

Strategies for the detection of copy number and other structural variants in the human genome

Andrew R. Carson,¹ Lars Feuk,¹ Mansoor Mohammed² and Stephen W. Scherer^{1*}

¹The Centre for Applied Genomics and Program in Genetics and Genomic Biology, The Hospital for Sick Children and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada

²CombiMatrix Diagnostics Corporation, Irvine, CA, USA

*Correspondence to: Tel: +1 416 813 7613; Fax: +1 416 813 8319; E-mail: swscherer@sickkids.ca

Date received (in revised form): 21st April 2006

Abstract

Advances in genome scanning technologies are revealing that copy number variants (CNVs) and polymorphisms, ranging from a few kilobases to several megabases in size, are present in genomes at frequencies much greater than previously known. Discoveries of additional forms of genomic variation, including inversions, insertions, deletions and complex rearrangements, are also occurring at an increased rate. Along with CNVs, these sequence alterations are collectively known as structural variants, and their discovery has had an immediate impact on the interpretation of basic research and clinical diagnostic data. This paper discusses different methods, experimental strategies and technologies that are currently available to study copy number variation and other structural variants in the human genome.

Keywords: copy number and structural variation, genome-scanning techniques

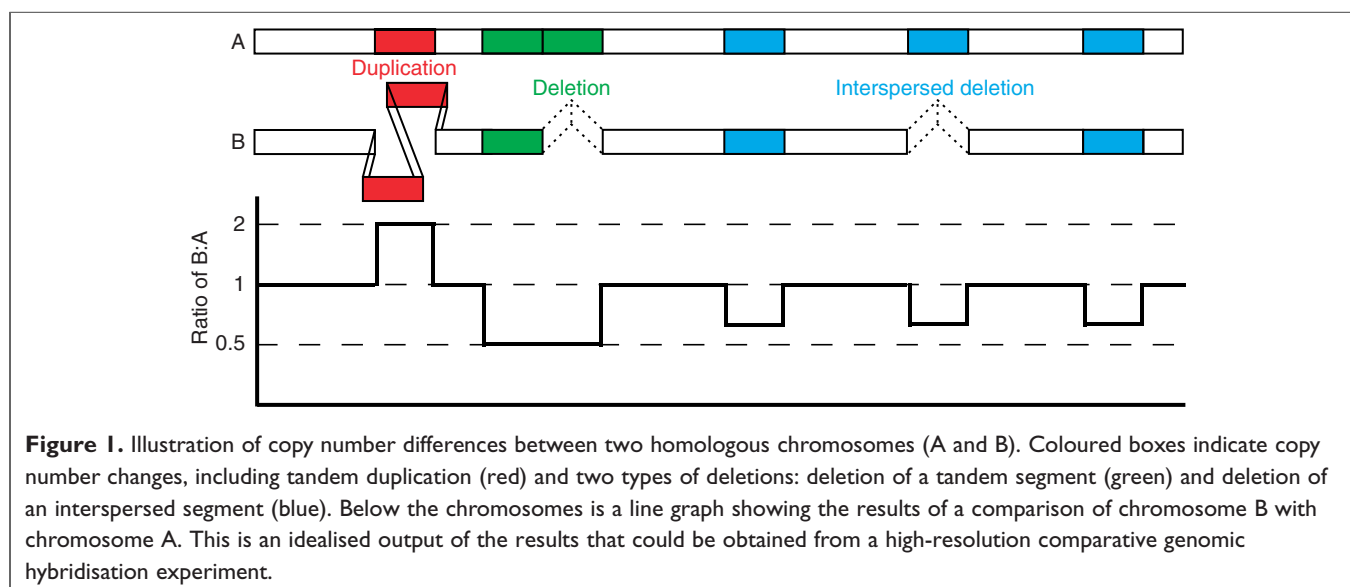
Introduction

The capacity for targeted or *en masse* detection of variation in the human genome is dictated by the resolution of the available technologies. In the early years of human genetics, variation was detected by studying chromosomes under microscopes, with notable observations of aneuploidy,^{1–3} heteromorphism⁴ and fragile sites,⁵ to name a few, dominating our knowledge base. The advent of molecular biology, and in particular nucleotide resolution analysis through DNA sequencing and genotyping, led to the discovery, characterisation and mapping of short tandem repeats (STRs; eg di-, tri- and tetranucleotide microsatellites)⁶ and single nucleotide polymorphisms (SNPs).^{7,8} STRs and SNPs, and the technologies used to detect them, are described in numerous comprehensive articles and reviews.^{9–12}

The latest advances in studying human genome variation have been in the examination of copy number variants (CNVs) (Figures 1 and 2) and other similarly sized structural changes along human chromosomes. In general, this class of variants refers to changes of an intermediate size, between micro- or minisatellites and microscopically visible changes (usually >1 kilobase [kb] and <3 megabases). Several recent investigations have found that these variants are much more frequent in the human genome than previously recognised^{13–20} and

in-depth descriptions of these variants and their properties can be found in several recent reviews.^{21–29}

Investigations of structural variants have been accompanied by a host of newly developed technologies and methodologies, with many of these latest-generation techniques being currently implemented and continually improved upon. As more techniques arise, the most commonly asked question seems to be, 'What is the most appropriate technique or experimental approach to address our specific question?' Several factors can be used to distinguish different techniques and should be considered before embarking on a new study. These include the scope of the technique (does the method assess targeted or genome-wide variation?) and the resolution of the technique (what types and sizes of variants need to be analysed?) (Figure 2). Here, the techniques and methodologies currently available for the analysis of structural variation (in particular, copy number variation) in the human genome are reviewed. Important factors in these analyses are highlighted, while, at the same time, recommendations for these strategies are made — in some cases based on the authors' own personal experiences. Fluorescence *in situ* hybridisation (FISH) is not discussed here because it is well established and has been previously reviewed in detail.³⁰ It should be noted, however, that FISH is often the only way to assess certain forms of structural variation along the chromosomes.



Methods for detecting and scoring copy number variation

The discovery and characterisation of structural variation in the human genome has been driven by advances in methods that allow comprehensive scanning of an entire genome, along with targeted scans of defined loci. Originally, technologies such as Southern blot hybridisation — using conventional and pulsed-field gel electrophoresis, FISH and microsatellite scanning offered the best methods for revealing changes in DNA copy number. In some cases, particularly for disease diagnostics, these methodologies are still the standard and may continue to offer the only way to correctly resolve a complex rearrangement. Each of these approaches lacks scalability, however, and may also require parental DNA to facilitate data interpretation. Notwithstanding, the prevailing approaches for identifying CNVs have been either array based or quantitative (primarily, polymerase chain reaction [PCR]-based) assays. The former is used primarily for genome scanning, while the latter is used for locus-specific testing (often with multiple loci screened simultaneously) in first-pass and confirmatory screening (for example, to confirm genome scan data). Additional methods for detecting CNVs, including what are referred to as computationally based and genotype-based assays, have come forth recently. A flowchart outlining the available methodologies is given in Figure 2, while some techniques are discussed in greater detail below and illustrated in Figure 3.

Array-based methods for detection of copy number variation

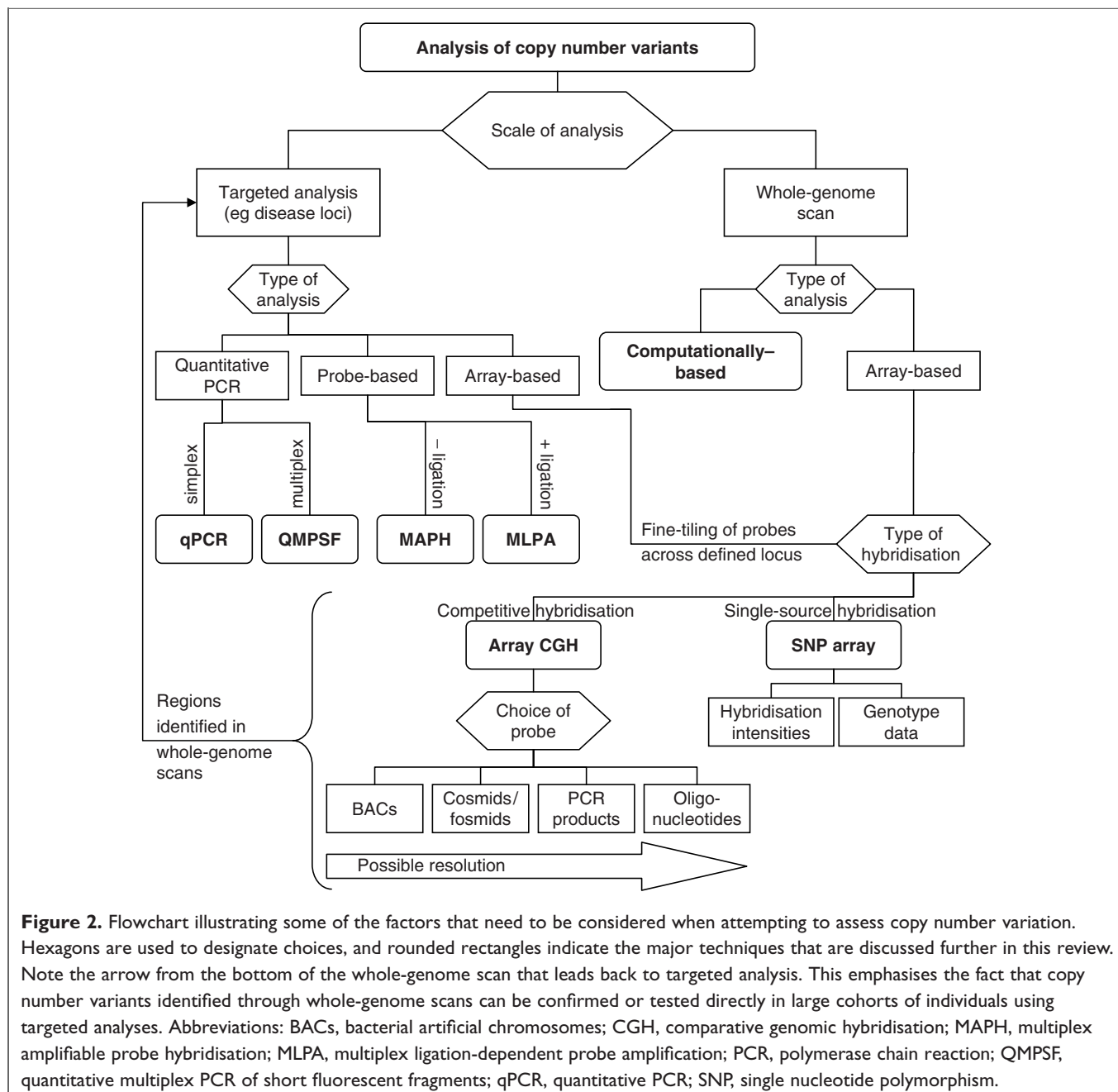
Array-based methods offer the most robust approach to scanning for CNVs in a genome-wide manner. The main differences between available platforms are in the type of DNA

molecule found on the array (genomic DNA clones, cDNA, PCR products or oligonucleotides) and the type of hybridisation used (competitive hybridisation versus single-source hybridisation). These different platforms each have inherent advantages and disadvantages.

Comparative genomic hybridisation

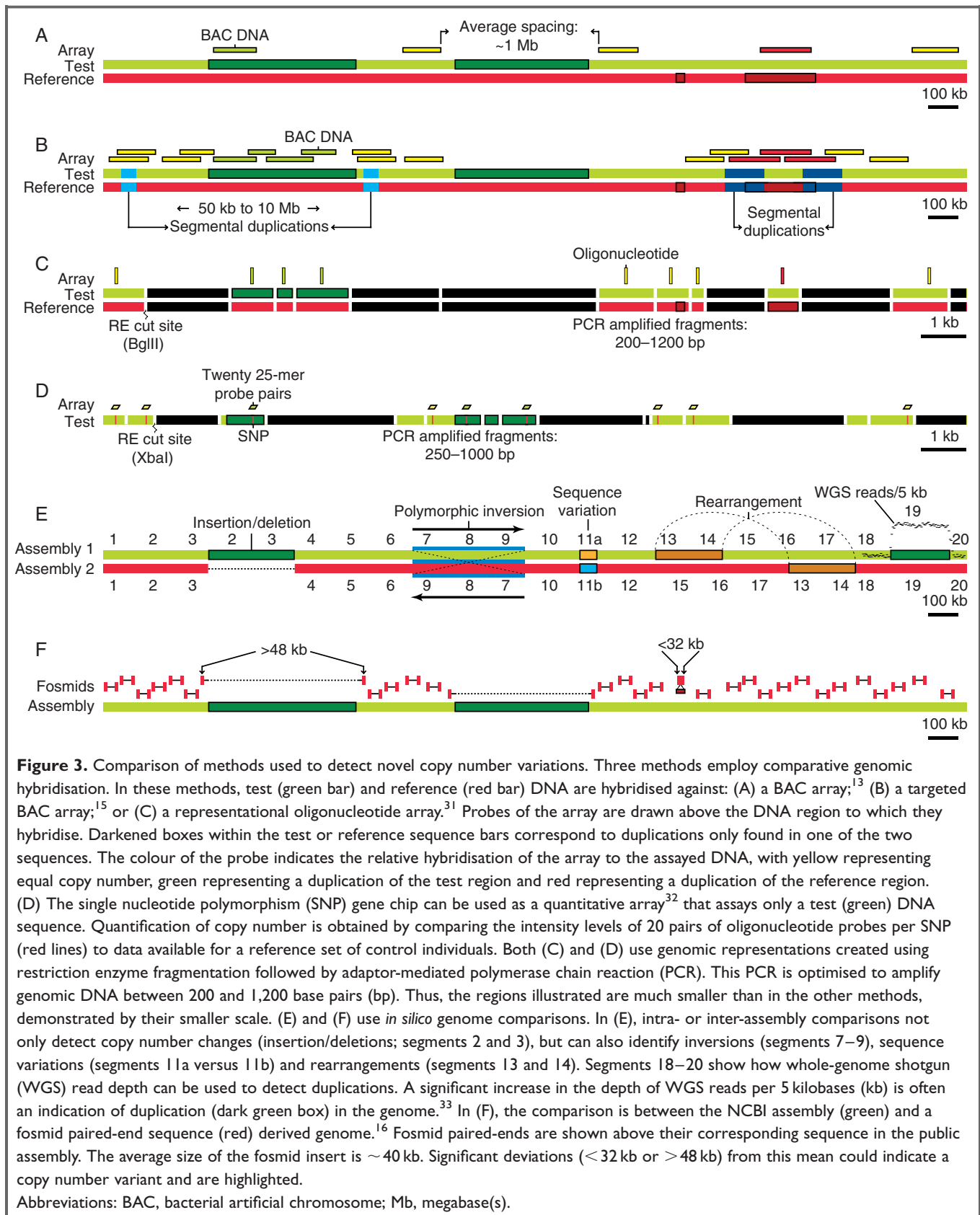
Detection of DNA copy number variation between differentially labelled test and reference genomes has long been feasible using competitive *in situ* hybridisation to metaphase spreads.³⁴ This approach, in which fluorescence ratios between the two hybridised DNA sources reveal regions of gain or loss, is referred to as comparative genomic hybridisation (CGH). With the advent of microarrays, sample DNA can now be hybridised to arrays spotted with a multitude of DNA sequences. This not only increases the specificity of CGH, but it also provides a dramatic increase in resolution. Array-based CGH (aCGH) was first described in 1997.³⁵ Two years later, the first experiment to employ a genome-wide scanning approach utilised an array spotted with cDNA clones.³⁶ As the technology evolved, bacterial artificial chromosomes (BACs) became the most common type of genomic clone for spotting on the arrays (Figure 3A).^{13,15,37–43} There are several advantages to using BAC clones, including their availability, coverage of the genome, known sequence identifiers (eg completed sequence or clone end-sequence) and FISH-readiness. Moreover, BACs and other large insert clones retain a genomic complexity that often results in a more robust signal-to-noise ratio profile.

There are now several BAC arrays available that provide partial or near-complete genome-wide coverage (genome-wide coverage is also referred to as clone ‘tiling-path’ resolution analysis).^{38–42} There are also commercial BAC arrays



available, both for genome-wide screening and for targeting regions involved in microdeletion syndromes (eg Spectral Genomics and Signature Genomics). Additionally, BAC arrays have been designed that target potential rearrangement hotspots that are flanked by highly identical segmental duplications¹⁵ (Figure 3B). BAC arrays do, however, have some inherent disadvantages. First, even with dense coverage, the resolution of BAC arrays is limited by the size of the clones themselves. Since BAC clones are generally 80–200 kb in size, it can be difficult to identify CNVs smaller than ~50 kb. Higher-resolution analysis can be achieved by spotting shorter

DNA molecules, such as cosmids, fosmids, cDNA,³⁶ PCR products^{44,45} or oligonucleotides on the array.^{14,31,46–49} Secondly, to the authors' knowledge, there are currently no available commercial services providing 'tiling-path' resolution arrays. Presumably, this is because of the inherent challenges in manufacturing this type of product. An academic enterprise from the British Columbia Cancer Research Centre (<http://www.bccrc.ca/arraycgh/arrays.html>) is the only provider of such arrays, although there may be other vendors that the authors are unaware of. Moreover, some groups offer access to their arrays on a collaborative basis.



Spotting an array with oligonucleotides is one way of achieving an increased resolution. The use of oligonucleotide CGH arrays for copy number detection was first implemented in an assay format called representational oligonucleotide microarray analysis (ROMA)^{14,31} (Figure 3C). Similarly to BAC arrays, ROMA utilises differentially labelled test and reference genome DNA competitively hybridised to an array. In order to increase the signal-to-noise ratio, however, the complexity of the input DNA is reduced. This is accomplished by a method called representation or whole-genome sampling.⁵⁰ In this method, the genomic DNA is fragmented with a specific restriction enzyme. The resulting fragments are then ligated with adaptors which act as primer binding sites in a subsequent PCR amplification. The amplification conditions are set only to amplify fragments up to ~1.2 kb. The initial version of this method used *Bgl*II digestion, which gives rise to an estimated 200,000 fragments in this size range.³¹ Oligonucleotides for the array are then designed to match sequences present in the (complexity-reduced) representation sample. The published ROMA arrays have contained 85,000 oligonucleotides that are each 70 base pairs (bp) in length, giving a resolution of ~30 kb throughout the genome. A potential disadvantage of these first-generation ROMA arrays is that, in general, only unique regions in the genome were represented. This means that the ~5 per cent of the genome covered by low copy repeats (LCRs — also called segmental duplications)^{33,51} would not generally be assayed, even though a large fraction of the existing copy number variation has been shown to reside in these regions.^{13,15,52} The omission of duplicate regions will most likely be remedied in higher resolution arrays, spotted with a larger number of oligonucleotides, which are expected to be forthcoming (ROMA platforms using arrays with >300,000 probes are now being developed). As reflected in the published literature, the ROMA technology seems to be primarily used by its inventor, Dr Michael Wigler, along with his collaborators, and has not yet been adopted by the majority of users.

In addition to the ROMA methodology, companies such as Agilent (<http://www.agilent.com>) and NimbleGen (<http://www.nimblegen.com>) have generated long oligonucleotide arrays for direct (non-representational) CGH analysis. Agilent now produces an array with ~43,000 60-mer probes with bias toward genic and non-repetitive euchromatic regions.^{43,53,54} Experiments using this array can be performed with as little as 100 ng of input DNA per sample and provides a resolution that can discern large chromosomal aberrations but lacks the power to pick up small CNVs. To address the need for greater resolution, Agilent is expecting to release next-generation arrays with ~185,000 and >300,000 probes early next year. NimbleGen offers both whole-genome and custom-targeted 'fine-tiling' arrays containing 385,000 probes.^{18,55} These provide a mean spacing of ~7–8 kb on the whole-genome array and can give a resolution better than 500 bp (with probes as densely spaced as 10 bp apart) in custom fine-tiling arrays.

NimbleGen uses probes that vary in size between 45–85 bp, such that their T_m s are equalised at 76°C ('isothermal'), thus providing more uniform probe performance. In the NimbleGen business model, however, the NimbleGen staff performs all of the hybridisation experiments, requiring customers to supply 1–3 µg of purified sample and reference genomic DNA. Data analysis is also heavily based on company-designed algorithms. Although this may be preferable for many laboratories, some may find it difficult to provide such a relatively large quantity of DNA. Other laboratories will not be able to provide DNA because of research ethics restrictions governing how their DNA samples can be handled and disseminated.

SNP chips

A slightly different approach to using oligonucleotide arrays for copy number detection is to use the hybridisation intensities from SNP genotyping platforms, such as those made by Affymetrix or Illumina.^{32,56–62} Although both of these approaches use oligonucleotides spotted on an array, the SNP arrays differ from CGH in a number of fundamental ways. Most notably, the SNP arrays were originally designed for genotyping, so they do not directly compare two DNA sources through competitive hybridisation, as is the case with CGH. Instead, hybridisation intensities from a single DNA source are adapted to provide information about copy number through subsequent comparisons with a set of reference values from control individuals.

With the Affymetrix SNP arrays, input DNA preparation is similar to ROMA in that the DNA to be hybridised is reduced in complexity by restriction digestion followed by PCR-based amplification of fragments in a specific size range. These DNA fragments are then hybridised to an array where each SNP is represented by 20 probe pairs representing the match and mismatch alleles⁵⁷ (Figure 3D). Comparison of the intensity values from these match and mismatch probes to reference values from control individuals is used to acquire copy number data. The highest resolution array currently available assays roughly 500,000 SNPs, with an average spacing of ~6 kb. This high density of probes allows consecutive SNPs to be used to estimate copy number and effectively increases the signal-to-noise ratio. Although the distribution of probes along the euchromatic regions is relatively uniform, there is under-representation within segmentally duplicated regions. As reflected in numerous publications demonstrating applicability, the most common publicly available algorithms used to extract the copy number intensity data include the Dchip⁶³ and CNAT⁵⁷ packages.

Instead of spotting oligonucleotides on glass slides, Illumina attaches probes to beads. The coated beads are pooled and then positioned randomly on an array, giving approximately 30-fold feature redundancy. Because the beads are positioned randomly, a series of hybridisations are needed to identify

the probe locations along the array. Illumina has two genotyping platforms that can be used for the detection of CNVs: GoldenGate and Infinium. Whereas GoldenGate can be used to multiplex genotyping experiments (up to 1,536-plex), the newer Infinium platform — which utilises 100K or 317K ‘beadchips’ — is the most viable option for whole-genome analyses. In this platform, input DNA is hybridised to the probes, this is followed by a single base extension at the SNP with differentially labelled nucleotides. Like the Affymetrix SNP arrays, the intensity values from this hybridisation are compared with values from control individuals to gain information about copy number. At the time of writing, there were only a few studies that used the Illumina platforms to identify CNVs. It is expected that this will soon change, however, as more initiatives incorporate copy number identification into the experimental design of their genetic studies.⁶⁴

Genotype-based methods for identifying deletions

Along with hybridisation intensities, SNP platforms also generate genotype data. These genotypes, when generated from multiple related individuals, afford additional opportunities to assess genomic deletions. With a single individual, genotyping will not detect deletions due to the fact that hemizyosity will be miscalled as homozygosity for the present allele (or, in the case of homozygous deletion, a null or failed genotype will be observed). By contrast, if parent–offspring trios are analysed, losses of heterozygosity^{65,66} can be discovered when Mendelian inheritance is violated. These can lend support to regions identified as deletions when assessing hybridisation intensities. Naturally, non-Mendelian inheritance can also signify other genomic alterations, including segmental or complete uniparental isodisomy or heterodisomy.^{67,68}

Additionally, large collections of genotype data, such as those generated by the International HapMap Consortium, can be used to identify deletion polymorphisms.^{18–20} Two recent studies have utilised this resource to identify deletions in the human genome by looking for violations of Mendelian inheritance, Hardy–Weinberg disequilibrium or null genotypes.^{19,20} An earlier study of patient samples offers a prototypic example of how this approach can be used in a disease study.⁶⁹ Certain regions of the genome have extensive marker coverage, giving this approach a high resolution with the power to detect relatively small (as small as 1 kb) deletions.

Interestingly, many of the deletions appear to be in high linkage disequilibrium (LD) with neighbouring SNPs.^{18,20,70} This could indicate that they have arisen on a single chromosome from an ancestral population. This is important, as it allows deletion carrier status to be inferred from SNP genotypes and further allows the investigation of associations between deletions and disease. It is also expected that other

CNVs, such as insertions/deletions of regions with copy number greater than two, will show significant LD with neighbouring SNPs. To date, however, no reports have been published to support this assumption. Additionally, LD associations are expected to be more complex in regions that have many copies (such as segmental duplications), making complex disease association analyses more difficult.

Quantitative methods for locus-specific testing

Quantitative methods offer an effective approach to assessing variation at targeted loci. These can be used to assess copy number changes at known or proposed disease regions, often with high-throughput screening of large cohorts of samples. Alternatively, these methods can be used to confirm regions that have been identified in whole-genome scans using array-based approaches. In general, the major difference between quantitative methods is in how they analyse the input DNA; do they directly assay the genomic DNA with PCR or do they first hybridise with a targeted probe? These differences are discussed in more detail below.

Quantitative PCR

Real-time quantitative PCR (qPCR) has been used for many years in the quantification of gene expression.⁷¹ The basis of qPCR is that the rate of amplification is proportionate to the number of template copies. By monitoring the amplification in ‘real-time’, it is possible to determine when the PCR reaction is in the exponential phase of amplification. It is in this phase that quantification of the starting template occurs; however, it is also necessary to use a control region with known copy number to adjust for variable starting amounts between samples. There are several types of qPCR assay, but they are all based on the same basic principle: an increase in PCR product is manifested as an increase in fluorescence, which can be monitored throughout the PCR reaction. Although most available protocols for real-time qPCR work well for scoring deletions and duplications,^{72,73} they are, generally, not suitable for multiplexing. To facilitate scoring of multiple target regions in a single experiment, some novel approaches have been developed. These assays are similar in that the final product is separated by size within a gel, which is inherently more amenable to multiplexing than spectral separation of products.

The simplest of these assays is called quantitative multiplex PCR of short fluorescent fragments (QMPSF).^{74,75} PCR assays are designed to amplify up to ten target regions in parallel, with each product varying in length. One primer for each target is labelled with a 6-FAM (6-carboxyfluorescein) moiety, while the other primer carries a short stabilising tail sequence. PCR amplification is stopped within the

exponential phase and products of different size are separated by electrophoresis. Each product is represented by a peak, and the peak height — relative to a reference — correlates with the amount of product. Because each reaction requires a labelled primer, experiments can become expensive if many regions are tested. To lower the cost, an alternative protocol has been developed,⁷⁶ in which all primers are designed with a specific hexa-decamer tail sequence. These tails then serve as templates in a subsequent amplification reaction, where universal labelled primers are used. Although this adds an extra step to the protocol, only a single FAM-labelled primer is needed, irrespective of the number of targets.

Probe-based multiplex assays

Multiplex amplifiable probe hybridisation (MAPH)^{75,77–79} and multiplex ligation-dependent probe amplification (MLPA)^{80–82} allow differences in copy number to be detected based on quantification of probes specific for a target region.

In MAPH, the test DNA is denatured, bound to a nylon filter and then hybridised with probes specific for the target region. Each probe has a different length but they all carry identical tail sequences, allowing subsequent amplification with fluorescently-labelled universal primers. After amplification, products are separated by size and quantified based on the fluorescence intensity ratio of target compared with control regions. Up to 40 loci can be interrogated simultaneously.⁸³ The main disadvantage of MAPH is the amount of work and optimisation needed to obtain a robust probe set. Each probe has to be cloned in order to add the universal tail sequences. Once a probe set has been developed, however, it can be used for high-throughput screening of all exons of a specific gene (eg *DMD*⁸⁴) for deletions or duplications in large patient cohorts.

MLPA is different from MAPH in that it is performed in solution, it is, however, still dependent on probe amplification. In MLPA, pairs of probes are made for each target. The two probes in each pair are designed to hybridise adjacent to each other at the target region. Through a DNA ligation step, a contiguous probe molecule is created. The probes carry a tail sequence that serves as a template for universal fluorescently-labelled primers in a subsequent amplification step. The resulting products can then be separated and quantified in the same manner as in MAPH. In the initial protocol for MLPA, each probe was cloned in a vector. A more recent advance demonstrates that synthesised probes work equally well, but there is a limit to the size of the probes that can be produced. This, however, can be overcome by introducing two different universal primer pairs, each labelled with a different fluorescent marker, thereby allowing for separation of the final products based on both size and wavelength of fluorescence.^{85,86} MLPA has been successful using up to 40 probes in a single experiment.⁸⁰ As with MAPH, once a probe

set has been developed, it works reproducibly in screening large cohorts of samples.⁸² There are more than 50 commercially available pre-tested MLPA probe kits designed for many of the known microdeletion syndromes, as well as for genes where intragenic deletions and duplications are common causes of disease (<http://www.mrc-holland.com>).

Computationally-based methods for detecting structural variants

While the above techniques physically assay DNA molecules to assess copy number variation, it is also possible to evaluate genomes *in silico* by comparing DNA sequences. As more sequence data become available, this option will become more viable and popular. Three main strategies have emerged, utilising different types of sequence data: (a) sequence assemblies (Figure 3E); (b) clone end sequences (Figure 3F); and (c) sequence read depths (Figure 3E), although these methods are largely limited to the analysis of data that are already publicly available from large-scale sequencing initiatives. This is due to the current, hugely prohibitive cost of generating full sequence coverage or redundant clone library end-sequence data from an individual's DNA.

Whole-genome or chromosome assemblies have the benefit of being able to detect practically any type of variation, even down to the single nucleotide. This provides an advantage over array-based methods, where the resolution is dependent on the density of probes spotted on the array. In this strategy, sequence assemblies from two sources are compared computationally, allowing differences in sequence, copy number or orientation to be annotated. Although the majority of these strategies compare two human assemblies, such as the two distinct and near-complete assemblies of chromosome 7^{87–89} and the HLA region,⁹⁰ it is also possible to detect large variations by comparing the human genome with its closest living relative, the chimpanzee. Although these interspecies comparisons primarily look for sequence variation between species, they can also identify polymorphic differences, such as inversions¹⁷ and copy number differences.⁹¹

Using clone end-sequences (eg fosmids) from a genomic library constructed from a single genome is another sensitive approach that allows for the detection of variants as small as 8 kb in one study.¹⁶ In this technique, end-sequences are anchored to the public genome assembly. The distance between the two ends can then be calculated, giving an observed or computed size of the clone. Because the approximate physical size of the clone is known (in the case of fosmids, the size is typically ~40 kb), large deviations between the computed and physical sizes can represent variations between the two genomes. Although some large deletions can be identified, this approach generally does not readily allow the detection of copy number increases larger than 40 kb. On top of copy number variation, this method is capable of

detecting some inversions by looking for end-sequences that have an incorrect orientation with respect to the public assembly. Along with genome assembly sequence comparisons, this represents the only published method (to the authors' knowledge) for genome-wide investigation of inversions. There is currently a National Institutes of Health-led large-scale genome initiative to end-sequence fosmid clones from numerous libraries prepared from different HapMap samples aimed at discovering structural variation in the human genome.

Following the example of scanning for segmental duplications,³³ the third approach to assessing sequence-read depth will probably become more relevant when whole-genome shotgun sequencing of multiple genomes becomes standard practice.⁹² The rationale behind this technique is that regions of the assembly with greater read depth may be present in multiple copies and, in certain instances, the copy number of the region may vary between individuals.

In all three of these sequence analysis strategies, the accuracy of the data will only be as good as the quality of the assemblies available and confirmation using a secondary method (eg quantitative methods listed above) will probably be required. It is also noteworthy that the study of SNPs located in LCRs,⁹³ some being paralogous sequence variants,⁵² can also be used to detect simple and complex copy number differences arising from duplication, deletion or gene conversion.

Clinical diagnostic implications

The ultimate aim of genetic diagnostics is to evaluate the genomic content of a cell or group of cells as completely and accurately as possible. The objective is to provide either diagnostic insight into the phenotype of a patient or, alternatively, to provide predictive or prognostic insight into the patient's disease or developmental outcome. The upshot is that there have always been some implicit assumptions as to what constitutes a normal genome and a concomitant normal phenotype and hence, by extension, what would constitute an abnormal genome and a deleterious phenotype. For example, the location of the *abl* gene and *BCR* locus (either by molecular genetic or cytogenetic means) on chromosomes 9q34 and 22q11.2, respectively, is considered normal, whereas their co-localisation on a Philadelphia chromosome is considered abnormal and suggestive of chronic myelogenous leukaemia. Similarly, the presence of two copies of chromosome 21 in a metaphase preparation is considered consistent with a normal karyotype, whereas the presence of a third copy would be consistent with Down syndrome.

Whereas the majority of approved diagnostic tests query specific genomic loci of confirmed pathogenicity, such as in the previously mentioned examples, others assess loci only suspected to be associated with disease (eg some sub-telomeric

deletions in mental retardation).^{94–100} In the latter scenario, if a genomic variant is found at the locus in question in a proband, current convention assumes that such a variation is pathogenic if: (i) there is an obvious accompanying phenotypic abnormality in the proband and (ii) the variant is absent in both parents. If one of these parameters is untrue, however, the diagnostician faces a dilemma as to the potential pathogenicity of the genomic aberration.¹⁰¹ The discovery of the phenomena of CNVs, particularly in the context of whole-genome array-based analyses, will challenge conventional understanding of the genome and necessitate a cautionary note regarding assumptions about phenotype/genotype associations. Many questions will emerge, such as: is the CNV detected in a proband truly the causative genomic aberration, or is it a benign CNV which deflected our attention from the more subtle malignant genomic aberration?' 'Is that genomic variant serendipitously detected in a normal healthy individual during the course of an unrelated genetic test cause for concern? Is it a prognosticator of a late-onset disease or, again, simply a benign structural variant?' 'Is that copy number polymorphism present in greater than 1 per cent of the population truly benign, or does it, in fact, have a clinical utility in identifying disease susceptibility?'

In the authors' estimation, the major factors influencing which of the current methods will be used broadly for the detection of CNVs in a clinical diagnostic (or basic research) setting will depend on many factors including accuracy, specificity, set-up time, assay cost, the extent of a genome required to be assayed and requirements for sample input. Broad-based implementation, including those at the regulatory level, will also be influenced by patent restrictions. The flowchart in Figure 2 summarises some factors that need to be considered regarding the technologies and approaches described in this review. To obtain the most comprehensive analysis of both the microscopic and sub-microscopic copy number variation of a genome, it is the authors' belief that the likely paradigm that will prevail will consist of: (i) a karyotype and array-based scan for global assessment of balanced and copy number-type unbalanced variants, respectively, followed by (ii) locus-specific confirmation using targeted FISH or quantitative PCR-based approaches. Obviously, where simple alterations are being tested for in defined phenotypes, such as the dosage-related microdeletion and duplication genomic disorders,^{102,103} only those techniques yielding locus-specific data would be required.

Unresolved issues and recommendations

Currently, there is no single approach that will allow all types of copy number or other structural variants to be identified. This is underscored by the surprisingly small degree of overlap between the published datasets, which in some cases assess

identical samples.²⁸ Short of comparing 'finished' sequence assemblies generated from unique donor (and preferably haploid) sources,²⁶ a multitude of approaches will be required to generate and validate variants in a comprehensive manner. There is also the possibility that new technologies, such as those that utilise direct counting and/or sequencing of single DNA molecules, will provide superior resolution of CNVs.^{92,104} Not only would these analyses be high resolution and high throughput but they could be highly informative, in that they would provide more accurate copy numbers, as opposed to the current techniques, which only annotate CNVs as gains or losses. Notwithstanding, the current repertoire of technologies has facilitated numerous significant studies of chromosome structure,¹⁰⁵ disease^{64,100} and clinical diagnostics.^{98,106,107}

For a newcomer to this field, the authors recommend reading papers relevant to the biological question at hand and then assessing which approach(es) is/are most conducive to success for that project. Even the best-funded laboratories will not be able to establish and maintain the complete range of technologies, so selection of one or a few platforms will need to be made. In a way, the copy number and structural variation field is currently in the same state of flux as the SNP genotyping field was in the late 1990s to early 2000s. A general trend in technologies seems to be moving towards the development of long (> 60-mers) oligonucleotide arrays containing very high probe coverage uniformly distributed across the genome (eg Agilent, Nimblegen platforms and others). It will be very interesting to see how the market dictates the success and evolution of this type of platform versus SNP-based approaches (eg Affymetrix and Illumina), which also allow extraction of copy number information but have the added information of genotypes. There will also be progress in identifying surrogate SNPs^{19,70} or other markers for the more complex structural variants, allowing typing in large sample collections. In some cases, the technology used will be dictated by what equipment already exists in local (core) laboratories. The authors consider this first review as providing a snapshot of what is happening in the field and foresee that a second edition summary in a year or so would look entirely different, due to rapid advances in the field.

When investing in the technologies described in this paper, there are considerations that extend outside those of most typical laboratory management decisions. For example, beyond the rather straightforward choice of targeted or genome-wide experiments (Figure 2) and then selection of the corresponding reagent (eg types of primers or microarrays), one must consider the training level of technicians, specialised equipment (eg hybridisation ovens, scanners), and extended warranty and service support for some expensive equipment that may require constant upgrading. Moreover, in the authors' experience, the weakest links in the current state of the technologies are the algorithms available for mining accurate structural variation data. Investments in this area are

growing, such that interpretation of any data generated now will only improve.

Many other practical issues are worth discussing, including:

- (1) DNA must be of very high-quality preparation and, for genetic studies, preferably from peripheral blood (lymphoblastoid lines can introduce some transformation-derived alterations). Low-quality DNA will lead to major problems, whereas whole-genome amplified DNA seems to be suitable;¹⁰⁸⁻¹¹⁰ however, this assertion will require more analysis to confirm that copy number ratios are maintained.
- (2) At present, for comparative array CGH approaches, there is no standard hybridisation control DNA sample that has been adopted (although some experiments compare an 'average' genome by using pooled DNA samples). Differences between control DNA can cause various problems in correlating results between projects. In a positive step, at a recent Wellcome Trust-sponsored meeting, a recommendation was made to contact the Genome Standards Department of the British Standards Institute regarding appropriate reference standards for CGH.
- (3) There are issues surrounding the content and quality of data in control databases. Currently, most structural variation data are available from two databases, namely the Database of Structural Variants (<http://projects.tcag.ca/variation/>)¹³ and the Human Structural Variation Database (<http://humanparalogy.gs.washington.edu/structuralvariation/>);¹⁵ however, neither of these datasets is currently displayed by the major genome browsers (NCBI, UCSC). The number of samples and entries in these databases are still rudimentary but are expected to grow substantially in 2006. The quality of the data in the databases is fairly heterogeneous, with only a small proportion of variants being validated using a second technology. This will become even more of an issue as large genotyping datasets become a significant source for discovery of new structural variants.⁶⁴
- (4) There are still limitations to all of the technologies. Most approaches readily allow resolution of a single gain or loss in copy number away from diploidy. With more complex deviations, however, where the normal copy number may be five or six, the interpretation becomes increasingly imprecise.
- (5) The accuracy of mapping the breakpoints of a structural variant can be fairly wide-ranging, due to limitations inherent in the technique or the density and distribution of probes on an array. Moreover, because many of the breakpoint regions can overlap segmental duplications or gaps in the human sequence assembly, more detailed targeted analysis is often required.

This review focuses on methodologies for the identification of CNVs in the human genome. If the goal is explicitly to fine-map the precise content and boundaries of the variable

regions, however, additional experiments are required. These include the elucidation of the DNA sequence content in each copy of a region. Due to the nature of their design, certain approaches are more amenable to this type of analysis. For example, in the fosmid end-sequencing strategy, fosmid clones overlapping the variable interval are automatically available for sequencing experiments. Also, some of these duplicated sequences may already be present in the human genome raw sequences (as seen in the human sequence read-depth analysis³³), but are collapsed into a single copy in the current genome build. For some variations, however, such as duplications >100 kb, this process could be more tedious and possibly require extensive library preparation and/or screening, mapping and re-sequencing.

The field of structural variation has now taken centre stage in human genetics and genomics research. The strengths and weaknesses of the different experimental and technical approaches discussed in this review, and also new ones, will be borne out through thorough scientific investigation. As more studies are undertaken, in particular through the investigation of the association between structural variants and disease,^{111–113} a greater understanding of the nature of the human genome and its variability in the population will be achieved. As in many other genomic studies, advances in the analysis of CNVs and structural variants in the human genome will set the precedent for the study of genomes in other species.

Acknowledgments

Research is supported by funds from The Centre for Applied Genomics, Hospital for Sick Children, Genome Canada/Ontario Genomics Institute, the Canadian Institutes of Health Research (CIHR) and the McLaughlin Centre for Molecular Medicine. L.F. is supported by the Swedish Medical Research Council, A.R.C. is supported through a scholarship from the Natural Science and Engineering Research Council and S.W.S. is an Investigator at CIHR and International Scholar of the Howard Hughes Medical Institute.

References

- Jacobs, P.A., Baikie, A.G., Court Brown, W.M. and Strong, J.A. (1959), 'The somatic chromosomes in mongolism', *Lancet* Vol. 1, p. 710.
- Edwards, J.H., Harnden, D.G., Cameron, A.H. et al. (1960), 'A new trisomic syndrome', *Lancet* Vol. 1, pp. 787–790.
- Patau, K., Smith, D.W., Therman, E. et al. (1960), 'Multiple congenital anomaly caused by an extra autosome', *Lancet* Vol. 1, pp. 790–793.
- Maegenis, R.E., Donlon, T.A. and Wyandt, H.E. (1978), 'Giemsa-11 staining of chromosome 1: A newly described heteromorphism', *Science* Vol. 202, pp. 64–65.
- Lubs, H.A. (1969), 'A marker X chromosome', *Am. J. Hum. Genet.* Vol. 21, pp. 231–244.
- Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991), 'DNA typing and genetic mapping with trimeric and tetrameric tandem repeats', *Am. J. Hum. Genet.* Vol. 49, pp. 746–756.
- Cargill, M., Altshuler, D., Ireland, J. et al. (1999), 'Characterization of single-nucleotide polymorphisms in coding regions of human genes', *Nat. Genet.* Vol. 22, pp. 231–238.
- Lander, E.S., Linton, L.M., Birren, B. et al. (2001), 'Initial sequencing and analysis of the human genome', *Nature* Vol. 409, pp. 860–921.
- Mir, K.U. and Southern, E.M. (2000), 'Sequence variation in genes and genomic DNA: Methods for large-scale analysis', *Annu. Rev. Genomics Hum. Genet.* Vol. 1, pp. 329–360.
- Syvanen, A.C. (2001), 'Accessing genetic variation: Genotyping single nucleotide polymorphisms', *Nat. Rev. Genet.* Vol. 2, pp. 930–942.
- Kwok, P.Y. and Gu, Z. (1999), 'Single nucleotide polymorphism libraries: Why and how are we building them?', *Mol. Med. Today* Vol. 5, pp. 538–543.
- Kwok, P.Y. and Chen, X. (2003), 'Detection of single nucleotide polymorphisms', *Curr. Issues Mol. Biol.* Vol. 5, pp. 43–60.
- Iafate, A.J., Feuk, L., Rivera, M.N. et al. (2004), 'Detection of large-scale variation in the human genome', *Nat. Genet.* Vol. 36, pp. 949–951.
- Sebat, J., Lakshmi, B., Troge, J. et al. (2004), 'Large-scale copy number polymorphism in the human genome', *Science* Vol. 305, pp. 525–528.
- Sharp, A.J., Locke, D.P., McGrath, S.D. et al. (2005), 'Segmental duplications and copy-number variation in the human genome', *Am. J. Hum. Genet.* Vol. 77, pp. 78–88.
- Tuzun, E., Sharp, A.J., Bailey, J.A. et al. (2005), 'Fine-scale structural variation of the human genome', *Nat. Genet.* Vol. 37, pp. 727–732.
- Feuk, L., MacDonald, J., Tang, T. et al. (2005), 'Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies', *PLoS Genet.* Vol. 1, p. e56.
- Conrad, D.F., Andrews, T.D., Carter, N.P. et al. (2006), 'A high-resolution survey of deletion polymorphism in the human genome', *Nat. Genet.* Vol. 38, pp. 75–81.
- Hinds, D.A., Kloek, A.P., Jen, M. et al. (2006), 'Common deletions and SNPs are in linkage disequilibrium in the human genome', *Nat. Genet.* Vol. 38, pp. 82–85.
- McCarroll, S.A., Hadnott, T.N., Perry, G.H. et al. (2006), 'Common deletion polymorphisms in the human genome', *Nat. Genet.* Vol. 38, pp. 86–92.
- Carter, N.P. (2004), 'As normal as normal can be?', *Nat. Genet.* Vol. 36, pp. 931–932.
- Buckley, P.G., Mantripragada, K.K., Piotrowski, A. et al. (2005), 'Copy-number polymorphisms: Mining the tip of an iceberg', *Trends Genet.* Vol. 21, pp. 315–317.
- Check, E. (2005), 'Human genome: Patchwork people', *Nature* Vol. 437, pp. 1084–1086.
- Lee, C. (2005), 'Vive la difference!', *Nat. Genet.* Vol. 37, pp. 660–661.
- van Ommen, G.J. (2005), 'Frequency of new copy number variation in humans', *Nat. Genet.* Vol. 37, pp. 333–334.
- Feuk, L., Carson, A.R. and Scherer, S.W. (2006), 'Structural variation in the human genome', *Nat. Rev. Genet.* Vol. 7, pp. 85–97.
- Nadeau, J.H. and Lee, C. (2006), 'Genetics: Copies count', *Nature* Vol. 439, pp. 798–799.
- Eichler, E.E. (2006), 'Widening the spectrum of human genetic variation', *Nat. Genet.* Vol. 38, pp. 9–11.
- Freeman, J.L., Perry, G.H., Feuk, L. et al., 'Copy number variation: New insights in genome diversity', *Genome Res.*, In Press.
- Speicher, M.R. and Carter, N.P. (2005), 'The new cytogenetics: Blurring the boundaries with molecular biology', *Nat. Rev. Genet.* Vol. 6, pp. 782–792.
- Lucito, R., Healy, J., Alexander, J. et al. (2003), 'Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation', *Genome Res.* Vol. 13, pp. 2291–2305.
- Bignell, G.R., Huang, J., Greshock, J. et al. (2004), 'High-resolution analysis of DNA copy number using oligonucleotide microarrays', *Genome Res.* Vol. 14, pp. 287–295.
- Bailey, J.A., Gu, Z., Clark, R.A. et al. (2002), 'Recent segmental duplications in the human genome', *Science* Vol. 297, pp. 1003–1007.
- Kallioniemi, A., Kallioniemi, O.P., Sudar, D. et al. (1992), 'Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors', *Science* Vol. 258, pp. 818–821.

35. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S. *et al.* (1997), 'Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances', *Genes Chromosomes Cancer* Vol. 20, pp. 399–407.
36. Pollack, J.R., Perou, C.M., Alizadeh, A.A. *et al.* (1999), 'Genome-wide analysis of DNA copy-number changes using cDNA microarrays', *Nat. Genet.* Vol. 23, pp. 41–46.
37. Locke, D.P., Segraves, R., Nicholls, R.D. *et al.* (2004), 'BAC microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications', *J. Med. Genet.* Vol. 41, pp. 175–182.
38. Ishkanian, A.S., Malloff, C.A., Watson, S.K. *et al.* (2004), 'A tiling resolution DNA microarray with complete coverage of the human genome', *Nat. Genet.* Vol. 36, pp. 299–303.
39. de Vries, B.B., Pfundt, R., Leisink, M. *et al.* (2005), 'Diagnostic genome profiling in mental retardation', *Am. J. Hum. Genet.* Vol. 77, pp. 606–616.
40. Paulsson, K., Heidenblad, M., Strombeck, B. *et al.* (2006), 'High-resolution genome-wide array-based comparative genome hybridization reveals cryptic chromosome changes in AML and MDS cases with trisomy 8 as the sole cytogenetic aberration', *Leukemia* Vol. 20, pp. 840–846.
41. Garnis, C., Lockwood, W.W., Vucic, E. *et al.* (2006), 'High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH', *Int. J. Cancer* Vol. 118, pp. 1556–1564.
42. Shadeo, A. and Lam, W.L. (2006), 'Comprehensive copy number profiles of breast cancer cell model genomes', *Breast Cancer Res.* Vol. 8, p. R9.
43. Miyake, N., Shimokawa, O., Harada, N. *et al.* (2006), 'BAC array CGH reveals genomic aberrations in idiopathic mental retardation', *Am. J. Med. Genet. A.* Vol. 140, pp. 205–211.
44. Dhami, P., Coffey, A.J., Abbs, S. *et al.* (2005), 'Exon array CGH: Detection of copy-number changes at the resolution of individual exons in the human genome', *Am. J. Hum. Genet.* Vol. 76, pp. 750–762.
45. Mantripragada, K.K., Tapia-Paez, I., Blennow, E. *et al.* (2004), 'DNA copy-number analysis of the 22q11 deletion-syndrome region using array-CGH with genomic and PCR-based targets', *Int. J. Mol. Med.* Vol. 13, pp. 273–279.
46. Barrett, M.T., Scheffer, A., Ben-Dor, A. *et al.* (2004), 'Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA', *Proc. Natl. Acad. Sci. USA* Vol. 101, pp. 17765–17770.
47. Brennan, C., Zhang, Y., Leo, C. *et al.* (2004), 'High-resolution global profiling of genomic alterations with long oligonucleotide microarray', *Cancer Res.* Vol. 64, pp. 4744–4748.
48. Urban, A.E., Korb, J.O., Selzer, R. *et al.* (2006), 'High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays', *Proc. Natl. Acad. Sci. USA* Vol. 103, pp. 4534–4539.
49. van den Ijssel, P., Tijssen, M., Chin, S.F. *et al.* (2005), 'Human and mouse oligonucleotide-based array CGH', *Nucleic Acids Res.* Vol. 33, p. e192.
50. Kennedy, G.C., Matsuzaki, H., Dong, S. *et al.* (2003), 'Large-scale genotyping of complex DNA', *Nat. Biotechnol.* Vol. 21, pp. 1233–1237.
51. Cheung, J., Estivill, X., Khajia, R. *et al.* (2003), 'Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence', *Genome Biol.* Vol. 4, p. R25.
52. Fredman, D., White, S.J., Potter, S. *et al.* (2004), 'Complex SNP-related sequence variation in segmental genome duplications', *Nat. Genet.* Vol. 36, pp. 861–866.
53. Wang, Y., Barbacioru, C., Hyland, F. *et al.* (2006), 'Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays', *BMC Genomics* Vol. 7, p. 59.
54. de Reynies, A., Geromin, D., Cayuela, J.M. *et al.* (2006), 'Comparison of the latest commercial short and long oligonucleotide microarray technologies', *BMC Genomics* Vol. 7, p. 51.
55. Selzer, R.R., Richmond, T.A., Pofahl, N.J. *et al.* (2005), 'Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH', *Genes Chromosomes Cancer* Vol. 44, pp. 305–319.
56. Zhao, X., Richmond, T.A., Pofahl, N.J. *et al.* (2004), 'An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays', *Cancer Res.* Vol. 64, pp. 3060–3071.
57. Huang, J., Wei, W., Zhang, J. *et al.* (2004), 'Whole genome DNA copy number changes identified by high density oligonucleotide arrays', *Hum. Genomics* Vol. 1, pp. 287–299.
58. Wong, K.K., Tsang, Y.T., Shen, J. *et al.* (2004), 'Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA', *Nucleic Acids Res.* Vol. 32, p. e69.
59. Shen, R., Fan, J.B., Campbell, D. *et al.* (2005), 'High-throughput SNP genotyping on universal bead arrays', *Mutat. Res.* Vol. 573, pp. 70–82.
60. Arinami, T., Ohtsuki, T., Ishiguro, H. *et al.* (2005), 'Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p', *Am. J. Hum. Genet.* Vol. 77, pp. 937–944.
61. Zeggini, E., Rayner, W., Morris, A.P. *et al.* (2005), 'An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets', *Nat. Genet.* Vol. 37, pp. 1320–1322.
62. Steemers, F.J., Change, W., Lee, G. *et al.* (2006), 'Whole-genome genotyping with the single-base extension assay', *Nat. Methods* Vol. 3, pp. 31–33.
63. Li, C., Wong, W.H. (2003), 'DNA-Chip Analyzer (dChip)', in Parmigiani, G., *et al.* Eds., *The Analysis of Gene Expression Data: Methods and Software*, Springer, New York, NY.
64. Feuk, L., Marshall, C.R., Wintle, R.F. and Scherer, S.W. (2006), 'Structural variants: Changing the landscape of chromosomes and design of disease studies', *Hum. Mol. Genet.* Vol. 15 (Suppl. 1), pp. R57–R66.
65. Lindblad-Toh, K., Tanenbaum, D.M., Daly, M.J. *et al.* (2000), 'Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays', *Nat. Biotechnol.* Vol. 18, pp. 1001–1005.
66. Mei, R., Galipeau, P.C., Prass, C. *et al.* (2000), 'Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays', *Genome Res.* Vol. 10, pp. 1126–1137.
67. Bruce, S.K., Leinonen, R., Lindgren, C.M. *et al.* (2005), 'Global analysis of uniparental disomy using high-density genotyping arrays', *J. Med. Genet.* Vol. 42, pp. 847–851.
68. Altug-Teber, O., Dufke, A., Poths, S. *et al.* (2005), 'A rapid microarray based whole genome analysis for detection of uniparental disomy', *Hum. Mutat.* Vol. 26, pp. 153–159.
69. Yu, C.E., Dawson, G., Munson, J. *et al.* (2002), 'Presence of large deletions in kindreds with autism', *Am. J. Hum. Genet.* Vol. 71, pp. 100–115.
70. Newman, T.L., Rieder, M.J., Morrison, V.A. *et al.* (2006), 'High-throughput genotyping of intermediate-size structural variation', *Hum. Mol. Genet.* Vol. 15, pp. 1159–1167.
71. Heid, C.A., Stevens, J., Livak, K.J. and Williams, P.M. (1996), 'Real time quantitative PCR', *Genome Res.* Vol. 6, pp. 986–994.
72. Bieche, I., Olivi, M., Champeme, M.H. *et al.* (1998), 'Novel approach to quantitative polymerase chain reaction using real-time detection: Application to the detection of gene amplification in breast cancer', *Int. J. Cancer* Vol. 78, pp. 661–666.
73. Ponchel, F., Toomes, C., Bransfield, K. *et al.* (2003), 'Real-time PCR based on SYBR-green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions', *BMC Biotechnol.* Vol. 3, p. 18.
74. Charbonnier, F., Raux, G., Wang, Q. *et al.* (2000), 'Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using multiplex polymerase chain reaction of short fluorescent fragments', *Cancer Res.* Vol. 60, pp. 2760–2763.
75. Vaurs-Barriere, C., Bonnet-Dupeyron, M.N., Combes, P. *et al.* (2006), 'Golli-MBP copy number analysis by FISH, QMPSF and MAPH in 195 patients with hypomyelinating leukodystrophies', *Ann. Hum. Genet.* Vol. 70, pp. 66–77.
76. Casilli, F., Di Rocco, Z.C., Gad, S. *et al.* (2002), 'Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments', *Hum. Mutat.* Vol. 20, pp. 218–226.
77. Armour, J.A., Sismani, C., Patsalis, P.C. and Cross, G. (2000), 'Measurement of locus copy number by hybridisation with amplifiable probes', *Nucleic Acids Res.* Vol. 28, pp. 605–609.

78. Hollox, E.J., Atia, T., Cross, G. *et al.* (2002), 'High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH)', *J. Med. Genet.* Vol. 39, pp. 790–795.
79. Hollox, E.J., Akrami, S.M. and Armour, J.A. (2002), 'DNA copy number analysis by MAPH: Molecular diagnostic applications', *Expert Rev. Mol. Diagn.* Vol. 2, pp. 370–378.
80. Schouten, J.P., McElgunn, C.J., Waaijen, R. *et al.* (2002), 'Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification', *Nucleic Acids Res.* Vol. 30, p. e57.
81. Langerak, P., Nygren, A.O., Schouten, J.P. and Jacobs, H. (2005), 'Rapid and quantitative detection of homologous and non-homologous recombination events using three oligonucleotide MLPA', *Nucleic Acids Res.* Vol. 33, p. e188.
82. Lalic, T., Vossen, R.H., Coffa, J. *et al.* (2005), 'Deletion and duplication screening in the DMD gene using MLPA', *Eur. J. Hum. Genet.* Vol. 13, pp. 1231–1234.
83. Armour, J.A., Rad, I.A., Hollox, E.J. *et al.* (2004), 'Gene dosage analysis by multiplex amplifiable probe hybridization', *Methods Mol. Med.* Vol. 92, pp. 125–139.
84. White, S., Kalf, M., Liu, Q. *et al.* (2002), 'Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization', *Am. J. Hum. Genet.* Vol. 71, pp. 365–374.
85. White, S.J., Vink, G.R., Kriek, M. *et al.* (2004), 'Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses', *Hum. Mutat.* Vol. 24, pp. 86–92.
86. Stern, R.F., Roberts, R.G., Mann, K. *et al.* (2004), 'Multiplex ligation-dependent probe amplification using a completely synthetic probe set', *Biotechniques* Vol. 37, pp. 399–405.
87. Scherer, S.W., Cheung, J., MacDonald, J.R. *et al.* (2003), 'Human chromosome 7: DNA sequence and biology', *Science* Vol. 300, pp. 767–772.
88. Scherer, S.W. and Green, E.D. (2004), 'Human chromosome 7 circa 2004: A model for structural and functional studies of the human genome', *Hum. Mol. Genet.* Vol. 13 (Spec. No. 2), pp. R303–R313.
89. Hillier, L.W., Fulton, R.S., Fulton, L.A.A. *et al.* (2003), 'The DNA sequence of human chromosome 7', *Nature* Vol. 424, pp. 157–164.
90. Beck, S. and Trowsdale, J. (2000), 'The human major histocompatibility complex: Lessons from the DNA sequence', *Annu. Rev. Genomics Hum. Genet.* Vol. 1, pp. 117–137.
91. Newman, T.L., Tuzun, E., Morrison, V.A. *et al.* (2005), 'A genome-wide survey of structural variation between human and chimpanzee', *Genome Res.* Vol. 15, pp. 1344–1356.
92. Shendure, J., Mitra, R.D., Varma, C. and Church, G.M. (2004), 'Advanced sequencing technologies: Methods and goals', *Nat. Rev. Genet.* Vol. 5, pp. 335–344.
93. Estivill, X., Cheung, J., Pujana, M.A. *et al.* (2002), 'Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome', *Hum. Mol. Genet.* Vol. 11, pp. 1987–1995.
94. Kok, K., Dijkhuizen, T., Swart, Y.E. *et al.* (2005), 'Application of a comprehensive subtelomere array in clinical diagnosis of mental retardation', *Eur. J. Med. Genet.* Vol. 48, pp. 250–262.
95. Rooms, L., Reyniers, E. and Kooy, R.F. (2005), 'Subtelomeric rearrangements in the mentally retarded: A comparison of detection methods', *Hum. Mutat.* Vol. 25, pp. 513–524.
96. Buckley, P.G., Mantripragada, K.K., Benetkiewicz, M. *et al.* (2002), 'A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications', *Hum. Mol. Genet.* Vol. 11, pp. 3221–3229.
97. Albertson, D.G. and Pinkel, D. (2003), 'Genomic microarrays in human genetic disease and cancer', *Hum. Mol. Genet.* Vol. 12 (Spec. No. 2), pp. R145–R152.
98. Rickman, L., Fiegler, H., Shaw-Smith, C. *et al.* (2006), 'Prenatal detection of unbalanced chromosomal rearrangements by array-CGH', *J. Med. Genet.* Vol. 43, pp. 353–361.
99. Aitman, T.J., Dong, R., Vyse, T.J. *et al.* (2006), 'Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans', *Nature* Vol. 439, pp. 851–855.
100. Vissers, L.E., Veltman, J.A., van Kessel, A.G. *et al.* (2005), 'Identification of disease genes by whole genome CGH arrays', *Hum. Mol. Genet.* Vol. 14 (Spec. No. 2), pp. R215–R223.
101. Rosenberg, C., Knijnenburg, J., Bakker, E. *et al.* (2006), '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.* Vol. 43, pp. 180–186.
102. Shaikh, T.H., Kurahashi, H., Saitta, S.C. *et al.* (2000), 'Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: Genomic organization and deletion endpoint analysis', *Hum. Mol. Genet.* Vol. 9, pp. 489–501.
103. Chen, K.S., Manian, P., Koeuth, T. *et al.* (1997), 'Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome', *Nat. Genet.* Vol. 17, pp. 154–163.
104. Service, R.F. (2006), 'Gene sequencing. The race for the \$1000 genome', *Science* Vol. 311, pp. 1544–1546.
105. Woodfine, K., Carter, N.P., Dunham, I. and Fiegler, H. (2005), 'Investigating chromosome organization with genomic microarrays', *Chromosome Res.* Vol. 13, pp. 249–257.
106. Wong, A., Lese Martin, C., Heretis, K. *et al.* (2005), 'Detection and calibration of microdeletions and microduplications by array-based comparative genomic hybridization and its applicability to clinical genetic testing', *Genet. Med.* Vol. 7, pp. 264–271.
107. Ylstra, B., van den Ijssel, P., Carvalho, B. *et al.* (2006), 'BAC to the future! or oligonucleotides: A perspective for micro array comparative genomic hybridization (array CGH)', *Nucleic Acids Res.* Vol. 34, pp. 445–450.
108. Ghazani, A.A., Arneson, N.C., Warren, K. *et al.* (2006), 'Limited tissue fixation times and whole genomic amplification do not impact array CGH profiles', *J. Clin. Pathol.* Vol. 59, pp. 311–315.
109. Little, S.E., Vuononvirta, R., Reis-Filho, J.S. *et al.* (2006), 'Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA', *Genomics* Vol. 87, pp. 298–306.
110. Wang, G., Maher, E., Brennan, C. *et al.* (2004), 'DNA amplification method tolerant to sample degradation', *Genome Res.* Vol. 14, pp. 2357–2366.
111. Le Caignec, C., Boceno, M., Saugier-veber, P. *et al.* (2005), 'Detection of genomic imbalances by array based comparative genomic hybridisation in fetuses with multiple malformations', *J. Med. Genet.* Vol. 42, pp. 121–128.
112. Jobanputra, V., Sebat, J., Troge, J. *et al.* (2005), 'Application of ROMA (representational oligonucleotide microarray analysis) to patients with cytogenetic rearrangements', *Genet. Med.* Vol. 7, pp. 111–118.
113. Bejjani, B.A., Saleki, R., Ballif, B.C. *et al.* (2005), 'Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: Is less more?', *Am. J. Med. Genet. A.* Vol. 134, pp. 259–267.