Published in final edited form as:

RSC Adv. 2013 June 21; 3(23): 8604–8617. doi:10.1039/C3RA40332G.

Identifying Distal *cis*-acting Gene-Regulatory Sequences by Expressing BACs Functionalized with loxP-Tn10 Transposons in Zebrafish

Pradeep K. Chatterjee^{1,*}, Leighcraft A. Shakes¹, Hope M. Wolf^{1,3}, Mohammad A. Mujalled¹, Constance Zhou¹, Charles Hatcher¹, and Derek C. Norford¹

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

Abstract

Bacterial Artificial Chromosomes (BACs) are large pieces of DNA from the chromosomes of organisms propagated faithfully in bacteria as large extra-chromosomal plasmids. Expression of genes contained in BACs can be monitored after functionalizing the BAC DNA with reporter genes and other sequences that allow stable maintenance and propagation of the DNA in the new host organism. The DNA in BACs can be altered within its bacterial host in several ways. Here we discuss one such approach, using Tn10 mini-transposons, to introduce exogenous sequences into BACs for a variety of purposes. The largely random insertions of Tn10 transposons carrying lox sites have been used to position mammalian cell-selectable antibiotic resistance genes, enhancer-traps and inverted repeat ends of the vertebrate transposon Tol2 precisely at the ends of the genomic DNA insert in BACs. These modified BACs are suitable for expression in zebrafish or mouse, and have been used to functionally identify important long-range gene regulatory sequences in both species. Enhancer-trapping using BACs should prove uniquely useful in analyzing multiple discontinuous DNA domains that act in concert to regulate expression of a gene, and is not limited by genome accessibility issues of traditional enhancer-trapping methods.

The case for using BACs for accurate expression of genes in higher vertebrates

DNA sequences that act in *cis* to regulate expression of a gene do so mostly by binding specific proteins to those sequences. Repeated sequences, such as di- and tri-nucleotide repeats and poly purine-pyrimidine tracts, found occasionally in the vicinity of gene control regions have been implicated in gene regulation [1-4]. Such sequences are often capable of adopting structures different from the regular double helix and shown to regulate

* corresponding author tel. # (919) 530-7017 fax # (919) 530-7998 email pchatterjee@nccu.edu.

³Present address: Department of Chemistry, University of North Carolina at Chapel-Hill, Chapel-Hill, NC 27599. USA

Leighcraft A. Shakes: lshakes@nccu.edu

Hope M. Wolf: hope.m.wolf@gmail.com

Mohammad A. Mujalled: mamujalled0829@email.campbell.edu

Constance Zhou: constancezhou@gmail.com

Charles Hatcher: chatche1@nccu.edu

Derek C. Norford: dnorford@nccu.edu

transcription of the gene *in vitro*. Although these sequences have been characterized biochemically to adopt a variety of alternate structures in solution [4], definitive examples of gene regulation *in vivo* by such sequences actually existing in triple helical or other structural variations of the double helix are rare (see review in [5]). The lack of evidence however does not imply that gene regulatory sequences perform their function *in vivo* existing as perfect double helices of DNA.

Proteins that bind a specific sequence in DNA makes contact not only with individual functional groups within the base pairs of the consensus sequence but usually also interact weakly with sequences flanking it. In the few cases that have been studied in detail, such weak interactions between the protein and DNA can extend till as much as two turns of the double helix flanking the consensus binding site [6, 7]. The contribution of flanking sequences cumulatively to the binding affinity of the protein to its consensus site can be substantial.

Gene regulatory proteins bound to different sites along the DNA need to interact with one another in order to regulate transcription of the gene. The distances separating such protein-bound sites along the DNA can sometimes be large, in the tens of thousands of base pairs in higher vertebrates. Interactions among proteins bound at different sites are thus likely to be influenced by both the length and sequence of the intervening DNA between sites. Such influences may include bending of the DNA with or without protein binding [8, 9]. Proper phasing of the regulatory protein bound sites, and the histones and other proteins that package the intervening DNA appear necessary for optimal regulation of the gene [10-12]. It is also likely that evolutionary pressures have fine-tuned the sequence variations allowed in these intervening DNA segments. The binding affinity of a protein to its DNA site and other proteins can thus be altered when consensus sequences for binding are taken out of their chromosomal contexts to construct small plasmid DNA vectors to express a gene. Considerations such as these should therefore caution us against inadvertently altering the complex protein-protein interactions when changing the intervening DNA sequences between protein binding sites. In situations where the gene regulatory protein factor exists in the cell as part of a larger protein complex, this problem is expected to be exacerbated.

Both the rate of a biochemical reaction in the cell, as well as the equilibrium that a reversible reaction establishes, are related inversely and logarithmically to the activation energy or the free energy difference between the reactants and products, respectively. It implies that small changes in the affinities of regulatory proteins for one another or their binding sites on DNA can profoundly affect the rates and equilibrium established within the cell. It is conceivable that some of this altered DNA-binding and/or protein-protein interactions resulting from binding to sites with modified flanking sequences could eventually manifest itself in altered tissue specificity of gene expression.

There is also the uncertainty of how one goes about selecting potential gene-regulatory sequences to stitch together in a small plasmid for expressing a gene. While one can make educated guesses one cannot be unbiased, as many of the players involved in the fine-tuning of gene expression remain unknown. In favorable circumstances, clues can be obtained by sequence comparisons of the gene regulatory DNA regions across different species [13].

However in situations where gene regulatory function is conserved without sequence similarity, choosing potential protein binding sequences can become a hit-or-miss affair [14-18]. Using BACs to explore regulation of genes by *cis*-acting elements appears to circumvent many of these hurdles. The DNA cloned in BACs, originally for mere faithful propagation of the DNA, also turns out to be probably the best templates for expressing genes in vertebrates. Many of the gene regulatory protein binding sites have been separated from one another along the DNA and from the start sites of genes owing to expansion of the genome in higher vertebrates throughout evolution. Thus it makes greater sense to use small pieces of the actual chromosome, represented in BACs, for accurate expression of the genes housed in them.

Functionalizing BACs

BACs are known to faithfully propagate approximately 300,000 bp of DNA from the chromosomes of a particular organism in bacteria as large single copy plasmids. Ideally, the genes contained in a BAC plasmid 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 evolved together for optimal function over millions of years. However, expressing the genes this way leaves one with no features to distinguish its expression from that of the organism's genome. A reporter gene to tag genes from the BAC plasmid is therefore desirable. Because many of the genes are expressed only in a limited number of tissue types, a reporter gene such as the Green Fluorescent Protein (GFP) gene is often used. The GFP gene is particularly desirable for organisms that have optical transparency for a significant part of their life cycle. The β -galactosidase gene is an important reporter gene for expressions in the mouse where the embryo is not optically transparent. It goes without saying that the expressed reporter gene protein should ideally not alter cell or tissue function; at least during the period of investigation.

BACs are incapable of replicating by themselves other than in the bacterial host where they are propagated as part of a large F-factor plasmid [19]. 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 of the animal.

Strategies for Functionalizing BACs

Approaches used for altering sequences in the large DNA inserts of BACs are fundamentally different from those traditionally used in small plasmids. There are two reasons for this difference. Common DNA modifying enzymes, such as restriction endonucleases that cut DNA at a specific sequence, have recognition sites that are small and therefore cut frequently in the large BAC DNA. Consequently the large number of DNA fragments generated by cutting with enzymes and the unavailability of robust separation techniques to isolate and keep track of the relative order of the pieces in the starting BAC; precludes using the “cut-and-paste” mechanism to alter sequences in BACs. Secondly, large duplex DNA

that is not packaged with DNA-binding proteins in the test-tube is also very brittle, and keeping it intact during manipulations *in vitro* is a challenge.

The way to circumvent these hurdles is to use DNA recombination. Thus DNA recombination *in vivo* has become the method of choice for modifying sequences in BACs. Although the net result again is cutting and re-joining of DNA ends, the entire process is concerted in the bacterial host with no free ends of DNA to go astray; and occurs primarily in a nucleoprotein complex that protects it from shear forces which otherwise would break the large DNA during manipulations *in vitro*.

Two independent approaches based upon using sequence homology for recombining DNA were developed to alter sequences in BACs. The first method introduced the major recombination function of *E.coli*, RecA [20], back into the severely recombination deficient host bacterium DH10B engineered to propagate vertebrate DNA in BACs. Recombination function, vital to all living organisms, was deliberately rendered low in this bacterial host to prevent vertebrate DNA, with substantial amounts of repetitive sequences, in BACs from rearranging. The second approach introduces recombination functions of phage λ , namely red α , red β and red γ into the DH10B host, and utilizes recombination of homologous sequences of shorter length to insert exogenous DNA cassettes into BACs [21]. Both methods have been widely used to engineer a variety of alterations in BAC DNA, such as introducing reporter gene cassettes into the genomic insert, mutating sequences at a target site, and introducing loxP sites [20-36]. 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 [27]. 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 [28]. 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 [31]. 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 [32]. 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 [33]. Procedures for constructing transgenic animals with functionalized BACs in mice [34], or zebrafish [35], have been reviewed recently. Highlights of using BAC transgenesis in zebrafish to locate distal gene-regulatory sequences were reported recently [36].

A different strategy for modifying BAC DNA also uses recombination, but does not require homologous sequences in genomic inserts of BAC to introduce exogenous DNA cassettes. Insertions of the bacterial transposon Tn10 can introduce exogenous DNA, including lox sites, at random locations in BACs [37, 38]. Site-specific recombination by the Cre-lox system on the other hand, can deliver reporter genes and other exogenous DNA cassettes such as sequencing primer sites, mammalian cell-selectable antibiotic resistance genes, enhancer-traps and sequences specifically recognized by the vertebrate transposon Tol2 precisely at the ends of genomic DNA inserts in BACs [39-41]. 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 should make the approach applicable to a wider variety of BACs in the public domain, including those containing high levels of repetitive sequences (see reference [39] for an example). 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 genome libraries appear to be random, demonstrating little sequence specificity for transposition [39]. 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 [42]. The minor sequence preferences for insertions of Tn10 observed in BACs probably have more to do with accessibility of sites for Tn10 insertions in HU protein-packaged vertebrate DNA in the bacterium than specificity for sequences. Transposition of Tn10 into a rare BAC clone occasionally displays apparent sequence selectivity [43]. However, this has been clearly shown to be due to a clonal selection process that arose from picking single colonies of transposon plasmid transformed BAC clones that had induced the transposase gene prior to actual induction with Isopropyl β -D-1-thiogalactopyranoside (IPTG). Slight modification of the procedure, that has since recommended inducing a large pool of transposon plasmid transformed BAC colonies instead of a single one, has rectified this potential problem completely [43]. A description of the transposon retrofitting approach for BACs follows.

Tn10 transposon retrofitting of BACs

The Tn10 transposon modification procedure for BACs is conceptually simple and easy to do, as illustrated in Figure 1. A loxP sequence is placed within the 70 base inverted repeat ends of the bacterial Tn10 mini-transposon (see lower panel of Figure 1).

Plasmid DNA carrying the transposon is introduced into the same cell that houses the BAC plasmid. Upon induction with IPTG the expressed transposase protein excises the DNA cassette flanked by the inverted repeats (shown as green and pink boxes L & R in the Tn plasmid) and inserts it into other nearby DNA in the bacterial host, including the bacterial genome and the BAC DNA. Insertions of Tn10 into BAC DNA occur in either orientation, and are irreversible because the transposase gene is left behind in the Tn plasmid and destroyed. Important considerations that dictate subsequent steps 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 (see reference [44] for a detailed description of the procedure).

Infection with P1 phage serves an additional purpose. The phage expresses Cre protein early in its life cycle to circularize the linear double-stranded phage DNA existing in the P1 head.

Newly synthesized Cre protein acts *in trans* to also recombine the loxP site transposed into the BAC DNA insert with the loxP endogenous to the BAC vector and located at one end of the insert DNA (see step 1, Figure 2). Thus the transposed loxP site of one orientation produces a deletion from one end of the genomic insert, while the loxP inserted in the opposite orientation simply inverts the DNA between it and the one endogenous to the BAC. Because BACs in all modern libraries carry insert DNA of average size ~ 160 kb, and because the amount of DNA that can be packaged in a P1 head is ~110 kb [45], the P1 headful packaging step serves as a selection strategy to isolate transpositions of loxP in the orientation identical to the endogenous one [46], as only that allows reducing the BAC DNA length to less than ~110 kb. Thus only those end-deletions that reduce the length of BAC DNA to less than ~110 kb are rescued, as shown in Figure 2.

Packaging of the end-deleted BAC DNA occurs from a large co-integrate plasmid in the cell which is described in detail elsewhere [46–48]. In co-integrate formation, the P1 phage DNA effectively contributes a packaging initiation site named the “pac site” which is recognized by the phage packaging proteins [shown in step 2, Figure 2]. The co-integrate DNA is thought to replicate by a “rolling circle” mechanism. A cut is made in the newly synthesized co-integrate DNA at the “pac site”, and the DNA end corresponding to piece loxP-B-A-C is stuffed into newly assembled empty heads of phage P1. Packaging of DNA from co-integrate is directional and the second cleavage is made after the P1 head is full, regardless of sequence, using a “headful-cleavage” mechanism [46–48]. Note that if Cre recombination of the transposed loxP with the loxP endogenous to the BAC did not reduce the length of BAC DNA segment B-A-C, to less than ~110 kb, this piece of BAC DNA inside the phage head would not be flanked by loxP on both ends [see step 3 of Figure 2]. In the absence of loxP sites flanking the truncated BAC DNA on both ends, the DNA would be unable to circularize by Cre recombination and would be destroyed when introduced into cells. Insertions of loxP in the other orientation, that cause inversion of the DNA, cannot reduce the size of segment B-A-C to less than 110 kb. Consequently, the BAC DNA in P1 heads from inversions is destroyed upon entry into cells.

The bacterial lysate containing P1 heads packaged with end-deleted BAC DNA is treated with chloroform. This treatment not only facilitates lysis of P1 infected recombination deficient DH10B but also kills all cells harboring intact transposon and BAC plasmids. This killing is important for selecting Tn10-inserted BACs in the next stage when antibiotic selection is employed, as otherwise cells with no transposon insertions into BACs but merely carrying the two intact plasmids would get selected. The lysate after chloroform treatment is used to infect fresh bacteria. Upon entry of the phage packaged BAC DNA into cells; Cre protein is expressed from the DNA end adjacent to the second loxP site (indicated as X in phage head and co-integrate), but only transiently, because the Cre gene is lost after CreloxP recombination to circularize the linear BAC DNA (Step 3, Figure 2). Cells are plated on LB plates containing chloramphenicol to select for clones of end-deleted BAC. Note that the phage DNA segment [P-G-E] from the co-integrate is incompletely packaged, if at all, and is destroyed after the loxP-Cre circularization in the new host. Even the rare phage particle containing a complete phage genome is unlikely to replicate or survive because of the chloramphenicol selection on the plates.

Selecting with a single antibiotic for the BAC DNA is sufficient for the first round of end-deletions, using either a loxP or a lox511 transposon. Additional selection for transposition of lox sites into BAC DNA is not necessary in this first round because the P1 headful packaging itself serves as the selection for the low frequency of lox-site insertions in the previous step (see references [46] for detailed discussions).

Analyses of retrofitted/ end-deleted BACs

BACs deleted from an end, 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 easily isolated by simple mini-prep procedures, digested with Not I enzyme and analyzed by FIGE [38, 49], as shown in Figure 3. High throughput formats for preparing DNA from BAC deletion clones in 96 well plates, suitable for subsequent FIGE analyses and end sequencing, have also been described earlier [38, 49].

Identifying end-deletions originating from authentic lox-Cre recombinations by unique size of BAC vector DNA upon Not I digestion

Whether the BAC DNA of reduced size arose from an authentic lox-Cre recombination, as opposed to an internal deletion between sequence repeats and/or other recombinogenic sites in the genomic DNA insert, is one of the early determinations that need to be made. Note that lox-Cre independent deletions can also be packaged by the P1 headful mechanism when there is no selection for loxP site transposition *per se*; such deletions are also efficiently isolated [46]. Therefore all our loxP and lox511 transposons have been designed to alter the size of the Not I-Not I BAC-vector DNA band that arises from a Not I enzyme digest of the deletion clone DNA (see references [49, 50] for discussion). An end-deletion from Cre-loxP recombination alters the size of BAC vector DNA, whereas deletions arising from within the genomic insert do not. End-deletions made with our lox511 transposons do not replace the Not I site at the lox511 end in the original BAC, resulting in a linear BAC with no separate vector DNA band [50].

Once authenticity of the retrofitted/ end-deleted BAC clone is established, the exact location of insertion of the Tn10 transposon that created the end deletion is determined by sequencing the BAC DNA directly with unique primer sites located on either the loxP or lox511 transposon ends that remain after deletion formation [38, 50].

Truncations from the opposite end of insert DNA in BACs

The genomic DNA insert in BACs is flanked by a loxP site and a mutant lox511 site [50, 51]. Modern genome libraries of DNA from numerous organisms such as the mouse, rat, human and zebrafish have used the vector pBACe3.6 or its derivatives (reference [50, 52], link for pTARBAC2.1 vector <http://bacpac.chori.org/ptarbac21.htm>), and all share this characteristic [53-56]. Thus one can make end-deletions specifically from the lox511 side of genomic insert in BACs by a similar approach using a lox511 Tn10 mini-transposon [50].

It is important to note that second round deletions from the opposite end of insert DNA requires selecting for the transposition event in addition to the BAC DNA: because first round deletions are selected by P1 headfull packaging, all starting BACs for second round deletions would be less than ~110 kb, and hence require selecting for the low frequency of transposition (see reference [46, 50] for detailed discussions).

Cre-mediated cross-recombination between loxP & lox511 sites does not occur in our method

The 34 bp sequence in the mutant lox site, lox511, differs by one nucleotide in the spacer region from that of wild type loxP (see references [50, 57], for actual sequence of lox511). Varying degrees of promiscuity in recombining different mutant *lox* sites, including the lox511 mutant, with wild type loxP have been reported in previous studies; using both partially purified Cre-extracts *in vitro* [57-60], or Cre over-expressed in cells [61-63]. For example, cross recombination between loxP and lox511 has been reported to occur at efficiencies ranging from 5 to 100 % under those experimental conditions that express Cre constitutively [61-63]. We have not observed such cross-recombination, and high levels of stringency *in vivo* in recombining identical *lox* sites [47, 50], or *lox* sites with at least identical spacers [64], have been achieved with Cre protein expressed from its native source namely, a phage P1 infection [47, 50, 64]. Depending on whether a loxP or a lox511 transposon is used, truncations from the corresponding lox- end can be made readily and exclusively with high stringency (Figure 4). Thus truncations of genomic DNA in a BAC from either end are not only efficient, but are highly specific for that end.

Exogenous DNA cassettes can be introduced precisely at one or both newly created end of BAC DNA

An important feature of the end-deletion technology is the ability to introduce reporter genes and other DNA cassettes precisely at the new end created in the BAC clone. The sequence in front of the loxP or lox511 arrows, as shown in Figures 1 and 4, is retained after the recombination event that creates the deletion (note that orientation of arrow refers to directionality of loxP and lox 511 sequences). This feature has been utilized to place numerous DNA cassettes, such as: i) mammalian cell-selectable antibiotic resistance genes in the BAC [39], ii) a basal promoter-containing EGFP gene in the BAC to serve as an enhancer-trap to functionally locate potential gene-regulatory elements further upstream [40], and iii) introduce iTol2 ends to generate GFP-functionalized BAC DNA ready for integration into zebrafish or mouse chromosomes [41].

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 a physical map of the chromosome using this technology [38, 39]. Long-range gene regulatory elements have also been mapped functionally using the end-deletion approach by either generating transgenic mice with EGFP functionalized BAC DNA [65], or electroporating the PAC-transgene DNA into patient-derived cells in culture [66]. The mouse *Nkx2-5* gene in a BAC was functionalized

with EGFP reporter gene using RedET-recombination, and then truncated progressively from the far upstream 5' end of the Nkx2-5 gene using a loxP transposon to identify transcription-enhancing factor binding sites [65]. Monitoring mRNA levels of the Hox11 gene allowed circumventing the use of a reporter gene such as EGFP in the PAC DNA, and a series of upstream deletions in a PAC clone was used to functionally identify potential far-upstream binding sites for transcription factors in the human Hox11 gene [66].

Inserting reporter gene & truncating BAC ends simultaneously

Introducing the EGFP reporter gene and truncating sequences downstream of the gene of interest in the BAC can be done simultaneously using lox-Tn10 transposons [40]. 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 (see panel A of Figure 5 for enhancer-trap transposon). Progressive truncations from one end of the genomic insert with this enhancer-trap transposon generated a library of BAC DNA molecules that differ in the position of the enhancer-trap cassette 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 [40].

Once a tentative DNA domain has been identified as enhancing transcription of the gene, progressive truncations from the opposite end can be made with a lox511 transposon to verify/substantiate the finding. Truncation from the opposite end and insertion of the ends of the vertebrate transposon iTol2, required for chromosomal integration of the BAC DNA [67], can be done simultaneously using a lox511-iTol2kan Tn10 transposon as illustrated below in Figure 5B (see reference [41]).

BAC transgenesis in zebrafish and mice using Tol2

BACs have been integrated into the germlines of zebrafish and mice using the vertebrate transposon Tol2 system [67, 68, 33, 41]. Although further removed from humans than mice evolutionarily, zebrafish has gained popularity in recent years due to a variety of reasons, that include optical transparency of its embryo. An unexpectedly high degree of conservation of non-coding DNA in vertebrate genomes has also popularized it as a model system to explore the function of highly conserved non-coding DNA in vertebrates [69].

Originally isolated from medaka fish, Tol2 is a vertebrate transposon system that functions effectively in zebrafish and mice [70]. Traditionally, Tol2 ends in the inverted orientation and flanking a 1 kb spacer DNA (iTol2) were introduced into the BAC DNA within the bacterial host using recombination of homologous sequences between the targeting vector and the genomic insert [67]. This way of introducing iTol2 cassettes can be potentially problematic as other preexisting sequence repeats in the BAC can rearrange. To avoid such potential complications from rearrangement of highly repetitive DNA in certain BACs (e.g. the mouse Npr3 gene locus [39]), a simpler and more widely applicable approach was developed using our Tn10 transposons to deliver iTol2 ends into BACs. The iTol2kan DNA cassette is inserted at one end of the BAC by a lox511Tn10 transposon, as shown schematically in Figure 5. It generates a large collection of integration ready BACs suitable for transgenesis into zebrafish or mouse embryos in a single experiment [41]. Progressive

truncations from an end of the genomic insert with this iTol2-Tn10 transposon generates a library of BAC DNA molecules that have varying lengths of upstream regulatory DNA of the Amyloid Precursor Protein (*appb*) gene in the BAC. The collection of well characterized BAC DNA was introduced into zebrafish embryos individually for expression of EGFP in neurons (Figure 6). Rapid functional identification of gene regulatory sequences located at large distances along the DNA from the start sites of genes is a distinct advantage of this approach [68]. Important transcription enhancing elements have been discovered in non-coding DNA from the intron, and 31 kb upstream of the *appb* gene. The clock regulated immune system modulator transcription factor E4BP4/ NFIL3 and Forkhead have been shown to regulate not only the zebrafish *appb* gene but also probably the human APP gene ([68] unpublished observations).

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

BACs offer a large span of DNA to house most, if not all, the sequences recognized by DNA-binding regulatory proteins that act in concert to regulate expression of a gene. But more importantly these protein-binding sites in the BAC DNA exist in their chromosomal contexts, ie. with intact flanking sequences that have evolved for millions of years. Therefore a general approach to retrofit BACs with enhancer-traps was developed such that *cis*-acting regulatory sequences can be identified by expressing the BACs in animals [40]. Comparisons between expressed enhancer-trap BACs in a set are probably more meaningful as the modifications in each remain constant. It is also important to note that this approach to identifying gene-regulatory elements is likely to be less biased than a targeted approach, because we do not select the sequences to test for regulatory potential, as is the case with other targeted mutagenesis approaches in BACs using recombineering [28-30, 32]. In situations where regulatory function is conserved without sequence similarity [14-18], choosing the correct sequence to test might be quite difficult in a targeted approach. Thus using enhancer-traps to “scan” entire lengths of BACs offers an effective and relatively unbiased alternative to other targeted approaches [68].

The BAC enhancer-trap technology described here has three distinctive features that should prove beneficial: i) allows sampling much larger sequences of DNA, including multiple discontinuous regulatory domains that act together to regulate expression of genes such as *appb*, where enhancer sequences in intron 1 and -31 kb upstream are separated by 35 kb of DNA [68]. This 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 in 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 [71-79]. The latter approach is “targeted”, and thus is not free of its biases. In the former approach the enhancer-trap need to be introduced into the fertilized egg at the one-cell stage for germline expression. However, most of the genome is inaccessible

for insertion of the enhancer-trap at this stage of development of the egg. Thus enhancers have been identified in only a very small percentage of the genome in animals studied, and vast regions of the genome appear to be refractory to enhancer-trapping by traditional means. By contrast, insertion of the trap occurs in isolated pieces of chromosomal DNA in BACs in the bacterial host in our approach. Consequently a much greater fraction of the genome is potentially accessible for study, because to date we have not encountered BACs refractory to Tn10 transposon insertions. Using enhancer-traps in individual BACs therefore has the potential to allow a more uniform coverage of the genome because the baseline efficiency of trap insertion is reset for individual BACs in the bacterial host.

Although the methodology does not allow generating internal deletions in the BAC, truncations from the end opposite to the enhancer-trap can be made with a lox511 transposon (as seen when iTol2 cassettes were introduced at the other end of BAC DNA, Figures 5 and 6) to explore functions of potential gene regulatory regions. Furthermore, specific mutations of transcription factor binding sites in gene regulatory domains can be introduced much more easily in a small Tn10 transposon plasmid, which can then be inserted into BAC DNA as described in the recent *apbb* gene regulation study [68].

Substitution of Lox sites flanking BAC inserts with Lox66

In a different application Tn10 transposons carrying ‘arm’ mutants of loxP, such as lox66, have been used to successfully replace the loxP site at one end of the genomic insert DNA in a BAC [64]. It should be possible also to substitute the lox511 site at the other end of insert DNA in a similar manner using a lox site with an identical ‘spacer’ as lox511 but carrying a different ‘arm’ sequence. The usefulness of such substitutions of loxP or lox511 sites in BACs with lox66 can be evaluated from the fact that the integration of loxP plasmids to a loxP site on the chromosome is reversible, with the equilibrium favored for excision [80]. Targeting lox66 substituted BACs specifically to a pre-positioned lox71 site, another arm-mutant of loxP, in chromosomes via lox-Cre recombination should be irreversible [80, 81]. Such strategies are likely to be useful in systems such as the zebrafish, where classical “knock-in” technology via homologous recombination cannot be used because of genome duplication in an ancestral teleost [82].

Advantages and Disadvantages of the Tn10 transposon retrofitting approach

Introducing exogenous DNA sequences into BACs with Tn10 transposons carrying loxP or lox511 sites do not rely upon sequences existing in the genomic inserts. As a result, the Tn10 transposons developed for a specific application are usable in all BACs in the public domain. This is unlike the vectors used for homologous recombination based approaches, where the targeting vectors need to be constructed anew for each BAC.

Because insertions of Tn10 transposons into genomic DNA in BACs appear to be largely random (see references [39] for discussion), large collections of BACs with ends progressively trimmed are generated in a single experiment. Such libraries, when made with Tn10 transposons carrying enhancer-traps or iTol2 cassettes, are uniquely suited for

functionally mapping long-range gene regulatory sequences by expressing them in animals. Thus multiple BACs from a contig spanning a genetic locus should allow such functional analyses to be extended over large sections of the genome. The approach does not require selecting sequences for mutational analyses to test their gene regulatory potential, and therefore it is unbiased. It enables enhancer-trap containing BACs, or BACs retrofitted with EGFP cassettes by sequence homology based recombination, to be converted into deletion libraries with integrated iTol2 ready for chromosome integration in the animal. Absence of undesirable rearrangement of insert DNA due to the high content of sequence repeats in certain BACs is an additional advantage with the approach. We note that vertebrate DNA in general and mammalian DNA in particular, has large numbers of DNA sequences that are repeated [39]. Use of radioactive isotopes is also avoided because no Southern blotting is required in retrofitting BACs with transposons.

A notable drawback of the Tn10 transposon based approach appears to stem from the very feature that makes it readily doable: the P1 headful packaging strategy used to isolate the functionalized BAC so easily also limits the size of BAC DNA that can be analyzed to ~ 110 kb. Although the genomic DNA insert is truncated in the process, the resulting 103 kb insert DNA size, the remainder being BAC vector, is unlikely to be a major disadvantage in most applications; because a majority of vertebrate genes can be housed in their entirety within this size limit. Almost half of evolutionarily conserved non-coding gene-regulatory sequences in vertebrate genomes [69], and probably a similar fraction of those that are conserved in function and shape but not in sequence [14, 15], are located within this span of DNA adjoining start sites of genes.

Lastly, both the lox sites flanking insert DNA in BACs can be readily substituted with lox sites with identical spacer but different arms using Tn10 transposons [64]. Such replacements would make the ~110 kb retrofitted BACs amenable to “knock-in” strategies of a slightly different type: BACs with lox66 substituted for loxP should be targetable to lox71 sites in chromosomes of organisms such as zebrafish, a system where conventional homologous recombination-mediated “knock-in” technology is unavailable due to genome duplication in an ancestral teleost during evolution [82].

Clearly the transposon retrofitting strategy and those based on homologous recombination have strengths that appear somewhat complimentary in nature. Thus a judicious approach might be to use a combination of the two methodologies, as demonstrated in earlier studies [65, 41]. Introducing the Tol2 inverted repeat ends to the ends of genomic DNA insert in a BAC that was previously functionalized with a reporter gene appears straightforward and easier to do with the Tn10 transposon approach. The libraries of iTol2 inserted BACs generated should facilitate integration of trimmed single-copy genes into the germline and also help to functionally map distal *cis*-acting gene regulatory sequences in animals. The approach should be applicable to a wider variety of BACs, including those with a high content of sequence repeats.

Future Applications

Strategy for BAC-vector free transgene integrations using Tol2

Zebrafish transgenesis using traditional approaches, namely homologous recombination to functionalize the BAC with the reporter gene and the iTol2 cassette, usually integrates the ~12 kb BAC vector DNA into chromosomes along with the genomic insert [20-35]. This can introduce multiple copies of the BAC vector DNA when more than one transgene is introduced into the same animal, leaving it susceptible to potential genome rearrangements when the transgenes insert into the same chromosome. With Tn10 transposons to deliver the inverted repeat ends of Tol2, there exists the ready option of placing each Tol2 repeat end (L or R) separately, at either the loxP or lox511 end of genomic insert. Thus one of the repeat ends of Tol2, say Tol2L, can be incorporated as part of one transposon and placed at the loxP end (as shown in Figure 7A), while the other repeat end, Tol2R, can be a part of the lox511 transposon for insertion at the lox511 end of genomic insert DNA as illustrated in Figure 7B. Introducing the inverted repeat ends of Tol2 this way, would allow only the genomic insert DNA to be inserted into chromosomes, while the BAC vector DNA along with the kanamycin resistance gene at the lox511 end and both lox sites, undesirable at times, would get eliminated during integration of the transgene (illustrated in Figure 7). The pair of 70 bp inverted repeat ends of the Tn10 transposon, which is indicated by the pink and green boxes in Figure 7, would not be recognized by Tol2 transposase.

Juxtaposing “far away” gene-regulatory elements located beyond that can be accommodated in BACs by using lox-Tn10 transposons

A substantial number (~30%) of long range regulators of transcription remain outside the size range of ~300 kb BACs to be housed together with the gene it regulates [69]. In such situations, it is conceivable that the long range regulatory sequence can be loaded onto a loxP- or lox511-Tn10 transposon and delivered to either end of the BAC DNA containing the gene and its other regulators, and tested quickly using our approach. This is an advantage of the Tn10 approach, whose potential remains to be tested.

Acknowledgments

The project described was supported by Award Number P20MD000175 from the National Center on Minority Health and Health Disparities (NCMHD) and funds from the North Carolina Biotechnology Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCMHD or the National Institutes of Health. We thank Ms. Rosalind Grays, Camilla Felton, Connie Keys, Crystal McMichael and Darlene Laws for support and encouragement. PKC would like to thank Drs. Ken Harewood and Sean Kimbro for encouragement and support.

REFERENCES

1. Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F. The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. *FASEB J.* 1988; 2:2939–2949. [PubMed: 3053307]
2. Ghosh MK, Katyal A, Chandra R, Brahmachari V. Targeted activation of transcription in vivo through hairpin-triplex forming oligonucleotide in *Saccharomyces cerevisiae*. *Mol. Cell. Biochem.* 2005; 278:147–155. [PubMed: 16180100]

3. Goni JR, Vaquerizas JM, Dopazo J, Orozco M. Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC Genomics*. 2006; 7:63. [PubMed: 16566817]
4. Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.* 2007; 282:32433–32441. [PubMed: 17785457]
5. Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* 2008; 36:5123–5138. [PubMed: 18676453]
6. Gartenberg MR, Ampe C, Steitz TA, Crothers DM. Molecular characterization of the GCN4-DNA complex. *Proc Natl Acad Sci U S A.* 1990; 87:6034–6038. [PubMed: 2201019]
7. Dalma-Weiszhausz DD, Gartenberg MR, Crothers DM. Sequence-dependent contribution of distal binding domains to CAP protein-DNA binding affinity. *Nucleic Acids Res.* 1991; 19:611–616. [PubMed: 1849264]
8. Fried MG, Crothers DM. CAP and RNA polymerase interactions with the lac promoter: binding stoichiometry and long range effects. *Nucleic Acids Res.* 1983; 11:141–158. [PubMed: 6306561]
9. Ulanovsky L, Bodner M, Trifonov EN, Choder M. Curved DNA: design, synthesis, and circularization. *Proc Natl Acad Sci U S A.* 1986; 83:862–866. [PubMed: 3456570]
10. Hebbar PB, Archer TK. Chromatin-dependent cooperativity between site-specific transcription factors *in vivo*. *J Biol Chem.* 2007; 282:8284–8291. [PubMed: 17186943]
11. Ganapathi M, Singh GP, Sandhu KS, Brahmachari SK, Brahmachari V. A whole genome analysis of 5' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. *Gene.* 2007; 391:242–251. [PubMed: 17321698]
12. Hardy S, Shenk T. E2F from adenovirus-infected cells binds cooperatively to DNA containing two properly oriented and spaced recognition sites. *Mol Cell Biol.* 1989; 9:4495–4506. [PubMed: 2531279]
13. Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O. Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Hum Mol Genet.* 2005; 14:3057–3063. [PubMed: 16155111]
14. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science.* 2006; 14:276–279. [PubMed: 16556802]
15. Parker SC, Hansen L, Abaan HO, Tullius TD, Margulies EH. Local DNA topography correlates with functional noncoding regions of the human genome. *Science.* Apr 17; 2009 324(5925):389–92. Epub 2009 Mar 12. [PubMed: 19286520]
16. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Bristow J, Ren B, Black BL, Rubin EM, Visel A, Pennacchio LA. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet.* 2010; 42:806–810. [PubMed: 20729851]
17. Kague E, Bessling SL, Lee J, Hu G, Passos-Bueno MR, Fisher S. Functionally conserved cis-regulatory elements of COL18A1 identified through zebrafish transgenesis. *Dev Biol.* 2010; 337:496–505. [PubMed: 19895802]
18. Taher L, McGaughey DM, Maragh S, Aneas I, Bessling SL, Miller W, Nobrega MA, McCallion AS, Ovcharenko I. Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res.* 2011; 7:1139–1149. [PubMed: 21628450]
19. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. 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 USA.* 1992; 89:8794–8797. [PubMed: 1528894]
20. Yang XW, Model P, Heintz N. Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol.* 1997; 9:859–865. [PubMed: 9306400]
21. Zhang Y, Buchholz F, Muylers JP, Stewart AF. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet.* 1998; 20:123–128. [PubMed: 9771703]

22. Jessen JR, Meng A, McFarlane RJ, Paw BH, Zon LI, Smith GR, Lin S. Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. *Proc Natl Acad Sci U S A*. 1998; 95:5121–5126. [PubMed: 9560239]
23. Muylers JP, Zhang Y, Testa G, Stewart AF. Rapid modification of Bacterial Artificial Chromosomes by ET recombination. *Nucleic Acids Res*. 1999; 27:1555–1557. [PubMed: 10037821]
24. Gong S, Yang XW, Li C, Heintz N. Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res*. 2002; 12:1992–1998. [PubMed: 12466304]
25. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res*. 2005; 33:e36. [PubMed: 15731329]
26. Yang Z, Jiang H, Chachainasakul T, Gong S, Yang XW, Heintz N, Lin S. Modified bacterial artificial chromosomes for zebrafish transgenesis. *Methods*. 2006; 39:183–188. [PubMed: 16828309]
27. Orford M, Nefedov M, Vadolas J, Zaibak F, Williamson R, Ioannou PA. Engineering EGFP reporter constructs into a 200 kb human beta-globin BAC clone using GET Recombination. *Nucleic Acids Res*. 2000; 18:E84. [PubMed: 10982895]
28. Mortlock DP, Guenther C, Kingsley DM. A general approach for identifying distant regulatory elements applied to the Gdf6 gene. *Genome Res*. 2003; 9:2069–2081. [PubMed: 12915490]
29. Jessen JR, Willett CE, Lin S. Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish. *Nat Genet*. 1999; 23:15–16. [PubMed: 10471489]
30. Carvajal JJ, Cox D, Summerbell D, Rigby PW. A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development*. 2001; 128:1857–1868. [PubMed: 11311165]
31. Gray M, Shirasaki DS, Cepeda C, Andre VM, Wilburn B, Lu X-H, Tao J, Yamazaki I, Li SH, Li XJ, Levine MS, Yang XW. Full length human mutant huntington with a stable polyglutamine repeat can elicit progressive and selective neuropathogenesis in BACHD mice. *J Neurosci*. 2008; 28:6182–6195. [PubMed: 18550760]
32. Collins PL, Chang S, Henderson M, Soutto M, Davis GM, McLoed AG, Townsend MJ, Glimcher LH, Mortlock DP, Aune TM. Distal regions of the human IFNG locus direct cell type-specific expression. *J Immunol*. 2010; 185:1492–1501. [PubMed: 20574006]
33. Bussmann J, Schulte-Merker S. Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development*. 2011; 138:4327–32. [PubMed: 21865323]
34. Fu Y, Maye P. Engineering BAC reporter gene constructs for mouse transgenesis. *Methods Mol Biol*. 2011; 693:163–179. [PubMed: 21080280]
35. Suster ML, Abe G, Schouw A, Kawakami K. Transposon-mediated BAC transgenesis in zebrafish. *Nat Protoc*. 2011; 6:1998–2021. [PubMed: 22134125]
36. Chatterjee S, Lufkin T. Fishing for function: zebrafish BAC transgenics for functional genomics. *Mol BioSystems*. 2011; 7:2345–2351.
37. Chatterjee PK, Coren JC. 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*. 1997; 25:2205–2212.
38. Chatterjee PK, Yarnall DP, Haneline SA, Godlevski MM, Thornber SJ, Robinson PS, Davies HE, White NJ, Riley JH, Shepherd NS. Direct Sequencing of Bacterial and P1 Artificial Chromosome Nested-deletions for Identifying Position-Specific Single Nucleotide Polymorphisms. *Proc. Natl. Acad. Sci. (USA)*. 1999; 96:13276–13281. [PubMed: 10557311]
39. Gilmore RC, Baker J Jr, Dempsey S, Marchan R, Corprew RNL Jr, Maeda N, Smithies O, Byrd G, Bukoski RD, Harewood KR, Chatterjee PK. Using PAC nested-deletions to order contigs and microsatellite markers at the high repetitive sequence containing Npr3 gene locus. *Gene*. 2001; 275:65–72. [PubMed: 11574153]
40. Shakes LA, Malcolm TL, Allen KL, De S, Harewood KR, Chatterjee PK. Context dependent function of APPb Enhancer identified using Enhancer Trap-containing BACs as Transgenes in Zebrafish. *Nucleic Acids Research*. 2008; 36:6237–6248. [PubMed: 18832376]

41. Shakes LA, Abe G, Eltayeb MA, Wolf HM, Kawakami K, Chatterjee PK. Generating libraries of iTol2-end insertions at BAC ends using loxP and lox511 Tn10 transposons. *BMC Genomics*. 2011; 12:351. [PubMed: 21736732]
42. Craig NL. Target site selection in transposition. *Annu Rev Biochem*. 1997; 66:437–474. [PubMed: 9242914]
43. Chatterjee PK, Briley LP. Analysis of a Clonal Selection Event during Transposon-Mediated Nested-Deletion Formation in Rare BAC and PAC Clones. *Analytical Biochemistry*. 2000; 285:121–126. [PubMed: 10998271]
44. Chatterjee, PK. Retrofitting BACs and PACs with LoxP Transposons to Generate Nested Deletions.. In: Zhao, Shaying; Stodolsky, Marvin, editors. *Bacterial Artificial Chromosomes. Methods in Mol. Biology*. Vol. 1. The Humana Press Inc.; 2004. p. 231-241.series vol 255
45. Yarmolinsky, MB.; Sternberg, N. Bacteriophage P1 in *The Bacteriophages*. Calendar, Richard, editor. Vol. 1. Plenum Publishing Corporation; 1988.
46. Chatterjee PK, Mukherjee S, Shakes LA, Wilson W III, Harewood KR, Byrd G. Selecting Transpositions of a Markerless Transposon Using Phage P1 Headful Packaging: New Transposons for Functionally Mapping Long Range Regulatory Sequences in BACs. *Analytical Biochemistry*. 2004; 335:305–315. [PubMed: 15556570]
47. Chatterjee PK, Shakes LA, Srivastava DK, Garland DM, Harewood KR, Moore KJ, Coren JS. Mutually Exclusive Recombination of Wild Type and Mutant loxP Sites *in vivo* Facilitates Transposon-Mediated Deletions from Both Ends of Genomic DNA in PACs. *Nucleic Acids Research*. 2004; 32:5668–5676. [PubMed: 15494454]
48. Chatterjee PK, Sternberg NL. Retrofitting high molecular weight DNA cloned in P1: introduction of reporter genes, markers selectable in mammalian cells and generation of nested deletions. *Genet Anal Biomol. Eng*. 1996; 13:33–42.
49. Chatterjee, PK.; Baker, JC, Jr.. *Bacterial Artificial Chromosomes. Methods in Mol. Biology series*. Vol. 1. Vol. 255. The Humana Press Inc.; Shaying Zhao & Marvin Stodolsky; 2004. Preparing Nested Deletions Template DNA for Field Inversion Gel Electrophoresis Analyses and Position-Specific End Sequencing With Transposon Primers.; p. 243-254.
50. Shakes LA, Garland DM, Srivastava DK, Harewood KR, Chatterjee PK. 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. *Nucleic Acids Research*. 2005; 33:e118. [PubMed: 16061933]
51. Frengen E, Weichenhan D, Zhao B, Osoegawa K, van Geel M, de Jong PJ. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics*. 1999; 58:250–253. [PubMed: 10373322]
52. Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, Catanese JJ, de Jong PJ. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics*. 1998; 52:1–8. [PubMed: 9740665]
53. Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res*. 2000; 10:116–128. [PubMed: 10645956]
54. Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res*. 2001; 11:483–496. [PubMed: 11230172]
55. Osoegawa K, Zhu B, Shu CL, Ren T, Cao Q, Vessere GM, Lutz MM, Jensen-Seaman MI, Zhao S, de Jong PJ. BAC resources for the rat genome project. *Genome Res*. 2004; 4:780–785. [PubMed: 15060022]
56. Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Girn N, Lee D, Masson A, Mayo M, Olson T, Pandoh P, Prabhu AL, Schoenmakers E, Tsai M, Albertson D, Lam W, Choy CO, Osoegawa K, Zhao S, de Jong PJ, Schein J, Jones S, Marra MA. A set of BAC clones spanning the human genome. *Nucleic Acids Res*. 2004; 32:3651–3660. [PubMed: 15247347]
57. Hoess RH, Wierzbicki A, Abremski K. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res*. 1986; 14:2287–2300. [PubMed: 3457367]

58. Lee G, Saito I. Role of nucleotide sequences of *loxP* spacer region in Cre-mediated recombination. *Gene*. 1998; 216:55–65. [PubMed: 9714735]
59. Missirlis PI, Smailus DE, Holt RA. A high-throughput screen identifying sequence and promiscuity characteristics of the *loxP* spacer region in Cre-mediated recombination. *BMC Genomics*. 2006; 7:73–85. [PubMed: 16595017]
60. Langer SJ, Ghafoori AP, Byrd M, Leinwand L. A genetic screen identifies novel non-compatible *loxP* sites. *Nucleic Acids Res*. 2002; 30:3067–3077. [PubMed: 12136089]
61. Sheren J, Langer SJ, Leinwand LA. A randomized library approach to identifying functional *lox* site domains for the Cre recombinase. *Nucleic Acids Res*. 2007; 35:5464–5473. [PubMed: 17702764]
62. Siegel RW, Jain R, Bradbury A. Using an *in vivo* phagemid system to identify non-compatible *loxP* sequences. *FEBS Lett*. 2001; 505:467–473. [PubMed: 11576551]
63. Wang Z, Engler P, Longacre A, Storb U. An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res*. 2001; 11:137–142. [PubMed: 11156622]
64. Chatterjee PK, Shakes LA, Stennett N, Richardson VL, Malcolm TL, Harewood KR. Replacing the wild type *loxP* site in BACs from the public domain with *lox66* using a *lox66* transposon. *BMC Res Notes*. 2010; 3:38. [PubMed: 20170521]
65. Chi X, Chatterjee PK, Wilson W III, Zhang S-X, DeMayo F, Schwartz RJ. Complex Cardiac Nkx2-5 Gene Expression Activated By Noggin Sensitive Enhancers Followed By Chamber Specific Modules. *Proc. Natl. Acad. Sci. (USA)*. 2005; 102:13490–13495. [PubMed: 16150722]
66. Brake RL, Chatterjee PK, Kees UR, Watt PM. The functional mapping of long-range transcription control elements of the *HOX11* proto-oncogene. *Biochem. Biophys. Res. Com.* 2004; 313:327–335. [PubMed: 14684164]
67. Suster ML, Sumiyama K, Kawakami K. Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics*. 2009; 10:477. [PubMed: 19832998]
68. Shakes LA, Du H, Wolf HM, Hatcher C, Norford DC, Precht P, Sen R, Chatterjee PK. Using BAC Transgenesis in Zebrafish to Identify Regulatory sequences of the Amyloid Precursor Protein gene in Humans *BMC. Genomics*. 2012; 13:451. [PubMed: 22947103]
69. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* Jan.2005 3:e7. (2005). [PubMed: 15630479]
70. Kawakami K. Transposon tools and methods in zebrafish. *Developmental Dynamics*. 2005; 234:244–254. [PubMed: 16110506]
71. O'Kane CJ, Gehring WJ. Detection in situ of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci U S A*. 1987; 84:9123–9127. [PubMed: 2827169]
72. Wilson C, Pearson RK, Bellen HJ, O'Kane CJ, Grossniklaus U, Gehring WJ. P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev*. 1989; 3:1301–1313. [PubMed: 2558051]
73. Korn R, Schoor M, Neuhaus H, Henseling U, Soininen R, Zachgo J, Gossler A. 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*. 1992; 39:95–109. [PubMed: 1283314]
74. Grabher C, Henrich T, Sasado T, Arenz A, Wittbrodt J, Furutani-Seiki M. Transposon-mediated enhancer trapping in medaka. *Gene*. 2003; 322:57–66. [PubMed: 14644497]
75. Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB, Welle Z, Ekker SC. Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics*. 2004; 5:62. [PubMed: 15347431]
76. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Developmental Cell*. 2004; 7:133–144. [PubMed: 15239961]

77. Ellingsen S, Laplante MA, Konig M, Kikuta H, Furmanek T, Hoivik EA, Becker TS. Large-scale enhancer detection in the zebrafish genome. *Development*. 2005; 132:3799–3811. [PubMed: 16049110]
78. Nagayoshi S, Hayashi E, Abe G, Osato N, Asakawa K, Urasaki A, Horikawa K, Ikeo K, Takeda H, Kawakami K. Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: *tcf7* and *synembryn-like*. *Development*. 2008; 135(1):159–169. [PubMed: 18065431]
79. Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F, Gómez-Skarmeta JL. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn*. 2009; 238:2409–2417. [PubMed: 19653328]
80. Araki K, Araki M, Yamamura K. Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Res*. 1997; 25:868–872. [PubMed: 9016639]
81. Liu*, Wei-yi; Wang*, Yun; Qin, Yao; Wang, Ya-ping; Zhu, Zuo-yan. Site-Directed Gene Integration in Transgenic Zebrafish Mediated by Cre Recombinase Using a Combination of Mutant Lox Sites. *Marine Biotechnology*. 2007; 9:420–428. [PubMed: 17503154]
82. Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonhaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V, Coutinho P. Subfunctionalization of duplicated zebrafish *pax6* genes by cis-regulatory divergence. *PLoS Genet*. 2008; 4:e29. [PubMed: 18282108]

BAC End-deletion Technology

Figure 1. Schematic representation of the BAC end-deletion technology using a loxP transposon

The inverted triangle represents the Tn10 transposon carrying a loxP site, which is shown in greater detail in the lower panel. The 70 bp inverted repeat ends of the Tn10 transposon are indicated by the green and pink boxes marked L and R respectively. The thick black arrow represents the loxP site in both the transposon and BAC vector DNA. The broken arrow represents the mutant lox511 site in BAC vector. Because transposon insertions are rare, single random transpositions into the BAC DNA occur when the transposase gene is induced. Upon Cre recombination, the transposon-inserted loxP of identical orientation to the loxP endogenous to the BAC generates a deletion of the DNA between them. Inversions are not shown. The pool of BAC DNA molecules therefore generates a library of end-deleted BACs from the random insertions of a loxP transposon into BAC DNA molecules.

Figure 2. Schematic representation of DNA recombinations occurring in the Tn10 transposon retrofitting and loxP-Cre end deletion of BACs

Step 1 shows creation of a deletion in the BAC DNA by Cre recombination of the transposed and endogenous loxP sites. **Step 2** shows the Cre recombination of circularized phage P1 DNA (P-G-E) with the end-deleted BAC DNA (B-A-C) to generate the co-integrate. The phage packaging site “pac site” and the Cre gene are indicated in the phage DNA by the thick solid bar and the X, respectively. **Step 3** shows the BAC DNA packaged in the phage P1 head, its recombination after entry into cells by newly synthesized Cre protein, and circularization into a BAC plasmid. If the length of DNA (B-A-C) exceeds the headful packaging capacity of ~110 kb, then the second loxP site (indicated by the thick arrows) would not fit in the phage head and the DNA cannot circularize by loxP-Cre recombination after entry into cells. This leads to the BAC DNA not being rescued. Note that the Cre gene is also lost upon circularization of the truncated BAC DNA.

Progressive truncations of genomic DNA from an end in BAC

Figure 3. FAGE display of an array of BAC deletion clone DNAs

DNA was isolated by the miniprep procedure from clones of a library of end-deleted BACs generated using a loxP transposon. The DNA was digested with Not I enzyme before electrophoresis. Lane 1 shows a 5 kb ladder, and lane 2 shows the DNA from the starting BAC clone.

Deleting Both Ends of DNA insert in BAC

With pTnLox511(B)markerless 1

With pTn(RSVneo 2)/loxP

Figure 4. Schematic representation of end-deletions generated from both ends of DNA insert in a BAC clone

The upper panel shows truncations from the lox511 end using a lox511 transposon. The largest clone from that library was then truncated from the loxP end with a loxP transposon. Figure is adapted from reference [50].

Figure 5. Schematic representation of deletion formation in starting BACs *appb* EGFP enhancer-trap by p*Tnlox511*-i*Tol2kan*

Panel A shows generation of starting BAC *appb* EGFP enhancer-trap by insertion of the enhancer-trap transposon (inverted triangle), and the location of *appb* gene within the BAC clone. **Panel B:** Shown schematically is the insertion of the Tn*lox511*-i*Tol2kan* transposon into the enhancer-trap *appb* BAC DNA. The thick black arrows represent loxP and lox511 sites (discontinuous arrow). DNA cassettes in front of the loxP or lox511 arrows are retained after the loxP-loxP or lox511-lox511 recombinations by Cre. The enhancer-trap DNA cassette is positioned in front of the loxP arrow (panel A), while the i*Tol2kan* cassette is located in front of the lox511 arrow. The red and purple arrowheads pointing outward in the i*Tol2kan* cassette correspond to the 200 bp inverted repeat end R and 150 bp inverted repeat end L of the Tol2 transposon (see Suster et al 2009 ref [67]). The kanamycin resistance gene is between the Tol2 inverted repeat ends. Adapted from reference [41].

iTol2-end insertion at lox511 end of BAC DNA

Figure 6. Panel A: FAGE analysis of an array of *apbb* BACs with EGFP containing Enhancer-trap at loxP end and iTol2kan cassette at lox511 end of DNA insert

The BAC DNA was digested with Not I before FAGE. Note that there is no vector DNA band in lanes 3-19 because the lox511-lox511 deletion with Cre removes both the Not I site in front of the BAC-lox511 and the Not I site behind the lox511 in the transposon. Lane 1 shows the starting enhancer-trap BAC which has the expected vector DNA band. **Panel B:** Germline transgenic F1 zebrafish obtained from injecting eggs with the BAC DNA shown by the blue arrowhead in lane 13.

Transposon plasmids for BAC vector-less insertions

Figure 7. Panels A: Schematic representation of Transposon plasmids for BAC vector-less insertions

Tn10 transposon plasmids pTn(Tol2L)/loxP and pTn(Tol2Rkan)/lox511 are drawn schematically to contain the Tol2L and Tol2R end sequences, indicated by the purple and red large arrowheads, respectively. **Panel B:** Resulting BAC DNA after transposition of the two transposons and lox-Cre recombinations.