ORIGINAL ARTICLE

Comparative genomic hybridisation using a proximal 17p BAC/PAC array detects rearrangements responsible for four genomic disorders

C J Shaw, C A Shaw, W Yu, P Stankiewicz, L D White, A L Beaudet, J R Lupski

J Med Genet 2004;**41**:113–119. doi: 10.1136/jmg.2003.012831

See end of article for authors' affiliations

Correspondence to: Dr James R Lupski, One Baylor Plaza, Room 604B, Houston, Texas 77030; USA; jlupski@bcm.tmc.edu

Received in revised form 29 August 2003
Accepted 10 September 2003

Background: Proximal chromosome 17p is a region rich in low copy repeats (LCRs) and prone to chromosomal rearrangements. Four genomic disorders map within the interval 17p11–p12: Charcot–Marie–Tooth disease type 1A, hereditary neuropathy with liability to pressure palsies, Smith–Magenis syndrome, and dup(17)(p11.2p11.2) syndrome. While 80–90% or more of the rearrangements resulting in each disorder are recurrent, several non-recurrent deletions or duplications of varying sizes within proximal 17p also have been characterised using fluorescence in situ hybridisation (FISH).
Methods: A BAC/PAC array based comparative genomic hybridisation (array-CGH) method was tested for its ability to detect these genomic dosage differences and map breakpoints in 25 patients with recurrent and non-recurrent rearrangements.
Results: Array-CGH detected the dosage imbalances resulting from either deletion or duplication in all the samples examined. The array-CGH approach, in combination with a dependent statistical inference method, mapped 45/46 (97.8%) of the analysed breakpoints to within one overlapping BAC/PAC clone, compared with determinations done independently by FISH. Several clones within the array that contained large LCRs did not have an adverse effect on the interpretation of the array-CGH data.
Conclusions: Array-CGH is an accurate and sensitive method for detecting genomic dosage differences and identifying rearrangement breakpoints, even in LCR-rich regions of the genome.

Comparative genomic hybridisation to metaphase chromosomes (metaphase-CGH or mCGH) has become a useful tool for genome wide comprehensive analysis of chromosomal imbalance. mCGH is a relatively rapid method for screening the entire genome, making the technique particularly attractive for the identification of acquired aberrations in tumour tissue and for detecting constitutional rearrangements in prenatal or postnatal samples. Although large chromosomal imbalances are readily detected by mCGH, resolution is limited to gains or losses of 2–10 Mb, rendering it unsuitable in the analysis of small segmental deletions and duplications.[1 2] A more sensitive technique designed for this purpose is microarray based CGH (array-CGH), in which individual BAC/PAC clones in arrays, instead of whole genomes as metaphases, are hybridised with genomic DNA to detect dosage changes with resolution down to 0.2 to 0.4 Mb. The increased resolution of array-CGH allows detection of deletions and duplications of a single BAC/PAC clone (≥50–200 kb).

The ability of array-CGH technology to identify small deletions has been tested previously at the neurofibromatosis type 2 (NF2) locus on chromosome 22q.[3] A critical region for congenital aural atresia also has been defined on chromosome 18q using array-CGH.[4] More recently, array-CGH was used to identify deletions, duplications, and triplications of chromosome 1p36.[5] These studies have proved that it is a reliable and sensitive method with which to detect genomic dosage changes in a single analysis. However, regions of the genome containing low copy repeats (LCRs) have not been investigated in detail. It is unknown whether array-CGH is limited in its ability to detect dosage changes when confounded by LCRs that are present in multiple copies throughout the genome.

Proximal chromosome 17p is an ideal region of the genome to analyse with array-CGH technology. This genomic interval has been extensively characterised; a complete BAC/PAC contig of proximal 17p has been assembled, and finished DNA sequence is available for nearly all clones.[6 7]

Four genomic disorders are caused by deletion or duplication of proximal 17p, and several LCRs have been identified in this region.[6 8 9] Duplication of band 17p12 leads to Charcot–Marie–Tooth disease type 1A (CMT1A), while deletion of the same segment causes hereditary neuropathy with liability to pressure palsies (HNPP).[10–13] LCRs termed CMT1A-REPs are located at the breakpoints of the majority of the deletions and duplications of this segment.[11] Smith–Magenis syndrome (SMS) results from a deletion of 17p11.2, while duplication of the same region leads to dup(17)(p11.2p11.2) syndrome.[8 14–17] Approximately 80–90% of SMS and dup(17)((p11.2p11.2) patients have a deletion/duplication with breakpoints mapping within LCRs termed proximal and distal SMS-REPs (middle SMS-REP maps between them).[16 18–20] In addition to the CMT1A-REPs and SMS-REPs, seven other LCRs (designated LCR17p A to G) have been identified recently in proximal 17p.[9]

We have collected a cohort of patients with varying sizes of deletions/duplications of proximal 17p, many of which have been characterised previously by fluorescence in situ hybridisation (FISH) analysis.[7 9 19 21–23] Of note, in the genomic DNA of these cell lines, the number of CMT1A-REP copies varies from two to six, and the number of SMS-REP copies varies from three to nine. Thus these particular

Abbreviations: array-CGH, microarray based comparative genomic hybridisation; CGH, comparative genomic hybridisation; FISH, fluorescence in situ hybridisation; HMM, hidden Markov model; HNPP, hereditary neuropathy with liability to pressure palsies; LCR, low copy repeat; MAD, median absolute deviation; mCGH, metaphase-CGH; PFGE, pulsed field gel electrophoresis; SMS, Smith–Magenis syndrome

rearrangements provide an ideal tool for testing how LCRs affect the interpretations of array-CGH technology for high resolution genome analysis of patients and normal controls. Furthermore, we have investigated the precision of rearrangement breakpoint mapping by array-CGH.

METHODS

Subjects

We analysed 25 individuals with deletions or duplications of proximal 17p using FISH and array-CGH. These included 15 patients with deletions, nine with duplications, and one with a deletion and a duplication. Control individuals (one male and one female) were unaffected parents of patients with deletions and had normal karyotypes. Peripheral blood samples from patients and family members were obtained after informed consent approved by the Baylor College of Medicine institutional review board.

FISH

FISH analysis of deletion patients was done as described before.[9] Dual colour FISH analysis of duplication patients was done on metaphase and interphase preparations of human peripheral blood lymphocytes and Epstein–Barr virus transformed lymphoblasts according to a modified procedure.[24]

Array-CGH

A minimal tiling path of 56 BAC/PAC clones from the centromere through the CMT1A region on 17p was included in the array, along with 16 control normalisation clones from chromosomes 2, 5, 9, 10, X, and Y. The DNA from BAC and PAC clones was prepared for array spotting as described.[5] Briefly, the DNA was chemically cross linked using (3-glycidoxypropyl)tri-methoxysilane (Sigma) and printed onto glass slides using an OmniGrid Accent microarrayer with Telechem Array II Chipmaker III pins (GeneMachine). Each clone was spotted in quadruplicate. Spotting was done in the Baylor College of Medicine microarray core facility. Patient DNA was isolated from peripheral blood using a Puregene kit (Gentra). The DNA was digested with *Dpn*II restriction enzyme (New England Biolabs) and RNase A (Roche), and then purified using a QIAquick gel extraction kit (Qiagen).

Patient and sex matched control DNA (250 ng) was differentially labelled with cyanine 3-dCTP and cyanine 5-dCTP (Perkin Elmer) using a BioPrime labelling kit (Invitrogen). Each pair of patient and control DNA samples was labelled twice with the dyes reversed and hybridised to the array at 37℃ for 24 hours. The microarray slides were washed at 42℃, scanned using a GenePix microarray scanner, and microarray image quantification was carried out using GLEAMS software (Nutec Sciences). Each BAC/PAC clone position was interrogated eight times; quadruplicate spottings were each examined twice with dye reversal.

Three control versus control hybridisations (two male, one female) were done and showed reproducible normalised $\log_2$(Cy3/Cy5) ratios.

Statistical analysis

Quantified array image files (.tiff) were subjected to single chip normalisation, and dye reversed array pairs were subjected to bi-chip scaling. All analysis was done on $\log_2$ ratios. The justification for single chip normalisation in spotted arrays is well documented.[25 26] The normalisation is used to remove systematic biases such as spatial and intensity artefacts, and bi-chip scaling is used to bring the dye reversed hybridisations to a common measurement scale to facilitate combining the microarray pairs for each patient. The bi-chip scaling factors for chip pairs are motivated by the MAD scaling approach.[26] Briefly, the median absolute deviation about 0 of the by-clone average normalised log-ratio is calculated for each chip in a dye reversed pair; we denote these MAD values as m_i for $i = 1,2$. Single chip scaling factors are then determined as $s^{-1} = m_i/(m_1 * m_2)^{0.5}$. Such scaling factors are well motivated statistically and they have the interpretation of drawing the dye reversed data to the $y = -x$ line. Once scaled, the data are sign changed and averaged to form a single value for each clone for each patient. These dye reversed average data are then used to make inferences regarding the gain/loss status of each clone for each patient. The inference uses a seven state hidden Markov model (HMM), which classifies each clone for each patient into the outcomes gain, loss, and no change (described below). The HMM technique considers the data at adjacent clones as statistically dependent to form an inference at each BAC/PAC locus. The HMM results were evaluated against the previous independent FISH analyses of a subset of the patients with rearrangements.

HMM inference method

The goal of inference in array-CGH is accurate detection of chromosomal change, which avoids false positive calls. An additional goal in our experiment is precise inference for the boundaries of chromosomal lesions, to call the breakpoints. The approach we took to accomplish these goals is to incorporate the adjacency information for the printed BACs/PACs into our statistical inference by use of an HMM.

HMMs are well studied inference methods for analysing dependent data, and these models have seen extensive use in biology in the field of sequence alignment. HMMs always consist of two components: a hidden sequence of unobserved states which are treated as a Markov chain, and a collection of observed emission data. An inference for the sequence of hidden states is formed using the observed emission values.

We developed a seven state HMM to call lesion boundaries and to infer the gain/loss status of each clone in each patient. The hidden states in our CGH array HMM were the gain/no change/loss status of each BAC in each patient. The emission values in our HMM were the observed normalised microarray values for each patient. The seven hidden states in our model are: initiate loss, loss, end loss, no change, initiate gain, gain, end gain. The use of initiate and end states yields better performance for calling lesion boundaries than a simpler model with fewer states. The emission distributions are all univariate Gaussian distributions determined separately for each clone, conditional on its gain/loss status in the FISH data.

We fitted and evaluated our model on 25 patients with known FISH data. To undertake fitting and to obtain inferences in an objective fashion, we took a cross validation approach. For each patient, we performed the model fit using all patients except the one under consideration. We then made inferences on the patient excluded from the fitting process. This step is done for each patient to generate the inference results.

In each fitting step, the HMM transition matrix is directly estimated from the observed transition frequencies in the remaining 24 patients. Emission distributions for each clone are determined by estimating a mean and variance value for each clone, conditional on the FISH outcome status in each patient. For clones in which no patient showed a gain/loss state, the emission distribution is estimated using the average value across all clones for which data were available. For the transition states, the emission mean was estimated to be the mean of the no change mean and the pure loss or gain mean, respectively, for the gain and loss transition states.

Inferences for each patient were made by using the Vitterbi algorithm, which is the well studied method for obtaining the highest probability path conditional on the observations and the model. A key feature is the ability of the HMM to make

correct inferences even in regions wherein data show high variance and might otherwise lead to mistaken conclusions.

RESULTS

Array-CGH validation of rearrangements previously mapped by FISH and PFGE

To test the reliability and accuracy of the array-CGH technology, we hybridised the DNA of 12 deletion patients and one duplication patient whose unusual rearrangements were previously characterised by FISH (fig 1).[9 23] In addition, three patients with common SMS deletions (1780, 1949, 1957), two patients with common dup(17)((p11.2p11.2) duplications (1789, 1913), two patients with CMT1A duplications (682, 723), and one patient with a common dup(17)((p11.2p11.2) duplication and a common HNPP deletion (1006) were analysed. All eight of these patients were documented previously to have a common deletion/duplication by the presence of a rearrangement specific pulsed field gel electrophoresis (PFGE) junction fragment or FISH analysis.[8 9 11 17 27] Those patients with common rearrangements analysed by PFGE only are assumed to have breakpoints mapping within the proximal and distal SMS-REPs (in the case of 17p11.2 deletions/duplications) or within the proximal and distal CMT1A-REPs (in the case of 17p12 deletions/duplications). Control individuals were hybridised to the array, and none of the clones was inferred as deleted or duplicated using the HMM analysis (fig 1A).

The array-CGH analysis was accurate in detecting clones displaying a gain or loss in each patient tested (figs 1–3). Importantly, both deletions (fig 1B) and duplications (fig 1C) were readily detected. Interestingly, in a rare patient with both a deletion and a duplication of 17p,[22] both rearrangements were easily discerned (fig 1D). As expected, all three SMS patients with common deletions revealed similar patterns in the array-CGH statistical plots (fig 2).

The deletion/duplication breakpoints detected using array-CGH were consistent with those previously mapped by FISH and PFGE (fig 3). The distal deletion breakpoint of patient 993, previously unmapped by FISH, was identified by array-CGH and subsequently confirmed by FISH. This breakpoint was mapped between clones RP11-64B12 and RP11-849N15 (figs 3A, 4A). The distal deletion breakpoint of patient 357 was confirmed to extend distally beyond clone RP11-350B3 (figs 3A, 4B).

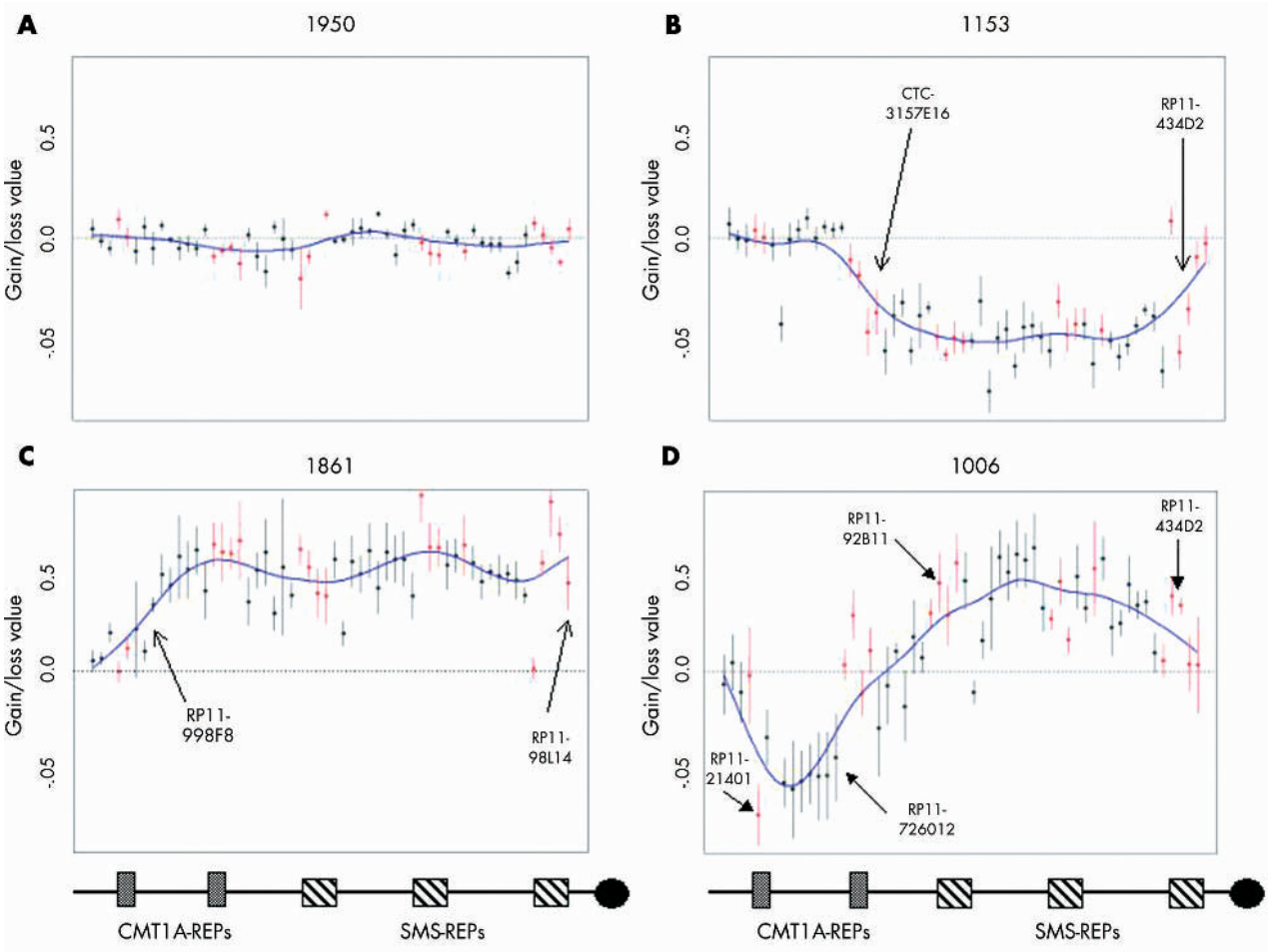

Figure 1 Plots of array-CGH data. The 17p clones are ordered from telomere (left) to centromere (right). Individual clone mean normalised $\log_2$(Cy3/Cy5) ratios of patient to control are plotted on the Y axis and represented by dots with error bars; clones containing LCRs >20 kb are highlighted in red. A blue smoothing spline is drawn between the points. Breakpoint clones are labelled and denoted with arrows. A diagram of proximal 17p is shown at the bottom; the black circle represents the centromere, the diagonally striped boxes represent the SMS-REPs, and the dotted boxes represent the CMT1A-REPs. (A) 1950; female control, showing no dosage gain or loss. (B) 1153; large SMS deletion from CTC-3157E16 through RP11-434D2. (C) 1861; large duplication from RP11-998F8 through RP11-98L14. (D) 1006; common dup(17)(p11.2p11.2) from RP11-92B11 through RP11-434D2 and HNPP deletion from RP11-214O1 through RP11-726O12. Note: these plots are intended only for visualisation purposes, and do not represent the final inferences based on the HMM analysis. All inferences presented in the results section are determined according to the seven state HMM described in Methods.

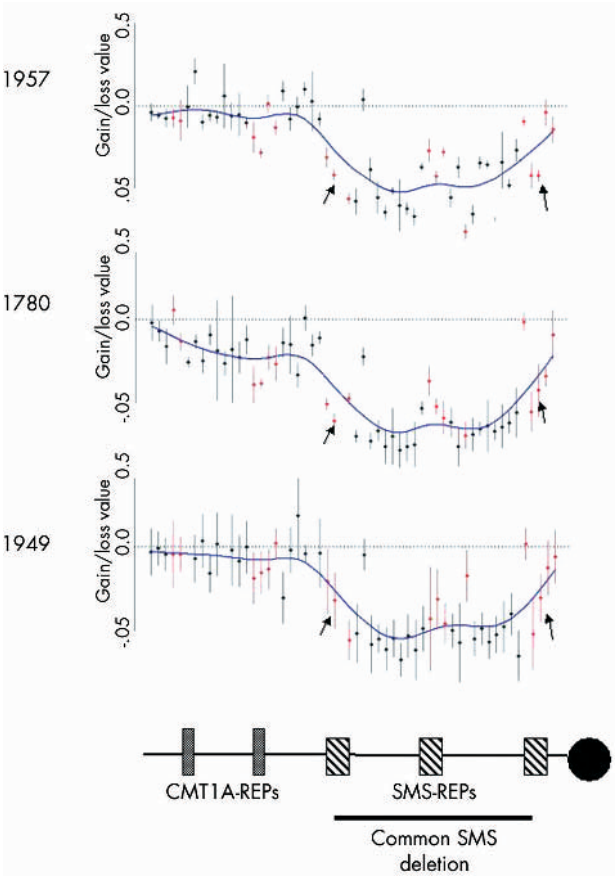

Figure 2 Statistical plots of array-CGH data on three patients with the common SMS deletion. The 17p clones are ordered from telomere (left) to centromere (right). Individual clone mean normalised $\log_2$(Cy3/Cy5) ratios of patient to control are plotted on the Y axis and represented by dots with error bars; clones containing LCRs >20 kb are highlighted in red. A blue smoothing spline is drawn between the points. Breakpoint clones are labelled and denoted with arrows. A diagram of proximal 17p is shown at the bottom; the black circle represents the centromere, the diagonally striped boxes represent the SMS-REPs, and the dotted boxes represent the CMT1A-REPs. The arrows denote deleted clones RP11-434D2 (right side), within proximal SMS-REP, through RP11-219A15 (left side), within distal SMS-REP.

FISH validation of array-CGH data on duplications of unknown size

To test the precision of the array-CGH technology further, DNA from four patients (527, 563, 1229, 1458) with large duplications of proximal 17p—which have not been characterised by FISH—were hybridised to the array (fig 3B). After tentative assignment of the duplication breakpoints based on the array data, FISH was undertaken with BAC/PAC clones flanking each prospective breakpoint. The array-CGH data indicated that patients 527 and 1458 were duplicated for all chromosome 17 clones in the array, from the most proximal clone through the CMT1A region (fig 3B). FISH was done on these cell lines with clones RP11-98L14 and RP11-350B3, both of which showed three signals in interphase cells (fig 4C, 4D). Based on the array data, patient 563 is duplicated from RP11-98L14 through RP11-726O12, within the CMT1A region. FISH done with both clones confirmed the array data (fig 4E). Array-CGH results indicated that patient 1229 was duplicated from RP11-98L14 through RP11-849N15, the clone that contains the *PMP*22 gene. FISH with these clones confirmed the array data (fig 4F).

Accuracy of array-CGH breakpoint mapping

The rearrangement breakpoints were accurately predicted by array-CGH, assigning 45/46 breakpoints (97.8%) correctly to within one overlapping adjacent clone of the breakpoint identified by FISH, and 100% to within two clones (fig 3). Previous FISH analysis of some of the clones spotted on the array showed a weak signal, indicating that those clones were partially deleted.[9] However, as it is unknown what portion of the clone must be deleted to show a loss using array-CGH, we considered either a no change or loss inference to be correct for clones shown to be partially deleted by FISH. Six of the cases analysed had one or both breakpoints located proximally or distally of the clones contained in the array (patients 357, 527, 563, 1229, 1458, and 1861). All eight of these breakpoints were correctly inferred by array-CGH to extend beyond the clones contained in the array (fig 3).

While the dosage changes and breakpoints were correctly inferred, a few clones did not perform as well as the majority when analysed independently (figs 1 and 2). RP1-836L9 did not show a gain or loss in any of the patients tested, indicating that the quality of the PAC DNA printed in the array was suboptimal for hybridisation. Likewise, in several cases, CTD-124H2, RP1-37N7, and RP11-48J14 appeared to have no dosage change, while they were shown to be deleted by FISH. Although these small inconsistencies are apparent in the plots of the statistical mean of the $\log_2$(Cy3/Cy5) fluorescence ratio (figs 1 and 2), they had no effect on the final inference (fig 3).

DISCUSSION

We have examined the capabilities of array-CGH to detect recurrent rearrangements (CMT1A duplication/HNPP deletion, dup(17)(p11.2p11.2)/SMS deletion) and map the breakpoints of unique non-recurrent rearrangements in proximal 17p, an LCR-rich genomic interval. Array-CGH allowed the detection of losses from deletion and gains from duplication. Furthermore, breakpoints for both recurrent and uniquely sized deletions and duplications were readily discerned using our inference analysis. Our study shows that array-CGH can detect dosage differences of 17p in patients compared with normal controls despite challenges introduced by complex genome architecture.

Comparison of array-CGH and FISH

While array-CGH has proved to be an accurate method for detecting genomic dosage change and mapping breakpoints, it remains to be determined whether array-CGH or FISH is a more sensitive technique. Previous work has estimated the resolution of array-CGH to be as low as 40 kb when using BAC/PAC clones in the array,[3] while the resolution of FISH is known to be as low as 1000 base pairs (bp) when cosmids and polymerase chain reaction (PCR) products are used as probes. However, finer resolution is probably feasible with array-CGH if smaller segments of DNA such as cosmids or PCR products are used to construct the array.

Array-CGH is a much more rapid and higher throughput technique than FISH, with the ability to test for hundreds or thousands of loci in a single analysis. Array-CGH is an exceptional tool for whole genome screening of dosage imbalances, some of which may remain undetected by standard disease specific FISH analysis. This technique has proved sensitive enough to detect triplications in chromosome 1p,[5] although the sensitivity threshold of array-CGH in mosaic cell lines is yet to be thoroughly investigated. Recently, array-CGH was shown to be sufficiently sensitive to detect duplication of all clones contained on a ring chromosome 18q that was present in 75% of cells.[4] However, the same study showed array-CGH was not sensitive enough to detect clones that were deleted in a mosaic cell line

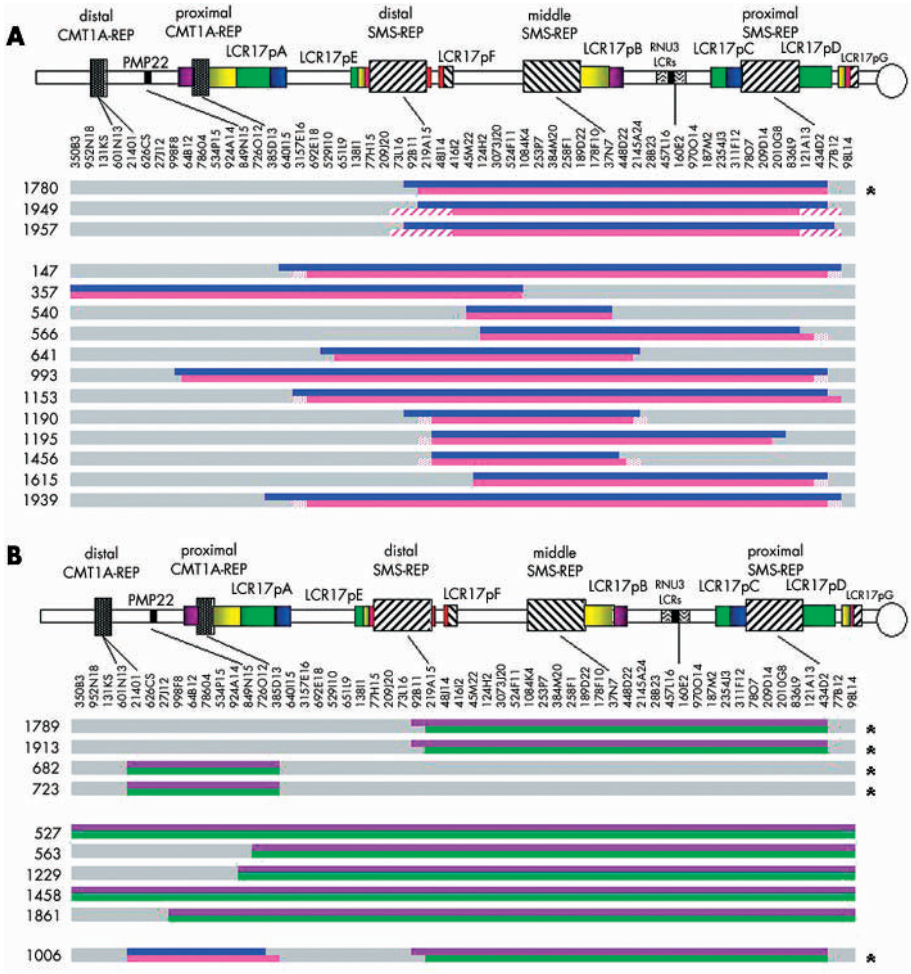

Figure 3 Summary of array-CGH and fluorescence in situ hybridisation (FISH) data. A schematic representation of proximal chromosome 17p, and the order of clones included in the array are shown at the top. Patient numbers are given on the left. Asterisks indicate patients whose breakpoints are assumed based on the presence of a rearrangement specific junction fragment when analysed by pulsed field gel electrophoresis. The grey bars represent clones that do not have a dosage imbalance. (A) Deletion analysis. The blue bars indicate clones that are deleted based on array-CGH data. The pink bars represent clones that were deleted based on FISH analysis. Diagonal stripes indicate clones that were not used as FISH probes; dotted boxes indicate clones that showed a partial signal when used as FISH probes. (B) Duplication analysis. The purple bars indicate clones that are duplicated based on array-CGH data. The green bars represent clones that were duplicated based on FISH analysis. Patient 1006 has both a deletion and a duplication.

carrying an 18q deletion in 33% of cells. This suggests that rearrangements present in less than 75% of cells may go undetected by array-CGH, while these cases are readily identified by classical cytogenetics and FISH. In addition, array-CGH is unable to detect balanced translocations. Thus it seems that although array-CGH is revolutionising the investigation of chromosomal rearrangements, there is still a great need for classical cytogenetics and FISH studies. For FISH studies, a region of suspected abnormality must be chosen for study, while array-CGH offers the potential to test for very large numbers of loci at once. From a clinical application standpoint, the array-CGH can enable a simultaneous high resolution analysis of the entire human genome. Abnormalities identified by such a screening approach could be confirmed by a locus specific FISH test.

Methods of array-CGH analysis

Different methods of analysis are applicable to array-CGH data, depending on the information desired from the experiment. An independent analysis method considers each clone separately when assigning a gain/loss/no change state. This method is accurate for predicting small interstitial changes, as the states of the clones near the breakpoint are not dependent on the states of the adjacent overlapping clones. However, individual clones may be assigned states that are inconsistent with adjacent clones, resulting in a deletion/duplication that appears non-contiguous. Although a handful of clones may perform less well than the majority when analysed using an independent method of analysis, these small inconsistencies vanish when a dependent analysis is implemented. This analysis method considers the state (gain/loss/no change) of the adjacent, overlapping clones and the number of consistently assigned clones when assigning a state to any particular clone. Thus a clone that has an inconsistent state when analysed independently will be consistent with adjacent, overlapping clones when analysed dependently. This method works well when information regarding contiguity of the deletion/duplication is desired, but may sometimes be less accurate at predicting the breakpoints of the rearrangement. The addition of intermediate states—begin and end states—facilitates the accurate call of breakpoints. Additionally, important information regarding the physical characteristics of particular clones may be concealed when analysed dependently. A

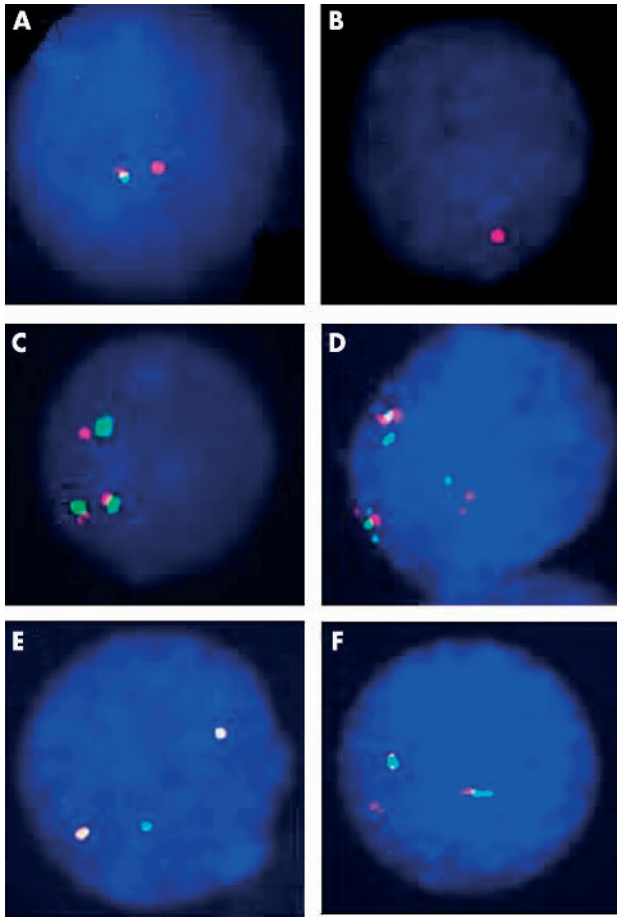

Figure 4 Fluorescence in situ hybridisation (FISH) mapping of breakpoints. (A) Patient 993; RP11-998F8 (red) shows two signals, while RP11-64B12 (green) shows one signal, indicating the deletion breakpoint maps between them. (B) Patient 357; RP11-350B3 shows one signal, indicating the deletion breakpoint is distal to this clone. (C) Patient 527; RP11-350B3 (green) and RP11-98L14 (red) show three signals, indicating the duplication encompasses all clones in the array, and may extend proximally or distally. (D) Patient 1458; RP11-350B3 (green) and RP11-98L14 (red) both have three hybridisation signals, showing the duplication extends beyond the clones within the array. This is a G2 cell, and thus each signal is present in duplicate. (E) Patient 563; two hybridisation signals are present for RP11-849N15 (red), while three are present for RP11-726O12 (green; two of which are superimposed on the red signals and appear yellow), indicating the duplication breakpoint is between them. (F) Patient 1229; RP11-849N15 (red) gives three signals, while RP11-924A14 (green) gives two, suggesting the duplication breakpoint maps between them.

dependent mode of analysis would be helpful in the construction of a clinical microarray designed to detect common deletions/duplications throughout the genome.

The method of analysis used in this paper merges the benefits of both independent and dependent analysis by assigning not only gain/loss/no change states to the clones, but also intermediate states. Thus transitions between a gain or loss state and a no change state (which occur at the breakpoints of the rearrangements) are more accurately predicted. In this report, 97.8% of the breakpoints identified using array-CGH matched those detected using FISH or PFGE (to within one adjacent clone on either side), while 100% fell within two adjacent clones when analysed using this method.

Performance of clones containing LCRs

The major advantage of array-CGH is to screen large numbers of clones at once to quickly focus on breakpoints and detect major regions of dosage imbalance. However, several of the most common deletion/duplication syndromes are associated with large LCRs,[28] which were anticipated possibly to introduce complications in the analysis of array-CGH data. Although BACs and PACs containing large LCRs could present analysis challenges (that is, gain/loss/no change inferences that are inconsistent with FISH data) owing to their homology with other clones, our data show these clones do not have a negative impact on the inference analysis. However, a few clones (RP1-836L9, CTC-124H2, RP1-37N7, RP11-48J14, and RP1-27J12) were observed to perform less well than the majority when the standard mean of the $\log_2$(Cy3/Cy5) fluorescence ratio was calculated (that is, the deviation from ''no change'' was less than expected; figs 1 and 2). While three of these clones contain LCRs greater than 20 kb (RP1-836L9, RP1-37N7, and RP11-48J14), the remaining two do not. In addition, a trend was observed for the clones adjacent to the middle SMS-REP (contained within clone RP1-37N7), showing the $\log_2$(Cy3/Cy5) fluorescence ratio to be smaller for those clones when some deletion patients were analysed (fig 2). The reason for the suboptimal performance of these clones is not obvious, although it may reflect underlying genomic architecture. While such clones did not affect the final interpretation of the array-CGH data, they may introduce difficulties in some studies and thus it is important to identify such clones for subsequent array-CGH experiments in any given genomic region.

Our data suggest that although clones containing LCRs are highly homologous to other regions of the genome, they must also contain sufficient unique sequence to hybridise specifically with the corresponding segment of genomic DNA. A comparison of the average variance (across 25 cases) of the $\log_2$(Cy3/Cy5) fluorescence ratio (patient/control) for each clone showed that BACs and PACs containing LCRs did not differ from unique sequence clones in this aspect. This observation is expected, as the LCR containing and unique sequence clones perform similarly in the experiment. A general trend showed that clones within the CMT1A region had a lower average variance value than clones within the SMS region, perhaps reflecting the increased size and number of LCRs in the SMS region. Thus these data show that array-CGH is an efficient method with which to detect genomic dosage change and map rearrangement breakpoints, even in regions of the genome laden with LCRs.

ACKNOWLEDGEMENTS

We thank the patients for their participation. We also thank Marjorie Withers for excellent technical assistance. This study was supported in part by grants from the National Institute of Child Health and Human Development (PO1 HD39420) and the Mental Retardation Research Center (HD24064).

Authors' affiliations
C J Shaw, C A Shaw, W Yu, P Stankiewicz, L D White, A L Beaudet, J R Lupski, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA

REFERENCES

1 **Bentz M**, Plesch A, Stilgenbauer S, Döhner H, Lichter P. Minimal sizes of deletions detected by comparative genomic hybridization. *Genes Chromosomes Cancer* 1998;**21**:172–5.
2 **Kirchhoff M**, Gerdes T, Maahr J, Rose H, Bentz M, Döhner H, Lundsteen C. Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. *Genes Chromosomes Cancer* 1999;**25**:410–13.
3 **Bruder CEG**, Hirvelä C, Tapia-Paez I, Fransson I, Segraves R, Hamilton G, Zhang XX, Evans DG, Wallace AJ, Baser ME, *et al*. High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum Mol Genet* 2001;**10**:271–82.
4 **Veltman JA**, Jonkers Y, Nuijten I, Janssen I, van der Vliet W, Huys E, Vermeesch J, Van Buggenhout G, Fryns J-P, Admiraal R, Terhal P, Lacombe D, van Kessel AG, Smeets D, Schoenmakers EFPM, van Ravenswaaij-Arts CM.

Definition of a critical region on chromosome 18 for congenital aural atresia by arrayCGH. *Am J Hum Genet* 2003;**72**:1578–84.
5 **Yu W**, Ballif BC, Kashork CD, Heilstedt HA, Howard LA, Cai W-W, White LD, Liu W, Beaudet AL, Bejjani BA, Shaw CA, Shaffer LG. Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. *Hum Mol Genet* 2003;**12**:2145–52.
6 **Inoue K**, Dewar K, Katsanis N, Reiter LT, Lander ES, Devon KL, Wyman DW, Lupski JR, Birren B. The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res* 2001;**11**:1018–33.
7 **Bi W**, Yan J, Stankiewicz P, Park S-S, Walz K, Boerkoel CF, Potocki L, Shaffer LG, Devriendt K, Nowaczyk MJM, Inoue K, Lupski JR. Genes in a refined Smith-Magenis syndrome critical deletion interval on chromosome 17p11.2 and the syntenic region of the mouse. *Genome Res* 2002;**12**:713–28.
8 **Chen K-S**, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Genet* 1997;**17**:154–63.
9 **Stankiewicz P**, Shaw CJ, Dapper JD, Wakui K, Shaffer LG, Withers M, Elizondo L, Park S-S, Lupski JR. Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am J Hum Genet* 2003;**72**:1101–16.
10 **Lupski JR**, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, Barker DF, Killian JM, Garcia CA, Chakravarti A, Patel PI. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 1991;**66**:219–32.
11 **Pentao L**, Wise CA, Chinault AC, Patel PI, Lupski JR. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat Genet* 1992;**2**:292–300.
12 **Chance PF**, Alderson MK, Leppig KA, Lensch MW, Matsunami N, Smith B, Swanson PD, Odelberg SJ, Disteche CM, Bird TD. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* 1993;**72**:143–51.
13 **Chance PF**, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum Mol Genet* 1994;**3**:223–8.
14 **Smith ACM**, McGavran L, Robinson J, Waldstein G, Macfarlane J, Zonona J, Reiss J, Lahr M, Allen L, Magenis E. Interstitial deletion of (17)(p11.2p11.2) in nine patients. *Am J Med Genet* 1986;**24**:393–414.
15 **Stratton RF**, Dobyns WB, Greenberg F, DeSana JB, Moore C, Fidone G, Runge GH, Feldman P, Sekhon GS, Pauli RM, Ledbetter DH. Interstitial deletion of (17)(p11.2p11. 2): report of six additional patients with a new chromosome deletion syndrome, *Am J Med Genet* 1986;**24**:421–32.
16 **Greenberg F**, Guzzetta V, de Oca-Luna MR, Magenis RE, Smith ACM, Richter SF, Kondo I, Dobyns WB, Patel PI, Lupski JR. Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). *Am J Hum Genet* 1991;**49**:1207–18.
17 **Potocki L**, Chen K-S, Park S-S, Osterholm DE, Withers MA, Kimonis V, Summers AM, Meschino WS, Anyane-Yeboa K, Kashork CD, Shaffer LG, Lupski JR. Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat Genet* 2000;**24**:84–7.
18 **Guzzetta V**, Franco B, Trask BJ, Zhang H, Saucedo-Cardenas O, Montes de Oca-Luna R, Greenberg F, Chinault AC, Lupski JR, Patel PI. Somatic cell hybrids, sequence-tagged sites, simple repeat polymorphisms, and yeast artificial chromosomes for physical and genetic mapping of proximal 17p. *Genomics* 1992;**13**:551–9.
19 **Juyal RC**, Figuera LE, Hauge X, Elsea SH, Lupski JR, Greenberg F, Baldini A, Patel PI. Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients. *Am J Hum Genet* 1996;**58**:998–1007.
20 **Park S-S**, Stankiewicz P, Bi W, Shaw C, Lehoczky J, Dewar K, Birren B, Lupski JR. Structure and evolution of the Smith-Magenis syndrome repeat gene clusters, SMS-REPs. *Genome Res* 2002;**12**:729–38.
21 **Elsea SH**, Purandare SM, Adell RA, Juyal RC, Davis JG, Finucane B, Magenis RE, Patel PI. Definition of the critical interval for Smith-Magenis syndrome. *Cytogenet Cell Genet* 1997;**79**:276–81. Erratum in: *Cytogenet Cell Genet* 1998;**81**:67.
22 **Potocki L**, Chen K-S, Koeuth T, Killian J, Iannaccone ST, Shapira SK, Kashork CD, Spikes AS, Shaffer LG, Lupski JR. DNA rearrangements on both homologues of chromosome 17 in a mildly delayed individual with a family history of autosomal dominant carpal tunnel syndrome. *Am J Hum Genet* 1999;**64**:471–8.
23 **Shaw CJ**, Stankiewicz P, Christodoulou J, Smith E, Jones K, Lupski JR. A girl with duplication 17p10–p12 associated with a dicentric chromosome. *Am J Med Genet* 2003; DOI 10.1002/ajmg.a.20355.
24 **Shaffer LG**, Kennedy GM, Spikes AS, Lupski JR. Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: implications for testing in the cytogenetics laboratory. *Am J Med Genet* 1997;**69**:325–31.
25 **Van Driessche N**, Shaw C, Katoh M, Morio T, Sucgang R, Ibarra M, Kuwayama H, Saito T, Urushihara H, Maeda M, Takeuchi I, Ochiai H, Eaton W, Tollett J, Halter J, Kuspa A, Tanaka Y, Shaulsky G. A transcriptional profile of multicellular development in Dictyostelium discoideum. *Development* 2002;**129**:1543–52.
26 **Yang YH**, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;**30**:e15.
27 **Wise CA**, Garcia CA, Davis SN, Heju Z, Pentao L, Patel PI, Lupski JR. Molecular analyses of unrelated Charcot-Marie-Tooth (CMT) disease patients suggest a high frequency of the CMT1A duplication. *Am J Hum Genet* 1993;**53**:853–63.
28 **Stankiewicz P**, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002;**18**:74–82.