metadata, citation and similar papers at core.ac.uk

brought to you by

provided by Archivio istituzionale della ricerca - Università di

doi:10.1093/annonc/mdm251

Analysis of extended genomic rearrangements in oncological research

L. De Lellis[1,2], M. C. Curia[1,2], G. M. Aceto[1,2], S. Toracchio[1,2], G. Colucci[3], A. Russo[4], R. Mariani-Costantini[1,2] & A. Cama[1,2]*

[1]*Department of Oncology and Neurosciences;* [2]*Center of Excellence on Aging, ''G. D'Annunzio'' University Foundation, University ''G. D'Annunzio'', Chieti;* [3]*Division of Medical Oncology, National Institute of Oncology, Bari;* [4]*Department of Surgical and Oncology, Section of Medical Oncology, Universitá di Palermo, Palermo, Italy*

Screening for genomic rearrangements is a fundamental task in the genetic diagnosis of many inherited disorders including cancer-predisposing syndromes. Several methods were developed for analysis of structural genomic abnormalities, some are targeted to the analysis of one or few specific loci, others are designed to scan the whole genome. Locus-specific methods are used when the candidate loci responsible for the specific pathological condition are known. Whole-genome methods are used to discover loci bearing structural abnormalities when the disease-associated locus is unknown. Three main approaches have been employed for the analysis of locus-specific structural changes. The first two are based on probe hybridization and include cytogenetics and DNA blotting. The third approach is based on PCR amplification and includes microsatellite or single nucleotide polymorphism (SNP) genotyping, relative allele quantitation, real-time quantitative PCR, long PCR and multiplex PCR-based methods such as multiplex ligation-dependent probe amplification and the recently developed nonfluorescent multiplex PCR coupled to high-performance liquid chromatography analysis. Whole-genome methods include cytogenetic methods, array-comparative genomic hybridization, SNP array and other sequence-based methods. The goal of the present review is to provide an overview of the main features and advantages and limitations of methods for the screening of structural genomic abnormalities relevant to oncological research.

Key words: copy number, gene dosage, locus-specific, molecular diagnosis, mutation detection, structural variations

introduction

Extended genomic rearrangements are recognized to play a pathogenic role in an increasing number of human genetic diseases and are relatively frequent, representing 7.4% of reported mutations in the January 2007 release of the Human Gene Mutation Database [1]. Therefore, screening for genomic rearrangements is a fundamental task in the genetic diagnosis of many inherited disorders including cancer-predisposing syndromes. In fact, rearrangements were reported to play a relevant role in familial cancer predisposition syndromes, such as hereditary nonpolyposis colorectal cancer [2], familial adenomatous polyposis coli and familial breast and ovarian cancers [3]. The relevance of genomic rearrangement analysis in oncological research is not limited to germline alterations since somatic alterations of genomic structure play a key role in the pathogenesis of cancer. The identification of somatic rearrangements in tumors is extremely important as it may provide novel markers for prediction of response to therapy and for prognosis and may uncover new potential therapeutic targets. Thus, the development of routine methods for the identification of genomic rearrangements is highly needed. A growing number of technologies are becoming available for this purpose. In this review we provide a summary of the main features, advantages and limitations of methods for the screening of structural genomic abnormalities that are employed in oncological research. Among the wide range of methods developed for the analysis of structural genomic abnormalities, some are targeted to the analysis of one or few specific loci, and others are designed to scan the whole genome. Locus-specific methods are used when the candidate loci responsible for the specific pathological condition are known. The advantage of locus-specific methods is their cost-effectiveness and the relatively high resolution achieved. Alternatively, whole-genome methods are used to discover loci bearing structural abnormalities when the disease-associated locus is unknown. These methods are currently expensive and the resolution achieved is generally lower, but both disadvantages are likely to be overcome by evolving technological developments.

*Correspondence to: A. Cama, Nuovo Polo Didattico, Edificio C, Università di Chieti, 66013 Chieti, Italy. E-mail: cama@unich.it

© 2007 European Society for Medical Oncology

locus-specific methods

Three main approaches have been employed for the analysis of locus-specific structural changes, the first two are based on probe hybridization and include cytogenetics and DNA blotting (Southern or dot blotting), and the third approach is based on PCR amplification. Because of the broadest availability of PCR, most of the methods developed are based on this approach.

fluorescent *in situ* hybridization

This methods is based on the hybridization of fluorescent probes to metaphase or interphase nuclei followed by analysis with a fluorescence microscope. FISH detects cancer-related defects in genomic structure involving altered copy number (deletions and duplications) or without net copy number changes (translocations and inversions) [4, 5]. The principle of FISH can be applied not only for locus-specific analysis but also for whole-genome analysis using multicolor probes (see below). The resolution of the method ranges from entire chromosomes to single loci and can be narrowed down to a few kilobases using fiber-FISH. The preparation of fiber-FISH is, however, not routinely available, making the resolution one of the limitations of this method. Moreover, FISH is relatively labor intensive compared with other methods.

Southern and dot blotting

Southern and dot blotting can be used to detect copy number changes (deletions and duplications) as compared with an internal standard. In addition, Southern blotting may reveal novel restriction fragments created by the rearrangements (e.g. junction fragments), which may indicate the presence of structural changes not associated to copy number changes (translocations or inversion). In this respect, polymorphisms may hamper the interpretation of the results of Southern blotting when a single restriction enzyme is employed. Carrying out independent experiments with different enzymes can overcome this problem. In general, these methods are considered of limited value as routine applications in cancer genetics, since they are laborious and time consuming, require large amounts of high-molecular weight DNA and their interpretation may be hampered by false-negative results [6].

PCR-based methods

Several PCR-based protocols were developed because these techniques are more suited than FISH and Southern blot for routine applications.

microsatellite or single nucleotide polymorphism genotyping and relative quantitation of alleles. Loss of a polymorphic allele marker in tumors (loss of heterozygosity) has been commonly employed to detect somatic deletions in tumors. Polymorphic microsatellite loci and/or single nucleotide polymorphism (SNP) markers have been employed for this purpose. Aneuploidies including supernumerary chromosomes [7] were investigated using quantitative PCR of short tandem repeats. SNPs are, however, more suited to achieve relative quantitation of alleles either by primer extension experiments [8, 9] or by pyrosequencing [10] and can be employed to detect changes in allele copy number (deletions and duplications), but not balanced rearrangements (translocations or inversion). All these methods are relatively simple, cost-effective and with the exception of pyrosequencing, can be implemented using standard equipment available in all molecular biology laboratories. They require heterozygous polymorphic markers in the genomic region of interest, which is not generally a limitation because the human genome contains millions of polymorphic SNP markers and a relatively large number of microsatellites.

real-time quantitative PCR. Real-time PCR is a practical and accurate way to achieve quantitation of a specific genomic target and can be applied to the determination of gene copy number [11]. A number of different applications using probes or intercalating dyes are available but the use of locus-specific probes appears more robust as it avoids nonspecific signals that may hamper the interpretation of results. The accuracy of the quantitation, the short hands-on time required for each determination and the closed system that helps to reduce contaminations are among the advantages of real-time PCR. The main disadvantages are the high cost of probes and limitations in the number of multiplex determinations in a single tube. Moreover, the analysis of duplications may be problematic since the signal increment contributed by one extra copy of a given gene would result in a maximum of 3:2 differences as compared with wild type.

long PCR. The method relies on the use of a mixture of two thermostable DNA polymerases, a proofreading and a nonproofreading [12]. This combination allows longer primer extension and larger size of amplicons to be amplified from good-quality genomic DNA (gDNA) (up to 30-kb fragments). Long-range PCR has been used to identify large deletions/duplications and chromosome breakpoints in several disorders because it allows unequivocal confirmation of rearrangements and provides a PCR-based diagnostic tool to search for the specific germline mutation in at-risk family members [6]. However, breakpoint characterization can be very time consuming and it is particularly challenging in case of rearrangements involving fragments too large to be amplified or encompassing the first or the last exon of a gene [6, 13–17]. Therefore, long-range PCR is not easily implemented in a routine mutation scanning.

multiplex PCR-based methods

These methods are designed to screen for copy number changes by comparative quantitation of simultaneously amplified DNA fragments representative of the multiple loci to be analyzed (Figure 1). They are among the best methods available for this purpose. Their limitation is that balanced rearrangements are not detected, but this is not a major limitation because copy number changes appear to be more frequently implicated in the pathogenesis of cancer than balanced rearrangements. Many methods were developed to detect copy number changes, including multiplex ligation-dependent probe amplification (MLPA), multiplex amplifiable probe hybridization (MAPH), quantitative multiplex PCR of short fluorescent fragments (QMPSF) and nonfluorescent multiplex PCR coupled to high-performance liquid chromatography (NFMP-HPLC) analysis. These semiquantitative PCR protocols rely either on capillary electrophoresis analysis of fluorescently labelled multiplex PCR (MAPH, MLPA and QMPSF) (Figure 1, panels A, B and C) [2, 18, 19] or, more recently, on the analysis of nonfluorescent multiplex PCR coupled to HPLC (NFMP-HPLC) analysis (Figure 1, panel E) [17]. A caveat of these methods is the possibility that nucleotide mismatches at the primer-binding site may interfere with primer hybridization, generating a false indication of exon deletion [20]. Therefore, for diagnostic purposes, the independent confirmation of the analyses using two semiquantitative methods or assays (Figure 2) substantially adds to the confidence of results, especially in cases where a single exon is involved in the rearrangement [17]. Another factor that should be considered to ensure unambiguous results in semiquantitative assays is the quantity and quality of gDNA templates [2, 6, 17]. In particular, low-quality gDNA templates may generate anomalous results that are easily recognized because control peaks are also affected and patterns are not reproducible in replicated experiments.

multiplex ligation-dependent probe. This assay (Figure 1, panel A) is based on hybridization of a set of probes (representing the loci of interest) to gDNA in solution followed by ligation of bound probes. Ligated products are amplified using fluorescent primers and PCR products are

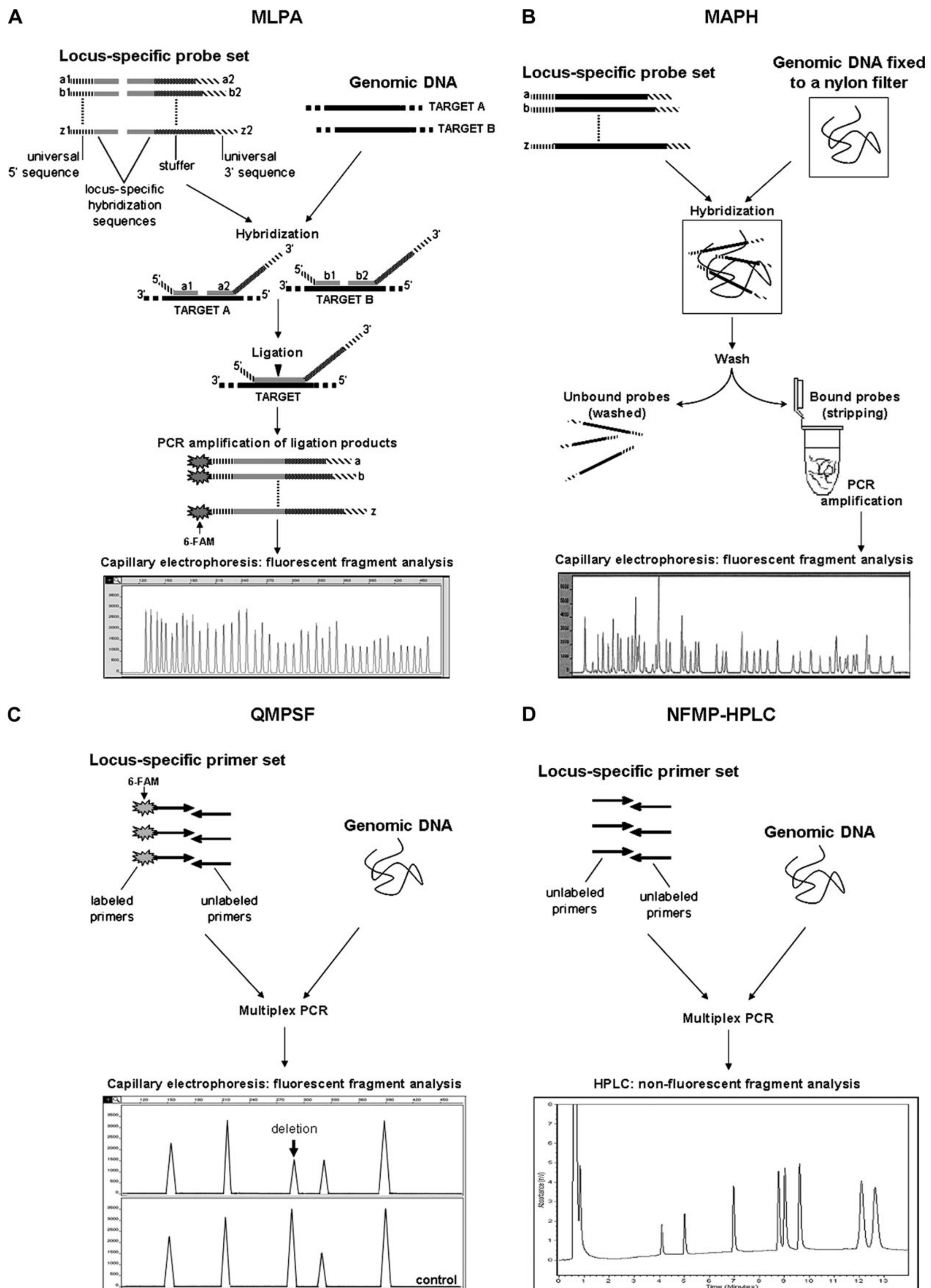

Figure 1. Multiplex PCR-based methods. Four methods are outlined in the panels (see text for details): (A) multiplex ligation-dependent probe amplification (MLPA), (B) multiplex amplifiable probe hybridization (MAPH), (C) quantitative multiplex PCR of short fluorescent fragments (QMPSF) and (D) nonfluorescent multiplex PCR coupled to high-performance liquid chromatography (NFMP-HPLC) analysis.

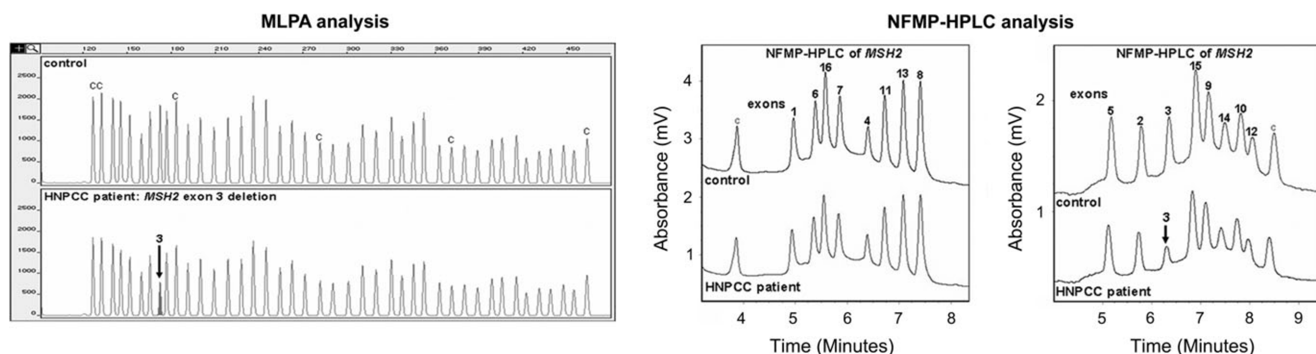

Figure 2. Comparison of results obtained by multiplex ligation-dependent probe amplification (MLPA) and nonfluorescent multiplex PCR coupled to high-performance liquid chromatography (NFMP-HPLC) analysis. Each panel shows representative examples of tracings obtained in a control individual (top) and in one hereditary nonpolyposis colorectal cancer (HNPCC) patient (bottom) by MLPA (left) and NFMP-HPLC (right) analyses. Control peaks are labelled 'c' and arrows indicate the putative deletion of *MSH2* exon 3 in this patient.

analyzed by capillary electrophoresis [21]. This method has the advantage of being available as a commercial kit and requires minimal hands-on time to be completed [19, 22]. In this regard, MLPA might be considered as an early step in molecular diagnosis for gene dosage measurements. Up to 40 probe sets have been successfully multiplexed and MLPA kits have been developed for many genes of interest for oncology research [23–25]. The costs and hands-on time required for each analysis are potential limitations of this fluorescent assay.

multiplex amplifiable probe hybridization. This method (Figure 1, panel B) is based on hybridization of a set of probes to gDNA fixed on a solid support, followed by washing of the unbound probes. Bound probes are PCR amplified with fluorescent primers and analyzed by capillary electrophoresis [21]. These probes are generated by cloning and are all flanked by the same vector sequences, allowing the use of universal PCR primers. This design simplifies the simultaneous amplification of multiple targets in a single PCR and reduces the cost for the setting of the method since only one labelled primer is required. Conversely, the presence of a solid support for DNA may pose a contamination risk. Thus, in addition to MLPA limitations, MAPH also requires accurate removal of unbound probes by a stringent washing step, with increased hands-on time.

quantitative multiplex PCR of short fluorescent. This assay (Figure 1, panel C) is based on the simultaneous amplification of short sequences corresponding to the different segments of the genomic region of interest using locus-specific labelled primers. QMPSF protocols have been successfully adapted to the analysis of different genes, which can be screened using several multiplex PCRs [2, 26–28]. The main disadvantage of QMPSF is the relevant investment for the initial setup of the method, because several fluorescent primers are in most cases required for the analysis of multiple gene fragments [2, 26].

nonfluorescent multiplex PCR coupled to high-performance liquid chromatography analysis. This method recently developed in our laboratory is similar to QMPSF, except that multiplex PCRs are nonfluorescent and are analyzed by HPLC [17]. This assay has several advantages including the short hands-on time required for each reaction, the low cost of nonfluorescent multiplex PCRs and HPLC analysis. Moreover, we have previously reported that it may complement MLPA to provide a robust and cost-effective strategy for screening and independent confirmation of putative genomic rearrangements [17] (Figure 2).

whole-genome methods

cytogenetic methods

These methods are useful for the identification of aneuploidies gross deletions and duplications. FISH with multicolor probes (e.g. chromosome painting and spectral karyotyping) helps to identify chromosomal rearrangements. Among cytogenetic analyses, comparative genomic hybridization (CGH) is a method employed to detect somatic alterations in genomic copy number (gains/losses) in tumor cells [29]. The method is based on the hybridization of test/proband (e.g. tumor) and normal DNA (fluorescently labelled with distinct colors) to normal human metaphase preparations. Quantitative image analysis of epiflourescence microscopy reveals regional differences in the fluorescence ratio of proband and control DNA, reflecting abnormal copy number in specific regions of proband cell genome. CGH has a low resolution and is not suited to detect balanced reciprocal translocations or inversions. More recently, CGH has been adapted to the use of array-based technology (see 'Array-CGH' section) which greatly increases resolution of the analysis. In general, cytogenetic methods are relatively labor intensive and even after using fluorescent probes, the size of the rearrangements that can be detected remains a limitation especially when alterations involve only portions of a gene.

array-based methods. Array-CGH. This technique is a development of the original cytogenetic CGH, but arrays are used instead of metaphases to hybridize differentially labelled test and normal DNA. Array-CGH represents one of the best methods for genome-wide screening of structural variation involving copy number changes [30]. Large-insert clones or long oligonucleotides (60–100 bp) are used to construct arrays. Oligonucleotides have the advantage to achieve an improved resolution as compared with long clones (from 50 kb down to a few kilobases). These methods are excellent for whole-genome analysis, but their present cost and resolution make them less useful for the routine diagnosis of limited rearrangements in specific genes, although this limitation may be overcome in the near future.

SNP array. This sequence-based array approach has the potential to analyze SNP genotypes on a whole-genome basis. Similar to what is described for genotyping of single SNPs, they are a useful tool to study somatic variations in copy number that generate allelic imbalances (e.g. deletions or amplifications in tumor versus normal DNA) [31]. In addition to somatic studies these arrays were also used to screen for germline variation in gene copy number [32, 33]. Using two independent SNP array platforms, incompatibility of array-generated data with Mendelian inheritance was successfully used to screen hemizygous germline deletions in parent–offspring trios [32, 33]. Moreover, using a bead array-based platform, the detection of clusters of null genotypes identified homozygous deletions, while deviations from Hardy–Weinberg equilibrium provided supporting evidence for the presence of either homozygous or heterozygous germline deletions [33]. SNP arrays are a powerful tool for the analysis of structural rearrangements in oncology, but their cost still limits a wide application of this technology.

other whole-genome methods

A number of other approaches are being pursued for the analysis of whole-genome structural variations including paired sequence analysis and comparative sequence analysis [34, 35]. Some of these methods are very powerful, but their description is beyond the scope of the present review because their cost and complexity greatly limits a wide application of this technology.

conclusion

Several studies have emphasized the significant role of large genomic rearrangements in a wide range of common diseases and phenotypes including cancer [36]. A major challenge, however, for the measurement of germline and somatic gene dosage changes is the availability of high-throughput assays in mutation-scanning studies. In fact, since many current detection methods have limitations in resolution, robustness or costs, their use in a routine diagnostic setting is often an unpractical option. Therefore, the proportion of structural genomic abnormalities playing a pathogenic role in cancer is likely to be underestimated. The recent development of routine methods designed for oncological research, such as commercial MLPA kits and NFMP-HPLC, should help to improve the screening of structural abnormalities, providing a more accurate estimate of these abnormalities. Also, the rapid evolution of array-based methods and the trend to produce lower cost chips holds the promise to widen their application in a clinical setting, which should improve our understanding of the role of structural abnormalities in cancer. The application of whole-genome assays, however, poses another important question that needs to be addressed to interpret correctly the results of these studies. According to our current knowledge, many structural variants do not appear related to genomic disorders or diseases [37, 38], while others can influence gene dosage and phenotype, predisposing to or causing disease [36]. Several studies are currently underway [39, 40] to analyze structural variation in apparently normal individuals. The results of these studies should provide a reference to recognize structural variations unique to cancer that are likely to play a pathogenic role. Additional studies will be also necessary to understand whether structural variation identified in apparently normal individuals may contribute to polygenic cancer predisposition.

In conclusion, the study of structural variation in human genome is an active field of research where developing technological and scientific advancements have the potential to provide important contributions to oncological research.

references

1. Stenson PD, Ball EV, Mort M et al. Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 2003; 21: 577–581.
2. Charbonnier F, Raux G, Wang Q et al. 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 2000; 60: 2760–2763.
3. Bunyan DJ, Eccles DM, Sillibourne J et al. Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. Br J Cancer 2004; 91: 1155–1159.
4. Grade M, Becker H, Liersch T et al. Molecular cytogenetics: genomic imbalances in colorectal cancer and their clinical impact. Cell Oncol 2006; 28: 71–84.
5. Fox JL, Hsu PH, Legator MS et al. Fluorescence in situ hybridization: powerful molecular tool for cancer prognosis. Clin Chem 1995; 41: 1554–1559.
6. Nakagawa H, Hampel H, de la Chapelle A. Identification and characterization of genomic rearrangements of MSH2 and MLH1 in Lynch syndrome (HNPCC) by novel techniques. Hum Mutat 2003; 22: 258.
7. Mansfield ES. Diagnosis of Down syndrome and other aneuploidies using quantitative polymerase chain reaction and small tandem repeat polymorphisms. Hum Mol Genet 1993; 2: 43–50.
8. Laan M, Gron-Virta K, Salo A et al. Solid-phase minisequencing confirmed by FISH analysis in determination of gene copy number. Hum Genet 1995; 96: 275–280.
9. Curia MC, Palmirotta R, Aceto G et al. Unbalanced germ-line expression of hMLH1 and hMSH2 alleles in hereditary nonpolyposis colorectal cancer. Cancer Res 1999; 59: 3570–3575.
10. Ronaghi M, Nygren M, Lundeberg J, Nyren P. Analyses of secondary structures in DNA by pyrosequencing. Anal Biochem 1999; 267: 65–71.
11. Hoebeeck J, Speleman F, Vandesompele J. Real-time quantitative PCR as an alternative to Southern blot or fluorescence in situ hybridization for detection of gene copy number changes. Methods Mol Biol 2007; 353: 205–226.
12. Barnes WM. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. Proc Natl Acad Sci USA 1994; 91: 2216–2220.
13. Wang Y, Friedl W, Lamberti C et al. Hereditary nonpolyposis colorectal cancer: frequent occurrence of large genomic deletions in MSH2 and MLH1 genes. Int J Cancer 2003; 103: 636–641.
14. Wang Y, Friedl W, Sengteller M et al. A modified multiplex PCR assay for detection of large deletions in MSH2 and MLH1. Hum Mutat 2002; 19: 279–286.
15. Wimmer K, Yao S, Claes K et al. Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. Genes Chromosomes Cancer 2006; 45: 265–276.
16. Charbonnier F, Baert-Desurmont S, Liang P et al. The 5' region of the MSH2 gene involved in hereditary non-polyposis colorectal cancer contains a high density of recombinogenic sequences. Hum Mutat 2005; 26: 255–261.
17. De Lellis L, Curia MC, Catalano T et al. Combined use of MLPA and nonfluorescent multiplex PCR analysis by high performance liquid

chromatography for the detection of genomic rearrangements. Hum Mutat 2006; 27: 1047–1056.
18. Armour JA, Sismani C, Patsalis PC, Cross G. Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res 2000; 28: 605–609.
19. Schouten JP, McElgunn CJ, Waaijer R et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002; 30: e57.
20. Wehner M, Mangold E, Sengteller M et al. Hereditary nonpolyposis colorectal cancer: pitfalls in deletion screening in MSH2 and MLH1 genes. Eur J Hum Genet 2005; 13: 983–986.
21. White S, Kalf M, Liu Q et al. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. Am J Hum Genet 2002; 71: 365–374.
22. Sellner LN, Taylor GR. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat 2004; 23: 413–419.
23. Taylor CF, Charlton RS, Burn J et al. Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. Hum Mutat 2003; 22: 428–433.
24. Montagna M, Dalla Palma M, Menin C et al. Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. Hum Mol Genet 2003; 12: 1055–1061.
25. Michils G, Tejpar S, Thoelen R et al. Large deletions of the APC gene in 15% of mutation-negative patients with classical polyposis (FAP): a Belgian study. Hum Mutat 2005; 25: 125–134.
26. Casilli F, Di Rocco ZC, Gad S et al. Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments. Hum Mutat 2002; 20: 218–226.
27. Bougeard G, Brugieres L, Chompret A et al. Screening for TP53 rearrangements in families with the Li–Fraumeni syndrome reveals a complete deletion of the TP53 gene. Oncogene 2003; 22: 840–846.
28. Audrézet MP, Chen JM, Raguenes O et al. Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms. Hum Mutat 2004; 23: 343–357.
29. Abdel-Rahman WM, Ollikainen M, Kariola R et al. Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. Oncogene 2005; 24: 1542–1551.
30. Pinkel D, Segraves R, Sudar D et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998; 20: 207–211.
31. Engle LJ, Simpson CL, Landers JE. Using high-throughput SNP technologies to study cancer. Oncogene 2006; 25: 1594–1601.
32. Conrad DF, Andrews TD, Carter NP et al. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 2006; 38: 75–81.
33. McCarroll SA, Hadnott TN, Perry GH et al. Common deletion polymorphisms in the human genome. Nat Genet 2006; 38: 86–92.
34. Feuk L, MacDonald JR, Tang T et al. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. PLoS Genet 2005; 1: e56.
35. Tuzun E, Sharp AJ, Bailey JA et al. Fine-scale structural variation of the human genome. Nat Genet 2005; 37: 727–732.
36. Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. Annu Rev Genomics Hum Genet 2002; 3: 199–242.
37. Barber JC, Joyce CA, Collinson MN et al. Duplication of 8p23.1: a cytogenetic anomaly with no established clinical significance. J Med Genet 1998; 35: 491–496.
38. Buckland PR. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. Ann Med 2003; 35: 308–315.
39. Iafrate AJ, Feuk L, Rivera MN et al. Detection of large-scale variation in the human genome. Nat Genet 2004; 36: 949–951.
40. Sebat J, Lakshmi B, Troge J et al. Large-scale copy number polymorphism in the human genome. Science 2004; 305: 525–528.