ORIGINAL PAPER

Bacterial artificial chromosomes as analytical basis for gene transcriptional machineries

**Junko Asami · Yukiko U. Inoue ·
Youhei W. Terakawa · Saki F. Egusa ·
Takayoshi Inoue**

Abstract Bacterial Artificial Chromosomes (BACs) had been minimal components of various genome-sequencing projects, constituting perfect analytical basis for functional genomics. Here we describe an enhancer screening strategy in which BAC clones that cover any genomic segments of interest are modified to harbor a reporter cassette by transposon tagging, then processed to carry selected combinations of gene regulatory modules by homologous recombination mediated systematic deletions. Such engineered BAC-reporter constructs in bacterial cells are ready for efficient transgenesis in mice to evaluate activities of gene regulatory modules intact or absent in the constructs. By utilizing the strategy, we could speedily identify a critical genomic fragment for spatio-temporally regulated expression of a mouse cadherin gene whose structure is extraordinarily huge and intricate. This BAC-based methodology would hence provide a novel screening platform for gene transcriptional machineries that dynamically fluctuate during development, pathogenesis and/or evolution.

The authors Junko Asami, Yukiko U. Inoue contributed equally to this work.

J. Asami · Y. U. Inoue · Y. W. Terakawa ·
S. F. Egusa · T. Inoue (✉)
Department of Biochemistry and Cellular Biology,
National Institute of Neuroscience, National Center of
Neurology and Psychiatry, Ogawahigashi 4-1-1, Kodaira,
Tokyo 187-8502, Japan
e-mail: tinoue@ncnp.go.jp

Introduction

In the post genome-sequencing era, genome-wide means of functional assay has become possible to systematically understand principles of the life and its changeful conditions (i.e., homeostatic regulations, disorders etc.). For example, by taking advantage of the DNA microarray screening system, one can describe which genes among tens of thousands are transcribed into mRNAs at any time and place of the developing and/or pathogenic organisms (reviewed in Dufva 2009; Lockhart and Winzeler 2000). The proteomics standard by which extremely sensitive detection capacity of mass spectrometry can characterize every small protein spot from multiple dimensional electrophoresis has also opened a door for high-throughput means of profiling translated protein dynamics in various contexts of organism's development and/or life conditions (Choudhary and Mann 2010; Pandey and Mann 2000 and references therein). Regarding the gene regulatory machineries, bioinformatics tools such as evolutionarily conserved region (ECR) browser (Ovcharenko et al. 2004) help list the non-coding genomic fragments highly conserved among species, which might be implicated to

play significant roles during organism's development, pathogenesis and/or evolution. The chromatin immunoprecipitation (ChIP) based technology (i.e., ChIP-sequencing; Robertson et al. 2007) further allows identification of those genomic fragments with which given transcriptional protein complexes interact and the ENCODE project (Thomas et al. 2006) recently unifies such comparative/experimental information on the whole human genome, providing a useful starting point to examine gene regulatory landscapes (reviewed in Natoli 2010). However, gene transcriptional regulations are elaborated by synergistic interactions among these functional genomic fragments that are sometimes scattered over a mega-base sized genomic territory and it is still of challenge to quickly and thoroughly investigate the value of non-coding genomic regions from any species in any in vivo contexts.

Bacterial Artificial Chromosome (BAC) contains an *Escherichia coli* derived F-factor replication origin, allowing stable propagation of a large exogenous DNA fragment (average size: ∼200 kb) by a single copy per a bacterial cell in a supercoiled circular form (Shizuya et al. 1992; Shizuya and Kouros-Mehr 2001). Yeast Artificial Chromosomes (YACs) could also keep mega-base sized DNA fragments in a long linear structure (Burke et al. 1987), yet YAC inserts are often chimeric with a higher rate of recombination, making it difficult to steadily engineer and maintain the YAC structure all the time (Shizuya et al. 1992; Shizuya and Kouros-Mehr 2001). From this reason, BACs rather than YACs had been utilized as rigid minimal components of genome-sequencing projects, and researchers can now easily obtain BAC clones that differentially cover any genomic regions of human, mouse, rat, chicken and so on. Recently, resourceful methods to precisely manipulate BAC clones by means of simple homologous recombination as well as transposon tagging in bacterial cells were developed (Yang et al. 1997; reviewed in Copeland et al. 2001). Since BAC transgenic system is efficient in wide variety of cell lines and fertilized eggs from mouse, frog and zebrafish (Poser et al. 2008; Montigny et al. 2003; Wade-Martins et al. 2001; Antoch et al. 1997; Kelly et al. 2005; Jessen et al. 1998), setting BACs as analytical basis would be a promising way for functional genomics at the post genome-sequencing era. Many research groups including ours (Spitz et al. 2003; Jeong et al. 2006; Inoue et al. 2008; Carvajal

et al. 2008) have certainly engineered BACs to survey large *cis*-regulatory landscapes in mouse. Nonetheless, investigative principles and/or conditions for BAC modification toolkits are even now mixtures of good and bad and extensive adjustments would be required for their effective use.

In the present study, we introduce an experimental strategy in which mouse or virtually any vertebrates' gene regulatory regions can be characterized from a mega-base sized genomic territory more comprehensively than ever. As the first step, transposon mediated reporter tagging was processed among BAC clones that differentially cover the genomic territory of interest in bacterial cells. These BACs are then microinjected into fertilized mouse eggs to select BAC clones that can recapitulate gene regulatory patterns of interest. Next, homologous recombination mediated systematic regional deletions from above selected BAC-reporter construct are carried out in bacterial cells and these modified constructs are again microinjected into mouse fertilized eggs to decisively evaluate the reporter activity in the developing mouse organisms: in this final step, rigorous assessment of how BAC deletion patterns affect reporter expression profiles reliably distinguishes critical regions for gene transcriptional regulations. By applying these methodological combinations, we could quickly identify a gene regulatory region for spatio-temporally restricted expression of a mouse cadherin gene whose organization had been too huge and complex to examine gene regulatory machinery with conventional methodology and/or bioinformatics. This strategy would thus provide a novel high-throughput screening platform for any non-coding genomic fragments that might harbor unknown significance in multi-cellular organisms' development, evolution and/or pathogenesis.

Materials and methods

Transposon mediated reporter tagging of BAC clones

A modified version of *Tn*1000 carrying the BGZ40 cassette used in this study was described previously (Morgan et al. 1996). Briefly, BAC clones *RP24-88H4*, *RP23-190L12* or *RP23-78N21* that differentially cover mouse cadherin-6 (*Cdh6*) gene locus

were electroporated into the male strain *MH1844* in which modified *Tn*1000 is resident on the sex factor plasmid *R388* (Morgan et al. 1996). Next, the BAC transformant was mated with the Streptomycin (Strp) resistant female strain DH10B on LB agar plate for 2 h at 37°C and those females received modified BAC with single random transposon insertion were selected on LB agar plate containing 100 μg/ml Strp and 12.5 μg/ml Chloramphenicol (Cam) at 37°C. After getting rid of *R388* from selected colonies by re-electroporation into DH10B strain, positions of transposon insertion were determined by direct sequencing of modified BACs from both ends of the transposon (Primers for direct sequencing were described in Morgan et al. 1996).

Homologous recombination mediated modification of BAC clones

A recombinogenic bacterial strain EL250 was used as previously described (Lee et al. 2001). For homologous recombination mediated BAC deletions, ∼1-kilo-base pair (kb) sized 5′ and 3′ arms were amplified via polymerase chain reaction (PCR) and cloned into the *pGEM-T-easy* vector (Promega) and in between these arms, an Ampicillin (Amp)-resistant (r) gene cassette or a Kanamycin (Kan)-r cassette put between two FRT sites was subcloned. Then the portion of arms with a selection cassette was released from the vector and kept at −30°C. Homologous recombination for BAC deletion was processed as was shown in Figs. 3a and 4b. Briefly, BAC clones for modification were electroporated into the recombinogenic EL250 strain and selected on L-agar plate containing 12.5 μg/ml Cam. Next, a selected single colony was inoculated in 50 ml of LB without antibiotics at 32°C until the OD_{600} reached to 0.5 and 15 ml of the culture was heat-shocked at 42°C for 15 min, while another 15 ml was maintained at 32°C for 15 min as negative control. Both cultures were placed on ice for 15 min, centrifuged and quickly washed by ice-cold water to keep the electro-competency and recombination activity. A few hundred nanograms of the arm-selection cassette fragment was then electroporated into the electro-competent cells and positive selection was carried out at 32°C on L-agar plate containing 12.5 μg/ml Cam and 12.5 μg/ml Kan or Amp for ∼40 h. Ordinarily, as

many as 100 colonies made growth from the heat-shocked competent cells, while few colony growth occurred from the control ones. After the homologous recombination, Amp-r or Kan-r cassette was excised by adding arabinose to the log phase liquid culture of the EL250 cells with the correctly modified BAC at 0.1% v/v for 1 h at 32°C to sufficiently induce *Flpe* (Fig. 3a). For confirmation of the deletion and excision, pulse filed gel electrophoresis (PFGE) by using the CHEF-DRII system (BioRad) and PCR based evaluation were performed. As for PFGE, *Not*I digested BAC constructs were applied in 1% agarose gel containing 0.5 x Tris-borate-EDTA (TBE) buffer for electrophoresis and the gel was placed under 6 V/cm electric field in cooled 0.5 x TBE buffer circulation with the field switch time being set at 1–20 s for 19 h. Small blocks of MidRange I PFG Marker (#N3551S, New England Biolabs) were simultaneously applied to the gel for estimation of the BAC fragment sizes.

Generation of BAC transgenic mice

All animal experiments in this study conform to Japanese governmental guidelines and have been approved by the Animal Care and Use Committee of the National Institute of Neuroscience, Japan (Project #2007022).

BAC transgenic mice were generated and genotyped as previously described (Inoue et al. 2008). Briefly, BAC DNA was purified from 1.5-l LB bacterial culture by equilibrium ultracentrifugation in continuous CsCl-ethidium bromide gradients. The purified BAC was linearized by PI-*Sce*I and dialyzed to the injection buffer. Final concentration and purity of the linearized BAC-reporter construct was checked by an O.D. meter (NanoDrop 1000, Thermo). For microinjections into pronuclei of fertilized eggs derived from the superovulated B6C3F1 mouse strain (Charles River, Japan), DNA was diluted to ∼2 ng/μl by the injection buffer and tails of transgenic founder (F0) pups or yolk sacs of F0 embryos were collected for the genotyping. For BAC transgenic founders/embryos, presence of *RP23/24*-BAC vector sequences immediately upstream and downstream of the PI-*Sce*I site was always confirmed by PCR to minimize the possibility that fortuitous deletions fell on the BAC construct integrated into the mouse chromosome as was described in Inoue et al. (2008).

Whole-mount detection of beta-galactosidase activities in the transgenic embryos

Timed pregnant mouse embryos precisely staging by Theiler (1972) were stained as described in Inoue et al. (2008). Photographs were taken under the binocular microscope (MZ8, Leica) equipped with a CCD camera (DFC300 FX, Leica).

Results and discussion

Transposon mediated reporter cassette insertions into a given BAC clone effectively trapped gene transcriptional machineries included within the BAC

Transposon is a powerful DNA delivery tool if its insertion and excision can strictly be controlled. In order to trap gene regulatory machineries included within a single BAC clone, we utilized a modified version of the *Tn*1000 bacterial transposable element in which a copy of BGZ40 reporter cassette was stably introduced (Fig. 1a; Morgan et al. 1996). The BGZ40 cassette contains a human beta-globin minimal promoter, *Escherichia coli* beta-galactosidase gene (*LacZ*) as a reporter, and a polyadenylation signal (pA) from simian virus 40 (SV40), virtually allowing rigid reporter activation after the transgenesis only when gene regulatory elements (i.e., enhancers) interactive with the human beta-globin minimal promoter were included in the modified BAC clone.

We initially examined how the BGZ40 modified transposon was integrated into a single BAC clone *RP24-88H4* that mainly covers 5′ upstream region of the mouse cadherin-6 (*Cdh6*) gene transcription start site (TSS; Fig. 1b). As was previously reported, transposon mediated integration of the reporter cassette into BACs can be achieved via simple 2 h' inter-strain mating of bacterial cells (Morgan et al. 1996) and we indeed found that every selected colony after the mating process harbors a single insertion of transposon at quasi random positions in the BAC clone (Fig. 1b). We usually obtained hundreds of colonies from the bacterial mating mix derived from a single 100-mm plate dish with ∼90% of the colonies yielding right transposon insertions. This demonstrates that a single mating experiment for a single

BAC clone is good enough to collect a series of reporter modified BAC clones by the transposon tool.

We then selected four reporter modified *RP24-88H4*-BAC clones (#166, #207, #208, #209; Fig. 1b) and generated transgenic mouse lines from each BAC to evaluate the reporter activity in vivo. As the results, we confirmed that all these four clones recapitulated exactly the same reporter expression patterns along differentiating Schwann cells (white arrows in Fig. 1c–f; Table 1), some other neural crest derivatives homed in branchial arches (white asterisks in Fig. 1c–f) as well as at the trunk region (Fig. 1c–f), mesenchymal cells in the limb buds (Fig. 1c–f) and cells along the dorsal midline at embryonic day e12.5 (Fig. 1c–f). This indicates that the BGZ40 reporter cassette is capable to capture some of the activities of gene regulatory elements included in the BAC clone *RP24-88H4* regardless of its integration sites.

We further evaluated to what extent the BGZ40 cassette can detect the activity of gene regulatory elements included in a given BAC clone. For this purpose, we chose another BAC clone *RP23-78N21* that covers both the *Cdh6* gene TSS-exon and ATG-exon to carry out two types of modifications on this clone: One is the transposon mediated insertion of BGZ40 cassette into the BAC clone and the other is the recombination mediated in-frame insertion of the *LacZ*-SV40pA cassette at the *Cdh6*-ATG-exon which is included in the BAC clone. In the latter configuration, the native *Cdh6* promoter element located at around the TSS is expected to activate the reporter gene in the modified BAC clone, logically capturing activities of all gene regulatory elements included in the clone after the transgenesis (Fig. 2a). As was shown in Fig. 2a, we prepared one transposon modified BAC clone and one homologous recombination modified BAC clone to generate transgenic mouse lines from each BAC-reporter construct. As the consequence, we found that e12.5 transgenic embryos from these two modified BACs recapitulated very similar reporter expression patterns along the neural crest derivatives at the trunk region as well as in the brainchial arches and so on (Fig. 2b, c; Table 1). Noticeably, expression levels of *LacZ* from the homologous recombination modified construct tended to be higher than those from the transposon modified construct probably due to the promoter interference
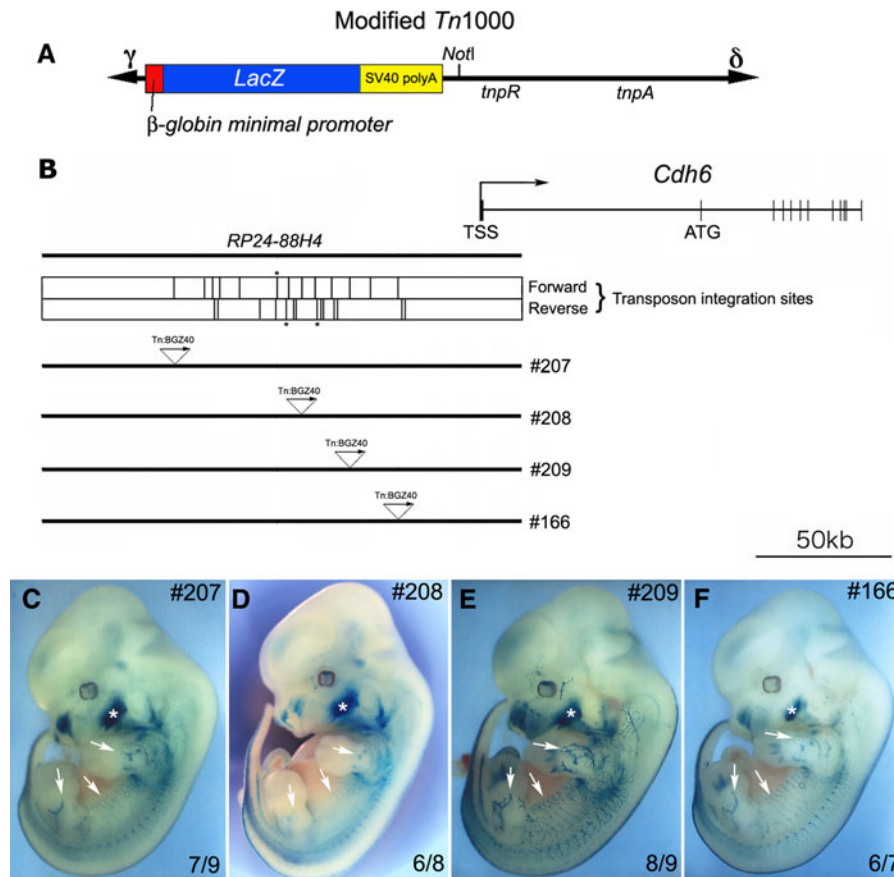
Fig. 1 Transposon mediated BAC modification to trap gene regulatory machineries. **a** Schematic of the modified *Tn*1000 carrying a BGZ40 cassette that contains the human beta globin minimal promoter, *Escherichia coli* beta-galactosidase gene (*LacZ*) and simian virus 40 polyadenylation signal (SV40 polyA). The transposon segment 3′ to a restriction enzyme *Not*I recognition site encodes two transposase genes *tnpR* and *tnpA* that are sufficient for the transposition activity. The end sequences γ and δ are required targets for transposon cleavage and maintained after the transposition, thus providing starting points for direct sequencing to determine the integration site. **b** An example of transposon mediated BAC modification. A BAC clone *RP24-88H4* shown by a *bold line* that covers 5′ territory of mouse cadherin-6 (*Cdh6*) gene locus was modified by the transposon. Size of the BAC clone and *Cdh6* gene can be compared with the 50 kilo-base pair (kb) *reference line* at the *bottom right corner* of the *panel*. *Short vertical lines* indicate positions of *Cdh6* exons along the genome. *TSS* Transcription start site, *ATG* Translation start codon. *Vertical lines* in *rectangles* put under the *bold line* represent sites for transposon integrations in the BAC clone *RP24-88H4*. Forward, transposon integration occurred in the forward direction to *Cdh6* gene; Reverse, transposon integration occurred in the reverse direction to *Cdh6* gene. *Asterisks* indicate integration hot spots where *RP24-88H4* got the transposition more than two times among sequenced 32 clones. Underneath the summary, four modified BAC clones selected for transgenesis were further depicted with their identification number. *Tn* Transposon. **c–f** Representative *LacZ* expression pattern from each BAC transgenic mouse line appears indistinguishable at e12.5. Identification number of the injected modified BAC clone is shown at the *upper right corner* of *each panel*. *White arrows*, expression along the differentiating Schwann cells; *white asterisks*, expression in the branchial arch. The ratio of embryos exhibiting reporter expression in the Schwann cells over the total number of independent transgenic mouse lines generated is shown at the *right bottom corner* of each *panel* (**c–f**) and in Table 1

between endogenous *Cdh6* promoter and BG minimal promoter and/or the intron-less feature of the BGZ40 cassette (Fig. 2b, c). The result demonstrates that the human beta globin minimal promoter in the BGZ40 cassette is good enough to detect virtually all activity of gene regulatory elements included in a given BAC clone after the transgenesis in mice.

Table 1 Summary of Tg mouse lines analysed in the present study

Figure	BAC transgene	Independent Tg mouse lines generated	Tg with the reproducible expression pattern as in Figures	Tg with the expected expression pattern
1-C	*RB24-88H4*, Tn:BGZ40 #207	9	7	7[s]
1-D	*RB24-88H4*, Tn:BGZ40 #208	8	6	6[s]
1-E	*RB24-88H4*, Tn:BGZ40 #209	9	8	8[s]
1-F	*RB24-88H4*, Tn:BGZ40 #166 (Fig. 3b #1)	7	6	6[s]
2-B	*RP23-78N21*, Rec:LacZpA	7	6	6
2-C	*RP23-78N21*, Tn:BGZ40	2	2	2
3-B	*RB24-88H4*, Tn:BGZ40 #166 Δ36 kb (#2)	7	7	0[s]
3-B, 3-E	*RB24-88H4*, Tn:BGZ40 #166 Δ10 kb (#3)	3	2	2[s]
3-B	*RB24-88H4*, Tn:BGZ40 #166 Δ10 kb (#4)	7	7	0[s]
3-B, 3-F	*RB24-88H4*, Tn:BGZ40 #166 Δ400 bp (#5)	4	4	0[s]

S Expression in Schwann cells, *Tg* Transgenic

Homologous recombination mediated systematic deletions from the transposon-modified BAC clones pinpoint the gene regulatory modules

We next sought to establish a reliable assay system in which systematic deletions of given genomic fragments from the reporter modified BAC clone by means of homologous recombination in bacterial cells. Since the transposon modified BAC *RP24-88H4* clones fairly recapitulated the Schwann cell specific expression of *Cdh6* gene at e12.5 (Fig. 1c–f), we tried to confine which genomic fragments in the BAC clone is required for the restricted gene expression. In the previous report, we selected five BAC clones that differentially cover the ∼350-kb genomic territory of huge *Cdh6* gene locus and found that each BAC with reporter modification fairly recapitulated distinct combinations of *Cdh6* mRNA expression profiles after the transgenesis (Inoue et al. 2008). From these results, we had already speculated that ∼36-kb genomic fragment covered with BAC *RP24-88H4* might be critical for *Cdh6* expression in Schwann cells (Inoue et al. 2008). We thus designed to generate a BAC based reporter construct in which ∼36-kb genomic fragment is completely deleted from a modified *RP24-88H4* clone #166 and examined if ∼36-kb genomic region is required for the reporter expression in Schwann cell populations after the transgenesis. For the deletion, we subcloned 1-kb 5′ upstream and 1-kb 3′ downstream fragments from the target region as homologous arms for recombination and further placed an Amp-r cassette

in between the arms. The configuration of this construct logically allows precise excision of the ∼36-kb genomic fragment from the BAC clone after the heat shock inducible homologous recombination event in the recombinogenic *Escherichia coli* strain EL250 (Fig. 3a, b). As the results, we obtained the right BAC-reporter construct with high efficiency (i.e., 4 colonies out of randomly selected 6 colonies contained the correctly modified clone), which was verified by PCR and pulse field gel electrophoresis (PFGE; Construct #2 in Fig. 3b, c).

To further make sure if any types of homologous recombination mediated BAC deletion can be processed with reasonable efficiency, we tried to generate several series of deletion constructs in which various size and/or combinations of deletions are introduced into the reporter modified BAC clones (Fig. 3b). Consequently, we confirmed by means of PCR and PFGE that any types of deletions can equally be placed into a given BAC clone (Fig. 3c, d). For instance, recombination efficiency for ∼400-base pairs (bp) and 77-kb deletions appeared to fall at a similar level (i.e., 4–5 colonies out of randomly selected 6 colonies contained BAC-reporter constructs with right deletions) in utilizing 1-kb-sized arms. Incidentally, arms as short as 100-bp yielded only small number of positive colonies, indicating that length of the recombination arms might be a critical factor for the homologous recombination efficiency (data not shown). Deletions at two positions could also be achieved by sequentially using two types of selection markers (Amp-r and Kan-r; Fig. 3b, d).
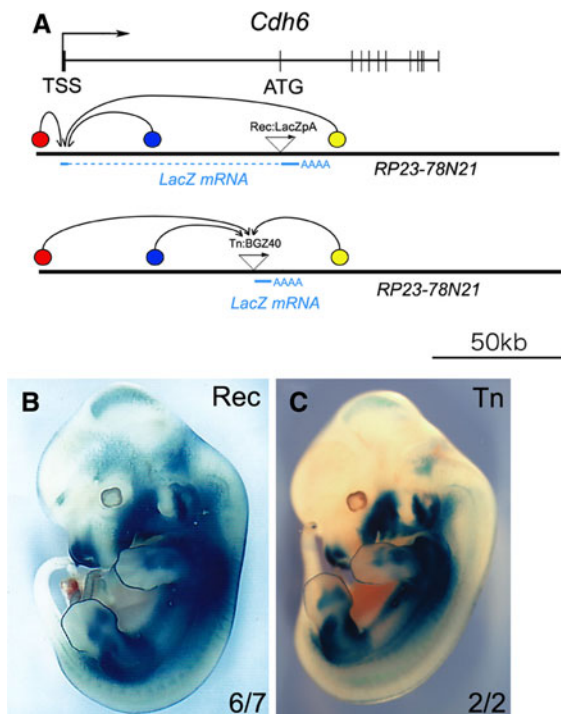
Fig. 2 Comparison of BAC modification strategy with reporter cassette. a *Top Cdh6* gene structure is shown as Fig. 1. *Middle* Depicted is a BAC-*LacZ* reporter construct generated by the homologous recombination (Rec) mediated in-frame introduction of a *LacZ* reporter cassette into the *Cdh6* ATG exon. After the transgenesis of this construct, three hypothetical gene regulatory elements (*colored circles*) included in the BAC activate the reporter expression by interacting with the endogenous *Cdh6* gene promoter that locates around the TSS exon. Note that *LacZ* mRNA (*light blue line*) is generated via splicing as *Cdh6* mRNA. *Bottom* Shown is a BAC-*LacZ* reporter construct generated by the transposon (Tn) mediated integration of BGZ40 cassette. After the transgenesis of this construct, three hypothetical gene regulatory elements (*colored circles*) included in the BAC activate the reporter expression by interacting with the human beta-globin minimal promoter in the transposon. Note that *LacZ* mRNAs (*light blue line*) is directly transcribed from the BGZ40 cassette without splicing. b, c Representative *LacZ* expression patterns from each BAC transgenic mouse line appear identical at e12.5. Rec, Transgenesis of the homologous recombination modified BAC-reporter construct (*upper construct in panel* a); *Tn* Transgenesis of the transposon modified BAC-reporter construct (*lower construct in panel* a). The ratio of embryos exhibiting the expected reporter expression pattern over the total number of independent transgenic mouse lines generated is shown at the *right bottom corner* of each *panel* (b, c) and in Table 1

We finally purified and transferred the BAC construct #166 with ~36-kb deletion (=Construct #2 in Fig. 3b) into mouse fertilized eggs to obtain

transgenic mouse lines. As the results, we found that the reporter expression in Schwann cell populations entirely disappeared from these mouse lines, while the expression in other regions such as neural crest derivatives in the trunk and cranial branchial arch was unchanged at e12.5 (Table 1; T. Inoue et al., manuscript in preparation). This indicates that the 36-kb territory is indeed required for the *Cdh6* expression in Schwann cell populations at e12.5. Other deletion BAC constructs except for construct #3 also abolished reporter expression in Schwann cell lineage after the transgenesis, further demonstrating that a ~400-bp module in the 36-kb territory that contains an ECR with several transcription factor binding sites is critical for the *Cdh6* expression at e12.5 (Fig. 3e, f; Table 1; T. Inoue et al., manuscript in preparation). Since the sets of BAC deletion constructs described in this study were able to be simultaneously prepared for the transgenesis within 3–4 weeks by a researcher, our strategy might provide a high-throughput screening platform for gene transcriptional machineries that dynamically fluctuate during development, pathogenesis and/or evolution.

A comprehensive strategy in characterizing gene transcriptional machineries

In the conventional method for examining gene transcriptional machineries (additive strategy: Fig. 4a), genomic fragments whose size is limited to several kilo-base pairs in plasmid constructions are linked to the reporter cassette and the reporter activities are monitored in cell lines and/or transgenic animals. In order to analyze a genomic region whose size is longer than several hundreds kilo-base pairs, one should hence prepare more than hundreds of plasmid based constructs to identify sufficient modules for spatio-temporally restricted gene expressions (Fig. 4a). While recent bioinformatics based predictions greatly help confine possible transcription factor binding modules in the genome, it is still needed to prepare tens of plasmids in verifying the physiological significance. Moreover, this strategy cannot identify those gene regulatory modules that are separated each other into a large genomic region yet synergistically work together for the gene activation/silencing.

From our experimental results described above, we now propose a systematic way (subtractive strategy: Fig. 4b) for narrowing down transcriptional regulatory
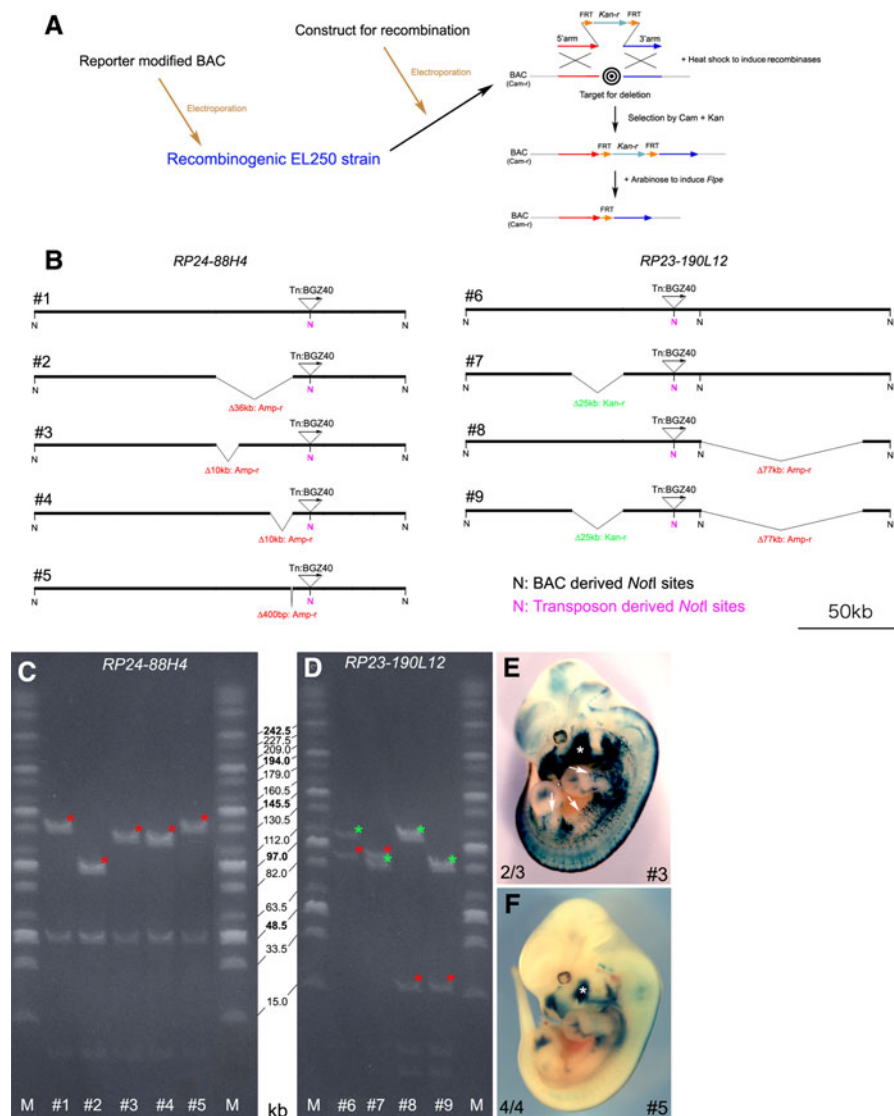
Fig. 3 BAC-reporter construction with efficient homologous recombination mediated deletions. **a** The flow chart to introduce deletions into BAC-reporter construct by homologous recombination. *Cam* Chloramphenicol, *Kan* Kanamycin, *Flpe* Flipase, *FRT* Flpe recombination target, *r* resistant gene. **b** Examples of BAC-reporter deletion constructs generated by homologous recombination. Constructs #2–#5 were generated from the transposon tagged RP24-88H4 (Construct #1 is identical to Construct #166 in Fig. 1b), while Constructs #7–#9 were originated from the transposon tagged RP23-190L2 (Construct #6). The size of deletion is depicted with the selection cassette used for the homologous recombination. Amp-r, ampicilin resistant gene cassette; Tn:BGZ40, integrated *LacZ* reporter cassette via transposition. **c, d** Band patterns confirmed by pulse field gel (PFG) electrophoresis. Each BAC-reporter construct was digested with *Not*I and the fragment size was estimated by comparison with the PFG markers (M) whose size components were indicated between the *panels*. *Bold*

characters correspond to the *bold marker bands*. In *panel* **c**, a band in *lane* #1 referred by *red asterisk* is modified by Amp-r integration. In *panel* **d**, a band in *lane* #6 referred by *red asterisk* is modified by Amp-r integrations, while that in *lane* #6 referred by *green asterisk* is modified by Kan-r integrations. Note that bands referred by corresponding asterisks in *lanes* #2–5 and #7–9 correctly shift with BAC modification patterns summarized in *panel* **b**. **e, f** Representative reporter expression patterns after the transgenesis of Construct #3 (**e**) and #5 (**f**). Note that the reporter expression in differentiating Schwann cells (*white arrows*) was totally abolished in BAC transgenic mouse embryos generated from Construct #5 (**f**), indicating value of the deleted 400-base pair (bp) sized region for the expression. *White asterisks*, expression in the branchial arch. The ratio of embryos exhibiting reproducible reporter expression over the total number of independent transgenic mouse lines generated is noted at the *left bottom corner* of each *panel* (**e, f**) and in Table 1

Fig. 4 Two strategies to examine gene transcriptional machinery in large genomic territories. **a** A conventional strategy with multiple reporter constructions (Additive strategy). To finally identify ∼5-kb sized critical regions for gene transcriptional regulation from a ∼200-kb sized territory covered with a BAC clone, many of plasmid-based reporter constructions and their transgenesis are essential. **b** A novel methodology with systematic deletion of BAC-reporter constructs (Subtractive strategy). In order to characterize ∼15-kb sized critical regions for gene transcriptional regulation from the ∼0.5-mega-base sized territory, only six BAC based constructions and their transgenesis are sufficient. See text for details
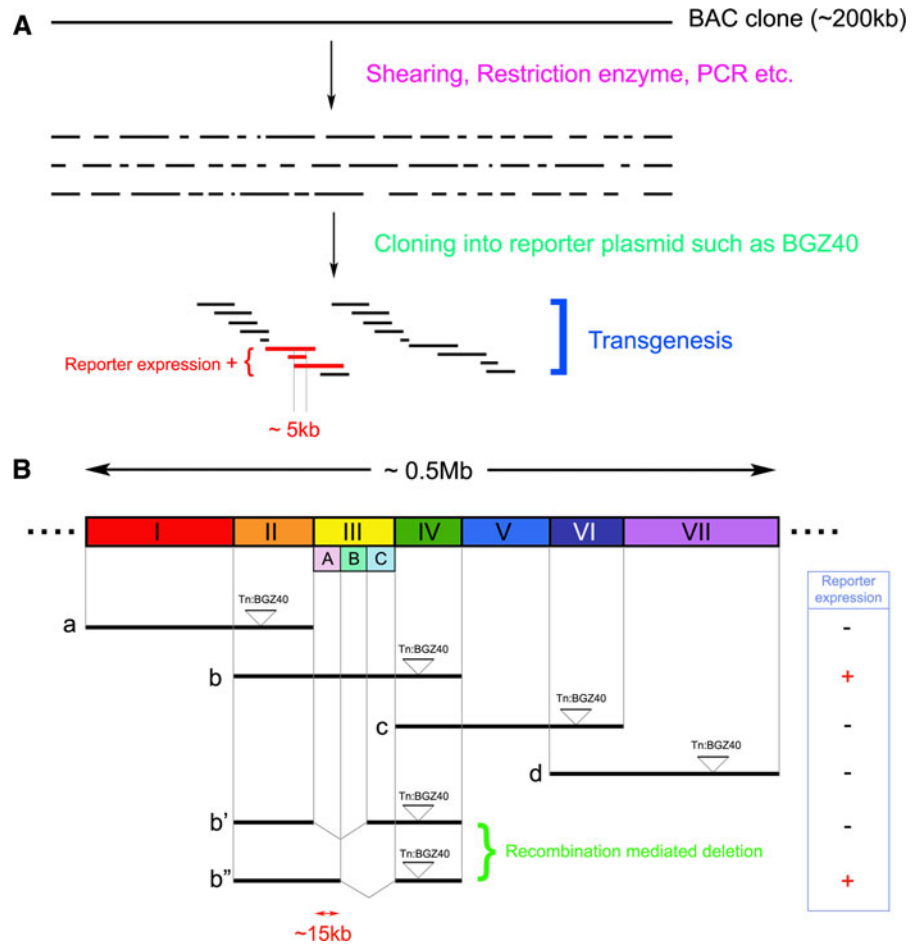
machineries into short territories from large genomic sections. As the first step, transposon mediated reporter integration into BAC contig is to be set as the analytical basis for enhancer screening. For instance, given that a ∼0.5-mega-base pair (Mb) sized genomic territory arises as the screening target for a gene regulatory machinery of interest, four BAC clones would be enough to differentially cover the territory (Constructs a–d in Fig. 4b) and it is straightforward to modify all of them by transposon tools compared to the recombination-based method that requires inclusion of transcription start sites in the target BAC clones. This could primarily capture almost all enhancer activity in these BAC clones (refer to Fig. 2) and if one can find out some of them recapitulated the gene expression patterns of interest after transgenesis, essential regulatory modules are speculated to be distributed within a ∼50-kb genomic segment (Fig. 4b; In this example, one can speculate that segment III is required for the

reporter expression because only Construct b yields the expression after the transgenesis). As the second step of evaluation, recombination mediated systematic deletions from the reporter modified BAC clone would be the most comprehensive way in identifying a critical genomic fragment from the speculated ∼50-kb genomic segment for the transcriptional regulation of interest. For instance, if two recombination mediated deletion constructs were generated from the reporter modified BAC clone with the deleted region being equally overlapped each other (Construct b' and b" in Fig. 4b), one can narrow the ∼50-kb segment into a fragment whose size is as short as the one third of the segment by examining the reporter activity in the transgenic animals (Fig. 4b; In this example, territory A in the segment III is speculated to be essential for the reporter expression because only Construct b" yields the expression after the transgenesis). It should be of note for this step that introduction of 'two' equally

overlapped deletions to a target segment of a selected BAC-reporter construct is a key to minimize experimental efforts the most: two equally overlapped deletions could narrow the critical region for transcriptional regulations into 1/3rd of the target segment (Fig. 4b), and two more cycles of experiments with similar deletion approaches would further confine the region into $(1/3)^3 = 1/27$th of the target segment. On the other hand, three equally overlapped deletions could confine the regulatory regions into 1/5th of the target segment and one more cycle of experiment would restrict the regions into $(1/5)^2 = 1/25$th of the target segment. Considering the facts that both examples described above require six BAC-reporter deletion constructs for the screening and that more overlapped deletions need more constructions, two-overlapped-deletion based screen is certainly the most efficient way to finally pinpoint the gene regulatory modules.

Taken together, only six BAC-reporter constructs (Constructs a–d, b' and b" in Fig. 4b) and their transgenesis stand logically sufficient to narrow the 0.5-Mb genomic territory into ~15-kb genomic fragment for gene transcriptional regulation (Fig. 4b) and here proposed new strategy indeed allowed us to identify many of *Cdh6* gene regulatory modules which cannot easily be examined by the conventional method for its huge and complex genomic organization (Fig. 3; our unpublished results).

Generally, the subtractive way of screening in characterizing critical factor(s) from tens of pooled candidates is a clear-cut and efficient option. For the immediate example, the four Yamanaka-factors that cooperatively induce pluripotent stem cells from somatic fibroblasts have been comprehensively identified by simultaneous gene transfections of any 39 factors among the pooled 40 candidate factors (Takahashi and Yamanaka 2006). In this case, only 40 rounds of experiments with 'subtractive' pools quickly solved out the four indispensable factors from possible trillions of combinations (=2^{40}). Hundreds of candidate gene regulatory modules scattered over a given genomic territory can be considered as the analogous situation. Notably, BAC clones derived from human or chick genomic library can similarly be modified and evaluated in mouse system after the transgenesis (our unpublished observation). Spitz et al. (2003) and Jeong et al. (2006) have already reported similar screen strategies for gene regulatory regions, yet the present study further confirms the

utility of BAC modification tools in evaluating gene regulatory elements (Figs. 1, 2, 3) and improves the way of evaluation that never primarily depends on weighting with bioinformatics and/or earlier experimental evidence-biased predictions (Fig. 4). Kokubu et al. (2009) have also innovated the transposon-based mouse chromosomal engineering method to broadly map *cis*-regulatory landscapes for any gene clusters. While this strategy surpasses transgenic-based systems including ours in the point that it allows integration of the reporter cassette into the natural genomic context, it would be time/money-consuming to attain the kilo-base-pair-sized analytical accuracy only with their local-hopping-transposon-tagging and deletion strategy in mouse embryonic stem cells. Our BAC-transgenic based methodology with systematic deletions (=subtraction of gene regulatory modules) in bacterial cells might complement the inventive Kokubu's method and would thus provide ideal foundations in assessment of any detailed gene regulatory landscapes most of which remain elusive.

Conclusions

The great majority of vertebrate genome is occupied by the non-coding DNA sequences that are thought to be involved in regulation of chromosomal structures and/or differential gene expressions. Mobile DNA elements and some other repetitive DNA sequences resident in the non-coding regions are also known to affect gene expression profiles. Recent studies have further revealed that human copy number variations as well as single nucleotide polymorphisms in the non-coding regions tightly link to pathogenic conditions. It must therefore be essential to be able to systematically annotate these non-coding DNA sequences and our BAC based methodology should help establish a novel high-throughput screening platform for these sequences that might harbor unspecified yet critical roles for organisms' development, pathogenesis and/or evolution.

References

Antoch MP, Song EJ, Chang AM, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS (1997) Functional identification of the mouse circadian clock gene by transgenic BAC rescue. Cell 16:655–667

Burke DT, Carle GF, Olson MV (1987) Cloning of large segment of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806–812

Carvajal JJ, Keith A, Rigby WJ (2008) Global transcriptional regulation of the locus encoding the skeletal muscle determination genes Mrf4 and Myf5. Genes Dev 22:265–276

Choudhary C, Mann M (2010) Decoding signaling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol 11:427–439

Copeland NG, Jenkins NA, Court DL (2001) Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet 2:769–779

Dufva M (2009) Introduction to microarray technology. Methods Mol Biol 529:1–22

Inoue T, Inoue YU, Asami J, Izumi H, Nakamura S, Krumlauf R (2008) Analysis of mouse Cdh6 gene regulation by transgenesis of modified bacterial artificial chromosomes. Dev Biol 315:506–520

Jeong Y, El-Jaick K, Roessler E, Muenke M, Epstein DJ (2006) A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133:761–772

Jessen JR, Meng A, McFarlane RJ, Paw BH, Zon LI, Smith GR, Lin S (1998) Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgensis. Proc Natl Acad USA 95:5121–5126

Kelly LE, Davy BE, Berbari NF, Robinson ML, El-Hodiri HM (2005) Recombineered Xenopus tropicalis BAC expresses a GFP reporter under the control of Arx transcriptional regulatory elements in transgenic Xenopus laevis embryos. Genesis 41:185–191

Kokubu C, Horie K, Abe K, Ikeda R, Mizuno S, Uno Y, Ogiwara S, Ohtsuka M, Isotani A, Okabe M, Imai K, Takeda J (2009) A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. Nat Genet 41:946–952

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73:56–65

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405:827–836

Montigny WJ, Phelps SF, Illenye S, Heintz NH (2003) Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells. Biotechniques 35:796–807

Morgan BA, Conlon FL, Manzanares M, Millar JB, Kanuga N, Sharpe J, Krumlauf R, Smith JC, Sedgwick SG (1996) Transposon tools for recombinant DNA manipulation: characterization of transcriptional regulators from yeast, Xenopus, and mouse. Proc Natl Acad USA 93:2801–2806

Natoli G (2010) Maintaining cell identity through global control of genomic organization. Immunity 33:12–24

Ovcharenko I, Nobrega MA, Loots GG, Stubbs L (2004) ECR browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res 32:W280–W286

Pandey A, Mann M (2000) Proteomics to study gene and genomes. Nature 405:837–846

Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods 5:409–415

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657

Shizuya H, Kouros-Mehr H (2001) The development and application of the bacterial artificial chromosome cloning system. Keio J Med 50:26–30

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (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 USA 89:8794–8797

Spitz F, Gonzalez F, Duboule D (2003) A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell 113:405–417

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676

Theiler K (1972) The house mouse: development and normal stages from fertilization to 4 weeks of age. Springer, Berlin

Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, Kuhn RM, Rhead BL, Smith KE, Thakkapallayil A, Zweig AS, The ENCODE Project Consortium, Haussler D, Kent WJ (2006) The ENCODE project at UC Santa Cruz. Nucleic Acids Res 35:D663–D667

Wade-Martins R, Smith ER, Tyminski E, Chiocca EA, Saeki Y (2001) An infectious transfer and expression system for genomic DNA loci in human and mouse cells. Nat Biotechnol 19:1067–1070

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