Mamm Genome (2007) 18:693–708
DOI 10.1007/s00335-007-9056-y

Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression

Kelly J. Chandler · Ronald L. Chandler · Eva M. Broeckelmann · Yue Hou · E. Michelle Southard-Smith · Douglas P. Mortlock

Received: 11 April 2007 / Accepted: 27 July 2007 / Published online: 20 September 2007

Abstract Bacterial artificial chromosomes (BACs) are excellent tools for manipulating large DNA fragments and, as a result, are increasingly utilized to engineer transgenic mice by pronuclear injection. The demand for BAC transgenic mice underscores the need for careful inspection of BAC integrity and fidelity following transgenesis, which may be crucial for interpreting transgene function. Thus, it is imperative that reliable methods for assessing these parameters are available. However, there are limited data regarding whether BAC transgenes routinely integrate in the mouse genome as intact molecules, how BAC transgenes behave as they are passed through the germline across successive generations, and how variation in BAC transgene copy number relates to transgene expression. To address these questions, we used TaqMan real-time PCR to estimate BAC transgene copy number in BAC transgenic embryos and lines. Here we demonstrate the reproducibility of copy number quantification with this method and describe the variation in copy number across independent transgenic lines. In addition, polymorphic marker analysis suggests that the majority of BAC transgenic lines contain intact molecules. Notably, all lines containing multiple BAC copies also contain all BAC-specific markers. Three of 23 founders analyzed contained BAC transgenes integrated into more than one genomic location. Finally, we show increased BAC transgene copy number correlates with increased BAC transgene expression. In sum, our efforts have provided a reliable method for assaying BAC transgene integrity and fidelity, and data that should be useful for researchers using BACs as transgenic vectors.

K. J. Chandler · R. L. Chandler · E. M. Broeckelmann · Y. Hou · D. P. Mortlock (✉)
Department of Molecular Physiology and Biophysics, Center for Human Genetics Research, Vanderbilt University School of Medicine, 519 Light Hall, 2215 Garland Avenue, Nashville Tennessee 37232, USA
e-mail: mortlock@chgr.mc.vanderbilt.edu

E. M. Southard-Smith
Division of Genetic Medicine, Department of Medicine, Vanderbilt University School of Medicine, 529 Light Hall, 2215 Garland Avenue, Nashville, Tennessee 37232, USA

Introduction

Bacterial artificial chromosomes (BACs) have been used extensively for mouse transgenesis (Giraldo et al. 2003; Heaney and Bronson 2006; Heintz 2001). Because of their large insert size, they can often accommodate the complete structure of genes of interest, including long-range *cis*-regulatory elements required for correct tissue-specific or temporal expression. They are also thought to be more resistant to position effects than smaller transgenes (Giraldo and Montoliu 2001; Gong et al. 2003). For these reasons, they are particularly useful for studying long-range *cis*-regulatory phenomena (Chandler et al. 2007; Mortlock et al. 2003) and for experiments where precise transgene expression is critical, such as *cre-recombinase* drivers (Copeland et al. 2001; Lee et al. 2001). In addition, BACs are increasingly used for rescue experiments or overexpression studies. In general, there is little published data that provide detailed documentation for potential correlations between BAC transgene copy number, expression, and structure. More data would be useful regarding the general variation of BAC copy number in transgenic mice and how this variation impacts BAC transgene expression and/or structure. However, the large size of BACs also makes it harder to analyze transgene

structure following integration into the genome. Founder animals or their transgenic progeny can provide large amounts of DNA for Southern blot analysis, although for some developmental studies where transgenic embryos generated by pronuclear injection are analyzed "transiently," little DNA (e.g., from yolk sacs) is usually available for analysis. PCR-based methods, while limited in scope to analyze large-scale transgene structure, can be useful for estimating transgene copy number. Quantitative PCR (Q-PCR) can also be easily applied to many DNA samples in parallel and provides results faster than traditional Southern blotting, with similar accuracy as we and others have shown (Ballester et al. 2004).

Our laboratory uses BAC transgenes to study long-range *cis*-regulatory elements of the BMP family genes *Gdf6*, *Bmp2*, and *Bmp4*. The nature of these experiments depends on verification of BAC transgene structure following transgenesis in mice. Toward this end we have generated numerous BAC transgenic mice using standard pronuclear injection methods and with several unique BAC transgenes that were useful for documenting trends in BAC copy number and integrity across independently created transgenic mice. In this article we present a straightforward method for estimating BAC transgene copy number in multiple *Bmp2* and *Bmp4* BAC transgenic lines and embryos using quantitative real-time PCR. In all, we analyzed copy number in 78 transiently generated BAC transgenic embryos or liveborn animals created by pronuclear injection and in 317 transgenic mice from 26 separate breeding lines established from liveborn founders. Eleven distinct *Bmp2* and *Bmp4* BAC constructs were used to generate this data. To our knowledge this is the most extensive analysis to date of copy number in BAC transgenic mice. Our method relies on comparing data from transgenic samples to a standard curve of calibrator samples that are generated by diluting purified BAC DNA over a range of known concentrations into wild-type mouse genomic DNA. This method is robust, conceptually simple, and amenable to processing large numbers of purified tail DNA samples in parallel. Our data allowed us to confirm stability of several BAC transgenic lines through germline transmission and to correlate copy number with strength of transgene expression. We also observed that transgenic lines carrying multiple BAC copies most likely carry one or more full-length BAC molecules. In general, transgene copy number was fixed in subsequent generations following germline transmission; however, we noticed several examples of striking discrepancies between founder copy number estimates and their transgenic progeny. We also clearly identified several founder animals that each transmitted two independently segregating transgene insertions. Although BACs are extremely useful as transgenic vectors and it is very feasible to create transgenic BAC lines that carry multiple, complete BAC molecules, BAC fragmentation and integration of BACs into separate genomic locations were observed at a frequency of 17% (3/18) and 12% (3/26), respectively. In summary, the monitoring of BAC transgene copy number can add useful information when interpreting BAC transgene expression and confirming stability of integrations through the germline.

Materials and methods

Transgenic mice

Bacterial artificial chromosome (BAC) vectors were modified using homologous recombination in *E. coli* essentially as previously described (Mortlock et al. 2003) to contain a *lacZ*:Neo (β-geo) fusion cassette into the *Bmp2* or *Bmp4* transcription unit. Briefly, mouse *Bmp2* BACs RP23-85O11 (239,101 kb) and RP23-409L24 (209,640 kb) were modified as previously described (Chandler et al. 2007). Mouse *Bmp4* BACs RP23-26C16 (227,097 kb) and RP23-145J23 (227,220 kb) were modified by inserting a GFP-(IRES)-β-geo cassette into the ATG start codon of *Bmp4*. Purified BAC DNA constructs were used for pronuclear injections to generate founder embryos or lines as previously described (Chandler et al. 2007). In brief, BAC DNA was harvested from 1 L of bacterial culture by alkaline lysis and subsequently purified using cesium chloride density centrifugation. Ethidium bromide was removed from purified BAC DNA by butanol extractions and BAC DNA was dialyzed against 3 L of microinjection buffer (10 mM Tris-HCl [pH 7.4], 0.15 mM EDTA [pH 8.0]) using 10,000-molecular-weight-cutoff Slide-A-Lyzer dialysis cassettes (Pierce) followed by additional dialysis and concentration of BAC DNA with 30,000-molecular-weight-cutoff Centriprep centrifugal filter devices (Millipore) to reduce the DNA solution final volume to less than 500 µl. BAC DNA samples were quantified as follows: digests with a rare-cutting restriction enzyme (*Nru*I) were analyzed by pulsed-field gel electrophoresis in 1% agarose/0.5× Tris-Borate-EDTA buffer for 15 h at 18°C (6 V/cm, 0.2–22-s switch time) alongside known quantities of lambda DNA-*Hin*dIII digests as mass standards. To determine BAC DNA concentration, the gel was stained with ethidium bromide and Quantity One® software (Bio-Rad) was used to quantify BAC DNA bands by comparison to a standard curve of the lambda DNA-*Hin*dIII band intensities. The stock concentration of uncut BAC DNA was back-calculated based on these estimates. Purified, circular BAC DNA was diluted to 1 ng/µl in microinjection buffer and used for pronuclear injections.

DNA isolation

DNA was extracted from mice tail biopsies or embryonic yolk sacs by overnight digestion in 500 µl of proteinase K buffer (10 mM Tris-HCl [pH 8.0], 100 mM NaCl, 10 mM EDTA [pH 8.0], 0.5% sodium dodecyl sulfate, 0.25 mg/ml proteinase K) with occasional vortexing. Following digestion, 250 µl of phenol and 250 µl of chloroform were added followed by vigorous vortexing to ensure thorough mixing of phenol:chloroform with the sample. Samples were immediately subjected to microcentrifugation at 16,249 rcf for 4 min to allow separation of the aqueous and organic layers. The aqueous layer was removed with a wide-bore pipette tip, paying careful attention to avoid the interface. Ethanol precipitation of the aqueous layer was performed and DNA pellets were washed with 70% ethanol followed by resuspension overnight in 200 µl (tail DNA) or 100 µl (yolk sac DNA) of TE (pH 7.4). Genomic DNA samples were quantified on a UV spectrophotometer at 260 nm and diluted to 10 ng/µl for real-time PCR.

Standard curve samples for real-time PCR

To create a standard curve of real-time PCR data from known amounts of BAC template, supercoiled BAC DNA was isolated by cesium chloride density centrifugation and quantified via gel electrophoresis, by comparing intensity of restriction-digested BAC DNA bands to lambda DNA-*Hin*dIII mass standards as described for the preparation of BAC DNA for pronuclear injections. Then, twofold dilutions of BAC DNA were spiked into 10 ng/µl genomic DNA (final concentration) that had been isolated from a C57BL6J × DBA2J F1 mouse liver by methods described above and quantified by UV spectrophotometry at 260 nm. This created a series of standard samples such that the ratio of BAC molecules ranged from approximately 1 to approximately 48 BAC copies per diploid mouse genome. Copy number standards were exposed to at least one freeze-thaw cycle prior to use, since tail and yolk sac DNA samples were also freeze-thawed before analysis.

Real-time PCR

Custom TaqMan® Assays-by-Design were used to generate primer and probe sets for Neo (present in β-geo fusion gene) and the mouse *Jun* gene (control) Applied Biosystems Inc. (ABI) Assay IDs 185300786 and Mm00495062_s1. The following primer pairs and probes were used: for Neo assay, forward primer: (5′-ATGACTGGGCACAACAGACAAT-3′); reverse primer: (5′-CGCTGACAGCCGGAACAC-3′); probe: (5′-FAM-CTGCTCTGATGCCGC-3′); for *Jun* assay, forward primer: (5′-GAGTGCTAGCGGAGTCTTAACC-3′); reverse primer: (5′-CTCCAGACGGCAGTGCTT-3′); probe: (5′-FAM-CTGAGCCCTCCTCCCC-3′).

Real-time PCR was performed on a GenAmp 9700 thermocycler and plates were scanned using the ABI PRISM® 7900HT sequence detection system. Two microliters (20 ng) of genomic DNA samples or copy number standards were analyzed in a 10-µl reaction volume with either primer-probe set (Neo, *Jun*). In addition, no-template controls were included in each experiment. All reactions were performed in duplicate or triplicate.

Copy number estimation

Copy number estimates were derived from ΔCt values for standard curve samples. To calculate ΔCt values, the average of duplicate Ct values generated with the Neo probe was subtracted from the average *Jun* Ct value. Using the scatterplot chart function in Microsoft Excel, ΔCt values for each standard were plotted (on the *Y* axis) against the known copy number of each standard (on the *X* axis) using a logarithmic scale. A logarithmic regression trendline and its corresponding equation were then generated to fit the slope. The resulting equation of the form $y = (\text{slope})\ln(x) + y$ intercept was used to estimate copy number of samples based on the ΔCt value. To solve for copy number (x), the base of the natural logarithm was raised to the power of x and multiplied by 2 to account for a diploid genome (estimated copy number = $2e^{((\Delta Ct\text{-}y\ \text{intercept})/\text{slope})}$).

Quantitative dot-blot hybridization

Genomic DNA samples were extracted from liver samples isolated from liveborn transgenic mice using standard genomic liver DNA isolation methods described above. Copy number values for the IRES-β-geo cassette were confirmed by dot-blot Southern hybridization using the following method: Copy number estimates were derived from standard curve samples. Standard curve samples were C57BL/6J × DBA/2J F1 hybrid genomic DNA samples spiked with known quantities of pIBG-Ftet plasmid DNA samples diluted to copy number equivalents (1, 2, 4, 8, 16, 32, 64, and 128 copies per diploid genome). Fifty microliters of standard curve and genomic DNA samples containing 10 µg of total DNA was added to 150 µl of denaturing solution (0.01 M EDTA [pH 8.0], 0.53 N NaOH). Samples were incubated at 95°C for 5 min and then placed on wet ice for 2 min. A Zeta-Probe GT membrane (Bio-Rad) was briefly washed twice with H_2O then once with 0.4 N NaOH for 5 min. The prewashed membrane was placed on a 96-well Minifold Vacuum Filtration Manifold apparatus (Schleicher

and Schuell), and the apparatus was assembled according to the manufacturer's instructions. Denatured DNA samples (200 µl total volume for each) were loaded onto the vacuum manifold and incubated for 30 min at room temperature. Following incubation, samples were vacuum filtered for 5 min until all of the samples had passed through the membrane. The membrane was then neutralized with 0.2 M Tris-HCl [pH 7.5], 2× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate) for 10 min and baked for 30 min at 80°C. Control genomic probe used in hybridizations was generated by PCR amplification of mouse genomic DNA using primers specific to the 3′ UTR of mouse *Gdf6* (forward prime: 5′-AAGCATGGAAAGAGGATGAAAGGG-3′; reverse primer: 5′-CGACCTCCAGTAACTTTAGTGTTGTCA-3′) and subsequent cloning into pCRII-TOPO (Invitrogen) followed by restriction enzyme digestion with *Not*I and *Spe*I to isolate an approximately 937-bp fragment. The transgene-specific probe used in hybridizations was generated from a 4.7-kb *Xba*I fragment containing the IRES-β-geo cassette from pIBG-Ftet (described above). Both control and transgene-specific probes were labeled using Ready-to-Go labeling beads (Amersham) and [$^{\alpha\text{-}32}$P]dCTP (Amersham). For control probe hybridizations, the membrane was washed twice with sterile H_2O and hybridized for 3 h with Rapid-Hyb buffer (Amersham) according to the manufacturer's instructions. The membrane was then exposed to a Kodak phosphor-imaging screen for 5 days and imaged using a Pharos FX imaging system and Quantity One® software (Bio-Rad). Immediately following exposure, the membrane was placed in strip buffer (10 mM Tris-HCl [pH 7.6], 1 mM EDTA [pH 8.0], 1% sodium dodecyl sulfate) and boiled for 10 min to remove bound probe. The membrane was neutralized and baked (described above), then processed for transgene-specific probe hybridization. Again, the membrane was washed twice with sterile H_2O and hybridized for 3 h with Rapid-Hyb buffer (Amersham) according to the manufacturer's instructions. The membrane was then exposed to a Kodak phosphor-imaging screen for 16 h and imaged using a Pharos FX imaging system and software (Bio-Rad). For both control probe and transgene-specific hybridizations, triplicate standard curve and genomic DNA samples were measured using Quantity One software. Copy number estimates were derived from standard curve samples. Control genomic probe hybridizations were used to calibrate total input DNA.

Preparation of agarose-embedded high molecular-weight DNA from BAC transgenic embryos

Mouse embryonic fibroblasts (MEFs) used for the generation of agarose-embedded high-molecular-weight DNA were isolated and cultured from e13.5 embryos generated by crossing BAC transgenic males with wild-type Crl:CD1(ICR) females as described previously (Chandler et al. 2007). Cells were embedded in 0.75% low-melting-point agarose (Invitrogen) using agarose plug molds (Bio-Rad) prior to restriction enzyme digestion and pulsed-field gel electrophoresis.

Southern analysis of high-molecular-weight transgenic DNA

Agarose plugs (isolated as described above) were washed twice in 50 ml of TEX buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA [pH 8.0], 0.01% Triton X-100) and once in 50 ml of TE (pH 8.0) (10 mM Tris-HCl [pH 8.0], 1 mM EDTA [pH 8.0]) at room temperature with agitation. Plugs were then transferred to 2-ml screw cap tubes (1 plug per tube) and equilibrated for 30 min at room temperature in 2 ml of 1× restriction enzyme buffer (NEB) containing 10 mM spermidine trihydrochloride with agitation. Once equilibrated, the solution was replaced with 800 µl of 1× restriction enzyme buffer containing 10 mM spermidine trihydrochloride and 200 units of *Mlu*I (NEB) and incubated for 6–8 h at 4°C on a three-dimensional rotator (Lab-line) to allow the enzyme to infiltrate the agarose plug. After 6–8 h, the tubes were transferred to a 37°C incubator and incubated for 12–16 h with agitation. Plugs were then washed twice in 2 ml of TEX buffer at 4°C for 30 min each with agitation and equilibrated at room temperature in 2 ml of 0.5× Tris-Borate-EDTA gel electrophoresis buffer. Restriction-digested high-molecular-weight DNA was then resolved by pulsed-field gel electrophoresis in 1% agarose/0.5× Tris-Borate-EDTA buffer for 20 h at 15°C (6 V/cm, 6–80-s switch time). DNA fragments were depurinated with 0.25 M HCl for 20 min and transferred by alkaline capillary transfer onto a Zeta-Probe GT membrane (Bio-Rad) with 0.4 N NaOH for 24 h. The membrane was neutralized with 0.2 M Tris-HCl [pH 7.5], 2× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate) for 10 min and baked for 30 min at 80°C. Probe used in hybridizations was generated from a 4.7-kb *Xba*I fragment containing the IRES-β-geo cassette from the pIBG-Ftet (described above) and was labeled using Ready-to-Go labeling beads (Amersham) and [$^{\alpha\text{-}32}$P]dCTP (Amersham). Membranes were washed twice with sterile H_2O and hybridized for 3 h with Rapid-Hyb buffer (Amersham) according to the manufacturer's instructions. Membranes were exposed to a Kodak phosphor-imaging screen for 16 h and imaged using a Pharos FX imaging system and Quantity One software.

Expression analysis of transgenic mice

To generate embryos for XGal staining, transgenic male mice were crossed to Crl:CD1(ICR) female mice to obtain

Table 1 Primer sequences used for polymorphic marker analysis

Target	Primer name	Forward primer (5′-3′)	Reverse primer (5′-3′)
5′ Flank	*D14Mit60*	AGGCTGCCCATAAAAGGG	GTTTGTGCTAATGTTCTCATCTGG
	D14Mit141	CCAGCATTCCGAAGTCATTT	AGGGAAAGAAGACAGCACGA
3′ Flank	*D14Mit56*	TGGCAAAGTTTTTTTTTTCCC	TCTGGGTAGAACTGTAATAGCACA
	D14Mit212	AACATGTGCACTGGAACAATG	TCATTTATCAATTTACTTTGGTGAGG
5′ Internal	C8	AGATACTCTAGCTGGGGC	GCTGTGCACGATTGTTA
	E8	CAATCCCCAGCTCAAAAC	GGAAGGTAGCTTTCCATC
	A9	CCATTACCCAGTCATGAC	AAGTAAGCCATTGCCTC
	C9	ACAGCTCACAGTTTGAGC	AGGTGTGTGAACTTGAAC
	E9	CAGGGTATCAACAGGAAC	CATGTAGCTAAATCTTGCC
	G9	CTGATGCTTCAAGTTACAC	CAAAGTTCCTTCTGAGGT
	C11	ACAGCAAAGGTCTCAGAC	GGGGTTTCAGCTCAGTAA
	E11	CTTGGCCCATTTCTTTAC	AGTGTGCATGTATGTGCA
Overlap	G6	TAGCTCCAGCACTTTGG	CAGAAGACAAGGTCATTCT
	A7	TGAGGGACAAGCAGTAGT	TTACAGCCTCCAATCCA
3′ Internal	A1	CATGTGAGATCTAGGCTC	CAGGCTGATAGTTCCTAAG
	E1	AGAACACTGGCTGCTCTT	GCTTGCTTGTATGTCATG
	G1	AGCAACAGCATCTTCTGG	GATGGCACTCATGCACTC
	A2	GGTATCTGCATACACATGC	CCAAACAGTGACCACTTT
	A3	GTTGAGATTCTATTGTCCC	GTCTCAGAAATGTTGAGAAG
	E3	GTCTCAGAAATGTTGAGAAG	ACGGAATTATTGGTAGCC
	G3	AGAAACCCATAGGGCTG	AGATGAGTGTTCCCCTTA
	1	GTACGTGTTTCTCAGACTC	CTGATTTGAGTTTCCTATC
	11	GTCCTCCATTTCTTCTT	GGCTCGATACAGAAAGCT
	E5	TTTCAACCATGAGTGGT	CATACACACTTGCATGCT
	A6	GGCATGGCATACACACTA	CGCCTGGTAGGATGTACT

timed pregnancies. Pregnant mice were killed by CO_2 inhalation and their embryos were harvested for XGal staining to detect *lacZ* expression. In brief, embryos were dissected into 1× phosphate-buffered saline (PBS) on ice then fixed with 10% neutral buffered formalin at 4°C with agitation. Embryos older than e14.5 were bisected to allow for reagent penetration after fixation. Next, embryos were processed for XGal staining essentially as described (DiLeone et al. 1998) with two minor changes: (1) 0.6 mg/ml XGal was used and (2) embryos were stained overnight at room temperature with agitation.

Polymorphic marker analysis of *Bmp4* BACs

Polymorphic marker analysis was performed along the length of each BAC using primers designed to amplify simple tandem repeats (STRs) that are polymorphic between C57BL/6J and DBA/2J strains. Because the BACs were derived from the C57BL/6J strain and the transgenic founders were (C57BL/6J × DBA/2J) F_2 hybrids, we could identify some transgenic founders that were fortuitously homozygous for DBA/2J alleles at the *Bmp4* locus by genotyping STR markers flanking the *Bmp4* region but outside the BACs. For some lines, animals were backcrossed to DBA/2J mice to generate the required DBA/2J homozygotes. Once we identified mice that were DBA/2J homozygotes for flanking STRs, markers designed to assay the C57BL/6J-derived BAC were used to interrogate the integrity of the transgene. The flanking markers (centromeric: *D14Mit212*, *D14Mit56*; telomeric: *D14Mit141*, *D14Mit60*) were identified using the JAX MGI database (www.informatics.jax.org). Internal STRs were identified within the BAC insert sequences obtained from the UCSC genome browser using custom software, and a subset of polymorphic STRs were identified by comparing PCR products from C57BL6/J and DBA/2J DNA samples for length differences. Primers were designed to amplify polymorphic STRs in both the flanking sequence and internal sequence (Table 1). Simple tandem repeat (STR) markers internal to the BMP BACs used in this study were identified by custom algorithms (K.M. Bradley and J.R. Smith) set to screen for tandem repeats with a sequence range of 2–6 and a minimum number of ten repeats present. Primers flanking these repeats were designed to amplify PCR products of less than 275 bp, which were then

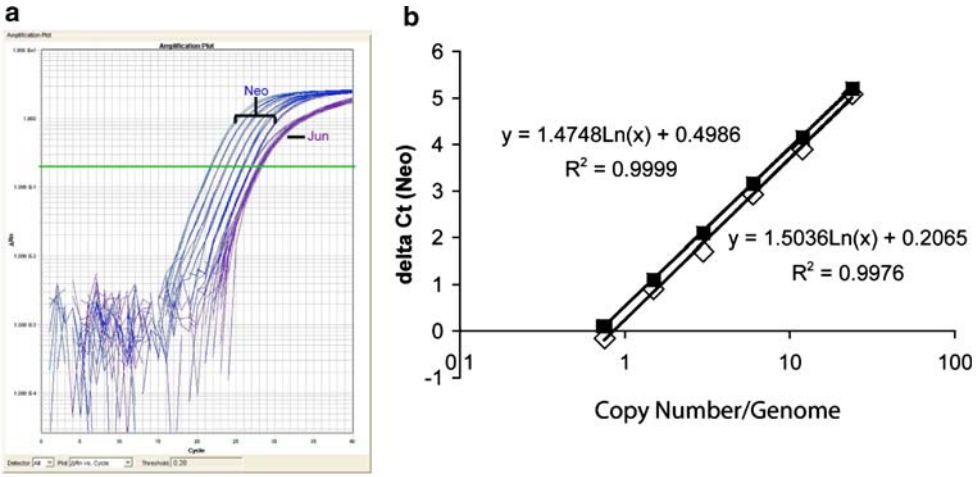

Fig. 1 BAC DNA copy number standards generate reproducible curves in real-time PCR. **a** Amplification plot depicting the Neo (blue) and *Jun* (purple) results for the copy number standards. As expected, each standard is approximately one cycle apart for the Neo assay and amplification plots are similar for all standards for the *Jun* assay. **b** Copy number standards were used to generate standard curves in real-time PCR using Neo and *Jun* primer/probe sets on two independent days (filled boxes = day 1, empty boxes = day 2). Replicate experiments indicate that the copy number standards provide highly reproducible standard curves (R^2 = 0.9999, R^2 = 0.9976) for estimating copy number of experimental samples

screened for length variations between C57BL/6J and DBA/2J mouse strains by electrophoresis on nondenaturing polyacrylamide gels (Deal et al. 2006). STR PCR products that displayed detectable variations in length were used to evaluate the presence of C57BL/6J *Bmp4* BAC sequences in transgenic animals.

Results

Validation of method for estimating BAC copy number by real-time PCR

Accurate estimation of copy number in unknown samples relies on the use of accurate standards and an effective linear range of detection. Our estimations were based on comparisons to a standard curve. We reasoned that by comparing samples of unknown copy number to a range of DNA copy number standards, each made with the same amount of mouse genomic DNA spiked with varying amounts of purified BAC DNA, we would be able to extrapolate copy number estimates in a manner that would help control for differences in amplification efficiency between PCR assays. This curve comprised real-time PCR data from standard template samples that contained two-fold dilutions of known quantities of BAC DNA in the same concentration of mouse genomic DNA. The BAC dilutions were designed to represent a range of approximately 1–48 copies of BAC molecules per diploid genome. For each standard, Ct values were generated using both an assay specific to the transgene (Neo) and an assay for a nontransgenic control gene (*Jun*). Amplification plots of the BAC copy number standards showed that all standards amplified similarly for the internal control (*Jun*) and that standards showed a stepwise one-cycle difference in Neo assay profiles, as expected (Fig. 1a). ΔCt values were plotted to generate a standard curve. Standard curves were highly similar in independent experiments with coefficients of determination (R^2) close to 1, indicating that the dilutions were made accurately, gave consistent results, and could be used to generate curve equations for estimating copy number from actual transgenic animals (Fig. 1b). Therefore, the copy number standards provided a method for generating robust standard curves.

We then tested dilutions of experimental samples to examine whether copy number estimates varied substantially depending on the amount of input genomic DNA used as template. Twofold dilutions of genomic DNA were prepared from two independent BAC transgenic animals from different lines and used as templates for real-time PCR (Fig. 2). This indicated that copy number estimates varied little over a linear range of input DNA from 4 to 32 ng. Therefore, 20 ng of genomic DNA (based on spectrophotometer readings) was used as input DNA for PCR because copy number estimates varied little with this amount of input DNA. Likewise, 20 ng of mouse DNA was used in the standard samples. We performed our initial real-time PCR experiments in triplicate and found very close data points across replicates. Therefore, we reasoned that if there was no significant difference between performing the experiment in duplicate reactions versus triplicate reactions, we would perform the remainder of our experiments in duplicate. To verify that there was no significant difference between the experiments performed in

duplicate versus triplicate, we performed a paired *t* test on an experiment encompassing 16 individual mice from several lines of varying copy numbers for which triplicate-averaged results were compared to duplicates. We found that results generated by real-time PCR performed in duplicate were not significantly different from results generated by triplicate reactions ($p > 0.14$). Therefore, we performed the bulk of our real-time PCR experiments in duplicate unless otherwise noted.

By fitting ΔCt values from experimental samples to the standard curve equation, we estimated the BAC copy number for a total of 78 transgenic founders (embryos and liveborn mice) and for 317 mice from 26 independent transgenic lines that were established from breeding some of the liveborn founders. We first looked for evidence to confirm consistency of our estimation method across samples. When estimating copy number for multiple transgenic littermates, we found copy number values were generally consistent among littermates within a line. For example, DNA samples isolated from yolk sacs or tails from littermates of independent lines produced similar copy number estimates, with minimal variability between DNA samples within a litter (Fig. 3). Lines with the highest copy numbers had sample estimates that fell outside the linear range of our standard curve; not surprisingly, these showed a wider range of copy number estimates (Fig. 3b).

Conventional methods for estimating copy number in transgenic lines include quantitative dot-blot hybridization or Southern blot analysis. To further validate the copy number estimates from real-time PCR data, we performed quantitative dot-blot hybridization on genomic DNA samples purified from livers from a limited number of mice

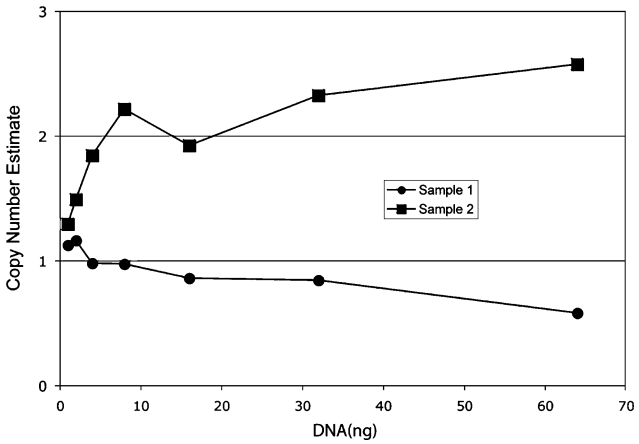

Fig. 2 DNA concentration has little impact on copy number estimates over a wide range of input DNA. DNA samples from two transgenic mice with copy number estimates of 2 (filled box) and 1 (filled circle) were each used to create a twofold dilution series of DNA templates, such that 1–64 ng DNA (total input) from each mouse was subjected to real-time PCR. The amount of template DNA versus copy number estimations indicate that copy number estimations vary little as input DNA ranged from 4 to 32 ng

from six *Bmp4* BAC lines (Fig. 4) and compared estimates based on dot blots to those generated by real-time PCR using the same individual liver DNA samples (Table 2). To control for differences in amount of input DNA, a control genomic probe was utilized (Fig. 4b). Both the dot-blot and the real-time PCR analyses of liver DNA samples were performed in triplicate. Statistical analysis revealed no significant difference between copy number values estimated by traditional dot-blot analysis and values estimated by our real-time PCR methods (paired *t* test, $p = 0.74$). The average real-time PCR copy number estimates generated from tail DNA samples of multiple transgenic mice within a line (Table 2, far right column) were close to the copy number averages estimated by dot blot from two mice. Therefore, the real-time PCR estimation method seemed suitable for application to many DNA samples from tail snips or embryonic yolk sacs.

Because real-time PCR results can be skewed by contaminating materials that adversely affect amplification, we examined our data set for evidence of consistent DNA quality. For 26 transgenic lines, we specifically computed the average copy number estimate for all transgenic animals (317 mice from 26 breeding lines), and we counted all animals having individual estimates that were within twofold of the initial average for the line. (For this, we considered only lines that were clearly segregating only single sites of transgene integration.) All estimates were based on duplicate Neo and *Jun* PCR reactions. We observed that 291 samples (47 yolk sac and 244 tail DNAs) of 317 DNA samples tested gave copy number estimates that were within this range. A limited number of samples (7 yolk sac and 19 tail DNA samples) gave estimates that were either twofold greater or less than the initial average and were considered poor-quality DNA isolations. These estimates may have been skewed by issues relating to impure DNA template; previous reports state that materials in mouse tail tissue or traces of phenol can inhibit PCR (Burkhart et al. 2002). The majority of both yolk sac and tail DNA samples were within twofold of their line average (291/317). Nevertheless, this was a limited problem that was easily overcome by examining multiple animals for a given transgenic line (Fig. 3). While we also reasoned that our copy number estimates based on single samples should be interpreted with caution, we concluded that this method would still be useful for analyzing copy number trends across many founder animals.

Distribution of copy number across breeding lines and founders

In our laboratory we have generated a number of BAC transgenic embryos and breeding lines as part of our efforts

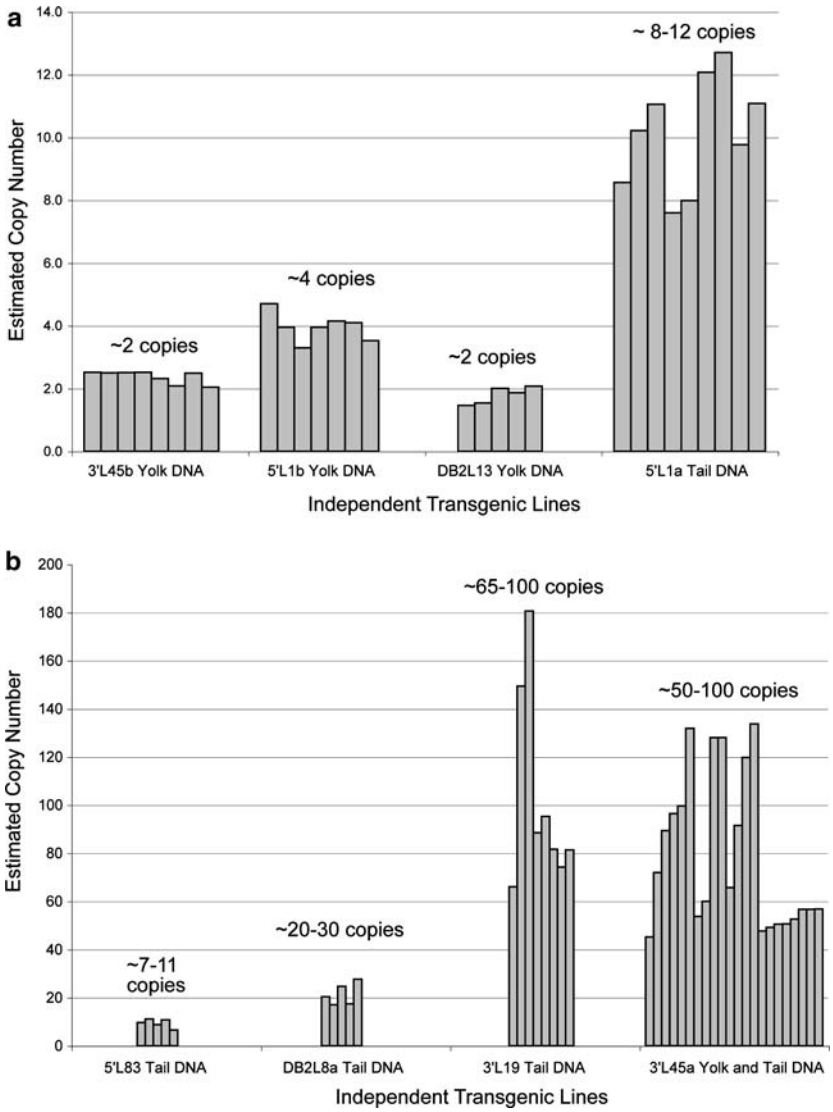

Fig. 3 Copy number estimates are consistent within independent transgenic lines. Shown are copy number estimates from individual mice as determined from yolk or tail DNA samples of multiple progeny from eight independent *Bmp4* BAC transgenic lines

to study regulation of the *Bmp2* and *Bmp4* genes. This involved 11 unique BAC constructs (6 *Bmp2* and 5 *Bmp4* BACs). While much of our data was based on progeny resulting from germline transmission of transgenes, some of our copy number data were based on DNA from midgestation transient transgenic embryos or liveborn mice generated directly from pronuclear injection. We refer to these as "transient transgenic" mice. Upon analyzing copy number estimates for 78 transient transgenic embryos or founder mice and for 26 breeding transgenic lines established from some of the founders, we were able to compile these data and create a distribution of copy number values for each independent founder or average values for each breeding line (Fig. 5). For this analysis, we recalculated copy number estimates for all transgenic lines after excluding the poor-quality DNA samples as defined above. As described below, we found several cases where two independently segregating insertions were clearly derived from one founder animal. In these cases, the separate insertions were considered as separate lines. For each line, samples from at least three mice were used to generate the average copy number value (average number of mice used for each line was 13).

Investigation of copy number in both transient transgenic embryos or mice and breeding lines allowed us to compare all BAC transgene integration events (Fig. 5). This distribution clearly shows that the majority of transient transgenic embryos, liveborn founder mice, and breeding transgenic lines contained one or more transgene copies per genome (Fig. 5). Not unexpectedly, real-time PCR analysis suggested that every breeding line contained one or more transgene copies. Approximately 22% of transient transgenic or founder animals (17/78) were estimated to carry less than one transgene copy per genome, suggesting that many of these animals were probably genetically mosaic for the transgene. However, such founders could often produce

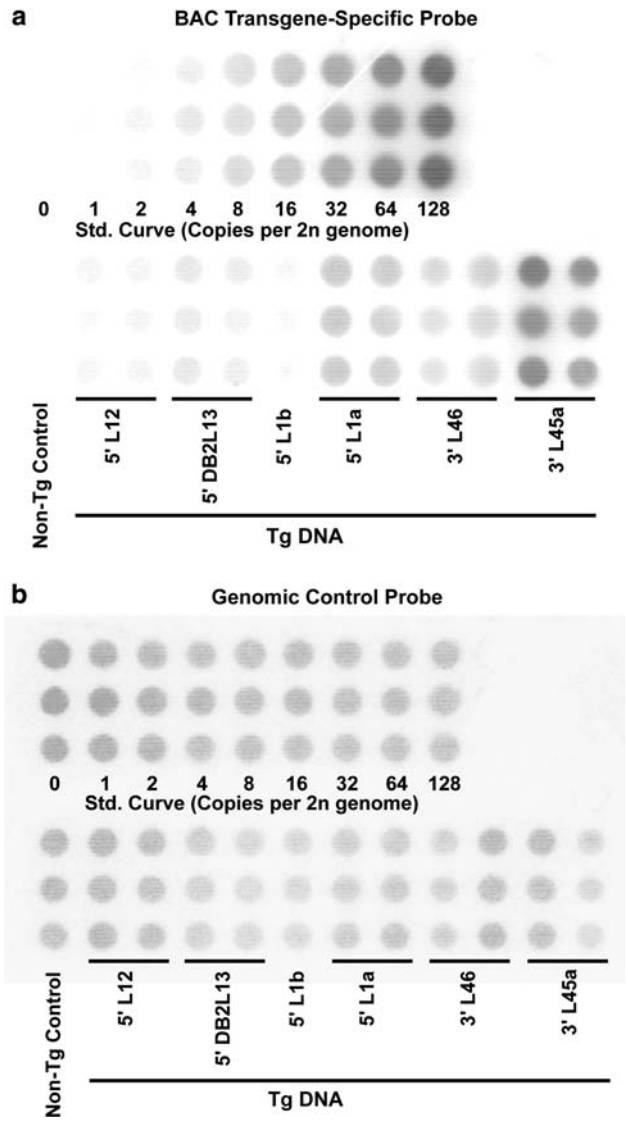

Fig. 4 Copy number estimation by dot-blot hybridization. **a** Dot blot hybridized with a BAC transgene-specific probe (IRES-βgeo). **b** The same dot blot stripped and reprobed with a genomic control probe (mouse *Gdf6* 3′ UTR fragment) to account for differences in input DNA. Standard curve (0–128 copies per diploid genome) and genomic DNA samples were assayed in triplicate. Copy number estimates for genomic DNA samples were derived by comparing the ratio of dot intensities for transgene-specific and control probe hybridizations in standard curve samples spiked with known quantities of pIBGFTet plasmid (see Materials and methods)

transgenic offspring, in some cases with rather high transgene copy numbers (not shown), suggesting that somatic mosaicism in the founder can preclude the ability to predict copy number estimates in their offspring before actually breeding the founder. Although most published reports suggest that BAC transgenes integrate as low copy concatamers (Giraldo and Montoliu 2001; Heaney and Bronson 2006; Jaenisch 1988), 18% of transient transgenic embryos or founders (14/78) had copy number estimates greater than approximately 25 BAC copies per genome. After breeding a subset of founders, 12% (3/26) of established lines also had more than approximately 25 copies, although founder estimates did not always predict the high copy numbers in offspring (see below).

Analysis of copy number in successive generations

Although most transgenic mice made via pronuclear injection have transgene DNA inserted at a single genomic location, integration into two separate, unlinked locations can occur (Fig. 6). As expected, most of our BAC founder animals ($N = 20$ of 23 bred founders) generated close to 50% transgenic and 50% nontransgenic progeny and copy estimates were similar among transgenic littermates, consistent with a single, stable site of BAC transgene integration in the founder. Our BAC lines are designed to drive *lacZ* expression during mouse development as a convenient reporter for *Bmp2* or *Bmp4* expression. Occasionally, when transgenic founders were bred to generate liveborn progeny and/or embryos for XGal staining, we noticed that roughly 75% of progeny were transgene-positive and that there were obviously two different levels of XGal staining intensity (i.e., "strong" vs. "weak"). This was observed for 3 of the 23 founder animals, suggesting that each founder could transmit at least two distinct, unlinked transgene insertions. Pedigree analysis of *Bmp4* 5′ BAC line L1 across two generations confirmed evidence for two independent integration events that segregated independently (Fig. 6; for clarity, the nontransgenic littermates are not shown; however, 19 of 25 weanlings from two litters were transgene-positive). Interestingly, one integration in this line had approximately ten BAC copies whereas the other integration had four copies. *lacZ* expression analysis confirmed that embryos generated from stud males containing the "high-copy" integration have more robust expression compared with embryos carrying the "low-copy" integration (Fig. 6a). In addition, the copy number estimate obtained from tail DNA of the pedigree founder was close to two, whereas copy number estimates of multiple progeny strongly suggest that the founder actually carried two integrations that each had more than two copies (copy number estimates from the founder female were determined from two independent tail snips to confirm these results; estimates were 1.8 and 2.6 copies, respectively). Although copy number estimates for the founder and F_1 progeny were different, estimates for successive generations were stable (Fig. 6a and data not shown).

Similar to *Bmp4* 5′ BAC L1, pedigree analysis of *Bmp4* 3′ BAC line L45 revealed two independent integration events (Fig. 6b). In this case, copy number estimates for the founder female were similar to copy number estimates of the "high-copy" integration event. A third founder also

Table 2 Comparison of copy number estimates generated by dot-blot analysis versus real-time PCR (RT-PCR) on *Bmp4* BAC transgenic mouse liver DNA samples from individual mice

Line	Mouse	Dot blot (Liver)	Avg.	RT-PCR (Liver)	Avg.	RT-PCR avg. all samples[a]
3′ L46	1	14	14	14	12	10
	2	13		10		
3′ L45a	1	76	77	75	83	76
	2	78		90		
3′ DB2 L13	1	7	8	4	5	4
	2	8		5		
5′ L12	1	2	3	2	3	2
	2	3		3		
5′ L1a	1	22	20	15	16	11
	2	17		16		
5′ L1b	1	5	N/A	4	N/A	4

[a] The average copy number computed from all samples using RT-PCR

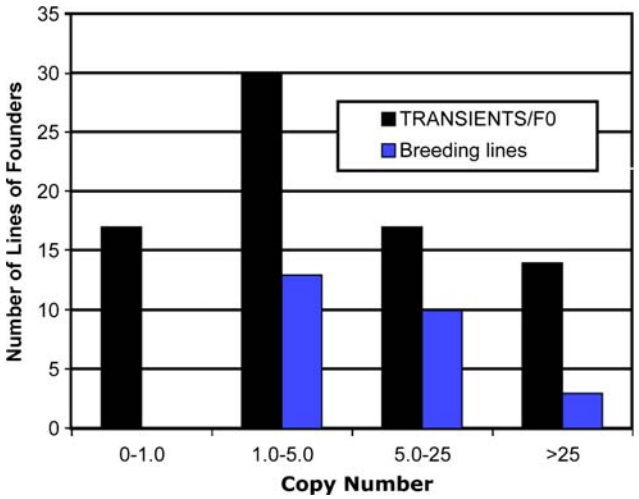

Fig. 5 The distribution of variation in copy number across stably breeding lines and transiently generated founder embryos or liveborn founder mice. While the majority of integration events contain 1–25 copies, all animals with estimates of fewer than one copy per genome were founder animals, suggesting somatic mosaicism

transmitted two segregating transgene insertions (not shown). Although these three founders each demonstrated multiple integration events, the majority of breeding founders analyzed (20/23 founders) had no evidence for multiple integrations.

Correlation between increased copy number and increased expression

It has been generally observed that for large transgenes, the correlation between strength of expression and copy number is more consistent than for small constructs. However, silencing of gene expression has been reported for large transgenes when present in "high" copy numbers (8–14 copies) (Li et al. 2000). To ensure staining differences were not due to varying protocols, embryos from different lines were stained for the same amount of time at the same temperature. We found that strength of XGal staining in *Bmp2* or *Bmp4* BAC transgenics correlated qualitatively with increased transgene copy number, as shown in Fig. 7, but because we did not measure expression quantitatively, we cannot determine if expression rigidly correlates to copy number in high-copy lines. As previously published, 3′ *Bmp2-lacZ* BAC transgenic embryos display a subset of endogenous *Bmp2* expression patterns such as whisker hair shaft, ventral footpads, osteoblast progenitors (bone), intervertebral discs, kidney, pelage hair follicle placodes, midbrain, and interdigital mesenchyme (Chandler et al. 2007). Deletion of a 40-kb segment of 3′ *Bmp2-lacZ* BAC (Del4) results in the tissue-specific loss of BAC-directed *lacZ* expression in the intervertebral discs and the midbrain (Chandler et al. 2007). *Bmp2* Del4 transient transgenic embryos show little to no *lacZ* expression when copy number is low (Fig. 7, bottom panel). As copy number increases, so does the apparent strength of *lacZ* expression. Of note, close inspection of transient transgenic embryo Del4-L41 showed a mosaic pattern of staining (not shown), thus, its overall staining strength appears to be relatively weaker than other embryos of similar copy number (Fig. 7, bottom panel). Although expression is decreased in embryos with modest copy number estimates, staining patterns were very similar to those in embryos with very high copy number. For example, *Bmp2-lacZ* BAC embryo Del4-L25 has expression in limb bones and digits as seen in *Bmp2-lacZ* BAC embryo Del4-L26, albeit at a significantly reduced level (Fig. 7, bottom). This transgene also drives expression in hair follicles such that the exterior of the embryo appears strongly stained at higher copy numbers (e.g., L48 and L26 embryos) (Chandler et al. 2007); lower-copy-number embryos had fainter, distinct expression in hair follicles (not shown). *Bmp4-lacZ* BAC transgenes direct expression faithfully in several tissues where *Bmp4* is endogenously expressed (K. Chandler et al, unpublished). For example, 3′ *Bmp4-lacZ* BAC transgenes direct expression in the craniofacial mesenchyme and whisker hair shafts (Fig. 7, top panel). Both structures are documented sites of *Bmp4* expression (Bitgood and McMahon 1995; Jones et al. 1991; Liu et al. 2005). Likewise, 5′ *Bmp4-lacZ* BAC transgenes direct expression in the developing forebrain, choroid plexus, and whisker primordia where *Bmp4* is known to be expressed (Fig.7, middle panel) (Bitgood and McMahon 1995; Furuta et al. 1997). Thus, analysis of *lacZ* expression in embryos

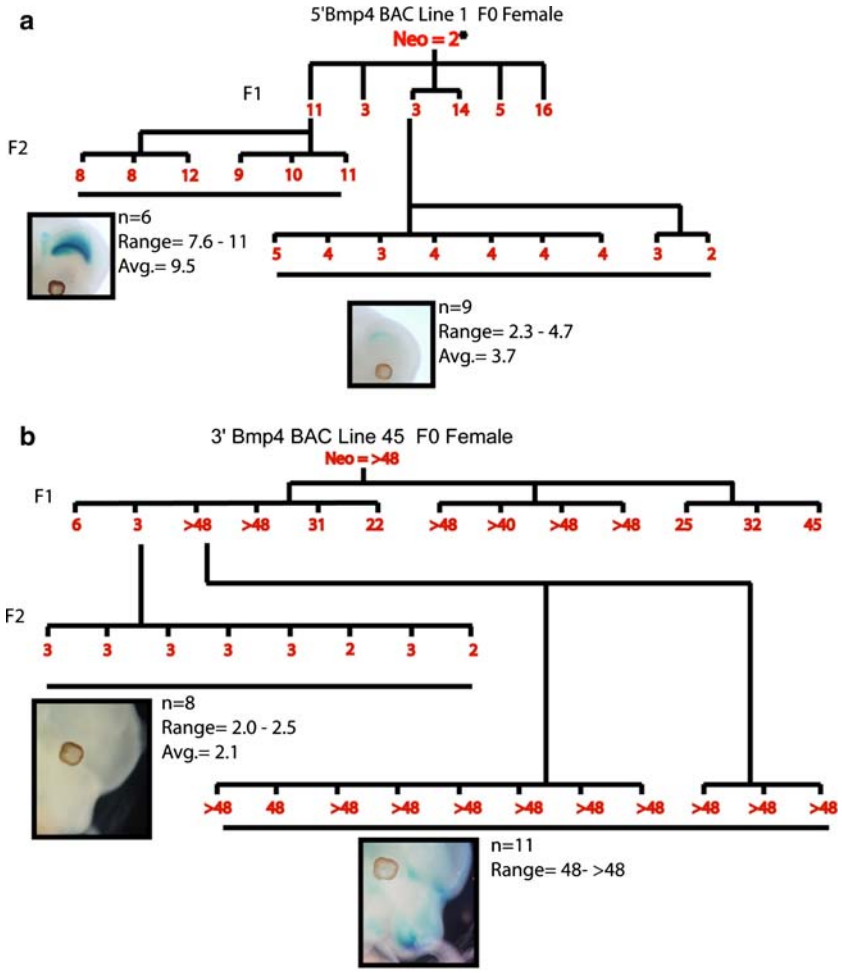

Fig. 6 Pedigree analysis of mice generated from two independent founder mice reveals that in both cases BAC transgenes have inserted in two separate, segregating locations in the genome, as demonstrated by number estimates. In both cases this was supported by an approximately 75% rate of transgenesis in F_1 progeny (nontransgenic littermates not shown). Copy number estimates for individuals are shown in red. Inset images show representative XGal-stained e12.5 embryos characteristic of each independent integration event. **a** 5′ *Bmp4* BAC line 1 founder female generated mice with "high" (average = 9.5) and "low" (average = 3.7) copy number estimates that segregate independently. For the founder, copy number estimates were based on the average of two independent tail biopsies/DNA preps (*). **b** 3′ *Bmp4* BAC line 45 founder female generated F_1 progeny with "high" (average ≥ 48) and "low" (average = 2.1) copy number estimates that segregate independently

generated from these three transgene constructs clearly showed increasingly robust *lacZ* expression as copy number increased, with no evidence of strong silencing effects with higher copy numbers. In contrast, lines with few transgene copies exhibited markedly reduced or completely absent *lacZ* expression. Similar trends were observed with no exceptions in a total of 26 breeding lines involving the constructs in Fig. 7 and in eight additional BAC constructs (data not shown).

Analysis of BAC transgene integrity

We investigated the possibility that internal deletions within the transgene might have occurred in lines that had minimal or absent expression. Transgenes that are introduced by pronuclear injection typically integrate into the genome as tandem concatamers by a mechanism involving homologous recombination between circularly permuted molecules (Bishop 1996; Bishop and Smith 1989; Hamada et al. 1993). Although large molecules can be prone to breakage before integration, and DNA molecules can be randomly broken after entering the cell (Bishop and Smith 1989), leading to insertion of fragmented transgenes, it has been reported that multiple-copy BAC insertions usually have at least one full-length monomer (Chandler et al. 2007; Gong et al. 2003); however, BAC transgene integrity and copy number are not always compared in published studies. One approach to monitor integrity of large transgenes is polymorphic marker analysis (Deal et al. 2006). We used this approach to analyze the presence of polymorphic markers across the length of transgenes in 18 *Bmp4* BAC lines (Fig. 8). This confirmed that multiple-copy BAC transgene integrations most often contain all segments of the transgene that were assayed, suggesting that these lines may have at least one intact copy of the BAC transgene. Twenty simple tandem repeat (STR) polymorphisms, with an average distance of 19 kb between each polymorphism, were assessed within the *Bmp4* BAC transgenes. The majority of lines (15/18) were shown to harbor all transgene-specific polymorphisms, suggesting integration of complete BAC molecules (Fig. 8). However, three lines lacked transgene-specific polymorphisms across one portion of the BAC, suggesting that these transgenes integrated into the mouse genome as partial fragments. Therefore, the frequency of lines in which part of the BAC

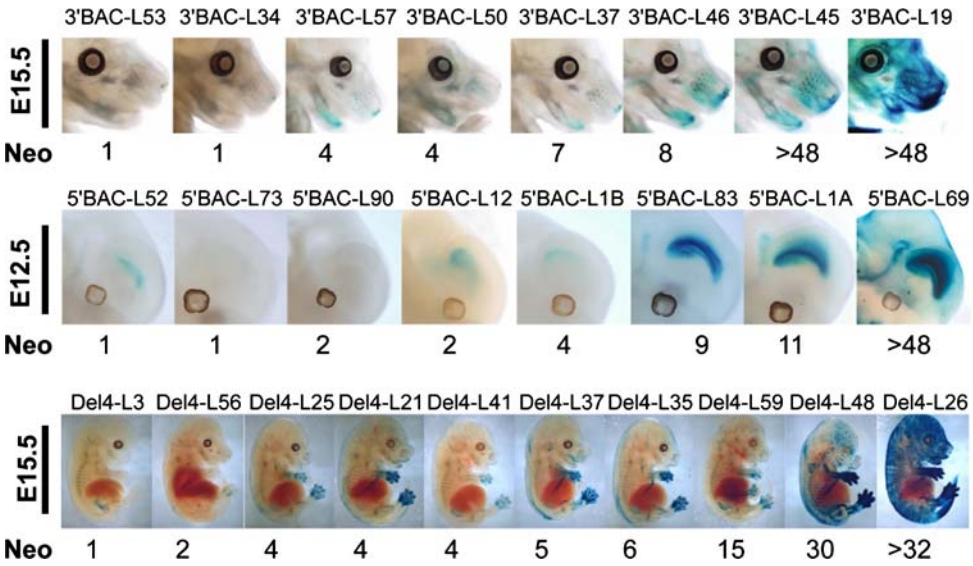

Fig. 7 Xgal-stained embryos generated from three distinct BAC transgene constructs suggest that increasing BAC transgene copy numbers correlates with increased transgene expression. Each row of images represents embryos from a separate BAC construct, arranged by increasing BAC copy number estimates. Top row: e15.5 embryos from independent *Bmp4* 3′ BAC lines. Middle row: e12.5 embryos from independent *Bmp4* 5′BAC lines. Bottom row: e15.5 *Bmp2* deletion-BAC transgenic founder embryos

was inadvertently deleted was 17% (3/18). Expression in two lines with internal deletions, *Bmp4* 5′ BAC L8b and *Bmp4* 3′ BAC L44, was completely undetectable by XGal stain (data not shown). In addition, pedigree analysis of founder "L8" revealed two independently segregating integrations, with one integration containing all BAC markers (line L8a) and the other integration being fragmented (line L8b). Copy number estimates for the "fragmented" lines indicated that each of these lines has an estimated one *lacZ* copy (Fig. 8). In contrast, copy number estimates for breeding lines with "intact" BAC transgenes ranged from more than 3 to more than 48 (Fig. 8). In addition, one founder carrying all BAC markers had a copy number estimate of 2.

In summary, for the 18 BAC lines or founders that we have analyzed for transgene structure, all those with two or more copies ($N = 15$) appeared to contain intact transgenes based on sampling for STR polymorphisms occurring at an average of every 19 kb, while those estimated to be single-copy integrants were each missing part of the transgene ($N = 3$). However, PCR analysis does not definitively address the linear structure of the BACs. To further investigate the integrity of both low-copy and high-copy BAC transgenic lines, we performed Southern blot analysis on high-molecular-weight DNA samples isolated from two individual mice each from a "low-copy" *Bmp4* BAC line (5′ BAC-L12) and from a "high-copy" *Bmp4* BAC line (5′ BAC-L1A). Each mouse DNA sample was digested with a rare cutting enzyme (*Mlu*I) and subjected to pulsed-field gel electorphoresis alongside a digest of purified BAC DNA (Fig. 9a). *Mlu*I cuts at two distinct locations in the 5′ BAC and digestion of purified 5′ BAC DNA yields two bands (Fig. 9a); however, one of the *Mlu*I sites is within a CpG island in the *Bmp4* promoter. Therefore, in the context of mouse genomic DNA, it is likely that only the promoter *Mlu*I site remains unmethylated and is sensitive to *Mlu*I digestion. *Mlu*I digestion of transgenic mouse DNA should then yield approximately 235-kb fragments from 5′ BACs integrated as tandem concatamers. After hybridization of the Southern blot with a transgene-specific probe, bands corresponding with the size of a full-length *Bmp4* BAC were visualized (Fig. 9b), suggesting that both the low- and high-copy BAC lines shown here most likely contain at least one intact molecule. Of note, two *Bmp2* BAC transgene lines previously analyzed by us each had copy numbers of 16 or more and were both shown to contain mostly concatamerized full-length BAC copies by Southern blot analysis (Chandler et al. 2007). Taken together, our data support the idea that multiple-copy BAC transgene insertions most likely contain one or more full-length BAC copies.

Discussion

BAC transgenic mice are increasingly used for biomedical research. However, few studies have addressed issues regarding BAC transgene copy number, integrity, or function across a large population of transgenic mice (Alexander et al. 2004; Gong et al. 2003). We have taken advantage of the BAC transgenic embryo founders and breeding lines produced in our lab to interrogate the copy number distribution among distinct and independent lines, the relationship between copy number and levels of transgene expression, the integrity of multiple BAC transgenes, and the fidelity of BAC transgenes across successive generations.

Several previous reports are also supportive of our observations. The potential for fragmentation of BAC transgenes prior to integration has been noted previously

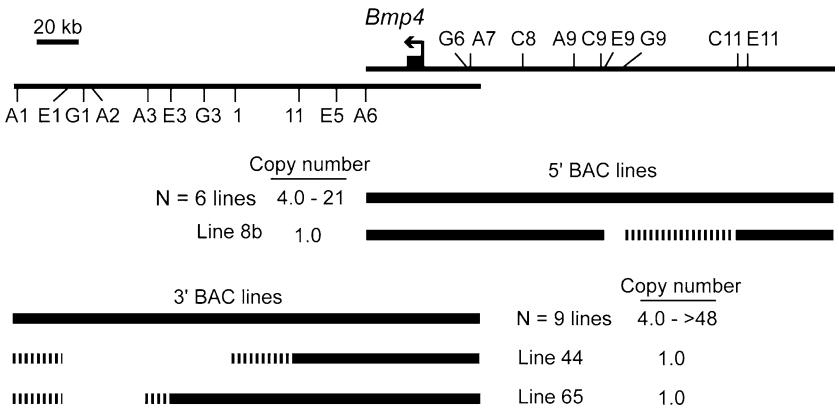

Fig. 8 Polymorphic marker analysis suggests that transgenic lines that have multiple BAC copies are likely to carry some intact BAC molecules. Polymorphic markers denoted along the length of *Bmp4* 5′ BAC (top right) and *Bmp4* 3′ BAC (top left; scale bar = 20 kb). Shown below each BAC are schematics representing lines for which all BAC markers are present, suggesting intact BAC transgenes (solid bars) and lines containing fragmented BAC transgenes (interrupted bars). Solid lines indicate the presence of contiguous transgene-specific markers. Open regions indicate loss of transgene-specific markers, and hatched regions indicate regions of potential breakpoints

and can even be used to refine potential *cis*-regulatory domains in some situations (Deal et al. 2006). In a very large set of independently generated BAC transgenic mice (the GENSAT project), Gong et al. (2003) reported that BAC transgene insertions having multiple copies invariably contained full-length copies as tandem arrays, although copy numbers were not reported for individual lines. In some cases, for conventional transgenes stability of transgene copy number has been followed over time in breeding colonies to document that sporadic loss of transgene copy number can occasionally occur, probably by internal recombinations (Alexander et al. 2004). We have no clear examples of loss of copies within our BAC transgenic lines, although we cannot rule out that it might occur sporadically. Monitoring BAC copy number in each breeding generation seems prudent.

The method we used to obtain copy number information on BAC transgenic lines used real-time PCR on phenol/chloroform-extracted yolk sac or tail biopsy DNA samples and a BAC copy number standard curve. Unlike the method described here, some previous studies describe real-time PCR methods to estimate copy number on liver biopsy DNA samples, which requires killing a transgenic mouse (Ballester et al. 2004). Therefore, our method is advantageous for use with valuable transgenic mice (such as founders) for which premature sacrificing is undesirable. In addition, our method estimates copy numbers based on a standard curve of known BAC copy number standards. Many other studies have used the $2^{-\Delta\Delta Ct}$ method to calculate copy number (Ballester et al. 2004; Tesson et al. 2002), which requires nearly equivalent PCR efficiencies between the unknown samples and the control. Others have reported that phenol/chloroform-extracted tail DNA samples contain PCR inhibitors (Burkhart et al. 2002) or cannot be quantified accurately by UV spectrophotometry due to phenol contamination (Alexander et al. 2004). However, our studies showed generally reproducible copy number estimates from littermates using tail biopsies prepared in this manner. In some instances, we found samples that gave estimates likely to be erroneous. In these cases, the estimates were more than two-fold greater or lower than the average for multiple transgenic littermates. This was observed for both tail and yolk sac DNA samples, and we suspect copy number estimates that fell outside of the twofold threshold were the result of phenolic contamination, PCR inhibitors not removed during extraction, or both. Alternatively, loss of transgene copies could result from recombination events in meiosis, as discussed above. This underscores the need to gather, when possible, multiple data points across transgenic littermates and generations for assessment of copy number in individual transgenic lines.

We showed a strong correlation between increased copy number and increased transgene expression in multiple independent transgenic mice derived from three distinct BAC transgene constructs. As copy number increased, so did the qualitative intensity of *lacZ* reporter expression, at least for *Bmp2* and *Bmp4* BACs. On the other hand, others have observed that even for BAC- or YAC-sized constructs, increased transgene copy number may not correlate with increased transgene expression and, in fact, may result in transgene silencing (Heaney and Bronson 2006). Distinct from epigenetic silencing, the sporadic deletion of integrated transgene copies following breeding can clearly reduce transgene expression compared with expression in preceding generations prior to deletion (Alexander et al. 2004). Although we did not measure transgenic mRNA or β-galactosidase activity quantitatively, we found no obvious evidence for the silencing of *Bmp2* or *Bmp4* BAC transgene expression when copy numbers were high

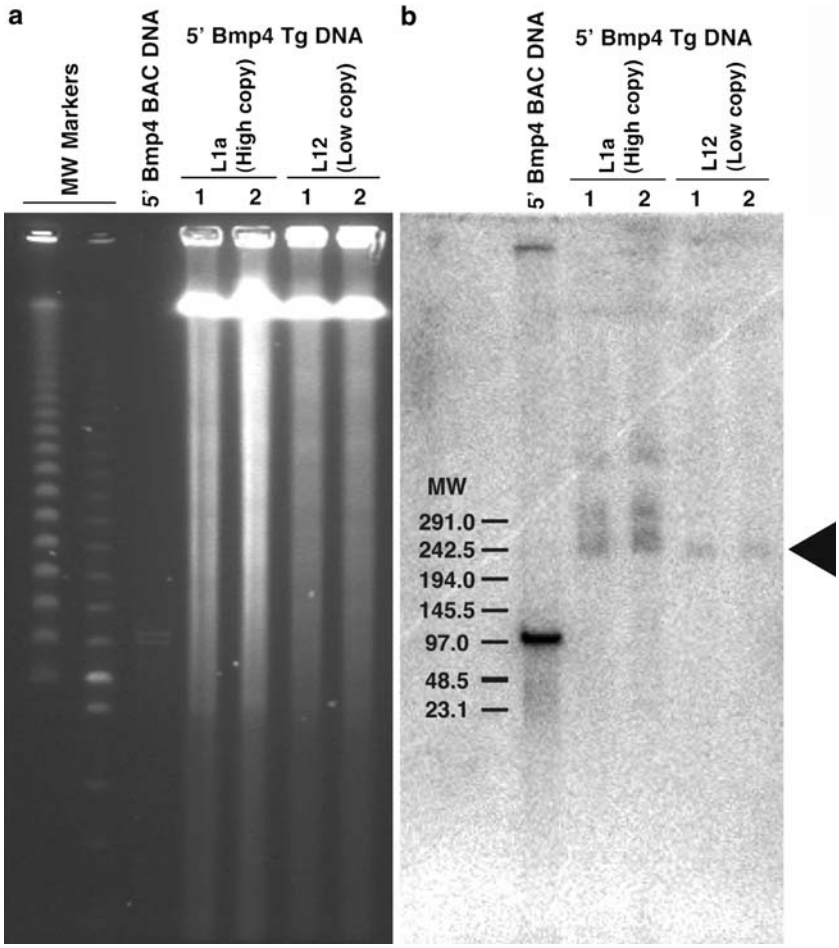

Fig. 9 Southern blot analysis on high-molecular-weight DNA samples from low-copy (5′ L12, average copy number = 2) and high-copy (5′ L1a, average copy number = 11) *Bmp4* BAC lines suggests intact transgene copies. **a** Image of ethidium bromide-stained pulsed-field gel electrophoresis of agarose-embedded and *Mlu*I-digested high-molecular-weight DNA isolated from embryos generated from 5′ BAC carrying stables lines (see Materials and methods). Also included are control digestions of purified 5′ BAC DNA (50 ng DNA per lane). **b** Phosphor image of gel shown in **a** following Southern transfer and hybridization with radiolabeled probe (IRES-*β*-geo cassette). The approximately 110-kb doublet in blot lane 1 represents the expected *Mlu*I fragments from purified (unmethylated) 5′ BAC DNA. In the lanes with transgenic mouse DNA digested with *Mlu*I, bands are evident (arrowhead) that are approximately the full-length size of the 5′ BAC transgenes (∼235 kb) (note that the high-copy line yields stronger bands than the low-copy line). This strongly suggests that one or more copies of transgenes are intact in both the high- and low-copy 5′ BAC lines

(Fig. 7). We found that BAC transgenic lines with at least approximately ten copies were ideal for analysis because transgenes were intact and expression of the reporter gene was robust.

Because BAC transgenes are significantly larger constructs compared with conventional transgenes and are more susceptible to fragmentation (Deal et al. 2006), it is imperative that the integrity of BAC transgenes is verified. In this article we showed that copy number estimates obtained from real-time PCR methods, coupled with marker genotyping and/or Southern blot analysis, revealed that transgenic lines having at least three copies might contain intact molecules as suggested by polymorphic marker analysis, whereas lines with fewer than two copies often contained only partial BAC fragments. We found that the majority of our BAC transgene lines contained all tested markers, and two lines were confirmed to contain at least one full-length molecule as demonstrated by Southern blot analysis. This is reassuring because of the general concern that BAC transgenes are easily fragmented. However, we still caution that careful preparation and handling of BAC DNA samples used for pronuclear injection is critical for efficient transgenesis.

Like smaller transgenes, it has been suggested that BACs typically incorporate in the genome as one to five copy concatamers within a single locus of the genome (Giraldo and Montoliu 2001; Heaney and Bronson 2006; Jaenisch 1988). Our data showed that 50% of transgenic lines had between one and five copies, and most lines had only single sites of transgene integration. However, we also

identified multiple lines that each had copy number estimates greater than approximately 48 or more. Although we are cautious in considering the raw values for these lines as definitive because they were outside the boundaries of our real-time PCR standard curve samples, one line repeatedly produced copy number estimates of nearly 100 (average copy number = 96; standard deviation [SD] = 28; $n = 38$) and an additional line had an average copy number of 76 (SD = 29; $n = 22$). Analysis of a limited number of DNA samples from the latter line (3′ BAC-L45A) by dot-blot hybridization corroborates this estimate (average copy number = 77; $n = 2$). To our knowledge these are the largest BAC copy number estimates ever reported in transgenic mice.

In addition, three transgenic founders transmitted two independent integration events, suggesting that BAC transgenes sometimes integrate in more than one loci of the genome at a reasonably high frequency. Analysis of lines having more than one integration event revealed that the independently segregating integrations often had distinct copy numbers in successive generations. Because we found that increased copy number strongly correlated with increased expression but also found that independent integrations often harbor distinct copy numbers, it is imperative that transgenic lines bred for successive generations are carefully characterized by analyzing copy number in multiple F_1 progeny to prevent the loss of valuable integration during subsequent breeding or expression data in F_2 progeny. In addition, the analysis of copy number in founders should be approached with caution without confirmation in F_1 progeny because of the possibility of mosaicism or multiple integration events that cannot be resolved unless the transgene is passed through the germline.

In closing, we presented a quick and reliable method for estimating copy number in BAC transgenic lines that has shown to be useful in characterizing multiple lines by analysis of limiting amounts of DNA. These methods are applicable to transient transgenic founder embryos and where limited tissue is available for DNA extraction (e.g., yolk sacs), albeit with the above caveats. To help reduce costs of reagents, we reduced TaqMan reaction volumes from 20 to 10 μl and used phenol/chloroform extractions to isolate DNA in place of kit-based methods. Finally, we have provided evidence for the importance of evaluating copy number across multiple progeny from BAC transgenic lines and have demonstrated the increased likelihood that multiple copy integrations typically contain one or more full-length BAC transgenes. We suspect these observations and techniques may be valuable to investigators as the demand for BAC transgenic mice increases.

Acknowledgments The authors thank the Vanderbilt CHGR DNA Resources Core for real-time PCR assistance and TaqMan probe/primer design; Jeff Smith and Kevin Bradley for assistance in identification of STR markers within the *Bmp4* BAC intervals; Jeff Innis and Maureen Gannon for helpful discussions on the manuscript; Nyk Reed for reading the manuscript; and Will Bush for statistical analysis and assistance. Kelly J. Chandler was supported by NIH Genetics Training Grant 1T32GM62758-03. Ronald L. Chandler was supported by NIH Developmental Biology Training Grant 5T32HD07502-08. Douglas P. Mortlock was supported by NIH Grants 1R01HD47880-01 and 5R01AR049529-04. Transgenic mice were generated by the Vanderbilt University VCSCV Transgenic Mouse and ES Cell Shared Resource, which is supported by the Vanderbilt Cancer, Diabetes, Kennedy, and Vision Centers. The authors acknowledge the use of the VUMC CHGR DNA Resources Core Facility.

References

Alexander GM, Erwin KL, Byers N, Deitch JS, Augelli BJ, et al. (2004) Effect of transgene copy number on survival in the G93A SOD1 transgenic mouse model of ALS. Brain Res Mol Brain Res 130:7–15

Ballester M, Castello A, Ibanez E, Sanchez A, Folch JM (2004) Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals. Biotechniques 37:610–613

Bishop J (1996) Chromosomal insertion of foreign DNA. Reprod Nutr Dev 36:607–618

Bishop JO, Smith P (1989) Mechanism of chromosomal integration of microinjected DNA. Mol Biol Med 6:283–298

Bitgood MJ, McMahon AP (1995) Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. Dev Biol 172:126–138

Burkhart CA, Norris MD, Haber M (2002) A simple method for the isolation of genomic DNA from mouse tail free of real-time PCR inhibitors. J Biochem Biophys Methods 52:145–149

Chandler RL, Chandler KJ, McFarland KA, Mortlock DP (2007) Bmp2 transcription in osteoblast progenitors is regulated by a distant 3′ enhancer located 156.3 kilobases from the promoter. Mol Cell Biol 27:2934-2951

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

Deal KK, Cantrell VA, Chandler RL, Saunders TL, Mortlock DP, et al. (2006) Distant regulatory elements in a Sox10-beta GEO BAC transgene are required for expression of Sox10 in the enteric nervous system and other neural crest-derived tissues. Dev Dyn 235:1413–1432

DiLeone RJ, Russell LB, Kingsley DM (1998) An extensive 3′ regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo. Genetics 148:401–408

Furuta Y, Piston DW, Hogan BL (1997) Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. Development 124:2203–2212

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

Giraldo P, Rival-Gervier S, Houdebine LM, Montoliu L (2003) The potential benefits of insulators on heterologous constructs in transgenic animals. Transgenic Res 12:751–755

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, et al. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425:917–925

Hamada T, Sasaki H, Seki R, Sakaki Y (1993) Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: sequence analysis of genome-transgene and transgene-transgene junctions at two loci. Gene 128:197–202

Heaney JD, Bronson SK (2006) Artificial chromosome-based transgenes in the study of genome function. Mamm Genome 17:791–807

Heintz N (2001) BAC to the future: the use of bac transgenic mice for neuroscience research. Nat Rev Neurosci 2:861–870

Jaenisch R (1988) Transgenic animals. Science 240:1468–1474

Jones CM, Lyons KM, Hogan BL (1991) Involvement of Bone Morphogenetic Protein-4 (BMP-4) and Vgr-1 in morphogenesis and neurogenesis in the mouse. Development 111:531–542

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

Li S, Hammer RE, George-Raizen JB, Meyers KC, Garrard WT (2000) High-level rearrangement and transcription of yeast artificial chromosome-based mouse Ig kappa transgenes containing distal regions of the contig. J Immunol 164:812–824

Liu W, Sun X, Braut A, Mishina Y, Behringer RR, et al. (2005) Distinct functions for Bmp signaling in lip and palate fusion in mice. Development 132:1453–1461

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

Tesson L, Heslan JM, Menoret S, Anegon I (2002) Rapid and accurate determination of zygosity in transgenic animals by real-time quantitative PCR. Transgenic Res 11:43–48