

**A GENETIC INVESTIGATION OF THE BONE  
MORPHOGENETIC PROTEIN SIGNALING  
PATHWAY IN CONGENITAL  
CARDIOVASCULAR MALFORMATION**

**Huay Lin Tan**

**Thesis submitted in partial fulfillment of the requirements for the  
degree of Doctor of Philosophy**

**Newcastle University  
Faculty of Medical Sciences  
Institute of Human Genetics  
December 2010**

# Abstract

## Background

Congenital cardiovascular malformations (CVM) affects approximately 7 in every 1000 live births and exhibits familial predisposition but the specific genetic factors involved are unknown. It was hypothesised that rare variants in genes of the bone morphogenetic protein (BMP) signaling pathway, which is known to play a key role in cardiac development, could influence the risk of CVM. Therefore, five genes of the BMP signaling pathway were surveyed for novel variants predisposing to CVM risk.

## Methods and Results

The exonic, splice site and untranslated regions of the *BMP2*, *BMP4*, *BMPRIA*, *BMPR2*, and *SMAD6* genes were sequenced in 90 unrelated Caucasian cases of CVM. One non-synonymous variant (p.C484F) in the MH2 domain of *SMAD6* was identified with predicted impaired protein function. Sequencing of the MH2 domain of *SMAD6* in an additional 348 cases showed two further non-synonymous variants, one of which (p.P415L) was also predicted to affect protein function. Both non-synonymous variants were absent in 1000 controls. *SMAD6* is an intracellular inhibitor of BMP signaling and functional effects of both variants were investigated using BRE-luciferase transcriptional reporter and alkaline phosphatase assays. Both *SMAD6* variants had significantly ( $p < 0.05$ ) lower activity than wild-type *SMAD6* in inhibiting BMP signaling in the BRE-luciferase assay. In addition, the p.C484F variant had a significantly ( $p < 0.05$ ) lower capacity to inhibit alkaline phosphatase generation in response to BMP signaling. This was consistent with the phenotype of the patient carrying the p.C484F variant, which featured aortic ossification with coarctation.

## Conclusions

Inadequate inhibition of BMP signaling due to genetic variation in *SMAD6* can be a cause of CVM in man. The potential significant *SMAD6* variants were identified in approximately 1% of the CVM patients or 6% in patients with coarctation and/or aortic stenosis. These findings show that *SMAD6* mutations in the MH2 domain may be important in abnormal cardiac development.

## Acknowledgements

I would like to thank my supervisors Bernard Keavney and Judith Goodship for their support and guidance throughout my study in Newcastle University. Thanks to Helen Arthur for providing invaluable support on the *in vitro* functional analysis of SMAD6. The constructs and help with the techniques were provided by Peter ten Dijke of Department of Molecular Cell Biology and Centre for Biomedical Genetics, Leiden University Medical Centre, Netherlands. Peter Avery provided invaluable statistical support. I also thank the Malaysian government, British Heart Foundation and EU Heart Repair for funding my studies.

I am grateful to Ana Topf for guiding me in the lab and being a translator. Thanks to Elise Glen for helping me with my functional studies; Darroch Hall for genotyping in control population; Raf Hussain, James Eden and Helen Griffin for providing support with sequencing and Sequenom assays and Caroline Daghish for introducing me to cloning and cell culturing. I appreciate the help of many people at the Institute of Human Genetics, Newcastle University who have helped me during my studies and the discussions in the coffee room.

I would like to thank my family and friends for their encouragement, for their moral support and for always being there for me. Above all I thank God for giving me the strength to complete my study.

## **Declaration**

I, Huay Lin Tan, declare that no portion of the work referred to in this thesis has been submitted in support of another degree or qualification at this or any other institute of learning.

I personally carried out all the work to develop and optimize the laboratory assays used in this project. I received assistance from colleagues in DNA extraction, MLPA, SNP genotyping and alkaline phosphatase assay, though I made substantial contributions to SNP genotyping and alkaline phosphatase assay. The statistical analysis was carried out by myself with the assistance of my supervisors and Dr. Peter Avery.

# Table of Contents

Abstract	i
Acknowledgements	ii
Declaration	iii
Table of Contents	iv
List of Figures	ix
List of Tables	xi
Abbreviations	xii
<b>1 Introduction</b>	<b>1</b>
1.1 Heart development	1
1.1.1 Heart tube formation – cellular sources	1
1.1.2 Looping of the heart	4
1.1.3 Cardiac neural crest cells	6
1.1.4 Anterior and secondary heart field	7
1.1.5 Chamber formation	9
1.1.6 Valve development	10
1.1.7 Aortic arch derivatives	12
1.2 Congenital cardiovascular malformation	14
1.2.1 Epidemiology of cardiovascular malformation	14
1.2.2 Non-inherited risk factors for cardiovascular malformation	22
1.2.3 Genetic basis for cardiovascular malformation	26
1.2.3.1 “Sporadic” CVM	26
1.2.3.2 CVM as part of particular genetic syndromes	28
1.2.3.3 Non-syndromic Mendelian families	31
1.3 Bone Morphogenetic Protein signaling pathway	32
1.3.1 BMP signaling during heart formation	32
1.3.2 <i>BMP2</i>	36
1.3.3 <i>BMP4</i>	39
1.3.4 <i>BMPRIA</i>	42
1.3.5 <i>BMPR2</i>	44
1.3.6 <i>SMAD6</i>	46
1.4 RNA processing	49
1.5 Mutation nomenclature	51

1.6	Aims and objectives of the study	52
<b>2</b>	<b>Materials and Methods</b>	<b>54</b>
2.1	Sample collection	54
2.1.1	CVM cases	54
2.1.2	Control populations	55
2.2	Screening of candidate genes	56
2.2.1	Sequencing primers	56
2.2.2	Polymerase Chain Reactions (PCR)	65
2.2.3	Gel electrophoresis	68
2.2.4	PCR clean up	68
2.2.5	Sequencing reactions	68
2.2.6	Sequence analysis	69
2.2.7	Identifying variants in parents	69
2.2.8	Genotyping rare variants	69
	2.2.8.1 Sequenom genotyping	70
	2.2.8.2 Taqman genotyping	74
2.3	<i>In silico</i> analysis of novel rare variants	75
2.4	Investigation of the effects of novel rare variants on splicing	77
2.4.1	Experimental design	77
2.4.2	Media and reagents	81
2.4.3	Cloning	81
	2.4.3.1 Primer designs and PCR	81
	2.4.3.2 Restriction enzyme digestions	84
	2.4.3.3 Dephosphorylation of vector	84
	2.4.3.4 Ligation	84
	2.4.3.5 Transformation of competent cells	84
	2.4.3.6 Screening PCR	85
	2.4.3.7 Sequencing of insertions	85
2.4.4	Transfection of HEK293 cells	86
2.4.5	RNA Extraction	86
2.4.6	Reverse Transcription PCR	87
2.5	Luciferase assay of SMAD6 variants	88
2.5.1	Experimental design	88
2.5.2	Cell lines and tissue culture	92

2.5.3	Constructs	92
2.5.4	Site-directed mutagenesis	96
2.5.5	Sequencing of constructs	98
2.5.6	Transfection of C2C12 cells	98
2.5.7	Measurement of luciferase activity	98
2.5.8	Immunoblotting	99
2.6	Alkaline phosphatase assay of SMAD6 variants	101
2.6.1	Experimental design	101
2.6.2	Transfection and stimulation of C2C12 cells with BMP ligands	104
2.6.3	Transfection of C2C12 cells with c.a.BMPRI1A	104
2.6.4	Measurement of alkaline phosphatase activity	104
2.7	Statistical analysis	106
<b>3</b>	<b>Results of mutation screening and <i>in vitro</i> splicing studies</b>	<b>107</b>
3.1	Overview	107
3.2	Phenotypes of the analyzed CVM patient cohort	108
3.3	<i>BMP2</i>	109
3.4	<i>BMP4</i>	112
3.4.1	Sequencing results	112
3.4.2	Investigation of effects on splicing of <i>BMP4</i>	117
3.4.2.1	<i>In silico</i> prediction	117
3.4.2.2	Minigene assays for <i>BMP4</i> variants	122
3.4.2.2.1	Novel variant c.-295G>A	122
3.4.2.2.2	Novel variant c.-136C>G	123
3.4.2.2.3	Novel variant c.-87C>T	127
3.5	<i>SMAD6</i>	129
3.6	<i>BMPRI1A</i>	138
3.7	<i>BMPRI2</i>	140
3.8	Summary	144
<b>4</b>	<b><i>In vitro</i> functional analysis of novel SMAD6 variants</b>	<b>145</b>
4.1	Overview	145
4.2	Optimization of transfection	146
4.3	Inhibitory effects of SMAD6 variants on BMP-Responsive element reporter	148

4.4	Inhibitory effects of SMAD6 variants on alkaline phosphatase activity	158
4.4.1	Inhibitory effects of SMAD6 variants with BMP ligands	158
4.4.2	Assessment of inhibitory effects of SMAD6 variants on alkaline phosphatase activity with c.a.BMPRI1A	163
4.5	Summary	165
<b>5</b>	<b>Discussion</b>	<b>166</b>
5.1	Overview	166
5.2	The <i>SMAD6</i> gene and CVM	169
5.3	The <i>BMP2</i> gene and CVM	174
5.4	The <i>BMP4</i> gene and CVM	177
5.5	The <i>BMPRI1A</i> gene and CVM	183
5.6	The <i>BMPRI2</i> gene and CVM	185
5.7	Possible mechanisms whereby SMAD6 variants identified in this study affect BMP signaling and CVM	189
5.7.1	BMPRI1A binding	189
5.7.2	Diminished competition with co-Smads	190
5.7.3	Impaired cooperative inhibition with Smurf1	191
5.7.4	Nuclear actions	191
5.8	This study in the context of other resequencing studies of CVM	193
5.8.1	Other genes in the BMP signaling pathway	193
5.8.2	Genes in other signaling pathways	195
5.9	Common polymorphisms and CVM	198
5.9.1	Effect sizes of rare and common variants	198
5.9.2	Previous studies of common genetic variants and CVM	201
5.9.3	GWAS studies of complex diseases	202
5.9.4	LFIP variants and complex diseases	203
5.10	Copy number variants (CNVs) and CVM	204
5.11	Strengths and limitations of this work	207
5.12	Future experiments	209
5.12.1	Characterization of SMAD6 mechanisms	209
5.12.2	Sequencing SMAD6 in larger samples	209
5.12.3	Further candidate gene sequencing	209
5.12.4	Common variants	210
5.12.5	CNV studies	211

5.13	Conclusions	212
<b>6</b>	<b>References</b>	<b>214</b>

## List of Figures

Figure 1.1	A brief schematic overview of the mouse heart development	3
Figure 1.2	Diagrams showing the formation of heart tube	5
Figure 1.3	Anterior and secondary heart fields	8
Figure 1.4	Heart valve development	11
Figure 1.5	Diagrams of the aortic arch system	13
Figure 1.6	BMP signaling pathway	35
Figure 1.6	The process of RNA splicing	50
Figure 1.7	Experimental process	53
Figure 2.1	<i>BMP2</i> Structure and Sequencing Primer Location	57
Figure 2.2	<i>BMP4</i> Structure and Sequencing Primer Location	58
Figure 2.3	<i>SMAD6</i> Structure and Sequencing Primer Location	59
Figure 2.4	<i>BMPRI1A</i> Structure and Sequencing Primer Location	60
Figure 2.5	<i>BMPRI2</i> Structure and Sequencing Primer Location	61
Figure 2.6	Map of pXJ41 vector	78
Figure 2.7	Overview of in vivo splicing analysis with the minigene approach	79
Figure 2.8	Example of the splicing pattern of TLE4 pre-mRNA	80
Figure 2.9	BMP signaling in C2C12 cells for luciferase assay	90
Figure 2.10	BRE-Luc inserted at <i>NheI</i> site in pGL3-basic vector circle map	93
Figure 2.11	c.a.BMPRI1A with HA tag inserted at <i>EcoRI</i> and <i>XhoI</i> sites in pcDNA3	94
Figure 2.12	SMAD6C wild type with FLAG tag inserted at <i>BamHI</i> and <i>XhoI</i> sites in pcDNA3	95
Figure 2.13	BMP signaling in C2C12 cells stimulated with BMP ligands for alkaline phosphatase assay	102
Figure 2.14	BMP signaling in C2C12 cells transfected with c.a.BMPRI1A for alkaline phosphatase assay	103
Figure 3.1	<i>BMP2</i> – Location of genetic variants	111
Figure 3.2	<i>BMP4</i> – Location of genetic variants	114
Figure 3.3	<i>BMP4</i> – Evolutionary conservation of the variant p.Gys174Ser (c.520G>A)	116
Figure 3.4	<i>BMP4</i> c.-295G>A <i>in silico</i> prediction	120
Figure 3.5	<i>BMP4</i> c.-136C>G <i>in silico</i> prediction	120
Figure 3.6	<i>BMP4</i> c.-87C>T <i>in silico</i> prediction	121

Figure 3.7	PCR screening of Ampicillin resistant colonies	124
Figure 3.8	Determination of fragment orientation	125
Figure 3.9	Splicing patterns of the variants	126
Figure 3.10	Amplified fragments of variant BMP4 c.-87>T using different primer pairs	128
Figure 3.11	<i>SMAD6</i> – Location of genetic variants	131
Figure 3.12	Predicted crystal structure of SMAD6 p.Cys484Phe (c.1244C>T)	133
Figure 3.13	SMAD6 – Evolutionary conservation of the variants	136
Figure 3.14	SMAD6 – Predicted miRNA binding site	137
Figure 3.15	<i>BMPRI1A</i> – Location of the genetic variant	139
Figure 3.16	<i>BMPRI2</i> – Location of the genetic variants	142
Figure 3.17	<i>BMPRI2</i> – Evolutionary conservation of the variant	143
Figure 4.1	Optimization of the ratio of transfection reagent to construct DNA	147
Figure 4.2	Optimization of the amount of construct DNA	147
Figure 4.3	Luciferase activity in C2C12 cells transfected with c.a.BMPRI1A and wild type SMAD6C	149
Figure 4.4	Overall SMAD6C variants inhibitory effect of three replicates and immunoblot images	150
Figure 4.5	Effect of a 50:50 mixture of SMAD6C p.Cys484Phe (c.1451G>T) variant with SMAD6C wild type constructs	153
Figure 4.6	Immunoblot images of phospho-Smad1/5/8 and Smad1	155
Figure 4.7	Immunoblot images of phospho-Smad1/5/8 and Smad1 treated with ExoSAP and cultured in different concentration of FBS	156
Figure 4.8	ALP activity of the cell lysates that were stimulated with ligands and cultured for 48 hours	159
Figure 4.9	ALP activity of the cell lysates that were stimulated with ligands and were cultured for 60 hours	160
Figure 4.10	Fluorescent microscope image of GFP transfected cells	162
Figure 4.11	ALP activities in the cell lysates that were transfected with c.a.BMPRI1A	164
Figure 5.1	<i>SPINK1</i> gene haplotypes (A) and <i>SPINK1</i> minigenes (B)	179
Figure 5.2	Location of the mutations in BMPRI2 identified by Roberts et al., 2004	186
Figure 5.3	The relationship between genetic effect size and allele frequency	200

## List of Tables

Table 1.1	Developmental timeline of normal human heart embryology	2
Table 1.2	Mean incidence of CVM per million live births	21
Table 1.3	Non-inherited risk factors for CVM	23
Table 1.4	Genetic syndromes associated with CVM	30
Table 1.5	Genes associated with non-syndromic CVM	31
Table 2.1	Primers used for sequencing	62
Table 2.2	PCR Conditions	66
Table 2.3	Primers used for Sequenom assay	73
Table 2.4	Primers and probes for TaqMan assay	74
Table 2.5	Cloning primers sequences and the combination of primer pairs	83
Table 2.6	Combination of constructs for luciferase assay	91
Table 2.7	Primers and reaction conditions used for the site-directed mutagenesis	97
Table 3.1	Phenotypes of the analysed CVM patients	108
Table 3.2	<i>BMP2</i> – Genetic variants	110
Table 3.3	<i>BMP4</i> – Genetic variants	113
Table 3.4	<i>BMP4</i> – Potential effect of variation on splicing	118
Table 3.5	<i>SMAD6</i> – Genetic variants	130
Table 3.6	MH2 domain of <i>SMAD6</i> – Genetic variants	134
Table 3.7	<i>BMPRIA</i> – Genetic variant	138
Table 3.8	<i>BMPR2</i> – Genetic variants	141
Table 5.1	Thirty-five previously unreported variants identified in the candidate genes	168

## Abbreviations

AA	Aortic Arch
AHF	Anterior Heart Field
ALP	Alkaline Phosphatase
ANOVA	Analysis of Variance
AoS	Aortic Sac
AS	Aortic Stenosis
ASD	Atrial Septal Defect
AV	Atrioventricular
AVC	Atrioventricular Canal
AVSD	Atrioventricular Septal Defect
BAV	Bicuspid Aortic Valve
BMP	Bone Morphogenetic Protein
bp	base pairs
BPV	Bicuspid Pulmonary Valve
BRE-Luc	BMP-responsive element luciferase reporter
c.a.	constitutively active
CGH	Comparative Genomic Hybridization
CHANGE	Congenital Hearts: A National Gene/Environment Study
CI	Confidence interval
CMV	Cytomegalovirus
CNV	Copy Number Variation
CoA	Coarctation of the Aorta
CVM	Cardiovascular Malformation
DMEM	Dulbecco's Modified Eagle Medium
(d)dNTPs	(Di)Deoxynucleotide Triphosphates
DORV	Double Outlet Right Ventricle
DOLV	Double Outlet Left Ventricle
EC	Endocardial Cushion
EDTA	EthyleneDiamineTetraAcetic Acid
EMT	Epithelial-Mesenchymal Transformation
ESE	Exonic Splice Enhancer
ESS	Exonic Splice Silencer
Exo-SAP	Exonuclease I and Shrimp Alkaline Phosphatase

FBS	Fetal Bovine Serum
FGF	Fibroblast Growth Factor
GFP	Green Fluorescent Protein
GWAS	Genome Wide Association Studies
hCMV	human Cytomegalovirus
HH	Hamburger-Hamilton
HLH	Hypoplastic Left Heart
HRP	Horseradish Peroxidase
IAA	Interrupted Aortic Arch
IFT	Inflow Tract
ISE	Intronic Splice Enhancer
ISS	Intronic Splice Silencer
IVC	Inferior Vena Cava
IVS	Interventricular Septum
LB	Lysogeny Broth
MAD	Mothers Against Decapentaplegic
MAF	Minor Allele Frequency
MALTI-TOF	Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight
MCS	Multiple Cloning Site
MES	2-(N-morpholino)ethanesulfonic acid
MH	MAD homology
miRNA	MicroRNA
MLPA	Multiplex Ligation-dependent Probe Amplification
NCC	Neural Crest Cells
NCCGP	North Cumbria Community Genetics Project
OFT	Outflow Tract
PA	Pulmonary Artery
PAA	Pharyngeal Arch Artery
PCR	Polymerase Chain Reaction
PDA	Patent Ductus Arteriosus
PFO	Patent Foramen Ovale
PH	Pulmonary Hypoplasia
PLA	Primitive Left Atrium
PPS	Peripheral Pulmonary Stenosis
PRA	Primitive Right Atrium

PS	Pulmonary Stenosis
PTA	Persistent Truncus Arteriosus
PT	Pulmonary Trunk
PV	Pulmonary Vein
RT-PCR	Reverse Transcription-PCR
RV	Right Ventricle
SBE	Smad binding elements
SDS	Sodium Dodecylsulphate
SHF	Secondary Heart Field
SMAD	Small Mother Against Decapentaplegic)
SNP	Single Nucleotide Polymorphism
SVC	Superior Vena Cava
TAE	Tris-acetate-EDTA
TAPVR	Total anomalous pulmonary venous return
TGA	Transposition of the Great Arteries
TGF- $\beta$	Transforming Growth Factor- $\beta$
TOF	Tetralogy of Fallot
VSD	Ventricular Septal Defect
wt	wild type

# **Chapter 1      Introduction**

## **1.1    Heart development**

The heart is the first organ to function in the development of the vertebrate embryo. Heart development is a continuous complex remodelling process that begins from the formation of the heart tube during the third week of gestation in man, followed by looping of the heart, chamber formation and valve development (Table 1.1). This process is induced by several interacting molecular signaling pathways that drive various cell populations to sculpt the heart. Formation of the heart is also governed by anterior-posterior (A-P), dorsal-ventral (D-V) and left-right (L-R) polarity.

Mouse has become the most popular animal model for understanding the normal or abnormal cardiac development. This is because the mouse and human heart are anatomically relatively similar throughout development apart from the differences in sizes and heart rate (Wessels and Sedmera 2003). Eventhough avians and mammals have four-chambered heart there are a few anatomical differences between both species. In birds, the sinus venosus is still present and the aortic arch is developed to the right whereas in mammals, the sinus venosus is almost gone and the aortic arch is developed to the left.

In this section, the major processes in heart formation and key cell populations involved is discussed briefly. The morphogenesis of the heart is summarised in Figure 1.1.

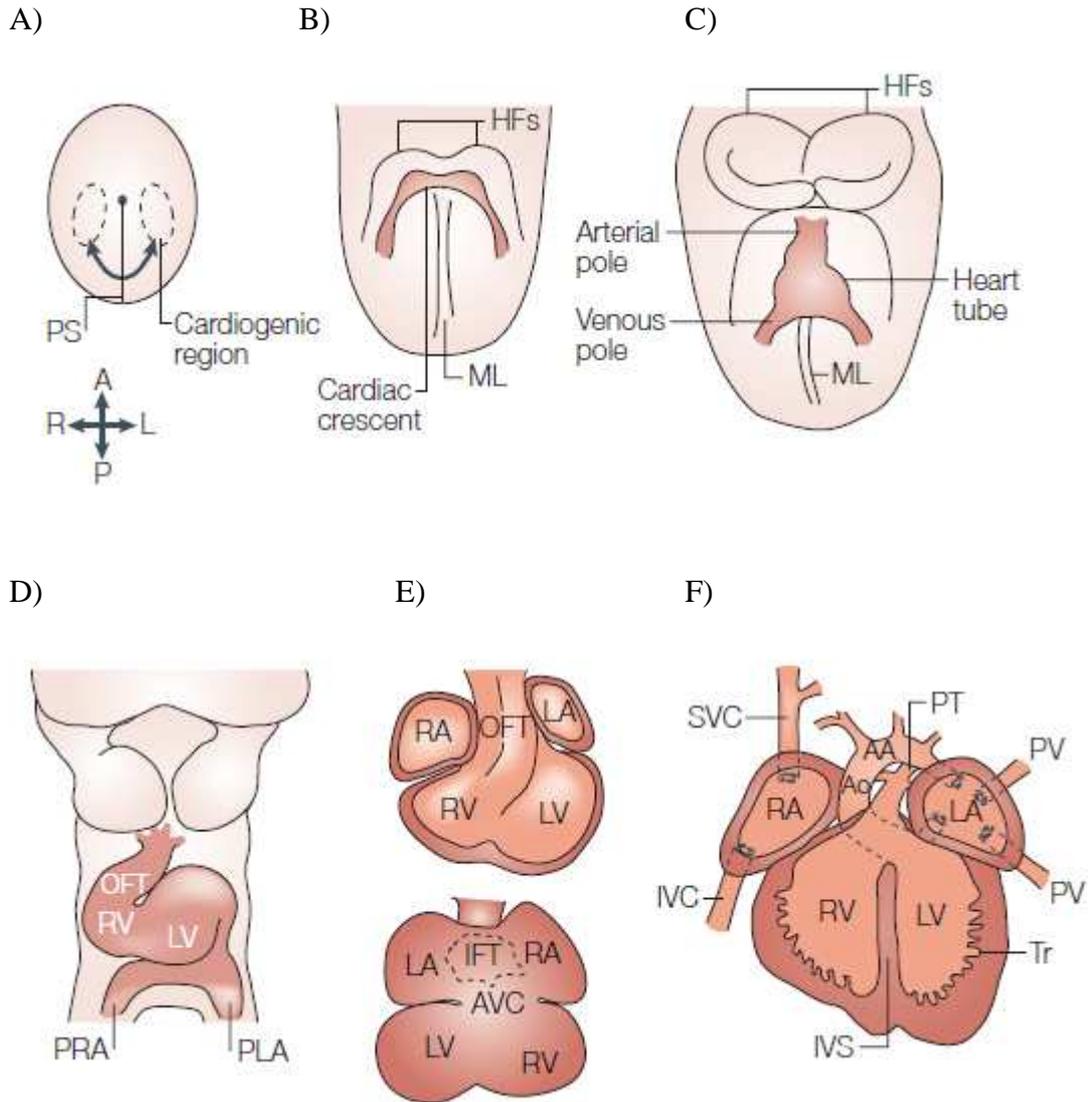
### **1.1.1    Heart tube formation – cellular sources**

Cardiac progenitor cells are located on both sides of the primitive streak and caudal to the Hensen's node in the the avian epiblast (Lopez-Sanchez et al. 2001). The cardiac precursors in the mouse are similar to chick, they are located laterally of the midline at E6.5 (Tam et al. 1997). These cells migrate in an anterior-lateral fashion to the lateral plate mesoderm where the formation of bilateral cardiogenic fields occurs (Yang et al. 2002). In the chick, the bilateral heart-forming regions are two separate entities (Colas et al. 2000), but in the mouse, the cardiogenic fields merge in the midline, forming the cardiac crescent (Cai et al. 2003). Formation of the heart tube begins when the

<b>Human development (days)</b>	<b>Developmental process</b>
13-14	Primitive streak formation
15-17	Formation of the three primary germ layers: ectoderm, mesoderm, and endoderm; primitive streak cells migrate to an anterior and lateral position to form the bilateral primary heart field
17-18	Lateral plate mesoderm splits into the somatopleuric mesoderm and splanchnopleuric mesoderm; splanchnopleuric mesoderm contains the myocardial and endocardial cardiogenic precursors in the regions of the primary heart field
18-26	Neurulation (formation of the neural tube)
20	Cephalocaudal and lateral folding brings the bilateral endocardial tubes into the ventral midline of the embryo Aortic arch is forming
21-22	Heart tube fusion
22	Heart tube begins to beat Aortic arches I and II are forming
24	Heart looping and the accretion of cells from the primary and secondary heart fields; neural crest migration starts Atria are beginning to bulge Right and left ventricles act like two pumps in series Outflow tract is distinguishable from right ventricle
Late fourth week	Sinus venosus is becoming incorporated into right atrium Endocardial cushions appear Early septum I appears between left and right atria Muscular interventricular septum is forming Truncocoanal ridges are forming Aortic arch I is regressing, aortic arch III is formed, aortic arch IV is forming
Early fifth week	Cardiac neural crest migrates through the aortic arches and enters the outflow tract of the heart Endocardial cushions are coming together; forming right and left atrioventricular canals Further growth of interatrial septum I and muscular interventricular septum occurs Truncus arteriosus is dividing into aorta and pulmonary artery Atrioventricular bundle is forming; possible neurogenic control of heart beat Pulmonary veins are becoming incorporated into left atrium Aortic arches I and II have regressed, aortic arches III and IV have formed aortic arch VI is forming Conduction system forms
Late fifth to early sixth week	Endocardial cushions fuse Interatrial foramen II is forming Interatrial septum I is almost contacting endocardial cushion Membranous part of interventricular septum starts to form Semilunar valves begin to form
Late sixth week	Interatrial foramen II is large Interatrial septum II starts to form Atrioventricular valves and papillary muscles are forming Interventricular septum is almost complete Coronary circulation is becoming established
57+	Outflow tract and ventricular septation complete (membranous part of interventricular septum is completed)
Birth	Functional septation of the atrial chambers, as well as the pulmonary and systemic circulatory systems

**Table 1.1 Developmental timeline of normal human heart embryology**

(adapted from Carlson, 2004 and Iazzo, 2009) (Carlson 2004; Iazzo 2009)



**Figure 1.1 A brief schematic overview of the mouse heart development**

A) The migration of the primitive streak (PS) to the anterior-lateral region of embryo at E6.5. B) Formation of the cardiac crescent at E7.5. C) Linear heart tube at E8. D) The heart tube elongates and loops rightward from E8.5-E10.5. E) The looped heart tube with well-defined cardiac chambers (upper panel, ventral view; lower panel, dorsal view). F) E12.5 four chambered heart.

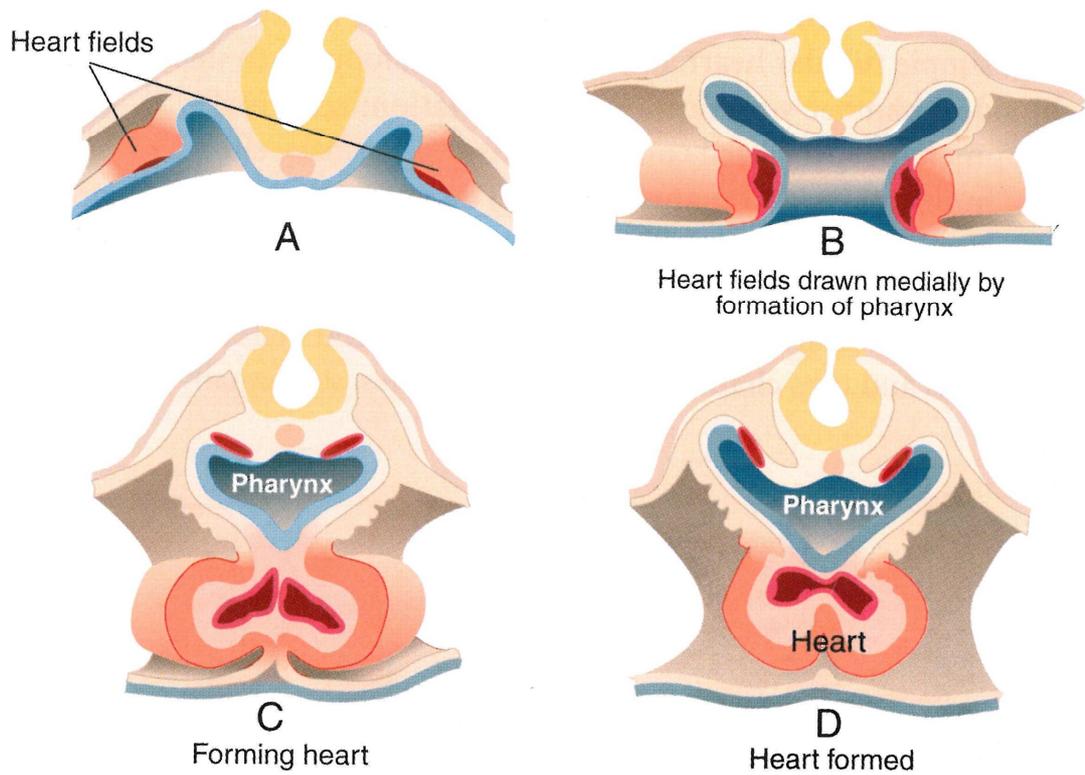
A, anterior; P, posterior; R, right; L, left; AA, aortic arch; Ao, aorta; AVC, atrioventricular canal; HF, head folds; IFT, inflow tract; IVC, inferior vena cava; IVS, interventricular septum; LA, left atrium; LV, left ventricle; ML, midline; OFT, outflow tract; PLA, primitive left atrium; PRA, primitive right atrium; PS, primitive streak; PT, pulmonary trunk; PV, pulmonary vein; RA, right atrium; RV, right ventricle; SVC superior vena cava; Tr, trabeculae. (Adapted from Buckingham et al. 2005) (Buckingham et al. 2005).

cardiogenic fields move to the ventral midline in a rostrocaudal orientation fusing from anterior to posterior (Arguello et al. 1975). Figure 1.2 shows the formation of the heart tube from bilateral cardiogenic fields. During the fusion process, cardiac jelly, a thick acellular matrix, is secreted, thus resulting in three layers in the cardiac tube: cardiac jelly layer sandwiched between an inner endocardial layer and outer myocardial layer (Kirby 2007).

Cells continue to be added to the arterial pole to form the outflow tract and to the venous pole to form the inflow region of the heart. In the arterial pole, mesenchymal cells in the pharyngeal splanchnic mesoderm caudal to the heart tube are mobilized to the outflow tract (Waldo et al. 2001). In the venous pole, nearby mesenchyme is recruited at each phase of development to form myocardial cells for maturation of the atria and veins (van den Hoff et al. 2001).

### **1.1.2 Looping of the heart**

The establishment of left-right axis occurs during the early gastrulation, before the looping process to induce asymmetric morphogenesis of the heart (Levin 2005). Disruption of the left-right axis results in randomised cardiac looping (Lowe et al. 2001). Initially, the heart tube is suspended at the inner curvature of the ventral foregut through dorsal mesocardium. The heart tube bends ventrally first to obtain C-shaped loop, and then undergoes dextral looping (D-looping) by rotating to the right where the left side of the tube is brought to the front and the inner curvature to the left (Manner 2000). Not much shifting is observed during this period as the intact dorsal mesocardium restricts the move. The dorsal mesocardium disappears as the looping process continues. During looping, lengthening of the heart tube occurs concurrently to form S-shaped loop (de la Cruz and Markwald 1998). The venous and arterial poles of the tube incorporate myocardium at both ends respectively. These myocardial cells express *Islet1* (*Isl1*) and originate from the second heart field (Cai et al. 2003). The migration of the tube from the ventral foregut happens earlier in mice than other vertebrate organisms, resulting in the formation of atria and right ventricle from the cells recruited from the poles of the tube (Cai et al., 2003). The process of convergence proceeds along S-shaped looping, closing the distance between outflow and inflow poles of the heart through the addition of cells from the secondary heart field (Waldo et al. 2005). The wedging process, where



**Figure 1.2 Diagrams showing the formation of heart tube**

Panels A-D showing the stages of a midline heart tube from bilateral primary heart fields in chick (adapted from Kirby, 2007) (Kirby 2007). The two bilateral heart fields are brought into the opposition by the formation of pharynx that fused at the ventral midline, thus giving rise to three layers: endocardial layer, cardiac jelly and myocardial layer.

the aorta moves behind the pulmonary trunk and becomes wedged between the mitral and tricuspid valves, takes place during septation (Yelbuz et al. 2002). In short, looping involves formation of the C-shaped loop, formation of the S-shaped loop by elongation, convergence of the arterial and venous poles of the tube and wedging of the aorta.

### **1.1.3 Cardiac neural crest cells**

Neural crest cells (NCC) are multipotential cells that arise from the dorsal neural tube. NCC then delaminate and migrate extensively throughout the body. NCC are divided into cranial and trunk regions depending on the cells' original location in the neural tube. Cardiac NCC is a sub-population of the cranial NCC that stem from the middle of the otic placode to the border of somite 3 (Kirby and Stewart 1983). Cardiac NCC give rise to ectomesenchymal cells. Ectomesenchymal cells are implicated in the development of the aorticopulmonary septum, the smooth muscle tunics of the great arteries and the connective tissue of the thymus, thyroid and parathyroids (Kirby et al. 1983). Cardiac NCC also contribute to the semilunar valves and the atrioventricular valves, especially in the septal leaflets (Nakamura et al. 2006).

The term 'cardiac neural crest' was derived from the ablation studies of the neural crest cells in quail-chick chimeras that showed this population of cells are important for normal heart development (Kirby et al. 1985). The cardiac NCC ablation model displayed various cardiovascular and non-cardiovascular defects. The cardiovascular phenotypes are abnormal outflow tract including persistent truncus arteriosus and outflow malalignment, aberrant patterning of the great arteries and myocardial dysfunction whereas the non-cardiovascular phenotypes are hypoplasia or aplasia of the thymus, parathyroids and sometimes the thyroid gland (Hutson and Kirby 2007; Nishibatake et al. 1987; Yelbuz et al. 2002).

During the looping process, myocardium is recruited to the outflow tract from the splanchnic mesoderm to lengthen the heart tube. Ablation of the cardiac NCC results in a straighter and shorter heart tube (Yelbuz et al. 2002). The myocardium that forms the outflow tract is generated by secondary heart field, indicating that abnormal cardiovascular phenotypes might involve the dysregulation of secondary heart field.

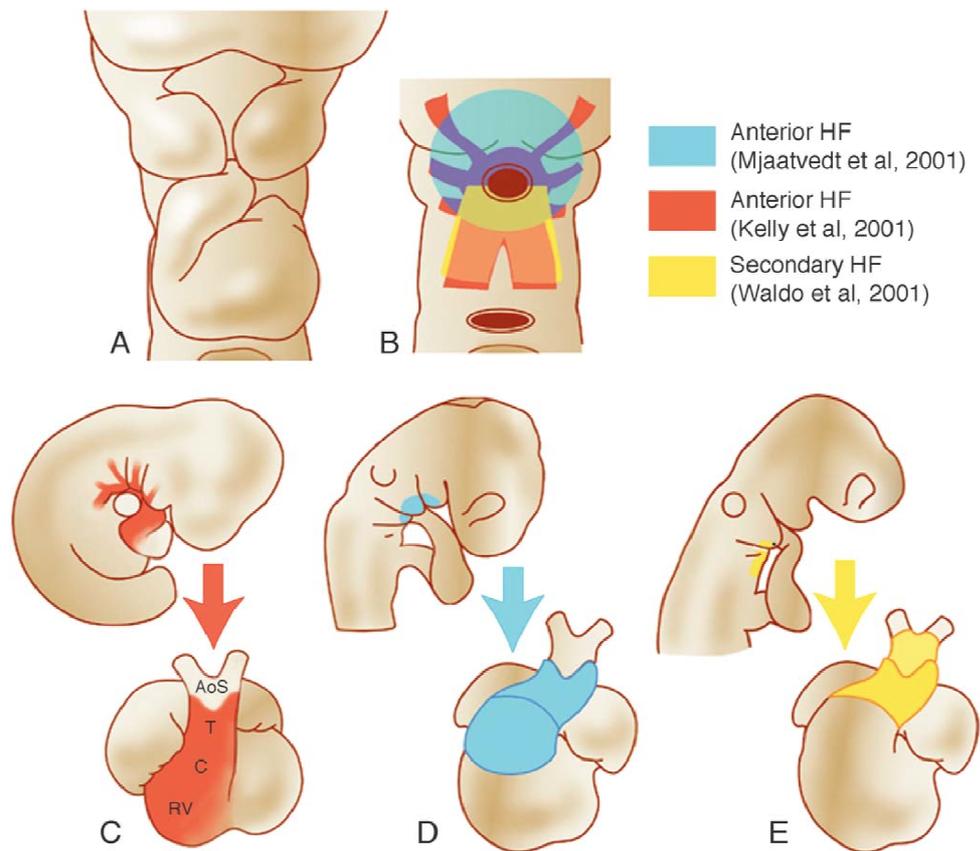
#### 1.1.4 Anterior and secondary heart field

The classical perception on the development of the heart was based on the segmental model of the pre-patterned myocardial cells in the primitive streak that give rise to the outflow tract, right ventricle, left ventricle, atrioventricular canal, right atrium and left atrium in the anterior-posterior axis (de la Cruz and Markwald 1998). However, recent experiments indicated the absence of segmental regionalization and proposed rather the concept of a heart field, where potential myocardial cells are located in the embryonic region that contribute cells to the developing heart. The heart field consisted of an anterior heart field (AHF) and a secondary heart field (SHF) (Buckingham et al. 2005) (Figure 1.3).

In the chick embryo, the AHF is situated immediately anterior to the developing heart tube and consisted of cranial pharyngeal mesoderm surrounding the aortic sac (Mjaatvedt et al. 2001). Fate mapping of the cells showed that they contributed to the conal and truncal regions of the outflow tract. Another group, Waldo and colleagues, described the SHF in the splanchnic mesoderm that underlies the caudal pharynx (Waldo et al. 2001). SHF cells are added to the outflow tract of the heart tube during looping. Cells in the SHF express *Nkx2.5* and *Gata4* before myocardial differentiation.

In the mouse embryo, the *lacZ* transgene with the myosin light chain promoter inserted 114kb upstream of *Fgf10* and proposed to be controlled by the *Fgf10* regulatory elements, showed the transcription of *lacZ* in part of the pharyngeal mesoderm (Kelly et al. 2001). The region that showed  $\beta$ -galactosidase ( $\beta$ -gal) activity from the transgene has been described as the AHF. Initially, the AHF cells lie medially to the cardiac crescent. The dynamic pattern of  $\beta$ -gal activity and cell labelling experiments indicated that the AHF cells moved from the pharyngeal mesoderm into the arterial pole of the heart between E8.25-E10.5 in the mouse embryo (Kelly et al., 2001). The experiments performed by Zaffran and colleagues using lineage tracing approach of the movement of AHF cells in the right ventricle and outflow tract myocardium (Zaffran et al. 2004). Retrospective clonal analysis of myocardial cells indicated two categories of clones: the second lineage contributes to the outflow tract and all other regions of heart, except the left ventricle, which was generated solely from the first lineage (Meilhac et al. 2004). The second lineage, marked by the *Isl1* expression, corresponds to the contribution of SHF cells (Cai et al. 2003). *Isl1* mutant mouse showed a decrease expression of BMP

(*BMP2*, 4, 6 and 7) and FGF genes (*FGF4*, 8, 10), indicating BMPs and FGFs are also expressed in secondary heart field (Cai et al. 2003). BMPs and FGFs are required for the activation of myocardial regulatory genes such as *NKX2.5* and *GATA4*, whereas WNT signaling exhibits inhibitory effect (Brand 2003).



**Figure 1.3 Anterior and secondary heart fields**

Panels A-E show the location, extend and contribution of the cells of anterior and secondary heart field that were identified by Mjaatvedt et al. (Mjaatvedt et al. 2001), Kelly et al. (Kelly et al. 2001) and Waldo et al. (Waldo et al., 2001) (Adapted from Abu-Issa et al. 2004), AoS, aortic sac; T, truncus; C, conus; RV, right ventricle.

### 1.1.5 Chamber formation

During the process of convergence, myocardium is added to the proximal part of the sinus venosus to form the atria. The expanding atria will be segregated into right and left atrium by a primary atrial septum. The primary atrial septum is generated from splanchnic mesoderm near the venous pole that differentiates into myocardium. The primary septum enlarges in the direction of the atrioventricular cushions thus separating the atrium into two chambers. As this septum grows, an opening between the lower rim of the septum primum and the endocardial cushions, ostium primum (foramen primum), is formed. The ostium primum becomes smaller and disappears as the primary septum fused with the endocardial cushions. Before the closure is completed, perforations appear in the central part of the primary septum. Perforations coalesce to form ostium secundum (foramen secundum), another opening that enables the blood to flow from the right to the left atrium (Moore and Persaud 1998). The primary septum grows on the right side of the pulmonary pit, resulting in the developing pulmonary vein being assigned to the left atrium (Webb et al. 2000). The secondary atrial septum completes the process of atrial septation. This secondary septum is located on the right of the primary septum and originates from the myocardium on the left (Franco et al. 2000). Secondary septum forms an incomplete partition between the atria, resulting in foramen ovale. The cranial part of the primary septum gradually disappears. The remaining of the septum becomes the valve of the foramen ovale. The leaflets of the primary and secondary septum fuse at birth, thus becoming a complete partition between the atria (Moore and Persaud 1998).

The atrium and the prospective left ventricle are connected by the lumen of the atrioventricular canal. Endocardial cushions form on the dorsal and ventral walls of the atrioventricular canal. These cushions are enlarged by the inclusion of mesenchymal cells. The endocardial cushions on the opposite ends approach each other and fuse, thus dividing the atrioventricular canal into right and left canal (Moore and Persaud 1998).

The right and left ventricles arise from a narrowed area that joins the inflow and outflow poles. The ballooning of the outer curvature myocardium expands the ventricular chambers (Christoffels et al. 2000). The trabeculae, which are myocardial diverticulae, grow towards the lumen from the outer wall of the myocardium, which is formed simultaneously, to separate the chamber into right and left ventricles (van Mierop and

Kutsche 1985). The ventricular septum develops from the merging of trabeculae (Ben-Shachar et al. 1985), becoming the muscular interventricular septum. The fusion between the free edge of the interventricular septum and the endocardial cushions is not complete, resulting in interventricular foramen, permitting communication between the right and left ventricles. This opening is later closed by the membranous interventricular septum. The closure of the foramen and the formation of the membranous interventricular septum is due to the fusion of tissues from the right bulbar ridge, the left bulbar ridge and the endocardial cushion (Moore and Persaud 1998).

The active proliferations of the mesenchymal cells invade the walls of the bulbus cordis results in the formation of bulbar ridges. At about the same time, similar ridges form in the truncus arteriosus. These ridges (bulbar and truncal ridges) are derived mainly from the neural crest mesenchyme (Kirby et al. 1983). The bulbar and the truncal ridges undergo 180° spiralling as they grow, resulting in the aorticopulmonary septum when the ridges fuse. This septum divides the bulbus cordis and the truncus arteriosus into an aortic and a pulmonary trunk. The bulbus cordis is incorporated into the walls of the definitive ventricles: the bulbus cordis forms the walls of the aortic vestibule inferior to the aortic valve in the left ventricle and forms the conus arteriosus in the right ventricle where the pulmonary trunk arises.

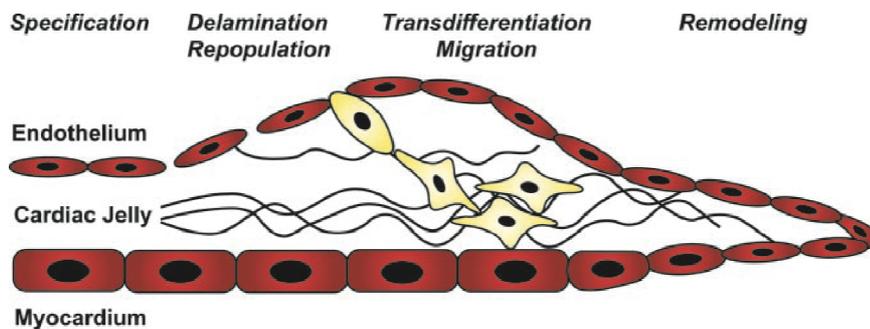
#### **1.1.6 Valve development**

After looping of the heart, in the region between the common atria and ventricle, the cardiac jelly expands into swellings called cardiac cushions (Anderson 2003) (Figure 1.4). Cardiac jelly separates the myocardium (outer layer) and endothelium (inner layer) of the heart tube. During valve formation, a subset of cells from the inner endothelium transform into mesenchymal cells that invade the cardiac jelly. This process is called epithelial-to-mesenchymal transformation (EMT) (Markwald et al. 1977). In short, EMT and cushion formation involve activation and delamination of endothelial cells which then migrate into the cardiac jelly and proliferate.

Atrioventricular cushion mesenchyme gives rise to the tricuspid and mitral valves. Chordae tendineae and the papillary muscles connect the valve leaflets to the atrioventricular cushion. The leaflets consist of three layers: the atrialis, the spongiosa

and the fibrosa from the atrial to the ventricular sides (Lincoln et al. 2006). The septal leaflets of the mitral and tricuspid valves stem from the dorsal and ventral atrioventricular cushions and delaminate from the ventricular septal wall while the mural leaflets develop from the myocardium of the atrioventricular canal and the myocardium below the leaflet subsequently disappear due to cell death (de Lange et al. 2004).

The aortic and pulmonary semilunar valves originate from the mesenchyme that formed truncal cushions during septation of the outflow tract. The cushions are composed of mesenchyme from various locations: cardiac neural crest from the pharyngeal arches, the pharynx, and epithelial-mesenchymal transformation of the endocardium. There are three leaflets in each valve. The leaflets form from three pairs of ridges extending into the lumen of the truncus. These early leaflets comprise of a main mesenchymal tissue overlaid with endocardium. The interaction of endocardium and the mesenchyme creates an arterial face of the leaflets thus detaching the leaflets (Hurle et al. 1980).



**Figure 1.4 Heart valve development**

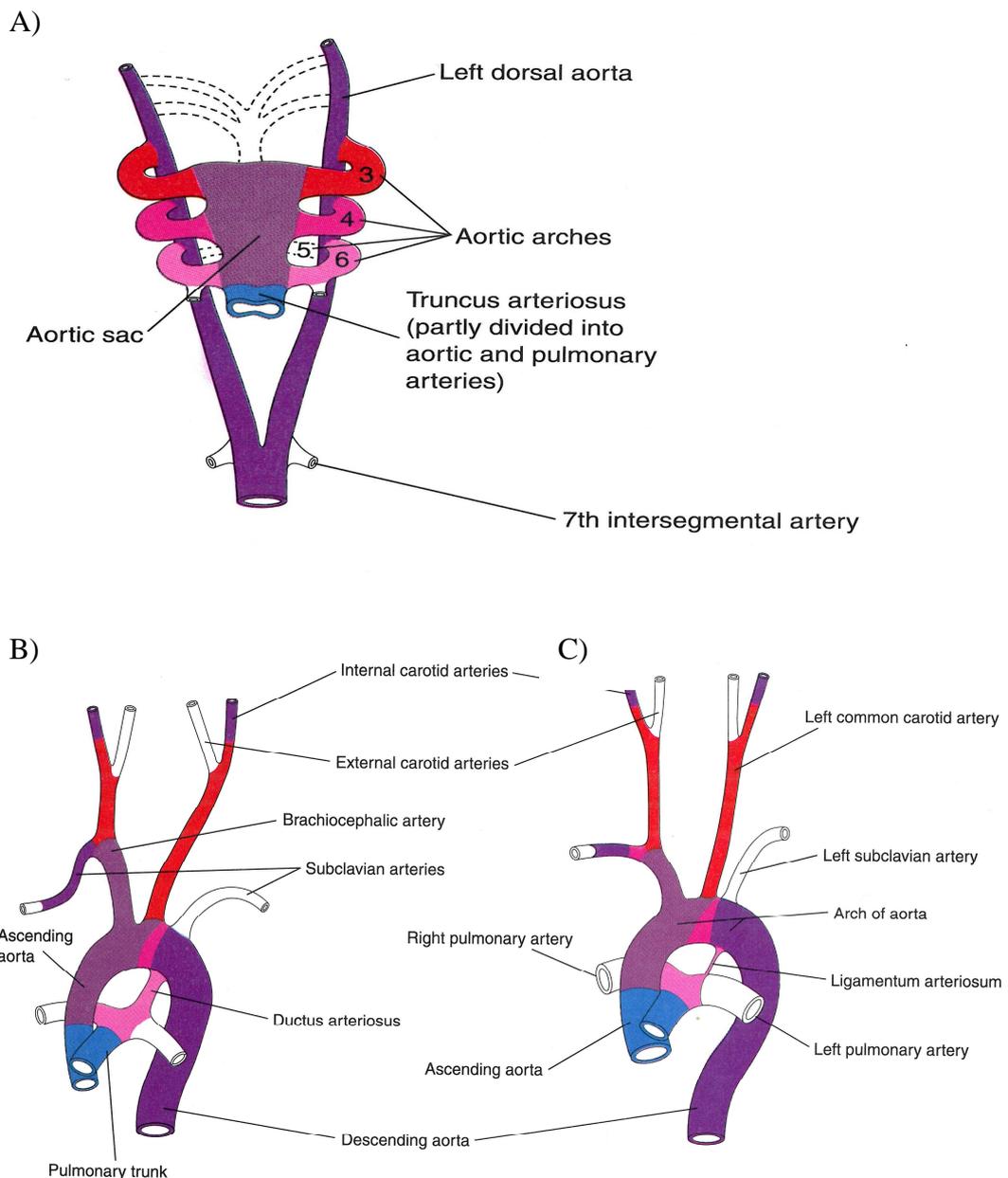
Diagrammatic overview of the heart valve development (Adapted from Armstrong and Bischoff, 2004) (Armstrong and Bischoff 2004).

### 1.1.7 Aortic arch derivatives

The pharyngeal arches are formed during the 4<sup>th</sup> and 5<sup>th</sup> week of development, they each received their own arteries known as aortic arches that arise from the aortic sac. The aortic arches begin from the aortic sac and terminate in the dorsal aorta. The aortic arch arteries begin as symmetrical sets of vessels. A total of six pairs of aortic arches are developed. However, these aortic arches are not all present at the same time. When the fourth and sixth arches are in the process of formation, the first two pairs have regressed (Moore and Persaud 1998).

The first aortic arch has largely disappeared, but a remaining small portion forms the maxillary artery and may also contribute to the formation of the external carotid arteries. The second aortic arch also regresses. The dorsal parts of these vessels remain and form the stapedial arteries, which are small vessels that run through the ring of the stapes, a small ear bone, in the embryo. The fifth pair aortic arch exists briefly and is never well developed. The remaining aortic arches: third, fourth and sixth, persist (Moore and Persaud 1998).

During further development of the aortic arches, the system gradually loses its symmetrical pattern of arteries due to remodelling that leads to the mature aortic arch and great vessel patterning (Kameda 2009) (Figure 1.5). The repatterning of the arteries into the asymmetric great arteries is due to the interaction of neural crest cells with the aortic arch arteries (Kirby 2007). The third aortic arch forms the common carotid artery and the first part of the internal carotid artery. The fourth aortic arch develops differently on the right and left sides (Moore and Persaud 1998). The left fourth aortic arch forms part of the aorta, between the left common carotid and the left subclavian arteries. On the right, it forms the proximal part of the right subclavian artery. The distal part of the subclavian artery is formed from the right dorsal aorta and the seventh intersegmental artery (Moore and Persaud 1998). The proximal part of the right sixth aortic arch becomes the proximal part of the right pulmonary artery whereas the distal part disappears. The proximal part of the left sixth aortic arch persists as the proximal part of the left pulmonary artery. The distal part forms the ductus arteriosus during fetal life (Moore and Persaud 1998). Irregularities of the embryonic heart such as interrupted aortic arch and coarctation of the aorta result from the persistence of parts of the aortic arches that usually disappear, or from disappearance of parts that normally persist.



**Figure 1.5 Diagrams of the aortic arch system**

A) Aortic arches at 6 weeks of embryonic development, the first two pairs of aortic arches have regressed. The obliterated components are indicated by broken lines.

B) Diagram of the aortic arches after remodelling at 8 weeks.

C) Diagram of the aortic arches in the adult.

The ductus arteriosus in B) is essential for the communication between pulmonary trunk and aorta, however, in C) the ductus arteriosus becomes the ligamentum arteriosum after birth (adapted from Moore and Persaud, 1998) (Moore and Persaud 1998).

## **1.2 Congenital cardiovascular malformation**

### **1.2.1 Epidemiology of cardiovascular malformation**

Congenital cardiovascular malformation (CVM) is a major group of birth defects that consists of developmental defects of the heart leading to impaired cardiovascular function. CVM affects approximately 7 in every 1000 live births (Hoffman and Kaplan 2002) and is the most common human birth defect worldwide (Christianson et al. 2006) (Table 1.1). The prevalence of CVM would be higher if the proportion of pregnancies which do not continue to term, and patients with bicuspid aortic valve who were diagnosed later in life, were included (Ward 2000). Advancement in medical and cardiac surgical care have resulted in most patients with CVM surviving to adulthood. There are now many adults living with CVM, and studies on CVM recurrence have shown that this patient population has increased risk of an affected child than the general population (Burn et al. 1998). CVM is thought to arise from the cumulative effects of genetic and environmental factors (Nora 1968). Some proportion of CVM occurs in patients with chromosomal abnormalities (7-14% in different series) and a handful of families with multiple affected individuals, due to Mendelian transmission, have been described (see below). Some genetic defects and certain environmental factors that contribute to CVM have been identified (Jenkins et al. 2007; Pierpont et al. 2007). However, the majority of the cause of CVM remains unknown.

Anomalies of the heart can be classified using several systems. In this project, CVM is categorised in individual patients based on the nature of the CVM that was judged clinically most important (usually, the malformation requiring surgery). It has been noted that classification systems do vary from study to study, especially on how multiple lesions are classified. The complex phenotypic spectrum of CVM may present challenges in identifying genetic predisposing factors (Pierpont et al. 2007). An alternative approach to classification involves grouping together of particular defects based upon what is known of developmental pathways in cardiogenesis. However, there remains considerable uncertainty regarding the processes involved in clinically observed CVM phenotypes. Moreover, new knowledge can substantially alter such classification schemes. For example, prior to the discovery of the second heart field, malformations of the cardiac outflow tract were generally thought to be mediated through cardiac neural crest cells functioning incorrectly during development. It has subsequently become clear that complex interactions, for example between neural crest

cells and cells derived from the second heart field, are critical to outflow tract morphogenesis. In the following paragraphs, the principal clinical CVM phenotypes are briefly listed and described.

Aortic stenosis (AS) occurs in 3-6% of CVM cases and is observed more frequently in males than in females, with the ratio of 4:1. This defect consists of variable degree of thickening of the valve that does not function properly, resulting in the impairment of the left ventricle to pump blood out of the heart into the aorta. Patent ductus arteriosus (PDA) and coarctation of the aorta (CoA) frequently coexist with AS (Braunwald 1997). Treatment using balloon dilatation or surgery is usually performed to palliate the symptoms and valve replacement will be carried out later in life (Hoffman et al. 2004).

Bicuspid aortic valve (BAV) is a congenital deformity where the aortic valve has only two cusps, instead of three cusps in normal heart development. BAV may become calcified later in life due to calcium deposition and lead to various degrees of severity of aortic stenosis and aortic regurgitation (Braunwald 1997). However, unlike other defects, BAV can function normally at birth and only develop stenosis or incompetence after 40 years of age. Therefore, the incidence of BAV is usually not recorded in paediatric population studies and varies from about 0.4-2.25% (Hoffman and Kaplan 2002). BAV can lead to calcified aortic valve with severe stenosis, aortic regurgitation, infective endocarditis and dilatation of the aortic root with dissection of the aorta. Therefore, surgical valve replacement is usually needed in middle life.

Coarctation of the aorta (CoA) could be detected during infancy if the defect is critical. This critical CoA accounts for one third of all CoA and if left untreated, leads to early death (Hoffman et al. 2004). In other cases, CoA would only manifest significant symptoms after 20 to 30 years (Braunwald 1997). This defect is a condition in which the aorta is significantly narrowed due to the thickening of the posterolateral aortic wall opposite the ductus arteriosus (Braunwald 1997). CoA might arise due to the small involution of the left aorta that moves cranially with the left subclavian artery during the remodelling of the aortic arches. The other common anomalies associated with this defect are BAV, VSD, and mitral stenosis or regurgitation. Surgery or balloon dilatation could relieve the symptoms but sometimes CoA may reoccur as the child grows.

Atrial septal defect (ASD) is one of the most common congenital cardiovascular malformations that occur with an average incidence of 941 per million live births (Hoffman and Kaplan 2002). The true incidence is higher as this malformation may be undetected in early life. ASD results in a left to right shunt of blood. The locations of ASD vary from ostium primum atrial defect, ostium secundum defect in midseptal region involving fossa ovalis and defects of the sinus venosus located high in the atrial septum (Braunwald 1997). Ostium primum defect is an atrioventricular septal defect (AVSD) and is discussed later. The most significant ASD is the ostium secundum defect, whereby a large opening between the left and right atria is present. This anomaly is due to excessive resorption of the septum primum or due to the inadequate development of the septum secundum. Patent foramen ovale (PFO) occurs when the foramen ovale failed to fuse with septum primum postnatally (Braunwald 1997). Patients with ASD may exhibit congestive heart failure, pulmonary vascular disease, mitral incompetence, atrial arrhythmias and thrombosis of large pulmonary arteries if left untreated. Uncomplicated ASD is suitable for closure with a device introduced while the patient is undergoing cardiac catheterisation. Surgery would be recommended for more severe cases (Braunwald 1997).

Atrioventricular septal defect (AVSD) occurs in 4-5% of CVM cases. AVSD is frequently observed in Down syndrome (Trisomy 21) patients, where the mothers involved are usually more than 34 years old (Hoffmann and Kaplan 2002), and Ellis-van Creveld syndrome and asplenia or polysplenia syndromes (Braunwald 1997). The endocardial cushions of the atrioventricular canal divide the canal into two (left and right) and also participates in the formation of the membranous portion of interventricular septum and in the closure of the ostium primum. Failure of fusion of the endocardial cushions results in AVSD. This anomaly is characterised by an atrial and a ventricular component separated by abnormal valve leaflets in the single atrioventricular orifice (Moore and Persaud 1998). If the endocardial cushions fuse only partially, where the ostium primum failed to fuse with endocardial cushions but the interventricular septum is closed, this defect is known as ostium primum defect. The defect is usually combined with deformed and incompetent AV valves (Braunwald 1997). Surgical repair is the recommended treatment for AVSD patients. If the left AV leaflet tissue is significantly deformed, mitral valve replacement would also be carried out (Braunwald 1997).

Double outlet right ventricle (DORV) occurs in average 157 out of one million live births (Hoffman and Kaplan 2002). The definition of double outlet right ventricle (DORV) varies but generally occurs when both the pulmonary and the aorta are exiting from the right ventricle (Braunwald 1997). More than 50% of the DORV patients display defects of the right AV valves; some patients have mitral atresia with a hypoplastic left ventricle and in a handful of patients exhibit tricuspid stenosis. Due to the various anatomic defects in the patients, infants are given palliation. Corrective surgery is carried out when the patient is age 1-2 years old. In contrast, double outlet left ventricle (DOLV) is one of the rarest cardiac anomalies (Braunwald 1997). DOLV refers to the origin of both great arteries from the left ventricle. Various defects are associated with DOLV, the most common are VSD and PS. Operative repair to close the VSD and positioning of a right ventricle-pulmonary artery conduit is usually performed on the patient.

Ebstein's anomaly occurs due to the anomalous attachment of the tricuspid leaflets, where the downward displacement of the tricuspid valve is located into the right ventricle (Braunwald 1997). Associated malformations such as patent foramen ovale or ASD is present in more than half the cases. This defect has a mean of 114 incidences per million live births (Hoffman and Kaplan 2002). Surgical approach such as valvoplasty or replacement of the tricuspid valve is used to treat the patients.

Hypoplastic left heart (HLH) occurs in average 266 cases in one million live births (Hoffman and Kaplan 2002). HLH consists of underdevelopment of the left cardiac chambers, atresia or stenosis of the aortic and/or the mitral orifices, and hypoplasia of the aorta; which usually cause heart failure during the first week of life. Pulmonary venous blood enters the right side of the heart through a PFO and the systemic circulation is sustained by PDA (Braunwald 1997). Reparative surgeries are performed in stages to alleviate the symptoms or cardiac transplantation is required.

Patent ductus arteriosus (PDA) is another common anomaly that varies with the gestational age of the subject. Preterm infants usually have higher incidence rate of PDA due to the abnormal physiology whereas in term infants have lower incidence rate. However, spontaneous closure can occur in some cases that may further inflate the incidence of PDA (Hoffman and Kaplan 2002). PDA is the abnormal existence of the ductus arteriosus after birth. The ductus arteriosus present in the fetus as a vessel

connecting the pulmonary trunk and the descending aorta distal to the left subclavian artery, to enable the fetus to bypass the pulmonary circulation. This channel is usually closed at birth. However, a small shunt of blood from the aorta to the left pulmonary artery is present in a normal full-term infant for 24 to 48 hours (Sansoucie and Cavaliere 1997). In the adult, the obliterated ductus arteriosus forms the ligamentum arteriosum. Closure of the ductus arteriosus in full-term infants is regulated by oxygen (Bernstein 1996). Bradykinin, a substance realised from the lungs during their initial inflation, causes the smooth muscle to contract. This substance is dependent on the high oxygen level of the aortic blood after aeration of the lungs at birth. PDA that persists weeks after birth is pathological entity. Surgical closure of PDA would be carried out to alleviate the symptoms.

Pulmonary atresia (PA) is an uncommon malformation that is highly lethal. PA is characterized by atretic pulmonic valve, due to the hypoplastic development of both the pulmonary valve and artery or occasionally accompanied by extremely narrowed right ventricle, resulting in obstruction of the pulmonary circulation (Braunwald 1997). Aggressive medical and surgical treatment would be prescribed during the neonatal period.

Pulmonary stenosis (PS) occurs in about 7% of CVM patients. However, this number does not mirror the true estimate of the incidence as there are unknown number of patients with very mild PS that do not require medical attention. PS results from fusion of the pulmonary valve cusps leading to obstruction of the pulmonary circulation and thickening of the right ventricle (Braunwald 1997). This anomaly is frequently observed in patients with Noonan syndrome. Surgical or balloon valvotomy is necessary for these patients.

Single ventricle is a lesion where both atrioventricular valves, or a common atrioventricular valve open into a single ventricular chamber (Braunwald 1997). The associated anomalies are abnormal great artery (L-malposition or D-malposition of the great artery), pulmonic valvular stenosis, subaortic stenosis, and coarctation of the aorta (Braunwald 1997). This lesion has an average incidence of 106 per million live births (Hoffman and Kaplan 2002). Surgical treatment using Fontan procedure is applied to the patients.

Tetralogy of Fallot (TOF) is one of the malformations of the cardiac outflow tract (OFT) and is the most common form of cyanotic CVM (Braunwald 1997). There are four heart defects in TOF: ventricular septal defect, right ventricular outflow obstruction to the lungs (pulmonary stenosis), an overriding aorta that receives blood from both ventricles and right ventricular hypertrophy. The fourth feature, right ventricular hypertrophy, arose due to the physiological consequence of the other three features as the right ventricular has to pump deoxygenated blood through the narrowed pulmonary valve. TOF patients would undergo open-heart surgery to correct these defects. However, the patients often have residual VSD and right ventricular outflow tract obstruction and pulmonic valve incompetence. Occasionally these patients suffer from complete heart block and sudden death.

Total anomalous pulmonary venous connection (TAPVC) occurs when connection of all the pulmonary veins to the right atrium or to the systemic veins, resulting in systemic venous circulation. Other cardiac malformations such as single ventricle, truncus arteriosus, and anomalies of the systemic veins occur in some TAPVC patients (Braunwald 1997). TAPVC occurs in average 94 patients out of one million live births. Reparative surgery is performed for the TAPVC patient as soon as possible as survival without treatment after 1 year of age is rare (Hoffman et al. 2004).

Transposition of the great arteries (TGA) occurs in 315 infants in one million live births (Hoffman and Kaplan 2002). TGA is a malformation characterized by the aberrant origin of the aorta from the right ventricle and the pulmonary artery from the left ventricle. This defect results in two parallel circulations communicate through ASD, PDA or VSD (Braunwald 1997). Additional cardiac malformations are small PDA and PFO that occurs in two thirds of TGA patients. The remaining of the patients exhibit either a large PDA or ASD, a VSD, subpulmonic stenosis or some combination of these defects (Hoffman et al. 2004). Arterial switch is usually performed on TGA patients.

The absence or fusion of the tricuspid valves, an interatrial communication, ventricular septal defect, hypoplasia of the right ventricle and hypertrophy of the left ventricle characterizes tricuspid atresia (Braunwald 1997). Additional CVMs, like TGA, PDA and CoA are usually present in patients with tricuspid atresia. The Fontan-Kreutzer procedure is carried out to correct the anomaly.

Truncus arteriosus is a rare but lethal lesion whereby the defect results from the failure of septation of the aortapulmonary trunk and is always accompanied by a VSD. Thus the single outflow tract vessel containing systemic, pulmonary and coronary arteries, overrides both ventricles and receives blood from both sides (Braunwald 1997). The operation would be carried out to close the VSD and the pulmonary arteries are excised from the truncus and an extracardiac valved conduit is used to establish the link between the right ventricle and the pulmonary arteries (Braunwald 1997).

Ventricular septal defect (VSD) is the most prevalent form of CVM, with an incidence rate of 2-5% (Hoffman and Kaplan 2002). The incidence might be higher as in some cases the defect closes within the first year of life. This defect occurs both as an isolated anomaly and also in combination with other anomalies. The ventricular septum consists of four compartments (the membranous septum, the inlet septum, the trabecular septum, and the infundibular septum). Defects that lead to VSD are due to deficiency of growth or failure of alignment in any of the compartment (Braunwald 1997). Incomplete closure of the interventricular foramen due to the failure of the membranous part of the interventricular septum to develop is the most common type of VSD (Moore and Persaud 1998). The severity of this defect depends on the size of the opening. If the opening is small, no treatment or operation is required as these defects usually close spontaneously. However, for larger defects causing significant left to right shunting, operation is required as large defects if left untreated would result in early congestive heart failure and death.

<b>Types of CVM</b>	<b>Mean Incidence</b>
Aortic stenosis	401
Atrial septal defect	941
Atrioventricular septal defect	348
Bicuspid aortic valve	13,556
Coarctation of the aorta	409
Double outlet right ventricle	157
Ebstein's anomaly	114
Hypoplastic left heart	266
Hypoplastic right heart	222
Patent ductus arteriosus	799
Pulmonary atresia	132
Pulmonary stenosis	729
Single ventricle	106
Tetralogy of Fallot	421
Total anomalous pulmonary venous connection	94
Transposition of the great arteries	315
Tricuspid atresia	79
Truncus arteriosus	107
Ventricular septal defect	3570

**Table 1.2 Mean incidence of CVM per million live births**

The mean incidences of major CVMs per million live births is shown in Table 1.2.1 (Hoffman and Kaplan 2002).

### 1.2.2 Non-inherited risk factors for cardiovascular malformation

Several non-inherited risk factors for CVM have been identified (Jenkins et al. 2007) (Table 1.2). One of the most potentially exciting discoveries was the reduced incidence of heart defects observed with maternal periconceptional use of multivitamins containing folic acid. In the Hungarian randomized trial study, maternal consumption of multivitamin supplements containing folic acid was associated with an approximately 60% reduction in risk for CVMs, with odds ratios of 0.42, confidence interval (CI) 0.19-0.98 (Czeizel et al. 1998). This discovery was confirmed by another independent study using a population-based case-control approach in Atlanta that showed an approximately 25% reduction in risk for CVMs in the offspring when periconceptional multivitamin was used with odds ratios of 0.76, CI 0.60-0.97 (Botto et al. 2000). A study has been carried out by Ionescu-Ittu and colleagues that showed reduced incidence rate of CVM by 6% per year for seven consecutive years after the implementation of the Canadian government policy of folic acid fortification of grain product when compared to the seven years before fortification (Ionescu-Ittu et al. 2009). This study supported the role of folate in reducing CVM but the required level of folate intake was unknown.

Folic acid is known to reduce risk of neural tube defects. This was conclusively shown in the MRC randomized folate trial (MRC 1991). Despite the studies of CVM referenced above, a recent consensus statement from the American Heart Association (Jenkins et al. 2007) described the relationship between folate and CVM as uncertain. This is a reasonable view, as the Hungarian and Atlanta studies used multivitamins rather than folate, and the Atlanta study was non-randomized. It is possible that the relationships observed in those studies were due to confounding. It would now be unethical to perform trials with an arm not including folate supplementation, so further randomized evidence will not be forthcoming. In this context, examining the relationship between genetic polymorphisms known to affect plasma folate and the risk of CVM could identify a causal association. Studies have shown the association of folate metabolism with homozygous 677C>T variant of methylenetetrahydrofolate reductase, *MTHFR* (Bailey and Berry 2005). *MTHFR* is an enzyme that catalyses the conversion of 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate, which is the bioactive form of folate. The homozygous TT genotype of 677C>T results in

<b>Non-inherited factors</b>	<b>Defects</b>
Multivitamin supplements (including folic acid)	Any defects, particularly: VSD
Phenylketonuria	Any defects
Maternal diabetes	Any defects, particularly: Complete AVSD Cardiomyopathy Laterality and looping TGA
Maternal rubella	Any defects, particularly: VSD PDA Pulmonary valve abnormalities Peripheral pulmonic stenosis
Maternal febrile illnesses	Any defects, particularly: Pulmonic stenosis Conotruncal defects Right-sided obstructive defects Tricuspid atresia All left-sided obstructive defects Coarctation of the aorta VSD
Maternal influenzae	Any defects, particularly: TGA All right-sided obstructive defects All left-sided obstructive defects Coarctation of the aorta VSD
Anticonvulsants	Any defects
Indomethacin tocolysis	PDA
Ibuprofen	Any defects, particularly: TGA VSD Bicuspid aortic valve
Thalidomide	Any defects

**Table 1.3 Non-inherited risk factors for CVM**

The non-inherited risk factors/ maternal exposure associated with CVM or specific type of CVMs (adapted from Jenkins et al., 2007)

reduced activity of the MTHFR enzyme and decreased level of plasma folate (Blom et al. 2006).

Thus, a causal relationship between plasma folate and CVM could be inferred if robust association between TT genotype and CVM could be established. Meta-analysis of thirteen studies on the relationship between *MTHFR* polymorphism and CMVs gave overall odds ratios of 1.3 (95%, CI 0.97-1.73) and 1.2 (95%, CI 0.83-1.74) for fetal and maternal MTHFR C/T genotypes respectively (van Beynum et al. 2007).

The caveats of van Beynum's study are the heterogeneity of the population of the thirteen studies and the small total number of cases involved. Not all of the studies analysed considered the intake of folic acid in the maternal diet. Analysing patients with various types of CVMs also complicated the study, as it might be possible that specific types of CVM could be associated with folate metabolism. Therefore, folate is thought to be still an open question that may be resolved by larger meta-analyses.

A number of maternal illnesses and conditions have been associated with an increasing risk of CVMs. Untreated maternal phenylketonuria resulting in high basal phenylalanine level ( $\geq 900\mu\text{M}$ ) was associated with a greater than six-fold-increased risk of having an offspring with CVM when compared to the metabolic control group (with phenylalanine level  $< 600\mu\text{M}$ ) (Levy et al. 2001). The most frequent defects observed are coarctation of the aorta and hypoplastic left heart syndrome.

Maternal pregestational diabetes has been consistently associated with an increased risk of CVM in offspring, with an odds ratio of 4.7 (95% CI 2.8-7.9) (Loffredo et al. 2001). The defects observed in the offspring are cardiomyopathy (OR=15.1, 95% CI 5.5-41.3), laterality and looping defects (OR=10.0, 95% CI 3.7-27.0), transposition of the great vessels (OR=6.6, 95% CI 3.2-13.3), and complete atrioventricular septal defects (OR =22.8, 95% CI 7.4-70.5).

Maternal rubella infection early in gestation causes CVM in offspring, in particularly PDA, pulmonary valve abnormalities, peripheral pulmonary stenosis and VSDs (Jenkins et al. 2007). Maternal influenza in the first trimester of pregnancy has been described as a probable risk factor for CVM (Scanlon et al. 1998; Zhang and Cai 1993). Other maternal febrile illnesses during the first trimester of pregnancy might lead to defects

such as pulmonic stenosis, tricuspid atresia, all right-sided obstructive defects, coarctation of the aorta, all left-sided obstructive defects, conotruncal defects and VSDs (Botto et al. 2001; Ferencz et al. 1997), though data are at present inconclusive.

Maternal therapeutic drug exposure has been associated with an increased risk of CVM. Maternal ingestion of drugs such as Thalidomide, Anticonvulsants, and nonsteroidal anti-inflammatory drugs (Indomethacin tocolysis, and Ibuprofen), will lead to offspring having various CVMs (Ericson and Kallen 2001; Norton 1997; Schardein 2000; Smithells and Newman 1992; Wilson et al. 1998).

Some findings have demonstrated a conclusive link of environmental factors to CVM in offspring, for example, rubella, Thalidomide, diabetes and anticonvulsants. However, the majority of the findings in this area remain inconclusive. This is because prospective epidemiological studies of environmental exposures are difficult to perform because the condition is uncommon and therefore pre-conceptual populations of hundreds of thousands of women would need to be enrolled for sufficient disease outcomes to be observed. On the other hand, retrospective studies have limitations as they utilize the possible exposure-outcome relationships from cases and controls that may be subject to recall or reporting bias. Studies of other adverse pregnancy outcomes have suggested that the effect of such biases could be substantial. Therefore, there may be many more hitherto undiscovered environmental factors that cause CVM. One potential use of genetic studies could be to identify some of these factors. If variants in genes mediating an individual's response to environmental exposures – for example through vitamin or xenobiotic metabolism – were found to be associated with CVM, then this would implicate the environmental exposure in disease causation.

### **1.2.3 Genetic basis for cardiovascular malformation**

#### **1.2.3.1 “Sporadic” CVM**

Although the incidence of CVM is higher when the fetus is exposed to non-inherited risk factors, most fetuses remain unaffected, suggesting there may be interaction between environmental exposures and genetic susceptibility factors. Familial recurrence and transmission risks indicate a significant genetic contribution to CVM (Burn et al. 1998; Oyen et al. 2009). Many studies have explored familial recurrence risks since the original proposal of a multifactorial aetiology to CVM by Nora in 1966 (Nora and Meyer 1966). However, almost all of these have issues with possible ascertainment bias that could have influenced their results (Burn et al. 1998; Lie et al. 1994; Nora and Nora 1987; Pradat 1994). Two contrasting studies, one British, one Danish, that attempted complete population ascertainment are more likely to have made accurate estimates of familial recurrence risks, and are discussed here in more detail. These studies addressed recurrence risk in the ~80% of cases that cannot be attributed to particular genetic syndromes or environmental causes.

The British offspring study by Burn and colleagues (Burn et al. 1998) showed the variable heritability among non-syndromic heart defects. They identified 727 individuals with disturbance of situs or segmental connection, with atrioventricular septal defect or with tetralogy of Fallot that survived surgery before 1970. Each individual was examined and interviewed. The CVM recurrence in the offspring of this population was 4.1%, which was 16 out of 393 live offspring, excluding miscarriages and terminated pregnancies. This risk was significantly different from sibling risk (2.1%, with  $p=0.021$ ) and substantially higher than the population baseline risk of 0.7%. When recurrence risk for specific cardiac defects was investigated, no cases of recurrent CVM in the offspring of probands with TGA were found, there was a 7.8% risk in the offspring of 88 individuals with AVSD and 3.1% recurrence risk in the offspring of 395 individuals with TOF. These results suggested AVSD and TOF were particularly influenced by genes. Computer modelling predicted that TGA is a sporadic defect, AVSD a single-gene defect and TOF a polygenic disorder of three genetic loci. However, the study conducted by Burn and colleagues might have included undiagnosed syndromic cases with CVM and did not include all of the CVM patients of reproductive age that have undergone surgery, achieving a ~70% participation rate among identified probands.

Oyen and colleagues utilized large and carefully annotated Danish population registries over three decades to investigate the recurrence risk for offspring of CVMs patients and to estimate the contribution of CVMs in families to the general population (Oyen et al. 2009). The recurrence risk ratio for all types of CVM was 3.21 among the first-degree relatives of a proband with CVM. However, this risk decreased when CVM was identified only in second and third-degree relatives. This result strongly suggested that genetic component played an important role to the manifestation of CVM. In a commentary on this paper, Shieh and Srivastava (Shieh and Srivastava 2009) reiterated the hypothesis that sporadic CVM is caused by genetic defects with variable penetrance that will only manifest as disease in conjunction with other genetic, epigenetic, environmental or hemodynamic factors. Oyen et al showed that specific CVMs demonstrated significant higher risk in first-degree relatives. Heterotaxia has a high recurrence risk ratio of 79.1 (95% CI 32.9-190), followed by right ventricular outflow tract obstruction with 48.6 (95% CI 2.5-85.6), atrioventricular septal defect with 24.3 (95% CI 12.2-48.7), left ventricular outflow tract obstruction with 12.9 (95% CI 7.48-22.2), conotruncal defects with 11.7 (95% CI 8.0-17.0), isolated atrial septal defect with 7.1 (95% CI 4.5-11.1) and isolated ventricular septal defect with 3.4 (95% CI 2.2-5.3). However, the contribution of multiple affected CVM families to the general population with CVM was 2.2% to 4.2% (excluding chromosomal cases), indicating that only a small proportion of CVM cases were from CVM families.

### 1.2.3.2 CVM as part of particular genetic syndromes

A number of CVMs have been associated with a single gene or chromosomal aberrations. However, CVM even within the syndromic and isolated cases demonstrated incomplete penetrance and variable expression. The 22q11 deletion syndrome has been described as several overlapping clinical syndromes including DiGeorge Syndrome, Velocardiofacial (Shprintzen) Syndrome or conotruncal anomaly face syndromes. Individuals with this deletion often display palate anomalies, feeding disorders, speech and learning disabilities, renal anomalies, behavioural disorders and CVMs. The common CVMs associated with a 22q11 deletion are tetralogy of Fallot, truncus arteriosus, conoventricular VSDs, and aortic arch anomalies (Goldmuntz et al. 1998; Marino et al. 2001; Momma et al. 1995). Approximately 6-28% of del22q11 cases are inherited from a parent in an autosomal dominant fashion while the remainder are *de novo* (Digilio et al. 2003). Not infrequently, a parent is identified with the deletion after a child with CVM has been cytogenetically diagnosed.

Williams-Beuren Syndrome is associated with specific CVMs, infantile hypercalcemia, skeletal and renal anomalies, cognitive deficits, “social personality” and elfin faces. The specific CVMs for this syndrome are supravalvular aortic stenosis, in conjunction with supravalvular pulmonary stenosis and peripheral pulmonary stenosis (Eronen et al. 2002). About 90% of the individuals affected with this syndrome can be diagnosed with microdeletion at chromosome 7q1.23 by fluorescence *in situ* hybridization (FISH).

The trisomies are chromosomal aberrations that are commonly associated with CVMs. Trisomy 13 (Patau syndrome) show a high incidence of atrial and ventricular septal defects, and disturbance of cardiac position. Patients diagnosed with Edwards syndrome (trisomy 18) have a 90-100% chance of having CVMs. About 40-50% of the patients with trisomy 21 (Down syndrome) have CVMs, and of these, almost half are diagnosed with AVSD (Pierpont et al. 2007).

There are other syndromes that are associated with single-gene defects. Alagille syndrome patients showed 20p12 deletion or *JAG1* mutation (located in 20p12.2-p11.23). *JAG1* encodes a Notch ligand protein (Krantz et al. 1997; Li et al. 1997). This syndrome is an autosomal dominant disorder. The clinical features of this syndrome are highly variable due to variable penetrance. The most common features are anomalies of

the liver, face, eyes, renal, neurons and heart. The specific heart defects occur in more than 90% of individuals with Alagille syndrome and include peripheral pulmonary hypoplasia, tetralogy of Fallot, and pulmonary valve stenosis (McElhinney et al. 2002).

Noonan syndrome is characterized by short stature, typical facial dysmorphism, webbed neck, chest deformity and CVMs. Approximately 80-90% of the individuals affected with Noonan syndrome are diagnosed with CVM. The most common observed cardiac anomalies are pulmonic valve stenosis and hypertrophic cardiomyopathy (Marino et al. 1999; Noonan 1994). Other defects such as secundum atrial septal defect, atrioventricular septal defect, mitral valve abnormalities, aortic coarctation and tetralogy of Fallot also occur. This genetically heterogeneous syndrome is associated with at least 3 genes: *PTPN11*, *SOS1*, and *KRAS* (Digilio et al. 2002; Roberts et al. 2007; Schubbert et al. 2006).

Holt-Oram syndrome is caused by mutations in the *TBX5* transcription factor gene located in chromosome 12q24.1 (Basson et al. 1997). This syndrome is characterized by upper-limb deformities and CVMs. Approximately 75% of these patients have ASD or VSD and/or progressive atrioventricular conduction disease.

Patients with CHARGE syndrome are characterized by coloboma of the eye, heart defects, atresia of the choanae, retardation, genital and ear abnormalities (Tellier et al. 1998). More than 50% of the patients were identified with mutation in the *CHD7* gene (Vissers et al. 2004). Mutations in the *TFAP2B* gene, a transcription factor expressed in neural crest cells, have been identified in two families with Char syndrome (Satoda et al. 2000). Individuals with Char syndrome are characterized by facial dysmorphism, hand anomalies and patent ductus arteriosus. These single-gene and chromosomal aberration show autosomal dominant inheritance and majority of these syndromes arise *de novo*. Table 1.3 summarizes the genetic syndromes associated with CVM.

<b>Syndrome</b>	<b>Causal gene/ loci</b>	<b>Type of CVMs</b>
Down syndrome	Trisomy 21	ASD, VSD, disturbance of cardiac position
Patau syndrome	Trisomy 18	ASD, VSD, PDA, TOF, DORV, D-TGA, CoA, BAV, BPV, polyvalvular nodular dysplasia
Edwards syndrome	Trisomy 13	AVSD, VSD, ASD
DiGeorge syndrome	22q11 deletion	IAA-B, TOF, PTA, VSD, isolated AA anomalies
Williams-Beuren syndrome	7q11.23 deletion	Supravalvar AS, PS, PPS
Alagille syndrome	20p12 deletion or <i>JAG1</i> mutation	PS, PH, TOF
Noonan syndrome	<i>PTPN11</i> , <i>SOS1</i> , and <i>KRAS</i>	Valvular PS, hypertrophic cardiomyopathy, secundum ASD, AVSE, mitral valve abnormalities, CoA, TOF
Holt-Oram syndrome	<i>TBX5</i>	ASD, VSD, progressive AV conduction disease
CHARGE syndrome	<i>CHD7</i>	TOF, AV canal defects, VSD, ASD, CoA, PDA
Char syndrome	<i>TFAP2B</i>	PDA

**Table 1.4 Genetic syndromes associated with CVM**

The genetic syndromes and the causal genes/ locus that are associated with CVM (adapted from Pierpont et al, 2007) (Pierpont et al. 2007).

### 1.2.3.3 Non-syndromic Mendelian families

Very few large families with autosomal dominant inheritance of non-syndromic CVM have been described. A number of genes have been identified by performing linkage analysis. Mutations in *NKX2.5* have been identified in 4 kindreds with atrial septal defects and atrioventricular conduction delay without syndromic defects (Schott et al. 1998). Two separate families with abnormal aortic valve that leads to severe valve calcification have *NOTCH1* mutations resulting in truncated polypeptide or altered polypeptide (Garg et al. 2005). Members of two families with ASD only or with ASD and other septal defects were identified with mutations in the *GATA4* transcription factor (Garg et al. 2003). In sporadic non-syndromic CVM cases, where no immediate family members were diagnosed with CVM, a handful of genes have been implicated. Screening of non-syndromic TOF patients performed by Goldmuntz and colleagues found more non-synonymous genetic variants of *NKX2.5* (Goldmuntz et al. 2001). Other investigations have identified mutations in cardiac candidate genes in CVM patients, for example *CRELD1* in AVSD cases (Zatyka et al. 2005) and *CONNEXIN43* in HLH cases (Dasgupta et al. 2001). Table 1.4 shows the genes that are associated with non-syndromic CVM. It must be emphasised that such non-syndromic Mendelian CVM families are very rare and contribute much less than 1% to the population burden of disease.

Gene	Type of CVMs
<i>NKX2.5</i>	Familial ASD, AV block
<i>GATA4</i>	ASD or VSD
<i>CRELD1</i>	AVSD
<i>CONNEXIN43</i>	HLH
<i>TAPVR1</i>	TAPVR

**Table 1.5 Genes associated with non-syndromic CVM**

Genes associated with autosomal dominant inheritance of non-syndromic CVM that have been described.

## 1.3 Bone Morphogenetic Protein signaling pathway

### 1.3.1 BMP signaling during heart formation

Bone morphogenetic proteins (BMPs) were originally named for their ability to induce bone formation (Urist 1965). Since then, BMPs have been implicated in a plethora of cellular functions during development and in adult life. Early heart development requires BMPs, particularly cardiomyocyte differentiation (Ghosh-Choudhury et al. 2003; Schlange et al. 2000; Shi et al. 2000). BMPs belong to the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily and comprise a subfamily of more than 20 members. BMPs are synthesized as large precursor proteins. After dimerization, these proteins are activated by endoproteolytic cleavage to produce mature dimer. The N-terminal of the protein controls the stability of the processed mature protein and the C-terminal adjacent to the cleavage site determines the cleavage efficiency (Constam and Robertson 1999). Activated BMPs are then glycosylated and secreted as homo- and heterodimers.

BMPs signal through the assembly of hetero-tetrameric serine/threonine kinase receptors, which comprise of two type I and two type II receptors (Figure 1.5). To date, three type I receptors and three type II receptors have been identified to mediate BMP signaling (Koenig et al. 1994; Liu et al. 1995; ten Dijke et al. 1994). These receptors are distinct from each other and form their own subgroups (Kawabata et al. 1998). However, they share sequence similarity among TGF- $\beta$  family receptors. The type II receptors are *BMPR2* (*BRK3*), *ACVR2A* (*ActRIIA*) and *ACVR2B* (*ActRIIB*) and the type I receptors are *ACVR1* (*ALK2*), *BMPRIA* (*ALK3*) and *BMPR1B* (*ALK6*). *BMPR2* is specific to BMP signaling whereas the other type II receptors *ACVR2A* (*ActRII*) and *ACVR2B* (*ActRIIB*) not only transduce BMP signals but also bind to activins for signaling (Kawabata et al., 1995). Among the type I receptors, *BMPR1B* is unlikely perhaps to play a part in cardiogenesis, as it is not expressed in the embryonic heart (Dewulf et al. 1995). Endothelial-specific *ACVR1* mutant mice generated using the *Cre/loxP* system display defects in the atrioventricular septa and valves, resulting from a failure of endocardial cells to differentiate into the mesenchyme in the atrioventricular canal (Wang et al. 2005a). *BMPRIA* is expressed in embryonic hearts and is discussed further in section 1.3.4. *BMPRIA* and *BMPR1B* preferentially mediate different BMP ligands from *ACVR1*. The BMP ligands that bind to *BMPRIA* and *BMPR1B* are *BMP2* and *BMP4* while *BMP7* binds to these receptors with low affinity. By contrast, *ALK2* primarily mediates *BMP5*, *BMP6* and *BMP7* signaling (de Caestecker 2004).

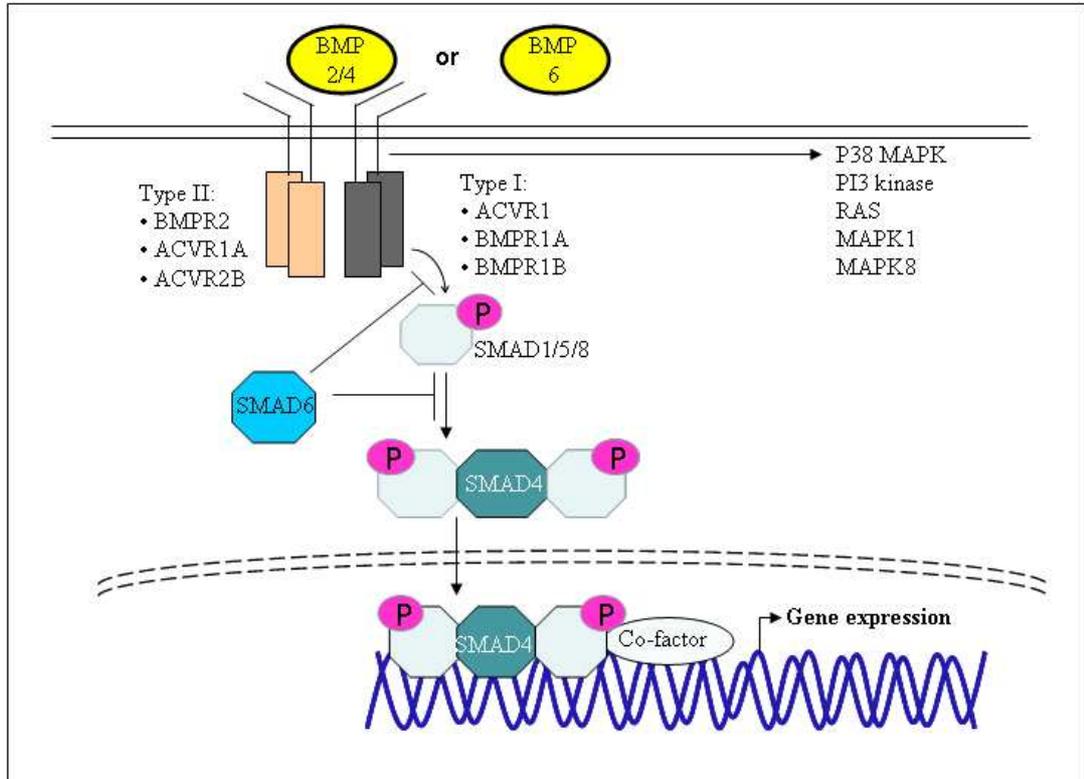
When BMP ligands bind to the heteromeric complex of type I and type II receptors, the type II receptor, which is constitutively in an active state, phosphorylates the type I receptor through the glycine-serine domain (Miyazono et al. 2005). Intracellular events after receptor activation have been shown to be mediated by SMAD (small mother against decapentaplegic) proteins (Heldin et al. 1997). Type I receptors in turn trigger intracellular signaling by phosphorylating the downstream components, which are receptor-activated SMADs (R-SMADs: *SMAD1/5/8*). After releasing from the receptor, the phosphorylated R-SMADs will form a complex with a common SMAD (*SMAD4*) and translocate into the nucleus where they participate in gene transcription by interacting with other transcription factors. *SMAD6* and *SMAD7* are inhibitory SMAD proteins (I-SMADs) induced by BMPs which inhibit the BMP signaling pathway through a negative feedback loop (Imamura et al. 1997; Souchelnytskyi et al. 1998).

Regulation of BMP signaling occurs at three levels: extracellularly, membrane level and intracellularly. At the extracellular level, Noggin (*NOG*) and other cysteine knot containing BMP antagonists bind directly to BMPs, thus inhibiting BMP signaling (Waldo et al. 2001). Co-receptors at the membrane level on one hand inhibit signal transduction by intercepting with receptor complex formation, e.g. *BAMBI* (Balemans and Van Hul 2002) and on the other hand may enhance signaling by presenting BMPs to receptors, e.g. *RGMB (DRAGON)* (Samad et al. 2005). Intracellular regulation occurs both in the cytoplasm and nucleus. *SMAD6* and *SMURF1* modulate BMP signals in the cytoplasm and also in the nucleus (Goto et al. 2007; Murakami et al. 2003). BMP signaling can be mediated by other intracellular proteins including *MAP3K7/MAP3K7IP1* to induce *p38 MAPK*, *PI3 kinase*, *RAS*, *MAPK1* and *MAPK8* (Kinbara et al. 2003; Nohe et al. 2004 ; Roux and Blenis 2004; Vanhaesebroeck et al. 2001).

BMP signaling is phylogenetically conserved in heart development. BMP is present not only in vertebrate, but also in *Drosophila* where it is known as *Decapentaplegic (Dpp)*, which is an orthologue of vertebrate *BMP2* and *BMP4*. *Dpp* induces both *tinman* and *pannier* (Sorrentino et al. 2005). *Tinman* is a transcription factor, whose vertebrate homolog is the critical cardiac transcription factor *Nkx2.5* whereas *pannier* is a Gata transcription factor family member. *Dpp* maintains the expression of *tinman* during differentiation of the cardioblasts (Xu et al. 1998). The observation that BMP2

expression in the anterior endoderm of drosophila, chick, zebrafish, frog and mouse is required to induce myocardial differentiation further strengthen the essential role of BMP signaling in heart formation throughout evolution (Andree et al. 1998; Brand 2003; Kishimoto et al. 1997; Lough and Sugi 2000; Schultheiss et al. 1997; Shi et al. 2000).

Five genes from the BMP signaling pathway were selected as candidates that might be involved in susceptibility to CVM in BMP signaling. *BMP2* and *BMP4* are BMP ligands at the extracellular level, *BMPR2* and *BMPRIA* at the membrane level and *SMAD6* at the intracellular level. These genes were selected based on their expression pattern during embryonic heart development, the cardiac abnormalities demonstrated by mutant animal models where available, and their role and specificity to BMP signaling. Detailed background on the selected candidate genes is presented in the following sections.



**Figure 1.6 BMP signaling pathway**

Activated BMP ligands bind to type I and type II receptors, type II receptor will then phosphorylate the type I receptor, which in turn phosphorylates receptor-activated SMADs (R-SMADs). The activated SMAD binds to common SMAD4 upon which the complexes are translocated to the nucleus and participate in transcription.

### 1.3.2 *BMP2*

Bone morphogenetic protein 2 (*BMP2*), also known as *BMP2A*, is located on chromosome 20p12 and has three exons (Figure 2.1 – Methods). The total transcript length is 3.6kb, producing a polypeptide of 396 amino acids in human. This gene has a TGF- $\beta$  propeptide in the N-terminal and a TGF- $\beta$  domain in the C-terminal that serves as a multifunctional peptide.

*In situ* hybridization in chicken embryonic hearts showed that the avian *Bmp2* gene was expressed in the endoderm at HH stage 5, and that it is essential for forming cardiac mesoderm (Somi et al. 2004). This experiment confirmed the previous study demonstrating that the delivery of *Bmp2* soaked beads to regions of non-cardiogenic mesoderm in vivo triggered the expression of cardiac transcription factors like *Nkx2.5* and *Gata4* (Schultheiss et al. 1997). At HH stage 8, *Bmp2* was expressed in the paired lateral heart-forming regions. *Bmp2* was later seen in the myocardium of the inflow tract at the venous pole of the heart tube at HH stage 10 (Somi et al. 2004). Inactivation of *Bmp2* by the addition of *Noggin* (an inhibitor that binds to TGF- $\beta$  family ligands [including BMP ligands] with high affinity to prevent the ligands from binding to receptors and thus inhibits signal transduction) to the explants at HH stage 5-8, inhibited expression of cardiac transcription factors like *Nkx2.5* and *eHand*. This further suggests that *Bmp2* is an important molecule for downstream cardiac specification (Schlange et al. 2000). At HH stage 16, in the arterial pole, *Bmp2* was expressed in the myocardium of the distal OFT and in the mesenchyme distal to the OFT but not in the pharyngeal arches. In the venous pole, *Bmp2* was expressed in the myocardial layer of the inflow tract. From HH stage 16-23, expression of *Bmp2* was restricted to the primary myocardium of the atrioventricular canal (AVC). At HH stage 23, *Bmp2* was expressed in the mesenchyme of the distal OFT ridges and the forming aorticopulmonary septum. *Bmp2* was expressed throughout the entire AVC at this stage. *Bmp2* was also present in the cardiomyocytes that protrude into the mesenchyme during muscularization of the AV cushions and tricuspid valve (Somi et al. 2004). In vitro cultures of cells from endocardial cushions of chick showed that *Bmp2* could phosphorylate *Smad1/5/8* and induce the expression of cartilage markers, such as *sox9* and *aggrecan*. This further reaffirmed the role of *Bmp2* in heart valve development (Lincoln et al. 2006). In the venous pole, *BMP2* was expressed in the distal myocardial border of caval vein and in the distal myocardium around the pulmonary vein (Somi et al. 2004).

*Tbx2* and *Tbx3* are T-box DNA binding proteins with T-domain proteins that are phylogenetically conserved and function as transcription factors to regulate a variety of developmental processes, including heart development (Naiche et al. 2005). The expression of *Tbx2* and *Tbx3* closely overlapped with the regions that express *Bmp2* in chicks. Treatment with *Bmp2* in non-cardiogenic embryonic tissue leads to the expression of *Tbx2* and *Tbx3*, demonstrating that *Bmp2* mediates heart development upstream of *Tbx2* and *Tbx3* (Yamada et al. 2000).

*MEF2A* (Myocyte-specific enhancer factor 2A) is a transcription factor initially implicated in the development of heart and cardiac contractility in zebrafish (Naya et al. 2002; Wang et al. 2005b). Investigation of *Bmp2* and *Mef2a* in zebrafish and neonatal rat cardiomyocytes showed that *Bmp2* is required for cardiac contractility and *Bmp2* could activate *Mef2a* transcriptional activity (Wang et al. 2007). This study indicates that *Bmp2* acts upstream of *Mef2a* in the development of functionally contractive cardiomyocytes.

To elucidate the function of *BMP2* in the development of the mouse heart, targeted deletion of *Bmp2* was performed (Zhang and Bradley 1996). The homozygous *Bmp2* knockout mouse died between E7.0 and E10.5, after formation of the heart tube. In the mutant mouse, the heart tube was formed inside the exocoelomic cavity instead of inside the amniotic cavity in the wild type. The expression of *Bmp2* was detected at E7.5 in the mesoderm of normal mouse (Zhang and Bradley 1996). At E9.5, *Bmp2* was presented in the outer myocardial cells of the AVC (Lyons et al. 1990; Zhang and Bradley 1996).

To overcome the embryonic lethality of homozygous *BMP2* knockout mouse, Ma and colleagues generated a *Bmp2* null allele by deleting the exon3 of *Bmp2* and crossed with *Nkx2.5<sup>cre</sup>* knock-in allele (Ma et al. 2005). *BMP* signaling was disrupted in the endocardial cells in the AVC at E9.5, resulting in severe AVC defects, a defect in cardiac jelly formation and the failure of mesenchymal cells to migrate to the AVC (Ma et al. 2005). The deletion of *Bmp2* in cardiac progenitors also results in changes of cell fate, where differentiated chamber myocardium was observed in heart-valve forming region (Rivera-Feliciano and Tabin 2006).

The expression and ablation of *BMP2* in the animal models suggest a role in the aetiology of conotruncal cardiovascular malformations and AVC defects. However, no re-sequencing of the *BMP2* in CVM patients had been performed at the start of this study. Therefore, investigation on whether mutations in *BMP2* could cause sporadic CVM in man was performed.

### 1.3.3 *BMP4*

Bone morphogenetic protein 4 (*BMP4*) is also known as *BMP2B*. This gene is located on chromosome 14q22-q23 and has three isoforms (Figure 2.2 – Methods). The 4 exons of *BMP4* generate a polypeptide of 408 amino acids. The functional domains of *BMP4* consist of a TGF- $\beta$  propeptide in the N-terminal and the TGF- $\beta$  domain in the C-terminal, similar to other BMP ligands.

In zebrafish, *Bmp4* has been shown to regulate left-right (LR) patterning both during early segmentation and during late segmentation (Chocron et al. 2007). During early segmentation, BMP signaling is required in the right lateral plate mesoderm to regulate both visceral and heart laterality; whereas during late segmentation, BMP signaling is required in the left lateral plate mesoderm to regulate left-sided gene expression and heart laterality. Restricted morpholino knock down experiments targeting the splice donor site between exon 1 and exon 2 of *Bmp4* to produce aberrant *Bmp4* protein in Kupffer's vesicle results in LR patterning defects (Chocron et al. 2007). In both zebrafish and *Xenopus*, *Bmp4* is expressed strongly on the left of the linear heart tube. Ectopic overexpression of *Bmp4* bilaterally in the heart tube results in randomized looping (Breckenridge et al. 2001).

Somi and colleagues carried out a study to identify which *Bmp* ligand might be principally involved in myocardium formation in intra- and extracardiac mesenchyme in vivo by visualizing the expression patterns of *Bmp2*, *Bmp4*, *Bmp5*, *Bmp6* and *Bmp7* mRNA using in situ hybridisation in avian embryonic hearts from Hamburger-Hamilton (HH) stage 5 to 44 (Somi et al. 2004). At HH stage 5, *Bmp4* was expressed in the ectoderm adjacent to the cardiac precursor cells (Schultheiss et al. 1997; Somi et al. 2004). The cardiac precursor cells are located in the mesoderm in the anterior bilateral regions of the embryo. When a straight heart tube is formed at HH stage 10, *Bmp4* was observed in the myocardium of the arterial pole and adjacent mesenchyme at both sides of the linear heart tube. *Bmp4* was expressed in the distal myocardial border (arterial pole) and the flanking mesenchyme at HH stage 10-23 and the expression was restricted to the mesenchyme later on. Expression of *Bmp4* was not seen in the inflow tract myocardium but it was present in the adjacent mesenchyme flanking the tract (Somi et al. 2004). Loss of function of *Bmp4* by the misexpression of *Noggin*, which binds to BMP ligands and inhibits BMP signaling in the developing heart from HH stage 10-12

revealed abnormalities of the outflow tract cushions development, resulting in a ventricular septal defect with double-outlet right ventricle or common arterial trunk (Allen et al. 2001). Between HH stage 16-23, when muscularization of the atrioventricular cushions and the tricuspid valve takes place, Bmp4 was expressed in part of the mesenchyme of the mitral valve and the myocardium of the tricuspid valve, at the ventricular level (Somi et al. 2004). Bmp4 expression was detected in the ventral layer of the right sinus horn at HH stage 13 and by HH stage 17, Bmp4 was expressed in the proepicardium at the venous pole of the developing heart (Schlueter et al. 2006). From HH stage 36 onward, Bmp4 was expressed in the endocardium throughout the entire heart (Somi et al. 2004). These studies showed the dynamic patterns of Bmp4 throughout chicken heart development.

RNA of Bmp4 was detected in mouse embryos from E6.5 to E10.5 using RT-PCR (Winnier et al. 1995). At E6.5, low levels of Bmp4 was expressed in the posterior primitive streak and by E7.5, the expression of Bmp4 was stronger. In the knockout mouse, where targeted deletion of Bmp4 without the first exon after amino acid position seven was generated by homologous recombination in embryonic stem (ES) cells, homozygous mutant embryos died between E6.5 and E9.5, during gastrulation (Winnier et al. 1995). In embryos heterozygous for the Bmp4<sup>lacZ</sup> allele, where the lacZ reporter gene was knocked into the Bmp4 locus, expression of Bmp4 was observed in the OFT at E8.5 and was maintained until birth; whereas in the inflow region, Bmp4 was detected in the sinus venosus at E8.5 (Jiao et al. 2003). At E9.0, Bmp4 was visualized in the dorsal midline of the common atrium and weakly in the atrioventricular canal (AVC). Cells from these regions will form the atrial septum primum (ASP) and inferior endocardial cushion (IEC). Bmp4 expression was later observed in cardiomyocytes covering the IEC from E9.5 to E12.5 and the ASP from E10.5 to birth, indicating that Bmp4 might play an important role during septation and valvulogenesis. In order to test this hypothesis and to circumvent the early lethality of Bmp knockout mice, a Cre/loxP approach, where the transcription of Cre recombinase is driven by the rat cardiac Troponin T (cTnT) promoter (Wang et al. 2001; Wang et al. 2000b), was crossed with floxed Bmp4 allele (*Bmp4*<sup>loxP-lacZ/oxP-lacZ</sup>) or (*Bmp4*<sup>loxP-lacZ/tm1</sup>), to specifically inactivate Bmp4 expression in cardiomyocytes was utilized. The generated mutant mice showed atrioventricular canal defect (AVCD), a single atrioventricular junction with a common valve at E15.5-16.5, similar to the human patient with phenotypic defect of AVCD (Jiao et al. 2003). Bmp4 mutants with different genotypes [*Bmp4*<sup>loxP-lacZ/tm1</sup>, cTnT-Cre;

*Bmp4*<sup>loxp-lacZ;loxp-lacZl</sup> and cTnT-Cre; *Bmp4*<sup>cre;loxp-lacZ/tm1</sup>]; also displayed cardiac defects, ranging from atrial septal defect (ASD), partial AVCD, complete AVCD to double outlet right ventricle (DORV), successfully recapitulating a variety of cardiac phenotypic defects in patients.

*Bmp4* expression was detectable in the OFT myocardium and within mesoderm ventral to the branchial-arch arteries (BAAs) at E9.0 (Liu et al. 2004). Deletion of *Bmp4* from myocardium of the developing heart in mouse using a conditional null allele of *Bmp4* and the *Nkx2.5*<sup>cre</sup> recombinase allele showed aberrant septation of the proximal OFT and reduced proliferation of OFT mesenchyme (Liu et al. 2004).

The cardiac defects observed in *Bmp4* animal models demonstrate that *Bmp4* has a pivotal role in heart development. Thus it remains to be elucidated whether mutations of *BMP4* in human result in CVM.

### 1.3.4 *BMPRIA*

*BMPRIA*, bone morphogenetic protein type 1A, is one of the three type I receptors that can mediate BMP signals. *BMPRIA* is also known as *ALK3*, *ACVRLK3*, and *CD292*. *BMPRIA* is located on chromosome 10q22.3 and has two isoforms (Figure 2.4 – Methods). This gene comprises 13 exons, encoding an approximately 3.6kb transcript, that generates a polypeptide of 532 amino acids in humans. There are four functional domains in the mature protein of BMPR1A: the extracellular ligand binding domain, the transmembrane domain, a serine/threonine kinase domain and the cytoplasmic domain.

In chick, Okagawa and colleagues showed the presence of mRNA of *Bmpr1a* by RT-PCR in chick AV endocardium at HH stage 14, during the transformation of endocardial endothelial cells into invasive mesenchymal cells (EMT) (Okagawa et al. 2007). No studies have been carried out to visualize the expression patterns of *Bmpr1a* mRNA in different stages of developing avian embryonic heart.

In mouse, *Bmpr1a* is broadly expressed during early mouse embryogenesis. *Bmpr1a* was detected throughout the epiblast and mesoderm at E7.0 and at E8.5, *Bmpr1a* was observed in the three germ layers and in most adult mouse tissues (Mishina et al. 1995). Homozygous mutant mice generated by deleting the first two exons of *Bmpr1a* died during embryogenesis (Mishina et al. 1995). At E7.0 and E8.5, abnormal or degenerative embryos were observed. The abnormal embryos did not have a primitive streak at E7.0, and were approximately half the size of the normal embryos. Histological and molecular analysis of the mutant embryos showed that no mesoderm had formed in the *Bmpr1a* mutant embryos (Mishina et al. 1995). At E8.0, no organized structures were observed in the abnormal embryos, which were much smaller in size than the advanced normal embryos that had formed head folds. No homozygous mutants were recovered at E9.5 (Mishina et al. 1995).

In a study performed by Klaus and colleagues, it was shown that the cardiac crescent and the primitive ventricle were absent in mice carrying a *Bmpr1a* deletion in cardiac mesoderm (Klaus et al. 2007). The mesoderm progenitors contribute to both first and second heart fields. This study showed that *Bmpr1a* is essential for the differentiation of the linear heart tube.

Deletion of exon 2 of *Bmpr1a* (encoding approximately 35% of the extracellular ligand-binding domain) from the myocardium using a Cre/lox system results in abnormal development of the trabeculae, compact myocardium, interventricular septum and endocardial cushions (Gaussin et al. 2002). In another knockout mouse with a deleted floxed *Bmpr1a* allele directed to the cardiac myocytes of the atrioventricular canal (AVC), the tricuspid mural leaflet and mitral septal leaflet were longer, the tricuspid posterior leaflet was displaced and adherent to the ventricular wall, and the annulus fibrosus was disrupted (Gaussin et al. 2005).

Deletion of *Bmpr1a* in the endocardium of mice resulted in loss of AV cushion formation, demonstrating the effect of BMP signaling on endocardium (Ma et al. 2005). The mutant mice showed the expression of phosphorylated-Smad1/5/8 in endocardial cells was decreased compared to the wild type and two-thirds of the mutants had hypoplastic cushions (Ma et al. 2005). Ablation of *Bmpr1a* in neural crest cells resulted in a shortened cardiac outflow tract and septation defects (Stottmann et al. 2004). The mutant embryos died in mid-gestation (at around E11.5-12) due to acute heart failure. Deletion of *Bmpr1a* in *Isl1*-expressing cells showed abnormal right ventricular and outflow tract from E9.5. At E13.5, persistent truncus arteriosus (PTA) and underdeveloped valves, severe aberrant morphology of the outflow tract, ventricular and atrial septal defects were seen in the mutant hearts (Yang et al. 2006).

The cardiac abnormalities observed in the mouse model with heart field specific ablation of *Bmpr1a* coupled with its role as a receptor in BMP signaling suggest *BMPR1A* as a candidate gene for CVM.

### 1.3.5 *BMPR2*

*BMPR2*, bone morphogenetic protein receptor type 2, is the type II receptor specific to BMP signaling. *BMPR2* is also known as *BMPR-II*, *BRK-3*, *T-ALK*, and *BMPR3*. *BMPR2* is located on chromosome 2q33-q34 and has three isoforms (Figure 2.5 – Methods). The genomic structure of *BMPR2* is approximately 190kb, comprising 13 exons, and the gene generates a polypeptide of 1,038 amino acids in humans. There are four functional domains in the mature protein of *BMPR2*: the extracellular ligand binding domain encoded by exons 2 and 3, the transmembrane domain generated by exon 4, a serine/threonine kinase domain from exon 5 to exon 11, and the cytoplasmic domain encoded by exons 12 and 13. Unlike the three other domains, the precise function of the *BMPR2* cytoplasmic domain at the C-terminal remains unknown, as a construct lacking this domain was able to activate Smad protein when expressed in vitro (Kawabata et al. 1995).

The pattern of expression of the other type II receptors (*ActR-IIA* and *-IIB*) in avian embryos can be observed in the primitive streak, indicating that these genes may be activated when mesoderm induction began (Stern et al. 1995). The *ActR-IIA* expression could be seen throughout the anterior mesoderm with the strongest expression in the lateral prospective heart cells (Ehrman and Yutzey 1999). However, no comprehensive studies have been carried out to investigate the expression patterns of *Bmpr2* in the developing avian heart.

In mice, the expression pattern of *Bmpr2* at E6.5-7.5 was uniform in embryonic and extraembryonic regions (Beppu et al. 2000). The *Bmpr2* mutant mice generated by gene targeting without the entire transmembrane domain and the amino-terminal region of the kinase domain (exons 4 and 5) died during embryogenesis (Beppu et al. 2000). However, the heterozygous mice were morphologically normal and fertile. The growth of the homozygous *Bmpr2* mutant embryos was arrested at E6.5, and they lacked a morphologically discernible primitive streak at E7.5. At E8.5, the sizes of the mutant embryos were similar to the E7.5 mutant embryos. No homozygous mutants were recovered at E9.5, suggesting that the development of *Bmpr2* mutant embryos were arrested before gastrulation (Beppu et al. 2000). The embryonic lethal *Bmpr2* knockouts during gastrulation obscure the possible role of *Bmpr2* in early heart formation.

A *Bmpr2* construct with deleted exon 2, resulting in a protein that lacked half of the extracellular ligand-binding domain was generated by Delot et al. This protein was expressed at levels comparable with the wild type construct, but has reduced signaling capability (Delot et al. 2003). Homozygous mutant mice died at mid-gestation with cardiovascular and skeletal defects (Delot et al. 2003). The onset of lethality occurred between E12 and birth. The most striking defect was in the outflow tract with the absence of septation of the conotruncus below the valve level. In addition, the knockout mice also revealed interruption of the aortic arch with varying severity, between the roots of the left common carotid and the left subclavian artery, resulting in communication between the pulmonary trunk and the descending aorta through the ductus arteriosus, which is similar to the phenotypic defect of interrupted aortic arch in humans. No semilunar valves was observed in the outflow tract of the mutant. However, the atrioventricular valves were normal, indicating that atrioventricular cushions can develop normally when BMP signaling is impaired in certain circumstances.

The early lethality and cardiac specific lesions observed in the mouse models showed the importance of *Bmpr2* in heart development. Therefore, mutations in *BMPR2* may give risk to CVM in man.

### 1.3.6 *SMAD6*

Human *SMAD6*, also known as *MADH6*, is located on chromosome 15q21-q22 and has three isoforms (Figure 2.3 – Methods). *SMAD6* is a member of the SMAD (small mother against decapentaplegic) protein family. It functions as an inhibitory SMAD (I-SMAD) to block BMP signaling intracellularly. In non-stimulated cells with BMP ligands, *SMAD6* is mainly located in the nucleus, however, when activated, *SMAD6* can be observed in the cytoplasm and the nucleus. There are two main functional domains in *SMAD6*: the carboxy terminal (C-terminal) and the amino terminal (N-terminal). The C-terminal and N-terminal regions are termed Mad homology 1 (MH1) and MH2 domains respectively. The MH2 domain is highly conserved among SMAD proteins and plays important roles in receptor recognition, interaction with transcription factors, dimerization (homo- and hetero-oligomerization among SMAD proteins) and their nuclear transport (Chacko et al. 2004). However, *SMAD6* lacks the SSXS motif at the end of the MH2 domain. This motif serves as a receptor phosphorylation site that is present in receptor-activated SMADs (R-SMADs). The MH1 domain of *SMAD6* is highly divergent from the MH1 domains of other SMADs. In addition, the similarity of MH1 domain between *SMAD6* and *SMAD7* (another I-SMAD) is only 36.7% (Hanyu et al. 2001). The MH1 domain determines the subcellular localization of *SMAD6* (Hanyu et al. 2001), exhibits sequence-specific DNA binding activity and negatively regulates the functions of the MH2 domain through physical interaction. The interaction between MH1 and MH2 domains is released when activated by type I receptor (Hata et al. 1998). The MH1 domain is not effective in antagonizing BMP signaling but the MH2 domain is responsible for the inhibition of BMP signaling, in which it is nearly as potent as the full-length *SMAD6* (Hanyu et al. 2001).

In chick, *Smad6* was expressed in the posterior primitive streak and the lateral margins of the embryos at Hamburger-Hamilton (HH) stage 4 (Yamada et al. 1999). At HH stage 5, expression was observed in the anterior lateral cardiogenic region and overlapped with the expression of *Nkx2.5*, *Bmp2*, *Bmp4* and *Bmp7*. The intensity of the expression of *Smad6* in the same region was increased by HH stage 7. At HH stage 8, *Smad6* expression was localized to the anterior lateral plate mesoderm and the underlying endoderm in the bilateral heart tubes. As the heart tube forms at HH stage 10, *Smad6* was highly expressed in the myocardium and in the endocardium, throughout the primitive heart tube. The expression of *Smad6* continued even after the process of

looping has begun and persisted at least to HH stage 28. At this time (HH stage 11), *Smad6* expression was similar to *Nkx2.5* and *Bmp7*, but not to *Bmp2* and *Bmp4* (Yamada et al. 1999).

At the later stages of chick heart development, at HH stage 20, *Smad6* was expressed throughout the heart, both in the myocardium and in the endocardium (Yamada et al. 1999). The expression was particularly high in the myocardium in the atrioventricular junction. By HH stages 26-28, when the septation process was nearly completed, *Smad6* was expressed throughout the heart: in the myocardium, endocardium, and atrioventricular (AV) endocardial cushion. The expression of *Smad6* was particularly strong in the regions of ventricular septum and the aorticopulmonary septum, but the outflow tract cushions did not express *Smad6*. Expression patterns of *Nkx2.5*, *Bmp2*, *Bmp4* and *Bmp7* covered all areas of *Smad6* expression, except the endocardial cushions (Yamada et al. 1999). Over-expression of *Smad6* in chick AV cushion endocardial cells resulted in a decrease of epithelial-mesenchymal transformation (EMT) (Desgrosellier et al. 2005) suggesting that *Smad6* is a negative regulator of EMT.

A *Smad6* knockout mouse was generated by inserting a nuclear-localized *LacZ*/neomycin resistance cassette into the 5' terminus of the MH2 domain of *Smad6* (Galvin et al. 2000). *Smad6* was expressed by E7. At E9.5-E13.5, expression of *Smad6* was high in the outflow tract and atrioventricular cushion regions of the heart in the homozygous mutant embryos. These expressions persisted at later embryonic stages but the staining pattern expanded to vascular endothelium of larger vessels (Galvin et al. 2000).

In the adult heart, the expression of *Smad6* was high in cardiac valves and outflow tracts, including both endocardial and endothelial cells, mesenchymal cells in the valves and vascular smooth muscle cells around the aortic and pulmonary roots (Galvin et al. 2000). There was no difference between the homozygous and heterozygous adult mice for *Smad6* expression generally, except in the epicardial and endocardial cells of the right atrium, where the expression of *Smad6* was stronger in the homozygous mutant mice (Galvin et al. 2000). However, the mechanism for this localized upregulation in the homozygous mutant mice was not determined.

Smad6 mutant mice demonstrated defects confined to the cardiovascular system, including hyperplastic valves and abnormal septation of the outflow tract, often resulting in narrow ascending aorta and large pulmonary trunk or vice versa (Galvin et al. 2000). These defects are in line with the Smad6 expression in the endocardium and cushion mesenchyme of the atrioventricular cushion and outflow tract. The development in homozygous null mice of Smad6 of valvular hyperplasia is consistent with the enhanced EMT in the endocardial cushions (Desgrosellier et al. 2005). Homozygous null mutant mouse examined at or after 6 weeks of birth showed ossification around the outflow tracts of the heart, but not in the wild type, indicating that Smad6 plays a role in the homeostasis of the adult cardiovascular system (Galvin et al. 2000).

The expression patterns of Smad6 during embryonic heart development in animal models and the defects in the cardiovascular development of Smad6 mutant mice indicate a prominent role of SMAD6 in cardiac development. Thus, mutation of *SMAD6* might cause CVM.

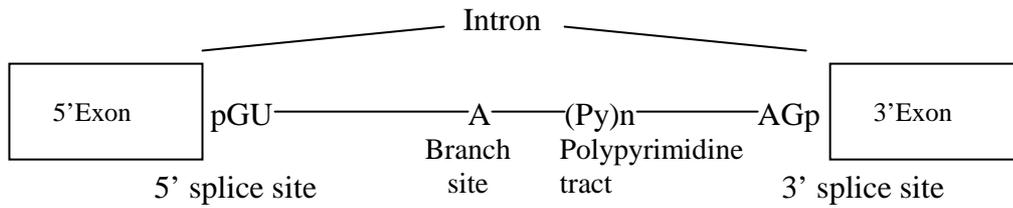
## 1.4 RNA processing

A gene is first transcribed into a pre-messenger RNA (pre-mRNA) that is complementary to the entire length of the genomic DNA, including both intronic regions and exonic sequences. The introns of the pre-mRNA are then excised while the exonic segments are joined end-to-end during a process known as RNA splicing, to produce a mature mRNA for translation.

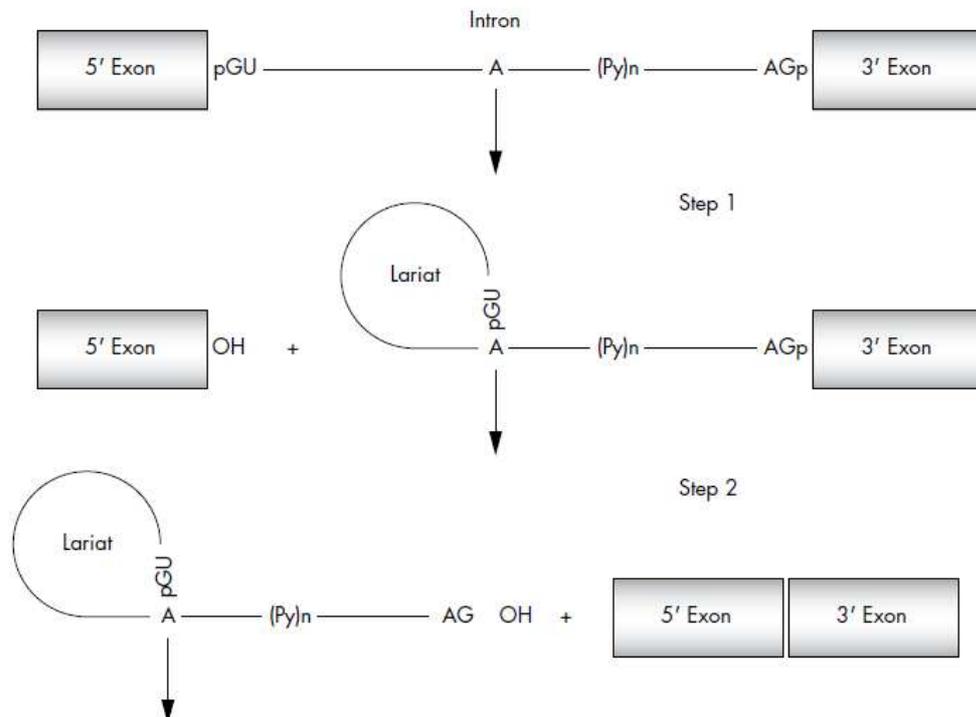
The splicing process takes place at splice sites, which are well-conserved motifs defined by consensus sequences (5' splice site, 3' splice site and branch site) (Figure 1.6). The two bases at the end of introns (5' splice site and 3' splice site) are highly conserved, where most of the introns start with a GU and ends with an AG, although they are not the only important bases (Aebi et al. 1987). Another conserved motif is the branch site that contains an A nucleotide that is universally conserved, followed by the polypyrimidine tract, both of which are located upstream of the 3' splice site (Kol et al., 2005). Other sequences are also present in exons and introns to promote splicing. These sequences act in a cis manner to increase (namely exonic or intronic splicing enhancers: ESE or ISE) or decrease (exonic or intronic splicing silencers: ESS or ISS) recognition of the splice sites (Cartegni et al. 2002).

The splicing mechanism involves a nucleophilic attack on the G nucleotide of the 5' splice site carried out by the 2' OH group on the conserved A nucleotide of the branch site, located upstream of the 3' splice site. This reaction results in a lariat formation, where the exon/intron junction at the 5' splice site is being cleaved. The 3'OH at the 3' end of the upstream exon of the 5' splice site performs a nucleophilic attack on the 3' splice site to generate splice products and to release the intronic RNA as a lariat-shaped structure (Baralle and Baralle 2005) (Figure 1.6).

A)



B)



**Figure 1.6 The process of RNA splicing**

A) The universally conserved motifs correspond to functionally important regions in splicing: the GU and AG dinucleotides spanning the 5' and 3' boundaries of the intron, the A nucleotide that serves as a branch site and the polypyrimidine tract (Py)n preceding the 3' AG are shown in two-exon pre-mRNA.

B) The mechanism of RNA splicing takes place in two catalytic steps. During step 1, the conserved A of the branch site carries out a nucleophilic attack on the G of the 5' splice site and results in the lariat structure (free exon 1 and lariat-exon 2). During step 2, the 3'OH at the 3' end of the exon 1 attacks the 3' splice site, causing the release of the intronic RNA (as a lariat-shaped structure) and spliced exon (adapted from Baralle and Baralle, 2005) (Baralle and Baralle 2005).

## 1.5 Mutation nomenclature

The descriptions of sequence variations were based on the guidelines for mutation nomenclature recommended by Human Genome Variation Society (HGVS), version 2.0, ([www.hgvs.org/mutnomen/recs.html](http://www.hgvs.org/mutnomen/recs.html)) (den Dunnen and Antonarakis 2000). Variants can be described at DNA, RNA and protein level in relation to a reference sequence, either a genomic or a coding DNA reference sequence. The letter in front of the nucleotide number indicates the type of the reference sequence used: “c.” for a coding DNA sequence (e.g. c.76A>T), “g.” for a genomic sequence, “m.” for a mitochondrial sequence, “r.” for an RNA sequence, “p.” for a protein sequence. The variants described at DNA level are in capitals, starting with a number referring to the first nucleotide affected (e.g. c.76A>T); at RNA level, the variants are in lower-case, starting with a number referring to the first nucleotide affected (e.g. r.76a>u); and at protein level, the variants are in capitals, starting with a letter referring to the first amino acid affected or three letter amino acid code (e.g. p.K76S or p.Lys76Ser), where the three letter amino acid code is preferred to described the amino acids.

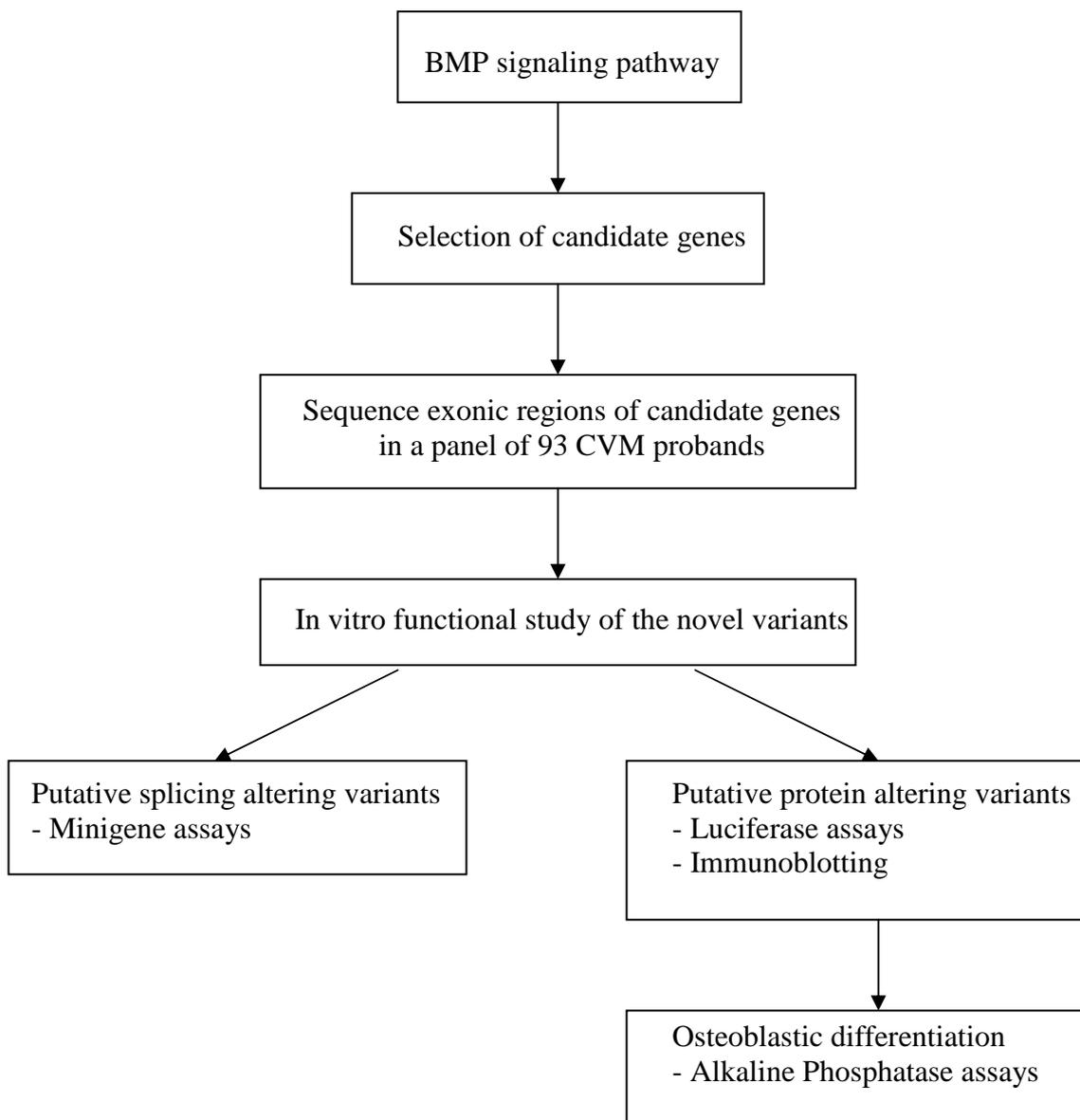
The symbol “>” indicates a substitution, e.g. c.76A>T denotes the change from the reference sequence A to a T. Deletions are designated by “del”, duplications are designated by “dup”, and insertions are designated by “ins”. The symbol “\_” (underscore) indicates a range of affected residues.

Nucleotide numbering for coding DNA sequences starts with +1, which is the A of the ATG-translation initiation codon. The nucleotide 5’ of +1 is numbered -1. There is no nucleotide zero. The nucleotide 3’ of the translation stop codon is \*1. Nucleotides at the beginning of the intron are positive numbered from the last nucleotide of the preceding exon, e.g. c.76+1C>T whereas nucleotides at the end of the intron are negative numbered upstream in the intron from the first nucleotide of the following exon, e.g. c.76-1A>G.

## 1.6 Aims and objectives of the study

The principal aim of the study was to identify and investigate the candidate genes in the BMP signaling pathway that could be involved in susceptibility to CVM in humans. Animal models of CVM have previously shown that certain genes in the BMP signaling pathway, when disrupted, cause CVM in the mouse. However, previous investigations of these genes in man either have not been previously conducted, or have been conducted in samples where parental information has been unavailable. Moreover, not all previous human mutation detection studies have involved functional investigation of identified mutations. Therefore, candidate genes were identified based on the phenotypes of the animal models in the literature. The exons of the candidate genes were then re-sequenced in a panel of non-syndromic CVM probands as it was believed that rare variants located in the protein coding regions that might affect protein function will have a large impact on disease phenotype. Parents (if available) of the probands with the newly identified novel variants were re-sequenced to establish whether the variant was inherited or was a *de novo* mutation. The variants were genotyped in an ethnically matched control population free of CVM to determine whether the variants are unique to the patient group.

Novel variants that were located at the intron/ exon boundaries and were predicted to affect splicing mechanisms were investigated using minigene assays. *In vitro* functional studies were carried out on novel variants that resulted in amino acid substitution which are predicted to be probably damaging to protein function. Figure 1.6 shows the sequential flow of the study.



**Figure 1.7 Experimental process**

Diagram illustrating the sequential steps taken in this study. Candidate genes of the BMP signalling pathway were selected for sequencing in a panel of 93 CVM probands. Novel variants identified in the patients cohort were investigated with minigene assays, luciferase assays and alkaline phosphatase assays.

## Chapter 2      Materials and Methods

### 2.1    Sample collection

#### 2.1.1    CVM cases

Patients with CVM and their parents (when available) were recruited in the Freeman Hospital, Newcastle. Ethical approval was given for this study by the Northern and Yorkshire Multicentre Research Ethics Committee. Recruited patients were of European ancestry. Each participating subject or, in the case of the minors, their parents gave informed consent. Those patients diagnosed with del22q11 syndrome, other known chromosomal abnormalities, recognised multi-organ malformation syndromes, known maternal exposure to teratogens during pregnancy and/or learning difficulties were excluded from the study. All CVM phenotypes were eligible for inclusion. Blood or saliva samples were collected. For the blood samples, DNA was extracted using the phenol/chloroform method (performed by Mr. James Eden). For the saliva samples, DNA was extracted according to the Oragene Self-Collection Kit protocol (performed by Mr. Rafiqul Hussain). DNA was then tested for del22q11 using the SALSA-MLPA kit (MRC Holland) following the manufacturer's protocol (performed by Dr. Ana Topf).

A total of 446 patients in 442 families were available for analysis in this study. Of these, 90 patients were re-sequenced for the coding sequences of the genes of interest (*BMP2*, *BMP4*, *BMPRIA*, *BMPR2* and *SMAD6*). Additional 180 patients were sequenced for *BMP4* and another 348 patients were resequenced for the MH2 domain of *SMAD6*.

An additional 357 patients in 354 families with tetralogy of Fallot (TOF) of European ancestry that were recruited by the host laboratory for the CHANGE study (Congenital Hearts: A National Gene/Environment Study), were also available for analyses focused on that specific CVM phenotype. Ethical approval was given for the CHANGE study and informed consent was obtained from patients or their parents. Blood or saliva samples of the patients were collected in Newcastle, Leeds, Bristol and Liverpool. Patients and their families that were available for analysis were of Caucasian ancestry. TOF patients diagnosed with del22q11 syndrome, other known chromosomal abnormalities and learning difficulties were excluded from the study. This population of patient was genotyped for the *SMAD6* p.Cys484Phe (c.1451G>T) variant.

### 2.1.2 Control populations

Two control populations were used in this study to exclude any novel variants detected being hitherto undiscovered moderate-frequency SNPs. Variants absent in controls were considered of greater interest for functional studies. Typically, in studies of Mendelian families, control samples of 100-200 chromosomes are used for this purpose. In complex inheritance, where the prior probability that a variant is causative of disease is substantially lower, larger control samples are needed. The number of chromosomes that must be sampled to provide support for the hypothesis that a variant detected in cases is not an uncommon polymorphism is dependent upon the allele frequency we wish to exclude at a given power, and can be calculated using the binomial distribution. Thus, if the population frequency of a variant was  $1/n$ , and  $n$  individuals were sampled, the probability that the variant would not be detected would be around 36%. However, if  $2n$  individuals were sampled, the probability of non-detection would fall to around 13% (that is, 87% power). Our study screened 3000 control chromosomes, therefore we had around 90% power to detect SNPs with an allele frequency of  $1/1500$ .

The first control population was a collection of 1425 British Caucasian individuals from 255 families that were ascertained through a proband with essential hypertension, as described by Palomino-Doza *et al.* (Palomino-Doza *et al.* 2008), who had no evidence of CVM at echocardiography. The families consisted of at least three siblings if at least one parent of the sibship was available and consisted of at least four siblings if no parent was available. Thus, screening of these families surveyed over 1000 founder chromosomes. This population was genotyped for the variants identified in *BMP2*, *BMP4* and *SMAD6* (except *SMAD6* p.Ala325Thr [c.973G>A] and *SMAD6* p.Pro415Leu [c.1244C>T]).

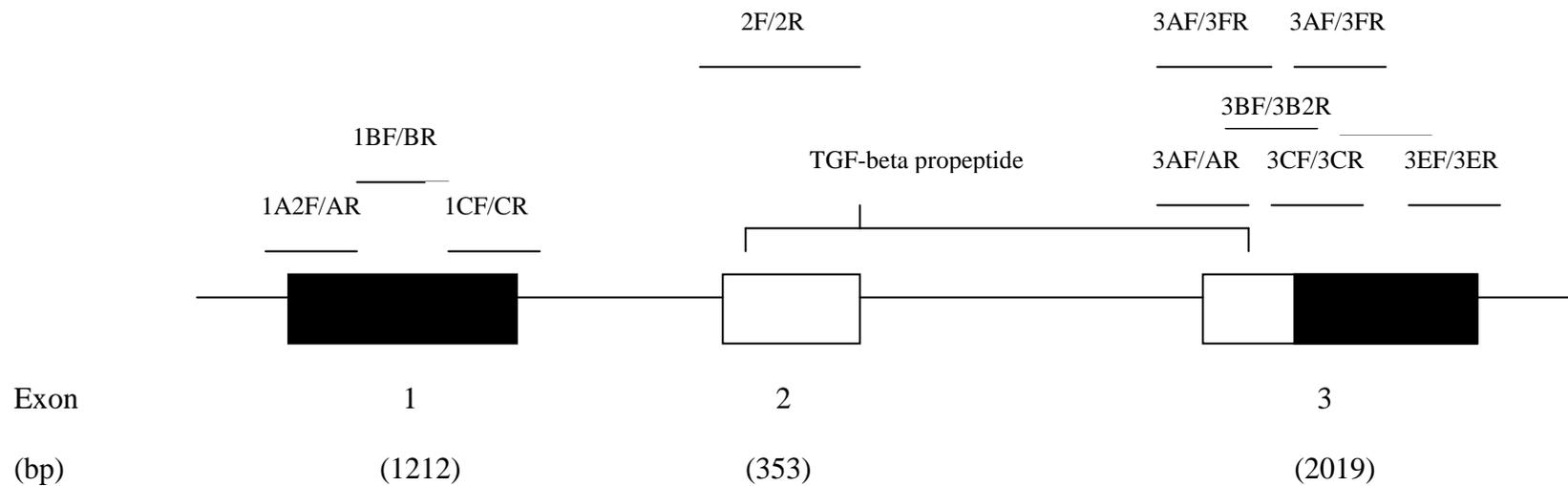
A second control population was derived from the North Cumbria Community Genetics Project (NCCGP). This population consisted of 1000 DNA samples of unrelated healthy offspring of healthy mothers born in Cumbria who were recruited between 1999 and 2002, as described by Dorak *et al.* (Dorak *et al.* 2009). This population was added in the screening of variants *SMAD6* p.Ala325Thr (c.973G>A) and *SMAD6* p.Prp415Leu (c.1244C>T) (see below). The NCCGP project was carried out to store DNA, plasma and viable cells from mother-child pairs in Cumbria for future epidemiological studies. As the population collected were healthy individuals, they have not undergone

echocardiogram, but no screened patient had been clinically diagnosed with CVM. This population was genotyped for SMAD6 p.Ala325Thr (c.973G>A) and SMAD6 p.Pro415Leu (c.1244C>T) variants (performed by Dr. Darroch Hall).

## **2.2 Screening of candidate genes**

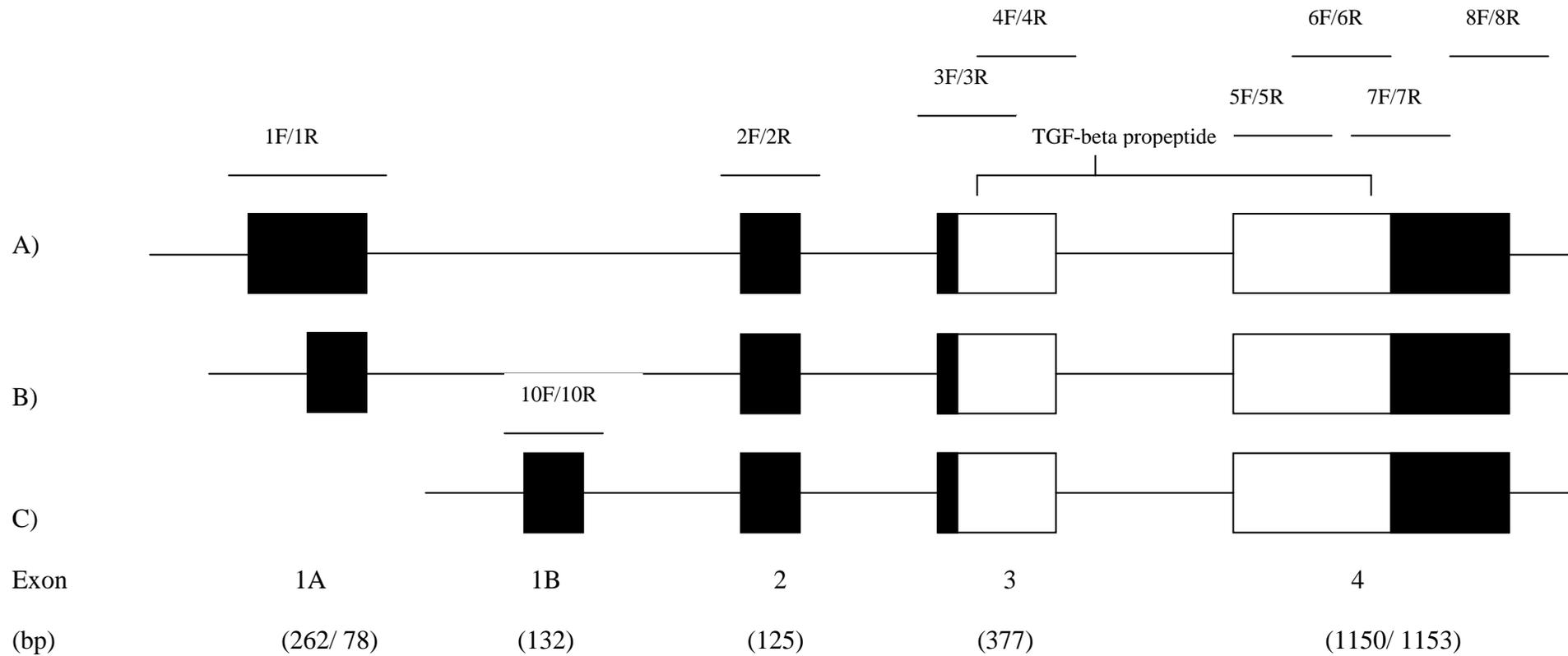
### **2.2.1 Sequencing primers**

Primers were designed to amplify the exonic and splice site regions of five genes of interest; *BMP2*, *BMP4*, *BMPRIA*, *BMPIR2* and *SMAD6* in a discovery cohort of 90 CVM patients. Sequences of the genes were obtained from Ensembl version 43 (<http://www.ensembl.org/index.html>). See Figures 2.1-2.5 for the location of the sequencing primers. Primers are listed in Table 2.1. Parameters for sequencing primers included: primers about 18-25 nucleotides long, similar GC content for both primers and no repetitive sequences. Base complementarity between the two primers was avoided to prevent primer dimer formation. Each primer was checked for specificity to the region desired using the BLAST programme (<http://www.ncbi.nlm.nih.gov>). Both primer pairs were designed approximately 80bp into the intronic sequence, so that good readable sequences at the beginning and at the end of exons could be obtained, as sequence data is often most accurate about 50bp from the primer. PCR products were usually between 350-700bp, as that was the achievable average read length of the MegaBACE sequencer. Larger exons (more than 700bp) were amplified in smaller amplicons, with overlapping 3' and 5' ends. The same primers were used for sequencing reactions.



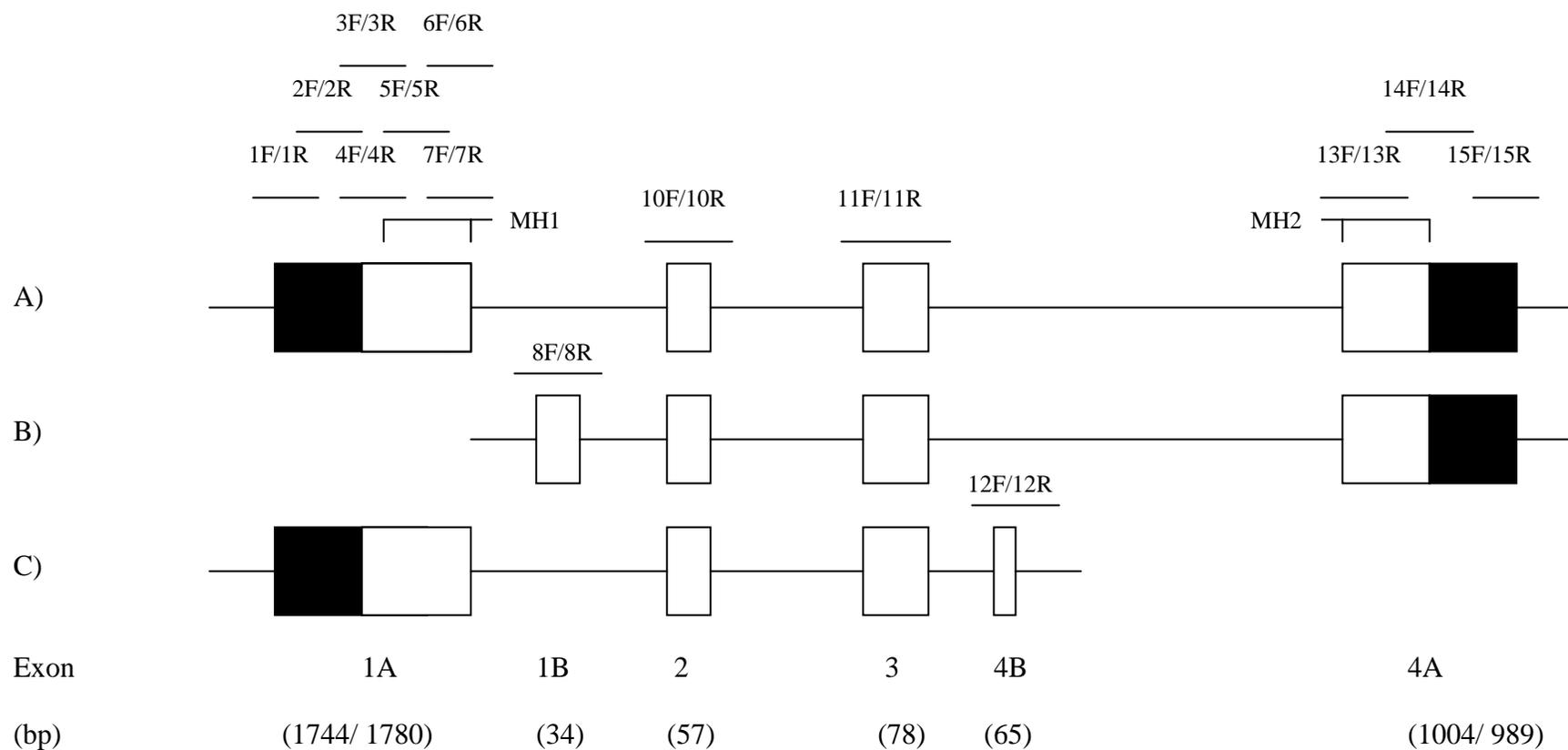
**Figure 2.1** *BMP2* Structure and Sequencing Primer Location

Diagram of BMP2 (ENST 00000378827). Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pair. Locations of the primers are located above the exons. 5'UTR and 3'UTR are in black colour. Location of the TGF-beta propeptide domain is labelled above the exons.



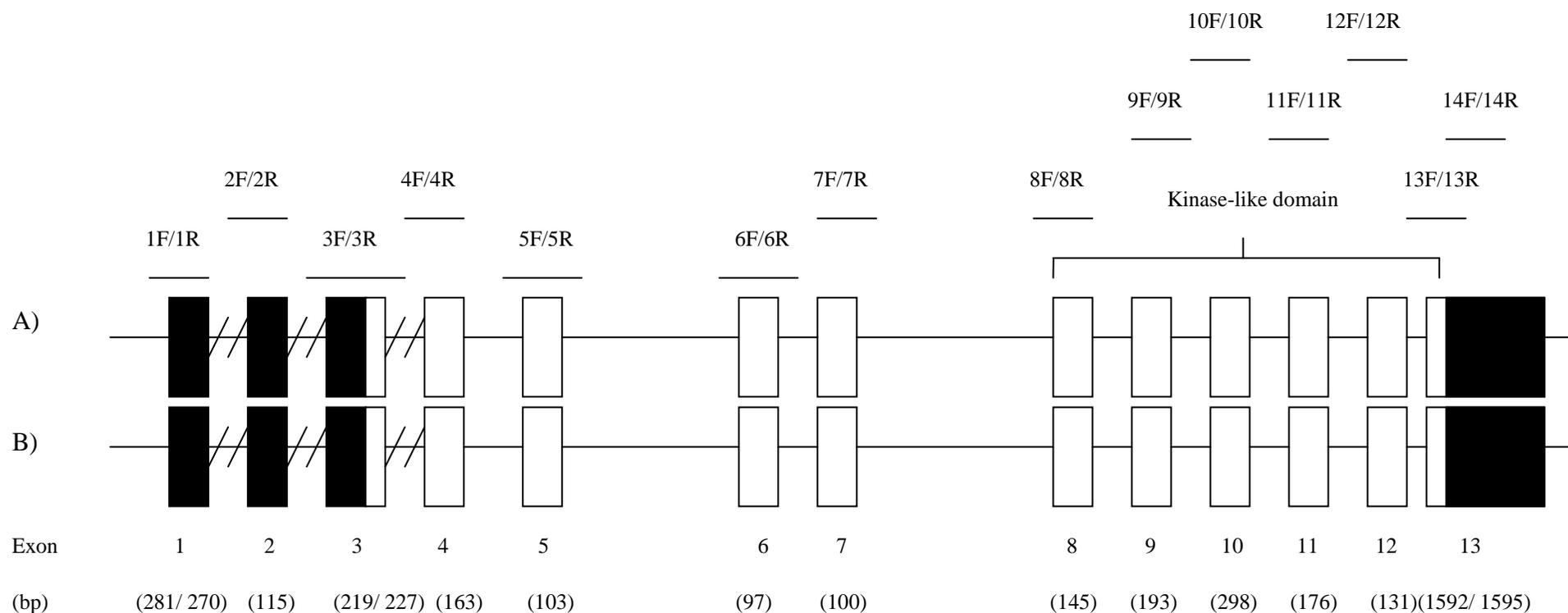
**Figure 2.2** *BMP4* Structure and Sequencing Primer Location

Diagram of *BMP4* isoforms A (ENST 00000245451), B (ENST 00000395600) and C (ENST 00000417573). Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pairs. Locations of the primers are located above the exons. 5'UTR and 3'UTR are in black colour. Location of the TGF-beta propeptide domain is labelled above the exons.



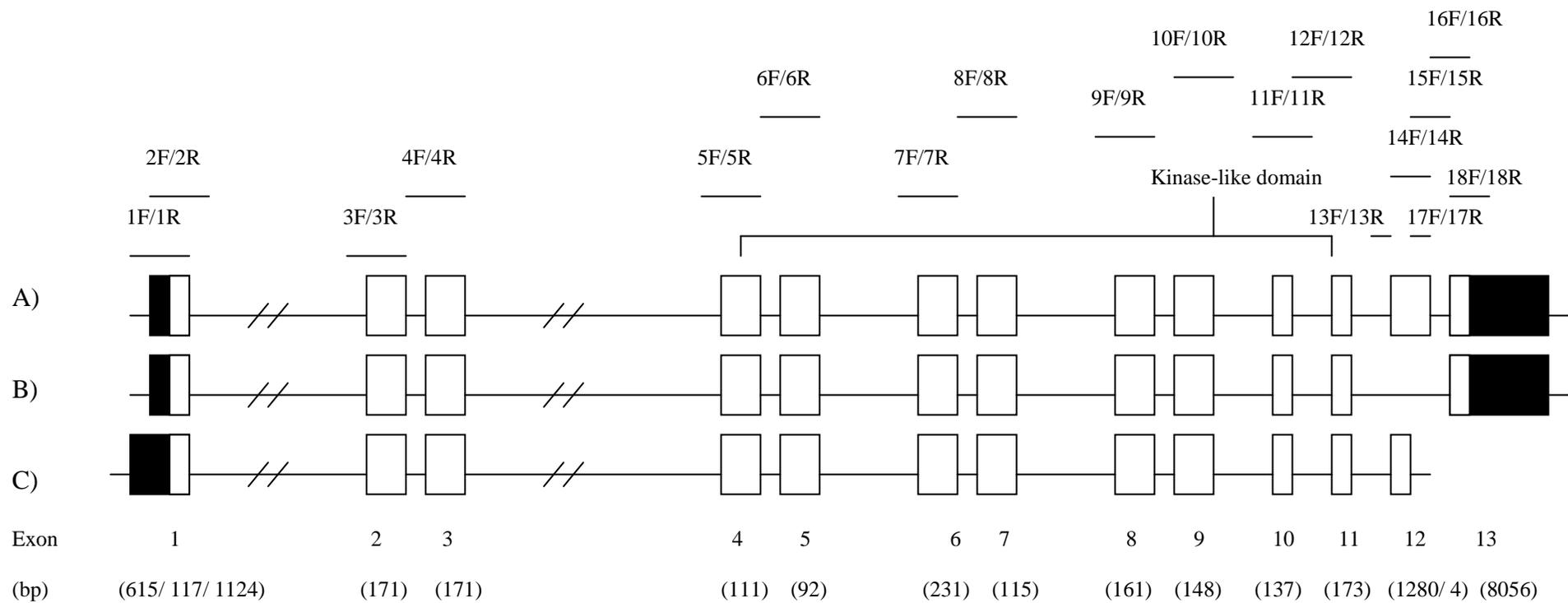
**Figure 2.3** *SMAD6* Structure and Sequencing Primer Location

Diagram of *SMAD6* isoforms A (ENST 00000288840), B (ENST 00000338426) and C (ENST 00000457357). Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pairs. Locations of the primers are located above the exons. 5'UTR and 3'UTR are in black colour. Location of the MH1 and MH2 domains are labelled above the exons.



**Figure 2.4** *BMPRIA* Structure and Sequencing Primer Location

Diagram of *BMPRIA* isoforms A (ENST 00000224764) and B (ENST 00000372037). Boxes represent exons and adjoining lines or broken lines represent introns. Exons are numbered and the sizes are indicated in base pairs. Locations of the primers are located above the exons. 5'UTR and 3'UTR are in black colour. Location of the kinase-like domain is labelled above the exons.



**Figure 2.5** *BMPR2* Structure and Sequencing Primer Location

Diagram of *BMPR2* isoforms A (ENST 00000374580), B (ENST 00000363702) and C (ENST 00000445231). Boxes represent exons and adjoining lines or broken lines represent introns. Exons are numbered and the sizes are indicated in base pairs. Locations of the primers are located above the exons. 5'UTR and 3'UTR are in black colour. Location of the kinase-like domain is labelled above the exons.

<b>Sequencing Primers</b>	<b>Sequence (5'→ 3')</b>
BMP2-1A2-F	TCCACACCCCTGCGCGCAGC
BMP2-1A-R	GCGCTAGGGATCGGCTCCGG
BMP2-1B-F	CAGCGCCGAGTGGATCACCG
BMP2-1B-R	CTGTCGTCCTCCTTAGGAACC
BMP2-1C-F	ATGGGCCGCGGCGGAGCTAG
BMP2-1C-R	CTGCAGGGCTCCGTTGCACG
BMP2-2F	GACTCGTGAGTGTGCCAAGC
BMP2-2R	GGCCAAGCTGCCTCTACAGG
BMP2-3A-F	TCCTGGTATGCAATGCATGATGAC
BMP2-3A-R	CAATGGCCTTATCTGTGACCAGC
BMP2-3B-F	TGGATTTCGTGGTGGAAGTGG
BMP2-3B2-R	AATATTAAGTGTCAACTGGGGT
BMP2-3C-F	CTATCAGGACATGGTTGTGGAGG
BMP2-3C-R	GTGAACTCAACAGTAGCACTGCAA
BMP2-3D-F	CAGCACATGAAGTATAATGGTCAG
BMP2-3D-R	TGAATACTTTGGCCAGATCAGCC
BMP2-3E-F	TTGGATAAGAACCAGACATTGCTG
BMP2-3E-R	TCCAGCCATTCATTACCTTTCCC
BMP2-3A-F	TCCTGGTATGCAATGCATGATGAC
BMP2-3F-R	CCACGTCACTGAAGTCCACG
BMP2-3G-F	ACTCAGTGCTATCTCGATGC
BMP2-3B2-R	AATATTAAGTGTCAACTGGGGT
BMP4-1F	GCCTGCTAGGCGAGGTCGG
BMP4-1R	TGCCTCGGTCACAGCCTGTG
BMP4-2F	AAGCCCTTTCGCCGCCTGC
BMP4-2R	AAGGGAGCGACAGTGCTGCC
BMP4-3F	TTGACTGTAAACAAGTTAGACC
BMP4-3R	CCGCATGTAGTCCGGAATGA
BMP4-4F	TAACCGAATGCTGATGGTCG
BMP4-4R	GGCTTTGATGTAACCCGAAC
BMP4-5F	CAGAATTCACCATTATTGCC
BMP4-5R	ATCGGCTAATCCTGACATGC
BMP4-6F	ACCACAATGTGACACGGTGG
BMP4-6R	CAGAATTGACCAGGGTCTGC
BMP4-7F	CCACTCGCTCTATGTGGACT
BMP4-7R	AAGTCCAGCTATAAGGAAGCA
BMP4-8F	CCTGGATGAGTATGATAAGGTG
BMP4-8BR	CTCCAACCTCACCTGAACAGAA
BMP4-10F	GCAGCATCTTCGATTAGTCAGG
BMP4-10R	GGCTCGAAGCTCCGGGCAG
Smad6-1F	GTCTAGACTGGCATATGATGG
Smad6-1R	TCCCTGGATTTGCATGCACG
Smad6-2F	CAGCCTCGGCCGCACGAAGC
Smad6-2R	AAGTTGTGCGGCTGGTCCATG
Smad6-3F	ATCAACAAGCCTGCCTTTCG
Smad6-3R	GAGGCGGTCTCCGGTTAC
Smad6-4BF	GCTCGGCGTGCGCGTGTCC
Smad6-4BR	ACTTCGGAGCGGCCGCAGCC

Smad6-5F	GGATATCGTATGTTTCAGGTCCA
Smad6-5R	AGCGCGCTTCGCGACTCCG
Smad6-6F	AGGCCCATGTTCGGAGCCAG
Smad6-6R	CACGGCGTGCTGCAGGTTCG
Smad6-7F	AACCGTCACGTACTCGCTGC
Smad6-7R	TGCGCCGACCCGCAAGTTC
Smad6-8F	CACCAGTCCAGATCCACAGG
Smad6-8R	ACCACCATCTAAACACAAGACA
Smad6-10F	GGCTTTAACTGTTGAGGGGA
Smad6-10R	CACAGCAACCAGTCCGGCT
Smad6-11F	GGTGACATCAGAAAGTTGGG
Smad6-11R	GCAACCTTCCCATCCAAGGA
Smad6-12F	ATGTGAAGGGTCAGAAGATTG
Smad6-12R	CACAGCCATCCAGACATGC
Smad6-13F	ACATGACTGCTGAATGCTGG
Smad6-13R	GAACACCTTGATGGAGTAGC
Smad6-14F	GAACGCGCAGCAAGATCGG
Smad6-14R	TCCATATACACACATTGCACC
Smad6-15F	CCGATGCCCAGAGACACAG
Smad6-15R	AACTGAGGTCAAGTTTCAAGG
BMPR1A-1F	CTTCCCAGTGCGCGCGCGCC
BMPR1A-1BR	CGCGCCCAGCTGCCTCTGCA
BMPR1A-2F	ATAGTTGTACATGCGAAACG
BMPR1A-2R	ATTACTTATGAACTCGTCAGG
BMPR1A-3F	AGCAAGGATACCTTTAATCTT
BMPR1A-3R	AGGCTGGGCCTAACTATTC
BMPR1A-4F	AGAAGAATCATAACTTGTCTC
BMPR1A-4R	CATGAAGACAGCTATATCAC
BMPR1A-5F	GCTACAATTATTGTGATAAGAA
BMPR1A-5R	AGTAACTCACCTATCAATCAT
BMPR1A-6F	GGCGTTGCCAATAAATCACG
BMPR1A-6R	ACATACAGCGATGTGTACAC
BMPR1A-7F	CCATCCACATAGATATGTTC
BMPR1A-7R	AATAATCTGACGTGTTCAATTC
BMPR1A-8F	AGTCGGAGCATGCTTCTCAA
BMPR1A-8R	CCTTAAGAGAACTATAGTCTAG
BMPR1A-9F	AACATGATACATAGAACTCTG
BMPR1A-9R	ACCTTGCAATACTGCATACC
BMPR1A-10F	AACTGGAGTTGGTTGGGTAC
BMPR1A-10R	GTATATAGTTGAACTCACACAT
BMPR1A-11F	TATCTGCACATGATACCTAAG
BMPR1A-11R	CCAACTATACATTTATGTA ACTA
BMPR1A-12F	GATGTCATGTAGTATGTAGCA
BMPR1A-12R	AAAGTTATATAACTGGAGCTG
BMPR1A-13F	ACATACATCTACTATTAAGAG
BMPR1A-13R	CTCGTAGACACTGAAAAGTA
BMPR1A-14F	GTGATGAAGTGAGTGGA ACT
BMPR1A-14R	TGAGAACCAAGTTAACATCC
BMPR2-1F	GACTGTGAGCTTGTCCATGG
BMPR2-1R	TCTTCTCTCCCGTGGACTG

BMPR2-2F	TGATATCGTGAAACTACGAGG
BMPR2-2R	ATGGCGAAGGGCAAGCACA
BMPR2-3F	TCATGAACAGAAGAACGTCAT
BMPR2-3R	AACAGGATTTTAACATACTCC
BMPR2-4F	CATAGCTTACACGTACTCTC
BMPR2-4R	ACGCCTGGCTTCAACCTTG
BMPR2-5AF	AGGAGCACATCTACTTGGTG
BMPR2-5AR	TAATCCAGTGGCATGGAAAG
BMPR2-5BF	ATCTATACAGAATAGTCTGTATG
BMPR2-5BR	ACAATACTGTCCATACGTGAT
BMPR2-6F	TTACTCCTATTGACATTAGGC
BMPR2-6R	TGCCTAGAATAGGCCTTGAC
BMPR2-7F	TCAGCCATACTAGAACAGAAT
BMPR2-7R	CACCTGCCTTAGCCTCCA
BMPR2-8F	TAGATCTTCATGGAATCCTAG
BMPR2-8R	CAACTGACTAATAATAAATCTCC
BMPR2-9BF	CTGAAGTGGCAGCATGTTTGT
BMPR2-9AR	CAAAGTACATCAGTGTGATACC
BMPR2-10F	TAATGACATGGTTAGGGTCAA
BMPR2-10R	CTTACATAAAAGTTGAGTTAGG
BMPR2-11F	GCTTACTTGGTATCAGAAATAC
BMPR2-11R	AACTATTAACGTTATTAACAGTC
BMPR2-12F	AGCATGTTCCGTAATCCTTG
BMPR2-12R	TCATTGAACTATTAGGCTGGT
BMPR2-13F	TATCAAGTTATTAATTGACACTTG
BMPR2-13R	CTCAGATATAGTAGTCATGCC
BMPR2-14F	CATCATACTGACAGCATCGTG
BMPR2-14R	TTGCAGTCTGTGTGAAGTCC
BMPR2-15F	TGTCTGCTTACAGCTGACAG
BMPR2-15R	TTGGCCAGATAGTACTGTCC
BMPR2-16F	CAAGAGACCTACTAGTTTGC
BMPR2-16R	GAATTAGTTCGGCCACCTTC
BMPR2-17F	CTGCCACAACCCAATATGC
BMPR2-17R	AAGGTAAATAATCACTAGTTGA
BMPR2-18F	CAAGAAATGTTTATTAGCACC
BMPR2-18R	TTCTGCATAGCTGAAATCTAT

**Table 2.1 Primers used for sequencing**

Sequences of the primers used to sequence the coding regions of *BMP2*, *BMP4*, *BMPRI1A*, *BMPI2* and *SMAD6*.

### 2.2.2 Polymerase Chain Reactions (PCR)

PCR were performed using HotStarTaq<sup>TM</sup> DNA Polymerase (Qiagen). PCR was conducted in a 15µl reaction mixture which comprised of 1x PCR Buffer, 1.5mM MgCl<sub>2</sub> or 1x Q solution, 0.2mM of each dNTPs, 0.5µM of each primer (forward and reverse), 2.5units of HotStarTaq DNA Polymerase, 20ng of DNA and ddH<sub>2</sub>O to a final volume of 15µl.

The PCR cycling programme started at 95°C for 15 minutes, to heat-activate the HotStarTaq DNA Polymerase, followed by 35 cycles of 95°C (or 97°C) for 30 seconds, the primer set specific annealing temperature for 30 seconds and 72°C for 1 minute, and a final extension of 5 minutes at 72°C.

In order to identify the optimum conditions for each of the primer sets, PCR were carried out in a gradient temperature from 50°C–70°C. PCR reactions were first tested with 1.5mM of MgCl<sub>2</sub>. If that failed to obtain a clear PCR band, Q solution was utilized. For regions with high GC content (60% or more), the denaturing temperature was increased to 97°C. The PCR conditions for each primer set are detailed in Table 2.2.

<b>Primer Pair</b>	<b>Denaturing Temp. (°C)</b>	<b>Annealing Temp. (°C)</b>	<b>MgCl<sub>2</sub> Concentration (mM)/ Q solution</b>
BMP2_1A2F/1AR	97	64.6	Q solution
BMP2_1BF/1BR	95	61.8	Q solution
BMP2_1CF/1CR	95	66.8	Q solution
BMP2_2F/2R	95	66.8	1.5
BMP2_3AF/3AR	95	68.4	1.5
BMP2_3BF/3B2R	95	61.8	1.5
BMP2_3CF/3CR	95	67.6	1.5
BMP2_3DF/3DR	95	66.8	1.5
BMP2_3EF/ER	95	61.8	1.5
BMP2_3AF/3FR	95	61.8	1.5
BMP2_3GF/3B2R	95	61.8	1.7
BMP4_1F/1R	95	68.4	1.5
BMP4_2F/2R	95	68.4	1.5
BMP4_3F/3R	95	61.8	1.5
BMP4_4F/4R	95	61.8	1.5
BMP4_5F/5R	95	58.4	1.5
BMP4_6F/6R	95	61.8	1.5
BMP4_7F/7R	95	58.4	1.5
BMP4_8F/8BR	95	61.8	1.5
BMP4_10F/10R	95	61.8	Q solution
Smad6_1F/1R	97	61.8	1.5
Smad6_2F/2R	95	69.6	1.5
Smad6_3F/3R	97	58.4	Q solution
Smad6_4BF/4BR	97	61.8	Q solution
Smad6_5F/5R	97	58.4	Q solution
Smad6_6F/6R	97	58.4	Q solution
Smad6_7F/7R	97	58.4	Q solution
Smad6_8F/8R	95	61.8	1.5
Smad6_10F/10R	95	61.8	1.5
Smad6_11F/11R	95	61.8	1.5
Smad6_12F/12R	95	64.6	1.5
Smad6_13F/13R	95	61.8	1.5
Smad6_14F/14R	95	61.8	1.5
Smad6_15F/15R	95	66.8	1.5
BMPR1A_1F/1BR	97	64.6	1.5
BMPR1A_2F/2R	95	58.4	1.5
BMPR1A_3F/3R	95	58.4	1.5
BMPR1A_4F/4R	95	58.4	1.5
BMPR1A_5F/5R	95	58.4	1.5
BMPR1A_6F/6R	95	58.4	1.5
BMPR1A_7F/7R	95	58.4	1.5
BMPR1A_8F/8R	95	58.4	1.5
BMPR1A_9F/9R	95	58.4	1.5
BMPR1A_10F/10R	95	58.4	1.5
BMPR1A_11F/11R	95	58.4	1.5
BMPR1A_12F/12R	95	58.4	1.5
BMPR1A_13F/13R	95	58.4	1.5
BMPR1A_14F/14R	95	58.4	1.5

BMPR2_1F/1R	95	61.8	1.5
BMPR2_2F/2R	95	58.4	Q solution
BMPR2_3F/3R	95	58.4	1.5
BMPR2_4F/4R	95	58.4	1.5
BMPR2_5AF/5BR	95	58.4	1.5
BMPR2_6F/6R	95	58.4	1.5
BMPR2_7F/7R	95	55.5	Q solution
BMPR2_8F/8R	95	58.4	1.5
BMPR2_9BF/9AR	95	58.4	1.5
BMPR2_10F/10R	95	58.4	1.5
BMPR2_11F/11R	95	58.4	1.5
BMPR2_12F/12R	95	58.4	1.5
BMPR2_13F/13R	95	58.4	1.5
BMPR2_14F/14R	95	58.4	1.5
BMPR2_15F/15R	95	58.4	1.5
BMPR2_16F/16R	95	58.4	1.5
BMPR2_17F/17R	95	58.4	1.5
BMPR2_18F/18R	95	58.4	1.5

**Table 2.2 PCR Conditions**

The optimum PCR conditions used for each of the primer sets.

### **2.2.3 Gel electrophoresis**

2% gels were prepared using SeaKem® LE Agarose (Cambrex), 1x TAE buffer, and 1x Gel Red (Biotium). 5µl of PCR product was loaded with 5µl of Orange G dye, alongside 2.5µl of 100bp DNA Ladder (Fermentas) and were run at 150V for 50 minutes. Gels were then visualized under UV light to check the size and approximate quantification of the amplified products.

### **2.2.4 PCR clean up**

ExoSAP-IT (GE Healthcare) was utilised to clean up PCR products by removing excessive dNTPs and primers. 2µl of ExoSAP-IT which contains two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase was added to 5µl of PCR product. This was then incubated at 37°C for 15 minutes followed by inactivation at 80°C for 15 minutes.

### **2.2.5 Sequencing reactions**

Sequencing was carried out using the DYEnamic ET dye terminator cycle sequencing kit for MegaBace DNA analysis systems (GE Healthcare). Reactions were conducted in a 96-well skirted plate and each reaction comprised 3µl of cleaned-up PCR product, 5pmol of forward or reverse primer, 4µl of DYEnamic ET terminator reagent premix and ddH<sub>2</sub>O to a final volume of 10µl. For regions with high GC content, a final concentration of 0.5M Betaine or 5% DMSO was added to the reaction.

The cycling programme started with 95°C for 2 minutes, followed by 35 cycles of 95°C for 30 seconds, 55°C for 15 seconds, and 60°C for 1 minute. Products were cleaned-up by isopropanol precipitation. 27µl of 80% isopropanol was added to each reaction, vortexed briefly, incubated for a minimum of 30 minutes at room temperature and centrifuged at 4°C for 30 minutes at 3313g. The supernatant was discarded by inverting and gently shaking the plate. The DNA pellets were washed in 100µl of 60% ice-cold ethanol and centrifuged for 10 minutes at 3313g. The supernatant was removed by inverting the plate and centrifuged upside down on tissue for 10 seconds at 200g, then

was air-dried for 5 minutes. The DNA pellets were re-suspended in 10µl of MegaBACE loading solution and run on a MegaBACE 1000 sequencing instrument (GE Healthcare) by the sequencing service at the Institute of Human Genetics (Newcastle University, UK).

### **2.2.6 Sequence analysis**

Sequences were analyzed using Staden Package software (<http://staden.sourceforge.net/>). Forward and reverse sequences were compiled against a text file of consensus sequence taken from ENSEMBL. Using the Pregap4 programme, sequences were pre-processed for Gap4 using Gap4 shotgun assembly configuration. Sequences were then aligned using Gap4 with a maximum mismatch of 20%. Single nucleotide polymorphisms (SNPs) and novel variations were detected by comparing all sequences together.

### **2.2.7 Identifying variants in parents**

Regions where novel variants were identified in probands were re-sequenced in their parents (when available) to determine whether the variants were inherited or *de novo*.

### **2.2.8 Genotyping rare variants**

Newly identified variants were genotyped in the control population. This was carried out either by Sequenom or Taqman methods (see below). Positive controls were included in the assay to confirm the reliability of the assay and the presence of the variants. Frequencies of the alleles identified in the controls and probands were calculated by genotype counting. Goodness of fit to Hardy Weinberg equilibrium was tested. The frequencies of the variants that were present in the controls were compared to the frequencies of the alleles in the probands, and the hypothesis that these were different in the two populations was tested using Fisher's exact test in Minitab software.

### 2.2.8.1 Sequenom genotyping

Multiplex assays were designed using the SEQUENOM RealSNP™ Programme (<https://www.realsnp.com/default.asp>) and Assay Design software. SNPs with 200bp of surrounding sequences were prepared in a text file and then uploaded into the ProxSNP programme to map for proximal SNPs that could interfere with assay design. Using the PreXTEND programme, PCR primers were designed and checked for their specificity to the desired regions. The output file was then run on the PleXTEND programme, to ensure the specificity of the design. Three plexes (15-plex, 11-plex and 5-plex) were designed to accommodate 31 novel genetic variants. The assay designs and the primer sequences are listed in Table 2.3.

10µl of multiplex reactions were set up using HotStarTaq™ DNA Polymerase (Qiagen). Multiplex PCR reactions consisted of 1.25x of HotStarTaq Buffer, 1.63mM of MgCl<sub>2</sub>, 0.5mM of dNTPs, 0.1µM of each primer, 1unit of HotStarTaq DNA Polymerase, 20ng of genomic DNA, pool of primer sets (primer PCR\_P1 and primer PCR\_P2) and ddH<sub>2</sub>O. This procedure amplified many individual loci of specific fragments of genomic DNA.

The PCR cycling programme consisted of an initial step at 95°C for 15 minutes, followed by 35 cycles of 97°C for 20 seconds, 59°C for 30 seconds and 72°C for 1minute and a final extension at 72°C for 3 minutes.

5µl of each PCR product were loaded into a 384-well plate (ABgene) and treated with Shrimp Alkaline Phosphatase (SAP) to dephosphorylate unincorporated dNTPs. SAP mixture was made up of 1x SAP buffer (Sequenom®), 1unit of SAP enzyme (Sequenom®) and ddH<sub>2</sub>O made up to 2µl. The reactions were gently vortexed and incubated at 37°C for 20 minutes and 85°C for 5 minutes in a thermocycler.

iPLEX extension reaction cocktail was prepared according to the number of primers in the plexes. Primers (extension primer [EXT]) were divided into 4 groups dependent on their mass: 0.625µM, 0.83µM, 1.04µM, and 1.25µM. Briefly, the mixture for each reaction was made up of 0.222x of iPLEX buffer (Sequenom®), 1x of iPLEX termination mix (Sequenom®), primer mix, 1x iPLEX enzyme (Sequenom®), and ddH<sub>2</sub>O to a final volume of 2µl. The iPLEX reaction was carried out to detect single-base polymorphisms or small insertion/deletion polymorphisms in amplified DNA.

During the iPLEX reaction, the primer is extended by one mass-modified nucleotide depending on the allele and the design of the assay.

The iPLEX reaction was cycled at 94°C for 30 seconds, 94°C for 5 seconds, 5 cycles of 52°C for 5 seconds and 80°C for 5 seconds, followed by another 39 cycles of 94°C for 5 seconds, 52°C for 5 seconds, and 80°C for 5 seconds. The cycle concluded with a final extension at 72°C for 3 minutes.

iPLEX reactions were desalted with 6mg of resin. 16µl of ddH<sub>2</sub>O was added to the reaction mix. The plate was inverted for 30minutes before centrifugation at 1864g for 10 minutes and stored at -20°C.

The 384-element SpectroCHIP bioarray and the operation of the MALDI-TOF machine were performed by the Sequenom facility at the Institute of Human Genetics (Newcastle University). The readout was viewed using MassARRAY Typer v3.4 software.

### 15-Plex Primers

SNP_ID	PCR_P1 Primer Sequence (5' → 3')	PCR_P2 Primer Sequence (5' → 3')	EXT Primer Sequence (5' → 3')
BMP4_7551C_T	ACGTTGGATGGTCCAGCTATAAGGAAGCAG	ACGTTGGATGACATACACCACACACACACG	GTTCCCATCCACTCACC
Smad6_80102T_C	ACGTTGGATGGAGGAAGGAGGAAGAGAAAAG	ACGTTGGATGGGACAAAACAAGAAAGACGC	AGACGCACTTTGGCTTA
BMP4_5348T_C	ACGTTGGATGGGTCTGTTTTATTATGCCAAG	ACGTTGGATGCGACTTTTTTCTTCCCCGTC	cCCCCGTCTCAGGTATCA
BMP4_4140C_T	ACGTTGGATGCAAACCTTGCTGGAAAGGCTC	ACGTTGGATGTCCTTTTAGGAGCCATTCCG	AGCCATTCCGTAGTGCCAT
Smad6_79800G_T	ACGTTGGATGCACTATCTGGGGTTGTTGAG	ACGTTGGATGTGCTACTCCCGGCAGTTCAT	tctGCAGTTCATCACCTCCT
Smad6_1520C_T	ACGTTGGATGACATACGATATCCTTTGGCG	ACGTTGGATGCCTCGCTGAGGGAACGGAC	AGGGAACGGACCCCCGGTAA
BMP2_11292G_A	ACGTTGGATGACAATCCAGTCATTCCACCC	ACGTTGGATGCGGAAACGCCTTAAGTCCAG	cctCCTTAAGTCCAGCTGTAA
BMP4_5619C_T	ACGTTGGATGAGCACTGGTCTTGAGTATCC	ACGTTGGATGAGAGACTGACCTTCGTGGTG	gttGTGGAAGCTCCTCACGGT
BMP4_3036C_T	ACGTTGGATGGAGTCGAGAAAACGGTGAAC	ACGTTGGATGAATAAGGGAAGCCGAGGCGA	acgaAGAGGTTCGAGCGCAGGC
BMP4_3071C_A	ACGTTGGATGTTACCGTTTTCTCGACTCC	ACGTTGGATGTGTGGCCGAGGTTTTACAAC	cgccGAAAGGAAATCCCACCAT
BMP2_12528delATTT	ACGTTGGATGAAATTGATATCTCGTGGCCC	ACGTTGGATGGCTCTCAGATGAGGTAAAGG	tAGGTAAAGGTAAAACACAAAT
Smad6_2274C_T	ACGTTGGATGAAGCTGTGGCAGCCGCAC	ACGTTGGATGTGCTGCTCGGCCGCTCTTT	aagaGCTGGCCCCGACCTGCAGCA
BMP4_795G_C	ACGTTGGATGTGAGGGACGCGAGCCTGAGA	ACGTTGGATGTTGGACGGGAATCCCATCG	aatctGGGAGACAAGCTAGATACT
Smad6_15336C_G	ACGTTGGATGCAGTGGTTTTCTGTGTCTCTC	ACGTTGGATGGACAGAGGACTCCAACCAC	cccaaTCCAACCACGGTTCCTCG
BMP4_2912G_C	ACGTTGGATGGAAGGAGTATTGCTGCGTG	ACGTTGGATGTTCTTCCCCACCCCCTTC	gaaCCTTCTCCGTTGCACCAGCAG

### 11-Plex Primers

SNP_ID	PCR_P1 Primer Sequence (5' → 3')	PCR_P2 Primer Sequence (5' → 3')	EXT Primer Sequence (5' → 3')
Smad6_1274G_C	ACGTTGGATGAGCGGCTACATGGACCACAG	ACGTTGGATGCCGCCTGCAGCCCCCTA	TCAACAACCTCCAGCCCC
BMP4_2912G_A	ACGTTGGATGTTCTTCCCCACCCCCTTC	ACGTTGGATGGAAGGAGTATTGCTGCGTG	CTCCGTTGCACCAGCAG
Smad6_1624G_A	ACGTTGGATGGACTTTGGCGAAGTCGTGTG	ACGTTGGATGTCCCATCTCGTCGCCGCCA	CGAAGTCGTGTGGTCCCC
BMP4_6756G_A	ACGTTGGATGGGGGCTTCATAACCTCATA	ACGTTGGATGTGGACCAGGGCCCTGATTG	TGTTTATACGGTGAAGC
BMP4_5670G_A	ACGTTGGATGTTTCACTGCCCTCTAGCCA	ACGTTGGATGTTCCACCACGAAGGTCAGTC	CCAGTCCCACCAGCCCTCC
BMP2_1044delTGT	ACGTTGGATGTTTCTCCCTTCCCTTCCCTGC	ACGTTGGATGAAGAAGTCCCAGCCAAGTG	TCCCCCTGCTCGCTGTTGT
BMP4_3071C_T	ACGTTGGATGTGTGGCCGAGGTTTTACAAC	ACGTTGGATGTTACCGTTTTCTCGACTCC	ccGAAAGGAAATCCCACCAT
BMP2_480C_T	ACGTTGGATGGGATCGGCTCCGGAGTCCC	ACGTTGGATGAGCGCCGTGGCCTCTGCT	TCCCCGGCGCTCTGCGCGCCC
BMP2_11547A_G	ACGTTGGATGTCTCGATGCTGTACCTTGAC	ACGTTGGATGACACCACAACCTCCACAA	CGAGAATGAAAAGGTTGTATT
BMP4_4140C_T	ACGTTGGATGGAACAAACCTTGCTGGAAAGG	ACGTTGGATGTTAGGAGCCATTCCGTAGTG	ccACTTGCTGGAAAGGCTCAGG
Smad6_1136A_T	ACGTTGGATGAAAGCCCATGTAGTTAAGCG	ACGTTGGATGGAACCCCTAAATGTGTCTGG	GGTTTAAAAAAAAGGCAAGGTA

### 5-Plex Primers

SNP_ID	PCR_P1 Primer Sequence (5' → 3')	PCR_P2 Primer Sequence (5' → 3')	EXT Primer Sequence (5' → 3')
Smad6_1683C_T	ACGTTGGATGTGGCGGCGACGAGGATGGGA	ACGTTGGATGTGGCGGCGACGAGGATGGGA	GAGGATGGGAGCTTGGG
Smad6_80159T_A	ACGTTGGATGTCTCTTCCTCCTTCCTCTTC	ACGTTGGATGTCTCTTCCTCCTTCCTCTTC	TCCTTCCTCTTCCTTACTTT
BMP2_713G_A	ACGTTGGATGCTCGACTCGCCGGAGAATG	ACGTTGGATGCTCGACTCGCCGGAGAATG	GAATGCGCCCGAGGACGACG
BMP4_3036C_G	ACGTTGGATGAATAAGGGAAGCCGAGGCGA	ACGTTGGATGAATAAGGGAAGCCGAGGCGA	acgaAGAGGTTCGAGCGCAGGC
BMP4_7611A_T	ACGTTGGATGGGTCAAGGTGAATGTTTAGG	ACGTTGGATGGGTCAAGGTGAATGTTTAGG	TGAATGTTTAGGATTTTTTCC

**Table 2.3 Primers used for Sequenom assay.**

PCR\_P1 primers and PCR P2 primers were used to amplify the fragments of interest and extension primers (EXT primer) for iPLEX reaction to identify the SNP alleles.

### 2.2.8.2 Taqman genotyping

SMAD6 c.1451G>T was genotyped using a Custom designed TaqMan® SNP Genotyping Assay (Applied Biosystems), following the manufacturer's design and ordering guide. A file of the SNP and about 300bp of sequence surrounding the SNP was prepared and verified that met the manufacturer's criteria for allele frequency, length, accuracy and uniqueness. The file was then uploaded into Filebuilder software (Applied Biosystems) and submitted to Applied Biosystems for Custom TaqMan® Assay Design. The sequences for the primers and probes are listed in Table 2.4.

The reactions were conducted in a 384-well skirted plate (ABgene). Each well contained a single 5µl reaction which consisted of 1x TaqMan® Universal PCR Master Mix, with no AmpErase® UNG (Applied Biosystems), 1x Assay Mix (Applied Biosystems), 20ng of DNA and ddH<sub>2</sub>O.

The thermal cycler conditions started with 95°C for 10 minutes, followed by 40 cycles of 92°C for 15 seconds and 60°C for 1 minute. The plates were then read with a 7900HT Fast Real-Time PCR System (Applied Biosystems).

Primers/ Probes	5'→ 3' Sequence
S6_79800GT-S6GTForward	AGCGTCCGCATCAGCTT
S6_79800GT-S6GTRreverse	GGCCCGCCACTATCTGG
S6_79800GT-S6GTV1_VIC	ACCTCCTGCCCCTGC
S6_79800GT-S6GTM1_FAM	CACCTCCTTCCCCTGC

**Table 2.4 Primers and probes for TaqMan assay**

Sequences of the PCR primers and probes used for the TaqMan assay to genotype SMAD6 c.1451G>T variant.

### 2.3 *In silico* analysis of novel rare variants

Non-synonymous genetic variants were analysed for evolutionary conservation using ClustalW2 programme from EMBL-EBI (European Bioinformatics Institute) (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Protein sequences of the organisms were taken from Ensembl and were uploaded into the programme for protein alignments.

The potential functional effect of the non-synonymous genetic variants were analyzed with the SIFT (<http://blocks.fhcrc.org/sift/SIFT.html>) and PolyPhen (<http://genetics.bwh.harvard.edu/pph/>) programmes to predict whether the substitutions were likely to affect protein structure and subsequently protein function.

The splicing effects for all of the genetic variants were predicted by the ESE finder 3.0 programme (<http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home>), to identify whether they were likely to result in changes of splice sites (Cartegni et al. 2003; Smith et al. 2006). This program predicts whether variants increase or decrease the threshold value in the splice donor site, splice acceptor site and branch site.

Another in-house programme generated by Dr. S. Grellscheid (Grellscheid, unpublished) was utilized to predict whether the genetic variants were likely to result in changes of splice sites. This programme plots the calculated strength of exonic splice enhancer (ESE) versus exonic splice silencer (ESS) sites using a custom Python script to generate a z-score for each nucleotide position for the wild type and mutant variants. These are compared graphically to estimate the effect of the variants on splicing.

Analysis of the crystal structure of the SMAD6 C484F variant was carried out in collaboration with Professor Rick Lewis of Institute for Cell and Molecular Biosciences, Newcastle University. Homology modelling of the MH2 domain of SMAD6 was implemented using the ‘Coot’ software package using SMAD2 as a template.

Variants that were identified in the 3'UTR of the candidate genes were checked for their miRNA target binding sites and expressions using MicroCosm Targets version 5 (formerly miRBase Targets) <http://www.ebi.ac.uk/enright->

[srv/microcosm/htdocs/targets/v5/](http://srv/microcosm/htdocs/targets/v5/) that contains computationally predicted targets for microRNAs in many species.

## 2.4 Investigation of the effects of novel rare variants on splicing

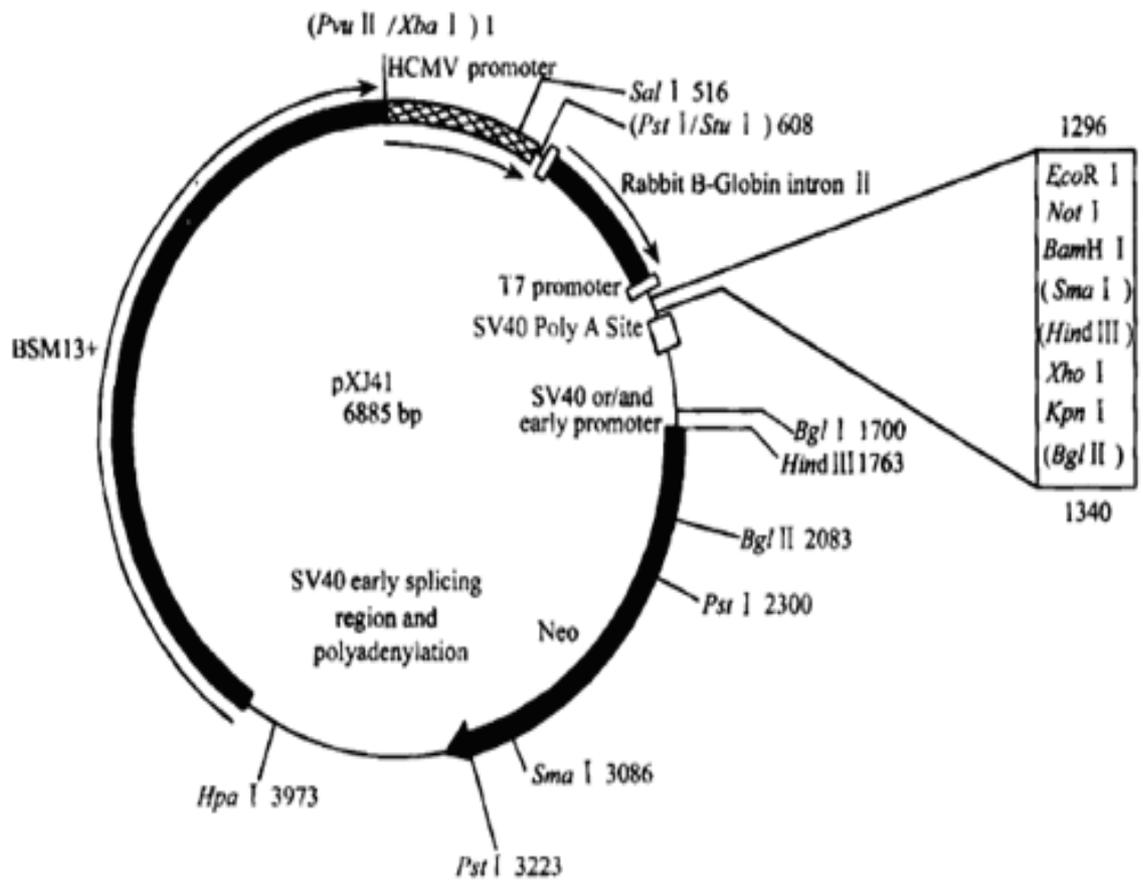
### 2.4.1 Experimental design

Minigene approach was utilized to investigate the effects of certain novel variants predicted to affect splicing. The minigene approach consists of inserting an exon of interest, namely the one in which the novel variant was found, and its flanking introns into a vector containing a constitutive small 'gene' and thus creating a new artificial minigene which includes the relevant exon.

The pXJ41 vector was used as backbone for the minigene construct. pXJ41 harbours a human cytomegalovirus (hCMV) promoter that drives the expression of the inserted gene; rabbit  $\beta$ -globin intron II containing an *Mfe*I site, which facilitates the splicing event and is flanked by fragments of rabbit  $\beta$ -globin exon 2 and exon 3; T7 promoter; a multiple cloning site (MCS); SV40 polyadenylation site; and BSM13+ sequence that include an origin of replication and the ampicillin resistance gene (Xiao et al. 1991) (Figure 2.6).

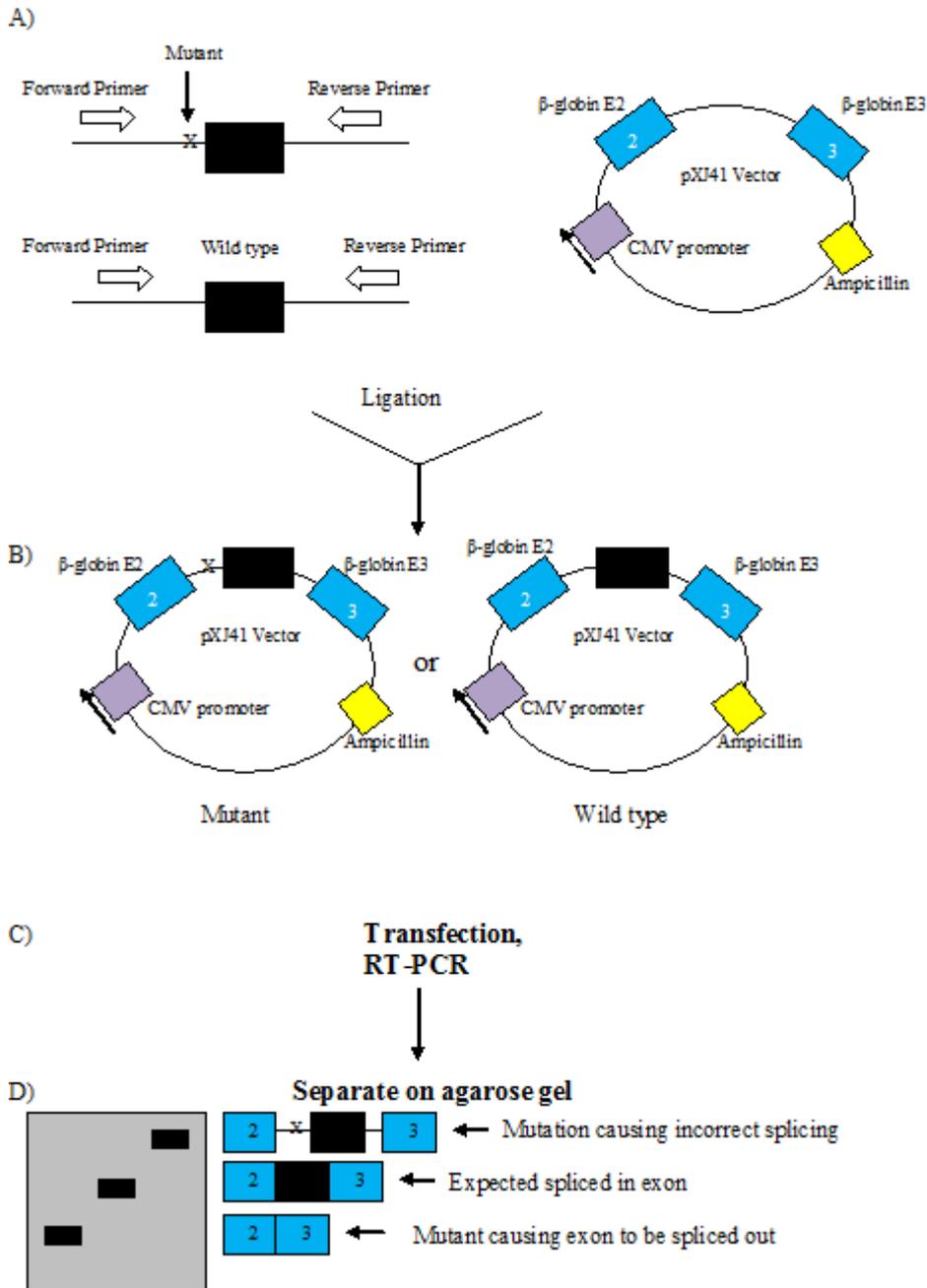
The fragment to be cloned into the vector is selected so that it comprises around 500 nucleotides upstream and downstream of the exon of interest. Oligonucleotides were designed so that after amplification a restriction site would be included at the flanking sites of the region of interest. Thus, the oligonucleotide consisted six nucleotides (GAGAGA) were included as the first nucleotides of the oligonucleotides, followed by the restriction sites and 18-20 nucleotides that were complementary to the fragment of interest. Six nucleotides were included because restriction enzymes do not cleave efficiently at the end of the molecule. PCRs were performed using high fidelity polymerase in order to obtain an error-free amplicon.

Both the wild type and the variant genomic regions (from control and patient's DNAs respectively) were amplified and cloned into the vector, in order to generate two "alleles" of the minigene construct. The effect of the relevant variants can be determined by comparing the mRNAs produced by a human cell line transiently transfected with the two alleles of the minigene construct. Figure 2.7 shows the method schematically and Figure 2.8 an example from the literature of this system in use.



**Figure 2.6 Map of pXJ41 vector**

Map of pXJ41 vector with a human cytomegalovirus (hCMV) promoter, rabbit  $\beta$ -globin intron II, T7 promoter, a multiple cloning site (MCS), SV40 polyadenylation site and BSM13+ sequence.



**Figure 2.7 Overview of in vivo splicing analysis with the minigene approach (Adapted from Stoss *et al.*, 1999) (Stoss *et al.* 1999)**

A) Using high-fidelity Taq polymerase, the alternatively spliced exon (black box) and its flanking intronic regions ( $\pm 500$ bps) were amplified from control and patient's genomic DNA and cloned into pXJ41. B) Resulting wild type and mutant minigene constructs, where the exon of interest was flanked by fragments  $\beta$ -globin exon 2 and exon 3. C) 24 hours after transfection, mRNA is analyzed by RT-PCR using primers of the flanking exons. D) Possible PCR products are shown, depending on whether the relevant exon was spliced out correctly.



**Figure 2.8** Example of the splicing pattern of TLE4 pre-mRNA (Extracted from Liu et al., 2009) (Liu et al. 2009)

RT-PCR analysis of the splicing pattern of the TLE4-T minigene from cells co-transfected with different splicing factors along with a constant amount of Tra2 $\beta$ . Tra2 $\beta$  activated the splicing to TLE4-T and also induced splicing inclusion of a further TLE4-B splicing isoform annotated as TLE4-B. The upper bands on the agarose gel correspond to the product including TLE4-B, the middle bands include TLE4-T and the lower bands correspond to the flanking exons.

## 2.4.2 Media and reagents

LB medium was prepared by adding 5g of tryptone, 2.5g of yeast extract and 5g of NaCl to 500ml of ddH<sub>2</sub>O. Medium was then autoclaved.

LB agar was prepared by adding 15g of agar to 1litre of LB media. The media was autoclaved and allowed to cool to around 55°C. 100mg/ml of ampicillin was then added to the cooled media, mixed well, and poured into petri-dishes. LB agar plates were stored at 4°C for later use.

NZY+ broth was prepared as described in the QuickChange® Site-Directed Mutagenesis Kit instruction manual. Briefly, NZY+ broth consisted of 10g of NZ amine, 5g of yeast extract and 5g of NaCl to 1litre of ddH<sub>2</sub>O. pH was adjusted to pH7.5 using 5M NaOH. Medium was then autoclaved. Prior to use, filter-sterilized supplements of 12.5ml of 1M MgCl<sub>2</sub>, 12.5ml of 1M MgSO<sub>4</sub>, and 20ml of 20% glucose were added to the medium.

## 2.4.3 Cloning

### 2.4.3.1 Primer designs and PCR

Primers were designed as in section 2.2.1 to amplify the exon of interest plus 500bp flanking intronic sequences. In view of the requirement for downstream applications, a restriction map for the fragment to be amplified was obtained. If the region to be amplified did not contain an *EcoRI* site, an *EcoRI* recognition sequence was added to the 5' end of both primers (5' – GAGAGAGAATTC – 3'). However, if the region contained an *EcoRI* site, an *MfeI* recognition sequence was added instead to the 5' end of both primers (5' – GAGAGACAATTG – 3').

PCR was carried out using Phusion High-Fidelity DNA Polymerase (Finnzymes). PCR reactions were set up on ice. The master mix comprised 1x of 5x Phusion GC Buffer (for regions with high GC content); 200µM of each dNTP, 0.5µM of each primer (forward and reverse), 0.02U/µl of Phusion DNA Polymerase, 30ng of DNA (proband with the variant identified) and ddH<sub>2</sub>O to a final volume of 50µl.

The PCR cycling programme started with initial denaturation at 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, annealing temperature (69.6°C) for 30 seconds and 72°C for 30 seconds (depending on the amplicon's length). A final extension of 72°C for 5 minutes concluded the programme. Primer sequences and primer pairs are listed in Table 2.5.

The PCR amplicon was then purified using QIAquick PCR Purification Kit (QIAGEN), following manufacturer's protocol. Concentration of the purified PCR amplicon was determined by NanoDrop spectrophotometer.

<b>Cloning Primers</b>	<b>Sequence (5'→ 3')</b>
MG-BMP4DF ( <i>EcoRI</i> )	GAGAGAGAATTCGGCATCCACGGTCGGGAGGG
MG-BMP4AR ( <i>EcoRI</i> )	GAGAGAGAATTCGCTCCAATGACAGCCAGTC
MG-BMP4EF ( <i>MfeI</i> )	GAGAGACAATTGAGCACGACGCACCGCGCCGA
MG-BMP4BR ( <i>MfeI</i> )	GAGAGACAATTGGCTAACCGGGTGGCTTGAGG
MG-BMP4ER ( <i>MfeI</i> )	GAGAGACAATTGGAAGCCGAGAAGGTGGTCTGG
MG-BMP4FF ( <i>MfeI</i> )	GAGAGACAATTGGTTCGAGCTGGGAGACGCAG
MG-BMP4FR ( <i>MfeI</i> )	GAGAGACAATTGTCCGGAGAGACCTCTGTGAG
MG-BMP4GF ( <i>MfeI</i> )	GAGAGACAATTGATCCTCTGGCTTCAGGTGGC
MG-BMP4GR ( <i>MfeI</i> )	GAGAGACAATTGCAGTTGGTTCAGAGCCAGAGG

<b>Novel Variants</b>	<b>Cloning Primer Pairs</b>
BMP4 c.-295G>A	MG-BMP4DF/4AR
BMP4 c.-136C>G	MG-BMP4DF/4AR
BMP4 c.-87C>T	MG-BMP4EF/4BR
BMP4 c.-87C>T	MG-BMP4EF/4ER
BMP4 c.-87C>T	MG-BMP4FF/4FR
BMP4 c.-87C>T	MG-BMP4GF/4GR
BMP4 c.-87C>T	MG-BMP4FF/4GR
BMP4 c.-87C>T	MG-BMP4GF/4FR

**Table 2.5 Cloning primers sequences and the combination of primer pairs**

Sequences of the primers used for cloning and the combination of primer pairs for each novel variants: BMP4 c.-295G>A, BMP4 c.-136C>G and BMP4 c.-87C>T.

#### **2.4.3.2 Restriction enzyme digestions**

The purified PCR amplicon was digested with *Mfe*I or *Eco*RI (NEB) as described in the manufacturer's protocol. The digestion reaction consisted of 5µl of NEBuffer 4, 1.5µl of *Mfe*I or *Eco*RI, (0.5µl of BSA was added for *Eco*RI reactions), 1µg of purified PCR product, and ddH<sub>2</sub>O to a final volume of 50µl. In parallel, 1µg of vector pXJ41 was digested with *Mfe*I. The vector pXJ41 was obtained from Prof. David Elliott, Institute of Human Genetics. The reaction was incubated at 37°C for 3 hours and deactivated at 65°C.

#### **2.4.3.3 Dephosphorylation of vector**

The linearized minigene vector pXJ41 was treated with phosphatase. 1x Antarctic Phosphatase Reaction Buffer (NEB) and 1µl of Antarctic Phosphatase (NEB) were added to 1µg of linearized vector. The reaction was mixed and incubated for 40 minutes at 37°C. The reaction was then heat inactivated for 20 minutes at 80°C.

#### **2.4.3.4 Ligation**

Before ligation, insert DNA and vector DNA were purified using the QIAquick PCR Purification Kit (QIAGEN), following the manufacturer's protocol. The vector and insert DNA were ligated using T4 DNA Ligase (Promega). The molar ratio of vector to insert DNA was 1 to 3. If the ligation was not successful, different molar ratios were used. The reaction consisted of 1x Ligase Buffer, 20units of T4 DNA Ligase, 100ng of vector DNA, 100ng of insert DNA and ddH<sub>2</sub>O to a final volume of 20µl. The reaction was incubated at 4°C overnight and was deactivated at 65°C for 20 minutes.

#### **2.4.3.5 Transformation of competent cells**

Ligation mixtures were transformed into JM109 competent cells (Promega), following the product's technical bulletin. Competent cells were thawed on ice and then 50µl of cells were transferred to a chilled eppendorf tube which contained 5µl ligation product

(approximately 100ng). The tube was flicked gently to mix and incubated on ice for 30 minutes. The cells were then heat-shocked for 90 seconds at 42°C and incubated on ice for 15 minutes. 500µl of LB medium was added to each transformation reaction and incubated for 60 minutes at 37°C with constant shaking at 200rpm. 50µl of cell culture was spread onto LB-Amp agar plate. The tube was spun down at 1864g for 1 min, most of the LB medium discarded, the pellet resuspended in the remaining LB medium and spread onto another LB-Amp agar plate. The plates were then incubated upside down at 37°C for 12-14 hours.

#### **2.4.3.6 Screening PCR**

Ampicillin-resistant colonies were individually picked and screened for insertion of the amplicon using colony PCR. A library of the colonies picked was kept. 1µl of colony DNA was used to perform PCR. PCR reactions were performed using HotStarTaq™ DNA Polymerase (Qiagen) as section 2.2.2, with vector primers pXJ\_B1 5' – GTAACCATATAAGCTGCAA – 3' and pXJ\_RTf 5' – GCTCCGATCGATCCTGAGAACT – 3', the annealing temperature was 58°C for 30 seconds and the elongation temperature was 72°C for 1 minute and 30 seconds. Amplified PCR products were then run on 1% gel alongside 1kb DNA Ladder (Fermentas) as in section 2.2.3. Positive colonies were identified by the size of the amplification products, which were around 1.5kb (when the exon was inserted). The size of the empty plasmid was around 200bp. Positive colonies underwent another PCR amplification using a forward primer complementary to the region of interest and a reverse vector primer (pXJ\_RTf) or a forward vector primer (pXJ\_B1) and a reverse primer on the region of interest. Only positive colonies containing the fragments inserted in the correct orientation will produce amplified products.

#### **2.4.3.7 Sequencing of insertions**

Positive colonies with the insert cloned into pXJ41 vector in the correct orientation were cultured in LB medium containing 50mg/ml Ampicillin at 37°C for 16 hours. DNA was then extracted using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's protocol. The plasmid was sent to MWG for sequencing with pXJ\_B1

and pXJ\_RTTF primers to confirm the inclusion and the correct direction of the region of interest.

#### **2.4.4 Transfection of HEK293 cells**

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS). Cells were grown at 37°C in a 5% CO<sub>2</sub> humidified atmosphere and were passaged every 3 days. HEK293 cells were seeded in 6-well plates (Nunc Corporation) for 24 hours before transfection. Cells were approximately 70% confluent when transfected. For each well, 500ng of plasmid DNA was added to 97µl of DMEM medium without serum and was mixed with 3µl of GeneJammer (Stratagene). The transfection reaction was incubated for 15 minutes at room temperature. The reaction was then added to the cells maintained in 2ml of DMEM with FBS drop by drop. Cells were incubated at 37°C in 5% CO<sub>2</sub> for another 24 hours after transfection.

#### **2.4.5 RNA Extration**

24 hours after tranfection, cells were trypsinised and centrifuged at 3300g for 2 minutes. The media was removed and cell pellets were resuspended in 100µl of Trizol (Invitrogen) in a hood and vortexed. 20µl chloroform was added and mixed vigorously for 15 seconds. The tube was then left at room temperature for 2 minutes and centrifuged at 15700g for 15 minutes at 4°C. The aqueous layer was transferred to a new eppendorf tube. RNA was precipitated by adding 50µl of 100% isopropanol and vortexed for a few seconds. The mixture was then left at room temperature for 15 minutes and centrifuged at 15700g for 10 minutes at 4°C. The supernatant was discarded and the pellet was washed with 100µl of 70% ethanol. The mixture was centrifuged at 9300g for 8 minutes at 4°C. Ethanol was then discarded and pellets were air-dried for 10 minutes. Pellets were resuspended in 60µl of RNase free water and the concentration of RNA was determined using NanoDrop spectrophometer.

#### **2.4.6 Reverse Transcription PCR**

RT-PCR was performed using OneStep RT-PCR Kit (Qiagen) following manufacturer's protocol. 5µl RT-PCR reaction consisted of 1x of OneStep RT-PCR buffer, 1x Q solution, 10mM of dNTPs, 0.2µl of HotStarTaq DNA Polymerase, 10mM of pXJ41\_B1 primer, 10mM of pXJ\_RTTF primer (the sequences of the primers are described in section 2.4.2.6) and 50ng of RNA.

The thermal cycler programme was 50°C for 30minutes, and then 95°C for 15minutes, followed by 25cycles of 94°C for 30seconds, 56°C for 30seconds and 72°C for 1minute, with a final extension at 72°C for 10minutes.

The PCR products were resolved on 1% agarose gel with gel red by electrophoresis to visualize bands under UV.

## 2.5 Luciferase assay of SMAD6 variants

### 2.5.1 Experimental design

The effects of the novel coding variants discovered in SMAD6 on BMP signaling were investigated by dual luciferase assay. The constitutively active form of the BMP type I receptor used in these experiments, c.a.BMPR1A, was described by Fujii and colleagues (Fujii et al. 1999). The BMPR1A cDNA was mutagenised so that an amino acid substitution was created at position 233 of the translated protein when subcloned into pcDNA3. The glutamine to aspartic acid change results in the activation of BMPR1A in the absence of ligands or the type I receptor. The c.a.BMPR1A was sequenced by me before use to confirm the presence of the mutation in the construct. The c.a.BMPR1A construct also contained a HA tag [YPYDVPDYA] at the C-terminal.

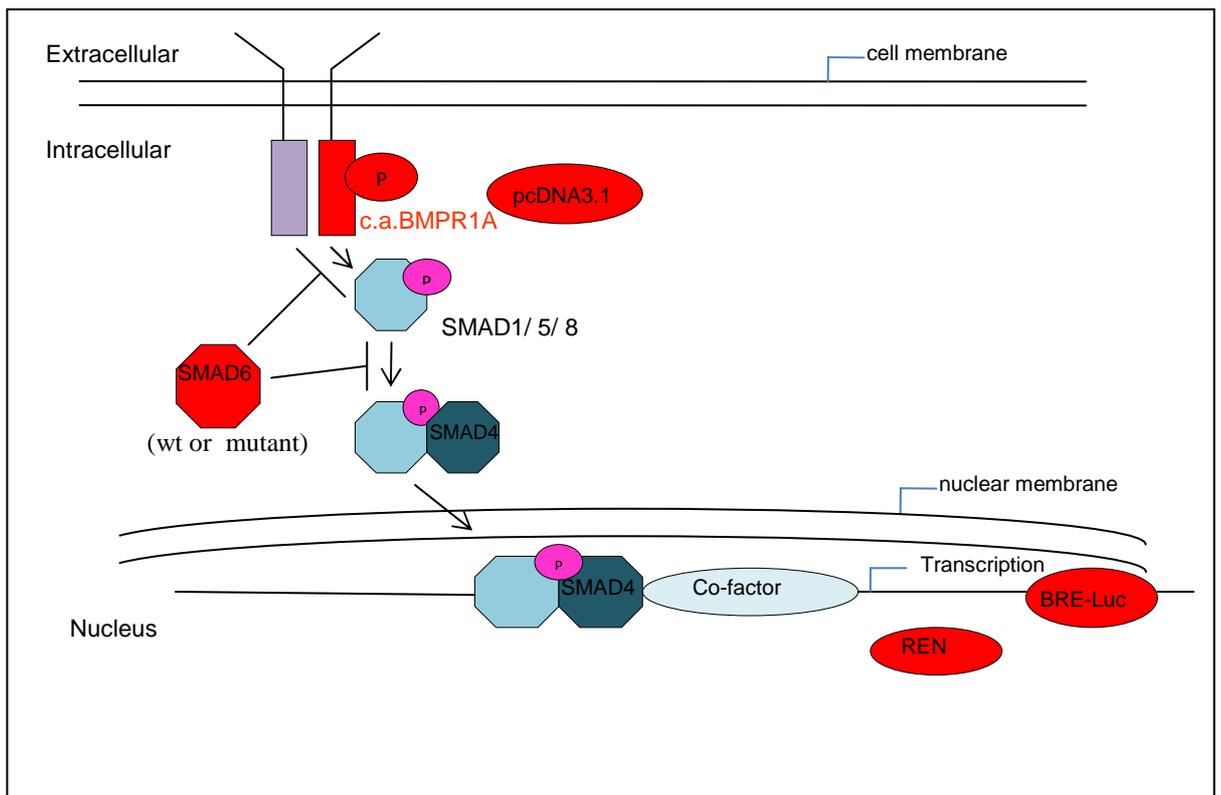
BRE-Luc was the luciferase reporter construct used to measure the transcriptional activation induced by BMP signaling (Korchynskiy and ten Dijke 2001). BRE-Luc consists of two Smad binding elements (SBEs) and a GGCGCC palindromic sequence flanked by two CAGC and two CGCC sites cloned into pGL3-basic vector. The pGL3 luciferase reporter vector contains a modified coding region for firefly (*Photinus pyralis*) luciferase. This vector lacked promoter and enhancer sequences, meaning that the luciferase reporter gene activity obtained would be solely driven by the inserted BMP-responsive elements. The Renilla luciferase cDNA, encoding the renilla luciferase from the marine organism sea pansy (*Renilla reniformis*), was contained within the pRL-TK vector. The luciferase activity was normalised against Renilla luciferase activity, which acted as an internal control for transfection efficiency.

Wild-type SMAD6C inhibits BMP signaling through at least three possible mechanisms (see below). Goto and colleagues have previously utilized this system to investigate the effects of inhibitory Smads (Smad6 and Smad7) on BRE-Luc induced by c.a.BMPR1A or c.a.ALK2 in C2C12 cells (Goto et al. 2007).

The genetic variants p.Ala325Thr (c.973G>A), p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T), were introduced into SMAD6 expression constructs using site-directed mutagenesis. The wild-type SMAD6 construct (SMAD6C) used which consists of the MH2 domain of SMAD6 inserted into a pcDNA3 vector, has been previously reported (Hanyu et al. 2001). The construct was FLAG tagged

[DYKDDDDK] at the N-terminal end of SMAD6. The presence of mutations was confirmed by sequencing. Each of the constructs was then transfected into the cell system along with BMP-responsive element luciferase reporter (BRE-Luc), *Renilla* luciferase reporter (REN) and constitutively active BMPR1A (c.a.BMPR1A). The experimental system is summarized in Figure 2.9.

There were six combinations of constructs used in this assay (Table 2.6). All of these combinations contained BRE-Luc and Renilla constructs. The first combination contained empty vector pcDNA3.1, which acted as a negative control. The second combination contained c.a.BMPR1A and pcDNA3.1, which acted as a positive control. The subsequent combination of constructs contained c.a.BMPR1A and either wtSMAD6C, SMAD6C p.Ala325Thr (c.973G>A), SMAD6C p.Pro415Leu (c.1244C>T) or SMAD6C p.Cys484Phe (c.1451G>T). This combination was selected to enable the identification of any difference in the ability of the wild-type and mutant constructs to inhibit the activation of BRE-Luc activity.



**Figure 2.9 BMP signaling in C2C12 cells for luciferase assay**

BMP signaling in the system of C2C12 cells that were transfected with different combination of constructs for luciferase assay. The constructs that were transfected are in red colour.

Combinations	1	2	3	4	5	6
Constructs						
BRE-Luc	✓	✓	✓	✓	✓	✓
Renilla	✓	✓	✓	✓	✓	✓
pcDNA3.1	✓	✓				
c.a.BMPRI1A		✓	✓	✓	✓	✓
SMAD6C wt			✓			
SMAD6C p.Ala325Thr (c.973G>A)				✓		
SMAD6C p.Pro415Leu (c.1244C>T)					✓	
SMAD6C p.Cys484Phe (c.1451G>T)						✓

**Table 2.6      Combination of constructs for luciferase assay**

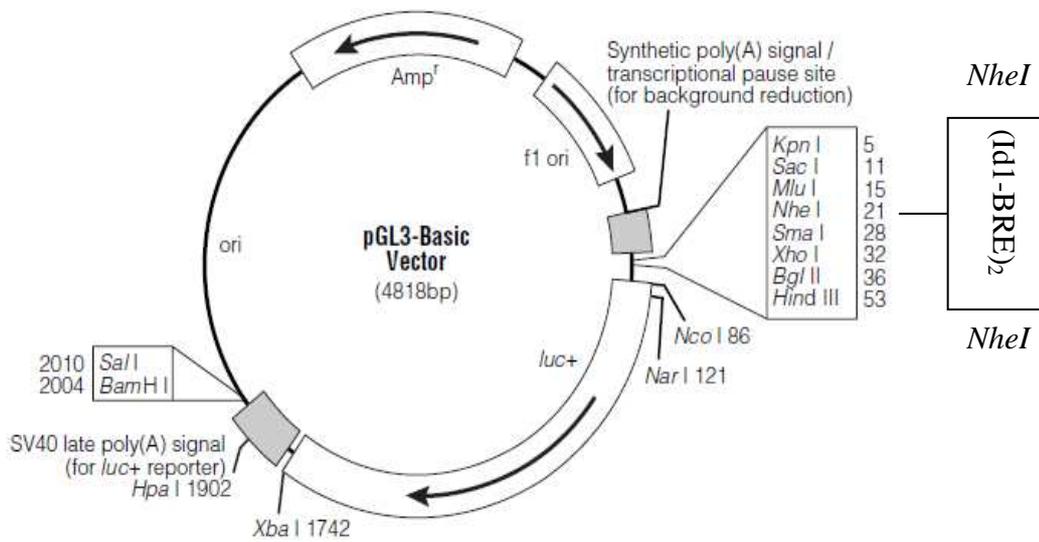
The combinations of constructs that were used in the luciferase assay.

### 2.5.2 Cell lines and tissue culture

Mouse myoblast C2C12 cells were maintained in Dulbecco's modified Eagle's medium (Invitrogen) in 10% fetal bovine serum (FBS) with 50 units/ml penicillin and 50µl/ml streptomycin (Invitrogen). Cells were grown at 37°C in a 5% CO<sub>2</sub> humidified atmosphere and were passaged every 2 days; split at a ratio of 1:10. The cells were detached from the flasks using 0.05% Trypsin (1X) with EDTA (Invitrogen).

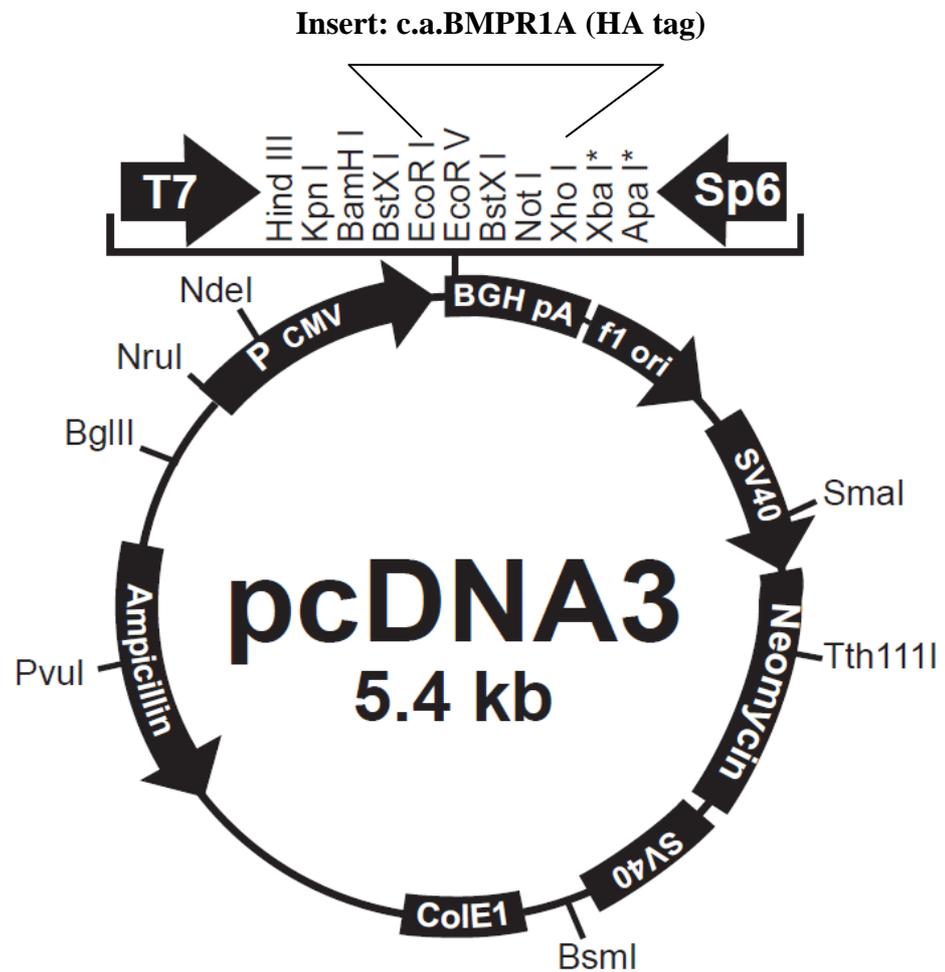
### 2.5.3 Constructs

The four constructs were as follow: (1) a luciferase reporter construct containing BMP-responsive elements (BRE-luc), which was previously described by Korchynskyi and ten Dijke, 2002 (Figure 2.10) (Korchynskyi and ten Dijke 2001); (2) pRL-TK *Renilla* luciferase reporter; (3) constitutively active BMPR1A, which was described by Miyazono 1999 (Figure 2.11) (Miyazono 1999); and (4) SMAD6C wild type, described by Hanyu et. al., 2001 (Figure 2.12) (Hanyu et al. 2001). All these were kindly donated by Prof. Peter ten Dijke, Leiden University Medical Center, Leiden, The Netherlands. Empty vector pcDNA3.1 (Invitrogen) was used as control. Constitutively active BMPR1A and SMAD6C wild type were sent to GATC to confirm their sequences.



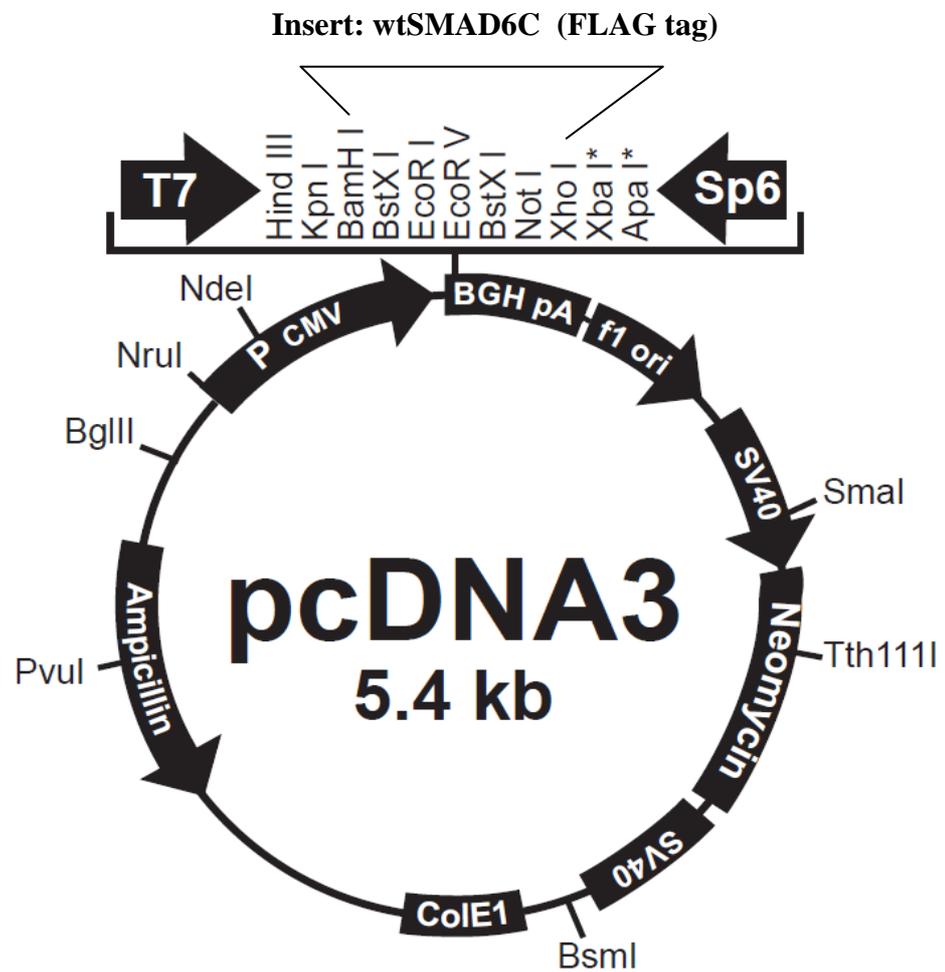
**Figure 2.10 BRE-Luc inserted at *NheI* site in pGL3-basic vector circle map**

Map of the BMP-responsive elements luciferase (BRE-Luc) vector inserted at *NheI* site in pGL3-basic vector, which lacks promoter and enhancer sequences.



**Figure 2.11** c.a.BMPRI1A with HA tag inserted at *EcoRI* and *XhoI* sites in **pcDNA3**

Map of the c.a.BMPRI1A vector with HA tag in pcDNA3. The BMPRI1A cDNA was mutagenised so that an amino acid substitution was created at position 233 in the serine/threonine kinase domain.



**Figure 2.12** SMAD6C wild type with FLAG tag inserted at *BamHI* and *XhoI* sites in pcDNA3

The wild type SMAD6 construct (SMAD6C) consists of the MH2 domain of SMAD6 inserted into a pcDNA3 vector.

#### 2.5.4 Site-directed mutagenesis

Site directed mutagenesis was performed using the QuickChange® Site-Directed Mutagenesis Kit (Stratagene). Mutagenic primers were designed using Stratagene's primer design programme at <http://www.stratagene.com/qcprimerdesign>. HPLC primers were ordered from Metabion. Primer sequences and conditions are listed in Table 2.7.

SMAD6 sample reactions comprised of 1x reaction buffer, 50ng of construct DNA, 125ng of each mutagenic primer, dNTP mix, 2.5U of *Pfu* DNA polymerase and ddH<sub>2</sub>O to a final volume of 50µl. 50µl control reaction comprised of 1x reaction buffer, 10ng of pWhitescript 4.5kb control plasmid, 125ng of each control primer, dNTP mix, 2.5U of *Pfu* DNA polymerase and ddH<sub>2</sub>O.

The thermal cycling programme started at 95°C for 30 seconds, followed by 12 cycles of 95°C or 97°C for 30 seconds (depending on the primer sets), 55°C for 1 minute and 68°C for 6 minutes or 10 minutes. 5µl of amplified product were run on 1% agarose gel to check for product amplification. 1µl of *Dpn* I restriction enzyme (10U/µl) was added to each reaction. Reactions were incubated at 37°C for 1 hour to digest the parental dsDNA. Resulting constructs were transformed into XL1-Blue supercompetent cells (Stratagene) as described in the QuickChange® Site-Directed Mutagenesis Kit instruction manual. 500µl of NZY<sup>+</sup> broth preheated to 42°C was added to the transformation reactions and was incubated at 37°C for 1 hour with shaking at 225rpm. 250µl of cell culture was spread onto LB-Amp agar plate. The plates were then incubated upside down at 37°C for 16 hours.

Primer pairs	Primer (5'→ 3' Sequence)		Denaturing Temp. (°C)	Elongation Time (min)
	Forward	Reverse		
SMAD6_C484F_G307T	ggcagttcatcacctcctccctgctgg	ccagcaggggaaggaggtgatgaactgcc	95	6
SMAD6_A325T_G475A	catgtctccggacaccaccaagccgag	ctcggcttgggtgtccggagacatg	97	10
SMAD6_P415L_C746T	ttcgtcaactccctgacgctggacgcg	cgcgtccagcgtcagggagttgacgaa	97	10

**Table 2.7 Primers and reaction conditions used for the site-directed mutagenesis.**

Sequences of the primers and the PCR conditions used for the site-directed mutagenesis to generate variants SMAD6 p.Ala325Thr (c.973G>A), SMAD6 p.Pro415Leu (c.1244C>T) and SMAD6 p.Cys484Phe (c.1451G>T).

### **2.5.5 Sequencing of constructs**

XL1-Blue cell positive colonies were picked and were incubated with NZY+ broth containing 50mg/ml Ampicillin at 37°C for 16 hours. Construct DNA was then extracted using the QIAprep Spin Miniprep Kit (Qiagen) following the protocol. The constructs were nanodropped to determine the concentration before sending to GATC for sequencing with T7 and Sp6 primers in order to identify the mutations.

### **2.5.6 Transfection of C2C12 cells**

C2C12 cells were seeded in 12-well plates (Nunc Corporation) for 24 hours before being transfected with plasmid DNA using FuGENE® HD Transfection Reagent (Roche) according to the manufacturer's protocol. The combination of constructs consisted of 450ng of BRE-luc (Firefly luciferase reporter with BMP Responsive Elements), 50ng of pRL-TK (Renilla luciferase reporter) and 500ng of pcDNA3.1 (empty plasmid as control) or 250ng of caBMPR1A (constitutive active BMPR1A) with either 250ng of pcDNA3.1, 250ng of SMAD6 wild type or 250ng of SMAD6 mutant constructs. Each combination of constructs was carried out in triplicate. Cells were incubated at 37°C in 5% CO<sub>2</sub> for another 24 hours after transfection.

### **2.5.7 Measurement of luciferase activity**

Cells were lysed and luciferase activity measurement was performed using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol. A Fluoroskan Ascent plate reader (Thermo Scientific) was utilised to read the luminescence activities of the cells. Firefly luminescence signal was measured first, followed by *Renilla* luciferase. The firefly luminescence readings were then normalized to the *Renilla* luciferase, to account for experimental variability caused by differences in cell viability or transfection efficiency.

### 2.5.8 Immunoblotting

1x NuPAGE® Sample Reducing Agent (Invitrogen) and 1x NuPAGE® LDS Sample Buffer (Invitrogen) were added to 15µl of previously lysed cells, which were then heated at 95°C for 10 minutes and allowed to cool. 4–12% NuPAGE® Novex Bis-Tris Gel (Invitrogen) was assembled in 1x MES SDS Running Buffer (Invitrogen) following NuPAGE® Technical Guide (Invitrogen). Samples were loaded into the wells alongside 10µl of PageRuler™ Prestained Protein Ladder (Fermentas) and were run at 200V until the yellow loading buffer tide line reached the bottom of the gel.

Two pieces of filter paper (8cm x 10cm), a piece of Hybond-C Extra nitrocellulose membrane (7cm x 9cm) (GE Healthcare) and 2 pieces of sponge were soaked in pre-cooled transfer buffer (2.93g Glycine, 5.81g Tris Base, 3.75ml 10% SDS, 200ml Methanol, made up to 1litre with ddH<sub>2</sub>O). The transfer apparatus was set up on the black part of the cassette in the following order: 1 piece of sponge, 1 filter paper, gel, membrane, 1 filter paper, 1 piece of sponge; to form the gel sandwich. The cassette was then placed into a transfer tank filled with chilled transfer buffer and an ice block. Samples were transferred onto the membrane at 350mA for 1 hour.

The membrane was then stained with Ponceau-S solution (Sigma) to check for protein transfer. The membrane was cut into two at around 40kDa and was rinsed with ddH<sub>2</sub>O until the background was clean. It was then incubated in TBST (8.75g NaCl, 6.5g Tris-HCl, 2ml Tween 20, made up to 1litre with ddH<sub>2</sub>O) with 5% marvel (non-fat dried milk) to block non-specific binding sites for 1 hour at room temperature or overnight in the cold room on a shaker. The membrane containing proteins of less than 40kDa was incubated with monoclonal anti-FLAG M2-Peroxidase (HRP) antibody produced in mouse (Sigma) at 1/10000 in blocking reagent for 2 hours at room temperature. The other part of the membrane was incubated with anti-HA-Peroxidase, High Affinity (3F10) (Roche) at 1/1000 in blocking reagent for 2 hours at room temperature. Three ten minutes washes in TBST were carried out and then both membranes were incubated with Supersignal® West Dura Extended Duration Substrate (Thermo Scientific) according to the manufacturer's instructions for 5 minutes. The membranes were placed against Biomax MR Film (Kodak) for 5 seconds, which was then developed in a compact X4 developer (Xograph Imaging Systems).

The membrane which contained proteins larger than 40kDa was washed 3 times with TBST, for ten minutes each. It was then incubated with stripping buffer (6.7ml of 0.5M Tris-HCl pH6.7, 10ml 20% SDS, 780 $\mu$ l  $\beta$ -mercaptoethanol, made up to 100ml with ddH<sub>2</sub>O) at 50°C for 30–40 minutes with gentle agitation. The membrane was rinsed with ddH<sub>2</sub>O three times followed by extensive washing in TBST for 15 minutes. The membrane was then blocked in TBST with 5% marvel for 1 hour at room temperature or overnight at 4°C on shaker. Monoclonal anti- $\alpha$ -tubulin primary antibody (Sigma) diluted in blocking reagent 1/1000 was added to the membrane and incubated for 2 hours at room temperature or overnight at 4°C. Three ten-minute washes in TBST were carried out. The membrane was then incubated for 1 hour with goat anti-mouse HRP secondary antibody (Thermo Scientific) diluted 1/1000 in blocking reagent. Following another three ten minutes washes in TBST, the relevant protein bands were detected by enhanced chemiluminescent (ECL) detection process as described above.

This membrane (which contained proteins larger than 40kDa) was then washed, stripped and checked with ECL to ensure that the previous antibodies had been removed. The blot was then blocked as described above. Smad1 antibody (Santa Cruz) diluted in blocking reagent 1/50 was added to the membrane and incubated overnight at 4°C. Three ten-minute washes in TBST were carried out. The membrane was then incubated for 1 hour with goat anti-mouse HRP secondary antibody (Thermo Scientific) diluted 1/300 in blocking reagent. After another three ten minutes washes in TBST, the protein bands were detected by the ECL detection process as above.

The same membrane underwent the same process (washed, stripped, checked with ECL and blocked), before incubating overnight with phospho-Smad1/ Smad5/ Smad8 antibody (Cell Signaling) at 1/1000 in blocking agent. Three ten-minute washes in TBST were carried out. The membrane was then incubated for 1 hour with goat anti-rabbit HRP secondary antibody (Jackson ImmunoResearch) diluted in 1/10,000 in blocking reagent. Following another three ten minutes washes in TBST, the protein bands were detected by ECL detection process.

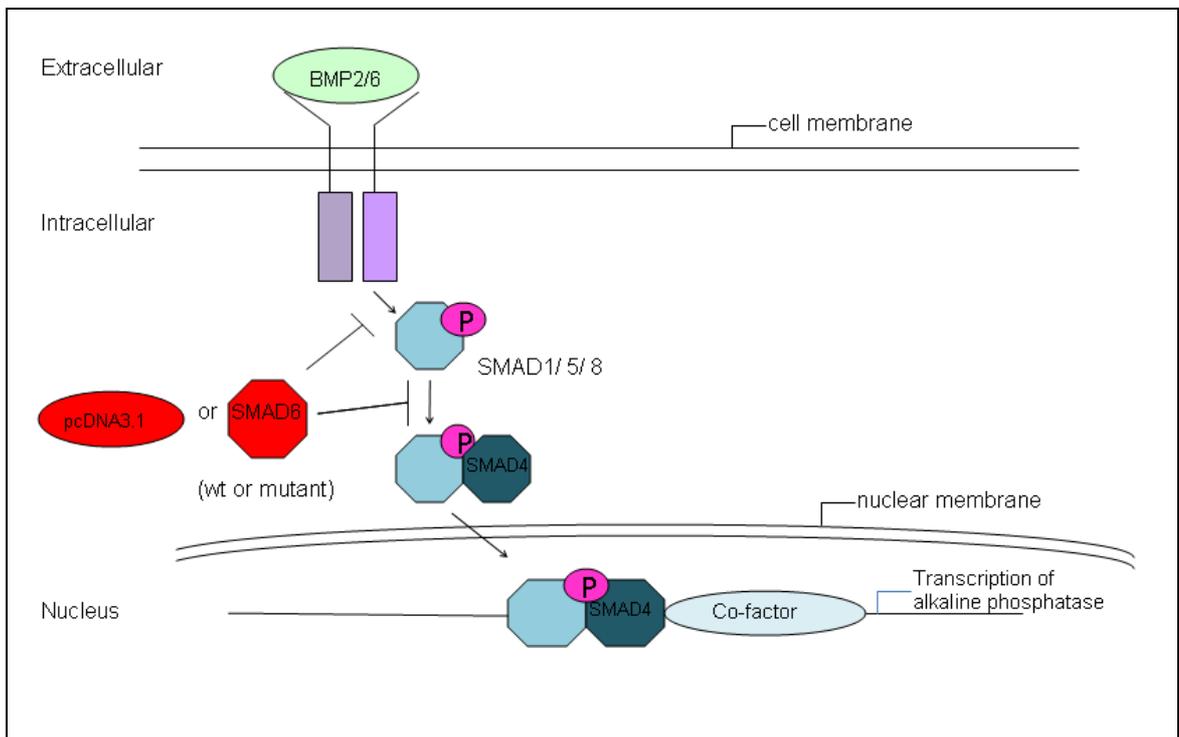
## 2.6 Alkaline phosphatase assay of SMAD6 variants

### 2.6.1 Experimental design

It has been reported by Katagiri and colleagues (Katagiri et al. 1994) that C2C12 cells will inhibit their myoblastic differentiation and convert to the osteoblastic differentiation pathway instead when treated with BMP2. The phenotype of one patient with a SMAD6 mutation included calcification of the aorta as a prominent feature. Therefore investigation on whether the SMAD6 mutations that were discovered failed to inhibit osteoblastic differentiation *in vitro* was performed. Extracellular bone matrix proteins and alkaline phosphatase (ALP) are expressed early during the commitment to the osteoblast lineage. To measure the inhibitory effect of wild-type and mutant SMAD6C constructs on osteoblastic differentiation, two experimental approaches were devised.

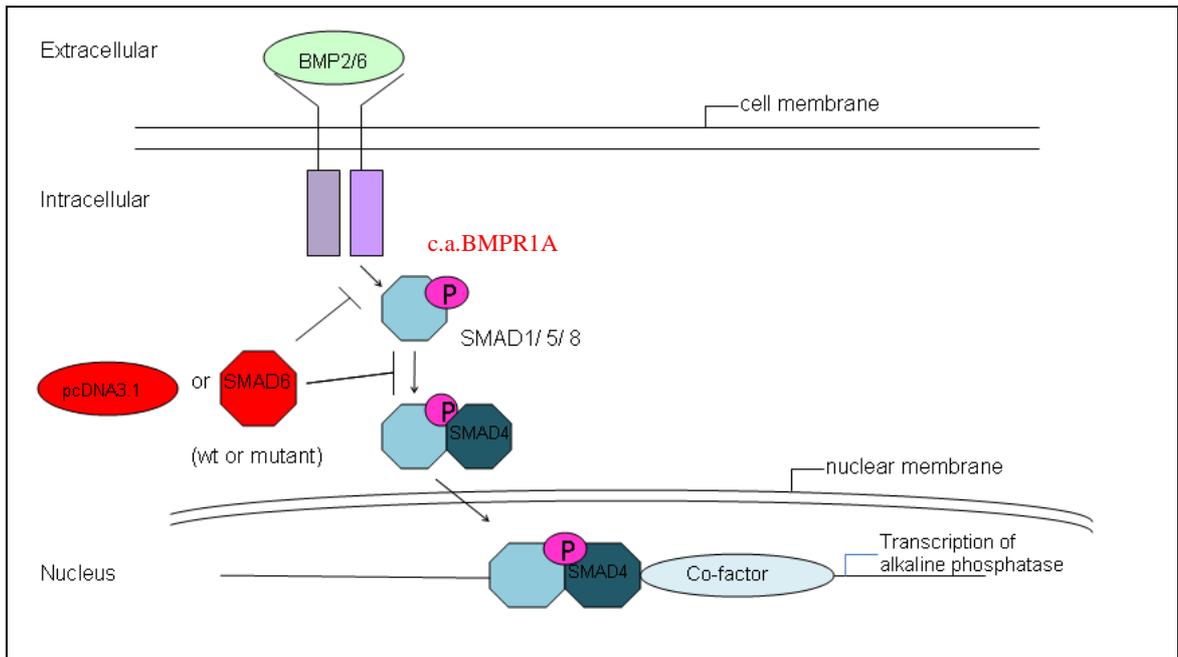
In the first experiment, cells were transiently transfected with either empty vector pcDNA3.1, SMAD6C wt, SMAD6C p.Pro415Leu (c.1244C>T) or SMAD6C p.Cys484Phe (c.1451G>T). The transfected cells were then treated with exogenous BMP2 or BMP6 ligands to stimulate alkaline phosphatase activity (Figure 2.13). Alkaline phosphatase is an osteogenic marker that will reflect the commitment of myoblast cells to osteoblastic differentiation. Both BMP2 and BMP6 were utilized to stimulate the cells as it has been reported that Smad6 inhibits BMP2 signaling more efficiently than BMP6 signaling but BMP6 induced more sustained signaling than BMP2 (Goto et al. 2007).

The second method of investigation of the inhibitory effect of wild type SMAD6 compared to the mutant SMAD6C constructs on ALP activity was carried out in collaboration with Dr. Elise Glen. Cells were transfected with c.a.BMPR1A and either SMAD6C wt, SMAD6C p.Pro415Leu (c.1244C>T) or SMAD6C p.Cys484Phe (c.1451G>T) as Fujii and colleagues have reported that c.a.BMPR1A induced ALP activity in C2C12 cells (Fujii et al. 1999) (Figure 2.14).



**Figure 2.13 BMP signaling in C2C12 cells stimulated with BMP ligands for alkaline phosphatase assay**

BMP signaling in the system of C2C12 cells that were transfected with either empty vector pcDNA3.1, SMAD6C wt, SMAD6C p.Pro415Leu (c.1244C>T) or SMAD6C p.Cys484Phe (c.1451G>T) and treated with BMP2 or BMP6 ligands to stimulate alkaline phosphatase transcription. The constructs that were transfected are in red colour and the stimulating molecules are in green.



**Figure 2.14 BMP signaling in C2C12 cells transfected with c.a.BMPR1A for alkaline phosphatase assay**

BMP signaling in the system of C2C12 cells that were transfected with c.a.BMPR1A and either empty vector pcDNA3.1, SMAD6C wt, SMAD6C p.Pro415Leu (c.1244C>T) or SMAD6C p.Cys484Phe (c.1451G>T) for alkaline phosphatase assay. The constructs that were transfected are in red colour.

### **2.6.2 Transfection and stimulation of C2C12 cells with BMP ligands**

The cell lines and constructs used for this experiment are described in sections 2.5.2 – 2.5.5. C2C12 cells were seeded in 12-well plates (Nunc Corporation) for 24 hours before transfecting with plasmid DNA using FuGENE® HD Transfection Reagent (Roche) according to the manufacturer's protocol. The cells were transiently transfected with either 1µg of pcDNA3.1, 1µg of SMAD6 wild type, 1µg of SMAD6 p.Cys484Phe (c.1451 G>T) or 1µg of SMAD6 p.Pro415Leu (c.1244C>T). Each transfection was carried out in duplicate. The growth medium was replaced with DMEM containing 2.5% FBS, 6 hours or 24 hours after transfection. Osteoblast differentiation of cells was induced by 300ng/ml recombinant human BMP-2 or BMP-6 in DMEM. The transfected and stimulated cells were then incubated at 37°C in 5% CO<sub>2</sub> for another 60 hours or 48 hours.

### **2.6.3 Transfection of C2C12 cells with c.a.BMPR1A**

C2C12 cells were seeded in 12-well plates (Nunc Corporation) for 24 hours before transfecting with plasmid DNA using FuGENE® HD Transfection Reagent (Roche) according to the manufacturer's protocol. The cells were transiently transfected with 0.5µg of c.a.BMPR1A and either 0.5µg of pcDNA3.1, 0.5µg of SMAD6 wild type, 0.5µg SMAD6 p.Cys484Phe (c.1451G>T) or 0.5µg of SMAD6 p.Pro415Leu (c.1244C>T). Each transfection was carried out in duplicate. The growth medium was replaced with DMEM containing 2.5% FBS, 24 hours after transfection. Osteoblast differentiation of cells was induced by the constitutively active BMPR1A. The transfected and stimulated cells were then incubated at 37°C in 5% CO<sub>2</sub> for another 24 hours.

### **2.6.4 Measurement of alkaline phosphatase activity**

After removing the culture medium, cells were washed with PBS, lysed with 250µl of lysis buffer (50mM Tris-HCl, 0.1% Triton-X, pH7.5) and sonicated to disrupt the cell membranes. ALP activity was determined in 20µl aliquots of each cell lysate in buffer containing 100mM glycine (Sigma), 1mM MgCl<sub>2</sub> (Sigma) and 0.1mM ZnCl<sub>2</sub> (Sigma),

pH 10.5, using *p*-nitrophenyl phosphate (Sigma) as a substrate. Each sample was measured in duplicate. The absorbance was measured at 405nm at different time point for 2 hours using Multiskan (Thermo Scientific). Immunoblotting was carried out on the cell lysates as described in section 2.5.8.

## 2.7 Statistical analysis

The mean reading and standard mean deviation (SEM) for each combination of constructs were used to identify the difference between the wild type and mutant constructs for both luciferase assay and alkaline phosphatase assay.

For the luciferase assay, the data set was assessed by one-way analysis of variance (ANOVA) in Minitab 15 to rule out any significant differences across the data set between the replicates, as the assay was performed three times independently. The pooled data was logarithmic transformed to conform to a Normal distribution. Using the transformed data, pairwise comparisons for the three replicates of each combination of plasmids were calculated.

For the alkaline phosphatase assay, the data set was assessed by two-way analysis of variance. The data from three independent experiments that were performed in duplicate was pooled together. Two way analysis of variance (ANOVA) was used to analyse the data between the wild type and mutant constructs, with each independent experiment as another factor. The group of cells that were transfected with constitutively active BMPR1A and empty vector pcDNA3.1, acted as a positive control. These were, however, excluded from the analysis examining differences between the wild-type and mutant SMAD6 constructs.

All p-values were two-sided and  $p < 0.05$  was accepted as the threshold for statistical significance.

## Chapter 3      Results of mutation screening and *in vitro* splicing studies

### 3.1 Overview

This chapter presents the results from the screening of candidate genes in the BMP signaling pathway for rare variants which could predispose to cardiovascular malformations (CVM). The results of the *in vitro* splicing assays are also presented in this chapter. The objectives of this part of the project were:

- To sequence the coding regions, including exon/intron boundaries of the candidate genes in the patient panel
- To determine whether the novel variants identified are transmitted or *de novo*
- To determine whether the novel variants identified are polymorphisms
- To analyse the identified novel variants using *in silico* approaches
- To analyse the effect of variants predicted to influence splicing using a minigene assay

Five genes that are involved in the BMP signaling pathway were chosen. These genes consisted of two ligands (*BMP2* and *BMP4*), two receptors (*BMPR1A* and *BMPR2*), and one signal transduction molecule (*SMAD6*). The exonic and the exon/intron boundaries of these genes were sequenced in a panel of 90 probands with various types of CVM. The dye-termination or Sanger sequencing method was utilized due to the state of the available technology at that time. The previously unreported genetic variants were then checked in their parents (whenever available) to determine inheritance and also were genotyped in a population of British Caucasian controls that are free of CVM.

The rare and novel variants were then analysed using *in silico* approaches to determine evolutionary conservation, the potential functional effect of the non-synonymous variants and the potential for splicing effects. Potentially significant splicing effects were assessed *in vitro* using the minigene splicing assay.

### 3.2 Phenotypes of the analyzed CVM patient cohort

A panel of 92 CVM patients were screened for del22q11 using Multiplex Ligation Probe-dependent Amplification (MLPA) (performed by Dr. Ana Topf). Two patients were identified with the deletion and were subsequently removed from the study. The remaining 90 patients were re-sequenced for the coding sequences, including exon/intron boundaries of *BMP2*, *BMP4*, *BMPR2*, *BMPRIA* and *SMAD6*. The cardiac phenotypes of the 90 patients were categorized into 16 main groups (Table 3.1). The phenotypes of a further 348 patients that were re-sequenced for the MH2 domain of *SMAD6* as a replication cohort are also listed in Table 3.1.

Cardiac phenotype	N=90	N=348	Total=438
Aortic stenosis	4	19	23
ASD	24	47	71
AVSD	8	11	19
Coarctation of aorta	5	18	23
DILV	1	1	2
Ebstein's anomaly	3	6	9
Hypoplastic left heart	2	4	6
PDA	3	15	18
PFO	4	6	10
Pulmonary atresia	0	4	4
Pulmonary stenosis	1	2	3
TGA	4	53	57
Tricuspid atresia	1	4	5
Truncus arteriosus	0	7	7
VSD	5	19	24
Complex/Multiple phenotypes	25	132	157

**Table 3.1 Phenotypes of the analysed CVM patients**

CVM patients were categorized into 16 main groups. The panel of 90 patients were re-sequenced for the coding sequences of *BMP2*, *BMP4*, *BMPR2*, *BMPRIA* and *SMAD6* whereas the panel of 348 patients were re-sequenced for the MH2 domain of *SMAD6* as a replication cohort.

(ASD: atrial septal defect, AVSD: atrioventricular septal defect, DILV: double inlet left ventricle, PDA: patent ductus arteriosus, PFO: patent foramen ovale, TGA: transposition of the great arteries, VSD: ventricular septal defect)

### 3.3 *BMP2*

The novel variants identified in *BMP2* in 90 patients are summarized in Table 3.2. The location of the variants is shown in Figure 3.1. These variants were genotyped in the first control population. Three variants were observed in the 5'UTR of exon one. The c.-906C>T variant was seen in a patient with membranous subaortic stenosis and was maternally inherited. This variant was not seen in the control population. The two other variants in the 5'UTR, c.-673G>A and c.-342\_-340delTGT, were both present in the controls with a minor allele frequency of 0.03 and 0.024 respectively. Two other novel genetic variants were found in the TGF-beta domain of the coding region in exon three. The TGF-beta domain is present in a variety of proteins: transforming growth factor beta, decapentaplegic proteins and bone morphogenetic proteins, all of which are members of the TGF-beta family. These variants were synonymous nucleotide substitutions, thus amino acid residues were not affected. The c.891G>A variant was paternally inherited, whereas the c.1146A>G variant was maternally inherited. Both variants were absent in the controls. In the 3'UTR region, a four-base pair deletion was identified in three probands and was present in the control population with a minor allele frequency of 0.19. For those variants present in cases and controls, the allele frequencies were compared using Fisher's exact test. There was a significant difference in allele frequency between cases and controls for the c.-342\_-340delTGT variant, which had a frequency of 0.044 in cases and 0.024 in controls (p=0.0057). There was no significant difference in allele frequency between cases and controls for the c.-673G>A and c.\*936-939delATTT variants.

The effect of the 5'UTR variants on splicing was predicted *in silico* using ESE Finder 3.0. The c.-906C>T variant resulted in the loss of the predicted branch site. However, since branch site sequences are very degenerative, and therefore loss of a single site is unlikely to affect splicing (see section 3.4 for more detailed discussion of this issue). There was no predicted effect on splicing for the c.-713G>A and c.-342\_-340del3 variants. Hence, no *in vitro* assessment of potential effects of these variants on splicing was carried out.

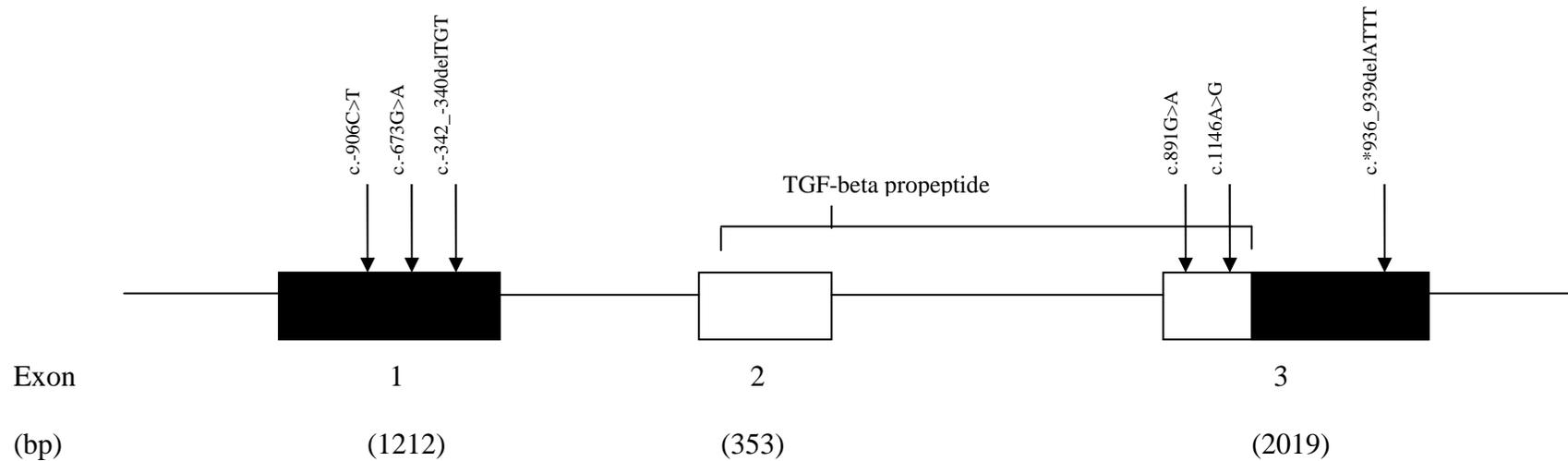
Using the MicroCosm Targets programme, the c.\*936\_939delATTT variant was not located at predicted sites for microRNAs.

Position	Nucleotide Change	Location	Amino Acid Substitution	Number of Patients	Type of CVMs	Inheritance	Controls (N=1425)	Fisher's exact test (p value)
Exon 1	c.-906C>T	5'UTR	-	1	Aortic stenosis	Mother	Absent	-
	c.-673G>A	5'UTR	-	4	1. Ebstein's anomaly 2. PFO 3-4. Complex	N/A	f(G)=0.97 f(A)=0.03	0.512
	c.-342_-340delTGT	5'UTR	-	8	1-4. ASD 5. TGA 6. AVSD 7-8. Complex	N/A	f(TGT)=0.976 f(delTGT)= 0.024	0.0057
Exon 3	c.891G>A	Exonic	p.Lys297Lys	1	TGA	Father	Absent	
	c.1146A>G	Exonic	p.Leu382Leu	1	ASD	Mother	Absent	
	c.*936_939delATTT	3'UTR		3	1. Tricuspid atresia 2. ASD 3. Complex	NA	f(ATTT)=0.981 f(delATTT)= 0.019	0.406

**Table 3.2** *BMP2* – Genetic variants

The genetic variants identified in *BMP2* in the panel of 90 probands.

(PFO, patent foramen ovale; ASD, atrial septal defect; TGA, transposition of the great arteries; AVSD, atrioventricular septal defect)



**Figure 3.1** *BMP2* – Location of genetic variants

Location of the genetic variation identified in *BMP2*, in the panel of 90 probands. Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pairs. 5'UTR and 3'UTR are in black colour. Location of the TGF-beta propeptide domain is labelled above the exons.

(TGF-beta: Transforming growth factor-beta)

## 3.4 *BMP4*

### 3.4.1 Sequencing results

Five exons (including the alternative first exon) of *BMP4* were sequenced and analysed for previously unreported genetic variants in 270 probands with cardiovascular malformation where 55 probands were diagnosed with Transposition of the Great Arteries (TGA) or TGA with other malformations. The numbers of patients sequenced were increased and the sequenced cohort for TGA patients was enriched, as during the period this work was conducted, Thienpont and colleagues (Thienpont et al. 2007) identified a deletion on 14q21.3, near *BMP4*, which is located at 14q22.2, in a patient with TGA, microtia, microphthalmia and polydactyly. A total of nine previously unreported genetic variants in *BMP4* were identified (Table 3.3) and the location of the variants is shown in Figure 3.2. The novel variants were genotyped in the first control population consisted of a collection of 1425 British Caucasian individuals.

A proband with complex transposition of great arteries with right sided atrioventricular valve atresia was interesting as three novel genetic variants were detected. The c.-283G>C variant was located in the 5'UTR of exon 1A, c.-295G>A variant was located in the intron/exon boundary of exon 1B and c.-87C>T variant was located in 5'UTR, exon two. Subsequent sequencing carried out on the parents indicated that the genetic variants were inherited from the father who is unaffected and has another unaffected child (not studied). All of these variants were absent in the controls.

Two variants, c.-171C>G and c.-136C>G, were observed in the alternative exon 1B in the 5'UTR. c.-171C>G was found in a total of 10 patients and was seen in the controls, but at a significantly lower frequency ( $p=0.00007$ ). The c.-136C>G variant was identified in a patient with TGA, pulmonary atresia and ventricular septal defect (VSD). This variant was inherited from the father and was absent in the controls.

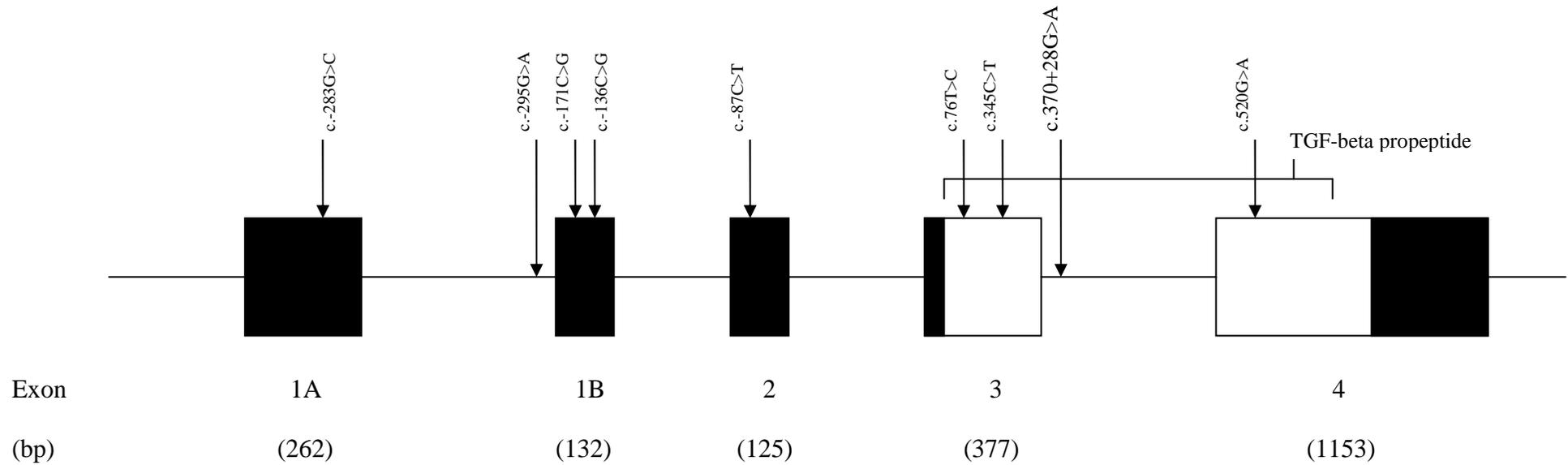
The non-coding variants that were absent in the controls were prioritized for *in silico* analysis. The c.-283G>C variant was not examined for its effect on *in silico* splicing as it was located in the first exon (see section 3.4.2.1). Other non-coding variants that were not observed in the controls (c.-295G>A, c.-136C>G and c.-87C>T) were investigated for effects on splicing (see section 3.4.2.1).

Position	Nucleotide Change	Location	Amino Acid Substitution	Number of Patients	Type of CVMs	Inheritance	Controls (N=1425)	Fisher's exact test (p value)
Exon 1A	c.-283G>C	5'UTR	-	1	TGA	Father	Absent	
Intron 1	c.-295G>A	Intron	-	1	TGA	Father	Absent	
Exon 1B	c.-171C>G	5'UTR	-	10	1-5. Complex 6-8. ASD 9. AVSD 10. Tricuspid atresia	NA	f(C)=0.983 f(G)=0.017	0.00007
	c.-136C>G	5'UTR	-	1	TGA	Father	Absent	
Exon 2	c.-87C>T	5'UTR	-	1	TGA	Father	Absent	
Exon 3	c.76T>C	Exonic	p.Leu26Leu	1	TGA	Father NA	f(T)=0.999 f(C)=0.001	0.282
	c.345C>T	Exonic	p.Asn115Asn	1	TGA	Mother	f(C)=0.998 f(T)=0.002	0.282
Intron 4	c.370+28G>A	Intronic		2	1. Coarctation of aorta 2. Hypoplastic left heart	Mother, Father	f(G)=0.963 f(A)=0.037	0.753
Exon 4	c.520G>A	Exonic	p.Gly174Ser	1	ASD	Mother	Absent	

**Table 3.3** *BMP4* – Genetic variants

The genetic variants identified in *BMP4* in the panel of 270 probands.

(TGA, transposition of the great arteries; ASD, atrial septal defect; AVSD, atrioventricular septal defect)



**Figure 3.2** *BMP4* – Location of genetic variants

Location of the genetic variation identified in *BMP4*, in the panel of 270 probands. Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pairs. 5'UTR and 3'UTR are in black colour. Location of the TGF-beta propeptide domain is labelled above the exons.

(TGF-beta: Transforming growth factor-beta)

Two synonymous variants were seen in exon three. The c.76T>C variant encoding leucine residue was not paternally inherited in a patient with TGA but the DNA from the mother was not available. The c.345C>T variant encoding for the residue asparagine was maternally inherited in a patient with TGA and VSD. Both of these variants were present in the controls. The c.370+28G>A variant was located in the intron of exon three and was seen in two patients, one with coarctation of the aorta and the other with hypoplastic left heart syndrome. This variant was present in the controls.

A non-synonymous substitution occurred in c.520G>A, in a patient with complete ASD, resulting in a glycine to serine change in residue number 174. This variant was maternally inherited and was not seen in the controls. The evolutionary conservation of the amino acid showed that the residue is highly conserved in all species examined, except fruitfly (Figure 3.3). However, *in silico* analysis of the protein function carried out using SIFT and PolyPhen predicted that the change is benign, therefore further *in vitro* investigation was not carried out.

**BMP4 p.G174S (c.520G>A)**

Human	GPDWERG <b>F</b> HRINIYEV
Chimpanzee	GPDWERG <b>F</b> HRINIYEV
Mouse	GPDWE <b>Q</b> G <b>F</b> HRINIYEV
Chicken	<b>SAA</b> WER <b>G</b> FHRINIYEV
Frog	<b>GPAWEE</b> G <b>F</b> HRINIYEV
Zebrafish	<b>DQTGDH</b> G <b>L</b> HRINIYEV
Fruitfly	<b>RSSANRTRYQVLVYDI</b>

**Figure 3.3 BMP4 – Evolutionary conservation of the variant p.G174S (c.520G>A)**

Evolutionary conservation of the non-synonymous unreported variant in BMP4 p.G174S (c.520G>A) and the surrounding sequences from human to fruitfly. (Blue: Non conserved region and red: amino acid residues of interest)

### 3.4.2 Investigation of effects on splicing of *BMP4*

#### 3.4.2.1 *In silico* prediction

The possibility that the four novel variants (c.-283G>C, c.-295G>A, c.-136C>G and c.-87C>T) that were not located in the translated region and were not found in the controls could affect splicing was investigated. The variant located in the first exon was considered to be unlikely to affect splicing, as exons located at the terminal ends of the gene, i.e. the first and the last exons, are known to require special mechanisms for their recognition (Berget 1995). Ohno and colleagues (Ohno et al. 1987) indicated that first exons end with a 5' splice site but have no processing signal in their beginning. They possess a modification at the beginning via the 7-methyl-guanosine cap which is essential for recognition and removal of the first intron. This cap also inhibits the formation of the spliceosome. The last exons are identified by factors recognizing the 3' splice site of the exons and interacting with factors recognizing poly(A) sites (Niwa et al. 1990). Therefore, it would be expected that the variants located in the first and last exons will not affect splicing.

The remaining three novel variants (c.-295G>A, c.-136C>G and c.-87C>T) were analysed using two *in silico* approaches. The first was the splice sites matrix library of ESE finder 3.0 (Cartegni et al. 2003; Smith et al. 2006) which is a web-based programme that analyses sequences to identify putative 5' splice sites, 3' splice sites and branch sites and to predict whether the mutations will disrupt these sites; the second was an in-house programme written by Dr. Sushma Grellscheid. The results of the ESE finder 3.0 analyses are shown in Table 3.4. A branch site was predicted to be created by the c.-295G>A variant. The branch site (BS) sequence is required for lariat formation and plays an important step in mRNA splicing. The computational method used a combination of both the branch site sequence and the polypyrimidine tract length to identify the putative branch site (Kol et al. 2005). However, the branch site sequences recognized by the splicing factors are known to be very degenerative (Gao et al. 2008). If a branch site were abolished, there would be other branch sites sequences upstream or downstream to facilitate splicing. Therefore, creation and or deletion of individual branch sites are unlikely to significantly affect splicing efficiency (Gao et al. 2008). The c.-136C>G variant was predicted to slightly decrease the affinity of 5' splicing if this nucleotide position were at a 5' splice site whereas the affinity of the branch site was predicted to decrease in variant c.-87C>T by the ESE finder programme. The matrices

<b>Position</b>	<b>Variant</b>	<b>ESE Finder 3.0</b>
Exon 1B	c.-295G>A	Branch site created
	c.-136C>G	5' splice site (donor) decreased
Exon 2	c.-87C>T	Branch site decreased

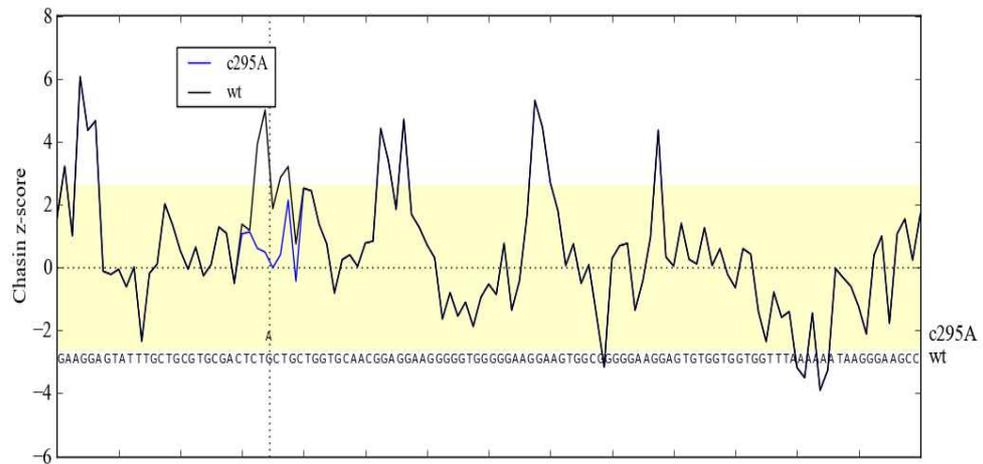
**Table 3.4** *BMP4* – Potential effect of variation on splicing

Effects of the genetic variants located in the 5'UTR or intronic region on putative exonic splicing elements predicted *in silico* using ESE Finder 3.0.

for 5' and 3' splice sites used by ESE finder were derived from dbCASE, a database of classified alternative splicing events (Zhang et al. 2007) to identify putative 5' and 3' splice sites. The sequences recognized by the splicing factors (branch sites, 5' and 3' splice sites) seem to be very degenerate; therefore it is difficult for the programme to discern the pseudo splice site from the authentic splice sites. When predicting the effect the variant may have on splicing, the computational method implemented in ESE finder 3.0 does not take into consideration the real exon/intron boundaries being analysed: it will analyse the sequence and will search for the possible 5' and 3' sites and attribute a score to those sites identified. Thus, it is possible for the programme to find a decreased 5' site where there is no 5' site in the input sequence. c.-136C>G is not at a 5' splice site so these results are difficult to interpret. Besides that, the default threshold values used by ESE finder are, although based on statistical analysis and empirical data, somewhat arbitrary. Given the ambiguity of these results from ESE finder, a second programme was used to analyse these variants for potential effects on splicing.

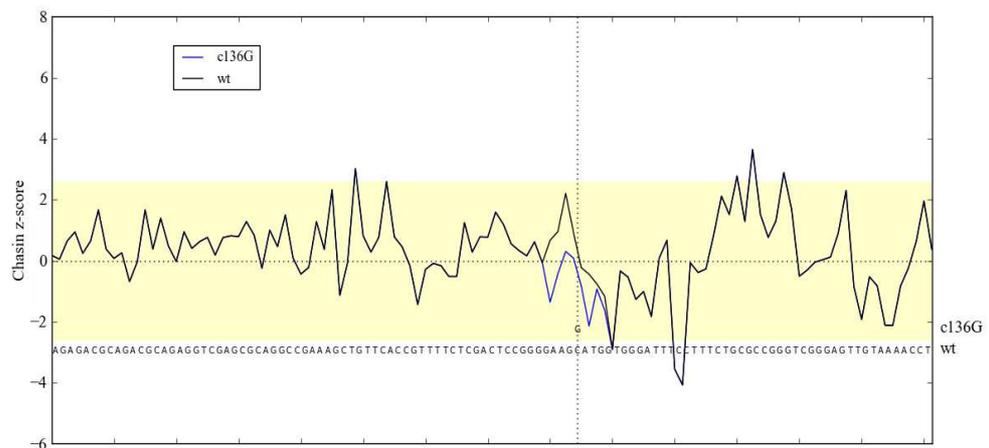
The second *in silico* analysis was performed using an in-house programme that was generated by Dr. Grellscheid. In figures 3.4 to 3.6, the x-axis shows the sequence, the positive and negative y-axis indicates ESE and ESS scores respectively. The threshold region within which any variation is not considered significant is shown in yellow.

The programme predicted that an ESE site that was present in the c.-295G variant was abolished when the variant was substituted to c.-295A (Figure 3.4). The change seemed to be significant, as a big peak (i.e. an important ESE) is lost. Since a potential ESE site is lost, this might imply that the spliceosome cannot bind to this site anymore. For the c.-136C>G and c.-87C>T variants, the effect of the substitution was not important, as the scores fell in the 'acceptable threshold region' (the yellow part) (Figure 3.5 and Figure 3.6). Therefore, based on the prediction, it would be expected that the c.-295 wild type allele (G) to splice correctly whereas the variant allele (A) might splice incorrectly, leading to the inclusion of the intronic sequence before the inserted exon 1B in the minigene assay. The c.-136C>G and c.-87C>T variants would be less likely to have effects on splicing. Nevertheless, these results generated by the programmes are prediction only, and cannot be considered conclusive. As these variants were absent in the 1000 controls, *in vitro* functional analysis using a minigene assay was carried out. The variant of principal interest was c.-295A>G, but having invested the effort to develop the assay, the c.-136C>G and c.-87C>T variants were investigated too.



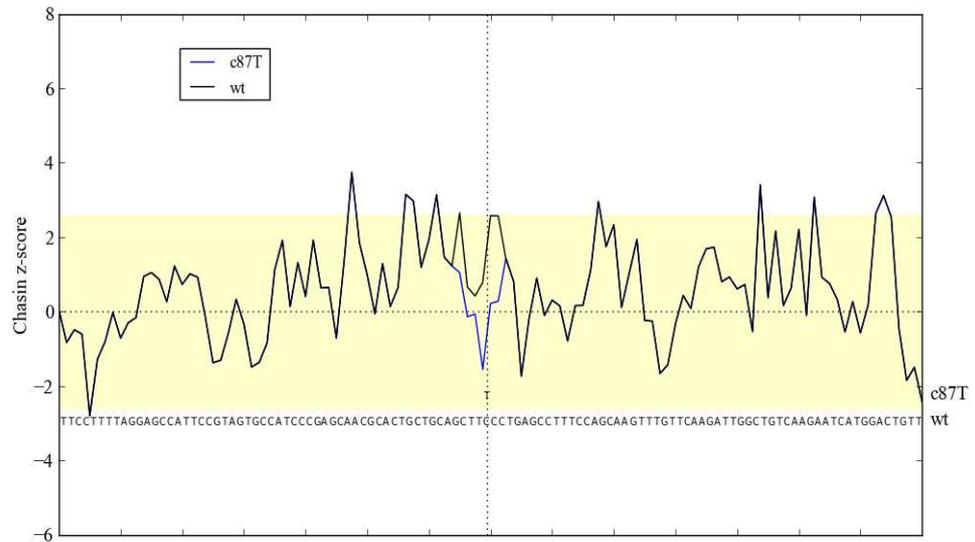
**Figure 3.4 BMP4 c.-295G>A *in silico* prediction**

*In silico* analysis of c.-295G>A variant using the programme created by Dr. Sushma Grellscheid showed the loss of an ESE site (the blue line).



**Figure 3.5 BMP4 c.-136C>G *in silico* prediction**

*In silico* analysis of c.-136C>G variant using the programme created by Dr. Sushma Grellscheid showed the decrease score of an ESE site but the reduction of the score (the blue line) was within the 'acceptable range' (yellow part).



**Figure 3.6 BMP4 c.-87C>T *in silico* prediction**

*In silico* analysis of c.-87C>T variant using the programme created by Dr. Sushma Grellscheid showed the reduction of the score (the blue line) but within the acceptable range (yellow part).

### 3.4.2.2 Minigene assays for *BMP4* variants

Three novel variants (c.-295G>A, c.-136C>G and c.-87C>T) that were predicted to affect splicing and were absent in normal controls were investigated further using minigene assays. The aims of this experiment were:

- To clone the genomic segments containing the wild-type and variant alleles into the minigene construct.
- To determine whether the allelic variants had any effect on splicing efficiency.

#### 3.4.2.2.1 Novel variant c.-295G>A

The novel variant c.-295G>A was predicted to create a branch site *in silico* by ESE finder 3.0. Further *in silico* analysis utilising the programme generated by Dr. Grellscheid predicted the variant to have a significant effect on splicing as an important ESE site is lost. Hence, an *in vitro* functional assay was carried out to determine whether the variant has any effect on splicing. The wild type (c.-295G) and the mutant (c.-295A) BMP minigene constructs were transfected into HEK293 cells. The variant was hypothesized might result in the inclusion of the intron between  $\beta$ -globin exon 2 and the inserted exon of BMP4, i.e. a failure of splicing between  $\beta$ -globin exon 2 and the inserted BMP4 exon (as shown in Methods chapter, Figure 2.7)

PCR-amplified genomic fragments containing the c.-295 variants were cloned into the pXJ41 vector using primers complementary to regions in BMP4. Positive colonies that were Ampicillin resistant were picked and amplified (Figure 3.7). A total of six colonies which contained the fragment of interest were identified. PCR of the six colonies carried out to determine the orientation of the insert showed that two colonies had the BMP4 fragment inserted in the correct orientation (Figure 3.8). These constructs were then sequenced to identify the wild type allele minigene construct and the mutant allele minigene construct. Fortuitously, one colony carried the wild type allele and one carried the mutant allele.

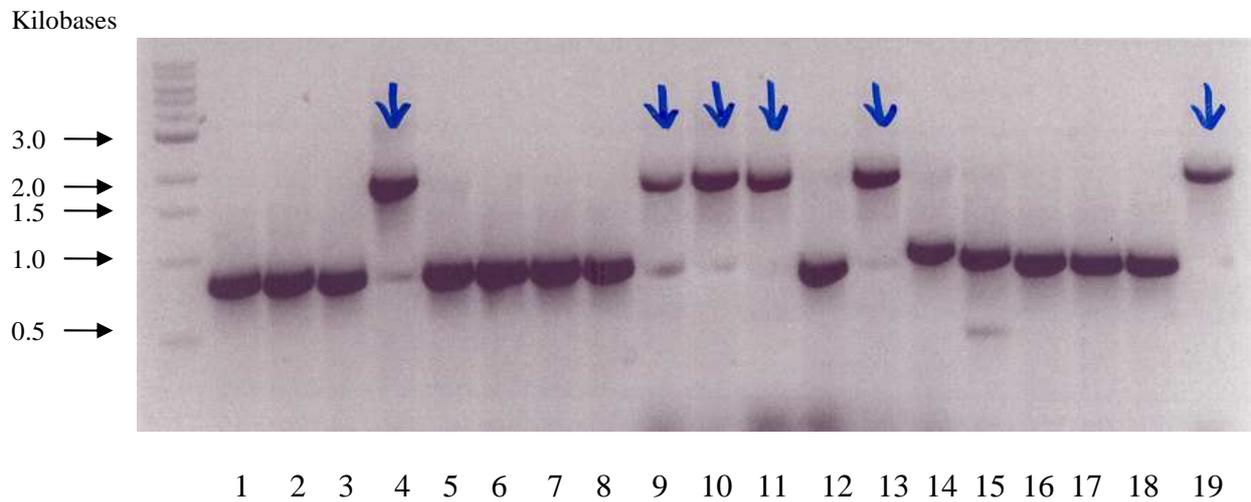
HEK293 cells were transfected with wild type and mutant minigene constructs, and RT-PCR of RNA extracted from these cells was used for comparison. The same size products were observed from both minigenes at about 200bps, showing that the

fragment of interest was correctly spliced, with  $\beta$ -globin exons (18bps of exon 2 and 53bps of exon 3) flanking the fragment (132bps) (Figure 3.9, lanes 1 and 2). Hence, this variant did not affect splicing in this assay.

#### **3.4.2.2.2 Novel variant c.-136C>G**

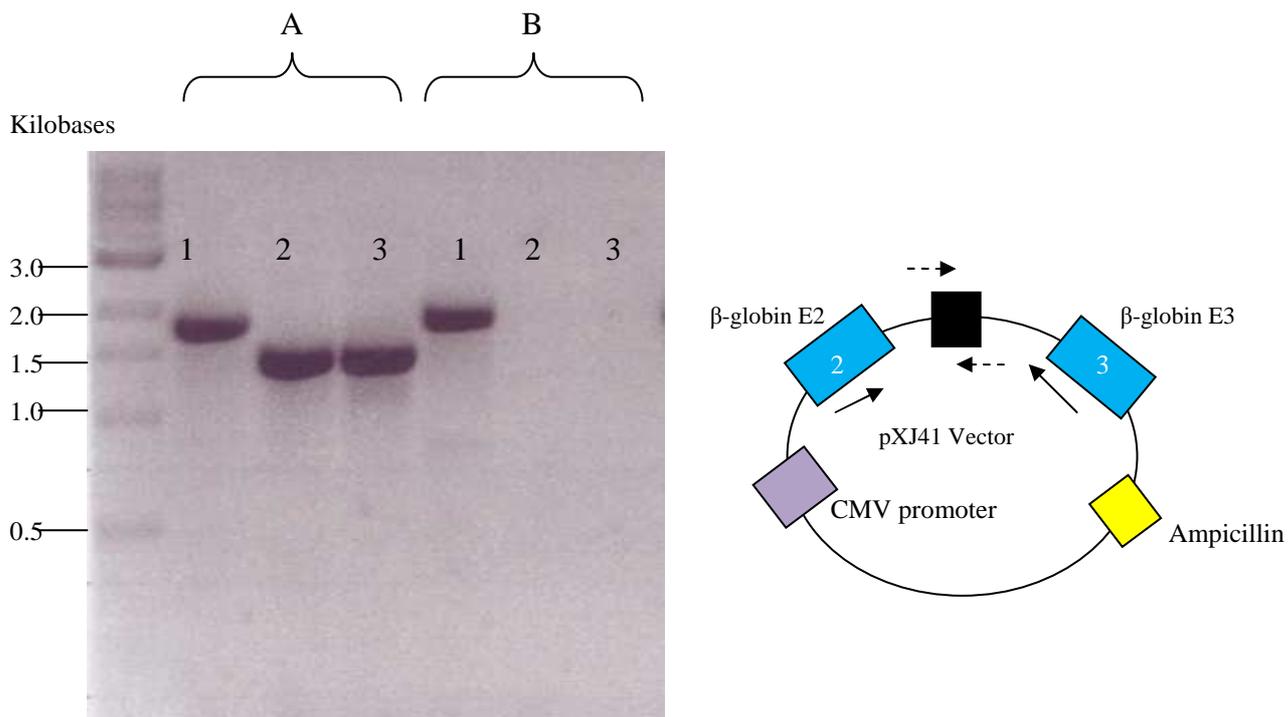
The novel variant c.-136C>G was predicted by ESE finder 3.0 to decrease the 5' splice site. However, the variant is not located at the exon/intron boundary. Therefore, the prediction of that programme is probably not relevant for this variant. Another *in silico* programme did not predict a strong effect of the variant on splicing. Nevertheless, the absence of this variant in 1000 controls warranted further investigation.

After transfection of wild type and mutant constructs, a total of eleven positive colonies that were Ampicillin resistant and contained the fragment of interest were identified. However, only three colonies had the fragment of interest inserted in the correct orientation. Constructs were sequenced to identify the allele of the minigene, one wild type and two mutant constructs were obtained. RT-PCR of RNA extracted from cells transfected with these two minigene constructs showed only one band at about 200bps, indicating no difference between the wild type and the mutant allele splicing pattern. Figure 3.9, lanes 3 and 4, shows the results of the splicing analysis after transfection in HEK 293 for the wild type (c.-136C) and mutant (c.-136G) minigene constructs. These results indicate that the variant did not have an effect on splicing *in vitro*.



**Figure 3.7 PCR screening of Ampicillin resistant colonies**

Example of colonies that contained the fragment of interest, approximately 1.5kb, are indicated with arrows (lanes 4, 9, 10, 11, 13 and 19). The smaller bands are colonies containing empty plasmids.



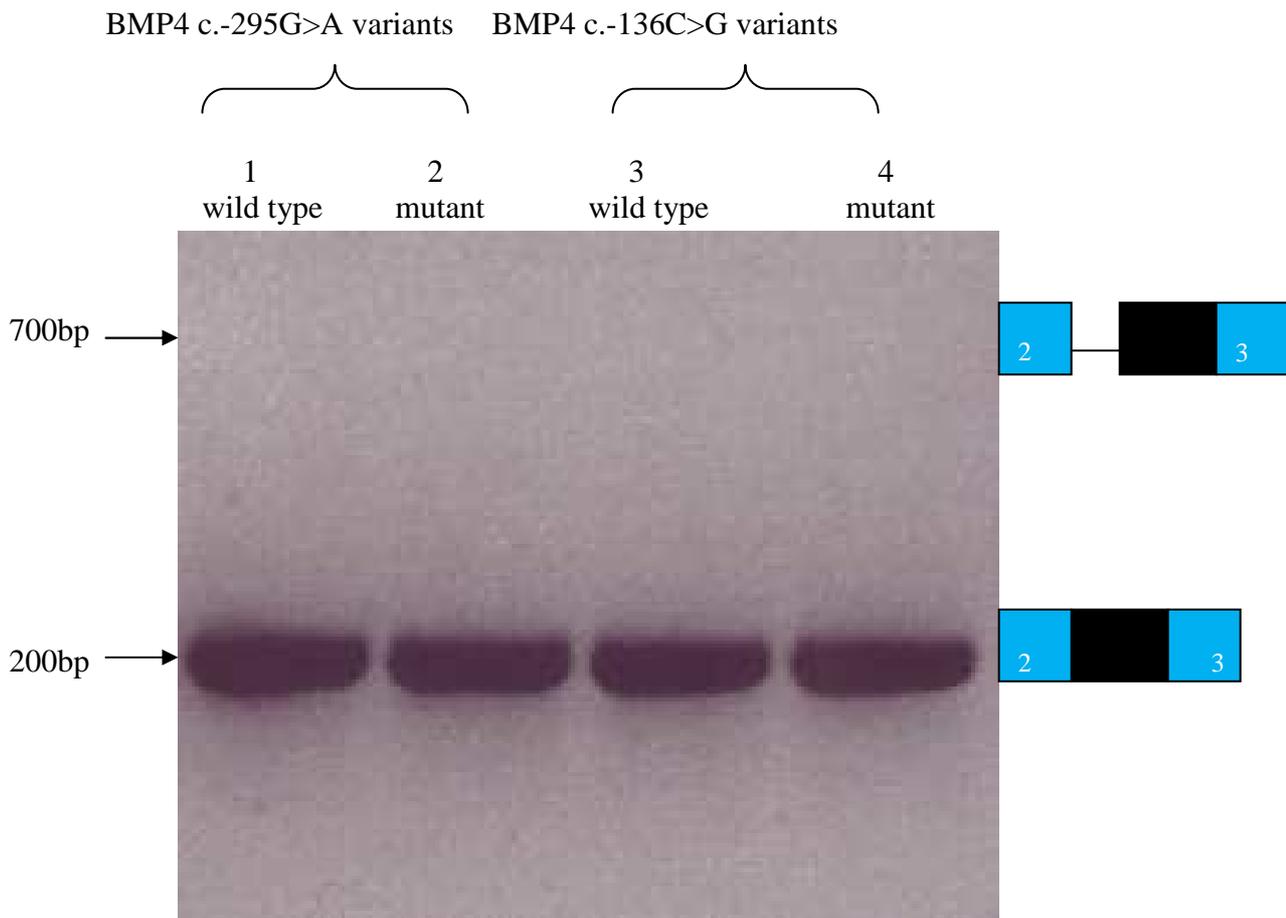
**Figure 3.8 Determination of fragment orientation**

On the gel, lanes A show a colony in which the construct (wild type or mutant) is in the correct orientation, and lanes B show a colony in which the construct is in the incorrect orientation.

Lanes marked 1 on the gel contain product generated using primers pXJ-Forward ( $\rightarrow$ ) and pXJ-Reverse ( $\leftarrow$ ) shown in schematic. Regardless of orientation, a  $\sim 1.7$  kb band is expected.

Lanes marked 2 on the gel contain product generated using pXJ-Forward ( $\rightarrow$ ) and a primer that is reverse complementary to the fragment of interest ( $\leftarrow$ ). A correctly oriented insertion will produce a  $\sim 1.5$  kb band, and an incorrectly oriented insertion no band.

Lanes marked 3 on the gel contain product generated using a forward primer complementary to the fragment of interest ( $\rightarrow$ ) and pXJ-Reverse ( $\leftarrow$ ). A correctly oriented insertion will produce a  $\sim 1.5$  kb band, and an incorrectly oriented insertion no band.



**Figure 3.9 Splicing patterns of the variants**

The splicing patterns of the wild type allele BMP4 c.-295G (lane 1), mutant allele BMP4 c.-295A (lane 2), wild type allele BMP4 c.-136C (lane 3) and mutant allele BMP4 c.-136G (lane 4) analysed using RT-PCR. (Ladders not shown). No band was seen at 700bp which was what would have been expected if splicing had not taken place correctly between  $\beta$ -globin exon 2 and the inserted exon 1B of BMP4. The blue boxes indicate  $\beta$ -globin exons and the black box indicates BMP4 exon.

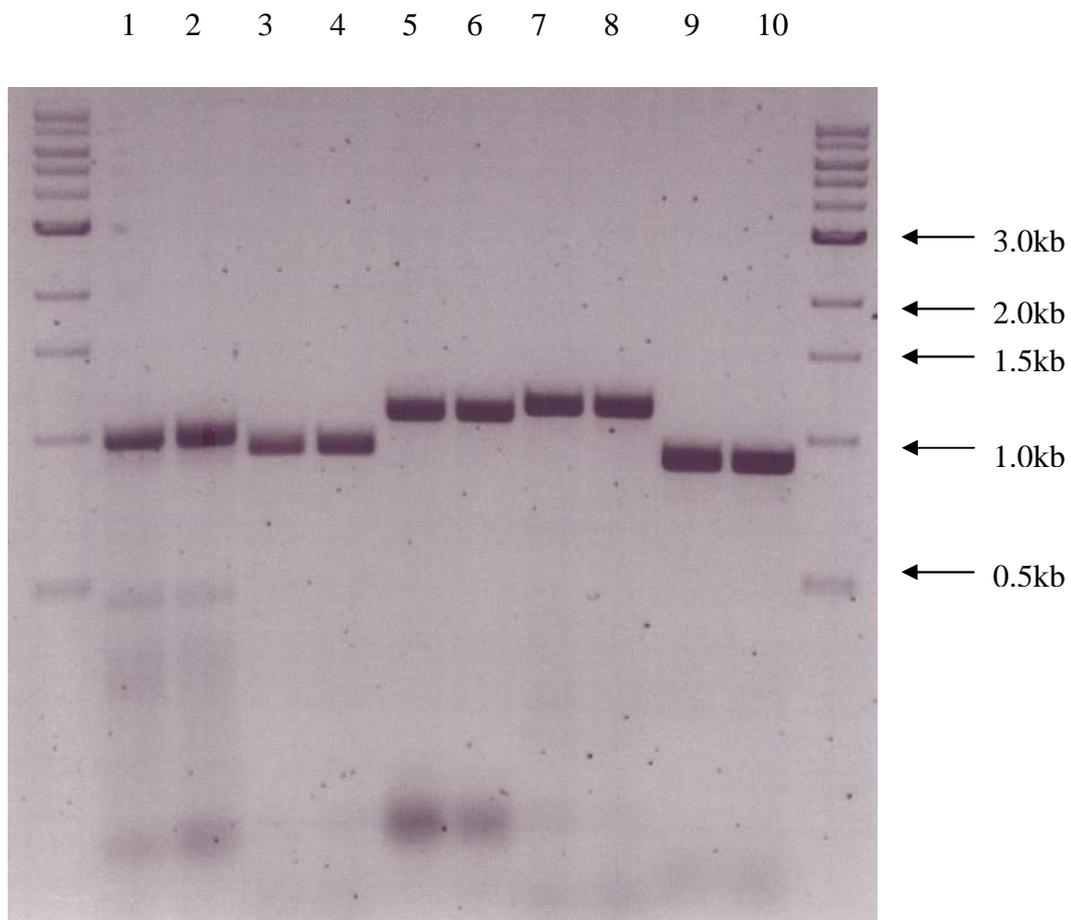
### 3.4.2.2.3 Novel variant c.-87C>T

The novel variant c.-87C>T was predicted by ESE finder 3.0 to decrease branch site efficiency. As stated previously, branch sites consensus sequences are degenerative in the genome. For this reason, the predicted branch site might be a false positive and have no effect on splicing. The second *in silico* programme did not predict a major effect of this variant on splicing. Nevertheless, since this variant is absent in 1000 controls, the minigene assay was proceeded.

The amplification of the fragment from the patient's DNA showed a clear specific band. This fragment was then digested with restriction enzyme and ligated with the vector and transformed into competent cells as described in chapter 2. The first screening of the transformed cells (approximately 50 colonies) showed that the fragment of interest was not inserted into the vector. The second screening PCR of the transformed cells also yielded negative results.

The next step taken was to use different ratios of vector to insert DNA during ligation. The ratios used were 1:3, 1:1 and 3:1. The different ligation reactions were then transformed into competent cells as described in chapter 2. Subsequent screening PCR of these cells showed that the cells did not contain the fragment of interest.

Alternative primer pairs were then designed to amplify the fragment of interest in order to optimize ligation. The fragments were successfully amplified using the Phusion High-Fidelity DNA Polymerase as shown in Figure 3.10. However, the screening PCR of the transformed cells again yielded negative results. It was concluded that more optimization of the cloning parameters would have been needed in order to clone this fragment. As time was a constraint and potentially more significant results had been found in other genes screened, this experiment further was not pursued further.



**Figure 3.10 Amplified fragments of variant BMP4 c.-87>T using different primer pairs**

All of the amplified fragments are around 1kb. Lanes 1 and 2 are of primers MG-BMP4EF/MG-BMP4ER. Lanes 3 and 4 are of primers MG-BMP4FF/MG-BMP4FR. Lanes 5 and 6 are of primers MG-BMP4GF/MG-BMP4GR. Lanes 7 and 8 are of primers MG-BMP4FF/MG-BMP4GR. Lanes 9 and 10 are of primers MG-BMP4GF/MG-BMP4FR.

### 3.5 *SMAD6*

The three isoforms of *SMAD6* with alternative first and fourth exons were sequenced in 90 probands. The genetic variants identified are recorded in Table 3.5 and the locations of the variants are shown in Figure 3.11. These variants were genotyped in the first control population. Ten previously unreported genetic variants were identified. Of these novel genetic variants, five did not result in amino acid changes; three variants were located in the 5'UTR and two variants in the 3'UTR region. In the 5'UTR region, variant c.-428A>T was identified in a patient with Patent Foramen Ovale (PFO) and variant c.-290G>C was found in a patient with complex phenotypes. Both variants were absent in the control population. However, as these variants (c.-428A>T and c.-290G>C) were located in exon 1 (see section 3.4.2.1), their effects on splicing were not investigated. Variant c.-44C>T was found in three patients: one patient with complex phenotypes and two patients with AVSD. This variant was seen in the controls with a minor allele frequency of 0.027. In the 3'UTR region, the c.\*262T>C variant was inherited from the parent and was present in the controls with a minor allele frequency of 0.023. The c.\*319T>A variant was identified in a patient with PFO and was not seen in the control population.

Three novel variants were identified in the MH1 domain. c.61G>A was seen in an ASD patient, resulting in p.Asp21Asn. It was inherited from the patient's father. Two synonymous single base substitutions occurred. c.120C>T was heterozygous in 7 probands and c.711C>T was seen in a patient with Ebstein's anomaly. All these three variants were present in the controls with a minor allele frequency of 0.022, 0.029 and 0.002 respectively.

A non-synonymous substitution was identified in the alternatively spliced region of *SMAD6* isoform C, in exon 4B, in a patient with large secundum ASD. The c.975C>G variant resulted in p.Ser325Arg substitution. DNA of the parents was not available. This variant was not seen in the controls. However, as this variant is located in the linker region with relatively less likely functional effects when compared to the MH1 and MH2 domains of *SMAD6*, further investigation was not performed.

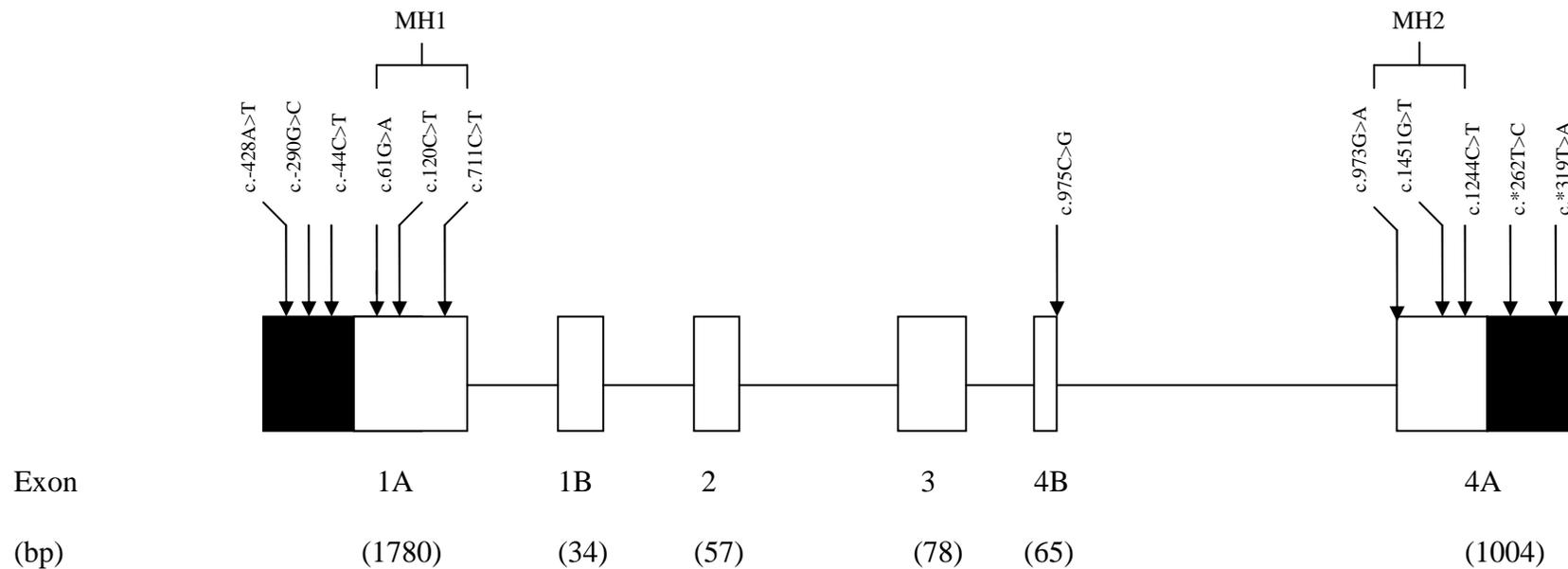
A non-synonymous variant, c.1451G>T, resulting in p.Cys484Phe, located in the MH2 domain, was identified in a patient with coarctation and calcification of the aorta. This

Position	Nucleotide Change	Location	Amino Acid Substitution	Number of Patients	Type of CVMs	Inheritance	Controls (N=1425)	Fisher's exact test (p value)
Exon 1	c.-428A>T	5'UTR	-	1	PFO	N/A	Absent	
	c.-290G>C	5'UTR	-	1	Complex	N/A	Absent	
	c.-44C>T	5'UTR	-	3	1. Complex 2-3. AVSD	N/A	f(C)=0.973 f(T)=0.027	0.722
	c.61G>A	Exonic	p.Asp21Asn	1	ASD	Father N/A	f(G)=0.978 f(A)=0.022	1.000
	c.120C>T	Exonic	p.Gly40Gly	7	1-2. VSD 3-4. PFO 5. Pulmonary stenosis 6-7. Complex	N/A	f(C)=0.971 f(T)=0.029	0.029
	c.711C>T	Exonic	p.His237His	1	Ebstein's anomaly	Mother	f(C)=0.998 f(T)=0.002	0.282
Exon 4B	c.975C>G	Exonic	p.Ser325Arg	1	ASD	N/A	Absent	
Exon 4A	c.1451G>T	Exonic	p.Cys484Phe	1	Coarctation of aorta	N/A	Absent	
	c.*262T>C	3'UTR	-	2	1. ASD 2. Aortic stenosis	Mother; Father N/A	f(T)=0.977 f(C)=0.023	1.000
	c.*319T>A	3'UTR	-	1	PFO	N/A	Absent	

**Table 3.5 SMAD6 – Genetic variants**

The genetic variants identified in *SMAD6* in the panel of 90 probands.

(PFO, patent foramen ovale; AVSD, atrioventricular septal defect; VSD, ventricular septal defect; ASD, atrial septal defect)



**Figure 3.11** *SMAD6* – Location of genetic variants

Location of the genetic variation identified in *SMAD6*, in the panel of 90 probands (except variants c.973G>A and c.1451G>T). Variants c.973G>A and c.1451G>T were subsequently identified in additional 348 probands. Boxes represent exons and adjoining lines represent introns. Exons are numbered and the sizes are indicated in base pairs. 5'UTR and 3'UTR are in black colour. Location of the MH1 and MH2 domains are labelled above the exons.

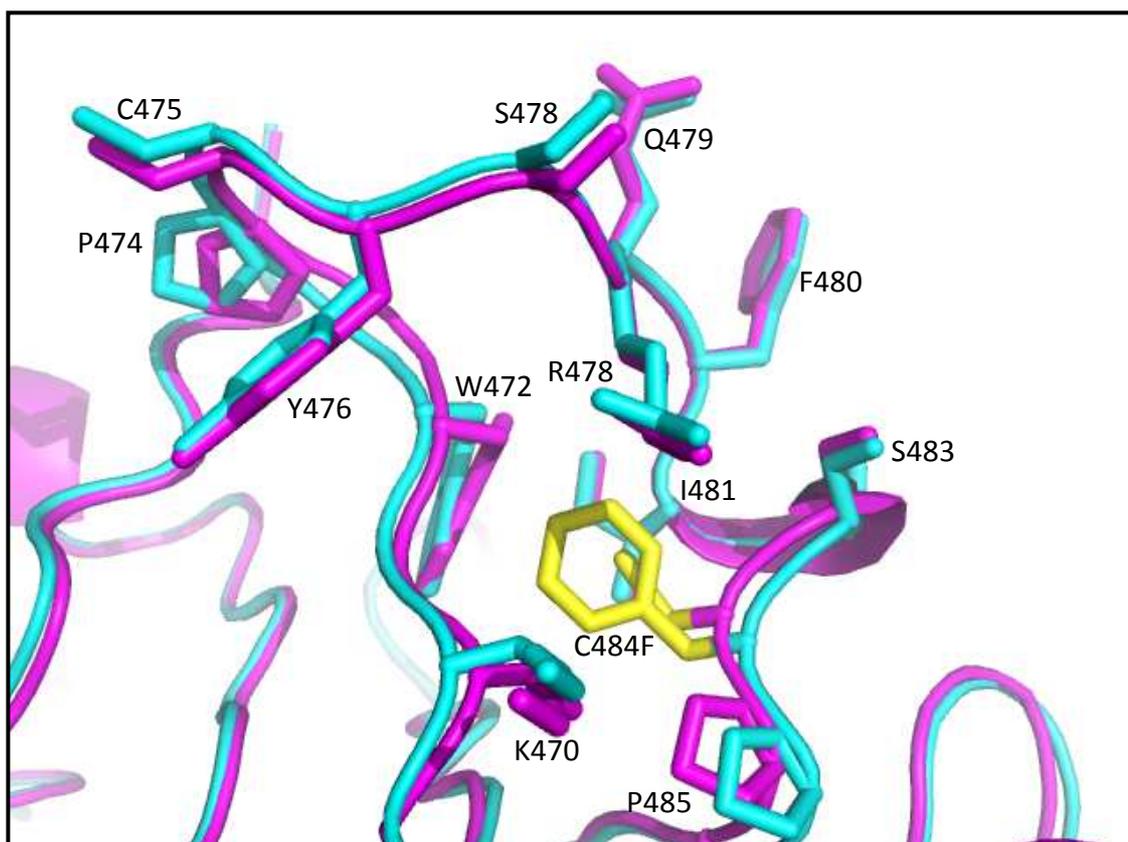
(MH1: Mad homology 1, MH2: Mad homology 2)

variant was not seen in the controls. The inheritance of the variant was not known as both parents were deceased.

The potential functional effect of the non-synonymous novel variants was analysed using SIFT and PolyPhen programmes. For p.Asp21Asn (c.61G>A) and p.Ser325Arg (c.975C>G) (alternatively spliced exon 4B) variants, SIFT predicted both variants to be poorly tolerated but PolyPhen predicted them to be benign. However, both algorithms predicted the p.Cys484Phe (c.1451G>T) variant to affect protein function.

Therefore, the crystal structure of the variant predicted to be damaging was further investigated *in silico*. The crystal structure of the SMAD1 homotrimeric complex (Protein Data Bank entry 1KHU), which is homologous to the MH2 domain of SMAD6, was utilised as a basis for modelling the SMAD6 structure and the p.Cys484Phe (c.1451G>T) variant. Models of the wildtype SMAD6 and the p.Cys484Phe (c.1451G>T) variant structures were generated using the same procedures for their comparison. The change of the cysteine to the phenylalanine resulted in significant changes in structure that were limited to the immediate vicinity of residue 484. In the wildtype model, Cys484 is buried from solvent by the 'L3 loop' of the MH2 domain and the thiol packs, in particular, against the indole ring of Trp472. To accommodate the extra bulk of a phenylalanine sidechain in the mutant protein, the structure in the immediate vicinity of the L3 adopted an altered conformation (Figure 3.12). Glycines 471 and 473, which immediately flank the tryptophan, provide the potential for the necessary flexibility in this part of the structure of the L3 loop. Since the L3 loop is known to interact with the L45 loop of BMPR1A (Imamura et al. 1997), it seems most likely that the p.Cys484Phe (c.1451G>T) variant affects binding of SMAD6 to BMPR1A, which is an inhibitory interaction.

Based on this finding, the MH2 domain of SMAD6 was re-sequenced in an additional 348 probands with a broad range of CVM. Two further non-synonymous variants were identified in SMAD6: c.973G>A resulted in the amino acid substitution of p.Ala325Thr in a patient with congenital mitral regurgitation and c.1244C>T that resulted in p.Pro415Leu substitution was found in a patient with congenital aortic stenosis. Both of these variants were inherited from phenotypically normal fathers (Table 3.6) (Figure 3.11). These two variants were genotyped in the second control population consisted of 1000 individuals that was derived from the North Cumbria Community Genetics Project



**Figure 3.12** Predicted crystal structure of SMAD6 p.Cys484Phe (c.1451G>T)

The homology models of wildtype (pink) and Cys484Phe (cyan) SMAD6 are superimposed, with residue 484 highlighted in both models in yellow. The individual amino acids in the L3 loop are drawn as stick models and the remainder of the structure as a ribbon cartoon. Note that Trp472 (W472) in particular has to move to accommodate the extra bulk presented by the mutation of Cys484 to Phe (performed by Prof. Rick Lewis).

<b>Position</b>	<b>Nucleotide Change</b>	<b>Location</b>	<b>Amino Acid Substitution</b>	<b>Number of Patients</b>	<b>Type of CVMs</b>	<b>Inheritance</b>
Exon 4A	c.973G>A	Exonic	p.Ala325Thr	1	Mitral valve regurgitation	Father
	c.1244C>T	Exonic	p.Pro415Leu	1	Aortic stenosis	Father

**Table 3.6 MH2 domain of *SMAD6* – Genetic variants**

The genetic variants identified in the MH2 domain of *SMAD6* in the replication cohort of 348 probands. Both variants were inherited from their respective father. These variants were genotyped in the second control population consisted of 1000 individuals. The p.Ala325Thr (c.973G>A) variant was not successfully genotyped in the controls but the p.Pro415Leu (c.1244C>T) variant was not observed in the controls (performed by Dr. Darroch Hall).

(NCCGP) (performed by Dr. Darroch Hall). The p.Pro415Leu (c.1244C>T) variant was successfully genotyped in the controls (not present in the controls) but the p.Ala324Thr (c.973G>A) variant failed.

*In silico* analyses by both SIFT and Polyphen programmes predicted the p.Ala325Thr (c.973G>A) variant to be benign but the p.Pro415Leu (c.1244C>T) variant to be probably damaging to protein function, as might be expected since prolines are generally accepted to be important for protein folding and stability. Similar comparative structural modelling of the p.Pro415Leu (c.1244C>T) variation suggests that H2, which, unusually, is N-capped by Pro415, is destabilised by mutation to leucine. Residue 415 is solvent accessible and on the same face of the MH2 domain as the L3 loop, perhaps suggesting that p.Pro415Leu (c.1244C>T) is also affected in receptor signaling (not shown). The p.Pro415Leu (c.1244C>T) variant was also absent in 1000 control individuals. The p.Ala325Thr (c.973G>A) variant falls outside the structurally characterised region of the MH2 domain of the SMAD family, and thus the likely effects of the mutation cannot be determined in this way.

The evolutionary conservation of the amino acid substitutions are shown in Figure 3.13. The aspartic acid of p.Asp21Asn (c.61G>A) was conserved only in chimpanzee. The p.Ser325Arg (c.975C>G) (alternative exon 4B) variant was not conserved. The alanine of the p.Ala325Thr (c.973G>A) substitution was conserved in chimpanzee, mouse and chicken. The proline residue of p.Pro415Leu (c.1244C>T) and cysteine residue of p.Cys484Phe (c.1451G>T) showed very high degree of conservation, from human to fruitfly.

The c.\*262T>C variant was predicted *in silico* to increase base pairing of hsa-miR-219-2-3p (Figure 3.14) but this variant was observed in the controls. No predicted microRNAs was located at the c.\*319T>A variant even though this variant was not observed in the controls. Therefore, both variants were not pursued further.

Based upon these results, functional analysis of the non-synonymous variants p.Ala325Thr (c.973G>A), p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) was carried out, results of which are presented in the next chapter.

### **SMAD6 p.D21N (c.61G>A)**

HUMAN	VRRLWRS-RVVPDREEGGSGGGG
CHIMPANZEE	-----DREEGGSGGGG
MOUSE	VRRLWRSRVVPDREEGSG--GGG
CHICKEN	VRRLWRSRVIPERDGGDGNGQS-
FROG	VRRLWRS-----
ZEBRAFISH	-----
FRUITFLY	-----

### **SMAD6 p.S325R (c.975C>G)**

HUMAN	-----SRG-----
CHIMPANZEE	SKEPDGVWAYNRGEHPHFVNS
MOUSE	SKEPDGVWAYNRGEHPHFVNS
XENOPUS	SREADGVWAYNRSDHPHFVNS
CHICKEN	SKEPDGVWAYNRSEHPHFVNS
ZEBRAFISH	SKEPDGVWAYNRSQHPHFVNS
FRUITFLY	SLENGDVWIYNRGNTTIFVDS

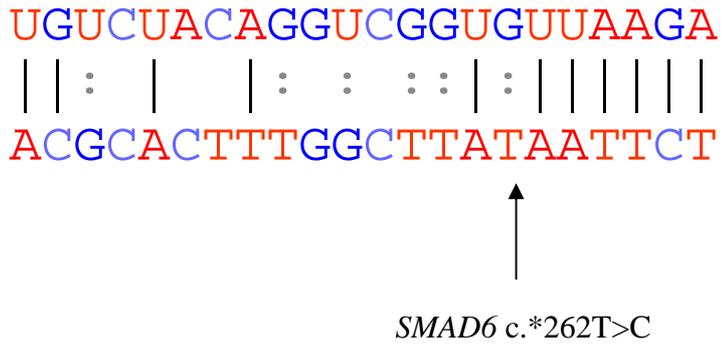
	<b><u>p.A325T</u></b> <b><u>(c.973G&gt;A)</u></b>	<b><u>p.P415L</u></b> <b><u>(c.1244C&gt;T)</u></b>	<b><u>p.C484F</u></b> <b><u>(c.1451G&gt;T)</u></b>
HUMAN	SMSPD <b>A</b> TKPSH	IFVNS <b>P</b> TLDAP	QFITS <b>C</b> PCWLE
CHIMPANZEE	SMSPD <b>A</b> TKPSH	IFVNS <b>P</b> TLDAP	QFITS <b>C</b> PCWLE
MOUSE	SMSPD <b>A</b> TKPSH	IFVNS <b>P</b> TLDAP	QFITS <b>C</b> PCWLE
CHICKEN	<b>S</b> TSPD <b>A</b> VKRSH	IFVNS <b>P</b> TL <b>D</b> I <b>P</b>	QFITS <b>C</b> PCWLE
XENOPUS	<b>S</b> LSPD <b>M</b> SK <b>Q</b> GH	IFVNS <b>P</b> TLDAP	<b>Q</b> MIT <b>S</b> <b>C</b> PCWLE
ZEBRAFISH	SMSP <b>S</b> SL <b>A</b> Q <b>N</b> H	IFVNS <b>P</b> T <b>L</b> E <b>H</b> H	QFITS <b>C</b> PCWLE
FRUITFLY	<b>D</b> G <b>K</b> D <b>H</b> N <b>I</b> N <b>S</b> <b>Q</b> V	IFV <b>D</b> S <b>P</b> T <b>L</b> S <b>E</b> N	<b>Q</b> D <b>I</b> M <b>G</b> <b>C</b> PCWLE

**Figure 3.13 SMAD6 – Evolutionary conservation of the variants**

Evolutionary conservation of the non-synonymous variants discovered in SMAD6:

p.Aps21Asn (c.61G>A), p.Ala325Thr (c.973G>A), p.Pro415Leu (c.1244C>T),

p.Cys484Phe (c.1451G>T) and p.Ser325Arg (c.975C>G). (Blue: Non conserved region and red: amino acid residues of interest)



**Figure 3.14 SMAD6 – Predicted miRNA binding site**

*In silico* prediction that the c.\*262T>C variant will increase base pairing of hsa-miR-219-2-3p. However, this variant was observed in the controls.

### 3.6 *BMPRIA*

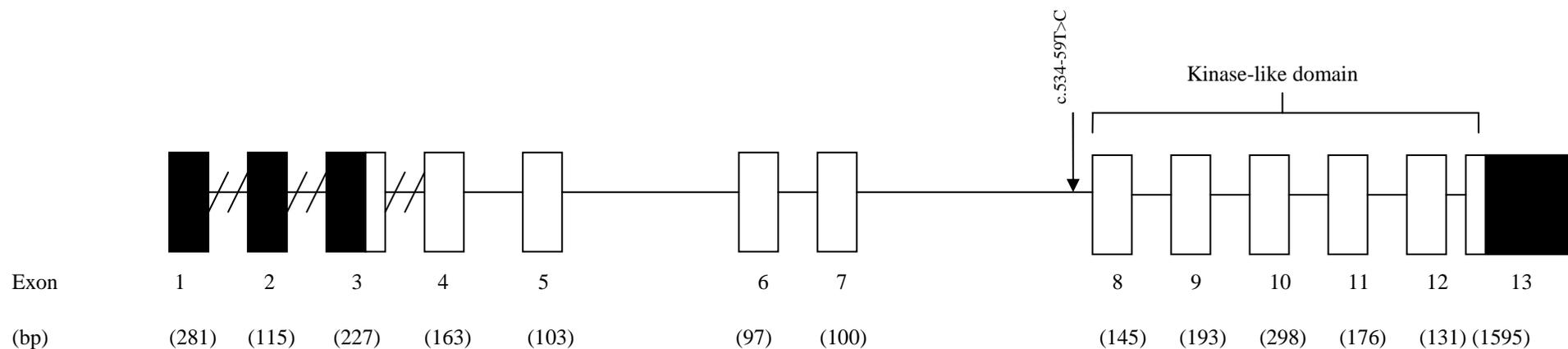
The genetic variants identified in *BMPRIA* are summarised in Table 3.7. Only one previously unreported variant was present in this gene in a patient with atrial septal defect (ASD) out of 90 patients that were sequenced. The location of the variant is shown in Figure 3.15. This variant was located in intron seven, 59 bases upstream of exon eight. This variant was not genotyped in the control population. However, data from the 1000 genome project (Durbin et al. 2010) that became available at a later date showed that this variant (rs118172017) is present in the normal population, indicating that this variant is a polymorphism.

Position	Nucleotide Change	Location	Type of CVM
Exon 1	-	-	-
Exon 2	-	-	-
Exon 3	-	-	-
Exon 4	-	-	-
Exon 5	-	-	-
Exon 6	-	-	-
Exon 7	-	-	-
Intron 7	c. 534-59T>C	intronic	ASD
Exon 8	-	-	-
Exon 9	-	-	-
Exon 10	-	-	-
Exon 11	-	-	-
Exon 12	-	-	-
Exon 13	-	-	-

**Table 3.7** *BMPRIA* – Genetic variant

The genetic variant identified in the intronic region of *BMPRIA* in the panel of 90 probands.

(ASD, atrial septal defect)



**Figure 3.15** *BMPRIA* – Location of the genetic variant

Location of the genetic variation identified in *BMPRIA*, in the panel of 90 probands. Boxes represent exons and adjoining lines or broken lines represent introns. Exons are numbered and the sizes are indicated in base pairs. 5'UTR and 3'UTR are in black colour. Location of the kinase-like domain is labelled above the exons.

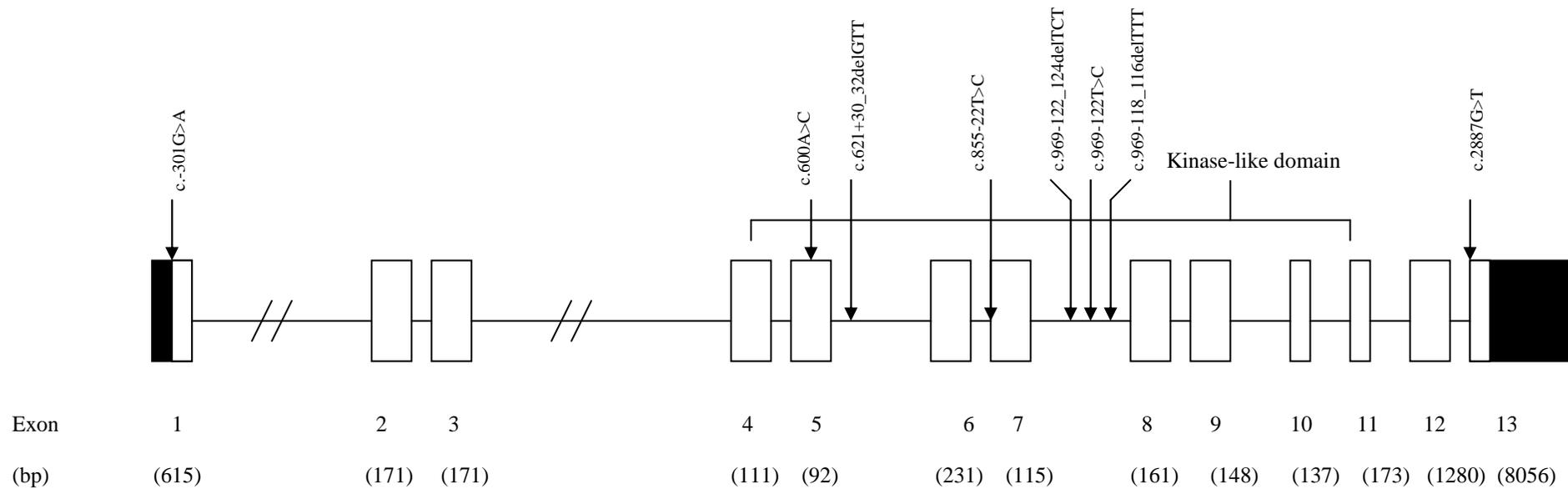
### 3.7 *BMPR2*

*BMPR2* was sequenced in 90 patients. A total of eight genetic variants were identified. This is summarized in Table 3.8. The locations of the variants are shown in Figure 3.16. However, these variants were not genotyped in the control population. Variants that were located in the coding region were: c.-301G>A in the 5'UTR, c.600A>C in exon five and c.2887G>T in exon thirteen. The c.600A>C variant resulted in a synonymous leucine substitution and was reported by Machado et al. (Machado et al. 2006), to be a polymorphism, with frequency of 2.4%. The c.2887G>T variant which results in the substitution p.Gly963Cys was found in an ASD patient. This variant was predicted by both SIFT and Polyphen programmes to be probably damaging. The glycine residue is conserved in chimpanzee, mouse, chicken and zebrafish but is substituted to serine in frog and arginine in fruitfly (Figure 3.17). A total of 5 variants were located in the intronic region. As rare non-synonymous variants located in the coding region are expected to have greater effect on disease susceptibility than non-coding variants (Kryukov et al., 2007), therefore these non-coding variants were not pursued. The detailed functional analysis of this gene, the p.Gly963Cys (c.2887G>T) variant, was not pursued as Rudarakanchana and colleagues (Rudarakanchana et al., 2002) have shown that *BMPR2* construct with truncated cytoplasmic tail or missense mutations in the cytoplasmic tail were able to transduce BMP signal in the functional analysis. Therefore, this variant might not obliterate BMP signaling.

Position	Nucleotide Change	Location	Amino Acid Substitution	Number of Patients	Type of CVMs
Exon 1	c.-301G>A	5'UTR		4	1-2. Ebstein anomaly 3-4. Complex
Exon 2	-	-			-
Exon 3	-	-			-
Exon 4	-	-			-
Exon 5	c.600A>C	exonic	p.Leu200Leu	1	ASD
Intron 5	c.621+30_32delGTT	intronic		1	ASD
Exon 6	-	-			-
Intron 6	c.855-22T>C	intronic		8	1. Coarctation of aorta 2. Ebstein's anomaly 3. ASD 4. Aortic stenosis, 5-8. Complex
Exon 7	-	-			-
Intron 7	c.969-122_124delTCT	intronic		2	1. Coarctation of aorta 2. VSD
Intron 7	c.969-122T>C	intronic		1	1. Ebstein's anomaly
Intron 7	c.969-118_116delTTT	intronic		25	
Exon 8	-	-			-
Exon 9	-	-			-
Exon 10	-	-			-
Exon 11	-	-			-
Exon 12	-	-			-
Exon 13	c.2887G>T	exonic	p.Gly963Cys	1	ASD

**Table 3.8** *BMP2* – Genetic variants

The genetic variants identified in *BMP2* in the panel of 90 probands. (ASD, atrial septal defect; VSD, ventricular septal defect.)



**Figure 3.16** *BMPR2* – Location of the genetic variants

Location of the genetic variations identified in *BMPR2*, in the panel of 90 probands. Boxes represent exons and adjoining lines or broken lines represent introns. Exons are numbered and the sizes are indicated in base pairs. 5'UTR and 3'UTR are in black colour. Location of the kinase-like domain is labelled above the exons.

**BMPR2 p.G963C (c.2887G>T)**

Human	SIQIGESTQD <b>G</b> KSGSGEKIKK
Chimpanzee	SIQIGESTQD <b>G</b> KSGSGEKIKK
Mouse	SIQIGESTQD <b>G</b> KSGSGEKIK <b>R</b>
Chicken	<b>S</b> <b>M</b> <b>Q</b> <b>L</b> <b>G</b> <b>D</b> <b>S</b> <b>S</b> <b>Q</b> <b>D</b> <b>G</b> KSGSGEKIKK
Frog	<b>S</b> <b>T</b> <b>Q</b> <b>C</b> - <b>E</b> <b>T</b> <b>L</b> <b>S</b> <b>D</b> <b>S</b> <b>R</b> <b>S</b> <b>G</b> <b>S</b> <b>G</b> <b>E</b> <b>K</b> <b>I</b> <b>K</b> <b>K</b>
Zebrafish	<b>S</b> <b>L</b> <b>L</b> <b>G</b> <b>E</b> <b>A</b> <b>V</b> <b>S</b> <b>Q</b> <b>E</b> <b>G</b> <b>K</b> <b>A</b> <b>G</b> <b>S</b> <b>A</b> <b>E</b> <b>K</b> <b>I</b> <b>K</b> <b>K</b>
Fruitfly	- <b>V</b> <b>F</b> <b>S</b> <b>G</b> <b>R</b> <b>G</b> <b>S</b> <b>S</b> <b>E</b> <b>R</b> <b>L</b> <b>R</b> <b>D</b> <b>P</b> <b>S</b> <b>E</b> <b>R</b> <b>V</b> <b>K</b> <b>T</b>

**Figure 3.17 BMPR2 – Evolutionary conservation of the variant**

Evolutionary conservation of the non-synonymous unreported variant in BMPR2 p.G963C (c.2887G>T) and the surrounding sequences from human to fruitfly. (Blue: Non conserved region and red: amino acid residues of interest)

### 3.8 Summary

Re-sequencing of five candidate genes in the BMP signaling pathway for rare variants in a panel of 90 patients with a wide range of CVMs successfully identified a number of novel variants. Six novel variants were identified in *BMP2*, nine novel variants in *BMP4* (in a total of 270 patients), ten novel variants in *SMAD6*, one novel variant in *BMPRI1A* and seven novel variants in *BMPRI2*. Further re-sequencing of the *SMAD6* MH2 domain yielded another two novel variants. In total, 35 unreported novel genetic variants were identified. Three novel variants of *BMP4* located in the non-coding region that were absent in the controls and were predicted to affect splicing were investigated further using minigene assays. However, two of the three variants that were successfully carried out showed no significant difference between wild type and variant alleles whereas another variant needs more optimization of the cloning parameters. Three novel variants identified in the MH2 domain of *SMAD6* resulted in non-synonymous substitution warrant further investigation for their functional effect as *in silico* prediction programmes, SIFT and Polyphen, predicted two variants (p.Pro415Leu [c.1244C>T] and p.C484F [c.1451G>T]) to be damaging and one variant (p.Ala325Thr [c.973G>A]) to be benign.

## Chapter 4      *In vitro* functional analysis of novel SMAD6 variants

### 4.1 Overview

*In vitro* functional studies have the potential to give us an insight on how mutant proteins function in the cellular environment. This chapter describes the *in vitro* functional analyses of the three non-synonymous SMAD6 variants reported in chapter 3. The aims of these experiments were:

- To assess the inhibitory effect of the mutant proteins compared to that of the wild type protein on BMP- responsive reporter transcriptional activity.
- To assess the inhibitory effect of the mutant proteins compared to that of the wild type protein on osteoblastic differentiation, assessed by alkaline phosphatase production in response to BMP signaling.

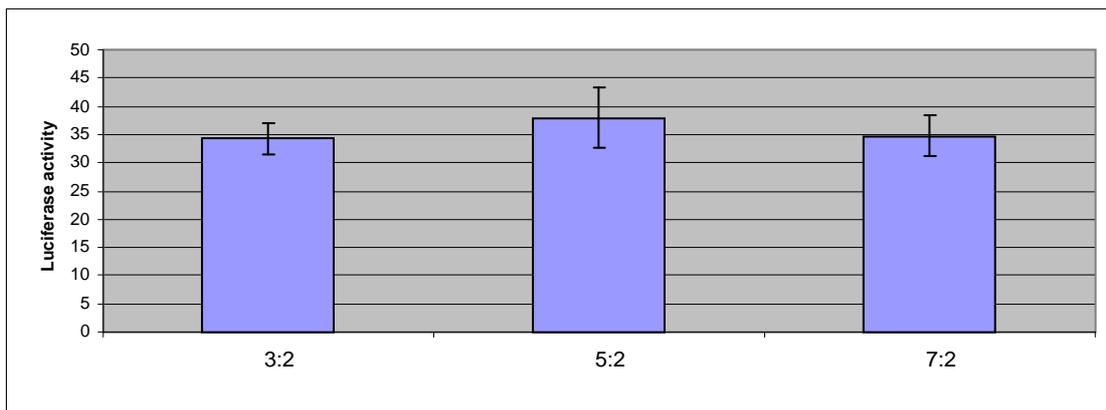
The C2C12 cells used in these experiments are mouse myoblast cells established by Yaffe and Saxel (Yaffe and Saxel 1977). These cells are capable of differentiation, form contractile myotubes and produce characteristic muscle proteins. This cell line has been used for many BMP signaling pathway *in vitro* functional studies (Fujii et al. 1999; Goto et al. 2007; Korchynskyi and ten Dijke 2001).

Two assays were selected to assess function of the mutant SMAD proteins. The first assay investigated their ability to inhibit transcription of a reporter under the control of a BMP-responsive element (BRE) and the second assay was to determine their ability to inhibit osteoblastic differentiation of C2C12 cells.

## 4.2 Optimization of transfection

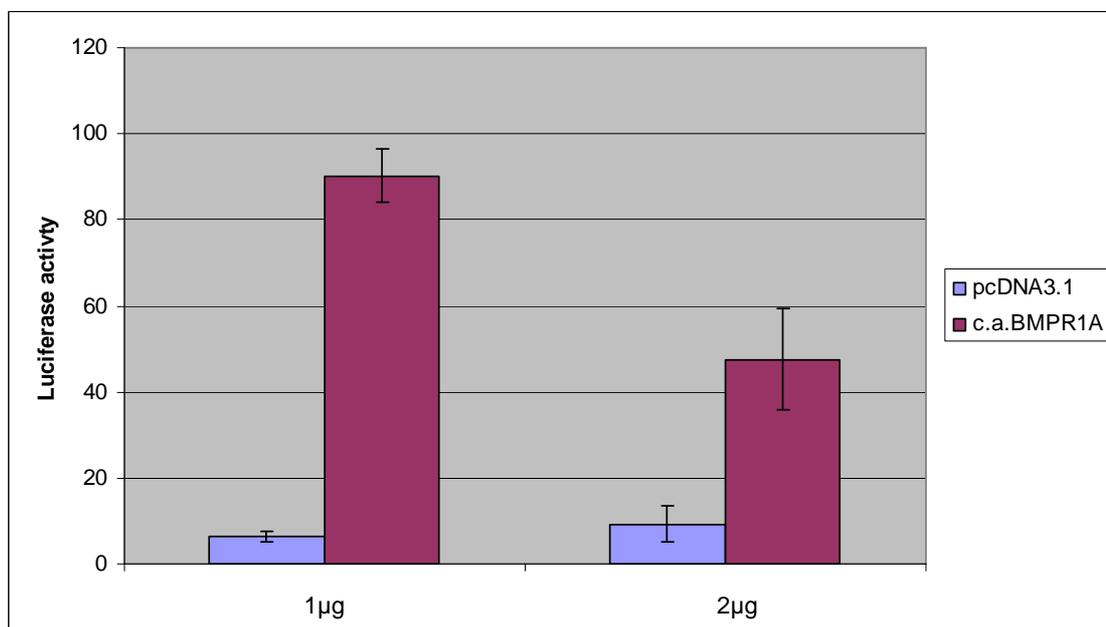
Initially, C2C12 cells were transiently transfected with 2 $\mu$ g of constructs (900ng of BRE-luciferase, 1 $\mu$ g of constitutively active BMPR1A and 100ng of Renilla luciferase construct) in different volumes of transfection reagent: 3 $\mu$ l, 5 $\mu$ l, and 7 $\mu$ l, to evaluate the effect of different transfection reagent to DNA ratios. The luciferase reading showed that there was no positive effect with higher volume of transfection reagent (Figure 4.1). Therefore, the lowest volume of transfection reagent (3 $\mu$ l) was used for subsequent experiments.

C2C12 cells were then transfected with 1 $\mu$ g (450ng of BRE-luciferase, 500ng of constitutively active BMPR1A [or empty vector] and 50ng of Renilla luciferase construct) or 2 $\mu$ g of constructs (900ng of BRE-luciferase, 1 $\mu$ g of constitutively active BMPR1A [or empty vector] and 100ng of Renilla luciferase construct) to evaluate the effect of different DNA concentrations on luciferase readout. Experiments were carried out in triplicate. Firefly luciferase reading was normalised to Renilla luciferase reading for each measurement. The mean reading for each combination of constructs are shown in Figure 4.2. The luciferase reading for c.a.BMPR1A compared to the empty control clearly shows that the constitutively active form of BMPR1A induced BMP signaling, in the transiently transfected cells. Optimal expression was achieved using the ratio of 3 $\mu$ l of transfection reagent to 1 $\mu$ g of constructs. Not only was the reading for c.a.BMPR1A higher but also the difference between empty vector (pcDNA3.1) and constitutively active BMPR1A is larger. This is crucial to be able to quantify the effect of the SMAD6C variants in the following experiments.



**Figure 4.1 Optimization of the ratio of transfection reagent to construct DNA**

The effect of 3 $\mu$ l, 5  $\mu$ l and 7 $\mu$ l of transfection reagent on the luciferase readout induced by constitutively active BMPR1A transfected with 2 $\mu$ g of construct DNA. The lowest volume of transfection reagent 3 $\mu$ l was used for subsequent experiments as there was no positive effect with higher volume of transfection reagent.



**Figure 4.2 Optimization of the amount of construct DNA**

The effect of different ratio of construct DNA (1 $\mu$ g of DNA to 2 $\mu$ g of DNA) with 3 $\mu$ l of transfection reagent on the luciferase reading induced by constitutively active BMPR1A (pcDNA3.1, an empty vector, acted as a negative control). Optimal expression was observed with the ratio of 3 $\mu$ l of transfection reagent to 1 $\mu$ g of DNA construct.

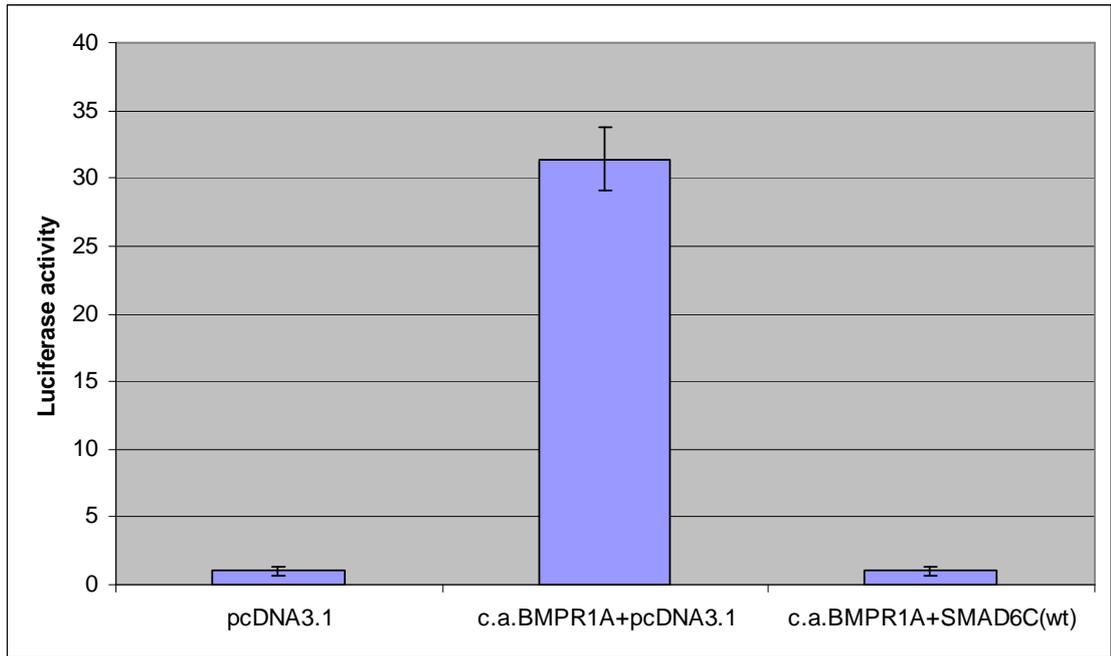
### **4.3 Inhibitory effects of SMAD6 variants on BMP-Responsive element reporter**

To confirm that wild type SMAD6C protein inhibits the transactivation of BRE-Luc induced by constitutively active BMPR1A, as previously described (Goto et al. 2007), luciferase assay of C2C12 cells transfected with BRE-Luc and renilla with either pcDNA3.1 or c.a.BMPR1A and pcDNA3.1 or c.a.BMPR1A and SMAD6C wild type constructs were performed. Reporter gene activity for each well was measured for both the firefly luciferase and the renilla activity. The mean relative luciferase activity for this experiment is displayed in Figure 4.3. Results of the relative luciferase activity suggest that the SMAD6C protein inhibits c.a.BMPR1A by approximately 30 fold.

The inhibitory activity of the mutant SMAD6C proteins compared to the wild type was assessed. Three independent experiments were performed in triplicate for each combination of constructs. Figure 4.4 shows the overall mean reading of all three replicates.

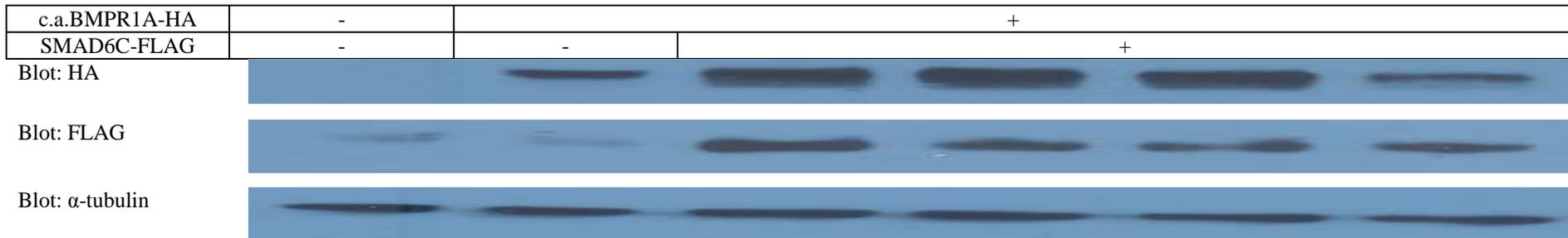
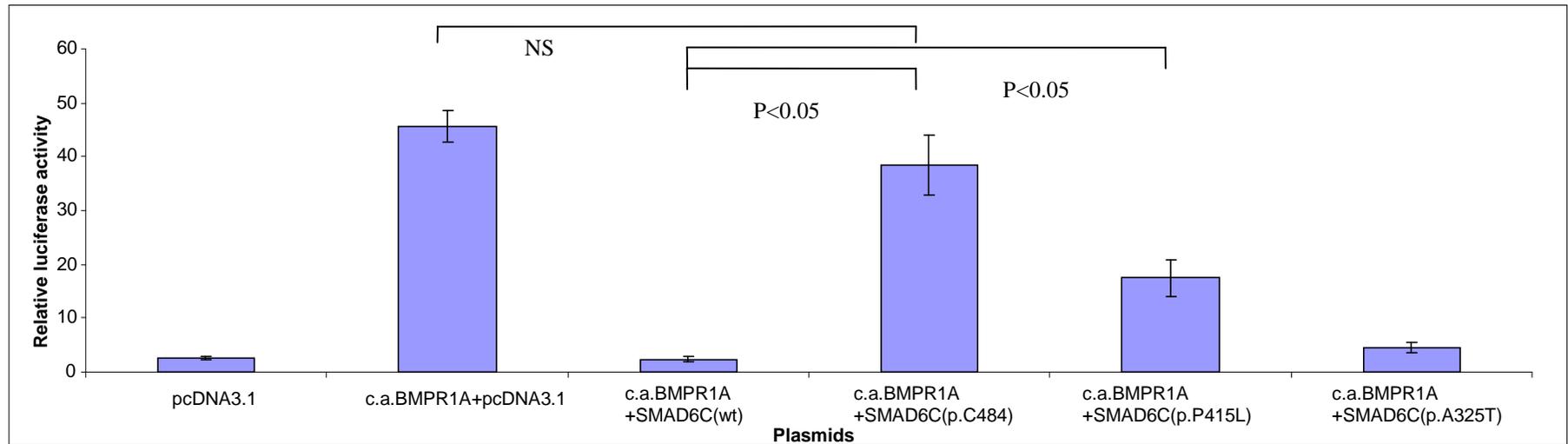
Mean relative luciferase activities for all nine readings for each combination of constructs were log transformed before one way ANOVA was performed. The mean relative luciferase activity of the p.Ala325Thr (c.973G>A) mutant was not significantly different from wild-type (p value > 0.05), indicating that this variant does not affect the inhibitory activity of SMAD6. However, both p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) mutants failed to inhibit BMP signaling compared to the SMAD6 wild type, with 6 fold increase in relative luciferase activity for p.Pro415Leu (c.1244C>T) mutant and 17 fold increase for p.Cys484Phe (c.1451G>T) mutant. The ANOVA results for both p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) mutants showed significant difference with a p value of less than 0.05 compared to wild type SMAD6. Comparing the relative luciferase activity of c.a.BMPR1A to the mutants, the p.Pro415Leu (c.1244C>T) mutant appeared to be hypomorphic, being significantly different from c.a.BMPR1A (p value < 0.05) whereas p.Cys484Phe (c.1451G>T) mutant appeared to be a null allele, not significantly different from the uninhibited c.a.BMPR1A (Figure 4.4).

Having shown the induction by c.a.BMPR1A and the inhibition by SMAD6 wild type in the luciferase assay, a western blot was performed using the same cell lysates that were



**Figure 4.3** Luciferase activity in C2C12 cells transfected with c.a.BMPR1A and wild type SMAD6C

The effect of wild type SMAD6C on the luciferase reading induced by constitutively active BMPR1A (pcDNA3.1, an empty vector, acted as a negative control). The SMAD6C protein inhibits the luciferase activity generated by c.a.BMPR1A by approximately 30 fold.



**Figure 4.4 Overall SMAD6C variants inhibitory effect of three replicates and immunoblot images**

The mean relative luciferase activity of three separate transfection experiments performed in triplicate for each combination of constructs (pcDNA3.1, an empty vector, acted as a negative control). One-way ANOVA was used to determine statistical significance of the readings between the SMAD6C(wt) and mutants. SMAD6C p.Cys484Phe (c.1245G>T) and p.Pro415Leu (c.1244C>T) were significantly different from the wild type but not the p.Ala325Thr (c.973G>A) mutant. The three panels below show levels of protein expression of c.a.BMPR1A (HA) (top blot), SMAD6C constructs (FLAG) (middle blot) and α-tubulin (loading control) (bottom blot) in the functional protein complexes from luciferase reading.

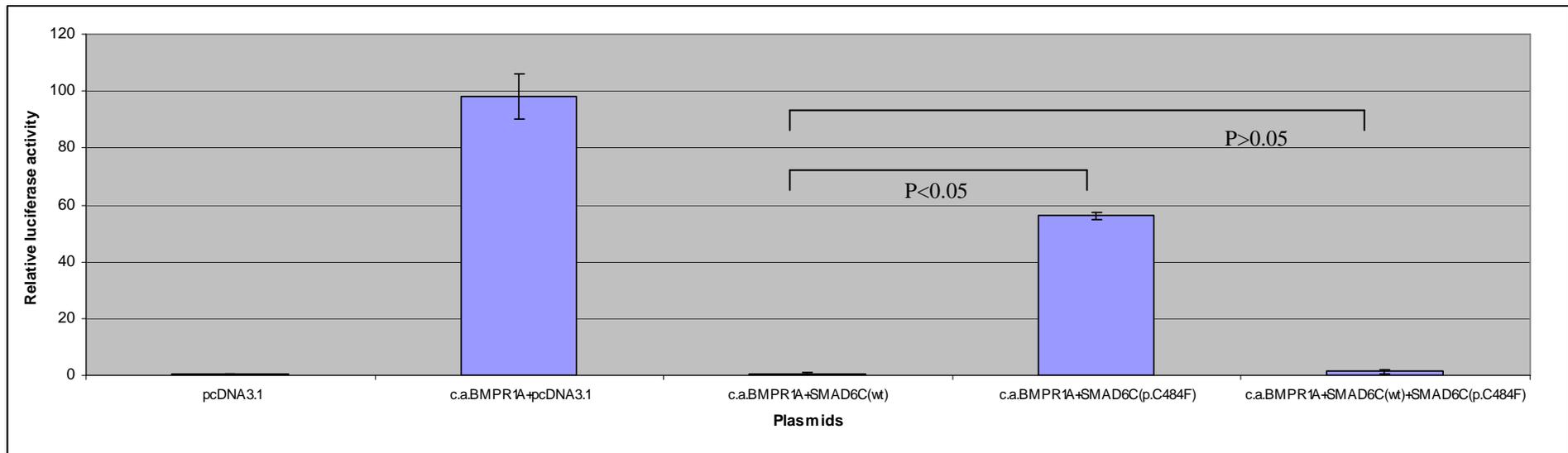
used for dual luciferase reading to determine that differences in readout could not be due to differences in levels, either of c.a.BMPR1A or SMAD6 proteins.

The molecular weight of the c.a.BMPR1A protein with a copy of the HA tag is approximately 62 kilodaltons (kDa). The molecular weight of the SMAD6C protein with a copy of the FLAG tag is approximately 29kDa. These tags do not modify the expression of the proteins to which they are appended (Brizzard et al. 1994; Kolodziej and Young 1991). Antibodies to the tags were used to detect c.a.BMPR1A and SMAD6 respectively. Endogenous tubulin was used as a loading control. The molecular weight of  $\alpha$ -tubulin is approximately 55kDa.

The c.a.BMPR1A bands (probed with anti-HA) for the cell lysates transfected with p.Cys484Phe (c.1451G>T) and p.Pro415Leu (c.1244C>T) and wild type SMAD6 constructs were equally strong. The BMPR1A band for the p.Ala325Thr (c.973G>A) construct was weaker. It could be argued that lower luciferase activity seen in these cells is actually a consequence of lower levels of c.a.BMPR1A being expressed. However, the intensity of the band is similar to the one from the group of cells that were transfected with c.a.BMPR1A and empty vector alone, which showed maximum luciferase activity indicating that the respective amount of c.a.BMPR1A being expressed was sufficient to induce maximum BMP signaling.

The intensity of the bands for SMAD6 proteins (probed with anti-FLAG) were of similar intensity. The faint bands that were seen in the first two lanes (cells that were transfected with empty vector only, and cells that were transfected with c.a.BMPR1A and empty vector) are non-specific bands, as there was no FLAG-tagged protein expressed in those two groups of cells. The blots demonstrated that the difference in the luciferase readout observed for the SMAD6 variants were not due to SMAD6 protein not being expressed (or being expressed in different amounts) but they truly represent differences in the inhibitory activities of such proteins (Figure 4.4). However, as the bands for the SMAD6 mutant proteins were slightly weaker than the wild type, in addition to the fact that the presence of the very weak non-specific bands in the cells not transfected with SMAD6, it would be better to quantify the intensity of the western bands in order to conclusively interpret the protein expression of SMAD6 mutants and wild type.

The mean relative luciferase activity of the cells transfected with the SMAD6C p.Cys484Phe (c.1451G>T) variant and the wild type SMAD6C construct in the presence of c.a.BMPR1A was not significantly different from the cells that were transfected with wild type only (p value > 0.05). Thus, when coexpressed with wild type SMAD6C, the SMAD6C p.Cys484Phe (c.1451G>T) variant did not serve as a dominant-negative mutant (Figure 4.5).



**Figure 4.5 Effect of a 50:50 mixture of SMAD6C p.Cys484Phe (c.1451G>T) variant with SMAD6C wild type constructs**

The SMAD6C p.Cys484Phe (c.1451G>T) variant did not show dominant-negative effect when coexpressed with wild type SMAD6C, since a 50:50 mix of variant and wild-type constructs was able to successfully inhibit BMP signaling (last column). (pcDNA3.1, an empty vector, acted as a negative control whereas c.a.BMPR1A activated BMP signaling)

The investigation on whether the diminution of the inhibitory action of SMAD6 affected the phosphorylation of the BMP-specific SMADs (SMAD1, SMAD5 and SMAD8) was performed using immunoblotting, in a manner analogous to that used by Goto and colleagues (Goto et al. 2007). The intensity of protein bands on western blots probed with antibody specific for phospho-SMAD1/5/8 and for unphosphorylated SMAD1 were compared between lysates for the six combinations of transfected plasmids. The molecular weight of the phospho-SMAD1/5/8 is 60kDa whereas the molecular weight of SMAD1 is approximately 56kDa. Results are presented in Figure 4.6. Under the culture conditions that were used, which involved 10% FBS, there was a high background level of SMAD phosphorylation, although unphosphorylated SMAD1 was also strongly detected. No differences in the ratio between phosphorylated and unphosphorylated SMADs between the different combinations of constructs were observed. This was probably due to a combination of the high background level of SMAD phosphorylation, and the inefficiency of transfection. Experiments described below suggested that only about 10% of cells were successfully transfected, and therefore any effects of SMAD6 construct transfection might have been obscured. Alternative explanations might have been either that the phospho-SMAD antibody was also detecting unphosphorylated SMAD proteins, or that the FBS used contained ligands which were influencing SMAD phosphorylation. It is of interest that the experiments of Goto et al. were described as having been conducted using 20% FBS; at that concentration, very little phosphorylation of SMADs was detected in the absence of BMP ligand stimulation. By contrast, however, unpublished data from our collaborator Dr. Helen Arthur's group produced during the period of this work suggested that at lower concentrations of FBS (0.2%-1%) phosphorylation of SMADs in the absence of ligand stimulation was observed. Therefore the possibility of nonspecific binding by the phospho-SMAD antibody and the effect of different concentrations of FBS on the SMAD/ phospho-SMAD ratio were investigated. Non-transfected cell lysates were cultured in medium containing 5% or 10% of FBS and treated with 1 $\mu$ l of ExoSAP-IT to remove the phosphate. These were probed with phospho-Smad1/5/8 and Smad1 antibodies respectively. The blots showed that phospho-Smad1/5/8 is indeed specific antibody as it does not bind to the unphosphorylated cell lysates (lysates that were treated with ExoSAP for 30 minutes and 1 hour) (Figure 4.7).

Combination of plasmids (BRE-Luc + Renilla)	pcDNA3.1	c.a.BMPR1A + pcDNA3.1	c.a.BMPR1A + SMAD6C (wt)	c.a.BMPR1A + SMAD6C (p.Cys484 Phe) (c.1451G>T)	c.a.BMPR1A + SMAD6C (p.Pro415Leu) (c.1244C>T)	c.a.BMPR1A + SMAD6C (p.Ala325Thr) (c.973G>A)
---	----------	-----------------------	--------------------------	--	---	--

Blot: phospho-Smad1/5/8



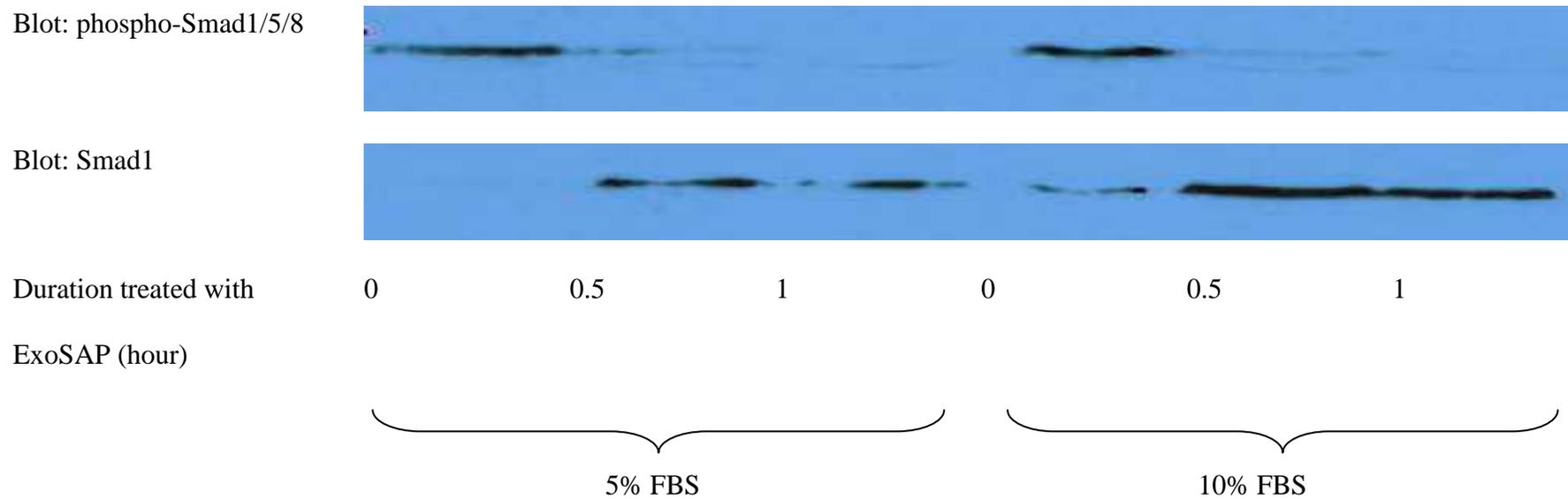
Blot: Smad1



**Figure 4.6 Immunoblot images of phospho-Smad1/5/8 and Smad1**

The panels above show protein expressions of phospho-Smad1/5/8 and Smad1 in the functional protein complexes from luciferase reading (in collaboration with Dr. Elise Glen). No differences in the ratio between phosphorylated and unphosphorylated SMADs between the different combinations of constructs were observed.

(pcDNA3.1, an empty vector, acted as a negative control; c.a.BMPR1A, activated BMP signaling; SMAD6C(wt), inhibited BMP signaling)



**Figure 4.7 Immunoblot images of phospho-Smad1/5/8 and Smad1 treated with ExoSAP and cultured in different concentration of FBS**

The panels above show protein expressions of phospho-Smad1/5/8 and Smad1 in the treated with 1 $\mu$ l of ExoSAP for 0 hour, 30 minutes and 1 hour in cell lysates that were cultured in medium with 5% or 10% of FBS (in collaboration with Dr. Elise Glen). The blots showed that phospho-Smad1/5/8 is indeed specific antibody as it does not bind to the unphosphorylated cell lysates (lysates that were treated with ExoSAP for 30 minutes and 1 hour).

At 10% FBS, as in Figure 4.7, both non-phosphorylated and phosphorylated SMADs are detectable. Somewhat surprisingly, at 5% FBS, there is more phosphorylated (i.e. active) SMAD signal detectable than non-phosphorylated SMAD. This suggests that possibly FBS contains both ligands and inhibitors of this signaling pathway, since based on the data generated and the results from Goto et al., the relationship between FBS concentration and SMAD activation does not seem to be monotonic.

Based on these experiments, it was concluded that further optimisation of culture conditions would be required to demonstrate differential effects of wild type and mutant SMAD6 on SMAD1/5/8/ phosphorylation. From the experiments performed, the possible mechanism whereby SMAD6 inhibits BMP signalling affected by the novel variants discovered was not conclusive.

## 4.4 Inhibitory effects of SMAD6 variants on alkaline phosphatase activity

### 4.4.1 Inhibitory effects of SMAD6 variants with BMP ligands

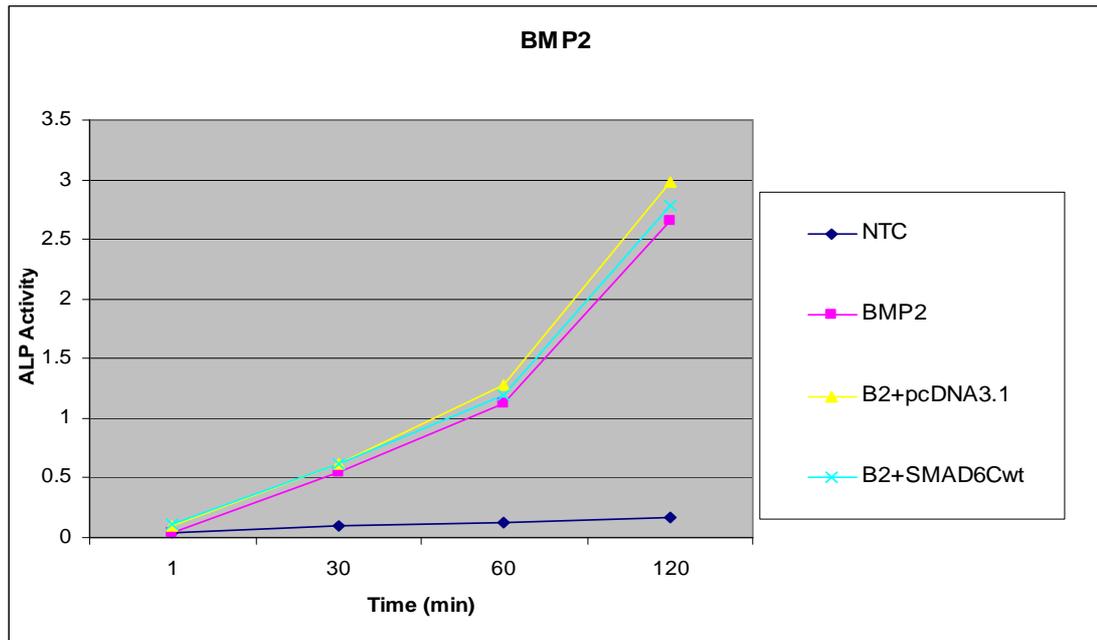
The inhibitory activities of the mutant SMAD6C proteins (p.Pro415Leu [c.1244C>T] and p.Cys484Phe [c.1451 G>T]) on alkaline phosphatase activity induced by 300 ng/ml of BMP2 or BMP6 were compared to that of the wild type SMAD6C, as previously described by Katagiri and colleagues (Katagiri et al. 1994). The p.Ala325Thr (c.973G>A) mutant was not tested in this assay as this mutant was not significantly different from the wild type protein in the luciferase assay. The ALP activity of the cell lysates was measured using *p*-nitrophenol phosphate as a substrate.

Various variables were assessed using empty vector pcDNA3.1 and wt SMAD6C before undertaking the experiment with the mutant constructs. Quantitative analysis of ALP activity yielded different results for cells that were stimulated with BMP2 or BMP6 for 60 hours or 48 hours. Generally, cells that were stimulated with BMP6 produced ALP activity around twice as high than BMP2 (Figure 4.8, upper and lower panels), which is in accordance with the results demonstrated by Goto and colleagues that BMP6 induced more sustained signaling (Goto et al. 2007). The non-BMP treated cells (NTC) yielded very little alkaline phosphatase activity for all sets of cells.

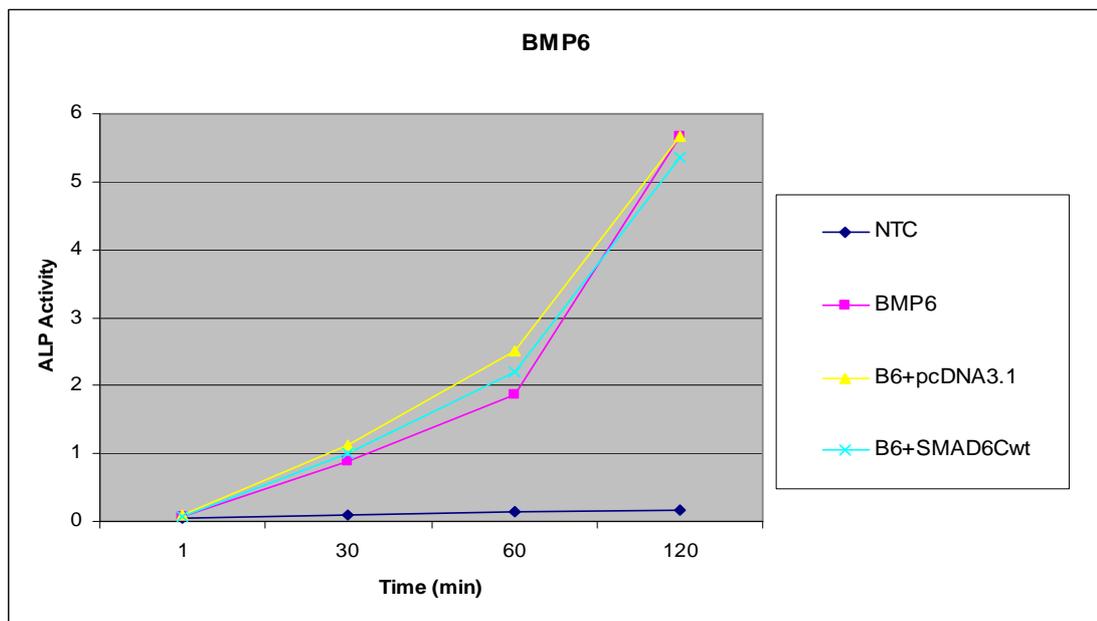
The ALP activity of the cells that were cultured for 48 hours after stimulation is shown in Figure 4.8. For both BMP2 and BMP6 stimulation, there was no apparent difference between untransfected controls and cells transfected either with empty vector or wild type SMAD6.

Therefore, cells were cultured for 60 hours after stimulation with the hope that the difference between the wild-type and the empty vector could be discerned (Figure 4.9). In this experiment, the expected inhibition of BMP-stimulated ALP expression with the SMAD6Cwt construct was not observed – indeed, the ALP expression when SMAD6Cwt was transfected appeared if anything higher than in the BMP treated but untransfected cells. Surprisingly, ALP expression when empty vector was transfected appeared marginally higher than in the BMP treated but untransfected cells (ANOVA *p*-value = 0.04). Since only an independent experiment was performed in duplicate for

A)



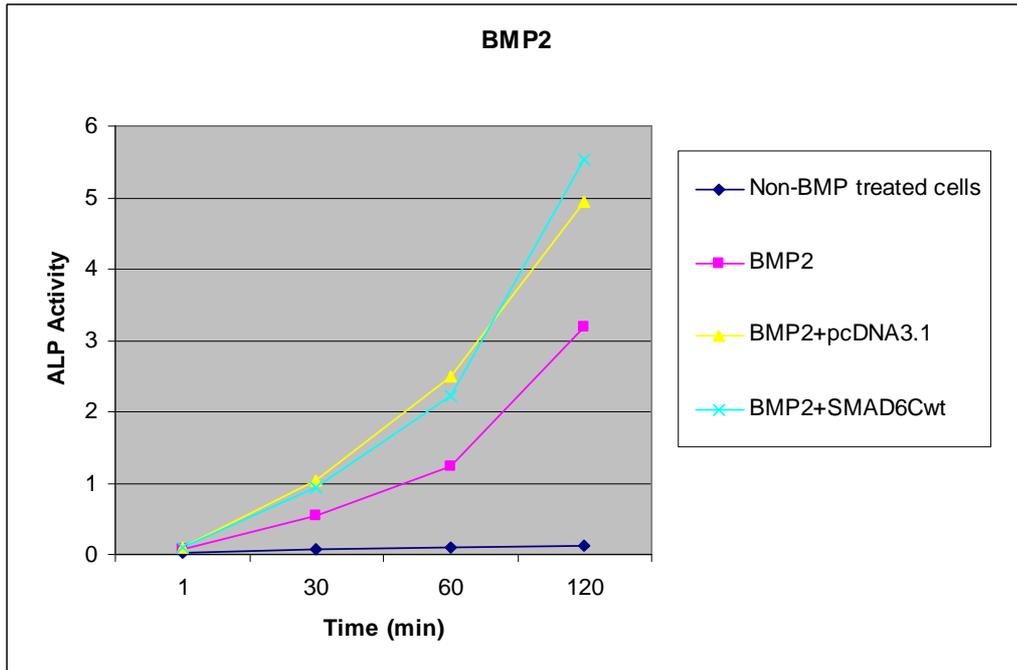
B)



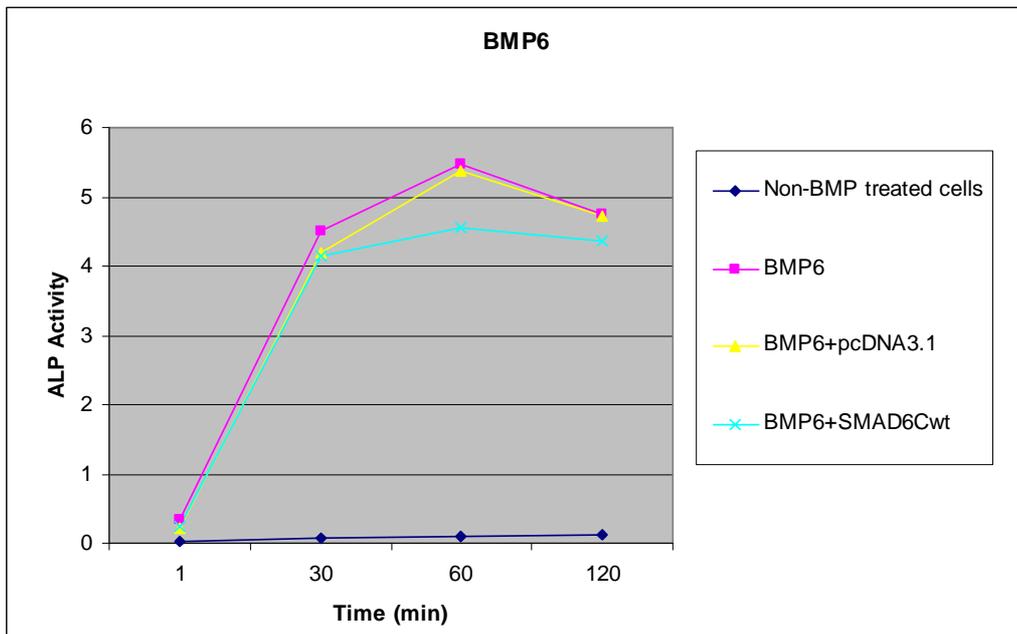
**Figure 4.8 ALP activity of the cell lysates that were stimulated with ligands and cultured for 48 hours**

The mean ALP activity of an independent experiment with duplicated cell lysates that were cultured for 48 hours. C2C12 cells were transfected with empty vector pcDNA3.1 and wtSMAD6C, then stimulated with 300ng/ml of BMP-2 or BMP-6 24 hours after transfection. A) ALP activity of the cells stimulated with BMP-2. B) ALP activity of the cells stimulated with BMP-6.

A)



B)



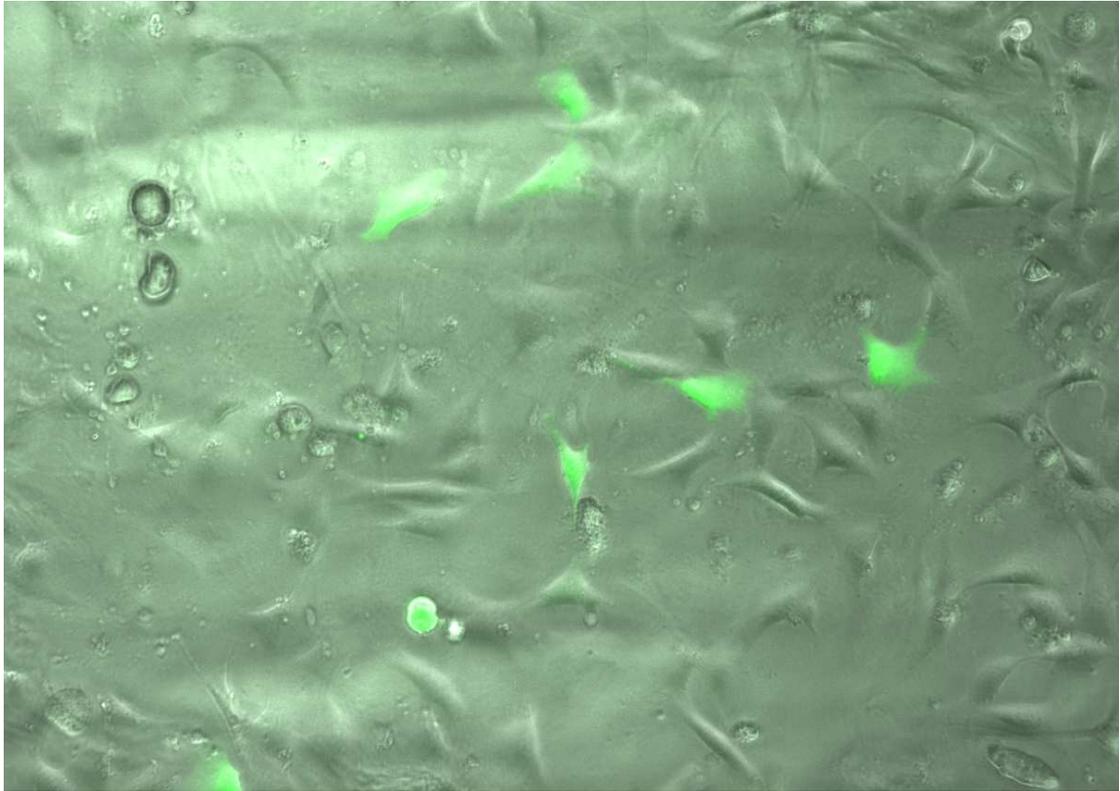
**Figure 4.9** ALP activity of the cell lysates that were stimulated with ligands and were cultured for 60 hours

The mean ALP activity of an independent experiment with duplicated cell lysates that were cultured for 60 hours. C2C12 cells were transfected with empty vector pcDNA3.1 and wtSMAD6C then stimulated with 300ng/ml of BMP-2 or BMP-6 6 hours after transfection. A) ALP activity of the cells stimulated with BMP-2. B) ALP activity of the cells stimulated with BMP-6.

each combination of constructs that were harvested at two different timepoint, this apparent difference could be due to a chance lower response in the BMP transfected cells; whatever the reason, it was clear that in this system SMAD6wt transfection was not suppressing ALP expression.

It was hypothesised that the transfection efficiency could be responsible for the failure of wild type SMAD6 to inhibit ALP in the system. To assess transfection efficiency, C2C12 cells were transfected with an expression plasmid encoding a visible reporter gene, a GFP construct that encodes green fluorescent protein, using FuGENE® HD Transfection Reagent as described in chapter 2. Images of the fluorescent cells were identified using a fluorescent microscope 24 hours post-transfection. Transfection efficiency was determined by comparing the number of cells expressing the reporter protein to the total number of cells in the population. Transfection efficiency was approximately 10% (Figure 4.10).

Given that the untransfected cells (90% of the cells present) also respond to BMP2/6, it was concluded that this was likely to mask any differences between the wild type SMAD6C construct and empty vector in the remaining 10% of cells. It was essential to have a system where a response would only be measured in cells transfected with wild type or mutant SMAD6. To limit BMP stimulation to transfected cells, the constitutively active BMPR1A construct was utilised.



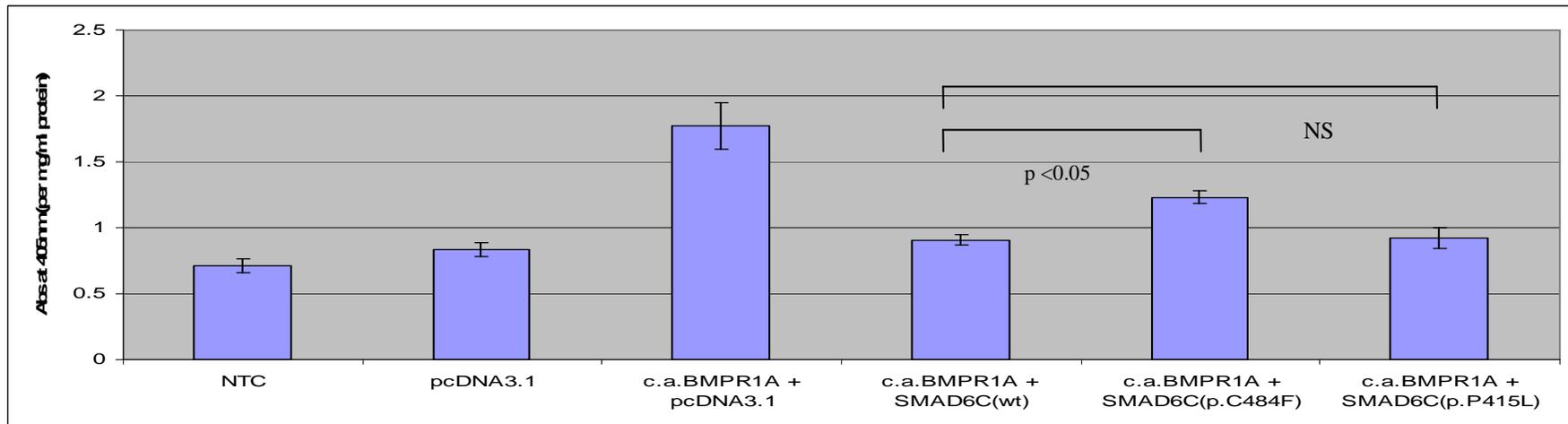
**Figure 4.10** Fluorescent microscope image of GFP transfected cells

Microscope image of transient transfection with GFP plasmid DNA produced fluorescent protein that was observed in C2C12 cells (magnification 10X). Transfection efficiency observed was approximately 10%.

#### **4.4.2 Assessment of inhibitory effects of SMAD6 variants on alkaline phosphatase activity with c.a.BMPR1A**

To overcome the limitations of transfection efficiency, the inhibitory activity of the mutant SMAD6C proteins on the inhibition of osteoblastic differentiation (ALP activity) compared to the wild type was assessed by transfecting c.a.BMPR1A and either empty vector, wild type SMAD6, p.Pro415Leu (c.1244C>T) or p.Cys484Phe (c.1451G>T). Three independent experiments were performed in duplicate for each combination of constructs. The mean reading of the ALP activities taken at the 13 hour time point is shown in Figure 4.11. Two-way ANOVA was used to determine the statistical significance of the differences between readings from the wild type SMAD6 and both mutants (p.Pro415Leu [c.1244C>T] and p.Cys484Phe [c.1451G>T]). SMAD6 p.Cys484Phe (c.1451G>T) showed 1.3 fold higher ALP activity compared to wild type with a p value of < 0.05. The ANOVA suggested a non-significant difference in the p.Pro415Leu (c.1244C>T) mutant compared to the wild type SMAD6 (p value > 0.05). This might be due to the size of the effect of p.Pro415Leu (c.1244C>T) mutant being too small to be detected with the number of replicates that were performed, or the effect of this mutant may not be optimally captured with the timepoint that was used. However, SMAD6 p.Cys484Phe (c.1451G>T) clearly failed to inhibit osteoblastic differentiation when compared to wild type SMAD6. This is in line with the results of the luciferase assay.

Similar protein expression levels of HA-tagged c.a.BMPR1A and FLAG-tagged SMAD6C wild type or mutants were observed when immunoblotting was carried out, as shown in Figure 4.11. The same intensity was observed in the loading control, nucleolin, showing that the concentration of the proteins for each combination of constructs was even. The blots demonstrated that the difference in the ALP activity observed for the SMAD6 variants were not due to the SMAD6 or c.a.BMPR1A protein being differentially expressed but rather due to differences in the inhibitory activities of the wild type and mutant SMAD6.



Blot: HA

Blot: FLAG

Blot: Nucleolin



**Figure 4.11 ALP activities in the cell lysates that were transfected with c.a.BMPR1A**

Alkaline phosphatase activity of c.a.BMPR1A transfected C2C12 cells with or without SMAD6C variants was measured at a single timepoint as absorbance of *p*-nitrophenyl at 405nm per mg/ml protein using *p*-nitrophenylphosphate as a substrate (in collaboration with Dr. Elise Glen). The results represent the average of three separate transfection experiments performed in duplicate. 2-way ANOVA was used to determine statistical significance of the readings at a single timepoint (+13 hours after start of assay) between the SMAD6C(wt) and both mutants. SMAD6C p.Cys484Phe (c.1451G>T) was found to show significantly greater ALP activity than SMAD6C(wt) after 13 hours ( $p < 0.05$ ) however SMAD6C p.Pro415Leu (c.1244C>T) was not found to show significantly different ALP activity to the SMAD6C(wt) at the 13 hour timepoint. The blots below show the protein expression of c.a.BMPR1A (HA), SMAD6 wild type or mutant constructs (FLAG) and nucleolin (loading control), demonstrating that the difference in the ALP activity observed for the SMAD6 variants were not due to the SMAD6 or c.a.BMPR1A protein being differentially expressed but rather due to differences in the inhibitory activities of the wild type and mutant SMAD6.

## 4.5 Summary

Using a site-directed mutagenesis technique, three non-synonymous SMAD6 variants were introduced into SMAD6C expression constructs. Dual luciferase assays showed that the SMAD6 p.Cys484Phe (c.1451G>T) variant resulted in abolition of the inhibitory effect of SMAD6, the p.Pro415Leu (c.1244C>T) variant reduced the inhibitory activity by half whereas the p.Ala325Thr (c.973G>A) variant did not differ from the wild type. Western blotting demonstrated that equal amounts of protein were present in each case. Therefore, the loss or partial loss of the inhibitory effect was not due to the absence of mutant proteins. Investigation of the p.Cys484Phe (c.1451G>T) and p.Pro415Leu (c.1244C>T) variants using an alkaline phosphatase assay showed that p.Cys484Phe (c.1451G>T) had diminished ability to inhibit osteoblastic differentiation, resulting in higher alkaline phosphatase activity than the wild type protein and providing a potential mechanistic explanation for the aortic calcification phenotype seen in the patient carrying the p.Cys484Phe (c.1451G>T) variant.

## Chapter 5 Discussion

### 5.1 Overview

This is the first study of the relationship between rare genetic variations in multiple key genes of the BMP signaling pathway and CVM. The BMP signaling pathway plays an important role during development and adult life. The five candidate genes, *BMP2*, *BMP4*, *BMPRIA*, *BMPR2* and *SMAD6* were selected based on previous reports that have shown them to be essential for normal heart development. It was hypothesized that rare functional variation in members of this pathway could predispose individuals to increased risk of CVMs. In order to investigate the presence of rare genetic variants, the exons of the candidate genes were sequenced in a panel of 90 CVM patients with a range of congenital heart disease phenotypes. The inheritance of the variants was determined when parents of the patient were available. The unreported genetic variants were genotyped in a collection of 1425 individuals from 255 nuclear families, contributing 1000 control chromosomes. A total of 35 unreported genetic variants were identified; of these, 27 variants were genotyped in the controls. All these variants were inherited either from mother or father. *De novo* variants were not identified in this study.

Among the previously unreported genetic variants identified in the candidate genes, the genetic variants seen in the 5'UTR or in the introns of *BMP4* were investigated further as they were absent in the controls and the location of the variants suggested they might affect splicing. The genetic variants were analyzed with two different programmes to identify any putative effects on pre-mRNA splicing and were investigated by minigene assay. The minigene assays for two of the three novel rare variants in *BMP4* showed that these two variants did not result in aberrant splicing, whereas the third assay was inconclusive.

A total of seven non-synonymous variants were identified in this study: one in *BMP4*, one in *BMPR2* and five in *SMAD6*. Two variants of *SMAD6* (p.Asp21Asn [c.61G>A] and p.Ser325Arg [c.975C>G]) were not evolutionarily conserved. Furthermore, *SMAD6* p.Asp21Asn (c.61G>A) was present in the controls. The potential functional effect of the remaining non-synonymous variants were analyzed with the SIFT and PolyPhen programmes. *BMPR2* p.Gly963Cys (c.2887G>T) and *SMAD6* p.Cys484Phe

(c.1451G>T) variants were predicted to affect protein function by both algorithms. BMPR2 p.Gly963Cys (c.2887G>T) was not further investigated as a study carried out by Rudarakanchana and colleagues showed that cytoplasmic tail of BMPR2 (the location of the variant) is not vital in transducing BMP signals (Rudarakanchana et al. 2002). Further re-sequencing of the MH2 domain of *SMAD6* in an additional 348 probands with a broad range of CVM phenotypes yielded two further non-synonymous variants. One of the variants, *SMAD6* p.Pro415Leu (c.1244C>T), was also predicted to affect protein function. Both non-synonymous variants of *SMAD6*, p.Cys484Phe (c.1451G>T) and p.Pro415Leu (c.1244C>T), were absent in the control population.

*In vitro* functional analyses of the three non-synonymous *SMAD6* variants that were located in the MH2 domain were carried out to assess the inhibitory effect of the mutant proteins compared to that of the wild type protein on the BMP-signaling pathway. Luciferase assay showed that p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) variants were significantly different from the wild type in the inhibition of BMP signaling ( $p < 0.05$ ). Based upon the known functions of the BMP signaling pathway and the phenotypes of the patients carrying these variants, these two variants were further investigated using an alkaline phosphatase assay to investigate osteogenic transformation during development. This assay revealed that the p.Cys484Phe (c.1451G>T) variant failed to inhibit alkaline phosphatase activity compared to the wild type ( $p < 0.05$ ). The inefficiency of the p.Cys484Phe (c.1451G>T) variant in inhibiting osteogenic differentiation is in accordance with the phenotype of the patient that exhibited calcification (and coarctation) of the aorta. The principal novel finding of this study is that inhibitory failure of BMP signaling due to genetic variation in *SMAD6* can be one of the factors that contribute to CVM in man. Table 5.1 summarizes the findings of this study in the five candidate genes sequenced.

Characteristic of the previously unreported variants	Number of variants	Variants
Absent in controls, coding region, predicted deleterious	2	<i>SMAD6</i> p.Pro415Leu (c.1244C>T); <i>SMAD6</i> p.Cys484Phe (c.1451G>T)
Absent in controls, coding region, predicted non deleterious	5	<i>SMAD6</i> p.Ser325Arg (c.975C>G) <sup>#</sup> ; <i>SMAD6</i> p.Ala325Thr (c.973G>A) <sup>∞</sup> <i>BMP4</i> p.Gly174Ser (c.520G>A) <i>BMP2</i> c.891G>A (p.Lys297Lys); <i>BMP2</i> c.1146A>G (p.Leu283Leu)
Absent in controls, non-coding region, putative regulation	8	<i>SMAD6</i> c.-428A>T; <i>SMAD6</i> c.-290G>C; <i>SMAD6</i> c.*319T>A <i>BMP4</i> c.-283G>C; <i>BMP4</i> c.-295G>A; <i>BMP4</i> c.-136C>G; <i>BMP4</i> c.-87C>T <i>BMP2</i> c.-906C>T
Present in controls, significant MAF difference in 180 cases/ 1000 control chromosomes	3	<i>SMAD6</i> c.120C>T <i>BMP4</i> c.-171C>G <i>BMP2</i> c.-342_-340delTGT
Present in controls, no MAF difference in 180 cases/ 1000 control chromosomes	9	<i>SMAD6</i> c.-44C>T; <i>SMAD6</i> c.61G>A; <i>SMAD6</i> c.711C>T; <i>SMAD6</i> c.*262T>C <i>BMP4</i> c.76T>C; <i>BMP4</i> c.345C>T; <i>BMP4</i> c.370+28G>A <i>BMP2</i> c.-673G>A; <i>BMP2</i> c.*936_939delATTT
Coding region, predicted deleterious	1	<i>BMP2</i> p.Gly963Cys (c.2887G>T)
Non-coding region	7	<i>BMP1A</i> c.534-59T>C <i>BMP2</i> c.-301G>A; <i>BMP2</i> c.621+30_32delGTT; <i>BMP2</i> c.855-22T>C; <i>BMP2</i> c.969-122_124delTCT; <i>BMP2</i> c.969-122T>C; <i>BMP2</i> c.969-118_116delTTT

<sup>#</sup> Alternatively spliced exon 4B of *SMAD6*, SIFT predicted deleterious but PolyPhen predicted non-deleterious

<sup>∞</sup> Requires another attempt of genotyping in 1000 controls

**Table 5.1 Thirty-five previously unreported variants identified in the candidate genes**

Summary of the findings in this study in the five candidate genes sequenced.

## 5.2 The *SMAD6* gene and CVM

In *SMAD6*, ten previously unreported variants were identified in a panel of 90 CVM patients. Five of these variants were present in the controls. The minor allele frequencies (MAF) in controls and cases for four variants: c.-44C>T, c.61G>A, c.711C>T and c.\*262T>C, were not significantly different, indicating that these variants are unlikely to confer substantial risks of CVM. However, they would require further study in more patients and controls to rule out small to moderate effects on risk of CVM. This is especially important for variant c.\*262T>C as this variant is predicted to increase base pairing of the miRNA has-miR-219-2-3p and it has been shown that SNPs that alter miRNA target binding are implicated in breast cancer susceptibility (Nicoloso et al. 2010). miRNAs are small non-coding RNAs (approximately 22 nucleotides in length that inhibit translation by base pairing with the 3'UTR of mRNAs. Nicoloso et al. have identified two SNPs (rs1982073-TGFB1 and rs1799782-XRCC1) that could modulate gene expression by modifying the binding of miRNA with mRNA.

The c.12-C>T variant was found in 7 patients with various phenotypic defects and was observed in the control population with a frequency of 0.029. The difference of MAF between controls and probands was significant ( $p=0.029$ ), providing weak evidence in forming of a relationship with CVM. However, further genotyping of the variant in a larger number of cases and controls would be necessary to confirm the association of the variant with CVM. It is possible that the relatively weak association observed was a chance finding due to multiple comparisons of allele frequencies between cases and controls.

Five variants were absent in the controls: c.-428A>T, c.-290G>C, c.975C>G, c.1244C>T and c.\*319T>A. Sequencing 1000 control chromosomes has 90% power to detect a polymorphism with 1/500 population frequency. However, the power to detect a very rare variant would be considerably lower (e.g. around 65% power to detect a variant of frequency 1/1000). There is an arbitrary element to the selection of the number of controls necessary to suggest a possible pathogenic role for a newly described variant (Collins and Schwartz 2002). Moreover, since the families recruited for this study did not show Mendelian inheritance patterns, the variants found were not expected to be “private” mutations. It is likely that if sufficient control chromosomes were sequenced, additional carriers would be found. Recognizing this, the absence of a

variant in 1000 control chromosomes was indicated as a variant of possible interest warranting further functional investigation. The control population was selected with great care: racial, ethnic and geographic background were similar to the cases and CVM had been ruled out by echocardiography.

The c.-428A>T and c.-290G>C variants were located in the 5'UTR of *SMAD6*. Such variants might have significant effects on gene expression. Variants in the promoter region of many genes have been shown to influence gene expression and in some cases disease susceptibility. For example, Signori and colleagues showed that a mutation in the 5'UTR of the *BRCA1* gene resulted in a 30-50% lower translation efficiency than the wild type allele and was associated with breast cancer risk (Signori et al. 2001). One potential drawback of classical transfection assays to assess the effect of promoter variants is that the variant is examined outside of its genomic context. Approaches such as allelic expression imbalance assays have the potential to overcome this issue. As a recent example, Cunnington and colleagues showed that chromosome 9p21 SNPs previously associated with coronary heart disease in genome wide association studies influence allelic expression of 3 nearby genes (Cunnington et al. 2010). Although the target tissue for these gene expression changes is likely to be vascular smooth muscle cells, that study provided strong evidence for cis-acting effects on gene transcription using RNA from peripheral blood white cells. Similar approaches could be adopted for the further investigation of the c.-428A>T and c.-290G>C variants in future studies. However, this project focus on those novel variants affecting protein coding.

A non-synonymous variant p.Ser325Arg (c.975C>G) in alternative exon 4B was identified in an ASD patient. The patient's parents were unavailable for investigation. The absence of this variant in the controls suggested the change might be a rare variant specific to a CVM family. However, as this variant is located in the linker region with relatively less likely functional effects when compared to the MH1 and MH2 domains of *SMAD6*, functional studies were not performed on this variant.

The c.\*319T>A variant was seen in a proband with PFO and was not seen in the control population. However, the variant was predicted not to bind to miRNA by the *in silico* programmes and it was therefore not prioritized for functional investigation.

The variant that was predicted to be deleterious to protein function in bioinformatic and structural analysis *in silico* was c.1451G>T, resulting in the amino acid substitution p.Cys484Phe. Unfortunately, it was not possible to establish the pattern of inheritance for this variant, as both parents of the patient were deceased. Screening of the siblings of this patient would give a better insight into the inheritance of the variant but that was not possible, as ethics of this study did not permit patients to be re-contacted. The absence of the variant in the controls indicates that variants in the MH2 domain might be associated with CVM. Therefore, the MH2 domain of *SMAD6* was further re-sequenced in a replication cohort of 348 patients. Two previously unreported variants were identified in the second screening. The c.973G>A variant was assumed to result in the amino acid substitution p.Ala325Thr in a patient with mitral valve regurgitation. This variant was inherited from a phenotypically normal father. The variant was not investigated in the control population as it was not successfully incorporated into the iPLEX sequenom assay. It might have been possible to design another sequenom assay for this variant or use a Taqman genotyping assay to check on the controls. However, since an *in vitro* functional assay system to investigate other *SMAD6* variants had been set up by this time, the study of p.Ala325Thr (c.973G>A) variant in that system directly was elected as an alternative approach. The p.Ala325 residue was evolutionarily conserved in human, chimpanzee, mouse and chicken but not in frog, zebrafish and fruitfly; indicating that this substitution was perhaps unlikely to have a major effect on the protein function. This was in accordance with the *in silico* analysis that predicted this variant not to affect protein function.

The c.1244C>T variant resulting in p.Pro415Leu substitution was identified in a patient with aortic stenosis and was inherited from an unaffected father. Therefore, the variant might interact with other sequence variants and environmental factors to cause CVM. The substitution was not detected in the control population. The p.Pro415 residue showed a complete conservation from human to fruitfly, and an effect on protein structure was predicted when mutated to leucine. In addition, both SIFT and Polyphen algorithms predicted this variant to be probably damaging.

The effects on the inhibitory activities of the variants were assessed by dual luciferase assay using a protein expression construct of the MH2 domain of *SMAD6*. No significant difference in inhibitory activity of the p.Ala325Thr (c.973G>A) variant was observed when compared to the wild type, as expected by the outcome of the prediction

programmes and also the lack of evolutionary conservation. The p.Pro415Leu (c.1244C>T) variant revealed significantly lower inhibitory ability than the wild type and the p.Cys484Phe (c.1451G>T) variant failed to inhibit the BMP signaling completely, appearing to be a null allele. The potential mechanisms of the impaired variants in the BMP signaling pathway are discussed below.

Based on the knowledge of BMP signaling in regulating calcification, the additional effects of the mutations (p.Pro415Leu [c.1244C>T] and p.Cys484Phe [c.1451G>T]) on osteogenic potential was investigated. The p.Cys484Phe (c.1451G>T) variant has reduced ability to inhibit alkaline phosphatase activity but not the p.Pro415Leu (c.1244C>T) variant. This suggests that p.Cys484Phe (c.1451G>T) variant might be less efficient in preventing tissue calcification, which is in accordance with the diagnosis of the patient. The patient was diagnosed with coarctation, bicuspid aortic valve and mild aortic stenosis following the discovery of hypertension as a young man. The patient had a coarctation repair at age 30 years. He subsequently developed significant aortic stenosis and underwent aortic valve replacement and repair of the aortic arch again. At the second operation, the transverse aortic arch proximal to the previous conduit, was found to be heavily calcified. There was no evidence of inappropriate calcification in non-cardiovascular tissues.

The p.Pro415Leu (c.1244C>T) variant was found in a patient who presented with a murmur at 18 months and was found to have a bicuspid aortic valve with moderate aortic stenosis. No evidence of coarctation or ectopic calcification was observed in this patient. Both patients demonstrated functionally significant SMAD6 variants in the luciferase assay and had phenotypes involving the aortic valve, suggesting that variation in SMAD6 function might particularly affect the development of the aortic valve. The prevalence of the pathogenic non-synonymous SMAD6 variants among the 47 patients with bicuspid aortic valve and/or aortic stenosis defects out of a total of 436 CVM patients was approximately 4%.

Galvin and colleagues generated Smad6 mutant mice by targeting insertion of a LacZ reporter into the 5' terminus of the Smad6 MH2 domain in embryonic stem cells (Galvin et al. 2000). Analysis of the mouse model revealed that Smad6 was expressed by embryonic day seven (E7) using northern-blot analysis. During development of Smad6 mutant embryos, X-gal staining showed that Smad6 was expressed in the developing

outflow tract and atrioventricular cushion regions of the heart (E9.5-E13.5). The X-gal staining pattern expanded to the blood vessels at later embryonic stages, indicating that Smad6 played an important role during the development of the heart, especially the outflow tract and atrioventricular cushion regions. Smad6 mutant mice showed hyperplastic thickening of the pulmonary and mitral valves, including misplaced septation of the cardiac outflow tract, resulting in an extremely narrow ascending aorta with a large pulmonary trunk or the reverse. Both patients with the functionally significant variants identified in this project had phenotypic defects involving the aorta and outflow tract. Genotyping ratios of the knockout mouse showed reduced numbers of the homozygous mutants than expected, indicating partial lethality. The viable mutant adult mice showed ossification of the outflow tracts of the heart at or after 6 weeks of age. These defects are in keeping with the phenotypic defects of the patient with variant p.Cys484Phe (c.1451G>T) (coarctation of the aorta) that completely failed to inhibit BMP signaling in the luciferase assay and also resulted in osteogenic expression in the alkaline phosphatase assay (calcification of the aorta that was discovered at the second operation a few years later); the patient with variant p.Pro415Leu (c.1244C>T) (aortic stenosis) was significantly unable to inhibit BMP signaling in the luciferase assay but did not reveal any enhanced osteogenic potential in the alkaline phosphatase assay.

This is the first description of a human disease associated with mutations in *SMAD6*. No previous studies have been carried out on the identification of *SMAD6* mutations in CVM. Sequencing of *SMAD6* in 100 patients with colorectal cancer (Fukushima et al. 2003), in 52 hepatocellular carcinoma samples (Kawate et al. 2001), in 30 human ovarian cancers (Wang et al. 2000a), and in 12 pancreatic cancer patients (Jonson et al. 1999), in previous studies has not identified any mutations of likely phenotypic relevance. Therefore, the functionally significant mutations that have been identified in patients with lesions of the aortic valve and aorta demonstrate that *SMAD6* might play an important role in the development and homeostasis of these structures.

Of the five genes investigated in the study, only targeted mutation of *SMAD6* MH2 domain shows a role in the development and homeostasis of the cardiovascular system. *BMP2*, *BMP4*, *BMPRIA* and *BMPR2* are more broadly expressed, and have important additional functions outside of the heart, as the knockout mice of these genes are non-viable. This is consistent with the known extra-cardiac phenotypic consequences of mutations in these genes in humans.

### 5.3 The *BMP2* gene and CVM

In *BMP2*, 6 novel variants were identified, of which three variants were absent in controls. The three variants that were present in the controls are: c.-673G>A, c.-342\_-340delTGT and c.\*936\_939delATTT. The minor allele frequencies (MAF) of controls and cases for both c.-673G>A and c.\*936\_939delATTT variants were not significantly different ( $p=0.512$  and  $p=0.406$  respectively). It is not possible to conclusively rule out association between these variants and CVM from the data. Studies of very much larger numbers of cases would be required to do this. The allele frequencies for both variants are in the range 0.02 to 0.03, and they would therefore be considered low frequency variants. Unless the phenotypic consequences of such low frequency variants are large, it will be very difficult to establish their definite involvement in disease. This is because amassing enough cases that carry the variant to be able to detect moderate odds ratios would require studies of tens of thousands of patients. With respect to functional annotation, the c.\*936\_939delATTT variant was predicted not to alter miRNA binding.

The c.-342\_-340delTGT variant in exon 1 was identified in 8 out of 90 probands and observed in controls with a frequency of 0.024. The difference of MAF between controls and probands was significant ( $p=0.0057$ ). This suggests that the c.-342\_-340delTGT variant might predispose to CVM. Further analysis of this variant in a larger number of cases and controls will be required to answer the question whether the association with CVM is robust.

One of the variants that were absent in the controls was c.-906C>T. This variant was located in 5'UTR of exon 1 and was maternally inherited by a proband with aortic stenosis, indicating incomplete penetrance. The variant was absent in the control population. It remains to be investigated by functional analysis whether the variant results in the increment/decrement of translation efficiency, allelic expression analysis or expression in tissue, whereby it could play a role in CVM susceptibility.

The synonymous c.891G>A and c.1146A>G variants were identified in exon 3. Variant c.891G>A was inherited from an unaffected father, whereas variant c.1146A>G was inherited from an unaffected mother. These variants were absent in the controls and only observed in families with affected CVM probands and healthy parents, indicating that these variants, if significant, are susceptibility factors rather than sufficient

causative factors. The manner in which these synonymous mutations might be causing an effect is however, unknown, and would need further investigation if these findings were to be pursued. Future work might include the generation of a BMP2 construct subcloned into an expression vector, and introduction of the variants by site-directed mutagenesis. Constructs would then be transiently transfected into a mammalian cell line followed by RNA extraction, reverse transcription and immunoblotting. The protein levels in the cell lysates for the variants would then be compared to that of the wild type.

No previous studies have been carried out on screening of *BMP2* mutations in patients with CVM. Previously studies have reported *BMP2* polymorphisms to be associated with certain other phenotypes. A missense polymorphism (p.Ser37Ala) and two SNP haplotypes (haplotype one is defined by three SNPs – TSC0271643(T), P9313(T) and rs235764(G), haplotype two is defined by two SNPs – rs1116867(A) and D35548(T)) in *BMP2* were shown to be associated with low bone mineral density (BMD) and associated fractures in Icelandic and Danish populations using linkage analysis (Styrkarsdottir et al. 2003).

Another study performed by Dathe and colleagues (Dathe et al. 2009) found CNV, but no point mutations in *BMP2* in individuals in two unrelated families affected with brachydactyly type A2 (BDA2), a limb malformation characterized by hypoplastic or aplastic middle phalanges of the second and fifth finger without any CVM. One of the pedigrees is part of a large Brazilian kindred of German origin (five generations, of which 26 individuals were clinically examined and molecular analysis was performed) while another family is smaller (three generations, only 2 individuals were investigated) and also of European origin. They detected a microduplication of approximately 5.5kb in a non-coding region 11kb downstream of *BMP2* in these families using high-density array CGH. They further investigated the evolutionary highly conserved noncoding duplicated region *in vivo*, using a transgenic mouse model, and showed that the duplicated region was able to regulate the expression of an X-Gal reporter construct in the limbs and resulted in increased expression of *BMP2*. This suggests that a regulatory element around *BMP2* that is affected by the duplication might affect the expression of *BMP2* in limb development, and that dysregulation of *BMP2* expression can lead to BDA2.

In conclusion, no definite association of the *BMP2* gene with CVM was identified. No previous studies have investigated this gene in human CVM. With respect to further studies, the findings of Dathe and colleagues on BDA2 have demonstrated the existence of important BMP2 regulatory elements distant from the coding sequence. This study was not designed to examine such genomic features. Although investigation of longer-range regulatory sequences on BMP2 expression and CVM could be considered, such work lay outside the primary hypothesis for this study.

## 5.4 The *BMP4* gene and CVM

In *BMP4*, nine previously unreported nucleotide variants were identified. Five of these variants were absent in controls whereas four variants were present in the controls. The minor allele frequencies (MAF) in controls and cases were not significantly different for three variants: c.370+28G>A variant in intron 3, c.76T>C and c.345C>T (synonymous variants), with p values of 0.753, 0.282 and 0.282 respectively. Since the allele frequencies observed in the controls for these variants are low, studies of many thousands of patients would be required to rule out odds ratios for diseases of 2 or less.

The c.-171C>G variant was identified in 10 out of 90 probands and present in the controls with a frequency of 0.017. The difference of MAF between cases and controls was significant, with a p value of 0.00007, indicating that the variant is more frequent in the probands. It would be interesting to genotype the c.-171C>G variant in further cases and controls to confirm the significance of this association. Genome wide association studies (GWAS) in over 2000 cases and controls are in progress in the host laboratory; for the purpose of this thesis and its focus on rare variants, it was decided not to pursue this preliminary, though interesting, observation in relatively small numbers of patients.

The c.-283G>C variant in exon 1A, c.-295G>A variant in intronic region of exon 1B and c.-87C>T variant in exon 2 were all identified in a patient with TGA and were all inherited from a phenotypically normal father indicating that these variants are part of a haplotype. Moreover, these variants were not seen in the controls. Another variant, c.-136C>G, located in the 5'UTR, was found in another patient with TGA and was inherited from the father but was absent in the controls. These variants were further investigated using the minigene approach.

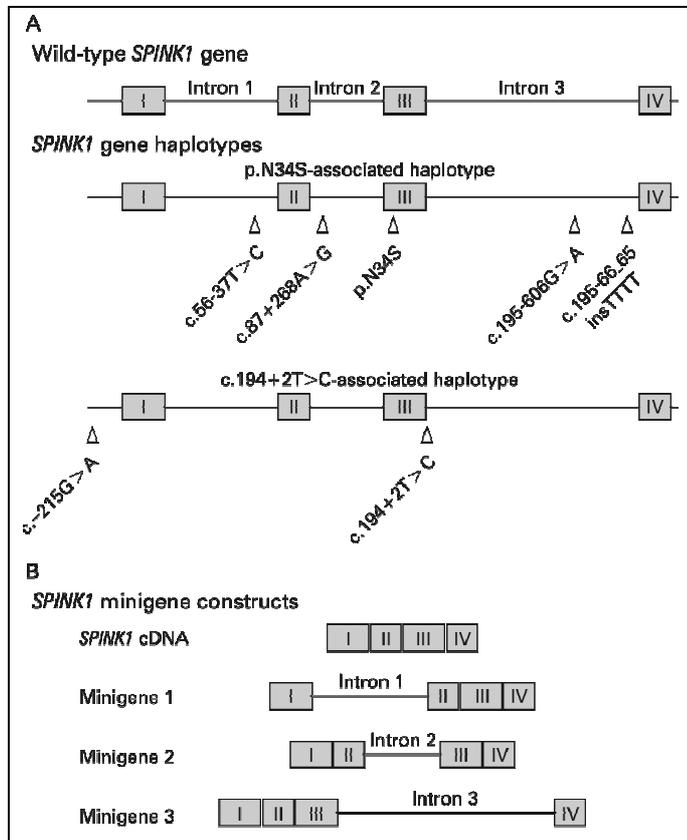
The minigene approach has been used in previous papers to analyse whether synonymous allelic variants found in patients affect splicing efficiency (Fernandez-Guerra et al. 2010; Wilson et al. 2009). This approach has the advantage that RNA is not required and genomic DNA can be used instead. For the c.-295G>A variant in intron 1 no aberrant splicing was observed when the variant was compared to the wild type, even though the ESE finder programme predicted the creation of a branch site in the variant and the programme generated by Dr. Grellscheid predicted the loss of an ESE site. This might be because this variant was part of a haplotype. It is possible that

c.-295G>A on its own does not affect splicing, whereas the full haplotype does. The approach that could be adopted to analyse this is to clone the segment of interest consisting of exon 1B and exon 2 with variant c.-295G>A and c.-87C>T including their flanking regions into the pXJ41 vector. A similar approach was used by Kereszturi and colleagues (Kereszturi et al. 2009) to study multiple intronic variants in common *SPINK1* (serine protease inhibitor Kazal type 1) haplotypes associated with chronic pancreatitis. Wild type and mutant version of introns 1,2 and 3 of the the two haplotypes were studied in parallel (Figure 5.1). They identified the haplotype consisting of the c.194+2T>C intronic alteration abolished *SPINK1* expression in the mRNA level. No detectable functional effect was observed in another haplotype comprising the p.Asn34Ser missense variant and four intronic alterations. However, it would be technically challenging to amplify a substantial fragment of genomic DNA such as the *BMP4* promoter region, and insert the fragment into the vector.

The ESE finder prediction of c.-136C>G variant in exon 1B was not accurate: the programme predicted a decrease of splicing efficiency but the variant was not located at the exon/intron boundary in the context of the gene. The minigene assay demonstrated that there was no difference between the wild type and the mutant variant. Further *in silico* analysis, using the programme by Dr. Grellscheid which only became available in the later stages of this work, showed the change was within the acceptable threshold, in accordance with the minigene results.

For the c.-87C>T variant, more optimization of the cloning parameters would be required in the future in order to clone the variant into the vector, if this line of experimentation were pursued. However, this would have low priority in further work, as both prediction programmes did not predict aberrant splicing for the variant.

Only one previously unreported variant that resulted in a non-synonymous substitution was identified, in exon 4 of *BMP4*. The c.520G>A variant resulted in a glycine to serine substitution (p.Gly174Ser). The variant was not detected in the controls. However, protein prediction programmes predicted this change not to affect protein function. This variant is evolutionarily conserved from human to zebrafish but not in fruitfly. It remains to be investigated by functional analysis; one possible approach would be the



**Figure 5.1** *SPINK1* gene haplotypes (A) and *SPINK1* minigenes (B)

Common pathogenic *SPINK1* gene haplotypes (A) and *SPINK1* minigenes used (B) in the study of Kereszturi et al., (Kereszturi et al. 2009).

study described by Tabatabaeifar and colleagues (Tabatabaeifar et al. 2009), who investigated the characteristic of *BMP4* mutations identified in patients with congenital anomalies of the kidney and urinary tract (CAKUT). Three missense mutations (p.Ser91Cys, p.Thr116Ser and p.Asn150Lys) found in five patients were studied *in vitro* using the full-length complement DNA clone of human *BMP4*. The mutants resulted in lower level of mRNA abundance of *BMP4* in the cellular systems when checked with real time reverse-transcription polymerase chain reaction (RT-PCR) and reaffirmed with western blotting. However, as the variant identified was predicted to be benign, it was not prioritized for functional analysis.

The coding region of *BMP4* had previously been analysed in 205 patients with congenital septal defects by Posch and colleagues (Posch et al. 2008), as cardiac septal defects are among the major CVMs. Mutations in transcription factors (*NKX2.5* and *GATA4*) had been previously identified in patients with septal defects (Garg et al. 2003; Schott et al. 1998). *BMP4* is required for development of the endocardial cushions and inactivation of this gene in a mouse model caused atrioventricular septal defects (AVSD). The study of Posch et al. was published after the commencement of this work. 110 patients in that study were diagnosed with ostium secundum atrial septal defects (ASDII) and 95 patients were diagnosed with different congenital septal defects and concomitant minor cardiac malformations. By contrast, in this study, *BMP4* was sequenced in a total of 270 patients with a broad range of CVM. Of these, 55 patients were diagnosed with TGA or TGA with other cardiac defects, and 65 patients with septal defects. Posch et al. identified 2 patients with a c.845+30G>A variant that had not been described previously. No investigation was performed on the inheritance of variants. Posch et al. also investigated the reported non-synonymous SNP, p.Val152Ala located in exon 4 of *BMP4* in cases and in 360 control alleles from unaffected individuals and found that the differences in allele frequencies between cases and controls was not significant. They concluded that patients with septal defects in general do not exhibit mutation in *BMP4*. A non-synonymous mutation, p.Gly174Ser (c.520G>A), was identified in a patient with ASD that was inherited from a phenotypically normal mother, suggesting that this allele, if causative, has reduced penetrance and might require environmental factors to cause CVM. This variant is absent in the controls but was predicted to be benign. Subsequent *in vitro* functional investigation on the rare variants located in the untranslated region that were absent in the controls showed that they did not abrogate splicing. Both this study and the study

conducted by Posch and colleagues, did not identify rare variants in patients with AVSD, the phenotype observed in the *BMP4* knockout mouse, but the number of AVSD patients were small in both studies (10 in this study and 16 in Posch et al.). This study focused on a larger number of TGA patients instead due to the work by Thienpont and colleagues that identified a deletion on 14q21.3, which is near *BMP4* (14q22.2), in a patient with TGA and other phenotypic defects (Thienpont et al. 2007). However, no *BMP4* mutations were found in TGA patients in this study.

*BMP4* has been associated with other phenotypic defects. A study performed by Suzuki and colleagues (Suzuki et al. 2009) identified a novel missense mutation, p.Ala346Val, in a patient with microform cleft lip and a cleft palate. The mutation was inherited from the father who had a bifid uvula and microform cleft lip. Suzuki et al. also identified two missense mutations, p.Ser91Cys and p.Arg28His, in parents with *orbicularis oris* muscle (OOM) defects and a child with unilateral cleft lip and palate. Another five previously unreported missense or nonsense mutations were identified in 968 patients with cleft lip and/or cleft palate (CL/P) in that study. One of these mutations, p.Arg162Gln, was inherited from a phenotypically normal father. None of these mutations were found in 529 controls. SIFT and Polyphen programmes predicted the p.Ser91Cys, p.Arg162Gln and p.Ala346Val variants to be possibly damaging. No functional investigation was performed on these variants to ascertain the effects. That study showed increased severity of the phenotype in children than in carrier parents, i.e. anticipation. However, the numbers of cases with parent-to-child anticipation identified were small in that study. If the anticipation seen in the Suzuki study were a general finding with *BMP4* mutations, the haplotype identified in this study (c.-283G>C, c.-295G>A and c.-87C>T) that was inherited from a phenotypically normal father (who did not have an echocardiogram evaluation) in a TGA proband might be exhibiting this phenomenon.

In another study, using a positional candidate gene approach, Bakrania and colleagues (Bakrania et al. 2008) sequenced the coding region of *BMP4* and identified a frame-shift mutation c.226del2, resulting in a truncated protein at amino acid 104, p.Ser76fs104X, in a proband who had anophthalmia-microphthalmia (AM), retinal dystrophy, myopia, poly- and/or syndactyly, and brain anomalies. The proband's mother, grandmother and maternal aunt carried the same mutation and had high myopia but not AM. The proband's normal two half-brothers did not carry the mutation. Further investigation

revealed that the proband also had a c.1-125G>A change in the 5'UTR of *SHH* (Sonic Hedgehog). *BMP4* interacts with *HH* signaling genes in animals in optic development. The *SHH* c.1-125G>A was also present in 16/191 controls. Other family members of this proband had a mutation in either *BMP4* or *SHH*, but not both, indicating that the combination of mutations might have contributed to the more severe defects observed in the proband. That study shows the potential for gene-gene interaction involving *BMP4* to affect the manifestation of developmental disorders. It is possible that such interaction is of importance in CVM.

In conclusion, previous studies have shown that mutations in *BMP4* were infrequently observed in CVM patients but were strongly associated with other phenotypic defects, suggesting that *BMP4* is important during development of multiple organ systems. This study did not identify association of *BMP4* with CVM (enriched with TGA). However, the suggestion of a link between AVSD with *BMP4* based on a mouse model would require further study in larger numbers of AVSD patients to clarify such relationship. Moreover, the ascertainment scheme for the present study would have tended to exclude patients with multiple and possibly syndromic phenotypes.

## 5.5 The *BMPRIA* gene and CVM

Only one previously unreported variant (c.543-59T>C) was identified in intron 7, 59 bases upstream from the 5' end of exon 8 in a patient with atrial septal defect (ASD). It remains to be investigated whether this variant is present or absent in the control population and also to determine its inheritance. This variant was located in an intron and no novel variant was observed in the coding region. As discussed previously, rare non-synonymous variants located in the coding region are expected to have greater effect on disease susceptibility than non-coding variants (Kryukov et al. 2007), and therefore this gene was not prioritized for further study.

*BMPRIA* mutations have been previously observed in patients with juvenile intestinal polyposis, which carries a risk of gastrointestinal cancers. Howe and colleagues established that between 35- 60% of juvenile polyposis syndrome (JPS) cases in North America were due to mutations in the *SMAD4* gene (Howe et al. 1998). To detect additional polyposis genes, they used genome wide screening in four multiplex families who were known to carry no mutation in *SMAD4*. Howe and colleagues (Howe et al. 2001) identified linkage with markers from chromosome 10q22-23, leading them to sequence the *BMPRIA* gene in those kindreds. Mutations in *BMPRIA* were discovered in each of the 4 families. One family had a four base pair deletion in exon one resulting in a stop codon at nucleotide 104-106. A second family had a one base pair deletion in exon 8 resulting in a stop codon, and two nucleotide substitutions causing stop codons (p.Gln238X and p.Trp271X) were found in the remaining two kindreds. None of these mutations were observed in 250 normal control individuals. The nonsense mutations were located at the serine-threonine kinase domain and were predicted to abrogate BMP signaling. However, no *in vitro* functional study was carried out on these mutations. Interrogation of other genes in the TGF-beta superfamily, including BMP signaling, (*SMAD1*, *SMAD2*, *SMAD3*, *SMAD5*, *SMAD7*, *BMPRI1B*, *ACVR1*, and *BMPRI2*) have yielded negative results (Bevan et al. 1999; Howe et al. 2004; Roth et al. 1999), suggesting that other genes predisposing to juvenile polyposis remain to be discovered.

In another study, principally focused on JPS, Zhou and colleagues (Zhou et al. 2001) found 2 patients with both cardiac anomalies (one with ventricular septal defects and the other with Ebstein's anomaly) and juvenile polyposis syndrome harboured mutations in *BMPRIA* without mutation in *SMAD4*. They studied 18 unrelated families and 7

unrelated patients. The mutation in the patient with ventricular septal defects was located in the 5' UTR (IVS1-3C>G). It was a splice site mutation resulting in the skipping of exon 1, and was predicted to result in truncated receptors. The patient with Ebstein's anomaly and JPS also had Wilms' tumour. The mutation discovered in that patient resulted in a cysteine to tyrosine substitution (p.Cys376Tyr) in the kinase domain of *BMPRIA*. These two mutations were absent in the 100 normal race-matched control chromosomes. However, Zhou and colleagues did not identify any *BMPRIA* mutation in two further patients with juvenile polyposis and cardiac anomalies: one diagnosed with ventricular septal defect and the other with aortic regurgitation. The frequency of patients with JPS and cardiac anomalies was not significantly different among mutation carriers (2/10 carriers) and non-carriers (2/15 non-carriers) due to the small sample size. Therefore, genotype-phenotype association of JPS and cardiac anomalies with mutation in *BMPRIA* could not be established.

In spite of this, the mouse model with the targeted deletion of *BMPRIA* in cardiac myocytes of the atrioventricular canal displayed tricuspid valve abnormalities reminiscent of Ebstein's anomaly (Gaussin et al. 2005). There were longer tricuspid mural leaflets and the tricuspid posterior leaflet was displaced and adherent to the ventricular wall. These defects potentially support a role for *BMPRIA* in human Ebstein's anomaly. As only 3 of the 90 sequenced probands were diagnosed with Ebstein's anomaly in this work, it remains of interest to sequence a larger number of patients with this defect in order to increase the power of detecting novel variants in *BMPRIA*.

In summary, this study did not find any association of *BMPRIA* gene with CVM. It might be that *BMPRIA* mutations might be found in a subset of patients with CVM and also predispose to cancer but not in patients with sporadic CVM only. The suggestion of a link between Ebstein's anomaly and *BMPRIA* would require further study to establish this relationship.

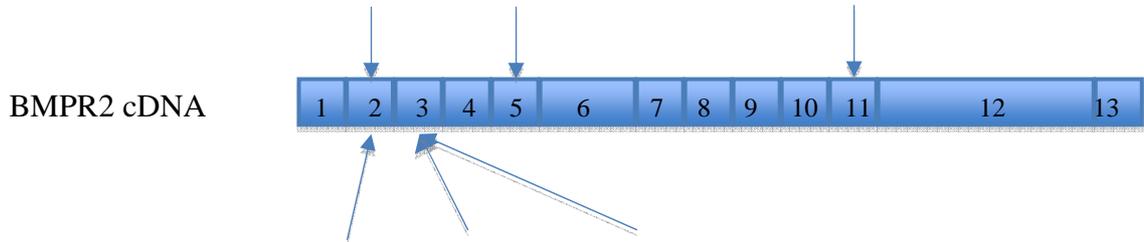
## 5.6 The *BMPR2* gene and CVM

Seven genetic variants were identified in *BMPR2*. One was located in the 5'UTR, 5 variants were seen in the intronic regions, and one non-synonymous variant in exon 13 resulted in a glycine to cysteine substitution. Due to the pressure of more significant results in other genes screened, these results were not pursued further. The next step in the investigation of these variants would be genotyping in the parents of cases, and in the control population. If the variant in the 5'UTR were not seen in the controls, functional analysis using luciferase assay could be performed to investigate whether the variant increases or decreases expression of the gene. The 5 variants located in the intronic regions could be investigated with prediction programmes and minigene assays. Depending on the outcome of the genotyping, if the c.2887G>T (p.Gly963Cys) variant seen in an ASD patient were not detected in the 1000 control chromosomes, it would be interesting to investigate by co-transfecting wild type or mutant *BMPR2* construct, *BMPR1A* construct and a p3TP-Lux promoter-reporter construct or BRE-Luc construct to determine whether this non-synonymous variant will affect the BMP signaling activity, as the *in silico* analysis predicted that the variant to be damaging. The *BMPR2* construct and p3TP-Lux promoter-reporter construct have been published (Rosenzweig et al. 1995; Wrana et al. 1992). The p3TP-Lux promoter-reporter is a luciferase reporter gene with the 3TP promoter that contains three consecutive TPA response elements (TREs) and a portion of the plasminogen activator inhibitor 1 (PAI-1) promoter region. These regions were chosen based on the responsiveness of these elements to TGF $\beta$ .

A previous study performed by Roberts and colleagues (Roberts et al. 2004) identified mutations in *BMPR2* in patients with both pulmonary arterial hypertension (PAH) and CVM, in whom the PAH was due to pulmonary vascular obstructive disease. The subjects of their study consisted of two cohorts: 40 adults and 66 children. A total of six non-synonymous mutations were found in the combination of 106 patients (Figure 5.2). Three mutations (p.Gln42Arg, p.Thr102Ala and p.Ser107Pro) located in exon 2 and 3 were identified in three out of four adult patients with atrioventricular canals. Another three mutations (p.Gly47Asn, p.Met186Val and p.Glu503Asp) located in exon 2, 5 and 11, were identified in the children. One of the children had an atrial septal defect and patent ductus arteriosus, one had an atrial septal defect, patent ductus arteriosus and

Mutations in  
children's cohort

p.Gly47Asn      p.Met186Val      p.Glu503Asp



Mutations in  
adult's cohort

p.Gln42Arg      p.Thr102Ala      p.Ser107Pro

Receptor domains



ECD Extracellular domain

TM Transmembrane

KD Kinase domain

CD Cytoplasmic domain

**Figure 5.2 Location of the mutations in BMPR2 identified by Roberts et al., 2004**

Location of the mutations in the BMPR2 identified by Roberts et al. with exons 1-13 indicated, and the receptor domains shown below the exons (Roberts et al. 2004).

partial anomalous pulmonary venous return; and one had an aortopulmonary window and a ventricular septal defect. These mutations were located in the extracellular domains (exon 2 and 3) and might interfere with heterodimer formation or ligand binding, in the kinase domain (exon 5) that might disrupt the phosphorylation and the long cytoplasmic tail (exon 11) with unknown function. It was noted that no children with atrioventricular canal defects had mutations in *BMPR2*. Roberts and colleagues did not exclude syndromic patients from their analysis. The patient with complete atrial ventricular canal defect and the p.Gln42Arg mutation also had Down syndrome and the patient with atrial septal defect and patent ductus arteriosus with the p.Met186Val mutation had a ring 14 chromosome. The cardiac defects in these patients might therefore have arisen principally from the chromosomal abnormalities rather than the *BMPR2* mutations. The p.Gly47Asn and p.Ser107Pro mutations were not inherited from the parents, demonstrating that these are *de novo* mutations. The altered amino acids of these two mutations are conserved only in man and mouse, in contrast to the other four mutations that are conserved across human, mouse, chicken, frog and puffer fish.

Other mutations were not checked for inheritance, as the parents' DNA were not available. However, Roberts and colleagues did not check the mutations that they found in a control population, giving rise to the question whether these mutations are polymorphisms or they are indeed rare mutations predisposing to defects. The mutations were interspersed in different domains of *BMPR2*, suggesting that there is no one particular domain that causes the defects. Even though the function of the long cytoplasmic tail domain (p.Glu503Asp in exon 11) in the study of Roberts et al. and the mutation that have been identified in this study (p.Gly963Cys [c.2887G>T] in exon 13) is still unknown, the presence of two mutations suggest that this domain might be of importance. Taken together, the findings of Roberts and colleagues, although suggestive, do not establish a conclusive role for *BMPR2* mutations in CVM. Since mutations in *BMPR2* are known to cause primary PAH, their study design would not have enabled mutations causing CVM to be distinguished from mutations that could cause PAH but were 'innocent bystanders' with regard to CVM.

Nevertheless, the study of Roberts and colleagues raises important questions. In the setting of CVM, the development of pulmonary hypertension due to left to right shunting is variable among patients with the same malformation. It remains unknown whether there is genetic susceptibility to the development of PAH in the context of

CVM. In order to address this question with the greatest power, it would be necessary to establish a large cohort of patients with the same phenotype, and relate their susceptibility to the development of PAH to common or rare genetic factors. While this is an approach that would potentially yield clinically significant and useful results, the practical difficulties of establishing phenotypically homogeneous cohorts of many thousands of CVM patients are substantial.

Mutations in *BMPR2* have been identified in many patients with pulmonary arterial hypertension (PAH), be it idiopathic PAH (IPAH) or familial PAH (FPAH). These heterozygous mutations comprise nonsense, missense, splice site defects, small insertions or deletions and duplications, and partial gene deletions or duplications (Machado et al. 2006). Cell-based *in vitro* analysis of mutations in *BMPR2* carried out by Rudarakanchana and colleagues (Rudarakanchana et al. 2002) to test their abilities to transduce BMP signals have shown that missense mutations located within the extracellular and kinase domains abrogated their signal-transducing abilities. However, the mutant *BMPR2* constructs with truncated cytoplasmic tail or missense mutations as observed in the PAH patients (Machado et al. 2001) were able to transduce a BMP signal, suggesting that the mutant might affect BMP signaling through an alternative pathway independent of Smads. They then examined the effect of the *BMPR2* mutants (including the cytoplasmic tail mutants) on p38<sup>MAPK</sup>, one of the MAP kinases activated by BMPs. It was shown that overexpression of the mutants lead to the activation of p38<sup>MAPK</sup> in the absence of ligand but the wild type receptor required the presence of BMP4 to activate p38<sup>MAPK</sup>. Therefore, the p.Gly963Cys (c.2887G>T) variant that have been identified in the cytoplasmic tail might disrupt p38<sup>MAPK</sup> signaling but not through SMAD signaling.

In conclusion, the preliminary findings of this study and the study of Roberts et al. (Roberts et al. 2004) suggest a possible association of *BMPR2* and CVM. However, as no previous investigation on the occurrence of *BMPR2* mutation in CVM patients without PAH have been performed, further investigation will be required to differentiate the role of *BMPR2* in PAH from any role of *BMPR2* in CVM. The mutations in *BMPR2*, especially the ones located in the cytoplasmic domain (p.Gly963Cys [c.2887G>T] in this study and p.Glu503Asp in Roberts' study), where both patients had ASD, might disrupt the BMP signaling involving p38<sup>MAPK</sup> activation rather than SMAD signaling resulting in heart malformation.

## 5.7 Possible mechanisms whereby SMAD6 variants identified in this study affect BMP signaling and CVM

### 5.7.1 BMPR1A binding

Several mechanisms have been reported for the inhibitory effect of SMAD6 on BMP signaling. SMAD6 has been demonstrated to preferentially bind to *BMPR1A* through the N-terminal lobe of the kinase domain and inhibit phosphorylation of BMP-specific R-Smads (Receptor activated Smads: Smad1, Smad5 and Smad8) to repress BMP signaling (Goto et al. 2007; Imamura et al. 1997). The residues in type I receptors that are involved in SMAD6 binding are located around the phosphate binding loop of the BMPR1A kinase domain, not on the L45 loop (Goto et al. 2007). The L45 loop is the structure that protrudes from the kinase domain of type I receptors that is responsible for intracellular signaling (Huse et al. 1999). It has been shown to activate specific R-Smad isoforms (Chen et al. 1998; Feng and Derynck 1997). Thus, it would be interesting to carry out studies on the mode of interaction between I-Smads (especially the mutant SMAD6s identified in this study) and BMPR1A, which would provide insights into how these mutants might affect the binding of SMAD6-BMPR1A and inhibit BMP signaling. The L3 loop is an important structure in the MH2 domain that determines SMAD recognition by BMPR1A (Lo et al. 1998). As the alteration to the crystal structure consequent upon the SMAD6 p.Cys484Phe (c.1451G>T) variant was predicted to disrupt the structure of the L3 loop, and p.Pro415Leu (c.1244C>T) was predicted to destabilize the structure located on the same face as the L3 loop, we hypothesized that these variants will affect binding of SMAD6 to BMPR1A. Lo and colleagues (Lo et al. 1998) have shown that mutating residues that are highly conserved in this loop in Smad2 resulted in greatly diminished or abolished Smad2 binding to the TGF- $\beta$  receptor complex. However, mutations of highly conserved residues in other structures, for example,  $\alpha$ -helix 2 and  $\alpha$ -helix 3 did not affect the binding of Smad2 to the receptor complex, indicating that these other regions are not crucial for SMAD-receptor interaction.

The cysteine 484 residue of SMAD6 is highly conserved from human to *Drosophila*, and this residue is also observed in SMAD7, another inhibitory Smad, suggesting that this residue is important in inhibiting BMP/TGF $\beta$  signals. This hypothesis is confirmed by the luciferase assays which demonstrated the complete loss of inhibitory effect of this mutant. The functional analysis and analysis of the crystal structure may also imply

that the change from cysteine to phenylalanine failed to inhibit BMP signaling due to the failure of the SMAD6 to stably bind to BMPR1A.

Another mutant that was identified was the proline 415 residue of SMAD6. This residue is highly conserved from human to drosophila for SMAD6 but differs in SMAD7, suggesting that this residue is specific to BMP signaling but not TGF $\beta$  signaling. The functional study performed on this variant showed significant loss of inhibitory effect in BMP signaling, but to a lesser extent than the p.Cys484Phe (c.1451G>T) variant. Analysis of crystal structure of both mutants suggested the most probable mechanism is due to the defective binding of the mutants to BMPR1A, resulting in loss of inhibitory effects. To test this mechanistic hypothesis, immunoblotting of SMAD1/5/8 and phospho-SMAD1/5/8 were carried out. If the SMAD6-BMPR1A complex were formed, it would be expected that the intensity of the phospho-SMAD1/5/8 bands to be lesser (or not visible at all) than the SMAD1/5/8 bands and vice versa. However, the difference was not visible as the serum in the culture medium and the low transfection efficiency masked the expression of the transfected cells.

### **5.7.2 Diminished competition with co-Smads**

Smad6 has been demonstrated to bind to phosphorylated Smad1, interfering with the complex formation of phospho-Smad1 and Smad4 by acting as a Smad4 decoy (Hata et al. 1998). Using a yeast two-hybrid system these authors showed that Smad1-Smad6 interaction was mediated by the C-domains (MH2 domains) of these proteins. In addition, these authors generated a Smad6 mutant, p.Gly471Ser, located in the L3 loop, and showed that the mutant did not interact with the phospho-Smad1 and failed to inhibit signaling. However, the mutant was able to bind to BMPR1B (type I receptor). The results of the western blot probed with SMAD1/5/8 and phospho-SMAD1/5/8 of the cell lysates from the luciferase assay that was performed was not conclusive. Further work to establish whether the mutant SMAD6s in this study failed to repress BMP signaling through this mechanism would be of interest.

### **5.7.3 Impaired cooperative inhibition with Smurf1**

Smurf1 (Smad ubiquitin regulatory factor) is a type of E3 ubiquitin ligase that induces the ubiquitination and degradation of the BMP-specific Smads (Smad1/5/8) independent of signal activation (Zhu et al. 1999). Murakami and colleagues (Murakami et al. 2003) demonstrated that Smad6 enhanced the interaction of Smurf1 with BMP type I receptors to prevent the phosphorylation of Smad1/5/8 and to remove the receptor complex through ubiquitin-dependent degradation. They also showed that Smad6 enhanced the interaction of phosphorylated R-Smads with Smurf1 and promoted proteasome-dependent degradation of this complex. These mechanisms raise the possibility that the functionally significant variants of SMAD6 were not able to link BMPRII to SMURF1 (due to the lack of interaction of SMAD6 and SMURF1) and subsequently promote degradation of the complex. Alternatively, it is possible that the mutant SMAD6s were not able to associate SMURF1 to phosphorylated-SMAD1/5/8 and thereby repress BMP signaling.

Smad6 recruits Smurf1 to degrade the Smad6-Tbx6 protein complex (Chen et al. 2009b), indicating that Smad6 and Smurf1 not only cooperate to regulate the Smad proteins but also to regulate other proteins. Tbx6 is a member of the family of transcription factors that share a common T-box domain which is highly conserved evolutionarily. The p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) variants that have been identified might impair the capacity of SMAD6 to interact with TBX6, or possibly other cardiac transcription factors. Further work would be necessary to establish the interaction between SMAD6 and other transcription factors downstream of BMP signaling. If this could be achieved, it would be interesting to investigate whether gene-gene interaction between SMAD6 and variants in other transcription factors was responsible for the variable penetrance we observed with SMAD6 variants in this study.

### **5.7.4 Nuclear actions**

In COS7 cells and HepG2 cells, Smad6 was observed in the nuclei as well as in the cytoplasm (Hanyu et al. 2001), indicating that the Smad6 mechanisms take place at the nuclear level in mammalian cells. Smad6 has also been shown to interact with transcription factors to inhibit BMP signaling in the nucleus (Bai et al. 2000). The

carboxyl-terminal domain (MH2 domain) of Smad6 was found to interact with Hoxc-8 transcription factor, thereby inhibiting Smad1-mediated transcriptional activity in the nucleus. Therefore, the SMAD6 mutants could affect interaction with transcription factors to inhibit BMP signaling. Such a mechanism would not necessarily imply that levels of phospho-SMAD1/5/8 would differ between wild-type and mutant, as the effect of the mutation would be downstream of co-SMAD phosphorylation.

This study raises questions on how the variants identified in SMAD6 affect the mechanism of SMAD6 repression of BMP signaling. This section has discussed the possible mechanisms of SMAD6 action and possible future experiments are discussed in section 5.12.

## 5.8 This study in the context of previous resequencing studies of CVM

### 5.8.1 Other genes in the BMP signaling pathway

Based on my results, further investigation of genes involved in BMP signaling is indicated. During the time this work was being done, Smith and colleagues (Smith et al. 2009) sequenced 32 candidate genes known to be important in atrioventricular septum (AVS) development in 190 patients with AVS defects and identified three missense mutations in *ACVRI* (*ALK2*). *ALK2* is a type I BMP receptor that transduces BMP signaling. Loss of *ALK2* in the endocardium of a knockout mouse was previously shown to result in hypoplastic endocardial cushions (Wang et al. 2005a). The results of Wang et al. supported the investigation of AVS defects in the study by Smith et al. as endocardial cushion contribute to the valves and septa of the heart during development. The three missense mutations discovered by Smith et al. were p.Ala15Gly, p.Arg307Leu and p.Leu343Pro. Nine of the 350 control individuals genotyped carried the p.Ala15Gly mutation, implying that this variant is common to the general population. However, the p.Arg307Leu and p.Leu343Pro mutations were absent in the 350 controls. Pedigree analysis of these mutations found the p.Arg307Leu mutation was inherited from a phenotypically normal father. The p.Leu343Pro mutation was inherited from the father who had a cardiac murmur but the inheritance of the mutation with a cardiac phenotype could not be established. Only the p.Leu343Pro mutation was predicted to affect protein function when the missense mutations were analyzed with the SIFT and PolyPhen programmes. Crystal structure of the p.Leu343Pro mutation modeled on transforming growth factor- $\beta$  receptor I showed that when leucine is substituted with proline, the  $\beta$  sheet would be disrupted, destabilizing the structure of the kinase domain of the receptor. Functional analysis showed that wild type *ALK2* was able to transduce signaling in a luciferase assay. Luciferase assay of the mutations showed that there was no significant difference in the ability of the p.Ala15Gly and p.Arg307Leu mutants to induce BMP signaling compared to the wild type. However, the p.Leu343Pro mutant resulted in significantly lower activity than the wild type. This is compatible with the dominant-negative form of the receptor where the mutant and the wild type constructs were co-expressed in the cells; the expression of the wild type construct was suppressed (Visser et al. 2001). This dominant-negative *ALK2* mutant was further investigated in vivo in a zebrafish model where wild type or mutant *ALK2* RNA was injected into zebrafish embryos at the single-cell stage. Mutant RNA revealed

severe dorsalization whereas wild type ALK2 RNA showed mild ventralization, indicating that ALK2 is important in dorsoventral patterning of the zebrafish embryo. The absence of the *SMAD6* p.Pro415Leu (c.1244C>T) and p.Cys484Phe (c.1451G>T) mutant in the control population, analysis of the crystal structures and the subsequently demonstrated functional significance in my study was similar to the study on *ALK2*. However, *ALK2* was sequenced in 190 phenotypically more homogenous cases of CVM, all of whom had septal defects whereas in this study, the MH2 domain of *SMAD6* in was sequenced in 438 patients with a broad range of phenotypic defects. The functionally significant *ALK2* mutation was present in one out of 190 patients (0.5%) with AVS defects. The two functionally significant *SMAD6* mutations were observed in two out of 438 CVM patients (0.5%) but the prevalence of these mutations is higher, 4.3%, in the subset of 46 patients with aortic stenosis or coarctation of the aorta that have been sequenced. Both the study of Smith et al. and my study identified mutations inherited from phenotypically normal fathers (*ALK2* p.Arg307Leu and *SMAD6* p.Pro415Leu), indicating that they likely interact with other variants and environmental factors to cause CVM. The study on *ALK2* and my study on *SMAD6* suggest that mutations in these genes might be causative for particular CVMs: *ALK2* for AVS defects and *SMAD6* for aortic stenosis or coarctation of the aorta. While the results of Smith et al showed that down regulation of BMP signaling by a dominant negative form of ALK2 could result in septal defects, this study has shown that an inappropriately high level of BMP signaling, due to loss of function mutations in the inhibitory *SMAD6*, could lead to malformations of the aortic valve and aorta. These results suggest a differential sensitivity to higher or lower levels of BMP signaling among the different anatomical regions of the developing heart, since low BMP signaling did not affect the aortic valve in the study of Smith et al., while overactivity of the pathway did not affect septal structures in the patients that were identified with mutations in *SMAD6*. This observation could be of importance for future studies. If up regulation or down regulation of particular signaling pathways only has significant implications for particular subgroups of CVM phenotypes, this will substantially complicate the identification of robust genotype-phenotype relationships.

### 5.8.2 Genes in other signaling pathways

A subset of genes has been implicated as non-syndromic causes of CVM in previous resequencing studies. Large families with autosomal dominant inheritance of CVM without other organ malformations have demonstrated that CVM can result from single-gene defects. Mutations in the homeodomain of homeobox transcription factor *NKX2.5* were found to cause non-syndromic Mendelian CVM in four large families with multiple affected individuals (23 individuals with secundum ASDs, 4 individuals with secundum ASDs and other CVM and 4 individuals with other structural heart malformations) (Schott et al. 1998). A multiplex family with aortic valve anomalies and severe valve calcification inherited in a Mendelian fashion was identified with premature stop codons in the *NOTCH1* gene (Garg et al. 2005). In *GATA4*, a p.Gly296Ser mutation was identified in a Mendelian family with septal defects and a frameshift mutation resulting in premature stop codon in a second family (Garg et al. 2003). These mutations segregated with the phenotypes and were not observed in controls, indicating that they are responsible for familial CVM. However, the incidence of large families with a Mendelian inheritance pattern of non-syndromic CVMs is rare. Most cases of CVM observed are sporadic cases, where no immediate family member is diagnosed with CVM. A number of resequencing studies in such non-syndromic cases have been performed. In general they have focused on transcription factors that are important in cardiac development. An example is the studies on *NKX2.5*. Here, mutations in the gene have been found, but they are present in less than 5% of CVM patients (Goldmuntz et al. 2001; McElhinney et al. 2003). Goldmuntz and colleagues (Goldmuntz et al. 2001) sequenced *NKX2.5* in 114 non-syndromic sporadic TOF patients and found 4 mutations that were not identified in 100 normal chromosomes. Mutations p.Glu21Gln, p.Arg216Cys and p.Ala219Val resulted in substitution of evolutionary highly conserved amino acids and therefore were considered to be significant missense mutations. The p.Arg216Cys and p.Ala219Val mutations were located in the functionally important NK2 domain. Some phenotypically normal family members of p.Glu21Gln and p.Ala219Val mutations were however found to carry these mutations. The fourth mutation, p.Arg25Cys, was identified in 3 unrelated probands and was inherited from affected or normal parents. This mutation is located in the coding region outside of conserved domains. Functional studies performed by Kasahara and colleagues (Kasahara et al. 2000) on this mutation showed impaired DNA binding, supporting a role in CVM. However, this mutation was identified largely in African-

American probands, and screening of this mutation in controls of the same ethnicity showed that it existed in 4.7% of normal subjects. Thus, the notion that this variant predisposes to CVM appears to be a false positive. This finding demonstrated the importance of ethnically matched cases and controls, in resequencing studies. There are other examples of false positive results due to ethnic heterogeneity. For example, in a mixed population of patients the *FIBRILLIN-1* p.Pro1148Ala substitution was believed to be a mutation causing Marfan syndrome as it was not seen in white or African-American controls, but it was found to be a polymorphism in the Asian population (Wang et al. 1997). My patients and controls are of Caucasian ancestry, therefore the question of population stratification between cases and controls should not arise. Subsequently McElhinney et al. sequenced *NKX2.5* in a panel of 608 patients with a broad range of CVM and showed that *NKX2.5* mutations occur in a small percentage of patients (3%) (McElhinney et al. 2003). 4% of the mutations were identified in patients with TOF (9 out of 201 TOF patients) and 4% (3 out of 71) in patients with a secundum ASD. These proportions are similar to those of the *SMAD6* mutations observed in this study, which occur in 4% of patients with defects in the aorta.

Another study that investigated non-syndromic sporadic CVM population with a broad range of defects for mutations in a single gene, similar to my study, was performed by Sperling and colleagues (Sperling et al. 2005). A cohort of 392 patients with various sporadic non-syndromic CVM was sequenced for *CITED2* mutations (Sperling et al. 2005). *CITED2* is a ubiquitously expressed transcriptional cofactor that coactivates TFAP2 and interacts with the histone acetylases CREBBP/EP300 through the C-terminus (Bhattacharya et al. 1999). This complex can be detected at the *Pitx2c* promoter in embryonic mouse heart, suggesting that *CITED2* plays a role in left-right patterning of the heart through Nodal-PITX2C pathway (Bamforth et al. 2004). The high affinity binding of *CITED2* with CREBBP/EP300 resulted in competitive inhibition of the interaction of EP300 and the transcription factor HIF1A. Seven novel variants in *CITED2* from 8 patients were identified; they were absent in 192 control individuals. Of these, three variants resulted in deletion or insertion of amino acids, namely p.Ser170\_Gly178del, p.Gly178\_Ser179ins9 and p.Ser198\_Gly199del, each in a patient with perimembranous ventricular septal defect, secundum atrial septal defect and sinus venosus atrial septal defect with abnormal pulmonary venous return to the right atria. These mutations were located in the serine-glycine rich junction but not in the EP300 binding domain. Analysis of these mutations using luciferase assays showed a

significant loss of ability to repress HIF1A transcriptional activity, suggesting that the variations in the serine-glycine rich junction might alter the activity of CITED2, affecting the EP300 binding domain to interact with CREBBP and EP300 or other cofactors. One shortcoming of this study is that the parents' DNA were not available for investigation. This study showed that 3 out of 392 CVM patients carried functionally significant variants (0.77%), which is similar to the functionally significant variants identified in SMAD6. This study also revealed that CITED2 is associated with CVM, in particularly septal defects and malrotations of the great arteries.

No studies have been performed systematically investigating a particular signaling pathway in a cohort of patients with a broad range of defects. In common with my work, previous resequencing studies that have had parental DNA available have consistently shown inheritance of potentially significant variants from unaffected parents. This observation highlights the difficulty of moving from Mendelian inheritance of rare variants with very strong variant-phenotype concordance, to the study of variants anticipated to have only intermediate penetrance. Using the Sanger sequencing methodology available to me for this study, it would not have been possible to sequence sufficient cases and controls to arrive at a statistically significant result comparing mutation frequency between case and control groups. Indeed, even with next-generation sequencing technology, this is anticipated to present significant challenges in data interpretation. Assuming that these functionally significant mutations in genes important for heart development are accepted as susceptibility factors, these observations show that gene-gene and gene-environmental interaction are likely to be important in determining whether individuals who carry such mutations develop CVM. This is in keeping with the 'complex disease' model of CVM causation.

## 5.9 Common polymorphism and CVM

My work has focused on the impact of rare variation in candidate genes on CVM. However, during the work, several new common polymorphisms with potential relationships to CVM risk were discovered in both the cases and the controls. It is therefore appropriate to consider the potential contribution of such variants to CVM risk in greater detail.

### 5.9.1 Effect sizes of rare and common variants

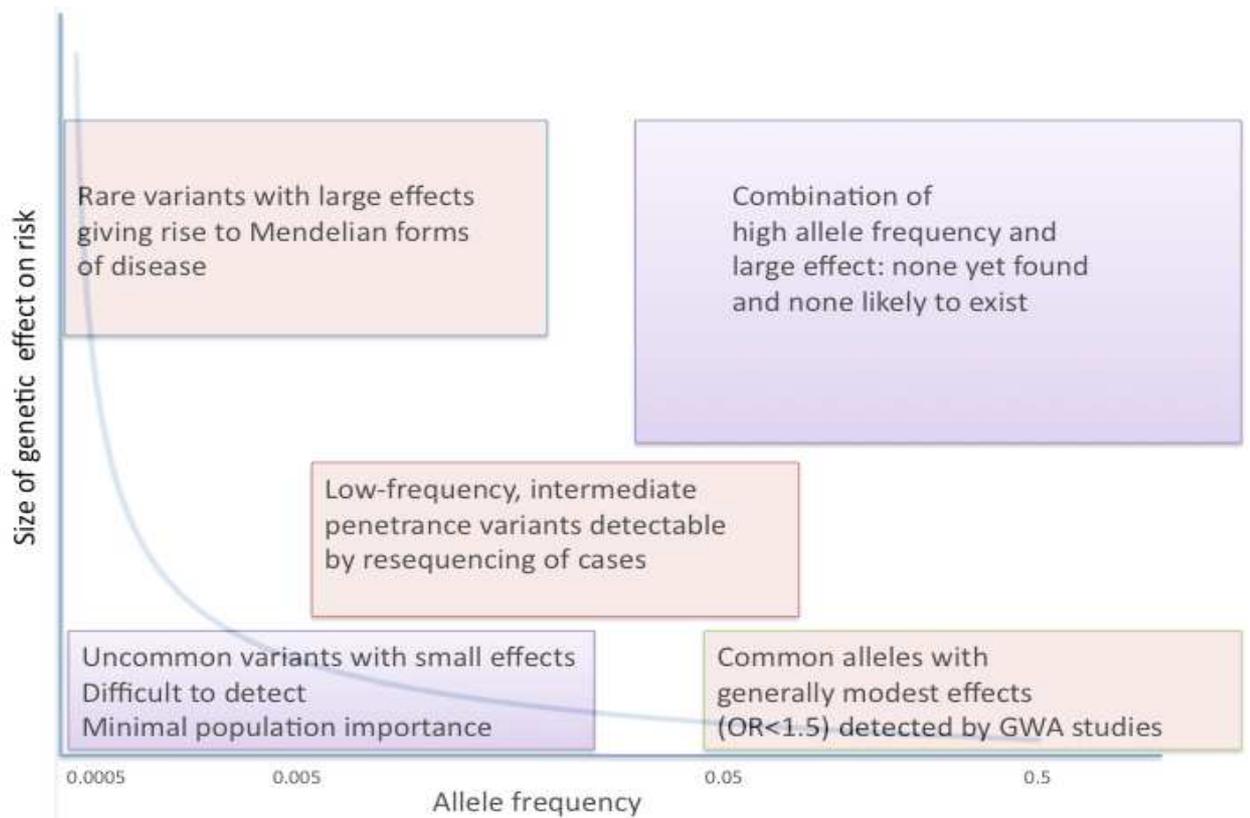
Definition of linkage disequilibrium relationships throughout the human genome in the HapMap project enabled the search for common variants that may cause susceptibility to CVM to study haplotype tagging SNPs in candidate genes. With the advent of array genotyping technology, the characterization of such common variants throughout the human genome using SNP arrays in large number of cases and controls to foster our understanding on the contribution of common variants to disease risk has been realized for many conditions e.g. coronary heart disease and diabetes. However, the architecture of the genetic component, i.e. the number of the effects and their sizes, that define disease susceptibility remains unclear in any condition. Population genetic theory suggests that alleles that are common in the population should have small biological effects, whereas large effects will be restricted to rare alleles. Figure 5.3 (adapted from Cunnington and Keavney, 2009) (Cunnington and Keavney 2009) shows the concept. The prediction has been demonstrated to be true for human and animal studies, apart from a few exceptional cases.

In Mendelian families (such as those with mutations in *NKX2.5*, *GATA4* and *NOTCH*), rare variants with substantially large effects on the function of a particular gene have been shown to be sufficient alone to confer increased risk for CVM. This type of variant is located in the upper left-hand portion of the curve in Figure 5.3. However, this type of variant is very hard to locate, as there are only a handful of families found with this type of phenotype. Also the contribution of such variants to the population attributing to risk of disease is low. Gene identification in Mendelian families is, however, an important avenue to pursue if suitable families are available. In addition to the new pathophysiological insight obtained, it is possible that rare non-private variants with

lesser effects on the gene in question may subsequently be found to contribute to disease risk in sporadic cases (as, for example, in the case of *NKX2.5*).

Common variants, on the contrary, would contribute significantly to the impact of CVM in the population but will have small effects on disease risk (lower right-hand portion of the curve). A study to characterize common variants will need large number of cases and controls in order to yield a better understanding of the contribution of these common variants to CVM risk.

The model of low frequency intermediate-penetrance (*LFIP*) variants (middle of the curve) was utilized in my work. The genome wide association studies could not capture these variants as they are not ‘tagged’ by common SNPs represented in genotyping chips, and the resequencing method is required to detect them. Common-SNP studies have failed to find variants that explain more than a small fraction of the heritability in many complex diseases. The search for the “missing heritability” is now largely focused as *LFIP* variants.



**Figure 5.3** The relationship between genetic effect size and allele frequency

### 5.9.2 Previous studies of common genetic variants and CVM

The most well known study of a common genetic variant and CVM is the association of the C677T SNP in *MTHFR* with CVM. The 5,10-methylenetetrahydrofolate reductase (MTHFR) plays a pivotal role in folate metabolism. This enzyme is necessary for metabolizing homocysteine into methionine by catalyzing the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate (5MeTHF), which in turn donates a methyl group to the intermediary metabolite homocysteine. The substitution of a cysteine (C) to thymidine (T) at position 677 has been shown to result in a reduction in the activity of the MTHFR enzyme, higher levels of homocysteine and lower levels of folate. The heterozygotes have around 65% of the wildtype enzyme activity but the homozygotes T allele have around 30% (Frosst et al. 1995). The effect of the C677T SNP on plasma levels of folate and homocysteine appears to be non-additive. Homozygous TT individuals have significantly higher homocysteine and lower folate than those with CC genotype. Homozygous CT individuals have levels that are similar to the CC genotype. The relationship between genotype and the plasma levels of folate and homocysteine is influenced by folate intake, such that the effect of the polymorphism is much more marked in conditions of folate deficiency. Folic acid has been strongly suspected to reduce the risk of CVM in offspring, but the evidence remains inconclusive. Multiple retrospective studies have suggested that folate supplementation reduces the risk of CVM, but such studies have substantial risk of confounding and/or recall bias. A Hungarian randomized trial on birth defects suggested a substantial effect of a multivitamin supplement containing 0.8mg of folic acid on CVM (OR 0.42, 95% CI 0.19-0.98) (Czeizel et al. 1998). It is not possible to definitively infer a causal role for folate, as opposed to other vitamins, from that study. Moreover, further randomized trials of folic acid to prevent CVM are not viable now as it is ethically unacceptable given the established association of neural tube defect and folate. In this context, genetic studies examining the association between the C677T SNP and the risk of CVM might shed light on whether or not plasma folate levels affect CVM risk. Such a study would use the “Mendelian randomization” approach which examines the congruence between associations of genotype with a hypothesized intermediate risk factor and with disease risk to assess causality of the risk factor (Keavney et al. 2006). Unfortunately, there are only relatively few studies published to date, and these are in quite small numbers of individuals. Two previous meta-analysis of this question have been negative, however, owing to the small numbers studied so far,

this remains an open question. Genotyping of the C677T SNP in large numbers of cases and controls, together with an updated meta-analysis of all data in the literature thus far, may yield a more conclusive result.

The work performed by Lambrechts and colleagues (Lambrechts et al. 2005) on the VEGF promoter haplotype using transmission disequilibrium testing in a cohort of approximately 250 patients identified a 1.8 fold increased risk for sporadic TOF. This promoter haplotype decreased the expression of VEGF. However, re-examination of this association in a larger number of cases and controls by Griffin et al. (Griffin et al. 2009), suggested no significant effect of the haplotype. This far, investigation of common genetic variation has been restricted to the analysis of one or a few candidate polymorphisms. However, the majority of the successes in other complex diseases that have been investigated for association with common genetic variants have come from genome wide association study, discussed in the following section.

### **5.9.3 GWAS studies of complex diseases**

Up till now, no genome-wide association studies (GWAS) on CVM have been performed. GWAS have been conducted in large numbers of unrelated cases of a wide variety of diseases and controls. The analysis compares the frequency of each of the genotypes at common SNPs throughout the genome between the cases and the controls. Among the best examples of GWAS of complex diseases is that of coronary heart disease (CHD). Four GWAS of CHD published in 2007 have shown association between common SNPs in chromosome 9p21 region (~100kb) and risk of CHD (Consortium 2007; Helgadottir et al. 2007; McPherson et al. 2007; Samani et al. 2007). These studies typed 100,000 or more SNPs in more than 1000 cases and more than 1000 controls. The positive results were confirmed in a further large number of cases and controls. Collectively, they found SNPs across >100kb in the 9p21 region associated with CHD with p values approximately  $1 \times 10^{-6}$  and odds ratio of 1.2-2.0. Therefore, using the rationale of previous studies on other complex diseases for investigation of CVM, one might hope to discover association of some common variants across the genome with CVM. The few variants that have been identified in this work, in particularly *BMP4* c.-171C>G, which was significantly different between cases and

controls with a p value of 0.00007, warrant further investigation by typing in replication cohorts of cases and controls.

#### **5.9.4 LFIP variants and complex diseases**

The work of Cohen, Hobbs and colleagues, taken in context with the results of recent GWAS studies of coronary heart disease, powerfully illustrates the fact that the genetic architecture of complex diseases can include both common and rare variants in the same genes (Cohen et al. 2006). It has invariably been observed that the rare variants have higher effects on diseases risk than do the common variants. Systematic sequencing of candidate genes in approximately 1000 cases of disease and an equal number of controls would give adequate power to identify (gene wise) individually rare variants which when taken together conferred odds ratios of less than 2. The recent developments in sequencing technology render these experiments potentially within reach, although they remain very expensive to perform and there are outstanding complex issues regarding analysis, particularly if whole-exome sequencing is used. Technical constraints did not permit such large scale sequencing of this project, as the technology to conduct these experiments has only recently become available. However, the results of the present study strongly suggest that the candidate genes that have been examined in this study would warrant further sequencing in larger numbers, particularly *SMAD6*.

## 5.10 Copy Number Variants (CNVs) and CVM

Another potential genetic mechanism that might predispose to CVM is copy number variation (CNV). CNV is defined as a DNA segment that is 1kb or longer with variable copy number in comparison with a reference genome. This submicroscopic copy number variation of DNA could range from kilobases to megabases involving tandem duplication or complex gains or losses of chunks of sequences in the genome. CNVs have been demonstrated to affect gene expression, phenotypic variation and adaptation (Feuk et al. 2006). CNVs contribute to human genetic and phenotypic diversity. Two models of CNV-phenotype associations have been suggested: one model involves common copy number polymorphisms (CNPs) whereas the other model involves rare CNVs that delete or duplicate larger genomic segments and appear in hemizygous or trisomic state. For example higher copy number of *CCL3LI* has been associated with lower HIV acquisition and progression (Gonzalez et al. 2005) (the first model), whereas 3 *de novo* CNV deletions were significantly associated with schizophrenia when checked in large numbers of cases and controls independently (Stefansson et al. 2008) (the latter model).

Before the SNP array technology to interrogate CNV was developed, comparative genomic hybridization (CGH) technology was utilized to characterize chromosomal abnormalities and subtle genomic aberrations (Barrett et al. 2004; Drazinic et al. 2005). De-Stefano and colleagues (De Stefano et al. 2008) investigated monozygotic twins with pulmonary atresia with intact ventricular septum (PA-IVS). PA-IVS is a rare CVM characterized by atretic pulmonary valve with the right ventricle varying in size and morphology among the patients. Using CGH technology, heterozygous 55-kb deletion on chromosome 20q13.12, which involves *WFDC8* and *WFDC9* genes, was identified in a pair of monozygotic twins. Further investigation in the family members revealed a *WFDC8* deletion in the mother and a *WFDC9* deletion in the father and an unaffected sibling. Thus, the twins possessed the maternal haplotype with a *WFDC8* deletion and the paternal haplotype with a *WFDC9* deletion. *WFDC8* and *WFDC9* genes encode a member of the WAP-type four-disulfide core (WFDC) domain family, which acts as protease inhibitors closely related to inherent immunity and inflammation (Clauss et al. 2005). De-Stefano et al. hypothesized that double haploinsufficiency of both genes in combination with environmental factors might be responsible for the CVM.

In another study, 5218 individuals with mental retardation, autism, or congenital anomalies were screened with CGH array and reaffirmed with custom oligonucleotide arrays for changes in CNVs (Mefford et al. 2008). Mefford and colleagues identified 25 individuals with recurrent 1.35Mb deletion in 1q21.1. Six of these individuals were diagnosed with CVMs. The microdeletion was *de novo* in 8 patients, inherited from a mildly affected parent in 3 patients, inherited from unaffected parent in 6 patients and of unknown inheritance in 9 patients. This deletion was absent in the controls, suggesting a degree of causative effect in the patient. However, the phenotypes observed in the patients showed considerable variability: mild to moderate mental retardation, microcephaly, cardiac abnormalities and cataracts. The wide range of phenotypic severity observed in these patients is similar to the 22q11 deletion, which is the most common microdeletion syndrome in humans, with an estimated prevalence of one in 4000 (Goodship et al. 1998; Kobrynski and Sullivan 2007). One of the features associated with the deletion is CVM. More than 70% of the affected individuals (but not all) diagnosed with the microdeletion had various CVMs (McDonald-McGinn et al. 1999). Phenotypic variability might arise due to variation in genetic background, epigenetic phenomena and environmental factor. As such, no concrete evidence could associate the 1q21.1 microdeletion to a specific phenotypic defect, in this case, CVM.

Recently, genome-wide CNV-detection was enabled using SNP chips (e.g. the Affymetrix Human Genome-Wide SNP Array 6.0) and Birdseye algorithm (Korn et al. 2008). A recent survey of 114 trios, consisting of a proband with non-syndromic Tetralogy of Fallot (TOF) and unaffected parents, identified 11 *de novo* CNVs at 10 unique loci that were absent or very rare in 2,265 controls (Greenway et al. 2009). Multiplex ligation-dependent probe amplification (MLPA) was used to validate the CNVs and to assess another second cohort of sporadic, non-syndromic TOF cases, in order to reaffirm the findings were replicable. In short, a total of 17 CNVs at 10 loci in 512 individuals with TOF that were absent or extremely rare in the controls were identified. Some of the loci encoded genes that have been previously implicated in cardiogenesis, such as *TBX1*, *NOTCH1* and *JAG1*, suggesting the need to assess gene dosage in mutation analyses of CVM associated genes. However, my study was unable to detect CNVs in my list of candidate genes as it was designed to investigate the sequence of the genes and not the dosage of the genes. Therefore, it would be interesting to carry out studies on CNVs in genes of the BMP signaling pathway. This

would provide fascinating insights into how CNVs in this pathway might influence the susceptibility to CVM.

## 5.11 Strengths and limitations of this work

Due to the time constraint and also the availability of technology at the start of my study, only five genes involved in BMP signaling were investigated in this thesis. Despite this limitation, the five genes were chosen with great care: ligands at the extracellular level (*BMP2* and *BMP4*), receptors at the membrane (*BMPRIA* and *BMPRII*) and SMADs at the intracellular level (*SMAD6*), encompassing each level of the pathway. These genes were also selected based on their functions in the BMP signaling pathway and defects exhibited in animal models in previous literature. Screening for genetic variants in the candidate genes was carried out on the exonic and exon-intron boundary regions, including the different isoforms of the genes. This measure had enabled a thorough interrogation of the coding regions for rare variants. There is the possibility that susceptibility factors for CVM might be within the promoter, intronic or downstream 3' regulatory regions of the candidate genes. However, variants in the coding region are hypothesized to have the greatest functional effects on phenotypes (Kryukov et al. 2007).

Animal models to illustrate how the *SMAD6* mutations resulted in CVMs were not investigated. However, luciferase assay was performed to show the functional significance of the mutations *in vitro*. The *SMAD6* knockout mouse generated by Galvin and colleagues (Galvin et al. 2000) demonstrated cardiac abnormalities as the patients with the *SMAD6* mutations. Integrating these results, we could conclude that mutation in *SMAD6* is pathogenic.

The panel of 90 probands was re-sequenced for genetic variants in the exons of the five candidate genes in the BMP signaling pathway. Previous studies on non-syndromic CVM demonstrated that the detectable mutations in genes that conferred risk factors for CVM occurred in about 2-3% of the cases: for example ~2% for *CITED2* (Sperling et al. 2005) to 3% for *NKX2.5* (McElhinney et al. 2003). Simple binomial probability suggests that if 2% of the cases with CVM were disease-associated mutations, re-sequencing of 90 probands would yield approximately 80% power to detect at least one such change. For this reason, sequencing more patients will have greater power to detect small percentage of patients with mutation in the gene of interest. However, initial investigation of the 90 probands was both a feasible number to start with and gave adequate power to detect mutation in the candidate genes.

In order to reduce phenotypic heterogeneity, non-syndromic CVM patients without other malformations were selected. The reduced phenotypic heterogeneity should have increased the power to detect variants responsible for non-syndromic CVM. Available parents of patients were collected to allow for further investigation of the rare variants detected in the patients with the aim to establish whether the variant might be a *de novo* mutation or the variant might possess reduced penetrance. The collection of probands was restricted to Caucasian individuals and the genotyping of the rare variants identified were performed in Caucasian population to avoid false positives due to genetic heterogeneity.

The broad phenotypic spectrum of the CVM of the patients collected is in tandem with the animal models with a single-gene defect that showed variable heart defects. Besides that, recruitment of patients with a particular defect would be slow and ideal number of patients would not be reached. The defects of the patients were documented in detail. Investigation of candidate genes on diverse phenotypic defects was performed in the hope that genotype-phenotype correlation of mutations would be observed.

With the high recruitment rate of CVM patients, a replication cohort was available for investigation for the rare variants that were identified in the first panel of 90 patients. This was especially important for the re-sequencing of *SMAD6* MH2 domain, as replication of the study was able to minimize false positive association and also to gain more power for genotype-phenotype correlation.

## **5.12 Future experiments**

### **5.12.1 Characterization of SMAD6 mechanisms**

The mutations identified in SMAD6 could be further investigated to illustrate the mechanism of the mutations in BMP signaling. There are four possibilities that might result in inactivation of inhibition: type I receptor-binding, competition with co-Smads, cooperation inhibition with Smurf1, and nuclear actions (as discussed in section 5.7). This would, in general, be amenable to investigation in cellular systems using transfection, immunoblotting, and luciferase assays.

### **5.12.2 Sequencing SMAD6 in larger samples**

It would be useful to recruit more patients and to sequence *SMAD6* in these patients. A larger cohort of patients might reveal more mutations in the gene and so accurately define the contribution of *SMAD6* in the pathogenesis of CVM. In addition, more controls could be genotyped for the rare variants identified and the known common variants. The increased number of patients and controls might enable the case-control comparisons to reach statistical significance, especially the common variants. It would be particularly interesting to resequence *SMAD6* in a large cohort of patients with malformation of the aortic valve.

### **5.12.3 Further candidate gene sequencing**

This study did not interrogate all of the genes in the BMP signaling pathway, but my findings strongly suggest that comprehensive investigation of all genes involved in BMP signaling be undertaken. It would be particularly interesting to sequence *SMAD7*, another inhibitor in the signaling pathway. *SMAD7* is not, however, specific to BMP signaling as it also inhibits TGF $\beta$  signaling, and it is for that reason that it was not included in this work. However, the demonstration of functional mutation in SMAD6, and the observation of Chen and colleagues in a SMAD7 MH2 domain knockout mouse, that showed majority of the mutant mice died *in utero* due to multiple defects in heart, exhibiting ventricular septal defect, non-compaction and outflow tract malformation, whereas the surviving adult mice displayed impaired cardiac functions and severe

arrhythmia (Chen et al. 2009a), published during the period of this work, strongly argue in favour of *SMAD7* as a candidate gene for CVM.

As genome-wide association studies require large samples and the common genetic variants interrogated explained a small fraction of disease susceptibility for any condition studied thus far, it has been hypothesized that LFIP variants account for a substantial proportion of the “missing heritability” not explained by common GWAS variants. Novel sequencing methods, collectively termed ‘next generation sequencing’ technology, which became available during the period of my study, will radically change the conduct of genotype/ phenotype association studies. Such technologies are able to sequence a human genome now less than 1 week on one instrument at a fraction of the cost of Sanger sequencing (Wheeler et al. 2008). Ng and colleagues developed a method to sequence all the protein-coding regions (‘exomes’) of the genome (Ng et al. 2009). The proof-of-concept experiment showed that this method is feasible and cost effective compared to whole-genome sequencing. They demonstrated that it was possible to detect previously identified causal genetic variants in individuals with Freeman-Sheldon syndrome (FSS), when compared with eight HapMap individuals from three populations. They then applied this concept to search for the gene causing a Mendelian disorder of unknown aetiology, Miller syndrome, in four affected individuals in three independent kindreds. Using a recessive model, they identified a single candidate gene, *DHODH*, that caused the syndrome (Ng et al. 2010). At the time of writing (Nov 2010) around 15 Mendelian conditions have been successfully solved using whole-exome sequencing. However, as noted above, the extension of this technique to non-Mendelian conditions has yet to yield novel definite genes, and the challenges of interpreting whole-exome data in the context of incomplete penetrance, genetic heterogeneity, and phenocopy remain very substantial.

#### **5.12.4 Common variants**

As noted above, genome-wide association studies have been successfully carried out to identify genes involved in common human diseases. The Wellcome Trust Case-Control Consortium (WTCCC) examined ~2,000 individuals for 7 major diseases (bipolar disorder, coronary artery disease, Crohn’s Disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes) and 3000 controls (Consortium 2007). They

successfully identified many new genes associated with the diseases. This study was possible due to a few factors: the information from the International HapMap resource facilitated the design and analysis of association studies, the availability of dense genotyping chips to probe hundreds of thousands of SNPs and most importantly, the assembly of large samples for each disease. In order to unravel the common variants predisposing to CVM using genome-wide association method, more CVM samples must be collected. The large sample size would require collaborations with other groups that were carrying out similar studies: well-characterized non-syndromic CVM patients of homogeneous ancestry. After genotyping, the massive number of statistical tests performed on the SNP data that have undergone stringent quality control procedures presents a high false positive rate if conventional statistical significance levels ( $p < 0.05$ ) are used. As a result, the P value of the study is set at approximately  $1.0 \times 10^{-8}$  or less. Besides that, an independent replication cohort of the same or larger size than the first set of samples is required for the SNPs selected to confirm the association. This project would enable us to identify the common variants across the genome that might be susceptible to CVM. The few variants that have been identified thus far with a  $p < 0.05$  would be an interesting starting point. Probing of more patients and controls for these variants will disseminate whether these variants reach statistical significance or not.

#### **5.12.5 CNV studies**

The study conducted by Greenway and colleagues (Greenway et al. 2009), identified de novo CNVs in patients with isolated sporadic tetralogy of Fallot. This study implies that rare CNV might be another interesting factor to investigate in CVM patients in the future. The idea is further strengthened by the recent published findings that showed common CNVs play very little role in contributing to the genetic basis of common human diseases (bipolar disorder, breast cancer, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes), which make a major impact on public health (Craddock et al. 2010). Therefore, rare or de novo CNV observed in CVM patients (if any) might be a causative factor, but common CNVs are perhaps a less promising avenue for investigation. CNVs could be investigated using Affymetrix arrays and validated using multiplex ligation-dependent probe amplification (MLPA) for the regions of interest. Here, the availability of trio data in our collection would be a significant strength.

## 5.13 Conclusions

This study and a study on *ALK2 (ACVRI)* (Smith et al. 2009) showed that mutation in the BMP signaling pathway plays an important part in cardiac malformation in man. CVM patients with rare non-synonymous variants located at important domains have been shown to abrogate or elevate BMP signaling showing that rare non-synonymous variants in the genes of the signaling pathway contribute to the predisposition to CVM. Further investigation of the genes involving in BMP signaling using next-generation sequencing technology is warranted.

Besides that, this study could have implications for clinical practice. Two functionally significant SMAD6 variants seen in two probands having phenotypes involving the aorta were identified out of 46 patients with coarctation or aortic stenosis defects. Therefore, the prevalence of non-synonymous SMAD6 variants in this category of CVM was around 4%. However, the prevalence of these variants was lower (~0.46%) when all of the 438 CVM patients sequenced for MH2 domain of SMAD6 were accounted. This percentage is quite similar to the study that was carried out by Sperling and colleagues (Sperling et al. 2005) where three functionally significant ins/del mutations in serine-glycine rich junction of CITED2 protein were identified out of 392 CVM patients with a broad phenotype spectrum, approximately 0.77%. If this pattern of predisposition of functionally significant variants to CVM was repeated throughout the human genome, taking into account the incomplete penetrance of such variants, it would be challenging to screen CVM patients and to provide genetic counseling to them regarding the recurrence risks to their offspring.

Cardiac development is a complex biological process requiring intrinsic and correct spatial temporal expression of many factors. Disruption of a number of signaling pathways, including BMP, transforming growth factor- $\beta$ , Wnt, Notch signaling have been implicated in cardiac development *in vivo* or *in vitro*. Studies that have demonstrated the cross-talk between BMPs and other signaling pathways, including certain transcription factors further confirmed the multifunctional role of BMPs in cardiac development (Chen et al. 2009b; Nakashima et al. 2005; Schwappacher et al. 2009; Takizawa et al. 2003). Mutational investigation of candidate genes carried out by various groups has shed some lights on cardiac development mechanism in man, paving the ground for more accurate genetic counselling and new approaches for prevention of

CVM in those who are susceptible. This is in keeping with the advancement of personalized genomic medicine, which aims to predict an individual's relative risk of developing a disease or transmitting a disease to the offspring based on the person's genome and to offer specific treatment for each person depending on how he or she might respond to drugs. The best example is the mutations in *BRCA* genes that are associated with breast cancer risk, which have helped to predict more accurately the risk prediction in some patients (Robson and Offit 2007). The era of personalized genomics is potentially within reach with the advent of next generation sequencing to sequence a human genome in search for an individual's risk of disease susceptibility and to personalize treatment; this project suggests that patients and families with CVM could be among those to benefit from this revolutionary change in healthcare practice.

## Chapter 6      References

- Aebi M, Hornig H, Weissmann C (1987) 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* 50: 237-46
- Allen SP, Bogardi J, Barlow AJ (2001) Misexpression of noggin leads to septal defects in the outflow tract of the chick heart. *Dev. Biol.* 235: 98-109
- Anderson RH (2003) Understanding the structure of the unicuspid and unicommissural aortic valve. *J Heart Valve Dis* 12: 670-3
- Andree B, Duprez D, Vorbusch B, Arnold HH, Brand T (1998) BMP-2 induces ectopic expression of cardiac lineage markers and interferes with somite formation in chicken embryos. *Mech Dev* 70: 119-31
- Arguello C, de la Cruz MV, Gomez CS (1975) Experimental study of the formation of the heart tube in the chick embryo. *J Embryol Exp Morphol* 33: 1-11
- Armstrong EJ, Bischoff J (2004) Heart valve development: Endothelial cell signaling and differentiation. *Circ. Res.* 95: 459-470
- Bai S, Shi X, Yang X, Cao X (2000) Smad6 as a transcriptional corepressor. *J Biol Chem* 275: 8267-70
- Bailey LB, Berry RJ (2005) Folic acid supplementation and the occurrence of congenital heart defects, orofacial clefts, multiple births, and miscarriage. *Am J Clin Nutr* 81: 1213S-1217S
- Bakrania P, Efthymiou M, Klein JC, Salt A, Bunyan DJ, Wyatt A, Ponting CP, Martin A, Williams S, Lindley V, Gilmore J, Restori M, Robson AG, Neveu MM, Holder GE, Collin JR, Robinson DO, Farndon P, Johansen-Berg H, Gerrelli D, Ragge NK (2008) Mutations in BMP4 cause eye, brain, and digit developmental anomalies: overlap between the BMP4 and hedgehog signaling pathways. *Am J Hum Genet* 82: 304-19
- Balemans W, Van Hul W (2002) Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev Biol* 250: 231-50
- Bamforth SD, Braganca J, Farthing CR, Schneider JE, Broadbent C, Michell AC, Clarke K, Neubauer S, Norris D, Brown NA, Anderson RH, Bhattacharya S (2004) Cited2 controls left-right patterning and heart development through a Nodal-Pitx2c pathway. *Nat Genet* 36: 1189-96
- Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence changes. *J Med Genet* 42: 737-48

- Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, Yakhini Z, Bruhn L, Laderman S (2004) Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* 101: 17765-70
- Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soultis J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J, Renault B, Kucherlapati R, Seidman JG, Seidman CE (1997) Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* 15: 30-5
- Ben-Shachar G, Arcilla RA, Lucas RV, Manasek FJ (1985) Ventricular trabeculations in the chick embryo heart and their contribution to ventricular and muscular septal development. *Circ Res* 57: 759-66
- Beppu H, Kawabata M, Hamamoto T, Chytil A, Minowa O, Noda T, Miyazono K (2000) BMP type II receptor is required for gastrulation and early development of mouse embryos. *Dev. Biol.* 221: 249-258
- Berget SM (1995) Exon recognition in vertebrate splicing. *J Biol Chem* 270: 2411-4
- Bernstein E (1996) The cardiovascular system. In Behrman, R.E, Kliegman, R.M., Arvin, A.M. (eds): *Nelson Textbook of Pediatrics* 15th ed.
- Bevan S, Woodford-Richens K, Rozen P, Eng C, Young J, Dunlop M, Neale K, Phillips R, Markie D, Rodriguez-Bigas M, Leggett B, Sheridan E, Hodgson S, Iwama T, Eccles D, Bodmer W, Houlston R, Tomlinson I (1999) Screening SMAD1, SMAD2, SMAD3, and SMAD5 for germline mutations in juvenile polyposis syndrome. *Gut* 45: 406-8
- Bhattacharya S, Michels CL, Leung MK, Arany ZP, Kung AL, Livingston DM (1999) Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. *Genes Dev* 13: 64-75
- Blom HJ, Shaw GM, den Heijer M, Finnell RH (2006) Neural tube defects and folate: case far from closed. *Nat Rev Neurosci* 7: 724-31
- Botto LD, Lynberg MC, Erickson JD (2001) Congenital heart defects, maternal febrile illness, and multivitamin use: a population-based study. *Epidemiology* 12: 485-90
- Botto LD, Mulinare J, Erickson JD (2000) Occurrence of congenital heart defects in relation to maternal multivitamin use. *Am J Epidemiol* 151: 878-84
- Brand T (2003) Heart development: molecular insights into cardiac specification and early morphogenesis. *Dev Biol* 258: 1-19

- Braunwald E (1997) Heart Disease - a textbook of cardiovascular medicine. 5th edition.  
Saunders
- Breckenridge RA, Mohun TJ, Amaya E (2001) A role for BMP signalling in heart looping morphogenesis in *Xenopus*. *Dev Biol* 232: 191-203
- Brizzard BL, Chubet RG, Vizard DL (1994) Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution. *Biotechniques* 16: 730-5
- Buckingham M, Meilhac S, Zaffran S (2005) Building the mammalian heart from two sources of myocardial cells. *Nature Reviews Genetics*: 826-835
- Burn J, Brennan P, J. L, Holloway S (1998) Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *The Lancet* 351: 311-316
- Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S (2003) *Isl1* identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* 5: 877-89
- Carlson BM (2004) Human embryology and developmental biology (3rd edition).  
Mosby
- Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3: 285-98
- Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31: 3568-71
- Chacko BM, Qin BY, Tiwari A, Shi G, Lam S, Hayward LJ, De Caestecker M, Lin K (2004) Structural basis of heteromeric smad protein assembly in TGF-beta signaling. *Mol Cell* 15: 813-23
- Chen Q, Chen H, Zheng D, Kuang C, Fang H, Zou B, Zhu W, Bu G, Jin T, Wang Z, Zhang X, Chen J, Field LJ, Rubart M, Shou W, Chen Y (2009a) Smad7 is required for the development and function of the heart. *J Biol Chem* 284: 292-300
- Chen YG, Hata A, Lo RS, Wotton D, Shi Y, Pavletich N, Massague J (1998) Determinants of specificity in TGF-beta signal transduction. *Genes Dev* 12: 2144-52
- Chen YL, Liu B, Zhou ZN, Hu RY, Fei C, Xie ZH, Ding X (2009b) Smad6 inhibits the transcriptional activity of Tbx6 by mediating its degradation. *J Biol Chem* 284: 23481-90

- Chocron S, Verhoeven MC, Rentzsch F (2007) Zebrafish Bmp4 regulates left-right asymmetry at two distinct developmental time points. *Dev. Biol.* 305: 577-588
- Christianson A, Howson CP, Modell B (2006) *Global Report on Birth Defects: The Hidden Toll of Dying and Disabled Children*. White Plains, NY: March of Dimes Birth Defects Foundation
- Christoffels VM, Habets PE, Franco D, Campione M, de Jong F, Lamers WH, Bao ZZ, Palmer S, Biben C, Harvey RP, Moorman AF (2000) Chamber formation and morphogenesis in the developing mammalian heart. *Dev Biol* 223: 266-78
- Clauss A, Lilja H, Lundwall A (2005) The evolution of a genetic locus encoding small serine proteinase inhibitors. *Biochem Biophys Res Commun* 333: 383-9
- Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 354: 1264-72
- Colas JF, Lawson A, Schoenwolf GC (2000) Evidence that translation of smooth muscle alpha-actin mRNA is delayed in the chick promyocardium until fusion of the bilateral heart-forming regions. *Dev Dyn* 218: 316-30
- Collins JS, Schwartz CE (2002) Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet* 71: 1251-2
- Consortium WTCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-78
- Constam DB, Robertson EJ (1999) Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases. *J Cell Biol* 144: 139-49
- Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczak K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P,

- Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464: 713-20
- Cunnington MS, Keavney BD (2009) Genetics of coronary heart disease In Evidence-Based Cardiology. BMJ Books
- Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B (2010) Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* 6: e1000899
- Czeizel AE, Dobo M, Dudas I, Gasztonyi Z, Lantos I (1998) The Hungarian periconceptual service as a model for community genetics. *Community Genet* 1: 252-9
- Dasgupta C, Martinez AM, Zuppan CW, Shah MM, Bailey LL, Fletcher WH (2001) Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). *Mutat Res* 479: 173-86
- Dathe K, Kjaer KW, Brehm A, Meinecke P, Nurnberg P, Neto JC, Brunoni D, Tommerup N, Ott CE, Klopocki E, Seemann P, Mundlos S (2009) Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am J Hum Genet* 84: 483-92
- de Caestecker M (2004) The transforming growth factor-beta superfamily of receptors. *Cytokine Growth Factor Rev* 15: 1-11
- de la Cruz MV, Markwald RR (1998) Living Morphogenesis of the Heart. Boston: Birkhauser
- de Lange FJ, Moorman AF, Anderson RH, Manner J, Soufan AT, de Gier-de Vries C, Schneider MD, Webb S, van den Hoff MJ, Christoffels VM (2004) Lineage and morphogenetic analysis of the cardiac valves. *Circ Res* 95: 645-54
- De Stefano D, Li P, Xiang B, Hui P, Zambrano E (2008) Pulmonary atresia with intact ventricular septum (PA-IVS) in monozygotic twins. *Am J Med Genet A* 146A: 525-8
- Delot EC, Bahamonde ME, Zhao M, Lyons KM (2003) BMP signaling is required for septation of the outflow tract of the mammalian heart. *Development* 130: 209-220
- den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15: 7-12

- Desgrosellier JS, Mundell NA, McDonnell MA, Moses HL, Barnett JV (2005) Activin receptor-like kinase 2 and Smad6 regulate epithelial-mesenchymal transformation during cardiac valve formation. *Developmental Biology* 280: 201-210
- Dewulf N, Verschueren K, Lonnoy O, Moren A, Grimsby S, Vande Spiegle K, Miyazono K, Huylebroeck D, Ten Dijke P (1995) Distinct spatial and temporal expression patterns of two type I receptors for bone morphogenetic proteins during mouse embryogenesis. *Endocrinology* 136: 2652-63
- Digilio MC, Angioni A, De Santis M, Lombardo A, Giannotti A, Dallapiccola B, Marino B (2003) Spectrum of clinical variability in familial deletion 22q11.2: from full manifestation to extremely mild clinical anomalies. *Clin Genet* 63: 308-13
- Digilio MC, Conti E, Sarkozy A, Mingarelli R, Dottorini T, Marino B, Pizzuti A, Dallapiccola B (2002) Grouping of multiple-lentiginos/LEOPARD and Noonan syndromes on the PTPN11 gene. *Am J Hum Genet* 71: 389-94
- Dorak MT, Mackay RK, Relton CL, Worwood M, Parker L, Hall AG (2009) Hereditary hemochromatosis gene (HFE) variants are associated with birth weight and childhood leukemia risk. *Pediatr Blood Cancer* 53: 1242-8
- Drazinic CM, Ercan-Sencicek AG, Gault LM, Hisama FM, Qumsiyeh MB, Nowak NJ, Cubells JF, State MW (2005) Rapid array-based genomic characterization of a subtle structural abnormality: a patient with psychosis and der(18)t(5;18)(p14.1;p11.23). *Am J Med Genet A* 134: 282-9
- Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. *Nature* 467: 1061-73
- Ehrman LA, Yutzey KE (1999) Lack of regulation in the heart forming region of avian embryos. *Dev Biol* 207: 163-75
- Ericson A, Kallen BA (2001) Nonsteroidal anti-inflammatory drugs in early pregnancy. *Reprod Toxicol* 15: 371-5
- Eronen M, Peippo M, Hiippala A, Raatikka M, Arvio M, Johansson R, Kahkonen M (2002) Cardiovascular manifestations in 75 patients with Williams syndrome. *J Med Genet* 39: 554-8
- Feng XH, Derynck R (1997) A kinase subdomain of transforming growth factor-beta (TGF-beta) type I receptor determines the TGF-beta intracellular signaling specificity. *EMBO J* 16: 3912-23

- Ferencz C, Loffredo CA, Correa-Villasenor A, Wilson PD (1997) Genetic and environmental risk factors of major cardiovascular malformations: the Baltimore-Washington Infant Study, 1981-1989. *Mt. Kisko*
- Fernandez-Guerra P, Navarrete R, Weisiger K, Desviat LR, Packman S, Ugarte M, Rodriguez-Pombo P (2010) Functional characterization of the novel intronic nucleotide change c.288+9C>T within the BCKDHA gene: understanding a variant presentation of maple syrup urine disease. *J Inherit Metab Dis*
- Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7: 85-97
- Franco D, Campione M, Kelly R, Zammit PS, Buckingham M, Lamers WH, Moorman AF (2000) Multiple transcriptional domains, with distinct left and right components, in the atrial chambers of the developing heart. *Circ Res* 87: 984-91
- Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP, et al. (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 10: 111-3
- Fujii M, Takeda K, Imamura T, Aoki H, Sampathw TK, Enomoto S, Kawabata M, Kato M, Ichijo H, Miyazono K (1999) Roles of Bone Morphogenetic Protein Type I Receptros and Smad Proteins in Osteoblast and Chondroblast Differentiation. *Molecular Biology of the Cell* 10: 3801-3813
- Fukushima T, Mashiko M, Takita K, Otake T, Endo Y, Sekikawa K, Takenoshita S (2003) Mutational analysis of TGF-beta type II receptor, Smad2, Smad3, Smad4, Smad6 and Smad7 genes in colorectal cancer. *J Exp Clin Cancer Res* 22: 315-20
- Galvin KM, Donovan MJ, Lynch CA, Meyer RI, Paul RJ, Lorenz JN, Fairchild-Huntress V, Dixon KL, Dunmore JH, Gimbrone Jr MA, Falb D, Huszar D (2000) A role of Smad6 in development and homeostasis of the cardiovascular system. *Nature genetics* 24: 171-174
- Gao K, Masuda A, Matsuura T, Ohno K (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 36: 2257-67
- 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 (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424: 443-7

- Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D (2005) Mutations in NOTCH1 cause aortic valve disease. *Nature* 437: 270-4
- Gaussin V, Morley GE, Cox L, Zwijsen A, Vance KM, Emile L, Tian Y, Liu J, Hong C, Myers D, Conway S, Jea (2005) Alk3/Bmpr1a receptor is required for development of the atrioventricular canal into valves and annulus fibrosus. *Circulation Research* 97: 219-226
- Gaussin V, Van de Putte T, Mishina Y, Hanks MC, Zwijsen A, Huylebroeck D, Behringer RR, Schneider MD (2002) Endocardial cushion and myocardial defects after cardiac myocyte-specific conditional deletion of the bone morphogenetic protein receptor ALK3. *Proc Natl Acad Sci U S A* 99: 2878-83
- Ghosh-Choudhury N, Abboud SL, Chandrasekar B, Ghosh Choudhury G (2003) BMP-2 regulates cardiomyocyte contractility in a phosphatidylinositol 3 kinase-dependent manner. *FEBS Lett* 544: 181-4
- Goldmuntz E, Clark BJ, Mitchell LE, Jawad AF, Cuneo BF, Reed L, McDonald-McGinn D, Chien P, Feuer J, Zackai EH, Emanuel BS, Driscoll DA (1998) Frequency of 22q11 deletions in patients with conotruncal defects. *J Am Coll Cardiol* 32: 492-8
- Goldmuntz E, Geiger E, Benson DW (2001) Nkx2.5 mutations in patients with tetralogy of fallot. *Circ. Res.* 104: 2545-2568
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell R J, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307: 1434-40
- Goodship J, Cross I, LiLing J, Wren C (1998) A population study of chromosome 22q11 deletions in infancy. *Arch Dis Child* 79: 348-51
- Goto K, Kamiya Y, Imamura T, Miyazono K, Miyazawa K (2007) Selective Inhibitory Effects of Smad6 on Bone Morphogenetic Protein Type I Receptors. *The Journal of Biological Chemistry* 282: 20603-20611
- Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, Ergul E, Conta JH, Korn JM, McCarroll SA, Gorham JM, Gabriel S, Altshuler DM, Quintanilla-Dieck Mde L, Artunduaga MA, Eavey RD, Plenge RM, Shadick NA, Weinblatt ME, De Jager PL, Hafler DA, Breitbart RE, Seidman JG, Seidman CE

- (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet* 41: 931-5
- Griffin HR, Hall DH, Topf A, Eden J, Stuart AG, Parsons J, Peart I, Deanfield JE, O'Sullivan J, Babu-Narayan SV, Gatzoulis MA, Bu'lock FA, Bhattacharya S, Bentham J, Farrall M, Granados Riveron J, Brook JD, Burn J, Cordell HJ, Goodship JA, Keavney B (2009) Genetic variation in VEGF does not contribute significantly to the risk of congenital cardiovascular malformation. *PLoS One* 4: e4978
- Hanyu A, Ishidou Y, Ebisawa T, Shimanuki T, Imamura T, Miyazono K (2001) The N domain of Smad7 is essential for specific inhibition of transforming growth factor-beta signaling. *J Cell Biol* 155: 1017-27
- Hata A, Lagna G, Massague J, Hemmati-Brivanlou A (1998) Smad6 inhibits BMP/Smad1 signaling by specifically competing with the Smad4 tumor suppressor. *Genes Dev* 12: 186-97
- Heldin CH, Miyazono K, ten Dijke P (1997) TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390: 465-71
- 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 (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 316: 1491-3
- Hoffman JI, Kaplan S (2002) The incidence of congenital heart disease. *J Am Coll Cardiol* 39: 1890-1900
- Hoffman JI, Kaplan S, Liberthson RR (2004) Prevalence of congenital heart disease. *Am Heart J* 147: 425-39
- Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G, Vogelstein B (2001) Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet* 28: 184-7
- Howe JR, Roth S, Ringold JC, Summers RW, Jarvinen HJ, Sistonen P, Tomlinson IP, Houlston RS, Bevan S, Mitros FA, Stone EM, Aaltonen LA (1998) Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science* 280: 1086-8

- Howe JR, Sayed MG, Ahmed AF, Ringold J, Larsen-Haidle J, Merg A, Mitros FA, Vaccaro CA, Petersen GM, Giardiello FM, Tinley ST, Aaltonen LA, Lynch HT (2004) The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. *J Med Genet* 41: 484-91
- Hurle JM, Colvee E, Blanco AM (1980) Development of mouse semilunar valves. *Anat Embryol (Berl)* 160: 83-91
- Huse M, Chen YG, Massague J, Kuriyan J (1999) Crystal structure of the cytoplasmic domain of the type I TGF beta receptor in complex with FKBP12. *Cell* 96: 425-36
- Hutson MR, Kirby ML (2007) Model systems for the study of heart development and disease. Cardiac neural crest and conotruncal malformations. *Semin Cell Dev Biol* 18: 101-10
- Iaizzo PA (2009) Handbook of cardiac anatomy, physiology and devices. Springer Sciences + Business Media
- Imamura T, Takase M, Nishihara A, Oeda E, Hanai J, Kawabata M, Miyazono K (1997) Smad6 inhibits signaling by the TGF-B superfamily. *Nature* 389: 622-626
- Ionescu-Ittu R, Marelli AJ, Mackie AS, Pilote L (2009) Prevalence of severe congenital heart disease after folic acid fortification of grain products: time trend analysis in Quebec, Canada. *BMJ* 338: b1673
- Jenkins KJ, Correa A, Feinstein JA, Botto L, Britt AE, Daniels SR, Elixson M, Warnes CA, Webb CL (2007) Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 115: 2995-3014
- Jiao K, Kulesa H, Tompkins K, Zhou Y, Batts L, Baldwin HS, Hogan BL (2003) An essential role of Bmp4 in the atrioventricular septation of the mouse heart. *Genes Dev* 17: 2362-7
- Jonson T, Gorunova L, Dawiskiba S, Andren-Sandberg A, Stenman G, ten Dijke P, Johansson B, Hoglund M (1999) Molecular analyses of the 15q and 18q SMAD genes in pancreatic cancer. *Genes Chromosomes Cancer* 24: 62-71
- Kameda Y (2009) Hoxa3 and signaling molecules involved in aortic arch patterning and remodeling. *Cell Tissue Res* 336: 165-78

- Kasahara H, Lee B, Schott JJ, Benson DW, Seidman JG, Seidman CE, Izumo S (2000) Loss of function and inhibitory effects of human CSX/NKX2.5 homeoprotein mutations associated with congenital heart disease. *J Clin Invest* 106: 299-308
- Katagiri T, Yamaguchi A, Komaki M, Abe E, Takahashi N, Ikeda T, Rosen V, Wozney JM, Fujisawa-Sehara A, Suda T (1994) Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage. *J. Cell Biol.* 127: 1755-1766
- Kawabata M, Chytil A, Moses HL (1995) Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-beta receptor. *J Biol Chem* 270: 5625-30
- Kawabata M, Imamura T, Miyazono K (1998) Signal transduction by bone morphogenetic proteins. *Cytokine Growth Factor Rev* 9: 49-61
- Kawate S, Ohwada S, Hamada K, Koyama T, Takenoshita S, Morishita Y, Hagiwara K (2001) Mutational analysis of the Smad6 and Smad7 genes in hepatocellular carcinoma. *Int J Mol Med* 8: 49-52
- Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R (2006) Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *Int J Epidemiol* 35: 935-43
- Kelly RG, Brown NA, Buckingham ME (2001) The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev Cell* 1: 435-40
- Kereszturi E, Kiraly O, Sahin-Toth M (2009) Minigene analysis of intronic variants in common SPINK1 haplotypes associated with chronic pancreatitis. *Gut* 58: 545-9
- Kinbara K, Goldfinger LE, Hansen M, Chou FL, Ginsberg MH (2003) Ras GTPases: integrins' friends or foes? *Nat Rev Mol Cell Biol* 4: 767-76
- Kirby ML (2007) *Cardiac Development*. Oxford University Press
- Kirby ML, Gale TF, Stewart DE (1983) Neural crest cells contribute to normal aorticopulmonary septation. *Science* 220: 1059-61
- Kirby ML, Stewart DE (1983) Neural crest origin of cardiac ganglion cells in the chick embryo: identification and extirpation. *Dev Biol* 97: 433-43
- Kirby ML, Turnage KL, 3rd, Hays BM (1985) Characterization of conotruncal malformations following ablation of "cardiac" neural crest. *Anat Rec* 213: 87-93
- Kishimoto Y, Lee KH, Zon L, Hammerschmidt M, Schulte-Merker S (1997) The molecular nature of zebrafish swirl: BMP2 function is essential during early dorsoventral patterning. *Development* 124: 4457-66

- Klaus A, Saga Y, Taketo MM, Tzahor E, Birchmeier W (2007) Distinct roles of Wnt/beta-catenin and Bmp signaling during early cardiogenesis. *Proc Natl Acad Sci U S A* 104: 18531-6
- Kobrynski LJ, Sullivan KE (2007) Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* 370: 1443-52
- Koenig BB, Cook JS, Wolsing DH, Ting J, Tiesman JP, Correa PE, Olson CA, Pecquet AL, Ventura F, Grant RA (1994) Characterization and cloning of a receptor for BMP-2 and BMP-4 from NIH 3T3 cells. *Molecular and Cellular Biology* 14: 5961-5974
- Kol G, Lev-Maor G, Ast G (2005) Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum Mol Genet* 14: 1559-68
- Kolodziej PA, Young RA (1991) Epitope tagging and protein surveillance. *Methods Enzymol* 194: 508-19
- Korchynskiy O, ten Dijke P (2001) Identification and Functional Characterization of Distinct Critically Important Bone Morphogenetic Protein-specific Response Elements in the Id1 Promoter. *The Journal of Biological Chemistry* 277: 4883-4891
- Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 40: 1253-60
- Krantz ID, Rand EB, Genin A, Hunt P, Jones M, Louis AA, Graham JM, Jr., Bhatt S, Piccoli DA, Spinner NB (1997) Deletions of 20p12 in Alagille syndrome: frequency and molecular characterization. *Am J Med Genet* 70: 80-6
- Kryukov GV, Pennacchio LA, Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80: 727-39
- Lambrechts D, Devriendt K, Driscoll DA, Goldmuntz E, Gweillig M, Vlietinck R, Collen D, Carmeliet P (2005) Low expression VEGF haplotype increases the risk for tetralogy of Fallot: a family based association study. *J Med Genet* 42: 519-522
- Levin M (2005) Left-right asymmetry in embryonic development: a comprehensive review. *Mech Dev* 122: 3-25

- Levy HL, Guldberg P, Guttler F, Hanley WB, Matalon R, Rouse BM, Trefz F, Azen C, Allred EN, de la Cruz F, Koch R (2001) Congenital heart disease in maternal phenylketonuria: report from the Maternal PKU Collaborative Study. *Pediatr Res* 49: 636-42
- Li L, Krantz ID, Deng Y, Genin A, Banta AB, Collins CC, Qi M, Trask BJ, Kuo WL, Cochran J, Costa T, Pierpont ME, Rand EB, Piccoli DA, Hood L, Spinner NB (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* 16: 243-51
- Lie RT, Wilcox AJ, Skjaerven R (1994) A population-based study of the risk of recurrence of birth defects. *N Engl J Med* 331: 1-4
- Lincoln J, Alfieri CM, Yutzey KE (2006) BMP and FGF regulatory pathways control cell lineage diversification of heart valve precursor cells. *Dev Biol* 292: 292-302
- Liu F, Ventura F, Doody J, Massague J (1995) Human type II receptor for bone morphogenetic proteins (BMPs): extension of the two-kinase receptor model to the BMPs. *Molecular and Cellular Biology* 15: 3479-3486
- Liu W, Selever J, Wang D, Lu MF, Moses KA, Schwartz RJ, Martin JF (2004) Bmp4 signaling is required for outflow-tract septation and branchial-arch artery remodeling. *Dev. Biol.* 101
- Liu Y, Bourgeois CF, Pang S, Kudla M, Dreumont N, Kister L, Sun YH, Stevenin J, Elliott DJ (2009) The germ cell nuclear proteins hnRNP G-T and RBMY activate a testis-specific exon. *PLoS Genet* 5: e1000707
- Lo RS, Chen YG, Shi Y, Pavletich NP, Massague J (1998) The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-beta receptors. *EMBO J* 17: 996-1005
- Loffredo CA, Wilson PD, Ferencz C (2001) Maternal diabetes: an independent risk factor for major cardiovascular malformations with increased mortality of affected infants. *Teratology* 64: 98-106
- Lopez-Sanchez C, Garcia-Martinez V, Schoenwolf GC (2001) Localization of cells of the prospective neural plate, heart and somites within the primitive streak and epiblast of avian embryos at intermediate primitive-streak stages. *Cells Tissues Organs* 169: 334-46
- Lough J, Sugi Y (2000) Endoderm and heart development. *Dev Dyn* 217: 327-42
- Lowe LA, Yamada S, Kuehn MR (2001) Genetic dissection of nodal function in patterning the mouse embryo. *Development* 128: 1831-43

- Lyons KM, Pelton RW, Hogan BL (1990) Organogenesis and pattern formation in the mouse: RNA distribution patterns suggest a role for bone morphogenetic protein-2A (BMP-2A). *Development* 109: 833-44
- Ma L, Lu M, Schwartz RJ, Martin JF (2005) Bmp2 is essential for cardiac cushion epithelial-mesenchymal transition and myocardial patterning. *Development* 132: 5601-5611
- Machado RD, Aldred MA, James V, Harrison RE, Patel B, Schwalbe EC, Gruenig E, Janssen B, Koehler R, Seeger W, Eickelberg O, Olschewski H, Elliott CG, Glissmeyer E, Carlquist J, Kim M, Torbicki A, Fijalkowska A, Szewczyk G, Parma J, Abramowicz MJ, Galie N, Morisaki H, Kyotani S, Nakanishi N, Morisaki T, Humbert M, Simonneau G, Sitbon O, Soubrier F, Coulet F, Morrell NW, Trembath RC (2006) Mutations of the TGF-beta type II receptor BMPR2 in pulmonary arterial hypertension. *Hum Mutat* 27: 121-32
- Machado RD, Pauciulo MW, Thomson JR, Lane KB, Morgan NV, Wheeler L, Phillips JAI, Newman J, Williams D, Nazzareno G, Manes A, McNeil K, Yacoub M, Mikhail G, Rogers P, Corris P, Humbert M, Donnai D, Martensson G, Tranebjaerg L, Loyd JE, Trembath RC, Nichols WC (2001) BMPR2 haploinsufficiency as the inherited molecular mechanism for primary pulmonary hypertension. *Am. J. Hum. Genet.* 68: 92-102
- Manner J (2000) Cardiac looping in the chick embryo: a morphological review with special reference to terminological and biomechanical aspects of the looping process. *Anat Rec* 259: 248-62
- Marino B, Digilio MC, Toscano A, Anaclerio S, Giannotti A, Feltri C, de Ioris MA, Angioni A, Dallapiccola B (2001) Anatomic patterns of conotruncal defects associated with deletion 22q11. *Genet Med* 3: 45-8
- Marino B, Digilio MC, Toscano A, Giannotti A, Dallapiccola B (1999) Congenital heart diseases in children with Noonan syndrome: An expanded cardiac spectrum with high prevalence of atrioventricular canal. *J Pediatr* 135: 703-6
- Markwald RR, Fitzharris TP, Manasek FJ (1977) Structural development of endocardial cushions. *Am J Anat* 148: 85-119
- McDonald-McGinn DM, Kirschner R, Goldmuntz E, Sullivan K, Eicher P, Gerdes M, Moss E, Solot C, Wang P, Jacobs I, Handler S, Knightly C, Heher K, Wilson M, Ming JE, Grace K, Driscoll D, Pasquariello P, Randall P, Larossa D, Emanuel BS, Zackai EH (1999) The Philadelphia story: the 22q11.2 deletion: report on 250 patients. *Genet Couns* 10: 11-24

- McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E (2003) NKX2.5 mutations in patients with congenital heart disease. *J Am Coll Cardiol* 42: 1650-5
- McElhinney DB, Krantz ID, Bason L, Piccoli DA, Emerick KM, Spinner NB, Goldmuntz E (2002) Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome. *Circulation* 106: 2567-74
- 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 (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316: 1488-91
- 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, Raber L, Gallati S, Striano P, Coppola A, Tolmie JL, Tobias ES, Lilley C, Armengol L, Spyschaert 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 (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359: 1685-99
- Meilhac SM, Esner M, Kelly RG, Nicolas JF, Buckingham ME (2004) The clonal origin of myocardial cells in different regions of the embryonic mouse heart. *Dev Cell* 6: 685-98
- Mishina Y, Suzuki A, Ueno N, Behringer RB (1995) Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation furin mouse embryogenesis. *Genes & Dev.* 9: 3027-3037
- Miyazono K (1999) Signal transduction by bone morphogenetic protein receptors: functional roles of Smad proteins. *Bone* 25: 91-3

- Miyazono K, Maeda S, Imamura T (2005) BMP receptor signaling: transcriptional targets, regulation of signals, and signaling cross-talk. *Cytokine Growth Factor Rev* 16: 251-63
- Mjaatvedt CH, Nakaoka T, Moreno-Rodriguez R, Norris RA, Kern MJ, Eisenberg CA, Turner D, Markwald RR (2001) The outflow tract of the heart is recruited from a novel heart-forming field. *Developmental Biology* 238: 97-109
- Momma K, Kondo C, Ando M, Matsuoka R, Takao A (1995) Tetralogy of Fallot associated with chromosome 22q11 deletion. *Am J Cardiol* 76: 618-21
- Moore KL, Persaud TVN (1998) *The Developing Human* (6th edition). W.B. Saunders Company
- MRC (1991) Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* 338: 131-7
- Murakami G, Watabe T, Takaoka K, Miyazono K, Imamura T (2003) Cooperative inhibition of bone morphogenetic protein signaling by Smurf1 and inhibitory Smads. *Mol Biol Cell* 14: 2809-17
- Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE (2005) T-box genes in vertebrate development. *Annu Rev Genet* 39: 219-39
- Nakamura T, Colbert MC, Robbins J (2006) Neural crest cells retain multipotential characteristics in the developing valves and label the cardiac conduction system. *Circ Res* 98: 1547-54
- Nakashima A, Katagiri T, Tamura M (2005) Cross-talk between Wnt and bone morphogenetic protein 2 (BMP-2) signaling in differentiation pathway of C2C12 myoblasts. *J Biol Chem* 280: 37660-8
- Naya FJ, Black BL, Wu H, Bassel-Duby R, Richardson JA, Hill JA, Olson EN (2002) Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med* 8: 1303-9
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42: 30-5
- 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 (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461: 272-6
- Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, Wojcik SE, Ferdin J, Kunej T, Xiao L, Manoukian S, Secretò G, Ravagnani F, Wang X,

- Radice P, Croce CM, Davuluri RV, Calin GA (2010) Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res* 70: 2789-98
- Nishibatake M, Kirby ML, Van Mierop LH (1987) Pathogenesis of persistent truncus arteriosus and dextroposed aorta in the chick embryo after neural crest ablation. *Circulation* 75: 255-64
- Niwa M, Rose SD, Berget SM (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* 4: 1552-9
- Nohe A, Keating E, Knaus P, Petersen NO (2004 ) Signal transduction of bone morphogenetic protein receptors. *Cell Signal* 16: 291-299
- Noonan JA (1994) Noonan syndrome. An update and review for the primary pediatrician. *Clin Pediatr (Phila)* 33: 548-55
- Nora JJ (1968) Multifactorial inheritance hypothesis for the etiology of congenital heart diseases. The genetic-environmental interaction. *Circulation* 38: 604-17
- Nora JJ, Meyer TC (1966) Familial nature of congenital heart diseases. *Pediatrics* 37: 329-34
- Nora JJ, Nora AH (1987) Maternal transmission of congenital heart diseases: new recurrence risk figures and the questions of cytoplasmic inheritance and vulnerability to teratogens. *Am J Cardiol* 59: 459-63
- Norton ME (1997) Teratogen update: fetal effects of indomethacin administration during pregnancy. *Teratology* 56: 282-92
- Ohno M, Sakamoto H, Shimura Y (1987) Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proc Natl Acad Sci U S A* 84: 5187-91
- Okagawa H, Markwald RR, Sugi Y (2007) Functional BMP receptor in endocardial cells is required in atrioventricular cushion mesenchymal cell formation in chick. *Dev. Biol.*
- Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M (2009) Recurrence of congenital heart defects in families. *Circulation* 120: 295-301
- Palomino-Doza J, Rahman TJ, Avery PJ, Mayosi BM, Farrall M, Watkins H, Edwards CR, Keavney B (2008) Ambulatory blood pressure is associated with polymorphic variation in P2X receptor genes. *Hypertension* 52: 980-5
- Pierpont ME, Basson CT, Benson DW, Jr., Gelb BD, Giglia TM, Goldmuntz E, McGee G, Sable CA, Srivastava D, Webb CL (2007) Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart

- Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 115: 3015-38
- Posch MG, Perrot A, Schmitt K, Mittelhaus S, Esenwein EM, Stiller B, Geier C, Dietz R, Gessner R, Ozcelik C, Berger F (2008) Mutations in GATA4, NKX2.5, CRELD1, and BMP4 are infrequently found in patients with congenital cardiac septal defects. *Am J Med Genet A* 146A: 251-3
- Pradat P (1994) Recurrence risk for major congenital heart defects in Sweden: a registry study. *Genet Epidemiol* 11: 131-40
- Rivera-Feliciano J, Tabin CJ (2006) Bmp2 instructs cardiac progenitors to form the heart-valve-inducing field. *Dev Biol* 295: 580-8
- Roberts AE, Araki T, Swanson KD, Montgomery KT, Schiripo TA, Joshi VA, Li L, Yassin Y, Tamburino AM, Neel BG, Kucherlapati RS (2007) Germline gain-of-function mutations in SOS1 cause Noonan syndrome. *Nat Genet* 39: 70-4
- Roberts KE, McElroy JJ, Wong WP, Yen E, Widlitz A, Barst RJ, Knowles JA, Morse JH (2004) BMPR2 mutations in pulmonary arterial hypertension with congenital heart disease. *Eur Respir J* 24: 371-4
- Robson M, Offit K (2007) Clinical practice. Management of an inherited predisposition to breast cancer. *N Engl J Med* 357: 154-62
- Rosenzweig BL, Imamura T, Okadome T, Cox GN, Yamashita H, ten Dijke P, Heldin CH, Miyazono K (1995) Cloning and characterization of a human type II receptor for bone morphogenetic proteins. *Proc Natl Acad Sci U S A* 92: 7632-6
- Roth S, Sistonen P, Salovaara R, Hemminki A, Loukola A, Johansson M, Avizienyte E, Cleary KA, Lynch P, Amos CI, Kristo P, Mecklin JP, Kellokumpu I, Jarvinen H, Aaltonen LA (1999) SMAD genes in juvenile polyposis. *Genes Chromosomes Cancer* 26: 54-61
- Roux PP, Blenis J (2004) ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol Mol Biol Rev* 68: 320-344
- Rudarakanchana N, Flanagan JA, Chen H, Upton PD, Machado R, Patel D, Trembath RC, Morrell NW (2002) Functional analysis of bone morphogenetic protein type II receptor mutations underlying primary pulmonary hypertension. *Hum Mol Genet* 11: 1517-25
- Samad TA, Rebbapragada A, Bell E, Zhang Y, Sidis Y, Jeong SJ (2005) DRAGON, a bone morphogenetic protein co-receptor. *J. Biol Chem* 280: 14122-14129

- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H (2007) Genomewide association analysis of coronary artery disease. *N Engl J Med* 357: 443-53
- Sansoucie DA, Cavaliere TA (1997) Transition from fetal to extrauterine circulation. *Neonatal Netw* 16: 5-12
- Satoda M, Zhao F, Diaz GA, Burn J, Goodship J, Davidson HR, Pierpont ME, Gelb BD (2000) Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus. *Nat Genet* 25: 42-6
- Scanlon KS, Ferencz C, Loffredo CA, Wilson PD, Correa-Villasenor A, Khoury MJ, Willett WC (1998) Preconceptional folate intake and malformations of the cardiac outflow tract. Baltimore-Washington Infant Study Group. *Epidemiology* 9: 95-8
- Schardein JL (2000) Anticonvulsants. In: *Chemically Induced Birth Defects*. 3rd ed. New York, NY: Marcel Dekker: 179-235
- Schlange T, Andree B, Arnold H, Brand T (2000) BMP2 is required for early heart development during a distinct time period. *Mechanisms of Development* 91: 259-270
- Schlueter J, Manner J, Brand T (2006) BMP is an important regulator of proepicardial identity in the chick embryo. *Dev. Biol.* 295: 546-558
- Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG (1998) Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281: 108-11
- Schubbert S, Zenker M, Rowe SL, Boll S, Klein C, Bollag G, van der Burgt I, Musante L, Kalscheuer V, Wehner LE, Nguyen H, West B, Zhang KY, Sistermans E, Rauch A, Niemeyer CM, Shannon K, Kratz CP (2006) Germline KRAS mutations cause Noonan syndrome. *Nat Genet* 38: 331-6
- Schultheiss TM, Burch JB, Lassar AB (1997) A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev* 11: 451-62
- Schwappacher R, Weiske J, Heining E, Ezerski V, Marom B, Henis YI, Huber O, Knaus P (2009) Novel crosstalk to BMP signalling: cGMP-dependent kinase I modulates BMP receptor and Smad activity. *EMBO J* 28: 1537-50

- Shi Y, Katsev S, Cai C, Evans S (2000) BMP signaling is required for heart formation in vertebrates. *Dev Biol* 224: 226-37
- Shieh JT, Srivastava D (2009) Heart malformation: what are the chances it could happen again? *Circulation* 120: 269-71
- Signori E, Bagni C, Papa S, Primerano B, Rinaldi M, Amaldi F, Fazio VM (2001) A somatic mutation in the 5'UTR of BRCA1 gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene* 20: 4596-600
- Smith KA, Joziassse IC, Chocron S, van Dinther M, Guryev V, Verhoeven MC, Rehmann H, van der Smagt JJ, Doevendans PA, Cuppen E, Mulder BJ, Ten Dijke P, Bakkers J (2009) Dominant-negative ALK2 allele associates with congenital heart defects. *Circulation* 119: 3062-9
- Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 15: 2490-508
- Smithells RW, Newman CG (1992) Recognition of thalidomide defects. *J Med Genet* 29: 716-23
- Somi S, Buffing AAM, Moorman AFM, van der Hoff MJB (2004) Dynamic patterns of expression of BMP isoforms 2,4,5,6, and 7 during chicken heart development. *The Anatomical Record Part A* 279A: 636-651
- Sorrentino RP, Gajewski KM, Schulz RA (2005) GATA factors in Drosophila heart and blood cell development. *Semin Cell Dev Biol* 16: 107-16
- Souchelnytskyi S, Nakayama T, Nakao A, Moren A, Heldin CH, Christian JL, ten Dijke P (1998) Physical and functional interaction of murine and Xenopus Smad7 with bone morphogenetic protein receptors and transforming growth factor-beta receptors. *J Biol Chem* 273: 25364-70
- Sperling S, Grimm CH, Dunkel I, Mebus S, Sperling HP, Ebner A, Galli R, Lehrach H, Fusch C, Berger F, Hammer S (2005) Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. *Hum Mutat* 26: 575-82
- Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J,

- Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemenev LA, Franke B, Sabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455: 232-6
- Stern CD, Yu RT, Kakizuka A, Kintner CR, Mathews LS, Vale WW, Evans RM, Umesono K (1995) Activin and its receptors during gastrulation and the later phases of mesoderm development in the chick embryo. *Dev Biol* 172: 192-205
- Stoss O, Stoilov P, Hartmann AM, Naylor O, Stamm S (1999) The in vivo minigene approach to analyze tissue-specific splicing. *Brain Res Brain Res Protoc* 4: 383-94
- Stottmann RW, Choi M, Mishina Y, Meyers EN, Klingensmith J (2004) BMP receptor IA is required in mammalian neural crest cells for development of the cardiac outflow tract and ventricular myocardium. *Development* 131: 2205-2218
- Styrkarsdottir U, Cazier JB, Kong A, Rolfsson O, Larsen H, Bjarnadottir E, Johannsdottir VD, Sigurdardottir MS, Bagger Y, Christiansen C, Reynisdottir I, Grant SF, Jonasson K, Frigge ML, Gulcher JR, Sigurdsson G, Stefansson K (2003) Linkage of osteoporosis to chromosome 20p12 and association to BMP2. *PLoS Biol* 1: E69
- Suzuki S, Marazita ML, Cooper ME, Miwa N, Hing A, Jugessur A, Natsume N, Shimozato K, Ohbayashi N, Suzuki Y, Niimi T, Minami K, Yamamoto M, Altannamar TJ, Erkhembaatar T, Furukawa H, Daack-Hirsch S, L'Heureux J, Brandon CA, Weinberg SM, Neiswanger K, Deleyiannis FW, de Salamanca JE, Vieira AR, Lidral AC, Martin JF, Murray JC (2009) Mutations in BMP4 are associated with subepithelial, microform, and overt cleft lip. *Am J Hum Genet* 84: 406-11
- Tabatabaieifar M, Schlingmann KP, Litwin M, Emre S, Bakkaloglu A, Mehls O, Antignac C, Schaefer F, Weber S (2009) Functional analysis of BMP4 mutations identified in pediatric CAKUT patients. *Pediatr Nephrol* 24: 2361-8
- Takizawa T, Ochiai W, Nakashima K, Taga T (2003) Enhanced gene activation by Notch and BMP signaling cross-talk. *Nucleic Acids Res* 31: 5723-31

- Tam PP, Parameswaran M, Kinder SJ, Weinberger RP (1997) The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development* 124: 1631-42
- Tellier AL, Cormier-Daire V, Abadie V, Amiel J, Sigaudy S, Bonnet D, de Lonlay-Debeney P, Morrissette-Durand MP, Hubert P, Michel JL, Jan D, Dollfus H, Baumann C, Labrune P, Lacombe D, Philip N, LeMerrer M, Briard ML, Munnich A, Lyonnet S (1998) CHARGE syndrome: report of 47 cases and review. *Am J Med Genet* 76: 402-9
- ten Dijke P, Yamashita J, Ichijo H, Frazen P, Laiho M, Miyazono K, Heldin CH (1994) Characterization of type I receptors for transforming growth factor- $\beta$  and activin. *Science* 264: 101-104
- Thienpont B, Mertens L, de Ravel T, Eyskens B, Boshoff D, Maas N, Fryns JP, Gewillig M, Vermeesch JR, Devriendt K (2007) Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur Heart J* 28: 2778-84
- Urist MR (1965) Bone: formation by autoinduction. *Science* 150: 893-899
- van Beynum IM, den Heijer M, Blom HJ, Kapusta L (2007) The MTHFR 677C->T polymorphism and the risk of congenital heart defects: a literature review and meta-analysis. *QJM* 100: 743-53
- van den Hoff MJ, Kruithof BP, Moorman AF, Markwald RR, Wessels A (2001) Formation of myocardium after the initial development of the linear heart tube. *Dev Biol* 240: 61-76
- van Mierop LH, Kutsche LM (1985) Development of the ventricular septum of the heart. *Heart Vessels* 1: 114-9
- Vanhaesebroeck B, Leever SJ, Ahmadi K, Timms J, Katso R, Driscoll PC, Woscholski R, Parker PJ, Waterfield MD (2001) Synthesis and function of 3-phosphorylated inositol lipids. *Annu Rev Biochem* 70: 535-602
- Visser JA, Olaso R, Verhoef-Post M, Kramer P, Themmen AP, Ingraham HA (2001) The serine/threonine transmembrane receptor ALK2 mediates Mullerian inhibiting substance signaling. *Mol Endocrinol* 15: 936-45
- Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG (2004) Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 36: 955-7

- Waldo KL, Hutson MR, Stadt HA, Zdanowicz M, Zdanowicz J, Kirby ML (2005)  
Cardiac neural crest is necessary for normal addition of the myocardium to the  
arterial pole from the secondary heart field. *Dev Biol* 281: 66-77
- Waldo KL, Kumiski DH, Wallis KT, Stadt HA, Hutson MR, Platt DH, Kirby ML (2001)  
Conotruncal myocardium arises from a secondary heart field. *Development* 128:  
3179-88
- Wang D, Kanuma T, Mizumuma H, Ibuki Y, Takenoshita S (2000a) Mutation analysis  
of the Smad6 and Smad7 gene in human ovarian cancers. *Int J Oncol* 17: 1087-  
91
- Wang J, Sridurongrit S, Dudas M, Thomas P, Nagy A, Schneider MD, Epstein JA,  
Karttinen V (2005a) Atrioventricular cushion transformation is mediated by  
ALK2 in the developing mouse heart. *Developmental Biology* 286: 299-310
- Wang M, Wang JY, Cisler J, Imaizumi K, Burton BK, Jones MC, Lamberti JJ, Godfrey  
M (1997) Three novel fibrillin mutations in exons 25 and 27: classic versus  
neonatal Marfan syndrome. *Hum Mutat* 9: 359-62
- Wang Q, Rao S, Topol EJ (2005b) Miscues on the "lack of MEF2A mutations" in  
coronary artery disease. *J Clin Invest* 115: 1399-400; author reply 1400-1
- Wang Q, Reiter RS, Huang QQ, Jin JP, Lin JJ (2001) Comparative studies on the  
expression patterns of three troponin T genes during mouse development. *Anat  
Rec* 263: 72-84
- Wang Q, Sigmund CD, Lin JJ (2000b) Identification of cis elements in the cardiac  
troponin T gene conferring specific expression in cardiac muscle of transgenic  
mice. *Circ Res* 86: 478-84
- Wang Y, Qian L, Liu D (2007) Bone morphogenetic protein-2 acts upstream of  
myocyte-specific enhancer factor 2a to control embryonic cardiac contractility.  
*Cardio. Res.* 74: 290-303
- Ward C (2000) Clinical significance of the bicuspid aortic valve. *Heart* 83: 81-5
- Webb S, Brown NA, Anderson RH, Richardson MK (2000) Relationship in the chick of  
the developing pulmonary vein to the embryonic systemic venous sinus. *Anat  
Rec* 259: 67-75
- Wessels A, Sedmera D (2003) Developmental anatomy of the heart: a tale of mice and  
man. *Physiol Genomics* 15: 165-76
- Wilson AM, Schlade-Bartusiak K, Tison JL, Macintyre G, Cox DW (2009) A minigene  
approach for analysis of ATP7B splice variants in patients with Wilson disease.  
*Biochimie* 91: 1342-5

- Wilson PD, Loffredo CA, Correa-Villasenor A, Ferencz C (1998) Attributable fraction for cardiac malformations. *Am J Epidemiol* 148: 414-23
- Winnier G, Blessing M, Labosky PA, Hogan BLM (1995) Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Gene Dev* 9: 2105-2116
- Wrana JL, Carcamo J, Attisano L, Cheifetz S, Zentella A, Lopez-Casillas F, Massague J (1992) The type II TGF-beta receptor signals diverse responses in cooperation with the type I receptor. *Cold Spring Harb Symp Quant Biol* 57: 81-6
- Xiao JH, Davidson I, Matthes H, Garnier JM, Chambon P (1991) Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* 65: 551-68
- Xu X, Yin Z, Hudson JB, Ferguson EL, Frasch M (1998) Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. *Genes Dev* 12: 2354-70
- Yaffe D, Saxel O (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature* 270: 725-727
- Yamada M, Revelli JP, Eichele G, Barron M, Schwartz RJ (2000) Expression of chick Tbx-2, Tbx-3, and Tbx-5 genes during early heart development: evidence for BMP2 induction of Tbx2. *Dev Biol* 228: 95-105
- Yamada M, Szendro PI, Prokscha A, Schwartz RJ, Eichele G (1999) Evidence for a role of Smad6 in chick cardiac development. *Dev. Biol.* 215: 48-61
- Yang L, Cai CL, Lin L, Qyang Y, Chung C, Monteiro RM, Mummery CL, Fishman GI, Cogen A, Evans S (2006) Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* 133: 1575-85
- Yang X, Dormann D, Munsterberg AE, Weijer CJ (2002) Cell movement patterns during gastrulation in the chick are controlled by positive and negative chemotaxis mediated by FGF4 and FGF8. *Dev Cell* 3: 425-37
- Yelbuz TM, Waldo KL, Kumiski DH, Stadt HA, Wolfe RR, Leatherbury L, Kirby ML (2002) Shortened outflow tract leads to altered cardiac looping after neural crest ablation. *Circulation* 106: 504-10
- Zaffran S, Kelly RG, Meilhac SM, Buckingham ME, Brown NA (2004) Right ventricular myocardium derives from the anterior heart field. *Circ Res* 95: 261-8
- Zatyka M, Priestley M, Ladusans EJ, Fryer AE, Mason J, Latif F, Maher ER (2005) Analysis of CRELD1 as a candidate 3p25 atrioventricular septal defect locus (AVSD2). *Clin Genet* 67: 526-8

- Zhang C, Hastings ML, Krainer AR, Zhang MQ (2007) Dual-specificity splice sites function alternatively as 5' and 3' splice sites. *Proc Natl Acad Sci U S A* 104: 15028-33
- Zhang H, Bradley A (1996) Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* 122: 2977-2986
- Zhang J, Cai WW (1993) Association of the common cold in the first trimester of pregnancy with birth defects. *Pediatrics* 92: 559-63
- Zhou XP, Woodford-Richens K, Lehtonen R, Kurose K, Aldred M, Hampel H, Launonen V, Virta S, Pilarski R, Salovaara R, Bodmer WF, Conrad BA, Dunlop M, Hodgson SV, Iwama T, Jarvinen H, Kellokumpu I, Kim JC, Leggett B, Markie D, Mecklin JP, Neale K, Phillips R, Piris J, Rozen P, Houlston RS, Aaltonen LA, Tomlinson IP, Eng C (2001) Germline mutations in BMPR1A/ALK3 cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes. *Am J Hum Genet* 69: 704-11
- Zhu H, Kavsak P, Abdollah S, Wrana JL, Thomsen GH (1999) A SMAD ubiquitin ligase targets the BMP pathway and affects embryonic pattern formation. *Nature* 400: 687-93