High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays

Alexander Eckehart Urban*†‡, Jan O. Korbel‡§, Rebecca Selzer¶, Todd Richmond¶, April Hacker‖, George V. Popescu*†, Joseph F. Cubells**, Roland Green¶, Beverly S. Emanuel††, Mark B. Gerstein§, Sherman M. Weissman†‡‡, and Michael Snyder*‡‡

*Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520; †Department of Genetics and §Molecular Biophysics and Biochemistry Department, Yale University School of Medicine, New Haven, CT 06520; ¶NimbleGen Systems, Inc., 1 Science Court, Madison, WI 53711; **Departments of Human Genetics, and Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA 30322; ††Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; and ‖The Children's Hospital of Philadelphia, Philadelphia, PA 19104

Contributed by Sherman M. Weissman, December 31, 2005

Deletions and amplifications of the human genomic sequence (copy number polymorphisms) are the cause of numerous diseases and a potential cause of phenotypic variation in the normal population. Comparative genomic hybridization (CGH) has been developed as a useful tool for detecting alterations in DNA copy number that involve blocks of DNA several kilobases or larger in size. We have developed high-resolution CGH (HR-CGH) to detect accurately and with relatively little bias the presence and extent of chromosomal aberrations in human DNA. Maskless array synthesis was used to construct arrays containing 385,000 oligonucleotides with isothermal probes of 45–85 bp in length; arrays tiling the β-globin locus and chromosome 22q were prepared. Arrays with a 9-bp tiling path were used to map a 622-bp heterozygous deletion in the β-globin locus. Arrays with an 85-bp tiling path were used to analyze DNA from patients with copy number changes in the pericentromeric region of chromosome 22q. Heterozygous deletions and duplications as well as partial triploidies and partial tetraploidies of portions of chromosome 22q were mapped with high resolution (typically up to 200 bp) in each patient, and the precise breakpoints of two deletions were confirmed by DNA sequencing. Additional peaks potentially corresponding to known and novel additional CNPs were also observed. Our results demonstrate that HR-CGH allows the detection of copy number changes in the human genome at an unprecedented level of resolution.

22q11DS | comparative genomic hybridization | copy number polymorphism | copy number variation

Copy number polymorphisms (CNPs) in the genomes of normal individuals are far more common than previously thought (1–3), and it is likely that such differences play an important role in determining phenotypic variation in human populations. In addition, deletions, insertions, and other aberrations of the genomic sequence such as partial polyploidies have been shown to be responsible for several developmental disorders (4). Although widespread, the exact number and frequency of CNPs that exist between human individuals are poorly understood, largely because the methods for detecting these polymorphisms are still limited. The same limitations hamper progress in understanding the role CNPs and chromosomal aberrations play in the etiology of cancer and developmental diseases. Rapid, unbiased, and comprehensive methods to investigate copy number differences in the genomic sequence at high resolution are expected to be important tools for understanding variations in human populations as well as providing insights into their role in human disease.

Comparative genomic hybridization (CGH) was originally developed as a method for detecting large deletions or expansions of chromosomal material by fluorescence microscopy (4). Subsequently, arrays of bacterial artificial chromosomes or cosmids were prepared and used for detecting deletions and duplications >100 kb. However, many variations in copy number are likely to be much smaller; and even for the analysis of larger CNPs, it is desirable to map the locations of their breakpoints with as high a degree of accuracy as possible.

Recently, methods for reducing the complexity of DNA samples have been coupled with oligonucleotide array analysis to detect deletions and duplications of ≈100 kb or larger (1, 5). However, facile high-resolution methods for unbiased and precise mapping of a wide range of CNPs, including small to moderately sized (≈0.5–10 kb) deletions and insertions, across large regions of the human genome using total genomic DNA from stable cell lines or patient blood samples have not been described previously.

Oligonucleotide arrays with 385,000 features can be synthesized by photolithography (6–8). By tiling large segments of genomic DNA, these arrays have the potential to map deletions at very high resolution. In addition, the sensitivity of suitably designed arrays is sufficiently high that total genomic DNA can be directly hybridized, thus avoiding bias that arises during selective PCR amplification of subsets of the DNA.

Here, we describe a microarray technology, high-resolution CGH (HR-CGH) for precise copy number mapping in mammalian cells. We applied this technology to the analysis of patients that have alterations in copy number in chromosome 22q11 and the pericentromeric region of chromosome 22. The patients include those with Cat-Eye syndrome (partial tetraploidy in 22q), Emanuel syndrome (also der22 syndrome, partial triploidy in 22q), Dup22 syndrome (duplication in 22q11), and 22q11 Deletion syndrome (22q11DS, also velocardiofacial syndrome or DiGeorge syndrome, deletions of variable size in 22q11) (9, 10). 22q11DS in particular is a very important model for diseases resulting from abnormal embryonal development. In addition to morphological abnormalities, a substantial proportion of 22q11DS patients develop learning disabilities, attention-deficit hyperactivity disorder, autistic-spectrum disorders, and schizophrenia and related psychoses (11). The 22q11 region contains several large (60- to 600-kb) clusters of segmental duplications [low-copy repeats (LCRs)] that are often the site of chromosomal rearrangement breakpoints (2, 12–14). However, patients with deletions that appear to be of the same size and location when analyzed by conventional methods, typically stretching from one LCR to another, show considerable variation in their

Conflict of interest statement: No conflicts declared.

Abbreviations: CNP, copy number polymorphism; CGH, comparative genomic hybridization; HR-CGH, high-resolution CGH; LCR, low-copy repeat.

Data deposition: The microarray data for this paper have been deposited in the Gene Expression Omnibus repository (accession no. GSE4240).

‡A.E.U. and J.O.K. contributed equally to this work.

‡‡To whom correspondence may be addressed. E-mail: sherman.weissman@yale.edu or michael.snyder@yale.edu.

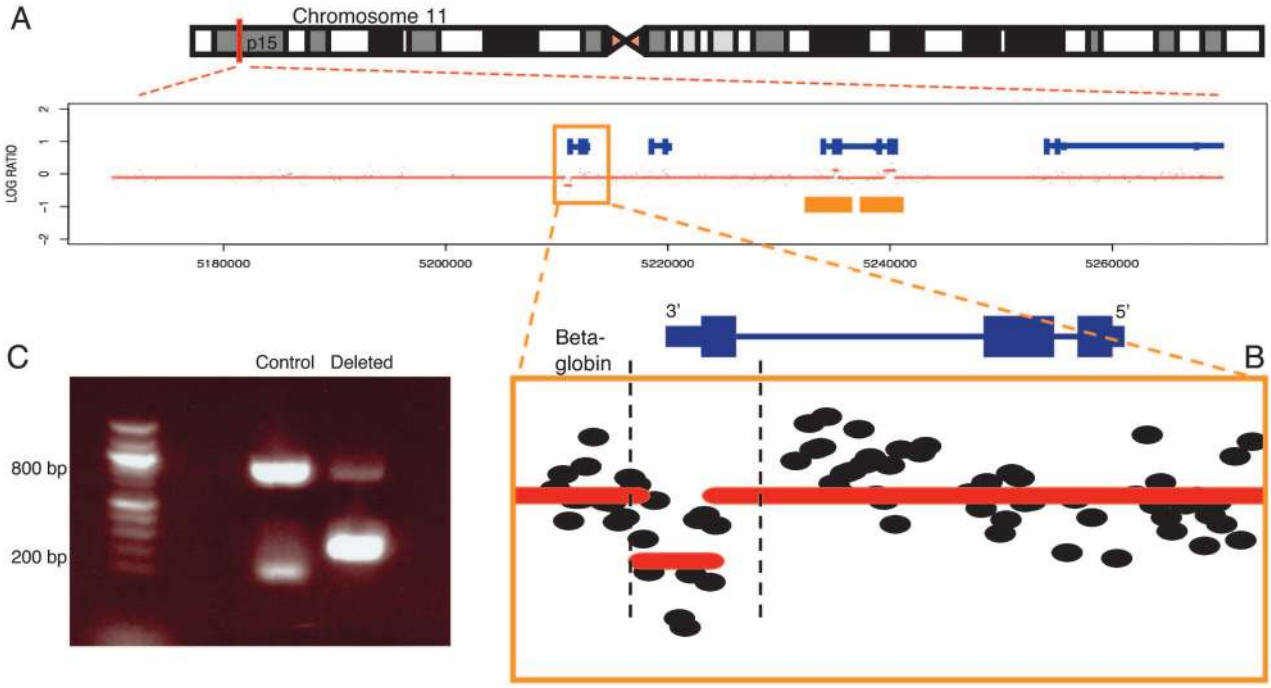

Fig. 1. Detection and analysis of a small heterozygous deletion in the β-globin locus. (*A*) The region flanking the β-globin gene, represented on the high-density HR-CGH array, and the signal obtained from probing the array with DNA from patient 05-029 (processed signal average of two probings). Blue, exons of known genes; orange, segmental duplication. (*B*) The signal indicating position and extent (indicated by dashed lines) of a ≈600-bp heterozygous deletion. (*C*) PCR analysis of the deletion locus yields an additional fragment from the patient sample ≈600 bp smaller in size than the only fragment from the control sample; its size was determined by DNA sequencing to be 622 bp.

symptoms. In addition, other patients with atypical deletion sizes and locations can be afflicted by a similar combination of symptoms. It is therefore important to determine the precise deletion endpoints in these patients. This has proven difficult because of the large degree of variability in the size of the deletions and the fact that the breakpoints are often located in regions containing segmental duplications.

Our analysis of DNA from patients containing copy number variation in the pericentromeric region of chromosome 22q has revealed remarkable heterogeneity in the location of chromosomal breakpoints. In addition, our results and those of an independent study (15) indicate that HR-CGH technology has the ability to rapidly map chromosomal alterations throughout large regions of the human genome. We demonstrate that HR-CGH can detect a heterozygous deletion of <700 bp, allows the mapping of deletions with end points residing within LCR regions, and can map chromosomal aberrations with a resolution of 50–200 bp, making it a useful tool for the precise mapping of a wide variety of CNPs.

Results

Detection of a Small Deletion in the β-Globin Locus. We first established HR-CGH technology and its resolution limits using high-density isothermal arrays and DNA from an individual containing a known heterozygous deletion in the β-globin locus. This deletion was reported to be 619 nucleotides long and to remove the 3′ exon and flanking noncoding sequences of the β-globin gene (16). DNA isolated from blood samples of this individual was labeled with Cy3 and hybridized to microarrays that tile through 100 kb of the β-globin locus. The arrays are of very high density, with overlapping isothermal oligonucleotides spaced 9 bp apart along the tiling path. As a control, a pool of DNA from seven individuals not known to contain chromosomal aberrations was labeled with Cy5, and hybridized simultaneously to the same array. Two independent probings of the β-globin array were performed. After normalization, the log ratio of the signals from the heterozygous β-globin deletion patients relative to the control pool sample was determined (see *Materials and Methods*).

We first analyzed the signal from two independent probings of the β-globin array. When averaging the processed signal from the two probings, most of the region exhibits similar signals between the DNA probe from the β-globin deletion patient and that of the control pool. However, one region that exhibited reduced DNA content in the patient sample was identified (Fig. 1). Using algorithms we developed to map alterations in copy number, we were able to locate the affected region and predict a deletion of 561 bp. A third independent probing of a β-globin array led to a prediction of a deletion in the same region and of a size of 725 bp. PCR primers were designed (16) to flank the deletion and amplify DNA from normal and mutant cells. DNA from the patient yields an additional fragment that is ≈600 bp shorter than the fragment amplified from control DNA (Fig. 1*C*). Sequence analysis of this fragment reveals the location of a 622-bp deletion precisely within the identified region. Thus, HR-CGH was able to accurately identify a 622-bp heterozygous deletion in mammalian DNA, and the breakpoints were found to lie within ≈50 bp of the predicted site.

Detection and Mapping of an Atypical Deletion in a Patient with 22q11DS. We next sought to determine whether heterozygous deletions could be detected and accurately mapped using a whole-chromosome tiling array. An isothermal array containing 45- to 85-bp oligonucleotides (average length 55 nucleotides) that tiles both strands of human chromosome 22q was designed. The start points of the probes are spaced 85 bp apart throughout the nonrepetitive sequences on each strand. The array was probed in three independent experiments with DNA isolated from a 22q11DS patient (patient 04-018) who had a heterozygous deletion in the 22q11 region. The deletion in this patient had been identified by using FISH and PCR-based methods; it was expected to be atypical for this disease, with only one of the two breakpoints located within an LCR, but its precise breakpoints were not known. Control DNA

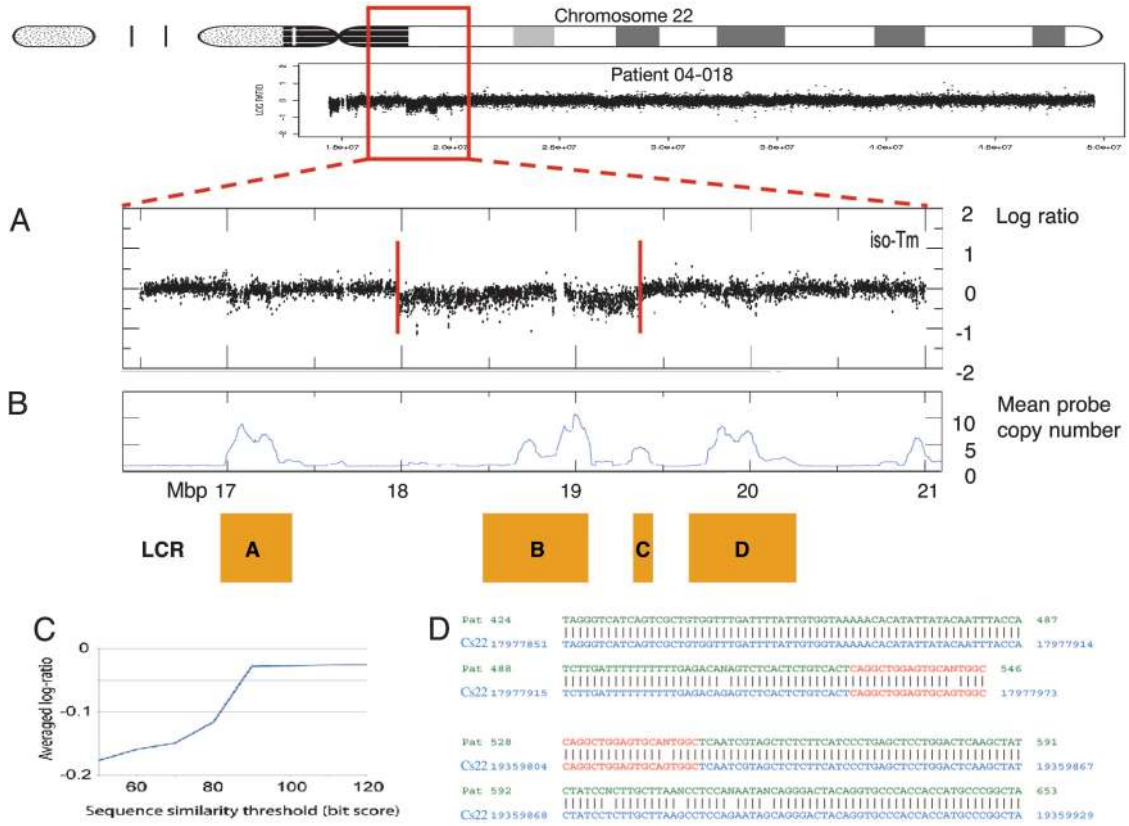

Fig. 2. Detection and analysis of a large deletion in chromosome 22q11 of patient 04-018. (*A*) Isothermal array interrogating the 22q11 region; breakpoint predictions resulted in an overall resolution (mean distance between predicted and verified breakpoints) of 620 bp. (*B*) Probe redundancy as determined by BLASTN (100,000-bp sliding window average; 80% minimum sequence identity) vs. chromosomal coordinates. Regions of high DNA repeat density coincide with one of the predicted breakpoints and with an additional region of less than clear deletion signal seemingly interrupting the 1.4 Mbp deletion at ≈19 Mbp (location of a known LCR that presumably causes crosshybridization). (*C*) Quantification of the contribution of crosshybridization originating from an LCR. Redundant oligomers were removed by applying BLASTN cutoff scores of varying stringency. Averaged log ratios are given for the entire LCR region from 18.7 to 19 Mbp. Using a threshold of 50 results in a log2 ratio of roughly −0.2, a value near the average log2 ratio of heterozygous deletions. This indicates that the region with less than clear deletion signal from ≈18.7 to 19.0 Mbp is due to crosshybridization, and that the deletion detected is contiguous. (*D*) Sequencing of the affected region verifies the 1.4-Mb deletion breakpoints. Pat, patient (green); CHR22, human reference genome (blue). The red sequence block corresponds to a 19-bp sequence framing the large deletion on both sides, suggesting a potential mechanism of deletion.

from the pool of seven normal individuals was used as a reference. A 1.4-Mbp region was identified that contained a reduced signal in the patient relative to DNA from normal patients (Fig. 2). The endpoints of this deletion region suggested that the breakpoint proximal to the centromere lies between LCRs A and B, whereas the distal breakpoint resides in or very near LCR C. As a control for the performance of the array, DNA from patient 04-018 was labeled with two different dyes and hybridized to the array; the ratio of the two different dyes was relatively uniform across the array (Fig. 5, which is published as supporting information on the PNAS web site).

Although diminished signals indicating the positions of the deletion are readily apparent, some chromosomal regions show a less clear deletion signal. These segments map to the LCR repeats, which are present in the pericentromeric regions of chromosome 22 (Fig. 2). The less-than-expected loss in signal in these regions is likely due to crosshybridization with sequences of homologous stretches of other LCRs. For example, heterozygous deletion of one copy of a three-copy LCR repeat would produce only a 17% (1/6) change in signal intensity.

Analysis of the signal from three replicate probings of the isothermal HR-CGH chromosome 22q array allowed us to predict the location of the deletion breakpoints in the genomic sequence. PCR analysis confirmed the presence of the breakpoint in the region predicted by HR-CGH. Sequence analysis of the breakpoint fragment revealed the precise location of the breakpoints to be within a pair of identical 19-bp sequence blocks, one of which resides in LCR-C, as expected (Fig. 2). The breakpoint prediction based only on HR-CGH data from any one of the three replicate probings was accurate to within, on average, 624 bp; of the six end-point predictions (three replicas from each of two ends), five were accurate to within a range from 178 to 628 bp, and one prediction was correct to within 1,642 bp (Table 1, which is published as supporting information on the PNAS web site). The repeat content in the LCR regions made processing of the data more challenging, but there were enough informative oligonucleotides present on the array even in those regions to allow for an unambiguous prediction of breakpoints. To our knowledge, a deletion endpoint lying within the boundaries of an LCR in a patient with 22q11DS has not been mapped before to nucleotide resolution. Thus, high-resolution breakpoint mapping could be achieved by HR-CGH based on oligonucleotide tiling arrays.

A Wide Spectrum of Heterozygous Deletions in 22q11 Can Be Mapped Using HR-CGH. We next used the comprehensive isothermal chromosome 22q tiling arrays to determine the breakpoints in samples in which deletions of various sizes have been predicted. DNA from seven additional patients containing heterozygous deletions in the 22q11 region were hybridized to the isothermal chromosome 22q array along with the control DNA pools, and the relative amounts of DNA were measured. Regions of reduced hybridization relative to control samples could be detected in each of the eight patients

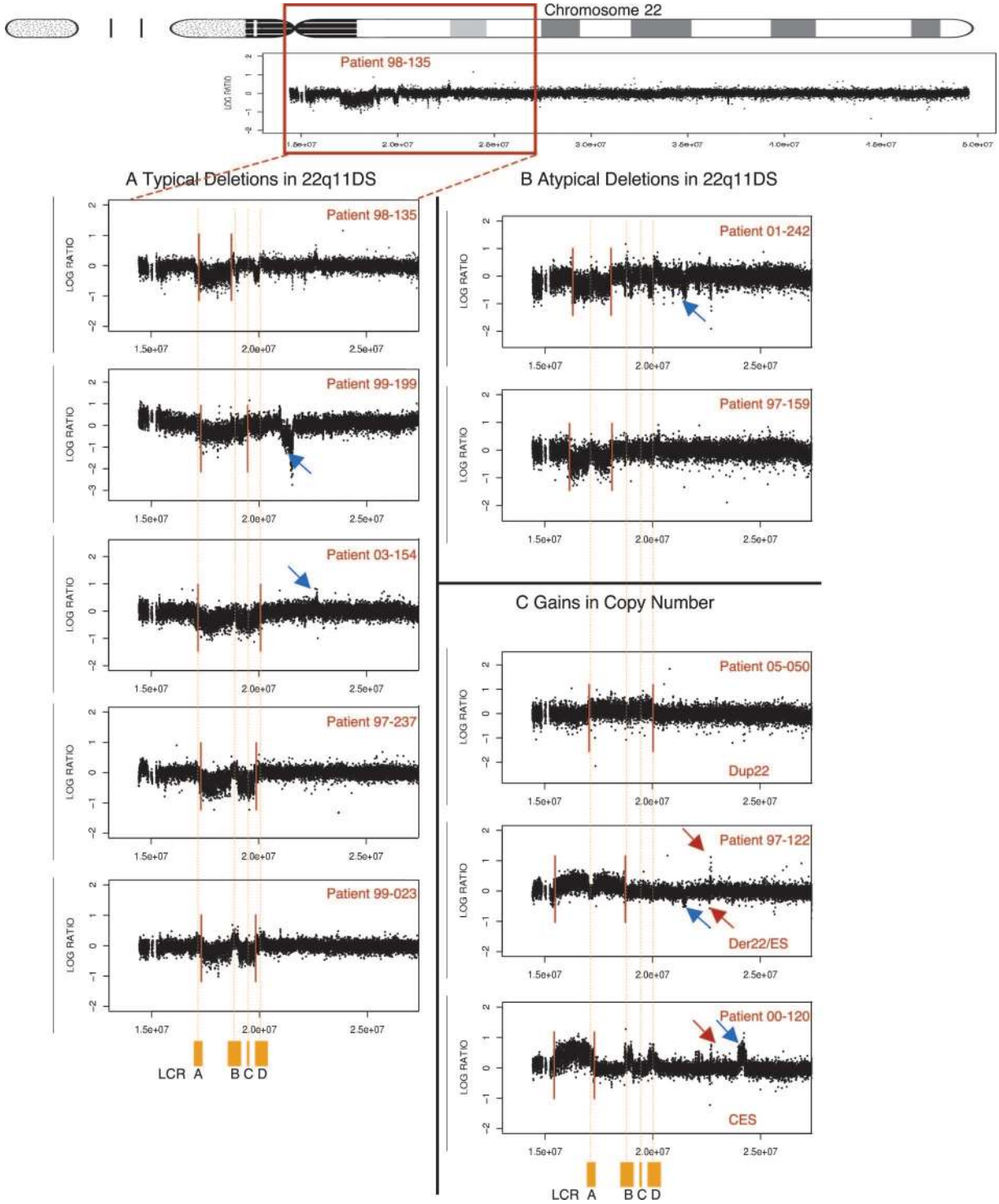

Fig. 3. Various types of deletion as well as gains in copy number can be detected with the chromosome 22q isothermal HR-CGH array. (*A*) Deletion types typical for 22q11DS. (*B*) Deletion types atypical for 22q11DS. (*C*) Three different cases of gains in copy number in 22q11. Vertical red lines indicate position of breakpoint prediction. Arrows indicate additional putative CNPs [blue, known CNP; red, new CNP (see Table 2); orange, LCR regions].

(Fig. 3). Based on the signal from these hybridizations, deletion sizes ranging from 1.4 to 3 Mbp are observed, and the endpoints of these deletions can be mapped (Fig. 3 and Table 1). Importantly, breakpoint prediction by HR-CGH can reveal differences in patients whose chromosomal abnormalities had been undistinguishable by other methods (Fig. 4). Two samples from patients with typical LCR A→ D deletions (patients 03-154 and 97-237, Fig. 4) that were indistinguishable by FISH analysis exhibit differences of up to 200 kb on each side of the breakpoint by HR-CGH. These differences affect at least seven known and seven predicted genes.

Detection of Increases in Copy Number Using Tiling CGH. In addition to examining patients who had chromosomal deletions, we also investigated whether increases in copy number can be detected. DNA from a patient with Emanuel syndrome (also known as der22 syndrome) and another with Cat-Eye syndrome, which are known

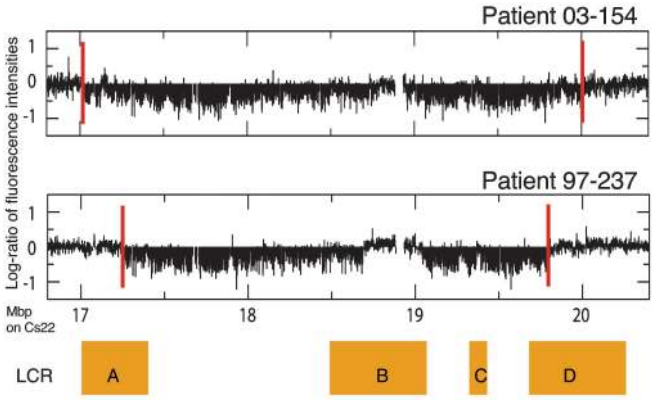

Fig. 4. The isothermal HR-CGH array covering chromosome 22q allows predicting substantially different deletion sizes for two 22Q11DS patients with typical A→D deletion that had not been distinguishable by conventional methods (breakpoint predictions are indicated by vertical red lines).

to result from partial triploidy or partial tetraploidy, respectively, as well as DNA from a patient with Dup22 syndrome, which results from a heterozygous duplication in 22q11, was labeled and hybridized to the isothermal chromosome 22q array. Increased levels of 22q11 sequences were detected in the DNA of patients with these syndromes (Fig. 3*C*). The signals in the single-copy regions are approximately proportional to the increased copy number of DNA in the patient. Thus, increases in genomic region copy number can be detected, mapped, and quantified using this approach.

Additional Putative CNPs in Chromosome 22q. Although CNPs were until recently considered an exception rather than common in human biology, recent studies suggest a complex pattern of copy number variation in the human population (1, 3). We therefore searched for additional chromosomal aberrations besides those identified as the main aberration in the 22q11 region and the β-globin locus on chromosome 11p15, respectively. We observed a large number of signals that correspond in their chromosomal coordinates to previously reported CNPs (3, 17). In addition to these putative known CNPs, in our samples we also found evidence for novel CNPs. Using a BLAST-based strategy, we compiled a set of 10 additional CNPs observed at various locations along chromosome 22q in our group of patients for which the signal is unlikely to be caused by crosshybridization (Fig. 3 and Table 2, which is published as supporting information on the PNAS web site). Half of these CNPs map to chromosomal coordinates for which there had been no CNP described previously; the remainder localize to positions where CNPs were already reported by others. These results indicate that HR-CGH allows the detection of known and novel CNPs and should have the potential to refine positional information to >1-kb accuracy.

Discussion

Detecting and mapping chromosomal copy number variation are of high importance for understanding human variation, evolution, and disease. Here we present the first high-resolution unbiased mapping method for detecting changes in DNA copy number using isothermal DNA tiling arrays and genomic DNA from patients with complex noncancerous genetic diseases. Previous studies to map such changes were either at low resolution (bacterial artificial chromosomes) and/or required the selective and probably nonuniform PCR amplification of a subset of regions from within the genomic DNA sequence (i.e., ROMA, representational oligonucleotide microarray analysis; ref. 1). The advantage of our approach is that the use of high-density tiling arrays allows precise determination of breakpoints (usually within several hundred base pairs or less), can be carried out by directly labeling DNA, and is quite reproducible. This simple method should thus allow the high-throughput mapping of any variation in chromosomal copy number throughout the human genome.

Using the tiling CGH method, we mapped breakpoints for nine heterozygous deletions, covering the spectrum of typical deletions in 22q11DS and mapping several atypical deletions as well. We also mapped a heterozygous duplication, a partial triploidy, and a partial tetraploidy in human DNA.

Thus HR-CGH is useful in most situations encountered in the course of studying copy number variations in the genomic sequence.

In most cases, the breakpoints resided in LCR sequences that are found at several locations on chromosome 22. These LCR sequences are suspected to play a role in the mechanisms that mediate chromosome breakage leading to deletion or insertion of chromosome fragments (18). It is likely that, in many and probably the majority of cases, the occurrence of a CNP correlated with an LCR sequence, regardless of whether the given CNP is directly involved in the etiology of a disease (2, 13). Although it is therefore to be expected that a breakpoint will often be located within the boundaries of an LCR, the repetitive nature of these regions has made it very difficult to precisely determine the extent of a CNP. We have shown that HR-CGH allows prediction of the extent of deletions that stretch into LCR regions, revealing that, between two patients with a hitherto undistinguishable genotype, the heterozygous gene complement might differ by more than a dozen genes. This ability should allow for substantial progress in explaining differences in phenotype between patients with chromosomal aberrations in the same region.

Although recent studies have uncovered a high incidence of copy number variations, the number and frequency of such variations that occur throughout the human genome are not known. In particular, before the development of HR-CGH, it was not possible to survey entire long stretches of genomes in high-throughput fashion for rearrangements significantly less than 100 kb in size, and there is no information as to whether the smaller deletions and insertions are more or less frequent than larger-scale copy number variations. Although the number of CNPs is probably less than the number of single-nucleotide polymorphisms, the total number of DNA bases involved in copy number changes may well be at least of the same order as the number of polymorphic bases in humans. This notion is supported by a recent comparison of the chimpanzee genome with the human reference genome sequence, which showed that most of the human–chimp sequence divergence is indeed due to large-scale structural variation (19). With the ability to survey the entire human genome (8) and the advent of even higher-density arrays, it should ultimately be possible to array the entire human genome on one or a few chips. Thus, it should be possible to analyze DNA from every person as well as somatic samples from disease tissue at a relatively low cost with HR-CGH.

Our results can be expected to have a strong impact on several fields. The detection of these genomic variations provides a tool for studying human population variation and flow and for genetically mapping disease loci. Mendelian disorders in humans are frequently the result of DNA deletions within the size range detected by the present methodology, permitting direct identification of such mutations. Conventional single-nucleotide polymorphism (SNP) approaches for studying genetic traits might often miss such copy number variations, either because they do not produce new SNPs or because they produce heterozygous patterns different from the 1:1 SNP ratios expected for a diploid organism. Importantly, common copy number variations might be expected to have subtle effects on the normal phenotype and perhaps contribute to the relative susceptibility to common diseases as well as to the sensitivity to pharmaceutical compounds. Thus, this approach should have broad medical utility.

Materials and Methods

Microarray Construction. We used NimbleGen Systems-manufactured maskless array synthesis (MAS) arrays (6) for two different

regions of the human genome for this study. The sequence of human chromosome 22q was tiled with oligonucleotides that start every 85 bp. A 100-kb region around the human β-globin locus on chromosome 11p15 was tiled with oligonucleotides starting every 9 bp. In each case, the repetitive parts of the sequence were identified by using REPEATMASKER (20), and oligonucleotide probes were designed to represent the nonrepetitive DNA sequence. The oligonucleotides were designed to be of variable length, adjusted to match a target melting temperature of 76°C according to the following formula: $T_m = 81.5 + 16.6*(\log_{10}([Na^+])) + 0.41 \times (\%GC) - 600/\text{length}$. Probe lengths were constrained to be a minimum of 45 and maximum of 85 bp.

Sample Labeling and Hybridization. The labeling of genomic DNA samples for HR-CGH was performed as described (15, 21), with modifications. Genomic DNA was randomly fragmented by sonication. Each DNA sample (1 μg) was denatured in the presence of 5′-Cy3- or Cy5-labeled random nonamer (TriLink Biotechnologies, San Diego) and incubated with 100 units (exo-) Klenow fragment (NEB, Beverly, MA) and dNTP mix [6 mM each in TE buffer (10 mM Tris/1 mM EDTA, pH 7.4; Invitrogen) for 2 h at 37°C. Reactions were terminated by addition of 0.5 M EDTA (pH 8.0), precipitated with isopropanol, and resuspended in water. The Cy-labeled test sample (Cy3) and a reference sample composed of a pool of genomic DNA from seven normal male individuals (Promega) (Cy5) were combined in 40 μl of NimbleGen Hybridization Buffer (NimbleGen Systems). After denaturation, hybridization was carried out on a MAUI Hybridization System (BioMicro Systems, Salt Lake City) for 18 h at 42°C at the NimbleGen Service Laboratory. The arrays were washed by using NimbleGen Wash Buffer System (NimbleGen Systems), dried by centrifugation, and scanned at 5-μm resolution by using the GenePix 4000B scanner (Axon Instruments, Union City, CA).

Data Analysis. Fluorescence intensity raw data were obtained from scanned images of the oligonucleotide tiling arrays by using NIMBLESCAN 2.0 extraction software (NimbleGen Systems). For each spot on the array, log2-ratios of the Cy3-labeled test sample versus the Cy5-labeled reference sample were calculated. Fluorescence intensity normalization was performed by using the Qspline algorithm (22) (normalize.qspline, www.bioconductor.org). After normalization, array data were subjected to a computational analysis procedure implemented in R and PERL, which is suited to process high-resolution data originating from HR-CGH experiments and to predict chromosomal breakpoints at <500-bp resolution (unpublished work).

Correlating Breakpoints with Repeat Density. To determine the extent to which DNA sequence repeat density correlates with the chromosomal aberrations as well as with the signal patterns from the arrays, we used BLASTN (23) searches of the oligonucleotide sequences present on the isothermal chromosome 22q array against the human genome (and also against separate chromosomes). With a sequence identity cutoff of 80%, this led to the identification of regions with high repeat density (Fig. 2), which largely correspond to previously described segmental duplications [see, for example, http://projects.tcag.ca/humandup (17)]. Some of these regions coincide well with the LCR regions in 22q, including those relevant for diseases resulting from aberrations in 22q11 (Fig. 2). Of the 24 predicted breakpoints reported here, 15 reside within LCR regions.

We analyzed the log-intensity ratios within the ≈1.4-Mb deletion region of patient 04-018 and found a good agreement of intensity ratio and repeat density (i.e., the number of BLASTN matches of the particular oligonucleotide sequence in the genome; Fig. 2). The repeat density can be used to reanalyze the probe intensities, allowing for partial compensation of deficiencies in signal-to-noise ratios.

Comparison with Known CNPs. To assess whether there is a correlation between CNPs reported by others and potential additional CNPs outside of the primary aberration in our samples, we selected the CNPs on chromosome 22 reported with strongest support in the Database of Genomic Variants (17), those CNPs mapped with a resolution of at least 200 kb. We considered copy number changes from this data set and potential additional CNPs in our samples as corresponding if both of their respective boundaries were separated by a distance <200 kb (see *Supporting Text*, which is published as supporting information on the PNAS web site).

Validation of Breakpoint Predictions. All of the samples used in this study had previously been analyzed by FISH and many of them by PCR-based methods.

For β-thalassemia patient 05-029, PCR was performed by using primers expected to flank the β-globin breakpoint (16). The resulting PCR product (Fig. 1*C*) was sequenced and matched to the genomic sequence to precisely determine the breakpoint. For 22q11DS patient 04-018, the regions surrounding the suspected breakpoint were amplified by vectorette PCR [Sigma-Aldrich (24)] by using DNA from lymphoblast cells derived from normal individuals (Promega pool) and patients. A unique 2-kb band that spans the suspected breakpoint region was identified and sequenced.

We thank H.-Y. Luo and D. H. K. Chui of Boston Medical Center (Boston) for providing genomic DNA with a 3′ deletion in the β-globin gene, B. E. Morrow of the Albert Einstein College of Medicine (New York) for 22q11DS cell lines for preliminary studies and advice on how to design the research, K. K. Kidd for critical comments on the manuscript, and N. J. Carriero for help with data processing. M.S. and S.M.W. are supported by grants from the National Institutes of Health. B.S.E. is supported by National Institutes of Health Grant CA39926. J.O.K. was supported by a European Molecular Biology Organization Long-Term Fellowship.

1. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., *et al.* (2004) *Science* **305,** 525–528.
2. Tuzun, E., Sharp, A. J., Bailey, J. A., Kaul, R., Morrison, V. A., Pertz, L. M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., *et al.* (2005) *Nat. Genet*. **37,** 727–732.
3. Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W. & Lee, C. (2004) *Nat. Genet*. **36,** 949–951.
4. Speicher, M. R. & Carter, N. P. (2005) *Nat. Rev. Genet.* **6,** 782–792.
5. Jobanputra, V., Sebat, J., Troge, J., Chung, W., Anyane-Yeboa, K., Wigler, M. & Warburton, D. (2005) *Genet Med*. **7,** 111–118.
6. Nuwaysir, E. F., Huang, W., Albert, T. J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J. P., Ballin, J., *et al.* (2002) *Genome Res*. **12,** 1749–1755.
7. Albert, T. J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E. F., Stengele, K. P. & Green, R. D. (2003) *Nucleic Acids Res.* **31,** e35.
8. Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., Tongprasit, W., Samanta, M., Weissman, S. M., *et al.* (2004) *Science* **306,** 2242–2246.
9. McDermid, H. E. & Morrow, B. E. (2002) *Am. J. Hum. Genet*. **70,** 1077–1088.
10. Saitta, S. C., Harris, S. E., Gaeth, A. P., Driscoll, D. A., McDonald-McGinn, D. M., Maisenbacher, M. K., Yersak, J. M., Chakraborty, P. K., Hacker, A. M., Zackai, E. H., *et al.* (2004) *Hum. Mol. Genet*. **13,** 417–428.
11. Fine, S. E., Weissman, A., Gerdes, M., Pinto-Martin, J., Zackai, E. H., McDonald-McGinn, D. M. & Emanuel, B. S. (2005) *J. Autism Dev. Disord*. **35,** 461–470.
12. Lindsay, E. A. (2001) *Nat. Rev. Genet*. **2,** 858–868.
13. Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U., Pertz, L. M., Clark, R. A., Schwartz, S., Segraves, R., *et al.* (2005) *Am. J. Hum. Genet*. **77,** 78–88.
14. Emanuel, B. S. & Shaikh, T. H. (2001) *Nat. Rev. Genet*. **2,** 791–800.
15. Selzer, R.R., Richmond, T. A., Pofahl, N. J., Green, R. D., Eis, P. S., Nair, P., Brothman, A. R. & Stallings, R. L. (2005) *Genes Chromosomes Cancer* **44,** 305–319.
16. Bhardwaj, U., Zhang, Y.-H., Lorey, F., McCabe, L. L. & McCabe, E. R. B. (2005) *Am. J. Hematol*. **78,** 249–255.
17. Cheung, J., Wilson, M. D., Zhang, J., Khaja, R., MacDonald, J. R., Heng, H. H. Q., Koop, B. F. & Scherer, S. W. (2003) *Genome Biol*. **4,** R47.
18. Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L. & Emanuel, B. S. (2000) *Hum. Mol. Genet*. **9,** 489–501.
19. Chimpanzee Sequencing and Analysis Consortium (2005) *Nature* **437,** 69–87.
20. Smit, A. F. A., Hubley, R. & Green, P. (1996–2004) REPEATMASKER OPEN (www.repeatmasker.org), Ver. 3.0.
21. Brennan, C., Zhang, Y., Leo, C., Feng, B., Cauwels, C., Aguirre, A. J., Kim, M., Protopopov, A. & Chin, L. (2004) *Cancer Res.* **64,** 4744–4748.
22. Workman, C., Jensen, L. J., Jarmer, H., Berka, R., Gautier, L., Nielsen, H. B., Saxild, H. H., Nielsen, C., Brunak S. & Knudsen, S. (2002) *Genome Biol*. **3,** research0048.
23. States, D. J. & Agarwal, P. (1996) *Proc. Int. Conf. Intell. Syst. Mol. Biol*. **4,** 211–217.
24. Lilleberg, S. & Patel, S. (1998) *Genosys Origins* **1,** 5–6.