Katholieke Universiteit Leuven
Group Biomedical Sciences
Faculty of Medicine
Department of Human Genetics
Laboratory for Genetics of Human Development

Improving our insight in the genetic origin

of congenital heart defects
using array comparative genome hybridization

A novel strategy for gene discovery

Bernard Thienpont

Jury:
Promotor: Prof. Dr. K. Devriendt
Copromotor: Prof. Dr. J.R. Vermeesch
Chair: Prof. Dr. B. De Strooper
Secretary: Prof. Dr. D. Lambrechts
Jurymembers: Prof. Dr. A. Rauch
                  Prof. Dr. M. Vikkula
                  Prof. Dr. A. Zwijsen
                  Prof. Dr. T. Vandendriessche

Doctoral Thesis in Medical Sciences

Leuven, 7th of May 2009

Aan mijn vader Marc Thienpont

Het verbeteren van ons inzicht in de genetische oorsprong van aangeboren hartafwijkingen door array CGH
samenvatting

Het chromosomenonderzoek vormt nog steeds een van de hoekstenen van het genetisch onderzoek bij kinderen met een aangeboren hartafwijking. In dit onderzoek worden de chromosomen in delende bloedcellen bekeken onder de microscoop. Op die manier wordt onderzocht of hele chromosomen of delen van chromosomen aanwezig zijn in te veel of te weinig kopijen. In deze studie hebben we onderzocht of chromosoomafwijkingen die niet zichtbaar zijn met de microscoop een frequente oorzaak voor aangeboren hartafwijkingen zijn. We hebben een nieuwe technologie geïntroduceerd ("array CGH") die toelaat om een profiel van het kopijaantal op te stellen van alle chromosomen. We toonden aan dat deze technologie een betrouwbare detectie van chromosoomafwijkingen mogelijk maakt, en dit met een resolutie die veel beter is dan het klassieke chromosomenonderzoek. Het toepassen van deze array CGH techniek in patiënten met een syndromale hartafwijking verhoogt de kans om de onderliggende genetische oorzaak te identificeren sterk: in 20% van deze patiënten wordt een chromosoomafwijking gevonden die onzichtbaar was met de klassieke technieken. Bovendien wordt op deze manier het genetische defect zeer precies afgelijnd waardoor de patiënt een meer specifieke en gepersonaliseerde diagnose krijgt. Een diagnose is van zeer groot belang voor de verdere follow-up van patiënten en hun families, vermits dit toelaat patiënt en ouders accuraat voor te lichten wat betreft herhalingsrisico en de mentale en fysieke ontwikkeling die verwacht kan worden. In bepaalde gevallen kan ook de behandeling van de patiënten verbeterd worden: complicaties die frequent voorkomen bij bepaalde genetische aandoeningen kunnen verhinderd of behandeld worden voordat ze optreden (bijvoorbeeld gehoorsverlies of complicaties aan het hart). We tonen dat de toepassing van deze hoge-resolutie technologieën de identificatie mogelijk maakt van kleine deletie- of duplicatiemutaties in enkelvoudige genen (bijvoorbeeld in het *FOXC1* of het *ATRX* gen). Deze verhoogde resolutie bemoeilijkt echter ook de interpretatie in verband met de oorzakelijkheid van de genetische afwijkingen die aangetroffen worden.

Chromosoomafwijkingen in het DNA van syndromale CHD patiënten wijzen ons de chromosoomregio's waar genen liggen die verantwoordelijk zijn voor de normale ontwikkeling van het hart. Soms weten we reeds welk gen in deze regio hartafwijkingen veroorzaakt. In andere gevallen was nog geen gen bekend en zijn alle genen die in deze regio

gelegen zijn in theorie mogelijk verantwoordelijk voor de hartafwijking. Onze doelstelling was om uit deze verschillende kandidaatgenen het gen te identificeren dat verantwoordelijk is voor de hartafwijking. We hebben geraffineerde strategieën ontwikkeld om de beste kandidaatgenen te selecteren. Hierbij werd gebruik gemaakt van de vele grote gegevensbanken die recent publiekelijk beschikbaar zijn gesteld. De 24 beste kandidaatgenen (afkomstig uit in totaal 6 chromosoomregio's) die via deze computeralgoritmes geïdentificeerd werden hebben we verder onderzocht. Ons uitgangspunt was dat het oorzakelijk gen voor een hartafwijking in het hart actief moet zijn tijdens de ontwikkeling van het embryo. Daarom werd de activiteit van alle genen onderzocht in verschillende stadia van de ontwikkeling van de zebravis. Slecht 2 van de 24 geselecteerde kandidaatgenen waren specifiek in het hart actief: *BMP4* and *HAND2*. Aangezien studies in de muis voor beide genen reeds aantoonden dat ze essentieel zijn voor de hartontwikkeling, zijn dit inderdaad excellente kandidaten om de hartafwijkingen van de patiënten te verklaren.

Bij één persoon met een hartafwijking werd op chromosoom 6 een bijkomende interessante regio geïdentificeerd. *TAB2* was daar het beste kandidaatgen volgens het computeralgoritme. Dit gen is bovendien afwijkend bij verschillende hartpatiënten die een afwijking op chromosoom 6 dragen. Wanneer we de activiteit van dit gen onderzochten in zebravis embryo's, bleek ook dat *TAB2* actief is in het ontwikkelende hart. Studies in muizen toonden nog geen betrokkenheid van dit gen in de hartontwikkeling, hoewel een verstoring van dit gen een hoge sterfte veroorzaakt vlak na de geboorte, gelijkaardig aan de verstoring in de patiënten. We hebben dit gen op een gelijkaardige manier uitgeschakeld in de zebravis, wat ook daar leidde tot ontwikkelingsafwijkingen. Bij 100 andere hartpatiënten konden we geen fouten in dit gen terugvinden, maar een Deense onderzoeksgroep waar we mee samenwerken vond wel een verstoring van dit gen in een familie met hartafwijkingen. Dit toont aan dat verstoring van *TAB2* een zeldzame oorzaak is voor hartafwijkingen.

Improving our insight in the genetic origin of congenital heart defects using array comparative genome hybridization

synopsis

Chromosome investigations are still an important part of the genetic investigations in children with congenital heart defects (CHDs). For this, chromosomes from dividing white blood cells are investigated under a microscope to check if certain chromosomes or parts of chromosomes are present in too many or too little copies. In the present work we have investigated whether submicroscopic chromosome imbalances are a frequent cause for CHDs. We introduced a novel genome-wide copy number profiling technique (aCGH) and showed that it enables a reliable detection of such imbalances at a resolution far surpassing the resolution of microscopic chromosome investigations. The application of this technique in patients with a syndromic CHD greatly enhances the chance of finding an etiological diagnosis. More precisely, in 20% of them, a disease-causing submicroscopic chromosome imbalance can be demonstrated. The correct delineation of chromosome aberrations by aCGH also entails a more accurate characterization of the genotype of the patient, permitting a more personalized, specific genetic diagnosis. A diagnosis is of the utmost importance for the follow-up of the patients and their families, as it allows more correct counseling of patient and parents regarding recurrence risks and the mental and physical development that can be expected. In some cases it also impacts treatment of the patients as complications associated with certain genetic conditions can be prevented or managed from a subclinical stage (e.g. hearing loss or cardiac complications). We showed that the application of higher-resolution platforms enables the genome-wide identification of indel mutations of single genes (e.g. in *FOXC1* or *ATRX*), but that this increased resolution is accompanied by an unexpected complexity in the evaluation of their causality.

The identification of submicroscopic indels in the DNA of syndromic CHD patients pinpoints regions that contain a gene responsible for heart development. We detected many imbalances that affect genes known to cause CHDs. Accordingly, imbalances identified in this way that do not affect known genes for CHDs pinpoint novel candidate regions. The use of advanced database mining strategies like ENDEAVOUR aids in ranking and selecting valuable candidate genes from these loci, and we showed that there is room for improvement by tailoring these tools to the needs of the underlying clinical or scientific question. We have

used expression analyses in zebrafish embryos to identify the most valuable candidates from a group of high-ranked candidate genes. Genes that showed a specific expression in the developing zebrafish heart were considered good candidate genes. Only 2 out of 24 candidate genes displayed such a pattern: *BMP4* and *HAND2*. Both genes are excellent candidates as they were already known to be involved in mammalian heart development through studies in mice.

In one person with a CHD we detected a deletion on the long arm of chromosome 6. In this region, our algorithm identified *TAB2* as the best candidate gene for causing heart defects. This gene is deleted in multiple patients with CHDs, is located in the critical deletion region and is ranked first as a candidate gene amongst over 100 genes from the region. Loss of a copy of this gene is described to be associated with a high mortality in newborn mice, and we have shown that it is associated with developmental defects in zebrafish. Although we could not identify pathogenic mutations in a group of 100 patients with isolated heart defects, others did find a disruption of this gene in 3 members of a small family that have heart defects. This shows that loss of a copy of *TAB2* is a rare cause of CHDs.

Table of content

CHAPTER 1   ETIOLOGY OF HUMAN HEART DEFECTS: INTRODUCTION

1. definition and clinical classifications

Congenital heart defects (CHDs) are a group of diseases characterized by a structural anomaly of the heart that is present at birth. Although they are often associated with severe morbidity and even mortality, they do not necessarily manifest themselves in the neonatal period and can even remain benign throughout life. The CHDs are distinguished from the cardiomyopathies, which are diseases of the heart muscle, and the cardiac rhythm disorders, diseases of the heart's pace generating and conduction system. These three cardiac diseases can and often will co-occur in one patient, and are sometimes interlinked. For example, structural heart defects can cause heart muscle overgrowth or dilatation, which in turn can lead to rhythm abnormalities. These three diseases can also co-occur due to a common underlying cause, without one being the functional consequence of the other. For example, *NKX2-5* mutation can provoke conduction abnormalities and/or structural defects[1], and *ACTC* mutations can cause cardiomyopathies and/or isolated septal defects[2,3].

CHDs arise from errors in cardiac development. Although attempts have been made to classify CHDs according to specific errors in cardiac developmental processes, a descriptive anatomical classification is generally favored, since this does not invoke the assumption that one knows what part of development is at the base of the CHD. Anatomically, various defects in the septa, valves, inflow and outflow regions of the heart are distinguished. An example of such a classification system is the AEPC list of cardiac codes (www.aepc.org/aepc/nid/European_Paediatric_Cardiac_Coding).

CHDs are found in almost 1% of live births, and the various types of CHDs are found at differing frequencies, with septal defects being the most common (Table 1). From a clinical perspective, one can distinguish syndromic and non-syndromic (or isolated) CHDs. This distinction is a practical one made by clinicians, as it reflects the risk for a patient to develop further complications unrelated to the CHD and since it is often the first step to reach an etiological diagnosis[4]. A first group of individuals considered to have a syndromic CHD are those with multiple congenital anomalies (MCA), i.e. having a second major congenital abnormality which is unrelated to the CHD (for example malformations of the kidney, skeleton or brain). Major malformations have clear medical, functional or esthetic

1

consequences for the individual[5]. Other individuals present multiple (three or more) minor anomalies (i.e. physical variants without functional or esthetic consequences that are found rarely in the normal population[6]). Depending on what types of heart defects are included in the study, syndromic CHDs make up an estimated 12-30% of all CHDs[4,7,8]. Syndromic and isolated CHDs are thought to have a different etiology: syndromes mostly arise from a single and often genetic cause, whilst isolated CHDs often have a multifactorial cause, arising at the interplay between multiple genetic and environmental factors. However, as will be outlined below, sometimes syndromes are caused by potent environmental factors like teratogens or by multiple genetic factors, and also isolated CHDs can in rare instances be ascribed to a single genetic cause. Of note, the distinction between syndromic and isolated CHDs is arbitrary since currently no universally used definitions for minor physical anomalies or firm criteria for dysmorphism exist. Moreover syndromic versus isolated CHDs not necessarily implicate different genes: for example, *TBX5* mutations cause CHDs and hand malformations in most patients (syndromic CHD), but in some patients with the same mutation, a CHD is the only apparent major expression (isolated CHD). CHDs occur at a frequency of around 1%, but the fraction of syndromic forms varies amongst different types of heart defects.

Table 1: Incidence of CHDs and frequency of associated malformations, averaged across multiple studies. (adapted from Hoffman & Kaplan (2002)[9], Greenwood (1975) [10] and Pradat *et al.* (2003)[8]). NA: not available.

CHD type	Incidence at birth [per 100 000 live births]	Frequency of associated malformations
Ventricular septal defect	357	27-31%
Patent ductus arteriosus	80	36%
Atrial septal defect	94	36-53%
Atrioventricular septal defect	35	41%
Pulmonary stenosis	73	19-29%
Aortic stenosis	40	17%
Aortic coarctation	41	23-26%
Tetralogy of Fallot	42	30-32%
d-type transposition of the great arteries	32	8-13%
Hypoplastic right heart	22	NA
Tricuspid valve atresia	8	18%
Ebstein anomaly	11	18%
Pulmonary atresia	13	21%
Hypoplastic left heart	27	12-21%
Truncus arteriosus	11	14-46%
Double outlet right ventricle	16	17-34%
Single ventricle	11	20-37%
Totally anomalous pulmonary venous connection	9	14-22%
All cyanotic	139	NA
All CHD*	960	25-30%
Bicuspid aortic valve	1355	NA

*CHDs do not include BAVs, isolated PAPVCs and silent PDAs

2. environmental causes

In a minority of cases, environmental factors are shown to cause the cardiac defect. Well documented examples include development of CHDs after antenatal exposure to teratogens such as alcohol or antiepileptic drugs. These have been reviewed extensively by Jenkins and colleagues[11].

2.1. diabetes

One of the major teratogenic risk factors for cardiovascular malformation is pregestational diabetes. It is associated with an up to 18 fold increase in the risk of a CHD at birth[11]. Overrepresented CHD subtypes include outflow tract defects (like transposition of the great arteries (TGA) and patent truncus arteriosus (PTA)), complete atrioventicular septal defect (AVSD) and also cardiomyopathies. Outflow tract obstruction defects seem less typical[12,13]. Also extra-cardiac malformations are overrepresented in children from mothers with uncontrolled diabetes, similar to those seen in the VACTERL association[13], i.e. Vertebral anomalies, Anal atresia, Cardiovascular anomalies, Tracheoesophageal fistula, Esophageal atresia, Renal and (preaxial) Limb anomalies (VACTERL).

2.2. twinning

A risk factor of a somewhat different nature in the fetal environment is the presence of a second fetus: twin pregnancies sharing the same chorion confer an increased risk for CHDs. This risk factor was the subject of a recent review by Manning[14]. The CHD risk is influenced by the chorionic and amniotic structures. The risk of CHDs in dichorionic monozygotic twins is unknown. Monochorionic diamniotic (MC/DA) twins however are at an increased risk for CHDs (5-7%), and the concordance rate in MC/DA twins for the presence of a CHD is about 25%[15,16]. One factor involved in this abnormal heart development is altered hemodynamics during development. Placental vascular anastomoses leading to twin-to-twin transfusion (TTT) syndrome are strongly implicated in this change in hemodynamics. A systematic literature review on the TTT effect showed an increase in CHD frequency from 3.4% in MC/DA twins without TTT to 8% in MC/DA twins with TTT[16]. In monoamniotic (MA) twins the risk for CHDs is thought to be even higher. However, since they occur rarely, reliable risk estimates are lacking. Manning and Archer[15] reported a CHD in 4 out of 7 MA twins. Interestingly, these included 2 right atrial isomerisms, a very rare defect. This suggests that

certain CHDs could also be related to the twinning process itself, especially in case of a late embryonic timing of the twinning event[14].

2.3. twin studies for CHDs

The frequency of CHDs in dizygotic twin pregnancies is thought to be 1% per fetus, similar to the population risk[14]. However, the concordance for CHDs in dizygotic twin pairs is estimated at 13.6%[17], which is higher than the recurrence risk for the nearest non-twin sibling (4% in the same study). This increased recurrence risk can either be secondary to the twinning (see above) and the concomitant altered fetal environment (*in utero* 'crowding') or a reflection of the more equal effects of external, environmental factors on cardiac development compared to non-twin siblings.

It should be evident from the above-mentioned observations that studies to unveil the genetic contributions to CHDs using twin cohorts do not provide unambiguous answers. However, it remains a challenge to explain the discordance in cardiac development observed in dichorionic monozygotic twins where altered hemodynamics do not play a role. Possible explanations are postzygotic mutations including differences in copy number variations (CNVs)[18], epigenetic differences originating postzygotically (e.g. differences in X-inactivation patterns) and stochastic factors. The latter include pure stochastic factors during cardiac morphogenesis[19,20] but also chance "catastrophic" events which may for instance lead to disorders such as oculo-auriculo-vertebral spectrum[21], a condition suspected to be caused by vascular disruption[22].

3. genetic causes

3.1. familial CHDs

Population studies of CHDs show that they occur at a rate of a little under 1%. Since they are more frequent in first degree relatives of individuals with a CHD, genetic factors likely contribute to their etiology. In Table 2, a compilation of most CHD family studies published in the last 30 years is provided. Several caveats apply to this table:

- Most studies describe only a small number of patients, and are thus likely to result in ascertainment bias.

- Some have been performed by mailing questionnaires or interview. They likely underestimate the frequency of CHDs: not all probands are aware of the cardiac status of all their relatives, and the cardiac status is sometimes unknown (an undiagnosed or subclinical CHD). Generally, such studies resulted in lower frequency estimates, and only some perform corrections by investigating also a control population.

- Mailing questionnaires are dependent on reply rates that may be influenced by factors like gender and by the presence of an affected relative.

- Included CHDs vary between studies. Where possible, functional, spontaneously resolving and subclinical CHDs were excluded from the table. However, they represent a significant fraction of recurrences: Whittemore *et al*. investigated the presence of CHDs in offspring of CHD parents directly by ECG, echocardiography or catheterization [23]. They found that the CHD was subclinical in almost half of the offspring with a CHD (7.5 of the 14.3%). The corresponding frequencies were adjusted to assemble Table 2. Nevertheless, their observation suggests that many relatives of CHD patients have undiagnosed, subclinical CHDs that are unrecognized in most studies. Their importance depends on their use: in a clinical setting, benign CHDs can probably be ignored in recurrence risk estimates; in a research setting, they can be essential for evaluation of the mode of inheritance of a CHD.

- Population studies classify complex CHDs as a single type. The obtained frequencies are thus probably inaccurate and heterogeneous estimates of the relative recurrence risk for most CHDs.

Some CHDs types are associated with a higher risk for offspring than for siblings of probands (Table 2), suggesting an autosomal dominant inheritance pattern. Examples include AVSD

and tetralogy of Fallot. For most other types, such as hypoplastic left heart (HLH), CHDs are more frequent in sibs (30%) than in parents (10.5%) of the proband. Also for aortic coarctation patients, sibs more often also have a CHD (12.8%) than parents (6.5%) (these frequencies were not included in Table 2 as they include Bicuspid Aortic Valves (BAV) as CHDs)[24]. However, not a single large study considers all first degree relatives of CHD patients. These numbers should therefore be interpreted with caution. The contrast between the incidence in children and in parents of CHD patients (Table 2) may seem contradictory, as it suggests different vertical transmission rates. However, it probably reflects the historical mortality and morbidity leading to a reduced reproductive fitness that was and is associated with CHDs. The high frequency of *de novo* pathogenic mutations that seems a logical consequence of the reduced fitness probably offers an additional explanation. These data moreover suggest that, together with the increasing CHD

Table 2: Frequency of congenital heart defects in the first degree relatives of probands with selected types of CHDs. Frequency estimates in brothers and sisters were combined into sibs, as this distinction was mostly not mentioned in the original study. CHD frequencies in parents of a proband were frequently subdivided between mothers and fathers and are also listed as such. CHDs in children of probands were rarely subdivided between those in sons and in daughters and are thus grouped. Risks for children of an affected adult did distinguish between affected mothers and fathers, and are listed here as mother and father proband. Only reports published after 1975 were included [23,25-46], as the inadequate cardiac imaging techniques available before this date may lead to an underestimate of CHD frequencies. BAVs were not included as a CHD, although they are overrepresented in relatives of hypoplastic left heart patients.

CHD type	Sibs	Parents	mothers	fathers	Children	mother is proband	father is proband
Aortic stenosis	0/12 0.00%	1/30 3.33%	1/15 6.67%	0/15 0.00%	28/632 4.43%	15/258 5.81%	14/392 3.57%
Atrial septal defect	41/648 6.33%	20/1288 1.55%	16/644 2.48%	4/644 0.62%	31/515 6.02%	25/486 5.14%	
Atrioventricular septal defect	1/112 0.89%				9/106 8.49%	21/209 10.05%	1/30 3.33%
Coarctation of the aorta	3/37 8.11%	3/108 2.78%	2/54 3.70%	1/54 1.85%	16/407 3.93%	14/308 4.55%	
Hypoplastic left heart	6/61 9.84%	1/125 0.80%	1/64 1.56%	0/61 0.00%			
Pulmonic stenosis	3/33 9.09%	2/116 1.72%	0/58 0.00%	2/58 3.45%	32/760 4.21%	19/450 4.22%	12/303 3.96%
Truncus arteriosus	7/107 6.54%	2/106 1.89%	0/4 0.00%	0/4 0.00%			
Transposition of the great arteries	20/1291 1.55%	7/1568 0.45%	4/784 0.51%	3/784 0.38%	0/56 0.00%	2/218 0.92%	0/6 0.00%
Tetralogy of Fallot	15/657 2.28%	1/278 0.36%	1/139 0.72%	0/139 0.00%	17/650 2.62%	13/439 2.96%	4/272 1.47%
Ventricular septal defect	5/95 5.26%	4/406 0.99%	3/203 1.48%	1/203 0.49%	45/1188 3.79%	23/591 3.89%	20/497 4.02%
Congenital heart defect	55/1586 3.47%	33/3145 1.05%	14/1388 1.01%	8/1388 0.58%	133/3294 4.04%	108/2451 4.41%	37/1119 3.31%

prevalence (because the improved survival rates following advances in surgical management of CHDs) the incidence of CHDs will increase: more CHD patients will have children. Consequently, the genetic, more penetrant causes of CHDs will increase, rendering a thorough understanding of the etiology of CHDs even more crucial.

The CHD risk for children of mothers with a CHD (4.41%) is consistently higher than for those of fathers with a CHD (3.31%) (Table 2, ref [25,27,37]). This risk seems especially higher for AVSD (10.05% versus 3.33%). There are several possible explanations for the increased transmission of CHD through females:

• One likely explanation is a lower susceptibility of females to develop a CHD. Indeed, the frequency of CHDs in females in lower than in males (CHD male/ female ratio = 1.25 vs population male/female = ratio 1.06, see also the discussion below on X-linked inheritance in chapter 0)[8]. Females apparently need an increased genetic/environmental burden to develop a CHD. Since this higher genetic burden is transferred to subsequent generations, it confers an increased recurrence risk through a mother with a CHD compared to through a father with a CHD.

• This may indicate a potential involvement of the mitochondrial genome. No mitochondrial mutations causing CHD have been identified so far. Especially for AVSDs, this might be an explanation for the major increase in transmission through females.

• Also parental imprinting of CHD causing genes could explain the increased transmission through the female germline. No evidence exists for such a mechanism.

• Finally, one could envisage transmitted epigenetic modifications that are caused by the altered environment in the mother or father with a CHD.

Also potential confounders should be taken into account:

▪ These studies are often performed using mailing questionnaires, and the response rate for fathers is consistently lower than for mothers.

▪ Also non-paternity might be a confounder. However, this cannot explain the entire increase in risk, and neither the observed CHD-dependent risk increase.

3.2. classifications of genetic causes

Several classifications exist for genetically caused CHDs. One can classify mutations based on:

• the genetic lesion: small mutations (monogenetic) versus chromosomal aberrations

- the inheritance pattern and the mechanism of pathogenesis of the mutation

- the number of loci involved: monogenic, oligogenic, polygenic

3.2.1. monogenetic and chromosomal CHDs

3.2.1.1.   difference between monogenetic and chromosomal CHDs

One classification is based on the genetic lesion, and distinguishes monogenetic and chromosomal CHDs. This distinction is historical and was based on the detection technique, i.e. the presence of a chromosomal aberration detected by karyotyping versus the detection of a small change at the nucleotide level detected by sequencing. This distinction between detection techniques is rapidly becoming blurred. Novel techniques for genome-wide copy-number determination such as aCGH (array comparative genomic hybridization) are enhancing the resolution of karyotyping and can detect chromosome imbalances that were previously only detectable by polymerase chain reaction (PCR) and sequence analysis. Moreover, the advent of genome-wide sequencing promises to also move the boundary of sequencing techniques into the field of chromosome aberration detection[47,48]. Despite this blurring of the historical boundary between detection techniques for monogenic and chromosomal CHDs, some of their characteristic features still warrant a distinction.

3.2.1.2.   chromosomal CHDs

Chromosomal CHDs are caused by large chromosome imbalances that are rare and affect multiple genes. Typical examples of chromosomal CHDs include:

- AVSDs in 40% of trisomy 21 (Down syndrome)[49]

- ASD, VSD and PS in 33% of terminal deletions of chromosome 4p (Wolf-Hirschhorn syndrome)[50-52]

- aortic abnormalities (CoA, AS, BAV and transverse arch elongation) in 50% of monosomy X (Turner syndrome)[53,54]

These types of genetic lesions are practically absent from the normal population: they are mostly incompatible with a normal development and therefore result in a syndromic presentation. Consequently, they occur mostly *de novo*. They can cause contiguous gene syndromes, where different parts of the patients phenotype can be explained by different genes affected by the imbalance. A classic example is Williams-Beuren syndrome (MIM 194050), caused by the recurrent deletion of chromosome 7q11.23. Haplo-insufficiency of one of the genes in the deleted region (*ELN*) is responsible for the supravalvar aortic stenosis

that is commonly present in these patients[55,56], while one or more other genes (*LIMK1*, *GTF2IRD1*, *GTF2I* and *CYLN2*) are responsible for the behavioral and developmental characteristics and dysmorphic features of patients[57]. Studies using aCGH have shown that insertions and deletions (indels) beyond the resolution of standard karyotyping can also cause syndromic CHDs[58]. These indels can affect one or more genes, and demonstrate the limitations of this distinction.

3.2.1.3. monogenetic CHDs

Monogenetic CHDs are (historically) considered to be caused by small nucleotide changes (<1kb) that affect a single gene. Each individual carries such small nucleotide variants throughout his genome, on average 1 every 1000 bp, and these typically are polymorphisms. Only a subset of these are disease-causing, when they functionally affect a gene critical for heart development. When this gene is also critical for other developmental processes, mutations can result in a syndromic CHD.

3.2.2. inheritance pattern of the CHD

Another classification of genetic causes of CHDs is based on the pattern of inheritance or segregation of the CHD. One can discriminate autosomal dominant, autosomal recessive, X-linked, (Y-linked), mitochondrial, and non-traditional patterns of inheritance. Examples of non-traditional patterns of inheritance include mosaicism, genomic imprinting or unstable mutations (e.g. myotonic dystrophy). This classification relies on a high penetrance of the CHD; otherwise no pattern of inheritance can reliably be deduced from the pedigree.

3.2.2.1. Autosomal dominant

Most mutations identified in CHD patients are inherited in an autosomal dominant fashion, meaning that the mutation affects a single autosomal allele (38/51 different genes described in July 2008 in a database on gene mutations associated with CHDs (*CHDWiki*) accessible at http://homes.esat.kuleuven.be/~bioiuser/chdwiki/). Gene mutations can affect gene function in different ways. This is most obvious in autosomal dominantly inherited disorders, and the different possible functional consequences of a gene mutation are therefore discussed below.

9

3.2.2.1.1. loss-of-function

The majority (22/38) of autosomal dominant gene mutations probably affects gene function through haploinsufficiency: the mutant allele is no longer functional and the remaining functional allele is insufficient to produce a normal function of the gene product during development. Genes susceptible to haploinsufficiency ("*dosage sensitive*" genes) typically belong to categories such as signaling molecules (4/21), transcription factors (11/21) and modulators of transcription, e.g. chromatin remodeling molecules (6/21) where a precise dosage is critical for a normal function. This also explains why both an increased and a decreased dosage of the gene may lead to a phenotype, as for instance seen for *TBX1*: both deletions and duplications of the 22q11.2 region have been found in CHD patients[59], and similarly both mutations that decrease (loss-of-function) and increase (gain-of-function) *TBX1* activity have been shown to cause DiGeorge syndrome[60].

3.2.2.1.2. dominant negative effect

Mutations causing dominantly transmitted CHDs can also function through a dominant negative effect, where the mutant protein inhibits the normal function of the wild-type protein formed by the other allele, e.g. during a process of dimerization. This has been described for some *JAG1* mutations in Alagille syndrome[61], and certain mutations in *TFAP2B* causing Char syndrome[62]. Other examples include genes encoding structural proteins of the heart muscle, such as *MYH6*, *MYH11* or *ACTC*[2,3,63,64]. Mutations in these sarcomere components can cause sarcomere disarray and concomitant muscle dysfunction. It has been suggested that this reduced cardiac muscle function may cause hemodynamic changes during development, resulting in CHDs[65].

3.2.2.1.3. gain-of-function

A third mechanism for the dominant effect of a mutation is through gain-of-function. In this case, the mutant protein acquires a new function causing it to be active ectopically or at a time where it normally is not active. Activating mutations in different components of the Ras/MAPK pathway are known to cause Noonan syndrome and related disorders such as cardio-facio-cutaneous (CFC) and Costello syndrome[66,67]. Most mutations in these genes alleviate the corresponding protein from its normal auto-inhibitive properties and encode a constitutively active form. Other examples of gain-of-function mutations in CHDs include rare instances of gain-of-function mutations in *TBX1* and *TBX5*[60,68].

3.2.2.2. autosomal recessive

When both alleles of an autosomal gene need to be mutated for a phenotype to become manifest, the phenotype will segregate in an autosomal recessive fashion. The patient carries the mutation in a homozygous (or compound heterozygous) form and usually both parents carry a mutation on one allele. Recessive mutations are usually loss-of-function mutations.

In some populations, recessive disorders are more frequent. In isolated populations (especially those that originate from a small number of ancestors) a few otherwise rare mutations can be present frequently, and are referred to as *founder mutations*. Also in children from consanguineous relationships recessive disorders are more frequent, as they are homozygous for multiple loci of their genome. CHDs are more frequent in children from consanguineous parents, with a reported relative risk of 1.35 to 1.8 for first cousin parents compared to non-consanguineous couples from the same region (Table 3)[69-72]. This can be explained by an increased genetic vulnerability of these children for any multifactorial malformation, or by the increased risk for an autosomal recessive disorder. Especially ASDs are overrepresented in children from consanguineous marriages[69,70]. Similarly, an increased frequency of consanguinity is seen in children with an ASD, ToF or AS[72]. Interestingly, ASDs are also more frequent in sibs than in parents of isolated ASD patients (6.3% vs 1.6%, Table 2), suggesting an important recessive component in the etiology of ASDs[17].

Table 3: Relative risk for CHDs in children of first cousin couples compared to non-consanguineous couples from the same region. Combined relative risks from three studies conducted in Lebanon and one in Saudi Arabia[69-72].

CHD type	relative risk
atrial septal defect	1.86
coarctation of the aorta	1.71
double outlet right ventricle	1.63
aortic stenosis	1.56
pulmonary atresia	1.54
tetralogy of Fallot	1.51
transposition of the great arteries	1.51
atrioventricular septal defect	1.48
ventricular septal defect	1.43
patent ductus arteriosus	1.39
pulmonic stenosis	1.31
congenital heart defects	**1.40**

There are only few known genetic causes of CHDs with a recessive inheritance pattern (8/51 genes), and practically all (7/8) are syndromic: mutations in *EVC*, *EVC2*, *LBR*, *MGP*, *NPHP3*, *ROR2* and *SLC2A10*. One exception are *NKX2-6* mutations, that were described by

Heathcote and colleagues in 3 patients with an isolated truncus arteriosus from a single consanguineous family[73,74].

3.2.2.3.   X-linked CHDs

3.2.2.3.1.   Mutations on the X chromosome

When a mutation affects a locus on the X chromosome and manifests a phenotype when a single allele is present (i.e. typically in an XY male), the disorder has an X-linked recessive inheritance pattern. Most of these mutations are associated with a loss-of-function.

In CHDWiki (in July 2008), there are six genes on chromosome X that - upon mutation - cause CHDs, and all are associated with a pleiotropic (syndromic) phenotype: *ATRX*, *BCOR*, *CFC1*, *GPC3*, *MID1* and *ZIC3*. Females mostly remain unaffected by these mutations as the functional copy on their other chromosome X is sufficient to sustain normal cardiac development or because the X chromosome carrying the mutation is preferentially inactivated, resulting in expression of only the wildtype allele. However, exceptions have been reported, including mildly affected females carrying a *MID1* mutation[75] and full ATR-X syndrome in a girl carrying an *ATRX* mutation on an X chromosome that was preferentially activated[76].

3.2.2.3.2.   Turner syndrome

Of interest, CHDs are present in approximately 40% of cases with monosomy X (Turner syndrome)[54], and other vascular abnormalities are even more common in this disorder. The most frequently observed cardiovascular malformations in Turner syndrome belong to the spectrum of left outflow tract abnormalities. They typically are elongation of the transverse arch (50%), coarctation of the aorta (12%), aortic valve abnormalities (such as stenosis, BAV and regurgitation) (16%) and also HLH (2-3%) [53,54].

The precise etiology of this increased CHD frequency remains elusive, and whether a single gene or multiple genes are implicated is currently unknown. In a normal 46,XX female, one X chromosome is inactivated, so also in a normal situation most genes are practically always transcribed from a single X chromosome. The absence of a second active chromosome X before inactivation is an unlikely explanation, as inactivation occurs before cardiac development. As these CHDs are more frequent in 45,X females than in 46,XY males and 46,XX females, and the pseudo-autosomal region - present both on the X and Y chromosome - escapes X inactivation, this region is a likely candidate to contain genes

responsible for the CHDs found in Turner syndrome. Also certain other genes across the X chromosome escape X-inactivation[54]. In combination with the sex-specific differences in hormone levels that could regulate their expression, these genes could also be involved in cardiogenesis and the concomitant high incidence of CHDs in Turner syndrome. CHDs seem less common in women with a 46X,i(Xq) (10%)[54], suggesting that the major genes responsible for CHDs lie on the long arm of the X chromosome. However, there is insufficient data for a conclusive genotype-phenotype correlation.

3.2.2.3.3. sex differences in CHD incidence

Sex differences in incidence of specific types of CHD have been observed (Table 4). When males are more frequently affected than females, this may point to a role of X-linked genes. However, other CHDs (ASDs & AVSDs) occur more frequently in females, and this cannot readily be explained by differences in the sex chromosomes. Also for many other multifactorial disorders (e.g. cleft lip and palate, pyloric stenosis, …) differences are observed in males versus females. This points to sex-dependent differences in embryonic development that are associated with different susceptibilities to various disturbances of normal development. The biological basis of these differences is not known but X or Y linked genes could play a role. Also hormonal differences could have an influence during later stages of development by acting on hormone-responsive genes involved in cardiovascular development.

Table 4: male/female incidence ratios of specific isolated CHD types that are significantly varying between sexes. (adapted from Pradat et al., 2003)[8].

CHD Type	Male/female ratio
aortic valve stenosis	2.41
d-type transposition of the great arteries	2.25
hypoplastic left heart	1.74
coarctation of the aorta	1.72
double outlet of the right ventricle	1.64
total anomalous pulmonary-venous return	1.56
pulmonary atresia	1.47
tetralogy of Fallot	1.44
atrial septal defect	0.86
atrioventricular septal defect	0.84

3.2.3. the number of loci involved

3.2.3.1. recurrence risk for CHDs

As mentioned earlier, the recurrence risk for CHDs is not compatible with a major monogenic component in most instances. Moreover, in sporadic (i.e. non-familial) patients with isolated CHDs, the frequency of *de novo* mutations in known genes and of *de novo*

insertion and deletion mutations (indels) in their genome is low[77,78]. These studies - despite having their limitations - suggest that mutations with a large effect and that result in fully penetrant monogenic causes of CHDs are rare. In most CHD patients, more than one locus will probably be responsible for the pathogenesis of the CHD. The contribution of each locus and the number of loci that contribute to CHD are discussed below.

3.2.3.2. effect size, relative risk, penetrance, recurrence risk & number of loci

The *effect size* of a specific mutant allele depends on how disturbing it is for normal heart development. Effect sizes vary between genes (depending on the importance of the gene for heart development) and between mutations of a gene (depending on the effect of the mutation on gene function). Effect size is reflected in the *relative risk* (RR) of the mutation for CHDs: small effect size mutations will rarely cause CHDs and are therefore associated with low RRs. On a population level, mutations with high RRs are incompatible with high population frequencies: they are generally subject to purifying selection and will thus be infrequent in the normal population. Low RRs can sometimes be found frequently in the normal population when they are associated with mutations that occurred many generations ago. New mutations with a low RR will however have a low population frequency (Figure 1). High population frequencies therefore imply mutations with low RRs or without effect.
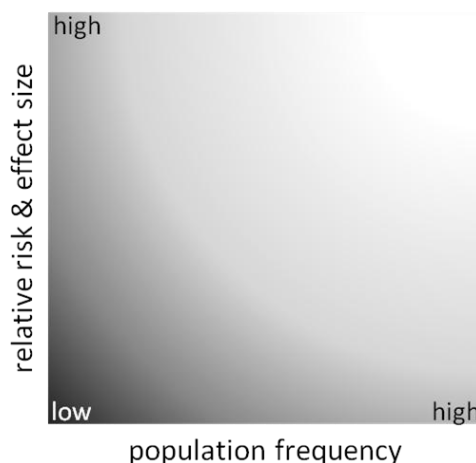
Figure 1: Relation between effect size, relative risk and frequency of a mutation in the population. Color intensities denote the number of CHD causing mutations with the corresponding characteristics in the population (white: none, dark: many). High population frequencies are incompatible with high relative risks and effects sizes. Mutations with lower effect size can persist more easily in a population and become frequent. Vice versa, mutations with high relative risks can exist at a low population frequency. Finally, there are probably a large number of mutations with a low relative risk that are infrequent in the population, as indicated by the dark color.

Penetrance of a genotype is defined by the probability that the genotype will cause a CHD[79], and thus the number of affected versus non-penetrant individuals with the genotype.

For dominant mutations, this corresponds to the RR. Like for RR, penetrance of any mutation largely depends on its effect size.

Recurrence risk refers to the probability that a new family member of the proband (sib, child, nephew, grandchild) will develop a CHD. Recurrence risks mainly depend on the degree of relation to the proband, the inheritance pattern of the CHD (autosomal, recessive, X-linked, reflecting the risk to have the pathogenic genotype), frequency of the phenotype and penetrance of the genotype. Recurrence risk is a term mainly used in clinical genetics, since in the counseling of a family recurrence risks are a major issue. As there is often no cause known for the CHD, empirical recurrence risk estimates are often used.

When the precise genetic etiology of a CHD is known, the recurrence risk depends on the number of loci that are causing the CHD. When many small effect size alleles converged to cause the CHD in the proband, the risk to have the same genetic constellation in a new member of the family (the genetic component of the recurrence risk) is low. When there was however a single highly penetrant mutation causing the CHD, the recurrence risk can be high depending on the inheritance pattern and the relationship to the proband. Given this importance of the number of loci, genetic causes of CHDs are also classified based on the number of alleles that need to co-occur and their effect size. One can discriminate a spectrum ranging from monogenic (a single locus is mutated, and the mutation is associated with a high penetrance of CHDs) over oligogenic (mutations in a few loci converge to disturb heart development[80] to polygenic (many alleles with a small effect size contribute to abnormal heart development). These notions were originally introduced to explain limited recurrence risks in highly genetic disorders. As for most classification systems, the distinction between monogenic-oligogenic-polygenic is partly artificial and arbitrary.

3.2.3.3.   monogenic causes for CHDs and variable expressivity

Mutations that are associated with a CHD but that are not found in the normal population are considered monogenic. Most known mutations that cause isolated CHDs have a high penetrance. This observation is probably subject to a severe ascertainment bias, since many genes identified so far were found via linkage studies in large families, necessarily having a high penetrance.

In most cases, mutations causing isolated and syndromic CHDs are associated with a variable expressivity, meaning that different CHD types or severities can be caused by a

mutation of the same gene. This can sometimes be explained by the type of mutation that is present. For example, loss-of-function mutations in *TBX5* cause septal defects in 70% of patients, while gain-of-function mutations have been described to cause atrial fibrillation but rarely septal defects[68]. However, often the variability can only be explained by modifying factors, either genetic or environmental. A typical example is the 22q11 deletion syndrome, where exactly the same genetics lesion can be found in multiple individuals that are affected to a variable extent and differ in their CHD[81]. Polymorphisms in the *VEGF* gene have been shown to explain part of the variability in susceptibility to a CHD in the del22q11.2 deletion syndrome[82]. Similarly, the increased frequency of AVSD amongst female and black Down syndrome patients suggests sex and race-dependent differences. All this points to a more complex pathogenesis of the phenotype, where genes on other chromosomes likely also influence the phenotypic outcome in these patients[49]. Indeed, mutations in *CRELD1* have been detected sporadically in Down syndrome patients with an AVSD[83]. Despite this variability, it is of interest that specific types of CHD are often associated with distinct syndromes. For clinicians this may aid in the diagnosis of syndromic CHD, and for researchers this can indicate that specific genes or genetic pathways are involved in the development of distinct components of the heart. In Down syndrome for instance, 39% of patients have an AVSD, whereas tetralogy of Fallot (ToF), which is considered to have a different pathogenesis, is observed in 6% of cases[49]. For the del22q11.2 syndrome, we have observed 89 patients with a CHD in a cohort of 147 patients (60%). The majority of these were conotruncal malformations (50% tetralogy of Fallot and 11% truncus arteriosus). Pulmonary stenosis usually makes up 9% of CHD types but was only found in one of 89 del22q11 CHD patients. It can similarly be noted that, even in families where there is only one recurrence of a CHD, there is a high concordance between both CHD types[25]. In conclusion, although single genetic causes for CHDs can result in different CHDs, specific genetic causes are associated with distinct types of CHDs.

3.2.3.4. oligogenic causes for CHDs

An emerging cause for CHDs are the so-called rare inherited variants. All these mutations are usually present below 1% in the normal population but are overrepresented in selected patients. They are typically inherited from a asymptomatic parent and are therefore not believed to be the only cause of CHDs in these patients. Other factors need to be present to

cause CHDs. The few genes that have been implicated in this inheritance mechanism in CHDs are all from the same Nodal pathway: *FOXH1*, *GDF1*, *CFC1* and *TDGF1*. Mutations in these genes have been found in 5% of patients from a group of patients with outflow tract septation defects like transposition of the great arteries, truncus arteriosus, double outlet right ventricle and tetralogy of Fallot[84]. As one would expect, a single patient was found to carry multiple alleles associated with reduced Nodal pathway strength. Further advances in high-throughput screening for mutations in this and other pathways will produce estimates on how often CHDs have an oligogenic cause. An issue in these studies is the evaluation of causality of these mutations. Given the expected low penetrance of each of these mutations, family segregation studies cannot be used to indicate pathogenicity. By using the zebrafish as a biosensor, Roessler and his colleagues demonstrated however that part of these mutations have a detrimental influence on protein function[84]. They moreover showed that these deleterious mutations are significantly overrepresented in a population of CHD patients compared to normal controls.

These mutations were found by a candidate gene approach: all these genes play a role in Nodal signaling, a pathway disturbed in some monogenic cases of CHDs and involved in left-right asymmetry establishment[85]. In order to have an impact on genetic counseling of CHD patients, more data should be gathered to establish the penetrance of these mutations. The contribution of rare inherited variants in other genes and pathways to the etiology of CHDs similarly remains to be investigated. Since this type of inheritance fits the sporadic nature of many CHDs, it is anticipated that many more of these oligogenic causes may still be identified.

3.2.3.5. polygenic and multifactorial causes for CHDs

A third group of causes is called complex, since multiple factors contribute to the development of the disorder. It is assumed that in many instances multiple genetic polymorphisms interact in disturbing normal cardiac development and/or function. In addition, to a lesser extent, the environment may influence this process. Virtually nothing is currently known about genes implicated in this multifactorial model of CHD. One of the few genes implicated is *VEGF*. Low expression haplotypes of this gene are overrepresented in tetralogy of Fallot patients[86].

It is unclear whether accumulation of deleterious polymorphisms in a single individual is as such sufficient for developing CHDs. It has however been shown that certain polymorphisms can modulate the penetrance of CHDs in monogenic disorders. The following genetic constellations can occur[79]:

- a polymorphism in the allele present in *cis* could influence the expression of the mutant protein, resulting in a modulation of phenotypic expression. This mechanism has thus far not been investigated in CHDs, but could be used for the identification of causal genes in common genetic lesions such as trisomy 21.

- a polymorphism in the allele present in *trans* could influence the expression or function of the wild-type protein. This was investigated for the remaining *TBX1* allele in del22q11 patients[87]. However, there was no association between any common variant and the presence of a CHD in the del22q11 patients

- a polymorphism elsewhere in the genome could contribute to the susceptibility of the patient to CHDs. It was shown that low expression *VEGF* haplotypes are overrepresented in del22q11 patients with a CHD[82].

To identify other polygenic causes of CHDs, we anticipate that genome-wide association studies will make a major contribution to the identification of novel relevant players in human heart development. Moreover, given the overlap between polygenic, oligogenic and monogenic causes of CHDs, such a genome-wide approach to identify polygenic causes may allow for the identification of mutations in the same gene with a much higher penetrance and clinical significance.

As mentioned, the frequency with which a specific gene mutation is associated with CHDs (i.e. the penetrance) will vary according to its effect size. The contribution of other influences such as environmental and stochastic factors will be relatively high when only small genetic effects contribute. A purely polygenic cause of a CHD is therefore much less likely than a multifactorial cause. In these cases, genetic and environmental influences both have an impact on cardiac development, either independently or cumulatively. An example of such a cumulative impact is the association between polymorphisms in the *NNMT* gene (encoding a protein that catalyzes nicotinamide), nicotinamide intake and CHDs[88]. Presence of the *NNMT* A-allele combined with maternal low nicotinamide intake at peri-conception confers a significantly increased risk for CHDs.

3.3. identifying and confirming genetic causes

A major challenge remains the confirmation of the role of a specific positional or functional candidate gene in the pathogenesis of CHD. The traditional approach of mutations analysis in patients faces many limitations:

- Given the genetic heterogeneity of causes for CHDs, large cohorts of patients need to be screened. Moreover, since expression of a mutation in a given gene can be variable, other types of CHD need to be included in the mutation screen than that of the index patient.

- The genetic component in CHD can be highly diverse, ranging from monogenic over oligogenic and polygenic to multifactorial. Current evidence shows that *de novo* mutations are exceptional in non-syndromic CHD patients. Rather, an increasing number of studies describe a higher number of rare (inherited) variants in individuals with CHD compared to normal controls. Therefore, in the future, we envision not only mutation screens in large cohorts of patients, but also simultaneous screening of a large number of genes in the same patients and in controls, to uncover how these individuals are genetically sensitized, and which pathways are disturbed in abnormal human heart development.

- Finally, evidence for the involvement of a gene in a multifactorial disorder is obtained from association studies. There is a lack of such studies in the field of CHDs at the present time.

- Mutation screens should not only include sequence analysis but also screening for CNV's.

- Functional studies, either in vitro or in vivo (animal models) will provide further evidence of a functional role of the found sequence alterations.

CHAPTER 2   RESEARCH INTRODUCTION

The aim of our studies was to advance our knowledge on the genetics of CHDs. To reach this aim, we sought to test an integrated, three step approach combining a novel positional cloning strategy, the newest candidate gene prioritization tools and a final validation step in zebrafish as an animal model as well as human mutation analysis. We sought to test the following hypotheses in this study:

1. Cryptic chromosomal insertions and deletions (indels) are a frequent cause of idiopathic syndromic CHDs

2. Identification of submicroscopic indels in syndromic CHD patients enables positional cloning of genes responsible for isolated and syndromic CHDs

3. Genes responsible for CHDs can be efficiently selected from a genome region using a tailored prioritization algorithm

4. Genes responsible for human CHDs can be discriminated by their expression pattern in the developing zebrafish

5. Novel genes responsible for human CHDs can be discovered using the following method: they can be located by positional cloning, selected by prioritization, identified by their expression pattern in zebrafish and validated by sequence analysis in CHD patients.

The following chapters describe how a novel molecular karyotyping technique, array comparative genome hybridization (aCGH), was implemented in the laboratory to establish an etiological diagnosis in patients with idiopathic syndromic CHDs and to serve as a positional cloning strategy. We describe how we performed gene selection using an adapted version of ENDEAVOUR and expression analyses in zebrafish. A selection of promising candidates were investigated in CHD patients by sequence analysis.

CHAPTER 3   MATERIALS AND METHODS

1. reagents and buffers

Alkaline phosphatase [AP] buffer

100mM   NaCl
50mM $MgCl_2$
100mM   Tris-HCl pH 9.5
1mM   levamisol hydrochloride   (Sigma-Aldrich, Bornem, Belgium)
0.1%   PlusOne™ Tween™ 20   (GE Healthcare, Chalfont St. Giles, UK)
H_2O to 50mL, always prepare fresh

2x Danio Lysis buffer

150mM   NaCl
50mM   Hepes, pH 7.5
1mM   EDTA
1%   Nonidet P-40
0.5%   Na-deoxycholate
0.1%   SDS
Add H_2O to 100 mL

Lysis buffer for nuclear protein extraction

0.2-0.5% Nonidet P-40 in RSB

Protein detection blocking buffer

25g   milk powder
50 mL   TBS 10x
450 mL dH_2O
500µL   PlusOne™ Tween™ 20   (GE Healthcare, Chalfont St. Giles, UK)

4% Paraformaldehyde [4% PFA]

2g   PFA
Add 1xPBS to 50mL
Heat to 70°C until dissolved, store maximally 1 week at 4°C

10x Phosphate-buffered Saline [PBS]

80g   NaCl
2g   KCl
17.8g $Na_2HPO_4.2H_2O$
2.4g KH_2PO_4
Add H_2O to 1L

PBS with tween 20 [PBST]

100mL   10x PBS
900mL H_2O
1mL   PlusOne™ Tween™ 20   (GE Healthcare, Chalfont St. Giles, UK)

Ponceau Red

2,5g   Ponceau   (Sigma-Aldrich, Bornem, Belgium)

5 mL   water-free (*glacial*) acetic acid
to 500mL with dH_2O

Radio Immunoprecipitation Assay [RIPA] buffer
50mM   Tris-HCl (pH 7.4)
150mM   NaCl
0.2 %   SDS
1%   Nonidet P-40   (Roche Applied Science, Mannheim, Germany)
protease inhibitor

RSB buffer
10mM   NaCl
10mM   Tris-HCl (pH 7.5)
3mM $MgCl_2$

20x Saline Sodium Citrate [SSC] buffer
In H_2O:
3 M   NaCl
0.3M Na_3Citrate
to pH 7.0 with 1M HCl

10x Tris Buffered Saline [TBS]
1500 mM   NaCl
200 mM   Tris
to pH 7.4 with 1M HCl

Whole mount *in situ* hybridisation Blocking Solution [WISH Blocking mix]
2%   sheep serum (heat inactivated)
2 mg/ml Bovine serum albumin [BSA]
1x PBST

Whole mount *in situ* hybridisation Hybridization Solution minus [WISH Hyb mix⁻]
50 mL   deionised formamide   (VWR, Batavia, Illinois, USA)
25 mL   20x SSC
100 µL   PlusOne™ Tween™ 20   (GE Healthcare)
5mg   heparin
Add H_2O to 100mL

Whole mount *in situ* hybridisation Hybridization Solution [WISH Hyb mix]
50 mg   yeast tRNA   (Sigma-Aldrich, Bornem, Belgium)
100mL   WISH Hyb mix⁻

Whole mount *in situ* hybridisation Staining Solution [WISH staining mix]
22.5µl   Nitro blue tetrazolium chloride [NBT] (Roche Applied Science, Mannheim, Germany)
17.5µl   5-Bromo-4-chloro-3-indolylphosphate [BCIP]   (Roche Applied Science)
Add AP buffer to 5mL

1X Transfer Buffer
50 mL   20x transfer buffer   (Invitrogen, Carlsbad, California, USA)
200 mL methanol

750 mL dH$_2$O (tap)

MOPS
50 mL   20x SDS
950 mL dH$_2$O (tap)

2. aCGH

aCGH was performed using 4 different platforms.

2.1.   1Mb

1Mb aCGH was performed using a micro-array containing 3625 BAC or PAC clones chosen genome-wide at 1Mb intervals (provided by the Wellcome Trust Sanger Institute, Hinxton, UK), plus a set of 58 clones that represent clinically important dosage-sensitive regions. Microarrays were constructed as described by Menten *et al.* [89]. Hybridization was done as described[89] using a three-way hybridization setup: DNA from one patient was labeled in different colors in 2 experiments and cohybridized with differentially labeled DNA from 2 other patients. DNA from both other patients was also differentially labeled and cohybridized in one additional experiment. In this way aCGH for all three patients was performed in a dyeswap replication experiment.

2.2.   full tiling

Full tiling path microarrays consisted of BAC and PAC clones from the Human 32K clone set obtained from the BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute (Oakland, California, http://bacpac.chori.org)[90]. Using these clones, micro-arrays were constructed as described[91]. CGH was performed on the microarrays with a single hybridization or dyeswap replication experiment using described methods[91]. The full tiling array for chromosome X was constructed as described[92].

2.3.   Nimblegen 385K

DNA probe labeling and CGH on 385K microarrays produced by Roche Nimblegen (Madison, Wisconsin, USA) was done as described by the manufacturer. Slides were scanned on a Genepix Axon 4000B microarray scanner (Molecular Devices, Sunnyvale, California, USA) at 5μm resolution with 4 replicates and full laser power. Slide image analysis and feature extraction was done by Nimblescan 2.4.27 at the advised predefined settings. Data was analysed and visualized using the Signalmap 1.9.0.03 software provided by Nimblegen.

2.4. Agilent 244K

2.4.1. DNA labeling

DNA was labeled by the BioPrime® Array CGH Genomic Labeling System (Invitrogen, Carlsbad, California, USA) using Cy3™- and Cy5™-labeled dCTPs (Amersham Biosciences, Boston, Massachusetts, USA) as recommended by the manufacturer with minor modifications: DNA denaturation was done at 98°C for 15 minutes rather than at 95°C for 5 minutes, and the labeling reaction by primer extension was allowed to continue overnight. Labeled DNA probes of patient and control were mixed and purified by ethanol precipitation. These mixed precipitated probes were solubilized in 159.5µL of TE Buffer, 1X, Molecular Grade, pH 8.0 (Promega, Leiden, the Netherlands) and DNA probe concentration and fluorescent label incorporation were quantified in 1.5µL with the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, Delaware, USA) using the microarray application.

2.4.2. probe preparation and hybridization

Probes were further prepared according to the manual of Agilent Human Genome CGH Microarray Kit 244A (Agilent, Santa Clara, California, USA) with a minor modification: some experiments were done using 50mL Human Cot-1 DNA (Roche Applied Sciences, Penzberg, Germany) at 0.2mg/mL rather than 1mg/mL. This reduction in Cot-1 was introduced to save costs and did not result in an obvious reduction of the quality of the results (data not shown). We denatured the probe mix for 10 minutes instead of 3 minutes at 95°C. The probe was hybridized on Agilent 244K microarray slides and incubated for approximately 40 hours in a Hybridization Oven (Agilent, Santa Clara, California, USA) at 65°C under constant rotation (20 rounds per minute, rpm). Agilent 244K microarray slides were reused up to 2 times by removing the hybridized probe from the slide (after image acquisition) using the NimbleGen Array Reuse Kit (Roche Nimblegen, Madison, Wisconsin, USA). Slides reused once yielded similar results as new slides, while slides hybridized for a third time yielded in a decrease in quality as indicated by an 1.5 fold increase in standard deviation of the resulting intensity Log$_2$Ratios. These slides were only used to delineate known aberrations. Slides were washed following manufacturers recommendations and scanned on a Genepix Axon 4000B microarray scanner (Molecular Devices, Sunnyvale, California, USA) or an Agilent Microarray scanner (Agilent, Santa Clara, California, USA) at 5 µm resolution.

2.4.3. Feature extraction and data processing

Extraction of spotted feature properties and signal intensities from the scanned slide images, and 2D lowess and dye intensity normalizations were done using the Agilent Feature Extraction Software (version 9.5.3.1) with the provided CGH-v4_10_Apr08 or the CGH-v4_95_Feb07 protocol and slide and probe annotation files. Resulting txt files were further processed according to the workflow presented in Figure 14 and discussed under section 5.3 (p 75).

3. FISH

Metaphase and interphase FISH was performed using DNA from BACs that were degenerate oligonucleotide primed (DOP) amplified for construction of the microarrays. For FISH, these products were labeled by a second DOP amplification[93] or by a random prime labeling (BioPrime® Array CGH Genomic Labeling System from Invitrogen, Carlsbad, California, USA) with SpectrumOrange™ or SpectrumGreen™ (Vysis, Abbott laboratories, Chicago, Illinois, USA). Hybridization of these clones was performed as described[89]. FISH with cosmid RT100[94] was used to confirm CBP duplication.

Multicolour banding FISH was performed using XCyte 4 according to the manufacturer's instructions (MetaSystems GmbH, Altlussheim, Germany).

4. real-time quantitation of DNA copy number

DNA copy number was quantified by real-time quantitative PCR (rtqPCR) on genomic DNA or on cDNA as described[89]. Locus-specific oligonucleotide primers were designed by the Primer Express software (V2.0.0., Applied Biosystems, Foster City, California, USA) in unique sequences of the human genome. These were selected by

- using repeat-masked sequences (http://www.repeatmasker.org) as input for Primer Express
- performing UCSC In-Silico PCR with the forward and reverse, twice the forward and twice the reverse primer as input. In-Silico PCR was required to yield for each combination respectively one, no and no result.

Oligonucleotides used as primers were synthesized by Eurogentec (Seraing, Belgium). The amplification results and the melting curves were analyzed with the ABI Prism 7000 SDS Software version 1.1 (Applied Biosystems, Lennik, Belgium). The DNA levels were normalized

to the number of measured copies of the p53 gene and relative differences were calculated according to the $2^{-\Delta\Delta CT}$ method[95]. Primer oligonucleotide sequences are described in Table 5.

Table 5: sequences of oligonucleotide primers used for rtqPCR on genomic DNA

Name	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
16p13_A	CCTTCATTGCCACTTGTGATACTG	TCAACGCTTTCCACTGACAAA
16p13_B	GGTCCCAGGGCTCTTAGTTTAGTT	AGTAGGTTCCCCTTTCACACATG
16p13_C	GCCAAGCACTGCCAAGAAA	TGGCGGAGCTTGTGTTTGA
16p13_D	CGGCCGAGAGCATCCA	TTCCGATTCGAGGGAGATTCT
16p13_E	TCTCATCCATCAGCCCAACTG	GCAGCCAGCAAGGGAAGAC
16p13_F	CCAGATTCCCACAGCACAGA	GATGGCCAAGTTGGATGCTT
ATRX_e1	GGCGCCCAGCAATCAC	CATGGGCTCAGCGGTCAT
ATRX_e2	TCGTGGAGGAGAACTTGTTTCTT	GTTGAATACATTGGTGCAGAAGCT
ATRX_e29	CCTACCTCACATTAGTTTCATCATTAAATT	GACTATTACCGTTTAGATGGTTCCACTAC
ATRX_e30	TCCAAAGCGATAAACTCTGAATATACTCT	AATCTGGTAGCTGCTAATCGAGTAATT
ATRX_e35	TCATTGTAGATTGCTTCTCTTCGTTT	CCCAAGTACAGGCGTTAGCATT
ATRX_e36	TTTTAAGTCAAGCAGGGCTACCA	GGAGGCCAGGTAATAAGGAAGCT
ATRX_i1-2	TCCACAGAACCACTTTAAGTAGCAA	GCCCAATCCTGCTTCTCCTT
ATRX_i2-3	AGGGTCTCACTCCATTGTCCAT	CCGGGAGGTTGAGGCTAAA
ATRX_i28-29	CATGAAAGATGAAGAATGGATAGGAA	TTCTGTCTGCCTGCAATGCT
ATRX_i8-9	CGGAAGGTGGAAACTTGATTTG	CCAAGGTTGCGTAGAATGCA
CCR_19_A	CAGTTGTGCCTTCGTGGTACTC	CTCCCCATCTTCTCCCTTGA
CCR_19_B	CAGCCTTACCCTTGGCAAAA	AGGTGCGCCCGAAGAGTAT
CCR_19_C	GGCTTGGCATCTTGAACAGAA	CACAGGACTCAGGGCTGCAT
CCR_19_D	AGGCCTCTGTCTATTGGAGCTAAG	TGGGTTGGACAGGCCATT
CCR_19_E	TGAGTCGCGTCGGCAAA	TCCTTGCAGGACTGGATGTG
CCR_19_F	CCTGGGCCAGGGATCTG	AACTTTTGCCCACGTTTTCCT
CCR_19_G	CCCTCTACATTTGCCCTGTTG	GTTAGCGTCGCCCTCCAAT
CCR_19_H	GGACCCCACCATGCTCAAG	ATCCTACCCCAACAGCAGATGT
CCR_4q	TGATGGCAGTTAAACAAAATGTCA	CTGTGCCTTAGGAGCAGAGGTT
CTC-281H14	GACCCTGAGTCCCCTGGATT	GCGCGTGGGACAGAAAAA
CTC-459H11	CACGCATTTCCGAAGTCCTT	GGGACTGGGTTGCATGTACAA
p53	CCCAAGCAATGGATGATTTGA	GAGCTTCATCTGGACCTGGGT
RP11-109D4	GCGGCAGCATGCAGTTG	TTCCCTTGGCTGTATTAGTCACTTC
RP11-469N6	TGTCACCCGGGCAGACA	CTGTTGCAGAGGTGGTCATGA
RP1-292H14	CACATCATGCAGCAGCTGAA	TGGACTGTGGTCATGTCAGACA
RP1-85M6	TGGCAGCGCCAGGAATT	GTTAGCAAAATCTTCACAATCATCTTG

5. polymorphic repeat length analysis

Polymorphic lengths of short tandem repeats were analyzed by PCR on 50 ng of genomic DNA using Expand Long Template (Roche) and buffer 3 according to the manufacturer's instructions. Primer are described in Table 6. Forward primers were 5' labeled with 6-CarboxyFluorescein-Aminohexyl Amidite [6-FAM]. 0.1µL to 1 µL PCR product was mixed with

13µL Hidi and 0.16µL ROX-labeled genotyping marker 100-500 (Applied Biosystems, Foster City, California, USA) as an internal

Table 6: sequences of oligonucleotides used as PCR primers for polymorphic repeat length analysis.

Name	Forward primer	Reverse primer
HUMARA	TCCAGAATCTGTTCCAGAGCGTGC	CTGGGACGCAACCTCTCTC
D4Sxxx	TGGATTGATGTCCTTCCTGA	TCCTTGGACCTACACTCTATAGTCA
D16S3027	ATATTTGGCATCTGGGG	CCAGCATGAGTTGCTTT
D16S2622	ACTGCATCCCTTTAAACACTT	TAGCTTGGGTGAAGGAGTGA
D16S3065	TACATCCATAAGTACCCTTAACAAT	TGAAGTGTATCCCCAGTATAGA
MS1	GGGTAAAAAGCCCCTTTTAGAG	GTCAGTGTGGTGCTGAGGAG
MS2	AGCCTAGGCAACAGAACGAG	GGGCCCTTTTCTAACTGGTC
MS4	CAGCAAGGAGACCACTAGGC	TGTGTGTGTCACCACTGCAC
D19S1037	CAGCAGATCCCACCTCCTAT	CCATGCAGCTATCCCTCATT
D19S252	TGGACCCAATTAACATTCCAA	ATGTGACCCCTCTACCCACA
D19S429	CGACCTTTATGTCCCAGCAT	GTGCTTGACACTCCCCTCAT
D19S899	TCAACCAAGCTGTGTGTGTGT	AGTTCAGGCCTCTGCACATT

size standard. Samples were denatured 95°C for 3 minutes and stored at -20°C until they were separated on an ABI PRISM 3100 automated DNA sequencer (Applied Biosystems, Foster City, California, USA). Results were analyzed with the GeneScan analysis software (V.3.7, Applied Biosystems, Foster City, California, USA) for peak position and surface area.

6. X inactivation pattern analysis

The inactivation pattern of chromosome X was investigated using a methylation-sensitive restriction digestion of the polymorphic *AR* repeat[96] by HpaII, followed by a polymorphic repeat length analysis with HUMARA primers (Table 6). As a control, each undigested sample was also analysed for the length of the polymorphic *AR* repeat. A male DNA sample was subjected to the same procedure as a positive control for restriction digestion. The amount of skewing of the X inactivation pattern was calculated as follows:

$$\frac{1}{1+ \dfrac{A^d/B^d}{A^u/B^u}} = \% \text{ activated allele A}$$

A or B: area under peak of allele A or B
^{d} or ^{u} : digested or undigested

7. RNA extraction, cDNA synthesis and gene mRNA expression analysis

Total RNA was extracted from human cell cultures using the RNeasy Mini Kit (QIAGEN, Venlo, Netherlands) and from zebrafish embryos using TRIzol® (Invitrogen, Carlsbad, California, USA) according to manufacturers instructions. After DNaseI (Roche Applied Science, Vilvoorde, Belgium) treatment, RNA was reverse transcribed with the Superscript II

kit or the Superscript III kit (Invitrogen, Carlsbad, California, USA). Depending on the application, a mixture of random, poly-A or gene specific oligonucleotide primers complementary to the mRNA (reverse primers) were used for cDNA synthesis. mRNA expression of a gene was quantified by rtqPCR on cDNA using specific primers designed in discrete exons of the gene or across exon boundaries, to avoid genomic DNA amplification. mRNA levels were quantified by comparison to mRNA levels of genes showing stable expression as described by Vandesompele and colleagues[97].

Name	Forward primer (5'→3')	Reverse primer (5'→3')
ATRX	AGCCGACAAGGCGTTCAA	ACCAATGTATTCAACTTGCTTTCACT
GAPD	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
β2M	TGCTGTCTCCATGTTTGATGTATCT	TCTCTGCTCCCCACCTCTAAGT
β-act	CACCCTGAAGTACCCCATCG	TGCCAGATTTTCTCCATGTCG

8. protein extraction and Western blotting

8.1. nuclear protein extraction

A nuclear protein fraction extraction protocol was adapted from McDowell et al.[98]. 10 million cells in suspension were pelleted by centrifugation @ 1200 rpm for 10 minutes, and washed 3 times in PBS by centrifugation and resuspension of the pellet. 1mL lysis buffer was added and the solution was vigorously vortexed for 30 to 60 seconds. Cells were kept on ice for 30 minutes and pulled through a syringe 4 to 5 times. Nuclei were collected in a pellet by 7 minutes low-speed centrifugation (1000g). The pellet was washed 3x in RSB and resuspended in 0.2mL RSB buffer. 1 unit of DNaseI (Roche Applied Science, Vilvoorde, Belgium) was added per 1 μL RSB, and DNaseI digestion was allowed to proceed for 1 hour at room temperature. Proteins were solubilised by adding 60μL RIPA buffer.

8.2. zebrafish protein extraction

Staged zebrafish embryos were dechorionated manually by use of a forceps, and the yolk was removed in icecold PBS by forcing them through a glass needle. Zebrafish tissue was centrifuged to a pellet and washed twice with ice cold PBS. The tissue was briefly centrifuged to a pellet and the supernatant was replaced with 200mL ice cold 1x Danio lysis buffer with freshly added Proteinase Inhibitor (Roche Applied Science, Mannheim, Germany) and Phosphatase Inhibitor Coctail 1 and 2 (Sigma-Aldrich, Bornem, Belgium). Cells were lysed by pulling them through a syringe ten times and by the subsequent heating to 98°C for 5 minutes. Cellular debri was removed by centrifugation at 15000g at 4°C for 20 minutes. The supernatant was stored at -80°C or immediately processed.

8.3. SDS PAGE, Western blotting and detection

Protein concentrations were checked by a colorimetric reaction using the Pierce 560 nm Protein Assay Reagent system according to the manufacturer's instructions. The absorbance generated by the colorimetric reaction was measured using the BCA-1.0sec-560nm protocol on a Victor plate reader (Perkin Elmer).

20μg protein was mixed with 3X loading dye and 3% β-mercapto-ethanol, heated for 10 minutes to 70°C and kept on ice until loading. The protein samples and a SeeBlue® Plus2 Pre-Stained Standard (Invitrogen, Carlsbad, California, USA) were loaded on a NuPAGE® 4-12% Bis Tris gel (1.5 mm, 15 well, Invitrogen) and sodium dodecyl sulfate polyacrylamide gel electrophoresis [SDS PAGE] was performed for 10 minutes at 120V and around 60 minutes at 180V to separate the proteins according to their electrophoretic mobility.

Proteins were blotted a nitrocellulose membrane by applying a 35mA current for 2 hours in blotting buffer. Efficacy of blotting was checked by staining the membrane with Ponceau Red. Unspecific protein binding of the membrane was blocked by incubation with blocking buffer for 1 hour at room temperature. The membrane was incubated with blocking buffer containing the primary antibody at the correct concentration for at least 4 hours. After 3 washes of 5 minutes in blocking buffer, the membrane was incubated for 1 hour with a 1 in 10 000 dilution in blocking buffer of the appropriate secondary antibody conjugated to horseradish peroxidase. Excess secondary antibody was removed by 2 washes of 15 minutes in blocking buffer and 2 washes of 10 minutes in TBS. Chemiluminescence was generated by the ECL kit and detected on photographic films.

The following primary antibodies were used:

- ATRX was detected using a 1/50 dilution of the monoclonal antibody 23C[98], a gift from Dr Garrick, (WIMM, John Radcliffe Hospital, Oxford, UK).

- SP1 was detected using a polyclonal antibody (PEP2, sc-59, Santa Cruz) as a positive control for nuclear protein extraction.

9. synthesis of digoxigenin-labeled RNA probes

RNA was extracted from zebrafish embryos at 12hpf, 18hpf, 22hpf, 30hpf, 36hpf and 48hpf using Trizol (Invitrogen) as recommended by the producer. RNA was reverse transcribed by the Superscript II or III Reverse Transcriptase Kit (Invitrogen) and a mix of random and poly-T primers. In a first round, we performed PCR using the primer pairs

described in Table 7. The reverse primers were tagged with a T7 or Sp6 transcription start site. If no single band was seen upon gel electrophoresis, we used a nested PCR strategy by first using the EXT_ primers described in Table 7. If no single product was obtained after nested PCR, we assumed that the gene of interest was not expressed. We verified whether the PCR product contained the gene of interest by direct Sanger sequencing.

PCR products were purified by a phenol-chloroform extraction, and transcribed using the T7 or Sp6 Transcription kit in the presence of DIG-labeled nucleotides (DIG label mix, Roche) and RNAse inhibitor (Roche). Subsequently, DNA was removed by treatment with DNAse (Roche), and the RNA was purified using a standard phenol-chloroform extraction (Eurogentec). A sample of the RNA product was size separated denaturing gel electrophoresis, and the generated RNA probes were stored at -70°C until use.

Table 7: sequences of oligonucleotides used as primers for DNA probe synthesis. EXT_ denotes the use of this primer pair in the first step of a nested PCR reaction (the *external* primer pair).

Name	Forward primer (5'→3')	Reverse primer (5'→3')
DVL1	ATCGCAGGGATGCTAGAAAA	TAATACGACTCACTATAGGGAGCTGAAGCTAAGGCTGTGC
RAI2	GGCTTGTGCCCTATCCTGTGGT	CATTTAGGTGACACTATAGAATGACGATCTCGATGTTTCCA
DLG7	CCACCTGTACCCAACAAACC	TAATACGACTCACTATAGGGCTCAGCATCCATCACTCCAA
HES4	GACACAAACGTCCTCAGCAA	TAATACGACTCACTATAGGGCGCAAGTCTACCAGGGTCTC
EWSR1_a	GGCGTCAACCAGCAACACTCAG	CATTTAGGTGACACTATAGAAGAAGCCACCTCTGTCTCCAG
KREMEN1	AACTTCTGCAGGAACCCAGA	TAATACGACTCACTATAGGGTGCGGTGATATTCACGATGT
REPS2	AGTTCTGCACTGCCTTCCAT	TAATACGACTCACTATAGGGGAGGCAACCAGCCACTTTTA
HMGB2b	TATGCGTTCTTCGTGCAGAC	TAATACGACTCACTATAGGGATCCTCGTCCTCCTCATCCT
EWSR1_b	CAGACTCAGTACGGCCAACA	TAATACGACTCACTATAGGGAACCTGACAACCTGCCAATC
AATKa	GGCTGCTGCACCTTCACAAAAA	CATTTAGGTGACACTATAGAAGTCTTCCTCTGCCTCACTGG
FOXK2	CCGTCCGGTGTTTCTCTACC	TAATACGACTCACTATAGGGGATGGTCAGCGGGGAGAT
No_ID	GGTGTGCCTGGTTATTCCGAAC	ATTTAGGTGACACTATAGAAGCATTCACTGGTTATGCTGGT
SKI_a	GAGGAAATCCAGGTGGACAA	TAATACGACTCACTATAGGGCTCAGACGTCCTGCTTCACA
SKI_b	GGTGAGACCATCTCGTGCTT	TAATACGACTCACTATAGGGGTCATCGTGTTTGGGCTTCT
HAND2	TACCATGGCACCTTCGTACA	TAATACGACTCACTATAGGGCAGATGGCCTCATTTCGTCT
AATKb	GCAATTCCACAGATCATCCA	CATTTAGGTGACACTATAGAACGGCTGATTCCCTAACTCAA
MAFG	ACTGAAGGTGAAGCGAGAGC	TAATACGACTCACTATAGGGTGTGCATTATGACCGTGCTT
ARID4A	GGTTCAAGCTCTTCCGATTGGT	CATTTAGGTGACACTATAGAACGTCCTGCTCTTCATCATCA
CSNK1D	GGACTGGGATGGAGCGTGAAC	ATTTAGGTGACACTATAGAAGCTTCACCATCAGAAAGGAAC
FOXK2	GGAGGGCTCGGTGGATGTTAG	CATTTAGGTGACACTATAGAATTTCAGACTGCGAGTTGTCG
HMGB2a	GGCGGGAGGAACACAAGAAGAA	CATTTAGGTGACACTATAGAAATCCTCGTCCTCCTCATCCT
OTX2	CCCCACAACCATCTTTAGCA	TAATACGACTCACTATAGGGGAAGTGGAACCAGCATAGCC
THOC5	GGCTTACAGCCTGGACTGCACA	CATTTAGGTGACACTATAGAACATACGGATGTCCGAGGTCT
VEGFC	GGGCAAACCTTGCTTTCGAGTC	CATTTAGGTGACACTATAGAATGATGTTCCTGCACTGAAGC
FBN2	CGGAATGTTCCAGGAAGCTA	TAATACGACTCACTATAGGGCCGAACGAACATCAACACAG
NKX2-5	GTGCGGGACATACTGAACCT	TAATACGACTCACTATAGGGTTGCTGTTGGACTGTGAAGG
NSD1	TTGGGGAACCTGAAGAACAG	TAATACGACTCACTATAGGGATTTGTACCGGGAAGCACAG
EHMT1a	CGAACCCTGTCCTGACAACT	TAATACGACTCACTATAGGGGGCTGATGTTCCCGTAGAAA
EHMT1b	AAGGGTCCAAGAAAGCGAAT	TAATACGACTCACTATAGGGGTTCCCATCTGGCTGACACT
NOTCH1b	CAGAATGAATGGGGAGAGGA	CATTTAGGTGACACTATAGAAGCGGTTCCGTATCAAATCT
NOTCH1a	CGGGATATTGCAGATCACCT	TAATACGACTCACTATAGGGCTGCCTGCTGTTGTTGTTGT
CREBBP	GAGAGTGGTTCAGCACACCA	TAATACGACTCACTATAGGGCATGCCATTCATGCTCATTC

Name	Forward primer (5'→3')	Reverse primer (5'→3')
ATRX	TGGTGTGCAGAAGGTGGTAA	TAATACGACTCACTATAGGGTTTTCAAGCGATTCCTCCAG
ext_DVL1	CGGATGTTGTTGACTGGTTG	GTGAGTTCTGGTGGGACGTT
ext_RAI2	TCCTTCAACATGCACTGCTC	AAAATAGCAGGCATGGCATC
ext_DLG7	AGACAAACCCCAGGAGGACT	AACACCATTCTGGGAACAGC
ext_HES4	GCCAGCCGATAATATGGAGA	AACACTTGCCCAACAAAAGG
ext_EWSR1_a	TAATGCTGCTTCAGCCACAC	AAGTCAGCGACTTCCTCCAA
ext_KREMEN1	AGACCAGTCTGCAAGGAGGA	ACAACAGAAAACCCCAGTGC
ext_REPS2	CGACCTCAATGCTCTCATCA	CAGAGTAGGGCGTGGTCAAT
ext_HMGB2b	AGACGTGAACAAACCCAAGG	CATCTTCCTCGTCCTCCTCA
ext_EWSR1b	CTACCTCAACAGCACCAGCA	CCCATATCACGATCCATTCC
ext_AATKa	CAGGTGGTGGTGAAGGAACT	CGAAGTGGTGGATGACATTG
ext_FOXK2	GTGTCCGTTCCGTCAGATTC	CGCAGGTGTGTGATGTTCTC
ext_no_ID	ATCTGATCGCAAGGAGGTTG	CGTCAGCCTTTCCCTGATAC
ext_SKI_a	CCATTGCTGCCTTCAAAAAT	CGTTCCTTCAGCAGGTCTTC
ext_SKI_b	GCAACACGATTTGCTCTTCA	GAAGGCTGACGGTCTCTGTC
ext_HAND2	ATGGGGAGACAGTTGGTGTC	GCGGGAAATTGCACATAAAT
ext_AATKb	AGCCCTACTGCTCCTCTTCC	TGGAGGTGGGTGTACTGTCA
ext_MAFG	CAAACAAGGCAAACAAAGCA	TTACAAATGTTCTCCGTTTGTG
ext_FBN2	GGATCCGGATATCTGCTCAA	CCGTACAGCAACACATGTCC
ext_NSD1	TCTTGTGAGCAGACCTGGTG	CGGTGTTGTGATTGTCCTTG
ext_EHMT1a	CTCGAAGGTCTGGAATGCTC	CGCCATAGTCAAACCCAAGT
ext_EHMT1b	GTCATCACCCACACAGCATC	CCTGTCCATTAAGGCGTTGT
ext_NOTCH1b	TCAGACATTTCGTTAATGGAC	ATCAGAGGTGTGGTGCCATC
ext_NOTCH1a	ACGGCTAAAGTCCTGCTTGA	GAAGGCGGAGTCAGAAACTG
ext_CREBBP	CAAGAGCCCACAGGAGAGTC	CGGGAGTAATGTTGCTGGTT
ext_ATRX	GATGAATTCCGAGGTCCTGA	CTGAAACTTGAGGCCTTTCG

10.   whole-mount *in situ* hybridization on zebrafish embryos

The zebrafish whole-mount *in situ* hybridization [WISH] protocol was based on a protocol generously provided by professor Bernard Thisse (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France) that was published recently[99].

10.1.   embryo production, growth and storage

Wild-type or transgenic *cmlc2*-GFP zebrafish[100] were kept at the fish facility or at the Glycobiology laboratorium of professor Guido David (K.U.Leuven, Leuven, Belgium). Embryos were produced by natural mating, collected and incubated in egg water at 28.5°C or temperatures in the range of 25°C to 32°C to accelerate or slow down embryonic development. Corresponding developmental stages at 28.5°C were calculated using the formula described by Kimmel and colleagues[101]:

$$HT = \frac{H}{0.055 \times T - 0.57}$$

T = temperature (°C)
HT = hours of development at temperature T
H = hours of development at 28.5°C

Before the 25 somite stage, embryos were transferred to fish water with 0.0045% 1-Phenyl-2-thiourea [PTU] to avoid embryo pigmentation[102]. Embryos were fixated at the desired morphological stage[101] by incubation in 4% PFA at 4°C overnight. Embryos were dehydrated by incubation in methanol for 15 minutes at room temperature and kept at -20°C for at least 2 hours or until further processing.

10.2. whole mount *in situ* hybridization and probe detection

Embryos were rehydrated by 5 minute incubations in a series of methanol in PBS dilutions (75%, 50%, 25%) and washed 4 times 5 minutes in PBST. Embryos were permeabilized by incubation in PBST with 10µg/mL proteinase K for a duration depending on their developmental stage (10 minutes for embryos of 12 to 24 hours post fertilization [hpf], 20 minutes for embryos up to 48 hpf and 30 minutes for older embryos). Digestion was stopped by incubation in 4% PFA for 30 minutes and PFA was removed by washing 5 times 5 minutes in PBST. A batch of 500 to 1000 unstaged embryos were transferred to WISH Blocking mix with a 1/400 dilution of anti–digoxigenin [DIG] antibody conjugated to alkaline phosphatase [AP] [α-DIG-AP] (Roche Applied Science, Mannheim, Germany) and kept overnight at 4°C under constant rotation until the detection step. Staged embryos were transferred to WISH Hyb mix and incubated for at least 2 hours at 70°C. These embryos were subsequently hybridized with 2µL of Dig-labeled RNA probe in 200µL WISH Hyb mix at 70°C overnight under mild agitation (40rpm). The probe in WISH Hyb mix can be used up to 3 times if it is stored at -80°C. Excess probe was washed away by consecutive 15 minute washes in a dilution series of WISH Hyb mix^{-} and 2x SSC ($^{3}/_{1}$, $^{1}/_{1}$, $^{1}/_{3}$, $^{0}/_{1}$) at 70°C, a 30 minute wash in 0.2x SSC at 70°C and a 30 minute wash in 0.2x SSC at room temperature. Embryos were transferred to WISH Blocking mix by consecutive 10 minute washes in a dilution series of 0.2x SSC and PBST ($^{3}/_{1}$, $^{1}/_{1}$, $^{1}/_{3}$, $^{0}/_{1}$) at room temperature. Embryos were transferred to WISH Blocking mix for at least 2 hours. Unstaged embryos were removed from the 1/400 α-DIG-AP dilution in WISH Blocking mix and discarded, and the 1/400 dilution was further diluted to 1/4000 in WISH blocking mix. Staged embryos were incubated in this 1/4000 dilution overnight at 4°C to allow DIG-labelled RNA probe detection. The next day, the 1/4000 dilution was removed. It can be used up to 3 times if stored at 4°C and if 0.02% NaN$_3$ is added. Embryos were washed 6 times 15 minutes in PBST and 2 times 5 minutes in AP buffer. They were transferred to WISH Staining solution for colorimetric probe detection.

Staining was regularly monitored under a staining microscope, but embryos were kept in the dark between monitorings. When the desired amount of staining was achieved, the staining reaction was stopped by 3 washes of 5 minutes in PBST, followed by incubation in 4% PFA at 4°C until tiff images of the mRNA expression patterns were captured using a Leica Fluo Combi stereomicroscope.

11. gene knockdown by morpholino injection into zebrafish embryos

Morpholinos were designed against the translation start site or a splice site of the gene of interest by Gene-Tools LLC (Philomath, Oregon, USA). The target site was sequenced to rule out imperfect binding of the morpholino due to mismatches compared to the reference sequence. The morpholino was ordered from Gene-Tools LLC (Philomath, Oregon, USA) and dissolved in a dextran rhodamine B (10,000 MW, neutral, invitrogen) dilution (1/50) to the desired concentration.

Morpholinos were injected in zebrafish embryos within 1 hour after being laid (before the 4-cell stage) using needles produced from 1.2mm capillaries with an internal glass fiber (Narishige International Limited, London, U.K.). 500pL (volume estimated by injection into an oil drop on a microscope slide with a microgrid) was injected into each embryo, and injections were verified by analysis of rhodamine fluorescence under a fluorescence microscope.

Table 8: Morpholinos used in this study

Morpholino name	Sequence
TAB2_sb_MO	ATCACTCTTGTTCTGAGGAAAGAAG
TAB2_tb_MO	ATCTGCTGGTTTCCCTGTGCCATTC

CHAPTER 4   POSITIONAL CLONING OF REGIONS FOR CHDS

1.   introduction

Positional cloning of genes for CHDs via chromosome aberrations is an established strategy for gene discovery. The underlying assumption is the following: the probability that 2 rare events (an error in development and a chromosome aberration) co-occur in a single patient by chance is small. It is more likely that both events are related and that the chromosome aberration causes the developmental error. Chromosome aberrations can be balanced (e.g. translocations or inversions) or unbalanced (e.g. deletions, duplications, triplications,…). Traditionally, both aberrations were detected by microscopic cytogenetics. If needed, mapping was performed, mostly by walking FISH. About 10 years ago however, a novel technique for detecting and delineating constitutional chromosome imbalances in a single experiment was developed: aCGH[103,104]. This allows simultaneous detection and mapping at a much improved resolution.

In an aCGH experiment, two DNA samples are differentially labeled and hybridized on a microscope slide that contains an array of DNA probes. These DNA probes can be large insert clones (BACs, PACs, fosmids)[103,104], or PCR amplified or in situ synthesized oligonucleotides[105,106]. Both array types were used in this study. Most experiments in this study were performed using an in house manufactured array slide with large insert clones chosen genome-wide with a median spacing of 1Mb (1Mb array, BAC/PAC clones provided by the Sanger Institute, Hinxton, UK), while other experiments used an array with BAC/PAC clones containing fragments of one or more specific chromosomes at tiling resolution. Finally, also a commercially available oligonucleotide array with a median probe spacing of 10kb was used (Agilent 244K array). See the materials and methods section for a description of the different platforms that were used.

2.   protocol development for 1Mb aCGH

As a first step in our studies, the protocol for performing and interpreting aCGH had to be improved, as the existing protocol generated an inconsistent quality of the experiments.

A major recurrent problem was a "wavy" pattern in the log2 ratios. This pattern appeared to correlate with the GC content of the underlying DNA sequence. Since the DNA labeling

protocol involves a DNA denaturation step, we hypothesized that a GC content dependent incomplete denaturation could cause this deviation. Indeed, upon denaturation at 98°C instead of 94°C, this bias largely disappeared. All DNA used in hybridizations in this study was therefore denatured at 98°C before labeling.

The quantity of labeled DNA was moreover made variable in order to consistently obtain sufficient hybridization signals for the Cy3 channel. Although the underlying problem (insufficient probe DNA spotted on the arrays) was not solved, consistently better results were obtained. In this way, a robust aCGH protocol was established in our laboratory.

Also aCGH data analysis needed optimization, as it involved many steps of manual copying and pasting rendering it error prone. It moreover resulted in a large number of false positives. To tackle these issues, a working template was built in Microsoft Excel, requiring only pasting a single time per hybridization, and twice per patient. Criteria for calling aberrant clones were further established by introducing a variable calling threshold and the requirement of consistent log2 ratio measurements per clone (both intra-experimental and inter-experimental)[89]. As a consequence, the presence of only 2 of 38 aberrations that were called could not be confirmed using an independent experiment. We could not confirm the presence of 9 out of 10 selected aberrations that did not meet the established criteria. This shows that our methodology can determine the presence of aberrations at a low cost of false-positive and false-negative results.

In the first series of experiments that were performed, 3 aberrations present in mosaic were detected[89]. For this reason, the ability of the aCGH technology to detect imbalances that are present in a mosaic state was verified. A statistical model was built in collaboration with the ESAT department (K.U.Leuven, Belgium) to calculate detection limits for mosaics depending on the number of clones that are imbalanced or not, the standard deviation of the experiment and the desired power of the test[107].

$$\frac{mD}{\sigma\sqrt{1/n_A + 1/n_E}} = t_{N-2,\alpha} - t_{N-2,\beta}$$

m = % of mosaicism
σ = pooled standard deviation
n_A = # of clones in the aneuploid region
n_E = # of clones in the euploid region

D = expected $\log_2$ ratio when all cells are aberrant
α = degree of certainty (<0.0001)
β = power of the test
N = # of clones on the array

To test this equation experimentally, a series of dilution experiments was performed where DNA from a patient carrying a trisomy 13 was mixed with a normal DNA sample. The

results are shown in Figure 2 and demonstrate that a 5% mosaic can readily be detected by aCGH. We thus showed that aCGH can also robustly detect imbalances that are present in only a fraction of the normal cells.
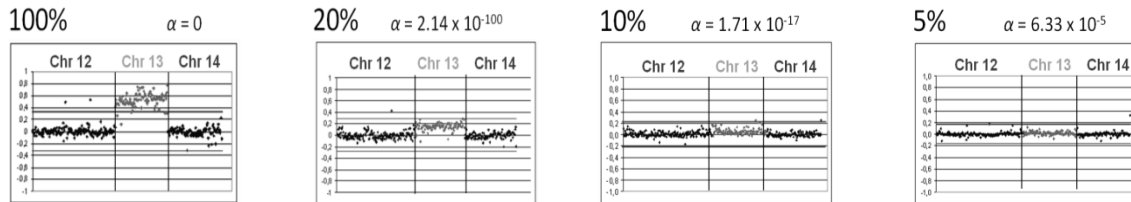

Figure 2: detection of mosaicism by aCGH. Results from a series of dilution experiments to test the ability of aCGH to detect a trisomy 13 in a decreasing fraction of cells are displayed. Dark dots indicate the log$_2$ ratios for chromosome 12 and 14, light for chromosome 13. The percentage above each image refers to the fraction of DNA originating from the trisomy 13 cell line, and the α value refers to the degree of certainty that is associated with the assignment of chromosome 13 to a separate copy number state (β = 0.95).

3.   1Mb aCGH for an improved diagnosis and positional cloning

3.1. patient selection and characterization

Once a robust protocol for the detection of micro-imbalances was established, this technology was used to analyze a group of patients with idiopathic syndromic CHDs. 99 patients were recruited from the congenital cardiology and genetic clinics of the University Hospital Leuven. All patients had a CHD and at least one of the following additional features:

(1) a second major malformation

(2) mental handicap or following special education

(3) three or more minor physical anomalies.

In none of them, an etiological diagnosis could be reached after

(1) expert dysmorphological examination

(2) routine karyotyping using G-banding analysis at least at ISCN +550 bands

(3) the appropriate, additional investigations such as FISH or gene sequencing to exclude well-defined genetic disorders.

3.2. results of aCGH and inheritance analyses

aCGH revealed copy number variations (CNVs) in most of the 99 patients, ranging in size from 0.15 to 49Mb. The vast majority of these had already been detected repeatedly in normal individuals, are annotated as phenotypically indifferent polymorphisms (http://projects.tcag.ca/variation/) [91,108] and were not further investigated. Some of the large aberrations were in retrospect visible on the microscopic karyotype.

23 patients (23%) carried imbalances that had not been described before in phenotypically normal individuals at the time of detection, with the exception of the duplication of chromosome 22q11.2 observed in patient 14 (Table 9 and Table 11). The presence of all these imbalances was confirmed using an independent technique: FISH or rtqPCR.

The *de novo* nature of these imbalances was investigated through the analysis of the parents using FISH or rtqPCR, except for patients 7 and 9. In 12 patients the imbalance was shown to have occurred de novo. In 9 patients it was inherited (Table 9 and Table 11); for these 9 patients, the parents were phenotypically normal, with the exception of the father of patient 14 who carried a duplication in chromosome 22q11.2 and had learning difficulties.

Patient 7 carried a terminal deletion of the long arm of chromosome 5 extending to the region responsible for the Sotos syndrome[109,110]. His phenotype (macrocephaly, large anterior fontanel, widely spaced nipples and low set ears) corresponds to the phenotype of terminal 5q deletions which include the Sotos syndrome gene *NSD1*[111]. The parents were reportedly normal, but parental samples were unavailable for this patient. Given the severity and high penetrance of Sotos syndrome and of the phenotype associated with terminal 5q deletions, chances that this deletion was inherited are considered as low.

Patient 9 carried an interstitial 10q deletion. Only a sample from the mother was available for this patient, and she was shown not to carry the deletion. The phenotype associated with 10q25q26 deletions in other patients is highly similar[112,113], suggesting this deletion is causal. Also the size of the deletion and the fact that it is not reported as a polymorphism support this conclusion.

In one additional patient a mosaic monosomy 7 was initially detected by aCGH. The level of mosaicism as observed by aCGH was calculated to be 5%. Interphase FISH analysis by two independent observers using a centromere 7 specific probe revealed a single signal in 10.5% of the nuclei of peripheral white blood cells of the patient while in a control sample a single signal was only observed in 3.5% of the nuclei. The difference between these two proportions is statistically significant (p<0.01) thus confirming the presence of the monosomy in approximately 7% of the patient's white blood cells. This monosomy was not found in fibroblasts and buccal smear cells. The *de novo* nature of this imbalance was not an issue[89]. This newborn presented with a coarctation of the aorta, absent thumbs, hydronephrosis and severe prenatal growth retardation, and was later diagnosed with

Rothmund-Thomson syndrome. The monosomy 7 is frequently found in association with myelodysplasia[114], a condition also seen in Rothmund-Thomson syndrome[115]. This chromosomal aberration is not considered as causal for the constitutional features, but did contribute to establishing a diagnosis.
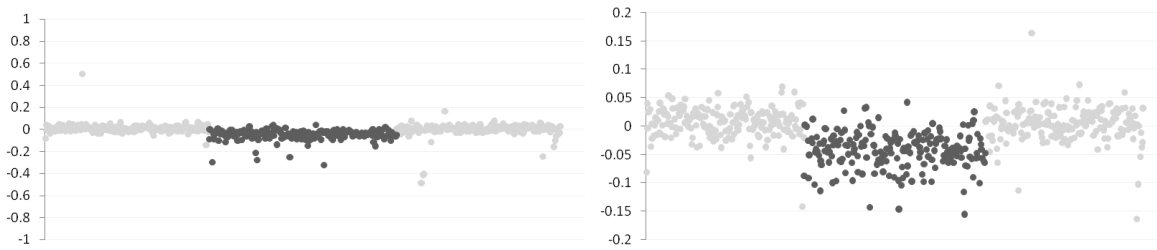

Figure 3: detection of a monosomy 7 present in +/- 7% of the white blood cells by aCGH. The left image shows the slightly decreased log$_2$ ratio (Y axis) of chromosome 7 (dark) in comparison to chromosomes 6 and 8 (lighter grey flanking dots respectively left and right). The right image shows the same profile but with a more narrow range of log$_2$ ratios.

3.3. causality of imbalances

The interpretation of the detected aberrations is not straightforward. In contrast to microscopic karyotyping, the mere detection of a variant does not imply causality. We had to establish robust criteria to evaluate the significance of an imbalance. We proposed the following criteria (the rationale behind each criterion is briefly explained).

criterion 1: has a similar aberration previously been reported in a patient with a similar phenotype?

The simultaneous occurrence of a rare chromosomal imbalance and a similar phenotype in multiple patients by chance is unlikely. For example, a submicroscopic deletion or duplication of chromosome 22q11.2 in multiple patients with heart defects led to its identification as the cause of the Velocardiofacial or DiGeorge syndrome or the delination of the duplication 22q11 syndrome[116]. In total fourteen patients in this study presented distinct phenotypic manifestations previously reported in other individuals with a chromosomal imbalance in the same chromosomal region (Table 10).

criterion 2: did the aberration affect a gene known to cause part of the patient's phenotype upon haplo-insufficiency or duplication?

The precise delineation of a chromosomal aberration by aCGH allows the identification of all affected genes. When haplo-insufficiency or over-expression of a gene located within the chromosomal imbalance is known to cause a distinct phenotype, the presence of this phenotype in the patient is a very strong argument in favor of a causal aberration. For this

reason, rtqPCR was performed on patient 3 to confirm the deletion of *NKX2.5*, a gene known to be involved in CHDs[1]. In 4 additional patients, genes known to cause CHDs were affected: *NSD1* (patient 4)[117] and *NOTCH1*[118] and *EHMT1*[119] (patient 5), *CREBBP*[120] (patient 14) and *ATRX*[121] (patient 16). Three patients carry a deletion that includes a gene causing a distinct phenotype different from CHDs when hemizygous. E.g. patient 7 had anophthalmia, and the deletion on chromosome 14q22.1q23.1 included the *OTX2* gene, known to be involved in anophthalmia[122] (Table 10).

criterion 3: did the aberration occur de novo*?*

In traditional cytogenetics, *de novo* chromosomal imbalances detected by routine karyotyping are large, encompass in general hundreds of genes and are therefore usually accepted to explain the patient's phenotype. At the present time, it is not certain whether this also applies to the smaller aberrations detected by aCGH. The frequency at which *de novo* aberrations at aCGH resolution occur in phenotypically normal individuals is not exactly known. Recently, it was found that copy number polymorphisms in genomic regions flanked by low copy repeats are typically inherited polymorphisms[123].

Data from deletions and duplications in the Duchenne muscular dystrophy gene, which has a size of 2.5 Mb, suggest that one in eight newborns carries a *de novo* imbalance of several kilobases[124]. This frequency is a likely overestimate as *DMD* seems particularly susceptible to indel mutations[125]. It is however likely that some *de novo* chromosomal imbalances will be phenotypically irrelevant. However, the majority of these imbalances are beyond the resolution of our BAC/PAC micro-array[126], and it is thus unlikely that the high frequency of *de novo* imbalances in the patients we studied is reflecting the naturally occurring rate of *de novo* phenotypically irrelevant imbalances.

criterion 4: were more than 20 genes affected by the imbalance?

As discussed before, the size of the chromosomal imbalance does matter. In one study, the size of *de novo* aberrations detected in patients (median 2.74Mb) was significantly larger than inherited aberrations (median size 0.43Mb)[127]. This size difference most likely reflects the increasing chance of affecting a dosage-sensitive gene with an increasing aberration size. The largest imbalance detected in a normal individual (the mother of patient 16) in the present study was maximally 3.7 Mb, affecting 17 genes at most. She is currently 42 years old, and never experienced cardiovascular manifestations. A recent medical examination did

not reveal any manifestations of a heart defect. The largest imbalance inherited from a normal individual in a previous study was 3.8 Mb[127], affecting 10 genes. Based on these data, the number of affected genes can be included as a criterion in the evaluation of causality, with a proposed threshold of 20 affected genes. By no means do we claim that imbalances affecting more genes have never been detected in normal individuals. However, these data suggest that imbalances affecting more genes will be rare in the normal population, an observation that is confirmed by a number of studies investigating the presence of CNVs in the normal population[128].

Initially, we considered each criterion equally important. If an aberration fulfilled two or more criteria, it was considered to be causal[58]. Later research results showed however that some imbalances affecting more than 20 genes are also compatible with normal development[129], that sometimes small imbalances involving only one or two genes can be causal[125,130]. Moreover, *de novo* imbalances also occur (infrequently, in 1%) in the normal population[131]. Although both criteria still suggest causality (aberrations that meet these criteria are more frequent in the syndromic CHD patient population than in the normal population), they no longer uniquely apply to aberrations found in the patient population. Only criterion (1) and (2) are sufficiently strong to be used as sole criterion in the evaluation of causality, and (3) and (4) as such are insufficient to prove causality. In 15 patients the imbalance scored positive on criterion (1) or (2). The majority of these aberrations also fulfilled criterion (3) and/or (4). Patient 13 carries a complex *de novo* aberration on chromosome 19p affecting over 100 genes. Based on the combination of these features, also this aberration was classified as causal. Therefore, in a total of 16/99 patients (16%) an aberration was detected that we considered to be causal. We refer to Table 9 for an overview of the aberrations found in these 16 patients. This study shows that the use of aCGH significantly improves the etiological diagnosis of patients with syndromic CHDs. The results on the first 60 patients as well as the first set of rules for causality evaluation were published in the European Heart Journal[58]. A manuscript describing the results from 90 additional CHD patients screened by 1Mb aCGH including 39 studied by us is in preparation. In this manuscript, a novel clinical algorithm for causality evaluation will be presented that takes into account the most recent findings concerning CNVs in normal patient populations.

Table 9: Summary of the genotype and other genetic data of the patients with a causal cryptic copy number variant detected by aCGH at 1Mb resolution. Size refers to the estimated maximal size. Y: yes, Pat: inherited from the father, Mat: inherited from the mother, ND: not determined, brtl: one of the parents carries a balanced reciprocal translocation

#	ISCN karyotype	imbalanced regions (kb)	size (kb)	de novo
1	46,XX .arr cgh 1pterp36.33 (RP11-465B22→RP1-37J18)x1	del 0→4,512-5,866	5,866	Y
2	46,XX .arr cgh 1q34qter(RP11-214M7→CTB-160H23)x1, 4q35.1qter(RP11-228F3→CTC-963K6)x3	del 234,259-235,542→247,233	12,974	Y
		dup 184,561-185,872→191,290	6,729	brtl
3	46,XY.ish(der4)(4pter→4q33::4q33→4q28::4q32.2→ 4q32.2::4q34.1→ 4q34.1::4q35.1→4qter)	dup 162,158-162,327→162,511-162,546	388	
		del 171,034-171,157→175,177-175,301	4,267	Y
		del 175,858-175,882→183,734-183,977	8,119	
4	46,XX .arr cgh 4q34.3qter(RP11-62B7→CTC-963K6)x1, 18q12.1(RP11-413I9→CTC-964M9)x3	del 181,215-182,025→191,273	10,058	Y
		dup 27,569-28,294→76,117	49,548	brtl
5	46,XX .arr cgh 5q23q23(CTC-461G12→RP1-241C15)x1	del 118,442-118,981→131,577-131,686	13,244	Y
6	46,XX .arr cgh 5q35.1(RP11-20O22→CTB-54I1)x1	del 170,440-170,455→172,758-172,785	2,345	Y
7	46,XY .arr cgh 5q35.3qter(CTB-87L24→CTC-240G13)x1	del 175,266-175,835→180,857	5,591	ND
8	46,XX .arr cgh 9q43.3qter(RP11-399H11→GS1-135I17)x1 20q13.33qter(RP5-836E13→CTB-81F12)x3 .ish der(9)t(9;20)(q34.3;q13.33)	del 135,650-137,112→140,273	4,623	Y
		dup 57,734-58,880→62,435	4,701	brtl
9	46,XX .arr cgh 10q25.2q26.11(RP11-426E5→ RP11-427L15)x1	del 113,136-113,860→120,485-121,545	8,409	ND*
10	46,XX .ish(der13)(13pter→13q21.31::13q21.31 →13q31.1::13q33.3→13q31.3::13q31.3→ 13qter)	del 59,687-62,001→62,167-62,949	3,262	
		del 80,958-82,035→91,449-92,499	11,541	Y
		dup 91,449-92,499→106,657-107,333	15,884	
11	46,XY .arr cgh 14q22.1q23.1(RP11-262M8→ RP11-224G6)x1	del 51,763-51,907→58,205-58,300	6,537	Y
12	46,XY .arr cgh 16p13.3(RT100→RT102)x3	dup 3,629-3,647→3,993-4,109	480	Y
13	46,XX .arr cgh 19p13.12p13.11(RP11-765H5→ RP11-33B5)x4,(RP11-171H5→RP11-97A20)x3, (RP11-512B16→RP11-96J2)x4	trip 15,684-15,739→16,793-16,933	1,249	
		dup 16,793-16,933→18,500-18,594	1,801	Y
		trip 18,500-18,594→19,560-19,608	1,108	
14	46,XX arr cgh 22q11.2q11.2(87H3→cHKAD-26)x3	dup 17,410-17,420→19,760-20,280	2,870	Pat
15	46,XX arr cgh 22q12.2q12.2(RP11-664C16→ RP11-794O14)x1	del 27,708-27,820→28,520-28,692	984	Y
16	46,XY arr cgh Xq21.1 (RP11-146M18→ RP11-217H1)x2, (RP5-1000K24→ RP13-213F13)x2	dup 76,647-76,663→76,884-76,928	281	
		dup 77,081-77,113→77,337-77,385	304	Mat

*the mother was shown not to carry this deletion.

Table 10: phenotype of patients carrying a causal cryptic chromosome imbalance detected by 1Mb aCGH. dysm: dysmorphism (3 or more minor anomalies), MR: mental retardation, IUGR: Intrauterine growth retardation, VPI: Velopharyngeal insufficiency. Heart defects are abbreviated as in the list of abbreviations

#	congenital heart defect	additional features	dysm	genes associated with this phenotype (partly)	similar genotype & phenotype report
1	muscular VSD, left-ventricular non-compaction	MR, IUGR	+	/	132
2	HLH	Club foot right	+	/	no
3	critical aortic valve stenosis	MR, cleft palate	+	/	133
4	perimembraneus VSD, multiple muscular VSDs apically	MR, corpus callosum hypoplasia, diafragmatic hernia	+	/	Trisomy 18
5	ASD type II	MR, cleft palate (robin sequence)	+	FBN2	134
6	ToF type DORV, infundibular PS, mitral valve dysplasia & mild prolapse, multiple VSDs	microcephaly, bronchomalacia (†)	+	NKX2-5	135
7	ToF, pulmonic valve atresia perimembraneus outlet VSD, functional bicuspid aortic valve	multicystic renal dysplasia	+	NSD1	111
8	narrow preductal coarctatio aortae, perimembraneous VSD	microcephaly, duplication cyst	+	EHMT1, NOTCH1	136
9	ODB	MR	+	/	113
10	ToF type DORV, right-turning aortic arch, small ODB	MR, microcephaly	+	/	137
11	TGA, ASD type II	microtia, anophthalmia, polydactyly	-	OTX2, BMP4	138
12	ASD type II, septum malposition (partly across VCI)	MR	+	CREBBP	139
13	multiple musc VSDs, one large midmusculair VSD, large ODB	MR, microcephaly, VPI	+	/	no
14	TA type I, quadricuspid functional truncus valve, left VCS via SC to RA, right VCS, rudimentaire VA	cleft lip and palate, hydronephrosis	+	TBX1	116
15	mild PS	MR, cleft uvula and VPI, high myopia	+	NF2	140
16	large ASD type II	severe hypospadias & micropenis, bil.bilobar lungs, no gallbladder	+	ATRX	125

3.2.   unclassified variants

With the advent of aCGH, a binary classification of chromosome imbalances as causal or benign has become problematic. A number of recent publications describe the recurrent identification of deletions or duplications of chromosomal regions (1q21.1, 15q11.2, 15q13.3 and 16p13.11) in populations of patients with developmental, mental and/or behavioural problems[141-147]. They are typically undetected or found at a significantly decreased frequency in the normal population, while they are overrepresented in the investigated patient populations. They are moreover frequently transmitted from a normal or mildly affected parent and probands are typically affected with different types of developmental abnormalities that can vary widely. These findings represent a paradigm shift in constitutional cytogenetics: a clear association of one chromosome aberration to one spectrum of phenotypes no longer stands. Certain chromosome aberrations can be linked to multiple seemingly independent phenotypes like heart defects, mental handicap or schizophrenia. They probably represent susceptibility factors of unknown effect size for several rare phenotypes, residing in the grey zone of the spectrum of effect sizes associated with mutations (see also Figure 1): their low population frequency and frequent *de novo* emergence suggests that they are subject to strong purifying selection, distinguishing them from the class of polymorphic mutations that are typically used in association studies, and that are inherited, frequent and have a relatively low effect size. On the other hand, their

Table 11: unclassified copy number variants detected in a group of 99 patients with idiopathic syndromic CHD. Imbalanced regions displayed in bold have been further delineated by means of an Agilent 244K array. Size refers to the maximal size of the imbalance. Mat: inherited from the mother, Pat: inherited from the father.

#	ISCN karyotype	imbalanced regions (kb)	size (kb)	*de novo*
17	46,XY .arr cgh 1q21.1(RP11-533N14→ RP11-301M17)x3	**dup 144,768→146,212**	1,444	Pat
18	46,XX .arr cgh 7q21.13(RP5-998H4)x3	dup 88,167-88,671→88,795-89,321	1,154	Mat
19	46,XY inv(10)(p112;q24) .arr cgh 9p22.2 (RP11-132E11)x1	del 16,325-16,995→17,138-17,839	1,514	Pat
20	46,XX .arr cgh 10q21.1 (RP11-430K23)x1	del 55,799-56,709→ 56,882-57,554	1,755	Mat
21	46,XX .arr cgh 15q21.3(RP11-215J7→ RP11-141F2)x3	**dup 52,620→52,739**	119	Pat
22	46,XX .arr cgh 16p13.11p12.3(RP11-489O1→ RP11-288I13)x3	**dup 15,404→17,583**	2,179	Mat
23	46,XY .arr cgh Xp21,3(RP11-37E19→RP6-27C10)x2	dup 28,427-29,091→29,104-29,347	920	Mat

inheritance pattern and the mild or normal phenotype of carrier parents preclude their classification as causing a Mendelian disorder, a hitherto typical feature of chromosome aberrations. They moreover show that not all syndromic disorders have a single monogenic cause.

We identified 7 unclassified variants (Table 11 and Table 12). We choose to classify the duplications of 16p13.11 and 1q21.1 (detected in patients 17 and 23, see Table 11 and Table 12) as unclassified rather than causal, since the genotype-phenotype correlation for each of these is at the moment unclear. Studies investigating their presence in control and affected populations yielded conflicting results: some report them as equally frequent in the normal versus patient population[141,145] and others as significantly overrepresented in affected populations[142,148]. The involvement of both these duplications in syndromic CHD has moreover not yet been studied. It is nevertheless intriguing that syndromic CHDs and the reciprocal deletion of 1q21.1 are frequently associated[142,147], and that some of the genes that lie in the duplicated regions are known to be involved in mammalian cardiovascular development (*MYH11*[63] in 16p11.2 and *GJA5*[149,150] in 1q21.1).

Arguably, the duplication of 22q11.2 (in patient 14, Table 9) is a susceptibility factor rather than a cause, since this duplication is also found in normal carrier parents[116]. This duplication was however found to be overrepresented in patients referred for DiGeorge

Table 12: phenotype of patients carrying an unclassified copy number variant (described in Table 11). dysm: dysmorphism (3 or more minor anomalies), MR: Mental retardation. Heart defects are abbreviated as in the list of abbreviations.

#	congenital heart defect	additional features	dysm	genes associated with this phenotype	similar genotype & phenotype reported?
17	UVH, dominant RV, small LV, Mitral valve prolapse, hypoplasia anterior left papillary muscle, moderate MS	single kidney (left), brain tumor (choroid plexus papilloma), hemivertebrae	-	/	*
18	critical aortic coarctation, hypoplastic aortic arch, muscular apical VSD	diafragmatic hernia, vertebral anomaly	-	/	/
19	VSD	cleft palate	-	/	/
20	PDA, small muscular VSD (spontaneously closed), small ASD II	Tracheomalacia	-	/	/
21	HLHS, severe MS and AS, LVOTO	-	+	/	/
22	ASD type II	coloboma, renal agenesis, choanal atresia	+	*MYH11*	*
23	left isomerism, mini VSD (spontaneously closed)	MR, microcephaly, epilepsy	+	/	/

*see discussion in main text

syndrome (typically syndromic CHD patients) and mental handicap. For this reason and in order not to complicate matters further, we choose to include this patient in the group of causal aberrations. However, this illustrates that the effects of each of these imbalances is situated on a spectrum of effects, ranging from small to large. The distinction therefore is somewhat arbitrary, since currently, no meaningful cut-off value exists. Preliminary data suggests that a similar phenomenon could apply to 16p13.3 duplications, as they have also been found in normal parents (unpublished results, manuscript in preparation). The high frequency of imbalance of these regions despite purifying selection in the investigated populations is due to the presence of the low copy repeats (LCRs) that flank these regions and rendering them susceptible for non-allelic homologous recombination (NAHR). There is however no reason why only LCR-flanked regions should be part of this class of heritable and frequently pathogenic copy number variants. Each of the other chromosome aberrations described in Table 11 could represent a similar case. For this reason we describe them as unclassified rather than benign.

4. individual patients

Our studies led to a number of additional valuable findings: an expansion of the phenotypic spectrum associated with known syndromes (left-ventricular non-compaction in del 1p36 syndrome[151], absent gallbladder in ATRX syndrome[125]), the identification of novel types of mutations in known disease-causing genes (intragenic duplications of *ATRX* in ATR-X syndrome[125]) and the identification of novel syndromes (*CPB* duplication[130]). We also demonstrated that a significant fraction of detected chromosome aberrations (4/16) do not consist of simple imbalances but rather are of complex nature, involving multiple deletions, duplications and inversions[125,152-154]. These additional findings resulted in a number of separate studies that are described below.

4.1. identification of intragenic duplications in *ATRX* as a cause for ATR-X syndrome

4.1.1. introduction

ATR-X syndrome is a rare X-linked syndrome characterized by severe to profound mental retardation, characteristic facial appearance, genital anomalies and alpha thalassaemia[155]. A mutation is identified only in a subset of patients with clinical suspicion of ATR-X syndrome. In this study we show that some of the patients suspected of having ATR-X syndrome but negative for mutation analysis, carry a large duplication inside ATRX, thereby extending the spectrum of mutation associated with this disorder. In one patient we show that this duplication causes a severe reduction of ATRX mRNA and absence of the ATRX protein.

4.1.2. clinical descriptions

4.1.2.1. family 1

Patient 1a (patient 16 in Table 9 and Table 10) was born at term, after a normal pregnancy (Figure 4). There was peripartal asphyxia. Biometry was at the 50th centile. The neonatal period was complicated by convulsions and severe hypotonia. There was hypertelorism, low set ears, inverted nipples and female external genitalia with small labia minores. Brain magnetic resonance imaging (MRI) was normal and he had an atrial septum defect type II. Karyotype was normal male, 46,XY in white blood cells and skin fibroblasts.

The child died at the age of 4.5 months from hypoventilation. Autopsy confirmed true male hermaphroditism with absent uterus and two dysgenetic testes. The gallbladder was absent.

During the second pregnancy, ultrasound examination at 15 weeks gestation indicated a female appearance of the external genitalia. Amniocentesis showed a normal male, 46,XY karyotype (patient 1b, Figure 4). At 19 weeks of gestation, ultrasound confirmed the female appearance of the external genitalia and detected an ASD. The pregnancy was interrupted. Ambiguous genitalia, with a severe hypospadias and a micropenis were present. Necropsy confirmed the presence of a large ASD type II. There was no uterus and the gonads were intra-abdominal. Also, the gallbladder was absent and bilateral bilobar lungs were noted. The normal male karyotype was confirmed on cultured fetal skin fibroblasts.

4.1.2.2.   family 2

Patient 2 (Figure 4) was born at term after caesarean section for prolonged labor. APGAR score was 6 after 10 min. Birth weight was 3.2kg (25-50th centile), length 53.5cm (50-75th centile) and head circumference 35cm (50th centile). He needed nasogastric tube feeding and there was failure to thrive. At age 6 months, weight was 4.53kg (3rd centile=5.3kg), length 61cm (3rd centile) and head circumference 39.2cm (3rd centile = 40). There was cryptorchidism, a small penis and bilateral clinodactyly of the fifth fingers. He had anteversion of the nostrils, a broad columella and hypertelorism with epicanthic folds. Karyotype was normal male 46,XY. Development was profoundly delayed with severe axial hypotonia and peripheral hypertonia. MRI scan of the brain revealed agenesis of the corpus callosum. He had a sensorineural hearing loss of 50dB and suffered from chronic unexplained anemia. Cardiac ultrasound was normal. He died unexpectedly at age 20 years.
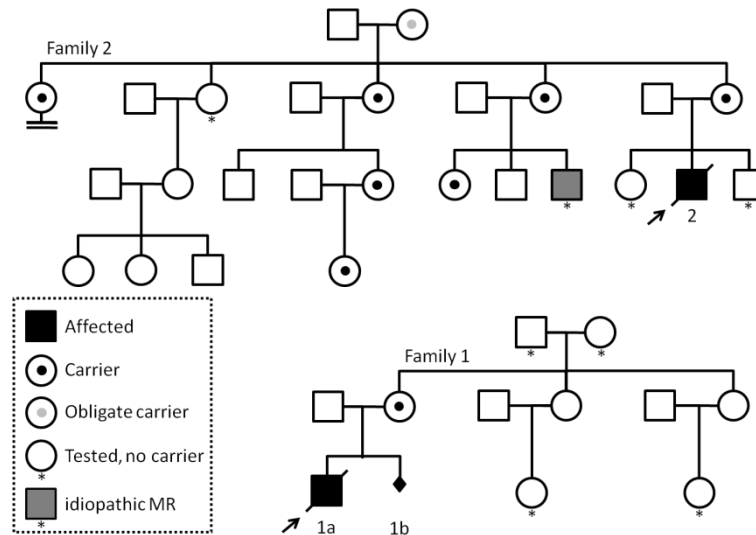
Figure 4: Pedigrees of the families where an intragenic *ATRX* duplication was found. The arrow indicates the proband, the individuals which were tested and shown not to carry a duplication indicated by an asterix.

4.1.3. results

In patient 1a, aCGH at 1Mb resolution revealed the duplication of two neighboring clones located on the X chromosome (RP5-875J14 and RP3-465G10). In his sibling (patient 1b), aCGH using the X-chromosome array revealed two non-contiguous duplications, one inside the ATRX gene, and one upstream (Figure 5a).

rtqPCR using primer pairs ATRX i1-2, i8-9 and i28-29 (Table 5) on a cohort of 50 patients with suspected ATR-X syndrome but without a detected sequence alteration in this gene revealed the presence of a duplication in one additional patient (patient 2). aCGH using the chromosome X microarray confirmed the presence of the duplication in patient 2 (Figure 5A). Consecutive rounds of rtqPCR using primer pairs ATRX e1, i1-2, e2, ,i2-3, i8-9, i28-29, e29, e30, e35 and e36 (Figure 5B, Table 5) demonstrated both duplications have different breakpoints. They span exon 2 to 35 in family 1 (222-281 kb) and exon 2 to 29 in family 2 (143-184 kb) (Figure 5C).

Further analysis by rtqPCR in the families indicated that both mothers carry the *ATRX* duplications. In family 1, the maternal grandparents did not carry the duplication. Likewise, in family 2, the healthy brother and sister of the index patient had no duplication. The 3 of the 4 sisters of the mother also carry the duplication (Figure 4).

Both carrier mothers had completely skewed X-inactivation (family 1: 100%, family 2: 99.35%). Analysis of the polymorphic AR repeat[96] (located approximately 11Mb from *ATRX*) revealed that both patients 1a and 1b carried the same allele, located on the inactivated X-chromosome of their mother. She inherited this chromosome from her father, indicating the
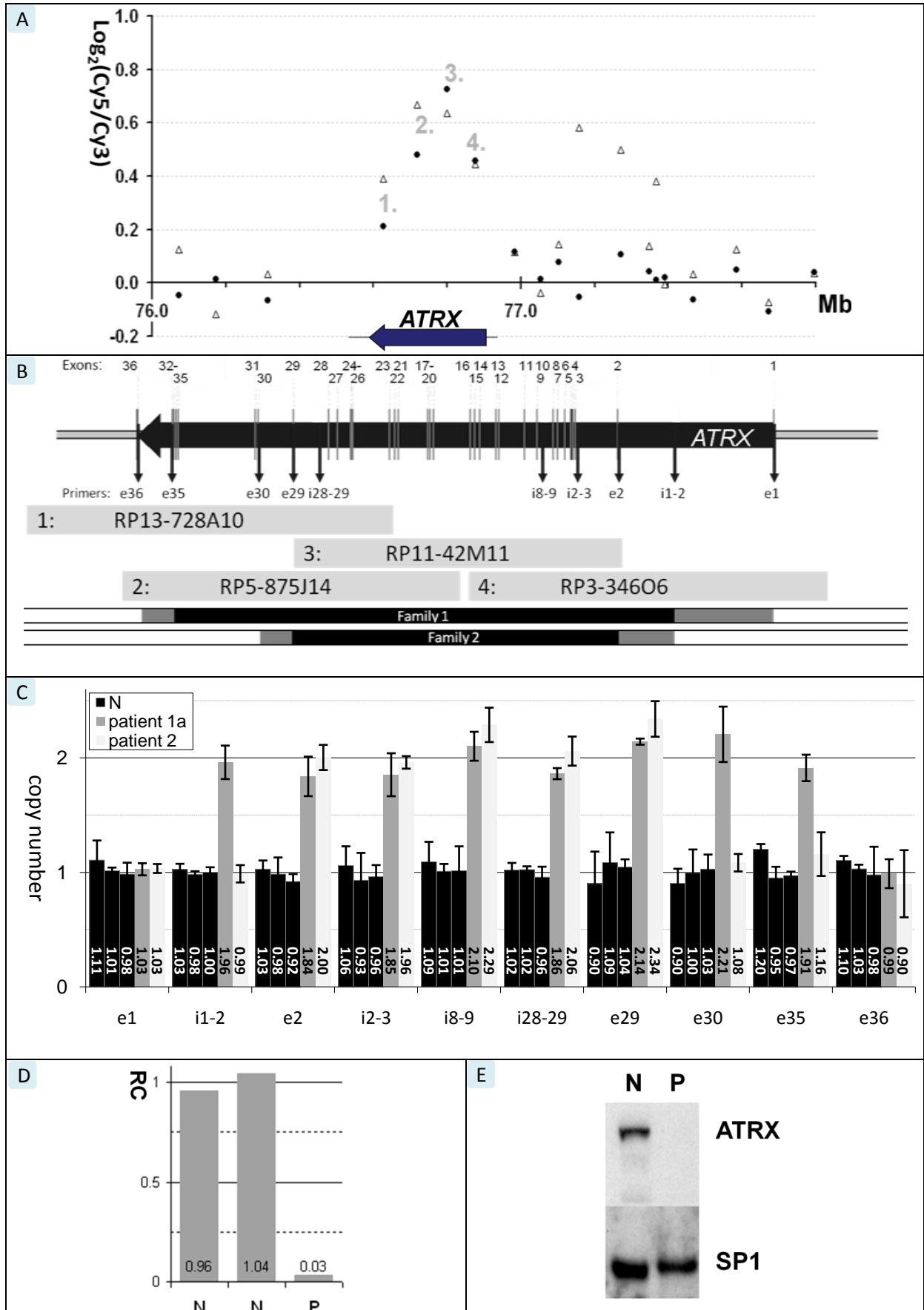
Figure 5 A: aCGH. aCGH analysis on patients 1a (triangle) and 2 (circle) showing an amplification of RP13-728A10 (1.), RP5-875J14 (2.), RP11-42M11 (3.) and RP3-346O6 (4.) in both patients and an additional duplication from RP3-465G10 to RP13-213F13 in patient 1a. The position of *ATRX* is indicated below. Patient DNA labeled with Cy5. **B: The ATRX region.**

The positions of BAC clones fully or partially duplicated on aCGH and all QPCR primers used in this study, relative to the ATRX gene (36 exons are indicated, with exon 7 spliced out. accession no. U72937). Primer sequences are listed in Supplementary Table 1. Black and grey bars below the region indicate the certainly and potentially duplicated region, respectively. C: rtqPCR. Results of rtqPCR analysis of three normal males (N), patient 1a and patient 2. D: mRNA expression. The expression of ATRX mRNA in cultured skin fibroblasts of patient 1b (P) is compared to normal 46,XY fetal skin fibroblasts (N) cultured in parallel. Values are normalized to N. RC=relative concentration. rtqPCR on cDNA was performed using primers expr_forw and expr_rev designed in exons 1 and 2. E: Western blot. The expression of the ATRX protein in cultured skin fibroblasts of patient 1b (P) is compared to that in normal 46,XY fetal skin fibroblasts (N) cultured in parallel. SP1=a nuclear protein serving as control for nuclear protein extraction.

duplication occurred in the grandfather. Also patient 2 and his mother's inactivated X-chromosome carry the same AR allele.

To assess the effect of this duplication on the functioning of *ATRX*, *ATRX* mRNA levels were quantified in a cell line from patient 1b, showing a reduction to 3% of the normal level (p < 0.01; Figure 5 D). RT-PCR on mRNA from his mother showed normal *ATRX* mRNA levels (data not shown). In addition, Western blot analysis revealed a total absence of ATRX in a nuclear protein extract from the fibroblast cell line from patient 1b (Figure 5 E).

4.1.4. discussion

In a family with two siblings presenting an unexplained disorder with cardiac and genital malformations, aCGH led to the identification of an intragenic duplication of the *ATRX* gene. In retrospect, the clinical features are fully compatible with the ATR-X syndrome. Less classical features, including true male pseudohermaphroditism and congenital heart defects have been described before[155]. However, we are unaware of previous reports of absent gallbladder in this syndrome.

Extending this study in a cohort of 50 additional patients suspected for ATR-X syndrome led to the identification of one additional patient carrying an intragenic *ATRX* duplication. Given the position of the duplications, they are expected to result in a loss of function. This was confirmed in skin fibroblasts from one of the patients, patient 1b. RT-PCR showed a drastic reduction of the level of ATRX mRNA resulting in a reduction in the level of ATRX protein below the detection limit of Western blot, demonstrating the detrimental nature of the intragenic duplication. Previous studies similarly showed a drastic reduction or even an apparent loss of the ATRX protein in some mutation carriers[156].

This observation of an intragenic *ATRX* duplication leading to gene disruption extends the spectrum of mutations causing the ATR-X syndrome. This has important practical consequences: traditional mutation analysis strategies relying on non-quantitative techniques for sequence analysis will need to be complemented by additional techniques

allowing the detection of copy number changes. In other genes checked for gross genomic rearrangements, deletions and duplications are detected but their frequency varies dramatically: deletions are detected in *DMD*, *TSC1*, *TSC2*, *CFTR* and *NF1* with reported frequencies of 65%, 0.45%, 6%, 1.5% and 2%, while intragenic duplications account for 7%, 0%, 0.24%, 0% and 0.3% of mutations[126,157-160]. The reason why duplications are in general less frequent is not known. One reason could be that the mechanisms generating duplications are more complex than those generating deletions. While most deletions are intrachromosomal rearrangements, duplication mechanistically are probably interchromosomal rearrangements. Moreover, certain whole gene duplications lead to a different phenotype than the loss-of-function phenotype associated with whole gene deletions, and will thus not be ascertained, as was seen for example in *MECP2*[161]. For the gene in which most duplications were hitherto characterized, *DMD*, no bias towards larger duplications was detected[157]. Because in the present study patients were not checked for duplications in each of the 36 exons of the *ATRX* gene, we expect that the frequency of duplication mutations may be higher than reported. Also in this perspective it is somewhat surprising that gross *ATRX* deletions are not yet detected in ATR-X patients, while we report on 2 different duplication events, one of which apparently causes loss-of-protein.

aCGH detected two flanking non-contiguous duplications in family 1: one inside *ATRX* and one spanning *ATP7A*, *PGAM4*, *PGK1* and *TAF9B*. We cannot exclude that this second duplication contributed to the phenotype of the siblings in family 1. Although uncommon, recent higher resolution analyses show that this type of complex intrachromosomal rearrangements occur more often than hitherto appreciated[153,157]. This implies caution for the extrapolation of copy-number measurements in discrete genomic regions to the regions in between. In this study, aCGH at 1Mb resolution showed 2 duplicated clones in family 1. Extrapolation would have implied the duplication of *ATRX* is not intragenic. Only higher resolution analysis revealed this were 2 non-contiguous duplications, with one disrupting *ATRX*.

In conclusion, this observation adds a novel type of mutations detectable in the *ATRX* gene, underscoring that quantitative analyses should be an integral part of mutation analysis in this and other disease genes.

4.2. left-ventricular non-compaction associated with monosomy 1p36

4.2.1. chromosomal anomaly

Routine karyotyping using G-banding analysis at ISCN +550 bands was performed using standard procedures, but revealed no cytogenetic abnormalities. aCGH at 1 Megabase (Mb) resolution was performed to detect submicroscopic imbalances[162], and identified a 4.6 – 5.9 Mb terminal deletion of the short arm of chromosome 1, with the breakpoint located between RP1-37J18, the most centromeric deleted clone, and RP11-49J3, the most telomeric normal clone. Fluorescence in situ hybridization on chromosome spreads from cultured lymphocytes from the patient and his parents was performed using a probe located in the 1p subtelomeric region, RP11-465B22. This analysis confirmed the presence of the described deletion and showed it had occurred *de novo*.

4.2.2. clinical description

The patient is the first child of healthy, unrelated parents. Family history is negative with regard to congenital malformations or developmental delay. She was born after an uneventful pregnancy, at 36.5 weeks, after premature rupture of the membranes. Birth weight was 1930 g (P5 = 2050 g), length: 44 cm (P5 = 44,5 cm), head circumference 30,5 cm (P5 = 30,8 cm).

There was a facial dysmorphism with a large anterior fontanel, low set ears with narrow auditory canals and upslanting of the palpebral fissures. Cardiac ultrasound revealed the presence of multiple small muscular ventricular septal defects and a cardiomyopathy of the non-compaction-type with non-compacted myocardium at the apex and left ventricular posterior wall (Figure 6). Initially there was a mild hypocontractility with a fractional shortening of 25%. At the age of 2 months during routine follow-up a more important left ventricular (LV) dysfunction was noted with LV dilatation (LV end diastolic diameter: 26 mm) and a fractional shortening of only 13%. There was mild mitral regurgitation. Clinically there was mild respiratory distress with hepatomegaly (liver 3 cm below right costal margin). At that time cardiac medication (digoxin, diuretics, lisinopril and carvedilol) was started. This resulted in normalization of cardiac function within 2 weeks. Cardiac function remained

stable during further follow-up. The cardiac medication was continued. The small muscular ventricular septal defects closed spontaneously by the age of 5 months.

At age 3 months, she presented convulsions (tonic-clonic with focal onset, myoclonic), controlled by means of carbamazepine treatment. MRI of the brain at age 3 months did not reveal any anomalies. Ophthalmologic examination was normal. There was a delayed psychomotor and growth development.
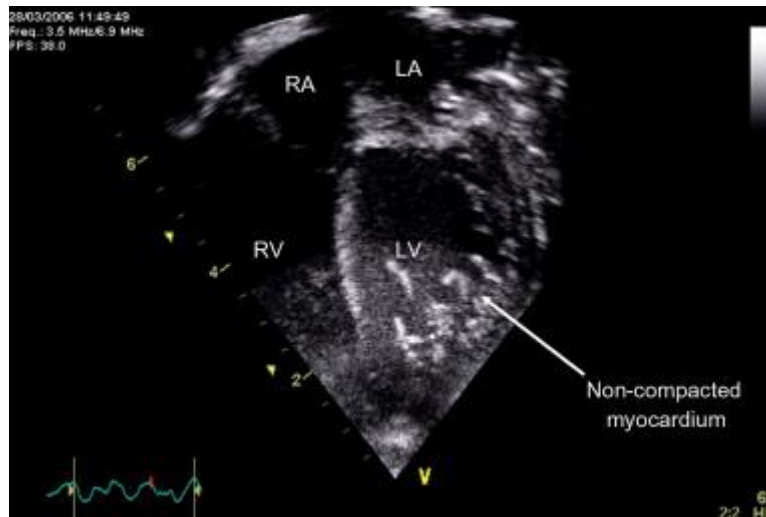
Figure 6: Four-chamber-view illustrating the apical and lateral non-compaction of the left ventricular myocardium. LV: left ventricle, RV: right ventricle, LA: left atrium, RA: right atrium.

4.2.3. discussion

We report on a new-born girl with left ventricular non-compaction (LVNC), dysmorphism and epilepsy. aCGH at 1Mb resolution revealed a deletion of the terminal 4.6 to 5.9 Mb of the short arm of chromosome 1. Deletions of chromosome 1p are amongst the most common terminal deletions, occurring with a reported frequency of up to 1 in 5000[163]. Around 40% of all chromosomal breaks in this region seem to occur 3-5Mb of the telomere, as in the present patient[164].

This is the first report describing LVNC in association with the 1p36 deletion syndrome. Cardiac abnormalities such as dilated cardiomyopathy (DCM) and structural cardiac defects are common findings in patients with monosomy 1p36. DCM has been described to occur in one in four patients, and structural abnormalities in one in two patients[164]. LVNC has been reported to occur in combination with dilated cardiomyopathy and structural congenital heart disease, but until now has not been reported in monosomy 1p36 syndrome. LVNC has been associated with other chromosomal abnormalities including unbalanced translocation with deletion of 1q43qter and duplication of 4q31qter[165], interstitial deletion of 1q43[166],

deletion of 5q35[167] and trisomy 13[168]. Apart from chromosomal abnormalities, mutations in multiple genes have been described in association with LVNC, including *AMPD1* (1p13.2), *LMNA* (1q21), *FXN* (9q13), *LDB3* (10q23.2), *PMP22* (17p11.2), *GAA* (17q25.2), *DTNA* (18q12.1), *DMPK* (19q13.3), *DMD* (Xp21) and *TAZ* (Xq28), as well as certain mitochondrial genes (for a review, see Finsterer *et al.*, 2006[169]). This multitude of genes and chromosomal regions indicates the marked genetic heterogeneity of NCLV.

The fact LVNC was until now unreported in monosomy 1p36 syndrome probably has multiple explanations. First, LVNC might have remained undetected, since its detection has long been technically challenging[170]. Second, mildly affected patients might have been evaluated as normal, since the degree of LVNC is reported to vary from a nearly normal myocardium to definitely pathological[171]. Finally, for many other genes associated with LVNC, only a subset of mutation carriers is affected[169]. It is therefore not surprising that not all monosomy 1p36 carriers have LVNC.

Whether LVNC is an acquired or a congenital defect is a matter of debate: from a developmental point of view, LVNC could be considered as an arrest in the development of trabecular compaction[170]. Alternatively LVNC could also be considered as a secondary compensatory change just like the hypertrophic response. Further research is required on the exact pathogenesis. The association of monosomy 1p and LVNC in this patient again links LVNC with DCM, which was previously associated with monosomy 1p36.

In conclusion, the finding of LVNC in this patient broadens the spectrum of cardiac anomalies found in association with the monosomy 1p36 syndrome. Conversely, in patients with syndromic forms of LVNC (i.e. associated features such as epilepsy, microcephaly, developmental delay or dysmorphism), cytogenetic studies are indicated, more specifically for monosomy 1p36[164]. Since deletions of 1p36 are often missed in routine karyotyping[163] as in this patient, preferentially a molecular detection technique is used. Given the multitude of loci implicated in LVNC, a genome-wide screening technique such as aCGH might be more successful in a patient with multiple congenital anomalies (features suggesting a chromosomal abnormality) but without an obvious diagnosis. Finally, this finding suggests that the unknown gene responsible for the cardiac defects in monosomy 1p36 syndrome is part of the group of genes mentioned above, causing both DCM, LVNC as well as structural cardiac defects.

4.2.4. *note on this report*

Since the publication of this report, a further 14 monosomy 1p36 carriers with LVNC have been reported[132,172,173] and the frequency of LVNC in monosomy 1p36 is estimated at 23%[132], showing that LVNC is indeed a frequent cardiopathie in monosomy 1p36 syndrome.

4.3. microduplication of *CBP*

Adapted from

4.3.1. introduction

Submicroscopic chromosomal imbalances are an important cause of congenital disorders. The occurrence of some recurrent chromosomal imbalances is explained by non-allelic homologous recombination between low copy repeats (LCRs) that flank the imbalanced region[174]. The identification of duplications as well as deletions of the same region in 17p11.2 suggested the existence of reciprocal imbalances to known recurrent microdeletions or microduplications[175,176]. The lack of recognizable phenotypic selection criteria made them however difficult to identify. The identification of duplications reciprocal to the deletion of 22q11.2 (DiGeorge)[177], 7q11.23 (Williams-Beuren)[178], and 17p11.2 (Smith-Magenis)[179] preceded the identification of the corresponding phenotype caused by these duplications, and some reciprocal duplications have been found to be benign variants[180].

To facilitate the detection of chromosomal imbalances causing recognizable syndromes as well as the discovery of their reciprocal imbalance, our whole genome micro-array was enriched with probes targeting regions known to be dosage-sensitive upon haplo-insufficiency or duplication (regardless of whether they were flanked or not by LCRs). Previous reports describe targeted microarrays dedicated to detect copy number changes in regions flanked by LCRs. Although they have proven successful[181], the present report highlights the importance of also looking for imbalances in other regions of the human genome, and primarily in regions that are known to be dosage sensitive upon deletion.

4.3.2. clinical description

The patient is the first-born male child of healthy, non-consanguineous Caucasian parents. Birth weight and length at term were 2,600 g (~3rd centile) and 44 cm (below 3rd centile). He was diagnosed with a type 2 atrial septal defect with displacement of the interatrial septum (partially overriding the inferior caval vein). His development was moderately delayed: there was little social contact at the age of 4 months and he walked independently at age 24 months. Now at age 15 years, he participates in special schooling

for persons with a mild to moderate mental handicap. At examination, his weight was 62 kg, length 173 cm, and head circumference 55.4 cm (all near the 75th percentile). Clinical evaluation showed minor anomalies including mildly protruding ears, a broad nose, hirsutism with synophrys and a low frontal hair line, mild cutaneous syndactyly of the fingers, camptodactyly of fingers 2-4, and incomplete extension of the elbows (Figure 7). The family history is negative for similarly affected individuals.

Figure 7: Patients' phenotype. Note the protruding ears, broad nose, synophrys, low frontal hair line, and camptodactyly of fingers 2-4. (Reprinted with permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc.)

4.3.3. results

In the absence of an etiological diagnosis, DNA from the propositus was subjected to aCGH analysis. This showed a duplication of two cosmids (RT100 and RT102) mapped to *CBP* (Figure 8a-d)[94]. rtqPCR analysis confirmed the presence of the duplication in the patient, and showed his mother does not carry the duplication (Figure 8c). Paternal DNA was not available. Sequential rounds of rtqPCR revealed the duplication is between 345 and 480 kb in size, spanning *TRAP1* and *CBP* entirely, and part of *ADCY9* and *DNASE1* (Figure 8d). FISH on metaphase chromosome spreads from the propositus showed a signal on both chromosomes 16, one weak and one stronger. The majority of interphase nuclei showed 3 signals (data not shown). To determine the parental origin of this duplication, marker analysis was performed. None of the six polymorphic markers inside the duplicated region showed more than two alleles, strongly suggesting the duplication arose intrachromosomally. Markers D16S2622 and MS2 showed that the duplicated allele corresponds to the maternal allele (Figure 8e and f). Since rtqPCR analysis demonstrated that the mother does not carry the duplication (Figure 8c), we conclude that it arose de novo.
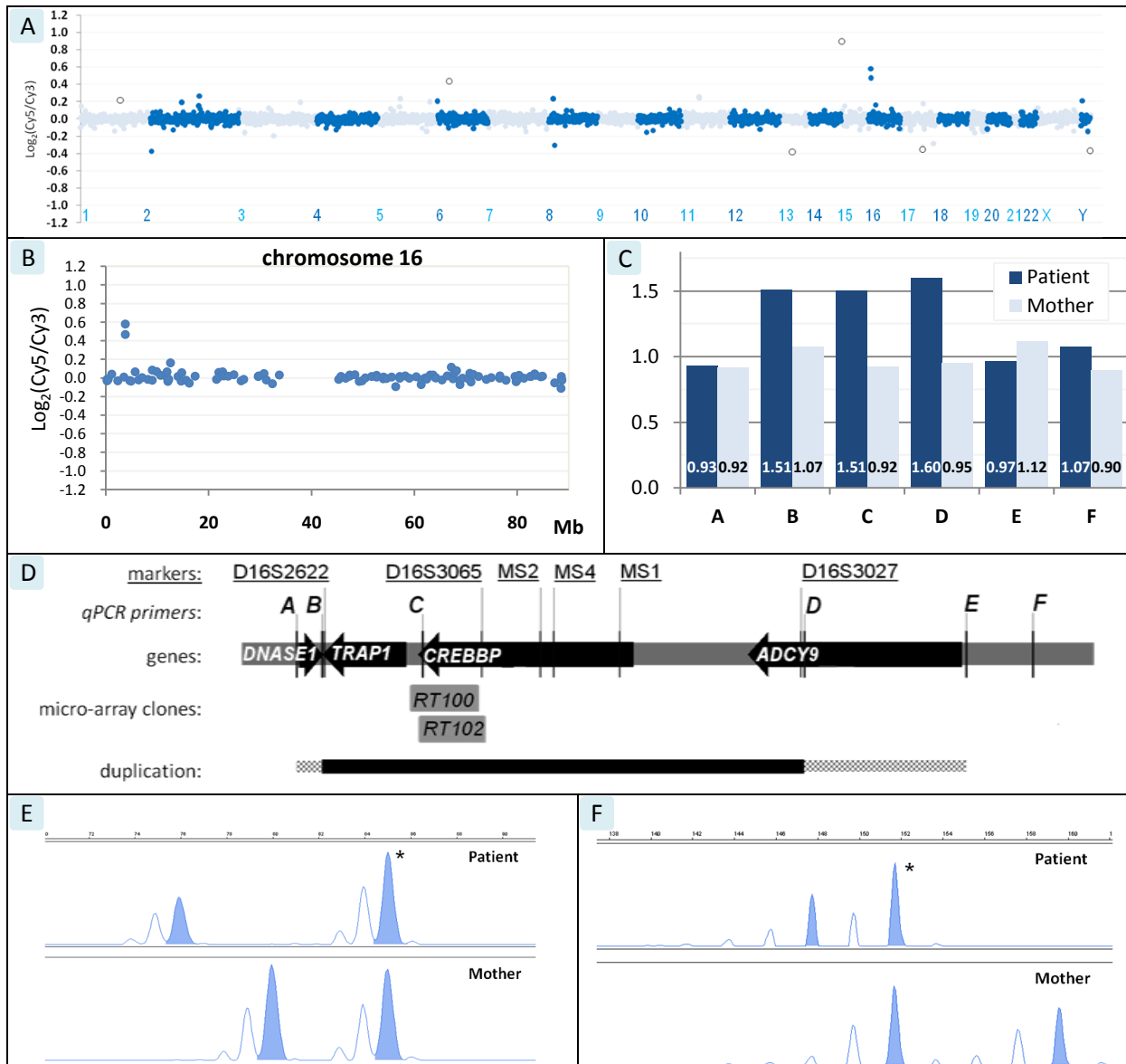
Figure 8: A: Molecular karyotype. X-axis: clones ordered by chromosome (p ⟶ q-arm); Y-axis: log$_2$ (Cy5/Cy3); unfilled dots: known polymorphic clones; dotted line: 4× standard deviation (SD) threshold; full line: log$_2$(1.5) - 2 × SD threshold[91]. **B:** Partial molecular karyotype for chromosome 16. Duplication of clones RT100 and RT102. Plot legend as in (A), except X-axis: clone position [Mb]. **C:** qPCR analysis and data processing was performed as described[89]. Intensities (indicated at the base) were normalized by comparison to three normal controls. X-axis: qPCR primer pairs used; Y-axis: relative copy number. **D:** Map of the *CBP* region. Physical location of polymorphic markers (underlined), qPCR primer pairs (italics), known genes (black arrow) and duplicated microarray clones (gray boxes). The shaded and the black bar below indicate, respectively the potentially duplicated and the duplicated area in the present patient. Results from polymorphic marker analysis for D16S2622 (E) and MS2[182] (F) showing that the duplicated allele (*) from the patient (upper box) is inherited from his mother (lower box). (Reprinted with permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc.)

4.3.4. discussion

We report on a patient with moderate mental retardation, a congenital heart defect and various minor physical anomalies. aCGH analysis of DNA from this patient revealed a duplication of the region encoding CREB-binding protein (CBP). rtqPCR and marker analyses showed this duplication was maternal and intrachromosomal in origin and occurred de novo.

Mutations in *CBP*[120] and in *EP300*[183], and micro-deletions of the *CBP* region[184] are known to cause Rubinstein-Taybi syndrome (RSTS). This is a rare congenital disorder affecting about one in 125,000 newborns[185]. Its main clinical features are developmental delay, heart defects, a characteristic face, and broad first toes and thumbs[186,187]. A duplication of only the first exon of *CBP* has been described in one patient with classical signs of RSTS, suggesting this to be an intragenic duplication causing haplo-insufficiency of CBP[183]. The duplication found in the present patient encompasses *CBP* and *TRAP1* entirely. The duplication of *TRAP1* most likely does not influence the patients phenotype: firstly because mutations in *TRAP1* are not yet described in association with human malformations, and secondly because some RSTS patients carry a microdeletion affecting *CBP* as well as *TRAP1*, without overt phenotypic difference, suggesting *TRAP1* is not dosage-sensitive upon haplo-insufficiency.

Duplication of the subtelomeric part of the short arm of chromosome 16 causes a recognizable phenotype characterized by psychomotor and growth retardation, microcephaly, seizures, cardiac and hand anomalies and a specific facial appearance[188]. The most telomeric band (16p13.3) is about 6 Mb in size and contains about 260 genes, making it particularly gene-rich. Duplication of this region has been found to be one of the more frequent subtelomeric duplications, occurring in seven patients out of a group of almost 12,000 selected for a variety of indications[189]. In seven patients, submicroscopic duplications of this region have already been sized beyond banding resolution using aCGH: five subtelomeric duplications were delineated[190,191], one interstitial duplication was reported[192] and one interstitial duplication was found in an aCGH screening study[193]. All duplications encompass *CBP*, except one unbalanced translocation between 16p and 17p. Comparison with the present patient indicates that duplication of *CBP* causes part of the clinical features associated with the 16p duplication phenotype, namely congenital heart defects and developmental delay. Other features commonly associated with subtelomeric 16p duplications (growth delay and microcephaly) were not found in the three patients carrying interstitial duplications. We conclude that one or more genes telomeric to *CBP* are also dosage-sensitive, and that their imbalance contributes to the more severe phenotype of other patients with a duplication of 16p13. Further support for this notion is provided by the absence of large deletions that span both *CBP* and *PKD1*, a gene located about 1.6 Mb more telomeric on 16p. This was suggested to indicate the presence of one or more hemizygous-lethal genes in the intervening gene-rich interval[194]. aCGH analysis of other cases will prove

essential for further dissection of this and other contiguous gene syndromes into their different causative facets.

Not surprisingly, the duplication of *CBP* in this patient affects similar parts of the body as haplo-insufficiency in RSTS (brain, heart). In general, duplications cause a less severe phenotype than deletions[174], as is observed here. We would not suggest a remarkable "reciprocal phenotype", as has been suggested e.g. for Williams-Beuren syndrome and its reciprocal 7q11.23 duplication[178].

CPB encodes a protein functioning in chromatin remodeling[195]. Another gene involved in chromatin remodeling, *MECP2*, is known to cause Rett syndrome upon deletion or mutation[196]. Recently duplications of this gene were shown in patients with severe to profound MR and progressive spacticity[161]. As for MeCP2, maintaining CBP levels within narrow limits seems essential for its proper functioning. This shows again that chromatin remodeling is exquisitely sensitive to dosage perturbations.

4.4. intrachromosomal rearrangements are frequently complex

Adapted from

4.4.1. introduction

4.4.1.1. complex chromosomal rearrangements

Complex chromosomal rearrangements (CCRs) have been defined as constitutional structural rearrangements involving three or more breakpoints[197]. CCRs have been classified based on their inheritance (familial or *de novo*), the number of breakpoints that are involved (less or more than 4) or the number of chromosomes involved. For a review we refer to Batanian and Eswara[198]. Intrachromosomal rearrangements involving 3 breakpoints are not regarded here as complex: although some may originate from mechanisms similar to those generating CCRs, most are terminal deletion-duplications (generated by breakage-fusion-bridge cycles[199,200]) and insertional translocations.

4.4.1.2. frequency

CCRs have long been considered rare, found only in a minute subset of patients that carry chromosomal imbalances[198]. Aberrant G-banding patterns are still intuitively attributed to simple deletions or duplications. The exact nature and extent of an imbalance is often not studied in detail using an independent technique (e.g. FISH) and often only one or a few loci are tested, because these are very laborious investigations. Very few aberrations detected

by microscopic karyotyping are thus mapped with a high resolution. Only when multiple chromosomes are visibly aberrant, it is evident that multiple chromosomal breaks were present in one patient. This probably explains why most reported CCRs detected by classic microscopy karyotyping are interchromosomal (affecting multiple chromosomes)[198] and only seven reports describe the detection of complex intrachromosomal rearrangements (CiCRs) by microscopy karyotyping. Single chromosomes involved in these reports were 2[201,202], 4[203], 10[204] and 21[205], and the number of breaks involved varied from 4 to 12.

4.4.1.3.   the impact of the introduction of aCGH

Following the development of aCGH[103,104] and its introduction as a routine constitutional karyotyping technique[89,206], a much more precise delineation of breakpoints of imbalances automatically ensued. As a consequence, an increasing amount of CiCRs have now been detected using this technique[58,125,127,153,157,207-217] affecting chromosomes 1 (3x), 4 (2x), 5, 7, 8, 9, 13, 17 (2x), 18, 19 and X (multiple cases). Similarly, also interchromosomal rearrangements are often found to be more complex upon closer examination[218,219], as they frequently appear to harbor cryptic indels.

4.4.2. clinical descriptions

4.4.2.1.   patient 3

This boy is the first child of unrelated parents, who both have mild mental handicap of unknown cause. Family history is otherwise negative with regard to mental handicap or congenital malformations. He was born at term pregnancy, with weight 3580 g (P75-97), length 51 cm (P75-97), and occipitofrontal circumference (OFC) 34 cm (P50-75). He had a critical aortic valve stenosis, a cleft of the soft palate and grade 1 hypospadias. There were multiple dysmorphic features, including transverse palmar creases, retrognathia, plagiocephaly, persistent oedema of the feet and a preauricular ear fistula in the right side. There was mild global body asymmetry. During follow-up, growth was on the 75th centile for weight and OFC and on the 75th-90th centile for length. Development was mildly delayed: at age 11 months, his development was 8 months according to the Bayley developmental scale (developmental index of 77), and at age 22 months, development was 16 months.

4.4.2.2. patient 10

This girl was born at 36 weeks, with birth weight of 2400 g, length of 44 cm and OFC of 31.8 cm (all parameters below the third percentile). Cardiac examination revealed a tetralogy of Fallot which was surgically corrected, and a right aortic arch. Long term clinical follow-up revealed a sensorineural hearing loss (-40db) at the age of 4 years, and she was integrated in a special school for children with mild to moderate mental handicap. Postnatal growth retardation was severe at that age: weight 10.4 kg (3rd centile 16 kg); height 87 cm (3rd centile 92 cm) and OFC 43.4 cm. After three episodes of generalized tonic-clonic convulsions antiepileptic treatment was started. Growth hormone therapy was started at the age of 5 years, with a significant catch up of growth and physical development. The fixed adducted position of the thumbs regressed spontaneously. Now at the age of 12.5 years height is 140 cm (3rd centile) and pubertal development is normal. She presents with Gilles de la Tourette syndrome.

4.4.2.3. patient 13

This girl was born as the second child of healthy, non-consanguineous parents. Family history is negative with regard to mental handicap and congenital malformations. Pregnancy was complicated by placental abruption. She was born at a gestational age of 37 weeks. Birth weight and length were not available. She was diagnosed at birth with multiple small ventricular septal defects (VSDs) and a large midmuscular VSD. Aged 2.5 years she underwent a surgical correction of strabismus. Carbamazepine was started at the age of 4 years because of myoclonic seizures. Brain MRI was normal. At the age of 5 she entered special education for severe psychomotor delay, scoring an IQ of 52 on the SON-R non-verbal intelligence test. Clinical examination at age 11 showed a biometry with all parameters below the third centile (weight = 22.8 kg , length = 124 cm, head circumference = 50 cm). There was facial dysmorphism with a broad nasal bridge, a thin upper lip, mild retrognathia and bilateral ptosis. She had long fingers, talipes valgus and a sacral dimple.

4.4.2.4. patient 16 and 16b

Clinical details on patients 16 and 16b are described in the report on the *ATRX* duplication (Chapter 4.1.2.1). They are respectively patients 1a and 1b from family 1.

4.4.3. results

Routine karyotyping of peripheral lymphocytes (patients 3, 10, 13, 16a and 16b) and cultured skin fibroblasts (patients 3 and 16b) did not reveal any chromosomal aberrations. In the absence of an etiological diagnosis, DNA from patients 3, 10, 13 and 16a was analyzed by whole genome aCGH at a 1Mb resolution.

4.4.3.1. patient 3

In patient 3, aCGH at a 1Mb resolution showed a 12-14Mb deletion of band 4q34. In retrospect this deletion was visible on the G-banded karyotype (Figure 9A). Interestingly, one clone (RP11-148L24) inside the deletion region showed a normal ratio, suggesting this clone was not deleted. FISH analysis confirmed this clone was still present on the aberrant chromosome 4, indicating the presence of two separate deleted regions rather than one large deletion. To characterize this rearrangement further, we performed aCGH using the full tiling path array platform for chromosome 4 (Figure 9 B). This analysis confirmed the presence of both flanking deletions (171.31Mb→175.71Mb and 176.05Mb→184.28Mb), and additionally showed a proximal 350kb duplication (162.40Mb→162.95Mb), in band 4q32.2. rtqPCR with a primer pair designed in the purported duplicated region (Table 5) confirmed the presence of the duplication. To examine the chromosomal organization of the duplication, we performed FISH. Surprisingly, 2 separate signals on the aberrant chromosome 4 were observed: one at about the expected position, and one more proximal at 4q28 or 4q31.1 (Figure 9 E). This finding prompted us to investigate the position and orientation of the chromosomal segment between the duplication and the deletion. FISH with probes RP11-1G8 (163.8 Mb) and RP11-275K4 (171.2 Mb) labeled with red and green fluorescence showed this segment was present in an inverted orientation. Multicolor banding FISH for chromosome 4 was performed to confirm the presence and determine the extent of the inversion. The inversion spans from 4q28 to 4q34 (Figure 9 C). The karyotype of the patient can therefore be summarized as follows: 46,XY .arr cgh der(4)(del(4)(q34.2→q34.3)del(q33→q34.1)dup(q32.2)

.ish(der4)(4pter→4q28::4q33→4q28::4q32.2→4q32.2::4q34.1→4q34.1::4q35.1→4qter)

To determine whether the rearrangement occurred *de novo*, FISH analysis of parental metaphase spreads was performed. FISH using clones RP11-213L8 and RP11-62B7 showed both clones to be present in the parents (Figure 9). rtqPCR of parental DNA showed the presence of two copies of the region duplicated in the child. Since chromosomal inversions

have been shown to be susceptibility factors for chromosomal rearrangements[220], the orientation of this chromosomal region was investigated in the parental chromosomes. Co-hybridization of differently labeled FISH clones (RP11-213L8, RP11-148L24, RP11-62B7) showed the expected order of probes as indicated by the human genome reference sequence (Figure 9). Similarly, multicolor banding FISH for chromosome 4 on parental metaphase spreads was normal (not shown). Therefore both deletions, the duplication and the inversion detected in the proband occurred *de novo*. STR length analysis using primer pair 4SXXX (Table 6) demonstrated that the duplication occurred on a single paternal chromosome (Figure 9).

4.4.3.2.   patient 10

aCGH at 1Mb resolution revealed 3 different imbalances on chromosome 13q: a small deletion in 13q21.31 called by a single clone (RP11-234O23), a deletion of 13q31 (first and last deleted clones: RP11-115N13 and RP11-94C14) and a duplication of 13q31.3q33.2 called by clones mapped immediately next to the deleted clones (RP11-632L2 to RP11-323K22) (Figure 10 A). Both deletions and the duplication were confirmed by FISH (Figure 10 middle panel). The duplicated fragment was shown to be inserted proximally an inverted orientation by FISH using differentially labeled RP11-632L2 and RP11-232K22. The chromosome imbalances were further delineated by aCGH on a Nimblegen whole genome 385K array. This demonstrated that the proximal deletion affected a segment starting at 62.01Mb and ending at 62.83Mb (size: 0.82Mb). The larger distal deletion affected a segment that is annotated to start at 82.07Mb and end at 91.89Mb (size: 9.82Mb). It is flanked telomerically by a small 0.76Mb segment of normal copy number. Finally, the large duplication ranged from 92.48 to 107.26 Mb. STR length analysis in the patient and his parents demonstrated that the duplication arose on a single paternal chromosome (Figure 10 upper right).
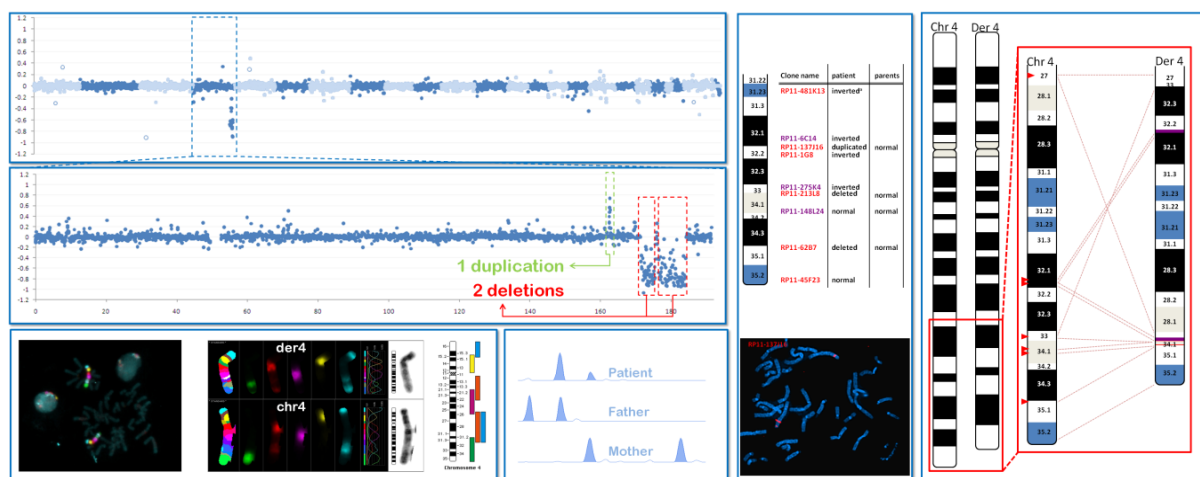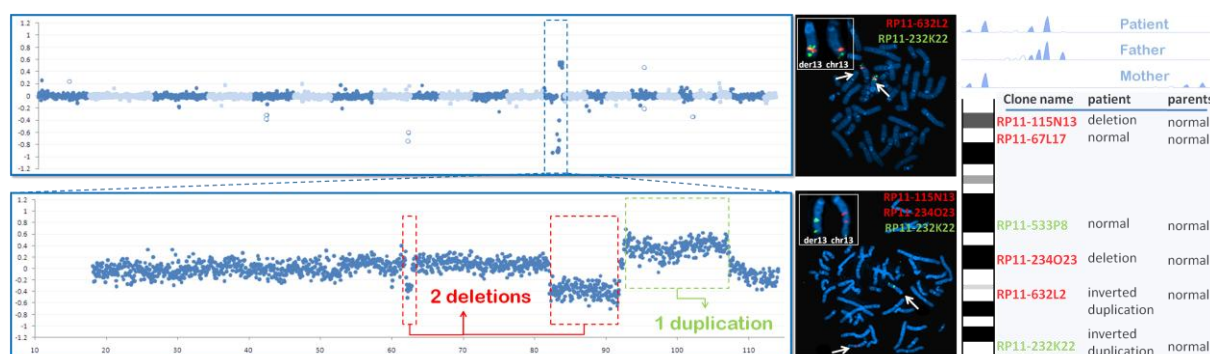
Figure 9: (A) Result of aCGH using a genome-wide micro-array with a 1Mb resolution, showing clones on chromosome 4 with a decreased copy-number. The y-axis shows $\log_2$ratios of Cy5 vs Cy3 signal intensities, with the patient is labeled in Cy5. Clones are ordered on the x-axis per chromosome (alternating colors) and based on their physical location. Empty dots indicate known polymorphic clones. **(B)** Result of aCGH using a chromosome 4 tiling-path BAC/PAC microarray, showing a region of duplication and 2 regions that of deletion on chromosome 4q32q34. Y axis as in (A). Clones are ordered on the X-axis based on their physical position on chromosome 4 [Mb]. **(C)** Overview of the probands M-band FISH results. left: A metaphase spread with M-band FISH probes in different pseudocolors. middle: summary of M-band FISH results for a normal (left) and derivative (right) chromosome 4. Right: overview of M-band pseudocolors. **(D)** analysis of an unannotated short tandem repeat located in the duplicated fragment for the patient and her parents, showing that the imbalance arose on a single maternal chromosome. **(E)** Overview of different FISH analyses. Clone names in blue and red indicate respectively SpectrumGreen™ or SpectrumOrange™ labeled clones. Results are shown for the proband and the parents. Below is shown the results for RP11-137J16, a probe inside the duplicated region showing hybridization on 2 different bands of der4. **(F)** Summary of the cytogenetic studies, showing the complexity of the rearrangement with at least 7 different breakpoints.

Figure 10: (A) Result of aCGH using a genome-wide micro-array with a 1Mb resolution, showing clones on chromosome 13 with an aberrant copy-number. The y-axis shows $\log_2$ratios of Cy5 vs Cy3 signal intensities, the patient is labeled in Cy5. Clones are ordered on the x-axis per chromosome (alternating colors) and based on their physical location. Empty dots indicate known polymorphic clones. **(B)** Partial result of aCGH using a Nimblegen 385K microarray, showing two deleted chromosome segments (13q21.31 and 13q31) and one duplicated segment (13q31.3q33.2). Y axis as in (A), but a sliding window of 7 median log2 values is plotted. Clones are ordered on the X-axis based on their physical position on chromosome 13 [Mb]. **(C)** Results of FISH on a metaphase spread of the proband using clones RP11-632L2 and RP11-232K22 labeled with Spectrum Orange and Spectrum Green respectively (pseudocolors red and green) showing an inverted duplication. **(D)** Results of FISH on a metaphase spread of the proband using clones RP11-115N13, RP11-234O23 and RP11-232K22 labeled with Spectrum Orange, Spectrum Orange and Spectrum Green respectively showing deletion of both proximal probes and a duplication of the distal probe. **(E)** Results of length analysis of an unannotated STR designed in the duplicated chromosome segment, showing that the duplication arose on a single paternal chromosome. **(F)** Summary of the results of FISH studies of the proband and her parents.

4.4.3.3.   patient 13

aCGH at 1Mb resolution revealed three contiguous clones mapping to 19p13.11 with a copy number significantly higher than expected, indicative of a duplication of this region

(CTD-2231E14, RP11-413M18, CTD-3149D2). rtqPCR using primer pairs described in Table 5 were used to confirm this aberration and to further map the breakpoints. This revealed a triplication, a duplication and normal copy number of the loci targeted by respectively primer pairs CCR19_B, CCR19_D and CCR19_G (11 C)[151]. To further analyze this unusual finding, aCGH using a full tiling path clone set for chromosome 19 was performed. This revealed the presence of a triplication of the region from 15.6 to 16.8 Mb, duplication of the flanking region until 18.1 Mb and again triplication until 18.9 Mb. The last normal clone (RP11-694H13) ends at 15.69, first and last triplicated clones (RP11-765H5 and RP11-33B5) start and end respectively at 15.63 and 16.84 Mb, the first and last duplicated clone (RP11-171H5 and RP11-97A20) seem to start and end respectively at 16.79 and 18.16 Mb, first and last clones of the second triplication (RP11-512B16 and RP11-96J2) seem to start and end respectively at 18.07 and 18.92 Mb, and the first clone that is again normal (RP11-744L24) starts at 18.86 Mb (Figure 12 A). These imbalances were confirmed by rtqPCR analyses using primer pairs CCR19_A, CCR19_C, CCR19_E, CCR19_F and CCR19_H (Figure 12 C).

FISH studies using either RP11-793A20 (starts at 16.2 Mb, triplicated), CTD-2278I10 (starts at 17.2 Mb, duplicated) and CTD-2017F19 (starts at 18.2 Mb, triplicated) on metaphase chromosome spreads from the patient showed one signal on each chromosome 19p13 (data not shown). FISH on interphase nuclei using these probes in the duplicated and triplicated regions showed the expected number of copies but did not allow a reliable deduction of the order of the different fragments.

The patient and her parents were typed for four STR markers located within the rearranged chromosomal fragment, to investigated the origin and mechanism of rearrangement (Table 6 and 11B). Marker D19S252 was not informative. Markers D19S899 and D19S429, located in a duplicated fragment according to the aCGH results, revealed also double dosage for the maternal allele. For marker D19S1037, a threefold higher dose was detected for one maternal allele compared to the paternal allele. This result places the proximal breakpoint of the second triplication at least 130 kb more proximal from the breakpoint predicted by the aCGH. These results also demonstrate that the aberration occurred on a single, maternal chromosome, and thus is the result of an intrachromosomal rearrangement.

All loci analyzed by qPCR were also analyzed in parental DNA samples (Figure 11C). Since all showed normal copy numbers, the CCR occurred *de novo*. This strongly suggests that the

aberration is causal for the phenotype of the proband. Also the size of the imbalance (3.26Mb) and the number of genes affected by the imbalance (102 Ensembl gene IDs in Ensembl release 46) support the causal nature. Small fractions of the aberrant region are described as variable in copy number in the Database of Genomic Variants[108], but none of these variable regions span a large part of the triplicated or duplicated region (all are smaller than 180kb).
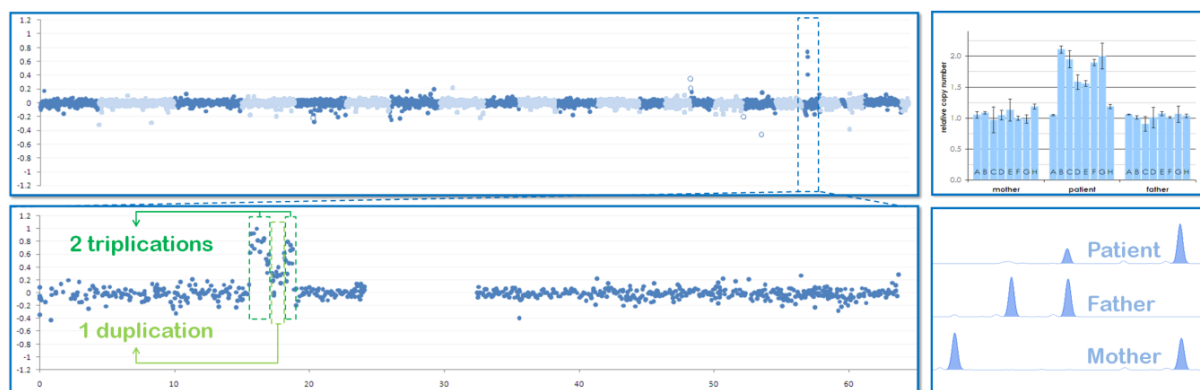

Figure 11: (A) Result of aCGH using a genome-wide micro-array with a 1Mb resolution, showing 3 consecutive clones on chromosome 19 with an increased copy-number. The y-axis shows log$_2$ratios of Cy5 vs Cy3 signal intensities, the patient is labeled in Cy5. Clones are ordered on the x-axis per chromosome (alternating colors) and based on their physical location. Empty dots indicate known polymorphic clones. (B) Result of aCGH using a chromosome 19 tiling-path BAC/PAC microarray, showing a region of triplication, duplication and triplication on chromosome 19p13.11. Clones are ordered on the X-axis based on their physical position on chromosome 19 [Mb]. (C) rtqPCR analyses in the patient and her parents confirm the presence of a *de novo* complex rearrangement showing different copy numbers. Primer sequences are listed in Table 5 as CCR19_. Combined results from at least 2 independent experiments are shown, error bars indicate 2 x SD[89]. (D) STR analysis of D19S1037 (located in the telomeric triplicated fragment) for the patient and her parents, showing that the imbalance arose on a single maternal chromosome.
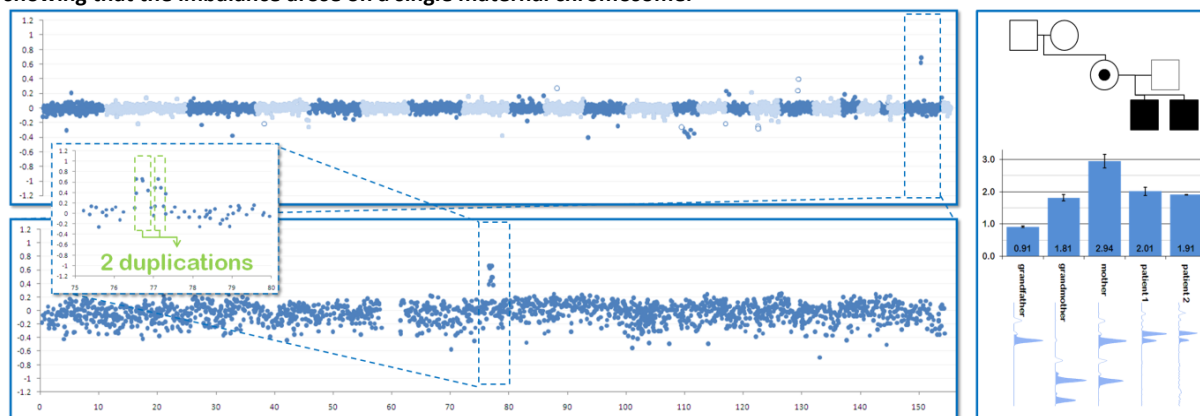
Figure 12: (A) Result of aCGH using a genome-wide micro-array with a 1Mb resolution, showing 2 consecutive clones on the X chromosome that indicate a duplication. The y-axis shows log$_2$ratios of Cy5 vs Cy3 signal intensities, the patient is labeled in Cy5. Clones are ordered on the x-axis per chromosome (alternating colors) and based on their physical location. Empty dots indicate known polymorphic clones. (B) Result of aCGH using a X chromosome tiling-path BAC/PAC microarray, showing two regions of duplication on chromosome X. The patient is labeled in Cy5. Clones are ordered on the X-axis based on their physical position on X chromosome. (C) Pedigree of family 1 (see main text and page 49), and results of rtqPCR analyses with the ATRX_i8-9 primer pair (Table 5) and DXS7498 STR length analysis (HUMARA primer pair designed in a region at 10 cMorgan from *ATRX*, Table 6) for the indicated family members, showing that the patient inherited the duplication from his mother, and that it arose *de novo* in her father's X chromosome.

4.4.3.4.   patient 16

Results of analyses in patient 16 and his sib have been described in results chapter 4.1.3 (page 51). Briefly, DNA from patient 16 was analyzed by aCGH at a genome wide 1Mb resolution, revealing 2 consecutive duplicated clones the X chromosome (Figure 12 A). DNA from his similarly affected sibling was analyzed by aCGH using a microarray constructed with a tiling BAC/PAC clones set[92], which showed the presence of two separate duplications (Figure 12 B). Length analysis of an STR located 10 cM from the duplications (DXS7498) showed they arose *de novo* in the maternal grandfather.

4.4.4. discussion

The presented results describe the frequent detection of complex intrachromosomal rearrangements. Out of the 16 causal constitutional imbalances we detected until now in a cohort of 99 patients with a syndromic CHD, 4 are CiCRs. Of note, an additional 2 CiCRs have been found in follow-up studies (unpublished results). Few other aCGH studies mention the detection of CiCRs[127]. Differences in phenotypic selection criteria are unlikely to explain this difference. It seems more plausible that some imbalances are - like in routine karyotyping - labelled as simple, while they are in fact more complex. Multiple reasons for misinterpretation could be envisaged:

- Some CiCRs are closely neighbouring imbalances. A single reporter suggesting a copy number different from the rest of the imbalance could be discarded as an irrelevant or bad measurement, or included in the imbalance by detection algorithms that use sliding windows[153].

- Large imbalances might divert the attention from small ones. Some CiCRs consist of large imbalance(s) and a small imbalance. Confirmation of the large imbalance could be viewed as sufficient, while the existence and *de novo* nature of the small imbalance is not further investigated[151].

- Out of the four CiCRs detected by our group until now, two were evident at a 1Mb resolution while two others (patients 13 and 16), became evident only upon higher-resolution analyses. Moreover, in both other cases additional complexity was revealed upon higher-resolution analysis. Since a large part of the large aCGH studies published until now use micro-arrays with a 1Mb resolution, it can be expected that some of the simple imbalances will appear to be CCRs upon higher-resolution analysis.

Many CCRs consist out of at least one duplication[58,125,127,153,157,207-209], and in *DMD*, complex duplications and triplications are found but no complex deletions[157]. Thus a closer examination of duplication imbalances for the presence of a CCR is recommended. Moreover, since duplications are less frequent than deletions, it is remarkable that duplications are found often in CiCR. It has been suggested that partial reversion of an unstable duplication to the normal situation might explain non-contiguous duplications[157]. The wide variety of types of CiCRs makes a single 'catastrophic' event causing multiple chromatin breaks that are resolved in chromatin clusters more likely[153,221]. Our findings moreover contrast with *de novo* translocations that are associated with microdeletions or duplications, which are described to be uniquely of paternal origin of: one out of four investigated CiCRs was maternal in origin, suggesting some CiCRs may arise by a mechanism different from that generating the translocations that are associated with cryptic imbalances[219].

The frequent identification of CiCRs is not only interesting from a mechanistic point of view, it also has important consequences for genotype-fenotype analyses. First, not all genes that appear to be imbalanced have by inference an abnormal copy number in the patient. In addition, faulty conclusions can be drawn on the functional consequences of an imbalance upon the genes they encompass: e.g. 1Mb analysis of a CiCR on Xq suggested a single duplication of at least the entire 5' UTR of the *ATRX* gene. Only higher-resolution analysis demonstrated the imbalance was more complex, with one duplication inside (and thus disrupting) the *ATRX* gene. Hence, the precise delineation of aberrations is essential for proper genotype-phenotype correlations.

5. implementation of high-resolution aCGH

5.1. rationale for screening at a higher resolution

Commercial platforms for molecular karyotyping offer an increasingly high resolution. Theoretically, aCGH using a 1Mb array allows for the detection of only a fraction of indels smaller than 1Mb (e.g. 17% of 150kb sized indels, or 50% of 0.5Mb sized indels) and is unable to detect indels smaller than 100kb. Arguably, an increasing resolution yields a proportional increase in number of causal imbalances detected. Two causal indels smaller than 1Mb were already detected using 1Mb aCGH. In one patient a 222–281 kb duplication was detected by a 1Mb clone located in *ATRX*[125]. In another patient 2 fosmids in *CBP* signaled the presence of a 345-480 kb duplication[130]. These fosmids are not part of the normal 1Mb clone set but of a clone set that was designed to detect small indels in known dosage-sensitive regions. Other groups have reported that indels smaller than 1Mb can be detected in syndromic patients[193]. These findings strengthened our suspicion that in a significant portion of patients that did not show an indel upon 1Mb aCGH, an indel smaller than 1Mb could be detected using higher-resolution arrays. To test if and to what extent indels smaller than 1Mb cause syndromic CHDs, we set out to reanalyze patients that were negative on 1Mb aCGH, now by CGH on high-resolution oligo-array slides.

5.2. platform selection

We evaluated two commercial oligonucleotide aCGH platforms that offer a resolution higher than 1Mb: the Nimblegen 385K array that contains 385 000 probes spread throughout the genome, and the Agilent 244k array that contains 244 000 probes spread throughout the genome. Despite the higher probe density of the Nimblegen array, the resolution was poorer than that of Agilent: due to a poor response (the deviation from the normal caused by an imbalances was less than half of what can be expected theoretically) and large standard deviation, reliable detection of deletions requires at least ten consecutive probes of the Nimblegen 385K array to indicate a deletion. Due to the better response and smaller standard deviation, one could rely on the signal generated by only two probes of the Agilent array to detect a deletion at an acceptable false-positive cost (<1 per experiment). The practical resolution of the Nimblegen 385K array is therefore about 80kb, while that of

the Agilent 244K array is around 20kb. We therefore choose to use the Agilent 244K platform.
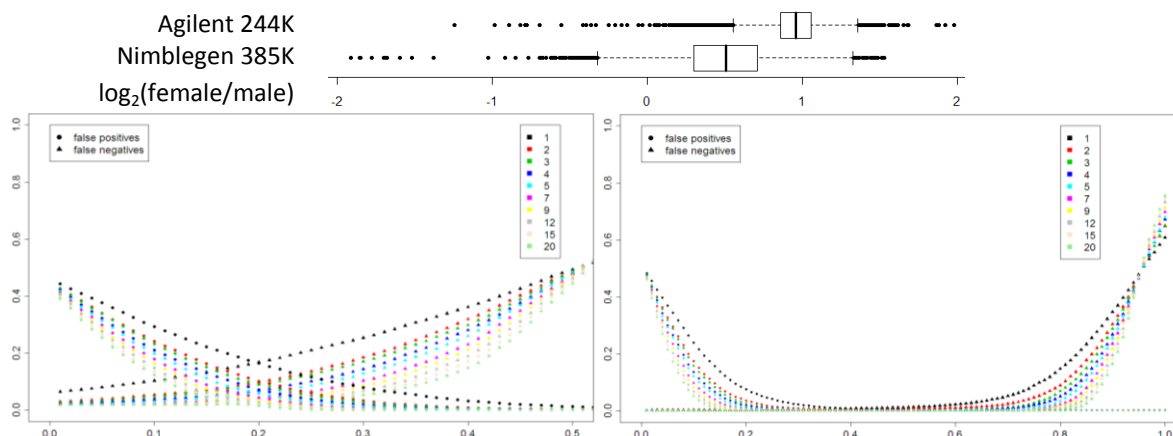
Figure 13: Comparison of Nimblegen 385K and Agilent 244K. A. Box plots of the distribution of log$_2$(ratio of the signals of labeled DNA from a female versus a male) values for clones on the X chromosome. Representative experiments are displayed. **B.** Frequency (y-axis) of false positive (•) and false negative (Δ) values when applying a varying median of sliding window threshold (x-axis) calling algorithm of a size as indicated by the colors, showing a resolution for deletion calling that is superior in terms of false positive values and false negative values for the Agilent 244K platform (right) in comparison to the Nimblegen 285K platform (left). For example, a false positive and false negative frequency of 5% can be reached by using a single Agilent probe. For Nimblegen, 7 consecutive probes are needed.

5.3. data processing

An Agilent 244K microarray contains over 60 times the number of probes of a 1Mb microarray. Simple analysis schemes based on copying and pasting data files in Excel are not suitable to handle and process this amount of data. We therefore designed a framework to manage and process the data that was generated using 244K arrays. A workflow of the strategy is shown in Figure 14.
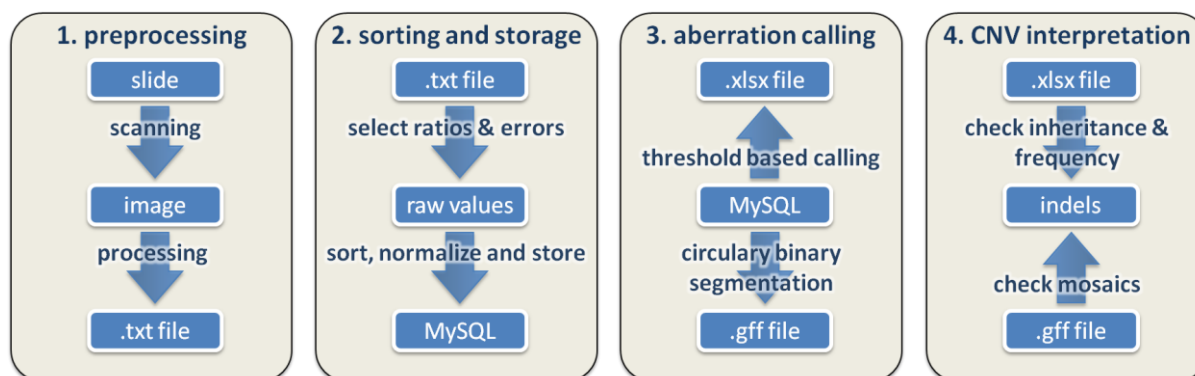
Figure 14: workflow depicting the different step used in Agilent data processing. More details on the exact procedures are described in the main text.

5.3.1. preprocessing

The preprocessing step consists of scanning the washed slide to generate a .tiff image file, and analyzing the image file using Feature Extraction (version 9.5.3.1, Agilent) to generate a .txt file containing the 2D Lowess normalized raw Log$_{10}$ ratios of each feature and

the associated errors (generated by a method undisclosed by the software provider) as well as other values and information that are not taken into account.

5.3.2. sorting, normalization and storage

Data management and processing is executed by scripts we designed in the R language and statistical environment (R version 2.8.0). R package RMySQL (V0.6) was used to extract the raw Log10 ratios and LogRatioErrors and store them in a separate SQL table for each experiment. From the MySQL database, the data is in a second step retrieved using RMySQL. Log_{10} ratios are transformed to the more customary Log_2 ratios. These are discarded if both their absolute value does not exceed 1 and their associated LogRatio error exceeds 0.2. LogRatios are sorted based on their annotated position in human genome (hg) build 17. Values originating from the ~1000 probes spotted in triplicate or duplicate are replaced with respectively the median of all 3 or mean of all 2. Genome coordinates are subsequently transformed to the current genome build (hg18). These data are subsequently normalized by a linear modeling to the GC content of the environment of the targeted regions. This environment is captured in a table containing the inverse of the GC content (1/GC%) of the probe itself and of 16 windows around the region targeted by the probe that increase in size from 100 bp to 60kb, extracted by the EMBOSS geecee tool implemented in Galaxy (main.g2.bx.psu.edu). Correlations of Log_2Ratio to GC content are thus removed (Figure 15).

A linear modeling is calculated of the resulting values and of those generated in 2 self-to-self experiments, mainly to remove the bias that is associated with the differences in dye used for labeling (Figure 15). The resulting Log Ratios are stored in a single table in MySQL to enable easy retrieval of relevant information and to minimized storage requirements. Standard deviation of probes from the autosomes and chromosome X separation are calculated and stored in a separate .txt file as quality control measures (Figure 14). aCGH experiments yielding a standard deviation of the normalized Log_2 ratios of the autosomal probes exceeding 0.20 were repeated.
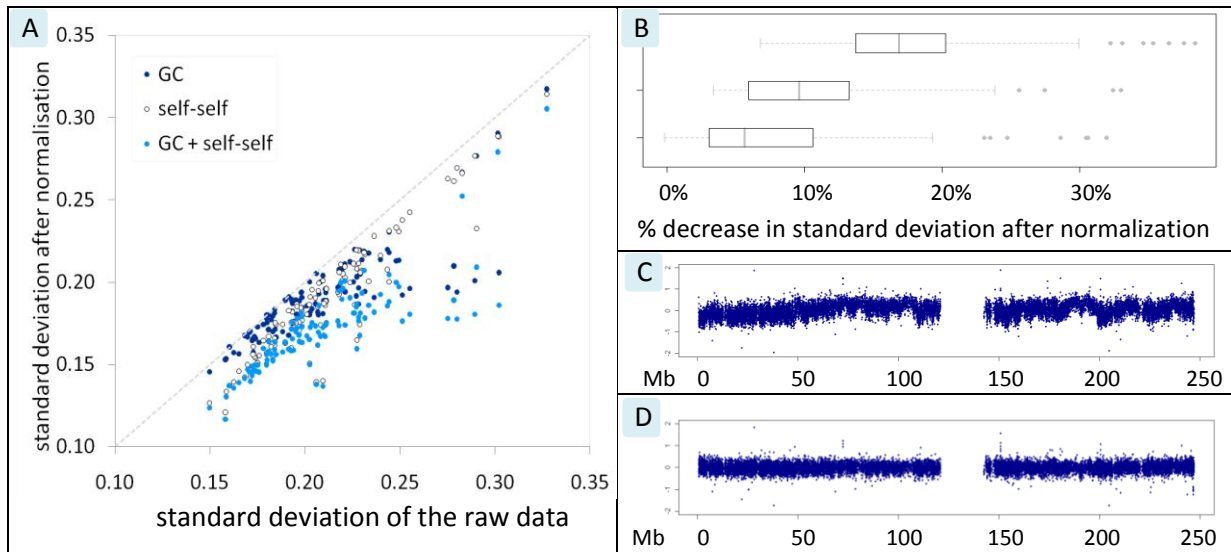
Figure 15: effect of normalization on results of aCGH experiments. A. Effect of different normalizations on the standard deviation of the raw data. Dotted line indicates no effect on standard deviation. GC-based normalizations have the largest impact on raw data with high standard deviations, while self-self normalizations have a similar effect in most experiments. **B.** Box plots of the relative decrease in standard deviation after normalization. Lower box plot: GC normalization; middle box plot: self-self normalization; upper box plot: GC and self-self normalization. **C. & D.** log2 ratio of signal intensities (Y axis) plotted versus physical position on chromosome 1 (X axis) before (D.) and after (E.) GC-based normalization of the results of an aCGH experiment showing extreme GC skewing. For details on normalization strategies, see main text.

5.3.3. aberration calling

Aberration calling is done by two methods: a threshold-based and a circular binary segmentation (CBS) method.

The threshold-based method was developed in the framework of this project. The frequency of false-positive results was estimated for various combined thresholds and standard deviations based on two self-self experiments (Figure 16). As there is no biological difference in copy number in this experiment, the result allows us to assess the false-positive rates associated with our algorithm. False positives resulting from the following conditions were tested: 2, 3 and 4 consecutive probes that each do not exceed a certain value (threshold maximum) and whose average or median does not exceed a certain value (threshold average). Based on these results, we defined thresholds that reduce the theoretical average number of false-positive results per experiment below 1 (Figure 17, Table 13).
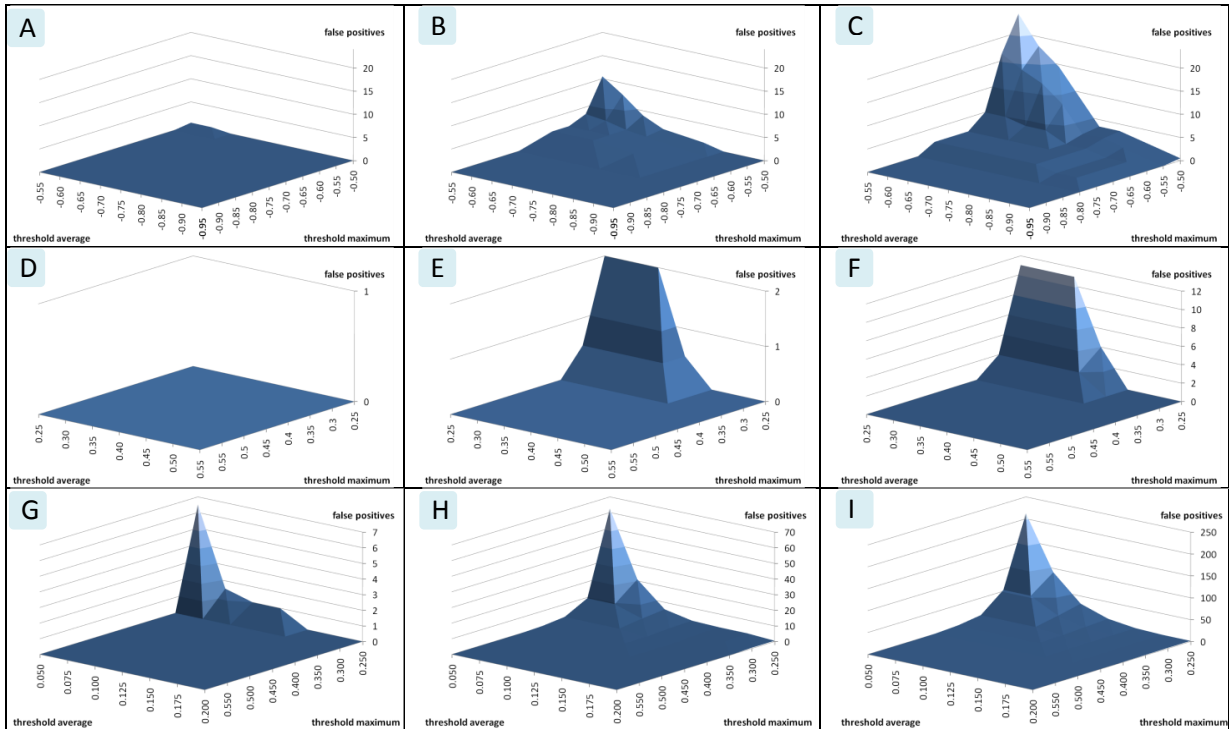
Figure 16: effect of threshold maximum, threshold average and standard deviation on the number of false positives, as assessed by 2 self-self experiments (A., D., G.). The standard deviation was altered by multiplying the $\log_2$ values by the appropriate number to obtain a standard deviation of 0.17 (B., E., H.) or 0.20 (C., F., I.). The threshold applies to 2 consecutive values for a deletion (A., B, C.) and 3 (D., E., F.) or 4 (G., H., I.) consecutive values for a duplication.
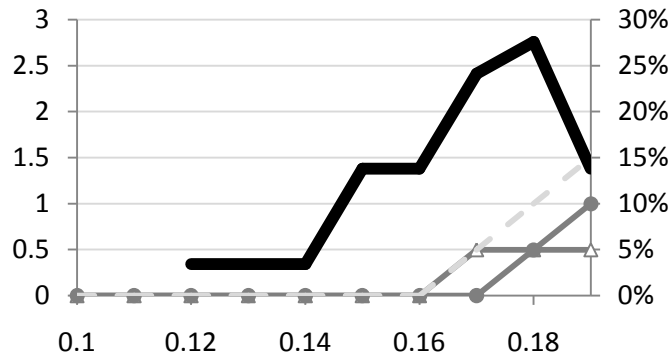

Figure 17: Estimated number of false positives (left Y-axis) based on 2 self-self hybridization experiments, with varying standard deviations (X-axis). Dots (•) indicate false positive duplications and triangles (Δ) deletions. The light-grey dotted line indicates the total number of false positives. The thick black line illustrates the distribution of the standard deviation of the experiments. Thresholds used for false positive estimation as in Table 13.

As a positive control experiment, two differentially labeled DNA samples were hybridized on an Agilent 244K array: one from a 46,XY male and one from an 46,XX female that carries a large duplication on chromosome 10p (size 11.7Mb). Chromosome X and chromosome 10q26 are present in a 2/1 and 3/2 ratio in this experiment. This allows us to assess the false-negative results associated with deletion and duplication calling (Table 13). Using both experiments, we set a number of criteria that allow aberration calling with a minimal number of false positive and false negative results. These criteria defined minimal or maximal median values of a number of consecutive clones, and minimal or maximal values

for a number of consecutive clones (minimal and maximal respectively for duplications and deletions), and their application results in a small number (<5%) of false negatives when applied (Table 13). The reported false negatives moreover are likely underestimates, as they are derived from biological material that undoubtedly contains CNVs and low copy repeats. For example, the false-negatives are based on the deletion of chromosome X, but the region between 89 and 90Mb always fails to indicate a deletion, as this region is paralogous to a region on chromosome Y.

Table 13: criteria for aberration calling using Agilent 244K arrays. Minimal and maximal (respectively for duplications and deletions) values for a number of consecutive clones and for their median, and the number of false negatives as estimated by a large duplication on 10p and a deletion of chromosome X.

aberration	consecutive clones	median of all values exceeds	all values exceed	false negatives
deletion	2	-0.8	-0.7	12.1%
deletion	3	-0.7	-0.5	4.3%
duplication	3	0.5	0.4	22.8%
duplication	4	0.4	0.35	3.1%

The CBS method was described by Venkatraman and Olshen[222] and was implemented in R using the following settings: α=0.002, 5000 permutations. We used the described hybrid method for p-value calculations. Splits were undone if they differed less than the 0.1 times the standard deviation (calculated after a 0.05 trim of the input data). The goal of these analyses was to identify mosaic imbalances as these will not always be detected by the threshold-based method.

5.3.4. CNV interpretation

All CNVs that were called by our algorithm were compared to known copy number polymorphisms (CNPs) of which we assume that they are not a major risk factor for developmental problems. This list was compiled from a list of common CNVs called in our own experiments and the list of the CNVs described by de Smith *et al.*[223], detected in a group of 50 normal males by use of an Agilent 185K platform, an aCGH platform similar to Agilent 244K. From the latter group we considered CNVs as common if they were detected more than twice in this population. All other CNVs we detected were compared to the data compiled in the database of genomic variants[108]. We verified if the region was listed there as copy number variable in a similar copy number state ("loss" or "gain"), and whether multiple independent studies reported it. If a CNV was described in 5 or less individuals, its presence was recorded in the patient overview. For each of these CNVs, inheritance was verified. Rare CNVs were included to avoid discarding CNVs that represent a mutation with a recessive

effect; several examples from our and other groups have already shown this can cause recessive disorders[224-226]. For each of the detected CNVs we also examined whether any of the affected genes are annotated to be associated with genetic disorders in the online mendelian inheritance in man® (OMIM) database, to avoid disregarding known genetic causes that are unjustly called as a benign CNV. We moreover examined whether any of the affected genes are described as being imprinted in a genome wide study.

6.   application of Agilent 244K aCGH to idiopathic syndromic CHD patients

6.1. study design

We reanalyzed in total 29 patients where no causal indel was detected using 1Mb aCGH. This group included 6 patients in whom an unclassified variant was identified by 1Mb aCGH. In contrast to the 1Mb aCGH which lead to the detection of indels in 32/99 patients, we expected to detect many CNVs in every patient at this resolution. Since each unknown CNV can in theory be a cause for the patients phenotype, each requires laborious investigations: its presence needs to be confirmed and its inheritance needs to be investigated by checking parental samples.

In an attempt to accommodate this foreseeable issue, we chose to analyze the patient and his parents together. Such an experimental design enables us to immediately assess the inheritance pattern of each aberration detected in a patient and to assess the frequency of false-positive results when using the aforementioned criteria we established for aberration calling.

6.2. study results

Each patient carried on average 49 CNVs in comparison to the reference DNA (range: 28-94). Almost all aberrations affect regions that are known to be CNVs and/or inherited from one or both of the parents. In all but two patients, we detected one or more previously unreported CNVs. These were 41 deletions and 34 duplications. 16 deletions affected only intergenic regions, and a further 8 were intronic. 7 duplications affected intergenic regions and 10 affected only the 5' or 3' end of a gene. Thus about 50% of duplications and 60% of deletions probably do not affect transcript integrity. The median maximal size of these aberrations was 63kb (Figure 18). There were no inherited heterozygous or homozygous disruptions of genes that are known to cause a recessive disorder that did explain part of the patients' phenotype, nor of known imprinted genes. Three of the 75 unknown CNVs were not inherited from a parent, and are further discussed below.
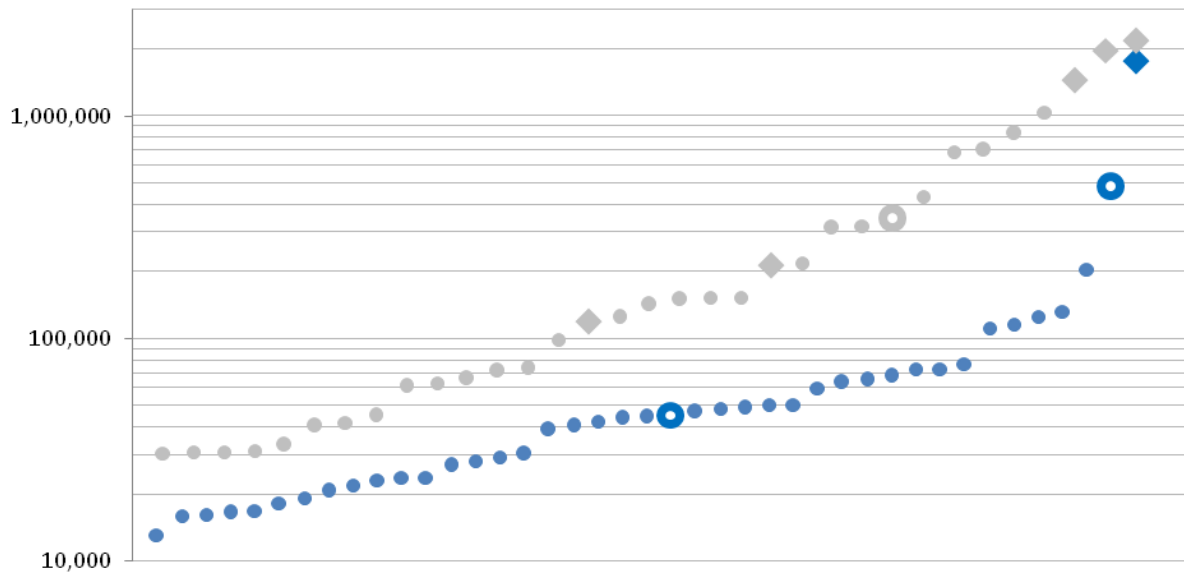
Figure 18: maximal sizes of previously unreported CNVs detected by aCGH on Agilent 244K arrays. CNVs are sorted on the X-axis by their maximal size, which is shown on the Y-axis (bp, $\log_{10}$ scale). 41 deletions are indicated in dark grey, 34 duplications in grey. Empty dots indicate *de novo* aberrations and diamonds aberrations already detected by aCGH on 1Mb arrays. This graph shows that most aberrations larger than 1Mb were already detected on 1Mb arrays, and that unreported duplications are in general larger than deletions. This difference can be partly explained by the lower detection threshold for deletions than for duplications. Duplications larger than 100kb are much more frequent than deletions (19 vs 7).

One *de novo* duplication affects a gene desert on 2p16.1 that is partially known to be copy-number variable. This boy was conceived by intracytoplasmatic sperm injection (ICSI) with a sperm from his biological father. Whether there is a link between the duplication event and the assisted reproduction by ICSI is unknown, although the increased occurrence of structural chromosome aberrations after ICSI is suggestive for a causal link[227]. Given the copy number variable nature of this region and the absence of annotated genes, chances are estimated to be small that this duplication is causal.

A second *de novo* deletion encompasses a region on 17q21.31. FISH studies confirm deletion of this region, which has recurrently been found to be deleted due to the presence of flanking LCRs. Deletions of this region have been found in multiple patients with a mental handicap and a variety of defects by three independent groups, which led to the identification of a novel so-called microdeletion syndrome[181,228,229]. Cardinal features of this microdeletion syndrome like developmental delay and facial dysmorphism were also found in the present patient, and other congenital malformations of this patient (a heart defects and agenesis of the corpus callosum) have also been described (in respectively 6 and 2 out of 22 patients)[230]. The deletion is therefore causal for the phenotype of this patient.

A third *de novo* deletion affects a region of maximally 45kb on 6q25.3 that encodes *FOXC1*. This deletion, confirmed by breakpoint-spanning PCR, was detected in a patient with bilateral congenital glaucoma and partial aniridia, club feet and an ASD. Deletions and loss-of-function mutations of *FOXC1* have been shown to cause eye malformations[231-233] and are occasionally also associated with heart defects[232,234]. We therefore classify this deletion as causal for the phenotype of this patient. The identification of this deletion again demonstrates the importance of obtaining an etiological diagnosis in these patients: mutations in *FOXC1* are rarely associated with developmental delay and were found to segregate through consecutive generations in multiple large families, permitting a favorable diagnosis regarding mental development of this newborn girl to her parents. This finding also shows that the chosen platform can detect causal indels below 1Mb and below 100kb resolution.

6.3. conclusions

aCGH using Agilent 244K arrays provides an increased resolution for the detection of cryptic chromosome imbalances compared to using 1Mb arrays. As discussed before for the 1Mb array, an increased resolution complicates the interpretation of the causality of the detected imbalances. The number of undescribed variants (73) detected in the 29 patients that were analyzed by far surpasses the number of undescribed variants detected in our 1Mb study (In 99 patients, 7 hitherto undescribed CNVs were detected). Despite the multitude of studies investigating the presence of CNVs in the human genome, a significant fraction of CNVs most likely still remains undescribed. This will be one of the major hurdles in applying aCGH at a resolution below 100kb. Availability of DNA samples and phenotypic data from the parents is essential for causality evaluation, and in theory further studies are needed before any undescribed CNV can be classified as benign or pathogenic. International collaborative databases that combine the generated genotypic data with the corresponding phenotypic data (such as DECIPHER, decipher.sanger.ac.uk) are essential to be able to identify which rare CNVs predispose to rare or incompletely penetrant genetic disorders.

CHAPTER 5   GENE SELECTION USING A CARDIAC ENDEAVOUR

1.   a need for gene prioritization

One of the goals of our aCGH studies was the identification of novel regions for CHDs through the detection of chromosomal imbalances in CHD patients. These imbalances pinpoint genomic intervals that contain one or more genes responsible for CHDs, thus pointing to novel candidate genes for human cardiac development. Indeed, in 8 of the 18 causal indels detected, a gene previously known to be involved in CHDs was affected. These genes are *NKX2.5*, *CBP*, *NSD1*, *FOXC1*, *ATRX*, *NOTCH1*, *EHMT1*, *FBN2* and *TBX1*. This finding demonstrates that indel detection by aCGH can indeed serve as a positional cloning strategy. In the remaining 10 patients, we assume that the imbalance is causal for the CHD and thus affects a novel gene involved in CHDs. However, even though these imbalances are much smaller than the cytogenetically detectable imbalances traditionally used to map disease genes, these regions still contain too many genes to allow direct functional studies on each of them. In order to select the best candidate genes in an unbiased way (i.e. without relying on preferences and knowledge of a single researcher or research group, but solely on most of the available knowledge) we used the ENDEAVOUR tool to prioritize candidate genes.

2.   adapting ENDEAVOUR to study CHDs: a CARDIAC ENDEAVOUR

2.1. introduction

In an attempt to identify the best candidate genes from these regions, we applied and improved an existing algorithm to prioritize candidate genes for CHDs based on sets of known genes. This existing algorithm, termed ENDEAVOUR, was recently developed by the group of Prof. Moreau (ESAT, K.U.Leuven, Belgium)[235]. The ENDEAVOUR algorithm uses a multitude of data sources to model a biological process of interest by profiling the genes known to be involved in that process, and compares these training gene profiles to candidate gene profiles in order to rank the latter (Figure 19). The regular ENDEAVOUR algorithm was further adapted to obtain better prioritizations and to accommodate the specificities and problems inherent to this project. This work was done in close collaboration

with the ESAT department (L.-C. Tranchevent from the group of Prof. Y. Moreau) and with

Dr. P. Van Loo (Department of Human Genetics, K.U.Leuven, Belgium)).
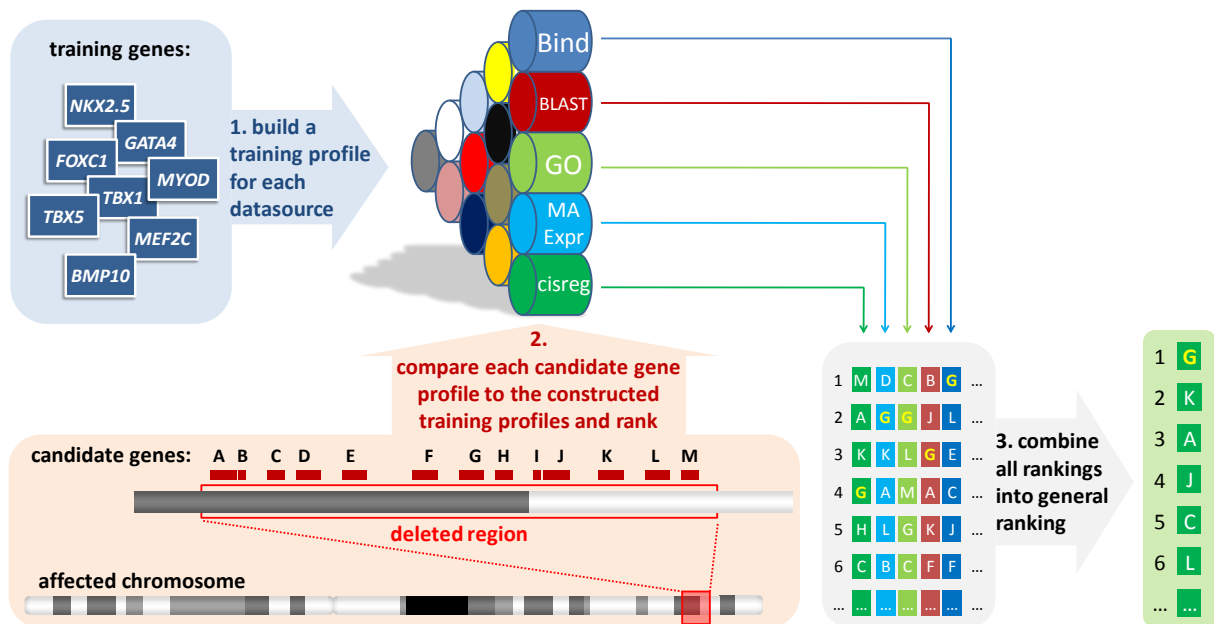
Figure 19: the regular ENDEAVOUR algorithm. In a first step, training genes that are involved in a process of interest (e.g. heart field specification) are profiled for a variety of data sources (indicated by the colored cylinders, only a subset of the available data sources are displayed in this image). Next, for each data source, candidate genes (e.g. genes from a candidate region identified by an aCGH-detected deletion) are ranked according to the way their profiles match the training gene profiles (only a subset of the rankings is displayed in this image). These rankings are subsequently merged to obtain a general ranking. In the example, gene "G" ranks first.

2.2. defining and using different training sets

Heart development is not a discrete developmental process, but depends on the combination of different discrete or interlinked developmental and cellular processes. Obtaining a specific signal for training genes involved in any aspect of heart development is not trivial. We therefore followed a different strategy, where we defined several different sets of training genes homogeneous with regards to their role in cardiac development.

Cells are added to the heart at different time points or originate from different regions in the embryo: cells from the first heart field, the second heart field and the neural crest all contribute to the adult heart. Genes involved in these processes are in training sets 1, 2 and 3 respectively. Moreover, defects in left-right axis determination (arising before heart development), in vascularisation and in heart valve formation have been shown to potentially cause CHDs (training sets 4-6). Disruption of genes involved in the contractile apparatus may also cause CHDs (besides cardiomyopathies) (training set 8). Finally, some genes that are associated with CHDs are involved in more general cellular processes (e.g. chromatin remodeling or cell proliferation) that are not directly tied to heart development.

All known genes causing CHD upon haploinsufficiency were selected and added as an additional training set (training set 7) to enrich our results for dosage sensitive genes implicated in human cardiac defects and to enable scoring of developmental processes that are not explicitly known to be associated to CHDs. By using these different sets representing discrete processes underlying heart development, we assumed to be able to obtain more specific signals for each of these varying processes. We included a combined set (training set 9) containing all genes from the aforementioned training sets, to accommodate detection of genes involved in the potential cross-talk between the each of the developmental processes. Also between the other sets there was some redundancy. Not all training sets were used for each candidate set. The rules we applied for selecting trainings sets are described in chapter 3 (page 92).

As in the regular Endeavour algorithm, prioritization using a number of different training sets generates a number of ranked lists of candidate genes. These rankings were combined using order statistics like the rankings obtained from the use of different data sources in the regular ENDEAVOUR algorithm[235] (visualized in step 3 in Figure 19). Sets are available at homes.esat.kuleuven.be/~bioiuser/chdwiki/index.php/Prioritization_datasets.

Table 14: Endeavour training sets, the number of genes and the AUC of the LOOCVs' ROC. Note that the AUC of _Combination_ and of _Causing CHD upon haplo-insufficiency_ (heterogeneous sets of genes) is much worse than that of most other (homogeneous) training sets

#	Training set name	# genes	AUC
1	First heart field	12	0.9791
2	Second heart field	12	0.9600
3	Neural crest	42	0.9764
4	Left-right axis determination	11	0.9018
5	Vascularisation	20	0.9865
6	Heart valve formation	14	0.9350
7	Causing CHD upon haploinsufficiency	33	0.8824
8	Cardiomyopathies	18	0.9666
9	Combination	157	0.9046

2.3. addition, removal and combination of data sources

2.3.1. selection of informative data sources

Whether or not a data source contains information on the process of interest can be assessed by a leave-one-out cross-validation (LOOCV). In such a test, each individual gene from a gene set of interest is removed once from the training set and added to a set of 99 randomly chosen genes. These 100 candidate genes are then ranked using the data source profile of the remaining genes in the training set. The rank of the left-out training gene among the 99 random genes reflects the information that is present in the data source on

the process of interest. LOOCVs can generate Receiver Operator Characteristic (ROC) curves that plot false positive versus true positive rates. The area under this ROC curve (AUC) is the measure that we used to assess the ability of the data source to efficiently rank genes according to the profile of the remaining training genes. An AUC around 0.5 is equivalent to random ranking of candidate genes, while an AUC verging upon 1 is equivalent to first-place rankings for most training set genes. We arbitrarily set a ROC AUC threshold of 0.6 as the alternative approach (testing whether each data source significantly improves the overall AUC) would bias towards data sources containing *a priori* knowledge. Data sources scoring below this threshold do not contain (enough) information (Figure 20). As these in theory only add noise to the predictions, we omit these data sources from our analyses to further improve the prioritizations. Indeed, upon combination training set 9 for example generated an AUC of 0.9451 instead of 0.9334.

Figure 20: performance of the different data sources upon LOOCV, used to decide which data sources were included in the prioritization. Performance of the data source in prioritizations depends on the training set, for example: the chdma (microarray expression data for heart development) is highly informative for first heart field and the neural crest cells, but performs poor for the other training sets. Similarly, Gene Ontology does not contain information to significantly prioritize genes involved in valve formation, while it does contribute to prioritizations according to the other training sets. Information on the standard data sources can be found on the Endeavour website

2.3.2. microarray expression in the murine heart

A second adaptation of the ENDEAVOUR algorithm was the addition of data sources. Data sources are the core of the ENDEAVOUR algorithm, and the availability of multiple independent and informative data sources is essential for optimal functioning. The first extra data source

was a set of genome wide expression analyses using microarrays that we retrieved from the gene expression omnibus (GEO, www.ncbi.nlm.nih.gov/geo) under accession number GSE1479. In this experiment, RNA expression profiles were obtained from the cardiac regions of mice sacrificed at specific developmental stages, to obtain a time-course of gene expression during murine heart development. Samples were taken at daily intervals from embryonic day (E) 10.5 to E14.5 and at E16.4 and E18.5. From E11.5, atrial and ventricular chambers were separated. We combined this data set with a set of random experiments from GEO done on the same platform (Affymetrix GeneChip Mouse Genome 430 2.0 Array) to obtain background values, and normalized these data using the robust multi-array average (RMA) method[236]. Multiple probe measurements per gene were combined into a single measurement by averaging them and data was transposed from murine to human genes by the orthology maps available in BioMart (www.ensembl.org/biomart). Quality of this expression data set was assessed by a LOOCV. This resulted in an AUC of 0.97 for the first heart field training set and of 0.82 for the neural crest cells training set, showing that this information can indeed be used. Comparison with regular microarray expression data sources demonstrates they are outperformed by this tailored data source for these training sets (resp. an average AUC of 0.75 and 0.56). As shown in Figure 20, for some other training sets this data source was uninformative. This is not surprising, as some biological processes that are modeled in the training sets (for example left-right asymmetry establishment, AUC = 0.42) were not investigated in these expression profiling experiments.

2.3.3. homology

A second added data source contained estimates of homology of human proteins to their orthologs in other species at different phylogenetic distances that are sequenced. Information on gene conservation is assumed to be related to conservation or changes in gene function. For example, gene duplication events and the subsequent changes in gene expression and/or protein structure have been linked to the evolutionary variations in heart structure and function[237]. Homology data was extracted from the HomoloGene database (www.ncbi.nlm.nih.gov/homologene), BioMart (www.ensembl.org/biomart) and the Inparanoid tool (inparanoid.sbc.su.se). Similar to expression data, this data was summarized using a vector representation and scoring was done using Pearson correlation. Missing data

was handled like for every other data source: it is not taken into account in the prioritization since the order statistics are based on rank ratios

Quality of the data as a data source for ENDEAVOUR was assessed by LOOCV and resulted in maximal AUCs of 0.77 for the first heart field training set using the Inparanoid data and 0.79 for the neural crest training set using the Biomart data. Conservation is more informative for first than for second heart field genes (an evolutionary more recent innovation in development), and – surprisingly - *left-right asymmetry establishment* does not generate a significant AUC.

2.3.4. combination of similar data sources

One of the issues of the existing ENDEAVOUR algorithm is the artificial inflation of signals generated by using similar data sources. When two data sources contain similar information, they will generate highly similar predictions that will have more impact on the final result than a single data source that is independent. To circumvent this problem, one can either remove one of the two similar data sources or combine prioritizations from data sources before they are combined with the other data sources. We choose the latter option, since the non-overlapping information in each of the data sources will still be taken into account when using this strategy. Prioritizations based on data sources that are significantly correlated and assess the same information are first combined separately before being combined with the other data sources of the same class, after which all are combined in an overall prioritization (Figure 21). Data prioritization using the CARDIAC ENDEAVOUR is thus done by a tree-like structure for the data sources (Figure 22). As discussed, only informative data sources are incorporated.
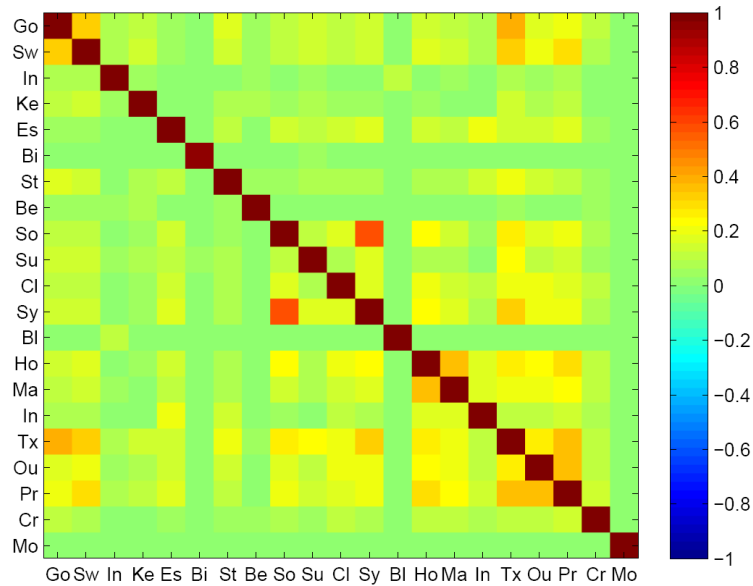
Figure 21: Example of Spearman rank correlations between data sources upon prioritization (here using the vascularisation training set). The Blast (Bl) data source is not strongly correlated with any other source, while the Textmining (Te) and the Gene Ontology (Go) data source, or both micro-array expression analysis data sources for adult human tissues (Sy and So) are strongly correlated. Obviously, all data sources are perfectly correlated with themselves.
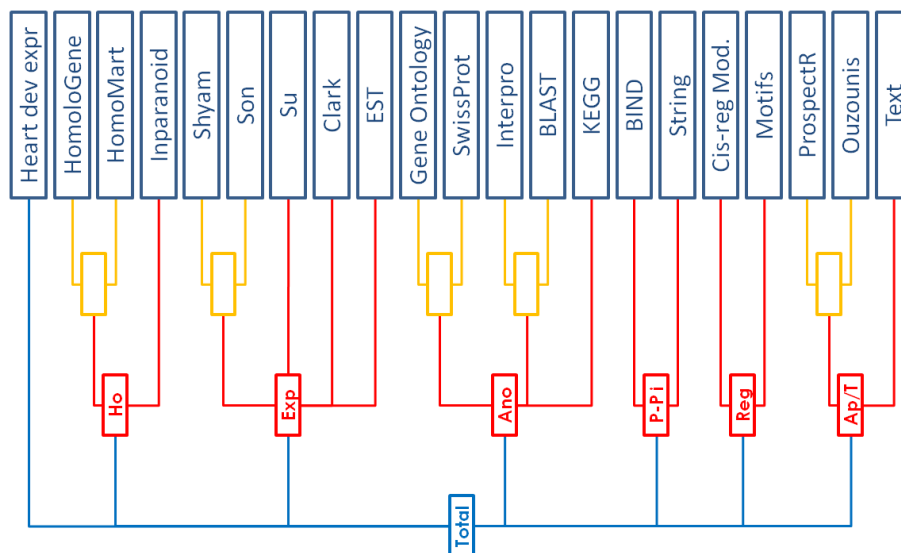
Figure 22: example of a tree-like structure used for prioritizations based on an individual training set. All data sources are displayed. As discussed in the main text, the use of a data source depends on the results of LOOCV.

2.3.5. incorporation of the CARDIAC ENDEAVOUR in CHDWiki

The gene prioritization strategy outlined above was incorporated into a Wiki that was developed in a collaboration between groups from the Center for Human Genetics and the ESAT department (K.U.Leuven). This Wiki (termed *CHDWiki*) is conceptualized as a portal for studying the genetics of CHDs. Apart from a very complete database on known genetic causes of CHDs, more advanced data analysis tools like protein-protein interaction database querying and the CARDIAC ENDEAVOUR algorithm are thus made available to the larger community studying cardiac development and CHDs. The speed of prioritizations is improved

because the predefined training sets enable querying the end result of any genome-wide prioritization already executed rather than needing to restart the algorithm each time based on different training sets. A manuscript describing the development of this portal is currently under review. This portal is available from homes.esat.kuleuven.be/~bioiuser/chdwiki.

3. applying the CARDIAC ENDEAVOUR algorithm to selected indels

3.1. a tailored use of training sets, indels and CARDIAC ENDEAVOUR

Not all training sets were used for each region containing candidate genes: sets 1 →3 and 5→7 were always used. When multiple patients were already described in the literature with a similar imbalance but without problems in the heterotaxia spectrum, we assumed there was no defect in left-right axis determination and omitted this set (del22q12.2 and delXp22). Likewise, when neither the patient under study nor other patients in the literature were affected by major valve malformations or a cardiomyopathy, the corresponding sets were not used. The cardiomyopathy set was thus only used for the 1p36.3 region, the valve development set only for the 6q25.1 region.

For each patient, we analysed whether a known CHD gene was affected by the imbalance, thus explaining the CHD observed in the patient. If so, the corresponding region was added to the group that served as a positive control. Otherwise, it was added to the list for further studies. This list does not correspond to the imbalances generated by the aCGH studies described in the previous chapter: part of the aCGH studies were done in parallel or even after the prioritisation and expression analysis studies presented in this and the following chapters. We moreover did not deem all detected imbalances suitable for further analysis. If an imbalance affected more than 150 genes or if genes were present in different abnormal copy number states (e.g. in unbalanced translocations), we choose not to continue the analysis. Two additional imbalance were detected by our group during studies performed outside the presented research[238]. Although they do not feature in Table 9, they were added to the list of candidate regions to be investigated (Table 15 and Table 16).

Finally, the format of the CARDIAC ENDEAVOUR we used for gene prioritisation does not correspond entirely to the format described above: while all adaptations we discussed are available in CHDWiki, not all modifications were available when genes were selected for further studies. Furthermore, many of the public data sources used by ENDEAVOUR are

constantly updated. The genes we further analysed were selected from the prioritizations available when we started our analysis in August 2006, when the modifications described in chapters 2.3.3 and 2.3.4 (page 89) were not yet available .

3.2. selection of genes from the prioritized lists of candidate genes

Genes affected by imbalances that include a known CHD gene were prioritized as a positive control. For each imbalance, a known CHD genes ranked on the first place. In 9q34.3, *NOTCH1* ranked first and *EHMT1* ranked fifth out of 106 genes. To identify novel CHD genes, genes from regions imbalanced in patients without a known CHD gene were prioritized by ENDEAVOUR using each individual training set. The results from each training set were combined in an overall prioritization by applying the same order statistics used in ENDEAVOUR. Obviously, obtaining a ranked list of candidate genes does not equal having a list of candidates, and we had to set criteria to select genes for further analysis:

1. From each candidate list, the 2 highest ranking genes from the overall prioritization were selected.

2. If the estimated penetrance of CHDs in similar aberrations was high (>25% or >50%), respectively the 3rd or 3rd and 4th ranking genes from the overall prioritization were selected.

3. If the p-value for the highest ranking gene in any prioritization using an individual training sets was higher than the median p-value of the LOOCV of this training set, this first-ranked gene was also selected.

4. Genes were removed from the candidate list if they were shown not to cause CHDs upon mutation in at least 2 independent reports (these include *CHEK2*, *RS1* and *CDKL5*).

In this way, we obtained a list of 22 genes for further analysis (listed in the next chapter in Table 16).

CHAPTER 6   VALIDATION OF ENDEAVOUR RESULTS BY EXPRESSION ANALYSIS IN DEVELOPING ZEBRAFISH EMBRYOS

1. rationale for gene expression analysis in zebrafish

In search of a high-throughput method to investigate the involvement of genes in cardiac development *in vivo*, we hypothesized that these can be distinguished from other genes by their expression pattern in the developing embryo, more specifically in the developing heart. The use of zebrafish as a model organism enables a fast and straightforward analysis of gene expression in multiple early developmental stages. Because of the high fecundity and the transparency of the developing embryo's, gene expression can moreover be evaluated in a high-throughput manner. Moreover, and in contrast to the fruitfly and the nematode, zebrafish are vertebrae and have a heart that structurally is closer to the human heart. Many of their gene functions are conserved with humans. Indeed, genes that require correct dosage in human (cardiac) development mostly appear to have a conserved function in zebrafish (cardiac) development.

2. design of gene expression analysis experiments

mRNA expression was analyzed by whole-mount *in situ* hybridization of DIG-labeled RNA probes on zebrafish embryos fixed at distinct developmental stages. We analyzed all the orthologues of the selected genes (27 genes unknown and 9 genes known to be involved in CHDs) in three to six different stages of development that are key to cardiac development:

12hpf	heart cell specification
18hpf	initiation of migration to the midline
22hpf	arrival at the midline, heart cone stage
30hpf	heart elongation, looping
36hpf	initiation of chamber specification, valve formation
48hpf	valve formation
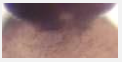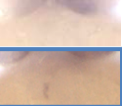
Probes were synthesized as described in the Materials and Methods section. Nested PCR to amplify the DNA templates for probe synthesis was performed on a cDNA mix that was prepared from RNA extracted from embryos at the aforementioned developmental stages.

When we were unable to amplify a specific product, we assumed that the gene was not expressed at the stages under study and therefore not a key player in heart development. We managed to amplify probe templates for all 34 genes for which we were able to assign an orthologue. We were unable to identify an orthologue for 2 genes: *ACTG1* (a gene now firmly implicated in hearing loss) and a hypothetical gene without an annotated name that is at present no longer annotated as a gene in Ensembl (Table 16).

3. positive controls

We first analyzed the expression of the 8 known CHD genes we identified. The results are summarized in Table 15. This analysis revealed that the zebrafish orthologues of three genes - *NKX2-5, TBX1* and *NOTCH1* - have a specific expression in the developing zebrafish heart (i.e., a restricted expression pattern that includes the developing heart) (Figure 23). *tbx1* expression in the developing zebrafish heart and branchial arches has repeatedly been described and we therefore did not reanalyze it in this study[82,239]. Since *NKX2-5* and *NOTCH1* mutations cause isolated CHDs in humans, the remainder of the features present in the patients with the indel must be caused by other genes in the region, and these thus probably represent contiguous gene syndromes.

Table 15: Imbalances encompassing known CHD genes and expression analysis of these CHD genes in the developing zebrafish heart

| imbalance (number of genes)	ref	CHD genes (prioritized rank)	Human phenotype (MIM number)	zebrafish orthologue	Heart expression (36-48hpf)
del5q23 (125)	58	FBN2 (1)	congenital contractural arachnodactyly (121050)	fbn2	
del5q35.1 (45)	58	NKX2.5 (1)	ASD with AV conduction defects (108900)	nkx2-5	
del5q35.3 (129)	58	NSD1 (1)	Sotos syndrome (117550)	nsd1	
del9q34.3 (106) dup20q13.33	58	EHMT1 (5)	chromosome 9q subtelomeric deletion syndrome (610253)	ehmt1a	
				ehmt1b	
		NOTCH1 (1)	aortic valve disease (109730)	notch1a	
				notch1b	
dup16p13.3 (3)	130	CREBBP (1)	Rubinstein-Taybi syndrome (180849)	crebbp	
dup22q11.2 (135)	58	TBX1 (1)	microduplication 22q11.2 (608363)	tbx1	heart & branchial arches
dupXq21.1 (5)	125	ATRX (1)	α-thalassemia/mental retardation syndrome, X-linked (301040)	atrx	
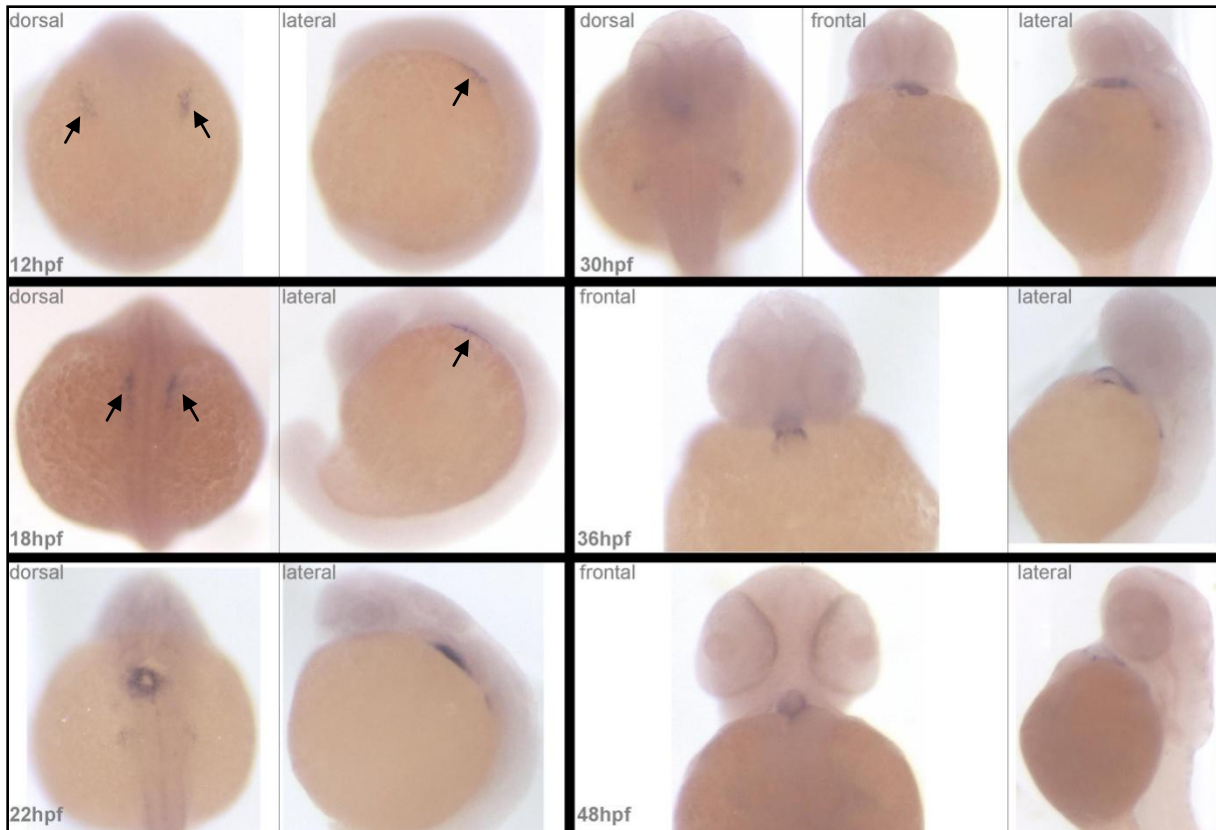
Figure 23: mRNA expression of *nkx2.5* in the developing zebrafish, showing expression in the (prospective) cardiac cells at all investigated developmental stages. Staging of the embryos was done by timed matings and development for a defined period (indicated in the lower left corner). Orientation of the embryo is indicated in the upper left corner. Arrows indicate early expression of *nkx2.5* in cells migrate towards the midline that will form the heart (heart primordia).
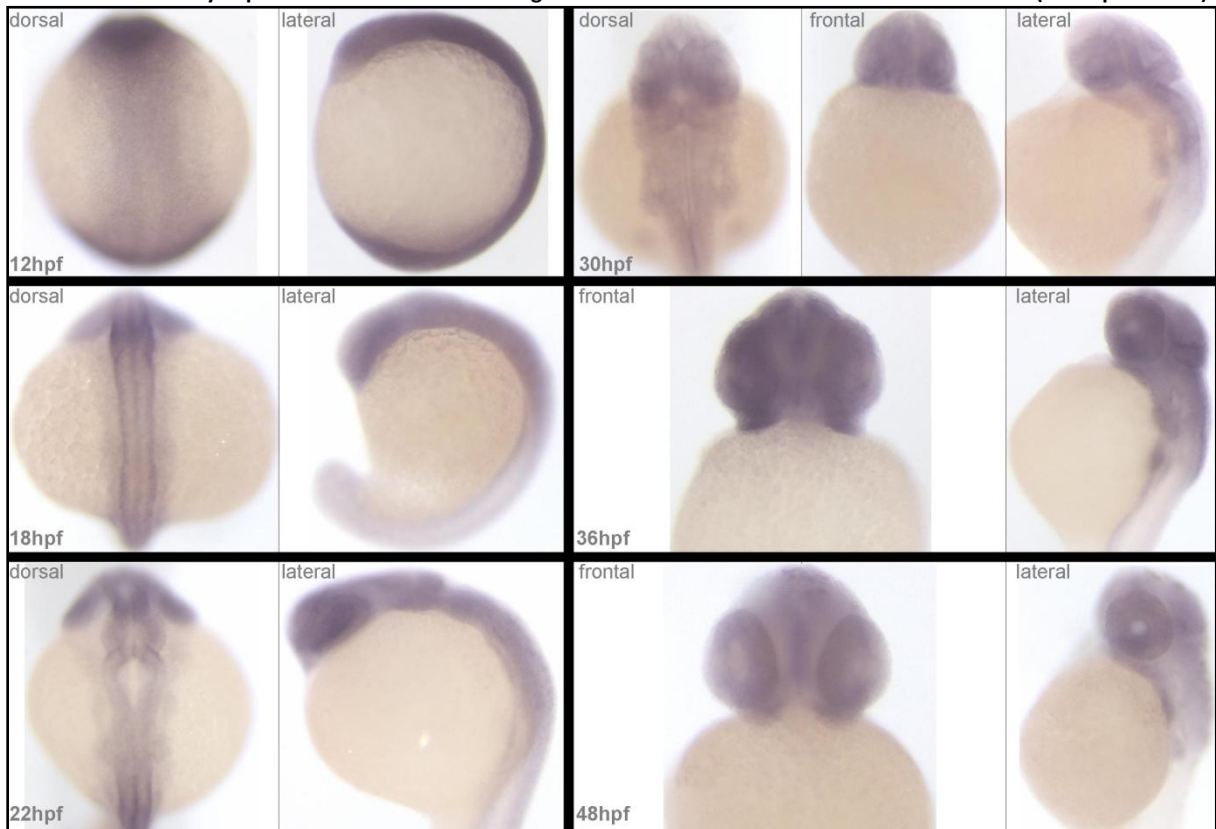

Figure 24: images of *cbp* mRNA expression is the developing zebrafish, demonstrating the ubiquitous nature of its expression. Legend as in Figure 23.

Mutations in *EHMT1*, *ATRX*, *CREBBP*, *NSD1,* and *FBN2* cause syndromic CHDs in humans[117,119,155,186,240]. In our studies, the orthologues of these genes appeared to be expressed "ubiquitously" i.e. throughout the developing zebrafish without a specific expression restricted to the developing zebrafish heart (for example: *cbp* in Figure 24). This pattern is consistent with the multisystem involvement of the syndromes caused by mutations in these genes. Nevertheless, genes that cause syndromic CHDs (such as *TBX1*, which causes DiGeorge syndrome[241]) can have specific expression in the developing heart. In conclusion, in 3 of the 8 known CHD genes, the specific cardiac expression suggests involvement of the gene in cardiac development. Five out of six syndromic CHD genes show a ubiquitous expression. As we will show in the next section, such an expression pattern is common to many genes in the developing zebrafish, and most genes for syndromic CHDs can therefore not be identified by their expression pattern.

4.   candidate genes

4.1. genes identified by ENDEAVOUR

The 22 candidate genes for CHDs that were selected using the cardiac ENDEAVOUR algorithm described above were analyzed in developing zebrafish embryos. Results are summarized in Table 16, and complete results can be consulted on CHDWiki on the respective gene pages (homes.esat.kuleuven.be/~bioiuser/chdwiki/). Two genes (*hand2* and *bmp4*) showed a specific expression in the developing heart. *hand2* is expressed at 12 hours post-fertilization (hpf) in the lateral plate mesoderm where the cardiac cells become specified. Later in development (22-48hpf) it is expressed in the primitive heart and the pharyngeal arches that give rise to the outflow tract and large vessels in humans (Figure 25). *bmp4* is specifically expressed at 18hpf in the specified cardiac cells as they migrate towards the midline, later in the developing heart and at 36hpf at the regions where valve development initiates (Figure 26). Knock-out studies in mice have shown that both Bmp4 and Hand2 are required for cardiac development in a dosage-dependent manner[242,243]. They thus are strong candidates to cause - upon haplo-insufficiency by deletion - the CHD observed in humans carrying a deletion of this gene. Finally, *vegfc* (involved in angiogenesis) is expressed in the vasculature (dorsal aorta, dorsal cardinal vein). Lack of a detectable

expression in the developing heart prompted us however to discard this gene for further studies despite the established role of its paralogue (*VEGFA*) in ToF[86].

Table 16: Chromosomal imbalances that do not contain a known CHD gene and expression analysis in the developing zebrafish heart of high-ranked genes after gene-prioritisation. NA: not available; No ID: no name for this gene annotated by the Human Gene Nomenclature Committee (HGNC)

Imbalance (number of genes)	ref	genes	zebrafish orthologue	Heart Expression (36-48 hpf)
del1p36.33 (97)	151	DVL1	dvl1	
		TP73	tp73	
		HES4	hes4	
		SKI	skia	
			skib	
del4q34 (41)	153	HAND2	hand2	
		HMGB2	hmgb2a	
			hmgb2b	
		VEGFC	vegfc	
del14q22q23.1 (61)	153	ARID4A	arid4a	
		DLG7	dlg7	
		BMP4	bmp4	
		OTX2	otx2	
dup17q25.3 (78)	NA	No ID	NA	
		AATK	aatka	
			aatkb	
		ACTG1	NA	
		CSNK1D	csnk1d	
		FOXK2	foxk2a	
			foxk2b	
		MAFG	mafg	
del22q12.2 (26)	153	THOC5	thoc5	
		EWSR1	ewsr1a	
			ewsr1b	
		KREMEN1	kremen1	
delXp22 (33)	238	RAI2	rai2	
		REPS2	reps2	
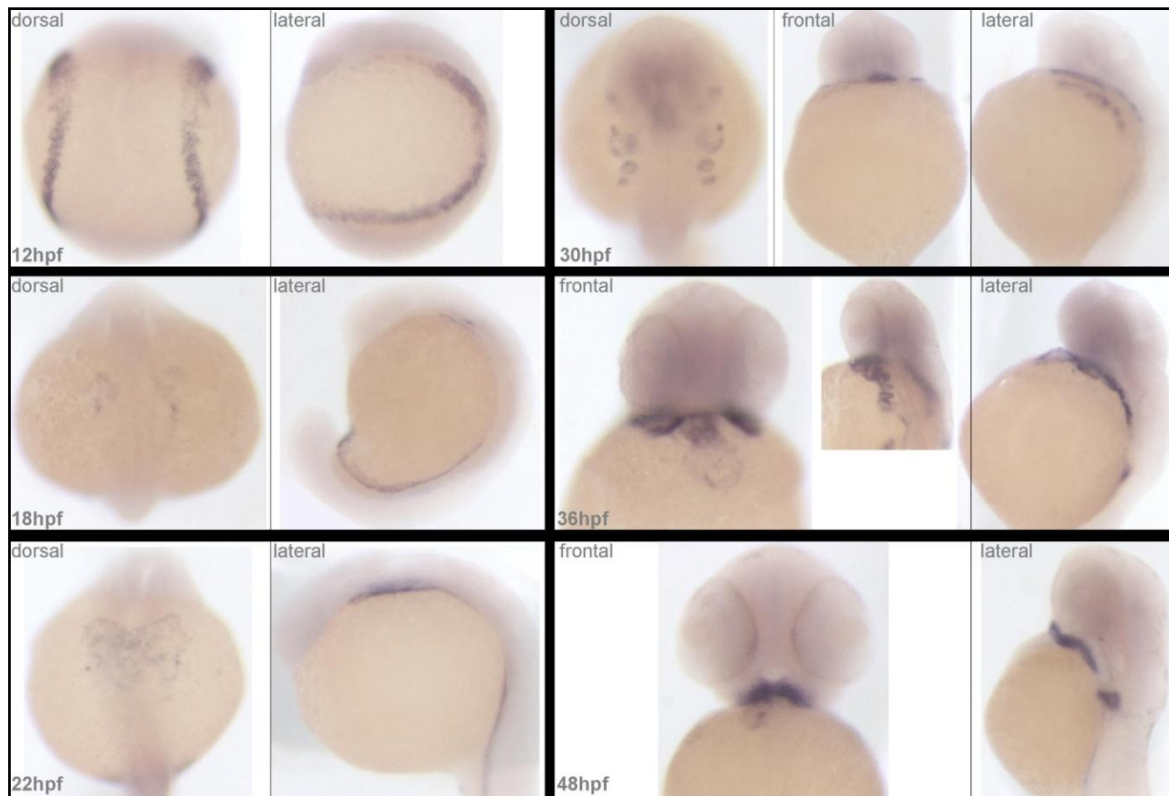
Figure 25: *hand2* mRNA expression in the developing zebrafish: expression in the (prospective) cardiac cells at all investigated developmental stages, as well as in the lateral plate mesoderm (12hpf), the ventral mesoderm (18hpf) , the branchial arches (30hpf and 36hpf (insert)) and the pectoral fin bud (36 & 48hpf). Legend as in Figure 23.
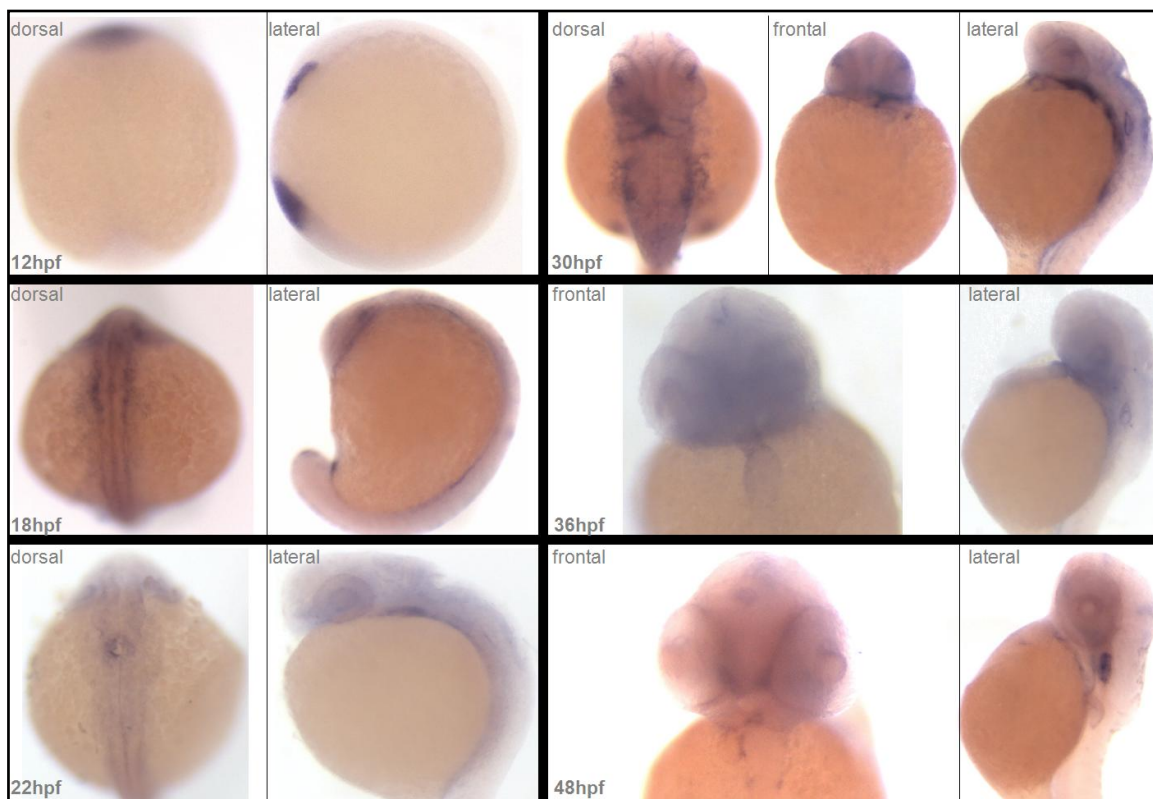
Figure 26: *bmp4* mRNA expression in the developing zebrafish: expression in the (prospective) cardiac cells from 18 to 36hpf, as well as in the posterior epidermis and the rudimentary hatching gland (polster) (12hpf), the ventral mesoderm (18hpf), the developing eyes (18 to 30 hpf), branchial arches (30hpf), otic placode (30hpf to 48hpf) and heart valves (36 & 48hpf) and pectoral fin bud (48hpf). Legend as in Figure 23.

4.2. TAB2

In a patient with a VSD, aortic coarctation and a hypoplastic aortic arch, we identified a *de novo* 6.76 Mb deletion in chromosome 6q24q25 (Figure 27, patient B). Several other patients with a CHD were reported with a deletion in the same region. In this study, we delineated the size and position of several of these deletions more precisely using aCGH (Figure 27A). Genotype and phenotype of these patients is described in Table 17.

TAB2, a gene located in the critical deletion region on 6q25.1, ranked first after prioritization by the CARDIAC ENDEAVOUR in a set of 76 candidate genes from 6q25 (Figure 29B). The corresponding p-value (4.17×10^{-7}) was highly significant. For no other gene in the critically deleted region, a significant p-value below 0.01 was obtained. Of interest, a collaborating group from Denmark studied a family (family A) in which *TAB2* is disrupted by a translocation breakpoint (Figure 27B). This translocation was found in 3 related individuals with a CHD (Aortic valve malformation), arrhythmia or cardiomyopathy (Figure 29A).

Whole-mount *in situ* hybridizations on wild-type zebrafish embryos that were staged and paraformaldehyde-fixed revealed ubiquitous *tab2* expression early in development (12hpf). Later in development a more restricted expression pattern was observed, with high expression in the developing bulboventricular valve, the dorsal aorta and common cardinal vein (Figure 28). The expression in the bulboventricular valve of the zebrafish (equivalent of the human aortic valve) further strengthens a role for *TAB2* in causing the CHD in patients with a 6q25 deletion.

Table 17: Reference of description, genotype and phenotype of patients with a molecularly delineated deletion of chromosome 6q24 and/or 6q25

	ref[a]	breakpoint regions on chromosome 6 (Mb)		heart defect	develop-mental delay	micro-ceph	IUGR	PNGR	CNS	oral	eye	hearing	Dysmorphism
		proximal	distal										
A	*	translocation		AS, arrhythmia, cardiomyopathy	N	N	N		nl	nl	nl	nl	N
B	*	143.71-143.80	149.98-150.25	VSD, hypopl AoA, CoAo	mild-mod	?	Y	Y	nl	nl	strabism	/	Hypospadias, epicantic folds, small chin, small ears,
C	*,I1	142.34-142.35	152.48-152.49	tricuspid & pulmonic valve dyspl, mvp	mild	N	Y	Y	nl	nl	astigmatism	/	Frontal bossing, bitemporal narrowing, PF short & slant up, hooded eyelids, sparse eyelashes, infraorbital folds, anteverted nares, thin upper lip , fetal pads (hands), low set dyspl ears
D	*,I2	146.17-146.18	156.55-156.59	ASD-II	mild	N	Y	Y	/	nl	retinal pigment changes, minimal optic nerve hypopl	/	Sacral dimple, narrow thorax, frontal bossing, bitemporal narrowing, PF short & slant up, sparse eyelashes, infraorbital folds, anteverted nares, thin upper lip, flat long philtrum, low posterior hair line, low set dyspl ears
E	II	148.70-148.78	151.38-151.35	ASD, PDA	mild	N	N	Y	nl	HP	nl	nl	Facial asymmetry, short PF, medial flare eyebrows, anteverted nares, thin upper lip, long & smooth philtrum, sandal gap, dyspl ears, cupshaped left ear
F	III1	149.09-149.23	156.06-156.30	VSD	mild	/	N	Y	/	nl	/	/	Short nose, full cheeks, smooth philtrum, thin upper lip, almond eyes, low set protruding ears, dyspl left ear
G	IV	151.98-152.03	155.29-155.34	N	/	/	/	/	nl	CLP	/	/	/
H	V3	149.95-149.96	160.27-160.28	N	Y	Y	N	Y	Acc	HP, BU	/	mild cHL	Midface hypopl, clinodact 5, plagiocephephaly, low set, posteriorly roated ears
I	*,I3	140.33-140.35	146.47-146.48	N	N	N	Y	Y	Nl	nl	/	nl	N
J	VI	151.16-152.37	157.79-158.68	N	mild-mod	Y	N	Y	Acc	CHP	nystagmus	/	Dolichocephaly, anteriorly placed anus, low set ears
K	V1	155.09	158.87-158.88	N	Y	Y	N	N	/	N	N	snHL	Long philtrum, micrognatia, PF slant down, epicanthus, low set ears

[a] *: this report; I:Nowaczyk[244]; II: Caselli[245]; III: Bisgaard[246]; IV: Osoegawa[247]; V: Nagamani[247]; VI: Pirola[248].

N = no, Y = yes, / = not available, nl = normal; heart defects as in *Abbreviations and Acronyms*

Acc = Agenesis of the corpus callosum, BU = bifid uvula, CHL = conductive hearing loss, CHP = cleft and high palate, CLP = cleft lip and palate, CNS = Central nervous system, dyspl = dysplastic, HP = high palate, hypopl = hypoplasia, IUGR = intra-uterine growth retardation. LS = low-set, mod = moderate, PNGR = post-natal growth retardation, SNHL = sensorineural hearing loss, PF = palpebral fissures.
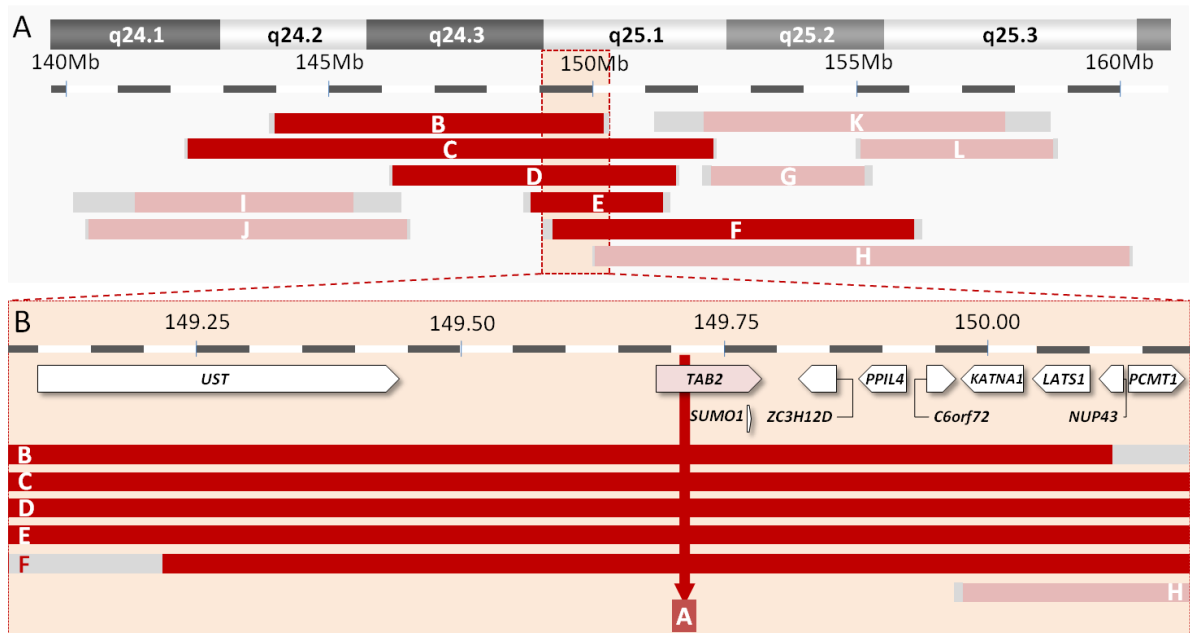
Figure 27: A: molecularly delineated deletions affecting the chromosomal region 6q24q25. Genotypes of patients B, C, D and J were further delineated in this study. Deletions associated with CHDs are depicted in dark red, while light red denotes deletions without CHDs. Regions containing deletion breakpoints are shown in grey. The critical region for CHDs is demarcated by a light red box and shown in more detail in B. **B:** The critically deleted region, and the location of the breakpoint on 6q of the translocation segregating with CHD in family A. The location and orientation of the genes in this region is indicated by the arrows. All CHD patients that carry a deletion in this region have a deletion that includes *TAB2*.
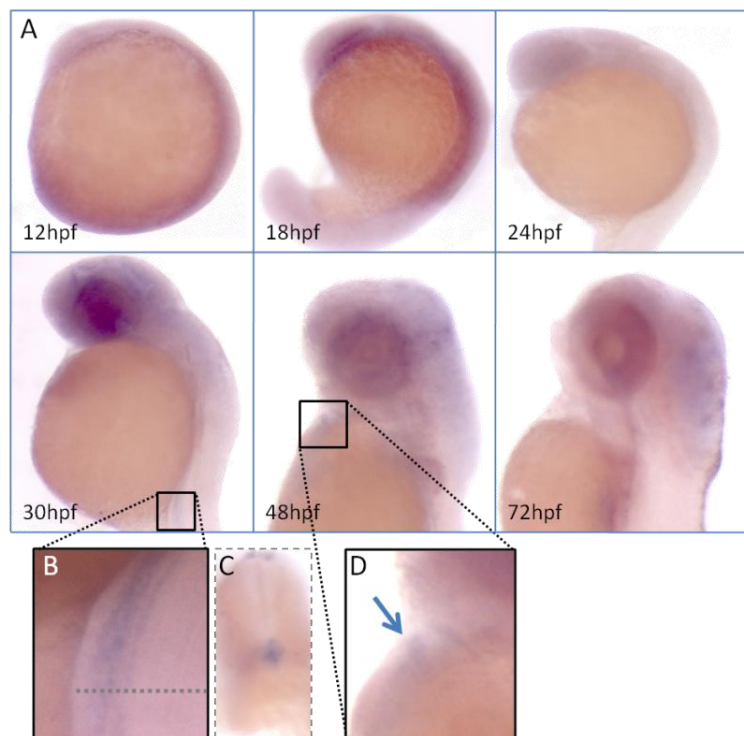
Figure 28: A: mRNA expression of *tab2* in the developing zebrafish at distinct developmental stages (indicated in the lower left corner). **B:** Magnification showing expression of *tab2* in the dorsal aorta and common cardinal vein. **C:** cross section through the hindbody of a zebrafish at 30hpf, showing restricted expression in the dorsal aorta and common cardinal vein. **D:** expression of *tab2* in the bulboventricular valve at 48 hpf (arrow).
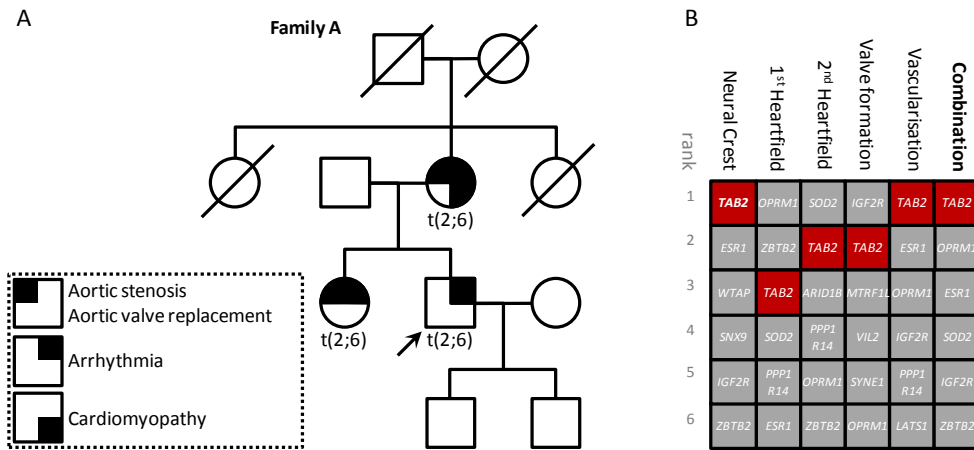
Figure 29: A. pedigree of family A, where a balanced translocation segregates with congenital heart disease (as coded in the box). B. prioritization of all 76 genes of 6q25, showing high ranking of *TAB2* in all models, and first-place ranking in the combined prioritization.

CHAPTER 7   SEQUENCING NOVEL CANDIDATE GENES FOR CHDS

From our studies described in the previous chapters, we identified three strong candidate genes for CHDs: *BMP4*, *HAND2* and *TAB2*. In an attempt to prove the involvement of these genes in human heart development and the pathogenesis of CHDs, we decided to investigate the presence of mutations in these 3 genes in patients with a CHD. The function of each gene is also briefly discussed in order to situate where in cardiac development they (may) function.

1. BMP4

BMP4 (bone morphogenetic protein 4) encodes a signaling molecule of the TGF-β family, and has been reported to be involved in mesoderm induction, tooth, limb, bone, eye and heart development[242,249,250]. Much effort has been put into the investigation of the phenotype caused by *Bmp4* mutations in mice. $Bmp4^{+/-}$ x wildtype intercrosses resulted in heterozygote frequencies below the expected mendelian ratios at weaning. 25% of $Bmp4^{+/-}$ died within 3 weeks after birth, and $Bmp4^{+/-}$ mice were affected to a variable degree and with a limited penetrance by cystic kidneys, polydactyly, and craniofacial, eye, male gland and heart defects[242,250].

Recently, a loss-of-function *BMP4* mutation was identified in a family with 6 individuals with autosomal dominantly inherited eye and hand malformations. The phenotypic expression of this mutation was highly variable (ranging from myopia to microphthalmia and from cutaneous webbing of the fingers to postaxial polydactyly, with occasionally also brain malformations) and the penetrance was reduced[251] (as was reported for the heterozygous mice).

Only one out of eleven Bmp4+/- mice was reported to have a VSD, indicating variable penetrance of CHDs in mice. Also in humans, CHDs may be a rare manifestation of haploinsufficiency of BMP4: no CHDs were reported in the family with a *BMP4* mutation, and only one out of the six 14q22 deletion patients reported had a cardiac involvement[138,251-254]. These data indicate that *BMP4* mutations may be a rare cause for CHDs. Several groups have searched for BMP4 mutations in humans with CHD. In over 200 isolated CHD patients, no mutations were detected thus far (Prof. J. Goodship, University of Newcastle, UK, personal

communication and reference [78]). However, given the phenotype of the BMP4 mutation observed in the single family, BMP4 mutations are more likely to be found in patients with hand and eye malformations, possibly associated with a CHD, rather than in isolated CHD patients. We analyzed 6 CHD patients with eye or hand malformations from our center for which DNA was available, but no mutations were identified

2. HAND2

HAND2 (Heart- and neural crest derivatives-expressed protein 2) encodes a transcription factor of the basic helix-loop-helix domain family. It plays a role in heart development and homozygous knockout of the gene in mice causes heart defects. *Hand2* displays partial redundancy with *Hand1*: heterozygous KO of *Hand2* in mice does not result in an overt phenotype, but heterozygous KO of *Hand2* in mice with a heart-specific homozygous KO of *Hand1* causes severe cardiac defects and death at E10.5. When 2 or more alleles of the paralogues (*Hand2* and *Hand1*) are knocked out, mice develop cardiac defects (except in *Hand1/Hand2* double heterozygote mice)[243,255]. Although these data suggest that *HAND2* is a dosage-sensitive gene, the absence of cardiac malformations in $Hand2^{+/-}$ mice indicate that *Hand2* is not dosage-sensitive upon haplo-insufficiency in mice.

To investigate whether *HAND2* mutations are responsible for CHDs in humans, we sequenced the coding part of the gene, the annotated miRNA-1-1 binding site in the 3' UTR[256] and the intron-exon boundaries in 35 patients with an isolated ASD. No pathogenic mutations were detected.

In two patients, the same nucleotide alteration was detected: in the polypyrimidine stretch between the branching site and the 3' splice site of the only intron of *HAND2*, there was a change from G to C (AC093849.2(HAND2):g.135915G>C). This is not an annotated polymorphism. This change is predicted to render the polypyrimidine stretch slightly less conforming to the consensus sequence. However, this is most likely a non-pathogenic mutation, as mutations of this stretch have never been shown to cause a genetic disorder[257].

3. TAB2

In contrast to HAND2 and BMP4, a role for TAB2 in heart development has not previously been reported. We will therefore discuss in more detail the roles of this protein and its binding partners.

3.1. structural and biochemical properties

TAB2 (TAK1-binding protein 2, also known as *MAP3K7IP2*) encodes a protein that associates with a member of the mitogen-activated protein kinase family, TAK1 (TGF-β activated kinase, also known as MAP3K7)[258]. The C-terminal coiled coil domain of TAB2 can bind TAK1[258], but TAB2 can also associate with itself or its paralogue TAB3[259]. The N-terminal CEU domain of TAB2 (**c**oupling of **u**biquitin conjugation to **E**R degradation) enables TAB2 to bind ubiquitinated proteins[260]. TAB2-TAK1 or TAB3 TAK1 complex formation and the coupled TAK1 activation is initiated by polyubiquitination of upstream signaling molecules[260-262]. TAK1 probably activates itself and downstream cascades of kinases by autophosphorylation on Thr-187 and/or Ser-192. TAK1 autophosphorylates when it is over-expressed *in vitro* [263], but it is assumed that under physiological conditions, this activation is enhanced or enabled by complex formation with the TAB proteins[264,265].
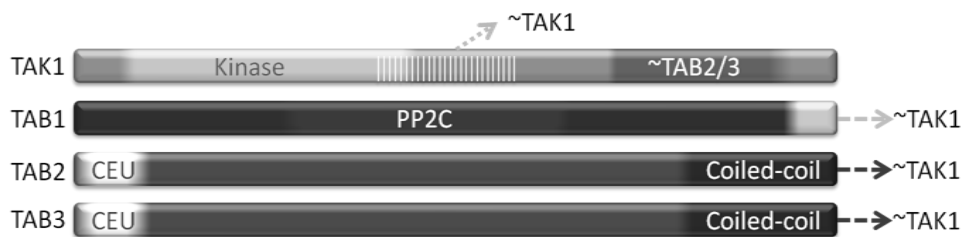
Figure 30: Protein domains of the TAK1 protein and the TAB proteins, and regions of interaction between these proteins.

The redundancy of TAB2 with TAB3 for TAK1 activation is still a matter of debate, and both likely play partially overlapping roles in different cell types and different pathways. TAB2 has been shown to be essential for TAK1 activation in mouse embryonic fibroblasts (MEFs) and MEK-293 cells[262] while in HeLa cells functional redundancy between TAB2 and TAB3 was demonstrated by co-knockdown[259]. Similarly, TAK1 was shown to be essential for response to TLR signaling in B-cells and IL-1- and TNF-response in MEFs.

TAB1 is not homologous to TAB2, and contains a PP2C (Protein phosphatase 2C) domain. TAB1 constitutively interacts with TAK1. TAB1 is able to activate TAK1 upon overexpression, but it is unlikely that TAB1 interaction is sufficient for TAK1 activation under physiological conditions. TAB1 is however shown to be essential for TAK1 activation upon some signals[263]. Although binding of TAB2 and TAB1 to TAK1 are not mutually exclusive, it is for the moment still unclear if both play partially redundant and independent roles in TAK1 activation, or whether both are necessary for TAK1 activation. TAB4 (also known as TIPRL), is a protein distantly related to TAB1. It has recently been discovered That it can have a role in TAK1 activation similar to and/or cooperatively with TAB1[266].
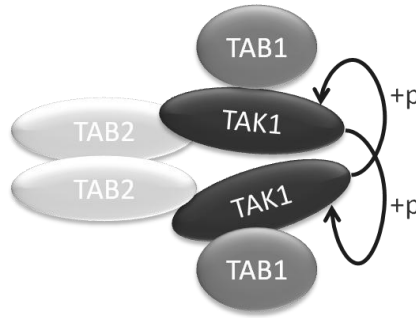
Figure 31: proposed model for TAK1 activation: TAB2 (or TAB3) proteins oligomerize upon upstream signaling, and bring together TAK1 proteins, that (aided by TAB1) are activated by autophosphorylation (after Chen[267]).

3.2. roles in cell biology and signal transduction

Numerous studies have investigated the signals that function upstream of TAK1 and TABs, and the roles of TAK1 and the TABs in the transduction of these signals to the vast array of downstream effectors. The results are summarized below. Most attention has been devoted to their role in the inflammatory and innate immune response, where TAK1 was shown to be responsible for the phosphorylation activation of IKK (causing NF-κB activation), JNK and p38 upon signaling by several ligands (cytokines like IL-1, TNF, LPS, or antigens) via the IL1 receptor, the TNF receptor, the TLR4 receptor or the T and B cell receptors respectively. These signals directly or indirectly cause (auto-) polyubiquitination of proteins like TRAF6 or TRAF2, providing a scaffold for binding with TAB2 and/or TAB3[262]. The TRAF2/6-TAB2/3 complex probably translocates to the cytosol to associate with TAK1-TAB1 and cause TAK1 activation[258]. For a review on this subject we refer to a recent paper by Chen[267].
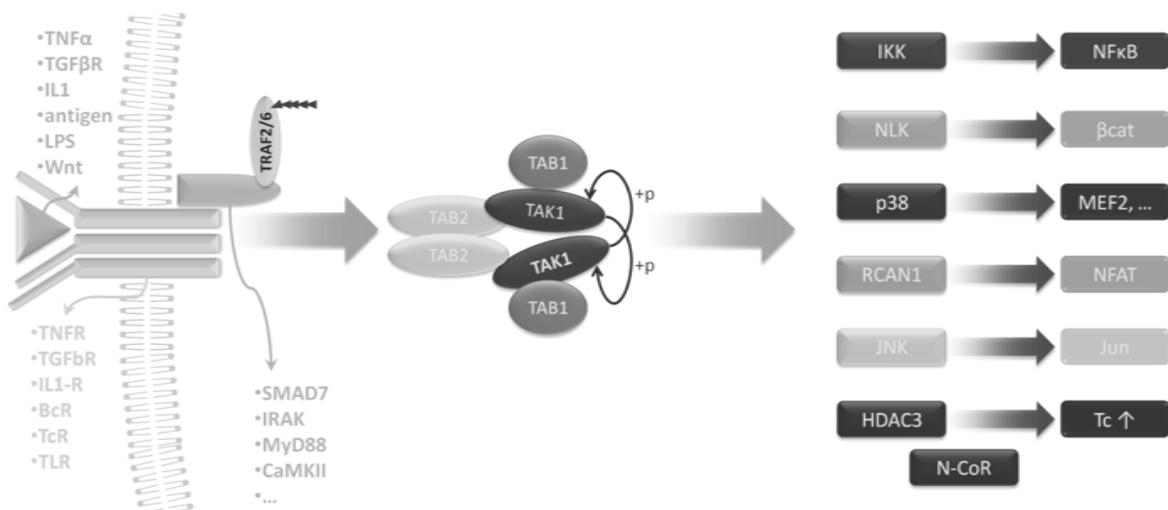
Figure 32: roles of the TAK-TAB complex in transduction of extracellular signals to the nucleus. From left to right: binding of extracellular ligands to membrane-bound receptors causes recruitment and activation of membrane-associated proteins. At the receptor, in the cytosol or in the nucleus, these signals activate the TAB-TAK complex. This complex can transduce signals to a variety of molecules that effectuate or modulate transcription of downstream target genes.

Several cell-based studies also clearly demonstrated roles of TAK1 besides inflammation, mainly in NLK (NEMO-like kinase) phosphorylation activation upon TGF-β or Wnt5a signaling. Active NLK can associate with NF-κB, CHD7 and/or TCF/LEF causing a decrease in β-catenin (canonical Wnt) and/or PPAR-γ-mediated transcription[268,269]. On the other hand, it was shown that Wnt5a signaling can also cause CaMKII activation, leading in turn to TAK1-TAB2 activation and NFAT signaling[270]. Finally, TAB2 was also shown to function in the nucleus in a repressor complex together with the p50 NF-κB, HDAC3 and N-CoR, where it (upon activation) exports HDAC3 from the nucleus, thereby relieving its repressive mark[271].

3.3. roles in development

Studies using *Tak1*[-/-], *Tab1*[-/-] and *Tab2*[-/-] knockout mice have shown that Tak1, Tab1 and Tab2 are essential for a normal embryonic development.

Tak1 is expressed ubiquitously early in development (E8), while expression is more restricted to specific organs later in development (nervous system, testis, kidney, liver, lung, pancreas and gut). *Tak1*[-/-] mice die around E10.5 to E12.5 with neural tube dysmorphogenesis and vascular defects [272,273]. Several mice with a tissue-specific modulation of *Tak1* activity were also generated. Most notably, Xie and colleagues reported on mice with cardiac-specific expression of a dominant-negative Tak1, showing that inhibiting Tak1 function in the heart leads to altered electrical conduction (shortened PR interval), impaired ventricular filling and cardiac hyperthropy[274]. They also discussed unpublished results that cardiac-specific deletion of Tak1 causes death at mid-gestation[274]. Zhang and colleagues reported on mice with cardiac-specific expression of activated Tak1, showing that they develop cardiac hypertrophy[275]. These combined results demonstrate that Tak1 is required for proper heart development, and that an increase or decrease of Tak1 signaling in the heart leads to cardiac disorders. In addition to this, tissue or time specific deletion of Tak1 showed its requirement in the T cells, B cells, keratinocytes, epidermis, enterocytes and hepatocytes[276-281].

Tab1 is expressed ubiquitously at mid gestation, and strongly at E12.5 in the heart ventricle and the lung mesenchyme[282]. *Tab1*[-/-] mice die at late gestation (around E18.5), and have cardiovascular and lung developmental defects. The aorta appeared thin and histologically disorganized. The heart ventricle displayed a thin wall, frequently VSDs, aberrant trabeculation and abnormal blood vessels. Expression of *Nppa* was decreased[282].

Tab2 shows a high expression in the fetal liver at E9.5[283], and in the dorsal aorta at E12.5[284]. Throughout heart development, Tab2 is significantly more expressed in the developing heart than its homologue Tab3, suggesting Tab3 is not redundant with Tab2 for heart development (Figure 33). *Tab2* is expressed ubiquitously in the adult mouse[283], with higher expression in the liver and heart. *Tab2*[-/-] mice die around E12.5 and display severe liver degeneration, with increased hepatocyte apoptosis. Interestingly, 70% of *Tab2*[+/-] mice die within 1 week after birth, but the reason for this increased mortality is unknown[283].
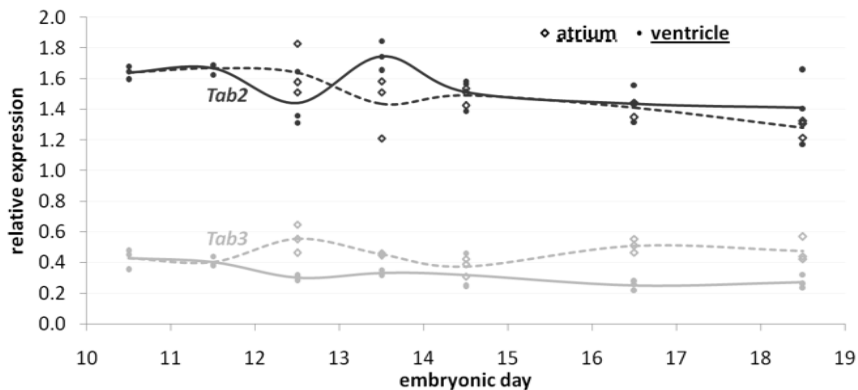
Figure 33: *Tab2* (dark) and *Tab3* (light) mRNA expression in the developing mouse heart (atrium and ventricle separately) on 7 different time points, as measured by micro-array expression analysis (data derived from GEO: GSE1479). Results from 3 replicates are displayed (data points), as well as averages per condition (lines).

3.4.   dosage-sensitivity in zebrafish

To investigate if the *tab2* gene is also dosage-sensitive in zebrafish and if it is involved in heart development, we knocked down its expression by injecting a morpholino (MO) directed against the splice acceptor site of the second (largest) coding exon (exon 3) into the yolk of embryos before they were at the 4-cell stage (Figure 34A). Analysis of this knock-down on mRNA extracted at 8hpf demonstrated that injection of between 2 and 3 ng of MO resulted in halving the amount of correctly spliced mRNA (Figure 34D). PCR using primers in exons 1 and 7 yielded 3 bands, of which 2 correspond to the wild-type situation. Sequencing the additional band revealed that exon 3 was absent from the new transcript. This is predicted to result in a premature stop codon after 6 nucleotides of exon 4 (Figure 34A). With less than 5% of the wild-type protein transcribed, this most likely represents a loss of function situation. To confirm this and to assess specificity of the phenotype, a second MO was injected that is directed against the translation start site of the tab2 gene. This translation-blocking MO (tbMO) yielded a similar phenotype as the sbMO, showing specificity of the knockdown. A phenotype was observed when the percentage of normal mRNA was below 60%, i.e. after injection of 2 ng of sbMO. This demonstrates that also in

zebrafish the *tab2* gene is dosage sensitive, corresponding to the haplo-insufficiency in the human (Figure 34B and D).
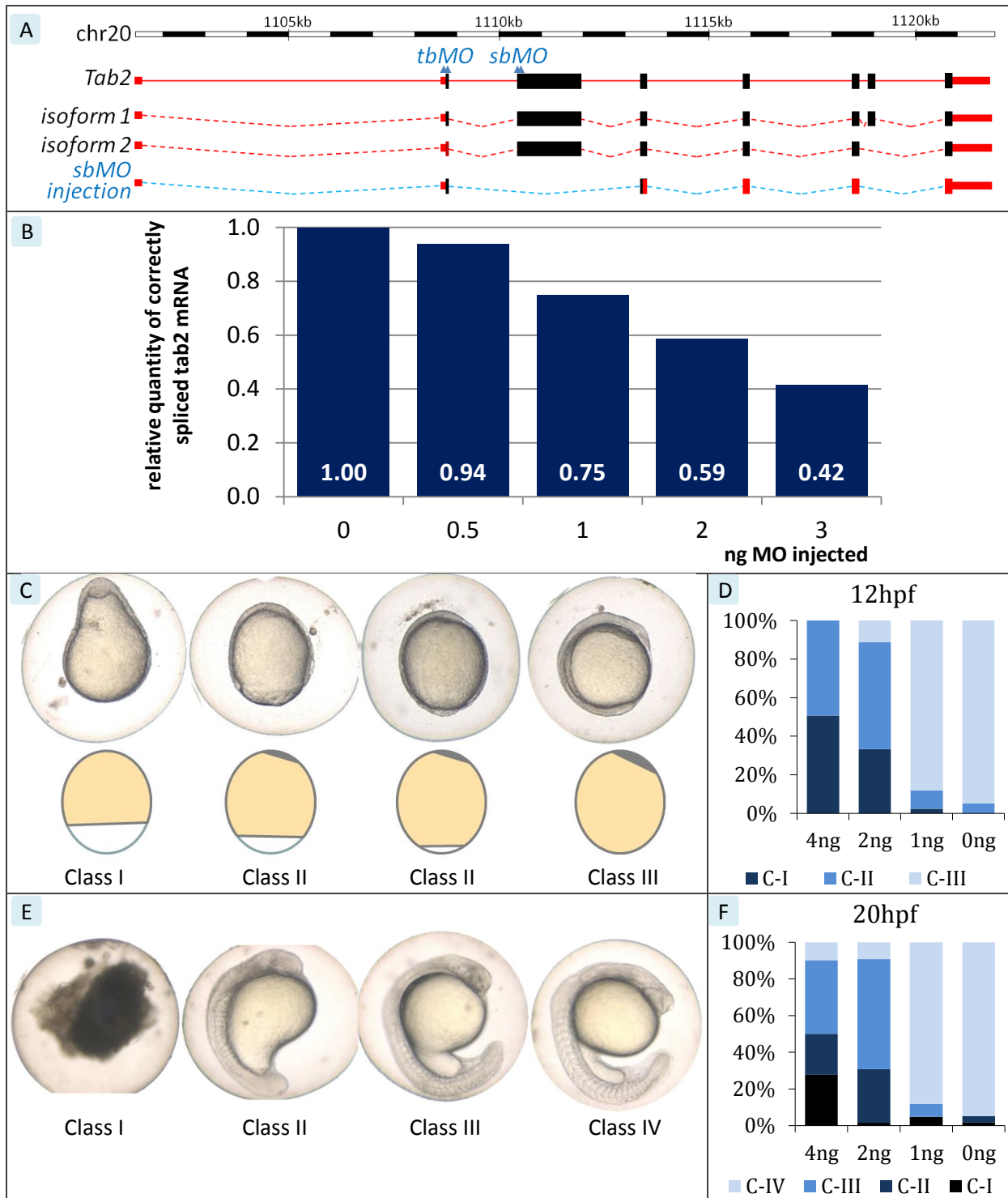
Figure 34: A: genomic organization of the *tab2* gene in the zebrafish. The position of all exons and the composition of the alternative transcripts is indicated. Coding sequences are indicated as black bars, non-coding sequence as red. The target sites of the sbMO and the tbMO used in this study are indicated by blue double arrowheads. The alternative transcript generated by sbMO injection is indicated below (blue introns). **B:** effect of sbMO injection on *tab2* mRNA splicing. **C:** phenotypes upon MO injection after 12hpf. In the middle, a cartoon illustrating the delayed progression of the yolk syncytial layer in 12hpf embryos. **D:** classification of embryos at 12hpf after sbMO injection at the dose as indicated on the X axis. Classification follows panel C. **E:** phenotypes upon MO injection after 20hpf. **D:** classification of embryos at 20hpf after sbMO injection at the dose as indicated on the X axis. Classification follows panel E.

The morphant embryos displayed early embryological defects (delayed epibolic movements during gastrulation). They further developed into abnormal larvae that were shorter, had a curved tail, abnormal somites, a dysfunctional heart and small head. The early defects observed in these embryos preclude a conclusive analysis of the effect of *tab2* knockdown on zebrafish heart development: although the hearts of morphant larvae were abnormally structured and displayed heart failure, we cannot exclude that these effect are secondary to the earlier defects in gastrulation. Other reports have similarly shown that knock-out of certain genes causing isolated heart defects in humans causes a much broader phenotype in zebrafish[84]. Probably tab2 has other roles during zebrafish development, for example in the epibolic movements that have no direct correlate in mammalian development. Nevertheless, the clear dosage-sensitivity of tab2 in zebrafish again demonstrates that a defined dosage of tab2 is required for normal development.

3.5. *TAB2*, a dosage-sensitive candidate gene for CHDs

In the previous chapters, we have shown that *TAB2* lies within a critical, 1.25Mb region on chromosome 6q that is associated with CHDs. Gene prioritization pinpointed *TAB2* as the best candidate gene for CHDs amongst all genes that lie within about 10Mb of TAB2. Moreover, *TAB2* is disrupted by a translocation in 3 individuals with congenital heart disease. These data suggest that human heart development is sensitive to dosage of *TAB2*. Also in mice, normal Tab2 dosage is essential for survival, although no cause for the increased mortality in haplo-insufficient mice could be identified. Tab2 is expressed in the heart and aorta of the developing mouse, and - like the expression of many other genes causing isolated CHDs - this expression is conserved in zebrafish. Also in zebrafish, tab2 is dosage-sensitive. Also TAB1 is required for proper heart development. Biochemical studies have shown that TAB2 is an essential mediator of TAK1 activation in some tissues, and embryological studies have shown that - for proper heart development - TAK1 activity is regulated within a defined range. Of interest, we identified a *TAK1* duplication in a girl with a pulmonic stenosis (data not shown). This duplication was inherited from her normal mother. Her brother had died early in childhood from a severe pulmonic valve stenosis as well, and was thus unavailable for study.

Taken together, these data suggest that normal *TAB2* gene dosage is required for human heart development. In an attempt to provide more conclusive evidence, we sequenced the

coding sequences and intron-exon boundaries of the *TAB2* gene in 100 patients with isolated CHDs (pulmonic or aortic stenosis). We did not identify any pathogenic mutation. The identified synonymous SNPs were in Hardy-Weinberg equilibrium and were present in frequencies comparable to those reported in a European control population. This lack of demonstrable mutations in *TAB2* is not entirely surprising, since other studies for other cardiac genes have shown that mutations causing CHDs are very rare.

CHAPTER 8   CONCLUSIONS AND FUTURE PROSPECTS

In the present work we have investigated whether cryptic chromosome imbalances are a frequent cause for CHDs. We introduced a novel genome-wide copy number profiling tool (aCGH) and showed that it enabled a reliable detection of such imbalances at a resolution far surpassing the resolution of metaphase karyotyping. The application of this technique in patients with a syndromic CHD greatly enhances the chance of finding an etiological diagnosis. More precisely, in 20% of them, a pathogenic submicroscopic imbalance can be demonstrated. The correct delineation of chromosome aberrations by aCGH also entails a more accurate characterization of the genotype of the patient, permitting a more personalized, specific genetic diagnosis. A diagnosis is of the utmost importance for the follow-up of the patients and their families, as it allows more accurate counseling of patient and parents regarding recurrence risks and the anticipated mental and physical development. In some cases it also impacts treatment of the patients as complications associated with certain genetic conditions can be prevented or managed from a subclinical stage. This impact depends however largely on prior knowledge, which is based on the experience of clinicians with patients that have similar conditions, and for this reason we have also published detailed reports on the genetic and clinical characteristics of some specific patients.

We showed that the application of higher-resolution platforms enables the genome-wide identification of (indel) mutations of single genes (e.g. in *FOXC1* or *ATRX*), but that this increased resolution is accompanied by an unexpected complexity in the evaluation of their causality. Based on our experiences with these high-resolution genome-wide screening techniques, we predict that the eminent introduction of methods that identify (indel) mutations genome-wide at a kilo-base or base-pair resolution will not only be accompanied by significant technical challenges concerning processing and management of the output, but that they will also cause tremendous challenges for the identification and interpretation of causal mutations. This challenge is illustrated clearly by the recent sequencing of the diploid genome of a healthy Asian male. The authors demonstrated that an astonishing 0.42 million single nucleotide mutations they identified in one genome were undescribed, as were over 80.000 indels and other structural genome mutations[285]. We and other have

demonstrated that the identification of a *de novo* (chromosome) mutation is insufficient to provide a diagnosis, and we confirmed that inherited mutations can also contribute to disease. It is moreover unknown at this time at what frequency these *de novo* mutations arise in each generation. Sifting through such numbers of uncharacterized mutations in a single patient to establish a diagnosis will present challenges for clinicians, technicians, bio-informaticians and biologists alike, and will require close collaboration between these disciplines and between research teams worldwide. We expect that open access platforms like the CHDWiki portal we developed will aid in diagnostic decision making by pooling and structuring clinical, genetic, cytogenetic and biological knowledge and by rendering it accessible, editable and minable in a straightforward way to people working in each of these disciplines. When no or many known genes are found to be mutated in these studies, prioritizing valuable candidate genes will be a challenge, and tools such as ENDEAVOUR are seminal in laying the foundations to perform such tasks.

The identification of (chromosome) mutations that predispose to rather than cause syndromic disorders highlights another challenge in the interpretation of a patient's genome: the monogenic view on some genetic disorders is probably too simplistic. The way combinations of mutations converge to cause disorders is a field that is just beginning to be explored. The advent of high-resolution genome profiles will open this field rapidly, but in order to advance, a clear insight in the contribution of different genes and mutations to the pathology needs to be established. A systems-biology approach that entangles the relations between genes seems necessary to enable assessment of how several mutations (and environmental factors) can disturb the biological networks that underlie development.

The identification of submicroscopic indels in syndromic CHD patients enables positional cloning of genes responsible for isolated and syndromic CHDs. We detected many imbalances that affect genes known to cause CHDs. Accordingly, regions identified in this way that do not contain known genes for CHDs represent novel candidate loci. Because some loci (e.g. on 4q, 22q12 or 6q25) are recurrently associated with CHDs, the burden of proof on these regions is much stronger, and these are unlikely not to contain novel genes responsible for CHDs. The use of advanced database mining strategies like ENDEAVOUR aids in selecting valuable candidate genes from these loci, and we showed that there is room for improvement by tailoring these tools to the needs of the underlying clinical or scientific question. This highlighted to us a need for an unbiased and genome-wide characterization of

how (heart) development proceeds at the molecular level, and how these processes are disturbed in specific disorders. Methods to tackle such questions are increasingly becoming available, and the generated output can undoubtedly leverage not only a better prioritization of candidate genes, but especially a better characterization of genes, to understand gene networks and how these are disturbed in specific disorders. On 6q25, CARDIAC ENDEAVOUR identified *TAB2* as the best candidate gene for causing heart defects, and follow-up and literature studies confirmed that this is indeed an excellent candidate. In other regions, genes known to be involved in heart development (*BMP4* and *HAND2*) were similarly identified.

We have used expression analyses in zebrafish embryos to identify the most valuable candidates from a group of high-ranked candidate genes. We hypothesized that genes responsible for human CHDs can be discriminated by their expression pattern in the developing zebrafish, and verified this by analyzing the expression of positive controls: genes that were affected by imbalances and that were known to be associated with CHDs. Our hypothesis proved true only for a subset of these genes: genes involved in syndromic forms of CHDs mostly showed a ubiquitous and thus uninformative expression pattern. This pattern is not surprising given the multi-systemic nature of the associated disorders. But as we started from regions identified in patients with syndromic forms of CHDs, this represented a true impediment: if the unknown CHD-associated gene causes a syndromic CHD, expression patterns often are non-informative. Nevertheless, positive control genes that are involved in isolated CHDs displayed a specific expression in the developing heart, and genes that showed a specific expression in the developing zebrafish heart were considered good candidate genes. Only 2 out of candidate 24 genes displayed such a pattern: *BMP4* and *HAND2*. Comfortingly, both genes were already known to be involved in mammalian heart development through studies in mice. Despite the aforementioned impediment, we consider a positive result of the gene expression analysis to be a strong argument in favor of the involvement of the gene in CHDs. Expression studies in zebrafish moreover enable a fast and high-throughput selection of candidate genes.

Finally, we have identified a novel candidate gene for CHDs: *TAB2*. This gene is deleted in multiple patients with CHDs, is located in the critical deletion region and is ranked first as a candidate gene amongst over 100 genes from the region. Haplo-insufficiency of this gene is described to be associated with a high mortality in newborn mice, and we have shown that

haplo-insufficiency is associated with developmental defects in zebrafish. Although we could not identify pathogenic mutations in a group of 100 patients with isolated heart defects, others did find a disruption of this gene co-segregating with CHDs in a small family. We therefore believe *TAB2* haplo-insufficiency causes CHDs.

aCGH	array comparative genomic hybridization aCGH
AEPC	association of European pediatric cardiologists
AP	alkaline phosphatase
AR	androgen receptor
AS	Aortic stenosis
ASD	atrial septal defect
ATRX	alpha-thalassaemia/mental retardation, X-linked
AUC	area under the curve
AV	atrioventicular
AVSD	atrioventicular septal defect
BAC	bacterial artificial chromosome
BAV	bicuspid aortic valve
CBS	circulary binary segmentation
CCRs	Complex chromosomal rearrangements
CFC	cardio-facio-cutaneous
CHDs	Congenital heart defects
CiCRs	complex intrachromosomal rearrangements
cM	centiMorgan
CNP	copy number polymorphism
CNV	copy number variant
DCM	dilated cardiomyopathy
DIG	digoxigenin
DNA	deoxyribonucleic acid
DOP	degenerate oligonucleotide primed
E	embryonic day
FISH	fluorescent in situ hybridisation
GEO	gene expression omnibus
GO	gene ontology
HLH	hypoplastic left heart
hpf	hours post fertilization
LCRs	low copy repeats
LOOCVs	leave-one-out cross-validations
LV	left ventricular
LVNC	left ventricular non-compaction
Mb	Megabase
MCA	multiple congenital anomalies
MC/DA	Monochorionic diamniotic

MEFs	mouse embryonic fibroblasts
MO	morpholino
MR	mental retardation
MRI	magnetic resonance imaging
NAHR	non-allelic homologous recombination
OMIM	online mendelian inheritance in man®
PAC	p1-derived artificial chromosome
PAPVC	partially anomalous pulmonary venous connection
PCR	polymerase chain reaction
PDA	patent ductus arteriosus
PS	pulmonic stenosis
PTA	patent truncus arteriosus
RMA	robust multi-array average
RNA	ribonucleic acid
ROC	Receiver Operator Characteristic
rpm	rounds per minute
RR	relative risk
RSTS	Rubinstein-Taybi syndrome
TGA	transposition of the great arteries
ToF	tetralogy of Fallot
TTT	twin-to-twin transfusion
UTR	untranslated region
VACTERL	Vertebral anomalies, Anal atresia, Cardiovascular anomalies, Tracheo-esophageal fistula, Esophageal atresia, Renal and preaxial Limb anomalies
VSDs	ventricular septal defects
WISH	whole-mount in situ hybridisation
α-DIG-AP	anti–digoxigenin antibody conjugated to alkaline phosphatase

1. Schott, J.J., Benson, D.W., Basson, C.T., Pease, W., Silberbach, G.M., Moak, J.P., Maron, B.J., Seidman, C.E. & Seidman, J.G. Congenital heart disease caused by mutations in the transcription factor *NKX2-5*. *Science* **281**, 108-11 (1998).

2. Monserrat, L., Hermida-Prieto, M., Fernandez, X., Rodriguez, I., Dumont, C., Cazon, L., Cuesta, M.G., Gonzalez-Juanatey, C., Peteiro, J., Alvarez, N., Penas-Lado, M. & Castro-Beiras, A. Mutation in the alpha-cardiac actin gene associated with apical hypertrophic cardiomyopathy, left ventricular non-compaction, and septal defects. *Eur Heart J* **28**, 1953-61 (2007).

3. Matsson, H. et al. Alpha-cardiac actin mutations produce atrial septal defects. *Hum Mol Genet* **17**, 256-65 (2008).

4. Meberg, A., Hals, J. & Thaulow, E. Congenital heart defects--chromosomal anomalies, syndromes and extracardiac malformations. *Acta Paediatr* **96**, 1142-5 (2007).

5. Leppig, K.A., Werler, M.M., Cann, C.I., Cook, C.A. & Holmes, L.B. Predictive value of minor anomalies. I. Association with major malformations. *J Pediatr* **110**, 531-7 (1987).

6. Frias, J.L. & Carey, J.C. Mild errors of morphogenesis. *Adv Pediatr* **43**, 27-75 (1996).

7. Stephensen, S.S., Sigfusson, G., Eiriksson, H., Sverrisson, J.T., Torfason, B., Haraldsson, A. & Helgason, H. Congenital cardiac malformations in Iceland from 1990 through 1999. *Cardiol Young* **14**, 396-401 (2004).

8. Pradat, P., Francannet, C., Harris, J.A. & Robert, E. The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. *Pediatr Cardiol* **24**, 195-221 (2003).

9. Hoffman, J.I. & Kaplan, S. The incidence of congenital heart disease. *J Am Coll Cardiol* **39**, 1890-900 (2002).

10. Greenwood, R.D., Rosenthal, A., Parisi, L., Fyler, D.C. & Nadas, A.S. Extracardiac abnormalities in infants with congenital heart disease. *Pediatrics* **55**, 485-92 (1975).

11. Jenkins, K.J., Correa, A., Feinstein, J.A., Botto, L., Britt, A.E., Daniels, S.R., Elixson, M., Warnes, C.A. & Webb, C.L. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* **115**, 2995-3014 (2007).

12. Wren, J.D. & Garner, H.R. Data-mining analysis suggests an epigenetic pathogenesis for type 2 diabetes. *J Biomed Biotechnol* **2005**, 104-12 (2005).

13. Loffredo, C.A., Wilson, P.D. & Ferencz, C. Maternal diabetes: an independent risk factor for major cardiovascular malformations with increased mortality of affected infants. *Teratology* **64**, 98-106 (2001).

14. Manning, N. The influence of twinning on cardiac development. *Early Hum Dev* **84**, 173-9 (2008).

15. Manning, N. & Archer, N. A study to determine the incidence of structural congenital heart disease in monochorionic twins. *Prenat Diagn* **26**, 1062-4 (2006).

16. Bahtiyar, M.O., Dulay, A.T., Weeks, B.P., Friedman, A.H. & Copel, J.A. Prevalence of congenital heart defects in monochorionic/diamniotic twin gestations: a systematic literature review. *J Ultrasound Med* **26**, 1491-8 (2007).

17. Caputo, S., Russo, M.G., Capozzi, G., Morelli, C., Argiento, P., Di Salvo, G., Sarubbi, B., Santoro, G., Pacileo, G. & Calabro, R. Congenital heart disease in a population of dizygotic twins: an echocardiographic study. *Int J Cardiol* **102**, 293-6 (2005).

18. Bruder, C.E. et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* **82**, 763-71 (2008).

19. Kurnit, D.M., Layton, W.M. & Matthysse, S. Genetics, chance, and morphogenesis. *Am J Hum Genet* **41**, 979-95 (1987).

20. Samoilov, M.S., Price, G. & Arkin, A.P. From fluctuations to phenotypes: the physiology of noise. *Sci STKE* **2006**, re17 (2006).

21. Digilio, M.C., Calzolari, F., Capolino, R., Toscano, A., Sarkozy, A., de Zorzi, A., Dallapiccola, B. & Marino, B. Congenital heart defects in patients with oculo-auriculo-vertebral spectrum (Goldenhar syndrome). *Am J Med Genet A* **146A**, 1815-9 (2008).

22. Hartsfield, J.K. Review of the etiologic heterogeneity of the oculo-auriculo-vertebral spectrum (Hemifacial Microsomia). *Orthod Craniofac Res* **10**, 121-8 (2007).

23. Whittemore, R., Wells, J.A. & Castellsague, X. A second-generation study of 427 probands with congenital heart defects and their 837 children. *J Am Coll Cardiol* **23**, 1459-67 (1994).

24. Loffredo, C.A., Chokkalingam, A., Sill, A.M., Boughman, J.A., Clark, E.B., Scheel, J. & Brenner, J.I. Prevalence of congenital cardiovascular malformations among relatives of infants with hypoplastic left heart, coarctation of the aorta, and d-transposition of the great arteries. *Am J Med Genet A* **124A**, 225-30 (2004).

25. Gill, H.K., Splitt, M., Sharland, G.K. & Simpson, J.M. Patterns of recurrence of congenital heart disease: an analysis of 6,640 consecutive pregnancies evaluated by detailed fetal echocardiography. *J Am Coll Cardiol* **42**, 923-9 (2003).

26. Nora, J.J. & Nora, A.H. Recurrence risks in children having one parent with a congenital heart disease. *Circulation* **53**, 701-2 (1976).

27. Burn, J. et al. Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet* **351**, 311-6 (1998).

28. Drenthen, W., Pieper, P.G., Roos-Hesselink, J.W., van Lottum, W.A., Voors, A.A., Mulder, B.J., van Dijk, A.P., Vliegen, H.W., Yap, S.C., Moons, P., Ebels, T. & van Veldhuisen, D.J. Outcome of pregnancy in women with congenital heart disease: a literature review. *J Am Coll Cardiol* **49**, 2303-11 (2007).

29. Rose, V., Gold, R.J., Lindsay, G. & Allen, M. A possible increase in the incidence of congenital heart defects among the offspring of affected parents. *J Am Coll Cardiol* **6**, 376-82 (1985).

30. Driscoll, D.J., Michels, V.V., Gersony, W.M., Hayes, C.J., Keane, J.F., Kidd, L., Pieroni, D.R., Rings, L.J., Wolfe, R.R. & Weidman, W.H. Occurrence risk for congenital heart defects in relatives of patients with aortic stenosis, pulmonary stenosis, or ventricular septal defect. *Circulation* **87**, I114-20 (1993).

31. Boughman, J.A., Berg, K.A., Astemborski, J.A., Clark, E.B., McCarter, R.J., Rubin, J.D. & Ferencz, C. Familial risks of congenital heart defect assessed in a population-based epidemiologic study. *Am J Med Genet* **26**, 839-49 (1987).

32. Caputo, S., Capozzi, G., Russo, M.G., Esposito, T., Martina, L., Cardaropoli, D., Ricci, C., Argiento, P., Pacileo, G. & Calabro, R. Familial recurrence of congenital heart disease in patients with ostium secundum atrial septal defect. *Eur Heart J* **26**, 2179-84 (2005).

33. Drenthen, W., Pieper, P.G., van der Tuuk, K., Roos-Hesselink, J.W., Voors, A.A., Mostert, B., Mulder, B.J., Moons, P., Ebels, T. & van Veldhuisen, D.J. Cardiac complications relating to pregnancy and recurrence of disease in the offspring of women with atrioventricular septal defects. *Eur Heart J* **26**, 2581-7 (2005).

34. Emanuel, R., Somerville, J., Inns, A. & Withers, R. Evidence of congenital heart disease in the offspring of parents with atrioventricular defects. *Br Heart J* **49**, 144-7 (1983).

35. Ferencz, C., Boughman, J.A., Neill, C.A., Brenner, J.I. & Perry, L.W. Congenital cardiovascular malformations: questions on inheritance. Baltimore-Washington Infant Study Group. *J Am Coll Cardiol* **14**, 756-63 (1989).

36. Hanna, E.J., Nevin, N.C. & Nelson, J. Genetic study of congenital heart defects in Northern Ireland (1974-1978). *J Med Genet* **31**, 858-63 (1994).

37. Romano-Zelekha, O., Hirsh, R., Blieden, L., Green, M. & Shohat, T. The risk for congenital heart defects in offspring of individuals with congenital heart defects. *Clin Genet* **59**, 325-9 (2001).

38. Thangaroopan, M., Wald, R.M., Silversides, C.K., Mason, J., Smallhorn, J.F., Sermer, M., Colman, J.M. & Siu, S.C. Incremental diagnostic yield of pediatric cardiac assessment after

fetal echocardiography in the offspring of women with congenital heart disease: a prospective study. *Pediatrics* **121**, e660-5 (2008).

39. Brenner, J.I., Berg, K.A., Schneider, D.S., Clark, E.B. & Boughman, J.A. Cardiac malformations in relatives of infants with hypoplastic left-heart syndrome. *Am J Dis Child* **143**, 1492-4 (1989).

40. Pierpont, M.E., Gobel, J.W., Moller, J.H. & Edwards, J.E. Cardiac malformations in relatives of children with truncus arteriosus or interruption of the aortic arch. *Am J Cardiol* **61**, 423-7 (1988).

41. Dennis, N.R. & Warren, J. Risks to the offspring of patients with some common congenital heart defects. *J Med Genet* **18**, 8-16 (1981).

42. Piacentini, G., Digilio, M.C., Capolino, R., Zorzi, A.D., Toscano, A., Sarkozy, A., D'Agostino, R., Marasini, M., Russo, M.G., Dallapiccola, B. & Marino, B. Familial recurrence of heart defects in subjects with congenitally corrected transposition of the great arteries. *Am J Med Genet A* **137**, 176-80 (2005).

43. Digilio, M.C., Casey, B., Toscano, A., Calabro, R., Pacileo, G., Marasini, M., Banaudi, E., Giannotti, A., Dallapiccola, B. & Marino, B. Complete transposition of the great arteries: patterns of congenital heart disease in familial precurrence. *Circulation* **104**, 2809-14 (2001).

44. Becker, T.A., Van Amber, R., Moller, J.H. & Pierpont, M.E. Occurrence of cardiac malformations in relatives of children with transposition of the great arteries. *Am J Med Genet* **66**, 28-32 (1996).

45. Digilio, M.C., Marino, B., Giannotti, A., Toscano, A. & Dallapiccola, B. Recurrence risk figures for isolated tetralogy of Fallot after screening for 22q11 microdeletion. *J Med Genet* **34**, 188-90 (1997).

46. Zellers, T.M., Driscoll, D.J. & Michels, V.V. Prevalence of significant congenital heart defects in children of parents with Fallot's tetralogy. *Am J Cardiol* **65**, 523-6 (1990).

47. Goossens, D., Moens, L.N., Nelis, E., Lenaerts, A.S., Glassee, W., Kalbe, A., Frey, B., Kopal, G., Jonghe, P.D., Rijk, P.D. & Del-Favero, J. Simultaneous mutation and copy number variation (CNV) detection by multiplex PCR-based GS-FLX sequencing. *Hum Mutat* (2008).

48. Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L. & Quake, S.R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci U S A* **105**, 16266-71 (2008).

49. Freeman, S.B., Bean, L.H., Allen, E.G., Tinker, S.W., Locke, A.E., Druschel, C., Hobbs, C.A., Romitti, P.A., Royle, M.H., Torfs, C.P., Dooley, K.J. & Sherman, S.L. Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genet Med* **10**, 173-80 (2008).

50. Wieczorek, D., Krause, M., Majewski, F., Albrecht, B., Horn, D., Riess, O. & Gillessen-Kaesbach, G. Effect of the size of the deletion and clinical manifestation in Wolf-Hirschhorn syndrome: analysis of 13 patients with a de novo deletion. *Eur J Hum Genet* **8**, 519-26 (2000).

51. Battaglia, A., Carey, J.C., Cederholm, P., Viskochil, D.H., Brothman, A.R. & Galasso, C. Natural history of Wolf-Hirschhorn syndrome: experience with 15 cases. *Pediatrics* **103**, 830-6 (1999).

52. Maas, N.M., Van Buggenhout, G., Hannes, F., Thienpont, B., Sanlaville, D., Kok, K., Midro, A., Andrieux, J., Anderlid, B.M., Schoumans, J., Hordijk, R., Devriendt, K., Fryns, J.P. & Vermeesch, J.R. Genotype-phenotype correlation in 21 patients with Wolf-Hirschhorn syndrome using high resolution array comparative genome hybridisation (CGH). *J Med Genet* **45**, 71-80 (2008).

53. Ho, V.B., Bakalov, V.K., Cooley, M., Van, P.L., Hood, M.N., Burklow, T.R. & Bondy, C.A. Major vascular anomalies in Turner syndrome: prevalence and magnetic resonance angiographic features. *Circulation* **110**, 1694-700 (2004).

54. Sybert, V.P. Cardiovascular malformations and complications in Turner syndrome. *Pediatrics* **101**, E11 (1998).

55. Li, D.Y., Toland, A.E., Boak, B.B., Atkinson, D.L., Ensing, G.J., Morris, C.A. & Keating, M.T. Elastin point mutations cause an obstructive vascular disease, supravalvular aortic stenosis. *Hum Mol Genet* **6**, 1021-8 (1997).

56. Tassabehji, M., Metcalfe, K., Donnai, D., Hurst, J., Reardon, W., Burch, M. & Read, A.P. Elastin: genomic structure and point mutations in patients with supravalvular aortic stenosis. *Hum Mol Genet* **6**, 1029-36 (1997).

57. Wu, Y.Q., Sutton, V.R., Nickerson, E., Lupski, J.R., Potocki, L., Korenberg, J.R., Greenberg, F., Tassabehji, M. & Shaffer, L.G. Delineation of the common critical region in Williams syndrome and clinical correlation of growth, heart defects, ethnicity, and parental origin. *Am J Med Genet* **78**, 82-9 (1998).

58. Thienpont, B., Mertens, L., de Ravel, T., Eyskens, B., Boshoff, D., Maas, N., Fryns, J.P., Gewillig, M., Vermeesch, J.R. & Devriendt, K. Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur Heart J* **28**, 2778-84 (2007).

59. Portnoi, M.F., Lebas, F., Gruchy, N., Ardalan, A., Biran-Mucignat, V., Malan, V., Finkel, L., Roger, G., Ducrocq, S., Gold, F., Taillemite, J.L. & Marlin, S. 22q11.2 duplication syndrome: Two new familial cases with some overlapping features with DiGeorge/velocardiofacial syndromes. *Am J Med Genet A* **137**, 47-51 (2005).

60. Zweier, C., Sticht, H., Aydin-Yaylagul, I., Campbell, C.E. & Rauch, A. Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions. *Am J Hum Genet* **80**, 510-7 (2007).

61. Boyer-Di Ponio, J., Wright-Crosnier, C., Groyer-Picard, M.T., Driancourt, C., Beau, I., Hadchouel, M. & Meunier-Rotival, M. Biological function of mutant forms of JAGGED1 proteins in Alagille syndrome: inhibitory effect on Notch signaling. *Hum Mol Genet* **16**, 2683-92 (2007).

62. Satoda, M., Zhao, F., Diaz, G.A., Burn, J., Goodship, J., Davidson, H.R., Pierpont, M.E. & Gelb, B.D. Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. *Nat Genet* **25**, 42-6 (2000).

63. Zhu, L., Bonnet, D., Boussion, M., Vedie, B., Sidi, D. & Jeunemaitre, X. Investigation of the MYH11 gene in sporadic patients with an isolated persistently patent arterial duct. *Cardiol Young* **17**, 666-72 (2007).

64. Ching, Y.H. et al. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet* **37**, 423-8 (2005).

65. Bruneau, B.G. The developmental genetics of congenital heart disease. *Nature* **451**, 943-8 (2008).

66. Aoki, Y., Niihori, T., Narumi, Y., Kure, S. & Matsubara, Y. The RAS/MAPK syndromes: novel roles of the RAS pathway in human genetic disorders. *Hum Mutat* **29**, 992-1006 (2008).

67. Denayer, E., de Ravel, T. & Legius, E. Clinical and Molecular Aspects of RAS-related disorders. *J Med Genet* (2008).

68. Postma, A.V., van de Meerakker, J.B., Mathijssen, I.B., Barnett, P., Christoffels, V.M., Ilgun, A., Lam, J., Wilde, A.A., Lekanne Deprez, R.H. & Moorman, A.F. A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circ Res* **102**, 1433-42 (2008).

69. Nabulsi, M.M., Tamim, H., Sabbagh, M., Obeid, M.Y., Yunis, K.A. & Bitar, F.F. Parental consanguinity and congenital heart malformations in a developing country. *Am J Med Genet A* **116A**, 342-7 (2003).

70. Yunis, K., Mumtaz, G., Bitar, F., Chamseddine, F., Kassar, M., Rashkidi, J., Makhoul, G. & Tamim, H. Consanguineous marriage and congenital heart defects: a case-control study in the neonatal period. *Am J Med Genet A* **140**, 1524-30 (2006).

71. Becker, S.M., Al Halees, Z., Molina, C. & Paterson, R.M. Consanguinity and congenital heart disease in Saudi Arabia. *Am J Med Genet* **99**, 8-13 (2001).

72. Chehab, G., Chedid, P., Saliba, Z. & Bouvagnet, P. Congenital cardiac disease and inbreeding: specific defects escape higher risk due to parental consanguinity. *Cardiol Young* **17**, 414-22 (2007).

73. Heathcote, K., Braybrook, C., Abushaban, L., Guy, M., Khetyar, M.E., Patton, M.A., Carter, N.D., Scambler, P.J. & Syrris, P. Common arterial trunk associated with a homeodomain mutation of NKX2.6. *Hum Mol Genet* **14**, 585-93 (2005).

74. Abushaban, L., Uthaman, B., Kumar, A.R. & Selvan, J. Familial truncus arteriosus: a possible autosomal-recessive trait. *Pediatr Cardiol* **24**, 64-6 (2003).

75. So, J. et al. Mild phenotypes in a series of patients with Opitz GBBB syndrome with MID1 mutations. *Am J Med Genet A* **132A**, 1-7 (2005).

76. Badens, C., Martini, N., Courrier, S., DesPortes, V., Touraine, R., Levy, N. & Edery, P. ATRX syndrome in a girl with a heterozygous mutation in the ATRX Zn finger domain and a totally skewed X-inactivation pattern. *Am J Med Genet A* **140**, 2212-5 (2006).

77. Erdogan, F., Larsen, L.A., Zhang, L., Tumer, Z., Tommerup, N., Chen, W., Jacobsen, J.R., Schubert, M., Jurkatis, J., Tzschach, A., Ropers, H.H. & Ullmann, R. High frequency of submicroscopic genomic aberrations detected by tiling path array CGH in patients with isolated congenital heart disease. *J Med Genet* (2008).

78. Posch, M.G., Perrot, A., Schmitt, K., Mittelhaus, S., Esenwein, E.M., Stiller, B., Geier, C., Dietz, R., Gessner, R., Ozcelik, C. & Berger, F. Mutations in GATA4, NKX2.5, CRELD1, and BMP4 are infrequently found in patients with congenital cardiac septal defects. *Am J Med Genet A* **146A**, 251-3 (2008).

79. Zlotogora, J. Penetrance and expressivity in the molecular age. *Genet Med* **5**, 347-52 (2003).

80. Badano, J.L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* **3**, 779-89 (2002).

81. Kobrynski, L.J. & Sullivan, K.E. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* **370**, 1443-52 (2007).

82. Stalmans, I. et al. VEGF: a modifier of the del22q11 (DiGeorge) syndrome? *Nat Med* **9**, 173-82 (2003).

83. Maslen, C.L., Babcock, D., Robinson, S.W., Bean, L.J., Dooley, K.J., Willour, V.L. & Sherman, S.L. CRELD1 mutations contribute to the occurrence of cardiac atrioventricular septal defects in Down syndrome. *Am J Med Genet A* **140**, 2501-5 (2006).

84. Roessler, E., Ouspenskaia, M.V., Karkera, J.D., Velez, J.I., Kantipong, A., Lacbawan, F., Bowers, P., Belmont, J.W., Towbin, J.A., Goldmuntz, E., Feldman, B. & Muenke, M. Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *Am J Hum Genet* **83**, 18-29 (2008).

85. Peeters, H. & Devriendt, K. Human laterality disorders. *Eur J Med Genet* **49**, 349-62 (2006).

86. Lambrechts, D., Devriendt, K., Driscoll, D.A., Goldmuntz, E., Gewillig, M., Vlietinck, R., Collen, D. & Carmeliet, P. Low expression VEGF haplotype increases the risk for tetralogy of Fallot: a family based association study. *J Med Genet* **42**, 519-22 (2005).

87. Rauch, A., Devriendt, K., Koch, A., Rauch, R., Gewillig, M., Kraus, C., Weyand, M., Singer, H., Reis, A. & Hofbeck, M. Assessment of association between variants and haplotypes of the remaining TBX1 gene and manifestations of congenital heart defects in 22q11.2 deletion patients. *J Med Genet* **41**, e40 (2004).

88. van Driel, L.M., Smedts, H.P., Helbing, W.A., Isaacs, A., Lindemans, J., Uitterlinden, A.G., van Duijn, C.M., de Vries, J.H., Steegers, E.A. & Steegers-Theunissen, R.P. Eight-fold increased risk for congenital heart defects in children carrying the nicotinamide N-methyltransferase polymorphism and exposed to medicines and low nicotinamide. *Eur Heart J* **29**, 1424-31 (2008).

89. Menten, B. et al. Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J Med Genet* **43**, 625-33 (2006).

90.   Krzywinski, M. et al. A set of BAC clones spanning the human genome. *Nucleic Acids Res* **32**, 3651-60 (2004).

91.   Vermeesch, J.R., Melotte, C., Froyen, G., Van Vooren, S., Dutta, B., Maas, N., Vermeulen, S., Menten, B., Speleman, F., De Moor, B., Van Hummelen, P., Marynen, P., Fryns, J.P. & Devriendt, K. Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis. *J Histochem Cytochem* **53**, 413-22 (2005).

92.   Bauters, M., Van Esch, H., Marynen, P. & Froyen, G. X chromosome array-CGH for the identification of novel X-linked mental retardation genes. *Eur J Med Genet* **48**, 263-75 (2005).

93.   Backx, L., Thoelen, R., Van Esch, H. & Vermeesch, J.R. Direct fluorescent labelling of clones by DOP PCR. *Mol Cytogenet* **1**, 3 (2008).

94.   Petrij, F. et al. Diagnostic analysis of the Rubinstein-Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations. *J Med Genet* **37**, 168-76 (2000).

95.   Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-8 (2001).

96.   Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M. & Belmont, J.W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* **51**, 1229-39 (1992).

97.   Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).

98.   McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J., Chapman, L., Rhodes, D. & Higgs, D.R. Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**, 13983-8 (1999).

99.   Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* **3**, 59-69 (2008).

100.   Huang, C.J., Tu, C.T., Hsiao, C.D., Hsieh, F.J. & Tsai, H.J. Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev Dyn* **228**, 30-40 (2003).

101.   Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**, 253-310 (1995).

102.   Karlsson, J., von Hofsten, J. & Olsson, P.E. Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. *Mar Biotechnol (NY)* **3**, 522-7 (2001).

103.   Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes, Chromosomes & Cancer* **20**, 399-407 (1997).

104.   Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W. & Albertson, D.G. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**, 207-11 (1998).

105.   Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C. & Vetrie, D. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* **76**, 750-62 (2005).

106.   Barrett, M.T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L. & Laderman, S. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* **101**, 17765-70 (2004).

107. Le Caignec, C., Spits, C., Sermon, K., De Rycke, M., Thienpont, B., Debrock, S., Staessen, C., Moreau, Y., Fryns, J.P., Van Steirteghem, A., Liebaers, I. & Vermeesch, J.R. Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Res* **34**, e68 (2006).

108. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. & Lee, C. Detection of large-scale variation in the human genome. *Nat Genet* **36**, 949-51 (2004).

109. Tatton-Brown, K. et al. Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with *NSD1* aberrations. *Am J Hum Genet* **77**, 193-204 (2005).

110. Tatton-Brown, K. et al. Multiple mechanisms are implicated in the generation of 5q35 microdeletions in Sotos syndrome. *J Med Genet* **42**, 307-13 (2005).

111. Rauch, A., Beese, M., Mayatepek, E., Dorr, H.G., Wenzel, D., Reis, A. & Trautmann, U. A novel 5q35.3 subtelomeric deletion syndrome. *Am J Med Genet A* **121**, 1-8 (2003).

112. McCandless, S.E., Schwartz, S., Morrison, S., Garlapati, K. & Robin, N.H. Adult with an interstitial deletion of chromosome 10 [del(10)(q25. 1q25.3)]: overlap with Coffin-Lowry syndrome. *Am J Med Genet* **95**, 93-8 (2000).

113. Waggoner, D.J., Chow, C.K., Dowton, S.B. & Watson, M.S. Partial monosomy of distal 10q: three new cases and a review. *Am J Med Genet* **86**, 1-5 (1999).

114. Haase, D. Cytogenetic features in myelodysplastic syndromes. *Ann Hematol* **87**, 515-26 (2008).

115. Pianigiani, E., De Aloe, G., Andreassi, A., Rubegni, P. & Fimiani, M. Rothmund-Thomson syndrome (Thomson-type) and myelodysplasia. *Pediatr Dermatol* **18**, 422-5 (2001).

116. Yobb, T.M. et al. Microduplication and triplication of 22q11.2: a highly variable syndrome. *Am J Hum Genet* **76**, 865-76 (2005).

117. Nagai, T. et al. Sotos syndrome and haploinsufficiency of *NSD1*: clinical features of intragenic mutations and submicroscopic deletions. *J Med Genet* **40**, 285-9 (2003).

118. Garg, V., Muth, A.N., Ransom, J.F., Schluterman, M.K., Barnes, R., King, I.N., Grossfeld, P.D. & Srivastava, D. Mutations in *NOTCH1* cause aortic valve disease. *Nature* **437**, 270-4 (2005).

119. Kleefstra, T., Brunner, H.G., Amiel, J., Oudakker, A.R., Nillesen, W.M., Magee, A., Genevieve, D., Cormier-Daire, V., van Esch, H., Fryns, J.P., Hamel, B.C., Sistermans, E.A., de Vries, B.B. & van Bokhoven, H. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am J Hum Genet* **79**, 370-7 (2006).

120. Petrij, F., Giles, R.H., Dauwerse, H.G., Saris, J.J., Hennekam, R.C., Masuno, M., Tommerup, N., van Ommen, G.J., Goodman, R.H., Peters, D.J. & et al. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* **376**, 348-51 (1995).

121. Gibbons, R.J., Picketts, D.J., Villard, L. & Higgs, D.R. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* **80**, 837-45 (1995).

122. Ragge, N.K. et al. Heterozygous mutations of OTX2 cause severe ocular malformations. *Am J Hum Genet* **76**, 1008-22 (2005).

123. Locke, D.P., Sharp, A.J., McCarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M. & Eichler, E.E. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am J Hum Genet* **79**, in press (2006).

124. van Ommen, G.J. Frequency of new copy number variation in humans. *Nat Genet* **37**, 333-4 (2005).

125. Thienpont, B., de Ravel, T., Van Esch, H., Van Schoubroeck, D., Moerman, P., Vermeesch, J.R., Fryns, J.P., Froyen, G., Lacoste, C., Badens, C. & Devriendt, K. Partial duplications of the ATRX gene cause the ATR-X syndrome. *Eur J Hum Genet* **15**, 1094-7 (2007).

126. White, S., Kalf, M., Liu, Q., Villerius, M., Engelsma, D., Kriek, M., Vollebregt, E., Bakker, B., van Ommen, G.J., Breuning, M.H. & den Dunnen, J.T. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. *Am J Hum Genet* **71**, 365-74 (2002).

127. de Vries, B.A.B. et al. Diagnostic Genome Profiling in Mental Retardation. *Am J Hum Genet* **77**, 606-616 (2005).

128. Redon, R. et al. Global variation in copy number in the human genome. *Nature* **444**, 444-54 (2006).

129. Balikova, I. et al. Subtelomeric imbalances in phenotypically normal individuals. *Hum Mutat* **28**, 958-67 (2007).

130. Thienpont, B., Breckpot, J., Holvoet, M., Vermeesch, J.R. & Devriendt, K. A microduplication of CBP in a patient with mental retardation and a congenital heart defect. *Am J Med Genet A* **143**, 2160-4 (2007).

131. Sebat, J. et al. Strong association of de novo copy number mutations with autism. *Science* **316**, 445-9 (2007).

132. Battaglia, A., Hoyme, H.E., Dallapiccola, B., Zackai, E., Hudgins, L., McDonald-McGinn, D., Bahi-Buisson, N., Romano, C., Williams, C.A., Brailey, L.L., Zuberi, S.M. & Carey, J.C. Further delineation of deletion 1p36 syndrome in 60 patients: a recognizable phenotype and common cause of developmental delay and mental retardation. *Pediatrics* **121**, 404-10 (2008).

133. Huang, T., Lin, A.E., Cox, G.F., Golden, W.L., Feldman, G.L., Ute, M., Schrander-Stumpel, C., Kamisago, M. & Vermeulen, S.J. Cardiac phenotypes in chromosome 4q- syndrome with and without a deletion of the dHAND gene. *Genet Med* **4**, 464-7 (2002).

134. Garcia-Minaur, S., Ramsay, J., Grace, E., Minns, R.A., Myles, L.M. & FitzPatrick, D.R. Interstitial deletion of the long arm of chromosome 5 in a boy with multiple congenital anomalies and mental retardation: Molecular characterization of the deleted region to 5q22.3q23.3. *Am J Med Genet A* **132**, 402-10 (2005).

135. Baekvad-Hansen, M., Tumer, Z., Delicado, A., Erdogan, F., Tommerup, N. & Larsen, L.A. Delineation of a 2.2 Mb microdeletion at 5q35 associated with microcephaly and congenital heart disease. *Am J Med Genet A* **140**, 427-33 (2006).

136. Iwakoshi, M., Okamoto, N., Harada, N., Nakamura, T., Yamamori, S., Fujita, H., Niikawa, N. & Matsumoto, N. 9q34.3 deletion syndrome in three unrelated children. *Am J Med Genet A* **126**, 278-83 (2004).

137. Brown, S., Gersen, S., Anyane-Yeboa, K. & Warburton, D. Preliminary definition of a "critical region" of chromosome 13 in q32: report of 14 cases with 13q deletions and review of the literature. *Am J Med Genet* **45**, 52-9 (1993).

138. Lemyre, E., Lemieux, N., Decarie, J.C. & Lambert, M. Del(14)(q22.1q23.2) in a patient with anophthalmia and pituitary hypoplasia. *Am J Med Genet* **77**, 162-5 (1998).

139. Marangi, G., Leuzzi, V., Orteschi, D., Grimaldi, M.E., Lecce, R., Neri, G. & Zollino, M. Duplication of the Rubinstein-Taybi region on 16p13.3 is associated with a distinctive phenotype. *Am J Med Genet A* **146A**, 2313-7 (2008).

140. Lybaek, H., Meza-Zepeda, L.A., Kresse, S.H., Hoysaeter, T., Steen, V.M. & Houge, G. Array-CGH fine mapping of minor and cryptic HR-CGH detected genomic imbalances in 80 out of 590 patients with abnormal development. *Eur J Hum Genet* **16**, 1318-28 (2008).

141. Hannes, F.D. et al. 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. *J Med Genet* (2008).

142. Mefford, H.C. et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* **359**, 1685-99 (2008).

143. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-41 (2008).

144. Ullmann, R. et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum Mutat* **28**, 674-82 (2007).

145. Stefansson, H. et al. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-6 (2008).

146. Brunetti-Pierri, N. et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* **40**, 1466-71 (2008).

147. Christiansen, J., Dyck, J.D., Elyas, B.G., Lilley, M., Bamforth, J.S., Hicks, M., Sprysak, K.A., Tomaszewski, R., Haase, S.M., Vicen-Wyhony, L.M. & Somerville, M.J. Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ Res* **94**, 1429-35 (2004).

148. Weiss, L.A. et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**, 667-75 (2008).

149. Kruger, O., Maxeiner, S., Kim, J.S., van Rijen, H.V., de Bakker, J.M., Eckardt, D., Tiemann, K., Lewalter, T., Ghanem, A., Luderitz, B. & Willecke, K. Cardiac morphogenetic defects and conduction abnormalities in mice homozygously deficient for connexin40 and heterozygously deficient for connexin45. *J Mol Cell Cardiol* **41**, 787-97 (2006).

150. Simon, A.M., McWhorter, A.R., Dones, J.A., Jackson, C.L. & Chen, H. Heart and head defects in mice lacking pairs of connexins. *Dev Biol* **265**, 369-83 (2004).

151. Thienpont, B., Mertens, L., Buyse, G., Vermeesch, J.R. & Devriendt, K. Left-ventricular non-compaction in a patient with monosomy 1p36. *Eur J Med Genet* **50**, 233-6 (2007).

152. Thienpont, B., Breckpot, J., Vermeesch, J.R., Gewillig, M. & Devriendt, K. A complex submicroscopic chromosomal imbalance in 19p13.11 with one microduplication and two microtriplications. *Eur J Med Genet* **51**, 219-25 (2008).

153. Thienpont, B., Gewillig, M., Fryns, J.P., Devriendt, K. & Vermeesch, J.R. Molecular cytogenetic characterization of a constitutional complex intrachromosomal 4q rearrangement in a patient with multiple congenital anomalies. *Cytogenet Genome Res* **114**, 338-341 (2006).

154. Maas, N., Thienpont, B., Vermeesch, J.R. & Fryns, J.P. Facial asymmetry, cardiovascular anomalies and adducted thumbs as unusual symptoms in Dubowitz syndrome: a microdeletion/duplication in 13q. *Genet Couns* **17**, 477-9 (2006).

155. Gibbons, R. Alpha thalassaemia-mental retardation, X linked. *Orphanet J Rare Dis* **1**, 15 (2006).

156. Abidi, F.E., Cardoso, C., Lossi, A.M., Lowry, R.B., Depetris, D., Mattei, M.G., Lubs, H.A., Stevenson, R.E., Fontes, M., Chudley, A.E. & Schwartz, C.E. Mutation in the 5' alternatively spliced region of the XNP/ATR-X gene causes Chudley-Lowry syndrome. *Eur J Hum Genet* **13**, 176-83 (2005).

157. White, S.J., Aartsma-Rus, A., Flanigan, K.M., Weiss, R.B., Kneppers, A.L., Lalic, T., Janson, A.A., Ginjaar, H.B., Breuning, M.H. & den Dunnen, J.T. Duplications in the DMD gene. *Hum Mutat* **27**, 938-945 (2006).

158. Kozlowski, P., Roberts, P., Dabora, S., Franz, D., Bissler, J., Northrup, H., Au, K.S., Lazarus, R., Domanska-Pakiela, D., Kotulska, K., Jozwiak, S. & Kwiatkowski, D.J. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum Genet* (2007).

159. Wimmer, K., Yao, S., Claes, K., Kehrer-Sawatzki, H., Tinschert, S., De Raedt, T., Legius, E., Callens, T., Beiglbock, H., Maertens, O. & Messiaen, L. Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes Cancer* **45**, 265-76 (2006).

160. Ferec, C. et al. Gross genomic rearrangements involving deletions in the CFTR gene: characterization of six new events from a large cohort of hitherto unidentified cystic fibrosis chromosomes and meta-analysis of the underlying mechanisms. *Eur J Hum Genet* **14**, 567-76 (2006).

161. Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gecz, J., Moraine, C., Marynen, P., Fryns, J.P. & Froyen, G. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet* **77**, 442-53 (2005).

162. Van Buggenhout, G., Van Ravenswaaij-Arts, C., Mc Maas, N., Thoelen, R., Vogels, A., Smeets, D., Salden, I., Matthijs, G., Fryns, J.P. & Vermeesch, J.R. The del(2)(q32.2q33) deletion syndrome defined by clinical and molecular characterization of four patients. *Eur J Med Genet* **48**, 276-89 (2005).

163. Heilstedt, H.A., Ballif, B.C., Howard, L.A., Kashork, C.D. & Shaffer, L.G. Population data suggest that deletions of 1p36 are a relatively common chromosome abnormality. *Clin Genet* **64**, 310-6 (2003).

164. Heilstedt, H.A., Ballif, B.C., Howard, L.A., Lewis, R.A., Stal, S., Kashork, C.D., Bacino, C.A., Shapira, S.K. & Shaffer, L.G. Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome. *Am J Hum Genet* **72**, 1200-12 (2003).

165. De Rosa, G., Pardeo, M., Bria, S., Caresta, E., Vasta, I., Zampino, G., Zollino, M., Zuppa, A.A. & Piastra, M. Isolated myocardial non-compaction in an infant with distal 4q trisomy and distal 1q monosomy. *Eur J Pediatr* **164**, 255-6 (2005).

166. Kanemoto, N., Horigome, H., Nakayama, J., Ichida, F., Xing, Y., Buonadonna, A.L., Kanemoto, K. & Gentile, M. Interstitial 1q43-q43 deletion with left ventricular noncompaction myocardium. *Eur J Med Genet* **49**, 247-53 (2006).

167. Pauli, R.M., Scheib-Wixted, S., Cripe, L., Izumo, S. & Sekhon, G.S. Ventricular noncompaction and distal chromosome 5q deletion. *Am J Med Genet* **85**, 419-23 (1999).

168. McMahon, C.J., Chang, A.C., Pignatelli, R.H., Miller-Hance, W.C., Eble, B.K., Towbin, J.A. & Denfield, S.W. Left ventricular noncompaction cardiomyopathy in association with trisomy 13. *Pediatr Cardiol* **26**, 477-9 (2005).

169. Finsterer, J., Stollberger, C. & Blazek, G. Neuromuscular implications in left ventricular hypertrabeculation/noncompaction. *Int J Cardiol* **110**, 288-300 (2006).

170. Freedom, R.M., Yoo, S.J., Perrin, D., Taylor, G., Petersen, S. & Anderson, R.H. The morphological spectrum of ventricular noncompaction. *Cardiol Young* **15**, 345-64 (2005).

171. Sasse-Klaassen, S., Probst, S., Gerull, B., Oechslin, E., Nurnberg, P., Heuser, A., Jenni, R., Hennies, H.C. & Thierfelder, L. Novel gene locus for autosomal dominant left ventricular noncompaction maps to chromosome 11p15. *Circulation* **109**, 2720-3 (2004).

172. Saito, S., Kawamura, R., Kosho, T., Shimizu, T., Aoyama, K., Koike, K., Wada, T., Matsumoto, N., Kato, M., Wakui, K. & Fukushima, Y. Bilateral perisylvian polymicrogyria, periventricular nodular heterotopia, and left ventricular noncompaction in a girl with 10.5-11.1 Mb terminal deletion of 1p36. *Am J Med Genet A* **146A**, 2891-2897 (2008).

173. Cremer, K., Ludecke, H.J., Ruhr, F. & Wieczorek, D. Left-ventricular non-compaction (LVNC): A clinical feature more often observed in terminal deletion 1p36 than previously expected. *Eur J Med Genet* (2008).

174. Devriendt, K. & Vermeesch, J.R. Chromosomal phenotypes and submicroscopic abnormalities. *Hum Genomics* **1**, 126-33 (2004).

175. Chance, P.F., Alderson, M.K., Leppig, K.A., Lensch, M.W., Matsunami, N., Smith, B., Swanson, P.D., Odelberg, S.J., Disteche, C.M. & Bird, T.D. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72**, 143-51 (1993).

176. Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., Chakravarti, A. & Patel, P.I. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* **66**, 219-32 (1991).

177. Ensenauer, R.E. et al. Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. *Am J Hum Genet* **73**, 1027-40 (2003).

178. Somerville, M.J., Mervis, C.B., Young, E.J., Seo, E.J., del Campo, M., Bamforth, S., Peregrine, E., Loo, W., Lilley, M., Perez-Jurado, L.A., Morris, C.A., Scherer, S.W. & Osborne, L.R. Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N Engl J Med* **353**, 1694-701 (2005).

179. Potocki, L. et al. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am J Hum Genet* **80**, 633-49 (2007).

180. Baris, H., Bejjani, B.A., Tan, W.H., Coulter, D.L., Martin, J.A., Storm, A.L., Burton, B.K., Saitta, S.C., Gajecka, M., Ballif, B.C., Irons, M.B., Shaffer, L.G. & Kimonis, V.E. Identification of a novel polymorphism--the duplication of the NPHP1 (nephronophthisis 1) gene. *Am J Med Genet A* **140**, 1876-9 (2006).

181. Sharp, A.J. et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* **38**, 1038-42 (2006).

182. Coupry, I., Roudaut, C., Stef, M., Delrue, M.A., Marche, M., Burgelin, I., Taine, L., Cruaud, C., Lacombe, D. & Arveiler, B. Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome. *J Med Genet* **39**, 415-21 (2002).

183. Roelfsema, J.H., White, S.J., Ariyurek, Y., Bartholdi, D., Niedrist, D., Papadia, F., Bacino, C.A., den Dunnen, J.T., van Ommen, G.J., Breuning, M.H., Hennekam, R.C. & Peters, D.J. Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am J Hum Genet* **76**, 572-80 (2005).

184. Breuning, M.H., Dauwerse, H.G., Fugazza, G., Saris, J.J., Spruit, L., Wijnen, H., Tommerup, N., van der Hagen, C.B., Imaizumi, K., Kuroki, Y. & et al. Rubinstein-Taybi syndrome caused by submicroscopic deletions within 16p13.3. *Am J Hum Genet* **52**, 249-54 (1993).

185. Hennekam, R.C., Stevens, C.A. & Van de Kamp, J.J. Etiology and recurrence risk in Rubinstein-Taybi syndrome. *Am J Med Genet Suppl* **6**, 56-64 (1990).

186. Hennekam, R.C. Rubinstein-Taybi syndrome. *Eur J Hum Genet* **14**, 981-5 (2006).

187. Rubinstein, J.H. & Taybi, H. Broad thumbs and toes and facial abnormalities. A possible mental retardation syndrome. *Am J Dis Child* **105**, 588-608 (1963).

188. Kokalj-Vokac, N., Medica, I., Zagorac, A., Zagradisnik, B., Erjavec, A. & Gregoric, A. A case of insertional translocation resulting in partial trisomy 16p. *Ann Genet* **43**, 131-5 (2000).

189. Ravnan, J.B., Tepperberg, J.H., Papenhausen, P., Lamb, A.N., Hedrick, J., Eash, D., Ledbetter, D.H. & Martin, C.L. Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J Med Genet* **43**, 478-89 (2006).

190. Wong, A., Lese Martin, C., Heretis, K., Ruffalo, T., Wilber, K., King, W. & Ledbetter, D.H. Detection and calibration of microdeletions and microduplications by array-based comparative genomic hybridization and its applicability to clinical genetic testing. *Genet Med* **7**, 264-71 (2005).

191. Martin, C.L., Waggoner, D.J., Wong, A., Uhrig, S., Roseberry, J.A., Hedrick, J.F., Pack, S.D., Russell, K., Zackai, E., Dobyns, W.B. & Ledbetter, D.H. "Molecular rulers" for calibrating phenotypic effects of telomere imbalance. *J Med Genet* **39**, 734-40 (2002).

192. de Ravel, T., Aerssens, P., Vermeesch, J.R. & Fryns, J.P. Trisomy of chromosome 16p13.3 due to an unbalanced insertional translocation into chromosome 22p13. *Eur J Med Genet* **48**, 355-9 (2005).

193. Friedman, J.M. et al. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* **79**, 500-13 (2006).

194. Hennekam, R.C., Tilanus, M., Hamel, B.C., Voshart-van Heeren, H., Mariman, E.C., van Beersum, S.E., van den Boogaard, M.J. & Breuning, M.H. Deletion at chromosome 16p13.3 as a cause of Rubinstein-Taybi syndrome: clinical aspects. *Am J Hum Genet* **52**, 255-62 (1993).

195. Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953-9 (1996).

196. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. & Zoghbi, H.Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185-8 (1999).

197. Houge, G., Liehr, T., Schoumans, J., Ness, G.O., Solland, K., Starke, H., Claussen, U., Stromme, P., Akre, B. & Vermeulen, S. Ten years follow up of a boy with a complex chromosomal rearrangement: going from a > 5 to 15-breakpoint CCR. *Am J Med Genet A* **118**, 235-40 (2003).

198. Batanian, J.R. & Eswara, M.S. De novo apparently balanced complex chromosome rearrangement (CCR) involving chromosomes 4, 18, and 21 in a girl with mental retardation: report and review. *Am J Med Genet* **78**, 44-51 (1998).

199. Tonk, V.S., Wilson, G.N., Yatsenko, S.A., Stankiewicz, P., Lupski, J.R., Schutt, R.C., Northup, J.K. & Velagaleti, G.V. Molecular cytogenetic characterization of a familial der(1)del(1)(p36.33)dup(1)(p36.33p36.22) with variable phenotype. *Am J Med Genet A* **139**, 136-40 (2005).

200. Ballif, B.C., Yu, W., Shaw, C.A., Kashork, C.D. & Shaffer, L.G. Monosomy 1p36 breakpoint junctions suggest pre-meiotic breakage-fusion-bridge cycles are involved in generating terminal deletions. *Hum Mol Genet* **12**, 2153-65 (2003).

201. Van Esch, H., Syrrou, M. & Lagae, L. Refractory photosensitive epilepsy associated with a complex rearrangement of chromosome 2. *Neuropediatrics* **33**, 320-3 (2002).

202. Shim, S.H., Wyandt, H.E., McDonald-McGinn, D.M., Zackai, E.Z. & Milunsky, A. Molecular cytogenetic characterization of multiple intrachromosomal rearrangements of chromosome 2q in a patient with Waardenburg's syndrome and other congenital defects. *Clin Genet* **66**, 46-52 (2004).

203. Romain, D.R., Columbano-Green, L.M., Parfitt, R.G., Chapman, C.J., Smythe, R.H. & Gebbie, O.B. A complex structural rearrangement of chromosome 4 in a woman without phenotypic features of Wolf-Hirschhorn syndrome. *Clin Genet* **28**, 166-72 (1985).

204. Tuck-Muller, C.M., Varela, M., Li, S., Pridjian, G., Chen, H. & Wertelecki, W. A complex five breakpoint intrachromosomal rearrangement ascertained through two recombinant offspring. *Am J Med Genet* **63**, 392-5 (1996).

205. Weise, A., Rittinger, O., Starke, H., Ziegler, M., Claussen, U. & Liehr, T. De novo 9-break-event in one chromosome 21 combined with a microdeletion in 21q22.11 in a mentally retarded boy with short stature. *Cytogenet Genome Res* **103**, 14-6 (2003).

206. Stankiewicz, P. & Beaudet, A.L. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. *Curr Opin Genet Dev* (2007).

207. Giorda, R., Ciccone, R., Gimelli, G., Pramparo, T., Beri, S., Bonaglia, M.C., Giglio, S., Genuardi, M., Argente, J., Rocchi, M. & Zuffardi, O. Two classes of low-copy repeats comediate a new recurrent rearrangement consisting of duplication at 8p23.1 and triplication at 8p23.2. *Hum Mutat* **28**, 459-68 (2007).

208. del Gaudio, D. et al. Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet Med* **8**, 784-92 (2006).

209. Vissers, L.E., Stankiewicz, P., Yatsenko, S.A., Crawford, E., Creswick, H., Proud, V.K., de Vries, B.B., Pfundt, R., Marcelis, C.L., Zackowski, J., Bi, W., van Kessel, A.G., Lupski, J.R. & Veltman, J.A. Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies. *Hum Genet* **121**, 697-709 (2007).

210. Manolakos, E., Kosyakova, N., Thomaidis, L., Neroutsou, R., Weise, A., Mihalatos, M., Orru, S., Kokotas, H., Kitsos, G., Liehr, T. & Petersen, M.B. Complex chromosome rearrangement in a child with microcephaly, dysmorphic facial features and mosaicism for a terminal deletion del(18)(q21.32-qter) investigated by FISH and array-CGH: Case report. *Mol Cytogenet* **1**, 24 (2008).

211. Chen, E., Obolensky, E., Rauen, K.A., Shaffer, L.G. & Li, X. Cytogenetic and array CGH characterization of de novo 1p36 duplications and deletion in a patient with congenital cataracts, hearing loss, choanal atresia, and mental retardation. *Am J Med Genet A* **146A**, 2785-90 (2008).

212. Girirajan, S., Williams, S.R., Garbern, J.Y., Nowak, N., Hatchwell, E. & Elsea, S.H. 17p11.2p12 triplication and del(17)q11.2q12 in a severely affected child with dup(17)p11.2p12 syndrome. *Clin Genet* **72**, 47-58 (2007).

213. Gijsbers, A.C., Bijlsma, E.K., Weiss, M.M., Bakker, E., Breuning, M.H., Hoffer, M.J. & Ruivenkamp, C.A. A 400kb duplication, 2.4Mb triplication and 130kb duplication of 9q34.3 in a patient with severe mental retardation. *Eur J Med Genet* **51**, 479-87 (2008).

214. Bernardini, L., Castori, M., Capalbo, A., Mokini, V., Mingarelli, R., Simi, P., Bertuccelli, A., Novelli, A. & Dallapiccola, B. Syndromic craniosynostosis due to complex chromosome 5 rearrangement and MSX2 gene triplication. *Am J Med Genet A* **143A**, 2937-43 (2007).

215. Lindstrand, A., Malmgren, H., Sahlen, S., Xin, H., Schoumans, J. & Blennow, E. Molecular cytogenetic characterization of a constitutional, highly complex intrachromosomal rearrangement of chromosome 1, with 14 breakpoints and a 0.5 Mb submicroscopic deletion. *Am J Med Genet A* **146A**, 3217-22 (2008).

216. Bernardini, L., Palka, C., Ceccarini, C., Capalbo, A., Bottillo, I., Mingarelli, R., Novelli, A. & Dallapiccola, B. Complex rearrangement of chromosomes 7q21.13-q22.1 confirms the ectrodactyly-deafness locus and suggests new candidate genes. *Am J Med Genet A* **146A**, 238-44 (2008).

217. Gajecka, M., Glotzbach, C.D. & Shaffer, L.G. Characterization of a complex rearrangement with interstitial deletions and inversion on human chromosome 1. *Chromosome Res* **14**, 277-82 (2006).

218. Gribble, S.M. et al. The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* **42**, 8-16 (2005).

219. De Gregori, M. et al. Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 cases. *J Med Genet* (2007).

220. Shaw, C.J. & Lupski, J.R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* **13 Spec No 1**, R57-64 (2004).

221. Aten, J.A., Stap, J., Krawczyk, P.M., van Oven, C.H., Hoebe, R.A., Essers, J. & Kanaar, R. Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92-5 (2004).

222. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657-63 (2007).

223. de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben-Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., Laderman, S., Froguel, P. & Blakemore, A.I. Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. *Hum Mol Genet* **16**, 2783-94 (2007).

224. Breckpot, J., Takiyama, Y., Thienpont, B., Van Vooren, S., Vermeesch, J.R., Ortibus, E. & Devriendt, K. A novel genomic disorder: a deletion of the SACS gene leading to Spastic Ataxia of Charlevoix-Saguenay. *Eur J Hum Genet* **16**, 1050-4 (2008).

225. Chabrol, B., Martens, K., Meulemans, S., Cano, A., Jaeken, J., Matthijs, G. & Creemers, J.W. Deletion of C2orf34, PREPL and SLC3A1 causes atypical hypotonia-cystinuria syndrome. *J Med Genet* **45**, 314-8 (2008).

226. Bisgaard, A.M., Kirchhoff, M., Nielsen, J.E., Kibaek, M., Lund, A., Schwartz, M. & Christensen, E. Chromosomal deletion unmasking a recessive disease: 22q13 deletion syndrome and metachromatic leukodystrophy. *Clin Genet* (2008).

227. Bonduelle, M., Van Assche, E., Joris, H., Keymolen, K., Devroey, P., Van Steirteghem, A. & Liebaers, I. Prenatal testing in ICSI pregnancies: incidence of chromosomal anomalies in 1586 karyotypes and relation to sperm parameters. *Hum Reprod* **17**, 2600-14 (2002).

228. Shaw-Smith, C. et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* **38**, 1032-7 (2006).

229. Koolen, D.A. et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* **38**, 999-1001 (2006).

230. Koolen, D.A. et al. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. *J Med Genet* **45**, 710-20 (2008).

231. Chanda, B., Asai-Coakwell, M., Ye, M., Mungall, A.J., Barrow, M., Dobyns, W.B., Behesti, H., Sowden, J.C., Carter, N.P., Walter, M.A. & Lehmann, O.J. A novel mechanistic spectrum underlies glaucoma-associated chromosome 6p25 copy number variation. *Hum Mol Genet* **17**, 3446-58 (2008).

232. Nishimura, D.Y., Swiderski, R.E., Alward, W.L., Searby, C.C., Patil, S.R., Bennet, S.R., Kanis, A.B., Gastier, J.M., Stone, E.M. & Sheffield, V.C. The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nat Genet* **19**, 140-7 (1998).

233. Nishimura, D.Y., Searby, C.C., Alward, W.L., Walton, D., Craig, J.E., Mackey, D.A., Kawase, K., Kanis, A.B., Patil, S.R., Stone, E.M. & Sheffield, V.C. A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. *Am J Hum Genet* **68**, 364-72 (2001).

234. Mirzayans, F., Gould, D.B., Heon, E., Billingsley, G.D., Cheung, J.C., Mears, A.J. & Walter, M.A. Axenfeld-Rieger syndrome resulting from mutation of the FKHL7 gene on chromosome 6p25. *Eur J Hum Genet* **8**, 71-4 (2000).

235. Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P. & Moreau, Y. Gene prioritization through genomic data fusion. *Nat Biotechnol* **24**, 537-44 (2006).

236. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. & Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).

237. Olson, E.N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922-7 (2006).

238. Van Esch, H., Jansen, A., Bauters, M., Froyen, G. & Fryns, J.P. Encephalopathy and bilateral cataract in a boy with an interstitial deletion of Xp22 comprising the CDKL5 and NHS genes. *Am J Med Genet A* **143**, 364-9 (2007).

239. Zhang, L., Zhong, T., Wang, Y., Jiang, Q., Song, H. & Gui, Y. TBX1, a DiGeorge syndrome candidate gene, is inhibited by retinoic acid. *Int J Dev Biol* **50**, 55-61 (2006).

240. Gupta, P.A., Putnam, E.A., Carmical, S.G., Kaitila, I., Steinmann, B., Child, A., Danesino, C., Metcalfe, K., Berry, S.A., Chen, E., Delorme, C.V., Thong, M.K., Ades, L.C. & Milewicz, D.M. Ten novel *FBN2* mutations in congenital contractural arachnodactyly: delineation of the molecular pathogenesis and clinical phenotype. *Hum Mutat* **19**, 39-48 (2002).

241. Yagi, H. et al. Role of TBX1 in human del22q11.2 syndrome. *Lancet* **362**, 1366-73 (2003).

242. Goldman, D.C., Hackenmiller, R., Nakayama, T., Sopory, S., Wong, C., Kulessa, H. & Christian, J.L. Mutation of an upstream cleavage site in the BMP4 prodomain leads to tissue-specific loss of activity. *Development* **133**, 1933-42 (2006).

243. McFadden, D.G., Barbosa, A.C., Richardson, J.A., Schneider, M.D., Srivastava, D. & Olson, E.N. The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189-201 (2005).

244. Nowaczyk, M.J., Carter, M.T., Xu, J., Huggins, M., Raca, G., Das, S., Martin, C.L., Schwartz, S., Rosenfield, R. & Waggoner, D.J. Paternal deletion 6q24.3: a new congenital anomaly syndrome associated with intrauterine growth failure, early developmental delay and characteristic facial appearance. *Am J Med Genet A* **146**, 354-60 (2008).

245. Caselli, R., Mencarelli, M.A., Papa, F.T., Uliana, V., Schiavone, S., Strambi, M., Pescucci, C., Ariani, F., Rossi, V., Longo, I., Meloni, I., Renieri, A. & Mari, F. A 2.6 Mb deletion of 6q24.3-25.1 in a patient with growth failure, cardiac septal defect, thin upperlip and asymmetric dysmorphic ears. *Eur J Med Genet* **50**, 315-21 (2007).

246. Bisgaard, A.M., Kirchhoff, M., Tumer, Z., Jepsen, B., Brondum-Nielsen, K., Cohen, M., Hamborg-Petersen, B., Bryndorf, T., Tommerup, N. & Skovby, F. Additional chromosomal abnormalities in patients with a previously detected abnormal karyotype, mental retardation, and dysmorphic features. *Am J Med Genet A* **140**, 2180-7 (2006).

247. Osoegawa, K. et al. Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. *J Med Genet* **45**, 81-6 (2008).

248. Pirola, B., Bortotto, L., Giglio, S., Piovan, E., Janes, A., Guerrini, R. & Zuffardi, O. Agenesis of the corpus callosum with Probst bundles owing to haploinsufficiency for a gene in an 8 cM region of 6q25. *J Med Genet* **35**, 1031-3 (1998).

249. Tucker, A.S., Matthews, K.L. & Sharpe, P.T. Transformation of tooth type induced by inhibition of BMP signaling. *Science* **282**, 1136-8 (1998).

250. Dunn, N.R., Winnier, G.E., Hargett, L.K., Schrick, J.J., Fogo, A.B. & Hogan, B.L. Haploinsufficient phenotypes in *Bmp4* heterozygous null mice and modification by mutations in *Gli3* and *Alx4*. *Dev Biol* **188**, 235-47 (1997).

251. Bakrania, P. et al. Mutations in BMP4 cause eye, brain, and digit developmental anomalies: overlap between the BMP4 and hedgehog signaling pathways. *Am J Hum Genet* **82**, 304-19 (2008).

252. Nolen, L.D., Amor, D., Haywood, A., St Heaps, L., Willcock, C., Mihelec, M., Tam, P., Billson, F., Grigg, J., Peters, G. & Jamieson, R.V. Deletion at 14q22-23 indicates a contiguous gene syndrome comprising anophthalmia, pituitary hypoplasia, and ear anomalies. *Am J Med Genet A* **140**, 1711-8 (2006).

253. Elliott, J., Maltby, E.L. & Reynolds, B. A case of deletion 14(q22.1-->q22.3) associated with anophthalmia and pituitary abnormalities. *J Med Genet* **30**, 251-2 (1993).

254. Bennett, C.P., Betts, D.R. & Seller, M.J. Deletion 14q (q22q23) associated with anophthalmia, absent pituitary, and other abnormalities. *J Med Genet* **28**, 280-1 (1991).

255. Togi, K., Yoshida, Y., Matsumae, H., Nakashima, Y., Kita, T. & Tanaka, M. Essential role of Hand2 in interventricular septum formation and trabeculation during cardiac development. *Biochem Biophys Res Commun* **343**, 144-51 (2006).

256. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214-20 (2005).

257. Gao, K., Masuda, A., Matsuura, T. & Ohno, K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* **36**, 2257-67 (2008).

258. Takaesu, G., Kishida, S., Hiyama, A., Yamaguchi, K., Shibuya, H., Irie, K., Ninomiya-Tsuji, J. & Matsumoto, K. TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway. *Mol Cell* **5**, 649-58 (2000).

259. Ishitani, T., Takaesu, G., Ninomiya-Tsuji, J., Shibuya, H., Gaynor, R.B. & Matsumoto, K. Role of the TAB2-related protein TAB3 in IL-1 and TNF signaling. *Embo J* **22**, 6277-88 (2003).

260. Kanayama, A., Seth, R.B., Sun, L., Ea, C.K., Hong, M., Shaito, A., Chiu, Y.H., Deng, L. & Chen, Z.J. TAB2 and TAB3 activate the NF-kappaB pathway through binding to polyubiquitin chains. *Mol Cell* **15**, 535-48 (2004).

261. Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C. & Chen, Z.J. Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351-61 (2000).

262. Wang, C., Deng, L., Hong, M., Akkaraju, G.R., Inoue, J. & Chen, Z.J. TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346-51 (2001).

263. Kishimoto, K., Matsumoto, K. & Ninomiya-Tsuji, J. TAK1 mitogen-activated protein kinase kinase kinase is activated by autophosphorylation within its activation loop. *J Biol Chem* **275**, 7359-64 (2000).

264. Shibuya, H., Yamaguchi, K., Shirakabe, K., Tonegawa, A., Gotoh, Y., Ueno, N., Irie, K., Nishida, E. & Matsumoto, K. TAB1: an activator of the TAK1 MAPKKK in TGF-beta signal transduction. *Science* **272**, 1179-82 (1996).

265. Besse, A., Lamothe, B., Campos, A.D., Webster, W.K., Maddineni, U., Lin, S.C., Wu, H. & Darnay, B.G. TAK1-dependent signaling requires functional interaction with TAB2/TAB3. *J Biol Chem* **282**, 3918-28 (2007).

266. Prickett, T.D., Ninomiya-Tsuji, J., Broglie, P., Muratore-Schroeder, T.L., Shabanowitz, J., Hunt, D.F. & Brautigan, D.L. TAB4 stimulates TAK1-TAB1 phosphorylation and binds polyubiquitin to direct signaling to NF-kappaB. *J Biol Chem* **283**, 19245-54 (2008).

267. Chen, Z.J. Ubiquitin signalling in the NF-kappaB pathway. *Nat Cell Biol* **7**, 758-65 (2005).

268. Takada, I. et al. A histone lysine methyltransferase activated by non-canonical Wnt signalling suppresses PPAR-gamma transactivation. *Nat Cell Biol* **9**, 1273-85 (2007).

269. Ishitani, T., Kishida, S., Hyodo-Miura, J., Ueno, N., Yasuda, J., Waterman, M., Shibuya, H., Moon, R.T., Ninomiya-Tsuji, J. & Matsumoto, K. The TAK1-NLK mitogen-activated protein kinase cascade functions in the Wnt-5a/Ca(2+) pathway to antagonize Wnt/beta-catenin signaling. *Mol Cell Biol* **23**, 131-9 (2003).

270. Liu, Q., Busby, J.C. & Molkentin, J.D. Interaction between TAK1-TAB1-TAB2 and RCAN1-calcineurin defines a signalling nodal control point. *Nat Cell Biol* (2009).

271. Baek, S.H., Ohgi, K.A., Rose, D.W., Koo, E.H., Glass, C.K. & Rosenfeld, M.G. Exchange of N-CoR corepressor and Tip60 coactivator complexes links gene expression by NF-kappaB and beta-amyloid precursor protein. *Cell* **110**, 55-67 (2002).

272. Shim, J.H., Xiao, C., Paschal, A.E., Bailey, S.T., Rao, P., Hayden, M.S., Lee, K.Y., Bussey, C., Steckel, M., Tanaka, N., Yamada, G., Akira, S., Matsumoto, K. & Ghosh, S. TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways in vivo. *Genes Dev* **19**, 2668-81 (2005).

273. Jadrich, J.L., O'Connor, M.B. & Coucouvanis, E. The TGF beta activated kinase TAK1 regulates vascular development in vivo. *Development* **133**, 1529-41 (2006).

274. Xie, M., Zhang, D., Dyck, J.R., Li, Y., Zhang, H., Morishima, M., Mann, D.L., Taffet, G.E., Baldini, A., Khoury, D.S. & Schneider, M.D. A pivotal role for endogenous TGF-beta-activated kinase-1 in the LKB1/AMP-activated protein kinase energy-sensor pathway. *Proc Natl Acad Sci U S A* **103**, 17378-83 (2006).

275. Zhang, D., Gaussin, V., Taffet, G.E., Belaguli, N.S., Yamada, M., Schwartz, R.J., Michael, L.H., Overbeek, P.A. & Schneider, M.D. TAK1 is activated in the myocardium after pressure overload and is sufficient to provoke heart failure in transgenic mice. *Nat Med* **6**, 556-63 (2000).

276. Liu, H.H., Xie, M., Schneider, M.D. & Chen, Z.J. Essential role of TAK1 in thymocyte development and activation. *Proc Natl Acad Sci U S A* **103**, 11677-82 (2006).

277. Wan, Y.Y., Chi, H., Xie, M., Schneider, M.D. & Flavell, R.A. The kinase TAK1 integrates antigen and cytokine receptor signaling for T cell development, survival and function. *Nat Immunol* **7**, 851-8 (2006).

278. Sato, S., Sanjo, H., Takeda, K., Ninomiya-Tsuji, J., Yamamoto, M., Kawai, T., Matsumoto, K., Takeuchi, O. & Akira, S. Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nat Immunol* **6**, 1087-95 (2005).

279. Tang, M., Wei, X., Guo, Y., Breslin, P., Zhang, S., Wei, W., Xia, Z., Diaz, M., Akira, S. & Zhang, J. TAK1 is required for the survival of hematopoietic cells and hepatocytes in mice. *J Exp Med* **205**, 1611-9 (2008).

280. Sato, S., Sanjo, H., Tsujimura, T., Ninomiya-Tsuji, J., Yamamoto, M., Kawai, T., Takeuchi, O. & Akira, S. TAK1 is indispensable for development of T cells and prevention of colitis by the generation of regulatory T cells. *Int Immunol* **18**, 1405-11 (2006).

281. Omori, E., Matsumoto, K., Sanjo, H., Sato, S., Akira, S., Smart, R.C. & Ninomiya-Tsuji, J. TAK1 is a master regulator of epidermal homeostasis involving skin inflammation and apoptosis. *J Biol Chem* **281**, 19610-7 (2006).

282. Komatsu, Y., Shibuya, H., Takeda, N., Ninomiya-Tsuji, J., Yasui, T., Miyado, K., Sekimoto, T., Ueno, N., Matsumoto, K. & Yamada, G. Targeted disruption of the Tab1 gene causes embryonic lethality and defects in cardiovascular and lung morphogenesis. *Mech Dev* **119**, 239-49 (2002).

283. Sanjo, H., Takeda, K., Tsujimura, T., Ninomiya-Tsuji, J., Matsumoto, K. & Akira, S. TAB2 is essential for prevention of apoptosis in fetal liver but not for interleukin-1 signaling. *Mol Cell Biol* **23**, 1231-8 (2003).

284. Orelio, C. & Dzierzak, E. Identification of 2 novel genes developmentally regulated in the mouse aorta-gonad-mesonephros region. *Blood* **101**, 2246-9 (2003).

285. Wang, J. et al. The diploid genome sequence of an Asian individual. *Nature* **456**, 60-5 (2008).

Peer-reviewed publications

B Thienpont, J R Vermeesch & J P Fryns
25 Mb deletion of 13q13.3→q21.31 in a patient without retinoblastoma
Eur J Med Genet, 48 (3) (**2005**) 363-366

B Thienpont, M Gewillig, J P Fryns, K Devriendt & J Vermeesch
Molecular cytogenetic characterization of a constitutional complex intrachromosomal 4q rearrangement in a patient with multiple congenital anomalies
Cytogenet Genome Res, 114 (3-4) (**2006**) 338-341

B Thienpont, J Vermeesch & K Devriendt
Anterior cervical hypertrichosis and mental retardation
Clin Dysmorphol, 15 (3) (**2006**) 189-190

B Thienpont, J Breckpot, M Holvoet, J R Vermeesch & K Devriendt
A microduplication of CBP in a patient with mental retardation and a congenital heart defect
Am J Med Genet A, 143A (18) (**2007**) 2160-2164

B Thienpont, T de Ravel, H Van Esch, D Van Schoubroeck, P Moerman, J R Vermeesch, J P Fryns, G Froyen, C Lacoste, C Badens & K Devriendt
Partial duplications of the ATRX gene cause the ATR-X syndrome
Eur J Hum Genet, 15 (10) (**2007**) 1094-1097

B Thienpont, L Mertens, G Buyse, J R Vermeesch & K Devriendt
Left-ventricular non-compaction in a patient with monosomy 1p36
Eur J Med Genet, 50 (3) (**2007**) 233-236

B Thienpont, L Mertens, T de Ravel, B Eyskens, D Boshoff, N Maas, J P Fryns, M Gewillig, J R Vermeesch & K Devriendt
Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients
Eur Heart J, 28 (22) (**2007**) 2778-84

B Thienpont, J Breckpot, J R Vermeesch, M Gewillig & K Devriendt
A complex submicroscopic chromosomal imbalance in 19p13.11 with one microduplication and two microtriplications
Eur J Med Genet, 51 (3) (**2008**) 219-225

P Van Loo, S Aerts, B Thienpont, B De Moor, Y Moreau & P Marynen
ModuleMiner - improved computational detection of cis-regulatory modules: are there different modes of gene regulation in embryonic development and adult tissues?
Genome Biol, 9 (4) (**2008**) R66

S Van Vooren, B Thienpont, B Menten, F Speleman, B De Moor, J Vermeesch & Y Moreau

Mapping biomedical concepts onto the human genome by mining literature on chromosomal aberrations
Nucleic Acids Res, 35 (8) (**2007**) 2533-2543

I Balikova, B Menten, T de Ravel, C Le Caignec, B Thienpont, M Urbina, M Doco-Fenzy, M de Rademaeker, G Mortier, F Kooy, J van den Ende, K Devriendt, J P Fryns, F Speleman & J R Vermeesch
Subtelomeric imbalances in phenotypically normal individuals
Hum Mutat, 28 (10) (**2007**) 958-967

J Breckpot, Y Takiyama, B Thienpont, S Van Vooren, J R Vermeesch, E Ortibus & K Devriendt, A novel genomic disorder: a deletion of the SACS gene leading to Spastic Ataxia of Charlevoix-Saguenay, Eur J Hum Genet, 16 (9) (**2008**) 1050-1054

D Castermans, B Thienpont, K Volders, A Crepel, J R Vermeesch, C T Schr&er-Stumpel, W J Van de Ven, J G Steyaert, J W Creemers & K Devriendt
Position effect leading to haploinsufficiency in a mosaic ring chromosome 14 in a boy with autism
Eur J Hum Genet, 16 (10) (**2008**) 1187-1192

T J de Ravel, I Balikova, B Thienpont, F Hannes, N Maas, J P Fryns, K Devriendt & J R Vermeesch
Molecular karyotyping of patients with MCA/MR: the blurred boundary between normal and pathogenic variation
Cytogenet Genome Res, 115 (3-4) (**2006**) 225-230

C Le Caignec, C Spits, K Sermon, M De Rycke, B Thienpont, S Debrock, C Staessen, Y Moreau, J P Fryns, A Van Steirteghem, I Liebaers & J R Vermeesch
Single-cell chromosomal imbalances detection by array CGH
Nucleic Acids Res, 34 (9) (**2006**) e68

N Maas, B Thienpont, J R Vermeesch & J P Fryns
Facial asymmetry, cardiovascular anomalies and adducted thumbs as unusual symptoms in Dubowitz syndrome: a microdeletion/duplication in 13q
Genet Couns, 17 (4) (**2006**) 477-479

N M Maas, G Van Buggenhout, F Hannes, B Thienpont, D Sanlaville, K Kok, A Midro, J Andrieux, B M Anderlid, J Schoumans, R Hordijk, K Devriendt, J P Fryns & J R Vermeesch
Genotype-phenotype correlation in 21 patients with Wolf-Hirschhorn syndrome using high resolution array comparative genome hybridisation (CGH)
J Med Genet, 45 (2) (**2008**) 71-80

B Menten, N Maas, B Thienpont, K Buysse, J V&esompele, C Melotte, T de Ravel, S Van Vooren, I Balikova, L Backx, S Janssens, A De Paepe, B De Moor, Y Moreau, P Marynen, J P Fryns, G Mortier, K Devriendt, F Speleman & J R Vermeesch
Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports
J Med Genet, 43 (8) (**2006**) 625-633

H Peeters, M L Voz, K Verschueren, B De Cat, H Pendeville, B Thienpont, A Schellens, J W Belmont, G David, W J Van De Ven, J P Fryns, M Gewillig, D Huylebroeck, B Peers & K Devriendt
Sesn1 is a novel gene for left-right asymmetry and mediating nodal signalling
Hum Mol Genet, 15 (22) (**2006**) 3369-3377

Participations to scientific manifestations

Presentations

2nd Marie Curie Conference on arrayCGH and Molecular Cytogenetics

Italy, Bari *Oct 19 – 22, 2005*

Detection of as low as 5% structural low grade mosaics by array CGH in patients with idiopathic mental retardation and multiple congenital aberrations - *B Thienpont, N Maas, B Menten, K Buysse, J Vandesompele, C Melotte, T de Ravel, S Van Vooren, IG Balikova, L Backx, S Janssen, A De Paepe, B De Moor, Y Moreau, P Marynen, J Fryns, G Mortier, K Devriendt, F Speleman & JR Vermeesch*

<u>abstract:</u> Eur J of Med Genet

Sixth annual meeting BeSHG

Belgium, Antwerp *Feb 17, 2006*

Array CGH a novel tool in genetic diagnosis of individuals with congenital heart defect - *B Thienpont, N Maas, L Mertens, B Eyskens, D Boshoff, JP Fryns, M Gewillig, JR Vermeesch & K Devriendt*

2nd International Meeting On Cryptic Chromosomal Rearrangements in Mental Retardation and Autism

Italy, Troina *April 7-8, 2006*

Array-CGH: a novel tool in genetic diagnosis of individuals with congenital heart defect - *B Thienpont, N Maas, L Mertens, B Eyskens, D Boshoff, JP Fryns, M Gewillig, JR Vermeesch & K Devriendt*

European Human Genetics Conference 2006

The Netherlands, Amsterdam *May 6-9, 2006*

Array-CGH: a novel tool in genetic diagnosis of individuals with congenital heart defect - *B Thienpont, N Maas, L Mertens, B Eyskens, D Boshoff, JP Fryns, M Gewillig, JR Vermeesch & K Devriendt*

<u>abstract:</u> Eur J Hum Genet 14 supp 1 C65 p96 (2006)

Third Marie Curie Conference on arrayCGH and Molecular Cytogenetics

Belgium, Leuven *Sep 13-16, 2006*

Array-CGH: a novel tool in genetic diagnosis of individuals with congenital heart defect - *B Thienpont, N Maas, L Mertens, B Eyskens, D Boshoff, JP Fryns, M Gewillig, JR Vermeesch & K Devriendt*

Weinstein **Cardiovascular Development Conference**

USA, Houston, JW Marriott Hotel *May 15-18, **2008***

Array-CGH: a novel tool to identify the genetic causes of CHDs - *Thienpont B, Breckpot J, Van Loo P, Tranchevent L, Gewillig M, Moreau Y, Vermeesch JR & Devriendt K*

European Human Genetics Conference 2008

Barcelona, Spain *May 31 - Jun 3, 2008*

Towards an improved genetic diagnosis of individuals with a congenital heart defect – *B Thienpont, L-C Tranchevent, P Van Loo, J Breckpot, M Gewillig, Y Moureau & K Devriendt*

<u>abstract:</u> Eur J Hum Genet 16 supp 2 C4.5 p21 (2008)

The second International Workshop on Machine Learning in Systems Biology

Belgium, Brussels *Sep 13-14, 2008*

Endeavour pinpoints genes causing heart defects in regions identified by aCGH - *Thienpont B, Breckpot J, Barriot R, Van Loo P, Tranchevent L, Gewillig M, Moreau Y & Devriendt K*

European Human Genetics Conference 2009
Austria, Vienna, Austria centre Vienna *May 23 - 26, 2008*

An aCGH screening study in 150 patients identifies a novel dosage-sensitive gene, TAB2, which is disrupted in multiple patients with cardiac defects
Thienpont B, Breckpot J, Zang L, Tranchevent LC, Van Loo P, Møllgård K, Tommerup N, Bache I, Tümer Z, Waggoner D, Gewillig M, Peeters H, Moreau Y, Vermeesch JR, Larsen LA & Devriendt K

Posters

Fifth annual meeting of the Belgian Society of Human Genetics
Belgium, Luik, CHU
Jan 28, 2005

Partial trisomy of chromosome 21 in an older mentally retarded patient - *B Thienpont, H Starke, JR Vermeesch, J Fryns & G Van Buggenhout*

Seventh Meeting of the Belgian Society of Human Genetics
Belgium, Marcinelle
Apr 20, 2007

Partial duplications of the ATRX-gene cause the ATR-X syndrome - *B Thienpont, T de Ravel, H Van Esch, D Van Schoubroeck, P Moerman, J R Vermeesch, J Fryns, G Froyen, C Badens & K Devriendt*

European Human Genetics Conference 2007
France, Nice
Jun 16-19, 2007

Partial duplications of the ATRX-gene cause the ATR-X syndrome - *B Thienpont, T de Ravel, H Van Esch, D Van Schoubroeck, P Moerman, J R Vermeesch, J Fryns, G Froyen, C Badens & K Devriendt*
<u>abstract</u>*:* Eur J Hum Genet 15 supp 1 P0028 p42 (2007)

4th Integrative Bioinformatics Workshop
Belgium, Ghent
Sep 10-12, 2007

Towards an improved genetic diagnosis of individuals with a congenital heart defect – *B Thienpont, P Van Loo, J Breckpot, B De Cat, L Tranchevent, M Gewillig & K Devriendt*

Eighth Meeting of the Belgian Society of Human Genetics
Belgium, Leuven
Apr 25, 2008

Complex intrachromosomal rearrangements – *B Thienpont, J Breckpot, J R Vermeesch & K Devriendt*

Courses

Cardiogenesis and congenital Cardiopathies: From Developmental Models to Clinical Applications
Italy**,** Bologna (organized by the European Genetics Foundation)
Jun 6 – 10, 2008

Contributions to Meeting Abstracts

Presentations

European Human Genetics Conference 2005
Czech Republic, Prague, Prague Congress Center *May 7, 2005*
  Molecular karyotyping detects structural low grade mosaics in ~4 % of patients with idiopathic mental retardation and multiple congenital aberrations - *M C Maas, B Menten, C Melotte, B Thienpont, K Buysse, G Froyen, P Marynen, A De Paepe, J Fryns, G Mortier, K Devriendt, F Speleman, J R Vermeesch*

Fifth European Cytogenetics Conference
Spain, Madrid *Jun 4 – 7, 2005*
  ArrayCGH detects mosaics in ~4% of patients with MR/CA - *NMC Maas, B Menten, C Melotte, B Thienpont, K Buysse, G Froyen, P Marynen, A De Paepe, JP Fryns, G Mortier, K Devriendt, <u>JR Vermeesch</u>, F Speleman*
<u>*abstract:*</u> Chromosome research, 13 supp 1: 144-O (2005)

XVIth European Dysmorphology Meeting
France, Strasbourg *Sep 8 – 9, 2005*
  Array-CGH a novel tool in genetic diagnosis of individuals with congenital heart defects - *B Thienpont, N Maas, L Mertens, B Eyskens, D Boshoff, JP Fryns, M Gewillig, J Vermeesch, <u>K Devriendt</u>*

3êmes Assises de Génétic Humaine et Medicale
France, Montpellier *Jan 26 – 28, 2006*
  Array-CGH : une nouvelle approche dans le diagnostic génétique de patients avec malformations cardiaques congénitales - *B Thienpont, L Mertens, B Eyskens, D Boshoff, N Maas, JP Fryns, M Gewillig, J Vermeesch, <u>K Devriendt</u>*

European Human Genetics Conference 2006
Amsterdam, the Netherlands *May 6-9, 2006*
  Single-cell chromosomal imbalances detection by array CGH - *<u>C Le Caignec</u>, C Spits, K Sermon, M De Rycke, B Thienpont, Y Moreau, J P Fryns, A Van Steteghem, I Liebaers, JR Vermeesch*
<u>*abstract:*</u> Eur J Hum Genet 14 supp 1 C49 p92 (2006)

  Subtelomeric imbalances in phenotypically normal individuals - *<u>I G Balikova</u>, T de Ravel, C Le Caignec, B Thienpont, B Menten, F Speleman, K Devriendt, J P Fryns, J R Vermeesch*
<u>*abstract:*</u> Eur J Hum Genet 14 supp 1 C69 p97 (2006)

41st Annual Meeting of the Association for European Paediatric Cardiology
Basel Convention Centre, Basel, Switzerland *May 24 – 27, 2006*
  Array-CGH a novel tool in genetic diagnosis for patients with congenital heart defects *<u>Mertens L</u>, Thienpont B, Eyskens B, Boshoff D, Maas N, Fryns JP, de Ravel T, Gewillig M, Vermeesch JR, Devriendt K*
<u>*abstract*</u>: Cardiology in the Young 16 p4O-46 (2006)

Third Marie Curie Conference on arrayCGH and Molecular Cytogenetics
Belgium, Leuven *Sep 13-16, 2006*
  Subtelomeric imbalances in phenotypically normal individuals - <u>I Balikova</u>, *T de Ravel, C Le Caignec, B Thienpont, B Menten, M Urbina, F Kooy, F Speleman, K Devriendt, J P Fryns, & J R Vermeesch*

Second Benelux BioInformatics Conference
The Netherlands, Wageningen *Oct 17-18, 2006*
Gene prioritization by genomic data fusion: applications – *P Van Loo, B Thienpont, S Aerts, B De Cat, G David, Y Moreau, K Devriendt & P Marynen*

European Mathematical Genetics Meeting 2007
Germany, Heidelberg, Conference Center DKFZ *Apr 12-13, 2007*
Gene prioritization by genomic data fusion - *P Van Loo, S Aerts, D Lambrechts, B Thienpont, S Maity, B Coessens, F De Smet, L-C Tranchevent, B De Moor, K Devriendt, P Marynen, B Hassan, P Carmeliet, Y Moreau*
abstract: Annals of Human Genetics, 71 (4) p550-559 (2007)

18th European Dysmorphology meeting,
France, Bisschoffsheim *Sep 6 – 7, 2007*
Genotype-phenotype correlation in trisomy 16p13.3 syndrome – *J Breckpot, B Thienpont, M Holvoet, J-P Fryns, J Vermeesch, T de Ravel, K Devriendt*

36st Annual Congres of the Belgian Society for Pediatric Medicine
Belgium, Schelle *Feb 29 – Mar 1, 2008*
Array CGH: a novel tool in the genetic diagnosis of Congenital Heart Defects (CHD) - *J Breckpot, B Thienpont, J R Vermeesch, B Eyskens, L Mertens, M Gewillig, K Devriendt*

Genomic Disorders 2008
Wellcome Trust Conference Centre, Hinxton, UK *March 17 – 19, 2008*
A novel genomic disorder: a deletion of the *SACS* gene leading to Spastic Ataxia of Charlevoix-Saguenay – *J R Vermeesch, J Breckpot, R Barriot, B Thienpont, S Van Vooren, B Coessens, L-C Tranchevent, M Gewillig, Y Moreau & K Devriendt*

A BioSapiens workshop : From Genome to Proteome and Biological Function
Université Libre de Bruxelles *April 2, 2008*
ENDEAVOUR: gene prioritization through genomic data fusion. *L-C Tranchevent, S Aerts, B Thienpont, P Van Loo, S Yu, B Coessens, R Barriot, S Van Vooren, B Hassam & Y Moreau*

Mini EURO Conference on Computational Biology, Bioinformatics and Medicine
Italy, Rome, National Research Council (CNR) Headquarters *Sep 15-17, 2008*
Gene prioritization through genomic data fusion: algorithm and applications - *L-C Tranchevent, S Aerts, B Thienpont, P Van Loo, S Yu, B Coessens, R Barriot, S Van Vooren, B Hassam, Y Moreau*

Posters

Fifth annual meeting of the Belgian Society of Human Genetics
Belgium, Luik, CHU *Jan 28, 2005*
Detection of low grade mosaics by array CGH - *Melotte C, Maas N, Thienpont B, Devriendt K, Vanbuggenhout G, Fryns JP, Vermeesch JR*

ArrayCGH analysis detects cryptic chromosomal aberrations in more than twenty percent of patients with idiopathic mental retardation and multiple congenital anomalies - *N MC Maas, B Menten, C Melotte, B Thienpont, K Buysse, G Froyen, P Marynen, A De Paepe, J Fryns, G Mortier, K Devriendt, JR Vermeesch, F Speleman*

Sixth annual meeting of the Belgian Society of Human Genetics
Belgium, Antwerp *Feb 17, 2006*

A gene for non-syndromic autosomal dominant microcephaly on chromosome 5q35 or 18q22 - *A Crepel, B Thienpont, A Vogels, J-P Fryns, JR Vermeesch & K Devriendt*

Molecular karyotyping reveals 12% submicroscopic imbalances in a selected group of institutionalized patients with idiopathic mental retardation - *IG Balikova, JR Vermeesch, N Maas, B Thienpont, R Thoelen, K Devriendt, B Dimitrov, JP Fryns & T de Ravel*

BSCDB Spring Neuroconnectivity Meeting
Belgium, Leuven *May 4-6, 2006*

A gene for non-syndromic autosomal dominant microcephaly on chromosome 5q35 or 18q22 - *A Crepel, B Thienpont, A Vogels, J-P Fryns, JR Vermeesch & K Devriendt*

European Human Genetics Conference 2006
The Netherlands, Amsterdam *May 6-9, 2006*

Emerging patterns of cryptic chromosomal imbalances in patients with idiopathic mental retardation and multiple congenital anomalies - *J R Vermeesch, N Maas, B Thienpont, T de Ravel, I Balikova, J Fryns, K Devriendt, Y Moreau, B Menten, F Speleman, G Mortier, A De Paepe*

Third Marie Curie Conference on arrayCGH and Molecular Cytogenetics
Belgium, Leuven *Sep 13-16, 2006*

Molecular Karyotyping of patients with MCA/MR: The blurred boundary between normal and pathogenic variation – *T de Ravel, I Balikova, B Thienpont, F Hannes, N Maas, JP Fryns, K Devriendt & JR Vermeesch*

Seventh Meeting of the Belgian Society of Human Genetics
Belgium, Marcinelle *Apr 20, 2007*

A microduplication of CBP in a patient with mental retardation and a congenital heart defect - *J Breckpot, B Thienpont, M Holvoet, J-P Fryns, J Vermeesch, T de Ravel, K Devriendt*

Human-mouse somatic cell hybrids: a method to study positional effect on gene expression by chromosomal translocations - *A Crepel, T Voet, B Thienpont, A Vogels, J Fryns, J Vermeesch, K Devriendt*

European Human Genetics Conference 2007
France, Nice *Jun 16-19, 2007*

A study of candidate genes for dominant non-syndromic microcephaly in a family with inherited chromosomal translocations t(5q35.2;18q22.3) - *A Crepel, T Voet, B Thienpont, A Vogels, J Fryns, J Vermeesch, K Devriendt*
abstract: Eur J Hum Genet 15 supp 1 P0179 p75 (2007)

15th Annual International Conference on Intelligent Systems for Molecular Biology & 6th European Conference on Computational Biology
Austria, Vienna *Jul 21-25, 2007*

Literature Mining for Constitutional Cytogenetics - *S Van Vooren, B Thienpont, B De Moor, J Vermeesch, Y Moreau*

Quatrièmes assises de génétiques humaine et médicale
France, Lille
Jan 17-19, 2008

A deletion of the *SACS* gene resulting in Spastic Ataxia of Charlevoix-Saguenay (ARSACS) – *J Breckpot, Y Takiyama, B Thienpont, J R Vermeesch, E Ortibus, K Devriendt*

Eighth Meeting of the Belgian Society of Human Genetics

Belgium, Leuven *Apr 25, 2008*

      Array-Comparative genomic hybridisation: a novel tool in the genetic diagnosis of individuals with congenital heart defects– *J Breckpot, B Thienpont, J R Vermeesch, B Eyskens, L Mertens, M Gewillig, K Devriendt*

      Gene prioritization through genomic data fusion – *L-C Tranchevent, S Aerts, B Thienpont, P Van Loo, Y Shi, B Coessens, R Barriot, S Van Vooren, B Hassam & Y Moreau*

European Human Genetics Conference 2008

Spain, Barcelona *May 31 – Jun 3, 2008*

      CHDWiki: a comprehensive tool to gather and manage cardiogenetic data - *J Breckpot, R Barriot, B Thienpont, S Van Vooren, B Coessens, L-C Tranchevent, J R Vermeesch, M Gewillig, Y Moreau, K Devriendt*

<u>abstract</u>*:* Eur J Hum Genet 16 supp 2 P08.15 p398 (2008)

European conference on Computational Biology 2008

Italy, Cagliari Sep 22-26, 2008

      Gene prioritization through genomic data fusion: algorithm and applications - *L-C Tranchevent, S Aerts, B Thienpont, P Van Loo, S Yu, B Coessens, R Barriot, S Van Vooren, B Hassam, Y Moreau*

Het cliché wil dat een doctoraat nooit door één iemand alleen gemaakt wordt. In dit geval is dat een ernstige onderschatting. Dit doctoraat was niet mogelijk geweest zonder de hulp van velen (collega's, vrienden en familie) die ik dan ook erg veel dank verschuldigd ben. Zonder te proberen volledig te zijn wil ik toch een aantal mensen bedanken voor hun hulp, ondersteuning en medewerking.

Als eerste wil ik de medewerkers van het Centrum voor Menselijke Erfelijkheid bedanken. We beseffen dit niet vaak genoeg, maar jullie houden de boel draaiende. Reinhilde, Lut en Christoffer, bedankt om mij te initiëren in de wereld van FISH en het praktische reilen en zeilen in het cytogenetica labo, en voor de vele FISH experimenten die door jullie handen gepasseerd zijn. Ik bedank ook de mensen van het DNA labo voor de stalen die ze zorgvuldig bewaren (basismateriaal voor dit onderzoek), en de mensen van het secretariaat voor de hulp, het verzamelen en bijhouden van de patiëntengegevens (en zo veel meer). De weefselkweek bedank ik voor de cellen van de patiënten die ze opgroeiden: zoals jullie kunnen zien zijn die voor sommige studies erg belangrijk geweest. Marleen, Veerle en Rita voor het regelen van de financiën, de administratie en al de andere dingen.

Ik bedank vanzelfsprekend ook de klinische genetica eenheid: professor Fryns om het CME draaiende te houden en mij zo de kans te geven om te werken in dit fascinerende en dankbare onderzoeksdomein. Griet, Hilde VE en Thomy voor samenwerking, voor het doorverwijzen van patiënten en voor de klinische onderzoekingen die jullie bij zovele patiënten deden, Maureen voor het volgen van patiënten in niet altijd vanzelfsprekende omstandigheden, professor Legius voor het inzicht in markeranalyse en het gebruik van het hybridisatiestation. Ik bedank ook van harte de eenheid kindercardiologie: Dr Gewillig, Dr Eyskens, Dr Mertens en Dr Boshoff. Het moge duidelijk zijn dat zonder jullie samenwerking dit onderzoek onmogelijk zou geweest zijn.

Ook de medebewoners van het 2^e verdiep en aanverwanten wil ik bedanken: Boyan, Ilse, Liesbeth, Elyes, Wendy, Thierry, Sigrun, Evelyne, Femke, Francois, Caroline, Cedric, Paul, Pascal,… voor de gezellige babbel, de koekjes, de koffiezetmachine (die dagelijks uren toegevoegd heeft aan het bewust beleefde deel van mijn dag, dank dus Hilde B), het printerpapier, … .

Alle collega's onderzoekers van het DME wil ik bedanken voor de niet aflatende discussies, het niet geloven van mijn onderzoeksresultaten en het mij zo dwingen alles opnieuw kritisch te bekijken. Ik excuseer mij bij deze ook voor deze houding ten opzichte van jullie ;). Marijke: ook bedankt voor de Xpertise, de Winnie the Poeh outfit en uitstaan van mijn chaotische vragen. Jelle voor het aangename cafégevoel op het CME, Idoya for the Aalst experience, Raf voor het ernstige geneuzel, Bart voor de joviale hulp en interesse, Katleen voor de visjes.

Ik wil ook de mensen van de zebravis faciliteit bedanken: jullie nemen werkelijk een belangrijke last van onze schouders, laten bovendien alles vlot verlopen en helpen waar mogelijk met de experimenten. Sophie en Fré, merci.

I sincerely thank our collaborators of the ESAT department for standing my non-binary way of thinking. Yves for passing by on the mountain, for bringing your PC and the associated students, knowledge and way of thinking. Leo and Peter VL for agreeing to consider with me the same question in about 50 different ways. Peter VL ook voor het

inzicht in CRMs, Daniela for the interesting concepts on weighed networks. Steven voor de coole tools met nog coolere lay-out, voor de ontelbare bugs die je opgelost hebt, voor Clone Wars, Bench, ABandApart en de boogjes. Bij wijze van reclame vermeld ik hier ook Cartagenia nog even (en Bert !). Ik bedankt ook Peter K om mij in R te introduceren – nooit gedacht dat ik daar plezier aan zou beleven.

I would like to thank the members of my jury: Prof. Anita Rauch, Prof. Miikka Vikkula, Prof. An Zwijsen, Prof. Diether Lambrechts, Prof. Thierry Vandendriessche and Prof. Bart De Strooper. Thank you for the encouraging and constructive comments on the manuscript and for granting me the honour to be part of my jury. Prof. Rauch and Prof. Vikkula, thank you for travelling to Leuven on this occasion.

My special thanks to the collaborators from Leuven, Belgium and across the world who provided help, patient data and material: Prof. Van Schoubroeck, Prof. Moerman, Prof. Buyse, Prof. Badens, Dr. Menten, Prof. Speleman, Katrine, Prof. Larsen, Prof. Waggoner, Dr. Zhang, Dr. Bena, Dr. Philip, Dr. Salamone, Dr. Kussmann, Dr. Villard, Prof. Bottani and Prof. Cheung. I also would like to thank patients and parents who agreed to participate in these investigations.

Ik wil vanzelfsprekend ook iedereen van onze onderzoeksgroep bedanken (onlangs omgedoopt tot het wel erg chique klinkende "Laboratory for Genetics of Human Development"). Koen: onnodig te zeggen dat jij met voorsprong het meest hebt bijgedragen tot dit werk en (veel belangrijker) tot het omscholen van deze bio-ingenieur tot een doctoraatsstudent in de medische wetenschappen, de introductie in het "vorsen" (zoals je steeds zei: research = search, en re-search, en re-search). Bedankt om mij de kans te geven te werken in deze uitermate boeiende discipline, voor je aanstekelijke enthousiasme, voor je steeds ambitieuze en vooruitstrevende plannen en zeer praktische kijk op hun uitwerking, om mij te leren zeer kritisch te zijn, voor je vertrouwen en voor je steun in moeilijke momenten. Het is ongetwijfeld dankzij jou dat ik nog geen dag het gevoel heb gehad dat ik moest komen werken. Jeroen: voor de verdubbeling van de krachten in het cardiogenetica team maar vooral ook voor de werksfeer (het aangename aan het nuttige paren, zoals men dat dan pleegt te noemen). An: toch even zeggen dat ik erg veel bewondering heb voor je doorzettingsvermogen: translocaties zijn duidelijke een tweesnijdend zwaard. Dries en Hilde, jullie reken ik ook tot deze groep. Bedankt en veel succes in jullie verdere exploten.

Mijn familie, vrienden en buren mag ik zeker niet vergeten te bedanken. Jullie hebben je neergelegd bij de eigenaardige werkuren die ik had, mijn soms vreemde hersenkronkels die het gevolg waren van overdosis genen. Dit is ook bij uitstek de plaats om mijn moeder en vader te bedanken: mams en papa, op jullie vertrouwen kon ik steeds terugvallen, dankzij jullie wist ik me steeds gesteund. Ik wil jullie dan ook van harte bedanken, voor alles. Clara, Benedicte, Ignace: dank voor jullie hulp en steun.

Tenslotte, Irina, thanks for the confidence, trust, support, smiles and so much more (I don't know where to start). I am looking forward to living an interesting live together, and never forget: obichamte japloefka.

Allemaal heel erg bedankt!