



## Advanced Integrated Mouse YAC Map Including BAC Framework

Leonard C. Schalkwyk, Brian Cusack, Ilona Dunkel, et al.

*Genome Res.* 2001 11: 2142-2150

Access the most recent version at doi:[10.1101/gr.176201](https://doi.org/10.1101/gr.176201)

---

### References

This article cites 31 articles, 8 of which can be accessed free at:  
<http://genome.cshlp.org/content/11/12/2142.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

# Advanced Integrated Mouse YAC Map Including BAC Framework

Leonard C. Schalkwyk,<sup>1,2</sup> Brian Cusack, Ilona Dunkel, Martina Hopp, Markus Kramer, Stefanie Palczewski, Jutta Piefke, Sabine Scheel, Michael Weiher, Gunther Wenske, Hans Lehrach, and Heinz Himmelbauer<sup>2</sup>

*Max-Planck-Institute of Molecular Genetics, Ihnestr  e 73, D-14195 Berlin, Germany*

Functional characterization of the mouse genome requires the availability of a comprehensive physical map to obtain molecular access to chromosomal regions of interest. Positional cloning remains a crucial way of linking phenotype with particular genes. A key step and frequent stumbling block in positional cloning is making a contig of a genetically defined candidate region. The most efficient first step is isolating YAC (Yeast Artificial Chromosome) clones. A robust, detailed YAC contig map is thus an important tool. Employing Interspersed Repetitive Sequence (IRS)-PCR genomics, we have generated an advanced second-generation YAC contig map of the mouse genome that doubles both the depth of clones and the density of markers available. In addition to the primarily YAC-based map, we located 1942 BAC (Bacterial Artificial Chromosome) clones. This allows us to present for the first time a dense framework of BACs spanning the genome of the mouse, which, for instance, can serve as a nucleus for genomic sequencing. Four large-insert mouse YAC libraries from three different strains are included in our data, and our analysis incorporates the data of Hunter et al. and Nusbaum et al. There is a total of 20,205 markers on the final map, 12,033 from our own data, and a total of 56,093 YACs, of which 44,401 are positive for more than one marker.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. BHI74059–BHI75013.]

The human genome sequence will provide us with a wealth of information including a catalog of human genes. Neither the sequence by itself nor *in vitro* studies (including expression profiles and proteomics), however, give direct access to previously unknown functions. Mutations, whether spontaneous or induced, are the geneticists' most important tool to link phenotypic effects to underlying genes and functional interactions. Positional cloning of human disease genes and mouse mutations has given us access to completely novel genes. In general, these have been rare, drastic, single-locus traits. Many more gene functions will become accessible through positional cloning of the genes behind more common but more subtle variation in many quantitative traits. Genes and their respective functions have mainly been conserved between mammals, such that model organisms can be employed for studies of gene function (Andersson et al. 1996; M  klos and Rubin 1996; Nadeau and Sankoff 1998; Denny and Justice 2000). For multiple reasons, the mouse is the most suitable model organism for functional studies. First, mice are small mammals with a short generation time and relatively low maintenance costs; second, a large number of different inbred strains exist that differ with respect to particular phenotypical aspects (Beck et al. 2000); and third, spontaneous and induced mutagenesis has produced a fund of thousands

of mutations (Bedell et al. 1997; Justice et al. 1999; Denny and Justice 2000). This provides us with a huge body of different entry points to understand in molecular terms what causes observed phenotypic differences. Of particular importance with respect to identifying causative genes is the availability of a clone-based physical map of the mouse genome, which is as complete as possible. Once genetic mapping has located candidate regions on the map, clones within these intervals can be used for the identification of candidate genes for mouse genomic sequencing and functional studies including the generation of transgenic animals.

We here present an advanced global physical map of the mouse genome. Our map contains 12,033 newly developed IRS-PCR markers, mostly YAC- and BAC-derived, and integrates the currently available physical mapping data in mouse (Hunter et al. 1996; Nusbaum et al. 1999). Sequence data is presented for a subset of IRS markers. Our map provides access to 56,093 YAC clones from four different YAC libraries originating from three mouse strains. Because 20% of the probes to assemble our map were derived from BAC clones, we also present a first global framework map of the mouse genome in BACs that can serve as a nucleus for mouse genome sequencing.

## RESULTS

### Generation of 12,033 Novel IRS-PCR Markers for Mouse Genome Mapping

The interspersed repetitive sequence (IRS)-PCR technology is based on the abundance of repeat elements in the genome of higher organisms. Repetitive sequence primers are used to amplify sequences that are flanked by repeat elements. For

<sup>1</sup>Present address: Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK.

<sup>2</sup>Corresponding authors.

E-MAIL [himmelbauer@molgen.mpg.de](mailto:himmelbauer@molgen.mpg.de); FAX 49-30-8413 1128.

E-MAIL [L.Schalkwyk@iop.kcl.ac.uk](mailto:L.Schalkwyk@iop.kcl.ac.uk); FAX 44-020-7848 0801

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.176201>.

instance, a single primer to a portion of the B1 repeat will amplify thousands of individual fragments from a mouse genomic DNA template. IRS-PCR can be used on any genomic DNA containing sample, that is, total genomic DNA, DNA from cell hybrids, from individual clones, or clone pools. Complex IRS-PCR reaction products can be cloned into plasmids. An alternative, which we used extensively, is IRS-PCR on YAC or BAC clones. Fragments generated on these low-complexity templates can directly be employed as markers. The generation of large numbers of IRS markers in this way is cheap, because there is no requirement to sequence markers or to design locus-specific primers. For this work, IRS-PCR probe fragments were generated with a single B1 repeat-derived primer from the following sources (Table 1): 10,620 IRS probes were amplified from random YAC clones of library WHTy917, also targeting the portion of the library for which microsatellite data were published (Nusbaum et al. 1999). Two thousand eight hundred eighty-nine IRS probes were generated from BAC clones. One thousand nine hundred twenty-five random IRS probes were from libraries that we prepared by cloning IRS-PCR product mixtures from C57BL/6 DNA or from somatic cell hybrid 167EJ (mouse chromosomes 17 and 3 on human background). The complete set of probes is publicly available. Sequence data for 955 markers have been deposited in GenBank.

### Streamlined Hybridization Procedures for High-Throughput Data Production on YAC Pools

Mapping reagents such as genetic crosses, radiation hybrid panels, and genomic libraries can be typed with IRS-PCR markers by hybridization. The target DNA (complex IRS-PCR products generated from the mapping reagent) is arrayed at high density onto a membrane and probed with one labeled marker fragment at a time. Each of the 15,434 IRS probes that we generated was hybridized against filters containing IRS-PCR products generated from YAC pools, representing the four available large insert mouse YAC libraries constructed at the ICRF and the Whitehead Institute (Table 2). Our pooling strategy (see Methods) allowed this high-representation genomic clone collection to be spotted on  $7 \times 11$ -cm reusable nylon filters, and >3000 such filters were produced. Hybridizations were carried out in sets of 96. Probes were labeled in microtiter plates and subsequently hybridized against YAC pool filters in parallel, with one probe per filter. Up to 576 probes could be tested per week, and a complete dataset of 15,434 IRS probes hybridized against YAC pool filters was generated in 18 months.

### Contig Building Using a Simulated Annealing Algorithm

Fifteen thousand four hundred thirty-four hybridization results were evaluated from autoradiograms, checked twice, and then stored in a database. Data entry was accomplished using a Java applet, which generated the image of a virtual filter in a Web browser. On this virtual filter, positive hybridization signals could be entered by clicking. Scoring errors were minimized by the design of the hybridization filters. The pattern in particular made it easy to recognize complete addresses (corresponding signals in three dimensions of the pooling system). The hybridization results were deconvoluted (addresses of positive YAC clones extracted from lists of positive YAC pools) and exported from the database for contig analysis with the software tool *wprobeorder* (Mott et al. 1993; Grigoriev et al. 1998). This program calculates distances between probes on the basis of positive target clones in common and orders probes using a simulated annealing algorithm. Once the optimal probe path is found and the path length of the probe order is minimized, the clones are fitted to the probe framework. However, even with simulated annealing, global analysis of the data began to reach the limits of our computing system when the probe number exceeded 10,000. We modified the *wprobeorder* program to use a compressed representation of the data internally, and eventually moved to a system of partitioning the dataset by chromosomes.

Our analysis used the data of Nusbaum et al. (1999) to provide links to the mouse genetic map. A starting point was the global analysis with *wprobeorder* of the first 9945 probes of our dataset together with the Nusbaum et al. data. This produced contigs that could be assigned to chromosomes on the basis of markers of known position that they contained. The chromosome-specific contig collections formed multiple anchor points for incremental data addition. Further probes were assigned to chromosomes using a program we developed called *increment*, which added new data to probe lists based on clones in common. The result was a set of probe lists, one for each chromosome, containing all probes that might be relevant to that chromosome (Table 3, column a). These crude probe lists contained many probes linked to the chromosome by chimeric YACs and hybridization-inherent noise. These probes could, to a large extent, be excluded on the basis of stringent contig-membership criteria, again using *wprobeorder*: *wprobeorder* declares a break wherever a probe has no YAC in common with the next. A break is, for example, introduced by a probe positive for only one of two clones spanning its neighbors. The data was thus broken into small groups of tightly linked probes, and those that could be

**Table 1. Markers on the Integrated MPI-MG YAC Map**

Marker designation	Marker type	No. of markers on map	Probe source	Reference
mbacr	IRS-PCR marker	2215	CITB mouse BAC clones	this study
bir	IRS-PCR marker	1247	IRS-PCR library from genomic C57BL/6 DNA	this study
173r	IRS-PCR marker	176	IRS-PCR library from cell hybrid 167EJ	this study
whtII	IRS-PCR marker	8395	WHTy917 mouse YAC clones	this study
D*Mit*	SSLP-STS	4314	MIT/WICGR	Dietrich et al. 1996
Various	STS	3534	MIT/WICGR	Nusbaum et al. 1999
HUN	IRS-PCR marker	324	K. Hunter	Hunter et al. 1996

**Table 2.** Large-Insert Mouse YAC Libraries

Library	No. of clones	Genome coverage	Average insert size	Strain of origin	Reference
ICRFy902	13,400	2x	700 kb	C3H, male	Larin et al. 1993
ICRFy903	5000	0.8x	720 kb	B10, female	Larin et al. 1993
WHTy910	20,000	4.5x	680 kb	B6, female	Kusumi et al. 1993
WHTy917	40,800	13x	820 kb	B6, female	Haldi et al. 1996

B10, C57BL/10; B6, C57BL/6.

mapped to a chromosome were retained in the list for that chromosome (Table 3, column b). To produce maps from these cleaned probe lists that could be easily integrated with other data, we used the Nusbaum et al. (1999) maps as a framework (Table 3, column c), to which we added our probe lists that included Nusbaum et al. and Hunter et al. data recursively in three rounds with the `increment` program (Table 3, columns d-f). This resulted in marker lists for each chromosome roughly double the size of those of Nusbaum et al. We then recovered IRS probes that were so far unassigned to any of the probe lists. To avoid incorrect chromosomal assignments caused by a small number of YACs positive for probes from many chromosomes, we ignored YACs positive with more than 15 probes (Table 3, column g). Finally, among the still-unassigned probes there remained a small number that created links between adjacent contigs of the Nusbaum et al. map. Because such links are valuable and will arise by chance only infrequently, they were assigned to the relevant chromosomes (Table 3, column h). The result of this process, before map editing, was the chromosomal assignment of 23,333 markers (data from all sources), 19,024 uniquely, the rest to two or more different chromosomes.

The map assembly to this point was automated, and the maps at this stage could be considered rough drafts. The final two stages of the analysis were map refinement and reconciliation with genetic data, which were done semimanually. Refinement on the integrated map had to take into consideration that the Nusbaum et al. and Hunter et al. datasets were based on a much smaller number of YACs. In addition, fewer positive YACs were obtained with each probe than would be expected from the genome coverage of the libraries used (average of 8.6 YACs per probe in our dataset). This was observed in each dataset, and is largely due to the use of a pooling strategy. If the stringent contig definition of `wprobeorder` described earlier is used, this would lead to high numbers of breaks. We therefore chose, by inspection, breakpoints between groups of `wprobeorder` contigs. These groups (map segments) were then individually ordered with `wprobeorder` and further edited by hand. The editing had the objectives of removing incorrectly assigned probes that did not belong to the chromosome, and to establish the locally optimal path of probes (see Fig. 1 for proximal MMU12, and for all other chromosomes see [www.molgen.mpg.de/~rodent/result/genome.html](http://www.molgen.mpg.de/~rodent/result/genome.html)). A summary of the progress of automated contig assembly, marker content of chromosomes, and map segment numbers is presented in Table 3. The final edited map consists of 595 map segments and contains 21,647 markers representing 20,205 unique markers (19,024 assigned once). Twelve thousand thirty-three of these markers are novel IRS-PCR markers generated and mapped by us.

## Integration of Mouse Physical and Genetic Map

The genomewide contig maps produced were compared to the 1999 mouse chromosome committee genetic maps. Chromosome committee maps were used because they are a synthesis of all available data, and

the value of the physical maps is maximized by being straightforwardly aligned with the comprehensive marker and gene lists that they provide. This integration depends on the committee members choosing positions for markers in a way that deals with the systematic differences between distances produced by different genetic crosses. The chromosome committee maps could be quite well reconciled with our data for most chromosomes, requiring only a small number of mismapped markers to be ignored and a few segments to be resorted per chromosome. Exceptions are chromosomes 2 and 15, where physically closely linked markers had widely differing genetic placements. For these two chromosomes we used data from the EUCIB cross (Breen et al. 1994; Rhodes et al. 1998).

## A First-Generation Global BAC Framework Map of the Mouse Genome

Two thousand two hundred fifteen IRS probes (from 1942 different BACs) were derived from random mouse BACs of clones from the CITB library. This mapping information can therefore be exploited to assemble a first BAC framework map of the mouse genome (Fig. 2) as a resource for further studies, both structural and functional.

## DISCUSSION

We here present an advanced mouse physical map that contains 21,904 markers, 12,033 of which represent new IRS-PCR markers generated and mapped by us. Sequence data are available for 955 markers. The chromosome maps that we present are integrated with the genome-wide contig data of the two previous major mouse mapping efforts by Hunter et al. and Nusbaum et al. Using a large number of common reference points, the integrated map can be compared to any of the existing maps of the mouse genome (genetic and RH maps), as well as to the consensus maps drafted by the various chromosome committees. The complete set of mouse contig data is available on our Web site at <http://www.molgen.mpg.de/~rodent>. This Web site can be used to access the data that we have generated, that is, integrated YAC contigs and BAC framework maps, and DNA sequences of markers.

The >50,000 YAC clones placed on the map represent more than 12-fold genome coverage. The map we present must thus be an essentially complete representation of the YAC cloneable portion of the genome. As discussed below, a minor portion of the genome is not accessed by our method because it is poor in B1 repeat elements. A quantitative assessment of coverage is difficult, but the majority of apparent gaps must be either small or negative in size (i.e., undetected overlaps). Integration of the mouse physical and genetic maps gives a minimum estimate of map coverage. The lengths of gaps between map segments were estimated in cM and com-

**Table 3.** Progress of Automated Contig Assembly, Marker Content, and Map Segments of MPI-MG Integrated Mouse YAC Map

Chr	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	Total (edited)	mbacr <sup>a</sup>	whtII <sup>a</sup>	bir <sup>a</sup>	173r <sup>a</sup>	Map segments <sup>b</sup>
MMU1	7396	1855	635	1153	1335	1340	1544	1550	1457	128	616	43	3	43
MMU2	7548	1835	600	1179	1377	1396	1640	1661	1565	157	685	77	1	52
MMU3	7943	1548	398	891	1150	1183	1369	1375	1277	132	616	29	72	34
MMU4	6243	1381	400	826	1026	1055	1257	1266	1153	148	515	64	0	39
MMU5	6676	1920	422	969	1095	1105	1347	1355	1278	151	538	129	0	36
MMU6	6295	1150	471	782	853	865	996	1009	964	77	380	23	0	30
MMU7	7252	1369	416	829	1007	1025	1235	1250	1142	128	498	64	0	30
MMU8	6789	1132	376	705	838	862	994	1009	919	110	340	64	0	29
MMU9	6542	1509	398	937	1221	1239	1465	1486	1300	158	547	144	0	27
MMU10	7814	1081	320	613	762	785	938	944	879	101	400	40	0	35
MMU11	7749	1997	518	1248	1506	1523	1755	1770	1626	206	674	174	1	24
MMU12	7810	1631	367	918	1211	1230	1469	1486	1311	212	615	87	0	27
MMU13	6797	1379	425	871	1040	1043	1184	1199	1155	136	507	60	0	28
MMU14	6714	1395	323	840	1063	1081	1298	1333	1172	152	569	81	0	30
MMU15	6400	1079	332	678	788	809	925	938	916	88	426	46	0	29
MMU16	4846	914	318	622	692	693	776	776	755	99	286	37	0	23
MMU17	5831	954	198	507	724	738	810	819	814	73	361	46	118	19
MMU18	7006	1001	270	530	699	708	875	885	841	90	410	33	1	18
MMU19	3861	535	158	315	395	402	451	460	444	56	190	18	0	15
MMUX	4204	767	241	553	645	646	740	750	683	37	307	82	0	27

Columns a–h give the number of markers for each chromosome at a particular stage of the analysis. Column stage (a) crude probe list; (b) sorted contigs; (c) Nusbaum et al. framework; (d) increment, first round; (e) increment, second round; (f) increment, third round; (g) addition of unassigned probes; (h) addition of linking probes. Total, marker content on final edited maps (sum of markers from MPI-MG, Hunter et al. [1996] and Nusbaum et al. [1999]).

<sup>a</sup>Number of MPI-MG markers (see Table 1) on edited chromosome maps.

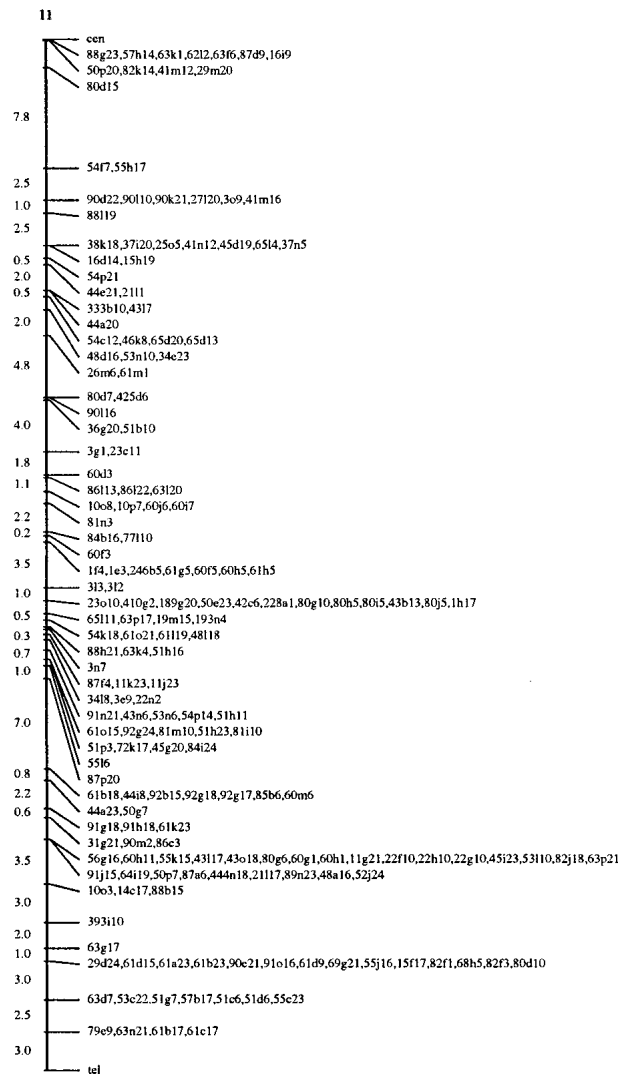
<sup>b</sup>See <http://www.molgen.mpg.de/~rodent/result/genome.html>

pared to the genetic length of chromosomes. For example, we have calculated that 79%, 83%, 75%, 76%, and 75% for mouse chromosomes 11, 12, 13, 14, and 18, respectively, are between the outermost genetically mapped markers of segments. These are low estimates, because map segments often extend beyond genetically mapped markers in both directions. Of the random 15,434 IRS markers that we have produced, we have placed 12,033 markers (78%) on the map. Nusbaum et al. estimated their map to cover 92% over the mouse genome based on STS screening. In contrast to our work, Nusbaum et al. include unanchored contigs in their calculations for map coverage. More comparable to the above figures is that they reported that 66% of STSs could be mapped into anchored contigs on the basis of YAC data alone (two independent clones in the same contig hit by one STS). Combining YAC and RH data, the portion of STSs that Nusbaum et al. could map to anchored contigs was at ~80%. Our map integrates the Nusbaum et al. dataset with our own data and the data of Hunter et al. It is, therefore, clear that the map provided by us is more comprehensive. This is demonstrated by the large number of links that we provide between hitherto separate clone contigs in the previously published datasets.

Our IRS-PCR markers are widely distributed across the genome. We have probes linked to the majority of MIT markers (SSLP and STS). Of 9877 MIT markers, 9036 (91.5%) have at least one YAC in common with one of our markers. We also have over 3000 markers that could not be linked to the microsatellite framework. Some of these must represent regions of the genome, which are multiple YAC lengths from the nearest microsatellite marker. There are also regions poor in IRS-PCR markers: this was, for instance, observed on chromosome 16 between markers *D16Mit5* and *D16Mit203* (cM 38–54). We believe these are genuinely B1-poor regions. The B1-

based primer B1R that we devised from sequence alignments matches the consensus of all six B1 subfamilies B1-A to B1-F. This, plus optimized reaction conditions, provides a maximum of product species and is particularly useful for amplification from low complexity templates, for example, YAC-DNA pools and individual genomic clones. Previous work by Boyle et al. (1990) has demonstrated that the most abundant dispersed repeat elements in the mouse genome are differentially distributed: although L1 elements are preferentially located within R bands, both B1 and B2 repeats are concentrated within G bands. Of these three repeat classes, B1 is the most frequent. Nevertheless, B1-FISH produced a less pronounced chromosomal banding pattern than observed with either B2 or L1 probes. This cannot be explained by sequence variation within the B1 repeat family, as a high degree of similarity has been reported between members of the individual B1 repeat subfamilies (Zietkiewicz and Labuda 1996). Technical reasons aside, this implies that the distribution of B1 repeats is less biased when compared to either L1 or B2.

We have generated a map framework of almost 2000 BACs that represents the largest collection of BACs so far mapped to high resolution. Assuming an average insert size of 150 kb per clone and no overlap between clones, this collection of BAC clones has a cumulative DNA content of 330 Mb, corresponding to >10% of the mouse genome. Because our clones were isolated from a segment of the library comprising only a little more than one genome equivalent, clones may be closely located, but will mostly be nonoverlapping (i.e., not derived from the same locus). This adds particular value to the framework we present. We propose to include these BACs for mouse genome sequencing for the following reasons: first, each clone is mapped to high resolution on the mouse physical map. Second, the DNA source for the CITB library is the



**Figure 1** BAC framework map for mouse chromosome 11 (MMU11). Shown are 189 BAC clones that we have incorporated into our physical map of MMU11 employing high-throughput IRS-PCR-based hybridization assays. For map display, microsatellite markers integrated from the Nusbaum et al. dataset were used as reference points to calculate positions of BACs on the genetic map in centimorgans. Higher map resolution within bins (several BACs that map into an interval defined by a pair of microsatellite markers) can be deduced from our contig data (see our public Web site at <http://www.molgen.mpg.de/~rodent/result/genome.html>).

129/Sv ES cell line CJ7. Genomic clones of a strain background identical to an ES cell line are superior targets for the generation of knock-out constructs. Third, the sequence from mapped clones will provide an excellent basis to generate SNP markers when compared to the C57BL/6 genotype, which has been selected for global genomic sequencing.

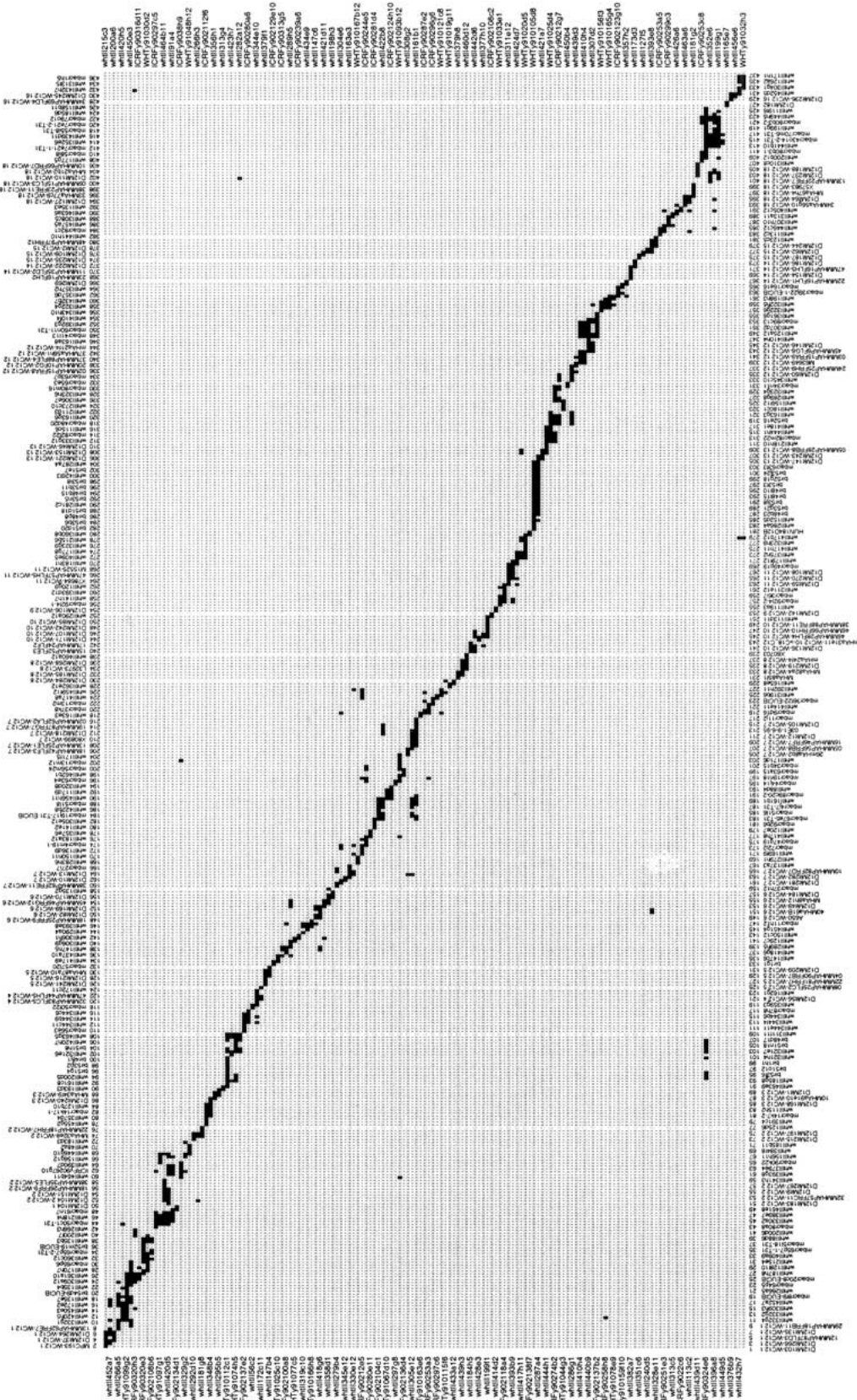
All our newly generated markers as well as the clones on our maps are in the public domain. A single mouse IRS-PCR primer was used to produce all of our IRS-PCR markers. If a researcher wants to obtain a particular marker fragment, the following options exist: first, a subset of probes is available as cloned fragments (sequence partly available). These probe fragments can be ordered from RZPD (<http://www.rzpd.de>).

Second, for IRS-PCR fragments not available as plasmid clones, a genomic clone (BAC, YAC) can be ordered from one of several Genome Centers and used to set up an IRS-PCR reaction, for example, with colony material. IRS-PCR is easy to perform and works reliably, and is particularly easy to do on low-complexity BAC and YAC templates (crude YAC-DNA preparations, colony material). Fragment isolation can be done with high throughput (96-well and 384-well formats). A PCR reaction with the IRS primer on a genomic clone template is sufficient to regenerate the marker fragment in any lab. Strictly speaking, cloning of IRS-PCR fragments is not necessary, because the markers can easily be regenerated by amplification from a particular genomic clone. In the case of probes derived from BAC clones, relatively few IRS products are produced. Different IRS products from the same BAC invariably give similar results when used as probes. This is not necessarily the case for YAC clones where the size of the clone and the probability of chimerism are both higher. In the case where it is important that a probe is recreated, the authors can supply size information or the physical fragment.

A physical map is a prerequisite for map-based (positional) cloning of genes. In particular, the sequence of the mouse genome will remain incomplete for years to come, and physical maps as presented by us will remain crucial instruments to proceed from a phenotype to the gene. We have compiled a list of 422 mouse phenotypes and traits listed in MGD (Blake et al. 2000) for which the causative genes are currently unknown, and assigned them to intervals on the map, or into gaps, respectively. This allowed us to identify candidate intervals for 80% of traits (data not shown; available on request).

As mouse genomic sequencing is still at an early stage, the map information that we provide will be enormously useful for contig construction efforts, gap closure, and gap size estimation on the emerging mouse genome working draft. BACs or other easily used *Escherichia coli*-based clones are a preferred material for sequencing or other studies. It is, therefore, worth emphasizing that the IRS-PCR technology is a convenient and fast tool to convert YAC contigs to BAC maps. Any researcher who wants to obtain a BAC contig located within a region of interest will first select YACs and BACs from our Web site, order them from a Genome Center, perform IRS-PCR on the clones, and hybridize IRS-PCR fragments against filters from the CITB library (Korenberg et al. 1999) or the larger insert (200 kb average) RPCI23 BAC library (Osoegawa et al. 2000). Because our maps are well integrated with other maps of the mouse genome, researchers can, before starting experimental work, inspect online resources, for example, <http://www.nih.gov/science/models/mouse/resources/index.html>, to obtain information regarding current contig building and sequencing activities within their intervals of interest. The Genome Sequence Center, Vancouver ([http://www.bcgsc.bc.ca/projects/mouse\\_mapping](http://www.bcgsc.bc.ca/projects/mouse_mapping)), is preparing global mouse BAC contigs by a restriction fingerprint method that robustly identifies large overlaps between clones. This will greatly facilitate BAC sequencing. Because of the size of the clones and the characteristics of the method, that work will not produce long-range contigs, nor does it directly produce links to the genetic or other physical maps. For some chromosomes (3, 4, 8, 10, 16, 18, X) numerous links to the genetic map have been produced by typing of BACs with microsatellite markers ([genomics.roswellpark.org/mouse/overview.html](http://genomics.roswellpark.org/mouse/overview.html)). This is not a density of anchors sufficient to make a continuous map, but does provide an initial

Physical map of mouse chromosome 12, interval cM 1.0 - cM 22.0



**Figure 2** YAC contig map of proximal mouse chromosome 12 (MMU12). YAC data shown correspond to the minimal number of YACs required to span the probe path as established using the *wpx:obex:dex* program. Marker origin is identified in Table 1. Additional information available for markers: T31, marker placed on mouse RH panel; EUC1B, marker genetically mapped on interspecific mouse backcross panel; WC12\*, marker placed on the physical map of Nusbaum et al. (1999). Map segments (breaks identified by blank vertical lines) are displayed sequentially with respect to their order along the chromosome. This order was deduced from genetically mapped markers within segments and their position on the MMU12 chromosome committee map (D'Eustachio and Riblet 1999).

framework. Other chromosomes have scarcely any links to the Vancouver BAC data; for example, chromosome 17 has only 24 (146 BACs). The BACs mapped in this work are thus a significant step ahead in starting points for BAC contig assembly.

## METHODS

### YAC and BAC Libraries, Cell Lines

Four large-insert YAC libraries available for the mouse were used to maximize both genomic coverage and integration of existing data (Table 2). ICRF YAC libraries ICRFy902 and ICRFy903 (Larin et al. 1993) were obtained from RZPD (<http://www.rzpd.de>). YAC libraries WHTy910 (Kusumi et al. 1993) and WHTy917 (Haldi et al. 1996) (Whitehead I and II) and the CITB 129/Sv mouse BAC library (Korenberg et al. 1999) were purchased from Research Genetics (<http://www.resgen.com>). Cell hybrid 167/EJ contains mouse chromosomes 17 and 3 on a human background (Pickford 1989).

### IRS-PCR

IRS-PCR was performed with a single mouse B1 element-derived 28mer primer (B1R; 5'-AGTTCAGGACAGC CAGGGCTAYACAGA-3'). This primer was the most successful (as judged by the number of products obtained from an arbitrary panel of YACs) of a series that was designed based on an alignment of 153 mouse B1 sequences obtained by a GenBank FASTA search. The position of the primer on the sequence was selected by adding 3' bases sequentially to a starting sequence and counting the number of exact matches in the sequence set at each step. The sequence contains one two-fold degeneracy at position 23 and overlaps primer B1MvsCH (Hunter et al. 1994). Reactions were carried out in 1 × PCR buffer (35 mM Tris-Base/15 mM Tris-HCl/0.1% Tween 20/50 mM KCl/125 μM of each dNTP/15 μM cresol red) and 1 μg (for genomic DNA templates) or 0.2 μg (BAC, YAC templates) of primer B1R. Cycling conditions were initial denaturation (94°C, 3 min) followed by (94°C, 30 s; 65°C, 30 s; 72°C, 180 s) × 35 and final incubation at 72°C, 10 min, using a 30-μL reaction volume in a PTC100 thermocycler (MJ instruments).

### Construction of IRS-PCR Fragment Plasmid Libraries and Insert Preparation

Genomic DNA (100 ng) from strain C57BL/6J (library 57R/bir) and cell hybrid line 167EJ were amplified with primer (CUA)<sub>4</sub>-B1R under conditions described above. Products were size-selected by gel electrophoresis. Low-melting point agarose gel slices containing fragments larger than approximately 500 bp were equilibrated in 10 mM Tris-HCl pH 7.6/30 mM NaCl/0.75 mM spermidine/30 mM spermine, molten at 68°C, and treated with 150 U/mL agarase (Sigma) for 2 h. The DNA was precipitated with isopropanol, treated with Uracil DNA Glycosylase, annealed with the tailed pAMP10 vector, and transformed into *E. coli* DH5α following the manufacturer's instructions (Gibco-BRL). Recombinant clones were picked into 384-well microtiter dishes using a custom-built picking robot (Maier et al. 1997). Inserts from clones for labeling were amplified in 1 × PCR buffer with 3 pmol of each vector primer 5/86 (5'-GCACGCGTACGTAAGCTTGGATCCTCTAG-3') and 3/86 (5'-CCGGTCCGGAATCCCCGGGT-3'). Amplification was carried out in 384-well polypropylene Q plates (Genetix Ltd.) in a waterbath thermocycling robot (Maier et al. 1997).

### Preparation of Template DNA from YAC Clones

Ninety-six-well plates containing 200 μL of uracil-free SD medium (2% glucose/0.7% yeast nitrogen base without ammonium sulfate/1.4% casamino acids/200 μg/mL tryptophan/100 μg/mL adenine hemisulphate/55 μg/ml tyrosine) were

inoculated from library plates with 96-pin plastic replicators and grown to saturation (48 h at 30°C) without shaking. DNA was prepared from the wells using a previously described protocol (Chumakov et al. 1992). DNA from these crude preparations was transferred into 96-well thermofast PCR plates (Advanced Biotechnologies) containing 30 μL of PCR mix and amplified as described above. Five microliters of each PCR product was electrophoresed on 1.2% low melting-point agarose and one to two individual products were harvested. The approximately 50 μL agarose fragments excised from the agarose gels were supplemented with 30 μL of water and melted; 12 μL aliquots of this mixture were prepared (also in thermofast plates) for labeling.

### Isolation of IRS-PCR Fragments from BAC Clones

Clones were stamped onto agar plates using a disposable multipin spotting device and grown overnight at 37°C. PCR reactions were inoculated from colony material in a 96-well format as described (Himmelbauer et al. 1998).

### IRS-PCR on YAC Clone Pools and Pool Filter Production

All libraries were pooled in similar three-dimensional (3D) schemes based on blocks of 8 × 96-well microtiter dishes (768 clones), each producing eight row pools, 12 column pools, and eight plate pools. All but the WHTy917 library were additionally pooled in a further three dimensions to introduce some redundancy and improve recovery of positive clones. This was achieved by reshuffling the plate stacks so that each plate had a completely different set of neighbors in the new stack. This produced new row and column pools, but the plate pools would remain the same. To produce partly reshuffled plate pools, partial plates were constructed containing 32 clones from each of two plates. The sources of the pooled DNAs were as follows: for ICRF libraries (ICRFy902, 903), six-dimensional pools were prepared by us. For library WHTy910, 3D pools were purchased from Research Genetics, and a further three dimensions were prepared by us as described above. For library WHTy917, 3D pools were obtained from the HGMP (<http://www.hgmp.ac.uk>). The clones were grown separately in 96-well plates as described above, and pools were assembled using a vacuum pump and manifold, which allowed a row or column to be harvested into a 50-mL centrifuge tube with a modified, two-port lid. The apparatus was extensively rinsed with distilled water between vacuuming operations, and the suspensions were processed into agarose blocks of 200 μL volume, containing 0.75% low melting point agarose, 1 M sorbitol, 1 mM EDTA, 100 mM sodium citrate pH 5.6, which were cast in filter-bottom microtiter plates (Millipore). The blocks were incubated at room temperature for an hour and then washed in NDS (0.5 M EDTA pH 8.0, 1 mM EDTA, 1% Sarkosyl, 100 μg/mL proteinase K) at 55°C for 48 h, in 0.5 M EDTA pH 8.0, and repeatedly in 1 × TE (1 mM EDTA, 10 mM Tris-HCl pH 8.0) at 4°C. The blocks were finally melted at 65°C with 1 mL of water. YAC pool DNAs were further diluted in TE (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA) so that 5- μL aliquots gave comparable PCR yields between libraries and aliquots prepared in 96-well thermofast PCR plates. Amplification was as described above, but in a quadruple recipe (120 μL volume). Each plate was checked for PCR product quality by agarose electrophoresis of 12 samples. Forty-five-microliter aliquots were transferred to 384-well microtiter plates for spotting. To compensate for rapid evaporation at the edges of the plate, edge wells were supplemented with 5 μL and corner wells with 10 μL of water. PCR products were spotted in a standard 5 × 5 interleave, duplicate spotting pattern with pins having a 250-μm diameter flat tips on Hybond N+ membranes (Amersham) using a robot from Linear Drives. Each position was spotted three

times, and 54 individual  $6.7 \times 11$ -cm filters were produced in each run lasting approximately 4 h. Each set of spotting plates was used for six spotting runs, so the volume of PCR product per spot was up to 0.07  $\mu$ L. During spotting the membranes were on top of two layers of filter paper (Whatman 3MM) saturated with 0.4 M NaOH. After spotting the membranes were sequentially washed with 0.4 M NaOH, 0.5 M sodium phosphate, pH 7.2, and 50 mM sodium phosphate pH 7.2. Membranes were UV treated on the DNA side using a Stratlinker (Stratagene) with an energy dose of 1200  $\mu$ J.

### Probe Preparation and Hybridization

Probes were labeled with 5  $\mu$ Ci of  $\alpha^{32}$ PdCTP by random priming (Feinberg and Vogelstein 1983). Reactions were incubated at room temperature overnight, diluted with one volume of  $1 \times$  TE, heat-denatured in a PCR machine for 10 min at 94°C and subsequently added to the YAC pool filters in hybridization buffer (Church and Gilbert 1984). Hybridizations were carried out in batches of 96 at a time, with filters rolled up and hybridized in 15-mL plastic tubes in a water bath set to 65°C for 16 h. Filters were washed at 65°C for 60 min each in  $2 \times$  SSC/0.1% SDS and  $0.1 \times$  SSC/0.1% SDS. Exposure against X-ray film was carried out for up to 1 wk at room temperature.

### Data Entry and Management

Autoradiograms were marked, checked for errors, and subsequently entered directly into a database. The database was implemented with Oracle 8.1.5, and the input tools largely with the PL/SQL cartridge of the Oracle application server. HTML-forms were used to record the textual experimental details (filter number, hybridization date, data quality assessment). A Java applet using JDBC allowed hybridization results to be entered (or edited) graphically by clicking on positions in a grid representing the spotting pattern. Positive filter coordinates were converted into clone names by the server using a deconvolution routine written in awk, which assigned a weight of 3 to complete, unambiguous positive clones, 2 to complete, ambiguous addresses (i.e., where there is more than one positive clone in a block of eight plates), and 1 to incomplete addresses (e.g., row and column, but no plate).

### Analysis

The analysis integrated published YAC (Hunter et al. 1996; Nusbaum et al. 1999) and genetic data (Breen et al. 1994; Rhodes et al. 1998; Blake et al. 2000), and was based on probe ordering by simulated annealing using *wprobeorder* (Mott et al. 1993; Grigoriev et al. 1998). Positions of IRS probes relative to the genetic map were determined experimentally on the EUCIB interspecific mouse backcross panel (Breen et al. 1994; Rhodes et al. 1998) and on the T31 mouse RH panel (McCarthy et al. 1997; Van Etten et al. 1999) according to published procedures (Himmelbauer et al. 1998, 2000). Additionally, the chromosomal localization of IRS probes was inferred on the basis of our hybridization results on YAC clones contained within contigs of Nusbaum et al. We distinguished three different cases: “rule1,” an IRS probe hybridized with at least two different YACs from a single Nusbaum et al. contig (strong positional information). “rule2,” an IRS probe hybridized directly to a YAC mapped into a Nusbaum et al. contig. In addition, further YACs identified by this probe were hit by different IRS probes generated from YACs of the same Nusbaum et al. contig (medium strong positional information). “rule3,” an IRS probe1 hybridized to several unmapped YAC clones. However, at least two IRS probes derived from different YAC clones from a single Nusbaum et al. contig hybridized against more than one YAC detected by IRS probe1 (weak positional information). For all rules1–3 the term “different clones and probes” implied that respective YACs had to be

derived from different library plates to minimize the risk of erroneous assignments. A similar concept was used to identify “linking probes” that join adjacent contigs in the Nusbaum et al. dataset. The analysis is discussed further in the Results section.

### Access to Raw Data, Clones, and Sequences

IRS-PCR libraries 57R/bir and 173R have been deposited at RZPD (<http://www.rzpd.de>). For BAC- and YAC-derived probes, markers can be regenerated with IRS-PCR on the respective clone with primer B1R. All large insert libraries are available from multiple genome centres, including RZPD. Raw data as published on our chromosome maps (<http://www.molgen.mpg.de/~rodent/result/genome.html>) are available for downloads. This set of data contains only YACs hit by more than one marker per chromosome. The full dataset that includes singly-hit YACs is available upon request from the authors. Sequence data are available at our Web site in FASTA format, and have also been deposited in GenBank.

### ACKNOWLEDGMENTS

We are grateful to L. Dang, M. Aurin, D. Grünwitzky, and M. Scheidig for data entry; K. Büsow for scripts useful in analyzing EUCIB data; D. Buczek for modifications on *wprobeorder*, R. Mühlhaus for help with Java; and M. Kirby and T. Kreidler for data management. We gratefully acknowledge the dedicated help of the MPI sequencing team (headed by Dr. R. Reinhardt), M. Clark, S. Keil, and T. Crnogorac-Jurcevic provided excellent help in early stages of the project. We thank Drs. M.-L. Yaspo and M. Burmeister for discussions and for critically reading the manuscript. This work was supported by the Max-Planck-Gesellschaft.

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

- Andersson, L., Archibald, A., Ashburner, M., Audun, S., Barendse, W., Bitgood, J., Bottema, C., Broad, T., Brown, S., Burt, D., et al. 1996. Comparative genome organization of vertebrates. *Mamm. Genome* **7**: 717–734.
- Beck, J.A., et al. 2000. Genealogies of mouse inbred strains. *Nat. Genet.* **24**: 23–25.
- Bedell, M.A., Largaespada, D.A., Jenkins, N.A., and Copeland, N.G. 1997. Mouse models of human disease. Part II: Recent progress and future directions. *Genes & Dev.* **11**: 11–43.
- Blake, J.A., Eppig, J.T., Richardson, J.E., and Davisson, M.T. 2000. The Mouse Genome Database (MGD): Expanding genetic and genomic resources for the laboratory mouse. *Nucleic Acids Res.* **28**: 108–111.
- Boyle, A.L., Ballard, S.G., and Ward, D.C. 1990. Differential distribution of long and short interspersed element sequences in the mouse genome: Chromosome karyotyping by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci.* **87**: 7757–7761.
- Breen, M., et al. 1994. Towards high resolution maps of the mouse and human genomes—A facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**: 621–627.
- Chumakov, I.M., et al. 1992. Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nat. Genet.* **1**: 222–225.
- Church, G.M. and Gilbert, W. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81**: 1991–1995.
- Denny, P. and Justice, M.J. 2000. Mouse as the measure of man? *Trends Genet.* **16**: 283–287.
- D'Eustachio, P. and Riblet, R. 1999. Mouse chromosome 12. *Mamm. Genome* **10**: 953.
- Dietrich W.F., et al. 1996. A comprehensive genetic map of the mouse genome. *Nature* **380**: 149–151.
- Feinberg, A.P. and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6–13.
- Grigoriev, A., Levin, A., and Lehrach, H. 1998. A distributed

- environment for physical map construction. *Bioinformatics* **14**: 252–258.
- Haldi, M.L., et al. 1996. A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mamm. Genome* **7**: 767–769.
- Himmelbauer, H., Wedemeyer, N., Haaf, T., Wanker, E.E., Schalkwyk, L.C., and Lehrach, H. 1998. IRS-PCR-based genetic mapping of the huntingtin interacting protein gene (HIP1) on mouse chromosome 5. *Mamm. Genome* **9**: 26–31.
- Himmelbauer, H., Schalkwyk, L.C. and Lehrach, H. 2000. Interspersed repetitive sequence (IRS)-PCR for typing of whole genome radiation hybrid panels. *Nucleic Acids Res.* **28**: e7.
- Hunter, K.W., et al. 1994. Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): Generation of a 5-cM, > 5 megabase contig on mouse chromosome 1. *Mamm. Genome* **5**: 597–607.
- Hunter, K.W., et al. 1996. Toward the construction of integrated physical and genetic maps of the mouse genome using interspersed repetitive sequence PCR (IRS-PCR) genomics. *Genome Res.* **6**: 290–299.
- Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B., and Bradley, A. 1999. Mouse ENU mutagenesis. *Hum. Mol. Genet.* **8**: 1955–1963.
- Korenberg, J.R., Chen, X.N., Devon, K.L., Noya, D., Oster-Granite, M.L., and Birren, B.W. 1999. Mouse molecular cytogenetic resource: 157 BACs link the chromosomal and genetic maps. *Genome Res.* **9**: 514–523.
- Kusumi, K., Smith, J.S., Segre, J.A., Koos, D.S., and Lander, E.S. 1993. Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome* **4**: 391–392.
- Larin, Z., Monaco, A.P., Meier-Ewert, S., and Lehrach, H. 1993. Construction and characterization of yeast artificial chromosome libraries from the mouse genome. *Methods Enzymol.* **225**: 623–637.
- Maier, E., Bancroft, D.R., and Lehrach H. 1997. Large-scale library characterization. In *Automation technologies for genome characterization* (ed. T.J. Beugelsdijk), pp. 65–88. John Wiley and Sons Inc., New York.
- McCarthy, L.C., et al. 1997. A first generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res.* **7**: 1153–1161.
- Miklos, G.L. and Rubin, G.M. 1996. The role of the genome project in determining gene function: Insights from model organisms. *Cell* **86**: 521–529.
- Mott, R., Grigoriev, A., Maier, E., Hoheisel, J., and Lehrach, H. 1993. Algorithms and software tools for ordering clone libraries: Application to the mapping of the genome of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **21**: 1965–1974.
- Nadeau, J.H. and Sankoff, D. 1998. Counting on comparative maps. *Trends Genet.* **14**: 495–501.
- Nusbaum, C., et al. 1999. A YAC-based physical map of the mouse genome. *Nat. Genet.* **22**: 388–393.
- Osoegawa, K., et al. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**: 116–128.
- Pickford, I. 1989. Ph.D thesis. Imperial Cancer Research Fund, London, UK.
- Rhodes, M., et al. 1998. A high-resolution microsatellite map of the mouse genome. *Genome Res.* **8**: 531–542.
- Van Etten, W.J., et al. 1999. Radiation hybrid map of the mouse genome. *Nat. Genet.* **22**: 384–387.
- Zietkiewicz, E. and Labuda, D. 1996. Mosaic evolution of rodent B1 elements. *J. Mol. Evol.* **42**: 66–72.

Received December 18, 2000; accepted in revised form September 27, 2001.