Trapping Enhancers by Transgenic Expression of BACs Engineered in Bacteria with loxP Transposons

Hope M Wolf[1,2], Kevin O Nyabera[1], Katya K De La Torre[1], Mugtaba A Eltayeb[1], Leighcraft A Shakes[1], Charles Hatcher[1], Derek C Norford[1] and Pradeep K Chatterjee[1*]

[1]Julius L. Chambers Biomedical/Biotechnology Research Institute & Department of Chemistry, North Carolina Central University, 1801 Fayetteville Street, Durham, NC 27707, USA

[2]Department of Chemistry, University of North Carolina at Chapel-Hill, Chapel-Hill, NC 27599, USA

Abstract

Bacterial Artificial Chromosomes (BACs) are large extra-chromosomal plasmids in bacteria that faithfully propagate large pieces of DNA from the chromosomes of organisms. Because they represent tiny contiguous pieces of the chromosome, BACs are ideally suited for expression of genes in their chromosomal contexts. Genes in BACs can be monitored for expression after the DNA is modified with reporter genes and other sequences that allow it to be stably propagated in the new host. Several methods have been developed to alter BAC DNA within its bacterial host. One approach uses Tn10 mini-transposons to introduce exogenous DNA into BACs. The random insertions of Tn10 carrying lox sites have directed mammalian cell-selectable antibiotic resistance genes, enhancer-traps and inverted repeat ends of the vertebrate transposon Tol2 precisely to the ends of genomic DNA inserts in BACs. Reporter gene expression from BAC DNA integrated into zebrafish or mouse chromosomes have resulted from such retrofitting. The methodology has been used extensively to analyze regulation of the Amyloid Precursor Protein (*appb*) gene in zebrafish. Functional identification of long-range regulatory sequences of *appb* has provided important clues for regulation of the APP gene in humans.

Introduction

DNA within cells is highly compacted primarily through wrapping around histone proteins in eukaryotes, or HU proteins in bacteria, and these nucleoprotein structures are further condensed in chromatin through higher orders of packaging as sketched in Figure 1. The energy required for bending the otherwise rigid double stranded DNA to wrap around proteins comes from favorable interactions between the packaging proteins and the DNA, and includes interactions from both the highly negatively charged phosphate backbone as well as the base pairs. Proteins that bind to specific sequences of DNA optimize these interactions as well.

Some sequences in double stranded DNA prefer to exist not as a straight helix but as curved or bent helixes in solution under physiological conditions [1]. The energetics of protein binding to curved double-helical DNA can become complicated for both histones in nucleosomes and those recognizing specific sequences [2,3]. Analyses of such interactions *in vitro*, as well as computational modeling, by several investigators have provided some understanding of their contribution to phasing nucleosomes along the DNA and positioning regulatory proteins in the neighborhood of genes [4-6]. They also highlight the contribution of flanking sequences to the affinity of gene-regulatory protein(s) for a consensus binding site [2-6].

A persistent challenge to our understanding of gene regulation is the mechanism by which regulatory proteins interact between themselves and the transcription machinery in a combinatorial fashion to bring about tissue- and time-specific expression of a gene. Testing the efficacy of individual regulatory elements to drive expression of reporter genes in small plasmid constructs has allowed a large number of tissue specific enhancers to be quickly identified in the mouse and zebrafish systems [7-9]. Mutational analyses of putative DNA binding sites of regulatory proteins have verified many of these findings. More specialized vectors using the Tol2 system, for efficient germline expression, have been developed [10]. However, expression analyses of
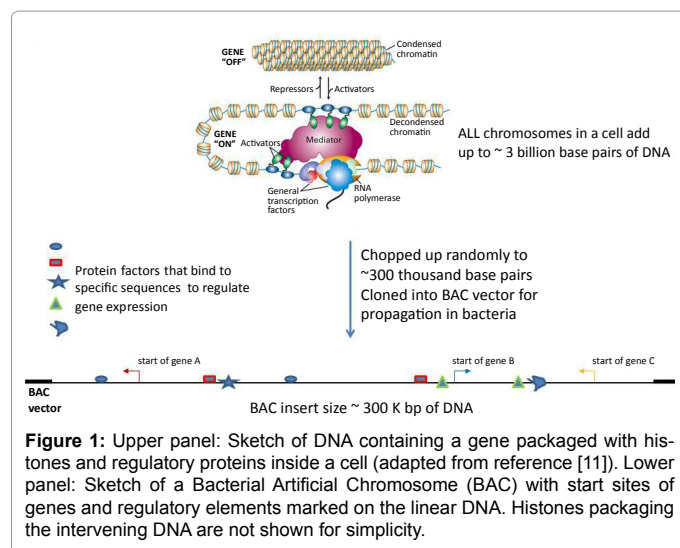

Figure 1: Upper panel: Sketch of DNA containing a gene packaged with histones and regulatory proteins inside a cell (adapted from reference [11]). Lower panel: Sketch of a Bacterial Artificial Chromosome (BAC) with start sites of genes and regulatory elements marked on the linear DNA. Histones packaging the intervening DNA are not shown for simplicity.

*Corresponding author: Pradeep K Chatterjee, Julius L. Chambers Biomedical/Biotechnology Research Institute & Department of Chemistry, North Carolina Central University, 1801 Fayetteville Street, Durham, NC 27707, USA, Tel: (919) 530-7017; Fax: (919) 530-7998; E-mail: pchatterjee@nccu.edu

individual regulators, one at a time, can be misleading. For example a single entity in a multicomponent regulatory system might not have activity by itself. The role of sequences adjoining a consensus site to the binding affinity of the protein can be elusive to track. Lack of strict sequence specificity by some proteins, as exemplified by the TATA binding protein or the High Mobility Group (HMG) proteins which appear to recognize specific features on DNA such as minor groove accessibility and/or flexibility [11] rather than a specific sequence, adds to the problem of identifying potential binding sites for mutational analyses.

Finally, it is important to note genes expressed from small plasmids integrated into the germline frequently suffer from position effects. Depending on the chromosomal environment, both gene silencing and altered expression can occur after transgenesis [12,13]. Vectors with "insulator sequences" to protect the minimal promoter from position effects have been developed [14]. Additionally, both "knock-in" technology using homologous recombination [11], as well as efficient site-specific transgenesis using loxP sites [15-18], or the PhiC31 integrase [19,20], have been described to overcome this hurdle and the latter method used recently for testing enhancer activity in medaka.

Although gene regulatory sequences comprise non-coding RNA, splicing elements and sequences that mark chromatin structure, here we focus our discussion only to those that bind regulatory proteins to direct expression of a gene in a spatio-temporal manner. Because many such regulatory components that fine-tune expression of a gene remain unknown, picking potential regulatory sequences to mutate can introduce bias. In favorable circumstances, clues can be obtained from sequence comparisons of gene regulatory DNA regions across different species [7-9,21]. Numerous studies in the past have used the concept successfully to identify novel *cis*-acting regulatory elements [7-9]. However in situations where gene regulatory function is conserved without sequence similarity, choosing potential protein binding sequences can become a hit-or-miss affair [22-26]. Thus a more inclusive approach for screening potential regulatory sequences is desired.

Regulatory proteins bound to different sites along the DNA must interact with one another to transcribe the gene. The distances separating such protein-bound sites along the DNA can sometimes be quite large, in the tens of thousands of base pairs in higher vertebrates [7-9,21]. Interactions among proteins bound at different sites are thus likely to be influenced by both the length and flexibility of the intervening DNA. Bending of the intervening DNA with or without protein binding, facilitates such interactions. Appropriate phasing of the regulatory protein bound sites with nucleosomes and other proteins that package the intervening DNA often appear necessary for optimal regulation of the gene [2-6]. It is also likely that evolutionary pressures have fine-tuned the sequence variations allowed in these intervening DNA segments. Because rules governing the sequence dependence of DNA flexibility remain inadequate, it is probable that the binding affinities of proteins to their DNA sites and other proteins are altered when consensus sequences are taken out of their chromosomal contexts to construct small plasmids for expressing a gene. If the gene regulatory protein exists in the cell as part of a larger protein complex, the problem can be exacerbated. Thus the usefulness of many of the above approaches using small plasmid constructs appears limited if the regulatory role of a protein depends on the context it finds itself with respect to other protein regulators participating in the expression. The tissue-specificity of an isolated expression-enhancing element tested out of its native context may be quite different from that in the context of its own gene.

Therefore it is probably important to test function of gene regulatory elements in their native context with large stretches of surrounding DNA sequences to capture their complete profile [27]. It requires special vectors to clone and capture large segments of genomic DNA for *in vivo* testing through creation of transgenic animals. Fortunately Bacterial Artificial Chromosomes, (BACs), developed originally for genome sequencing projects appear to fulfill that role [28-30].

Several elegant approaches have been taken over the years to address this complexity in the interactions between regulatory elements and their protein counterparts in gene expression:

1. Enhancer-trapping in animals using the entire genome

2. Mutating putative protein binding sites in BACs using a) homologous recombination, or b) Transposition followed by progressive deletions from BAC ends

3. Enhancer-trapping using BACs

Inserting Enhancer-traps Directly into Genomes of Organisms: Traditional Enhancer-trapping

One of the earliest approaches taken to identify expression-enhancing elements of a gene in the context of their chromosomal settings, before the era of BACs, involved inserting transposable elements directly into the genomes of animals. A basal-promoter driven reporter gene, such as GFP, was engineered into the transposable element for easy screening of transgenic lines. Thus enhancer-trap modified Ty-elements and retro-viral vectors inserted directly into the whole-genomes of drosophila, or mice and zebrafish embryos, respectively, were used to identify and isolate numerous enhancer-trap reporter lines [31-39]. However for reasons discussed below, whole-genome enhancer-trapping appears not to be ideally suited for functionally dissecting the contribution of individual regulatory elements regulating expression of a gene.

Using BACs for Accurate Expression of Genes in Vertebrates: Mutational Analyses in BACs

Many of the hurdles encountered in functionally dissecting regulators of gene expression can be circumvented by using BACs. Genomes of higher vertebrates expanded during evolution and physically separated many regulatory factor binding sites from one another along the DNA and transcription start sites of genes (Figure 1). Thus BACs, which represent tiny pieces of the chromosome originally intended for mere faithful propagation of the DNA cloned in them, now appear ideally suited not only for accurate expression of the genes housed in them but also for dissecting the long-range interactions between gene regulators through mutational analyses.

Functionalizing BACs

BACs faithfully propagate ~300,000 bp pieces of DNA from the genome of an organism in bacteria as large single copy plasmids [28-30]. Ideally, genes contained in a BAC should be expressed in the same organism from which the chromosomal DNA was obtained for constructing the BAC library. This is largely because both the proteins and the sequences separating binding sites of protein complexes along the DNA have co-evolved for optimal function over millions of years. However, expressing genes in the same host leaves one with no identifying features to distinguish its expression from that of the organism's genome. A reporter to tag genes from the BAC is therefore desirable, and the GFP gene is useful for organisms that exhibit optical transparency for the early part of their life cycle such as zebrafish.

BACs are propagated as part of a large F-factor plasmid in bacteria and cannot replicate in the animal [28]. Chromosomal integration of the reporter gene-tagged BAC DNA is thus required for stable expression. In order for the BAC DNA to be truly germline, integration into the chromosome must occur at the one-cell stage of the fertilized egg during development.

Strategies for Functionalizing BACs

a) Using homologous recombination

Procedures for altering sequences in small plasmids developed during the 70's and 80's are not suitable for modifying the large DNA inserts in BACs. There are at least two important reasons why strategies for altering BAC DNA differ fundamentally from small plasmids: common restriction endo-nucleases that cut DNA at specific sequences have small recognition sites, and cut frequently in the BAC DNA to generate large numbers of fragments. Separation techniques robust enough to isolate and keep track of the relative order of these pieces in the BAC are unavailable, which precludes using the "cut-and-paste" method to alter sequences. Secondly, large duplex DNA that is not packaged with DNA-binding proteins in the test-tube is very brittle, and keeping it intact during *in vitro* manipulations is difficult.

Hurdles such as these can be overcome using DNA recombination *in vivo*, where the entire process of cutting and re-joining occurs in a concerted manner within the bacterial host with no free ends of DNA to participate in unintended biochemical reactions in the cell. The nucleoprotein complex in which this occurs also protects the large DNA from shear forces.

There are two independent methods that use sequence homology for recombining DNA to alter sequences in BACs. The first approach introduced the major recombination function of *E. coli*, RecA [40], back into the severely recombination deficient host bacterium DH10B engineered to propagate vertebrate DNA. Recombination is a vital function in all living organisms, and was deliberately rendered low in this bacterial host in order to prevent the vertebrate DNA in BACs from rearranging due to large amounts of repetitive sequences. The second approach introduced the recombination functions of phage λ, namely red α, red β and red γ into the DH10B host. It has the advantage of using shorter homologous sequences for recombination to introduce exogenous DNA cassettes into BACs [41]. Both methods have been widely used to engineer a variety of alterations in BAC DNA, such as introducing reporter gene cassettes in frame into the first exon of the target gene, mutating sequences at a chosen site, and introducing loxP sequences [42-52]. Thus BACs of ~200 kb size and containing the β-globin gene were functionalized with EGFP reporter gene using RedET-recombination and expressed in stable erythropoietic cell lines [47]. RecA mediated homologous recombination was used to introduce a LacZ reporter gene cassette into the Gdf6 transcription unit and long-range regulatory sequences mapped using a series of BAC deletions in transgenic mice [48]. Using RecA mediated homologous recombination, 97 mixed CAA-CAG repeats were introduced into the human htt gene in a BAC to build a mouse model of huntingtin's disease [51]. Transgenic mice carrying human BAC clones with 1 kb deletions made by RedET-recombination were used to functionally identify cell lineage-specific regulatory elements 30 kb upstream of the IFNG gene [52]. Recently a moderate throughput BAC modification procedure has been described that uses RedET recombination to introduce the GFP cassette and iTol2 for integration into zebrafish chromosomes [53]. Procedures for constructing transgenic animals with functionalized BACs in mice [54], or zebrafish [55], have also been reviewed recently.

b) Transposition followed by progressive deletions from BAC-ends

A different way to alter sequences in BAC DNA uses transposons. It does not require sequence homology between vector and genomic inserts of BACs to introduce exogenous DNA cassettes. Insertions of the bacterial transposon Tn10 can introduce exogenous DNA, including lox sites, at random locations in BACs [56,57]. Furthermore, site-specific recombination by the Cre-lox system can deliver these reporter genes and other exogenous DNA precisely to the ends of genomic DNA inserts in BACs [58-60]. Such DNA cassettes include sequencing primer binding sites, mammalian cell-selectable antibiotic resistance genes, enhancer-traps and sequences specifically recognized by the vertebrate transposon Tol2. It is significant that the recombinases involved in either of these procedures, Tn10-transposase and Cre protein respectively, do not act upon sequence repeats and/or other recombinogenic sites in the genomic DNA insert to rearrange it. This particular characteristic makes the approach applicable to a wider variety of BACs in the public domain, including those containing high levels of repetitive sequences [58]. Because DNA of higher vertebrates, especially mammalian DNA, contains high levels of sequence repeats the latter approach appears less susceptible to complications from intra-BAC DNA rearrangements.

Insertions of the Tn10 transposon into BAC DNA from a wide variety of vertebrate genomes appear to be random, demonstrating little sequence specificity for transposition [58]. It is unclear whether this lack of sequence specificity arises from the absence of selective pressure evolutionarily for insertions of a prokaryotic transposon into vertebrate DNA, because insertions of Tn10 into prokaryotic DNA have long been known to prefer a somewhat degenerate nevertheless consensus insertion site [61]. The minor sequence preferences for insertions of Tn10 observed in BACs probably have more to do with the accessibility of sites for Tn10 insertions than specificity for sequences. Note that *E. coli* has histone-like proteins, known as HU protein, which package the vertebrate DNA in the BAC and could generate differences in accessibility for transposon insertions. Transposition of Tn10 into a rare BAC clone occasionally displays apparent sequence selectivity [62]. However, this has been clearly shown to arise from a clonal selection process that arose from picking single colonies of Tn10 plasmid-transformed BAC clones that had induced the transposase gene prior to actual induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG). Induction of a large pool of Tn10 plasmid-transformed BAC colonies, instead of a single clone, has completely rectified this potential problem [62].

Deleting BAC ends with loxP-Tn10 transposons

The bacterial transposon Tn10 is used to introduce exogenous DNA sequences into the genomic insert of BACs as illustrated in Figure 2. A loxP sequence is placed within the 70 base inverted repeat ends of the Tn10 mini-transposon (Figure 2, Panels A and B).

Plasmid DNA carrying the transposon is introduced into the cell containing the BAC plasmid. Upon induction with IPTG the expressed transposase protein excises the DNA cassette flanked by the inverted repeats (shown as black and pink boxes L and R in the Tn10 plasmid) and inserts it into other nearby DNA in the bacterial host, including the bacterial genome and the BAC DNA (shown in Panel C). Insertions of Tn10 into BAC DNA occur in either orientation, and are irreversible because the transposase gene is left behind as linear DNA and destroyed. Important considerations that dictate subsequent steps
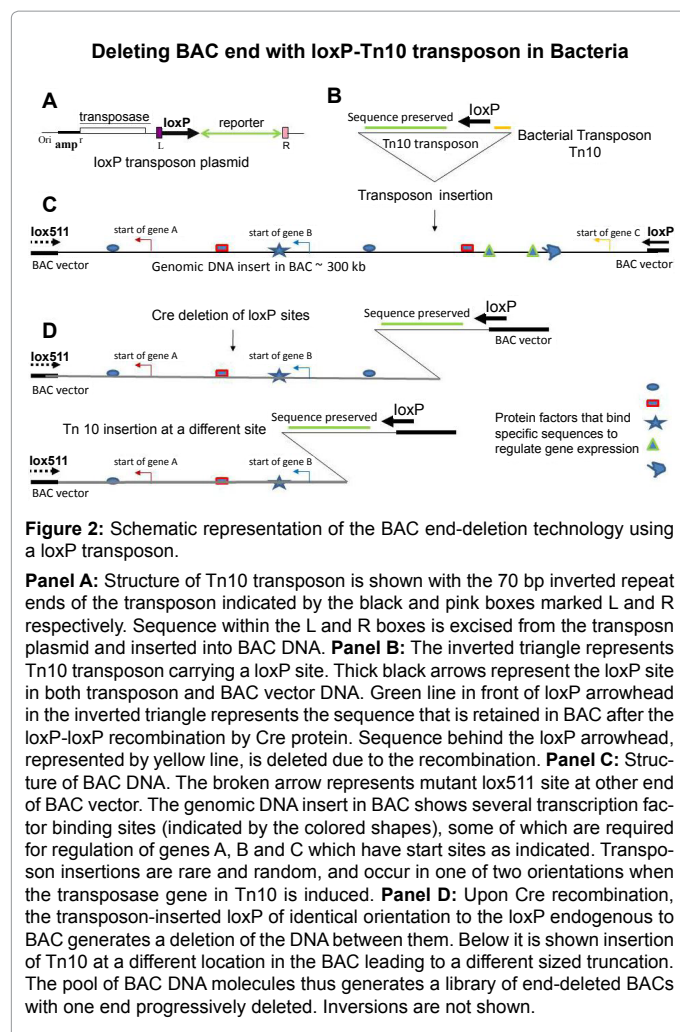
Figure 2: Schematic representation of the BAC end-deletion technology using a loxP transposon.

Panel A: Structure of Tn10 transposon is shown with the 70 bp inverted repeat ends of the transposon indicated by the black and pink boxes marked L and R respectively. Sequence within the L and R boxes is excised from the transposn plasmid and inserted into BAC DNA. **Panel B:** The inverted triangle represents Tn10 transposon carrying a loxP site. Thick black arrows represent the loxP site in both transposon and BAC vector DNA. Green line in front of loxP arrowhead in the inverted triangle represents the sequence that is retained in BAC after the loxP-loxP recombination by Cre protein. Sequence behind the loxP arrowhead, represented by yellow line, is deleted due to the recombination. **Panel C:** Structure of BAC DNA. The broken arrow represents mutant lox511 site at other end of BAC vector. The genomic DNA insert in BAC shows several transcription factor binding sites (indicated by the colored shapes), some of which are required for regulation of genes A, B and C which have start sites as indicated. Transposon insertions are rare and random, and occur in one of two orientations when the transposase gene in Tn10 is induced. **Panel D:** Upon Cre recombination, the transposon-inserted loxP of identical orientation to the loxP endogenous to BAC generates a deletion of the DNA between them. Below it is shown insertion of Tn10 at a different location in the BAC leading to a different sized truncation. The pool of BAC DNA molecules thus generates a library of end-deleted BACs with one end progressively deleted. Inversions are not shown.

of the procedure include 1) damage to the host genome from Tn10 insertions, and 2) the fraction of BACs actually modified is relatively low ~1 in 10,000. Efficient steps are thus required to recover the few BACs with Tn insertions and transfer these into a new host. Both these challenges are met by packaging the Tn10-modified BAC DNA in phage P1 heads. Therefore cells containing the BAC DNA with Tn10 insertions are infected with P1 phage after induction with IPTG.

An additional purpose is served by the P1 phage Infection. The phage expresses Cre protein early in its life cycle to circularize the linear double-stranded phage DNA existing in the P1 head. Cre protein acts in *trans* to also recombine the loxP site transposed into the BAC DNA insert with the loxP endogenous to BAC vector and located at one end of insert DNA (Figure 2, Panel D). Thus the transposed loxP of one orientation produces a deletion from one end of genomic insert, while the loxP inserted in the opposite orientation simply inverts the DNA between it and the one endogenous to the BAC [63].

Analyses of retrofitted/end-deleted BACs

BACs deleted from an end and/or inserted with exogenous DNA cassettes at the newly created ends of genomic inserts, are analyzed by isolating their DNA and separating them by Field Inversion Gel Electrophoresis (FIGE). The DNA in BAC clones is isolated by simple mini-prep procedures, digested with Not I enzyme and analyzed by

FIGE [63,64]. High throughput formats for preparing DNA from BAC deletions in 96-well plates, suitable for subsequent FIGE analyses and end sequencing, have also been described [57,64].

Truncations from the opposite end of insert DNA in BACs

Genomic DNA in BACs is flanked by loxP and mutant lox511 sites [29,65]. Genome libraries in the public domain of numerous organisms such as the mouse, rat, human and zebrafish have used the vector pBACe3.6 or its derivatives, link for pTARBAC2.1 vector (http://bacpac.chori. org/ptarbac21. htm), and all share this characteristic [30,66-68]. Thus end-deletions specifically from the lox511 side of genomic insert were made using a similar approach with lox511-Tn10 mini-transposons [65].

Cross- recombination between loxP and lox511 sites does not occur with Cre expressed by phage P1 infection

The 34 bp sequence of mutant lox511 differs from the wild type loxP by one nucleotide in the spacer region (for lox511 sequence) [65,69]. Previous studies had reported recombination between loxP and lox511 to occur at efficiencies ranging from 5 to 100% under experimental conditions that express Cre constitutively [70-75]. Such cross-recombination was not observed when Cre protein was expressed in a phage P1 infection *in vivo* [65,76,77], and high levels of stringency in recombining identical lox sites [65,76], or lox sites with at least identical spacers was achieved [77]. Truncations of genomic DNA in a BAC from either end are not only efficient, but are highly specific for that end.

Introducing Exogenous DNA cassettes precisely at one or both newly created BAC end(s)

The end-deletion technology can introduce reporter genes and other DNA cassettes precisely at the new end created in the BAC. Sequence in front of the loxP arrow, shown in green in Panels B and D of Figure 2, is retained after the recombination event that creates the deletion, while that shown in yellow (Panel B, Figure 2) is eliminated. Note that orientation of arrow refers to directionality of loxP sequence. This characteristic has been used to target a variety of DNA cassettes to BAC ends, such as: i) mammalian cell-selectable antibiotic resistance genes [57,58], ii) a basal promoter-containing EGFP gene to serve as an enhancer-trap to functionally locate potential gene-regulatory elements further upstream [59], and iii) introduce end sequences of the vertebrate transposn iTol2 to generate integration-ready GFP-functionalized BAC DNA for zebrafish or mouse transgenesis studies [60,78].

The ability to truncate either end of a BAC insert progressively while keeping the other end intact has been very useful to a variety of mapping studies. Thus genetic markers and polymorphisms have been mapped on physical maps of chromosomes using the technology [57,58,79]. Long-range regulatory elements of the mouse Nkx2-5 gene were mapped functionally using end-deleted EGFP functionalized BACs to generate transgenic mice [80].

Introducing Reporter Gene and Truncating BAC End Simultaneously: Enhancer-trapping using BACs

Truncation of sequence downstream of a gene while introducing the EGFP reporter can be achieved simultaneously in BACs using lox-Tn10 transposons. Potential regulatory sequences identified upstream of the gene this was can then be conformed by truncating BAC DNA from the opposite end using a lox511 Tn10 transposon. These strategies have been successfully used to decipher the role of sequences in two
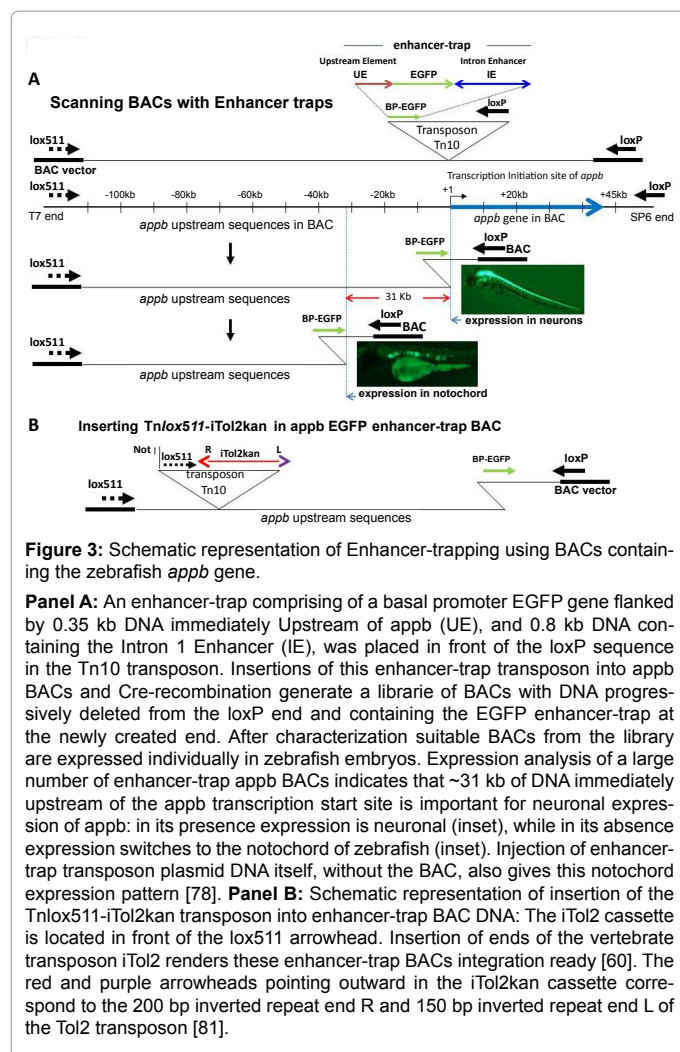
Figure 3: Schematic representation of Enhancer-trapping using BACs containing the zebrafish *appb* gene.

Panel A: An enhancer-trap comprising of a basal promoter EGFP gene flanked by 0.35 kb DNA immediately Upstream of appb (UE), and 0.8 kb DNA containing the Intron 1 Enhancer (IE), was placed in front of the loxP sequence in the Tn10 transposon. Insertions of this enhancer-trap transposon into appb BACs and Cre-recombination generate a librarie of BACs with DNA progressively deleted from the loxP end and containing the EGFP enhancer-trap at the newly created end. After characterization suitable BACs from the library are expressed individually in zebrafish embryos. Expression analysis of a large number of enhancer-trap appb BACs indicates that ~31 kb of DNA immediately upstream of the appb transcription start site is important for neuronal expression of appb: in its presence expression is neuronal (inset), while in its absence expression switches to the notochord of zebrafish (inset). Injection of enhancer-trap transposon plasmid DNA itself, without the BAC, also gives this notochord expression pattern [78]. **Panel B:** Schematic representation of insertion of the Tnlox511-iTol2kan transposon into enhancer-trap BAC DNA: The iTol2 cassette is located in front of the lox511 arrowhead. Insertion of ends of the vertebrate transposon iTol2 renders these enhancer-trap BACs integration ready [60]. The red and purple arrowheads pointing outward in the iTol2kan cassette correspond to the 200 bp inverted repeat end R and 150 bp inverted repeat end L of the Tol2 transposon [81].

non-contiguous domains of the zebrafish *appb* gene in regulating its tissue-specific expression [59,78]. A basal promoter-containing EGFP gene was designed to serve as an enhancer-trap, and placed in front of the loxP arrowhead such that the cassette was retained after the end-deletion (Panel A, Figure 3). Deleting progressively from one end of BAC insert containing the *appb* gene with this enhancer-trap transposon (the right hand end in this case) generated a library of BACs that differ in location of the enhancer-trap with respect to the start site of transcription of the gene. The collection of well characterized BAC DNA was introduced individually into zebrafish embryos for expression of the gene in specific tissues.

Once a distal DNA domain at -31 kb had been tentatively identified as enhancing transcription of the *appb* gene, progressive truncations from the opposite end were made with a lox511 transposon to confirm the finding [78]. Truncation from the opposite end with the lox511 transposon and insertion of the ends of the vertebrate transposon iTol2, required for chromosomal integration of the BAC DNA [60,81], were done simultaneously as illustrated in Figure 3B.

Integrating BAC DNA into the germline of zebrafish and mice using Tol2

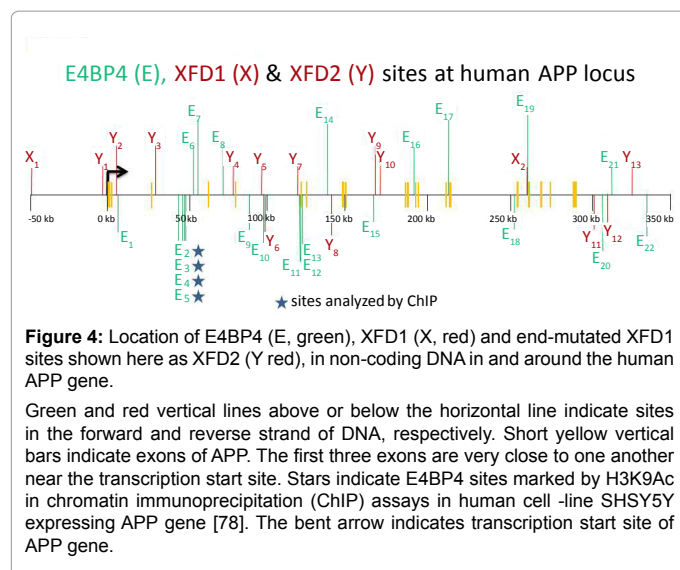Germline integration of BAC DNA in zebrafish was achieved

using Tol2 [60,78,81]. Originally isolated from medaka fish, Tol2 is a vertebrate transposon system that functions effectively in zebrafish and mice [81]. Traditionally, Tol2 ends in the inverted orientation and flanking a 1 kb spacer DNA (iTol2) were introduced into BAC DNA using recombination of homologous sequences between targeting vector and the genomic insert [81]. To avoid potential complications from rearrangement of highly repetitive DNA that occurs in certain BACs from higher vertebrate genome libraries (see for example the mouse Npr3 gene locus analyzed earlier), a simpler and more widely applicable approach was developed using Tn10 transposons to deliver iTol2 ends into BACs [58,60]. The iTol2kan DNA is inserted by a lox511Tn10 transposon to the lox511 end of the enhancer-trap BAC as shown schematically in Figure 3B. A large collection of end-deleted integration-ready BACs is generated in a single experiment. Such a collection is highly desirable for mapping distal regulatory sequences of a gene functionally by expressing individual members in zebrafish or mouse embryos [78].

Mechanistic dissection of amyloid precursor protein (*appb*) gene regulation in zebrafish

Application of several components of the Tn10 transposon BAC retrofitting technology described above to functionally identify distal non-contiguous regulatory elements of the zebrafish *appb* gene is illustrated above. Zebrafish has gained popularity as a model system in recent years due to a variety of reasons, which include optical transparency of its embryo. This is despite the fact that it is further removed from humans than mice evolutionarily. The model has been used to explore the function of highly conserved non-coding DNA in vertebrates, found tens of thousands of base-pairs removed from the start sites of genes [7-9,21]. However, functional non-coding DNA is sometimes not conserved in many developmentally regulated genes including *appb* [22-26,59]. Identifying regulatory DNA domains in the *appb* gene has therefore been a challenge. The enhancer-trap approach is particularly useful in such circumstances because it allows one to scan a series of overlapping BACs to identify gene-regulatory sites without having to guess potential regulatory sites to test. Thus, using only a basal promoter driven EGFP gene in a loxP-Tn10 enhancer-trap transposon to scan *appb* gene-containing BACs the requirement of an intron enhancer was established initially. In the absence of this enhancer there is no expression of a BAC transgene that contained approximately 100 kb of 5' sequences [59]. It is an important necessary condition to establish for the existence of an enhancer downstream of the transcription start site of *appb* gene. Subsequent enhancer-traps were therefore built to include the minimal intron 1 enhancer element.

Binding sites of Transcription factor E4BP4/NFIL3 required for regulating *appb*

Enhancer-traps built with the minimal intron 1 enhancer expressed *appb* in a pattern similar to the endogenous gene in neurons of zebrafish [59]. These experiments highlighted the importance of a second regulatory region located between 28 and 31 kb upstream of the gene for guiding *appb* expression to neurons. Deleting this upstream region shifted the expression pattern from being neuron-specific to notochord-specific, which is the default pattern observed with the basal promoter plus intron enhancer combination. Consistent with this finding expression of GFP in neurons of enhancer-trap BACs with a full complement of upstream sequences reverted to specific expression in the notochord when the three E4BP4 sites at -31 kb of zebrafish *appb* was deleted with a Tnlox511-iTol2kan transposon from the opposite

Figure 4: Location of E4BP4 (E, green), XFD1 (X, red) and end-mutated XFD1 sites shown here as XFD2 (Y red), in non-coding DNA in and around the human APP gene.

Green and red vertical lines above or below the horizontal line indicate sites in the forward and reverse strand of DNA, respectively. Short yellow vertical bars indicate exons of APP. The first three exons are very close to one another near the transcription start site. Stars indicate E4BP4 sites marked by H3K9Ac in chromatin immunoprecipitation (ChIP) assays in human cell -line SHSY5Y expressing APP gene [78]. The bent arrow indicates transcription start site of APP gene.

end of BAC DNA (Figure 3). These upstream sites also bound the E4BP4 DNA binding protein domain, expressed in *E. coli*, efficiently and selectively in vitro [78].

Expression of enhancer-trap BACs containing mutated intron 1 enhancers was analyzed next. A select group of enhancer-trap BACs from 18 libraries, constructed with mutated intron 1 enhancers was expressed in zebrafish. While generating mutations in the small Tn-plasmid is rapid, creating and characterizing an enhancer-trap BAC deletion library with the mutant transposon is not. It takes about a month by an individual to get a set of four BAC libraries done [63]. Expression analyses of selected clones from these 18 libraries indicated binding sites of at least two known transcription factors are important for function. They are the clock-regulated immune system modulator transcription factor E4BP4/NFIL3 and members of the Forkhead gene family (XFD1). A search of non-coding DNA in introns and the 50 kb sequence surrounding the *appb* gene for additional binding sites revealed a ~ 8-fold and ~ 11-fold greater frequency than statistical of E4BP4 and XFD1 sites, respectively. Although comparison of zebrafish and human APP did not reveal substantially conserved non-coding sequences that could represent regulatory elements, conservation of gene expression via the use of the same transcription factors was hypothesized and explored. Over representation of E4BP4 binding sites at the human APP locus supports the proposal, as seen in Figure 4. Strikingly, one such cluster of four E4BP4 sites in the fourth intron of the human APP gene was marked by a peak of acetylated histones in a human neuroblastoma cell line that expressed APP [78]. Thus it appears that E4BP4/NFIL3 may regulate human APP expression via binding to distal regulatory sequences. Higher than statistical frequency of E4BP4/NFIL3 sites have also been noted in both mouse (13 sites), and rat (10 sites) APP (unpublished observations).

Transcription factor E4BP4/NFIL3 is clock regulated and intricately linked with the immune system, where it is required for protecting Natural Killer (NK) T cells and regulates IL-12 p40 in macrophages [82-85]. If it were to also regulate human APP expression then it would help us understand in molecular terms not only the circadian pattern observed for β-amyloid levels in mice brains, but also the importance of immunological and inflammatory processes that underlie the onset of Alzheimer's Disease in humans [86,87]. These characteristics have

been deduced thus far only from the pathology and physiology of Alzheimer's patients.

Regulation of *appb* in Zebrafish and APP in Humans by the Forkhead family of transcription factors

Earlier studies had indicated certain members of the forkhead family of transcription factors, (fkd1, fkd2 and fkd4), are expressed during gastrulation in the zebrafish, with high levels of fkd1 and fkd4 mRNA accumulating exclusively in the notochord during somitogenesis [88]. It suggested that the fkd1, fkd2 and fkd4 proteins are available only in the notochord and could explain the expression of EGFP exclusively in the notochord when the cluster of three E4BP4/NFIL3 sites at -31 kb are deleted in the enhancer-trap *appb*-BACs [78]. A suppressor role for E4BP4/NFIL3 is consistent with this scenario and supported by recent reports [82,89]. The forkhead binding site in intron 1 enhancer required for function and the much higher than statistical frequency of forkhead sites in and around the *appb* gene in zebrafish suggest a regulatory role for this family of transcription factors [78].

Over-representation of fkd sites is however absent in human APP: although there are only two sites with the XFD1 consensus sequence, sites with 8 of 9 bases identical (consecutively) to the consensus exist far more abundantly in human APP (13 additional such sites were identified, designated as XFD2 and shown below in Figure 4). It is tempting to speculate that Fkd protein complexes potentially capable of binding to such sites might have evolved in higher vertebrates to accommodate the single end-nucleotide change. We are currently assessing this hypothesis.

Green and red vertical lines above or below the horizontal line indicate sites in the forward and reverse strand of DNA, respectively. Short yellow vertical bars indicate exons of APP. The first three exons are very close to one another near the transcription start site. Stars indicate E4BP4 sites marked by H3K9Ac in chromatin immunoprecipitation (ChIP) assays in human cell -line SHSY5Y expressing APP gene [78]. The bent arrow indicates transcription start site of APP gene.

A novel interplay between Fkd and E4BP4/NFIL3 is proposed to restrict expression of *appb* to neurons of zebrafish, either directly or indirectly through other proteins. The findings, taken together, would suggest that *appb* in zebrafish and APP in humans may follow the same regulatory logic using the same set of transcription factors, despite a lack of sequence similarity in their regulatory DNA. It suggests potential regulatory pathways for the human APP gene may be discovered not on the basis of comparing DNA primary sequences with zebrafish *appb*, but on the model of conservation of transcription factors.

Unique advantages of using enhancer-trap BACs for exploring gene regulation by distal *cis*-acting discontinuous DNA Domains

Because BACs represent tiny pieces of the chromosome, they offer a large enough contiguous span of DNA to house most of the sequences required to regulate expression of a gene. But more importantly, the protein-binding sites in the BAC DNA exist in their chromosomal contexts, ie. sequences flanking important transcription factor binding sites, which have evolved for millions of years, are preserved. Therefore a general approach to retrofit BACs to equip them with enhancer-traps was developed such that *cis*-acting regulatory sequences could be easily identified by expressing a series of end-truncated BACs in animals [59]. Because all members of the deletion library contain identical changes at the loxP or lox511 end, comparisons of expression patterns between any two enhancer-trap BACs from the same deletion series are more

meaningful as the modifications in each remain constant. Thus BACs can be "scanned" with enhancer-traps to identify distal regulatory elements; and this approach is likely to be less biased than other methods, where mutations are targeted to specific sites in the BAC, because there is no need to select sequences to test their regulatory potential. In situations where regulatory function is conserved without sequence similarity [22-26], choosing the correct sequence to test might be very difficult as many of the players involved in fine tuning the regulation of gene expression remain unknown.

The BAC enhancer-trap technology has three distinctive features that should prove beneficial: i) much larger sequences of DNA can be probed, including multiple regulatory domains that are discontinuous, and often act together to regulate expression of higher vertebrate genes such as *appb*. Note that the enhancer sequences in intron 1 and the -31 kb upstream region in *appb* are separated by some 35 kb of DNA [78]. Gene regulation of this type would be far more tedious to analyze using other BAC recombineering approaches. ii) is less biased compared to targeted approaches as noted above, and iii) compared to traditional enhancer-trapping that use whole genomes of animals, our approach is not affected by genome accessibility issues of the trap during the one-cell stage of the embryo. Traditional methods for enhancer-trapping have used either a retroviral vector or other vertebrate transposable element to insert the trap directly into whole genomes of animals, or integrate test-sequences cloned in a small vector into the germline using a vertebrate transposon such as Tol2 [10,14,31-39]. Using test-sequences is a "targeted" approach, and thus not free of the investigators' biases. Additionally, the traditional methods require the enhancer-trap be introduced into the fertilized egg at the one-cell stage for germline expression. Some regions of the genome are more accessible than others for insertion of the enhancer-trap at this stage of development of the egg, and because one cannot have multiple insertions of the same trap for an unambiguous analyses, most of the genome remains unreachable. Consequently enhancers have been identified in only a very small percentage of the genome in animals studied, and vast regions appear refractory to enhancer-trapping by traditional means. By contrast, the Tn10 approach requires inserting the trap in isolated pieces of chromosomal DNA in BACs in the bacterial host. Therefore a much greater fraction, probably all, of the genome is potentially accessible for study because to date BACs refractory to Tn10 transposon insertions have not been encountered. Although the functional context of transcription factor binding sites identified this way is limited to the size of BACs, using enhancer-traps to scan individual BACs allows a more uniform coverage of the genome. This is because the baseline efficiency of trap insertion is reset for individual BACs in the bacterial host. Additionally, mutations of transcription factor binding sites in the BAC can be more easily introduced via a small Tn10 transposon plasmid, which can then be inserted into BAC DNA as demonstrated during functionally dissecting the intron 1 enhancer of *appb* gene [78].

However when comparing the enhancer-trap approach as used in BACs to that when used in the entire genome, such as in whole-genome screens, it is probably important to also remember that the goals in the two are usually different. With a BAC the goal is usually to generate a specific reporter for a gene of interest or to functionally dissect regulatory elements for a specific gene. The whole-genome approach on the other hand is usually for high-throughput screening *in vivo* to recover numerous useful enhancer-trap reporter lines.

Acknowledgement

References

1. Ulanovsky L, Bodner M, Trifonov EN, Choder M (1986) Curved DNA: design, synthesis, and circularization. Proc Natl Acad Sci U S A 83: 862-866.

2. Fried MG, Crothers DM (1983) CAP and RNA polymerase interactions with the lac promoter: binding stoichiometry and long range effects. Nucleic Acids Res 11: 141-158.

3. Gartenberg MR, Ampe C, Steitz TA, Crothers DM (1990) Molecular characterization of the GCN4-DNA complex. Proc Natl Acad Sci U S A 87: 6034-6038.

4. Dalma-Weiszhausz DD, Gartenberg MR, Crothers DM (1991) Sequence-dependent contribution of distal binding domains to CAP protein-DNA binding affinity. Nucleic Acids Res 19: 611-616.

5. Ganapathi M, Singh GP, Sandhu KS, Brahmachari SK, Brahmachari V (2007) A whole genome analysis of 5' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. Gene 391: 242-251.

6. Hebbar PB, Archer TK (2007) Chromatin-dependent cooperativity between site-specific transcription factors in vivo. J Biol Chem 282: 8284-8291.

7. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444: 499-502.

8. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3: e7.

9. Shin JT, Priest JR, Ovcharenko I, Ronco A, Moore RK, et al. (2005) Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. Nucleic Acids Res 33: 5437-5445.

10. Nagayoshi S, Hayashi E, Abe G, Osato N, Asakawa K, et al. (2008) Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: tcf7 and synembryn-like. Development 135: 159-169.

11. Lodish H, Berk A, Kaiser C, Krieger M, Scott M, et al. (2008) Molecular Cell Biology. (6thedn), WH Freeman, NY, USA.

12. Bronson SK, Plaehn EG, Kluckman KD, Hagaman JR, Maeda N, et al. (1996) Single-copy transgenic mice with chosen-site integration. Proc Natl Acad Sci U S A 93: 9067-9072.

13. Giraldo P, Montoliu L (2001) Size matters: use of YACs, BACs and PACs in transgenic animals. Transgenic Res 10: 83-103.

14. Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, et al. (2009) Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev Dyn 238: 2409-2417.

15. Sauer B, Henderson N (1990) Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. New Biol 2: 441-449.

16. Albert H, Dale EC, Lee E, Ow DW (1995) Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J 7: 649-659.

17. Araki K, Araki M, Yamamura K (1997) Targeted integration of DNA using mutant lox sites in embryonic stem cells. Nucleic Acids Res 25: 868-872.

18. Liu WY, Wang Y, Qin Y, Wang YP, Zhu ZY (2007) Site-directed gene integration in transgenic zebrafish mediated by cre recombinase using a combination of mutant lox sites. Mar Biotechnol (NY) 9: 420-428.

19. Mosimann C, Puller AC, Lawson KL, Tschopp P, Amsterdam A, et al. (2013) Site-directed zebrafish transgenesis into single landing sites with the phiC31 integrase system. Dev Dyn 242: 949-963.

20. Kirchmaier S, Höckendorf B, Möller EK, Bornhorst D, Spitz F, et al. (2013) Efficient site-specific transgenesis and enhancer activity tests in medaka using PhiC31 integrase. Development 140: 4287-4295.

21. Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O (2005) Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. Hum Mol Genet 14: 3057-3063.

22. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 14: 276-279.

23. Parker SC, Hansen L, Abaan HO, Tullius TD, Margulies EH (2009) Local DNA topography correlates with functional noncoding regions of the human genome. Science 324: 389-392.

24. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, et al. (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42: 806-810.

25. Kague E, Bessling SL, Lee J, Hu G, Passos-Bueno MR, et al. (2010) Functionally conserved cis-regulatory elements of COL18A1 identified through zebrafish transgenesis. Dev Biol 337: 496-505.

26. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, et al. (2011) Genome-wide identification of conserved regulatory function in diverged sequences. Genome Res 21: 1139-1149.

27. Chatterjee S, Lufkin T (2012) Regulatory genomics: Insights from the zebrafish. Curr Top Genet 5: 1-10.

28. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, et al. (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.

29. Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, et al. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52: 1-8.

30. Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, et al. (1999) A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics 58: 250-253.

31. O'Kane CJ, Gehring WJ (1987) Detection in situ of genomic regulatory elements in Drosophila. Proc Natl Acad Sci U S A 84: 9123-9127.

32. Wilson C, Pearson RK, Bellen HJ, O'Kane CJ, Grossniklaus U, Gehring WJ (1989) P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. Genes Dev 3: 1301-1313.

33. Korn R, Schoor M, Neuhaus H, Henseling U, Soininen R, et al. (1992) Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. Mech Dev 39: 95-109.

34. Burns JC, Friedmann T, Driever W, Burrascano M, Yee JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. Proc Natl Acad Sci U S A 90: 8033-8037.

35. Lin S, Gaiano N, Culp P, Burns JC, Friedmann T, et al. (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. Science 265: 666-669.

36. Grabher C, Henrich T, Sasado T, Arenz A, Wittbrodt J, et al. (2003) Transposon-mediated enhancer trapping in medaka. Gene 322: 57-66.

37. Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB, Welle Z, et al. (2004) Enhancer trapping in zebrafish using the Sleeping Beauty transposon. BMC Genomics 5: 62.

38. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, et al. (2004) A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev Cell 7: 133-144.

39. Ellingsen S, Laplante MA, König M, Kikuta H, Furmanek T, et al. (2005) Large-scale enhancer detection in the zebrafish genome. Development 132: 3799-3811.

40. Yang XW, Model P, Heintz N (1997) Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. Nat Biotechnol 15: 859-865.

41. Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet 20: 123-128.

42. Jessen JR, Meng A, McFarlane RJ, Paw BH, Zon LI, et al. (1998) Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. Proc Natl Acad Sci U S A 95: 5121-5126.

43. Muyrers JP, Zhang Y, Testa G, Stewart AF (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res 27: 1555-1557.

44. Gong S, Yang XW, Li C, Heintz N (2002) Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. Genome Res 12: 1992-1998.

45. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33: e36.

46. Yang Z, Jiang H, Chaichanasakul T, Gong S, Yang XW, et al. (2006) Modified bacterial artificial chromosomes for zebrafish transgenesis. Methods 39: 183-188.

47. Orford M, Nefedov M, Vadolas J, Zaibak F, Williamson R, et al. (2000) Engineering EGFP reporter constructs into a 200 kb human beta-globin BAC clone using GET Recombination. Nucleic Acids Res 28: E84.

48. Mortlock DP, Guenther C, Kingsley DM (2003) A general approach for identifying distant regulatory elements applied to the Gdf6 gene. Genome Res 13: 2069-2081.

49. Jessen JR, Willett CE, Lin S (1999) Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish. Nat Genet 23: 15-16.

50. Carvajal JJ, Cox D, Summerbell D, Rigby PW (2001) A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. Development 128: 1857-1868.

51. Gray M, Shirasaki DS, Cepeda C, Andre VM, Wilburn B, et al. (2008) Full length human mutant huntington with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. J Neurosci 28: 6182-6195.

52. Collins PL, Chang S, Henderson M, Soutto M, Davis GM, et al. (2010) Distal regions of the human IFNG locus direct cell type-specific expression. J Immunol 185: 1492-1501.

53. Bussmann J, Schulte-Merker S (2011) Rapid BAC selection for tol2-mediated transgenesis in zebrafish. Development 138: 4327-4332.

54. Fu Y, Maye P (2011) Engineering BAC reporter gene constructs for mouse transgenesis. Methods Mol Biol 693: 163-179.

55. Suster ML, Abe G, Schouw A, Kawakami K (2011) Transposon-mediated BAC transgenesis in zebrafish. Nat Protoc 6: 1998-2021.

56. Chatterjee PK, Coren JC (1997) Isolating large nested deletions in bacterial and P1 artificial chromosomes by in vivo P1 packaging of products of Cre-catalyzed recombination between the endogenous and a transposed loxP site. Nuc Acids Res 25: 2205-2212.

57. Chatterjee PK, Yarnall DP, Haneline SA, Godlevski MM, Thornber SJ, et al. (1999) Direct Sequencing of Bacterial and P1 Artificial Chromosome Nested-deletions for Identifying Position-Specific Single Nucleotide Polymorphisms. Proc Natl Acad Sci U S A 96: 13276-13281.

58. Gilmore RC, Baker Jr J, Dempsey S, Marchan R, Corprew Jr, RNL, et al. (2001) Using PAC nested-deletions to order contigs and microsatellite markers at the high repetitive sequence containing Npr3 gene locus. Gene 275: 65-72.

59. Shakes LA, Malcolm TL, Allen KL, De S, Harewood KR, et al. (2008) Context dependent function of APPb Enhancer identified using Enhancer Trap-containing BACs as Transgenes in Zebrafish. Nucleic Acids Res 36: 6237-6248.

60. Shakes LA, Abe G, Eltayeb MA, Wolf HM, Kawakami K, et al. (2011) Generating libraries of iTol2-end insertions at BAC ends using loxP and lox511 Tn10 transposons. BMC Genomics 12: 351.

61. Craig NL (1997) Target site selection in transposition. Annu Rev Biochem 66: 437-474.

62. Chatterjee PK, Briley LP (2000) Analysis of a Clonal Selection Event during Transposon-Mediated Nested-Deletion Formation in Rare BAC and PAC Clones. Analytical Biochemistry 285: 121-126.

63. Chatterjee PK (2014) Directing Enhancer-traps and iTol2 End Sequences to Deleted BAC ends with loxP- and lox511-Tn10 transposons. In: Kumaran Narayanan (ed.) Bacterial Artificial Chromosomes Methods in Molecular Biology, Humana Press/Springer.

64. Chatterjee PK, Baker Jr JC (2004) Preparing Nested Deletions Template DNA for Field Inversion Gel Electrophoresis Analyses and Position-Specific End Sequencing With Transposon Primers. "Bacterial Artificial Chromosomes" 243-254.

65. Shakes LA, Garland DM, Srivastava DK, Harewood KR, Chatterjee PK (2005) Minimal Cross-recombination between wild type and loxP511 sites in vivo facilitates Truncating Both Ends of Large DNA Inserts in pBACe3.6 and Related Vectors. Nuc Acids Res 33: e118.

66. Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, et al. (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome Res 10: 116-128.

67. Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, et al. (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 11: 483-496.

68. Osoegawa K, Zhu B, Shu CL, Ren T, Cao Q, et al. (2004) BAC resources for the rat genome project. Genome Res 14: 780-785.

69. Hoess RH, Wierzbicki A, Abremski K (1986) The role of the loxP spacer region in P1 site-specific recombination. Nucleic Acids Res 14: 2287-2300.

70. Lee G, Saito I (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. Gene 216: 55-65.

71. Missirlis PI, Smailus DE, Holt RA (2006) A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. BMC Genomics 7: 73.

72. Langer SJ, Ghafoori AP, Byrd M, Leinwand L (2002) A genetic screen identifies novel non-compatible loxP sites. Nucleic Acids Res 30: 3067-3077.

73. Sheren J, Langer SJ, Leinwand LA (2007) A randomized library approach to identifying functional lox site domains for the Cre recombinase. Nucleic Acids Res 35: 5464-5473.

74. Siegel RW, Jain R, Bradbury A (2001) Using an in vivo phagemid system to identify non-compatible loxP sequences. FEBS lett 505: 467-473.

75. Wang Z, Engler P, Longacre A, Storb U (2001) An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. Genome Res 11: 137-142.

76. Chatterjee PK, Shakes LA, Srivastava DK, Garland DM, Harewood KR, et al. (2004) Mutually Exclusive Recombination of Wild Type and Mutant loxP Sites in vivo Facilitates Transposon-Mediated Deletions from Both Ends of Genomic DNA in PACs. Nuc Acids Res 32: 5668-5676.

77. Chatterjee PK, Shakes LA, Stennett N, Richardson VL, Malcolm TL, et al. (2010) Replacing the wild type loxP site in BACs from the public domain with lox66 using a lox66 transposon. BMC Res Notes 3: 38.

78. Shakes LA, Du H, Wolf HM, Hatcher C, Norford DC, et al. (2012) Using BAC transgenesis in zebrafish to identify regulatory sequences of the amyloid precursor protein gene in humans. BMC Genomics 13: 451.

79. Brake RL, Chatterjee PK, Kees UR, Watt PM (2004) The functional mapping of long-range transcription control elements of the HOX11 proto-oncogene. Biochem Biophys Res Commun 313: 327-335.

80. Chi X, Chatterjee PK, Wilson W 3rd, Zhang SX, Demayo FJ, et al. (2005) Complex cardiac Nkx2-5 gene expression activated by noggin-sensitive enhancers followed by chamber-specific modules. Proc Natl Acad Sci U S A 102: 13490-13495.

81. Suster ML, Sumiyama K, Kawakami K (2009) Transposon-mediated BAC transgenesis in zebrafish and mice. BMC Genomics 10: 477.

82. Cowell IG (2002) E4BP4/NFIL3, a PAR-related bZIP factor with many roles. Bioessays 24: 1023-1029.

83. Bozek K, Relógio A, Kielbasa SM, Heine M, Dame C, et al. (2009) Regulation of clock-controlled genes in mammals. PLoS One 4: e4882.

84. Kamizono S, Duncan GS, Seidel MG, Morimoto A, Hamada K, et al. (2009) Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo. J Exp Med 206: 2977-2986.

85. Kobayashi T, Matsuoka K, Sheikh SZ, Elloumi HZ, Kamada N, et al. (2011) NFIL3 is a regulator of IL-12 p40 in macrophages and mucosal immunity. J Immunol 186: 4649-4655.

86. Kang J-E, Lim MM, Bateman RJ, Lee JJ, Smyth LP, et al. (2009) Amyloid-ß Dynamics Are Regulated by Orexin and the Sleep-Wake Cycle. Science 326: 1005-1007.

87. Popović M, Caballero-Bleda M, Puelles L, Popović N (1998) Importance of immunological and inflammatory processes in the pathogenesis and therapy of Alzheimer's disease. Int J Neurosci 95: 203-236.

88. Odenthal J, Nüsslein-Volhard C (1998) fork head domain genes in zebrafish. Dev Genes Evol 208: 245-258.

89. Smith AM, Qualls JE, O'Brien K, Balouzian L, Johnson PF, et al. (2011) A distal enhancer in Il12b is the target of transcriptional repression by the STAT3 pathway and requires the basic leucine zipper (B-ZIP) protein NFIL3. J Biol Chem 286: 23582-23590.