Université d'Ottawa • University of Ottawa

Examining the repressive nature of D4Z4 repeats and their role in the pathogenesis of Facioscapulohumeral Muscular Dystrophy

By

Darren J. Yip

THESIS
Submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry, Microbiology and Immunology,
Human and Molecular Genetics
Faculty of Medicine
University of Ottawa

ABSTRACT

Molecular analysis of Facioscapulohumeral Muscular Dystrophy (FSHD) patients has demonstrated that the disorder is associated with deletions of multiple copies of a subtelomeric 3.3 kb unit termed D4Z4, which lie at 4q35. Sequence analysis of the D4Z4 repeat has failed to identify a functional transcript; thus, it has been postulated that FSHD results from position effects on a neighboring gene(s). The purpose of our investigation was to determine the function that these repeats may hold in the pathogenesis of FSHD by evaluating their regulatory effect on reporter gene expression. To address this, we developed LacZ expression constructs containing 1 to 4 D4Z4 repeats and assessed reporter gene activity using transient and stable transfection assays in C2C12 myoblasts. Our results have revealed inconclusive evidence to support a repressive effect involving the D4Z4 repeats in either transient transfections or stable clones grown under normal or myogenic differentiation conditions. However, during differentiation it was observed that the D4Z4 repeats confer a mild fusion defect onto the multinucleated myofibres. The presence of D4Z4 not only accompanied the reduced fusion competence of myoblast cultures resulting in the formation of less myotubes, but also corresponded to the heightened incidence of deformed myotubes. While the results of these experiments do not provide irrefutable evidence, they have provided some insight towards an association between the D4Z4 repeat and repression and have increased our understanding of the role D4Z4 repeats may play in the etiology of Facioscapulohumeral Muscular Dystrophy.

ACKNOWLEDGMENTS

Where to start, there are so many people who had a hand, or two in the success and completion of this project. First to my supervisor, Dr. Picketts, for providing me with support, guidance and an opportunity to see the bigger picture, Dave has imparted on to me an enthusiasm for research and science that I am sure will follow me in future endeavors. I must also thank the present and past members of the Picketts' lab, especially Nathalie, Maribeth, Dom, Magda, Todd and Shaobo for providing an environment which, not only nurtured the scientific process but also made it fun to go to work every morning. The support from my thesis advisory committee members, Drs. Dennis Bulman and Robert Korneluk has been much appreciated. The friends I have made during my time as a graduate student are too numerous to name, but all have had a role providing a positive outlook and help my get through some personal tragedies over the last few years. I must especially thank my parents and family, especially my mom who has helped edit the rough drafts of this manuscript, and now knows even less about what I do then before she started.

I must also acknowledge Dr. R.R. Frants (Leiden, The Netherlands) for graciously providing us with the cY13 cosmid. As well I must thank Dr. Robin Parks (Ottawa, ON) for providing not only the use of the AdCre vector (AdMA26), but for opening up his lab for my experiments and providing scientific guidance.

This thesis is dedicated to the memory of my grand parents, Kew Dock and Victoria Yip, who even from afar have lent me their support, good will and a drive to succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS [1]

A	Adenine
$A_{\#}$	Absorbance (at wavelength #)
Ad	Adenovirus
ATP	Adenosine Triphosphate
β-actin	Beta-actin
β-Gal	Beta-galactosidase
BOS	Eflα ubiquitous promoter
bp	Base Pairs
c	Signifies clone (e.g. cDY044; clone 4, derived from pDY04)
C	Cytosine
^{o}C	Degrees Celsius
cDNA	Complementary DNA
Ci	Curie
CK	Creatine Kinase
CMV	Cytomegalovirus
CPE	Centromere Position Effect
Cre	Cre-recombinase
DAB	3,3'-Diaminobenzidine
*DUX*1	Double Homeobox Gene 1 (also 2, 3 and 4)
dATP	Deoxyadenosine Triphosphate
dCTP	Deoxycytidine Triphosphate
DEPC	Diethyl Pyrocarbonate
dGTP	Deoxygausine Triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DMI	Deformed Myotube Index
DMSO	Dimtheyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside Triphosphate
E(var)	Enhancer of Variegation
EDTA	Ethylenediamine Tetraacetic Acid

[1] Abbreviations referring to genes are italicized (*e.g. DMD*). When referring to a human gene all letters are capitalized (*e.g. CPP*32) and for murine genes a capital is used only for the first letter (*Snf*2h). When referring to both human and mice the human designation is used by default.

FBS	Fetal Bovine Serum
FSHD	Facioscapulohumeral Muscular Dystrophy
FGF	Fibroblast Growth Factor
*FRG*1	Facioscapulohumeral Muscular Dystrophy Region Gene 1
G	Guanine
H	Histone (as in histone 1, 2A, 2B, 3 or 4)
HEFTI	Helicase-like Transcription Factor Target
HP1	Histone Protein 1
HRP	Horseradish Peroxidase
ICF	Immunodeficiency, Centromeric instability and Facial abnormalities Syndrome
IMR	Idiopathic Mental Retardation
kb	Kilobases
LacZ	Lac Operon gene for β-galactosidase
LB	Luria-Broth
LoxP	Cre-recombinase consensus excision sites
M, mM	Molar (moles per litre), Milimolar
MCS	Multiple Cloning Site
MD	Muscular Dystrophy
MEF2	Myocyte Enhancer Factor-2
MF20	Antibody for Myosin Heavy Chain
MFI	Myotube Fusion Index
min	Minutes
mL	Millilitre
mRNA	Messenger RNA
nm	Nanometer
N	any nucleoside base (A, T, G or C)
Neo^R	Neomycin Resistance Gene
ONPG	o-nitrophenyl-β-D-galactopyranoside
ORF	Open Reading Frame
p	Signifies Plasmid (e.g. pDY04)
PBS	Phosphate Buffered Saline
PcG	Polycomb Group
PCR	Polymerase Chain Reaction
PEV	Position Effect Variegation
PRE	Polycomb Response Element

RIGS	Repeat Induced Gene Silencing
RNA	Ribonucleic Acid
RNaseA	Ribonuclease A
RT	Reverse Transcriptase
RT-PCR	Reverse Transcriptase - Polymerase chain reaction
s	Seconds
Su(var)	Suppressor of Variegation
SD	Standard Deviation
SDS	Sodium Dodecly Sulfate
SRY	Sex-determining Region Y Chromosome
SSC	Saline Sodium Citrate
STAR	Sub-telomeric Anti-silencing Region
T	Thymine
TE	Tris-EDTA
TPE	Telomere Position Effect
Tris	Tris(hydroxymethyl) aminomethane
TrxG	Trithorax Group
U	Units
µL	Microlitres
UV	Ultraviolet
wt	Wild Type
X-Gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside
YAC	Yeast Artificial Chromosome

1.0 – Introduction

1.1 – Facioscapulohumeral Muscular Dystrophy

Facioscapulohumeral Muscular Dystrophy (FSHD, OMIM: 158900) is the third most prevalent neuromuscular disease, after Duchenne and Myotonic Muscular Dystrophies. FSHD is an autosomal dominant disorder that affects roughly one in 20,000 individuals [1-4]. In the North American population (~300 million) this translates into approximately 15,000 cases, the majority of which are familial; however *de novo* mutations account for 10-30% [5-9]. The causative disease gene(s) has not yet been identified. Treatment regimes consist primarily of increasing physical activities to slow the progression of deterioration and to administer treatment for specific symptoms.

1.1.1 – The Clinical Features of FSHD.

Although previously observed by Duchenne, FSHD was first described as a separate disorder by Landouzy and Desjerine in 1885 ([10] and ref cited). It is characterized as a progressive and clinically heterogeneous myopathy initially affecting facial and shoulder girdle muscles [1, 2, 5, 11]. During progression of the disease other skeletal muscle groups develop weakness and atrophy, preferentially the scapula fixators (shoulder), upper-arm (humeral, biceps and triceps) and abdominal muscles. Eventually the disease advances to include the foot-extensor and pelvic-girdle regions [2, 5, 12, 13]. Approximately 9-19% of patients with this disorder will eventually be confined to a wheelchair [3, 5, 12]. Some

FSHD patients have been observed to sleep with their eyes open, have noticeable difficulty smiling and have a pouting facial expression as a direct result of their facial muscle weakness [12]. In addition to the unique facial characteristics, there is an asymmetry of the scapula and upper limb muscles, which often result in a winged-like appearance (clinical term *genu recuvatum*) [5, 12]. Other symptoms observed in patients include retinal vasiopathies, which result in retinal talengectasia (microanyeurisms) and abnormal vascularization, high-frequency hearing impairment, mental retardation and epilepsy [1, 9, 14-19]. Congenital heart defects, hepatamegaly and mitochondrial myopathies are also associated with this disorder [1, 13, 18, 20]. Patients exhibit an extreme variability in the age of onset and in the clinical phenotypes expressed, ranging from asymptomatic to severely handicapped [2, 5, 7, 12, 16, 19, 21, 22]. The penetrance of this disease increases with maturity and is estimated at less than 5% for ages 0-4, 21% between the ages 5-9, 58% for the ages 10-14, 86% between 15-19 years and an estimated 95% penetrance by the age of 20 [3, 5]. There appears to be no loss in life span or fertility; however a higher rate of miscarriages has been suggested [5, 6, 9]. There are no clear indications of gender bias in the incidence of this disease; however males do tend to exhibit the more severe manifestations of the disorder [23, 24].

Affected muscle tissues are characterized by disorganized myotube formation, extensive cytoplasmic swelling, frequent muscle inflammation and deterioration, appearing as irregular sized moth-eaten structures [12, 20, 25-27]. At the biochemical level, the expression of the apoptotic enzyme Caspase 3 (*CPP32*) and the overexpression of Fibroblast Growth Factors 1 and 2 (*FGF1*, *FGF2*) and *FGF* receptor 4 have also been suggested to play a role in the muscle deterioration [25, 27]. An increase in the level of

serum Creatine Kinase (*CK*), that is commonly associated with many myopathies such as Duchenne, Limb-girdle, Scapulperoneal, Becker's and Myotonic Muscuar Dystrophies, is less elevated among FSHD patients [5, 12, 13, 25, 28]. Facioscapulohumeral MD was historically diagnosed based on four criteria, the clinical presentation of the facial muscle weakness and scapular involvement, familial history, electromyography and histochemical analysis of muscle biopsies. The facial muscle weakness associated with FSHD is important for clinical diagnosis as the disorder presents very similarly to a mild case of Limb-girdle MD [25] or as a very similar syndrome called Scapuloperoneal MD [29, 30]. Both of these diseases show weakness in the shoulder-girdle and peroneal muscle groups [2, 5, 13, 31].

1.1.2 – The Molecular Genetics of FSHD.

The gene for FSHD was genetically mapped to the subtelomeric region of chromosome 4 in 1990 [32, 33]. Subsequently, linkage analysis using a large set of markers, fluorescence *in situ* hybridization and radiation hybrid mapping established a more refined localization to between the DNA marker D4S139 and the telomeric repeats of chromosome 4 at q35 [33-36]. Several positional cloning and sequence analysis techniques have been applied; however a disease gene has not yet been identified. Molecular analysis of FSHD patients has associated the disease with large deletions of a 3.3kb tandem repeat termed D4Z4 [2, 14, 32, 36-39]. These repeats lie within a highly polymorphic *Eco*RI fragment, distal to the DNA marker D4S139 at 4q35 [2, 24, 28, 32, 37, 38, 40, 41]. The size of the polymorphic region can be assessed by hybridizing for this fragment in an *Eco*RI genomic digest with the probe p13E-11 (D4F104S1) located just 5kb proximal to the D4Z4

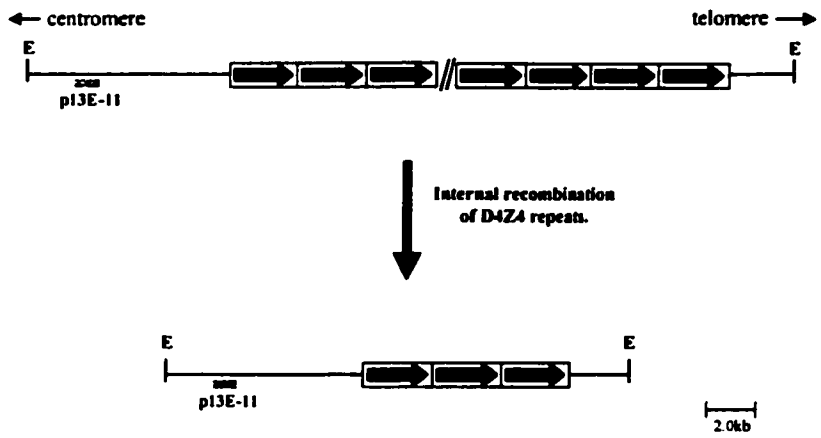
repeats [3, 8, 37, 38, 42]. Deletion of the D4F104S locus has no pathogenic effect on FSHD [9]. Typically a normal individual would have between 10-100 copies of this repeat generating a *Eco*RI fragment of ~38-350kb, while affected individuals generally have fragments ranging from 10-35kb or approximately 1-10 D4Z4 repeat copies [8, 9, 37, 39, 43, 44], refer to (Figure 1(A)). The variability observed in size of the internal deletions suggests that the fragments arise from independent rearrangements [3, 28]. There is a correlation between the age of onset and severity of symptoms to the size of the deletion. Lunt *et al.*, [3, 5] found that the larger the deletion, the more severe phenotypes and the earlier the age of onset (Figure 1(B)).

The probe p13E-11 has been found to detect a second polymorphic locus at 10q26 and a smaller 9.5 kb chromosome Y-specific band; however these and the small arrays generated by recombination on chromosome 10 are non-pathogenic [7, 35, 41, 44-46]. It has been noted that subtelomeric regions such as 4q35 and 10q26 are not only highly unstable and prone to recombination, but also prone to translocation and crossover rearrangement events [6, 8, 17, 37, 44, 47]. The repeats localized at 10q26 are 98% identical with the repeats on 4q35 and are prone to the same type of internal recombinations, and so create a more complicated analysis for molecular diagnosis by generating bands ranging from 20-300kb [2, 44]. However, sequence analysis of the 10q26 associated 3.3kb repeats revealed a novel *Bln*I restriction site not present within the 4q35 repeats [41, 47]. Therefore, double digestion of genomic DNA with *EcoR*I and *Bln*I provided a means for the discrimination of p13E-11 *EcoR*I fragments from chromosome 4 or 10. Restriction fragment length polymorphism (RFLP) of *EcoR*I and *EcoRI/Bln*I genomic digests followed by hybridization with p13E-11 can thus be used as a reliable molecular diagnostic test for

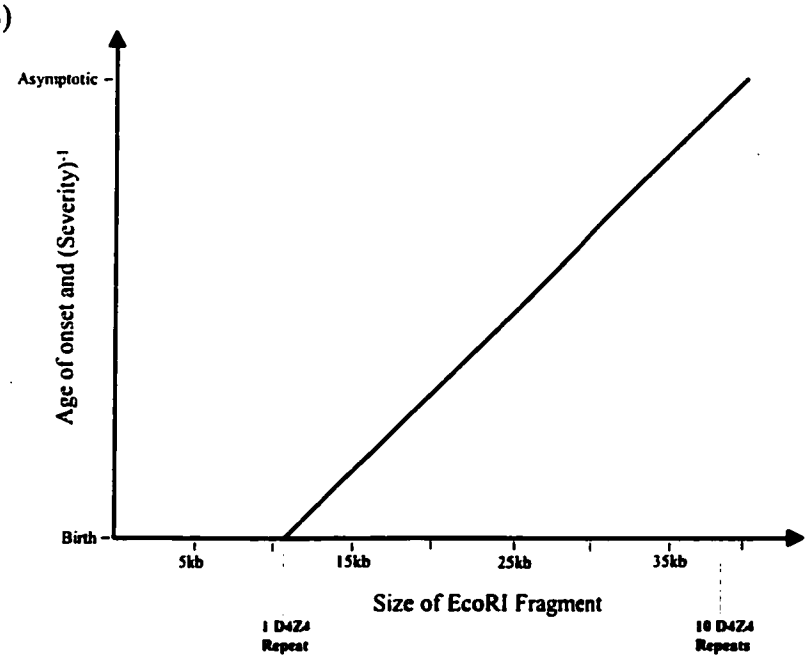
Figure 1. FSHD is caused by a deletion of D4Z4 repeats within a polymorphic *Eco*RI fragment.

(A) Molecular analysis of FSHD patients has associated the disease with large deletions of the 3.3kb D4Z4 repeats (boxed arrows). Internal recombinations within the D4Z4 array create a highly polymorphic *Eco*RI fragment, which can be visualized using the probe p13E-11 (gray box). [E] = *Eco*RI restriction site. (Adapted from [31]) (B) Correlation of the age of onset and disease severity to the size of the polymorphic *Eco*R1 fragment starting at 10.2kb (a patient containing 1 D4Z4 repeat). Due to the wide variability in symptoms many people genetically predisopsed to FSHD may remain asympotic throughout their lives. This figure was compiled from data previously published [3].

(A)


centromere
telomere
E
E
p13E-11
Internal recombination of D4Z4 repeats.
E
E
p13E-11
2.0kb

(B)


Asymptotic
Age of onset and (Severity)-1
Birth
5kb
15kb
25kb
35kb
Size of EcoRI Fragment
1 D4Z4 Repeat
10 D4Z4 Repeats

FSHD [35, 41, 44, 48]. To further obscure the molecular diagnosis there is evidence of subtelomeric exchange between the 4q35 and 10q26 repeat units. Such exchanges result in a mixture of the two repeats at both loci and occurs naturally in 20% of the population [44, 46, 47]. The implications of these events suggest that both molecular and clinical techniques must be evaluated for FSHD diagnosis. In the majority of FSHD families, a short *Eco*RI fragment co-segregates with the disease [7-9, 38, 42]. However in approximately 5% of FSHD patients the disease is not linked to 4q35. These cases are presumed to result from mutations in a causative gene(s) or be due to misdiagnosis [1, 2, 12]. Further complicating the diagnosis of FSHD is the high incidence of somatic mosaicism, wherein the causal mutation is a deletion of the D4Z4 repeats in a proportion of mitotically dividing cells [2, 7, 23, 24, 40]. The age of onset and the severity of symptoms in mosaic patients is linked directly to the size of the integral deletion of D4Z4 repeats and the proportion of somatic cells, which possess this FSHD genotype [3, 23]. It has been observed that while the incidence of mosaic males is less than females, males present the more severe manifestations [15, 22, 23].

In order to discern the molecular basis of Facioscapulohumeral Muscular Dystrophy the D4Z4 repeat, the repeat arrays and adjacent regions have been scrutinized for expressed sequences [40, 43, 44, 49-51]. Based on evidence involving the deletion of D4Z4 repeats, three potential models have been explored:

[1] – There is a gene contained within the D4Z4 repeats whose normal function is compromised as a result of the deletion.

[2] – The repeats normally function as a buffer preventing telomeric gene silencing. The reduced size of repeat arrays in FSHD patients would allow telomeric position effect (TPE) to influence expression of neighbouring genes.

[3] – The repeating D4Z4 units possess repressive elements, which constitute a catalyst necessary for induced formation of heterochromatin. Heterochromatin may regulate the expression of centromeric genes or key regulatory elements. The loss of these repeats could therefore cause the inappropriate upregulation or continued expression of gene(s), the product of which may be toxic to muscle cells.

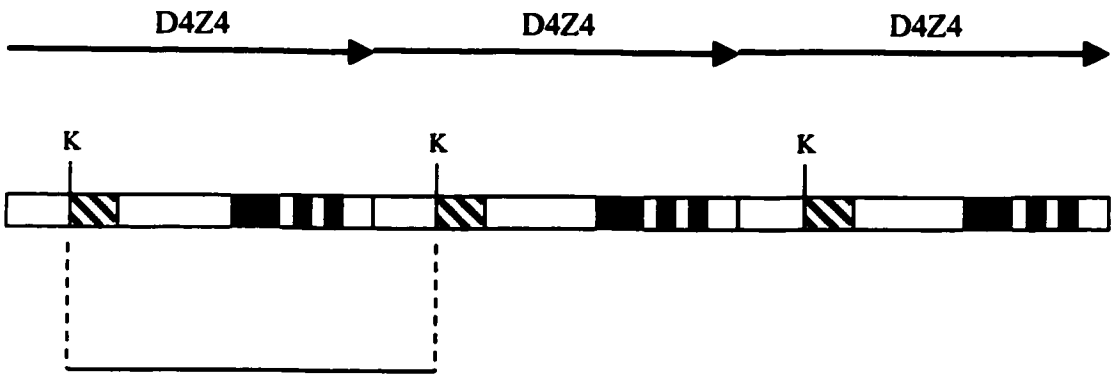
1.2 – The D4Z4 Repeats and Facioscapulohumeral Muscular Dystrophy.

There are many examples of how expanded repeats are associated with a genetic disease by interfering with gene stability or function. Fragile X Syndrome, Myotonic Dystrophy, Huntington's disease and Spinobulbar Muscular Atrophy, Spinocerebellar Ataxia (type 1, 2, 3, 6, 7, 8 and 12) are examples of how expansions in trinucleotide repeats destabilize a known gene product or its genomic region [52-61]. As well, there is an example where the addition of larger repeats (5.5kb) to the apolipoprotein(a) gene (*APO*(a)) can form additional Kringle IV repeat subunits, leading to an increased risk of atherosclerosis [62, 63]. Since FSHD has been causally linked to large deletions of integral copies of the D4Z4 repeat, the possible presence of a gene or gene-modifying elements within the repeat has been evaluated. Several groups determined the sequence of D4Z4 repeats [40, 43, 64]; Acc.L32607, L24543, D38024). Although the sequence of each repeat analyzed was not 100% identical they do appear highly similar. The 3.3kb or D4Z4 repeats are GC rich (approximately 71% C + G) [35, 43, 51, 65-67] and each repeat contains a double-homeobox (or double-homeodomain) related sequence motif. Also contained within the repeat are an LSau middle-repetitive and a hhsmp3 low-copy type repetitive sequence,

which are preferentially localized in heterochromatic regions [2, 35, 39, 40, 43, 51, 65]. For a schematic diagram of the D4Z4 repeat refer to Figure 2. Analysis by Hewitt *et al.*, [43] showed that the paired sequences were incorporated into a 405bp open reading frame (ORF), and they further suggested the GC-rich LSau repetitive element contained a putative promoter. More recently, analysis of the D4Z4 sequence has uncovered a second larger ORF with a putative promoter [51]. The potential promoter region of this 1.2kb ORF includes 2 MEF2 and a MyoD muscle specific enhancer binding domains, a GC-box, a TACAA box and 3 consensus-binding sites for the ubiquitous transcription factor SP1. The TACAA sequence is a variation of the more common TATAA box (TATA box), necessary for RNA polymerase II initiation of transcription [51, 68]. The myocyte enhancer factor-2 (MEF2) and myogenic determination factor MyoD are identified by muscle specific DNA-binding activity and have been identified for their importance in skeletal and cardiac muscle gene expression [69-72]. The MEF2-binding site, $CTA(A/T)_4TAG$, and the MyoD specific binding site, CANNTG (the so called E-Boxes), have been shown to be essential for the expression of many muscle specific genes [69, 73-76]. Myo*D* and MEF2 act together with the other myogenic determination factors, myogenin, MYF5 and MRF4 to regulate the genes necessary for directing and maintaining terminal myogenic differentiation [69-73, 77]. It was found that this putative D4Z4-promoter could indeed drive transient expression of a fused luciferase reporter gene, with strong activity [51]. This data is corroborated by previous analysis of the HEFT1 promoter from *DUX*1 and *DUX*2 [66]. *DUX*1 and 2 are novel transcription factors that contain a double-homeobox domain. These genes are found as a single open reading frame within 3.3kb repeats dispersed throughout the genome on acrocentric chromosomes; however neither *DUX*1, *DUX*2, nor a third member *DUX*3 were

Figure 2. The positions of sequence elements within the D4Z4 repeat.

A schematic drawing of three complete D4Z4 repeats (arrows, top), depicting internal repetitive elements (LSau, hhsmp3 and homeobox motifs). 3.3kb *Kpn*I unit cloned for further experiments is shown. [K] = *Kpn*I restriction site.



3.3kb unit cloned for experiments

FIGURE KEY

- LSau repeats
- hhsmp3 repeats
- Homeobox motif

mapped to the 4q35 repeats [51, 66]. Ding and colleagues [66] did discover that the 182bp HEFT1 promoter shows 87% homology to sequence within the D4Z4 repeat, suggesting that a *DUX*1-like protein may be encoded within the 4q35 repeats. *DUX*4, as this double-homeobox motif was later designated, and its 3 other homologues (*DUX*1, 2, 3) share strong homology to the *Drosophila paired* and *orthodenticale* homeodomains [2, 38, 40, 43, 64, 66]. Moreover, these domains are associated with a group of basic helix-turn-helix proteins commonly characterized as developmentally important transcriptional regulators. These so called Homeobox genes, which include the PAX, HOX, EMX and MSX families, and many other novel genes, are highly conserved throughout evolution [73, 78-81]. Interestingly, comparative sequence analysis shows that the D4Z4 repeats are only found in the primate species lineage, arising approximately 25 million years ago [45, 64, 81]. Sequence similarity between the D4Z4 repeat and the muscle specific *PAX*3 and *MHox* paired-homeobox motifs provides insight towards a muscle specific role [40, 82-84]. Moreover, the presence of the muscle-specific factor binding domains for MEF2 and MyoD suggest that there may be a muscle related role for *DUX*4 [51, 66, 69, 71, 85].

Two disease models can be derived from these findings. Firstly, if there is a gene contained within the D4Z4 repeat, a deletion of integral copies from the D4Z4 array may reduce the number of active genes. This would result in a lower dosage of gene products resulting in FSHD pathology through a "loss of function" mutation. Conversely, as the D4Z4 repeats lie within a putatively heterochromatin region [17, 52, 65, 86, 87], it suggests the *DUX*4 gene is normally non-functional. A partial deletion of integral repeats may destabilize the heterochromatin structure of the region allowing expression of the *DUX*4 gene in a subset of muscle cells representing a "gain of function" mutation. The product of

DUX4 may prove to be toxic when expressed under abnormal conditions, possibly due to a strong dimerization potential common to *DUX1* and *DUX2* [51, 66]. Despite the presence of this putative promoter, no functional gene transcript has ever been isolated from the D4Z4 sequence [40, 43, 49-51, 65, 67]. Furthermore, D4Z4-homologous repeats have been localized to other regions of the genome including chromosomes 1p, 2p, 4p, 9q, 12p, 13q, 18p, 21q, and 22q as well as the pericentromeric regions of chromosomes 1, 3, 9, 10 and Y [45, 50, 64-66, 86]. Taken together, this data suggests that the disease gene is not contained within the repeats themselves but may be located within the neighbouring proximal region. Therefore the reduction in the size of the repeat array may be suggestive of a role in regulating expression of a nearby gene through a process called position effect variegation (see below).

1.3 – Position Effect Variegation and Facioscapulohumeral Muscular Dystrophy.

Position effect variegation (PEV) is a term encompassing a wide assortment of processes whereby the expression of a gene is regulated in part by the surrounding chromatin environment [88-92]. In chromatin, DNA is organized and packaged within the genome by structural proteins called histones [47, 68, 93]. The nucleosome, which is the fundamental and structural unit of chromatin, is comprised of 146bp of DNA wrapped around a histone (H) octamer containing 2 of each: H2A, H2B, H3 and H4 proteins [94, 95]. Additional protein subunits such as H1 act as nucleosome linkers. Chromatin has several layers of organizational structure; the linear arrays of nucleosomes that comprise the primary structure of chromatin can be folded and condensed into higher-order structures

through protein-protein and protein-DNA interactions [96]. At the highest level of condensation there are two clear domains, euchromatin and heterochromatin ([96-98] and references within). Heterochromatin generally refers to non-transcribed, densely packed chromatin that is cytologically defined as regions of the genome, which remain condensed throughout the cell cycle [68, 91, 92, 96, 99, 100]. Heterochromatin tends to be concentrated at the centromeric and telomeric regions (constitutive heterochromatin); however a molecular signal can result in the stable, yet reversible formation of a heterochromatin (temporal or facultative heterochromatin) and this can occur anywhere [101-105]. Both forms of heterochromatin are involved in many functions including nuclear organization, chromosome segregation and gene silencing [89, 92, 97, 106-108]. Euchromatin, on the other hand, is a less condensed organization of chromatin where the vast majority of actively expressing genes tends to reside [68, 93]. In general, DNA in heterochromatin tends to be heavily methylated while euchromatin contains more acetylated histones [103, 109-112]. Heterochromatin is one of the primary means of transcriptional repression [95, 113-117]. Therefore, for gene expression (spatially, temporally, and quantitatively) a gene requires not only intact coding sequence and functional regulatory elements but also the appropriate chromatin environment. In 1930, while studying *Drosophila* genetics, Muller noticed that when the euchromatic gene, *white* was placed near a heterochromatin boundary the gene was silenced, but only in a portion of cells ([106] and references within). The resulting observation was a mixture of white and red eye facets; however if the white gene is completely knocked out, 100% white facets are formed [118, 119]. This is the classic example of position effect, which simply means that a gene's expression is modified as a result of the surrounding chromatin environment wherein it lies [116, 120]. Position effect variegation is thus the stochastic on/off or mosaic expression of

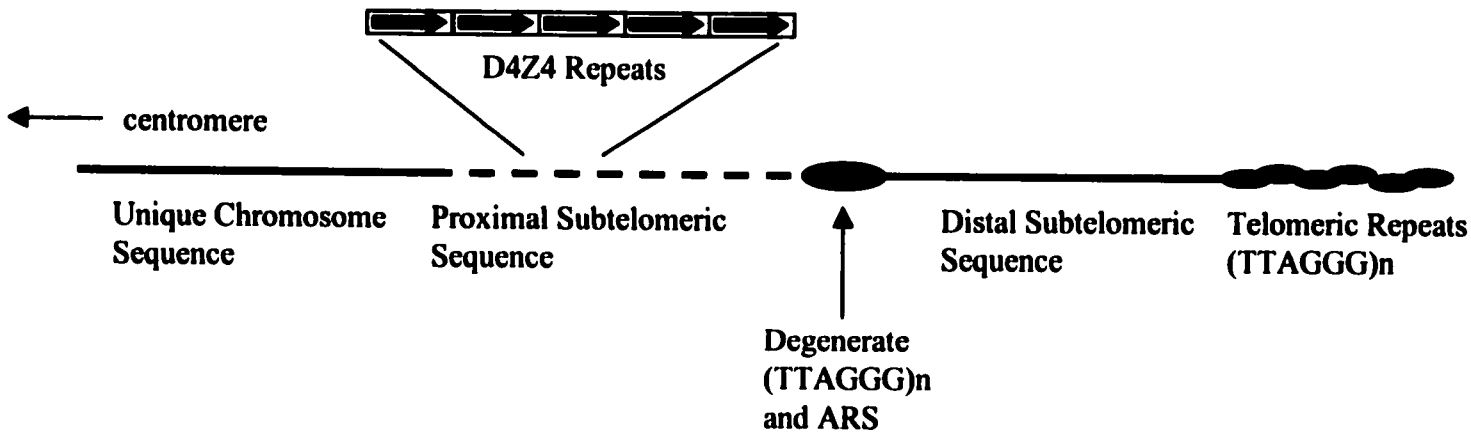
a gene placed in or adjacent to a junction between euchromatin and heterochromatin [88, 100, 113, 121-123]. The result is an allele-specific transcriptional repression through chromatin remodeling on a cell by cell basis. The following models, which are proposed for FSHD pathogenesis, can be grouped into those that are the result of constitutive chromatin or facultative chromatin position effects.

1.3.1 – Telomeric Silencing and FSHD (Constitutive Heterochromatin).

Constitutive heterochromatin, which is concentrated at the telomeres and centromeres of chromosomes, was originally thought of as "junk" DNA with no coding potential and is inextricably linked to late-replication [49, 101, 104, 124]. The centromeres are the sites of spindle attachment and are the loci responsible for poleward movement during mitosis and meiosis [68, 101, 104, 124]. Primate centromeres are prominent heterochromatic regions built up of repetitive DNA sequences, primarily the 171bp basic α-satellite, and a protein scaffold [91, 93, 102]. Telomeres are the natural ends to the chromosomes and comprise between 4-15kb of repeating units of TTAGGG. They serve a primary role as a sealer and protector of the chromosome ends from degeneration and as a center for the initiation of chromosome duplication [68, 125-130]. Between these hexamer repeats and the most distal euchromatin domain lies the subtelomeric region, ranging from several to a few thousand kilobases. This region contains mostly middle and highly repetitive sequences but also tandem arrays of unique repeats [17, 125, 126, 131]. It is in this subtelomeric region that the Facioscapulohumeral Muscular Dystrophy associated-D4Z4 repeats are located [15, 35, 45, 46, 49]. For an overview of the 4q35 subtelomeric region refer to Figure 3. The boundaries of the centromeric and telomeric regions tend to be

Figure 3. Organization of the subtelomeric sequence.

The subtelomeric region is divided into proximal and distal regions by the degenerate telomeric repeats and an origin of replication (ARS). Between the telomeric repeats and the degenerate telomeric repeats (gray ovals) lies the distal subtelomeric sequences. The predicted location of the D4Z4 repeats (boxed arrows) is within the proximal subtelomeric sequence (dashed line). The unique chromosome sequence contains gene rich euchromatin. This figure is modified from [131].

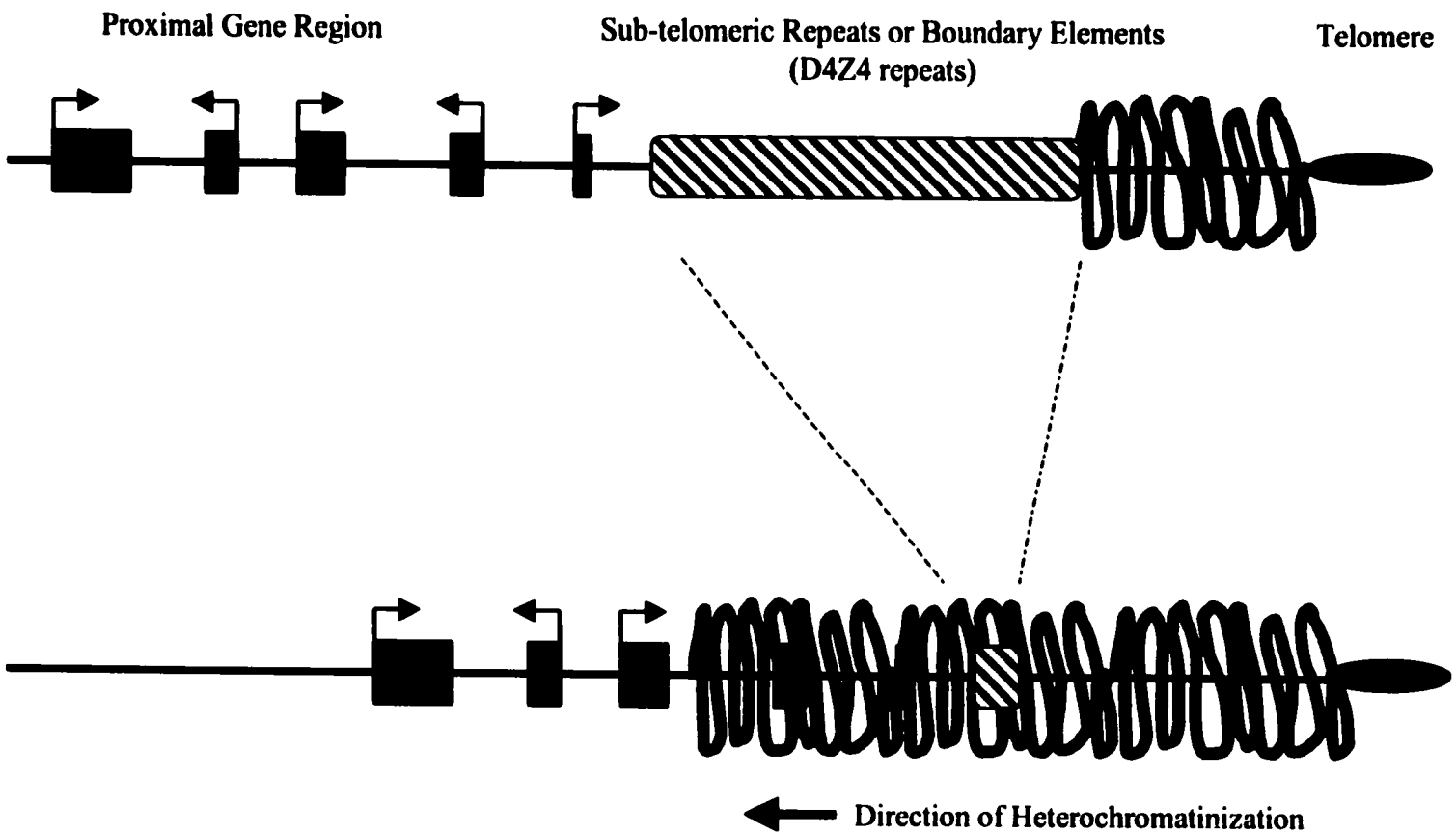

D4Z4 Repeats
centromere
Unique Chromosome Sequence
Proximal Subtelomeric Sequence
Distal Subtelomeric Sequence
Telomeric Repeats (TTAGGG)n
Degenerate (TTAGGG)n and ARS

very ambiguous, containing similar repetitive elements and both constitutive and facultative heterochromatin [101, 104, 106]. Expression of genes located near such heterochromatin centres tends to be regulated by long range position effects called centromere position effect (CPE) or telomere position effect (TPE) [125, 132, 133]. Sometimes these genes can lie up to 900kb away from the mutation breakpoint, such as in the case of X-linked deafness, discussed below [116]. The process by which the heterochromatin from the centromeres or telomeres is able to spread out and engulf an euchromatic gene is called heterochromatinization (Refer to Figure 4). Heterochromatinization occurs when extensive chromatin packaging can transcriptionally silence a gene by (i.) inducing a later replication time point which will put the gene at a disadvantage in the competition for transcription factors or (ii.) by generating a more tightly condensed structure of chromatin which itself inhibits transcription [86, 92, 134-136].

There are some examples of genetic diseases that are caused by position effect through heterochromatinization associated with constitutive chromatin (CPE or TPE). For example, the deletion of an 8kb interstitial fragment approximately 900kb away from the gene *POU*3F4 is the possible cause of a spread of heterochromatin shutting down expression of this gene and causing a form of X-linked deafness (DNF3; [116, 137]). It is possible that this specific 8kb region contains a buffer or boundary elements, which would normally prevent the spread of heterochromatin from affecting the expression of *POU*3F4. A second example of a disease causing mutation involving the spread of constitutive chromatin in humans is XY female sex reversal. It has been shown that partial deletions within the array of 2.6kb Sx1 repeats causes the spread of centromeric heterochromatin to affect the level of expression of the testis-determining gene *SRY* (sex-determining region Y

Figure 4. Cis-spreading heterochromatinization model for Facioscapulohumeral Muscular Dystrophy.

One model for FSHD suggests that the D4Z4 array (represented by a striped box) normally functions as a Sub-Telomeric Anti-silencing Region (STAR) to protect proximal genes from inappropriate silencing due to telomere position effect (top figure). In the cases with a large deletion of D4Z4 repeats (FSHD), the remaining STAR elements are insufficient to retain a protective boundary from cis-inactivation (bottom figure). Dark gray boxes with arrows represent active genes within the proximal region, gray boxes without arrows represent inactive genes. The light gray oval depicts the placement of telomeres and the thick dark line represents the DNA condensation effect of heterochromatinization. (Figure adapted from [31]).


Proximal Gene Region
Sub-telomeric Repeats or Boundary Elements
(D4Z4 repeats)
Telomere
Direction of Heterochromatinization

chromosome) [138, 139]. Experiments in mice have shown that deletions within the Sx1 array to below 50 repeating units bring the *Sry* gene from 35kb away to approximately 10kb from the Y centromere [139-141]. It has been suggested that these Sx1 repeats may provide a buffer or boundary to the heterochromatinization of the *Sry* locus [116, 139-142]. In yeast, proximity of a gene to the telomeres has been shown to repress expression as well as slow replication through TPE [133, 143]. Also, there are several examples how sub-telomeric chromosomal rearrangements and chromosome-end truncations cause human idiopathic mental retardation (IMR) syndromes. It has been suggested that 6-8% of IMR cases are the result of such telomeric position events [44, 125, 144, 145]. A consequence of the DNA rearrangement could be that the genes normally located hundreds of kilobases away from the telomere become located adjacent to the TTAGGG repeats. These repeats may then be able to initiate heterochromatin spread thus affecting gene expression.

Do the D4Z4 repeats act as a sub-telomeric anti-silencing region (STAR)? If so, then a deletion within the repeat array to below the threshold of 10 units may disrupt the functional organization of the 4q telomeric region [2, 15, 65, 86, 146]. If the repeats do indeed work as a telomeric spacer, it is hypothesized that the disease results from the inability of the smaller D4Z4 arrays to prevent cis-telomeric heterochromatinization from entering into the gene-rich proximal region and not a result of the removal of a D4Z4 gene through deletion. Under the premise of an autosomal dominant pattern of inheritance for FSHD only one chromosome 4 is necessary to contain the deleted D4Z4 allele. This telomere position effect silencing model therefore invokes a haploinsufficiency pathology to explain the observed inheritance. The region immediately proximal to the D4Z4 repeats then became a focal point for the study for the discovery of the novel FSHD disease gene.

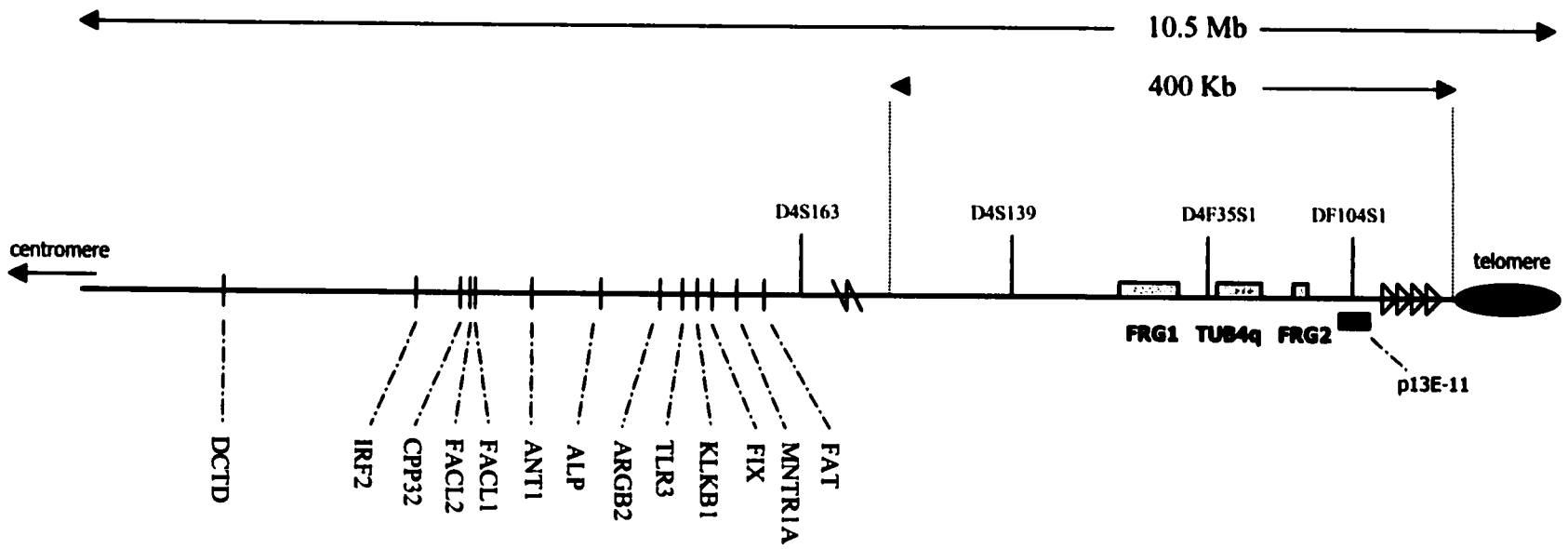
This search initially characterized two novel gene sequences, *FRG*1 (FSHD region gene 1) and *FRG*2 which are expressed genes with unknown functions and a pseudogene *TUB*4q [45, 49, 81, 86]. *FRG*1 is located 100kb proximal to the D4Z4 repeats and is abundantly expressed in most human tissues, including skeletal muscle; however *FRG*2 (37kb proximal to the repeats) has only been found to be expressed temporally in human and murine fibroblast cells undergoing forced myogenesis [45, 49, 86]. *FRG*2 expression was not detected in FSHD patient biopsies [49, 81]. *TUB*4q, which is not expressed, lies 80kb away from the D4Z4 repeats [49]. The hypothesis that the repeats act as, or contain STAR elements is consistent with several observations. Firstly, the variability in severity of phenotypes, which has been linked to the larger deletions (shorter *Eco*RI fragments) suggests that the extent of heterochromatinization can be cumulative based on the number of repeats. Secondly, in this model several proximal genes may be affected giving rise to numerous symptoms. There are other noteworthy genes which lie in the proximal region to the D4Z4 repeats. They include the apoptotic enzyme Caspase 3 (*CPP*32) and the skeletal muscle Adenine Nucleotide Translocator-1 (*ANT*1), which is important for muscle energy dynamics [18, 20, 27]. Despite the implications, mutations in the expression of these genes may have for muscle cell dynamics, they are approximately 5.5 Mb away from the D4Z4 repeats and as such are not considered the cause of FSHD. For an overview of the 4q35 genomic structure refer to Figure 5.

If the D4Z4 repeats at 4q35 act as a STAR for the prevention of TPE, then the resulting deletion identified in FSHD patients would, in effect, down-regulate expression of genes lying immediately centromeric and thereby representing a "loss of function" mutation. Following exclusion of a direct mutation within the coding sequence of *FRG*1, an

Figure 5. Physical map of the human chromosome 4q35 region.

Chromosome 4q35 spans approximately 10.5Mb. The genes *FRG*1, *FRG*2 and the pseudogene *TUB*4q which lie immediately proximal to the D4Z4 array (white triangles) are represented as light gray boxes. Other proximal genes of interest located beyond 400kb from the repeats are shown (to scale). The placement of the probe p13E-11 (black box) and specific DNA markers loci (D4S163, D4S139, D4F35S1 and D4F104S1) are labeled above.

Chromosome 4q35



allele-specific single nucleotide polymorphism was used to differentially assess *FRG*1 expression from both chromosomes of an FSHD patient [86]. Van Deutokom *et al.* [86] discovered that there was a slight increase in FRG1 expression from the D4Z4 deleted chromosome 4 over the normal allele. Furthermore, there are indications that other genes, such as *ANT*1 and *CPP*32, which lie within the D4Z4 proximal region, are upregulated in FSHD patients [18, 27]. This is surprising since this model would predict that proximal genes juxtaposed to the D4Z4 region would be silenced, and as such TPE cannot account for FSHD. Moreover, the report of a family monosomic for the 4q35-qtel region yet phenotypically normal with respect to FHSD suggests that haploinsufficiency of distal 4q35 genes cannot account for this disease [147]. The noticeable upregulation of the most distal 4q35 genes (*FRG*1, *CPP*32 and *ANT*1) in FSHD patients suggests that the D4Z4 repeat may normally have a mechanistic effect in repression of transcriptional activity within this genomic interval [18, 27, 86].

1.3.2 – Chromatin Repression and FSHD (Facultative Heterochromatin).

It has been said that the human genome is like a desert, with vast tracks of lifeless space interrupted only by the occasional oasis of activity [97]. Under this model it is the heterochromatin that represents the lifeless space by repressing gene activity. As previously noted constitutive heterochromatin is concentrated at the centromere and telomeres while facultative heterochromatin can form anywhere along the chromosome [103, 105, 124]. A fundamental difference between constitutive and facultative heterochromatin is that

constitutive heterochromatin tends to extend the reach of a centromere or telomere over a larger region, while facultative heterochromatin is more prone to affect specific and targeted loci [94, 102, 148]. A well studied example of facultative heterochromatin is X-inactivation, where the cis-acting Xist gene coordinated the silencing of nearly all genes on one of the two X chromosomes in female mammals (reviewed in [102, 105, 148, 149]. Although some genes do lie in regions immediately proximal to constitutive heterochromatin, the vast majority of genes lie within euchromatin regions [100, 106]. Many processes govern a gene's transcriptional status. Given the dynamic nature of chromosomes, it is unlikely that a gene would remain in a fixed active or inactive state [89, 120]. Moreover, during development, subsets of genes are activated in a temporally and spatially restricted manner [77, 150-152]. In order for cell specification and development or differentiation to occur, there has to be a critically precise mechanism(s) in place to allow some gene loci to have constitutive or inducible activation, whereas others remain silent or become repressed. One modulator of these processes is the dynamic organization of chromatin. The presence of heterochromatin restricts the access of transcriptional machinery and other regulating elements to the DNA [97, 108, 153]. There must therefore be mechanisms in place, which form heterochromatin over regions destined to be repressed, or disband heterochromatin at other intervals. These mechanisms also fall under the wide spectrum of Position Effect Variegation (PEV). There are a variety of processes at play during PEV: histone modification, ATP-dependant nucleosome remodeling, DNA methylation and protein structures [154]. In an attempt to elucidate the proteins responsible for such control, several genetic screens were carried out in *Drosophila*, isolating both suppressors and enhancers of variegation (Su(var) and E(var); [90, 115]). Among those identified were members of the structurally diverse Polycomb and Trithorax group (PcG and

TrxG) proteins along with other structural proteins [88, 106, 155, 156]. PcG proteins form complexes which functionally maintain transcriptional repression, while the TrxG proteins play a critical role in maintaining, but not initiating, a transcriptionally active state, through the modification of the local chromatin environment [120, 151, 157-160].

In higher eukaryotes, repetitive DNA arrays are often targeted for facultative heterochromatin formation. It has also become clear that transgene arrays are susceptible to this form of repression as well [161-166]. Mechanisms for the formation of facultative heterochromatin have been observed in all eukaryotes and frequently involve the inclusion of Heterochromatin Protein 1 (*HP*1 or the Drosophila Su(var)2-5) in a dosage dependant manner and for some genes the incidence of Polycomb Response Elements (PRE; [88, 92, 167, 168]. PREs have been shown through biochemical and genetic experiments to be the finite target sites for assembly of PcG complexes [91, 107, 120, 167]. PREs are sequences that act like a switch to control whether a gene is transcriptionally active or inactive and are often contained within regulatory elements necessary for gene transcription [152, 158, 169, 170]. PcG mediated repression is a balance between the activity of a PRE and the chromosomal context of the region [119, 120, 169]. The PcG and TrxG form a primary means of control for the regulation of pattern specific regulation of homeotic genes, such as those from the HOX family [171-175]. Mutations in key Polycomb genes result in homeotic transformations and severe developmental mutations especially in the body plan of higher eukaryotes [176-180]. Mutations in some PcG genes, such as B*mi*-1 have also been shown to cause severe proliferative effects in lymphoid cells, which induce the formation of lymphomas [181].

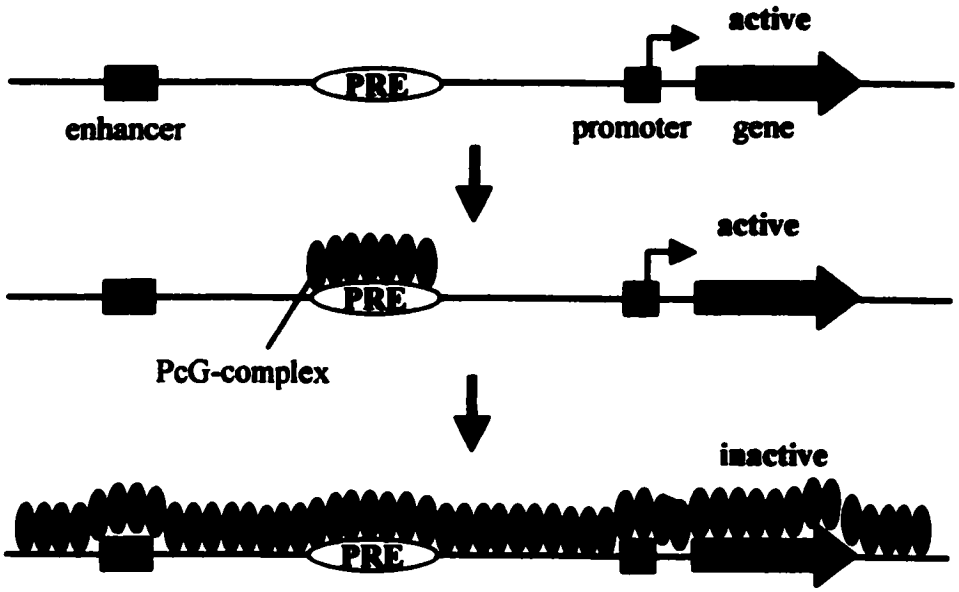
Much of our knowledge of the Polycomb activities comes from research in *Drosophila*; however PREs and both the PcG and TrxG mechanisms have been identified in vertebrates [158, 159, 171, 175, 176, 181, 182]. Out of the estimated 30-40 loci, there are greater than 13 known members of *Drosophila* PcG, all with known mammalian homologues (reviewed in [119, 132, 152, 175]). It has been shown that PcG proteins work together within complexes comprised of different proteins, each member having a specific role in repression [183-186]. These PcG repressive complexes are assembled as a chain of recruitment; stability is dependant on concentration of components and the permisiveness of the chromatin environment [177, 185, 187]. It is not yet clear which proteins are responsible for PcG complex recruitment to the PREs [173, 188, 189]. However there are two candiates, the drosphila *pleiohomeotic* (*Pho*) and the GAGA factor. *Pho*, and its human homologue *YY1*, is the only PcG member thus far which has been shown to have DNA binding properties with a minimal consensus sequence GCCATTAC [174, 180, 186, 188, 190]. However, not all PREs contain these binding domains [170, 188]. A second recruitment element has been suggested, the GAGA factor (also *Trl* for Trithorax-like) is a member of the TrxG family, but has been linked to both activation and repression, in both cases as a recruitment factor [108, 189, 191-193]. Researchers have found that GAGA factor binds to sites which overlap PREs and contain a high proportion of repeating GA residues [191, 194, 195]. Furthermore GAGA has been inextricably linked to PcG complexes through its ability to be co-immunoprecipitated with Pc [170, 189]. The PRE can thus be thought of as a mosaic of multiple protein binding sites for the recruitment PcG complexes [170].

There are two basic mechanisms proposed for PcG repression (see Figure 6). Based on the evidence of homolgy between HP1 and the PcG protein *polycomb* (*Pc*), it has been suggested that Polycomb repression acts at the chromatin level by forming multimeric repressive complexes over an entire domain as shown in Figure 6(A), [117, 168, 182]. The result is that a broad region undergoes heterochromatinization, similar to that observed in TPE and CPE [117, 151, 196]. In this model, the PcG complex nucleates from the PRE and then spreads by further PcG-PcG and PcG-chromatin interactions to cover a region, which may contain enhancer and transcriptional elements, if not the gene itself [156, 171]. However, through crosslinking experiments Strutt et al., [189] found that the PcG proteins do spread out from a PRE, but only over 1-2kb, and so this model alone cannot account for the widespread transcriptional repression associated with a PRE [119]. The second proposed mechanism (Figure 6(B)) is a looping model that was proposed by Pirrotta and Rastelli [119]. This model suggests that there are two forms of PREs, a strong or central core PRE that has a high affinity for the formation of stable PcG complexes, and outlying secodary sites or "S" elements which have less affinity and may only bind weaker less stable PcG complexes. The centrally bound PcG complex would interact with the S-associated complexes thus drawing the DNA into a looping bundle which is in sharp contrast to the current model for cis-telomeric and cis-centromeric spreading (reviewed in [119, 122, 156, 187]). Some evidence which supports the "looping" model came from Schlossher et al., [197]. They found that while transcriptional activity was repressed, PcG complexes had no significant effect on the availability of the DNA for restriction endonuclease digest. This suggested that only certain regions were actually bundled up and protected from restriction digest by protein.

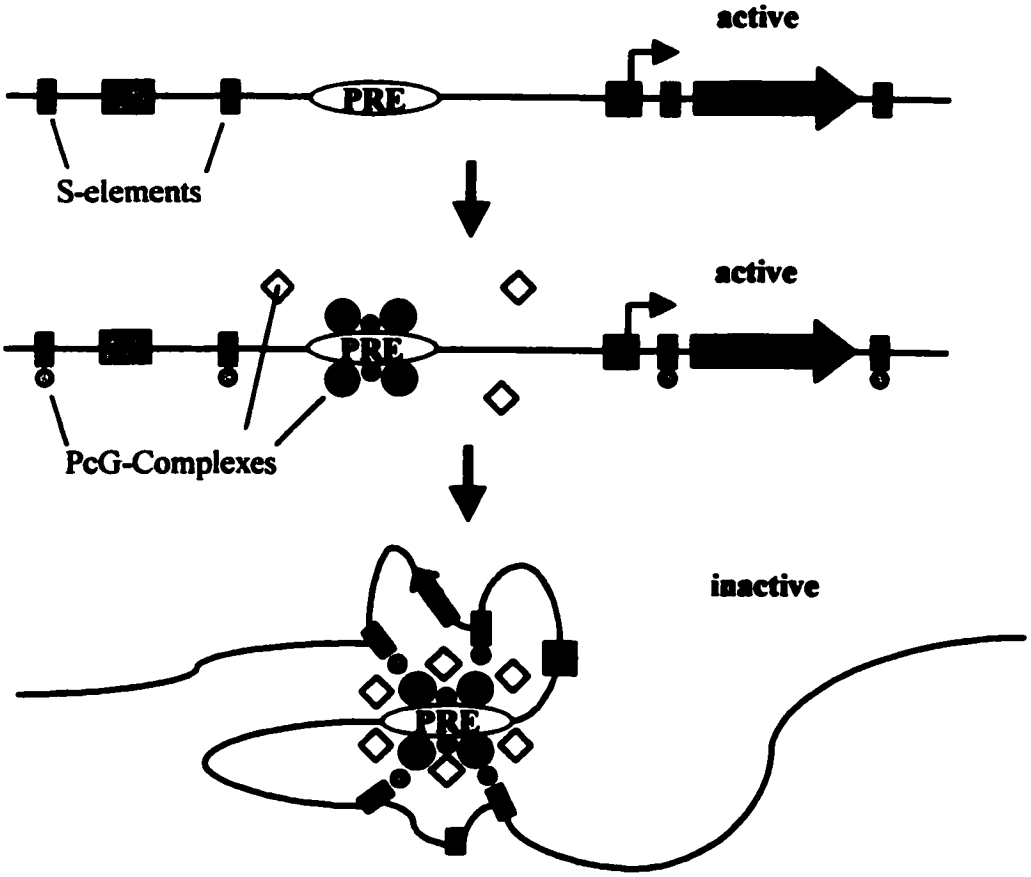
Figure 6. Possible mechanisms which lead to Polycomb Group-mediated repression.

(A) In the "spreading model", initiation of repression occurs as PcG complexes (green ovals) bind and nucleate out from the PRE (white oval). In a mechanism reminiscent of the Centromeric and Telomeric Position Effect cis-spreading heterochromatinization model the repressive PcG complexes build up through and silence gene expression by expanding to cover an entire coding unit (gene, promoter and enhancer elements, shown as blue (arrow), black and orange rectangles, respectively). (B) Another view of PcG repression is the "looping model." In this scenario a central PRE binds PcG complexes with high affinity, while outlying S elements, in purple are less efficient PREs and bind PcG complexes (blue/green circle and yellow diamonds) with lower affinity. The repressive looping structure is then formed as a cooperative action between the S-bound and central PRE-bound complexes, drawing the DNA into a bundle. Both models result in a tightly bound heterochromatin structure, which results in transcriptional inactivation.

(A)


active
PRE
enhancer
promoter
gene
active
PRE
PcG-complex
inactive
PRE

(B)


active
PRE
S-elements
active
PRE
PcG-Complexes
inactive
PRE

The D4Z4 repeat contains repetitive elements such as the hhsmp3 low copy and LSau middle copy type repeats which tend to be found in heterochromatic regions [2, 35, 39, 40, 43, 51, 65]. Furthermore, multiple copies of the 68 base pair Sau3 satellites, which are often associated with heterochromatin, are located just outside of the repeats [43]. It is hard not to imagine that the D4Z4 array is typically contained within a heterochromatin domain. Since Polycomb-group regulation is often associated with Homeotic gene regulation and the D4Z4 repeat potentially contains such a gene (*DUX*4), we examined the repeat sequence for Polycomb Response Elements. Within each repeat we identified seven consensus PRE sequences. It may be that these putative PRE elements form a repressive structure, which normally suppresses the region at 4q35. Under this model we postulate that the D4Z4 repeats represent a site of nucleation for the formation of a heterochromatin complex. Perhaps a larger array of repeats is able to form a more stable repressive structure, and that the long-range mechanisms consistent with position effect variegation is the result. The inherent variability of symptoms between patients and even monozygotic twins is indicative of the stochastic and mosaic effect of PEV [21, 22]. This suggests that there is a critical threshold of repeats, under which this repressive structure is unable to form or is ineffective in the regulation of neighbouring gene expression. We hypothesize that deletion within the D4Z4 array leads to continuous or misappropriate expression of one or more genes from the vicinity, within skeletal muscle, and is the underlying mechanism that causes the Facioscapulohumeral Muscular Dystrophy disorder. Furthermore, several of the genes from the 4q35 region have functions that may account for several of the FSHD symptoms, observed in severe cases. For example, *DFNA*24 has been associated with a non-progressive autosomal dominant form of hearing loss; *ANT*1 is involved in muscle cell mitochondrial energy metabolism and over expression of *CPP*32 may lead to muscle cell

deterioration by apoptosis [14, 18, 20, 27]. This suggests that FSHD is a syndrome composed of many separate and distinct phenotypes as a result of inappropriate transcriptional regulation of many loci.

There are other examples of human genetic diseases that are caused by a failure to control transcription through chromatin structure. The Immunodeficiency, Centromeric instability and Facial abnormalities (ICF) syndrome is an extremely rare autosomal recessive disorder caused by mutation in the DNA methyltransferase gene *DNMT*3B [198-200]. Mutations of this gene lead to the wide-spread hypomethylation of satellite DNA resulting in instability of the pericentromeric heterochromatin of chromosomes 1, 9, 16 and X [198-201]. At these regions the heterochromatin is undercondensed, and as a result prone to varying degrees of stretching and breakage [67, 202]. It has also been suggested that key genes in these regions may be inappropriately upregulated causing the craniofacial, cerebral and immunological development [51, 198]. Coincidentally, one of the two non-satellite DNA repeats found to be strongly hypomethylated in ICF patients was the D4Z4 repeat, implicated in FSHD [67, 87].

1.4 – Rationale and Research Objectives.

Facioscapulohumeral Muscular Dystrophy is a debilitating genetic disorder that affects approximately 1:20,000 individuals. The incidence of FSHD has been irrefutably linked to internal deletions of D4Z4 repeats at 4q35; however, over the last decade the role of these repeats in the pathogenesis of this disease has not been elucidated. Our working hypothesis is that a reduction in the number of these repeats below a critical threshold

prevents the formation of a repressive heterochromatin complex. This inability to form facultative heterochromatin complex affects the genomic environment of the surrounding region, and ultimately alters local gene expression. The purpose of our experiments is to determine the function these repeats hold in the pathogenesis of FSHD by evaluating their regulatory effect on reporter gene expression. There has been much speculation as to the repressive nature of the D4Z4 repeats, yet few functional tests have been undertaken to define the role of D4Z4 in the pathogenesis of FSHD. It is therefore imperative to define the molecular mechanism underlying this disease. To address this potential mechanism, we have examined the expression of a reporter gene in plasmids containing 1 to 4 concatamerized copies of the D4Z4 repeat. Our aim was to determine if D4Z4 repeats had a repressive effect on neighbouring genes, muscle function, or myoblast differentiation. To address this we proposed the following objectives:

[1] To construct the LacZ expression plasmids containing cloned arrays of D4Z4 repeats.

[2] To investigate and measure the effect that D4Z4 repeats have on LacZ gene expression and to compare it to LacZ expression from a construct without D4Z4 repeats in transient transfection assays using C2C12 myoblasts.

[3] To determine if the D4Z4 repeats are able to initiate gene repression following stable integration of plasmid constructs into the C2C12 genome

[4] To assess if myogenic differentiation plays a critical role in D4Z4 repeat-induced regulation of gene expression.

Such information will provide additional insight into the role that the D4Z4 repeats play in the pathogeneis of FSHD. This in turn will provide valuable information, which would broaden our understanding of how chromatin structure regulates human genetic diseases.

2.0 – Materials and Methods

2.1 – Materials

2.1.1 – Sources of Chemicals and Reagents

Agarose, proteinase K, Lipofectamine Reagent and DNA molecular weight ladders (1 kb plus Ladder) were purchased from Gibco BRL (Burlington ON). ScintiSafe scintillation fluid, sarkosyl, EDTA and dextran sulfate, were purchased from Fisher Scientific (Toronto ON). Formamide and lysozyme were obtained from Roche Diagnostics Corp. (Indianapolis IN). Ethidium bromide, placental DNA, carbenicillin, DEPC, X-gal, ONPG, 3,3'-diaminobenzidine, guanidium thiocyanate, ferricyanide, ferrocyanide, glutaraldehyde and β-mercaptoethanol were purchased from Sigma Chemical Co. (St. Louis MO). Phenol, bacto-agar, bacto-tryptone and yeast extract, chloroform and isoamyl alcohol were purchased from VWR International (Ville Mont-Royal QB). Reporter Lysis Buffer and purified β-gal protein standard were obtained from Promega (Madision WI). Stock dNTPs, Sephadex G-50, Kodak Biomax MR-1 film and Random Hexamers (used in cDNA synthesis) were purchased from Amersham Pharmacia Biotech Inc. (Baie d'Urfe, QB). All cell culture plates, dishes and flasks are from Falcon (Franklin Lakes NJ) or Nunc (Denmark). Cell culture media was purchased from Gibco BRL and cell culture serum (fetal bovine serum) was purchased from Cansera International, (Rexdale ON). All other common laboratory chemicals were obtained from Sigma, Fisher, Gibco BRL or VWR.

2.1.2 – Enzymes

Restriction endonucleases *Asp*718 and *Tru*91 (Roche Diagnostic), *Hind*II and *Spe*I (New England Biolabs Ltd, Mississauga ON), *Kpn*I, *Bgl*II, *Nhe*I, *Nde*I, *Eco*RI and *Eco*RV (MBI Fermentas, Burlington ON) were used under the conditions provided by the supplier. Taq DNA Polymerase used in PCR reactions was purchased from Invitrogen. Superscript Reverse Transcriptase was purchased from Gibco BRL. The 5' and 3' end modifying enzymes Klenow DNA Polymerase I and T4 DNA Polymerase were purchased from New England Biolabs. RNaseA and Lysozyme were from Roche Diagnostics, Proteinase K was from Gibco BRL and RNA guard, a ribonuclease inhibitor, was purchased from Amersham Pharmacia Biotech Inc.

2.1.3. – Radiochemicals

All radiolabeled [α-^{32}P]dCTP isotopes (10μCi/μL) were purchased from Amersham Pharmacia Biotech Inc. (Baie d'Urfe, QB).

2.1.4 – Oligonucleotides

Unless otherwise stated, all oligonucleotides used in experiments for stable clone confirmation or reverse transcription-PCR were synthesized by Invitrogen Custom Primers (Frederick MD), Gibco BRL Custom Primers (Ottawa ON) or Cortec DNA Service Laboratories Inc. (Kingston, ON). The sequences of the oligonucleotides used are listed below.

LacZ Reporter Gene Specific (cDNA) (Designed by R. Kothary, Ottawa ONT)

RAS 39: 5'- ttg gcc tga act gcc agc tgg cgc agg - 3'

RAS 40: 5'- tcc cgc agc gca gac cgt ttt cgc tcg - 3'

D4Z4 Repeat Specific

bD4Z4-3026f: 5'- gga agc acc cct cag cg - 3'

bD4Z4-3219r: 5'- gaa ggc agg aat ccc agg c - 3'

D4Z4-f1905: 5'- cga gcc tgc ttt gag cg - 3'

D4Z4-r2385: 5'- cgt ggg gtg cgg gaa - 3'

Sequencing Primers

T3: 5' - att aac cct cac taa ag - 3'

T7: 5' - gat atc act cag cat aa - 3'

2.2 – Methods

2.2.1 – Plasmid Constructs

To evaluate the potential effect D4Z4 has on promoter activity, a complete D4Z4 repeat was cloned from the cosmid cY13 (gift from RR Frants, Leiden, The Netherlands).

The cosmid cY13, which contains three D4Z4 repeats, was previously derived from chromosome 4q35-specific YAC 25C2E [203]. DNA from cY13 was purified and digested with *Kpn*I to release the D4Z4 repeats. The 3.3 kb band corresponding to the D4Z4 repeat was agarose gel purified using the GFX PCR DNA and Gel Band Purification Kit (Amersham Pharmacia Biotech Inc.). D4Z4 repeats were then concatamerized and ligated into pBluescript II KS +/- phagemid (Stratagene, La Jolla CA) using the Rapid DNA Ligation Kit (Roche Diagnostics Corp.), with modifications. Briefly, D4Z4 fragments (200ng) were allowed to self-ligate for 20 min to form concatamers and then 50ng of *Kpn*I digested pBluescript was added to the mixture for 30 min at room temperature. In this method D4Z4 repeats in arrays of 1-4 were obtained and determined to be in uniform direction through DNA restriction endonuclease digestion with *Bgl*II and agarose gel electrophoresis. *Tru*91 digestion followed by treatment with 2U of Klenow DNA Polymerase I at room temperature for 20 minutes to fill recessed 3' termini [204] was used to release concatamerized repeat arrays from these intermediate plasmids (pDYInt01$^{1-4}$). Following gel purification of the appropriately sized bands, the arrays were further digested with *Spe*I and cloned into *Eco*RV / *Spe*I double digested pBluescript. This additional cloning step generated two *Hind*III restriction sites flanking the cloned D4Z4 repeat arrays and corresponds to pDYInt02$^{1-4}$. The promoter expression construct (pDYInt04) was developed by use of a band containing the MCS and flanking LoxP recognition sequences (Ac. AX359675) was released from pBS246 (gift of M. McBurney, Ottawa ON) by *Eco*RI digestion. Five μg of the resulting *Eco*RI fragment was then treated with 2U Klenow DNA Polymerase I, to fill recessed 3' termini then digested further with *Spe*I. Following gel purification, this fragment was cloned into pBluescript, which had been digested by *Kpn*I

and end-treated with 2U (for 5μg DNA) of T4 DNA Polymerase for 20 minutes at room temperature to remove 5' protruding ends [204], then digested with *Spe*I. Following construction of pDYInt03, the Eflα (BOS) promoter / LacZ reporter cassette was released by successive *Kpn*I digestion, blunted by T4 DNA Polymerase and then *Not*I digestion from pRP2044 (gift of R Parks, Ottawa ON). This fragment was cloned 93 bp downstream of the second LoxP sequence of pDYInt03 by blunt-sticky ligation to Klenow DNA Polymerase I treated *Spe*I and *Not*I sites, forming pDYInt04. D4Z4 arrays were then gel purified with *Hind*III digestion from pDYInt02[1-4] and cloned by Rapid DNA Ligation into the BOS promoter expression construct (pDYInt04) to generate the final constructs pDY01, pDY02, pDY03 and pDY04. The plasmid pRP2044, which does not have any D4Z4 repeats, was used as control. For an overview of the cloning strategy refer to Appendix A. Sequence analysis, using T3 and T7 primers confirmed that the inserts were D4Z4 repeats and confirmed the validity of the pDY01-pDY04 experimental constructs (Ottawa Regional Cancer Center Sequence Facilities, Ottawa ON). The D4Z4 repeats were only obtained in one orientation.

2.2.2 – Bacterial Transformations

Liquid media (SOB, SOC and LB) and solid media (LB-agar) were prepared as described [204]. Following ligation, plasmid DNA was used to transform Competent *E. coli* cells. Competent Stbl2 cells (Life Technologies Inc. Burlington ON) were prepared as described in [205]. Briefly, 10-15 colonies of Stbl2 cells that had grown overnight at 37°C on an LB-agar plate were inoculated into 250mL of SOB media and grown to an optical

density (A_{600}) of 0.6 at room temperature with shaking at approximately 200-250 rpm in the dark. The cultured cells were then pelleted and resuspended in ice cold TB media (10mM Pipes, 55mM $MnCl_2$, 15mM $CaCl_2$ and 250mM KCl) for 10 min. Cells were then pelleted and resuspended in 20mL of TB and DMSO (final concentration of 7%). Following a 10 min incubation on ice, 500μL aliquots were cryogenically frozen and stored at -80°C until use. Transformation was performed as follows: an aliquot of the ligation mix (3-7μL) was added to 150μL of thawed competent cells and incubated for 20 minutes on ice in 15mL polypropylene tubes. The mixture was then heat pulsed for 45s at 42°C without agitation and transferred to an ice bath for 2 additional minutes. One milliliter of LB media was then added and the tube(s) were placed in a 37°C shaking incubator for a 1 hour recovery. Following recovery a desired portion (ranging from 75- 150μL) of the mixture was spread onto LB-agar plates containing 20μg/mL, a synthetic derivative of Ampicillin and grown overnight at 37°C. Individual colonies were chosen for growth in 4mL overnight cultures at 37°C with shaking (250rpm) using LB media containing 20μg/mL Carbenicillin. Plasmid DNA was isolated in order to screen for positive clones (see below).

2.2.3 – Plasmid DNA Preparation

All plasmid DNA was isolated using a Boil Lysis mini prep protocol [204] or by using the GFX Micro Plasmid Prep Kit (Amersham Pharmacia Biotech. Inc.) according to the manufacturer's suggested protocol. Large-scale plasmid DNA preps of positive plasmid clones were performed using Qiagen Midi and Maxi Prep Kits (Valenci CA) according to the manufacturer's protocol. All DNA was resuspended in 35-50μL (Mini) or 80-100μL

(Midi/Maxi) of water or TE (10mM Tris, HCL, 1mM EDTA, pH 8.0) containing 10μg/mL RNAse A (Roche Diagnostics). Absorbance at 260nm (A_{260nm}) in a Beckman Coulter DU 640 spectrophotometer (Fullerton CA) was used to determine the DNA concentration of each sample. Quality of the DNA sample was assessed as the DNA purity index, and was determined as the absorbance at 260nm divided by absorbance at 280nm (A_{260nm}/A_{280nm}) [201]. Typically only plasmid DNA having a purity index of greater than or equal to 1.7 was used.

2.2.4 – Genomic DNA Isolation from Mammalian Cells

Total genomic DNA was extracted from pelleted cells using a tail lysis buffer (50mM Tris-Cl (pH 8.0), 10mM EDTA, 10mM NaCl, 1% SDS) containing 20μg/mL freshly added Proteinase K (Gibco BRL). Following 24-48 hours incubation in Lysis Buffer at 50°C, DNA was isolated from protein and cell debris by three consecutive phenol: chloroform:isoamyl alcohol (25:24:1) extractions the aqueous phase (containing DNA) was then separated from the organic phase (containing cell debris) by centrifugation at 4000 rpm for 15 min and saved; DNA was precipitated using 1/10 volume 3M sodium acetate and 3 volumes of 95% ethanol. Pelleted DNA was washed twice with 70% ethanol and resuspended in water containing 10μg/mL RNAseA (Roche Diagnostics). Absorbance at 260nm in a Beckman Coulter DU 640 spectrophotometer (Fullerton CA) was used to determine the DNA concentration of each sample. Quality of the DNA sample was assessed as the DNA purity index, and was determined as the absorbance at 260nm divided by

absorbance at 280nm [204]. Typically only DNA having a purity index of greater than or equal to 1.7 was used.

2.2.5 – Cell Culture

C2C12 myoblasts (American Type Culture Collection, Manassas VA) were maintained at subconfluency under conditions of 37°C with 5% CO_2 in growth media containing Dulbecco's Modified Eagle's Medium, [DMEM] (Gibco BRL) and 10% heat inactivated fetal bovine serum [FBS] (Cansera International). Cells were cryogenically preserved at -80°C in DMEM, 10% FBS and 10% DMSO. Stable cell clones were maintained at subconfluency under conditions of 37°C with 5% CO_2 in selection media containing DMEM, 10% FBS and 0.8mg/mL G418 (Gibco BRL) or cryogenically preserved at -80°C in DMEM, 10% FBS, 0.8mg/mL G418 and 10% DMSO. All cells were harvested for DNA, protein or RNA analysis by treatment with a minimal volume (1/10 volume of culture media) of 1X Trypsin-EDTA (Gibco BRL) for five minutes and resuspended in 50mM Tris-Cl (pH 7.6) following a wash with 1X phosphate buffered saline (PBS).

2.2.6 – Myoblast Differentiation

Cells were grown to 80-90% confluency in 6 well dishes or 35 mm dishes from Falcon (Franklin Lakes NJ) or Nunc (Denmark) before being induced to differentiate through mitogen deprivation by serum withdrawal. Differentiation media comprised DMEM with 2% heat inactivated horse serum (Sigma Chemical Co.) or for stable clones included 0.8mg/mL G418. Differentiation medium was changed every 1-2 days during the

experimentation period, lasting from 0-4 days. This method for C2C12 myoblast differentiation has been has been previously described [57, 58, 206].

2.2.7 – Adenovirus Infection

R. Parks (Ottawa, ON) has graciously provided us with a first generation adenovirus (AdMA26) which confers expression of Cre recombinase. For an overview of the virus construction refer to Anton and Graham, (1995) [207]; AdMA26 is a derivative of AdCre1 with a murine cytolamegalovirus (CMV) promoter rather than a human CMV promoter. For Cre recombinase treatment, monolayer cultures of stable clone cells (approximately 10^5 cells) in 35mm dishes were incubated with 200μL of PBS containing a concentration of 100 AdMA26 virus particles per cell, or moi (multiplicity of infection). Following a 1 hour incubation at 37°C with occasional agitation, cells were gently rinsed in PBS and returned to growth medium (DMEM with 10% FBS) for 24 hours after which cells were harvested for protein analysis (see below).

2.2.8 – Transfections of Mammalian Cells

For transient transfections, approximately 10^4 C2C12 myoblasts were seeded in six well plates or 35mm dishes overnight to 50-80% confluency. The following day each dish was transfected with equal molar ratios of experimental or control plasmid DNA using 2μL of Lipofectamine Reagent (Gibco BRL). Four hours following transfection, cells were transferred to DMEM media containing 10% FBS for recovery. Subsequent to a 48-hour

recovery period, when cultures were confluent, replicate plates were either harvested, induced to differentiate for various times (12, 24, 48, 72 or 96 hours), or assayed for transfection efficiency using β-gal *in situ* analysis.

To establish stable transfected clones, 10^4 C2C12 myoblasts were seeded overnight into 100mm dishes (Falcon, Franklin Lakes NJ). The following day cells were transfected with 2μg of experimental or control plasmid DNA, and 0.5μg of the neomycin resistance marker plasmid, pGTneo (New England Biolabs) using 8μL of Lipofectamine Reagent (Gibco BRL) as recommended by the manufacturer. Four hours following transfection, cells were cultured in DMEM media containing 10% FBS for recovery. After a 24 hour recovery period, cells were transferred to selection media containing DMEM, 10% FBS and 0.8mg/mL G418 (Gibco BRL) to select for stable integration of plasmid DNA. G418 resistant colonies were selected by scraping individual colonies from 100mm dishes with a pipette tip and then transplanted to a 12 well plate (single well per colony). Following successive expansion under growth conditions to larger dishes or cell culture flasks clones were screened for reporter expression (by *in situ* β-galactosidase Assay), PCR amplification of marker sequences and Southern hybridization for copy number and integration sites (see below).

2.2.9 – *In situ* β-galactosidase Assays

In situ assays with the 5-bromo-4-chloro-3-indoyl-β-D-galactoside [X-gal] (Roche Diagnostics) substrate are described in [111]. Cells were washed twice in PBS and fixed

with PBS containing 0.2% gluteraldehyde and 2% formaldehyde for 10 min at room temperature then rinsed with several times with PBS. The colour reactive substrate was then added (PBS containing 10mM X-Gal, 5mM potassium ferrocyanide, 5mM potassium ferricyanide and 1mM $MgCl_2$) and cells were incubated in the dark at room temperature in for 8-16 hours to allow for colour development. Stained cultures were rinsed several times and stored in PBS at 4°C, indefinitely.

2.2.10 – PCR Screening of Stable cells

Polymeraze chain reaction primers directed toward the D4Z4 repeat (bD4Z4-3026f and bD4Z4-3219r) and LacZ gene (RAS 39 and RAS 40) were used to determine whether the constructs had integrated into the C2C12 genome by screening genomic DNA isolated from stable clones. Conditions for 50µL amplification reactions using 0.05µg of experimental DNA or <0.025µg of plasmid DNA (positive controls), PCR buffer (20mM Tris-Cl (pH 8.4), 50mM KCl and 1.5mM $MgCL_2$), 0.2mM dNTPs (mix of dCTP, dATP, dGTP and dTTP), 0.5mM primers and Taq DNA polymerase (Invitrogen) were as follows: 95°C for 2 min, then 35 cycles of 94°C for 30s, 56°C (or 52°C for D4Z4 primers) for 30s, 72°C for 1 min, followed by 10 min at 72°C. The products from the PCR (5-10µL) were resolved and visualized on a 1.5% agarose gel with ethidium bromide staining.

2.2.11 – Southern Blot Hybridization Analysis

Transgene copy number was determined for a selection of clones using Southern blot hybridization and normalized to band intensity of the resulting autoradiographs. Genomic DNA (15μg) from C2C12 or stable clones was digested in a *Asp*718 (1 U/μg) and *Nde*I (1U/μg) at 37°C, separated on a 0.8% agarose gel. DNA fragments were then transferred by capillary action to a nylon membrane (GeneScreen Plus; NEN Life Science Products, Boston MA) as previously described [208]. An internal DNA probe directed towards a 2.3 kb *NdeI*/*Nhe*I fragment of the LacZ gene within each construct was labeled with [α-^{32}P]dCTP, using Rediprime II Random Prime Labeling System (Amersham Pharmacia Biotech Inc.). Prehybridization and hybridizations were performed in 50% formamide with 1X SSC, 1X Denhardt's solution, 2% SDS, 20ug/mL denatured salmon sperm and 6% dextran sulfate at 42°C. For hybridization, 1x10^6 counts per minute of denatured probe and 18μg of sheared and denatured human placental DNA (repetitive sequence competitor) was added to the solution and incubated overnight. Hybridized membranes were washed in 0.5X SSC for 30 min at room temperature and then in 0.1% SDS / 0.1X SSC for up to 1 hour at room temperature. Membranes were exposed to Kodak BioMax autoradiography film for 4-8 days, at -80°C. Intensity of the size specific band was calculated using the UVP–White/UV Transillumeter (Diamed, Mississauga ON). Normalized band intensities from autoradiographs was used to approximate relative copy numbers. To further correlate copy number a second digest was performed. *Nhe*I, which cuts within the 3' end of the expression cassette and *Eco*RV, which does not cut the construct, were used in a double digest to ascertain the boundary between transgenic insertion and genomic DNA. The number of bands is indicative of the number of random

integration sites within the genome. Relating the number of insertion sites to the intensity of the previous internal band allowed for the approximation of the relative copy number. To assess equal loading the membranes were stripped in 0.1% SDS at 65°C for 1 hour and rehybridized with a probe for the mouse *Snf*2h gene (gift from M. A. Lazzaro, Ottawa, ON).

2.2.12 – Antibodies and Immunoreactive Cell Staining

Following growth and differentiation (for 96 hours) of stable or transiently transfected C2C12 cultures, cells were washed thoroughly with ice cold PBS and fixed with 90% methanol and 10% PBS (-20°C) for 6 min on ice. After a final wash with PBS, the fixed cells were then blocked at room temperature for 1 hour in PBS plus 5% fat free dry milk (blocking buffer). The primary antibody MF20 (DSHB, Iowa City IA) is a monoclonal antibody raised against the mouse myosin heavy chain, an early marker for muscle differentiation. MF20 was used as a 1:50 dilution in blocking buffer and incubated for 1 hour at room temp. Following three washes with PBS (to cover), a polycolonal goat anti-mouse secondary antibody conjugated to horseradish peroxidase [HRP] (Gibco BRL) was added in a 1:1000 dilution in 5% milk plus 50mM Tris-Cl (pH 7.6). Following 3 more washes in PBS (to cover), the cells were stained with approximately 0.2mg 3,3'-Diaminobenzidine [DAB] (Sigma) in 50mM Tris-Cl (pH 7.6) plus 0.03% H_2O_2 for approximately 2 hours and washed several times with PBS. Cell nuclei were then counter-stained by rinsing cells with Harris modified Hematoxylin with acetic acid (Fisher Diagnostics, Fair Lawn NJ) for 2 min, followed by several washes using PBS. Plates containing stained cells were stored indefinitely at 4°C in PBS.

2.2.13 – Analysis of Myotube Fusion

The relative fusion competence of wild type C2C12, cultures of transiently transfected cells and stable clones induced to differentiate was assessed using the Myotube Fusion Index (MFI) [209]. MFI was calculated as the ratio of nuclei incorporated into myotubes from total nuclei within a field of view, using phase-contrast microscopy in the method described by Gorza and Vitadello, [210]. For the purposes of these experiments a myotube is defined as a multinucleated structure containing 3 or more nuclei and staining positive for MF20. Twenty or more fields of view per sample with 4 replicate plates per construct were used, with the exception of cRP2044n5, were only ten fields of view were scored. The MFI was calculated for C2C12 (control), pRP2044, pDY02 and pDY04 transiently transfected cells and the stable clones: cRP2044n5, cRP2044n7, cRP2044n31, cDY042, cDY04n2 and cDY04n/3.

Similarly, the Deformed Myotube Index (DMI) was calculated as the ratio of myotubes showing a deformed morphology from the total myotubes within a field of view. For the purpose of these experiments a deformed myotube structure is defined as a multinucleated structure containing 3 or more nuclei and staining positive for MF20 with a shape that deviates significantly from structures observed in wild type C2C12 myotubes cultures. Twenty or more fields of view per sample with 4 replicate plates per construct were used for this calculation. The DMI was calculated for C2C12 (control), pRP2044, pDY02 and pDY04 transiently transfected C2C12 cells and the stable clones: cRP2044n5, cRP2044n7, cRP2044n31, cDY042, cDY04n2 and cDY04n/3.

2.2.14 – RNA Isolation from Mammalian Cells

Approximately $2x10^5$ cells (two confluent 10cm dishes) from stable clones or C2C12 cells in either growth of differentiation media were harvested for RNA. Total RNA was isolated using a guanidium thiocyanate buffer (4M guanidinium thiocyanate, 20mM sodium acetate pH5.4, 0.5% sarkosyl) and phenol isolation protocol previously described [211]. Isolated RNA was then used for reverse transcription followed by PCR amplification.

2.2.15 – Protein Isolation and Analysis of β-gal Activity

Whole cell extracts were obtained using the Reporter Lysis Buffer and protocol (Promega), with modifications. Following experimental treatments as described above, cells were washed thoroughly with PBS and incubated in 300μL of 1X Reporter Lysis Buffer for 15 minutes, then scraped from the plates and collected. Samples were stored at -15°C until use, upon which they were freeze-thaw treated and centrifuged for 15 min at 12,000g to precipitate cell debris and genomic DNA. Aliquots of 25 μL were allocated for protein quantification through Bio-Rad DC Protein Assay (Bio-Rad Laboratories, Hercules CA) with spectrophotometric quantification (A_{750nm}) according to the manufacturer's recommendations. Excess protein samples were stored at -15°C, indefinitely. Equivalent protein samples (approximately 5ng) from each experimental condition were assayed for βgal activity using quantitative ONPG assay (o-nitrophenyl-β-D-galactopyranoside) (Sigma) adapted from [204]. Briefly, reactions were left to incubate at 37°C for approximately 15-20 min or until a control reaction containing approximately 5μg of purified β-gal standard (Promega) turned yellow, at which point all reactions were stopped

with 1M Na_2CO_3. Relative reporter activity was determined using spectrophotometric quantification (A_{420nm}). All reporter assays were normalized to transfection efficiency for transient experiments.

2.2.16 – Reverse Transcriptase-PCR (RT-PCR)

First strand cDNA synthesis was carried out on total RNA prepared from C2C12 and stable clone cells by the following method. For each sample, 5μg of total RNA and 300ng or random hexamers (Gibco BRL) were allowed to anneal by incubation at 65°C for 10 minutes (total volume 41μL with DEPC-treated H_2O). Synthesis of first strand cDNA was carried out by adding 5μL of 10X first-strand buffer (Gibco BRL), 1μL (40U/μL) RNA Guard (Amersham Pharmacia), 200mM dNTPs with 1μL or Superscript Reverse Transcriptase (Gibco BRL) at 42°C for 1 hour. Absorbance at 260nm (A_{260nm}) in a Beckman Coulter DU 640 spectrophotometer (Fullerton CA) was used to determine the DNA concentration of each sample. Quality of the DNA sample was assessed as the DNA purity index and was determined as the absorbance at 260nm divided by absorbance at 280nm [204]. 0.05μg of cDNA or negative control samples (prepared without Superscript RT) were used for RT-PCR experiments with the following primer pairs: D4Z4-f1905 / D4Z4-r2385 (experimental primers), RAS 39 / RAS 40 (internal control primers for LacZ) and β-Actin competimers (external control) (Ambion Inc. Austin, TX). Conditions for the 50μL reactions containing PCR buffer (20mM Tris-Cl (pH 8.4), 50mM KCl and 1.5mM $MgCL_2$), 0.2mM dNTPS, 0.5mM primers (each) and Taq DNA polymerase (Invitrogen) were as follows: 95°C for 2 min, then 30 cycles of 94°C for 30s, 56°C (or 52°C for D4Z4 primers)

for 30s, 72°C for 1 min, followed by 10 min at 72°C. Following PCR amplification 5-10 μL of the products were resolved on a 1.5% agarose gel and visualized with ethidium bromide staining. No amplification was observed in control (no RT) samples.

2.2.17 – Computational and Sequence Analysis

All genomic and cDNA sequence analysis (restriction sites and primer development) was performed using MacVector 6.5.3, Oxford Molecular Group plc (1999). Analysis of open reading frames (ORFs), minimal polycomb Response Elements [PREs] (Acc. L32205 and [188]) and GAGA factor binding sites [189, 191, 194] of the D4Z4 repeat (Acc. D38024) was performed using alignment tools in MacVector 6.5.3, Oxford Molecular Group plc (1999). The search for potential transcription factor consensus binding regions within the D4Z4 repeat was aided by MatInspector V2.2 available with Transfac 4.0 online (http://bioinformatics.weizmann. ac.il/transfac/). The genomic region of 4q35 was analyzed for proximal region gene structure using the UCSC human Genome Project Working Draft Browser, available online (Http://genome.uscs.edu/ndex.html) and the Online Mendelian Inheritance in Man (OMIM) Gene map (http://www.ncbi.nlm.nih.gov/htbin post/Omim/getmap). All primer pairs were checked against homologous sequence databases using the standard nucleotide-nucleotide Basic Local Alignment Search Tool [BLASTn] [212] available online (http://www.ncbi.nlm.nih.gov/BLAST/).

3.0 – RESULTS

3.1 – The generation of D4Z4 - LacZ expression constructs.

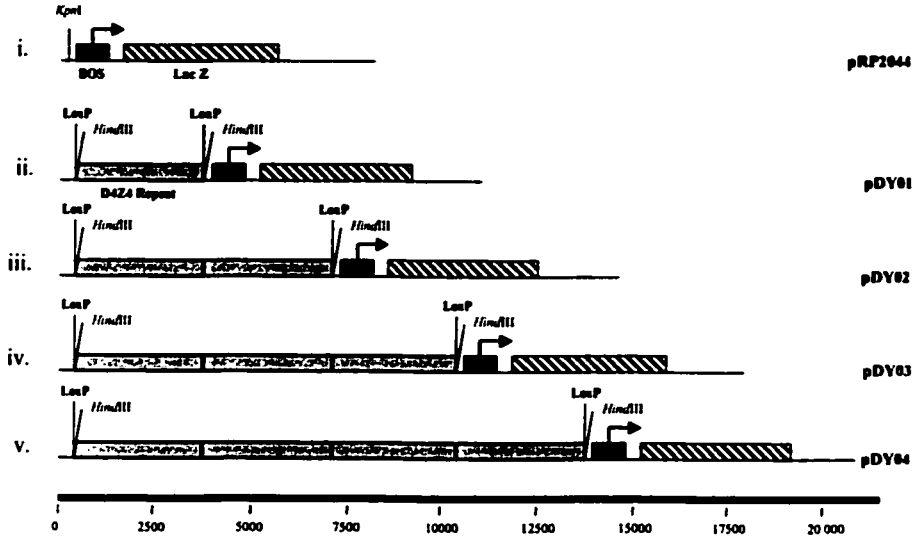
To assess the effect D4Z4 repeats have on gene expression, we have generated DNA expression constructs with varying numbers of the D4Z4 repeat upstream of a reporter gene cassette. A series of LacZ reporter constructs were generated, containing 1-4 copies of the 3.3 kb DZ4 repeats, flanked by LoxP recombination sites (see Figure 7(A)). The floxed D4Z4 arrays were cloned 93 bp upstream of the Eflα (BOS) promoter and the LacZ reporter gene. These plasmids are termed experimental constructs: pDY01 pDY02, pDY03 and pDY04 (Figure 7(A)). The control reporter construct pRP2044 (Figure 7(A)) from which the experimental constructs were derived was the gift of R. Parks (Ottawa ON). Sequencing of the cloned repeats confirmed them to be D4Z4 repeats (data not shown). To assess the number of repeats in each construct, a *Hind*III digest was performed to release the D4Z4 array. The number of repeats within each band was determined based on gel electrophoresis mobility in relation to the migration of molecular weight markers (Figure 7(B) and data not shown).

Based on the presence of putative repressive elements termed Polycomb Respones Elements (PRE; see 3.5.1 and Discussion), we hypothesized that the D4Z4 repeats are repressive in nature and that the introduction of D4Z4 upstream of the BOS ubiquitous promoter would repress expression of the LacZ reporter. To assess whether D4Z4 repeats repress LacZ transgene expression we transfected our plasmids into C2C12 cells an

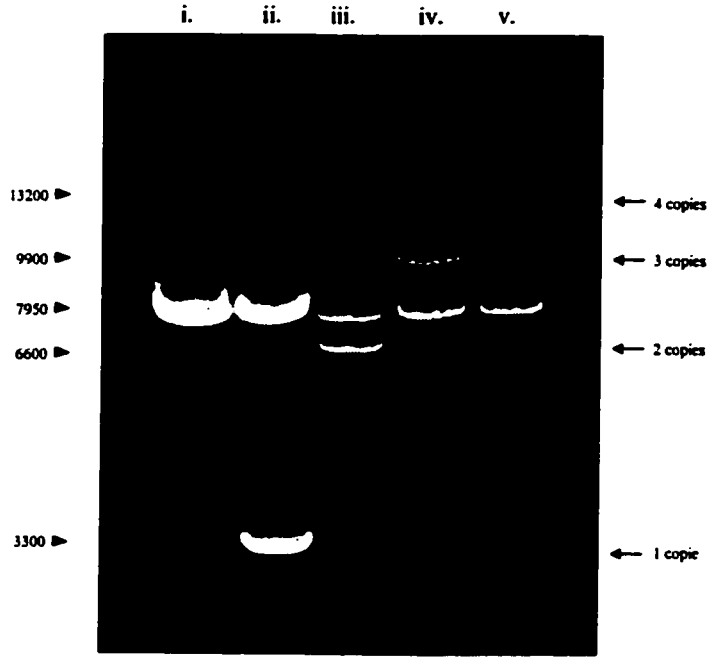
Figure 7. Construction of reporter plasmids containing D4Z4 repeats.

(A) A schematic diagram depicting the structure of the control reporter plasmid pRP2044 (i.), and the experimental plasmids pDY01, pDY02, pDY03 and pDY04 (ii.-v.) which contain 1-4 copies of the D4Z4 repeat flanked by LoxP consensus sites. Floxed D4Z4 repeats (gray boxes) lie 93bp from the BOS (Ef1α) promoter (black box) that drives the expression of the LacZ reporter gene cassette (striped box). (B) Plasmids ii.-v. were digested with *Hind*III to release a fragment containing the D4Z4 repeat arrays and separated on a ethidium bromide-stained agarose gel to confirm repeat number. Plasmid i. (pRP2044) does not contain a *Hind*III site and was linearized with *Kpn*I digestion. The size of each band is shown on the left.

(A)


i.
KpnI
BOS
Lac Z
pRP2044
ii.
LoxP
HindIII
LoxP
HindIII
D4Z4 Repeat
pDY01
iii.
LoxP
HindIII
LoxP
HindIII
pDY02
iv.
LoxP
HindIII
LoxP
HindIII
pDY03
v.
LoxP
HindIII
LoxP
HindIII
pDY04
0
2500
5000
7500
10000
12500
15000
17500
20 000

(B)


i.
ii.
iii.
iv.
v.
13200
9900
7950
6600
3300
4 copies
3 copies
2 copies
1 copie

analyzed cell extracts for β-gal activity after 48 hours (transient transfection) or following the selection of stable clones under both growth and differentiation conditions (see Figure 8). C2C12 cells were originally derived from mouse myosatellite cells and were chosen as an *in vitro* model as they are the best characterized cell culture model to study myogenesis [57, 58, 73].

3.2 – The effect of D4Z4 repeats on transient reporter gene expression.

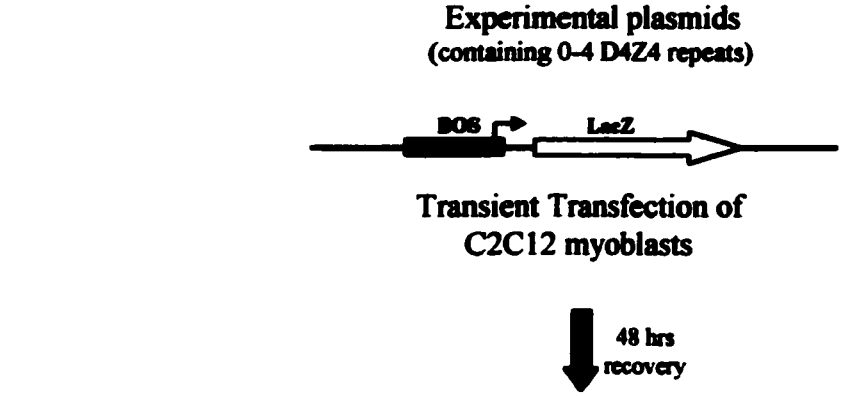
3.2.1 – Analysis of transient LacZ expression during growth conditions.

To assess if the D4Z4 repeats have any effect on transient LacZ expression; C2C12 cells were transiently transfected with the control plasmid (pRP2044) or the experimental plasmids (pDY01, pDY02, pDY03 and pDY04). Protein samples were isolated from cells 48 hours after transfection and assayed for β-gal activity with the quantitative ONPG assay. The gray bars in Figure 9 represent the mean values of β-gal activity relative to the control-transfected cells (pRP2044) (N = 18). Although there appears to be a slight increase between pRP2044 and the experimental plasmids this was not statistically significant for pDY01, pDY02 and pDY04. There is a statistically significant increase in relative β-gal activity among cells transfected with plasmid pDY03 (3 D4Z4 repeats, P = 0.006), however the increase between pDY02 and pDY03 is not significant, suggesting that pDY03 represents an outlier and this value reflects the variability of the assay. Since we anticipated that the D4Z4 repeat would inhibit reporter gene expression these results suggested that the repressive effect might only be detectable following myogenic differentiation.

Figure 8. A schematic representation of transfection strategies.

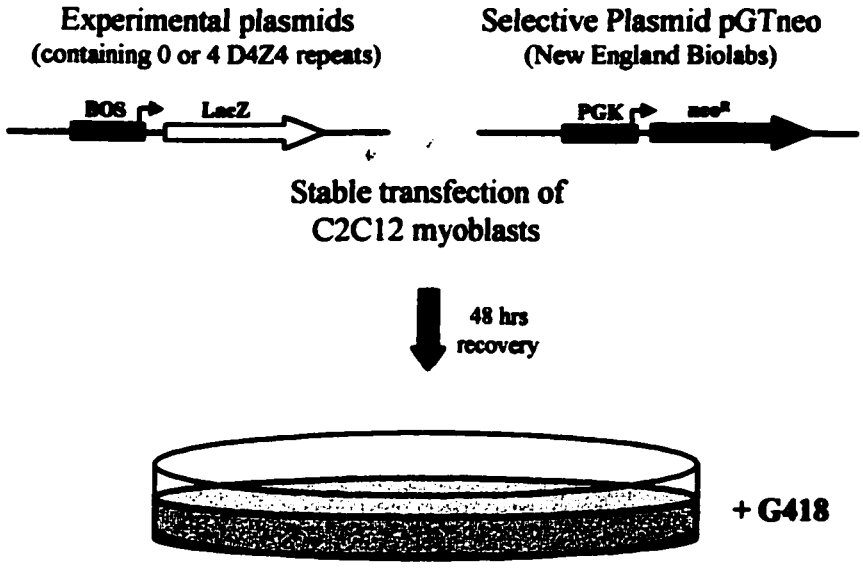
(A) C2C12 myoblasts were transiently transfected with control or D4Z4 containing plasmids then assayed 48 hours later for β-gal activity. Cell extracts isolated from transfected cells under growth or differentiation conditions were assessed by quantitative ONPG assays. For each plasmid, relative transfection efficiency was calculated as the ratio of β-gal positive cells from the total cell population. (B) Stable plasmid expression in C2C12 myoblasts was achieved by generating G418 selectable clones following co-transfection of experimental or control plasmids with the drug selectable plasmid pGTNeo (New England Biolabs). Drug-resistant clones were examined by PCR for stable integration of plasmid DNA. Positive stable clones were assayed for β-gal activity during growth and following Cre recombinase-mediated excision of floxed D4Z4 repeats, or after differentiation, by quantitative ONPG assays. For further details of both strategies refer to Materials and Methods.

(A)



1. Determination of transfection efficiency
2. Measurement of β-gal activity (quantitative ONPG assay) or
3. Measurement of β-gal activity following differentiation (quantitative ONPG assay)

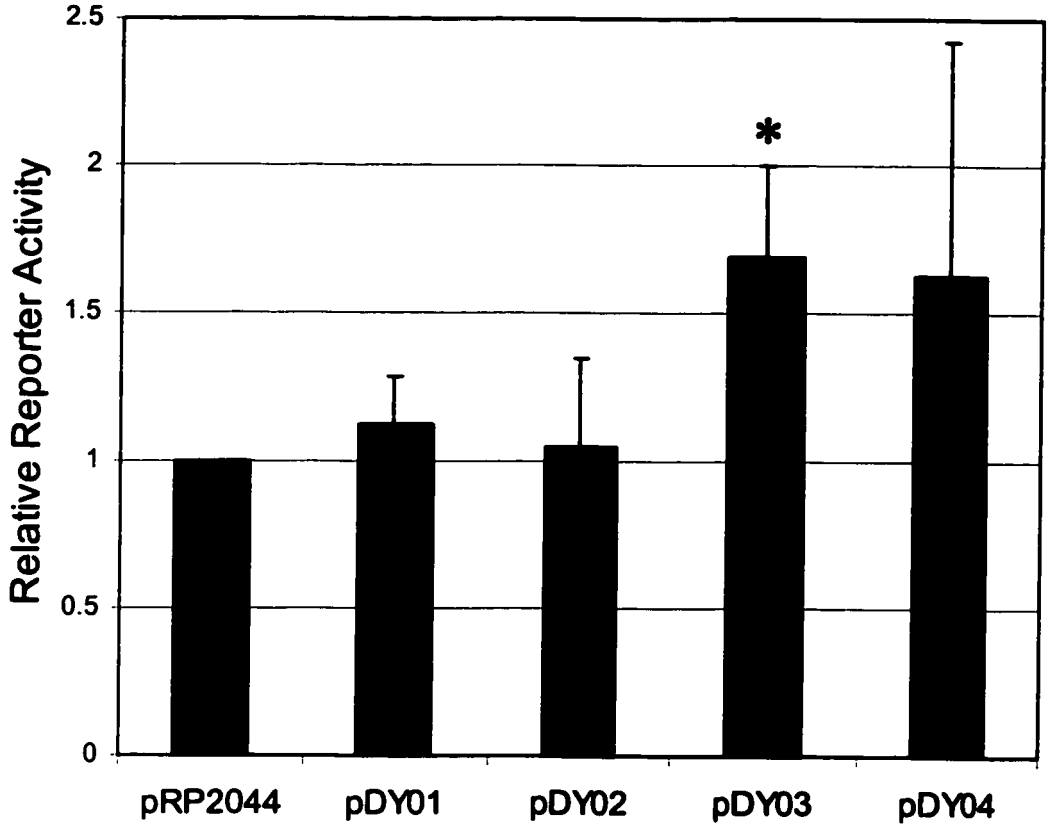
(B)



1. Expansion of G418 resistant single cell clones (characterize; PCR and Southern hybridization)
2. Measurement of β-gal activity during growth or following treatment with Cre recombinase (quantitative ONPG assay)
3. Measurement of β-gal activity during growth or following myotube differentiation (quantitative ONPG assay)

Figure 9. Reporter gene expression in transiently transfected C2C12 myoblasts.

Subsequent to a 48-hour recovery period from transfection, cell extracts were assayed for β-gal reporter activity with spectrophotometric ONPG assays. Bars represent the mean value of 18 experiments (+/- SD) for reporter activity. (*, P-value = 0.006). All values have been corrected for transfection efficiency and normalized to pRP2044 values.

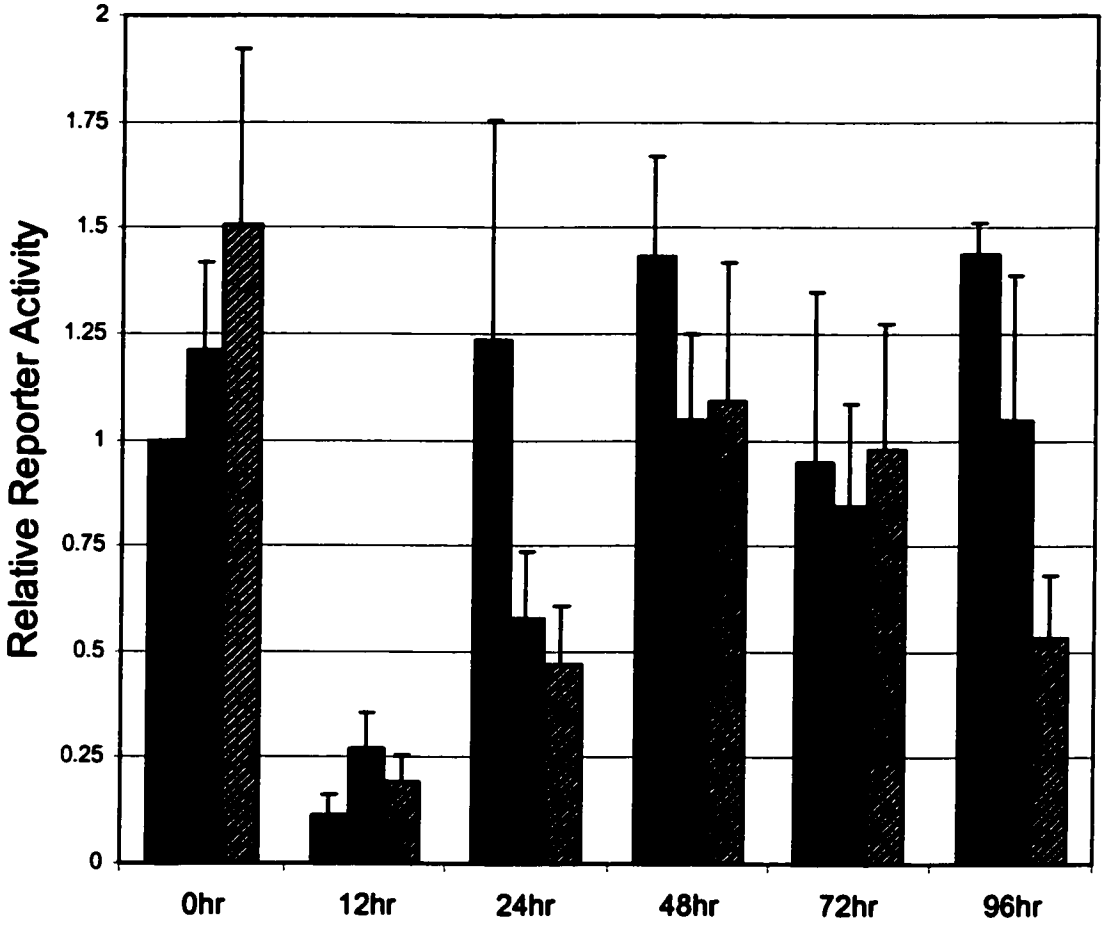

Relative Reporter Activity
2.5
2
1.5
1
0.5
0
*
pRP2044
pDY01
pDY02
pDY03
pDY04

3.2.2 – Analysis of transient LacZ expression during myogenic differentiation.

To assess whether there is a muscle-dependant molecular event involving the D4Z4 repeats in the regulation of gene expression, transiently transfected C2C12 myoblasts were studied during myogenesis. The mean values of β-gal reporter activity from transiently transfected C2C12 myoblasts during differentiation are presented in Figure 10, relative to undifferentiated pRP2044 cells (control) for three experiments. Following a 48 hour recovery, cells transfected with the control plasmid pRP2044 and the experimental constructs pDY02 and pDY04, were introduced to differentiation media (containing low serum) resulting in mitogen deprivation and induced myogenesis (terminal differentiation). Protein samples were isolated over the course of 96 hours at time points: 12, 24, 48, 72 and 96 hours. β-gal activity was then assessed with the quantitative ONPG assay. Initially, after 12 hours under differentiation conditions, the relative level of β-gal activity was significantly decreased in all tested plasmids by greater than 4.5-fold relative to control or basal values (undifferentiated control, t = 0). This decrease in reporter activity coincides with a high level of cell apoptosis, a common response to the initial stress of mitogen deprivation [213-215]. C2C12 cells transfected with pRP2044 quickly return to produce basal levels of β-gal activity within 24 hours, and this activity level is maintained throughout the time course (dark bars). In pDY02 and pDY04 transfected cells (light and striped, respectively) there is a time lag in recovery from 24 to 48 hours compared to the pRP2044 cultures. Myoblast fusion is typically observed in C2C12s between 3 and 4 days [206, 213, 214, 216]. By the end of the time course (96 hours), pRP2044 transfected cultures had an approximately 0.5-fold increase in β-gal activity, while pDY02 cultures were essentially constant with their respective basal measurements. However C2C12 cells

Figure 10. Reporter gene expression in transiently transfected C2C12 myoblasts during differentiation.

Cell cultures transiently transfected with plasmid DNA from pRP2044 (dark bars), pDY02 (light bars) and pDY04 (striped) were induced to differentiate by serum withdrawal. Reporter gene activity was then assessed from cell extracts isolated from differentiating cultures at 12, 24, 48, 72 and 96 hours with ONPG assays. Bars represent the mean value of three experiments (+/- SD) for reporter activity. All values have been corrected for transfection efficiency and normalized to pRP2044 control values (time =0).


Relative Reporter Activity
2
1.75
1.5
1.25
1
0.75
0.5
0.25
0
0hr
12hr
24hr
48hr
72hr
96hr

transfected with pDY04 plasmid DNA demonstrated β-gal activity that was approximately 30% the level of their initial undifferentiated expression. At 96 hours, the pDY02 and pDY04 cultures had β-gal activities, which were approximately 86% and 33% relative to the pRP2044 culture (96 hrs) β-gal activity at that same time point, suggesting that the D4Z4 repeats may have a cumulative effect on the repression of LacZ during differentiation. Time course analysis was halted after 96 hours due to the high rate of death and cell turnover resulting in loss of transfected cells. It is also possible that cells containing the plasmids pDY02 and pDY04 fail to differentiate properly and as such are more prone to undergo apoptosis which could also lead to a seemingly lower β-gal activity.

3.3 – The effect of stable D4Z4 repeats on reporter gene expression.

The potential repressive effect of the D4Z4 repeats may require genomic integration of transgenes for the formation of a stable heterochromatin complex, signifying the importance of the chromatin environment on gene silencing [89, 156]. To assess this, we studied expression of the LacZ reporter gene from control and experimental plasmid constructs stably integrated into the host C2C12 cell genome. Due to the inherent clone differences, such as position of integration and copy number, which are known to affect transgene expression, we selected several C2C12 lines containing the control plasmid pRP2044 and experimental plasmid pDY04 in order to reduce clonal variability and identify consistent trends in reporter expression [57, 163, 164, 217-219].

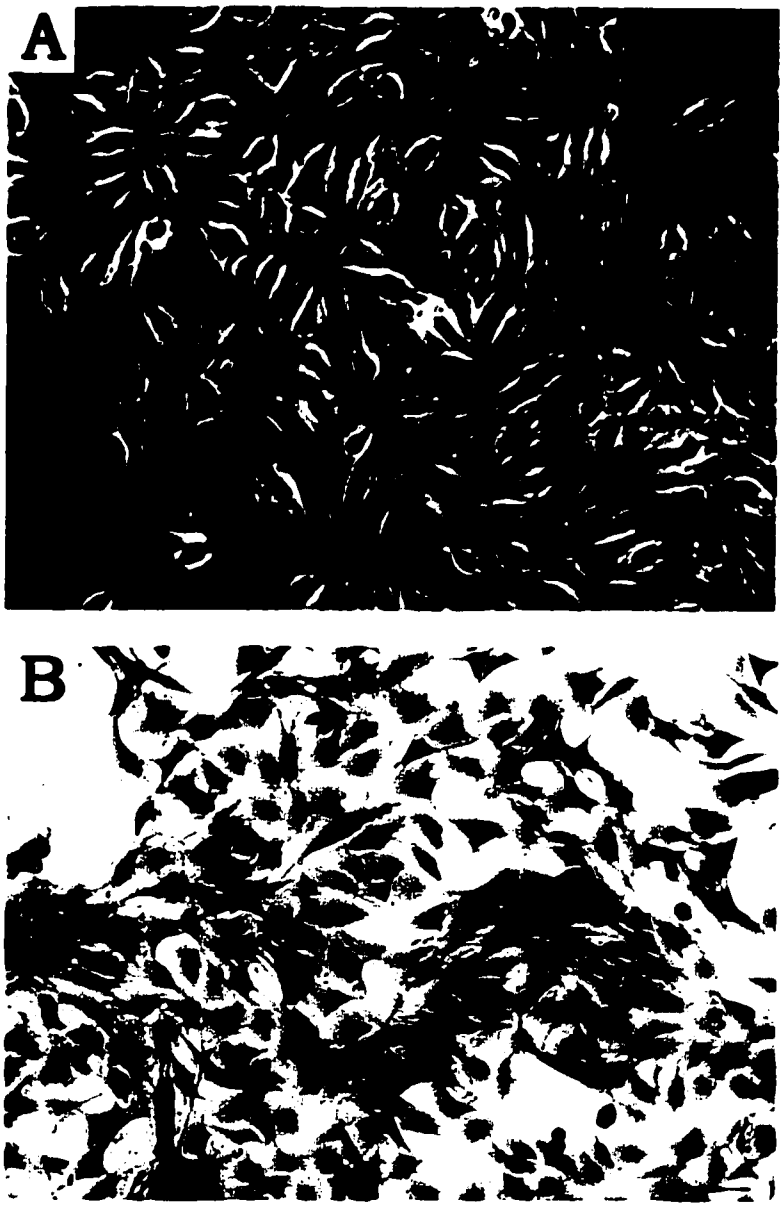
3.3.1 – Selection of stable clones in a C2C12 background.

It has been previously noted from *in vitro* studies that the drug selectable marker gene conferring neomycin resistance (Neo^R) can infer a negative effect on a nearby reporter gene promoter when it is contained within the same plasmid [219]. To avoid these potential Neo^R gene effects, we chose to co-transfect the experimental or control plasmid of interest with the selectable marker plasmid pGTNeo (New England Biolabs) for the generation of stable clones. Our hope was to develop single copy clones, which would allow for the direct assessment of the potential regulatory effects D4Z4 repeats have on a LacZ reporter at one locus. All stable clones were generated using low passage C2C12 cultures (<9 passages), in an effort to reduce the occurrence of fusion defects that can occur spontaneously through selection [58]. Expanded G418 (a neomycin derivative) resistant clones were screened for control or experimental plasmid insertion, analyzed for intact expression cassettes and characterized for transgene copy number and plasmid insertion frequency.

Since we observed that transfections containing four D4Z4 repeats (pDY04) created the greatest effect for reporter gene expression in the transient studies, we focused our efforts on identifying pRP2044 and pDY04 stable clones. Following co-transfection of plasmids pRP2044 and pDY04 with pGTNeo, several G418 resistant colonies were selected and expanded (data not shown). A percentage of these selected clones were expected to contain only the pGTNeo selectable plasmid and not the co-transfected LacZ constructs; consequently an inspection for LacZ expression was performed using *in situ* β-gal

Figure 11. Stable clones are homogeneous with wild type C2C12 myoblast morphology.

(A) Phase microscopy photograph of C2C12 myoblasts. (B) Phase microscopy photograph of stable clone cDY043 cells depicting an identical cell morphology to its C2C12 progenitor and homogeneous LacZ expression, as determined by cleavage of the colour reactive substrate X-gal. This analysis was comparable for all stable clones with 85-100% homogeneity. (Scale bars represent 50μm)


A
B

assays. Figure 11 is a phase microscopy image depicting wt C2C12 cells (Figure 11(A)) and cells from line cDY043 (Figure 11(B)) following *in situ* staining for β-galactosidase. This image shows that stably transfected clones present the prototypical C2C12 morphology with, in this case, homogeneous LacZ expression. Although only half of the 14 characterized clones demonstrated 100% homogeneity for expression of LacZ (refer to Table 1), typically each clone selected had a high incidence of β-gal expressing cells (85-95%) and were thus used for additional experiments. Despite the presence of LacZ activity, positive confirmation of the D4Z4 repeat sequence was necessary to rule out the possibility that the repeats were excised during integration of plasmid in the generation of the stable lines. A polymerase chain reaction based screening method was employed to verify the presence of both LacZ and D4Z4 sequences. Amplification of LacZ and D4Z4 sequence was performed using the RAS 39 / RAS40 and bD4Z4-3026f / bD4Z4-3219r primer pairs. The position of the primer sequences are shown as arrows in Figure 12(A). A LacZ specific band (~200 bp) was amplified from all clones and plasmid control lanes (Figure 12(B), upper panel) while the D4Z4 specific band (~200 bp) was only amplified in the pDY04 containing clones, and the pDY04 plasmid control lane (Figure 12(B), lower panel). The diffuse bands located around the 100 bp marker in the LacZ amplification experiment are most likely the product of primer multimers and do not represent an amplified secondary band. The negative control lanes using mock treatment with no DNA (Blank) and control C2C12 genomic DNA failed

Table 1. Summary of transgene copy and insertion site number for stable clones.

CLONE	Copy Number*	Insertions*	B-Gal Expression**
cRP20442	3	ND	95%
cRP20444	4	ND	100%
cRP2044n5	2	1	100%
cRP2044n7	3	1	95%
cRP2044n15	ND	ND	93%
cRP2044n24	4	1	100%
cRP2044n31	1	1	100%
cDY041	1	1	100%
cDY042	3	ND	90%
cDY04n2	12	1	95%
cDY043	2	2	100%
cDY044	2	2	85%
cDY04n/3	2	1	100%
cDY04n/23	1	1	95%

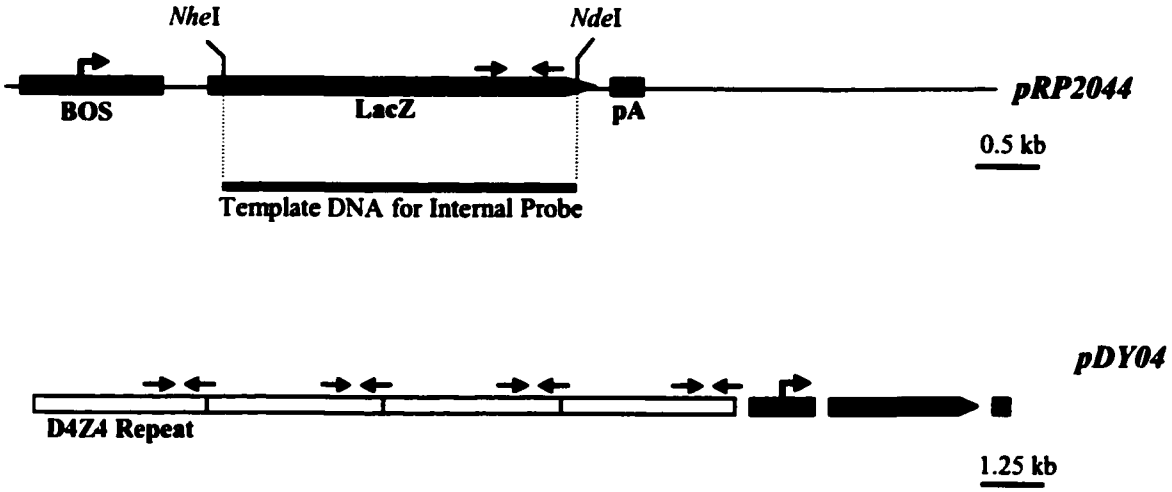
* Copy and Insertion numbers estimated from Southern blot analysis. Refer to Materials and Methods (data not shown) ND, not determined

** Percentage of cells in culture expressing B-gal

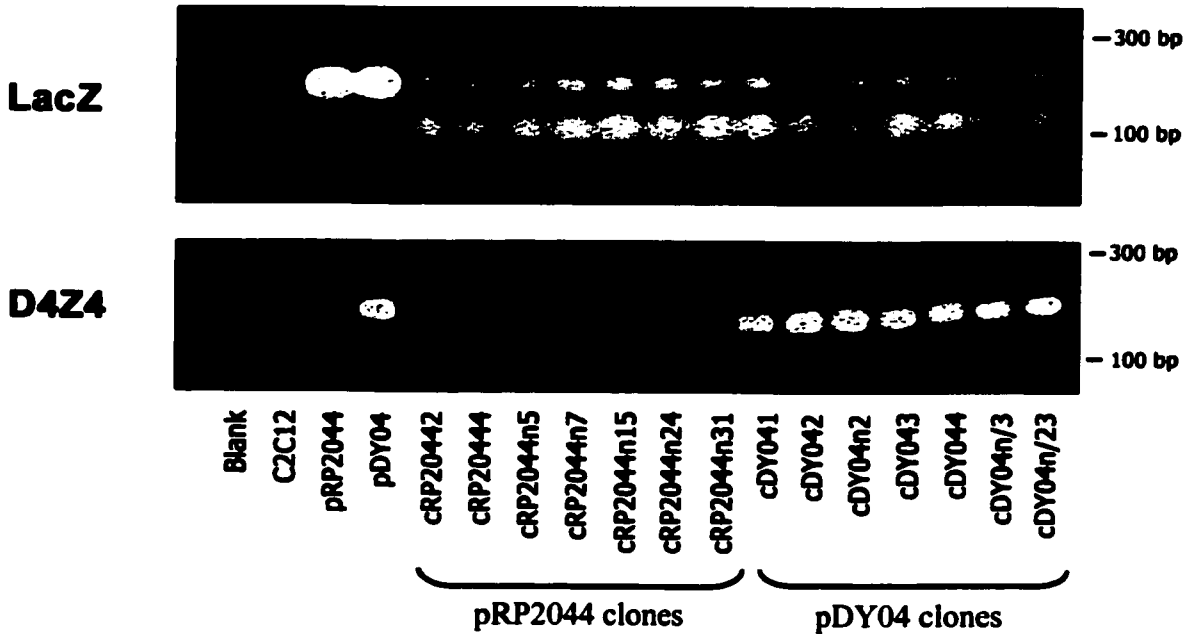
Figure 12. Isolation of stable clones containing the experimental or control plasmids.

(A) Schematic diagrams of the pRP2044 and pDY04 constructs depicting the placement of LacZ (dark gray arrow) and D4Z4 (light gray Bars) sequence. Arrows above LacZ and D4Z4 repeats correspond to primer pairs used to amplify these genes. The black bar represents the *Nhe*I / *Nde*I fragment of LacZ, which was used as a template for the internal probe used in Southern analysis of stable clones. (B) PCR amplification of genomic DNA isolated from stable clones (shown at the bottom). In the upper panel primer pairs directed toward LacZ sequence detected a band (~200bp) in all clones and in pRP2044/pDY04 plasmid controls. The diffuse non-specific band (~100-120bp) represents primer dimers. In the lower panel a D4Z4 specific band (~200bp) was only detected in clones containing the D4Z4 repeat and its plasmid control lane (pDY04). Mock reactions with no DNA or C2C12 genomic DNA did not amplify. The placement of molecular weight size markers is shown on the right.

(A)


NheI
NdeI
BOS
LacZ
pA
pRP2044
0.5 kb
Template DNA for Internal Probe
pDY04
D4Z4 Repeat
1.25 kb

(B)


LacZ
D4Z4
300 bp
100 bp
300 bp
100 bp
Blank
C2C12
pRP2044
pDY04
cRP20442
cRP20444
cRP2044n5
cRP2044n7
cRP2044n15
cRP2044n24
cRP2044n31
cDY041
cDY042
cDY04n2
cDY043
cDY044
cDY04n/3
cDY04n/23
pRP2044 clones
pDY04 clones

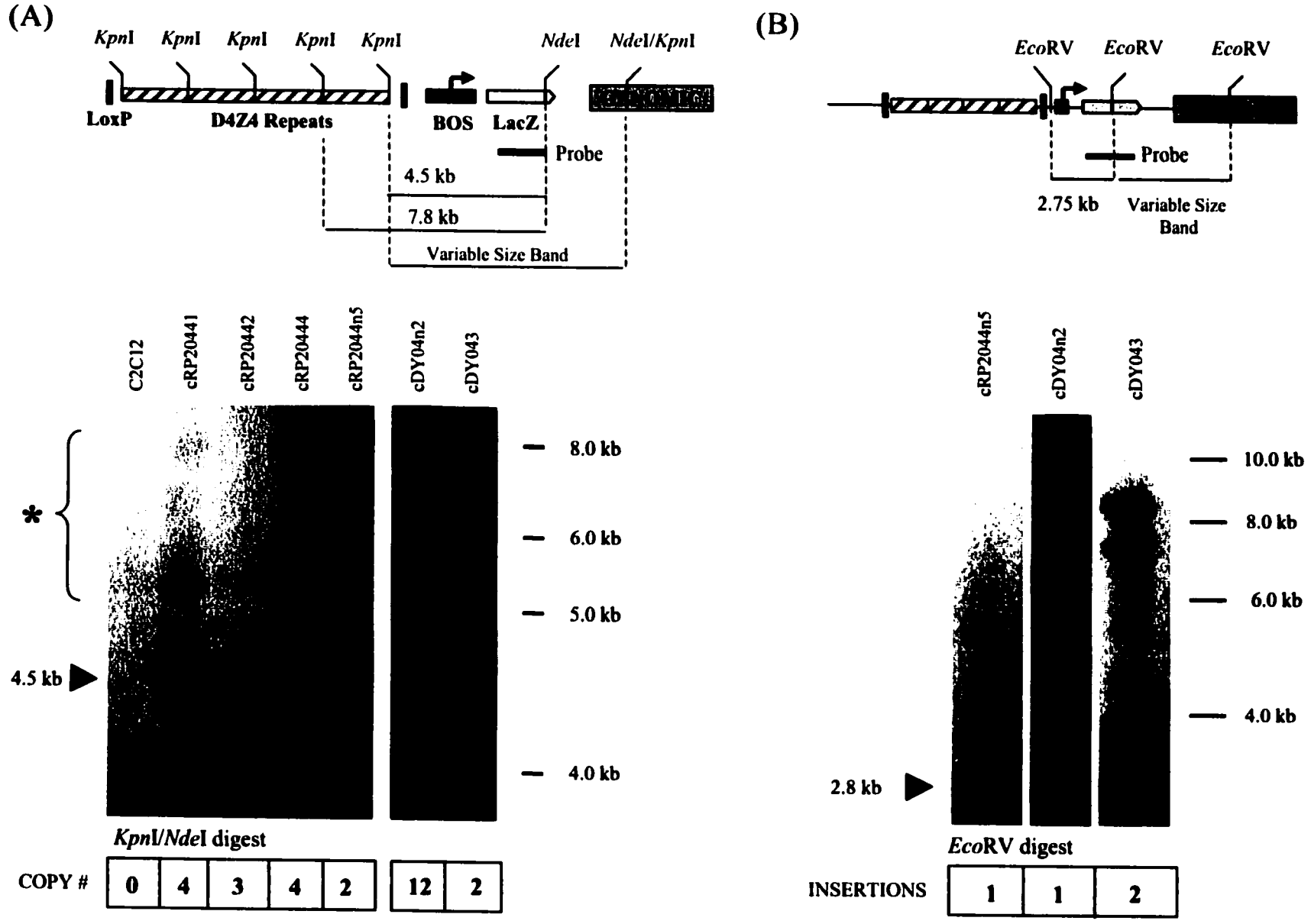
to amplify either product, as expected. These results indicate that not only is the LacZ reporter gene present, but also that the entire expression construct has been integrated. Significant variation in LacZ expression was observed among clones; it has been previously suggested that copy number and insertion frequency of transgenes along with position effects play key roles in the expression of reporter genes [57, 163, 217-219]. Therefore, each stable clone generated was analyzed for transgene copy number and whether the clone integrated at a single or multiple sites by Southern hybridization analysis with a probe generated from a 2.3 kb *Nde*I/*Nhe*I fragment of the LacZ gene (Figure 12(A), 13(A)). The relative intensity of a 4.5 kb *Kpn*I/*Nde*I DNA fragment detected by the 2.3 kb probe was used to determine the copy number for each clone. The blot was then striped and rehybridized with a probe for *Snf*2h, as a loading control for comparison. Figure 13(A) shows Southern analysis of the stable clones cRP20441, cRP20442, cRP20444, cRP2044n5, cDY02n2, and cDY043. In order to assess for transgene insertion, this same LacZ probe was used to detect a 2.8kb internal and a variable sized *Eco*RV band following digestion of genomic DNA isolated from stable cell clones cRP2044n5, cDY04n2 and cDY043, as shown in Figure 13(B). A summary of copy number and integration frequency is shown in Table 1. In addition, Table 1 also delineates the homogeneity status of each clone.

3.3.2 – Clones containing pDY04 demonstrate a lower frequency of LacZ expression.

During the selection of stable clones it was noticed that the ratio of G418 resistant clones that were positive for the LacZ gene was higher for the plasmid pRP2044 than it was for pDY04. It was further noted that a proportion of stable clones positive for the LacZ gene by PCR did not contain any β-gal activity. To investigate these circumstances further,

Figure 13. Determination of the number of transgene copies and insertion sites by Southern blot analysis.

(A) A schematic representation (top) showing positions of restrictions sites for *Kpn*I and *Nde*I within the genomic DNA integrated-pDY04 cassette. The size of predicted DNA fragments are shown below. Genomic DNA isolated from stable cell cultures (listed above each lane) and C2C12 a control were digested with *Kpn*I and *Nde*I, then subjected to Southern blot analysis (bottom). The blot was first hybridized with the ^{32}P-labelled LacZ internal probe (black bar), then striped and rehybridized with a DNA probe that detects mouse *Snf2*h gene as a loading control (data not shown). A 4.5 kb signal which corresponds to an internal fragment (arrowhead) was used to assess copy number by relative intensity. The copy number is given at the bottom of each lane. Larger signals (marked by *) are the result of incomplete digestion such as the 7.8 kb and variable sized DNA fragments shown in the schematic drawing. (B) A schematic representation (top) showing positions of *Eco*RV restriction sites within the genomic DNA integrated-pDY04 cassette. The size of predicted DNA fragments from *Eco*RV digestion are shown. Genomic DNA isolated from stable cell cultures (listed above each lane) and a C2C12 control, was digested with *Eco*RV, then subjected to Southern blot analysis (bottom) using the same LacZ probe as in (A). A 2.75 kb internal fragment (arrowhead) and variable size band was detected. The number of variably sized fragments detected represent the number of random insertions for each clone (listed below each lane). Bands located at ~10 kb and 11 kb in the cDY04n2 lane represent partial digestions, as judged by the intensity differentiation between the ~6.5kb and 2.8kb band.


(A)
KpnI KpnI KpnI KpnI KpnI
NdeI
NdeI/KpnI
LoxP
D4Z4 Repeats
BOS
LacZ
Probe
4.5 kb
7.8 kb
Variable Size Band
C2C12
cRP20441
cRP20442
cRP20444
cRP2044n5
cDY04n2
cDY043
8.0 kb
6.0 kb
5.0 kb
4.0 kb
*
4.5 kb
KpnI/NdeI digest
COPY #
0
4
3
4
2
12
2
(B)
EcoRV
EcoRV
EcoRV
Probe
2.75 kb
Variable Size Band
cRP2044n5
cDY04n2
cDY043
10.0 kb
8.0 kb
6.0 kb
4.0 kb
2.8 kb
EcoRV digest
INSERTIONS
1
1
2

several additional clones were generated with the pRP2044 and pDY04 plasmids. Selected clones were screened by PCR for the LacZ gene to evaluate for integration. Cell extracts from positive clones were screened for β-gal activity using the more sensitive spectrophotometric ONPG assay. We observed a 2-fold higher incidence of clones positive for LacZ by PCR with pRP2044 than with pDY04 (Table 2). This suggests that stable integration of the plasmid pDY04 (~21 kb) integration is less efficient than for the plasmid pRP2044 (~8 kb), which is probably a function of plasmid size. Secondly, we observed that only 36% of pDY04 clones that contained the LacZ gene by PCR analysis actually expressed the β-gal protein. This was 20% lower than pRP2044 (see Table 2), suggesting that pDY04 clones may be more susceptible to position effects or transgene silencing (see discussion).

3.3.3 – The effect of stable D4Z4 repeats on reporter expression assays.

3.3.3.1 – Analysis of D4Z4 involvement on the stable expression of LacZ.

The purpose of the LoxP sites in the context of the experimental plasmid was to allow for stable Cre-recombinase mediated site-specific excision of the D4Z4 repeats. By introducing Cre-recombinase through adenoviral infection to stable clone myoblast cultures, we were able to assess direct effects of D4Z4 on LacZ expression within the same genomic region and thereby independent of position effects. Quantin B *et al.*, [220], showed that despite the many difficulties of gene transfer into muscle cells, adenoviral vectors produced moderate to high expression of Ad-reporter genes in myogenic cell lines as well as *in vivo*.

Table 2. Summary of stable clones that contain and express the LacZ reporter gene.

PLASMID	Clones Selected	Positive by PCR**	Positive for Reporter Activity *
pRP2044	17	13 (76%)	7 (54%)
pDY04	91	33 (36%)	12 (36%)

* for a representational selection of raw data see Appendix B

** confirmed by using primer pairs RAS39/RAS40 and bD4Z4-f3026/bD4Z4-r3219

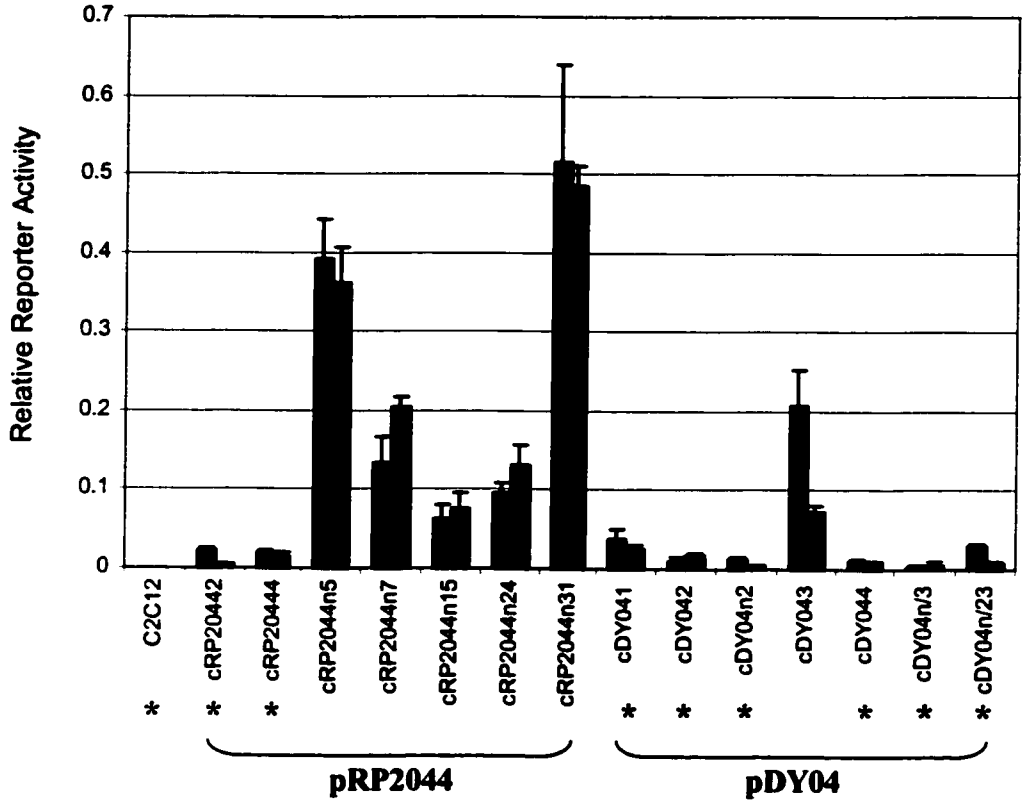
Previous experiments have also shown that adenovirus vectors are a viable and efficient means for the introduction of over-expressing transgenes in C2C12 cultured cells [215, 221]. We infected stable clone cultures with the adenovirus AdMA26 (gift of R. Parks, Ottawa ON) that expresses Cre-recombinase under the control of the murine cytomegalovirus (mCMV) promoter. Figure 14 shows the mean β-gal activity of seven different pDY04 stable clones and seven unique clones containing the control plasmid (pRP2044) from eight replicate experiments. The first observation we made is that each clone has an independent level of reporter gene activity under growth conditions (dark bars); for example, compare cDY043 and cDY04n2. Similarly, expression of control plasmids is also susceptible to this variability. Several clones (denoted by * in Figure 14(A)) had a relatively low level of expression. Figure 14(B) contains data from these clones plotted on a different scale. The second observation that we made was that the pDY04 clones predominantly express at a lower level than the pRP2044 clones. These observations suggest that the differences in gene expression are the result of the site of integration.

Next we examined LacZ expression following Cre-excision of the D4Z4 repeats. We expected to see an increase in β-gal activity upon Cre-excision (light bars), if the D4Z4 repeats were functionally repressive. We observed an increase in β-gal activity following Cre-treatment in two clones (cDY042 and cDY04n/3, Figure 14(B)); however this was not a consistent trend. Furthermore, significant decreases were observed in clones cDY043 (Figure 14(A)) and cDY04n/23 (Figure 14(B)). Treatment with AdMa26 was not expected to affect clones containing the control plasmid pRP2044 as they do not contain LoxP

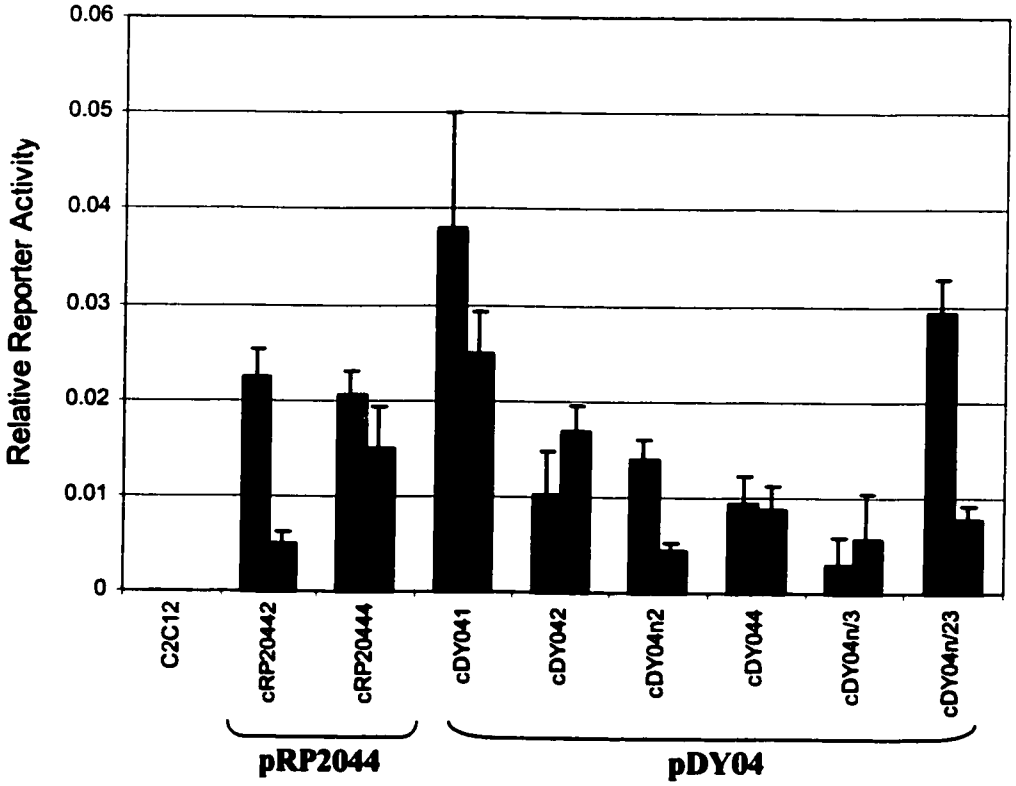
Figure 14. Reporter gene activity of pRP2044 and pDY04 stable clones under growth conditions.

The average β-gal activity per LacZ transgene copy is plotted for pRP2044 and pDY04 clones (listed below). (A) Cell extracts isolated from stable clones were assayed for β-gal activity during growth (dark bars) or following Cre recombinase-mediated excision of floxed D4Z4 repeats (light bars) by quantitative ONPG assays. (B) Clones from panel (A) with low activity (denoted by *) were re-plotted using an appropriate scale. Bars represent the mean value of 8 replicate experiments (+/- SD) for reporter activity.

(A)


Relative Reporter Activity
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
C2C12
cRP20442
cRP20444
cRP2044n5
cRP2044n7
cRP2044n15
cRP2044n24
cRP2044n31
cDY041
cDY042
cDY04n2
cDY043
cDY044
cDY04n/3
cDY04n/23
pRP2044
pDY04

(B)


Relative Reporter Activity
0.06
0.05
0.04
0.03
0.02
0.01
0
C2C12
cRP20442
cRP20444
cDY041
cDY042
cDY04n2
cDY044
cDY04n/3
cDY04n/23
pRP2044
pDY04

recognition sequence. Nonetheless, clone cRP20442 (Figure 14(B)) shows a significant decrease in the level of reporter activity. The other control clones (cRP20444, cRP2044n5, cRP2044n7, cRP2044n15, cRP2044n24 and cRP2044n31) remain relatively constant in their expression. These results suggest that under growth conditions the D4Z4 repeats are unlikely to have a repressive effect on gene expression in a genomic context. Moreover, clones cDY04n2 (Figure 14(B)), cDY043 (Figure 14(A)) and cDY04n/23 (Figure 14(B)) that demonstrated a decrease in reporter activity following Cre-excision, contained multiple transgene arrays. Thus, the reduced activity may reflect excision of the internal LacZ reporter genes. This would result in a superficially lower level of LacZ expression while not representing true gene repression (see discussion).

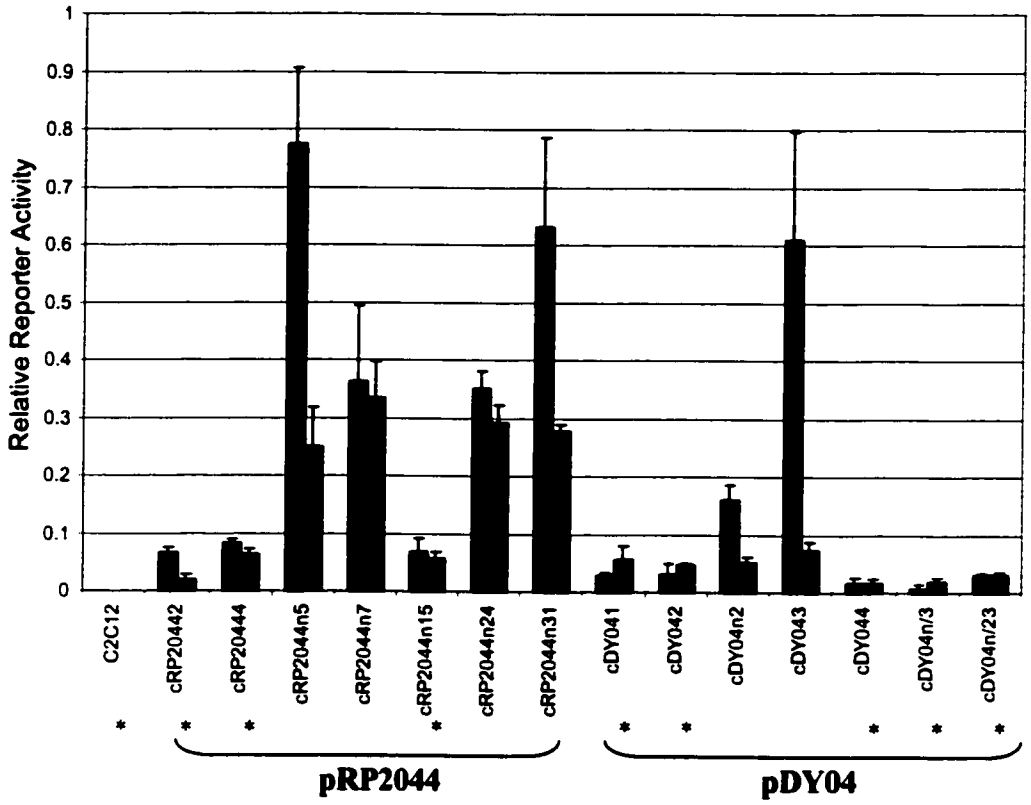
3.3.3.2 – Analysis of D4Z4 involvement on the stable expression of LacZ during differentiation.

Based on preliminary results from transient assays, we hypothesized that repression by D4Z4 repeats would only become evident during myogenesis, possibly through chromatin domain rearrangements or expression of muscle-specific factors. From transient assays, we had observed that the D4Z4 repeat caused a lag in the recovery of LacZ expression during differentiation and that at 96 hours of differentiation, the cells transfected with D4Z4 repeats had a noticeable decrease in β-gal activity (Figure 10). To determine if stable clones displayed the same reduced level of LacZ expression we measured the average β-gal activity per transgene copy following 96 hours of differentiation compared to the values obtained under myoblast growth conditions. Figure 15 shows the mean β-gal activity of seven pDY04 stable clones and seven clones containing the control plasmid (pRP2044)

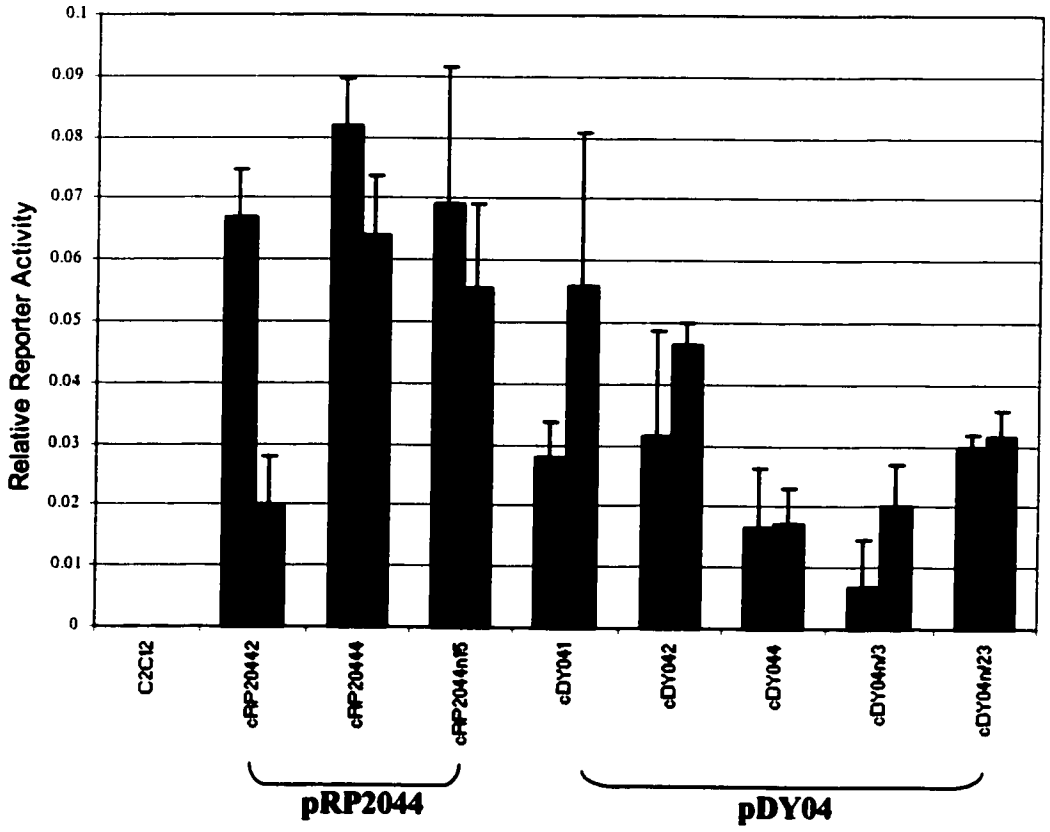
Figure 15. Reporter gene activity following differentiation of pRP2044 or pDY04 stable clones.

The average β-gal activity per LacZ transgene copy is plotted for each clone (listed below). (A) Cell extracts isolated from stable clones were assayed for β-gal activity during growth (dark bars) or following differentiation (light bars) using ONPG assays. (B) Clones from panel (A) with low activity (denoted by *) were re-plotted using an appropriate scale. Bars represent the mean value of 8 replicate experiments for growth and 6 replicate experiments for differentiated cells (+/- SD) for reporter activity. All clones were differentiated for 96 hours.

(A)


Relative Reporter Activity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
C2C12
cRP20442
cRP20444
cRP2044n5
cRP2044n7
cRP2044n15
cRP2044n24
cRP2044n31
cDY041
cDY042
cDY04n2
cDY043
cDY044
cDY04n/3
cDY04n/23
pRP2044
pDY04

(B)


Relative Reporter Activity
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
C2C12
cRP20442
cRP20444
cRP2044n5
cDY041
cDY042
cDY044
cDY04n/3
cDY04n/23
pRP2044
pDY04

from eight replicate experiments. As observed with the previous experiment, several clones (denoted by * Figure 15(A)) had a low level of expression, and were thus plotted on an appropriate scale (Figure 15(B)) for presentation. We expected that LacZ expression would decrease in clones containing D4Z4 repeats under differentiation conditions. Decreases in β-gal reporter activity were detected for the D4Z4 containing clones cDY043 (Figure 15(A)), cDY04n2 (Figure 15(B)) and cDY04n/23 (Figure 15(B)). However, these trends were not consistent among all clones as cDY041 (Figure 15(A)), cDY042 (Figure 15(B)) and cDY043 (Figure 15(A)) showed a measurable increase in reporter expression. Moreover in control clones cRP20442 (Figure 15(B)) and cRP2044n5 (Figure 15(B)), where we expected no change in gene expression, a 3.5-fold decrease in reporter activity was observed. These results suggest that under differentiation conditions the D4Z4 repeats do not uniformly have a repressive effect on gene expression. However other molecular events such as chromatin rearrangements, which occur during myogenesis [77], may have a D4Z4-independent effect on the reporter gene regulation which would be dependent on the transgene integration site (see discussion).

3.4 – Morphology is affected by D4Z4 repeats during myogenesis in both transient transfections and stable lines.

It became increasingly clear that under differentiation conditions the LacZ reporter gene expression was being influenced at the molecular level, but not necessarily in a repressive nature. It was also clear from personal observations of the myoblast cultures that there was the emergence of a trend in phenotype associated with the presence of D4Z4 repeats. To further explicate these observations we investigated the morphology of

differentiating myoblasts in transient cultures and stable clones. Since the differentiation response of C2C12 myoblasts is known to be sensitive to cell confluency, culture conditions and serum quality [57], each transient culture and stable clone was tested multiple times. For all transient assays the transfection efficiency is represented by the average ratio of X-gal positive cells to total cell counts. The resulting ratios have been incorporated into the myotube fusion calculations.

3.4.1 – Analysis of myoblast fusion competence and morphology in transient transfections.

We first evaluated transient C2C12 cultures quantitatively for the ability to form myotubes by growing low passage wild type C2C12 to 80% confluence, transiently transfecting them, and then culturing the transfected cells in differentiation media. After 4 days of differentiation, cultures were fixed and stained for the expression of the striated muscle myotube marker, myosin heavy chain (MF20). Counter-staining with Harris modified Hematoxylin allowed for elucidation of nuclei incorporated in the multinucleated myotubes from single nucleated myocytes. For the purposes of these experiments, a myotube was defined as a multinucleated structure containing 3 or more nuclei and staining positive for MF20. Typical myotube formation, as observed in wild type C2C12 cells, appear as elongated multinucleated structures (brown) with centralized nuclei (purple) (Figure 16(A-C) and Figure 17(A)). During differentiation, transient cultures transfected with pRP2044 (Figure 17(B)) and experimental plasmids pDY02 (Figure 17(C)) and pDY04 (Figure 17(D)) demonstrated a lower rate of myotube fusion and an increasing incidence of abnormal morphology from the typical myotubes which are highlighted by arrows.

Figure 16. Morphology of C2C12 myoblasts following differentiation by serum withdrawal.

Phase microscopy photographs (A), (B), and (C) depict different magnifications (1x, 30x and 60x, respectively) of typical myotube structures observed when wild type C2C12 mouse myoblast cells are induced to differentiate by serum withdrawal (elongated with centralized nuclei). Immunostaining with Myosin Heavy Chain antibody (MF20) shows brown myotubes. Nuclei are counter-stained purple with Harris modified Hematoxylin. Bars represent approximately 1000μm, 34μm and 17μm in figures (A), (B) and (C), respectively.

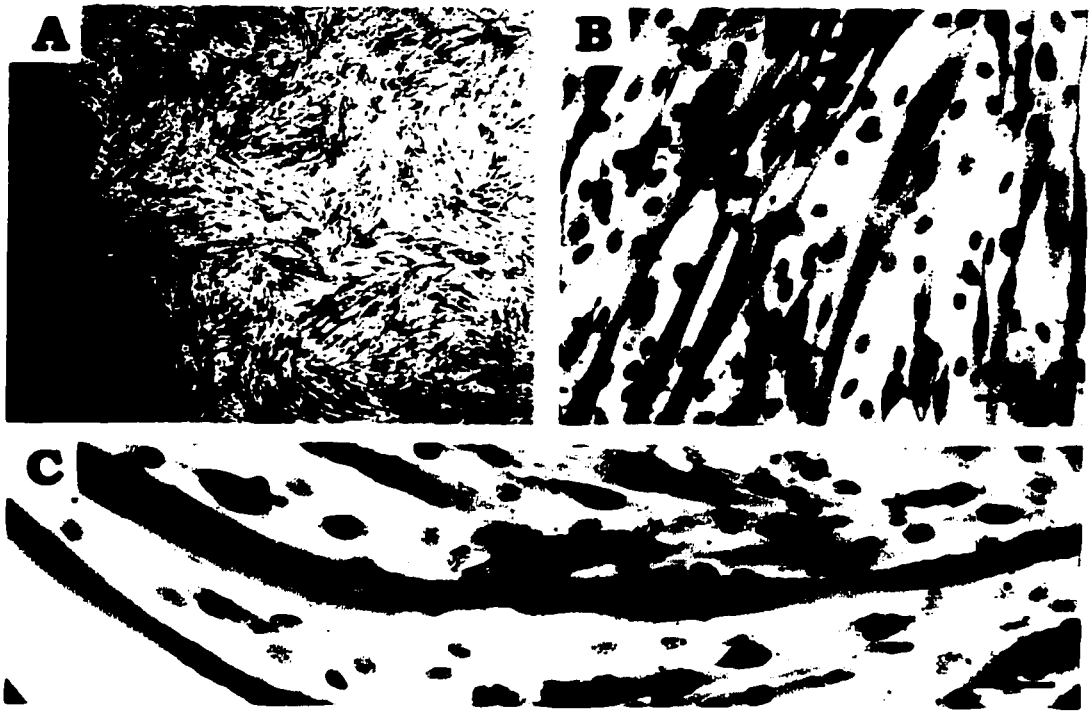

A
B
C

Figure 17. C2C12 fusion competence and morphology are affected by the D4Z4 repeats in transiently transfected C2C12 myotubes.

Phase microscopy images of untransfected C2C12 myotube cultures (A), and cells transfected with the pRP2044 (B), pDY02 (C) and pDY04 (D) experimental plasmids. Following a 48 hour recovery from transfection, cultures were induced to differentiate through mitogen deprivation for 96 hours. Myotube structures were immunostained (brown) with a Myosin Heavy Chain antibody (MF20) and nuclei are counter-stained with Harris modified Hematoxylin (purple). Arrows denote specific myotube structures with abnormal or deformed morphology (swelling with non-centralized masses of nuclei). The fusion competence of the cultures decreases with plasmid transfection, and between plasmids of increasing repeats as shown by the relative abundance of brown stained myotubes from individual nuclei. (Scale bars represent approximately 110μm)

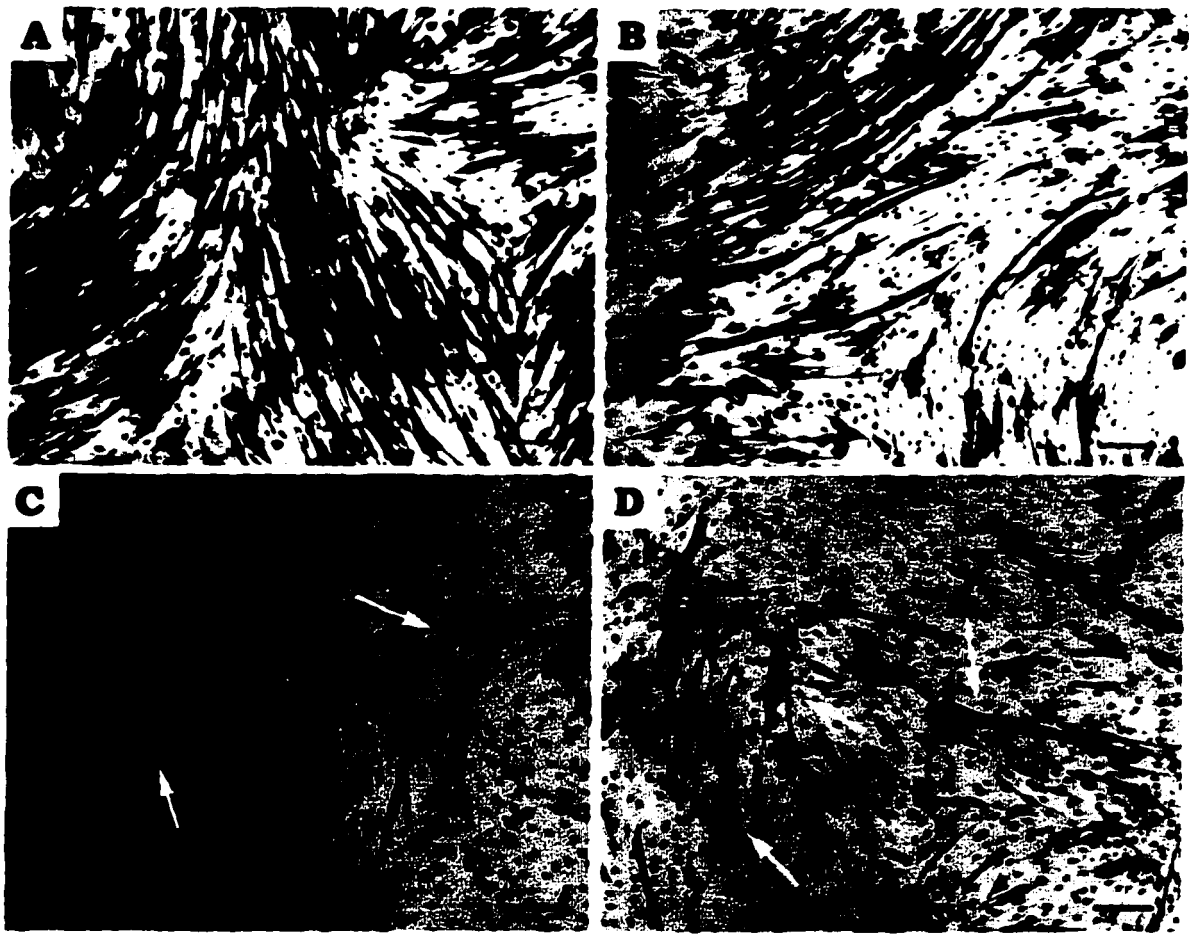
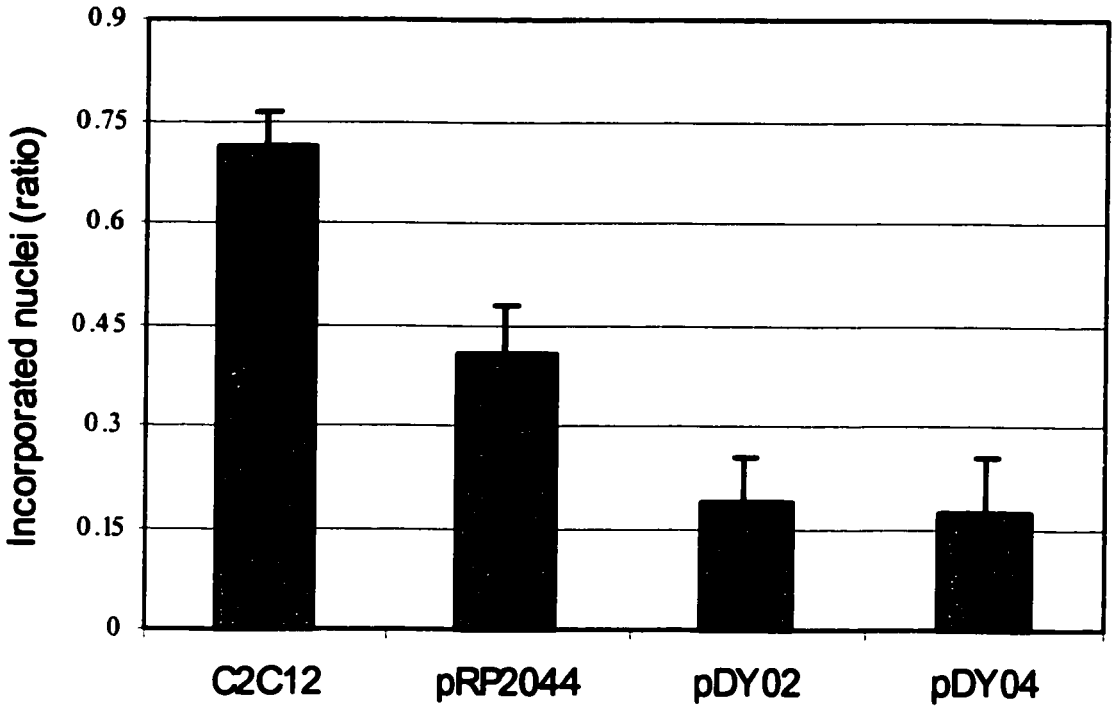

A
B
C
D

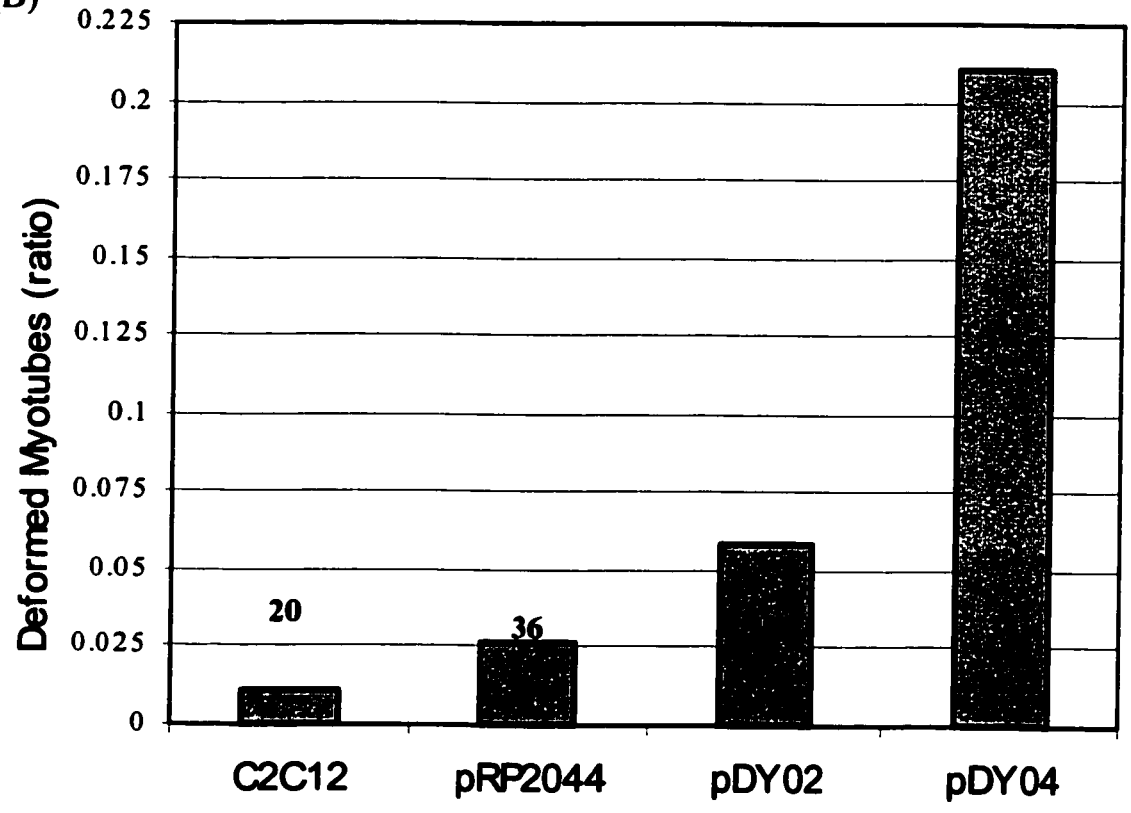
Figure 18. C2C12 cells transiently transfected with D4Z4 repeats exhibit abnormal myotube fusion.

(A) C2C12 cells transiently transfected with control (pRP2044) or the D4Z4 repeat-containing plasmids (pDY02 and pDY04) demonstrated a reduced myotube fusion potential as represented by the myotube fusion index (myotube incorporated nuclei/total nuclei) when compared to differentiation of untransfected wild type C2C12 cells. Bars represent the mean MFI of scored fields of view from 4 replicated plates (+/- SD). The number of scored fields is written inside each bar. (B) Total myotubes were counted and plotted as a ratio of normal (elongated, with centralized nuclei as shown in figures 9 and 10) to abnormally formed myotube structures (swelling with non-centralized masses of nuclei as pointed out in figure 10). D4Z4 transfected cells (pDY02 and pDY04) displayed a greater number of deformed myotubes compared to controls (wild type C2C12 and pRP2044). Bars represent the mean DMI of scored fields of view from 4 replicated plates. The total scored fields for samples is written inside each bar.

(A)


0.9
0.75
0.6
0.45
0.3
0.15
0
Incorporated nuclei (ratio)
C2C12
pRP2044
pDY02
pDY04

(B)


0.225
0.2
0.175
0.15
0.125
0.1
0.075
0.05
0.025
0
Deformed Myotubes (ratio)
20
36
C2C12
pRP2044
pDY02
pDY04

The proportion of incorporated nuclei from unincorporated nuclei provides an accurate measurement for the relative fusion competence of the culture and is presented as the Myotube Fusion Index (MFI [209]). Figure 18(A) shows that the MFI ranges from 70% in wild type C2C12 cells down to approximately 20% in D4Z4 containing transfections (pDY02 and pDY04). However, there is also an approximately 2-fold decrease in myotube fusion between C2C12 control and the pRP2044 cultures suggesting that transfection alone, in the absence of D4Z4 repeats also disrupts normal myotube fusion. Upon completion of the MFI analysis, we observed the appearance of a trend in myotube formation; not only was there a decrease in fusion competence (MFI), but it correlated with an increase in aberrantly formed myotubes in cultures. To quantitatively analyze these results, we developed the Deformed Myotube Index (DMI). The DMI is the relative proportion of myotubes (MF20 positive with three or more incorporated nuclei) which have the abnormal phenotype compared to normal myotubes. Figure 18(B) shows the observed trend of increasing deformity of myoblast fusion in transient cultures. We observed that wild type C2C12 cells have a low rate of deformed myotubes at approximately 1%. There is a small increase in deformed myotubes in the pRP2044 transfected cells (2.6%) suggesting that transfection with non-D4Z4 plasmids has little effect in the formation of deformed structures. However, we observed a very significant increase in the rate of deformed myotube formation between pDY02 (6%) and pDY04 (21%). These trends suggest that while the addition of D4Z4 repeats in both pDY02 and pDY04 confer a 4-fold decrease in ability of C2C12 to form myoblasts (Figure 18(A)), there is a significant increase in the rate of abnormal myotube formation, 6-fold for pDY02 and 21-fold for pDY04 (Figure 18(B)). *In situ* β-gal staining shows that transfected cells contributed to both differentiated myotubes (normal and deformed) and undifferentiated myocytes (data not shown).

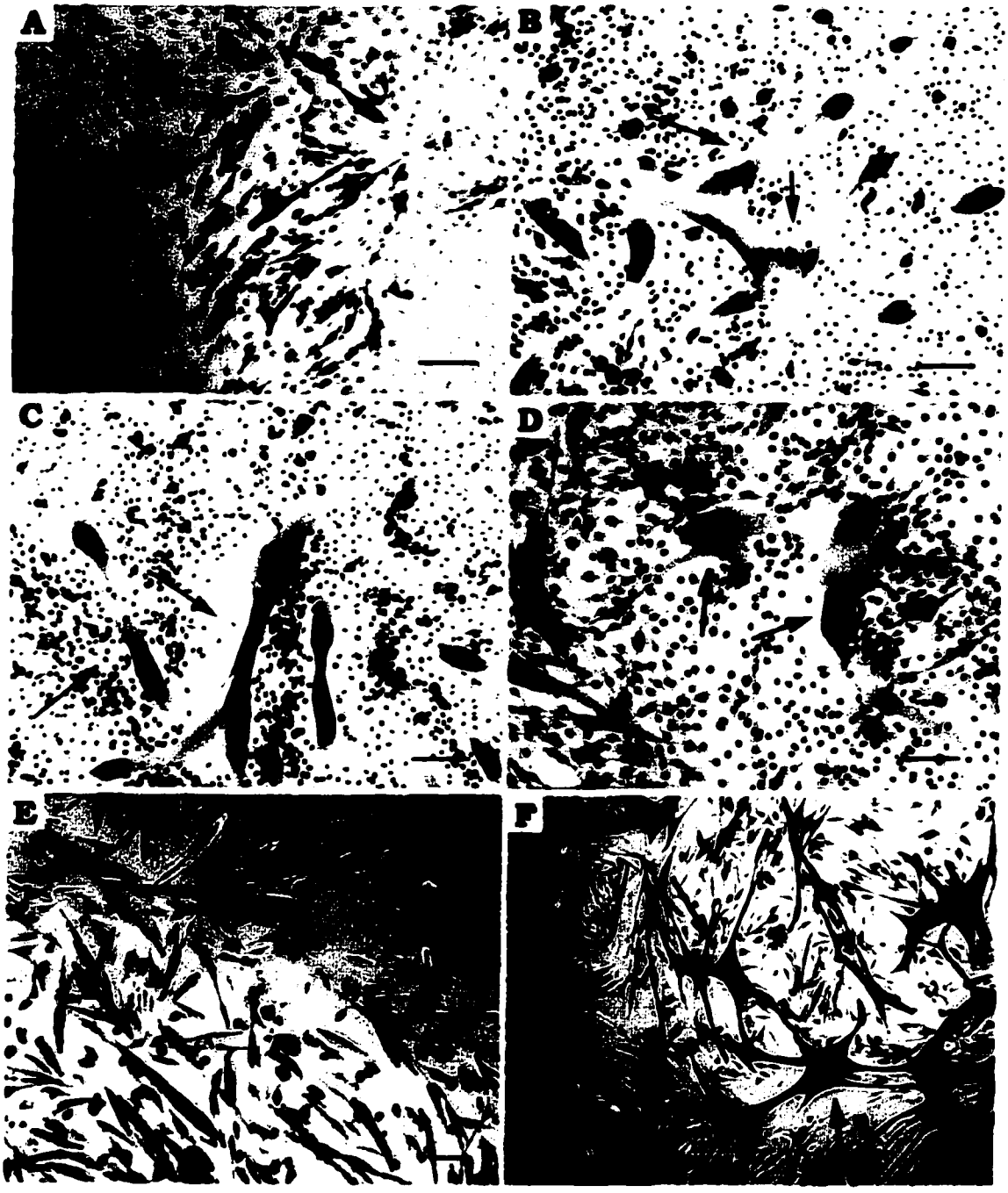
3.4.2 – Analysis of myoblast fusion competence and morphology in stable clones.

To confirm that the fusion defect observed in transient assays were a result of the D4Z4 repeats, we repeated this experiment with the stable clones. Figure 19 shows a selection of photographs depicting differentiated stable clones with low fusion competence and a higher rate of aberrant fusion morphology. Figure 19(A), (B), (C) and (D) show clones cRP20444, cDY042, cDY042 and cDY04n3. MF20 positive myotubes are stained brown. In clones containing the D4Z4 repeat, a higher percentage of myotubes are formed with much larger masses of concentrated nuclei and no apparent structural uniformity (denoted by arrows). From Figures 19(B), (C) and (D) it is also apparent that there are far fewer differentiated myotubes; with normal morphology (compare to Figure 19(A)). In addition, the morphological defect is more pronounced for the stable clones than the transiently transfected cultures (compare Figures 16, 17 and 19). Figure 19(E) and (F) show that all cells from undifferentiated myoblasts to the multinucleated myotubes of clones cRP2044n5 and cDY04n2 are positive for β-gal activity following *in situ* staining. It should be mentioned that some clones (cRP20442, cRP2044n15 and cDY044) fail to differentiate at all, and continue to grow and divide under low mitogen conditions (data not shown).

To quantitatively assess these defects we performed MFI and DMI calculations from pDY04 stable clones and the control (pRP2044) clones and compared them to C2C12 cells (Figure 20). In these experiments we observed that the average MFI for wild type C2C12 cells was approximately 60%. Excluding clone cRP2044n5 (discussed below), we see a consistent 2-fold decrease in myotube fusion between wt C2C12 and pRP2044 clones

Figure 19. Morphological analysis of myotube formation in D4Z4 stable clones.

Phase microscopy images of stable clones cRP20444, cDY402, cDY042, cDY04n3 cRP2044n5 and cDY04n2 (a-f), respectively, were taken following a 96 hour serum withdrawal-induced differentiation. In Panels (A, B, C, and D) myotube structures were immunostained (brown) with a Myosin Heavy Chain antibody (MF20). Nuclei are counter stained purple with Harris modified Hematoxylin. Panels (E and F) show that β-gal activity (blue) is observed in myotubes and undifferentiated myoblasts alike. Arrows denote specific myotube structures with abnormal or deformed morphology (swelling with non-centralized masses of nuclei). Bars represent approximately 93μm.


A
B
C
D
E
F

(cRP2044n7 and cRP2044n31), suggesting that stable clones derived from non-D4Z4 plasmids also have an effect on myotube fusion. In stable clones containing pDY04 (cDY042, cDY04n2 and cDY04n/3) we saw a 3-fold decrease in the myotube fusion index compared to C2C12 cells (60% reduced to ~20%). Unincorporated myoblasts from clone cRP2044n5 (denoted by *, Figure 20(A)) displayed an abnormally high level of apoptosis under differentiation conditions, and as such the MFI value was significantly skewed in the absence of unincorporated nuclei. It has been shown previously that the low mitogen conditions necessary for differentiation cause myoblasts to be at a higher risk for apoptosis [213-215, 221]. The clone cRP2044n5 is especially susceptible to this form of programmed cell death. Despite the mild effect that D4Z4 repeats have on the MFI, it is quite clear from the phenotype and DMI values presented in (Figure 20(B)) that the D4Z4 repeats are affecting the morphology of the myotubes. Similar to transient studies, the wt C2C12 cells have a low DMI value of approximately 1%. The pRP2044 stable clones display an elevated DMI value of 9.5% (average from clones cRP2044n5, cRP2044n7 and cRP2044n31). Interestingly, the percentage of deformed myotubes in pDY04 clones was greatly elevated over C2C12 cells and pRP2044 clones to an average of 55% (average of clones cDY042, cDY04n2 and cDY04n/3). The similar trends observed between transient and stable assays involving the D4Z4 repeats suggest that a common disruption in myotube formation is correlated to increasing D4Z4 repeats.

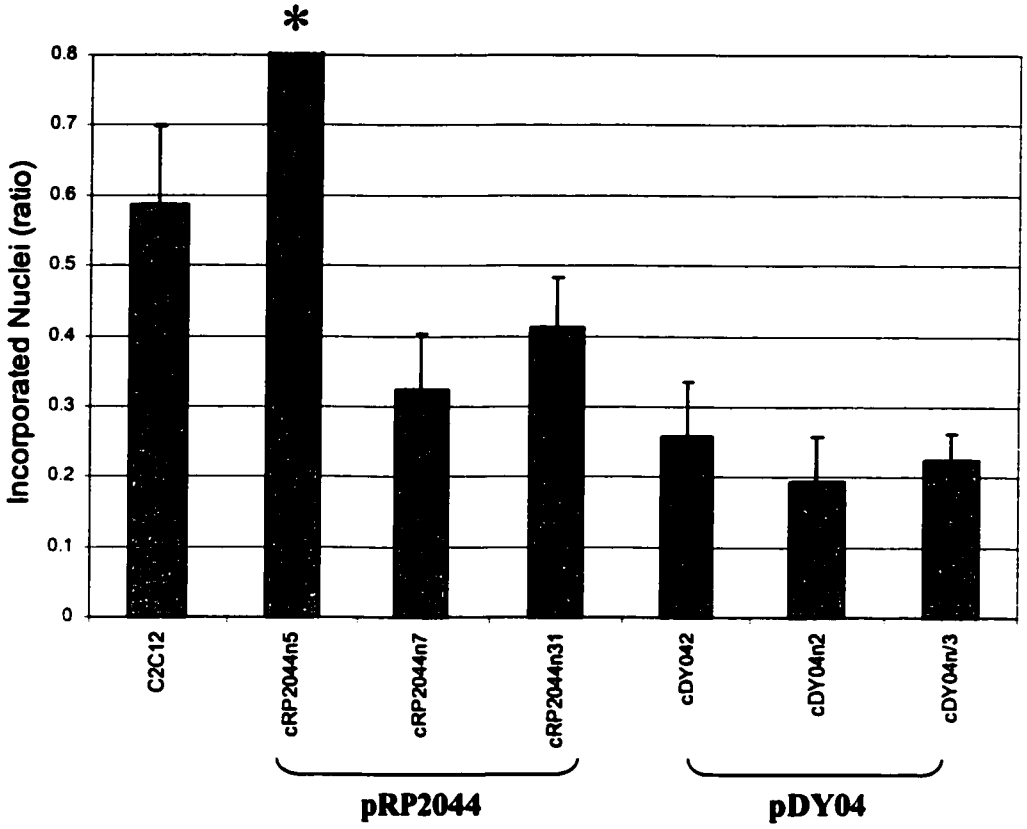
3.5 – Is there a D4Z4 specific sequence expressed in stable clones?

In an attempt to elucidate the molecular defect that the D4Z4 repeats present in differentiating myoblasts, we undertook a thorough investigation of the D4Z4 sequence to

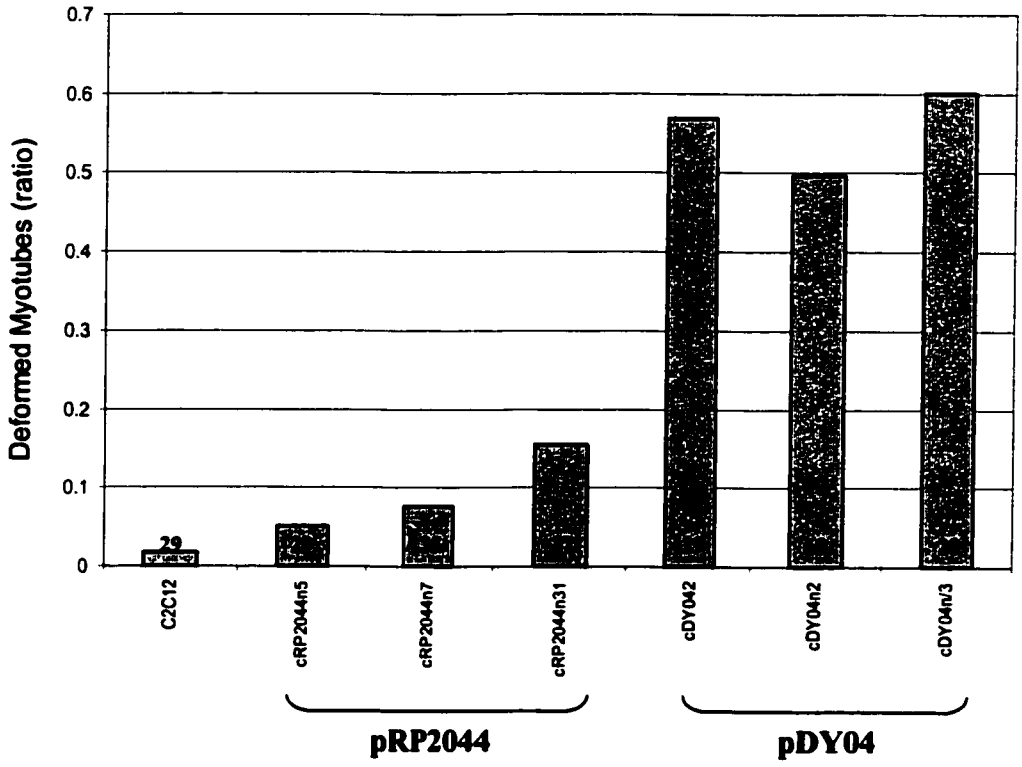
Figure 20. Stable clones containing D4Z4 repeats exhibit an abnormal myotube fusion.

(A) A selection of stable clones containing control plasmid DNA (cRP2044n5, cRP2044n7 and cRP2044n31) or D4Z4 repeat containing experimental plasmid (cDY042, cDY04n2 and cDY04n/3) demonstrated a reduced myotube fusion potential as represented by the myotube fusion index (myotube incorporated nuclei/total nuclei) when compared to differentiation of untransfected wild type C2C12 cells. Bars represent the mean MFI of from 6 (or more) replicated plates (+/- SD). The total number of scored fields for each sample is written inside each bar. Unincorporated myoblasts from clone cRP2044n5 (denoted by *) displayed an abnormally high level of apoptosis under differentiation conditions and as such the MFI value was significantly skewed. (B) Total myotubes from stable cultures were counted and plotted as a ratio of normal (elongated, with centralized nuclei as shown in figures 9 and 12) to abnormally formed myotube structures (swelling with non-centralized masses of nuclei as pointed out in figure 12). D4Z4 repeat containing clones (cDY042, cDY04n2 and cDY04n/3) displayed a greater number of deformed myotubes compared to controls (wild type C2C12, cRP2044n5, cRP2044n7 and cRP2044n31). Bars represent the mean DMI from 6 (or more) replicated plates. The total number of fields scored for each sample is written inside the bar.

(A)


*
Incorporated Nuclei (ratio)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
C2C12
cRP2044n5
cRP2044n7
cRP2044n31
cDY042
cDY04n2
cDY04n/3
pRP2044
pDY04

(B)


Deformed Myotubes (ratio)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
29
C2C12
cRP2044n5
cRP2044n7
cRP2044n31
cDY042
cDY04n2
cDY04n/3
pRP2044
pDY04

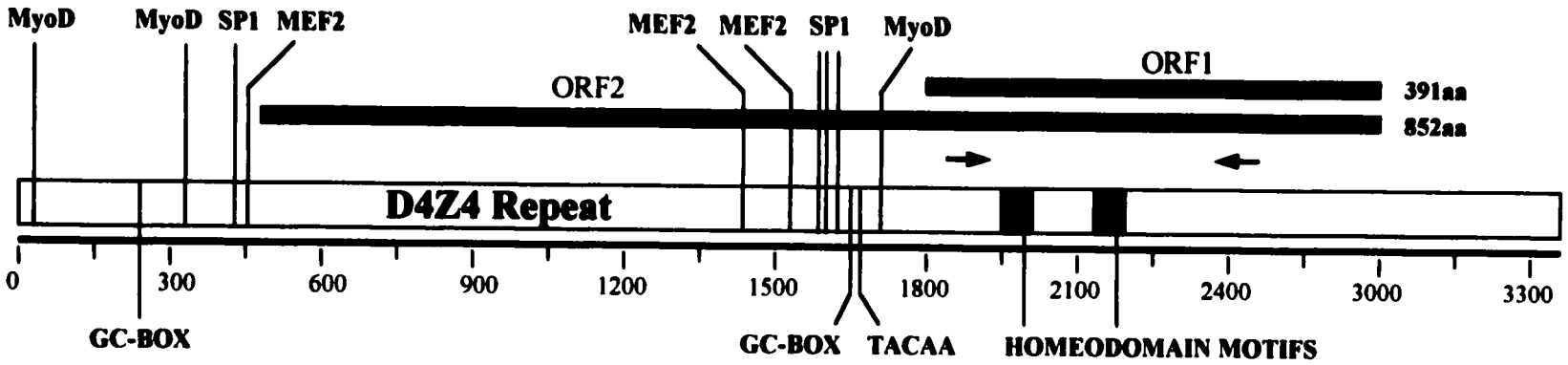
search for potentially expressed regions. Observations from the DMI calculations, which showed that the percentage of deformed myotubes was correlated to the presence of D4Z4 repeats, suggested that under differentiation conditions a sequence might be expressed from the D4Z4 repeats which lie in the pDY02 and pDY04 constructs. Several groups cite that there may be a gene encoded within the D4Z4 repeat, yet there have been no expressed sequences isolated [35, 40, 49, 51, 65-67]. Despite the fact that there has not been an expressed sequence isolated from the D4Z4 repeat, this still remains a viable theory in the pathogenesis of FSHD. A reduction of repeats below a certain threshold may diminish the potential D4Z4 repression, allowing for ectopic expression of D4Z4 transcripts, which may be toxic to myofibres.

3.5.1 – Computational and sequence analysis of the D4Z4 repeat.

Two large open reading frames (ORF), which incorporate the double homeodomain motif, have been identified within the sequence of the D4Z4 repeat. Figure 21 provides a schematic overview of the results from computational analysis of the D4Z4 repeat sequence. ORF1 (blue bar) was first described by Gabriels et al., [51]. They found a large open reading frame that encoded a putative 391 amino acid peptide, incorporating the homeobox motifs corresponding to the putative gene *DUX4* [2, 51, 66]. Upstream from this ORF lies a putative promoter region that includes two MEF2 and a MyoD binding site, a GC-box, a TACAA box and 3 consensus-binding sites for the ubiquitous transcription factor SP1. In their studies Gabriels et al., [51], isolated this promoter region and found that it could clearly drive gene expression, using the luciferase reporter. Our sequence analysis revealed a second larger ORF, which encompasses ORF1. ORF2 (red bar) encodes a putative 852-

Figure 21. Schematic diagram of the D4Z4 repeat displaying the position of two potential Open Reading Frames (ORFs) and their putative promoters.

Each complete 3.3kb D4Z4 sequence (Acc. D38024) represented by the yellow bar contains two potential ORFs, which incorporate the homeodomain motifs (black boxes). ORF1, represented by the blue bar was proposed by Gabriels et al. [25]; it encodes a predicted 391 amino acid product. Our analysis detected a second open reading frame, ORF2 (represented by the red bar). ORF2 encodes a predicted 852 amino acid product. Putative promoter elements are located upstream of each ORF. The position of GC-boxes, MEF2 and MyoD consensus binding sites and the TACAA box (specific to OFR1) are labeled. The positions of consensus binding sites for the ubiquitous transcription factor SP1 are also shown. The placement of D4Z4 primers D4Z4-f1905 and D4Z4-r2385 used in RT-PCR experiments are shown as arrows.


MyoD
MyoD
SP1
MEF2
MEF2
MEF2
SP1
MyoD
ORF2
ORF1
391aa
852aa
D4Z4 Repeat
0
300
600
900
1200
1500
1800
2100
2400
3000
3300
GC-BOX
GC-BOX
TACAA
HOMEODOMAIN MOTIFS

residue polypeptide. Upstream of this ORF lies a putative promoter region, which includes two MyoD and a MEF2 binding site, and a GC-box and a SP1 consensus-binding site.

In order to determine if the D4Z4 repeat possesses potentially repressive elements, we screened the D4Z4 sequence for possible Polycomb Response Elements (PRE). Polycomb (PcG) proteins have often been associated with repression of homeotic genes [108, 171, 172]. Using the consensus binding sequence (Acc. L32205 and [188]) for the Drosophila PcG gene *pleiohomeotic* (*pho*), the only known PcG shown to actively bind to DNA, we found four potential clusters within the D4Z4 repeat. Subsequent screening for GAGA factor binding sites [189, 191, 194] showed overlap with these four regions. Previous examples have shown that these two binding sites are found commonly in association at PREs, suggesting that *pho* (or its human homologue YYI) and GAGA factor acts as recruitment factors for PcG complexes [108, 174, 188, 189]. This evidence suggests that the D4Z4 repeats may possess repressive elements necessary for the formation of a stable repressive structure normally required for the control of proximal gene regulation.

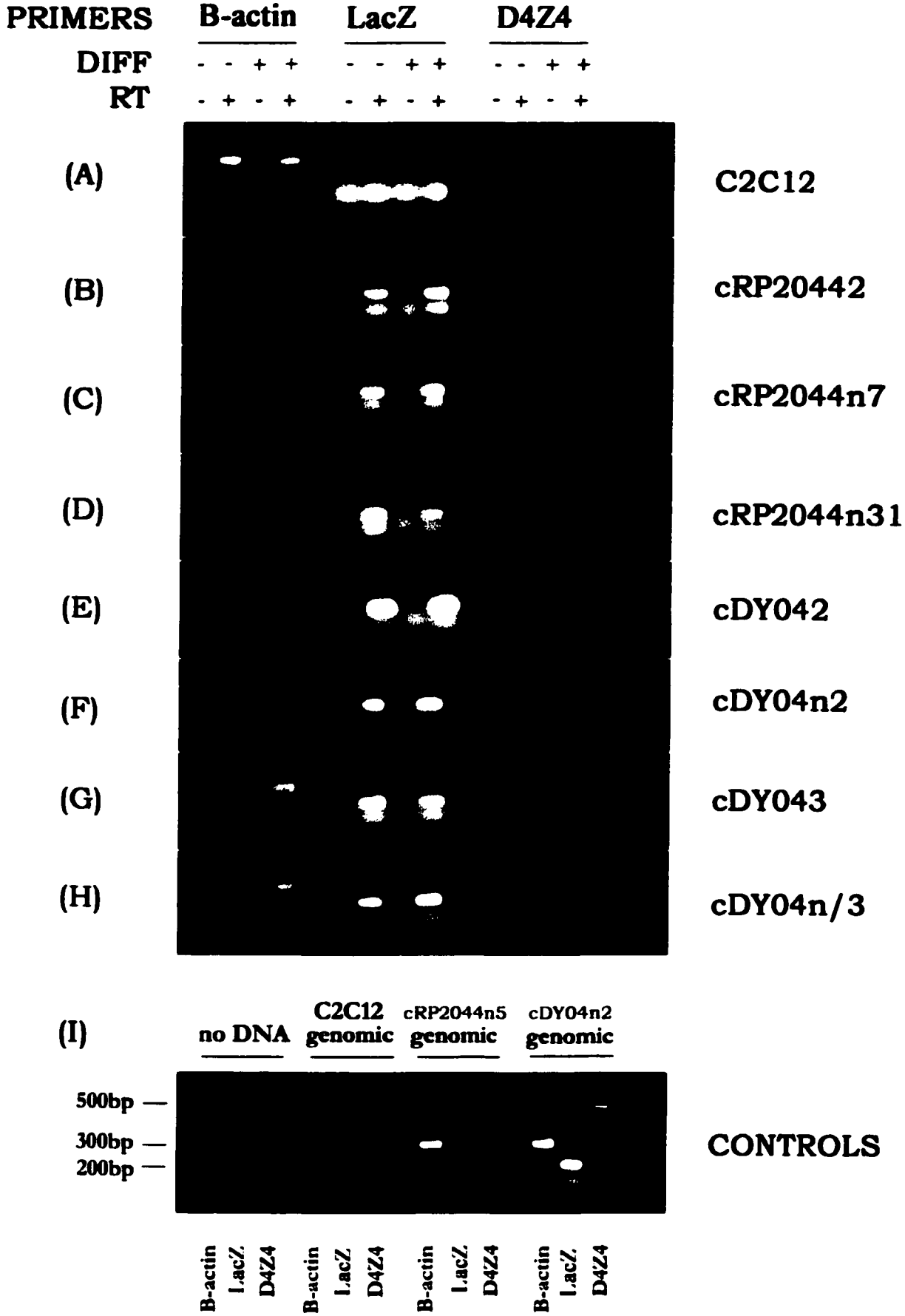
3.5.2 – Is the *DUX4* sequence expressed during myogenic differentiation?

We hypothesized that the resulting increase in DMI may be the result of ectopic expression of the putative homeobox gene (termed *DUX4*) contained within each D4Z4 repeat [51, 66, 87]. To assess this, RNA was prepared from C2C12 cells and a selection of stable clones during growth or differentiation conditions (96 hours). Utilizing RT-PCR and primers specific to a non-repetitive region common to both ORFs (shown as arrows on Figure 21), we analyzed the clones for expression from the D4Z4 repeats as shown in Figure

22. Panel I shows the control reactions for the PCR primers. An approximately 300 bp band corresponding to the external control β-actin gene is amplified from genomic DNA isolated from C2C12 cells and stable clones cRP2044n5 and cDY04n2. A 200 bp LacZ specific band was only amplified from the genomic DNA isolated from stable clones and a 500 bp D4Z4 specific band was only amplified from genomic DNA isolated from pDY04 clones. There was no PCR amplification from the negative control lanes using mock treatments with no DNA (blank). Panels A though H represent amplified products derived by RT-PCR from RNA isolated from C2C12 and a selection of stable clones. These clones contained either the D4Z4 containing plasmid (pDY04) or the control plasmid (pRP2044) and were grown under growth or differentiation conditions (+/- DIFF). The +/- RT columns correspond to experimental reactions (+ Reverse Transcriptase) or negative control reactions (-RT) to test for genomic DNA contamination. There was no amplification in any -RT reaction columns. The external control primers directed toward β-actin amplified a product in all clones and C2C12 cells under growth and differentiation conditions demonstrating the quality of all cDNA samples. The internal control primers directed toward the LacZ gene (contained within each construct) amplified a specific band under both growth and differentiation conditions. However we were unable to amplify the D4Z4 specific sequence corresponding to ORF1 and ORF2 (Figure 22) in any sample. These results suggest that the effect of D4Z4 repeats on myotube differentiation is not due to expression from within the D4Z4 repeat.

Figure 22. D4Z4 specific sequence is not expressed in stable clones during growth or differentiation.

Gel electrophoresis images of RT-PCR analysis for β-actin, LacZ and D4Z4 specific sequences. β-actin competimers (Ambion Inc.) were used as an external control for cDNA synthesis. LacZ primers (RAS 39 and RAS 40) represent an internal control for reporter expression and the D4Z4 specific primers (D4Z4-f1905 and D4Z4-r2385) were used to amplify an expressed sequence. Panels (A) through (H) represent analyses of expression using RT-PCR on total RNA isolated before and following differentiation of C2C12 (control) from a selection of stable clones (listed on the right). +/- DIFF refers to cDNA samples made from RNA isolated from differentiated (+) or untreated cells (-). There was no amplification in samples that lacked (-) Reverse Transcriptase (RT), a control for genomic DNA contamination. Panel (I) represents control reactions using genomic DNA from C2C12 or stable clones (cRP2044n5 and cDY04n2) or mock reactions with no DNA. The placement of molecular weight size markers is shown on the left.


PRIMERS
B-actin
LacZ
D4Z4
DIFF
- - + + - - + + - - + +
RT
- + - + - + - + - + - +
(A)
C2C12
(B)
cRP20442
(C)
cRP2044n7
(D)
cRP2044n31
(E)
cDY042
(F)
cDY04n2
(G)
cDY043
(H)
cDY04n/3
(I)
no DNA
C2C12 genomic
cRP2044n5 genomic
cDY04n2 genomic
500bp
300bp
200bp
CONTROLS
B-actin
LacZ
D4Z4

4.0 – Discussion

The main focus of this project was to determine if the D4Z4 repeats function as repressive sequence elements. It has long been known that the disease-causing mutation for FSHD is a deletion of an integral number of repeats from the D4Z4 array on chromosome 4q35 to below a critical threshold of 10 repeats [37-39]. The genotype-phenotype correlation between loss of repeats and severity and age of onset is evidence of D4Z4 involvement; however the exact mechanisms by which these repeats cause the disease and the molecular outcomes of the deletion are not quite so clear [3, 5]. Our working hypothesis was that a reduction of the number of repeats below a critical threshold level of 10 prevents the formation of a stable repressive heterochromatin complex. The inability to form a heterochromatin complex would affect the genomic environment of the surrounding region, ultimately altering local gene expression [64, 65, 86]. Although this is the favoured model in the FSHD field, there is no substantial evidence to suggest that the D4Z4 repeats are repressive in nature, or even that they facilitate the formation of heterochromatin. As such, we decided to isolate and clone copies of the D4Z4 repeats into a BOS-LacZ reporter construct to determine if they can directly repress the transcription of the reporter gene following transient and stable transfections in C2C12 mouse myoblasts.

4.1 – D4Z4 repeats do not repress gene expression in transient or stable clones under growth conditions.

Initially we performed transient experiments with 0-4 copies of the D4Z4 repeats adjacent to the reporter gene. We observed no significant change in LacZ expression,

suggesting that the D4Z4 repeats do not function as a classically defined repressor element. However, it is clear that the D4Z4 locus is involved, as the clinical phenotype associated with FSHD is only present in individuals with 10 or less repeats [37, 43, 44]. If the threshold number of repeats is necessary for a repressive effect on neighbouring genes, then the 4 copies used in these experiments may not have been sufficient and could explain why we detected no differences. Yet in the genomic context of FSHD the nearest genes on 4q35, *FRG*1 (100 kilobases) and *FRG*2 (37 kilobases) lie a significant distance proximal to the repeats [45, 49, 86]. In the design of our experiments we postulated that the D4Z4 threshold may be necessary for repression over a larger area, however, we surmised that a minimal number of repeats would be sufficient to influence expression of the reporter gene, which lies 100 base pairs from the repressive elements. Nevertheless, repression of the reporter was not observed in transient assays. It is therefore likely that a repressive complex originating from the D4Z4 repeats is based on a requirement for a defined number of repeats and suggests a mechanism involving Polycomb response elements (see below).

An alternate possibility was that repression is dependent upon the integration of the plasmid to establish a chromatin domain [89, 156]. To assess whether D4Z4 repression required such an event we isolated stable clones containing pDY04 and pRP2044 DNA for analysis of reporter gene activity. We observed a high degree of variability in β-gal activity between clones. This was not surprising, as sequences and chromatin structure adjacent to the site of integration can influence promoter activity; this situation is commonly referred to as position effect [163, 217, 219]. It was noted that during the generation of stable clones that the ratio of control clones positive for the LacZ gene (by PCR) and exhibiting β-gal

activity was higher than that observed for pDY04 generated clones (Table 2). This suggests that the stable integration of the control plasmid was more efficient than pDY04 and that the pDY04 clones may be more susceptible to position effects and transgene silencing through heterochromatinization.

Our original goal was to generate clones containing single-copy transgene insertions, yet previous studies have shown that stable transgenes have a tendency to insert as multiple arrays rather than single copies [164-166]. These researchers have also shown that there are several molecular and cellular mechanisms, which result in position effect variegation. The permanent repression of many of these copies occurs through spreading from a neighbouring constitutive heterochromatin domain or from self-forming facultative heterochromatinization [100, 162, 222]. Some common examples are homology-dependant silencing or repeat-induced gene silencing (RIGS), through DNA:DNA pairing resulting in increased methylation and leading to transcriptional repression [164, 223, 224]. Moreover, Heterochromatin Protein 1 (HP1) and Polycomb group proteins have been shown to localize over a silenced transgene array, suggesting an active cellular process for inactivation involving the formation of heterochromatin [92, 165, 167, 174, 190, 225]. If the transgene arrays provide a natural target or a platform for heterochromatinization and transcriptional repression we would have expected a notable if not complete decrease of β-gal activity from clones containing multiple arrays of the plasmid construct. In our experiments, it is possible that multiple copy arrays could also be informative by providing a scenario more similar to a normal non-FSHD individual. Hypothetically, if an array contains three pDY04 (~21kb) insertions, it would contain not only three copies of the LacZ gene, but also 12 copies of the D4Z4 repeat within a 63kb locus. This puts the total number of potentially repressive

repeats in the region to above the 10-repeat threshold for FSHD [3, 28, 38]. For example, clone cDY04n2 contains 12 copies of the pDY04 construct at 1 insertion site, suggesting that there are 48 D4Z4 repeats present, which is significant as it may create conditions close to a normal non-FSHD genomic environment. From this clone we observed a very high level of expression under growth conditions, suggesting that the repeats do not repress reporter gene activity and this is consistent with our transient data.

We observed a high variability of activity between each clone (multiple and single integrations), so to directly assess the level of LacZ activity within a specific clone independent of position effects we generated the LacZ expression plasmids with LoxP consensus sequences flanking the inserted repeat arrays. This permitted the Cre site-specific recombinase to excise the D4Z4 repeats from each clone allowing us to search for consistent trends in gene expression between clones, independent of position effect. Following Cre excision we expected to observe an increase in LacZ expression; however, in clones cDY04n2 and cDY043, which contain multiple transgene arrays we observed an unexpected decrease in β-gal activity. The decrease in β-gal activity is likely a result of Cre-excision between LoxP sites from two different plasmids, such that the LacZ gene is also lost. Removal of several copies of the reporter gene, in addition to the D4Z4 repeats would also lead to a decrease in β-gal activity observed in ONPG assays. However, we also observed a decrease in clone cDY04n/23, which contains only one single insertion, suggesting that changes in reporter gene expression are not directly associated with the D4Z4 repeats and may be greatly influenced by their integration site. In addition, clones cDY042 and cDY04n/3 showed increases in β-gal activity following Cre-excision. Thus, no clear trends

of repression among stable clones under growth conditions or following Cre-excision were observed. The combined results from transient and stable assays suggested that the anticipated repression might only emerge upon myogenic differentiation or that the repeats themselves are not repressive in nature.

4.2 – D4Z4 repeats have no effect on gene expression, but do confer a reduced fusion competence and abnormal myotube morphology during C2C12 terminal muscle differentiation.

FSHD is a muscle disease; therefore we reasoned that repression by D4Z4 repeats may only become evident during myogenesis. Following transfection we cultured transiently transfected C2C12 cells in differentiation medium over the span of 96 hours. By determining β-gal activity at 12, 24, 48, 72 and 96 hours, we found that plasmids containing D4Z4 repeats lagged in their recovery of β-gal activity immediately following transfection. At the final time point (96 hours) a decrease in LacZ expression is clearly associated with the presence of D4Z4 repeats. When we differentiated stable clones for 96 hours we observed both significant increases and decreases in reporter gene expression among control and experimental clones suggesting that in our model the D4Z4 repeats do not directly repress gene expression. In the stable clones it is possible that position effects and chromatin rearrangements may be the cause of a D4Z4-independent effect on the reporter gene expression. Chromatin remodeling is a common form of gene regulation, and has specific importance during myogenesis [77, 226]. Several changes in chromatin structure, brought about by changes in methylation and acetylation status, have been associated with the process of muscle gene regulation [227]. This process is often utilized to establish a

cell-type-specific chromatin pattern, which regulates gene expression during the process of differentiation by providing an open or closed state for the accessibility of transcription factors [150]. Furthermore, the process of cell differentiation is actually the result of an equilibrium between silencing and activation that produces a set of active genes characteristic of a cell lineage [163]. A sudden change in chromatin structure of a region between growth and differentiation conditions results in an altered genomic environment, for example the reversal of a euchromatin domain to a heterochromatin domain may physically repress the transgene's expression. Alternatively, differentiation conditions prompt the expression of muscle specific transcription factors, which may elicit a D4Z4 response by activating specific muscle transcripts from within the repeat.

We observed that during the differentiation of both transient and stable clone cultures there was a lower efficiency of fusion that was inversely correlated to the number of D4Z4 repeats. We also observed a higher percentage of abnormally fused myotubes, which resulted in a globular or bag-like structure with a mass of concentrated nuclei (Figure 17 and Figure 19). In both stable and transiently transfected cells this trend was directly correlated to the presence of D4Z4 repeats. In differentiation assays of pDY04 transient and stable cultures, only a small percentage were capable of myotube fusion, as calculated by the proportion of nuclei incorporated into myotubes compared to total nuclei (the Myotube Fusion Index, MFI). While decreases in the MFI of both stable clones and transient cultures generated with the control plasmid pRP2044 were observed, these values were further decreased in the experimental lines and cultures, prompting the suggestion that the D4Z4 repeats do effect cell fusion. Typically fused C2C12 myotubes present as elongated structures with centralized nuclei [57, 58]. However, we observed a high incidence of

deformed myofibres in differentiating cultures containing D4Z4 repeats as calculated using the proportion of deformed structures from normal myotubes (the Deformed Myotube Index, DMI). In both stable cell lines and transiently transfected C2C12s we detected a sharp increase for DMI values in the presence of D4Z4 repeats. Furthermore in transient assays we observed this increase to be directly proportional to the number of D4Z4 repeats. Despite the similarity in trends between stable and transient experiments, one clear difference was noted; stable clones tended to produce abnormally large myotubes with several centralized nuclei while the transiently transfected cells produced larger numbers of smaller myotubes with less nuclei. This finding may represent the differences between a heterogeneous culture of transfected and wild type cells and a highly homogeneous stable clone.

Typically C2C12 cells exit the cell cycle and begin to differentiate and fuse into myotubes after 72 hours in differentiation media, or they eventually undergo programmed cell death (apoptosis) [213, 214, 221]. The regulation of muscle differentiation is coordinated by the precise control of overlapping muscle specific genes by the myogenic determination factors (MDF) [73, 206]. There are four basic MDFs: MyoD, myogenin, Myf-5 and MRF4 [69-71]. MyoD and Myf5 are important as early determination factors to initiate myogenic differentiation, followed consecutively by myogenin and MRF4, which maintain terminal differentiation ultimately providing resistance to apoptosis [213].

The deformed myotube structures were not present in the control C2C12 sample, and few cells with this morphology were observed in the pRP2044 clones, suggesting that the D4Z4 repeats are the cause of the phenotype of myotubes. Some of the characteristics

applied to the clinical diagnosis of FSHD are the presence of severe muscle inflammation, as well as disorganized and variably sized myofibres in patient biopsies [5, 12, 20]. In our culture model, the presence of the deformed myotube structures may be a manifestation of these defects, thus providing a physical link between FSHD phenotypes and the D4Z4 repeat. There are two possible mechanisms involving the D4Z4 repeats, which could account for the aberrant morphology. First, in FSHD cases the low repeat copy arrays may prevent the formation of a heterochromatin complex, creating a condition where, in the open capacity, there is a gene (*DUX4*) transcribed from within the repeats. This scenario would mimic a "gain of function" mutation created by the ectopic expression of a D4Z4 transcript from the deleted 4q35 chromosome region containing the short (< 35kb) allele and is consistent with the autosomal dominant pattern of inheritance observed among FSHD families [1, 147]. Secondly, perhaps the introduction of D4Z4 repeats, which contain putative MyoD and MEF2 binding sites, might interfere with the function of these transcription factors leading to disruptions in the differentiation potential of the cells. It is entirely possible that the ensuing morphology in our *in vitro* cell culture model are the result of trans-effects caused by the presence of open D4Z4 repeats. Perhaps the reason we observed a D4Z4 repeat-copy number dependent effect in deformed myotube fusion is the result of the D4Z4 repeat sequence acting as a sink for important factors, such as MyoD and MEF2. The idea that DNA repeats may cause abnormal cell fusion during C2C12 myoblast differentiation through squelching of transcription factors has previously been suggested for *in vitro* studies involving Myotonic Dystrophy [57, 58]. In these studies they found that transgene arrays of (CUG)n repeats alone could cause abnormal C2C12 fusion, possibly by interfering with the differentiation process, in a trans-dependant manner.

4.3 – The putative D4Z4 gene (*DUX4*) is not expressed in stable C2C12 clones during induced differentiation.

One of the current models for D4Z4 involvement in FSHD is that there is a gene transcript contained within the repeat sequence. It is proposed that the repeats are normally encased within a heterochromatin structure and that following the internal deletion this transcript is activated through the inability to be silenced from the limited number of repeats. Through sequence similarity between the D4Z4 homeodomains and other known Homeobox genes, the putative *DUX4* transcript is the focal point of this model [51, 66]. Moreover, a sequence corresponding to a putative promoter region has been identified within the D4Z4 repeat. This promoter has been shown to drive expression of a Luciferase reporter gene with high expression [51]. However, numerous attempts have been made to isolate this putative transcript, but to no avail [40, 43, 49-51, 65, 67]. In an attempt to elucidate the molecular defect causing FSHD, we screened the D4Z4 repeat sequence (Acc. D38024) for potential coding regions. We discovered a large open reading frame (ORF) which overlapped a shorter ORF previously described by Gabriels et al [51]. Both of these ORFs, contain two homeodomain motifs and potential promoter regions (including *MEF2*, *Myo*D, SP1 binding sites and GC-box sequences, see Figure 21). Interestingly, the sequence shift causing the *Bln*I restriction site specific to the 10q26 repeats also obstructs both ORFs by creating a stop codon just before the homeobox domains [35, 47, 48]. We hypothesized that the ectopic expression of the putative D4Z4 gene (*DUX4*) from the 4q35 repeats might result in the higher Deformed Myotube Index (DMI) observed in experiments using both stable clones and transient transfection experiments. However, we were unable to amplify either D4Z4 ORF following RT-PCR of RNA isolated from stable clones grown under

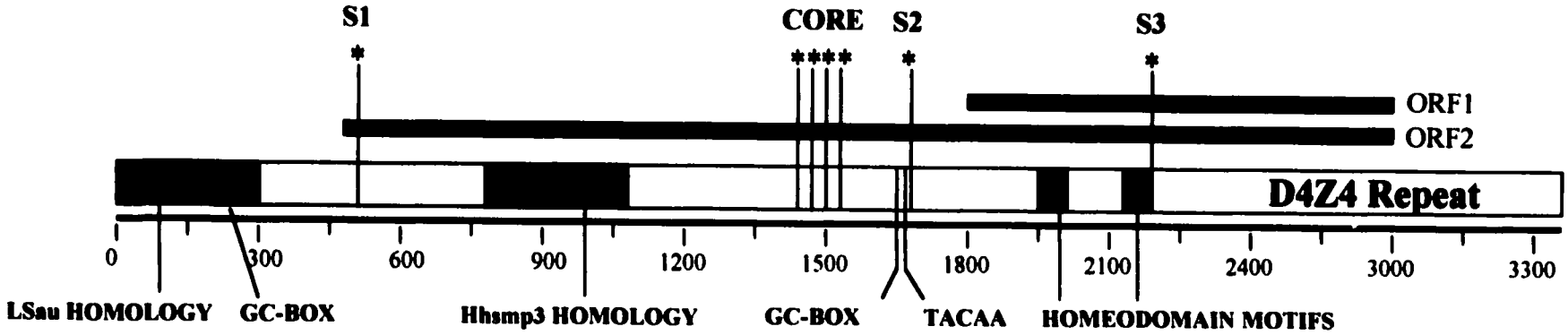
normal growth or differentiation conditions. These results suggest that the effects observed in myotube cultures are not the direct result of expression from within the D4Z4 repeat, but rather a function of the repeat sequence itself.

4.4 – The D4Z4 repeat and putative Polycomb-mediated repression in humans.

The D4Z4 repeats are located in a heterochromatin region, and were therefore proposed to represent a non-functional buffer from the telomere or, from later work, a domain important for correct regulation of downstream genes [51, 86]. The presence of two homeodomains within the repeats suggested that polycomb group (PcG) proteins might be associated with repression at 4q35 [19, 38, 40, 64, 171, 196]. Transcriptional regulation of HOX genes is mediated by the antagonistic roles of Polycomb group and Trithorax group proteins [160, 176]. Studies have shown that the deletion of key PcG proteins causes the upregulated or continued expression of HOX genes leading to mutations termed homeotic transformations [155, 171, 176, 178, 184]. In *Drosophila*, PcG complexes have been shown through both genetic and biochemical means to bind to regulatory units termed Polycomb Response Elements (PRE). PREs comprise a short sequence involving a concentration of overlapping binding domains for the PcG protein *pho* and GAGA factor [155, 170, 174, 188, 189, 228]. Our analysis uncovered four distinct regions containing the consensus binding sites for *pho* and GAGA factor within each of the 3.3 kb D4Z4 repeats (see Figure 23). The discovery of putative PREs within the D4Z4 repeat provides a means for the formation of a stable repressive complex involving the D4Z4 array.

Figure 23. The position of potential PcG binding sites and Polycomb Response Elements (PRE) and their relation to structurally important D4Z4 characteristics.

Each complete 3.3kb D4Z4 sequence (Acc. D38024) represented by the yellow and black bar contains four potential PREs regions (labeled as S1, S2, S3 or CORE). The position of consensus binding sites for the YYI Polycomb group protein (homologue of Drosophila *pho*) and the overlapping GAGA factor sites are both represented by asterisks (*). The homeodomain motifs and other sequence characteristics of the D4Z4 repeats are labeled (black boxes). ORF1 and ORF2 are represented by the blue and red bars. The position of GC-boxes and the TACAA box (specific to OFR1) are labeled.


S1
CORE
S2
S3
ORF1
ORF2
D4Z4 Repeat
0
300
600
900
1200
1500
1800
2100
2400
3000
3300
LSau HOMOLOGY
GC-BOX
Hhsmp3 HOMOLOGY
GC-BOX
TACAA
HOMEODOMAIN MOTIFS

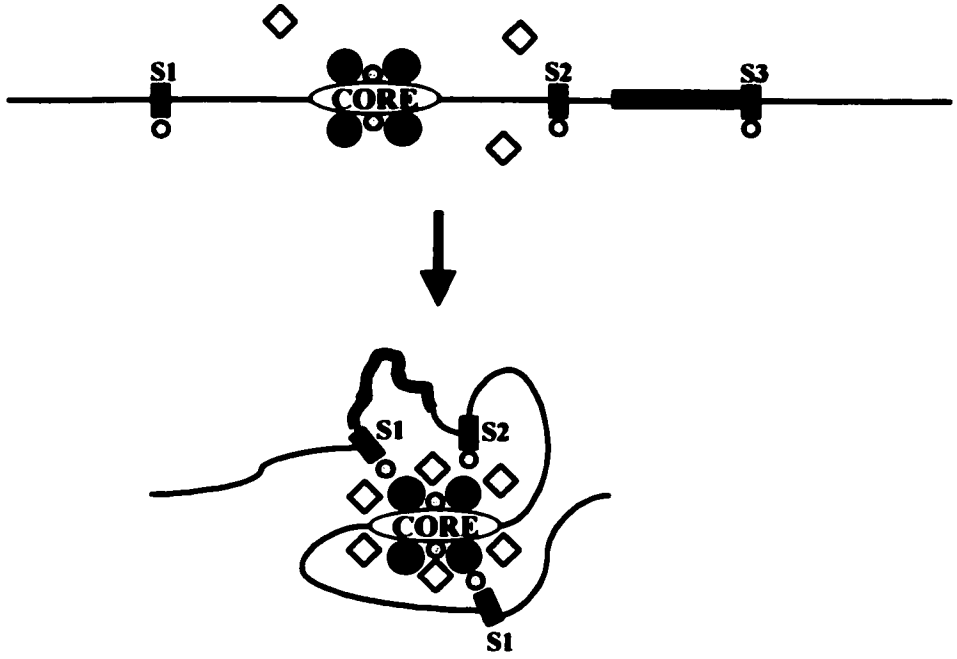
Pirotta and Rastelli [119] suggest that there are two forms of PREs: a strong central mediator or core site, which has a very high affinity for PcG complexes, and secondary "S"-sites which have less affinity and bind weaker and less stable complexes. The central core bound PcG complex initiates repression by interaction with the S-associated units, thus drawing the DNA into a looping bundle of DNA and protein [119, 132, 156]. By applying this model to the D4Z4 repeat, we propose that the central grouping of 4 putative binding sites is the central or "CORE" PRE and the three outlying single sites may be the so-called S-elements. Figure 24(A) shows how the combination of these elements may be the route to the formation of a repressed single repeat. Given that non-FSHD individuals have between 10-100 of the D4Z4 repeats, Figure 24(B) describes a scenario wherein multiple repeats may form a larger repressed region which might initiate an event involving the repression of a larger region. It has previously been noted that a combination of many PREs will work together to repress a larger area, suggesting that the D4Z4 arrays with many potential PREs are able to bind up a larger region than just a few repeats [122, 169]. We propose that the deletion of D4Z4 repeats causes FSHD by the inability to form this repressed heterochromatin domain. Inability to form the repressed structure results in inappropriate or continued expression of nearby genes. Such inappropriate expression could mimic a "gain of function" mutation leading to the FSHD phenotype.

In our transient and stable *in vitro* model we did not observe repression to be associated with the D4Z4 repeats, and two potential circumstances may be applicable. First, it is possible that our previous assumption that four repeats would be sufficient to repress over a short-range may have been incorrect. Perhaps, in order for the D4Z4 repeat to act as

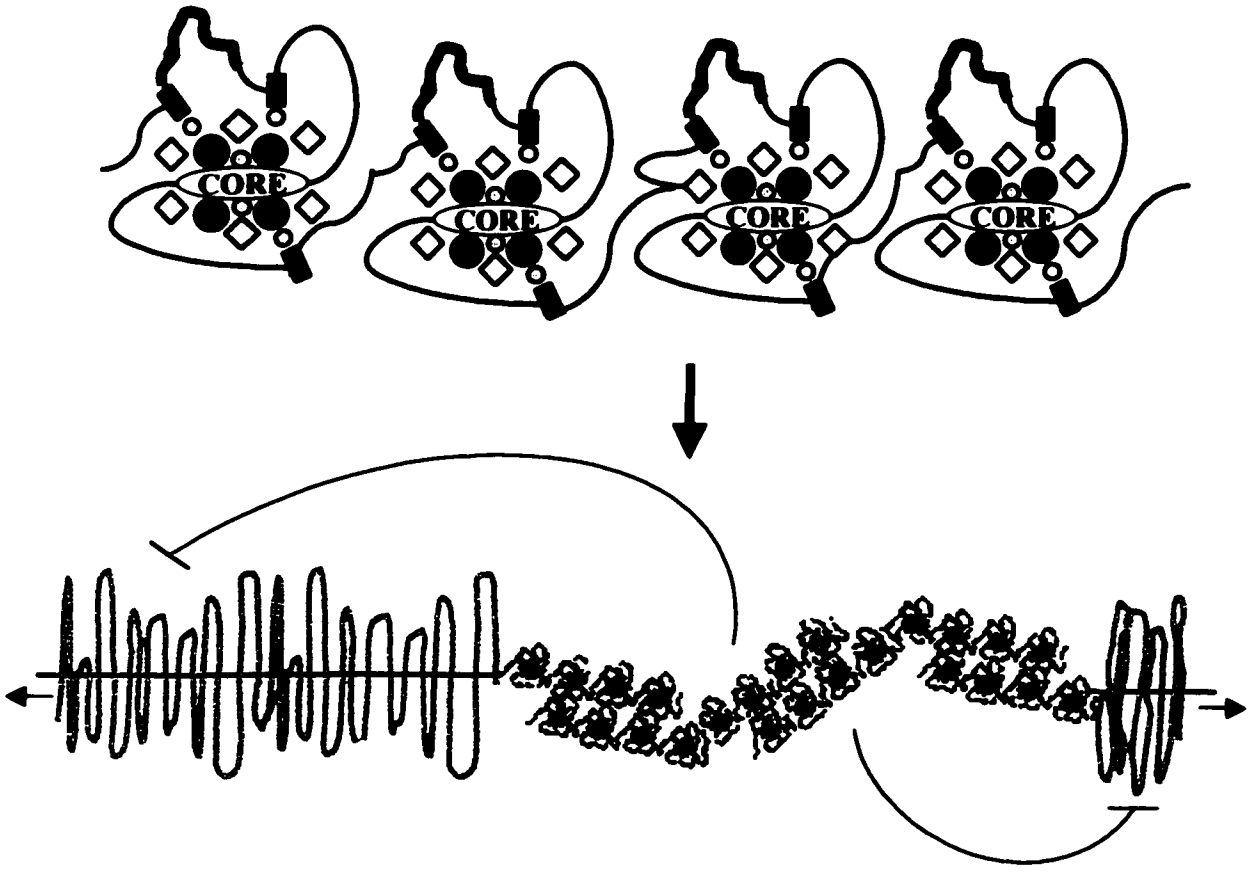
Figure 24. The assembly model for silencing complexes at the D4Z4 repeat and an example for cooperative interaction in the formation of facultative heterochromatin.

(A) This figure represents the cooperative assembly of interactive Polycomb Response Elements (PRE) involved in the repression of one D4Z4 repeat. In this model the D4Z4 repeat contains a central PRE (CORE) and three weaker PcG interacting "S"–elements; S1, S2 and S3 (purple boxes). Following initial recruitment of PcG complexes (green and blue circles and yellow diamonds) the weakly bound S-complexes are stabilized through interaction with the CORE, resulting in a bundle of DNA and protein. The gray box represents the region of the D4Z4 repeat containing the homeobox motifs. This D4Z4 looping model is based on a PRE model proposed by Pirotta and Rastelli [119]. (B) When several repeating units of the D4Z4 repeat are bound up together the effect creates a region of tightly bound DNA which creates the initial substrate necessary for the formation of facultative heterochromatin (thick blue line) over the surrounding region, repressing nearby genes.

(A)



(B)



the proposed catalyst for the formation of heterochromatin, there does indeed need to be a larger array of repeats, such that we were unable to observe repression. Secondly, it has been shown that the mouse genome does not contain D4Z4 repeats [81]. Even though Polycomb group-mediated repression and Polycomb response elements have been well documented in mice, they may not contain the subset of PcG proteins necessary for binding directly to the D4Z4 PREs specifically, however, mice do contain *pho* and GAGA homologues [171, 176, 179, 181, 229].

4.5 – Future directions.

The major limitation of this study has always been the number of D4Z4 repeats in the plasmid arrays. The normal non-FSHD individual has a D4Z4 array consisting of between 10-100 repeats lying hundreds of kilobases away from potential disease causing genes. We were only able to generate arrays containing a maximum of 4 repeats in our plasmids. We can argue that if each repeat is a separate repressive unit then four repeats may be sufficient to repress a promoter that lies one hundred base pairs away from the repeats. However, we cannot completely exclude the possibility that 4 repeats is insufficient to cause repression. In order to accurately assess this theory a future experiment may be to clone the BOS-LacZ reporter cassette into BAC or PAC clones generated from DNA containing various numbers of D4Z4 repeats, and to repeat the described experiments using larger repeat arrays.

Even though the current theories for FSHD suggest a role for the D4Z4 repeats in the formation of a facultative heterochromatin and repression, our experiments have not

provided a direct link between the two. Many questions need to be addressed in future experiments to completely explore the functional role of the D4Z4 repeats in Facioscapulohumeral Muscular Dystrophy. We have identified putative PREs within the D4Z4 repeat, and suggested a possible model for repression. Future experiments could be designed to functionally assess these sites for PcG binding potential. Two methods which may provide insight are electromobility shift assays (EMSA), which could be used to see if the D4Z4 PRE is a functional binding domain for the PcG protein YYI (*pho*) and/or the GAGA factor (*Trl*), the proposed PcG complex recruitment factors [170, 177, 186, 188, 189, 228]. A second potential means of assessing the validity of these PREs is to perform chromatin immunoprecipitations (ChIP) with the stable clones generated for this study or even endogenous human muscle cells. We would use an *in vivo* formaldehyde cross-linking technique to link DNA-bound proteins in place [230]. Through immunoprecipitation using PcG protein antibodies followed by PCR analysis or slot blot hybridization of the D4Z4-specific PRE site sequence, we would be able to further characterize these elements.

In order to follow up on the trans-effect theory for the obstruction of normal cell fusion during differentiation, we could make use of a quantifiable *in vitro* transcription assay. We would generate a G-Free transcript, which contains only A, C, and T residues, under the control of a muscle specific promoter. A muscle specific promoter, such as the *myogenin* promoter, must be chosen in order to provide competition with the D4Z4 repeat for muscle specific transcription factors, which include *Myo*D and *MEF*2 [231-234]. This vector would be incubated with nuclear extracts from C2C12 or pDY04 cells as a source of transcription factors to regulate transcription. Similar experiments have been used to examine the effect of various mutations on gene transcription and the activity of cis-

regulatory elements [235, 236]. We propose to use this system to test the theory that the D4Z4 repeats cause the abnormal myotube fusion phenotype by squelching away important transcription factors. In this reconstituted *in vitro* system the absence of GTP will drive the expression of only the G-Free transcript, and by adding radiolabeled UTP we have a means of quantification by acrylamide gel electrophoresis. In measuring the amount of G-Free product following the addition of purified D4Z4 repeats we would be able to determine if the D4Z4 repeat is able to effectively squelch out important factors as a result of the trans-effect, which is potentially the cause of the abnormal myotube fusion we observed. We are simply suggesting that if the D4Z4 repeats act as a competitor by creating a sink for transcription factors, then it would diminish the level of G-Free transcript expression.

4.6 – Conclusion.

For more than ten years, large deletions of the D4Z4 repeats have been associated as the causal mutation for Facioscapulohumeral Muscular Dystrophy (FSHD). Many mechanisms have been proposed; however there is relatively little functional data involving the D4Z4 repeat and its association with FSHD. The purpose of our investigation was to determine the function these repeats hold in the pathogenesis of FSHD. We therefore evaluated the potential regulatory effect that these repeats might have by cloning them into LacZ expression constructs and assessing the effects that these repeats may impart upon the reporter's expression in stable and transient transfection assays.

The results of this investigation have revealed inconclusive evidence to support a direct repressive effect on a reporter gene by the D4Z4 repeats in either transient transfections or stable clones grown under normal or myogenic differentiation conditions. However, during differentiation it was found that the D4Z4 repeats imparted an abnormal morphology onto the multinucleated myofibres. It was found that the presence of D4Z4 not only accompanied the reduced fusion competence of myoblast cultures resulting in the formation of less myotubes, but also corresponded to the heightened incidence of deformed "bag-like" atypical myotubes. In conclusion, while these experiments do not provide irrefutable evidence, they have provided some insight towards an association between the D4Z4 repeat and repression and have increased our understanding of the role D4Z4 repeats may play in the etiology of Facioscapulohumeral Muscular Dystrophy.

5.0 – REFERENCES

1. Padberg, G.W., et al., Facioscapulohumeral muscular dystrophy in the Dutch population. *Muscle Nerve*, 1995. **2**(4): p. S81-4.

2. Fisher, J. and M. Upadhyaya, Molecular genetics of facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscul Disord*, 1997. **7**(1): p. 55-62.

3. Lunt, P.W., et al., Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet*, 1995. **4**(5): p. 951-8.

4. Altherr, M.R., et al., Efforts toward understanding the molecular basis of facioscapulohumeral muscular dystrophy. Muscle Nerve, 1995. **2**(8): p. S32-8.

5. Lunt, P.W. and P.S. Harper, Genetic counselling in facioscapulohumeral muscular dystrophy. *J Med Genet*, 1991. **28**(10): p. 655-64.

6. Zatz, M., et al., High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *Am J Hum Genet*, 1995. **56**(1): p. 99-105.

7. Bakker, E., et al., Diagnostic, predictive, and prenatal testing for facioscapulohumeral muscular dystrophy: diagnostic approach for sporadic and familial cases. *J Med Genet*, 1996. **33**(1): p. 29-35.

8. Weiffenbach, B., et al., Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromsome 4q35 recombination events. *Nat Genet*, 1993. **4**(2): p. 165-9.

9. Passos-Bueno, M.R., et al., No evidence of genetic heterogeneity in Brazilian facioscapulohumeral muscular dystrophy families (FSHD) with 4q markers. *Hum Mol Genet*, 1993. **2**(5): p. 557-62.

10. Kissel, J.T., Facioscapulohumeral dystrophy. *Semin Neurol*, 1999. **19**(1): p. 35-43.

11. Furukawa, T., Neurogenic FSH muscular atrophy. *Muscle Nerve*, 1995. **2**(7): p. S96-7.

12. Brouwer, O.F., et al., Facioscapulohumeral muscular dystrophy in early childhood. *Arch Neurol*, 1994. **51**(4): p. 387-94.

13. Toniolo, D. and C. Minetti, Muscular dystrophies: alterations in a limited number of cellular pathways? *Curr Opin Genet Dev*, 1999. **9**(3): p. 275-82.

14. Hafner, F.M., et al., A novel locus (DFNA24) for prelingual nonprogressive autosomal dominant nonsyndromic hearing loss maps to 4q35-qter in a large Swiss German kindred. *Am J Hum Genet*, 2000. **66**(4): p. 1437-42.

15. van Overveld, P.G., et al., Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum Mol Genet*, 2000. **9**(19): p. 2879-84.

16. Padberg, G.W., et al., On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve*, 1995. **2**(80): p. S73-80.

17. Stout, K., et al., Somatic pairing between subtelomeric chromosome regions: implications for human genetic disease? *Chromosome Res*, 1999. **7**(5): p. 323-9.

18. Haraguchi, Y., et al., Genetic mapping of human heart-skeletal muscle adenine nucleotide translocator and its relationship to the facioscapulohumeral muscular dystrophy locus. *Genomics*, 1993. **16**(2): p. 479-85.

19. Funakoshi, M., K. Goto, and K. Arahata, Epilepsy and mental retardation in a subset of early onset 4q35- facioscapulohumeral muscular dystrophy. *Neurology*, 1998. **50**(6): p. 1791-4.

20. Wijmenga, C., et al., The human skeletal muscle adenine nucleotide translocator gene maps to chromosome 4q35 in the region of the facioscapulohumeral muscular dystrophy locus. *Hum Genet*, 1993. **92**(2): p. 198-203.

21. Tawil, R., et al., Extreme variability of expression in monozygotic twins with FSH muscular dystrophy. *Neurology*, 1993. **43**(2): p. 345-8.

22. Griggs, R.C., et al., Monozygotic twins with facioscapulohumeral dystrophy (FSHD): implications for genotype/phenotype correlation. FSH-DY Group. *Muscle Nerve*, 1995. **2**(5): p. S50-5.

23. van der Maarel, S.M., et al., De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am J Hum Genet*, 2000. **66**(1): p. 26-35.

24. Busse, K., et al., An inherited 4q35-EcoRI-DNA-fragment of 35 kb in a family with a sporadic case of facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscul Disord*, 2000. **10**(3): p. 178-81.

25. Saito, A., et al., An overexpression of fibroblast growth factor (FGF) and FGF receptor 4 in a severe clinical phenotype of facioscapulohumeral muscular dystrophy. *Muscle Nerve*, 2000. **23**(4): p. 490-7.

26. Arahata, K., et al., Inflammatory response in facioscapulohumeral muscular dystrophy (FSHD): immunocytochemical and genetic analyses. *Muscle Nerve*, 1995. **2**(66): p. S56-66.

27. Sandri, M., et al., Caspase 3 expression correlates with skeletal muscle apoptosis in Duchenne and facioscapulo human muscular dystrophy. A potential target for pharmacological treatment? *J Neuropathol Exp Neurol*, 2001. **60**(3): p. 302-12.

28. Vitelli, F., et al., Inheritance of a 38-kb fragment in apparently sporadic facioscapulohumeral muscular dystrophy. *Muscle Nerve*, 1999. **22**(10): p. 1437-41.

29. Felice, K.J., et al., FSH dystrophy 4q35 deletion in patients presenting with facial-sparing scapular myopathy. *Neurology*, 2000. **54**(10): p. 1927-31.

30. Tawil, R., et al., Scapuloperoneal syndromes. Absence of linkage to the 4q35 FSHD locus. *Arch Neurol*, 1995. **52**(11): p. 1069-72.

31. van Deutekom, J.C., Towards the Molecular Mechanism of Facioscapulohumeral Muscular Dystrophy. 1996, Leiden University: Leiden, The Netherlands.

32. Upadhyaya, M., et al., DNA marker applicable to presymptomatic and prenatal diagnosis of facioscapulohumeral disease. *Lancet*, 1990. **336**(8726): p. 1320-1.

33. Wijmenga, C., et al., Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet*, 1990. **336**(8716): p. 651-3.

34. Winokur, S.T., et al., A radiation hybrid map of 15 loci on the distal long arm of chromosome 4, the region containing the gene responsible for facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet*, 1993. **53**(4): p. 874-80.

35. Cacurri, S., et al., Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric KpnI repeat units implicated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet*, 1998. **63**(1): p. 181-90.

36. Grewal, P.K., et al., The mouse homolog of FRG1, a candidate gene for FSHD, maps proximal to the myodystrophy mutation on chromosome 8. *Mamm Genome*, 1997. **8**(6): p. 394-8.

37. Wijmenga, C., et al., Pulsed-field gel electrophoresis of the D4F104S1 locus reveals the size and the parental origin of the facioscapulohumeral muscular dystrophy (FSHD)-associated deletions. *Genomics*, 1994. **19**(1): p. 21-6.

38. Wijmenga, C., et al., Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet*, 1992. **2**(1): p. 26-30.

39. van Deutekom, J.C., et al., FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet*, 1993. **2**(12): p. 2037-42.

40. Lee, J.H., et al., Characterization of a tandemly repeated 3.3-kb KpnI unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. *Muscle Nerve*, 1995. **2**(13): p. S6-13.

41. Deidda, G., et al., Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet*, 1996. **33**(5): p. 361-5.

42. Upadhyaya, M., et al., Molecular analysis of British facioscapulohumeral dystrophy families for 4q DNA rearrangements. *Hum Mol Genet*, 1993. **2**(7): p. 981-7.

43. Hewitt, J.E., et al., Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet*, 1994. **3**(8): p. 1287-95.

44. Lemmers, R.J., et al., Inter- and intrachromosomal sub-telomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet*, 1998. **7**(8): p. 1207-14.

45. van Geel, M., et al., Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics*, 2002. **79**(2): p. 210-7.

46. van Deutekom, J.C., et al., Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum Mol Genet*, 1996. **5**(12): p. 1997-2003.

47. van der Maarel, S.M., et al., A new dosage test for subtelomeric 4;10 translocations improves conventional diagnosis of facioscapulohumeral muscular dystrophy (FSHD). *J Med Genet*, 1999. **36**(11): p. 823-8.

48. Bakker, E., et al., The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve*, 1995. **2**(44): p. S39-44.

49. van Geel, M., et al., The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. *Genomics*, 1999. **61**(1): p. 55-65.

50. van Deutekom, J.C., et al., Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. *Muscle Nerve*, 1995. **2**(26): p. S19-26.

51. Gabriels, J., et al., Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene*, 1999. **236**(1): p. 25-32.

52. Otten, A.D. and S.J. Tapscott, Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A*, 1995. **92**(12): p. 5465-9.

53. Wohrle, D., et al., A microdeletion of less than 250 kb, including the proximal part of the FMR-I gene and the fragile-X site, in a male with the clinical phenotype of fragile-X syndrome. *Am J Hum Genet*, 1992. **51**(2): p. 299-306.

54. Petersen, A., et al., Expanded CAG repeats in exon 1 of the Huntington's disease gene stimulate dopamine-mediated striatal neuron autophagy and degeneration. *Hum Mol Genet*, 2001. **10**(12): p. 1243-54.

55. Davies, S. and D.B. Ramsden, Huntington's disease. *Mol Pathol*, 2001. **54**(6): p. 409-13.

56. Cummings, C.J. and H.Y. Zoghbi, Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genomics *Hum Genet*, 2000. **1**: p. 281-328.

57. Amack, J.D. and M.S. Mahadevan, The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum Mol Genet*, 2001. **10**(18): p. 1879-87.

58. Amack, J.D., A.P. Paguio, and M.S. Mahadevan, Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum Mol Genet*, 1999. **8**(11): p. 1975-84.

59. Koob, M.D., et al., An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet*, 1999. **21**(4): p. 379-84.

60. Kawaguchi, Y., et al., CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet*, 1994. **8**(3): p. 221-8.

61. Holmes, S.E., et al., Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat Genet*, 1999. **23**(4): p. 391-2.

62. Wade, D.P., et al., 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc Natl Acad Sci U S A*, 1993. **90**(4): p. 1369-73.

63. Marcovina, S.M. and M.L. Koschinsky, Lipoprotein(a) concentration and apolipoprotein(a) size: A synergistic role in advanced atherosclerosis? *Circulation*, 1999. **100**(11): p. 1151-3.

64. Winokur, S.T., et al., The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum Mol Genet*, 1996. **5**(10): p. 1567-75.

65. Lyle, R., et al., The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics*, 1995. **28**(3): p. 389-97.

66. Ding, H., et al., Characterization of a double homeodomain protein (DUX1) encoded by a cDNA homologous to 3.3 kb dispersed repeated elements. *Hum Mol Genet*, 1998. **7**(11): p. 1681-94.

67. Kondo, T., et al., Whole-genome methylation scan in ICF syndrome: hypomethylation of non- satellite DNA repeats D4Z4 and NBL2. *Hum Mol Genet*, 2000. **9**(4): p. 597-604.

68. Lewin, B., Genes VI. 1997, New York: Oxford University Press.

69. Rawls, A. and E.N. Olson, MyoD meets its maker. *Cell*, 1997. **89**(1): p. 5-8.

70. Sabourin, L.A. and M.A. Rudnicki, The molecular regulation of myogenesis. *Clin Genet*, 2000. **57**(1): p. 16-25.

71. Arnold, H.H. and B. Winter, Muscle differentiation: more complexity to the network of myogenic regulators. *Curr Opin Genet Dev*, 1998. **8**(5): p. 539-44.

72. Molkentin, J.D., et al., Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell*, 1995. **83**(7): p. 1125-36.

73. Engel, A. and C. Franzini-Armstrong, Myology, Basic and Clinical. 2nd ed. Section 1: Development, Anatomy, and Cell Biology, ed. S.D. .Hauschka. 1993, Toronto: McGraw-Hill Inc.

74. Wheeler, M.T., et al., An E-box within the MHC IIB gene is bound by MyoD and is required for gene expression in fast muscle. *Am J Physiol*, 1999. **276**(5 Pt 1): p. C1069-78.

75. Lu, J., et al., Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. *Mol Cell*, 2000. **6**(2): p. 233-44.

76. Olson, E.N., M. Perry, and R.A. Schulz, Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. *Dev Biol*, 1995. **172**(1): p. 2-14.

77. McKinsey, T.A., C.L. Zhang, and E.N. Olson, Control of muscle development by dueling HATs and HDACs. *Curr Opin Genet Dev*, 2001. **11**(5): p. 497-504.

78. Chi, N. and J.A. Epstein, Getting your Pax straight: Pax proteins in development and disease. *Trends Genet*, 2002. **18**(1): p. 41-7.

79. Langston, A.W. and L.J. Gudas, Retinoic acid and homeobox gene regulation. *Curr Opin Genet Dev*, 1994. **4**(4): p. 550-5.

80. Finnerty, J.R. and M.Q. Martindale, The evolution of the Hox cluster: insights from outgroups. *Curr Opin Genet Dev*, 1998. **8**(6): p. 681-7.

81. van Geel, M., Facioscapulohumeral muscular dystrophy, a complex human disease entangled in dymanic molecular evolution. 2001, Leiden University: Leiden, The Netherlands.

82. Cserjesi, P., et al., MHox: a mesodermally restricted homeodomain protein that binds an essential site in the muscle creatine kinase enhancer. *Development*, 1992. **115**(4): p. 1087-101.

83. Bober, E., et al., Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development*, 1994. **120**(3): p. 603-12.

84. Goulding, M., A. Lumsden, and A.J. Paquette, Regulation of Pax-3 expression in the dermomyotome and its role in muscle development. *Development*, 1994. **120**(4): p. 957-71.

85. Tupler, R., et al., Profound misregulation of muscle-specific gene expression in facioscapulohumeral muscular dystrophy. *Proc Natl Acad Sci U S A*, 1999. **96**(22): p. 12650-4.

86. van Deutekom, J.C., et al., Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum Mol Genet*, 1996. **5**(5): p. 581-90.

87. Tsien, F., et al., Methylation of the FSHD syndrome-linked subtelomeric repeat in normal and FSHD cell cultures and tissues. *Mol Genet Metab*, 2001. **74**(3): p. 322-31.

88. Eissenberg, J.C. and S.C. Elgin, The HP1 protein family: getting a grip on chromatin. *Curr Opin Genet Dev*, 2000. **10**(2): p. 204-10.

89. Cockell, M. and S.M. Gasser, Nuclear compartments and gene regulation. *Curr Opin Genet Dev*, 1999. **9**(2): p. 199-205.

90. Karpen, G.H., Position-effect variegation and the new biology of heterochromatin. *Curr Opin Genet Dev*, 1994. **4**(2): p. 281-91.

91. Lohe, A.R. and A.J. Hilliker, Return of the H-word (heterochromatin). *Curr Opin Genet Dev*, 1995. **5**(6): p. 746-55.

92. Nielsen, A.L., et al., Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. *Mol Cell*, 2001. **7**(4): p. 729-39.

93. Wolffe, A.P., Chromatin, Structure and Function. 2nd ed. 1995, Toronto, Ont: Academic Press.

94. Berger, S.L. and G. Felsenfeld, Chromatin goes global. *Mol Cell*, 2001. **8**(2): p. 263-8.

95. Rea, S., et al., Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 2000. **406**(6796): p. 593-9.

96. Woodcock, C.L. and S. Dimitrov, Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev*, 2001. **11**(2): p. 130-5.

97. Ringrose, L. and R. Paro, Gene regulation. Cycling silence. *Nature*, 2001. **412**(6846): p. 493-4.

98. Sharp, J.A., et al., Yeast histone deposition protein Asf1p requires Hir proteins and PCNA for heterochromatic silencing. *Curr Biol*, 2001. **11**(7): p. 463-73.

99. Lupo, R., et al., Drosophila chromosome condensation proteins Topoisomerase II and Barren colocalize with Polycomb and maintain Fab-7 PRE silencing. *Mol Cell*, 2001. **7**(1): p. 127-36.

100. Lyko, F. and R. Paro, Chromosomal elements conferring epigenetic inheritance. *Bioessays*, 1999. **21**(10): p. 824-32.

101. Henikoff, S., K. Ahmad, and H.S. Malik, The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 2001. **293**(5532): p. 1098-102.

102. Hennig, W., Heterochromatin. *Chromosoma*, 1999. **108**(1): p. 1-9.

103. Lloyd, V.K., D.A. Sinclair, and T.A. Grigliatti, Genomic imprinting and position-effect variegation in Drosophila melanogaster. *Genetics*, 1999. **151**(4): p. 1503-16.

104. Ahmad, K. and S. Henikoff, Centromeres are specialized replication domains in heterochromatin. *J Cell Biol*, 2001. **153**(1): p. 101-10.

105. Marahrens, Y., X-inactivation by chromosomal pairing events. *Genes Dev*, 1999. **13**(20): p. 2624-32.

106. Weiler, K.S. and B.T. Wakimoto, Heterochromatin and gene expression in Drosophila. *Annu Rev Genet*, 1995. **29**: p. 577-605.

107. Eissenberg, J.C., et al., Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proc Natl Acad Sci U S A*, 1990. **87**(24): p. 9923-7.

108. Elgin, S.C., Heterochromatin and gene regulation in Drosophila. *Curr Opin Genet Dev*, 1996. **6**(2): p. 193-202.

109. Rice, J.C. and C.D. Allis, Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr Opin Cell Biol*, 2001. **13**(3): p. 263-73.

110. Kelly, T.J., et al., Type B histone acetyltransferase Hat1p participates in telomeric silencing. *Mol Cell Biol*, 2000. **20**(19): p. 7051-8.

111. Guy, L.G., R. Kothary, and L. Wall, Position effects in mice carrying a lacZ transgene in cis with the beta- globin LCR can be explained by a graded model. *Nucleic Acids Res*, 1997. **25**(21): p. 4400-7.

112. Garrick, D., et al., Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Res*, 1996. **24**(24): p. 4902-9.

113. Vyas, P., et al., Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell*, 1992. **69**(5): p. 781-93.

114. Dillon, N. and F. Grosveld, Chromatin domains as potential units of eukaryotic gene function. *Curr Opin Genet Dev*, 1994. **4**(2): p. 260-4.

115. Csink, A.K. and S. Henikoff, Genetic modification of heterochromatic association and nuclear organization in Drosophila. *Nature*, 1996. **381**(6582): p. 529-31.

116. Kleinjan, D.J. and V. van Heyningen, Position effect in human genetic disease. *Hum Mol Genet*, 1998. **7**(10): p. 1611-8.

117. Cavalli, G. and R. Paro, Chromo-domain proteins: linking chromatin structure to epigenetic regulation. *Curr Opin Cell Biol*, 1998. **10**(3): p. 354-60.

118. Talbert, P.B. and S. Henikoff, A reexamination of spreading of position-effect variegation in the white-roughest region of Drosophila melanogaster. *Genetics*, 2000. **154**(1): p. 259-72.

119. Pirrotta, V. and L. Rastelli, white Gene Expression, Repressive Chromatin Domains and Homeotic Gene Regulation in Drosophila. *Bioessays*, 1994. **16**(8): p. 549-56.

120. Farkas, G., B.A. Leibovitch, and S.C. Elgin, Chromatin organization and transcriptional control of gene expression in Drosophila. *Gene*, 2000. **253**(2): p. 117-36.

121. Ahmad, K. and S. Henikoff, Modulation of a transcription factor counteracts heterochromatic gene silencing in Drosophila. *Cell*, 2001. **104**(6): p. 839-47.

122. Pirrotta, V., Transvection and chromosomal trans-interaction effects. *Biochim Biophys Acta*, 1999. **1424**(1): p. M1-8.

123. Clark, D.V., J.F. Sabl, and S. Henikoff, Repetitive arrays containing a housekeeping gene have altered polytene chromosome morphology in Drosophila. *Chromosoma*, 1998. **107**(2): p. 96-104.

124. Henikoff, S., et al., Heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci U S A*, 2000. **97**(2): p. 716-21.

125. Ofir, R., et al., Position effect of human telomeric repeats on replication timing. *Proc Natl Acad Sci U S A*, 1999. **96**(20): p. 11434-9.

126. Flint, J., et al., The relationship between chromosome structure and function at a human telomeric region. *Nat Genet*, 1997. **15**(3): p. 252-7.

127. Shore, D., Telomeric chromatin: replicating and wrapping up chromosome ends. *Curr Opin Genet Dev*, 2001. **11**(2): p. 189-98.

128. Dubrana, K., S. Perrod, and S.M. Gasser, Turning telomeres off and on. *Curr Opin Cell Biol*, 2001. **13**(3): p. 281-9.

129. Bryan, T.M. and T.R. Cech, Telomerase and the maintenance of chromosome ends. *Curr Opin Cell Biol*, 1999. **11**(3): p. 318-24.

130. Blackburn, E.H., Switching and signaling at the telomere. *Cell*, 2001. **106**(6): p. 661-73.

131. Flint, J., et al., Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Hum Mol Genet*, 1997. **6**(8): p. 1305-13.

132. Pirrotta, V., PcG complexes and chromatin silencing. *Curr Opin Genet Dev*, 1997. **7**(2): p. 249-58.

133. Baur, J.A., et al., Telomere position effect in human cells. *Science*, 2001. **292**(5524): p. 2075-7.

134. Park, Y. and A.J. Lustig, Telomere structure regulates the heritability of repressed subtelomeric chromatin in Saccharomyces cerevisiae. *Genetics*, 2000. **154**(2): p. 587-98.

135. Imai, S., et al., Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature*, 2000. **403**(6771): p. 795-800.

136. Flavell, R.B., Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc Natl Acad Sci U S A*, 1994. **91**(9): p. 3490-6.

137. de Kok, Y.J., et al., Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum Mol Genet*, 1996. **5**(9): p. 1229-35.

138. Mahadevaiah, S.K., et al., Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. *Hum Mol Genet*, 1998. **7**(4): p. 715-27.

139. Laval, S.H., et al., Y chromosome short arm-Sxr recombination in XSxr/Y males causes deletion of Rbm and XY female sex reversal. *Proc Natl Acad Sci U S A*, 1995. **92**(22): p. 10403-7.

140. McElreavey, K., et al., Loss of sequences 3' to the testis-determining gene, SRY, including the Y pseudoautosomal boundary associated with partial testicular determination. *Proc Natl Acad Sci U S A*, 1996. **93**(16): p. 8590-4.

141. McElreavey, K., et al., XY sex reversal associated with a deletion 5' to the SRY "HMG box" in the testis-determining region. *Proc Natl Acad Sci U S A*, 1992. **89**(22): p. 11016-20.

142. Capel, B., et al., Deletion of Y chromosome sequences located outside the testis determining region can cause XY female sex reversal. *Nat Genet*, 1993. **5**: p. 301-7.

143. Ferguson, B.M. and W.L. Fangman, A position effect on the time of replication origin activation in yeast. *Cell*, 1992. **68**(2): p. 333-9.

144. Giraudeau, F., et al., Use of a set of highly polymorphic minisatellite probes for the identification of cryptic 1p36.3 deletions in a large collection of patients with idiopathic mental retardation. *J Med Genet*, 2001. **38**(2): p. 121-5.

145. Wu, Y.Q., et al., Molecular refinement of the 1p36 deletion syndrome reveals size diversity and a preponderance of maternally derived deletions. *Hum Mol Genet*, 1999. **8**(2): p. 313-21.

146. Bi, X. and J.R. Broach, Chromosomal boundaries in S. cerevisiae. *Curr Opin Genet Dev*, 2001. **11**(2): p. 199-204.

147. Tupler, R., et al., Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J Med Genet*, 1996. **33**(5): p. 366-70.

148. Peters, A.H., et al., Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat Genet*, 2002. **30**(1): p. 77-80.

149. Boumil, R.M. and J.T. Lee, Forty years of decoding the silence in X-chromosome inactivation. *Hum Mol Genet*, 2001. **10**(20): p. 2225-32.

150. Muller, C. and A. Leutz, Chromatin remodeling in development and differentiation. *Curr Opin Genet Dev*, 2001. **11**(2): p. 167-74.

151. Orlando, V. and R. Paro, Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev*, 1995. **5**(2): p. 174-9.

152. Satijn, D.P. and A.P. Otte, Polycomb group protein complexes: do different complexes regulate distinct target genes? *Biochim Biophys Acta*, 1999. **1447**(1): p. 1-16.

153. Wallrath, L.L., Unfolding the mysteries of heterochromatin. *Curr Opin Genet Dev*, 1998. **8**(2): p. 147-53.

154. van der Vlag, J. and A.P. Otte, Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet*, 1999. **23**(4): p. 474-8.

155. Mollaaghababa, R., et al., Mutations in Drosophila heat shock cognate 4 are enhancers of Polycomb. *Proc Natl Acad Sci U S A*, 2001. **98**(7): p. 3958-63.

156. Pirrotta, V., Polycombing the genome: PcG, trxG, and chromatin silencing. *Cell*, 1998. **93**(3): p. 333-6.

157. Ringrose, L. and R. Paro, Remembering silence. *Bioessays*, 2001. **23**(7): p. 566-70.

158. Brock, H.W. and M. van Lohuizen, The Polycomb group--no longer an exclusive club? *Curr Opin Genet Dev*, 2001. **11**(2): p. 175-81.

159. Saurin, A.J., et al., The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol*, 1998. **142**(4): p. 887-98.

160. Hanson, R.D., et al., Mammalian Trithorax and polycomb-group homologues are antagonistic regulators of homeotic development. *Proc Natl Acad Sci U S A*, 1999. **96**(25): p. 14372-7.

161. Pal-Bhadra, M., U. Bhadra, and J.A. Birchler, Cosuppression of nonhomologous transgenes in Drosophila involves mutually related endogenous sequences. *Cell*, 1999. **99**(1): p. 35-46.

162. Wolffe, A.P., When more is less. *Nat Genet*, 1998. **18**(1): p. 5-6.

163. Martin, D.I. and E. Whitelaw, The vagaries of variegating transgenes. *Bioessays*, 1996. **18**(11): p. 919-23.

164. Dorer, D.R., Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res*, 1997. **6**(1): p. 3-10.

165. Garrick, D., et al., Repeat-induced gene silencing in mammals. *Nat Genet*, 1998. **18**(1): p. 56-9.

166. Dorer, D.R. and S. Henikoff, Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell*, 1994. **77**(7): p. 993-1002.

167. Fanti, L., et al., Heterochromatin protein 1 binds transgene arrays. *Chromosoma*, 1998. **107**(5): p. 286-92.

168. Smothers, J.F. and S. Henikoff, The hinge and chromo shadow domain impart distinct targeting of HP1- like proteins. *Mol Cell Biol*, 2001. **21**(7): p. 2555-69.

169. van der Vlag, J., et al., Transcriptional repression mediated by polycomb group proteins and other chromatin-associated repressors is selectively blocked by insulators. *J Biol Chem*, 2000. **275**(1): p. 697-704.

170. Horard, B., et al., Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Mol Cell Biol*, 2000. **20**(9): p. 3187-97.

171. Bel, S., et al., Genetic interactions and dosage effects of Polycomb group genes in mice. *Development*, 1998. **125**(18): p. 3543-51.

172. O'Carroll, D., et al., The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol*, 2001. **21**(13): p. 4330-6.

173. Mihaly, J., et al., In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. *Development*, 1997. **124**(9): p. 1809-20.

174. Fritsch, C., et al., The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. *Development*, 1999. **126**(17): p. 3905-13.

175. Gould, A., Functions of mammalian Polycomb group and trithorax group related genes. *Curr Opin Genet Dev*, 1997. **7**(4): p. 488-94.

176. Akasaka, T., et al., Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development*, 2001. **128**(9): p. 1587-97.

177. Poux, S., D. McCabe, and V. Pirrotta, Recruitment of components of Polycomb Group chromatin complexes in Drosophila. *Development*, 2001. **128**(1): p. 75-85.

178. Akasaka, T., et al., A role for mel-18, a Polycomb group-related vertebrate gene, during theanteroposterior specification of the axial skeleton. *Development*, 1996. **122**(5): p. 1513-22.

179. Takihara, Y., et al., Targeted disruption of the mouse homologue of the Drosophila polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development*, 1997. **124**(19): p. 3673-82.

180. del Mar Lorente, M., et al., Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development*, 2000. **127**(23): p. 5093-100.

181. Jacobs, J.J., et al., The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature*, 1999. **397**(6715): p. 164-8.

182. Strouboulis, J., et al., Transcriptional repression by XPc1, a new Polycomb homolog in Xenopus laevis embryos, is independent of histone deacetylase. *Mol Cell Biol*, 1999. **19**(6): p. 3958-68.

183. Kyba, M. and H.W. Brock, The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. *Mol Cell Biol*, 1998. **18**(5): p. 2712-20.

184. Sewalt, R.G., et al., Characterization of interactions between the mammalian polycomb-group proteins Enx1/EZH2 and EED suggests the existence of different mammalian polycomb-group protein complexes. *Mol Cell Biol*, 1998. **18**(6): p. 3586-95.

185. Shao, Z., et al., Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*, 1999. **98**(1): p. 37-46.

186. Garcia, E., et al., RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *Embo J*, 1999. **18**(12): p. 3404-18.

187. Pirrotta, V., Chromatin complexes regulating gene expression in Drosophila. *Curr Opin Genet Dev*, 1995. **5**(4): p. 466-72.

188. Brown, J.L., et al., The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell*, 1998. **1**(7): p. 1057-64.

189. Strutt, H., G. Cavalli, and R. Paro, Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *Embo J*, 1997. **16**(12): p. 3621-32.

190. Birchler, J.A., M.P. Bhadra, and U. Bhadra, Making noise about silence: repression of repeated genes in animals. *Curr Opin Genet Dev*, 2000. **10**(2): p. 211-6.

191. Hodgson, J.W., B. Argiropoulos, and H.W. Brock, Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol Cell Biol*, 2001. **21**(14): p. 4528-43.

192. Katsani, K.R., M.A. Hajibagheri, and C.P. Verrijzer, Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *Embo J*, 1999. **18**(3): p. 698-708.

193. Henikoff, S. and D. Vermaak, Bugs on drugs go GAGAA. *Cell*, 2000. **103**(5): p. 695-8.

194. Jimenez-Garcia, E., et al., The GAGA factor of Drosophila binds triple-stranded DNA. *J Biol Chem*, 1998. **273**(38): p. 24640-8.

195. O'Brien, T., et al., Distribution of GAGA protein on Drosophila genes in vivo. *Genes Dev*, 1995. **9**(9): p. 1098-110.

196. Breiling, A., et al., The Drosophila polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. *Mol Cell Biol*, 1999. **19**(12): p. 8451-60.

197. Schlossherr, J., et al., Gene inactivation in Drosophila mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Mol Gen Genet*, 1994. **243**(4): p. 453-62.

198. Hansen, R.S., et al., Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. *Hum Mol Genet*, 2000. **9**(18): p. 2575-87.

199. Wijmenga, C., et al., Localization of the ICF syndrome to chromosome 20 by homozygosity mapping. *Am J Hum Genet*, 1998. **63**(3): p. 803-9.

200. Xu, G.L., et al., Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, 1999. **402**(6758): p. 187-91.

201. Hassan, K.M., et al., Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication. *Hum Genet*, 2001. **109**(4): p. 452-62.

202. Hendrich, B. and W. Bickmore, Human diseases with underlying defects in chromatin structure and modification. *Hum Mol Genet*, 2001. **10**(20): p. 2233-42.

203. Wright, T.J., et al., Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum Mol Genet*, 1993. **2**(10): p. 1673-8.

204. Sambrook, J., E.F. Fritsch, and T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd edt. 1989, Cold Spring Harbor, NY: Cold Spring Harbor Laboratoru Press.

205. Inoue, H., H. Nojima, and H. Okayama, High efficiency transformation of Escherichia coli with plasmids. *Gene*, 1990. **96**(1): p. 23-28.

206. Liu, C., et al., MyoD-dependent induction during myoblast differentiation of p204, a protein also inducible by interferon. *Mol Cell Biol*, 2000. **20**(18): p. 7024-36.

207. Anton, M. and F.L. Graham, Site-specific recombination mediated by an adenovirus vector expressing the Cre recombinase protein: a molecular switch for control of gene expression. *J Virol*, 1995. **69**(8): p. 4600-6.

208. Southern, E.M., Detection od Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis. *J. Mol. Biol.*, 1975. **98**: p. 503-517.

209. Langen, R.C., et al., Inflammatory cytokines inhibit myogenic differentiation through activation of nuclear factor-kappaB. *Faseb J*, 2001. **15**(7): p. 1169-80.

210. Gorza, L. and M. Vitadello, Reduced amount of the glucose-regulated protein GRP94 in skeletal myoblasts results in loss of fusion competence. *FASEB J*, 2000. **14**(3): p. 461-75.

211. Chomczynski, P. and N. Sacchi, Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Anal Biochem*, 1987. **162**: p. 156-59.

212. Altschul, S.F., et al., Basic Local Alignment Search Tool. *J. Mol. Biol.*, 1990. **215**: p. 403-410.

213. Wang, J. and K. Walsh, Resistance to apoptosis conferred by Cdk inhibitors during myocyte differentiation. *Science*, 1996. **273**(5273): p. 359-61.

214. Rios, R., et al., Myostatin regulates cell survival during C2C12 myogenesis. *Biochem Biophys Res Commun*, 2001. **280**(2): p. 561-6.

215. Fujio, Y., et al., Activation of Akt2 Inhibits anoikis and apoptosis induced by myogenic differentiation. *Cell Death Differ*, 2001. **8**(12): p. 1207-12.

216. Li, Y.P., et al., Skeletal muscle myocytes undergo protein loss and reactive oxygen-mediated NF-kappaB activation in response to tumor necrosis factor alpha. *Faseb J*, 1998. **12**(10): p. 871-80.

217. Day, C.D., et al., Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. *Genes Dev*, 2000. **14**(22): p. 2869-80.

218. Robertson, G., et al., Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res*, 1996. **24**(8): p. 1465-71.

219. Mei, Q., R. Kothary, and L. Wall, The tkNeo gene, but not the pgkPuro gene, can influence the ability of the beta-globin LCR to enhance and confer position-independent expression onto the beta-globin gene. *Exp Cell Res*, 2000. **260**(2): p. 304-12.

220. Quantin, B., et al., Adenovirus as an expression vector in muscle cells in vivo. *Proc Natl Acad Sci U S A*, 1992. **89**(7): p. 2581-4.

221. Fujio, Y., et al., Cell cycle withdrawal promotes myogenic induction of Akt, a positive modulator of myocyte survival. *Mol Cell Biol*, 1999. **19**(7): p. 5073-82.

222. Kooter, J.M., M.A. Matzke, and P. Meyer, Listening to the silent genes: transgene silencing, gene regulation and pathogen control. *Trends Plant Sci*, 1999. **4**(9): p. 340-347.

223. Selker, E.U., Gene silencing: repeats that count. *Cell*, 1999. **97**(2): p. 157-60.

224. Faugeron, G., Diversity of homology-dependent gene silencing strategies in fungi. *Curr Opin Microbiol*, 2000. **3**(2): p. 144-8.

225. Henikoff, S. and L. Comai, Trans-sensing effects: the ups and downs of being together. *Cell*, 1998. **93**(3): p. 329-32.

226. Surani, M.A., Reprogramming of genome function through epigenetic inheritance. *Nature*, 2001. **414**(6859): p. 122-8.

227. Marks, P.A., V.M. Richon, and R.A. Rifkind, Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells. *J Natl Cancer Inst*, 2000. **92**(15): p. 1210-6.

228. Poux, S., R. Melfi, and V. Pirrotta, Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev*, 2001. **15**(19): p. 2509-14.

229. Roseman, R.R., et al., Long-range repression by multiple polycomb group (PcG) proteins targeted by fusion to a defined DNA-binding domain in Drosophila. *Genetics*, 2001. **158**(1): p. 291-307.

230. Kuo, M.H. and C.D. Allis, In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, 1999. **19**(3): p. 425-33.

231. Cheng, T.C., et al., Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science*, 1993. **261**(5118): p. 215-8.

232. Edmondson, D.G., et al., Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. *Mol Cell Biol*, 1992. **12**(9): p. 3665-77.

233. Black, B.L., J.F. Martin, and E.N. Olson, The mouse MRF4 promoter is trans-activated directly and indirectly by muscle-specific transcription factors. *J Biol Chem*, 1995. **270**(7): p. 2889-92.

234. Johanson, M., et al., Transcriptional activation of the myogenin gene by MEF2-mediated recruitment of myf5 is inhibited by adenovirus E1A protein. *Biochem Biophys Res Commun*, 1999. **265**(1): p. 222-32.

235. Picketts, D.J., C.R. Mueller, and D. Lillicrap, Transcriptional control of the factor IX gene: analysis of five cis- acting elements and the deleterious effects of naturally occurring hemophilia B Leyden mutations. *Blood*, 1994. **84**(9): p. 2992-3000.

236. Gorski, K., M. Carneiro, and U. Schibler, Tissue-specific in vitro transcription from the mouse albumin promoter. *Cell*, 1986. **47**(5): p. 767-76.

Appendix A - Cloning strategy for experimental constructs.

This appendix is meant to serve as a visual aid, depicting the steps taken during the complicated cloning strategy presented in Materials and Methods section 2.2.1. (Representative figures are not drawn to scale)

Figure Legend:

→ D4Z4 Repeats

→ MCS from pBluescript$_{A-C}$

→ MCS from pBS246

→ Neo^{R} from pBluescript$_{A-C}$

→ LoxP site

→ LacZ gene from pRP2044

→ BOS Promoter from

E → *EcoR*I restriction site

EV → *EcoR*V restriction site

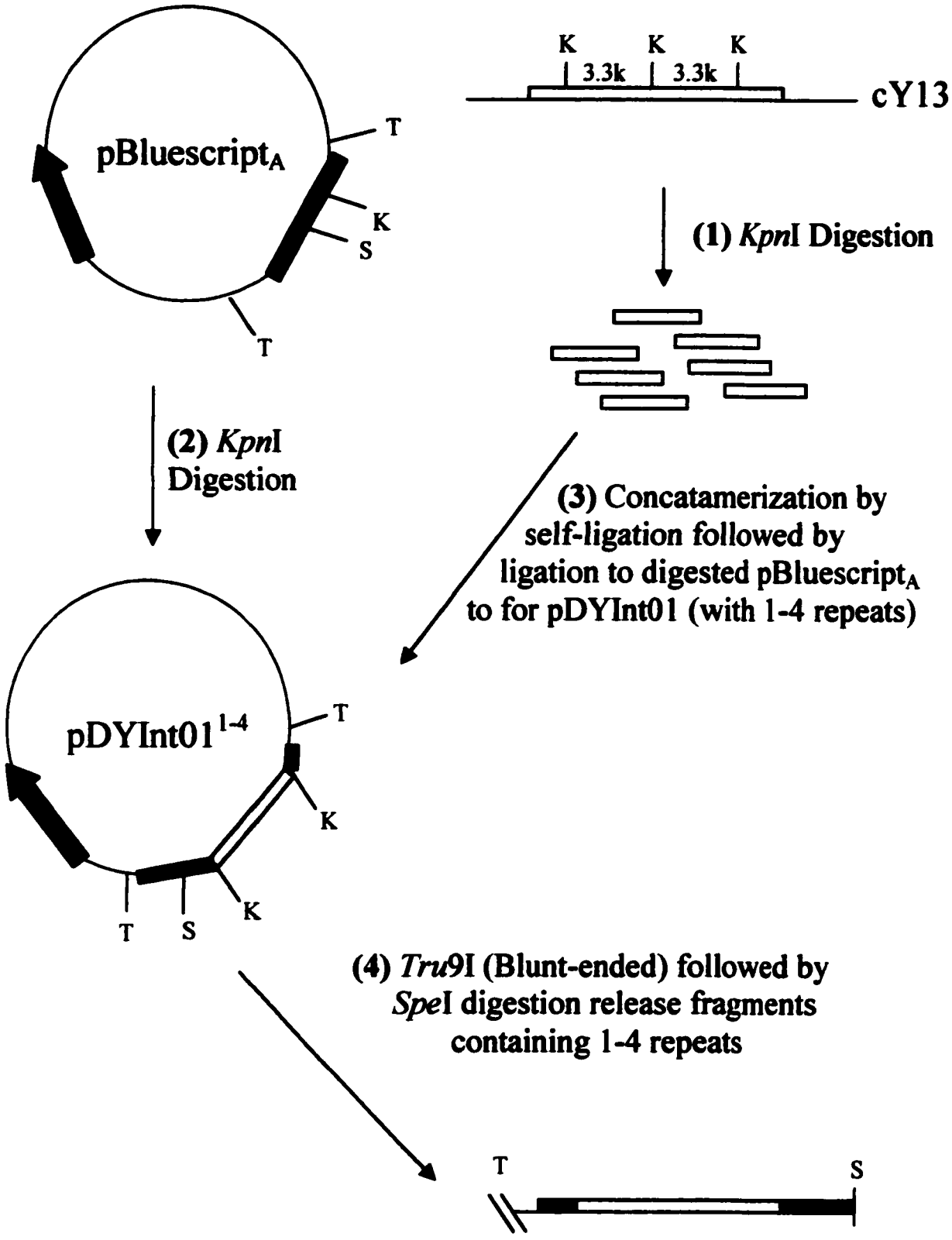
H → *Hind*III restriction site

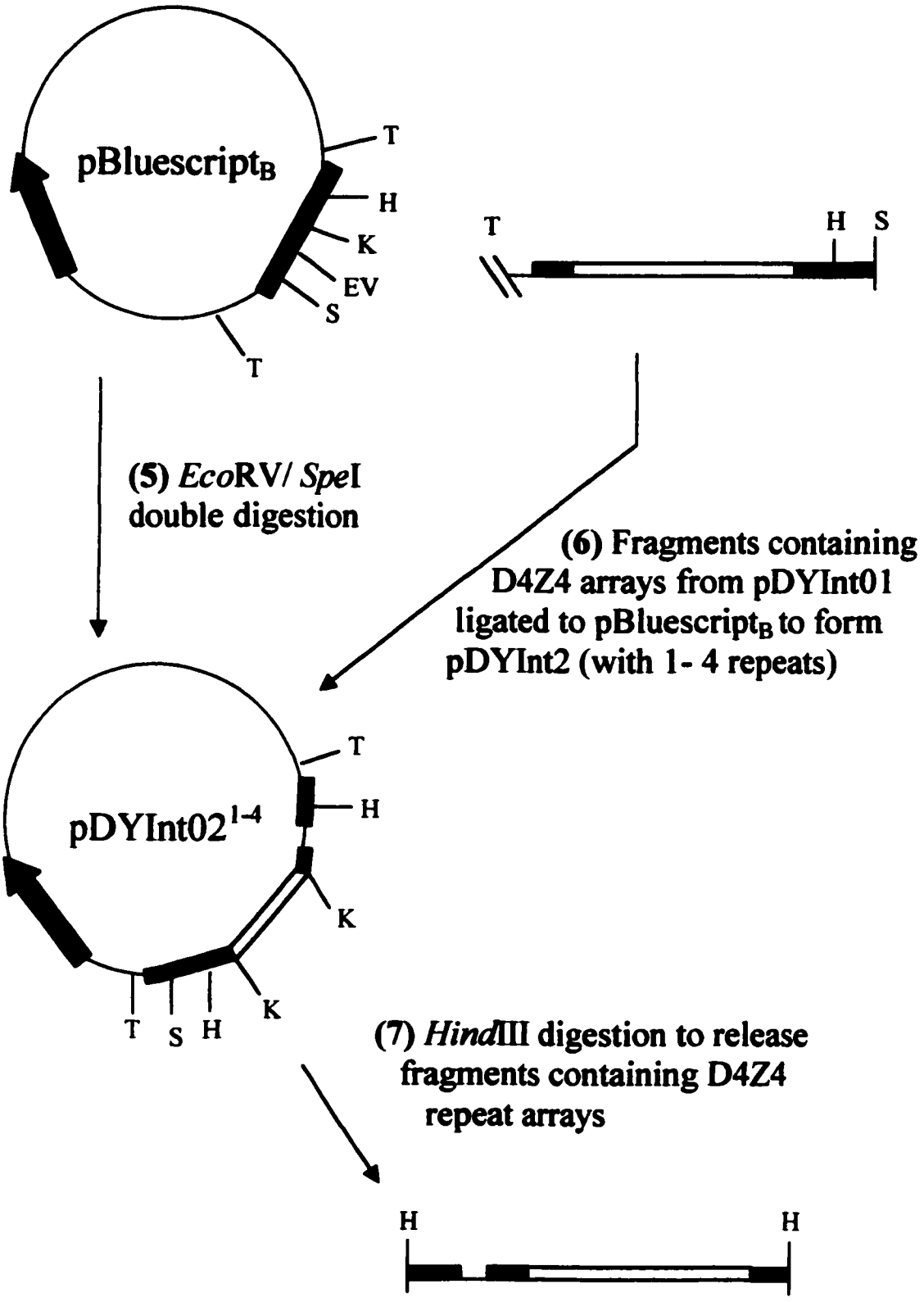
K → *Kpn*I restriction site

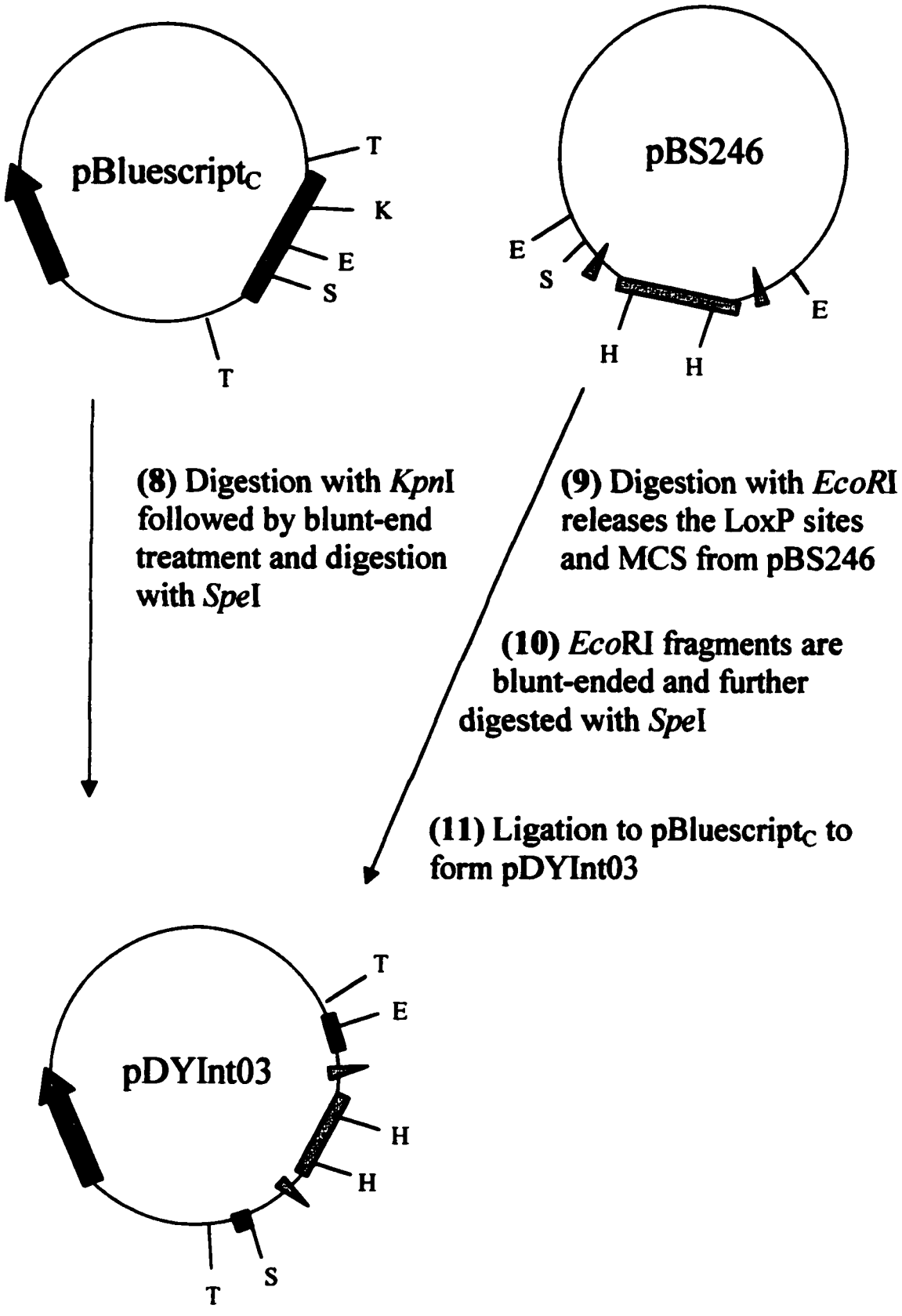
N → *Not*I restriction site

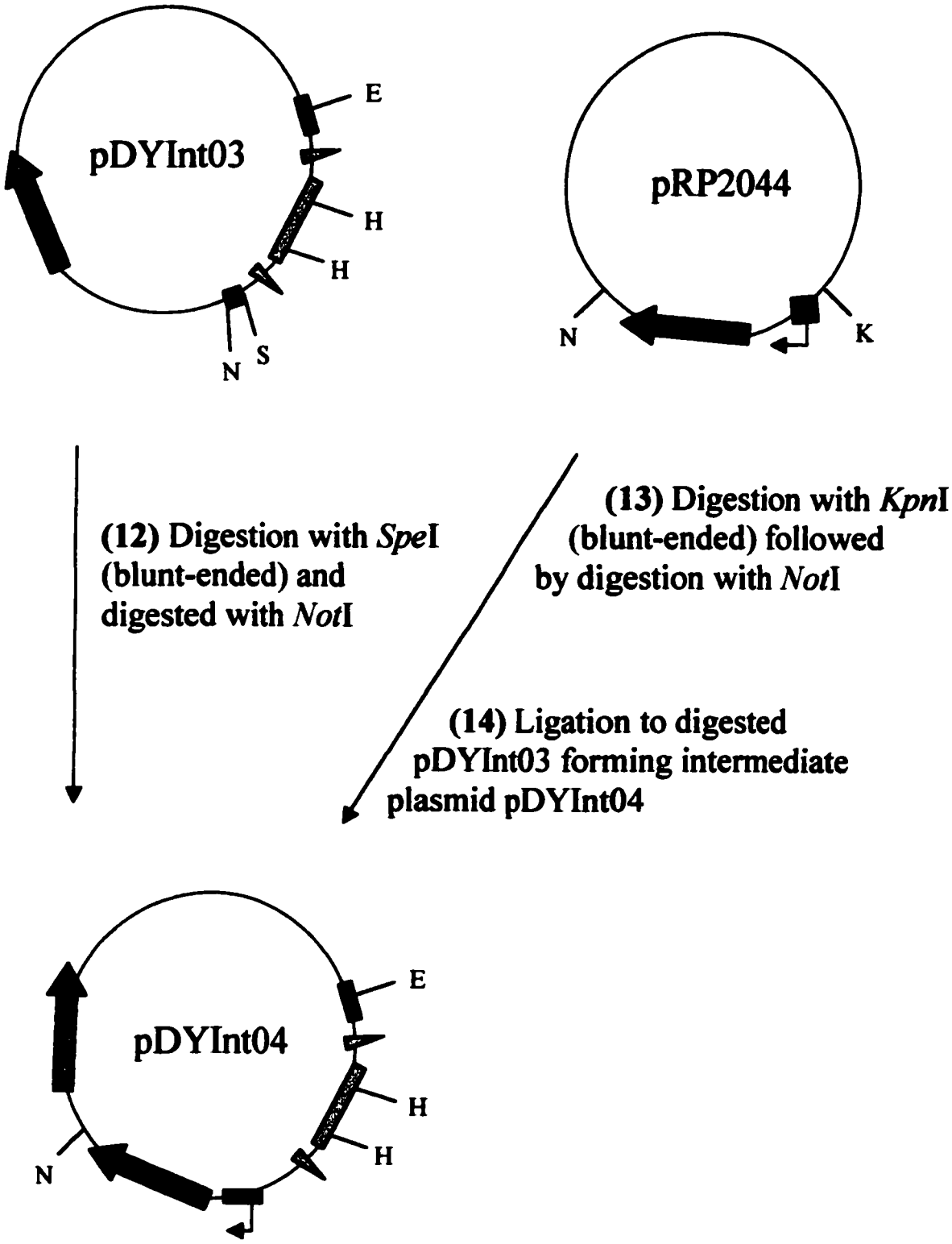
S → *Spe*I restriction site

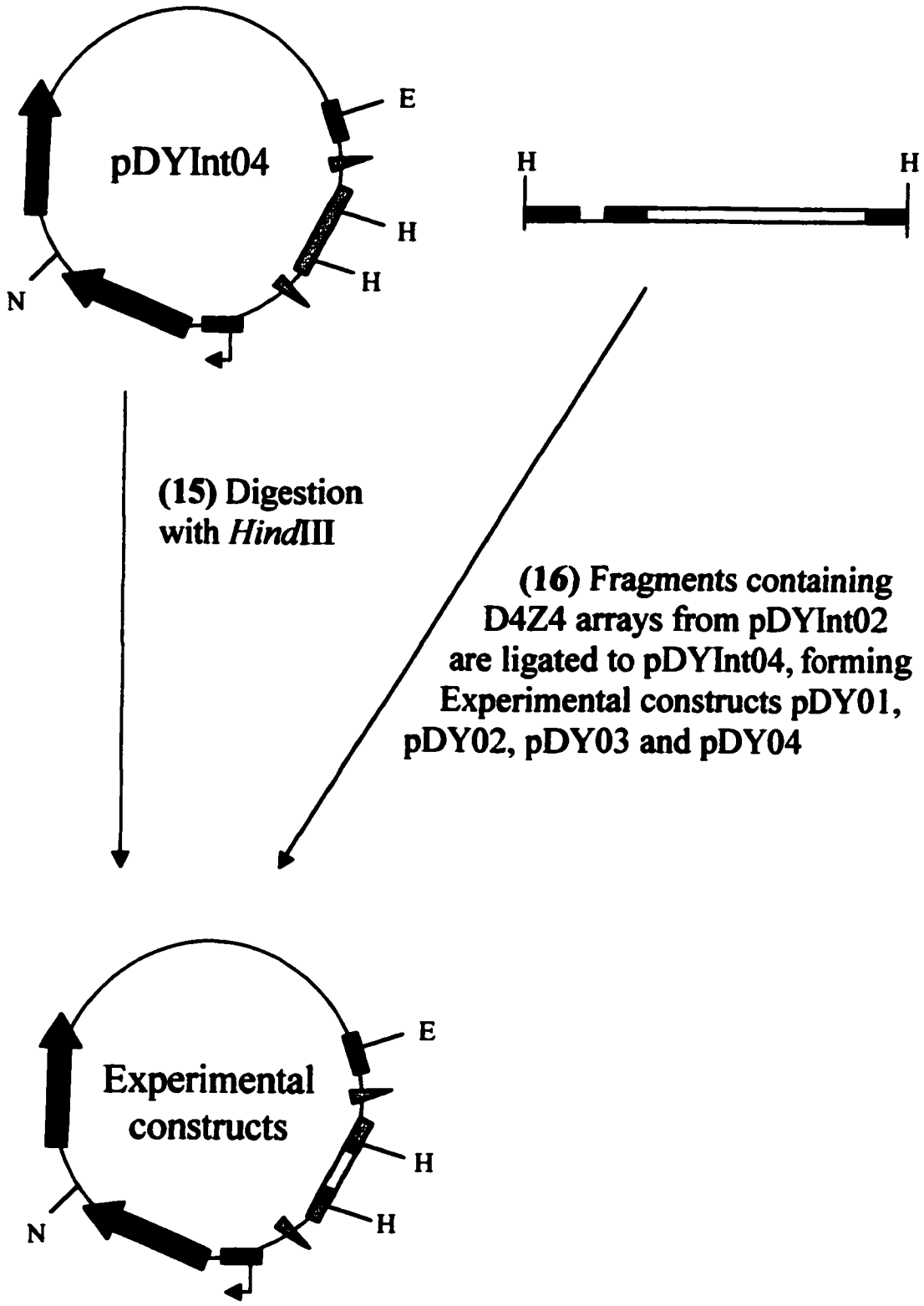
T → *Tru*9I restriction site


K
K
K
3.3k
3.3k
cY13
pBluescript$_A$
T
K
S
T
(1) *Kpn*I Digestion
(2) *Kpn*I Digestion
(3) Concatamerization by self-ligation followed by ligation to digested pBluescript$_A$ to for pDYInt01 (with 1-4 repeats)
pDYInt01$^{1-4}$
T
K
T
S
K
(4) *Tru*9I (Blunt-ended) followed by *Spe*I digestion release fragments containing 1-4 repeats
T
S


pBluescript$_B$
T
H
K
EV
S
T
T
H S
(5) *Eco*RV/ *Spe*I
double digestion
(6) Fragments containing
D4Z4 arrays from pDYInt01
ligated to pBluescript$_B$ to form
pDYInt2 (with 1- 4 repeats)
pDYInt02$^{1-4}$
T
H
K
T S H K
(7) *Hind*III digestion to release
fragments containing D4Z4
repeat arrays
H
H


pBluescriptC
T
K
E
S
T
pBS246
E
S
H
H
E
(8) Digestion with KpnI followed by blunt-end treatment and digestion with SpeI
(9) Digestion with EcoRI releases the LoxP sites and MCS from pBS246
(10) EcoRI fragments are blunt-ended and further digested with SpeI
(11) Ligation to pBluescriptC to form pDYInt03
pDYInt03
T
E
H
H
T
S


pDYInt03
E
H
H
N
S
pRP2044
N
K
(12) Digestion with *Spe*I (blunt-ended) and digested with *Not*I
(13) Digestion with *Kpn*I (blunt-ended) followed by digestion with *Not*I
(14) Ligation to digested pDYInt03 forming intermediate plasmid pDYInt04
pDYInt04
E
H
H
N


pDYInt04
E
H
H
N
H
H
(15) Digestion with *Hind*III
(16) Fragments containing D4Z4 arrays from pDYInt02 are ligated to pDYInt04, forming Experimental constructs pDY01, pDY02, pDY03 and pDY04
Experimental constructs
E
H
H
N

Appendix B: A representational selection for raw data aquired from ONPG analyisis for Table 2.

Sample	ONPG Value*	Pos / Neg
C2C12 (10 Ug)	0.0001	NEG
C2C12 (20 Ug)	0.0023	NEG
Bgal protein (5 Ug)	0.1489	POS
Bgal protein (10 Ug)	0.2598	POS

pRP2044 Clones	ONPG Value*	Pos / Neg
D3	0.0014	NEG
D4	0.0007	NEG
D8	0.0018	NEG
D9	0.0216	POS
D11	0.0346	POS
D24	0.0003	NEG
D34	0.0421	POS
D39	0	NEG
D43	0.0287	POS

pDY04 Clones	ONPG Value*	Pos / Neg
E6	0.0005	NEG
E7	0	NEG
E13	0.0392	POS
E27	0.0025	NEG
E32	0.0036	NEG
E39	0.004	NEG
E35	0.021	POS
E41	0.0337	POS
E49	0.0046	NEG

*ONPG values correspond to Absorbance as calculated by the optical density reading at 420nm wavelenght using the Beckman Coulter DU 640 spectrophotometer (Fullerton CA).