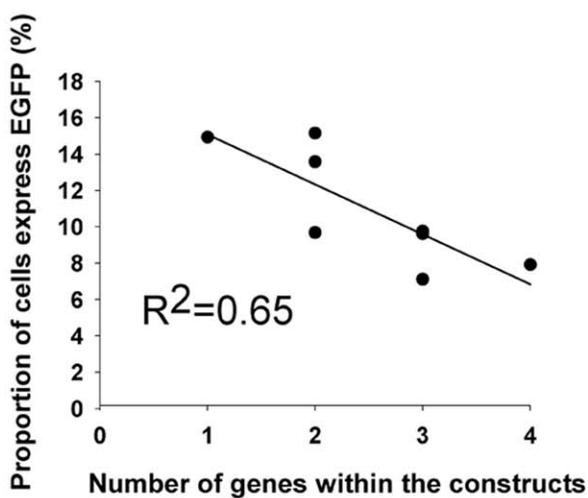
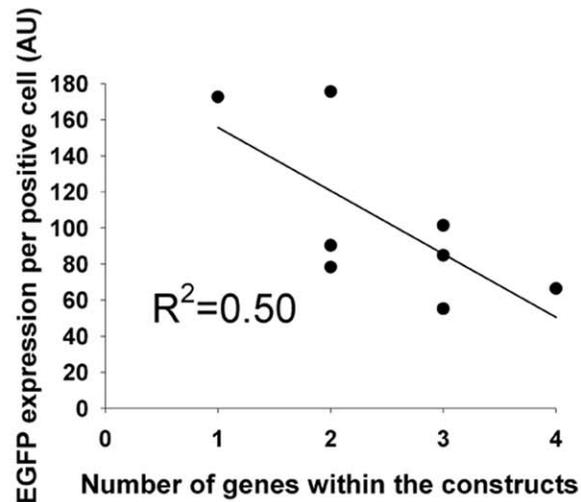
A**B****C**

Figure 3. Analysis of EGFP expressing mouse ES cells 24 h post transfection. (A) Comparative analysis was conducted to determine the relationship between the mean EGFP expression per positive cell () and the proportion of cells expressing EGFP (). Data presented as mean \pm SEM from 3 independent experiments. AU = arbitrary units. (B) The relationship between the proportion of cells expressing EGFP and number of genes contained within the constructs. Each point on the graph depicts the mean percentage of EGFP expressing cells transfected by one type of vector.

Data was obtained from 3 independent experiments. (C) The relationship between EGFP expression per cell and number of genes within constructs. An inverse linear relationship between the number of genes contained within each vector and the mean EGFP expression levels per positive cell was observed, within an increased number of genes associated with decreased EGFP expression levels. R^2 = linear regression coefficient. AU = arbitrary units.
doi:10.1371/journal.pone.0048668.g003

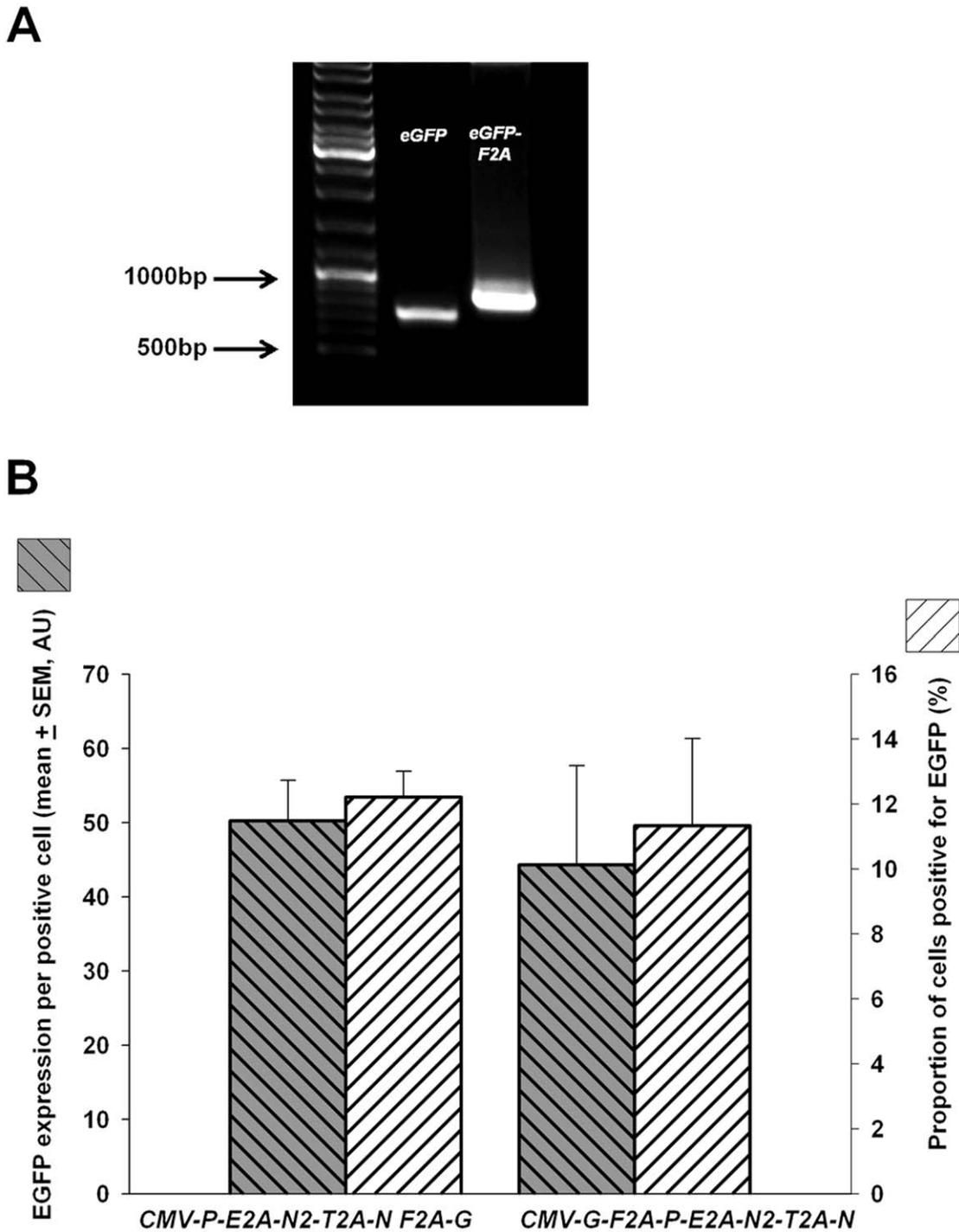


Figure 4. The expression of a gene at different locations within a 2A polycistronic expression cassette. (A) *eGFP-F2A* was amplified using PCR, and the difference in size compared to the native *eGFP* can be observed using gel electrophoresis. (B) The expression of EGFP was assessed when it was moved from the final position to the first position within a four-gene polycistronic construct. No statistical difference in EGFP expression was found (left axis), and no difference was found in the proportion of the cells expressing EGFP 24 hr after the cells were transfected with the two constructs (right axis). Data presented as mean \pm SEM from 3 independent experiments. AU = arbitrary units.
doi:10.1371/journal.pone.0048668.g004

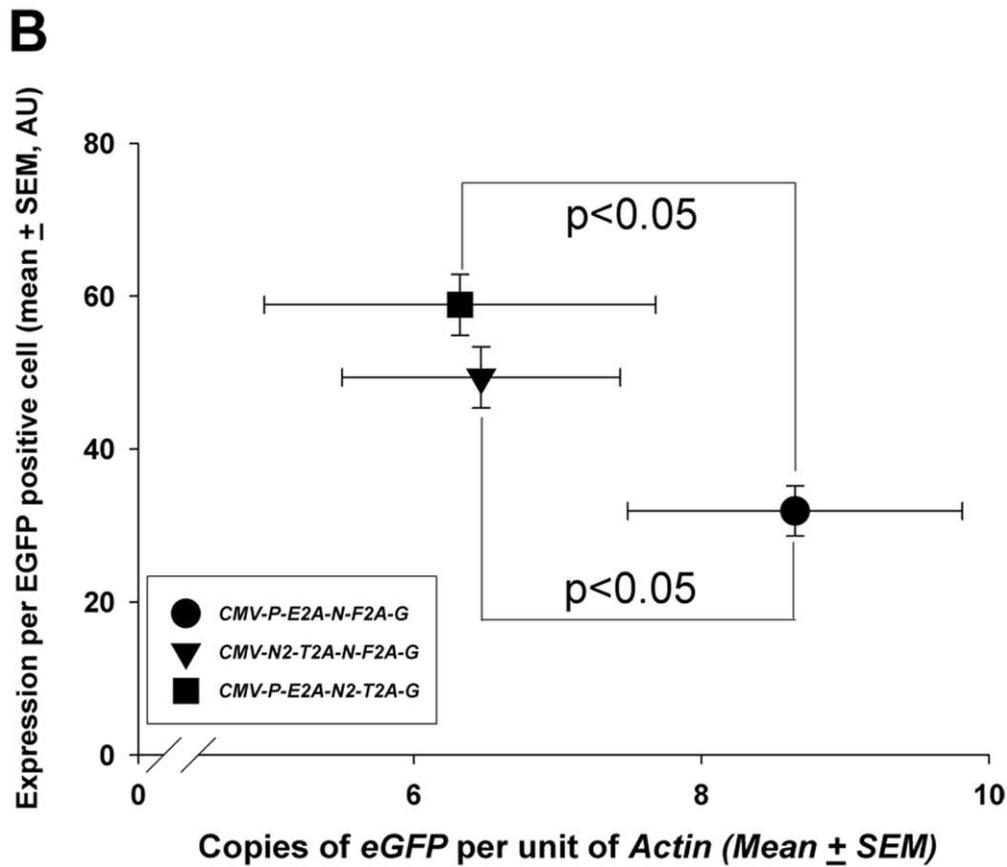
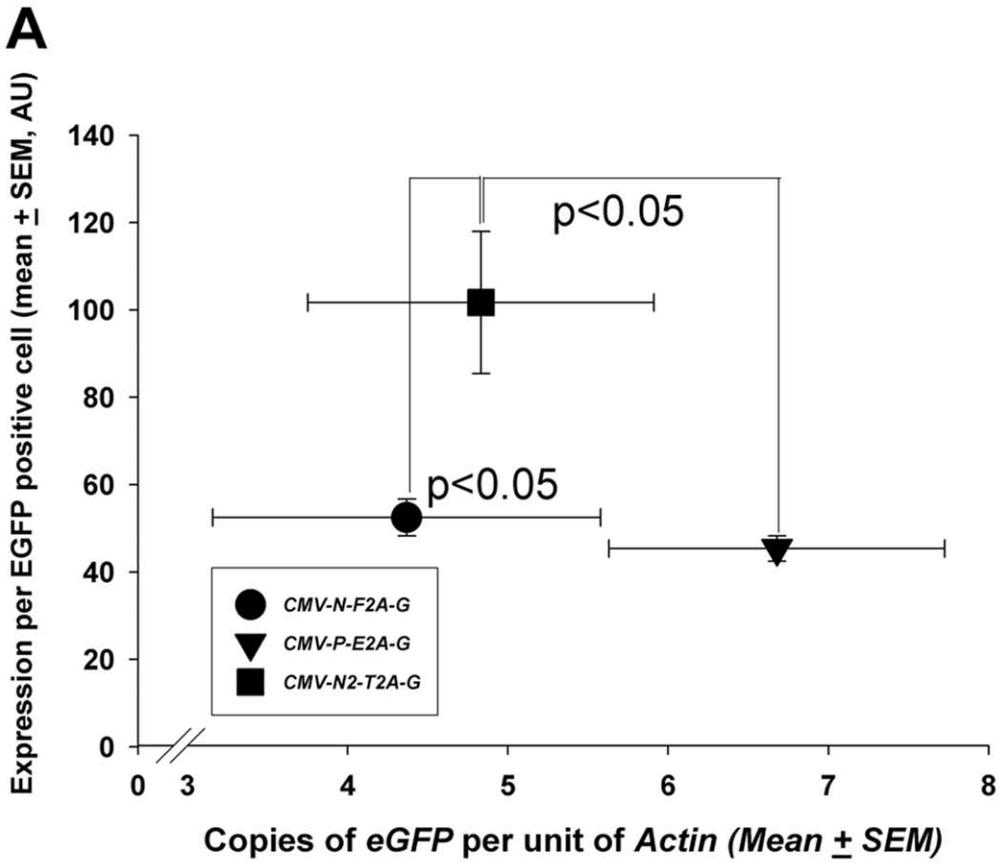


Figure 5. Analysis of vector DNA copy numbers using qPCR. Difference in EGFP expression in mouse ES cells transfected with two-gene and three-gene vectors compared to vector DNA copy numbers in the transfected cells. Vector DNA copy number was determined using qPCR and normalised to *Actin*. Even though significant differences in the mean EGFP expression per positive cell was observed between T2A containing constructs and non-T2A containing constructs for both two-gene (A) and three-gene (B) polycistronic vectors, the increased EGFP expression was not correlated with high vector DNA copy number. Data presented as mean \pm SEM from 3 independent experiments. AU = arbitrary units. doi:10.1371/journal.pone.0048668.g005

expression was detected at an excitation frequency of 488nm. The final data was analysed using the BD Cell Quest Pro software (BD Biosciences).

Determining Vector DNA Copy Numbers

In order to analyse the vector copy number in the transfected mouse ES cells, DNA from the cells was extracted using the Wizard Genomic isolation System (Promega) following the manufacturer's instructions. An equal amount of DNA from each sample was added to each quantitative PCR (qPCR) reaction. The qPCR primer sequences for *eGFP* were obtained from a previous study [29]. *Beta-Actin* was used as a control (Table 1). All samples were then amplified using a Rotorgene RG3000 (Corbett Life Science, Concord, NSW, Australia). The Ct value of the *eGFP* was normalised against its corresponding β -*Actin* Ct value. The results are represented as the "units of *eGFP* DNA per unit of *Actin*". This provides a measure of relative transfection efficiency.

Western Blot

To investigate the cleavage efficiency of each of the 2A peptides used in the study, protein samples were extracted from an equal number of cells transfected with *Pdx1-E2A-eGFP*, *Nkx2.2-T2A-eGFP* and *Ngn3-F2A-eGFP* constructs. Constructs were separated by 12% (w/v) SDS-PAGE and transferred onto nitrocellulose membranes. The membranes were blocked using 5% skim milk in TBS-Tween (TBST, Tris-Buffered Saline with 0.05% v/v Tween 20) before incubation with GFP primary antibody (1:5000) (Abcam, Cambridge, MA, USA) overnight at 4°C and horseradish peroxidase conjugated secondary antibody (1:5000) (Abcam) for 1 h at room temperature. The samples were developed using Super-Signal West Femto Chemiluminescent Substrate (Thermo Fisher Scientific, Scoresby, VIC, Australia) and visualised using chemiluminescence on a Fujifilm LAS-4000 (GE Healthcare, Waukesha, WI).

Statistical Analysis

Statistical analysis was conducted using SPSS statistics software (IBM, Armonk, NY). Students T-tests were conducted to compare the difference in EGFP expression level between each transfected sample and vector DNA copy numbers between T2A and non-T2A containing vectors. Univariate analysis was conducted to determine the effect of different construct designs containing T2A on the expression levels of EGFP when *eGFP* was the terminal gene within the construct. The difference between results was considered significant if $p < 0.05$.

Results

The present study constructed a range of polycistronic expression cassettes by linking a number of *Gene-2A* sequences using unique restriction enzyme sites. This contrasts with the conventional recombinant PCR method for generating 2A polycistronic expression cassettes [24]. The method made use of the multicloning site of the *pCMV-Script* plasmid (Figure 1A). A 2A peptide was assigned to each gene (F2A was assigned to *Ngn3*, E2A to *Pdx1* and T2A to *Nkx2.2*) and amplified using conventional PCR. The size difference between the native gene and the gene

with a 2A tag can be visualised using gel electrophoresis (Figure 1B). This approach allowed each component within a 2A polycistronic expression cassette to be edited by conventional restriction digestion/ligation (Figure 1C).

All vectors generated in the study can be found in Table 3. Gene expression from the generated vectors was first screened using FACS detection of the expression of EGFP protein, which was incorporated as the last gene within each of the constructs. The cells transfected with the *pCMV-Script* backbone plasmid were used as negative controls. An equal amount of each vector was used to transfect mouse ES cells in all samples and EGFP expression was assessed 24 h post transfection by FACS analysis. All vectors generated using the restriction enzyme digestion/ligation method resulted in the expression of EGFP in a proportion of the transfected population (Figure 2A). Western blot analysis of EGFP was performed on cells transfected with one of three different two-gene constructs (Figure 2B). It was expected that if cleavage was incomplete a chimeric protein of larger molecular mass would result. In each case, all of the EGFP detected was of the expected size, showing efficient cleavage by all three 2A sequences joined by the restriction enzyme method. These results show that each of the 2A sequences tested had equally high cleavage efficiencies and that the measurement of EGFP was a reliable proxy for assessing vector expression and 2A cleavage efficiency in this model.

No relationship was found between the mean EGFP expression per positive cell and the proportion of cells positive for EGFP (Figure 3A). There was an inverse linear relationship ($R^2 = 0.65$) between the number of genes contained within each vector and the proportion of cells that expressed detectable EGFP (Figure 3B), and a similar relationship with the level of expression in those cells that were positive for EGFP signal ($R^2 = 0.50$) (Figure 3C).

In order to investigate whether the position of a gene located in a 2A polycistronic expression cassette has an impact on its protein expression. An *eGFP-F2A* sequence was amplified using PCR (Figure 4A) and ligated into the first position of a four-gene expression cassette. Moving the *eGFP* from the last to first position within the construct did not change the level of EGFP expressed (Figure 4B), and confirms the utility of the flanking restriction enzyme approach for modifying the constructs.

As well as the inverse relationship between expression and the number of genes within a construct, there was also marked variability in EGFP expression between different two and three-gene constructs. To assess whether this was accounted for by differences in transfection efficiency, the *eGFP* copy number (relative to β -*actin*) was measured by qPCR and compared to the level of EGFP expression. This showed that the level of EGFP expression did not show an obvious relationship to the gene copy number. It was observed that a construct containing the T2A peptide had the highest level of expression relative to copy number (Figure 5A, 5B). The constructs tested, however, contained a range of genes so it was not possible to determine whether the observed variability was accounted for by the efficiency of transcription/translation of the different genes or due to the presence of different 2A sequences. We therefore used the flanking restriction enzyme approach to cassette re-design to change the constructs to have all

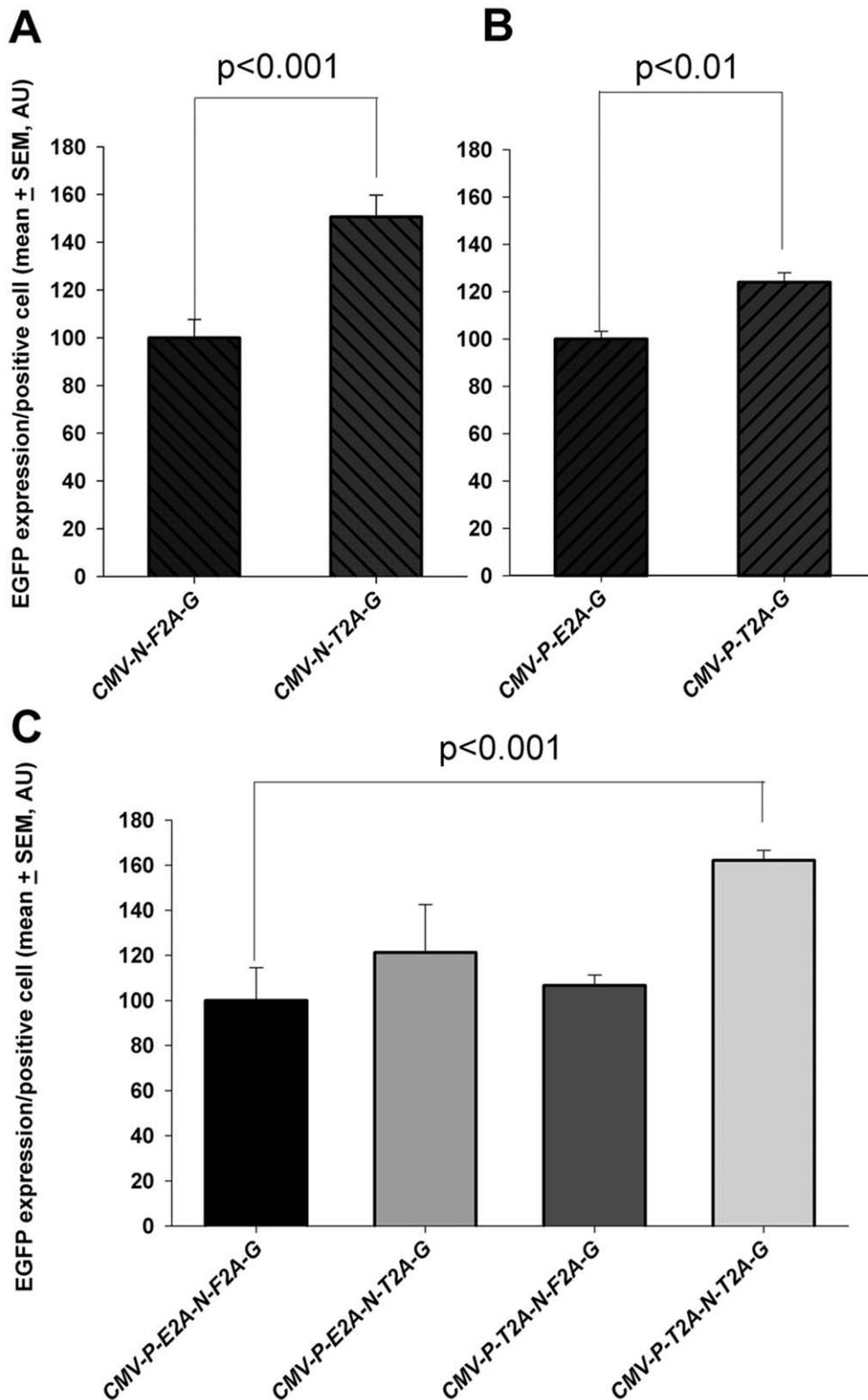


Figure 6. Analysis of the difference between T2A, E2A and F2A on EGFP expression by cells. (A) Replacing F2A peptide with T2A peptide in a two-gene construct caused a significant increase in mean EGFP expression per positive cell. (B) Replacing E2A with T2A in a two-gene construct caused a significant increase in the mean EGFP expression per positive cell. (C) In a three-gene construct, the expression of EGFP was not altered after replacing either a single single F2A or E2A with T2A. A significant increase in EGFP expression was detected when both F2A and E2A were replaced with T2A. Data presented as mean \pm SEM from 3 independent experiments. AU = arbitrary units.
doi:10.1371/journal.pone.0048668.g006

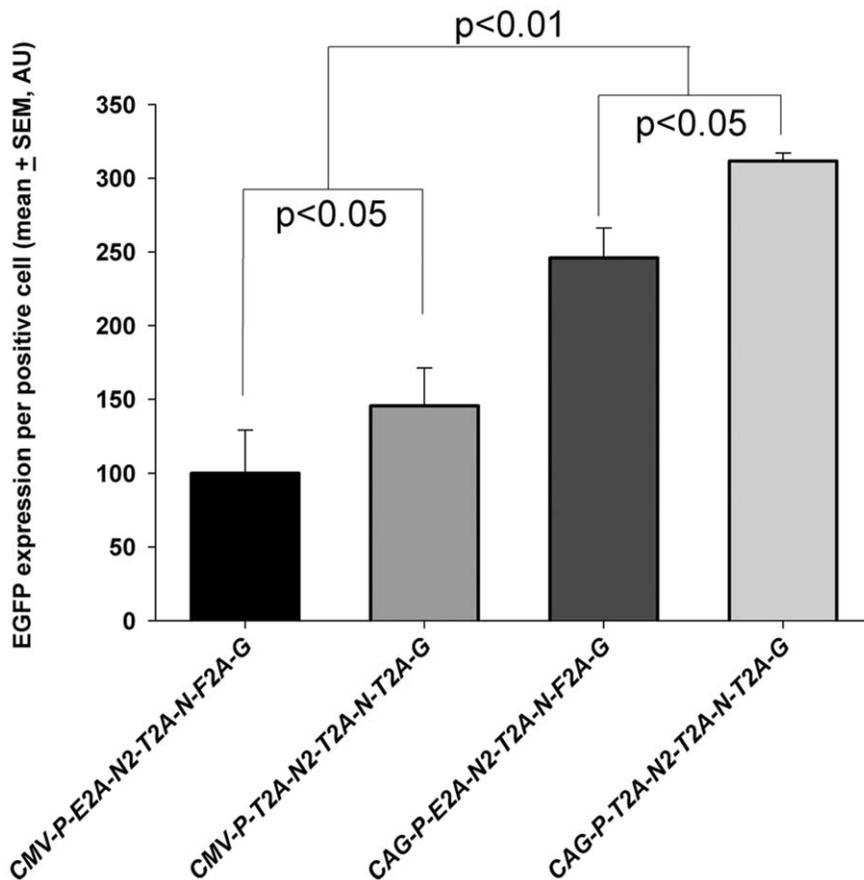


Figure 7. The effect of a different promoter on the beneficial effect of T2A on EGFP expression in four-gene constructs. A significant increase ($P < 0.05$) in mean EGFP expression per positive cells can be observed when comparing a four-gene construct linked with only T2A peptides and a four-gene construct linked with a mixture of E2A, T2A and F2A peptides, irrespective of promoter. The CAG promoter in place of the CMV promoter increased expression of EGFP, irrespective of the 2A sequences ($p < 0.01$). There was a significant interaction effect between 2A sequence and promoter ($p < 0.05$), showing that CAG together with all T2A gave the highest level of expression. Data presented as mean + SEM from 3 independent experiments. AU = arbitrary units. doi:10.1371/journal.pone.0048668.g007

genes linked by T2A. Thus, *Pdx1-T2A* and *Ngn3-T2A* were used to directly replace *Pdx1-E2A* and *Ngn3-F2A* in the original cassettes. The expression of these new vectors in ES cells showed that in otherwise identical two-gene cassettes the use of a T2A-linker increased expression compared to F2A or E2A (**Figure 6A**). Using this same approach to reconstruction of the cassette in three-gene constructs, replacing only one 2A peptide (E2A or F2A) with T2A did not result in a significant change in EGFP expression, but exclusive use of T2A in this three-gene vector resulted in an approximate 60% increase in EGFP expression (**Figure 6B**).

CMV is not always an efficient promoter in ES cells and it was therefore of interest to determine whether the increased efficiency of T2A was also manifested when the construct was under a stronger promoter. We re-designed four-gene cassettes to be expressed under either CMV or the CAG enhancer/promoter. Constructs were designed to have either different 2A sequences between each gene or linked exclusively by T2A sequences. EGFP expression was significantly higher under the CAG promoter than CMV vectors, irrespective of the 2A sequences used (including those linked exclusively with T2A) (**Figure 7**). The result shows that the beneficial effect of T2A sequence on expression was independent of the efficiency of the promoter used.

Discussion

The present study constructed various polycistronic expression cassettes by linking a number of *Gene-2A* sequences, each flanked by unique restriction enzyme sites. This method allowed the generation of a combinatorial array of constructs containing up to four genes. This provides for a more time and cost-effective approach to the construction of multiple constructs and provides for a rapid method of their re-engineering as required [24]. Additionally, this method allowed the use of multiple copies of the same 2A peptide within an expression cassette without the need for use of sequence variants [30]. This eliminated any potential changes in expression or cleavage efficiency caused by using sequence variants of the same peptide [22], and therefore reduces the level of validation of the constructs required.

This study confirms several reports that the efficiency of transcription decreases with an increase in the number of genes contained in a 2A polycistronic vector [7,8]. We found that a significant component of this loss of efficiency could be ameliorated by changing the identity of the 2A sequence used. Expression cassettes containing the T2A sequence were transcribed more efficiently than those containing either F2A or E2A in mouse ES cells, and four-gene cassettes linked exclusively with

T2A peptide had higher expression levels than the same genes linked by a mix of 2A sequences.

It is reported that the *CMV* promoter is not consistently effective in ES cells [31,32,33,34], so we were interested to examine whether the apparent beneficial effect of the T2A peptide was also manifested under the control of a stronger promoter. The *CAG* promoter/enhancer is reported to act efficiently in ES cells [35] and we found a range of construct designs were more efficiently expressed under the control of the *CAG* promoter compared with *CMV*. Despite this increased efficiency there was still a further beneficial effect of the T2A peptide.

It is been well documented that similar levels of gene expression can be achieved for all genes within a 2A polycistronic vector [4,8,11,17,18,36], therefore, it is reasonable to assume that the expression of EGFP is representative of the expression levels of all the proteins within a single expression cassette. The advantage of using a reporter like EGFP for this study is that it provides a relatively cost effective method for detecting and quantifying protein expression levels and isolation of transfected cells by FACS.

There are conflicting reports on the relative cleavage efficiency of the 2A sequences. Some reports show a greater efficiency of T2A compared to F2A and E2A [19,23] and another that F2A was better than T2A and E2A [11]. In those studies GSG linkers on the N-terminus of the 2A peptides were not included in the design. This linker can facilitate almost complete cleavage of the peptides [15,37], and in this study we found that each of the three 2A peptides used had an equally high cleavage efficiency in mouse ES cells. The observation that EGFP production was the same whether it was first or last in the sequence of an expression cassette provides confirmation that in this model cleavage efficiency was not a limiting factor to efficiency of expression. The results show that the identity of the 2A sequences was a primary determinant of transcription and/or translation of the polycistronic expression

cassettes. The mechanisms involved require further investigation and their understanding may provide further insights into strategies of optimisation of gene expression. The P2A sequence was recently shown [23] to have high cleavage efficiency in a polycistronic expression system *in vivo*, further illustrating the need for continued exploration of these options.

This study demonstrates the improved utility and efficiency of a flanking restriction enzyme based construction approach for 2A polycistronic expression cassettes for the generation of a combinatorial range of vectors. It enabled rapid analysis of the effects of a range of factors on the expression efficiency of various designs. This allowed the identification that the 2A sequence was a major determinant of the efficiency of expression of these vectors, and provides for a convenient method of use of the same 2A sequence to link multiple genes. The study points to the need for careful analysis of this aspect of design for each use. Further analysis is required to assess whether this will apply to all target cells and all expressed genes. The results provide an efficient platform for the rapid development and testing of polycistronic constructs for use in the budding field of cellular reprogramming.

Acknowledgments

We thank Dr H.D. Morgan for valuable assistance in the design and construction of the 2A containing vectors, Dr M Morris for assistance with Western Blot analysis and L. Peters for assistance with flow cytometer analysis.

Author Contributions

Conceived and designed the experiments: CO SG. Performed the experiments: SG. Analyzed the data: CO SG. Contributed reagents/materials/analysis tools: MJ. Wrote the paper: CO SG. Initiated research program: MJ.

References

- Jang SK, Krausslich HG, Nicklin MJ, Duke GM, Palmberg AC, et al. (1988) A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J Virol* 62: 2636–2643.
- Kim KJ, Kim HE, Lee KH, Han W, Yi MJ, et al. (2004) Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Sci* 13: 1698–1703.
- Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126: 663–676.
- Tang W, Ehrlich I, Wolff SB, Michalski AM, Wolf S, et al. (2009) Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. *J Neurosci* 29: 8621–8629.
- Radcliffe PA, Mitrophanous KA (2004) Multiple gene products from a single vector: 'self-cleaving' 2A peptides. *Gene Therapy* 11: 1673–1674.
- Arbab AS, Yocum GT, Wilson LB, Parwana A, Jordan EK, et al. (2004) Comparison of transfection agents in forming complexes with ferumoxides, cell labeling efficiency, and cellular viability. *Mol Imaging* 3: 24–32.
- Hasegawa K, Cowan AB, Nakatsuji N, Suemori H (2007) Efficient multicistronic expression of a transgene in human embryonic stem cells. *Stem Cells* 25: 1707–1712.
- Chinnasamy D, Milsom MD, Shaffer J, Neuenfeldt J, Shaaban AF, et al. (2006) Multicistronic lentiviral vectors containing the FMDV 2A cleavage factor demonstrate robust expression of encoded genes at limiting MOI. *Virology* 3: 14.
- Doronina VA, de Felipe P, Wu C, Sharma P, Sachs MS, et al. (2008) Dissection of a co-translational nascent chain separation event. *Biochem Soc Trans* 36: 712–716.
- Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, et al. (2012) 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res* 40: 3143–3151.
- Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, et al. (2004) Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat Biotechnol* 22: 589–594.
- Ryan MD, King AM, Thomas GP (1991) Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* 72 (Pt 11): 2727–2732.
- Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, et al. (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J Gen Virol* 82: 1013–1025.
- Holt J, Szymczak-Workman AL, Vignali KM, Burton AR, Workman CJ, et al. (2006) Generation of T-cell receptor retrogenic mice. *Nat Protoc* 1: 406–417.
- Kaji K, Norby K, Paca A, Mileikovsky M, Mohseni P, et al. (2009) Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* 458: 771–775.
- Ryan MD, Drew J (1994) Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *Embo J* 13: 928–933.
- Woljten K, Michael IP, Mohseni P, Desai R, Mileikovsky M, et al. (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458: 766–770.
- Yusa K, Rad R, Takeda J, Bradley A (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat Methods* 6: 363–369.
- Donnelly ML, Hughes LE, Luke G, Mendoza H, ten Dam E, et al. (2001) The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J Gen Virol* 82: 1027–1041.
- de Felipe P, Luke GA, Brown JD, Ryan MD (2010) Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnol J* 5: 213–223.
- Anderson RP, Voziyanova E, Voziyanov Y (2012) Flp and Cre expressed from Flp-2A-Cre and Flp-IRES-Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. *Nucleic Acids Res* 40: e62.
- Ibrahimi A, Vande Velde G, Reumers V, Toelen J, Thiry I, et al. (2009) Highly efficient multicistronic lentiviral vectors with peptide 2A sequences. *Hum Gene Ther* 20: 845–860.
- Kim JH, Lee SR, Li LH, Park HJ, Park JH, et al. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* 6: e18556.
- Szymczak-Workman AL, Vignali KM, Vignali DA (2012) Generation of 2A-linked multicistronic cassettes by recombinant PCR. *Cold Spring Harb Protoc* 2012: 251–254.

25. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA (2008) In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 455: 627–632.
26. Miyazaki J, Araki K, Yamato E, Ikegami H, Asano T, et al. (1990) Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* 127: 126–132.
27. Torres V, Barra L, Garces F, Ordenes K, Leal-Ortiz S, et al. (2010) A bicistronic lentiviral vector based on the 1D/2A sequence of foot-and-mouth disease virus expresses proteins stoichiometrically. *J Biotechnol* 146: 138–142.
28. Matsuda T, Cepko CL (2004) Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci U S A* 101: 16–22.
29. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, et al. (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 312: 436–439.
30. Szymczak-Workman AL, Vignali KM, Vignali DA (2012) Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harb Protoc* 2012: 199–204.
31. Bagchi B, Kumar M, Mani S (2006) CMV promoter activity during ES cell differentiation: potential insight into embryonic stem cell differentiation. *Cell Biol Int* 30: 505–513.
32. Chung S, Andersson T, Sonntag KC, Bjorklund L, Isacson O, et al. (2002) Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines. *Stem Cells* 20: 139–145.
33. Liew CG, Draper JS, Walsh J, Moore H, Andrews PW (2007) Transient and stable transgene expression in human embryonic stem cells. *Stem Cells* 25: 1521–1528.
34. Ward CM, Stern PL (2002) The human cytomegalovirus immediate-early promoter is transcriptionally active in undifferentiated mouse embryonic stem cells. *Stem Cells* 20: 472–475.
35. Alexopoulou AN, Couchman JR, Whiteford JR (2008) The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors. *BMC Cell Biol* 9: doi:10.1186/1471-2121-1189-1182.
36. Shao L, Feng W, Sun Y, Bai H, Liu J, et al. (2009) Generation of iPSC cells using defined factors linked via the self-cleaving 2A sequences in a single open reading frame. *Cell Res* 19: 296–306.
37. Holst J, Vignali KM, Burton AR, Vignali DA (2006) Rapid analysis of T-cell selection in vivo using T cell-receptor retrogenic mice. *Nat Methods* 3: 191–197.