The human genome; you gain some, you lose some

Kriek, M.

Chapter IV-2

Summary

For decades, trypsin Giemsa banding of metaphase spreads has been the standard diagnostic method to detect chromosomal rearrangements. The method has several advantages; all chromosomes are seen under the microscope, and individual cells can be karyotyped, permitting clonal analysis (Caspersson, Lomakka, and Zech 1972; Yunis 1976). A major limitation is the fact that due to the contraction of chromosomes during metaphase, and the resolution of the light microscope, G banding is not capable of identifying rearrangements that are smaller than 3-5 Mb.

Fluorescence in Situ Hybridisation (FISH) (Landegent *et al*. 1985) (Ried *et al*. 1990) partly overcame this problem, allowing direct testing for the presence or absence of a specific genomic region. This method is especially used for the identification of micro deletion syndromes and subtelomeric rearrangements. It has several drawbacks though, as detecting rearrangements using FISH analysis is only possible when an obvious, specific phenotype is present that is recognized by a specialist, and when a specific FISH probe is available. Multi-colour methodologies have been developed (Knight *et al*. 1997) (Engels *et al*. 2003), however, the number of loci that can be analysed simultaneously remains limited.

Assays for gene copy number or gene dosage have long been utilized in the clinical molecular genetic laboratories. For many years, Southern blotting analysis (Southern 1975) followed by densitometry was the main assay available to assess for a small number of copy number variations. The development of real-time Polymerase Chain Reaction (PCR), Multiplex Amplifiable Probe Hybridisation (MAPH) (Armour *et al*. 2000) and Multiplex Ligation-dependent Probe Amplification (MLPA) (Schouten *et al*. 2002) allowed more widespread analysis of gene dosage. MAPH and MLPA are PCR-based methods to simultaneously determine the copy number of a large set, currently up to 60, of different chromosomal loci. The advantage of MAPH/MLPA compared to other techniques, such as FISH, is that the resolution of detection is limited only by the size of the probes used (100-500 bp) and it facilitates the parallel screening of several tens of patients at many different loci in one experiment.

Chapter II-1 describes the use of MAPH probe sets covering different genomic loci, including subtelomeric regions, regions involved in microdeletion syndromes and a set of genes evenly spread through out the rest of the genome. Using these probe sets, 184 mentally retarded patients were screened. Results included the detection of rearrangements in subtelomeric and pericentromeric regions, as well as several interstitial alterations, indicating that submicroscopic alterations with a higher frequency were not limited to the ends of the chromosomes.

In **Chapter II-2**, a MAPH assay was designed containing exon-specific single copy

sequences from within a selection of the 169 regions flanked by duplicons that were identified, at a first pass, in 2001. Subsequently, the frequency of chromosomal rearrangement among patients with mental retardation (MR) and/or congenital malformations (CM) was determined. The same study population was tested for rearrangements in regions with no known duplicons nearby, using a set of probes derived from function-selected genes. As expected, the alteration frequency per unit of DNA is much higher in regions flanked by duplicons (fraction of the genome tested: 5.2%) compared to regions without known duplicons nearby (fraction of the genome tested: 24.5%). Thus, the data supported the emerging hypothesis that regions flanked by duplicons are enriched for copy number variations.

Recently, technologies have been developed that provide both genome-wide and high resolution analysis. The first arrays used relatively large fragments of DNA (~150 kb) isolated from mainly Bacterial Artificial Chromosome (BAC) (Solinas-Toldo *et al.* 1997; Pinkel *et al.* 1998; Snijders *et al.* 2001). A newer format uses oligonucleotide probes of 25 to 60 nt in length. Due to the smaller size of these probes and the much larger number of loci analysed, it is possible to detect much smaller copy number variations (CNVs) compared to the variations that can be revealed using BAC-PAC clone array. The 25-mer probe arrays were originally designed for use in genome-wide SNP analysis, for linkage and association studies. As successive arrays have provided ever increasing coverage, currently up to 500-1000K, they were quickly used to estimate copy number changes by using both the signal strength and SNP score. The 60 nt oligo based arrays give stronger signal intensity and CNVs can therefore be detected using solely the signal intensity.

To assess the value of MLPA and array based techniques in clinical diagnosis, 58 developmentally delayed (DD) patients with a normal karyotype were independently tested with both array-CGH and MLPA. The results are described in **chapter II-3**. It shows that both methods are effective and represent an improvement to classical and molecular cytogenetics, as currently applied. Considerations balancing cost-efficiency and complexity promote a format where MLPA screening precedes array-CGH analysis. In addition, an alternative screening strategy, encompassing MLPA testing prior to karyotyping for unselected samples is described. However, at this moment, whole genome array analysis has become affordable, making MLPA analysis as a first step not necessary anymore.

Chapter III-1 highlights the value of using different genomic approaches to unravel chromosomal alterations and their phenotypic impact. Albeit was possible to identify a previously unreported rearrangement within the 22q11 region, e.g. a dele-

tion of the proximal part of chromosome band 22q11. It is argued that a deletion in this chromosome band is unrelated to the phenotypic trait seen in the index patient. The other chromosome 22 homologue carries a duplication of the Velocardiofacial/DiGeorge syndrome (VCFS/DGS) region. In addition, a previously undescribed deletion of 22q12.1, located in a relatively gene-poor region, was identified.

Chapter III-2 describes the identification of the gene involved in Peters Plus syndrome. This shows how the application of new techniques leads to the identification of the pathogenetic mutations of this autosomal recessive syndrome.

Chapter III-3 shows that MLPA testing is capable of fine mapping the breakpoints of different interstitial 16p deletions. As the clinical features are not very distinctive, the screening of the ATR-16 region in patients presenting with mild to moderate MR and microcytic hypochromic anemia with normal ferritin levels is proposed.

Finally, four platforms (a 3 K BAC clone array, a 44 K microarray and two SNP-based arrays (250-500K, and 317K) have been compared for their ability to identify the breakpoints in four patients with different sizes of interstitial 2p deletions, all localised within 2p16.1-p21(**chapter II-4**). All platforms identified the deletion and the results were comparable cross-platform.

It is evident that the genome-wide high resolution arrays provide an enormous improvement of the resolution of chromosome analysis. However, preliminary studies indicate that the extent of 'normal' copy number variation in the human genome may amount to at least 12 % (Redon *et al*. 2006). This percentage is still rather small compared to the number of variants that will be revealed using next generation sequencing. This sequence-based whole genome genotyping of a patient will soon be possible on a large scale. Consequently, the problem we are now facing with CNVs based on the outcome of array screening will be amplified substantially. Collecting data of sequence variation in very large groups of healthy individuals as well as of well-characterised patients will be needed to understand the results in the near future.

Thus, for each apparent rearrangement detected, it is necessary to determine its phenotypic consequences.