Published in final edited form as:

Chromosoma. 2013 October ; 122(5): 351–361. doi:10.1007/s00412-013-0418-9.

Characterization of a large human transgene following invasin-mediated delivery in a bacterial artificial chromosome

Austin E. Gillen^{1,2}, Catherine A. Lucas³, Pei Ling Haussecker⁺, Steven T. Kosak³, and Ann Harris^{1,2,*}

¹Human Molecular Genetics Program, Lurie Children's Research Center, Northwestern University Feinberg School of Medicine Chicago, IL. USA

²Department of Pediatrics, Northwestern University Feinberg School of Medicine Chicago, IL. USA

³Department of Cell and Molecular Biology, Northwestern University Feinberg School of Medicine Chicago, IL. USA

Abstract

Bacterial artificial chromosomes (BACs) are widely used in transgenesis, particularly for the humanization of animal models. Moreover, due to their extensive capacity, BACs provide attractive tools to study distal regulatory elements associated with large gene loci. However, despite their widespread use, little is known about the integration dynamics of these large transgenes in mammalian cells. Here, we investigate the post-integration structure of a ~260 kb BAC carrying the cystic fibrosis transmembrane conductance regulator (*CFTR*) locus following delivery by bacterial invasion and compare this to the outcome of a more routine lipid-based delivery method. We find substantial variability in integrated copy number and expression levels of the BAC *CFTR* transgene after bacterial invasion-mediated delivery. Furthermore, we frequently observed variation in the representation of different regions of the *CFTR* transgene within individual cell clones, indicative of BAC fragmentation. Finally, using fluorescence in situ hybridization (FISH), we observed that the integrated BAC forms extended, megabase-scale structures in some clones that are apparently stably maintained at cell division. These data demonstrate that the utility of large BACs to investigate *cis*-regulatory elements in the genomic context may be limited by recombination events that complicate their use.

Keywords

Invasin-mediated gene transfer; *CFTR* gene expression; BAC recombineering BAC integration

Introduction

Genetic methods to determine the function and regulatory mechanisms of human genes often necessitate delivery of the complete locus into heterologous cellular expression systems. For example, this may be required to generate transgenic animal model systems or to study regulatory elements in appropriate cellular expressions systems. Bacterial artificial chromosomes (BACs) are widely used as genomic context vectors to propagate large (up to 300 kb) DNA fragments and introduce them as transgenes into mammalian cells (Shizuya and Kouros-Mehr 2001). These vectors, which are based on the *E. coli* F-plasmid, are easily

*To whom correspondence should be addressed: ann-harris@northwestern.edu, Tel: 001-773-755-6525; Fax 001-773-755-6593.

⁺Former member of Harris laboratory in Weatherall Institute of Molecular Medicine, University of Oxford, UK.

propagated (O'Connor et al. 1989), and their extensive cloning capacity is particularly useful for the analysis of large genes or gene clusters, which encompass critical intronic or intergenic *cis*-regulatory elements. BACs are a preferred tool for the generation of transgenic lines of insects, fish and mammals (Giraldo and Montoliu 2001), and may have potential as gene therapy vectors. They are also useful for the 'humanization' of animal models whereby a human gene replaces a mutant/silent endogenous gene, and the model can then be used to test pharmacological therapeutics (reviewed in (Devoy et al. 2012)).

The successful delivery of intact BACs by microinjection into mouse embryonic stem cells has resulted in the generation of many animal models carrying functional BAC transgenes (Giraldo and Montoliu 2001). However, given the apparent efficacy of these protocols, few data have been generated on the integrity and stability of the BAC transgene (Chandler et al. 2007). Moreover, the molecular basis of the lower success rates in delivering BACs into somatic cells in culture has not been well studied. Detailed genomic mapping and thorough cytogenetic analysis of integrated BACs are rarely performed. Standard transgene delivery protocols for cultured cells, including lipofection, nucleofection, and electroporation, are known to be associated with rearrangement of the BAC, albeit at low frequency (Magin-Lachmann et al. 2004). In contrast, bacterial invasion, which uses modified *E. coli* to deliver BACs directly to mammalian cells without the need for purification of BAC DNA, was reported to consistently deliver intact, unrearranged BACs in excess of 250 kb to cultured mammalian cells (Perez-Luz et al. 2007; Kotzamanis et al. 2009). This observation was supported by the recovery from mammalian cells of intact BACs carrying the OriP/EBNA-1 cassette, which allows the vector to be maintained as an episome (Cheung et al. 2012). However, despite its potential, the BAC integration dynamics and copy number in stable cell clones generated by bacterial invasion have not been rigorously investigated.

Here, we use lipid-mediated transfection or invasin-mediated delivery to introduce a large BAC carrying the cystic fibrosis transmembrane conductance regulator (*CFTR*) locus into intestinal epithelial cells (Kotzamanis et al. 2009; Laner et al. 2005) to investigate the integrity of the transgene and thus its potential as a therapeutic tool. The *CFTR*_{1,2,3} BAC (BAC123s) was generated by homologous recombination of three overlapping BACs spanning the *CFTR* locus (Kotzamanis et al. 2009) (coverage shown in Fig. 1). In addition to encompassing the whole *CFTR* gene, the 260 kb BAC123s includes 40.1 kb 5' to the *CFTR* transcriptional start site and 25 kb 3' to the end of the locus and is thus comparable to a *CFTR* YAC (yCFTR), which we used previously. yCFTR, when introduced as a transgene into *cfr*^{-/-} mice, restored a normal phenotype (Manson et al. 1997). This suggests that the most important conserved regulatory elements in the human gene are found within yCFTR, and by extension, in BAC123s. BAC123s was further modified to include a *Cla*I restriction enzyme site in place of an endogenous *Bcl*I site in the 3' UTR (BAC123s *Cla*I), to enable discrimination of the associated transcript from endogenous human *CFTR* mRNA. In addition to wild-type BAC123s, two derivative constructs were evaluated in which critical regulatory elements were deleted by recombineering. These elements include a strong intestinal selective enhancer within intron 11 (BAC123s *DHS*11 *Cla*I) and an enhancer-blocking insulator element at +6.8 kb from the last coding base (BAC123s *DHS*+6.8kb *Cla*I). The wild type BAC alone was evaluated by lipid-mediated transfection. All three BACs were introduced into CMT-93 mouse intestinal epithelial cells, and BAC-derived *CFTR* expression was measured by qRT-PCR from multiple selected clonal cell lines and correlated to integrated BAC copy number in each clone. We demonstrate, by detailed genomic and fluorescence in situ hybridization (FISH) analysis, that variable copy number between different regions of this large 260 kb BAC is a common event upon integration. Moreover, invasin-mediated gene transfer often generated multiple BAC copies at single sites within the genome, rather than achieving the low copy number integration that was expected. These multiple integration events generally correlated with suppression of normal

transgene expression. Our data are particularly important in the context of using large BACs to generate transgenic animal models and in evaluating their potential as therapeutics.

Results

Generation of BAC123s *Cla*I stable clones in CMT-93 using a lipid reagent

CMT-93 murine rectal carcinoma cells were transfected with high quality BAC123s *Cla*I DNA, confirmed by pulsed field gel electrophoresis (PFGE), using Lipofectamine 2000 and subsequently placed under zeocin selection. The CMT-93 cell line was chosen for these experiments since it expresses murine *Cftr* and was used previously to introduce unmodified BAC123s (Kotzamanis et al. 2009). Moreover, the human transgene can be readily detected using species-specific copy-number and expression primers. Expression of human *CFTR* in the resulting stable clones was assayed using RT-PCR with human-specific primers located in exons 11–13 (C1R/C2L) and the 3' UTR (GV4579/5508) (Chalkley and Harris 1991; Vassaux et al. 1997). Forty-one zeocin resistant clones were screened (Table 1), but none expressed *CFTR* mRNA containing the middle and 3' UTR of the locus. One clone (100A13; Fig. 2a) produced a partial human *CFTR* mRNA (detected by the GV4579/5508 amplicon but not the C1R/C2L amplicon). RT-PCR with primers for the murine survival of motor neuron (*Smn*) gene, which is ubiquitously expressed, was used to verify the quality of the RNA and cDNA (Gogliotti et al. 2010). RT-PCR with the ZeoF/ZeoR primer set established that a subset of clones expressed the *Streptoalloteichus hindustanus* bleomycin gene (*Sh ble*), which confers zeocin resistance (Fig. 2b–c) (Calmels et al. 1991; Drocourt et al. 1990). PCR on genomic DNA isolated from five clones showed that they all lacked human *CFTR* exon 4, and only clone 100A13 contained exon 24 and the 3' UTR, consistent with the partial *CFTR* RNA detected in this line.

Generation of BAC123s *Cla*I stable clones in CMT-93 using bacterial invasion

Next, BAC delivery by bacterial invasion was used to try and circumvent the BAC fragmentation observed with Lipofectamine 2000-mediated transfection. BAC123s *Cla*I, BAC123s DHS11 *Cla*I and BAC123s DHS+6.8kb *Cla*I (recombineering to generate these BACs is shown in Online Resource 1, Fig. S1), were introduced into the bacterial invasion strain *E. coli* BM4573. Integrity of the BACs in BM4573 was evaluated by PFGE (Fig. 3, inset). For each bacterial invasion experiment 1.2×10^6 CMT93 cells (6 wells at 2×10^5 cells/well) were used, and a total of 41 zeocin-resistant clones were generated in four separate experiments. These included 12 BAC123s *Cla*I (WT), 12 BAC123s DHS11 *Cla*I (DHS11) and 17 BAC123s DHS+6.8kb *Cla*I (DHS+6.8 kb) clones which were given the following identifiers: BAC construct (Plate, letter)(clone number) (e.g. WTA3, 11B1, 6.8F2, etc.).

Characterization of CMT-93-BAC123s *Cla*I stable clones generated by bacterial invasion

Initial screening of the zeocin-resistant CMT-93 clones to identify those expressing a BAC123s-derivative transgene was performed by qRT-PCR using a human-specific TaqMan primer/probe set in exons 5–6a. Positive clones (summarized in Table 2) were then assayed for expression of the 3' end of the human *CFTR* transcript using SYBR green qRT-PCR with a *Cla*I-specific primer set in the 3' UTR. As shown in Fig. 3, 3' UTR expression was evident in all clones with detectable exon 5/6a expression. A total of 15 clones had no detectable expression including four WT, three DHS11 and seven DHS+6.8 kb. BAC-derived *CFTR* expression levels were extremely variable, spanning a >10,000-fold range with both assays (Fig. 3). The ratio of exon 5/6a expression to 3' UTR expression was also highly variable, spanning an equivalent range from the clone with the lowest ratio (WTG1; 1:60.9) to that with the highest (6.8G1; 159.9:1).

To determine the mechanism underlying this wide range of expression levels, we first determined BAC copy number by SYBR green qPCR with primers in exon 1, intron 10 and intron 20, normalized to 18S genomic DNA (Fig. 4; primer locations shown in Fig. 1). The BAC copy number showed wide variation not only between clones but also between primer sets within a single clone, suggesting that not all integrated BAC copies were intact. In contrast, normal human genomic DNA from 2 individuals (G1, G2) had a copy number of 2 across the locus and the Caco-2 cell line that is known to have an average chromosome number of 89, including frequent duplications of chromosome 7 (Melcher et al. 2002), had a copy number of 4 across the locus, as expected. Three zeocin-resistant clones with no detectable *CFTR* expression were evaluated, and all three were lacking at least a portion of the *CFTR* locus. In clone WTA1, a qPCR product was evident from exon 1 but not introns 10 or 20, while clones 11D2 and 6.8B2 had no detectable amplification with any of the three primer sets suggesting integration of the BAC vector was accompanied by loss of most of the *CFTR* locus. A further three clones in which *CFTR* expression was measured by both the exon 5/6a and the 3' UTR assay were also found to be missing a portion of the *CFTR* locus. There was no detectable product with one of the three PCR primer sets (intron 20 and intron 10, respectively) in clones WTF1 and 11A2, while clone 6.8G2 lacked 9 exon 1 and intron 10. Since other primer sets generated robust products in the same DNA samples these results are unlikely to be caused by the failure of the PCR reactions, rather they probably reflect relative depletion of certain *CFTR* regions from the BAC. These data illustrate general structural instability in the integrated BAC123s *Cla*I transgenes.

To determine BAC copy number by an independent, non PCR-based method, we evaluated each clone by Southern blot, following cleavage of genomic DNA with *Hind*III. Blots were probed separately with *CFTR* exons 4, 9 and 24 (3' UTR) as shown in Fig. 5a (probe locations shown in Fig. 1) and exposed to phosphorimager screens. Densitometric quantitation (by ImageJ) is shown in Fig. 5b (gray bars), along with relative qPCR copy numbers (black bars) for the clones shown on the Southern blots and the results generated by the two methods are generally consistent in each clone. The apparent discrepancy between the Southern blot and qPCR data for the two very high copy number clones shown in Fig. 5 (WTE1 and WTG1) is due to the intense signal being outside the linear range of the phosphorimager screen.

Finally, BAC copy number and integration site of the CMT-93-BAC123s *Cla*I clones were evaluated by fluorescence *in situ* hybridization (FISH) analysis on nine clones. BAC123s *Cla*I and the cW44 and A0726 cosmids (Rommens et al. 1989; Nizetic et al. 1994; Smith et al. 2000) were used as probes. These cosmids encompass the 5' and 3' end of the *CFTR* locus, respectively (probe locations shown in Fig. 1). The nine clones (WTA3, WTG1, WTH3, 11B2, 11B3, 11H1, 6.8A1, 6.8F1 and 6.8H3) covered the full range of expression and copy-number variation, and representative images are shown in Fig. 6. All clones had a single FISH signal on metaphase spreads (excluding spreads with >80 chromosomes; the CMT-93 cells used here were hyperdiploid, with a modal chromosome number of 50 (data not shown), consistent with the published karyotype (Franks and Hemmings 1978)), suggesting a single integration point or several integration points in a confined area. Clones with average qPCR-based copy numbers of less than six had discrete focal FISH signals on both metaphase and interphase nuclei, with their intensity correlating directly with copy number measured by qPCR and Southern blot (for example clone WTH3, Fig. 6a–b). In contrast, three clones with average qPCR copy numbers of greater than six that were assayed showed extensive regions of hybridization, often encompassing half of the chromosome on which they were located, and this signal corresponded to extended nuclear areas in interphase nuclei (for example clone 6.8F1, Fig. 6a–b). These unusual and apparently stable blocks of BACs suggest extensive replication accompanying or following integration into the mouse chromosome. Interestingly, despite the presence of these

structures and a weak correlation between high copy number and low BAC-derived *CFTR* expression, excluding high-copy clones from the data analysis did not decrease total variability in BAC-derived *CFTR* expression (data not shown).

Impact of BAC instability on the analysis of regulatory elements in the *CFTR* locus

One of our initial aims in this project was to assess whether we could successfully utilize bacterial invasion-mediated gene transfer of a 260kb BAC encompassing the whole *CFTR* locus to determine the contribution of critical *cis*-regulatory elements in the locus. These elements include a strong intestinal selective enhancer within intron 11 and an enhancer-blocking insulator element at +6.8 kb from the last coding base. We first determined that the intron 11 element showed hypersensitivity to DNase I in the WT BAC123s *Cla*I integrated in CMT-93 cells (Online Resource 1, Fig S2). This suggested that trans-acting factors necessary to drive the intron 11 enhancer element were present in this murine rectal cell line. Next we evaluated *CFTR* expression levels in a subset of CMT-93 clones that carried <6 BAC copies by qRT-PCR. Five clones with the WT BAC123s *Cla*I, 5 clones with BAC123s DHS11 *Cla*I and 8 clones with BAC123s DHS+6.8kb *Cla*I were analysed (Online Resource 1, Fig. S3). Though there is an apparent reduction in *CFTR* expression from both the deletion clones in comparison to the WT BAC, the large variation between copy number, expression levels and BAC integrity result in statistical significance not being reached for any construct. These results further illustrate the limitations of using invasion-mediated gene delivery as a method to investigate the *in vivo* function of *cis*-acting regulatory elements in a large BAC.

Discussion

BACs have been used extensively to develop genomic context vectors for several applications, including the generation of transgenic animal models and also in gene regulation studies. However, few studies have attempted a detailed characterization of the integration dynamics of these large vectors (Rostovskaya et al. 2012; Dubose et al. 2013). Here, we describe the delivery, by bacterial invasion, of three closely related BACs carrying the human *CFTR* locus into the murine rectal carcinoma cell line CMT-93. Multiple CMT-93 clones were obtained that stably expressed WT BAC123s *Cla*I or one of the two deletion constructs (BAC123s DHS11 *Cla*I and BAC123s DHS+6.8kb *Cla*I). We also compared the efficiency of bacterial invasion of the WT BAC123s with delivery into CMT-93 cells by Lipofectamine 2000. Among 41 zeocin-resistant clones generated by lipid delivery, only one contained and expressed a part of the *CFTR* locus, though all clones tested contained the zeocin-resistance gene. Bacterial invasion was much more efficient at mediating delivery and integration of the BAC constructs. However, the resultant clones showed a much more variable range of human *CFTR* expression than expected, based on a low copy number of BAC integrations. Moreover, not only was there substantial variation between clones carrying the same construct, the abundance of the 5' and 3' end of the *CFTR* transcript differed within individual clones, implicating instability within the transgene. Expecting that the overall clonal variation in *CFTR* expression was likely due in part to BAC copy number, we used qPCR, Southern blotting and FISH to measure this variation. *CFTR* copy number was also found to vary widely, both between clones and across the *CFTR* locus within individual clones. This variation is thought to be due to post-integration duplications or amplifications, given the very low probability that a single CMT-93 cell may be invaded by more than one BM4573 bacterial cell (in which the BAC is maintained at a copy number of 1–2 per *E. coli* genome) (Shizuya et al. 1992). Interestingly, in experiments performed concurrently with our study, others observed a similar phenomenon of variable copy number across BACs integrated into ES cells (Rostovskaya et al. 2012). In this case BACs were introduced into ES cells by lipid-based or electroporation delivery methods

(Lipofectin or Nucleofection), but the results are consistent. Also, development of a technique for generating high-resolution maps of large integrated transgenes recently revealed that structural rearrangement may be more common, and more severe, in BAC transgenic models than previously thought (Dubose et al. 2013).

Our data extend these observations by using FISH to directly visualize BAC transgene integrants. In addition to validating the copy number for low-copy clones, the FISH data provided two useful insights into the integration of BAC123s into CMT-93 cells. First, the FISH signal on interphase and metaphase nuclei was restricted to one site on a single chromosome, demonstrating that each clonal cell line (with a consistent CMT-93 karyotype, $n \approx 50$) resulted from either a single BAC integration event or multiple integrations in a restricted genomic area. The single chromosome carrying the BAC did not appear to be consistent between clones (based on chromosome size alone), suggesting a significant degree of randomness in the integration site. The observation of a single, random site even in clones with more than one copy of the transgene concurs with previous data using circular BAC DNA delivered by microinjection (Gong et al. 2003). Second, FISH revealed that clones with copy numbers of six or more (measured by qPCR) frequently exhibited unusual hybridization patterns when probed with the BAC123s *Cla*I BAC. On metaphase spreads, these patterns ranged from intense focal signals, suggesting a high concentration of BAC sequence in a small area, to broad signals covering up to half of a chromosome. Moreover, the presence of these unusual extended blocks of integrated BAC123s, and high BAC copy number in general, showed a negative correlation with transgene expression. This apparent inhibition of *CFTR* expression might result from a feedback mechanism which silences transcription of the gene when it is overexpressed from multiple copies of the locus. Alternatively the extensive fragmentation of the gene that occurs during the generation of the extended transgene blocks may repress transcription. These possibilities could be investigated further by combining high resolution mapping of the integrated transgenes with immuno-FISH to evaluate active and inactive histone marks. The mechanism whereby the transgene structures are stably maintained through cell division over many generations warrants further investigation.

Our data, suggest that BACs may be inherently unstable when integrated (or while integrating) into mammalian genomes, consistent with very recent observations by others (Rostovskaya et al. 2012; Dubose et al. 2013). While this phenomenon may not be restricted to BACs, there is evidence that other large genomic context vectors can be recovered intact following integration. Notably, yCFTR, a 310 kb YAC which completely overlaps BAC123s, was present in at least one correctly sized, full-length copy in transgenic mice (Manson et al. 1997). Moreover, it appears that the instability is directly related to integration, and not delivery, as BACs delivered using bacterial invasion which do not integrate into the genome can be recovered intact (Cheung et al. 2012). This suggests that the peri- or post-integration instability observed may be mediated by bacterial sequences present on the BAC vector backbone. This BAC-specific instability probably occurs during integration, rather than post-integration, since a transposon based integration system (which does not rely on random, homologous recombination mediated integration) resulted in stable integration of low-copy, intact BAC transgenes (Rostovskaya et al. 2012). Whatever the cause of the instability, these results highlight the need for careful characterization of large integrated transgenes (particularly BACs), prior to use in functional studies.

Materials and Methods

BAC recombineering with pBAC-RT

BAC123s *Cla*I BAC recombineering was performed in *E. coli* EL350 (Lee et al. 2001) using the pBAC-RT plasmid as described in (Dolphin and Hope 2006).

Cell Culture

The human colon carcinoma cell line Caco-2 (Fogh et al. 1977) and murine rectal carcinoma cell line CMT-93 (Franks and Hemmings 1978) were grown by standard methods in Dulbecco's Modified Eagle's Medium (DMEM) with 10% FCS. CMT-93 cells were used as they had previously been used for apparently successful BAC-mediated CFTR gene delivery (Kotzamanis et al. 2009)

Lipofectamine 2000 Transfection in CMT-93

CMT-93 cells were co-transfected 48 h after plating in 6-well dishes (2×10^5 cells/well) with 1 $\mu\text{g/mL}$ of BAC DNA using Lipofectamine 2000 (Life Technologies). After 24 h, the cultures were passaged into 10 cm dishes, and after a further 24 h, 100 $\mu\text{g/mL}$ zeocin was added. After 2–3 weeks under zeocin selection, individual colonies were isolated via tower cloning (P1) and expanded in 24-well plates. These cultures were passaged into 6-well plate wells (P2), which were subsequently passaged into T25 flasks (P3).

Bacterial invasion and stable cloning in CMT-93

Bacterial invasion in CMT-93 was performed in 6-well plates as described in (Kotzamanis et al. 2009) using *E. coli* BM4573 (Laner et al. 2005; Grillot-Courvalin 2010) (generous gift of Dr. Catherine Courvalin). Briefly, 2×10^5 CMT-93 cells/well were plated in six-well plates. 48 hours later, BM4573 containing the appropriate BAC were added to the cultures at a multiplicity of infection (MOI) of 100 in the presence of 20 μM chloroquine. The plates were centrifuged at $180 \times g$ for 5 min. and incubated at $37^\circ\text{C}/5\% \text{ CO}_2$ for 2 h. The cells were washed twice with PBS and incubated for 24 h in media supplemented with 20 $\mu\text{g/mL}$ Kanamycin. The cells were then transferred to 10 cm culture dishes and 100 $\mu\text{g/mL}$ zeocin selection was added after one more day incubation. After 2–3 weeks under Zeocin selection, individual colonies were isolated via tower cloning (P1) and expanded in 24-well plates. These cultures were passaged into two 6-well plate wells each (P2, 1:10), which were subsequently passaged into three T25 flasks and one T75 flask (P3, 1:7.5). At this passage, RNA was isolated from one T25 each using TRIzol, gDNA was isolated from the T75, and one T25 was fixed for FISH.

qRT-PCR

Total RNA for qRT-PCR was isolated using TRIzol (Life Technologies). *CFTR* expression was assayed by TaqMan qPCR with a probe and primer set spanning exons 5 and 6 (TAQEX5/6) (Mouchel et al. 2004) and by SYBR green qPCR using a primer set specific for the *Clal* polymorphism in the 3' UTR. 18S rRNA was used as an endogenous normalization control in both TaqMan and SYBR green qPCR. All qPCRs were performed in duplicate on at least two separate occasions.

Copy Number Measurement

Genomic DNA was isolated by standard methods (Strauss 2001). *CFTR* copy number was assayed by SYBR green qPCR using primer sets in exon 1, intron 10 and intron 20. 18S genomic DNA was used as an endogenous normalization control. Copy number was also assayed by Southern blot (using standard protocols (Brown 2001b, a)). Genomic DNA was digested with *Hind*III, separated on a 0.8% agarose gel and blotted onto a Hybond-N+ nylon membrane (GE Healthcare) by neutral transfer. Blots were probed with [^{-32}P]dCTP-labeled probes from *CFTR* exon 4, exon 9 and the 3' UTR as described in (Rowntree et al. 2001). The RA2.2 probe (Vyas et al. 1992) was used as an endogenous normalization control. Primers and probes used for copy number analysis are shown in Fig. 1.

Fluorescence *in situ* Hybridization (FISH)

Cells were prepared by briefly treating with a hypotonic solution (75 mM KCl) and then fixing in at least five changes of 3:1 absolute methanol:glacial acetic acid before spreading. FISH was performed according to standard methods (Espinosa and Le Beau 1997). Probes used for FISH analysis are shown in Fig. 1.

Primer Sequences

Sequences and genomic locations for primers used for (q)RT-PCR, copy number qPCR, Southern probe generation, and DNase I qPCR are shown in Online Resource 1, Table S1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by NIH R01HD68901 and R01HL094585 (PI:AH), the Cystic Fibrosis Foundation (PI:AH), a CABS Award from the Burroughs Wellcome Fund and an NIH/NIGMS New Innovator Award (DP2 OD008717) (PI:STK). We are grateful to Drs C. Huxley and G. Kotzamanis for sharing BAC123s; Drs C. Courvalin and S. Goussard for the *E. coli* BM4573 invasin strain; Dr C. DiDonato for the murine *Smn* primers; also Dr N. Blackledge and N. Gosalia for assistance.

References

- Brown T. Hybridization analysis of DNA blots. Curr Protoc Mol Biol. 2001a; Chapter 2(Unit2): 10.1002/0471142727.mb0210s21 [PubMed: 18265178]
- Brown T. Southern blotting. Curr Protoc Mol Biol. 2001b; Chapter 2(Unit2 9A)10.1002/0471142727.mb0209as21
- Calmels T, Parriche M, Durand H, Tiraby G. High efficiency transformation of Tolypocladium geodes conidiospores to phleomycin resistance. Current genetics. 1991; 20 (4):309–314. [PubMed: 1934136]
- Chalkley G, Harris A. Lymphocyte mRNA as a resource for detection of mutations and polymorphisms in the CF gene. Journal of medical genetics. 1991; 28 (11):777–780. [PubMed: 1770535]
- Chandler KJ, Chandler RL, Broeckelmann EM, Hou Y, Southard-Smith EM, Mortlock DP. Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. Mammalian genome : official journal of the International Mammalian Genome Society. 2007; 18(10):693–708.10.1007/s00335-007-9056-y [PubMed: 17882484]
- Cheung W, Kotzamanis G, Abdulrazzak H, Goussard S, Kaname T, Kotsinas A, Gorgoulis VG, Grillot-Courvalin C, Huxley C. Bacterial delivery of large intact genomic- DNA-containing BACs into mammalian cells. Bioengineered bugs. 2012; 3(2):86–92.10.4161/bbug.18621 [PubMed: 22095052]
- Devoy A, Bunton-Stasyshyn RK, Tybulewicz VL, Smith AJ, Fisher EM. Genomically humanized mice: technologies and promises. Nature reviews Genetics. 2012; 13(1):14–20.10.1038/nrg3116
- Dolphin CT, Hope IA. Caenorhabditis elegans reporter fusion genes generated by seamless modification of large genomic DNA clones. Nucleic Acids Res. 2006; 34(9):e72.10.1093/nar/gkl352 [PubMed: 16717278]
- Drocourt D, Calmels T, Reynes JP, Baron M, Tiraby G. Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. Nucleic Acids Res. 1990; 18 (13):4009. [PubMed: 1695734]
- Dubose AJ, Lichtenstein ST, Narisu N, Bonnycastle LL, Swift AJ, Chines PS, Collins FS. Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. Nucleic Acids Res. 201310.1093/nar/gks1463

- Espinosa R 3rd, Le Beau MM. Gene mapping by FISH. *Methods in molecular biology*. 1997; 68:53–76. [PubMed: 9055250]
- Fogh J, Wright WC, Loveless JD. Absence of HeLa cell contamination in 169 cell lines derived from human tumors. *J Natl Cancer Inst*. 1977; 58 (2):209–214. [PubMed: 833871]
- Franks LM, Hemmings VJ. A cell line from an induced carcinoma of mouse rectum. *The Journal of pathology*. 1978; 124(1):35–38.10.1002/path.1711240108 [PubMed: 722371]
- Giraldo P, Montoliu L. Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic research*. 2001; 10 (2):83–103. [PubMed: 11305364]
- Gogliotti RG, Hammond SM, Lutz C, Didonato CJ. Molecular and phenotypic reassessment of an infrequently used mouse model for spinal muscular atrophy. *Biochemical and biophysical research communications*. 2010; 391(1):517–522.10.1016/j.bbrc.2009.11.090 [PubMed: 19961830]
- Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*. 2003; 425(6961):917–925.10.1038/nature02033 [PubMed: 14586460]
- Grillot-Courvalin, CLA. Attenuated Invasive E. Coli Strains and Applications Thereof as Intracellular Vector for Therapeutic Molecule. US Patent US 2010/0216233. 2010. p. A1
- Harris A, Harris H, Hollingsworth MA. Complete suppression of tumor formation by high levels of basement membrane collagen. *Molecular cancer research : MCR*. 2007; 5(12):1241–1245.10.1158/1541-7786.MCR-07-0200 [PubMed: 18171981]
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res*. 2002; 12(6):996–1006.10.1101/gr.229102.ArticlepublishedonlinebeforeprintinMay2002 [PubMed: 12045153]
- Kotzamanis G, Abdulrazzak H, Gifford-Garner J, Haussecker PL, Cheung W, Grillot-Courvalin C, Harris A, Kittas C, Kotsinas A, Gorgoulis VG, Huxley C. CFTR expression from a BAC carrying the complete human gene and associated regulatory elements. *J Cell Mol Med*. 2009; 13(9A):2938–2948.10.1111/j.1582-4934.2008.00433.x [PubMed: 18657227]
- Laner A, Goussard S, Ramalho AS, Schwarz T, Amaral MD, Courvalin P, Schindelhauer D, Grillot-Courvalin C. Bacterial transfer of large functional genomic DNA into human cells. *Gene therapy*. 2005; 12(21):1559–1572.10.1038/sj.gt.3302576 [PubMed: 15973438]
- Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*. 2001; 73(1):56–65.10.1006/geno.2000.6451 [PubMed: 11352566]
- Magin-Lachmann C, Kotzamanis G, D'Aiuto L, Cooke H, Huxley C, Wagner E. In vitro and in vivo delivery of intact BAC DNA -- comparison of different methods. *The journal of gene medicine*. 2004; 6(2):195–209.10.1002/jgm.481 [PubMed: 14978773]
- Manson AL, Trezise AE, MacVinish LJ, Kasschau KD, Birchall N, Episkopou V, Vassaux G, Evans MJ, Colledge WH, Cuthbert AW, Huxley C. Complementation of null CF mice with a human CFTR YAC transgene. *EMBO J*. 1997; 16 (14):4238–4249. [PubMed: 9250667]
- Melcher R, Koehler S, Steinlein C, Schmid M, Mueller CR, Luehrs H, Menzel T, Scheppach W, Moerk H, Scheurlen M, Koehle J, Al-Taie O. Spectral karyotype analysis of colon cancer cell lines of the tumor suppressor and mutator pathway. *Cytogenetic and genome research*. 2002; 98(1):22–28. 68544. [PubMed: 12584437]
- Mouchel N, Henstra SA, McCarthy VA, Williams SH, Phylactides M, Harris A. HNF1alpha is involved in tissue-specific regulation of CFTR gene expression. *Biochem J*. 2004; 378(Pt 3):909–918.10.1042/BJ20031157 [PubMed: 14656222]
- Nizetic D, Monard S, Young B, Cotter F, Zehetner G, Lehrach H. Construction of cosmid libraries from flow-sorted human chromosomes 1, 6, 7, 11, 13, and 18 for reference library resources. *Mammalian genome : official journal of the International Mammalian Genome Society*. 1994; 5 (12):801–802. [PubMed: 7894163]
- O'Connor M, Peifer M, Bender W. Construction of large DNA segments in Escherichia coli. *Science*. 1989; 244 (4910):1307–1312. [PubMed: 2660262]

- Perez-Luz S, Abdulrazzak H, Grillot-Courvalin C, Huxley C. Factor VIII mRNA expression from a BAC carrying the intact locus made by homologous recombination. *Genomics*. 2007; 90(5):610–619.10.1016/j.ygeno.2007.07.005 [PubMed: 17822869]
- Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. 1989; 245 (4922):1059–1065. [PubMed: 2772657]
- Rostovskaya M, Fu J, Obst M, Baer I, Weidlich S, Wang H, Smith AJ, Anastassiadis K, Stewart AF. Transposon-mediated BAC transgenesis in human ES cells. *Nucleic Acids Res*. 2012; 40(19):e150.10.1093/nar/gks643 [PubMed: 22753106]
- Rowntree RK, Vassaux G, McDowell TL, Howe S, McGuigan A, Phylactides M, Huxley C, Harris A. An element in intron 1 of the CFTR gene augments intestinal expression in vivo. *Hum Mol Genet*. 2001; 10 (14):1455–1464. [PubMed: 11448937]
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A*. 1992; 89 (18):8794–8797. [PubMed: 1528894]
- Shizuya H, Kouros-Mehr H. The development and applications of the bacterial artificial chromosome cloning system. *The Keio journal of medicine*. 2001; 50 (1):26–30. [PubMed: 11296661]
- Smith DJ, Nuthall HN, Majetti ME, Harris A. Multiple potential intragenic regulatory elements in the CFTR gene. *Genomics*. 2000; 64(1):90–96.10.1006/geno.1999.6086 [PubMed: 10708521]
- Strauss WM. Preparation of genomic DNA from mammalian tissue. *Curr Protoc Mol Biol*. 2001; Chapter 2(Unit2):2.10.1002/0471142727.mb0202s42 [PubMed: 18265182]
- Vassaux G, Manson AL, Huxley C. Copy number-dependent expression of a YAC-cloned human CFTR gene in a human epithelial cell line. *Gene therapy*. 1997; 4(6):618, 623.10.1038/sj.gt.3300442 [PubMed: 9231079]
- Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, Higgs DR. Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell*. 1992; 69 (5):781–793. [PubMed: 1591777]

Fig. 1. BAC123s *Cla*I BAC coverage at the *CFTR* Locus

Diagram showing the coverage of BAC123s *Cla*I (dark gray) at the *CFTR* locus, along with the location of the primers, probes and cosmids (light gray) used to determine BAC123s copy number by qPCR, Southern blot and FISH, respectively. Primers used for DNase I qPCR (supplemental data) are also shown. Figure based on output from <http://genome.ucsc.edu> (hg19 assembly) (Kent et al. 2002).

Fig. 2. Partial *CFTR* mRNA expression from a BAC123sCla1 CMT-93 clone generated by Lipofectamine 2000

(a) Expression of human *CFTR*, measured by RT-PCR using the primers indicated on the left. Note that clone CMT-93–100A13 has detectable product with the GV primer set (3' UTR), but not with C1R/C2L (exons 11–13). (b) Expression of murine SMN, measured by RT-PCR. (c) Expression of the zeocin resistance gene from BAC123s, measured by RT-PCR. D98AP2 15:58 (Harris et al. 2007) is included as a positive control. (d) Schematic representation of the *CFTR* mRNA (dark gray, with 5' and 3' untranslated regions in light gray), with the location of primers used in RT-PCR and qRT-PCR indicated below.

Fig. 3. BAC-derived *CFTR* expression in CMT-93 BAC123s *Cla* I clones

Relative expression of hCFTR on a log₁₀ scale measured by TaqMan qPCR using a primer/probe set in exon 5/6a (black) and by SYBR green qPCR using a *Cla* I-specific primer set in the 3' UTR (white). Both data sets are normalized to 3' UTR expression in clone 11B4, which has the lowest detectable expression among clones included in this analysis. Clones are arranged by BAC and in order of increasing exon 5/6a expression (left to right). Error bars show SEM. Inset shows pulsed field gel analysis of *Nru* I digests of BAC123s *Cla* I ("WT") in *E. coli* EL350 and BM4753, along with BAC123s DHS11 *Cla* I ("DHS11") and BAC123s DHS+6.8kb *Cla* I ("DHS+6.8kb") in *E. coli* BM4573. Visible ladder markers are 97kb (lower) and 145.5kb (upper).

Fig. 4. BAC123s copy number in CMT-93 BAC123s *Cla* I clones

BAC123s *Cla* I copy number, measured by SYBR green qPCR using primer sets in exon 1 (CFTR-gDNA-5 ; black), intron 10 (CFTR-gDNA-Mid; gray) and intron 20 (CFTR-gDNA-3 ; white). Error bars show SEM.

Fig. 5. Representative CMT-93 BAC123s *Cla* I southern blots

(a) Southern blots of *Hind* III-digested genomic DNA from parental Caco-2, CMT-93, and CMT-93 BAC123s *Cla* I clones, probed with [32 P]dCTP-radiolabeled DNA probes from *CFTR* exon 4, exon 9 and 3'-UTR. The RA2.2 probe (Vyas et al. 1992) was used as a loading control. (b) Relative *CFTR* transgene quantitation by densitometry from 'a' (gray bars) with corresponding qPCR-derived copy number (black bars). Gray bars represent probes for exon 4, exon 9 and the 3' UTR (left to right for each triplet). Each probe, normalized to RA2.2, is shown relative to clone Δ6.8H3. Black bars are PCR products from exon 1 (CFTRgDNA-5'), intron 10 (CFTR-gDNA-Mid) and intron 20 (CFTR-gDNA-3') (left to right for each triplet). Error bars show SEM.

Fig. 6. FISH in low- and high-copy CMT-93 BAC123s *Cla* I clones

FISH using the BAC123s *Cla* I BAC (red) and A0726 cosmid (green) on CMT-93 BAC123s clones WTH3 (low copy) and $\Delta 6.8F1$ (high copy) on (a) interphase nuclei and (b) metaphase spreads. Error bars show SEM.

Table 1CMT-93 BAC123s *Cla*I Lipofectamine 2000 clone summary

CMT-93 BAC Clone	hCFTR+ (RNA)	hCFTR+ (DNA)	SMN+	Zeo+
CMT-93 BAC123s <i>Cla</i> I	1 [*] /41	1 [*] /5	41/41	10/10

* Human CFTR mRNA and genomic DNA in clone 100A13 was only detected at Exon 24/3' UTR.

Table 2Summary of Zeo^R CMT-93 clones generated by BAC123s *Cla*I bacterial invasion.

CMT-93 BAC 123s <i>Cla</i> I Clone	hCFTR Expression (vs. 11B4)		Average Copy Number (qPCR)	FISH
	Exon 5/6a	3' UTR		
WTA3	64382.0	1469.0	1.6	Y
WTH3	15023.8	34437.2	1.6	Y
WTC1	8260.0	7041.8	1.3	
WTF1	4386.2	755.6	0.5	
WTE1	1050.8	143.0	14.5	
WTB1	703.6	782.6	5.9	
WTA2	170.8	1238.4	2.1	
WTG1	3.6	253.8	17.3	Y
WTA1	0.0	N/A	0.2	
11B2	8267.0	1329.6	0.8	Y
11A3	1814.6	103.2	2.7	
11B1	128.2	471.0	12.4	
11B5	10.4	144.4	4.7	
11B4	7.0	1.0	5.7	
11B3	3.2	3.2	5.4	Y
11A2	2.8	16.8	0.4	
11H1	2.8	192.4	10.4	Y
11D2	0.0	N/A	0	
6.8K2	24570.0	3958.2	2.6	
6.8I1	15851.2	5866.0	2.1	
6.8E1	8582.0	2576.6	2.3	
6.8A1	3411.2	1678.0	3.2	Y
6.8H3	2157.0	932.4	0.6	Y
6.8G1	1718.4	12.8	1.3	
6.8J1	23.2	739.4	0.9	
6.8F1	18.4	29.4	7.4	Y
6.8G2	5.0	280.0	0.2	
6.8H1	2.0	2.0	1.1	
6.8B2	0.0	N/A	0	