



1-Mb Resolution Array-Based Comparative Genomic Hybridization Using a BAC Clone Set Optimized for Cancer Gene Analysis

Joel Greshock, Tara L. Naylor, Adam Margolin, et al.

Genome Res. 2004 14: 179-187

Access the most recent version at doi:[10.1101/gr.1847304](https://doi.org/10.1101/gr.1847304)

References

This article cites 10 articles, 6 of which can be accessed free at:
<http://genome.cshlp.org/content/14/1/179.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

1-Mb Resolution Array-Based Comparative Genomic Hybridization Using a BAC Clone Set Optimized for Cancer Gene Analysis

Joel Greshock,¹ Tara L. Naylor,¹ Adam Margolin,¹ Sharon Diskin,¹
 Stephen H. Cleaver,² P. Andrew Futreal,³ Pieter J. deJong,⁴ Shaying Zhao,⁵
 Michael Liebman,¹ and Barbara L. Weber^{1,6}

¹Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; ²Georgia State University, Atlanta, Georgia 30303, USA; ³Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK; ⁴Children's Hospital Oakland-BACPAC Resources, Oakland, California 94609, USA; ⁵The Institute for Genomic Research, Rockville, Maryland 20850, USA

Array-based comparative genomic hybridization (aCGH) is a recently developed tool for genome-wide determination of DNA copy number alterations. This technology has tremendous potential for disease-gene discovery in cancer and developmental disorders as well as numerous other applications. However, widespread utilization of aCGH has been limited by the lack of well characterized, high-resolution clone sets optimized for consistent performance in aCGH assays and specifically designed analytic software. We have assembled a set of ~4100 publicly available human bacterial artificial chromosome (BAC) clones evenly spaced at ~1-Mb resolution across the genome, which includes direct coverage of ~400 known cancer genes. This aCGH-optimized clone set was compiled from five existing sets, experimentally refined, and supplemented for higher resolution and enhancing mapping capabilities. This clone set is associated with a public online resource containing detailed clone mapping data, protocols for the construction and use of arrays, and a suite of analytical software tools designed specifically for aCGH analysis. These resources should greatly facilitate the use of aCGH in gene discovery.

[Supplemental material is available online at www.genome.org.]

Initial studies revealed the enormous potential of genomic copy number profiling using array-based comparative genomic hybridization (aCGH) as a tool for identifying candidate disease genes and cancer classification (Pinkel et al. 1998; Hodgson et al. 2001; Fritz et al. 2002; Pollack et al. 2002). However, a high-resolution collection of human clones with sequence-based mapping annotation optimized for aCGH use has not been publicly available. We selected bacterial artificial chromosome (BAC) clones to develop this clone set for CGH array construction. BAC clones are thought to be superior to cDNAs and oligonucleotides as aCGH probes because their large genomic size (~150 Kb) confers high and consistent binding specificity (Pinkel et al. 1998; Hodgson et al. 2001), and thus more accurate copy number determination. The large BAC clone collection recently extensively characterized by the BAC Resource Consortium (Cheung et al. 2001), although a very valuable resource, was not optimized for even spacing, aCGH hybridization performance, or the presence of cancer genes on BAC clones. Thus we collected, optimized, and expanded a conglomerate BAC clone set from multiple sources specifically for aCGH-based gene discovery, and we developed a set of associated aCGH-specific analytic tools to facilitate their use.

Unlike probes for global gene expression profiling, those used for aCGH must flank one another by small and consistent genomic distances. Minimization of this spacing provides sufficient resolution for detecting regions of DNA copy number

change small enough to be useful for subsequent gene identification. For analysis of malignant tissues, the ability to directly measure the copy number of known or suspected cancer-related genes by including BAC clones containing those genes is also desirable. Finally, like all hybridization probes, clones used for aCGH must provide reproducible data across multiple experiments. Thus, the initial BAC clone set we collected from multiple sources was optimized for performance in aCGH assays by including only those clones with: (1) unambiguous mapping data across genome builds based primarily on range-defining sequence anchors; (2) the capacity for reproducible, sensitive copy number determination; and (3) even spacing across the genome, so that the collection has no gap greater than 2 Mb and a mean spacing of less than 1 Mb. In meeting these criteria, we specifically selected clones that contained at least one of ~400 known and candidate cancer genes (list available in Supplemental Data available online at www.genome.org).

RESULTS

Initial Clone Selection

Five existing public and private BAC clone collections were accessed for development of the initial, unoptimized aCGH clone collection, serving as a conglomerate library on which the final compilation was based (Table 1). We selected these clones (n = 4275) by virtue of being the most comprehensive for targeted genomic regions, as well as for complementary coverage of the genome as a whole. The initial clone set was then subjected to a series of optimization steps described below and summarized in Figure 1.

Corresponding author.

E-MAIL weberb@mail.med.upenn.edu; FAX (215) 662-7617.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1847304>. Article published online before print in December 2003.

Table 1. Clone Collections by Source

Source	Clone numbers		
	Initial collection ^a	Post-alignment filter ^b	Optimized set ^c
The BAC Resource Consortium (Cheung et al. 2001)	3304	2454	2135
Wellcome Trust Sanger Institute (S. Gregory)	144	45	36
BAC-PAC Resource (CHORI)	355	330	320
NCI-Cancer Chromosome Aberration Project ^d	276	188	174
Cedar Sinai Medical Center (J. Korenberg)	196	165	144
TIGR (S. Zhao)	0	0	1325
Total	4275	3182	4134

^aThe number of clones received from each source. These include 130 duplicates.

^bThe number of clones after filtering each set for misaligned and ambiguously mapped clones.

^cThe aligned, quality adjusted 1 Mb resolution clone set containing 402 cancer-related genes.

^d<http://cgap.nci.nih.gov/Chromosomes/CCAP>.

Clone Optimization Based on Genome Alignment

As an initial step, 130 clones were removed for being redundant in two or more of the original clone sets. Next, the 4036 clones in the initial set for which sequence-tagged site (STS) mapping information was available to us were evaluated. These clones were identified by 2152 unique STSs. Clones identified by the same STS but not having the same clone name were retained because of the possibility of differential performance on the array, which we evaluated in a subsequent step. Clones identified only by STSs that could not be reliably aligned to the genome were removed. To this end, all STS locations were identified by aligning electronic polymerase chain reaction (ePCR) products (Schuler 1998, <http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi>) to the human genome sequence (build: Nov 2002). STS alignments were considered successful, and clones retained, if there was a single placement in a euchromatic region. Next, 402 STS-mapped clones from the initial collection that could not be mapped by ePCR were end-sequenced for genome placement.

The evaluation of end-sequenced and fully sequenced BAC clones was done using BLAT alignments (Kent 2002). Publicly available clone-identifying sequences were downloaded using GenBank accession numbers, and nonpublic sequences were acquired from the original clone source. Sequences that uniquely identified each clone were then used to query genome alignment (build: Nov 2002). Existing BLAT results for fully sequenced clones were downloaded directly from the UCSC Human Genome Browser (<http://www.genome.ucsc.edu>), and all were retained. End-sequenced clones were retained if alignments produced a single contig placement.

Thus clones were retained after the alignment analysis if their identifying sequences met at least one of the following criteria: (1) unambiguous STS placement by ePCR alignment, (2) unambiguous end-sequence pair placement within a plausible genomic distance (~50–250 kb), (3) a single end sequence in a concordant location with a unique STS placement, and/or (4) complete clone sequence was available. Based on these criteria, 963/4036 nonduplicate clones (24%) were removed from the initial set. Specifically, of the clones only anchored by an STS marker, 888 were eliminated as a result of 598 ambiguous STS alignments (~30% of all STS mapped clones). BLAT alignments eliminated 368 end-sequenced clones, primarily because of paired ends aligning to incompatible locations (e.g., different chromosomes). A small number of clones with a unique single end-sequence alignment were retained (n = 148) to maximize coverage in poorly represented genome locations. Finally, clones mapped only by fluorescence in situ hybridization (FISH) were removed (n = 109; Table 2).

Clone Optimization Based on aCGH Performance

After the clone set was optimized for genome alignment, aCGH arrays were fabricated using all clones remaining in the set (n = 3182) to evaluate individual clone performance. Normal DNAs (n = 50) and cancer cell line DNAs (n = 50; American Type Culture Collection) were cohybridized to the arrayed clone set in pairs, either as normal:normal or normal:cancer cell line cohybridizations, for the purpose of removing clones that yielded inconsistent results. Normal:normal DNA cohybridizations (n = 50) were done to optimize clone performance for reproducible detection of normal two-copy sequence. These experiments were done using DNA prepared from an individual donor (test, labeled with Cy3-dCTP) cohybridized to pooled normal DNA containing that individual (reference, labeled with Cy5-dCTP). Cancer cell:normal DNA cohybridizations (n = 50) provided a means of assessing clone performance under conditions where complex copy-number aberrations occur at multiple genomic sites. All experiments were done with two cohybridizations (“dye-swap”), one where cell line DNA was labeled with Cy3, and another where cell line DNA was labeled with Cy5. Pooled normal human male DNA served as the reference in all cases and was labeled with the opposite dye. Each parameter used for analysis was calculated for Cy3 and Cy5 channels separately.

Three hybridization traits, that is, reliability, reproducibility, and sensitivity, were used to measure the performance of each clone. Reliability was assessed using normal:normal cohybridization data where the Cy3/Cy5 intensity ratio for every clone was expected to be 1.0. Reproducibility was measured by two parameters: (1) signal variation relative to the mean [$1 - \text{standard deviation (sd)}/\text{mean}$], and (2) correlation between mean and median spot intensities. Sensitivity was defined by the

Table 2. Sequence Alignment Features of the Initial and Optimized Clone Sets

	Initial set		Optimized set	
	N	% Total	N	% Total
FISH only	109	2.5	0	—
STS-marker only	2335	54.6	1219	29.5
End-sequence	1579	36.9	2793	67.5
Fully-sequenced	122	2.9	122	2.9
Duplicate clones	130	3.1	0	—
Total	4275		4134	

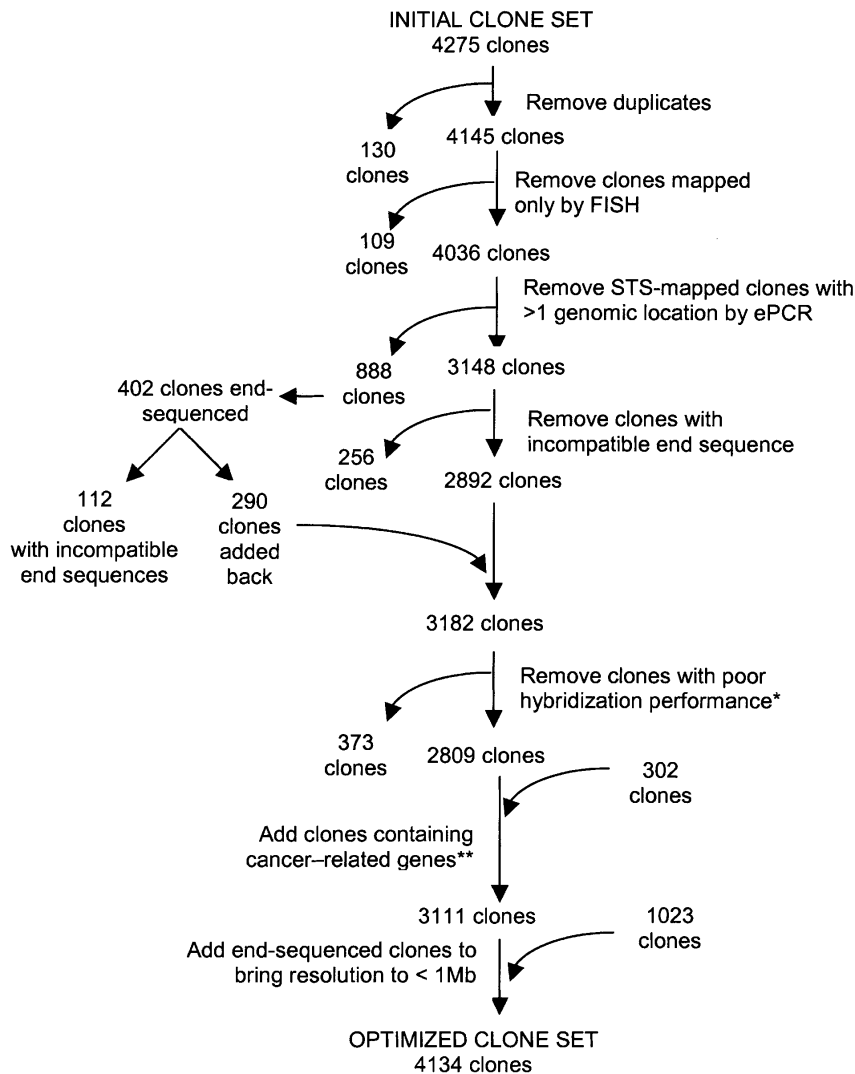


Figure 1 Flow chart describing the optimization process and the number of clones removed or added at each step. *Poor hybridization performance is defined in the text. **A list of the cancer-related genes can be found in the Supplemental data.

percent of pixels greater than two sd above background, and clones were eliminated if one of the following was true in the normal:cancer cell line cohybridizations: (1) the spot pixel signal variation was more than two sd above the mean for either of the measured parameters, (2) the spot pixel signal variation was more than one sd from the mean for both of the measured parameters,

and/or, (3) the percent of pixels more than two sd above background did not exceed a calculated threshold. That threshold was defined as more than 70% of pixels greater than two sd above mean background in more than 25% of experiments. Clones also were eliminated if, in normal:normal cohybridizations, the mean signal intensity ratio sd across all experiments was more than two sd above that of the entire clone set (Table 3).

Overall, 373 clones were eliminated due to poor aCGH performance (Fig. 1). Differing measures of signal variation eliminated largely the same clones, whereas differing measures of hybridization efficiency targeted different clones (overlap = 34 clones). Of note, clone elimination based on hybridization kinetics was done prior to assessing signal intensity ratios and sd from normal:normal hybridizations. This step included most of the clones that would have been eliminated due to varying ratios (overlap = 249 clones). In other words, poor hybridization resulted in inconsistent Cy3/Cy5 ratios.

Overall, 373 clones were eliminated due to poor aCGH performance (Fig. 1). Differing measures of signal variation eliminated largely the same clones, whereas differing measures of hybridization efficiency targeted different clones (overlap = 34 clones). Of note, clone elimination based on hybridization kinetics was done prior to assessing signal intensity ratios and sd from normal:normal hybridizations. This step included most of the clones that would have been eliminated due to varying ratios (overlap = 249 clones). In other words, poor hybridization resulted in inconsistent Cy3/Cy5 ratios.

Clone Optimization for Gene Coverage and Resolution

At this point in the optimization process, the set contained 2809 clones in 2025 contigs covering ~438.7 Mb (14.36%) of the human genome, with a mean gap size of 1.29 Mb ($\sigma = 1.87$ Mb). This intermediate set included clones containing 86 of 501 genes we designated as cancer-related based on existing literature (see Supplemental data).

With the goal of decreasing the mean gap size to less than 1 Mb and including as many of the cancer-related genes as possible, we drew 302 clones containing 337 cancer-related genes from among ~70,000 end-sequenced clones available from The Institute for Genome Research (TIGR; http://www.tigr.org/tdb/at/abe/bac_end_search.html). End-sequenced BACs containing the remaining 78 cancer-related genes were not available. Clone coverage then was optimized using the UCSC tiling path. All gaps greater than 2.0 Mb ($n = 379$, $\mu = 8.83$ Mb, $\sigma = 2.88$ Mb) were reduced below that size

Table 3. Performance Features of Initial and Optimized Set Represented by a Combined Score for Cy3 and Cy5 Channels

Quality parameter	Number filtered ^a	Initial set		Removed clones		Optimized set	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Signal variance	148	0.5	0.1	0.3	0.1	0.5	0.1
Mean-median correlation	73	0.8	0.2	0.3	0.8	0.8	0.1
Percent pixels >2 SD above background	203	88.8	10.2	58.2	13.8	90.1	7.1
Ratio std. dev. >0.2	336	1.1	0.5	1.5	1.4	1	0.1
Total	373						

^aMany clones were removed based on multiple criteria.



Figure 2 UCSC Human Genome Browser view of chromosome 2 (<http://www.genome.ucsc.edu>) with coverage of the optimized clone set (top track), the initial clone set (middle track), and the BAC Resource Consortium set (bottom track, from Cheung et al. 2001). The improved resolution of the optimized set is representative of that seen genome-wide.

limit using an additional 1023 end-sequenced clones (Fig. 1). Centromeric and subtelomeric regions were not filled. Representative coverage for chromosome 2 is shown in Figure 2.

The Optimized Clone Set

As a result of the process described above, the optimized BAC clone set consists of 4134 clones that directly cover 567 Mb of unique sequence (18.4% of the human genome; Tables 4, 5). Coverage occurs in 2711 contigs with some clone overlap as a result of aggregation in areas of high gene density. The mean gap size in the optimized set is 0.92 Mb ($\sigma = 1.17$ Mb), representing a statistically significant improvement over the initial collection of mapped clones in both gap size and deviation ($\text{prob} < |t| = 0.001$). Supplementing the initial collection with 1615 end-sequenced clones (290 clones selected from the initial set for end-sequencing and 1325 clones supplied by TIGR) raised the proportion of end-sequenced clones in the set from ~37% to ~68% (2793/4134), thereby reducing the proportion of clones mapped only by an STS marker from ~55% to ~30% (1219/4134; Table 2). The mean clone size of these end-sequenced clones is 177.5 kb ($\text{sd} = 30.7$ kb). No gap in gene-rich regions is larger than 1.15 Mb (Table 4), but subtelomeric regions and other areas of high repeat density inflate the sd of distance between covered regions and coverage variation between chromosomes ($\sigma = 2.79$, unit = %). This is especially true for acrocentric chromosomes and those with large heterochromatic regions (e.g., chromosomes 9, 13, and 14). In addition, gene-rich chromosomes have enhanced coverage because of supplementation of the set with cancer gene-specific clones (e.g. chromosomes 3, 7, and 19).

Online Resources and Analytic Software

Current mapping information and clone annotations for the optimized set are available at <http://acgh.afcri.upenn.edu>, a curated Web site that is updated with every genome release. This site provides an interface for querying the clone collection by name, feature (e.g., genes), or location. All clone data files are available for download. This site also contains detailed protocols for clone replication, array construction, hybridization, and initial data analysis.

In addition, this Web site contains downloadable software for multi-experiment analysis of DNA copy number data from aCGH or other array-based copy number determination platforms, such as "SNP chips". This software suite, called CGHAnalyzer, has three primary capabilities: (1) several systems of copy number determination, (2) multiple data display windows, and (3) analytical tools for integrating genomic data with aCGH results. Each component is customizable and geared toward automating data display (e.g., Veltman et al. 2003) and spot-calling (e.g., Hodgson et al. 2001). CGHAnalyzer, built on TIGR's Multiple Experiment Viewer (MeV) platform, contains all clustering algorithms available in MeV (<http://www.tigr.org/software/tm4/mev.html>) and is compatible with Windows, Mac OSX, and Sun Solaris operating systems.

Data from multiple experiments can be loaded into CGHAnalyzer from several standard array data formats or a database. The genomic location of clones also can be read from a file, which, if using the optimized clone set described here, can be queried remotely from the online database (<http://acgh.afcri.upenn.edu>). Copy number can be determined using signal intensity ratio thresholds or a clone-specific method derived from the normal distribution of the normal:normal DNA cohybridizations. These copy number determination methods can be used alone or in combination.

The main window of the CGHAnalyzer is designed for multiple experiment analysis (Fig. 3). Copy number profiles are de-

Table 4. Genome Coverage of the Initial and Optimized Clone Sets

	Initial set	Optimized set
Clones (N)	3182	4134
Gaps (N)	2048	2726
Mean gap size (kb)	1288	920
Gap size std. dev. (kb)	1865	1170
Total direct coverage (Kb)	438469	566968
% Direct coverage	14.3	18.4
Cancer gene coverage (N)	86 genes	423 genes

pictured as vertical color-coded bars parallel to an ideogram of each chromosome. Experiments can be arranged from left to right in customizable order, and gains and losses can be displayed separately on either side of the ideogram (annotated by genome coordinates), generating a view typical of metaphase CGH data display. The starting point for analysis in this view is individual clone data; however, these data can be transformed into genomic regions using a 'flanking region' approach. In this mode, flanking regions include all sequence between the nearest clones of different copy number to either side of the clone(s) that detect altered copy number. In addition, customized gene lists can be loaded to annotate the main CGHAnalyzer view. Tabular results can be exported for further analysis.

CircleViewer is the lowest-resolution display module, designed to view the copy number profile of the entire genome for a single aCGH experiment (Fig. 4). The third display module, CGHBrowser, displays aCGH data as line graphs in either linear order or scaled by genome coordinate (Fig. 5). Chromosomes and

experiments may be viewed individually, or the entire genome and multiple experiments may be seen in a single view. All display tools are mouse-driven and linked to one another such that the equivalent display of any active view is accessible in all other modules. Customized display windows can be exported as images for presentation purposes.

DISCUSSION

Widespread use of aCGH for high-throughput genomic analysis of cancer and other human diseases is currently limited primarily by the lack of a publicly available, unambiguously mapped high-resolution clone set that performs consistently in aCGH hybridizations. The optimized set described here has significant improvements over currently available clone collections with respect to density and distribution of genome coverage, as well as aCGH functionality. This clone set is publicly available, and may be obtained by requesting the UPenn aCGH BAC Clone Collection from CHORI (<http://www.chori.org/bacpac/>).

Another major limitation of aCGH studies has been the inability to easily summarize data on a genome-wide level with region-specific annotations across multiple experiments. To our knowledge, there are no publicly available analytical tools developed specifically for displaying and summarizing large aCGH data sets. The software we developed for this purpose provides a scope-independent medium for displaying and analyzing multiple aCGH experiments. The main window of CGHAnalyzer provides high-resolution perspective such that regions of commonality can be identified and integrated with genes and their annotations. CGHBrowser further increases the resolution of data viewing to allow more critical evaluation of data quality, and CGHCircleViewer decreases the resolution enough to view the

Table 5. Clone Composition of Initial and Optimized Sets

Chr	Initial set					Optimized set				
	Clones (N)	Mean gap size (kb)	Gap size std. dev. (kb)	Direct coverage (kb)	Percent direct coverage	Clones (N)	Mean gap size (kb)	Gap size std. dev. (kb)	Direct coverage (kb)	Percent direct coverage
1	246	930	1704	41350	16.8	372	831	1515	45827	18.6
2	292	1570	1673	29199	12	370	941	486	43101	17.8
3	209	857	1076	36124	18.1	242	743	727	40739	20.5
4	170	1741	1759	19673	10.2	266	935	489	32996	17.1
5	167	1587	1569	23640	12.9	240	947	553	34313	18.6
6	177	1485	1639	20316	11.9	224	962	530	28951	17
7	219	761	799	34103	21.6	237	678	553	37482	23.8
8	151	1572	1785	21143	14.4	188	999	687	28998	19.7
9	116	1521	3035	15295	11.6	159	1030	1789	21224	16
10	140	1300	1227	17639	13.1	189	910	553	23656	17.6
11	141	1377	1631	18984	13.8	194	848	578	25598	19.4
12	172	1297	1398	21449	16	216	897	682	27561	20.5
13	99	1931	2641	11079	9.8	136	1159	1508	16126	14.2
14	128	1521	2011	13092	12.6	151	1120	1524	16926	16.2
15	160	1272	2074	15241	15.3	151	1049	1581	17335	17.5
16	102	1158	2530	22165	24.1	129	850	1823	25403	27.7
17	84	945	1303	18936	22.2	109	743	727	21885	25.7
18	112	1229	1331	11177	14.4	127	845	576	15002	19.3
19	57	1569	1673	11076	18.5	84	748	712	12169	20.2
20	60	1240	1440	8293	13.1	82	723	629	11149	17.7
21	31	1613	2616	4408	9.9	45	1137	1825	6057	13.5
22	34	1216	2702	6390	13.4	49	1069	2501	7141	15
X	99	1669	1540	16078	10.6	147	1017	610	24108	15.9
Y	16	5675	9690	1618	2.8	27	2626	6457	3217	5.5
Uncertain	1093					0				
Total	4275					4134				

The number of clones and gap std. dev. are not perfectly related due to varying sizes of heterochromatic regions.

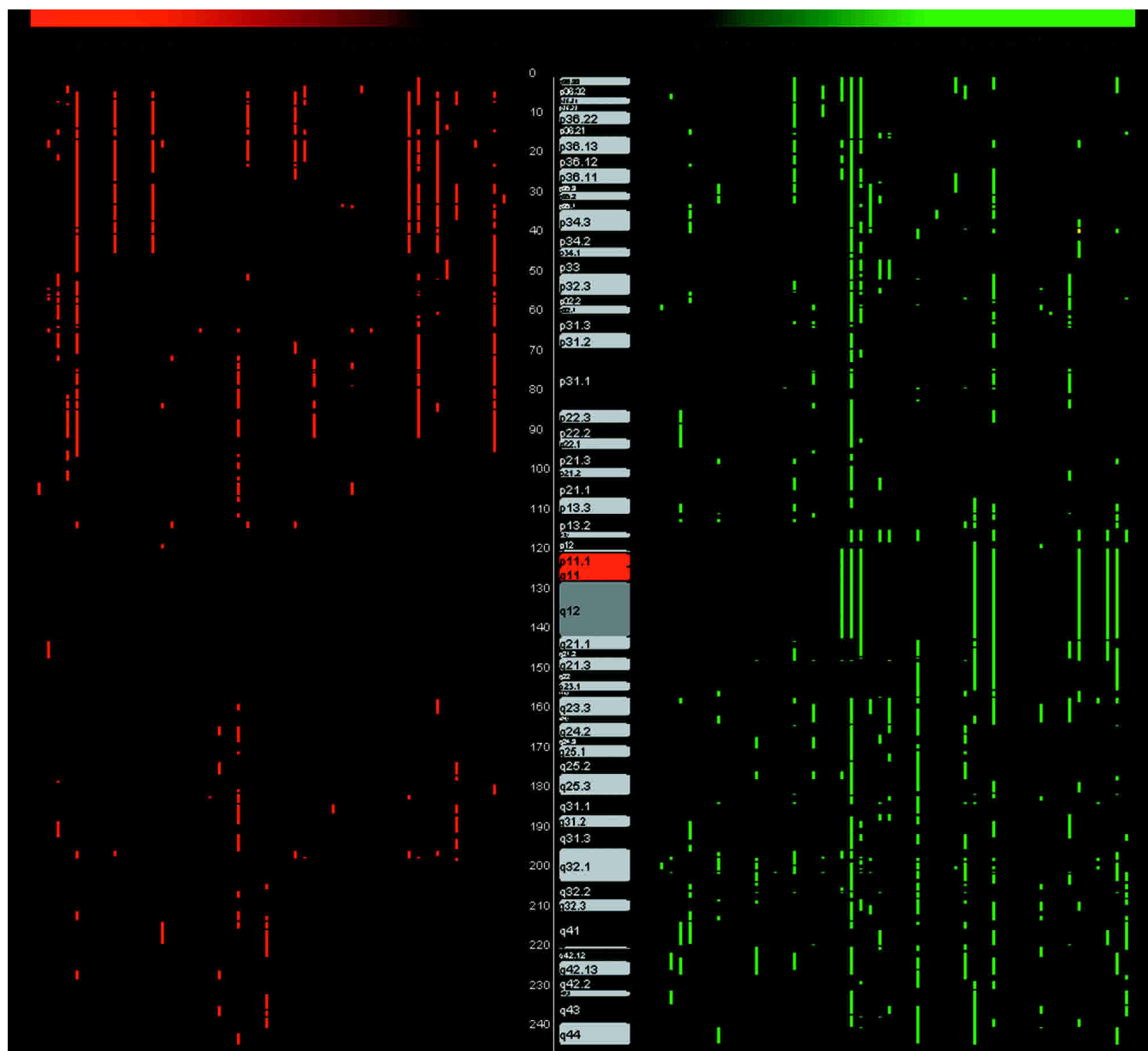


Figure 3 Cancer cell line dye-swap experiments loaded into the CGHAnalyzer main window (chromosome 1). Columns depict the copy number profile of each cell line with respect to its genome position on chromosome 2. Copy number gains are shown in green, losses are shown in red. Selecting a clone or region gives access to these data and the raw data in another window.

entire genome as a single sample on one screen. This view allows rapid assessment of the overall amount of chromosomal instability in a given sample and highlights large regions of copy number change. Combined with the clustering algorithms available from the TIGR MeV platform, these modules provide a comprehensive data analysis suite that greatly simplifies evaluation of aCGH data.

In summary, we have collected and evaluated a BAC clone set with comprehensive high-resolution coverage of the human genome. This set is optimized for aCGH performance and curated with cancer analysis in mind. We also have developed Web-based resources and a software suite to enhance the functionality and analysis of this clone set. These resources are easily modifiable for use with other clone sets from humans, mice, or other organisms and may be integrated with global expression analysis data as

well. These reagents and tools, now publicly available, should facilitate the widespread dissemination of aCGH, a powerful tool in the investigation of human disease.

METHODS

Array Construction

Arrays were constructed using a modification of a previously published protocol (Hodgson et al. 2001; see <http://acgh.afcri.upenn.edu>). BAC clone starter cultures were grown in 700 μ L 2X YT broth with 12.5 μ g/mL chloramphenicol at 37°C for ~15 h. Full cultures of 2.5 mL broth (as above) were inoculated from the starter cultures and grown for 15 h in 48-well blocks. DNA was extracted using high-throughput 96-well blocks (REAL prep kits, QIAGEN, yield ~35 μ g/culture). BAC DNA was then

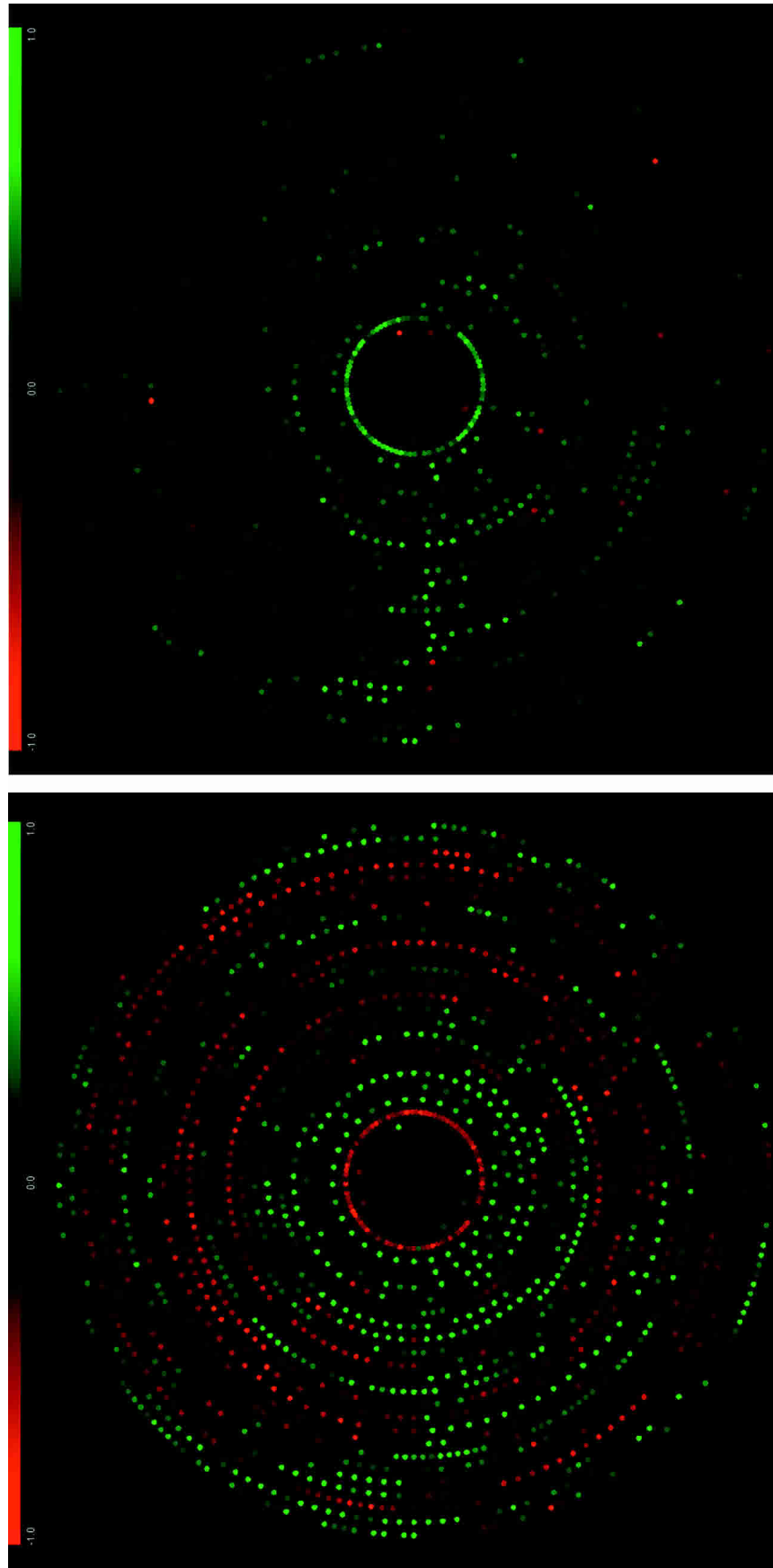


Figure 4 CircleViewer screen shots summarizing data from two cancer cell lines (NB13, left; PA1, right). The concentric circles represent chromosomes in order of size, and each spot represents a BAC clone. Gains (green) and losses (red) are derived from intensity ratios above or below an adjustable threshold (here, deletion ≤ 0.8 , gains ≥ 1.2). BAC clones can be identified and raw data accessed interactively.

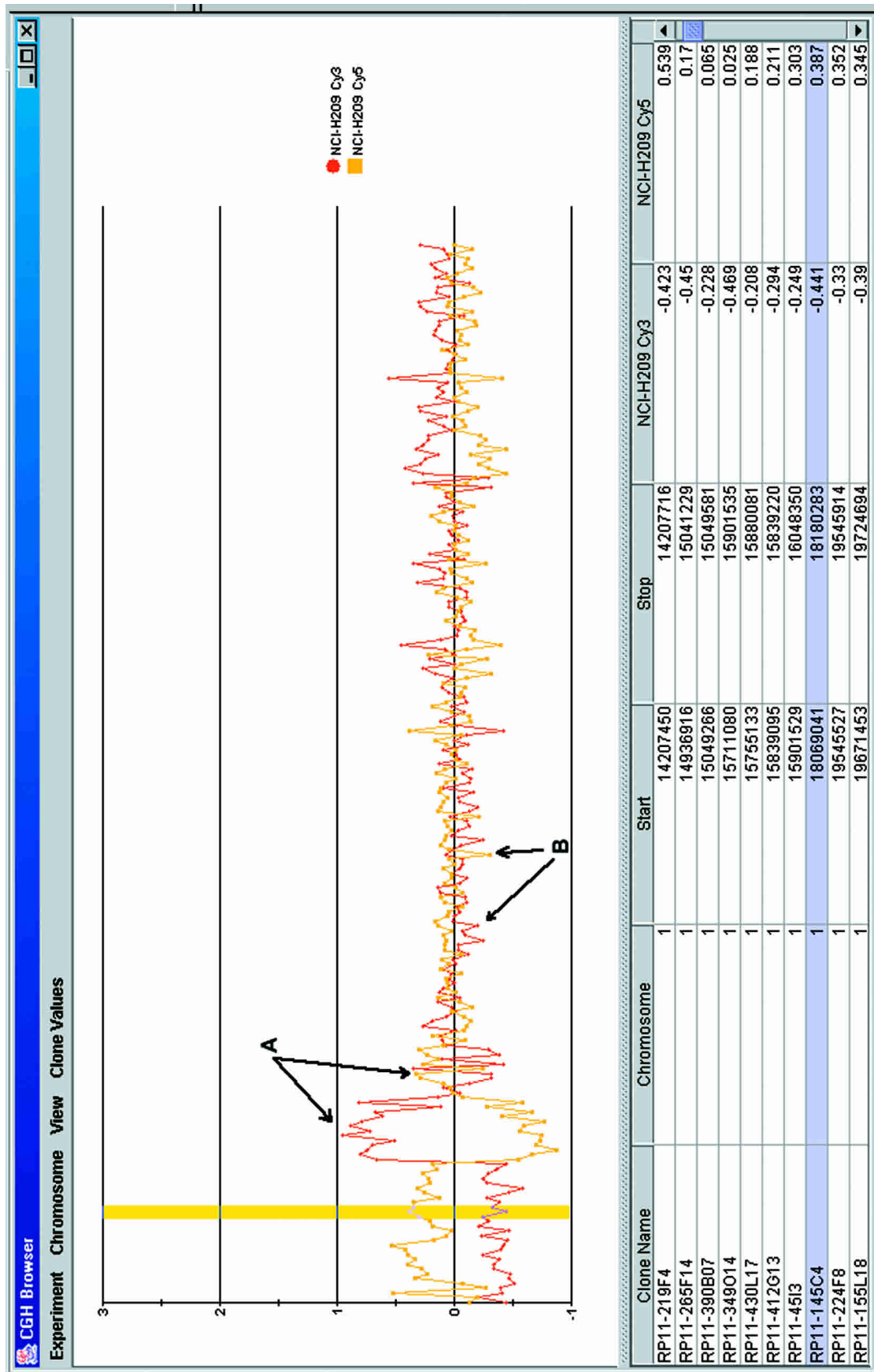


Figure 5 CGH Browser screen shot depicting a chromosome 1p deletion flanked by a copy number increase in the lung cancer cell line NCI-H209. This view can be used to access raw data and estimate the amplitude of copy number changes. Scores from any spot-calling algorithm and Cy3/Cy5 ratios in a log₂ and linear scale can also be visualized. The abscissa can be set to display data scaled to genome coordinates or in linear order (shown here). Intensity ratios with tumor DNA labeled with Cy3-dCTP are red; the same DNA labeled with Cy5-dCTP is represented in yellow. Ratios above or below the threshold for normal variation that represent true copy number gain or loss should have a dye-swap match of equal magnitude above or below y = 1 or y = -1, respectively (arrow A). Clones not conforming to this pattern consistently in normal: normal cohybridizations were eliminated from the set (arrow B).

amplified using degenerate oligonucleotide primers (Cheung and Nelson 1996). The primer used was 5'-CCGACTCGAGNNNNNATGTGG-3'. Each BAC was amplified twice at different annealing temperatures (58°C and 60°C), enhancing the overall PCR product coverage for each clone. PCR products were pooled and purified with 96-well QiaQuick kits (QIAGEN). The final yield of DNA for each BAC was typically between 10 and 15 µg/clone. Each clone DNA was lyophilized and resuspended in 50% DMSO/water such that the final concentration of DNA was ~300 ng/µL. A minimum of two replicates per clone were printed on each slide using the aqueous DMSO buffer as spotting solution. During testing, several printing configurations on three microarray spotters (Molecular Dynamics GenIII, Molecular Dynamics Lucidea, and Omnigrad 2000) were used to reduce positional and mechanical bias. Arrays were printed on Corning CMT Ultra-Gap slides.

Array Hybridization

DNA from peripheral blood and cancer cell lines was prepared using alkaline lysis. All DNAs were labeled using a random primer labeling kit (Invitrogen) with human Cot-1 DNA added to the hybridization solution to reduce nonspecific binding. Arrays were hybridized for 48–72 h at 37°C on a slowly rotating platform. Slide washing was done using published protocols (Hodgson et al. 2001). All images were scanned with an Affymetrix 428 microarray scanner (Affymetrix) and analyzed with Genepix software (Axon). A print-tip-based data adjustment was done on each array to account for variations in dye labeling efficiency and scanning variation such that the genomic copy number ratio was equal to 1 (Yang et al. 2001).

ACKNOWLEDGMENTS

This work was supported by funds from the Abramson Family Cancer Research Institute and the Breast Cancer Research Foundation (B.L.W.), The Institute for Genomic Research (S.Z.) and The Wellcome Trust Sanger Institute (P.A.F.).

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

Cheung, V.G. and Nelson, S.F. 1996. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of DNA. *Proc. Natl. Acad. Sci.* **93**: 14676–14679.

Cheung, V.G., Nowak, N., Jang, W., Kirsch, I.R., Zhao, S., Chen, X.N.,

Furey, T.S., Kim, U.J., Kuo, W.L., Olivier, M., et al. 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409**: 953–958.

Fritz, B., Schubert, F., Wrobel, G., Schwaenen, C., Wessendorf, S., Nessling, M., Korz, C., Rieker, R.J., Montgomery, K., Kuchlerlapati, R., et al. 2002. Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.* **62**: 2993–2998.

Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., Albertson, D.G., Pinkel, D., Collins, C., et al. 2001. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat. Genet.* **29**: 459–464.

Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12**: 656–664.

Pinkel, D., Segreaves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., et al. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**: 207–211.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., and Brown, P.O. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci.* **99**: 12963–12968.

Schuler, G.D. 1998. Electronic PCR: Bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol.* **16**: 456–459.

Veltman, J.A., Fridlyand, J., Pejavar, S., Olshen, A.B., Korkola, J.E., DeVries, S., Carroll, P., Kuo, W.L., Pinkel, D., Albertson, D., et al. 2003. Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.* **63**: 2872–2880.

Yang, L., Tran, D.K., and Wang, X. 2001. BADGE, Beads array for the detection of gene expression, a high-throughput diagnostic bioassay. *Genome Res.* **11**: 1888–1898.

WEB SITE REFERENCES

<http://www.ncbi.nlm.nih.gov/genome/sts/epcr.cgi>; National Center for Biotechnology Information ePCR search.

<http://www.genome.ucsc.edu/>; Human Genome Sequence browser.

http://www.tigr.org/tdb/at/abe/bac_end_search.html; TIGR human BAC clone resource.

<http://acgh.afcri.upenn.edu/>; Weber Lab aCGH resource site containing all clone data and analytic software.

<http://www.tigr.org/software/tm4/mev.html>; TIGR Multi Experiment Viewer software.

<http://www.chori.org/bacpac/>; CHORI BAC PAC Resource Web page for requesting the optimized clone set.

<http://cgap.nci.nih.gov/Chromosomes/CCAP/>; Cancer Chromosome Aberration Project.

Received August 4, 2003; accepted in revised form October 20, 2003.