The human genome; you gain some, you lose some

Kriek, M.

Chapter IV

Discussion
Summary
Nederlandse Samenvatting

Chapter IV-1

Discussion

We are currently able to apply genome-wide screening tools with an unprecedented resolution to detect progressively smaller variants. It can be argued that this will improve the basis of genetic counselling significantly, as the probability of finding variants that may be related to the impairment of development and health in a patient will increase, and more information can be given about recurrence risks. It also enables us to verify assumptions that were made long before karyotyping and molecular diagnostic tools were invented. In these early days, geneticists have categorised large patient groups as having a multifactorial cause for their developmental delay. With the identification of variants that show a clearly detectable, but incomplete association with MR, one can now 'prove' on a molecular basis that the assumptions made were correct.

Although the identification of new variants is gratifying, it is accompanied by a progressively more difficult task for the people working in human and clinical genetics. After the introduction of karyotyping, a relatively small number of heteromorphisms (e.g. variants not related to human disease) were recognised and documented. This is in contrast to the current situation, where the number of variants with an unknown contribution to genomic disorders is huge. It has recently become clear that, by zooming in on the human genome using array-based platforms, variations exist at an unexpectedly high frequency among healthy individuals; as much as 12% of the human genome show CNVs that are probably not related to any clinical feature (Redon *et al.* 2006).

So, the more we learn about the human genome, the more we are confronted with questions about the implications of new findings. Does it involve a disease-causing alteration or is it a neutral variant?

In many reports the authors have only considered *de novo* variants to be causative. As soon as it became clear that one of the parents carried the same CNV, as the one detected in the affected child, it was thought to be a neutral variant. This is not always a correct assumption, as familial variants might be related to genomic disorders due to phenotypic variability (Ullmann *et al.* 2007), the presence of an autosomal recessive disorder (chapter III-2), or related to a deletion of an imprinted region that may be silent in one parent and disease-causing in the next generation. To complicate the picture even further, genetic disorders can also originate by a combination of two or more variations inherited from two parents, where each of which alone will not result in disease (Klopocki *et al.* 2007) (Lupski 2007). In addition, we can expect that some of the regions showing CNVs among healthy individuals contribute to genetic disease. This would indicate that CNVs present in regions described in the Human Variation database are not always neutral variants. In contrast, as pointed out in chapter I-5, the

finding of a *de novo* variation in an affected individual does not automatically mean that the alteration is disease causing. All these observations impact greatly on genetic counselling and this also underlines the main drawback of using the new platforms, as we sometimes lack the knowledge to adequately inform the patient and the family of the consequence of any finding. To resolve this, it is of great importance to collate CNV data in databases that are accessible to everyone working in this field. Two of such databases are available at this moment (ECARUCA and DECIPHER).

New tools for genome analysis reach the diagnostic laboratory at a quick pace. As a result, one can consider several technical approaches to help diagnose the patient with mental retardation and / or congenital malformation. In chapter II-3, we proposed a strategy in which MLPA covering the chromosome ends and regions related to micro deletions/ duplications should be used first, and if uninformative be followed by whole genome analysis. As pointed out by Rauch *et al.* (2006), this first step will detect an aberration in 5-20% of the MR patients, depending on the criteria used for selection. Since these rearrangements are also readily detected by currently available genome-wide screening tools (arrays), the use of these arrays as a first step now seems a more logical way to go, if it is possible to implement this in the diagnostic setting. Due to the necessity of guaranteeing Standard Operating Procedures in the diagnostic laboratories, it is often difficult to implement the most recent technologies that have proven to be efficient in a research setting. The rapid evolution of technology demands constant adaptation from both the clinician (who has to explain the outcome of the screening towards the patient) and the laboratory (validation and implementation of a new technique) in order to continue applying state of the art diagnostic methods.

At this moment, there is no golden standard available for determining which genome-wide screening platform provides the most relevant data for diagnostic purposes. The advantages of both CGH array based screening and high-density SNP genotyping have been discussed in section I.6.3.5. A recent study (Redon *et al.* 2006) has shown that in addition to the SNP-arrays, array-CGH analysis is required to cover all CNV regions in the human genome, with at least one third of CNVs >50 kb otherwise being missed. New arrays of both Affymetrix and Illumina are closing this gap by combining both SNP- and non-SNP probes on one array. In addition, Nimblegen now has a 42 M non-SNP array available enabling the detection of variants as small as 500 bps.

Although we already struggle to arrive at correct and comprehensive interpretation of high resolution array analysis in a diagnostic setting, a next generation of technical advance is approaching.

Recently, the whole genome of Nobel laureate Jim Watson was sequenced, revealing as many as 600,000 variants that had not been reported before. The cost involved of this project was substantial and therefore this way of screening the human genome is not yet applicable on large scale. It can be expected, however, that affordable sequence-based whole genome genotyping will become possible within the coming 2-5 years. As a result, SNP typing and array-CGH will be superseded fairly soon by next generation sequencing. The first step towards the implementation of genome wide sequencing would be increasing the knowledge about "harmless" variations in a large group of normal individuals, since on average 1 in 1000 nucleotides in the genome of a two healthy individuals varies. In addition, screening large cohorts of affected individuals with well-defined clinical features is essential to be able to interpret this new data (Ropers *et al*. 2007).

The possibility of 'reading' the whole human genome at the nucleotide level will also provide information about susceptibility for diseases that are not related to the patients' reason for consulting a specialist. This issue should be discussed with the patient or the parents during the counseling prior to genome-wide testing. One might choose to communicate only the variants that are thought to be causally related to the patients' phenotype or those that are well known to have a great potential influence on the patients' health (for example inactivation of tumor-suppressor genes). Two examples of alterations in tumor-suppressor genes detected after screening the human genome for MR-related CNVs are described in chapter III-2 and III-4. The patients in chapter II-2, carrying an interstitial 2p deletion, have a high chance of developing a HNPCC related tumor, as the deletion includes the *MSH6* gene. In chapter III-4, the mother and the maternal grandmother of the two index patients with Peters Plus syndrome were found to have a 1.5 Mb deletion encompassing a part of the *BRCA2* gene. Despite the fact that both women already developed breast cancer, they are now confronted with a high recurrence risk and a moderate increased risk of developing ovarian cancer. These 'side effects' of screening can't be avoided. However, a positive consequence of this knowledge is the fact that these patients can now be included in a screening program.

In summary, we can conclude that the plasticity of the genome creates a conundrum of Babylonic proportions. Nevertheless, it is expected that the implementation of new screenings technologies will give greater insight into a range of genetic diseases, and will hopefully lead to a better understanding of the many different causes of intellectual disability and congenital malformations.