The Molecular Dissection of Contiguous Gene Syndromes
With A Focus On 4p16 Deletion Syndrome

Femke HANNES

Doctoral thesis in Biomedical Sciences

Leuven, 2010
The Molecular Dissection of Contiguous Gene Syndromes
With A Focus On 4p16 Deletion Syndrome

Femke HANNES

Jury:

Promoter: Prof. dr. Joris Robert Vermeesch
Co-promoter: Prof. dr. Jean-pierre Fryns
Chair: Prof. dr. Koen Devriendt
Jurymembers: Prof. dr. Guy Froyen
Prof. dr. Dieter Lambrechts
Prof. dr. Orsetta Zuffardi
Prof. dr. Pawel Stankiewicz
Prof. dr. Frank Kooy

Doctoral thesis in Biomedical Sciences

Leuven, 2010
Dedication to all mentors who have nurtured my passion for science and more specific human genetics
Table of Contents

Abbreviations............................................................................................................................................................ 9

GENERAL INTRODUCTION AND AIMS OF THE STUDY .............................................................................................. 11
1. GENOTYPE-PHENOTYPE CORRELATIONS ...................................................................................................... 13
   1.1. Microscopically visible chromosomal rearrangements....................................................................... 13
   1.2. Submicroscopic rearrangements......................................................................................................... 14
2. BROKEN CHROMOSOMES LEADING TO TERMINAL DELETIONS .................................................................... 21
   2.1. Telomeres............................................................................................................................................ 21
   2.2. Rescue of telomeres ............................................................................................................................ 22
   2.3. Telomere length control ...................................................................................................................... 23
   2.4. Position-effect variegation or PEV....................................................................................................... 24
3. AIMS OF THE STUDY...................................................................................................................................... 25

GENOTYPE-PHENOTYPE CORRELATIONS ........................................................................................................... 27
Paper 1: Recurrent reciprocal deletions and duplications of 16p13.11 .................................................................. 29
Paper 2: Genotype-phenotype correlation in 21 patients with Wolf-Hirschhorn syndrome .................................. 49
Paper 3: Duplication of the Wolf-Hirschhorn syndrome critical region causes neurodevelopmental delay........... 71
Paper 4: A familial microdeletion 600 kb proximally of the WHSCR reveals a mild WHS phenotype..................... 83
Paper 5: Benign and pathogenic copy number variation on the short arm of chromosome 4............................. 101

MECHANISM GENERATING 4P TERMINAL DELETIONS........................................................................................ 113
Paper 6: Telomere healing following DNA polymerase arrest induced breakages is likely the main mechanism generating chromosome 4p terminal deletions ................................................................................................... 115

INVESTIGATING THE ROLE OF POSITION-EFFECTS IN 4p16 MICRODELETION SYNDROME ................................. 137
Paper 7: Silencing of telomere flanking genes on chromosome 4p16 suggests telomere position-effects .......... 139

GENERAL DISCUSSION & CONCLUSION................................................................................................ ............ 157
1. Genetic dissection of common features in microdeletion syndromes............................................................... 159
   1.1. The advantage of molecular karyotyping .......................................................................................... 159
   1.2. The use of animal models in dissecting CGS ..................................................................................... 161
   1.3. The use of induced pluripotent stem cells ........................................................................................ 162
   1.4. Monitoring for additional mutations ................................................................................................ 162
2. Origin of phenotypic variability....................................................................................................................... 162
3. Telomere position-effects ............................................................................................................................... 163
   3.1. Monitoring gene-expression outside the deletion boundaries........................................................ 163
4. Overall conclusion....................................................................................................................................... 166

SUMMARY/SAMENVATTING .......................................................................................................... ................ 167

REFERENCES ...................................................................................................................................................... 173

CURRICULUM VITAE ........................................................................................................................................... 187
List of publications................................................................................................................................................ 189
Acknowledgements-dankwoord........................................................................................................................... 191
Abbreviations

A
aCGH
array comparative genome hybridisation
ALT
alternative lengthening mechanism
azaC
2'Deoxy-5-azacytidine

B
BAC
bacterial artificial chromosome
BIR
breakage induced replication

C
CGH
comparative genome hybridization
CGS
contiguous gene syndrome
CHD
congenital heart defects
CL/P
cleft lip and/or palate
CNVs
copy number variants

D
DGV
database of genomic variants
DNA
Desoxyribo Nucleic Acid
dTRF1
dominant negative mutation of TRF1

E
EBV
epstein barr virus
ES
embryonic stem cells

F
FISH
Fluorescence In Situ Hybridisation
FoSTeS
Fork stalling and Template switching

H
HP1
heterochromatin protein

I
iPS
induced pluripotent stem
ISCA
International standards for cytogenomic arrays
IUGR
Intra-Uterine Growth Retardation

L
LCR
low copy repeat
LINEs
long interspersed nuclear elements

M
MLPA
multiplex ligation probe amplification
MMBIR
microhomology-mediated break-induced replication
MR
mental retardation
MR/MCA
mental retardation and multiple congenital anomalies

N
NAHR
non-allelic homologous recombination
NF1
neurofibromatosis
NHEJ
non-homologous end joining

O
OFC
Occipital-frontal circumference

P
PAC
P1-derived artificial chromosome
Pev
position-effect variegation

Q
qPCR
quantitative PCR
RNA
Ribosomal Nucleic Acid

RT-qPCR
real-time quantitative PCR

SD
standard deviation

SINEs
short interspersed nuclear elements

SNPs
single nucleotide polymorphisms

SPPR
subtelomeric pairing with proximal recombination

STR
simple tandem repeat

TERC
telomerase RNA component

TERRA
telomeric-repeat containing RNAs

TERT
telomerase reverse transcriptase

TPE
telomere position-effect

TRF1-2
telomere repeat binding factors 1 and 2

TSA
Trichostatin-A

UCSC
University of California-Santa Cruz

WHS
Wolf-Hirschhorn syndrome

WHSC1
Wolf-Hirschhorn syndrome candidate 1

WHSC2
Wolf-Hirschhorn syndrome candidate 2

WHSCR1
Wolf-Hirschhorn syndrome critical region 1

WHSCR2
Wolf-Hirschhorn syndrome critical region 2
GENERAL INTRODUCTION AND AIMS OF THE STUDY

Chapter I
General Introduction
1. GENOTYPE-PHENOTYPE CORRELATIONS

1.1. Microscopically visible chromosomal rearrangements

In 1956, the discovery of 46 chromosomes in man was reported [1]. Different chromosomes could be categorised on behalf of their differences in length and asymmetry based on the position of the centromere. Soon thereafter, the first association between a numerical abnormality (trisomy 21) and disease was noticed in a patient with Down Syndrome [2]. Besides trisomy 21, other numerical imbalances including trisomy 13 (Patau syndrome) [3] and trisomy 18 (Edwards syndrome) [4] were identified. These findings were the first steps towards understanding the genetic basis of congenital diseases.

In 1963, the first structural rearrangement in terms of the deletion of the complete short arm of chromosome 5 in a patient suffering from mental retardation and a distinct cat-like cry, known as Cri-du-Chat syndrome, was detected [5]. The search for other structural aberrations was challenging due to the inability to analyze the chromosomes in greater detail. There was a need for better cytogenetic screening techniques to detect structural imbalances with greater accuracy. To this end, in 1968, a fluorescent dye, known as Quinacrine, was used to stain the chromosomes what resulted in a chromosome specific banding pattern (Q-banding) [6]. Later, other dyes were used to visualize different banding patterns also referred to as G-, C- and R-banding. The G-banding or Giemsa banding is the most popular staining method. This staining method colors the G-rich bands dark and the GC poor regions light [7]. Standard metaphase spreads visualize approximately 450 bands whereas the prometaphase staining reveals 550-850 bands what renders it possible to detect smaller structural imbalances (Figure 1).

![Figure 1: A G-banded karyotype of a patient with Cri-du-Chat syndrome del (5)(p-). The arrow indicates the location of the deletion.](image)
This banding approach enables the detection of deletions, duplications, inversions and balanced rearrangements to a resolution of 5 Mb. Karyotyping was introduced as a routine diagnostic tool to diagnose patients with developmental delay. The diagnostic yield is estimated to be around 5% in these individuals [8, 9].

A syndrome is defined as a combination of particular clinical manifestations that are recurrent and well recognizable and assumed to be the result of a single cause [10]. Several syndromes turned out to be characterized by large structural chromosomal imbalances. Since the size of the imbalances identified was large, it was believed that the specific syndromal manifestations were originating from several, physically linked, unrelated genes within the imbalanced region. This concept was referred to as contiguous gene syndromes (CGS) [11]. Each syndrome is defined by a recognizable pattern of malformations, often in association with mental retardation and abnormalities of growth [12]. Some well known examples of contiguous gene syndromes are Wolf-Hirschhorn syndrome and Cri-du-Chat syndrome mainly caused by terminal deletions of the short arms of chromosomes 4 and 5, respectively.

One approach to gain insight in the molecular pathogenesis of contiguous gene syndromes, is by collecting patients with similar phenotypes but imbalances of variable size. The identification of the minimal region of overlap associated with a specific phenotypic feature enables the localization of the genes of that pathology. This method is referred to as genotype-phenotype correlations. The use of karyotyping in genotype-phenotype correlations has led to a first phenotypic map in patients with Cri-du-Chat syndrome [13]. These first studies triggered the systematic collection and description of chromosomal syndromes (e.g. the catalogues of unbalanced chromosome aberrations in man from Albert Schinzel [14].

Although karyotyping was instrumental in the identification of many CGS, it was usually impossible to molecularly dissect the contribution of the genes within the imbalance. First, conventional karyotyping provides a genome wide analysis but has only limited resolution. Thus, aberration boundaries are inaccurately mapped, small deletions overlooked, and complex aberrations improperly identified. Secondly, large deletions convey often very complex phenotypes which make it impossible to dissect the different phenotypic anomalies. Therefore, there was a need for higher resolution molecular methods enabling the identification of smaller chromosomal imbalances and which accurately map variation breakpoints.

1.2. Submicroscopic rearrangements

1.2.1. Targeted detection of copy number variations

Fluorescence in Situ Hybridization (FISH) and Multiplex ligation probe amplification (MLPA) enable the detection of chromosomal rearrangements and/or copy number variations in the genome. FISH relies
on the hybridization of fluorescently labeled locus specific probes on genomic DNA [15] whereas MLPA identifies the target sequence by the hybridization of two adjacent probes to the genome that are joined by a ligation, subsequently amplified and size fractionated [16, 17]. FISH is been used for both the detection of numerical as well as structural chromosomal defects whereas MLPA can only detect copy variations [18, 19]. The resolution of FISH was initially limited because of the use of metaphases, but with advanced cytogenetic techniques, interphase nuclei or even artificially extended chromatin fibers could be used which improved the resolution to respectively 50-100 kb and 5 kb.

FISH was applied for the diagnosis of recurrent genomic disorders. The mapping of the human telomeres enabled the development of subtelomere FISH which made use of a unique set of telomere clones corresponding to each of the human chromosome arms within approximately 500 kb of the end of the chromosome arm [20, 21]. By systematically screening patients with mental retardation using subtelomere FISH, it was shown that cryptic translocations and imbalances involving the telomere were a significant cause of mental retardation [22, 23]. Cryptic subtelomeric imbalances were shown to be the cause of 3-6% of the patients with mental retardation [8, 24, 25].

1.2.2. Genome wide submicroscopic copy number detection

Initially, a variant of the FISH technique was developed by Kallioniemi and coworkers in 1992, known as comparative genome hybridization (CGH) [26]. Instead of hybridizing the labeled probe to metaphases of the patient, genetic material of both the patient and a reference are differentially labeled and co-hybridized on metaphase spreads of a normal individual. A difference in copy number variation will be detected by a fluorescence ratio difference between patient and reference at a certain locus. However, the poor reproducibility, the labor intensity and the limited resolution made this technique cumbersome to detect submicroscopic rearrangements.

Human genome wide libraries of large insert clones (mainly bacterial artificial chromosomes or BACs) were used for the completion of The Human Genome Project [27]. With the availability of those large insert clones, the development of array CGH or array comparative genome hybridization was initiated. Instead of using metaphase spreads, large insert clones were introduced as targets for the

**Figure 2: The principle behind aCGH**
hybridisation of differentially labelled patient and reference material [28, 29] (Figure 2). This approach combined the strength of a targeted high resolution approach such as FISH with that of a whole genome approach like karyotyping. Deletions, duplications or amplifications of segments less than 5Mb could be detected. The first genome wide BAC array contained approximately 3500 clones in which the clones were separated 1 Mbp from each other [30, 31] and soon followed by the development of higher resolution tiling BAC arrays constructed with 32,000 clones covering the whole genome with a resolution of one clone per approximately 100 kb [32-34]. Nowadays, companies have developed micro-arrays containing over two million probes covering the whole genome (or even targeted regions) [35].

Array CGH was rapidly used as a tool to diagnose patients with MR/MCA. 5-20 % of the patients were carrier of de novo submicroscopic rearrangements [30, 36-38].

1.2.3. Genotype-phenotype correlations

Array CGH enabled another approach to perform genotype-phenotype correlations. Instead of starting from a group of patients with similar phenotypes to delineate the minimal region of overlap, large cohorts of patients with unidentified phenotypes could be easily genotyped. Subsequent collection of patients with common genotypes enabled the characterization of similar phenotypes and thus the identification of novel microdeletion/microduplication syndromes. Recently, this latter approach is referred to as the ‘genotype-first’ approach [39]. Furthermore, to clarify the extent to which each gene/region underlying the rearrangement contributes to the phenotype in CGS, the identification of different sized, well mapped submicroscopic imbalances is indispensable. To this end, high resolution molecular techniques enabled the delineation of true CGS, whereas in other CGS, a single gene was identified which is associated with the majority of symptoms linked to a specific syndrome.

Identification of novel recurrent genomic disorders

Recently, it was shown that the regional genomic architecture can predispose to recurrent rearrangements. The homologous blocks with >97% identity and 10-400 kb in size, also known as region-specific low copy repeats or LCRs, can serve as substrates for homologous recombination. The size, sequence identity and orientation of the LCRs renders the region unstable such that when two LCRs happen to misalign, a deletion, duplication, inversion or a translocation is created. The misalignment between LCRs is known as non-allelic homologous recombination (NAHR) [40, 41]. Since NAHR was considered as the major cause of genomic disorders, targeted micro-arrays including regions flanking site specific LCRs were developed and used to investigate patients with mental retardation and multiple congenital anomalies (MR/MCA). An ever increasing number of recurrent microdeletion/microduplication syndromes were identified e.g. on different chromosomes [42-44].
Molecular dissection of contiguous gene syndromes

Genotype-phenotype correlations using array CGH on certain well-known CGS have enabled more accurate molecular delineation of the regions associated with specific phenotypic features in 1p36 microdeletions [45], 1q41 microdeletions [39], 5p microdeletions [46], 9q34 microdeletions [47, 48], 18q microdeletions [49] and 22q13 microdeletions [50, 51]. Typically, in true CGSs different genes underlying the deletion are responsible for the phenotype and thus genotype-phenotype correlations will eventually lead to a genetic dissection of different phenotypic manifestations related to the haplo-insufficiency of different, contiguous genes within the deletion or duplication.

Wolf-Hirschhorn syndrome (WHS) is caused by large contiguous gene deletions of the telomeric end of the short arm of chromosome 4. Interstitial deletions do occur. However, those are less frequently observed [52-54]. Early mapping efforts identified a minimal region of overlap marked as the Wolf-Hirschhorn syndrome Critical Region 1 (WHSCR1) [55]. Further finemapping efforts resulted in a size reduction, containing a 165 kb sequence including two genes, the Wolf-Hirschhorn syndrome candidate 1 (WHSC1) and 2 (WHSC2) gene. This region was suggested to be responsible for at least two of the core clinical manifestations of WHS including the facial gestalt and developmental delay. In 2003, the detection of a 1.9 Mb terminal deletion in a WHS patient partially including the WHSCR1 suggested a second critical region 2 (WHSCR2) partially overlapping the original WHSCR1 at the distal site [56]. Both critical regions contain one common gene, WHSC1, which is partially or completely deleted in every WHS patient reported, so far. The confirmation of previously detected critical regions and, in addition, the identification of certain novel regions of overlap for certain phenotypic manifestations such as microcephaly and cleft lip and palate was demonstrated using micro-array analysis in 5 mildly affected WHS patients [57] (Figure 3).

**Figure 3:** Assembly of the Wolf-Hirschhorn syndrome genotypes and phenotypes (modified from Van Buggenhout et al. [57])
Gene identification within contiguous gene syndromes

Although the identification and the delineation of critical regions in relation to CGS have been established, the identification of single causative genes underlying CGSs remained challenging. It mainly depended on the description of key patients with a specific chromosomal imbalance such as submicroscopic deletions or translocations. For several disorders it was shown that apparently contiguous gene syndromes could be the result of the disruption of one particular gene embedded in a larger chromosomal region, for example, in the majority of patients with neurofibromatosis (NF1), a mutation within a single gene (NF1 gene) is present whereas in 5-10% of the patients a microdeletion is responsible for pathogenesis [58, 59]. Patients with Angelman syndrome and mutations in the UBE3A gene are clinically almost indistinguishable of those carrying a microdeletion containing the UBE3A gene [60, 61]. The main clinical features in Sotos syndrome are predominately caused by one particular gene, NSD1 [62-64]. However, even if a single gene is dominantly influencing the main features of a syndrome, often minor features are influenced by haplo-insufficiency of other genes in the region. For all these syndromes, microdeletions are described in patients with more pronounced facial gestalt, developmental delay and/or a higher incidence of seizures or more severe microcephaly. This suggests that other genes within the deletion contribute to the additional complexity of those phenotypes.

1.2.4. Origin of genomic rearrangements

Three major mechanisms have been proposed that underlie genomic rearrangements in the human genome. Those rearrangements can alter the copy number state of dosage sensitive genes within the region involved or disrupt genes leading to genomic disease. When those pathogenic rearrangements occur during meiosis, it is referred to as human disorders, whereas when those pathogenic rearrangements occur in somatic cells, cancers or mosaicism are generated. All three molecular mechanisms leading to recurrent or non-recurrent rearrangements will be discussed further below.

Recurrent rearrangements

Recurrent rearrangements are most often caused by non-allelic homologous recombination (NAHR) and are recognised by their common size and breakpoint clustering or, in other words, NAHR results in the reshuffling of DNA sequences located in-between two paralogous sequences. Those paralogous sequences are known as segmental duplications or low copy repeats (LCRs), most often vary from 10 kb to 300 kb in size and typically contain > 95% sequence homology. Up to 5% of the human genome account for those paralogous sequences and therefore it is believed to be the most common cause of genomic disorders and benign CNVs [65].
Due to their high degree in sequence homology, non-allelic sequences can misalign and subsequently crossover, leading to recurrent rearrangements and depending on the orientation of both LCRs at which NAHR takes place, different types of recurrent rearrangements are formed. When both LCRs are in direct orientation on the same chromosome, simple deletions or duplications may be generated. When both LCRs are in opposite orientation on the same chromosome, inversions might be generated. NAHR between two LCRs on different chromosomes leads to translocations (Figure 4). Well known examples of recurrent translocations are t(4;8) and t(4;11) likely mediated by NAHR between two olfactory receptor gene cluster LCRs [66, 67]. Recently, Zhishuo and Stankiewicz provided molecular evidence to support NAHR between interchromosomal LCRs (Ou Zhishou and Stankiewicz P, accepted in Genome Research).

Non-recurrent rearrangements

Among a number of CNVs occur at random in the genome without recurrent breakpoints. Those events are called non-recurrent rearrangements. The process of repair is likely to be non-homologous end-joining (NHEJ), a recombination-based mechanism of repair. This mechanism is characterized by the absence of LCRs flanking the breakpoints and often the presence of an “information scar” or microhomology at the site of joining. This information scar contains usually several nucleotides of unknown origin. Both deletions, duplications and translocations can be deduced from this mechanism (Figure 4) [65].

Figure 4: Three mechanisms generating genomic rearrangements (Taken from Gu et al. [65]).
When, on the other hand, non-recurrent CNVs are observed overlapping a single locus, Fork Stalling and Template Switching (FoSTeS) or Microhomology-Mediated Break-Induced Replication (MMBIR) is likely the underlying the mechanism of repair. As opposed to recombination-based mechanisms including NAHR and NHEJ, FoSTeS and MMBIR are replication-based mechanisms. Both mechanisms are deduced from the same principle such a stalling or collapse of the DNA replication fork during DNA replication generating single-stranded ends. Those single-stranded ends invade then a second DNA replication fork and anneals by virtue of microhomology at the 3’-end. Switching to a DNA replication fork downstream results in a deletion, whereas invading a DNA replication fork upstream leads to a duplication. When multiple series of such events occur, very complex rearrangements can be build. In addition, depending on whether the leading of lagging strand was invaded and copied, the orientation differs from direct to inverse compared to the original orientation of the fragment copied (Figure 4) [65].
2. BROKEN CHROMOSOMES LEADING TO TERMINAL DELETIONS

2.1. Telomeres

The primary role of the telomeres is to protect chromosome ends from recombination, fusion and from being recognized as damaged DNA [68]. Besides their protective role, telomeres serve as a template to adhere to the nuclear envelope to form the bouquet formation during meiosis [69].

The protective function of the telomere relies mainly on the telomere-specific DNA conformation, the higher chromatin organisation as well as the telomere-associated proteins. First, a specific telomere DNA conformation is formed by a single-stranded 150-200 bp 3’ nucleotide overhang which then folds back, integrates into the double-stranded telomere sequence and as such forms T- and D-loops (Figure 4) [68]. The structure avoids the chromosome being recognised as a break and protects the ends from exonuclease degradation and telomere fusions. When the telomeres become too short, the overhang cannot properly fold back, and as such provides a template region for the telomerase enzyme which will lengthen the telomeres by the de novo synthesis of telomeric repeats [70]. Secondly, telomeric DNA in mammals is packaged with histones to form highly ordered structures. Those histones are the basic elements to form chromatin. Moreover, the flexible N-terminal tails of the histones are targets for a variety of post-translational modifications, including acetylation and methylation. Depending on the set of specific histone tail modifications, transcriptionally active chromatin domain (euchromatin) or alternatively transcriptionally repressed chromatin domains (heterochromatin) are created. Heterochromatin assembly at the telomere and subtelomere regions in mammals requires a specific set of marks such as trimethylation of H3K9 and H4K20, DNA hypermethylation, histone acetylation and binding of heterochromatin 1 protein (HP1). The formation of their heterochromatic structure has an important role in telomere length regulation and vice versa [71]. Finally, the telomere-associated proteins or the shelterin complex are involved in shaping and defining the final structure of the telomere [72]. In addition, they associate with proteins involved in different pathways such as damage signaling [72], telomere length regulation [73, 74] and heterochromatin assembly [75].
Telomeres are located at the end of linear chromosomes. Their structure is conserved among almost all eukaryotes. The DNA sequence is free of genes and consist of GT-rich tandem repeats: \((G(2-3)TG(1-6)T)n\) in \textit{Sacharomyces Cerevisiae} and hexamere repeat \((TTAGGG)n\) in humans and mice [76]. In humans, the telomere length varies between approximately 5-15 kb ending in a G-rich single-stranded overhang between 150-200 nucleotides. Furthermore, the telomeric repeats are docking sites for a multiprotein complex consisting out of 6 proteins. Those telomere associated proteins, the shelterin proteins, are pivotal in correct formation of the telomere. Three proteins recognize the telomere sequence and two of them bind double-stranded DNA such as telomere repeat binding factors 1 and 2 (TRF1-2) [77, 78], whereas the protein POT1 binds to single-stranded G-rich overhangs [79]. TRF1-2 and POT1 recruit and are interconnected by three additional proteins TPP1, TIN2 and RAP1 that allow cells to distinguish the telomeres from sites of damage [72]. Beside this nucleoprotein structure, a recent discovery has identified telomeric-repeat-containing RNAs (TERRA) at telomeres in addition to the proteins [80, 81].

The subtelomeric region is a transition region between the telomeric cap and the chromosome specific sequence. Those regions consist out of a mosaic patchwork of sequence blocks that are shared among different subtelomeric regions [82]. The subtelomeres contain genes, genefamilies and pseudogenes. Although their function in the genome is unknown, the subtelomeric regions are proposed to diversify the sequence due to the continued shuffling of sequence between the highly homologous blocks and as such create potential new genes and functions. This may help the organism to adapt to new environmental conditions [83].

2.2. Rescue of telomeres

Telomeres are essential for maintaining chromosome integrity. Broken chromosomes have lost their protective ends, the telomeres, and when not properly repaired, this will result in genomic instability, cell senescence and finally in apoptotic cell death. Thus, in order to maintain the broken chromosome during multiple cell divisions, stabilization of the break is pivotal. Two main pathways have been observed whereby the broken chromosome can either retain or acquire a new telomeric cap to become structurally stable within the cell. Both pathways, known as telomere capture and healing, will be discussed further below.

First, during the process of telomere capture, the telomere from another chromosome, sister chromatid or homologous chromosome will be captured. Different models have been proposed by which a telomeric end can be acquired including telomere capture by non-allelic homologous recombination (NAHR), by subtelomeric pairing with proximal recombination (SPPR) or by break-induced replication (BIR). Models like NAHR can be explained by the misalignment between two LCRs whereas in SPPR model, pairing of LCRs at the subtelomeres could facilitate recombination at more proximal regions containing only microhomology [22, 40, 82]. The BIR model proposes the stabilization of a double-stranded break by invading and subsequently replicating sequences from another chromosome end.
The final result will be a derivative chromosome that contains in addition to the telomeric sequence also the subtelomere sequence and possibly even more. Breakpoint mapping of different chromosome arms have shown a variety of rearrangements suggesting telomere capture events as a mechanism in generating derivative chromosomes [87, 88].

The second process by which broken chromosomes are rescued is telomere healing. Hereby, a chromosomal break is stabilized by the direct addition of telomeric repeats either by the telomerase-mediated de novo addition or by the telomerase-independent, recombination-based mechanisms [89-91]. The outcomes are pure terminal truncations. Telomere healing has been proposed to occur in vivo and was detected at terminal truncations of chromosomes 16p and 22q [92, 93]. Consistent with this observation, in vitro studies using human cancer cell extracts demonstrated the addition of telomeric repeat sequences onto the ends of oligonucleotides. The transition region at the breakpoint is characterized by a few bp of sequence in frame with the telomeric repeat [94]. Inversely, in vitro studies have shown the addition of pre-existing telomeres at the ends of a linearized plasmid when transferred to a telomerase negative cell line (SV40 transformed human fibroblast cell line) [95, 96]. Furthermore, in mammalian cells, de novo addition of telomeric repeat sequences by telomerase has been demonstrated to be a common event at induced DSBs in wild type mice embryonic stem cells (ES) whereas chromosome healing rarely occurs in ES cell lines that are deficient in telomerase [97]. Despite this evidence that de novo addition of telomeres can occur in mammalian cells, the understanding of the process leading to terminal deletion syndromes is limited mainly due to the limited number of breaks analyzed at basepair level. To this end, a more in-depth study is required.

### 2.3. Telomere length control

In most human somatic cells, telomeres shorten with every round of DNA replication, eventually leading to cellular senescence or apoptosis, whereas non-dividing cells are capable of maintaining their telomeric DNA over time. This shortening is the result of an asymmetric replication of the lagging and leading strand and subsequent resection of the newly synthesized ends (Figure 5) [98]. This erosion of telomeric DNA can be counteracted mainly by the action of telomerase. This enzyme consists out of two subunits involving a reverse transcriptase (TERT) that synthesizes and adds telomeric repeats to the end of the chromosome using a template molecule also known as telomerase RNA component (TERC) [99, 100]. Telomerase is absent in nearly all somatic cells whereas it is highly expressed during embryonic development, in most cancers and stem cells [101]. In the absence of

---

[Figure 5: The end-of-replication problem]
telomerase, some cancer cells can rescue critically short telomeres by the alternative lengthening (ALT) mechanism, a recombination between telomeric sequences resulting in a heterogeneous telomere length [102].

### 2.4. Position-effect variegation or PEV

Heterochromatin is a highly condensed, inaccessible structure organised in highly ordered nucleosomal arrays. This highly condensed genomic structure has the ability to spread across the initial nucleation site, thereby exerting control over the neighbouring locus. This control manifests itself by the transcriptional silencing of genes subjected to this spreading. In a population of cells, this spreading of heterochromatin can deviate from cell to cell leading to transcriptionally silenced in some cells but active in others and as such producing a mosaic or variegated phenotype. This process is also referred to as position-effect variegation [103].

PEV is seen nearby loci that are highly enriched for repeats and can be induced either by the expansion of repeats or the relocation of transcriptional active gene near a transcriptional inactive, closed region. The repeat expansion causes a heterochromatin-induced silencing of the disease gene. Depending on the length of the repeat expansion, the disease progression might accelerate. Some well known human disorders such as Friedreich’s ataxia, myotonic dystrophy type 1 and fragile X syndrome are likely to be result of PEV [104]. Alternatively, PEV can be induced upon the relocation of a transcriptional active gene near a transcriptional inactive, closed region. The latter event has been mainly studied with use of transgenes. Typically, those transcriptional inactive regions represent the centromeres or telomeres. Those chromosomal structures are known for their constitutive gene poor heterochromatin. Moreover, when repression is initiated by the telomere, it also known as telomere position-effect (TPE).

Telomere-induced silencing at broken chromosomes ends is the best studied of the PEV-related phenomena. This type of silencing in *Saccharomyces cerevisiae* involves linear spreading of repressive chromatin structures originating from the relocated telomere sequence by means the extended binding of heterochromatic proteins proximally from the distal end. A model was proposed as follows [105, 106];

- the initial binding of the *RAP1* protein to the nonnucleosomal telomere sequence
- followed by the recruitment of the SIR complex that has the potential to recruit and bind SIR complexes bound to the nucleosomal subtelomeric sequence
- upon binding, the SIR complex will deacetylate the histone tails of the nucleosomes
- finally leading to a closed structure spreading further into euchromatic telomere flanking region.

Similar as in yeast, telomere position-effects have also been identified in human cancer cells [107, 108]. However, thus far, no TPE has been shown to be involved in human disorders.
3. AIMS OF THE STUDY

The global aim of this thesis is to investigate the contribution of chromosomal rearrangements to human disease and to genetically dissect contiguous gene syndromes.

Initially, we focused on copy number variations identified by microarrays in patients with mental retardation and multiple congenital anomalies. The goal was to determine the association of a phenotype with the imbalanced chromosomal region and to unravel the association between the genes within the imbalance and the clinical features (Chapter II). To this end, we focused on copy number variations identified on chromosome 16p13 (part I) and 4p (part II).

The majority of patients with WHS carry terminal deletions. How terminal deletions arise remains largely unknown. We took advantage of the collection of terminal 4p deletions to characterize a series of terminal breakpoints at sequence level. This approach enabled us to gain more insight into the molecular mechanism causing breakage and stabilization of broken chromosomes (Chapter III).

Telomeres are known to exert position-effect variegation (PEV) on genes flanking the telomeres. The telomere sequence of the chromosome arm with the terminal deletions is relocated to another location and as such may exert PEV on the genes nearby. More recently, it has also been shown that interstitial deletions might influence expression of genes flanking the imbalance. We wondered whether PEV might exist at terminal and interstitial 4p deletions and as such might confound genotype-phenotype correlations. The collection of cell lines derived from 4p deletion patients and the identification of the breakage sites enabled us to study potential position-effects and in particular telomere position effects (Chapter IV).
GENOTYPE-PHENOTYPE CORRELATIONS

Chapter II
The identification of novel microdeletion-duplication syndromes depends on an accurate genotype-phenotype correlation. In the past, those correlations were mainly driven from a collection of patients with a similar and detailed clinical description and later common or overlapping chromosomal rearrangements were uncovered.

Nowadays, with the advent of microarrays, direct mapping of chromosomal imbalances such as deletions, duplications or even amplifications to the genomic sequence is straightforward. Moreover, the higher the resolution of the microarray, the more accurate the imbalance could be defined. Today, because of the ease of which a genotype can be defined, the idea of genotype-phenotype correlation is reversed and is driven from a well defined genotype in a collection of patients with idiopathic mental retardation. This novel workflow is referred to as the ‘genotype-first’ approach [106]. When applying this approach in our screen of MR/MCA patients using microarrays, we noted several interesting submicroscopic rearrangements as described in part I and II.

Part I: Recurrent copy number variations on chromosome 16p13

Part II: Variable sized copy number variations on chromosome 4p
   Paper 2: Genotype-phenotype correlation of 21 WHS patients
   Paper 3: A microduplication of the WHSCR in a patient with MR/MCA
   Paper 4: A microdeletion outside the WHSCR in a patient with mild WHS features
   Paper 5: The construction of a molecular map of the short arm of chromosome 4
Paper 1: Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant


* Both authors contributed equally to this manuscript

Abstract

Genomic disorders are frequently caused by non-allelic homologous recombination between segmental duplications. Chromosome 16 is especially rich in a chromosome 16 specific low copy repeat, termed LCR16. Our BAC array CGH screen of 1027 patients with mental retardation and/or multiple congenital anomalies (MR/MCA) detected 5 patients with deletions and 5 with apparently reciprocal duplications of 16p13 and covers 1.65 Mb, including 15 RefSeq genes. In addition, 3 atypical rearrangements overlapping or flanking this region were detected. Fine mapping by high resolution oligonucleotide arrays suggests that these deletions and duplications result from non-allelic homologous recombination (NAHR) between distinct LCR16 subunits with >99% sequence identity. Deletions and duplications were either de novo or inherited from unaffected parents. To determine whether these imbalances are associated with the MR/MCA phenotype or whether they might be benign variants, a population of 2014 normal controls was screened. The absence of deletions in the control population proves that 16p13.11 deletions are significantly associated with MR/MCA (p=0.0048). Despite phenotypic variability, common features were observed: 3 deletion patients presented with MR, microcephaly and epilepsy (2 of which had also short stature), while 2 other deletion carriers ascertained prenatally presented with cleft lip and midline defects. In contrast to its previous association with autism, the duplication appears to be a common variant in the population (5/1682, 0.29%). These findings demonstrate that deletions inherited from clinically normal parents are likely to be causal for the patients’ phenotype whereas the role of duplications (‘de novo’ or inherited) in the phenotype remains uncertain. This difference in knowledge regarding the clinical relevance of the deletion and the duplication causes a paradigm shift in (cyto)genetic counseling.
Introduction

Non-allelic homologous recombination (NAHR) between neighboring intrachromosomal segmental duplications (also termed Low Copy Repeat, or LCRs) is the major mechanism underlying genomic disorders. Several recurrent clinical syndromes are caused by either gains or losses of sequences flanked by segmental duplications. Following the initial discovery that Charcot-Marie-Tooth disease and HNPP are caused by the duplication and the reciprocal deletion at 17p11.2 respectively, many other recurrent clinical syndromes were shown to be caused by NAHR [1]. The finding that about 4% of the human genome is made up of intrachromosomal segmental duplicons led to speculations that many more unrecognized recurrent rearrangement syndromes might exist [2]. The advent of array CGH has enabled the genome wide screening for copy number variations in large patient populations leading to the identification of several novel low copy repeat-mediated rearrangements [3, 4].

Chromosome 16 is especially rich in intrachromosomal segmental duplications. During recent primate evolution, chromosome 16 has undergone intense segmental duplication activity, and over 10% of the euchromatic region of the p-arm is composed of highly complex low copy repeats [5, 6]. During evolution, these blocks were generated in a stepwise fashion, generating multiple subunits termed LCR16a-t ranging in size from 20-600 kb and sharing >97% sequence similarity [7]. The complex architecture of chromosome 16p therefore suggests it as an excellent candidate region for novel microdeletion syndromes [3]. Recently, Ballif et al. [8] did indeed uncover a novel microdeletion syndrome of 16p11.2-p12.2 which appears to be mediated by NAHR between 16p segmental duplicons. In addition to this novel syndrome, different studies analyzing copy number variation of MR/MCA patients imbalances of 16p13 have been reported [9-11]. However, the relevance of these findings remains unclear.

During the screening of 1027 patients with MCA and/or MR, we identified 6 deletions and 7 duplications of 16p12-p13, apparently caused by NAHR between LCR16. Our detailed analyses of these imbalances define 16p13.11 as a region of recurrent microdeletion/duplication and suggest that the deletion is a risk factor for MR/MCA while the clinical relevance of the duplication is uncertain.

Materials and methods

Selection of patients tested by BAC array CGH

This study collected patients with a mental handicap and/or multiple congenital anomalies, obtained from 3 sources. Patients were ascertained (i) by the clinical geneticist of Leuven, Belgium (n=500), (ii) children and young adults from a variety of UK clinical genetics centers, community learning disability teams and other sources, including hospital neuro-pediatricians (n=372) [3], and (iii) DNA from fetuses with one of more congenital anomalies that underwent autopsy at Children’s Hospital and Regional Medical Center (Seattle, WA) after elective termination or demise (n=155) [12]. All were reported to have a normal karyotype at 550 G-band resolution, and in many cases cryptic subtelomeric rearrangements and other specific genetic abnormalities had been excluded.

Array comparative genome hybridisation
Patients were analyzed using two different BAC array CGH platforms. Patients from Leuven, Belgium (n=500) were hybridized to a custom BAC array with clones spaced at approximately 1Mb intervals throughout the genome, according to the protocol of Menten et al.[13] Regions were scored as copy number variant if one clone passed the threshold of 4x SD, and if two or more flanking clones passed the threshold of log2(3/2)-2 x SD [14]. Patients from the UK (n=372) and USA (n=155) were hybridized to a custom BAC array consisting of ~2000 clones targeted to regions of the genome flanked by segmental duplications [10]. Regions were scored as copy number variant if the log2 ratio of two or more consecutive clones each exceeded twice the standard deviation of the autosomal clones in dye-swap replicate experiments [3].

An additional patient with a 16p13.11 duplication was identified by array comparative genomic hybridization to human 185k genome-wide oligonucleotide arrays (Agilent Technologies) according to manufacturer’s instructions. In brief, genomic DNAs from the patient and from a single sex-matched reference were separately double-digested using the restriction endonucleases Alu I and Rsa I (Promega) and purified using Microcon centrifugal filter devices (Millipore Corporation). 1.5 µg of the digested products were differentially labeled by random priming with Cy3-dUTP and Cy5-dUTP (Perkin Elmer) and co-hybridized to the array for 48 hrs at 65˚C in a rotating oven. Parental DNAs were also hybridized in this way. Hybridized arrays were washed and scanned using an Agilent Microarray Scanner. Image data were extracted using Agilent Feature Extraction software version 8.5 and the data analyzed using Agilent CGH Analytics software version 3.4 (z-score method setting).

To refine breakpoints of these rearrangements, we utilized oligonucleotide arrays. One oligo array consisting of 385,000 isothermal custom made probes (length 45-75 bp) covering a number of chromosomal regions, including this 5 Mb region of chromosome 16p (mean density, 1 probe per 131 bp) (NimbleGen Systems, Madison WI). Hybridizations were performed as described previously [15] and a single normal male was utilized as a reference (GM15724, Coriell, Camden, NJ).

A second oligo array used, was the Affymetrix GeneChip Human Mapping 262K NspI that contains 262,264 25-mer oligonucleotides. 250 ng genomic DNA was processed according to the Affymetrix GeneChip Human Mapping 500K Manual (http://www.affymetrix.com). Copy number was assessed using DNA-Chip Analyzer (dChip) software 2006.[16] Regions of copy number gain and loss were detected using the hidden Markov model output of dChip.

**Real-time quantitative PCR (RT-qPCR)**

RT-qPCR was performed as previously described [13] with minor modifications. Primers were designed from RepeatMasked sequence (www.repeatmasker.org/) using Primerexpress 2.0.0 oligo design software (Applied Biosystems, Lennik, Belgium), and validated with in silico PCR and BLAT (http://genome.ucsc.edu/). SNPs were excluded from the primer sequence (SNP track in UCSC Browser). RT-qPCR was performed using Q-PCR mastermix Plus for SYBR Green I without UNG (Eurogentec, Liege, Belgium) according the manufacturer’s instructions. Each reaction was performed in duplicate in a final volume of 15 µl containing 6-30 ng/µl genomic DNA, 1.25 µM forward/reverse primer, and 7.5 µl SYBR Green mastermix.
Control populations

The first control population comprised 722 unrelated individuals from Belgium who had been referred for clinical genetic testing for hemochromatosis or cystic fibrosis. Genomic DNA of each individual was extracted from blood lymphocytes according to standard procedures, and assayed by RT-qPCR using primerpair 3 for detection of copy number changes of 16p13.1. Thresholds were set at a fold difference of 0.8 and 1.3. All samples surpassing these thresholds were analysed twice to confirm the presence of copy number changes.

A second control population, comprising 960 unrelated Caucasian adults (age 40-70 years) from the USA, were genotyped using the HumanHap300 Genotyping BeadChips (Illumina, San Diego, California), comprising ~317,000 HapMap SNPs distributed throughout the genome. Each individual was enrolled in the PARC study (http://www.pharmgkb.org/network/members/parc.jsp#team) which aims to identify genetic contributors to the variable efficacy of statin drugs on cardiovascular disease risk. Hybridizations, data analysis and copy number analysis, focused on this region of 16p, were performed according to published protocols [17].

Case reports

Patients carrying 1.65 Mb 16p13.11 deletions

Case 1: This adult patient is the only affected member of 5 siblings. She has severe mental retardation, therapy resistant epilepsy and behaviour problems. She has short stature (143 cm = <P3; P3 = 155 cm) and microcephaly (OFC 51 cm; P3 = 52.2 cm). She has a short nose, smooth philtrum, wide mouth and fine palpebral fissures. She has difficulty in expressive language and has an ataxic gait.

Case 2: The adult proband and his brother were referred with severe mental retardation. The proband has short stature (150 cm; P3 = 168 cm), microcephaly (OFC 51 cm; P3 = 52.2 cm), a pectus excavatum and spasticity of the extremities. He is treated for epilepsy.

Case 3: This adult man with moderate mental retardation is the only child to a non-consanguineous couple. His stature and OFC are normal and he is obese (103 kg; >P97). He is not dysmorphic, extremely talkative and displays intermittent verbal aggression and self-mutilation.

Case 4: The dizygotic twin pregnancy occurred via in vitro fertilization with intracytoplasmic sperm injection (ICSI). At 16 weeks gestation, ultrasound examination demonstrated an abnormal female fetus. The pregnancy went to term. The male fetus has a VSD, a right-sided aorta and frequent infections as a result of a low CD3 count – he has a 22q11 deletion. The 1856 g cyclopic female fetus had a crown rump length of 28.5cm and a head circumference of 26 cm (proportionally microcephalic). Holoprosencephaly, agenesis of the nose, a midline upper lip notch and a median cleft palate were present together with 2 pre-auricular tags on the left, and a right dysplastic ear with a pre-auricular tag and an atretic auditory canal.
Case 5: Prenatal ultrasound at 18 wks gestation showed a cleft lip on the right, a possible intracranial abnormality with dilated lateral ventricles and thinning of the cortical mantle. Following termination at 21 weeks, autopsy showed post-hemorrhagic hydrocephalus with marked ventriculomegaly, cortical thinning, a hypoplastic falx cerebri, cleft lip on right side, 2 preauricular skin tags on the right and cleft T1 and T3 vertebral bodies. Physical growth parameters were consistent with gestational age.

Patients carrying 1.65 Mb 16p13.11 duplications

Case 6: The second child to a non-consanguineous couple, he was born after an uneventful pregnancy. He has moderate mental retardation. He was constipated from birth and at 6 months diagnosed with Hirschsprung disease, for which he had a resection and an end-to-end anastomosis. Behaviour problems developed at 5 years of age and he has followed special education. He is adult, has a normal biometry, bushy eyebrows and a mandibular overbite.

Case 7: This patient was originally reported by Kriek et al.[9] in a MLPA screen of 105 patients selected for developmental delay and/or congenital malformations. He has learning problems in primary school, but is not clearly dysmorphic and has no serious behaviour problems. The extent of the duplication was verified using array CGH, and FISH. MLPA and RT-qPCR of the parents confirmed the de novo nature as well as the size of this duplication.

Case 8: The patient has severe learning disabilities with limited use of language, poor vocabulary and repetitive speech. She displays challenging, agitated behaviour marked by shouting, hand clapping, kicking, hitting and throwing objects at people, although this has improved with age. She has epilepsy.

Case 9: The patient is the only child of non-consanguineous parents. He walked at 13 months, and later on was described as clumsy but not dyspraxic. Speech has always been problematic. His delay was moderate (IQ 55), and he had behavioural problems such as increased impulsivity and limited attention control. He had grommets inserted for upper airway infections, but otherwise there were no major physical problems. At 14 ¾ yr, OFC was 54.7cm (50th centile). He was noted to have large simple ears, thick lips, a large tongue, large and somewhat puffy hands with small nails. CT-scan of the brain was normal. Metabolic screen of urine showed no abnormalities. EEG was normal. Ophthalmologic examination showed bilateral astigmatism. FRAXA testing showed a full expansion in the patient and a premutation in his mother, which had been inherited from her father.

Case 10: An ultrasound examination at 20 weeks of the pregnancy had revealed microcephaly. This was confirmed after delivery and at 4 years of age the OFC was 44cm (P3 = 48.6 cm). She had brachycephaly, telecanthus, abnormal eyebrows, deep set eyes, epicanthic folds, a pinched nasal tip, prominent nose and small jaw. She also had a VSD, an umbilical hernia, deep creases on the palms and soles, and clinodactyly. An MRI scan showed a small brain with delayed myelination and prominent extra axial fluid filled spaces. She has delayed speech and a hyperactive behaviour with
aggressive episodes, and at 6 years of age, attends normal school with much support. Her mother also has a small head, mild learning difficulties and poor temper control. Her mother, (the proband’s maternal grandmother), has a small head and mild learning difficulties, whereas one developmentally and behaviourally normal maternal aunt carries the duplication and one aunt with mild learning difficulties and aggressive behaviour is cytogenetically normal.

Patients carrying larger atypical 16p13.11 rearrangements

*Case 11:* The only child of a non-consanguineous couple, she was born at term with a birth weight of 3200g. She had feeding and respiratory difficulties in the neonatal period. This adult has an asymmetric face with a left facial nerve paresis, a short neck with reduced mobility, bilateral epicanthal folds, strabismus, choroid colobomata and had an atresia of the right choana. An atrial septal defect type 2 closed spontaneously and she has a unique right kidney with double ureters. She has hearing difficulties and a CT scan showed aplasia of the semicircular canals and abnormal middle ear bones. Her behaviour is normal. *CHD7* mutation analysis was normal.

*Case 12:* This patient was reported in Sharp *et al.*[3]

*Case 13:* This sixth child to a pre-eclamptic mother weighed 1800 g at full-term. She had neonatal seizures and marked developmental delay, walking and talking only at 4 years of age. This adult has short stature (150 cm; P3 = 155 cm) and microcephaly (OFC 50.5 cm; P3 = 52.2 cm). She has an IQ of 38 and suffers from cyclic depression.

**Results**

**Recurrent microdeletion/duplication of 16p13.11**

Array CGH was performed on 1027 patients or fetuses with unexplained mental retardation (MR) and/or multiple congenital anomalies (MCA) using one of two custom BAC arrays [10, 13]. In total, 13 patients with 16p12-p13 rearrangements were detected (1.3%), 6 of whom carried a deletion and 7 a duplication. Ten of these rearrangements (5 deletions and 5 duplications, Patients 1-10) appeared to involve the same ~1.65 Mb region of 16p13.11, suggesting that they likely represent the reciprocal events of NAHR (Figure 1). Three additional patients carried atypical larger rearrangements. These were up to 3.4 Mb in size, and either overlapped or flanked the more common 1.65 Mb 16p13.11 deletions/duplications seen in the other 10 patients. RT-qPCR was used to confirm the results of the array CGH analysis, and clearly distinguished patients with deletions and duplications from normal control individuals (Figure 2). For one patient with a typical deletion (Case 4), testing of parental DNA showed that the microdeletion was inherited from the phenotypically normal father. By contrast, for one patient with a typical duplication (Case 10), testing revealed three generations in which the duplication initially appeared to segregate with a mild/more severely abnormal phenotype, but testing of additional family members showed the duplication and phenotype did not co-segregate.
Genotype-phenotype correlations

Figure 1: Molecular overview of recurrent deletions and reciprocal duplications in cases 1 to 10. (A) Ensembl overview (freeze: 24-04-2007) which visualises 1 Mb BAC clones, the genes involved in the imbalance, the copy number variable regions [19] and chromosome bands. (B) The location of RT-qPCR primers used to finemap the breakpoints, are depicted. (C) The extent of the deletion and duplication is shown by red and green bars, respectively. (D) Organization of the segmental duplication structure at the distal and proximal breakpoints of recurrent 16p13.1 rearrangements. Each colored bar represents a pairwise alignment with > 98% identity.

Rearrangement breakpoints localise to LCR16 segmental duplication blocks containing a positively selected gene family

Each rearrangement was mapped at higher resolution using either Affymetrix 262K NspI SNP arrays (Figure 3) or a custom Nimblegen oligonucleotide array (Figure 4). This data confirmed the results of BAC array CGH, and localised the breakpoints of each rearrangement to segmental duplications composed of large clusters of LCR16 [7]. Significantly, 10 cases (Patients 1-10) showed both proximal and distal breakpoints which localised to the same intervals (distal breakpoints, 14.7-14.75 Mb, proximal breakpoints 16.3-16.77 Mb). These duplication blocks which define the common breakpoint regions were also observed to be highly polymorphic in copy number in normal control individuals (Figure 4). These polymorphic breakpoint regions correspond to LCR16a/LCR16b motifs, which have been shown to have undergone rapid proliferation during primate evolution and contain a gene family, called Morpheus, which shows a signature of extreme positive selection [18]. RT-qPCR using primers flanking these segmental duplications confirmed that that the breakpoints occurred within these LCRs, summarised in Figure 2.
**Figure 2**: RT-qPCR results with different primer sets flanking the breakpoint regions. (A) In all individuals carrying the common deletion/duplication (patients 1-10), the distal breakpoints occur between primers pp1 and pp2, while the proximal breakpoint is defined by primers pp4 and pp5. (B) The extent of the atypical duplication and deletion identified in patients 11 and 13 were delineated by primers pp2 and pp3 and pp4 and pp5, respectively, for distal breakpoints and between pp6 and pp7 for the proximal breakpoint.

The presence of 2 copies for a locus is defined by a fold difference of 1 whereas a fold difference of 0.5 or 1.5 corresponds to a deletion or a duplication, respectively.

**Figure 3**: Detection of 16p12-p13 imbalances defined by the 262k NspI SNP array. Data from the common 16p13.11 1.65 Mb rearrangements is shown. Each plot shows physical probe position on 16p (x-axis) against probe intensity ratio (y-axis). Red shading corresponds to the common deleted region, and green shading indicates the duplication observed in cases 1,2,3,4, 6 and 7.

There are 15 RefSeq genes in the common 1.65 Mb recurrent rearrangement region. Given that microcephaly and brain malformations are the major recurrent characteristic in our deletion patients, NDE1 was believed to be a good candidate dosage-sensitive gene which might underlie the features of these deletions. In order to test the hypothesis that the deletion acts by unmasking the presence of a recessive mutation on the remaining allele, we sequenced the NDE1 gene in 4 of our deletion...
patients (Cases 1-4). Primers covering all 8 exons and the 5' and 3' UTRs were used for direct sequencing of genomic DNA, but no mutations were detected.

**The microdeletion associates with the MR/MCA phenotype**

Although neither 16p13.11 duplications or deletions have been detected in 210 unrelated HapMap individuals [19] nor in 122 other individuals sampled from the normal population [10, 20, 21], a larger and population matched sample size was needed in order to draw a statistically meaningful conclusion as to their pathogenic significance. Therefore, the copy number of the commonly rearranged 16p13.1 region in 722 population matched control individuals ascertained from the Belgian population, and in a further 960 Caucasian controls drawn from the USA, were evaluated. These analyses did not detect any deletions of this region in the 1682 controls tested, but did reveal the presence of 5 16p13.11 duplications. Although we do not have access to detailed clinical information for these 5 control patients, their method of ascertainment suggests that they are unlikely to have significant developmental abnormalities, However, we cannot rule out the possibility that they might have abnormal learning, memory or behavioural characteristics. Combining this data with previously published analysis of 332 control individuals results in a total control population of 2014, including 5 carriers of 16p13.1 duplications, but no deletions. This compares to 5 deletions and 7 duplications ascertained from 1027 MR/MCA patients. We therefore conclude that deletions of 16p13.1 are significantly associated with patient phenotype (p=0.0048, Pearson Chi square test with simulated p-value (based on 10000 replicas)), but as demonstrated by their presence in some apparently unaffected relatives, show incomplete penetrance. The incidence of duplications in association with disease is not statistically significantly different from that in controls (p=0.1273, Pearson’s Chi-squared test). This observation leads us to a number of possible conclusions regarding the duplication including (i) that it might be truly clinically benign or (ii) that it might be compatible with a phenotype that has passed unnoticed (or has not been excluded) in the control population tested or (iii) that it might work in combination with other predisposing factors to give an MR phenotype.

**Atypical rearrangements of chromosome 16p12-p13**

In addition to the common recurrent 1.65 Mb microdeletion/duplication of 16p13.11 observed in 10 patients, 3 atypical rearrangements were detected. Two patients (cases 11 and 12) carried a larger duplication of ~3.4 Mb in size, overlapping the typical duplication. The distal breakpoint of these atypical duplications is located between 15.0-15.4 Mb and the proximal breakpoint located within a third LCR16 cluster (18.3-18.4 Mb) (Figure 4). RT-qPCR with primer pair 2 is normal while RT-qPCR with primer pair 3 shows the presence of a duplication (Figure 2). This result confirms that the duplication starts in a more proximal LCR16 as compared with the typical duplication. This duplicated region includes 12 RefSeq genes. In patient 11, the normal mother was found to be a carrier of this same duplication.
Figure 4: High-resolution oligonucleotide array mapping of seven 16p12.3-p13.11 rearrangements. 4 of seven cases show a common distal breakpoint (14.7-14.75 Mb). For those four patients with the common 1.65 Mb rearrangement (represented by red shading), the proximal breakpoints also map to a second LCR16 cluster (16.3-16.77 Mb). 3 additional patients show an atypical rearrangement. Patients 11 and 12 show an atypical larger duplication, with the distal breakpoint between 15.0-15.4 Mb and the proximal breakpoint located within a third LCR16 cluster (18.3-18.4 Mb). Patient 13 shows an atypical deletion with proximal breakpoint in the third LCR16 cluster and distal breakpoint in the second cluster. Data from normal control individuals shows significant copy number polymorphism of the LCR16 clusters which define these three breakpoint regions. Note that the high degree of homology between these LCR16s also results in false-positive signals from probes which are identical to those within the true deletion/duplication in Patients 5, 8 and 9. The image shows a 5 Mb region of 16p12-p13 (chr16:14,000,000-19,000,000). For each individual, deviations of probe log2 ratios from zero are depicted by grey/black lines, with those exceeding a threshold of 1.5 standard deviations from the mean probe ratio coloured green and red to represent relative gains and losses, respectively. Segmental duplications of increasing similarity (90-98%, 98-99%, and >99%) are represented by grey/yellow/orange bars, respectively.

One further patient (case 13) carried a 1.6-2.1 Mb deletion which flanked the common rearrangement region. This atypical deletion has distal breakpoint in the second cluster and the proximal breakpoint in the third LCR16 cluster (Figure 4). The region includes 2 RefSeq genes. This imbalance was inherited from a phenotypically normal mother. A summary of all thirteen rearrangements detected, is shown in table 1.
Genotype-phenotype correlations
<table>
<thead>
<tr>
<th>Patient (Deciphercode)</th>
<th>Phenotype</th>
<th>Pattern of Inheritance</th>
<th>Type of imbalance</th>
<th>Distal breakpoint (Mb)</th>
<th>Proximal breakpoint (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (CHG00002371)</td>
<td>Severe mental retardation, IQ=38, short stature (143cm), microcephaly (51 cm), epilepsy, ataxia</td>
<td>Parents not available for testing</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>2 (CHG00002372)</td>
<td>Severe mental retardation, short stature (150 cm), microcephaly (51 cm), epilepsy, pectus excavatum, spastic extremities</td>
<td>Parents not available for testing</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>3 (CHG00002374-</td>
<td>Moderate mental retardation, normal stature (183 cm), normal head circumference (55.3 cm), behaviour problems</td>
<td>Parents not available for testing</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>4 (CHG00001230)</td>
<td>Fetus at term, autopsy showed holoprosencephaly, nose agenesis, midline upper lip notch, midline cleft palate, dysplastic external ear and atretic auditory canal on right, preauricular skin tags bilaterally, relative microcephaly</td>
<td>Parents not available for testing</td>
<td>Duplication</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>5 (CHG00001046)</td>
<td>Moderate mental retardation, normal stature (176 cm), normal head circumference (57.4 cm), behaviour problems, Hirschsprung</td>
<td>Mother does not carry duplication</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>6 (CHG00001046)</td>
<td>Moderate mental retardation, normal stature (176 cm), normal head circumference (57.4 cm), behaviour problems, Hirschsprung</td>
<td>Father unavailable for testing</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>7 (LEI00002370)</td>
<td>Severe mental retardation, mild developmental delay, learning disabilities. Originally reported by Kriek et al.[9]</td>
<td>Severe learning disabilities with limited use of language, poor vocabulary and repetitive speech. Challenging, agitated behaviour marked by shouting, hand clapping, kicking, hitting and throwing objects at people, although this has improved with age, epilepsy.</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>8 (LEI00002370)</td>
<td>Moderate mental retardation, behavioural problems – increased impulsivity, limited attention span. Large simple ears, thick lips, large tongue, large puffy hands and small nails. Also has an expansion of the FRAXA triplet repeat. Microcephaly detected at 20 weeks gestation. At 4 years of age OFC 44cm (P3=48.6cm). Brachycephaly, telecanthus, abnormal eyebrows, deep set eyes, epicantthic folds, pinched nasal tip, prominent nose, small jaw, VSD, umbilical hernia, deep palmar and plantar creases, speech delay, hyperactive behaviour with aggressive episodes. Mother has small head (no OFC).</td>
<td>Parents not available for testing</td>
<td>Duplication</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>9 (LEI00002370)</td>
<td>Moderate mental retardation, behavioural problems – increased impulsivity, limited attention span. Large simple ears, thick lips, large tongue, large puffy hands and small nails. Also has an expansion of the FRAXA triplet repeat. Microcephaly detected at 20 weeks gestation. At 4 years of age OFC 44cm (P3=48.6cm). Brachycephaly, telecanthus, abnormal eyebrows, deep set eyes, epicantthic folds, pinched nasal tip, prominent nose, small jaw, VSD, umbilical hernia, deep palmar and plantar creases, speech delay, hyperactive behaviour with aggressive episodes. Mother has small head (no OFC).</td>
<td>Parents not available for testing</td>
<td>Duplication</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>10 (CHG00000993)</td>
<td>Mental retardation, multiple congenital anomalies. Originally reported by Sharp et al.[10]</td>
<td>Phenotypically normal mother carries duplication</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>11 (CHG00000993)</td>
<td>Feeding and respiratory problems as neonate. Asymmetric face with left facial nerve paresis, short neck with reduced mobility, bilateral epicanthic folds, strabismus, choroid colobomata, atresia of right choana, ASDII, unique right kidney with double ureters, aplasia of semicircular canals, abnormal middle ear bones</td>
<td>Phenotypically normal mother carries duplication</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>12 (CHG00002373)</td>
<td>Mental retardation, multiple congenital anomalies. Originally reported by Sharp et al.[10]</td>
<td>Parents not available for testing</td>
<td>Duplication</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
<tr>
<td>13 (CHG00002373)</td>
<td>Marked developmental delay, IQ=38, short stature (150cm), microcephaly (50.5cm), neonatal seizures</td>
<td>Phenotypically normal mother carries deletion</td>
<td>Deletion</td>
<td>14.7-14.75</td>
<td>16.3-16.77</td>
</tr>
</tbody>
</table>
**Discussion**

We describe 5 patients carrying identical 1.65 Mb deletions of 16p13.11 encompassing 15 genes. In addition, we detected 5 patients carrying apparently reciprocal duplications of this same region. The rearrangement breakpoints are located in low copy repeats implying that non-allelic homologous recombination between these flanking LCRs mediates these rearrangements. Three further patients presenting with larger atypical rearrangements showed breakpoints which also mapped to clusters of LCR16.

Despite careful parental analysis, observations of rearrangements of 16p13.11 make it difficult to distinguish between disease causing events (generally presumed to be de novo) and benign variants (that do not contribute to a phenotype). Two deletions and one duplication were found to be inherited from an apparently normal parent while another duplication was inherited from a clinically mildly abnormal parent. Conversely, we also observed one duplication that occurred de novo. Previous reports have shown 16p13.11 deletions to have occurred both de novo and inherited from a normal parent, while duplications were reported to be inherited in 2 families from normal parents [11, 22, 23].

One interpretation of these observations could be that rearrangements of this region are benign variants, and that the observed phenotypes are co- incidental with the presence of the imbalance. However, a case-control study shows a significant higher incidence of the del16p13.11 in MR/MCA patients implying that the del16p13.11 is a risk factor contributing to the MR/MCA phenotype. The duplication is present in equal frequency in the normal and the patient population indicating this variant is compatible with a normal phenotype. The identification of this dup16p13.11 in a fragile X patient (case 9) lends support to this hypothesis. Recently, Ullmann et al. [11] reported 3 dup 16p13.11 carriers in a cohort of 182 autistic individuals. Since the duplication was detected in 5 out of 2014 normal individuals in our study, the duplication appears overrepresented in the autistic populations (p=0.023, Fisher exact test). More studies are needed strengthen this association.

Three of the four adult patients with 16p13.11 deletions had both microcephaly and seizures, whilst the 2 fetuses with the deletions had brain anomalies, 1 of which also had relative microcephaly. Short stature is also present in the same 3 adult patients. Small stature and dysmorphic features were also a feature of one of the patients described by Ullman et al. [11] In contrast, a fourth of our adult patients and 1 deletion carrier described previously [22] had a normal head circumference (55.3 cm) while 1 patient was reported with macrocephaly [11].

While the typical 16p13.11 duplication may be a benign variant, it is striking that 4 of the duplication carriers not only presented with mental retardation but also behaviour problems. Similarly, members of 2/3 families with duplications reported by Ullman et al.[11] had similar behaviour problems. Although we do not have behaviour phenotypes of all MR/MCA patients tested using the arrays, the occurrence of this type of behaviour in this patient population seems higher than average. Hence, it seems plausible that dup16p13.11 carriers may have a predisposition for aggressive behaviour.
Thus far, most genomic imbalances are classified as either benign or pathogenic and most microdeletion syndromes are presumed to be well defined clinical entities. However, even well known genomic disorders can be phenotypically heterogeneous and more variable than originally thought due to incomplete penetrance or variable expression. The variability of the del22q11 phenotype originally led to their different clinical classification as DiGeorge syndrome (heart and thymus defects), Sprintzen syndrome (speech difficulties) or VCFS syndrome (conotruncal anomaly face) (MIM 188400). Recently, several reports have appeared on atypical patients with 22q11 deletions [24, 25]. Equally, in dup22q11 carriers the phenotype may range from severe mental retardation through to completely unaffected, and different minor developmental anomalies are noted [26, 27]. Also 22q11 duplications can be inherited from apparently normal individuals [26, 28]. Therefore, it is possible that 16p13.11 duplications might also be causative and the heterogeneous phenotype of our patients explained in part by (1) the unbiased selection criteria (2) variability due to other genetic or environmental determinants, (3) incomplete penetrance, (4) variable expression or (5) unmasking of recessive alleles [29]. In particular due to the presence of del16p13.11 in a normal parent of one of our probands, we set out to test the latter hypothesis. Since microcephaly was observed in 2 of the 3 adult typical deletion patients and in 1 of the 2 deletion fetuses, NDE1 was considered an excellent candidate gene for this phenotype. NDE1 is highly expressed in brain [30] forms complexes with LIS1, a dosage-sensitive gene which is crucial for neuronal migration and cerebral development, and which is known to underlie the Miller-Dieker lissencephaly syndrome (MIM #247200). Furthermore, Nde1-null mice show microcephaly [31]. However, sequencing of all exons of NDE1 in 4 deletion patients did not reveal any mutation on the remaining allele, suggesting that this is not the mechanism responsible for the phenotype in our patients. Another plausible candidate gene is NTAN1 (asparagine-specific N-terminal amidase). Mouse models deficient for this enzyme showed altered activity, social behaviour and memory [32]. However although this region is known as copy number variable in the Redon database (http://projects.tcag.ca/variation/), learning and memory defects and aberrant social behaviour could not be excluded among controls and therefore NTAN1 gene remains a good candidate gene.

Imprinting is a mechanism that could potentially explain the presence of these rearrangements in unaffected relatives [11]. However, there are no known imprinted genes on chromosome 16 (http://www.geneimprint.com/site/gensess-by-species). Furthermore, for both deletions and duplications of 16p we observed the inheritance of these imbalances from normal parents through both the maternal and paternal germ lines, indicating that imprinting does not significantly influence patient phenotype of this disorder.

In addition to the common 1.65 Mb rearrangement observed in the 10 patients reported here, 3 “atypical” chromosomal imbalances either overlapping or flanking this common region were detected in patients with MR/MCA. Breakpoints for all 3 imbalances were also located within LCR16 sequences. However, the atypical 3.4 Mb duplication appears to be mediated by different LCR16s while the atypical deletion is mediated by a third LCR16. Therefore, the complex structure of the LCR16s in this region appears to be involved in generating a variety of different chromosomal rearrangements. The
finding of different sized rearrangements on chromosome 16p is similar to that observed for other recurrent genomic disorders, such as the Prader-Willi/Angelman syndrome [33], Smith-Magenis syndrome [34], and the 15q24 deletion syndrome [35], where recombination within alternate LCRs can result in recurrent deletions and duplications of different size.

In conclusion, we report here a novel genomic disorder likely caused by NAHR between copies of LCR16. Although in some cases inherited from a normal parent, we demonstrate a strong association of the deletion with developmental disorders. Reciprocal duplications were observed as both inherited and de novo events, and were also identified in several controls, suggesting that the duplication by itself confers either no phenotype at all or a range of phenotypes of varying severity. Alternatively, the duplication may require additional predisposing factors to have a phenotypic effect. Our findings have important implications for genetic counseling. Traditionally, chromosomal imbalances inherited from a normal parent were considered benign, while de novo chromosomal imbalances were considered pathogenic. While our results suggest that the inherited 16p13.11 deletion is likely causal for the phenotype, the clinical significance of both de novo and inherited duplications remains uncertain and they may be benign variants. The study of additional patients and normal individuals with 16p13.11 rearrangements is required to reinforce this hypothesis and to obtain a better insight in the potential pathology associated with the observed microdeletion and microduplication events.

Acknowledgements

We are grateful to Dr. Ronald Krauss and the Pharmacogenomics, Risk of Cardiovascular disease project (PARC) for the use and analysis of Illumina SNP genotyping data, the MicroArray Facility, Flanders Interuniversity Institute for Biotechnology (VIB) for their help in the spotting of the arrays and the Mapping Core and Map Finishing groups of the Wellcome Trust Sanger Institute for the initial clone supply and verification. We also thank Dr Abdur Khan (Northamptonshire NHS Healthcare Trust, UK) and Dr Angela Barnicoat (Great Ormond Street Hospital for Children NHS Trust, UK) for their clinical contribution.

This work was supported in part by grants from the NIH (HD043569, EEE), Merck Research Laboratories (AJS), Oxford Genetics Knowledge Park and the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the Department of Health’s NIHR Biomedical Research Centres funding scheme (the views expressed are those of the authors and not necessarily those of the Department of Health (RR and SJLK), the Catholic University of Leuven (GOA/2006/12 and Centre of Excellence SymBioSys, Research Council K.U.Leuven EF/05/007). EEE is an Investigator of the Howard Hughes Medical Institute.
References


Paper 2: Genotype-phenotype correlation in 21 patients with Wolf-Hirschhorn syndrome using high-resolution array CGH

Nicole M.C. Maas, Griet Van Buggenhout, Femke Hannes, Bernard Thienpont, Damien Sanlaville, Klaas Kok, Alina Midro, Joris Andrieux, Britt-Marie Anderlid, Jacqueline Schoumans, Roel Hordijk, Koenraad Devriendt, Jean-Pierre Fryns, Joris R. Vermeesch

Journal of Medical Genetics. 2008; 45(2): 71-80

My contribution to this work has led to the molecular characterization and/or confirmation of chromosomal rearrangements found in nearly all patients discussed here.
Abstract

The Wolf-Hirschhorn syndrome (WHS) is usually caused by terminal deletions of the short arm of chromosome 4 and is phenotypically defined by growth and mental retardation, seizures, and specific craniofacial manifestations. Large variation is observed in phenotypic expression of these features. In order to compare the phenotype with the genotype, we localized the breakpoints of the 4pter aberrations using a chromosome 4 specific tiling BAC/PAC array. In total, DNA from 21 patients was analyzed, of which 8 had a cytogenetic visible and 13 a submicroscopic deletion. In addition to classical terminal deletions sized between 1.9 and 30 Mb, we observed the smallest terminal deletion (1.4 Mb) ever reported in a patient with mild WHS stigmata. In addition, we identified and mapped interstitial deletions in four patients. This study positions the genes causing microcephaly, intrauterine and postnatal growth retardation between 0.3 and 1.4 Mb and further refines the regions causing CHD, CL/P, oligodontia, and hypospadias.
**Introduction**

Wolf-Hirschhorn syndrome (WHS), first described by Wolf et al. [1] and Hirschhorn et al. [2], is usually caused by a partial deletion of the short arm of chromosome 4. It is a well known syndrome with growth and mental retardation, microcephaly, seizures, 'Greek helmet' facies, and major malformations such as cleft lip and/or palate (CL/P), coloboma of the eye, congenital heart defects (CHD), and hypospadias. Large variation is observed in phenotypic expression. These features and especially the facial appearance change with ageing [3-6]. The Pitt-Roger-Danks syndrome was described as the milder end of the clinical spectrum of the WHS [7]. Although WHS affects around 1 per 50,000 births, it is suspected that the syndrome is more frequent, because not all patients can be diagnosed with a standard chromosomal investigation [7]. In case of a clinical suspicion of WHS in a patient with normal chromosomes, additional fluorescence in situ hybridization (FISH) studies of the subtelomeres and the WHS critical region (WHSCR) are usually performed.

Different genes probably play an important role in the complex phenotype of WHS. One approach to understand the role of different genes is to compare the phenotypes of patients with differently sized deletions and correlating the genotype with the phenotype [4, 8-13]. Earlier genotype-phenotype correlation studies identified a WHSCR of approximately 165 kb. This WHSCR covers the entire WHS candidate gene 2 (WHSC2), and part of the WHS candidate gene 1 (WHSC1) [14, 15]. In a recent study, this 165 kb WHSCR was not deleted in one patient with a typical WHS face which prompted the investigators to assign a new critical region of WHS, 'WHSCR2'. This region includes the LETM1 and partially the WHSC1, but not the WHSC2 gene [10]. We reported on five mild WHS patients with small deletions of chromosome 4p covering or flanking the WHSCR which also pinpointed the WHSC1 as the main candidate for causing the facial WHS appearance [12].

Genotype-phenotype correlation studies suggest that hemizygosity of genes other than WHSC1 in the region contribute to some of the phenotypic aspects. An overview of the genes in the region and their potential contribution to the WHS phenotype was described before [13]. CL/P, CHD, renal abnormalities and severe mental retardation are rare in patients with a small terminal microdeletion, but are common in patients with a larger deletion extending to proximal of 4.4 Mb [9]. However, which of these genes is responsible for which phenotype remains unknown, with the exception of LETM1 which is the most likely candidate gene for epilepsy in WHS patients [16].

In order to advance phenotype-genotype correlation efforts, we present genotype-phenotype correlations of 8 previously reported and 13 new patients with WHS phenotypic features and characterized the location and size of the deletions by full tiling chromosome 4 BAC array CGH for all 21 patients. Previous genotype-phenotype correlations of this region have been hampered by the presence of other imbalances in part of the WHS patients which likely confused some of the correlations. In this study, only patients with pure 4p deletion were included, which enables us to further refine the 4p phenotypic map.
**Materials and methods**

**Patients**
All 21 patients were diagnosed and clinically examined by clinical geneticists of Brussels and Leuven (Belgium), Groningen (The Netherlands), Stockholm (Sweden), and Paris and Lille (France). A personal and family history was obtained from each patient. Most patients were regularly reevaluated during several years and follow up data could be recorded. The summary of the clinical features for each patient is provided in Tables 1 and 2. The phenotypic data of the five patients with clear WHS stigmata but in whom no 4p deletion was detected by means of cytogenetic, array CGH, and FISH studies are not included in the tables. Patient 5, 6, 8-11, 20 and 21 have been described elsewhere [12, 17-21].

**Cytogenetic analysis**
Karyotyping was performed on metaphase spreads prepared from peripheral blood lymphocytes by routine standard cytogenetic procedures [22].

**Fluorescence in situ hybridization (FISH)**
The deletion for all but two patients was confirmed using the commercial LSI WHS region/CEP 4 control (dual color) (Abbott Inc., Downers Grove, IL, USA). The WHSCR probe covering the entire 165 kb WHSCR between loci D4S166 and D4S3327 (Cytocell technologies Ltd., Oxford, UK), was used to analyze the metaphase spreads of patient 2. The metaphase spreads of patient 1 have been tested with probe D4S96 (Oncor, Parsippany, NJ USA) located at ~1.2 Mb from the telomere. Locus specific BAC or PACs were Spectrum Orange labeled as described before [22].

**Array CGH**
Genome wide array CGH at 1 Mb resolution was performed as described before [22]. A chromosome 4 tiling BAC array containing 1903 targets was generated as described elsewhere [23]. Cy5 labeled patient DNA was co-hybridized versus Cy3 labeled reference DNA of a healthy individual. The fluorescence intensities measured were first background subtracted.
Normalization was performed by dividing each log₂ transformed intensity ratio by the mean of the log₂ transformed intensity ratios of all targets derived from the long arm of chromosome 4. In about 1/3 of the hybridizations, an additional 2D normalization was performed using Bioconductor. (http://www.bioconductor.org) [24, 25]. If successive clones have intensity values below 4x SD of all intensity ratios, the region is considered deleted. Hybridization efficiencies of the chromosome 4 tiling array were around 97 %. The average standard deviation of the log₂ intensity ratio per experiment was 0.08.
Results

Genotype-phenotype correlation
In this study, we aim to correlate the WHS phenotypes with 4p deletion sizes and therefore excluded patients carrying unbalanced translocations or other chromosomal rearrangements which may influence the phenotypic features [11]. The presence of unbalanced translocations was excluded by subtelomeric FISH for patients 5, 6, 10, 11, and 21 [12, 19], and both subtelomeric and interstitial chromosomal rearrangements were excluded by 1 Mb array CGH in the other patients. Thus, only those patients with pure 4p deletions were retained in this study. Subsequently, DNA from patients was hybridized on a chromosome 4 tiling array.

Figure 1: Results of the array CGH analysis of two patients. The Y axis represents the log₂ of the intensity ratios of the combined dye swap experiments of patient versus control DNA. In the X axis the spotted clones are ordered from the 4p telomere to the centromere. The distance from the 4p telomere is indicated in Mb. The arrows indicate the location of the breakpoints and the distance from the telomere is indicated. (A) DNA from a patient with a small terminal deletion. (B) DNA from a patient with a cytogenetically visible interstitial deletion.
## Genotype-phenotype correlations

### Table 1 Clinical findings in 9 WHS patients with de novo cytogenetically visible deletions

<table>
<thead>
<tr>
<th>Case:</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (female/male)</td>
<td>f</td>
<td>f</td>
<td>m</td>
<td>m</td>
<td>m</td>
<td>F</td>
<td>f</td>
<td>f</td>
<td>5/3</td>
</tr>
<tr>
<td>Facial appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek Warrior Helmet facies</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
</tr>
<tr>
<td>Hypertelorism</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8/8</td>
</tr>
<tr>
<td>Prom glabella</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
</tr>
<tr>
<td>High forehead</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8/8</td>
</tr>
<tr>
<td>epicanthus</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>5/8</td>
</tr>
<tr>
<td>Broad nose</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>6/8</td>
</tr>
<tr>
<td>Beaked nose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4/8</td>
</tr>
<tr>
<td>Short philtrum</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8/8</td>
</tr>
<tr>
<td>Oligodontia</td>
<td>n.i.</td>
<td>+</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>too young</td>
<td>too young</td>
<td>1/2</td>
</tr>
<tr>
<td>Micrognathia</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8/8</td>
</tr>
<tr>
<td>Downturned shape mouth</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>8/8</td>
</tr>
<tr>
<td>Cleft lip/palate</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>-/</td>
<td>2/8</td>
</tr>
<tr>
<td>Colobomata of iris</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3/8</td>
</tr>
<tr>
<td>Narrow lacr ducts</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
<td>2/4</td>
</tr>
<tr>
<td>Strabismus</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
</tr>
<tr>
<td>Dysplastic ears</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
</tr>
<tr>
<td>Preaur tag/pit</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2/8</td>
</tr>
<tr>
<td>Birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational weeks at birth (weeks,days)</td>
<td>33</td>
<td>38</td>
<td>38</td>
<td>40.2</td>
<td>38</td>
<td>37</td>
<td>41.5</td>
<td>37.2</td>
<td>Ø 38</td>
</tr>
<tr>
<td>Weight (g)</td>
<td>1180 (&lt;P10)</td>
<td>1630 (&lt;P10)</td>
<td>1900 (&lt;P10)</td>
<td>3660 (P75-P90)</td>
<td>1930 (&lt;P10)</td>
<td>1500 (&lt;P10)</td>
<td>2610 (&lt;P10)</td>
<td>1550 (&lt;P10)</td>
<td>Ø 1995</td>
</tr>
<tr>
<td>Length in cm</td>
<td>39 (&lt;P10)</td>
<td>42 (&lt;P10)</td>
<td>49 (P50-P75)</td>
<td>52 (P50-P75)</td>
<td>47.5 (P25-P50)</td>
<td>48 (P10-P25)</td>
<td>42 (&lt;P10)</td>
<td>Ø 46</td>
<td></td>
</tr>
<tr>
<td>IUGR</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
</tr>
<tr>
<td>OFC (cm)</td>
<td>27 (&lt;P10)</td>
<td>30.5 (&lt;P10)</td>
<td>30 (&lt;P10)</td>
<td>4 weeks: 36 (P10-P25)</td>
<td>34.5 (P75-P90)</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>Ø 30.5</td>
</tr>
<tr>
<td>microcephaly at birth</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>+</td>
<td>5/7</td>
</tr>
<tr>
<td>Postnatal examination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Age at examination</strong></td>
<td>6</td>
<td>12</td>
<td>7</td>
<td>3.6</td>
<td>5</td>
<td>50</td>
<td>1.6</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td><strong>height (cm)</strong></td>
<td>89 (&lt;P3)</td>
<td>142 (P10-P25)</td>
<td>&lt;P3</td>
<td>9 (-1SD)</td>
<td>&lt;P3</td>
<td>120 (&lt;P3)</td>
<td>72.5 (&lt;P3)</td>
<td>47.5 (&lt;P3)</td>
<td></td>
</tr>
<tr>
<td><strong>weight (kg)</strong></td>
<td>8.0 (&lt;P3)</td>
<td>30 (P10)</td>
<td>&lt;P3</td>
<td>12.6 (-2SD)</td>
<td>&lt;P3</td>
<td>25 (&lt;P3)</td>
<td>6.9 (&lt;P3)</td>
<td>2.5 (&lt;P3)</td>
<td></td>
</tr>
<tr>
<td><strong>OFC (cm)/ microcephaly (+/-)</strong></td>
<td>44.2 (&lt;P3)</td>
<td>48 (&lt;P3)</td>
<td>+</td>
<td>47.5 (-2SD)</td>
<td>43 (&lt;P3)</td>
<td>46 (&lt;P3)</td>
<td>42.2 (&lt;P3)</td>
<td>35 (&lt;P3)</td>
<td></td>
</tr>
<tr>
<td><strong>Mental retardation</strong></td>
<td>severe</td>
<td>severe</td>
<td>+</td>
<td>IQ 74</td>
<td>severe</td>
<td>Profound</td>
<td>profound</td>
<td>profound</td>
<td></td>
</tr>
<tr>
<td><strong>Sitting without support [y,m]</strong></td>
<td>3.6</td>
<td>1.5</td>
<td>3.6</td>
<td>1.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Walking without support [y,m]</strong></td>
<td>-</td>
<td>3.5</td>
<td>6</td>
<td>3.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3/8</td>
<td></td>
</tr>
<tr>
<td><strong>Speech [y,m]</strong></td>
<td>-</td>
<td>-</td>
<td>7</td>
<td>1.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>hypotonia</strong></td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>6/8</td>
<td></td>
</tr>
<tr>
<td><strong>Seizures</strong></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>7/8</td>
<td></td>
</tr>
<tr>
<td><strong>Congenital heart defects</strong></td>
<td>+</td>
<td>(ASDII, Pst)</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>(AVSD)*</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Clubbing of fingers/toes</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0/8</td>
<td></td>
</tr>
<tr>
<td><strong>Scoliosis</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>2/8</td>
<td></td>
</tr>
<tr>
<td><strong>Club feet</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>2/8</td>
<td></td>
</tr>
<tr>
<td><strong>Renal anomalies</strong></td>
<td>renal reflux</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Genital anomalies</strong></td>
<td>-</td>
<td>n.i.</td>
<td>hypospadias</td>
<td>+*</td>
<td>hypospadias</td>
<td>-</td>
<td>+</td>
<td>ventrally spaced anus</td>
<td></td>
</tr>
<tr>
<td><strong>Sacral dimple</strong></td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5/7</td>
<td></td>
</tr>
<tr>
<td><strong>deletion (start-end in Mb)</strong></td>
<td>8.8</td>
<td>1.8-10.1</td>
<td>10.9</td>
<td>2.7-14.8</td>
<td>14.8</td>
<td>16.5</td>
<td>19.5</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis by</strong></td>
<td>G-banding</td>
<td>1Mb array (Spectral Genomics)</td>
<td>G-banding</td>
<td>G-banding</td>
<td>G-banding</td>
<td>G-banding</td>
<td>G-banding</td>
<td>G-banding</td>
<td></td>
</tr>
<tr>
<td><strong>Subtelomeric screening by</strong></td>
<td>1 Mb array</td>
<td>interstitial deletion</td>
<td>1 Mb array</td>
<td>interstitial deletion</td>
<td>1 Mb array</td>
<td>1 Mb array</td>
<td>FISH subtel</td>
<td>1 Mb array</td>
<td></td>
</tr>
<tr>
<td><strong>Last normal clone (interstitial) (Mb)</strong></td>
<td>RP11-138D6 (1.7-1.9)</td>
<td>RP11-444J4 (2.6-2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Genotype-phenotype correlations

<table>
<thead>
<tr>
<th>First deleted clone (interstitial) (Mb)</th>
<th>RP11-1170P16 (1.8-1.9)</th>
<th>RP11-372F2 (2.9-3.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last deleted clone (Mb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-689P11 (8.5-8.7)</td>
<td>RP11-572M24 (10.1-10.2)</td>
<td>RP11-518.6 (10.9-11.1)</td>
</tr>
<tr>
<td>First normal clone (Mb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP11-637J21 (8.6-8.8)</td>
<td>RP11-61G19 (10.2-10.4)</td>
<td>RP11-775H15 (11.0-11.2)</td>
</tr>
<tr>
<td>Cytogenetic band</td>
<td>4p16.1</td>
<td>4p16.1</td>
</tr>
</tbody>
</table>

n.i.: not investigated; n.r.: not reported; f: female; m: male; Ø: average; ASD: atrial septal defect; VSD: ventricular septal defect; Pst: pulmonic stenosis. * see clinical description.
Of eight patients with cytogenetically visible terminal 4p deletions, six carried a terminal 4p deletion. Patient 17 carried an interstitial deletion 1 Mb proximal to the WHSCR1 from 2.7 to 14.8 Mb (Figure 1), and patient 15 had an interstitial deletion spanning between 1.8 and 10.1 Mb from the telomere. An overview of the genotypes and phenotypes of the eight patients is shown in table 1 and figure 3A. Hypertelorism, prominent glabella, high forehead, short philtrum, typical down turned or carp shaped mouth, and microcephaly were present in all patients. The typical facial features, intra-uterine growth retardation (IUGR) and seizures were present in all patients except for patient 17 (2.7 to 14.8 Mb interstitial deletion). In two of these eight patients a CL/P was observed (patient 20 and 21), with terminal deletions of respectively 19.5 Mb and 37 Mb. Strabismus was present in five of the seven investigated patients (not in patient 15 (1.8 to 10.1 Mb deletion) and 20 (0 to 19.5 Mb deletion)). Colobomata of the iris were observed in three patients with deletions of at least 10.9 Mb from the telomere. Two patients (patients 18 and 21 with respectively a 14.8 and 37 Mb terminal deletion) were found to have narrow lacrimal ducts. A pre-auricular tag or pit was observed in patients with deletions of 14.8 Mb and larger. All patients had mental, postnatal growth, and psychomotor retardation. Of six patients for which renal ultrasound was performed, a renal defect was only found in one patient (patient 14 with a 8.8 Mb deletion). Fifteen of the seventeen investigated patients had sacral dimples. All three boys with a large deletion presented with genital defects. Patients 16 and 18 had hypospadias (10.9 and 14.8 Mb terminal deletions), and patient 17 had a normal migrated left testis and a non-descended right testis (2.7 to 14.8 Mb interstitial deletion). Hypotonia was present in six of eight individuals except patient 15 and 20 (1.8-10.1 and 0-19.5 Mb deletion). An overview of all patients with microscopic visible deletions is shown in Figure 2.

![Figure 2: Results of the detected deletions using the full tiling array CGH of chromosome 4 in 8 WHS patients with microscopically visible deletions. The bars show the sizes of the deletions.](image)
Table 2. Clinical findings in 13 WHS patients with *de novo* microdeletions

<table>
<thead>
<tr>
<th>Case:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>m</td>
<td>m</td>
<td>f</td>
<td>m</td>
<td>f</td>
<td>m</td>
</tr>
<tr>
<td>Facial appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek Warrior Helmet facies</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hypertelorism</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Prom giabella</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>High forehead</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
</tr>
<tr>
<td>epicanthus</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Broad nose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Beaked nose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Short philtrum</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Oligodontia</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
</tr>
<tr>
<td>Micrognathia</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Down mouth</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Cleft lip/palate</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>+/-</td>
<td>n.i.</td>
<td>+/-</td>
</tr>
<tr>
<td>Colobomata of iris</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Narrow lacr ducts</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Strabism</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dysplastic ears</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Preaur tag/pit</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Birth

<table>
<thead>
<tr>
<th>Gestational weeks at birth</th>
<th>40</th>
<th>36</th>
<th>n.i.</th>
<th>39</th>
<th>37</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (g)</td>
<td>2080</td>
<td>1570</td>
<td>n.i.</td>
<td>2640</td>
<td>2270</td>
<td>2050</td>
</tr>
<tr>
<td>(P3)</td>
<td>(&lt;P3)</td>
<td>(&lt;P3)</td>
<td></td>
<td>(&lt;P3)</td>
<td></td>
<td>(&lt;P3)</td>
</tr>
<tr>
<td>Length (cm)</td>
<td>47</td>
<td>n.i.</td>
<td>n.i.</td>
<td>47.5</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>IUGR</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>OFC (cm)</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>33</td>
<td>29</td>
<td>0y1m: 34cm</td>
</tr>
<tr>
<td>(P25-P50)</td>
<td></td>
<td></td>
<td></td>
<td>(&lt;P10)</td>
<td>(&lt;P10)</td>
<td>(&lt;P3)</td>
</tr>
<tr>
<td>Microcephaly at birth</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Neonatal feeding difficulties</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
</tr>
</tbody>
</table>

Postnatal examination

<table>
<thead>
<tr>
<th>Age at examination (y,m)</th>
<th>11,8</th>
<th>2,4</th>
<th>23</th>
<th>12</th>
<th>16</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (cm) / short stature (+/-)</td>
<td>+</td>
<td>87</td>
<td>n.i.</td>
<td>151</td>
<td>+</td>
<td>172</td>
</tr>
<tr>
<td>(P10-P25)</td>
<td>(P50-P75)</td>
<td></td>
<td>(P50-P75)</td>
<td></td>
<td>(P50-P75)</td>
<td></td>
</tr>
<tr>
<td>weight (kg) / low weight (+/-)</td>
<td>+</td>
<td>9,2</td>
<td>-</td>
<td>31</td>
<td>+</td>
<td>41,3</td>
</tr>
<tr>
<td>(&lt;P3)</td>
<td>(P10-P25)</td>
<td></td>
<td>(P10-P25)</td>
<td></td>
<td>(P10-P25)</td>
<td></td>
</tr>
<tr>
<td>OFC (cm) / microcephaly (+/-)</td>
<td>+</td>
<td>46.3</td>
<td>n.i.</td>
<td>+</td>
<td>50</td>
<td>(+P3)</td>
</tr>
<tr>
<td>(&lt;P3)</td>
<td>(P25-P50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental retardation</td>
<td>moderate</td>
<td>severe</td>
<td>severe</td>
<td>moderate</td>
<td>+</td>
<td>mild / moderate</td>
</tr>
</tbody>
</table>
### Facial appearance

<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>f</td>
<td>f</td>
<td>m</td>
<td>f</td>
<td>m</td>
<td>m</td>
<td>5f/8m</td>
</tr>
</tbody>
</table>

#### Birth

<table>
<thead>
<tr>
<th>Birth Averages</th>
<th>1800 (P50-P75)</th>
<th>2460 (P25)</th>
<th>2010 (&lt;P3)</th>
<th>3250 (P50)</th>
<th>1900 (P3-P10)</th>
<th>2650 (P75-P90)</th>
<th>1900 (&lt;P3)</th>
<th>Ø 2224</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Averages</td>
<td>42 (P50)</td>
<td>46 (&lt;P10)</td>
<td>44 (P25)</td>
<td>50 (P50-P75)</td>
<td>45 (&lt;P10)</td>
<td>49 (P90)</td>
<td>n.i.</td>
<td>Ø 46</td>
</tr>
<tr>
<td>Birth Averages</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>11/13</td>
</tr>
<tr>
<td>Birth Averages</td>
<td>30 (P75-P90)</td>
<td>32 (&lt;P10)</td>
<td>31 (&lt;P25)</td>
<td>36,2 (&gt;P90)</td>
<td>30 (&lt;P10)</td>
<td>n.i.</td>
<td>n.i.</td>
<td>Ø 31</td>
</tr>
<tr>
<td>Birth Averages</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>4/8</td>
</tr>
<tr>
<td>Birth Averages</td>
<td>+</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
<td>2/2</td>
</tr>
</tbody>
</table>

#### Postnatal examination

<table>
<thead>
<tr>
<th>Postnatal Averages</th>
<th>129 (&lt;P3)</th>
<th>128 (+/-3SD)</th>
<th>155 (P25-P50)</th>
<th>149 (&lt;P3)</th>
<th>153 (&lt;P3)</th>
<th>93 (&lt;P3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postnatal Averages</td>
<td>21 (-2SD)</td>
<td>23 (&lt;P3)</td>
<td>32,1 (P3-P10)</td>
<td>36,7 (&lt;P3)</td>
<td>-</td>
<td>11,5 (&lt;P3)</td>
</tr>
<tr>
<td>Postnatal Averages</td>
<td>48,5 (-2SD)</td>
<td>49 (&lt;P3)</td>
<td>53,4 (P25)</td>
<td>11y: 47,3</td>
<td>52 (&lt;P3)</td>
<td>44,8 (&lt;P3)</td>
</tr>
<tr>
<td>Postnatal Averages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Moderate: severe, severe, mild / moderate, moderate / severe, severe, moderate
### Table 2 continued

<table>
<thead>
<tr>
<th>Case:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting without support [y,m]</td>
<td>+</td>
<td>-</td>
<td>n.i.</td>
<td>0,1</td>
<td>n.i.</td>
<td>1,2</td>
</tr>
<tr>
<td>Walking without support [y,m]</td>
<td>1,6</td>
<td>-</td>
<td>n.i.</td>
<td>2,1</td>
<td>n.i.</td>
<td>1,6</td>
</tr>
<tr>
<td>Speech [y,m]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>5,5</td>
<td>n.i.</td>
<td>4</td>
</tr>
<tr>
<td>hypotonia</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Seizures</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Deafness</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Congenital heart defects</td>
<td>-</td>
<td>+ small VSD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clubbing of fingers/ toes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Scoliosis/hyperk</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>n.i.</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Club feet</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Renal anomalies</td>
<td>-</td>
<td>Leftsided kidney duplication</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
</tr>
<tr>
<td>Genital anomalies</td>
<td>-</td>
<td>Hypospadias</td>
<td>-</td>
<td>Hypospadias</td>
<td>n.i.</td>
<td>Hypospadias</td>
</tr>
<tr>
<td>Sacral dimple</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>extra</td>
<td>-</td>
<td>Hip dysplasia, nystagmus</td>
<td>-</td>
<td>Scoliosis</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

#### Delineation of the deletion

<table>
<thead>
<tr>
<th>Deletion (start-end in Mb)</th>
<th>1.4</th>
<th>1.9</th>
<th>2.0</th>
<th>2.2</th>
<th>2.3</th>
<th>1.1-2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis by means of</td>
<td>FISH Oncor WHS probe</td>
<td>FISH Cytocell WHS probe</td>
<td>FISH Vysis WHS probe</td>
<td>FISH Vysis WHS probe</td>
<td>FISH Vysis WHS probe</td>
<td>FISH Vysis WHS probe</td>
</tr>
<tr>
<td>Subtelomeric screening by</td>
<td>1 Mb array</td>
<td>1 Mb array</td>
<td>1 Mb array</td>
<td>1 Mb array</td>
<td>FISH subtel</td>
<td>FISH subtel</td>
</tr>
<tr>
<td>Last normal clone (Mb)</td>
<td>RP11-20I20 (1.1-1.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First deleted clone (Mb)</td>
<td>RP11-386I15 (12.1-1.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last deleted clone (Mb)</td>
<td>RP11-1398P2 (1.5-1.6)</td>
<td>RP11-318G6 (2.0-2.2)</td>
<td>RP11-478C1 (2.1-2.3)</td>
<td>RP11-478A6 (2.4-2.6)?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First normal clone (Mb)</td>
<td>RP11-1398P2 (1.5-1.6)</td>
<td>RP11-318G6 (2.0-2.2)</td>
<td>RP11-478C1 (2.1-2.3)</td>
<td>RP11-503N18 (2.3-2.5)</td>
<td>RP3-474M20 (2.6-2.7)?</td>
<td></td>
</tr>
<tr>
<td>Chr. 4 band</td>
<td>4p16.3</td>
<td>4p16.3</td>
<td>4p16.3</td>
<td>4p16.3</td>
<td>4p16.3</td>
<td>4p16.3p16.3</td>
</tr>
</tbody>
</table>

n.i.: not investigated; n.r.: not reported; f: female; m: male; Ø: average; OFC: occipitofrontal circumference; ASD: atrial septal defect; VSD: ventricular septal defect; Pst: pulmonic stenosis
<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>n.i.</td>
<td>5.0</td>
<td>+</td>
<td>0.9</td>
<td>n.i.</td>
<td>3</td>
<td>+</td>
<td>8/9</td>
</tr>
<tr>
<td>2.6</td>
<td>-</td>
<td>n.i.</td>
<td>1.3</td>
<td>n.i.</td>
<td>6</td>
<td>1.2</td>
<td>7/9</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>n.i.</td>
<td>+</td>
<td>+</td>
<td>4</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>n.i.</td>
<td>+</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>Pst</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>n.i.</td>
<td>+</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Vesico-urethral reflux</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>-</td>
</tr>
<tr>
<td>Hypospadias</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>n.i.</td>
<td>Clitoromegaly</td>
<td>Hypo- spadias</td>
</tr>
<tr>
<td>Kyphosis</td>
<td>-</td>
<td>n.i.</td>
<td>n.i.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Delineation of the deletion**

| 2.7 | 2.7 | 2.7 | 1.8-3.6 | 3.7 | 4.8 | 5.3 |
| FISH Vysis | FISH subtel. | FISH Vysis | FISH Vysis | FISH Vysis | FISH Vysis | FISH Vysis |
| WHS probe | Probe Vysis | WHS probe | WHS probe | WHS probe | WHS probe | WHS probe |
| 1 Mb array | FISH subtel | 1 Mb array | FISH subtel | 1 Mb array | 1 Mb array | 1 Mb array |
| RP11-1197E19 | (1.8-1.9) |
| RP11-262P20 | (1.8-2.1) |
| RP11-444J4 | (2.6-2.7) | RP11-444J4 | (2.6-2.7) | RP11-444J4 | (2.6-2.7) | RP11-357G3 | (3.2-3.4) | RP3-513G18 | (3.6-3.7) | RP11-359G4 | (4.7-4.9) | CTD-2265H21 | (5.3-5.5) |
| RP11-372F2 | (2.9-3.1) | RP11-372F2 | (2.9-3.1) | RP11-529E10 | (3.4-3.6) | RP11-489M13 | (4.0-4.2) | RP11-32619 | (4.8-5.0) | RP11-653L17 | (5.5-5.7) |
In 13 patients, a submicroscopic deletion was initially detected by FISH. In 11 of them, the deletion was observed by FISH using the locus specific probe detecting the WHSC1 deletion. However, in patient 1, the deletion was detected using the Oncor WHS probe located distally from WHSCR1. Subsequently, array CGH was performed to determine the deletion boundaries. This patient had a 1.4 Mb terminal deletion not including the WHSCR1, confirmed by FISH with the flanking clones RP11-1244E8 (1.35 to 1.42 Mb, deleted) and RP11-1398P2 (1.46 to 1.61 Mb, normal) and the deleted telomeric clone CTC-36P21. In patient 2, FISH using the Cytocell WHS probe presented a weak signal in one chromosome 4 and a strong signal in the other chromosome 4 in all metaphases, suggesting a partial deletion. Array CGH pinpointed the breakpoint between RP11-21I14 (1.91 to 2.10 Mb, deleted) and RP11-318G6 (2.00 to 2.19 Mb, normal) (Ensembl release 42, Dec 2006). In two patients (patient 6 and 10), small interstitial deletions were observed. An overview of the genotype-phenotype correlation of these 13 patients is shown in Table 2 and Figure 3. Patients with deletions of the WHSCR1 all present with the characteristic facial features, except for patient 3. Patient 1 with the 1.4 Mb terminal deletion not involving the WHSCR1 had mild clinical stigmata and no typical WHS face. Patient 10 with the 1.8 to 3.6 Mb interstitial deletion presented with the typical WHS face, but lacked most other WHS stigmata.

**Figure 3:** Results of the detected deletions using the full tiling array CGH of chromosome 4 in 13 WHS patients with submicroscopic deletions. The bars show the sizes of the deletions. The thick black bars show the locations of the candidate regions for each phenotypic feature. The asterisk stands for IUGR, growth retardation, epilepsy, microcephaly, short stature. The red bars show the locations of the critical regions for hypospadias and CL/P defined by Estabrooks et al. [26]
Discussion

Genotype-phenotype correlation studies may enable the identification of the role of the different 4p genes in the etiology of WHS. However, several issues remain to be resolved. The number of patients analyzed is still limited and the resolution at which the location of the breakpoints has thus far been analyzed is low. Second, most genotype-phenotype correlation studies are derived from patients with large cytogenetically visible deletions spanning several Mb. Third, since several of the clinical features of this syndrome have considerable variable expressivity or penetrance, the phenotypic characterization of many more WHS patients will be required to pinpoint the genes involved in these more rare aspects of the WHS phenotype. Fourth, the role of position effects due to the relocation of the telomeres is unknown. Finally, since WHS is a contiguous gene syndrome, some of the features seen in WHS could be a result from haploinsufficiency of more than one gene in the region. Unraveling the contribution of each of the genes in the region to these multigenic phenotypes provides a new challenge for geneticists.

This report presents the largest genotype-phenotype correlation analysis of WHS patients thus far. Using a chromosome 4 tiling path array, both interstitial and terminal 4p deletion breakpoints were fine mapped. In addition to the classical terminal deletions, sized between 1.9 and 30 Mb, we identified a 1.4 Mb terminal deletion, the smallest deletion ever to be reported in a patient with WHS phenotypic features. In addition, we identified and mapped four interstitial deletions. Both the patients lacking all classic WHS features and the atypical deletions advanced the dissection of the molecular features leading to the different WHS characteristics.

The main characteristic of WHS is the typical face, usually referred to as a ‘Greek warrior helmet face’. Previous studies hypothesized hemizygosity of the \textit{WHSC1} as the most likely cause of the facial phenotype [9, 12, 27]. This notion was confirmed by the present study. All, except in one patient with a \textit{WHSC1} deletion, did have the typical facial features. The patient with the \textit{WHSC1} deletion but without the typical face is 23 years old. Because of advancing age the facial features could have coarsened over time (patient 3, Figure 3C,D) [6, 28]. The two patients in this study without the facial WHS features do not have a deletion of \textit{WHSC1}: one patient had only a 1.4 Mb terminal deletion not covering the \textit{WHSC1} (patient 1, Figure 3A,B), while another patient had a large interstitial deletion upstream of the \textit{WHSC1} (patient 17, Figure 3E,F).
Figure 3: Facial features of 5 patients with WHS. A and B: patient 1 from frontal and aside. C and D: patient 3 from frontal and aside. E and F: patient 17 from frontal and aside. G: patient 15 from frontal.

The molecular features of WHSC1 suggest that the gene might function as a chromatin-remodeling enzyme because SET domains have been shown to function as histone methylases [29]. A deficiency in chromatin remodeling could deregulate the expression of a variety of genes and hence, lead to pleiotropic effects. Recent studies indicated that haploinsufficiency of other chromatin remodeling enzymes cause syndromatic phenotypes such as the Chromodomain helicase DNA-binding protein 7 (CHD7) gene, causing the CHARGE (Coloboma, Heart anomaly, Choanal atresia, Retardation, Genital and Ear Anomalies) syndrome [30], the Nipped-B-like (NIPBL) gene causing the Cornelia de Lange syndrome [31] and heterozygous mutations in the V-ha-ras Harvey rat sarcoma viral oncogene homolog (HRAS) gene causing the Costello syndrome [32]. Because of this possibility, we sequenced the WHSC1 gene in five WHS patients without 4p deletions (data not shown). Intriguingly, no mutations were detected in this study nor in previous reports [29]. Either mutations elsewhere in the genome can cause phenocopies of the WHS or WHSC1 is the wrong target gene and another gene in the region is key for the phenotype.

Patient 17, a boy with a large interstitial deletion ranging from 2.7 to 14.8 Mb has normal growth parameters, he had no seizures, and mental delay was mild with an IQ of 74. Sitting and walking without support and speech were only mildly delayed compared to the other patients with large deletions. In contrast, all patients with large 4pter deletions, including our 6 patients with large terminal deletions sized between 8.8 and 37 Mb present with pre- and postnatal growth retardation, the typical face, seizures, microcephaly, severe to profound mental and psychomotor retardation and muscular hypotonia (the latter except patient 20). Hence, haploinsufficiency of genes in the 2.7 Mb 4pter region causes these WHS main features.
Not only patient 17 (2.7-14.8 Mb deletion), but also patient 10 with an interstitial deletion (1.8-3.6 Mb deletion) has normal growth parameters. A patient with a normal height and a 191 kb deletion encompassing the WHSC1 and WHSC2 genes was described [14]. In contrast, the patient with a 1.4 Mb terminal deletion did present with short stature and IUGR. Therefore, this study locates a short stature candidate region in the terminal 1.4 Mb.

In our previous study, a healthy female with multiple miscarriages carried a 0.3 Mb terminal deletion [12]. Therefore, the 0.3 Mb terminal region is not a candidate region for the genotypic map. Previously, we speculated that the WHS associated microcephaly might be the result of a contiguous deletion involving at least two genes, one located in a 1.8 Mb terminal region and one between 2.2-2.5 Mb. This hypothesis resulted from patients carrying a 1.9 and 2.2 terminal deletion without microcephaly. This latter observation may also result from incomplete penetrance or genetic modifiers. The microcephaly in the patient with a 1.4 Mb terminal deletion as well as in the other three investigated patients with deletion sizes up to 2.3 Mb is however more consistent with a gene localized in this terminal 1.4 Mb region.

Hemizygosity of LETM1, an ubiquitous Ca\(^{2+}\) binding protein involved in Ca\(^{2+}\) homeostasis, located at 1.8 Mb from the 4p telomere has been suggested to cause seizures [10, 33]. However, the patient with an interstitial deletion encompassing LETM1 (patient 6) did not present with seizures while the patient with the 1.4 Mb terminal deletion did present with seizures. Therefore, another gene in this terminal region may cause the epilepsy.

Besides these main characteristics features, WHS is characterized by several minor features. Oligodontia occurred in 6 out of 9 investigated WHS patients in the present study. Nieminen et al. [34] showed that a deletion of the msh homeobox 1 gene (MSX1), located at 4.9 Mb, might underlie this defect. In contrast to this hypothesis, three patients in this study with deletions in the terminal 2.7 Mb or smaller had oligodontia. Since the carrier of a deletion from 1.1 to 2.5 Mb has oligodontia, the candidate gene has to be sought in this interval. Cleft lip and palate (CL/P) has been observed in 9/29 patients described in the literature [8, 9] and a candidate region for CL/P was mapped between 150 kb (D4F26) and 2.3 Mb (D4S43) [26]. We observed CL/P in a patient with a terminal microdeletion of 3.7 Mb, and in 2/8 patients with large deletions. We defined the critical region from 0.3 Mb to 3.7 Mb, but after interpretation of the results of Estabrooks et al. [26], this CL/P interval may be refined between 0.3 and 2.3 Mb. Hypospadias was detected in all boys but patient 1 (1.4 Mb deletion). Estabrooks et al. [26] mapped the critical region for hypospadias between D4S127 (3.0 Mb) and D4S10 (4.0 Mb). Hence our results are concordant with these previous findings. Estabrooks et al. [26] mapped the critical region for congenital heart defect (CHD) between locus D4S43 and D4S241 proximal from 2.3 Mb. Zollino et al. [9] and Wieczorek et al.[8] found CHD in only 13/19 patients with large deletions. In the present study, 4 out of 8 patients from both large and small deletions presented with a CHD when having a deletion of 3.7 to 14.8 Mb. We refined the region for heart defect proximal from 3.7 Mb.
Genotype-phenotype correlations

In conclusion, the WHS is a syndrome with a spectrum of phenotypic features, from very subtle and mild to a wide range of severe aberrations. The gene(s) causing the IUGR, postnatal growth retardation and microcephaly are localized in the 0.3 to 1.4 Mb 4pter region. It is conceivable that a single gene might cause all these features. Finally, this study refines the candidate regions for CHD, CL/P, oligodontia, and hypospadias. The phenotypic characterization of more WHS patients will be required to delineate regions involved in these more rare aspects of the WHS phenotype.

Acknowledgements

We wish to thank the Micro Array Facility, Flanders Interuniversity Institute for Biotechnology (VIB) for their help in the spotting of the arrays and Children’s Hospital Oakland Research Institute (CHORI) for the initial clone supply and verification. This work was made possible by grants OT/O2/40, GOA/2006/12 and Center of Excellence SymBioSys (Research Council K.U.Leuven EF/05/007), Catholic University of Leuven. We wish to thank Mrs. R. Thoelen for her technical support in FISH analyses and Drs. E. Chabchoub for his help in screening patients with the 1 Mb array.
References


[33] Endele S., Fuhry M., Pak S. J., Zabel B. U. and Winterpacht A. LETM1, a novel gene encoding a putative EF-hand Ca(2+)-binding protein, flanks the Wolf-Hirschhorn syndrome (WHS) critical region and is deleted in most WHS patients. Genomics. 1999;60(2):218-25

**Paper 3: Duplication of the Wolf-Hirschhorn syndrome critical region causes neurodevelopmental delay**

Femke Hannes§, Malgorzata Drozniewska§, Joris R. Vermeesch, Olga Haus

§ Both authors contributed equally to this manuscript.

*European Journal of Medical Genetics. 2010;53(3):136-40*
Abstract

Wolf-Hirschhorn syndrome (WHS) is caused by deletions on chromosome 4p and is clinically well defined. Genotype-phenotype correlations of patients with WHS point to a critical locus to be responsible for the main characteristics of this disorder. Submicroscopic duplications of this region, however, are not known. Here we report a patient with an interstitial 560 kb duplication overlapping this critical locus. The present case shows that not only deletions but also duplications of the Wolf-Hirshhorn critical region cause mental retardation and multiple congenital anomalies. Interestingly, the duplication phenotype overlaps partially with the deletion phenotype. However, his facial phenotype differs from the typical WHS gestalt.
Introduction

Terminal deletions of chromosome 4p cause the Wolf-Hirschhorn syndrome (WHS MIM 194190). WHS is characterised by severe to profound mental retardation, microcephaly, distinct facial gestalt, pre- and postnatal growth retardation and epilepsy. In addition, heart and kidney anomalies, cleft lip and palate, oligodontia, hypospadias, eye abnormalities and structural brain anomalies can contribute to the variability of the phenotype. Patients with WHS have variable sized deletions and the syndrome is considered a contiguous gene syndrome [1-3]. In order to unravel the contribution of the genes in the deletion regions to the disease phenotype, extensive genotype-phenotype correlations have been recently carried out. These analyses led to the identification of critical regions for facial gestalt and mental retardation, short stature, microcephaly, hypospadias, oligodontia and heart malformations [3-5]. The minimal critical region responsible for the typical facial gestalt of WHS was initially determined by the minimal region of overlap between the deletions of two WHS patients. This region is called the WHS critical region or WHSCR and contains two genes, \( \text{WHSC1} \) and \( \text{2} \) [6]. The exact contribution of those genes to the disease phenotype remains elusive. The smallest de novo deletion ever described in a patient with a very mild expression of WHS phenotype encompasses 191.5 kb and covers the WHSCR [7].

Trisomies of chromosome 4p are clinically defined by severe mental retardation associated with muscular hypertonia, pre- and postnatal growth retardation and a distinct craniofacial dysmorphology including a broad nose with depressed nasal bridge, microphthalmia, large and low set ears, pointed chin, choanal atresia, short neck, low frontal hairline. Trisomy 4p is a rare chromosomal disorder. It was first described 39 years ago by Wilson et al. 1970 and thus far, only a handful of reports have been published [8-10]. Previously, two patients have been described with duplications of the WHSCR however both are associated with terminal deletions [11, 12]. Those deletions will influence the phenotype. Though, no submicroscopic pure interstitial duplications covering the WHSCR have yet been described.

Here we report on a unique patient with a submicroscopic interstitial duplication of 560 kb involving the WHSCR.

Case report

The male propositus was the first child of healthy non-consanguineous parents without a family history of congenital malformations or development delay. The father had 3 healthy children from the previous marriage. The mother was 31 years old and the father 44 years at time of his birth. Pregnancy was normal and birth was at 42 weeks of gestation. Birth weight was 3.660 g and length 50 cm. He scored 10 at the Apgar scale. A malformation of right hand (shortening of proximal phalanges of fingers II and III of the right hand) and contractures of fingers III and IV bilaterally were noted at birth. There was a general hypotonia. At the age of 11 months he was hospitalized because of
suspicion of congenital encephalopathy with psychomotor retardation, glaucoma of the left eye and the presence of minor phenotypic features. At that time he could not sit nor turn aside and presented severe axial hypotonia, and speech delay was present. At the age of 14 months he was hospitalised for the second time. The neurologic examination showed the psychomotor development delay of 10 months and the ultrasound examination showed brain myelinisation delay.

He was referred for genetic counselling at the age of 15 months. His weight was < 11th centile. He presented with axial hypotonia, facial dysmorphic features (Figure 1) (bilateral long and narrow palpebral fissures, epicentral epicanthus, low set and dysmorphic ears), dental caries, high arched palate. He had a short neck, small hands and feet and a transverse palmar crease bilaterally. Clinical examination at the age of 23 months revealed an unproved general development. He started to speak at the age of 18 months and was not able to sit without support. He suffered from seizures. Ultrasound of the heart was unchanged. The TSH level was normal. Hypoacusis of the right ear was discovered. His mother described him as a friendly child with “laughing” behaviour.

Materials and methods

Metaphase chromosome analysis with G-banding was performed on peripheral blood lymphocytes according to standard procedures. Karyotype analysis was performed on metaphase spreads of the patient and both parents. For each individual 20 mitoses were analysed, 10 mitoses were studied in detail.

A whole painting probe for chromosome 4 (WCP4) (Q-Biogene, UK) was used following the manufactures protocol in order to investigate ectopic hybridisation. In addition, FISH analysis with the commercial Wolf-Hirschhorn syndrome probe covering the Critical Region (WHSCR) was used according the manufactures protocol (Q-biogene, UK). In total, 10 metaphases and 50 interphase nuclei of both the patient and his parents were analyzed.

A multicolour banding with partial painting probes specific for chromosome 4 was conducted with minor modifications according the manufacturer’s instructions (XCyte 4 MetaSystems GmbH, Alt lusheim, Germany). Metaphase spreads were prepared as described before [13]. Finally, the MetaSystems ISIS software was used to capture and process the differential labelled regions of chromosome 4.

A genome wide 1Mb BAC array was performed for the patient and his parents [14]. Data extraction and analysis was performed via basic statistic analysis in excel followed by 2D normalisation using Bioconductor. One single clone was called aberrant when passing the threshold of log(3/2)-2SD and two or more contiguous clones are called aberrant when passing the threshold of log(3/2)-4SD [15].
A homemade tiling array containing 1903 targets derived from chromosome 4 was performed as described before [4, 5]. DNA of the patient labelled in Cy5 was hybridised against that of a healthy control individual labelled in Cy3. Data-analysis was done similar to that described for the 1Mb BAC array.

**Results**

Chromosome analysis revealed the presence of a pericentric inversion of chromosome 4 in the patient (Figure 2A). Parental analysis showed normal karyotypes. To exclude a translocation, whole chromosome 4 painting FISH analysis was performed, but no ectopic hybridisation of chromosome 4 could be detected (Figure 2B).

![Figure 2: Results of different cytogenetic and molecular techniques used on material of the proband.](image)

(A) The karyotype shows the aberrant chromosome 4 indicated by an arrow. (B) A whole chromosome 4 paint. (C) FISH analysis with a commercial probe covering the WHSCR indicated in a red colour and a control probe indicated in a green colour. (D) A multicolour banding specific for chromosome 4. The arrows pinpoint to the inversion breakpoints. (E) Full tiling array CGH analysis on chromosome 4. The array points to the duplicated clones. The line indicates the location of the submicroscopic duplication.
Subsequently, FISH analysis with a specific probe covering the WHSCR showed three signals: one signal on the normal chromosome 4, one signal in the short arm and a signal in the long arm of chromosome 4. FISH analysis on metaphase spreads of both parents revealed one signal for each allele. Since one extra signal was detected on the q-arm, we suspected that the inversion breakpoint might coincide with this duplication (Figure 2C). To investigate the chromosome 4 rearrangement in more detail, we performed multicolour banding or mBAND analysis. The presence of an inversion at 4q22 was confirmed and the position of the inversion breakpoint further refined (Figure 2D). A 1Mb BAC array CGH was conducted with DNA of both parents and the patient and demonstrated a duplication of 3 clones (RP11-572O17, Cancer_1G10, RP11-1197E19) in the patient whereas both parents were normal for this locus, which indicates that the rearrangement occurred de novo. Subsequently, chromosome 4 full-tiling BAC array CGH was performed to further delineate the rearrangement. A submicroscopic interstitial duplication of 7 clones covering 560 kb including the WHSCR was detected (Figure 2E). The duplication spans from clone RP11-1398P2 to clone RP11-21114. The first normal clones flanking the duplication region are RP11-1244E8 and RP11-318G6. The karyotype is as follows 46,XY,inv(4)(p16.3q22). arr 4p16.3 (1,458,385-1,907,425) x3 dn.

Table 1: Comparison of the clinical features in two patients with submicroscopic deletions and one patient with a duplication covering the WHSCR.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aberration</td>
<td>WHSCR del</td>
<td>1.2-2.4Mb del</td>
<td>1.4-2.0Mb dup</td>
</tr>
<tr>
<td>Height</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Weight</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Speech delay</td>
<td>Yes</td>
<td>NR</td>
<td>Yes</td>
</tr>
<tr>
<td>IQ</td>
<td>72-74</td>
<td>68</td>
<td>NR</td>
</tr>
<tr>
<td>Developmental delay</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MRI</td>
<td>NR</td>
<td>Delayed myelinisation</td>
<td>Delayed myelinisation</td>
</tr>
<tr>
<td>Ears</td>
<td>NR</td>
<td>Low set</td>
<td>Low set</td>
</tr>
<tr>
<td>Seizures</td>
<td>NR</td>
<td>NR</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Discussion

Genotype-phenotype correlations in patients with WHS established that deletions of the WHSCR are causal for the main characteristics and pinpoint WHSC1 and LETM1 as the main dosage sensitive genes. The phenotypic consequences of duplications of these genes are not yet known. Here, we report on the identification of a submicroscopic de novo 560 kb interstitial duplication including the WHSCR. Interestingly, the duplication phenotype reveals an overlap with the deletion phenotype. Table 1 depicts the features that partially overlap in the deletion-duplication phenotype within the WHSCR. Within this table 1, clinical features of three patients (two patients with a deletion and a third patient referring to our case reported here) (Figure 3) are compared. It is known that hemizygosity of
the WHSCR result in neurological implications, however not necessarily leading to severe mental retardation [7]. In addition, speech development may also be impaired by the aberration of this locus. The duplication phenotype in our case corresponds to speech delay, low weight for height, seizures, a general delay in development and a facial dysmorphism with low set ears.

Taken together, both the deletion and duplication carriers of the WHSCR are characterized by similar features like developmental delay, low set ears, speech delay, low weight and a delay in myelinisation. Interestingly, most of the features above such as low weight for height, seizures, speech delay, developmental delay were reported for a case presented by Rosello and colleagues [12]. Those similarities demonstrate that the phenotypic features result from the duplication of the WHSCR rather than the deletion distally to the duplication.

Why the clinical features between the deletion- and the duplication phenotype overlap is unclear. Several hypotheses can be formulated. First, the gene itself can result in a similar clinical outcome either whether the gene is over or underexpressed. Also for 22q11 deletions/duplication a similar phenotype is observed: Mice over- and underexpressing Tbx1 result in phenotypes similar to that of Velocardiofacial syndrome or VCFS [16]. In addition, Zweier et al. [17] reported in patients with DGS gain of function mutations in TBX1 resulting in stabilization of the protein dimer. The increased transcriptional activity in those mutation carriers resembles the overexpression seen in duplication carriers [17, 18]. Second, the structural rearrangements of the region of interest can deregulate the expression of a principal gene flanking this region of interest. For example, Crolla and van Heyningen have described multiple rearrangements flanking the major disease gene for eye anomalies like aniridia, thereby altering the regulatory elements upstream of the PAX6 gene resulting in an inactivation position effect [19].
Since in this patient a pericentric inversion coincides with a duplication, it is not excluded that the pericentric inversion contributes to the patients phenotype by disrupting a gene at the inversion breakpoint.

Apparently balanced chromosomal rearrangements often hide cryptic aberrations that cause the patients phenotype. While it cannot be excluded that the inversion has interrupted a dosage sensitive gene, it seems likely that the duplication is the major cause of the phenotype. Several studies demonstrated that deletions rather than gene rearrangements can account for the abnormal phenotype in about 40% of patients carrying apparently balanced translocations [20-23]. Noticeably, all rearrangements were deletions but none described the co-occurrence of a duplication. Only one recent report describes a patient with a cryptic duplication on the affected chromosome, however this duplication was not directly associated with the breakpoint [24]. How the inversion lead to the occurrence of the duplication remains unclear.

In total, 10 refseq genes are residing within the duplication region. Two genes are of particular interest since they have been postulated to be responsible for the facial pattern, developmental delay and epilepsy seen in patients with WHS [25]. First, Wolf-Hirschhorn syndrome candidate 1 (\textit{WHSC1}) is involved in chromatine remodelling and regulating gene expression. It is ubiquitously expressed in early development. Data from genotype-phenotype correlations pinpoints this gene as the major candidate gene for facial- and neurodevelopment. However, the duplication of this region ends up in a totally different facial spectrum. This raises a question whether the \textit{WHSC1} gene is solely the major candidate gene for the facial development. Recent reports have shown that WHS patients with deletions flanking the \textit{WHSCR} can display the specific facial gestalt (Personal observations and [26, 27]). Those results indicate a) that position effects could play a role on the expression of the \textit{WHSC1} gene of \textit{WHSCR} flanking regions or b) that multiple genes, including the \textit{WHSC1} gene, cooperate together to result in a specific facial gestalt and thus hemizygosity of the \textit{WHSC1} gene alone is not sufficient to display the unique facial gestalt. Second, the leucine zipper EF-hand-containing transmembrane protein 1 or \textit{LETM1} gene is a human protein homologue of the yeast \textit{MDM38} gene involved in mitochondrial morphology, protein transport from the mitochondrial matrix and exchange of mitochondrial K+/H+ [28-30]. Moreover, \textit{LETM1} knockdown caused the disassembly of several complexes of the respiratory chains, leading to a decreased membrane potential [31]. Disruption of genes involved in regulation of membrane potential are known to cause epilepsies. Interestingly, our patient suffered from asphyctic seizures suggesting that overexpression of \textit{LETM1} also result in epilepsy.

Transforming acidic coiled-coil containing protein 3 or \textit{TACC3} is involved in regulating microtubule dynamics during cell division. Overexpression of \textit{TACC3} in HeLa cells resulted in partially destabilized microtubules, defects in chromosome alignment and finally resulting in mitotic arrest [32]. Thus far, several studies have highlighted the importance of microtubule dynamics during mitotic spindle assembly for proper neurogenesis in the cerebral cortex [33, 34]. Moreover, the severe neurodevelopment delay seen in this patient may be enhanced by the duplication of the \textit{TACC3} gene.
In order to elucidate the function of genes in disease development, the description of patients with submicroscopic, well characterised aberrations covering only one or a few genes can help to unravel the function. Hence, the description of this patient contributes to this functional annotation process of genes located in and flanking the WHSCR and their role in trisomy 4p syndrome. It seems likely that high resolution array CGH will soon reveal more patients with such imbalances, enabling a more accurate clinical description.

**Acknowledgments**

We thank the MicroArray Facility, Flanders Interuniversity Institute for Biotechnology (VIB) for their help in the spotting of the arrays. This work was supported in part by grants of the K.U.Leuven (GOA/2006/12) and the Centre of Excellence SymBioSys, Research Council K.U.Leuven EF/05/007).
References


Genotype-phenotype correlations


Paper 4: A familial microdeletion 600 kb proximally of the currently accepted WHSCR reveals a mild Wolf-Hirschhorn syndrome phenotype

Femke Hannes, Peter Hammond, Koenraad Devriendt, Oliver Quarrell, Jean-Pierre Fryns, Joris Robert Vermeesch

In preparation
Abstract

Wolf-Hirschhorn syndrome (WHS) is characterised by a plethora of manifestations most often including variable developmental delay, short stature, microcephaly, seizures and a distinctive face, and is molecularly defined by the deletion of a critical region. High resolution array CGH detected a maternally inherited 432 kb deletion located 600 kb proximal to the Wolf-Hirschhorn critical region (WHSCR) in a patient with mild WHS. 3D facial analysis using dense surface modelling confirmed mild facial characteristics of WHS in the index patient and partially in the mother. Three healthy siblings were shown to have no WHS facial characteristics. To determine whether long range position effects influenced gene expression within the WHSCR, we analysed quantitative PCR in EBV cell lines derived from patient lymphoblasts. This showed expression levels in the deletion region were halved whereas expression was normal within the WHS critical region as well as in other flanking genes. Sequencing of both candidate genes, WHSC1 and WHSC2, did not reveal any mutations. We hypothesize that either (1) this locus harbours regulatory sequences which affect gene expression in the WHSCR in a defined temporal and spatial developmental window or (2) that this locus is additive to deletions of WHSCR increasing the phenotypic expression. This observation illustrates the value of 3D facial analysis in genotype-phenotype correlations.
Introduction

Wolf-Hirschhorn Syndrome (WHS) is associated with deletions of chromosome 4p and is characterized by moderate to severe mental retardation, pre- and postnatal growth retardation, microcephaly, seizures, various organ malformations (e.g., cleft lip and palate) and characteristic facial features. Overall, there is high clinical variability among WHS patients [1-3]. Distinctive facial features constitute one of the major diagnostic markers of WHS. The characteristic facial gestalt of WHS patients, also referred to as the Greek warrior helmet facies, is defined by a broad forehead, high nasal bridge, prominent glabella, high arched eyebrows, protruding eyes, hypertelorism, short philtrum, downturned corners of the mouth, and micrognathia. Moreover, the facial features are more pronounced with severity of the global phenotype [4]. Age, sex and medication lead to additional variability of the facial gestalt.

WHS is most often caused by terminal deletions involving chromosome 4p16.3 and may extend as far as 4p14 [5]. Interstitial deletions are less frequently reported [6-9], but are of particular interest since they facilitate genotype-phenotype correlations and thus the search for causative genes. Early mapping efforts identified two different sized overlapping deletions defining the Wolf-Hirschhorn Syndrome Critical Region (WHSCR1). Further fine mapping resulted in a 165 kb critical region including two genes, WHSC1 and WHSC2 [10]. This region was suggested as responsible for at least two of the core clinical manifestations of WHS, the facial gestalt and developmental delay. Later, two additional patients with partial deletions of the WHSCR1 extending more distally defined a new critical region WHSCR2 [11, 12]. Overall, both critical regions 1 and 2 overlap and have one gene, WHSC1, in common. Thus far, nearly all of the reported deletions include WHSC1 [13, 14].

A recent study demonstrated growth retardation and various WHS-like midline defects in whsc1 haploinsufficient mice. The craniofacial defects similar to those seen in WHS were detected in some whsc1+/- mice. In addition, double heterozygous mutants, whsc1+/- and nkx2-5+/-, developed more heart malformations such as atrial and ventricular septal defects (ASD/VSD). Those results indicate that the deletion of WHSC1 alone is not sufficient to account for the entire phenotype, but is modified by additional genetic factors [15].

Here, we report on a patient clinically diagnosed with mild WHS features. Molecular karyotyping revealed a maternally inherited submicroscopic deletion 600 kb proximal to the currently accepted WHSCR. Objective 3D facial analysis confirmed a mild WHS facial gestalt in the patient and partially in his mother. Our results suggest that the WHSCR flanking sequence contributes directly or indirectly to the severity of WHS. Mutation analysis of both WHSC1 and WHSC2 was normal and we could not show a position-effect on gene-expression within the WHSCR nor the surrounding region in an EBV cell line.
**Case report**

The male proband is the second of four siblings. He was born at term after an uneventful pregnancy with birth weight 2.980g. He presented with feeding difficulties and at the age of 5 yrs 3 months was referred because of mildly delayed development and attention deficit disorder. He presented with several minor anomalies, such as hypertelorism, cutaneous webbing of the fingers, long fingers, and prominent eyes. He has a high pitched voice. His height then was 101 cm (-2.2 SD), weight 15 kg (-2.4 SD) and head circumference 51 cm (p42). At the age of 7 yrs 8 months, he entered special education for children with mild learning disability. He had developed several facial tics, and was taking Ritalin for attention deficit. More recent clinical examination revealed similar features as before. Height was 117.5 cm (p3), weight 19 kg (p3=20 kg) and OFC 53 cm (p75). His mother followed normal professional schooling but had attention problems. She has a similar facial phenotype to the proband. Her height was 155 cm, head OFC 53.3 cm. The siblings presented without learning difficulties, followed regular schooling and had a normal biometry.

**Material and methods**

**Cytogenetic and molecular analysis**

Fluorescence in situ hybridisation (FISH) with a commercial WHS probe (Cytocell) was performed on metaphase spreads of peripheral blood lymphocytes from the index patient and both parents. DNA was extracted from peripheral blood lymphocytes by the Qiagen DNA extraction kit (Qiagen NV, Venlo, The Netherlands). Array CGH was performed using a homemade chromosome 4 tiling BAC array as previously reported [9]. A custom-made 385k Nimblegen array containing 60-mer probes spaced every 20 bp was designed to target the region of chromosome 4p16 between 1.0 Mb to 11.8 Mb (UCSC hg18), excluding the highly repetitive regions. Labelling and hybridisation followed the instructions of the manufacturer (Nimblegen). Data extraction, analysis and visualisation were done with Nimblescan 2.5 and Signalmap 1.9 (Nimblegen) respectively. To confirm the results of both arrays and to check parental inheritance, FISH analysis was conducted on interphase nuclei of buccal smears of the patient, three siblings and the mother, and metaphase spreads derived from blood lymphocytes of the father. The clone RP11-201O13 (2.679.772 bp- 2.747.829 bp), located within the deletion region, was used as a probe for the FISH experiments.

**3D facial analysis**

3D face images of the proband and family members were collected using a commercial photogrammetric device and each was landmarked by one individual (PH) using 22 reproducible landmarks [16]. For the patient and siblings, a comparative image dataset of 200 controls and 100 individuals under 20 years with a genetically confirmed diagnosis of WHS was selected from an existing collection. 3D dense surface models (DSMs) of face shape have delineated common facial features in a range of neurodevelopmental conditions, often, in addition, establishing accurate
discriminating characteristics or assisting the determination of phenotype-genotype correlations [17]. For this younger group, 20 dense surface models (DSMs) were computed using stratified, randomly selected training sets of 90% of the combined dataset of 300 individuals. In each case, the remaining 10% was used for blinded discrimination between control and WHS using closest mean classification. Accuracy of classification was estimated as the mean of the areas under the 20 associated receiver operating characteristics (ROC) curves. This corresponds to the probability of correctly classifying a randomly chosen pair of faces, one from each of the control and WHS subgroups. A mean, blinded classification position and 95% confidence interval was computed for the proband and siblings using the DSMs from the 20-folded cross-validation. Finally, a single DSM was computed for the proband, siblings and 387 controls with a wide age range to enable a facial heat map comparison of the proband and each sibling with an age and sex matched mean suitably scaled where necessary.

Within the recruited WHS group, only 16 were over 16 years of age. This was too few for an adult-specific analysis of the mother along the lines for the patient and siblings where 100 WHS images were available. In particular, such a small number of WHS images undermines accurate DSM synthesis of dysmorphic facial features. To compensate, we built DSMs using all 108 WHS images of children and adults, as well as 387 controls. Then, within the models, the 40 oldest WHS patients (9.9 yrs to 32.3 yrs, mean 16.2 yrs) and 200 age matched controls (6.2 yrs to 33.2 yrs, mean 16.1 yrs) were used to analyse the mother's face.

To confirm the clinical diagnosis of mild WHS, 22 clinical geneticists had been asked which chromosomal deletion syndrome might best fit the patient’s phenotype. Three geneticists suggested WHS; others suggested Aarskog, velocardiofacial, Kabuki and Smith-Magenis syndromes. Because of these alternative diagnoses we tested the patient against images of children with a FISH-confirmed diagnosis of Smith-Magensis (n=85) and velocardiofacial (n=89) syndromes. No images were available for testing for facial features of Aarskog and Kabuki syndromes.

**Cell culture, RNA and cDNA preparations**

Epstein barr virus (EBV) transformed lymphoblasts derived from the patient and 10 control individuals were grown in DMEM/F-12 media supplemented with 10% fetal calf serum. Total RNA was prepared in parallel from logarithmic growth-phase cells with RNeasy Mini kit (Qiagen) in accordance with the manufacturer’s instructions. DNAse treatment (Roche) was performed and subsequently converted to cDNA with the use of Superscript III (Invitrogen) primed with a combination of both oligo d(T) (Invitrogen) and random primers (Invitrogen).

**Real-Time quantitative PCR (RT qPCR) and Data analysis**

Primers for RT qPCR were designed to span an intron boundary when possible using the Universal ProbeLibrary Assay Design Center (Roche Applied science). Primers containing SNPs or repetitive sequences were excluded using the dbSNP and blat tool from the UCSC browser. The runs were
performed in the Lightcycler 480 instrument I using the LightCycler® 480 SYBR Green I Master mix (Roche Applied science). In total, 22 genes located from the telomere of chromosome 4p to approximately 4.5 Mb were selected. Three genes, GUSB, CLK2 and ACTB were used for normalisation of the expression data. Final working volume of 20 µl contained 0.5 µM of primers, 10 µl SybrGreen I Master mix and 5 µl of template cDNA (in total 33.3 or 66.7 ng) and each sample was conducted in replicate. The complete experiment was repeated two times. The raw data was extracted from the Lightcycler 480 software and subsequently analysed and stored via qBase (qBase-user manual v1.3.5).

**Breakpoint spanning PCR and Sequencing**

Breakpoint spanning PCR was performed with both primers (Eurogentec) designed in normal copy number regions flanking the breakpoints. The amplification was conducted in a total volume of 50 µl including 200µM dNTPs, 250 nM of both primers, 1.5 mM MgCl2, 2.5 U platinum Taq polymerase (Invitrogen) and 100 ng template DNA. Following amplification program was used; one cycle of 5 min on 95°C; 30 cycles of 95°C - 30 s, 5X°C - 30 s, 72°C – 2 min; and one cycle of 72°C – 10 min. The product was subsequently Sanger sequenced using the BigDye terminator system. The resulting labeled fragments were size separated using an ABI 3130xl sequencer.

**Sequencing of WHSC1 and 2 gene**

DNA extracted from peripheral blood lymphocytes was used to sequence both WHS candidate genes. Primer design within the intronic region was done via Primer3 (http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html). PCR was optimised and the products were subsequently sequenced using the BigDye terminator system and size separated.

**Results**

Following clinical examination, mild WHS was suggested. FISH analysis using a commercial probe covering both WHSCRs on metaphase spreads of the patient was normal. Subsequent molecular karyotyping with a homemade chromosome 4 BAC tiling array detected a submicroscopic interstitial deletion 600 kb proximal to the WHSCR. The boundaries of the deletion were defined by two normal flanking clones, RP11-448F22 (2.384.345 bp) and RP11-355N4 (3.094.606 bp), and two deleted clones, RP11-677N20 (2.646.413 bp) and RP11-705L9 (2.996.839 bp). Array CGH using a targeted 385k targeted Nimblegen array redefined the deletion size to 432 kb (Figure 1A). Breakpoint-spanning PCR combined with sequence analysis revealed the exact location of both distal and proximal breakpoints at 2.605.128 bp and 3.036.901 bp, respectively (Human genome build 18). Both distal and proximal breakpoints are located within repeat sequences including a long interspersed nuclear element (LINE) at the distal site and a short interspersed nuclear element (SINE) at the proximal site. Interspersed between both breakpoints are 5 nucleotides of unknown origin, suggesting that the deletion was repaired by non-homologous-end-joining (NHEJ) (Figure 1C). Within this deletion region,
7 refseq genes are located including FAM193A, TNIP2, SH3BP2, ADD1, NOP14, GRK4, and MFSD10. This deletion has not been reported so far. Analysis of the database of genomic variants reveals three copy number variations in 2502 normal individuals partially underlying the deletion including two genes TNIP2 and SH3BP2 [18, 19]. No clinical details could be obtained regarding WHS-like facial morphology or other manifestations as seen in our patient. Therefore, we cannot exclude them from the set of candidate genes.

FISH using clone RP11-201O13, located within the deletion region, was hybridised on interphase nuclei of buccal smear from the mother and the four siblings (Figure 1B) as well as on metaphase spreads of both father and mother. FISH analysis confirmed a deletion in the patient. Interestingly, a deletion was detected in the nuclei of the mother who presented with equally mild WHS features. The phenotypically normal siblings and father were diploid for this locus. Taken together, these results indicate a maternal inheritance of the deletion.

![Image](image.png)

**Figure 1:** Overview of the cytogenetic and molecular data. A) The results of the high resolution oligonucleotide microarray of Nimblegen targeting a region of 1.0 Mb to 11.8Mb. Each dot represents a combination of 10 probes targeting a region of approximately 200 bp. The red box depicts the location of the WHSCR and the two other grey boxes indicate the locations of both olfactory receptor gene clusters (ORC). B) The red colour represents a clone RP11-201O13 and the green dot, a subtelomere 4q. C) Sequence analysis of the breakpoint junction.
Genotype-phenotype correlations

The 20-folded facial analysis of the younger dataset produced an estimated classification accuracy of 0.999 on a scale of [0.1]. The results for the proband and siblings are shown in Figure 2A. The background scatter is included solely to indicate typical classification positions and was not the referant population. It is computed from a single DSM for all 300 younger subjects. The proband’s face is classified unseen as more WHS-like than control-like. In contrast, his three siblings are classified as more-control-like than WHS-like. The single DSM based on controls (n=200) and individuals with one of three syndromes (WHS-100; SMS-85; VCFS-89) produced a three-way 2-D scatter (Figure 2B) that classifies the proband’s face unseen as more like WHS than the other two. A four-way control-WHS-SMS-VCFS closest mean comparison resulted in a 3D scatter animated in supplementary video S1. This classifies the patient unseen within the WHS group some distance from the SMS and VCFS clusters, and, the three siblings within the control cluster. Supplementary Figure S2 shows heat maps reflecting differences parallel to three orthogonal axes of the mean WHS face, the proband and three siblings compared to age/sex/ethnicity matched control means computed from 50 contiguously aged controls. The first column comparing WHS mean to a size-adjusted control mean highlights the characteristic features. The third, fourth and fifth columns don’t reveal any WHS-like features in the three siblings and corroborate their earlier quantitative closest mean classifications as control-like. The second column highlights the proband’s mild WHS-like features: relatively wide palpebral fissures and mild hypertelorism (red-green on exocanthi in A); upsweep to the supra-orbital region (blue on supraorbit in B), short and curled philtrum (blue on lip centre in B), prominent glabella (blue on forehead in C) and, broad nasal bridge (blue on nasal bridge in C).

Taken together, these results strongly suggest a causal relationship between the mild WHS facial features in the proband and the chromosome 4p deletion.

Within DSMs based on 108 WHS subjects and 387 controls, the face of the patient’s mother was compared unseen with a subset of 40 WHS patients and 200 controls. A closest mean classification based on 33 modes of the underlying DSM (corresponding to 99% variance coverage) classified her face on the periphery of the WHS and control clusters but slightly closer to the control mean. On inspection, her face synthesis did not accurately reflect the shape of her nose. In contrast, a DSM for just a nasal patch and the same population did and classified her nose shape as WHS with DSM modes corresponding to 99% variance coverage (supplementary Figure S3). Extending the number of modes derived from the face DSM by 3 to 99.1% variance coverage improved the accuracy of her face synthesis and moved her classification position marginally, still peripheral to both groups but now more WHS-like than control-like (Figure 2C). These results suggest that the mother has some WHS-like facial dysmorphism, especially around the nasal region. However, their sensitivity to small changes in variance coverage emphasises that the small number of available adult WHS images cannot support the same robust analysis completed for the patient. A more complete analysis will have to await additional adult WHS recruitment.
Figure legend: see next page
Genotype-phenotype correlations

Figure 2: Overview of the facial analysis. Classifications showing A) face of patient and sibs respectively as more WHS-like and more control-like; B) mother’s face as borderline control-WHS but slightly more WHS-like and C) patient’s face as more WHS-like than Smith-Magenis-like (SMS) or velocardiofacial-like (VCFS).

All previous evidence indicates that haploinsufficiency of WHSC1 and/or WHSC2 are essential to cause mild WHS, so the finding of a more proximal deletion is surprising. One explanation is the presence of a co-segregating intragenic mutation with the deletion in this family. A second possibility is a long range position-effect that alters gene-expression within the WHSCR1 and surrounding genes. Real-Time quantitative PCR was performed on cDNA extracted from EBV cultured lymphoblasts obtained from the patient and ten control individuals. For five out of seven hemizygous genes, the expression levels were stable in controls and expression was approximately half of that in controls (Figure 3). The expression levels of 13 genes located distal to the deletion and 4 genes located proximal to the deletion showed normal expression levels and were similar to those seen in controls. For one gene, TACC3, a minimal decrease in the expression value was noticed (0.76) (Figure 3). TACC3 expression in normal individuals remained within the threshold values (0.8 +/- 0.14) and thus the value in the index patient falls within the normal range. Hence, no long range position effect originating from the deletion region on either both WHSCR1-2 nor any other genes flanking the critical regions could be observed in EBV cell lines. Conversely, an effect of the WHSCR on gene expression in the deletion region was excluded by investigating a patient with a terminal 4p deletion encompassing the WHSCR but not the newly identified deletion.
To evaluate the possibility of a co-segregating intragenic mutation in the \textit{WHSC1} and/or 2 gene, we sequenced the complete coding sequence in the patient and a control individual. No mutations were detected. Moreover, the tiling resolution array excluded a small intragenic WHSCR1 and 2 deletion.

**Discussion**

Thus far, deletions in the WHS critical region are considered the molecular hallmark of WHS. Here, we identify a submicroscopic interstitial deletion 600 kb proximal to the WHSCR in a patient and his mother who both have mild WHS features. Since the clinical assessment of the facial gestalt is subjective, we applied dense surface modelling analysis which confirmed the WHS-like facial features. Several mechanisms can be envisioned on how this deletion might contribute to the WHS phenotype.
First, the deletion region might harbour control elements that regulate the expression of genes in WHSCRs. Disruption of cis-regulatory control elements of genes like PAX6, FOXL2 and SOX9, lead to aniridia [20], blepharophimosis syndrome (BPES) [21] and Pierre Robin sequence (PRS) [22], respectively. This deregulation is described on both sides of the gene and caused by disruptions 10 kb to 1.5 Mb away. We investigated similar long-range effects by expression analysis in EBV cell lines. This did not reveal any aberrant expression of genes within WHSCR1 and 2 and flanking genes approximately 600 kb away from the deletion. Another possibility is that loci within the WHSCR1 and WHSCR2 exert a long range effect on one or more genes in the newly identified deletion region and that in fact reduced expression of these genes explains the WHS phenotype. This would explain why mutations in WHSC1 or WHSC2 have never been observed in WHS-like patients [18, 23]. However, expression analysis in an EBV cell line with a terminal deletion including WHSCR1 and 2 did not reveal reduced expression of the genes in this novel deletion region. Since WHS is a developmental disorder, gene expression in lymphoblasts is not necessarily representative of the expression of those genes in the different tissues shaping the face during embryonal development. Therefore, we cannot exclude such a position-effect with certainty. To study this hypothesis further, animal models will be essential.

Another possibility is that the deletion could be fortuitous and another genetic cause is co-segregating with the phenotype. Since the most likely candidate genes to cause WHS are the WHSC1 and 2 genes, we sequenced both of them. This analysis did not reveal any mutation which is in line with previous sequencing efforts in WHS-like patients [18, 23].

Two genes, WHSC1 and LETM1, have been confirmed as being involved in the pathogenesis of WHS. WHSC1 is thought to cause developmental delay and the specific WHS facial gestalt, because, until recently, all patients were carriers of a partial or complete deletion of this particular gene [8]. However, a few reports have described patients who retain WHSC1, but who are suggestive for the facial phenotype [13, 14]. The deletions were all distally located from the WHSCR1. Taken together, this demonstrates that the hemizygosity of the WHSC1 gene alone is not sufficient and thus the interplay with other loci is necessary to express the complete distinct facial phenotype seen in WHS patients. There is already evidence for additive effects. Firstly, genotype-phenotype correlations have shown that manifestations such as microcephaly, cleft lip and palate and mental retardation are probably the result of haplo-insufficiency of more than one gene in the region [9]. Secondly, several genes located on chromosome 4p16 are known to interact. For example, TACC3 and CTBP1 both interact with the transcriptional regulator FOG-1 [24, 25] and SLBP and the NELF complex (a multi-subunit complex that includes WHSC2) are both players in the same pathway involving the translation of replication-dependent histones [26]. Taken together, those data provide evidence that WHS is in fact a true contiguous gene syndrome and that haplo-insufficiency of multiple genes is necessary for the presentation of the full blown WHS phenotype.
Deletion of either this region or the WHSCR leads to mild phenotypic effects, whereas deletion of both loci produces a more pronounced facial phenotype. The DSM based analysis identified mild WHS facial features in the patient and partial WHS features in the mother.

The deletion region itself contains 7 refseq genes. Thus far, none of those genes has been associated with the aetiology of WHS. One gene, *FAM193a*, is a hypothetical protein with unknown function. *ADD1* encodes for a cytoskeleton protein. Whereas *NOP14* is involved in processing of the pre-18S-rRNA fraction. *GRK4* encodes a G protein coupled receptor kinase, which is presumably linked to hypertension. *MFSD10* is a major facilitator superfamily of transporter proteins and likely plays a role in the efflux of organic anions. Mutations in *SH3BP2* are well known to cause cherubism and finally, *TNIP2* is an inhibitor of the NF-Kappa-B activation. The mouse phenotype derived from a null allele for *Tnip2* results in an impaired IL-1 response and macrophage physiology.

Genotype-phenotype correlation aims to improve gene identification involved in the pathogenesis of the genetic syndrome. However, due to the high phenotypic variability between the different WHS patients, genotype-phenotype correlations are not always informative. Therefore, unravelling the factors that contribute to the phenotypic variability is important for understanding the disease process and for patient counselling. One approach to pinpoint genes involved in biological variation is to study expression profiles of the loci of interest and their surroundings. Since no altered expression pattern could be detected in the patient, we wondered whether WHS patients who do carry at least a deletion of the *WHSC1* gene, but are nonhemizygous for the deletion of interest, have altered expression patterns of genes within the deletion of interest. Thus expression analysis was performed on cDNA of two additional WHS patients with terminal deletions of 2.1 Mb and 2.3 Mb in size (Figure 3). None of the genes within our deletion of interest showed an altered expression pattern compared to normal individuals. Hence, those results exclude regulatory potential between the two loci in both directions in EBV transformed lymphoblasts. However, those results do not exclude position-effects resulting in an altered spatial and temporal expression pattern in other cell types.

In conclusion, we report that a microdeletion 600 kb proximally of the WHSCR leads to a borderline WHS phenotype and propose that this locus predisposes to WHS when an extra mutation or aberration of the WHSCR is present.
References


deletion telomeric to the WHSCR and WHSCR 2 regions. Eur J Hum Genet. 2009;17(1):129-32


Supplementary video and figures

Supplementary Video S1: Video animating rotation of 3D scatter plot comparing patient and sibs unseen to controls and individuals with WHS and Smith-Magenis and velocardiofacial syndromes.

Supplementary Figure S2: Facial heat maps showing mild WHS facial features in patient and their absence in his sibs.

Supplementary Figure S3: Classification of nasal region of mother as more WHS-like than control-like.
Paper 5: Benign and pathogenic copy number variation on the short arm of chromosome 4

Femke Hannes and Joris Robert Vermeesch

Cytogenetics and Genome Research. 2008;123(1-4):88-93
Abstract

The terminal deletion of the short arm of chromosome 4 causing the Wolf-Hirschhorn syndrome is one of the first pathogenic copy number variations (CNVs) ever described. Since this first discovery, a large number of 4p CNVs causing variable phenotypes have been described. Here, we present an overview on those benign and pathogenic visible and submicroscopic 4p imbalances. Interestingly, some CNVs can be, dependent on their copy number state, both benign and pathogenic. In addition, we show how the collection of both phenotypes and genotypes of 4p terminal deletions is leading towards the genetic dissection of the Wolf-Hirschhorn syndrome.
The short arm of chromosome 4p is about 50 Mb in size and contains 379 genes. Early genome mapping efforts of this chromosome arm were triggered by two genetic disorders: on the one hand the search of the gene causing Huntington’s chorea, which has now been identified as huntingtin (HTT), located 3 Mb away from the telomere, and on the other hand, the study of Wolf-Hirschhorn syndrome (WHS), a syndrome characterized by terminal 4p deletions. By now, nine disease associated genes have been mapped to chromosome 4p (Table 1). More recently, with the advent of array CGH and high throughput sequencing technologies, also a large number of submicroscopic benign and pathogenic copy number variants are being uncovered.

In this review, we gather knowledge about pathogenic and benign CNVs. We first provide an overview of known pathogenic 4p imbalances and will subsequently focus on WHS. WHS stands as a model on how the study of variable sized deletions can help to unravel genetic factors contributing to contiguous gene syndrome phenotypes. Secondly, we critically review the recent information about benign CNVs in the chromosome 4pter region. Finally, we show that apparent benign CNVs can turn pathogenic.

**Table 1: Overview of known genetic diseases on 4p**

<table>
<thead>
<tr>
<th>Disease</th>
<th>MIM ID</th>
<th>Gene</th>
<th>Dominant/recessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retinitis pigmentosa</td>
<td>612095</td>
<td>PROM1</td>
<td>recessive</td>
</tr>
<tr>
<td>Cherubism</td>
<td>118400</td>
<td>SH3BP2</td>
<td>dominant</td>
</tr>
<tr>
<td>Ellis-Van Creveld syndrome</td>
<td>225500</td>
<td>EVC</td>
<td>recessive</td>
</tr>
<tr>
<td>Weyers acrofacial dysostosis</td>
<td>193530</td>
<td>EVC/EVC2</td>
<td>dominant</td>
</tr>
<tr>
<td>Catsh* syndrome</td>
<td>610474</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Crouzondermo-skeletal syndrome</td>
<td>612247</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Thanatophoric dysplasia</td>
<td>187600</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Hypochondroplasia</td>
<td>146000</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Achondroplasia</td>
<td>100800</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Lacrimoauriculodentodigital (LADD) syndrome</td>
<td>149730</td>
<td>FGFR3</td>
<td>dominant</td>
</tr>
<tr>
<td>Huntington disease</td>
<td>143100</td>
<td>HTT</td>
<td>dominant</td>
</tr>
<tr>
<td>Hurler syndrome</td>
<td>607014</td>
<td>IDUA</td>
<td>recessive</td>
</tr>
<tr>
<td>Night blindness</td>
<td>163500</td>
<td>PDE6B</td>
<td>dominant</td>
</tr>
<tr>
<td>Phenylketonuria II</td>
<td>261630</td>
<td>QDPR</td>
<td>recessive</td>
</tr>
<tr>
<td>Deafness, nonsyndromic sensorineural 6</td>
<td>600965</td>
<td>WFS1</td>
<td>dominant</td>
</tr>
</tbody>
</table>

(*) Camptodactyly, Tall stature and hearing loss syndrome

**Pathogenic CNVs**

In 1965 visible chromosomal deletions on chromosome 4 were for the first time associated with a clinical defined phenotype called Wolf-Hirschhorn syndrome [1]. Since then, an increasing number of pathogenic aberrations of 4p were reported. In 2000 up to 100 cases were described with deletions (from 4pter to 4p15) as the sole chromosomal anomaly. Two distinct phenotypes have been associated with deletions on the short arm of chromosome 4: the Wolf-Hirschhorn syndrome, which is likely associated with terminal deletions and the proximal 4p deletion syndrome, which is likely associated with interstitial deletions (4p16→p12) [2-9]. In addition to the deletions also several
Genotype-phenotype correlations

duplications linked to a well-characterised chromosome disorder have been reported [3, 10-16]. An overview of all published pathogenic CNVs in 4p is presented in Figure 1. In addition, multiple other chromosomal aberrations including translocations, inversions, ring chromosomes, and deletions associated with duplications have been reported. Due to their complexity, these are not covered here; however, they are assembled in the ECARUCA database (www.ecaruca.net). Interestingly, apart from chromosome 22, chromosome 4 has the highest number of pathogenic deletions reported in ECARUCA. On the short arm of chromosome 4 (4p16→p11), 157 pathogenic deletions and 51 pathogenic duplications have been reported since May 2008.

Since novel high resolution screening techniques paved their way in diagnostics, an increasing number of submicroscopic pathogenic CNVs have been reported. Array CGH data of patients with constitutional anomalies are collected in DECIPHER (www.sanger.ac.uk/PostGenomics/decipher); however the pathogenicity of those imbalances often remains uncertain. As of May 2008, a total 28 submicroscopic apparently pathogenic CNVs on chromosome 4p, 25 deletions, and 3 duplications are reported in DECIPHER. Their sizes vary from 120 kb to 8.25 Mb with a majority of deletions affecting the terminal part of chromosome 4p causing the Wolf-Hirschhorn syndrome.

Deletions: the Wolf-Hirschhorn syndrome

Wolf-Hirschhorn syndrome is recognized by a specific facial dysmorphology often referred to as the ‘Greek helmed facies’ and severe to moderate mental retardation, microcephaly, epilepsy and pre- and postnatal growth retardation. The distinctive facial appearance is the combined result of microcephaly, broad forehead, hypertelorism, a prominent glabella, high arched eyebrows, short philtrum, and micrognatia. In a smaller number of the cases heart and renal defects, midline defects, teething problems, and hearing defects are also reported.

WHS is usually associated with a chromosome 4p segmental terminal deletion. First through conventional karyotyping, later by FISH, and more recently by array CGH, these deletions have been mapped with ever increasing precision. There is a linear relation between the severity of the syndrome and the 4p16 deletion size [28]. The overwhelming majority of the deletions are terminal and vary in size up to 20 Mb. About 20% of the cases are restricted to 4p16.3, representing the most terminal 5 Mb [29], which are below the resolution of conventional karyotyping. No common breakpoints could be identified [30-32].
Figure 1: Overview of the copy number variations on the short arm of chromosome 4. On top, an Ensemble view visualizes the CNVs stored in the DGV (first row) and published by Redon et al. [17] (second row). Benign and pathogenic variants are collected from published and our own data. Undet stands for undetermined and indicates a CNV not yet known to be benign or pathogenic. Deletions are indicated in red and duplications in green. The size of the colored boxes indicates the size of the deletion/duplication. Numbers next to the boxes correspond to a reference. WHS: Wolf-Hirschhorn syndrome. 1) Liehr et al., [18]; 2) Rodriguez et al., [19]; 3) Van Buggenhout et al., [20]; 4) Balikova et al., [21]; 5) Van de Graaf G et al., [22]; 6) Jabs W et al., [23]; 7) Chitayat et al., [24]; 8) Fryns JP et al., [25]; 9) White et al., [9]; 10) Moller et al., [26]; 11) Ishikawa et al., [5]; 12) Davies J et al., [27]; 13) Tonk et al., [8]; 14) South et al., [7]; 15) Kozma et al., [6]; 16) Estabrooks et al., [4]; 17) Basinko a et al., [2]; 18) Kakinuma et al., [12]; 19) Wyandt et al., [16]; 20) Sabaratnam et al., [15].

Based on the minimal region of overlap of the deletions detected in two WHS patients, a Wolf-Hirschhorn Critical Region (WHSCR) was defined and overlaps with a 165 kb interval [33]. This critical region contains two genes, Wolf-Hirschhorn syndrome candidate 1 (WHSC1) and 2 (WHSC2) gene. In 2003, based on the detection of a 1.9 Mb terminal deletion in a WHS patient, a second critical region (WHSCR2) flanking the original WHSCR at the distal site was defined [34]. Both critical regions contain the WHSC1 gene, and since this gene is (partially) deleted in every WHS case reported, haploinsufficiency of this gene is probably causing the facial characteristics. However, there are some arguments against this hypothesis. First, sequencing of WHSC1 in WHS-like patients without deletions at or near the WHSCR did not reveal any mutations [30, 35]. Also, a few recent reports describe patients with a mild WHS phenotype which includes the recognizable facies, but with deletions not covering the WHSCR [30, 31] (Hannes et al., in preparation). This implicates that other neighboring genes contribute to the etiology of this syndrome. Finally, studies in mice show that the deletion of the Whsc1 gene alone is not sufficient for WHS [36].

Extensive genotype-phenotype correlations have delineated the critical regions causal for the other aspects of the WHS phenotype. Intrauterine and postnatal growth retardation, microcephaly, cleft lip
and palate, oligodontia, congenital heart disease, and hypospadias have now been assigned to small intervals, but the genes causing these traits have yet to be identified [20]. Since those critical regions contain multiple genes, other approaches are needed to come one step closer to identify causative genes. Possible approaches are bioinformatic tools to identify the most likely candidate genes, animal models, and improved phenotyping. Finally, correct genotype and phenotype data storage in designated databases like DECIPHER or ECARUCA would enlarge the WHS study population with mildly affected WHS patients and eventually identify small causal imbalances [37].

**Duplications: the trisomy 4p syndrome or duplication 4p syndrome**

Not only deletions but also several 4pter duplications have been reported and are generally known as trisomy 4p syndrome or dup 4p syndrome [11]. This syndrome is defined by large duplications containing 2/3 of the short arm of chromosome 4. The severity of the syndrome correlates with the size of the duplication. Around 30% of the young patients die within one year of age. Common features are the specific facial appearance with a prominent glabella, bulbous nose with flat or depressed nasal bridge, retrognathia, pointed chin, short neck with low hairline, enlarged ears with abnormal helix and antihelix, and other features like rocker-bottom feet with prominent heel, arachnodactyly and camptodactyly [14]. Recently, some large duplications are shown to reside in the genome in an inverted orientation often accompanied with small cryptic deletions [38, 39].

**Benign copy number variants**

**Cytogenetically visible euchromatic variants**

Phenotypically neutral euchromatic imbalances are called euchromatic variants and an increasing number of such variants is being uncovered (for review see [40]). This collection of euchromatic variants is accessible at the website of the National Genetics Reference Laboratory (http://www.ngrl.org.uk/Wessex/collection.htm). Among all chromosomes, the highest variability is demonstrated by the Y chromosome [41]. On chromosome 4p only two such euchromatic variants have been reported: 1) The chromosomal study of a healthy father and daughter showed the presence of an excess of material in the short arm of one chromosome 4. Subsequent FISH studies determined the 4p16.1 locus to be approximately 3.3 Mb in size and stretch from 8.6 Mb to 11.6 Mb from the telomere [19]. 2) Liehr et al. [18] performed a cytogenetic study on a large group of small supernumerary marker chromosomes where some of them are associated with a normal clinical phenotype. Thereby the euchromatic region of 4p13→p12 is described as a benign variant.

**Submicroscopic copy number variation**

CNVs present in the normal healthy human population are collected in the Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/). This database reports 333 CNVs covering 10.72 Mb (21.08%) on the whole p-arm and 134 CNVs covering 4.66 Mb (46.6%) on the distal 10 Mb of 4p (Table 2). Up to 50% of the genes in the 10-Mb terminal part are located in CNV loci compared to 30% on the whole p-arm. Aside from the strikingly high incidence of apparently benign CNVs, it is
remarkably that relatively more CNVs are reported on the distal 10 Mb part of the chromosome compared to the whole p-arm. Our own analysis of both 1 Mb low resolution BAC array data and high resolution 244K Agilent array suggests that the percentage of benign CNVs is much lower than indicated by DGV.

**Table 2:** Comparison of the variation reported on the whole arm of chromosome 4p and the terminal 10 Mb region of chromosome 4p. Data is extracted from the Database of Genomic Variants (DGV).

<table>
<thead>
<tr>
<th></th>
<th>4p</th>
<th>0-10 Mb distal part of 4p</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (Mb)</td>
<td>50.86</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Number of genes</td>
<td>379</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>Total number of variable regions</td>
<td>577</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>Number of variable loci</td>
<td>333</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>Number of genes in CNV</td>
<td>114</td>
<td>76</td>
<td>50% of the genes on the distal part and 30% on 4p are located in CNV loci</td>
</tr>
<tr>
<td>Largest CNV locus (Mb)</td>
<td>1.31</td>
<td>1.31</td>
<td>relative to 4p</td>
</tr>
<tr>
<td>Total sequence CNV (Mb)</td>
<td>10.72</td>
<td>4.66</td>
<td>relative to 10 Mb of the distal part</td>
</tr>
<tr>
<td>% CNV</td>
<td>21.08</td>
<td>9.16</td>
<td>Striking! Only 53% of the terminal 10 Mb is unique sequence according to the DGV</td>
</tr>
<tr>
<td>Unique sequence (Mb)</td>
<td>40.14</td>
<td>5.33</td>
<td></td>
</tr>
</tbody>
</table>

The reason for the discrepancy between the apparent paucity of benign CNVs observed by our own analyses and the high CNV rate observed in DGV might have both a technical and/or a biological cause. First, the technical variation can be explained by the different platforms used. Initial mapping of benign variation was based on relatively low resolution arrays. Thus, the size of some CNVs may be overestimated. BAC insert sizes are reported to be CNV while high resolution oligo-arrays pinpoint smaller CNV regions [42]. A second attribution to technical variation is the use of different data-analysis algorithms. The false positive and negative rates may vary amongst laboratories. One biological cause of variation might be the use of different reference samples. An overview of all benign CNVs is provided in Figure 1.

To determine whether an imbalance is benign or rather causing developmental delay, studies of large patient cohorts are imperative. Only the recurrent association of an imbalance with normal individuals and/or a similar phenotype will allow a CNV to be assigned. Also, the serendipitous study of rare families may assign certain loci as benign or causing developmental delay based on their copy number state. Below we describe one locus for which these conditions are met.

The terminal 300 kb deletion of chromosome 4 has now several times been reported in healthy individuals and can be considered to be benign when hemizygous [43, 44].
**Benign variation turning pathogenic**

While the recurrent association of a copy number variant with a normal phenotype pinpoints the locus as a benign variant, thus far the number of copies tolerated is not (yet) reported. There are already several examples of benign hemizygous deletions which can cause recessive disorders when nullizygous. Recently, Flipsen-ten Berg et al. [45] described an atypical patient with a terminal 4p deletion. Not only was a terminal 8.4 Mb deletion detected but also an associated inverted duplication. Interestingly, the authors showed that the terminal 4p deletion unmasked a mutation of *WFS1* in the other allele, thus contributing to the atypical WHS phenotype.

Recently, the first example of a 4p amplification associating with a disorder was reported [21]. The olfactory receptor gene cluster (ORGC) at 9.07 Mb is also a well known benign CNV according to the Olfactory Receptor DataBase or ORDB (http://senselab.med.yale.edu/ordb/default.asp). However, the amplification of this locus has been associated with autosomal dominant inheritance of microtia, eye coloboma, and imperforation of the nasolacrimal duct [21]. A genome wide linkage analysis and array CGH revealed an amplification of 750 kb on 4p16. The variation is microscopically visible and co-segregates with the phenotype in the family. Surprisingly, upon molecular characterization it was shown that this region encompasses the ORGC described above. In addition, a 290 kb region flanking the ORGC but also described as copy number variable was co-amplified. All affected family members carry five copies of this locus.

**Conclusion**

Since the introduction of array CGH, we have seen an exponential increase in the number of both (apparent) benign and pathogenic imbalances detected at chromosome 4p. It seems evident that in the near future, more small 4p imbalances will be uncovered. To increase our insights in the consequences of these imbalances it is imperative to collect both genotype and phenotype information of large cohorts of persons with imbalances. Such analyses will enable us to unravel the genetic determinants contributing to the WHS phenotype and other 4p syndromes. Such data can always be deposited at the databases indicated in this text. It can be anticipated that the mere collection of CNV data will not be sufficient to generate clinically relevant knowledge and that eventually curation by (groups of) experts will be required.
References


Genotype-phenotype correlations


Feenstra I., Brunner H. G. and van Ravenswaaij C. M. Cytogenetic genotype-phenotype studies: improving genotyping, phenotyping and data storage. Cytogenet Genome Res. 2006;115(3-4):231-9


MECHANISM GENERATING 4P TERMINAL DELETIONS

Chapter III
Telomere healing
Paper 6: Telomere healing following DNA polymerase arrest induced breakages is likely the main mechanism generating chromosome 4p terminal deletions

Femke Hannes, Jeroen Van Houdt, Oliver W. Quarrell, Martin Poot, Ron Hochstenbach, Jean-Pierre Fryns and Joris R. Vermeesch

Human mutation. 2010
Abstract

Constitutional developmental disorders are frequently caused by terminal chromosomal deletions. The mechanisms and/or architectural features that might underlie those chromosome breakages remain largely unexplored. Since telomeres are the vital DNA protein complexes stabilizing linear chromosomes against chromosome degradation, fusion and incomplete replication, those terminal deleted chromosomes acquired new telomeres either by telomere healing or by telomere capture. To unravel the mechanisms leading to chromosomal breakage and healing, we sequenced nine chromosome 4p terminal deletion boundaries. A computational analysis of the breakpoint flanking region, including 12 previously published pure terminal breakage sites, was performed in order to identify architectural features that might be involved in this process. All terminal 4p truncations were likely stabilized by telomerase mediated telomere healing. In the majority of breakpoints multiple genetic elements have a potential to induce secondary structures and an enrichment in replication stalling site motifs were identified. These findings suggest DNA replication stalling induced chromosome breakage during early development is the first mechanistic step leading towards terminal deletion syndromes.
Introduction

Terminal chromosomal deletions are the most common class of subtelomeric abnormalities and are often associated with mental retardation and multiple congenital anomalies [1]. The most frequent terminal deletion syndromes include the 1p36 deletion syndrome (MIM 607872), the 4p terminal deletion leading to Wolf-Hirschhorn syndrome (MIM 194190), the 5p terminal deletions causing Cri-du-Chat syndrome (MIM 123450), the 16p terminal deletion leading to alpha thalassaemia (MIM 141750), 9q34 deletion syndrome (MIM 610253) and the 22q terminal deletion syndrome (MIM 606232). Despite their clinical importance, it remains largely unknown when during development those terminal deletions arise, what the mechanisms are causing the chromosomal breakages leading up to those terminal deletion syndromes or which mechanisms rescue the broken chromosomes.

Recent studies show that many apparent simple terminal deletion breakpoints have proved difficult to sequence [2-4]. In a recent study on chromosome 9q deletions, only 2 out of 14 apparently pure terminal deletions could be sequenced [4]. The inability to analyze the breakpoints may relate to the complexity of the genomic sequence at the breakpoint such as inverted and tandem duplications or intervening sequence of unknown or ectopic origin [1, 2, 5-8]. Another reason for the inability to identify precisely the genomic location of the telomeric breakpoint is its location within repetitive sequences [3]. Overall, relatively few breakpoints have been analyzed at the sequence level and systematic studies mapping larger series of broken chromosomes are nearly non-existent. Moreover, the finding of mainly pure terminal deletions at chromosome 16pter versus more complex rearrangements at 1p, 22q and 9qter might imply that different chromosomes predispose to different mechanisms leading to the chromosome breakage and subsequent rescue.

Apparently simple 4p terminal deletions are the most frequent cause of WHS (55%-75%) [9-12]. Thus far, no chromosome 4 terminal deletions breakpoints have been mapped at sequence level. Here, we sequence systematically 9 consecutive terminal deletions of the short arm of chromosome 4 and explore the genomic architecture surrounding the breakpoints. In contrast to studies at other chromosomes, we were able to map and sequence all 9 telomere junction sites. Our data suggest that all telomeres were added de novo to the chromosome specific sequence. Furthermore, analysis of the breakage sites provide evidence that specific sequence motifs potentially induce chromosomal breaks leading up to terminal deletion syndromes.

Materials and Methods

Human subjects

Patients with developmental anomalies and mental retardation were ascertained over a period of five years of screening. Blood samples were collected after informed consent. Seven patients were seen in Leuven (BE), one in Sheffield (UK) and one in Utrecht (NE). Except for the latter patient were all patients diagnosed with Wolf-Hirschhorn syndrome. The phenotype and genotype of the latter patient have been described elsewhere [13]. All patients, but one, had normal high resolution G-banded chromosome analysis and were ascertained through either abnormal fluorescent in situ
hybridization (FISH) results using a commercially available WHS probe (Cytocell) \[14\] and/or 1 Mb BAC array following the protocol as described elsewhere \[15\]. Unbalanced 4p translocations as well as interstitial 4p deletion carriers were excluded.

**Ultra-high resolution oligonucleotide microarray**

A tiling path oligonucleotide microarray containing 385,000 probes (Roche NimbleGen systems) covering the region between 1.0 Mb to 11.8 Mb on chromosome 4p was custom designed. Both olfactory receptor gene clusters, located respectively around 3.9 Mb and 8.8 Mb, were avoided from the design. The average probe resolution was 30 bp. In addition, for one patient with the largest deletion, a HD2 whole genome oligonucleotide microarray with 2.1 million probes was used (Roche NimbleGen systems). For this array the average probe resolution was 1.5 kb. The labeling, hybridization, washing and scanning were performed according the manufacturer’s instructions (Roche NimbleGen systems). Data extraction and calculation was done via NimbleScan (Roche NimbleGen systems). The log2 ratios were calculated following the SegMNT algorithm. A 10x average window was used for breakpoint determination and the visualization was done via SignalMap V1.9 (Roche NimbleGen systems).

**Quantitative PCR (qPCR)**

qPCR primers were designed with Universal ProbeLibrary (Roche Applied Science). Primers were chosen free from any single nucleotide polymorphisms or repeats using repeatmasker provided by the UCSC browser (http://genome.ucsc.edu/). qPCR was performed using the lightcycler 480 instrument (Roche Applied Science) working in a total volume of 15 µl including 7.5 µl of mastermix, 2.5 µl of primermix (2.5 µM) and 50 ng template DNA. The following amplification conditions were used: 95°C for 5 min, and 40 cycles at 95°C for 10 s and 60°C for 20 s. After the amplification protocol, a melting curve was obtained at 95°C for 30 s, 60°C for 30 s and 95°C for a continuous mode (5 acquisitions per °C) and finally cooling down to 40°C for 30 s. Next, data was analyzed with Excel (Microsoft) according to the comparative ddCt method (Sequence Detection System bulletin 2 [Applied Biosystems]).

**Telomere anchored PCR and breakpoint sequencing**

A unique primer, located in the normal copy region, was chosen per breakpoint in combination with a telomere specific primer (5'TATGGATCCCTAACCCTGACCCTAACCC3') to amplify across the junction. A total volume of 50 µl contained platinum Taq polymerase (Invitrogen), 1.5 mM MgCl₂, 200 µM of dNTPs, 0.25 µM forward and reverse primer (Eurogentec) and 100 ng of template DNA. The PCR program is defined as follows: 95°C-5 min, 35 cycles of (95°C-30s, 58°C-30s, 72°C-1.30 min) with a final extension of 72°C-10 min. After amplification, the PCR products were analyzed on a1% agarose gel and sequenced on an ABI 3130 automated capillary DNA sequencer using the BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems).
In silico analysis of breakpoint flanking sequences

The Blat tool provided by the UCSC browser (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) [16] was used to map the obtained sequences to the human genome for determining the exact location of the breakpoint. Furthermore, we collected from the literature, 14 additional pure terminal breakpoints that have been mapped at the sequence level including seven breakpoints at chromosome 16p [5, 17], one breakpoint cluster at chromosome 22q containing three terminal deletion breakpoints [18, 19], two at chromosome 9q34 [4] and two single breakpoints at respectively chromosomes 7q [6] and 1p36 [7]. To exclude a bias during our statistical analysis when using all three terminal breakpoints at chromosome 22q, only one breakpoint flanking region [18] was taken into account. In total, 12 terminal breakpoints were used to complete our analysis (Supplementary table 1).

All breakpoint flanking regions for further in silico analyses were defined by 150 bp including 75 bp flanking both sides of the breakpoint. In order to assess whether the breakpoint regions (4p and others) were enriched for certain genomic features, we randomly selected 500 sequences of 150 bp from the human genome (build 18), distributed over the different chromosomes and avoiding gaps and centromeric regions. Different online tools were used (online resources) to analyze these 150 bp random human sequences and the breakpoint sequences surrounding each junction (including 75 bp at both sides of the exact breakpoint).

RepeatMasker (http://www.repeatmasker.org/) (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-3.0) was used to identify the highly repetitive sequences like long-interspersed nuclear elements (LINEs), Short-interspersed tandem nuclear elements (SINEs) long terminal repeats (LTRs) and simple tandem repeats (STRs). Non-B structural elements including left-handed DNA and tetraplex structures were detected by respectively ZHUNT (http://bioinfo.cgrb.oregonstate.edu/zDNA/) [20] and QGRS (http://bioinformatics.ramapo.edu/QGRS/analyze.php) [21]. REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer/submission.html) [22] allows identifying different repeat types, we recorded direct, mirror, complementary and inverted repeats. The presence of known motifs was analyzed with Fuzznuc (http://inn-temp.weizmann.ac.il/cgi-bin/emboss/fuzznuc) (EMBOSS package, [23]). The motifs presented by Abeysinghe and co-workers were used for further analysis [24]. Melina (http://melina2.hgc.jp/public/index.html) [25, 26] was used for identifying novel common motifs and Weblogo (http://weblogo.berkeley.edu/logo.cgi) [27] was used to generate sequence logo’s which are the graphical representation of a nucleotide multiple sequence alignment. Weblogo enabled us to screen for the presence of common nucleotides at certain positions from the exact breakpoint.

For all features we recorded the number of observations in the breakpoint sequences (4p and others from literature) and the random human sequences (Table 1). In both sets of sequences we counted the number of sequences with at least one hit for a given feature (Table 1). We performed a Fisher exact test in order to determine whether the occurrence of a feature was significantly different for the breakpoint sequences and the random human sequences.
**Results**

**Sequence analysis of terminal breakpoints**

DNA of 8 patients were hybridized on specific tiling path oligonucleotide microarrays of chromosome 4p which enabled further fine mapping of the breakpoints to an approximately 200-1500 bp interval (Figure 1A). The breakpoint in one larger terminal deletion, mapped at approximately 15 Mb according the 1 Mb BAC array, was further characterized using a whole genome HD2 NimbleGen array. This breakpoint is mapped to an approximately 12 kb interval.

![Figure 1. Overview of 9 terminal breakpoints at chromosome 4p16. A) Results of a targeted oligonucleotide microarray (Nimblegen) starting from 1.0 Mb to 11.8 Mb. Y-axis represents the log2 values of the intensity ratios and the X-axis, the position of the oligonucleotides on chromosome 4p. Breakpoint n°9 is analyzed with a HD2 oligo array (Nimblegen) of which only a part of chromosome 4p is visualized. B) The junction sequence. The red arrow indicates the start of the telomere repeat sequence. The green arrow refers to a TTAGGG repeat. C) Global alignment of the junctions and indication of the microhomology with TTAGGG sequence in yellow. Upper line refers to the reference sequence, the second line is patient derived sequence at the breakpoint, and third line is the telomere sequence.](image)

Since terminal deletions are often reported to be associated with submicroscopic terminal duplications [28] the probes flanking the contiguous stretch of probes with reduced intensity ratios due to the terminal deletion were inspected for increased intensity ratios. No increased intensities were identified near the breakpoints thus excluding the presence of inverted deletion duplications. Nevertheless, in breakpoint n° 6, 7 and 8, a group of probes around 4 Mb, more specifically between
3,854,200 bp and 4,322,200 bp, showed an apparently normal copy number within a larger terminal deleted region (Figure 1A). The reason for this elevation in log2 values is most likely the presence of repetitive regions flanking the olfactory receptor gene clusters which have multiple copies scattered within the genome. Hence, a deletion of a few copies will not be detected and will appear as normal copy number value.

To validate the results of the high resolution arrays and further fine map the location of the breakpoint, a series of qPCRs were performed in the purported breakpoint region. Quantitative PCR using different primer sets showed a contiguous stretch present as a single allele while a flanking set of primer pairs showed the presence of two alleles and thus no additional complex rearrangements flanking the breakpoint could be registered. The region in-between those primer pairs with altered copy number contains the terminal breakpoint. Once the breakpoint region was refined down to 200-800 bp, a breakpoint spanning PCR was designed. To this end, telomere anchored PCR was performed using a telomere specific primer combined with the reverse primer from the first primer pair with a diploid copy number state. A positive control without a 4p deletion was performed in parallel to distinguish the correct PCR product from the other bands as seen in the positive control. All amplicons were smaller than 600 bp and subsequently sequenced (Figure 1B). In all sequenced products, 4 to 27 TTAGGG hexamers were obtained (Figure 1B). The chromosome specific sequences flanking the telomeric repeats were aligned with the human genome reference sequence and revealed the immediate proximity of the telomeric repeats to the chromosome specific sequence in all fragments (Figure 1C). De novo telomere addition to a DNA fragment is thought to be mediated by telomerase. Telomerase mediated addition of telomeric repeats onto a DNA fragment is stimulated by the presence of a telomeric repeat sequence to which the RNA template of telomerase can bind [29-31]. Hence, we determined whether such telomeric repeats are located at or near the telomeric breakpoint. In 7/9 junctions we observed, in frame with the TTAGGG repeat, microhomology of 2 to 5 bp with the RNA component of telomerase (Figure 1C and Table 2). In addition to terminal breakages at chromosome 4p16, microhomology in frame with the telomere sequence was also detected in 9/12 breakpoint flanking sequences previously reported.

Table 2. Overview of the exact terminal breakpoints at chromosome 4p obtained from 9 different patient samples. STR: simple tandem repeats, LINE: long interspersed nuclear elements, SINE: short interspersed nuclear elements.
The enzymatic part of telomerase also known as TERT contains mainly 3 functional domains: the active site located at the C-terminus, the TERC binding site is located centrally, and the N-terminus interacts with the ssDNA end. This N-terminal anchor site will determine the efficiency by which telomerase can conduct his function. Recently, Sealy and colleagues showed a ssDNA length and sequence dependency of the telomerase enzyme within a short template sequence. Moreover the interaction of the N-terminus preferentially interacts with oligonucleotides of minimal 13 nt in length and a G-rich character at certain positions in the oligonucleotide [32]. In order to characterize such a sequence specificity flanking the breakpoint at the junctions that could define the interaction with telomerase, a 16 bp fragment flanking each breakpoint was analyzed via Weblogo. No common nucleotides at certain positions flanking the breakpoint could be detected on chromosome 4p16 or others.

**In silico analysis of the breakpoint**

**Low and high copy repeats**

The mechanism(s) generating terminal chromosomal deletions remain(s) largely unknown. If the genomic architecture influences the generation of chromosomal breaks leading up to terminal deletions, specific sequences or DNA structures might be present near the deletion breakpoint both at chromosome 4p16 and others. To investigate this possibility, in silico analysis was conducted on the 150 bp breakpoint junction interval defined by 75 bp on either side of the breakpoint. The breakpoint flanking sequences were first analyzed for the presence of low copy repeats (LCRs) and highly repetitive elements. LCRs are mediators of non-allelic homologous recombination [33], and could be involved in chromosome instability leading up to terminal deletions. No LCRs were detected in the nine 4p deletion breakpoint sequences, whereas six of the nine contained highly repetitive elements. In 3 breakpoint flanking sequences we observed long-interspersed nuclear elements (LINEs), in two other fragments short-interspersed nuclear elements (SINEs) were present and one sequence contained a simple tandem repeat (STR)(See Table 2). In light of the 12 previously reported breakpoints, in total 4 breakpoint flanking sequences contained repetitive elements including one breakpoint flanking sequence contained a SINE element, the other two contained STRs and one breakpoint flanking sequence contained a LCR.

**Non-B-conformations genetic elements**

Certain genetic elements in the genome can adopt non-B conformations such as triplexes, tetrplexes, cruciforms, slipped hairpin structures and left-handed Z-DNA. Those conformations can hamper DNA replication and as a consequence may result in a break [34, 35]. To investigate the potential role of non-B structural elements residing in the sequence surrounding the terminal breakpoint, different applications such as REPuter were used to investigate the presence of direct, inverted, complementary, mirror repeats and ZHUNT was used to search for left handed Z-DNA conformations. In total, 6/9 (67%) breakpoint flanking sequences on chromosome 4p16 contained one or more non-B structural elements. Moreover, direct repeats can potentially form slipped hairpin structures which in turn may trigger the rearrangement. Multiple direct repeats of minimal 8 bp were
found within one breakpoint flanking sequence. Mirror repeats of minimal 8 bp, potentially leading to triplex structures, were detected in 2 breakpoint flanking sequences and complementary repeats were identified in 3 breakpoint flanking sequences. Finally, 1 breakpoint flanking sequence contained an inverted repeat potentially adopting cruciform conformations and is proposed to be responsible for genome rearrangements and/or gross deletions [35]. 12 other previously reported simple terminal deletions were analyzed in parallel with those on chromosome 4p. We observed direct, inverted and mirror repeats from at least 8 bp in respectively 4, 3 and 2 breakpoint flanking sequences. To investigate whether the finding of a certain element is significant or accidental, similar in silico analysis was performed for randomly chosen sequences extracted from the human genome and subsequently compared to the combined number at terminal breakpoints both at chromosome 4p and others. An overrepresentation of individual structural elements such as mirror repeats (p=0.106) and alternating purine/pyrimidine sequences (p=0.118) leading to respectively triplex structures and left-handed Z-DNA were identified (Table 1). Although those elements were often seen in breakpoint flanking sequences, their p-values did not reach statistical significance (p<0.05) upon frequency determination.

Figure 2. Overview of all the elements potentially leading towards a non-B conformation or specific motifs that hamper DNA replication or facilitate recombination. Both groups of terminal deletions are represented including aberrations at chromosome 4p (left) and 12 other chromosomal terminal deletions previously reported in the literature (right).
Furthermore, the local flipping of small regions of right-handed B-DNA to left-handed Z-DNA in the sequence requires negative supercoiling making the sequence prone to breakage. This B to Z transition is strongly favored in alternating purine/pyrimidine sequences and can be in silico determined. ZHUNT analysis predicted at chromosome 4p in one breakpoint flanking sequence n° 4 a 15 bp sequence (TGGTGCGTCCCGTA) 21 bp away from the breakpoint inducing a left-handed structure. Equally, within the group of 12 previously reported breakpoint flanking sequences, one contained a 15 bp sequence (AAGTGTGCACGCTCG) 56 bp proximal to the breakpoint. Finally, oligo(G)n tracts were identified which may form tetraplex structures potentially inducing breakage. Among nine breakpoint flanking sequences on chromosome 4p, 4 different oligo(G)n tracts of 15 to 29 bp were identified in three breakpoint flanking sequences, whereas in the 12 other breakpoint flanking sequences 7 contained at least one oligo(G)n repeat with a size varying between 11 and 26 bp (Table 1 and Figure 2).

**Known sequence motifs**

Besides the complex architectural features and repeats inducing altered secondary structures, certain sequence motifs are also known to predispose to breakage. Overrepresentation of certain sequence motifs are found in the vicinity of the converted region which can promote recombination not by their sequence per se but because their potential to form non-B conformations leading to a break [36]. We evaluated the presence of known motifs within the breakpoint flanking sequences on both the 5' to 3' and the 3' to 5' strands. In total, 17/40 motifs were obtained from the nine breakpoint flanking sequences. Moreover, 3 of those motifs such as the deletion hotspot consensus (TGRRKM), the DNA polymerase arrest site (WGGAG) and the Murine parvovirus recombination hotspot (CTWTTY) were found to be frequent (> 50%), however not necessarily simultaneously present. Amongst the nine breakpoint flanking sequences, 5 contained such a deletion hotspot recombination hotspot motif (Table 1). All of the three motifs are scattered throughout the 150 bp. Whereas among the 12 previously reported breakpoint flanking sequences, in 9/12 one or more DNA polymerase arrest site(s), in 7/12 Ig heavy chain class switch repeats 5 and in 4/12 a Ig heavy chain class switch repeats 3 were found (Table 1). Comparison using 500 sequences generated at random, not derived from the human genome sequence, showed statistically significant p-values for the DNA polymerase arrest site (p=0.026) and Ig heavy chain class switch repeat 5 (p=0.01).

However the same analysis with 500 randomly generated sequences from the human genome, showed that these particular motifs such as the DNA polymerase arrest site (p=0.119) and Ig heavy chain class switch repeats 5 and 3 (p=0.102 and 0.08) were overrepresented, although not statistically significant, in the group of in total 21 terminal breakpoints including 9 on chromosome 4p and 12 other terminal breakpoints previously reported.

**Novel sequence motifs**

Alternatively, the identification of novel common motifs was performed using Melina. Among the breakpoint flanking sequences at chromosome 4p, one particular motif of 5 bp (KGGMA) was commonly found in all breakpoint flanking sequences. This novel motif was tested in the 12 previously reported breakpoint junctions. In 9/12 breakpoint flanking sequences, 19 hits were obtained in both
sense and antisense direction. Investigation of this novel motif at 500 randomly generated sequences of the human genome, revealed no statistical overrepresentation or an enrichment within the breakpoint flanking sequences.

**Discussion**

To unravel the mechanisms leading up to terminal deletion syndromes, we mapped and sequenced the breakpoints of nine terminal 4p deletions. We demonstrate that all broken chromosomes contain pure TTAGGG repeats flanking chromosome specific sequence. Similar findings were previously reported for 12 pure terminal breakpoints at other chromosome arms. Microhomology with the TTAGGG repeat was found in 16/21 (76.2%) terminal breakpoints. Furthermore, we analyzed the breakpoint flanking region for the presence of repeats, potential genomic architectural elements and sequence motifs that might ultimately lead to chromosome breakage. Herein, we demonstrate an enrichment of DNA polymerase arrest sites and non-B structural elements that might impede the progression of DNA polymerase during DNA replication. Hence, our data suggest that DNA polymerase pausing may lead to chromosomal breaks and those breaks are subsequently healed by telomerase.

**Healing of broken chromosomes**

Broken chromosomes can acquire a telomere sequence either by capturing the telomere or by telomerase mediated telomere healing. There are two lines of evidence that imply that all chromosome 4p telomeres analyzed were healed by telomerase [1]. In vitro studies have demonstrated that telomerase preferably adds a segment of the telomeric hexamer repeat unit onto a DNA fragment at its 3’ end. A microhomology in frame with the telomere repeat of 2-5 bp with the telomerase template in 7/9 (77.7%) breakpoints was demonstrated. This result is concordant with previous published terminal breakpoints at other chromosomes in which this type of microhomology is detected at 9/12 (75%) breakpoints [4-6, 18]. When a telomere is captured from another chromosome, usually not only the telomere but also subtelomeric sequences are captured [37]. No such subtelomeric sequences were detected. In addition, telomeric repeats flanking the unique euchromatic sequence have been replicated during multiple cell divisions and thus acquire at the most proximal location many permutations of the hexameric TTAGGG sequence such as e.g. TTGGG, TGAGGG and TAGGG. Hence, when a telomere sequence is captured from another chromosome, those variants of the TTAGGG sequence would be present. On the other hand, when the telomere is acquired via telomerase only exact TTAGGG hexamers will be present. In all our junctions fragments exact TTAGGG repeats were flanking the breakpoint junction. Hence, all 4p broken chromosomes are likely stabilized via telomerase promoted repair and thus suggest a telomere healing event.

Indirectly, we can deduce when during development breaks arise. First, it is known that telomerase activity is low in mature oocytes and spermatozoa, but highly expressed from the blastocyst stage onwards [38- 40]. If a breakage occurs in the oocyte, spermatocyte or in the zygote other mechanisms of DSB repair will be activated rather than repair via telomerase leading possibly to
other types of rearrangements [41]. Otherwise, when a break occurs in early development, telomerase is present and will interfere with the DSB repair processes such as non allelic homologous recombination (NAHR) and non homologous end joining (NHEJ). Moreover, it has recently been shown that such a de novo addition via telomerase is negatively regulated by DNA damage signaling preventing aberrant healing of broken DNA ends by telomerase. This shows that healing via telomerase is subordinate to both non allelic homologous recombination (NAHR) and non homologous end joining (NHEJ) [42]. Hence, telomeric deletions are often stabilized as derivative chromosomes rather than pure deletions [9]. Secondly, our results show an extensive microhomology at the terminal breakpoints, suggesting an important role in telomerase dependent repair. Finally, it was recently shown that genomes of early cleavage stage embryos are prone to chromosome instability and that a high number of blastomeres carry chromosomal deletions [43]. Since chromosome instability is high in early cleavage stage embryos and since telomerase is most active during early development, it seems likely that terminal deletions arise during early human development. This implies that the origin of generation and stabilization occurs during mitosis rather than meiosis and thus, it seems likely that the broken 4p chromosomes are generated and stabilized during early embryonic development.

Besides that approximately 76% of the terminal deletion breakpoints contained microhomology with the RNA template of telomerase, a small proportion of the breakpoints did not contain microhomology. It is known that binding and elongation of the telomerase enzyme involves base pairing with the telomerase RNA template and the substrate [44, 45]. In vitro studies have recently shown that binding of telomerase to a substrate also depends on the G-rich character at certain positions in the substrate [46]. According to our data, no obvious G-rich position nearby the breakpoint could be defined in a 16 bp sequence fragment flanking the terminal breakpoints without any microhomology. This suggests that in vivo telomerase can elongate broken chromosomes without having a complement to the RNA template nor a certain enrichment for a nucleotide near the healing site.

**Mechanism of chromosomal breakage**

Sequences such as repeats, non-B structural elements or certain motifs can predispose to DNA breakage [24, 36, 47-49]. Here, the enrichment of particular motifs was observed at terminal breakpoints. A DNA polymerase arrest site was identified to impede the progress of DNA polymerase alfa at phiX174 DNA resulting in stalling of the replication fork [50]. Replication arrest actually stimulates slippage and can trigger recombination and thus the finding of a replication arrest motif can be a key factor in genome instability [51-53]. In E.coli a mechanistic link between a replication arrest site and homologous recombination has been established [54]. In about 50% of the breakpoint flanking sequences such an arrest site was present. Likewise, an overrepresentation of DNA polymerase arrest motifs (p=0.01) are found at sequences flanking the gene conversion region [36]. Furthermore, both Ig heavy chain switch class repeats 3 and 5 were enriched at the terminal breakpoints. These repeats mediate a process of nonhomologous recombination that introduces variation in the constant region of an immunoglobulin heavy chain gene. Those switch repeats have been noted at translocation breakpoints associated with a number of hematological malignancies [55].
Moreover, repeat 5 was also shown to be significantly overrepresented within close proximity at translocation breakpoints \(p=0.05\) [24].

Besides the specific motifs, repetitive sequences inducing altered DNA conformations are frequent within the surroundings of the rearrangement breakpoints [34]. Such elements can elevate the rate of single-strand lesions, and hence contribute to their conversion to double-strand breaks [47, 56]. Within the terminal breakpoint flanking sequences of both chromosome 4p and others, direct, mirror, inverted and complementary repeats were found even as puridine/pyrimidine and oligo(G)n tracts. Interestingly, the fragments were likely to induce more left-handed Z-DNA \((p=0.118)\) and triplex \((p=0.106)\) structures than what we would expect from the random genome. Other studies with similar findings conclude that both structures are significantly associated with the site of DNA breakage at translocation junctions [24, 34].

In addition, the presence of repetitive elements including high and low copy repeats might play a role in stimulating the formation of non-recurrent breakpoints [57]. No low copy repeats were identified at or near the breakpoints whereas highly repetitive elements were abundant. 6/9 breakpoints are embedded within a repetitive element most often involving interspersed sequences (Table 2). In addition, 4/12 pure terminal breakpoints extracted from the literature are within segmental duplications, simple tandem repeats or SINEs (Table 1). Overall, the repetitive content of the breakpoint flanking sequences is heterogeneous and the fragments do not \textit{per se} contain LCR nor repetitive elements. Hence, in our dataset, 50% of the breakpoints associate with LCR or highly repetitive sequences which are comparable with that of the human genome and thus there is not strong evidence of the contribution of those elements in the generation of breakages.

The relatively high frequency of DNA polymerase arrest sites at the breakpoint flanking sequences suggests one plausible mechanism of DNA breakage following DNA replication stalling. The presence of other structures rendering the DNA prone to DNA breakage might be witnesses from other breakage mechanisms. It is well known that dicentric chromosomes can be formed (either during meiosis or during mitosis). Following mitosis, the dicentric chromosome would break leading to two daughter cells having either a deleted 4p or an inv dup del (4p) and it can be expected that the breaks would occur at the “weakest” loci [28]. No inv del dupl (4p) were detected during our study. Another mechanism that might induce breakages is homologous recombination intermediates. Larger groups of terminal deletions need to be sequenced to determine the involvement of specific motifs in the breakage process and to discriminate whether the occurrence of different motifs might be relics of different mechanisms.

\textbf{Acknowledgments}

We thank all families who participated in the study. This work was made possible in part by grants GOA/2006/12, and the SymBioSys Center of Excellence (Research Council, K.U.Leuven, EF/05/007) to J.R.V.
Table 1. Overview of the non-B structural elements and motifs tested at chromosome 4p and others

<table>
<thead>
<tr>
<th></th>
<th>4p-</th>
<th>Other chr</th>
<th>Combined</th>
<th>Human sequences</th>
<th>Random sequences</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n° seq all hits %</td>
<td>n° seq all hits %</td>
<td>n° seq all hits %</td>
<td>n° seq all hits</td>
<td>n° seq all hits</td>
<td>COM/HUMAN</td>
</tr>
<tr>
<td>Human sequences</td>
<td>n=9 seq=9 hits=9 %</td>
<td>n=12 seq=12 hits=12 %</td>
<td>n=21 seq=21 hits=21 %</td>
<td>n=500 seq=500</td>
<td>n=500 seq=500</td>
<td></td>
</tr>
<tr>
<td>Random sequences</td>
<td>n=9 seq=9 hits=9</td>
<td>n=12 seq=12 hits=12</td>
<td>n=21 seq=21 hits=21</td>
<td>n=500 seq=500</td>
<td>n=500 seq=500</td>
<td></td>
</tr>
<tr>
<td>MICROHOMOLOGY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zDNA</td>
<td>7</td>
<td>7</td>
<td>16</td>
<td>76.2</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>Oligo(G)n</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>14.3</td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>Direct repeats</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>25.0</td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td>Mirror repeat</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>25.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complementary repeats</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>25.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverted repeats</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>25.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA polymerase arrest site</td>
<td>WGGAG</td>
<td>6</td>
<td>10</td>
<td>75.0</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>deletion hotspot consensus</td>
<td>TGRKRM</td>
<td>5</td>
<td>11</td>
<td>75.0</td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td>Murine parvovirus recombination hotspot</td>
<td>CTWTTY</td>
<td>5</td>
<td>5</td>
<td>55.6</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Ig Heavy chain class switch repeat 5</td>
<td>TGGAC</td>
<td>4</td>
<td>5</td>
<td>58.3</td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td>DNA polymerase a/b frameshift hotspot 2</td>
<td>TGGNGT</td>
<td>3</td>
<td>3</td>
<td>33.3</td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td>Ig Heavy chain class switch repeat 3</td>
<td>GGGGT</td>
<td>3</td>
<td>3</td>
<td>33.3</td>
<td></td>
<td>0.13</td>
</tr>
<tr>
<td>Ig Heavy chain class switch repeat 4</td>
<td>TGGGG</td>
<td>3</td>
<td>5</td>
<td>50.0</td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td>translin binding site</td>
<td>GCCCWSS W</td>
<td>2</td>
<td>2</td>
<td>22.2</td>
<td></td>
<td>0.80</td>
</tr>
<tr>
<td>Ig Heavy chain class switch repeat 1</td>
<td>GAGCT</td>
<td>2</td>
<td>2</td>
<td>22.2</td>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td>Ig Heavy chain class switch repeat 2</td>
<td>GGGCT</td>
<td>2</td>
<td>2</td>
<td>22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>consensus scaffold attachment region 3</td>
<td>WADAWAY AWW</td>
<td>2</td>
<td>2</td>
<td>22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>consensus scaffold attachment region 4</td>
<td>TWWTDTT WWW</td>
<td>2</td>
<td>2</td>
<td>22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA polymerase a frameshift hotspot 2</td>
<td>CTGGCG</td>
<td>1</td>
<td>1</td>
<td>11.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DNA polymerase b frameshift hotspot 1</td>
<td>ACCCWR</td>
<td>1</td>
<td>1</td>
<td>11.1</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>DNA polymerase a/b frameshift hotspot 1</td>
<td>ACCCCA</td>
<td>1</td>
<td>1</td>
<td>11.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Human minisatellites conserved seq/X-like element</td>
<td>GGWGGWWWG</td>
<td>1</td>
<td>1</td>
<td>11.1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Vaccinia Topoisomerase I consensus</td>
<td>YCCTT</td>
<td>1</td>
<td>1</td>
<td>11.1</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>DNA polymerase a frameshift hotspot 1</td>
<td>TCCCCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Translin binding site 1</td>
<td>ATGCAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>X-element E.coli</td>
<td>GCTGGTTG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ad6-M26</td>
<td>ATGACGT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARS consensus S cerevisiae</td>
<td>WRTTATRTTWW</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARS consensus S Pombe</td>
<td>WRTTATTTAW</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Consensus scaffold attachment region 1</td>
<td>AATAAAAAYA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Consensus scaffold attachment region 2</td>
<td>TTWTWTWTT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Drosophila topoisomerase 2 consensus</td>
<td>GTNWAYATVTATNR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Heptamer recombination signal</td>
<td>CACAGTG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Human hypervariable minisatellites seq 1</td>
<td>GGAGGTT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Human hypervariable minisatellites seq 2</td>
<td>AGAGGTTG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Human minisatellites core sequence</td>
<td>GGGCAGGARG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Human replication origin consensus</td>
<td>WAWTDDWWWDDHWWGHMAWTTT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LT-IS motif</td>
<td>TGGAAATCCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mariner transposon-like element</td>
<td>GAAAAATGAGCTATTTACCCAGGAA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Murine MHC recombination hotspot</td>
<td>GAGRCAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nonamer recombination signal</td>
<td>ACAAAAAAC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Category</td>
<td>Sequence</td>
<td>n°seq</td>
<td>All hits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------------</td>
<td>-------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pur binding site</td>
<td>GGNGAGGGAGARRR</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recombination hotspot</td>
<td>CCNCCNTNCNCNC</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrotransposon</td>
<td>TCATACACACGTCGAGGGGTAGAGGACT</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertebrate topoisomerase II</td>
<td>RNYNNXNNGYNGKTNY</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XY32 homopurine-pyrimidine</td>
<td>AAGGGAGARGGGTATAGGGAGAGGGAA</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend: n°seq is the number of sequences where at least one hit was recorded. All hits refers to the complete set of hits that have been found over all the sequences indicated.
References


Telomere healing


[41] Ballif B. C., Wakui K., Gajecka M. and Shaffer L. G. Translocation breakpoint mapping and sequence analysis in three monosomy 1p36 subjects with der(1)t(1;1)(p36;q44) suggest mechanisms for telomere capture in stabilizing de novo terminal rearrangements. Hum Genet. 2004;114(2):198-206


**Supplementary data**

**Supplementary Table 1**

<table>
<thead>
<tr>
<th>N°</th>
<th>Chromosome</th>
<th>Breakpoint (bp)</th>
<th>Repeat at breakpoint</th>
<th>Microhomology with TTAGGG (bp)</th>
<th>Reference</th>
<th>Patient initials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7q32</td>
<td>131,449,009</td>
<td>No</td>
<td>1</td>
<td>[6]</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>22q13.3</td>
<td>49,469,068</td>
<td>STR</td>
<td>0</td>
<td>[18]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16p</td>
<td>1,886,763</td>
<td>No</td>
<td>4</td>
<td>[5]</td>
<td>BO</td>
</tr>
<tr>
<td>4</td>
<td>16p</td>
<td>147,681</td>
<td>No</td>
<td>3</td>
<td>[5]</td>
<td>CMO</td>
</tr>
<tr>
<td>5</td>
<td>16p</td>
<td>144,446</td>
<td>No</td>
<td>3</td>
<td>[5]</td>
<td>IDF</td>
</tr>
<tr>
<td>6</td>
<td>16p</td>
<td>142,819</td>
<td>SD</td>
<td>4</td>
<td>[5]</td>
<td>TAT</td>
</tr>
<tr>
<td>7</td>
<td>16p</td>
<td>112,851</td>
<td>STR</td>
<td>3</td>
<td>[5]</td>
<td>TI</td>
</tr>
<tr>
<td>8</td>
<td>1p36</td>
<td>4,788,530</td>
<td>No</td>
<td>2</td>
<td>[7]</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9q34</td>
<td>137,329,926</td>
<td>SINE</td>
<td>5</td>
<td>[4]</td>
<td>P40</td>
</tr>
<tr>
<td>10</td>
<td>9q34</td>
<td>139,129,925</td>
<td>No</td>
<td>2</td>
<td>[4]</td>
<td>P3</td>
</tr>
<tr>
<td>11</td>
<td>16p</td>
<td>210,918</td>
<td>No</td>
<td>1</td>
<td>[17]</td>
<td>HW</td>
</tr>
<tr>
<td>12</td>
<td>16p</td>
<td>212,156</td>
<td>No</td>
<td>5</td>
<td>[17]</td>
<td>BR</td>
</tr>
<tr>
<td>13</td>
<td>22q13.3</td>
<td>49,469,074</td>
<td>STR</td>
<td>(0)</td>
<td>[19]</td>
<td>Subj 1</td>
</tr>
<tr>
<td>14</td>
<td>22q13.3</td>
<td>49,469,086</td>
<td>STR</td>
<td>(3)</td>
<td>[19]</td>
<td>Subj 2</td>
</tr>
</tbody>
</table>

*Not included in the analysis, because of similar locus as described previously by Wong and colleagues [18].
Position-effect
INVESTIGATING THE ROLE OF POSITION-EFFECTS IN 4p16 MICRODELETION SYNDROME

Chapter IV
Position-effects
Paper 7: Silencing of telomere flanking genes on chromosome 4p16 suggests telomere position-effects

Femke Hannes, Thiery Voet, Jean Pierre Fryns and Joris Vermeesch

In preparation
Abstract

Telomere position-effects (TPE) have been demonstrated at artificial telomeres from yeast to human cancer cells. To determine whether telomere position-effects play a role in human constitutional disorders, we studied the expression of 21 genes in 7 lymphoblastic cell lines, 3 with interstitial and 4 with terminal 4p deletions. All microdeletions have occurred de novo. At one terminal deletion, a diploid gene flanking the newly seeded telomere showed half of the normal expression. We demonstrate that a loci flanking the promoter region was hypermethylated and subsequent reactivated upon azaC treatment. These findings suggest that the downregulation is most likely caused by a telomere position-effect. This is the first demonstration of a potential telomere position-effect in a human disorder and it shows that TPE should be taken into consideration when performing genotype-phenotype correlations.
**Introduction**

Position-effects is the term used to describe altered expression of genes flanking an imbalance [1]. The transcriptional deregulation can presumably be explained by two mechanisms: (1) through modification of the chromatin structure of the gene locus and (2) through a change of trans-acting factors on cis-regulatory sequences [2]. Overall, it is not clear how often and to what extent position-effects do play a role in human genetic disease. Nevertheless, a handful reports have tried to unravel this question. Merla et al. showed that not only genes residing within the 7q11.23 deletion but in addition some genes flanking both sides of the deletion showed reduced expression despite their diploid copy-number state [3]. Also, studies in mice have shown that interstitial genetic imbalances may influence nearby gene-expression which may have a biological significance [4]. Telomeres are also known to be able to influence gene expression, a phenomenon known as telomere position-effect (TPE). TPE is the result of telomeric heterochromatin spreading into telomere flanking sequence resulting in gene silencing. This latter mechanism is reversible and has been demonstrated at artificial telomeres from yeast to human cancer cells [5-7]. This propagation of heterochromatin varies upon both the distance to the telomere and telomere length [8, 9]. A few case reports have suggested a role for TPE influencing human phenotypes. A submicroscopic insertional translocation of the PLP1 gene adjacent to the telomere sequence of chromosome 19q was detected in a healthy male. Duplications of PLP1 cause Pelizaeus–Merzbacher disease. Since the extra copy of the PLP1 gene is flanking the telomere sequence in a healthy individual, the expression of the duplicated PLP1 gene is likely to be silenced [10]. Comparison between the clinical phenotype of patients carrier of a terminal 14q deletions and ring chromosome 14 both with the same breakpoint on 14q, demonstrates that patients with a ring chromosome 14 present additional clinical features like seizures and retinitis pigmentosa. Those additional features can originate from the presence of the telomere on 14p which induces a silencing effect on natural repositioned telomere flanking genes on 14q [11]. However, thus far, no direct link between human genetic disease and TPE has been proven yet.

Wolf-Hirschhorn syndrome, a clinically well defined contiguous gene syndrome, and its milder representation, known as 4p16 microdeletion syndrome, are characterized by both interstitial and most often terminal deletions on chromosome 4p16. We recently mapped and sequenced the breakpoints of a series of patients with pure terminal deletions and another series with interstitial deletions. Here, we use lymphoblastic cell lines from patients with these well characterized interstitial as well as pure terminal 4p deletions to investigate whether both telomere position-effects and/or position-effects exerted by interstitial deletions exist and whether those position-effects could play a role in the phenotype of the patients. To this end, the relative gene-expression level of 21 genes was examined in 7 lymphoblastic cell lines including 4 patients carrier of terminal deletions and 3 with interstitial deletions at chromosome 4p16. We demonstrate that submicroscopic deletions at chromosome 4p16 can influence the expression of genes flanking the rearrangement in lymphoblastic cell lines derived from patients with WHS/4p16 microdeletion syndrome. Moreover, we show for the first time a reversible silencing of a gene flanking the relocated telomere in a cell line with a terminal deletion. This
suggests TPE involvement in human terminal deletion syndromes and hence should be taken into account when performing genotype-phenotype correlations.

**Material and Methods**

**Selection of subjects**
Lymphoblastic cell lines of 7 patients carrier of terminal or interstitial deletions on chromosome 4p16 were selected. All deletions were further characterized via a 385k custom made oligonucleotide micro-array targeted to a region starting from 1 Mb to 11.8 Mb at chromosome 4p16. The probe resolution was 1 probe every 20/30 bp. The clinical details were described previously [12-14].

**Cell culture, RNA extraction and cDNA synthesis**
Lymphoblastic cell lines of 11 controls and 7 patients with WHS/4p deletion syndrome were grown in RPMI 1640 media supplemented with 10% fetal bovine serum. Total RNA from all samples was prepared in parallel from logarithmic growth–phase cells with the use of RNeasy Mini Kit (Qiagen), in accordance with the manufacturer's instructions. Subsequently, a DNase treatment was performed on total extracted RNA from all samples (Roche). Quality assessment of the RNA was done via the Agilent 2100 Bioanalyzer (Agilent Technologies). Next, 2 µg of good quality RNA was converted to cDNA with the use of Superscript III reverse transcriptase (Invitrogen) and a combination of both oligo d(T)12-18 (Invitrogen) and random primers (Invitrogen).

**Real-time quantitative PCR (RT qPCR)**
Real-time quantitative PCR with SybrGreen detection was used to measure differences in gene expression levels using the Lightcycler 480 instrument I (Roche Applied Science). Initially, via the Unigene website (http://www.ncbi.nlm.nih.gov/unigene) 58 genes at chromosome 4p16 were chosen to be expressed in blood cells. Next, primers were designed using the Universal ProbeLibrary Assay Design Center (Roche Applied Science) to span an intron boundary when possible. All primer pairs were computationally analyzed for SNPs or repetitive sequences using dbSNP and Repeatmasker via the University of California Santa Cruz genome browser (UCSC-human genome build 18). Blat analysis (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) was performed to ensure the specificity of the primers. All primersets were obtained from Eurogentec. The mastermix used is LightCycler® 480 SYBR Green I Master (Roche Applied Science).

GeNorm was exploited to determine the three best housekeeping genes for normalization in 10 lymphoblastic cell lines of normal individuals as described [15, 16]. Subsequently, GUSB, CLK2 and ACTB were selected for further analysis. The amplification efficiency coefficients or E-values were obtained using dilution series containing a pool of cDNA samples, as described elsewhere [17]. To this end, 21 stably expressed genes were selected with an E-value between 0.95 and 1.05 (Table 1). The remaining genes were not or not stably expressed in lymphoblastic control cell lines. Taken together, a correct working SybrGreen assay as indicated above was obtained for 36 % of the selected genes.
All RT qPCRs were performed in a final volume of 20 µl and two replicates per sample in a 96-well plate format. A final primer concentration of 0.5 µM, 10 µl SybrGreen I Master mix (Roche Applied Science) and 5 µl of template cDNA (in total 83.3 or 166.7 ng) were used. The runs were performed according to following amplification conditions: 95°C for 5 min, and 40 cycles at 95°C for 10 s and 60°C for 20 s. After the amplification protocol, a melting curve was obtained at 95°C for 30 s, 60 for 30 s and 95°C for a continuous mode (5 acquisitions per °C) and finally cooling down to 40 °C for 30 s. Each plate contained all the samples, including all patient and control samples for that particular gene to exclude any variability between the different runs.

The raw Cp points were automatically calculated from the Lightcycler 480 software via the absolute quantification/2nd derivative option and subsequently imported as a text file in qBASE for further analysis, data storage and management. In addition, melting curves for all combinations (sample/gene) were analyzed.

**DNA methylation analysis**

Hypermethylated DNA was prepared as suggested by the supplier’s protocol. Briefly, 5 µg control DNA was dissolved in the presence of SAM (S-adenosylmethionine-32 mM) (New England Biolabs) and SSSI CpG methyltransferase (New England Biolabs-4U/µl). This mixture was incubated overnight at 37°C followed by an ethanol purification.

Bisulfite conversion of 500 ng genomic DNA (patient 1 and 4 control individuals) and hypermethylated DNA (used as a positive control) was performed using an EZ DNA methylation kit according to the instructions of the manufacturer with minor modifications (Zymoresearch). The temperature was increased to 42 °C with an extension of the incubation time from 15 min to 30 min. After purification, 4 µl of converted DNA was used as a template to amplify the regions of interest with platinum Taq polymerase (Invitrogen) using specific primers. Primer design was done via Methyl Primer Express v1.0 using standard parameters. PCR conditions applied here started with a denaturation step at 95°C for 5 min followed by the amplification for 35 cycles at 95°C for 30 s, 53°C for 30 s and 72°C for 1 min and a final extension at 72°C for 7 min.

**Treatment with 2'-Deoxy-5-azacytidine (azaC)**

A 10 mM stock solution of azaC (Sigma) was prepared in sterile water and stored at -20 °C. Lymphoblastic cell lines derived from patient 1 and five control individuals were counted and seeded at an initial concentration of 0.4 *10^6 cells/ml in a total volume of 20 ml/flask. This cell suspension was grown in RPMI 1640 media supplemented with 10 % fetal bovine serum and treated with azaC at a final concentration of 1 µM. A negative control was left untreated. The media were replenished every 48 h by the replacement of half the conditioned media with fresh media and drug [18-20]. After 2-4-8 days of treatment, all the cells were harvested and DNA and RNA extracted.
Results

Molecular analysis of three out of four terminal deletions has been proven to be without any additional complexity at the breakpoints according to their array CGH results, breakpoint spanning PCR and subsequent sequencing results.Briefly, patient 1 has a 1.691.685 bp terminal deletion with the breakpoint located within the TMEM129. Patients 2 and 3 are carrier of a 2.161.732 bp and 2.424.502 bp terminal deletion, respectively. The breakpoint of patient 2 is embedded within the POLN and a LINE repeat. The terminal breakpoint of patient 3 is located within a gene desert and a SINE repeat. For one additional interstitial deletion (patient 5) both distal and proximal breakpoints were determined at 2.605.128 bp and 3.036.901 bp and are embedded within FAM193a and GRK4, respectively. The breakpoints of patients 4, 6 and 7 are not mapped at sequence level. Nevertheless, high resolution oligonucleotide microarray could accurately delineate the breakpoints. The terminal breakpoint of case 4 is located within a high copy number variable region at the border of the olfactory gene cluster family at approximately 3.851.000 bp. For the two remaining interstitial deletions 6 and 7, both their distal and proximal breakpoints were determined at approximately 1.169.000 bp and 2.496.000 bp and 1.873.000 bp and 3.450.600 bp, respectively.

Relative expression levels of genes within and flanking seven different 4p16 deletions were obtained in EBV transformed lymphoblasts. In total, 21 genes with a correct E-value between 0.95-1.05 were withheld for further analysis (Table 1). For 23 genes, the expression level was too low and variable and they were excluded for further analysis. Finally, for 14 genes no signal could be detected after 35 cycles using RT-qPCR. All genes, but one, located in the deletion region showed half of the normal expression. One outlier in patient 2, also known as KIAA1530, had a relative expression value of 0.83, which is in-between the expected expression value of a normal copy number state ($X_{kiaa1530-normal}$, $SD_{kiaa1530} = 0.904$) and one copy number state ($X_{kiaa1530-hap} + SD_{kiaa1530} = 0.56$) (Table 1). Inversely, 86% of the diploid genes flanking the deletion regions had relative expression values within the normal range. The relative expression values from the remaining 14% were either detected above or under the expected relative expression values (Table 1). Furthermore, the relative expression values of the major candidate genes for WHS, known as WHSC1 and LETM1, were expressed as expected according their copy number status in lymphoblastic cell lines.

Briefly, in 4 patients an abnormal expression pattern (not expected according to their copy number status) flanking either terminal or interstitial deletions was observed. The different cases will be elaborated on further below.
Table 1: Relative expression levels within and flanking terminal and interstitial 4p deletions. All genes are organized from telomere (top) to centromere (end). The grey cells indicate the genes located in the deleted region. The red numbers indicate the relative expression values surpassing the threshold for that specific gene with a normal diploid copy number state.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Patient 1</th>
<th>Patient 2</th>
<th>Patient 3</th>
<th>Patient 4</th>
<th>Patient 5</th>
<th>Patient 6</th>
<th>Patient 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PIGG</td>
<td>0.480</td>
<td>0.602</td>
<td>0.598</td>
<td>0.576</td>
<td>1.056</td>
<td>1.233</td>
</tr>
<tr>
<td>2</td>
<td>ATP5I</td>
<td>0.545</td>
<td>0.595</td>
<td>0.545</td>
<td>0.647</td>
<td>1.130</td>
<td>1.426</td>
</tr>
<tr>
<td>3</td>
<td>GAK</td>
<td>0.533</td>
<td>0.653</td>
<td>0.544</td>
<td>0.492</td>
<td>1.042</td>
<td>1.148</td>
</tr>
<tr>
<td>4</td>
<td>TMEM175</td>
<td>0.455</td>
<td>0.542</td>
<td>0.388</td>
<td>0.429</td>
<td>1.122</td>
<td>0.958</td>
</tr>
<tr>
<td>5</td>
<td>CTBP1</td>
<td>0.531</td>
<td>0.434</td>
<td>0.421</td>
<td>0.508</td>
<td>0.945</td>
<td>0.396</td>
</tr>
<tr>
<td>6</td>
<td>MAEA</td>
<td>0.440</td>
<td>0.367</td>
<td>0.341</td>
<td>0.523</td>
<td>1.036</td>
<td>0.345</td>
</tr>
<tr>
<td>7</td>
<td>KIAA1530</td>
<td>0.359</td>
<td>0.833</td>
<td>0.343</td>
<td>0.481</td>
<td>1.063</td>
<td>0.437</td>
</tr>
<tr>
<td>8</td>
<td>SLBP</td>
<td>0.625</td>
<td>0.632</td>
<td>0.451</td>
<td>0.654</td>
<td>1.179</td>
<td>0.465</td>
</tr>
<tr>
<td>9</td>
<td>TACC3</td>
<td>0.310</td>
<td>0.222</td>
<td>0.243</td>
<td>0.429</td>
<td>0.741</td>
<td>0.177</td>
</tr>
<tr>
<td>10</td>
<td>LETM1</td>
<td>0.951</td>
<td>0.487</td>
<td>0.478</td>
<td>0.554</td>
<td>1.092</td>
<td>0.482</td>
</tr>
<tr>
<td>11</td>
<td>WHSC1</td>
<td>1.233</td>
<td>0.478</td>
<td>0.400</td>
<td>0.589</td>
<td>1.031</td>
<td>0.467</td>
</tr>
<tr>
<td>12</td>
<td>MXD4</td>
<td>0.576</td>
<td>1.033</td>
<td>0.502</td>
<td>0.500</td>
<td>0.785</td>
<td>0.417</td>
</tr>
<tr>
<td>13</td>
<td>TNIP2</td>
<td>0.710</td>
<td>1.277</td>
<td>1.226</td>
<td>0.410</td>
<td>0.439</td>
<td>0.868</td>
</tr>
<tr>
<td>14</td>
<td>SH3BP2</td>
<td>0.599</td>
<td>0.999</td>
<td>1.285</td>
<td>0.399</td>
<td>0.357</td>
<td>0.784</td>
</tr>
<tr>
<td>15</td>
<td>ADD1</td>
<td>1.155</td>
<td>1.618</td>
<td>0.908</td>
<td>0.384</td>
<td>0.499</td>
<td>0.955</td>
</tr>
<tr>
<td>16</td>
<td>NOP14</td>
<td>0.970</td>
<td>1.038</td>
<td>0.846</td>
<td>0.467</td>
<td>0.552</td>
<td>0.811</td>
</tr>
<tr>
<td>17</td>
<td>MFSD10</td>
<td>0.807</td>
<td>0.981</td>
<td>1.130</td>
<td>0.429</td>
<td>0.375</td>
<td>0.876</td>
</tr>
<tr>
<td>18</td>
<td>HTT</td>
<td>1.187</td>
<td>1.066</td>
<td>1.067</td>
<td>0.507</td>
<td>1.147</td>
<td>1.123</td>
</tr>
<tr>
<td>19</td>
<td>LRPAP1</td>
<td>0.860</td>
<td>1.063</td>
<td>1.148</td>
<td>0.416</td>
<td>0.881</td>
<td>0.830</td>
</tr>
<tr>
<td>20</td>
<td>TMEM128</td>
<td>0.977</td>
<td>1.298</td>
<td>1.075</td>
<td>1.003</td>
<td>1.165</td>
<td>0.999</td>
</tr>
<tr>
<td>21</td>
<td>STX18</td>
<td>0.922</td>
<td>1.166</td>
<td>1.098</td>
<td>0.969</td>
<td>0.965</td>
<td>1.089</td>
</tr>
</tbody>
</table>

Patient 1 is of particular interest since expression is halved for 4 diploid genes. Note that the first gene, TACC3, is located approximately 1.37 kb away from the relocated telomere. The relative expression level of TACC3 (0.31) was similar to that obtained in cell lines with a deletion of TACC3 (0.27). This amount in reduction in relative expression suggests a complete silencing of one allele. Besides the TACC3, a significant decrease in expression levels of three contiguous genes MXD4, TNIP2 and SH3BP2, approximately 540 kb proximal to the relocated telomere was detected. Despite being diploid, the relative expression values, 0.576 (MXD4) - 0.710 (TNIP2) - 0.599 (SH3BP2) were found within their expected range when hemizygous (-SD and +SD) (Table 1 and Figure 1). Note that this particular region containing three contiguous genes covers approximately 600 kb and is embedded within a larger region containing genes with normal expression levels. These results demonstrate a discontinuous silencing pattern in lymphoblastic cell lines from patients with terminal deletion syndromes (Table 1 and Figure 1).

In patient 2, the relative expression value for one gene, ADD1 (1.618), passed the upper threshold in normal individuals (X_{add1-normal} + SD_{add1} = 1.496). This gene is located at 684 kb away from the relocated telomere (Table 1 and Figure 1).
Figure 1: Overview of the relative expression levels in lymphoblastic cell lines of 7 patients with 4p deletion syndrome. Those graphs include 21 stably expressed genes located in 4p16.3 and are orientated from the telomere (left) towards the centromere (right). The genes linked to the numbers can be found in table 1. The red bars under the graphs pinpoint the genes which are located in the hemizygous region. The Y-axis shows the fold difference in comparison with controls.

Also, similar findings were obtained for two other patients 6 and 7 characterized by interstitial deletions. The relative expression level of ATPI5, distally located outside both interstitial deletions (1.426 (patient 6) and 1.283 (patient 7)), is elevated compared to 10 normal control individuals ($X_{atpi5-normal} + SD_{atpi5} = 1.205$) (Table 1 and Figure 1). In patient 7, three other genes were aberrantly expressed such as TMEM128, KIAA1530 and TACC3. The relative expression level of TMEM128 (1.347) exceeded the upper threshold of relative gene expression within 10 normal individuals ($X_{tmem128-normal} + SD_{tmem128} = 1.291$) (Table 1 and Figure 1). Also, the relative expression values of two other genes, KIAA1530 (0.783) and TACC3 (0.539), were surpassing the lower threshold limits ($X_{kiaa1530-normal} - SD_{kiaa1530} = 0.903$, $X_{tacc3-normal} - SD_{tacc3} = 0.620$) (Table 1 and Figure 1). Note that none of those genes were directly flanking interstitial deletions.
**DNA methylation analysis of silenced genes**

The reduced expression of TACC3 could result from either the deletion of a regulatory sequence telomeric to the start of the transcription site or could be the result of a telomere position-effect. Telomere position-effect is induced by the hypermethylation of the telomere flanking region. To determine whether DNA hypermethylation could be responsible for the altered expression pattern, bisulfite treatment followed by sequencing of the CpG islands in TACC3 and MXD4 were performed. TACC3 is directly flanking the terminal breakpoint whereas MXD4 is located approximately 540 kb away from that breakpoint (Figure 2A). Within the TACC3, two amplicons were chosen within one large CpG island and 464 bp apart from each other. The first amplicon contains the potential promoters and the first non-coding exon (Figure 2A). Despite the proximity of the relocated telomere, no differential methylation pattern within 19 CpG’s could be identified between patients 1 and 4 other control samples. The second amplicon is located in-between non-coding exon 1 and exon 2. Interestingly, 8 clearly differentially methylated CpG’s were detected (Figure 2B). This finding might suggest that the hypermethylation is functional and responsible for silencing. In addition to the TACC3 gene, 2 other genes further away from the breakpoint were investigated including LEMT1 that was used as a control locus whereas for MXD4, for which lower levels of expression had been detected. Amplicons 3 and 4 located within LETM1 and MXD4 contained 39 and 38 CpG’s, respectively. No differential methylation pattern could be observed at those locations, further away from the relocated telomere.

**Figure 2A:** Schematic representation of the genes relative to the relocated telomere in three patients with 4p terminal deletions. The arrows indicate actively transcribed genes. The lollipops indicate the silenced genes. The triangle indicates that no gene expression was observed. The green dot indicates no difference in DNA methylation pattern between patient 1 and the control individuals, whereas the red dot refers to the hypermethylated CpG locus in the patient.
Reactivation of silenced TACC3

To determine whether this locus of differentially methylated DNA within TACC3 was functional, lymphoblastic cell lines derived from patient 1 and three controls were treated with azaC. As a positive control, a lymphoblastic cell line of a patient with fragile X syndrome was also treated in parallel. Fragile X syndrome is caused by an expansion of the CGG repeat tract within the FMR1 gene which triggers the hypermethylation of the FMR1 promotor resulting in silencing of the transcription. Treatment of these FMR1 mutated cells with azaC results in the loss of methylation marks at the CGG tract and reactivates transcription of the gene [18].

Initially, the optimal azaC working concentration for the lymphoblastic cell lines was determined using media at a final concentration of 0.5-1.0-1.5 µM azaC. At day four, total RNA was extracted and subsequent transcript levels were evaluated for both the TACC3 and FMR1 gene in the positive control and normal individuals. A final concentration of 0.5 to 1.0 µM azaC was not toxic to the cells and no remarkable variation in relative transcription was observed between different treatments. Moreover, when an increase in the gradient of azaC was applied to our positive control, a 10 fold increase in relative transcript level for the FMR1 gene was observed (Figure 3A). These results indicate that our treatment was affective, albeit not toxic.

All cell lines were treated with a final concentration of 1.0 µM azaC for ten days in total. After day four and day ten, both the DNA and RNA were extracted from EBV transformed cells derived from patient 1 and three control individuals. After 4 days of treatment, a complete reactivation of TACC3 was
obtained in patient 1 resulting in an increase of relative expression from 0.36 to 0.69 using respectively 0 µM and 1.0 µM azaC (Figure 3B) whereas no particular difference in relative transcription was observed in the other control individuals. Note that 0.5 µM azaC was sufficient to completely relieve the silencing effect. This in contrast with the expression levels of three genes further proximal to the TACC3 gene which did not react upon azaC treatment and thus their lower expression level did not alter significantly between 0 µM and 1µM azaC (Figure 3C). Those later results are concordant with the absence of methylation at the promoters of those genes. Taken together, those results proof that TACC3 flanking the newly seeded telomere can be reactivated upon azaC treatment suggesting that it is not the deletion of regulatory elements but rather the methylation that silenced the expression of TACC3.
Figure 3: Reactivation of TACC3 expression upon four days of azaC treatment. (A) A gradient of azaC was applied to a lymphoblastic cell line derived from a patient carrier of a FMR1 mutation. FMR1 expression was measured via RT qPCR. (B) Lymphoblastic cell line derived from patient 1 was treated for 4 days with 1.0 µM azaC. TACC3 expression was measured via RT qPCR. (C) The relative expression levels of three silenced genes (MXD4-TNIP2-SH3BP2) proximally located to TACC3 were determined after 1.0 µM azaC treatment. The Y-axis shows the fold difference in comparison with the controls.
**Discussion**

Although the compelling evidence for TPE at yeast telomeres and artificial telomeres from yeast to human cancer cells [5-7], there is limited evidence for TPE at regular human telomeres nor for a role of TPE in human disease. Here, we show the reversible silencing of TACC3, a gene flanking a newly seeded telomere. A model has been proposed in yeast in which TPE can spread along the telomere flanking regions depending on the binding of the SIR 2-3-4 complex. This complex has the ability to either deactylate specific histone tails causing compaction of the underlying sequence and self-self interaction that potentially induces the propagation of heterochromatin [21-23]. Recently, a protein of the family of SIRT proteins, SIRT6, has been described in humans to specifically interact with telomeres and exert a histone specific deactylase activity. Moreover, SIRT6 would contribute to the spreading of the heterochromatin, similar as that has been seen in yeast [24, 25]. Overall, DNA methylation which as such can induce silencing is likely a generally property of TPE.

First, there is a lack of extensive studies that investigated the potential link of TPE in human congenital disorders. One study showed a delay in replication of genes flanking a breakpoint in a patient with a terminal 22q deletion. Unfortunately, no link with silencing, DNA hypermethylation nor condensation could be made [26]. Furthermore, a few case reports suggest the involvement of the telomere in silencing nearby genes [10, 11], but did not investigate TPE. Secondly, the difference in constitution between telomeres of different species has tempered the search for a decent model for TPE in humans.

TPE is characterized by the reversibility of repression due to the heterochromatinization of the region. To demonstrate this, cells were treated with different drugs to alter specific heterochromatin markers genome wide [5, 7]. Hereby, treatment with trichostatin A (TSA), an inhibitor of histone deacetylalatas, and not azaC, a DNA demethylation drug, has led to an increased transgene-expression in human cancer cells. Notably, no complete silencing of the transgene could be observed. In contrast, studies at artificial telomeres in mice ES-cells have shown that treatment with TSA did not produce a significant reversal whereas treatment with azaC returned the expression of the telomeric transgenes close to their original levels [27]. Similarly, upon the introduction of human artificial chromosomes in mice with different backgrounds containing short and long telomeres, silencing of telomere flanking genes was noticed and is DNA methylation dependent (Voet et al., personal communication). The studies in mice complemented our results whereby a complete reactivation was obtained after azaC treatment. This contrast in results obtained after drug treatments between different cell lines could be manifold. First, a difference in telomere length could influence the mechanism of TPE between different cell types. Mouse telomeres are long (> 50 kb) containing many heterochromatic markers whereas telomeres in human cancer cells tend to be rather short (> 2-4 kb). Secondly, mouse ES cells are capable of de novo DNA methylation of the integrated transgenes whereas in cancer cells this de novo DNA methylation is strongly diminished or nonexistent. In fact, enzymes capable of de novo DNA methylation are suppressed in cancer cells and somatic cells whereas in mouse ES cells, high levels...
of de novo DNA methylation activity are recorded [28, 29] and thus this may reflect a different response to azaC treatment. Overall, the mechanisms involved in the formation of TPE and the distance that TPE extends from the end of the chromosome may therefore differ between normal human cells and human cancer cells.

The degree of TPE appears to depend both on the telomere length itself and on the distance between the telomere sequence and the gene [6]. Recently, it has been shown that the distance of silencing can reach as far as 100 kb of the human telomere [30] and the closer to the telomere, the stronger the effect can be. In patients 3 and 4, no TPE could be observed. Here, the first measurable genes are located over 100 kb from the telomere. In patient 2, the first measurable gene was located at 72 kb and thus less than 100 kb from the relocated telomere. Also, no TPE could be observed. For patient 1, a position-effect is observed at 1.37 kb from the telomere whereas no position-effect could be observed of the first gene flanking TACC3 74 kb proximal from the relocated telomere. Those results indicate that position-effects may not reach beyond 70 kb away from the telomere. On the other hand, by altering the telomere length in the cells derived from the patients, TPE may be altered on telomere flanking genes and as such establish the link between the position-effect and the proximity of the telomere.

Also, from studies in yeast and mammalian cells it is known that TPE can influence gene expression in a discontinuous manner. Such a discontinuous expression pattern was noticed at native telomeres in yeast [31] and later on in human fibroblasts. In this latter background, up until 620 kb away from the natural telomere a discontinues expression pattern of 8 consecutive genes was illustrated by Ning et al. in young versus old human fibroblasts [32]. Our work revealed a similar discontinuous pattern of at least three consecutive genes in a domain located approximately 540 kb away from the terminal breakpoint in patient 1, although they had a normal copy number status. In addition, we were able to demonstrate that the low relative expression values are methylation independent. One possible explanation for the diploid genes escaping silencing in-between both the relocated telomere and the proximal region of aberrant expression could be the presence of natural insulators that shelters the genes from the effects of silencers [33, 34]. Another explanation could be the physical separation of a general cis-regulator due to the deleted sequence. Although a general locus control element that potentially thrives the expression of the diploid genes cannot explain the lower expression values in this particular case, since larger terminal deletions up to 2.1 Mb do not show altered expression of those diploid genes. Overall, silencing of the diploid genes further upstream the relocated telomere observed here operates likely through telomeric heterochromatin structure.

Besides potential TPE, another rationale was to study whether gene expression flanking interstitial deletions was also influenced. In patient 5, no aberrant expression of genes up to 2.1 Mb distal and 1.1 Mb proximal from the deletion region was observed. However in patient 6 and 7, a few genes escaped this rule. In patient 6, one gene, ATP5I, 510 kb distally located from the interstitial deletion region was overexpressed. Similarly, this latter gene was also overexpressed, albeit less evident, in patient 7 and is 1.2 Mb distally located from the deletion. Also, in patient 7, three genes located at both
sides of the deletion show aberrant relative expression values either over- or underexpression. The molecular mechanism responsible for the aberrant gene expression flanking those interstitial 4p deletions is still not clear. No evidence of a gradient in aberrant relative expression values away from the deletion region could be detected as described previously [3]. Hence, gene expression analysis of more CNVs flanking regions need to be analyzed.

Wolf-Hirschhorn syndrome (WHS) is a contiguous gene disorder characterised by terminal or interstitial deletions of chromosome 4p. In the past, genotype-phenotype correlations have been pivotal in refining the candidate regions for different features of WHS [12, 13]. However, previous extensive correlations have been hampered by phenotypic variability among patients with similar deletions (personal observations). One way to explain this phenotypic variability is the presence of position-effects. Our analysis indicates that genes flanking the genomic variation in human disease can alter in expression status and as such potentially contributes to the disease variation. Hence, genotype-phenotype correlations are not sufficient to unravel Wolf-Hirschhorn syndrome and possibly other genetic syndromes. Thereby, the expression of genes nearby the genomic variation should be taken in account to adjust genotype-phenotype correlations and as such to complete the search for genes in relation to disease. Our results strongly suggest that telomere position-effects can affect gene expression and potentially play a role in human terminal deletion syndromes.
References


GENERAL DISCUSSION & CONCLUSION

Chapter V
General discussion and conclusion
Understanding the process of genetic diseases is crucial to gain insight in human development and to develop strategies for treatment. One major cause of genetic diseases are copy number changes often generating rearrangements that imply multiple genes leading to contiguous gene syndromes (CGS). Contiguous gene syndromes convey complex phenotypes often associated with high inter-individual clinical variability. This thesis has focused on the genetic factors contributing to CGS and more specifically to WHS and 4p microdeletion syndromes.

1. Genetic dissection of common features in microdeletion syndromes

1.1. The advantage of molecular karyotyping

Genotype-phenotype correlations are aimed to identify the contribution of each of the genes in a CNV to the phenotypic features. In the past, conventional karyotyping was used leading to associations of very large regions with diseases [13]. As a consequence, the ability to identify the molecular basis behind the disease was very difficult. Nevertheless, micro-arrays have enabled to speed up the molecular identification process in genotype-phenotype correlations and have proven to be successful both in the identification of novel microdeletion or microduplication syndromes [42, 43, 109] as well as in redefining the causative genes/regions of well known diseases due to accurate breakpoint mapping [110-113].

1.1.1. The identification of pathogenic submicroscopic imbalances

With the advent of micro-arrays we identified novel pathogenic submicroscopic imbalances on chromosome 4p16. Larger deletions covering this whole region are known to cause WHS. The finding of smaller submicroscopic deletions in patients with milder phenotypes is crucial to genetically dissect this syndrome and to identify smaller regions and/or genes. Alternatively, a sporadic, de novo microduplication covering the WHSCR1 is shown here to be pathogenic, as the patient presents developmental delay and multiple congenital anomalies. The finding of this microduplication pinpoints to the dosage sensitivity of the WHSCR1 and its role in human development.

Although haplo-insufficiency of WHSC1 gene is proposed to cause the developmental delay and the distinct facial WHS gestalt, one patient challenged previous hypothesis of one predominantly causative gene. The patient was carrier of a maternally inherited submicroscopic 4p16 deletion proximal to the critical region and shown to segregate with the WHS facial phenotype. This latter case illustrates the potential additive effect of two or more hemizygous regions within chromosome 4p16. One haplo-insufficient region can contribute to a mild phenotypic representation whereas both result in the complete full blown phenotype.

Since micro-arrays have enabled the identification of submicroscopic deletions at chromosomes 4p16 in patients with some but not all of the WHS phenotypic features, it might be more appropriate to redefine the disorder as 4p16 microdeletion syndrome. The identification and further characterization of more unique patients will be pivotal to further unravel this contiguous deletion syndrome.
1.1.2. The identification of benign submicroscopic imbalances

A benign CNV is defined as a CNV not contributing to the disorder studied. In the study of developmental anomalies, CNVs embedded in the normal population or/and inherited from a normal parent are often considered as benign. Such benign CNVs are excluded from genotype-phenotype. We identified a 300 kb terminal deletion at chromosome 4p16 in a normal individual who was referred to the clinic because of multiple miscarriages [110]. Integration of this case in our genotype-phenotype correlations enabled to refine the distal boundary of the critical region for growth retardation and microcephaly with an additional 300 kb, and reinforces the importance to include apparently benign variants in genotype-phenotype correlations.

1.1.3. When a benign submicroscopic imbalance turns pathogenic

Imbalances called originally benign in one individual, may turn to a disease state in another individual dependent on the environment, the copy number state, the inheritance pattern and/or other genetic factors. Recently, these CNVs were referred to as Mendelian copy number variants (CNVs) [114]. Recurrent 16p13.1 microdeletions in patients with mental retardation, microcephaly, epilepsy, short stature were significantly enriched in the patient cohort versus the normal population (p-value = 0.0048). Later, different follow-up studies have demonstrated a significant association between this recurrent microdeletion and neurocognitive disorders more specifically idiopathic generalized epilepsy [115-117]. Collectively, 34 microdeletions involving the 16p13.1 loci were detected in 5563 patients with idiopathic generalized epilepsy while only in 2 out of 6814 control individuals (p-value = 0.002*10^-7), confirming our previous findings. The microdeletion was shown to be either de novo or inherited from parents not suffering from a neurocognitive disorder. Hence, the low population frequency of this recurrent microdeletion in conjunction with the presence of de novo events, may suggest a negative selective factor and thus a strong influence on the disease phenotype. Another study using DNA of patients suffering of a neurocognitive disorder such as schizophrenia did not reveal a significant increase of this recurrent microdeletion in the patients versus the normal population [118]. Inversely, the reciprocal 16p13.1 microduplication was described in our study not to be associated with MR/MCA phenotype. Ullmann et al. reported this duplication variant to be associated with autism and later, this microduplication was shown to be significantly associated with schizophrenia [118, 119]. Since our patient population was investigated for epilepsy and intellectual disability and not for autism or any other neurocognitive disorders, we cannot confirm whether or not other neurocognitive disorders are present in our patient population.

Taken together, those reports indicate common etiological factors for seemingly diverse diseases such as intellectual disability, autism, schizophrenia and epilepsy. In addition, this region represents one of the first susceptibility loci involved in neurocognitive disorders. To know whether disruption of this locus, either the microdeletion or microduplication, exerts a strong or rather a mild susceptibility to neurocognitive disorders, larger groups of well defined patients and controls should be analyzed in great detail using a high resolution molecular techniques such as aCGH or next generation sequencing. Finally, those mendelian CNVs underscore the additional complexity in making a proper
diagnosis. A general collective database is essential to characterize the pathological consequences of those mendelian CNVs.

1.1.4. Functional impact of submicroscopic imbalances

Understanding the clinical relevance of submicroscopic imbalances will be the main challenge of the near future. To enhance genotype-phenotype correlations, it is pivotal to continuously collect all copy number variable regions in combination with a comprehensive phenotypic analysis. This assembly will be vital in interpreting copy number variations necessary for the confirmation of pathogenicity and for the proper counseling of the patients and their families. Proper examples of such collections are publicly available databases like DECIPHER, ECARUCA, ISCA as well as in-house developed databases including those from Cartagenia, Baylor and Signature Genomics. Furthermore, those data will lead to an increased insight into the molecular pathways behind mental retardation and will successively contribute to a better understanding of human development.

1.2. The use of animal models in dissecting CGS

Since most CGS are rare and are often characterized by larger rearrangements containing multiple genes, finding the patients with unique rearrangements containing only a few genes is a difficult task. Moreover, functional analysis and study of the pathology in patients is often impossible. Animal models can be used to phenocopy the disease specific manifestations and therefore will speed up the process in dissecting a complex phenotype as seen in CGS. The most obvious model is mice in that it complements the phenotype in humans. Murine models have provided insights into the disease process either by elucidating the predominant gene or evaluating the effects of modifier loci in Smith-Magenis syndrome [120], DiGeorge syndrome [121-123] and Charcot-Marie-tooth syndrome [124, 125]. Nested deletions within the WHS syntenic region of mice have been created to recapitulate different manifestations of WHS [126]. Although this study was informative to the understanding of the disorder, no single genes causing the WHS phenotype could be identified so far.

Another approach in mice to dissect CGS and more specifically WHS and 4p microdeletion syndromes would be the creation and subsequent study of targeted polygenic mouse models [127]. Via this method potential genetic interactions of genes within the commonly deleted region can be revealed. In this perspective, at least two genes within the commonly deletion region of WHS, CTBP1 and TACC3, share a common interactor (FOG-1). It is not impossible that deletion of one gene causes a mild to normal phenotype whereas deletion of both genes is necessary to express the full-blown phenotype as we know in patients with classical WHS [127]. This polygenic mouse could be created either by crossing two mice heterozygous for a certain candidate gene resulting in a compound heterozygote mouse or inducing deletions containing more than one gene using the Cre-LoxP recombination system allowing to control the size and the position of the deletions.
1.3. The use of induced pluripotent stem cells

Animal models of human congenital disorders are invaluable but sometimes fail to represent a similar phenotype to the human pathophysiology. These drawbacks could be counteracted by the use of other disease models.

Recently, induced pluripotent stem (iPS) cells were generated by epigenetic programming of human somatic cells, in particular fibroblasts [128]. These cells, like embryonic stem cells, have the unique capacity to continuously self-renew and their potential to give rise to all cell types in the human body. The iPS derived fibroblast cells provide a reservoir of different cell types that can be obtained from the patients. Hence, the iPS cells can serve as a powerful platform to model human disease in culture with regards of patient specific genetic background.

In-vitro derived cardiomyocytes from patients with Leopard syndrome were shown to be hypertrophic compared to those derived from the healthy brother and thus pinpoint a mutated gene, *PTPN11*, to be responsible [129]. Furthermore, in light of CGS, such patient-specific stem cells can offer an unprecedented opportunity to recapitulate both normal and pathologic human tissue formation in vitro and as such will enable disease investigation.

1.4. Monitoring for additional mutations

One approach to gain deeper understanding of the phenotypic variability in patients with WHS/4p16 microdeletion syndrome is to make use of target-enrichment strategies of the complete 4p16 region combined with next generation sequencing to screen for additional recessive mutations within the deletion region [130]. Furthermore, it has been demonstrated in mice that additional mutations at other loci than 4p16 are required to develop heart disease [131]. On behalf of this idea, complete exome sequencing of the genetic material of WHS patients carrier of 4p16 deletions could provide a method to screen for additional mutations, besides those on chromosome 4p16, responsible for specific clinical manifestations. Finally, the use of complete exome sequencing might unravel potential mutations in patients with WHS phenocopies. This latter approach has been proven to be successful to identify a genetic cause in Miller syndrome [132].

2. Origin of phenotypic variability

Our results have shown that genotype-phenotype correlations are often hampered by the high phenotypic variability among patients presenting with 4p16 microdeletion syndrome. Note that this high clinical variability is also seen in other subtelomeric deletion syndromes as well as other genomic disorders (eg microdeletion syndrome 16p13.1). Since genotype-phenotype correlations are mainly aimed at genes underlying the rearrangement, accurate breakpoint mapping at basepair level to capture all the genes underlying the rearrangements in gene rich regions like subtelomeric regions are crucial. To this end, 4p terminal breakpoints were systematically analyzed at basepair level. *In silico*
analysis revealed a direct addition of telomere sequence to the breakages, also known as telomere healing (Figure 1).

**Figure 1:** Native versus relocated telomeres and their flanking genes. The red box refers to the subtelomere sequence, the blue and green circles to respectively the old versus relocated telomere sequence, the arrows refer to the actively transcribed genes and the lolypops to the inactive genes. The break is indicated by a yellow star.

Since haplo-insufficiency of the deleted genes alone cannot explain the clinical manifestations in some cases, other modifying factors must contribute to phenotypic variability. Those modifiers can be either of (epi)genetic origin or environmental factors [133]. The genetic factors influencing the phenotype could be mutations or variations in the genes located in the remaining allele. For example, Cohen syndrome is a recessive disorder characterised by mental retardation and multiple congenital anomalies and, most often, mutations in the $\text{COH1}$ gene. However, recently, in minority of the patients a single mutation in the $\text{COH1}$ gene in combination with a deletion of the other allele result in a compound heterozygosity [134]. In Bernard-Soulie r syndrome (BSS), unmasking of a mutation on the non-deleted allele has been demonstrated to contribute to the recessive disorder [135]. Furthermore, besides the deletion underlying mutations/alterations, (epi)genetic modifiers outside the rearrangement might play a role including additional rearrangements or mutations somewhere else in the genome or position-effects. Those latter modifiers will be discussed further below in relation to WHS/4p microdeletion syndrome.

3. Telomere position-effects

3.1. Monitoring gene-expression outside the deletion boundaries

Genomic structural rearrangements might lead to a modification of gene-expression levels within and surrounding chromosomal structural variation. Besides the Williams-Beuren syndrome, no other disease causing CNVs have been systematically explored. To this end, we used patients with WHS/4p
deletion syndrome as a model to study position-effect in relation to two types of structural variation i.e. interstitial and terminal deletions. Dosage specific expression was shown for nearly all genes. However, a few genes escaped this rule. It was demonstrated that not only hemizygous but also certain diploid genes flanking the deletions are aberrantly expressed. Moreover, one case with a terminal deletion was of particular interest to us and will be elaborated on further below.

3.1.1. Position-effects flanking terminal deletions are indicative for TPE

A combined approach including terminal breakpoint mapping, expression and DNA methylation analysis within and surrounding the deletion region, demonstrated a reversible silencing of telomere flanking genes. Reversible silencing of telomere flanking genes is known as telomere position-effects (TPE). This process was detected yeast telomeres to artificial telomeres in human cancer cells [107, 108, 136, 137]. Even more, an investigation of the endogenous genes located near native telomeres in young versus old human fibroblasts demonstrated telomere position-effects in a discontinuous manner indicating that genetic elements residing within the subtelomeric region could shield them from TPE [138] (Figure 2). Similarly, a group of at least three contiguous genes approximately 600 kb away from the relocated breakpoint was shown to be inactive. Possibly, insulators are presented to shield the intervening region from being silenced (Figure 2). Despite those findings, the link between TPE and human congenital disease remains questionable. Thus far, TPE has not been implicated in human genetic disease. One report has evaluated the occurrence of TPE in a patient with a pure terminal deletion of chromosome 22q, but no telomere flanking gene inactivation could be detected at mRNA level [139]. Possibly, the distance of the gene from the relocated telomere was too large (> 50 kb). Nevertheless, a growing body of evidence exists that telomere flanking regions are involved in pathogenesis from studies in patients carrier of ring chromosomes and complex rearrangements [140, 141].

Figure 2: Model of repression at native versus relocated human telomeres. The red box represents the subtelomere region containing insulators (rectangles), actively transcribed (arrows) and silenced (lolypops) genes. The original versus relocated telomere sequence is visualised by respectively blue and green circles. The level of silencing starting from the telomere is indicated by a blue bar.
3.1.2. Future experiments to determine the involvement of TPE in human terminal deletion syndromes

Although we show a reversible silencing of telomere flanking genes, the involvement of the telomere has not yet been proven. In light of this, further experiments need to be designed in order to establish a link between the telomere and nearby silencing.

TPE is influenced by the telomere length and the heterochromatic markers [71]. One approach to address this question would be to alter the telomere length in the cell lines of the patient to demonstrate a link between the position-effects in our patient and TPE. Longer telomeres contain more heterochromatic markers and as such induce silencing further downstream whereas shorter telomeres contain less heterochromatic markers and as such convert to an open state and relieve the silencing (Figure 3). One gene and its negative mutant are known to alter telomere length when overexpressed. Moreover, telomere repeat binding factor 1 (TRF1) is a negative regulator of telomere length and needs to form homodimers in order to bind the telomere sequence. Alternatively, the dominant negative mutant, dTRF1d, lacks the C-terminal DNA binding domain but retains the dimerization domain. Overexpression of dTRF1d will lead to a depletion of TRF1 proteins to the telomere sequence and as such results in telomere lengthening [142]. This approach has been successfully used by Koering and coworkers to relieve the silencing of a telomere flanking gene upon shortening of telomere length in human cancer cells [108].

Currently, such an overexpression model is being established in lymphoblastic cell lines derived from the patients and control individuals with 4p terminal deletion syndromes.

**Figure 3:** Interplay between telomere length and gene expression. Lolypops indicate the silenced genes whereas the arrows refer to the actively transcribed genes. The amount of silencing spreading from the telomere towards the centromere is indicated by a blue bar.
4. Overall conclusion

Contiguous gene syndromes convey complex phenotypes associated with larger copy number changes containing multiple genes. To understand the contribution of small copy number changes and eventually individual genes to the disease aetiology, genotype-phenotype correlations are usually performed. Here, we focussed on a large group of patients (> 1000) with intellectual disabilities and multiple congenital anomalies and studied the effect of copy number change in relation to disease. This investigation enabled us (i) to characterise a novel recurrent microdeletion/duplication on chromosome 16p13.1 in a group patients with intellectual disabilities and multiple congenital anomalies, (ii) to confirm and further delineate critical regions linked to certain features of Wolf-Hirschhorn syndrome (WHS) and (iii) to identify and further characterise patients with 4p microdeletion syndrome meaning patients with atypical genotypes and phenotypes overlapping the region linked to WHS. This latter group of patients made us realise that genotype-phenotype correlations based upon the molecular karyotype alone are not sufficient to decipher the contribution of all genes towards the 4p microdeletion syndrome. Given the large phenotypic heterogeneity among those patients, other factors need to be taken into account to result in a complete understanding of the phenotype. Other genomic factors could be either unmasking mutations or variants including mutations and CNVs at other loci in the genome. Next generation sequencing techniques will help to elucidate those variants. Besides genomic data, our results have shown that position-effects are involved and that factors flanking the copy number variable region should be taken into account when performing genotype-phenotype correlations. In other words, genotype-phenotype data should be enriched with expression and/or epigenetic data of at least genes flanking the region to unravel the complete set of genes that contribute in pathogenesis.
Summary

Genotype-phenotype correlations were initially used to identify the minimal region of overlap between differently sized rearrangements in association with a specific phenotypic feature. This approach enabled the molecular dissection of the disease. However, the detection of the genotype by low resolution cytogenetic techniques hampered accurate genotype-phenotype correlations and the search for causative genes. With the advent of micro-array technology, the genotypes were more easily obtained, previously known imbalances were molecularly delineated and novel, previously unknown imbalances could be identified. This genome wide approach enabled the collection of patients with similar genotypes and subsequent identification of common clinical features in order to detect novel microdeletion/microduplication syndromes. This latter design is referred to as the ‘genotype-first’ approach.

In the first part of the thesis, this ‘genotype-first’ approach was applied to a large group of patients with developmental disabilities and multiple congenital anomalies and enabled the identification and characterization of a microdeletion/microduplication syndrome at chromosome 16p13.1. Association studies have shown a significant overrepresentation of the 1.55 Mb microdeletion in the patient versus the control population and thus indicates a causative role in the etiology of developmental delay. Furthermore, the genome wide approach enabled the identification of uncommon submicroscopic rearrangements and further delineation of known imbalances on chromosome 4p16. Taken together, this research has enabled I) the identification and further delineation of minimal regions of overlap associated with certain manifestations typical for WHS, II) the characterization of a submicroscopic deletion outside of both currently accepted critical regions in a patient with mild representation of WHS and III) the identification of a submicroscopic duplication including the critical region in a patient with developmental delay and multiple congenital anomalies.

In the second part of the thesis, terminal deletions on chromosome 4p16 were characterized at the sequence level to gain more insight into the origin of constitutional breakage and the mechanism of telomere rescue at broken chromosomes leading up to terminal constitutional deletions. We developed a combined strategy of high resolution micro-array analysis and telomere anchored PCR followed by Sanger sequencing. This strategy enabled us to characterize 9 terminal breakages at basepair level. Detailed analysis revealed a direct addition of telomere sequence adjacent to the breaks and a microhomology of 2 to 5 bp in phase with the telomere repeat at nearly all broken chromosomes. Additional in silico analysis has shown an enrichment of DNA polymerase arrest sites nearby the breakpoint. Those data suggest that an arrest in DNA replication leads to broken chromosomes followed by a template dependent healing via telomerase.

In the third part of the thesis, we investigated the role of position-effects at chromosome 4p16 deletion syndrome since it is known from studies in yeast and human cancer cells that regions flanking the relocated telomere can be subjected to aberrant gene expression. To this end, 21 stably expressed
genes from the tip to 4.5 Mb away from the telomere were studied in lymphoblastic cell lines of patients with interstitial and terminal deletions. Several genes flanking both interstitial and terminal deletions revealed variable expression patterns compared to normal individuals. In particular, telomere flanking gene inactivation was reversible and methylation dependent. Those data strongly suggest that telomere position-effects (TPE) play a role in pathology and should be taken into account when performing genotype-phenotype correlations.


**Samenvatting**

Genotype-fenotype correlaties zijn initieel erop gericht om de genetische oorzaak in patiënten met duidelijke gedefinieerde syndromale fenotypes verder te verfijnen. Hierbij vertrekt men van een groep patiënten met éénzelfde syndroomaal phenotype om vervolgens een minimale chromosomale regio van overlap te kunnen definiëren met betrekking tot één of meerdere gemeenschappelijk(e) klinisch(e) kenmerk(en). Met de ontwikkeling van micro-arrays, werd het steeds eenvoudiger om het genotypie van een patiënt in kaart te brengen. Deze genoomwijde techniek heeft het mogelijk gemaakt om nieuwe, voordien ongekende syndromen te identificeren. Daar waar in het verleden gebaseerd werd op een duidelijk gemeenschappelijk klinisch fenotype, laat deze nieuwe genoomwijde techniek ons toe om patiënten met een eerder heterogeen fenotype en een gemeenschappelijk genotype te verzamelen. Achteraf wordt naar gemeenschappelijk klinische kenmerken gezocht en een minimale chromosomale regio van overlap. Deze laatste aanpak wordt ook het ‘genotype-eerst’ benadering genoemd.

*In een eerste deel van de thesis*, hebben wij tevens gebruik gemaakt van deze recent ontwikkelde micro-array techniek om enerzijds via de ‘genotype-eerst’ benadering een nieuw, ongekend 16p13.1 microdeletie/duplicatie syndroom in kaart te brengen. Uit associatiestudies is gebleken dat de 1.55 Mb microdeletie vaker teruggevonden wordt in de patiëntengroep dan in de normale populatie en dus verantwoordelijk kan geacht worden voor de verstondelijke beperking en/of dysmorfe kenmerken. Anderzijds heeft deze genoomwijde micro-array techniek tot de identificatie van ongewone submicroscopische herhangschikkingen op chromosoom 4p16 geleid. Deze deleties lieten ons toe om een beter inzicht te verwerven in de bijdrage van de verschillende genen in deze regio tot het WHS fenotype. Zo heeft dit onderzoek geleid tot I) het verder verfijnen van reeds gekende alsook de identificatie nieuwe minimale regio’s van overlap voor de verschillende fenotypische aspecten van het WHS, II) de karakterisatie van een submicroscopische interstitiële deletie gedetecteerd buiten de algemeen aangenomen kritische regio in een patiënt met een mild klinisch WHS fenotype en III) de identificatie van een submicroscopische duplicatie overlappend met de algemene kritische regio in een patiënt met ontwikkelingsvertraging en andere aangeboren afwijkingen.

*In een tweede deel van de thesis*, zijn we gaan inzoomen op de terminale breukpunten in patiënten met WHS/4p16 microdeletie syndroom om meer inzicht te verwerven in het ontstaan en stabilisatie van terminale breuken. We zijn initieel gestart met het ontwikkelen van een nieuwe procedure bestaande uit een combinatie van de hoge resolutie micro-array technologie en telomere specifieke amplificatie gevolgd door Sanger sequenering. Door deze werkmethode toe te passen, hebben we 9 terminale breuken op basenpaar niveau in kaart gebracht en aangetoond dat deze breuken gestabiliseerd worden door directe aanvulling van telomere herhalingen. Bovendien zijn kleine gelijkenissen teruggevonden in fase met de RNA component van telomerase op de locatie van de breuk. Verder hebben computer gebaseerde analyses een aanrijking aan polymerase arrest sites gevonden in de nabijheid van de breukpunten. Tenslotte, is een model gedefinieerd waarbij initieel
de breuken kunnen ontstaan door een pauze in DNA replicatie gevolgd door een template afhankelijke replicatie met behulp van telomerase.

In *een derde deel van de thesis*, hebben we de expressie van genen zowel in de gedeleteerde regio alsook erbuiten flankerend aan de deletie verder bestudeerd. In totaal, zijn expressie profielen van een 20-tal genen verspreid van de telomeer tot 4.5 Mb op chromosoom 4p in kaart gebracht in cellijnen afgeleid van patiënten met WHS/4p16 microdeletie syndroom. Deze data tonen aan dat zowel interstitiële alsook terminale deleties de expressie van genen flankerend aan de deleties kunnen beïnvloeden. Bij verder onderzoek hebben we aangetoond dat telomere flankerende gen expressie, omkeerbaar en DNA methylatie afhankelijk is. Deze gegevens suggereren dat telomere positie-effecten (TPE) een rol kunnen spelen in het ziektebeeld. Tenslotte kan hieruit geconcludeerd worden dat genotype-fenotype correlaties louter berustend op genotype data onvoldoende zijn om een volledig klinisch beeld zoals WHS volledig te kunnen ontrafelen.
REFERENCES

Chapter VII


[9] Shaffer L. G. American College of Medical Genetics guideline on the cytogenetic evaluation of the individual with developmental delay or mental retardation. Genet Med. 2005;7(9):650-4


[34] Stankiewicz P. and Beaudet A. L. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev. 2007;17(3):182-92


[41] Lupski J. R. Genome structural variation and sporadic disease traits. Nat Genet. 2006;38(9):974-6


[73] Loayza D. and De Lange T. POT1 as a terminal transducer of TRF1 telomere length control. Nature. 2003;423(6943):1013-8


[114] Knight Genetics of mental retardation. Karger. 2010


[121] Paylor R. and Lindsay E. Mouse models of 22q11 deletion syndrome. Biol Psychiatry. 2006;59(12):1172-9


References


[140] Inoue K., Osaka H., Thurston V. C., Clarke J. T., Yoneyama A., Rosenbarker L., Bird T. D., Hodes M. E., Shaffer L. G. and Lupski J. R. Genomic rearrangements resulting in PLP1
deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. Am J Hum Genet. 2002;71(4):838-53


PERSONALIA

Name: Femke Hannes
Address: Uilenstraat 64, bus 9 B-1150 Brussels Belgium
Birth date - place: °15 October 1982 - Neerpelt (BE)
Nationality: Belgian

EDUCATION

PhD in Biomedical Sciences in the laboratory for Cytogenetics and Genome research (2005-2010)
Molecular and Developmental Genetics (MDG) programme
University of Leuven
Oude markt 13, 3000 Leuven

Advanced course in Cellular Biotechnology (2004-2005)
(Summa cum laude)
University of Leuven
Oude markt 13, 3000 Leuven

Master degree in biochemical engineering-Focus medical bio-engineering (2000-2004)
(Magna cum laude)
Campus Vesalius
Andreas Vesaliusstraat 13
3000 Leuven
List of publications

2010


2009


2008


2007


2006

2005
ACKNOWLEDGEMENTS-DANKWOORD

Een drukke, ongelooflijk boeiende, onder-de-huid-kruipende ervaring, zo kan ik mijn afgelopen vijf jaar samenvatten. Het onderzoek heeft me doen inzien dat je plezier en kracht moet putten in het zoeken naar verklaringen eerder dan het focussen op het behalen van mooie resultaten. Daarbij hebben vele mensen gedurende deze periode wel op één of andere manier mijn wetenschappelijk pad gekruist. Hoog tijd dus om deze personen in de ‘picture’ te plaatsen en een woordje van dank uit te spreken.

Dit gevarieerde werk was nooit mogelijk geweest zonder de ondersteuning van de diverse groepen in het departement. Met zeer uiteenlopende vragen kon ik bij hen terecht. Vragen zoals rond epigenetica kon terecht bij de groep van Guy Froyen, cloneringen bij de groep van Thijs Baens, transfecties bij de groep van Jan Cools, cellbiologie bij de groep van Pascal Zimmerman/Guido David, DNA extracties en Southern blotting bij de groep van Gert Mattheis, western blotting bij de groep van Sabine Tejpar, lentivirale transducties van de groep van Marinee Chuah/Thierry Vandendriessche, tal van FISH-experimenten bij de routine onder Joris Vermeesch en Peter Vandenberghe en als laatste maar zeker en vast niet de minste, de routine weefselkweek…. Allen, een enorme dank-je-wel!

Bovendien wil ik mezelf richten tot de hele groep van professoren binnen het Centrum Menselijke Erfelijkheid voor het bijwonen van tal van vergaderingen en meer specifiek de donderdag avond vergaderingen. Jullie input is uitermate belangrijk geweest om mijn projecten en mezelf in de goede richting te sturen. Toch wil ik een speciaal woordje van dank uiten aan Griet Van Buggenhout, Koen Devriendt, Thomy de Ravel en Thierry Voet. Jullie goede, hechte samenwerking was onontbeerlijk en leidde tot mooie resultaten en verscheidene publicaties.

Ook wil ik me richten tot de mensen die de technische en administratieve kant hebben behartigd: Gaande van het vangen van spinnen in het labo tot reserveren van vluchten. Ik moest mijn vraag geen twee keer stellen. Altijd waren ze bereid om me te helpen. Bedankt Erik, Rita, Marleen en Veerle.

Tal van personen hebben me niet enkel wetenschappelijk ondersteund, maar ook de ‘fun’-factor voor hun rekening genomen. Beata en Jeroen VH, het woord dat ik verkies om jullie te omschrijven is veelzijdigheid. Ik kon bij jullie terecht met vragen over wetenschap alsook over meer alledaagse zaken zoals paarden of hoe-train-je-jie-collega. Tezamen hebben we heel wat afgelachen. Liesbeth B, vanaf het begin stonden we in dezelfde schoentjes. We waren voorbestemd om samen doorheen het proces van doctoreren te gaan. Je maturiteit heeft mij ook geleerd om volwassen te worden en van tijd tot tijd te relateren. Katrien VR, ik was keer op keer geboeid door je verhalen en de grote dosis energie die je meebreng naar het labo. Zelfs na de moedige nachtjes voor het printen van je thesis. Barbara, jij stond er altijd wel om me met een glimlach voort te helpen. Evelyne en An C, jullie talent voor planning en organisatie is onuitputtelijk. Onze fameuze teambuilding tweedaagse was daar een goed voorbeeld van. Caroline R, bedankt om me altijd te helpen wat betreft de sequencing. Sigrun, een echte
Limburgse in hart en nieren. Kortom, allemaal straffe wetenschappers met hun eigen visie die ik allen hoop in de toekomst nog eens terug te zien.

Ook Rasoul, Rob, Bart, Jef en Veerle, jullie wil ik bedanken om mij altijd met open armen te ontvangen met natuurlijk de nodige bijkomende dosis suiker en humor.

Ook over de studenten die ik kon begeleiden, heb ik nooit mogen klagen, maar er was er één die er met kop en nek boven uitstak. Greet, ik ben blij dat je hebt besloten om voor onze groep te blijven werken. Ik ben ervan overtuigd dat je je stempel zal kunnen drukken op het werk in Joris zijn groep.

Natuurlijk kan ik dit dankwoord niet beëindigen zonder mijn ouders en vrienden te bedanken. Mama en papa, jullie zou ik willen bedanken voor de opvoeding die ik als kind heb genoten en ook de mogelijkheid die ik heb gekregen om me te ontplooien. Alles wat ik van jullie heb geleerd, heb ik gebruikt om te komen waar ik de dag van vandaag sta. Ik hoop dat jullie fijner op me kunnen zijn.

Nico en Karen (Ambertje en Daan), mijn vrienden en mijn tweede thuis. Jullie hebben voor de vele momenten van ontspanning gezorgd. Geen weer was te ruw om naar Rotselaar te rijden en de barbecue aan te steken om onze magen te vullen. En Nico, ‘mijn hero’, ervaring heeft me geleerd dat je altijd paraat zal staan om me te beschermen 😊.

Ook zou ik de vriendengroep in Brussel willen bedanken. Zoë en Batte, Steph en Frans, Nima en Jean, Jan en Heidi, Alex en Arnauld, Loïc, Mieke en Ides, jullie zou ik willen bedanken voor de woorden van moed die ik broodnodig had als de tijd wat moeilijker werden. Allen bedankt om er voor mij te zijn!

Joris, je hebt vanaf het prille begin in mij geloofd en me de kans gegeven om te mogen starten in het Centrum Menselijke Erfelijkheid. Uw onvoorwaardelijk enthousiasme voor de wetenschap, alsook uw energieke persoonlijkheid zal me altijd bij blijven. We hebben destijds een poster gecreëerd met de boodschap “Inspirational”. Wel, ik sta hier nu nog steeds volledig achter. Ook mijn co-promoter, Professor Fryns, u heeft dit proces een zekere humane toets gegeven. De mogelijkheid die u bezit om tussen de lijnen te lezen, heeft me van vaak gerust kunnen stellen en de moed gegeven wanneer ik voor hete vuren stond. Bovendien was een schouderklopje er nooit teveel aan. Een oprechte dank-je-wel!

Tot slot, mijn vriend Stéphane, die ruim 4,5 jaar aan mijn zijde heeft gestaan, mij heeft gesteund door dik en dun en steeds weer de moeite heeft genomen om mijn gedachten op een ander spoor te plaatsen. Je eeuwige optimisme, je goedlachs karakter en je enorme lust voor het leven zijn slechts enkele van je sterke eigenschappen. Door de gemeenschappelijke liefde voor verre reizen en het natuurlijk schoon hebben we samen nieuwe continenten en, nog belangrijker, ook steeds weer elkaar ontdekt. Ver van alle zorgen, enkel jij en ik met de rugzak en ons tentje hebben we van tijd tot tijd echt kunnen genieten van het samen zijn…. Dit alles in sterk contrast met het feit dat je me de laatste jaren ook vaak hebt moeten missen. Ik wil je voor deze onvoorwaardelijke steun, begrip en empathie van harte bedanken. Kortom, met jou langs mijn zijde ben ik de grootste geluksvogel!