GEZONDHEIDSZORG VESALIUS
HOGESCHOOL GENT
LID VAN DE ASSOCIATIE UNIVERSITEIT GENT

Genomic analysis with array CGH

Comparison between saliva and blood as DNA sources

Biomedical Laboratory Technology
Option in Pharmaceutical and Biological Laboratory Technology
Professional Bachelor Project 2009
Supervisor: Dr. Nigel P. Carter
Co-supervisor: Dr. Stefan J. Vermeulen

Mieke Debaere

Acknowledgements

The past few months have flown by. I could not have wished for a better work placement than The Wellcome Trust Sanger Institute in Cambridge. Accompanied by team 70 at the Sanger Centre has thrown open the area of molecular cytogenetic research and techniques. I am grateful to each one of team 70 for the warm welcome, to pick me up in much pleasant and instructive working days, as well as being ready to answer all my questions. Particular acknowledgement goes to Dr. Nigel P. Carter and Diana Rajan. Their guidance, kindness, valuable time and contributions of knowledge out of strong experience were more than appreciative. Thank you also Dr. Nigel Carter for providing financial support. In addition, a special note of thanks goes to Tom Fitzgerald, who plotted the resulting microarray data.

I would like to extend my acknowledgement to Ancy Leroy and Dr. Stefan J. Vermeulen who made all of these possible. Especially Dr. Stefan J. Vermeulen has been a great support and encouragement throughout the internship, who I could always reach. Many thanks. Furthermore, indispensable gratitude goes to all of my friends and family for their advise and understanding during the quarterly internship abroad and completion of the thesis.

Finally, thank you Stefan Vermeulen and Diana Rajan for comments and suggestions on an earlier version of the thesis.

Table of contents

List of figures and tables

Abstract

If you can imagine that your DNA and my DNA - globally taken - differs in approximately 0.1 % of their total sequence, how could it then be possible that 2 randomly chosen humans can be such resembling creatures, without even any phenotypic similarity? Contemporary research is unveiling the phenotypic consequences due to a newly identified type of genetic variation, called copy number variation (CNV). Using microarray technology it has become apparent that submicroscopic deletion, insertion, duplication and more complex variations occur frequently in the human population. For studying these variants, array-based CGH is the most powerful technique for the genome-wide detection of copy number changes in comparison with reference DNA. In this thesis we compare human whole genome samples prepared from blood or saliva utilizing array CGH. We suggest saliva as a new source for collecting total genomic DNA in a user-friendly way. The Oragene™ DNA Self-Collection kit, for which the donor simply spits into the collection vial, has become widely available in recent times. DNA extracted from saliva by the Oragene method has several advantages over the usual method of DNA extraction from normal blood. Oragene saliva DNA and normal blood DNA were collected from two patients. We used an Agilent oligonucleotide array CGH platform to compare array CGH results from both DNA sources. Beyond one single blood related discrepancy located in chr14 q11.2, comparing the array CGH results between paired samples panned out very well. We conclude that the Oragene™ DNA Self-Collection kit can be used for array CGH. Moreover the Oragene kit offers an noninvasive method over blood collection for high resolution array CGH.

Keywords: *Oragene DNA Self-Collection Kit, saliva, DNA extraction, oligonucleotide array CGH*

1 The way to digital molecular cytogenetics

Patterns of genetic variability have been numerously examined in both cytogenetic and molecular genetic analysis [1]. Various forms, including variable number of tandem repeats (VNTRs), single nucleotide polymorphisms (SNPs), transposable elements and structural variants, would represent one thousandth of the human genome [2]. Nevertheless, an abundance of newly recognized structural variants have appeared resulting in a major source of common genetic variation.

The very first genomic differences were revealed thanks to karyotyping and chromosomal banding**.** These original cytogenetic techniques detected mainly rare abnormal variants in the quantity or structure of chromosomes, such as aneuploidy, ring chromosomes, chromosome translocations, inversions, duplications, deletions and fragile sites, which often turned out to be associated with specific diseases [3-6]. A refined characterization of these variants became clearer with the advent of fluorescence *in situ* hybridization (FISH) , a molecular cytogenetic technique allowing fluorescent labelled DNA probes to visualize target chromosome regions. Initially, metaphase FISH techniques were able to detect chromosome rearrangements at a resolution of ~5 Mb. In subsequent years the ability to locate FISH probes in interphase nuclei and on stretched chromatin DNA (fibre FISH) identified microrearrangements down to 50 kb and 5 kb respectively [1]. The increased resolution authorized an accurate characterization of microduplications (gains) and microdeletions (losses) which are a significant cause of disorders as well [7,8]. However, throughout these locus-specific studies the sensitivity of FISH is shown to be limited, as it requires prior knowledge of the target region in order to design complementary probes. Technical innovations introduced as multicolour FISH-based karyotyping, including multiplex FISH, spectral karyotyping, and comparative genomic hybridization (CGH) overcame these limitations by providing simultaneous analysis of all chromosomes.

Conventional comparative genomic hybridization (CGH) for genome-wide detection of DNA sequences that vary in copy number among individuals has been developed in the early 90s [9]. This approach, in which differentially labelled genomic DNA from a test and reference sample compete for *in situ* hybridization onto normal metaphase spreads, has proven useful in assessing chromosomal regions that are repeatedly gained or lost in tumours [10]. Copy number variation (CNV) is determined by the fluorescence ratio

between corresponding DNA sequences from hybridized test and reference DNA. Hence, another limited number of studies identified the presence of specific gains or losses that are not related to diseases at all [11,12]. Despite the fact conventional CGH has escalated the sensitivity to detect copy number changes, the resolution afforded by metaphase spreads limits this method of screening CNVs to less than ~5 Mb.

In array CGH, metaphase chromosomes are replaced by mapped DNA-sequences or oligonucleotides that are spotted robotically onto glass slides. For this methodology, the resolution is restricted only by the density of spotted sequences and by their size. Over the past 10 years, arrays have been constructed using large-insert clones (40 - 200 kb in size), small insert clones (1.5 – 4.5 kb), cDNA fragments (0.5 – 2 kb), and genomic PCR amplicons (100 bp - 1.5 kb) [13]. More recently, on-chip synthesis technology of oligonucleotides in the 25-80 bp range has achieved a previously unattainable detection resolution. In this way, array-based CGH combines the whole-genome screening capacity of conventional CGH with a hugely enhanced resolution.

Table 1| Comparison of probe size and density in CGH arrays.

CGH array	Sequence size	Target DNA sequences	Resolution	Genome coverage
cDNA array	0.5 – 2 kb	cDNA clones	276 kb	all genes
BAC array	80-200 kb	3.000 BAC clones	1 Mb	whole genome
ROMA array	70 bases	85.000 oligonucleotides	30 kb	reduced-genome
Oligonucleotide array	60 bases	30.000 oligonucleotides	63 kb	whole genome
BAC tiling array	80-200 kb	36.000 overlapping BAC clones	50 kb	whole genome
PCR-product array	100 bp-1.5 kb	60.000 PCR products	15 kb	whole genome
In situ synthesized oligonucleotide array	60 bases	> 244.000 oligonucleotides	< 12 kb	whole genome

Array CGH methods have been widely employed to detect copy number alterations in patients with solid tumours, mental retardation, subtelomeric rearrangements and other unbalanced constitutional rearrangements [14-18]. Although several molecular techniques (MLPA, MAPH, and other) can be used for CNV identification, array CGH currently offers the most cost-effective and robust methodology.

2 Copy number variation in the human genome

2.1 Definitions

2.1.1 DNA copy number variation

Copy number variation (CNV) is defined as a DNA segment that is minimum 1 kilobase to several kilobases or megabases in length, for which the copy number is variable relative to a reference genome. The definition of copy number variation contradicts what previously had been thought; it was assumed that human genes are mostly present in two copies, with one copy inherited from each parent, till recent technology has uncovered genes occurring in one, three or more copies on an intermediate-scale.

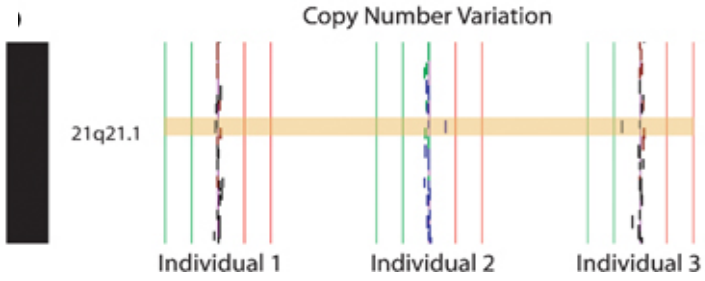


Figure 1| Illustration of copy number variation in 3 healthy individuals at chromosomal region 21q21.1 [19]. At the same chromosomal region and relative to the same reference genome, no copy number variation is shown in individual 1, whereas individual 2 and 3 demonstrate respectively duplication and deletion events.

Different subtypes of CNV comprise deletions, duplications, insertions and multi-allelic CNVs. Some CNVs are strictly defined by DNA breakpoints, others overlapped by copy number variant regions (CNVRs) originating from various molecular mechanisms [20].

2.1.2 DNA structural variation

Structural variation is the term that encompasses microscopic variants (≥ 5 Mb in size) while submicroscopic variants range from ~1 kb to 5 Mb [21]. The submicroscopic variations, including copy number variation, intermediate translocations and inversions, and segmental duplication bridge the gap between microscopically visible variants and small-scale sequence variants (1-700 bp) [22].

2.2 The extent of copy number variation

Small-scale sequence variations have been readily detected by DNA sequencing analysis in the field of molecular genetics, involving variants altered at a single base pair location, namely, single nucleotide polymorphisms (SNPs). Until recently, it was presumed that SNPs were the most prevailing form of human genetic variation and constituted much normal phenotypic variation. However, hardly 5 years has passed since the ubiquity of CNVs in healthy individuals was unveiled [23,24]. The low frequency of earlier DNA copy number variation observed by traditional cytogenetic analysis has been extended to a recognized prevalence through applying whole-genome scanning array technologies. Using 2 different microarray platforms, Redon *et al.* established the first-generation global map of CNV's and reported an astonishing 12 % of the human genome that is present in a variable copy number [25]. Clearly, copy number variation composes a substantial portion of genetic variation.

2.3 The importance of copy number variation

2.3.1 Overall influences of structural variation on phenotype

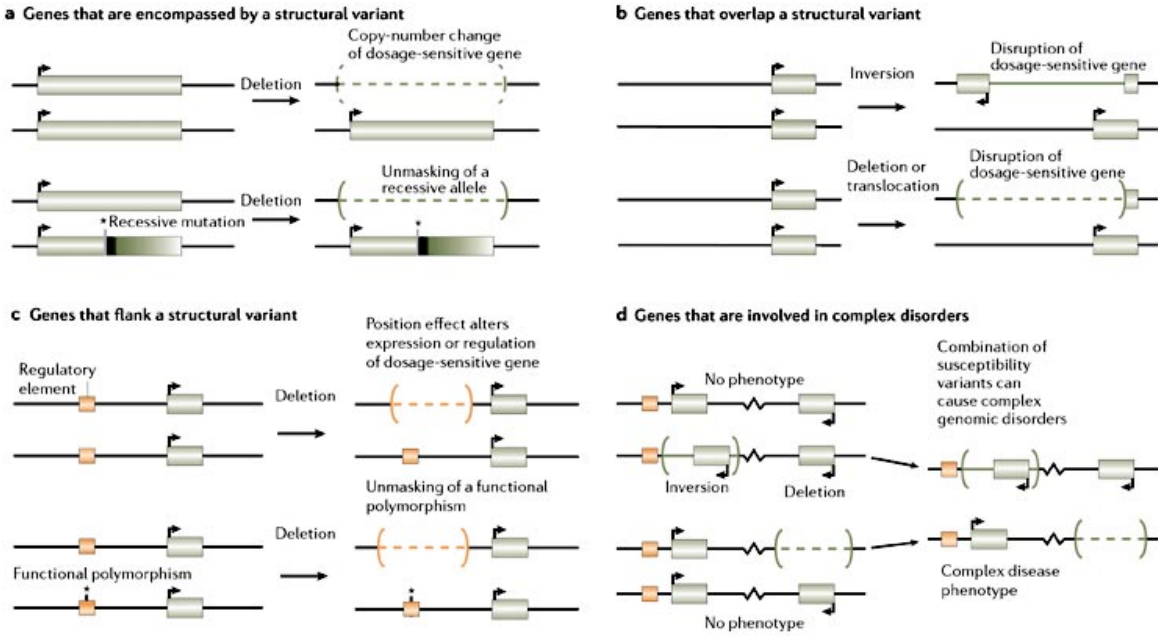


Figure 2| Functional impact of structural variation in the human genome [21]. a. Dosage-sensitive genes can produce higher or lower levels of gene product under the influence of a structural deletion (upper panel) or duplication. Dosage-insensitive genes may expose a recessive allele if the gene is deleted in one of the paired alleles (lower panel). **b.** Coding sequences containing inversion (upper panel), deletion or translocation (lower panel) variants, become disrupted in a manner of reduced gene expression. **c.**Structural variants also indirectly affect gene expression in the vicinity, either by deleting gen regulation or uncovering functional polymorphisms. **d.**Two susceptibility variants, expressing no phenotype of onset, can contribute to disease in a next generation

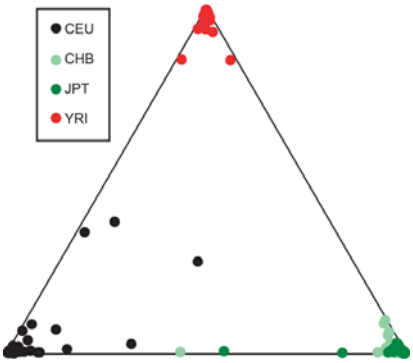
2.3.2 Phenotypic effects of pathogenic CNVs

Owing to their size, spanning thousands to millions of base pairs, it is likely that regions of copy number variation often contain and/or disturb functional DNA sequences. In this manner CNVs influence gene expression, alter gene dosage, cause genomic disorders, etc., all of which can lead to dramatic phenotypic consequences. Disease-causing CNVs often are implicated with critical developmental genes. Mental retardation, schizophrenia and autism are just a few human diseases emanating from recurrent rearrangements located within imbalanced genomic regions, which are listed under the term 'genomic disorder'. An interactive web-based database called DECIPHER shares CNVs associated with their clinical conditions from anonymous patients across the world to facilitate diagnosis [26].

2.3.3 Phenotypic effects of beneficial CNVs

While array CGH has been mainly applied in the detection and mapping of copy number changes, its current implementation in the detection of CNVs and the phenotypic association might prove valuable towards human evolution and uniqueness. There appears to be an enrichment of sensory-, immune- and inflammatory receptor genes in CNVRs in such a way that CNV may account for individual drug response, disease resistance and susceptibility [24,27]. Most CNVs have been found surrounding such genes. And although research is still in its infancy, variable copy number is expected to be contributing significantly to inter-individual expression variation [28].

2.4 Population-specific copy number variation



The global CNV map generated by Redon *et al* is the map with the highest genome coverage so far [25]. By screening 270 individuals of 4 worldwide HapMap populations [42] on copy number variants already revealed the presence of stable copy number polymorphisms (CNPs) in ancestral populations of America, Europe and Africa.

Figure 3 | Triangle plot showing 3 clusters, respectively in the corners of the triangle, of CNV genotypes for 4 populations[25]. HapMap populations include Africa (YRI), Europe (CEU), Japan (JPT) and China (CHB)

The Database of Genomic Variations (DGV) is in the process of recording all copy number variations reported in certain populations [29]. A well-known example of specific CNVs in the African ancestry is located on the functional CCL3L1 chemokine receptor genes. In this particular case low copy number of CCL313 genes has been associated with an increased susceptibility for HIV viruses [30].

2.5 How to quantify copy number variation ?

Array CGH experiments measure copy number changes in a test genome by pinpointing intensity differences in the hybridization patterns of test (patient) and reference (normal) DNA. Usually microarray data is shown in logarithms with base 2. The transformation of raw intensity values between typically 1 and 2^{20} into log intensities provides a compressed fold-change range. Any difference between the logarithms of the test and reference signal intensities (the log ratio) reveals e.g. one, two or three-fold changes in DNA copy number. Log_2ratio (test/reference) intensities are digitally quantified for any chromosomal region. Amplified chromosome regions are screened with a positive log_2ratio value, as opposed to deleted regions which are represented by negative log_2ratios throughout the genome. For example single copy gains in the test genome typically have a log_2 ratio (3/2) of 0.58, while single copy losses in the test genome can be detected by a log_2 ratio (1/2) of minus 1.

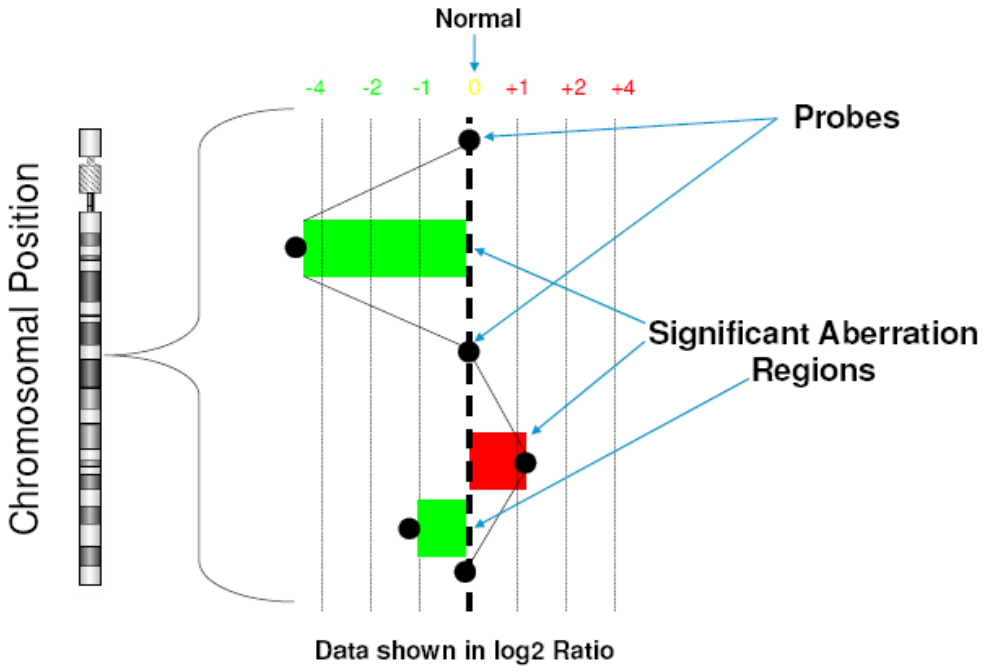


Figure 4 |Quantification of copy number changes in a scatter plot of log_2 (test/reference) intensities [31]. Log_2 (test/reference) intensities are described for all hybridization targets (probes) in the array CGH technique. Intensity ratios deviating from zero indicate potential copy number variations in each chromosome of interest.

3 Principle of Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) has been recognized as a valuable approach for detecting copy number imbalances across an entire genome. The general principle in CGH is based on the co-hybridization of differentially labelled test and reference DNA to a third source of DNA referred to as probes and attached onto 75 x 25 mm glass slides. Usually the test and reference DNA become labelled with respectively red and green fluorophores that emit fluorescence at a distinct wavelength. Thereby, the addition of unlabelled Cot-1 DNA blocks repetitive sequences. If equal amounts of test and normal reference DNA are then applied to one glass slide, *in situ* hybridization of the two samples performs in proportion. Fluorescent signals will point out genomic regions of the test sample that are hybridized in more or less extent compared to the reference; red fluorescence is predominantly at regions where test DNA is amplified, while green fluorescence dominates where the test genome is deleted. Genomic regions where test and reference are hybridized in equal amounts appear yellow. The relative intensities of red and green fluorescent signals along the chromosome must be quantified in order to reveal differences in intensity ratios ; copy number changes. Over the past 20 years, assays for CNV detection based on the CGH principle have evolved from using conventional CGH to the increasingly more comprehensive array-based CGH.

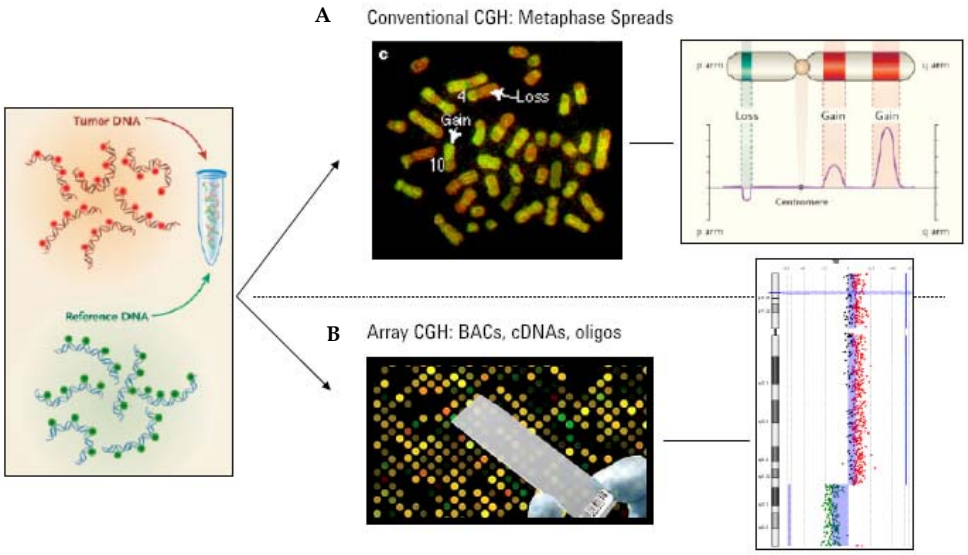


Figure 5 | Enhanced performance from classical CGH to array-based CGH [31]. A. *LABELLING DYES*: rhodamine or Texas red (red) and FITC (green) *TARGET DNA:* normal metaphase chromosomes *LIGHT SOURCE:* lamp, *DETECTION:* epifluorescence microscope and digital image analysis, *RESOLUTION:* 5 Mb, **B.** *LABELLING DYES*: Cy5 (red) and Cy3 (green) *TARGET DNA:* BAC clones, cDNA clones or oligonucleotides, *LIGHT SOURCE:* laser beams, *DETECTION:* scanning and image analysis, *RESOLUTION:*1 Mb -15 kb

4 Oligonucleotide array CGH

At present, oligonucleotide probes provide the highest detection resolution for array CGH. While in array CGH the resolution is determined by the number and size of sequences on the array, densely packed oligonucleotides of typically 60-70mers length obtain an optimal resolution for CNV detection. Microarrays containing hundreds of thousands oligo probe sequences can be designed to unique regions known across the entire human genome [32].

4.1 Early oligonucleotide arrays

Originally oligonucleotide arrays were designed using oligonucleotides of 20 to 30 bases in order to detect SNPs. Even though SNP arrays reveal information for both SNP and CNV, it is unlikely to avoid cross-hybridization by multiple short oligonucleotide probes [33]. Instead long oligonucleotides (60-70 bases) provide much more sequential combinations to represent the whole human genome sequence, thereby improving hybridization specificity.

Spotting microarrays with 70-mer oligonucleotide probes, was first performed in a method called representational oligonucleotide microarray analysis (ROMA) [34]. With ROMA, the complexity of the input DNA is reduced prior to hybridization. In this way the probability of cross-hybridization further decreases as well as background noise outcome is tempered. Although copy number variation is better assayed with improved signal-to-noise ratios, the complexity-reduced sample and reference genome no longer represents the entire genome.

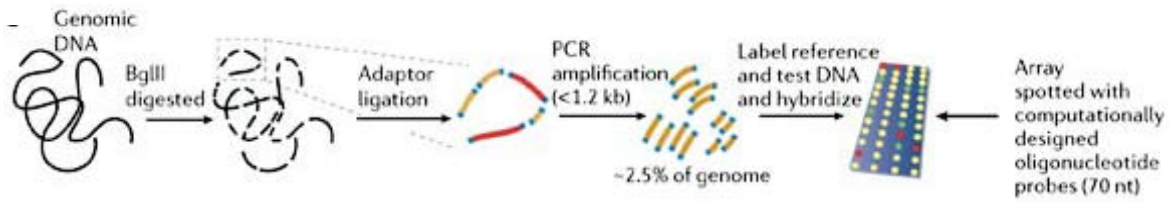


Figure 6 |Schematic overview of representational oligonucleotide microarray analysis(ROMA)[21]. In this approach, sample and reference gDNA are fragmented with B*gl*II restriction enzymes, fragments are ligated to adapters, which are then amplified by linker-mediated PCR using universal primers. However, PCR conditions are set only to amplify fragments smaller than usual B*gl*II fragments of 1.2 kb (yellow). Fragments of greater size (red) are lost, reducing the complexity of DNA samples. Approximately 2.5% of the sample and reference genomes is left to be hybridized onto oligo arrays, which are designed to match the representation fragments.

4.2 Agilent oligo CGH microarray

4.2.1 Agilent array technologies

Agilent is amongst one of the world's largest manufacturers of microarrays who supplies a complete package from reagents to software necessary for array CGH performance. Unlike ROMA arrays, their technology has generated long oligonucleotide arrays for direct CGH analysis of whole genomic DNA samples. The production of 60-mer length oligonucleotide microarrays covering the entire human genome is brought forth from ink-jet technology . This methodology enables on-chip synthesis of oligonucleotide probes with an extremely high density, as well as in any desired oligo sequence.

4.2.2 Microarray design

Researchers can apply for either catalog or custom microarrays. The array probe design of the former is determined by Agilent technologies. Catalog arrays have been developed to cover the human genome sequence in a gene-centric way, since all gene bases are certainly known. For the customized arrays a web-based application called eArray is available. The eArray online tool lets researchers design custom microarrays by choosing from a major database of validated probe sequences. For example in CNV determination, probes should be spaced more evenly throughout the genome including regions of segmental duplication and other repetitive sequences. Depending on experimental needs and taking advantage of Agilent's flexible microarray platform, any array design can be created.

4.2.3 Microarray format

Agilent offers a variety of microarray formats. In a way to greatly reduce experimental costs, multiple arrays may be manufactured on a single 75 x 25 mm microarray slide.
For example : 4 x 44k formats provide 4 individual arrays of 44.000 probes on one single microarray slide. Therefore 1 x 244k, 2 x 105k, 4 x 44k or 8 x 15k formats allow the simultaneous examination of one, two, four or eight samples respectively. The highest resolution is obtained in format 244k microarray, containing 244.000 oligonucleotides that cover the whole human genome with one probe every 12 kb.

Lately 1M microarrays have been launched with an increased density up to one million probe sequences, supplying an even more comprehensive array.

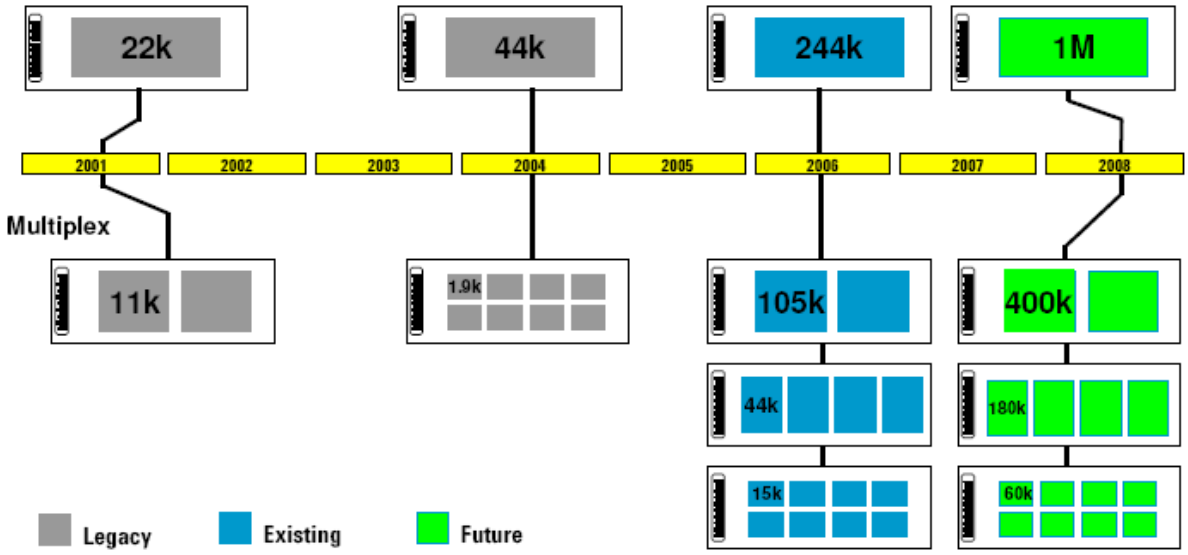


Figure 7 | The history of microarray formats supplied by Agilent [31]. The first microarray formats released in 2001 and 2004 are not longer available. For now, four standard formats of 1x 224k, 2 x 105k, 4 x 44k and 8 x 15k provide a resolution up to 12 kilobases. The latest third generation of Agilent microarrays achieves an even higher precision, involving arrays of 1 million probes and several multipack formats: 2 x 400k, 4 x 180k, 8 x 60k.

4.2.4 Microarray fabrication

4.2.4.1 Immobilization of oligonucleotides on a glass substrate

The chemistry employed for fixing oligonucleotides on a glass surface is accomplished via phosphoramidite nucleosides. Phosphoramidite nucleosides include nucleic acid analogues modified on reactive amine and hydroxyl groups, in particular on the 5'-hydroxyl group of deoxyribose bearing a dimethoxytrityl (DMT) protecting group. The microarray surface on which the very first layer of nucleosides will bind, should be covered with a hydrophobic and density-optimized substrate [35]. Automated synthesis of oligonucleotides directly onto the specially-prepared glass slides is a three-step process. The first step involves the coupling of a phospharamidite nucleoside to the glass substrate; the 2nd step involves the oxidation of the bound molecule and the 3rd step involves the activation of the nucleotide. Thereafter this cycle is repeated to allow the growth of the desired oligonucleotide probe.

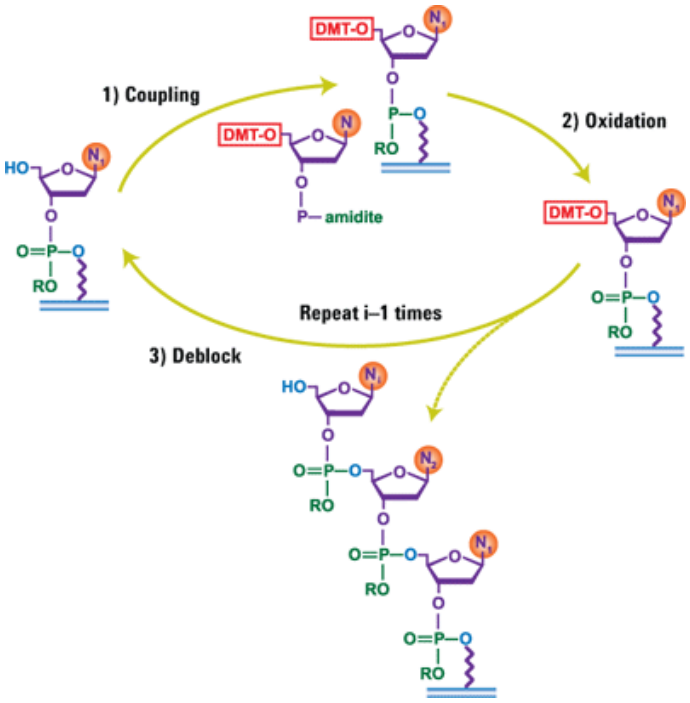


Figure 8 | General cycle of oligonucleotide synthesis via phosphoramidite nucleosides [36]. 1) COUPLING The first phosphoramidite nucleosides become attached to an available –OH group on the array substrate. With the DMT group at the 5'-OH location, nucleosides are protected from coupling to 3'-groups on other incoming nucleosides. OXIDATION After the first layer of nucleosides is linked to appropriate locations on the substrate, every trivalent phosphite group at the 3' location is oxidized into a pentavalent phosphate group. **3)** DEBLOCKING By means of de-blocking agents DMT groups are removed and oxidized. Unprotected 5'-OH ends may react with specific nucleosides of the next coupling reaction. The cycle needs to be repeated per desired monomer.

Phosphoramidite chemistry is ideal to synthesize short nucleic acid chains. This type of synthesis in the 3'-5' direction allows sequences up to 100 bases. For constructing 60-mer length oligonucleotide arrays, the process shown in Figure 8. requires 60 repeated cycles. In between each step, excessive reagents must be washed away in order to prevent random reaction later in the synthesis. The entire array is flushed with wash solution, oxidizing and de-blocking agents through a microfluidic channel that exposes uniform doses of reagents. The microarrays undergo a final deblocking step so that all protecting groups are removed.

4.2.4.2 *In situ* synthesis printing process

The four different nucleosides, A, C, G, and T are delivered in picoliter volume droplets using non-contact inkjet printing. Agilent's SurePrint technology employs 4 inkjet heads and reservoirs containing the four types of phosphoramidites nucleosides separately.
As the printing process is similar to the printing technology of Hewlett Packard on paper, oligonucleotides become 'printed' onto the microarray surface by base-to-base synthesis. The order in which the oligonucleotide chains grow one nucleotide at a time, is obtained from digital sequence files. As a result, multiple layers of all the four nucleosides printed

simultaneously build *in situ* oligonucleotide probes with each a specific DNA sequence. This type of inkjet printing not only offers a high-precision spot placement, reproducible construction, denser coverage or non surface-contact anomalies, but clearly as well a great flexibility. Only a different digital sequence file is required to print another array design.

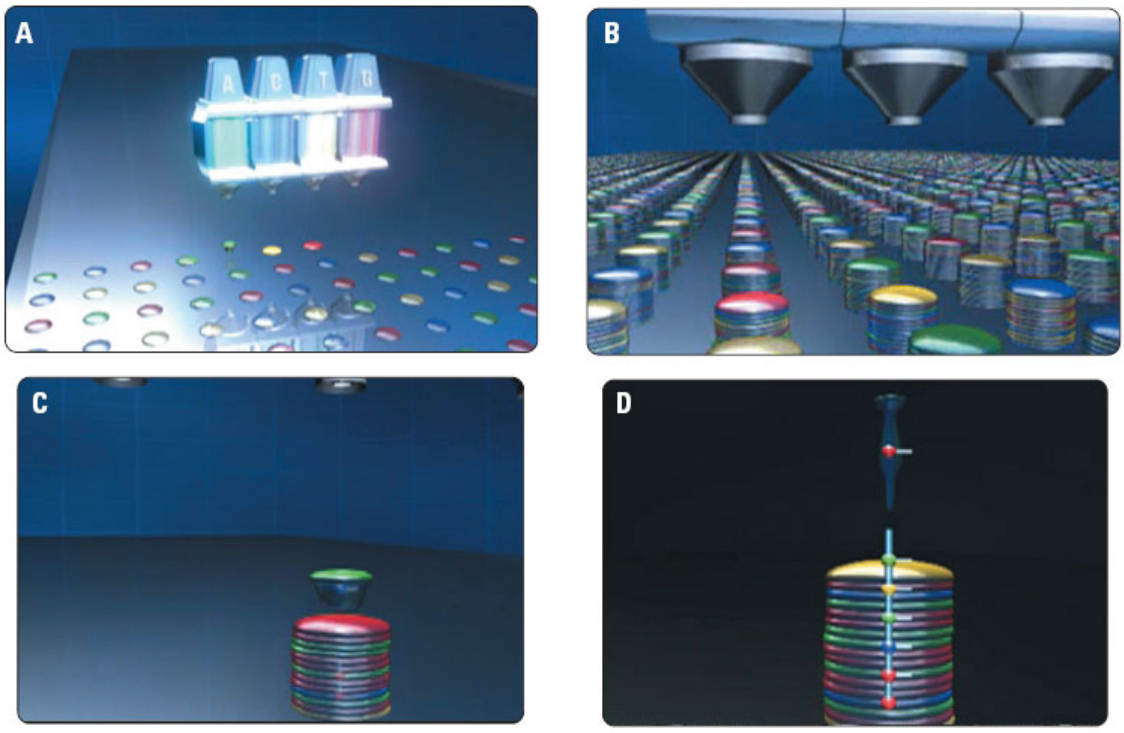


Figure 9 | I*n situ* synthesis printing process by SurePrint technology [36]. A. The first layer of nucleoside phosporamidites is being spotted onto the microarray surface. **B.** Unique oligonucleotide sequences have been printed precisely and are growing in multiple layers onto the array. **C.** A droplet of nucleoside is being added to one oligonucleotide chain. **D.** The addition of the new droplet to the oligonucleotide chain assembly in close-up.

4.3 Advantages and drawbacks

Oligonucleotide-based array CGH enables the whole genome analysis for copy numbers at an unprecedented resolution. I*n situ* synthesized oligo arrays overcame the limitation of microarrays that were spotted with presynthesized oligonucleotides, having a maximum number of 60.000 probes. Moreover *in situ* synthesis processes offer new collaborations between the researcher and manufacturer regarding the experimental design.
A major drawback of oligonucleotide arrays, however, is the poor signal/noise ratio. High background noise levels occur easily when sample and reference genomes not specifically hybridize to arrayed oligonucleotides. Although increasing the length of probes improves the hybridization specificity in oligoarrays, BAC clone-based arrays usually have a fivefold lower signal to signal/noise ratio.

5 Assessment of microarray data quality

Microarray data quality can be affected by several factors including the biological source and quality of DNA samples, slide handling, environmental conditions, scanner sensitivity. Systematic variability such as dye effects can be corrected by normalization procedures, whereas other influences produce noise. The assessment of data quality is an essential step in microarray analysis.

5.1 Quality control parameters

5.1.1 Quality control report

A Quality Control (QC) report generated by Feature Extraction software provide statistics for each microarray to evaluate the wet lab performance: DNA labelling, hybridization and washing. The Feature Extraction program has been designed to identify high quality signal intensities, distinguish outliers, remove background noise, and normalize intensity values. Following quality control statistics are only appropriate for Agilent CGH arrays.

5.1.1.1 Spot finding of the four corners of the array

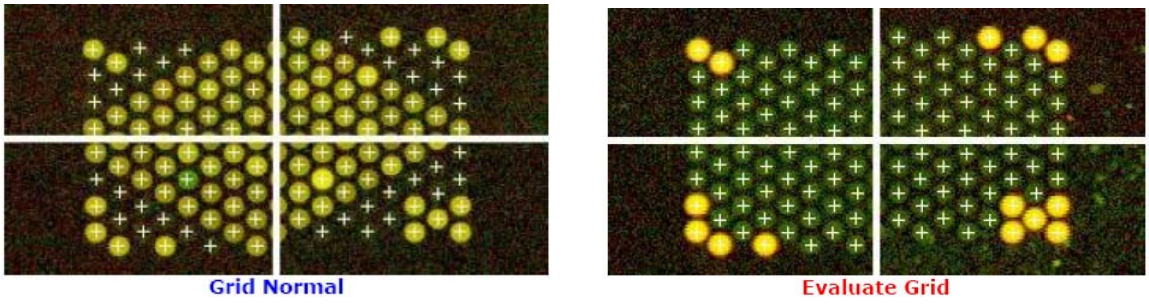


Figure 10 | QC Report – Microarray grid alignment. The two images involve the detection of spots in the 4 corners of the microarray. The words 'Grid normal' appear below the image in case grid marks line up exactly over the array spots. 'Grid evaluation' is recommended if grid marks appear arbitrarily at the 4 array corners.

Based on the barcode of the array, the Feature Extraction program assigned a grid template to the scanned microarray image file. The spots have been located properly if grid marks can be seen in the four corners of the array. If not, you may have to run the extraction with a new grid.

5.1.1.2 Outlier statistics

Feature	Red	Green	Any	% Outlier
Non Uniform	94	101	103	0.04
Population	51	61	83	0.03

Figure 11 | QC Report – Outlier measurements. Outliers deviate significantly from the rest of the array data. The number of non uniform outlier values within a spot or background population outliers lead to incorrect log ratio results and need to be excluded

Outlier values are the source of experimental noise. Only high numbers of outliers require attention and indicate the need to check the hybridization or washing steps.

5.1.1.3 Spatial distribution of all outliers on the array

a. **b.**

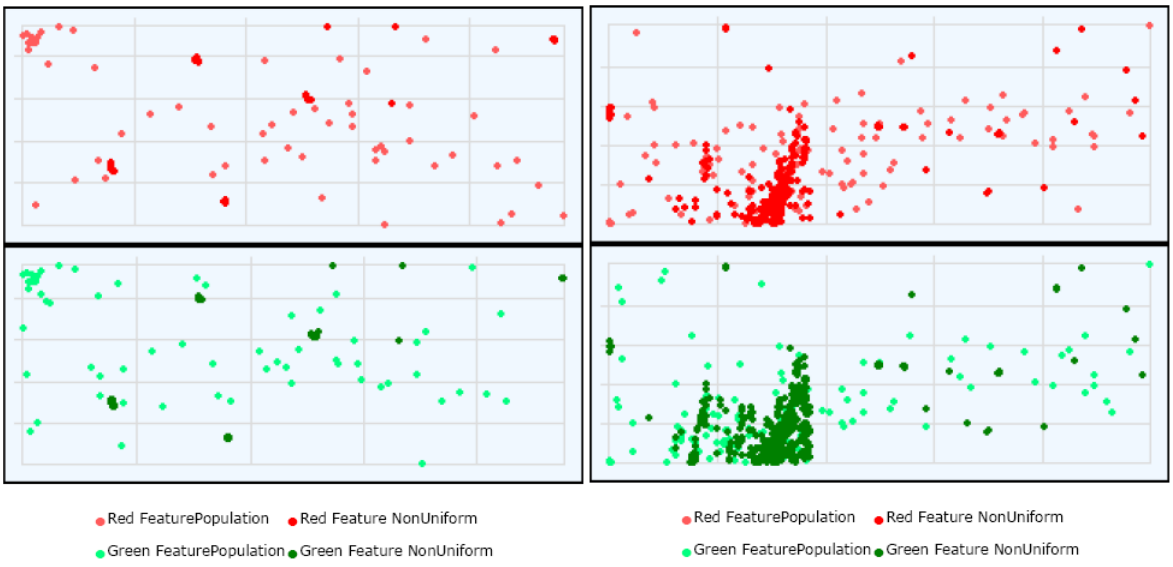


Figure 12 | QC Report – Spatial distribution of all outliers. **(a.)** Even distribution **(b.)** Uneven distribution**.**

The QC report displays the location of all outliers on the array for the red (Cy5) and green (Cy3) channel. The positions of both population and non uniform outliers are shown in the two plots. For each array, outliers are expected to be few in number and appear at random. Unevenly distributed outliers caused by non uniform outliers clustering together usually points at wash artifacts onto the array.

5.1.1.4 Spatial distribution of positive and negative log ratios

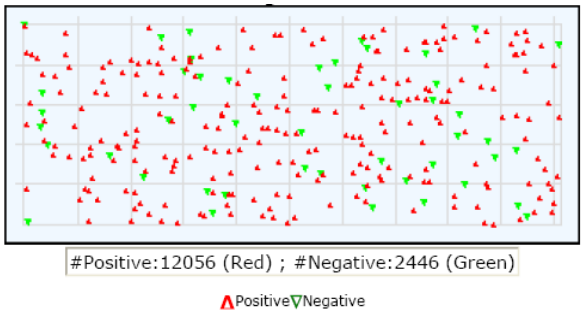


Figure 2 | QC Report – Spatial distribution of positive and negative log_2 ratios

In Figure 13 you can view the distribution of the significant positive and negative log_2 ratio measurements as they are found on the actual array. If the microarray data contains more than 5000 data points, the Feature Extraction program randomly selects 5000 log_2 ratios in the same proportion. The total number of spots producing either positive log_2 ratios (red) or negative log_2 ratios (green) is shown below the plot.

5.1.1.5 Histogram of signals plot

a. **b.**

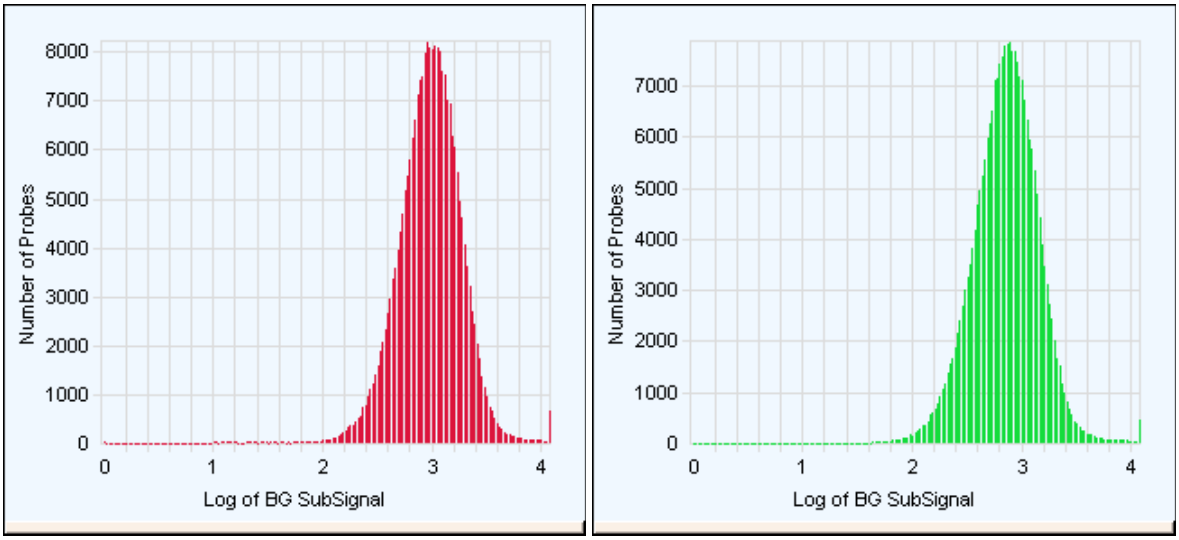


Figure 3 | QC Report – Histogram of signal plots. a. Red signals (Cy5) plot **b.** Green signals (Cy3) plot

When background intensity is subtracted by Feature Extraction protocols, the number of spots are plotted against base 2 logarithms of each intensity value. Reporting normalized intensity distributions for the red (Cy5) and green (Cy3) signals across the array monitors the hybridization quality. Ozone exposure and non-specific hybridization respectively may lower and broaden the intensity distribution.

5.1.1.6 Red and green background-corrected signals

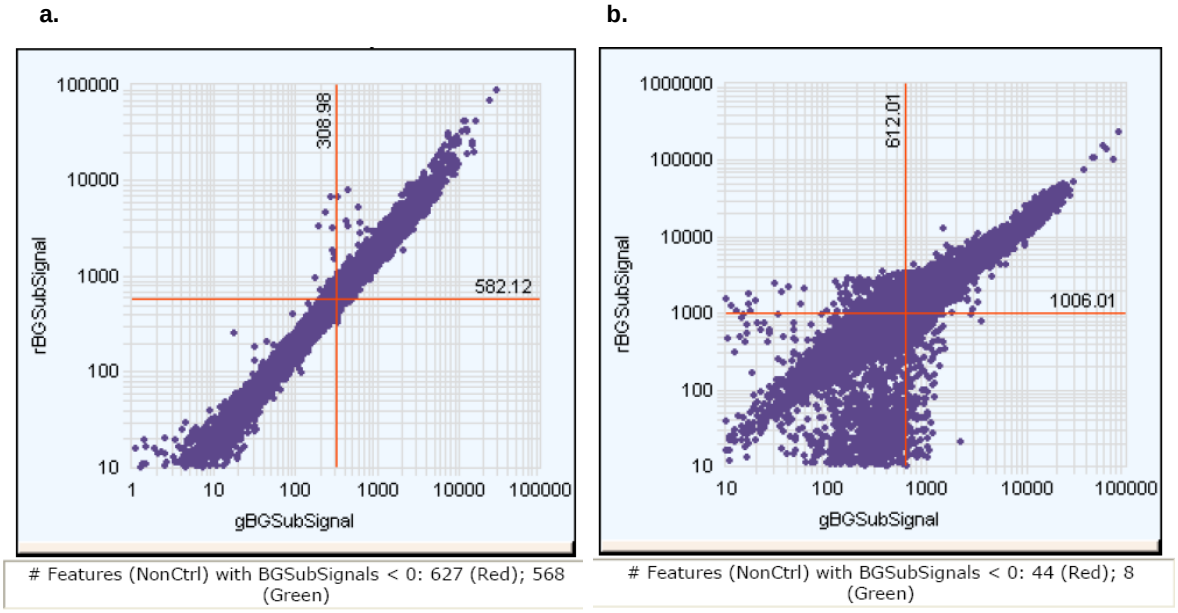


Figure 4 | QC Report – Plots of background-corrected signals. a. Self-self hybridization (female-female) **b.** female(Cy5) versus male (Cy3) hybridization

Figure 15 presents two logarithmic (base 2) plots of normalized intensity values with the red signals on the vertical axis and the green signals on the horizontal axis for inlier data. Ideally all data points, each representing a single spot, should have equivalent intensities. Clusters of data points deviating from the resulting linear curve with a slope of 1 indicate false-positive or false-negative results. Except high Cy3 intensities of the Y chromosome in comparison with female sex chromosomes are clustered together beneath the curve. Values below each plot notify the inliers that have a background-corrected signal less than zero, pointing out the number of high non-specific hybridization signals.

5.1.2 Quality control metrics

Microarray QC metrics, which are measurements of the overall data quality, are exported to a table in the Feature Extraction QC report and in DNA Analytics. These metrics rely on positive and negative control probes integrated in every microarray probe design from Agilent. In Table 2 the standard guideline is presented by which excellent, good and poor quality data can be validated. All metrics are computed after outliers have been removed.

Table 2 | The normal ranges of quality control metrics only appropriate for Agilent CGH arrays.

QC Metrics	Excellent	Good	Poor
BGNoise	<5	5 - 10	>10
Signal Intensity	>150	50 - 150	<50
Signal to Noise	>100	30 - 100	<30
Reproducibility	<0.05	0.05 - 0.2	>0.2
DLRSpread	<0.2	0.2 - 0.3	>0.3

5.1.2.1 Background noise

Negative control probes were printed onto the array for an estimation of background noise. This metric is calculated for each channel as the standard deviation of the negative control probes. High background noise disturb the specific signal intensities and can result from poor slide handling, long hybridization time, contaminated reagents or problems with the wash procedures.

5.1.2.2 Signal Intensity

For each channel, the median signal intensity value is reported from which background intensity has been subtracted. The array cannot be interpreted if the signals are too low. Low signal intensities are likely due to poor-quality DNA samples, ozone degradation, or losses of DNA during the labelling and purification step.

5.1.2.3 Signal/noise ratio

The signal-to-noise ratio, resulting from dividing signal intensity over background noise, has an important effect on the ability to detect copy number changes. Only if this metric exceeds a value greater than 100, copy number variations can be quantified accurately.

5.1.2.4 Reproducibility

To evaluate the reproducibility of signal intensities across the array, microarray designs of Agilent include spike-in probes. These oligonucleotides were spotted in multiple copies on the array such that signal intensities can be compared at regions with similar copy number. Serving as a positive control, a number of 200 probes are replicated across the entire array.

At least 3 inliers are required to calculate a coefficient of variation for one probe sequence. The reproducibility value is determined by the median of those coefficients and reported for each colour channel. High reproducibility is indicative of poor hybridization mixing, immobile bubbles or leakage between microarray and gasket slide.

5.1.2.5 Derivative log ratio spread

The most important quality parameter in array CGH is the Derivative Log Ratio Spread. With this metric, a more robust measurement of noise has been made by calculating the standard deviation of intensity ratios. Hereby the spread of intensity ratio differences is determined from probe to probe so that, because most pairs of adjacent probes have the same copy number, the majority of differences between intensity ratios are just noise. In that manner, the DLRSpread is a representing value for the overall data quality of each independent array. The higher the standard deviation, the lower the detection sensitivity.

5.2 Environmental impact on microarray data quality

Environmental factors such as light, ozone and high humidity levels may have deleterious effects on microarray data quality due to the susceptibility of Cy3 and Cy5 fluorescent dyes [37]. The stability of cyanine fluorophores on the dry surface of microarray glass slides can be easily affected by exposure to high humidity and room fluorescent light. However one of the most common problems is the degradation of Cy5 signals caused by external ozone. Cyanine 3 is much less sensitive to ozone just because the ozonolysis reaction is thought to occur at the carbon-carbon double bonds joining the two aromatic rings, which is one bond shorter than in the structure of cyanine 5.

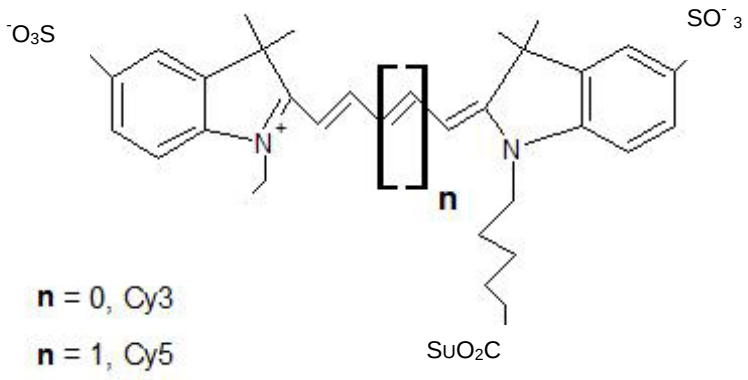


Figure 5 | Chemical structures of Cy3 and Cy5.

Particularly during summer months higher ozone levels will rapidly (10-30 seconds) oxidize Cy5 dye molecules. Throughout the laboratory ozone levels should be monitored to keep the concentrations as low as 5 to 10 ppb.

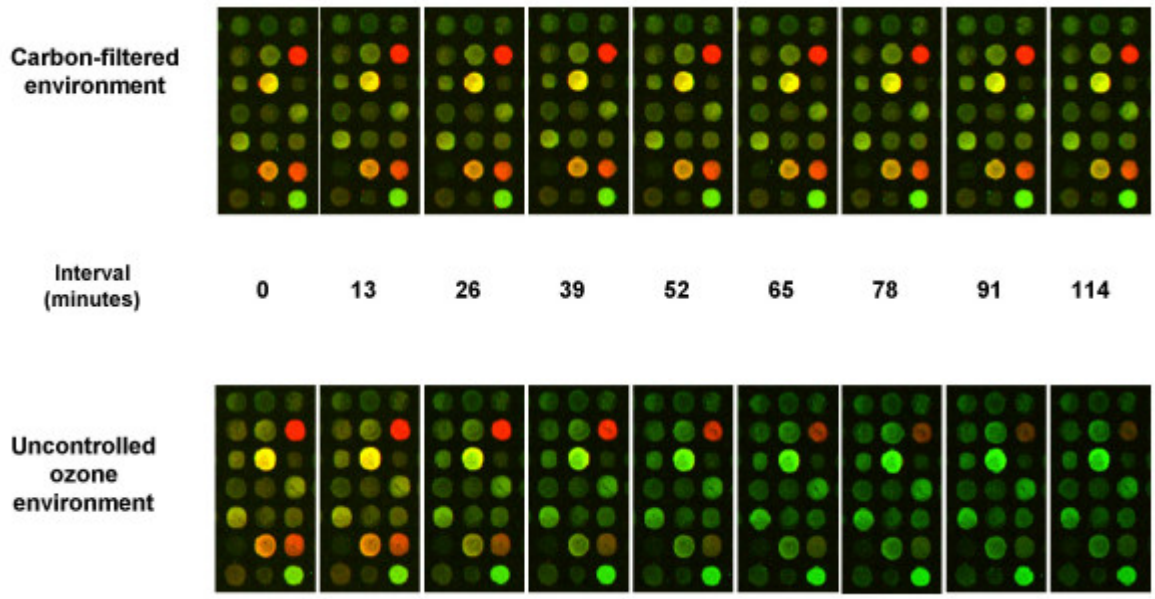


Figure 17 | Degradation of Cy5 signals for arrays kept in an ozone controlled and uncontrolled lab [37]. Two 20 K microarrays were scanned at several intervals during 114 minutes. With time, the array kept in the carbon-filtered lab remained unaltered, while the array exposed to ozone shows dominantly green spots (cy3).

6 DNA extracted from 2 different sources: saliva vs. blood

6.1 Quality and quantity of sample DNA for array CGH

The quality and quantity of DNA samples has a major impact on microarray data quality depending on the selected array platform. While large insert BAC clone arrays efficiently pick up low-quality DNA, oligonucleotide and small PCR-product arrays require easily accessible DNA sequences. For that reason, the ROMA approach has considered sample DNA preparation though array CGH results were as much improved in specificity as variable in reproducibility. Typically an amount of 300 ng genomic DNA is released on long-oligonucleotide arrays.

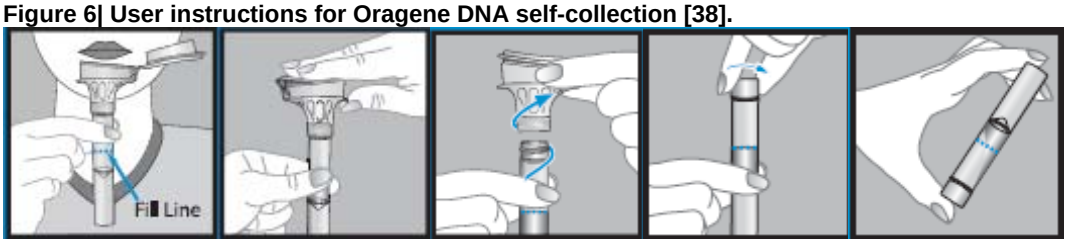
6.2 The Oragene™ DNA Self-Collection Kit

Recently the Oragene DNA Self-Collection kit has been launched on the market sector by DNA Genotek Inc. The commercial kit provides a self-employed DNA collection method. According to the manufacturer, saliva can be used as an alternative source of genomic DNA that is equivalent in quality and quantity to DNA as usually extracted from blood. The procedure allows the long-term sample storage at room temperature, involves simple DNA purification, produces high yields and is noninvasive method for DNA collection.

6.2.1 Step 1 – DNA collection

6.2.1.1 DNA collection from spitters

Figure 6| User instructions for Oragene DNA self-collection [38].



1. Spit saliva until the blue 'fill line' shown on the picture is reached (for 2ml saliva)

2. Close the lid by firmly pushing until you hear a loud click. The fluid in the lid will then release.

3. Hold the tube upright. Unscrew the funnel from the tube.

4. Use the small cap attached to close the tube.

5. Shake the capped tube 5 seconds to mix the preserving fluid well with the sample.

6.2.1.2 DNA collection from non-spitters

Figure 7|Material for collecting Oragene saliva DNA from infants or young children [39]. Up to 5 sponges per kit are necessary to collect saliva in the cheek pouches of the non-spitter individual.

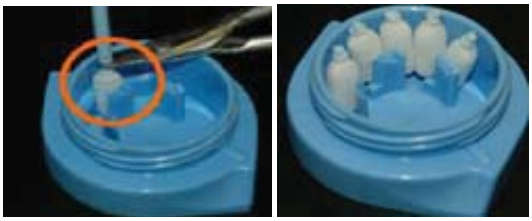


Figure 80|Method for collecting Oragene DNA from infants or young children using sponges [39]. Sponges are subsequently cut with scissors into the Oragene Kit vial to preserve the DNA at room temperature.

6.2.2 Step 2 – Storage of collected saliva

The Oragene DNA self-collection kit allows the storage of saliva DNA at room temperature due to the stabilizing fluid mixed with the samples when DNA collection was performed.

6.2.3 Step 3 – DNA purification

Both the Oragene DNA purification protocol and QIAamp purification protocol (Qiagen) are successful procedures to extract pure DNA from Oragene samples [40]. The former is based on a simple alcohol precipitation, while the latter uses a silica-gel membrane column.

6.3 Comparison between DNA from human blood and saliva

Table 3 | Comparison between blood and oral DNA collection by DNA Genotek [41]. Of all oral collection methods shown in the table (mouthwash, buccal swabs and Oragene DNA), Oragene DNA collection provides the highest DNA yield and lowest bacterial content. Compared to DNA collection from venous blood, Oragene DNA extraction gives rather equivalent results. Blood spotting yields much less DNA. The main advantage of Oragene DNA collection over traditional DNA collection methods is the non invasiveness; the Oragene saliva DNA self-collection method does not require painful blood draw or uncomfortable and unreliable cheek scrapes.

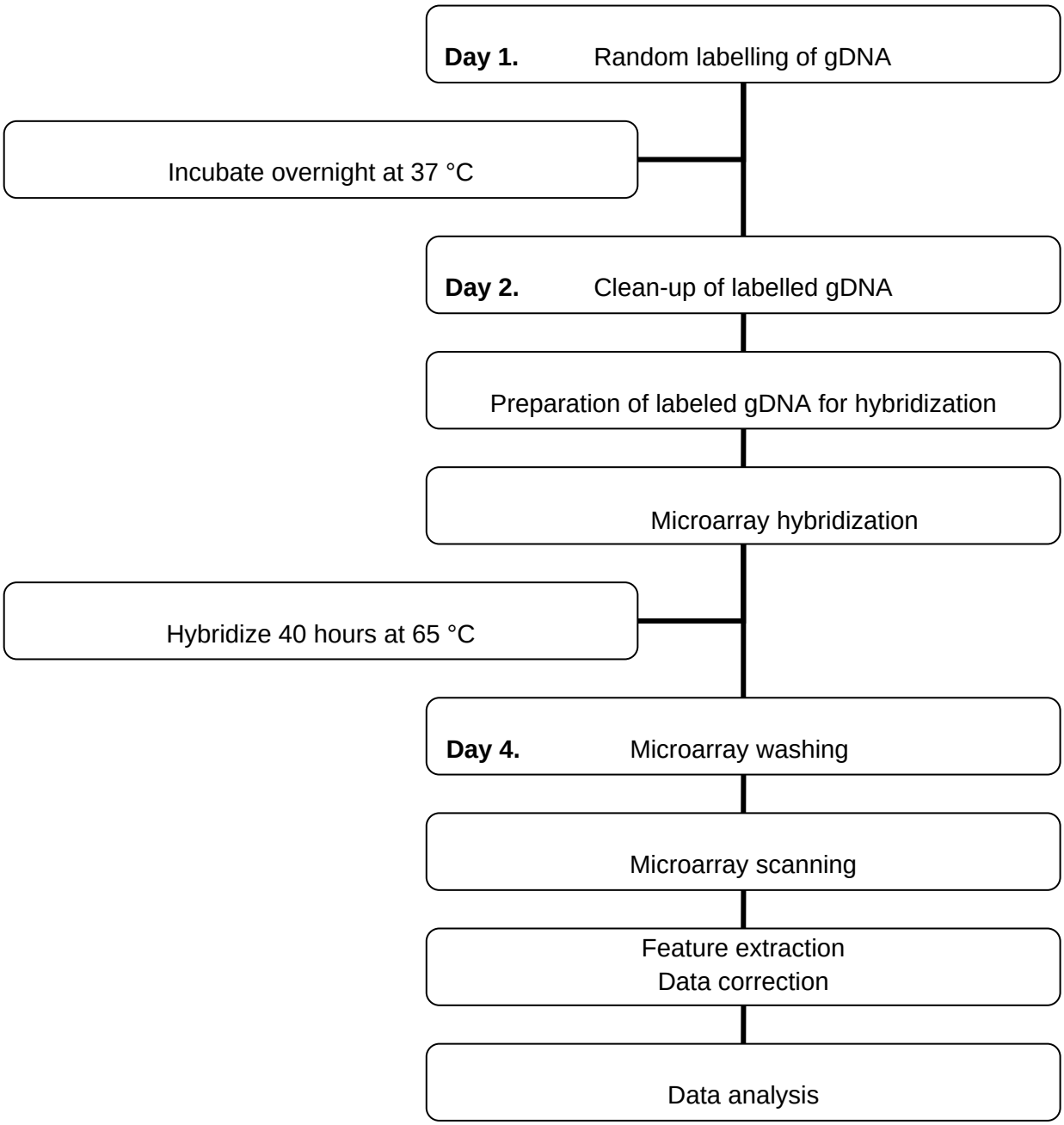
	Blood Collection		Oral Collection		
Attributes	Blood Spots	Venous Blood	Mouthwash	Buccal Swabs	Oragene•DNA (OG-500)
Specimen stability at room temperatue	Years	Weeks	Weeks	Days	Years
Median DNA yield	2 µg/card	30 µg/1 ml	35 µg/10 ml	2 µg/swab	110 µg/2 ml
Molecular weight	< 23 kb	>23 kb	> 23 kb	< 23 kb	>23 kb
Completely non-invasive	×	×	×	×	✓
Standardized format for high throughput processing	×	✓	×	×	✓
No special shipping or handling required	✓	×	✓	✓	✓
Low infectious risk	✓	×	✓	✓	✓
Low bacterial content (<12%)	✓	✓	×	×	✓

7 Agilent 244K Array-Based CGH protocol

The following protocol is obtained from The Wellcome Trust Sanger Institute in Cambridge and adapted from Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis.

7.1 Process flow diagram

Below is a flow diagram showing the workflow in one week to achieve the highest signal strength during scanning on Agilent 244K Microarrays.



7.2 Materials

7.2.1 Equipment

Pipetman micropipettes (P-10, P-20, P-200, P-1000)	Gilson
Pre-sterilized and aerosol-resistant Pipette tips	Rainin
Powder-free Gloves	Shieldskin™ Orange Nitrile
UV-Vis spectrophotometer	Nanodrop™ ND-8000
Medical wipes	Kimberly Clark
Vortex mixer	Fisherbrand®
Microcentrifuge	Eppendorf Centrifuge 5415D
Heat block for 1.5 ml eppendorf tubes	Techne
Ozone detector	Eco Sensors
Ozone scrubbers	SciGene
Air Conditioning	Adcock
Dehumidifier	Amcor
37°C incubator (Hybaid Shake `n` Stack)	Thermo Scientific
Microcon YM-30 filter	Millipore
1.5-ml microfuge tubes	Eppendorf
Desiccator cabinet	
Microarray Hybridization Chambers	Agilent
Microarray Hybridization Gasket Slides	Agilent
244K Custom Human Whole Genome Microarray	Agilent
Hybridization oven set to 65°C	
Hybridization oven rotator rack	
250 ml capacity dish for washing slides	Raymond A Lamb
Slide-racks	Raymond A Lamb
Magnetic stir bars	Camlab
Magnetic stir plate	Camlab
Computer workstation	HP XW4600
DNA Microarray Scanner	Agilent
Scan Control software v. 7.0	Agilent
Feature Extraction software v. 10.5	Agilent
DNA Analytics v. 4.0	Agilent
SignalMap software v.1.9	Nimblegen

7.2.2 Reagents

BioPrime® Labeling Kit, with:	Invitrogen cat # 18094011
2.5x Random Primers Solution	p/n Y01393
Water	p/n 50837
Klenow Fragment 40 U/µl	p/n Y01396
Stop Buffer	p/n 50690
10x dNTP-mix (1mM dCTP, 2mM dATP, 2mM dGTP 2mM dTTP in TE buffer)	cat # AB-0315/A
1mM Cy3-dCTP	Amersham p/n PA53021
1mM Cy5-dCTP	Amersham p/n PA55021
1x TE buffer, pH 8.0	Promega cat # V6231
Oligo aCGH Hybridization Kit, with:	Agilent
10X Blocking agent	p/n 5188-6416
2X Hybridization Buffer	p/n 5188-6420
Cot-1 Human DNA (1.0 µg/µl)	Roche cat # 11581074001 Lot No. 70152122
Oligo aCGH wash buffer 1	Agilent p/n 5188-5221
Oligo aCGH wash buffer 2	Agilent p/n 5188-5222

7.3 Method

7.3.1 Random labelling of genomic DNA

In order to label the genomic DNA of test and reference sample a mixture of random hexamer primers is used together with Cy-labelled cytosine nucleotides and Klenow polymerase.

1. Measure the DNA concentration using a Nanodrop spectrophotometer.

   Make sure the extracted DNA samples are pure. Only a ratio of absorption 260/280 nm higher than 1.8 and a 260/230 nm ratio higher or equal to 2 are acceptable as ideal.

2. Calculate the required volume, containing 300 ng gDNA, per sample.

3. Mix the following reagents in an autoclaved tube:

2.5x Random Primers Solution	60 µl
Water	70.5 µl – x µl
Required volume gDNA	x µl
Total	130.5 µl

4. Vortex and spin down briefly.
5. Denature at 100°C for 10 min and set immediately on ice.
6. While on ice, add the following reagents in the order indicated:

Perform all subsequent work in an ozone-controlled environment (≤ 5 ppb). Avoid exposure to light to retain the fluorescent signal from the Cy-dyes as much as possible.

10x dNTP-mix	15 µl
Cy3 –dCTP (to reference sample)	1.5 µl
Cy5 –dCTP (to test sample)	1.5 µl
Klenow Fragment	3 µl

{Make up a mastermix for more than 4 samples.}

7. Vortex and spin down briefly.
8. Incubate the reaction at 37°C overnight over activated charcoal.
9. Stop every reaction with 15 µl Stop Buffer. Mix contents by pipetting up and down.

7.3.2 Clean-up of labelled genomic DNA

The labelled DNA needs to be purified of dirty components causing background noise.

10. Add 330 µl of 1x TE buffer to each reaction tube.
11. Transfer each labelled gDNA to a Microcon YM-30 filter in a 1.5-ml microfuge tube.
12. Spin 10 minutes at 8.0 rcf at room temperature.
13. Discard the flow-through of the column.

Pay attention to the colour of the flow-through. It can either be too blue or too red if the filter is damaged, for samples labelled in Cy5 and Cy3 respectively.

14. Add 480 µl of 1x TE buffer to each filter.
15. Spin 12 minutes at 8.0 rcf at room temperature.
16. Discard the flow-through of the column.
17. Invert filter into a fresh 1.5-ml tube.
18. Spin 1 minute at 8.0 rcf at room temperature to collect the purified sample.
19. Bring the sample to a total volume of 80.5 µl with 1x TE buffer.
20. Determine the yield and specific dye activity by using the Nanodrop, program `Microarray`.

Specific CyDye Activity = [pmol/µl Cy dye] / [µg/µl DNA]

Table 4 | Expected DNA yield after labelling and clean-up.

Input gDNA (µg)	Yield (µg)	Specific Cy-3 Activity (pmol/µg)	Specific Cy-5 Activity (pmol/µg)
0.5	5 to 7	25 to 40	20 to 35

21. Combine the Cy3-labelled sample (control) and Cy5-labelled sample (test) in a new 1.5 ml heat-resistant nuclease-free tube.
22. Flick to mix the contents and spin down briefly.
23. Labelled DNA can be stored at -20°C in the dark.

7.3.3 Preparation of labelled genomic DNA for hybridization

Pre-annealing the labelled DNA probes with Cot-1 DNA is necessary to block repeated DNA sequences, thereby effectively suppressing non-specific hybridization to targets.

24. Add the following components to each tube of Cy 5- and Cy 3-labelled gDNA mixture in the order shown:

Cot-1 DNA	50 µl
10 x Blocking Agent	52 µl
2 x Hybridization Buffer	260 µl

25. Mix the sample by pipetting up and down. Spin down briefly.
26. Transfer sample tubes to 95°C for 3 minutes of denaturation and cover the tubes with a heat-resistant box to minimize light exposure.
27. Transfer sample tubes to 37°C for 30 minutes of pre-hybridization and cover the tubes with a heat-resistant box to minimize light exposure.

7.3.4 Microarray hybridization

Denatured probe DNA will associate with complementary sequences onto the array.

28. Load a clean gasket slide, with the gasket label facing up, into the chamber base.
29. Dispense 490 µl of hybridization sample mixture.
30. Place the active side of a microarray slide with the Agilent labelled barcode down onto the gasket slide.
31. Assemble the chamber and hand-tighten the clamp.
32. Vertically rotate the assembled chamber to check for immobile air bubbles. If necessary, tap the chamber on a hard surface to move stationary bubbles.
33. Place the chamber in the rotator rack in a hybridization oven.
    Make sure to balance the loaded chambers with empty chambers. Rotate at 20 rpm.
34. Hybridize at 65°C for 40 hours.

7.3.5 Microarray washing (without Stabilization and Drying Solution)

Remove any unbound material of the hybridization results. Stabilization and Drying Solution are useless if you wash microarray slides in an ozone-controlled environment.

35. Establish the following wash conditions:

Table 5 | Microarray wash conditions.

	Dish	*Wash buffer*	*Temperature*	*Time*
Disassembly	*No. 1*	*Oligo aCGH Wash buffer 1*	*Room temperature*	
1st wash	*No. 2*	*Oligo aCGH Wash buffer 1*	*Room temperature*	*5 min*
2nd wash	*No. 3*	*Oligo aCGH Wash buffer 2*	*37°C*	*1 min*

{Wash a maximum of 5 slides at once. Use fresh wash buffers for each group.}

36. Disassemble the hybridization chamber.
37. Handle the slides from their ends as you quickly transfer the sandwich to dish No.1
38. Submerge the gasket-array sandwich completely in dish No. 1 with Oligo aCGH Wash buffer 1 and open the sandwich from the barcode end only.
39. Let the gasket slide drop to the bottom of dish No. 1.
40. Transfer the microarray slide into dish No.2 with Wash buffer 1.
41. Stir 5 minutes.
42. Transfer the microarray slide into dish No.3 with prewarmed Wash buffer 2 at 37°C
43. Stir 1 minute.
44. Take 5 to 10 seconds to slowly remove the slide out of wash buffer 2.
45. Assemble the slides into the slide holders, with the Agilent labelled side facing up.

7.3.6 Microarray scanning using Agilent Scanner G2565BA

To visualize the microarray, the two spectrally distinct fluorescent dyes Cy3 (532 nm) and Cy5 (635 nm) are excited with a laser beam of required wavelengths. Scan slides as soon as possible to minimize degradation of signal intensities.

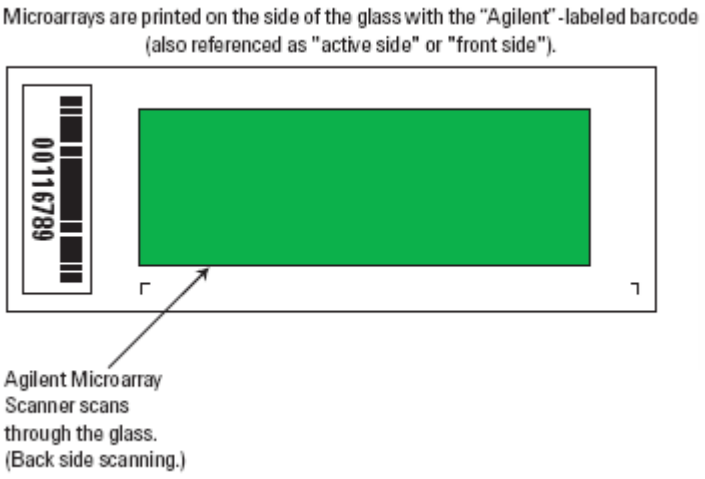


Figure 9 | Agilent microarray slide.

46. Place assembled slide holders into the scanner carousel.
47. Agilent 244K arrays require 5 µm scan resolution.

    Use AGILENT SCAN CONTROL SOFTWARE V. 7.0, with following scan settings:

Scan Area	61 x 21.6 mm
Scan resolution	5 µm
Dye channel	Red & Green
PMT power	100%
TIFF dynamic range	20 bit

48. Select Slot *m*, where the first slide is located, and Slot *n*, where the last slide falls within.
49. Save the images.
50. Once the lasers have warmed up, the Scanner status in the main window will say `Scanner Ready`.
51. Click `Scan slot m-n`.

7.3.7 Feature extraction and data correction

Software has been applied to extract and correct data from raw microarray image files.

52. Import scanned files to FEATURE EXTRACTION SOFTWARE V.10.5. Feature Extraction (FE) will compare assay data to a design file which contains coordinates of all probes synthesized on the microarray slide.
53. Review the output QC report (PDF files), where the derivative log ratio spread (DLRSpread) is considered as the main quality index.
54. Correct Feature Extraction text files with an in-house script. This will remove any probes showing dye-bias from the arrays, and correct any waves introduced in the hybridisation profile due to GC-content. As a result, the number of false positive data points is reduced.

7.3.8 Data analysis

Microarray data is ready to be analyzed for any chromosome, any sequence and any abberation using DNA Analytics and SignalMap software.

55. Import the corrected text files into DNA ANALYTICS V.4.0, create and activate an experiment to analyze and visualize the data.

Use the Aberration Detection Method 2 (ADM-2) algorithm for aberration detection. Specify aberration and feature filter conditions that is a 2-probe filter and the default feature level filter, respectively.
Export an Interval-based aberration summary report for subsequent analysis.

56. View the corrected data in SIGNALMAP to assess how reliable the probes involved in detected aberrations are. Probes are colour coded according to quality from black (high quality), blue, green, yellow to red (poor quality) for viewing in SignalMap. Especially in repetitive regions of the genome that are associated with poorer quality probes, CNV calls in these regions are generally not very informative in terms of copy number.

8 Comparison of array CGH results from saliva and blood

8.1 *Aim of the Oragene project*

As DNA Genotek recently came out with a promising Oragene DNA saliva collection kit, we compare Oragene saliva DNA to normal blood DNA from two patients and evaluate the hybridization quality, reproducibility, dose response and CNV detection for analysis with array CGH. Blood DNA was provided already extracted by an external collaborator and saliva DNA was extracted in-house using the Oragene kit and QIAamp purification. We performed comparative genomic analysis of total DNA against a male HapMap sample NA10851 using Agilent custom 244k whole genome CGH arrays. A well-characterized European-American male NA10851 is selected from the 4 populations of the International HapMap study [42] and used as reference sample for all experiments.

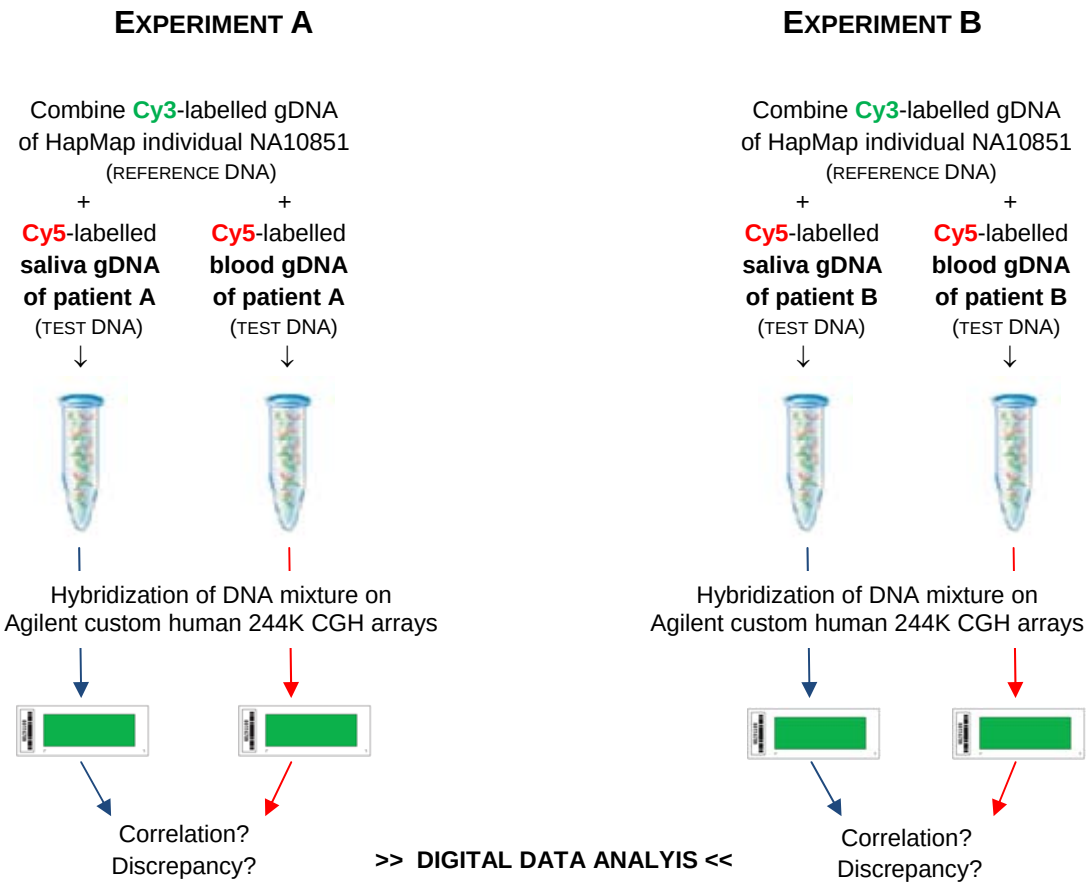


Figure 10 |Experimental design. Our project starts with the comparison of saliva DNA and blood DNA from patient A for analysis with high-resolution array CGH. We tried to confirm the first experiment in the same way using DNA samples from patient B. Applying digital analysis we qualify the array CGH results of the paired blood and saliva samples and determine if saliva is a viable alternative source of DNA for such experiments.

8.2 Microarray data quality

8.2.1 Quality Control (QC) report

8.2.1.1 The use of Agilent Feature Extraction software v.10.5

Feature Extraction software version 10.5 was used to extract and normalize data from microarray image files (TIFF file) of scanned Agilent CGH microarrays.

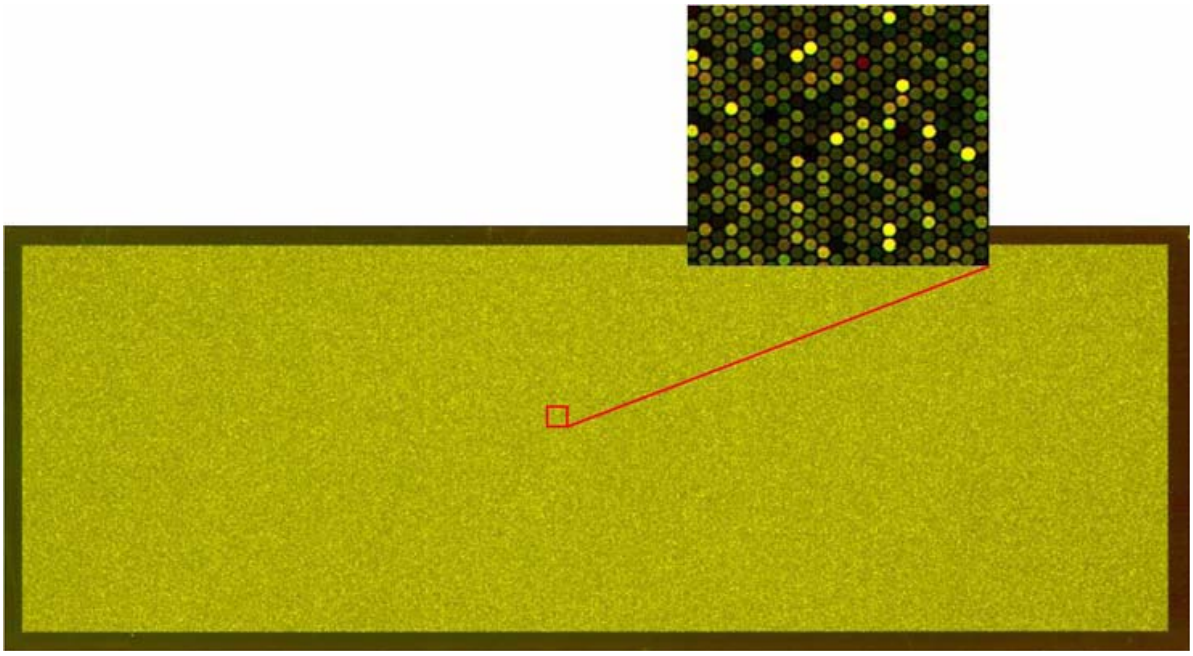


Figure 11 | Microarray image opened in Feature Extraction software v.10.5 (full and zoomed view). Green spots – copy number gain in test sample, red spot – copy number loss in test sample, yellow spot – no change in copy number relative to the reference sample.

Following default settings, the Feature Extraction Program generated several output files for each microarray. One of these is a QC report which describes general statistics of each microarray data set (see chapter 5.1.). Feature statistics can help you to evaluate microarray performance. QC metrics are a valuable guideline for assessing the relative quality among multiple microarrays in an experiment or to indicate potential processing errors.

8.2.1.2 Results

QC reports for each sample are attached as appendices 1, 2,3 & 4.

8.2.1.3 Conclusion

Of all 4 microarrays Quality Control metrics are within the normal ranges as presented in Table 2. Despite the DLRSpread of 0.17 from the array ran with the Oragene saliva DNA sample from patient A being slightly higher than the DLRSpread of the other 3 arrays, all microarray experiments were performed successfully and are mutually comparable.

8.2.2 Quality Control (QC) Metrics

8.2.2.1 The use of Agilent DNA Analytics software v.4.0

In preparation for analysis by DNA Analytics, Feature Extraction text files containing all normalized data sets were processed with an in-house-script (see chapter 7.3.7.). As we proceed with corrected TEXT files, slightly corrected QC metrics are displayed after uploading in DNA Analytics software.

In DNA Analytics software version 4.0, the QC metrics can be shown in a histogram plot which gives a quick overview how each quality parameter varies in each data set.

8.2.2.2 Results

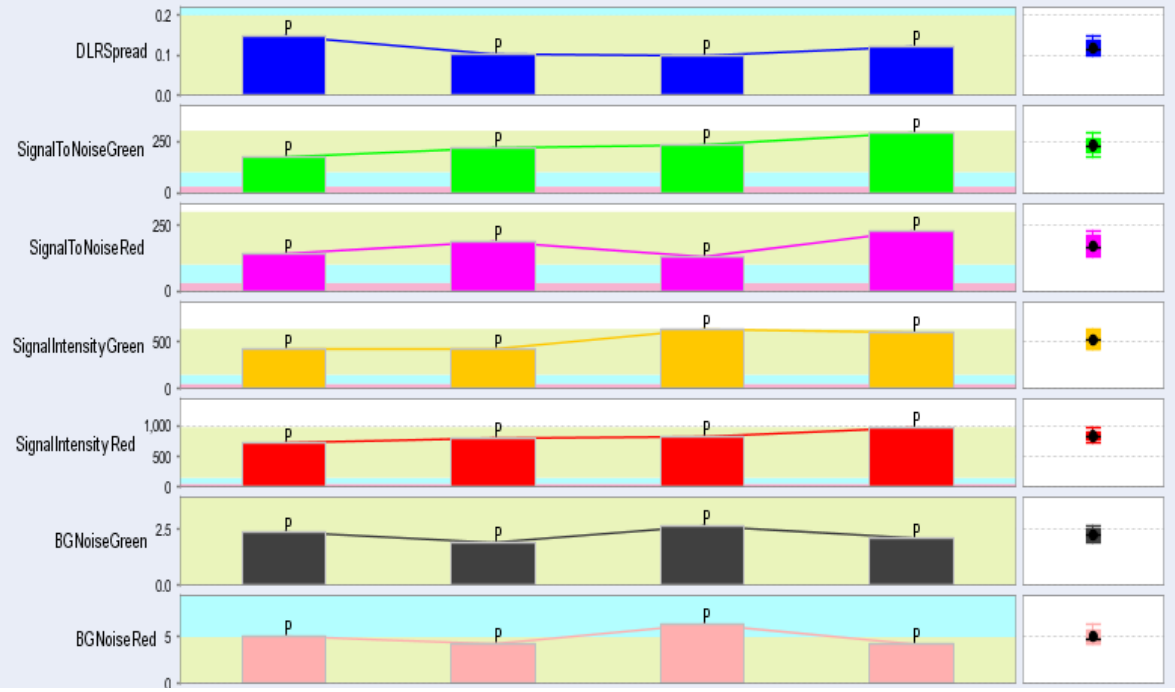


Figure 24 | QC metrics plot. QC metrics of the saliva and blood array data from patient A, and the saliva and blood array data from patient B are plotted from left to right. Note the higher DLRSpread of saliva sample A.

Table 6 | QC Metrics table. QC metric values are shown for respectively the saliva and blood sample from patient A, followed by the values for the saliva and blood sample from patient B. Yellow boxes indicate excellent QC metrics, blue boxes assign good quality values and red boxes would point out a poor quality hybridization.

DLRSpread	SignalToNoiseGreen	SignalToNoiseRed	SignalIntensityGreen	SignalIntensityRed	BGNoiseGreen	BGNoiseRed
0.149823	176.366917	139.757501	412.751495	720.469482	2.340300	5.155140
0.102844	219.709318	188.239049	418.961502	808.402008	1.906890	4.294550
0.100153	234.671028	128.221784	624.239014	828.372986	2.660060	6.460470
0.122073	287.101387	225.469379	597.001495	971.401001	2.079410	4.308350

8.2.2.3 Conclusion

Reviewing the QC metrics in DNA Analytics we observe a mild increased background noise in both saliva samples, particularly in the red channel (Cy5 channel). Although the values for background noise are still in the normal range, the Cy5 signal-to-noise ratio is better for blood samples.

8.2.3 Chromosome X dose response

8.2.3.1 Description

Another way to qualify microarray data, makes use of the average $\log_2$ratio for the whole chromosome X. Since we have applied a male reference DNA sample to co-hybridize, the extra chromosome X in the female test DNA sample is present in a 2:1 ratio, which should give a theoretical average $\log_2$ratio of 1 across chromosome X. The higher the $\log_2$ratio is, as determined by the aberration summary report generated in DNA Analytics, indicates a better chromosome X dose response.

8.2.3.2 Results

Table 7| Chromosome X dose response values.

DNA source	*Patient sample*	CHRX DOSE RESPONSE
Oragene saliva	Patient A	**0,94**
Blood	Patient A	**0.96**
Oragene saliva	Patient B	**0.93**
Blood	Patient B	**0.93**

8.2.3.3 Conclusion

As the chromosome X dose response usually ranges from 0.75 to 0.95, saliva and blood provide high quality DNA sources for array CGH based on the dose response values.

8.3 Correlation analysis of paired saliva and blood samples

8.3.1 Agreements in CNV detection

8.3.1.1 The use of Agilent DNA Analytics software v.4.0

Copy number analysis was performed in DNA Analytics software v. 4.0. The software includes several algorithms to choose from, which determine the detection sensitivity. The different algorithms are tailored to factors such as the quality and type of sample being investigated (e.g. ability to detect large aberrations in tumour samples, and small aberrations in CNV studies). Within those algorithms, different thresholds can be applied to filter the quality/quantity of probes being called based on the extent they deviate from the baseline, and the number of probes included in the aberrant region. For this project, CNV calls were made using the ADM-2 algorithm (Aberration Detection 2) set at a calling threshold of 5 applied with a 2-probe filter.

CNVs detected in both samples were visualized in genome, chromosome and gene view. CNV calls made on each chromosome are displayed next to their ideograms. DNA gains are illustrated by a coloured bar orientated to the right and DNA losses by a coloured bar to the left. The height of the coloured bars correspond to the average $\log_2$ratio of the probes within the CNV call. Longer bars correspond to probes with a higher average $\log_2$ratio and so deviate more significantly from the baseline. We indicated CNV calls made on saliva samples with a blue colour, while CNV calls made on blood samples were coloured in red. Purple bars resulted from the overlay of blue and red CNV calls, i.e. CNVs corresponding both in saliva as well as in blood. The aberration report summarizes all CNV's properties.

8.3.1.2 Results

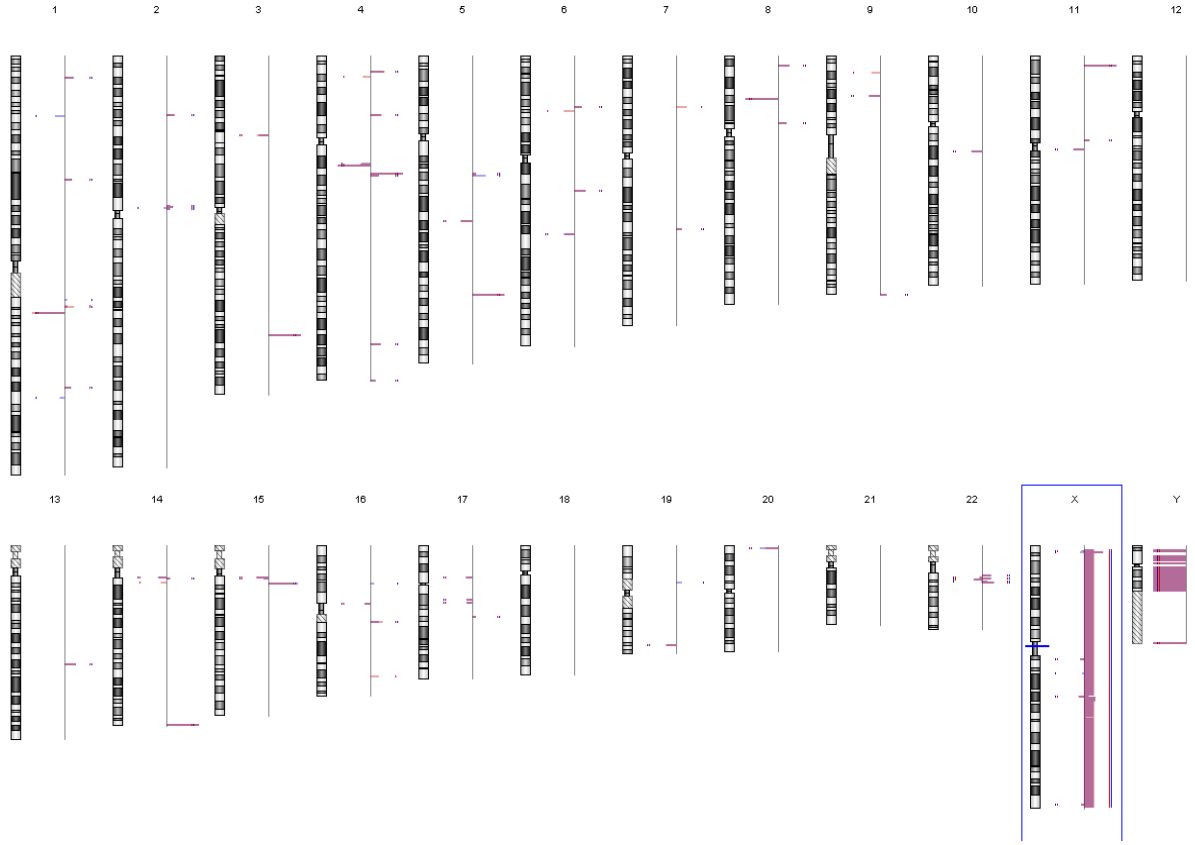


Figure 12 | Screenshot of the genome view in DNA Analytics (patient A).

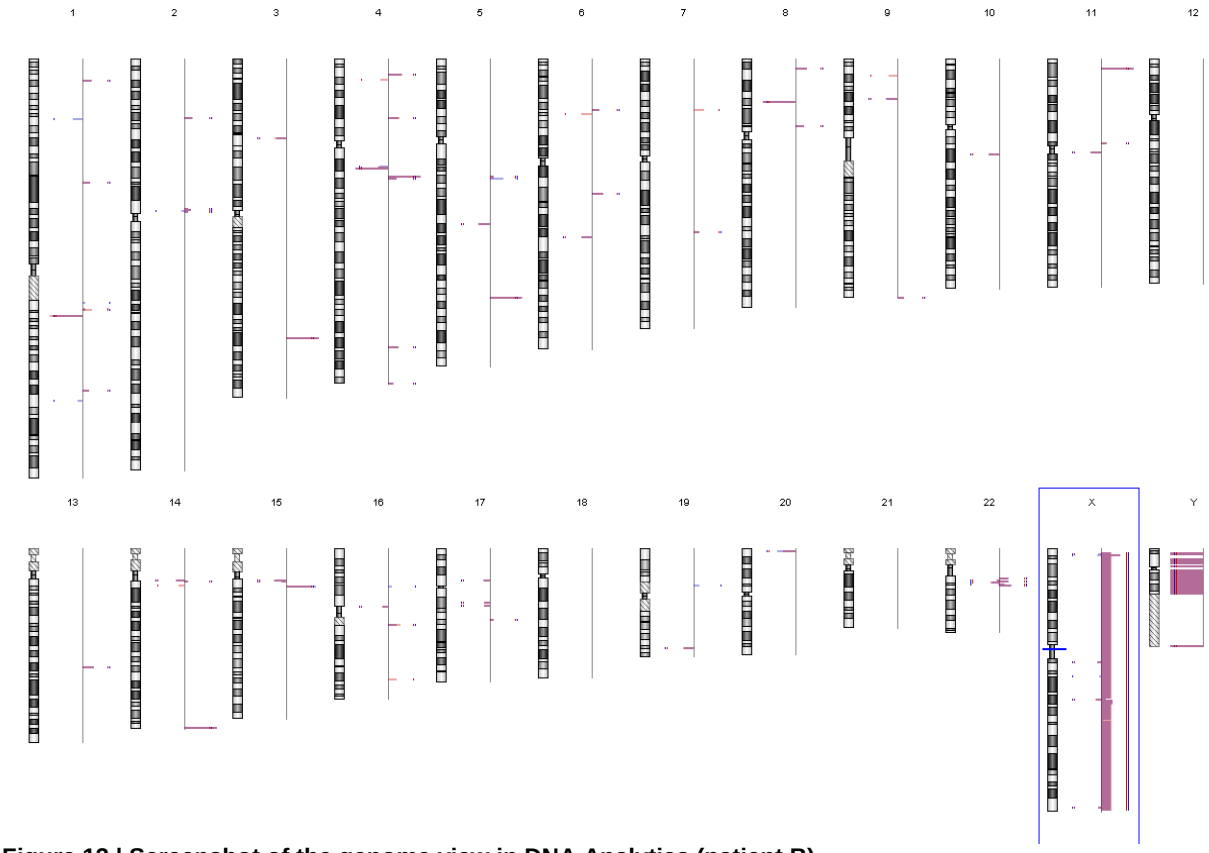


Figure 13 | Screenshot of the genome view in DNA Analytics (patient B).

8.3.1.3 Conclusion

For patient A, 54 autosomal CNV calls matched on the saliva and blood sample. The DNA sources made 6 CNV calls that weren't made in the other. For patient B, 41 autosomal calls matched, while 2 extra calls were made on saliva gDNA and 17 extra calls on blood gDNA.

8.3.2 Correlation values

8.3.2.1 Description

To determine how well saliva array data statistically matches the paired blood array data, we quantified their linear relationship or correlation. The correlation coefficient *R* measures the linearity between two variable data sets in a value between -1 to 1. A perfect correlation occurs in ± 1, a strong correlation is greater than 0.5, while any *R* values less are weak [43].

8.3.2.2 Results

Table 8 | Correlation values between genomic DNA extracted from blood and saliva

Chromosome number	*Blood-saliva R (Patient A)*	*Blood-saliva R (Patient B)*
Chr1	0.64	0.62
Chr2	0.65	0.67
Chr3	0.73	0.79
Chr4	0.74	0.77
Chr5	0.59	0.69
Chr6	0.61	0.64
Chr7	0.71	0.74
Chr8	0.66	0.59
Chr9	0.70	0.78
Chr10	0.54	0.54
Chr11	0.62	0.67
Chr12	0.52	0.59
Chr13	0.60	0.55
Chr14	0.77	0.83

Chr15	0.73	0.72
Chr16	0.57	0.59
Chr17	0.66	0.62
Chr18	0.50	0.55
Chr19	0.61	0.54
Chr20	0.57	0.57
Chr21	0.67	0.72
Chr22	0.87	0.89
Chr23	0.84	0.87
Chr24	0.96	0.96
Total genome	0.93	0.95

8.3.2.3 Conclusion

For all 24 human chromosomes present in saliva DNA and blood DNA, few correlation values were less than 0.6 and not any value describes a weak correlation. Computing the correlation coefficient of the entire genome in paired saliva and blood samples indicates a highly positive correlation percentage of 93% in patient A and of 95% in patient B.

8.3.3 The log_2 ratio correlation in whole-genome array CGH profiles

8.3.3.1 Array CGH profile

The log_2 ratios of (Cy5/Cy3) normalized probe signal intensities are plotted against their chromosomal positions in the form of array CGH profiles. We plotted the log_2 ratio profiles for the saliva and blood samples, as well as their correlation. Figure 27 and 28 show three genome profile plots for each patient's 24 chromosomes types. The upper profile describes the log_2ratios for blood (red), the bottom profile describes the log_2ratios for saliva (green) while the plot in between scatters the difference in the two data sets at consecutive probes (black). Copy number ratios which may be clearly discerned by red and green array CGH profiles, are no longer seen when similar log_2 ratios have been produced from the 2 samples. Theoretically, the plot of the difference in log_2 ratios between perfectly matching samples, should equal zero.

8.3.3.2 Results

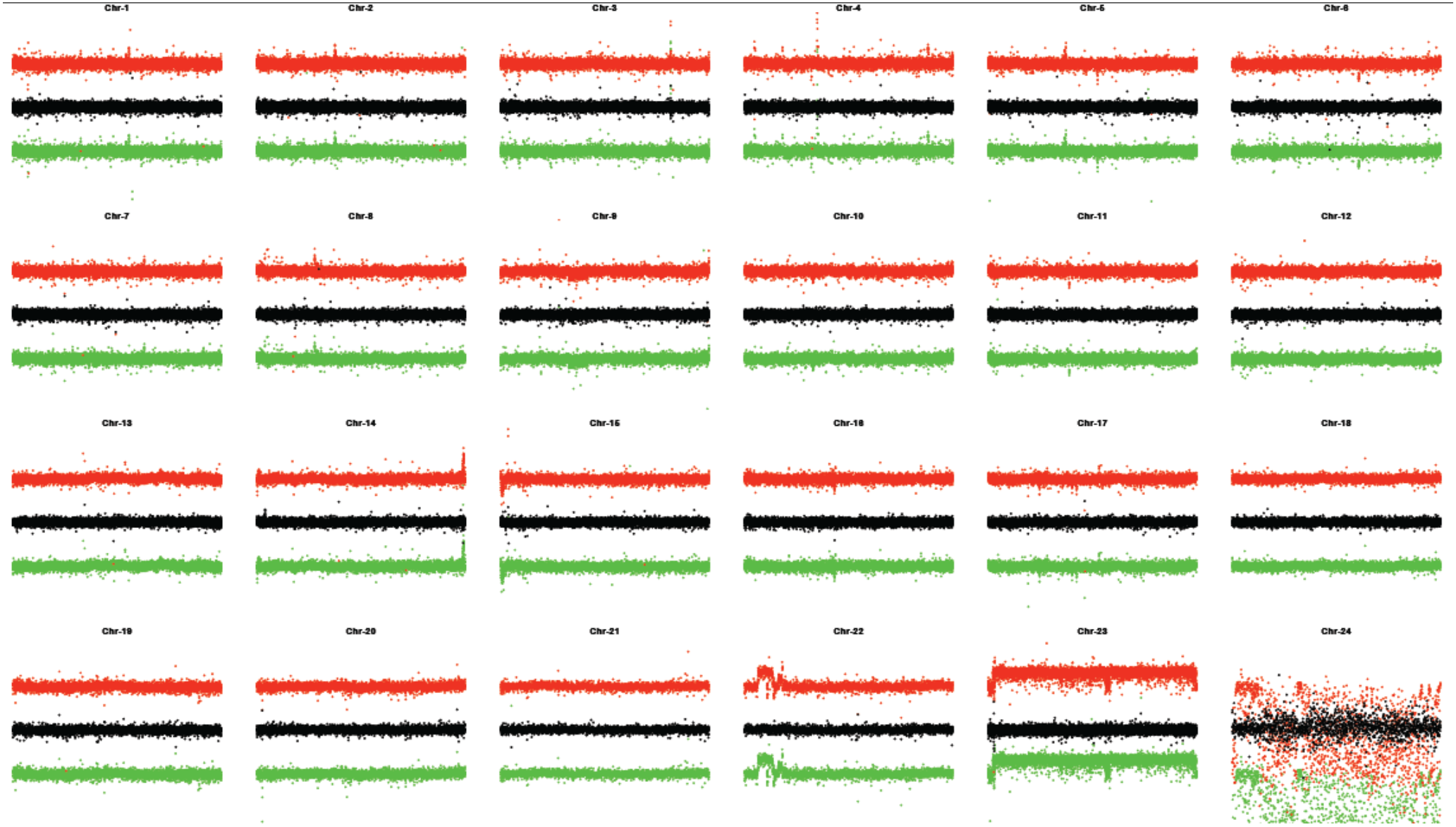


Figure 14 | Whole genome array CGH profile of patient A. In **red**: $\log_2$ ratio for blood sample, In **green**: $\log_2$ ratio for saliva sample, In **black**: difference in $\log_2$ ratio between paired samples.

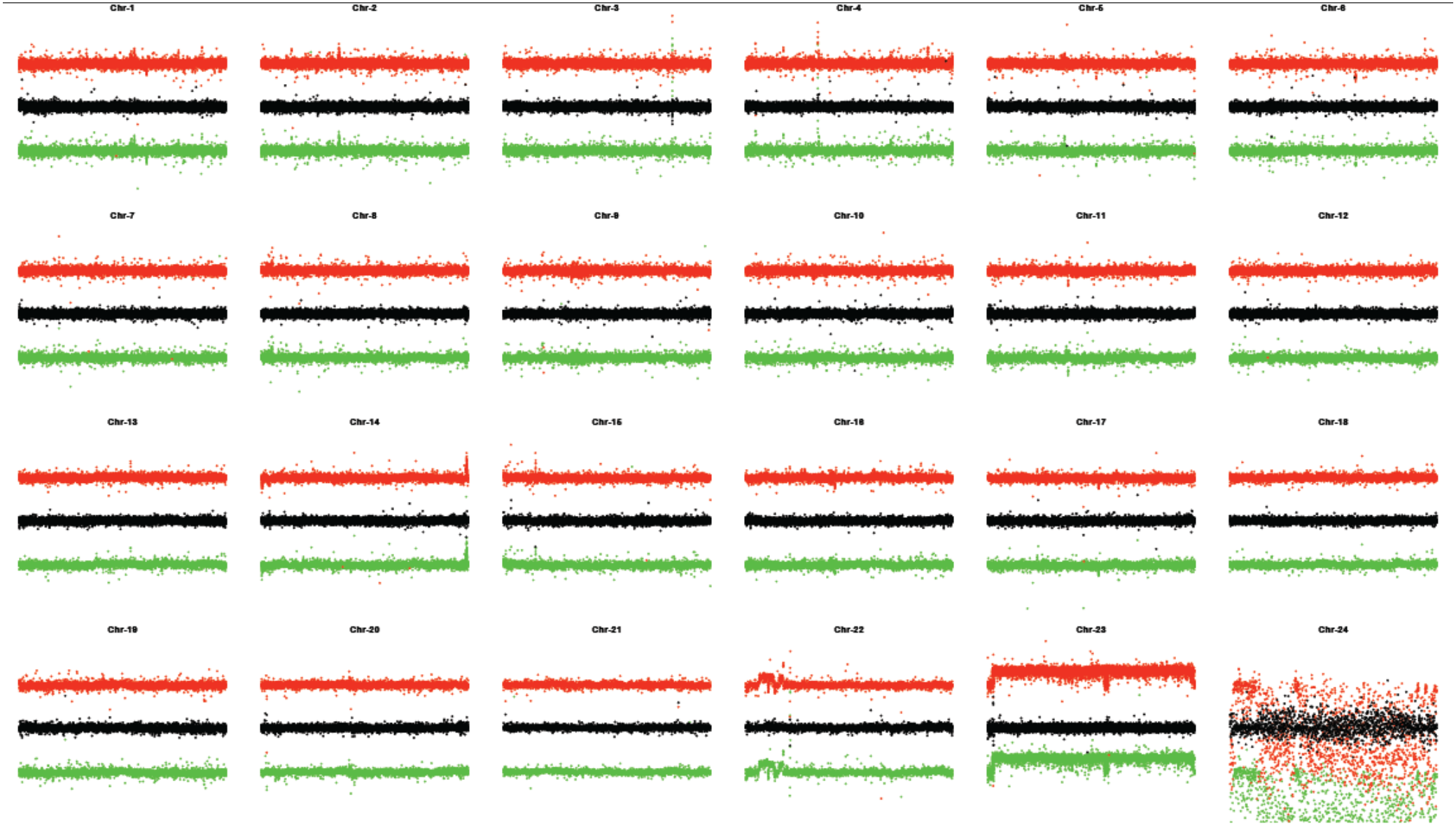


Figure 15 | Whole genome array CGH profile of patient B. In **red**: $\log_2$ ratio for blood sample, In **green**: $\log_2$ ratio for saliva sample, In **black**: difference in $\log_2$ ratio between paired samples.

8.3.3.3 Conclusion

The black scatter plots in Figure 27 and 28 imply little difference in log_2ratios between paired blood and saliva samples across the whole genome. The closer black points are plotted to make a straight line around zero, the more similar the log_2ratios are between blood and saliva samples.

8.4 Discrepancies between saliva and blood paired samples

8.4.1 Disagreements in CNV detection

As we have counted in the aberration summary report (DNA Analytics) for patient A, each DNA source made 6 CNV calls that haven't been made in the other. For patient B, 2 extra CNV calls are made on the saliva sample and 17 extra calls are made on the blood sample. . The confidence of every CNV call is assessed based on p-values, log_2ratios and the number of probes involved, given in the summary report as well. Although we observed that these CNVs are slightly overcalled due to the higher DLRS of one sample and through regions of poor quality probes, we found one significant discrepancy related to blood DNA.

8.4.2 Results

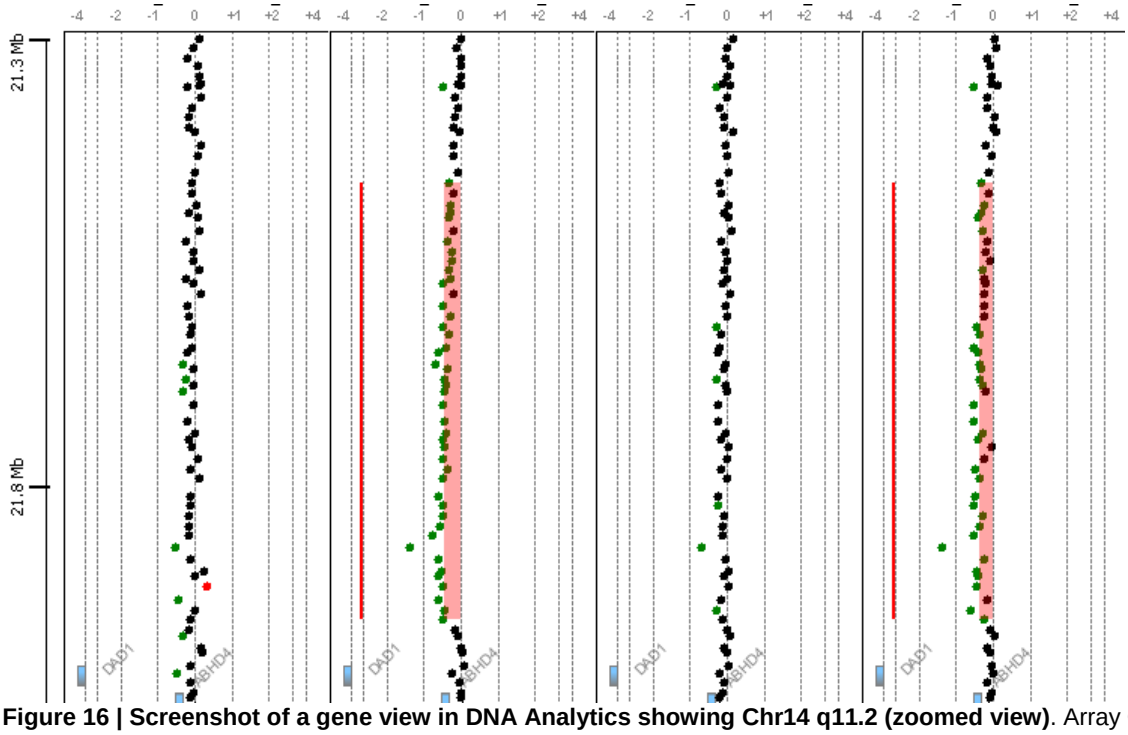


Figure 16 | Screenshot of a gene view in DNA Analytics showing Chr14 q11.2 (zoomed view). Array CGH profiles are shown in the following order from left to right: saliva and blood sample from patient A, and again for patient B. Chr14 q11.2 was detected as a deleted region in both blood samples. Green and red dots are seen if the log_2ratio of an individual probe is greater than ± 0.25.

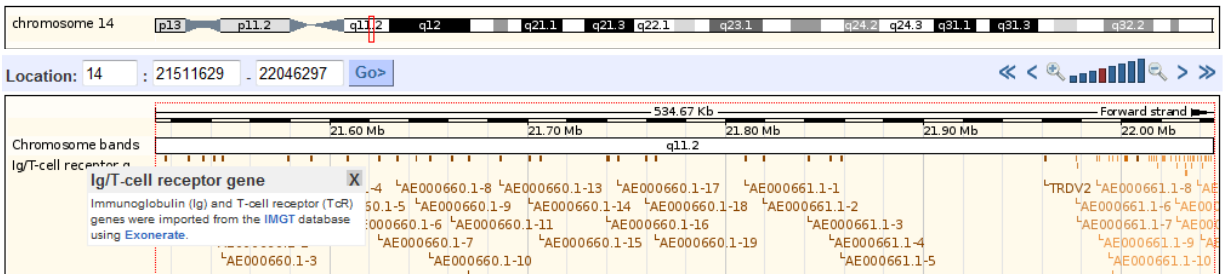


Figure 170 |Screenshot of Chr14 q11.2 in Ensembl genome browser. We entered the chromosomal location of the CNV calls only made in blood (14:21511629..22046297) into Ensembl, and found that this region seems to encode for a T-cell receptor.

8.4.3 Conclusion

Evidence shows that T-cell receptor regions are concentrated in blood whereas salivary cell types lack a T-cell receptor. Remaining discrepant CNV calls are due to poor probe signals (assessable in SignalMap) or the slightly noisier hybridization as indicated by the higher DLRSpread 0.17 on saliva from patient A.

8.5 CyDye signal intensity and distribution

Previously, in the assessment of data quality we noticed a lower signal-to-noise ratio for the Cy5 channel when using saliva samples. An increased background noise, causing low signal-to-noise ratios, usually indicates non-specific hybridization. Even though the noise threshold is just slightly exceeded we observed this only with Cy5-labelled Oragene saliva samples. As the signal/noise ratio is an important parameter in reliable CNV detection, we investigated any signal bias across the array associated with the cyanine dyes.

8.5.1 Intensity histogram

8.5.1.1 Description

Intensity histograms show the distribution of Cy dye signals across the array, by plotting the number of probes at each available intensity level. For a 20-bit microarray image 2^{20} or 1 048 576 intensity levels would be graphically displayed on the x-axis. However, data intensities only lie within a small percentage of the available range. In Figure 18 all the raw intensity values, corrected by background-subtraction [44], are set out on an adjusted x-axis scale which is useful to assess the distribution of Cy5 and Cy3 signal intensities.

8.5.1.2 Results

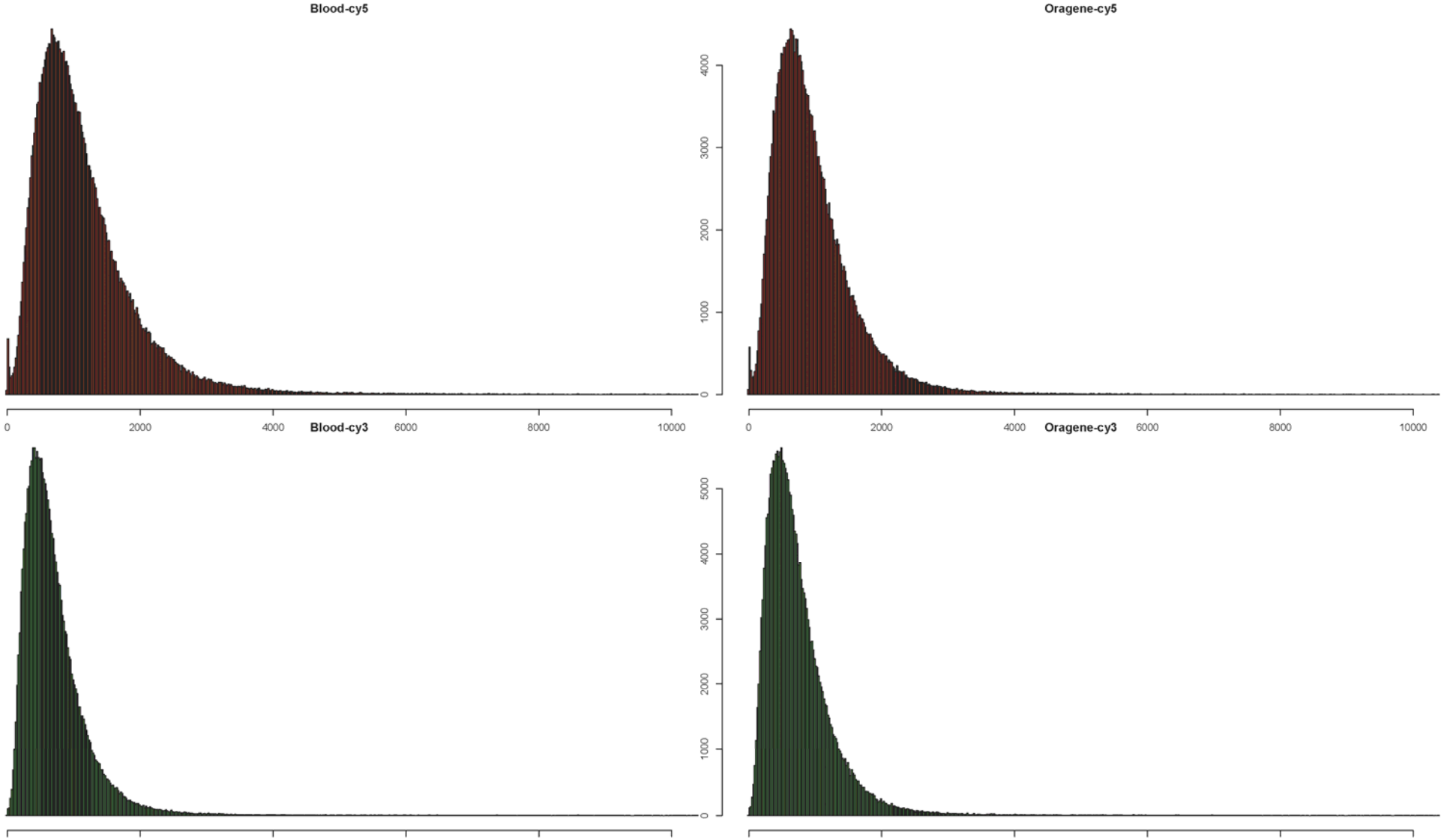


Figure 18| Intensity histograms comparing the Cy3 and Cy5 channel between paired arrays. *Top left*–red = Cy5 intensity distribution across the whole array (blood, patient B), *Top right*–red = Cy5 intensity distribution (saliva, patient B), *Bottom left*–green = Cy3 intensity distribution (blood, patient B), *Bottom right*–green = Cy3 intensity distribution (saliva, patient B)

8.5.1.3 Conclusion

No significant differences have been seen between the distributions of Cy5 and Cy3 on the two arrays. It is known that throughout the intensity range, Cy5 signals tend to be higher than Cy3 signals. This bias is due in part to the difference in labelling affinity of the two fluorophores.

8.5.2 MA-plot

8.5.2.1 Description

To determine the quality of normalisation between the Cy3 and Cy5 intensity channels, an MA-scatter plot of the log-ratios against the mean log-intensities is used. In an MA- plot, *M* values are taking the $\log_2$ratio of Cy5 and Cy3 intensities, while *A* values represent the average of the two $\log_2$intensities for each probe. Across the entire array M and A values are typically computed by [45]

$$\mathbf{M} = \log_2(\text{Cy5} / \text{Cy3}) \qquad \text{(Y-AXIS)}$$

$$\mathbf{A} = 1/2(\log_2(\text{Cy5}) + \log_2(\text{Cy3})) \qquad \text{(X-AXIS)}$$

In good quality normalization, the majority of data points should be centered around zero on the x-axis. The majority of spots have similar intensity values in both the reference (Cy3) and test (Cy5) channels with a $\log_2$ ratio around zero. Up-regulated and down-regulated values from M=0 indicate that the Cy3 and Cy5 channels are behaving differently. Because spot intensities are measured using a 20 bit image, the maximum possible A value is 20. In Figure 32 and 33, MA-plots are shown after data normalization.

8.5.2.2 Results

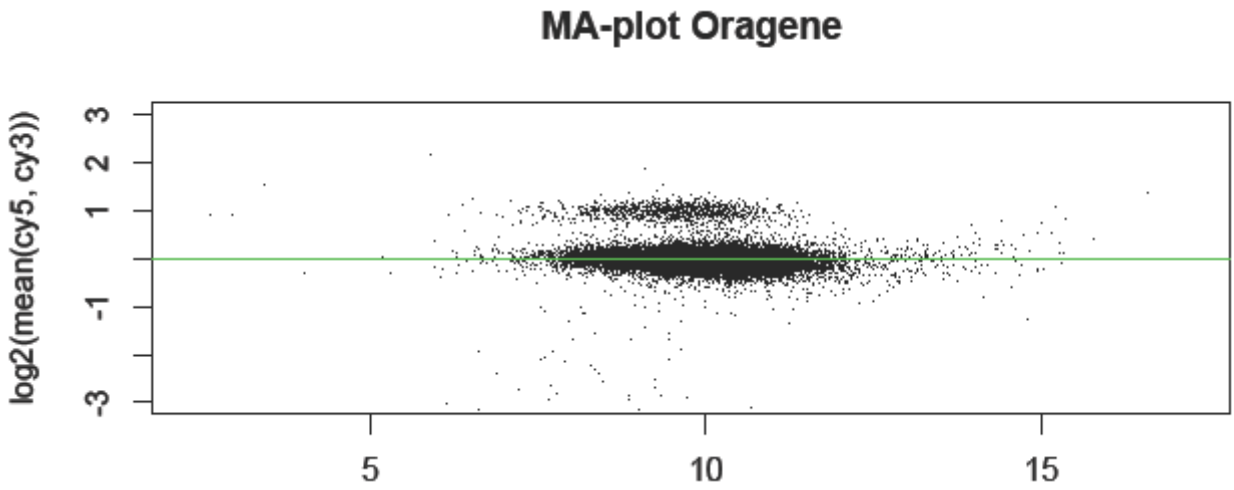


Figure 192 | MA-plot comparing the Cy3 and Cy5 channel within a single array (saliva, patient B).

Figure 203 | MA-plot comparing the Cy3 and Cy5 channel within a single array (blood, patient B). The red coloured line illustrates a linear trend performed where the two channels are behaving similarly.

8.5.2.3 Conclusion

In each MA-plot the $\log_2$ (Cy5/Cy3) ratios are evenly distributed around the x-axis across all intensities. As this is a sex-mismatch experiment, the signal intensity ratios of the Y chromosome have only been determined by the male reference (Cy3) which are clustered in a small point cloud above the main point cloud. For saliva, intensity data appears tight around M=0. For blood, the major cloud of data is not exactly symmetrically scattered around zero.

9 Conclusion of experiments

We evaluated DNA from blood and saliva using a newly introduced extraction method in the laboratory. Both biological DNA sources were evaluated with array CGH. Both samples had a similar reproducibility in signal intensity and quality on Agilent custom 244k CGH arrays. Further experiments may be necessary to examine whether the slightly increased background noise for Cy5-labelled saliva DNA is a random or systematic measurement error. However, their signal-to-noise ratio still falls within an excellent range. In the two experiments we observed a blood-saliva correlation of 93% and 95% between total genome samples extracted from the same patient. Only one chromosomal region (chr14q11.2) gave discrepant results, namely the T-cell receptor region. This is a logical discrepancy since the T-cell receptor region is highly variable and unique to leukocyte DNA. In conclusion, the array CGH results from patient's blood DNA and Oragene saliva DNA appear highly comparable.

References

1. Speicher, M. R., Carter, N. P. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Rev. Genet.* **6**, 782–792 (2005).
2. Venter, J. C. *et al*. The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Jacobs, P.A., Baikie, A.G., Court Brown, W.M. and Strong, J.A. The somatic chromosomes in mongolism. *Lancet* **1**, 710 (1959).
4. Lucas, M., Kemp, N.H., Ellis, J.R., Marshall, R. A small autosomal ring chromosome in a female infant with congenital malformations. *Ann. Hum. Genet.* **27**, 189-195 (1963).
5. Rowley, J.D. A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290-293 (1973).
6. Lubs, H.A. A marker X chromosome. *Am. J. Hum. Genet.* **21**, 231–244 (1969).
7. Florijn, R. J. *et al.* High-resolution DNA fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. *Hum. Mol. Genet.* **4**, 831–836 (1995).
8. Knight, S. J., Fling, J. The use of subtelomeric probes to study mental retardation. Methods Cell Biol. 75, 799-831 (2004).
9. Kallioniemi, A. *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).
10. Weiss M.M., Kuipers, E.J., Meuwissen, S.G., van Diest, P.J., Meijer, G.A. Comparative genomic hybridization as a supportive tool in diagnostic pathology. *J. Clin. Pathol.* **56**, 522-527 (2003).
11. Barber, J.C.K., Joyce, C.A., Collinson, M.N., *et al.* Duplication of 8p23.1: A cytogenetic anomaly with no established clinical significance. *J. Med. Genet.* **35**, 491–496 (1998).
12. Lin, H., Pizer, E., Morin, P.J. A frequent deletion polymorphism on chromosome 22q13 identified by representational difference analysis of ovarian cancer. *Genomics* **69**, 391-394 (2000).
13. Carter, N.P. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature* **39**, S16-S21 (2007).
14. Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nature Genet.* **34**, 369–376 (2003).
15. Shaw-Smith, C., Redon, R., Rickman, L. *et al.* Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J. Med. Genet.* **41**, 241-248 (2004).

16. Rosenberg, C., Knijnenburg, J., Bakker, E. *et al.* Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents*. J. Med. Genet.* **43**, 180–186 (2006).
17. Menten, B. *et al.* Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J. Med. Genet.* **43**, 625–633 (2006).
18. Veltman, J. A. *et al.* High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.* **70**, 1269–1276 (2002).
19. Lockwood, W.W., Chari, R., Chi, B., Lam, W.L. Recent advances in array comparative genomic hybridization technologies and their applications in human genetics. *European Journal of Human Genetics* **14**, 139–148 (2006).
20. Wenli, G., Zhang, F., Lupski, J.R. Mechanisms for human genomic rearrangements *PathoGenetics* **1**(1), 4 (2008).
21. Feuk, L., Carson, A. R., Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet. **7**, 85-97 (2006).*
22. Scherer, S.W., Lee, C., Birney, E., *et al.* Challenges and standards in integrating surveys of structural variation. *Nature Genet.* **39**, S7-S15 (2006).
23. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
24. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
25. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
26. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., Carter, N.P. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524-533 (2009).
27. Ouahchi, K., Lindeman, N., Lee, C. Copy number variants and pharmacogenomics *Pharmacogenomics* **7**, 25-29 (2006).
28. McCarroll , S.A., Hadnott, T.N., Perry, G.H., Common deletion polymorphisms in the human genome. *Nat. Genet.* **38**, 86-92 (2006).
29. Database of Genomic Variants - http://projects.tcag.ca/variation/
30. Gonzalez, E., Kulkarni, H., Bolivar, H. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434 -1440 (2005).

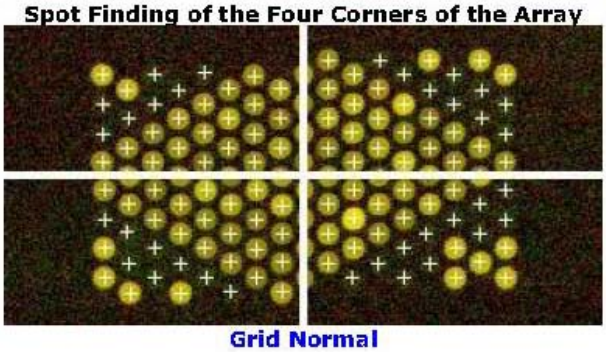
31. Kishawi, I. Agilent array technology and custom capabilities. [Internet] *Agilent Technologies*, published in 2008. Available from: http://www.chem.agilent.com/Library/posters/Public/Agilent%20custom.PDF
32. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **15**, 860-921 (2001).
33. Bignell GR, Huang J, Greshock J et al: High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* **14**, 287-295 2004
34. Lucito, R. Et al. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291-2305 (2003).
35. LeProust, E. Agilent's microarray platform: How high-fidelity DNA synthesis maximizes the dynamic range of gene expression measurements. [Internet] *Agilent Technologies*, published in 2008. Available from: http://www.imgm.com/fileadmin/IMGM/pdfs/agilent_application_note_5989-9159en_2008-08-31.pdf
36. SurePrint technology. [Internet] *Agilent technologies*, updated in 2009. Available from: http://www.chem.agilent.com/en-US/Products/Instruments/dnamicroarrays/Pages/gp557.aspx
37. Branham, W.S., Melvin, C.D., Han, T. *et al.* Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements *BMC Biotechnology* **7**, 8 (2007).
38. OG-500 Donor instructions.[Internet] *DNA Genotek*, updated in 2009. Available from http://www.dnagenotek.com/pdf_files/ART-PD-PR-061_Issue_2.0_OG-500_User_Instructions.pdf
39. Using saliva sponges to collect DNA samples from infants & young children. [Internet] *DNA Genotek*, published in 2006. Available from: http://www.dnagenotek.com/pdf_files/PD-PR-018_DNA%20collection%20from%20non-spitters%20protocol_issue%201_2.pdf
40. Comparison of DNA purified with Oragene DNA and the QIAamp™ Mini Kit. [Internet] *DNA Genotek*, published in 2004. Available from: http://www.dnagenotek.com/pdf_files/MKAN004_QIAmp.pdf
41. Oragene™ DNA Self-Collection Kit. The new "gold standard" for DNA collection, stabilization and preparation.[Internet] *DNA Genotek*, published in 2008 Available from: http://www.dnagenotek.com/pdf_files/PD-BR-012_Issue_1.0_OG-500_Brochure.pdf
42. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789-96 (2003).
43. Cohen, J. Statistical power analysis for the behavioral sciences (2nd ed.). (1988).
44. Finkelstein D, et al. Microarray data quality analysis: lessons from the AFGC project. *Plant. Mol. Biol.* **48**, 119–131 (2002).
45. Dudoit, S., Yang, Y. H, Speed, T. P., Callow, M. J. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12**, 111-140 (2002).

APPENDICES

QC Report - Agilent Technologies : 2 Color CGH

Date	Friday, March 27, 2009 - 11:00	Sample(red/green)	
User Name	md8	FE Version	10.5.1.1
Image	US22502573_252090510173_S01	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	020905_D_F_20080704	Dye Norm	Linear

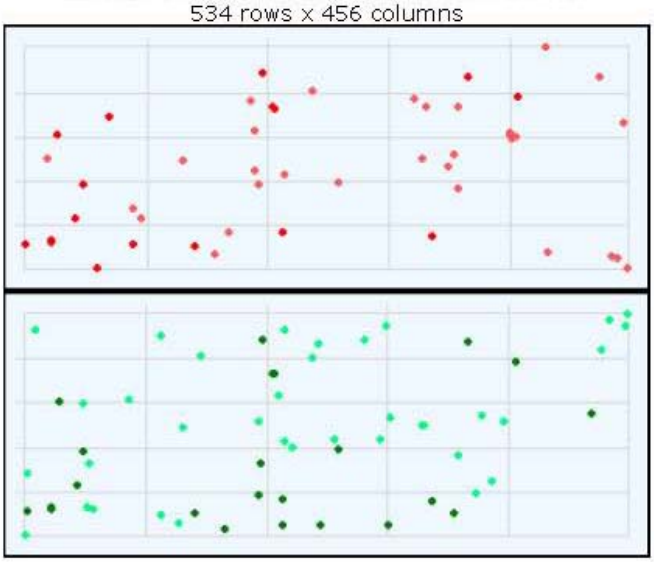
Spot Finding of the Four Corners of the Array



Grid Normal

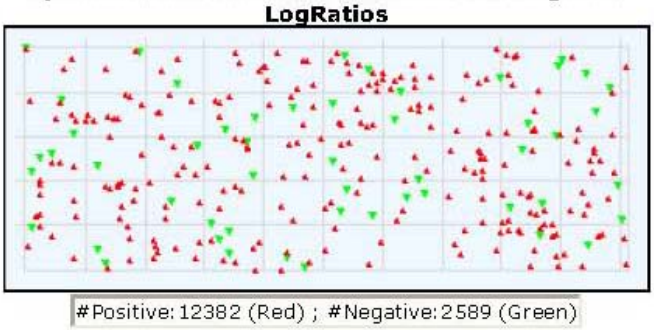
Outlier Numbers with Spatial Distribution

534 rows x 456 columns



Feature	Red	Green	Any	% Outlier
Non Uniform	18	26	30	0.01
Population	31	37	62	0.03

Spatial Distribution of the Positive and Negative LogRatios



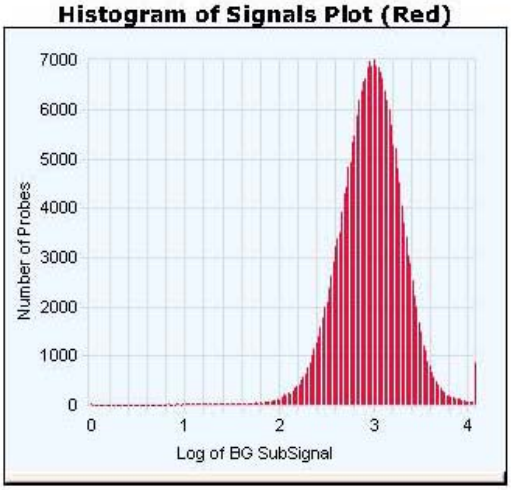
#Positive: 12382 (Red) ; #Negative: 2589 (Green)

Positive Negative

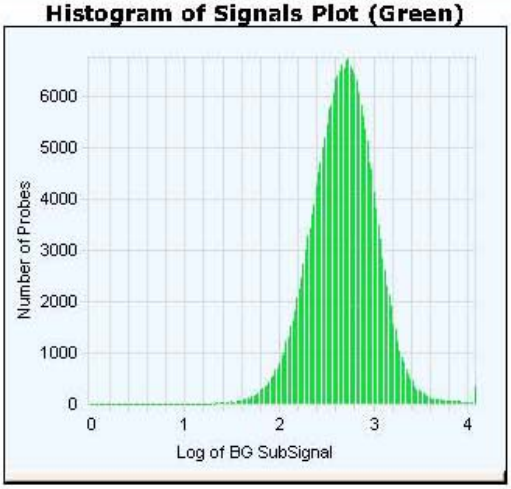
CGH_QCMT_Dec08 : (10 of 10) QCMetrics InRange

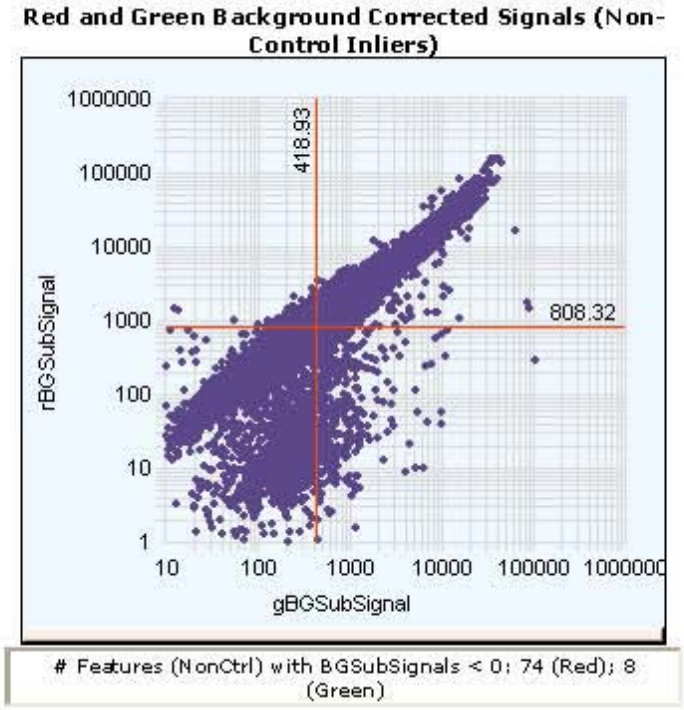
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.01	1.00	NA
DerivativeLR_Spread	0.12	0.30	NA
gRepro	-0.01	0.20	NA
g_BGNoise	1.91	15.00	NA
g_Signal2Noise	219.69	NA	30.00
g_SignalIntensity	418.93	NA	50.00
rRepro	-0.01	0.20	NA
r_BGNoise	4.29	15.00	NA
r_Signal2Noise	188.22	NA	30.00
r_SignalIntensity	808.32	NA	50.00

Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)

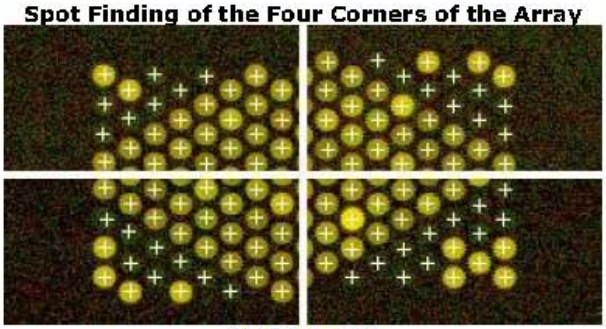



Red and Green Background Corrected Signals (Non-Control Inliers)
1000000
100000
10000
1000
100
10
1
418.93
808.32
rBGSubSignal
10
100
1000
10000
100000
1000000
gBGSubSignal
Features (NonCtrl) with BGSubSignals < 0: 74 (Red); 8 (Green)

QC Report - Agilent Technologies : 2 Color CGH

Date	Friday, March 06, 2009 - 10:12	Sample(red/green)	
User Name	ep2	FE Version	10.5.1.1
Image	US22502573_252090510127_S01	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	020905_D_F_20080704	Dye Norm	Linear

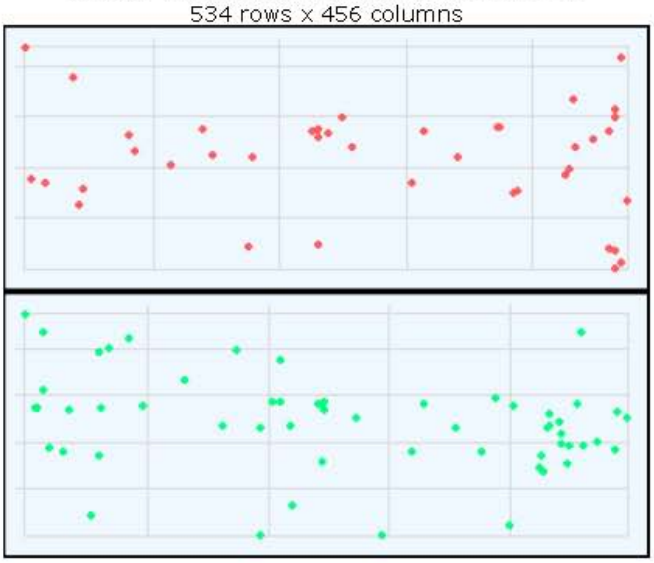
Spot Finding of the Four Corners of the Array



Grid Normal

Outlier Numbers with Spatial Distribution

534 rows x 456 columns

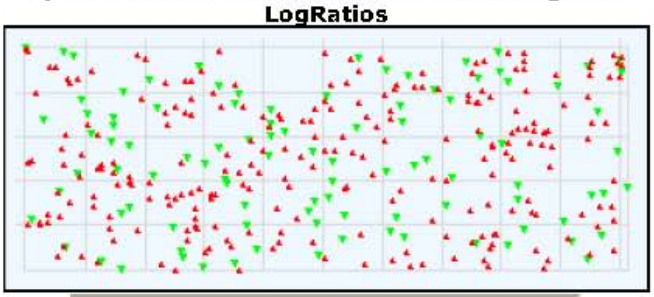


Red FeaturePopulation Red Feature NonUniform

Green FeaturePopulation Green Feature NonUniform

Feature	Red	Green	Any	% Outlier
Non Uniform	0	0	0	0.00
Population	42	57	84	0.03

Spatial Distribution of the Positive and Negative LogRatios



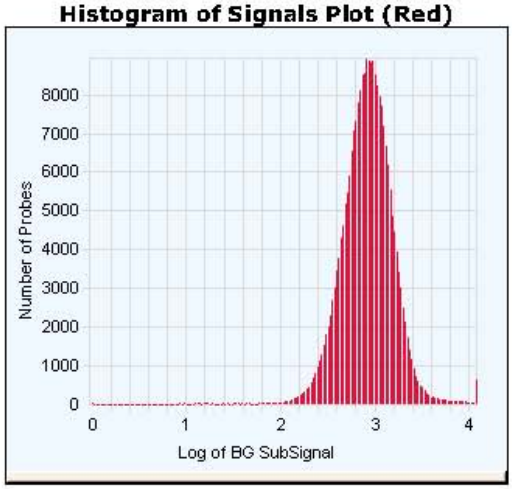
#Positive: 11931 (Red) ; #Negative: 4917 (Green)

Positive Negative

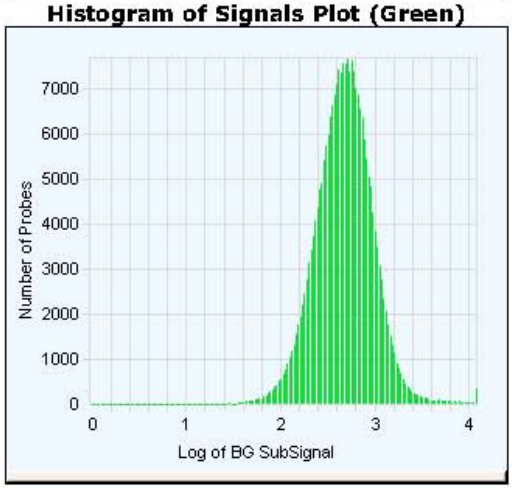
CGH_QCMT_Dec08 : (10 of 10) QCMetrics InRange

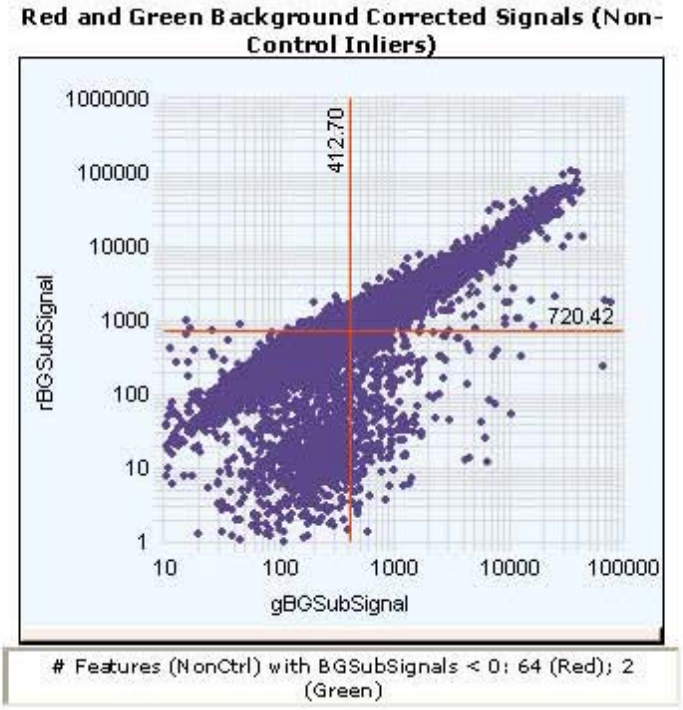
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.00	1.00	NA
DerivativeLR_Spread	0.17	0.30	NA
gRepro	-0.01	0.20	NA
g_BGNoise	2.34	15.00	NA
g_Signal2Noise	176.35	NA	30.00
g_SignalIntensity	412.70	NA	50.00
rRepro	-0.01	0.20	NA
r_BGNoise	5.16	15.00	NA
r_Signal2Noise	139.75	NA	30.00
r_SignalIntensity	720.42	NA	50.00

Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)

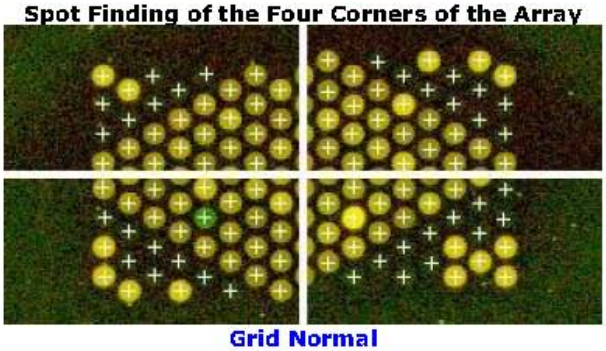



Red and Green Background Corrected Signals (Non-Control Inliers)
1000000
100000
10000
1000
100
10
1
rBGSubSignal
412.70
720.42
10
100
1000
10000
100000
gBGSubSignal
Features (NonCtrl) with BGSubSignals < 0: 64 (Red); 2 (Green)

QC Report - Agilent Technologies : 2 Color CGH

Date	Thursday, March 19, 2009 - 10:18	Sample(red/green)	
User Name	md8	FE Version	10.5.1.1
Image	US22502573_252090510203_S01	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	020905_D_F_20080704	Dye Norm	Linear

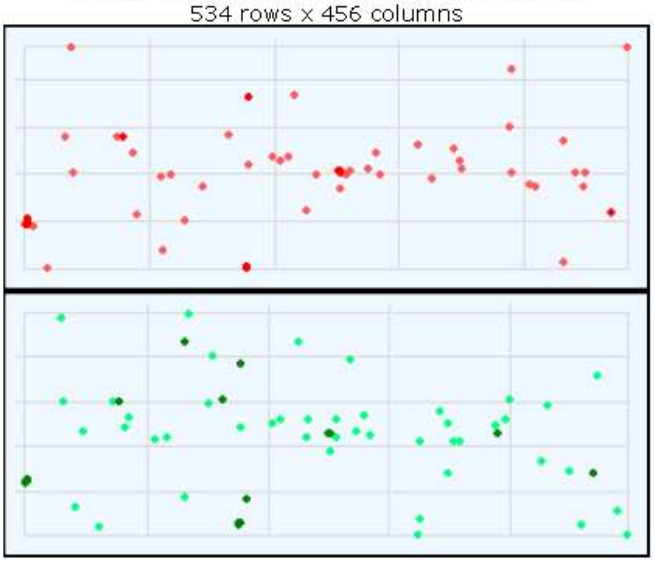
Spot Finding of the Four Corners of the Array



Grid Normal

Outlier Numbers with Spatial Distribution

534 rows x 456 columns

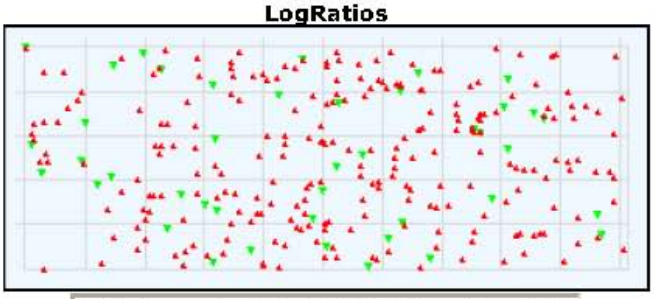


Red FeaturePopulation Red Feature NonUniform

Green FeaturePopulation Green Feature NonUniform

Feature	Red	Green	Any	% Outlier
Non Uniform	24	20	28	0.01
Population	44	47	71	0.03

Spatial Distribution of the Positive and Negative LogRatios



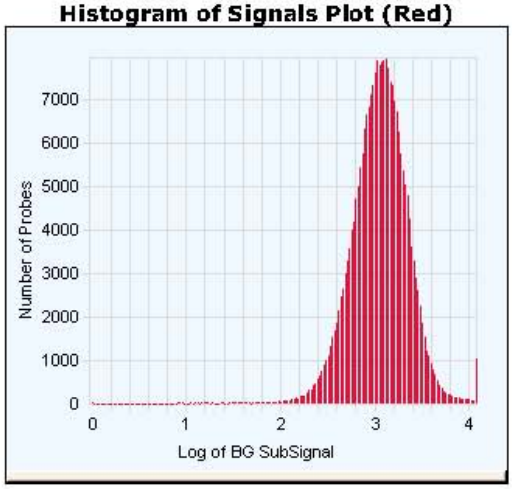
#Positive: 12199 (Red) ; #Negative: 2066 (Green)

Positive Negative

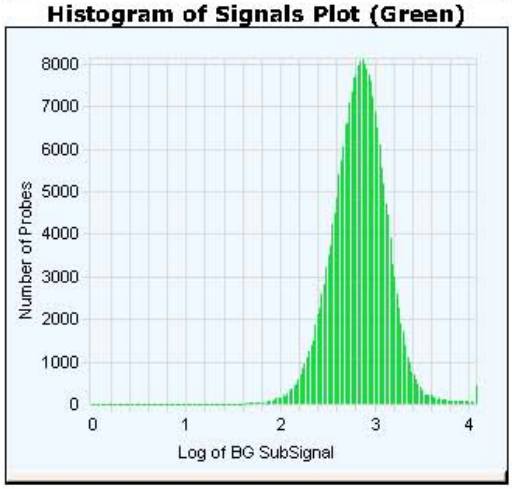
CGH_QCMT_Dec08 : (10 of 10) QCMetrics InRange

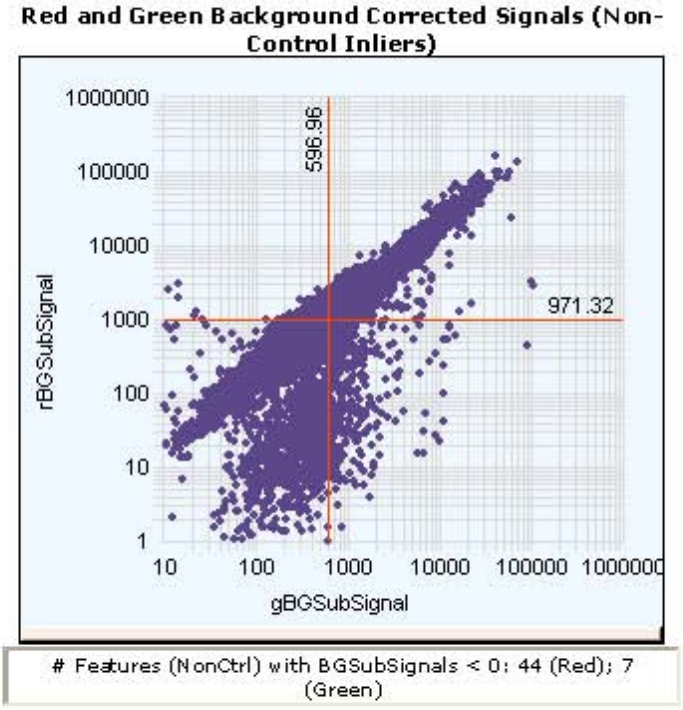
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.01	1.00	NA
DerivativeLR_Spread	0.13	0.30	NA
gRepro	-0.01	0.20	NA
g_BGNoise	2.08	15.00	NA
g_Signal2Noise	287.08	NA	30.00
g_SignalIntensity	596.96	NA	50.00
rRepro	-0.01	0.20	NA
r_BGNoise	4.31	15.00	NA
r_Signal2Noise	225.45	NA	30.00
r_SignalIntensity	971.32	NA	50.00

Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)

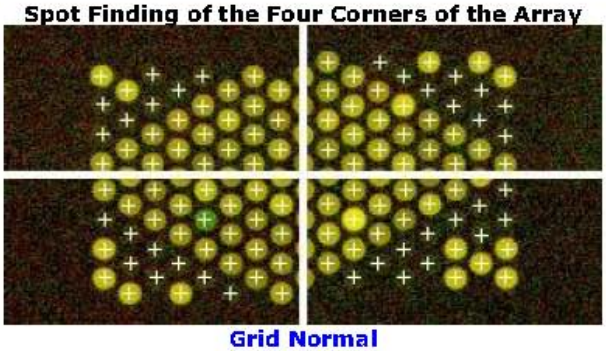



Red and Green Background Corrected Signals (Non-Control Inliers)
1000000
100000
10000
1000
100
10
1
rBGSubSignal
596.96
971.32
10
100
1000
10000
100000
1000000
gBGSubSignal
Features (NonCtrl) with BGSubSignals < 0: 44 (Red); 7 (Green)

QC Report - Agilent Technologies : 2 Color CGH

Date	Thursday, February 26, 2009 - 13:22	Sample(red/green)	
User Name	ep2	FE Version	10.5.1.1
Image	US22502573_252090510125_S01	BG Method	Detrend on (NegC)
Protocol	CGH_105_Dec08 (Read Only)	Multiplicative Detrend	True
Grid	020905_D_F_20080704	Dye Norm	Linear

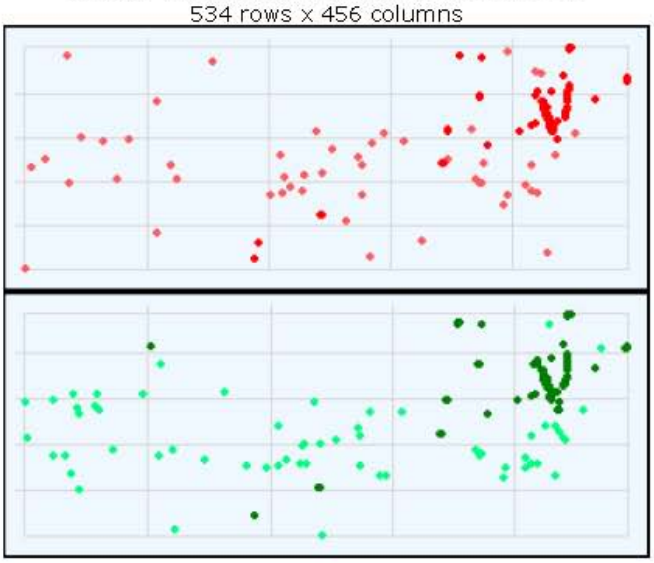
Spot Finding of the Four Corners of the Array



Grid Normal

Outlier Numbers with Spatial Distribution

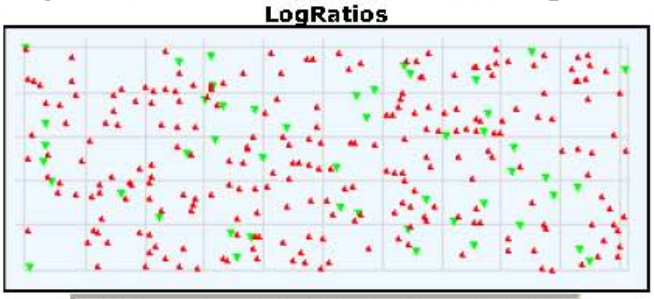
534 rows x 456 columns



Red FeaturePopulation Red Feature NonUniform
Green FeaturePopulation Green Feature NonUniform

Feature	Red	Green	Any	% Outlier
Non Uniform	94	101	103	0.04
Population	51	61	83	0.03

Spatial Distribution of the Positive and Negative LogRatios



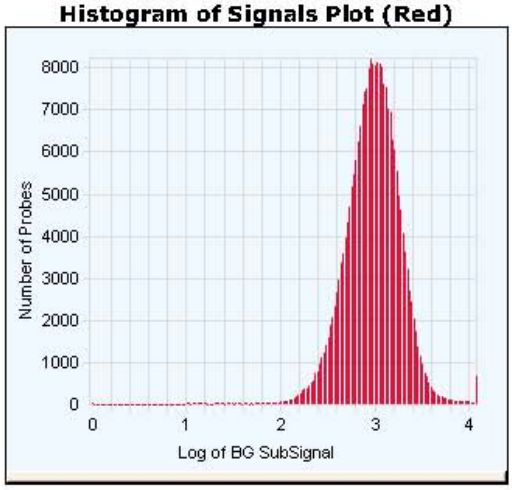
#Positive: 12056 (Red) ; #Negative: 2446 (Green)

Positive Negative

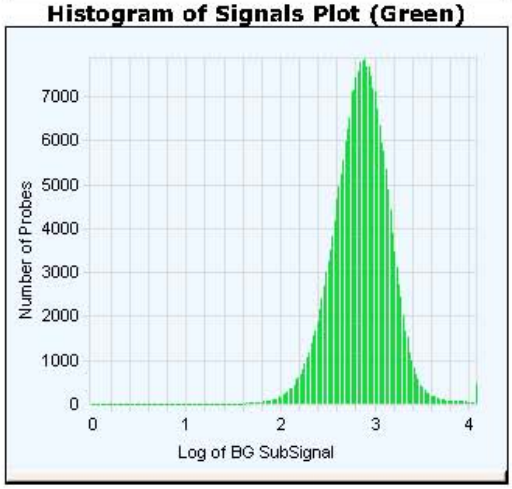
CGH_QCMT_Dec08 : (10 of 10) QCMetrics InRange

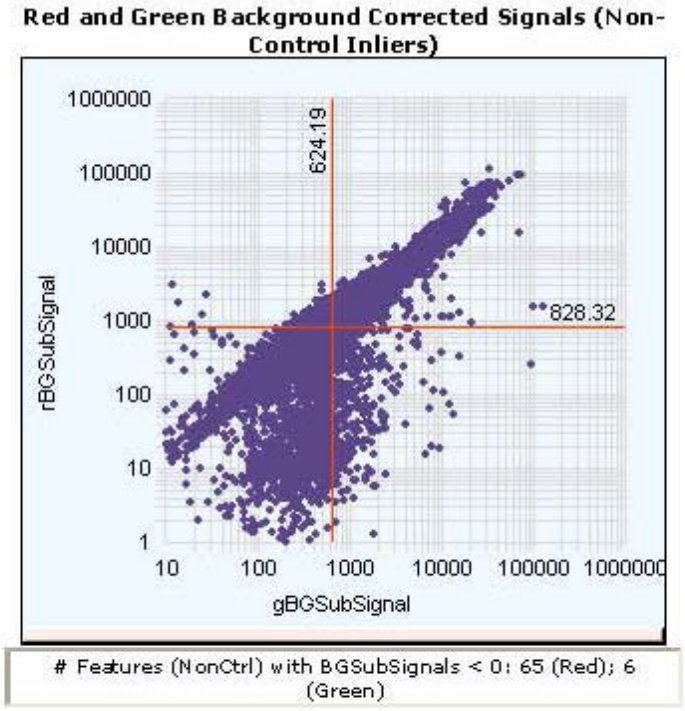
Metric Name	Value	UpLim	LowLim
AnyColorPrcntFeatNonUnifOL	0.04	1.00	NA
DerivativeLR_Spread	0.11	0.30	NA
gRepro	-0.01	0.20	NA
g_BGNoise	2.66	15.00	NA
g_Signal2Noise	234.65	NA	30.00
g_SignalIntensity	624.19	NA	50.00
rRepro	-0.01	0.20	NA
r_BGNoise	6.46	15.00	NA
r_Signal2Noise	128.21	NA	30.00
r_SignalIntensity	828.32	NA	50.00

Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)




Red and Green Background Corrected Signals (Non-Control Inliers)
1000000
100000
10000
1000
100
10
1
624.19
828.32
rBGSubSignal
10
100
1000
10000
100000
1000000
gBGSubSignal
Features (NonCtrl) with BGSubSignals < 0: 65 (Red); 6 (Green)