Genetic variation in the non-canonical Wnt signalling pathway and predisposition to Tetralogy of Fallot

ADDISON JULIAN PALOMINO DOZA

A thesis submitted to Newcastle University for the degree of doctor of Philosophy
08/12/2010

To my grandfather who learnt English by mail and taught me the value of work, the weight of dignity, the color of honesty

To my grandmother who endured all the mail courses and loved me more than anyone else

ABSTRACT

Introduction

Congenital cardiovascular malformation (CCM) is a prevalent group of conditions in humans. Tetralogy of Fallot (TOF) is the commonest cyanotic heart defect affecting 0.24/1000 newborns worldwide. Human CCM etiology has a genetic component with 17% of CCM attributable to particular genetic syndromes. The remaining 80% of "sporadic" cases nevertheless show a high heritability, suggesting complex polygenic inheritance modulated by environmental factors. Multiple previous studies have explored the involvement of the non-canonical Wnt signaling pathway, also known as the planar cell polarity pathway, in cardiogenesis but none has explored the relationship between genetic variation within the pathway and predisposition to TOF in humans.

Methods

Exonic sequences of 7 key genes within the pathway (ROCK1, ROCK2, Wnt11, Wnt5a, Dvl3, Dvl1l1 and ANKRD6) were obtained in 93 TOF probands as an initial mutational screen. All novel variants were genotyped in at least 465 TOF probands and 1465 controls and analyzed using in silico methods to determine impact in structure and splicing. Three variants were functionally assessed. Genotypes for tag SNPs in ROCK1 were also obtained.

Results

Twelve uncommon previously unreported variants were found within the seven genes: four non synonymous, 3 synonymous, 3 intronic (near exon/intron boundaries) and 2 located at untranslated regions. Two novel common variants were also found: one synonymous and one intronic. None of the variants was proven to be *de novo*. None of the probands carried more than one novel variant. Statistically significant differences in allele frequencies were found for ROCK1 C807T and WNT11 207 +47 G>T. WNT5A Asp119Ser was found as a singleton in one proband and not in controls. Functional experiments did not show splicing changes associated with ROCK1 C807T and WNT11 207 +47 G>T. Rs288979, an intronic common variant (mAF=0.051) was significantly associated with TOF (p=0.000015 OR: 0.61).

Conclusions

Results of this work suggest that genetic variation in the non canonical Wnt signaling genes is involved in the etiology of TOF in humans. The genetic predisposition to TOF seems to be given by a combination of rare variants (ROCK1 C807T, WNT11 207 +47 G>T and WNT5A Asp119Ser) associated with high ORs and common variants (rs288979) associated with low ORs. Further deep sequencing work in larger samples should unravel the remaining rare variants associated with TOF, while genome-wide association study (GWAS) may reveal the role for additional common variants.

ACKNOWLEDGMENTS

I first would like to thank my mother for her support and encouragement during all these years.

I am indebted to my supervisor Prof. Bernard Keavney for his guidance, support, help, and encouragement since the first day I stepped onto British soil; to my co-supervisors Prof. Judith Goosdhip and Prof. Deborah Henderson for all their guidance and advice.

I would like also to thank Andrea Barrera for her company, care, support and for bearing all these scientific discussions during these years.

I am grateful to the faculty of Medicine, Universidad de los Andes, Colombia due to their support during the first years of this work; also to Newcastle University for the Overseas Research Scholarship and the Newcastle University International Scholarship that supported me all the way long.

I am also grateful to all the people in Prof Keavney laboratory in the Institute of Human Genetics for their guidance patience and friendship: Ana Topf, Huay Tan, Mike Cunnington, Darroch Hall, James Eden, Rafiqul Hussein and Thahira Rahnman. I would also like to thank all the people in the Institute specially Caroline Dalgliesh, Helen Blair and Dr. Bill Chaudhry for their help.

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ASD	Atrial septal defect
AVSD	Atrioventricular septal defect
BAV	Bicuspid aortic valve
BLAST	Basic Alignment Search Tool
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
CAD	Coronary artery disease
CCM	Congenital cardiac malformation
CDCV	Common disease common variant
CDRV	Common disease rare variant
CE	Convergent extension
CHD	Congenital heart disease
CNC	Cardiac neural crest cells
CNV	Copy number variant
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco Modified Eagle Media
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DORV	Double outlet right ventricle
Dvl	Dishevelled
EDTA	Ethylenediaminetetraacetic acid
ES	Embryonic stem cells
EST	Expressed sequence tag
FCS	Foetal calf serum
FDR	False discovery rate
FHF	First heart field
FITC	Fluorescein isothiocyanate
FZ	Frizzled
GI	Gastrointestinal
GWA	Genome wide association
HRC	Human random controls
HRP	Horseradish peroxidise
HUVECs	Human Umbilical Vein Endothelial Cells
IQR	Interquartile range
LD	Linkage disequilibrium
LFIP	Low frequency intermediate penetrance
Lp	Loop tail
mAF	Minor allele frequency
MAF	Major allele frequency
MLPA	Multiplex ligation-dependent Probe Amplification

mRNA	Messenger ribonucleic acid
NCC	Neural crest cells
NEAA	Non essential aminoacids
OFT	Outflow tract
OR	Odds risk
PAR	Population attributable risk
PBS	Phosphate buffered saline
PCP	Planar cell polarity pathway
PCR	Polymerase chain reaction
PDA	Patent ductus arteriosus
PTA	Patent truncus arteriosus
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reactior
RV	Right ventricle
RVOT	Right ventricle outflow tract
SHF	Second heart field
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
Taq	*T. Aquaticus*
TBST	Tris buffered saline- Tween 20
TEMED	N',N',N',N'-Tetramethylethylenediamine
TOF	Tetralogy of Fallot
UTR	Untranslated region
VSD	Ventricular septal defect
WTCCC	Welcome trust case control cohort

1. INTRODUCTION

1.1. Genetics of complex diseases - background

A complex disease or trait corresponds to a condition in which risk cannot be ascribed to a single genetic or environmental factor (1). Inheritance in this type of conditions does not correspond to Mendelian models and multiple genes act to increase or reduce the risk of developing them (2). In complex diseases individual risk is determined by an unknown combination of genetic, environmental and stochastic factors (3). Typically, no single factor, genetic or environmental, is either sufficient or necessary to develop the disease.

In contrast to the relative success finding the genetic aetiology of Mendelian diseases, research on genetic predisposition to complex diseases had proven difficult until very recently (4). Initially, only a few genes were consistently linked to conditions and many of the associations were characterized by non reproducibility. However, since the completion of ground breaking genome wide associations (GWA) studies, a large number of consistent associations have been described and reproduced (5).

The evolution of research into complex diseases departed from the point where success was achieved in mapping Mendelian diseases. The first approach to unravel genetic susceptibility was linkage analysis. The strategy was extremely successful in Mendelian diseases, with approximately 2200 disorders linked to a gene today (6). When applied to common diseases, linkage mapping proved a valuable tool to discover rare Mendelian forms of common diseases like hypertension, breast cancer or diabetes, but the proportion of heritability linked to the loci identified was minimal. In terms of discovering variants causing the non-Mendelian forms of these and other complex diseases, linkage analysis either in extended pedigrees or in affected sibling-pairs was an almost complete failure. Failure of linkage analysis in common disease was primarily due to the fact that the method is dependent on the size of the risk associated with the variant being large, and as a consequence that there is a close correlation between allele transmission and the presence of the phenotype in affected families. In 1996, Risch and Merikangas published

in Science perhaps the most influential article in genetics of the past twenty years. This was a theoretical explanation of the failure of linkage studies of common diseases. It was shown mathematically that for variants conferring a relative risk of less than 2 for disease, the probability of being detected by linkage mapping was minimal, and that an association approach using single nucleotide polymorphisms, the commonest form of variation in the human genome, would be required. (7).

The next step was association studies of candidate genes. In this case the constraints of family studies and transmission imbalance were solved but the method faced other problems. There was only a small prior probability of any one of 20,000 candidate genes being associated with the conditions, and the studies needed a high methodological rigor to avoid false positives due to population stratification and multiple test analysis (4, 6). Some encouraging results were obtained, often with genes where the prior probability of involvement could be inferred to be large (for example, the ApoE E2/E3/E4 polymorphism, which affects plasma lipid levels, and myocardial infarction) but still there was an overwhelmingly large portion of heritability not being explained by the loci found, and most of the positive results failed to be reproduced.

Two main factors influenced development of the next tool to investigate complex disease genetics in the “post genomic era” following the publication of the human reference genome: description and build of comprehensive genomewide human genetic variation databases; and development of commercial massively high througput genotyping methods, chiefly “SNP chips”. Using these databases and genotyping methods, studies in where more than a hundred thousand to a million SNPs across the genome are genotyped at thousands of individuals were made possible. These studies were called genome wide association studies (GWAs) (5, 8).

GWAs were believed for some years to be the tool to take forward genetics of complex diseases, but it was only in 2007 with the publication in Nature of the Welcome Trust Case-Control Consortium (WTCCC) study that their full potential became apparent. The WTCCC was a joint GWA research project conducted in a British population examining 7 diseases in 2000 probands and 3000 shared controls. This work identified 24 independent association signals in case control comparisons for rheumatoid arthritis, type 1 and 2 diabetes, bipolar disorder, Crohn’s disease and coronary artery disease (9). Today,

hundreds of consistently associated loci with a number of different diseases have been identified using this approach.

GWAs were designed to capture most, if not all, common genetic variation in the genome. However, despite encouraging results most of the heritability is still not explained by the associated loci found (8). A number of explanations have been suggested to explain the "missing heritability" in GWAs, including the possibility of larger numbers of variants of smaller effect not found yet, the larger effect of uncommon variants that are poorly detected by available genotyping arrays, rare structural variants poorly captured on existing chips (like copy number variants (CNVs)) and low power of studies to detect gene/gene, and gene/environment interaction.

In general it remains unclear how associated variants explain genetic susceptibility to complex diseases. One of the striking results of GWAs for all conditions is that the overwhelming proportion of identified loci do not correspond to missense or nonsense variants, and are not even located in known protein reading frames. Research into the functional implications of these non coding variants is related to the understanding of genetics of gene expression and the effects of these variants on transcription and translation (10).

1.2. Genetic architecture of complex diseases

Allelic or genetic architecture of complex diseases refers to the number of alleles and their frequencies affecting susceptibility to a given condition at a given disease locus (11). The importance of this allelic architecture is based on the marginal risk increase associated with each allele at a disease gene, and therefore the size of the statistical signal associated with it (7). If this architecture is understood, study design can be adapted, increasing the probability of success.

The structure and implications of "genetic architecture" have been one of the most debated topics in complex disease genetics. Two main hypotheses: the common disease-common variant (CDCV) and common disease-rare variant (CDRV) theories have been advanced to explain genetic contribution to individual diseases susceptibility.

The CDCV hypothesis supports the idea that genetic variation or susceptibility in complex traits is due to a common variant or a small number of them with small effects on quantitative phenotypes related to disease risk, or with very low penetrance (odds ratios generally less than 1.5) in the population. In this case allelic heterogeneity at any locus would be low, and susceptibility alleles should be found using linkage disequilibrium strategies (11, 12).

These alleles are ancient variants that due to different genetic effects have become common among the populations. As ancient variants they would be segregated in “blocks” and would be linked to other variants that can be used as markers. The reasons explaining a deleterious variant reaching high frequencies at a population include random genetic drift and fluctuating selection pressures if at some time the allele has granted selective advantage. This might be the case for the principal allele causing cystic fibrosis (ΔF508). Some theories have linked the high prevalence of this clearly deleterious variant to resistance to secretory diarrhoea transmitted by cattle in the early stages of civilization in Europe (13, 14). The CDCV hypothesis is certainly to some extent true, having been borne out by the recent results of GWA studies; however, as noted above, variants detected so far account for only a small proportion of the calculated heritability of any disease studied. Two well known examples, from before the GWAs era, that support the CDCV theory are the APOE and the PPARγ loci. In the first one a common allele, known as ε4, with frequencies ranging between 0.05 and 0.4 predisposes to Alzheimer disease (15). In the second one, a single frequent variant (ProAla12), seems to decrease risk of type 2 Diabetes (16).

Due to the success of GWAs and the number of alleles being discovered with the approach, a catalogue to group the findings has been built (17). The last available catalogue (09/07/2010) lists 631 papers and 3073 significantly associated SNPs. The description of these findings is beyond the scope of this thesis. See Table 1 for a list of a few conditions with the number of loci found to be associated at GWAs and heritability explained by the variants. For a graphical example of the number of loci associated with 148 traits see Figure 2.

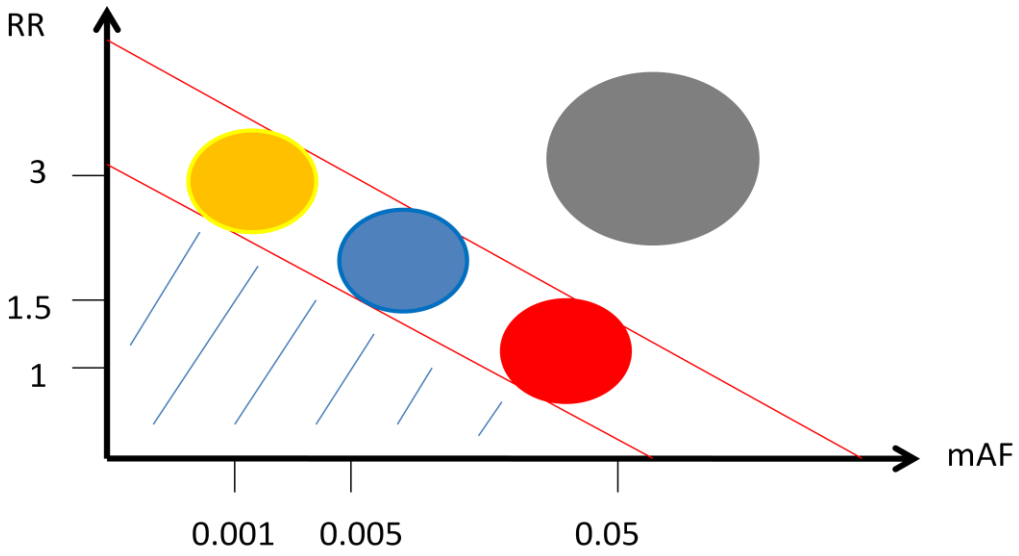

Figure 1. Graphical representation of feasibility of identifying susceptibility variants depending on their effect (RR) and their minor allele frequency. The area between the red lines represents the variants commonly detected on genetic research in complex diseases. Yellow circle represents those very rare variants with high effects, typically associated with Mendelian forms of disease, detectable historically by linkage mapping, and more recently by whole-exome sequencing. Blue circle represents those rare variants that confer intermediate risk and can be detected on mutational screenings or association studies. Red circle represents those common variants often associated with low risk and that are detected by GWAs or association studies. The area under the red lines with parallel blue lines represent those rare variants with that confer low risks and probably lie below the threshold for detection. The grey circle represent the very few common variants that confer high RRs. Adapted from Manolio et al. Nature 2009; 461(7265):747-753

On the other hand, the CDRV hypothesis supports that genetic susceptibility to complex traits is principally due to numerous uncommon variants associated with high risk of suffering the condition. Allelic heterogeneity in this case would be high, and the probability of finding susceptibility alleles using LD would be low. These uncommon variants are therefore a consequence of mutational events, and they correspond to recently acquired mutations subjected to evolutionary pressures with frequencies that depend on the deviation of fitness associated with them.

There is evidence to support both, CDCV and CDRV hypotheses, and the genetic architecture of disease susceptibility may, in general, be a mixture between common and uncommon variants. The genetic architecture of particular diseases might be determined by conditions like mutation rate and evolutionary pressures (11).

The achievements of GWAs have completed a broader picture of genetic susceptibility to complex diseases. However, the GWA studies are designed to fit the CDCV theory and as a consequence findings correspond to common genetic variants and the risk associated to uncommon ones is still under investigation.

Trait	Heritability explained	Number of loci
Fasting glucose	1.50%	4
Early myocardial infarction	2.80%	9
Height	5%	40
HDL cholesterol	5.20%	7
Type 2 diabetes	6%	18
Systemic lupus erythematosus	15%	6
Crohn's disease	20%	32
Age related macular degeneration	50%	5

Table 1. List of 8 traits, the number of loci associated with them at GWAs and the portion of heritability explained by the variants. Note that there are traits, like height, where a high number of associated variants have been found but the percentage of heritability explained is low. On the list there are also traits with a small number of variants associated explaining a big portion of heritability (eg. Age related macular degeneration)

To illustrate evidence for CDRV we can take the example of research on NOD2, a gene implicated in the susceptibility to Crohn´s disease (CD). The Caspase-Activation Recruitment Domain containing protein 15 , also known as NOD2, was the first susceptibility gene identified to be associated to the risk of developing CD (18). Despite around 30 loci having been associated to the condition, the highest risk is still linked to variants at NOD2 (19). Three relatively common (mAF around 5%) variants in NOD2 have been associated with CD: R702W, G908R and 1007fsinsC (20). However, mutational screening experiments and a metanalysis have shown numerous uncommon variants being associated with the trait and explaining a portion of heritability (21) .

As clearly explained in a model by Pritchard et al (11), it would be expected that genetic susceptibility alleles would be subject to purifying selection and, at expected human mutation rates, genetic susceptibility loci would have high genetic heterogeneity and a high number of uncommon and relatively new variants responsible for phenotypic variation. It seems, nonetheless, that in certain diseases allele frequencies can be subjected to unexpected effects due to random drifts or selective advantages (11, 22).

Knowing the genetic architecture of each disease locus would be the first task for an optimal research design in complex diseases. Still, most of the factors involved are unknown and the debate is open.

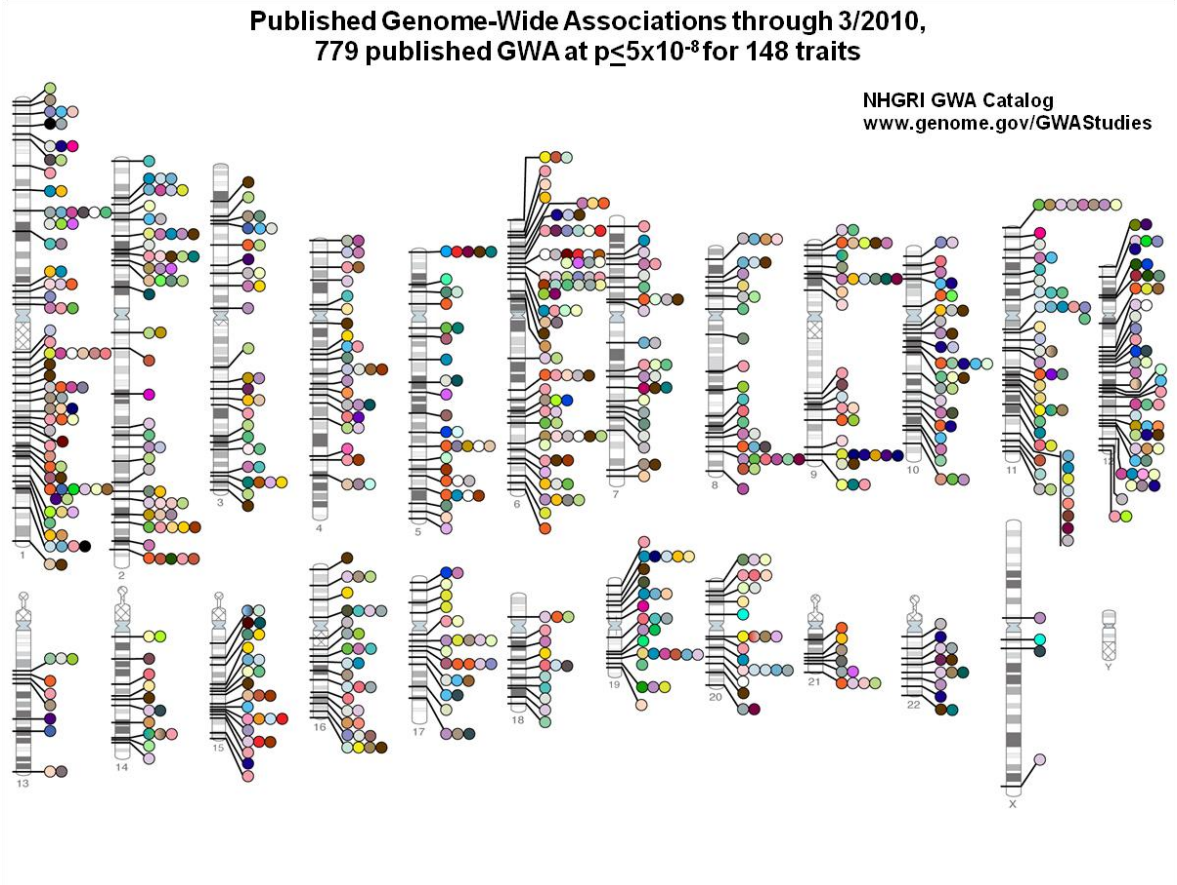

Figure 2. Scheme representing by colours association signals on the 24 chromosomes for 148 traits. Taken with permission from www.genome.gov/gwastudies. Accessed [3/09/2010].

1.3. Epidemiology of Congenital Cardiovascular Malformations

The prevalence and incidence of CCM varies from cohort to cohort. This variation depends not only in the proband's genetic background but also in the ascertainment criteria and diagnostic methodology used.

Hoffmann in 2002 (23) analysed the available data from research works after 1955 in incidence and prevalence of CCM, concluding that the incidence of CCM is approximately 8 per 1000 newborns. This incidence increases to 20 if the Bicuspid aortic valve (BAV) and silent patent ductus arteriosus (PDA) are taken into account. According to Hoffman, variation between different estimates was related to the relative frequency of VSDs, and this in turn was related to how early diagnosis was made. His conclusion was that given

the source of variation, incidence of CCM is homogeneous between geographical regions and time points.

Findings in Hoffmann's study were higher than estimates from two of the biggest previous studies: the Baltimore-Washington Infant Study, which reported a 6 year incidence of 4.34/1000 live births (24), and the Pradat et al. study which reported an incidence of 2.85/1000 live births including data from three large registries from France, Sweden and California (25). The difference was mainly due the detection and inclusion of small VSDs and BAV.

A good approximation for many western countries is that the incidence of CCM is approximately 7/1000 live births or 1 in 145 newborns, with variations depending in the methodology used (26). Adult CCM prevalence is more difficult to determine. It is thought that in the UK at least 150,000 adults carry a form of CCM. This would give an approximate prevalence of 250 cases per 100,000 (26). In 2009 Oyen et al. published a paper in Circulation exploring CCM epidemiology in a large population-based study in Denmark. Births between 1977 and 2005 were analysed. From 1,793,591 persons born in the period, 18,708 showed at least one form of CCM with a calculated prevalence of 103 per 100,000 live births (27).

Being the most common form of congenital malformation(28), and the leading cause of mortality in the paediatric population (29), CCM constitutes not only a challenging group of diseases for researchers but also for clinicians. Despite the fact that 50 years ago only 25% of the patients survived beyond childhood (30), the mortality for this group of diseases in the paediatric population calculated by Wren in 2001 was 18%, and in the adult population it was 13.9%(31). A linear increase is therefore expected in the number of patients reaching maturity. CCM mortality is associated with acute surgical complications, arrhythmias, endothelial dysfunction, secondary structural damage and heart failure (30). In conclusion, CCM is a group of highly prevalent conditions, with great impact in terms of morbidity and mortality in the community.

1.4 Aetiology of Congenital Cardiovascular Malformation

Congenital Cardiovascular Malformations (CCM) are defined as a group of defects present at birth within the cardio-circulatory structure and that are actually or potentially of functional relevance (27, 32). However, a percentage of these anomalies are not detected at birth and some experts include in this group diseases not present but with a latent predisposition at birth, like cardiomyopathies. Approximately 60% of CCM are diagnosed in babies less than 1 year old, 30% in infants and 10% in adult life (26)

These defects are caused by aberrant patterns of development or failure to progress from an early stage in normal cardiogenesis (32). In the next pages I will review the process of normal cardiogenesis, the genetic factors involved and how the normal process is either truncated or distorted to produce CCM.

The crucial time of cardiac development occurs between gestation week 6 and 12 and therefore most of the factors involved in CCM exert their influence at this stage. Additionally, some other anomalies can be developed later in pregnancy, like valve diseases (26) . A broad spectrum of factors, acting at crucial stages of cardiogenesis has been implicated as contributors to CCM. These contributors could be grouped into “genetic” and “environmental”. Nonetheless, the majority of cardiac malformations are neither purely genetic nor environmental, with most of CCM being classified as multifactorial or “complex diseases”. As an example, Ferencz et al concluded in 1989, that cardiac malformations associated with non cardiac abnormalities are caused by chromosome abnormalities in 12.9%, single mutations in 4.6%, and teratogenic factors in 8.3% (33). However, in the remaining group without non cardiac abnormalities the causes remained unknown in 73.2%. Most of the subsequent studies have concluded similarly, with approximately 17% of CCM being “purely genetic” or as I will call them thereafter “syndromic” and the remaining cases called “sporadic” behaving as complex traits.

It is important to remember that families in this large group of “sporadic cases” present higher familiar recurrence rates than general population, pointing to genetic factors being involved (34). In 1998 Burn et al. (35) published the results of a cohort of patients

affected by CCM, where the recurrence rate was 2.1% and the overall risk for first degree relatives was 4%. Most previous and subsequent cohorts have found higher recurrence rates among relatives of CCM patients when compared to general population, with rates varying depending on patient selection and ascertainment methods. A meta-analysis published in 1994 showed an overall 5.8% risk of transmission if the mother suffers CCM, and 3.1% if the father is affected (36).

In the Oyen et al. population based study calculated relative risks for monozygotic and dizygotic twins were 15.17 and 3.33 respectively. Recurrence risk for singletons with an affected first degree CCM affected family member was 3.21. Interestingly, “same heart defect” recurrence risk was calculated in 8.15 for first degree relatives overall. When probands with known chromosomal aberrations and extracardiac malformations were excluded the risk only slightly changed (8.15-8.61), indicating that the increased risk corresponded to non syndromic CCM. When data for dissimilar defects were analysed recurrence risk among first degree relatives was calculated as 2.68. However, in this cohort the population attributable risk for CCM family history was only 2.2%. The low PAR can be attributed to the preponderance of environmental factors or low penetrance genetic variants requiring either a combination of them or an additional environmental trigger (27).

Numerous environmental exposures are suspected to be responsible for approximately 8% of CCM cases. Among these we count: Infections like rubella; maternal diseases like diabetes, lupus, epilepsy or phenylketonuria; drugs like alcohol, hydantoin, thalidomide, retinoids, cocaine, and solvents; or environmental pollutants like nitrofen (37, 38). In summary, CCM is a heterogeneous group of diseases, in which the contributing factors have been elucidated for a fifth of the cases, including Mendelian and environmental causes. Research results in the remaining 80% of the cases of unknown aetiology suggest a complex interaction between genetic and environmental factors.

1.4.1. Syndromic CCM

Numerous syndromic forms of CCM have been described. Many of them are rarely seen and some are relatively common. Among the commoner syndromic CCMs is the association of Down's syndrome and atrio-ventricular septal defects (AVSD). Down's syndrome coexists with AVSDs in 35% of cases and 75% of AVSDs encountered clinically are present in Down's syndrome sufferers (26). Other chromosome number anomalies like trisomy 18, trisomy 13, Klinefelter Syndrome and Turner Syndrome (38, 39) are consistently associated to CCM.

Nonetheless, not only gross genetic anomalies at the chromosome level have been associated to CCM. Despite a deep review of the previously described forms of syndromic CCM being beyond the scope of this introduction, I will describe some of the commoner forms with a special section for the 22q11.2 deletion. Some of the commoner and best described syndromic forms of CCM are:

- Holt Oram Syndrome: Characterized by limb and cardiac anomalies like atrial septal defects (ASD), TOF and atrioventricular conduction defects. The syndrome is caused by mutations at TBX5 (40).
- Char syndrome: Characterized by limb and cardiac anomalies, predominantly patent ductus arteriosus (PDA). The defects are associated to genetic defects at the so called Char locus, leading to a deficiency in the activity of the TFAPβ protein (40).
- Noonan syndrome: Comprising characteristic facial features, chest deformities and pulmonary valve stenosis. It has been linked to mutations at the PTPN11 and K-RAS genes.

Syndrome	Gene	Cardiac malformation
CHARGE	Sema3E/CHD7	ASD, VSD, mitral valve defects
Noonan	K-Ras/PTPN11/Shp2	PS
Cardio-Fascio-Cutaneous	K-Ras	PS
Costello	H-Ras	PDA, PS
Cardio-Fascio-Cutaneous	MEK1/MEK2/B-Raf	PS
DiGeorge modifier	VEGF promoter	VSD, PTA, IAA, TOF
DiGeorge	TBX1	VSD, PTA, IAA, TOF
Char	TFAP2β	PDA
Ellis-Van-Creveld	EVC/EVC2	ASD
Heterotaxy syndrome	CFC1/ZIC3	d-TGA with Heterotaxy
Alagille	NOTCH2/JAG1	PS, TOF
Holt-Oram	TBX5	ASD, VSD, conduction
William	Elastin	SVAS
Neurofibromatosis-Noonan	NF1	PS
Neurofibromatosis	NF1	PS
Marfan	TGFBR2 and Fibrilin 1/2	Aortic aneurysm

Table 2. Commonest syndromic CCM and their genetic aetiologies. PS: Pulmonic valve stenosis, SVAS: supravalvular aortic stenosis, ASD: atrial septal defect, VSD: ventricular septal defect, TOF: tetralogy of Fallot, TGA: transposition of great arteries, PDA: patent ductus arteriosus, PDA: patent ductus arteriosus, IAA: interrupted aortic arch.

Genetic aetiologies have also been found for multiple syndromes associated with CCM as a secondary feature like: Marfan (Fibrilin 1/ 2, TGFBR2), neurofibromatosis (NF1), Williams" (elastin), Alagille (JAG1, NOTCH2), Ellis Van Creveld (EVC) and CHARGE association (CHD7).

1.4.1.1. 22q11.2 deletion

Another well known form of syndromic CCM is the Di George syndrome, also called CATCH22, 22q11 deletion syndrome, and conotruncal anomaly face syndrome. The nomenclature of the syndrome is sometimes confusing, because although the clinical diagnosis of velocardiofacial syndrome (pharyngeal dysfunction, cardiac anomalies, and dismorphic facies) or Di George syndrome (cardiac anomalies, hypoparathyroidism, and immunodeficiency) is straightforward, only 35-90% of Di George patients and 80-100% of the velocardiofacial patients carry the 22q11.2 hemyzygous deficiency (41).

Estimates of the prevalence of the 22q11.2 deletion syndrome are around 1 in 4000 live births. The group of Goodship reported a prevalence of one in 3900 live births in the UK

(42). In the biggest population based study in Atlanta metropolitan area, USA, Botto et al. reported an overall prevalence of 1 in 5900 live births (43).

Of the recognized 22q11.2 deletion syndrome patients 49-83% present a form of CCM of which 17-22% is TOF, 14-15% interrupted aortic arch, 13-14% ventricular septal defect (VSD) and 7-9% truncus arteriosus (41). Most of the defects, not only the cardiovascular manifestations, are related to impaired development of fourth branchial arch (41). Pathophysiologically most of the defects seem to be related to haploinsufficiency of the TBX1 gene, despite the commonly deleted region containing 35 genes. This hypothesis is strongly supported by the phenotype exhibited by the TBX1 knockout mouse. In 2001 Lindsay et al published a letter to Nature describing a series of experiments comprising a nested deletion strategy at the mouse homologous region to 22q11.2. The conclusion was that TBX1 deletion was responsible for the phenotype observed at the mouse carrying deletion of the whole region. Phenotype was fully penetrant at E10.5 and consisted in defective development of fourth branchial arch arteries (44).

Also in 2001 Jerome and Papaioannou published a paper describing the phenotype of the TBX1 -/- mouse. At E11.5 17% of TBX KO heterozygotes showed either absent or reduced fourth branchial arch. No living homozygotes were found among the offspring of heterozygote mice. Nevertheless, cesarean sections found homozygotes in Mendelian ratios. TBX -/- were edematous indicating heart failure and showed abnormal facial structure, shortened necks and low set ears. All homozygotes showed single outlet outflow tract (PTA) and in some cases right or double aortic arch (45).

TBX1 is expressed in the secondary heart field which, as described below, is involved in the development of structures like the outflow tract (OFT) and the right ventricle. TBX1 seems also to be expressed at some cell populations at the anterior heart field fated to be cardiomyocytes in the OFT (46). These fated cells seem to be involved at the neural crest cell migration in OFT formation. The pathophysiology of the parathyroid and thymic defects associated with TBX haploinsufficiency has been extensively researched. TBX1 deficiency leads to decrease of proliferation of endodermal cells at branchial arches and therefore to defects at the structures derived from them. Endodermal proliferation depends on the activation of a transcription factor cascade that involves TBX1. TBX1 directly activates FGF8 and 10, factors involved in proliferation of surrounding cells; MYO1 and MYF5,

factors involved in the development of branchiomeric muscles; and other factors including PAX9, GCM2, GBX2, EYA1 or HOXA3 involved in cardiac, thymic and parathyroid development (47).

On the other hand, cardiac defect development in 22q11 deletion is not as well understood as the branchial defects are. It is, however, recognized that a similar proliferation factor cascade is truncated with TBX1 deficiency: Islet 1 (ISL1) regulates sonic hedgehog (SHH). SHH in turn enhances the expression of forkhead family members (FOX). FOX family members bind to TBX tissue specific enhancers. TBX activates FGF8 and 10 as previously described (48).

Perhaps the most striking feature of 22q11.2 deletion syndrome is the phenotypic variability. Despite the syndrome being located at the “genetic end” of the CCM spectrum, its phenotype is not uniform and it may be related to background- environmental-stochastic factors (41). This is the case argued by experiments showing that despite TBX1 knockout mice showing little or no thymic defects initially, when they were bred into different strains these defects became obvious and highly prevalent. Also, despite the fourth branchial arch defect being fully penetrant at early stages in heterozygotes only a few present the phenotype at birth (49). This background-environment- genetic interactions might explain the case of 22q11.2 deletion syndrome phenotypic differences between human populations (50).

1.4.2. Non-syndromic genetic forms of CCM

Despite most of the genetic causes of CCM remaining unsolved, a few predisposing allele have been consistently identified in syndromic forms of CCM. As considered before, it might be that these genes posses a simpler architecture with less allelic heterogeneity and fewer common associated alleles. These facts make these genes good candidates to start the search in non syndromic forms of the disease using linkage tools and family studies.

Among the commonest forms of non syndromic CCM is the bicuspid aortic valve. In 2005 Garg et al. described that premature stop codons at NOTCH1 segregated with bicuspid aortic valve and early onset valve calcification in two families (51).

Associations between genetic variants at Nkx2.5, atrial septal defects (ASD) and atrioventricular conduction block have also been described. Between 8-19% of familial ASDs and 1-4% of sporadic ASDs are related to mutations at NKx2.5 (52). Multiple groups have described associations between sequence variants at NKx2.5 and the phenotype (40). However, it remains impossible to predict the phenotype associated with the variants. Moreover, nonsense mutations at NKX2.5 have been also linked to other phenotypes in small samples, like hypoplastic left heart, VSDs or conotruncal defects. The phenotypic heterogeneity of the NKX 2,5 variants is not yet fully understood leaving open questions about the protein role and the interactions at cardiogenesis.

Other gene previously associated with heart septation defects is GATA4. Variants at GATA4 have been linked to ASDs, VSDs and AVSDs in families. Research at the variants, starting with the phenotypic similarity to the defects apparently caused by NKX2.5 and TBX5, has revealed that the variants seem to disrupt the interaction between these three proteins. The complex formed by NKX2.5, GATA4 and TBX5 is apparently a cornerstone for human heart septation (40).

A search done at the catalogue of published genome wide association studies (www.genome.gov/gwastudies) revealed that up to August 2010 there have not been publications showing results on GWAs exploring CCM.

In summary, predisposition to non- syndromic CCM have been extensively researched giving some insights, not only into CCM pathogenesis, but also at the embryological process of cardiogenesis. Nonetheless it is important to clarify that most of genetic predisposition to CCM is still unsolved and most of the genetic architecture, embryological implications and interactions are still unanswered questions.

1.5. Tetralogy of Fallot

1.5.1. Introduction

The defect now known as Tetralogy of Fallot was first described in 1671 by Stenson, but it first precise anatomical description was illustrated by William Hunter at St George's Medical School in 1784: "…the passage from the right ventricle into the pulmonary artery, which should have admitted a finger, was not so wide as a goose quill; and there was a hole in the partition of the two ventricles, large enough to pass the thumb from one to the other. The greatest part of the blood in the right ventricle was driven with that of the left ventricle into the aorta, or great artery, and so lost all the advantage which it ought to have had from breathing" (53). The modern definition of TOF was refined by Etienne Louis Fallot, but the term is attributed to the Canadian Maude Abbott in 1924 (54).

Tetralogy of Fallot is a congenital heart defect of the outflow tract characterized by right ventricular outflow tract obstruction, ventricular septal defect between the anterior and posterior limbs of the trabecular septal band, dextroposition (over-riding) of the aorta and right ventricular hypertrophy(55) Figure 3. The condition is associated often with a number of other cardiocirculatory anomalies like coronary malformations or pulmonary atresia(38). That is the reason why now we regard the condition as a family of diseases characterised by intracardiac anatomy but variable in terms of associated abnormalities, pulmonary anatomy and outcomes (54).

About 3.5% of all infants suffering CCM have TOF (54). TOF is the commonest cyanotic heart defect affecting nearly 0.24/1000 newborns worldwide, with both sexes being affected equally [2]. The calculated prevalence in the United Kingdom is nearly 0.31/1000 newborns(56).

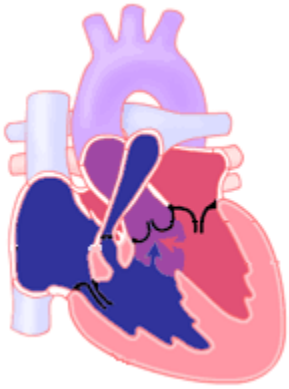

Figure 3. Tetralogy of Fallot Anatomy. Taken from the heart centre encyclopaedia (Public domain), www.cincinnatichildrens.org (57). Oxygenated blood is depicted red and de deoxygenated in blue. Note the shunt/mixture of deoxynated/oxygenated blood in the aorta consequence of the overriding position and the VSD. Pulmonary stenosis is also represented and is the determining factor for the right-left shunt.

1.5.2. Clinical features and management

The VSD in TOF is almost always non restrictive, with the pressure in both ventricles being equal. As a consequence the direction and magnitude of flow between ventricles depends on the degree of right ventricular outflow tract obstruction (54).

Diagnosis of TOF, as in many complex CCM, is frequently made during foetal life using ultrasonography. The diagnosis is only made when outflow tract images are obtained (58). However suspicion can be raised with abnormal views of the vessels at the upper mediastinum or a smaller diameter of the pulmonary artery. Nevertheless, most of patients are diagnosed during early life. The clinical presentation, as the degree of flow through the VSD, depends on the right sided outflow tract obstruction (58). The condition is suspected due to its clinical presentation characterised by: a variable degree of cyanosis, palpable right ventricular impulse and thrill; an early systolic aortic ejection sound; an usually single second heart sound; a systolic murmur which intensity and duration depends inversely on the severity of right ventricular outflow tract obstruction: with extreme degrees of obstruction the murmur is almost inaudible (32).

Before the development of surgical intervention, 50% of TOF patients died during early years of life with almost none reaching maturity (54). From 1955 the treatment for the condition was surgical, with the age of patients receiving primary surgical correction gradually decreasing (54).

Most centres offer primary surgical correction between 3-6 months with a transatrial, transpulmonary approach sparingly using RVOT patches. However, in some centres a staged approach consisting in first constructing a right to left shunt, and afterwards complete repair, is also offered (38). Detractors of the two staged approach argue about the consequences of the long lasting right ventricle pressure overload and the effects of sustained cyanosis.

However, it seems that depending on the expertise of the surgeons at the Blalock-Taussig shunt operation technique, the construction of these left to right shunts can favour growth of pulmonary arteries facilitating delayed repair (58). Irrespective of timing, ideal surgical repair should consist in closure of VSD, preservation of right ventricle structure without outflow tract obstruction and preservation of a functional pulmonary valve (54). Unfortunately, the nature of the RVOT obstruction usually makes this impossible.

Modern surgical repair techniques have shifted from a “complete obstruction relief” policy to sparing techniques where the pulmonary valve is preserved even at the cost of some residual stenosis (54). This trend has reduced the adverse late effects of pulmonary incompetence and aneurysmal dilation. Despite the great advances in surgical technique, intensive care and diagnosis, TOF patients survival rate is approximately 86% at 32 years (59). The 3 leading causes of late death are: heart failure, arrhythmia and complications of re-operation (60).

In the early postoperative period most children have an uneventful recovery. The most common complication is “restrictive right ventricle physiology”. This type of physiology is apparently related to the amount of right ventricle manipulation and transannular patch usage. The complication is characterized by a low cardiac output state and longer duration of intensive care stay and inotropic support (54).

Perhaps the most serious post surgical complication is pulmonary regurgitation. Until recently pulmonary incompetence was regarded as an unimportant consequence of surgical repair. Nevertheless, it has been shown that the amount of incompetence is directly related to exercise intolerance, ventricular arrhythmia, sudden death and right ventricular failure (54). The amount of pulmonary regurgitation can nowadays be measured by the gold standard cardiac magnetic resonance.

There is now a consensus that the timing of pulmonary valve replacement after TOF repair has been too late most of the time (58). Prevention of the late consequences of pulmonary regurgitation has shifted the opinion to accept earlier indication for replacement, either surgical or percutaneous, in the presence of significant pulmonary regurgitation (>30%) and: subjective or objective deterioration of exercise tolerance, right ventricle dilation, ventricular tachycardia, pregnancy contemplated, QRS duration >180 msec or if cardiac surgery is indicated for other reason (58).

There is also a trend to contemplate the effects of right ventricle dilation failure and dilation over the left ventricle. It seems that the poor left ventricular performance is related to biventricular dys-synchrony. Functional improvement has been reported with biventricular pacing and resynchronisation (61). It has been reported that 6% of surgically repaired TOF patients suffer sudden death in the long term (59). The risk of sudden cardiac death seems to be related to QRS duration but not to non sustained VT episodes during 24h ECG monitoring. Some groups have advocated to the use of implantable cardiac defibrillators in high risk patients (58).

1.5.3. Introduction to TOF genetics

Most cases of TOF seem to be sporadic. However, almost 20% of TOF patients show a syndromic form of the disease (39). In this group, the 22q11 microdeletion and velocardiofacial syndromes have been described several times. Johnson et al. described the presence of the 22q11 deletion in 14% of the probands in a cohort of Tetralogy of Fallot patients (62).

TOF can arise in the context of maternal infections, illness or exposure to teratogens, Boon (1972) calculated the heritability of the condition as 54%, with a recurrence risk in siblings of 2% for TOF and 1% for any cardiac defect (63). Digilio encountered a similar rate (64). Burn et al studied patients undergoing cardiac surgery at 12 centres in the UK (35). His group encountered 395 TOF patients who had 223 children with a recurrence rate of CCM (not only TOF) of 3.1%, which is around six times the general population risk. Risk was greater for mothers compared to fathers.

TOF have been linked to dominant mutations causing decrease of gene product dosage. These associated genes correspond to cardiac transcription factors (NKX2.5, TBX1, TBX5 and GATA4), transmembrane receptors NOTCH1 and NOTCH2, and their ligand JAG1. Cytogenetic anomalies like 22q11.1 deletion and trisomy 21 account for approximately 15% and 7% of TOF respectively.

In non syndromic cases of TOF, a number of genes have been linked with TOF in small number of patients. For example: Eldadah encountered 9 TOF patients carrying a mutation in the JAG1 gene, and Goldmuntz in 2001 described 4 TOF patients with mutations in the NKX 2.5 homeo-box. Other genes previously linked to the condition are: CSX, ZPFM2 and VEGF (64). However, in the majority of patients the causing factor is unknown, with some authors suggesting that TOF is a “complex” disease with a probable oligogenic origin (35).

In 2009 Greenway et al. published a paper describing the results of a genome copy number variant (CNV) survey on 121 TOF trios (proband and two unaffected parents) using an Affymetrix assay. De novo CNVs of more than 20Kb found were compared to those found in a 2265 controls cohort. CNVs present in more than 0.1% of the controls were discarded, the rest were validated using MLPA. At the end, 11 rare de novo CNVs were found in the TOF cohort and 4 were found among controls, the difference was not statistically significant. Analysis of inherited CNVs showed 3 rare CNVs, not present among controls, inherited from unaffected parents. Further analysis explored CNVs in 9 loci in 398 non syndromic TOF probands. A total of 17 rare CNVs not present in controls were identified. In summary 1% of non syndromic TOF sufferers presented a CNV on 1q21.1; additional CNVs were identified at 3p25.1, 7p21.3 and 22q11.2; in a single individual 6 CNVs were identified, two of them on genes previously reported to be

associated with TOF (NOTCH1 and JAG1). Accordingly to this research results 10% of sporadic TOF can be explained by rare CNVs.

According to the published research thus far, TOF genetic architecture might correspond to an entity with most of its predisposition being based on uncommon variants conferring a high risk to carriers. This would be the case if the lack of publications associating common variants to the condition, especially GWAs, corresponds to a biological fact and is not a consequence of methodology and sample collection.

1.6. Cardiogenesis

1.6.1. Morphological aspects of cardiogenesis

The heart is the first organ to form during development, and begins to pump at 3 weeks of gestation (40, 65).Cardiogenesis is a extremely delicate process and involves different cell types and the interaction of genetic and epigenetic factors (40). One interesting fact about cardiac development is its evolutionary conservation in terms of morphological and molecular keystones (66). Human cardiogenesis resembles phylogenetic heart evolution, and its molecular similarities and differences with other organisms establish a parallel between acquisition of new structures and functional specialization (66).

The process of heart development follows a conserved line of morphological changes among vertebrates. The cardiac progenitor cells have been identified in the anterior region of the embryo primitive streak. These cells of mesodermal origin migrate to each side of the midline, fusing and forming the structure known as the cardiac crescent (28, 65). At human embryonic day 21-23, the cardiac progenitors fuse forming a tube, start expressing cardiac markers and develop contractile function. Subsequently, the cardiac tube loops to the right and rotates in the anterior-posterior axis allowing the convergence between the outflow and inflow tracts. Cardiac cells undergo expansion, and the septation process begins forming mature chambers and the outflow tract. Finally, valves, coronary arteries, and conduction tissue are formed (65). Figure 4 shows a graphic summary of cardiogenesis.

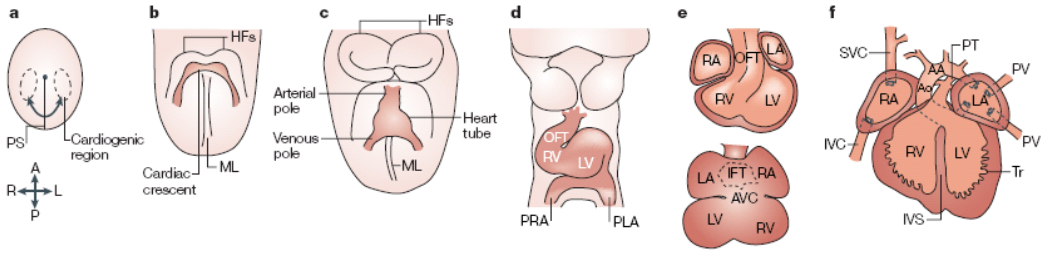

Figure 4. Graphic summary of the major stages in cardiogenesis. Taken with permission from Buckingham et al (65).

Scientific opinions about cardiogenesis have changed with time. Initially, heart development was believed to follow a tight spatial patterning of cell fate. Indeed, early experiments established correspondence between crescent regions and mature structures (65, 67). That was followed by the discovery of two group of cells named primary or first heart field (FHF) and secondary heart field (SHF), and the new goal was to build maps of cell fates based on those groups (28).

The finding of cardiac neural crest cells contribution, and new animal models have radically changed scientific opinion about cardiogenesis, showing overlapping of regions and difficulties defining zones (28) . Both heart fields express different markers which will be explored in following chapters (67). These differences allowed mapping the cells and identification of regional relevance. It is now believed that the primary heart field constitutes a general scaffold, and contributes essentially to the left ventricle (28). The secondary heart field which is formed by a group of cells in the anterior pharyngeal mesoderm, contributes almost ubiquitously to the chambers but is highly important to OFT and right ventricle (RV) development (28). The contribution from the NCC is mainly to the formation of the OFT and will be explored in the OFT section.

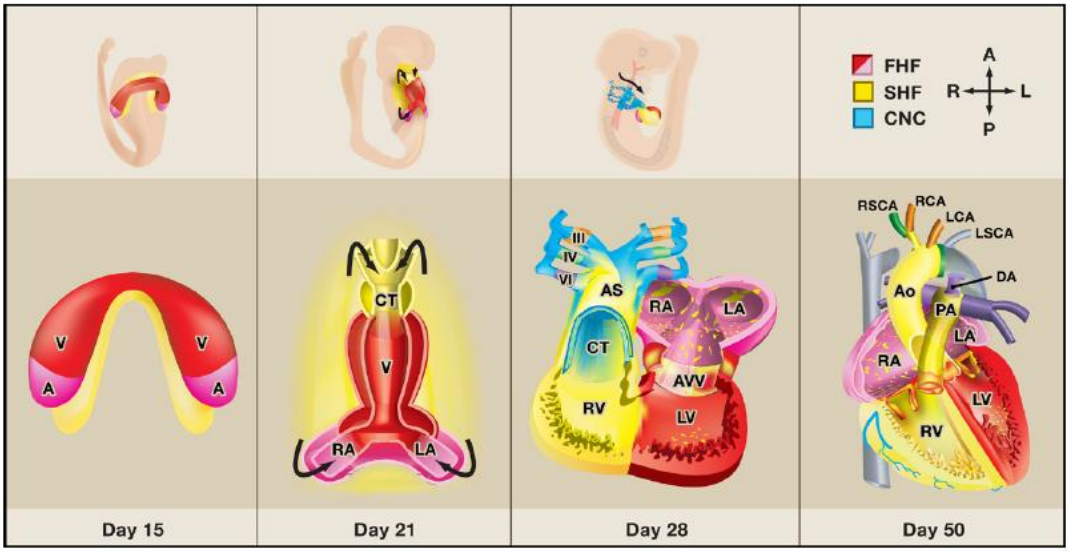

Figure 5. Heart development and contribution of Primary Heart field (FHF), Secondary Heart Field (SHF), and Cardiac neural crest cells (CNC). Taken with permission from Srivastava (2006) (28).

1.6.2. Molecular aspects of cardiogenesis

1.6.2.1. Introduction

Cardiogenesis refers to the cellular and molecular mechanisms that lead to the commitment of undifferentiated mesodermal progenitors into the cardiac lineage (68) and the formation of the complex structure we call the heart. This differentiation process responds to the influence of molecular signals within the mesodermal cells and from non mesodermal adjacent tissues.

Previous works in heart development and comparative genetics have shown a conserved gene regulatory network from simple heart like structures in invertebrates to complex chambered hearts in vertebrates. It has been therefore hypothesised that the addition of structures and increasing complexity of the heart are due to the expansion and modification of that genetic network.

The first heart structure is believed to have been developed by a billaterian approximately 500 million years ago. This heart structure was apparently a monolayer structure with peristaltic contraction similar to the ones encountered in tunicates. From that monolayered structure the heart evolved to be able to separate oxygenated from deoxygenated blood in a multichambered complex structure exhibiting synchronous contraction like the human heart (66). All this structure addition (valves, chambers, conduction system) corresponds not only to an increasing complexity of the genetic network controlling the process but also the acquisition of additional cell lineages within the process (such as the neural crest cells). Genetic evolution of cardiogenesis is thought to have been driven by three key processes: gene duplication, changes in cis regulatory sequences and mutation in the genes involved resulting in new function.

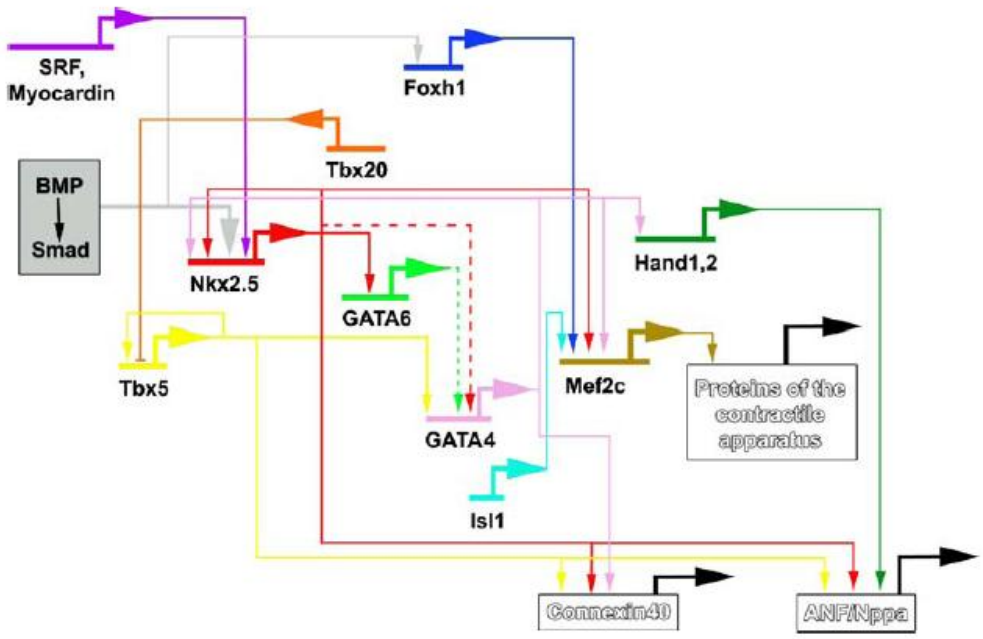

Figure 6. Scheme representing the basic transcriptional circuit for cardiogenesis in vertebrates. Taken with permission from Munoz-Chapuli R et al. Cardiogenesis: an embryological perspective. Journal of Cardiovascular Translational Research; 3(1):37-48

Heart development in vertebrates seems to be controlled by two groups of genes. The first is a core of evolutionary conserved transcription factors, including NK2, MEF2, GATA, TBX and Hand. These proteins regulate cell fate, expression of contractile machinery and morphogenesis, also regulating expression of each other and reinforcing the cardiac gene development program (66). Other accessory transcriptional factors also contribute to cardiogenesis. Disruption of each one of these core cardiac inductors does not completely suppress cardiac differentiation but rather alters the fusion in the midline of the heart fields, or impairs cell diversification, chamber formation and maturation (68). The second group of genes is an accessory group related to the function of the regulatory core and to the acquisition of new structures. As an example we can cite Isl2 which is a protein related to the secondary heart field and the expression of Mef2c (66).

The pattern of expression of the previously mentioned genes in the heart fields follows this line: cardiac induction; expression of specific activators and regulatory transcription factors; construction of contractile machinery; and finally growth and patterning control (66). It is also known that cis regulatory elements control the expression of different genes

in a spatially specific fashion creating specialized groups of cells among the cardiac progenitors contributing to the complexity of the developed heart.

As it is understood at the moment, the first step is the induction of the mesoderm to cardiac fate during gastrulation. This induction seems to be exerted by the endoderm through the expression of Bone Morphogenetic Proteins (BMPs) and Fibroblast Growth Factors (FGFs) (67). The activation of the cardiac progenitors by BMPs through the SMAD pathway induces the expression of Nkx 2.5, a domain homeobox ortholog to the *Drosophila* Tinman, expressed in cardiac mesoderm and the adjacent endoderm that serves as the signal to initiate cardiogenesis. Forced expression of Nkx2.5 in zebrafish was shown to induce size increase of the heart field and cardiac profile protein expression in fibroblasts (69). Dominant negative Nkx2.5 constructs can block cardiogenesis in Xenopus (66). However, targeted disruption in mice of Nkx2.5 disrupted cardiogenesis at a much later stage, with failure of looping at the linear stage but without compromise of contractile machinery expression (70).

Mef2 is a protein that belongs to the MADs- box family and is perhaps the most ancient myogenic transcription factor. Mef2 participates in the differentiation of muscle cell types with the subtype MEf2c being implicated in the expression of specific cardiac proteins and with the regulation of gene transcription in the secondary heart field. The GATA family of genes regulate the expression of contractile protein genes and upstream genes including Nkx2.5 and Hand (66). Three of the six GATA genes in vertebrates are expressed in the developing heart. Forced expression of GATA4 in *Xenopus* results in premature induction of cardiogenesis and more importantly have been implicated in non syndromic forms of CCM in humans (See section 1.4.1).

Seven genes of the Tbx family are so far known to be expressed in the mammalian developing heart. Mice lacking Tbx5 present defects in the posterior region of the heart tube. Tbx5 disruption using dominant negative constructs cause failure to develop early cardiac heart fields in *Xenopus.* Additionally, defects in TBX5 are the cause of Holt Oram syndrome in humans (40). The transcription factor's action leads to the expression of muscle genes and genes controlling growth and patterning (like our pathway of interest: the Wnt family of proteins).

Table 3 shows some of the key genes with their expression pattern and mutant phenotype in animal models.

When the cardiac precursors initiate differentiation, specific activators of the heart fields induce the expression of the core transcription factors. In the primary heart field the activators are genes NKx2.5 and GATA4, whereas in the secondary heart field they are Isl1 and Foxh (28, 65-67). At this stage, the anterolateral population of cardiac progenitors expressing Nkx2.5/ BMP are called the primary heart field, and the posteromedial region of progenitors expressing Isl2/ FGF8 are known as the secondary heart field. Not only activators but also inhibitors like the Wnt canonical proteins family are essential to this stage of development (71).

Gene	Mutant phenotype	Cardiac Tube	Secondary field	Crescent expression
Nkx2.5	Single atria and ventricle. No Hand1	Yes	Yes	Yes
Tbx5	Hypoplastic LV, Sinoatrial defects	Left Ventricle, Atrium, Infundibulum	No	Yes
Hand1	LV disrupted	Left ventricle	No	Yes
Fgf10	No	No	Yes	No
Fgf8	Outflow tract defects	No	Yes	No
Tbx1	Outflow tract defects	No	Yes	Yes
Isl1	Single atria and ventricle. Atrial venous pole abnormal	All	Yes	No
Foxh1	Outflow tract defect absent, Right ventricle does not develop	No	Yes	No
Hand2	RV abnormalities	Right ventricle	Yes	Yes
Tbx20	Chambers don't develop, Outflow tract disrupted	Outflow tract, ventricles	Yes	Yes

Table 3. Relevant genes in cardiac embryogenesis, expression and mutant phenotype. Adapted from Buckingham et al (65).

1.6.2.2. The heart fields and the outflow tract

The first experiment describing a second population of cells that were added late in cardiogenesis, after the formation of the heart tube, and contributed to the outflow tract formation, was made as early as 1977 by de la Cruz et al. (72). In this classic experiment an iron oxide marking technique was done to the most cranial part of the chick embryo. If the labelling was made at HH12 the region was incorporated into the right ventricle trabeculae. Labelling at stage HH22 identified the marked region in the myocardium below the semilunar valves. Nevertheless, it was only in this century that the molecular bases of OFT development were discovered.

The development of the OFT is, like cardiogenesis as a whole, a complex process involving the interplay of genetic and environmental factors in which different groups of cells have a key role. The OFT is the final result of remodelling a tubular structure called the conotruncus (73). This process includes the formation of cardiac cushions, migration of endothelial cells, myocardialization of the endothelial cells, fusion of the cushions, septation and remodelling (73).

In 2001 three different groups identified, using different experimental approaches, a novel cell population to the FHF that contributed to the OFT formation that was named the anterior heart field. Kelly et al made an experiment with a FGF 10- nlacZ reporter mouse determining that the right ventricle and the OFT are derived from the splanchnic mesoderm and the pharyngeal arch core in stages E8.25-E10.5 (74). Mjaatvedt labelled cells anterior to the heart tube using mitotracker or an adenovirus and later found them in the OFT. This same group ablated the bilateral heart fields anterior to the heart tube resulting in development of a rudimentary heart (75). Wal et al., after indentifying expression of GATA4 and Nkx2.5 in the mesoderm anterior to the heart tube, labelled the region, finding the cells in the proximal OFT (76). Kirby et al. also had shown in 1983 that removal of mesodermal cells anterior to the heart tube produce a failure in outflow tract septation (77).

Another key discovery in the understanding of cardiac embryogenesis and particularly the embryogenesis of the OFT was the description of the pattern of Isl1 expression. Isl1 is expressed in the rostrolateral mesoderm of the chick embryo and its expression is lost in the fusing differentiating myocardium remaining at the undifferentiated regions (78).

Reporter experiments of Isl1 show expression of the gene in the OFT, the atria and the right ventricle. Cai et al reported in 2003 the cardiac phenotype for the Isl1 KO mouse, which roughly is characterized by absence of OFT, part of the right ventricle and atria (79), regions that correspond to the previously labelled expression domain. This group of experiments showed that there are two progenitor cell lineages separated by time; the first and the second lineages. The primitive left ventricle seemed to be exclusively derived from the first lineage whereas the right ventricle was from the second one. In the rest of the cardiac chambers the two lineages intermixed. These cell lines were not spatially identifiable in the heart fields but the first seemed to undergo early differentiation while the second remained undifferentiated and contributed late to the arterial and venous pole of the developing heart (78).

It seems that the heart fields are not different group of cells but a unique region separated by spatial and temporal patterns of expression, the terms of first, second, anterior and secondary heart field are useful. Summarizing, the first heart field refers to the first group of cells that differentiate and express muscle specific proteins. These cells are the first to down-regulate Isl1 expression. The second heart field are the cells added to the cardiac poles when the heart tube is formed. The anterior heart field corresponds to the cells labelled with a Fgf10-lac reported and which exclusively contribute to the OFT and right ventricle. Finally, the secondary heart fields refers to the splanchnic mesoderm caudal to the OFT that gives rise to the most distal part of the OFT and the media of the great arteries (78).

Importantly, addition of myocardium to the OFT derived from the secondary heart field adds length to the structure. This added length allows OFT to rotate in order to anneal with the respective ventricles. Failure in this key process, as illustrated by the results of ablation of the secondary heart fields, produces mis-alignment of the aorta and pulmonary atresia-stenosis, a phenotype that closely resembles TOF (80).

1.6.2.3. The cardiac neural crest cells

An important element in the OFT formation is the population of cardiac neural crest cells (CNC). The first article describing the contribution of this cell group to cardiogenesis was

published in 1983 (81). Early experiments were performed using ablation and quail-chick chimeras. The role of this group of cells was revealed when ablation of premigrating CNC showed outflow tract defects and hypoplastic or aplastic parathyroid, thymus and thyroid glands (81). Strikingly the phenotype derived from CNC ablation closely resembles DiGeorge syndrome in humans.

From the seminal ablation works, experiments using fate mapping of the CNC determined their contribution to cardiogenesis in mammals. Waldo et al. used a Cx43 (gene expressed at CNC)- β galactosidase expressing transgenic mouse showing CNC presence within the pharyngeal arches, the OFT and ventricular myocardium (82). Subsequent revision of the matter showed that Cx43 can be also expressed in myocardial cells. Several works have performed fate mapping of CNC using Cre- Lox models. MacMahon et al. produced a transgenic mice using the Wnt1 promoter regulating expression of Cre in CNC cells (83). The model was used by the group of Sucov et al to fate map CNC. Labelled CNC were identified invading the aortic arch and the pharyngeal arches at E10. CNC were found invading the OFT by embryonic day 10.5 and contributing to the aorto-pulmonary septation(83).

During the last decade multiple CCM mouse models have been described, with many of them showing defects in CNC development. Most of our knowledge of the embryological involvement of these cells in cardiogenesis and the genetic pathways involved are derived from these models. (78). For a summary of the pathways implicated in CNC development see Table 4. Bone morphogenetic proteins are members of the transforming growth factor beta superfamily known to be important in CNC development. BMPs have been shown to be key regulators during morphogenesis in general and cardiogenesis in particular. Kaartinen reported in 2004 a Cre/Loxp experiment specifically disrupting Alk2, a BMP receptor, in CNCs. Disruption of the receptor resulted in failure of OFT septation and differentiation of CNC derived muscle (84). Disruption of type II and 1A BMP receptors also result in outflow tract defects (83).

Notch receptors are transmembrane proteins involved in a wide number of cell specification and differentiation processes. The activation of the pathway leads to a proteolytic cascade resulting in liberation of the Notch intracellular domain that translocates into the nucleus activating the expression of downstream targets as the basic

helix-loop-helix transcription factors (83). As previously mentioned in the section on syndromic CCM, mutations in Jagged1 are the cause for the Alagille syndrome. In mouse inactivation of Jagged1 causes severe vascular defects, and double heterozygotes for Jagged1/Notch2 show severe cardiac abnormalities including pulmonary stenosis. However, despite the high likelihood of these defects being linked to the CNC migration, there is still lack of evidence in the matter.

Family	Genes
Bone morphogenetic proteins	Alk2, Bmpr1a, Bmpr2, Msx2
Transforming growth factor beta	TGFβ2, perlecan
Fibroblast growth factors	Fgf8
Notch, Jagged	Jagged1, Notch2, Presenilin1, Ht2
Wingless/Wnt	Dvl2, Wnt1, Wnt3a, APC
Endothelins	EdnrA, ECE-2, ET-1
Semaphorins, neuropillins, plexins	Npn1, PlexinA2, Sema3C, PlexinD1
Neurotrophins	Trk-C, Neurotrophin 3
Vascular endothelial growth factos	VEGF165, Hif1-alpha
Transcriptional regulators	Ap2α, Cited2, Foxc1, Foxc2, FoxD3, Nkx2.6, Pax3, Sox9, Tbx1
Adhesion molecules	NCAM, N- cadherin, Connexin-43

Table 4. Summary of genes and pathways implicated in neural crest development. Adapted from Stoller J, Epstein J. Cardiac neural crest. Seminars in cell and developmental biology 2005; 16:704-715

1.7. The WNT pathway

1.7.1 Introduction to the Wnt pathway

The *wnt* family of genes is a highly evolutionary conserved group of proteins characterized by their homology with the wingless gene in *Drosophila*, the first described member of the family (85). This family of secreted lipid modified glycoproteins comprises 19 genes in the mammalian genome (71), controlling numerous processes like cell migration, cell fate and cell proliferation or apoptosis (71, 86).

Initially, it was thought that this group of proteins exerted their function via a unique pathway. However, nowadays we recognize at least 2 functionally distinct groups of interacting proteins: the canonical or β catenin dependent and the non canonical or β catenin independent pathway (85). As a consequence, we can divide the *wnt* proteins into two functional groups: the first one comprising Wnt1, 3a and 8 which trigger just the canonical pathway, and the second one comprising Wnt4, 5a and 11 which in turn activate the non-canonical pathway and seem to have inhibitory properties over the canonical one (71). However, due to the high degree of interaction between the pathways some models are attempting to unify the *wnt* signalling as a whole interacting signalling pathway (71, 87)

Equally, the importance of the *wnt* inhibitors during development has been recognized. Two big groups of these molecules are distinguished: the secreted frizzled related proteins (sFRP) and the Dickkopf proteins (Dkk) (71).

1.7.2. The canonical pathway

Within the canonical pathway, the signal is started by the coupling of the Wnt protein to a Frizzled receptor (Fz). The activation of the Fz receptor and its Lipoprotein related protein co receptor (LRP) activates one of the proteins that will be explored during this research, Dishevelled, inducing disassembly of the complex between Glycogen synthase kinase 3 (GSK3), Axin and Adenomatosis Polyposis Coli (APC). This process blocks the

phosphorylation of β -catenin, decreasing its degradation. The accumulation of β catenin allows its interaction with the LEF/TCF family of transcription factors. The complex translocates to the nucleus and regulates gene transcription trough interaction with TCF/LEF transcription factors (71, 85). Certain target genes such as c-myc and cyclinD1 allow the canonical pathway to regulate cell proliferation. Latest experimental results have linked protein Gβq and Gβo subunits in canonical function (88). See Figure 7

1.7.3. The non- canonical pathway

Research in the non-canonical Wnt signalling, or Planar Cell Polarity Pathway (PCP), started in 1983 when Gubb et al described a small set of genes controlling polarity of hairs and bristles in *Drosophila (89)*. Studies of PCP have included multiple model organisms and a summary of the most relevant findings will be summarized in this section.

PCP research in *Drosophila* has been based in the orientation of eye and wing structures. Interestingly mutations in the PCP genes do not cause random distribution of the hairs in the wing but rather patterns, for example one resembling impressionist brushstrokes and motivating the mutant name *Van Gogh*. There are approximately 10 core PCP genes in *Drosophila*. The accepted sequence of activation of the noncanonical Wnt pathway follows this line: Wnt proteins activate Frizzled receptors which in turns recruits the adaptor protein Dishvelled (Dvl) as in the canonical pathway (71). Dvl activates, in a way still poorly understood, a Jun-Rac-Rho kinase pathway controlling cytoskeleton dynamics (89). The mechanism explaining how neighbouring cells communicate PCP signals and how global organism polarity is determined is still under research. The most widely accepted theory is that two Cadherin proteins (Dachsous and Fat), together with a Golgi protein called Fj, set up a global polarity signal that is translated by the asymmetric assembly of a protein complex including Fz receptors, Vang, Dvl, Stan and intracellular adaptor protein *Diego*. Interestingly Wnt protein loss of function and overexpression experiments have failed to produce PCP phenotypes in *Drosophila*, a fact that might be explained by functional overlapping between the Wnt proteins.

There are 5 processes known to be PCP controlled in vertebrates: convergent extension, neural tube closure, eyelid closure, hair bundle orientation and follicle orientation in the

inner ear cells. Additional to the previously mentioned PCP gene core, derived from the research in *Drosophila,* there are also a number of associated proteins that are essential to the PCP pathway involving for example transporters and chaperones. All genes that play a role in PCP signalling are homologs to one of the core *Drosophila* PCP genes (89).

Convergent extension (CE) movements correspond to medial migration and intercalation of mesodermal cells, movements that are controlled by lamellipodia formation. These CE movements can be disrupted in Zebrafish by mutation of Vangl2 (*trilobite),* prickle, or wnt11 (*siberblick)* and in *Xenopus* by disruption of any of the core PCP genes. For example disruption of Dvl in *Xenopus* produces a failure in CE movements (89). In mice loss of Vangl2, Celsr1, Ptk7 and simultaneous loss of two of the Dvl genes leads to neural closure defects, shortened embryo and CCM (90, 91).

The accumulation of the non canonical effector proteins in certain subcellular regions lead to the activation of the small G proteins Rho and Rhac, Phospholipase C, Protein Kinase C, and intracellular calcium release (87, 92). This complex interaction of processes control cell polarity, adhesion and motility through the action of the Rho induced kinases (85, 92) that results in control of cell elongation, cell body orientation and protrusive processes (93). Expression of dominant negative forms of Rok2 decreases dorsal cell migration in *Zebrafish (93).* Also Rok2 overexpression in *siberblick* Wnt11 zebrafish mutants suppresses the defects shown by the mutation, suggesting a regulatory role in the pathway downstream of Wnt11.

Finally, transcriptional changes also take place due to the modulation of the JNK transcription factor (71, 85, 92, 94). Rac binds the domain of Dvl and directly activates Jnk. In *Xenopus* morpholino knockdown of Jnk impairs CE movements, and produces cell adhesion defects (93).

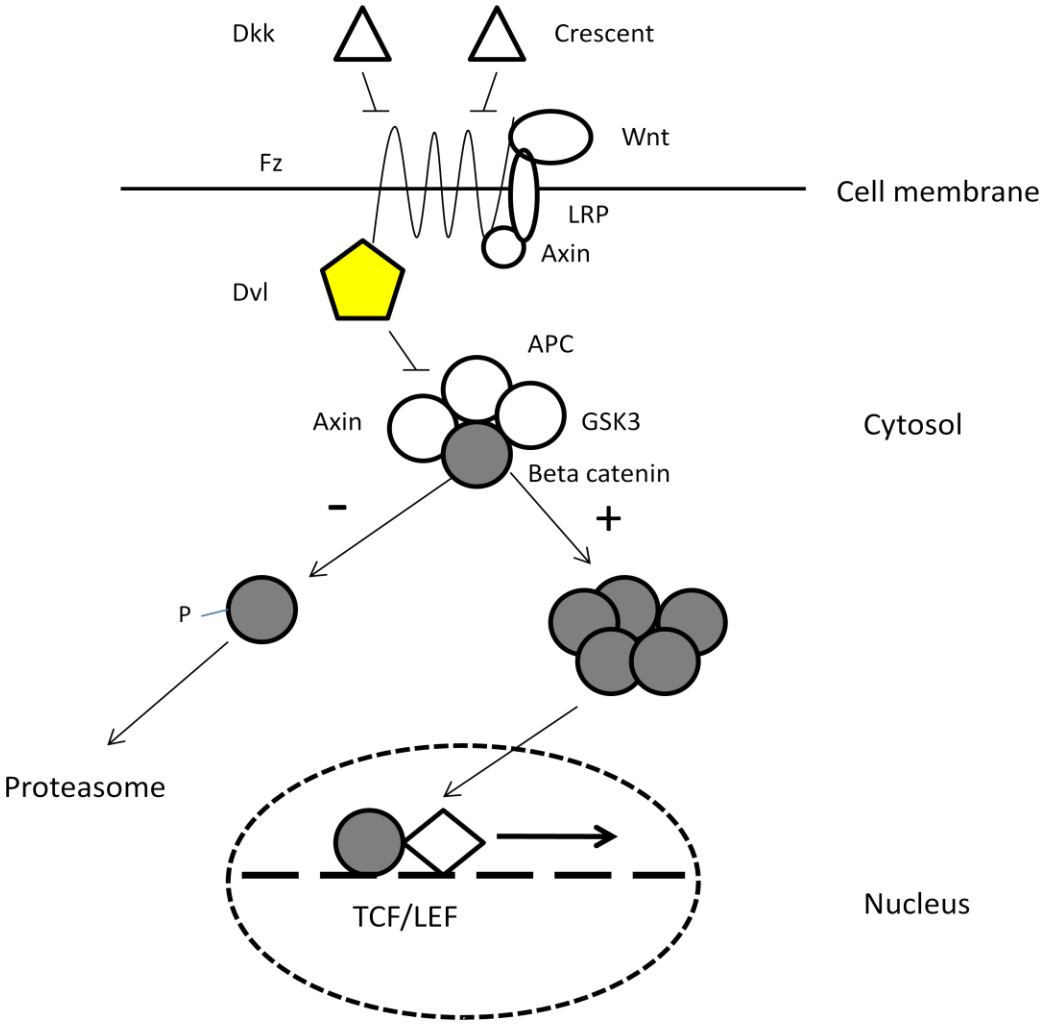

Figure 7. The canonical *Wnt* pathway. Colour filled figures represent the genes involved in this research. In yellow dishevelled.

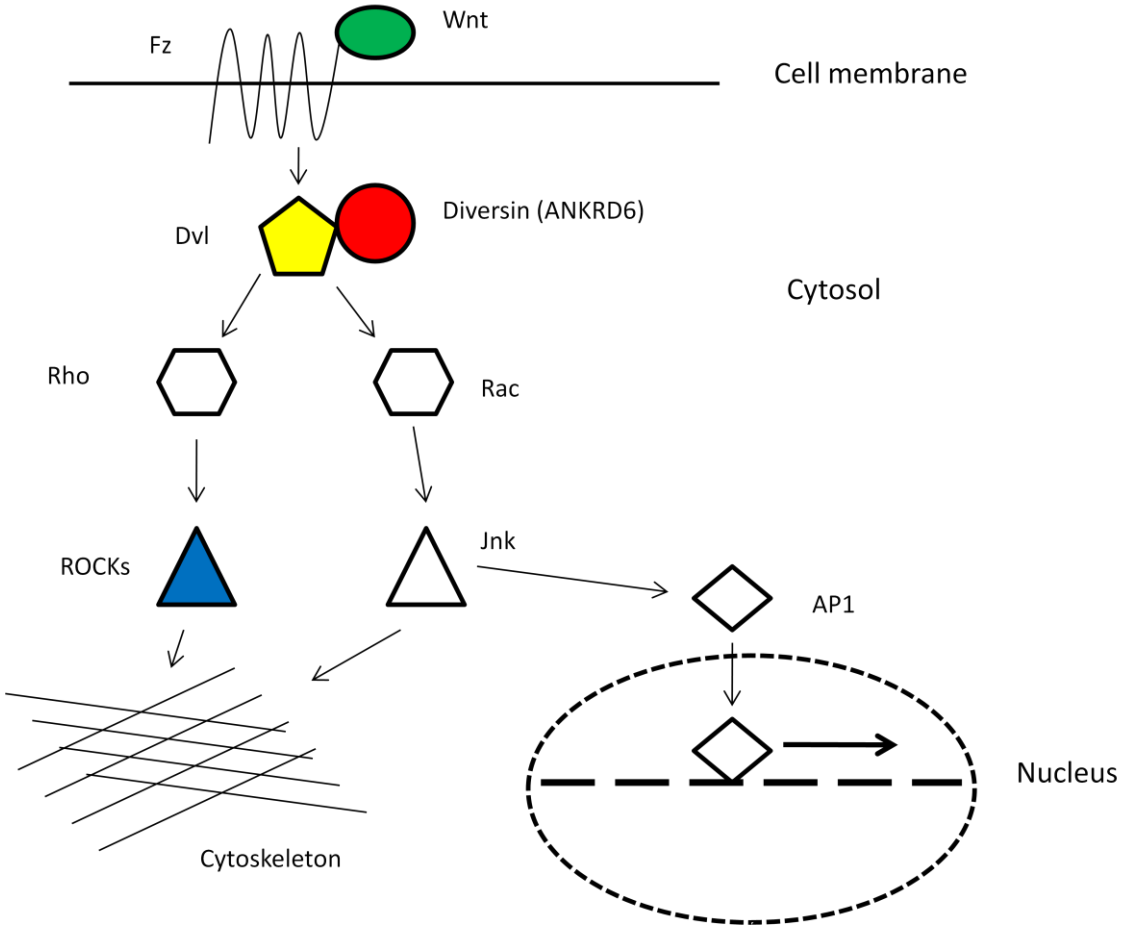

Figure 8. The non-canonical *wnt* pathway. Colour filled figures represent the genes involved in this research. In green the Wnt protein; In yellow dishevelled; In red ANKRD6 or its mouse ortholog Diversin; And in blue Rho associated kinases.

1.7.4. Wnt signalling and body axis patterning

One of the main processes in development is the regionalization of embryo sections in two axes, dorsoventral and anteroposterior, providing an orthogonal system for regional patterning. Wnt signalling, canonical and non canonical, is key for axes development. Research into body axes development has shown some conserved features among species, like the Hox gene expression, but also has revealed that this process has unique features between species (95).

Early processes in axis configuration can vary widely among species due to selective conditions like internal or external development. The basic body plan and axis of vertebrates is established during gastrulation by coordinated movements that originate the three germ layers (96). Among these movements, convergent extension is key to the correct configuration of the axes. One of the most important pieces of evidence of Wnt signalling being involved in axis configuration are the phenotypes of known mutants as *silberblick (slb), pipetail (ppt) and trilobite (tri)*. All of these previously mentioned mutants show shortened bodies with evidence of defects in axis development due to defective convergent extension movements. The mutated genes responsible for these three mutants, Slb, ppt and tri are part of the non canonical Wnt signalling in Zebrafish.

In developmental terms, polarization of embryos can occur as early as egg fertilization. Nowadays is recognized that polarized expression of genes in the visceral endoderm directs the orientation of antero-posterior axis in vertebrates (97). Formation of the primitive streak due to mesenchymal transformation of endodermal cells in the epiblast define, perhaps for the first time, the anterior and posterior poles of amniote vertebrate embryos. Asymmetric expression of Wnt3/Cer1-Lefty1 is involved in the migration of visceral endoderm and organization of the axis. Wnt3 expression marks the posterior embryonic region with the anterior visceral endodermal cells migrating towards regions that express Cer1-Lefty1 (97). Wnt3 and β catenin KO mice fail to form the primitive streak (98, 99). Dorso-ventral axis determination, in a simplistic view, apparently depends on a gradient of Bmp proteins expression that takes different forms depending on the organism

(95). For example in frogs, Wnts expression results in the formation of the Spearman's organizer at the dorsal region. The organizer express BMP inhibitory proteins.

Despite that in mammals the left and right sides of the embryo are determined when the primary A-P and D-V axes are developed, the first event showing left right asymmetry in the embryo is the leftward fluid flow at a cavity called the node (99). More evident asymmetry starts to be shown at 6-8 somite stage when the embryo begins to rotate along its rostro-caudal axis. Axial rotation is accompanied by ventral closure. During this rotation organ situs start to become apparent and left right differences between paired organs emerge. Differences in paired organs are intrinsic to the organ itself with explanted organs maintaining their developing pattern (100).

Numerous human situs defects have been reported ranging from single organ to complex conditions like the immotile cilia syndrome. In mouse the *inversus viscerum* mutation causes randomization of heart looping and body situs whereas the *inversion of embryonic turning (inv)* mutation show complete inversion of body situs and turning (100). It seems that early disruptions cause randomization or complete inversion of situs. The insertion that causes the *inv* phenotype disrupts the coding sequence of the *inversin* gene, an ortholog of ANKRD6 and a key signalling protein of the non canonical Wnt signalling.

To account for the directionality of the left right axis development Brown and Wolpert proposed a theoretical model based on a chiral molecule (101). Lately, research has focused on cilia, which contain a chiral motor structure. Evidence of ciliar involvement in laterality came from a human disease model, Kartagener syndrome, characterized by inmotile cilia and high incidence of left right inverted organs (101). The first laterality sign, the leftward fluid in the node, due to coordinated ciliar movement, must be transmitted to subsequent stages of development. In that sense, differential expression of genes at each side of the midline directs the asymmetric development of the left-right axis.

A conserved characteristic among vertebrate embryos is the left-sided expression of *Nodal* at the left side of the midline as a consequence of repression of Shh expression at the right side and therefore up regulation at he left one. There are several proteins downstream of nodal that are differentially expressed, making possible left-right

asymmetry development, like LEFTY1 and LEFTY2 or Pitx2. A deep review of this pattern of protein expression is over the scope of this introduction.

1.7.5. Wnt signalling and early cardiogenesis

Cardiogenesis is a complex process that can be subdivided in overlapping phases such as linear heart tube formation, cardiac looping, chamber formation, septation and maturation. In all these phases Wnt proteins, Fz receptors and Wnt inhibitors have been shown to be expressed in a consistent fashion in different model organisms (102) (85, 102). Eleven WNT orthologs have been encountered in the heart forming fields and early cardiac tube of mouse, chicken, quail and *xenopus* (85, 103, 104). Also Wnt 3a, 5a, 5b, and 8c are expressed in the primitive streak of avian and mouse embryos (71). Wnt 8, 2b, 5a, 7a and 14 have been shown to be expressed on the tubular heart of chicken embryos (71, 85). Also the mouse receptors Fz-2, 4, 9, the *Xenopus* receptors Fz 7, 8, 10a, 10b, and the human Fz receptors 1, 2, 7, 8 and 9 have been found within the heart fields mesoderm, cardiac neural crest cells and adult heart (85).

Works in the function of the canonical *Wnt* pathway in early cardiogenesis were initially confusing. A positive involvement of the pathway had been described in *Drosophila* whereas inhibition of the pathway in *Xenopus* using Crescent or Dkk seemed to induce cardiogenesis (105). The endothelial specific disruption in mouse of β catenin resulted in the formation of multiple heart forming fields (102). This particular conflict seems to have been answered by later experiments showing that canonical Wnt signalling can exert different effects depending on the time and location of action. In 2007 Ueno et al. published an experiment showing that shock inducible *Zebrafish* transgenic embryos expressing Wnt8 before gastrulation resulted in more cardiomyocyte formation while expression of Wnt8 after gastrulation resulted in the complete opposite (106). Some late experiments suggest that even this temporal biphasic model is too simple with different progenitor cell populations showing different responses to canonical Wnt signalling depending on their differentiation status and their molecular make up (102).

It is important to notice the relationship between mesodermal specification and β catenin signalling. Huelsken published in 2000 the phenotype of the β catenin KO mouse showing an early defect of anterior-posterior axis formation with complete absence of subsequent mesodermal tissue formation (99). In the same direction Wnt3 -/- mouse shows no primitive streak, failure of mesoderm formation, and failure to develop a posterior-anterior axis (98).

Mesp1 is the earliest cardiovascular lineage marker. The gene is upstream of Isl1 among other genes. Another important marker is Flk1, a tyrosine kinase receptor that can, in a time dependent fashion, induce the differentiation of cardiac progenitors formation. Consistent with the early mesodermal positive influence of Wnt/β catenin signalling influence on early mesodermal commitment, depletion of this pathway results in absence of Mesp1 expression and downregulation of Flk1 (102).

A number of observations also link canonical Wnt signalling to roles in the differentiation of cardiac progenitors in the SHF. Tissue specific down regulation of β catenin leads to a reduction in Isl1 cells in this field and the loss of SHF derived structures (102). Moreover, chromatin immunoprecipitation experiments have shown that β catenin binds directly and positively regulates the promoter of Isl1 (102).

With respect to the non canonical Wnt pathway, the best investigated factor is Wnt11. In *Xenopus,* loss of function experiments have shown that Wnt11 is essential for cardiogenesis and that its overexpression can cause contractile tissue formation (107). Loss of function experiments in *Xenopus*, artificially blocking non canonical ligands such as Wnt 11 or proteins like Dishevelled lead to defects in heart tube formation including cardia bifida (103, 107). Also important is the antagonistic effect at this early stage between Notch signalling and canonical Wnt signalling. It has been described that Notch4 activation can redirect Flk1 + cells fated to become blood progenitors into the cardiac lineage, an effect that is dependent on the expression of secreted Wnt inhibitors as Sfrp1 and Sfrp5. This effect was abolished with the addition of Wnt3 but not with Wnt5a (102).

Experiments in embryonic stem (ES) cells have also shown the relevance of non canonical Wnts in cardiogenesis. In ES cells, as in endothelial progenitor cells (EPCs), a cardiogenic program can be triggered trough the induction of Notch signalling. In turn, Notch signalling

upregulates Wnt ligands, specially Wnt5a (102). Also Wnt5a seems to be upregulated in response to Mesp1. It is however interesting that the Wnt5a KO mouse reported by Tai et al and Yamaguchi et al. do not exhibit a cardiac phenotype (108, 109). Downstream identity of the proteins involved in the cardiac induction exerted by the non canonical ligands Wnt11 and Wnt5a has also been researched. The activation exerted by Wnt11 can be inhibited trough the inhibition of PKC and JNK in *Xenopus (102).*

In summary, during early cardiac differentiation Wnt involvement can be elegantly described in 4 phases as in the paper of Gessert et al.: 1. Wnt β catenin signalling is needed for mesoderm/endoderm formation; 2. For cardiac specification, canonical signalling must be low. Canonical inhibitors as Srpfs and non-canonical Wnt pathway ligands as Wnt11 and Wnt5a might be involved in the process; 3. Canonical signalling is needed for the proliferation of certain cardiac progenitors, especially in the SHF; 4. Finally, for cardiomyocyte final maturation canonical signalling should be low. Non canonical signalling is involved in the process (102).

1.7.6. Wnt signalling and the formation of the four chambered mature heart.

The *wnt* signalling pathway has been shown to be involved in later stages of cardiogenesis in the formation of OFT, cardiac valves and the conduction system(85). Cardiac morphogenesis is a complex process involving tube formation, looping, chamber formation and heart maturation, all these processes require regulation of cell adhesion. Mediating this cell to cell interaction a key protein is N-cadherin, a calcium dependent adhesion molecule. It is widely accepted that non canonical Wnt signalling regulates N-cadherin mediated cell adhesion. Wnt11 KO mice show OFT defects and thinner ventricular walls, and exhibit defects in the localization, organization and intercellular contacts mediated by N-cadherin (110). Loss of function experiments in *Xenopus*, artificially blocking non canonical ligands such as Wnt 11 or downstream proteins such as Dishevelled lead to defects in heart tube formation like cardia bifida illustrating the relevance of the non canonical Wnt signalling for cell adhesion, heart tube formation and looping (103, 107).

As previously described cardiac looping is followed by the process of formation of the 4 heart chambers, process that includes: ballooning of the atria and ventricles, trabeculation of the ventricles, septation and valve formation. Expression analysis in cardiac cushions have identified active canonical signalling gene expression (102). Loss of APC function in zebrafish leads not only to defects in heart looping but also to excessive endocardial cushions due to increased proliferation and endocardial-mesenchymal transition. In contrast, overexpression of APC or inhibition of Dkk, inhibits cardiac cushion formation (111). Also Wnt9a and the inhibitor Frzb1 are expressed in chick cardiac cushions. Wnt9a is found in endocardial cells while Frzb1 is in cells undergoing EMT, showing the proliferative effect of Wnt9a canonical signalling and the inhibitory, pro non canonical influence of Frzb1 (102).

As described before, CNCs are implicated in OFT development. Wnt1 and Wnt3a are expressed in CNCs. The CNC cell type specific proliferation has been linked with β catenin dependent expression of regulators like Pitx2 (85). Cell proliferation seems to be upregulated by this signal. Loss of Wnt1 in an experiment with Wnt1-Cre in mice showed reduced proliferation of CNCs (102). In 2001, the group of Brault demonstrated that mice lacking beta-catenin, even in the presence of activating signals, show defective NCC proliferation (112). Experiments by Kioussi et al. showed that inhibition of β-catenin produces OFT malformations due to reduced CNC proliferation. Also, Dvl loss of function lead to OFT tract malformations related to NCC proliferation defects (113).

In terms of the non canonical pathway Wbt5a and Wnt 11 are expressed in the OFT myocardium. Disruption of the non canonical *wnt* pathway in the Vangl2 -/- mice known as loop tail (Lp), produces profound OFT malformations (114). The Vangl2 protein is co-expressed with Wnt-11 and Dvl2. Coexpression of ROCKs and Vangl2 and the polarization of migrating cardiomyocytes is disrupted in the Lp mouse (114). In the same way, Wnt5a loss of function results in OFT septation defects. Schleiffarth et al. showed PTA associated to a mutation in Wnt5a in mice. Also this group described the expression of Wnt5a in the pharyngeal mesoderm adjacent to CNC cells and in the myocardial cell layer of the conotruncus at the time when CNC cells begin to form the aortopulmonary septum (115).

The relevance and functional implications of *wnt* pathways in human development is still under research. Part of the objective of this work is to link the available information to human cardiac phenotypes.

Gene	Ortholog	Organism	Model	Phenotype
Diversin	ANKRD6	Zebrafish	Dominant Negative	Cardia Bifida
Apc	APC	Mice	Knockdown	GI tumors
			Knockdown Homozygous	Lethal
		Zebrafish	Knockdown	Looping failure, valve abnormalities
Vangl2	VANGL2	Mice	Knockdown	Outflow tract
RhoA	RHOA	Chick	RNAi	Cardia Bifida
		Mice	RhoA Inhibition	Defects in looping, septation and maturation
Fzd3	FZD3	Mice	Knockout	Neural defects, no mention to heart
Fzd4	FZD4	Mice	Knockout	Cerebelar and esophageal defects
Fzd5	FZD5	Mice	Knockout	Vasculogenesis defects
Rock2	ROCK2	Mice	Knockout	Placental malformations. Note: Cardiac expression
Rock1	ROCK1	Mice	Inhibition	Looping and chamber formation defects
Wnt11	WNT11	Xenopus	RNAi	Fusion, cardia bifida, duplication, thickening of Myocardium
Wnt1	WNT1	Mice	RNAi	Pericardial edema, ventricle enlargment, narrowing outflow tract
Wnt3	WNT3	Mice	RNAi	Pericardial edema, ventricle enlargment, narrowing outflow tract
Dvl2	DVL2	Mice	Knockout	Lethal:Double outlet, transposition, peristent truncus. Neural tube defects
Dvl1	DVL1	Mice	Knockout	Social disturbances. Cardiac phenotype when disrupted with Dvl2

Table 5. Animal models of Wnt pathway disruption. (103, 111, 113, 114, 116-122).

1.8. The genes

The genes investigated in this thesis are Rock1, Rock2, Wnt5a, Wnt11, Dvl2, Dvl3, Dvl1l1 and ANKRD6.

1.8.1. ROCKs

The proteins of the Rho family of small GTPases are key cell behaviour regulators and have been shown to be involved in several processes as smooth muscle contraction, cell migration and division, apoptosis, cytoskeleton organization, focal adhesions and myocyte hypertrophy (121, 123). Rocks, also known as Rho kinases, are the main and first described effectors of the small Rho GTPases (123).

Two isoforms of Rocks have been characterised in the mammalian genome: Rock I also known as Rockβ, and Rock II or Rockα (123). Rock 1 is a gene located on chromosome 18 of the human genome, comprising 32 exons and translating into a protein of 1354 aminoacids. Rock2 is a gene located on chromosome 2 of the human genome, which has 32 exons and translates into a 1388 aminoacid long protein (39, 124) .

Rocks are a family of serine/threonine kinases of approximately 160 kDa similar to myotonic dystrophy kinase, Cdc42 binding kinases and citron kinases. Structurally, Rocks have an amino-terminal kinase domain, followed by a coiled-coiled region with a Rho binding motif and functional motif at C terminus. These motifs comprise a pleckstrin homology domain (PH) involved in subcellular location (123). See Figure 9.

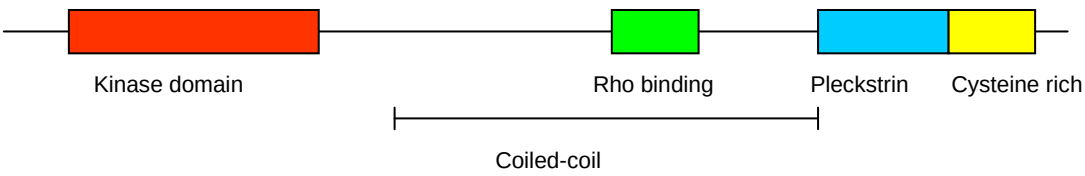

Figure 9. Scheme of Rock protein domains and regions.

The kinase activity of the proteins is triggered by the binding of Rho GTPases, and some lipids like arachidonic acid. When activated, the protein changes its configuration decreasing the autoinhibitory activity of the C-terminus leading to the phosphorylation of downstream proteins. Their physiological function is exerted via phosphorylation of proteins like myosin light chain kinase, myosin light chain phosphatase, LIM kinases and Adducin among others. The overall effect is an increase in contractility and organization of the cytoskeleton (123).

The mRNA of Rock1 and Rock2 have been ubiquitously identified in rat and mouse tissues (125). However, Rock2 is most strongly expressed in brain and muscle tissues (123). In the same way, Rock1 and 2 expression has been detected in early (121) and late cardiogenesis (126), in mice and chick embryos.

These proteins which are the primary effectors of the non canonical *Wnt* pathway (127) have been linked in several ways to development and especially to cardiac morphogenesis. They have been identified as key factors to organogenesis in vertebrates like mouse and chick (126), and to regulate movements of convergent extension and gastrulation in *Xenopus*, *Drosophila* and *C. Elegans* (128, 129). Interestingly knockdown of Rock 2 in mice causes lethality due to vascular malformations mainly in the placenta (120).

In early cardiogenesis, Zhao in 2003 showed that inhibition of Rocks in mouse embryos produces failure of heart looping and chamber formation, his group argued as the main explanation a decrease in cardiomyocyte proliferation (121). In the same year, the group of Kaarbo demonstrated that the inhibition with RNAi of RhoA in chick embryos causes cardia bifida (130).

Zhao also showed that Rocks are expressed in mouse endocardial cushions and that the inhibition of Rho kinases causes developmental arrest of these structures (131). In the same experiment, in vitro inhibition of ROCKs produced decrease in endocardial cell migration (131), finding which is similar to the one of Sakabe's group using chick embryos (132).

It is interesting that disruption of Rho kinases (ROCKs) seems to not produce cardiac malformations in mice. The ROCK1 KO mouse phenotype is characterized by eyelid

defects with 50% showing fully open eyes and the rest varying degrees of eyelid fusion and omphalocele. No obvious heart defects have been described in the ROCK1 -/- mouse (133). The ROCK2 KO mouse phenotype is characterized by intrauterine growth retardation and lethality associated with abnormalities and thrombus formation at the labyrinth layer of the placenta. As in the case of the ROCK1 -/- mouse, no heart abnormalities have been described (120)

It is also relevant to mention the numerous works relating the expression and activities of Rho kinases in cancer cells and their relationship with metastasis generation. Despite not being directly related with the cardiogenesis subject, research into the cytoskeleton changes due to ROCK effects is relevant to the physiology of cell migration and adhesion giving insights into some functional aspects of these proteins (134). ROCKs primary action due to the activation exerted by Rho GTPases is to induce specific types of cytoskeleton rearrangements in the cell and to modulate local dynamics of microtubules. The typical cytoskeleton rearrangement induced by these proteins are stress fibers and the contractile ring in the mitotic cell, both actomyosin bundles of anti-parallel actin fibers cross-linked by myosin fibers (134).

To migrate, cells polarize protrusions towards the stimuli and make adhesions to the extracellular matrix pulling the cellular body forward. At the rear cell tail should be detached and recoiled. Cytoskeleton and particularly actin, myosin and micro-tubules are key to this process. Two Rho GTPases, Cdc42 and Rac, seem to control cell polarity and membrane protrusions respectively (134). The first experiments like the ones by Worthylake et al. or Yoshinaga showed in cell migration models that ROCK was necessary to the retraction of the rear tail but that its inhibition did not decrease cell polarity and protrusions (134). Further research into the mechanism of tail detachment showed that it was dependent on cell adhesion attenuation. Despite the vision of that ROCK is not indispensable for forward protrusion formation in cell migration, some experiments like the one by Grosse et al. have found that Rho colocalizes with mDia (a coeffector) at the front of the migrating cell and that siRNA mDia inhibition interferes with cell migration and microtubules stabilization (134).

The previous experimental data in addition to evidence of Rho proteins expression in migrating neural crest cells (135), make Rho kinases ideal aetiological candidate genes for CCM.

1.8.2. ANKRD6

Ankirin repeat domain containing protein 6 (ANKRD6) is a human ortholog for the protein Diversin in vertebrates (136). Diversin is in turn a protein related to Diego in *Drosophila*, which has a recognized role in the non canonical *Wnt* signalling pathway (116, 137). Recent findings have also linked Diversin as a regulator of the canonical *wnt* pathway in mouse (138).

The ANKRD6 gene is located on chromosome 6, comprising 16 exons and coding for a protein of 727 amino acids (124). ANKRD6, like Diversin, is a modular protein, mainly encountered in the cytoplasm, has an amino terminal Ankirin repeat domain, a central casein kinase binding domain and a conductin binding domain at C terminus (116) . The casein kinase binding domain and the conductin binding domain seem to be involved in canonical Wnt signalling whereas the region involved in non canonical signalling has not yet been identified.

In 2002 the group of Scharwz-Romond described the effect of Diversin in cell culture and *Xenopus* embryos, showing that the protein increases the degradation of B-catenin and blocks the formation of the secondary body axis (138) . The same group also found that the expression inhibition of this gene produces convergent extension defects in Zebrafish embryos. In summary, the data points to an inhibitory function in the canonical pathway and a regulatory function in the non-canonical one. In terms of interaction, Diversin and Diego seem to cooperate with proteins like Dishevelled, Strabismus or Van Gogh in the localization of the complex responsible for the action of the non canonical *wnt* signalling pathway (116, 137).

Moeller in 2006 showed that the injection of Diversin mRNA lacking the ankirin repeat domain produces cardia bifida in a dominant negative fashion, demonstrating also that the protein directly interacts with Dishevelled in PCP movement control during development

(116). Due to its involvement in the non canonical *wnt* pathway, interaction with Dvl and phenotypes in animal models ANKRD6 is a strong aetiological candidate gene for CCM.

1.8.3. Dvls

Three Dvl genes have been described in mammals: Dvl1, Dvl2 and Dvl3 (139). Additionally, in 1996 the group of Pizzuti reported a homologue gene to Dvl, located within the DiGeorge syndrome region: DVL1L1(140). This is one of the candidate genes explored in this work.

The genes, which code for 500-600 amino acid long proteins, show a high degree of structural and functional conservation among vertebrates and invertebrates. Structurally, the proteins are characterised by three domains: The DIX, functionally related to the canonical *wnt* pathway, the PDZ involved in both canonical and non-canonical pathways, and the DEP which is mainly non-canonical related (141). See Figure 10.

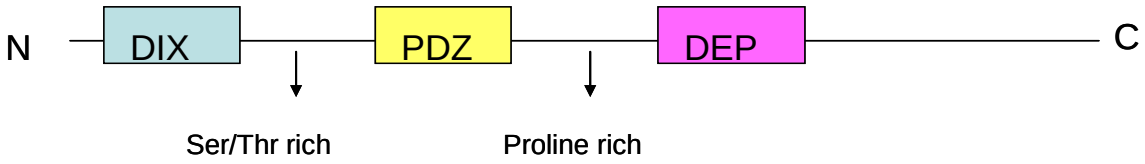

Figure 10. Dvl proteins structure scheme. Adapted from Wallingford et al. (141)

The DVL genes chosen as candidate genes in this research are: DVL2, DVL3 and DVL1L1. DVL2 is a gene located on chromosome 17, containing 15 exons and coding for a 737 aa long protein. DVL3 is a 15 exon long gene located in chromosome 3, coding for a 716 amino acid long protein. Finally, DVL1L1 is a single exon gene located on the short arm of chromosome 22, without characterised product in the proteome yet (39, 124).

Initially, it was thought that Dvl was mainly involved in the canonical pathway signal transduction, but now its involvement in both canonical and non-canonical pathways has been recognized (141). Despite that the nature of the changes leading to Dvl activation and the processes behind its selectivity for one of the pathways in a given context are under research, some of the molecular interactions have been described (141).

In terms of the canonical pathway, Dishevelled seem to bind and inhibit Axin in a *Wnt* dependent fashion blocking the phosphorylation of β catenin and thus activating the pathway (141). On the other hand, non canonical involvement seems to be related to two parallel pathways. As mentioned before, the signal in the PCP pathway is triggered by the accumulation of an incompletely understood group of proteins in an asymmetrical fashion within the cell(142). In this context Dvl activates DAAM1 (Dishevelled associated activator of the morphogenesis 1 and Rhac leading to ROCKs and JNK signalling respectively (143).

Experiments in animal models have shown that in *Drosophila* Dsh (Dvl) is involved in the PCP pathway controlling hair and bristle polarity (141). Park in 1996 showed that the gene is able to rescue cardiac development in *Drosophila* wingless mutants demonstrating that the protein is downstream in the pathway (144).

In terms of vertebrate development, Dvl involvement in convergent extension and the canonical *wnt* pathways has been largely recognized (141). The gene has been proven asymmetrically expressed in *xenopus* gastrulation, controlling anterior-posterior body axis patterning (145). In 2004, Ewald and coworkers demonstrated that Dvl inhibition in *Xenopus* embryos cause gastrulation defects like blastospore closure failure(146). Furthermore, cell studies trying to elucidate this fact have shown that disruption of dishevelled genes block the formation of lamellipodia in migrating cells (147).

In 2002, Hamblet and coworkers showed that knockout mice lacking Dvl2 had 50% perinatal lethality, Dvl2 -/- often failed to thrive and were cyanotic, showed reduced mobility and did not feed. Pups that survived beyond 24h developed into adulthood but 25% of them showed kinked tails and also exhibited scoliosis and rarely vestigial tail. Strikingly, most of the Dvl -/- pups that survived were females. Embryo analysis showed that also a minority of Dvl2 -/- embryos showed incomplete thoracic neural tube defects and exencephaly (113).

Lethality was due to cardiac defects, especially double outlet right ventricle and persistent truncus arteriosus. The same group also demonstrated that the cardiac defects were due to failures in migrating NCC. Dvl2 KO mice developed neural tube closure defects and

somite segmentation defects, reinforcing the theory of Dvl involvement in vertebrate convergent extension movements (113). Interestingly mice lacking only Dvl1 do not show cardiac abnormalities. The phenotype is characterised rather by poor social interaction and gating defects (148).

Due to the animal model findings, biology and relevance as a central branch point of *wnt* pathways, I believe that dishevelled genes are strong candidates as causal factors for CCM.

1.8.4. Wnts

The Wnts are a family of cysteine rich glycoproteins comprising 19 genes in the mammalian genome (71). Two of the non canonical pathway *wnt* activator proteins are part of this research: wnt5a and wnt11. Wnt 5a is a gene located on chromosome 3 of the human genome, comprising 4 exons and coding for a 380 aminoacids long protein (124). Wnt 11 is a gene located on chromosome 11 of the human genome, comprising 4 exons and translating into a 253 aminoacids long protein (124).

Expression of these two genes throughout cardiac development has been reported in the literature: Wnt 5a has been detected in the primitive streak, precardiac mesoderm, outflow tract and primitive ventricle of chick and mouse embryos (71). Expression of wnt11 has been shown in the primitive streak, precardiac mesoderm, primitive heart tube, outflow tract myocardium and conduction system of mice, chick and frog embryos (71).

In 2007 Schleiffarth et al (115) analysed the cardiac phenotype Wnt5a -/- mouse. Wnt5a -/- mice were grossly abnormal and died shortly after birth. OFT defects were present in all Wnt5a mice, including PTA (Majority of cases), TGA and double outlet right ventricle. Large ventricular septal defects were present in all specimens. These findings are similar to the previous results from the Oishi group who encountered VSDs at Wnt5a -/- mice (149). Angiographic analysis of Wnt5a embryos showed other aortic arch associated abnormalities like interrupted aortic arch, right sided aortic arch and coarctation of the aorta. The observed phenotype combining persistent truncus arteriosus and ventricular septal defects not only constitutes an OFT malformation but grossly resembles TOF.

Schleiffarth et al also showed that Wnt5a is expressed in the mouse secondary heart field mesoderm. Wnt5a was not found to be expressed at migrating neural crest cells. Wnt5a knockout mice did normally express primary and secondary heart field markers. In contrast, neural crest cell markers as *Sox10* and *Crabp1* expression were decreased in Wnt5a -/- mice. The hypothesis the authors advanced to explain the phenotype was that Wnt5a plays a regulatory role in neural crest cell migration and is expressed primarily in the secondary heart field.

In 2007 Zhou et al. published a work describing several experiments related to the regulation of morphogenesis exerted by the PCP pathway, including the description of the cardiac phenotype of WNT11 -/- mice (150). They found that Wnt11 KO mice, similar to Pitx -/- mice, present cardiac conotruncal defects. Pitx is a downstream effector of the canonical Wnt signaling pathway, and Wnt11 expression seems to be also regulated by Pitx. At 9.5 Wnt11 -/- mice OFT were shorter than in wild type controls and at 11.5 Wnt11 -/- OFT were still shorter and lacked the rightward curve present in wild type controls. Wnt11 -/- mice exhibited OFT defects with a penetrance of 100%. OFT defects were accompanied by VSDs.

1.9. Hypothesis and Aims

The hypothesis of this project is:

- Common and/or uncommon variants within the *Wnt* non-canonical pathway genes ROCK1, ROCK2, DVL2, DVL3, DVL1L1, WNT5a, WNT11 and ANKRD6 are aetiologically responsible for a fraction of non-syndromic TOF heritability in a human population of British origin.

With the hypothesis in mind, the aims of this project are:

- To identify and quantify the contribution of rare genetic variants within the *Wnt* non-canonical pathway genes ROCK1, ROCK2, DVL2, DVL3, DVL1L1, WNT5a, WNT11 and ANKRD6, to the risk of suffering TOF.
- To identify and quantify the contribution of common genetic variants within the *Wnt* non-canonical pathway genes ROCK1, ROCK2, DVL2, DVL3, DVL1L1, WNT5a, WNT11 and ANKRD6, to the risk of suffering TOF.
- To explore the contribution of gene-gene interaction within the *Wnt* signalling pathway to the risk of developing TOF.
- To perform functional studies based on the genetic findings in order to explain the mechanisms of associated candidate gene variants.

2. MATERIALS AND METHODS

2.1. POPULATIONS

2.1.1. CHANGE

The majority of the TOF probands included in this study were ascertained within the frame of the CHANGE study (Congenital Hearts: A National Gene/Environment study), a research effort led by my host laboratory with the aim of collecting a substantial number of trio families composed by a Tetralogy of Fallot proband and his/her parents (151). The sample was recruited during the experimental stage of this thesis and as a consequence the number of samples available was not completely uniform depending on the availability of samples.

The ascertainment strategy was as follows: Caucasian patients with Tetralogy of Fallot or pulmonary atresia with ventricular septal defect were recruited through the practice of Cardiology Consultants in 5 centres: Bristol Royal Hospital for Children, Royal Liverpool Children's Hospital NHS Trust, Alder Hey Hospital in Liverpool, Leeds General Infirmary and Newcastle Freeman Hospital.

The cardiologists identified the patients and informed them about the study. If the patients and the family were interested, a nurse met them to explain in detail the research protocol. Informed consent was obtained, a blood or saliva sample or mouth swab was taken from the proband and the parents when available. Finally, a form including key questions (See Appendix A) and a pedigree was filled for each family (151). Patients classified as TOF associated with known syndromes, 22q11 microdeletion, known maternal teratogen exposure, learning difficulties or with a high probability of being associated with a mendelian transmitted condition were excluded. Ethical approval was granted by the Northern and Yorkshire MREC.

2.1.2. CONCOR

The CONCOR project is a initiative in the Netherlands for the development of a National registry and a DNA bank of patients with congenital heart disease(152).

The Netherlands Heart foundation and the Interuniversity Cardiology Institute of the Netherlands (ICIN) with the collaboration of all 8 University Medical Centers in the Netherlands initiated the CONCOR project and now more than 20 non-university hospitals have also joined(152). Ethical approval was granted.

Patients with congenital structural heart defects or Marfan syndrome are eligible for the CONCOR cohort. Patients are asked to enter the study by their cardiologist or nurse during their outpatient visit. Probands are informed about the aims and characteristics of the study and informed consent is signed. Twenty cm^3 of peripheral blood is taken and DNA extracted and stored. Clinical data is entered by trained research nurses in a centralized database(152).

The DNA of 468 Tetralogy of Fallot patients were kindly facilitated by the CONCOR Steering committee to our laboratory. Tests in our laboratory identified probands carrying 22q11 deletion who were excluded from the analysis.

2.1.3. Oxford TOF Cohort

Samples from Caucasian patients with Tetralogy of Fallot ascertained as part of the GO-CHD (Genetic Origins of Congenital Heart Disease) study were used during this research, through collaboration with Professor Shoumo Bhattacharya, Oxford University. Ascertainment strategy consisted of recruitment of European Caucasian cases with non-Mendelian, non-chromosomal congenital heart disease sufficiently serious to require cardiac catheterisation and/or surgical correction. Probands were collected from five paediatric cardiology units when attending outpatient clinics, catheterisation or surgery. Informed consent was obtained and a blood sample taken. Family history of congenital

heart disease and consanguinity was obtained. Diagnostic data were mapped to European Paediatric Cardiac Codes (EPCC). Permission to collect parental blood samples and parental attendance for transthoracic echocardiography was sought where possible. Ethical approval was granted.

2.1.4. Nottingham TOF Cohort

Samples from Caucasian patients with Tetralogy of Fallot, for the Nottingham Family Study of Congenital Heart Disease were made available for this research through collaboration with Dr. Frances Bu'Lock (Leicester University) and Professor David Brook (Nottingham University). The ascertainment strategy consisted of recruitment of European Caucasian cases with non-mendelian, non-chromosomal congenital heart disease sufficiently serious to require cardiac catheterisation and/or surgical correction. Trio families consisting of TOF patients and their parents when available were recruited at Glenfield Hospital, Leicester. Informed consent was taken and a blood sample was obtained. Ethical approval was granted.

2.1.5. HTO

One of the control populations used in this research was a large cohort comprising 1428 patients from 248 nuclear families (that is, over 1000 control chromosomes). The recruitment process of the HTO population has been described in the literature(153). Briefly, families were recruited through a proband diagnosed with essential hypertension. Families were required to consist of at least three siblings clinically assessable for blood pressure if at least one parent of the sibship was available to give blood for DNA analysis, and to consist of at least four assessable siblings if no parent was available for DNA analysis. The majority of the individuals in the family collection therefore have blood pressures within the conventionally accepted "normal range". The families were subjected to a cardiological evaluation and echocardiography, and Congenital Cardiovascular Malformation was excluded. Ethical approval was granted.

2.1.6. ECACC

The commercially available Human Random controls population (HRC) from ECACC was kindly facilitated by Dr Tim Goodship. The HRC is a DNA collection from a control population of 480 randomly selected, non-related UK Caucasian blood donors (154).

DNA is obtained from lymphoblastoid cell lines derived by EBV transformation of peripheral blood lymphocytes from the blood samples and is organized in five panels. Each panel contains samples from 96 separate individuals (154).

Population	Subjects
HTO	1428
ECACC	480
CHANGE	465
CONCOR	468
Oxford+Nottingham	140

Table 6. Number of total subjects in each one of the populations used. Cases and controls were genotyped as specified in each one of the chapters.

2.2. MUTATIONAL SCREENING

2.2.1. Sequencing

2.2.1.1. PCR

The published reference genomic sequence of ROCK1, ROCK2, DVL3, DVL1L1, WNT5A and WNT11 was downloaded from the Ensembl webpage release number 38 (39). Primers aimed to amplify exonic sequences and 100 bp at each side were designed using the web based resource Primer 3 (155). The configuration was set up to obtain primers with an optimum Tm of 60, minimum length of 18bp and minimum complementarity. Each primer was reviewed manually and blasted to rule out unspecific annealing. Redesigning

of primers with poor PCR performance was done using the Primer Blast tool on the NCBI web page (155). PCR amplification of 2 Dvl2 exonic regions, completing a previous screening performed in our laboratory, in 93 TOF patients was performed during this research work. The method used was the same as for the other PCR reactions. See primers in Table 12.

Primers were purchased from Metabion TM; the dNTPs were ordered from Metabion TM, the Taq polymerase, the Buffer, Magnesium Chloride and Q solution were purchased with the Hotstar Taq polymerase kit from Qiagen TM. The machines used during the PCR protocols were DNA engine TETRAD TM.

Standardized PCR reactions were performed using the following concentrations: 1mM of buffer, 2.5 mM Magnesium chloride, 0.2 µm dNTPs, 0.2 µm of each primer and 0.2 units of Taq polymerase per well. Optimization was achieved performing a PCR reaction with a gradient of annealing temperatures for each exon and checking it in agarose gel electrophoresis.
Table 7 shows the generic temperature program used.

No of cycles	Temperature and time
1	95C for 15 min
40	95C for 20 sec T depending on exon for 30 s 72C for 1 min
1	72C for 3 min

Table 7. Generic PCR program used

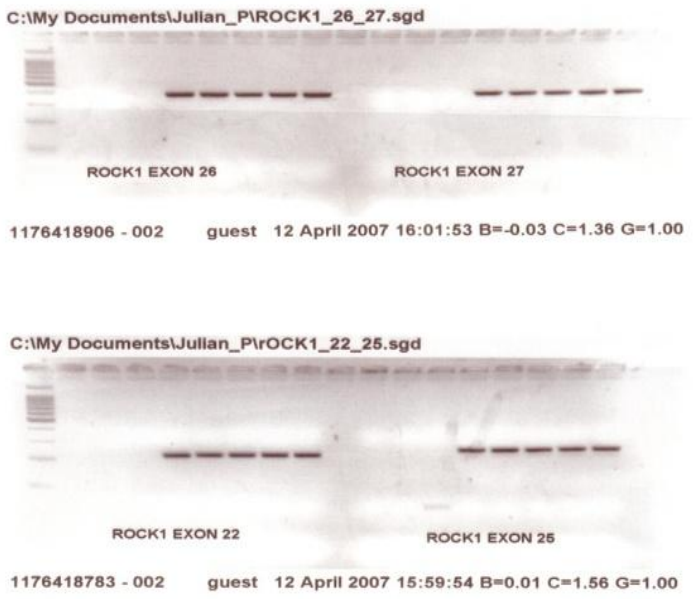

Figure 11. Agarose gel to check PCR reactions.

Region	Forward Primer	Reverse Primer	Opt annealing T	Product length
1	CGGAGGAAGTTGGTTGAAAT	TGCAGCGAACCAGACTAATG	64.5	243
2	CACAAGATGTGTTAGGGGAAA	AAAAATAAGGGCTTGAATGACA	61.5	275
3	CATGAAGTTTGTTGTTTTTCATTG	AAATAATATGTTAAACACACAAGTGGA	61.5	288
4	TGGAATAATTTGGCAGAAGAGAA	AAACACAAACTAAGCTTCTCAGACA	61.5	370
5	TTGCAACTTCTGTAATTTGATTATTT	ACGGAAGAAGAATGGGTGAA	61.5	399
6	CCTAGATTGAGGATGCTTGACC	TCCAGCTATGGTTAGTAAATTCTTACA	61.5	299
7	CAAAGATTTGGAGTCCCATAGGT	ACGAAGCAGGTTTGGATCAT	62.5	385
8	CGGGTAACTTTGACTCTTCCTT	TTACATCAGGTTTCACTGTCATTT	64.5	352
9	TGGGCTGATACTATGCTGACA	CCCATATAGTATTTGTGGCTTCTG	64.5	369
10	ACCATCAGTAGTTGGTATGGAA	ATGAAATACTGGAAATATCATGC	55.5	394
11	AGGCGTGTAGTATTTATTATTGGTT	CAGCTAACCACAGCTTCACA	61.5	257
12	TTGTATTTCTAATGGCCTAACG	TCAATTCTAGCCGAGACACA	58.5	445
13	GGCCATAGGCTTCACTTGAC	TCAGGAAACGAAACAATTTGAA	58.5	272
14	GCGTTTGAAACAGAACATTCA	CCTGAGACCTGGGTTATTTAGC	61.5	409
15	GGCTTTGGGAGTGGACCTA	GCTGTGTCCGATTCTGTCCT	64.5	367
16	AGTGTTTCAAATGAGGTTCCTT	TGATGATGGTTGCATACATTG	58.5	469
17	TCGGATCAGGAAACTTAATACA	CCACATACAACCAATCAAACA	54.5	370
18	AAGGCATCATATCTCTTCCAGA	GATGGGATTACAGGCATGAG	54.5	399
19	GAAACCAGTTTACAGCCCATA	GTGAACACAGACCCATGTCA	54.5	379
20	AAACACGTGGAACAATGAAGA	CCTGAAAACCTTTGCAAGTAA	58.5	499
21	TGTGGCCTGACCAAGTTAAT	ATGCCAAGAAACAAATGGAC	54.5	405
22	TTTGCATTTGTCATGGGTAA	CCATAAAGGCAAATTTCAGC	58.5	393
23	TTTTATAGAAAGTTTGCAATGGA	ACTTGCACACACACTTACCC	61.5	410
24	CTGGAAACCAGACATTTGCT	TCTTAGAAACCTAACACAACATAGG	61.5	309
25	TCATCATGTTAAATCAAATTCAGTG	GCCTGGCCTTGAAATTCTAT	58.5	496
26	GTCCTGCTTCCTGATTTTGA	ACAGCCATATCCCAATTTCA	58.5	431
27	AATCTGTGGTAGCAAACATTCA	GGAATCCATGGACAAAATGA	58.5	400
28	TTTGTGTGATTGAAGCATGG	CAAGAGCAACATTAACTTGGTT	58.5	402
29	AGTGGCAAATATTATGCTTATCA	GGTAACAGTGCTGAAGTAACACA	60.5	275
30	TCACCCATATGTCCTCTTTAGTG	CATTCAGGAATGGTGAAAGG	58.5	279
31	AGAATCTAAGTCCTAAGCGTTTT	CCCCACCAGTGTCAAAATA	61.5	491
32	TGCATTTGATAGGTTGGATTT	AAATGATGCCCTCTGTGGT	61.5	384

Table 8. Rock1 primers, optimal annealing temperature and PCR product length.

Region	Forward Primer	Reverse Primer	Opt annealing T	Product length
Exon 1	CCAGTCCAGTCGGGAGAGT	GTAGCGAGGCAAGTCGAAAC	62	469
Exon 2	tcccagtttaggatggagga	ccagctacaatctcagtgatgc	62	249
Exon 3, 4	atgccttgctggaagtgaac	ggtagctgggtggaaagtga	67	480
Exon 5	acccaccttcttgccctact	cccaacacaaagcccataag	62	369
Exon 6,7	gcttatgggctttgtgttgg	gggcagggaacatattctacg	61.5	571
Exon 8	tcagttcctgtccctttcca	gcctctgaaggagcagggta	62	310
Exon 9,10	ggaagactccctgacctacca	gggttctgggtgagagaaga	66.5	395
Exon 11	tggatgggagactcagtccta	gtctccacatggaccctcat	65	373
Exon 12	cccagcattgatcccataat	cagatcaaaggaagggtcca	61.5	338
Exon 13	tggacccttcctttgatctg	cagagtctgctgccacttcc	61.5	354
Exon 14	ctctcatccagagcccacct	gactgcctgtaggaggcaaa	64.5	400
Exon 15A	GAGGAAGGCTATAGGTGGGC	CCTGCTCACATCACATCCAC	62.5	520
Exon 15B	ATCGGAGGAAGGAGAAGGAC	CTGGTCCCAGAGACAAAAGG	61.5	675

Table 9. Dvl3 primers, optimal annealing temperature and PCR product length.

Region	Forward Primer	Reverse Primer	Opt annealing T	Product length
1A	GAAGGGAAAGAAAGGGAGGA	CCTCCAGCTTCCTCTGGC	53	736
1B	CGCGAATCTGACTGAGGG	ACAGAATCGTTTGGTCTCCG	53	565
2	ggaaaccagttaaagctgttgg	ttgaataaatcaaatggtaagcaaa	61.5	393
3	aaaccatagagagttttggttctt	catgggtgggcatagaaatta	61.5	388
4	aaatgttctcctttgtttctttca	tgaagcataagacttggcaca	61.5	400
5	tccattacatacgcctgagtct	tggttatatcttgtttgggtgct	62.5	567
6	tcagcagtgaagttctgttgg	ttccctacgacaaaggcaat	65	379
7	tttgttcttccctgcatgtg	gggtaatgagaggcagactga	58.5	355
8	tcacatactgtaacttgggattgc	ccacagaacaaatttgaaatgaa	58.5	411
9	gtgaatcgataaaatggcaaa	tgagtctcaaaataacacatcatct	61.5	374
11	tgcaaagctaaatcatacatgg	ttgggaacttctctcttggg	53	324
12	cattgttttctttaagcctctca	attacaggtgtgagccacca	61.5	249
13	ggcaacagagcaagacttcg	aaggtaggtttcaaggcacg	64	296
14_15	tttttatctgtttggaatagcaagt	gcctttgcaacactgtca	62	495
16	aattgacagtgttgcaaaagg	tccagcacactatccatcatct	66	500
17	gctcacaaacaacagctcca	gaatccatgagcatgtggaa	64.5	334
18	ggctgcaatgagccataatc	ttttccctccaattcccaat	61.5	433
19	acaaagagctgagattgcgt	agttaaaccaacaggttgcaga	53	518
20	catcaccatattttataccaacttca	tgctcaaaagaaaatccacaga	53	356
21	cccagaccatactttgagagc	tttgcccagctgctcttt	62	246
22	tcatttgctcggggacttag	tcattttaatggttcagttttcac	58.5	384
23	gctcctagaacaatgcctggta	caacattgagaaggtggcagt	68	461
24_25	aagactcacctgggtgacaga	cattcgatgactactgctaccaa	64	548
26	gccctgtatgtagggaagga	ttggatgcaacataactcaaca	58.5	452
27	tttcatgtctgtgttttactggtt	tccttcagagcagtcagtcg	62	376
28	tgatgtatagtgataaattctaaagga	tcctacaactacaaaagccaaa	58.5	282
29	aatttattgttattttaaggaagactg	tgaagcaaatcacaggaaaa	62	386
30	tctgagttcctaacatgggtca	cctgaagagggtcaaagaatg	58.5	315
31	gattaagtttgcgtaatcaagagc	CAGATTCTTTGCCGTTGAAA	64	545
32	GCACCTTGCAAAGgtaaatga	catgcacattaagacagttccaa	58.5	471

Table 10. Rock2 primers, optimal annealing temperature and PCR product length.

Region	Forward Primer	Reverse Primer	Opt anealling Temp	Product length
Exon 1	gcggtgccacctctctctc	cgtcctccccgacctgag	Pending	415
Exon 2A	caacatgcaagcagcagaat	GAGAAGGCGCTCTGAAAGTG	62	290
Exon 2B	CTTGTCTACCGCTTCCCTGA	gctgtgaaccaaacacagca	61,5	251
Exon 3	gagggagtgatcatccgaaa	cctcagagcaggtcttccag	66,5	386
Exon 4	gggtccttgtgtccaagtgt	tcgaggactcaaaagtttcca	62	285
Exon 5	caattgagaaaggccgtctg	ccactctgggctctcactct	65	320
Exon 6	agaatgtgagcgttgccttt	ctcacatgcctgaatgatcc	62	389
Exon 7	ttggggtggagcatctatgt	cccgtttgcactaacagacc	62	330
Exon 8	cacacctaggccaggtaatga	gctgtggttccagcagagat	62	313
Exon 9	tgtgactgggtgtggaactg	ccccatgcatgcctatacct	58,5	331
Exon 11	ctcgtgggtcccttaccata	cagccaagtgagtggctttt	55,5	348
Exon 12	tggtgtcagtcttgcagagg	ccaggaatgccgttaaacaa	55,5	318
Exon 13	ccatcatctctatcccctggt	ggaagggtgtgctaaagtcg	62	348
Exon 14	caggggctcctacttggttc	gggagtctcagcctctgagtt	62	318
Exon 15	tgaaaagtccatccttgatgc	ttcccacagccctagcatac	62	327
Exon 16A	aaatactgacccgcagccata	GGCACGATGCCTTGTTGG	62	360
Exon 16B	GCCCTCAACTCCACTGCTAC	GGCTTCGTGCTTCTTCAATC	62	385
Exon 16C	GGAAGAGTGGGCCAACAAG	GTTGACTCCTTGGGCATAGC	62	396

Table 11. Ankrd6 primers, optimal annealing temperature and PCR product length

Region	Forward Primer	Reverse Primer	Opt anealling Temp	Product lenght
Exon 11	gctgagtcagggagggac	ggtaaccagagtcagggatcg	61,5	303
Exon 15A	ggatctgagggtggatgaga	ccaccatcatggggttgta	61,5	362
Exon 15B	agcaagtgggactagcga	gagcccaggcactgtaagag	61,5	394

Table 12. Dvl2 primers, optimal annealing temperature and PCR product length

Region	Forward Primer	Reverse Primer	Opt anealling Temp	Product length
Exon 1	AAGGTCTTTTGCACAATCACG	GGGAAATGGAGGGATAGGAA	58.5	523
Exon 2	ttaagagaccccggatgctt	attgccgcatcatacaaaca	61,5	300
Exon 3	tgaggacaagcaggagagaaa	aaggcatcctccttcatgc	61,5	480
Exon 4	GCCAGTGATCCCTTGTCCT	TAGCAAAGGAGTGGCAGAGG	58.5	432
Exon 5A	catggagaaggtcgaggaga	GTCTGCACGGTCTTGAACTG	59	461
Exon 5B	GTGATGCCCTGAAGGAGAAG	gttgcaattcttggggaaaa	62	471

Table 13. Wnt5A primers, optimal annealing temperature and PCR product length

Region	Forward Primer	Reverse Primer	Opt anealling Temp	Product lenght
Exon 1	cgacccctcctttgtaatttg	gaaacagaggctgagggatg	58,5	410
Exon 2	gcaaagtccattgagggaaa	cctgtatccacccgtcattc	61,5	486
Exon 3	cgtcaccttggttacagcag	gaagggctgaggatgaggat	67	540
Exon 4	ACTCCAGCAGGCAGTGACTC	CTCAACTGAGCAGGGTCTCC	61.5	538
Exon 5	TCCCCAAAGATCAAGACAGC	GCTTCCAAGTGAAGGCAAAG	61.5	471

Table 14. Wnt 11 primers, optimal annealing temperature and PCR product length

Region	Forward Primer	Reverse Primer	Opt anealling Temp	Product length
Exon 1A	ccttcaggttactgatgttcattg	AAGCAGCATTTACACAGAAGCA	61,5	419
Exon 1B	CTAGACATCCTGCCCACTGC	tccagcttgggtgacagagt	61.5	473

Table 15. DVL1L1 primers, optimal annealing temperature and PCR product length.

After the optimization process, each exon was amplified in 93 TOF patients. The reaction was checked in 3% Agarose Gel Electrophoresis. See Figure 11. PCR products were cleaned using ExoSap ® purchased from USB Corporation. 5 ul of each PCR products were mixed with ExoSap ® and incubated at 37°C for 45 minutes.

2.2.1.2. Sequencing reaction

Sequencing was performed using MegaBACE DNA Analysis. The DYEnamic ET Dye Terminator Cycle Sequencing Kit for MegaBACE DNA Analysis Systems ® was purchased from Amersham Biosciences TM. The DYEnamic ET terminator reagent premix containing the Thermo Sequenase II DNA polymerase, dNTPs and labeled ddNTPs was used for the

amplification PCR reaction. A mixture containing, 2.56 ul of premix, 2ul of PCR product, 0,25pM of primer and water up to a volume of 10ul, was loaded for each sample and run on the thermocycler at the following conditions: 35 cycles of 95ºC for 20 seconds, 50ºC for 15 seconds and 60ºC for 1 minute. The products were precipitated using isopropanol.

The ABD files produced by the MegaBACE machine were analyzed using Staden Package v1.0 TM software(156). Databases were constructed with groups of 12 sequences. These files were assembled in the minimum number of contigs possible for analysis. Sequences not aligned were reviewed manually to verify the quality of trace and basecall. When the quality of the sequence was poor, defined as background or low confidence values, the files were discarded and the process repeated.

Contigs assembled with GAP4 software were edited manually. The files for the forward and reverse strand were compared with the consensus sequence from the Human Genome Project. Differences were reviewed, and deleted or confirmed when necessary.

2.2.1.3. Statistical analysis of variation found on Sequencing

Allele and genotype frequencies of the previously known SNPs and the newly discovered ones were manually calculated using Microsoft Office Excel 2003. The degree of Fitness with Hardy-Weinberg equilibrium was measured for each SNP using Chi Square tests performed in Minitab 14 ™ software.

The allele and genotype frequencies for the previously described genetic variants were downloaded from the NCBI SNP database builds 125-128 and Hapmap database phase 3 (39, 124, 157). Frequencies obtained from the sample were compared to the published ones using Chi square tests performed in Minitab 14 ™ software.

2.2.1.4. In silico analysis of impact of variants

Location of the variants within the protein was performed by manual analysis of the Ensembl database (124). Sequence of uncommon variants and previously undescribed common variants was analyzed using three web based resources: Polyphen (158) which evaluates the probability of structural damage based on known structural data about the

protein; ESEfinder (159) which determine the presence of known motifs for Intronic and Exonic splicing enhancers; and finally Spliceview (160) which checks the relative strength of splicing donor and acceptor sites within a given sequence.

2.2.2. MALDI TOF mutational screening

In an attempt to speed up the analysis and amplification steps of the mutational screening, Rock2 exons 10,11,13 and 28 were screened using a combination of Meta-PCR and MALDI TOF strategy. 48 TOF samples were screened using this method.

Meta PCR is a technique of constructing a DNA artificial molecule with a combination of PCR fragments in any order. There are two main steps to this: the first one is a multiplex PCR using primers with 5' complementary linkers at limited concentrations. At the beginning of the reaction the linkers are incorporated to the fragments, and at the end primers exhaustion produces overlapping amplification and full length product. The second step is a round of PCR using internally nested primers to amplify the full region (161). The nested primers were designed to contain 5'-3' oriented T7 RNA polymerase promoters for the forward and reverse strand, splitting the second step into two reactions: one for the forward and one for the reverse strand.

The products were amplified using T7 RNA polymerase and a mixture of nucleotides containing either CTP or TTP. Subsequently the products were digested with RNAase A producing a group of distinctive fragments with different masses (162). The reactions were fixed in a special resin, transferred to a SpectroChip™, and run in a Bruker Autoflex MALDI TOF spectrometer. The mass spectra patterns were analysed with the Sequenom™ SNP detection software. Manual review of the spectra was performed. See Figure 12 for spectra examples.

The sequences were downloaded from the Ensembl web page (124).Primers aimed to amplify coding sequences and 100 bp at each side were designed using the web based resource Primer 3 (155). The configuration was set up to obtain primers with an optimum Tm of 57, minimum length of 18bp and minimum complementarity. Linkers previously

tested by the National Genetics Reference Laboratory in Manchester were added to the designed primers (161).

Type	Exon	Primer1	Primer2
Meta-PCR 1	10	cagcacatagattaaaatgaaatg	gacgcggcggcaattaaaaacatggggattactaaaa
Meta-PCR 2	11	tgccgccgcgtcaaagtcacattaaggaggtacaaat	cgggcgcgcagaatgagaggcttaaagaaaacaatg
Meta-PCR 3	13	tctgcgcgcccgaaaggagagctaccaagtaattatg	ggccgggcgcttatccttaggtaagtcagacatcaaa
Meta-PCR 4	28	aagcgcccggccaagcctcacatgttgtgctta	tcctacaactacaaaagccaaa
Nested 1	ALL	cagtaatacgactcactatagggagaaggctcagcacatagattaaaatgaaatg	aggaagagagtcctacaactacaaaagccaaa
Nested 2	ALL	aggaagagagcagcacatagattaaaatgaaatg	cagtaatacgactcactatagggagaaggcttcctacaactacaaaagccaaa

Table 16. Primers for Meta PCRs and Nested PCRs. Linkers in blue, spacers in red and T7 promoter in green.

Primers were purchased from Metabion®. Taq polymerase, PCR buffer and magnesium were obtained with the Hotstar Taq kit purchased from Qiagen.

PCR reactions were performed using the following concentrations: 1mM of buffer, 2.5 mM Magnesium chloride, 0.2 µm dNTPs, 0.4 µm of each primer (2mM) and 0.2 units of Taq polymerase per well. Optimization was achieved performing a PCR reaction with a gradient of annealing temperatures for each exon and checking it in agarose gel electrophoresis. Optimal temperature for multiplexing was 54 C°. The PCR program used in both PCR rounds was as follows: 95 C° for 10 min.; 40 cycles of 95 C° for 20 sec., 54 C° for 30 sec and 72 C° for 1 min; finally 72 C° for 3 min.

RNA amplification and MALDI TOF reactions were performed by specialized technicians.

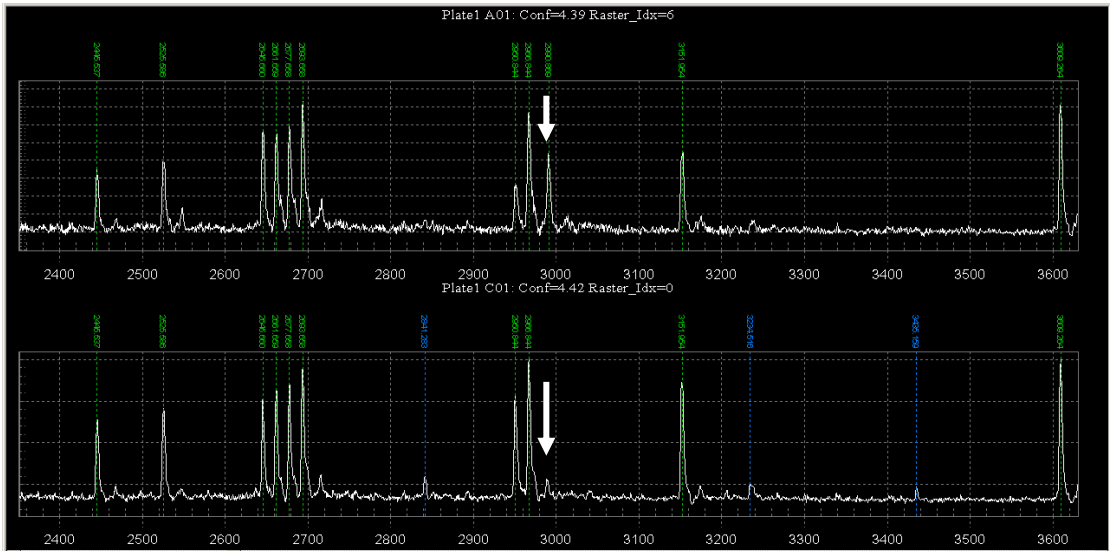

Figure 12.Example of Mass spectra differences between samples. White arrows denote the key difference.

2.3. CONTROL GENOTYPING

All previously undescribed variants and the previously described SNPs that showed significant allele frequency differences when compared to previously published data were genotyped in at least 1 control population.

2.3.1. Sample calculation

The number of controls needed to be genotyped for any variant to find a significant difference if the variant is common, or to probe that is absent in controls if it is uncommon, was calculated *a priori*. Traditional probability algorithms were used as an initial approach, see Equation 1. Power calculation was also performed using Shaun Purcell's online Genetic Power Calculator and the software Quanto (163)

$$n = \frac{\log(\beta)}{\log(1 - mAF)}$$

Equation 1. N is the number of samples needed to obtain a β error to detect a variant with a given minor Allele frequency mAF for any variant.

The number of controls was calculated using the variant's minor alleles frequencies, an hypothetical OR of 2 and the published incidence in Caucasian British population of TOF (0.31/1000 newborns)(56) . Power calculated for each of the variants is specified in the results section

2.3.2 Sequenom

A total of 16 variants consisting of: 10 non previously described uncommon variants, 3 previously described SNPs and 2 non previously described common variants were genotyped using the iPlex™ method on the HTO cohort which comprises 1428 subjects distributed in 248 families and the complete set of CHA probands. 13 variants (W1

experiment) were also genotyped on 279 CHA parents, the ECACC cohort (n=465), the Oxford and Nottingham Fallot subjects (n= 209) and the CONCOR Fallot subjects (n=465).

The iPlex™ is a MassExtend based genotyping method which uses the Sequenom® platform. Summarizing, it consists of three steps: a multiplex PCR amplification of fragments containing the variant; an extension reaction which specifically extends a predesigned extension primer for each variant; and a mass spectrometry to analyse the mass spectra of the assay to determine the basecall for each one of the variants depending on their mass (164).

The sequence for each one of the variants and 200bp upstream and downstream was obtained from either the Ensembl web page (124) or our previous sequencing work. The sequences were uploaded to the Real SNP web based resource in order to design the assays (165). This software groups the variants depending on how suitable are they to be amplified as a multiplex and how their masses relate to each other for the mass spectrometry. The result was an assay with 13 variants (13 plex) and other one containing 3 variants (3 plex), the assays were called W1 and W2 respectively. PCR primer sequence, extension primer sequence and mass for each variant are listed on Table 17 and 18.

Multiplex PCRs were performed following the standardized process for iPLEX® experiments. The taq polymerase, buffer, and magnesium chloride were purchased as the Hotstar Taq polymerase kit from Qiagen. Desalted primers were purchased from Metabion. A total volume of 10ul was mixed for each reaction containing 1ul of 10X buffer, 1.63mM of Magnesium chloride, 0.5mM of dNTPs, 0.1uM of each primer and 1 unit of Taq polymerase. Cycling conditions are specified in Table 19.

Assay	Gene	SNP_ID	Forward Primer	Reverse Primer	UEP_MASS	Extension Primer	Call 1	EXT1_MASS	Call 2	EXT2_MASS
W1	ANKRD6	Pro636Leu	ACGTTGGATGCTCAAACTAAGAAGTCTGGG	ACGTTGGATGTGCCCACAGGTGCTGCTG	5140.4	GCATCGTGCCCAGCAAC	C	5387.5	T	5467.4
W1	ANKRD6	rs2273238	ACGTTGGATGAGGTGGCCAAAATCTTACTG	ACGTTGGATGAGAAAGCAATGTTGCAACTC	5293.5	GGAAGCCGGAGCAGATA	C	5540.6	T	5620.6
W1	WNT11	720 +6 C>T	ACGTTGGATGAACTGATGCGTCTACACAAC	ACGTTGGATGCACAAGCACAACTATCTGGG	5324.5	GTGGGGAGACAGGTAAC	C	5571.7	T	5651.6
W1	WNT5A	Asp119Ser	ACGTTGGATGCAATTCCGACATCGAAGGTG	ACGTTGGATGTCCTACCTATCTGCATCACC	5868.8	ggCTGCAGCACTGTGGATA	A	6140.0	G	6156.0
W1	WNT11	207 +47 G>T	ACGTTGGATGATCCAGGGGACCCCAAAAC	ACGTTGGATGTGTGAGTAGGGACCCCGCGT	5951.9	tACCCCAAAACGCCCTCCGC	G	6199.1	T	6223.1
W1	ANKRD6	Leu192Phe	ACGTTGGATGGCACTGACCTGGTTCTTTTC	ACGTTGGATGGCGCGCTATAATCACTTGTC	5983.9	aAGAAAGCAGTGAGGAGGA	T	6255.1	C	6271.1
W1	DVL3	1339 +17 A>G	ACGTTGGATGACACAGAGCGTGAGTGTCC	ACGTTGGATGATGGACCCTCATCAGTCCC	6053.9	cAGCGTGAGTGTCCCACCCT	A	6325.1	G	6341.1
W1	ROCK1	807 C>T	ACGTTGGATGCTTCCTTACTCACCTACAAG	ACGTTGGATGGGAAGAGAATGTGACTGGTG	6590.3	ctgcCTCACCTACAAGCATTTC	T	6861.5	C	6877.5
W1	ROCK1	1785 T>G	ACGTTGGATGGCTTGCAGCTGGTAATAATC	ACGTTGGATGAACAGAGAGTTGCAAGAGAG	6688.3	acTCTTTGTCTGTTTGTGACTT	G	6935.5	T	6959.6
W1	ROCK1	Thr773Ser	ACGTTGGATGGATCCCAATTATTTTTCTC	ACGTTGGATGACAGCCGCTTATTTGATTCC	7045.6	TCATTCATAGGTTAAGAATCTAA	C	7292.8	G	7332.8
W1	ANKRD6	rs9344950	ACGTTGGATGGGGTGCAGAAGGAAAATTAG	ACGTTGGATGTTATGCAGTTGCAAAACTGG	7122.7	ggAGGAAAATTAGCACCAATAAA	A	7393.9	G	7409.9
W1	ANKRD6	rs9353687	ACGTTGGATGTCTCCAGAAAGCTTCTGCTC	ACGTTGGATGTAAGATATCCTTGGTGGATG	7261.7	CTCCAGATTCTGAATCTTTAACAT	T	7533.0	C	7549.0
W1	ROCK2	672 -25 A>T	ACGTTGGATGAGGACCAAGGAATTTAAGCC	ACGTTGGATGAATGAATGCTTGGAAATTG	7474.9	CTGTTAAGAAATAAAAAGAGGAAA	T	7746.1	A	7802.0

Table 17. PCR primers, extension primers, masses and base call for the W1 sequenom experiment.

Assay	Gene	SNP_ID	Forward Primer	Reverse Primer	UEP_MASS	Extension Primer	Call 1	EXT1_MASS	Call 2	EXT2_MASS
W1	DVL3	446 T>A	ACGTTGGATGATAACCCATCGGAGCTGCC	ACGTTGGATGACTCACTGGAAGGATGGGG	5429.5	ATCGGAGCTGCCACCACC	A	5700.7	T	5756.6
W2	ANKRD6	ANK_7_T176N	ACGTTGGATGGGAGCCTAATGATGGACAAG	ACGTTGGATGTGATTTTTGTCCGCAGGCAG	5197.4	CAGCAACGTGCAAACAG	C	5484.6	A	5524.5
W2	ROCK2	rs2271621	ACGTTGGATGTCTGTCTGTGTAACTGGTCG	ACGTTGGATGCCTTGGGTGATATTTTATTCC	7674.1	ACATGAAATAACTTGCTATAAAAAA	G	7961.3	T	8001.2

Table 18. PCR primers, extension primers, masses and base call for the W2 sequenom experiment.

N of cycles	Temperature and time
1	95°C for 15 minute
34	97°C for 20 seconds 59°C for 30 seconds 72°C for 1 minute
1	72°C for 3 minutes

Table 19. PCR program used for the iPLEX multiplex amplification PCR

PCR products were cleaned with shrimp alkaline phosphatase (SAP) from Sequenom Inc. This step cleaves one phosphate on the unincorporated dNTPs converting them into dNDPs. Manufacturer instructions were followed: 5ul of the PCR product were transferred to a 384 well format plate and mixed with 0.17 ul of SAP buffer and 0.3 ul of SAP enzyme. The reaction was incubated for 45 min at 37°C. Due to the volume format used for the reactions a Multimek liquid handler was programmed and used to mix the components.

Subsequently, a volume of 2ul containing 0.22X of iPLEX buffer, 1X iPLEX termination mix, a mixture of the extension primers depending on their mass (0,625 uM- 1,25uM), and 0.041 ul of iPLEX enzyme was added to each one of the reactions. Extension cycling conditions were performed following manufacturer conditions. See Figure 13.

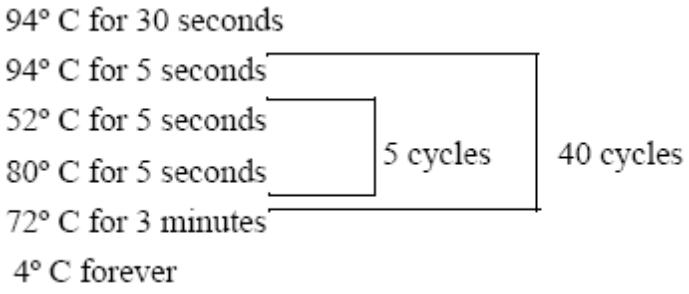

Figure 13. Cycling conditions for iPLEX extension reactions (164).

A spectroCLEAN enzyme was used to remove salt particles from the samples. The resin and 10 ul of water were added to each sample and the plates were agitated for 30 min. A Samsung TM nanodispenser was used to transfer 15 nl of sample to an SpectroChip

(166). The SpectroChip was introduced and analysed into a Bruker Autoflex MALDI TOF spectrometer, producing the spectra and reporting the alleles for each sample and SNP.

2.3.3. RFLPs

A total of 175 TOF patients and 1304 probands were screened for the mutation described by our group in 2006 in exon 14 of Dvl2 gene, 536(T/I), mRNA 1886 (C/T).

The sequence of Dvl2 exon 14 was uploaded to the web based resource Web Cutter (167). Enzymes able to distinguish between the possible alleles were analyzed manually. The ECO RV enzyme was chosen. The enzyme cuts in the consensus sequence GAT/ATC which corresponds to the non-wild type allele. Three bands are produced, one of 386 bp for the wild type allele and two of 148 and 248 for the non wild type allele. See Figure 14. Exon 14 was amplified using previously optimized primers. Table 20 shows the oligos.

Region	Forward Primer	Reverse Primer	Product lenght	Opt anealling Temp
Exon 14	AACTGTTCCCAGTTCCTCTCC	GGGCAACTGAGTCCTCACC	396	64.5

Table 20. Dvl2 exon 14 primers

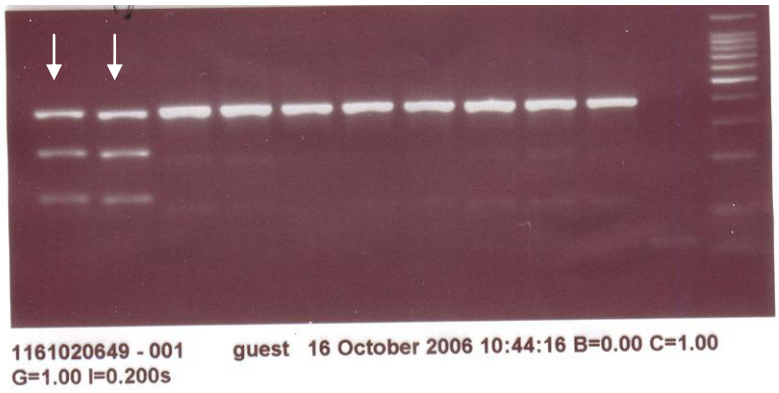

Figure 14. Agarose gel showing RFLPs designed for the mutation at position 1886 in the Dvl2 gene. The arrows denote heterozygous subjects.

The primers were purchased from Invitrogen TM. ECO RV enzyme was ordered from Fermentas TM. Taq polymerase and PCR buffer were obtained with the GoTaq polymerase kit purchased from Promega.

PCRs were run at optimal temperature in Peltier Thermal Cyclers, DNA engine DYAD TM. Products were checked through agarose gel electrophoresis. Subsequently, enzyme digestion assays were performed following manufacturers recommendations. The reaction contained 1.5 ul of Buffer 10X, 0.15 ul of BSA 100X, 1 enzyme unit, 3 ul of DNA and water up to a volume of 15 ul. Assays were incubated for 12 hours at 37 C°, and heated at 80 C° for 20 minutes. Digestion products were run in 3% agarose gels, and genotypes obtained manually and stored as Microsoft Excel TM files.

2.3.4. Taqman assays

The rs2074216 marker and the variant Pro136Arg, previously reported in DVl2, (168) were genotyped in the HTO population using the Taqman platform. The technique used was the fast real time 7900HT, Taqman, by Applied Biosystems. Briefly, the allelic discrimination assays using this technique consists in two steps, the first one of amplification or preread-run and the second one of analysis or post-read run (169).

The polymorphic region is amplified using PCR primers previously designed and optimized by the company and a polymerase enzyme with exonuclease activity. Each reaction has, in addition, two probes which specifically hybridises with one of the alleles. At each end of the probe there is a reporting fluorescent dye molecule, either VIC® or FAM®, and a quencher molecule respectively. When the probe is attached to the DNA template, the polymerase excises the quencher and the dye from the probe in the amplification step, making possible the fluorescence emission when faced to light stimulation. In the analysis step, the samples are grouped in four clusters depending on the fluorescence spectra: failing, homozygous wild type, heterozygous and homozygous non-wild type (169). See Figure 15.

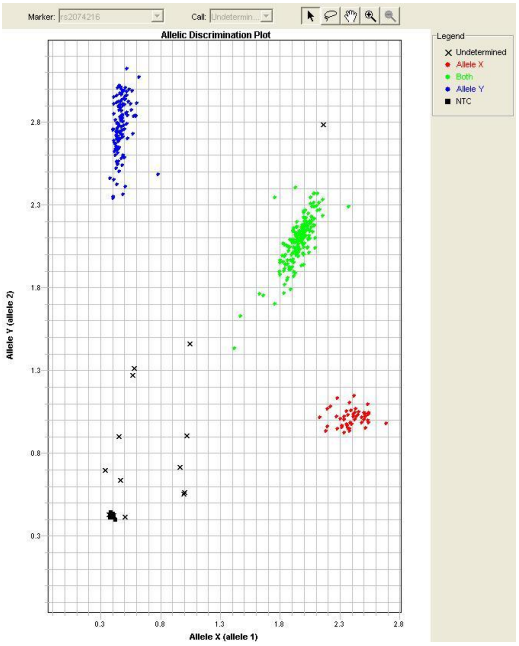

Figure 15. Taqman graphic result. Homozygous wild type sample are shown as blue dots, heterozygous as green dots, and homozygous non wild type as red dots.

Universal master mix DNA taq polymerase UNG® no Amperase was purchased from Applied biosystems®. Probes and primers for rs2074216 were ordered from Applied Biosystems®, as pre-designed genotyping assays. Sequences for the primers and probes for this assay are not known. The variant in DVL2 residue 136 was genotyped using a custom made assay ordered from Applied Biosystems ®. See Table 21

Assay	Forward Primer	Reverse Primer	Reporter 1	Reporter 2
M136	AGAAACCGAGTCAGTAGTGTCACT	CCATGCTCACTGCTGTCTCT	CTGCGAGGCCGCT	CTGCGAGGCCACT

Table 21. Taqman probes and primers for Pro136Arg in Dvl2

Reactions were prepared following instructions from the manufacturer. Each well contained: 2.5ul of Universal master mix, 0.061 of 40X primers and probes mixture, 1.37 ul of water, and 20 ng of DNA. The amplification run was performed in a 384 well plate using thermal cyclers DNA engine Tetrad TM. The standard program used consists of an activation step of 10 minutes at 95 C°, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C.

Post-read run analysis was performed using the 7900HT Applied Biosystems® machine. Frequencies obtained were stored as Microsoft Excel™ files.

2.3.5. Kaspar assays

Due to optimization difficulties 2 variants: ANKRD6 1 -304 A>G and WNT5A 1461 +38 C>T, were genotyped out of our laboratory by Kbiosciences laboratory services (170). The sequences were obtained from our sequencing work and sent to the company to design the probes and genotype 465 CHA TOF proband and 1248 HTO controls.

The technique used by Kbiosciences is similar to the Taqman®: the polymorphic region is amplified using PCR primers specifically designed and optimized by the company, and a polymerase enzyme with exonuclease activity. Two probes which specifically hybridises with one of the alleles are added to the reaction. At each end of the probe there is a reporting fluorescent dye molecule and a quencher molecule respectively. When the probe is amplified, the polymerase excises the quencher and the dye from the probe in the amplification and the fluorescence emission is detected when stimulated with a specific wavelength light. The samples are grouped in four clusters depending on the fluorescence spectra: failing, homozygous wild type, heterozygous and homozygous non-wild type (169). Results were electronically sent from K biosciences as Excel files.

2.3.6. Statistical analysis of case and control genotypes

The data with the genotypes and the pedigree were stored as Microsoft Office Excel 2003 files. The files containing family-proband data and genotypes were merged together using Microsoft Office Access 2003. The genotypes were checked for Mendelian errors using Pedstats ® (171), the resulting errors were reviewed and deleted when necessary. Additional checks based on unlikely recombination patterns within families were carried out using the error-checking option in MERLIN version 1.1.1 (172). Errors were corrected when possible by reference to the raw genotyping data, and when this was not possible, genotypes were excluded from analysis. The degree of fitness with Hardy Weinberg equilibrium was measured for each marker using the Pedstats ® software.

Population variable frequencies were calculated with SPSS v15.0 software. Allele frequencies were calculated using Pedstats software. Haplotypes were obtained when appropriate using Haploview 4.0 software (173). Extremely rare variants (N=1 or 2) were

compared manually. Initial ANOVAs comparing allele frequencies were using SPSS v15.0. Discrete association analysis was calculated with the algorithm present at Haploview software (174) and LAMP v0.0.9 software (175). Graphics were performed using SPSS v15.0 software.

To take some account of the risk of false-positive findings because of multiple comparisons, the results were interpreted based on the false discovery rate (176). The program QVALUE was used running on top of the statistical package R to determine q values (177). An arbitrary false discovery rate threshold of 0.05 was adopted.

2.4. ROCK1 duplication analysis

2.4.1. Bioinformatics analysis

Due to higher than expected heterozygosity rates in the SNPs at Exon 31 of ROCK1, a bioinformatics analysis was undertaken to explore the reasons for this.

The SNP rs35881519 mapped to two regions when Ensembl web page was consulted at the time. A search of duplicated regions was performed for ROCK1 at the Human Genome Segmental Duplication database (178). The original overlapping clones for ROCK1: AC021727, AC036178, AC131287, AC135340, AC139704 and the clone for the potential duplication: AP001005 were downloaded from the Entrez Nucleotide database at the NCBI web page (179). Sequences were stored as Microsoft Office Word 2003 files.

Dot plots (180) were constructed using the web based resource DNA dot (181) to compare each one of the sequence to the duplicated one. The dot plot is a graphical representation of a matrix comparison of two sequences. A user defined “window” spanning a number of bases is compared with the entire length of the subject sequence. If identity is found a dot is placed in the corresponding X-Y coordinate.

Using dot plots the sequence spanning the duplication was identified. The sequences were copied into Microsoft Wordpad software files. The files were aligned using the Clustal function present at the MEGA 4 software (182, 183). The graphic interface was used to

identify regions with differences within the duplication. A search using sequences characteristic to each one of the clones was performed at nucleotide BLAST against the expressed sequence tags EST database (184). Results of the BLAST search were manually analysed to determine if the duplication was previously identified as an EST.

2.5. MINIGENES

2.5.1. PCR and E. coli transformation

The method used for the minigene experiments has been described before in the literature (185). Exon 8 of ROCK1 and exon 1 of WNT11 including 300 nucleotides of intronic flanking region were amplified using the phusion PCR kit (NewEngland Biolabs ®) from DNA samples of heterozygote subjects for the ROCK1 uncommon variant 807 C>T and the WNT11 207 +47 G>T uncommon variant. Primers designed using the web based software Primer 3 were modified to include an ECORI site and a Poli-A tail. HPLC purified primers were purchased from Metabion®. See Table 22.

PCR reactions were performed following manufacturer guidelines. A mixture of 100uM of dNTPs, 0.5uM of each primer, 1,5 ul of DNA template (20mg/ul) and 0.5ul of polymerase were cycled 35 times under the conditions specified in Table 23. PCR reactions were performed on a DNA engine Tetrad2 thermal cycler. The concentration of PCR product was measured using a Nanodrop® spectrophotometer.

Variant	Forward Primer	Reverse Primer	Product lenght
ROCK1 807 C>T	AAAAAAAAAGAATTCTTGCACTGGCCTGTGTACTT	AAAAAAAAAGAATTCTCAGTTTCAAAGGAAGAGTCAAA	823
WNT11 207 +47 G>T	AAAAAAAAAGAATTCCCTGGACTCCTGATTCCTCA	AAAAAAAAAGAATTCAGGTCCTTTCAGGAGCCTCT	1126

Table 22. Minigene primers

N of cycles	Temperature and time
1	98°C for 1 minute
34	98°C for 20 seconds 62°C for 10 seconds 72°C for 30 seconds
1	72°C for 10 seconds

Table 23. Phusion PCR cycling conditions

The PCR products were cleaved with ECORI (New England Biolabs ®), and purified using the Qiaquick PCR purification kit from Qiagen®. 1ug of PCR was mixed with 5ul of 10X buffer, 1 U of ECORI and 44 ul of H2O and incubated at 37°C for 4 hours. Qiagen® instructions for PCR purifications were followed (186): 250ul of buffer PB were added to 50ul of reaction; the mixture was transferred to a Qiaquick spin column and centrifuged at 13000 rpm for 30 seconds in a tabletop microcentrifuge; flow through was discarded and the column washed with 0.75 ml of buffer PE; the column was centrifuged twice at 13000 rpm for 30 seconds in a tabletop microcentrifuge, flow through was discarded; DNA was eluted applying 30 ul of buffer EB to the column and centrifuging at the same conditions for 1 minute. The concentration of product was measured using a Nanodrop® spectrophotometer.

The purified fragments were inserted into the MfeI cloning site of the non commercial vector Pxj41 (187) kindly gifted by Prof David Elliot. Briefly, digestion of the circular vector was performed adding 1ug of it to 5ul of 10X buffer, 1 U of MfeI enzyme, and 44 ul of H2O and incubated at 37°C for 4 hours. MfeI enzyme was purchased from New England Biolabs.

After digestion the vector was purified using the Qiagen PCR purification kit following the procedure explained before. Both digestion fragments, the PCR products and the digested vector, were ligated using T4 DNA ligase from New England Biolabs. 1ul of vector, 8ul of PCR product, 2 ul of DNA ligase, 2 ul of T4 DNA ligase buffer 10X and 7 ul of water were incubated at room temperature for 6 h. The cloning site in the Pxj41 vector is located midway in the Intron between the constitutively expressed mouse B-globin exons 2 and 3 (185). See Figure 16.

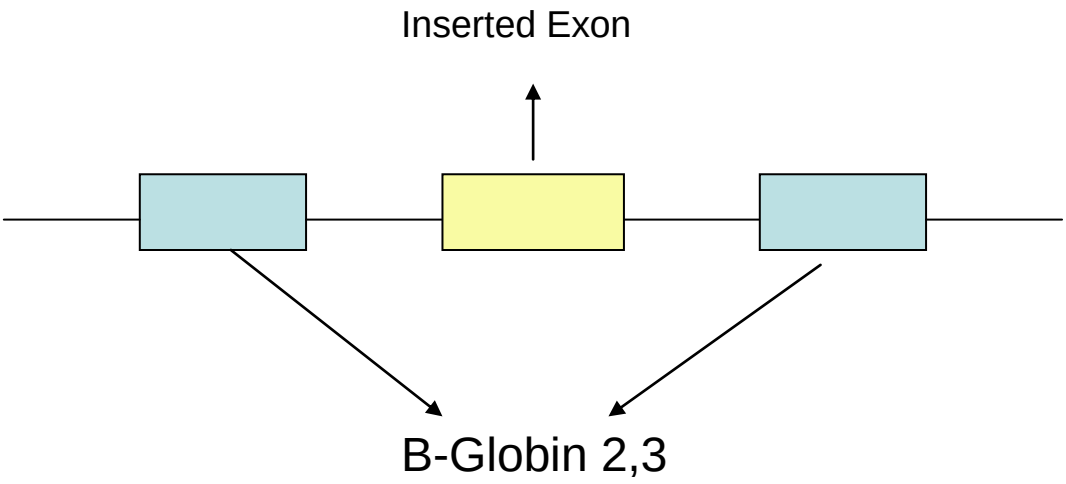

Figure 16. Pxj41 globin exons and their relation with the inserted exon.

The ligated fragments were cloned transforming competent E Coli cells JM109 from Promega ®. Briefly, 5ul of ligation were mixed with 50ul of competent cells and incubated on ice for 30 minutes. Subsequently cells were heat shocked at 42°C for 90 seconds and incubated again on ice for 5 minutes. After a recovery incubation time at 37°C cells were spreaded on agar-LB media plates (See Table 24 for preparation instructions) with Ampicilin at a concentration of 100 mg/ml to select transformed cells. Plates were incubated at 37°C overnight.

Ingredient	Quantity
Tryptone	10 g
Yeast	5 g
NaCl	5 g
1M NaOH	1 ml
Agar	15 g
Water	1 L

Table 24. Preparation of LB media agar plates

Colonies were picked after 12 h of incubation and DNA was extracted using a simple heating procedure: each colony was manually picked with a pipette tip and loaded into 50 ul of water. The mixture was incubated at 100°C for 5 min.

PCR amplification of the B-globin and inserted exon was performed on each colony using the vector specific flanking primers PXJRTF and PXJB1. The PCR was performed using the Hotstar Taq polymerase kit from Qiagen. Reaction mixtures had the same

concentrations as described in the section 2.2.1.1 and were cycled at the conditions described in Table 19. PCR product length was checked on a 2% agarose gel electrophoresis. See Figure 17

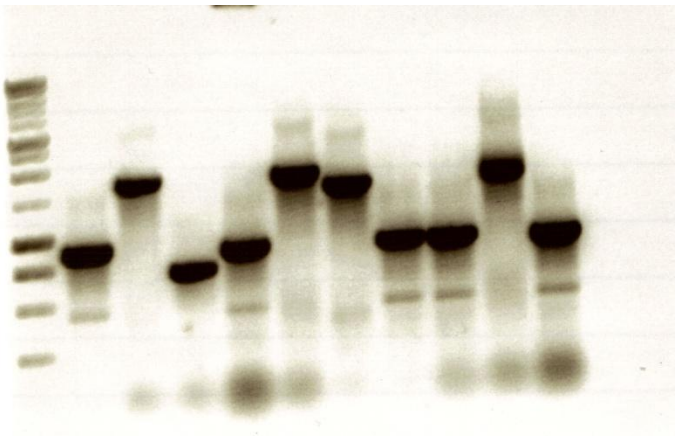

Figure 17. Agarose gel electrophoresis of Pxj41 transformed colonies

Colonies with the expected length were subsequently tested for insert orientation, using a PCR with a combination of primers from the insert and vector. For example, forward PXJRTF and reverse ROCK1 exon 8. PCR was performed using the Hotstar Taq polymerase kit from Qiagen. Reaction mixtures had the same concentrations as described in the section 2.2.1.1 and were cycled at the conditions described in Table 19.

Colonies with the insert in the correct orientation were seeded and cultured on 50 ml of LB-ampicilin media for 12 hours in a shaker at 37°C. DNA from the culture was extracted using the Miniprep kit from Qiagen® : cultures were spinned at 3000 rpm for 5 minutes; pelleted bacterial cells were resuspended in 250ul of P1 buffer and transferred into an 1,5 ml microcentrifuge tube; 250 ul of buffer P2 were added; 350 ul of Buffer N3 were added; samples were spinned at 13000 rpm on a tabletop microcentrifuge for 10 min.; the supernatant was applied to a Qiaprep column and centrifuged for 30 sec at 13000 rpm on a tabletop microcentrifuge; flow-through was discarded, the column was washed with 0,5 ml of PB buffer and spinned at the same conditions; flow through was discarded and the column was also washed with 0,75 ml of PE buffer and spinned again at the same conditions; after another 1 min spinning cycle, DNA was eluted adding 50 ul of EB buffer to the membrane and centrifuged at 13000 rpm for 1 min at a table-top microcentrifuge. The concentration of product was measured using a Nanodrop® spectrophotometer.

The DNA obtained was sequenced outside the Institute by Eurofins MWG Operon. Sequences were manually analyzed to identify colonies carrying the wild type and the mutated allele.

2.5.2. HEK 293 transfection and splicing analysis

2.5.2.1. HEK 293 culture and DNA Transfection

All cell culture procedures were performed under clean conditions on an isolated room and cell manipulation was always performed at a laminar flow hood. Human Embryonic kidney cells 293 (HEK293) were used for this set of experiments. The cell line was generated by the transformation of normal embryonic kidney cells with Adenovirus 5 DNA (188).

HEK293 cells were cultured under standard conditions: DMEM media + 10% fetal calf serum at 37°C. DMEM media was purchased from GIBCO and Fetal calf serum from Sigma. Approximately 2.5 X 10^5 cells were seeded at each well on 6 well (10 cm^2) plates and incubated at 37°C overnight to allow 70-80% confluency. HEK 293 cells were kindly gifted by Caroline Dalgliesh from Prof David Elliot's group

Transfection with the Pxj41+ ROCK1 (Exon 8)/WNT11 (Exon 1) DNA was made using standard procedures. Briefly, 97 ul of DMEM media without serum was mixed with 3ul of the Transfection reagent Gene Jammer from Stratagene ® and incubated at room temperature for 5 minutes. Around 500 ng of DNA was added to the mixture and incubated for 10 minutes at room temperature. The mixture was added to each well and incubated for 24h at 37°C.

After the incubation, cells from each well were washed with 1 ml of PBS and trypsinised: 3 drops of GIBCO® Trypsin-EDTA (0.25% Trypsin; 1mM EDTA) were added to each well and plates were incubated at 37°C for 5 min. The reaction was stopped adding 1 ml of DMEM containing 10% fetal calf serum. The content of each well was transferred to an eppendorf tube and spinned on a bench centrifuge at 6000 rpm for 2 minutes. Supernatant was discarded and pellets properly labeled and stored at -20°C until RNA extraction.

2.5.2.2. RNA extraction and qualitative analysis

RNA from the samples was extracted using the Trizol® reagent from Invitrogen: samples were vortexed and 20ul of chloroform was added to each one. The mixture was vortexed again and incubated for 2 minutes at room temperature. Afterwards, the samples were spun on a bench centrifuge for 15 min at 15000 rpm at 4°C. The aqueous layer obtained was mixed with isopropanol and incubated at room temperature for 10 minutes. After of a new spin of 10 min under the same conditions, supernatant was discarded. The pellet was resuspended on 70% ethanol. The samples were spun for 8 min and ethanol removed by pipette. The pellet air dried and resuspended on RNAse free water. RNA concentration was measured using a Nanodrop® spectrophotometer.

Splicing of the exon inserted was analyzed using RT-PCR. The one step RT-PCR Qiagen kit was used to perform the reaction with the PXJRTF and PXJB1 primers. Manufacturer instructions were followed: 1ul of 5X Qiagen buffer, 400 uM of each dNTP, 1 ul of Qiagen Q solution, 0.2 ul of the enzyme mixture, primers at a concentration of 0.6 uM and 100 ug of the extracted RNA were mixed and cycled following the program specified at Table 25.

N of cycles	Temperature and time
1	30°C for 30 minutes
1	95°C for 15 minutes
34	97°C for 30 seconds 55°C for 30 seconds 72°C for 1 minute
1	72°C for 10 minutes

Table 25. Cycling conditions for One Step RT-PCR

Qualitative comparisons in splicing between the wild type containing exon and the mutated one were performed using agarose electrophoresis. Confirmation of RNAm sequence was obtained by sequencing. RT-PCR products were sent to Eurofins MWG operon to be sequenced. The files received were manually reviewed using the Staden package software.

2.6. ROCK1 tag SNP genotyping

2.6.1. Marker selection and assay design

A search for ROCK1 was performed on the Hapmap database (157). Allele frequencies for the SNPs in the 28 first ROCK1 exons and a region of 15 Kb upstream were dumped from the database (Chromosome 18: 16792848..16960818) for the Utah residents with Northern and Western European Ancestry (CEU) population. The file comprising 167 Kb and 109 SNPs was analysed using Haploview 4.0 software. SNPs with a mAF of less than 0.05 were excluded from the analysis. See Figure 18

Tagging strategy was set for an r^2 threshold of 0.8. A total of 34 SNPs were included and, according to the tagging analysis, captured using 12 SNPs. The software was forced to include the non synonymous SNPs rs2292296 (Leu1097Phe), rs2271255 (Lys222Glu) and rs45449301 (Ile432Val).

The sequence for each one of the variants and 200bp upstream and downstream was obtained from the NCBI web page (39). Sequences were uploaded to the Real SNP web based resource in order to design the assays (165). Assays were designed as described in section 2.3.2. See Table 26 for primer sequences and mass.

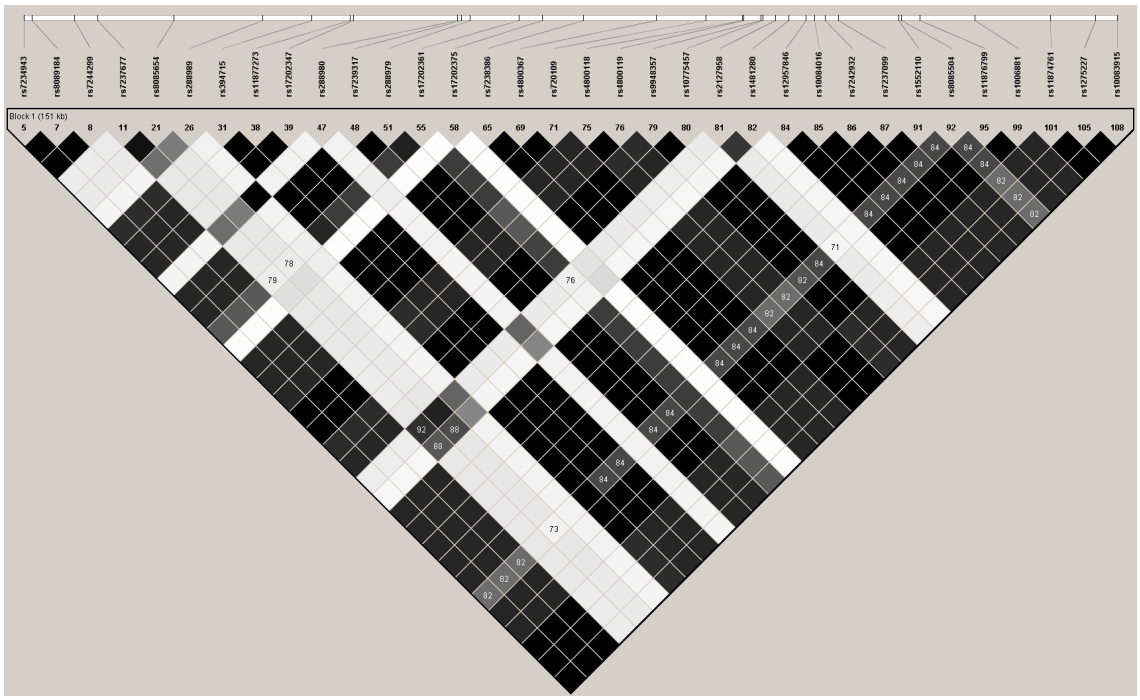

Figure 18. LD plot showing all 34 SNPs included on the tagging analysis of ROCK1. Gray scales denotes R^2 between markers.

SNP	Forward primer	Reverse primer	EP Mass	Extension primer	Allele1	Allele1 Mass	Allele 2	Allele2 mass
rs2292296	ACGTTGGATGTCAGCACTAGGAAAACTAGC	ACGTTGGATGTGAGCAATTGCGTGCTAAAC	5507,6	CACTTGTAGAATCCGAGA	A	5778,8	G	5794,8
rs17202375	ACGTTGGATGCATTATGTCCTTTAGAGCTT	ACGTTGGATGCTGTCTCTATTTAAAAAAAAA	5817,8	TCCTTTAGAGCTTGAAGTA	G	6065	A	6144,9
rs1481280	ACGTTGGATGATCTGAGGTCCTTTGGGATA	ACGTTGGATGCCAATCTCTCTGTTAAACCC	6174	CATGAATAACTAGATGGCAG	C	6421,2	A	6445,3
rs398528	ACGTTGGATGCAGTTCGGTGTTACTGAACAT	ACGTTGGATGCAATGGAGCCAGACCCTGT	6355,2	CATTTTCAAGGCCTTCTTAAA	T	6626,4	A	6682,3
rs8085504	ACGTTGGATGAGGAGGAGGAGAGATTGGTC	ACGTTGGATGCCCCTTCCATCTCTTTTACC	6531,3	AGGAGAGATTGGTCTTTGTAT	G	6778,4	A	6858,4
rs7237677	ACGTTGGATGGGCTGAAATTTGCCTTTATGG	ACGTTGGATGGTCAGTTTAATGAATCTCTG	6726,4	AGTGCTATAAATCAACACTGTA	G	6973,6	C	7013,6
rs288979	ACGTTGGATGACAAAGCATGTAAAGAAGC	ACGTTGGATGCACGCCTGGTGGTTTCTAAC	7155,7	AAAGCATGTAAAGAAGCAAGAAA	A	7426,9	G	7442,9
rs7227454	ACGTTGGATGTTCACCTGTCCTAGAATTTC	ACGTTGGATGGCTGAGTGAAAGAAACCACG	7237,7	CCTGTCCTAGAATTTCATCTACAT	T	7508,9	C	7524,9
rs288989	ACGTTGGATGTCTTTAGGAATTTGCAGTGG	ACGTTGGATGCATCTCTAGAGTAGTATTCC	7323,8	AGTGGTGATCATCTATTTTTATCC	A	7595	T	7650,9
rs2271255	ACGTTGGATGGTTAGCAGATTTTGGTACT	ACGTTGGATGCCTTGAATTAGCAAAAACT	7733	GTTAGCAGATTTTGGTACTTGTATG	T	8004,2	C	8020,2
rs17202368	ACGTTGGATGGCAGCAATACCAGATGCTAC	ACGTTGGATGGATTATAAATTGGAAGTTCA	7950,2	GCAGCAATACCAGATGCTACCAAAAA	C	8197,4	A	8221,4
rs45449301	ACGTTGGATGATTAAAGCAGGAAAGTTTG	ACGTTGGATGGCATTTCATTATGCAGCTGT	8053,3	TAAAGCAGGAAAGTTTGCAAAAAACA	T	8324,5	C	8340,5

Table 26. PCR primer extension primer and mass for the ROCK1 tagged SNPs Sequenom assay.

2.6.2 iPLEX genotyping

All 13 variants were genotyped, using the iPLEX® platform, in 465 CHA probands, 744 CHA parents and 1428 HTO controls. Multiplex PCRs were performed following the standardized process for iPLEX® experiments as described in section 2.3.2. PCR products were cleaned using ExoSAP and extension reactions were performed as described in section 2.3.2.

Reactions were cleaned with a spectroCLEAN enzyme. A Samsung TM nanodispenser was used to transfer 15 nl of sample to an SpectroChip (166). The SpectroChip was introduced and analysed into a Bruker Autoflex MALDI TOF spectrometer, producing the spectra and reporting the alleles for each sample and SNP.

2.6.3. Statistical analysis

Genotypes and pedigrees were stored as Microsoft Office Excel 2003 files. The files containing family-proband data and files containing genotypes were merged together using Microsoft Office Access 2003. The genotypes were checked for Mendelian errors using Pedstats ® (171), the resulting errors were reviewed and deleted when necessary. Additional checks based on unlikely recombination patterns within families were carried out using the error-checking option in MERLIN version 1.1.1 (172). Errors were corrected when possible by reference to the raw genotyping data, and when this was not possible, genotypes were excluded from analysis. The degree of fitness with Hardy Weinberg equilibrium was measured for each marker using the Pedstats ® software.

Population variable frequencies were calculated with SPSS v15.0 software. Allele frequencies were calculated using Pedstats software. Haplotypes were obtained when appropriate using Haploview 4.0 software (173). Extremely rare variants (N=1 or 2) were compared manually. Initial ANOVAs comparing allele frequencies were using SPSS v15.0. Discrete association analysis was calculated with the algorithm present at Haploview software (174) and LAMP v0.0.9 software (175). Graphics were performed using SPSS v15.0 software.

To take some account of the risk of false-positive findings because of multiple comparisons, the results were interpreted based on the false discovery rate (176). The program QVALUE was used running on top of the statistical package R to determine q values (177). An arbitrary false discovery rate threshold of 0.05 was adopted.

3. ROCK1 RESULTS

3.1. CHANGE POPULATION DESCRIPTION

A total of 565 TOF probands recruited to the CHANGE study were available for genetic analysis. For 39 of these probands it was impossible to recruit either parent. For 526 at least one of the parents was available and was recruited, 329 were trios. Two hundred and ninety six probands (52%) were males and 269 (48%) were females. Mean age for the 565 probands was 20 years with an interquartile range (IQR) of 16.71 years. Mean age of mothers was 46.9 years (IQR: 14.41). Mean age of mothers when the proband child was born was 28.6 years (IQR: 8).

Data about personal and family history was obtained through interviews with a genetics trained nurse. In these interviews a total of 94 (16.6%) of the probands reported at least one second degree family member affected by any kind of congenital heart disease. One hundred of the total of TOF probands (17%) showed at least one other congenital abnormality, including developmental and blood disorders. Twenty six of the total 100 (26%) showing other congenital abnormalities, presented 2 or more of them. Classifying the total of 126 other malformations in 13 anatomical-physiological groups, including one for cleft lip and palate: Twenty three (18.25%) corresponded to genitourinary malformations; fifteen (11.9%) to malformations that could be classified as skeletal, including craniosynostosis; fifteen (11.9%) to gastrointestinal malformations; fourteen (11.11%) were soft tissue malformations, including hernias; thirteen (10.3%) to malformations in lungs or airways; eleven (8.73%) corresponded to other cardiovascular malformations additional to TOF; nine (7.14%) were abnormalities in the sense organs; nine (7.14%) corresponded to CNS and developmental abnormalities; eight (6.35%) were malformations in vessels from second degree branching; four were cleft lip or palate (3.17%) cases; three (2.38%) were glandular abnormalities; and two suffered congenital blood abnormalities. See Table 27.

Group	N	Percentage
Genito- urinary	23	18.25
Skeletal	15	11.90
GI	15	11.90
Muscular- soft tissue	14	11.11
Pulmonary- airways	13	10.32
Other CCM	11	8.73
Senses	9	7.14
CNS- developmental	9	7.14
Other circulatory	8	6.35
Cleft lip and palate	4	3.17
Glands	3	2.38
Blood	2	1.59

Table 27. Type of malformations and number of affected individuals at the TOF population. N: Number of affected individuals. Percentage refers to the percentage of the total malformations represented by each group

Nine out of 707 TOF probands' siblings were also affected with TOF (1.2%), giving a λ_S (Sibling relative risk) of 40. One of 62 (1.62%) TOF probands offspring was also affected with TOF, giving a λ relative risk of 54. Calculations were made taking into account the UK population TOF incidence (0.301/1000 live borns).

In terms of maternal conditions, 36 (6.8%) of mothers (526 available) of the probands suffered of pregnancy induced hypertension and 37 (6.8%) suffered any degree of vaginal bleeding during pregnancy, with two of these suffering both conditions. Ten (1.8%) of the mothers were diabetic prior to pregnancy and seven (1.3%) developed gestational diabetes. Of these 17 mothers suffering diabetes 7 were controlled with diet, 9 were receiving insulin and 1 was receiving oral hypoglycemic agents. Diagnosis criteria for diabetes could not be verified in this retrospective study, and the quality of metabolic control during pregnancy is also unknown. Thirteen (2.4%) mothers suffered epilepsy during pregnancy with 11 of them being diagnosed before pregnancy. Five (38%) of the epileptic mothers were receiving medication for the condition.

3.2. Previously undescribed uncommon variants

Sequences of 32 exons in 93 patients were obtained for Rock1. A total of 4065 bp in each individual were screened using methods described in section 2.2.1. The genetic duplication in ROCK1 which I discovered during the course of the work, described in the methods section, prevented the correct identification of genetic variation from exon 27 to 32. Three previously undescribed uncommon variants were encountered at ROCK1. See Table 28

Exon	Variant	Alleles	Function	AA
7	807 C>T	C/T	S	L
16	1785 T>G	T/G	S	S
20	Thr773Ser	C/G	NS	T/S

Table 28. Previously undescribed uncommon variants. N: Number of carriers; AA: Aminoacid; S: Synonymous; NS: No synonymous; L: Leucine; S: Serine; T: Threonine.

3.2.1. ROCK 1 807 C>T

The first uncommon variant is a synonymous change (C/T) in exon 7, mRNA position 807. The change was encountered in 1 patient in heterozygous state at the first mutational screening and is located in a codon translated into Leucine. See Figure 19.

Despite the single base change not being in the consensus splice sequence, computational analysis of the variant in exon 7 showed interesting results. The mutation decreases the strength of the acceptor site at the end of the exon and produces disappearance of a SrP40 binding site located at its position. Additionally, 807C>T is located within the kinase motif of the protein, which increases the probability of functional alterations if this mutation causes splicing defects.

The variant was inherited, in the first proband detected to be carrying this variant, from the mother who does not show an abnormal cardiac phenotype.

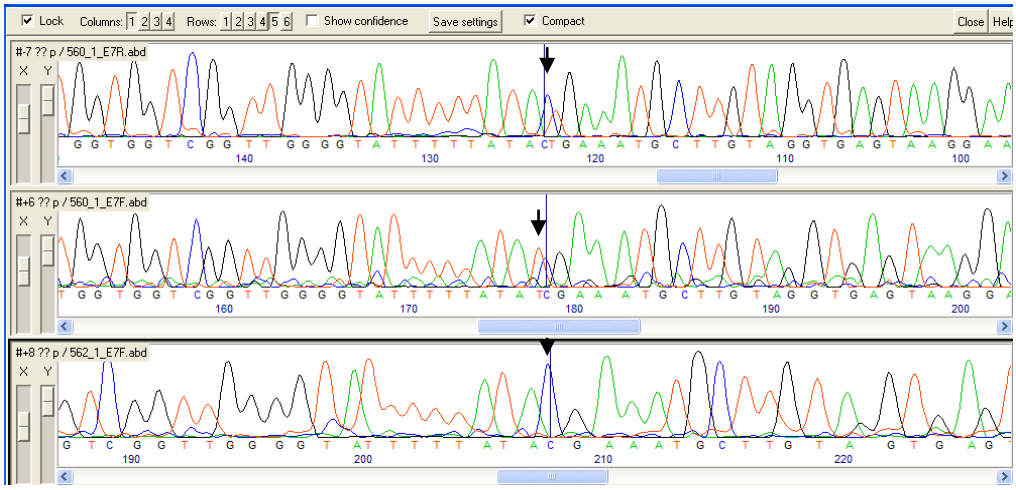

Figure 19. ROCK1 807C>T variant trace display. Upper and medium panel show patient sequence, bottom shows a normal trace.

3.2.2. ROCK 1 807 C>T is associated with TOF

A total of 1908 genotypes for ROCK1 807 C>T were obtained from controls (ECACC, HTO and CHANGE parents) following methods described in section 2.3.2. One thousand and seventy three genotypes (CHANGE, CONCOR, Oxford and Nottingham) were obtained from TOF probands, comprising cases from the CHANGE, Oxford, Nottingham and CONCOR cohorts. Power was calculated as 75% for a general case-control analysis, hypothesising a relative risk of 2 for the heterozygotes and alpha error of 0.05. Genotyping was successful for 98% of the samples. The estimated genotype miscall rate was <1% for the SNP. ROCK1 807 C>T was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Table 29 shows allele counts and frequencies for the ROCK1 807 C>T variant. The variant was present in our population and there are no previously described frequencies in the databases. Thirty three heterozygotes were encountered among probands, whereas only 12 in controls. Four of the probands carrying the variant were males and 9 were females; sex was unknown in 20 of the probands due that they were identified in cohorts which I did not have that information at the time of analysis. Only 2 out of the 13 families, where demographic data were available, showed family history of CHD: one of the probands has a cousin with an unspecific CHD and the sister of another proband is affected with dextrocardia. Two out of the 13 probands carrying the 807C>T variant, where other medical data were available, showed other congenital abnormalities: one of them was born with a hypoplastic left lung and another has eye coloboma.

Using data from 1073 TOF cases and 1908 controls, an association was found between TOF and genotypes at the uncommon variant ROCK1 807 C>T (p=0.000019, LOD: 3.15). Thirty three heterozygotes were found among TOF cases, whereas only 12 were encountered at the control population. Heterozygotes for ROCK1 807 C>T carry a calculated OR for TOF of 4.89 (95% CI: 2.51-9.50). This association was still significant after correction with Qvalue for multiple testing (15 SNPs), using a 5% false discovery rate level. Figure 30 shows a graphical comparison of the proportion of the genotypes in each subpopulation.

	C/C	C/T	T/T	MAF	mAF
Probands	1073	33	0	98.5%	1.5%
Controls	1908	12	0	99.7%	0.3%

Table 29. ROCK1 807 C>T allele counts and frequencies for probands and controls

3.2.3. ROCK1 807 C>T does not affect splicing in a Minigene model

A minigene including ROCK1 exon 8 was constructed following the methods described in section 2.5.1. The minigene was expressed in HEK 293 cells following the methods described in section 2.5.2. Qualitative analysis on agarose gels of RT-PCR products, from total mRNA, using specific primers for the minigene showed no difference in splicing (transcript length and number) when mutant allele constructs were compared to wild type. See

Figure 20. The bands correspond to the mature mRNA transcribed from the Pxj41+ ROCK1 (Exon 8) wild type and Pxj41+ROCK1 (Exon 8 (807C>T)) carrying the variant. In Figure 21 a scheme is shown explaining the possible results of the Minigene experiment.

Sequencing of the RT-PCR product of both minigenes (807 C>T and wild type) showed 100% identity, showing not only an equal product size but also the same RNA nucleotide sequence after transcription for both 807 C>T and wild type.

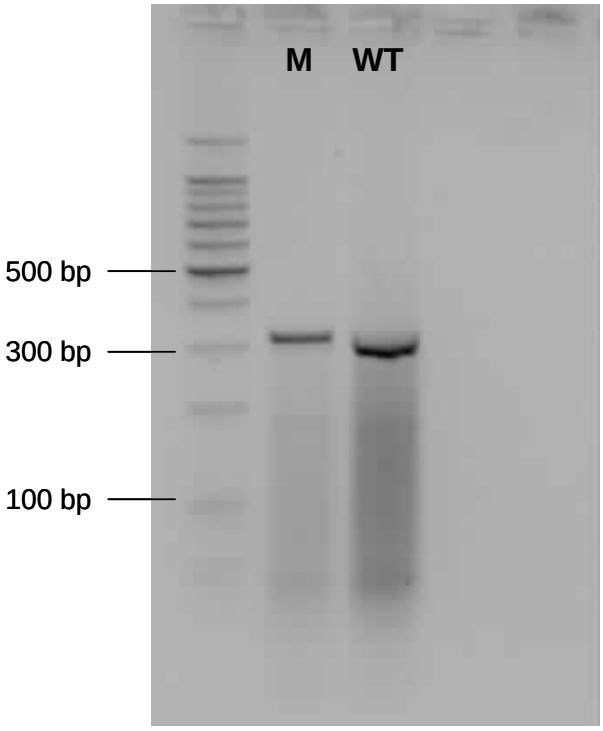

Figure 20. Agarose gel showing the RT-PCR product for minigenes carrying (M=mutant) and not carrying (WT=wild type) the ROCK1 807 C>T variant.

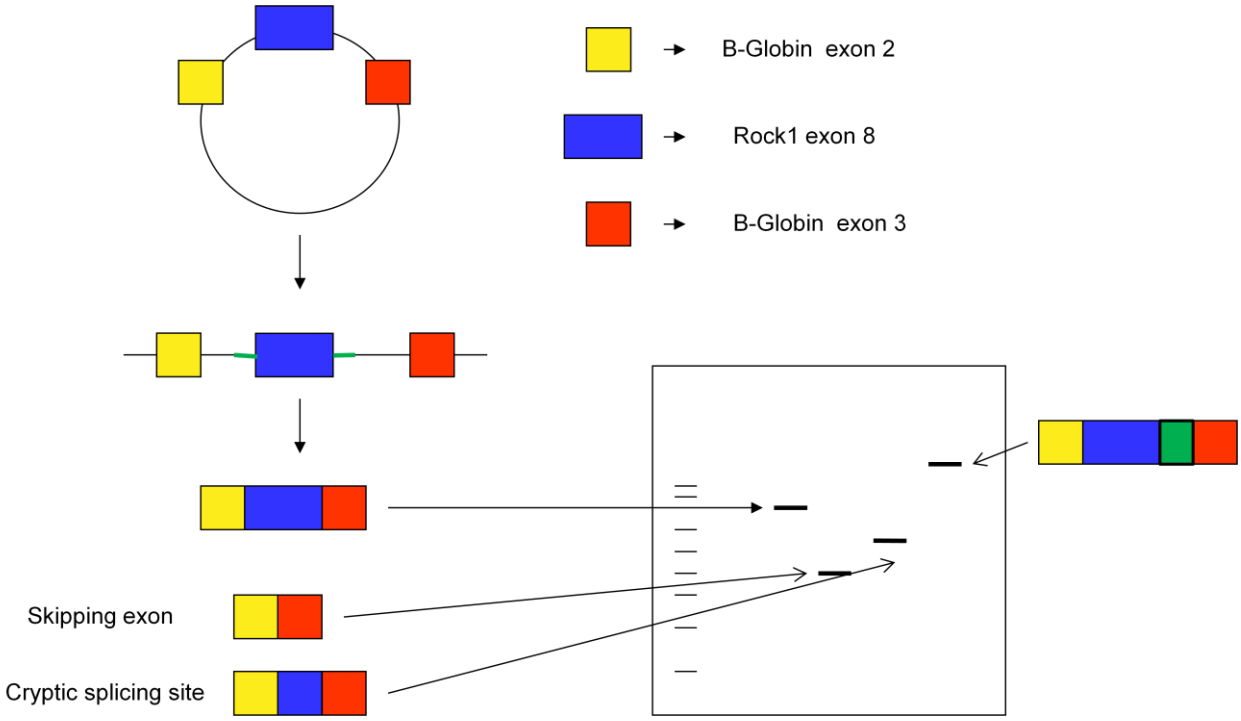

Figure 21. Representation of Minigene product qualitative analysis. The plasmid is represented at the top left corner. In yellow, rabbit β globin exon 1 and in red β globin exon 2. Between β globin exons, in blue, ROCK1 exon 8 with the flanking intron regions in green. At the bottom right a representation of the possible results for the agarose gel. If the exon inserted is skipped, a lighter band proportionally to the exon skipped length is expected. If an exonic cryptic donor splicing site a shorter product will also be expected. In the case of the inclusion of an intronic pseudoexon a larger product would be expected

3.2.4. ROCK1 1785 T>G

The second uncommon variant is a synonymous change (T/G), in exon 16, mRNA position 1785. The variant was identified in 1 patient in the heterozygous state in the initial mutational screening. The codon translates into Serine (535). Serine number 535 is located in the coiled coil region of the protein, and does not belong to recognized conserved domains. The variant is not located in consensus splicing sequences and it does not disrupt known splicing cis regulatory sequences. In silico analysis showed no splicing changes associated.

Samples for both parents were available for the first proband detected to be carrying ROCK1 1785 T>G. Sequences were obtained. The variant was inherited from the mother who does not have an abnormal cardiac phenotype. See Figure 23 for the ABI sequencing traces of the parents of the first proband detected to be carrying the variant.

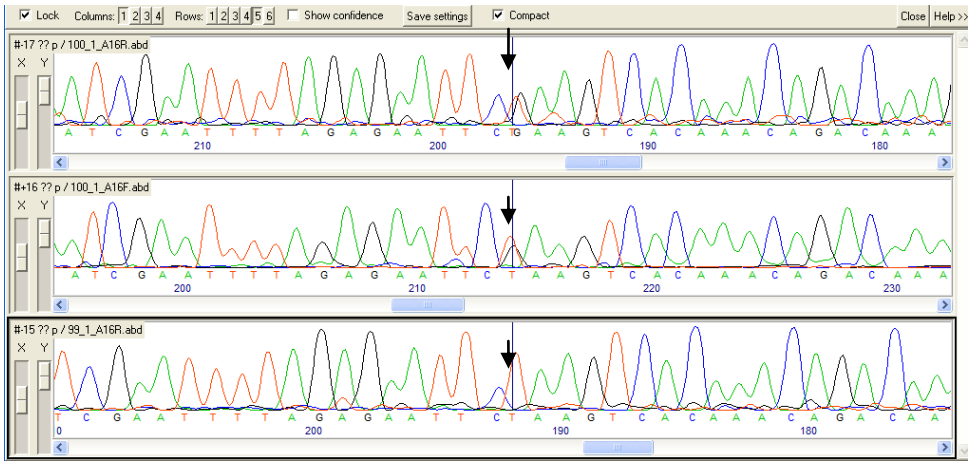

Figure 22. ROCK1 1785 T>G variant. Trace display. Upper and medium panel show patient sequence. Bottom panel shows a normal trace.

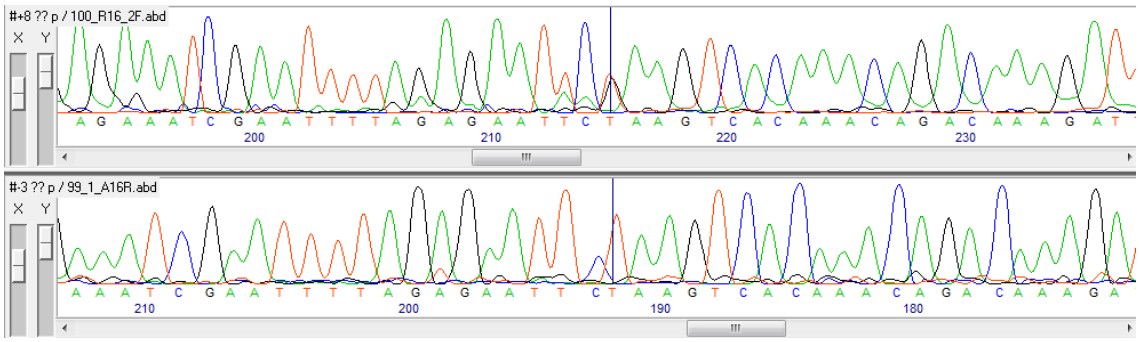

Figure 23. ABI sequencer traces for the parents of the carrier of ROCK1 807C>T. Upper panel shows mother's sequence. Bottom panel shows father's sequence.

3.2.5. ROCK1 1785 T>G is present in controls

One thousand three hundred nineteen HTO controls, 457 probands and 169 CHA parents were genotyped using a Sequenom assay following methods described in section 2.3.2 to assess the presence of 1785 T>G. Ten heterozygotes and no homozygotes were found among controls.

Five heterozygotes and no homozygotes were found among the CHA probands. Three of the probands carrying the mutation were females and 2 males. Only one out of these 5 probands had a family history of CHD with the father and brother being affected by unspecified congenital defects. Both populations are in Hardy Weinberg equilibrium. There are no significant differences when allele frequencies are compared between probands and controls (p=0.499). See Table 30.

.

	CHA probands			HTO controls			
Variant	**n**	**MAF**	**mAF**	**n**	**MAF**	**mAF**	**p value**
1785 T>G	457	0.99453	0.00547	1319	0.99621	0.00379	0.499

Table 30. ROCK1 1785 T>G allele frequencies and p value calculated for the genotype difference.

3.2.6. ROCK1 Thr773Ser

The third uncommon variant encountered is non synonymous (C/G), and is located in exon 20. The variant causes a translational substitution from Threonine to Serine in residue number 773, and was found in 1 patient in the heterozygous state.

Residue number 773 is located at the Tropomyosin recognizing motif of the protein (124). Samples from both parents of the proband carrying the variant were available. The variant was paternally inherited. See Figure 25 for ABI sequencing traces from the parents of the first detected carrier. Abnormal cardiac phenotype was not observed within the family.

To determine the theoretical degree of structural and functional impact of the variant, the Polyphen prediction web based resource was used (158). The variant in aminoacid number 773 appeared benign using this approach.

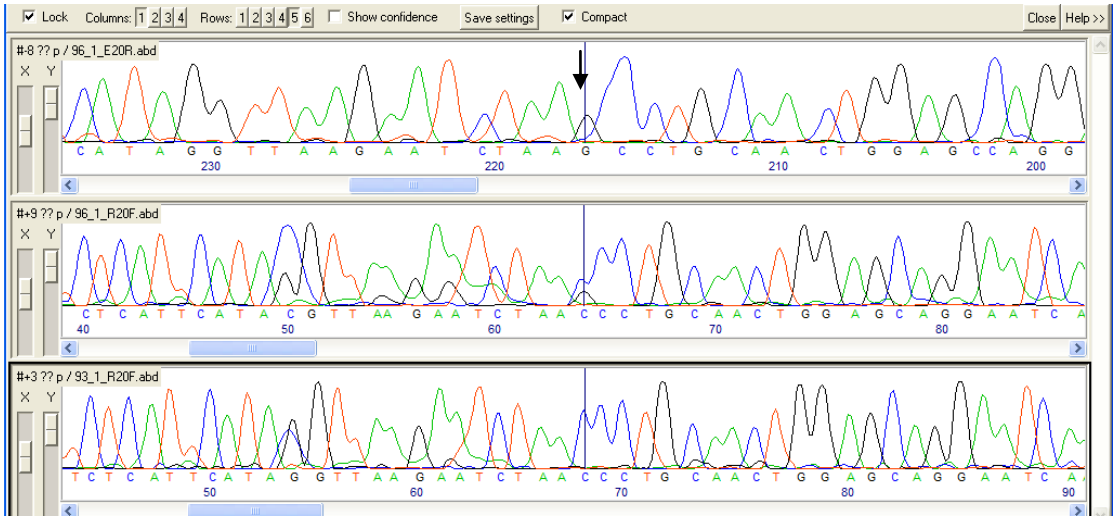

Figure 24. ROCK1 Thr773Ser variant trace display. Upper and medium panel show patient sequence. Bottom panel shows a normal trace.

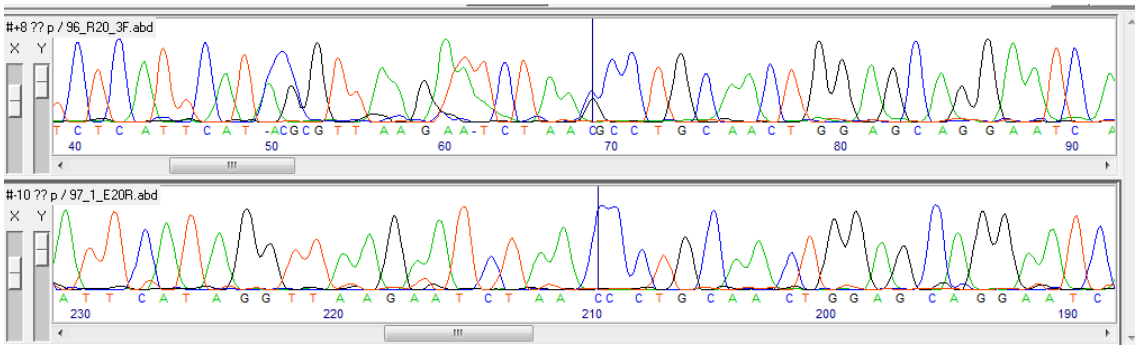

Figure 25. ABI sequencer traces for the parents of the carrier of ROCK1 Thr773Ser. Upper panel shows father's sequence. Bottom panel shows mother's sequence.

3.2.7. ROCK1 Thr773Ser is present in controls

One thousand three hundred and thirty one HTO controls, 458 probands and 169 CHA parents were genotyped using a Sequenom platform and following methods described in section 2.3.2. Forty six heterozygotes and no homozygotes were found among controls.

Seventeen heterozygotes and no homozygotes were found among the CHA probands. Both populations are in Hardy Weinberg equilibrium. Ten of the probands carrying the mutation were females and 7 males. Five out of 17 probands carrying the Thr773Ser variant show other congenital abnormalities: one pulmonary atresia, 2 laryngomalasia, 1 plagiocephaly and 1 present major aorto pulmonary collateral arteries. Six out of the 17 families carrying the variant had a

history of CHD: one of the probands has a cousin and an uncle affected by TOF, 2 probands have cousins with unspecified CHD, 1 proband has a brother with patent truncus arteriosus (PTA), 1 proband has an aunt presenting complete situs inversus and the father of one of the probands has an unspecified CHD.

There are no significant differences when allele frequencies are compared between probands and controls (p=0.77). See Table 31 for allele frequency of the variant in probands and controls and the p value calculated for the difference between them.

	CHA probands			HTO controls			
Variant	**n**	**MAF**	**mAF**	**n**	**MAF**	**mAF**	**p value**
Thr773Ser	458	0.98144	0.01856	1331	0.98272	0.01728	0.77

Table 31. ROCK1 Thr773Ser allele frequencies and p value calculated for the genotype difference.

3.3. Three exons of ROCK1 are duplicated on chromosome 18 and are possibly expressed

Due to 100% heterozygosity of all SNPs located in exons 30-32 a search of the human segmental duplication database was performed identifying a duplication on chromosome 18. The probable source was located at 16774338..16793806 (ROCK1) and the target located at 102547..121692. The human genome clones in ROCK1 (AC021727, AC036178, AC131287, AC135340, AC139704) and the duplicated region (AP001005) were downloaded. The pseudogene was named as Q49AA3 in Ensembl release number 38. The resource DNA dot was used to identify the region overlapping between clones (181). See Figure 26.

The regions identified as duplicated between clones were annealed using the mega 4 software (183) to define the limits of the duplications and differences between clones. The duplicated region comprises a total of 19413 bp. However, the region is not completely identical: a first stretch of sequence of 603 bp located at ROCK1 intron 29, with 97% identity, and 6 gaps is separated by a AT rich region from a second stretch of duplication of 18565 bp, 98% identical and with 87 gaps. The duplicated region ends 9391 bp downstream from the end of Rock1 exon 33. The pseudogene shows a calculated pairwise distance of 0.01 with ROCK1.

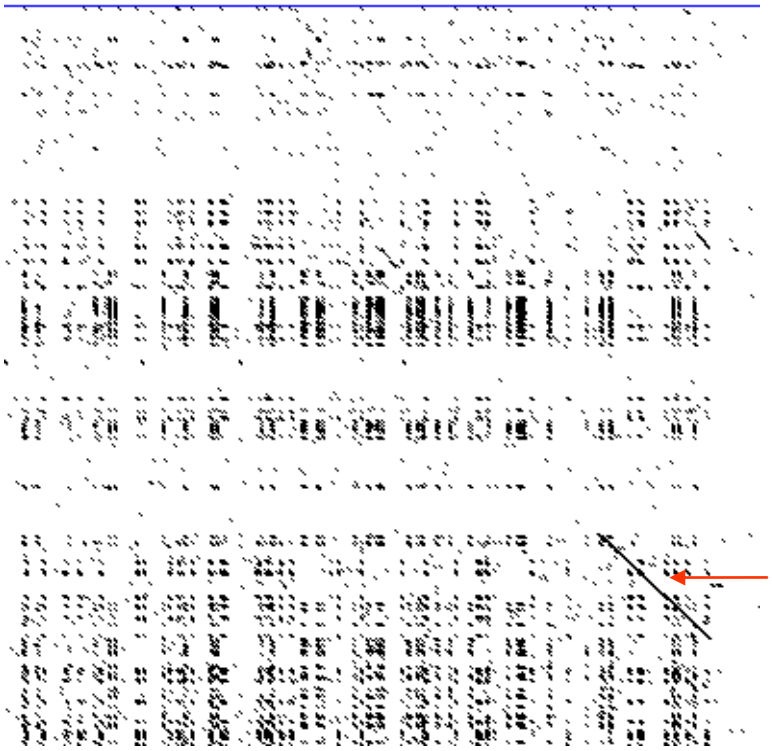

Figure 26. Dot plot for AC021727 and AP001005. Each axis represents a DNA sequence. A dot is placed for each 3 base pair window of identity. Red arrow identifies the duplicated region.

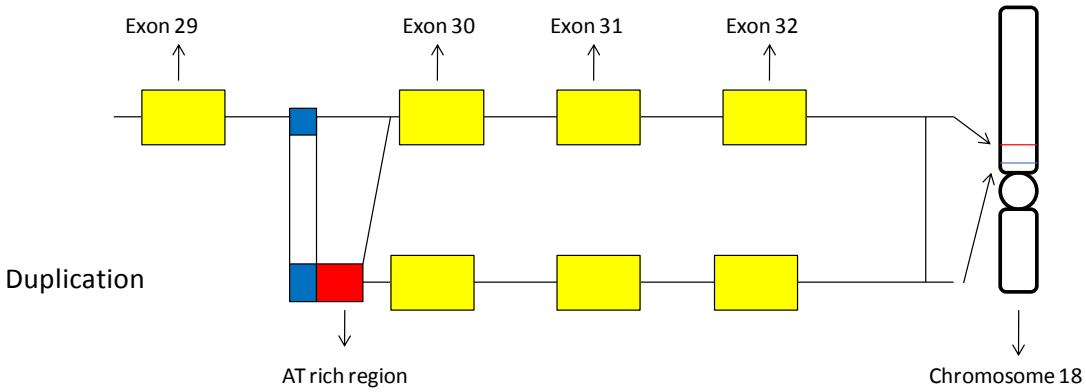

Figure 27. Schematic view of the relationship between ROCK1 and the duplication and their relative position in chromosome 18. In the upper part of the graph Exon 29-31 of ROCK1 are represented as yellow boxes. The blue box corresponds to an intronic region separated by an AT rich region, represented as a red box, from the stretch of exon 30-32 and a portion corresponding to the 5'end of the gene. Rock1 is orientated on the minus strand.

Exonic regions were identified within the pseudogene (Q49AA3). Differences in exonic regions between ROCK1 and the duplication were identified. A region in exon 30 with 2 differences separated by 4bp was used to perform a BLAST search at the human genome EST database for the sequence corresponding to ROCK1 and Q49AA3. The difference in this region allowed differentiation between the products of the pseudogene and ROCK1. Seven BLAST hits showed transcripts with sequences specific to the Q49AA3 pseudogene. The majority of RNA libraries where the transcript was identified were obtained from testis samples. Protein expression has not been documented.

One variant initially thought to be a SNP, 3649 G>C, located in a codon translated into Serine in ROCK 1 exon 31, was detected to be a difference between ROCK1 and the pseudogene. The variant showed 100% heterozygotes in 93 TOF probands. To determine if the duplication was also present in all controls exon 31 was sequenced in 93 HTO controls. Heterozygosity was also 100% in this sample subset for 3649 G>C.

3.4. Tagging SNP genotyping in ROCK1. Common genetic variation in ROCK1 is associated with TOF

ROCK 1 common genetic variation was genotyped using the tagging strategy described in section 2.6.1. Thirty seven SNPs located in the first 28 ROCK1 exons and 15Kb upstream (See Table 32) were captured genotyping 12 SNPs. As described in section 2.6.1, the Haploview 4.0 software was forced to include the previously described non synonymous SNPs rs2292296 (Leu1097Phe), rs2271255 (Lys222Glu) and rs45449301 (Ile432Val). Figure 28 shows a Haploview plot representing all SNPs in ROCK1 and their LD relationships.

Genotyping was successful for at least 98% of the samples for all of the SNPs. The estimated genotype miscall rate was 1% for all of the SNPs. All 13 of the markers were in Hardy-Weinberg equilibrium at the 5% significance level. Table 33 shows genotype frequencies and counts for the population. The minor allele frequencies ranged from 0% to 46.0%, yielding marker heterozygosities from 0% to 46.8%. Allele frequencies were highly similar to previously published Hapmap data for CEU population, see Table 33.

SNP	mAF
rs7234943	0.05
rs8089184	0.05
rs7244299	0.052
rs7237677	0.417
rs7227454	0.045
rs8085654	0.405
rs288989	0.442
rs394715	0.061
rs11877273	0.06
rs17202347	0.058
rs288980	0.442
rs7239317	0.058
rs288979	0.059
rs17202361	0.076
rs17202368	0.025
rs17202375	0.076
rs7238386	0.06
rs4800367	0.058
rs720109	0.058
rs4800118	0.05
rs4800119	0.05
rs9948357	0.056
rs10775457	0.058
rs2127958	0.433
rs1481280	0.375
rs12957846	0.058
rs10084016	0.058
rs7242932	0.058
rs7237099	0.058
rs1552110	0.058
rs8085504	0.058
rs11876799	0.058
rs1006881	0.058
rs11874761	0.05
rs398528	0.027
rs1275227	0.05
rs10083915	0.05

Table 32. ROCK1 SNPs captured with the tagging strategy. mAF: minor allele frequency.

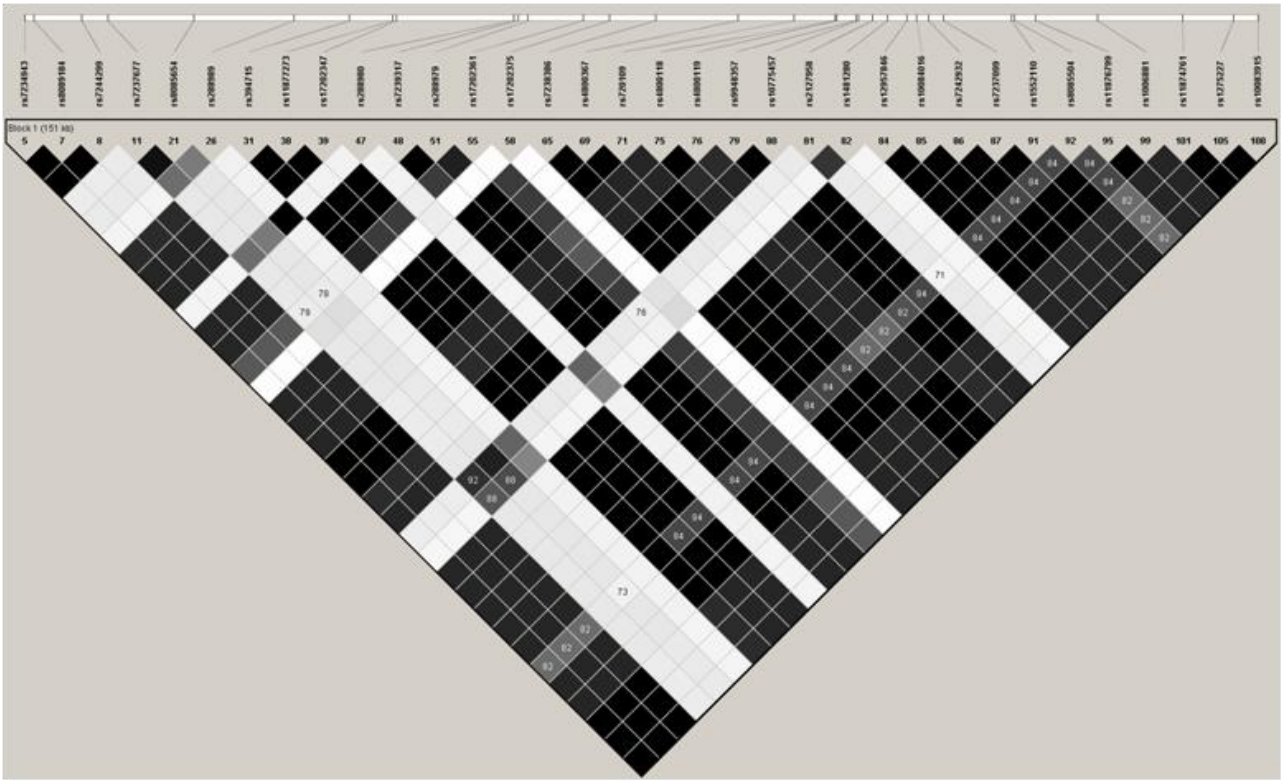

Figure 28. LD plot showing all 34 SNPs included on the tagging analysis of ROCK1. Gray scales denotes R^2 between markers.

		Experimental population						Hapmap CEU	
SNP	**Genotypes**	**Hom WT**	**Heterozygotes**	**Hom NWT**	**mAF**	**MAF**	**Het**	**mAF**	**MAF**
rs2292296	2238	2238	0	0	0.000	1.000	0.000	0.000	1.000
rs7237677	2220	819	1040	361	0.397	0.603	46.800	0.417	0.580
rs7227454	2208	1790	401	17	0.099	0.901	18.200	0.052	0.948
rs288989	2192	698	973	521	0.460	0.540	44.400	0.167	0.833
ROCK1 807 C>T	3026	2981	45	0	0.007	0.993	1.500	*	*
rs45449301	2227	2226	1	0	0.000	1.000	0.000	0.025	0.975
rs288979	2225	2045	167	13	0.043	0.957	7.500	0.059	0.941
rs17202368	2074	1993	79	2	0.020	0.980	3.800	0.025	0.975
rs17202375	2099	1809	278	12	0.072	0.928	13.200	0.076	0.924
rs2271255	2223	2222	1	0	0.000	1.000	0.000	0.000	1.000
rs1481280	2220	901	988	331	0.372	0.628	44.500	0.364	0.636
rs8085504	2228	2061	161	6	0.039	0.961	7.200	0.058	0.942
rs398528	2243	2137	105	1	0.024	0.976	4.700	0.973	0.027

Table 33. Counts and allele frequencies for our population and Hapmap CEU data. Hom WT: Homozygotes wild type; Hom NWT: Homozygotes not wild type; mAF: minor allele frequencies; MAF: major allele frequencies; Het: Heterozygotes.

The three non-synonymous SNPs included in the genotyping showed very low frequencies at our population, similar to previous Hapmap CEU data. Rs2292296 was not found in the population. Rs45449301 and rs2271255 were found only in one heterozygote control.

Correlation between our SNPs was, as expected, modest. Figure 29 shows linkage disequilibrium relationships between the typed SNPs. Seven haplotype blocks containing all SNPs in ROCK1, defined using the four gamete rule (189), with frequencies of more than 1%, account for more than 95% of genetic variation within the population. Haplotype blocks included 9 SNPs: rs7227454, 807 C>T, rs288979, rs288989 rs17202368, rs17202375, rs1481280, rs8085504 and rs398528. Common haplotype composition and frequencies were in close agreement to those described in previous published Hapmap CEU data. Haplotype composition and frequencies are detailed in Table 35.

There was no evidence of strong linkage disequilibrium between any of the tagged SNPs and 807 C>T. Table 34 show LD values between 807 C>T and the tagged SNPs. Tests using LAMP software showed no evidence for linked variants explaining the association found between 807 C>T and TOF.

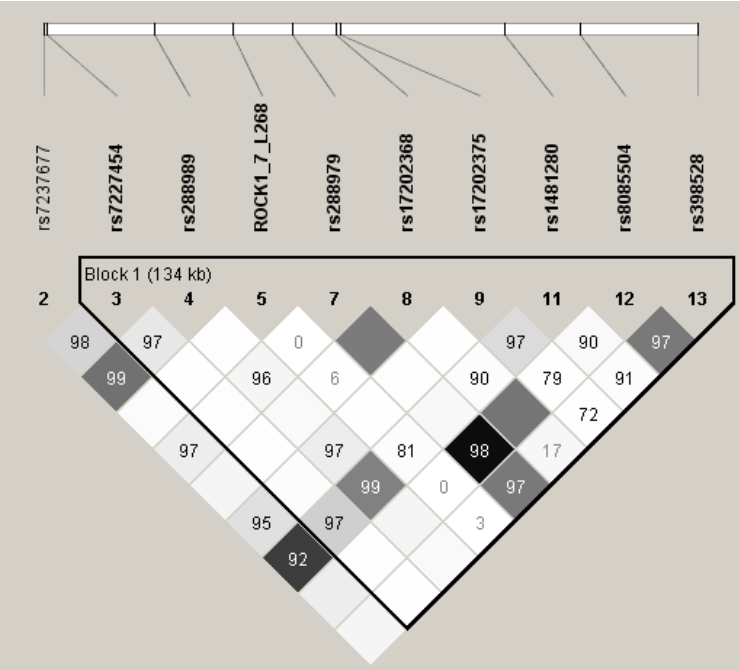

Figure 29. SNPs genotyped (mAF>0.001) and relationships between them. The linkage disequilibrium relationships between SNPs are represented by Haploview triangle plots; darker small squares denote higher linkage disequilibrium (represented by r2), and the numbers in the small squares show the significant pairwise D_ values derived from the present study. Haplotype block is delimited by a line, and designed as Block 1.

SNP	r-square
rs7237677	0,012
rs7227454	0,001
rs288989	0.001
rs288979	0,000
rs17202368	0,002
rs17202375	0,001
rs1481280	0,008
rs8085504	0,000
rs398528	0,000

Table 34. Quantified linkage disequilibrium for our population between the tagged SNPs (mAF<0.001 not included) and 807 C>T. LD is expressed as R-square

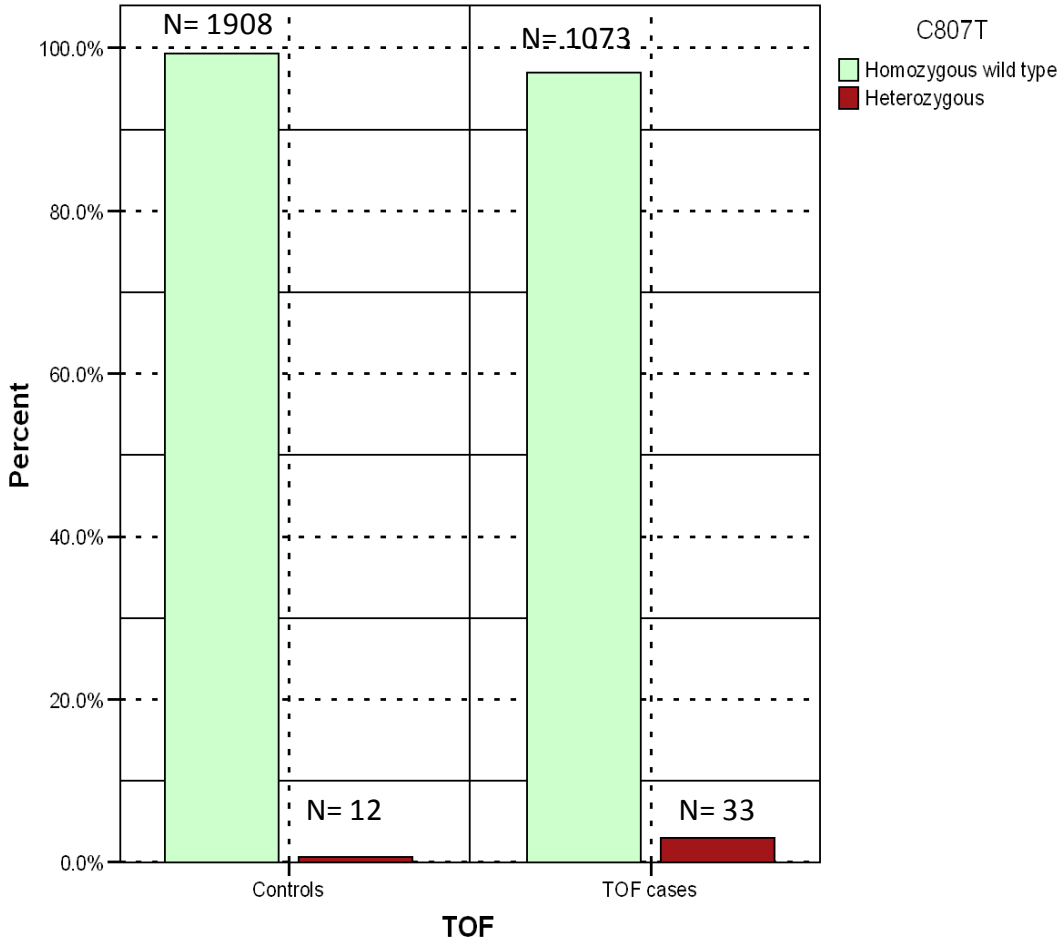

Figure 30. Comparison of the proportion of genotypes, for ROCK1 807 C>T between TOF patients and controls. Allele frequencies show a significant difference (p=0.000019) between probands and controls.

Haplotype	Frequency
GAGAGAGCA	0.468
GTGAGAACA	0.180
GTGAGAGCA	0.128
TTGAGAACA	0.098
GTGAGCACA	0.074
GTGCGAGAT	0.019
GTGCAAGAA	0.019

Table 35. Common Haplotypes (Frequency >1%) for ROCK1 within our research population. Genotypes are specified for the following SNPs: rs7227454, 807 C>T, rs288989, rs288979, rs17202368, rs17202375, rs1481280, rs8085504 and rs398528.

Haplotype	p value
GAGAGAGCA	0.3925
GTGAGAACA	0.2666
GTGAGAGCA	0.6174
TTGAGAACA	0.037
GTGAGCACA	0.4239
GTGCGAGAT	0.2762
GTGCAAGAA	0.272

Table 36. Chi square p values for association between common haplotypes and TOF.

Association analysis, using Chi square tests, for the common Haplotypes found in the population showed no association between blocks and TOF. Table 36 shows common Haplotypes and calculated p values.

In the allele frequencies analysis with LAMP software, an association between rs288979 and Tetralogy of Fallot (p=0.000015) was found. Rs288979 is an intronic uncommon SNP (mAF= 0.043) located in ROCK1 intron 9, at 8189 bp from exon8/Intron boundary. Case/control ratios encountered were: AA: 1666/395 AC: 121/40 CC: 3/3. See Figure 31 for a graphical comparison of genotype frequencies. The minor allele (C) confers a protective effect with a calculated OR of 0.6127 (95% CI: 0.434-0.866) for TOF. In silico analysis of the variant showed no evidence of splicing changes associated with it. This association was still significant after correction with Qvalue for multiple testing (12 SNPs), using a 5% false discovery rate level.

There was no evidence for linkage between rs288979 and the other tagged SNPs and tests using LAMP software showed no evidence for linked variants explaining the association found. Rs288979 is not in linkage disequilibrium with ROCK1 807 C>T. Only one proband carrying ROCK1 807C>T also carries the allele C of rs288979 in the heterozygote state. No other significant association was found between any of the variants and TOF. Table 37 shows the p value calculated using LAMP for the differences between cases and controls in allele frequencies for each SNP genotyped.

SNP	p value
rs7237677	0,690000
rs7227454	0,230000
rs288989	0,151600
rs288979	0,000015
rs17202368	0,390000
rs17202375	0,530000
rs1481280	0,119000
rs8085504	0,116000
rs398528	0,240000

Table 37. Calculated p value for a model comparing the null hypothesis of no association with association models for the tagged genotyped SNPs.

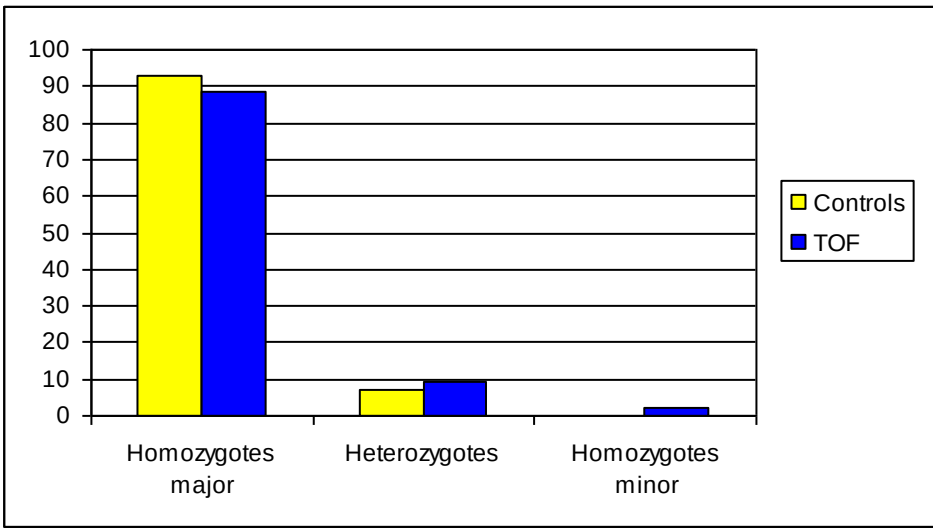

Figure 31. Bar plot comparing genotype frequencies for rs288979. Yellow bars represent control subjects and blue bars TOF subjects.

3.5. DISCUSSION

To our knowledge this is the first study to examine common and uncommon genetic variation at ROCK1 and TOF. Also, our sample is the largest reported and the only thus far with sufficient power to explore genetic variation and its relationship with non-syndromic TOF.

It would be expected that the loss or decrease of function in ROCK1 would be associated with cardiac malformations, as illustrated by previous experimental data showing expression of the protein during cardiogenesis (121, 126) and regulatory functions of the protein in convergent extension movements in *Xenopus*, *Drosophila* and *C. Elegans* (128, 129). In 2002 Zhao and cols published a paper showing that inhibition of Rho kinases with Y27632 on cultured murine embryos causes failure in chamber septation, heart looping and cardiac cushion proliferation (131). His group also showed that Rho kinases are expressed throughout cardiogenesis, data in accordance to the work published by Wei and cols (126). However, ROCK1 knockout mice do not show cardiac phenotype (133), showing perhaps the functional overlapping between ROCK1 and 2.

Three previously undescribed variants were found at ROCK1. The first one 807 C>T is a synonymous variant located at exon 7, at a codon translating into Leucine. Despite the variant not being located at consensus splice sites, automated splice analysis showed that the minor allele may decrease the splicing acceptor site strength and causes disappearance of a Srp40 binding site. The variant is uncommon, with a minor allele frequency of 0.7% in the population (n= 3026, cases frequency: 1.5%; controls frequency: 0.3%). Allele frequencies are significantly different between cases and controls (p=0.000019), and the minor allele confers an OR of 4.89 for >TOF. Qualitative analysis of families carrying the variant showed no evidence of higher incidence of cardiac or other congenital defects compared to other proband families not carrying the variant. Also, genotyped parents carrying the variant did not show abnormal cardiac phenotype.

However, in our splicing experimental model the variant does not cause aberrations, leading to the possibility of 807 C>T being linked to a different and unknown causal variant or causing its effects via different mechanism like quantitative changes in expression levels. At the initial mutational screening (n=93), I did not encounter any other uncommon variant in ROCK1 or in any other of the screened genes, to be linked to 807 C>T. Nevertheless, the mutational screening sample size would be insufficient to detect variants with allele frequencies of roughly less than 1/186.

To explore the hypothesis that the association between 807C>T is due to a linked variant, a tagging strategy (190) was designed to capture genetic variation within the gene and linkage disequilibrium relationships between markers. Previously described missense SNPs were forced to be included. Previously described missense SNPs were practically absent in our sample.

No evidence of LD was found between 807 C>T and the other markers, no haplotype was associated with TOF and remarkably another intronic common variant (mAF: 0.051) showed association with TOF. Rs288979 is significantly associated with TOF (p=0.000015) and its minor allele confers a protective effect with an OR of 0.61. In silico analysis of 807 C>T did not showed a functional splicing explanation for the association. Again, as in the association for 807 C>T, there is neither a linked variant to explain an indirect association and further work is necessary to elucidate the mechanism underlying the association found.

The two relatively uncommon variants are associated with TOF may be considered “low frequency intermediate penetrance” variants. However it would be also expected that these variants would be subjected to purifying selection in the population depending on their risk association (11, 22). In this respect the two variants differ in a way: whereas rs288979 confers a near to “expected” modest protective effect (OR= 0.61), the variant 807 C>T confers a high risk to suffer TOF (OR= 4.89). This high risk associated with the variant would make the purifying selection, that the SNP is subjected to, greater than mild and its frequency would accordingly be (without an extraordinary mutation rate) very low as we found (191).

Purifying selection in non-syndromic TOF sufferers has been greatly influenced by modern surgical therapy. About 86% of TOF affected individuals survive beyond 30 years (59). Supposing a hypothetical model where TOF predisposing variants would only be deleterious due to the increased probability of cardiac malformation, natural selection over the variants would be greatly decreased by successful cardiac surgery and even highly deleterious variants might reach detectable levels at the population. Nevertheless, the number of generations after the development of successful surgical techniques is low and even if the fitness of TOF patients is close to 1, not taking into account an improbable selective advantage, the only quantifiable effect may be the maintenance of the allele frequencies dependent on the mutation rates. In a sort of mixture between common and uncommon variant theory, decrease of natural selection forces could make the TOF population a group of individuals where common and uncommon predisposing variants coexist.

The effect size for rare variants is predicted by theory to be larger than that for commoner variants. If mAF frequencies are examined, 807 C>T is almost sevenfold less common than rs288979, frequencies are inversely related to the effect conferred by each one of them (807 C>T= 4.89; rs288979= 0.61).

Functional mechanisms were explored for 807C>T with the hypothesis that (due the findings of the in silico analysis and being a variant predicted to be synonymous) the most likely functional explanation for the association was a splicing change. However, the minigene did not show splicing changes associated with the variant. A functional explanation for the association with rs288979 is also yet to be found, the most likely mechanism, due to its intronic location, also an influence over splicing. A negative minigene experiment, however, does not discard the hypothesis of the variant exerting its influence through splicing changes.

With the premise that the rs288979 also influences splicing, a minigene model for the variant as a starting point is needed. However, to confirm or discard this idea with 807C>T either a direct RT PCR amplification of RNA from a subject carrying the variant, sample type that was not available, or the expression of a bigger construct, perhaps the complete gene, in a mammalian cell line with quantitative and qualitative analysis of RNA would be required to give the definite answer. Additionally, the variants might also influence the rate of expression of ROCK1. The previously described experiments with abundance quantification of ROCK1 RNA might clarify that hypothesis.

In the case of indirect association being the cause of the associations found, future work should explore both common and uncommon genetic variation to discover causative alleles. First, mutational screening of the pathway, involving multiple genes or even complete genomes/exomes in bigger samples using next generation sequencing platforms could unravel uncommon genetic variants associated and not detected in our mutational screening.

Even if a functional explanation for the associations found is discovered, they would only account for a small fraction of the heritability of the trait. Systematic approaches that are recently feasible utilising the increased understanding of human genetic variation anticipated to come from the 1000 genomes project, the increasing capacity of genotyping chips to incorporate intermediate-frequency (0.5% to 5%) variants, and next-generation resequencing studies for rarer variants, will be required to identify the remaining heritability of the trait.

3.6. CONCLUSIONS

I found that common and uncommon genetic variation at ROCK1 is associated with TOF in Caucasian subjects. Three previously undescribed uncommon variants were encountered in the sample. One of the newly described uncommon variants, 807 C>T, a synonymous SNP, is significantly associated with TOF. 807 C>T confers an OR of 4.5 for TOF. Splicing functional data did not give an explanation for the association found. A tagging strategy was adopted to explore

the possibility of 807C>T being linked to another variant. No variant was found to be linked to 807C>T.

Another association was encountered between rs288979, an intronic SNP, and TOF. In silico analysis of the variant did not show splicing aberrations associated with the variant. Rs288979 confers an OR of 0.61 for TOF. Despite the low likelihood of a protective effect being exerted by a relatively uncommon variant, this is not the only explanation possible for the OR found. Natural selection or indirect associations can explain the decreased number of TOF probands carrying the variant.

In conclusion one uncommon synonymous SNP and a relatively uncommon intronic SNP located at ROCK1 are associated with TOF. No functional explanation was found for direct or indirect association with the variants. It remains a question if whether the associations found cause splicing changes not detected by the experimental means used or are explained by linkage disequilibrium with undetected uncommon variants. The fact that two variants are significantly associated with TOF, neither of which were in significant LD with each other, independently confirm that there is an association of ROCK1 with TOF. Due that the strongest a animal model, the ROCK1 KO mouse, does not show a cardiac phenotype is unlikely that isolated genetic variation at ROCK1 can be responsible for the predisposition to TOF

This study represents the first large sample to explore genetic variation at ROCK1 in association with TOF. Further work is needed to elucidate the functional implications of the associations found.

4. ANKRD6 AND WNT11

4.1 ANKRD6

4.1.1. Previously undescribed uncommon variants

Sequences of 16 exons, including the 5' UTR region, were obtained for ANKRD6 in 93 CHANGE probands. A total of 2490 bp were screened using methods described in section 2.2. Two previously undescribed uncommon variants were encountered at ANKRD6. See Table 38.

Exon	Variant	Alleles	Function	AA
1	1 -304 A>G	A/G	UTR (-304)	
7	Leu192Phe	C/T	NS	L/F
7	Thr176Asp	C>A	NS	T/N

Table 38. ANKRD6 Previously undescribed uncommon variants. AA: Aminoacid; NS: Non synonymous; L: Leucine; F: Phenylalanine; N: Asparagine; T: Threonine.

4.1.1.1. ANKRD6 1-304 G>A

The first novel variant found was a single base substitution (A/G) in the ANKRD6 5' UTR region, position 1 -304. The variant was found in a single CHANGE proband at the initial mutational screening. See Figure 32 for the ABI sequencer traces from the carrier and a proband not carrying the variant.

In silico analysis of the variant using the web resource NSITE (192) to identify regulatory motifs in the sequence revealed that despite being located at the 5' UTR the variant is not located at any known regulatory motif. It was not possible to determine if the variant was inherited or de novo, as we had no DNA sample from either mother or father. Neither of the parents presented an abnormal cardiac phenotype.

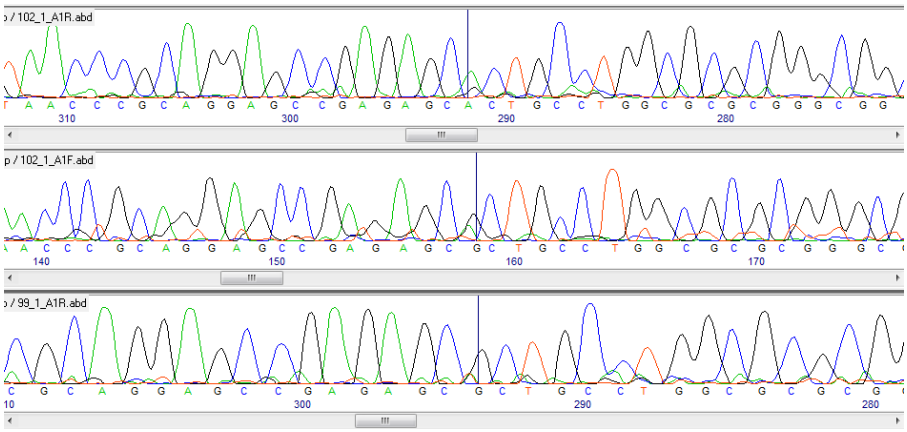

Figure 32. ANKRD6 1-304 G>A variant trace display. Upper and medium panel show patient sequence, bottom shows a normal trace.

4.1.1.2 ANKRD6 1-304 G>A is present in controls

A total of 1744 genotypes were obtained from cases and controls using the Sequenom platform and the methods described at 2.3.2. . The estimated genotype miscall rate was less than 1% for the SNP. 1- 304 G>A was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations. Genotyping was successful for 98% of the samples.

Five heterozygotes were found in both populations: probands and controls. I did not find a significant difference at genotype or allele frequencies between cases and controls for ANKRD6 1-304 G>A (Chi-squared [1df] = 0.052; p= 0.061). For allele counts and genotype frequencies see Table 39.

	G/G	G/A	A/A	MAF	mAF
Probands	411	5	0	0.9940	0.0060
Controls	1323	5	0	0.9981	0.0019

Table 39. Genotype counts and allele frequencies for ANKRD6 1-304 G>A. MAF: Major allele frequency, mAF: minor allele frequency

4.1.1.3. ANKRD6 Leu192Phe

The second novel variant found was a single base substitution (C/T) in ANKRD6 exon 7, producing a change from Leucine to Phenylalanine at codon number 192. This change from a non-polar aminoacid to a similar non-polar one is located at the end of the ankirin repeats domain, which is the described functional domain of the protein. In silico analysis of the variant using the web based resource Polyphen did not indicate that it is structurally damaging. The variant was originally encountered in a single CHANGE proband in the heterozygous state. Figure 33. shows ABI sequencer traces for the wild type allele and the variant

It was not possible to determine if Leu192Phe was inherited or de novo, as we had no DNA sample from either mother or father. Apparently neither of the parents presented an abnormal cardiac phenotype.

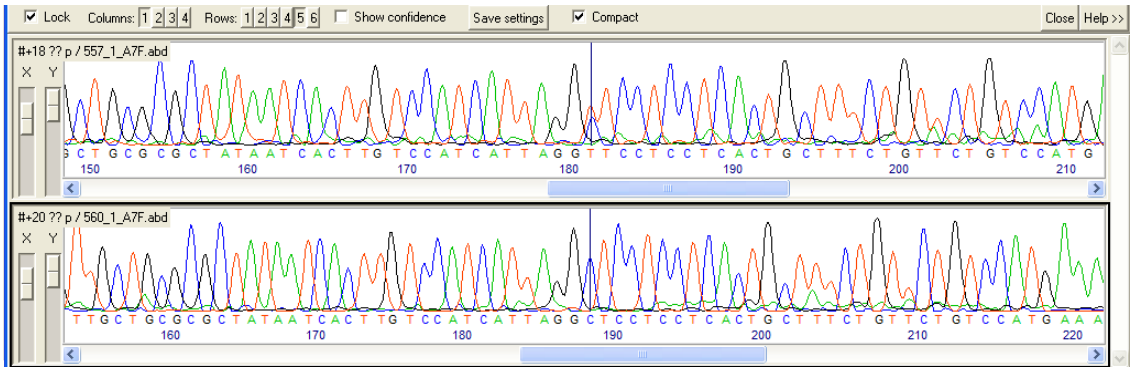

Figure 33. ABI sequencer traces for Leu192Phe. Upper panel shows the variant carrier and the bottom one am homozygous wild type subject.

4.1.1.4. ANKRD6 Leu192Phe is present in controls

A total of 2779 genotypes, in 1062 TOF probands (CHANGE, Oxford, Nottingham and CONCOR) and 1718 controls (ECACC and HTO), were obtained using the Sequenom platform and the methods described at 2.3.2. Genotyping was successful for 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Leu192Phe was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

A total of 2 heterozygotes were found in the population, with one heterozygote found in both populations: probands and controls. I did not find a significant difference in genotype or allele frequencies between cases and controls for ANKRD6 Leu192Phe (Chi square: 1.737, p value: 0.41). For allele counts and genotype frequencies see Table 40. No sample was available from parents of the control subject carrying the variant.

	C/C	C/T	T/T	MAF	mAF
Probands	1060	1	1	0.9995	0.0005
Controls	1717	1	0	0.9997	0.0003

Table 40. Genotype counts and allele frequencies for ANKRD6 Leu192Phe. MAF: Major allele frequency, mAF: minor allele frequency

4.3.1.5. ANKRD6 Thr176Asp

The third uncommon variant found was a single base substitution (C>A) in ANKRD6 exon 7, producing a change from Threonine to Asparagine at codon number 176. The change from a polar aminoacid to a similar polar one is located in the ankirin repeats domain, which is the described functional domain of the protein. In silico analysis of the variant using the web based resource Polyphen indicated that the variant might be structurally damaging. The variant was originally encountered in a single CHANGE proband in the heterozygous state. Figure 34 shows ABI sequencer traces for the wild type allele and the variant.

Sequences of exon 7 on the parents of the proband carrying the variant revealed that it was maternally inherited. The mother carrying the variant does not have an abnormal cardiac phenotype. Figure 34 below shows the sequencer traces for the parents. There was no family history of CCM among the family of the carrier.

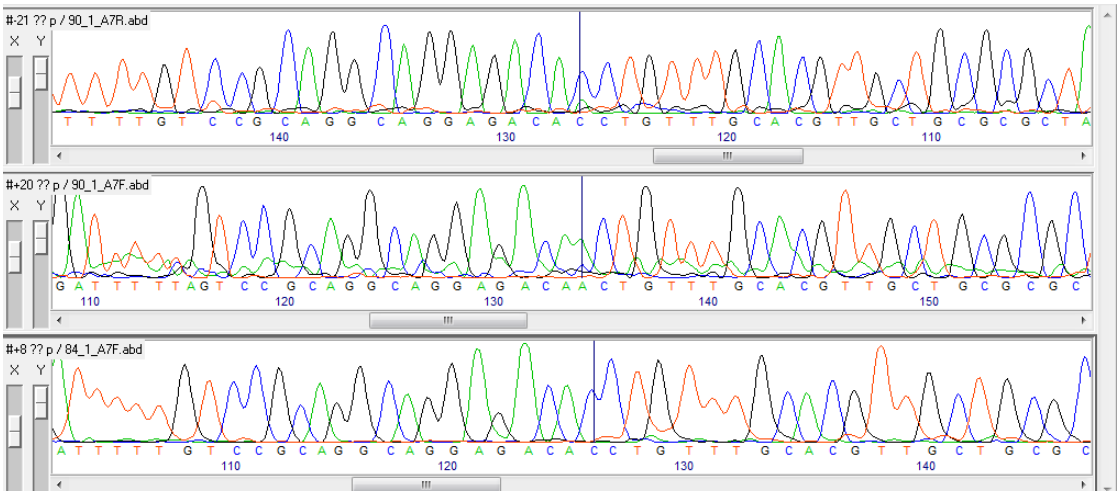

Figure 34. ABI sequencer traces for ANKRD6 Thr176Asp. Upper and medium panels show the variant carrier and the bottom one a homozygous wild type subject.

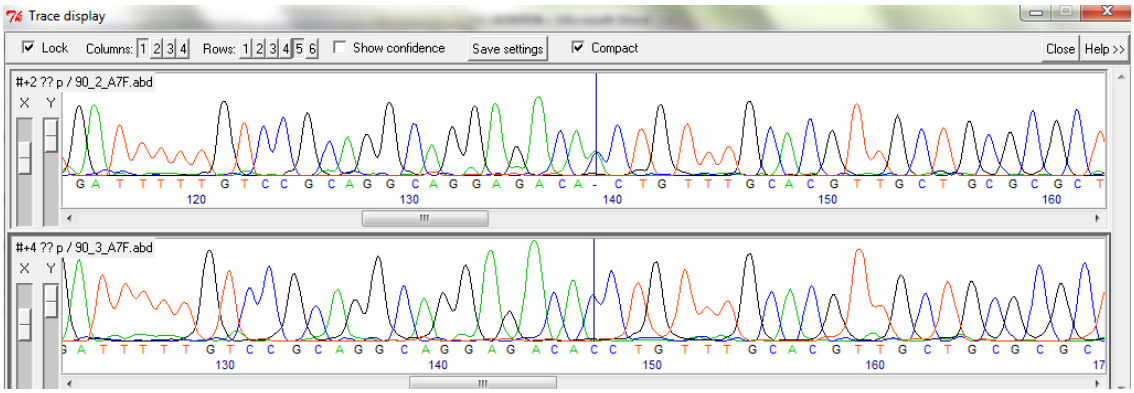

Figure 35. ABI sequencer traces for ANKRD6 Thr176Asp in the parents of the carrier. Upper panel shows mother's sequence. Bottom panel shows father's sequence.

4.3.1.6. ANKRD6 Thr176Asp is present in controls

One thousand six hundred and fifty two genotypes were obtained from CHANGE TOF probands (N=459) and HTO controls (N=1193) using the Sequenom platform and the methods described at 2.3.2. Genotyping was successful for more than 95% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Thr176Asp was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Three heterozygotes were found at the population, with one heterozygote found among the TOF probands and 2 among the controls. I did not find a significant difference at genotype or allele frequencies between cases and controls for ANKRD6 Thr176Asp (Fisher exact test p value: 0.624). Allele counts and genotype frequencies are shown at Table 41.

	C/C	C/T	T/T	MAF	mAF
Probands	458	1	0	0.9989	0.0011
Controls	1191	2	0	0.9992	0.0008

Table 41. Genotype counts and allele frequencies for ANKRD6 Thr176Asp. MAF: Major allele frequency, mAF: minor allele frequency.

4.1.2. Previously unreported common variants

One previously undescribed common variant was encountered in the mutational screening of the first 93 CHANGE probands.

4.1.2.1. ANKRD6 Phe636Leu

The common variant found was a single base substitution (C/T) in ANKRD6 exon 16. The variant produces a change from phenylalanine to leucine at residue 636. The change from a non-polar aminoacid to a similar non-polar one is not located in conserved domains on the protein sequence. However, in silico analysis of the variant using the web based resource Polyphen indicated that the variant is probably structurally damaging. Additional in silico analysis showed that it probably does not produce splicing changes. The residue has not been reported as functionally essential.

The variant was originally encountered in 21 heterozygotes and 1 homozygote of the CHANGE proband population (N=93). Figure 36 shows ABI sequencer traces for the wild type allele and the variant.

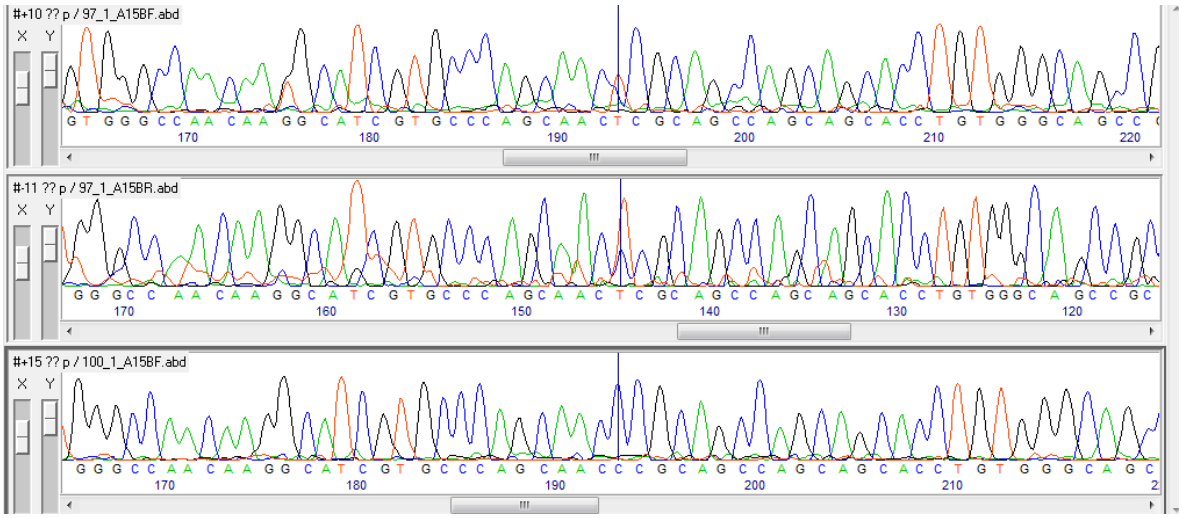

Figure 36. ABI sequencer traces for ANKRD6 Phe636Leu. Upper and medium panels show the variant and the bottom one a homozygous wild type subject.

4.1.2.2. ANKRD6 Phe636Leu is present in controls

One thousand seven hundred and fifty eight genotypes were obtained from CHANGE TOF probands (N=442) and HTO controls (N=1316) using the Sequenom platform and the methods described at 2.3.2. Given the number of cases and controls and minor genotype number, power was calculated as 100% for a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.05. Genotyping was successful for more than 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Phe636Leu was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Three hundred and eighty three heterozygotes were found in the population. A significant difference was found in genotype frequencies between cases and controls for ANKRD6 Phe636Leu (Chi square: 8.51, p value: 0.014). However, after correcting using q value for multiple testing the result was not significant at a 0.05 cutoff (q value: 0.11). Allele frequencies and genotype counts are shown at Table 42.

Increasing the sample, including Oxford, Nottingham and CONCOR probands, and ECACC controls, to 1686 controls and 1041 cases showed no association between Phe636Leu and TOF (Chi square: 5.41, P value: 0.067). For allele frequencies and genotype counts see table 6.

	C/C	C/T	T/T	MAF	mAF
Probands	315	118	9	0.8462	0.1538
Controls	1026	265	25	0.8803	0.1197

Table 42. Genotype counts and allele frequencies for ANKRD6 Phe636Leu. MAF: Major allele frequency, mAF: minor allele frequency.

	C/C	C/T	T/T	MAF	mAF
Probands	766	258	17	0.8598	0.1402
Controls	1686	354	32	0.8991	0.1009

Tabla 43. Genotype counts and allele frequencies for ANKRD6 Phe636Leu including CHA, Oxford, Nottingham and CONCOR probands, and ECACC and HTO controls. MAF: Major allele frequency, mAF: minor allele frequency.

4.1.3. Common previously described variants

A total of nine previously described SNPs were encountered in the mutational screening on 93 TOF CHANGE probands. Most of them were in Hardy Weinberg equilibrium and showed no significant differences in allele frequencies when compared to previously reported data on Caucasian populations (CEU population at Hapmap). For a table showing all SNPs, their minor and major allele frequencies, and their p values for fit to the Hardy Weinberg equilibrium, and for comparisons to the published CEU Hapmap data (N=90) see Table 44.

Three common previously described variants at ANKRD6 were genotyped in the control population and the total of available CHANGE samples due to their lack of Hardy Weinberg equilibrium and also the lack of previous published allele frequencies data: rs2273238, rs9344950 and rs9353687.

One thousand seven hundred and ninety eight genotypes were obtained from CHANGE TOF probands (N=453) and HTO controls (N=1345) for rs2273238; 1769 for rs9344950 (Probands: 455, controls: 1314) and 1772 (Probands: 449, controls: 1323) for rs935368, using the Sequenom

platform and the methods described at 2.3.2. Power was calculated as 100% for all 3 SNPs, using a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.05. Genotyping was successful for more than 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Rs2273238, rs9344950 and rs9353687 were in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

SNP	Function	AA	mAF	MAF	HW p value	P value comp
rs10944453	Int		0.2000	0.8000	0.9800	0.6800
rs16881983	NS	122(Q/E)	0.0100	0.9900	0.9900	0.6800
rs3748085	NS	128(I/V)	0.1400	0.8600	0.8100	0.8100
rs1179906	Int		0.1400	0.8600	0.9700	0.0680
rs2273238	NS	233(T/M)	0.0400	0.9600	0.0510	NI
rs17292811	S		0.2200	0.7800	0.9600	0.9400
rs9344950	NT 3'		0.0400	0.9600	0.0000	NI
rs9353687	S	C	0.1279	0.8721	0.0432	NI
rs3210511	S	S	0.4505	0.5495	0.9754	0.4832

Table 44. Previously described SNPs in ANKRD6, Allele frequencies (N=93) and p values for Hardy Weinberg equilibrium and for comparisons with previous published data for Caucasian population when available. AA: Amino acid; mAF: minor allele frequency; MAF: major allele frequency; HW: Hardy Weinberg; p value comp: Chi square p value when compared with previous data; Int: Intronic; NS: Non synonymous; S: Synonymous; NT3': Non translated 3'.

For the SNPs rs2273238 and rs9344950 no significant allele frequency differences were encountered in our sample. In the case of rs9353687 a significant difference was found between cases and controls (Chi square: 6.48, p value: 0.039). However, after correcting for multiple testing the result was not significant at a 0.05 cutoff (q value: 0.182). Allele counts and genotype frequencies are shown in Table 45.

Increasing the number of genotyped samples for rs9353687 to 2740 including Oxford, Nottingham and CONCOR cases, and ECACC controls (Probands: 1016, controls: 1694), completely abolished the association (Chi square: 1.13, p value: 0.56). For genotype counts and allele frequencies see Table 46.

	Probands			Controls				
	Hom WT	Het	Hom N WT	Hom WT	Het	Hom N WT	Chi square	p value
rs2273238	1245	99	1	423	29	1	1.15	0.573
rs9344950	916	363	35	337	105	13	3.59	0.16
rs9353687	914	373	36	337	99	13	6.48	0.039

Table 45. Common ANKRD6 SNPs genotyped at a control population. Hom WT: Homozygotes wild type; Het: Heterozygotes; Hom N WT: Homozygotes non wild type.

	C/C	C/T	T/T	MAF	mAF
Probands	740	276	30	0.8394	0.1606
Controls	1171	478	45	0.8323	0.1677

Table 46. Genotype counts and allele frequencies for rs9353687. CHANGE, Oxford, Nottingham and CONCOR cases included. HTO and ECACC controls included. MAF: Major allele frequency, mAF: minor allele frequency.

Rs9353687 is a single base change (C>T) located at exon 11 of ANKRD6. The SNP is synonymous and is located at a residue coding for cysteine (Residue number 506) in the conserved ankirin repeat domain. In silico analysis of the SNP sequence showed that it probably does not produce splicing changes.

4.2. WNT11

4.2.1. Previously unreported uncommon variants

Sequences of 4 exons, including the 5' UTR region, were obtained for WNT11 in 93 CHANGE probands. Eight hundred and eighty five bp were screened using methods described in section 2.2. Two previously undescribed uncommon variants were discovered. See Table 47

Variant	Alleles	Function
207+ 47 G>T	G/T	INT
720 +6 C>T	C/T	INT

Table 47. WNT11 Previously undescribed uncommon variants. Both variants were intronic and located in the flanking regions included in the mutational screening.

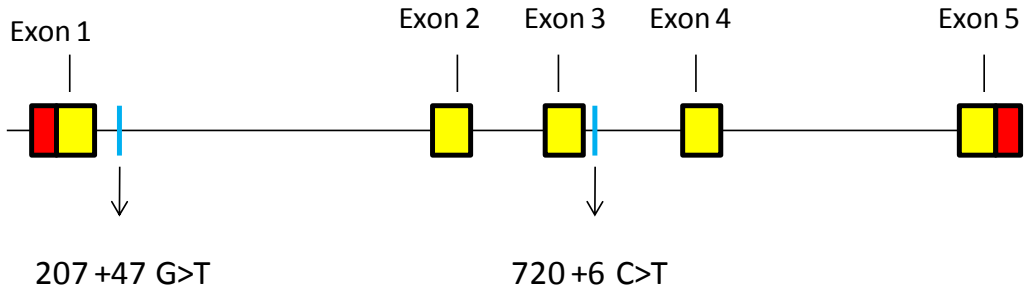

Figure 37. Schematic view of the location of the novel uncommon variants in Wnt11. Transcribed exons are depicted in yellow. Red boxes represent untranslated regions. Both variants are marked in blue.

4.2.1.1. WNT 11 207 +47 G>T

The first uncommon variant found was a single base substitution (G/T) at WNT11 intron 1, position 207 +47. The variant was found in a single CHANGE proband in heterozygote state at the initial mutational screening. See Figure 38 for the ABI sequencer traces from the carrier and a proband not carrying the variant.

This single base change is not in the consensus splice sequence, however computational analysis of the variant with spliceview suggested that the variant decreases the strength of the donor site at the end of exon 1 and produces disappearance of two splicing modulator binding sites located at its position: SC35 and SF2. 207 +47 G>T is located at an Intron within the WNT11 superfamily conserved domain of the protein, which increases the probability of functional alterations if this mutation causes splicing defects.

Sequenom genotyping of parents of the proband carrying the variant showed that it was paternally inherited (For methods see section 2.3.2). The father carrying the variant does not show an abnormal cardiac phenotype. There was no history of CCM in the family carrying the variant.

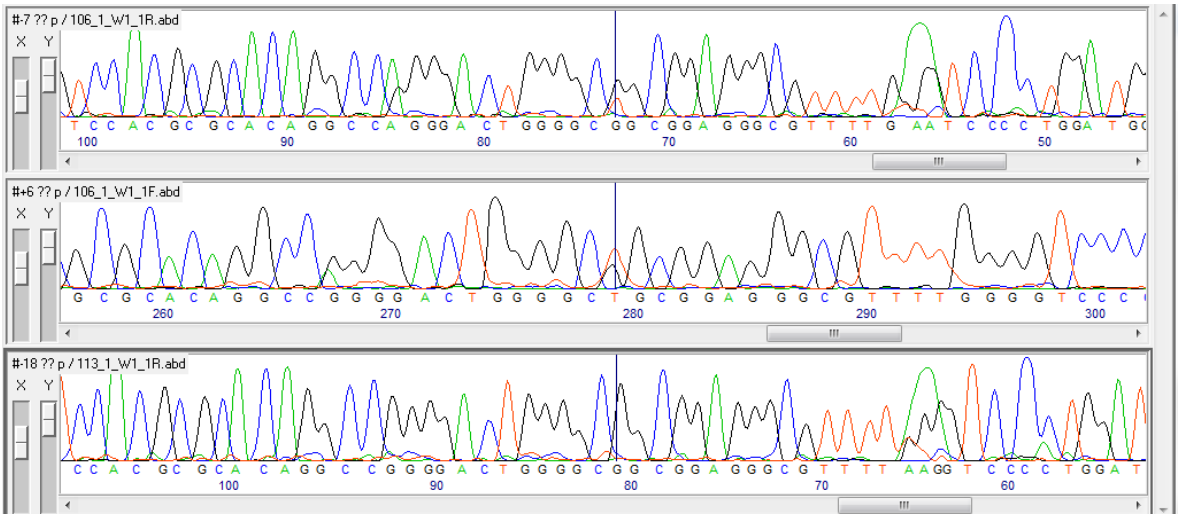

Figure 38. ABI sequencer traces for WNT11 207 +47 G>T. Upper and medium panels show the variant and the bottom one a homozygous wild type subject.

4.2.1.2. WNT 11 207 +47 G>T is not present in controls

A total of 1674 control genotypes from HTO and ECACC samples and 931 TOF proband genotypes (CHANGE, CONCOR, Oxford and Nottingham) were obtained for WNT 11 207 +47 G>T using a Mass extend Sequenom experiment (For methods see section 2.3.2), to check for the presence of the variant in controls. Genotyping was successful for 98% of the samples. The estimated genotype miscall rate was 1% for the SNP.

Nine TOF probands among CHANGE, Oxford, Nottingham and CONCOR populations were found to be carrying the 207 +47 G>T variant, none of the HTO or ECACC controls. The frequency difference between cases and controls reached significant levels (Fisher p value: 0.00009)

4.2.1.3. WNT 11 207 +47 G>T does not cause splicing changes

Despite the intronic location of the 207 +47 G>T mutation, the positive results from in silico splicing analysis and the significant difference in allele frequency between cases and the control population prompted me to carry on further functional investigation of this variant.

A minigene including WNT11 exon 1 and Intron 1 was constructed following the methods described in section 2.5.1. The minigene was expressed in HEK 293 cells following the methods

described in section 2.5.2. Qualitative analysis on agarose gels of RT-PCR products, from total mRNA, using specific primers for the minigene showed no difference in splicing (transcript length and number) when mutant allele constructs were compared to wild type. See Figure 39. The bands correspond to the mature mRNA transcribed from the Pxj41+ WNT11 (Exon 1+Intron 1 207 +47 G>T) wild type and Pxj41+Wnt11 (Exon 1+Intron 1) carrying the variant. See Figure 21 for a graphical explanation of the minigene experiment.

Sequencing of RT-PCR product for both minigenes, wild type and WNT11 207 +47 G>T variant showed complete identity, with splicing taking place at the same donor and acceptor sites.

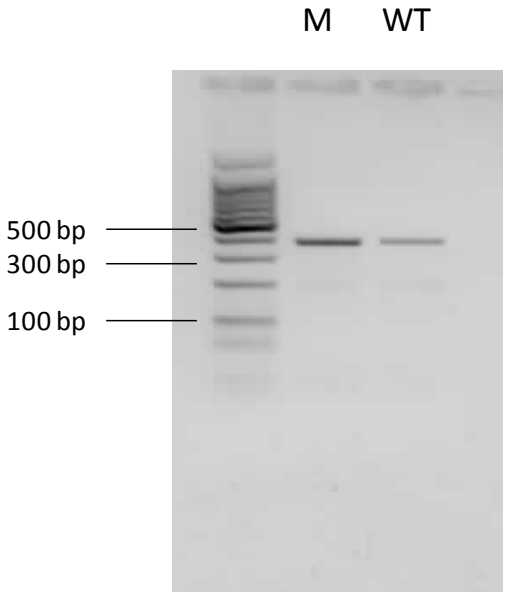

Figure 39. Agarose gel showing the RT-PCR product for minigenes carrying (M=mutant) and not carrying (WT=wild type) the WNT11 207 +47 G>T variant. Both lines show the expected length of mRNA including rabbit beta globin exons 2 and 3, and human Wnt11 exon 1 (467bp)

4.2.1.4. WNT 11 720 +6 C>T

The second uncommon variant encountered was a single base substitution (C/T) at WNT11 intron 3, position 720 +6. The variant was found in a single CHANGE proband in homozygote state at the initial mutational screening. See Figure 40 for the ABI sequencer traces from the carrier and a proband not carrying the variant.

This single base variant is not located at known consensus splice sequences. In silico analysis of the variant using ESEfinder and SpliceView showed that the variant does not decrease the strength of acceptor and donor splicing sites and is not located at known modulator sequence motifs. WNT 11 720 +6 C>T is not located at known conserved domains within the protein.

Genotyping of parents of the proband carrying the variant was performed using a Sequenom experiment (For methods see section 2.3.2) Despite the proband carrying the variant being a homozygote, only the mother was detected to be carrying the variant, with the father showing an homozygous wild type genotype. Due to the results obtained on the control genotyping, which showed the variant to be present in the control population, no further investigation was made for this SNP. The finding was interpreted as a possible case of non-paternity although a deletion in the father could not be ruled out. The proband carrying the variant did not show other associated conditions and there was no family history of CCM or other inherited conditions.

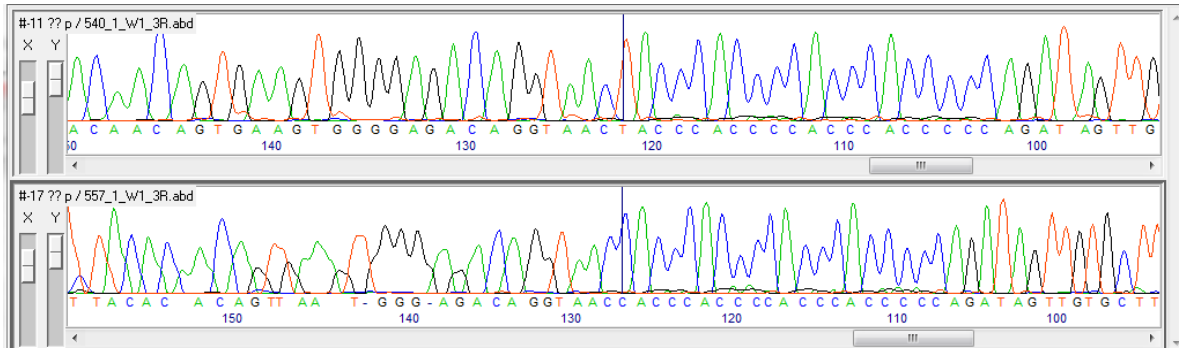

Figure 40. ABI sequencer traces for WNT11 720 +6 C>T. Upper panel show the variant (homozygote) and the bottom one a homozygous wild type subject.

4.2.1.5. WNT 11 720 +6 C>T is present in controls

A total of 1339 control genotypes from HTO samples and 458 CHANGE TOF proband genotypes were obtained for WNT 11 720 +6 C>T using a Mass extend Sequenom experiment (For methods see section 2.3.2).

Genotyping was successful for more than 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Hardy-Weinberg analysis at the total population and each of the subgroups showed that 720 +6 C>T was in equilibrium at the 5% significance level for both the total and the subgroups.

Twelve heterozygotes were found in the population. Strikingly, the only homozygote encountered was the initial proband detected to be carrying the variant. I did not find a significant difference in genotype or allele frequencies between cases and controls for WNT 11 720 +6 C>T (Chi square: 0.392; p value: 0.531). Although the intention of control genotyping was to test for the presence of the variant in controls power was calculated as 25% for a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.05. Approximately 2200 TOF cases would be needed to obtain 80% power. Allele counts and genotype frequencies are shown in Table 48.

	C/C	C/T	T/T	MAF	mAF
Probands	454	4	1	0.9935	0.0065
Controls	1331	8	0	0.9970	0.0030

Table 48. Genotype counts and allele frequencies for WNT 11 720 +6 C>T. MAF: Major allele frequency, mAF: minor allele frequency.

4.3. DISCUSSION

4.3.1. Summary of ANKRD6 variants

Three previously undescribed uncommon variants were found at ANKRD6 in the screening of 93 CHANGE TOF probands. The first uncommon variant found, 1-304 G>A, was a single base substitution (A/G) at ANKRD6 5' UTR region. It was impossible to determine if the variant was de novo. The variant was not located at known cis regulatory motifs. There was no allele frequency differences between TOF cases and controls for ANKRD6 1-304 G>A.

The second uncommon variant found at ANKRD6 was a single base substitution (C/T) at ANKRD6 exon 7, producing a change from Leucine to Phenylalanine at codon number 192. It was impossible to determine if the variant was de novo. Phe192Leu is located at the Ankirin repeats domain of the protein. In silico analysis of the variant did not indicate that it is structurally damaging. There was no allele frequency difference between TOF cases and controls at Leu192Phe.

The third uncommon variant found was a single base substitution (C>A) at ANKRD6 exon 7, producing a change from Threonine to Asparagine at codon number 176. Thr176Asp in silico analysis showed that the variant might be structurally damaging. The variant was maternally inherited. I did not detect allele or genotype frequencies differences between TOF cases and controls. Since all these three variants could be demonstrated to be present in controls, they were given a low priority for functional investigation.

One previously undescribed common variant was encountered at ANKRD6. The variant is a single base substitution (C/T) at ANKRD6 exon 16 producing a change from phenylalanine to leucine at residue 636. Phe636Leu in silico analysis showed that the variant might be structurally damaging and probably does not cause splicing changes. Although a nominally significant difference was demonstrated in genotype frequencies between cases and controls for ANKRD6 Phe636Leu (Chi

square: 8.51, p value: 0.014) q value correction for multiple tests deemed the association not significant and increasing the sample to 1686 probands and 1041 controls did not confirm the association.

Three previously described common variants (rs2273238, rs9344950 and rs9353687) were genotyped in the total of available CHANGE TOF samples and HTO control samples. For the SNPs rs2273238 and rs9344950 no significant allele frequency differences between cases and controls were encountered. Significant genotype frequencies differences were encountered for rs9353687 (Chi square: 6.48, p value: 0.039). However, after correcting for multiple testing the result was not significant at a 0.05 cutoff (q value: 0.182). Moreover, increasing the number of genotypes to 1694 controls and 1046 probands did not confirm the association (Chi square: 1.13, p value: 0.56).

4.3.2. ANKRD6 results in context of previous studies

To our knowledge this research work is the first to explore the relationship between human genetic variation at ANKRD6 and TOF. Several pieces of previous work support the hypothesis that protein function variation at ANKRD6 might be related to cardiac malformations on humans.

Previous studies indicate that loss or decrease of function of ANKRD6 could cause defects in cardiogenesis due to its known involvement at the PCP pathway. ANKRD6 is a human ortholog for the protein Diversin in vertebrates (136). Diversin is in turn a protein related to Diego in *Drosophila*, which has a recognized role in the non canonical *Wnt* signaling (116, 137). Diego, like frizzled, is a non canonical Wnt agonist preventing Prickle from binding to dishvelled (193).

No ANKRD6 knockout mouse has yet been described in the literature. However, most of the work and animal models on Diversin and Diego supports the idea that ANKRD6 suppression is likely to produce a cardiac phenotype. In 2002 the group of Moeller et al. reported that injection of ANKRD6 mRNA lacking the ankirin repeats (Div-ΔANK) into early zebrafish embryos produces cardia bifida (116). Apparently the phenotype was caused by a dominant negative effect of this form of diversin. The phenotype was not reproduced by injecting morpholinos against diversin. It is also interesting that the cardiac phenotype in this model is rescued by the coinjection of an activated form of RhoA, indicating a downstream function of this small GTPase protein. The same group reported that injection of full size forms of mouse diversin or the ankirin repeats rescued the phenotype (cardia bifida) induced by Wnt5A-Wnt11 morpholinos, inferring that the action of the protein is located downstream in the pathway from the Wnt secreted proteins.

4.3.3. Summary of Wnt11 variants

Two previously undescribed uncommon variants were found in the mutational screening. The first uncommon variant found was a single base substitution (G/T) at WNT11 intron 1, position 207 +47. The variant is located at the Wnt11 conserved domain. In silico analysis of the variant showed that it can cause decrease of strength of the acceptor site at the end of the exon and also disappearance of SC35 and SF2 binding sites. 207 +47 G>T was paternally inherited at the proband found to be carrying it. The variant was a found in nine probands and none of the controls. A minigene experiment showed no difference in splicing (transcript length and number) between the variant and wild type alleles.

The second uncommon variant encountered was a single base substitution, on homozygous state, (C/T) at WNT11 intron 3, position 720 +6. WNT 11 720 +6 C>T is not located at known conserved domains within the protein and does not cause splicing changes at in silico analysis. I did not detect allele or genotype frequencies differences between TOF cases and controls for WNT 11 720 +6 C>T.

4.3.4. Wnt11 in context of previous work

To our knowledge this work is the first to explore the relationship between genetic variation at WNT11 and TOF. Several pieces of previous work support the hypothesis that protein function variation at WNT11 might be related to cardiac malformations in humans.

Wnt 11 and its homologues have been shown to be consistently expressed in the heart forming fields in quail chicken mouse and xenopus (85, 194), and in the mature heart in humans. Its expression in the primary heart tube has been also demonstrated in mouse, chicken and *xenopus* embryos. Wnt11 has been shown to induce cardiogenesis in Xenopus and Zebrafish (103, 107). In quail embryo mesodermal explants, Wnt11 expression or addition produces cardiomyocyte differentiation (104).

In an exquisite set of experiments Pandur et al. in 2002 showed that the addition of dominant negative Wnt11 constructs to *Xenopus* embryos decreased the expression of cardiac markers like Nkx2.5 (107). Injection of Wnt11 inhibiting morpholinos exerted the same effect.

4.3.5. A functional explanation for 207 +47 G>T

WNT 11 207 +47 G>T is a synonymous variant detected in 9 probands. Taking into account the proportion of strongly and mildly deleterious missense uncommon variants we might also think that this is probably a medically deleterious variant. According to my findings it is likely that the variant does not affect splicing. If the variant is truly pathogenic, it is unlikely that is a marker of a second linked pathogenic variant due to its low mAF. Wnt11 +47 G>T influence in protein function can be exerted in two ways: the first is a splicing effect not detected in the minigene experiment. A minigene is a simplification of a complex network of factors interacting to define exon and intron boundaries. This type of experiment does not take into account the genomic environment of the exon, the usual gene promoter and regulating factors located farther than a couple hundred of base pairs, the splicing change could be not detected with a minigene.

4.4. Conclusions

The major finding of this part of my research was the discovery of an intronic uncommon variant in 9 TOF probands that was not found in controls. Despite being an intronic variant, functional analysis for WNT11 207 +47 G>T failed to show a splicing effect of the variant. Also a common variant (Phe636Leu) in ANKRD6 was initially found to be associated with TOF but p values were not significant after multiple test correction. Increasing the sample ruled out the association.

Despite that theoretical research have shown a high probability of uncommon missense variants in ANKRD6 and Wnt11 of being medically detrimental and previous animals models have shown the implications of both genes in cardiogenesis, my research have found several genetic variants without finding significant associations with TOF. However, the results are not conclusively negative. For most of the uncommon variants the sample needed for detecting a difference between cases and controls, if the risk associated is low, would be difficult to collect. Increasing the sample for variants like WNT11 720 +6 C>T, ANKRD6 1-304 G>A, Leu192Phe or Thr176Asp can confirm differences between cases and controls.

An alternative approach to the functional implication of WNT11 207 +47G>T like quantitative RT PCR or reverse transcriptase amplification of total RNA from a carrier of the variant can give definite answers to the question of whether or not this variant influence protein expression and splicing. Due to the important role of these two genes (ANKRD6 and Wnt11) in cardiogenesis, and the finding of uncommon and common novel variants, potentially associated to TOF, it is worth to undertake bigger sequencing experiments in bigger samples, not only to describe the remaining uncommon genetic variation yet to be discovered but also to confirm or rule out the associations between the already discovered variants.

5. ROCK 2, DVL1L1, DVL2, AND DVL3

5.1. DVL1L1

Sequence of the only DVL1L1 exon was obtained in 93 CHANGE probands. A total of 350 bp were screened using methods described in section 2.2. I did not find any previously unreported variants either common or uncommon. Two previously reported SNPs were found at the exon: rs807429, a SNP located at the 3' end of the pseudogene; and rs812088, an exonic apparently synonymous SNP. Both SNPs were in Hardy-Weinberg equilibrium at the 5% significance level. The SNPs were not typed in control populations because there were not significant differences with previously reported allele frequencies in Caucasian populations. For allele counts and frequencies see Table 49.

	HWT	Het	HNWT	MAF	mAF
rs807429	69	20	1	0.8700	0.1300
rs812088	80	11	0	0.8800	0.1200

Table 49. Genotype counts and allele frequencies for rs807429 and rs812088. mAF: Minor allele frequencies; MAF major allele frequencies

5.2. ROCK2

5.2.1. Previously undescribed common variants

Sequences of 32 exons, not including the UTR, were obtained for ROCK2 in 93 CHANGE probands. A total of 4283 bp were screened using methods described in section 2.2. I did not find any novel rare variants among the 93 TOF probands screened.

5.2.1.1. ROCK2 145 -25 A>T

The only previously undescribed common variant found was a single base substitution (A>T) in ROCK2 intron 1. The variant was found in eight CHANGE probands in heterozygote state on the fist mutational screening.

Intron 1 is not a conserved region. The variant is not located at consensus donor or acceptor splicing sites. In silico analysis of the variant using ESE finder showed that the variant does not alter the strength of theoretical splicing regulator sites around the region. Analysis using spliceview showed no differences in the strength of donor and acceptor splicing sequences between wild type and mutant alleles.

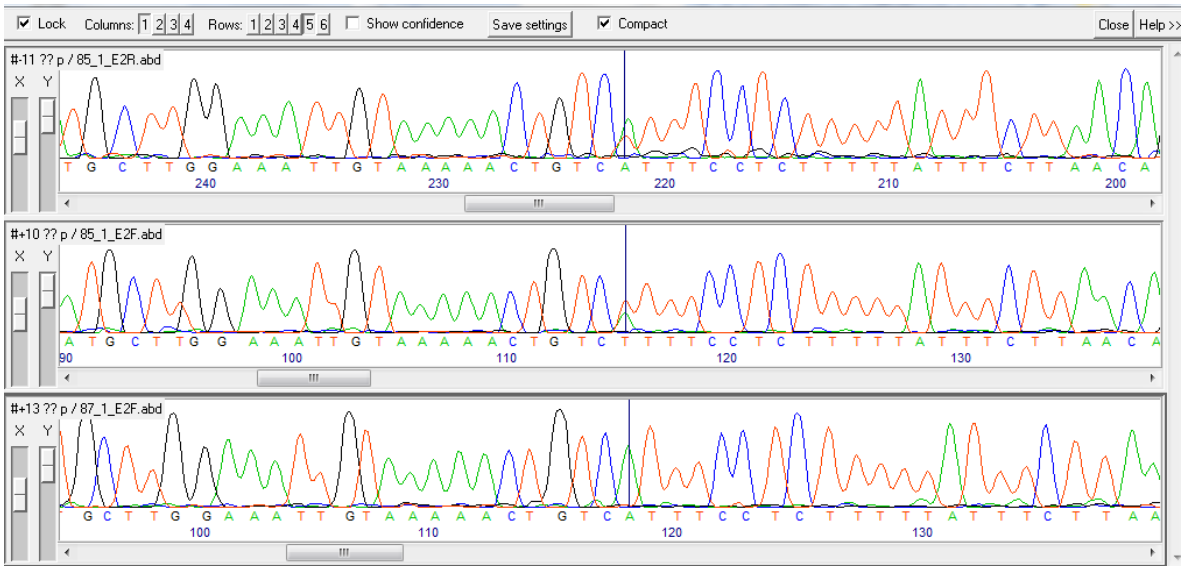

Figure 41. ABI sequencer trace displays for ROCK2 145 -25 A>T variant. Upper and medium panel show heterozygote sequences, bottom shows a normal trace.

A total of 1772 genotypes were obtained from cases and controls using the Sequenom platform and the methods described at section 2.3.2. . Due to its allele frequency and given the number of cases and controls, power was calculated as 100% for a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.1. Genotyping was successful for more than 98% of the samples used. The estimated genotype miscall rate was less than 1% for the SNP. 145 -25 A>T was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

An almost equal percentage of heterozygotes were found in both populations: probands and controls. I did not find a significant difference in genotype or allele frequencies between cases and controls for 145 -25 A>T (X^2= 0.22; p= 0.89). Allele counts and genotype frequencies are shown in Table 50.

	A/A	A/T	T/T	MAF	mAF
Probands	410	45	1	0.9485	0.0515
Controls	1193	120	3	0.9521	0.0479

Table 50. Genotype counts and allele frequencies for ROCK2 145 -25 A>T. MAF: Major allele frequency, mAF: minor allele frequency.

5.2.2. Previously described common variants

Four previously described SNPs were encountered at ROCK2 mutational screening on 93 TOF CHANGE probands. All but one of them were in Hardy Weinberg equilibrium and showed no significant differences when compared to previously Hapmap data on Caucasian population (CEU population). ROCK2 SNPs found at the initial mutational screening, their allele frequencies, measurements for fitness to the Hardy Weinberg equilibrium and comparisons for allele frequencies with the published CEU data see Table 51.

SNP	Function	AA	mAF	MAF	HW p value	P value comp
rs41264193	S	114(S)	0.0417	0.9583	0.9237	NI
rs9808232	NS	431(T/N)	0.4000	0.6000	0.8186	0.7093
rs41264191	Int		0.0444	0.9556	0.9072	NI
rs2271621	Int		0.3708	0.6292	0.5969	0.0000

Table 51. Previously described SNPs on ROCK2, allele frequencies and p values for Hardy Weinberg equilibrium and for comparisons with previous published data with Caucasian population when available. AA: Amino acid; mAF: minor allele frequency; MAF: major allele frequency; HW: Hardy Weinberg; p value comp: Chi square p value when compared with previous data; Int: Intronic; NS: Non synonymous; S: Synonymous; NT3': Non translated 3'.

One common previously described SNP (rs2271621) at ROCK2 was genotyped in the control population and the total of available CHANGE samples due to its significant allele frequency difference with previously published data (p value: 0.00005).

One thousand six hundred and fifty four genotypes were obtained from CHANGE TOF probands and HTO controls for rs2271621, using a mass Extend experiment (methods described at section 2.3.2). Power was calculated in 100% for a case control allele frequency comparison for the SNP to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.01. Genotyping was successful for more than 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Rs2271621 was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations. However, I did not find significant differences in genotype or allele frequencies between cases and controls (X^2 square: 0.024; p value: 0.988).

	A/A	A/T	T/T	MAF	mAF
Probands	241	188	29	0.7314	0.2686
Controls	630	488	78	0.7308	0.2692

Table 52. Genotype counts and allele frequencies for rs2271621. MAF: Major allele frequency, mAF: minor allele frequency

5.3. DVL3

5.3.1. Previously undescribed uncommon variants

Sequences for the 15 DVL3 exons, including the 5' UTR region, were obtained in 93 CHANGE probands at the initial mutational screening. Two thousand two hundred and seventy five bp were screened in each proband using methods described in section 2.2. Two previously undescribed uncommon variants were encountered at DVL3. See Table 53.

Exon	Variant	Alleles	Function	AA
3	446 T>A	T/A	S	P
11	1339 +17 A>G	G/A	INT	

Table 53. Not previously described uncommon variants encountered at DVL3. AA: Aminoacid; S: Synonymous; INT: Intronic; P: Phenylalanine.

5.3.1.1. DVL3 446 T>A

The first uncommon variant found was a single base substitution (T>A) located at DVL3 exon 3, mRNA position 446. The SNP is located within a codon that translates a phenylalanine and does not produce changes at the protein sequence level. The variant was found in a single CHANGE proband in heterozygous state at the initial mutational screening. See Figure 41 for the ABI sequencer traces from the carrier and a proband not carrying the variant.

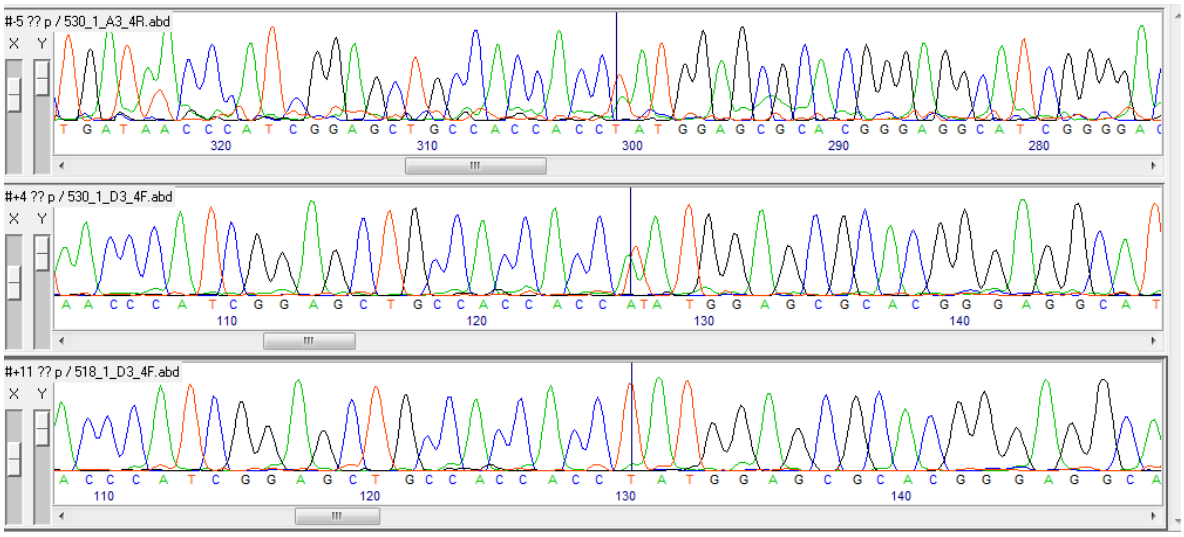

Figure 42. ABI sequencer trace displays for DVL3 446 T>A variant. Upper and medium panel show heterozygote sequences, bottom shows a normal trace.

The variant is not located at conserved domains of the protein, and does not disrupt known splicing donor or acceptor sites. In silico analysis of the variant showed that the variant does not change the relative strength of splicing donor and acceptor sites and also does not influence any theoretical binding sites for known splicing regulators.

The variant was inherited, in the first proband detected to be carrying it, from the father who does not show an abnormal cardiac phenotype. The carrier did not show evidence of other inherited conditions. There is no history of CCM within the family.

5.3.1.1. DVL3 446 T>A is present in controls

A total of 1538 genotypes were obtained from cases and controls through a Sequenom assay following methods described at 2.3.2. According to the number of probands and controls power was calculated in 81% for a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.1. Genotyping was successful for >98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. 446 T>A was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Forty four heterozygotes were found in total at both populations: probands and controls. No homozygotes for the variant were found. Minor allele frequency for the entire population was 1.4%. I did not find significant differences at genotype or allele frequencies level between cases and controls for DVL3 446 T>A (X^2= 0.128; p= 0.864). For allele counts and genotype frequencies see Table 54.

	A/A	A/T	T/T	MAF	mAF
Probands	410	11	0	0.9869	0.0131
Controls	1084	33	0	0.9852	0.0148

Table 54. Genotype counts and allele frequencies for DVL3 446 T>A. MAF: Major allele frequency, mAF: minor allele frequency.

5.3.1.3. 1339 +17 A>G

The second uncommon variant found was a single base substitution (A>G) located at DVL3 intron 11, seventeen bp from the 3' end of the exon. The SNP is not located at consensus splicing donor or acceptor sites and the intron does not belong to known conserved domains. The variant was found on a single CHANGE proband in heterozygous state at the initial mutational screening. See Figure 43 for the ABI sequencer traces from the carrier and a proband not carrying the variant.

Figure 43. ABI sequencer trace displays for 1339 +17 A>G variant. Upper panel shows a heterozygote for the variant, bottom shows a homozygote wild type.

In silico analysis of the variant showed that the variant does not change the relative strength of splicing donor and acceptor sites and is not located within any theoretical binding site for known splicing regulators.

In the first proband carrying the variant was impossible to determine if the variant was inherited. The mother of the carrier does not carry the variant and we do not have sample from the father. In the controls, the three heterozygotes are members of the same family and the original carrier is the father.

5.3.1.4. 1339 +17 A>G is present in controls

One thousand seven hundred and seventy five genotypes were obtained for DVL3 1339 +17 A>G including 455 TOF probands and 1320 controls. Genotypes were obtained using a mass Extend assay on the Sequenom platform following methods described in section 2.3.2. Power calculation according to the number of cases and controls included, showed a power of 22% to detect a difference between cases and control in a general case and control analysis hypothesizing a relative risk of 2 for heterozygotes and taking into account an alpha error of 0.1. Nevertheless, as mentioned previously, control genotyping was performed to check the presence of the uncommon variants in the control population, as to design a case and control study on variants as uncommon as 1339 +17 A>G would require an unfeasible number of cases: approximately 3663 for 80% power.

Five heterozygotes were found among both, cases and controls: two among the probands and three among the controls. No homozygotes for the variant were found. Minor allele frequency for the entire population was 0.14%. No statistically significant difference was found between cases and controls in allele frequency at 1339 +17 A>G (Fisher exact test p value: 0.608). For allele counts and genotype frequencies see Table 55.

	A/A	A/G	G/G	MAF	mAF
Probands	453	2	0	0.9978	0.0022
Controls	1317	3	0	0.9989	0.0011

Table 55. Genotype counts and allele frequencies for DVL3 1339 +17 A>G. MAF: Major allele frequency, mAF: minor allele frequency.

5.4. DVL2

Mutational screening for two exons (11 and 15) was performed for this gene during the experimental work of this thesis. The remaining 13 were screened during previous work at our laboratory. A total of 1102 bp were screened. I did not find new or previously described common variants during the mutational screening.

5.4.1. Control genotyping for not previously reported uncommon variants

5.4.1.1. DVL2 Thr536Ile

A non synonymous (Thr536Ile) maternally inherited single nucleotide variant (C/T) was found in heterozygote state in one TOF proband in a previous mutational screening performed for Dvl2 at our laboratory. Arg163Trp was genotyped in 1244 controls and 142 CHANGE TOF probands using an RFLP following methods detailed in section 2.3.3. The variant is not located at conserved domains within the protein. In silico analysis of the variant showed that it is unlikely to be structurally damaging and it does not alter cis splicing regulatory elements.

Power calculation according to the number of genotypes included showed a power of 80% to detect a difference in a general case and control analysis hypothesizing a relative risk of 2 for heterozygotes and taking into account an alpha error of 0.1. Genotyping was successful for >98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Thr536Ile was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Seventy three heterozygotes were found in total, cases and controls: sixty three among the probands and ten among controls. One homozygote was found in each one of the populations. The variant was found in 3% of the alleles of the global population. An association was found between the variant and TOF (p value: 0.0006). Correction with Qvalue for multiple testing, using a

5% false discovery rate level, yielded the association not significant. For allele counts and genotype frequencies see Table 56.

	C/C	C/T	T/T	MAF	mAF
Probands	131	10	1	0.9845	0.0155
Controls	1180	63	1	0.9824	0.0176

Table 56. Genotype counts and allele frequencies for DVL2 Thr536Ile. MAF: Major allele frequency, mAF: minor allele frequency.

5.4.2. Control genotyping for previously reported common variants

5.4.2.1. rs2074216

A synonymous SNP located at codon 135 of Dvl2 protein was decided to be genotyped on a control population due to its differences in allele frequencies at a previous mutational screening when compared to previously published data on Caucasian populations and allele frequency failure at Hardy Weinberg equilibrium tests. The SNP is located in exon 3 of the protein and does not belong to known conserved domains within the protein. The variant does not disrupt consensus splicing sequences and in silico analysis does not show disruption of cis splicing regulatory sequences with the variant.

Rs2074216 was genotyped in 1236 HTO controls and 175 CHANGE TOF probands using a Taqman platform experiment and following methods detailed in section 2.3.4. Minor allele frequency in the total population was 0.37. Power calculation according to the number of cases and controls included showed a power of 100% to detect a difference in a general case and control analysis hypothesizing a relative risk of 2 for heterozygotes and taking into account an alpha error of 0.05. Genotyping was successful for >98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. Rs2074216 was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations. There was no difference between TOF cases and controls at the allele frequencies level (p value: 0.14). For genotype and allele frequencies at TOF probands and controls see Table 57.

	C/C	C/T	T/T	MAF	mAF
Probands	71	75	29	0.6200	0.3800
Controls	490	589	157	0.6347	0.3653

Table 57. Genotype counts and allele frequencies for rs207416. MAF: Major allele frequency, mAF: minor allele frequency.

5.5. DISCUSSION

5.5.1. Summary of ROCK2 findings

I did not find uncommon variants in ROCK2. One previously undescribed common variant was found in the mutational screening. ROCK2 145 -25 A>T is an intronic single nucleotide variant located at intron 1 of the protein. In silico analysis of the variant showed that does not cause splicing changes. ROCK2 145 -25 A>T is not located at conserved domains within the protein. The minor allele frequency of the variant was 0.048 and there was no allele frequency difference between probands and controls.

One previously described common variant in ROCK2 (rs271621) was genotyped in a control population due to allele frequency differences with previously published data in the first mutational screening. Rs2271621 is an intronic variant not located at consensus splicing sequences and not previously described as of clinical relevance. Minor allele frequency at the total population is 0.26 without significant differences between probands and controls. The initial differences encountered between CHA cases and previous data were probably due to sample size.

5.5.2. ROCK2 findings in context of previous studies

Previous studies indicate that Rho kinases in general and ROCK2 function in particular are crucial for cardiogenesis. To our knowledge this is the first study to examine common and uncommon human genetic variation at ROCK2 and its relationship with TOF.

ROCK2 or ROCKα is one of the Rho GTPases effectors, with multiple functions at development and that exerts its function through phosphorylation of multiple proteins like myosin light chain kinase, myosin light chain phosphatase, LIM kinases and Adducin (195) Previous animal and molecular experiments have consistently linked ROCK2 with PCP and cardiogenesis (127, 130). The results of those studies indicate that a function disarrangement of the protein is highly likely to produce cardiac malformations. ROCK2 expression has been reported in murine and chick embryos at stage E7.5-9.0 (126). Zhao et al. also reported expression of the protein in murine myocardium and trabeculae in tissue sections at E9.5-E12 (121).

Zhao et al. showed that inhibition of Rho kinases using Y27632 (an unspecific Rho associated kinases inhibitor) in cultured mouse embryos prevented heart looping, septation and atria formation. Embryos treated with Y27632 showed at E9.5 a bulbous like primitive structure with a

conotruncus like structure anterior and posterior to it. Primitive atria and AV canal did not form in these embryos (121). Also when Y27632 was added to embryos at later stages of development cardiogenesis seemed to be further arrested showing smaller hearts and dilated pericardium. Examination of tissue sections of these hearts showed a condensed myocardium mass without development of the trabeculae and failure to develop the compact myocardium of the ventricular wall. Part of the development failure was apparently due to inhibition of cardiomyocyte proliferation as measured on the inhibition treated embryos.

Wei et al. also researched on the effect of Rho kinases inhibition using Y27632 in chick embryos. Their experiments showed that inhibition of Rho kinases produced a failure to fuse in the midline with severity that depended on the stage at which the inhibitor was added. The conclusion was that Rho kinases regulated the process controlling cell migration (126). Zhao's group also showed that Rocks are expressed in mouse endocardial cushions. At the same experiment this group tested the effect of Y27632 on monolayer and 3D collagen cultures of endocardial cushion explants(131). Again, inhibition of ROCKs produced decrease in endocardial cell migration, a finding that was partly explained by the inhibition of the transformation of endocardial cells into a mesenchymal phenotype (131). Sakabe et al. reported a similar finding using chick embryos (132).

It is however striking that ROCK2 KO mouse does not show cardiac phenotype. Its phenotype is rather characterized by intrauterine growth retardation and lethality associated with abnormalities and thrombus formation at the labyrinth layer of the placenta. The placental abnormalities occur mainly at the maturation and adaptation stage taking place at E12-E16 and are related to regulation of protein expression, in particular of prolactin like GTPase 1 or PAI I (120). A possible explanation that is given by Thumkeo et al. for the absence of other phenotypes is the functional overlapping of the two ROCK isoforms. However, compensating overexpression of ROCK1 was not detected in the ROCK2 KO mouse.

Despite ROCK2 being a good candidate gene to cause a portion of TOF in human population, and findings in this research linking a portion of TOF inheritance to common and uncommon variation at the other Rho kinase isoform ROCK1, only one novel common variant was discovered in the mutational screening in ROCK2. It is however likely that due to the low allele frequencies of the variants associated with cardiac phenotypes in humans, this mutational screening was underpowered to detect such variants in our population. Screening of 186 chromosomes only give us a power of around 80% to detect variants present in 2.5% of the probands, not taking into account random effects. To definitively rule out the hypothesis of uncommon genetic variation at ROCk2 being linked to TOF, a bigger sample should be sequenced. Given the next generation sequencing platforms now available, the possibility of sequencing not only this gene but a number of them in bigger samples is not only possible but the logical next step.

5.5.3. Summary of Dvl findings

Two previously undescribed uncommon variants were encountered at DVL3. The first uncommon variant found was a synonymous single base substitution located at mRNA position 446, DVL3 446 T>A. The variant is not located at consensus splicing sequences and in silico analysis does not show that it disrupts cis splicing modulators. DVL3 446 T>A is present in controls without significant differences in allele frequency.

The second uncommon variant found was a single base substitution (A>G) located at DVL3 intron 11, 1339 +17 A>G. The SNP is not located at consensus splicing donor or acceptor sites and according to in silico analysis does not disrupt splicing modulators. Allele frequency of the variant did not show statistical significant differences when compared to controls.

One novel uncommon and one novel common variant described in a previous mutational screening on Dvl2, performed in our laboratory, were genotyped in a control cohort. The uncommon variant is a non synonymous maternally inherited SNP, Thr536Ile. The variant is not located at consensus splicing sequences. In silico analysis of the variant showed that is unlikely to be structurally damaging. Thr536Ile was found in controls at an apparently significantly different frequency when compared to probands. Nevertheless when corrected for multiples tests the difference was not significant.

Also a Dvl2 synonymous SNP, rs2074216, was genotyped in a control population. The SNP is located at exon 3 of the protein and does not belong to known conserved domains. The variant does not disrupt consensus splicing sequences and in silico analysis does not show disruption of cis splicing regulatory sequences with the variant. I did not find significant differences between cases and controls at the allele frequency of this SNP.

5.5.4. Dvl findings in context of previous studies

This is the first study to explore genetic variation in Dvls and its relationship with TOF. In the light of previous findings, it would be expected that a loss or decrease of function at the Dvl proteins causes developmental abnormalities, in particular in cardiogenesis. It is also expected that the genetic influence of a variant in a single gene of these group of paralogs will produce an identifiable phenotype independent of the functional overlapping between members of the protein

family. This fact is illustrated by findings of Hamblet et al., a group that described the phenotype of Dvl1 and Dvl2 KO mouse showing that penetrance of cardiac defects was similar between single and double knockout mice (113).

Hamblet et al. published a paper in 2002 describing the phenotype of Dvl1-2 KO mouse. Dvl2 mouse showed 50% perinatal lethality. The surviving KO mice were almost invariably females and developed normally to adulthood. Lethality was explained by a 50% penetrance of cardiac defects, with the majority exhibiting DORV and VSD, 12.5% TGA and 25% PTA. Cardiac conotruncal septation defects were explained according to the authors by neural crest cell migration defects. Whole mount probe hybridization showed that Dvl2 embryos did not show Pitx expression, a neural crest cell marker, at the 4^{th} branchial arch, neural crest cells or the developing outflow tract (113). Interestingly Dvl1 -/- mice do not show cardiac abnormalities. The Dvl1 -/- phenotype is characterized by poor social interaction and gating defects (148).

Dvl isoforms' involvement in convergent extension movements at embryogenesis and its regulatory role in *wnt* pathways, including canonical, non canonical and calcium controlled, has been largely recognized (141). See Figure 44 for a graphical illustration of Dvl involvement in the Wnt pathways. Itoh showed that explants of *Xenopus* caps not only elongated in a Xdsh (*Xenopus* paralog of Dvl) dose dependent manner but also developed different types of tissue depending on the amount of the mRNA added. (145). In 2004, Ewald and coworkers demonstrated that Dvl inhibition in *Xenopus* embryos cause gastrulation defects(146). Cell studies trying to elucidate this fact have shown that disruption of disheveled genes block the formation of lamellipodia in migrating cells (147).

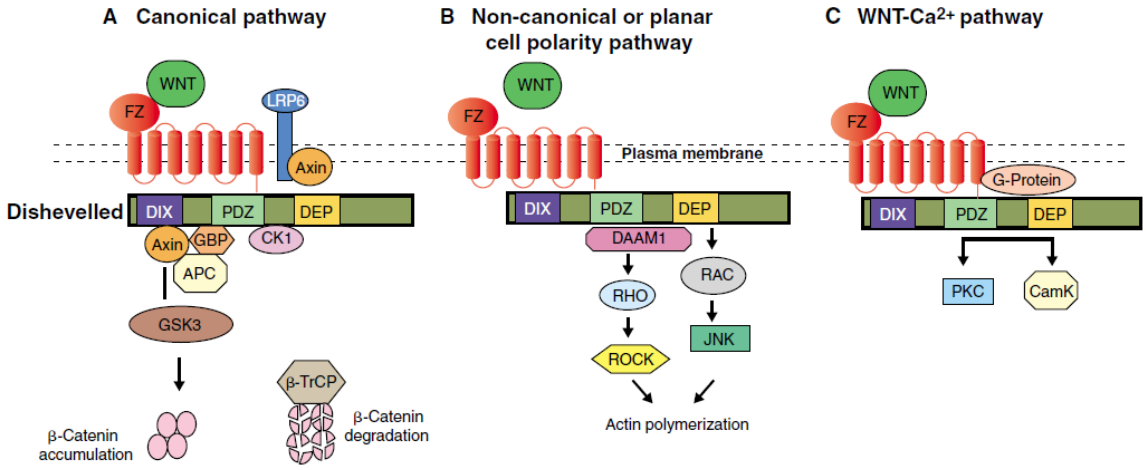

Figure 44. Dishvelled involvement in the Wnt signalling pathways. A. Activation of the canonical pathway depends on the FZ (Frizzled) receptor and the LRP co receptor, activation of Dvl produces disassembly of the complex of Axin, adenomatous polyposis coli (APC) and Glycogen synthase kinase 3 (GSK3) blocking β catenin phosphorylation and therefore decreasing is degradation. B. A signal is transduced by a FZ receptor activating Dvl that in turns activates the small GTPases Rho and Rac that exerts their function trough their respective kinases. C. WNT signalling through the FZ receptors and Dishevelled and G-proteins produces the release of intracellular calcium and activation of Phospholipase C, CamK2 and PKC. Graph reproduced with permission from Wallingford, J.B. and R. Habas, *The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity.* Development, 2005. 132(20): p. 4421-36.

The previous literature evidence suggests that a Dvl loss of function would likely cause incompletely penetrant defects in cardiac outflow tract and may cause other congenital defects consequence of abnormal convergent extension movements. However, only one partially positive result was found. A non synonymous SNP, DVL2 Thr536Ile, apparently a structurally benign variant, showed significant differences in allele frequency between cases and controls. The difference was not significant when corrected for multiple testing. Failing multiple test corrections is a fact that could be derived from the risk associated with the variant and sample size. Thr536Ile was present in 3% of chromosomes and therefore a low OR is expected to be associated with the variant, given the risks associated to variants of similar frequency reported in the literature for other complex diseases. This is explained by the evolutionary pressures that a variant is subjected to depending on their fitness impact. Without random genetic drift or reproductive advantages of heterozygotes, the frequency of a variant with influence on fitness is proportional to the amount of this so called influence. The risk associated with Th536Ile might be as expected modest and perhaps a bigger sample is needed to confirm or rule out the possible implication of uncommon genetic variation at Dvl2 in the pathogenesis of TOF.

However, it is interesting that Thr536Ile was present in a higher frequency in controls, giving a calculated OR of 0.7. In other words the variant seems to be protective and decrease the risk of suffering TOF. The mechanisms explaining the possible protective effect of the variant are complicated. Dvl protein function or expression variation, due to its extensive implications in multiple embryogenetic processes might have profound effects on development. This fact leaves three possible explanations for a protective effect given by a non synonymous variant in a key protein such as Dvl: the protective effect is the result of a gene to gene interaction; the protective effect manifests itself due to gene-environment interactions; or the protective effect is not exerted by Dvl2 but the marker is in LD with the locus-loci responsible for the protection.

5.6. CONCLUSIONS

One novel common variant was found in the mutational screening at ROCK2. 145 -25 A>T is an intronic variant found in 0.048% of chromosomes that seems to be benign according to in silico analysis. Comparison of allele frequency between TOF cases and controls did not show significant differences.

Two uncommon variants were found at DVL3 during the mutational screening: 446 T>A, a synonymous variant, and 1338 +17 A>G, an intronic variant. Both of Dvl3 uncommon variants did not show positive results at in silico analysis and were found in controls at similar frequencies when compared to cases**.**

Two Dvl2 variants, a novel uncommon non synonymous (Thr536Ile) and a previously described common (rs2074216) were genotyped in the HTO controls. There was not significant allele frequency difference between TOF probands and controls for rs2074216. However, for Thr536Ile an apparently benign change an association was found but p values were not significant after multiple test correction. Due to the crucial involvement of Dvl2 in embryogenesis and taking into account that this is the first study to explore human genetic variation at Dvls in relationship with TOF the next logical step is to increase the sample to confirm or rule the possible protective association between Thr536Ile and TOF. The mechanistic and functional implications of the variant are more complicated and should be explored if the association is confirmed

It is likely due to the functional overlapping of Dvls that a single uncommon variant can solely explain a cardiogenesis disarrangement as complex as TOF, excepting in the case of dominant negative forms of the proteins. It is also likely that the truly deleterious variants in ubiquitous

proteins as Dvls are subjected to such a negative evolutionary pressure that they are difficult to reach detectable levels even if they cause any form of TOF.

In the case of ROCK2 it seems to be not associated with TOF. As it was mentioned, previous knockout mouse models have shown no evidence of a cardiac phenotype, and all previous works reporting cardiac malformations associated with the inhibition of ROCKs have unespecifically inhibited both forms of Rho kinases in the experiments.

6. WNT5A

6.1. Previously unreported uncommon variants

Sequences of 5 exons in 93 patients were obtained for WNT5A. Two previously undescribed uncommon variants were encountered at WNT5A. See Table 28.

Exon	Variant	Function	AA	Residue	N
2	Asp119Ser	NS	D/S	212	1
Non translated 3'	1461 +38 C>T	UTR			3

Table 58. Previously undescribed uncommon variants. N: Number of carriers; AA: Aminoacid; UTR: untranslated region; NS: No synonymous; D: Aspartic acid; S: Serine; T: Threonine.

6.1.1. Asp119Ser

The first uncommon variant found was a non synonymous change in exon 2 detected in 1 patient in heterozygous state. The A/G mutation produces a change in residue number 119 from Aspartate to Serine. The residue is located within the Wnt conserved domain. The residue has not been reported as functionally essential. In silico analysis of the mutation indicate that it is not structurally damaging and does not produce splicing changes. Figure 45 shows the ABI sequencer traces for the wild type allele and the variant. Analysis of the residue shows that it is conserved among chick, mouse and human.

It was not possible to determine if the variant was inherited or de novo, as we had no DNA sample from either mother or father. Apparently neither of the parents presented an abnormal cardiac phenotype.

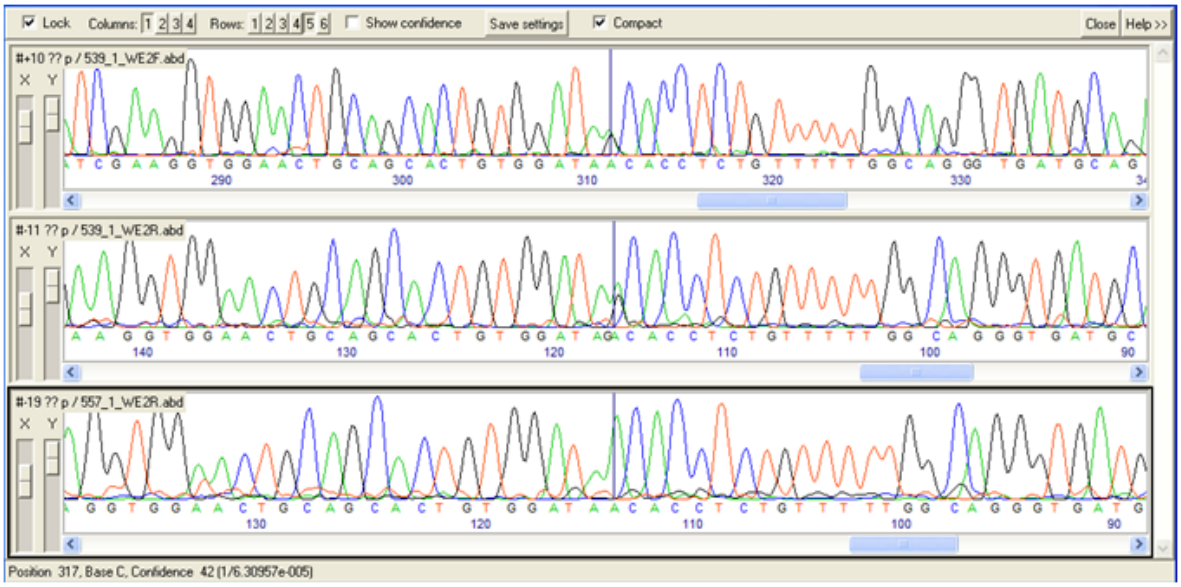

Figure 45. WNT5a mutation 363A>G trace display. Upper and medium panel show patient sequences. Bottom panel shows a normal trace.

6.1.2. Wnt5A Asp119Ser is not present in controls

A total of 1336 control genotypes from HTO samples and 457 CHANGE TOF probands genotypes were obtained for 363 A>G using a Sequenom platform experiment and following the methods specified in section 2.3.2. Allele frequencies for probands and controls are shown in Table 59.

Despite that control genotyping was performed in our sample to check for the presence of the variant, power was calculated as 20% to detect genotype frequencies differences for Asp119Ser between cases and controls, hypothesising a relative risk of 2 for the heterozygotes and alpha error of 0.05. Genotyping was successful for 98% of the samples. The estimated genotype miscall rate was 1% for the SNP.

Only the initial individual identified on the mutational screening was found to be carrying the Wnt5A 363 A>G variant, none of the other cases or the controls was found to be carrying it.

6.1.3. Wnt5A 1461 +38 C>T

The second uncommon variant was found at the 3´ untranslated region of the gene: 1461 +38 C>T. The C/T change was encountered in 3 patients in the heterozygous state. The mutation is not located within known cis regulatory sequences and according to in silico analysis is unlikely to

cause splicing changes. Figure 46 shows the ABI sequencer traces for the wild type allele and the variant.

Obtaining sequences of the region on the parents of the probands carrying the variant revealed that it was paternally inherited in two of them and maternally inherited in one. The parents carrying the variant do not show an abnormal cardiac phenotype Figure 47 shows the sequencer traces for the parents. Only one of the families has family history of CCM with one uncle having a nonspecific history of a “hole in the heart”, that could not be further investigated in the context of this study.

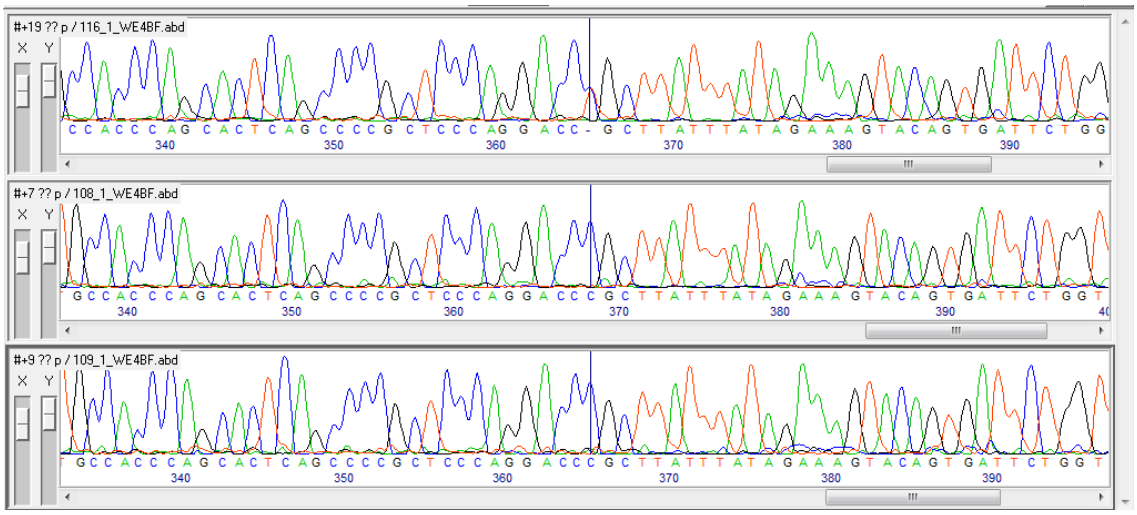

Figure 46. WNT5a mutation 1461 +38 C>T trace display. Upper panel shows variant trace. Medium and bottom panel shows wild type traces.

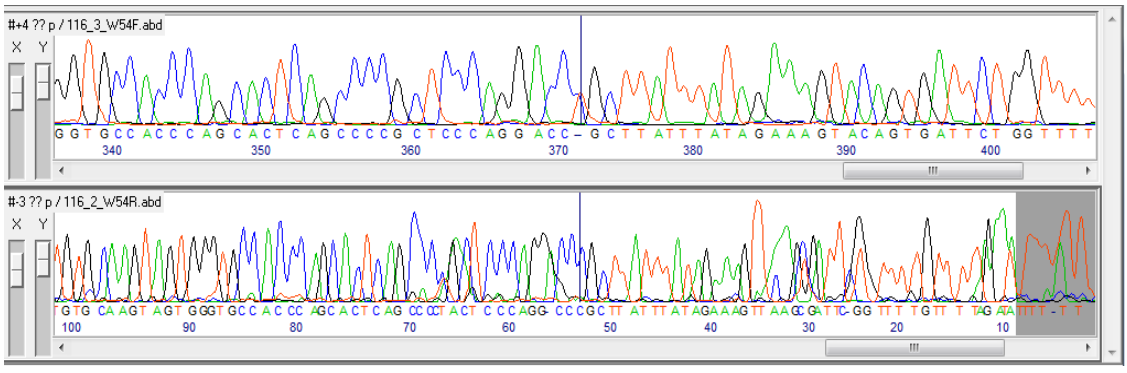

Figure 47. WNT5a mutation 1461 +38 C>T parents trace display. Upper panel shows father’s trace. Bottom panel shows mother’s trace.

	CHA		HTO	
	mAF	MAF	mAF	MAF
Asp119Ser	0.9989	0.0011	1.0000	0.0000
1461 +38 C>T	0.9988	0.0012	0.9805	0.0195

Table 59. Allele frequencies for Wnt5A Asp119Ser and 1461 +38 C>T. mAF: Minor allele frequency; MAF: Major allele frequency

6.1.4. Wnt5A 1461 +38 C>T is present in controls

Three hundred and ninety five CHANGE TOF proband genotypes and 1293 HTO control genotypes were obtained for 1461 +38 C>T using a Sequenom experiment and following methods detailed in section 2.3.2.

For 1461 +38 C>T power was calculated in 90% for a general case-control analysis, hypothesising a relative risk of 2 for the heterozygotes and alpha error of 0.05. Genotyping was successful for 98% of the samples. The estimated genotype miscall rate was less than 1% for the SNP. 1461 +38 C>T was in Hardy-Weinberg equilibrium at the 5% significance level for the global population and for each of the subpopulations.

Fourteen heterozygotes and 1 homozygote for the 1461 +38 C>T allele were found among probands. Fifty heterozygotes for the T allele were identified within control samples. I did not find a significant difference at genotype or allele frequencies between cases and controls for Wnt5A 1461 +38 C>T (χ^2(2 df) = 0.21; p= 0.64)

6.2. Discussion

Wnt 5a has been detected in the primitive streak, precardiac mesoderm, outflow tract and primitive ventricle of chick and mice embryos (71). Also, as previously discussed in this thesis, both canonical and non-canonical Wnt signalling are critical for cardiogenesis and their disruption has been previously linked to cardiac malformations in animal models (90, 114).

Two previously undescribed uncommon variants were found at our mutational screening on Wnt5A. The first uncommon variant found was a non synonymous A>G change that produces a change in residue number 119 from Aspartate to Serine at the Wnt conserved domain. In silico analysis of the mutation using the web based resource Polyphen indicates that it is benign. The

variant is uncommon. Only the initial individual identified on the mutational screening was found to be carrying the variant, none of 1336 controls or any other of the probands (456) was found to be carrying it. It was impossible to determine if the variant was de novo due to that we had no DNA sample from the parents.

The association between Wnt5a and TOF phenotype is complicated to elucidate. However it is well known the involvement of the protein in convergent extension processes. The relevance of the non canonical Wnt pathway derives from the seminal research in zebrafish mutagenesis that resulted on the description of the mutant fish named *siberblick (slb), pipetail (ppt), trilobite (tri), and knypek (kny)* (196) . Similar evidence were subsequently gathered from works on *xenopus* (197, 198). It is also widely known that Wnt5A (*Drosophila* ppt) in one of the principal activators of the Wnt PCP pathway (86).

Sequencing studies of many hundreds if not thousands of cases and controls would be required to statistically compare the frequency of uncommon variants (less than 1% frequency) between cases and controls in any given gene. While such studies are now entering the realms of technical feasibility, this was not the case during the conduct of my project. Therefore, having discovered a rare variant absent in a large number of controls, the only way to proceed is to carry out functional investigation. As outlined above, previously published animal model data strongly supports the idea that a loss of function at Wnt5A might be associated with OFT malformations. However, early attempts to test the functional implications of Wnt5a Asp119Ser were unsuccessful (data not shown). It is necessary, in the future, to design and perform a series of experiments to explore the functional implications of the variant at the developmental and cellular levels. To test the impact of Wnt5a Asp119Ser in cell migration using direct stimulation or knockdown experiments might be the objective of subsequent experiments.

The second uncommon variant found at the initial mutational screening was a single nucleotide change at the 3'untranslated region (1461 +38 C>T). The mutation is not located within known cis regulatory sequences and according to in silico analysis is unlikely to cause splicing changes. No significant differences were found at allele frequencies between 395 CHANGE TOF probands and 1293 HTO control genotypes. The variant was considered of minor pathological relevance.

6.3. Conclusions

I described two previously undescribed uncommon variants at Wnt5A. The first of them is a non synonymous A>G change that produces a change in residue number 119 from Aspartate to Serine

at the Wnt conserved domain. In silico analysis showed that the variant might be benign. This non synonymous variant was found as a “singleton”: only the initial individual identified on the mutational screening was found to be carrying the variant, none of 1336 controls or any other of the probands (456) was found to be carrying it. Due to the findings previously exposed, Asp119Ser is thought to be probably associated with TOF. Previous published works indicate that, due the frequency detected, the variant might be de novo, and at least mildly medically detrimental (191).

Despite a variant found as a singleton (Asp119Ser) in Wnt5a and not found in controls, I did not find any other evidence of uncommon genetic variation within the gene being associated with TOF. Even if the functional implications of Asp119Ser are elucidated and it is in fact a functionally deleterious variant, the PAR explained by this extremely uncommon variant will be very low. It is unlikely that a significant portion of TOF heritability is explained by uncommon genetic variation in Wnt5a.

7. DISCUSSION

7.1. Overview

This chapter summarizes the findings of this thesis in the broader context of human and animal genetic studies of CCM. Detailed discussion for each of the genes has been provided within the body of the results chapter. Here I also discuss the results of my sequencing based studies in the context of approaches targeting common genetic polymorphisms and copy number variation that have emerged during the time this work has been undertaken.

The principal positive result of my work was the detection of a novel LFIP variant in ROCK 1, 807C>T, presence of which was significantly associated with the risk of TOF. The variant conferred an OR of 4.89 (95% CI: 2.51-9.50), and the association was significant at p=2x10^-5; this association remained significant after correction for multiple testing using QVALUE with an FDR of 5%. The variant appeared not to affect splicing in a functional assay; a possible effect on the regulation of ROCK1 expression during development remains to be confirmed. An association between a previously described common variant in ROCK1 (rs288979) and TOF was also found.

I found one non-synonymous variant in the genes I screened that was absent in 1000 chromosomes from a population free of CVM; this was the Asp119Ser substitution in the Wnt5A gene. Functional investigation of the effect of this Wnt5A variant on cell motility was undertaken, but this was hampered by the difficulties in the detection of the protein (Data not shown). I also found an intronic variant in WNt11, 107 +47 G>T, in 9 TOF subjects and none of more than 1000 control chromosomes. The variant appeared not to affect splicing in a minigene experiment.

7.2. Summary of uncommon variants found in Wnt pathway genes

The first approach of this work was to explore uncommon genetic variation within the non-canonical Wnt pathway by performing a mutational screening on the exons of seven key genes within the pathway. The rationale of doing this was derived from two key facts that have been previously mentioned but that will be discussed further in this chapter. First, the amount of evidence associating the PCP pathway with the process of cardiogenesis. Second, the results from previous research suggesting that a portion of genetic predisposition to non syndromic cardiac malformation may be related to uncommon genetic variants of intermediate penetrance.

Twelve uncommon variants were found at the 6 genes in the mutational screening using samples from 93 TOF probands. Four of them were non synonymous, 2 were located at untranslated regions, 3 were intronic and 3 synonymous. None of the 465 probands genotyped for all uncommon variants carried more than one of them. Findings of the screening are summarized at Table 60.

None of the variants was proven to be de novo, but parent samples were missing for 4 of the probands, making it impossible to discard that hypothesis. Statistically significant differences in allele frequencies were found for two variants, ROCK1 C807T and WNT11 207 +47 G>T. Estimates derived from the 1000 genomes project are that the substitution probability in the human genome is $1.2X10^{-8}$ per base pair, per generation. Positive in silico assays for structure disruption or splicing changes were found in four cases: ANKRD6 Thr176Asp, WNT11 207 +47 G>T, ROCK1 807 C>T and 1785 T>G. The decision of going further into functional analysis was made in three of the variants: ROCK1 807 C>T, a synonymous variant with significant allele frequency differences between cases and controls, showing a plausible influence on splicing; WNT5A Asp119Ser, a missense variant found as a singleton and not present in controls: and WNT11 207 +47 G>T, an intronic variant found in 9 probands and not present in controls with positive results on the in silico analysis for splicing disruption. Each case will be discussed in detail below.

	Variant	Function	De novo?	Number	mAF	Present in controls	mAF in controls	p value for controls	In silico results	Functional tests
ANKRD6	1 -304 A>G	UTR (-304)	U	5	0.0020	Y	0.0060	0.06100	Neg	N
	Leu192Phe	NS	U	1	0.0005	Y	0.0003	0.42000	Neg	N
	Thr176Asp	NS	M	1	0.0008	Y	0.0011	0.62400	Pos	N
WNT 5A	Asp119Ser	NS	U	1	0.0005	N	0.0000	NA	Neg	Y
	1461 +38 C>T	UTR	P	14	0.0170	Y	0.0195	0.64000	Neg	N
WNT11	207 +47 G>T	INT	P	9	0.0048	N	0.0000	0.00009	Splicing,SC35	Y
	720 +6 C>T	INT	M	9	0.0047	Y	0.0026	0.53000	Neg	N
DVL3	446 T>A	S	P	11	0.0140	Y	0.0130	0.86000	Neg	N
	1339 +17 A>G	INT	U	2	0.0033	Y	0.0018	0.60800	Neg	N
ROCK1	807 C>T	S	M	33	0.0150	Y	0.0030	0.00002	Splicing, SrP40	Y
	1785 T>G	S	M	5	0.0024	Y	0.0029	0.49000	SC35	N
	Thr773Ser	NS	P	33	0.0155	Y	0.0176	0.77000	Neg	N

Table 60. Uncommon variants encountered at mutational screening on 6 genes at the Non-canonical Wnt pathway. First column lists the genes, with the variants in each one of them listed on the second one. Relevant data about the variants is specified in the table: the genetic functional implications of the variants, if they are de novo, the number of TOF probands carrying the variant, the minor allele frequency in probands, if they were found on controls and the minor allele frequency in that case on them, p value for allele frequency comparison between cases and on controls, results of in silico assays and finally if functional tests were performed on them. Number of cases and controls vary slightly for each variant, see each chapter. UTR: untranslated region; NS: Non synonymous; INT: Intronic; S: Synonymous; U: Unknown; M: Maternally inherited; P: Paternally inherited; Y: Yes; N: No; NA: Not applicable; Neg: Negative; Pos: Positive.

7.3. Weaknesses of association studies

It is expected that in complex systems, like the factors influencing the risk of suffering a common disease, the interactions between the elements (environment and genes) should be taken into account to design effective research strategies. However, genetic association studies are focus on the additive effects of genes simplifying the etiological models of common diseases. Before the GWA era, one of the main problems of associations was the lack of reproducibility of results that can be explained by a series of problems inherent to association studies. Most of the results showed in previous pages correspond to a case control analysis and therefore might have some of the limitations of this type of research.

One of the main problems explaining spurious positive results in association analysis is selection bias. Bias can occur when the population is composed by strata in terms of allele frequencies with probabilities of being selected as a case or control varying between strata. Population substructure is composed by a number of these strata with different risks of suffering the disease. Additionally, measurement of the exposure can be biased if genotyping is performed at different times with different genotyping errors or techniques. Bias can also occur in outcome measurement if genotypes are known before classification.

In terms of my research there is the small possibility of population stratification in the selection of controls. The main control population used was the HTO, a population recruited based in one hypertensive patient. By definition, the HTOs correspond to a subpopulation and therefore it is possible, but unlikely, that they have different allele frequencies when compared to TOF subjects solely due to their selection. In that respect, a vast amount of genotyping have been performed (and published) in the HTO population showing that the group is not a strata (168, 199-201). However, it is important to not forget that my control population was selected due to one hypertensive proband, and even if most of the subjects were normotensive the ideal control population would have been a ramdomly selected sample from general population.

Other of the common problems encountered in the old case control studies was the lack of power. Sample size was oftenly calculated with an overestimation of the true genetic effect and recruitment of patients was a constraint in many diseases (like TOF) by low disease incidence. Additionally, as it was previously mentioned, effects of common variants in common diseases (when present) are associated with small effects, requiring enormous sample sizes to be detected. In this respect, power calculations were performed and described for each one of the variants tested in probands and controls. In summary, most of the common variants tested showed more than enough power to detect differences with controls using a sample size as the one available to us. However, when uncommon variants were tested the power attained was insufficient to confidently discard the hypothesis of association.

For example power was calculated in 20%, 14% and 17% for ANKRD6 1-304 G>A, Leu192Phe and Thr176Asp respectively for a general case-control analysis, to detect a relative risk of 2 for heterozygotes and supposing an alpha error of 0.1. The variants showed mAFs of 0.004, 0.003 and 0.001. For the uncommon variants like Asp119Ser approximately 17000 TOF cases would be needed to assume with 80% power that the variant is absent on 51000 controls. Due to TOF prevalence it would be extremely difficult to collect that sample. For a prevalence of 0.31/1000 liveborns there would be approximately 18000 TOF cases in the UK, supposing a 100% survival rate and including syndromic cases. It is important to say that the control genotyping was performed to screen for the presence of the variants in controls and not to perform a statistical test of association. In general I believe that power was sufficient for the tests comparing genotype frequencies of common variants between cases and controls. However in the case of including a major number of SNPs or to perform a whole genome screening for uncommon variants, a bigger sample will be needed as it is mentioned in other sections of this chapter.

Spurious results in association analysis can also arise solely by chance. If we take into account the number of tests performed in a trait, a proportion of positive results would be produced by random effects. In my research positive results were corrected taking into account the number of tests performed, using Q value, decreasing the possibilities of positive associations due to chance.

7.4. Mutational load and other strategies

In 2004 Cohen and cols published a paper that is one of the seminal works of the analysis of mutational load in complex diseases. In that work the sequence of three candidate genes (ABCA1, APOA1, and LCAT) previously identified as associated with rare genetic forms of HDL deficiency, was obtained in 32 individuals from general population of 4 ethnic origins and each sex, representing the upper and lower 5% of the plasma HDL levels. The number of non-synonymous variants in the low HDL group was significantly higher when compared to the low HDL individuals and 16% of the subjects in the low-HDL group had variants not present in the high-HDL group. The findings were reproduced in a Canadian cohort (202).

The strategy used in the Cohen and cols work is called mutational load and can be used to determine the contribution of rare genetic variants to common genetic traits, especially quantitative traits. The advantage of analyzing the mutational load in quantitative traits strives in the enrichment of uncommon highly penetrant variants in the extremes of the distribution. Tetralogy of Fallot constitutes an example of discrete trait with difficult quantization of severity due that up to date there is not an accepted form of classifying the condition.

Different strategies and approaches to the question about the contribution of uncommon variants to the predisposition to TOF could have been used and might be the objective of future work. A mutational load in the non canonical Wnt signaling comparing TOF sufferers with general population can show significant differences in the number of uncommon variants reinforcing the idea that the variants found are indeed disease causing alleles.

7.5. Uncommon variants previously associated with CCM

If a search is performed at OVID Medline, from 1988 to 2004 including the terms Cardiac congenital malformation and limiting to the genetics subheading in humans, a total of 1201 articles are cited. Much of the published literature corresponds to positive results on syndromic CCM. Findings in non syndromic CCM are illustrated by three seminal works: the association of NKX2.5 and ASD, TOF and JAG1, and CCM and GATA4.

Schott et al. published in 1998 in Science the results of a genetic linkage screen of 4 families showing autosomal dominant transmission of CCM with the majority of affected individuals showing secundum ASDs and conduction defects. Results showed a linkage peak at 5p35, where NKX2.5 is located. Mutational screening of the gene showed that affected individuals in two of the families shared a missense variant (Thr178Met), and the other two of the families carried a nonsense variant (Gln170ter) (203). Subsequent work showed that mutations on the homeodomain were associated to CCM and often to AV block.

Two works by the group of Goldmuntz reinforced the idea that uncommon variants at the homeo domain were responsible for a percentage of "sporadic" CCM with great phenotypic variability. In 2001 this group screened 114 TOF and 100 control patients describing 4 uncommon variants on probands. Goldmuntz et al. found 3 novel (Glu21Gln, Arg216Cys, and Ala219Val) and 1 previously reported (Arg25Cys) uncommon variants in TOF probands. The three novel variants were not found in controls and the previously reported variant was found in black controls. According to the authors approximately 4% of TOF might be attributed to uncommon variants at NKX2.5 (204). However, Arg25Cys was found at similar frequency in controls and the number of controls genotyped for the novel ones might be insufficient to discard their presence in normal subjects given the frequency of the variants.

In 2003 the same group published a screening on 608 patients including conotruncal anomalies, left sided lesions, ASDs and Ebstein's anomaly. Twelve variants in Nkx2.5 were identified in 18 patients (3%) (205). Ten of the variants were non synonymous, one was an in frame deletion and one was synonymous. It is important to say that despite the number of patients screened, the number of probands in each subgroup is small. It is interesting that most of the variants found were missense rather than synonymous, a fact that is contradictory to evolutionary predictions and opposite to my findings. Recent work from the 1000 Genomes project showed significant evidence of selection pressure against potentially damaging non-synonymous SNPs, since these SNPs are found throughout the genome at lower allele frequencies than would be predicted from an evolutionary neutral model. Finally, despite the undeniable value of the work of Goldmuntz and colleagues the number of controls (n= 50) is certainly too small to categorically deny the presence of the

variants in a normal population. Such a number of controls would only have about 66% power to detect a variant of 1% population frequency, and far lower power to detect variants that would be classified as rare (ie less than 0.5% frequency). Although the contribution of Nkx2.5 to CCM is definite from the Mendelian families described, and the studies of sporadic CCM patients are consistent in suggesting a role for LFIP variants, it seems likely that the contribution of this gene has been overestimated in the literature to date.

In 2000 Eldadah et al published a paper with a genetic linkage search for the cause of non syndromic TOF in a large kindred segregating TOF as an autosomal dominant trait. Data was consistent with linkage at 20p12 on JAG1, a gene previously known to cause Alagille syndrome (See introduction). Analysis of the gene in the kindred revealed a missense mutation G274D being carried by 13 members of the family (206). The variant was not found in 100 unrelated control chromosomes. The interesting fact is that JAG1 haploinsufficiency caused by nonsynonymous mutations seems to cause clinically milder forms of Alagille syndrome including TOF. The fact that disease genes implicated in the pathogenesis of syndromic CCM are involved in the non syndromic forms is important, but not directly explored in this work, since to date there is no syndromic form of human CCM reported to be caused by mutations of the non-canonical Wnt genes that I have studied.

Vidu et al. reported in 2003 the results of linkage analysis in a large family with 16 individuals affected by CCM. A linkage peak was discovered in 8p22-23 were GATA4 is located. Mutational screening of the gene showed a non synonymous variant (G296S) with complete penetrance being carried by all affected individuals. The variant was not found in non affected family members and 3000 unrelated healthy controls. NKX2.5 and TBX5 were also screened without finding any variant, a finding that indicates the genetic heterogeneity of CCM, with different kindreds getting the predisposition from different genes-alleles (207).

At our hypothesis based on findings at NKx2.5, GATA4 and JAG1, as main examples, susceptibility to non syndromic CCM and particularly TOF may be given by a number of uncommon variants not fully penetrant, and subjected to selection pressure proportional to the risk conferred. Taken the evidence previously cited the first step was a mutational screening.

7.6. Splicing functional analysis

After the mutational screening the next step was to test the most likely pathogenic variants in functional tests to link the statistical facts to pathophysiological explanations. Rock 807C>T was the first variant to be tested: a relatively uncommon variant, that confers a high calculated OR, with statistically significant (multiple test corrections < 0.05) differences in genotype frequencies when compared to a large control sample and with results of in silico analysis which showed a high likelihood of causing splicing changes. Splicing was tested using a minigene approach, but the experiment comparing the mutant allele with the wild type failed to show splicing differences.

Wnt11 207 +47 G>T was also functionally tested. This intronic uncommon variant was found in nine probands, 1 TOF in the CHANGE population, 7 TOFs in the Nottingham and Oxford population and another one on the CONCOR population. The variant was not found in controls (p= 0.00009, multiple test corrections < 0.05) and in silico analysis showed that is likely to produce splicing changes. Nevertheless, the minigene experiment testing splicing consequences of the variant was negative: no splicing difference was found between the mutant allele and the wild type. It is important to stress that a major limitation of my minigene experiments is that despite historical positive controls using the same vector and conditions were taken into account, no parallel positive control experiment was performed during the experimental stages of this thesis leaving the possibility of methodological factors preventing the recognition of splicing changes associated with the variants.

The association found for both variants, ROCK1 807 C>T and Wnt11 207 +47 G>T might be explained by two additional reasons, obviating the necessity to explore further the splicing hypothesis (See below). The first explanation is an effect on transcriptional activity. It has been proposed that a big portion of phenotypic variability among humans is due to variation in RNA transcription and processing. Data in the literature supports the idea that synonymous variants influence the transcriptional activity of a gene resulting in a phenotype. In 2006 Wang et al. published a study showing the results of allelic expression imbalance (AEI) experiments in a few genes to correlate previous associations apparently

synonymous SNPs and their phenotypes (208). Wang et al. investigated the relevance of the synonymous SNP C3435T at ABCB1 in transcriptional activity. Recent works have shown that genetic variation in ABCB1 can be correlated to individual response to certain drugs. This group showed a two fold increase in transcriptional activity for the T allele. C3435T was not in linkage disequilibrium with any variant located in the promoter. A next step to take with the synonymous variants found to be significantly associated in our sample might be to analyze in heterozygotes the allelic expression imbalance to confirm or discard a possible influence exerted over transcriptional activity.

The second possible explanation for an association is LD in the region. Exploring the region were ROCK1 is located the nearest gene is ESCO1 a gene that belongs to a conserved family of acetyltransferases involved in sister chromatid cohesion. ESCO1 has been associated with some cancers (209) but it has not previously been associated with CCM. Other genes in the region include SNRPD1 and ABHD3. The two nearest described genes to WNT11 are PRKRIR, a p58 inhibitor, and UVRAG. UVRAG is a gene associated with UV radiation resistance, vitiligo, tumorogenicity decrease in colon cancer and strikingly to left right malformations in a case report (210). Despite LD with a variant in the region can explain the associations found, the likelihood of the nearest genes being associated with TOF is lower than the non canonical wnt genes being responsible themselves for the associations found.

7.7. How likely is it that uncommon variants are pathogenic?

The negative functional results can be explained starting with the premise that these three variants have a high likelihood to be pathogenic. Not only does biologically sound literature link Rock1 and Wnt11 to cardiogenesis, but also previous theoretical approaches to the problem of rare missense mutations and human disease back up the idea that these uncommon genetic variants not found in controls could be pathogenic (191, 211).

The variants discussed: ROCK1 807 C>T, Wnt5A Asp119Ser and Wnt11 207 +47 G>T are uncommon variants not found or found at significantly lower frequencies in controls. The question of the theoretical likelihood of an uncommon variant being deleterious can be

seen not only using the perspective of the known data about structure and function but also taking into account the so called “fitness distribution” of human genetic variation.

Some previous research has investigated the distribution of fitness in humans. Eyre-Walker et al. published a paper in 2006 where they tested a method to estimate the fitness distribution of alleles subjected to selective pressure. They tested the distributions produced by theoretical expressions predicting the number of polymorphisms segregating in any given sample, against sequencing results of 320 human genes to infer the fitness distribution associated with SNPs of different frequencies. However, their estimates are based on the assumption that intronic variants are evolutionary neutral, a fact that is not completely accurate if we take into account the possibility of either splicing changes or changes in gene expression levels that may be visible to natural selection. Both exonic and intronic variants in their sample were skewed towards the uncommon, a finding that they interpreted as due to selection in the exonic case and due to population growth and admixture in the intronic ones. Eyre et al. estimated that 19% of mutations are neutral and 14% slightly deleterious (segregated but not fixed) (212).

Kryukov and his team published a research paper in 2007 where they analyzed the theoretical probability of missense and nonsense variants being medically detrimental according to comparisons between data on human disease causing mutations, chimpanzee divergence and modern human genetic variation. They calculated that approximately 20% of de novo missense mutations are strongly detrimental. This finding is in accordance to the previously published data by Yampolsky (211). Following a similar approach, they estimated that 27% of missense mutations are effectively neutral and 53% are mildly deleterious. According to their calculations 50-70% of de novo variants showing a frequency <1% are at least mildly deleterious.

Recently, in October 2010, a paper that undoubtedly will change the way we interpret association data in humans, specially for uncommon variants, was published in Nature by the 1000 genomes project (213). The work shows the summary of results coming from three projects: low coverage whole genome sequencing of 179 individuals from 4 populations; high coverage sequencing of two mother-father-child trios; and exon targeted sequencing of 697 individuals. Variant discovery number depended on the sequencing strategy with 14.4 million SNPs, 1.3 million indels and 20000 large structural variants being

detected in the low coverage sequencing. Summarizing most of the common SNPs were previously described in public databases but evidently uncommon and structural variants were mostly novel, reflecting poor previous characterization. In terms of functional implications of the variants, they quantified that an average human subject differs from the human genome reference sequence in 10000-12000 synonymous and 10000-12000 non synonymous SNPs. The number of strongly detrimental variants in an average human being is, however, much lower: 190–210 in-frame indels, 80–100 premature stop codons, 40–50 splice-site-disrupting variants and 220–250 frame shift deletions. Also they quantified that a typical human subject is heterozygote for 50-100 variants previously classified as causing inherited disorders.

Given the data published in the era preceding next-generation sequencing, it can be inferred that uncommon variants, like the ones found in the non-canonical Wnt pathway, are subjected to evolutionary pressures and therefore a proportion of them are likely to be pathogenic. However, new data coming from enormous sequencing projects like the 1000 genomes project depict a more complex picture with hundreds and even thousands of these deleterious alleles being carrying by each individual. Gene-gene and gene-environment interactions should now be researched to explain the individual contribution of complex disease pathogenesis in general and TOF in particular.

7.8. Complexity of alternative splicing

One of the difficulties of demonstrating the functional implications of the Wnt11 207 +47 G>T and ROCK1 807 C>T variants is the complexity of alternative splicing in the human genome. Genomic variation leading to splicing changes represent almost 50% of the pathogenic variants among human diseases (214). It has been also demonstrated that the majority of multiexon genes are alternatively spliced making this process the greatest source of proteomic variation. A study published in 2004 researching on the exon-exon junctions in 10000 human genes in 52 tissues and cell lines, using microarrays, concluded than more than 74% of the genes were alternatively spliced (215).

Exon recognition by the splicing machinery on the primary transcripts where small exons should be picked among vast intronic sequences represents a delicate process with numerous factors implicated. The first recognizable factor would be the conserved sequences denoted exon- intron junctions or 5'-3'splicing sites. Recognition of these consensus sequences not only depends on them but also on cis acting sequences within introns and exons (214). These regulatory motifs have been called intronic and exonic enhancers and silencers. Enhancers and silencers influence exon definition trough the recruiting of splicing machinery and the interaction between them.

Most of the reported splicing variants in databases such as the Human Gene Mutation Database correspond to mutations at the consensus splicing sequences causing "exon skipping". For example, this is the operative mechanism in the case of the synonymous transversion described by Montera et al. in 2001 in a family presenting with familial adenomatous polyposis (FAP). This single nucleotide change in nucleotide position 1869 of the APC gene, produces complete skipping of APC exon 14 that translates into FAP (216).

However, less frequently a cryptic splicing site is created and other complex patterns of protein splicing can also be created due to the interactions of the different regulating factors (217). An example of this aberrant splicing pattern is the single nucleotide change in BRCA1 R71G described in several families presenting a familial form of breast cancer. cDNA analysis of the non synonymous (R71G) variant showed that it causes deletion of the last 22bp of BRCA1 exon 5 and the reading frame alteration produces a stop codon in the first nucleotides of exon 6 at codon position 64 (218).

Although disruption of exon and intron junction is not the case for our suspicious variants, located at 47 bp from the 3' junction (Wnt11 207 +47 G>T) and 13 bp from the 3'end of an exon (ROCK1 807 C>T), the theoretical disruption of cis regulatory elements and the repercussion of such changes on splicing caused by these variants was the reason for carrying forward functional investigation on them.

It is worth mentioning that not only local factors are involved in the process of exon recognition. Elements of the wider context within which the exon is found, such as the splicing events taking place before the exon of interest, the type of promoter and the

presence of other regulatory elements throughout the gene (214) should be taken into account to interpret the results of a highly simplified model of splicing such as a minigene. A minigene only includes the variant within the exon or intron of interest and a shorter context within an artificial gene composed by non human exons with a different promoter. If the changes exerted by the variant do not grossly disrupt the splicing junction, a high likelihood of not being recognized might be expected in this type of experiments.

There are multiple examples in the literature of pathogenic variants which could be misclassified as synonymous harmless SNPs, if enhancers, silencers, RNA secondary structure and the wider splicing context are not taken into account (217). Two works illustrate in a mechanistic way the pathogenic implications of this type of splicing variation: In 2000 Ars et al. published a paper exploring the RNA implications of published genomic variants in Neurofibromatosis type 1. They performed a mutational screening in the NF1 gene in 80 NF1 patients identifying 44 previously undescribed variants. Out of the total, 14 variants caused splicing changes with 1 frameshift, 2 nonsense and 2 missense also causing splicing variation. Twelve variants were located at introns with only 6 of them disrupting splicing consensus sites (219). In 1999 a similar study was published by Teroaka et al. analyzing the possible splicing implications of mutations found in a screening of 49 ataxia telangiectasia disease patients. Sixty two variants were found during the screening. Ten of the variants were located at splicing consensus sequences causing mainly exon skipping. However, three of the variants were found out of splicing consensus sites and created cryptic splicing sites. Also, two nonsense variants produced splicing aberrations detected on RT PCR sequencing (220).

It is evident in the previously exposed examples that misclassification of variants during mutational screening due to solely analysis of genomic sequence is possible. Not only is a high level of awareness of this fact required but it is also important to design optimal experiments to test the hypothesis at the splicing level. In my case testing the splicing consequences of these two highly likely pathogenic variants needed the construction of a model of splicing due to the lack of RNA samples from the probands carrying the mutations. As previously mentioned, the model is too simplistic to be taken as a definite confirmation of absence of splicing consequences being associated with the variants. Reverse transcriptase RNA amplification and direct sequencing of the RNA from patients, or from cell lines expressing the entire gene, carrying the variants might be the most direct

way to test the splicing implications of these two variants (Wnt11 207 +47 G>T and ROCK1 807C>T). However, obtaining RNA from patients requires to collect a new sample with the technical and ethical drawbacks associated. This strategy, however, does not explore the impact of alternative splicing and splicing changes taking place and being relevant only at certain stages of heart development.

7.9. Common genetic variants and the risk of TOF

Although my work principally focused on uncommon genetic variation, I also explored the impact of common variation at my candidate genes on the risk of TOF. Common genetic variants were genotyped in the complete CHA population and controls when they failed Hardy Weinberg equilibrium tests (since this might indicate the presence of a copy number variable region) or if they showed significant differences in allele frequency when compared to previous data. Four common variants - ROCK2 rs2271621, ANKRD6 rs2273238, rs9344950 and rs9353687 - were genotyped. No significant differences in allele frequencies were found between cases and controls for any of the previously described common variants.

Two previously undescribed common variants were found in the mutational screening. One of the common variants described is a single base substitution (A>T) at ROCK2 intron 1 (ROCK2 145 -25 A>T). The variant is not located in a conserved domain and in silico analysis did not show changes at splicing cis regulatory sequences. Genotyping of controls for ROCK2 145 -25 A>T did not show significant differences with TOF cases (mAF in probands and controls 0.13). The second common variant found in our sample is a single base substitution (C/T) at ANKRD6 exon 16. The variant produces a change from phenylalanine to leucine at residue 636. The variant is not located in a conserved domain within ANKRD6. In silico analysis of Phe636Leu using the web based resource Polyphen indicated that the variant might be structurally damaging. Genotyping of controls did not show significant allele frequencies differences between cases and controls (mAF: Probands: 0.14: Controls: 0.10).

To elucidate a possible genomic explanation for the strong association found between ROCK1 807 C>T and TOF a tagging strategy (190) was designed to capture common

genetic variation within the gene and to explore linkage disequilibrium relationships between markers. Previously described missense SNPs in the gene were forced to be included. The hypothesis was that the variant was only indirectly associated with the condition and that was the explanation of the negative result at the splicing experiment. No evidence of LD was found between 807 C>T and the other markers, and no common haplotype was associated with TOF. Rs288979, an intronic common variant (mAF=0.051) was significantly associated with TOF (p=0.000015) with a protective effect and OR calculated for its minor allele of 0.7. In silico analysis of the variant did not show splicing changes associated with the SNP. None of the common genotyped variants was in disequilibrium with either rs288979 or ROCK1 807 C>T. Both the uncommon and the common variants could represent independent associations or be linked by a third causal variant. Sequencing of bigger samples using new platforms might confirm their independent associations or find other causal variants not detected up to this moment.

To my knowledge this is the first study to examine human common and uncommon genetic variation at the Wnt pathway and TOF. Also, our sample is the largest reported and the thus far the only with sufficient power to explore genetic variation and its relationship with non-syndromic TOF.

7.10. The genetic architecture of TOF

It is reasonable to expect genetic architecture (number, type, frequency and effect size of susceptibility variants) to vary between traits. Research into the allelic spectrum of human disease has progressed enormously in the postgenomic era and some insights from previous research can give clues about the architecture of TOF. Susceptibility to some monogenic disorders like cystic fibrosis is associated to a relatively small number of alleles with high population attributable risk (PAR) within a frame of less common alleles with lower PARs. Some other monogenic disorders like β thalassemia have multiple associated alleles with some groups presenting a small number of them (221).

In the case of common diseases a good example is Alzheimer disease. A metanalysis published in the pre-GWA era by Betram et al. illustrates the genetic architecture of the condition. Literature review covering 789 publications up to 2005 confirmed the association

with the well known high risk allele ApoE4 with ORs ranging from 2.7 to 5.2, and a population attributable risk of approximately 20%. However, 20 additional positively associated alleles in 16 non ApoE loci were also described with modest ORs associated with them (222). Three GWAs published from 2007 have confirmed this architecture, at least with respect to common alleles, with one allele conferring high OR and accounting for most of the heritability with multiple minor signals accounting for lower risks. In the first GWA published by Coon et al. in 2007 the association with rs4420638, a SNP located 14Kb from ApoE, was the strongest signal detected (OR=4), confirming the previously described association (223). However, subsequent GWAs, one performed in 1411 Alzheimer disease sufferers carrying ApoE4 found a significant association with rs2373155 a SNP located in GAB2 (224), and one published recently showed a novel SNP associated located in MTHFD1L (225).

Another example with a different character is the case of the variants associated with type 2 Diabetes. Multiple GWA have identified numerous loci associated with type 2 diabetes confirming the hypothesis that genetic susceptibility for this condition is given by a heterogeneous group of variants associated with low risk. Up to date 22 GWAs have been published exploring genetic predisposition to type 2 diabetes, describing 101 significantly associated loci.

Due to the findings of this thesis it might be concluded that CCM, and particularly TOF, is a group of diseases where common and uncommon variants coexist and are both responsible for genetic predisposition. Obviating the necessity of external reproduction of my findings, the allelic architecture of TOF in the Non-canonical Wnt pathway genes comprises a common variant in ROCK1 (rs288979) conferring a modest apparently protective OR (0.7), but due to its mAF, being associated with a PAR of 11%; and two uncommon variants at ROCK1 (807C>T) and Wnt5A (207 +47 G>T) apparently conferring high risk to suffer the disease but with low PARs associated. It is tempting to speculate on the consequences of surgical repair on the genetic architecture of CCM and particularly on TOF. It is however unlikely due to the number of generations since the introduction of surgery that a quantifiable effect on allele frequencies have been exerted by the late fitness increase of TOF sufferers. It is also to note that given the negative selective pressure associated with TOF susceptibility alleles, surgical repair could have, as the

most, increased the fitness to 1 and under normal conditions this would mean that allele frequencies will be maintained on the population.

7.11. Genome wide association studies in complex diseases

During the experimental work of this thesis the landscape in complex disease genetics entirely changed. Due to the modest results of the candidate gene approach in research of complex human traits, specific and powerful tools were needed to explore the contribution of common and uncommon genetic variants in large samples and with a “genomic” orientation. Development of human genetic variation databases and commercial highthrougput genotyping methods like the “SNP chips” allowed research efforts exploring hundreds of thousands of common genetic variants in samples of thousands of patients and controls. This type of research study is usually termed Genome wide association analysis (5, 8).

The WTCCC study, published in Nature in 2007, was the first of its kind that show the potential of GWAs. The WTCCC was a joint GWA study made on a British population examining 7 diseases on 2000 probands and sharing 3000 controls, identifying 24 independent association signals for rheumatoid arthritis, type 1 and 2 diabetes, bipolar disorder, Crohn’s disease and coronary artery disease (9). Up to date 674 published works describing the association with 3275 SNPs are listed on the GWAs catalogue (17). The results of GWAs represent a wealth of data with some optimistic views considering that in 3 years most of the common genetic variation associated with human complex diseases will be identified. A review of the GWAs accomplishments is beyond the scope of this chapter but a good example are the advances in coronary artery disease (CAD) genetics. The first breakthrough in CAD was the publication in Science of two works reporting the first and most important locus associated with CAD: 9p21.3. In the work by McPherson and colleagues a GWA was performed in a 312 case cohort of Canadian subjects suffering early onset CAD. Significant results were tested in a second sample comprising 311 additional cases. Fifty SNPs significantly associated in this second cohort were tested in a third sample comprising 1347 cases and 9054 controls. Two SNPs, located in 9p21 20Kb apart, were associated with CAD: rs10757274 and rs2383206. The association was validated in three additional cohorts (226).

A second work by Helgadottir et al. also reported an association between CAD and a LD block of 190 Kb located in 9p21 represented in their analysis by three SNPs: rs1333040, rs2383207 and rs10116277. The work by the group of Helgadottir was performed in an Icelandic population comprising 1607 myocardial infarction sufferers and 6728 healthy controls without CAD history (227). It is interesting that, like many of the associated SNPs in GWAs, the variants in 9p21 associated with CAD are located in non coding regions. However, the variants seems to be involved in the production of a short form of a downstream gene (ANRIL) and probably influence the expression of cyclin dependent kinases (228). The region (9p21.3) has also been associated with aneurismal dilation of abdominal aorta and cerebral vessels, diabetes and numerous cancers opening the possibilities of multiple pathogenic implications. In a work from our laboratory, published in March 2009, Cunnington et al. tested the association between the SNPs located at the region and carotid intimal thickness and known cardiovascular risk factors without finding any significant association. Despite the robustness of the association between 9p21 and CAD, the functional implications of the variants and the clinical relevance for daily practice are still open questions.

It is important to say that up to the moment of this thesis no GWA in CCM has been published. The finding in this work of a common variant in ROCK1 (Rs288979) significantly associated with TOF and previous findings in numerous complex traits support the idea that a least a portion of heritability-predisposition to CCM, and particularly to TOF, is due to common genetic variation. Despite the difficulties in recruiting a sample of any CCM big enough to attain acceptable power, is evident that one of the necessary steps to follow is to perform a GWA with TOF probands. The sample of the CHANGE population recruited by our laboratory can be the starting point for this work.

GWAs were designed to capture most, if not all, common genetic variation on the genome, but it has become clear in the last three years that most of complex disease heritability is still not explained by the associated loci (8). A number of explanations have been suggested to explain this “missing heritability” from GWAs. Importantly, the possibility of larger numbers of variants of smaller effect not found yet; the effect of uncommon variants poorly detected by available genotyping arrays; structural variants not captured (like copy

number variants (CNVs)); low power in studies thus far to detect gene to gene interaction; and the still to be researched interaction between genetics and environment.

Since the description of the first copy number variant (CNV), trisomy 21, fifty years ago, a vast amount of developments have changed our understanding of this type of genetic variation. Despite CNV corresponding to any variation of the dogma stating that genes come in pairs, from small indels to large chromosomal variants, traditionally the term CNVs have been used to describe DNA segments of 1Kb or more that show copy number variation when compared to reference sequence (229). In the last 5 years recognition of CNV prevalence in our genome have been catalogued and documented in databases as the Database of Genomic Variants (DGV) (230). The DGV lists 66741 CNVs in October 2010 (http://projects.tcag.ca/variation/). In terms of abundance, it is still debated if whether or not CNVs are more abundant than SNPs in the human genome. Nevertheless, the amount of genetic variation due to CNVs is far greater than that due to SNPs, with a group quantifying that 1.28% of nucleotide differences between humans is due to CNVs (231). In terms of phenotypic impact of CNVs, this depends on the variant size, with microscopically visible CNVs almost always being associated with a phenotype. Smaller CNVs can be clinically relevant, being influential to normal human phenotypic variation or being functionally silent (229).

In 2008 Eichler et al. published the results of a screening for CNVs at the 1q21.1 region in 5218 patients suffering unexplained mental retardation, autism or congenital anomalies. This group identified 25 unrelated probands carrying an overlapping deletion and 9 additional probands carrying the reciprocal duplication in the region. Analysis of the same region in 4737 controls showed no deletion and only two small duplications. The deletion seems to be associated not only to cognitive impairment but also to a number of phenotypes as cataracts, joint hipermobility, seizures or hypotonia. Interestingly 6 of the 25 patients carrying the deletion showed CCM. As explanations for the phenotypic variability associated with the deletion Eichler et al hypothesized that factors as genetic imprinting, epigenetic influences, genetic background or recessive variants expressed in the case of deletions (232). This might be the case for the associated, not fully penetrant, relatively uncommon variants found, as ROCK1 807 C>T. Epigenetic, background or associated recessive uncommon variants can influence the development of the fully TOF phenotype.

This ground breaking paper opens the door to explore the possible pleiotropic effects in development of such key genes as the non canonical wnt pathway.

In 2009 a paper published by Greenway et al. reported the results of a genome copy number variants (CNV) survey on 121 TOF trios (proband and two unaffected parents) using an Affimetrix 6.0 chip assay. A total of 17 rare CNVs not present in controls were identified. One percent of non syndromic TOF sufferers presented a CNV on 1q21.1; additional CNVs were identified at 3p25.1, 7p21.3 and 22q11.2; in a single individual 6 CNVs were identified, two of them on genes previously reported to be associated with TOF. These results illustrate the relevance of the previously mentioned structural variants not captured in most GWAs, a field needing to be explored at common disease genetics. One of the next steps needed to be taken to unravel all components of heritability of TOF, especially at the non canonical Wnt pathway is to explore CNVs in our population.

7.12. Environmental factors in complex diseases

In 2006 Hemminki et al. published a paper exploring the implications of environment in complex disease susceptibility (233). Epidemiological research has shown that environmental population attributable risk in complex diseases, like cancer is 90%, and coronary artery disease is over 70% (234). If we take into account that environmental risk factors' effect is not additive and that the individual response to an exposure also depends on the genetic makeup, research into these environmental factors is complex. A good example is the mental retardation associated to phenylketonuria, depending on the point of view the disease can be seen as 100% environmental (dietary) or genetic. Additionally, there are problems determining the environmental components of heritability. A non quantifiable portion of family clustering in common diseases is given by shared environmental exposures. It is worth asking how much of the heritability in CCM is due to shared environment and establish not only a strategy to include environmental factors into genetic research but also the maximum possible of the heritability expected to be linked solely to genetic variants. Lately the National Institutes of Health announced a gene-environment initiative to boost the inclusion of environmental exposures in genetic research.

Due to the difficulties in obtaining accurate environmental information, and due to that the design of sample recruitment was not made to specifically test gene-environment interactions, the amount of information in this thesis about environmental influences is scarce. However, some facts are interesting, in our sample pregestational diabetes prevalence is 1.2%, higher than the 0.4% previously reported for UK population (235). At the same time the prevalence of gestational diabetes in CHANGE mothers is lower than the reported for general population. The etiological implications of these facts are unknown, and they can be purely unrelated facts. Designing a strategy of data mining in our TOF CHANGE sample to obtain sufficient and reliable environmental data is key to quantify the effect of such factors, and to identify the ones likely to interact with genotypes to increase the risk of suffering TOF.

7.13. Next generation sequencing

As previously discussed, GWAs are designed to capture common genetic variation (mAF>5%) and its relationship to human complex diseases. Despite encouraging results, the majority of trait heritability is not explained by the associated SNPs found in GWAs. An example of this fact is the case of Diabetes type 2. A metanalysis of the available GWAs in diabetes mellitus type 2 was published in 2008 comprising the data from a total of 4549 cases and 5579 controls (236). Despite the finding of 18 significantly associated SNPs, only 6% of heritability has been explained. Another interesting fact derived from the results of GWAs is that despite multiple variants being significantly associated with a myriad of traits, only rarely have the functional implications/causal variants been identified, and in many cases uncommon variants found in the associated blocks have been overlooked due to the presumption of a common variant being the causal one (237). It is expected that after the description and classification in databases of variants with frequencies between 1-5%, as a result of projects like the 1000 genomes, these variants are going to be included in GWAs and another portion of heritability will be explained.

Three of our positively associated variants are uncommon, and they explain only a small fraction of heritability in TOF. There is a strong possibility of numerous uncommon variants across the genome contributing to the risk of suffering TOF, and this hypothesis should be

explored. Discovery of variants with mAF<1% can only be achieved by sequencing complete genomes-exomes in large samples, and that task was until recently regarded as impossible.

During the period of my work, a number of new technologies called “high throughput sequencing” or “next generation sequencing” were made available. These technologies outperform the traditional Sanger technology by a factor of 100-1000 with a cost that is 0.1-4% of the old Sanger sequencing cost. Four types of these next generation sequencing are now on the market: pyrosequencing on the Roche/454 GS FLX Titanium sequencer, the first of these platforms, allowing sequencing of 750Mb a day at a cost of about U$20 per Mb; the Illumina Genome Analyzer with a capacity of approximately 5000Mb per day at a cost of U$0.5 per Mb; the Applied Biosystems SOLiD with a capacity and price similar to the Illumina platform; and Helicos HeliScope, the only platform able to sequence complete molecules, with a performance of approximately 4250 Mb a day at a cost of U$ 0.33 per Mb but, at present, very poor accuracy (238). All of these platforms have the capacity of deep resequencing of large samples at achievable costs and in reasonable time spans.

The idea of studies that involve whole genome/exome sequencing of all subjects is now possible, and is reasonable to expect that they will performed as the GWAs. In fact various examples have shown that is possible to identify rare disease causing variants using this type of approach. Two studies by Ng et al. have demonstrated reported the identification of the causal variants of two mendelian conditions, the Freeman Sheldon syndrome and the Miller syndrome, using whole exome sequencing (239, 240).

However, although the strategy of whole genome sequencing seems straightforward, saving technical issues, it remains the question on how to approach the research, meaning who should be sequenced and how to confront the daunting task of analyzing the amount of data produced by this type of study. One of the proposed answers is to sequence multiple affected individuals in families looking for co-segregation. Nevertheless, the amount of co segregating variants, even if only the functionally relevant ones are taken, would be high making it necessary to narrow down depending on gene functions or previous linkage/association studies results (237). Another possible strategy is to

sequence the extremes of the phenotype distribution, knowing that in such extremes rare causative variants are enriched.

The possibility of correctly identifying uncommon genetic variants associated with common diseases is now feasible with the use of next generation sequencing. The discovery of variants with large effects on gene function and influence in disease risk not only will allow the calculation of risk, but more importantly might give therapeutic targets for intervention in common diseases The idea of resequencing complete exomes or at least all genes at the Non-canonical Wnt pathway, in large samples will complete the puzzle of the involvement of uncommon genetic variation in TOF.

7.14. Clinical perspectives

As it is apparent from the results of this thesis not only multiple processes but, among a single pathway, numerous genetic variants, common and uncommon, might be involved in the pathogenesis of CCM and particularly of TOF. The implications of such variants are still under research and other relevant variants should be discovered in the future using new strategies designed to detect common and uncommon genetic variants.

There are three main fields where the complete description of the genetic variants responsible for non syndromic TOF development is relevant in day to day clinical practice and patient care. It is important to note that a long way is still needed to be walked and the results of this thesis only represent one of the first steps towards a clinical application of basic research in the etiology of TOF. Firstly, it might be possible that with the identification of the complete genetic architecture of TOF, genetic risk could be calculated. It is however impossible to say if an accurate risk calculation would be attainable in the near future. Risk calculation depends largely in PAR attributable to genetics, environment and all possible relationships between them. According to the results of this thesis, risk of suffering TOF is given by a combination of common and uncommon variants with low PARs, the nature and number of the variants yet undiscovered is still unknown. The completion of the picture with whole genome sequencing and GWA studies will unravel if genetic risk calculation is possible or if in the other hand risk is heavily dependent on numerous uncommon variants making calculations for specific cases difficult.

Identification of genetic variants associated with non syndromic TOF might also be important in determining prognosis. However, phenotype-genotype relationships have not yet been established. It is important to say that due that our sample was not recruited with that objective in mind, and given the frequency of the variants associated, no genotype-phenotype correlation could be inferred.

Finally and perhaps the last clinical goal of genetic research in common diseases is making possible therapeutic interventions to prevent in high risk groups the incidence of CCM. Or in a more broad sense to provide therapeutic targets to either truncate the anomalous embryogenetic process or to specifically treat the complications. It is impossible to say at this moment if such interventions are possible. Only the completion of the genetic research in TOF and the discovery of the majority of implicated genetic variants can give us targets to begin working.

8. BIBLIOGRAPHY

1. Kiberstis P, Roberts L. It's not just the genes. *Science* 2002; 296:685:686.

2. Haines J, Pericak-Vance M. *Overview of mapping common and genetically complex disease genes.* New York: Wiley and sons; 1998.

3. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nature Immunology* 2001; 2(9):803-809.

4. Colhoun HM, McKeigue PM. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; 361:865-872.

5. Altshuler D, Daly M. Guilt beyond a reasonable doubt. *Nature Genetics* 2007; 39(7):813-815.

6. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; 322(5903):881-888.

7. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273(5281):1516-1517.

8. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; 461(7265):747-753.

9. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447(7145):661-678.

10. Hardy J, Singleton A. Genomewide association studies and human disease. *New England journal of Medicine* 2009; 360(17):1759-1768.

11. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Human Molecular Genetics* 2002; 11(20):2417-2423.

12. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development* 2009; 19(3):212-219.

13. Alfonso-Sanchez MA, Perez-Miranda AM, Garcia-Obregon S, Pena JA. An evolutionary approach to the high frequency of the Delta F508 CFTR mutation in European populations. *Medical Hypotheses*; 74(6):989-992.

14. Wiuf C. Do delta F508 heterozygotes have a selective advantage? *Genetical Research* 2001; 78(1):41-47.

15. Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *American Journal of Human Genetics* 2000; 67(4):881-900.

16. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genetics* 2000; 26(1):76-80.

17. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 2009; 106(23):9362-9367.

18. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; 411(6837):599-603.

19. Van Limbergen J, Wilson DC, Satsangi J. The genetics of Crohn's disease. *Annual Review of Genomics & Human Genetics* 2009; 10:89-116.

20. Oostenbrug LE, Nolte IM, Oosterom E, van der Steege G, te Meerman GJ, van Dullemen HM, Drenth JP, de Jong DJ, van der Linde K, Jansen PL, Kleibeuker JH, Drenth JPH, Jansen PLM. CARD15 in inflammatory bowel disease and Crohn's disease phenotypes: an association study and pooled analysis. *Digestive & Liver Disease* 2006; 38(11):834-845.

21. Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *American Journal of Gastroenterology* 2004; 99(12):2393-2404.

22. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics* 2001; 69(1):124-137.

23. Hoffman JIE, Kaplan S. The incidence of congenital heart disease. *Journal of the American College of Cardiology* 2002; 39(12):1890-1900.

24. Rubin J, Ferencz C, McCarter R. Congenital cardiovascular malformations in the Baltimore Washington Area. *MD Med J* 1985; 34(11):1079-1083.

25. Pradat P, Francannet C, Harris JA, Robert E. The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. *Pediatric Cardiology* 2003; 24(3):195-221.

26. Eugene Braunwald, Michael Gatzoulis, Lorna Swan, Therrien J. *Adult Congenital Heart Disease: a Practical Guide*. 1st ed: WileyBlackwell; 2005.

27. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. Recurrence of congenital heart defects in families. *Circulation* 2009; 120(4):295-301.

28. Srivastava D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* 2006; 126(6):1037-1048.

29. Hoffman JIE, Kaplan S, Liberthcon R. Prevalence of congenital heart disease. *American Heart Journal* 2004; 147(3):425-439.

30. Warnes C. The adult with congenital heart disease. Born to be bad? *J Am Coll Cardiol* 2005; 46:1-8.

31. Wren C, O´Sullivan JJ. Survival with congenital heart disease and need for follow up in adult life. *Heart* 2001; 85:438-443.

32. Zipes D, Libby P, Bonow R, Braunwald E, eds. *Braunwald´s Heart Disease*. 7th ed. Philadelphia: Elsevier Saunders; 2005.

33. Ferencz C, Boughman J, Neill C, Brenner J, Perry L. Congenital cardiovascular malformations: Questions on inheritance. *J Am Coll Cardiol* 1989; 14:756-763.

34. Uebing A, Steer P, Yentis S, Gatzoulis M. Pregnancy and congenital heart disease. *BMJ* 2006; 332:401-406.

35. Burn J, Brennan P, Little J, Holloway S, Coffey R, Somerville J, Dennis NR, Allan L, Arnold R, Deanfield JE, Godman M, Houston A, Keeton B, Oakley C, Scott O, Silove E, Wilkinson J, Pembrey M, Hunter AS. Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet* 1998; 351(9099):311-316.

36. Nora J. From generational studies to a multilevel genetic-environmental interaction. *J Am Coll Cardiol* 1994; 23:1468-1471.

37. Mone S, Gillman M, Miller T, Herman E, Lipshultz S. Effects of Environmental Exposures on the Cardiovascular System: Prenatal Period Trough Adolescence. *Pediatrics* 2004; 113:1058-1069

38. Freedom RM, ed. *Neonatal heart disease*
New York: Springer-Verlag; 1992.
39. National Institutes of Health. National Center for Biotechnology Information. **www.ncbi.nlm.nih.gov.**
40. Ransom J, Srivastava D. The genetics of cardiac birth defects. *Semin Cell Dev Biol* 2007; In press.
41. Kobrynski LJ, Sullivan KE. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* 2007; 370(9596):1443-1452.
42. Goodship J, Cross I, LiLing J, Wren C. A population study of chromosome 22q11 deletions in infancy. *Archives of Disease in Childhood* 1998; 79(4):348-351.
43. Botto LD, May K, Fernhoff PM, Correa A, Coleman K, Rasmussen SA, Merritt RK, O'Leary LA, Wong LY, Elixson EM, Mahle WT, Campbell RM. A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. *Pediatrics* 2003; 112(1 Pt 1):101-107.
44. Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, Jurecic V, Ogunrinu G, Sutherland HF, Scambler PJ, Bradley A, Baldini A. Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 2001; 410(6824):97-101.
45. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nature Genetics* 2001; 27(3):286-291.
46. Xu H, Cerrato F, Baldini A. Timed mutation and cell-fate mapping reveal reiterated roles of Tbx1 during embryogenesis, and a crucial function during segmentation of the pharyngeal system via regulation of endoderm expansion. *Development* 2005; 132(19):4387-4395.
47. Kelly RG, Jerome-Majewska LA, Papaioannou VE, Kelly RG, Jerome-Majewska LA, Papaioannou VE. The del22q11.2 candidate gene Tbx1 regulates branchiomeric myogenesis. *Human Molecular Genetics* 2004; 13(22):2829-2840.
48. Ilagan R, Abu-Issa R, Brown D, Yang YP, Jiao K, Schwartz RJ, Klingensmith J, Meyers EN, Ilagan R, Abu-Issa R, Brown D, Yang Y-P, Jiao K, Schwartz RJ, Klingensmith J, Meyers EN. Fgf8 is required for anterior heart field development. *Development* 2006; 133(12):2435-2445.
49. Taddei I, Morishima M, Huynh T, Lindsay EA. Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/del22q11

syndromes. *Proceedings of the National Academy of Sciences of the United States of America* 2001; 98(20):11428-11431.

50. Munoz S, Garay F, Flores I, Heusser F, Talesnik E, Aracena M. Heterogeneidad de la presentacion clinica del sindrome de microdelecion del cromosoma 22. *Revista medica de Chile* 2001; (129):515:521.

51. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D. Mutations in NOTCH1 cause aortic valve disease. *Nature* 2005; 437(7056):270-274.

52. Leong FT, Freeman LJ, Keavney BD. Fresh fields and pathways new: recent genetic insights into cardiac malformation. *Heart* 2009; 95(6):442-447.

53. Hunter W. Medical observations and inquiries. London: Private publication; 1784. p. 417-419.

54. Apitz C, Webb GD, Redington AN. Tetralogy of Fallot. *Lancet* 2009; 374(9699):1462-1471.

55. National Institutes of Health. National Center for Biotechnology Information. **www.ncbi.nlm.nih.gov.**

56. Wren C, Richmond S, Donaldson L. Temporal variability in birth prevalence of cardiovascular malformations. *Heart* 2000; 83:414-419.

57. Cincinnati Children's Hospital medical Center. Tetralogy of Fallot. **www.cincinnatichildrens.org2006).**

58. Shinebourne EA, Babu-Narayan SV, Carvalho JS. Tetralogy of Fallot: from fetus to adult. *Heart* 2006; 92(9):1353-1359.

59. Murphy JG, Gersh BJ, Mair DD, et al. Long-Term outcome in patients undergoing surgical repair of tetralogy of Fallot. *New England journal of Medicine* 1993; 329:593-599.

60. Williams RG, Pearson GD, Barst RJ, Child JS, et al. Report of the National Heart, Lung, and Blood Institute Working Group on Research in Adult Congenital Heart Disease. *2006* 2006; 47:701-707.

61. Kirsh JA, Stephenson EA, Redington AN. Images in cardiovascular medicine. Recovery of left ventricular systolic function after biventricular resynchronization pacing in a child with repaired tetralogy of Fallot and severe biventricular dysfunction. *Circulation* 2006; 113(14):e691-692.

62. Johnson MC, Hing A, Wood MK, Watson MS. Chromosome abnormalities in congenital heart disease. *American Journal of Medical Genetics* 1997; 70:292-298.

63. Boon AR, Farmer MB, Roberts DF. A family study of Fallot's tetralogy. *Journal of Medical Genetics* 1972; 9:179-192.

64. National Institutes of Health. OMIM. **www.ncbi.nlm.nih.gov.**

65. Buckingham M, Meilhac S, Zaffran S. Building the mammalian heart from two sources of myocardial cells. *Nature* 2005; 6:826-835.

66. Olson EN. Gene regulatory networks in the evolution and development of the heart. *Science* 2006; 313(5795):1922-1927.

67. Kirby M. Molecular Embryogenesis of the Heart. *Paediatric and Developmental Pathology* 2002; 5:516-543.

68. Munoz-Chapuli R, Perez-Pomares JM. Cardiogenesis: an embryological perspective. *Journal of Cardiovascular Translational Research*; 3(1):37-48.

69. Chen JN, Fishman MC. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. *Development* 1996; 122(12):3809-3816.

70. Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP. Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes & Development* 1995; 9(13):1654-1666.

71. Eisenberg LM, Eisenberg CA. Wnt signal transduction and the formation of the myocardium. *Developmental Biology* 2006; 293(2):305-315.

72. de la Cruz MV, Sanchez Gomez C, Arteaga MM, Arguello C. Experimental study of the development of the truncus and the conus in the chick embryo. *Journal of Anatomy* 1977; 123(Pt 3):661-686.

73. Restivo A, Piacentini G, Placidi S, Saffirio C, Marino B. Cardiac outflow tract: a review of some embryogenetic aspects of the conotruncal region of the heart. *Anatomical Record* 2006; Part A, Discoveries in Molecular, Cellular, & Evolutionary Biology. 288(9):936-943.

74. Kelly RG, Brown NA, Buckingham ME. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Developmental Cell* 2001; 1(3):435-440.

75. Mjaatvedt CH, Nakaoka T, Moreno-Rodriguez R, Norris RA, Kern MJ, Eisenberg CA, Turner D, Markwald RR. The outflow tract of the heart is recruited from a novel heart-forming field. *Developmental Biology* 2001; 238(1):97-109.

76. Waldo KL, Kumiski DH, Wallis KT, Stadt HA, Hutson MR, Platt DH, Kirby ML. Conotruncal myocardium arises from a secondary heart field. *Development* 2001; 128(16):3179-3188.

77. Kirby M, Gale T, Stewart D. Neural crest cells contribute to normal aorticopulmonary septation. *Science* 1983; 220:1059-1061.

78. Dyer LA, Kirby ML. The role of secondary heart field in cardiac development. *Developmental Biology* 2009; 336(2):137-144.

79. Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Developmental Cell* 2003; 5(6):877-889.

80. Ward C, Stadt H, Hutson M, Kirby ML. Ablation of the secondary heart field leads to tetralogy of Fallot and pulmonary atresia. *Developmental Biology* 2005; 284(1):72-83.

81. Hutson MR, Kirby ML. Neural crest and cardiovascular development: a 20-year perspective. *Birth Defects Research Part C, Embryo Today: Reviews* 2003; 69(1):2-13.

82. Waldo KL, Lo CW, Kirby ML. Connexin 43 expression reflects neural crest patterns during cardiovascular development. *Developmental Biology* 1999; 208(2):307-323.

83. Stoller J, Epstein J. Cardiac neural crest. *Seminars in cell and developmental biology* 2005; 16:704-715.

84. Kaartinen V, Dudas M, Nagy A, Sridurongrit S, Lu MM, Epstein JA, Kaartinen V, Dudas M, Nagy A, Sridurongrit S, Lu MM, Epstein JA. Cardiac outflow tract defects in mice lacking ALK2 in neural crest cells. *Development* 2004; 131(14):3481-3490.

85. Brade T, Manner J, Kuhl M. The role of Wnt signalling in cardiac development and tissue remodelling in the mature heart. *Cardiovascular Research* 2006; 72(2):198-209.

86. Katoh M. WNT/PCP signalling pathway and human cancer (review). *Oncology Reports* 2005; 14(6):1583-1588.

87. Keller R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* 2002; 298:1950-1954.

88. Katanaev V, Ponzielli R, Semeriva M, Tomlinson A. Trimeric G protein dependent frizzled signalling in drosophila. *Cell* 2005; 120:111-122.

89. Wang Y, Nathans J, Wang Y, Nathans J. Tissue/planar cell polarity in vertebrates: new insights and new questions. *Development* 2007; 134(4):647-658.

90. Henderson DJ, Conway SJ, Greene ND, Gerrelli D, Murdoch JN, Anderson RH, Copp AJ. Cardiovascular defects associated with abnormalities in midline development in the Loop-tail mouse mutant. *Circulation Research* 2001; 89(1):6-12.

91. Curtin JA, Quint E, Tsipouri V, Arkell RM, Cattanach B, Copp AJ, Henderson DJ, Spurr N, Stanier P, Fisher EM, Nolan PM, Steel KP, Brown SD, Gray IC, Murdoch JN.

Mutation of Celsr1 disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse. *Current Biology* 2003; 13(13):1129-1133.

92. Saburi S, McNeill H. Organising cells into tissues: new roles for cell adhesion molecules in planar cell polarity. *Current Opinion in Cell Biology* 2005; 17(5):482-488.

93. Roszko I, Sawada A, Solnica-Krezel L. Regulation of convergence and extension movements during vertebrate gastrulation by the Wnt/PCP pathway. *Seminars in Cell & Developmental Biology* 2009; 20(8):986-997.

94. Liao G, Tao Q, Kofron M, Chen J-S, Schloemer A, Davis RJ, Hsieh J-C, Wylie C, Heasman J, Kuan C-Y. Jun NH2-terminal kinase (JNK) prevents nuclear beta-catenin accumulation and regulates axis formation in Xenopus embryos. *Proceedings of the National Academy of Sciences of the United States of America* 2006; 103(44):16313-16318.

95. Petersen CP, Reddien PW, Petersen CP, Reddien PW. Wnt signaling and the polarity of the primary body axis. *Cell* 2009; 139(6):1056-1068.

96. Tada M, Concha ML, Heisenberg CP. Non-canonical Wnt signalling and regulation of gastrulation movements. *Seminars in Cell & Developmental Biology* 2002; 13(3):251-260.

97. Tam PP, Loebel DA, Tanaka SS. Building the mouse gastrula: signals, asymmetry and lineages. *Current Opinion in Genetics & Development* 2006; 16(4):419-425.

98. Liu P, Wakamiya M, Shea MJ, Albrecht U, Behringer RR, Bradley A. Requirement for Wnt3 in vertebrate axis formation. *Nature Genetics* 1999; 22(4):361-365.

99. Huelsken J, Vogel R, Brinkmann V, Erdmann B, Birchmeier C, Birchmeier W. Requirement for beta-catenin in anterior-posterior axis formation in mice. *Journal of Cell Biology* 2000; 148(3):567-578.

100. Beddington RS, Robertson EJ. Axis development and early asymmetry in mammals. *Cell* 1999; 96(2):195-209.

101. Ibanes M, Izpisua Belmonte JC, Ibanes M, Izpisua Belmonte JC. Left-right axis determination. *Wiley interdisciplinary reviews Systems biology & medicine* 2009; 1(2):210-219.

102. Gessert S, Kuhl M. The multiple phases and faces of wnt signaling during cardiac differentiation and development. *Circulation Research*; 107(2):186-199.

103. Garriock RJ, D'Agostino SL, Pilcher KC, Krieg PA. Wnt11-R, a protein closely related to mammalian Wnt11, is required for heart morphogenesis in Xenopus. *Developmental Biology* 2005; 279(1):179-192.

104. Eisenberg CA, Eisenberg LM. WNT11 promotes cardiac tissue formation of early mesoderm. *Developmental Dynamics* 1999; 216(1):45-58.

105. Marvin MJ, Di Rocco G, Gardiner A, Bush SM, Lassar AB. Inhibition of Wnt activity induces heart formation from posterior mesoderm. *Genes & Development* 2001; 15(3):316-327.

106. Ueno S, Weidinger G, Osugi T, Kohn AD, Golob JL, Pabon L, Reinecke H, Moon RT, Murry CE. Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 2007; 104(23):9685-9690.

107. Pandur P, Lasche M, Eisenberg LM, Kuhl M. Wnt-11 activation of a non-canonical Wnt signalling pathway is required for cardiogenesis. *Nature* 2002; 418(6898):636-641.

108. Tai CC, Sala FG, Ford HR, Wang KS, Li C, Minoo P, Grikscheit TC, Bellusci S. Wnt5a knock-out mouse as a new model of anorectal malformation. *Journal of Surgical Research* 2009; 156(2):278-282.

109. Yamaguchi TP, Bradley A, McMahon AP, Jones S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* 1999; 126(6):1211-1223.

110. Nagy, II, Railo A, Rapila R, Hast T, Sormunen R, Tavi P, Vainio SJ. Wnt-11 signalling controls ventricular myocardium development by patterning N-cadherin and beta-catenin expression. *Cardiovascular Research*; 85(1):100-109.

111. Hurlstone AF, Haramis AP, Wienholds E, Begthel H, Korving J, Van Eeden F, Cuppen E, Zivkovic D, Plasterk RH, Clevers H. The Wnt/beta-catenin pathway regulates cardiac valve formation. *Nature* 2003; 425(6958):633-637.

112. Brault V, Moore R, Kutsch S, Ishibashi M, Rowitch D, McMahon AP. Inactivation of the beta-catenin gene by Wnt1-Cre mediated deletion results is dramatic brain malformation and failure of craniofacial development. *Development* 2001; 2001:1253-1264.

113. Hamblet NS, Lijam N, Ruiz-Lozano P, Wang J, Yang Y, Luo Z, Mei L, Chien KR, Sussman DJ, Wynshaw-Boris A. Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Development* 2002; 129(24):5827-5838.

114. Phillips HM, Murdoch JN, Chaudhry B, Copp AJ, Henderson DJ. Vangl2 acts via RhoA signalling to regulate polarized cell movements during development of the proximal outflow tract. *Circulation Research* 2005; 96(3):292-299.

115. Schleiffarth JR, Person AD, Martinsen BJ, Sukovich DJ, Neumann A, Baker CV, Lohr JL, Cornfield DN, Ekker SC, Petryk A. Wnt5a is required for cardiac outflow tract septation in mice. *Pediatric Research* 2007; 61(4):386-391.

116. Moeller H, Jenny A, Schaeffer H-J, Schwarz-Romond T, Mlodzik M, Hammerschmidt M, Birchmeier W. Diversin regulates heart formation and gastrulation movements in development. *Proceedings of the National Academy of Sciences of the United States of America* 2006; 103(43):15900-15905.

117. Fodde R, Eddelman W, Yang K, van Leeuwen C. A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Procedures National Academy of Sciences USA* 1994; 91:8969-8973.

118. Wang Y, Thekdi N, Smallwood PM, Macke JP, Nathans J. Frizzled-3 is required for the development of major fiber tracts in the rostral CNS. *Journal of Neuroscience* 2002; 22(19):8563-8573.

119. Ishikawa T, Tamai Y, Zorn AM, Yoshida H, Seldin MF, Nishikawa S, Taketo MM. Mouse Wnt receptor gene Fzd5 is essential for yolk sac and placental angiogenesis. *Development* 2001; 128(1):25-33.

120. Thumkeo D, Keel J, Ishizaki T, Hirose M, Nonomura K, Oshima H, Oshima M, Taketo MM, Narumiya S. Targeted disruption of the mouse rho-associated kinase 2 gene results in intrauterine growth retardation and fetal death. *Molecular & Cellular Biology* 2003; 23(14):5043-5055.

121. Zhao Z, Rivkees SA. Rho-associated kinases play an essential role in cardiac morphogenesis and cardiomyocyte proliferation. *Developmental Dynamics* 2003; 226(1):24-32.

122. Agustine K, Edison TL, Sadler TW. Antisense attenuation of Wnt-1 and Wnt3a expression in whole embryo culture reveals roles for these genes in craniofacial, spinal chord and cardiac morphogenesis. *Developmental Genetics* 1993; 14:500-520.

123. Riento K, Ridley AJ. Rocks: multifunctional kinases in cell behaviour. *Nature Reviews Molecular Cell Biology* 2003; 4(6):446-456.

124. The European Bioinformatics Institute and Genome Research limited. Ensembl. **www.ensembl.org2006-2007).**

125. Leung T, Chen XQ, Manser E, Lim L. The p160 RhoA-binding kinase ROKa is a member of a kinase family and is involved in the reorganization of the cytoskeleton *Molecular and Cellular Biology* 1996; 16(10):5313-5327.

126. Wei L, RoBerts W, Wang L, Yamada M, Zhan S, Zhao Z, Rivkees SA, Schwartz RJ, Imanaka-Yoshida K. Rho kinases play an obligatory role in vertebrate embryonic organogenesis. *Development* 2001; 128:1953-1962.

127. Kim G-H, Han J-K. JNK and ROKalpha function in the noncanonical Wnt/RhoA signalling pathway to regulate Xenopus convergent extension movements. *Developmental Dynamics* 2005; 232(4):958-968.

128. Zhu S, Liu L, Korzh V, Gong Z, Low BC. RhoA acts downstream of Wnt5 and Wnt11 to regulate convergence and extension movements by involving effectors Rho kinase and Diaphanous: use of zebrafish as an in vivo model for GTPase signalling. *Cellular Signalling* 2006; 18(3):359-372.

129. Verdier V, Guang Chao C, Settleman J. Rho-kinase regulates tissue morphogenesis via non-muscle myosin and LIM-kinase during Drosophila development. *BMC Developmental Biology* 2006; 6:38.

130. Kaarbo M, Crane DI, Murrell WG. RhoA is highly up-regulated in the process of early heart development of the chick and important for normal embryogenesis. *Developmental Dynamics* 2003; 227(1):35-47.

131. Zhao Z, Rivkees SA. Rho-associated kinases play a role in endocardial cell differentiation and migration. *Developmental Biology* 2004; 275(1):183-191.

132. Sakabe M, Ikeda K, Nakatani K, Kawada N, Imanaka-Yoshida K, Yoshida T, Yamagishi T, Nakajima Y. Rho kinases regulate endothelial invasion and migration during valvuloseptal endocardial cushion tissue formation. *Developmental Dynamics* 2006; 235(1):94-104.

133. Shimizu Y, Thumkeo D, Keel J, Ishizaki T, Oshima H, Oshima M, Noda Y, Matsumura F, Taketo MM, Narumiya S. ROCK-I regulates closure of the eyelids and ventral body wall by inducing assembly of actomyosin bundles. *Journal of Cell Biology* 2005; 168(6):941-953.

134. Narumiya S, Tanji M, Ishizaki T. Rho signaling, ROCK and mDia1, in transformation, metastasis and invasion. *Cancer & Metastasis Reviews* 2009; 28(1-2):65-76.

135. Henderson DJ, Ybot-Gonzalez P, Copp AJ. RhoB is expressed in migrating neural crest and endocardial cushions of the developing mouse embryo. *Mechanisms of Development* 2000; 95(1-2):211-214.

136. O'Brien R, Sonnhammer E. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research* 2005; 33:D476-D480.

137. Das G, Jenny A, Klein TJ, Eaton S, Mlodzik M. Diego interacts with Prickle and Strabismus/Van Gogh to localize planar cell polarity complexes. *Development* 2004; 131(18):4467-4476.

138. Schwarz-Romond T, Asbrand C, Bakkers J, Kuhl M, Schaeffer H-J, Huelsken J, Behrens J, Hammerschmidt M, Birchmeier W. The ankyrin repeat protein Diversin recruits Casein kinase epsilon to the beta-catenin degradation complex and acts in both canonical Wnt and Wnt/JNK signalling. *Genes & Development* 2002; 16(16):2073-2084.

139. Wang J, Hamblet NS, Mark S, Dickinson ME, Brinkman BC, Segil N, Fraser SE, Chen P, Wallingford JB, Wynshaw-Boris A. Dishevelled genes mediate a conserved mammalian PCP pathway to regulate convergent extension during neurulation. *Development* 2006; 133(9):1767-1778.

140. Pizzuti A, Novelli G, Mari A, Ratti A, Colosimo A, Amati F, Penso D, Sangiuolo F, Calabrese G, Palka G, Silani V, Gennarelli M. Human homologue sequences to the Drosophila dishevelled segment-polarity gene are deleted in the Digeorge syndrome. *American Journal of Human Genetics* 1996; 58:722-729.

141. Wallingford JB, Habas R. The developmental biology of Dishevelled: an enigmatic protein governing cell fate and cell polarity. *Development* 2005; 132(20):4421-4436.

142. Axelrod JD. Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. *Genes and Development* 2001; 15(10):1182-1187.

143. Habas R, Dawid IB. Dishevelled and Wnt signalling: is the nucleus the final frontier? *Journal of biology (Online)* 2005; 4(1):2.

144. Park M, Wu X, Golden K, Axelrod JD, Bodmer R. The wingless signaling pathway is directly involved in Drosophila heart development. *Developmental Biology* 1996; 177(1):104-116.

145. Itoh K, Sokol SY. Graded amounts of Xenopus dishevelled specify discrete anteroposterior cell fates in prospective ectoderm. *Mechanisms of Development* 1997; 61(1-2):113-125.

146. Ewald AJ, Peyrot SM, Tyszka MJ, Fraser SE, Wallingford JB. Regional requirements for Dishevelled signalling during Xenopus gastrulation: Separable effects on blastopore closure, mesendoderm internalization and archenteron formation. *Development* 2004; 131(24):6195-6209.

147. Wallingford JB, Rowning BA, Vogeli KM, Rothbacher U, Fraser SE, Harland RM. Dishevelled controls cell polarity during Xenopus gastrulation. *Nature* 2000; 405:81-85.

148. Lijam N, Paylor R, McDonald MP, Crawley JN, Deng CX, Herrup K, Stevens KE, Maccaferri G, McBain CJ, Sussman DJ, Wynshaw-Boris A. Social interaction and sensorimotor gating abnormalities in mice lacking Dvl1. *Cell* 1997; 90(5):895-905.

149. Oishi I, Suzuki H, Onishi N, Takada R, Kani S, Ohkawara B, Koshida I, Suzuki K, Yamada G, Schwabe GC, Mundlos S, Shibuya H, Takada S, Minami Y. The receptor tyrosine kinase Ror2 is involved in non-canonical Wnt5a/JNK signalling pathway. *Genes to Cells* 2003; 8(7):645-654.

150. Zhou W, Lin L, Majumdar A, Li X, Zhang X, Liu W, Etheridge L, Shi Y, Martin J, Van de Ven W, Kaartinen V, Wynshaw-Boris A, McMahon AP, Rosenfeld MG, Evans SM. Modulation of morphogenesis by noncanonical Wnt signaling requires ATF/CREB family-mediated transcriptional activation of TGFbeta2. *Nature Genetics* 2007; 39(10):1225-1234.

151. Keavney B. CHANGE study outline. Newcastle upon Tyne; 2006.

152. van der Velde ET, Vriend JW, Mannens MM, Uiterwaal CS, Brand R, Mulder BJ. CONCOR, an initiative towards a national registry and DNA-bank of patients with congenital heart disease in the Netherlands: rationale, design, and first results.[erratum appears in Eur J Epidemiol. 2005;20(10):885]. *European Journal of Epidemiology* 2005; 20(6):549-557.

153. Keavney B, Baker M. Association between common polymorphisms of the proopiomelanocortin gene and body fat distribution. *Diabetes* 2005; 54:2492-2496.

154. Health Protection Agency. European Collection of Cell cultures. **http://www.hpacultures.org.uk/products/dna/hrcdna/hrcdna.jsp2009).**

155. Rozen S, Skaletsky HJ. *Primer 3 on the www for general users and for biologist programmers*. Totowa: Humana Press; 2000.

156. Bonfield J, Beal K, Jordan m, Cheng Y, Staden R. The Staden Package Mini-manual. Medical Research Council; 2002.

157. The International Hapmap Consortium. The international hapmap project. *Nature* 2003; 426:789-796.

158. Ramensky V, Bork P, Sunyaev S. Human non-synonimous SNPs: server and survey. *Nucleic Acids Research* 2002; 30(17):3894-3900.

159. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Research* 2003; 31:3568-3571.

160. Nalla VK, Rogan PK. Automated splicing mutation analysis by information theory. *Human Mutation* 2005; 25:334-342.

161. National Genetics Reference Laboratory. Mutation scanning in complex genes using Meta-PCR and DNA sequencing. In: Department of health, ed.; 2004.

162. Ragoussis J, Gareth E, Kaur K, Colella S. Matrix-Assisted Laser Desorption/Ionisation, Time-of-Flight Mass Spectrometry in Genomics Research. *Plos Genetics* 2006; 7(2):920-929.

163. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19(1):149-150.

164. Sequenom Inc. IPLEX Gold application guide. 2007.

165. Sequenom Inc. RealSNP assay database. **https://www.realsnp.com/default.asp2009).**

166. Eden J. SEQUENOM tm MassARRAY/MassEXTEND SNP Typing Facility. **www.ncl.ac.uk/ihg/research/facilities/sequenom.htm.**

167. Heiman M. Web Cutter. **http://www.firstmarket.com/cutter/cut2.html2006).**

168. Palomino Doza J. Dvl2 mutational screening in tetralogy of Fallot. Newcastle upon Tyne: Newcastle University; 2006.

169. Applied Byosistems. *Allelic Discrimination Assays Getting Started Guide*: Applied Biosystems; 2005.

170. Kbiosciences. Kbiosciences. Services and techniques for Genomics and Drug discovery. **http://www.kbioscience.co.uk/.**

171. Abecasis G, Wiggington J. PEDSTATS: descriptive, statistics, graphics and quality assesment for gene maping data. *Bioinformatics* 2005; 21:3445-3447.

172. Abecasis G, Cardon L, Cookson W, Cherny S. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* 2002; 30:97-101.

173. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; 21(2):263-265.

174. Purcell S, Sham P, Daly MJ. Parental phenotypes in family-based association analysis. *American Journal of Human Genetics* 2005; 76(2):249-259.

175. Li M, Boehnke M, Abecasis GR. Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *American Journal of Human Genetics* 2006; 78(5):778-792.

176. Benjamini Y, Hochberg Y. Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics* 1997; 24(3):407-418.

177. Storey J. The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics* 2003; 31:2013-2035.

178. Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biology* 2003; 4(4):R25.

179. National Institutes of Health. Entrez Nucleotide database. **http://www.ncbi.nlm.nih.gov/nuccore/9155448?report=GenBank.**

180. Maizel JV, Jr., Lenk RP. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the National Academy of Sciences of the United States of America* 1981; 78(12):7665-7669.

181. Colorado State University. DNA dot. **http://www.vivo.colostate.edu/molkit/dnadot/index.html.**

182. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 2008; 9(4):299-306.

183. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology & Evolution* 2007; 24(8):1596-1599.

184. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology* 1990; 215(3):403-410.

185. Venables JP, Bourgeois CF, Dalgliesh C, Kister L, Stevenin J, Elliott DJ. Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Human Molecular Genetics* 2005; 14(16):2289-2303.

186. Qiagen. Qiaquick spin handbook. 2008.

187. Bourgeois CF, Popielarz M, Hildwein G, Stevenin J. Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Molecular & Cellular Biology* 1999; 19(11):7347-7356.

188. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *Journal of General Virology* 1977; 36(1):59-74.

189. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *American Journal of Human Genetics* 2002; 71(5):1227-1234.

190. Cardon LR, Abecasis GR. Using haplotype blocks to map human complex trait loci. *Trends in Genetics* 2003; 19(3):135-140.

191. Kryukov GV, Pennacchio LA, Sunyaev SR, Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *American Journal of Human Genetics* 2007; 80(4):727-739.

192. Softberry. NSITE Recognition of regulatory motifs. **http://linux1.softberry.com/berry.phtml?topic=nsite&group=programs&subgroup=promoter2010).**

193. Adler PN. Planar signaling and morphogenesis in Drosophila. *Developmental Cell* 2002; 2(5):525-535.

194. Ku M, Melton DA. Xwnt-11: a maternally expressed Xenopus wnt gene. *Development* 1993; 119(4):1161-1173.

195. Nakagawa O, Fujisawa K, Ishizaki T, Saito Y, Nakao K, Narumiya S. ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice. *FEBS Letters* 1996; 392(2):189-193.

196. Masazumi T, Masatake K. Noncanonical Wnt/PCP Signaling During Vertebrate Gastrulation. *Zebrafish* 2009; 6(1).

197. Axelrod JD. Interaction between wingless and notch signalling pathways mediated by dishevelled. *Science* 1996; 271(5257):1826-1832.

198. Boutros M, Paricion N, Strutt DI, Mlodzik M. Dishevelled activates JNK and discriminates between JNK pathways in planar polarity and wingless signaling. *Cell* 1998; 94(1):109-118.

199. Cunnington MS, Mayosi BM, Hall DH, Avery PJ, Farrall M, Vickers MA, Watkins H, Keavney B. Novel genetic variants linked to coronary artery disease by genome-wide association are not associated with carotid artery intima-media thickness or intermediate risk phenotypes. *Atherosclerosis* 2009; 203(1):41-44.

200. Gaukrodger N, Avery PJ, Keavney B, Gaukrodger N, Avery PJ, Keavney B. Plasma potassium level is associated with common genetic variation in the beta-subunit of the epithelial sodium channel. *American Journal of Physiology - Regulatory Integrative & Comparative Physiology* 2008; 294(3):R1068-1072.

201. Rahman T, Baker M, Hall DH, Avery PJ, Keavney B. Common genetic variation in the type A endothelin-1 receptor is associated with ambulatory blood pressure: a family study. *Journal of Human Hypertension* 2008; 22(4):282-288.

202. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; 305(5685):869-872.

203. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 1998; 281(5373):108-111.

204. Goldmuntz E, Geiger E, Benson DW. NKX2.5 mutations in patients with tetralogy of fallot. *Circulation* 2001; 104(21):2565-2568.

205. McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E. NKX2.5 Mutations in Patients with Congenital Heart Disease. *Journal of the American College of Cardiology* 2003; 42(9):1650-1655.

206. Eldadah ZA, Hamosh A, Biery NJ, Montgomery RA, Duke M, Elkins R, Dietz HC. Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Human Molecular Genetics* 2001; 10(2):163-169.

207. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 2003; 424(6947):443-447.

208. Wang D, Wolfgang S. Searching for polymorphisms that affect gene expression and mRNA processing: example ABCB1 (MDR1). *AAPS Journal* 2006; 8(3):E515-520.

209. Enjuanes A, Benavente Y, Bosch F, Guerrero I, Colomer D, Alvarez S, Reina O, Ardanaz MT, Jares P, Orad A, Pujana MA, Montserrat E, Campo E. Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia. *Cancer Research* 2008; 68(24):10178-10186.

210. Lida A, Emi M, Matsuoka R, Hiratsuka E, Okui K, Ohashi H, Inazawa J, Fukushima Y, Imai T, Nakamura Y. Identification of a gene disrupted by inv(11)(q13.5;q25) in a patient with left-right axis malformation. *Human Genetics* 2000; 106(3):277-287.

211. Yampolsky LY, Kondrashov FA, Kondrashov AS. Distribution of the strength of selection against amino acid replacements in human proteins. *Human Molecular Genetics* 2005; 14(21):3191-3201.

212. Eyre-Walker A, Woolfit M, Phelps T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 2006; 173(2):891-900.

213. The 1000 genomes project consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467:1061:1073.

214. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. *Nature Biotechnology* 2004; 22(5):535-546.

215. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD, Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 2003; 302(5653):2141-2144.

216. Montera M, Piaggio F, Marchese C, Gismondi V, Stella A, Resta N, Varesco L, Guanti G, Mareni C. A silent mutation in exon 14 of the APC gene is associated with exon skipping in a FAP family. *Journal of Medical Genetics* 2001; 38(12):863-867.

217. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* 2002; 3(4):285-298.

218. Vega A, Campos B, Bressac-De-Paillerets B, Bond PM, Janin N, Douglas FS, Baena M, Pericay C, Alonso C, Carracedo A, Baiget M, Diez O. The R71G BRCA1 is a founder Spanish mutation and leads to aberrant splicing of the transcript. *Human Mutation* 2001; 17(6):520-521.

219. Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1.[Erratum appears in Hum Mol Genet 2000 Mar 1;9(4):659]. *Human Molecular Genetics* 2000; 9(2):237-247.

220. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA, Concannon P. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *American Journal of Human Genetics* 1999; 64(6):1617-1631.

221. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends in Genetics* 2001; 17(9):502-510.

222. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nature Genetics* 2007; 39(1):17-23.

223. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, Halperin RF, Marlowe L, Kaleem M, Walker DG, Ravid R, Heward CB, Rogers J, Papassotiropoulos A, Reiman EM, Hardy J, Stephan DA. A high-density whole-genome association study reveals that APOE is the major susceptibility

gene for sporadic late-onset Alzheimer's disease. *Journal of Clinical Psychiatry* 2007; 68(4):613-618.

224. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC, Alexander GE, Caselli RJ, Kukull W, Papassotiropoulos A, Stephan DA. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 2007; 54(5):713-720.

225. Naj AC, Beecham GW, Martin ER, Gallins PJ, Powell EH, Konidari I, Whitehead PL, Cai G, Haroutunian V, Scott WK, Vance JM, Slifer MA, Gwirtsman HE, Gilbert JR, Haines JL, Buxbaum JD, Pericak-Vance M. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. *Plos Genetics* 2010; 6(9).

226. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; 316(5830):1488-1491.

227. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007; 316(5830):1491-1493.

228. Roberts R, Wells GA, Stewart AF, Dandona S, Chen L, Roberts R, Wells GA, Stewart AFR, Dandona S, Chen L. The genome-wide association study--a new era for common polygenic disorders. *Journal of Cardiovascular Translational Research*; 3(3):173-182.

229. Lee C, Scherer SW, Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Reviews in Molecular Medicine*; 12:e8.

230. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nature Genetics* 2004; 36(9):949-951.

231. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC, Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AWC, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers Y-H, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. *Plos Biology* 2007; 5(10):e254.

232. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla JA, Baralle D, Collins A, Mercer C, Norga K, de Ravel T, Devriendt K, Bongers EM, de Leeuw N, Reardon W, Gimelli S, Bena F, Hennekam RC, Male A, Gaunt L, Clayton-Smith J, Simonic I, Park SM, Mehta SG, Nik-Zainal S, Woods CG, Firth HV, Parkin G, Fichera M, Reitano S, Lo Giudice M, Li KE, Casuga I, Broomer A, Conrad B, Schwerzmann M, Gallati S, Striano P, Coppola A, Tolmie JL, Tobias ES, Lilley C, Armengol L, Spysschaert Y, Verloo P, De Coene A, Goossens L, Mortier G, Speleman F, van Binsbergen E, Nelen MR, Hochstenbach R, Poot M, Gallagher L, Gill M, McClellan J, King MC, Regan R, Skinner C, Stevenson RE, Antonarakis SE, Chen C, Estivill X, Menten B, Gimelli G, Gribble S, Schwartz S, Sutcliffe JS, Walsh T, Knight SJ, Sebat J, Romano C, Schwartz CE, Veltman JA, de Vries BB, Vermeesch JR, Barber JC, Willatt L, Tassabehji M, Eichler EE. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *New England journal of Medicine* 2008; 359(16):1685-1699.

233. Hemminki K, Lorenzo Bermejo J, Forsti A. The balance between heritable and environmental aetiology of human disease. *Nature Reviews Genetics* 2006; 7(12):958-965.

234. Willett WC. Balancing life-style and genomics research for disease prevention. *Science* 2002; 296(5568):695-698.

235. Diabetes UK. Diabetes in the UK 2010: Key statistics. **http://www.diabetes.org.uk/Documents/Reports/Diabetes_in_the_UK_2010.pdf.**

236. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostr, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H,

Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, rgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* 2008; 40(5):638-645.

237. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*; 11(6):415-425.

238. Kircher M, Kelso J, Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. *Bioessays*; 32(6):524-536.

239. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; 461(7261):272-276.

240. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*; 42(1):30-35.

APPENDIX A

The Change Study Questionnaire

Questionnaire to be completed by nurse with mother,

Family code number:..... Individual code number.....

Mother's date of birth:.....

Ethnic group: .....

Occupation*(s)* - before pregnancy:.....

.....

Occupation - during early part of pregnancy: .....

Partner's occupation:.....

Ethnic group: .....

Family tree:

XXX570

The Change Study Questionnaire

Complete this sheet for each affected child

Does your child have any other medical problems?.....

.....

.....

We are interested in the month before you became pregnant and the first three months of your pregnancy. The following information helps us to identify when your pregnancy started.

Child's date of birth:.....

Gestation at birth: .....

Was your pregnancy planned? YES/NO

Calculated: LMP: .....

Hence one month prior to LMP: .....

How far into pregnancy were you when you realised you were pregnant?: .....

1. Did you have Diabetes prior to pregnancy YES/NO

   Did you have Epilepsy prior to pregnancy YES/NO

   Did you have Diabetes during to pregnancy YES/NO

   Did you have Epilepsy during pregnancy YES/NO

   Any other medical problems prior to pregnancy

   Month of diagnosis/Gestation..... Condition .....

   ....

   ....

   ....

   ....

   ....

The Change Study Questionnaire

2. Were you taking any regular medication in the month before you became pregnant or the first three months of pregnancy?

   List *(medication and dose)*: .....
   .....
   .....
   .....
   .....
   .....

   If taking medication, including prescribed folate obtain consent to contact GP/hospital for information on dosage if necessary.

   Consent given: YES/NO/not applicable

3. Did you smoke in the month before you became pregnant or the first three months of the pregnancy?

   YES/NO Number of cigarettes per day:.....

4. How much alcohol did you drink per week before you knew you were pregnant?

   Conversion to units/week:.....

   How much alcohol did you drink per week after you knew you were pregnant?

   Conversion to units/week:.....

5. Were you taking folic acid before you knew you were pregnant?

   NO
   YES - prescribed - YES/NO

   Did you start taking folic acid when you found out you were pregnant?

   NO
   YES - prescribed - YES/NO

6. Were you taking iron tablets before you knew you were pregnant?

   NO
   YES - prescribed - YES/NO .....Brand

   Did you start taking iron tablets when you found out you were pregnant?

   NO
   YES - prescribed - YES/NO .....

7. Were you taking a multivitamin supplement before you knew you were pregnant?

   NO/YES

   Did you start taking a multivitamin supplement when you found out you were pregnant?

   NO/YES Brand: .....

8. Dietary factors in the month before pregnancy or the first trimester:

   Was your diet vegetarian? YES/NO

   Was your diet vegan? YES/NO

   Were you dieting to lose weight? YES/NO

9. Do you have any current medical problems?:

   Diabetes YES/NO .....
   Epilepsy YES/NO .....
   Other: .....

10. Are you taking any prescribed medication at the moment?

    YES/NO If yes list below:
    .....
    .....
    .....

11. Are you currently taking any vitamin supplements?

    YES/NO If yes list below:
    .....
    .....
    .....

12. ARE YOU Pre- menopausal
    or Post- menopausal
    Pregnant

Delete the ones which do NOT apply

Do you have any current medical problems? .....
.....
.....
.....
.....
.....
.....
.....