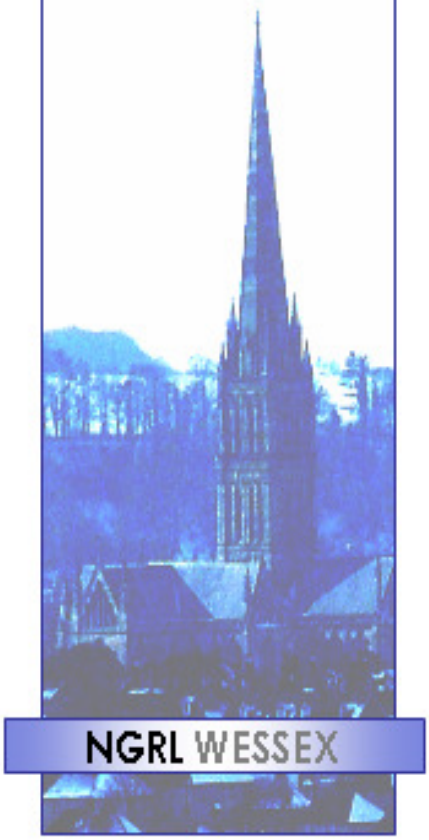
National Genetics Reference Laboratory
(Wessex)

NHS

**Technology Assessment**




# Comparison for cytogenetics array platforms hardware and software for use in identifying copy number aberrations in constitutional disorders

**July 2010**

| | |
|---|---|
| **Title** | Comparison for cytogenetics array platforms hardware and software for use in identifying copy number aberrations in constitutional disorders |
| **NGRL Ref** | NGRLW_aCGH_1.0 |
| **Publication Date** | July 2010 |
| **Document Purpose** | Dissemination of information about cytogenetics array platforms hardware and software for use in identifying copy number aberrations in constitutional disorders |
| **Target Audience** | Laboratories performing or setting up aCGH |
| **NGRL Funded by** |  |



**Contributors**

| Name | Role | Institution |
|---|---|---|
| Dr Shuwen Huang | Senior Scientist | NGRL ( Wessex) |
| Dr John Crolla | Head of Cytogenetics | WRGL |

**Conflicting Interest Statement**

The authors declare that they have no conflicting financial interests. All companies have been given the opportunity to comment on the content of the report

**How to obtain copies of NGRL (Wessex) reports**

An electronic version of this report can be downloaded free of charge from the NGRL website (http://www.ngrl.org.uk/Wessex/downloads)

or by contacting

Prof Nick Cross
National Genetics Reference Laboratory (Wessex)
Salisbury District Hospital
Odstock Road
Salisbury
SP2 8BJ
UK

E mail: ncpc@soton.ac.uk
Tel: 01722 429080
Fax: 01722 338095

# Comparison for cytogenetics array platforms hardware and software for use in identifying copy number aberrations in constitutional disorders

Shuwen Huang and John Crolla

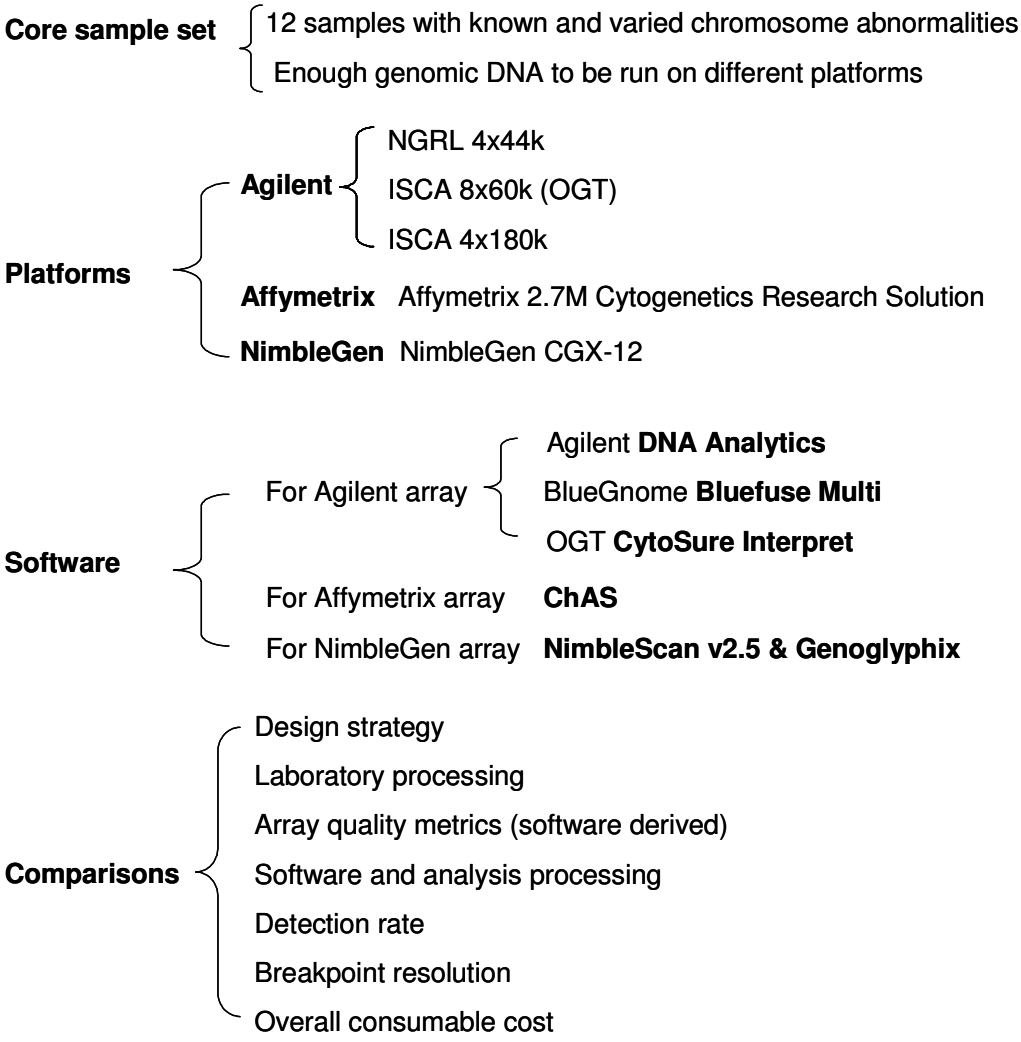
National Genetics Reference Laboratory (Wessex), Wessex Regional Genetics Laboratory, Salisbury District Hospital, Odstock Road, Salisbury, Wiltshire, UK.

## Abstract

Copy number variants have been shown to be associated in a significant proportion of cases presenting with developmental delay and a range of other phenotypes including dysmorphic features with or without congenital anomalies, mental retardation, and autistic spectrum disorders. Array Comparative Genomic Hybridization (aCGH) is now the first line test for patients with these disorders in many laboratories throughout the developed world [1]. Several commercial platforms are available which can either be adopted directly or customised to meet specific diagnostic criteria. An evaluation and comparison of these different platforms is presented here to help inform testing laboratories which platform(s) are most suitable in the clinical cytogenetics diagnostic setting.

We selected 12 cytogenetically abnormal samples originally analysed by conventional karyotyping and/or using our customised National Genetics Reference Laboratory (NGRL) 4x44k constitutional arrays. We ran these same 12 samples on (1) the Affymetrix 2.7M Cytogenetics Research Solution, (2) the Agilent International Standard Cytogenomic Array Consortium (ISCA) custom 4x180k array, (3) Oxford Gene Technology's (OGT) CytoSure™ ISCA 8x60k array printed by Agilent, and (4) NimbleGen CGX-12 array. Comparisons have been focused on evaluating several parameters, viz. (a) the optimal design for constitutional aCGH, (b) laboratory processing, (c) aCGH quality metrics, (d) software and analytical processing, (e) detection rate, (f) breakpoint resolution, and (g) overall consumable costs (Figure 1). Comparison data from these different platforms and software packages for the 12 core samples are presented. Based on the information obtained from this comparison and the evaluation project, combined with recent technical improvements in our array laboratory, we present a strategy that uses a semi-automated aCGH workflow and the OGT's CytoSure ISCA 8x60k array platform to provide significant advances in both processing and quality whilst achieving significant consumable cost savings.

**Figure 1**. The basic components used for the comparison




## Selection of 12 core sample set

The core comparison set includes 12 samples with sufficient genomic DNA to run on multiple platforms. These were selected based on the results from conventional cytogenetics and/or the NGRL 4x44k arrays (Table 1). The selection criteria for these cases include: 1) The size of the region called and/or poor probe coverage, but involving potentially important genes, e.g. case 1 and case 4 where 4x44k gave equivocal results. 2) Single abnormalities including deletions and duplications, e.g. cases 2, 3, 7, 9, and 12, which were selected because the cytogenetic abnormalities reflected different regions, and where the comparison is therefore focused on breakpoint resolution. 3) Deletion and duplication in a single patient but on different chromosomes, e.g. case 8 and case 11. 4) Complex abnormalities: case 5 shows different $\log_2$ ratios within the amplified region on chromosome X; case 6 has a chromosome

4 deletion and a mosaic marker chromosome 13 which was originally detected by conventional karyotyping, but was not detected by the DNA Analytics software on 4x44k array using ADM-2 algorithm. 5) 4x44k provided preliminary evidence of separate non-contiguous deletions instead of a contiguous deletion, e.g. case 10.

**Table 1**. The 12 core sample set

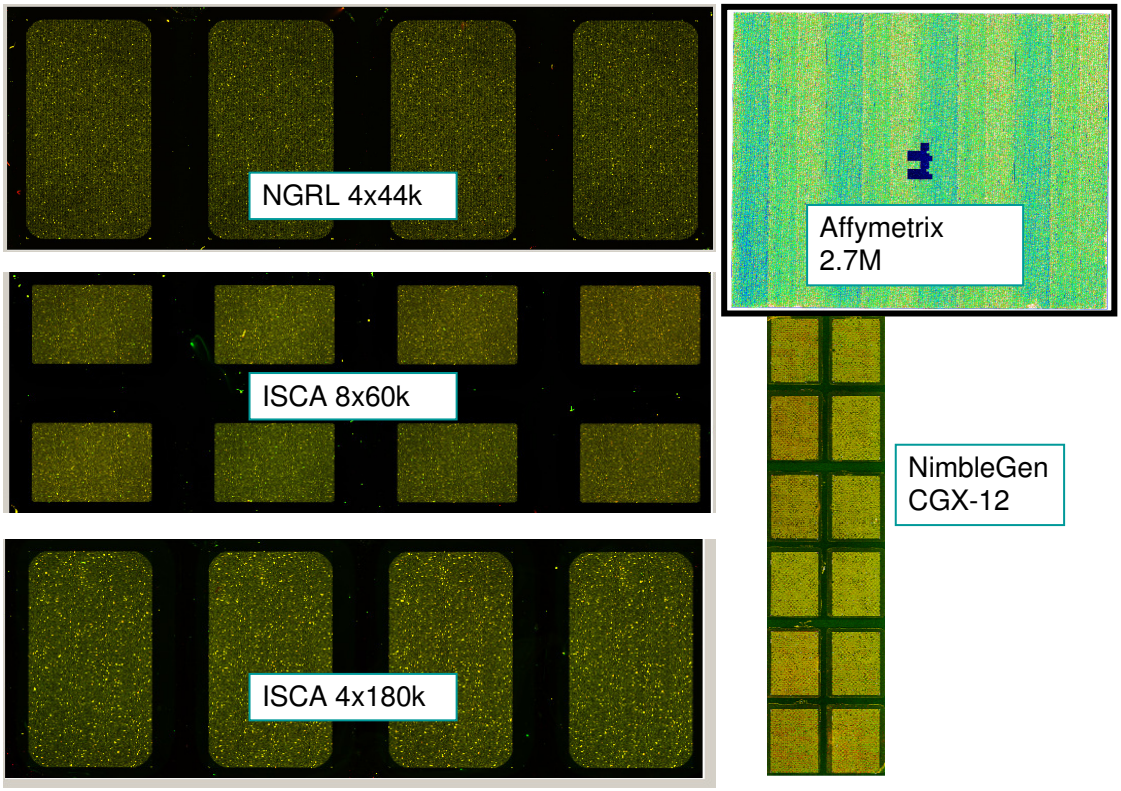
| Core Sample Set | | | |
|---|---|---|---|
| **Sample No** | **Gender** | **Copy number changes detected on 4x44k** | **Note** |
| Case 1 | F | 2 probes deleted on *A2BP1* gene (chr16) | Non-optimal coverage on *A2BP1*. |
| Case 2 | F | del(15)(q13.2q13.3)mat | Single abnormality deletion |
| Case 3 | F | dup(11)(q13.4) | Single abnormality duplication |
| Case 4 | F | chr18 *TCF4* gene del (3 probes) | Non-optimal coverage. |
| Case 5 | M | dup(X)(q28)mat | Chromosome X complex amplification |
| Case 6 | M | der(13)(q12.11q12.2)de novo; del(4)(q13.3)de novo | Del(4) not visible by karyotyping. Probe density in 44k did not detect mosaic SMC(13) using ADM-2. |
| Case 7 | M | del(17)(q23.1q23.2) | Single abnormality deletion |
| Case 8 | F | del(15)(q13.2q13.3); dup(17)(q12) | Complex abnormality (deletion and duplication on different chromosomes) |
| Case 9 | F | del(15)(q13.1q13.2)pat | Single abnormality deletion |
| Case 10 | M | del(15)(q13.1q13.3) ; dup(16)(p11.2) | Suspected non-contiguous deletion. |
| Case 11 | M | del(17)(q12) + dup(6)(q25.1) | Complex abnormality (deletion and duplication on different chromosomes) |
| Case12 | F | dup(12)(p11.23p11.22) | Improve breakpoint resolution. |

## Platforms and software for comparison

The core samples were run on five platforms for comparison: 1) Agilent **4x44k** custom array, National Genetics Reference Laboratory (NGRL), Wessex. 2) Agilent **4x180k** custom array, International Standard Cytogenomic Array (ISCA) Consortium. 3) OGT's CytoSure ISCA **8x60k** array printed by Agilent. 4) NimbleGen **CGX-12** custom array, Signature Genomics, and 5) Affymetrix 2.7M Cytogenetics Research Solution , Affymetrix. The scan images for the above platforms are shown in Figure 2.

For the **Agilent arrays**, three software packages have been evaluated: Agilent's **DNA Analytics 4.0.76**, OGT's **CytoSure Interpret v3.0.6**, and BlueGnome's **BlueFuse Multi v2.1.** For the **NimbleGen CGX-12 array**, **NimbleScan v2.5 & Genoglyphix v2.4** was used. The **Chromosome Analysis Suite (ChAS) v1.0.1** software is used for **Affymetrix 2.7M array.**

**Figure 2.** Examples of scan images for the compared platforms

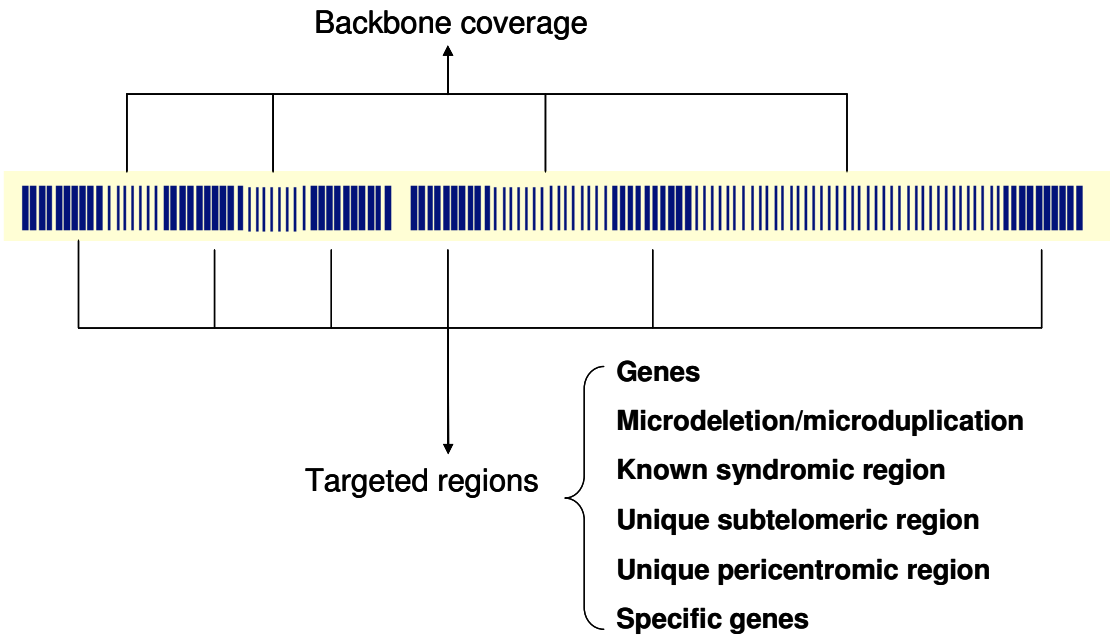



### Design strategy

All five platforms were designed specifically for the cytogenetics community, which in general terms uses a strategy comprising an overall backbone coverage combined with focused targeted regions (Figure 3). Agilent based array platforms provide facilities for the customers to be directly involved in the design by using Agilent's free on-line eArray software. Affymetrix 2.7M array and NimbleGen CGX-12 arrays were designed by the companies in collaboration with clinical cytogeneticists. Users are unable to make any direct design changes, but comments and recommendations can be submitted to the company for consideration to be included in later iterations.

NGRL 4x44k array was designed in November 2006, based on the "off the shelf" Agilent 44B commercial array content. By reducing the coverage density of cancer related genes in the 44B design, large genomic intervals with no probes in regions of ≥250kb were populated with probes and, in addition, 155 target regions (known deletion and duplication intervals) were given higher probe density coverage. The 155 original targeted regions can be found at http://www.ngrl.org.uk/Wessex/microdel_collection.htm.

Figure 3. Whole genome plus targeted array design




The ISCA 8x60k and 4x180k have an even backbone probe coverage of 60kb and 25kb, respectively, and high density coverage of ~ 500 targeted regions with the spacing of 5kb per probe or at least 20 probes per gene region. These targeted regions include telomere and unique centromere FISH clone regions, microdeletion/duplication regions, genes of known haploinsufficiency, and X-linked mental retardation regions. The 4x180k version utilises 140,000 probes thereby allowing 40,000 probes for individual customisation if required.

The Affymetrix 2.7M array has evenly distributed whole genome backbone coverage of ~ 1kb spacing, with dense intragenic markers of about 690bp spacing. It has 2.3 million non-polymorphic markers for the detection of copy number variants and 400,000 SNP markers for identifying loss of heterozygosity (LOH), uniparental disomy (UPD), and regions identical-by-descent.

The NimbleGen CGX-12 array covers regions known to be involved in cytogenetic abnormalities, including over 200 syndromes, the pericentromeric regions, and subtelomeres, with a maximum probe spacing of one probe every 35 kb throughout the genome and one probe every 10 kb in clinical regions. Table 2 shows the basic design information of these five platforms.

**Table 2.** Comparison table of array design strategies

| | Affymetrix 2.7M | NimbleGen CGX-12 | ISCA 4x180k (Agilent) | ISCA 8x60k (OGT/Agilent) | NGRL 4x44k (Agilent) |
|---|---|---|---|---|---|
| Backbone coverage | 1kb/probe | 35k/probe | 25kb/probe | 60kb/probe | 75kb/probe |
| Target region coverage | 690bp/probe | 10kb/probe | 5kb/probe or 20 probes per targeted region (or on average, 50 probes per targeted region) | On average, 40 probes per targeted region | Two to five extra probes per targeted regions |
| Backbone resolution | Depend on number of markers, size and confidence | 175kb (5 contiguous probes to make a call) | 100kb (4 contiguous probes to make a call) | 240kb (4 contiguous probes to make a call) | 225kb (3 contiguous probes to make a call) |
| Target region resolution | If the filter set as 20 markers within 50kb, the resolution should be 20x690 for the targeted region | 50kb (5 contiguous probes to make a call) | 20kb (4 contiguous probes to make a call) | ~48kb (4 contiguous probes to make a call) | ~60kb (3 contiguous probes to make a call) |
| Probe types | 2.3 million non-polymorphic markers for CNV, and 400,000 SNP markers for LOH, UPD, and regions identical-by descent | Non-polymorphic markers for CNV | Non-polymorphic markers for CNV | Non-polymorphic markers for CNV | Non-polymorphic markers for CNV |
| Probe size | 49 mer | 60 mer | 60 mer | 60 mer | 60 mer |
| How many targeted regions are included | Coverage (%) if using filter of 20 markers with minimum size of 50 kb<br>RefSeq genes (18,701) 18,533 (99.1%)<br>Cancer genes (318) 318 (100%)<br>Cytogenetics relevant/haploinsufficiency genes (559) 548 (98.0%)<br>X chromosome genes (801) 786 (98.1%)<br>OMIM genes (12,341) 12,242 (99.2%) | Over 700 genes, 200 recognized genetic syndromes, 41 unique subtelomeric regions, 43 unique pericentromeric regions. Regions tested could be viewed at http://www.signaturegenomics.com/disorders_tested.html | 501 targeted regions | 498 targeted regions | 155 targeted regions (http://www.ngrl.org.uk/Wessex/microdel_collection.htm) |
| Customer self design availability | No | No | Yes | Yes | Yes |

## Laboratory processing

Four array platforms were processed in the NGRL microarray lab, Affymetrix 2.7M Cytogenetics Research Solution, NGRL 4x44k custom array, ISCA 4x180k array and OGT's CytoSure ISCA 8x60k array. The NimbleGen CGX-12 arrays were carried out by Signature Genomics in the USA using the 12 core DNA samples provided by the NGRL. For the arrays processed at the NGRL, Promega's male and female pooled control DNA samples were used as references, and a sex matched strategy was applied. For the NimbleGen CGX-12 array, individual male and individual female control samples were selected as appropriate.

Laboratory processing comparisons are focused on the following: 1) the amount of genomic DNA required; 2) whether a DNA amplification procedure is required; 3) overall time to complete the procedure; 4) hands-on time; 5) scanner resolution; 6) volume of data generated per sample; 7) time per sample to process data including scanning and data extraction (Table 3).

**Table 3.** Comparison table of laboratory processing

| Platform | Affymetrix | NimbleGen CGX | Agilent | | |
|---|---|---|---|---|---|
| **Array** | **2.7M Cytogenetics Research Solution** | **CGX-12** | **ISCA 4x180k** | **ISCA 8x60k (OGT)** | **NGRL 4x44k** |
| Amount of DNA needed | 0.1µg | 1ug | 0.5-1.5µg | 0.5-1.0ug | 0.5-1.5µg |
| DNA amplification procedure needed? | Yes | No | No | No | No |
| Overall time to complete procedure | 3 days | 3-4 days | 2-3 days | 2-3 days | 2-3 days |
| Hands-on time | ~3 hours for 8 samples | ~7 hours | ~7 hours for 8 samples | ~7 hours for 8 samples** | ~7 hours for 8 samples |
| Scanner resolution | 2.5µm | 5µm*** | 3µm | 3µm | 5µm |
| Volume of data generated per sample | 455Mb | 250Mb | 756Mb | 710Mb | 239Mb |
| Scan time (per slide) | 15mins | 12min | 15mins | 15mins | 8 mins |
| Extraction of data time (per sample) | 14mins | 3mins | 3.5mins | 1.9mins | 1.2mins |
| Overall time* spent processing data (per sample) | 29min | 15mins | 18.5mins | 16.9mins | 9.2mins |

* Overall time = scan time + data extraction time
** Significantly reduced by semi-automation (see page 14)
*** Scanner resolutions of 2, 2.5 and 3µm are also available.

From Table 3, it can be seen that the amount of input DNA for Affymetrix array is ~ 100ng, but a DNA amplification procedure is needed. The recommended amount of DNA for the remaining platforms is ~ 1000ng, but without DNA amplification. It is possible however to obtain an analysable result using less DNA, for example, we have used 200ng for the Agilent platforms and obtained very good results.

The processing speeds for all platforms are essentially the same using the recommended manufacturers' protocols. The overall time to complete the procedure for all platforms is ~ 3 days. The hands-on bench time for the Affymetrix arrays is ~ 3 hours for 8 samples, while the other platforms required ~ 7 hours. The data collected here are based on manual methods, and these hands-on times can be significantly reduced by semi-automation (see below).

The resolution of the scanner available will determine the type (s) of array (s) which can be used locally. Table 3 gives details of both scanner resolution and also the size of data files generated. This is an important consideration in the context of local IT infrastructure planning.

## Array quality metrics

Each platform has its own measure of array quality, but there are some common parameters shared by the platforms. For example, the Derivative Log Ratio Spread (DLRS), the standard deviation of the differences between adjacent points divided by the square root of 2, is used by Agilent's DNA Analytics, OGT's CytoSure Interpret, and BlueGnome's Bluefuse Multi

software. We compared the same data set of 4x44k arrays using the above software packages and obtained very similar average DLRS scores, 0.16, 0.15 and 0.14 respectively for the DNA Analytics, CytoSure Interpret and Bluefuse Multi. Standard Deviation (SD), a measure of the dispersion of a set of data from its mean, is used by both Bluefuse Multi and Genoglyphix. The two major QC metrics for Affymetrix array are the Median Absolute Pairwise Difference (MAPD) score which applies to copy number probes and the SNP QC which applies to SNP probes. Table 4 lists the major QC metrics for each platform and software package, the reference values for a good quality array, and the average QC metrics for our 12 core samples on each platform and/or software. As shown in the table, there is no significant difference between these platforms/software using the manufacturers' standard procedures.

**Table 4.** Comparison table for array QC metrics*

| **Platform** | **Affymetrix** | **NimbleGen** | **Agilent** | | |
|---|---|---|---|---|---|
| **Software** | **ChAS*** (2.7M) | **NimbleScan v2.5 & Genoglyphix** (12x135k) | **DNA Analytics*** (4x180k, 8x60k, 4x44k) | **Bluefuse Multi*** (4x44k) | **CytoSure Interpret*** (4x44k) |
| Major parameters and high quality array reference values | **MAPD < 0.27** | **SD<0.14** | **DLRS <0.2** | **SD<0.1** | **DLRS <0.2** |
| | **SNP QC >1.1** | Mad.1dr <0.23 | Signal to noise green >100 | **DLRS <0.15** | Green Signal to Noise Ratio > 100 |
| | Interquantile range (iqr) | Interquartile density 0.5-1.8 | Signal to noise red >100 | Mean Ch1 Spot Amplitude: 400-700 | Red Signal to Noise Ratio > 100 |
| | Median-raw-intensity >2000 | Ratio range <1.0 | Signal intensity green >150 | Mean Ch2 Spot Amplitude: 400-700 | Green Signal Intensity > 150 |
| | Antigenomic ratio <0.4 | Signal range <1.5 | Signal intensity red >150 | SBR Ch1: 5-20 | Red Signal Intensity > 150 |
| | Waviness-seg-count <100 | | BG noise Green <5 | SBR Ch2: 5-20 | Green Background Noise < 5 |
| | Waviness-sd <0.1 | | BG Noise Red <5 | | Red Background Noise < 5 |
| | | | Reproducibility Green <0.05 | | Signal Intensity Ratio > 0.7 |
| | | | Reproducibility Red <0.05 | | Non-Uniform Features < 0.005 (0.5%) |
| | | | | | Saturated Features < 0.005 (0.5%) |
| Value comparison of major parameter for the 12 core samples | **MAPD=0.18 SNP QC=2.28** | **SD=0.15** | **DLRS 4x44k = 0.16 8x60k = 0.12 4x180k = 0.15** | **SD=0.1 DLRS=0.14** | **DLRS=0.15** |

* These are obtained under our laboratory conditions and so may vary when used locally.

## Software and analysis processing

Five software packages were compared and evaluated and comparisons were focused on the following: 1) total analytical time per case; 2) aberration filters used; 3) algorithms used; 4) tracks included; and 5) flexibility provided by adding custom tracks.

Within the definition of total analytical time, we have included time spent on analysis, checking, and authorization. Generally, more time is required for higher resolution arrays and more complicated abnormalities. As shown in table 5, the total analytical time per case under our conditions is around 30-45 minutes for all the platforms evaluated.

For most of the array platforms, aberration filter defines the minimum number of probes in the region. For example, for the 4x44k arrays, the aberration filter of 3 contiguous probes was used to make a call; for the 4x180k and 8x60k arrays, 4 contiguous probes were applied. The NimbleGen CGX-12 array aberration filter was set as 5 contiguous probes.

The aberration filter for Affymetrix's ChAS is composed of three parts, the marker count, the mean marker distance and the confidence values. Although there are recommended settings, users need to try different combinations in order to find the most suitable filters for their specific applications. In order to compare the data we collected from other platforms and software packages, we decided the optimal aberration filter for copy number changes were as follow: with the confidence value set at 85%, 85 markers within 100kb along the whole genome, and 20 markers within 50kb in each of the 559 "cytogenetics relevant/haploinsufficiency genes" as defined by Affymetrix. Full details of these regions/genes are not directly available via the Affymetrix website but can be obtained by direct contact with the company. In this way the software will call an aberration only when there are at least 85 contiguous probes abnormal within 100kb along the whole genome, or 20 contiguous probes abnormal within 50kb in the "cytogenetics relevant/haploinsufficiency genes", which has a confidence value of over 85%. In addition to copy number variant detection, ChAS also incorporates the function for detecting long contiguous stretches of homozygosity (LCSH), which can be used for identifying LOH, UPD, and regions identical-by-descent. We used the company's recommended 3 contiguous markers within 5Mb and 85% confidence as the aberration filter. For mosaicism detection, we used 500 markers within 5Mb and an 85% confidence limit.

**Table 5**. Comparison table for software and analysis processing

| **Array Company** | **Affymetrix** | **NimbleGen** | **Agilent** | | | | |
|---|---|---|---|---|---|---|---|
| **Software** | **ChAS v1.0.1** | **Genoglyphix v2.4 (Signature Genomics)** | **DNA Analytics v4.0** | | | **BlueFuse Multi v2.1 (BlueGnome)** | **CytoSure Interpret v3.0.6 (OGT)** |
| **Array platform** | **2.7M** | **CGX-12** | **4x180k** | **8x60k** | **4x44k** | **4x44k** | **4x44k** |
| Total analytical time per case* (Average) in minutes | 40 | 30 | 45 | 35 | 30 | 45 | 30 |
| Aberration filters | 100kb with 85 markers for amp + del, 50kb with 20 markers for cyto relevant regions<br>5Mb for LCSH$^{\Psi}$ with 3 markers and mosaicism with 500 markers, 85% Confidence | 5 contiguous probes | 4 contiguous probes | 4 contiguous probes | 3 contiguous probes | 3 contiguous probes | 3 contiguous probes |
| Algorithms | | NG packager segmentation algorithm | ADM-2 | | | Multi v1.0 | CBS |
| Tracks included | BACs, DGV, FISH Clones, Genes, OMIM, Segmental Duplications, sno/miRNA | FISH probes, Sequence Gaps, Segmental Duplications, GC Content, SignatureSelect Clones, SignatureSelect OS 105K Probes, NimbleGen CGX Probes, SignatureSelect OS 44K Probes, Abnormal Region(s), MyGCAD, Community GCAD, GCAD, Benign CNVs, Genes, RefSeq Genes, SGL GPS, SGL CNVs, DGV | Genes, DGV, CpGIsland, miRNA, PAR | | | Disease, Genes, BlueFISH, BAC Gain/ BAC Loss, Oligo Gain/ Oligo Loss, DGV Gain/ DGV Loss, BG Gain/ BG Loss | Syndrome, Gene, Exon, CHOP CNV, ECARUCA, Recombination hotspot, DGV, Confirmation (FISH and MLPA probes), DECIPHER, Redon CNV |
| Possibility of adding custom tracks | Yes | Yes | Yes | Yes | Yes | No | Yes |

* Total analytical time including analysing, checking and authorising
$^{\Psi}$ LCSH= Long contiguous stretches of homozygosity

All software packages use different analytical algorithms and the algorithm for ChAS was not provided by the company; therefore we are unable to list these in Table 5. The Genoglyphix software from Signature Genomics applied an NG packager segmentation algorithm, while Agilent software provided z-score, ADM-1, ADM-2 and CBS (circular binary segmentation). We used the ADM-2 algorithm in our analysis. OGT's CytoSure Interpret applied CBS as its algorithm, while BlueGnome's Bluefuse Multi used an algorithm specific to its own software, named multi v1.0 for the software version used in our comparisons.

Additional annotation tracks incorporated into the software are very useful providing additional information by clicking the annotation tracks relating to the abnormal regions. Table 5 lists the main annotation tracks of the five platforms. Another factor to be taken into consideration is the facility of adding custom tracks to the software, e.g. in house FISH and MLPA probes, a flexibility provided by most packages.

It is worth noting that all of the software packages have many positive features and from our comparison it is clear that none of the software packages can automatically perform all the tasks required for a full and comprehensive analysis. Human interaction (and interpretation) is still required.

### Detection rate and breakpoint resolution

All five software platforms detected copy number changes with varying levels of resolution, and mosaicism was detected to a level of ~ 30% abnormal cells. The Affymetrix array can also detect LCSH. Within the 12 samples tested, we detected LCSH in three patients using the filter settings defined above (page 10).

A total of 15 abnormalities from the 12 test samples as defined by our original 4x44k results were used for determining the inter-software detection rates. As shown in Table 6, the 2.7M Affymetrix array, ISCA 4x180k and NimbleGen CGX array detected all 15 abnormalities with 100% detection rate. The ISCA 8x60k failed to record a duplication of ~165kb on chromosome 6 q25.1 giving a 93% detection rate. The NGRL 4x44k array failed to detect an intragenic deletion within the *A2BP1* gene on chromosome 16 due to the low density of only 2 probes in the ~158kb deletion region..

Another factor affecting the detection rate is the reference DNA used. In all experiments, we used the Promega's pooled male and female control DNA samples for all the platforms we ran in our lab. For the NimbleGen CGX-12 array, Signature Genomics used an individual male and an individual female control.

Table 6. Comparison table for detection rate and breakpoint resolution

| Platform | Affymetrix | NimbleGen | Agilent | | |
|---|---|---|---|---|---|
| **Array** | **2.7M Cytogenetics Research Solution** | **CGX-12** | **ISCA 4x180k** | **ISCA 8x60k (OGT)** | **NGRL 4x44k** |
| What can be detected | Copy Number Changes | Copy Number Changes | Copy Number Changes | Copy Number Changes | Copy Number Changes |
| | Mosaicism | Mosaicism | Mosaicism | Mosaicism | Mosaicism |
| | LCSH* | | | | |
| Overall detection rate | 100% | 100% | 100% | 93%<br>Missing dup (6)(q25.1) on case 11 | 93%<br>Missing A2BP1 gene deletion on case 1) |
| Average breakpoint resolution (not mediated by known segmental duplications ) (bp) | 7665 | 27670 | 18307 | 43255 | 73073 |
| Average breakpoint resolution (mediated by known segmental duplications) (bp) | 43952 | 510138 | 24228 | 74030 | 272029 |
| Overall breakpoint resolution (bp) | 25809 | 268904 | 21268 | 58643 | 172551 |
| Average number of calls per patient | 3.5 | 9.67 | 34.2 | 7.3 | 2.67 |
| Calls on genic regions | 3.2 | 7.17 | 25.5 | 5.5 | 2.41 |
| Calls on non-genic regions | 0.3 | 2.5 | 8.7 | 2 | 0.16 |

* LCSH= long contiguous stretches of homozygosity

Different software settings, algorithms and/or annotation files used could also affect the detection rate and breakpoint resolution. For example, a mosaic chromosome 13 deletion in case 6 was not detected by the DNA Analytics on the NGRL 4x44k array using algorithm ADM-2. However, by changing the algorithm to ADM-1, the software did detect the abnormality.

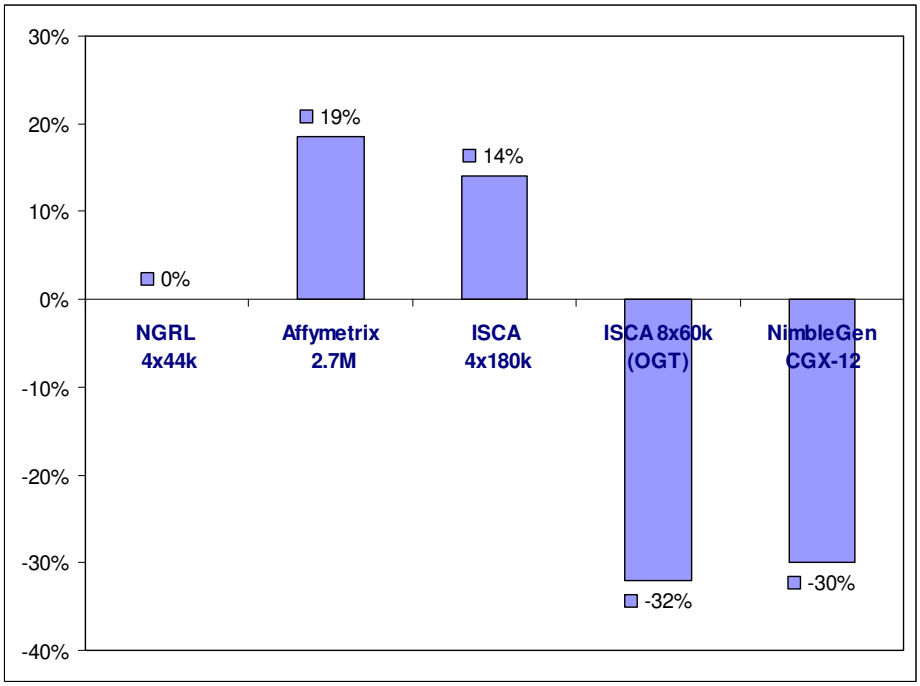
The average breakpoint resolution was calculated in two groups i.e. regions not mediated by known segmental duplications and regions mediated by known segmental duplications. From Table 6, it can be seen that the average breakpoint resolution in the non segmental duplicated regions is higher than those in the segmental duplicated regions and that it is also consistent with the array resolution. The Affymetrix 2.7M array gave the highest breakpoint resolution of 7665 base pairs on average. It also shows that the breakpoint resolution for the cases analysed in this study is not necessarily consistent with the platform's resolution, especially in the segmental duplication mediated regions. For example, the Affymetrix 2.7M array has the highest array resolution, but the breakpoint resolution in the segmental mediated regions is only about 44kb, which is lower than the ISCA 4x180k array (~24kb). This in part is explained by the use of a repeat masker with the Affymetrix software which makes analysis easier, but at the price of lowering the breakpoint resolution in these regions.

## Overall consumable cost

In addition to the optional resolution required for diagnostic aCGH, the consumable cost per sample is a factor when reaching a decision as to which platform(s) to use. We compared the overall consumable cost for the five platforms and expressed these by setting the cost for the NGRL 4x44k platform as the baseline against which the cost for all the other four platforms were compared (Figure 4). From this it can be seen that the Affymetrix 2.7M genomic solution and ISCA 4x180k arrays are the most expensive, while the OGT's CytoSure ISCA 8x60k and the NimbleGen CGX-12 arrays are significantly cheaper than the 4x44k array. Please note that the price comparisons are based on the published retail prices and include the costs of all consumables involved in aCGH. Further discounts may be available to users depending on the volumes being purchased.

Other factors that need to be taken into account include the laboratory infrastructure (scanner resolution, IT infrastructure, liquid handling robots etc) which may also significantly affect the final decision on which platform is to be used.

**Figure 4.** Comparison of the overall consumable costs

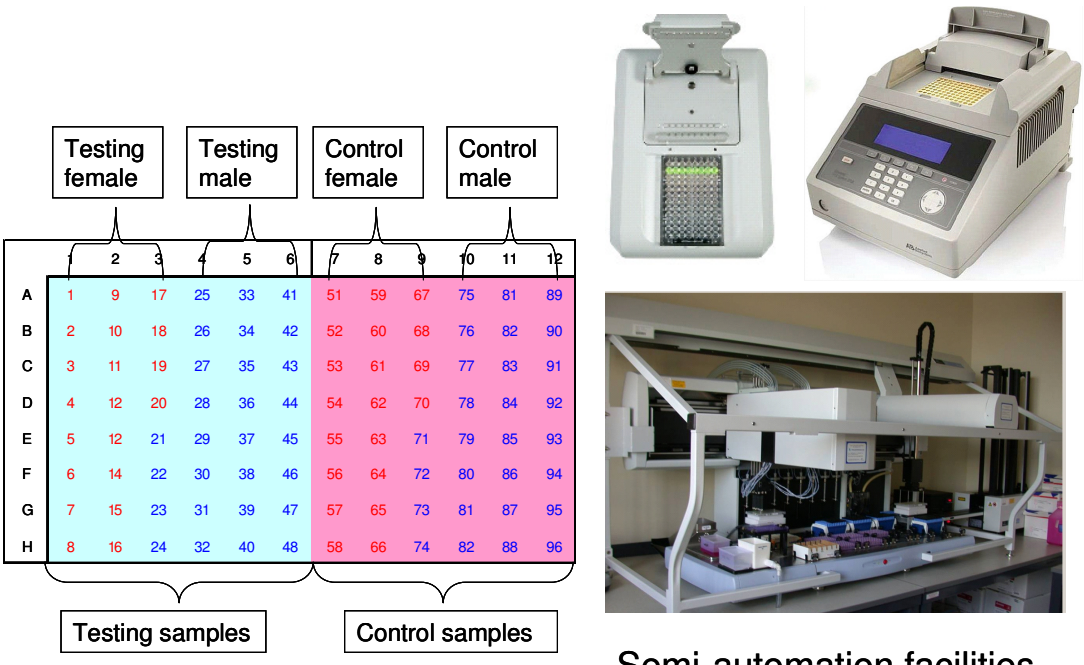



## Semi-automated, high throughput aCGH workflows

Based on the information derived from this project, and combined with the technological improvements recently achieved in the NGRL array laboratory, in February 2010, we developed, tested, validated and implemented a semi-automated workflow using the OGT's CytoSure ISCA 8x60k array platform.

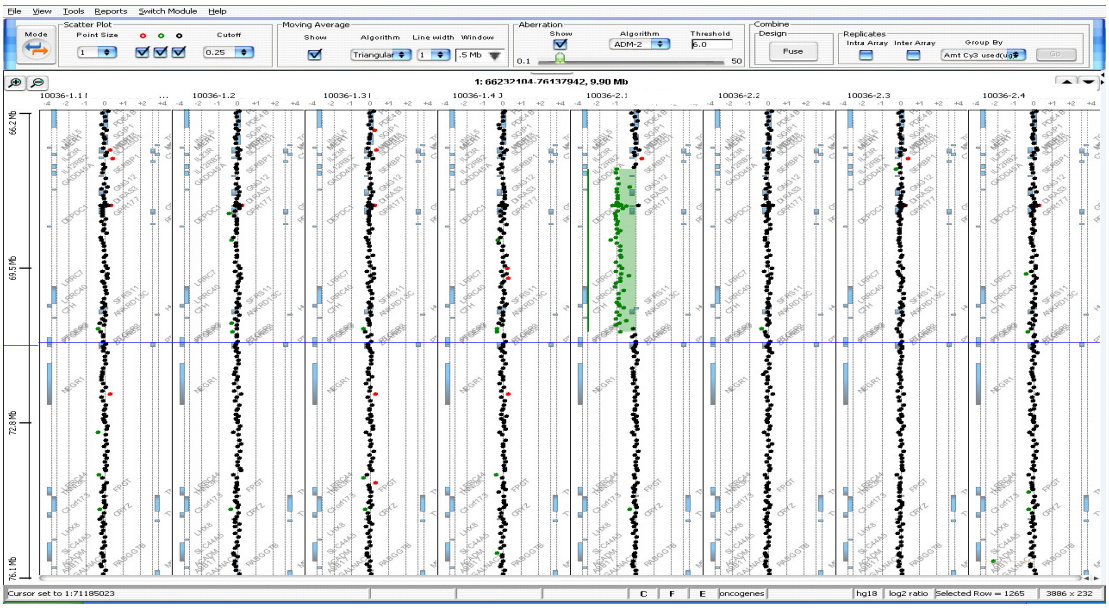
Prior to this work, aCGH was carried out essentially as a manual process involving multiple tube labelling, followed by transferring and pipetting, by one technologist handling a minimum of 8 cases simultaneously. We scaled up array processing by using a 96 well plate and a 96 well PCR machine for labelling, a NanoDrop® 8-Sample UV-Vis Spectrophotometer for DNA quality and dye incorporation evaluation, and a 96 well purification module for labelled DNA purification. In collaboration with the Wessex Regional Genetics Laboratory's high throughput laboratory, we aliquot equal amounts of DNA into 96 well plates robotically, which completes the process within minutes compared with hours required by comparable hand pipetting. The hands-on time for 48 samples using the semi-automated workflow is 5 hours (6.25 minutes per sample) compared with 7 hours for 8 samples manually (52.5 minutes per sample) i.e. an 8.4 times improvement in handling times. Figure 5 shows the basic elements of the semi-automated workflow.

**Figure 5.** Basic elements of semi-automated workflow




The robotic process produces stable and reproducible results. Figure 6 shows an image of 8 samples from one 8x60k array slide processed by the semi-automated workflow, from which the array quality is very high with an average DLRS for these 8 samples 0.11. Similar consistency has been achieved in over 500 consecutive samples with an overall failure rate of ~1%.

Figure 6. An example of OGT's CytoSure ISCA 8x60k array processed by semi-automation



## Conclusions

The evaluation provided here shows that all the platforms tested provided robust, accurate and good quality clinical aCGH results. This reflects significant investment by multiple vendors resulting in improvements in array design and quality together with rapidly developing software platforms (most of which have evolved further since this evaluation was carried out). The final decision on which platform to use locally may depend on a variety of factors, including cost, performance, local infrastructure and often local research and development interests.

At the WRGL we decided to migrate from the NGRL 4x44k to the ISCA 8x60k array platform provided by OGT and printed by Agilent. Using this configuration we have achieved an improved breakpoint resolution of ~ 3 times (Table 6), with an overall consumable cost reduction of 32% (Figure 4). Array CGH applications to other cytogenetic referrals are being investigated in the NGRL microarray laboratory, for example, evaluation of the 8x15k array platform for solid tissue samples is ongoing. More efficient and economic workflows are also being investigated.

One of the roles for the NGRL is to disseminate NGRL findings to the major diagnostic laboratories and the research communities in the UK. Please contact us if you have any queries related to this study or more broadly to the diagnostic application of aCGH.

## Contact details:

Shuwen Huang email: shuwen.huang@salisbury.nhs.uk

or John Crolla email: john.crolla@salisbury.nhs.uk

National Genetics Reference Laboratory (Wessex), Salisbury District Hospital, Odstock Road, Salisbury, SP2 8BJ

Tel: 01722 429098 Fax: 01722 338095

## Acknowledgements:

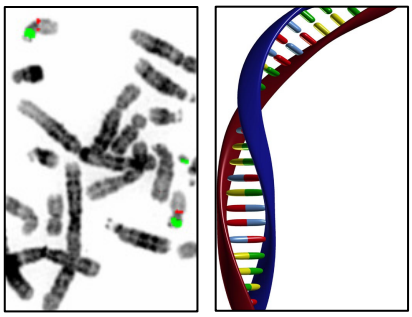
**1. All members of array team (in alphabetic order)**

Sarah Beal, Jocelyn Charlton, Annette Cockwell, Morag Collinson, John Crolla, Elizabeth Hird, Sarah Jones, Viv Maloney, Joanne Morris, Derek Richardson, Emma Jane Taylor

**2. All companies involved in this project**

## Reference

1. David T. Miller, Margaret P. Adam, Swaroop Aradhya, Leslie G. Biesecker, Arthur R. Brothman, Nigel P. Carter, Deanna M. Church, John A. Crolla, Evan E. Eichler, Charles J. Epstein, W. Andrew Faucett, Lars Feuk, Jan M. Friedman, Ada Hamosh, Laird Jackson, Erin B. Kaminsky, Klaas Kok, Ian D. Krantz, Robert M. Kuhn, Charles Lee, James M. Ostell, Carla Rosenberg, Stephen W. Scherer, Nancy B. Spinner, Dimitri J. Stavropoulos, James H. Tepperberg, Erik C. Thorland, Joris R. Vermeesch, Darrel J. Waggoner, Michael S. Watson, Christa Lese Martin, David H. Ledbetter, Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies, The American Journal of Human Genetics - 14 May 2010 (Vol. 86, Issue 5, pp. 749-764)



**National Genetics Reference Laboratory (Wessex)**
**Salisbury District Hospital**
**Salisbury SP2 8BJ, UK**

**www.ngrl.org.uk**