DEVELOPMENT OF NOVEL HUMAN REPORTER CELL LINES USING RECOMBINANT ADENO-ASSOCIATED VIRUS MEDIATED HOMOLOGOUS RECOMBINATION

by
Sandra L. Fernandez

A THESIS

Presented to the Department of Cell and Developmental Biology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Master of Science
June 2007

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the M.S. thesis of

Sandra L. Fernandez

has been approved

[redacted]
Mentor/Advisor

[redacted]
Member

[redacted]
Member

[redacted]
Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

This work was funded by a grant from Shriners Hospitals for Children Research to Peter J. Hurlin and grants from the U.S. National Institutes of Health to David W. Russell. Special thanks to Moroni Scott Fernandez for his support throughout my graduate research.

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 (rAAV) 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 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

Eukaryotic transcriptional regulation

Regulation of gene expression can occur at every step in the process of converting DNA into functional proteins. Individual genes display different combinations of expression control at the levels of transcription, mRNA localization, processing, and stability, translational control, and post-translational modifications. Transcriptional control is a key regulatory point for many genes since blocking this step allows the tightest control of gene expression by ensuring that no superfluous gene products will be made (Alberts 2002).

Transcription initiation in eukaryotes requires recruitment of the RNA polymerase II holoenzyme via binding of general transcription factors to the core promoter region just upstream of a gene's transcriptional start site (Fig. 1). Gene specific transcription factors, both activating and repressing, bind to proximal and distal cis acting DNA regulatory elements positioned upstream, intergenic and downstream of the gene and influence the ability of RNA polymerase II to begin transcription (Alberts 2002). In addition, activating factors enhance transcription through the recruitment of histone acetyl transferase proteins, which covalently modify histones leading to nucleosome remodeling. This activity opens the chromatin allowing gene regulatory factors to bind as well as the providing acetyl-lysine binding sites for bromodomains of the general transcription factors (Razin 2007).

Methods used to study transcription

There are two main approaches used to study transcription activation. The first approach is to directly detect the presence of a specific transcript. This is done using

methods such as northern blot and qRT-PCR assays on RNA molecules isolated from cells or tissues. A second approach for studying transcription is to monitor the expression of a reporter gene that is under the control of the promoter of interest. Generally, a minimal promoter is cloned into a reporter plasmid and then transfected into cells. The presence of activating transcription factors is then determined through the detection of reporter molecules.

Northern blot and qRT-PCR provide the most direct methods for detecting transcripts, however they are laborious. Care must be taken when preparing and manipulating the volatile RNA to prevent degradation. For northern analysis the RNA must first be separated on a gel and then fixed to a membrane followed by detection using radiolabled probes. The final readout from these experiments can take about a week to obtain. qRT-PCR provides a more rapid analysis of transcripts but requires many manipulations between harvesting the RNA and getting a final qRT-PCR curve. At each step pipetting error can introduce huge differences in apparent gene expression. Also, qRT-PCR requires extensive optimization to ensure only the proper PCR product is being examined for quantification.

Detection of a reporter molecule, such as Luciferase, as a readout of promoter activation is a quick and easy solution to the limitations of detecting transcripts directly. The Luciferase molecules are rapidly detected directly from cell lysates and provide a quantitative readout of gene expression. However, the reporter gene is being expressed under the control of a minimal or partial promoter, which omits multiple or far upstream signals that are present in native chromatin. Furthermore, transcription within the

artificial context of a plasmid lacking any of the secondary structures such as nucleosomes ignores many of the intricacies known to regulate endogenous genes.

High-throughput screening methods can lead to discovery

A promising approach employed by the pharmaceutical and biotechnology industries is high-throughput screening. High-throughput applications allow researchers to rapidly screen library arrays containing huge numbers of molecules such as cDNAs, siRNAs, or small molecules. The unbiased nature of blindly screening thousands of potential regulatory molecules in high-throughput applications is a major advantage inherent to the approach and can lead to the discovery of new therapeutic compounds. High-throughput assays ideally encompass a rapid and straightforward readout for the modulation of a biochemical pathway while maintaining the endogenous cells function. The methods described above for the study of transcriptional activation are not well suited for high-throughput applications due to assay length (northern blot), multiple manipulations and the associated pipetting error (qRT-PCR), and omission of huge portions of a gene's native regulatory regions (reporter plasmids).

Creation of human reporter cell lines

The goal of the research described herein is to develop a human cellular system amenable to high-throughput applications that allows discovery and verification of transcriptional regulatory molecules. The approach is to take advantage of the facile reporter gene assay while maintaining all of the native regulatory elements for the gene of interest by inserting a reporter molecule directly into a genetic locus. Specifically, a promoterless EGFP-Luciferase-SV40 poly A (EGFP-Luc) fusion gene will be inserted into a native locus directly downstream of a transcriptional start site under the control of

the endogenous promoter. Human somatic cells, both primary and immortal, will be used in this approach thus maintaining direct relevance to human gene regulation or disease of any newly discovered regulatory molecules.

Inserting reporter genes into the genome through targeted homologous recombination

Introduction of the EGFP-Luc reporter gene into a specific locus in human somatic cells presents a technical challenge. Plasmid based homologous recombination has been shown as a viable method for introducing genetic changes into human somatic cells (Sedivy 1999A). Unfortunately, plasmid based gene targeting in somatic cells is hampered by very low rates of targeted homologous recombination (Sedivy 1999B). In addition, somatic cells do not respond well to the positive and negative selection methods that have improved the gene targeting efficiencies in mouse embryonic stem cells (Vasquez 2001). Although plasmid based gene targeting and selection are superior in mouse embryonic stem cells, our purpose is to develop a high-throughput cellular system that accurately reflects human gene regulation. Thus, it is not desirable to use mouse embryonic cells for this application.

Recombinant adeno-associated virus mediated homologous recombination

Recent studies have shown high efficiency gene targeting through the use of recombinant Adeno-Associated Virus (rAAV) mediated homologous recombination in human somatic cells (Hirata 2002, Porteus 2003, Chamberlain 2004). The efficiency of gene targeting by rAAV vectors surpasses traditional plasmid-based methods by several fold (Bunz 2002, Topaloglu 2005), with up to 1% of unselected cells undergoing targeting (Russell 1998, Hirata 2002). A number of rAAV-mediated targeting strategies

have been successfully employed (for examples see Hirata 2000, Hirata 2002, Kohli 2004, Liu 2004, Hendrie 2005, Topaloglu 2005, Vasileva 2006).

Wildtype AAV is a single stranded DNA parvovirus with a 4.7 kb genome flanked by two palindromic sequences, that form hairpin structures, called inverted terminal repeats (ITR). The wild-type genome is comprised of two open reading frames, *rep* and *cap,* encoding non-structural and structural genes, respectively, which can be replaced in recombinant versions of the virus and supplied in trans for recombinant virion production. AAV is not associated with any human pathology and is naturally replication deficient, requiring a helper virus, such as adenovirus, to produce infectious particles. Helper virus genes can also be supplied in trans (adenovirus E2A, E4, VA) making rAAV a superior virus for gene introduction due to its high efficiency transduction and superior biosafety (Vaslieva 2005).

In the absence of helper virus wild-type AAV integrates into a specific region of chromosome 19 and persists as a latent infection. rAAV vectors with genomic homology can also integrate specifically into the genome at the corresponding site of homology. The current model for targeted AAV integration, at chromosome 19 or elsewhere, is that the AAV genomes are bound by DNA double strand break repair proteins such as KU86 and RAD52 which then mediated the integration of the viral single stranded DNA into complementary portions of the genome through gene conversion or crossover events. Non-specific insertion can also occur at double strand break points through non-homologous end joining (Vaslileva 2005).

Limitations of using rAAV in gene targeting

The vectors used in rAAV targeting strategies have the general features of viral ITRs flanking introduced homologous arms that mediate recombination to a specific genomic locus, which in turn flank a drug selection cassette (Fig. 2). 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 (Vasileva 2005). 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 base pairs (bp) into the rAAV targeting vector. Our reporter gene of choice is EGFP-Luc, which has the advantageous property of both visual (EGFP) and quantifiable (Luc) readouts of gene expression. Unfortunately, EGFP-Luc-pA is 2.7 kilobase pairs (kbp) long making it unusable in conjunction with a drug selectable marker in traditional rAAV targeting vectors. Even with the omission of Luciferase the EGFP-pA at around 900 bps would be pushing the limits for rAAV mediated gene targeting using the drug selection method.

A new selection strategy for rAAV gene targeting events

Since the inclusion of a standard selectable marker in rAAV gene targeting vectors preclude the use of many reporter genes due to size constraints, we have developed a new selection strategy for rAAV targeting. Our method relies on

manipulation of known gene regulatory conditions to directly select for the introduction of a reporter gene into a specific endogenous locus (Fig. 3). rAAV particles are introduced to cells when the gene target is being either actively transcribed or replicated leaving the locus open and available for homologous recombination. Infected cells are switched to conditions where the target gene is silenced, removing background reporter gene expression. The cells are then introduced to cellular conditions where the target gene is activated and the reporter gene is expressed. Cells that express the EGFP-Luc reporter gene as expected for the given cellular conditions are selected using fluorescence activated cell sorting (FACS). Cells collected must be screened to verify the presence of the targeted allele and can then be used for transcriptional regulation studies. This strategy is applicable for targeting any gene with well-defined growth conditions where the target gene can be induced from low levels. Here we describe the use of this method to develop primary and immortal human cells in which a promoter-less EGFP-Luc fusion gene has been introduced into the *c-Myc* locus under the direct control of the endogenous *c-Myc* promoter.

Proof of concept: Targeting EGFP-Luc into the *c-Myc* locus

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 (Grandori 2000). 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 (Nesbit 1999,

Christoph 1999, Erisman 1985). Indeed the *c-Myc* gene has been found to be induced by a wide variety of mitogenic proteins and suppressed by anti-mitogenic proteins (Grandori et al. 2000). Although *c-Myc* transcription is known to be regulated at a variety of levels (Spencer 1991), it remains unclear how the various mitogenic and anti-mitogenic signals converge on the *c-Myc* promoter to control gene expression (Weber 2005). Therefore, to facilitate mechanistic studies of *c-Myc* gene regulation in human cells we will create primary human foreskin fibroblasts (HFF) and immortal U2-OS cells that express EGFP-Luc as a reporter 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 (Fig. 4, Persson et al.1985, Rabbitts et al. 1985, Dean 1986) to select, by FACS sorting, for rAAV-mediated knockin of an EGFP-Luc reporter gene into the *c-Myc* locus (Fig. 3).

Chapter 1

The information in this chapter is taken largely from the manuscript in prep entitled "Development of human reporter cell lines using rAAV mediated homologous recombination"

Sandra L. Fernandez, David W. Russell, Peter J. Hurlin

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 (rAAV) 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 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.

Materials and Methods

Reagents

Klenow DNA polymerase and all restriction endonucleases were obtained from New England Biolabs and used according to the manufacturer specifications. T4 DNA ligase (Invitrogen) was used for all cloning ligations according to the manufacturer

specifications. All primers were obtained from the Shriners Research DNA core (Portland, OR).

Tissue culture

Primary HFFs were generously provided at passage one by Carla Grandori (Fred Hutchinson Cancer Research Center). AAV-293 cells were obtained from Stratagene. All cells 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) unless otherwise indicated. All cultures were maintained in incubators at 37°C and 5% CO_2.

Serum stimulation

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, Fig. 5). A promoter-less EGFP-Luciferase-SV40 pA fusion gene was obtained from pEGFPLuc (Clontech). 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, seen in Figure 6, 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 (Stratagene) vector backbone. ITR flanked *c-Myc* targeting vector integrity was confirmed by AhdI restriction mapping (Fig. 7).

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 (Fig. 8, Veldwijk 2002).

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 (Trobridge 2005,Vasileva 2005). 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 (MOI) of ~0.2 particles/cell in 8 mls 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/mL, 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. 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. 9). 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. 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. 9). To identify random integrations, an EGFP-specific probe was used (Fig. 9). 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. Sequencing was performed on PCR product spanning the LHA and RHA regions where the product amplified included regions outside of the targeting construct to insure we weren't sequencing random integrants.

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) according to the respective manufacturer's protocols. 1 μl cDNA was used as RT-PCR template to detect *c-Myc* and *EGFP-Luc* mRNA expression in a standard Platinum Taq (Invitrogen) PCR reaction supplemented with 5mM $MgCl_2$, 3% DMSO, 1M Betaine, and 0.4 μM of each primer. The *c-Myc* wildtype allele transcripts were detected using the RT1 forward primer, GCTCGCCCAAGTCCTGC, which anneals in exon 2, and the RT2 reverse primer, GCTGATGTGTGGAGACGTGG, which anneals in exon 3 (Fig. 6). EGFP-Luc transcripts were detected from a primer pair annealing 3' in the EGFP-Luc coding region. The EGFP-Luc forward primer, RT3, is TATGGGCTCACTGAGACTACATCA and the reverse primer, RT4, is TCAGAGACTTCAGGCGGTCAA (Fig. 6). As a control, β-actin was amplified from the prepared cDNAs using TGAAGATCCTCACCGAGCG as the forward primer and ACTGTGTTGGCGTACAGGTCTTT as the reverse primer. PCR conditions were: 30 seconds at 95°C, 30 seconds at 52°C, and 30 seconds at 72°C repeated for 20 cycles.

Western blot

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

Luciferase activity was determined by the Tropix Dual Light system (PE Applied Biosystems). Cells were collected from 10 cm dishes into 500 μL Tropix lysis solution supplemented with fresh DTT at 0.5 mM. 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 the Tropix TR717 Microplate Luminometer (PE Applied Biosystems) using 100 μL Dual Light Buffer B injection, 5 second delayed injection intervals, and 5 second measurement time intervals. Time course samples were run in triplicate and the average data +/- SEM are given.

FACS analysis

EGFP fluorescence activity was determined by FACS analysis. Cells were harvested by trypsinization and filtered through a 40 μm mesh to obtain a single cell suspension. Cell suspension concentration was adjusted to ~$2x10^6$ cells/mL in cold PBS and kept on ice prior to sorting on the FACSCalibur 4-color analyzer (Becton Dickinson). Percent EGFP-Luciferase positive cells were determined using BD CellQuest Pro version 5.2. Time course changes of EGFP-Luciferase expressing populations were compared under the same gates.

Cyclohexamide assay

To determine the half-life of the EGFP-Luciferase fusion protein a cyclohexamide degradation assay was performed. The pEGFPLuc plasmid (Clontech) was transiently transfected into U2-OS cells (ATCC HTB-96) at 95% confluence using lipofectamine 2000 (Invitrogen) as prescribed by the manufacturer. 5 hrs post transfection the cells were trypsinized from three 10-cm dishes, pooled, and plated into eighteen 35-mm dishes

roughly maintaining the surface area of the original transfection plates so little expansion occurred. 24 hrs post transfection the cells were treated with 50 μg/mL cyclohexamide to block *de novo* protein synthesis. Cells were harvested at 0, 3, 5, 8, 12, 26, 31, and 48 hours post-cyclohexamide treatment. Two plates were harvested at each time point in 100 μL of either protein lysis buffer or Tropix lysis buffer for western blot and Luciferase activity analysis, respectively.

Results and Discussion

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 (Fig. 9). 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. 3). These cells were stimulated to re-enter the cell cycle by the addition of 20% serum for 6 hours and subjected to FACS sorting 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, Fig. 10) than the short, 20-30 minute half life of *c-Myc* (Ramsay et al. 1984, Hann et al. 1985, Rabbitts et al. 1985). Of $5x10^5$ cells sorted, 48 cells (0.01% of the parent population) showed expression of EGFP-Luc (Fig. 11). 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. 12, 13). 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 Figure 14. These results are consistent with previous results showing that AAV vectors can precisely insert DNA at specific genomic loci (Hirata 2002).

rAAV-mediated gene targeting vectors can also integrate at random locations (Hirata 2002). 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 (Fig. 9). This analysis showed three of the 24 clones (clones 3,6,&9) with random integration events and that Clone 9 had one random integration event (Fig. 15) and one correctly targeted *c-Myc* allele (Fig. 12, 13, 15). The other 4 positive clones showed only a single correctly targeted *c-Myc* allele (Fig. 12, 13) and no random integration events (Fig. 15).

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.

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 serum stimulation experiment using the same conditions used in the original selection scheme (Fig. 3,4). FACS sorting 6 hours following serum stimulation yielded 20.2% of the MR1 population (Fig. 16A bottom panel) and 7.29% of the MR2 population (Fig. 16B bottom panel) 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. 16 top panel). 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 serum stimulation of MR1 and MR2 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 (Persson et al. 1985, Ramsay et al. 1984, Hann et al. 1985, Rabbitts et al. 1985, Dean 1986). Serum stimulation

experiments were conducted using either MR1 or MR2 cells collected at 0, 2, 4, 8, 16, and 24 hours post stimulation. The data shown in Figures 17-20 delineate the MR2 response to serum stimulation, however MR1 yielded identical results. *c-Myc* and *EGFP-Luc* transcripts were induced upon serum stimulation and subsequently declined with near-identical kinetics (Fig. 17). EGFP-Luc protein was also strongly induced, but its induction to measurable levels appeared to be delayed compared to the c-Myc protein (Fig. 18). 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. 18). The progressive accumulation of EGFP-Luc is a reflection of the much longer half-life of EGFP-Luc fusion protein (Fig. 10) compared to the 20-30 minute half-life of c-Myc. Thus, although only a portion of the *c-Myc* reporter cell respond to serum stimulation conditions the population as a whole has maintained its ability to recapitulate the expected induction 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. 19) and in a robust increase in the percentage of EGFP-positive cells (Fig. 20). Noteably, the fold induction of Luciferase activity (4.5 fold) was lower that that of EGFP (14 fold) (Fig. 19, 20). 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 easily detected amplification of events, like serum stimulation, that trigger *c-Myc* transcription.

Chapter 2

Abstract

The clonal primary *c-Myc* reporter strains MR1 and MR2 described in Chapter 1 have theoretical utility in elucidation of *c-Myc* transcriptional regulatory pathways. However, primary cells have a long but limited capacity for expansion in culture. Furthermore, the harsh conditions associated with cloning a single primary cell results in the clonal population undergoing senescence at a premature mean population doubling (MPD). Thus it is desirable to re-derive the *c-Myc* reporter cell strain using conditions that preserve the maximum expansion potential of the primary HFF cells. To accomplish this, we have devised a new strategy to establish a *c-Myc* reporter strain from a mixed "bulk" population through a series of enrichment steps using FACS sorting. The elimination of a cloning step produced a *c-Myc* reporter strain that had undergone fewer MPDs and displayed cellular morphology and growth rates consistent with a younger population. The bulk *c-Myc* reporter strain responded to serum stimulation conditions as expected and will be a useful reagent for studying *c-Myc* transcriptional regulation.

Introduction

Primary cells have long been known to have a limited capacity for expansion in culture (Hayflick 1961). After explantation, primary cells undergo a period of rapid proliferation. After repeated subculture the primary cells begin to grow larger displaying a changed morphology, debris accumulation, and slowed growth. The cells ultimately lose their ability to proliferate. This phenomenon is referred to as replicative senescence (see Cristofalo 1993 and Goldstein 1990 for review). The number of population doublings a cell strain is able to undergo in culture is thought to be proportional to the

lifespan of the species and to the age of the animal they were explanted from (Rhome 1981). Indeed, reports from various laboratories have confirmed that neonatal HFF can undergo ~60-80 MPDs before reaching replicative senescence (Garkavtsev 1997).

At the onset of this research we deemed the use of a primary diploid human cell to be a critical tool for studying normal human transcriptional regulation. Although primary HFF cells have a limited lifespan, a single targeted HFF that undergoes 60-80 MPDs should produce a sufficiently large population to be used in transcriptional studies following the cloning process. However, through routine subculture of the MR1 and MR2 cells strains it became clear that these cloned primary HFF strains were prematurely ageing and not likely to complete the expected 60-80 MPDs.

We believe that the accelerated ageing is due to the harsh conditions associated with cloning. Anti-proliferative cellular programs are activated in the absence of cell growth signals. While FBS supplies some growth factors in the culture media, the growth signals obtained through matrix adhesions and neighbor secreted growth factors are absent in the early cloning process. Although some cells can eventually overcome these limitations, it appears that it occurs to the detriment of longevity. Thus it was necessary to devise a new strategy for collecting the targeted cells in order to preserve their maximum lifespan potential.

Since low cell density is not well tolerated by primary cells, we will collect all EGFP positive cells together as a "bulk" or mixed population. The collected population will be a mixture of untargeted, targeted, and randomly targeted cells. Collection of the bulk population into the smallest appropriate culture dish will help to maximize cell

density and promote proliferation. The proliferative potential will also be aided by supplying cell specific growth factors through the use of conditioned media.

Although it is likely that our studies do not have a rigid requirement for 100% purity, our population must be comprised primarily of targeted cells in order to obtain useful levels of reporter gene expression after a transcription activation event. Thus, the bulk population must be enriched for cells that contain the targeted allele. The enrichment process is merely one or more round of serum stimulation on the quiescent bulk population followed by collection of EGFP expressing cells by FACS sorting.

We expect the collected cells to largely be expressing the reporter gene under the control of the *c-Myc* promoter due to the targeted stimulation of the *c-Myc* promoter that occurs during serum stimulation. However, it does remain possible that through random integration events the reporter gene could be inserted downstream of a promoter that also responds to serum stimulation. These cells would be included in the enriched population and cannot be purged from the population using these selection methods. Likewise, random integration could also occur behind a strong constitutively active promoter. Since our selection relies solely on reporter gene expression these cells would also remain in the population and would likely confound future transcriptional studies. Although these events are not likely, we must rigorously verify that the bulk population contains the properly targeted *c-Myc* allele and that they respond properly to serum stimulation conditions.

Materials and Methods

Cell culture

Primary HFF cells 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). After FACS selection the targeted cells were grown in conditioned media until they reached confluence in a 10 cm dish. To maintain maximal cell density the FACS selected cells were expanded in incremental tissue culture plates from a 96 well plate up to a 10-cm dish. Similarly, the FACS selected cells were never diluted more than 1:2 or 1:3 during expansion and routine culture. All cultures were maintained in incubators at 37°C and 5% CO_2.

Conditioned media preparation

Early passage primary HFFs were plated at 50% confluence (2.5×10^6 cells in a 10 cm dish) in 20 mL of media. The cells were allowed to grow for 7-14 days until the media changed to an orange color. The media was then collected from the cells and purified through a low protein binding 0.2 μm PES filtration unit (Nalgene). Before use the conditioned media was diluted 1:1 with 20% FBS DMEM.

Construction and packaging rAAV gene targeting vectors

The *c-Myc* EGFP-Luc targeting vector described in Chapter 1 was used to produce high titer vector stocks that were kindly provided by David W. Russell (University of Washington). rAAV viral particles were prepared as described in Zolotukhin *et. al.* (1999) and Hirata *et. al.* (2002). The rAAV vector stocks were

provided at 4.3 x 10^8 particles/μL as determined by alkaline southern blot (Inoue 1998 and Hirata 2002).

Gene targeting in human fibroblasts

Primary HFFs at second passage were grown to 30% confluence (~1.5x10^6 cells) in 10 cm plates. HFFs were then infected overnight using 50 μL *c-Myc* rAAV vector stock at a MOI of ~14,000 particle/cell in 5 mls fresh medium. Infected cells were given fresh media 24 hours post infection and driven into quiescence by confluence arrest and serum deprivation as described in Chapter 1.

To select for reporter gene insertion events into the *c-Myc* locus, quiescent cells were stimulated with 20% serum for 4 or 6 hours. Stimulated cells were then trypsinized, strained through a 40 μM mesh to generate a single cell suspension at ~5x10^6 cells/mL, and then sorted on a FACS Vantage with DiVa (Digital Vantage) upgrade (Becton Dickinson). The FACS Vantage was programmed to deliver all EGFP-Luc positive cells to individual wells of a 96 well tissue culture plate. Bulk sorted populations were then expanded for genotype analysis.

Enrichment of bulk *c-Myc* reporter strains

To purify the bulk sorted mixed genotype populations, a subsequent round of serum stimulation and FACS sorting selection was performed. The original bulk populations were driven into quiescence through density arrest and serum withdrawal as described above. EGFP-Luc positive cells from the unstimulated and serum stimulated samples were bulk sorted as individual populations using the FACS Vantage with DiVa upgrade. The enriched populations were expanded for genotype analysis and characterization of their response to serum stimulation as described in Chapter 1.

MPD and growth rate determination

At each passage, cells suspensions were counted using an Improved Neubauer hemacytometer (Hausser Scientific) to determine the total population size and the amount of cells to be seeded for subculture. MPD were calculated according to the formula: $2^n = P - P_o$ where n = the mean population doubling, P = the population size, and P_o = the amount of cells seeded. Growth rates were calculated as the MPD/day.

Results and Discussion

The first sign that the MR1 and MR2 cells were prematurely ageing was a changed cellular morphology where the cells became enlarged and spindly. The larger cell size was reflected in a lower cell count at confluence than observed for non-clonal HFF populations (Table 1). The cells also experienced a slow in growth that is reflected in a dramatic decrease in growth rates (Table 2, Fig. 21). "Middle aged" MR1 and MR2 strains (45-48 MPD) had growth rates comparable to that of a non-clonal (WT HFF) population near senescence (75 MPD). Some fluctuations are seen in the growth rate curves in Figure 21 and are likely due to passaging differences and/or counting errors as these data were collected for a cursory tally of population age and not as part of a precise side-by-side ageing study. However, the overall trend of a prematurely slowed growth rate is clear for both clonal populations compared to the non-clonal population.

Generating *c-Myc* gene reporter bulk HFF strains

For the reasons discussed above, we will re-create the primary HFF *c-Myc* reporter cells using conditions that preserve the longevity of the cell strain. HFF cells at passage 2 were infected with the high titer *c-Myc* rAAV at a MOI of 14,000 for 24 hours. The use of the rAAV at this high MOI is consistent with previous successful studies

(Hirata 2002) and should help to improve the targeting efficiency for these experiments. Following infection, the cells were driven into quiescence as described in Chapter 1. Quiescent populations were then stimulated with 20% serum for either 4 or 6 hours. Following stimulation, the EGFP-Luc positive cells were selected using FACS sorting as described in Chapter 1 with the exception that all EGFP-Luc positive cells were collected together as a single bulk population (Fig. 22). The inclusion of the 4 hour time point in this experiment was an attempt to capture the purest possible bulk population. Selection of cells at the earliest detectable point of reporter gene expression following serum stimulation will exclude cells spuriously expressing EGFP-Luc at any later time points. 4 hours is the earliest point EGFP-Luc is detectable as determined by western blot (Fig. 18). The 4 and 6 hour selected populations were designated as MR3 and MR4 respectively. 94 EGFP-Luc positive cells were collected for MR3 from approximately 5.7 x 10^5 cells (0.016%). Similarly, 86 EGFP-Luc positive cells were selected for MR4 from 5.4 x 10^5 cells (0.016%). It appears that there was no significant benefit or disadvantage to selecting the EGFP-Luc cells at the 4 hour time point over the 6 hour time point that returned viable targeted populations in the experiments described in Chapter 1. After FACS selection the cells were routinely cultured in conditioned media until they reached density in a 10-cm dish following step-wise expansion. The bulk population was never diluted more than 1:2 or 1:3 during expansion and subculture in order to preserve optimal density in an attempt to promote growth and longevity.

The mixed genotype bulk population needed to undergo an enrichment process to increase the percent targeted cells. Enrichment of the MR3 and MR4 bulk populations was achieved through a second round of serum stimulation followed by selection using

FACS. Both MR3 and MR4 were stimulated for 6 hours with 20% serum and then EGFP-Luc positive cells were collected as single bulk populations (Fig. 23). MR3 experienced a 1.4 fold increase in EGFP-Luc positive cells when comparing unstimulated and serum stimulated populations under the same gate (21.1% and 28.8% respectively). MR4 showed fewer cell overall responding to serum stimulation with only 12.3% EGFP-Luc positive cells. However, MR4 experienced a better enrichment with a 2 fold increase in EGFP-Luc positive cells in the serum stimulated population (12.3%) compared to unstimulated control (6.2%).

The MR3 and MR4 pre-enriched and enriched populations were both subjected to PCR and sequence genotype analysis as described in Chapter 1 (Fig. 24 and data not shown). Although PCR is not quantitative, approximate trends in targeting can be inferred from the genotypes seen in Figure 24 since the PCR reactions were run simultaneously using identical template amounts. As expected we observed an increase in the *c-Myc* targeted allele in the enriched MR4 population compared to the pre-enriched MR4 population. Interestingly, the MR3 pre-enriched population showed a stronger representation of the *c-Myc* targeted allele compared to the MR4 pre-enriched population but during the enrichment process MR3 experienced a decrease in targeted alleles. These results suggest that MR3 experienced a random integration event that placed the EGFP-Luc reporter gene under the control of a promoter that is either constitutively active or serum responsive. It is of note that the promoter controlling EGFP-Luc in the MR3 cells is likely stronger than *c-Myc* since the MR3 positive cells have a greater shift in green fluorescence (Fig. 23). Targeted cells become heterozygous for *c-Myc* and subsequently experience a slowed growth. This results in a growth advantage of randomly targeted

cells homozygous for *c-Myc* over targeted cells, which explains why MR3 lost representation of the targeted allele during enrichment. Sequence analysis confirmed that the pre-enriched and enriched populations of MR3 and MR4 all contained an inframe insertion of the EGFP-Luc reporter gene into the *c-Myc* locus as expected.

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

To test whether the MR4 population was sufficiently enriched to be a useful reagent in *c-Myc* transcriptional studies we assessed a time course response to serum stimulation conditions as described in Chapter 1. Although the majority of the MR3 population is likely comprised of randomly targeted cells they will be included in these experiments for a comparison of how randomly targeted cells respond to serum stimulation conditions. Quiescent cells were stimulated with 20% serum and then harvested at 0, 2, 3, 4, 6, 8, 16, and 24 hours post stimulation for a side-by-side comparison of reporter gene expression at the level of RNA, protein, Luciferase activity, and EGFP fluorescence as described in Chapter 1.

The serum stimulation profile for MR4 recapitulates the results from the MR1 and MR2 serum stimulation studies (Fig. 17-20). Enriched MR4 cells experience a robust induction of *c-Myc* and *EGFP-Luc* transcripts in serum stimulated cells followed by a gradual decay of both transcripts over a 24 hour period (Fig. 25). MR4 also experiences a strong induction of c-Myc and EGFP-Luc proteins (Fig. 26) showing a similar delay in EGFP-Luc expression as observed in the MR1 and MR2 cells (Fig. 18). Again, we believe that that this delay is likely due to differences in processing of the exogenous transcripts and proteins relative to c-Myc and/or differences in antibody detection limits in western blot. As observed for MR1 and MR2, the reporter gene accumulates after

serum stimulation due to the long half-life of the EGFP-Luc protein (Fig. 10). The accumulation of reporter genes was confirmed by the 3.3 fold increase in Luciferase activity (Fig. 27) and the 16.4 fold increase in EGFP fluorescence (Fig. 28) that occurred in serum stimulated MR4 cells over a 24 hour time course. The accumulation of reporter genes may prove to be useful when screening gene regulatory libraries since a single activating event is amplified to easily detectable levels over the convenient 16-24 hour time frame.

MR3 shows unregulated expression of EGFP-Luc transcripts and protein over the 24 hour time course. Evidence to support a random integration of the EGFP-Luc reporter behind a constitutively active promoter in the MR3 population is given in Figure 25 where the EGFP-Luc reporter transcript is strongly expressed in quiescent cells (0 hr) and its expression is not enhanced by the addition of serum. Similarly, the MR3 cells show a lack of response to serum stimulation conditions through steady expression of EGFP-Luc protein over the 24 hour period as seen by western blot, Luciferase activity, and EGFP fluorescence (Fig. 26-28). Also, it is of note that the MR3 Luciferase activity hovers around the 0 hr background activity level of MR4 while the EGFP activity for MR3 is on average 38 fold higher than the 0 hr background activity level of MR4. The source of this discrepancy is not clear but it may be due to the differences in the detection of a Luciferase enzyme driven light reaction compared to the fluorescence of EGFP that appears to amplify the signal.

Growth rates for the bulk *c-Myc* reporter strain

MR4 performed as expected in the serum stimulation studies and is expected to be a useful reagent for studying *c-Myc* transcriptional activation. Growth rates of the MR4

population were examined to determine whether these cells were experiencing a premature senescence similar to that of the MR1 and MR2 clonal populations. The growth rates of the MR4 strain at MPD 41 is roughly equivalent to that of a similarly aged wildtype HFF population (Table 3, Fig. 29). Although growth rate data isn't available for any MPDs past 41 it is expected that the MR4 cells will continue to mirror the growth rates observed for wildtype HFFs as these cells appear to be experiencing healthy growth and expansion. There are some large variations seen in growth rates over the MPD ranges shown which may be due to differences in how the targeted cells were passaged compared to the wildtype population again since these data were not collected from a precise side-by-side ageing study. Also, variations in the growth rates are likely due to experimental manipulations such as the large growth rates seen at MPD 30, which are associated with the enrichment FACS selection of MR3 and MR4. Several vials of MR3 and MR4 cells at MPD 38 have been cryo-preserved for use in future *c-Myc* transcriptional studies. These cell strains appear to have not undergone any premature ageing and can be expanded as a useful reagents.

Chapter 3

Abstract

Creation of the *c-Myc* EGFP-Luc knockin reporter lines in primary human cells provides a platform to study endogenous human *c-Myc* gene regulation with the ease of a reporter system. However, since all primary HFFs have a finite lifespan in culture it is desirable to create a *c-Myc* reporter cell line using immortal cells. Thus we aim to create a *c-Myc* EGFP-Luc knockin cell line from U2-OS human osteosarcoma cells. The U2-OS cell line has been chosen since it has retained a normal *c-Myc* response to serum stimulation while possessing an unlimited expansion potential. U2-OS cells do harbor chromosomal aberrations but since they have maintained a normal response to serum stimulation it is thought that most of the *c-Myc* regulatory elements are intact and useful for elucidating *c-Myc* transcriptional regulatory pathways. The use of an immortal cell line permits multiple selection/enrichment schemes that were not possible with the age limited primary strains. Many purification schemes were employed to establish a U2-OS *c-Myc* reporter cell line however none of the isolated populations have acquired the targeted *c-Myc* allele. It is unclear why targeting the *c-Myc* allele has remained elusive in U2-OS cells when targeting *c-Myc* in primary HFFs was accomplished with relative ease. Possible blocks to the targeted insertion of the EGFP-Luc reporter gene in U2-OS cells may be due to the chromosomal instability inherent to tumor cell lines. Since rAAV preferentially integrates at random double strand break sites, genomic instability can result in a prevalence of random insertion events detrimentally diluting any targeted reporter gene insertion events. Another possibility is that the high level of auto-

fluorescence the U2-OS cells display interferes with the FACS selection methods described herein.

Introduction

Normal diploid primary human cells with their intact chromosomes offer the most natural system to study endogenous human gene regulation. The major limitation in using primary cells is their finite ability to be expanded in culture. At the outset of this project our goal was to develop reporter cell lines to be used in high-throughput studies of gene regulation. Although the primary MR4 cell strain developed in Chapter 2 has retained much of its potential lifespan, it will eventually undergo replicative senescence just like all primary cell strains. Therefore it is desirable to create a *c-Myc* EGFP-Luc knockin in immortal cells. The limitless expansion possibilities of immortal cells are a useful feature to incorporate into a high-throughput approach where huge numbers of regulatory molecules will be screened requiring a similarly huge number of cells. Although the immortal cells are more practical for high-throughput studies, primary cells still offer the best insight into normal human gene regulation. Therefore the combined use of both immortal and primary cell lines, that have equivalent expression of a reporter gene, will be valuable in high-throughput screening (immortal cells) and confirmation of the screen results (primary cells).

The gene targeting selection methods delineated in the preceding chapters have an absolute requirement for defined cellular conditions that turn-off and turn-on the targeted gene of interest. Serum stimulation conditions provide an ideal setting for targeting *c-Myc* since *c-Myc* is largely silenced in quiescent cells and has a strong induction following serum stimulation. Any candidate cell lines used to create a *c-Myc* reporter

line using our method of gene targeting selection must recapitulate these defined *c-Myc* expression conditions. U2-OS cells are a widely used immortal cell line originally isolated from a moderately differentiated osteosarcoma of the tibia (Ponten 1967). When taken through the serum stimulation protocol the U2-OS cells (Fig. 30) replicate the expected *c-Myc* expression kinetics observed in primary cells (Fig. 4) including a strong induction of RNA and protein 2 hours after serum stimulation followed by a decrease in both molecules over a 24 hour period. It is presumed that the U2-OS cells have likewise maintained many of the cis regulatory elements and trans acting signaling pathways that control *c-Myc* gene expression leaving them as viable prototypes for studying *c-Myc* gene regulation. Thus the immortal U2-OS cell line is a good candidate for creating our *c-Myc* EGFP-Luc reporter line.

Using immortal cells to derive our *c-Myc* reporter line affords new possibilities for the establishment and enrichment of the line. The most obvious benefit is the ability to clone individual targeted cells to produce homogeneous populations that aren't afflicted with a shortened lifespan. Immortal cells can also be subjected to serial enrichment sorts without the worry of returning a more aged population. Serial enrichments can include both positive and negative selections via FACS sorting. Specifically, quiescent populations can be sorted prior to a serum stimulation event and the EGFP-Luc *negative* cells collected, ridding the population of cells expressing reporter genes under nonspecific conditions. The negative population can then be followed up with a serum stimulation enrichment to harvest those cells which express EGFP-Luc under those conditions that *c-Myc* is expressed.

Materials and Methods

Cell Culture

U2-OS cells were obtained from ATCC (HTB-96) and 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). After FACS selection the targeted U2-OS cells were grown in conditioned media until they reached confluence in a 10 cm dish. To maintain maximal cell density the FACS selected cells were expanded in incremental tissue culture plates from a 96 well plate up to a 10-cm dish. All cultures were maintained in incubators at 37°C and 5% CO_2.

Gene targeting in U2-OS cells

U2-OS cells recently acquired from ATCC were allowed to reach to 30% confluence (~$2x10^6$ cells) in 10 cm plates. U2-OS cells were then infected overnight using 33.3 μL *c-Myc* rAAV vector stock described in Chapter 2 at a MOI of ~7,000 particles/cell in 5 mls fresh medium. Infected cells were given fresh media 24 hours post infection and driven into quiescence by confluence arrest and serum deprivation as described in Chapter 1.

To select for reporter gene insertion events into the *c-Myc* locus, 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 ~$7x10^6$ cells/mL, and then sorted on a FACS Vantage with DiVa (Digital Vantage) upgrade (Becton Dickinson). The FACS Vantage was programmed to deliver 96 single EGFP-Luc positive cells to individual wells of a 96 well tissue culture plate and then the remaining

suspension was bulk sorted into a single well of a 6 well tissue culture plate. Clonally and bulk sorted populations were then expanded for genotype analysis.

To improve the purity of selected populations, quiescent cells were subjected to a pre-sort to remove cells improperly expressing reporter genes. The quiescent population was harvested and sorted as described above. The FACS Vantage was programmed to collect the "bulk negative" population from those cells that did not display any EGFP-Luc expression. The collected bulk negative population was expanded for subsequent enrichments.

Results and Discussion

To introduce the EGFP-Luc reporter gene into the *c-Myc* locus of U2-OS cells, we infected 2 x 10^6 cells with 33.3 μL (MOI ~7000 particles/cell) of the high titer vector stock described in Chapter 2. The infected cells were driven into quiescence through density arrest and serum withdrawal followed by serum stimulation and selection via FACS sorting. The serum stimulated population yielded 0.022% EGFP-Luc positive cells (Fig. 31B). The return of EGFP-Luc positive cells from this sort is encouraging with yields 1.38 to 2.2 fold better than the yields returned from the primary HFF selection sorts (0.010%, 0.016%). 96 of the infected U2-OS cells were sorted into individual wells of a 96 well tissue culture plate and clonally expanded. The remaining EGFP-Luc positive cells from the sort were collected together as a bulk "pre-enriched" population (Fig. 31A).

The pre-enriched, mixed genotype, population requires an enrichment step to increase the amount of cells in the population that express EGFP-Luc upon serum stimulation and thus likely under the control of the *c-Myc* promoter. The pre-enriched

population was expanded and brought again to quiescence. Cells were then serum stimulated and the EGFP-Luc positive cells were collected. Serum stimulation of the pre-enriched population resulted in 2.9% of the cell expressing EGFP-Luc (Fig. 32 right). There does not appear to be a significant enrichment of serum stimulation dependent EGFP-Luc expressing cells in this experiment since there is not a significant increase in the percent EGFP-Luc in the serum stimulated sample compared to the unstimulated quiescent population (2.4%) (Fig. 32 left).

Since there appears to be a high background level of EGFP-Luc expression in the pre-enriched population we collected a "bulk-negative" population (Fig. 31A) to rid the population of cells contributing to background EGFP-Luc expression. The bulk-negative population was established from those cells in the quiescent population that did not have any shift in green fluorescence (Fig. 32 left in purple). The bulk negative cells were expanded and brought to quiescence for subsequent enrichment via serum stimulation (Fig. 33). The enrichment of the bulk negative population does not appear to have significantly enriched the population for serum responsive EGFP-Luc reporter gene expression when comparing serum stimulated bulk negative populations (0.87%) to the background expression of EGFP-Luc (0.85%) seen in the unstimulated bulk negative population (Fig. 33A). Interestingly, there appears to be a significant shift in green fluorescence for a large portion of the bulk negative population (Fig. 33B) after serum stimulation which is not reflected in the histograms shown in Figure 33A. The discrepancy in green fluorescence between the two types of histograms shown in Figure 33 is not clear but may be due to instrumental compensation for autofluorescence. When the background level of U2-OS autofluorescent cells, contained in the M1 region of the

U2-OS untreated histogram (18.77%), is subtracted from the bulk negative enrichment samples there remains an ~ 8 fold increase of green fluorescing cells in the serum stimulated bulk negatives compared to the unstimulated bulk negatives. Unfortunately the strong influence of autofluorescent cells (18.77% of untreated U2-OS) in that region of green fluorescence intensity precludes the selection of those cells as a pure population. However, considering the poor results obtained using the FL1-A versus FL2-A histograms (Fig. 32, 33A) for selection of the U2-OS cells it may be desirable to select EGFP-Luc expressing cells using the FL1-A versus cell number histogram (Fig. 33B) to select for EGFP-Luc positive cells even with an inherent error of 18%.

The various enrichment schemes do not appear to have worked as seen by the lack in any significant increase in EGFP-Luc expression above background levels for the various U2-OS populations. However, this conclusion needs to be substantiated by genotyping the collected populations. Genomic DNA was collected for the pre-enriched, enriched stimulated, enriched unstimulated, and bulk negative populations. The genotype of the populations was then determined by PCR (Fig. 9, 34). None of the collected populations display the *c-Myc* targeted allele, which confirming the dismal FACS sort results (gels on left). To verify that these populations were expressing EGFP-Luc, a PCR genotype, which amplified the EGFP portion of the construct, was run. Consistent with the region of the FACS histograms the populations were selected from, the pre-enriched, enriched stimulated and enriched unstimulated populations all are positive for EGFP while the bulk negative population does not have any detectable amounts of EGFP (gels on right). These genotype results suggest that the original population was comprised

mainly of randomly integrated cells, which express low but selectable levels of EGFP-Luc independent of the media serum levels.

The 96 cells collected from the FACS selection were expanded as clonal populations. Of the 96 cells collected only 31 grew out as viable colonies. These 31 clones were tested for the presence of the *c-Myc* targeted allele but none of the clones displayed the targeted reporter gene (Fig. 35). To detect the presence of randomly integrated reporter genes the PCR for the internal EGFP was also conducted. Of the 31 clones tested only 2 clones tested positive for EGFP (data not shown). This low representation of even random integration events suggests that the FACS Vantage was not programmed with enough stringency. Care could be taken upon repeat of these experiments to insure that only very green cells are selected for during the clonal sort. However, these results in combination with the bulk population results suggest that the stringency of the sort was only part of the problem.

Since none of the populations from the preceding experiments were targeted it is necessary to try to re-derive the U2-OS *c-Myc* reporter line. U2-OS cells were prepared for selection as described above. Quiescent populations were serum stimulated and then EGFP-Luc expressing cells were selected using FACS sorting (Fig. 36). The serum stimulated population returned 0.03% EGFP-Luc positive cells, the highest yield we have observed to date. The percent EGFP-Luc positive cells was the same for the serum stimulated and unstimulated samples, however since an unstimulated sample was not included with any of the preceding experiments it is not clear whether this indicates a failed experiment. The EGFP-Luc positive cells were collected for both the unstimulated

and serum stimulated populations are currently being expanded in culture and require subsequent enrichments.

The bulk negative cells were collected from the unstimulated population (Fig. 36 left purple). These cells were expanded and followed-up with an enrichment FACS sort. The enrichment of the bulk negative population returned 0.007% and 0.006% EGFP-Luc positive cells for the unstimulated and serum stimulated samples, respectively. Again, no significance difference between unstimulated and serum stimulated samples was observed for this enrichment scheme but since the unstimulated control was not included in the primary HFF experiments it is not clear whether this result indicates the absence of the *c-Myc* targeted allele. The EGFP-Luc positive cells are currently being expanded and should be genotyped to determine whether any *c-Myc* targeted alleles are present.

It is unclear why the U2-OS cells have failed to return a *c-Myc* EGFP-Luc targeted population. Percent EGFP-Luc positive cells returned from the selection sorts (Fig. 31 & 36) following infection were better than those obtained for the primary HFFs (Fig. 11 & 22) but the U2-OS EGFP-Luc positive cells are not responsive to serum stimulation (Fig. 32 & 33) and do not possess detectable levels of the *c-Myc* targeted allele (Fig. 34 & 35). The difficulty in targeting the population may be due to cell specific biology that is preventing the targeting event from occurring or being selected.

Randomly integrated rAAV occurs at a rate of around 10% while targeted integration occurs at rates of up to 1% (Vasileva 2005, 2006). Integration of rAAV is known to occur preferentially at double strand break sites, whether homologous or random, through the homologous recombination or non-homologous end joining branches of the DNA double strand break repair pathway (Vasileva 2006). U2-OS cells

have a disrupted genome with chromosome counts in the hypertriploid range and very few normal chromosomes remaining (ATCC). The apparent instability of the U2-OS genome may result in an abundance of double strand breaks, leaving these cells more prone to random insertions. U2-OS cells may also compensate for genomic instability with hyperactive DNA repair systems such as non-homologous end joining that could similarly result in a prevalence of random integration events. However, DNA repair studies in U2-OS cells employ DNA damaging methods such as ionizing radiation to cause double strand breaks and it is not clear whether the background level of DNA damage is higher in the U2-OS cells compared to primary cells.

The U2-OS cells display a high level of autofluorescence compared to primary HFF cells. When the clonal selection and enrichment sorts were performed for the primary HFF (Fig. 11 & 16) and the U2-OS cells (Fig. 31 & 32) there were noticeable differences in the population wide trends in green fluorescence (Fig. 38). The successful targeting selection of the primary HFF cells occurred from populations with low background fluorescence levels. The U2-OS selections were completed over a high autofluorescence background. EGFP-Luc expression driven off the *c-Myc* promoter is weak, giving fluorescence signals in the range where autofluorescence is detected for wildtype U2-OS cells. The high level of autofluorescence displayed by the U2-OS cells may prevent them from being useful in these targeting methods. A different immortal cell line could be used but care should be taken upfront to insure that the cells do not display a high level of autofluorescence.

Conclusion and Summary

We demonstrate here a method for rapidly producing primary human gene reporter cells by rAAV-mediated gene insertion, and characterize *c-Myc* gene reporter cell strains 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.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell 4th ed. Garland Sciences 2002. Chapter 7.

ATCC (American Type Culture Collection). http://www.atcc.org/common/catalog/numSearch/numResults.cfm?atccNum=HTB-96#22509

Bunz F. Human cell knockouts. Curr Opin Oncol. 2002 Jan;14(1):73-8.

Chamberlain, J.R., Schwarze, U., Wang, P., Hirata, R.K., Hankenson, K.D., Pace, J.M., Underwood, R.A., Song, K.M., Sussman, M., Byers, P.H., et al. (2004). Gene targeting in stem cells from individuals with osteogenesis imperfecta. Science 303, 1198-1201.

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 Apr 20;84(2):169-73.

Cristofalo VJ, Pignolo RJ. Replicative senescence of human fibroblast-like cells in culture. Physiol Rev. 1993 Jul;73(3):617-38.

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 Aug;5(8):1969-76.

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 Jul 15;261(20):9161-6.

Garkavtsev I, Riabowol K. Extension of the replicative life span of human diploid fibroblasts by inhibition of the p33ING1 candidate tumor suppressor. Mol Cell Biol. 1997 Apr;17(4):2014-9.

Goldstein S. Replicative senescence: the human fibroblast comes of age. Science. 1990 Sep 7;249(4973):1129-33.

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;16:653-99.

Hann SR, Thompson CB, Eisenman RN. c-myc oncogene protein synthesis is independent of the cell cycle in human and avian cells. Nature. 1985 Mar 28-Apr 3;314(6009):366-9.

Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. Exp Cell Res. 1961 Dec;25:585-621.

Hendrie, P.C., and Russell, D.W. (2005). Gene targeting with viral vectors. Mol Ther 12, 9-17.

Hirata R, Chamberlain J, Dong R, Russell DW. Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. Nat Biotechnol. 2002 Jul;20(7):735-8.

Hirata RK, Russell DW. Design and packaging of adeno-associated virus gene targeting vectors. J Virol. 2000 May;74(10):4612-20.

Inoue N, Russell DW. Packaging cells based on inducible gene amplification for the production of adeno-associated virus vectors. J Virol. 1998 Sep;72(9):7024-31.

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 Jan 2;32(1):e3.

Liu, X., Yan, Z., Luo, M., Zak, R., Li, Z., Driskell, R.R., Huang, Y., Tran, N., and Engelhardt, J.F. (2004). Targeted correction of single-base-pair mutations with adeno-associated virus vectors under nonselective conditions. J Virol 78, 4165-4175.

Nesbit CE, Tersak JM, Prochownik EV. MYC oncogenes and human neoplastic disease. Oncogene. 1999 May 13;18(19):3004-16.

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 Nov;5(11):2903-12.

Ponten J, Saksela E. Two established in vitro cell lines from human mesenchymal tumours. Int J Cancer. 1967 Sep 15;2(5):434-47.

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 May;23(10):3558-65.

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 Aug;4(8):2009-15.

Ramsay G, Evan GI, Bishop JM. The protein encoded by the human proto-oncogene c-myc. Proc Natl Acad Sci U S A. 1984 Dec;81(24):7742-6.

Razin SV, Iarovaia OV, Sjakste N, Sjakste T, Bagdoniene L, Rynditch AV, Eivazova ER, Lipinski M, Vassetzky YS. Chromatin domains and regulation of transcription. J Mol Biol. 2007 Jun 8;369(3):597-607.

Rohme D. Evidence for a relationship between longevity of mammalian species and life spans of normal fibroblasts in vitro and erythrocytes in vivo. Proc Natl Acad Sci U S A. 1981 Aug;78(8):5009-13
Russell, D.W., and Hirata, R.K. (1998). Human gene targeting by viral vectors. Nat Genet 18, 325-330.Sedivy JM, Dutriaux A. Gene targeting and somatic cell genetics--a rebirth or a coming of age? Trends Genet. 1999A Mar;15(3):88-90.

Sedivy JM, Vogelstein B, Liber H, Hendrickson E, Rosmarin A. Gene Targeting in Human Cells Without Isogenic DNA. Science 1999B Jan 1;283:9a

Spencer CA, Groudine M. Control of c-myc regulation in normal and neoplastic cells. Adv Cancer Res. 1991;56:1-48.

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 Oct 7;33(18):e158

Trobridge G, Hirata RK, Russell DW. Gene targeting by adeno-associated virus vectors is cell-cycle dependent. Hum Gene Ther. 2005 Apr;16(4):522-6.

Vasquez KM, Marburger K, Intody Z, Wilson JH. Manipulating the mammalian genome by homologous recombination. Proc Natl Acad Sci U S A. 2001 Jul 17;98(15):8403-10.

Vasileva A, Jessberger R. Precise hit: adeno-associated virus in gene targeting. Nat Rev Microbiol. 2005 Nov;3(11):837-47.

Vasileva A, Linden RM, Jessberger R. Homologous recombination is required for AAV-mediated gene targeting. Nucleic Acids Res. 2006 Jul 5;34(11):3345-60. Print 2006.

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 Aug;6(2):272-8.

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 Jan;25(1):147-61.

Zolotukhin S, Byrne BJ, Mason E, Zolotukhin I, Potter M, Chesnut K, Summerford C, Samulski RJ, Muzyczka N. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther. 1999 Jun;6(6):973-85.

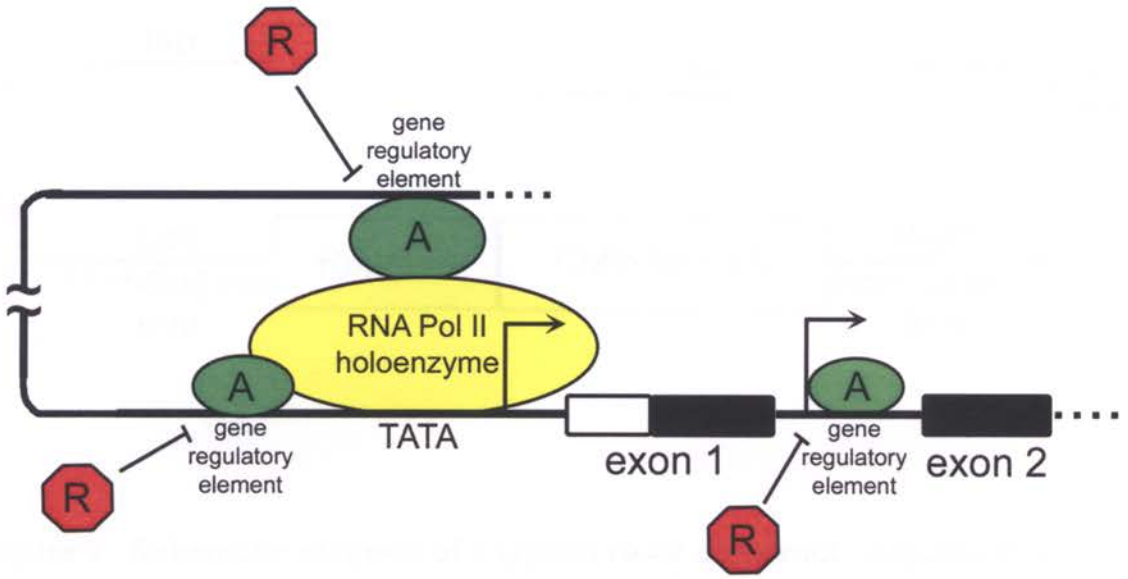


Figure 1. Diagram of eukaryotic transcriptional activation.
Gene specific activating transcription factors compete with repressive transcription factors for binding at cis acting gene regulatory elements. Proximal and distal gene regulatory elements are located upstream, downstream, and intergenically. Activating transcription factors aid in RNA Polymerase II holoenzyme binding of core promoter elements and speed the rate of transcriptional activation.

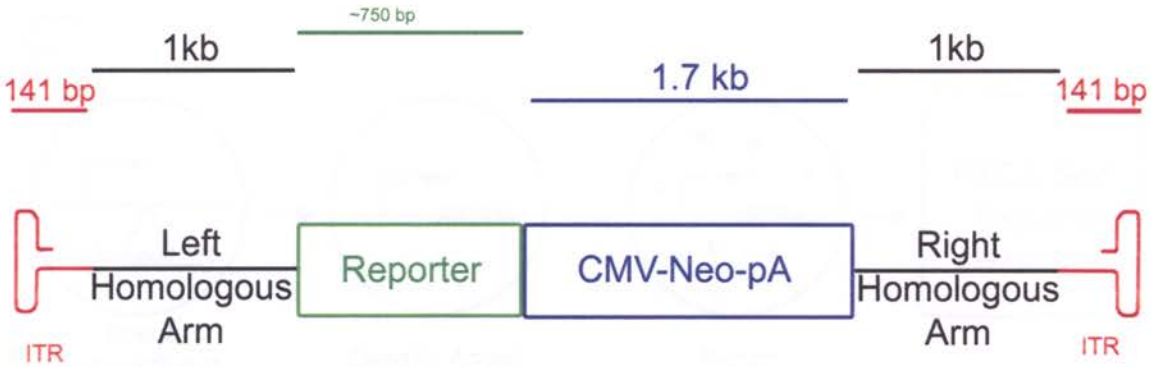


Figure 2. Schematic diagram of a typical rAAV construct. A typical rAAV construct contains a drug selection cassette to allow isolation of transduced cells. Targeted genetic changes require the presence of homologous arms that are usually around 1kb. Left and right inverted terminal repeats (ITR) are required for rAAV packaging and transduction. The inclusion of all typical rAAV vector components limits any introduced reporter gene to a size of ~750 bps.

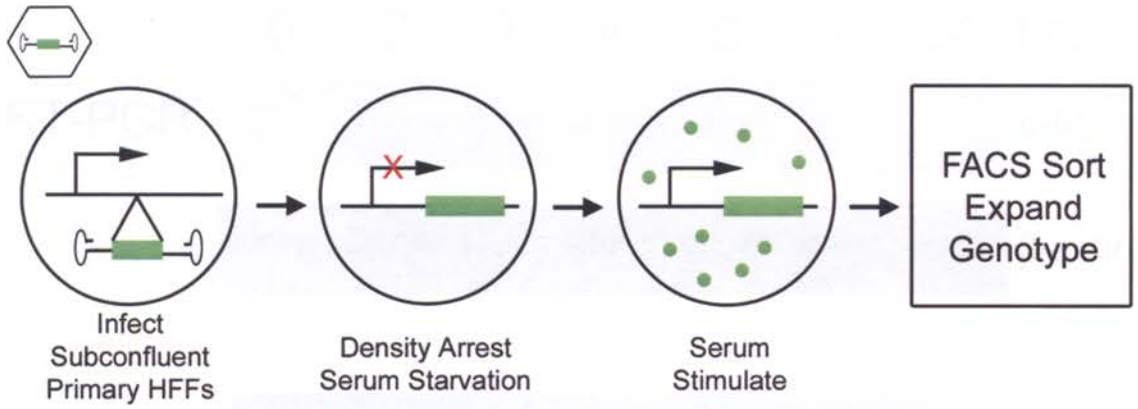


Figure 3. Direct selection of targeted insertion of the EGFP-Luc reporter gene into the c-Myc locus using known c-Myc regulatory conditions. Subconfluent primary HFF cells are infected with rAAV targeting vectors when c-Myc is being actively transcribed or replicated in proliferative cells. Background 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.

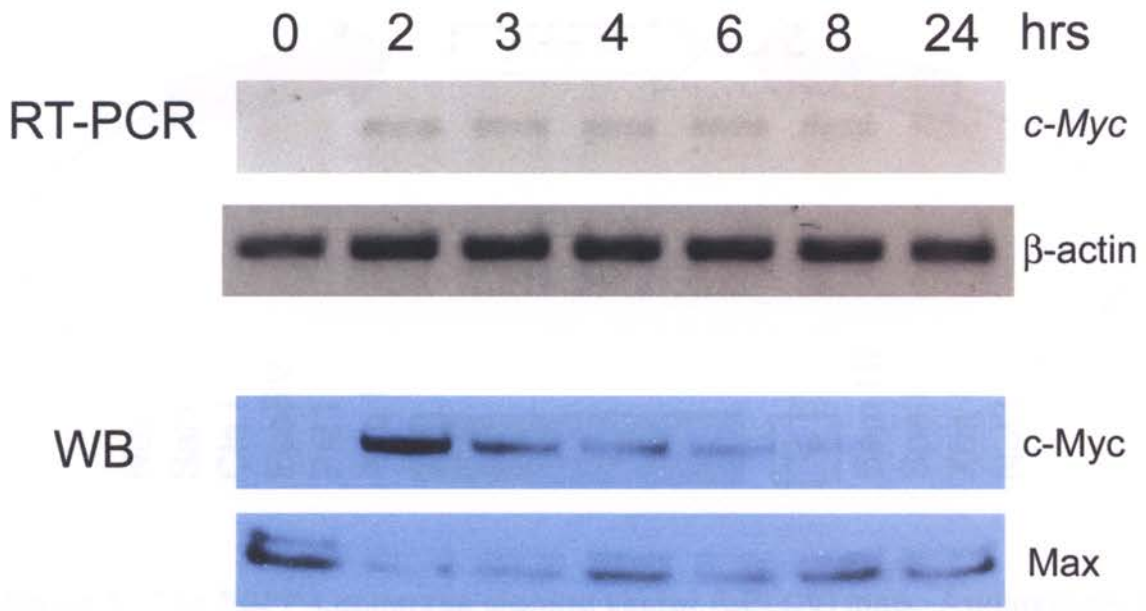


Figure 4. Induction of c-Myc transcripts and protein following serum stimulation. Cells were brought to quiescence through density arrest and serum withdrawal. c-Myc experiences a rapid induction of transcripts (top) and protein (bottom) following addition of serum which decay to pre-stimulation levels over a 24 hour period. β-actin was used as a control for RT-PCR and Max was used as a control for western blot.

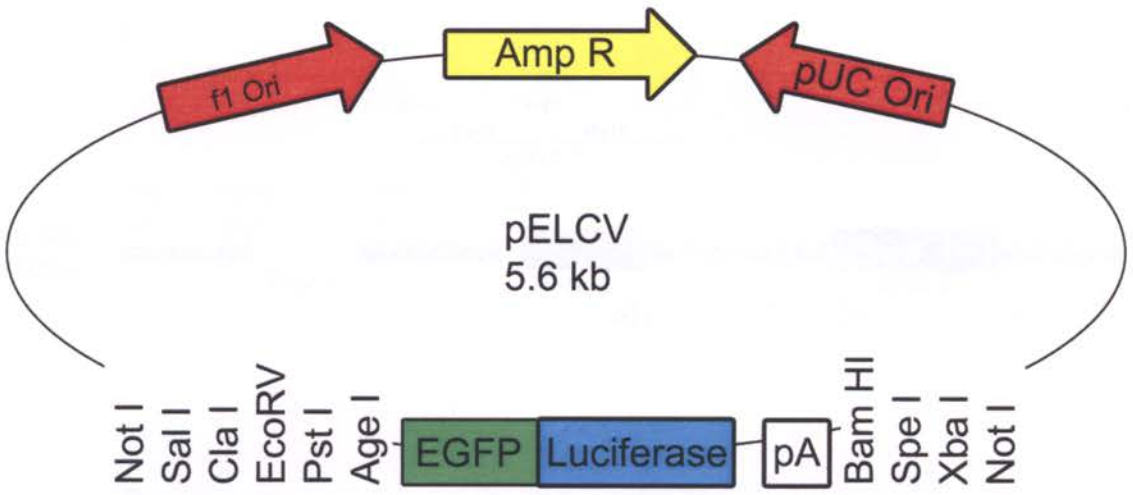


Figure 5. The EGFP-Luciferase cloning vector (pELCV) map. A promoterless EGFP-Luciferase fusion gene is surrounded by the unique restriction sites shown that allow easy introduction of homologous arms. Dual NotI sites allow the targeting construct to be subcloned into most rAAV ITR containing plasmids.

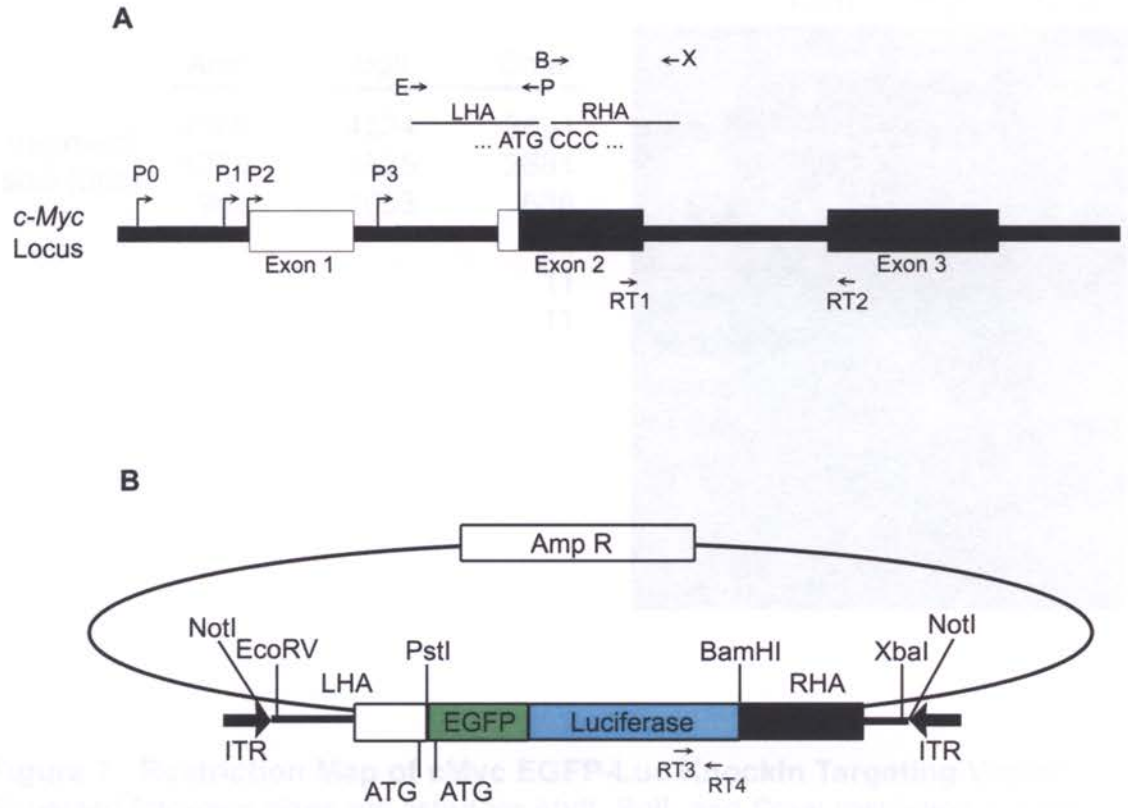


Figure 6. c-Myc EGFP-Luciferase Knockin Targeting Vector. (**A**) The position of the left homologous arm (LHA) and right homologous arm (RHA) within the c-Myc locus is shown. The LHA ends with the c-Myc start codon contained in exon 2 and the RHA begins with the first base following the ATG. (**B**) The LHA and RHA were amplified with primers containing the unique restriction sites shown and subcloned into pELCV creating a fusion with the EGFP-Luc reporter gene that is inframe with the c-Myc ATG. Locations of the RT-PCR primers sets are shown for c-Myc (**A**) and EGFP-Luc (**B**). E, EcoRV. P, PstI. B, BamHI. X, XbaI. P0, P1, P2, and P3 indicate locations of the known c-Myc promoters.

	AhdI	BglI	SmaI
fragment size (bps)	4566	4624	3891
	1773	1625	2681
	963	1053	686
			22
			11
			11

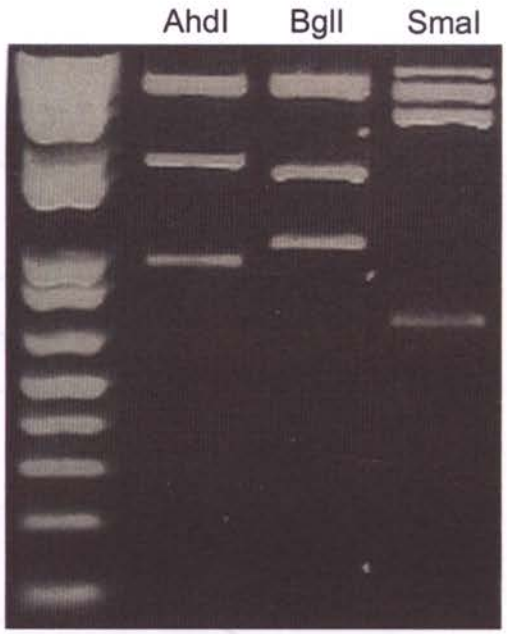


Figure 7. Restriction Map of cMyc EGFP-Luc Knockin Targeting Vector. Expected fragment sizes are listed for AhdI, BglI, and SmaI restriction digests. The restriction enzyme map of the c-Myc EGFP-Luc targeting vector corresponds exactly to the expected fragment sizes. The AhdI restriction map verifies viral ITR integriy. The BglI and SmaI maps verify insert orientation.

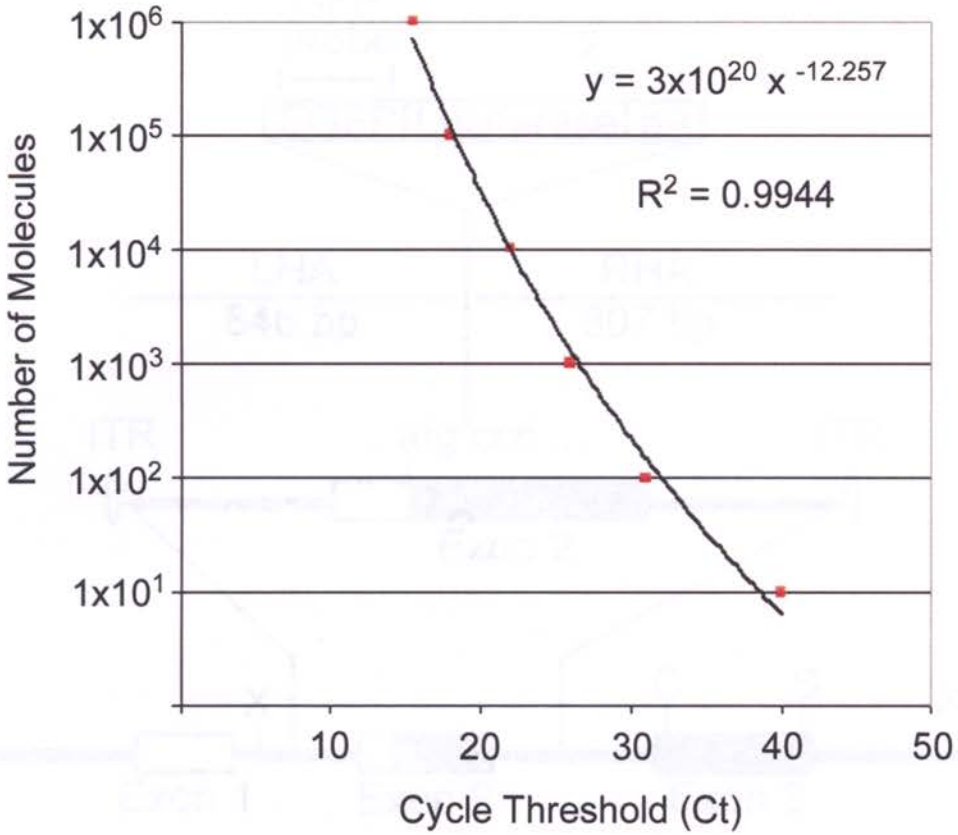


Figure 8. c-Myc EGFP-Luc Knockin Targeting vector titration. The standard curve was generated via RT-PCR using SYBR green I dye in the Applied Biosystems 7900HT sequence detection system. 3uL of c-Myc rAAV viral stock was used as template in the RT-PCR reaction and returned a Ct of 24.4 which corresponds to a vector stock concentration of $1x10^6$ particles/mL.

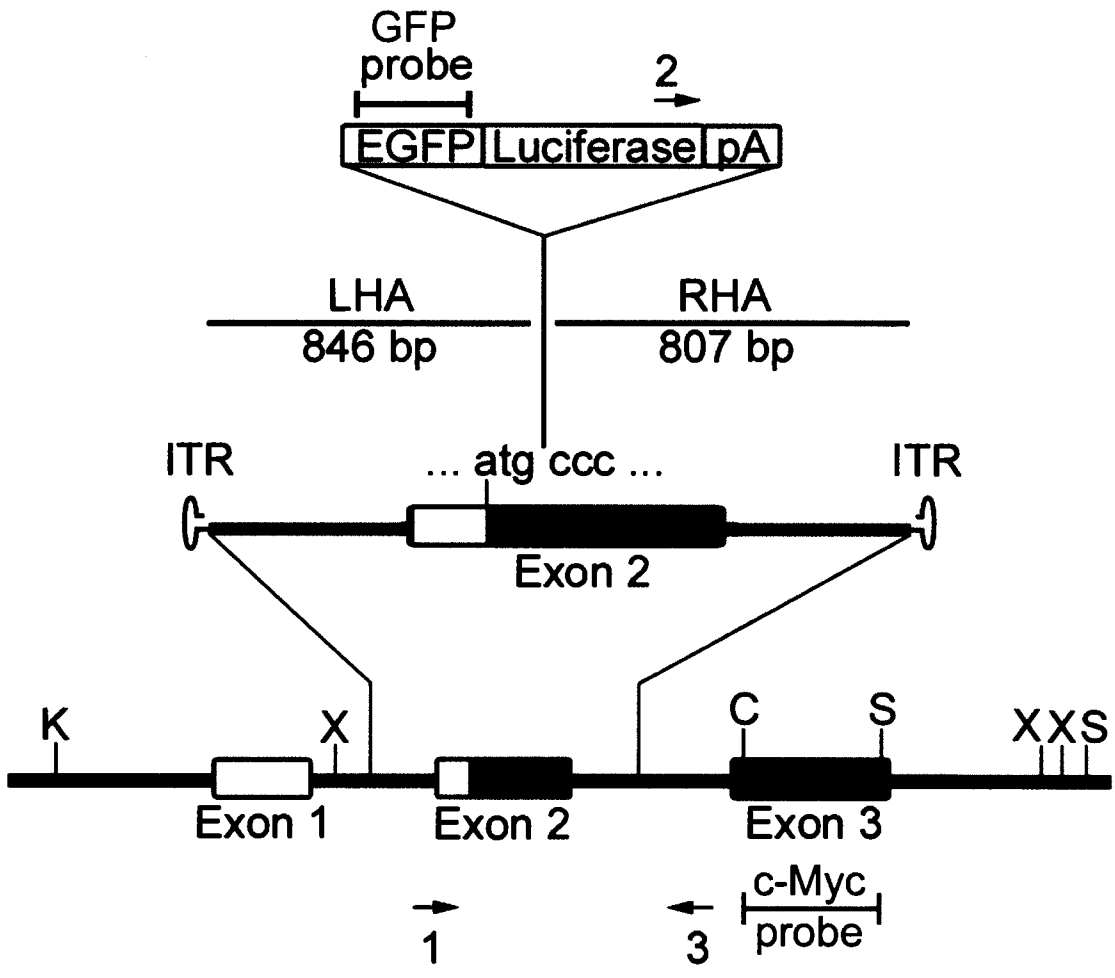
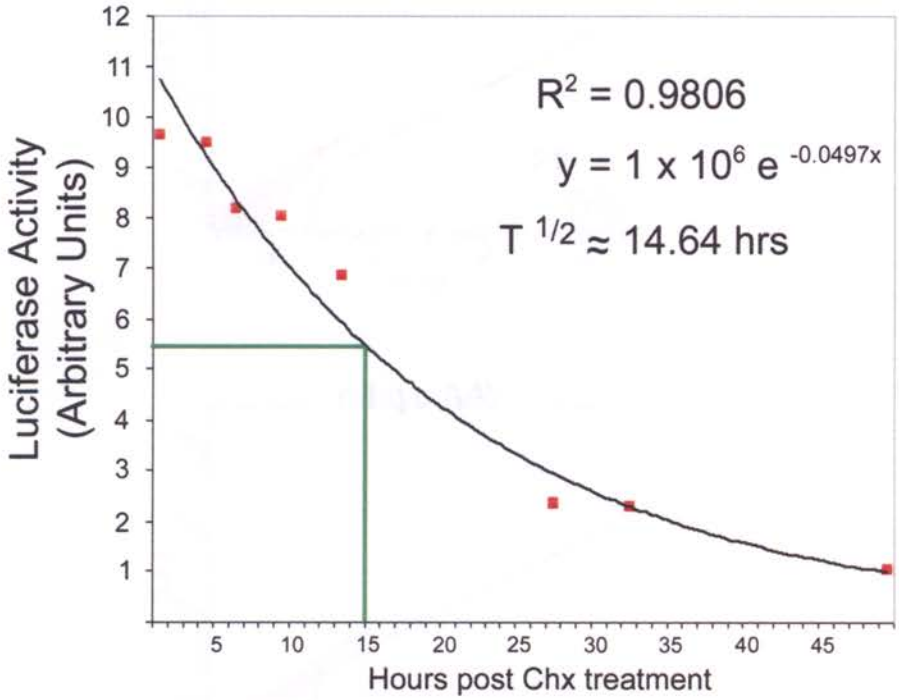


Figure 9. Diagram of the c-Myc rAAV targeting and genotype strategies. The targeting region of the endogenous c-Myc locus is shown. The LHA, RHA, and ITR portions of the rAAV vectors facilitate insertion of the EGFP-Luciferase fusion gene directly between the first and second codon of the c-Myc gene. Primer binding sites for PCR genotyping are indicated for primers 1-3. Positions of the c-Myc exon 3 and the GFP Southern probes are shown. K, KpnI. X, XbaI. C, ClaI. S, SspI.

A



B

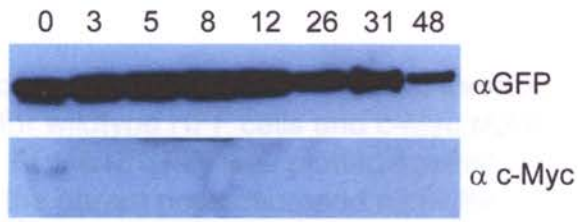


Figure 10. EGFP-Luciferase fusion protein stability. The stability of the EGFP-Luc protein was determined in a cyclohexamide (Chx) protein degradation assay following a transient transfection of pEGFPLuc into U2-OS cells. Samples were collected at the indicated times post Chx treatment and EGFP-Luc stability was assessed through Luciferase activity (A) and western blot (B). The c-Myc blot was included to show its relative stability compared to the reporter gene. The 14.6 hour halflife of EGFP-Luc was determined from the formula shown in A and corresponds to 50% of maximal Luciferase activity.

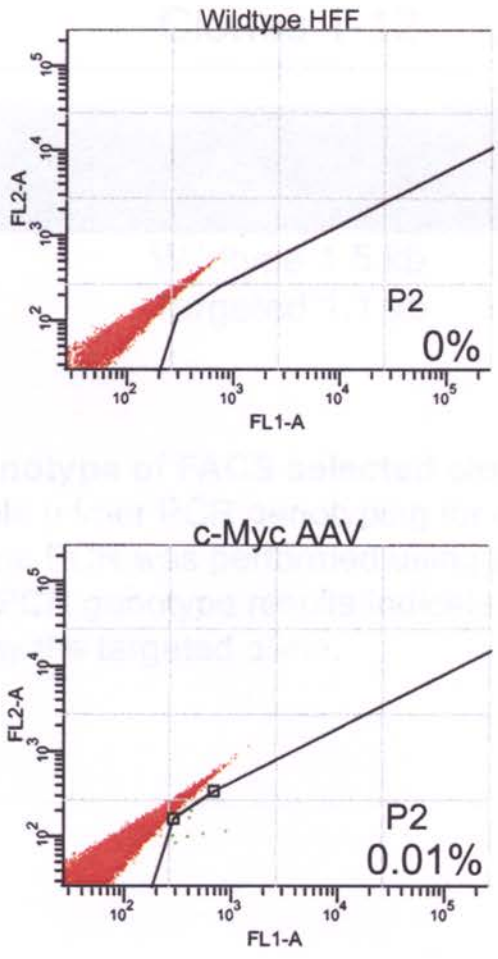


Figure 11. FACS selection of c-Myc EGFP-Luc targeting events. FACS histograms are shown for wildtype HFF cells and c-Myc rAAV targeted cells following CCE. FL1-A (EGFP) was plotted against FL2-A (no fluorofor) to center the parent population and allow for selection of the dim EGFP-Luciferase positive cells. The P2 region of the plot indicates the gating used for selection of the EGFP-Luciferase positive portion of the parent population.

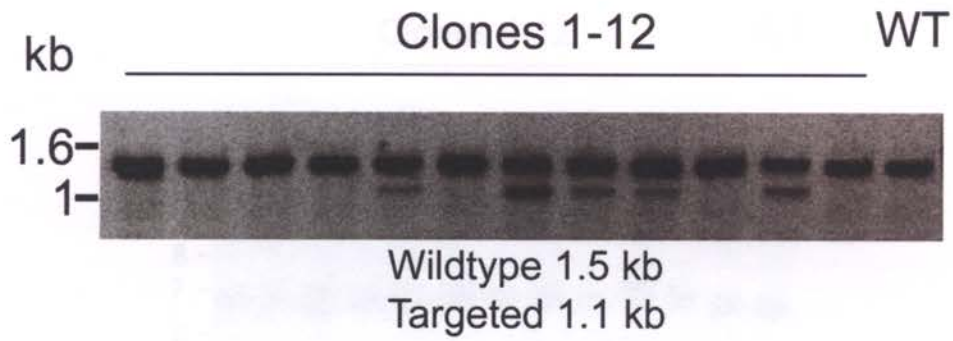


Figure 12. PCR genotype of FACS selected clones 1-12. Results from the triple primer PCR genotyping for clones 1-12 are shown. Genotype PCR was performed using primers 1-3 shown in Figure 9. PCR genotype results indicate five clones (5, 7, 8, 9, 11) display the targeted allele.

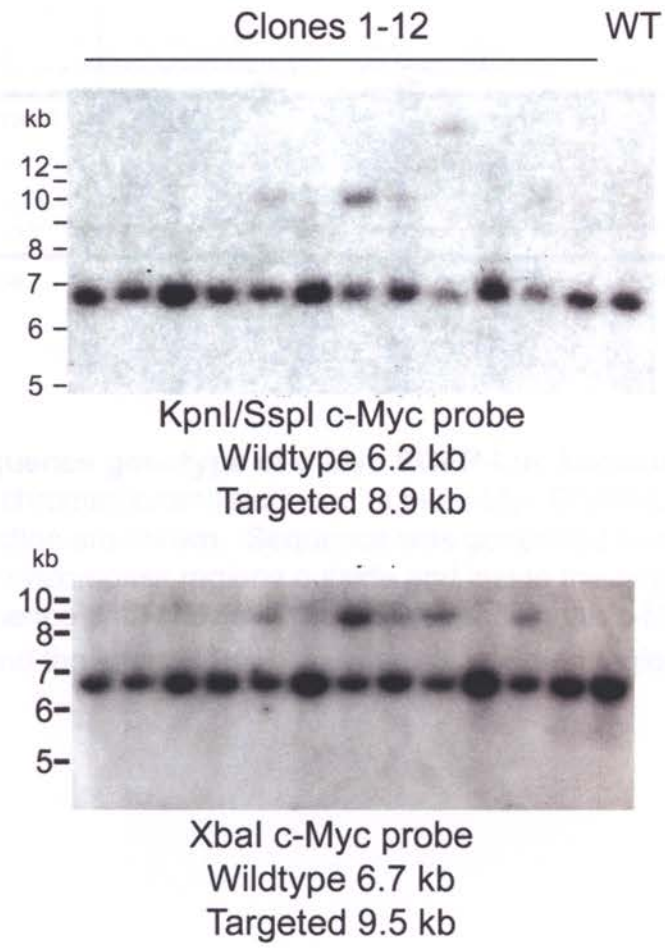


Figure 13. Southern Blot Genotype of FACS selected clones.
The Southern blots using the c-Myc exon 3 probe are shown. Southern blotting was performed on samples digested with either KpnI and SspI (top) or XbaI (bottom). The XbaI blot confirms the five positive clones from the PCR genotype. The KpnI/SspI blot verifies proper targeting in clones 5, 7, 8, and 11 but yields an unexpected banding pattern for clone 9.

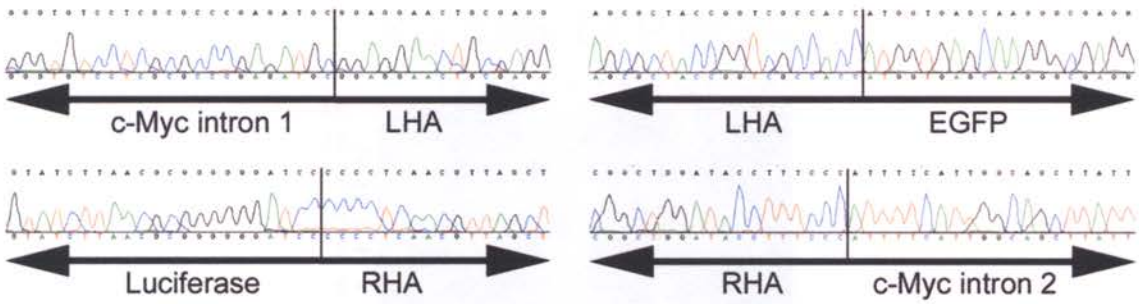


Figure 14. Sequence genotype of c-Myc EGFP-Luc knockin clones. Representative chromatograms from one of the c-Myc EGFP-Luc knockin allele sequencing reaction are shown. Sequence was generated from PCR products, which encompass regions outside and inside the targeting construct. Sequence data shows the regions spanning the c-Myc rAAV insertion sites and the internal EGFP-Luciferase insertion regions.

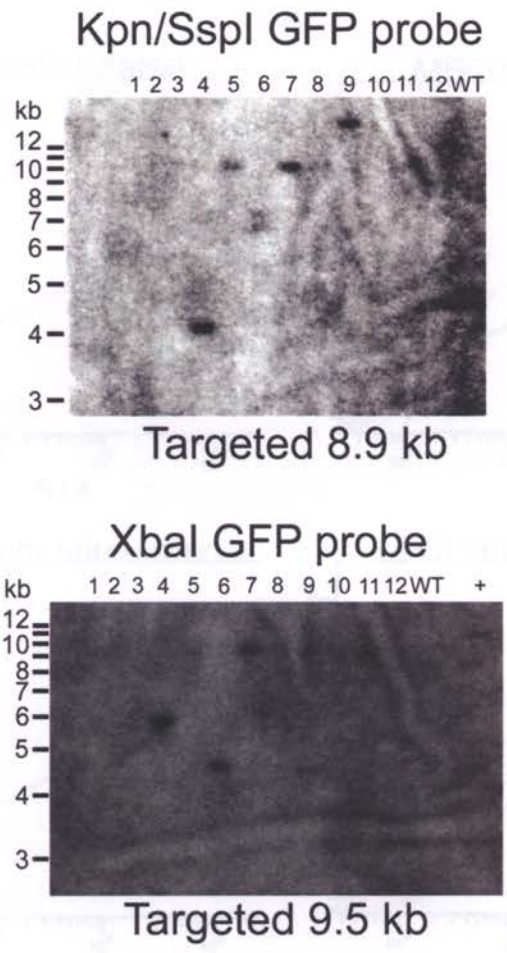


Figure 15. Southern blot genotype of random integration events. Southern blot for FACS selected clones 1-12 are shown. Genomic DNA was either digested with KpnI and SspI or XbaI and then blotted with a probe specific to the GFP region of the reporter gene. Clones 4, 6, and 9 show bands that are not consistent with a targeted integration event. 5, 7, 8, and 11 all display a single band corresponding to a targeted integration of the reporter gene. Clone 9 displays improper bands for both the KpnI/SspI and XbaI blots suggesting it underwent a random integration event in addition to a targeted integration event. +, positive control is 25 copies of the 7.3 kb c-Myc targeting construct.

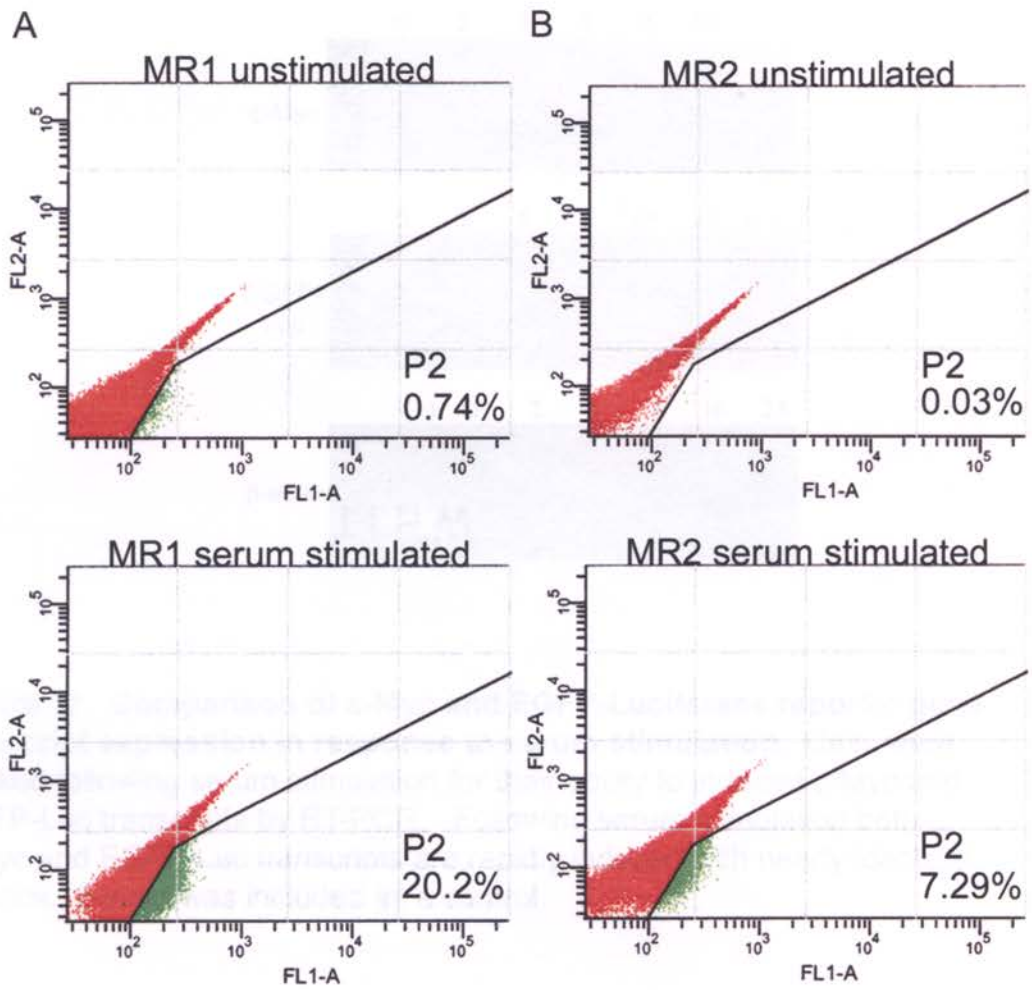


Figure 16. MR1 and MR2 response to serum stimulation.
MR1 (A) and MR2 (B) clonal populations were subjected to serum stimulation conditions and then evaluated for EGFP-Luc reporter gene expression. The unstimulated background control populations are shown in the top panels. Serum stimulated populations are shown in the bottom panels where the EGFP-Luciferase expressing cells are located in the P2 gated region. 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.

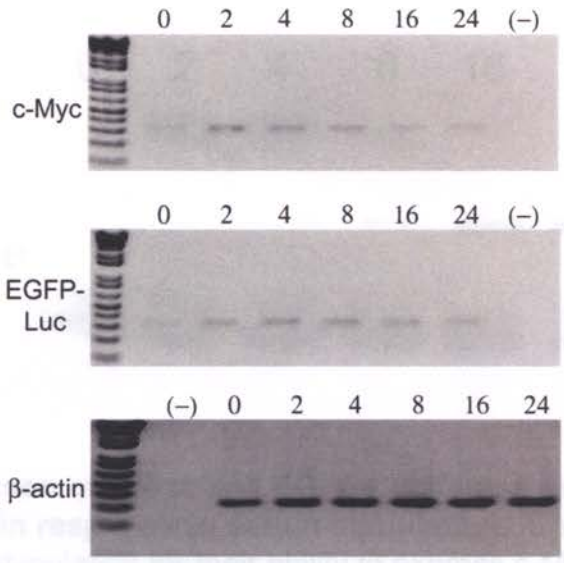


Figure 17. Comparison of c-Myc and EGFP-Luciferase reporter gene transcript expression in response to serum stimulation. Cells were tracked following serum stimulation for their ability to induced c-Myc and EGFP-Luc transcripts by RT-PCR. Following serum stimulation both c-Myc and EGFP-Luc transcripts are rapidly induced with nearly identical kinetics. β-actin was included as a control.

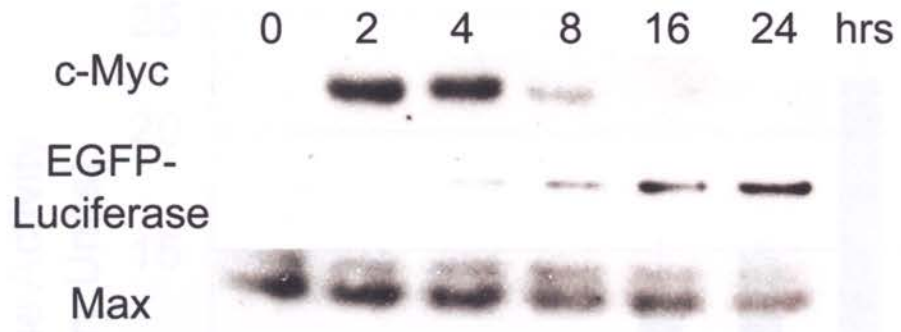


Figure 18. Comparison of c-Myc and EGFP-Luciferase reporter gene protein expression in response to serum stimulation. Cells were tracked after serum stimulation for their ability to express c-Myc and EGFP-Luc proteins. The c-Myc protein is induced following expected serum stimulation kinetics. EGFP-Luc is induced with delayed kinetics compared to c-Myc and then accumulates over the 24 hour period. Max is used as a loading control.

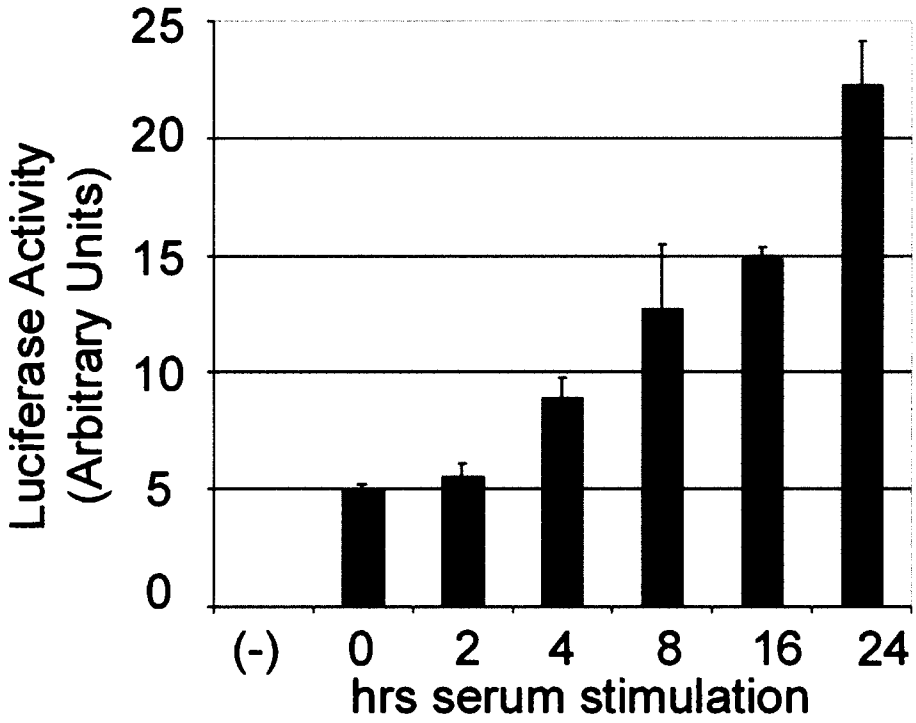


Figure 19. Luciferase reporter gene activity kinetics in response to serum stimulation. Luciferase activity was assessed following serum stimulation. Luciferase activity progressively increases over the 24 hour period with an overall 4.5 fold increase in activity. Data shown is the average of triplicate samples +/- SEM.

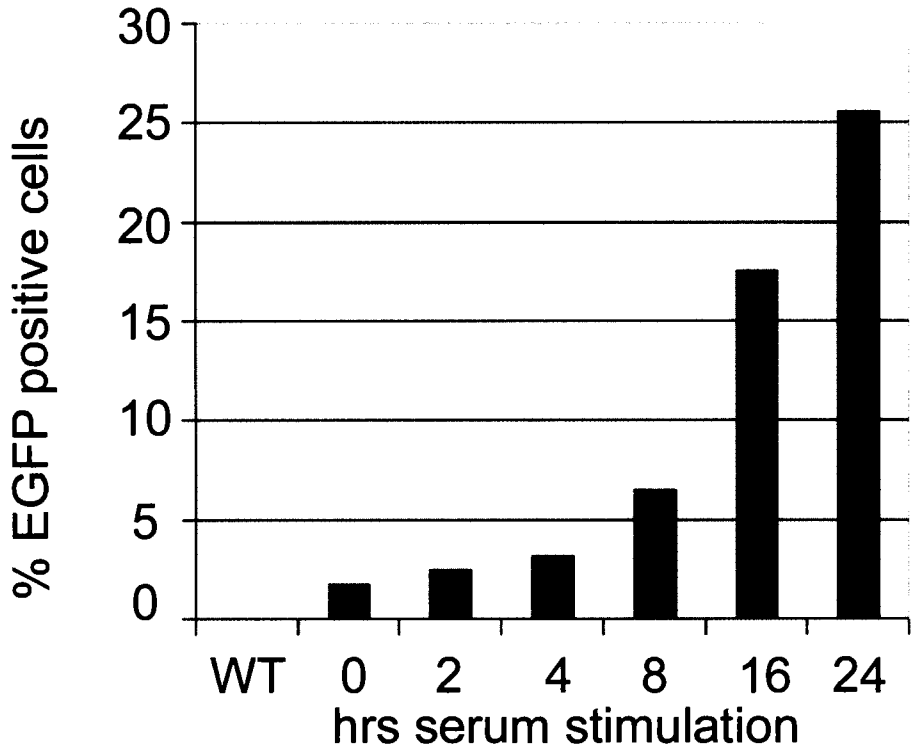


Figure 20. EGFP reporter gene activity kinetics in response to serum stimulation. EGFP fluorescence activity was assessed following serum stimulation. FACS Calibur data is shown where the EGFP-Luciferase positive percent of the parent population was determined using BD CellQuest Pro v. 5.2. Time course changes of the percent EGFP-Luciferase expressing populations were compared under the same gates.

	WT HFF	MR1	MR2
Avg. Cell Count	5.1×10^6	1.8×10^6	1.2×10^6

Table 1. Comparison of total cell counts of MR1 and MR2 at confluence. Cell counts are listed for wildtype HFFs, MR1, and MR2 are the average total cells present at density in a 10-cm dish for MPDs 41-54.

MPD	WT HFF	MR1	MR2
30-31	0.80	0.65	0.65
41	0.57	0.33	0.29
45-46	0.45	0.40	0.13
48	0.47	0.14	0.14
54	1.08	0.10	n/a
64	0.33	n/a	n/a
75	0.13	n/a	n/a

Table 2. Growth Rate Comparison of MR1 and MR2 clonal populations. Growth rates are listed as MPD/days in culture. MPDs were calculated according to the formula: $2^n = P - P_O$ where n = the MPD, P = total cell count at density, and P_O = number of cells seeded.

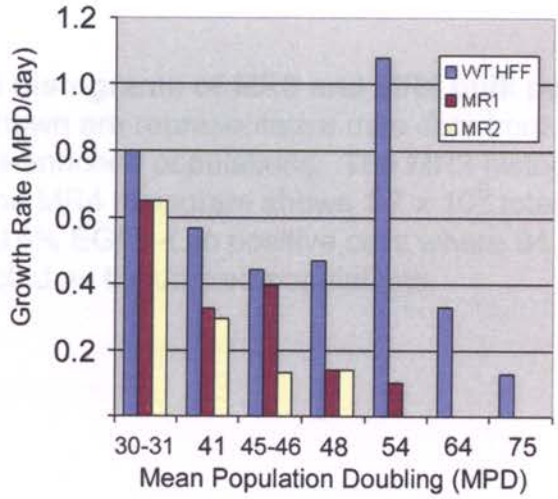


Figure 21. Growth Rate Comparison of MR1 and MR2 clonal populations. Graphical representation of the growth rates listed in Table 2. The MR1 and MR2 clonal populations at MPD 45-48 show growth rates comparable to the WT HFF non-clonal population near senescence at MPD 75.

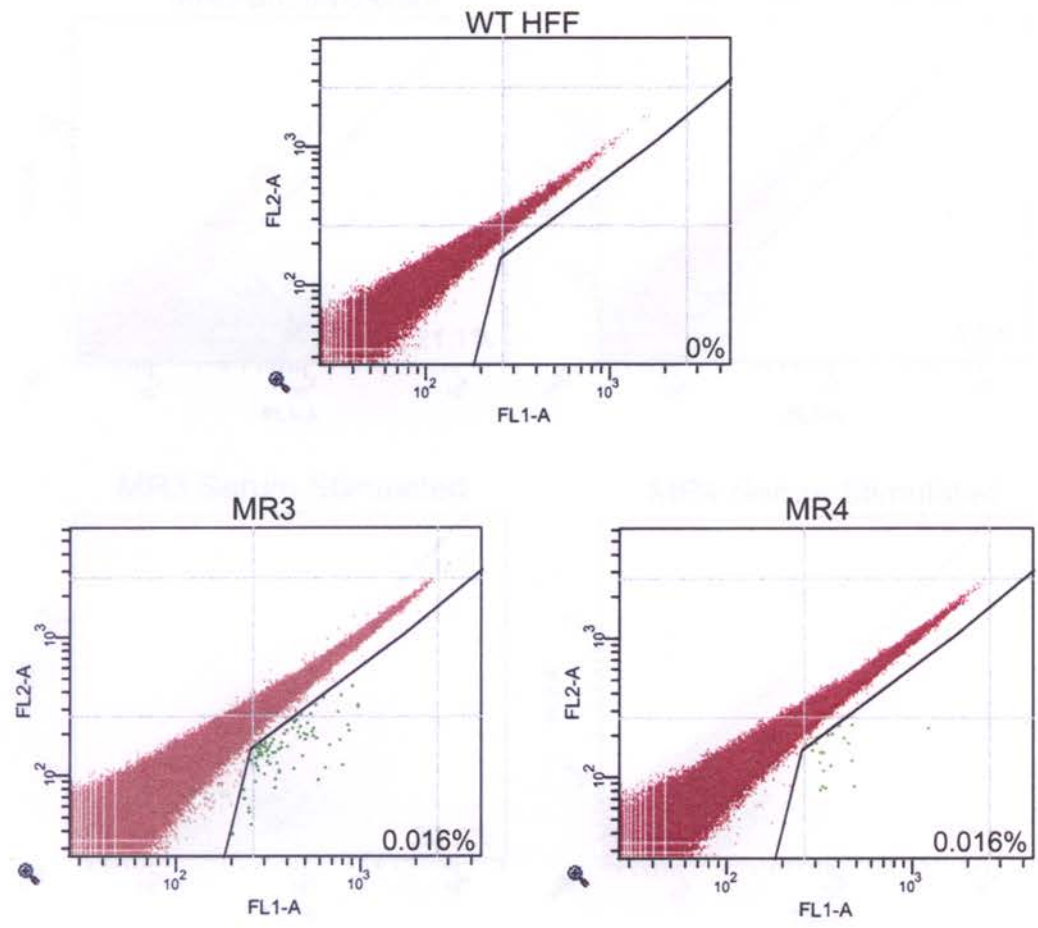


Figure 22. FACS histograms of MR3 and MR4 bulk population selection. The histograms shown are representative data sets from the selection sort of the MR3 and MR4 pre-enriched populations. The MR3 histogram details 5.7 x10^5 total events and the MR4 histogram shows 1.7 x 10^5 total events. Both MR3 and MR4 returned 0.016% EGFP-Luc positive cells where 94 and 86 total cells were respectively collected as bulk/mixed populations.

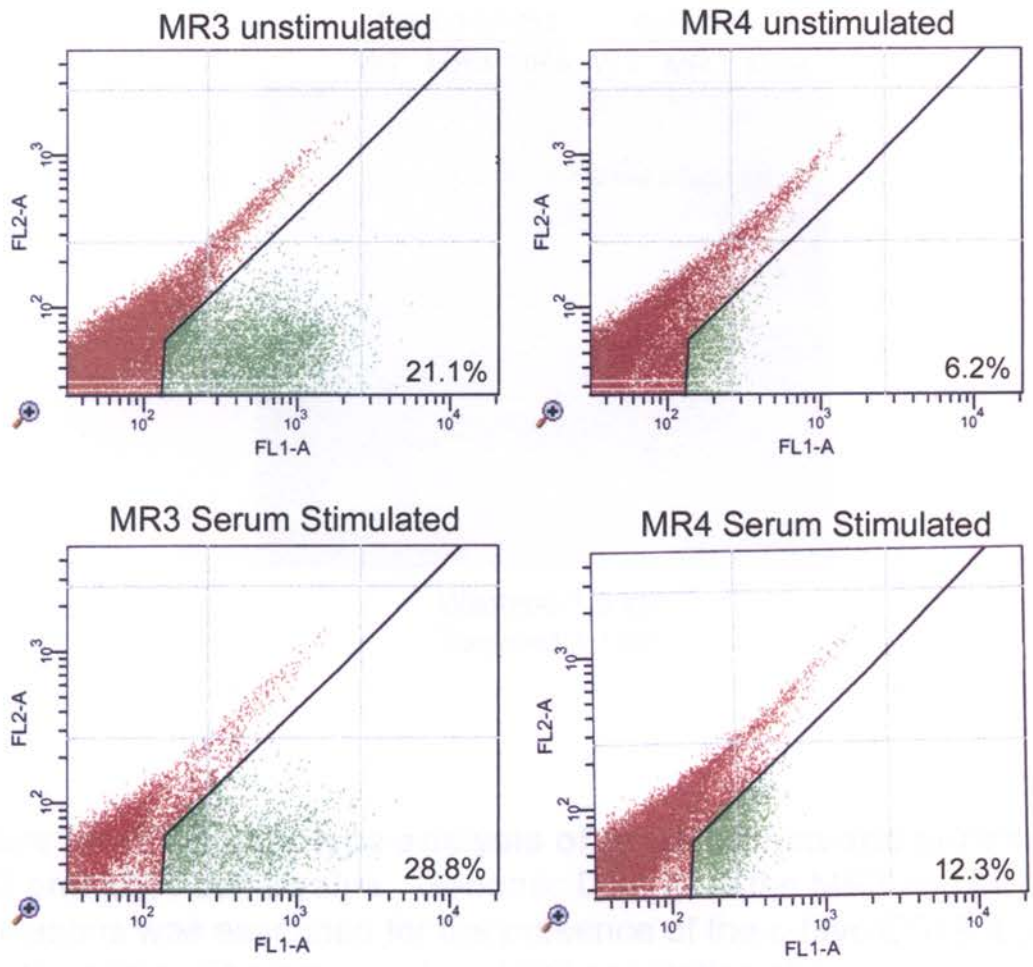


Figure 23. Enrichment sort of MR3 and MR4 bulk populations. Quiescent MR3 and MR4 populations were stimulated with 20% for 6 hours. EGFP-Luc positive cells for either MR3 or MR4 were collected as a pooled population. Serum stimulated MR3 cells showed a 1.4 fold increase in EGFP-Luc positive cells. The serum stimulated MR4 population experienced a 2 fold increase in EGFP-Luc positive cells.

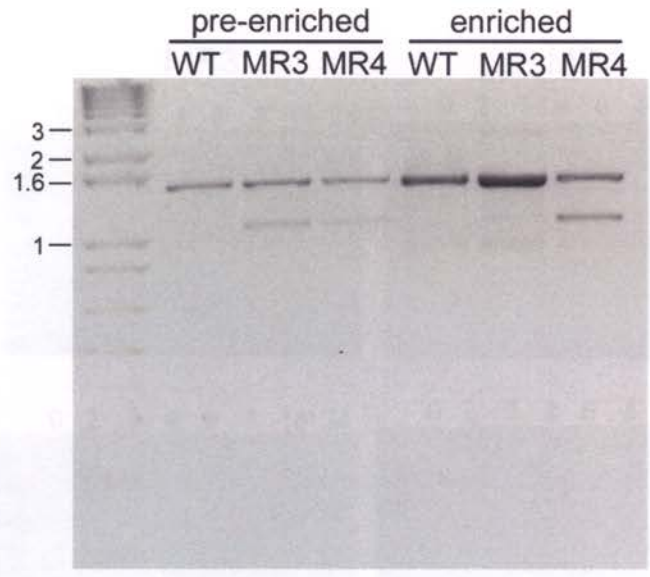


Figure 24. PCR genotype analysis of pre-enriched and enriched MR3 and MR4 cell strains. Genomic DNA from the MR3 and MR4 populations was examined for the presence of the c-Myc EGFP-Luc targeted allele. The pre-enriched MR3 population shows a strong band for the targeted allele which is largely lost in the MR3 enriched population. The MR4 pre-enriched population shows only a modest representation of the targeted allele which is strongly enhanced in the enriched MR4 population.

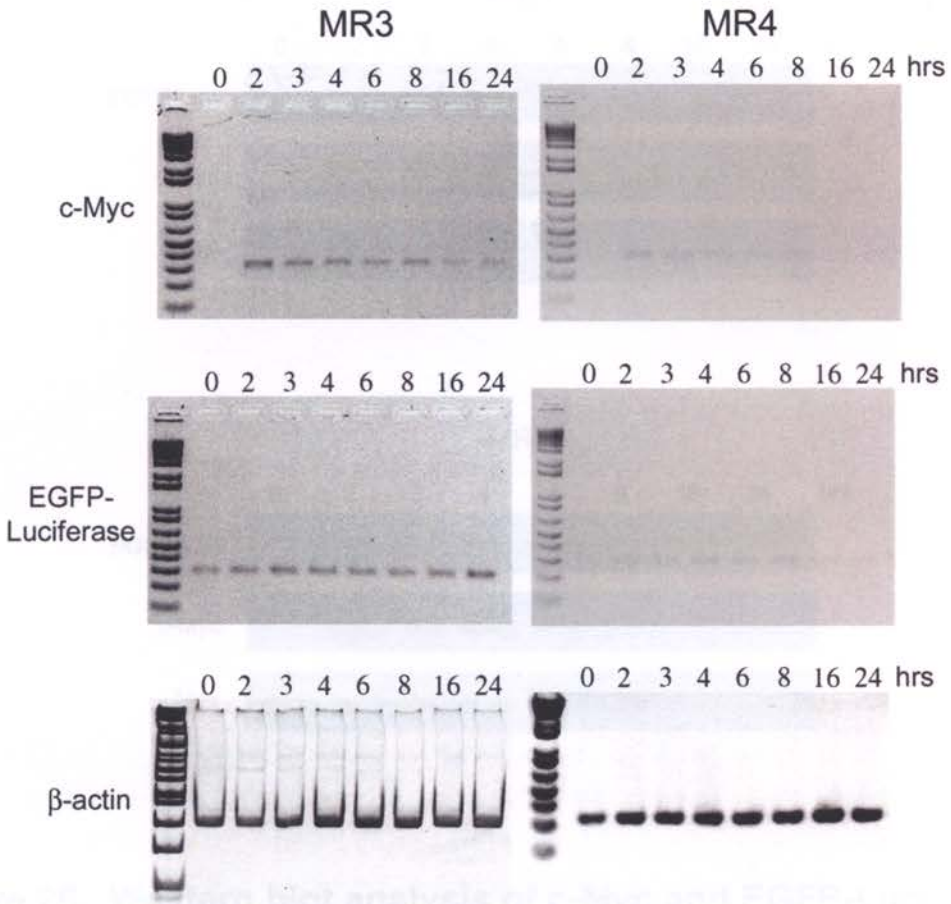


Figure 25. RT-PCR analysis of c-Myc and EGFP-Luc transcript expression following serum stimulation. RT-PCR was performed on cDNA synthesized from 1μg total RNA. MR3 and MR4 both show a rapid induction of c-Myc transcripts following serum stimulation. MR4 shows less overall c-Myc transcript expression which is due to its heteorzygosity at the c-Myc allele. MR3 shows constitutive expression of EGFP-Luc transripts indicating a strong representation of randomly targeted cells in the population. MR4 shows EGFP-Luc induction with similar kinetics to c-Myc transcript induction indicating the majority of reporter genes are being expressed under the serum responsive c-Myc promoter. PCR product was resolved on 1% SeaKem LE agarose 1XTBE gels with the exception of the β-actin sample for MR3 which was resolved on a 8% 1XTBE polyacrylamide gel.

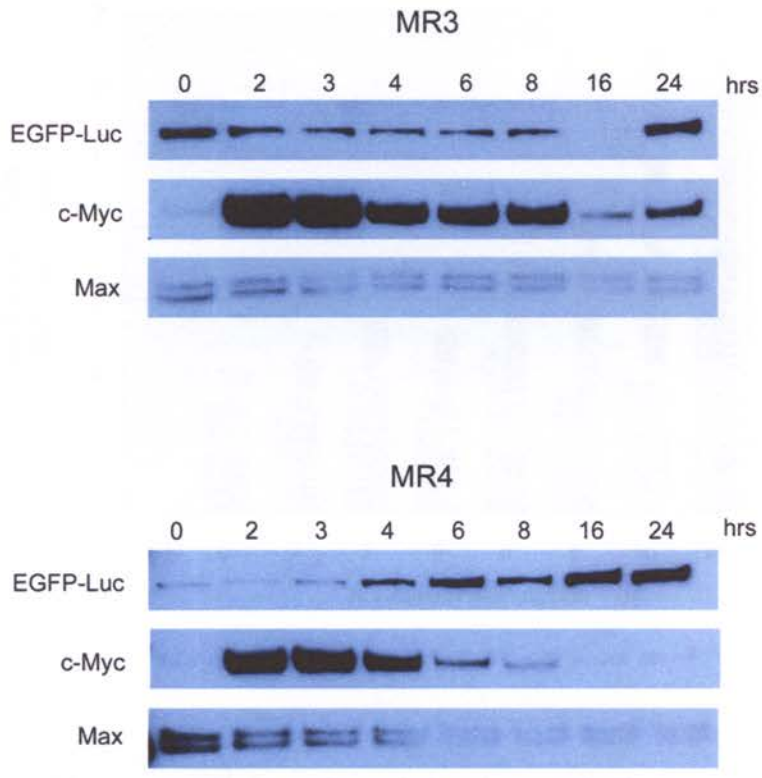


Figure 26. Western blot analysis of c-Myc and EGFP-Luc protein expression following serum stimulation. Western blot analysis was performed on protein lysates from the MR3 and MR4 cells following serum stimulation. MR3 and MR4 cells both display the expected c-Myc induction kinetics following serum stimulation. MR3 experiences constitive expression of EGFP-Luc while MR4 displays the expected accumulation of reporter molecules.

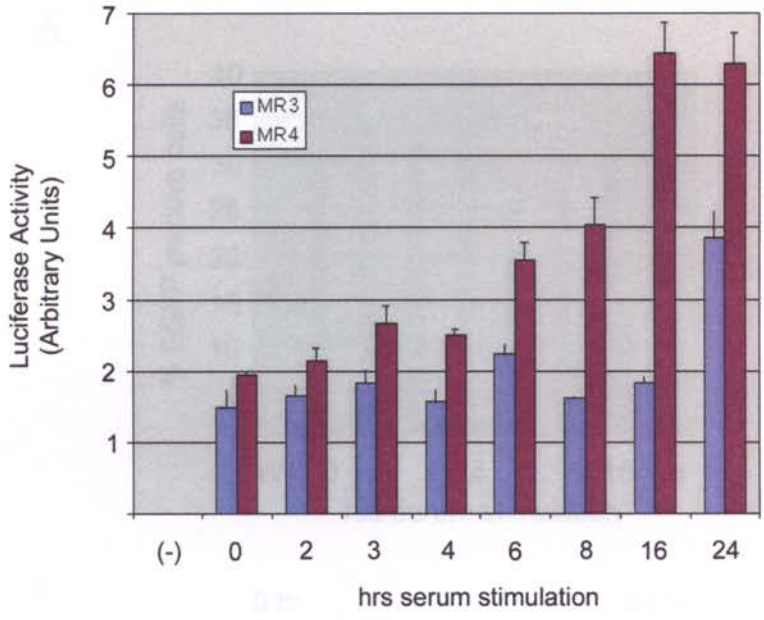


Figure 27. Luciferase activity of enriched populations following serum stimulation. Luciferase activity following serum stimulation of quiescent cells was assessed for the MR3 and MR4 populations. The MR3 population shows virtually unchanged Luciferase activity over a 16 hour period. The increase in MR3 activity at 24 hours is not thought to be directly due to serum stimulation. Luciferase activity is accumulated in the MR4 population which displays a 3.3 fold increase in activity over 24 hours. Data shown is the average of triplicate samples +/- SEM.

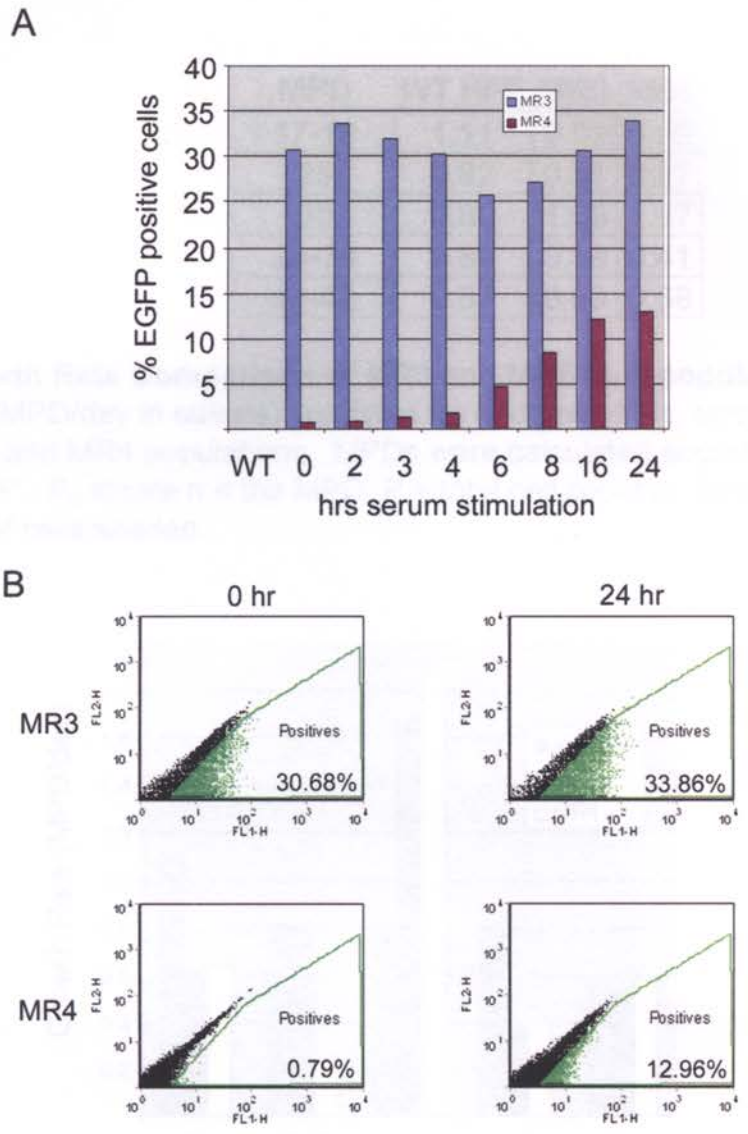


Figure 28. EGFP fluorescence activity of enriched populations following serum stimulation. Quiescent MR3 and MR4 populations were serum stimulated over a 24 hour period. Live cells were collected for FACS analysis. (A) The MR3 population shows a high level of deregulated expression of EGFP-Luc over the 24 hour period. The MR4 population shows a serum responsive accumulation of EGFP-Luc fluorescence over the 24 hour period. (B) Representative FACS histograms are shown for the MR3 (top) and MR4 (bottom) populations at 0 (left) and 24 (right) hours. Percent EGFP-Luc positive cells were determined for time course samples under the same gate using BD CellQuest Pro v. 5.2.

MPD	WT HFF	MR3	MR4
17-18	1.11	0.63	0.65
25	0.82	0.81	0.27
30	0.80	1.69	1.87
34-36	0.80	0.38	0.41
41-42	0.57	0.53	0.58

Table 3. Growth Rate Comparison of MR3 and MR4 bulk populations. Growth rates (MPD/day in culture) are listed for wildtype HFFs, and the enriched MR3 and MR4 populations. MPDs were calculated according to the formula: $2^n = P - P_o$ where n = the MPD, P = total cell count at density, and P_o = number of cells seeded.

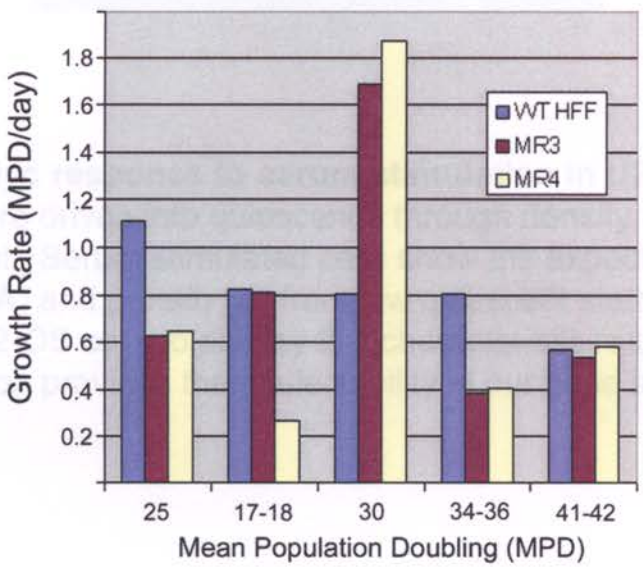


Figure 29. Growth Rate Comparision of MR3 and MR4 bulk populations. Graphical representation of the growth rates listed in Table 3. The growth rates fluctuate for MR3 and MR4 over the MPDs shown but overall they are more consistent with the wildtype HFF population than the MR1 and MR2 populations (Fig 21).

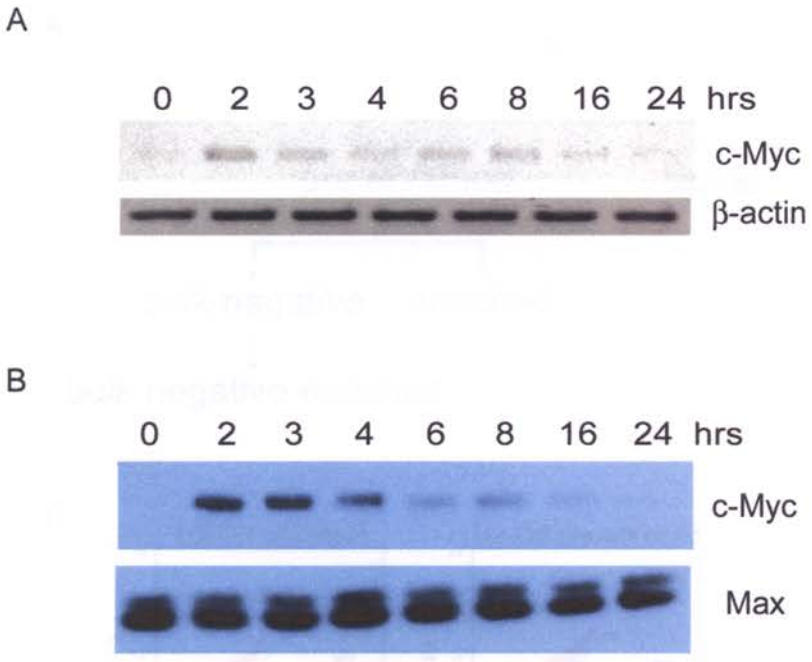


Figure 30. c-Myc response to serum stimulation in U2-OS cells. U2-OS cells were driven into quiescence through density arrest and serum withdrawl. Serum stimulated cells show the expected induction of c-Myc RNA (**A**) and protein (**B**) from low quiescent state levels. The ability of U2-OS cells to display this characteristic response to serum stimulation provides their selectability in our gene targeting system.

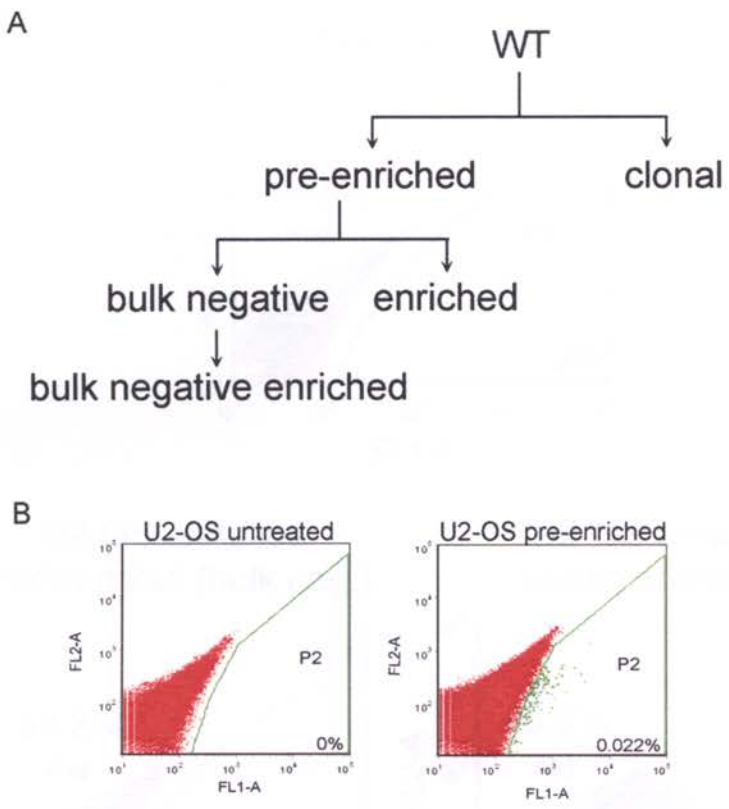


Figure 31. Selection of the pre-enriched c-Myc EGFP-Luc targeted U2-OS cells. (**A**) Diagram of cell lineage nomencalture for U2-OS cell targeting. (**B**) Infected U2-OS cells were brought to quiescence and then serum stimulated. EGFP positive cells were selelcted using FACS sorting. 96 cells were plated individually to establish clonal targeted populations. The remaining cells were collected as the bulk "pre-enriched" population. The infected U2-OS cells returned 0.022% EGFP positive cells which is 1.38 fold better than the percent EGFP cells returned from any of the primary HFF targeting experiments.

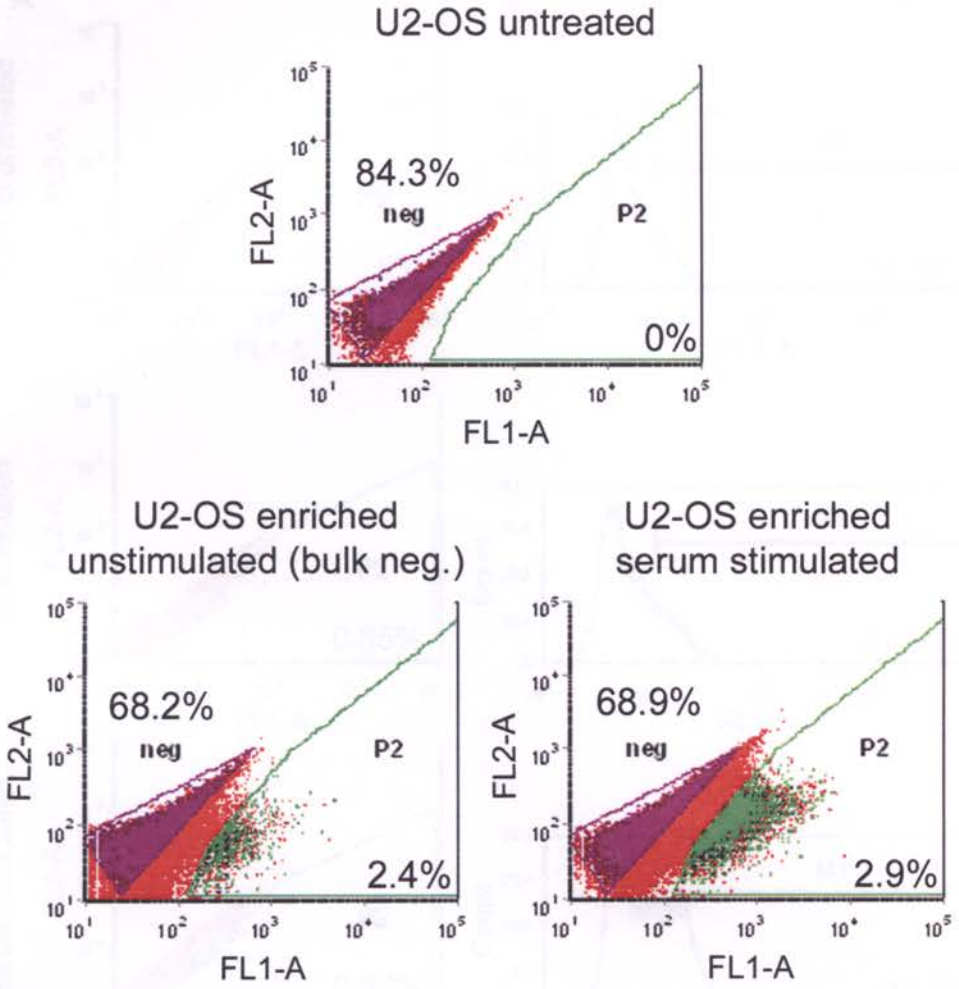


Figure 32. FACS enrichment of the U2-OS pre-enriched population. The bulk sorted U2-OS pre-enriched population was enriched through a round of serum stimulation. The EGFP-Luc positive cells (P2/green) were collected as single bulk populations for the unstimulated and serum stimulated samples. The enrichment did not return any significant difference in the percent EGFP-Luc positive cells. The bulk negative (neg/purple) cells were collected from the unstimulated population for subsequent enrichment.

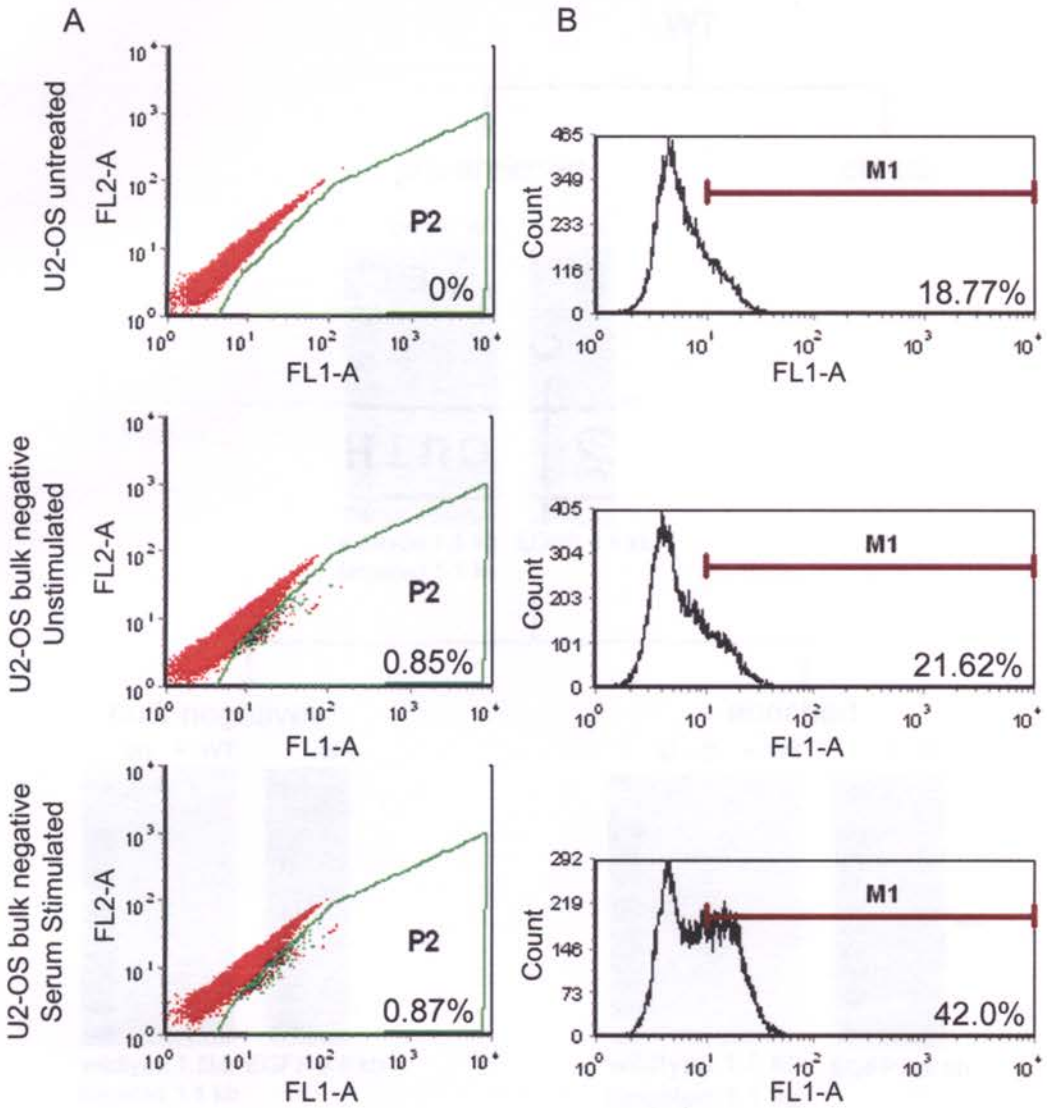


Figure 33. FACS enrichment of the U2-OS bulk negative population. The bulk negative population was enriched through a round of serum stimulation. The EGFP-Luc positive cells were collected as bulk populations for the unstimulated and serum stimulated samples. (A) The enrichment did not yeild any significant increase in EGFP-Luc expression for the stimulated bulk negative population. (B) The number of cells with a shift in green fluorescence (FL1-A) are shown. After subtracting background autofluorescence, the bulk negative serum stimulated sample shows an 8 fold increase in EGFP-Luc expression compared to the unstimulated sample. These cells cannot be collected as a pure population due to a large representation of autofluorescent cells at this intensity of green fluorescence.

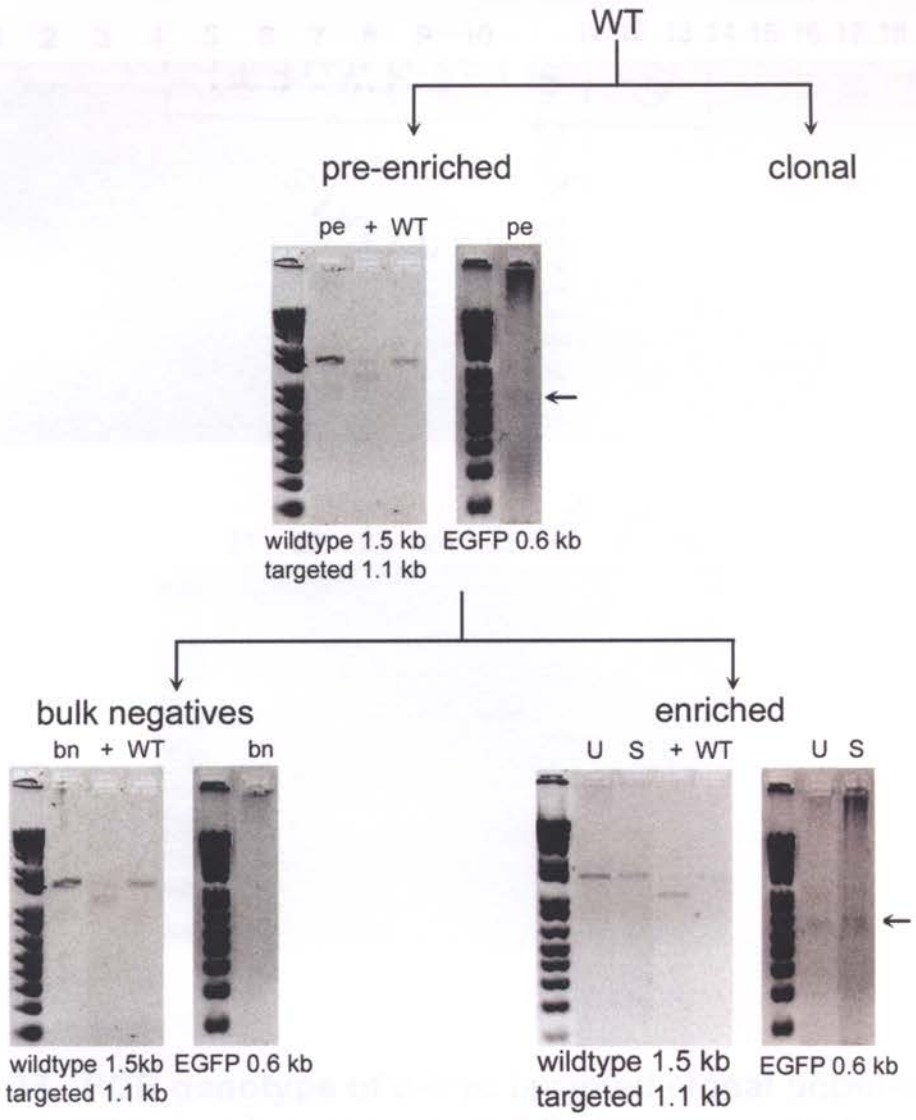


Figure 34. PCR genotype of bulk selected U2-OS populations. Genomic DNA from the bulk populations were collected for genotype analysis by PCR. The genotype PCR for the c-Myc targeted allele is seen in the gels on the left. The presence of targeted or randomly integrated EGFP-Luc reporter gene in the population is detected by PCR using internal EGFP primers (gels on right). pe, pre-enriched. bn, bulk negative. U, unstimulated. S, stimulated. +, MR1 positive control.

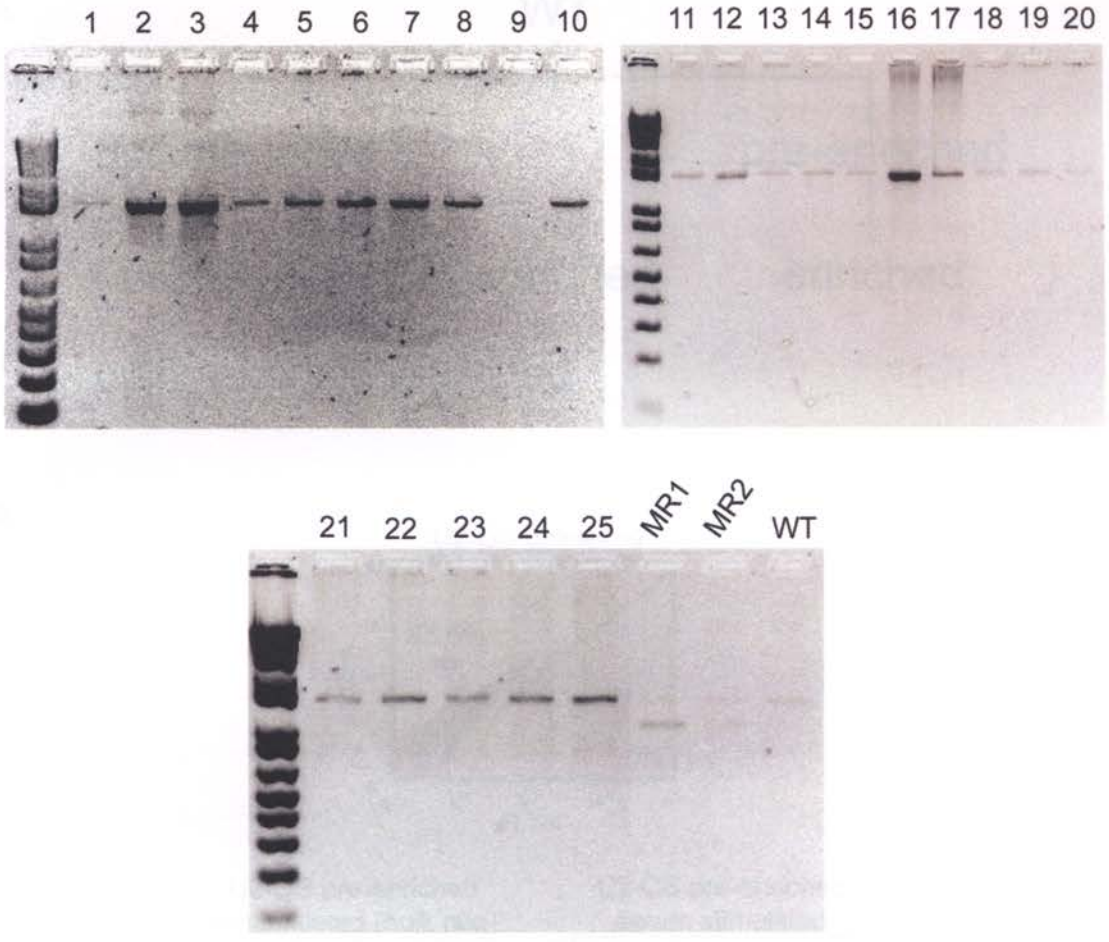


Figure 35. PCR genotype of c-Myc targeted clonal populations. Genomic DNA was collected from U2-OS clonal populations was subjected to genotype PCR. 31 clones grew out from the 96 cells collected at the "pre-enrichment" sort. None of the 31 clones displayed the targeted allele (25 clones shown). MR1 and MR2 were included as positive controls.

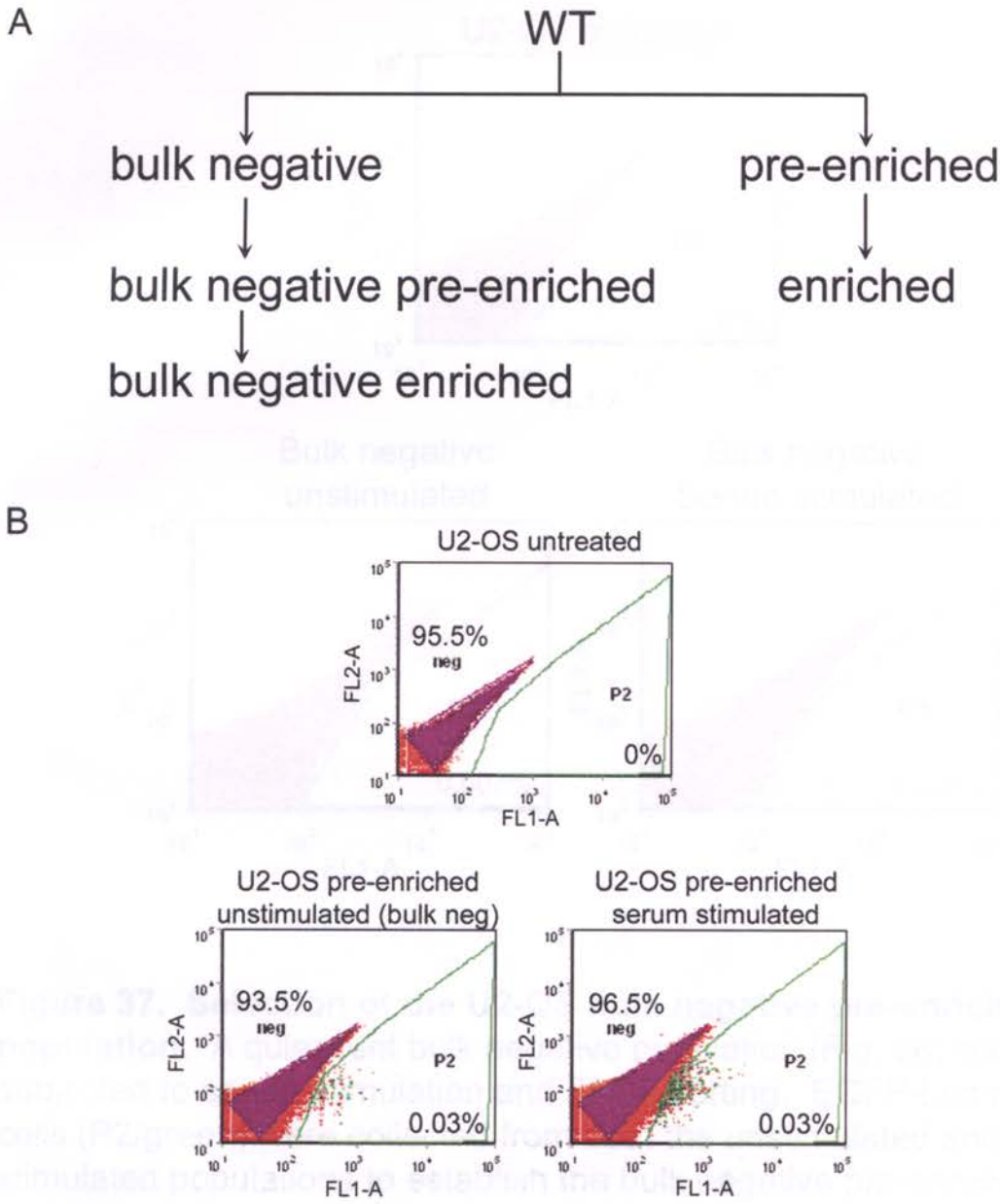


Figure 36. FACS seleIction of the U2-OS pre-enriched and bulk negative populations. (A) Diagram of population lineage nomenclature for the establishment of a U2-OS c-Myc EGFP-Luc reporter line. (B) Quiescent and serum stimulated U2-OS cells were selected by FACS following infection with the c-Myc EGFP-Luc targeting vector. A quiescent population was sorted for comparison of background EGFP-Luc expression as well as for the establishment of a bulk negative population that will be enriched for EGFP-Luc positive cells in subsequent selections (purple). The EGFP-Luc positive cells were collected from the serum stimulated population to establish the pre-enriched population (green).

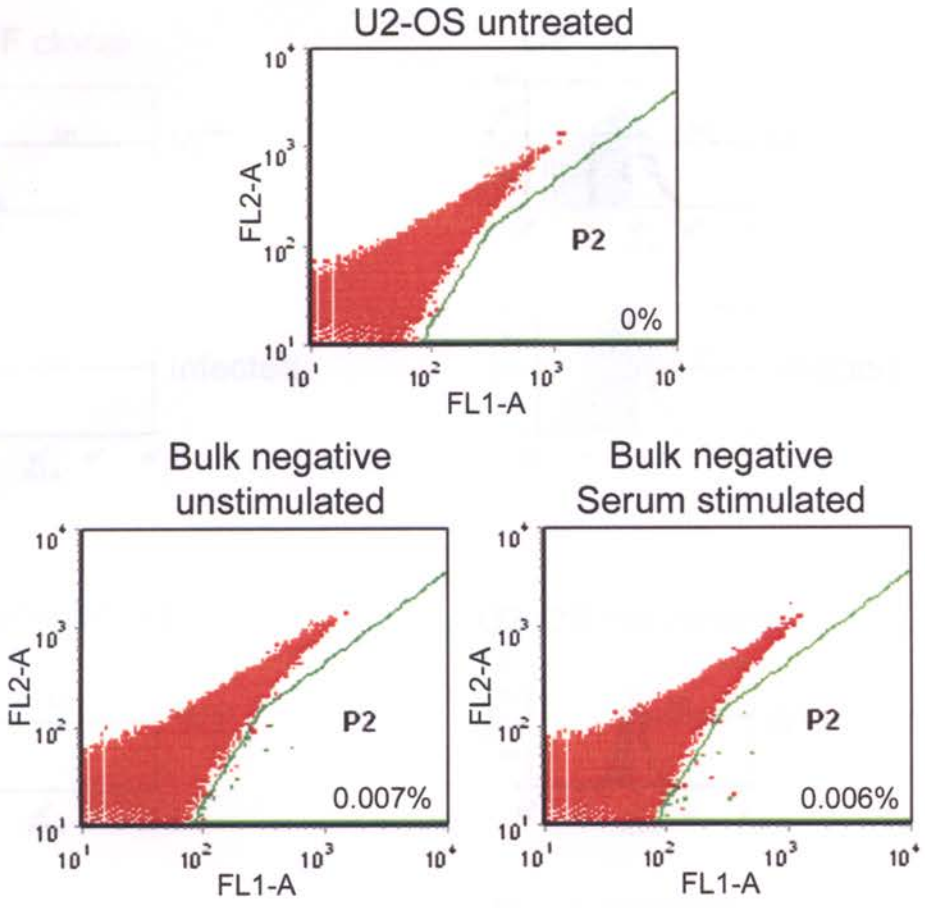


Figure 37. Selection of the U2-OS bulk negative pre-enriched population. A quiescent bulk negative population (Fig. 36) was subjected to serum stimulation and FACS sorting. EGFP-Luc positive cells (P2/green) were collected from both the unstimulated and serum stimulated populations to establish the bulk negative pre-enriched population. The serum stimulated bulk negative population shows EGFP-Luc expression at background levels.

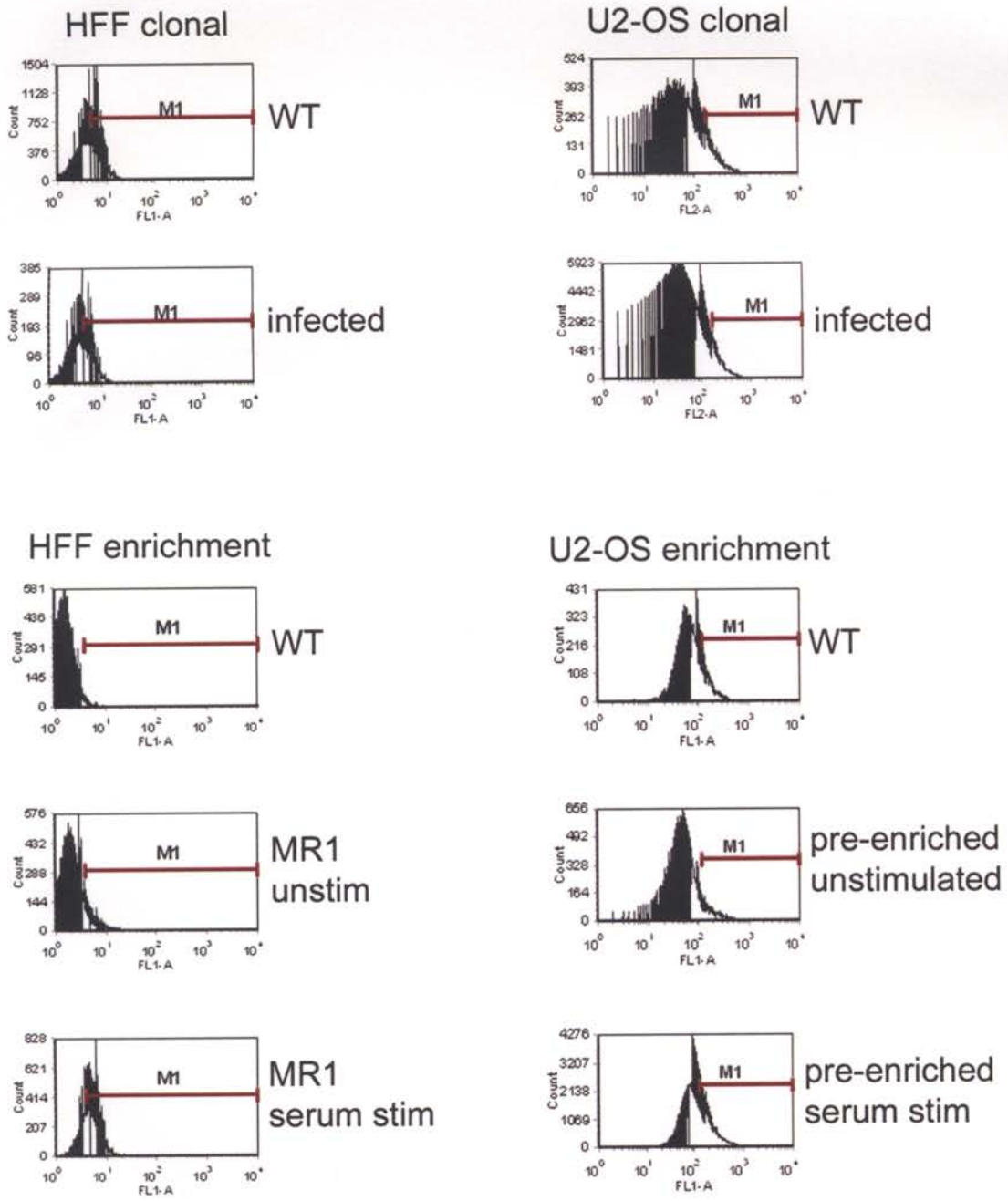


Figure 38. Background fluorescence comparison in the primary HFF and U2-OS cells. Histograms from representative FACS selections are shown for the HFF and U2-OS cells. The U2-OS cells show a large population wide shift in green fluorescence that move the population into the intensity range where the EGFP-Luc positive cells were selelcted from in the successful HFF targeting experiments.