



Bacterial Artificial Chromosome Libraries for Mouse Sequencing and Functional Analysis

Kazutoyo Osoegawa, Minako Tateno, Peng Yeong Woon, et al.

Genome Res. 2000 10: 116-128

Access the most recent version at doi:[10.1101/gr.10.1.116](https://doi.org/10.1101/gr.10.1.116)

References This article cites 25 articles, 6 of which can be accessed free at:
<http://genome.cshlp.org/content/10/1/116.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Bacterial Artificial Chromosome Libraries for Mouse Sequencing and Functional Analysis

Kazutoyo Osoegawa,^{1,5} Minako Tateno,^{1,2} Peng Yeong Woon,^{1,3} Eirik Frengen,^{1,4} Aaron G. Mammoser,¹ Joseph J. Catanese,^{1,5} Yoshihide Hayashizaki,² and Pieter J. de Jong^{1,5,6}

¹Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, New York 14263 USA; ²Genome Science Laboratory, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), Tsukuba, Ibaraki 305 Japan; ³The Wellcome Trust Centre For Human Genetics, University of Oxford, Headington, Oxford OX3 7BN, UK; ⁴The Biotechnology Centre of Oslo, University of Oslo, N-0349 Oslo, Norway

Bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) libraries providing a combined 33-fold representation of the murine genome have been constructed using two different restriction enzymes for genomic digestion. A large-insert PAC library was prepared from the 129S6/SvEvTac strain in a bacterial/mammalian shuttle vector to facilitate functional gene studies. For genome mapping and sequencing, we prepared BAC libraries from the 129S6/SvEvTac and the C57BL/6J strains. The average insert sizes for the three libraries range between 130 kb and 200 kb. Based on the numbers of clones and the observed average insert sizes, we estimate each library to have slightly in excess of 10-fold genome representation. The average number of clones found after hybridization screening with 28 probes was in the range of 9–14 clones per marker. To explore the fidelity of the genomic representation in the three libraries, we analyzed three contigs, each established after screening with a single unique marker. New markers were established from the end sequences and screened against all the contig members to determine if any of the BACs and PACs are chimeric or rearranged. Only one chimeric clone and six potential deletions have been observed after extensive analysis of 113 PAC and BAC clones. Seventy-one of the 113 clones were conclusively nonchimeric because both end markers or sequences were mapped to the other confirmed contig members. We could not exclude chimerism for the remaining 41 clones because one or both of the insert termini did not contain unique sequence to design markers. The low rate of chimerism, ~1%, and the low level of detected rearrangements support the anticipated usefulness of the BAC libraries for genome research.

[The sequence data described in this paper have been submitted to the GenBank data library under accession numbers AQ797173–AQ797398.]

Laboratory mice represent an extraordinarily valuable model organism to seek genes involved in diseases and investigate the disease mechanisms (Wynshaw-Boris 1996; Smith et al. 1997). The laboratory mice offer a manipulative biological model permitting genome structure-function analysis. Bacterial artificial chromosome (BAC; Shizuya et al. 1992) and P1-derived artificial chromosome (PAC; Ioannou et al. 1994) cloning procedures have been developed to generate large-insert genomic libraries with acceptable genomic fidelity as compared with yeast artificial chromosome (YAC) cloning systems. In addition to the increased clone fidelity and low levels of cloning artifacts, BACs and PACs can more easily be separated from host DNA than YACs. Consequently, BACs and PACs provide excellent templates for shotgun sequencing strategies

and have become the main sequence-ready clone resources for use in large-scale mapping and sequencing efforts (Venter et al. 1996; Gregory et al. 1997; Marra et al. 1997). 129S6/SvEvTac and C57BL/6J female mice (Simpson et al. 1997) were selected as the most useful initial inbred strains from a number of different inbred strains. The PAC and BAC libraries were both constructed to provide each 10-fold genome equivalents representing 99.995% statistical probability of obtaining at least one clone containing unique sequence (Woon et al. 1998; Dunham et al. 1999). Traditionally, fluorescent in situ hybridization (FISH) has been used as the preferred approach to examine possible chimerism of YAC clones (Green et al. 1991; Haldi et al. 1994). Multiple cytogenetic map locations for a particular YAC clone result most often from unrelated genomic segments cocloned into the same YAC. At a much lower level of likelihood, multiple YAC hybridization sites are caused by duplicated genomic regions such that the same genomic segments are present in multiple locations. Therefore, the FISH approach is less

⁵Present address: Children's Hospital, Oakland Research Institute, 747 Fifty Second Street, Oakland, California 94609-1809 USA.

⁶Corresponding author.

E-MAIL Pieter.deJong@wl.com; FAX (510) 749-4266.

useful for examining the anticipated low levels of chimerism in BACs and PACs. This is mainly because the chimeric clone levels are possibly lower than the occurrence of the genomic duplication background. In addition, the mechanism leading to chimeric clones in BACs or PACs is probably different. Hence, there is no reason to presume that unrelated genomic segments in chimeric clones can have random size ratios as frequently observed for YACs. It is perhaps more likely that chimeric BACs consist of one very large and one very small fragment. This is due to the much stronger size bias in bacterial transformation as compared with yeast transformation. Hence, chimeric BACs with two large fragments would presumably only be generated at very low efficiency. A priori, it appears therefore more likely that large chimeric BACs consist of one big and one small fragment. It would be difficult to determine the chimeric nature of such BACs using a FISH mapping strategy as is commonly used (Ioannou et al. 1994; Woon et al. 1998; Korenberg et al. 1999). The most conclusive way to determine whether a clone is not chimeric is to show that both ends of the clone map to the same small genomic region, for instance, to other clones within a confirmed BAC or PAC contig. This approach is relatively independent of low-level genomic duplications, because only the members of the confirmed contig are considered. In addition, chimerism would be detectable even when the unrelated genomic segments in a postulated chimera differ in

size by an order of magnitude. FISH would be less sensitive to identify such unbalanced chimeric clones because the smaller fragment would contribute a relatively weak signal to the hybridization picture. After our extensive characterization, it has become clear that chimeric clones are very rare and occur on the order of ~1%, and it has also been determined that BAC and PAC clones are rather stable.

RESULTS

Construction of Mouse PAC and BAC Libraries and Their Characteristics

Table 1 summarizes the characteristics of one PAC and two BAC libraries prepared from the 129S6/SvEvTac and C57BL/6J strains. The RPCI-21, 129-murine PAC library was prepared from six separate large-scale ligations of *Mbo*I partially digested genomic DNA and the *Bam*HI-cut pPAC4 vector. The resulting 254,217 clones have been picked into 384-well plates. For the first half of this PAC library, the genomic DNA was isolated after a single sizing step, resulting in a rather wide distribution of clone insert sizes with an average of 127 kb. The second segment of the library was prepared from *Mbo*I fragments subjected to two sizing steps, as recently described (Osoegawa et al. 1998). The resulting 130,000 clones have a decreased number of small clones, hence an increased average insert size of 147 kb. Two mouse BAC libraries, designated RPCI-22 and RPCI-23, were

Table 1. Mouse Libraries Described in this Paper

RPCI-21						
DNA source: 129S6/SvEvTac female spleen						
Cloning vector: pPAC4						
Segment	Cloning enzyme	Total clones	Number of plates	Noninsert clones (%)	Insert size (kb)	Genomic representation
1	<i>Mbo</i> I	125,318	336	0.5	127	5.3
2	<i>Mbo</i> I	128,899	336	0.5	147	6.3
Total		254,217		0.5	137	11.6
RPCI-22						
DNA sources: 129S6/SvEvTac female spleen						
Cloning vector: pBACe3.6						
1	<i>Eco</i> RI	102,914	288	5.5	155	5.3
2	<i>Eco</i> RI	109,039	288	0.1	154	5.6
Total		211,953		2.7	154	10.9
RPCI-23						
DNA source: C57BL/6J kidney and brain						
Cloning vector: pBACe3.6						
1	<i>Eco</i> RI	80,661	240	11.3	198	5.3
2	<i>Eco</i> RI	89,973	240	1.3	197	5.9
Total		170,634		6.0	197	11.2

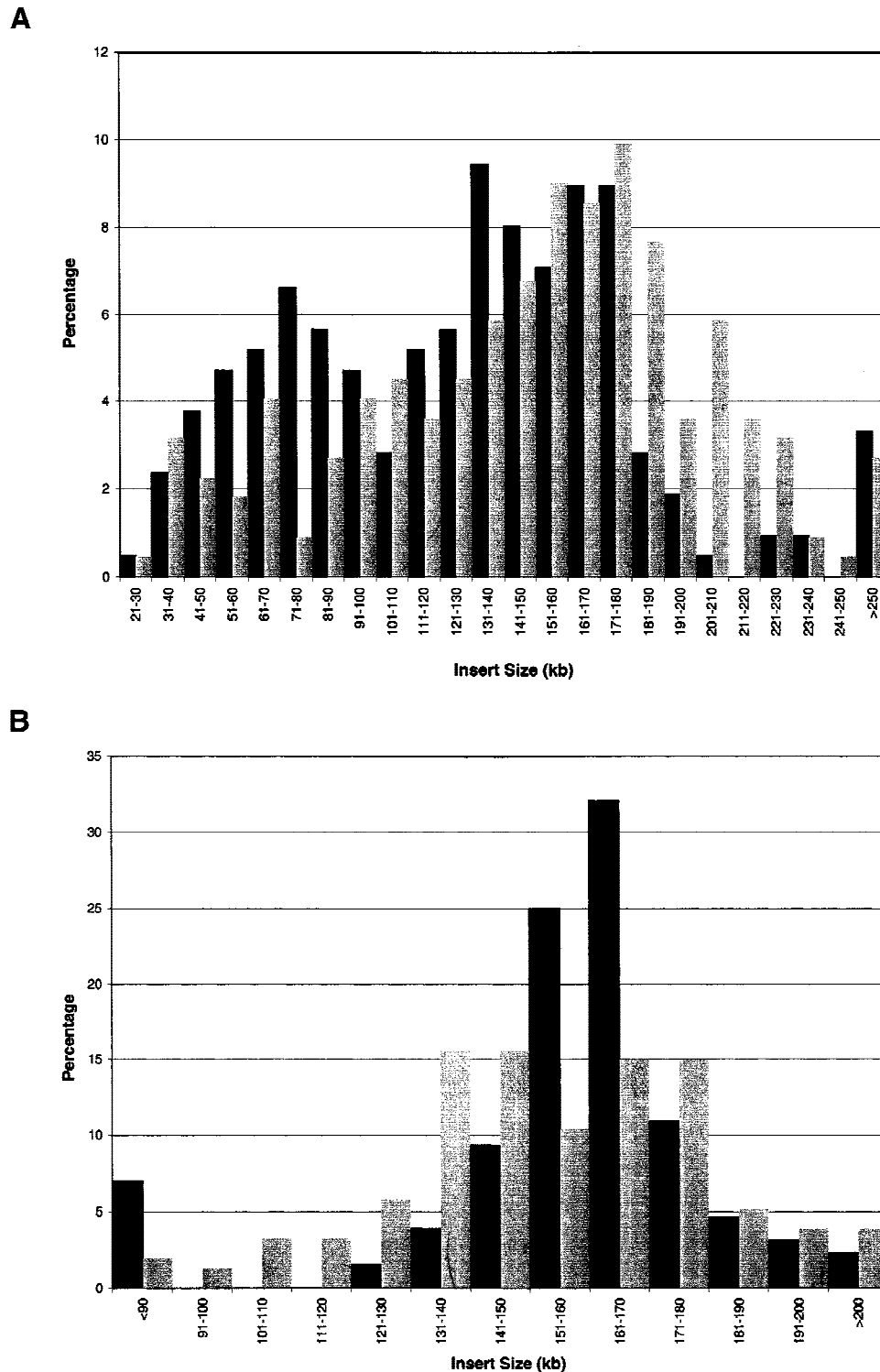


Figure 1. (Continued on page 119.)

constructed in the pBACe3.6 vector from *EcoRI*-digested DNA from the 129S6/SvEvTac and C57BL/6J strains, respectively. The RPCI-22 library was constructed from two separate large-scale ligations, and

the RPCI-23 library was prepared using nine ligations. The average insert sizes of all libraries were determined to be 137 kb (RPCI-21), 154 kb (RPCI-22), and 197 kb (RPCI-23). The size distributions of genomic inserts for

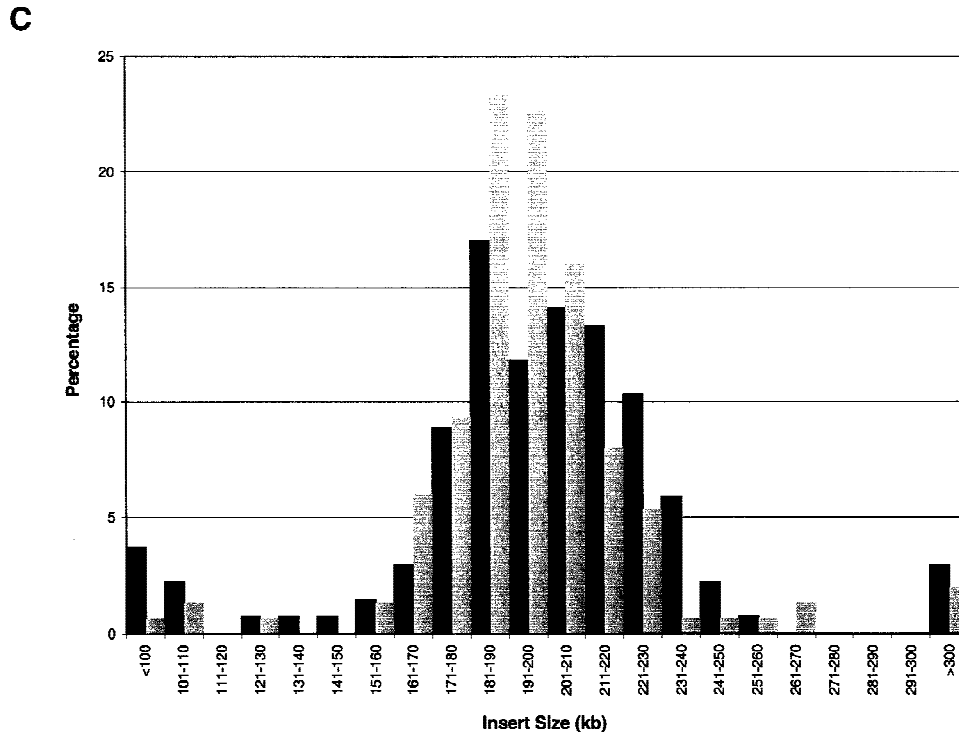


Figure 1 The insert size distributions of RPCI-21 (A), RPCI-22 (B), and RPCI-23 (C) libraries. A total of 434 clones from RPCI-21, 282 from RPCI-22, and 285 from RPCI-23 were selected randomly, taking separate ligations and transformations into consideration. The horizontal axis refers to the size range of insert DNA, and the vertical axis indicates the percentage of clones corresponding to each size range. Black bars and gray bars correspond to segment 1 and segment 2, respectively.

each library are shown in Figure 1. A small proportion of the clones lacked inserts and contained a smaller vector fragment. Such deleted vector molecules have lost part of the positive selection mechanism depending on the *SacB* gene expression, hence explaining the growth of the colonies in the presence of sucrose. The proportion of noninsert clones was determined to be 0.5%, 2.7%, and 6.0% of the total clone numbers for the three libraries, respectively. It was determined that 3.4% wells contain multiple clones by comparing duplicate restriction enzyme fingerprint patterns as described below. Based on the number of clones and average insert sizes, we estimate that the libraries provide 11.4-fold (RPCI-21), 10.9-fold (RPCI-22), and 11.2-fold (RPCI-23) mouse genome representation.

Screening the Libraries

Comparative anchor tagged sequence (CATS) markers are derived from the conserved sequences of unique genes (Lyons et al. 1997). Twenty-two markers were successfully amplified from 129-strain genomic DNA as single PCR products and then used for hybridization-based screening of the three libraries. An unusually large number of positive clones were identified in the RPCI-21 library using the *AMY1A*, *GAPD*, and *LDHA* probes. Most of these clones could also be con-

firmed by PCR analysis. Nevertheless, most of the abundant clones belonged to multiple contigs as determined by fingerprint and Southern mapping (not shown). This indicates that these markers are not unique in the mouse. The remaining 19 probes were used to evaluate the genome representation in all libraries. The PCR products amplified from the PAC templates were sequenced and analyzed using the BLASTN program (Altschul et al. 1990) to ascertain whether the PCR products correlate with the original gene. The product generated with the *BGN* primers did not correspond to the *BGN* sequence. The sequences derived from the "*GBA*" and "*IGH*" PCR products revealed homology with glucocerebrosidase (*GCB*) and major histocompatibility complex (MHC) class II *H2-IE- α* genes. The other 16 PCR-product sequences showed strong homology with their corresponding genes from human, mouse, or rat. All positive clones derived from these 19 probes were analyzed by Southern hybridization, allowing the assembly of 19 small contigs consistent with the idea that unique markers were used to identify all clones (Figure 2). An additional nine unique markers described in Methods were used to screen the libraries. The average genome redundancies of the RPCI-21, RPCI-22, and RPCI-23 libraries using all 28 unique markers were determined to be 9.3, 10.6, and 13.9 per marker with s.d.s of 3.5, 3.9, and 6.0,

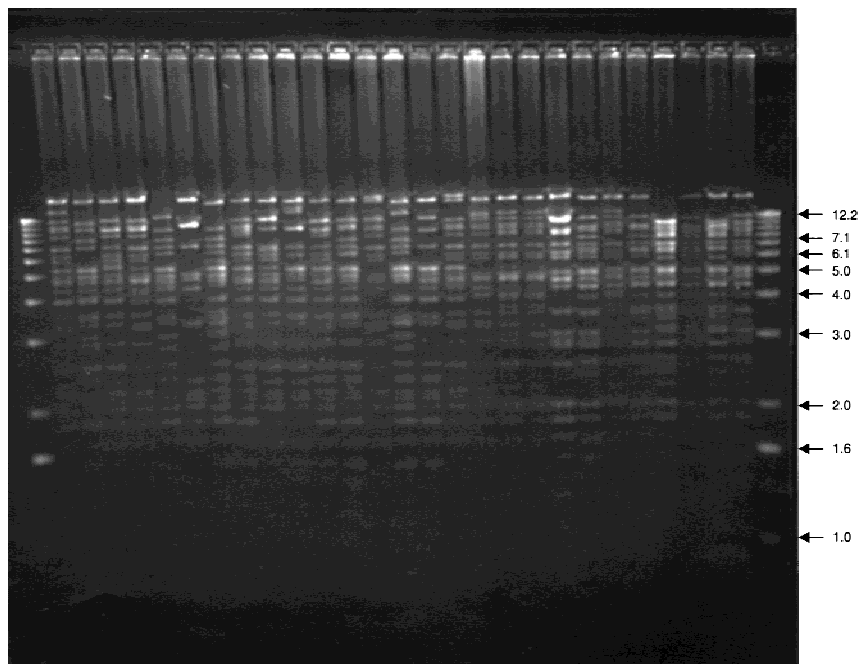


Figure 2 Fingerprinting and Southern hybridization have been applied to confirm positive clones and clone integrity. PAC and BAC DNAs from RPCI-21 and RPCI-23 clones from the CAT region were isolated and digested with *EcoRI*. The 1-kb ladder DNA marker was loaded on both sides of the gel.

respectively. Table 2 summarizes screening results from RPCI-21, -22, and -23.

End Sequences from BACs and PACs

The *CAT*, *IGFBP1*, and *MLR* gene contigs were chosen to investigate whether the clones faithfully represent the genomic DNA and whether cloning sites are used randomly. Detailed comparison of cloned genomic DNA with the uncloned genome is laborious and not practical for large genomic regions. To circumvent this problem, we presume that the rearrangements in the genomic DNA occur in an irreproducible fashion. If the clones misrepresent the genomic DNA, then different clones would have different rearrangements. Hence, our goal is the detailed comparison of many related clones derived from independent cloning events. Consistent genomic DNA patterns between clones provide confidence that the genome is faithfully represented in the libraries. The strategy is to prepare many additional unique markers derived from the contig members and use these as probes to characterize the overlapping clones for the presence or absence of the marker sequence. After single colony isolation, all end sequences from BAC and PAC clones belonging to the three contigs were determined utilizing T7 and SP6 promoter flanking sequences as described in Methods. The end sequences can be found in GenBank under accession numbers AQ797173–AQ797398. A total of 76 ends from 38 PACs from three contigs were sequenced. Only

two insert ends appeared to be identical: the sequence for RPCI-21 402G5 (SP6-end) was the same as RPCI-21 581N18 (SP6-end) (Fig. 3B). This demonstrates there are no preferential digestion at any specific *MboI* site in the three regions. For the 75 BACs in the three contigs, we sequenced 150 ends. A small majority of the end sequences (86 out of 150) represent unique sites in the collection. The remaining 64 *EcoRI* cloning ends were derived from 26 *EcoRI* sites. Two of the clones (RPCI-23 98E13 and RPCI-23 453A4) in the *IGFBP1* and two of the clones (RPCI-23 142P1 and RPCI-23 298M17) in the *MLR* region had identical end sequences for both insert ends. These clones were derived from different ligation and transformation reactions, indicating that these identical clones were derived from independent cloning events. For one of the contigs, for the *IGFBP1* marker, 6 of 28 BACs started at the same *EcoRI* site,

suggesting preferential cleavage of this site in the *IGFBP1* region.

Clone Chimerism

To determine whether any of the clones in the three contigs contained chimeric inserts, we used the BAC and PAC end sequences to design hybridization probes for screening all the overlapping clones. Hybridization probes were prepared as 3'-overlapping oligonucleotide (overgo) probes filled in with radioactive nucleotide triphosphates using the procedure designed by John McPherson (1999). Essentially, the end sequences are first screened against the rodent repeat database using RepeatMasker (A.F.A. Smit and P. Green, unpubl.), and the remaining unique sequences are used for the design of unique overgo probes with similar duplex stability. Some end sequences cannot be used because they contain only repetitive sequences. All clone-end probes should map to the overlapping clones on colony hybridization filters, except if the clone end is also the contig end or if the clone end is derived from an unrelated genomic region in the case of a chimeric clone. Therefore, we decided to map all unique end probes to the overlapping clones from three contigs to ascertain the possibility of chimerism. This also permits us to characterize the clones for consistency by comparing the empirical hybridization results with the results predicted from the linear map order of clones and markers. Inconsistent marker hy-

Table 2. Screening Libraries Using Various Single Locus Markers

Markers	RPCI-21	RPCI-22	RPCI-23	Locus
Comparative anchor-tagged sequences (CATS)				
Aminolevulinate, δ , syntase 1 (<i>ALAS2</i>)	10	8	14	X.63
Biglycan (<i>BGN</i>)*	7	9	10	
Catalase (<i>CAT</i>)	17	7	10	2.57
Ceruloplasmin (ferroxidase) (<i>CP</i>)	6	13	6	9.53
Cytochrome P-450, subfamily XIX (<i>CYP19</i>)	4	5	13	9.31
Dopamine Receptor (<i>DRD2</i>)	10	11	9	9.28
Factor 9 (<i>F9</i>)	8	6	11	X.22
Fibronectin 1 (<i>FN1</i>)	9	9	23	1.36
Glucosidase, β ; acid (<i>GBA</i>)*	11	8	24	
Glucose-regulated protein (78 kD) (<i>GRP78</i>)	13	12	3	2.23
Insulin-like growth factor binding prot. 1 (<i>IGFBP1</i>)	16	14	17	11.02
Immunoglobulin heavy chain gene cluster (<i>IGH</i>)*	12	17	16	
Mineralcorticoid (aldosterone) receptor (<i>MLR</i>)	7	10	20	8.35
Moloney murine sarcoma oncogene homol. (<i>MOS</i>)	8	10	16	4.00
Myeloperoxidase (<i>MPO</i>)	7	7	19	11.50
Neurofibromin 1 (<i>NF1</i>)	13	16	18	11.46
Paired box homeotic gene (<i>PAX8</i>)	17	4	14	2.10
Avian reticuloendotheliosis viral onc. hom. (<i>REL</i>)	9	11	21	11.13
Thy-1 cell surface Antigen (<i>THY1</i>)	4	15	19	9.26
Exon				
HPRT	10	12	10	X.17
End probes from BAC/PAC				
3-17D3-T	7	10	23	2.57
3-209M19-T	9	12	6	2.57
1-439F16-T	11	14	15	11.02
3-20C9-T	7	20	11	11.02
3-233F12-S	6	13	6	8.35
3-458o24-S	7	12	15	8.35
EN21	5	9	3	
EN31	11	4	17	
Average number of positive clones	9.3	10.6	13.9	

The locus column indicates the mouse chromosome location of each probe. For example, the *HPRT* gene is located on chromosome X, at the 17-cM position. The *BGN*, *GBA*, and *IGH* probes are marked with an asterisk (*). No homology was found to the *BGN* sequence using the database, whereas the *GBA* and *IGH* PCR products showed homologies against glucocerebrosidase (*GCB*) and MHC class II *H2-IE- α* genes. However, the original names are used here.

bridization results can provide an indication of clone rearrangements and deletions. Detailed clone and marker analysis was applied to the three previously discussed contigs established for the *CAT*, *IGFBP1*, and *MLR* markers.

Figure 3A shows the results for the *CAT* contig, which includes 34 PACs or BACs. Of the 68 clone ends, 4 ends could not be used to design a unique hybridization overgo probe. The remaining 64 clone ends represent 58 distinct end sequences (Fig. 3A). For 31 of the 34 clones, we were able to design unique overgo probes for both ends. Most of these overgo pairs (29 out of 31) map back to overlapping contig members, thus confirming the 29 clones to be nonchimeric. For the two remaining pairs, only one end probe maps to the other contig members. However, the other end probes represent the two ends of the contig. We independently mapped these ends by expanding the contig in both directions (Table 2) and found that these two clones

were also nonchimeric (data not shown). The final 3 (out of 34 clones) could not be ascertained with respect to the possibility of chimerism because at least one of the two end probes was repetitive within the contig region.

Figure 3B shows the results for the *IGFBP1* contig, which includes 43 PACs and BACs. Out of the 86 insert ends, 54 could be used for the design of 45 distinct overgo probes. The other ends either lack enough unique sequence (21 out of 86) or represent multiple occurrences of the same clone-end sequences within the contig (11 out of 86). We were able to design unique overgo probes from both ends for only 14 out of the 43 clones. Thirteen probe pairs map to the other contig members, thus confirming the 13 clones to be nonchimeric. From the remaining probe pair (clone RPCI-23 20C9), one of the probes could not be mapped because it represented a contig end. We independently mapped this end by expanding the contig (Table 2)

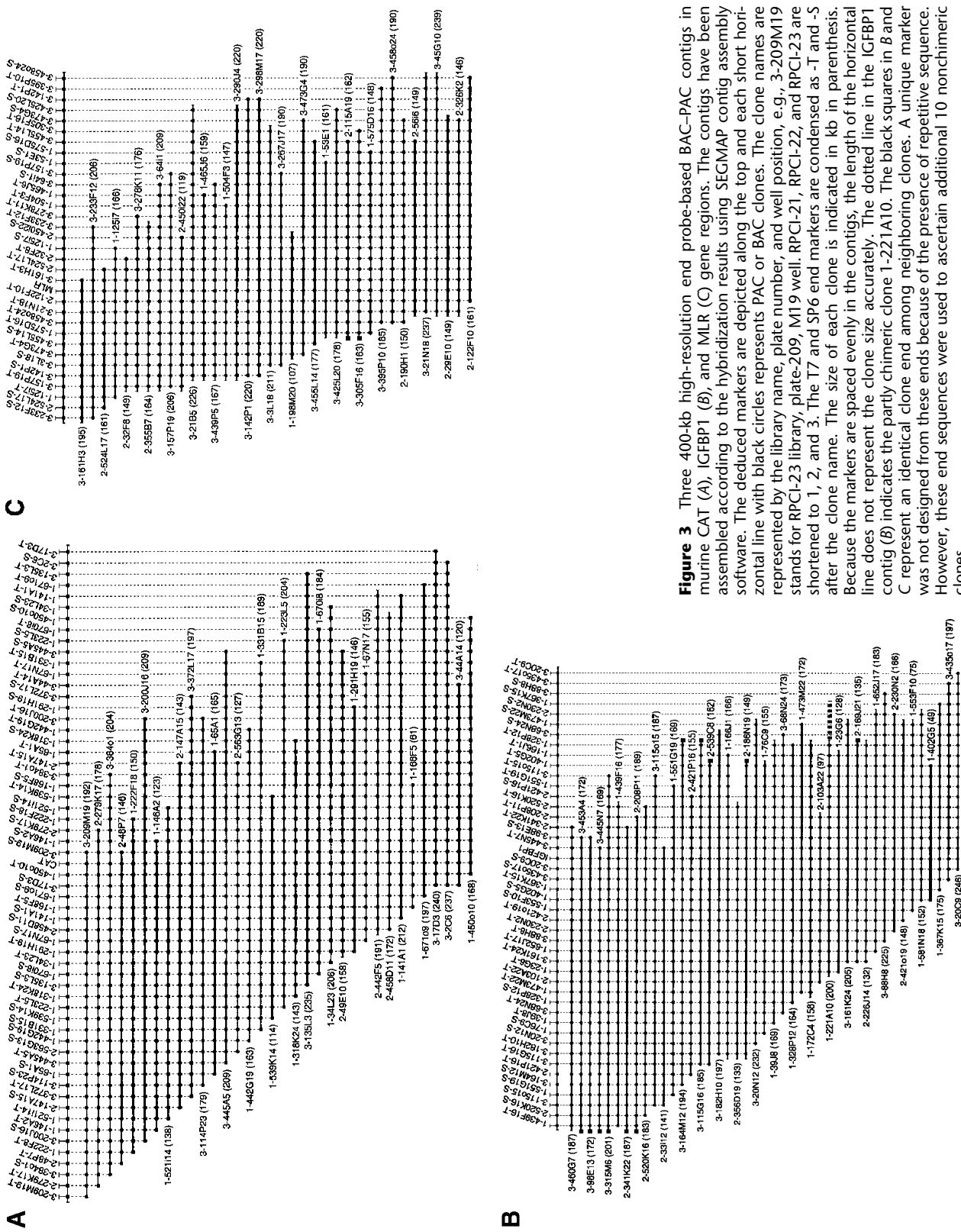


Figure 3 Three 400-kb high-resolution end probe-based BAC-PAC contigs in murine CAT (A), IGFBP1 (B), and MLR (C) gene regions. The contigs have been assembled according to the hybridization results using SEGMAP contig assembly software. The deduced markers are depicted along the top and each short horizontal line with black circles represents PAC or BAC clones. The clone names are represented by the library name, plate number, and well position, e.g., 3-209M19 stands for RPC1-23 library, plate-209, M19 well. RPC1-21, RPC1-22, and RPC1-23 are shortened to 1, 2, and 3. The T7 and SP6 end markers are condensed as -T and -S after the clone name. The size of each clone is indicated in kb in parenthesis. Because the markers are spaced evenly in the contigs, the length of the horizontal line does not represent the clone size accurately. The dotted line in the IGFBP1 contig (B) indicates the partly chimeric clone 1-221A10. The black squares in B and C represent an identical clone end among neighboring clones. A unique marker was not designed from these ends because of the presence of repetitive sequence. However, these end sequences were used to ascertain additional 10 nonchimeric clones.

and found that this clone was also nonchimeric (data not shown). For the other contig end, we were unable to design a probe because of the absence of unique sequences. Although it was not possible to design unique end probe pairs for 29 of the 43 contig clones, we were nevertheless able to determine that 10 out of 29 clones were also nonchimeric. This was determined based on the presence of identical sequence ends shared between multiple clones in the contig. Six of these 10 clones share the same identical repetitive end, which may be the left end of the contig in Figure 3B. As a result, 24 of the 43 clones were confirmed to be nonchimeric. Eighteen clones (out of 43) could not be ascertained with respect to the possibility of chimerism because at least one of the two ends lacked unique sequences. The final clone out of 43 (RPCI-21 221A10) was determined to be chimeric. This clone had sufficient unique sequence at one end to design a unique overgo probe, whereas the other end lacked unique sequence. Clone 221A10 definitely belonged to the contig because it was positive for 23 contig markers. However, the overgo probe from 221A10-T7 did not map back to other contig members and was also not at the contig end. We thus conclude that the clone is chimeric.

Figure 3C shows the third contig established around the *MLR* marker, which includes 36 BACs and PACs. For 14 out of 36 clones, it was possible to design unique probes for both ends. These probes all mapped back to other contig members, confirming these 14 clones as nonchimeric. Two additional clones (out of 36) were determined to be nonchimeric because one end probe mapped back to the contig and the other end shared the identical repetitive end sequence with other contig clones. Twenty (out of 36) clones could not be ascertained with respect to the possibility of chimerism because at least one of the two ends lacked unique sequences.

Possible Clone Rearrangements

A total of 59, 45, and 33 distinct markers including the original markers have been assigned to the *CAT*, *IGFBP1*, and *MLR* contigs, respectively (Fig. 3). Based on approximate contig sizes of 400 kb, the average marker spacings are calculated to be 6.8, 8.9, and 12.1 kb, respectively. Because most of the markers are derived from clone ends, it was possible to determine an unambiguous linear order for all markers using the SEG-MAP program. This allowed for the establishment of consistent colony hybridization patterns. In other words, if a BAC or PAC clone is positive for two distant markers, then it should also be hybridization positive for all markers internal to the distant markers. A clone deletion would be indicated by a negative hybridization result for one or more internal markers. Using all of the clones and all of the 137 markers in colony hy-

bridizations, we have not found a single inconsistent hybridization result. This indicates the absence of large deletions spanning more than one marker interval. However, these screening results do not exclude deletions much smaller than the average 10-kb marker intervals. To increase the sensitivity of our screening for deleted clones and also to permit rearranged clones to be detected, Southern blot hybridizations were performed for all clones and 77 randomly selected markers. For most of markers (73 out of 77), we did not detect any differences between the *EcoRI* restriction fragment sizes for hybridization-positive clones. For the remaining markers (4 out of 77), we observed two distinct *EcoRI* fragment sizes. However, these distinct sizes were always shared with multiple clones and were strain specific (129S6/SvEvTac vs. C57BL/6J), indicating polymorphisms rather than rearrangements. The Southern blot screening results indicate the absence of small deletions or rearrangements detectable within the resolving power of the agarose gel electrophoresis (few hundred base pairs up to a few kilobase pairs).

In addition to the Southern hybridization, fingerprinting analysis (Marra et al. 1997) has been applied to detect rearrangements within the three contigs that include 113 BAC and PAC clones. The fingerprinting is based on the *EcoRI* restriction fragment patterns from duplicate single-colony isolates for each clone. Clonal rearrangements can be detected as differences in the duplicate fingerprints resulting from clonal heterogeneity. Rearrangements can also be detected by comparing different clones from the same contig for consistent fragment patterns. PACs from the RPCI-21 library were prepared from *MboI* partially digested DNA. Consequently, the *EcoRI* fingerprints of overlapping clones will always have different vector-insert junction fragments. The two BAC libraries were both constructed using *EcoRI* partially digested DNA. Hence, the BAC clones will never generate *EcoRI* vector-insert junction fragments. Three small rearrangements were detected in the contigs as heterogeneity between duplicate subcolonies (Fig. 4). Four additional fingerprint inconsistencies were found within single clones by comparing them with all their corresponding overlapping clones. These inconsistencies are probably not caused by genetic polymorphisms because all of the overlapping BACs and PACs constructed from the same inbred strains were consistent. Three of the four fingerprint inconsistencies affected only one or two fragments per clone. However, one PAC clone (RPCI-21 221A10) contained five unique restriction fragments not found in any of the overlapping clones. This 200-kb PAC clone was determined previously to be chimeric based on the STS-content mapping. The estimated accumulated sizes of the unique fragments are 10–15 kb; hence, the unrelated genomic segments in this clone are quite dissimilar in size: 185 and 15 kb. In summary, 7 clones

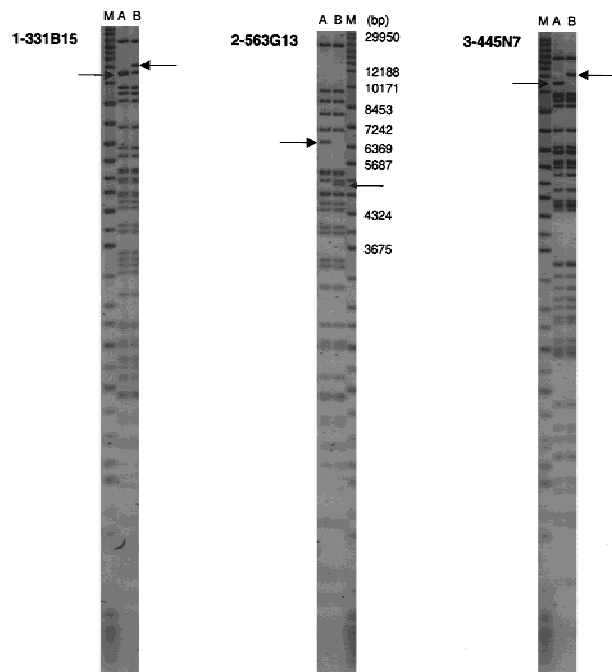


Figure 4 The clone heterogeneity from fingerprinting gel images. (M) Marker, A and B are derived from different single colonies from the same clone. (*center*) Some sizes of marker bands are indicated. These fingerprints show a deletion that took place during replication. Arrows indicate inconsistent bands.

out of 113 were found to have rearrangements: one due to chimerism and the others due to alterations to a single genomic fragment during or after the cloning process.

DISCUSSION

Mammalian model organisms such as mice provide the ability to identify, manipulate, and eventually understand genes related to homologous genes in human. Comparative analysis of genome structure and gene function between mouse and human facilitates the identification and understanding of human pathways affected in diseases. High-quality BAC and PAC clone resources consisting of large genomic fragments are a prerequisite to genome analysis in mice and the eventual sequencing of the mouse genome. In addition, individual clones often contain complete genes embedded in their genomic milieu. The clones can thus be used for functional studies in cell lines or in transgenic applications (Yang et al. 1997; Probst et al. 1998). The ability to reintroduce these clones back into mammalian cells also offers a splendid opportunity to study the effects of modified genes. Recent developments in site-directed mutagenesis of BAC clones permit the efficient modification inside *Escherichia coli* (Yang et al. 1997; Zhang et al. 1998). In view of all these applications of BACs and PACs, it is very important to have BAC libraries with minimal levels of cloning artifacts, a

high degree of clone stability, and a fairly unbiased representation of the entire genome. The BAC and PAC cloning approaches have been developed originally by Shizuya et al. (1992) and Ioannou et al. (1994). During the recent years, improvements in vectors and cloning procedures allow the preparation of better clone libraries at increased efficiencies (Kim et al. 1996; Strong et al. 1997; Osoegawa et al. 1998; Frengen et al. 1999). The three libraries described here were prepared during a time of improvements, which are reflected in the increased insert size and narrow insert size distribution seen in the most recent BAC library prepared from the C57BL/6J mouse strain. BAC clones in this RPCI-23 library have average inserts of ~200 kb, with 89% of the BACs between 160 kb and 250 kb (Table 1; Fig. 1C). The libraries have been prepared from two strains: the 129 inbred strain (129S6/SvEvTAC) widely used for transgenic applications (Simpson et al. 1997) and the standard laboratory mouse (C57BL/6J) that is more commonly used for genome and genetic analysis. In view of the widespread interest in the scientific community to use these clone collections, we analyzed the libraries to determine the level of cloning artifacts, for example, chimeric clones, the stability of the clones, and the randomness of the cloning process. Chimeric clone levels in genomic libraries are usually determined by using the clones as probes for FISH analysis on metaphase-spread chromosomes. Multiple cytogenetic map locations identified by a particular probe are an indication of chimerism within the probe. However, such an analysis only provides clear answers if the chimeric cloning levels are disturbingly high and obvious, and well above the level of duplications in the genome. At artifact levels <10%, it is not feasible to use FISH to ascertain the precise level of chimeric clones. To circumvent this problem, a more sensitive and laborious assay was designed based on the mapping of markers derived from BAC or PAC insert ends. Thus, we sequenced the insert terminals for all BACs and PACs included in three very redundant contigs. Not all end sequences are suitable for the design of unique markers, and only 59 of the 113 end-sequenced clones could be used to create markers for both ends. In addition, we were able to determine that another 12 clones were also nonchimeric because of the presence of identical sequence ends shared between multiple clones in the contig. Surprisingly, the end sequencing also revealed a bias in the cloning process for the libraries constructed from *EcoRI* partially digested DNA. In one particular contig of 28 BACs, six clones shared one identical end and two clones shared both ends. These clones were derived from different transformations and therefore represent independent cloning events. Although an obvious cloning bias was demonstrated with respect to a single *EcoRI* site in one of the three contigs, no *EcoRI* cloning bias was found in the other

two contigs. It is conceivable that this *EcoRI* recognition site is preferentially cut. However, these "hot sites" do not block chromosome walking because other BACs have been identified spanning the particular hot site (Table 2). Because these clones were part of the contig, it was possible to map the end markers. All but one of the end sequences detected confirmed contig clones. One of the end-derived markers did not identify the overlapping clones and is derived from a 200-kb chimeric PAC clone consisting of two genomic fragments of ~185 and 15 kb. Other cloning artifacts can result from rearrangements within a single genomic fragment during the initial cloning steps or during propagation resulting in clonal heterogeneity. Initial rearrangements should become apparent upon careful comparison of clones with overlapping clones by restriction fingerprinting. Inconsistent bands in the fingerprint of one particular clone that are lacking in all other contig clones indicate possible rearrangements. Four out of 113 clones analyzed demonstrated inconsistent patterns. Three of these four clones had apparently rearranged during the cloning process, whereas the fourth was the 200-kb chimeric clone, possibly resulting from coligation of two fragments. An additional 3 of the 113 clones demonstrated clonal heterogeneity as evident from minor differences in the fingerprints of duplicate subcolonies (Fig. 4).

In summary, it is clear that the BAC/PAC chimeric clones occur very infrequently at levels ~1%. Other clonal rearrangements occur at low levels well below 10%. In addition, the data indicates a good genomic representation with 28 markers detecting positive clones in each of the three libraries. These RPCI BAC and PAC libraries should be useful resources for mouse genome mapping, sequencing projects, and functional analysis.

METHODS

Markers for Electrophoresis

Low Range PFG Marker (New England Biolabs) containing a mixture of λ DNA-*HindIII* fragments and λ concatemers was used as marker DNA for pulsed-field gel electrophoresis. λ DNA-*HindIII* and 1-kb ladder fragments (Life Technologies) was used for agarose gel electrophoresis.

Media and Plates

LB medium (1% Bacto-tryptone, 0.5% yeast extract, 0.5% NaCl) containing 25 μ g/ml kanamycin or 20 μ g/ml chloramphenicol was used to grow PAC or BAC clones, respectively. LB medium containing 7.5% glycerol and appropriate antibiotics was used to store the clones in 384-well plates at -80°C . LB/agar (1.5% Bacto-agar, 5% sucrose, and antibiotics) plates were used for colony picking, and LB/agarose (1.5%; Ultrapure, Life Technologies) plates were used for preparing high-density replica filters. SOC medium (Life Technologies; 2% Bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl_2 , 10 mM MgSO_4 , 20 mM glucose) was used to

induce the expression of the antibiotic resistant gene after electroporation.

BAC/PAC Library Construction

The pPAC4 (E. Frengen, B. Zhao, S. Howe, D. Weichenhan, K. Osoegawa, E. Gjernes, J. Jessee, H. Prydz, C. Huxley, and P.J. de Jong, unpubl.) and pBACe3.6 vectors (Frenge et al. 1999) were used for construction of PAC and BAC libraries, respectively. DNA plugs generated from the spleens of 5-week-old 129S6/SvEvTac female mice (Taconic) were partially digested with *MboI* for PAC library construction. Two BAC libraries were generated using the same DNA plugs used for the construction of the PAC library as well as DNA plugs isolated from the kidneys and brains of 5-week-old C57BL/6J female mice (Jackson Laboratory). DNA digested with *MboI* or a combination of *EcoRI* and *EcoRI* methylase was size-fractionated in a clamped homogeneous electrical field (CHEF; Chu et al. 1986) apparatus (Bio-Rad) and recovered by electroelution prior to ligation with *BamHI* sites in the pPAC4 or *EcoRI* sites in the pBACe3.6, respectively. The ligated DNA was transformed into *E. coli* DH10B cells (Life Technologies) via electroporation procedure. The transformed cells were incubated in SOC medium shaking at 200 rpm for 1 hr at 37°C . The cell suspension containing 10% glycerol was frozen in ethanol dry ice bath and stored at -80°C prior to colony picking. A detailed protocol for constructing PAC and BAC libraries has been described previously (Osoegawa et al. 1998).

Arraying Recombinant Clones

Prior to picking the recombinant colonies, the frozen cell suspension was thawed on ice and spread with sterile glass beads onto LB/agar plates containing sucrose and antibiotics in 22×22 -cm tray to an optimum density of 1500 colonies per plates based on the estimated titer. The bacterial colonies were picked into 384-well plates containing LB medium with glycerol and antibiotics using an automatic colony-picking machine (Q-bot; GENETIX). The 384-well plates were covered with plastic wrap and incubated at 37°C for 20 hr. The empty wells were checked manually the following day, and the plates were stored at -80°C .

Insert Size Analysis of PAC and BAC DNA

Randomly picked PAC or BAC clones were grown overnight at 37°C in a shaking incubator, followed by DNA extraction using an automated plasmid isolation machine (AutoGen 740, Integrated Separation Systems). DNA (0.5–1 μ g) was dissolved in 100 μ l of TE (pH 8.0), and ~50 ng was digested with 0.1 units of *NotI* (New England Biolabs) to separate the vector and insert DNA fragments. The insert sizes were determined using CHEF or field-inversion gel electrophoresis (FIGE; Bio-Rad) analysis under the following conditions: 1% agarose gel (Ultrapure, Life Technologies) in $0.5 \times$ TBE buffer, at 14°C , 6 V/cm for 16 hr, with a 0.1- to 40-sec pulse time at a 120° angle for CHEF; and 1% agarose gel in $0.5 \times$ TBE buffer, at room temperature, 180 V (forward voltage), 120 V (reverse voltage) for 16 hr, with a 0.1- to 14-sec switch time linear shape for FIGE. The gels were stained with ethidium bromide, and DNA was visualized using an Alpha Innotech IS1000 digital imager.

Preparing High-Density Replica Filters

The PAC or BAC clones were gridded onto 22×22 -cm nylon filters (0.45- μ m pore size; Schleicher & Schuell) that were placed on LB/agarose plates containing antibiotics. The grid-

ding was accomplished using an automatic colony gridding machine (*BioGrid*, *BioRobotics*). Each filter contains 36,864 colonies that represent 18,432 (48×384) independent clones spotted in duplicate in a 4×4 clone array. Colonies were lined up into 16×24 small sections, where each section represents a 4×4 dot array of 16 colonies or duplicates of eight individual clones in specific vector patterns. The gridded colonies were incubated at 37°C overnight and processed. Each filter was placed on a paper (Whatman 3MM) soaked in alkaline solution (0.5 M NaOH, 1.5 M NaCl) for 4 min at room temperature, followed by 4 min at 80°C, then settled in neutralization solution (0.5 M Tris-HCl at pH 7.5, 1.5 M NaCl) for 4 min at room temperature. After drying the filter on a paper, the filter was soaked in a 200-ml solution containing 50 mM Tris-HCl (pH 8.5), 50 mM EDTA, 100 mM NaCl, 1% *N*-lauroyl sarcosine, and 0.5 mg/ml Pronase (Boehringer Mannheim) and incubated at 37°C for 2 hr with periodic shaking. The filter was dried; DNA was then cross-linked to the filter (Bio-Rad). The percentage of nonrecombinant clones containing the pUC-stuffer fragment was determined by hybridization of the gridded filters with pPAC4 vector for RPCI-21 and pBACe3.6 vector for RPCI-22 and RPCI-23 as a probe.

Screening the Libraries Using Various Probes

Twenty-two CATS (Lyons et al. 1997) primer sets were amplified by PCR using mouse genomic DNA as template. The primer sets used in this work, kindly provided by Dr. Lyons, are as follows: aminolevulinate, δ -, synthase 1 (*ALAS2*), amylase, α -1A; salivary (*AMY1A*), biglycan (*BGN*), catalase (*CAT*), ceruloplasmin (ferroxidase) (*CP*), cytochrome P-450, subfamily XIX (*CYP19*), dopamine receptor (*DRD2*), factor 9 (*F9*), fibronectin 1 (*FNI*), glyceraldehyde-3-phosphate dehydrogenase (*GAPD*), glucosidase, β ; acid (*GBA*), glucose-regulated protein (78 kD) (*GRP78*), insulin-like growth factor binding protein 1 (*IGFBP1*), immunoglobulin heavy chain gene cluster (*IgH*), lactate dehydrogenase A (*LDHA*), mineralcorticoid (aldosterone) receptor (*MLR*), Moloney murine sarcoma oncogene homolog (*MOS*), myeloperoxidase (*MPO*), neurofibromin 1 (*NFI*), paired box homeotic gene (*PAX8*), avian reticuloendotheliosis viral oncogene homolog (*REL*), and Thy-1 cell surface antigen (*THY1*). The sequences of these primer sets were described previously (Lyons et al. 1997). The location of each marker is obtained from Lyons et al. (1997) or from the Mouse Genome Database (MGD) 3.1, Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine (<http://www.informatics.jax.org/>). The PCR reaction was performed using 10 ng of mouse genomic DNA as template with 0.15 μ M each primer, 250 μ M dNTPs, $1 \times$ PCR buffer, and 0.4 units *Taq* DNA polymerase (Boehringer Mannheim) in a 10- μ l reaction volume. The amplification was performed at 94°C for 1 min, then 94°C for 15 sec, 56°C for 15 sec, and 72°C for 30 sec for 5 cycles, followed by 94°C for 15 sec, 54°C for 15 sec, and 72°C for 30 sec for 33 cycles, and finally 72°C for 1 min in the GeneAmp PCR system 9600 (Perkin-Elmer Cetus). The PCR products were separated in a 1% agarose gel and identified by ethidium bromide staining. DNA was excised from the gels, eluted by repeating freezing and thawing three times. The DNA was extracted with 50–100 μ l of TE (pH 8.0) and diluted 100 times with sterile distilled deionized water. The PCR products were reamplified and labeled with [α -³²P]dCTP (3000 Ci/mmol; Amersham). The reaction was performed with 0.15 μ M each primer, 50 μ M dATP, 50 μ M dGTP, 50 μ M dTTP, 20 μ M dCTP, 10 μ Ci [α -³²P]dCTP, $1 \times$ PCR buffer, and 0.4 units of *Taq* DNA polymerase in a 10- μ l reaction volume. Following

the reaction, 50 μ l of TE buffer (pH 8.0) was added into the mixture, and the solution was mixed and transferred into 1.5-ml microtubes. The PCR products were confirmed by exposure on an X-ray film after agarose gel electrophoresis. One microliter of the solution was spotted on two DE81 papers (Whatman), and one filter was washed with 0.5 M sodium phosphate buffer (pH 7.0). The incorporation of ³²P was estimated using liquid scintillation apparatus (Beckman) comparing the count of the washed filter with that of the unwashed filter. The labeled probes were pooled and added to hybridization solution containing 0.25 M sodium phosphate buffer (pH 7.0), 7% SDS, and 1 mM EDTA. Filters (5–7) were placed in a 7.5-cm-diam. bottle with nylon mesh placed between the filters in 30-ml hybridization buffer. Hybridization was performed at 65°C overnight in a rotating DNA oven. The washing was carried out with $0.25 \times$ SSC–0.1% SDS solution five times at 65°C. The filters were enclosed in a plastic bag (PE FLAT BAGS, Fisherbrand) and exposed to X-ray film (Kodak) overnight at –80°C. Positive clones were assembled into a 96- and 384-well plate sublibrary containing medium with antibiotics and spotted onto nylon membranes for rescreening with individual probes. The sequence of each PCR product using a positive clone as template was determined and analyzed using the BLASTN homology search program (Altschul et al. 1990). The RPCI-22 and RPCI-23 libraries were similarly screened. In addition to these CATS probes, a hypoxanthine phosphoribosyltransferase (*HPRT*) gene exon amplified by PCR, six overlapping oligonucleotide (overgo) probes, which were designed from PAC or BAC ends, and two overgo probes designed from Moloney murine leukemia virus insertion sites in mammary tumors (*EN21* and *EN31*) were applied to screen the three libraries with the double screening procedure as described above. The overgo probing technology and a designing script have been developed at the Genome Sequencing Center in Washington University, St Louis (McPherson 1999). The six overgo probes derived from PAC or BAC ends were localized to three contig ends as described below (Fig. 3). The probes are summarized in Table 2. DNA from positive clones was isolated, and ~100 ng (18 μ l) of DNA was treated with RNase A (final concentration, 10 μ g/ml) and 1 unit of *Eco*RI restriction enzyme (New England Biolabs) together in a 20- μ l volume. The reactions were performed in flexible 96-well plates at 37°C for 2 hr. The digested DNA was separated with 0.7% agarose gel (25 \times 30-cm gel tray size) in $0.5 \times$ TBE buffer at room temperature. The gel was stained with ethidium bromide, and the image was captured by the Alpha Innotech IS1000 digital imager (Fig. 2A). The separated DNA was transferred onto nylon membranes (Hybond N+, Amersham) using the Southern blotting alkaline transfer method (Sambrook et al. 1989). The Southern hybridization was performed to confirm the positive clones using either the PCR product or clone itself. Positive clones were also confirmed with PCR.

End Sequencing and Designing End Probes

Single colonies were obtained from positive clones for *CAT*, *IGFBP1*, and *MLR* loci. Five colonies from each clone were picked, into 96-well plates, and analyzed with a CHEF or FIGE apparatus using the same conditions as described above. All clone ends were sequenced with the use of 5'-GCCGCTAATACGACTCACTATAGGGAGAG-3' or 5'-TAATACGACTCACTATAGGG-3' for the T7 end and 5'-GTTTTTGGGATCTGCCGTTTC-3' for the SP6 end. The direct BAC and PAC end sequence procedure, developed at The Institute for Genomic Research, was modified as follows: Cell

suspension (R1), lysis (R2), and neutralization (R3) buffers were purchased from Qiagen. Three microliters of single colony culture fluid was inoculated in three 96-deep-well blocks (Beckman) containing 1.6 ml of LB medium and appropriate antibiotics into each well and incubated for 20 hr at 37°C in a shaker at 325 rpm. Bacterial cells were pooled from the three blocks into a single block. Pellets were suspended in 300 μ l of buffer R1 with RNaseA (0.2 mg/ml) and RNaseT1 (100 U/ml). After the addition of 300 μ l of lysis buffer R2, the samples were mixed gently and allowed to stand at room temperature for 5 min. The R3 buffer (300 μ l) was added; the sample was merged gently and placed on ice for 5 min. Fifty microliters of ProCipitate reagent (Ligochem, Inc.) was added, and samples were moderately mixed several times during a 5-min incubation at room temperature. The lysates were transferred to a Qiagen Turbofilter plate (Qiagen) positioned over a 2.0-ml deep well collection block inside a vacuum manifold, and then, the filtrates were collected. Equal volumes of isopropanol were added, mixed well, and held at -20°C for 1 hr. DNA was collected by centrifugation at 4°C for 1 hr at 1900g. After decanting the supernatant, the pellet was washed with 70% ethanol and spun at 1900g for 10 min. The ethanol was removed, and the pellet was dried prior to resuspension in 53 μ l of 1 mM Tris-HCl buffer (pH 8.0). Cycle sequencing reactions were carried out in a 96-well format using the dRhodamine Terminator Cycle Sequencing chemistry (Perkin-Elmer Applied Biosystems). Eight microliters of template DNA was mixed with 6 μ l of dRhodamine premix and 1 μ l of 10 pmoles/ μ l primer to produce a 1.25 \times reaction with 15 μ l of final volume. An initial 96°C incubation is followed by 100 cycles of 96°C for 10 sec, 50°C for 10 sec, 60°C for 4 min, as performed at the University of Oklahoma, Advanced Center for Genome Technology. Unincorporated dye terminators were removed from the reaction products by gel exclusion chromatography. Each well of a Prepsep 96 deep-well spin filter block (Lida, Inc.) was fitted with a 20- μ m polyethylene frit (Lida, Inc.) and loaded with 1.5 ml of a 50 mg/ml slurry of Sephadex G50F (DNA grade; Pharmacia). After removing excess water by centrifugation, the reaction products were transferred to the Sephadex containing block. The block was spun, and the eluted samples were collected in a 96-well tray. The samples were dried and resuspended in 1–1.5 μ l of formamide loading buffer that was prepared by combining one part of an aqueous solution containing 25 mM EDTA and 50 mg/ml Blue Dextran 2000 (Pharmacia) with five parts of deionized formamide (Amresco). The suspended samples were denatured at 96°C for 3 min, placed on ice immediately, and loaded on the sequencing gel of a Perkin-Elmer Applied Biosystems 377 DNA Sequencer. Primer sets for overgo probes have been designed using the script after masking repetitive sequences utilizing the RepeatMasker (A.F.A. Smit and P. Green, unpubl.; <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Clone Rearrangements

Three independent contigs have been assembled from *CAT*, *IGFBP1*, and *MLR* regions using end sequence-based overgo probing hybridization to assess clone fidelity. The filters prepared from the sublibraries derived from each library were used for the hybridization. Hybridization was performed using the same buffer as described above at 60°C for the overgo probes. Washes were performed three times in 0.5 \times SSC–0.1% SDS solution at 60°C. The contigs were assembled by SEGMAP based on the hybridization data. Southern hybrid-

ization was performed with PAC and BAC DNA isolated from positive clones of each library. The *Eco*RI digested DNA was separated on 0.7% agarose gel in 0.5 \times TBE buffer (Fig. 2A). The separated DNA was transferred onto nylon membranes and hybridized using the probes as described above. In addition to the Southern hybridization analysis, the fingerprinting analysis was applied to identify clone rearrangements. The detailed protocol was described previously (Marra et al. 1997). In brief, BAC or PAC DNA was isolated using the automatic plasmid isolation machine (AutoGen 740) from 2 ml of LB medium containing appropriate antibiotics. The DNA was dissolved in 100 μ l of TE buffer. Five microliters of DNA was digested in a 10- μ l reaction containing 50 ng of RNaseA (Sigma), 5 units *Eco*RI and 1 \times *Eco*RI reaction buffer (New England Biolabs) at 37°C for at least 2 hr. When the digestion was completed, 2 μ l of 6 \times loading buffer [15% Ficoll (Sigma), 0.25% bromophenol blue, 0.25% xylene cyanol FF] was added. The *Eco*RI-digested DNA (1.75 μ l of 12 μ l) was run on a 1% agarose gel (Seakem LE, FMC Bio Products) in 1 \times TAE buffer at 90 V for 15 min, then at 45 V for 12 hr. The electrophoresis was performed in Horizon 20-25 electrophoresis tanks (Life Technologies) at 16°C using a VWR recirculator. Marker contains 12.5 ng/ μ l Analytical Marker DNA, Wide Range Ladder (Promega), and 2 ng/ μ l Marker V (Boehringer Mannheim). The marker was loaded in every fifth lane. After electrophoresis, the gel was stained in a Rubbermaid container containing 500 ml of a 1 : 10,000 dilution of SYBR Green I (FMC Bio Products) in 1 \times TAE for 20–30 min. The gel was scanned using FluorImager 595 (Molecular Dynamics). The image was transferred to a SUN workstation. The gel image was analyzed using a program called Image 3.9d (Sulston et al. 1988, 1989). When the image analysis of the gels was completed, the image files were archived on the computer, and an FPC project [FPC v.4.3 (Gregory et al. 1996; Soderlund et al. 1997, 1998)] was automatically created. The fingerprints were analyzed to identify clonal rearrangements using FPC v.4.3.

ACKNOWLEDGMENTS

Current information on mouse genomic PAC and BAC libraries is available through our World Wide Web home page located at <http://bacpac.med.buffalo.edu>. We thank Dr. Leslie Lyons for providing the CATS primers, Dr. John D. McPherson for instructing the overgo probing techniques, Dr. Marco A. Marra for instructing the fingerprinting techniques, and Dr. Norma J. Nowak for reviewing the manuscript. We are grateful to Amy F. Beck, Alfred Cairo, Gregory P. Caldwell, Melanie Hierl, David W. Kois, Yukiko Osoegawa, Beth A. Palka, Chung Li Shu, Barbara Swiatkiewicz, and Baohui Zhao for their technical assistance. K.O. has been supported with a Postdoctoral Fellowship for Research Abroad from the Japan Society for the Promotion of Science (JSPS). This work was funded by grants from the National Institutes of Health (1R01HL55700), the National Human Genome Research Institute (1R01RG01165), and the U.S. Department of Energy (DE-FG02-94ER61883).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altshul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman.

1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Chu, G., D. Vollrath, and R.W. Davis. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric field. *Science* **234**: 1582–1585.
- Dunham, I., K. Dewar, U.-J. Kim, and M.T. Ross. 1999. In *Genome analysis: A laboratory manual*, vol. 3 (eds. B. Birren, E.D. Green, P. Hieter, S. Klapholz, R.M. Myers, H. Riethman, and J. Roskams), pp. 3–6. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Frengen, E., D. Weichenhan, B. Zhao, K. Osoegawa, M. van Geel, and P.J. de Jong. 1999. A modular positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics* **58**: 250–253.
- Green, E.D., H.C. Riethman, J.E. Dutchik, and M.V. Olson. 1991. Detection and characterization of chimeric yeast artificial chromosome clones. *Genomics* **11**: 658–669.
- Gregory, S., C. Soderlund, and A. Coulson. 1996. Contig assembly by fingerprinting. In *Genome mapping: A practical approach* (ed. P.H. Dear), pp. 227–254. Oxford University Press, Oxford.
- Gregory, S.G., G.R. Howell, and D.R. Bentley. 1997. Genome mapping by fluorescent fingerprinting. *Genome Res.* **7**: 1162–1168.
- Haldi, M., V. Perrot, M. Saumier, T. Desai, D. Cohen, D. Cherif, D. Ward, and E.S. Lander. 1994. Large human YACs constructed in a rad52 strain show a reduced rate of chimerism. *Genomics* **24**: 478–484.
- Ioannou, P.A., C.T. Amemiya, J. Garnes, P.M. Kroisel, H. Shizuya, C. Chen, M.A. Batzer, and P.J. de Jong. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**: 84–89.
- Kim, U.-J., B.W. Birren, T. Slepak, V. Mancino, C. Boysen, H.-L. Kang, M.I. Simon, and H. Shizuya. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**: 213–218.
- Korenberg, J.R., X.N. Chen, K.L. Devon, D. Noya, M.L. Oster-Granite, and B.W. Birren. 1999. Mouse molecular cytogenetic resource: 157 BACs link the chromosomal and genetic maps. *Genome Res.* **9**: 514–523.
- Lyons, L.A., T.F. Laughlin, N.G. Copeland, N.A. Jenkins, J.E. Womack, and S.J. O'Brien. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* **15**: 47–56.
- Marra, M.A., T.A. Kucaba, N.L. Dietrich, E.D. Green, B. Brownstein, R.K. Wilson, K.M. McDonald, L.W. Hillier, J.D. McPherson, and R.H. Waterston. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**: 1072–1084.
- McPherson, J.D. 1999. In *Genome analysis: A laboratory manual*, vol. 4 (eds. B. Birren, E.D. Green, P. Hieter, S. Klapholz, R.M. Myers, H. Riethman, and J. Roskams), pp. 207–213. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Osoegawa, K., P.-Y. Woon, B. Zhao, E. Frengen, M. Tateno, J.J. Catanese, and P.J. de Jong. 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* **52**: 1–8.
- Probst, F.J., R.A. Fridell, Y. Raphael, T.L. Saunders, A. Wang, Y. Liang, R.J. Morell, J.W. Touchman, R.H. Lyons, K. Noben-Trauth, T.B. Friedman, and S.A. Camper. 1998. Correction of deafness in shaker-2 mice by an unconventional myosin in a BAC transgene. *Science* **280**: 1444–1447.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* **89**: 8794–8797.
- Simpson, E.M., C.C. Linder, E.E. Sargent, M.T. Davisson, L.E. Mobraaten, and J.J. Sharp. 1997. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat. Genet.* **16**: 19–27.
- Smith, D.J., M.E. Stevens, S.P. Sudanagunta, R.T. Bronson, M. Makhinson, A.M. Watanabe, T.J. O'Dell, J. Fung, H.-U.G. Weier, J.-F. Cheng, and E.M. Rubin. 1997. Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates *minibrain* in learning defects associated with Down syndrome. *Nat. Genet.* **16**: 28–36.
- Soderlund, C., I. Longden, and R. Mott. 1997. FPC: A system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13**(5): 523–535.
- Soderlund, C., I. Longden, and I. Dunham. 1997. Sequence ready clones. In *Guide to human genome computing* (ed. M.J. Bishop), pp. 151–179. Academic Press.
- Strong, S.J., Y. Ohta, G.W. Litman, and C.T. Amemiya. 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. *Nucleic Acids Res.* **25**: 3959–3961.
- Sulston, J., F. Mallett, R. Staden, R. Durbin, T. Horsnell, and A. Coulson. 1988. Software for genome mapping by fingerprinting techniques. *Comput. Appl. Biosci.* **4**(1): 125–132.
- Sulston, J., F. Mallett, R. Durbin, and T. Horsnell. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* **5**(2): 101–106.
- Venter, J.C., H.O. Smith, and L. Hood. 1996. A new strategy for genome sequencing. *Nature* **381**: 364–366.
- Woon, P.-Y., K. Osoegawa, P.J. Kaisaki, B. Zhao, J.J. Catanese, D. Gauguier, R. Cox, E.R. Levy, G.M. Lathrop, A.P. Monaco, and P.J. de Jong. 1998. Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosomal library. *Genomics* **50**: 306–316.
- Wynshaw-Boris, A. 1996. Model mice and human disease. *Nat. Genet.* **13**: 259–260.
- Yang, X.W., P. Model, and N. Heintz. 1997. Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotech.* **15**: 859–865.
- Zhang, Y., F. Buchholz, J.P. Muyrers, and A.F. Stewart. 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**: 123–128.

Received August 24, 1999; accepted in revised form November 16, 1999.