Biol. Proced. Online 2007; 9(1): 84-90
doi:10.1251/bpo136
December 24, 2007

# Development of human gene reporter cell lines using rAAV mediated homologous recombination

Sandra L. Fernandez[1,2], David W. Russell[3] and Peter J. Hurlin[1,2,*]

[1]Shriners Hospitals for Children. [2]Department of Cell Developmental Biology, Oregon Health & Science Portland, OR 97201, USA.
[3]Department of Pediatrics, Division of Genetics and Developmental Medicine, University of Washington Seattle, WA 98195-7720, USA.

[*]Corresponding Author: Peter J. Hurlin, Department of Cell Developmental Biology, Oregon Health & Science Portland, OR 97201, USA. Phone: (503) 221 3438; Fax: (503) 221 3451; E-mail: pjh@shcc.org

Submitted: September 10, 2007; Revised: October 4, 2007; Accepted: November 19, 2007

Indexing terms: Dependovirus, Proto-Oncogene Proteins c-myc

Abbreviations: EGFP, enhanced green fluorescent protein; HFF, human foreskin fibroblast, rAAV, recombinant adeno-associated virus

## ABSTRACT

Understanding mechanisms of gene regulation has broad therapeutic implications for human disease. Here we describe a novel method for generating human cell lines that serve as reporters of transcriptional activity. This method exploits the ability of recombinant adeno-associated virus to mediate the insertion of exogenous DNA sequences into specific genomic loci through homologous recombination. To overcome the severe size limitation of the rAAV for carrying exogenous DNA, an enhanced green fluorescent protein (EGFP)-Luciferase fusion gene was used as both a selectable marker and gene expression reporter. EGFP was used for selection of correctly targeted alleles by taking advantage of known regulatory conditions that activate transcription of specific genes. Using this method, we describe the generation of primary human fibroblasts that express EGFP-Luciferase under the control of the c-Myc oncogene.

## INTRODUCTION

Gene regulation is a complex process that involves the binding and recruitment of transcription factors to a wide variety of regulatory elements, including distal enhancers and proximal promoter sequences. Studies of gene regulation often employ artificial plasmid reporter systems that fuse regulatory regions to cDNAs that encode reporter genes such as luciferase or chloramphenicol acetyl transferase. Although it is well established that gene regulatory elements can function at long distance (>50kb) and may exist 5′, 3′ or be intergenic, most plasmid-based reporter systems require that only a relatively small DNA fragment(s) be used. Furthermore, the analysis of regulatory regions or presumptive regions is compromised in plasmid reporter systems, whether *in vivo* or *in vitro*, by the artificial regulatory environment cultivated in this approach. While these limitations can be overcome by analysis of reporter genes that have been introduced into endogenous loci by plasmid-based homologous recombination (1), the efficient generation of reporter cells by this method is hampered by a low efficiency of homologous recombination. For these reasons, we sought to develop a method that facilitates the efficient generation of reporter cell lines, which accurately reflects endogenous gene regulation.

Recent studies have shown high efficiency gene targeting through the use of recombinant Adeno-Associated Virus (rAAV)-mediated homologous recombination in human somatic cells (2, 3, 4). The efficiency of rAAV targeting surpasses traditional plasmid-based methods by several fold (4, 5), with up to 1% of unselected cells undergoing targeting (2, 6). Wild-type AAV is a single stranded DNA parvovirus with a 4.7 kb genome flanked by two inverted terminal repeats (ITR) (7). A number of rAAV-mediated targeting strategies have been employed (2, 4, 8, 9, 10, 11, 12, 13), and the general features of the rAAV targeting virus include viral ITRs with introduced homologous arms that mediate recombination to specific genomic sites flanking a drug selection cassette. Importantly, the small packaging size of AAV (4.7 kb) severely restricts the amount of exogenous DNA that can be included in rAAV gene targeting vectors (7). For example when a typical drug selection cassette such as the ~1.7 kb PGK-Neo-pA is used in the targeting vector, there remains space for inclusion of only 3 kb of targeting sequence, which is usually arranged as homology arms of at least 1kb. The high efficiency targeting of rAAV occurs despite the relatively small amounts of homologous sequence that can be included. However, the inclusion of homologous arms together with a drug selection cassette restricts or precludes the addition of other sequences, such as reporter genes, larger than ~750 bps into the rAAV targeting vector.

To overcome the need to use a standard selectable marker gene, we have developed a new selection method for rAAV targeting. Our method relies on manipulation of known gene regulatory conditions to directly select for the introduction of a transcriptional reporter gene into a specific endogenous genetic locus. Here we describe the use of this method to develop primary human fibroblasts in which a promoter-less EGFP-Luciferase (EGFP-Luc) fusion gene has been introduced into the *c-Myc* locus in frame with, and immediately downstream of the ATG translational start site.

These cells will be useful for studying signal transduction pathways and specific transcription factors that regulate *c-Myc* transcription. Further, these and other reporter cell lines developed using techniques described here may provide a robust platform for conducting high-throughput screens for cDNAs, siRNAs, and small molecules that modulate endogenous gene expression.

## MATERIALS AND METHODS

### HFF culture and cell cycle entry assays

Primary human foreskin fibroblasts (HFFs) were generously provided at passage one from Carla Grandori (Fred Hutchinson Cancer Research Center). HFFs were routinely cultured in high glucose Dulbecco's modified Eagle medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (Hyclone), 100 U/mL penicillin, and 100 µg/mL streptomycin (Invitrogen) in incubators at 37°C and 5% $CO_2$. For cell cycle entry assays, HFFs were first driven into quiescence by maintaining cells at confluence for 3 days in medium containing 10% FBS followed by 3 days of culture in DMEM containing 0.1% FBS. Quiescent cells were stimulated to enter the cell cycle by the addition of medium containing 20% FBS.

### Construction and packaging rAAV gene targeting vectors

To facilitate production of rAAV targeting vectors that insert an EGFP-Luciferase reporter gene into target loci of interest we created the pEGFP-Luciferase cloning vector (pELCV). A promoter-less EGFP-Luciferase-SV40 pA fusion gene was obtained from pEGFPLuc (Stratagene). To preserve a unique XbaI site in the final pELCV multiple cloning region, the pEGFPLuc XbaI site in the 3' portion of the EGFP-Luciferase fusion gene was destroyed by XbaI digestion, Klenow fill-in, and blunt end ligation. The 2690 bp EGFP-Luciferase-pA restriction fragment was then removed from pEGFPLuc at NheI and MluI sites and made blunt by a Klenow fill-in reaction. This fragment was blunt end ligated into the SmaI site of pBluescript II SK+ (Stratagene). Addition of rAAV ITRs requires DNA fragments with 5' and 3' NotI ends. Therefore, a second NotI site was introduced into pELCV between the KpnI and XhoI recognition sequences using a KpnI NotI XhoI oligo linker to complement the existing NotI in the pBluescript II SK+ multiple cloning site.

Construction of the *c-Myc* EGFP-Luc knockin targeting vector was accomplished by PCR amplification of the *c-Myc* left and right homologous arms from the human genomic BAC RP11 237 F24 (Invitrogen) using primers containing unique exogenous restriction enzyme sites (underlined). The 846 bp left homologous arm (LHA) was amplified using an EcoRV forward primer (GGTCAGATATCGGAGGAACTGCGAGGAGC) and a PstI reverse primer (CTCGGTCCTGCAGCATCGTCGCGGGAGGCTGCTG) that ends with the c-Myc ATG. The 807 bp right homologous arm (RHA) was amplified using a BamHI forward primer (GGTCAGGATCCCCCCTCAACGTTAGCTTCACC) that starts with the first bp after the ATG and a XbaI reverse primer (CTCGGTCTAGAGAAGGGATGGGAGGAAACGC).

PCR product was restriction digested and sequentially ligated into pELCV. In-frame fusion of the *c-Myc* ATG with the start codon of the EGFP-Luc fusion gene was confirmed by sequencing. The AAV ITRs were introduced by ligation of the NotI targeting fragment into the pAAV-hrGFP vector backbone (Stratagene). ITR flanked *c-Myc* targeting vector integrity was confirmed by AhdI restriction mapping.

rAAV vector stocks were prepared by cotransfection of the *c-Myc* targeting vector, pAAV-RC (Stratagene), and pHelper (Stratagene) into 60% confluent AAV-293 cells (Stratagene) with lipofectamine (Invitrogen) as detailed in the manufacturer's protocol. rAAV particles were collected 3 days post-transfection by scraping cells from the 10 cm dish into 1 mL PBS pH 7.4 (Invitrogen) followed by 4 freeze/thaw cycles between a dry ice/ethanol bath and a 37°C water bath. Vector stocks were clarified by centrifugation and used fresh or stored at –80°C. The titer of the rAAV stock was ~$1x10^6$ viral particles/mL as verified by RT-PCR (25).

## Gene targeting in human fibroblasts

Homologous recombination is most efficiently facilitated by rAAV when target loci are in a euchromatic state due to active replication or transcription (7, 26). Therefore, to bias rAAV integration for homologous targeting instead of random integration events, rAAV was introduced to HFF cultures in log phase proliferation. Primary HFFs at second passage were grown to 40% confluence (~2x10$^6$ cells) in 10 cm plates. HFFs were then infected overnight using 333 µL *c-Myc* rAAV vector stock at a multiplicity of infection of ~0.5 particles/cell in 8 mL fresh medium. Infected cells were given fresh media 24 hours post infection and driven into quiescence by confluence arrest and serum deprivation as described above. To select infected cells that have the EGFP-Luc gene correctly inserted immediately downstream of the c-Myc translation start site and therefore under the control of *c-Myc* regulation, quiescent cells were stimulated with 20% serum for 6 hours. Stimulated cells were then trypsinized, strained through a 40 µM mesh to generate a single cell suspension at ~5x10$^6$ cells, and then sorted on a FACS Vantage with DiVa (Digital Vantage) upgrade (Becton Dickinson). The FACS Vantage was programmed to deliver single EGFP-Luc positive cells to individual wells of a 96 well tissue culture plate (Fig. 1B). Individual clones were expanded for genotype analysis.

## Genotype analysis

Genomic DNA was isolated from individual EGFP-Luc positive clones. For PCR genotyping, triplex PCR was performed using one primer specific for the LHA upstream of the EGFP-Luc insertion site (primer 1), one primer specific to the 3′ end of the EGFP-Luc fusion gene (primer 2), and one primer that recognizes sequence 3′ of the RHA that is outside the targeting construct (primer 3 - Fig. 2A). Primers 1 and 3 amplify a 1.5 kb product from the wildtype *c-Myc* allele. Primers 2 and 3 amplify a 1.1 kb product that indicates the EGFP-Luc fusion gene has been knocked into the *c-Myc* allele (Fig. 2B). For Southern blot analysis, 20 µg genomic DNA was digested with XbaI overnight, separated on a 0.8% 1xTAE agarose gel, and transferred to Hybond XL nylon membrane (Amersham). To identify correctly targeted *c-Myc* alleles membranes were hybridized with an α-$^{32}$P dCTP labeled *c-Myc* exon 3 probe (Fig. 2A and C). To identify random integrations, an EGFP-specific probe was used (data not shown). Targeted clones were finally confirmed by sequencing across the regions where the targeting construct juxtaposes the genomic DNA as well as the EGFP-Luc insert regions within the targeting construct to insure the reporter gene was inserted inframe (Fig. 2D).

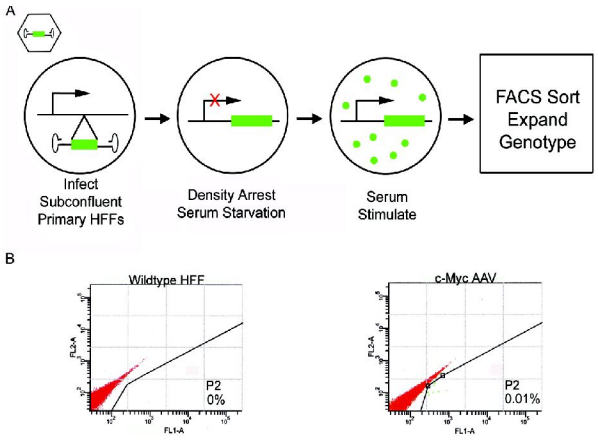


**Fig. 1: Direct selection of targeted EGFP-Luciferase reporter gene insertion into the *c-Myc* locus using known *c-Myc* regulatory conditions. (A)** Subconfluent primary human foreskin fibroblast cells are infected with rAAV targeting vectors when *c-Myc* is being actively transcribed or replicated in proliferative cells. Background reporter gene expression is silenced through density arrest and serum withdrawal. Reporter gene expression is induced by serum stimulation. Single EGFP-Luciferase positive cells are selected using FACS. Cloned cells are expanded and screened for gene targeting events by PCR, Southern blot, and sequencing. **(B)** FACS histograms are shown for wildtype and *c-Myc* rAAV targeted cells following cell cycle entry. FL1-A (EGFP) was plotted against FL2-A (no fluorofor) to center the parent population and allow for selection of the dim EGFP-Luciferase positive cells. The P2 region of the plot indicates the gating used for selection of the EGFP-Luciferase positive portion of the parent population.

## RT-PCR

Total RNA was isolated using Trizol reagent (Invitrogen) and cDNA synthesized in a 20 µL reaction from 1 µg total RNA using Superscript III (Invitrogen) and random nonamer primers (Takara). 1 µL cDNA was used as RT-PCR template to detect *c-Myc* and *EGFP-Luc* mRNA expression. The *c-Myc* wildtype allele transcripts were detected using a forward primer, GCTCGCCCAAGTCCTGC, which anneals in exon 2, and a reverse primer, GCTGATGTGTGGAGACGTGG, which anneals in exon 3 (Fig. 4A). EGFP-Luc transcripts were detected from a primer pair annealing 3′ in the EGFP-Luc coding region. The EGFP-Luc forward primer is TATGGGCTCACTGAGACTACATCA and the reverse primer is TCAGAGACTTCAGGCGGTCAA.

## Western Blot

Protein was collected in 50mM Tris pH 7.4, 1% NP40, 150mM NaCl, 1mM EDTA, 1mM $Na_3VO_4$, 1mM NaF, 1X

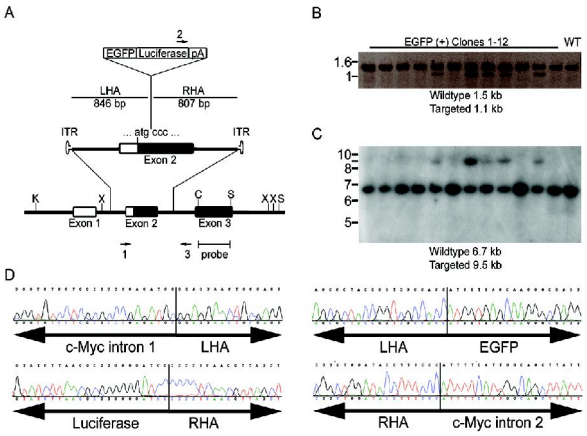


**Fig. 2:Genotype analysis of EGFP-Luciferase positive clonal populations. (A)** Diagram of the *c-Myc* rAAV targeting and genotype strategies. The targeting region of the endogenous *c-Myc* locus is shown. The left and right homologous arms (LHA & RHA), and viral inverted terminal repeat portions of the rAAV vectors facilitate insertion of the EGFP-Luciferase fusion gene directly between the first and second codon of the *c-Myc* gene. K, KpnI. X, XbaI. C, ClaI. S, SspI. **(B)** Triple primer PCR genotyping was performed using primers 1-3 shown in (A). **(C)** Southern blot analysis using the *c-Myc* exon 3 probe shown in (A). **(D)** Sequence data showing the regions spanning the *c-Myc* rAAV insertion sites and the internal EGFP-Luciferase insertion regions.

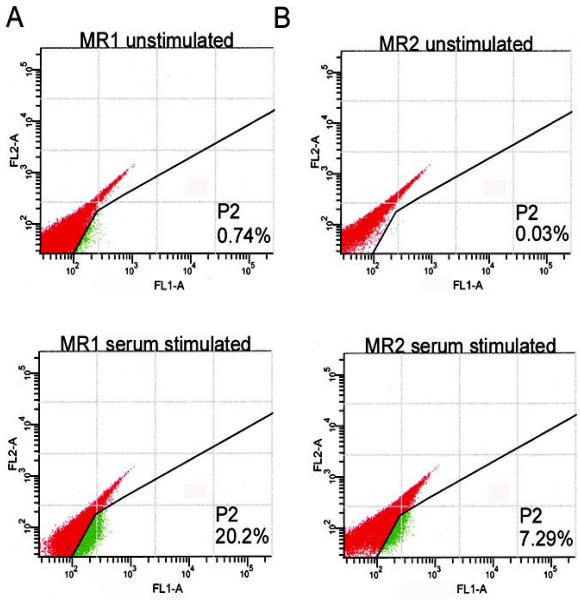


**Fig. 3: MR1 and MR2 response to cell cycle entry.** MR1 **(A)** and MR2 **(B)** clonal populations were subjected to cell cycle entry conditions and then evaluated for EGFP-Luc reporter gene expression. The unstimulated background control populations are shown in the top panels. Serum stimulated populations are shown in the bottom panels where the EGFP-Luciferase expressing cells are located in the P2 gated region.

Complete Protease Inhibitor (Roche). 10 µg of protein was separated by electrophoresis in 4-12% Bis-Tris NuPage gels (Invitrogen) and transferred onto nitrocellulose membrane (Nitrocellulose Transfer Membrane - Bio-Rad). Membranes were probed with 1:200 anti-c-Myc (9E10) (Santa Cruz), 1:2000 anti-GFP (JL-8) (Clontech), and 1:500 anti-Max (C-17) (Santa Cruz) antibodies.

### Luciferase Assays

Cells were collected from 10 cm dishes into 500 µL Tropix lysis solution supplemented with fresh 0.5 mM DTT. 10 µL of cellular lysate was added to 25 µL Dual Light Buffer A in a 96 well plate. Samples were assayed for Luciferase activity on a Tropix TR717 Microplate Luminometer (PE Applied Biosystems). Samples were run in triplicate and the average data are reported +/- SEM.

## RESULTS AND DISCUSSION

Transcription of the c-*Myc* gene is tightly regulated and events such as viral transduction, viral integration, chromosomal translocations and gene amplification that deregulate c-*Myc* transcription activate its well-characterized oncogenic potential (14). In addition, *c-Myc* mRNA and protein levels are deregulated or elevated in many tumors that show no physical disruption at the gene level. In the latter cases, deregulated c-*Myc* expression is thought to be due to oncogenic activation of mitogenic signal transduction pathways that regulate *c-Myc* gene expression (15, 16, 17). Indeed the *c-Myc* gene has been found to be induced by a wide variety of mitogenic proteins and suppressed by anti-mitogenic proteins (14). Although *c-Myc* transcription is regulated at a variety of levels (18), it remains unclear how the various mitogenic and anti-mitogenic signals converge on the *c-Myc* promoter to control gene expression (19). Therefore, to facilitate mechanistic studies of *c-Myc* gene regulation in human cells we devised a strategy, depicted in Fig. 1A, to generate primary human foreskin fibroblasts (HFF) that serve as reporters of *c-Myc* transcription. In this strategy, we take advantage of the well-documented induction of *c-Myc* transcription that occurs upon serum stimulation of quiescent cells (20, 21, 22) to select, by FACS sorting, for rAAV-mediated knockin of an EGFP-Luc reporter gene into the *c-Myc* locus. As described in Materials and Methods, the rAAV vector was designed to insert the 2690 bp promoter-less EGFP-Luc fusion gene into exon 2 of the *c-Myc* locus in frame with the major *c-Myc* translational start codon. Insertion of the reporter gene is predicted to produce a chimeric transcript comprised of the *c-Myc* 5′ untranslated region fused to EGFP-Luc. A

polyadenylation signal on the EFGP-Luc cDNA prevents read-through transcription of downstream sequences and therefore its insertion inactivates the targeted *c-Myc* allele. However, the reporter gene knockin is predicted to result in minimal disruption of the native cis regulatory sequences that govern *c-Myc* transcription so that the reporter gene will provide an accurate readout of endogenous *c-Myc* activity.

## Generating *c-Myc* gene reporter human fibroblast cell lines

Primary HFFs at passage 2 were infected with the *c-Myc* rAAV targeting vector for 24 hours. The infected cells were then driven into quiescence by combined confluence arrest and serum deprivation (Fig. 1A). These cells were stimulated to reenter the cell cycle by the addition of 20% serum for 6 hours and subjected to fluorescence activated cell sorting (FACS) for cells that expressed EGFP. Although, *c-Myc* is maximally induced between 2 and 4 hours following serum stimulation (Fig. 4), the 6 hour time point was chosen for sorting to ensure that cells had entered the cell cycle and because EFGP-Luc has a much longer half life (~15 hrs, data not shown) than the short, 20-30 minute half life of *c-Myc* (21, 23, 24). Of $5x10^5$ cells sorted, 48 cells (0.01% of the parent population) showed expression of EGFP-Luc (Fig. 1B). EGFP-positive cells were automatically collected and sorted into individual wells of a 96 well dish by the flow cytometry instrument. Of the 48 cells plated, 24 grew out to formed viable colonies. The 24 clonal populations were expanded, DNA extracted and had PCR and Southern blot genotyping performed. Five clones (21%) were found to carry a single targeted *c-Myc* allele (Fig. 2B and C). DNA from these five clones was subjected to sequence analysis to confirm that the EGFP-Luc gene was inserted correctly. All five showed correct in-frame insertion of the reporter beginning after the ATG of the major c-Myc translation start site. Sequence at the insertion junctions for a representative clone is shown in Fig. 2D. These results are consistent with previous results showing that AAV vectors can precisely insert DNA at specific genomic loci (2).

rAAV-mediated gene targeting vectors can also integrate at random locations (2). To determine whether random integration of the EGFP-Luc gene occurred in our selected clones, Southern blots were performed using a probe specific to the EGFP-Luc gene (not shown). This analysis showed three of the 24 clones with random integration

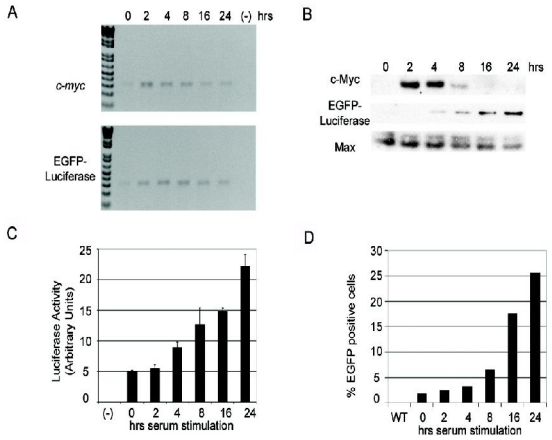


**Fig. 4: Comparison of *c-Myc* and EGFP-Luciferase reporter gene expression and activity kinetics in response to cell cycle entry.** The MR1 cells were tracked through CCE for their ability to induced *c-Myc* and *EGFP-Luciferase* transcripts by RT-PCR **(A)** and protein by Western blot **(B)**. Max protein expression is used as a loading control. **(C)** Luciferase activity of the MR1 population over CCE. Data shown is the average of triplicate samples +/- SEM. **(D)** FACS Calibur data is shown where the EGFP-Luciferase positive percent of the parent population was determined using BD CellQuest Pro v. 5.2. Time course changes of the percent EGFP-Luciferase expressing populations were compared under the same gates.

events and that Clone 9 had one random integration event (not shown) and one correctly targeted *c-Myc* allele (Fig. 2B and C). The other 4 positive clones showed only a single correctly targeted *c-Myc* allele (Fig. 2B and C) and no random integration events (not shown).

We conclude from these results that even though the rate of return of EGFP-Luc positive cells following infection of HFFs with the rAAV- *c-Myc* targeting vector is low, FACS provides a powerful selection method that can rapidly yield productive cell lines from very rare, but precise targeting events. It is of note that we have found that use of a high titer virus at a high multiplicity significantly improves the percent yield of EGFP-Luc positive cells compared to the low titer stocks used in these experiments.

## Comparison of endogenous *c-Myc* expression to *c-Myc* reporter activity

Clones 7 and 8 (designated *c-Myc* reporter 1 [MR1] and 2 [MR2]) were chosen for further characterization because these populations showed correct targeting and did not contain any random rAAV integration events. We first performed a cell cycle entry experiment using the same conditions used in the original selection scheme (Fig. 1A). FAC sorting 6 hours following serum stimulation yielded

20.2% of the MR1 population (Fig. 3A) and 7.29% of the MR2 population (Fig. 3B) positive for EGFP-Luc. This is compared to 0.74% and 0.03% of EGFP-positive cells observed in the unstimulated MR1 and MR2 populations respectively (Fig. 3). Thus, MR1 cells show a 27-fold induction and MR2 cells show a 243-fold induction of the EGFP reporter 6 hours after serum stimulation. Although there is a strong induction of EGFP, it was clear that not all cells induce the reporter by 6 hours. We therefore further examined the induction kinetics of the reporter gene compared to *c-Myc* expression during cell cycle entry of MR1 cells to determine whether the inserted EGFP reporter genes recapitulated expression characteristics of the *c-Myc* gene.

Induction of *c-Myc* transcription and protein following serum stimulation of quiescent fibroblasts peaks between 2-4 hours, and subsequently declines to low, but measurable levels by 24 hours and throughout the cell cycle (20, 21, 22, 23, 24). Cell cycle entry experiments were conducted using MR1 cells collected at 0, 2, 4, 8, 16, and 24 hours post stimulation. *c-Myc* and *EGFP-Luc* transcripts were induced upon serum stimulation and subsequently declined with near-identical kinetics (Fig. 4A). EGFP-Luc was also strongly induced, but its induction to measurable levels appeared to be delayed compared to c-Myc protein (Fig. 4B). It is not clear why EGFP is not detected by 2 hour after serum stimulation, as c-Myc is, but this may have to do with the relative strength of the antibodies, or due to a slower rate of translation of the *EGFP-Luc* fusion mRNA compared to *c-Myc* mRNA. EGFP-Luc protein levels also differed from endogenous c-Myc protein levels in that whereas c-Myc levels declined after 4 hours, EGFP-Luc continued to accumulate throughout the 24 hour period monitored (Fig. 4B). The progressive accumulation of EGFP-Luc is a reflection of the much longer half-life of EGFP-Luc fusion protein (~15hrs) compared to the 20-30 minute half-life of c-Myc.

The accumulation of EGFP-Luc protein following serum stimulation was also reflected in a progressive increase over the 24 hour period in Luciferase activity (Fig. 4C) and in an robust increase in the percentage of EGFP-positive cells (Fig. 4D). Notably, the fold induction of Luciferase activity (4.5 fold) was lower that that of EGFP (10 fold) (Fig. 4C and D). The reason for this discrepancy is not clear, but may reflect underlying differences in the detection instruments. Importantly, the low background of EFGP-Luciferase in quiescent cells together with the progressive accumulation of EGFP-Luc signal, appear to provide conditions that allow an amplification of events, like serum stimulation, that trigger *c-Myc* transcription.

### Summary

We demonstrate here a method for rapidly producing human gene reporter cells by rAAV-mediated gene insertion, and characterize a *c-Myc* gene reporter cell strain that we developed using this method. Our method for reporter insertion is applicable to any gene whose expression can be induced from a low basal level. For example, a number of genes such as cyclins and cell cycle inhibitor/checkpoint genes have well-defined conditions in which their expression can be induced from a low level. Reporter cells for these genes and in particular genes that are involved in cancer and other diseases offer a possible platform for the identification of proteins and molecules with potential therapeutic value.

## ACKNOWLEDGMENTS

This work was funded by a grant from Shriners Hospitals for Research to PJH and grants from the U.S. National Institutes of Health to DWR.

## REFERENCES

1. Sedivy JM, Dutriaux A. Gene targeting and somatic cell genetics--a rebirth or a coming of age? Trends Genet. 1999;15:88-90.
2. Hirata R, Chamberlain J, Dong R, Russell DW. Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. Nat. Biotech. 2002;20:735-738.
3. Chamberlain JR, Schwarze U, Wang P, Hirata RK, Hankenson KD, Pace JM, Underwood RA, Song, KM, Sussman M, Byers PH, Russell DW. Gene targeting in stem cells from individuals with osteogenesis imperfecta. Science 2004;303:1198-1201.
4. Topaloglu O, Hurley PJ, Yildirim O, Civin CI, Bunz F Improved methods for the generation of human gene knockout and knockin cell lines. Nucleic Acids Res. 2005;33:e158
5. Bunz F. Human cell knockouts. Curr. Opin. Oncol. 2002;14:73-78.
6. Russell DW, and Hirata RK. Human gene targeting by viral vectors. Nat. Genet. 1998;18:325-330.
7. Vasileva A, Jessberger R. Precise hit: adeno-associated virus in gene targeting. Nat Rev. Microbiol. 2005;3:837-847.
8. Hirata RK, Russell DW. Design and packaging of adeno-associated virus gene targeting vectors. J. Virol. 2000;74:4612-4620.

9. Porteus MH, Cathomen T, Weitzman MD, Baltimore D. Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. Mol. Cell. Biol. 2003;23:3558-3565.
10. Kohli M, Rago C, Lengauer C, Kinzler KW, Vogelstein B. Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. Nucleic Acids Res. 2004;32(1):e3.
11. Liu X, Yan Z, Luo M, Zak R, Li Z, Driskel, RR, Huang Y, Tran N, Engelhardt, JF. Targeted correction of single-base-pair mutations with adeno-associated virus vectors under nonselective conditions. J. Virol. 2004;78;4165-4175.
12. Hendrie PC, and Russell DW. Gene targeting with viral vectors. Mol. Ther. 2005;12:9-17.
13. Vasileva A, Linden RM, Jessberger R. Homologous recombination is required for AAV-mediated gene targeting. Nucleic Acids Res. 2006;34:3345-3360.
14. Grandori C, Cowley SM, James LP, Eisenman RN. The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu. Rev. Cell Dev. Biol. 2000;16:653-699.
15. Nesbit CE, Tersak JM, Prochownik EV. MYC oncogenes and human neoplastic disease. Oncogene 1999;18:3004-3016.
16. Christoph F, Schmidt B, Schmitz-Drager BJ, Schulz WA. Over-expression and amplification of the c-myc gene in human urothelial carcinoma. Int. J. Cancer 1999;84:169-173.
17. Erisman MD, Rothberg PG, Diehl RE, Morse CC, Spandorfer JM, Astrin SM. Deregulation of c-myc gene expression in human colon carcinoma is not accompanied by amplification or rearrangement of the gene. Mol. Cell. Biol. 1985;5(8): 1969-1976.
18. Spencer CA, Groudine M. Control of c-myc regulation in normal and neoplastic cells. Adv. Cancer Res. 1991;56:1-48.
19. Weber A, Liu J, Collins I, Levens D. TFIIH operates through an expanded proximal promoter to fine-tune c-myc expression. Mol. Cell. Biol. 2005;25:147-161.
20. Persson H, Gray HE, Godeau F. Growth-dependent synthesis of c-myc-encoded proteins: early stimulation by serum factors in synchronized mouse 3T3 cells. Mol. Cell. Biol. 1985;5:2903-2912.
21. Rabbitts PH, Watson JV, Lamond A, Forster A, Stinson MA, Evan G, Fischer W, Atherton E, Sheppard R, Rabbitts TH. Metabolism of c-myc gene products: c-myc mRNA and protein expression in the cell cycle. EMBO J. 1985;4:2009-2015.
22. Dean M, Levine RA, Ran W, Kindy MS, Sonenshein GE, Campisi J. Regulation of c-myc transcription and mRNA abundance by serum growth factors and cell contact. J Biol. Chem. 1986;261:9161-9166.
23. Ramsay G, Evan GI, Bishop JM. The protein encoded by the human proto-oncogene c-myc. Proc. Natl. Acad. Sci. U S A 1984;81:7742-7746.
24. Hann SR, Thompson CB, Eisenman RN. c-myc oncogene protein synthesis is independent of the cell cycle in human and avian cells. Nature 1985;314:366-369.
25. Veldwijk MR, Topaly J, Laufs S, Hengge UR, Wenz F, Zeller WJ, Fruehauf S. Development and optimization of a real-time quantitative PCR-based method for the titration of AAV-2 vector stocks. Mol Ther. 2002;6:272-278.
26. Trobridge G, Hirata RK, Russell DW. Gene targeting by adeno-associated virus vectors is cell-cycle dependent. Hum Gene Ther. 2005;16:522-526.

# PROTOCOL

## Generating reporter knockin cell line

1. Determine gene to be targeted based on its ability to be induced under defined cell culture conditions (i.e. cell cycle entry).
2. Clone target gene-specific arms into pELCV vector. The 5′ arm should include the translation start codon of the target gene and cloning should be engineered to generate in-frame fusion with the GFP-Luciferase reporter in pELCV. The goal being to maintain endogenous gene regulatory sequences and context while introducing the reporter gene.
3. Sequence vector to ensure junction sequences are correct.
4. Generate infectious recombinant AAV.
5. Infect cells under logarithmic growth to favor sequence-specific recombination.
   Activate gene transcription of GFP-Luc reporter targeted gene (e.g. cell cycle entry for c-Myc).
6. Used FAC sorting to select cells that show induction of GFP that mimics expression of the targeted gene.
7. Directly plate GFP-positive single cells in 96 well plate or pool positive cells for expansion.
8. Expanded GFP-positive clones or pooled populations can either be genotyped at this point or run through an “enrichment” step. For enrichment, cells are put through a second round of selection using the gene-inducing conditions followed by FAC sorting of GFP-position cells. Enriched cell populations are expanded and genotyped.
9. Correctly targeted cells are used in assays in which expression of the targeted gene, as determined by either GFP (e.g. using FACS) or luciferase activity is the readout. Either approach could be used in a high-throughput screening format.