Array-based Comparative Genomic Hybridization for the Differential Diagnosis of Renal Cell Cancer

Mónica Wilhelm, Joris A. Veltman, Adam B. Olshen, Ajay N. Jain, Dan H. Moore, Joe C. Presti, Jr., Gyula Kovacs, and Frederic M. Waldman

Abstract

Array-based comparative genomic hybridization (CGH) uses multiple genomic clones arrayed on a slide to detect relative copy number of tumor DNA sequences. Application of array CGH to tumor specimens makes genetic diagnosis of cancers possible and may help to differentiate relevant subsets of tumors, biologically and clinically, which would allow better prognostic and therapeutic decision making. In this study, we have used array-based CGH to detect DNA copy number alterations in distinct types of renal cell carcinomas for diagnostic purposes. We were able to correctly diagnose 33 of 34 malignant tumors by automated computational means and to group together eight benign neoplasms and normal kidney samples. These results indicate that array-based CGH is capable of diagnosing the vast majority of renal cell carcinomas based on their genetic profiles.

Introduction

Array-based CGH is similar to preexisting chromosomal CGH protocols but uses large-insert DNA clones (1–3) or cDNAs (4, 5) as chromosome-specific hybridization targets in an arrayed format instead of normal metaphase chromosomes. Currently, libraries of these large-insert clones have been integrated into the draft sequence of the human genome (6) and, thus, represent ideal targets to be used for array CGH. Copy number alterations as detected by array CGH can be directly related to sequence information and will dramatically accelerate the identification of novel cancer-causing genes. Quantitative measurements of DNA copy number changes by high-resolution array CGH in breast cancer have been successfully used for oncogene and tumor suppressor gene mapping and for the localization and identification of novel tumor-related genes. At present, array CGH in breast cancer has been successfully used for oncogene delineation (3), which demonstrates the efficient use of this technology for the localization and identification of novel tumor-related genes. The purpose of this study was to evaluate the potential of array CGH in differential tumor diagnosis. Renal cancer was chosen as a model system because previous genetic studies have identified specific combinations of genetic changes that are characteristic of subtypes of renal cancer (7–9). The study was performed in a double-blinded fashion, and classification was performed by both manual and automated computational means.

Materials and Methods

Patient Material

Genomic DNA was isolated from 40 renal tumors of distinct genetic types (16 conventional RCCs, 13 papillary RCCs, 5 chromophobe RCCs, and 6 oncocytomas) and 2 normal kidneys (frozen tissue from chromophobe RCCs, short-term cell cultures from other tumor types) by standard protocols. Normal DNA, isolated from lymphocytes of healthy persons, was used as reference for two-color array analysis.

Array-based CGH

The array fabrication and hybridization was performed as described by Pinkel et al. (2) with modifications. Two alternative arrays were used: one consisting of 94 target clones (6 cosmids, 20 BACs, and 67 P1 clones), and the second array consisting of 235 clones (the first 94 plus an additional 67 BACs and 75 P1 clones). Genomic target DNA was isolated from bacterial cultures using QIAgen maxi-kits (QIAgen, Valencia, CA) following the instructions of the manufacturer. Ten μg of DNA were sheared, precipitated, and dissolved in 1 μl of water, followed by the addition of 4 μl of DMSO containing 0.4 μg/μl nitrocellulose. DNA was robotically spotted in quadruplicate onto amino-silane-coated glass slides. Test and reference DNA samples were labeled by nick translation with fluorescein-12-dUTP (DuPont NEN Life Sciences, Boston, MA) and Alexa 568-5-dUTP (Molecular Probes, Eugene, OR) or by random priming with fluorolink cy3-dUTP and cy5-dUTP (Amersham Pharmacia, Piscataway, NJ). Unincorporated fluorescent nucleotides were removed using Sephadex G-50 spin columns. Labeled test and reference DNA samples (0.5–1.0 μg) were mixed with 100 μg of Cot-1 DNA (Life Technologies, Inc., Gaithersburg, MD), precipitated, and resuspended in 30 μl of a hybridization solution containing 50% formamide, 10% dextran sulfate, 2× SSC, and 100 μg of tRNA. The hybridization solution was heated to 70°C for 10 min to denature the DNA, then incubated for 1 h at 37°C to allow blocking of the repetitive sequences. Twenty-seven kidney samples were hybridized to the 94-clone array and 15 renal samples were hybridized to the 235-clone array. Hybridization was performed for 48 h in a moist chamber on a slowly rocking table, followed by a 15-min posthybridization wash in 50% formamide/2× SSC at 45°C, and 10 min in phosphate buffer at room temperature. Slides were mounted in 90% glycerol in phosphate buffer containing 4',6-diamidino-2-phenylindole (DAPI; 0.3 μg/ml). Sixteen-bit fluorescence intensity data were obtained using a charge coupled device camera (Sensys, Photometrics; equipped with a Kodak KAF 1400 chip) coupled to a ×1 optical system, as described by Pinkel et al. (2).

For each set of arrays used in this study, six to eight normal-versus-normal hybridizations were performed to define the normal variation in T:R ratio for each target clone. The average T:R ratio of the quadruplicate of each clone was calculated and divided by the median T:R ratio of all of the targets present on the array to center the T:R values at 1.0. A slight clone-to-clone variability in the intensity ratios was observed (overall coefficient of variation, <10%), and this was reproducible for each target in the normal-versus-normal replicate hybridizations (Fig. 1A). This intrinsic clone-to-clone variability was corrected for by the use of a normal clone and a reference clone.

Received 10/1/01; accepted 1/2/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.


1 The abbreviations used are: CGH, comparative genomic hybridization; RCC, renal cell carcinoma; BAC, bacterial artificial chromosome; T:R, test-reference (ratio).

2 Supported by National Cancer Institute Grant CA47537.

3 The abbreviations used are: CGH, comparative genomic hybridization; RCC, renal cell carcinoma; BAC, bacterial artificial chromosome; T:R, test-reference (ratio).

for by dividing each T:R ratio by the mean T:R ratio of that particular clone in the normal-versus-normal hybridizations. The threshold for gains and losses for each target was calculated as the mean ± 2 × the SD (obtained from the normal-versus-normal hybridizations) from the centered mean of 1.

Computational and Statistical Analysis

Three types of analysis were performed: statistical correlation with tumor subtype, hierarchical clustering, and multiclass pattern recognition.

**Statistical Correlation with Tumor Subtype.** For the subtype correlation, we restricted the analysis to the 47 target clones that exhibited the most substantial deviation from normal across the 42 tumor samples (the top one-half of the clones ranked by the sum of the absolute value of the log relative copy number over all samples). For each class, we computed the rank correlation (Kendall’s Tau) of copy number at each locus with a binary characterization of class (1, in the class; 0, not in the class). Significance thresholds for each correlation were computed based on permutation analysis using 10,000 randomizations of the tumor ID to data mapping. Correlations for each subtype were identified as significant only if their magnitude exceeded 95% of the maximal magnitude correlations from the random trials (thus corresponding to \( P < 0.05 \)).

**Statistical Correlation with Hierarchical Clustering.** Hierarchical clustering of the tumor samples was performed using standard methods, with Euclidean distance as the distance metric (10). We restricted the clones considered in the distance computation to those that were significantly correlated with class.

**Statistical Correlation with Multiclass Pattern Recognition.** We assessed the performance of a fully automatic classification system to assign genetic type to the samples. Performance was quantified using leave-one-out cross-validation (construct a classifier using all but one sample for training; test on that sample; repeat for all samples). We used K-nearest neighbor classification (10), which associates class with a test sample based on the distance of its array CGH profile to the K nearest profiles of training samples. Euclidean distance, with \( K = 3 \), and 20 target clones in the distance computation, yielded the best performance. The 20 clones chosen for each iteration of cross-validation were those with the highest F-statistic in the iteration’s training set. Classification performance was not very sensitive to different combinations of distance metric, number of clones \((K)\), or clone selection method.

Results and Discussion

Clones were selected to represent all autosomal chromosomes and the X chromosome, with the majority of additional targets on those chromosomes most frequently altered in renal cancer (7–9). We tested the performance of the technology by using sex-mismatched normal-versus-normal hybridizations on arrays consisting of 94–235 large-insert clones spotted in quadruplicate (Fig. 1A). On average, 94% of all targets spotted on the different arrays could be analyzed. Spots composed of less than 10 pixels, showing correlations of the two fluorescent dyes below 0.5, or showing autofluorescent particles over the target, were removed as inadequate. The average coefficient of variation of the T:R intensity ratio for the quadruplicate spots of each target was 3%. The SD of intensity ratios was calculated for each target clone in this set of controls. The mean SD for all target clones was 0.04 around a centered mean of 1.0. This normal variation is considerably smaller than that observed in chromosomal CGH (11). Detection of single copy number changes was tested by analyzing the fluorescence intensity ratio of the X chromosome targets in the sex-mismatched hybridizations. Male-versus-female and female-versus-male hybridizations resulted in average X chromosome fluorescence intensity ratios of 0.55 and 1.69, respectively.

To test the utility of array-based CGH in differential diagnosis of
renal cancer, DNA from 42 renal samples was hybridized to these arrays in the presence of normal reference DNA. Statistical analysis was done using the T:R ratios of 91 non-X-chromosome targets analyzed for all of the 42 samples, to identify a subset of target clones the copy number of which was significantly correlated with one of the four tumor subclasses (Fig. 2). In total, 24 of the 91 non-X-chromosome target clones significantly correlated with one of the four subgroups ($P<0.05$). We accounted for the problem of multiple comparisons by using permutation analysis to estimate statistical significance (see “Materials and Methods”). We tested whether relative copy number at these 24 clones, considered together as a vector for each sample, could recapitulate the subclasses using hierarchical clustering. Fig. 3 shows the clustering of the tumor samples using similarity between clone copy-numbers to build the clustering tree. The 34 carcinomas segregated nearly perfectly by type, and the normal samples and benign oncocytomas were grouped together. The
degree to which samples with identical genetic changes cluster together based on copy number suggests that such data may be used for accurate prediction. Systematic classification experiments were then performed with cross-validation to formally estimate the predictive accuracy of automatic class assignment based solely on copy-number measurements. Four classes were used for this analysis (see “Materials and Methods”). Normals and oncocytomas were combined into one class because oncocytomas (benign neoplasms) generally show few karyotypic alterations (8, 9). K-nearest-neighbor classification was used using Euclidean distance. Classification performance was 41/42 (98% correct) as described in detail above. This experiment shows that without previous knowledge of the impact of genetic changes on the diagnosis, a genetic classification is possible when using array-based CGH analysis covering the entire genome.

Subjective interpretation of the array CGH results was also used to classify the renal tumors, independently of automatic methods. The normalized ratios were plotted for all targets (Fig. 1B). Specific patterns of copy number gains and losses were obtained. Genetic alterations were manually scored if the majority of the targets mapped to a specific chromosomal arm or to a whole chromosome showed the same change. Samples were classified on the basis of the Heidelberg classification model (9) without prior knowledge of the diagnosis. Rules for diagnosis were: (a) conventional RCCs were defined by a deletion of chromosome 3p, or a gain of chromosome 5q combined with deletions of at least two of the chromosomes 6q, 8p, 9p, or 14q; (b) papillary RCCs were defined by a gain of at least two of the chromosomes 3q, 7, 8, 12, 16, 17, or 20, and a lack of 3p deletion; (c) chromophobe RCCs were defined by deletions of at least two of the chromosomes 1, 2, 6, 10, 13, or 17; (d) if a genetic profile did not match with any of these three carcinoma groups, or if no genetic changes were observed, the sample was classified as renal oncocytoma/normal/other. Fifteen profiles typical of conventional RCC were identified, each showing the loss of chromosome 3p. Thirteen profiles typical of papillary RCC were seen, each showing a combination of gains at chromosomes 7 and 17. The five genetic profiles typical for chromophobe RCC included loss of at least four chromosomes known to be specific for this type of tumor. An additional nine cases did not match these three renal tumor subgroups and were included in the oncocytoma/normal/other category. Isolated loss of chromosome 14q was seen in two of these cases, whereas loss of chromosome 8 or chromosome 11 was seen in one case each. (these four cases with minimal changes were presumed to be oncocytomas). Four samples showed no copy number changes in any of the targets used (presumed normal or oncocytoma). One additional case (case 45) showed gains of 1p, 4q, 8q, 9q, 10q, 13q, 17p, and loss of chromosomes 3p and 17q. When compared to the other cases, this sample showed the highest level of copy number imbalances by array-based CGH. Recent publications have shown the power of large-scale gene expression analysis for the classification of leukemias and lymphomas (12, 13) or solid tumors (14–16). Genomic analysis is more suitable for diagnostic applications than is expression analysis because DNA is more stable and is more readily available than is RNA. Smaller amounts of DNA can be used for analysis without amplification procedures. In addition, DNA extracted from microdissected archival formalin-fixed and paraffin-embedded tumors can be used for array CGH. In this study the histological heterogeneity of renal carcinoma was used as a model for array-based tumor diagnosis, correctly separating and subtyping 33 of 34 malignant tumors by hybridizations on a limited set of target clones. These results show that DNA array analysis allows a precise differential diagnosis in almost all cases of renal cancer. Additional studies will define whether the genetic diagnosis yields information beyond that represented by histology alone, including alterations associated with clinical outcome and response to therapy. This will be especially useful for tumors with mixed or overlapping morphological patterns. Finally, we demonstrate that automatic, high-resolution, genome-wide screening of copy number changes can become a feasible approach for cancer diagnosis.

Acknowledgments

We thank Sandy DeVries and Beth Gum for their technical support.

References

Array-based Comparative Genomic Hybridization for the Differential Diagnosis of Renal Cell Cancer

Mónica Wilhelm, Joris A. Veltman, Adam B. Olshen, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/4/957

Cited articles
This article cites 15 articles, 6 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/4/957.full.html#ref-list-1

Citing articles
This article has been cited by 31 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/4/957.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.