BAC TRANSGENE ARRAYS AS A MODEL SYSTEM FOR STUDYING LARGE-SCALE CHROMATIN STRUCTURE

BY

QIAN BIAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Andrew S. Belmont, Chair
Professor Taekjip Ha
Professor Paul R. Selvin
Assistant Professor Fei Wang

ABSTRACT

The folding of interphase chromatin into large-scale chromatin structure and its spatial organization within nucleus has been suggested to have important roles in gene regulation. In this study, we created engineered chromatin regions consisting of tandem repeats of BAC transgenes, which contain 150-200 kb of defined genomic regions, and used them as a model system to study the mechanisms and functional significance of large-scale chromatin organization.

The BAC transgene arrays recapitulated several important features of endogenous chromatin, including transcription level and intranuclear positioning. Using this system, we showed that tandem arrays of housekeeping gene loci form open large-scale chromatin structure independent of their genomic integration sites, including insertions within centromeric heterochromatin. This BAC-specific large-scale chromatin conformation provided a permissive environment for transcription, as evidenced by the copy-number dependent and position independent expression of embedded reporter mini-genes. This leads to the development of a novel method for reliable transgene expression in mammalian cells, which should prove useful in a number of therapeutic and scientific applications.

We also demonstrated that BAC transgene arrays can be employed as an effective system for dissecting sequence determinants for intranuclear positioning of gene loci. We showed that in mouse ES and fibroblast cells a BAC carrying a 200 kb human genomic fragment containing the beta-globin locus autonomously targets to the nuclear periphery. Using BAC recombineering, we dissected this 200kb region and identified two genomic regions sufficient to target the BAC transgenes to nuclear periphery. This study represents a first step towards elucidation of the molecular mechanism for the nuclear peripheral localization of genes in mammalian cells.

DEDICATION

献给我的父亲， 卞文峰和我的母亲， 赵秀君，
感谢他们鼓励我走上科研的道路, 和他们无私的爱。

To my dear wife, Yiren Xu (徐意人)，
for her love and support along the way, and for making me a better man.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor, Dr. Andrew S. Belmont, for opening the door of cell biology for me, for always being supportive, and for guiding me through the excitements and frustrations of science. I also thank my committee members, Dr. Paul R. Selvin, Dr. Taekjip Ha and Dr. Fei Wang for their help and precious advice.

I am grateful to all members Belmont lab for their help and friendship during my graduate career, and for offering me opportunities to learn from them. I especially want to thank Matt Plutz for sharing his expertise in molecular cloning with me, and Dr. Paul Sinclair for teaching the techniques related to embryonic stem cells. I was also very lucky to work together with an excellent undergraduate student, Jurgis Alvikas, who helped me tremendously on BAC recombineering. I would like to thank Dr. Edith Heard (Curie Institute) and Dr. Neal Copeland (National Cancer Institute) for providing valuable reagents.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

For cells to function and propagate, the information stored in DNA molecules needs to be precisely retrieved and duplicated. Understanding the mechanisms for regulation of transcription and replication is therefore one of the main themes of current biological research. Advances in sequencing technology in last twenty years have led to the rapid expansion of genomic databases and improved the understanding of the function of genome. However, to fully elucidate the mechanisms for genome regulation, one must realize that in eukaryotic cells the genome exists in the form of chromatin, a highly compacted and organized structural entity, rather than in the form of one dimensional DNA molecules. The organization and dynamics of chromatin structure plays a critical role in gene regulation.

The compaction of over 1 meter of DNA into chromatin fibers which fit within the limited space of nucleus is achieved via a hierarchy of folding events (Woodcock, 2006). The basic building block of chromatin is the nucleosome, consisting of 146 bp of DNA wrapping around a core histone octamer (Luger et al., 1997). *In vitro* studies have shown that purified or reconstituted nucleosomes arrays can fold into 30nm chromatin fiber and higher order chromatin fiber upon the addition of cation (Hansen, 2002; Simpson et al., 1985). While the 30nm fiber has been well accepted as an intermediate level of chromatin structure, the chromatin folding above 30nm fiber in interphase nuclei, defined as large-scale chromatin structure, has been poorly characterized and understood (Maeshima et al., 2010; Woodcock and Ghosh, 2010). The relationship between large-scale chromatin structure and the regulation of nuclear processes involving DNA, although implied by many studies, is still not well established (Belmont et al., 1999).

On the other hand, the highly folded chromatin is arranged in the three dimensional space of the nucleus in a non-random fashion. The nuclear radial positioning of chromatin domains and genes and their location relative to specific nuclear subcompartments or other genomic loci have been shown to have comprehensive implications on gene regulation (Ferrai et al., 2010; Misteli, 2007).

Research in the Belmont laboratory focuses on understanding the organization and functional significances of higher-order chromatin structure. Since the mid 1990s, the laboratory has been working on the development of new strategies for visualizing chromatin structure *in vivo* and studying the ultrastructure and dynamics of chromatin. The studies in this thesis are based on a new generation of experimental system which recapitulates the chromatin structure of endogenous loci at a higher degree by using tandem repeats of bacterial artificial chromosome (BAC) transgene arrays (Bian and Belmont, 2010; Hu et al., 2009a; Sinclair et al., 2010). Using this method, I investigated the functional implication of "open" large-scale chromatin structure at the microscopic level and developed a new strategy for expressing transgenes at high level. I also performed a study to identify the sequence determinants for the localization of a genomic region at nuclear periphery. This study is a first step towards better understanding the molecular mechanism for the nuclear peripheral localization of genes in mammalian cells.

Hierarchical folding of chromatin

A nucleosome core particle is formed by wrapping 146 base pairs of DNA double helix around an octameric histone core (two copies of histone H2A, H2B, H3 and H4) in 1.65 left handed superhelical turn (Luger et al., 1997). The tails of the core histone protrude from the histone octamer and are subject to covalent modifications.

Such posttranslational histone modification can alter the interactions between histones and DNA and other chromatin-bound proteins and therefore may play critical roles in regulating chromatin structure (Horn and Peterson, 2002). In addition to the core histones, another species of histone, linker histone, is present in most metazoan cells. Linker histone binds to the linker DNA between nucleosomes and protects about 20bp of linker DNA from nuclease digestion (Parseghian et al., 2001). Multiple lines of evidence has suggested linker histone plays important roles in mediating interactions between nucleosomes and chromatin fibers and promoting the formation of higher-order chromatin structure (Bednar et al., 1998; Fan et al., 2005; Woodcock et al., 2006).

The studies on purified or reconstituted oligo nucleosome arrays have suggested a hierarchical folding model for chromatin structure. In low ionic environment, the oligo nucleosome arrays exhibit an extended "beads on a string" conformation with a diameter of about 10nm. Upon the addition of 1-2 mM of bivalent cation, nucleosomes on the arrays interact with each other and lead to the formation of a compact, fiber-like structure with a diameter of 30nm (Hansen, 2002). Although there is a general agreement that such "30 nm" fiber is a bona fide chromatin structure and an intermediate state of *in vitro* chromatin folding, how the nucleosomes and linker histones are arranged with in the 30nm fiber structure still remains controversial (Maeshima et al., 2010; Woodcock and Ghosh, 2010). Separate lines of evidence support either of two models for the folding path of 30nm fiber- the one start solenoid model and the two start zigzag model, with some recent evidence suggesting the two modes of folding may coexist in the same 30nm fiber under certain conditions (Grigoryev et al., 2009; Maeshima et al., 2010). Further increases of cation concentration increase the interactions between 30nm fibers and leads to the

formation of high molecular weight aggregates of 10 and 30nm fibers. Importantly, the linker histone H1 and some other chromatin architectural proteins such as MeCP2, MENT and polycomb have been shown to regulate the compaction of the oligo nucleosome arrays and the formation of the thicker chromatin fibers *in vitro*, which may mimic the *in vivo* regulation of chromatin folding (Francis et al., 2004; Georgel et al., 2003; Springhetti et al., 2003; Woodcock, 2006).

In contrast to this reproducible generation of the 30nm chromatin fiber in vitro, there is lack of solid evidence for the existence of 30nm fiber in vivo. Uniform 30nm fibers could only be observed in a few cell types, and in some cases only after a carefully controlled swelling procedure (Horowitz et al., 1994; Woodcock and Horowitz, 1995). On the other hand, a recent study using cryo-electron microscopy of thin section of vitrified mitotic chromosomes did not reveal any ordered fiber structure. In fact, the results suggested the chromatin form a homogenous "grainy" mass (Eltsov et al., 2008). Such a "melted" chromatin state may be due to the extremely high compaction in mitotic chromosomes and therefore may not resemble the chromatin state in less compacted euchromatic regions in interphase nuclei. Still, the gap between the *in vitro* observation and limited *in vivo* evidence complicates the interpretation of folding mechanisms of chromatin.

Models of large-scale chromatin folding

In the interphase nucleus where transcription and DNA replication occur, DNA primarily exists in structures well above the 30nm fiber. Such chromatin structures are usually referred as “large-scale chromatin structure” (Belmont et al., 1999). Under light microscopy, the fluorescent *in situ* hybridization (FISH) signals of 100-200kb genomic regions usually appear as spots below the diffraction limit (Lawrence et al.,

1990; Yokota et al., 1995). Even at transcriptionally active, less condensed chromosomal regions, the compaction level of chromatin is several fold higher than the 1:40 linear compaction ratio expected from 30nm fiber (Mahy et al., 2002; Muller et al., 2004). How the lower-level chromatin units fold into the higher-order chromatin structure remains mysterious due to technical limitations of current methods available for structural analysis of nuclei and chromosomes.

Several models have been used to explain large-scale chromatin folding. A radial-loop, helical folding model for interphase chromatin has been extended from the radial-loop model of mitotic chromatin folding (Manuelidis, 1990). On the other hand, the numerical modeling of the distances between the FISH signals of regions with different genomic separations suggested a giant loop, random walk model (Munkel et al., 1999; Yokota et al., 1995). In this model, chromatin is organized in giant loops several megabase pairs in size. Such giant loops are then attached to a flexible backbone. Finally, an alternative "chromonema" model is supported by the ultrastructure analysis on interphase chromatin using *in vitro* light microscopy and transmission electron microscopy (Belmont and Bruce, 1994; Li et al., 1998; Robinett et al., 1996). In this model, 30nm fibers first organize into chromonema fibers with the diameter of 60-80 nm. These fibers can be further coiled into thicker fibers with the diameter of 100-130 nm.

Most previous studies on chromatin structure using light microscopy, limited by the resolution, could not provide sufficient evidence to confirm any of the proposed models for large-scale chromatin folding. For instance, a FISH study on two genomic loci with the sizes of 300-400 kb revealed that the chromatin regions decondensed into elongated structures consisting of a series of adjacent puncta, or bead-like domains after transcription activation (Muller et al., 2004). However, such

features of chromatin organization can be explained with any of the random-walk, radial-loop and chromonema models. Therefore, technical improvements on light microscopy and electron microscopy which allow the structural analysis of specific chromatin regions at higher resolution will be essential for elucidating the folding paths of large-scale chromatin structure.

Large-scale chromatin structure and gene regulation

Nucleosome positioning, DNA methylation and histone modifications play critical roles in gene regulation by affecting local chromatin structure (Zhou et al., 2011). However, it has been increasingly recognized that the large-scale chromatin structure also has important functional implications on gene regulation.

The correlation between chromatin compaction level and transcriptional activity has been observed at multiple levels of chromatin organization. It has long been noticed that the active genes are often found in less condensed euchromatin while more condensed heterochromatin is enriched in transcriptionally inactive sequences. It has been shown that in mammalian cells the gene-rich, transcriptionally active chromosomes occupy larger nuclear space than the gene-poor chromosomes (Croft et al., 1999). On the other hand, chromatin decondensation was found accompanying transcription activation in different organisms. In insect polytene chromosomes, dramatic chromatin decondensation occurs at highly transcribed chromosome regions, known as the phenomenon “puffing” (Callan, 1982; McKenzie et al., 1975). In mammalian cells, chromatin decondensation was observed at certain genomic regions with sizes from several hundred Kbp to several Mbp after gene activation (Mahy et al., 2002; Ragoczy et al., 2003; Volpi et al., 2000). Such decondensation has also been recapitulated in multiple artificial transgene arrays (Hu et al., 2009a; Muller et al.,

2001; Tumbar et al., 1999).

Large-scale chromatin structure may affect transcription by acting as an accessibility barrier for regulatory factors (Dillon and Festenstein, 2002). Such a model, although attractive, is challenged by observations in yeast and mammalian cells that chromatin-associated proteins such as swi6 and HP1 can rapidly exchange at sites within highly compact heterochromatin (Cheutin et al., 2004; Cheutin et al., 2003; Festenstein et al., 2003). Another study showed inert probes with a radius of gyration of 6 nm, which corresponds to spherical proteins with a molecular weight of more than 1400 KD, could diffuse freely into condensed heterochromatin domains (Verschure et al., 2003). These results suggest that higher-order chromatin compaction itself does not provide an insurmountable barrier for the diffusion of large protein complexes, although it may still slow down the rate of this diffusion. Whether the lower level chromatin structures within different chromatin domains restrain the accessibility of nuclear proteins remains to be tested.

Another mechanism for gene regulation via chromatin organization is chromatin looping (Fraser, 2006). Chromatin loops have been considered as one of the most important structural and functional units in chromatin organization. In a chromatin loop, genes can be placed in proximity with distant regulatory elements, thereby positively or negatively affecting gene expression. The most studied examples for such a regulatory role of chromatin loops are the mouse and human beta-globin loci. The expression of beta-globin genes are regulated by a distant regulatory element called the locus control region (LCR), which is located ~50 kb upstream of beta-globin cluster (Grosveld et al., 1987). The LCR contains several DNase I hypersensitive sites, which may function as enhancers and/or chromatin insulators. Chromatin loops mediate the direct interaction between different beta-globin genes

with LCR during development and give rise to a precise temporal expression pattern of beta-globin genes. The existence of direct regulatory interactions *in vivo* between the DNase hypersensitive sites within LCR and the beta-globin genes has been confirmed by several assays such as RNA tagging and recovery of associated proteins (RNA TRAP) (Carter et al., 2002) and chromatin conformation capture (3C) (Tolhuis et al., 2002).

Furthermore, organizing chromatin into large-scale loops may contribute to gene regulation. The large-scale chromatin loops may physically segregate large genomic regions with distinct transcription patterns away from their genomic neighborhoods and place them to the proximity of specific nuclear compartments. For instance, giant chromatin loops containing several megabases of DNA protruding from chromosome territory were observed at highly transcribed major histocompatibility complex (MHC) gene clusters in several human cell types (Volpi et al., 2000). The upregulation of MHC genes resulted in higher frequency of the formation of such large-scale chromatin loops, suggesting a functional link between organization of large-scale chromatin loops and transcription. At the mouse Hox locus, the formation of large-scale chromatin loops was observed upon the induction of Hox genes (Chambeyron and Bickmore, 2004). The remarkably synchrony between the protrusion of chromatin from chromosome territory and the temporal expression pattern of Hox genes implies a direct regulatory role of the large-scale chromatin loops.

Spatial organization of chromatin within nucleus

The highly ordered chromatin is arranged in the 3D nuclear space in a non-random way. This spatial organization pattern of chromatin may have

important implications on gene expression. Early electron microscopic studies suggested condensed, transcriptionally inactive heterochromatin is usually found near nuclear periphery and surrounding nucleolus, while euchromatin is located in the interior of nucleus. The correlation between nuclear radial positioning of whole chromosomes, chromosome regions and individual genes and their gene density, transcriptional activity and replication timing has been observed in a number of studies. In human lymphocytes, the gene-poor chromosome 18 is usually located toward nuclear periphery while the gene-rich chromosome 19 is usually positioned in nuclear interior (Croft et al., 1999). The genomic regions in early-replicating R band are usually located toward the interior of nucleus compared with the regions in late-replicating G bands (Ferreira et al., 1997; Gilbert et al., 2004). At the single gene level, genes like beta-globin and immune-globin heavy chain are located at the nuclear periphery when they are inactive (Kosak et al., 2002; Ragoczy et al., 2006). Upon gene activation during development these genes relocate to a more interior nuclear position. These investigations point to a link between the spatial organization of chromatin and genome function.

Is nuclear radial positioning a determining factor for gene expression? The answer is likely no. In mammalian cells, the two alleles of the same gene may exhibit different radial positions but usually show similar transcriptional activity (Takizawa et al., 2008a; Takizawa et al., 2008b). This observation is consistent with the fact that there is no radial gradient of concentration for most transcription regulators. The exact functional significance of nuclear radial positioning remains to be revealed.

One extreme case for nuclear radial positioning is the attachment of genomic regions to the nuclear lamina. The nuclear lamina is formed by a layer of specialized

intermediate filaments and underlays the inner nuclear membrane in higher eukaryotes (Dechat et al., 2008). In addition to providing mechanical support to nucleus, the nuclear lamina may play an important role in genome organization and transcription regulation. The mutation of lamin genes leads to aberrant gene expression patterns and a number of laminopathies such as Hutchison- Gilford progeria syndrome and Emery- Dreifuss muscular dystrophy (Gruenbaum et al., 2005). Chromatin associated with the nuclear lamina is depleted of RNA polymerase II and active chromatin markers such as histone acetylation and histone H3 lysine 4 dimethylation (Luco et al., 2008; Sadoni et al., 1999). A recent effort to map the genomic regions preferentially associated with the nuclear lamina identified more than 1000 domains which comprise about 30% of mammalian genome (Guelen et al., 2008). These lamin-associated domains (LADs) are enriched of transcriptionally inactive genomic regions and repressive chromatin markers such as histone H3 lysine 27 trimethylation. Therefore, the nuclear lamina may function in gene repression.

Three recent studies tested whether the association with nuclear lamina had a direct silencing effect on gene expression. All these studies employed an artificial tethering system to relocate transgene arrays or whole chromosomes containing tandem repeats of lac operator (LacO) sequence to the nuclear periphery by expressing fusion proteins consisting of lac repressor (LacI) and lamin protein or lamin associated peptides. Interestingly, different effects of nuclear periphery attachment on transcription were observed in these three independent studies.

In one study, a doxycycline-inducible transgene array was tethered to the nuclear lamina using a Lamin B1-LacI fusion protein (Kumaran and Spector, 2008). This peripheral tethering did not result in changes in the kinetics of doxycycline stimulated transcription or decrease of transcript levels after induction. Rather, the percentage of

cells responding to the induction dropped from 90% to 70%.

In another study, an entire chromosome containing a single LacO array was repositioned to nuclear periphery by expressing a Lap2 β -LacI fusion protein (Finlan et al., 2008). Reduction of transcription was observed at several genes close to the anchoring site. However, most genes in the chromosome did not show detectable changes of transcription level. Finally, a study used an Emerin-LacI fusion to tether a plasmid transgene array to the nuclear lamina (Reddy et al., 2008). This tethering leads to a 75% reduction of the selectable marker gene next to the LacO repeat, the most dramatic change reported so far.

The different results observed in these studies may be due to the different fusion proteins used to tether genes to nuclear periphery but may also reflect gene-specific repressive effects upon attachment to nuclear lamina.

Although gene silencing was observed in all these artificial tethering studies, it is important to note the direct tethering may not recapitulate the peripheral association of genes *in vivo*. It is still unclear whether the peripheral localization of genes is the result of direct interaction with the nuclear lamina or lamin-associated peptides. FISH studies suggested genes within LADs typically show peripheral localization in 30-50% of the cells, implying that the interactions between chromatin with the nuclear lamina may be transient and/or stochastic (Guelen et al., 2008). In contrast, the artificial tethering system resulted in peripheral localization in nearly 100% of cells. To what extent the effects observed with these systems reflect the real influence of nuclear peripheral localization is therefore questionable.

Besides the radial organization of chromatin inside the nucleus, another important theme of the spatial organization of the genome is the association of specific genomic regions with different nucleoplasmic compartments. The proteins involved in

different nuclear activities may accumulate in the nucleus, giving rise to a variety of nuclear subcompartments or nuclear bodies. For instance, transcription is compartmentalized in the nucleus with RNA polymerase II complexes accumulating at many transcription hot spots to form " transcription factories" (Wansink et al., 1993). The proteins involved in splicing of transcripts accumulate into distinct nuclear speckles (Hall et al., 2006). Other nuclear bodies, such as the Cajal body, the PML body, and the polycomb body, all have specialized functions in certain nuclear processes (Bernardi and Pandolfi, 2007; Gall, 2003; Schuettengruber et al., 2007). In mouse B lymphocytes, the Myc gene, upon transcriptional activation preferentially moved to transcription factories preoccupied by the Igh gene which is located on a different chromosome (Osborne et al., 2007). A more recent study revealed that genes regulated by a specific transcription factor Klf1 co-associated at a limited number of transcription factories containing Klf1, again suggesting genome organization could facilitate transcription regulation (Schoenfelder et al., 2010). Similarly, genes regulated in the same pathway in skeleton muscles or erythroid cells tend to co-associate at nuclear speckles (Smith et al., 1999; Xing et al., 1995). Therefore, the spatial organization of genome relative to these functional compartments may have important implications in coordinating nuclear activities and optimizing genome function.

Technologies for studying chromatin and nuclear structure

To understand the organization of large-scale chromatin structure and its functional relationship with transcription and replication, methods for visualizing specific chromatin regions at high resolution are needed. The traditional and most widely used method for visualizing chromatin structure is DNA fluorescent *in situ*

hybridization (FISH), in which DNA probes are used to bind to their complementary sequences in the genome (Lawrence et al., 1990; Yokota et al., 1995). This method is relatively convenient and can be used to detect any genomic regions with nearly any size. However, the DNA hybridization in this procedure requires heat denaturation of nuclear DNA. Such a harsh treatment, while not necessarily producing obvious structural changes of chromatin under the light microscope, results in poor ultrastructure preservation of chromatin and the nucleus at the electron microscopic structural level (Solovei et al., 2002).

One of the goals of the research in Belmont lab has been to develop methods to study the structure and dynamics of chromatin organization. In the 1990s, the Belmont lab first created an *in vivo* tagging system to visualize specific chromatin regions (Belmont and Straight, 1998; Robinett et al., 1996). This method employs the specific interaction between the well studied lac repressor protein (LacI) and the lac operator DNA (LacO). DNA constructs containing direct repeats of LacO sequence were used to generate transgene arrays in mammalian cells. By expressing LacI-GFP fusion protein in living cells or staining with LacI in fixed cells, the chromosome regions tagged with LacO repeats could be visualized.

One obvious advantage of the *in vivo* tagging system is the improvement of chromatin structure preservation. Without the need of heat denaturation, the chromatin structure visualized by LacI binding more likely reflects the native state of chromatin. More importantly, the *in vivo* tagging system possesses several unique advantages which can not be achieved by using other systems. First, this system makes it possible to study the dynamics of specific chromatin regions in living cells. In a previous study, a transgene array tagged with LacO repeats was observed to move from the nuclear periphery to the nuclear interior after tethering a transcriptional

activator VP16 to the array (Tumbar and Belmont, 2001). Using live cell imaging, we demonstrated that this long range movement is unidirectional along a curvilinear path and provided important insights in understanding chromatin reorganization upon transcription activation (Chuang et al., 2006).

Second, the functions of chromosomal proteins such as transcriptional activators, repressors and chromosome architectural proteins can be studied by tethering these proteins to LacO tagged chromatin regions (Carpenter and Belmont, 2004). For instance, a study showed that acidic transcriptional activator VP16 induced dramatic chromatin decondensation at a gene-amplified heterochromatic chromatin region, producing extended large-scale chromonema fibers (Tumbar et al., 1999). Other acidic activators, such as Gal4, Hap4 and p65 were shown to have similar chromatin decondensation activity (Carpenter et al., 2005).

Third, this *in vivo* labeling system is valuable for ultrastructural studies of specific chromatin regions by the electron microscopy. Chromatin regions tagged with LacO repeats can be detected using Nanogold-conjugated antibodies. In a recent study, Nanogold-conjugated IgG was microinjected into the cells before embedding without the need of detergent treatment (Kireev et al., 2008). Such a method not only improved the ultrastructure preservation of chromatin but also significantly improved the labeling efficiency of the Nanogold-coupled antibodies.

Since the establishment of the *in vivo* chromosome tagging system, it has been adapted by a number of groups to study chromatin structure and dynamics in mammalian cells, prokaryotic cells, yeast, Drosophila and C. elegans. Meanwhile, the Belmont lab has been working to improve this system by making it better resemble the native chromatin. The earliest version of the *in vivo* tagging system was used to visualize gene-amplified chromatin regions consisting of hundreds to

thousands copies of plasmid DNA and co-amplified genomic DNA (Li et al., 1998; Robinett et al., 1996; Tumbar et al., 1999). These engineered chromosome regions exhibited high heterogeneity of chromatin compaction levels and might not represent the conformation of endogenous chromatin.

A more physiological relevant system instead used tandem repeats of plasmid transgenes that contained endogenous promoters and/or cis-regulatory elements of genes (Dietzel et al., 2004; Muller et al., 2001). However, variation of chromatin conformation and transcription activation kinetics was still observed among different cell clones carrying the plasmid transgene arrays, suggesting the behavior of these plasmid arrays was under the influence of chromatin position effect. In addition, multi-copy plasmid transgenes were often subject to transgene silencing and heterochromatinization (Dorer and Henikoff, 1997; Garrick et al., 1998). Electron microscopic studies on several cell lines carrying plasmid arrays showed that they form chromatin structures more condensed than surrounding chromatin (Kireev et al., 2008). These drawbacks of the plasmid transgene arrays systems lead to the development of the bacterial artificial chromosome (BAC) transgene array system for studying chromatin structure and dynamics.

BACs are large DNA constructs which can carry 100-300kb of mammalian genomic DNA (Shizuya et al., 1992). They usually contain not only the genes and promoters but also other cis-regulatory sequences important for gene regulation and chromatin organization. BACs have been widely exploited as vectors for gene expression. It has been shown that BAC transgenes in mammalian cells and transgenic animals reproduce expression patterns and levels of endogenous genes. We reasoned that the BACs might also recapitulate the large-scale chromatin structure of endogenous genomic regions.

Recently published studies from the Belmont lab showed that tandem repeats of DHFR, MT or Hsp70 BAC transgenes which consist of several to tens of copies of BACs tagged with LacO repeats appeared as extended arrays of LacI-GFP spots in CHO cells (Hu et al., 2009a). All different stable cell clones containing these BAC transgene arrays showed similar conformation, with the LacI spots separated by 0.2-0.3 μm and each LacI spot containing similar copy number of BAC for each type of BAC transgene. Electron microscopic studies revealed that these BAC transgenes form large-scale chromatin fibers indistinguishable to the surrounding euchromatin regions (Kireev et al., 2008). The BAC transgenes were expressed at levels within several fold of their endogenous counterparts. In the case of Hsp70 BAC transgenes, the BAC transgenes showed very similar rate of gene activation to endogenous Hsp70 genes upon heat shock. Such a synchronous and homogenous pattern of gene activation was not observed in the plasmid transgene system. Together these results suggest that BAC transgene arrays recapitulate the behavior of endogenous chromatin better than previously used gene amplification and plasmid based systems.

The BAC transgene array system provides us some unique opportunities to study chromatin structure. Multi-copy 100-300kb BAC transgenes will allow us to analyze changes in chromatin structure upon transcription activation that may be too subtle to be visualized at endogenous genomic loci. By combining the *in vitro* modification of the BAC DNA and the *in vivo* analysis we will have the capability to identify cis-determinants for a number of chromatin properties and behaviors. By studying the conformation and dynamics of the tandem repeats of BAC transgenes we will be able to move one step closer to understanding the mechanisms for large-scale organization of endogenous chromatin and their functions in regulating nuclear processes.

CHAPTER 2

BAC TG-EMBED: ONE STEP METHOD FOR HIGH-LEVEL, COPY NUMBER DEPENDENT, POSITION INDEPENDENT TRANSGENE EXPRESSION

ABSTRACT

Chromosome position effects combined with transgene silencing of multi-copy plasmid insertions lead to highly variable and usually quite low expression levels of mini-genes integrated into mammalian chromosomes. Together, these effects greatly complicate obtaining high-level expression of therapeutic proteins in mammalian cells or reproducible expression of individual or multiple transgenes. Here we report a simple, one-step procedure for obtaining high level, reproducible mini-gene expression in mammalian cells. By inserting mini-genes at different locations within a BAC containing the DHFR housekeeping gene locus, we obtain copy number dependent, position independent expression with chromosomal insertions of one to several hundred BAC copies. These multi-copy DHFR BAC insertions adopt similar large-scale chromatin conformations independent of their chromosome integration site, including insertions within centromeric heterochromatin. Prevention of chromosome position effects, therefore, may be the result of embedding the mini-gene within the BAC-specific large-scale chromatin structure. The expression of reporter mini-genes can be stably maintained during continuous, long-term culture in the presence of drug selection. This method is extendable to reproducible, high level expression of multiple mini-genes, providing improved expression of both single and multiple transgenes. Finally, we show that the removal of a divergent promoter region from the DHFR BAC results in ~10 folds increase of reporter mini-gene

expression level. We suggest the BAC TG-EMBED should be useful in a wide range of applications involving transgene expression.

INTRODUCTION

Plasmid based expression cassettes using cDNA mini-genes driven by viral or cloned eukaryotic promoters represent the most common method for expression of transgenes. Stable expression is usually achieved by integration of these cassettes into the host eukaryotic genome. However, expression levels of these mini-genes are typically greatly influenced by the chromatin structure surrounding the integration site, producing chromosome position effects, sometimes accompanied by variegation of expression(Karpen, 1994). Transgenes integrated into repressive chromatin regions are expressed at low levels and tend to be silenced over time. This effect is particularly pronounced in mammalian cells. A second phenomenon contributing to low levels of transgene expression is multi-copy transgene silencing, observed for most plasmid mini-genes(Dorer and Henikoff, 1997; Garrick et al., 1998). In transgene silencing, expression per gene copy tends to decrease with increasing transgene copy number such that transgene expression levels do not increase proportionally with copy number and very high copy number insertions may express at levels comparable or even lower than single copy insertions.

The combined impact of chromosome position effects and transgene silencing makes typical transgene expression in mammalian cells both unpredictable and unstable, and has hindered both industrial and clinical applications as well as biomedical research applications. These problems are compounded when cell lines expressing multiple transgenes are required.

As just one example, a number of recombinant proteins are important therapeutic

reagents with enormous market value. Mammalian cell culture has been the dominant expression system for therapeutic protein production as it facilitates both proper protein folding and posttranslational modifications(Andersen and Krummen, 2002; Wurm, 2004). In the absence, however, of a robust, single-step method for reliable, high-level, multi-copy transgene expression, gene amplification remains the method of choice for obtaining high expressing cell clones(Gandor et al., 1995). This process of gene amplification, in which cell mutants carrying hundred of copies of an inserted mini-gene are gradually selected, requires repeated rounds of cell selection, subcloning, and clone characterization over a period of many months. Even then, selection of amplified cell clones with high level, stable expression can be difficult and unpredictable, in many cases requiring a year or more for clone development and stabilization.

To improve the efficiency and minimize the unpredictability of transgene expression, various cis-regulatory elements have been used to flank transgenes to maintain accessible chromatin structure and counteract chromosome position effects. Locus control regions (LCRs)(Grosveld et al., 1987), insulators(Chung et al., 1993; Pikaart et al., 1998), ubiquitous chromatin opening elements (UCOEs)(Antoniou et al., 2003; Williams et al., 2005), Scaffold/Matrix associated regions (SAR/MARs)(Kim et al., 2004; Zahn-Zabal et al., 2001), and antirepressor elements (STAR) (Kwaks et al., 2003)have all been shown to improve transgene expression to some degree(Kwaks and Otte, 2006).

Locus control regions were identified as DNA sequences conferring copy number dependent, position independent expression close to endogenous gene expressions in transgenic mice(Li et al., 2002). Specifically, this property of copy number dependent expression is used as an operational definition distinguishing LCR

elements from enhancers. Known LCRs generally function in a tissue-specific manner, restricting their use to certain cell types. Also, as demonstrated for the beta-globin LCR, they are thought to act by interacting with only one promoter at any one moment in time, limiting their use to promote the expression of multiple transgenes simultaneously(Wijgerde et al., 1995). A “mini-locus” cosmid construct piecing together the LCR, 3’ beta-globin DNase I hypersensitivity site, and other beta-globin regulatory sequences showed highly linear, copy number dependent beta-globin expression over a 1-94 range in cosmid copy number(Grosveld et al., 1987). Copy number dependent expression has also been reproduced after random insertion of 244 kb or 155 kb yeast artificial chromosomes containing the intact beta-globin locus(Peterson et al., 1998). Instead, a 100 kb DNA genomic DNA region containing the entire beta-globin locus and LCR on a bacterial artificial chromosome (BAC) produced variable position effects in transgenic mice, with BAC transgenes expressing anywhere from 0-105% the level of the endogenous genes, with most lines expressing at ~1/3 the level of the endogenous genes(Kaufman et al., 1999). To reduce the original cosmid mini-LCR to a size allowing the use of more traditional cloning vectors, including viruses, “micro-LCRs” have been constructed using various combinations of the small DNA regions corresponding to the DNase I hypersensitivity sites of the LCR. These micro-LCRs are necessary but not sufficient for conferring chromosome position independent expression, with additional and different regulatory regions present in the introns and intergenic regions of specific globin genes, possibly arranged in a specific spatial pattern, required to confer copy number dependent expression(Bharadwaj et al., 2003; Guy et al., 1996; Kang et al., 2004; Rubin et al., 2000; Stamatoyannopoulos et al., 1997). For all of these reasons, use of LCRs in cloning vectors as a general tool for

conferring high-level protein expression or multi-gene transgenesis may be limited.

Insulators are another important class of cis-regulatory elements used to shield transgenes from chromosome position effects created by spreading of activating or repressive chromatin marks from regions flanking the transgene insertion site(Chung et al., 1997). Most published studies using insulator sequences have not included careful analysis of the degree to which these sequences confer copy number dependent expression(Chung et al., 1997; Sekkali et al., 2008). In at least one case where such analysis was done, the chicken HS4 insulator was not sufficient to confer copy number dependent transgene expression in B cells(Truffinet et al., 2005). Copy number dependent expression in this system required additional, LCR-like cis regulatory sequences in the transgene together with the insulator sequence. Most of the other cis-regulatory elements described above also are not sufficient to confer both position independent and copy number dependent transgene expression.

Here we describe a new transgene expression system based on embedding mini-gene constructs within large, cloned mammalian genomic DNA regions. When these DNA constructs are integrated into the mammalian genome they appear to create a reproducible, favorable chromatin environment for transcription independent of their chromosome integration site. We have implemented this system using bacterial artificial chromosomes (BACs) as the cloning vector. Using GFP and mRFP reporter mini-gene constructs, we show that by inserting these reporter genes into a BAC containing the mouse DHFR locus we obtain position independent, copy number dependent expression of reporters in cells clones carrying one to several hundred copies of the BAC. This contrasts with the typical position dependent, copy number independent, and lower expression values observed after directly transfecting cells with the very same mini-genes, with or without flanking chicken HS4 core

insulator sequences.

Therefore in a single stable transformation we can isolate cell clones expressing mini-genes amplified up to several hundred folds. These cell clones show stable chromosome karyotype, transgene copy number, and expression over many passages in the presence of selection. Moreover, we also demonstrate position independent, copy number dependent, simultaneous expression of both RFP and GFP reporter genes inserted within a single BAC construct, showing the ability to use this method for multi-gene transgenesis. This BAC TG-EMBED (BAC TransGene EMBEDded) method provides a new methodology for single and multi-transgene expression capable of facilitating a wide range of applications.

RESULTS

Establishing the BAC TG EXPRESS system in mammalian cells

BACs can carry 100-300kb of eukaryotic genomic DNA insertions (Shizuya et al., 1992) and are therefore large enough to contain not only the coding regions of genes but also many of their distal regulatory elements. In transgenic animals, genes contained within large genomic BAC inserts usually express within several fold of the endogenous gene expression level and show similar tissue expression patterns (Antoch et al., 1997; Heintz, 2000; Kaufman et al., 1999; Yang et al., 1997). We hypothesized that a BAC containing a housekeeping gene locus would assemble into a large-scale chromatin structure permissive for transcription when integrated into the chromosome of any mammalian cell line. Furthermore, we predicted that insertion of typical mini-gene constructs anywhere within this open large-scale chromatin environment would allow high level, reproducible mini-gene expression, independent

of the BAC chromosomal insertion site.

To test this idea we used Tn5 mediated transposition to insert a reporter gene cassette into a BAC with a 175 kb mouse genomic DNA insert containing the full length DHFR gene and most of the Msh3 gene. The DHFR gene is expressed in proliferating cells with a significant upregulation seen at the late G1/ early S phase boundary(Slansky and Farnham, 1996). Expression decreases in quiescent fibroblasts and after differentiation of myoblasts into myotubes(Schmidt and Merrill, 1989a; Schmidt and Merrill, 1989b). The DHFR and Msh3 genes are transcribed from a divergent promoter. In NIH 3T3 cells both the DHFR and Msh3 genes were shown to show similar, growth regulated expression patterns(Schilling and Farnham, 1995), whereas in rat embryonic fibroblasts and human fibroblasts the expression of Msh3 was reported to be constitutive(Iwanaga et al., 2004).

To help visualize the large-scale chromatin structure of the BAC transgene, we inserted a Tn5 transposon containing a 256mer lac operator direct repeat plus a kanamycin selectable marker(Levi et al., 2005). We then used Tn5 transposition again to insert a reporter gene cassette containing an mRFP mini-gene driven by a CMV promoter plus a prokaryotic / eukaryotic zeocin selectable marker (Fig. 2.1a).

A BAC with the lac operator transposon inserted within the Msh3 gene was chosen as the template for transposition with the second reporter gene transposon. Two BAC clones with insertions of the reporter gene transposon within different Msh3 gene introns were selected for further analysis (Fig. 2.1b). In clone C27 the reporter gene is transcribed in the same direction as the Msh3 gene, while in clone C4 the reporter gene is transcribed in the opposite direction. Reporter gene insertion sites of both clones were more than 50kb away from the lac operator repeat to minimize possible silencing effects from repetitive DNA. Linearized BAC DNA

was transfected into mouse NIH 3T3 cells and stable transformants were selected using zeocin. As controls, we transfected just the reporter gene cassette alone, with and without flanking insulator sequences (2 copies of the chicken β-globin core HS4 insulator(Chung et al., 1997) (Fig. 2.1a). Flanking mini-genes with this core insulator sequence previously was shown to reduce chromosome position effects(Emery et al., 2000; Recillas-Targa et al., 2002).

After 3 weeks of selection, reporter gene expression in mixed clonal populations of stable transformants was measured for each DNA construct using flow cytometry (Fig. 2.1c). Flow cytometry of nontransfected cells established the fluorescence background levels. A large fraction of cells transfected with the reporter cassette fragment alone showed no detectable fluorescence above background levels. A broad fluorescent intensity distribution biased towards low expression levels with a long tail extending to higher fluorescence values was observed. Cells transfected with the reporter gene cassette flanked by insulator sequences showed a somewhat decreased fraction of non-fluorescent cells. The distribution was more symmetric with an increased fraction of fluorescent cells.

Strikingly, the reporter gene cassette embedded in either location within the DHFR BAC resulted in essentially 100% of cells with fluorescence values above background levels. Moreover the intensity modes for the fluorescence distributions corresponding to the reporter cassette embedded within the BACs were 1-2 orders of magnitude higher than that observed for the reporter cassette alone or with flanking insulators.

We next examined variability of reporter gene expression within individual cell clones isolated from these mixed clonal populations (Fig. 2.1d-f). Flow cytometry was performed on two clones with comparable mean reporter gene expression levels

from each cell population. Both cell clones with the reporter gene embedded within the BAC transgene showed noticeably more homogeneous expression (Fig. 2.1f), as compared to cell clones containing just the reporter gene cassettes with (Fig. 2.1e) or without (Fig. 2.1d) flanking insulators. In particular, the plasmid based reporter genes produced clones with very broad intensity distributions, with long tails skewed towards lower fluorescence values. Notably, this tail of low expressing cells, indicative of variegated expression within individual cell clones, was absent in cell clones carrying the BAC transgenes.

BAC embedded mini-genes show copy number dependent and position independent expression

A key test of transgene protection from chromosome position effects is to verify copy number dependent, position independent expression. We used qPCR to determine transgene copy number in 7-12 individual cell clones for each transfected DNA construct. We then plotted mean expression levels, measured by flow cytometry, versus mini-gene copy number for each clone (Fig. 2.2a-c). As expected, the reporter gene cassette alone shows poor linearity of reporter gene expression with copy number, indicative of copy number independent expression (Fig. 2.2a, correlation coefficient R^2= 0.153, fixing the y-intercept to the fluorescence background level of non-transfected cells). Somewhat improved linearity is observed for the reporter gene cassette construct flanked by insulators (Fig. 2.2b, R^2=0.657).

However, there is a striking increased linearity for the reporter gene cassette embedded at either location within the BAC. In fact, cell clones derived from BACs containing the reporter gene inserted at either of two locations within the BAC

showed nearly identical expression levels per BAC copy (Fig. 2.2c). The net correlation coefficient calculated using data from all 15 clones (Fig. 2.2c, R^2=0.930), demonstrates a markedly increased copy number dependence produced by embedding the reporter gene within the BAC.

Position independent expression of the transgenes embedded in the BAC is implied by the observed copy number dependence, assuming each clone has an independent chromosome insertion site. We verified different chromosome integration sites for a handful of clones using mitotic chromosome spreads and fluorescence *in situ* hybridization (FISH), revealing single insertion sites for multi-copy BAC integrations, including clones with hundreds of BAC copies (Fig. 2.2d). In contrast to the high degree of chromosome instability observed with gene amplification, all clones showed a single chromosome integration site and uniform size of the multi-copy BAC insertions. Insertions were mapped cytologically both to the middle and ends of chromosomes, including insertions cytologically close to the telomere (Fig. 2.2d, top left) and even one which inserted within the centromeric heterochromatin as confirmed by two-color FISH using a DHFR BAC probe together with a pan-centromeric probe (Fig. 2.2d, top right, and Fig. 2.2e), yet all showed position independent expression proportional to BAC copy number.

BAC transgenes show similar large-scale chromatin structures independent of chromosome integration sites

The design of the new transgene expression system was motivated by our hypothesis that large mammalian genomic inserts cloned within BACs would create a BAC specific, large-scale chromatin conformation independent of chromosome insertion site. Previous work from our laboratory characterizing two different CHO

(Chinese Hamster Ovary) cell lines containing multiple integrated copies of the same DHFR BAC used in this study revealed a large-scale chromatin fiber-like conformation similar to surrounding euchromatic chromosome regions(Hu et al., 2009a).

To examine the large-scale chromatin structure of the multi-copy insertions of the DHFR BAC containing the mRFP reporter gene in a number of independently derived NIH 3T3 cell clones, we performed two-color 3D FISH with both lac operator and DHFR BAC probes (Fig. 2.3). Lac operator FISH signals appeared as arrays of dot-like structures while the BAC FISH signals produced more continuous staining patterns. Particularly for cell clones with smaller copy number BAC insertions, these lac operator FISH signals frequently showed linear configurations, with the BAC FISH signal forming a more continuous, fiber-like conformation. All NIH 3T3 cell clones examined showed similar FISH patterns.

To avoid possible changes in structure induced by DNA denaturation during the FISH procedure, in the cell clones characterized in Fig. 2c-d we transiently expressed an EGFP- lac repressor (dimer)- NLS fusion protein to visualize the lac operator repeats in these cells(Robinett et al., 1996). Interphase nuclei showed string-like chains of GFP stained dots similar in conformation for each clone (Fig. 2.4), including in cells from a clone with the BAC copies inserted within centromeric heterochromatin (Fig. 2.2). Further characterization of these cell clones has revealed a similar spacing between GFP spots (0.33-0.4μm) and a similar ratio of number of BAC copies to GFP spots(Sinclair et al., 2010), suggesting similar large-scale chromatin compaction levels in all cell lines independent of the chromosome integration site.

Stability of BAC transgene array structure and embedded mini-gene expression

To test the long-term stability of transgene arrays formed by chromosomal integration of multiple BAC copies, we continuously passaged three NIH 3T3 cell clones (C4-2, C27-6, C27-13) over a 63-day period, replating cells every three days in the presence of drug selection (75μg/ml Zeocin). Expression levels of mRFP were monitored by flow cytometry after 18, 43, and 63 days of passaging. Expression levels remained at ~80-100% of the starting values for all three clones over this time period, demonstrating stable reporter gene expression with continuous, long-term culture (Fig. 2.5a). After 63 days we also examined mitotic chromosome spreads by FISH to assay the stability of the chromosomal BAC transgene arrays. For all three clones examined, the BAC transgene arrays showed no detectable chromosomal rearrangements, with the chromosomal location and size of the BAC array unchanged, as assayed by the distinctive appearance of the chromosome carrying the BAC insertion and the ratio of the chromosomal length of the BAC transgene array to the total chromosome arm length (Fig. 2.5b).

In parallel, the same cell clones were passaged for 60 days in the absence of drug selection. In this case, mRFP expression levels dropped by 30% to 80% from the starting levels for the three clones examined (Fig. 2.6a). Silencing was not due to loss of BAC transgene copies as mitotic spreads demonstrated that the chromosome location and size of the BAC transgene array remained constant over this same time period (Fig. 2.6b). Silencing also was not accompanied by global changes in large-scale chromatin compaction of the BAC transgene array. Staining of the BAC transgene array by transient expression of EGFP-lac repressor revealed similar numbers of GFP spots and separation between GFP spots in nuclei from silenced clones to that observed in nuclei from parental cells, prior to long-term passaging

without drug selection (data not shown). The observed time course for silencing of the reporter gene in the absence of selection is similar to previously observed reporter gene silencing in plasmids without insulator sequences(Hino et al., 2004; Recillas-Targa et al., 2002), which has been attributed, at least in part, to promoter DNA methylation(Mutskov et al., 2002; Pikaart et al., 1998).

Copy number dependent, position independent co-expression of multiple mini-genes

As described previously, we obtained strikingly similar values of reporter gene expression per transgene copy for NIH 3T3 clones carrying BACs with the reporter cassette inserted at two different locations within the BAC. qRT-PCR data revealed copy number dependent expression both for the kan/neo selectable marker linked to the reporter gene as well as the DHFR gene itself (data not shown). This is consistent with our hypothesis that the BAC DNA creates a global large-scale chromatin conformation permissive for reporter gene expression, while also suggesting that the same BAC embedding approach could be extended to reliable, copy number dependent expression of multiple transgenes.

To test this idea explicitly, we used BAC recombineering(Warming et al., 2005) to insert two different reporter gene cassettes into the parent DHFR BAC. We inserted a CMV promoter-driven EGFP reporter gene cassette containing the kanamycin/neomycin selectable marker, into Msh3 intron 8, the same intron into which the original reporter gene transposon inserted in BAC clone C27. We then inserted the original CMV promoter-driven mRFP reporter gene / zeocin selectable marker cassette into Msh3 intron 19, the same intron into which the original reporter gene transposon inserted in BAC clone C4. We transfected NIH 3T3 cells with the

linearized two-reporter BAC and selected for stable transformants with zeocin. As a control, we also transfected cells with a plasmid construct containing both the mRFP and EGFP expression cassettes.

As observed for the single reporter gene constructs, transfection with the dual reporter gene fragments alone resulted in most stable transformants showing only background fluorescence levels (Fig. 2.7a). Examining this mixed population of cell clones, cells with higher than background fluorescence levels showed poor correlation of mRFP and EGFP reporter gene expression, with most cells showing higher EGFP versus mRFP expression. Only a small fraction of cells showed a linear relationship between EGFP and mRFP expression.

In contrast, in the mixed population of stable colonies derived from transfection of the two-reporter BAC nearly all cells expressed both mRFP and EGFP at higher than background levels (Fig. 2.7b). The average BAC expression levels were roughly 2 orders of magnitude higher than the average plasmid expression levels. Moreover, flow cytometry revealed a striking linear correlation (R^2=0.8262) of mRFP versus EGFP fluorescence intensities. 5/5 individual clones containing the BAC transgene showed a similar ratio of mRFP/ EGFP expression to the mixed clonal population (Fig. 4c, 4 clones shown). These results demonstrate the potential of this BAC transgene embedding method for simultaneous expression of multiple transgenes at reproducible relative expression levels independent of the chromosome integation site, with expression levels proportional to BAC copy number.

Multiple locations with DHFR BAC support similar levels of reporter mini-gene expression

Previously, we showed very similar expression levels for two different locations

of the mRFP reporter gene within the DHFR BAC (Fig. 2.7c). The two-reporter approach described in the previous section provided a rapid way to evaluate the potential of multiple locations within the DHFR BAC to support reporter gene expression and to explore the variability of reporter gene expression when placed at different locations within the BAC.

We randomly inserted the CMV-mRFP-Zeocin transposon into the DHFR BAC already containing the CMV-EGFP reporter, thereby generating a number of BAC clones containing EGFP at the same fixed position but mRFP at different locations (Fig. 2.7d). Six randomly chosen BAC clones (057-GN-RZ-1, 2, 3, 4, 6 and 9) showed CMV-mRFP-Zeocin transposon insertions within different introns of the Msh3 gene, mapping from ~3-80 kb away from the EGFP reporter gene (Fig. 2.7e). Each of these BACs was then transfected into independent cultures of NIH 3T3 cells and mixed populations of stable colonies (dozens to hundreds per flask) from each transfection were selected and propagated together.

After selection, the expression levels of both reporters in these mixed clonal cell populations of cells were measured by flow cytometry (Fig. 2.7e) and compared to stably transfected cells carrying the original two color reporter BAC construct (057-GN-RZ) (Fig. 2.7b). Nearly 100% cells of all mixed clonal populations showed high expression levels of both EGFP and mRFP reporters and excellent linear correlation of EGFP and mRFP expression (Fig. 2.7e). The ratio of mRFP to EGFP fluorescence from each mixed population showed no more than a 2.4 fold variation with no correlation observed between the mRFP/EGFP ratio and the DNA distance between reporters. This small variation in reporter fluorescence ratio not only confirms the previously described chromosome position independent expression of the BAC-TG-EMBED system, but also suggests that mini-gene expression is

relatively independent of location within the BAC, at least throughout the Msh3 gene region.

Deletion of the divergent DHFR promoter increases the expression level of reporter mini-genes

Next, we aim to understand the molecular basis for the transcriptionally permissive environment established on the DHFR BAC transgene arrays. As the first step towards this goal, we want to examine whether there are any sequence elements required for the protection of reporter mini-genes from chromosome position effect and whether they play a role in the establishment of the "open" large-scale chromatin structure.

The expression of DHFR and Msh3 genes are controlled by a divergent promoter region. This region contains a major and a minor promoter 500bp apart from each other, which produce ~90% and 10% of total DHFR transcripts, respectively (Linton et al., 1989). It has been shown that some divergent promoters function as chromatin open elements and can prevent the spreading of surrounding heterochromatin (Antoniou et al., 2003; Williams et al., 2005). Therefore, we would like to investigate whether the divergent DHFR promoter plays a direct role in establishing the permissive environment for the transcription at the BAC transgene arrays.

Using BAC recombineering, we deleted a ~2.8 kb region, which includes both the major and minor divergent DHFR promoters and the first exon of DHFR and Msh3 genes from the DHFR BAC clone C27 (Fig. 2.8a). This modified BAC was introduced into NIH/3T3 cells and stable cell clones containing the tandem repeats of the promoter-less BAC arrays were isolated as described before.

We next examined whether the deletion of the divergent promoter region eliminated the transcription of DHFR and Msh3 genes in the BAC transgene arrays using RNA FISH. The DNA of unmodified DHFR BAC was used to prepare the FISH probes. To our surprise, RNA FISH signals were still detected at the BAC transgene arrays for all clones examined, although at a lower level (data not shown). One possible explanation to this is there are other previously uncharacterized promoters in DHFR BAC and therefore the divergent promoter region is not responsible for 100% of total transcription at the BAC transgene. In fact, previous genome-wide analysis indicated there are other species of Msh3 transcript species originating from sites more than 80kb away from the major divergent promoter region. In addition, it has been suggested that it is common for genes to utilize alternative promoters located at 5' distal regions (Denoeud et al., 2007). And the removal of the major promoters from the BAC could increase the possibility of using alternative promoters. More carefully analysis on the transcripts from the BAC transgene will be needed to understand the transcription events occurring at the promoter-less BAC.

To test whether the divergent promoter deletion affects the expression of the embedded mRFP reporter, flow cytometry analysis was performed on 8 independent subclones carrying the promoter-less BAC. All 8 clones showed high level and homogenous expression of mRFP, similar to the clones carrying the intact BAC transgenes. We then measured the transgene copy number in these clones and fit the mRFP expression levels to the transgene copy number (Fig. 2.8b). The result showed a good linear correlation (R^2= 0.810) which is comparable to the intact DHFR BAC (R^2=0.930), therefore suggesting the mRFP reporter was expressed in a copy number dependent, position independent fashion.

Strikingly, we observed much higher mRFP expression level per copy for the

clones containing the promoter-less DHFR BAC. The slope for the linear fitting between the mRFP fluorescent level and transgene copy number suggested the mRFP, when embedded in the promoter-less BAC, was expressed about ten times higher than embedded in the intact BAC (12.65 versus 1.34) (Fig. 2.8b). Therefore, BAC TG-EMBED method as an approach for transgene expression can be significantly improved by removing the divergent promoter from the DHFR BAC.

The mechanisms underlying this dramatic increase of reporter expression remain under investigation. One hypothesis is there are sequence elements in the divergent promoter region of DHFR which function as transcription repressors on mRFP reporter. Alternatively, the transcription originating from the divergent promoter may inhibit the otherwise very strong transcription of mRFP reporter via some competition mechanisms. Whether this is a general mechanism for the divergent promoters or specific to the DHFR locus is unknown.

We next wanted to analyze whether the removal of the divergent promoter region has any effects on large-scale chromatin structure of the BAC transgene arrays. After transfecting cells with the EGFP- lac repressor fusion protein, the BAC transgenes in all clones appeared as extended arrays of GFP spots which are indistinguishable to the previously characterized intact DHFR BAC transgene arrays (Fig. 2.8c). The separations between neighboring GFP spots were also similar to observed with the intact BAC transgenes. However, more detailed analysis on three clones suggested that each GFP spot represent 12-15 copies of promoter-less DHFR BAC, about two times higher than the 6-7 copies per spot in the clones containing the intact DHFR BAC(Sinclair et al., 2010). Therefore, although the deletion of the divergent promoter region does not seem to alter the general large-scale chromatin fiber-like conformation, it indeed increases the chromatin condensation level within

the BAC transgene arrays. We reason this increased linear DNA compaction ratio may be related to the seemingly decreased transcription level at the promoter-less BAC transgene arrays indicated by RNA FISH. In a previous study, we reported that the DHFR BAC transgene arrays in CHO cells showed a similar two-fold condensation upon the G0 arrest(Hu et al., 2009a). However, this change was not observed on intact DHFR BAC arrays in 3T3 cells. Whether there is a causal relationship between the transcription repression and the chromatin condensation will be further investigated.

DISCUSSION

Here we have described a novel transgene expression scheme that essentially eliminates chromosome position effects in mouse NIH 3T3 cells by embedding mini-transgene constructs within large genomic fragments cloned within BACs. Both the BAC array and expression of the mini-transgene constructs embedded within the BAC remain stable during continuous growth in the presence of drug selection. We suggest this effect is related to the ability of BACs containing housekeeping gene loci to maintain a euchromatin-like large-scale chromatin conformation conducive to transgene expression. Similar levels of expression were observed for multiple locations of the mini-transgene within the BAC. Moreover we showed that by embedding two mini-reporter genes at different locations within the same BAC that our method conferred reproducible, chromosome position independent expression for both reporter genes simultaneously, suggesting the usefulness of this method for multi-gene transgenesis. Although so far only the DHFR BAC was tested, we speculate that BACs with genomic inserts containing other active gene loci may behave similarly. Experiments are now in progress to better define the cis elements

within these genomic loci that confer this behavior.

Our BAC TG-EMBED method is both simple and fast, with a transposon reaction typically requiring just 2 days to insert an expression cassette into a BAC and a single transfection and selection yielding mammalian cell clones stably expressing transgenes at levels up to hundreds of fold higher than a single transgene copy. Therefore by selecting individual cell clones with different BAC copy numbers, a wide range of expression levels should be obtainable from a single stable transfection experiment.

Using our method, we have demonstrated chromosome position independent, copy number dependent expression, with linearly proportional increases in expression over more than a 100 fold variation in transgene copy number. While insulator sequences have generally been used successfully to protect transgenes against chromosome position effects, in a direct comparison the BAC TG-EMBED method produced nearly a 10 fold improved linear correlation coefficient of reporter gene expression versus copy number than that observed by flanking the same reporter gene with two copies of the chicken HS4 core insulator sequence (0.80 versus 0.09).

We anticipate that the BAC TG-EMBED method will be able to be combined with the use of additional cis elements adjacent to the mini-transgene, chosen to maximize or stabilize transgene expression(Kwaks and Otte, 2006), to provide the benefits of both approaches. Indeed, whereas the chicken HS4 insulator sequence is capable of protecting against reporter gene silencing during long-term cell growth in the absence of selection(Pikaart et al., 1998), we observed significant reduction in reporter gene expression using the BAC TG-EMBED method in some clones after 60 days of continuous cell passaging without drug selection. This suggests that the mechanism by which the BAC TG-EMBED method shields reporter genes from chromosome

position effects is different from the mechanism underlying insulator action. Experiments are now in progress to determine the mechanism underlying reporter gene silencing over time in the absence of selection and to test whether flanking the mini-reporter genes with insulator sequences can overcome the observed silencing during long-term cell growth in the absence of selection. An obvious alternative approach would be to test whether the use of eukaryotic rather than viral promoters to drive transgene expression could overcome this effect.

We showed that the expression level of reporter genes could be further improved by about ten folds by a simple deletion of the DHFR divergent promoter region, while the reporter genes were still expressed in a copy number dependent, position independent manner. Although the exact molecular mechanism for this dramatic improvement is still unclear, this promoter deletion method will undoubtedly facilitate a wide range of applications of the BAC TG-EMBED method. We note that in the case of DHFR BAC, the deletion of the divergent promoter did not completely block the transcription at the DHFR BAC transgenes and an "open" large-scale chromatin structure was still observed. These features may be specific to the DHFR locus. Therefore, whether the removal of the divergent promoter can be used as a general method to improve reporter genes embedded at other housekeeping gene loci remains to be seen.

On a broader scope, experiments are also in progress to better understand the basic molecular mechanism by which the BAC TG-EMBED method protects mini-gene constructs from chromosome position effects. Our data indicated the open large-scale chromatin structure formed over the DHFR BAC transgene arrays is correlated with the favorable environment for transcription. Whether the open large-scale chromatin structure is sufficient or necessary for establishing such an

environment is still unknown. The results from the promoter-less BAC transgene arrays suggested that shielding of transgenes from position effect can still occur at large-scale chromatin structure two times more condensed than the intact BAC transgene arrays. Future experiments will be aimed at testing our hypothesis that this protection is related to the ability of the BAC transgene arrays to adapt an open large-scale chromatin conformation and/or to position near specific nuclear compartments. These studies should also improve our understanding on the functional significance of the open large-scale chromatin structure. We note that the methodology described here to compare the expression levels of multiple reporter genes contained within BAC transgene arrays can be adapted to dissect the molecular origin of chromosome position effects produced by specific cis sequences contained within the BAC.

Although we used the CMV promoter for our reporter gene constructs, the DHFR gene contained within the BAC, driven by its natural promoter, also appeared to show copy number dependent expression. We speculate that the BAC TG-EMBED method may work with many other promoters, therefore allowing considerable flexibility for transgene expression. By choosing a BAC containing a cloned genomic region that will assume an open chromatin conformation in the desired target cells, this method may be extendable to expression of transgenes in a wide range of cell types.

With regard to production of therapeutic proteins in mammalian cells, the BAC TG-EMBED method offers several additional key advantages over the current prevailing method, gene amplification. Using a single transfection step, the equivalent of several hundred-fold gene amplification can be obtained. Whereas gene amplification requires the use of transformed cell lines capable of gene amplification, the BAC TG-EMBED method should be applicable to a wide range of

cell types. Finally our results indicate that in contrast to the genome instability inherent in gene amplification, even large BAC transgene arrays show genome stability, as assayed by the stability of the BAC transgene array chromosome integration site and by the size of the transgene array after long-term cell passaging. Similar stability of multi-copy BAC transgene arrays has also been observed in CHO and murine embryonic stem cells (data not shown).

We note that an alternative BAC based expression system for increased expression of a gene present in the genome of the host cell, would be simply to increase the copy number of this gene by transfecting an unmodified BAC containing the actual genomic locus coding for the protein to be expressed. We expect that at least for low copy number insertions this approach would also yield copy number dependent, position independent expression when transfected into a cell type normally expressing this protein. However, the transcription factors regulating the genomic locus might be present at low concentrations, preventing copy number dependent expression at high BAC copy numbers. The endogenous promoter might be weak, preventing high-level expression even at high copy number. Finally, the cell type normally expressing the target transgene might not be amendable to large-scale protein production methods or even in vitro culture. The BAC TG-EMBED method has the advantage of allowing choice of the promoter and cell type to use for expression.

In principle, BAC recombineering could be used to replace the coding region of a housekeeping gene, in a BAC containing the housekeeping genomic locus, with the cDNA for the target transgene. Again though one would still be restricted in the choice of promoters to optimize expression level and control over gene induction and the required BAC recombineering would be more complicated as compared to transposon mediated insertion of a mini-gene using the BAC TG-EMBED approach.

Moreover, this BAC recombineering approach would not work as well as the BAC TG-EMBED method for expression of multiple transgenes.

In conclusion, the BAC TG-EMBED expression method provides a novel, complimentary approach to current transgene expression methodologies with several key advantages for specific applications. Future extensions of this methodology should prove useful for industrial production of therapeutic proteins, production of recombinant proteins for biochemical studies, improved transgene expression for gene therapy, and multi-transgene expression for cell and tissue engineering.

MATERIALS AND METHODS

Plasmid Constructions

Plasmid p[CMV-mRFP-Zeo] (pCRZ) was created from the EZ-Tn5™ pMOD-2<MCS> transposon construction vector (Epicentre Technologies). The SV40-zeocin fragment was cut from pSV40/Zeo2 (Invitrogen) and cloned into the SacI and HindIII sites of pMOD-2 vector to generate an intermediate construct p[Zeo]. The CMV-mRFP DNA fragment was obtained by PCR amplification (forward primer, 5'-CGA GCT CTG AGC TAT GAG AAG CGC CA-3'; reverse primer, 5'-CCC TCG AGT GCC GAT TTC GGC CTA TTG GTT-3') from vector pmRFP(Hu et al., 2009a), digested with SacI and XhoI, and cloned into p[Zeo] to generate pCRZ.

Plasmid pBSKS-cHS4(Wuebbles et al., 2009), a gift from Dr. Peter Jones (University of Illinois, Urbana-Champaign), contains two copies of a 2 copy cHS4 core element (Chung et al., 1997) cloned into the BamH1-Not1 digested pBluescript KS (+) (Stratagene). The CMV-mRFP-Zeo expression cassette was cut from pCRZ by PshAI digestion and cloned into the AflII site of pBSKS-cHS4 to generate

pcHS4-CRZ or into the AflIII site of vector pEGFP-N1 (Clontech) to generate plasmid pGNRZ.

BAC constructions

A 256mer lac operator direct repeat and kanamycin selectable marker was inserted into mouse DHFR BAC 057L22 (CITB mouse library) using a Tn5 transposon as described previously(Levi et al., 2005). BAC Clone 057-k8.32-C29 containing this lac operator transposon inserted 75kb downstream of the Msh3 transcription initiation site was selected for subsequent engineering. The [CMV-mRFP-Zeo] transposon generated by PshAI digestion of pCRZ was inserted into 057-k8.32-C29 using Tn5 transposase (Epicentre Technologies) following the manufacturer's directions and bacterial clones were selected using 25μg/ml zeocin. Primer pMOD-Seq-For (5'-GCC AAC GAC TAC GCA CTA GCC AAC-3') was used to sequence the [CMV-mRFP-Zeo] transposon insertion site; BAC clones C4 and C27 with the [CMV-mRFP-Zeo] transposon inserted into nucleotides 117695 and 23426 of Msh3 gene, respectively, were selected for use in these experiments (Sup. Fig. 1).

DHFR BAC 057-GN-RZ carrying both EGFP and mRFP reporters was constructed by λ Red-mediated homologous recombination(Warming et al., 2005). A EGFP-Kan/Neo DNA fragment flanked by homology arms was created by PCR from vector pEGFP-C1 (Clontech Labratories, Inc.) using 60-bp primers (forward, 5 – ATG AAT GCA CAT CTG TAC ATG CAT TAT TCA TTG TTC TAT GTT TTT GTG ATG CTC GTC AGG - 3'; reverse, 5' – ATT ATA CCA AGA GCA ACT TCA GAA TAA GTT TCC TAG AAT TGG TGG GGA AAA GGA AGA AAC - 3'). Each primer contained a 17bp template homology sequence and a 43bp BAC target site homology sequence. Recombination protocols followed standard protocols

(Warming et al., 2005) and recombinants were selected with 20μg/ml kanamycin. BAC clone 057-GN carrying the EGFP-Kan/Neo DNA fragment was chosen for further recombineering. To insert RFP, a 1.3kb homology region flanking the BAC target site was PCR amplified (forward primer, 5' - GGA ATT CGT TTA AAC CAT GGG TAC TTG GGA GCA CT - 3'; reverse primer, 5' - GGA ATT CGT TTA AAC AAA ACA CAT CTG CCC AGG TC - 3') and cloned into EcoRI digested pMOD-2<MCS>. The mRFP expression cassette was cut from pCRZ using PshAI and inserted into the homology region, within pMOD-2<MCS>, at the EcoRV site. The mRFP-Zeo plus homology region fragment generated by PmeI digestion was used for a 2nd recombination reaction followed by 20μg/ml kanamycin and 25 μg/ml zeocin selection.

To construct BACs containing reporter genes EGFP at a fixed position and mRFP at different locations, the [CMV-mRFP-Zeo] transposon was used to randomly transpose mRFP into BAC 057-GN as described above, generating BAC 057-GN-RZ clones 1-9. Transposon insertion sites and directions for clones were mapped by DNA sequencing. BAC clones 057-GN-RZ-1, 2, 3, 4, 6 and 9 with the transposon insertions into nucleotides 42769, 30688, 27130, 73522, 104954, and 28045 of the Msh3 gene, respectively, were selected for transfection into NIH/ 3T3 cells.

To delete the divergent promoter region from the BAC, the GalK selectable marker was PCR amplified from vector pGalK, a gift from Dr. Neal G. Copeland (National Cancer Institute), using primers 5’- CTG CCT CAC TGA AGA CCA CAG CCT TCA TCC CAC CGT GTC TTC CCG ACG GCC AGT GAA TTG- 3’ and 5’- TTC CTG CTG TCA CCT TCT GTC ACC TGT ATC GGG AAG GTT GGA ATG CTT CCG GCT CGT ATG- 3’. Each of the primers contains a 17bp sequence homologous to pGalK template and a 43bp sequence homologous to the region

flanking the target region. Homologous recombination was performed per standard protocols and the recombinants in which the target region is substituted by GalK were selected on minimal media with 0.2% galactose as the only carbon source.

After each BAC modification, the integrity of the BAC constructs and the length of the lac operator repeat were checked by a HindIII and XhoI digest. Observed gel band sizes were compared with the predicted digestion patterns generated by the Gene Construction Kit software (Textco BioSoftware). Only BAC clones showing the correct digestion pattern and a full length lac operator repeat were chosen for future modifications. BAC DNA for transfection into mammalian cells was prepared with the Qiagen Large Construct Kit.

Cell culture and transfection

NIH 3T3 cells (CRL-1658, ATCC) were grown in Dulbecco modified Eagle's medium (Invitrogen) plus 10% Bovine Growth Serum (HyClone). pCRZ was digested with PshAI and pcHS4-CRZ digested with EcoRV and NotI. DNA fragments containing the CMV-RFP-Zeo expression cassette were gel purified prior to transfection. EcoRV linearized pGNRZ was used for transfection. BAC clones C4 and C27 were linearized with SrfI and BAC 057-GN-RZ was linearized with BsiWI prior to transfection. The linear integrity of all BACs after restriction digestion was confirmed by PFGE. All DNA fragments were ethanol precipitated and transfected into NIH 3T3 cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's directions. Mixed clonal populations of stable transformants were obtained after 3 weeks of zeocin (75μg/ml) selection; individual subclones were obtained by serial dilution. A lentivirus expression construct (Paul Sinclair, Andrew Belmont) was used to transiently express EGFP-LacI. EGFP-LacI binding to the lac

operator repeats localized on each BAC copy provided an outline of the large-scale chromatin conformation of the BAC transgene array in well-preserved nuclei. Three days after lentivirus transduction, cells were plated on coverslips, fixed with 1.6% paraformaldehyde (Polysciences) in calcium, magnesium free PBS buffer and stained with 0.2 ug/ml DAPI (Sigma-Aldrich). BAC transgene array large-scale chromatin conformation was also visualized by 2 color DNA FISH (described below). Deconvolution wide field light microscopy was carried out as described previously(Hu et al., 2009a).

Flow cytometry

Reporter gene expression levels were measured on a MoFlo flow cytometer (Cytomation) using 584 nm and 488 nm laser excitation for mRFP and EGFP respectively. Emission filters centered at 607 nm and 507 nm were used for mRFP and EGFP respectively. Rainbow fluorescent beads RFP-30-5A (Spherotech, Inc.) were used for calibration of both mRFP and EGFP fluorescence. Untransfected NIH 3T3 cells were used to establish background fluorescence levels. The linear fitting of mean RFP expression level versus transgene copy number for each group of clones was performed using Microsoft Excel fixing the y-intercept, a, to the fluorescence background level of non-transfected cells. The correlation coefficient R^2 when the y-intercept is fixed is defined as: $R^2 = bb'$, where $b = (\Sigma x_i y_i - a\Sigma x_i)/\Sigma x_i^2$ and $b' = (\Sigma x_i y_i - a\Sigma x_i)/\Sigma (y_i - a)^2$.

Transgene copy number measurement

Genomic DNA from individual clones was isolated using the Purelink Genomic DNA mini-kit (Invitrogen) and DNA concentration measured with the Fluorescent

DNA Quantitation kit (Bio-Rad). The BAC or plasmid transgene copy number was determined by real-time quantitative PCR using a SYBR-Green PCR master mix (Applied BioSystem) and the iCycler machine (Bio-Rad). Primers used to measure BAC copy number (5'- GAA CTG CCT CCG ACT ATC CA-3' and 5'-CGA GGA GCT CAT TTT CTT GC-3') amplify a 106 bp region within DHFR exon 4. Primers used to measure plasmid copy number (5'-ATG AGG CTG AAG CTG AAG GA-3' and 5'-GTC CAG CTT GAT GTC GGT CT-3') amplify a 114 bp region within the mRFP coding region. Standard curves were measured using serial dilutions of a plasmid containing both mRFP and DHFR (pSV2-DHFR-CRZ).

Mitotic spreads and FISH

Metaphase spreads were prepared and FISH was performed according to standard protocols(Hu et al., 2009a). For mapping BAC transgenes, a biotin labeled probe was prepared from BAC 057L22 DNA using the BioNick Labeling Kit (Invitrogen). For simultaneous detection of the BAC transgene and centromeric sequences in NIH 3T3 clone C4-10, the digoxigenin BAC probe was hybridized to metaphase chromosomes together with a biotin labeled pan-centromeric chromosome paint (Cambio). For two color 3D FISH, a plasmid pSV2-DHFR-8.32 containing the 256mer lac operator repeat(Robinett et al., 1996) was used to generate a biotin labeled lacO probe and the 057L22 DNA was used to generate a digoxigenin labeled BAC probe using the BioNick Labeling Kit and digoxigenin conjugated dUTP (Roche Applied Science). Biotin and digoxigenin labeled probes were detected with Alexa-594 tagged streptavidin (Invitrogen) and fluorescein conjugated anti-digoxigenin antibody (Roche Applied Science), respectively.

ACKNOWLEDGEMENTS

This work was supported by grant number GM58460 and GM42516 from the National Institute of General Medical Sciences awarded to ASB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

We thank Edith Heard (Curie Institute) for providing DHFR BAC (clone 057L22 from CITB mouse library) and Neal Copeland (National Cancer Institute) for providing the reagents for BAC recombineering.

FIGURES

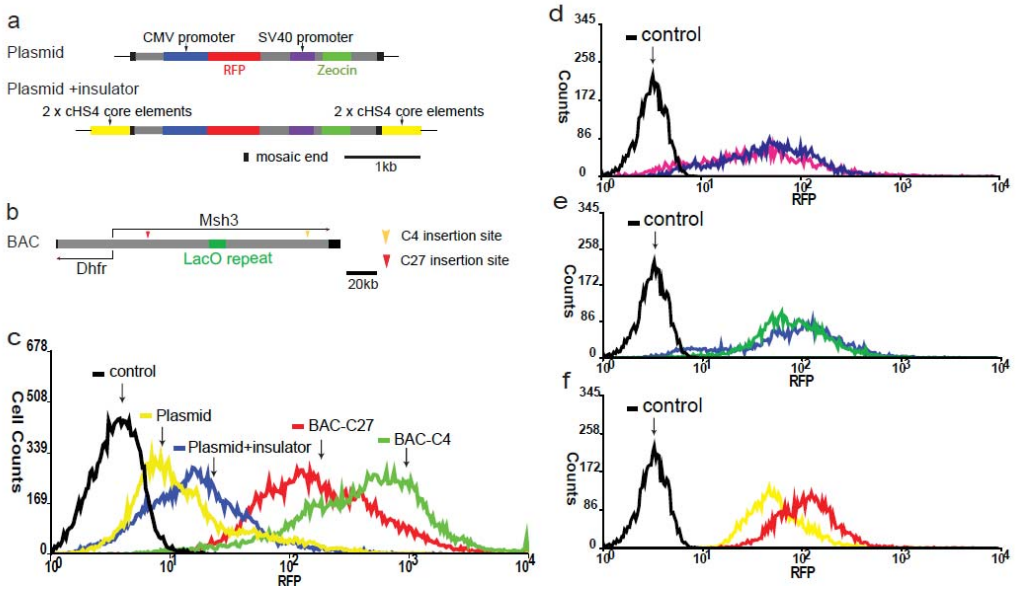


Fig. 2.1 Outline of experimental system. (a) The CRZ (top) and cHS4-CRZ (bottom) expression cassettes showing the 19 bp Tn5 mosaic ends, mRFP and zeocin mini-genes, and the cHS4 insulator sequences. (b) DHFR BAC map showing BAC vector backbone (black), mouse genomic sequence (gray), 256mer lac operator transposon (green), and mRFP expression cassette transposon insertion sites (arrowheads) for C4 (orange) and C27 (red) clones. (c) mRFP expression levels (x-axis) measured by flow cytometry show higher and more uniform expression levels for NIH 3T3 mixed clonal populations of stable transformants after transfection with BAC transgenes (green and red) versus plasmid constructs with (blue) or without (yellow) cHS4 insulator sequences. Nontransfected cells (black) establish the background fluorescence level. (d-f) BAC transgenes produce more homogeneous expression within individual cell clones. Expression distribution for 2 stable cell clones with similar average expression levels selected from each transfection experiment: CRZ plasmid (d), cHS4-CRZ (e), BAC clone C4 (f). Controls (black) correspond to nontransfected cells. (c-f) Fluorescence measured in the same arbitrary units.

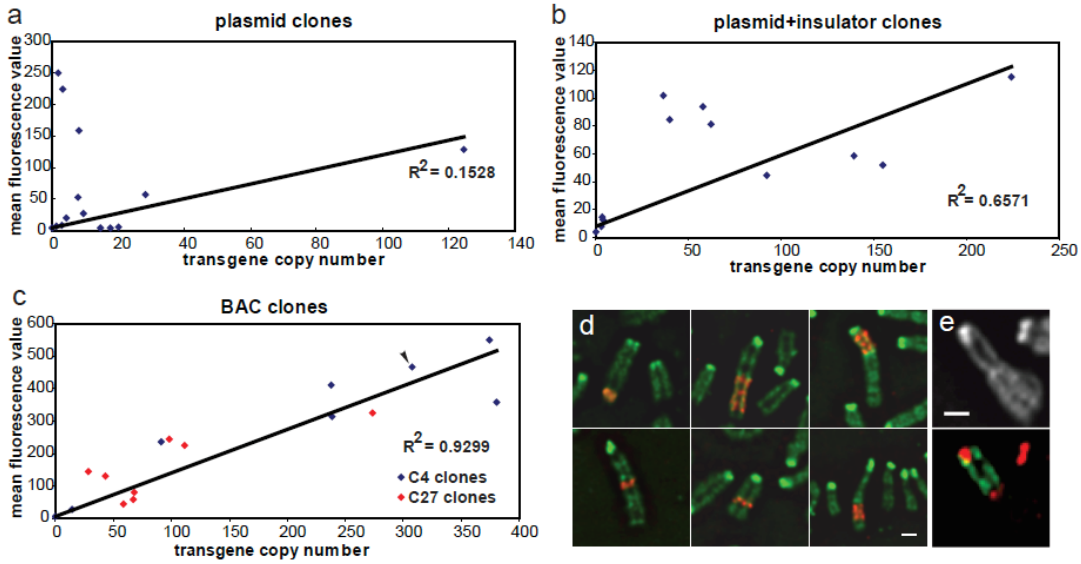


Fig 2.2 Mini-genes embedded in BAC transgenes are protected from position effects. (a-c) Mini-genes embedded in BAC transgenes, but not mini-genes alone, show copy number dependent, position independent expression. Average mRFP expression levels (y-axis) of individual cell clones are plotted versus mini-gene copy number (x-axis) for CRZ mini-gene (a), cHS4-CRZ mini-gene (b), and BAC C4 (blue) or C27 constructs (orange). Linear regression fits (black line) with y-intercepts set at background fluorescence levels are shown with corresponding correlation coefficients. (d) BAC insertion sites and size of inserts are shown by FISH on metaphase chromosomes for cell clones C4-2, C4-6, and C4-10 (top, left to right) and C27-5, C27-12, and C27-13 (bottom, left to right) using a BAC hybridization probe (red) and DAPI (green) counterstain. Interruption of DAPI bright staining (d, top right) and splitting (e) of pan-centromeric FISH signal (red) by BAC FISH signal (green) shows BAC insertion into centromeric heterochromatin in C4-10 clone. Copy number dependent expression in clone C4-10 is maintained (c, arrowhead).

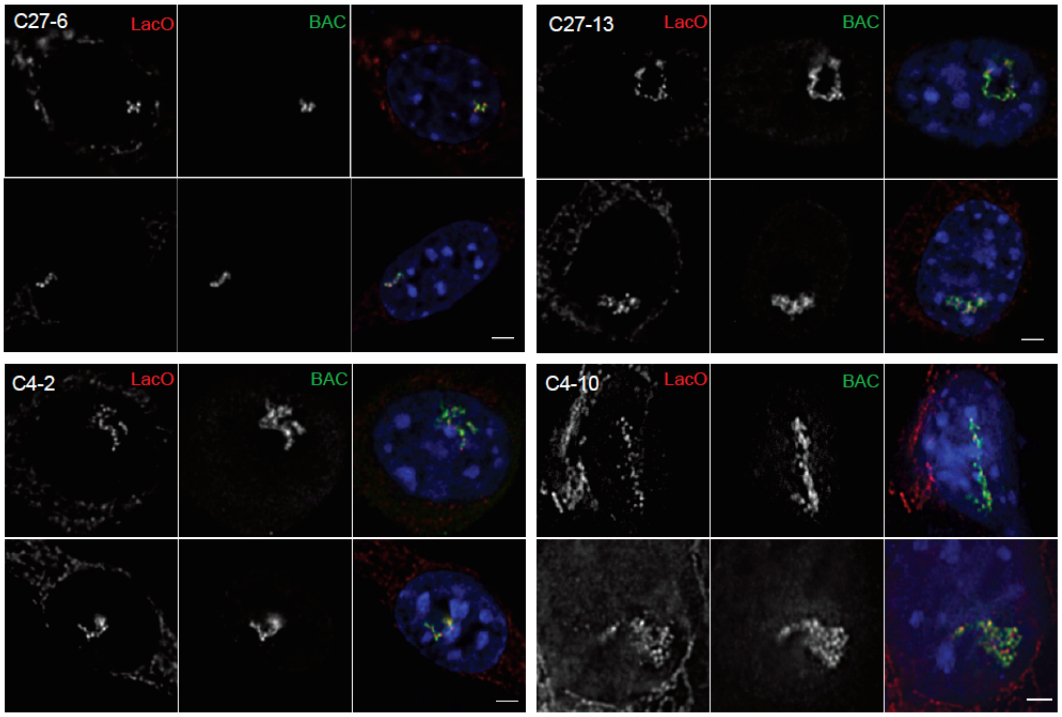


Fig. 2.3 Two color DNA FISH reveals similar, extended large-scale chromatin conformations formed by BAC transgenes in independently derived NIH/3T3 clones. Two examples from each of four NIH/3T3 clones are shown. Lac operator FISH (red), BAC FISH (green), and DNA DAPI staining (blue) show large BAC transgene arrays typically located within the nuclear interior. Extended, linear configurations are frequently observed. Spacing between lac operator FISH foci is similar in different clones. Scale bar = 2 μm.

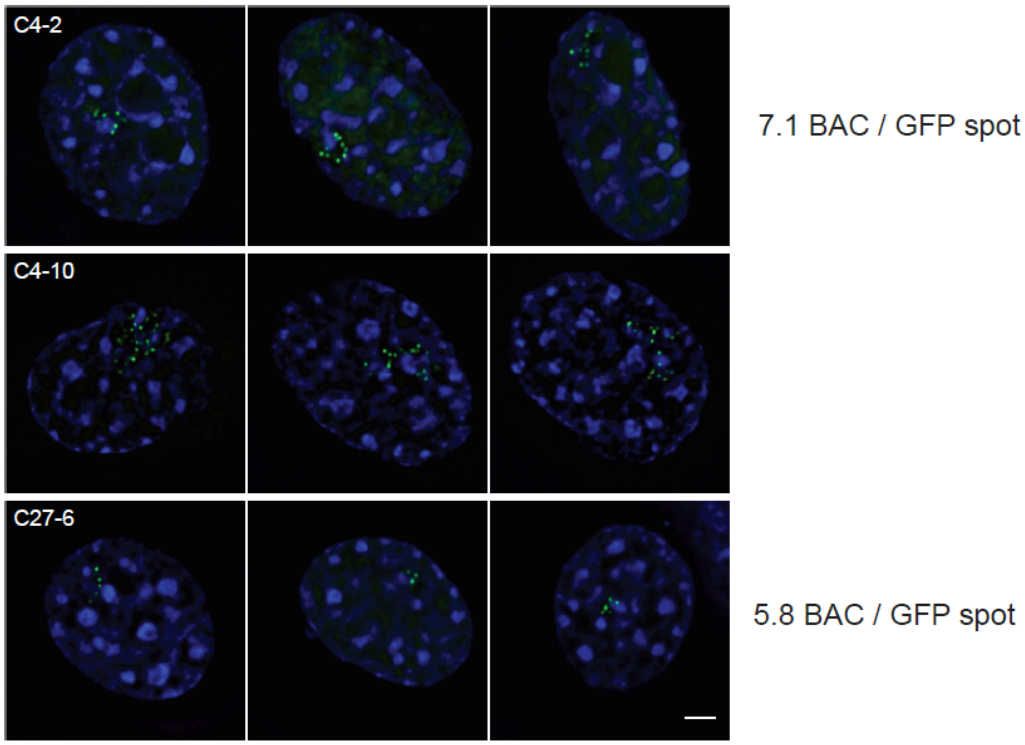


Fig. 2.4 Mini-genes are located within similar large-scale chromatin structures formed over BACs independent of chromosomal insertion site. Similar interphase large-scale chromatin conformation is observed for BAC transgenes in three cell clones as judged by spacing between GFP-LacI (Green) spots. DNA is stained by DAPI (blue). Scale bar = 2 μm.

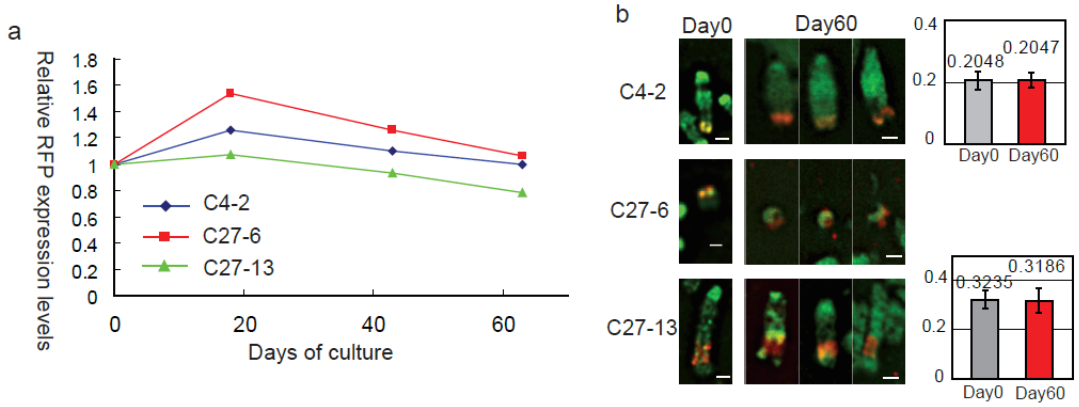


Fig. 2.5 BAC transgenes and expression of embedded reporter genes are stably maintained with selection. (a) Expression levels of mRFP reporter gene in three NIH/3T3 clones over 60 days of continuous passaging are plotted relative to values measured on day 1. Values represent the average cellular fluorescence as measured by flow cytometry. (b) Sizes and chromosomal locations of BAC transgene arrays within metaphase chromosomes remain stable before and after 60 days of culture as visualized by FISH using DHFR BAC probe (red) and DAPI staining (green). Scale bar = 1 μm. Right- BAC transgenes in clone C4-2 and C27-13 map close to the telomere, while the BAC transgenes in clone C27-6 are located on a mini-chromosome. Left- For clones C4-2 and C27-13, the ratios between the chromosomal arm length of the BAC transgene array relative to the total chromosome arm length were measured from more than 10 examples for each time point. Mean values and standard deviations of the ratios are plotted. No obvious changes in BAC trangene sizes are observed.

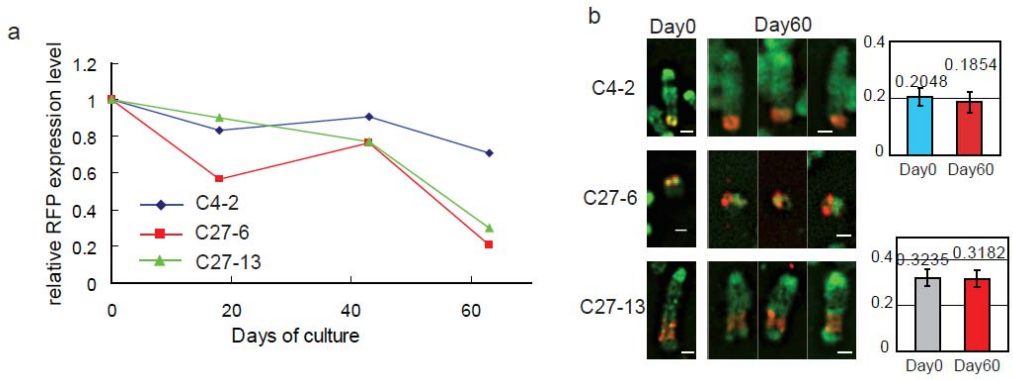


Fig. 2.6 Decrease of reporter gene expression over time in the absence of drug selection despite chromosomal stability of the BAC transgene array. (a) Three NIH/3T3 clones were propagated in the absence of Zeocin for 60 days. mRFP mean expression levels, as measured by flow cytometry, are plotted relative to initial values at day 1. A several fold drop in expression is observed between days 40 and 60 for 2 of the three cell clones. (b) Left- Metaphase chromosome location and size of BAC transgene arrays are analyzed by DNA FISH (green- DAPI DNA staining, red- BAC FISH). Right- The sizes of the BAC transgene arrays, measured by the ratio of the length of the chromosome region containing the BAC transgene array relative to the total chromosome arm length, show no obvious changes after 60 days of cell passaging without drug selection.

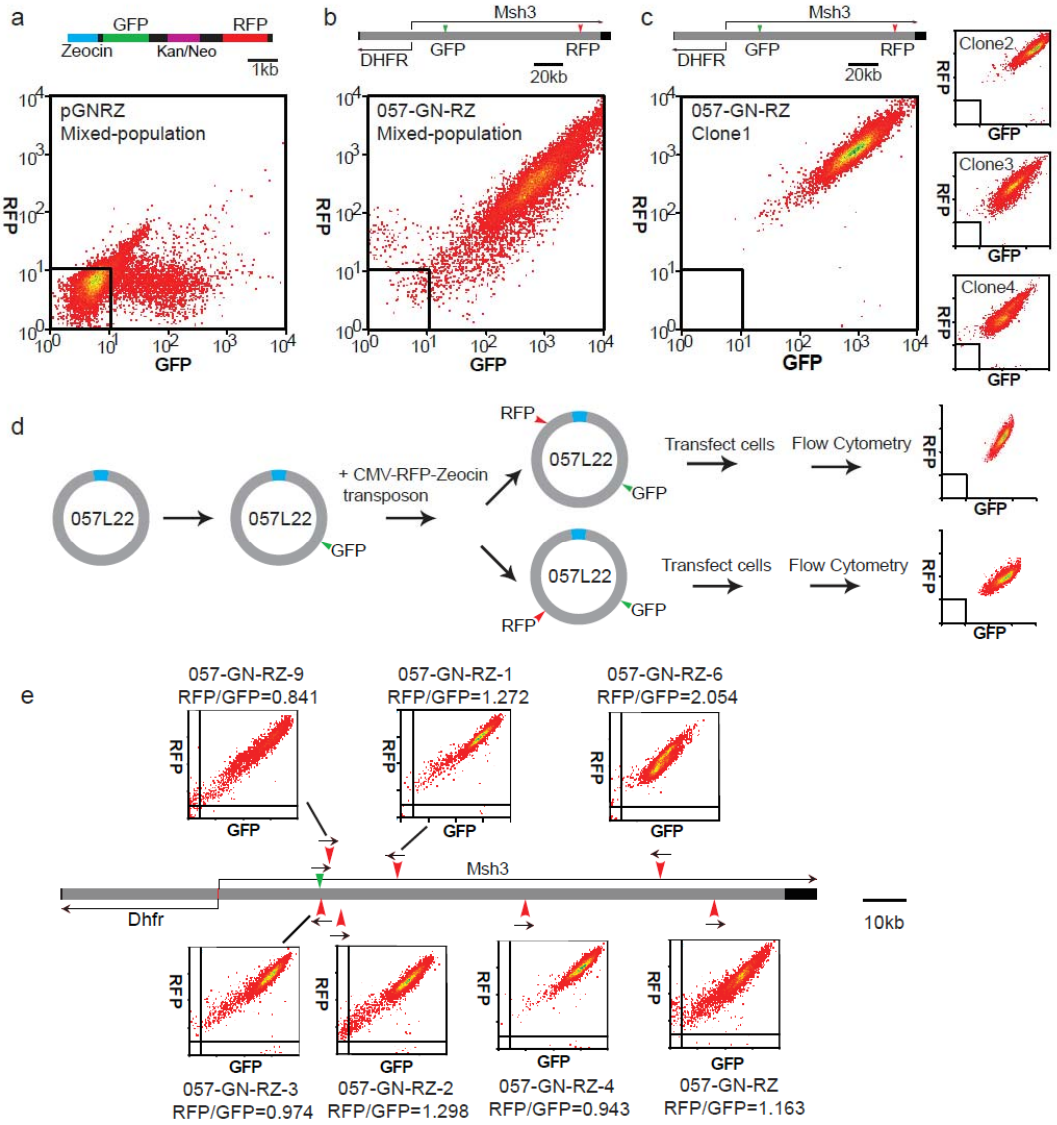


Fig. 2.7 Two reporter mini-genes embedded within single BAC show near constant expression ratios and higher and more uniform expression levels than plasmid constructs containing both mini-genes. (a & b) Scatter-plots showing EGFP (x-axis) and mRFP (y-axis) expression levels measured by flow cytometry of mixed clonal populations of NIH/3T3 cells stably transfected with pGNRZ plasmid construct (a) or DHFR BAC 057-GN-RZ containing both mini-gene reporters (b). Each spot represents a single cell measurement. Black rectangles (lower left) demarcate ranges of background green and red fluorescence values measured for control, nontransfected cells. Majority of cells carrying plasmid construct show only background mRFP or GFP fluorescence levels; of the cells showing above average background fluorescence the majority show higher than background GFP expression but background levels of mRFP expression. Only a small fraction of cells with

Fig. 2.7 (con't)
plasmids show linearly correlated GFP and mRFP expression. This contrasts with high level, linear coexpression of both reporter genes in cells carrying the BAC construct. (c) High level, linearly correlated, more uniform co-expression levels for reporter genes are observed in four different cell clones carrying DHFR BAC 057-GN-RZ. (d) The two-reporter gene BAC can be used to measure the relative expression levels of reporters inserted at different locations within the BAC. A BAC carrying the GFP reporter at a fixed site was used as the template for random transposition to insert the mRFP reporter at different BAC locations. After separate transfection of BACs carrying the mRFP at different locations, stable, mixed clonal populations derived from each of these BACs were analyzed by flow cytometry. (e) Ratios between the mean values of mRFP and EGFP reporter expression of mixed clonal cell populations derived from a total of 7 two-reporter BACs are summerized. The green arrowhead indicates the insertion site of the EGFP reporter in all BAC constructs. The red arrowheads indicate the insertion sites of mRFP reporters in different BAC constructs. Arrows indicate the directions of reporter gene transcription. All mixed populations show high level and linearly correlated expression for both reporters, with expression ratios varying less than 2.4 fold across the BAC.

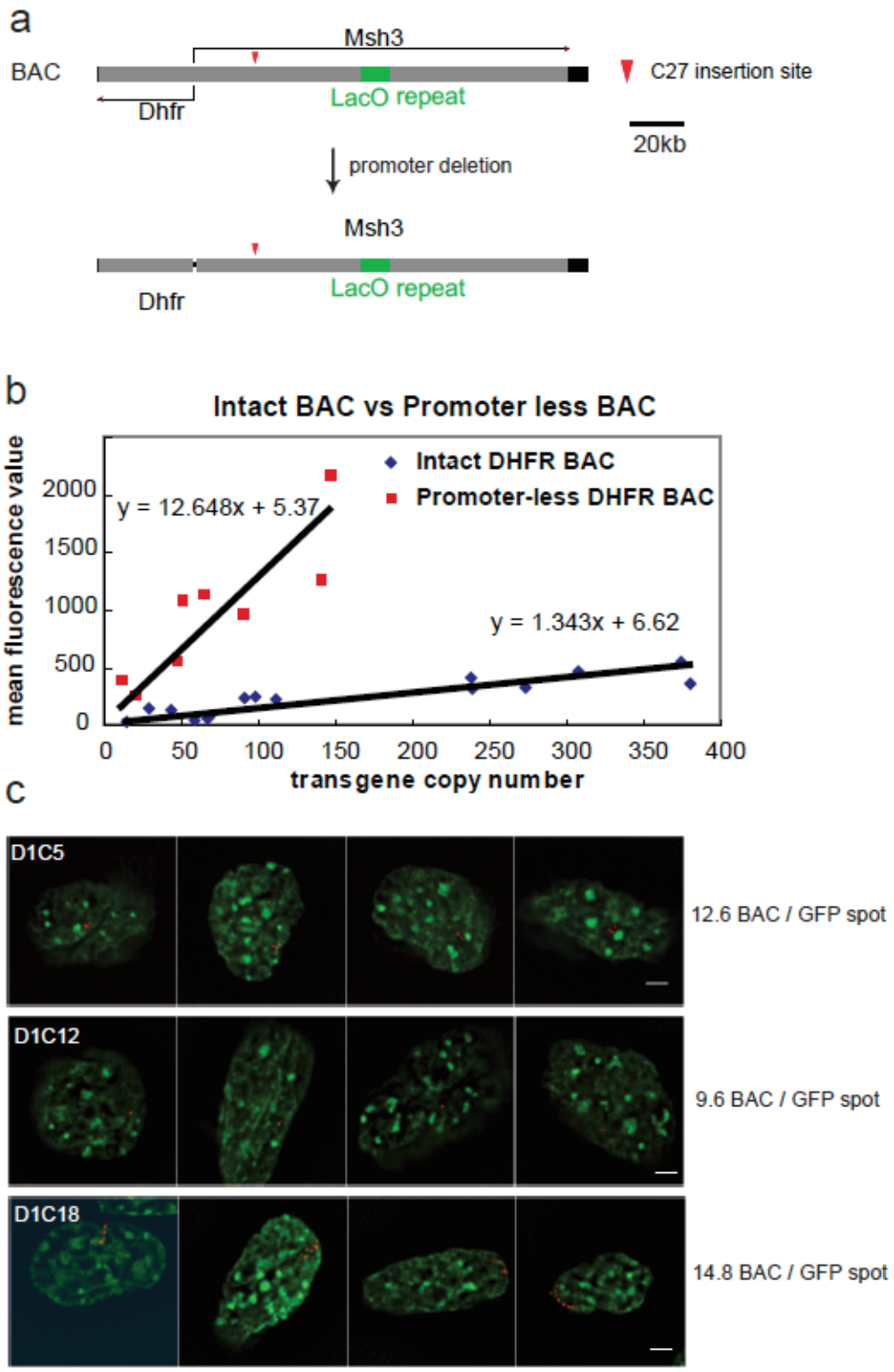


Fig. 2.8 The deletion of divergent DHFR promoter increases the expression of reporter mini-genes. (a) The 2.8kb divergent promoter region was removed from the DHFR BAC containing mRFP reporter. (b) The mRFP mini-genes embedded in the promoter-less BAC show copy-number dependent, position independent expression. Average mRFP expression levels (y-axis) of individual cell clones are

Fig. 2.8 (con't)
plotted versus mini-gene copy number (x-axis) for promoter-less BAC (red) and the full-length DHFR BAC (blue). Linear regression fits (black line) with y-intercepts set at background fluorescence levels are shown. The slopes for the regression lines indicate mRFP embedded in the promoter-less BAC express at higher level than embedded in the intact DHFR. (c) The promoter-less BAC transgenes in three stable cell clones appear as extended arrays of LacI-GFP spots (Green) similar to the intact DHFR BAC transgenes. Average BAC copy number represented by each LacI-GFP spot is listed. DNA is stained by DAPI (blue). Scale bar = 2 μm.

CHAPTER 3

IDENTIFYING DNA SEQUENCES REQUIRED FOR LOCALIZATION OF THE HUMAN BETA-GLOBIN LOCUS TO THE NUCLEAR PERIPHERY

ABSTRACT

The intranuclear positioning of chromatin has been suggested to have an important role in gene regulation. In higher eukaryotic cell nuclei, the nuclear periphery is enriched in heterochromatin. Recent studies have demonstrated that attachment of genes to the nuclear periphery can repress their transcription. However, the mechanisms underlying the nuclear peripheral localization of endogenous genes have not yet been elucidated. Here we report our initial work aimed at identifying cis elements required for the peripheral localization of the human beta-globin locus. Previously, we had demonstrated that in mouse ES cells a BAC carrying a 200 kb human genomic fragment containing the beta-globin locus targets to the nuclear periphery independent of the chromosome integration site. We now show this autonomous targeting of the beta-globin BAC to the nuclear periphery is also observed in mouse fibroblasts. Using BAC recombineering, we have demonstrated that globin related sequences are not required for the nuclear peripheral localization of the beta-globin BAC transgene arrays. Additional work identified two genomic regions sufficient to target the BAC transgenes to nuclear periphery. Moreover, we demonstrated the association of the BAC transgenes with centromeric heterochromatin is also sequence-dependent and is regulated by mechanisms different to nuclear peripheral localization. Finally, we showed that targeting the transcriptional activator VP16 to beta-globin BAC arrays resulted in a relocation of the arrays to the nuclear interior, suggesting that transcriptional activators can reverse

this peripheral localization. Ongoing experiments are in progress to further dissect this region to identify specific sequences required for peripheral localization as a first step towards elucidation of the molecular mechanism for the nuclear peripheral localization of genes in mammalian cells.

INTRODUCTION

The spatial organization of eukaryotic genome in 3D space of nucleus is not random. In several normal and tumor human cell lines, gene-rich chromosomes were shown to locate towards the nuclear interior while the gene-poor chromosomes adopted a more peripheral nuclear localization (Cremer et al., 2003; Croft et al., 1999). Multiple lines of evidence also have suggested a general correlation between nuclear radial positioning and gene activity, with active genes preferentially located in the nuclear interior and silent genes located at the nuclear periphery(Takizawa et al., 2008a; Zink et al., 2004).

The localization of genes at the nuclear periphery has been linked with gene repression. Electron micrographs have revealed that transcriptionally inactive, highly compacted heterochromatin accumulates adjacent to the nuclear envelope in typical mammalian cells (Akhtar and Gasser, 2007). Consistent with these observations, the nuclear periphery has been shown to be relatively depleted of markers of active transcription such as histone acetylation (Sadoni et al., 1999), histone H3 lysine 4 dimethylation (H3K4me2) and RNA polymerase II (Luco et al., 2008). These results suggest that the nuclear periphery is either relatively depleted of active genes and/or creates a repressive environment for transcription. Moreover, several developmentally regulated genes have been shown to locate at the nuclear periphery in their silent state and reposition to the nuclear interior upon gene

activation; examples include the beta-globin genes during murine erythroid maturation (Ragoczy et al., 2006), the IgH gene during murine B cell differentiation (Kosak et al., 2002) and the Mash1 gene during murine neuronal differentiation (Williams et al., 2006). These results suggest that the nuclear peripheral localization of genes may play a role in maintaining the gene repressive state during development.

The functional consequences of nuclear peripheral localization have been directly tested experimentally in yeast (Andrulis et al., 1998) and mammalian cells (Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008) by artificially tethering chromosome sites to the nuclear lamina. In yeast, tethering to the nuclear periphery did not result directly in gene silencing per se but did appear to strengthen silencing associated with telomere position effects, restoring gene repression to a defective silencer. In the three mammalian systems tested to date, the actual fusion proteins used for tethering, the reporter constructs and genes used to monitor changes in expression, and the number of copies of the tethering sequence have all differed making it difficult to directly compare results from each study. However, a general conclusion emerging from these studies viewed together is that the degree of gene repression associated with tethering to the nuclear periphery is gene and promoter specific, and typically quantitative, causing a modulation in transcription, rather than qualitative (i.e. turning transcription from on to off).

To fully understand the physiological significance of the nuclear peripheral positioning of genes, the molecular mechanisms underlying such positioning needs to be unveiled. Recent studies using the DamID method (Greil et al., 2006; Vogel et al., 2007) have mapped the preferred genome-lamina interactions in Drosophila and cultured mammalian cells (Guelen et al., 2008; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). In human fibroblasts, more than 1300 sharply defined domains with

sizes of 0.1- 10 megabases were shown to preferentially interact with the nuclear lamina. These lamina-associated domains (LADs) are enriched in genes with low expression-levels and repressive chromatin marks such as H3K27me3 and H3K9me2, while depleted of the active chromatin mark H3K4me2.

The discovery of LADs revealed important features of the spatial organization of mammalian genome. However, our understanding on how specific genomic regions are located to nuclear periphery is rudimentary. Whether there is a causal relationship between the direct molecular interaction with nuclear lamina components and the peripheral localization of genomic regions at the microscopic level remains unknown. It is also unclear whether specific sequences within the LADs serve as the "anchor points" for nuclear periphery targeting or whether targeting to the nuclear periphery is a general property for all the sequences within the LADs. To date, a functional assay to directly test the ability of specific sequences to target to the nuclear periphery is lacking.

Here we describe a BAC transgene system for identifying sequence determinants of nuclear peripheral localization of specific genomic regions. BACs are large DNA constructs, typically carrying 100-300 kb mammalian genomic DNA fragment (Shizuya et al., 1992). Previous studies in our lab have suggested that transgene arrays consisting OF tandem repeats of the BACs recapitulate several important features of endogenous genes, including similar expression levels to the corresponding endogenous genes, chromatin decondensation after gene activation, and localization of the BAC transgenes to the same specific nuclear compartments as the endogenous gene locus (Bian and Belmont, 2010; Hu et al., 2009a; Sinclair et al., 2010). Using a BAC recombineering system (Warming et al., 2005) to delete specific sequences from the BAC, in theory we should be able to identify specific

genomic regions within the BAC sequence required for gene targeting to specific nuclear compartments.

In this study, we dissected a ~200kb human genomic region containing the beta-globin locus. Beta-globin genes are specifically expressed in erythroid cells. It has been well documented that the beta-globin locus is preferentially associated with the nuclear periphery in cell types where it is inactive, such as ES cells (Hepperger et al., 2008; Ragoczy et al., 2006). Previously we have shown that a BAC containing a 200kb human genomic region including the beta-globin locus autonomously targeted to nuclear periphery in murine ES cells (Sinclair et al., 2010). Here we demonstrated the same BAC also targeted to nuclear periphery in murine fibroblasts. By systematically deleting sequences from the 200kb BAC, we showed the beta-globin gene cluster itself and its regulatory elements, including the locus control region (LCR), the upstream Hispanic region (UHR) and 3' DNase I hypersensitive site (3' HS1) are not required for the peripheral localization of the genomic region. We identified two genomic regions, 32 and 48 kb in size, which independently are sufficient to target the BAC transgenes to nuclear periphery. We also observed the association of these beta-globin BAC transgenes with mouse chromocenter, another heterochromatic compartment, changed upon sequence alterations. Finally, we showed that these beta-globin BAC transgene arrays relocated from the nuclear periphery to the interior upon tethering of the acidic transcriptional activator VP16. Further sequence dissection is in progress to define the minimal sequence elements for determining the nuclear peripheral localization of this locus. Further analysis of these cis-elements should eventually lead to the elucidation of at least one molecular mechanism for nuclear peripheral localization of genes in mammalian cells.

RESULTS

A BAC containing the human beta-globin locus autonomously targets to the nuclear periphery in mouse fibroblasts.

Previous studies in our lab had shown that the transgene arrays consisting of multi-copies of a 207kb BAC containing the human beta-globin locus targeted to the nuclear periphery in mouse ES cells independent of the chromosomal insertion sites (Sinclair et al., 2010). In three randomly selected stable clones carrying the beta-globin BAC, BAC transgene arrays partially overlapped with the nuclear pore staining signals in 50%-80% cells. This extreme peripheral localization is specific to the beta-globin BAC, as DHFR and alpha-globin BAC with similar size to beta-globin BAC did not show obvious nuclear periphery targeting. These results suggested that the 207kb human beta-globin genomic fragment may contain functional DNA elements responsible for autonomous targeting of this locus to the nuclear periphery.

As a strategy to identify the cis requirements for this peripheral targeting of the beta-globin locus, we decided to test systematically the ability of different BACs with specific deletions within the beta-globin genomic insert for their ability to target autonomously to the nuclear periphery. Our approach was to use BAC recombineering techniques to remove specific sequences from the BAC and then introduce this modified BAC DNA into cells using stable transfection to establish independent cell lines with the BAC DNA integrated at different chromosome locations. Use of a parental cell line already stably expressing GFP-LacI would facilitate visualization of the location of these beta-globin BAC transgenes relative to the nuclear periphery (Fig. 3.1).

Due to the technical difficulty of obtaining stable ES cell lines stably expressing

GFP-LacI, and the resulting difficulty of screening of ES clones carrying BAC transgene arrays, we decided to perform this sequence interrogation study in mouse NIH/3T3 fibroblasts. We first determined whether the beta-globin BAC would autonomously target to the nuclear periphery in these 3T3 mouse fibroblasts as previously observed in mouse ES cells. The beta-globin BAC containing a 256mer lac operator repeat was transfected into a NIH/3T3 cell line which expressed GFP-LacI at a suitably low level allowing sensitive detection of BAC transgenes. Stable clones carrying BAC transgenes were isolated and clones containing only one BAC transgene insertion site were selected for further analysis.

In 4 out of 5 randomly selected NIH/3T3 cell clones, the beta-globin BAC transgenes were located within 0.5μm from the nuclear periphery (defined by 4', 6-diamidino-2-phenylindole (DAPI) staining) in 30-64% of the cells (Fig. 3.2 a,d). Association with the nuclear periphery was measured in two dimensions from single optical sections. The percentages of peripheral localization of BAC transgenes in all 5 clones are significantly higher than the 14.8% predicted for an average-sized 2D nucleus with diameters of 16 and 11μm (Fig. 3.2 d). Similar patterns of nuclear peripheral localization were also observed using Lamin A or nuclear pore immunostaining to define the nuclear periphery (Fig. 3.2 b,c). In contrast, three previously established NIH/3T3 clones carrying DHFR BAC transgene arrays showed 5-16% of peripheral localization (Fig. 3.2 d). Therefore, we concluded that the human beta-globin BAC confers BAC-specific, autonomous nuclear peripheral localization in mouse NIH/3T3 fibroblasts.

The peripheral localization pattern of the beta-globin BAC transgenes recapitulates the behavior of the endogenous human beta-globin locus in human fibroblasts. Fluorescent *in situ* hybridization (FISH) using the human beta-globin

BAC as a probe following immunostaining against Lamin A was performed in human fibroblasts BJ-hTERT cells. Measurements of the distance of the endogenous human beta-globin locus from the nuclear periphery in BJ-hTERT cells were compared to measurements of the distance of the human beta-globin BAC transgenes from the nuclear periphery, also defined by Lamin A immunostaining, in NIH/3T3 cell clone HBB C3. The BAC beta-globin transgene in the HBB C3 clone showed a pattern of localization relative to the nuclear periphery strikingly similar to the endogenous beta-globin, with more than 50% of spots localizing within 0.5 um to the nuclear periphery as measured in 2D optical sections (Fig. 3.2 e,f). In contrast, no significant nuclear peripheral localization was observed for the human alpha-globin locus in BJ-hTERT cells (Fig. 3.2 d).

Beta-globin genes, locus control region (LCR), upstream Hispanic region (UHR) and 3' HS1 are not required for the nuclear peripheral localization of the beta-globin BAC transgene arrays.

There has been a long standing theory that the nuclear peripheral localization of genes may have a functional impact on gene regulation. Moreover, the beta-globin locus moves from the nuclear periphery to the nuclear interior during erythrocyte maturation (Ragoczy et al., 2006). Therefore we first tested whether known regulatory elements of the beta-globin locus were required for the peripheral association of the beta-globin locus in non-erythroid cells.

The human beta-globin gene cluster consists of five globin genes (HBE1, HBG1, HBG2, HBD and HBB). A 39.5 kb deletion upstream of the beta-globin gene cluster, named the Hispanic deletion, causes profound alteration of beta-globin gene expression. The most important regulatory region within the Hispanic deletion is the

LCR (locus control region) containing six DNase I hypersensitive sites (HSs). The LCR is located 6-22 kb upstream of the HBE1 gene and is required for the high expression level of all beta-globin genes. The 27.5kb region upstream of the LCR which was also removed by Hispanic deletion is called the UHR (upstream Hispanic region) and may also contain functional elements for beta-globin gene regulation. Two erythroid-specific DNase I hypersensitive sites (HS), the 3'HS1 located at the 3' side of beta-globin gene cluster between the HBB and olfactory receptor OR51V1 genes and HS-110 located 110kb upstream of HBE1 between olfactory receptor genes OR51B6 and OR51M1, also have been demonstrated as important for beta-globin regulation.

We used a λ-red based BAC recombineering system and a GalK based dual selection system to remove specific sequences from the beta-globin BAC (Warming et al., 2005). The use of GalK as a selectable marker for both positive and negative selection allowed us to remove the GalK after one deletion and use it again in the selection for a subsequent deletion. A single deletion (D5) removed a ~116kb region between olfactory receptor genes OR51V1 and OR51B2, deleting the entire beta-globin gene cluster as well as the LCR, UHR and 3'HS1 sequences (Fig. 3.3 a). The HS-110 site, which is equivalent to the HS-62.5 site in mouse, could not be removed in the same deletion due to the insertion of the LacO repeat between OR51B2 and OR51B5 in previous BAC engineering.

We transfected the deleted BAC (HBBD5) into NIH/3T3 cells and analyzed the nuclear localization of the BAC transgenes in stable clones. All 5 randomly picked clones showed at least 30% peripheral localization of BAC transgenes, with 4 out of 5 clones showing more than 50% peripheral localization (Fig. 3.3 b,f). Therefore, HBBD5 BAC transgenes still conferred the autonomously localization to nuclear

periphery and we concluded that the beta-globin genes themselves, LCR, UHR and 3'HS1 were not required for the nuclear peripheral targeting of the human beta-globin locus.

Each of two genomic regions located upstream of beta-globin genes is sufficient for nuclear peripheral localization of human beta-globin locus.

To examine whether there are functional sequence elements outside of the D5 region responsible for the nuclear peripheral localization of the human beta-globin BAC, we made a second deletion starting with the HBBD5 BAC to remove the ~80kb region between the LacO repeat and BAC vector backbone (Fig. 3.3 a). As the result, this double deletion HBBD5D7 BAC contains only the BAC vector backbone, the LacO repeat, and an ~5kb human genomic fragment which includes the olfactory receptor genes OR51V1 and OR51B2. None of 5 randomly selected clones carrying the HBBD5D7 BAC transgene showed more than 20% peripheral localization (Fig. 3.3 c,f), demonstrating that specific sequences on the HBB BAC located within the D7 deletion region are required for peripheral targeting.

To further confirm the existence of elements for peripheral localization in the D7 region, we made the D4 deletion removing a 56kb sub-region of D7 from the full-length beta-globin BAC (Fig. 3.3 a). The D4 region contains olfactory genes OR51B5, OR51B6 and OR51M1 and the HS-110 site. Only 1 out of 5 randomly selected clones carrying this HBBD4 BAC showed greater than 20% peripheral localization (Fig. 3.3 c,f), demonstrating the existence of functional DNA sequences located in the olfactory receptor region upstream of the beta-globin genes which are responsible for the nuclear peripheral localization of the human beta-globin BAC. We further dissected this region to narrow down the possible region containing

peripheral targeting elements. Previous studies in human lung fibroblasts using the DamID technique have mapped the position of lamin-associated domains (LADs) which are large genomic domains preferentially interacting with nuclear lamina. The high-resolution LAD map indicates a 32kb LAD region at the 3' end of the beta-globin BAC used in this study (Fig. 3.3 a). LADs are often demarcated by CTCF binding sites (Guelen et al., 2008) . CTCF is a protein binding to insulator sequences which has been implicated to have an important role in organizing global chromatin architecture. Available ChIP-on-chip data indicates the presence of multiple CTCF binding sites, some of which coincide with the HS-110 site, in an ~10kb region 5' to the LAD (Fig. 3.3 a). To investigate whether the LAD and/or CTCF enriched boundary region are the determinants for the nuclear peripheral localization of the beta-globin BAC, we deleted either the 32kb LAD region (deletion D8) or the 43kb region including both the LAD and the LAD boundary region (deletion D9) (Fig. 3.3 a).

Both the HBBD8 and HBBD9 BAC transgenes still showed a high percentage of nuclear peripheral localization, suggesting either that the LAD region does not contain a peripheral targeting sequence or that the LAD region contains a peripheral targeting sequence but that at least one additional peripheral targeting sequence is present outside of the LAD region within the D4 deletion (Fig. 3.3 d,f).

To test whether there was a redundant peripheral targeting sequence located outside of the LAD region plus a targeting sequence inside of the LAD region, we then made a new deletion D10 which removed the ~48kb region outside of the LAD region in the D7 deleted region (Fig. 3.3 a). Similar to deletion D8, D10 also did not change the ability of the BAC transgenes to target to nuclear periphery (Fig. 3.3 e,f). Therefore, our results suggested there aretwo functionally redundant regions within

the BAC responsible for the nuclear peripheral localization.

To determine whether these two genomic regions are sufficient by themselves to confer the nuclear peripheral localization of the locus or instead need to function together with other regulatory elements of the beta-globin locus, we deleted the D8 or D10 regions from the HBBD5 BAC (Fig. 3.3 a). There were no significant differences in nuclear peripheral localization between the double deletions HBBD5D8 or HBBD5D10 BACs versus the full-length beta-globin BAC (Fig. 3.3 e,f), suggesting that both the 32kb D8 region and 48kb D10 region contain functional sequence elements sufficient to confer nuclear peripheral localization.

Co-transfection of beta-globin and DHFR BAC results in nuclear interior localization of BAC transgenes.

Our data have demonstrated that the beta-globin BAC transgenes, when randomly integrated into the genome, exhibited significant nuclear peripheral localization in most of the examined cell clones. However, it is still unclear whether the peripheral localization signals present in the beta-globin locus have a dominant effect over nearby transcriptionally active genomic loci, which are usually found in the interior of the nucleus.

To directly test this, we co-transfected the beta-globin BAC and a BAC containing the mouse housekeeping gene DHFR locus into NIH/3T3 cells. The 175kb DHFR BAC contains the full-length DHFR gene and most of the Msh3 gene (Fig. 3.4 a). Both the DHFR and Msh3 genes are expressed in proliferating cells and are transcribed from the same divergent promoter (Schilling and Farnham, 1995). Previously we have shown the DHFR BAC transgenes were preferentially positioned to the nuclear interior in mouse ES and NIH/3T3 cells (Sinclair et al., 2010). The

multi-copy DHFR BAC transgenes support copy-number dependent and position independent expression of the DHFR gene as well as mini-gene reporters embedded in the BAC (Bian and Belmont, 2010).

The DHFR BAC was tagged with a 256mer lac operator repeat and a zeocin selectable marker using the Tn5 transposon system as described previously (Bian and Belmont, 2010). The linearized DHFR and beta-globin BACs were mixed together and transfected into NIH/3T3 cells. Stable transformants containing both BACs were selected with the combination of G418 and zeocin. Stable clones containing a single, visible insertion, suggesting that the DHFR and beta-globin BACs were integrated into the same chromosome site, were selected for further analysis. None of the five examined clones showed substantial peripheral localization, with the highest peripheral localization percentage being 16% and the lowest being 0% (Fig. 3.4 b). This is significantly different from BAC transgenes arrays containing only the beta-globin BAC (p-value=0.006). Therefore, when present within the same multi-copy BAC transgene array, the housekeeping DHFR locus prevented the sequence elements in the beta-globin BAC from targeting the BAC transgenes to nuclear periphery.

The interior localization of beta-globin BAC transgenes upon sequence modifications is accompanied by elevated chromocenter association.

In addition to the nuclear periphery, another important nuclear compartment related to gene silencing is centromeric heterochromatin. The association of inactive genes with the mouse centromeric heterochromatin has been observed in a number of cell types (Brown et al., 1999; Brown et al., 1997; Francastel et al., 2001). A previous study has suggested that the inactive beta-globin locus preferentially

localized to both the centromeres and the nuclear periphery in murine erythroleukemia (MEL) cells(Francastel et al., 2001). However, the mechanisms underlying the differential distribution of inactive genes between different heterochromatin compartments are unknown.

During our dissection of the human beta-globin locus, we also examined effects of different sequence deletions on the association of the BAC transgenes with the chromocenters. The full-length beta-globin BAC transgenes did not exhibit significant association with the chromocenters, with only one out of five clones showing association in more than 30% of the nuclei (Fig. 3.5 a). Interestingly, a general anti-correlation between chromocenter association and the peripheral localization of the BAC transgenes was observed, with clones showing a higher degree of nuclear peripheral localization also showing lower degree of chromocenter association (Fig. 3.5 c). A similar pattern was also observed in clones containing HBBD5 BAC transgenes, which were also preferentially localized at nuclear periphery (Fig. 3.5 a).

Previously we have shown that the HBBD4 BAC, in which a region upstream of the beta-globin gene cluster was deleted, lost the ability to autonomously target to nuclear periphery. Surprisingly, we found that these HBBD4 BAC transgenes showed an accompanying increase in chromocenter association compared with the full-length beta-globin BAC (p-value= 0.01) (Fig. 3.5 a,b). The anti-correlation between chromocenter association and nuclear peripheral localization was also observed in these clones (Fig. 3.5 c). Therefore, our data indicated that the association of genomic regions with centromeric heterochromatin is governed by signals different to the ones targeting genes to the nuclear periphery. Moreover, changes of sequence composition at the beta-globin locus can alter its distribution

between these two heterochromatin compartments.

We also examined the clones carrying the HBBD5D7 BAC transgenes, in which the beta-globin genes and most of the olfactory receptor genes were removed. The HBBD5D7 BAC transgenes retained the ability to target to chromocenters to some degree, with two out of five clones showing higher than 40% chromocenter association (Fig. 3.5 a). In fact, we found that a plasmid containing the 256 mer lac operator repeat alone could give rise to about 30% chromocenter association in mixed populations of stably transfected cells (data not shown), suggesting that the lac operator direct repeat, and/or vector backbone sequence we used as the tag for microscopic visualization, may partially contribute to the chromocenter association of the BAC transgenes. Notably, we observed no obvious correlation between chromocenter association and nuclear periphery localization in the HBBD5D7 clones, with three out of five clones showing lower than 30% association with both heterochromatic compartments (Fig. 3.5 c). In this way the chromatin regions consisting of the HBBD5D7 BAC behaved differently from the other BAC transgenes.

We have shown that co-transfection with the transcriptionally active DHFR locus abolished the nuclear periphery targeting of beta-globin locus. Does this also change the nuclear localization pattern of the BAC transgenes in terms of chromocenter association? To our surprise, the clones carrying both the DHFR and beta-globin BAC transgenes exhibited a high degree of chromocenter association, similar to that observed in HBBD4 clones and significantly higher than the full-length beta-globin BAC clones (p-value=0.006) (Fig. 3.5 d,e). This suggests that the ability of truncated beta-globin BAC transgenes to target to centromeric heterochromatin is insensitive to the signals from surrounding active genes, which were sufficient to

change the radial positioning pattern of the transgenes.

It is noteworthy that in the clones analyzed in this study, the total percentage of the BAC transgenes associated with nuclear periphery or chromocenter was usually 60-90% but never 100%. This may be partially attributed to the dynamic nature of the chromatin in interphase nucleus. On the other hand, we have evidence showing that most of the BAC transgene arrays co-localized to heterochromatin foci highlighted by H3K9me3 immunostaining (Fig. 3.6). In some cases, the BAC transgenes showing no association with the nuclear periphery and chromocenters co-localized with small foci of H3K9me3 staining in the nuclear interior (Fig. 3.6 c), suggesting that the inactive loci may find alternative heterochromatin compartments other than the predominant nuclear periphery and centromeres.

Tethering of the acidic transcriptional activator VP16 relocates the beta-globin BAC transgenes from the nuclear periphery to the nuclear interior

A number of developmental regulated genomic loci, including beta-globin locus, have been shown to reposition from nuclear periphery to interior upon gene activation during development. This suggests that the nuclear peripheral localization of genes can be reversed and this reversion may be related to transcription activity. Previously in our lab, we have described an inducible movement of an engineered chromosome site from the nuclear periphery to interior upon targeting the transcription activator VP16 to a peripherally located plasmid gene array (Chuang et al., 2006; Tumbar and Belmont, 2001). We therefore set out to test whether this repositioning can be recapitulated at peripherally located beta-globin BAC transgene arrays.

A fusion protein consisting of mRFP, Lac repressor and the acidic activation

domain of VP16 was constructed and transfected into several NIH/3T3 clones carrying full length HBB or HBBD5 BAC transgenes. Two days after transient transfection, the intranuclear positioning of the BAC transgenes in RFP expressing cells was examined. Due to the competition of the overexpressed mRFP-LacI-VP16, the LacI-GFP spots appeared dimmer and became invisible in some clones. Nevertheless, a more interior localization of the BAC transgenes was observed in multiple clones. For example, in clone HBBD5 C43, the percentage of BAC transgenes at nuclear periphery decreased from 52% to 14% after VP16 tethering (Fig. 3.7 b,d).

We note that the Lac operator repeat, which is the tethering site for VP16, is more than 3kb away from the nearest olfactory receptor promoter. Therefore, it is unclear whether transcription is required for such a redistribution of the BAC transgenes. Previously we have shown a synthetic acidic peptide (DELQPASIDP), which did not activate transcription, could also induce the repositioning of a periphery located plasmid transgene array (Chuang et al., 2006; Tumbar and Belmont, 2001). We then tested whether this peptide caused the redistribution of beta-globin BAC transgenes. Indeed, the transfection of a mCherry-LacI- DELQPASIDP construct into clone HBBD5C43 resulted in a redistribution of BAC transgenes toward nuclear interior similarly to VP16 (Fig. 3.7 c,d). Our data thus suggests the recruitment of certain transcription activator is sufficient to change the nuclear peripheral localization of genes. We reason the beta-globin BAC transgene arrays provide a valuable system for studying the intranuclear movement of genomic loci in mammalian cells.

DISCUSSION

The three-dimensional organization of eukaryotic genome within the nucleus has

been implied to have an important regulatory role on transcription by a large collection of studies. However, the molecular mechanisms controlling the spatial organization of genome are largely unknown. Also whether the intranuclear positioning of genomic loci is coded by specific sequences has not been directly tested. In this study, we demonstrate we can recapitulate the nuclear peripheral localization of human beta-globin locus using engineered chromosome regions consisting of multi-copy BAC transgenes. By combining this BAC transgene system with BAC recombineering techniques we identify specific genomic regions responsible for the nuclear peripheral targeting of the beta-globin locus. This likely opens the door for future mechanistic studies on nuclear peripheral localization.

An effective system for dissecting the intranuclear positioning of specific chromosome regions

Previously, we have shown that the multi-copy BAC transgene arrays recapitulate several features of endogenous chromosomes, including intranuclear localization (Hu et al., 2009a; Sinclair et al., 2010). Here we demonstrate we can further employ this system to dissect the sequence elements responsible for nuclear positioning to the nuclear periphery.

This BAC transgene array system possesses some unique advantages over the traditional, plasmid transgene based system for identifying sequence determinants for intranuclear positioning of genomic loci. It has been well documented that plasmid transgenes are susceptible to chromosome position effect and repeat-induced transgene silencing. Previous electron microscopic studies in our lab revealed that multi-copy plasmid transgene arrays often appeared as chromatin structures more compacted than surrounding endogenous chromatin (Kireev et al., 2008). Such a

tendency of heterochromatinization of plasmid arrays may impair their ability to recapitulate the intranuclear positioning patterns of endogenous genomic loci. In contrast, in previous studies and this study, we have shown several different BAC transgenes localized to the same nuclear compartment as their endogenous counterparts, such as the Hsp70 BACs to the nuclear speckles after heat shock and the beta-globin BACs to the nuclear periphery (Hu et al., 2009a; Sinclair et al., 2010). The large sequence capacity of BACs provides another apparent advantage for sequence dissection, as the intranuclear positioning of genomic loci may be determined by large sequence elements or synergistic activities of multiple sequence motifs. The sizes of these elements may exceed the capacity of typical plasmid constructs, therefore limiting the effectiveness of the plasmid transgene system in dissection of functional sequences determining intranuclear positioning.

We note an alternative approach for dissecting the sequence determinants for intranuclear positioning is to remove specific sequences from endogenous genomic loci and examine the effects. However, this approach requires significantly more work for generating proper cell lines for the analysis. More importantly, we reason the BAC transgene arrays, as a reductionist system, reduces the complexity for the sequence dissection. The interphase chromatin is highly compacted, with several hundred kilobases of DNA detected by FISH usually appearing as a spot under the diffraction limit of light microscopy (Lawrence et al., 1990; Yokota et al., 1995). The large-scale chromatin structure is also organized to accommodate a large number of interactions between distal sequences, which are poorly characterized. Therefore, the intranuclear localization of an endogenous genomic locus could be influenced or constrained by the features of other sequences in its spatial proximity. Given that there may be redundancy of signals for determining nuclear positioning, the

likelihood of seeing a phenotype after removing several or tens of kilobases of specific sequences from an endogenous locus may be low, and the interpretation of phenotypes will be difficult. By using the multiple-copy BAC transgene tandem arrays, which form engineered chromosome regions with size from several hundred kilobases to several megabases, the influence of neighboring genomic regions and other long-range chromatin interactions is diluted. Moreover, the intranuclear localization of these BAC transgene arrays is more likely determined by the sequences contained within the BAC. As a result, the chances of identifying smaller sequence determinants for intranuclear positioning by extensive sequence dissection will be significantly enhanced.

The nuclear peripheral localization of human beta-globin locus is determined by its genomic neighborhood.

The relationship between the transcription activation of beta-globin locus and its intranuclear positioning pattern has been extensively investigated. However, little is known about the mechanisms underlying the remarkable nuclear peripheral localization of inactive beta-globin locus. In this study, by dissecting a ~200kb BAC containing the beta-globin locus, we demonstrated that the inactive beta-globin genes themselves, as well as the known regulatory elements for beta-globin expression, are not required for the nuclear peripheral localization of this locus in fibroblasts. Further dissection leads to the identification of two genomic regions responsible for localization of the multi-copy BAC transgenes to nuclear periphery. Both of these regions reside in the olfactory receptor gene clusters flanking the beta-globin locus. Therefore, our data suggests the preferential nuclear peripheral localization of the beta-globin locus is dictated by its genomic neighborhood.

Although we have not yet narrowed down the signals for nuclear periphery targeting to the olfactory receptor genes, it is noteworthy that the genomic structure of the beta-globin region, with the beta-globin genes and LCR embedded in the middle of a large cluster of olfactory receptor genes, is highly conserved in mammals. Olfactory receptor genes were also found in the neighborhood of beta-globin genes in chicken, but not in lower vertebrates. It is possible that this particular genomic arrangement reflects an evolutionary mechanism to establish the nuclear peripheral localization of inactive beta-globin locus, therefore reinforce its correct expression pattern.

Competition between nuclear peripheral and nuclear interior positioning signals

Active and inactive genomic loci usually display different intranuclear positioning pattern, with active loci preferentially located in the interior and inactive loci located at the periphery of the nucleus. However, whether the interior or the peripheral localization has a dominant effect over the other is poorly understood, due to the lack of the knowledge for their respective molecular mechanisms. A recent study in *C.elegans* examined the positioning of small transgene arrays containing two differentially expressed tissue specific promoters (Meister et al., 2010). The results suggested the active promoter dictated the intranuclear positioning over the inactive promoter, resulting in internalization of the transgene arrays in specific tissues. Here we reported that the DHFR housekeeping gene loci has a similar effect on the positioning of inactive loci in mammalian cells.

In our initial studies, we observed that the transgene arrays containing tandem repeats of the beta-globin BAC targeted to nuclear periphery independent of their chromosomal integration sites, at least in the clones examined. However, when we

co-transfected the beta-globin BAC with a BAC containing housekeeping gene DHFR locus, the transgene arrays containing both BACs were predominantly positioned in the interior of the nucleus. Although not directly mapped, we assume the two types of BACs were arranged in an interleaved fashion within each engineered chromosome region. This suggests the 200kb genomic locus carried by the BAC was not sufficient to establish the nuclear peripheral positioning when flanked by highly transcribed housekeeping locus.

The internal positioning of the transgenes may be due to a mechanism actively targeting the housekeeping gene loci to nuclear subcompartments in the interior. Alternatively, the surrounding housekeeping gene loci may change the properties of the periphery targeting elements within the inactive loci and block their functions. It is also possible that the localization of a locus is determined by the integration of different positioning signals over a large chromatin domain. Nonetheless, our data suggest the nuclear peripheral localization of inactive loci can be profoundly influenced by adjacent active loci. This is also compatible with the observation that most of the identified LADs are large genomic domains with a size larger than 1Mb and depleted of highly transcribed genes.

Localization of beta-globin BAC transgenes to the nuclear periphery or chromocenter is determined by different mechanisms

In addition to the nuclear periphery, centromeres exist as another important docking site for heterochromatin. Multiple studies have reported the association of inactive genomic loci with centromeric heterochromatin. However, whether there is preferential association of inactive loci with specific heterochromatic nuclear compartments has not been investigated in detail.

In this study, we showed the 200kb BAC containing beta-globin genes and neighboring olfactory receptor genes, when introduced into cells as tandem transgene arrays, preferentially targeted to nuclear periphery. The removal of sequence elements required for peripheral targeting resulted in the interior localization of the BAC transgenes arrays. Interestingly, these modified BAC transgenes displayed a high degree of association with chromocenters. Similar preferential association with chromocenters was also observed in the transgene arrays containing both the beta-globin BAC and DHFR BAC. Our data therefore suggests the association of the inactive beta-globin locus with the nuclear periphery versus chromocenters is regulated by different mechanisms. Our results indicate that the inactive loci may be organized to preferentially locate to certain accumulation sites for heterochromatin instead of a random distribution. And the distribution between different heterochromatic compartments may be changed by modifying the sequence composition.

We speculate that the association of inactive loci with the nuclear periphery may require the functions of some specific sequence elements, while the association with chromocenters may be a general feature for heterochromatin. The 200kb genomic region containing beta-globin locus, when flanked by active housekeeping locus, was not able to confer the nuclear periphery targeting but may still retained certain features of heterochromatin sufficient for the chromocenter targeting. It is conceivable that the association of inactive loci with the nuclear periphery is a more "stringent" nuclear organization as the active transcription machinery is largely excluded from nuclear periphery. On the other hand, the interior localized chromatin may have the flexibility to associate with both active and inactive nuclear subcompartments.

Future directions

In this study, we report a system for identification of sequence determinants for the nuclear peripheral localization of genomic loci. Sequence dissection on a ~200kb region reveals two redundant regions sufficient to confer the nuclear peripheral targeting of human beta-globin locus. Our future studies will aim to further narrow down these genomic regions to identify the minimal motifs controlling nuclear peripheral localization of chromatin. The follow-up functional interrogation of those motifs via bioinformatic and biochemical studies will serve as a necessary step toward the eventual understanding of the molecular mechanisms underlying the spatial organization of genome within nucleus. Finally, the periphery-targeting sequences identified in this study also provide a means to study the functional consequences of tethering genes to nuclear periphery in a more natural context.

MATERIALS AND METHODS

BAC Recombineering

Human BAC CTD-2643I7 was obtained from Invitrogen. A 256mer lac operator direct repeat and kanamycin/neomycin selectable marker was inserted into the BAC using a Tn5 transposon as described previously (Bian and Belmont, 2010; Levi et al., 2005). BAC clone 2643I7-C4, which contains the lac operator repeat between olfactory receptor genes OR51B2 and OR51B5 was selected for subsequent engineering.

A λ red-mediated BAC recombineering system and GalK-based dual selection scheme was used to delete specific regions from the beta-globin BAC. The purified 2643I7-C4 DNA was transformed into an E.coli strain SW102, in which the

expression of λ red recombination machinery can be induced by shifting temperature from 32˚C to 42˚C. The DNA fragments for recombination were prepared by PCR from plasmid template pGalK using 60bp primer pairs (Table 3.1). Each primer consists of a 43bp homology sequence flanking the region to be deleted from the BAC and a 17bp sequence for amplifying the GalK selectable marker (5'-cgacggccagtgaattg-3' for the forward primer and 5'-tgcttccggctcgtatg-3' for the reverse primer). The recombination was performed according to standard protocols(Warming et al., 2005). Following the recombination, the recombinants were selected at 32˚C on minimal medium in which galactose was supplied as the only carbon source. Following the galactose selection, the recombinants were screened by PCR using 20bp primers outside of the target regions (CHECK-F and –R primers, Table 3.1).

To make subsequent deletions from a previous deletion, the GalK selectable markers were removed from the BAC. The DNA fragments for the removal of GalK were generated by PCR using 60bp primers partially complementary to each other (GkRm-F and –R, Table 3.1). Each of the 60bp primers consists of a 52bp homology region flanking the GalK marker and an 8 bp sequence complementary to the last 8 bp of the other homology region. The 100bp PCR products were introduced into SW102 cells to replace the GalK. Following recombination, negative selection was performed using minimal medium containing 2-deoxy-glactose (DOG) and the recombinants were again screened using the CHECK-F and –R primers. After the removal of GalK was confirmed, another round of recombination was performed as described above.

After each recombineering step, the integrity of the BAC constructs and the length of lac operator repeats was checked by restriction fingerprinting. Small amount of

BAC DNA was purified using the buffers from Qiagen Large Construct Kit (Qiagen). The DNA was then digested with AvaI and HindIII and ran on a 0.5% agarose gel. The restriction digestion patterns were compared with the predicted digestion patterns generated by program Gene Construction Kit (Textco BioSoftware). Only BAC clones showing correct digestion pattern and also containing full-length 256mer lac operator repeats were chosen for further modifications.

DHFR BAC 057L22 (CITB mouse library) is a gift from Edith Heard (Curie Institute). The insertion of lac operator repeat and a CMV-RFP-SV40-Zeocin reporter was performed using a Tn5 transposon system as described before(Bian and Belmont, 2010).

Cell culture and establishment of BAC cell lines

NIH 3T3 cells (CRL-1658, ATCC) were grown in Dulbecco modified Eagle's medium (Invitrogen) plus 10% Bovine Growth Serum (HyClone) at 37 ^{o}C with 5% CO_2. To create a cell line stably expressing EGFP-dimer lac repressor fusion protein, the plasmid p3'SS-d-LacI (Robinett et al., 1996) was linearized by XmnI and transfected into 3T3 cells using lipofectamin 2000 (invitrogen) according to manufacturer's directions. 24 hours after transfection, the cells were plated onto 94mm hybridoma dishes (Greiner Bio-One) and then incubated overnight before selection with 200 μ g/ml Hygromycin B. Two weeks after transfection, the subclones were transferred from the dishes to 96 well plates and propagated. A subclone 3T3_LG_C29 expressing EGFP-dimer lac repressor at desirable level was chosen for the subsequent BAC transfection.

The full-length and modified 2643I7 BACs were purified using the Qiagen large-construct kit. Purified BAC DNA was linearized with PI-SceI,

ethanol-precipitated and transfected into 3T3_LG_C29 cells. The stable transfectants were selected using 800 µ g/ml G418 and 200 µ g/ml Hygromycin B for two weeks and subclones were picked. To examine the conformation and localization of BAC transgenes in individual subclones, the cells were plated on coverslips, fixed with 3 % paraformaldehyde (Polysciences) in calcium, magnesium free PBS buffer for 10 min and stained with 0.2 µ g/ml DAPI (Sigma-Aldrich). Deconvolution wide field light microscopy was carried out as described previously.

For co-transfection of beta-globin BAC and DHFR BAC, the beta-globin BAC and DHFR BAC was linearized with PI-SceI and BsiWI, respectively, and transfected into 3T3_LG_C29 as described above. Following transfection, the stable clones were selected using 800 µ g/ml G418, 200 µ g/ml Hygromycin B and 75 µ g/ml Zeocin.

BJ-hTERT cells were obtained as a gift from Dr. Peter Adams (Beatson Institute/ Glasgow University) and grown in DMEM supplemented with 10% FBS, non essential amino acids, vitamins and 0.54 µ g/ml Puromycin.

Immunofluorescence

NIH/3T3 cells were fixed with 3% paraformaldehyde in CMF-PBS for 10min at room temperature and washed three times with CMF-PBS. Following fixation the cells were permeabilized with 0.5% Triton X-100 in CMF-PBS for 5min at room temperature and blocked with 5% normal goat serum for 30min. Primary antibody was applied for 4 hours at room temperature at the following dilutions in CMF-PBS containing 0.1% Triton X-100: mouse monoclonal antibody RL1 against nuclear pore O-linked glycoprotein, 1:500 (ABR-Affinity BioReagents); anti–lamin A antibody, 1:500 (obtained from B. Goldman, Northwestern University, Evanston, IL) and rabbit

anti-H3K9 trimethylation (Upstate), 1:1000. Cells were washed with CMF-PBS for 3 times and stained with secondary goat anti-rabbit or anti-mouse antibody conjugated with Texas red (Jackson ImmunoResearch Laboratories) at 1:500 dilution overnight at 4°C.

Immunostaining followed by DNA FISH

BJ-hTERT cells were fixed with 3% paraformaldehyde for 10min at room temperature and permeabilized with 0.5% Triton X-100 in PBS for 5 min on ice. 5% normal goat serum diluted in PBS was used for blocking. Rabbit anti-lamin-A antibody (obtained from Bob Goldman, Northwestern University, IL) was used at 1:1000 dilution and secondary goat anti-rabbit antibody conjugated to Texas-Red (Jackson Labs) were used at a 1:500 dilution. Following immunostaining, cells were postfixed in 3% paraformaldehyde in PBS for 10 min at room temperature and permeabilized in 0.1M HCl / 0.7% Triton X-100 in 2xSSC for 10 min on ice. DNA FISH was then performed as described previously (Hu et al., 2009b). The probes for detecting *α-globin* and *β-globin* loci were prepared from BACs RP11-344L6 and CTD-2643I7 using nick translation.

Microscopy and image analysis

A personal deconvolution microscope system (DeltaVision; Applied Precision) was used with a 60× NA 1.4 lens for data collection. XY and Z optical displacement between different filter sets was determined experimentally using fluorescent microspheres (Tetraspeck; Invitrogen). Deconvolution on the optical sections was performed using an enhanced ratio iterative-constrained algorithm(Agard et al., 1989).

The distances between LacI-GFP spots and nuclear edge defined by DAPI staining

were measured using ImageJ (National Institutes of Health) open-source software. For defining nucleus edge, the optical section of DAPI staining in which the LacI-GFP is in focus was selected. Segmentation of nucleus was then performed using the OtsuThresholding 16Bit and k-mean clustering plugins with the following parameters: number=2, cluster=0.00010000 and randomization=48. For BAC transgenes consisting of multiple LacI-GFP spots, measurements were taken from the EGFP-LacI signal closest to the nuclear edge.

ACKNOWLEDGEMENTS

This work was supported by grant number GM58460 and GM42516 from the National Institute of General Medical Sciences awarded to ASB. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

We thank Neal Copeland (National Cancer Institute) for providing the reagents for BAC recombineering.

FIGURES AND TABLE

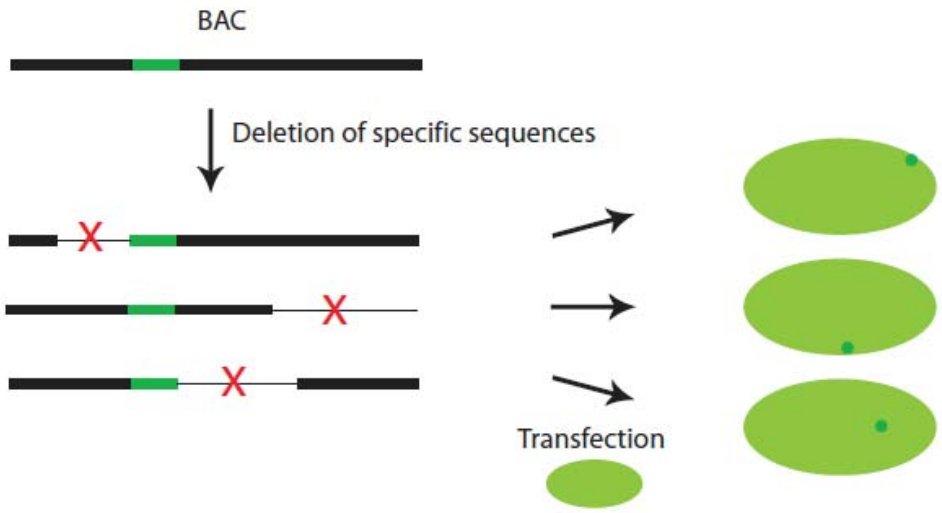


Fig. 3.1 Experimental system for dissecting sequence elements required for nuclear peripheral localization. BAC recombineering is used to remove specific regions from a BAC containing lac operator repeat. Modified BACs are introduced to mouse fibroblast NIH/3T3 cells stably expressing LacI-GFP and the intranuclear positioning of the BAC transgenes can be assessed.

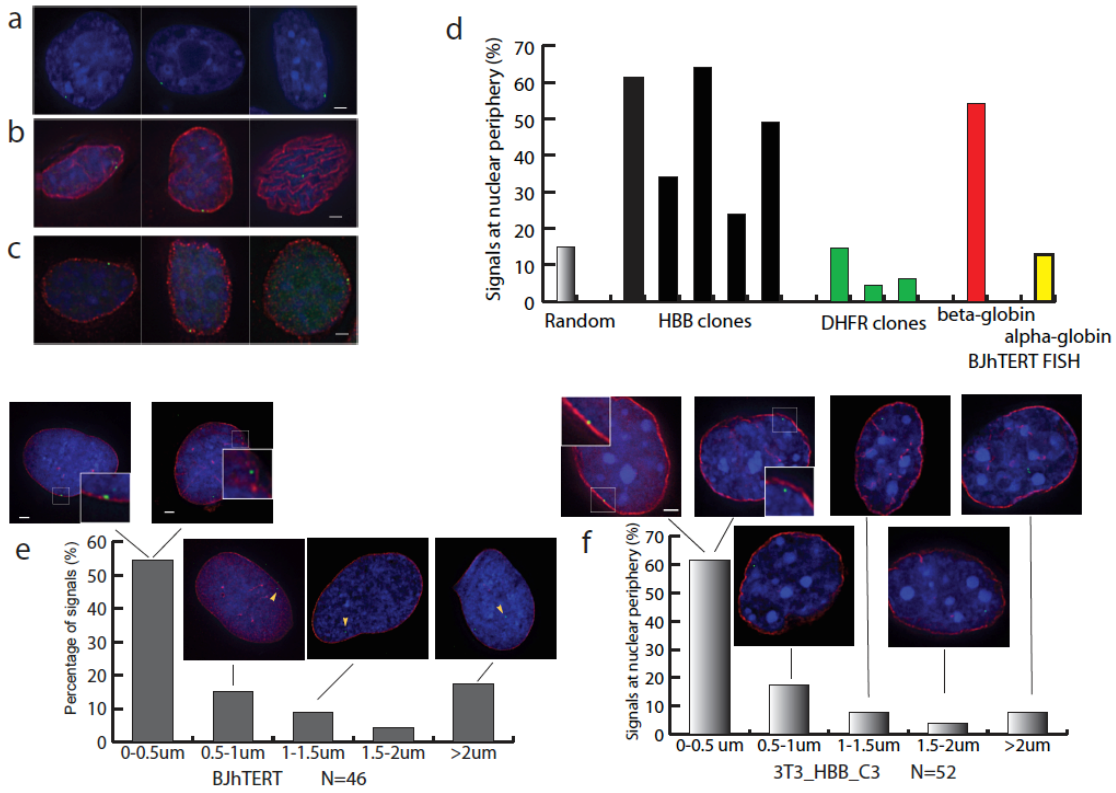


Fig. 3.2 The ~200kb human beta-globin BAC autonomously targets to nuclear periphery in mouse fibroblasts. (a-c) Representative examples from NIH/3T3 clone HBB_C3 showing that the BAC transgenes, appearing as LacI-GFP spots (green), target to nuclear periphery defined by DAPI counterstaining (a, blue), Lamin A staining (b, red) or nuclear pore staining (c, red). (d) The beta-globin BAC transgenes in five independent NIH/3T3 clones show significantly higher nuclear peripheral localization compared with either the DHFR BAC transgenes in NIH/3T3 clones or the endogenous alpha-globin locus in human fibroblast BJ_hTERT cells. A transgene is defined as "peripheral" when the distance between the transgene and nuclear edge defined by DAPI staining is shorter than 0.5μm. At least 50 cells from each NIH/3T3 clone and at least 45 BJ_hTERT cells were analyzed. (e,f) Detailed analysis suggests the beta-globin BAC transgenes in clone HBB_C3 recapitulate the radial nuclear localization pattern of endogenous human beta-globin locus in BJ_hTERT cells. The distance between the BAC transgene in HBB_C3 highlighted by LacIGFP binding (e, green) or the endogenous beta-globin locus labeled by FISH signal (f, green) and the Lamin A signal (e and f, red) was measured and summarized as histograms. All scale bars= 2μm.

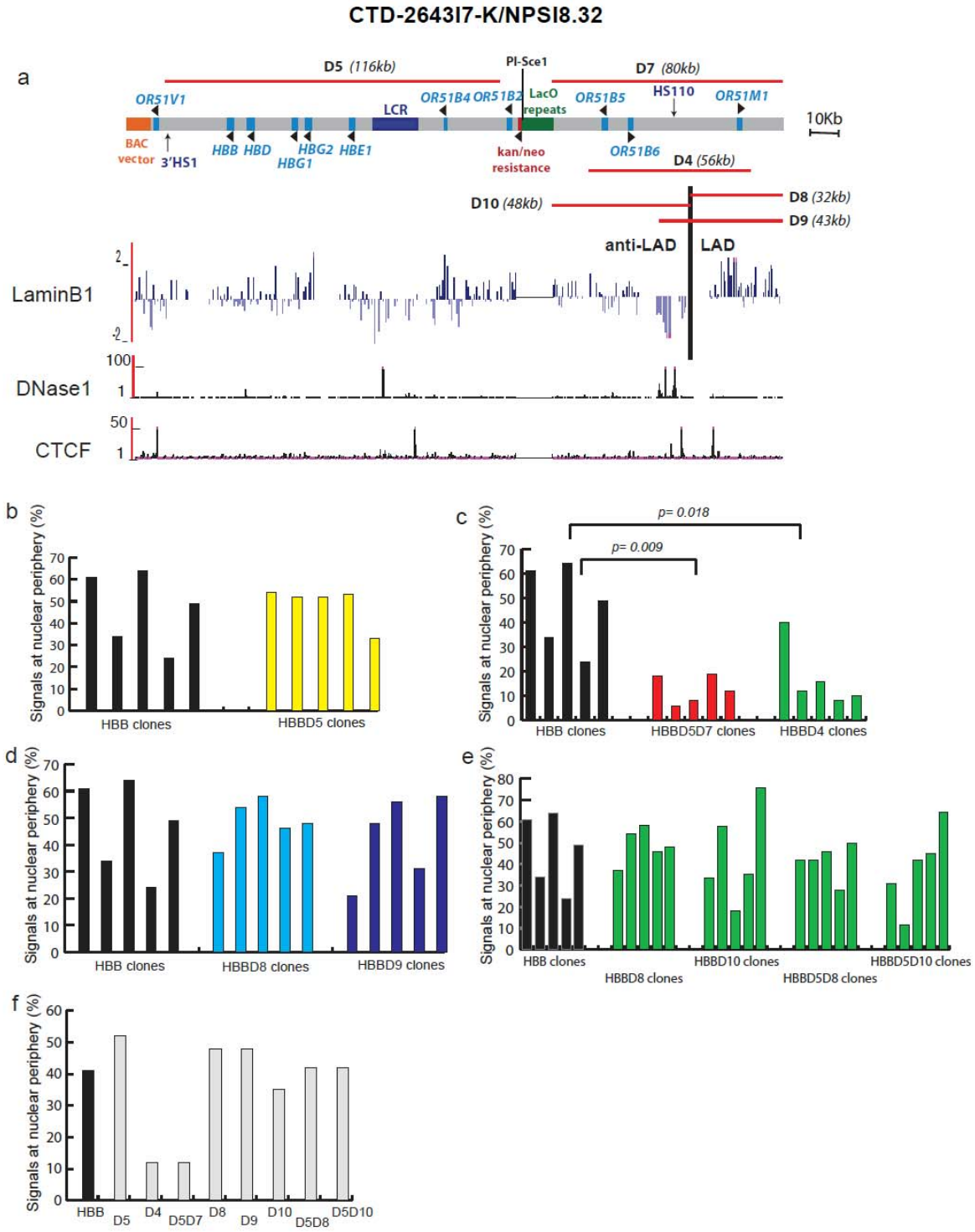


Fig. 3.3 Sequence dissection reveals two redundant regions in the ~200kb beta-globin BAC responsible for the nuclear peripheral localization. (a) Map of the 200kb BAC containing beta-globin locus. Red lines indicate the regions to be deleted. The map of the BAC is aligned with the Lamin B1 DamID map in human

Fig. 3.3 (con't)

fibroblast Tig3 cells, the DNaseI hypersensitivity map in normal human lung fibroblasts (NHLF) and the CTCF binding map in NHLF, all available in UCSC Genome Browser. (b) The beta-globin genes, LCR, UHR and 3'HS1 are not required for the peripheral localization of the locus. Histogram shows the percentages of BAC transgenes at nuclear periphery in five NIH/3T3 clones carrying full-length beta-globin BAC (HBB) and five clones carrying HBBD5 BAC. (c) There are functional elements required for the nuclear peripheral localization of the 200kb region, as HBBD5D7 and HBBD4 BAC lose the ability to target to nuclear periphery. (d) Previously defined LAD /LAD Boundary is not the only region responsible for peripheral targeting, as HBBD8 and HBBD9 BACs show nuclear periphery targeting similarly to HBB BAC. P-value is from two-tailed, two sample unequal variance t-test. (e) Each of two redundant regions sufficient to target the BAC transgenes to nuclear periphery, as HBBD8, HBBD10, HBBD5D8 and HBBD5D10 BACs all display similar level of nuclear peripheral localization. (f) Summary of the sequence dissection. The median level of peripheral targeting of the five cell clones for each BAC is shown. For each NIH/3T3 cell clone, the distance between BAC transgenes and nuclear periphery was measured in at least 50 cells.

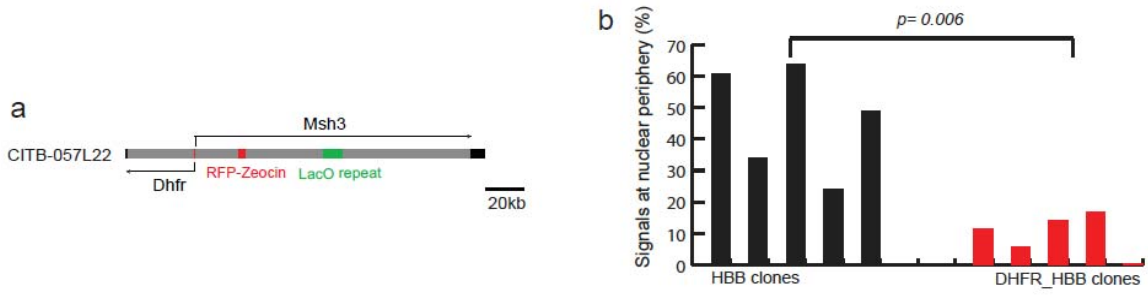


Fig. 3.4 Nuclear interior localization of BAC transgenes consisting of both beta-globin and DHFR BACs. (a) Map of the ~180kb DHFR BAC containing a 256mer lac operator repeat and a Zeocin selectable marker. (b) Transgenes consisting of both DHFR and beta-globin BACs in five independent NIH/3T3 clones show significantly lower percentage of nuclear peripheral localization compared with the beta-globin BAC alone.

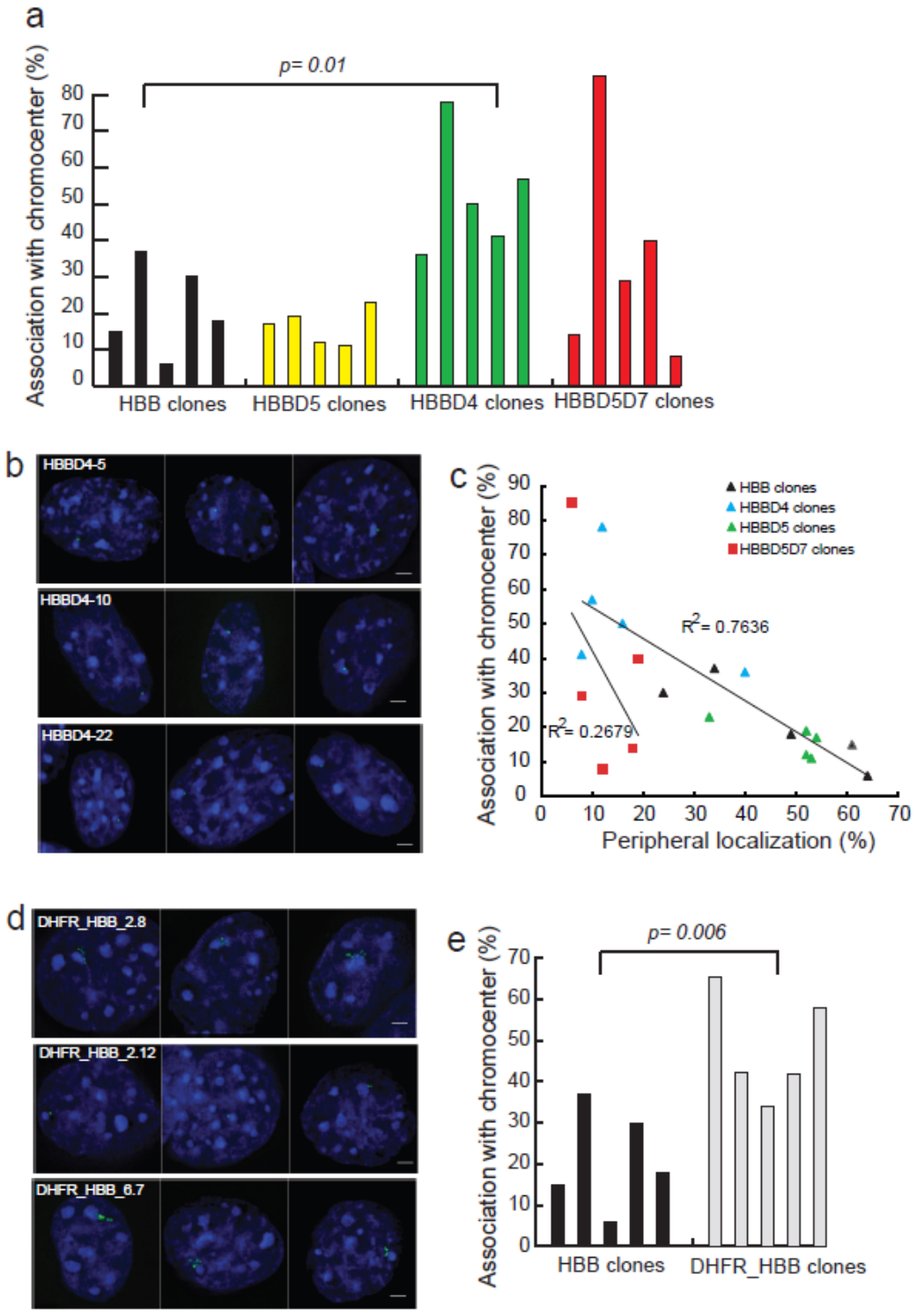


Fig. 3.5 Association of beta-globin BAC transgenes with chromocenters. (a) HBBD4 and HBBD5D7 BAC transgenes exhibit significantly higher level of chromocenter association compared with HBB and HBBD5 BAC transgenes. (b) Representative cells from three independent clones showing the association of HBBD4 transgenes (green) with chromocenters. (c) The percentages of

Fig. 3.5 (con't)
chromocenter association are plotted against the percentages of nuclear peripheral localization. Linear fitting of the HBB clones (black triangles), HBBD4 clones (blue triangles) and HBBD5 clones (green triangles) shows remarkable correlation as indicated by R2 value while the HBBD5D7 clones (red squares) shows no obvious linear correlation. (e) Transgenes consisting of both DHFR and HBB BACs (green) show remarkable chromocenter association regardless of the size of transgenes. (f) The percentages of chromocenter association for DHFR_HBB transgenes are significantly higher than HBB transgenes. P-values are calculated from two-tailed, two sample unequal variance t-test. DNA is stained with DAPI (blue) in all images. All scale bars= 2μm.

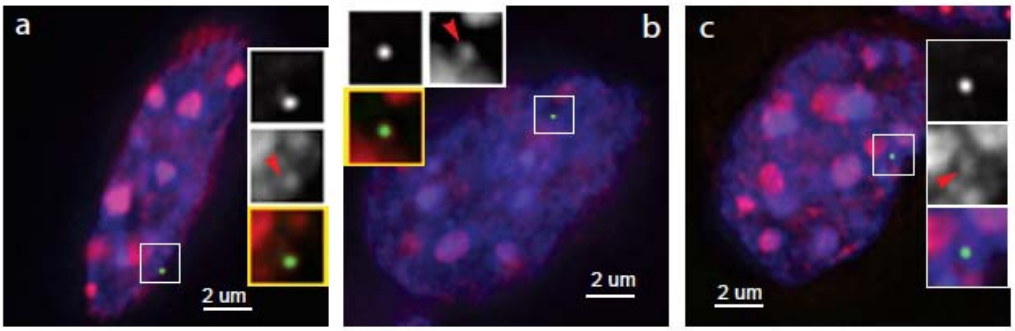


Fig. 3.6 The association of beta-globin BAC transgenes with heterochromatin compartments enriched in H3K9me3. Examples from clone HBB_C3 showing the association of the BAC transgenes (green) with H3K9me3 stained foci (red) regardless the intranuclear positioning of the BAC transgenes. (a) BAC transgene at nuclear periphery. (b) BAC transgene adjacent to a chromocenter. (c) BAC transgene is located in interior of nucleus and not associated with chromocenter. DNA is stained with DAPI and scale bars=2μm.

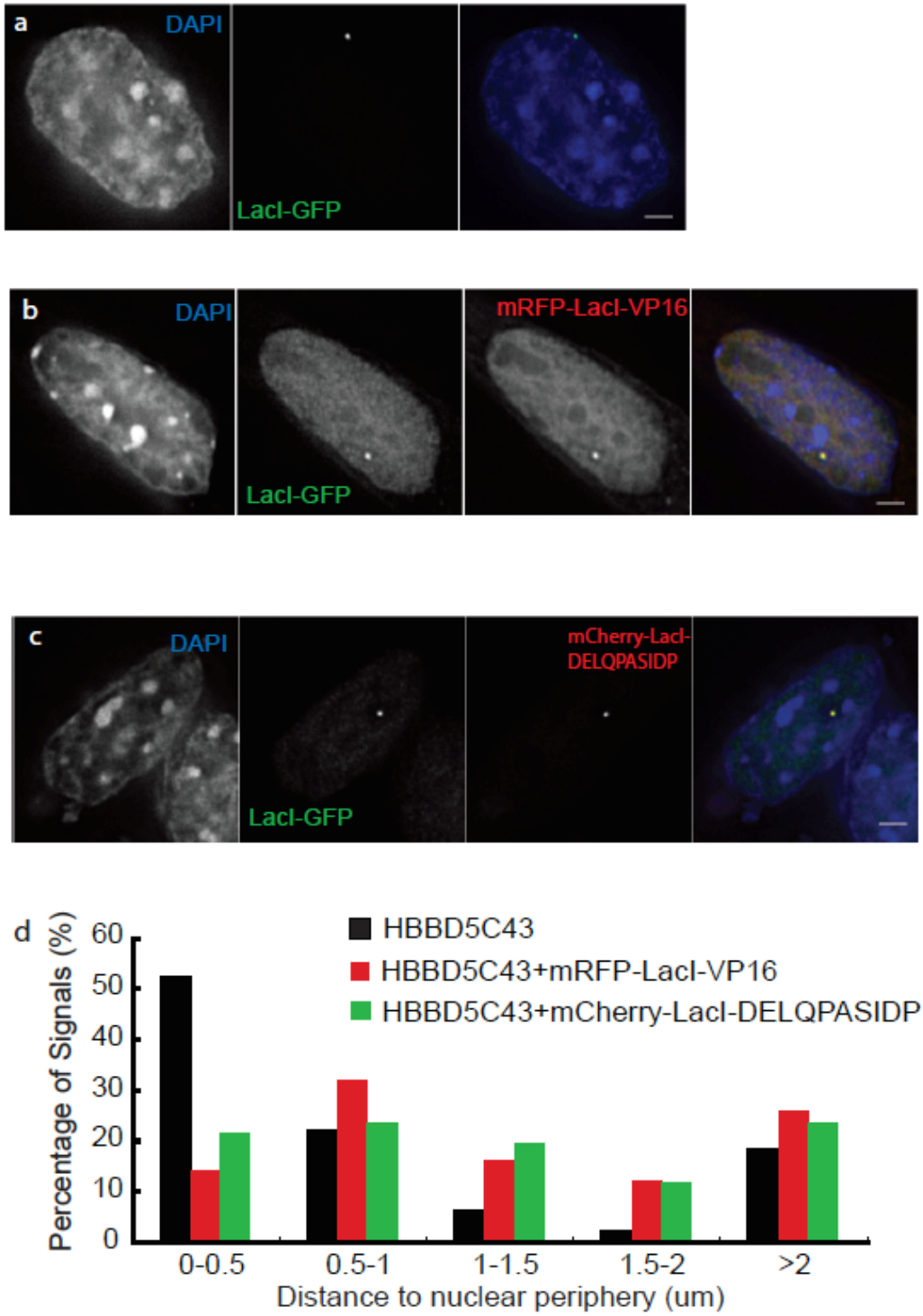


Fig. 3.7 Peripheral localization of BAC transgene arrays can be reversed by recruitment of transcriptional activator. (a) A representative cell of clone 3T3_HBBD5_C43. The BAC transgene (green) shows nuclear peripheral localization. After transient transfection with mRFP-LacI-VP16 (b) or mCherry-LacI-DELQPASIDP (c), the transcriptional activator VP16 or the short peptide (red) is targeted to the BAC transgene. This resulted in the relocation of the

Fig. 3.7 (con't)
BAC transgenes to nuclear interior. DNA is stained with DAPI (blue). Scale bar = 2μm. The distances between BAC transgenes and nuclear periphery defined by DAPI staining are measured in at least 50 cells before or after transfection. The result is summarized in (d).

Table 3.1 Primers for BAC recombineering

Primer Name	Primer Sequence
HBBD4-F	5'-gaaatgtcttcttcagaaaatttgtatacggcaaagatagagacgacggccagtgaattg-3'
HBBD4-R	5'-gtttgtgtgtttcttttgtaaatatgtctaagttcctcatagatgcttccggctcgtatg-3'
HBBD5-F	5'-ccctgatcccagaaatttatttgatgtgtctctcatagcacaccgacggccagtgaattg-3'
HBBD5-R	5'-ttcttcagactgactacaaaggcttcagactagctacaaagcttgcttccggctcgtatg-3'
HBBD7-F	5'-aattcaagaatgaagccgaggaccctcaaggtgattgttacagcgacggccagtgaattg-3'
HBBD7-R	5'-gcagtcttatttaacgagaaattcagaatcctggaggcaagtctgcttccggctcgtatg-3'
HBBD8-F	Same as HBBD7-F
HBBD8-R	5'-acacatagaccaaattgaacagaataaagagcccagaaataaatgcttccggctcgtatg-3'
HBBD9-F	Same as HBBD7-F
HBBD9-R	5'-agcctagtatccatcagaaaatgtatgcaaatccttataaacctgcttccggctcgtatg-3'
HBBD10-F	5'-tttatttctgggctctttattctgttcaatttggtctatgtgtcgacggccagtgaattg-3'
HBBD10-R	Same as HBBD7-R
D4-Check-F	5'-aaacgagtggtcctgtttgg-3'
D4-Check-R	5'-tgcaaccttgctagcatctg-3'
D5-Check-F	5'-tcttgccccattaccaatgt-3'
D5-Check-R	5'-agctcaggagttggagacca-3'
D7-Check-F	5'-gttttcccagtcacgacgtt-3'
D7-Check-R	5'-gttttcttggttgccacgat-3'
D8-Check-F	Same as D7 check F
D8-Check-R	5'-gcagcacacgttctgatttc-3'
D9-Check-F	Same as D7 check F
D9-Check-R	5'-tttgacatgaggcacaatgc-3'
D10-Check-F	5'-caggaaagggaactccatga-3'
D10-Check-R	D10 CHECK R: same as D7 check R
D5-GkRm-F	5'-atttcttctggaaaatgttccctgatcccagaaatttatttgatgtgtctcttctcccaa-3'
D5-GkRm-R	5'-ataattagctgggcatggcagtgtgctcctgtaatcccagttacttgggagaagagacac-3'
D8-GkRm-F	5'-tatagggcgaattcaagaatgaagccgaggaccctcaaggtgattgttacagccggaagc-3'
D8-GkRm-R	5'-acacatagaccaaattgaacagaataaagagcccagaaataaatgcttccggctgtaaca-3'

REFERENCES

Agard, D.A., Y. Hiraoka, P. Shaw, and J.W. Sedat. 1989. Fluorescence microscopy in three dimensions. *Methods Cell Biol*. 30:353-377.

Akhtar, A., and S.M. Gasser. 2007. The nuclear envelope and transcriptional control. *Nat Rev Genet*. 8:507-517.

Andersen, D.C., and L. Krummen. 2002. Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol*. 13:117-123.

Andrulis, E.D., A.M. Neiman, D.C. Zappulla, and R. Sternglanz. 1998. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*. 394:592-595.

Antoch, M.P., E.J. Song, A.M. Chang, M.H. Vitaterna, Y. Zhao, L.D. Wilsbacher, A.M. Sangoram, D.P. King, L.H. Pinto, and J.S. Takahashi. 1997. Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell*. 89:655-667.

Antoniou, M., L. Harland, T. Mustoe, S. Williams, J. Holdstock, E. Yague, T. Mulcahy, M. Griffiths, S. Edwards, P.A. Ioannou, A. Mountain, and R. Crombie. 2003. Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics*. 82:269-279.

Bednar, J., R.A. Horowitz, S.A. Grigoryev, L.M. Carruthers, J.C. Hansen, A.J. Koster, and C.L. Woodcock. 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci U S A*. 95:14173-14178.

Belmont, A.S., and K. Bruce. 1994. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol*. 127:287-302.

Belmont, A.S., S. Dietzel, A.C. Nye, Y.G. Strukov, and T. Tumbar. 1999. Large-scale chromatin structure and function. *Curr Opin Cell Biol*. 11:307-311.

Belmont, A.S., and A.F. Straight. 1998. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol*. 8:121-124.

Bernardi, R., and P.P. Pandolfi. 2007. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol*. 8:1006-1016.

Bharadwaj, R.R., C.D. Trainor, P. Pasceri, and J. Ellis. 2003. LCR-regulated transgene expression levels depend on the Oct-1 site in the AT-rich region of beta -globin intron-2. *Blood*. 101:1603-1610.

Bian, Q., and A.S. Belmont. 2010. BAC TG-EMBED: one-step method for high-level, copy-number-dependent, position-independent transgene expression. *Nucleic Acids Res*. 38:e127.

Brown, K.E., J. Baxter, D. Graf, M. Merkenschlager, and A.G. Fisher. 1999. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol Cell*. 3:207-217.

Brown, K.E., S.S. Guest, S.T. Smale, K. Hahm, M. Merkenschlager, and A.G. Fisher. 1997. Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*. 91:845-854.

Callan, H.G. 1982. The Croonian Lecture, 1981. Lampbrush chromosomes. *Proc R Soc Lond B Biol Sci*. 214:417-448.

Carpenter, A.E., and A.S. Belmont. 2004. Direct visualization of transcription factor-induced chromatin remodeling and cofactor recruitment in vivo. *Methods Enzymol*. 375:366-381.

Carpenter, A.E., S. Memedula, M.J. Plutz, and A.S. Belmont. 2005. Common effects of acidic activators on large-scale chromatin structure and transcription. *Mol Cell Biol*. 25:958-968.

Carter, D., L. Chakalova, C.S. Osborne, Y.F. Dai, and P. Fraser. 2002. Long-range chromatin regulatory interactions in vivo. *Nat Genet*. 32:623-626.

Chambeyron, S., and W.A. Bickmore. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*. 18:1119-1130.

Cheutin, T., S.A. Gorski, K.M. May, P.B. Singh, and T. Misteli. 2004. In vivo dynamics of Swi6 in yeast: evidence for a stochastic model of heterochromatin. *Mol Cell Biol*. 24:3157-3167.

Cheutin, T., A.J. McNairn, T. Jenuwein, D.M. Gilbert, P.B. Singh, and T. Misteli. 2003. Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science*. 299:721-725.

Chuang, C.H., A.E. Carpenter, B. Fuchsova, T. Johnson, P. de Lanerolle, and A.S. Belmont. 2006. Long-range directional movement of an interphase chromosome site. *Curr Biol*. 16:825-831.

Chung, J.H., A.C. Bell, and G. Felsenfeld. 1997. Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A*. 94:575-580.

Chung, J.H., M. Whiteley, and G. Felsenfeld. 1993. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*. 74:505-514.

Cremer, M., K. Kupper, B. Wagler, L. Wizelman, J. von Hase, Y. Weiland, L. Kreja, J. Diebold, M.R. Speicher, and T. Cremer. 2003. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol*. 162:809-820.

Croft, J.A., J.M. Bridger, S. Boyle, P. Perry, P. Teague, and W.A. Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 145:1119-1131.

Dechat, T., K. Pfleghaar, K. Sengupta, T. Shimi, D.K. Shumaker, L. Solimando, and R.D. Goldman. 2008. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev*. 22:832-853.

Denoeud, F., P. Kapranov, C. Ucla, A. Frankish, R. Castelo, J. Drenkow, J. Lagarde, T. Alioto, C. Manzano, J. Chrast, S. Dike, C. Wyss, C.N. Henrichsen, N. Holroyd, M.C. Dickson, R. Taylor, Z. Hance, S. Foissac, R.M. Myers, J. Rogers, T.

Hubbard, J. Harrow, R. Guigo, T.R. Gingeras, S.E. Antonarakis, and A. Reymond. 2007. Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res*. 17:746-759.

Dietzel, S., K. Zolghadr, C. Hepperger, and A.S. Belmont. 2004. Differential large-scale chromatin compaction and intranuclear positioning of transcribed versus non-transcribed transgene arrays containing beta-globin regulatory sequences. *J Cell Sci*. 117:4603-4614.

Dillon, N., and R. Festenstein. 2002. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet*. 18:252-258.

Dorer, D.R., and S. Henikoff. 1997. Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics*. 147:1181-1190.

Eltsov, M., K.M. Maclellan, K. Maeshima, A.S. Frangakis, and J. Dubochet. 2008. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc Natl Acad Sci U S A*. 105:19732-19737.

Emery, D.W., E. Yannaki, J. Tubb, and G. Stamatoyannopoulos. 2000. A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci U S A*. 97:9150-9155.

Fan, Y., T. Nikitina, J. Zhao, T.J. Fleury, R. Bhattacharyya, E.E. Bouhassira, A. Stein, C.L. Woodcock, and A.I. Skoultchi. 2005. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*. 123:1199-1212.

Ferrai, C., I.J. de Castro, L. Lavitas, M. Chotalia, and A. Pombo. 2010. Gene positioning. *Cold Spring Harb Perspect Biol*. 2:a000588.

Ferreira, J., G. Paolella, C. Ramos, and A.I. Lamond. 1997. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol*. 139:1597-1610.

Festenstein, R., S.N. Pagakis, K. Hiragami, D. Lyon, A. Verreault, B. Sekkali, and D. Kioussis. 2003. Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. *Science*. 299:719-721.

Finlan, L.E., D. Sproul, I. Thomson, S. Boyle, E. Kerr, P. Perry, B. Ylstra, J.R. Chubb, and W.A. Bickmore. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*. 4:e1000039.

Francastel, C., W. Magis, and M. Groudine. 2001. Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc Natl Acad Sci U S A*. 98:12120-12125.

Francis, N.J., R.E. Kingston, and C.L. Woodcock. 2004. Chromatin compaction by a polycomb group protein complex. *Science*. 306:1574-1577.

Fraser, P. 2006. Transcriptional control thrown for a loop. *Curr Opin Genet Dev*. 16:490-495.

Gall, J.G. 2003. The centennial of the Cajal body. *Nat Rev Mol Cell Biol*. 4:975-980.

Gandor, C., C. Leist, A. Fiechter, and F.A. Asselbergs. 1995. Amplification and expression of recombinant genes in serum-independent Chinese hamster ovary cells. *FEBS Lett*. 377:290-294.

Garrick, D., S. Fiering, D.I. Martin, and E. Whitelaw. 1998. Repeat-induced gene silencing in mammals. *Nat Genet*. 18:56-59.

Georgel, P.T., R.A. Horowitz-Scherer, N. Adkins, C.L. Woodcock, P.A. Wade, and J.C. Hansen. 2003. Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J Biol Chem*. 278:32181-32188.

Gilbert, N., S. Boyle, H. Fiegler, K. Woodfine, N.P. Carter, and W.A. Bickmore. 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*. 118:555-566.

Greil, F., C. Moorman, and B. van Steensel. 2006. DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. *Methods Enzymol*. 410:342-359.

Grigoryev, S.A., G. Arya, S. Correll, C.L. Woodcock, and T. Schlick. 2009. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A*. 106:13317-13322.

Grosveld, F., G.B. van Assendelft, D.R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*. 51:975-985.

Gruenbaum, Y., A. Margalit, R.D. Goldman, D.K. Shumaker, and K.L. Wilson. 2005. The nuclear lamina comes of age. *Nat Rev Mol Cell Biol*. 6:21-31.

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M.B. Faza, W. Talhout, B.H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 453:948-951.

Guy, L.G., R. Kothary, Y. DeRepentigny, N. Delvoye, J. Ellis, and L. Wall. 1996. The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. *Embo J*. 15:3713-3721.

Hall, L.L., K.P. Smith, M. Byron, and J.B. Lawrence. 2006. Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol*. 288:664-675.

Hansen, J.C. 2002. Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct*. 31:361-392.

Heintz, N. 2000. Analysis of mammalian central nervous system gene expression and function using bacterial artificial chromosome-mediated transgenesis. *Hum Mol Genet*. 9:937-943.

Hepperger, C., A. Mannes, J. Merz, J. Peters, and S. Dietzel. 2008. Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma*. 117:535-551.

Hino, S., J. Fan, S. Taguwa, K. Akasaka, and M. Matsuoka. 2004. Sea urchin insulator protects lentiviral vector from silencing by maintaining active chromatin structure. *Gene Ther*. 11:819-828.

Horn, P.J., and C.L. Peterson. 2002. Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science*. 297:1824-1827.

Horowitz, R.A., D.A. Agard, J.W. Sedat, and C.L. Woodcock. 1994. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J Cell Biol*. 125:1-10.

Hu, Y., I. Kireev, M. Plutz, N. Ashourian, and A.S. Belmont. 2009a. Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J Cell Biol*. 185:87-100.

Hu, Y., I. Kireev, M.J. Plutz, N. Ashourian, and A.S. Belmont. 2009b. Large-scale chromatin structure of inducible genes- transcription on a linear template *J. Cell Biol.* 185:87-100.

Iwanaga, R., H. Komori, and K. Ohtani. 2004. Differential regulation of expression of the mammalian DNA repair genes by growth stimulation. *Oncogene*. 23:8581-8590.

Kang, S.H., P.P. Levings, F. Andersen, P.J. Laipis, K.I. Berns, R.T. Zori, and J. Bungert. 2004. Locus control region elements HS2 and HS3 in combination with chromatin boundaries confer high-level expression of a human beta-globin transgene in a centromeric region. *Genes Cells*. 9:1043-1053.

Karpen, G.H. 1994. Position-effect variegation and the new biology of heterochromatin. *Curr Opin Genet Dev*. 4:281-291.

Kaufman, R.M., C.T. Pham, and T.J. Ley. 1999. Transgenic analysis of a 100-kb human beta-globin cluster-containing DNA fragment propagated as a bacterial artificial chromosome. *Blood*. 94:3178-3184.

Kim, J.M., J.S. Kim, D.H. Park, H.S. Kang, J. Yoon, K. Baek, and Y. Yoon. 2004. Improved recombinant gene expression in CHO cells using matrix attachment regions. *J Biotechnol*. 107:95-105.

Kireev, I., M. Lakonishok, W. Liu, V.N. Joshi, R. Powell, and A.S. Belmont. 2008. In vivo immunogold labeling confirms large-scale chromatin folding motifs. *Nat Methods*. 5:311-313.

Kosak, S.T., J.A. Skok, K.L. Medina, R. Riblet, M.M. Le Beau, A.G. Fisher, and H. Singh. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*. 296:158-162.

Kumaran, R.I., and D.L. Spector. 2008. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol*. 180:51-65.

Kwaks, T.H., P. Barnett, W. Hemrika, T. Siersma, R.G. Sewalt, D.P. Satijn, J.F. Brons, R. van Blokland, P. Kwakman, A.L. Kruckeberg, A. Kelder, and A.P. Otte. 2003. Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat Biotechnol*. 21:553-558.

Kwaks, T.H., and A.P. Otte. 2006. Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol*. 24:137-142.

Lawrence, J.B., R.H. Singer, and J.A. McNeil. 1990. Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science*.

249:928-932.
Levi, V., Q. Ruan, M. Plutz, A.S. Belmont, and E. Gratton. 2005. Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys J*. 89:4275-4285.
Li, G., G. Sudlow, and A.S. Belmont. 1998. Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. *J Cell Biol*. 140:975-989.
Li, Q., K.R. Peterson, X. Fang, and G. Stamatoyannopoulos. 2002. Locus control regions. *Blood*. 100:3077-3086.
Linton, J.P., J.Y. Yen, E. Selby, Z. Chen, J.M. Chinsky, K. Liu, R.E. Kellems, and G.F. Crouse. 1989. Dual bidirectional promoters at the mouse dhfr locus: cloning and characterization of two mRNA classes of the divergently transcribed Rep-1 gene. *Mol Cell Biol*. 9:3058-3072.
Luco, R.F., M.A. Maestro, N. Sadoni, D. Zink, and J. Ferrer. 2008. Targeted deficiency of the transcriptional activator Hnf1alpha alters subnuclear positioning of its genomic targets. *PLoS Genet*. 4:e1000079.
Luger, K., A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 389:251-260.
Maeshima, K., S. Hihara, and M. Eltsov. 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol*. 22:291-297.
Mahy, N.L., P.E. Perry, and W.A. Bickmore. 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol*. 159:753-763.
Manuelidis, L. 1990. A view of interphase chromosomes. *Science*. 250:1533-1540.
McKenzie, S.L., S. Henikoff, and M. Meselson. 1975. Localization of RNA from heat-induced polysomes at puff sites in Drosophila melanogaster. *Proc Natl Acad Sci U S A*. 72:1117-1121.
Meister, P., B.D. Towbin, B.L. Pike, A. Ponti, and S.M. Gasser. 2010. The spatial dynamics of tissue-specific promoters during C. elegans development. *Genes Dev*. 24:766-782.
Misteli, T. 2007. Beyond the sequence: cellular organization of genome function. *Cell*. 128:787-800.
Muller, W.G., D. Rieder, G. Kreth, C. Cremer, Z. Trajanoski, and J.G. McNally. 2004. Generic features of tertiary chromatin structure as detected in natural chromosomes. *Mol Cell Biol*. 24:9359-9370.
Muller, W.G., D. Walker, G.L. Hager, and J.G. McNally. 2001. Large-scale chromatin decondensation and recondensation regulated by transcription from a natural promoter. *J Cell Biol*. 154:33-48.
Munkel, C., R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer, and J. Langowski. 1999. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol*. 285:1053-1065.
Mutskov, V.J., C.M. Farrell, P.A. Wade, A.P. Wolffe, and G. Felsenfeld. 2002. The

barrier function of an insulator couples high histone acetylation levels with specific protection of promoter DNA from methylation. *Genes Dev*. 16:1540-1554.

Osborne, C.S., L. Chakalova, J.A. Mitchell, A. Horton, A.L. Wood, D.J. Bolland, A.E. Corcoran, and P. Fraser. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*. 5:e192.

Parseghian, M.H., R.L. Newcomb, and B.A. Hamkalo. 2001. Distribution of somatic H1 subtypes is non-random on active vs. inactive chromatin II: distribution in human adult fibroblasts. *J Cell Biochem*. 83:643-659.

Peric-Hupkes, D., W. Meuleman, L. Pagie, S.W. Bruggeman, I. Solovei, W. Brugman, S. Graf, P. Flicek, R.M. Kerkhoven, M. van Lohuizen, M. Reinders, L. Wessels, and B. van Steensel. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*. 38:603-613.

Peterson, K.R., P.A. Navas, Q. Li, and G. Stamatoyannopoulos. 1998. LCR-dependent gene expression in beta-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. *Hum Mol Genet*. 7:2079-2088.

Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod, and B. van Steensel. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet*. 38:1005-1014.

Pikaart, M.J., F. Recillas-Targa, and G. Felsenfeld. 1998. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev*. 12:2852-2862.

Ragoczy, T., M.A. Bender, A. Telling, R. Byron, and M. Groudine. 2006. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev*. 20:1447-1457.

Ragoczy, T., A. Telling, T. Sawado, M. Groudine, and S.T. Kosak. 2003. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res*. 11:513-525.

Recillas-Targa, F., M.J. Pikaart, B. Burgess-Beusse, A.C. Bell, M.D. Litt, A.G. West, M. Gaszner, and G. Felsenfeld. 2002. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A*. 99:6883-6888.

Reddy, K.L., J.M. Zullo, E. Bertolino, and H. Singh. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*. 452:243-247.

Robinett, C.C., A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, and A.S. Belmont. 1996. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol*. 135:1685-1700.

Rubin, J.E., P. Pasceri, X. Wu, P. Leboulch, and J. Ellis. 2000. Locus control region activity by 5'HS3 requires a functional interaction with beta-globin gene

regulatory elements: expression of novel beta/gamma-globin hybrid transgenes. *Blood*. 95:3242-3249.

Sadoni, N., S. Langer, C. Fauth, G. Bernardi, T. Cremer, B.M. Turner, and D. Zink. 1999. Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. *J Cell Biol*. 146:1211-1226.

Schilling, L.J., and P.J. Farnham. 1995. The bidirectionally transcribed dihydrofolate reductase and rep-3a promoters are growth regulated by distinct mechanisms. *Cell Growth Differ*. 6:541-548.

Schmidt, E.E., and G.F. Merrill. 1989a. Maintenance of dihydrofolate reductase enzyme after disappearance of DHFR mRNA during muscle cell differentiation. *In Vitro Cell Dev Biol*. 25:697-704.

Schmidt, E.E., and G.F. Merrill. 1989b. Transcriptional repression of the mouse dihydrofolate reductase gene during muscle cell commitment. *J Biol Chem*. 264:21247-21256.

Schoenfelder, S., T. Sexton, L. Chakalova, N.F. Cope, A. Horton, S. Andrews, S. Kurukuti, J.A. Mitchell, D. Umlauf, D.S. Dimitrova, C.H. Eskiw, Y. Luo, C.L. Wei, Y. Ruan, J.J. Bieker, and P. Fraser. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*. 42:53-61.

Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc, and G. Cavalli. 2007. Genome regulation by polycomb and trithorax proteins. *Cell*. 128:735-745.

Sekkali, B., H.T. Tran, E. Crabbe, C. De Beule, F. Van Roy, and K. Vleminckx. 2008. Chicken beta-globin insulator overcomes variegation of transgenes in Xenopus embryos. *Faseb J*. 22:2534-2540.

Shizuya, H., B. Birren, U.J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A*. 89:8794-8797.

Simpson, R.T., F. Thoma, and J.M. Brubaker. 1985. Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell*. 42:799-808.

Sinclair, P., Q. Bian, M. Plutz, E. Heard, and A.S. Belmont. 2010. Dynamic plasticity of large-scale chromatin structure revealed by self-assembly of engineered chromosome regions. *J Cell Biol*. 190:761-776.

Slansky, J.E., and P.J. Farnham. 1996. Transcriptional regulation of the dihydrofolate reductase gene. *Bioessays*. 18:55-62.

Smith, K.P., P.T. Moen, K.L. Wydner, J.R. Coleman, and J.B. Lawrence. 1999. Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J Cell Biol*. 144:617-629.

Solovei, I., A. Cavallo, L. Schermelleh, F. Jaunin, C. Scasselati, D. Cmarko, C. Cremer, S. Fakan, and T. Cremer. 2002. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp Cell Res*. 276:10-23.

Springhetti, E.M., N.E. Istomina, J.C. Whisstock, T. Nikitina, C.L. Woodcock, and S.A. Grigoryev. 2003. Role of the M-loop and reactive center loop domains in the folding and bridging of nucleosome arrays by MENT. *J Biol Chem*. 278:43384-43393.

Stamatoyannopoulos, J.A., C.H. Clegg, and Q. Li. 1997. Sheltering of gamma-globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the A gamma-globin gene. *Mol Cell Biol*. 17:240-247.

Takizawa, T., P.R. Gudla, L. Guo, S. Lockett, and T. Misteli. 2008a. Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev*. 22:489-498.

Takizawa, T., K.J. Meaburn, and T. Misteli. 2008b. The meaning of gene positioning. *Cell*. 135:9-13.

Tolhuis, B., R.J. Palstra, E. Splinter, F. Grosveld, and W. de Laat. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*. 10:1453-1465.

Truffinet, V., L. Guglielmi, M. Cogne, and Y. Denizot. 2005. The chicken beta-globin HS4 insulator is not a silver bullet to obtain copy-number dependent expression of transgenes in stable B cell transfectants. *Immunol Lett*. 96:303-304.

Tumbar, T., and A.S. Belmont. 2001. Interphase movements of a DNA chromosome region modulated by VP16 transcriptional activator. *Nat Cell Biol*. 3:134-139.

Tumbar, T., G. Sudlow, and A.S. Belmont. 1999. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol*. 145:1341-1354.

Verschure, P.J., I. van der Kraan, E.M. Manders, D. Hoogstraten, A.B. Houtsmuller, and R. van Driel. 2003. Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep*. 4:861-866.

Vogel, M.J., D. Peric-Hupkes, and B. van Steensel. 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc*. 2:1467-1478.

Volpi, E.V., E. Chevret, T. Jones, R. Vatcheva, J. Williamson, S. Beck, R.D. Campbell, M. Goldsworthy, S.H. Powis, J. Ragoussis, J. Trowsdale, and D. Sheer. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci*. 113 (Pt 9):1565-1576.

Wansink, D.G., W. Schul, I. van der Kraan, B. van Steensel, R. van Driel, and L. de Jong. 1993. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol*. 122:283-293.

Warming, S., N. Costantino, D.L. Court, N.A. Jenkins, and N.G. Copeland. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res*. 33:e36.

Wijgerde, M., F. Grosveld, and P. Fraser. 1995. Transcription complex stability and

chromatin dynamics in vivo. *Nature*. 377:209-213.
Williams, R.R., V. Azuara, P. Perry, S. Sauer, M. Dvorkina, H. Jorgensen, J. Roix, P. McQueen, T. Misteli, M. Merkenschlager, and A.G. Fisher. 2006. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci*. 119:132-140.
Williams, S., T. Mustoe, T. Mulcahy, M. Griffiths, D. Simpson, M. Antoniou, A. Irvine, A. Mountain, and R. Crombie. 2005. CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol*. 5:17.
Woodcock, C.L. 2006. Chromatin architecture. *Curr Opin Struct Biol*. 16:213-220.
Woodcock, C.L., and R.P. Ghosh. 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*. 2:a000596.
Woodcock, C.L., and R.A. Horowitz. 1995. Chromatin organization re-viewed. *Trends Cell Biol*. 5:272-277.
Woodcock, C.L., A.I. Skoultchi, and Y. Fan. 2006. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res*. 14:17-25.
Wuebbles, R.D., M.L. Hanel, and P.L. Jones. 2009. FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. *Dis Model Mech*. 2:267-274.
Wurm, F.M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. 22:1393-1398.
Xing, Y., C.V. Johnson, P.T. Moen, Jr., J.A. McNeil, and J. Lawrence. 1995. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol*. 131:1635-1647.
Yang, X.W., P. Model, and N. Heintz. 1997. Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol*. 15:859-865.
Yokota, H., G. van den Engh, J.E. Hearst, R.K. Sachs, and B.J. Trask. 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J Cell Biol*. 130:1239-1249.
Zahn-Zabal, M., M. Kobr, P.A. Girod, M. Imhof, P. Chatellard, M. de Jesus, F. Wurm, and N. Mermod. 2001. Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol*. 87:29-42.
Zhou, V.W., A. Goren, and B.E. Bernstein. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*. 12:7-18.
Zink, D., M.D. Amaral, A. Englmann, S. Lang, L.A. Clarke, C. Rudolph, F. Alt, K. Luther, C. Braz, N. Sadoni, J. Rosenecker, and D. Schindelhauer. 2004. Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol*. 166:815-825.

APPENDIX

BAC RECOMBINEERING PROTOCOLS

Bacterial artificial chromosomes (BACs) are a class of DNA constructs carrying 100-300 kb of eukaryotic genomic fragments and have been used extensively in a variety of genomic sequencing projects. BACs have also been proven to be valuable tools for study gene functions and regulations and confer several remarkable advantages over typical plasmid constructs. First, because the BAC can carry a much larger size of insertion, it can be used to study the function of large genes. Furthermore, the BACs can contain not only the coding regions of genes but also their distal regulatory elements, which may be hundreds of kb away from the genes. It has been shown in transgenic animals and cell lines the genes contained in the big BAC insertions can be expressed at similar levels to their endogenous counterparts. A recent study has also suggested BACs can be used as tools to survey the expression patterns and functions of genes in a high-throughput way by inserting sequences coding for fluorescent tags to the UTRs of genes contained in the BACs. In addition, by making specific deletions within the BACs, the functions of regulatory elements important for gene expression can be investigated.

Most of the applications of BACs require addition, deletion or introducing point mutations into the BAC sequences. However, due to its large size, it is usually difficult to find appropriate restriction sites to perform the sequence manipulation using standard recombinant DNA technology. Instead, recent technical advances have made it possible to use recombination to engineer BAC DNA sequences. Our lab has been using the BAC recombineering (recombination-mediated genetic engineering) technique to effectively make serial insertions and deletions in BACs. Here I will briefly review the BAC recombineering schemes and introduce the technical

modifications we made on the common recombineering strategies.

BAC recombineering using λ-red encoded genes

An early version of recombineering scheme was designed based on the activity of E.coli RecA gene, which mediates homologous recombination between circular DNA constructs. However, this method has a severe limitation for BAC recombineering as the RecA has been known to cause instability of BAC constructs. BACs carry large eukaryotic genomic inserts and usually contain repetitive sequences and RecA may induce unwanted recombination between repeats and generate rearranged BACs. The screening for correctly recombined BACs may be laborious.

It has later been discovered bacteriophage λ also contains a homologous recombination system which is termed red. λ- red system is more powerful than bacterial recombination system and requires much shorter homology arms to perform recombination at high fidelity. Another advantage of the λ-red system is that it can perform recombination with linear DNA fragment, therefore make it possible to use a PCR fragment for recombineering. The first generation of λ-red system was developed by transforming a RecA negative E.coli strain with a plasmid encoding three proteins: Exo, Bet and Gam. Exo possesses 5' to 3' exonuclease activity and generate 3' end overhangs at the ends of the linear fragment. Bet then binds to the overhangs and mediate recombination between the linear fragment and the target construct. The expression of Gam inhibit the activity of RecBCD proteins from the host bacteria strain, which have linear double strand DNA exonuclease activity, therefore protect the linear fragment from being degraded. The λ-red system is so far the most standard and efficient recombineering system.

The expression of phage recombination machinery may still cause unwanted

rearrangement of BAC constructs. To improve the stability of the λ-red system, the three phage genes are expressed from a λ phage which is integrated into *E.coli* genome. The expression of the three genes is under tight control of temperature sensitive λ-*cI857* repressor. This engineered bacteria strain was normally cultured at 32˚C and the expression of the λ-red recombination machinery is suppressed by the active λ-*cI857* repressor. When the temperature is shifted to 42˚C for 15 minutes, the repressor is inactivated and the recombination machinery is expressed to a high level. The amount of the recombinant machinery will ensure high frequency of recombination.

GalK for efficient positive and negative selection

The selection for correct recombinants after induced recombineering is used achieved by using drug resistant selectable markers. However, for many applications, the removal of the selectable marker after the first round of recombineering is necessary. For instance, if a point mutation or a deletion is to be made in a gene contained in a BAC, the selectable marker needs to be removed to avoid disruption of the coding region. In addition, the selectable markers used for selection in bacterial recombination are usually prokaryotic genes and driven by prokaryotic promoter. The introduction of these exogenous prokaryotic sequences into eukaryotic genomic regions may cause unwanted chromatin effects such as accumulation of DNA methylation, thus give rise to artifacts. Furthermore, in some applications multiple rounds of addition or deletion of sequences need to be made in the same BAC. Although different selectable markers can be used in each round of recombination, this will limit the flexibility of the BAC recombineering. Therefore, selectable markers which can be used in both positive and negative selection need to be

developed for "seamless" sequence manipulations in BACs.

The *sacB-neo* is one of the most widely used positive/negative selectable marker. Neomycin resistance is used for the positive selection. And glucose toxicity in the presence of SacB protein can be used in negative selection to specifically select the clones in which the s*acB-neo* is removed. The drawback of this selection strategy is that the s*acB* is prone to point mutations in *E.coli,* which results in high background in negative selection.

More recently, g*alK* has been developed as a new selectable marker for both positive and negative selection. *E.coli galK* gene encodes protein galactokinase, which is essential for utilizing the galactose as the carbon source. On the other hand, *galK* gene can catalyze the phosphorylation of a galactose isoform 2-deoxy-galactose (DOG), and eventually generate the 2-deoxy-galactose-1-phosphate which is toxic to the cells. To employ *galK* gene as the selectable marker, the endogenous galK gene was removed from the bacterial genome to generate a specialized strain. The BAC to be modified needs to be introduced to this strain. A linear DNA fragment containing galK gene can then be integrated to the BAC by homologous recombination. The recombinants are selected on minimal medium in which the galactose is supplied as the only carbon source. Only the bacteria cells containing the exogenous *galK* can grow. The *galK* can then be removed from the correct recombinants by another round of recombination. And the negative selection is performed with minimal medium containing DOG. Only the recombinants in which the *galK* is removed can survive, therefore resulting in efficient negative selection. Although spontaneous mutations within *galK* can still give rise to background clones in negative selection, the ratio of correctly recombined clones to background clones is relatively high, making the screening very feasible.

Revised recombination scheme for seamless insertion into BAC

Using the GalK selection system, efficient point mutations and deletions within BACs have been achieved. The established scheme for making point mutation in genes within BACs contains two steps. In the first step, a *galK* fragment flanked by about 50bp of homology arms is generated by PCR. The fragment is transformed into the *E.coli* strain carrying the target BAC. After recombination, E.coli clones carrying the BAC which contains the *galK* insertion at the target location are selected on plates containg galactose as the only carbon source. In the second step, a PCR product containing the appropriate homology arms and the point mutation was introduced to replace the *galK*. The negative selection is performed on minimal plates which contain glycerol as the carbon source and also DOG to select for the loss of *galK*. Up to 67% of clones surviving the negative selection are correct recombinants.

We initially adopted this system to insert a DNA fragment about 4-5kb into DHFR BAC. However, after we did the recombination to replace the galK with the insert at the site of interest, we failed to get any correct recombinants from screening a total of 40 colonies after the negative selection with DOG. All the background clones we checked are due to gross deletions within the BAC or the loss of GalK activity without removing it from the BAC. We reasoned that the big size of our DNA fragment may result in low rate of replacement of *galK*, and the correct recombinants may be overwhelmed in the background clones.

To obtain the ratio of correct recombinants to background clones, we replaced the *galK* insertion in the BAC with a 1.7 kb Kan/Neo fragment, and selected the clones either on DOG plates or on DOG plates containing 20ug/ml Kanamycin. The colonies growing on the latter represent the recombinants in which the Kan/Neo selectable

marker replaced the *galK*. We observed that the number of colonies on DOG/Kan plates is only ~2% of the number of colonies on DOG plate. We suspect this percentage may even decrease when using a bigger fragment to replace *galK*. This will make the screening of clones very difficult.

To solve this problem, we revised the scheme of the two-step recombination to achieve the "seamless" insertion of bigger DNA sequences into BACs (Fig). A targeting fragment containing three homology arms A, B and C, each with the size of 45-50bp, is generated. The *galK* selectable marker is then inserted between homology boxes Band C. And the DNA sequence to be added in to the BAC, X, is cloned into the sites between A and B. An alternative method to generate the master targeting fragment is by doing recombinant PCR. This fragment is then introduced to bacteria carrying the target BAC and recombination will occur between homology boxes A and C, resulting the simultaneous insertion of both *galK* and X. The recombinants can be efficiently selected for the presence of galK. To remove *galK*, a DNA fragment containing homology boxes B and C is generated by PCR using the BAC DNA as template, and is used to replace *galK*. Because of the small size of the fragment, the recombination is usually very efficient. As a result, 20%-50% of total colonies after DOG selection are correct recombinants, which can be screened relatively easily.

Summary of the applications of BAC recombineering in Belmont lab

In Belmont lab, we are most interested in using the engineered BAC to create transgene arrays to study the behavior of chromatin. By adding fluorescent tag or MS2 repeats, the mRNA and/or the protein product from genes contained in the BACs can be visualized, making live cell imaging feasible. In addition, by introducing mini-gene reporters to the BAC, we are able to study the influence of large-scale

chromatin folding on gene expression.

The DNA sequences can be conveniently inserted to specific locations within the BAC using the recombineering approaches. The selectable marker for the additon of sequences may or may not need to be removed. If the marker do not need to be removed, any common drug resistance gene can be used as selectable marker. For instance, a DNA fragment containing a GFP reporter driven by a CMV promter and the Kan/Neo resistant gene driven by SV40 promoter was introduced into a BAC containing mouse DHFR locus and the recombinants were selected on LB plates containing Kanamycin. If the marker need to be removed to ensure the flexibility for further manipulation or to minimize effects of prokaryotic sequences, the galK should be the choice of selectable marker. As an example, four minigenes were inserted into different locations within DHFR BAC sequentially. The galK was used as selectable marker for both positive and negative selection for the insertion of each of the four genes.

We are also exploiting BACs to identy cis DNA elements regulating the behavior of chromatin. For instance, by dissecting specific genomic loci we want to identify DNA elements required for the specific nuclear localization of the genes or the chromatin decondensation accompanying transcriptional activation.

We have been able to efficiently delete genomic regions up to 75kb from a 200kb BAC containg human beta-globin locus. Similar to adding sequences to BAC, one can use different selectable markers depending on whether they need to be removed. If galK is used as the selectable marker, it can be removed to facilitate the next round of manipulation in the same BAC.

Details about the recombineering procedures are summarized in the protocols attached.

Protocol 1.

Transform BAC into E.coli SW102 strain for recombineering

E.coli SW102 strain was obtained from Neal Copeland lab (NCI) and is derived from DY380 strain. This strain is specifically engineered for the λ-red mediated DNA recombination and GalK selection. To suppress the expression of recombination machinery, the strain should be culture at 32˚C. Culturing at elevated temperature may cause genomic instability of the strain.

To modify a BAC with recombination, the first step is to introduce the BAC DNA into the SW102 strain. In this protocol, SW102 E.coli cells are made electrocompetent and the BAC DNA is transformed into the cells by electroporation.

1. Before transformation, the BAC DNA should be purified by mini-prep or maxi-prep. The quality and the integrity of the BAC DNA should be checked by restriction digestion pattern generated by appropriate enzymes.
2. Autoclave one 250ml flask and four 30ml centrifuge tubes the day before transformation.
3. The day before transformation, inoculate a 2ml LB culture with the frozen glycerol stock of the SW102 cell. Grow the culture overnight in a 32˚C shaking waterbath.
4. Inoculate a 100ml LB culture in a flask with 1ml of the overnight culture and grow the culture at 32˚C. It will take approximately 3 hours for the culture to reach the OD600 value of 0.55-0.6. Start monitoring the OD600 of the culture from the 2hr45min mark to ensure the culture is not overgrown.

5. Prepare an ice-water bath before the culture is ready. Cool down ddH2O and the centrifuge tubes on ice.
6. When the culture is ready, briefly cool it down on ice. Then aliquot 25ml culture to each centrifuge tube.
7. Centrifuge the culture at 5,000 rpm for 5min at 4˚C withe the Sorvall SS34 rotor. Pour off the supernatant. Put the tubes back to the ice-water bath. Then add 1ml of ice-cold water to each tube. Swirl the tube in the ice-water slurry until the pellet is completely resuspended. This may take 1-2 minutes for each tube. Add 9ml more ice-cold water to the tube to bring the total volume to 10ml, then swirl briefly.
8. Repeat step 7. The pellet will become much looser after this centrifugation. The supernatant should be removed with a 10ml serological pipette with great caution instead of poured off. Remove as much supernatant as you can without disturbing the pellet. Because the pellet is loose, it should be easily resuspended.
9. Centrifuge again. After this centrifugation, remove as much supernatant as you can, this usually leaves 0.5-1 ml of supernatant in the tube. Resuspend the pellet with a pipetteman, then transfer the suspension in each centrifuge tube to a clean 1.5ml microcentrifuge tube. Keep the cell suspension on ice.
10. Centrifuge the suspension using a benchtop centrifuge machine in 4˚C cold room at 1,000 g for 5min. Remove supernatant to reduce the volume to about 200µl per tube. This will give you 200ul electrocompetent SW102 cells per 25 ml of culture or 800µl competent cells per 100 ml culture. Combine the cells in all 4 tubes and aliquot them to 500ul microcentrifuge tubes at 50ul per tube.

Unused SW102 competent cells can be flash frozen with liquid nitrogen or ethanol/ dry ice and stored at -80˚C.

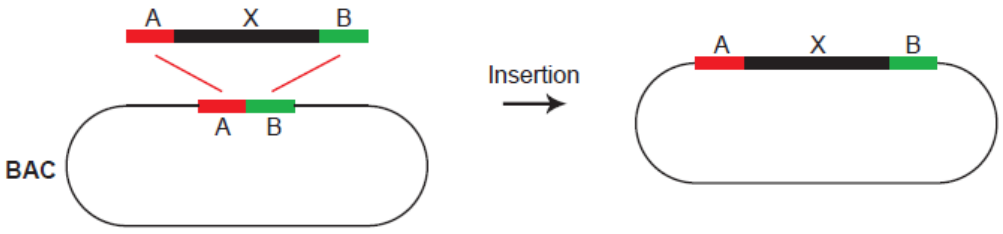
11. Mix the BAC DNA to be transformed with 50ul SW102 competent cells. Use 10-50ng of maxi-prepped or 1ul of mini-prepped BAC for each transformation. Mix DNA with competent cells thoroughly by pipetting.
12. Perform electroporation following standard protocol. Using 0.2 cm electroporation cuvette and the Bio-Rad electroporator with the following settings: Resistance=200 Ω, Capacitance= 25 μF and Voltage= 2.5kV.
13. Recover the cells by adding 950ul of SOC medium. Shake in 32˚C waterbath for 1hr.
14. Plate 100ul (1/10 of the total cells) on a LB plate containing 12.5ug/ml chloramphenicol. Incubate the plate at a 32˚C incubator. The colonies should be evident after ~24 hours incubation. Depending on the quality of DNA there should be tens to hundreds of colonies on the plate.
15. Inoculate 5ml LB cultures containing chloramphenicol with several colonies. Use 4ml of the overnight culture for mini-prep. Check the restriction digestion patterns of the DNA to make sure there are no unwanted rearrangements with in the BAC. Freeze down the clones with correct digestion patterns.

Protocol 2.

Insertion of DNA sequences into BACs using drug resistant selectable marker

This protocol describes the most straightforward recombineering, which is to insert a fragment containing a prokaryotic selectable marker. The correct recombinants will be selected using antibiotics, which is usually very efficient. However, the drawback of this protocol is the selectable marker can not be removed subsequently. We have successfully inserted fragments up to 4.5 kb containing a GFP or RFP reporter together with a Kan/Neo or Zeocin selectable marker into the BAC of interest using this protocol.

This protocol consists of three parts: the generation of DNA fragment for recombination, the recombination and the screening of recombinants.



Protocol 2.1 Generation of DNA fragment containing drug resistant selectable marker for recombination

A DNA fragment containing the selectable marker and the homology boxes is usually generated by PCR. First, the target location in the BAC for the insertion needs to be determined. The DNA sequences flanking the targeting site will be the homology boxes for the recombination. The PCR primers should be designed to contain a 43-47 bp homology box (43 bp is the minimal length for efficient recombination) and a 13-17 bp sequence complementary to the ends of the sequence

to be inserted. The total length of the primers should be 60bp (this is the most economical length for oligo synthesis).

1. Amplify the sequence to be inserted from a plasmid template using standard PCR protocol. The PCR program should consist of two parts. For the first 5 cycles, the annealing temperature should be set at about 3 degrees lower the melting temperatures of the 13-17 bp DNA stretch within the primers. This is followed by 35 more cycles with the primer annealing temperature set at about 3 degrees lower than the melting temperature of the full length primers.
2. After the PCR amplification, run 2ul of the samples from a 50ul PCR reaction on a mini-gel. If the PCR works, add 2ul of DpnI directly to the PCR reaction and incubate at 37˚C for 2-4 hours. The DpnI enzyme is a frequent cutter which digests the plasmid template but not the PCR product.
3. After the digestion, gel-purify the PCR product. The gel should be run slowly. I usually set the voltage at ~100V to get better separation of the bands.
4. Run the recovered PCR product on a mini-gel to check the purity of the PCR product and estimate the concentration. The concentration of the DNA fragment should be at least 20ng/µl and usually the higher the better. The purified PCR product can be stored at -20˚C.

Protocol 2.2 Insertion of DNA sequences of interest into BAC by homologous recombination

In this protocol, the SW102 E.coli cells already carrying the BAC will be induced by heat-shock to express the recombination machinery, then made electrocompetent.

DNA fragment generated in protocol 2.1 will then be transformed into these cells by electroporation. The recombinants will be selected by antibiotics. As the negative control for the recombination, the same amount of the DNA fragment needs to be transformed into uninduced SW102 cells and selected with the same antibiotics. The colonies grown on the control plates indicate the background from the carry-over of the undigested plasmid template used in previous PCR.

1. The day before recombination, inoculate a 5ml LB culture containing 12.5ug/ml chloramphenicol with the SW102 bacteria clone carrying the BAC to be modified. Grow the culture overnight in a 32˚C shaking waterbath.
2. Inoculate a 50ml LB culture with 1ml of the overnight culture and shake at 32˚C. Grow the culture to the OD600 of about 0.55-0.6. Monitor the OD600 value as described before.
3. After the culture reaches the optimal density, split the culture into halves. Transfer 25ml of the culture into a 30ml centrifuge tube and leave it on ice. This will be used as the uninduced control for recombination. Heat-shock the other half of the culture at 42˚C for 15min in a shaking waterbath. After heat-shock, briefly cool down the culture on ice, then transfer it to a 30ml centrifuge tube.
4. Follow steps 7-10 in protocol 1. Wash the bacteria with ice-water to make both the induced and uninduced SW102 cells electrocompetent. Reduce the volume of each 25ml culture to 150-200ul, this will be enough for 3-4 electroporation reactions.

5. Transform 5ul of the DNA fragment prepared in protocol 2.1 into 50ul of induced or uninduced SW102 competent cells following standard procedures. Recover the cells with 950ul of SOC medium and shake in 32˚C waterbath for 1hr. The recombination reaction will occur in the bacteria during this period.
6. After recovery, plate the induced and uninduced samples on LB plates containing chloramphenicol and the appropriate antibiotics for selection. The recombination efficiency may vary depending on the homology arms and the size of the DNA fragment. I usually plate 1/10 and the rest (9/10) of the culture of both induced and uninduced samples on plates.
7. Incubate the plates at a 32˚C incubator. The colonies should become obvious after 24hrs, which is much faster than the selection using GalK and minimal medium. Compare the numbers of colonies on the plates of induced and uninduced samples. There are usually at least 5 times more colonies on the plate of induced sample versus uninduced sample.

Protocol 2.3 Screening for clones carrying correctly recombined BACs

After drug selection, PCR will be used to confirm whether the DNA sequence of interest is inserted into the correct site within the BAC. The PCR primers bind to regions flanking the homology boxes can be used to amplify the regions between them in the engineered BAC. The size of the product will indicate whether there is insertion in the desired location. However, when the insertion is big (>4kb), this PCR may not work efficiently and likely give rise to high background. To efficiently screen for correct recombinants, we usually use two primer pairs to check the ends of the insertion. Each of the primer pair consists of one primer binding to a region outside of the homology box within the BAC and one primer binding to a region within the

insertion. The primers should be designed in a way that the product size will be 100-500 bp. The specific PCR product from either pair of the primers will only be obtained when the insertion occurs at the correct location.

After the verification of the recombinants by PCR method, the routine restriction digestion should be performed to check whether there are unwanted rearrangements of the BACs and, in case of the presence of lac operator repeats, the shortening of repetitive sequences.

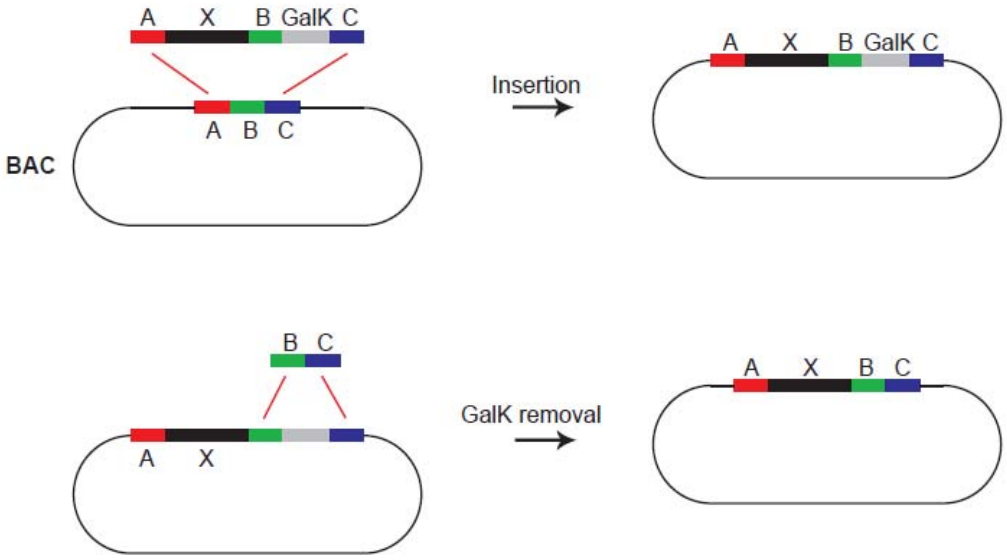
1. Draw grids on the bottom of a LB-Chloramphenicol plate and number the grids. Pick a colony from the experimental plate with a pipette tip and dab onto the gridded plate. Transfer 20-30 colonies for each experiment. Incubate the gridded plate at 32˚C overnight.
2. After the colonies grow up on the gridded plate, scrape some bacteria from each colony with a pipette and transfer the bacteria into a 500ul eppendorf tube containing 25ul ddH_2O. Resuspend the bacteria thoroughly by pipetting or vortexing.
3. Lyse the bacteria by heating at 99˚C for 15min on a thermocycler. Briefly cool down the tubes on ice then spin the cells down at top speed for 2min with a benchtop centrifuge. Transfer 20ul of the supernatant to a clean tube for each clone. This will be used as the template for PCR.
4. Perform PCR following standard protocols. Use 1ul of the colony prep as template for each 25ul PCR reaction. Also set up a control reaction using unmodified BAC DNA as PCR template. Run agarose gel to check the size of

the PCR products. The percentage of correct recombinants may vary from 10-100%.

5. For the clones showing correct PCR products, inoculate 5ml LB+ Chloramphenicol cultures with colonies on the gridded plate. Grow the cultures overnight at 32˚C. Mini-prep the BAC DNA and check the restriction digestion pattern with appropriate enzymes.
6. Freeze down the clones carrying the correctly modified BACs.

Protocol 3. Seamless insertion of DNA sequences into BAC using *galK* as the dual selectable marker

As discussed before, the use of common drug resistant selectable marker usually does not allow the removal of marker afterwards. To achieve seamless insertion of the sequence of interest, we use *galK* as the selectable marker for both positive and negative selection. The *galK* will be inserted along with the sequence of interest into the target BAC. The recombinants are selected with minimal medium containing galactose. A second recombination will then be performed to remove the *galK* and the recombinants are selected with minimal medium containing DOG. This scheme is particularly useful for sequential insertion of multiple DNA sequences into different locations in the same BAC.



Protocol 3.1 Generation of a DNA fragment containing homology regions, *galK* selection cassette and the sequence to be inserted

As described before, a DNA fragment consisting of five segments: homology

regions A, B and C, the *galK* selection cassette and the sequence X of interest needs to be generated. The homologous regions should be at least 43bp in length for efficient recombination. The sequence X is placed between homology regions A and B. The *galK* selection cassette is placed between homology regions B and C and will eventually be removed.

This five-element fragment can be generated by routine molecular cloning. To speed up the cloning procedure, an oligo containing homology regions A, B and C and appropriate restriction sites can be synthesized. Then the *galK* and sequence X can be inserted at the restriction sites sequentially. Theoretically, this fragment can also be generated by a PCR based method called recombinant PCR. However, we haven't been able to test this method because all the DNA sequences we have inserted with the *galK* strategy are relatively large (ranging from 4.5-6kb) and the PCR amplification of them is relatively inefficient.

1. Design the oligo containing homology regions A, B and C. First decide the site of insertion in the BAC. The 50bp regions flanking the targeting site can then be picked as homology regions A and B. And the 50bp region downstream of B can then be picked as homology region C. Check the restriction sites with in the regions A, B and C and make sure they do not contain the sites will be used later for cloning. Then add the sequences of the restriction sites between the homology regions (I usually use site EcoRI to insert *galK* cassette between B and C). Also add a rare restriction site to either end of the oligo (I usually add blunt-end restriction sites PmeI to both ends).

2. A pair of primers which match the ends of homology region B and C should also be designed, this will be used to generate a DNA fragment consisting B and C for the removal of *galK*.
3. Transform the commercially synthesized oligo, which is already been inserted in a vector of choice, to an appropriate *E.coli* strain. Purify the plasmid DNA. Insert the *galK* selection cassette, which can be digested from plasmid pGalK, into the site between homology regions B and C. Screen for correct clones
4. Insert the sequence X into the site(s) between homology regions A and B. Screen for correct clones.
5. Purify the plasmid DNA. Digest the plasmid with the restriction enzymes which cut at the ends of the oligo to release the five-element fragment.

Protocol 3.2 Insertion of the sequence of interest to the BAC followed by Gal selection

The experiment procedure for the insertion of the sequence of interest to the target BAC is almost identical to protocol 2.2. The only difference is that the SW102 bacteria need to be washed with minimal medium for multiple times before plating to remove the traces of rich medium. The selection of recombinants on minimal plates is slower than common drug selection, which takes 3-4 days.

Recipes for the M9 medium and the M63 minimal plates (GAL or DOG) can be found at:

http://web.ncifcrf.gov/research/brb/protocol/Protocol3_SW102_galK_v2.pdf

1-5. Same as Protocol 2.2

6. After recovery, transfer the 1ml cultures to autoclaved eppendorf tubes. Spin the cells down with a benchtop centrifuge at top speed for 1min.
7. Carefully remove the supernatant with pipettes as some of the pellets may be loose. Resuspend the pellet thoroughly with 1ml of M9 minimal medium. Then centrifuge again.
8. Repeat step 7.
9. Remove the supernatant and resuspend the pellet with 1ml M9 minimal medium. Plate 1/10 and 9/10 of the induced and uninduced samples on M63-Gal minimal plates.
10. Incubate the plates in a 32˚C incubator. The plates should be placed into humid containers as they may dry out. The colonies should become obvious by day 3 of selection and reach desirable size at day 3 or day 4.

Protocol 3.3 Screening for clones carrying correctly recombined BACs

The clones will be screened by PCR to check whether the insertion occurs at the correct location. The strategy and experimental procedures are the same as protocol 2.3.

Protocol 3.4 Generation of DNA fragment for removing *galK*

The DNA sequence of interest, X, and the *galK* selection cassette have been inserted into the BAC. And *galK* can be removed by a second round of recombination followed by negative selection. Because *galK* is inserted between homology regions B and C, a DNA fragment consisting only homology regions B and C can be generated by PCR and used to remove *galK*.

1. Perform standard PCR using the primers designed in Protocol 3.1 and the unmodified BAC DNA as template.
2. Purify the PCR product using Qiagen PCR purification kit. Gel purification is not necessary. The concentration of purified fragment should be at least 20ng/μl.

Protocol 3.5 Removal of *galK* followed by negative selection with DOG

The *galK* selection cassette will be removed and the recombinants are selected with DOG, which is toxic to the bacteria in the presence of GalK enzyme. However, the loss-of-function mutations of the *galK* gene or the loss of the *galK* by unwanted deletions of the BAC can also give rise to clones surviving the DOG selection. As a result, the number of background colonies after negative selection is significantly higher than positive selection. As the control for the *galK* removal, instead of transfecting DNA fragment into uninduced SW102 bacteria, the same volume of ddH_2O or EB buffer is transfected into heat-shocked sample to evaluate the level of spontaneous loss of *galK*.

Another possible source of background could be the carry-over of the SW102 cells containing unmodified BAC. In the positive selection using galactose as the only carbon source, cells not expressing GalK stop growing, instead of being killed. Therefore, when colonies are picked from the GAL plate, there may be small numbers of "hitchhikers" also being picked, which can recover later in normal media and survive the DOG selection following the *galK* removal. To eliminate the contamination of the "hitchhikers", the *galK*-positive clone should be streaked onto a

special indicator plate (MacConkey indicator plates). And only bright red colonies should be picked for subsequent removal of *galK*.

Recipe for MacConkey indicator plates: Add 20g Difco MacConkey Agar Base (BD, catalog# 281810) and 5g galactose to 500ml water, autoclave to sterilize. Let cool down to about 50˚C. Add 500ul of 12.5 mg/ml Chloramphenicol. Pour plates, 20ml per plate.

3. Streak the glycerol stock *galK*-positive clone onto a MacConkey plate to obtain single colonies. Incubate the plate at 32˚C. Bright red colonies should become obvious after 24 hours.
4. Pick one bright red colony from the MacConkey plate and inoculate a 5ml culture containing Chloramphenicol. Shake overnight at 32˚C.
5. Inoculate a 25 ml culture with 0.5 ml overnight culture. Grow the culture to the OD600 of 0.55-0.6.
6. Heat shock the culture at 42˚C for 15min.
7. Make the cells electrocompetent as described before.
8. Electroporate 50ul of the electrocompetent cells with 5µl of the DNA fragment prepared in Protocol 3.4. As control, electroporate another 50µl of cells with 5µl of EB buffer. Recover the cells in 10ml LB for 4.5 hours at 32˚C.
9. As in protocol 3.2, pellet 1ml of the recovered culture and wash twice with 1xM9 salts. Then make two serial dilutions with 1xM9 salts. Plate 100ul of the resuspended pellet (1/100 of the total cells), 100µl of 1:10 dilution (1/1000 of the total cells) and 100µl of 1:100 dilutions (1/10000 of the total cells) on

DOG plates. Incubate at 32˚C for 3 days. The number of colonies on the recombination plates may not be significantly different to the control plates (could vary from 1:1 to 100:1, according to literature). Screen 20-30 colonies from the recombination plate.

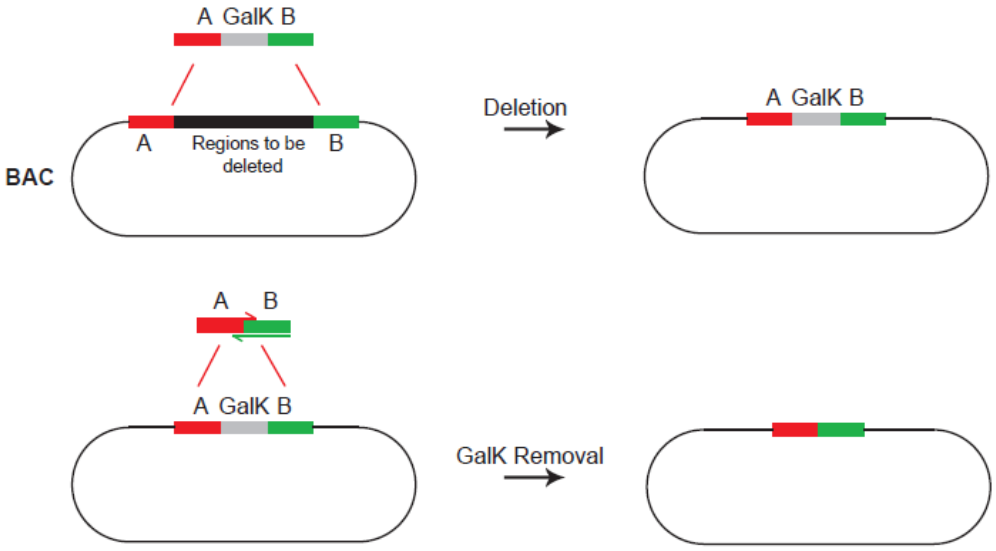
Protocol 3.6 Screening for clones carrying correctly recombined BACs

Following *galK* removal, the recombinants should again be screened using PCR. I usually use primers outside of the *galK* sequence, with one primer upstream of homology box B and in the insertion X and the other primer downstream of homology box C, for the screening. The correct product size will not only indicate the *galK* has been removed but not confirm the sequence X is still at the correct location. The experiment procedures for PCR screening are similar to protocol 2.3.

Following the PCR screening the restriction fingerprinting should be performed to confirm the correct arrangement of the BAC sequences. By choosing proper enzyme the loss of the ~1.5 kb *galK* fragment should be obvious.

Protocol 4. Deletion of specific BAC sequences using *galK* as the dual selectable marker

BAC recombineering provides a convenient system for deleting specific sequences from the BAC. Using this method I have deleted regions ranging from 2.8kb (the promoter region in DHFR BAC) to 116kb (the beta-globin genes in HBB BAC). The use of *galK* as selectable marker again allows the removal of *galK* and making subsequent deletions from one deletion.



Protocol 4.1 Generation of a DNA fragment containing homology regions and *galK* using PCR

The recombination fragment for making deletions can be conveniently generated using PCR. The sequences flanking the region to be deleted are chosen as the homology boxes. And a plasmid containing *galK*-pGalK (from Neal Copeland lab) is used as the PCR template. The PCR primers should consist of two parts, the 17bp sequence homology to the template pGalK and the 43 bp homology box, similar to the primes described in protocol 2.1.

The 17bp universal sequences I have been using to PCR pGalK are:

Forward primer: 5'-cgacggccagtgaattg-3'

Reverse primer: 5'-tgcttccggctcgtatg-3'

As described in protocol 2.1, it is critical to completely remove the pGalK template from the PCR product by doing DpnI digestion and gel extraction as this will contribute to most of the background for the positive selection.

Protocol 4.2 Deletion of sequences followed by Gal selection

The experimental procedure is exactly the same as protocol 3.2. The uninduced control should give nearly zero colonies while the one tenth of the induced sample should give dozens to hundreds of colonies after the selection.

Protocol 4.3 Screening for clones carrying correctly recombined BACs

PCR will be used to screen the recombinants. The PCR primers flanking the deleted region should be used. And the BAC before the deletions should be used as the negative control. Due to the usually big size of the deleted regions, the control BAC should not give rise to any PCR product. And the correct recombinants should produce a PCR product containing the *galK,* which should have the size between 1.5-2.5kb, depending on the primers used.

Following the PCR screening, the restriction fingerprinting should be performed.

If the desirable recombinants are obtained, the BAC can be stored as glycerol stock at -80˚C or streaked on a MAC plate for subsequent removal of *galK* and further deletions.

Protocol 4.4 Generation of DNA fragment for removing *galK*

The DNA fragments for the removal of GalK were generated by PCR using 60bp primers partially complementary to each other (GkRm-F and –R). Each of the 60bp primers consists of a 52bp homology region flanking the *galK* marker and a 8bp sequence complementary to the last 8 bp of the other homology region. This will give rise to a 16bp complementary region between the two primers. For PCR, no extra template is needed. The PCR program should consist of 5 cycles with annealing temperature lower than the melting temperature of the 16bp stretch and 35 cycles with annealing temperature lower than the melting temperature of the full length primers, as described in protocol 2.1.

Protocol 4.5 Removal of *galK* followed by negative selection with DOG

The experimental procedures are exactly the same as protocol 3.5.

Protocol 4.6 Screening for clones carrying correctly recombined BACs

The same pair of primers designed for screening in protocol 4.3 can be used to screen for the loss of *galK*. The PCR products should have the sizes of several hundred bp after the removal of *galK*.