The human genome; you gain some, you lose some

Kriek, M.

Chapter III-4

Comparison of four genome wide platforms using four overlapping interstitial 2p alterations

Marjolein Kriek[1], Claudia A. L. Ruivenkamp[1], Yavuz Ariyurek[1], Margot E. Kalf[1], Jeroen Knijnenburg[2], Arie van Haeringen[1], Emanuela Lucci-Cordisco[3], Maurizio Genuardi[4], Carla Rosenberg[5], Ana C. Krepischi-Santos[5], S.R. Sanders[6], Stefan J. White[1], K. Szuhai[2], Martijn H. Breuning[1], Johan T. den Dunnen[1]

[1]Center for Human and Clinical Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands
[2]Dept. Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands
[3]EL-C: Institute of Medical Genetics, Catholic University "A. Gemelli" School of Medicine, Rome, Italy
[4]MG: Medical Genetics Unit, Department of Clinical Pathophysiology, University of Florence Medical School, Florence, Italy
[5]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.
[6]Section of Genetics & Metabolism, Health Science Centre, Winnipeg, Canada

Submitted

Abstract

Molecular karyotyping by array-based techniques represents a giant leap forward compared to microscopic metaphase banding. We compared the performance of four different array based platforms to identify and map the breakpoints in four patients with different interstitial 2p deletions, all localised within 2p16.1-p21. Currently, there are two main array formats, array-CGH and SNP-based. For array-CGH the probes used are (3K–32K) genomic clones or up to 244K oligonucleotides, with the size and number determining the resolution of analysis. SNP arrays, containing 10K–1000K loci have proven to facilitate, in addition to genome-wide association studies, the detection of deletions and duplications. The resolution of these arrays depends on the number of SNP loci present and on their distribution across the genome.

In this study, the platforms used include a 3K large genomic insert clone array, a 44K (long oligo) microarray and two SNP- based arrays (250-500K, and 317K). Our analysis showed that the size of the 2p deletions varied, from ~10.6Mb in patient 1, to ~2.4Mb in patient 4. The minimum region of overlap of the deletions was ~1.3Mb encompassing 8 genes. The *MSH6* gene was deleted in minimally three out of four patients, indicating that they have a 60-90% chance of developing colon carcinoma. No clear genotype/phenotype correlation emerged from the comparison of the four patients.

Comparing cross-platform the breakpoint mapping gave similar results in the majority of cases.

Introduction

For decades trypsin Giemsa banding of metaphase spreads has been the standard diagnostic method to detect chromosomal rearrangements. The method has several advantages; all chromosomes are seen under the microscope, and individual cells can be karyotyped, permitting clonal analysis[1,2] and the study of mosaicism. A major limitation is the fact that due to the contraction of chromosomes during metaphase and the resolution of the light microscope, G banding is not capable of identifying rearrangements smaller than 3-5 Mb.

Fluorescence in Situ Hybridization (FISH)[3,4] partly overcomes this problem, allowing direct testing for the presence, absence or amplification of specific genomic regions. This method is especially used for the confirmation of microdeletion syndromes and the analysis of potential subtelomeric rearrangements. FISH analysis can also be used for the detection of mosaicism to a very low level, depending on the number of cells analysed. However, it has several drawbacks, as detecting rearrangements using FISH analysis is only possible when cells are available, an obvious, specific phenotype is present that is recognized by a specialist, and when a specific FISH probe exists. Finally, although multi-colour methodologies have been developed[5,6,7] the number of loci that can be analysed simultaneously is limited.

Recently, array-based technologies have been developed that provide both genome-wide and high resolution analysis. In contrast to FISH, where fragments of DNA are labeled and hybridized to chromosome spreads, array-based approaches label genomic DNA, which is then hybridized to DNA spotted on a solid support, typically a glass slide. The size of the DNA probe and the number of probes on the array determine the resolution of analysis.

The first arrays used relatively large DNA fragments (~150 kb) isolated from Bacterial Artificial Chromosome (BAC) or P1 derived Artificial Chromosome (PAC) clones.[8-10] A newer format uses oligonucleotide probes of 25 to 60 nt in length.[11,12] Due to the smaller size of these probes and the much larger number of loci analysed, it is possible to detect much smaller copy number variations (CNVs) with greater precision compared to those that can be revealed using BAC-PAC clone arrays. The 25-mer probe arrays were originally designed for SNP analysis. However, they were quickly used to estimate copy number changes by using both signal strength and allele scoring. Initial studies used the Affymetrix 10K array, which demonstrated the principle that the arrays could provide quantitative data.[13] Subsequent work has taken advantage of

higher resolution chips, currently up to 500-1000K.[14] In practice, these arrays have an effective resolution below 10 kilobases. However, despite their extremely high resolution, it should be noted that these tools can not be used to detect copy neutral rearrangements like translocations and inversions.

In this study, we have analysed four patients with different sizes of interstitial 2p deletions, all localised within the chromosome region 2p16.1-p21. We have compared different platforms for identifying the deletions as well as their ability to define breakpoints. In addition, we have collated and compared the clinical data of these patients. It appears that psycho-motor delay is the only common clinical feature that corresponds to a deletion within this area.

Methods

Array- Comparative Genomic Hybridisation (Array-CGH)

The array-CGH procedures were performed as previously described.[15] The clones were provided by the Wellcome Trust Sanger Institute (UK), and information regarding the full set is available at the Ensembl web site.

The array contained ~3500 large genomic insert clones spaced at ~1 Mb intervals over the genome, meaning that the resolution of the array varies between 0.2-3 Mb. Profiles were displayed by using the Log(2) ratio of test and reference sample. The thresholds were set at -0.3 and 0.3. BACs with a Log(2) ratio outside this interval were considered to be altered.[15,16]

Agilent microarray

Agilent Human Genome CGH Microarrays consist of ~44,000 60-mer oligonucleotide gene focused probes that span coding and non-coding sequences with an average spatial resolution of ~35 kb. Both genes with known function and hypothetical genes were included in the array. We used a loop-hybridisation design to analyse six DNA samples, including three patients with an interstitial 2p deletion. In a loop hybridisation design DNA sample 1 and 2 are differently labelled (Cy5 and Cy3, respectively) and subsequently hybridised on the same array. The second array includes DNA sample 2 and 3 that are labelled using Cy5 and Cy3, respectively. In this way, one sample is hybridised twice on an independent array in two different colors (= dye swap), enabling its own internal quality control. Arrays were hybridised according to the recom-

mendations of the supplier (www.agilent.com). Data were analysed using the Agilent CGH Analytics 3.4 software with a moving averaging window of 1Mb. The size of the three different deletions was calculated using the Log(2) ratio. The thresholds were set at -0.3 and 0.3.

Affymetrix 500K Genechip

The Genechip Human Mapping 500K array set was used. The procedure was performed as described in the Affymetrix GeneChip Human Mapping 500K Manual (http://www.affymetrix.com). The set comprises two SNP arrays contain ~250.000 25-mer oligonucleotides each. Using this protocol, the human genome is cut by restriction enzymes (NspI and StyI); one restriction enzyme is used per array, enabling the analysis of 250,000 loci. The use of a second restriction enzyme is necessary for the analysis of another 250,000 loci. For data analysis, DNA-Chip Analyzer (dChip) software (version release 02-16-06) was used.[17,18] Regions of copy number gain and loss were detected using the hidden Markov model output of dChip. The thresholds for this platform were set between 1.6 and 2.4 using a linear scale, in where 2.0 represents two copies of a given locus.[11,19]

Illumina 317K beadchip

The Illumina humanhap 317K genotyping beadchip work up has been performed as suggested by the manufacturer (www.illumina.com). The SNP array consists of 317,000 25-mer oligonucleotide probes. For data analysis, the beadstudio data analysis software provided by Illumina was used.

In this platform, the regions for CNVs are detected based on the LogR ratio. This tool combines data of both heterozygosity (SNP call) and signal strength. The thresholds were set at -0.3 and +0.3. In addition to the Log R ratio, the data analysis software also provides B allele frequency, Loss of Heterozygosity (LOH) and Copy number (CN) score.

Patient samples

The four patient samples were gathered from the Netherlands (patient 1), Canada (patient 2), Brazil (patient 3) and Italy (patient 4), respectively. Two of them have been described previously.[20,21]

The DNA of the patients was applied to each platform once, except for the Agilent array (due to the dye swap procedure). This study was approved by the Institutional Review Board of the Leiden University Medical Center, conforming with Dutch law.

Table 1. Overview of the BAC array results.

Intnl Clone name	Chrom.	Patient 1	Patient 2	Patient 3	Patient 4
RP11-204D19	2				
RP11-24I5	2				
RP11-421J10	2				
RP11-27C22	2				
RP11-110G2	2				
RP11-1084a21	2				
RP11-436K12	2				
RP5-960D23	2				
RP11-19A8	2				
RP11-436L21b	2				
RP11-436L21	2				
RP11-460M2	2				
RP11-319N5	2				
RP11-5M9	2				
RP11-391D19	2				
RP11-389K20	2				
RP11-335O22	2				
RP11-7H13	2				
RP11-508L23	2				
RP11-30C22	2				

Minimum region of overlap

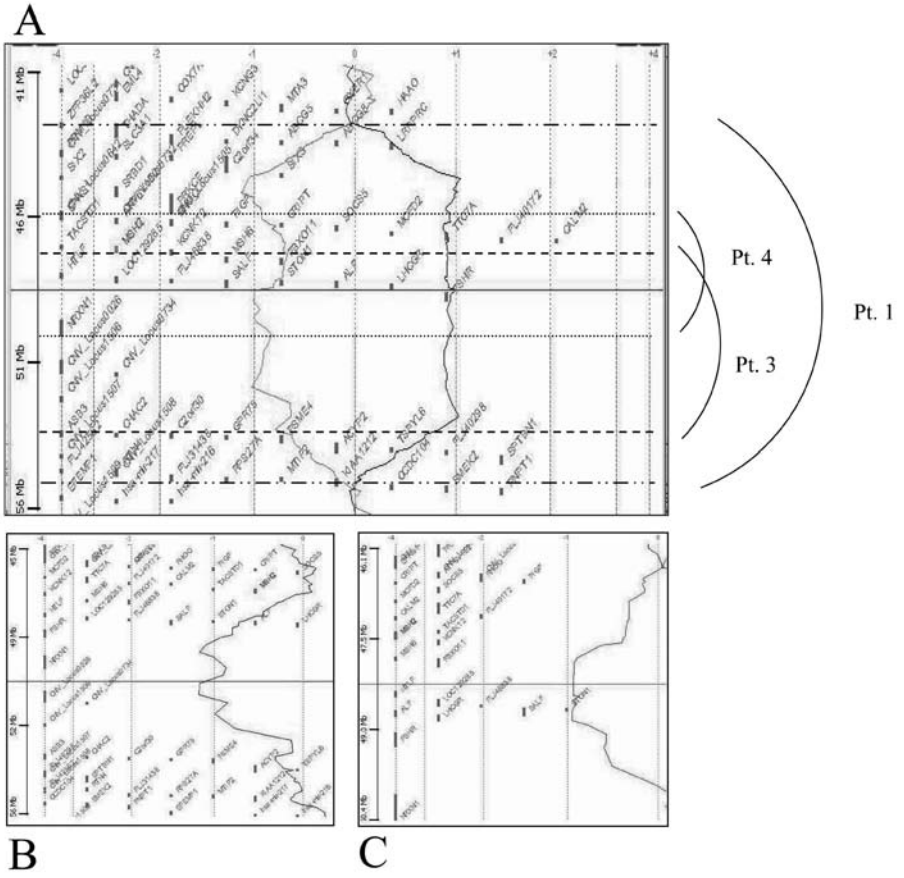
Deleted BACs per patient, depicted as grey bars.

Results

Initial chromosome analysis of patient 1 did not reveal any abnormalities. However, by using both Multiplex Amplifiable Probe Hybridisation (MAPH)[22] and 3K array-CGH, it was possible to identify a deletion of chromosome region 2p16.2-p21. Retrospective analysis of the karyogram (G-banding) did detect the interstitial 2p deletion. The banding pattern of the short arm of chromosome 2 of this patient was compared to that of the previously described patient with a deletion within this region,[20] and was found to be similar (data not shown).

To study deletions in this region and their phenotypic consequences in more detail, we collected DNA from three additional patients with overlapping interstitial 2p deletions. These DNA samples were hybridised on four different array platforms to test the performance of these platforms and to map the deletion breakpoints as precisely as possible.

Figure 1. The result of patients using Agilent microarray platform.



(A) Due to the dye swap, the deletion is depicted in two colors resulting in a symmetrical profile pattern. All genes localized within the deleted region are visualized using the Agilent software tool. The deleted region of patient 3 (B) and patient 4 (C) are outlined by the dashed line and the dotted line, respectively. For the size of the deletion see table 2. Pt= patient. [See appendix: colour figures.]

Platform 1: 3K BAC array

A summary of the results obtained by array-CGH analysis is shown in table 1. The deletion of patient 1 closely resembles that of patient 2 although it extends one centromeric BAC further. The deletion of patient 4 is the smallest. The minimal region of overlap is defined by the telomeric breakpoint of patient 3 and the centromeric breakpoint of patient 4 and it is estimated to be 1.4 - 1.5 Mb.

Table 2. Localisation of the breakpoints of the four deletions identified by the 3K BAC array, the 44K micro-array of Agilent, the 500K Affymetrix Genechip and the 317K beadchip of Illumina.

Patient 1	**Last probe +**	**Genome position**	**First probe -**	**Genome position**	**Last probe -**	**Genome position**	**First probe +**	**Genome position**	**Maximum size**	**Minimum size**
aCGH	RP11-204D19	43,517Kb	RP11-24I5	44,273Kb	RP11-508L23	53,793Kb	RP11-30C22	54,808Kb	11,291Kb	9,520Kb
Agilent	A_14_P115860	44,049Kb	A_14_P119582	44,076Kb	A_14_P135447	54,658Kb	A_14_P126243	54,703Kb	10,654Kb	10,582Kb
Affymetrix	rs6736282	44,058Kb	rs17031803	44,066Kb	rs4387841	54,608Kb	rs4671950	54,613Kb	10,555Kb	10,542Kb
Illumina	rs6723119	44,041Kb	rs11124960	44,069Kb	rs10496032	54,577Kb	rs11896012	54,631Kb	10,590Kb	10,508Kb
Patient 2	**Last probe +**		**First probe -**		**Last probe -**		**First probe +**			
aCGH	RP11-204D19	43,517Kb	RP11-24I5	44,273Kb	RP11-7H13	52,903Kb	RP11-508L23	53,793Kb	10,276Kb	8,630Kb
Agilent	N.D.		N.D.		N.D.		N.D.			
Affymetrix	rs7562014	43,991Kb	rs4953037	44,003Kb	rs1318578	53,557Kb	rs6727792	53,558Kb	9,567Kb	9,554Kb
Illumina	rs4953037	44,003Kb	rs12712900	44,010Kb	rs10164725	53,560Kb	rs4672456	53,588Kb	9,585Kb	9,550Kb
Patient 3	**Last probe +**		**First probe -**		**Last probe -**		**First probe +**			
aCGH	RP5-960D23	47,773Kb	RP11-460n15	47,795Kb	RP11-335O22	52,192Kb	RP11-7H13	52,903Kb	5,130Kb	4,397Kb
Agilent	A_14_P122998	47,627Kb	A_14_P102713	47,658Kb	A_14_P131858	52,224Kb	A_14_P115721	52,790Kb	5,163Kb	4,566Kb
Affymetrix	rs13401500	48,028Kb	rs6729916	48,028K	rs1861980	53,624Kb	rs4672481	53,627Kb	5,599Kb	5,596Kb
Illumina	rs2651767	47,926Kb	rs2134056	47,927Kb	rs2287511	53,626Kb	rs6746107	53,639Kb	5,713Kb	5,699Kb
Patient 4	**Last probe +**		**First probe -**		**Last probe -**		**First probe +**			
aCGH	RP11-110G2	46,104Kb	RP11-1084a21	?	RP11-436L21	49,158Kb	RP11-460M2	49,302Kb	3,198Kb	?
Agilent	A_14_P103946	46,819Kb	A_14_P105713	46,847Kb	A_14_P111706	49,272Kb	A_14_P101515	49,560Kb	2,741Kb	2,425Kb
Affymetrix	rs17035674	46,884Kb	rs2289929	46,898Kb	rs4971697	49,269Kb	rs12713041	49,270Kb	2,386Kb	2,371Kb
Illumina	rs1053952	46,903Kb	rs1374274	46,909Kb	rs6743414	49,278Kb	rs12622540	49,290Kb	2,387Kb	2,369Kb

Agilent data was not obtained from patient 2. The size of the deletion of patient 3 using Affymetrix genechip was calculated based on 250K resolution. Three out of six of the breakpoint locations defined by Agilent were not in agreement with the results provided by the SNP arrays (proximal and distal breakpoint of patient 3 and the distal breakpoint of patient 4). In three of the breakpoints there is a small difference in localisation of the breakpoints obtained by Affymetrix and Illumina (distal breakpoint of patient 2 and 4; proximal breakpoint of patient 4). One exception includes the distal breakpoint of patient 3, in which there is a difference of five SNPs encompassing 100 Kb of genomic sequence. Data were based on the Ensemble website, assembly December 2006.
?: the localisation of this BAC is unknown in Ensemble. N.D.: not determined

Platform 2: 44K Agilent Technologies microarray

At the time this study was performed, the Agilent 44K oligo array was the only Agilent array available, covering only gene-based sequences of the human genome. Three of the four interstitial 2p deletions were tested using this oligo array (figure 1). Patient 2 could not be tested, as there was not enough material available.

Due to the loop-hybridisation set up (see Material and Methods) in combination with a dye swap, the samples were effectively analysed in two independent hybridisations. The analysis maps the proximal breakpoint of patient 3 to the region between the *NRXN1* and the *ASB3* genes, a large region (2.8 Mb) devoid of known genes. Consequently, the estimation of this breakpoint might be less accurate. The results of this platform agree with the outcome of the array-CGH. Deletion size varied, from a maximum in patient 1 (10.6 – 10.7 Mb) to a minimum in patient 4 (2.4 – 2.7 Mb) (table 2). The size of the minimum region of overlap calculated based on the Agilent data is 1.6 Mb.

Platform 3: Affymetrix 250K / 500K Genechip

Three out of four patients were analysed using a 500K Genechip (patients 1, 2, 4). Patient 3 was analysed only by the 250K Genechip using the NspI restriction enzyme (figure 2). Interestingly, the sizes of the deletions calculated based on 250K analysis were comparable with those obtained from the combined data of both arrays (500K) (data not shown), indicating that for the calculation of large CNV the use of only one restriction enzyme can be sufficient.

The minimal region of overlap between the different interstitial deletions on chromosome band 2p is 1.2 Mb (table 2).

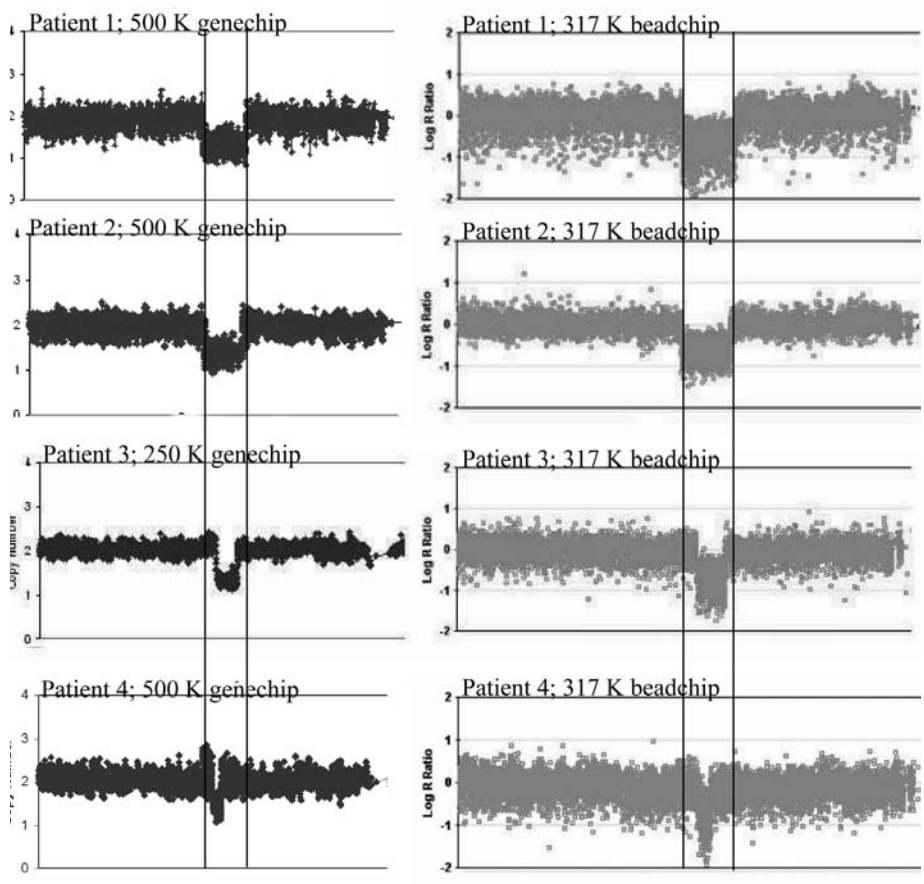
Platform 4. Illumina 317K beadchip

The results obtained for all four patients are depicted in figure 2. The results regarding the sizes are in agreement with the results obtained using the other 3 platforms (table 2). The minimal region of overlap, based on the results of the beadchip, is 1.4 Mb.

Discussion

In this study, different high resolution genome wide screening platforms were compared, including array-CGH using large insert clones, the long-oligo array of Agilent, the Affymetrix Genechip and the beadchip of Illumina. The genechip and the beadchip are SNP based arrays and they both use short-oligos.

Figure 2. The interstitial 2p deletion of the four patients analysed by Affymetrix genechip (left) and the beadchip of Illumina (right).



The deletions of the different patients are shown separately. Patient 3 was only analysed using 250K NspI genechip. A normal copy number of two is represented by a copy number between 1.6 and 2.4 for the Affymetrix genechip or by a LogR ratio between –0.3 and +0.3 for the beadchip of Illumina. The vertical lines represent the size of the largest deletion. In general, the variation of the data points obtained by the beadchip is larger than that of the genechip. Especially in patient 3, the difference in variation is remarkable. [See appendix: colour figures.]

Comparing across platforms, we found that the localisation of both proximal and distal breakpoints was largely in agreement (table 2). Nearly all BACs that showed 2 copies did not have overlap with regions that were deleted according to the results obtained by the SNP arrays and vice versa. One exception was the proximal breakpoint in patient 3 in which BAC RP11-7H13 should have been deleted according to the data

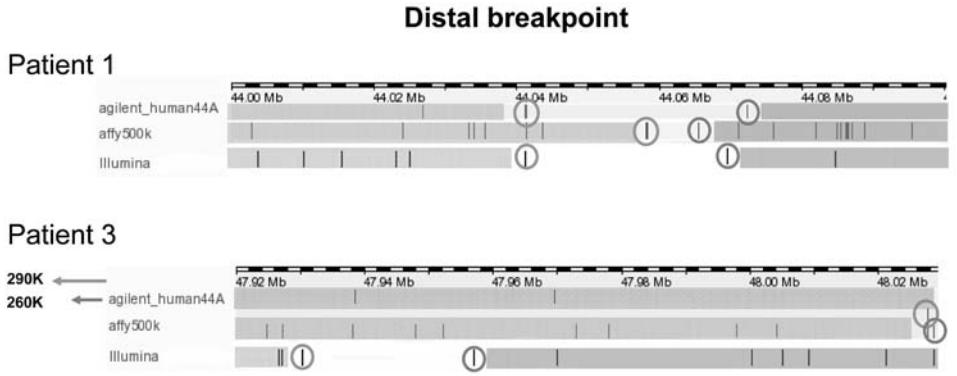
obtained by two SNP platforms. Notably, an 'aberrant' Agilent result was present at the same breakpoint as was the 'aberrant' BAC (proximal breakpoint in patient 3). In fact, the breakpoint mapping of the two array-CGH platforms was similar (breakpoint at ~52.8Mb), as was the outcome of both SNP platforms for the proximal breakpoint of patient 3 (localised at ~53.6Mb)(table 2). This might be explained by the difference in probe density near the breakpoints localised by the different platforms (see also Results). This idea is strengthened by the fact that, based on in silico data of the 244K array, (an improved version of the Agilent array), the number of oligonucleotides localised near the proximal breakpoint of patient 3 was significantly increased; 25 probes were localised within the breakpoint interval determined by the 44K array. The number of 'extra' probes present at the rest of the breakpoints defined by the 44K array is 5-10.

There is also some discrepancy between the outcomes of the two SNP platforms. In general, the data obtained using the beadchip showed more variation in all patients compared to that of genechip (figure 2). The maximum number of SNPs that were in discordance between the two SNP arrays was five (the distal breakpoint of patient 3) (figure 3). These five SNPs are in a genomic region covering more than 100 Kb. The number of data points for both SNP based arrays is similar at this breakpoint (based on the Ensembl database) (figure 3), indicating that both SNP platforms should be equally informative. We do not have a satisfactory explanation for this difference. The other differences in breakpoint mapping between the two SNP based tools included either only one or two SNPs (distal breakpoint of patient 2 and 4) or the differences in localisation of the breakpoints were very small (proximal breakpoint of patient 4; breakpoint mapping difference 7 kb). The observed difference in breakpoint mapping can be related to the use of different scoring algorithms that differ between platforms. This indicates that sequencing of the breakpoints is still needed to obtain information about the exact localisation of the breakpoint.

Patient 1, 3 and 4 did not show any copy number alterations outside chromosome region 2p16.2-p21. Patient 2, however, had a deletion on chromosome band 6q22.31 of ~1.2 Mb in size. This was identified using the BAC-array (BAC clone RP11-475J3) and the two SNP-based platforms. It has previously been found within the healthy population[23] and there are no known genes within the deleted region. It is therefore thought to be a neutral variant.

Currently, there is no golden standard available to determine which platform is the most accurate. It might be argued that high density SNP genotyping would be the

Figure 3. Overview of the distal breakpoints of patient 1 and patient 3 defined by Agilent, the Affymetrix genechip and the beadchip of Illumina.



The deleted region is depicted in red, whereas regions showing two copies are depicted in green. A green circle represents the last data point that showed a normal copy of two. A red circle represents the first data point that showed a deletion.
The number of data points per platform is comparable at the location of the distal breakpoint of patient 1 and 3. In patient 1, the breakpoint mapping of all platforms is concordant. In contrast, there is a huge difference in breakpoint mapping in patient 3. According to the results obtained by Agilent platform, the distal breakpoint of the deletion is located 290-260K outside the most distal point of the picture (47,92 Mb) (green and red arrow). The results of the Affymetrix platform show that the deleted region starts more proximally at ~48.03Mb (black arrow). The beadchip of Illumina defines the distal breakpoint of the deletion between these two points. [See appendix: colour figures.]

most appropriate to implement for the screening copy number alteration, as this tool offers the simultaneous measurement of copy number changes and copy-neutral loss of heterozygosity (i.e uniparental disomy). On the other hand, the SNPs have been selected based on criteria such as heterozygosity, confirmation with Hardy-Weinberg equilibrium. Although these features are important for association studies, where SNPs need to be informative, they are less critical for copy number analysis where even spacing is more important. Indeed, many regions prone to rearrangements (e.g. duplicons) are lacking or are underrepresented on these arrays, as the associated SNPs did not meet the required quality criteria. This is in contrast to array-CGH in which the location of the oligonucleotides is not limited to known SNPs, and, therefore, it is possible to analyse regions of the genome where no validated SNPs are available.[24] Calculating the spacing between the consecutive data points per platform within chromosome region 2p16.2-p21, shows that the median spacing of genechip was 2.40 kb, with a maximum of 65.10 kb, the median spacing of the beadchip was 4.57 kb (with a maximum of 71.85 kb) and finally, that of Agilent using the 244K was 9.85 kb (with

a maximum of 47.40 kb). Thus, although the median spacing of Agilent is the largest (as it is gene-based), this platform might be the most valuable tool for investigating CNVs, depending on the genome region of interest (based on the maximum spacing of the three platforms).

Indeed, a previous study[24] has shown that in addition to the SNP-arrays, array-CGH analysis is required to cover all CNV regions in the human genome, with at least one third of CNVs >50 kb otherwise being missed. New arrays of both Affymetrix and Illumina are closing this gap by combining both SNP- and non-SNP probes on one array.

The beadchip has several clear advantages over the genechip, such as a higher SNP call rate, which is important when the expected size of the CNV is small. In our study, about 5-10%, sometimes even more, of all SNPs on the Affymetrix platform could not be scored (data not shown), resulting in a significant reduction of its resolution. Of course, the cause of such reduction might lie in a suboptimal quality of the DNA, however the identical DNA was used on the Illumina arrays. In addition, the genechip needed two arrays (this experiment) for a resolution comparable to that of the beadchip, which is especially of interest for the detection of small CNVs, and nearly all steps of the Illumina protocol can be automated. At the time these experiments were performed, only Illumina provided customer friendly software. Recently, however, software enabling easy calculation of the data generated by Affymetrix has become available, demonstrating the fast adaptation of products and application within this field. An important argument in favor of the genechip is the fact that they have started to validate these arrays to allow implementation in a diagnostic setting.

Looking at the breakpoints of the four patients, it can be concluded that the deletion of patients 1, 2 and 4 includes both the *MSH2* and the *MSH6* genes. The distal breakpoint of patient 3 is localised within or nearby (depending on the platform applied) the *MSH6* gene; the *MSH2* gene is not deleted in patient 3. This means that at least three out of four patients have a twenty fold increased chance of developing colon cancer or other Lynch syndrome-related tumors[25] compared to the healthy population. For this reason, it is of high clinical interest to diagnose the breakpoints of interstitial 2p deletions. However, when comparing the phenotype of the four patients (table 3), the only feature in common is mental retardation, which is a non-specific feature of nearly all chromosome anomalies. The lack of a common phenotype could be due to the different ages at observation (table 3) of the patients and the difference in size of the deletion.

Table 3. Overview of clinical features of the four patients with different sizes of interstitial 2p deletions.

	Patient 1	**Patient 2**	**Patient 3**	**Patient 4**
Localisation	2p16.2-p21	2p16.2-p21	2p16.3	2p16.3-p21
Cytogenetically visible	Yes	Yes	No	No
De novo	Yes	Yes	N.D.	Father not tested
Age of examination	6, 13, 36 months	5, 13.6, 17 month	7 years	37 years
Psychomotor delay	**Present**	**Present**	**Present**	**Present**
Length	Short stature (-2 SD)	Tall stature	75th percentile	Short stature
Weight	10th percentile	95th percentile	50-75th percentile	> 97th percentile
Size of skull	< 25th percentile	Microcephaly	50th percentile.	< 25th percentile
Shape of skull	Flattening of the occipital region	Flattening of posterior parietal region	Turricephaly	Brachycephaly with narrow forehead
Others	Aorta descendens P97. Palatoschisis Cataract, Nystagmus, Strabismus convergens	ASD Mild astigmatism Hypothelorism	Joint hyperextensibility with tendency to dislocation, High arched, narrow palate oblong face, large mouth, thin upper lip	
Colon cancer	Too young	Too young	Too young	Yes

N.D.: not determined

The minimum region of deletion overlap is localised between the distal breakpoint of patient 3 and the proximal breakpoint of patient 4. This region is ~1.3 Mb in size and encompasses 8 genes, from *MSH6* to *FSHR*. So far, this region was not found altered among healthy individuals.[24] The *FOXN2* gene, located between *MSH6* and *FSHR* might be of interest in relation to the phenotype of the patients. It is known that deregulation of FOX family genes can lead to congenital disorders in addition to its involvement in several types of cancer. Furthermore, the *FBXO11* gene coding for F-box protein family, might be involved in some of the developmental anomalies, as it related to phosphorylation-dependent ubiquitination. Mutations within the *LHCGR* and the *FSHR* genes are related to aberrant external and/or internal genital organs. No mutations with specific pathogenetic consequences have been reported for the remaining two genes (*CCDC128, STON1*).

Recently, the whole genome of Nobel laureate Jim Watson was sequenced (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi), revealing as much as 600,000 single nucleo-

tide variants that had not been reported before. The cost involved of this project was substantial and therefore this way of screening the human genome is not applicable on large scale yet. It can be expected, however, that affordable sequence-based whole genome genotyping will become possible within the coming two years. As a result, SNP typing and array-CGH will be superseded fairly soon by next generation sequencing. The first step towards the implementation of genome wide sequencing in a diagnostic setting would be to type "harmless" variations in a large group of normal individuals, since on average 1 in 1000 nucleotide on the human genome of a healthy individual varies. In addition, screening large cohorts of affected individuals with well-defined clinical features is essential to be able to interpret this new data.[26]

References

1. Caspersson T, Lomakka G, Zech L. The 24 fluorescence patterns of the human metaphase chromosomes - distinguishing characters and variability. *Hereditas* 1972: **67**(1):89-102.
2. Yunis JJ. High resolution of human chromosomes. *Science* 1976: **191**(4233):1268-1270.
3. Landegent JE, Jansen in dW, van Ommen GJ, Baas F, de Vijlder JJ, Van Duijn P, Van der PM. Chromosomal localization of a unique gene by non-autoradiographic in situ hybridization. *Nature* 1985: **317**(6033):175-177.
4. Ried T, Mahler V, Vogt P, Blonden LAJ, Van Ommen GJB, Cremer T, Cremer M. Direct carrier detection by in situ suppression hybridization with cosmid clones of the Duchenne/Becker muscular dystrophy locus. *Hum Genet* 1990: **85**:581-586.
5. Raap AK, Van De Corput MPC, Vervenne RAW, Van Gijlswijk RPM, Tanke HJ, Wiegant J. Ultra-sensitive FISH using peroxydase mediated deposition of biotin- or fluorochrome tyramides. *Hum Mol Genet* 1995: **4**:529-534.
6. Liehr T, Starke H, Heller A, Kosyakova N, Mrasek K, Gross M, Karst C, Steinhaeuser U, Hunstig F, Fickelscher I, Kuechler A, Trifonov V, Romanenko SA, Weise A. Multicolor fluorescence in situ hybridization (FISH) applied to FISH-banding. *Cytogenet Genome Res* 2006: **114**(3-4):240-244.
7. Liehr T, Starke H, Weise A, Lehrer H, Claussen U. Multicolor FISH probe sets and their applications. *Histol Histopathol* 2004: **19**(1):229-237.
8. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997: **20**(4):399-407.
9. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998: **20**(2):207-211.
10. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001: **29**(3):263-264.
11. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, Girard L, Minna J, Christiani D, Leo C, Gray JW, Sellers WR, Meyerson M. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004: **64**(9):3060-3071.
12. Scott DA, Klaassens M, Holder AM, Lally KP, Fernandes CJ, Galjaard RJ, Tibboel D, de Klein A, Lee B. Genome-wide oligonucleotide-based array comparative genome hybridization analysis of non-isolated congenital diaphragmatic hernia. *Hum Mol Genet* 2007: **16**(4):424-430.
13. Herr A, Grutzmann R, Matthaei A, Artelt J, Schrock E, Rump A, Pilarsky C. High-resolution analysis of chromosomal imbalances using the Affymetrix 10K SNP genotyping chip. *Genomics* 2005: **85**(3):392-400.
14. Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J, Liu G, Ihara S, Nakamura H, Hurles ME, Lee C, Scherer SW, Jones KW, Shapero MH, Huang J, Aburatani H. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 2006: **16**(12):1575-1584.
15. Knijnenburg J, Szuhai K, Giltay J, Molenaar L, Sloos W, Poot M, Tanke HJ, Rosenberg C. In-

sights from genomic microarrays into structural chromosome rearrangements. *Am J Med Genet* A 2005: **132**(1):36-40.

16. Vissers LE, De Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, Straatman H, Van D, V, Huys EH, Van Rijk A, Smeets D, Ravenswaaij-Arts CM, Knoers NV, Van DB, I, De Jong PJ, Brunner HG, van Kessel AG, Schoenmakers EF, Veltman JA. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am *J Hum Genet* 2003: **73**(6):1261-1270.
17. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 2001: **98**(1):31-36.
18. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004: **20**(8):1233-1240.
19. Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Asano J, Bailey DK, Barber S, Birch P, Brown-John M, Cao M, Chan S, Charest DL, Farnoud N, Fernandes N, Flibotte S, Go A, Gibson WT, Holt RA, Jones SJ, Kennedy GC, Krzywinski M, Langlois S, Li HI, McGillivray BC, Nayar T, Pugh TJ, Rajcan-Separovic E, Schein JE, Schnerch A, Siddiqui A, Van Allen MI, Wilson G, Yong SL, Zahir F, Eydoux P, Marra MA. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* 2006: **79**(3):500-513.
20. Sanders SR, Dawson AJ, Vust A, Hryshko M, Tomiuk M, Riordan D, Prasad C. Interstitial deletion of chromosome 2p16.2p21. *Clin Dysmorphol* 2003: **12**(3):183-185.
21. Lucci-Cordisco E, Zollino M, Baglioni S, Mancuso I, Lecce R, Gurrieri F, Crucitti A, Papi L, Neri G, Genuardi M. A novel microdeletion syndrome with loss of the MSH2 locus and hereditary non-polyposis colorectal cancer. *Clin Genet* 2005: **67**(2):178-182.
22. White S, Kalf M, Liu Q, Villerius M, Engelsma D, Kriek M, Vollebregt E, Bakker B, van Ommen GJ, Breuning MH, Den Dunnen JT. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* 2002: **71**(2):365-374.
23. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science* 2004: **305**(5683):525-528.
24. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature* 2006: **444**(7118):444-454.
25. Hendriks YM, de Jong AE, Morreau H, Tops CM, Vasen HF, Wijnen JT, Breuning MH, Brocker-Vriends AH. Diagnostic approach and management of Lynch syndrome (hereditary nonpolyposis colorectal carcinoma): a guide for clinicians. *CA Cancer J Clin* 2006: **56**(4):213-225.
26. Ropers HH. New perspectives for the elucidation of genetic disorders. *Am J Hum Genet* 2007: **81**(2):199-207.