Detection of Genomic Duplications and Deletions

Graham R. Taylor1 and Carol A. Delaney2

1Regional Genetics Laboratory and CR-UK Mutation Detection Facility, St James’s University Hospital, DNA Laboratory, Leeds, Yorkshire LS9 7TF, UK;
2SE Scotland Cytogenetics Laboratory, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

12.1 INTRODUCTION

Gene and chromosome duplications have been implicated as fundamental evolutionary mechanisms (Ohno et al., 1968; Li and Gojobori, 1983; Schughart et al., 1989). For the purpose of this chapter, duplications and deletions are defined as those too large to be routinely detected by sequencing or simple PCR-based methods (>100 bp), but too small to be detected by conventional cytogenetic metaphase analysis (<106 kb). Unlike other types of rearrangement, for example LINE insertions or balanced translocations, a duplication or deletion will be associated with a direct change in gene dosage. This four-order size range includes small fragments of genes, exons, entire genes, and multiple genes, and the same methods can, of course, be used to measure changes in chromosome count.

Historically, supernumerary chromosomes are the earliest “gene” duplications reported, revealed by cytogenetic techniques (Lejeune, 1960; Patau et al., 1961). Soon afterward, evidence for pathogenic α-globin gene deletions was reported (Ose and Bush, 1962), although more than 10 years elapsed before direct confirmation was achieved (Ottolenghi et al., 1974). β-Globin gene deletions were reported soon afterward (Kan et al., 1975), followed by gene deletions in other hemoglobinopathies and culminating in the use of Southern blotting for the prenatal diagnosis of globin deletions (Orkin et al., 1978). Gene dosage changes were next reported in the immunoglobulin genes (Rabbitts et al., 1980; van Loghem et al., 1980) and the first deletion of a tumor suppressor gene (in retinoblastoma) was reported soon afterward (Juni et al., 1982). The identification of repetitive DNA associated with a gene deletion was described in a form of hereditary persistence of fetal hemoglobin (Jagadeeswaran et al., 1982). A pathogenic role was attributed to repetitive DNA elements by Hess and coworkers (1983), who suggested that DNA insertion elements may disrupt gene correction processes in the two duplication units containing HBA2 and HBA1 genes. Although the widespread use of Southern blotting (Southern, 1975) from the mid-1970s until the late 1980s may have facilitated the detection of deletions and duplications, the application of the polymerase chain reaction (PCR; Mullis et al., 1986) may unintentionally have produced an ascertainment bias away from them, the analysis of dystrophin gene deletions in males being a notable exception (Beggs et al., 1990). This is because a typical PCR is not designed for quantitative analysis, but for optimal purity and yield. However, the observation of germ-line deletions in a wide range of genetic conditions has required the development of techniques that can detect gene dosage changes in hemizygotes. These techniques include Southern blotting and PCR modifications, as well as newer methods. Techniques that measure gene dosage can also be adapted to quantify the somatic mosaicism and PCR failure (allele dropout, see also Chapter 33).

12.2 MECHANISMS

Deletions and duplications can be mediated by homologous recombination involving recombinogenic elements, for example PMP22 (Inoue et al., 2001); or gene duplications, for example the type IV collagen genes COL4A5 and COL4A6, paired head to head on chromosome Xq22 deletions in X-linked Alport syndrome. BRCA1 gene deletions involving a head-to-head of a partial pseudogene have been reported (Brown et al., 2002), as well as Alu-mediated intragenic deletions (Puget et al., 1997; Rohlf et al., 2000). A 26 bp core sequence in two out of five α-thalassemia deletions has been reported (Harteveld et al., 1997), supporting the idea that Alu repeats stimulate recombination events not
only by homologous pairing, but also by providing binding sites for recombinogenic proteins. Deletion of subtelomeric repeats has been implicated in facioscapulohumeral muscular dystrophy (FSHD) where a remnant fragment, the result of a deletion of tandemly arrayed 3.3 kb repeat units (D4Z4) on 4q35 can be detected (Lemmers et al., 1998). Non-homologous recombination also has been implicated in gene deletions (Hu and Worton, 1992; Suminaga et al., 2000).

12.3 PATHOLOGICAL CONSEQUENCES

12.3.1 Supernumerary Chromosomes

In practice, when considering alterations in autosomal chromosome copy number, only supernumerary chromosomes need to be addressed, as autosomal monosomy is not compatible with life (unless the individual has monosomy mosaicism). Monosomy of the X chromosome can occur (Turner syndrome), although it is estimated that only 1 in 300 conceptuses with monosomy X survive (Kajii et al., 1980). Triploidy and trisomies of other autosomes are also a common cause of spontaneous abortion. From a clinical perspective, prenatal detection of triploidy and trisomies 13 and 18 are important as fetuses may survive until birth, but most die soon after due to congenital malformations. Trisomies that have less severe consequences are trisomy 21 (Down syndrome), trisomy X (rarely show any physical abnormalities), and Klinefelter syndrome (XXY). Supernumerary chromosomes are readily detectable by conventional cytogenetics (see also Chapter 10); however, molecular techniques may be faster and less expensive.

12.3.2 Microdeletions

Pathogenic partial deletions and duplications have been widely reported for all chromosomes, and cause a range of symptoms. Some deletions are easily detectable by conventional cytogenetics, others are smaller (microdeletions) and can be difficult or impossible to detect by conventional cytogenetics, and other cytogenetic techniques such as FISH or molecular techniques are required. Some of the more common microdeletions with associated syndromes (summarized in Table 12.1) are 22q11.2 (Di George syndrome), 15q11.2/q12 (Prader-Willi and Angelman syndromes), 17p (Miller-Dieker and Smith-Magenis syndromes), 4p16.3 (Wolf-Hirschhorn syndrome), and 5p15 (Cri du chat syndrome).

12.3.3 Subtelomeric Deletions

Several syndromes that are caused by microscopically visible chromosomal deletions and duplications, including the subtelomeric region, have been known to cause idiopathic mental retardation (Flint et al., 1995; Knight et al., 1999) (e.g. 1p, 4p, 5p, 9p). Despite their clinical relevance, screening for more cryptic alterations in the subtelomeres cannot be readily observed by G-banding analysis since most terminal bands are G-band negative. More recent studies using molecular methods such as detection of loss of hypervariable DNA polymorphisms or microsatellite markers and multiprobe FISH that detect submicroscopic subtelomeric deletions have shown that these can account for up to 5% of cases with mental retardation (Flint et al., 1995; Knight et al., 1999; Rio et al., 2002). Subtelomeric deletion analysis of all chromosomes can be performed rapidly and cost effectively using new molecular techniques such as MAPH (Sismani et al., 2001), MLPA (Rooms et al., 2006; Palomares et al., 2006), and array CGH (de Vries et al., 2005).

12.3.4 Gene Deletions or Duplications

A proportion of some types of cancers are caused by inherited germ-line mutations in tumor suppressor genes, and full or partial gene deletions account for a significant number of these mutations. There are a range of other single gene disorders where deletion or duplication of part or all of the gene accounts for a significant proportion of detected mutations, such as Duchenne muscular dystrophy, spinal muscular atrophy, Charcot-Marie-Tooth disease,
Fanconi anemia, congenital adrenal hyperplasia, and rare metabolic disorders such as non-ketotic hyperglycinemia. Some examples of genes for which an estimate has been obtained for the proportion of germ-line mutations that are deletions or duplications are shown in Table 12.1.

Since the majority of mutations identified in the tumor suppressor genes lead to the production of a truncated product (Couch and Weber, 1996), until recently most studies performed mutation analysis on genomic DNA using PCR-based techniques such as sequencing, heteroduplex analysis (Chapter 4), or the protein truncation test (PTT; Chapter 19). Deletions (Petrij-Bosch et al., 1997; Puget et al., 1997; Swensen et al., 1997) or duplications (Puget et al., 1999) within the BRCA1 gene, for example, would have been missed by conventional PCR-based methods and were detected by either reverse transcriptase PCR (RT-PCR) or Southern blotting. Although both PCR and Southern blotting have been adapted to provide quantitative data, PCR has become the method of choice for genetic testing. Estimates of gene dosage typically have been based on comparisons with a reference standard. Other approaches, including the study of junction fragments or microsatellite inheritance and more recently long-range PCR (Coulter-Mackie et al., 1998), FISH (Voskova-Goldman et al., 1997) and array-CGH (Bruder et al., 2001), have also been employed.

The overall contribution of deletions and duplications recorded in the May 2000 Human Gene Mutation database was 5.5% of reported mutations. As of March 2008 the proportion had increased to 22.7%, comprising 16% deletions and 6.5% duplications (http://www.hgmd.cf.ac.uk; Stenson et al., 2003). Given the greater technical difficulties in identifying deletions and duplications, this is still likely to be an underestimate. This emphasizes the importance of including the measurement of gene dosage in any comprehensive mutation scan. Numerous polymorphisms have been described involving deletion or duplication of large chromosomal segments, sometimes involving entire genes; for example, the common deletion polymorphisms of the cytochrome P450 gene CYP2D6 (Meyer and Zanger, 1997), the theta-class glutathione S-transferase gene GSTT1 (Wiencke et al., 1995), and the mu-type gene GSTM1 (Brockmoller et al., 1992).

Regions close to telomeres are especially prone to inter-chromosomal rearrangements that can lead to different forms of presence/absence polymorphism, including the multiallelic variation in the structure of the 16p telomere (Wilkie et al., 1991), and the deletion polymorphism near the 12q telomere (Baird et al., 2000). The assembly of the human genome sequence is complicated by the presence of duplication/deletion polymorphisms, either because only one form of a region of variable structure is recorded in the sequenced chromosome (Siniscalco et al., 2000) or because the presence of unsuspected polymorphism for tandem duplication can lead to misassembly (Bailey et al., 2001, 2002).

12.4 DIAGNOSTIC TECHNIQUES

Classical cytogenetic genetic techniques have been valuable in identifying supernumerary chromosomes and large deletions. However, the resolution of these methods is limited to several megabases and they would miss many known submicroscopic deletions. Fluoresence in situ hybridization (FISH) techniques have greatly expanded the capabilities of cytogenetics (see also Chapter 10); single copy probes allow detection of microdeletions, which would be impossible to detect by conventional cytogenetics; however, due to the specialized equipment and expertise required and the low throughput, detection of submicroscopic deletions has largely become the realm of specialized molecular genetic diagnostics. The detection of submicroscopic duplications by cytogenetics remains a challenge; the use of interphase FISH together with image analysis is being used increasingly.

12.4.1 Southern Blotting

Southern blotting still has a role in a comprehensive diagnostic service, both for gene dosage measurement and for the measurement of allele expansions. Southern blots can detect deletions or duplications by the identification of novel restriction fragments created by the rearrangement or by measurement of band intensity compared to a control fragment. Fragment size changes are seen because the region of genomic DNA created by a deletion or duplication may have gained or lost restriction enzyme sites. In those cases, a Southern blot of a genomic digest using an appropriate restriction enzyme will detect novel fragments. There is a risk that some altered fragment sizes may simply be due to restriction fragment length polymorphisms, although the risk can be reduced by performing separate digestions with a different enzyme. Direct visualization of fragment sizes has long been used for detecting α-globin duplications and deletions (Orkin et al., 1978), and the increased size range available in pulse field gel electrophoresis (PFGE) allows detection of deletions and duplications even in genes as large as the dystrophin gene (Kenrick et al., 1987).

Estimation of gene dosage by measuring the intensity of probe hybridization (usually in comparison with a control) has identified several instances of gene deletions (Bonifas and Epstein, 1990), though few studies report details of the dose–response curve (such as its linear range) and internal controls are often lacking. Dystrophin deletion carrier testing is possible by assessing relative band intensity (van Essen et al., 1997). Accurate measurement of band intensity requires a phosphoimager, which gives a linear dose response over a wider range than film emulsion. Figure 12.1 shows the application of Southern blotting to measure the relative gene dosage of CYP21A and CYP21B, which can be deleted in congenital adrenal hyperplasia. The risk of inaccuracies introduced by uneven transfer of DNA to the membrane or incomplete washing of the probe requires
that an additional probe for a control locus should be included as a standard.

12.4.2 Microsatellites and SNPs

Microsatellites have been used to detect supernumerary chromosomes (Mansfield, 1993) and large deletions or duplications (Brice et al., 1992). This approach is limited by the fact that semi-quantitative PCR methods do not work well using microsatellites, so only cases in which the allele lengths of the microsatellites differ over chromosomal regions of interest can be interpreted unambiguously. Peak heights or areas of microsatellites can be hard to interpret because of preferential amplification of the smaller allele and the stuttering effect that produces a series of minor peaks immediately adjacent to the major peak. The difficulties with multiplex quantitative fluorescent PCR (QF-PCR) are reduced by selection of tetranucleotide repeat markers and is finding increasing use in the rapid diagnosis of common aneuploidies (Donaghue et al., 2003). Loss of heterozygosity of single nucleotide polymorphisms (SNPs) using high-throughput Affymetrix human SNP arrays has been proposed as a rapid means of identifying allele imbalance caused by genomic deletions in tumor cells (Hoque et al., 2003).

12.4.3 Polymerase Chain Reaction

Neubauer and colleagues (1990) described the co-amplification of two amplicons followed by quantitation on the basis of staining intensity as a means of estimating gene dosage. Differential PCR-based methods are semi-quantitative: they determine relative concentrations of two amplicons, but not absolute molar amounts. Taking advantage of the increased sensitivity of fluorescent detection methods Yau and coworkers (1996) described a multiplex fluorescent PCR that was able to detect female carriers of deletions or duplications in the dystrophin gene. Using peak areas, they were able to give statistical estimates of their assay, likely to be essential for diagnostic applications. Similar methods have been used for testing for hereditary motor and sensory neuropathy (HMSN) duplications (Rowland et al., 2001) and APC deletions (Flintoff et al., 2001). Casilli and coworkers (2002) refined the basic multiplex fluorescent PCR method by selecting shorter fragments, tagging the primers with common tags of 16 nucleotides, and using a modified PCR buffer containing DMSO. The method was termed quantitative multiplex PCR of short fluorescent fragments (QMSFPCR) and enables rapid design of amplicons to fine-map the limits of a deletion or duplication (Casilli et al., 2002). Although differential PCR methods offer high orders of multiplicity (e.g. 8–15 amplicons per reaction), end-point multiplex PCR assays rely on the equivalence of amplification of each fragment in the multiplex in the test sample as well as a control sample. This equivalence may be lost if the starting template DNA concentration is too variable, since different fragments within the multiplex may be amplified with different efficiencies. In some situations it may be necessary to choose a more robust, lower order multiplex base on real-time PCR. Real-time PCR (also known as quantitative PCR; qPCR) provides a means for continuous detection and quantification of product throughout the amplification process, and as such can dispense with a gel separation stage and operate in a closed system (see also Chapter 7). The accumulation of PCR product is monitored by staining using interchelating dyes (e.g. SYBR Green) or by dual-labeled probes such as TaqMan (Laurendeau et al., 1999), molecular beacons (Tyagi and Kramer, 1996), or other fluorescent detection systems. Real-time PCR is becoming widely used as a method for measuring gene dosage (Feldkotter et al., 2002; Bertin et al., 2003; Covault et al., 2003; Gaikovitch et al., 2003; Kim et al., 2003). Real-time quantitative PCR has been used for screening large gene and chromosomal rearrangements (Ariani et al., 2004). Additionally, real-time PCR can increase the sensitivity of mutation detection especially in double-copy genes such as X-linked methyl-CpG-binding protein 2 (MECP2) gene, where current mutation scanning techniques such as DGGE, SSCP, DHPLC and direct sequencing are prone to miss gross rearrangements.

12.4.4 MAPH and MLPA

The multiplex amplifiable probe hybridization (MAPH) method for copy-number measurement (Armour et al., 2000) combines hybridization as the primary step to detect copy number with end-point multiplex PCR to amplify the
hybridized probes. Sets of short probes corresponding to the segments to be tested, each flanked by the same primer-binding sites, are hybridized with the test genomic DNA immobilized on a solid support. After washing, each specifically bound probe will be present in an amount proportional to its copy number. All probes can then be amplified simultaneously with a single primer pair, and quantified after electrophoretic separation. This study demonstrated the simultaneous assessment of copy number in a set of 40 human loci, including detection of deletions causing Duchenne muscular dystrophy and Prader-Willi/Angelman syndromes. The high order multiplex achieved may be limited only by the need to have probes of varying size for electrophoretic separation. MAPH probes are generated by cloning the target sequences into a plasmid vector, amplifying the cloned sequence using primers directed to the vector with the result that all probes are then flanked with the same sequence. Probes that are intended to be multiplexed must be of sufficient size difference to be resolved by electrophoresis. The membranes are then washed rigorously to remove unbound probe, and the remaining specifically bound probe will be present in an amount proportional to its target copy number. The probes are then stripped from the membrane by boiling, and amplified simultaneously with the universal primer pair. Products are then separated by electrophoresis, and a relative comparison is made between the peak heights. Reduced peak heights compared to internal control probes indicate a reduction in gene copy number (deletion) and an increase in gene copy number (duplication) produces increased peak heights. The assay can be completed in two to three days, requiring one overnight hybridization followed by membrane washing, a PCR step, and a product detection step. The system works well, but the manipulation of small nylon filters presents some difficulties in sample handling and labeling.

A similar technique called multiplex ligatable probe amplification (MLPA) avoids the use of filters by using single-stranded ligatable probes and a thermo-stable DNA ligase to produce the amplifiable target (Schouten et al., 2002; see also Chapter 13). In the MLPA technique, genomic DNA is hybridized in solution to probe sets, each of which consists of two halves. One half consists of a target-specific sequence (20–30 nucleotides) flanked by a universal primer sequence, and can be generated synthetically. The other half also has a target-specific sequence at one end (25–43 nucleotides) and a universal primer sequence at the other, but has a variable length stuffer fragment in between (19–370 nucleotides) to generate the size differences necessary in the probes to allow electrophoretic resolution. This larger probe part is generated by cloning the target-specific sequence into M13 derived vectors that already contain the variable length fragments; single-stranded DNA is then purified from the phage particles and made double stranded at two sites by annealing short oligonucleotides in order that the desired probe fragment be liberated by restriction enzyme digestion. The two probe halves are designed such that the target-specific sequences bind adjacent to the target DNA, and can then be joined by use of a ligase. This generates a contiguous probe flanked by universal primer binding sites that can then be amplified by PCR, whereas unbound probe halves cannot be amplified, and hence eliminates the need for removal of excess probe by washing. The amounts of ligated probe produced will be proportional to the target copy number, and after PCR amplification the relative peak heights indicate deletion or duplication of target sequence.

In diagnostic use, applications of these high order multiplex methods include the detection of aneuploidies, unbalanced cryptic translocations, and whole or partial gene duplications or deletions. Commercial kits based on MAPH and MLPA are available (http://www.mrc-holland.com). The high multiplicity of MAPH and MLPA, plus their use of standard genetic laboratory apparatus, make it highly likely that these techniques will become widely used. Both techniques offer a rapid means of scanning up to 40 loci for gene dosage, and are likely to be used widely in research and diagnostic settings. MAPH represents a conceptual breakthrough for the analysis of gene dosage, but the handling of small filter discs is difficult and could pose sample tracking problems in routine medium throughput settings. The liquid-phase solution to sample handling offered by MLPA and the ready availability of robust commercial kits has led to rapid acceptance by diagnostic laboratories worldwide. MLPA uses M13 (single-stranded) probes that are more technically challenging to construct than MAPH probes, which can be made from PCR products.

There is an ever-growing range of MLPA probe sets and kits supplied by MRC-Holland. However, it is also possible to construct additional probes to enable fine mapping and identification of the junction fragments/deletion breakpoints by designing your own synthetic MLPA probes. The advantage of synthetic probes, as compared to the phage M13-derived probes made by MRC-Holland, is that the cloning step is omitted, so they are available sooner, but disadvantages are that fewer probes can be used in a single MLPA reaction, approximately 11 bespoke probes can be added to a probemix. Protocols assisting design of synthetic probes are available from the MRC-Holland website or using AlleleID probe design software (http://www.premierbiosoft.com/datafiles/alleleIDWin.exe). Alternatively, additional loci could be examined by adapting MLPA to an array detection setting, so that fragments were identified by sequence rather than size. This would enable all MLPA probes to be chemically synthesized. Array adaptations of both MAPH and MLPA might enable much higher order multiplexes than are currently possible. However, even 40-plex assays represent a substantial gain in the multiplex order typically available in PCR. MAPH or MLPA could provide competition for the array-based CGH approaches: a 96-well array of 40 probes could interrogate over 3,000 loci, representing a better than 1 cM coverage of the entire human genome or much higher single chromosome resolution. Two color MLPA kits are
available increasing the number of loci examined to some extent by including two probe sets which are amplified by two sets of primers: one labeled with FAM and the other with HEX. An example is the MLPA kit to detect deletions and duplications in the CBP (Crebb binding protein) gene. In all, this kit has synthetic probes for 20 out of 32 exons of CBP, and also five control probes for unlinked loci. The probes span the CBP gene from exons 2 to 32 (the coding region of the gene) and, due to the number of probes, is able to detect deletions and duplications in CBP in cases with Rubinstein-Taybi syndrome previously undetectable by FISH (Roelfsema et al., 2005).

12.4.5 Long-range PCR
Long-range PCR (Barnes, 1992) is a modified PCR protocol that includes a proof-reading polymerase and short denaturation times to enable the size of amplicons to be increased from 3–5 kb to beyond 30 kb. This opens the possibility of screening for deletions of this scale by direct PCR, with the advantage of producing the junction fragment for further analysis. Long-range PCR has been used successfully to identify deletions including mitochondrial DNA (Fromenty et al., 1996), C4 gene deletions in the MHC complex (Grant et al., 2000), the CYP2D6 deletion allele, and LDL deletions (Kim et al., 1999). Despite the simplicity in principle of a long-range PCR approach, in practice it is often difficult to design a robust assay. Direct detection of duplications by amplification of entire duplicated regions is likely to prove difficult in diploid genomes since the unduplicated allele will have a significant advantage during the amplification process. However, full characterization of a deletion or duplication requires the sequence of the junction fragments. This can be achieved either by using long-range PCR or by fine mapping, for example using QMSF-PCR followed by conventional PCR and sequencing.

12.4.6 Array-Based CGH
Microarrays have become widely used tools for gene expression studies, and now encouraging developments are taking place in array-based comparative genome hybridization (Array-CGH) as a tool for measuring gene duplications and deletions. Comparative genomic hybridization originally used metaphase chromosomes as targets for differentially labeled probes (e.g. Cy5, Cy3) from control and test samples. Gene dosage changes could be detected by variation of the relative intensity of the two labels (Kallioniemi et al., 1992). By replacing the metaphase spread with microarrayed BAC DNA, Pinkel and colleagues (1998) reported the use of array-based CGH analysis to investigate chromosome 20 gene dosage alterations in breast cancer. DNA purified from BAC clones spaced at approximately 3 Mb intervals along the entire chromosome together with some X chromosome controls were arrayed on nitrocellulose membranes as 200–400 micron spots in duplicate. Test DNA from breast cancer cell-lines was labeled by nick translation with fluorescein, control DNA with Texas Red, and the spots were counterstained with DAPI. Scanning used custom-built mercury-arc illumination and CCD detection. Other studies have used different dyes (with Cy3/Cy5 being a popular choice) and omitted counterstaining. Many detection systems use confocal laser scanners, which, although more restricted in fluorochrome options, are brighter and have fewer problems with light scattering.

Growing high numbers (>104) of cloned DNA are expensive and labor intensive. To circumvent large-scale cultures and DNA purifications, array features have been made using degenerate oligonucleotide PCR (DOP-PCR) of BAC templates using 5’ amine-linked primers (Hodgson et al., 2001). Fiegler and coworkers (2003) improved the standard DOP-PCR to reduce non-specific host and vector amplicons from BAC clones, improving the signal-to-noise ratio. A ligation-mediated PCR BAC labeling method has also been reported that improved signal-to-noise ratios (Snijders et al., 2001). Array CGH is still in development (see also Chapters 10 and 12), but strong commercial and academic interest are likely to result in an increase in research and diagnostic applications in the near future. The supply of robust competitor (Cot1) DNA has been a major problem reported by many array-CGH users (Carter et al., 2002). Cot1 DNA is used to block the hybridization of repetitive DNA in BAC-derived amplicons. Buckley and coworkers (2002) constructed a comprehensive microarray representing a human chromosome for analysis of DNA copy-number variation. The chromosome 22 microarray covered 34.7 Mb with an average resolution of 75 kb using a sequence-defined, repeat-free, and non-redundant strategy for array preparation. This enabled an increase in array resolution and eliminated the need for Cot1 DNA. Array targets were made using phi29 DNA polymerase synthesis.

As the technology becomes more widely used, commercial suppliers are starting to emerge; for example, the Cytochip array from BlueGnome (http://www.cytochip.com) is designed for diagnostic use. This is a BAC microarray with an average genome coverage of approximately 565 kb, with median coverage of 250 kb in subtelomeric regions and an average resolution of 100 kb in 90 known clinically relevant regions (microdeletion/duplication syndromes). Higher resolution oligonucleotide-based microarrays for the detection of copy-number changes are becoming increasingly popular allowing genome-wide analysis which can map chromosomal imbalance breakpoints at exon level resolution, including imbalances that are single copy-number genomic alterations (Selzer et al., 2005). Commercial suppliers of oligonucleotide-based arrays include Agilent Technologies Inc. (http://www.agilent.com). Several Agilent array CGH platforms exist including arrays with 44,000 60-mer oligonucleotide probes covering both coding and
non-coding human sequences with an average genome resolution of 50kb through to 244,000 probes with a median probe spacing of 9kb. Other suppliers of high-resolution oligonucleotide-based arrays to detect gene dosage changes include Illumina (http://www.illumina.com) and Affymetrix (http://www.affymetrix.com). Additionally, genome centers, for example, Leiden University in Holland and the Sanger Centre in the UK, are using their resources to manufacture genome- or chromosome-specific arrays.

Since array-CGH enables high-resolution detection of structural variation, this technique is moving from a research tool into diagnostic service provision. Confirmation of the array CGH results are often needed prior to issuing a diagnostic report. Several copy number/dosage techniques confirm array CGH results, including FISH, MLPA, STR analysis, real-time PCR/qPCR, long-range PCR, and, ultimately, DNA sequencing. The choice of confirmation method will depend upon several factors, including the size of the region involved, the precise location of the abnormality, availability of FISH or MLPA probes, the density of STRs, the material available from the proband and the parents (DNA, metaphase spreads, etc.), and access to equipment. Often there is an interplay between several techniques to determine copy-number variation per case (Fig. 12.2).

One of the findings of using higher-resolution array CGH for genomic-wide copy-number changes is the awareness of the significant degree of natural copy-number variation in the genome. At 100kb array CGH resolution Redon and coworkers (2006) reported an average of 70 polymorphisms per patient and a total of 1,116 unique polymorphisms on over 270 patients. Similarly, de Vries and coworkers (2005) reported 4–9 copy-number variants per patient in a cohort of mental retardation patients, where 96% of these variations were found to be inherited polymorphisms.

12.5 NOMENCLATURE

Recently, the Human Genome Variation Society (http://www.hgvs.org) has taken on the responsibility of establishing a standardized mutation nomenclature that includes deletions and duplications. A nomenclature reference is maintained at this site by Dr. Johan Den Dunnen, Leiden University Medical Centre. The current recommendation for exon or multi-exon deletion nomenclature varies, depending on whether or not the breakpoint has been identified. If the breakpoints are not sequenced (e.g. detected on Southern blot or by MLPA), exonic deletions are described as c.88-?_923\_del, indicating a deletion starting at an unknown position in the intron 5’ of cDNA nucleotide 88 and ending at an unknown position in the intron 3’ of cDNA nucleotide 923. A genomic or cDNA reference (e.g. Genbank, EMBL, DDJB) should be cited, including the version number. If the junction fragments are known, then the form g.390_1458del or (g.390_1458del1069) should be used for a genomic reference sequence or c.13-23_301-143del (c.13-23_301-143del1069) for a cDNA reference sequence.

12.6 GENE DOSAGE APPLICATIONS IN TUMOR PROFILING

Partial chromosomal losses occur relatively frequently in a large number of tumor types, and this is readily detectable by demonstrating loss of heterozygosity (LOH) of polymorphic
microsatellite markers. Such techniques have been widely employed to investigate various aspects of tumor development:

1. Clonality. Loss of chromosomal regions is essentially an irreversible event in tumor cell development, and all subsequent subclones would be expected to demonstrate the same deletions, and may have accumulated more. LOH analysis of multiple tumors from the same patient can be used to determine whether tumors, either synchronous or metachronous, are clonal in origin, and in the latter case may be used to determine whether subsequent tumors are recurrences or metastases of previous tumors or independent primary tumors. For example, in patients with multiple synchronous lung tumors it is important for treatment decisions to discriminate multicentric lung cancers from intrapulmonary metastases, and this can be aided by clonality studies (Shimizu et al., 2000).

2. Identification of tumor suppressor genes. It is hypothesized that the reason that tumors frequently undergo LOH, and that LOH of particular regions is commonly associated with specific tumor types, is that the regions lost harbor tumor suppressor genes. Identification of such tumor suppressor genes may provide useful tools for diagnosis, prognosis, and possible therapies. Hence, numerous studies have been undertaken to identify regions of LOH specific to individual tumor types and the genes present in these regions. For example, chromosome 9 is the most frequently deleted chromosome in transitional cell carcinoma of the bladder, and candidate genes have been identified in three of the four regions of minimal deletion (Knowles, 1999). In prostate cancer, gains at Q1 and 1q are among the most common chromosomal alterations, and amplification of the AR (Xq12), MYC (8q24), and EYS3 (8q23) genes have been found in a large fraction of hormone-refractory prostate cancers (Nupponen and Visakorpi, 2000).

3. Diagnosis. LOH analysis has been employed both in detection of the presence of a tumor, as well as differential diagnosis and typing of tumors. For example, LOH has been reliably demonstrated in DNA extracted from urine sediments in bladder cancer patients, and may have a role in the non-invasive diagnosis of bladder cancer (Linn et al., 1997; Berger et al., 2002). Differential diagnosis between renal oncocytomas and renal cell carcinomas can be difficult due to morphological similarities, but is important due to their different prognoses. LOH analysis has shown that they can be differentiated on the basis of spectrum of chromosomal loss (Herbers et al., 1998). LOH analysis has also been used to characterize the aggressive intraductal carcinoma of the prostate, to differentiate it from the less aggressive high-grade dysplasia (prostatic intraepithelial neoplasia, PIN), and to provide evidence that it does not represent invasion of Gleason grade 3 cancers into the ductal/acinar system (Dawkins et al., 2000).

4. Prognosis. In several tumor types, LOH of specific regions have been hypothesized to be key events in tumor evolution and progression. LOH analysis of tumor specimens has been used to determine their value as prognostic markers. It has been found in colorectal cancers that high level LOH correlated with earlier onset and lymphatic invasion, and hence a poorer prognosis, and low level LOH was more common in earlier stage disease and predicted a more favorable outcome (Choi et al., 2002). In breast cancers, specific chromosomal regions have been identified for which LOH is a significant predictor of lymph-node metastasis and hence may serve as a negative prognostic indicator (Nagahata et al., 2002).

12.7 SUMMARY AND FUTURE DEVELOPMENTS

Gene dosage is a significant contributor to the overall burden of the germ-line and somatic mutations in man. Any comprehensive mutation screen therefore should include measurement of gene dosage. The permeation of genomics into medical practice will increase the demand for mutation screening with applications in diagnosis, predictive testing, and treatment. This, in turn, will encourage the development of more highly automated approaches, the introduction of robust statistical analysis, and quality control of results that may be used in a diagnostic context. Those techniques are most easily adaptable to robust laboratory processes and automated data handling will have a competitive advantage. Ease of handling, hence more robust results, perhaps explains the greater uptake of MLPA compared with MAPH. Both MAPH and MLPA produce data using automated DNA sequencers that are spreadsheet-ready. Statistical parameters may then be easily applied as quality measures (Taylor et al., 2003). The same is true of array scanner output and real-time PCR systems, in contrast to classical genetic techniques like Southern blotting or cytogenetics. Whereas MLPA, real-time PCR, QMSF-PCR, and MAPH approaches are focused around relatively low numbers of targets, the array methods offer very high orders of multiplicity, the so-called “hypothesis-free” approach. It seems likely that the two approaches will converge as arrays become increasingly targeted toward regions or functional clusters of interest and the PCR-based approaches move toward an array-like format, either in capillary arrays or high density microtiter arrays.

Finally, as clonal sequencing technology matures it may displace the current generation of array technology. This reflects advantages in the clonal/digital nature of the output in contrast to the analog output from arrays, the unbiased readout, not dictated by what has been arrayed, and the ability to read 30–400 base strings of DNA sequence rather than infer sequence results from hybridization to arrayed sequence of 20 to 80 bases. Structural variations, including copy-number variants, can be readily detected using paired
reads, although the sensitivity and specificity of this method has not yet been established. Campbell and coworkers (2008) characterized 306 germ-line structural variants and 103 somatic rearrangements to the base-pair level of resolution.

REFERENCES


