

Rare and private variations in neural crest, apoptosis and sarcomere genes define the polygenic background of isolated Tetralogy of Fallot

Marcel Grunert^{1,3,†}, Cornelia Dorn^{1,3,4,†}, Markus Schueler^{1,3}, Ilona Dunkel¹, Jenny Schlesinger^{1,3}, Siegrun Mebus^{5,‡}, Vladimir Alexi-Meskishvili⁶, Andreas Perrot³, Katharina Wassilew⁷, Bernd Timmermann², Roland Hetzer⁶, Felix Berger⁵ and Silke R. Sperling^{1,3,4,*}

¹Group of Cardiovascular Genetics, Department of Vertebrate Genomics and ²Next Generation Service Group, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany ³Cardiovascular Genetics, Experimental and Clinical Research Center, Charité—Universitätsmedizin Berlin and Max Delbrück Center (MDC) for Molecular Medicine, Berlin 13125, Germany ⁴Department of Biology, Chemistry and Pharmacy, Free University of Berlin, Berlin 14195, Germany ⁵Department of Pediatric Cardiology, German Heart Institute Berlin and Department of Pediatric Cardiology, Charité—Universitätsmedizin Berlin, Berlin 13353, Germany ⁶Department of Cardiac Surgery and ⁷Department of Pathology, German Heart Institute Berlin, Berlin, Germany

Received November 5, 2013; Revised December 17, 2013; Accepted January 19, 2014

Tetralogy of Fallot (TOF) is the most common cyanotic congenital heart disease. Its genetic basis is demonstrated by an increased recurrence risk in siblings and familial cases. However, the majority of TOF are sporadic, isolated cases of undefined origin and it had been postulated that rare and private autosomal variations in concert define its genetic basis. To elucidate this hypothesis, we performed a multilevel study using targeted re-sequencing and whole-transcriptome profiling. We developed a novel concept based on a gene's mutation frequency to unravel the polygenic origin of TOF. We show that isolated TOF is caused by a combination of deleterious private and rare mutations in genes essential for apoptosis and cell growth, the assembly of the sarcomere as well as for the neural crest and secondary heart field, the cellular basis of the right ventricle and its outflow tract. Affected genes coincide in an interaction network with significant disturbances in expression shared by cases with a mutually affected TOF gene. The majority of genes show continuous expression during adulthood, which opens a new route to understand the diversity in the long-term clinical outcome of TOF cases. Our findings demonstrate that TOF has a polygenic origin and that understanding the genetic basis can lead to novel diagnostic and therapeutic routes. Moreover, the novel concept of the gene mutation frequency is a versatile measure and can be applied to other open genetic disorders.

INTRODUCTION

Congenital heart defects (CHDs) are the most common birth defect in human with an incidence of almost 1% of all live births (1). Approximately one-third of CHDs are associated with non-cardiac syndromes such as Trisomy 21 (Down syndrome

[MIM 190685]). Most CHDs occur sporadically (70%) and do not follow Mendelian heritage (2). There are many different phenotypes ranging from a single septal defect up to a univentricular heart. Already in 1968, Nora suggested a multifactorial inheritance with genetic–environmental interactions (2). Since then, many genes have been identified harboring functional

*To whom correspondence should be addressed at: Department of Cardiovascular Genetics, Experimental and Clinical Research Center, Charité—Universitätsmedizin Berlin and Max Delbrück Center (MDC) for Molecular Medicine, Lindenberger Weg 80, 13125 Berlin, Germany. Tel: +49 30450540123; Email: silke.sperling@charite.de

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

‡Present address: Department of Pediatric Cardiology and Congenital Heart Disease, German Heart Center of the Technical University Munich, Munich 80636, Germany.

mutations in patients and were classified as CHD genes (3). Useful resources have been familial cases; however, the large proportion of non-familial cases still awaits genetic and molecular work-up.

Tetralogy of Fallot (TOF [MIM 187500]) is the most common form of cyanotic congenital heart disease with a prevalence of 3 per 10 000 live births, accounting for 7–10% of all CHDs (4). The characteristics of TOF were first described in 1671 and later named after Etienne-Louis Fallot. TOF is regarded as a family of diseases characterized by four cardiac features: ventricular septal defect, overriding aorta, right ventricular outflow tract obstruction and right ventricular hypertrophy (Fig. 1A) (5). Accordingly, the TOF heart shows hemodynamic settings different from a normal heart, such as shunting via the septal defect and an increased pressure in the right ventricle. Additional panels of cardiovascular abnormalities like atrial septal defects or pulmonary artery malformations as well as non-cardiac abnormalities are often associated with the disease. TOF is a well-recognized subfeature of syndromic disorders such as DiGeorge syndrome (MIM 188400) (6), Down syndrome (7), Alagille syndrome (MIM 610205) (8) and Holt-Oram syndrome (MIM 142900) (9). Interestingly, it has been shown that differences in the clinical outcome of TOF after corrective surgery depend on the associated abnormalities (10).

That TOF has a genetic basis is demonstrated by an increased recurrence risk in siblings of ~3% and a number of documented familial cases (11). A panel of copy number variations (CNVs) is associated with isolated TOF cases and more recently two genetic loci harboring common disease variants were identified (12,13). However, the majority of TOFs are isolated, non-syndromic cases whose precise causes are unknown, which is also the situation for the majority of CHDs and many serious non-Mendelian diseases with a clear genetic component.

It has been assumed that CHDs might also be caused by rare autosomal recessive variations in concert with private variations (3,14), which might individually show minor functional impairment but in combination could be disease causing (15). In this concept, multiple mutations in different genes can lead to disturbances of a molecular network that result in a common phenotypic expression. However, a great challenge is the discrimination of variations and genes causative for a disease in a particular individual from deleterious variations being tolerated. Here, we introduce a novel approach to discriminate causative genes considering the frequency of a gene's affection by deleterious variations in a cohort (gene mutation frequency, GMF). We show that TOF is caused by combinations of rare and private mutations in neural crest (NC), apoptosis and sarcomere genes. This finding is in agreement with the hypothesis that sub-features of TOF, namely a ventricular septal defect, might result from premature stop of cardiomyocyte proliferation. Furthermore, genes coincide in a functional interaction network and show continuous expression during adulthood, which, e.g. in case of sarcomeric genes known to cause cardiomyopathy, could potentially explain well-known differences in the long-term clinical outcome of phenotypically similar cases. Our findings demonstrate that TOF has a polygenic origin and that understanding the genetic basis might lead to novel diagnostic and therapeutic routes.

RESULTS

TOF cohort and study approach

We studied 26 well-defined individuals of which 22 are patients with TOF (Fig. 1A) and 4 are healthy controls. These TOF cases were selected based on our previous gene expression analysis and phenotypic evaluations such that these are sporadic cases without any additional cardiovascular or other abnormalities (16,17). We conducted a multilevel study of these cases with the aim to gather insights into rare or private variations that might define a molecular network underlying the development of TOF. To analyze genomic variations, we applied targeted re-sequencing using genomic DNA from blood and selected genes and microRNAs of known or potential interest for cardiac development and function by combining different data resources and bioinformatics approaches; details are given in the Supplementary Material, Tables S1 and S2. This resulted in 867 genes and 167 microRNAs to be assessed. Further, we obtained expression profiles of transcripts and microRNAs in cardiac tissues using Illumina sequencing, and studied histological sections of endomyocardial specimen of selected cases. Supplementary Material, Table S3 gives an overview of samples and different analyses performed.

Genomic variations observed in TOF

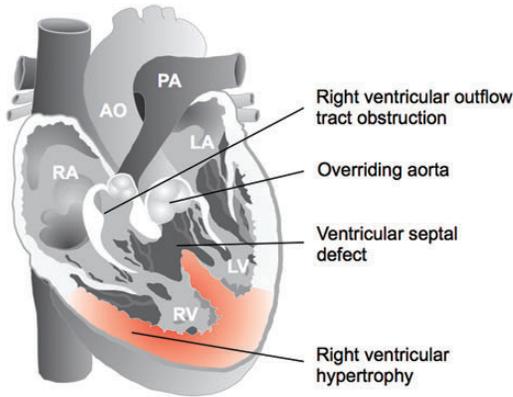
Single nucleotide variation (SNV) and insertion/deletion (InDel) calling and filtering in TOF cases resulted in a total of 223 local variations altering the coding sequence of 162 genes classified as damaging ($n = 146$), nonsense ($n = 3$), frameshift ($n = 61$) or splice site ($n = 12$) mutations as well as amino acid deletion ($n = 1$) (Supplementary Material, Fig. S1 and Table S4). In general, variations were equally distributed over all chromosomes (Fig. 1B). No relevant mutations were observed in microRNA mature sequences.

Discrimination of causative genes by considering the frequency of a genes's affection

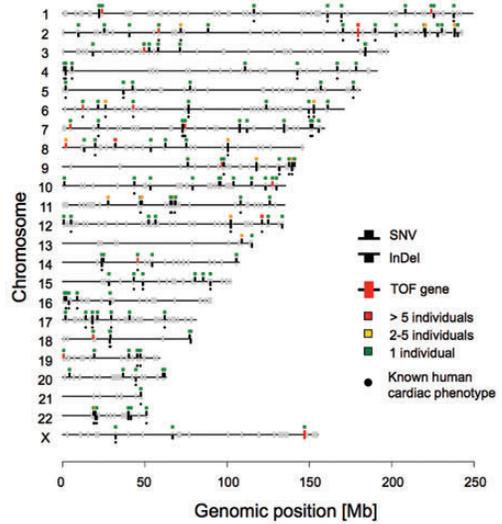
We propose that multiple private and/or rare genetic variations could contribute to TOF. However, a great challenge has been the establishment of tools to discriminate variations and genes causative for a disease in a particular individual from deleterious variations being tolerated in the individual context. With the increasing number of individuals being genotyped, previously called private mutations now are also rarely found in controls (18–20). This questions our previous concept, where the proof of a mutation–phenotype association was based on its private finding in the diseased versus healthy cohort, where the latter consisted of few hundred individuals (3,21).

Along this line, we developed a concept that would overcome the limited focus on individual mutations and instead consider at a whole all deleterious mutations in a distinct gene; having in mind that genes associated with a disease would have more deleterious mutations in patients than controls. Thus, we introduce the GMF, which can be seen as an analog to the minor allele frequency (MAF) that is based on single variations and used for example in genome-wide association studies. The GMF is calculated based on the number of individuals harboring deleterious mutations in relationship to the total number of individuals

A Tetralogy of Fallot



B



C Gene Mutation Frequency (GMF)

$$GMF = \frac{\text{Number of individuals with mutations}}{\text{Number of individuals with sufficient genotype information}} / \text{Captured gene length} \times 1000$$

$$GMF = \frac{n_M}{n - n_{N/A}} / l \times 1000$$

n - Total number of individuals in the cohort
 n_M - Number of individuals with at least one mutation
 n_{N/A} - Number of individuals without sufficient genotype information
 l - Captured gene length in bp

Example of GMF calculation with individual genotypes for n=10: $GMF = \frac{3}{10-2} / 2500 \times 1000 = 0.15$

Individual Mutation		Individual									
		1	2	3	4	5	6	7	8	9	10
Gene A (2500 bp)	I	0	1 ^{het}	0	0	0	0	N/A	N/A	0	0
	II	0	0	0	1 ^{het}	0	0	1 ^{hom}	N/A	0	0
	III	N/A	0	0	1 ^{hom}	0	0	0	N/A	0	0
	Summary	N/A	1	0	1	0	0	1	N/A	0	0

D Maximal GMF (GMF_{MAX})

Applicable if no individual genotypes are available for a cohort

$$GMF_{MAX} = \frac{\text{Calculated max. number of individuals with mutations}}{\text{Calculated min. number of individuals with sufficient genotypes}} / \text{Captured gene length} \times 1000$$

$$GMF_{MAX} = \frac{\lambda}{\max\left(\lambda, n - \sum_{m \in \text{Gene}} n'_{N/A}\right)} / l \times 1000, \lambda = \min\left(n, \sum_{m \in \text{Gene}} n'_M\right)$$

n - Total number of individuals in the cohort
 n'_M - Number of individuals with a mutation M
 n'_{N/A} - Number of individuals without genotype information for a mutation M
 l - Captured gene length in bp

Example of GMF calculation without individual genotypes for n=10: $GMF_{MAX} = \frac{\min(10, 4)}{\max(\min(10, 4), 10 - 5)} / 2500 \times 1000 = \frac{4}{5} = 0.32$

Genotype Mutation		Genotype				
		hom	het	∑ _{hom/het}	wt	N/A
Gene A (2500 bp)	I	0	1	1	7	2
	II	1	1	2	7	1
	III	1	0	1	7	2
				∑ n' _M =4		∑ n' _{N/A} =5

with sufficient genotype information (Fig. 1C). The GMF is normalized by the gene length and kilobase-scaled to allow for comparison between genes of different lengths. To overcome the limitation that individual genotype information are not directly provided in public data sets, we introduce a so-called maximal GMF (GMF_{MAX}), which is based on the calculated maximal possible number of individuals with mutations (Fig. 1D). Deleterious mutations are defined by filtering settings, which can vary depending on the study focus; however, same settings should be applied to case and control data. In the following, we use the NHLBI-ESP genomic data as the control data set, which represents the largest exome data set of control individuals currently available and includes 4300 exomes of European American ancestry (EA controls).

To verify the appropriateness of the GMF, we conducted a retrospective study for hypertrophic cardiomyopathy (HCM). We re-analyzed eight studies, which identified relevant mutations in five genes (*MYH7*, *TNNT2*, *TNNI3*, *MYL2*, *ACTC1*) causing HCM (22). We calculated GMFs for the different HCM cohorts based on the number of identified deleterious mutations. We compared these GMFs against the GMF_{MAX} calculated for the respective gene in the EA control data set (Table 1), which was accordingly filtered for deleterious mutations. The GMFs obtained for the HCM cohorts were in general at least 5-fold higher than the GMF_{MAX} of the controls and its significance was underlined by a one-sided Fisher's exact test. This holds true not only for large-scale studies of more than 190 patients but also for smaller studies below 50 cases or even down to 15 cases. The latter cohort is characterized by a very specific phenotype description, which might reduce noise in the data and reflects the situation of our TOF cohort. Finally, we assumed that the GMF could be a valuable measurement to identify disease-related genes harboring deleterious mutations in a broad range of cohort sizes.

Isolated TOF caused by polygenic variations

In our TOF cohort, we found 103 genes harboring exclusively SNVs, in 18 genes SNVs and InDels, and in 41 genes only InDels. Of these, 50 were private SNVs and 66 private InDels, which have not been observed in controls or dbSNP (v137). For 121 genes affected by SNVs, GMFs were calculated and for 107 of these, sufficient sequence information was available in EA controls enabling a comparison. We found 47 genes with an at least 5-fold higher GMF in the TOF cohort compared with the EA controls (Supplementary Material, Fig. S2 and Table S5). To substantiate this finding, we evaluated a Danish control cohort consisting of exome data for 200 ethnically matched individuals with individual genotype information (19). In this data set, sufficient information was provided for 42 out of the 47 genes confirming all our results obtained with the EA controls (data not shown). Further, we statistically evaluated the occurrence of deleterious SNVs in

the TOF cohort applying a Fisher's exact test. This resulted in 15 genes with a significantly higher GMF in the TOF cohort compared with EA controls ($P < 0.05$, Fig. 2A) and a mean GMF ratio of 30 (Table 2).

The assessment of the sarcomere gene titin (*TTN*) using the GMF approach was hindered as it is extraordinary long (captured exonic length of 110 739 bp) and thus, a high number of SNVs (1016 deleterious SNVs) was identified in the 4300 EA controls (Supplementary Material, Table S5). The high number of SNVs in *TTN* leads to a strong reduction of the available genotypes even if the sequencing quality for individual SNVs is sufficient. For comparison, the second longest gene affected by SNVs in our cohort is *SYNE1* (captured exonic length of 30 235 bp), which shows 242 SNVs in the EA controls and for which a GMF_{MAX} could be calculated. To overcome this problem, we performed an exon-by-exon approach by calculating the mutation frequency for individual exons (exon mutation frequency, EMF). We found seven out of nine affected exons with a significantly higher EMF in the TOF cohort compared with EA controls ($P < 0.05$, Supplementary Material, Table S6). To ensure that both approaches lead to the same results, we also calculated the EMF for *SYNE1* and found that neither the GMF ($P = 0.9$) nor the EMF ($P = 0.1$) is significantly higher in the TOF cases compared with the EA controls.

In summary, we identified a total of 16 genes (called 'TOF genes') based on a significantly higher GMF or EMF (Fig. 2A). Out of these, 11 genes could further be confirmed in comparison with the Danish controls, which might be biased by a far lower coverage in the Danish study (not confirmed are *BARX1*, *FMRI*, *HCN2*, *ROCK1*, *WBSCR16*). Out of the 16 TOF genes, 6 genes have known associations with human cardiac disease and 7 genes show a cardiac phenotype when mutated or knocked out in mice. Five of the TOF genes had not, to our knowledge, previously been associated with a cardiac phenotype at all, and 11 not with human CHDs (Fig. 2A and Supplementary Material, Table S7). For the case TOF-08, no deleterious SNVs were found in significant TOF genes; however, histological assessment of a cardiac biopsy showed that a deleterious mutation in the cardiomyopathy gene coding for myosin binding protein C3 (*MYBPC3*) might be causative (Fig. 4). For *MYBPC3*, a GMF calculation in controls is hindered due to insufficient genotype information in EA controls. In addition, TOF-08 harbors a deleterious mutation in the armadillo repeat gene deleted in velocardiofacial syndrome (*ARVCF*), which shows a 3-fold higher GMF in TOF compared with EA controls.

Confirmation of genomic variations using RNA-seq and Sanger sequencing

In addition to DNA sequencing, we gathered mRNA profiles from right ventricles of respective patients to study expression

Figure 1. Genes affected in TOF are distributed over all chromosomes and were subjected to GMF calculation. (A) Schematic representation of Tetralogy of Fallot. AO, aorta; LA, left atrium; LV, left ventricle; PA, pulmonary artery; RA, right atrium; RV, right ventricle. (B) Genomic positions of affected genes. Genes targeted by sequencing are shown in grey. A black bar above or below the line marks genes with detected SNVs and InDels, respectively. The 16 defined TOF genes are shown in red. The box above each affected gene indicates the number of TOF patients, which have at least one local variation in that gene. Dots below genes indicate known human cardiac phenotypes curated from the literature (Supplementary Material, Table S7). (C) Calculation of GMF with individual genotype information. An example based on 10 individuals is given. Homozygous and heterozygous mutations are denoted by 'hom' and 'het', respectively. Zero indicates the wild-type (wt) and 'N/A' if no genotype information is available. (D) Calculation of maximal GMF if no individual genotypes are available. The provided example is based on the same 10 individuals and genotypes as given in (C). As expected, the maximal GMF (0.32) is higher than the GMF (0.15).

Table 1. GMF analysis identifies genes known to cause hypertrophic cardiomyopathy

Gene	HCM patients			GMF	Reference (PMID)	Patient recruitment	NHLBI-ESP EA controls		Min. geno- types	GMF _{MAX}	GMF (HCM)/ GMF _{MAX} (controls)	P
	Unique SNVs	Affected patients	Screened patients				Filtered unique SNVs	Max. affected individuals				
<i>MYH7</i> (6087 bp)	84	125	758	0.027	–	–	37	67	4267	0.003	10.5	9.59×10^{-59}
	38	48	197	0.040	12707239	F					15.5	1.47×10^{-36}
	23	28	192	0.024	20624503	F					9.3	6.23×10^{-17}
	13	13	90	0.024	19035361	DK					9.2	5.27×10^{-9}
	12	18	88	0.034	16858239	I					13.0	2.16×10^{-14}
	10	12	80	0.025	16199542	AUS					9.6	1.24×10^{-8}
	1	1	50	0.003	16754800	USA					1.3	0.550
	2	2	46	0.007	12818575	S					2.8	0.167
	3	3	15	0.033	16267253	USA/CDN/GB					12.7	0.002
<i>TNNT2</i> (7281 bp)	13	19	758	0.003	–	–	13	22	4298	0.001	4.9	1.64×10^{-6}
	6	6	197	0.004	12707239	F					6.0	0.001
	3	6	192	0.004	20624503	F					6.1	9.68×10^{-4}
	2	3	90	0.005	19035361	DK					6.5	0.014
	2	2	88	0.003	16858239	I					4.4	0.083
	1	1	80	0.002	16199542	AUS					2.4	0.346
	1	1	50	0.003	16754800	USA					3.9	0.234
	0	0	46	–	12818575	S					–	–
	0	0	15	–	16267253	USA/CDN/GB					–	–
<i>TNNI3</i> (2032 bp)	14	19	670	0.014	–	–	6	6	2874	0.001	13.6	8.13×10^{-10}
	5	5	197	0.012	12707239	F					12.2	3.47×10^{-4}
	5	6	192	0.015	20624503	F					15.0	3.76×10^{-5}
	2	3	90	0.016	19035361	DK					16.0	0.0020
	3	3	80	0.018	16199542	AUS					18.0	0.0014
	1	1	50	0.010	16754800	USA					9.6	0.114
	0	0	46	–	12818575	S					–	–
	1	1	15	0.033	16267253	USA/CDN/GB					31.9	0.0358
<i>MYL2</i> (1362 bp)	8	8	478	0.012	–	–	5	18	4300	0.003	4.0	0.0029
	4	4	197	0.015	12707239	F					4.9	0.014
	4	3	90	0.024	19035361	DK					8.0	0.0085
	0	0	80	–	16199542	AUS					–	–
	0	0	50	–	16754800	USA					–	–
	1	1	46	0.016	12818575	S					5.2	0.183
	0	0	15	–	16267253	USA/CDN/GB					–	–
<i>ACTC1</i> (4639 bp)	2	3	281	0.002	–	–	0	0	4300	0.000	>>1	2.28×10^{-4}
	1	1	90	0.002	19035361	DK					>>1	0.0205
	0	0	80	–	16199542	AUS					–	–
	0	0	50	–	16754800	USA					–	–
	0	0	46	–	12818575	S					–	–
	1	2	15	0.029	16267253	USA/CDN/GB					>>1	1.13×10^{-5}

First line of each gene denotes the summary of all studies (given in the respective rows below). For each gene, the non-overlapping exonic length in bp is given in brackets (based on hg19/Ensembl v.72). The gene mutation frequency is normalized for the non-overlapping exonic length of the particular gene. *P*-value is based on Fisher's exact test of GMF (HCM) versus GMF_{MAX} (EA controls). Note that the important HCM gene *MYBPC3* could not be assessed due to insufficient genotype information in the EA controls. HCM, hypertrophic cardiomyopathy; GMF, gene mutation frequency; PMID, PubMed ID.

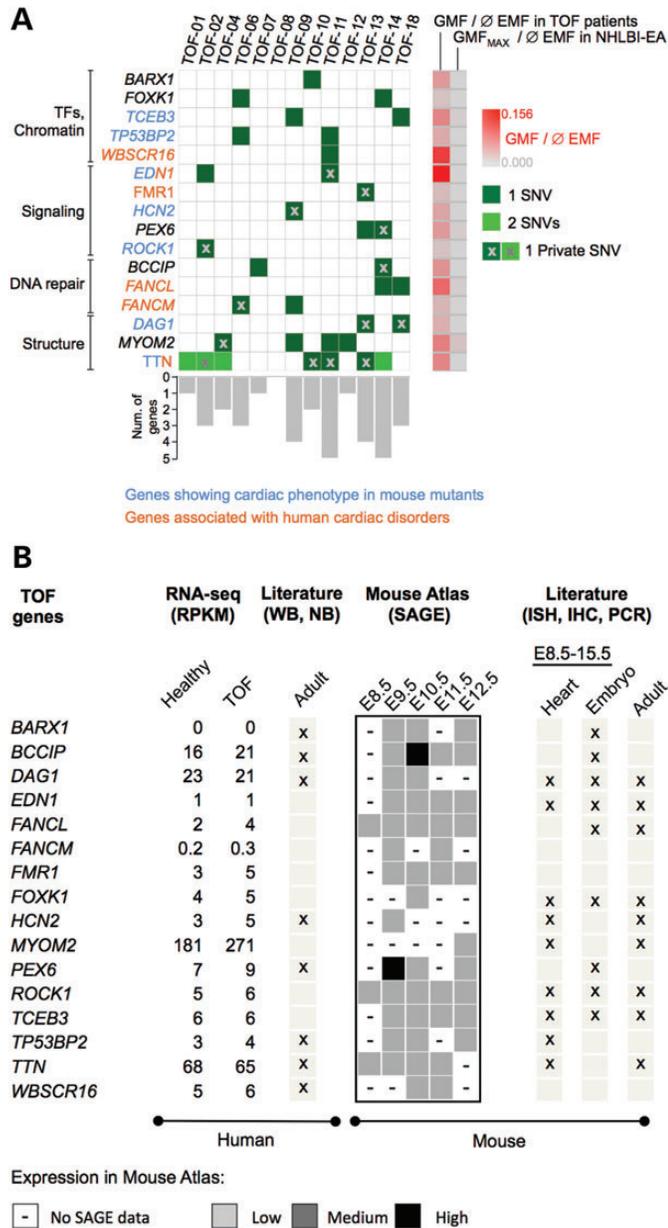


Figure 2. TOF genes and their expression in human and mouse heart. (A) Distribution of SNVs found in the 16 significantly affected TOF genes ($P < 0.05$) in TOF subjects. Private mutations are marked by 'x'. Gene-wise frequencies of SNVs are represented by grey bars. GMF in TOF cases and EA controls are indicated by a grey-to-red gradient. For *TTN*, the average exon-mutation frequency (EMF) over all significantly over-mutated exons is given. EMF, exon mutation frequency; GMF, gene mutation frequency; SNV, single nucleotide variation. (B) Cardiac expression of TOF genes in human and mouse. RNA-seq: average RPKM normalized expression levels in postnatal TOF and healthy unaffected individuals measured using mRNA-seq. Mouse Atlas: SAGE expression tag data of different developmental stages taken from Mouse Atlas. If several different heart tissues have been measured, the maximum expression is shown. SAGE level is grouped into no (0), low (1–3), medium (4–7) and high (>7) expression. Literature: availability of published mRNA or protein expression data sets in mouse heart development (E8.5–E15.5) as well as human and mouse adult hearts based on literature search (the most frequently found methods are indicated). 'Embryo' indicates that expression relates to whole embryo. The full list of data sets and corresponding publications can be found in the Supplementary Material, Table S10. RPKM, reads per kilobase per million; SAGE, serial analysis of gene expression; WB, western blot; NB, northern blot; ISH, *in-situ* hybridization; IHC, immunohistochemistry; PCR, polymerase chain reaction.

of the mutated alleles (Supplementary Material, Table S3). Of the local variations covered at least $10\times$ in mRNA-seq, 94% could be confirmed (Supplementary Material, Fig. S3 and Table S8). This underlines the functional relevance in case of deleterious mutations. In addition, all 35 SNVs observed in TOF genes (Table 2) as well as selected variations in additional affected genes (*ACADS*, *ARVCF*, *MYBPC3*) were confirmed using Sanger sequencing (Supplementary Material, Table S9).

TOF genes are expressed during development and adulthood

Since TOF is a developmental disorder, the genes causing it must have functions during embryonic development. We performed a thorough literature analysis (Supplementary Material, Table S10) and evaluated embryonic profiles using the Mouse Atlas (23) (Fig. 2B). All of these genes show a cardiac embryonic expression in at least one stage of the crucial developmental phase (E8.5–E.12.5) and the majority has a continued expression in adult heart. Based on gene expression profiles obtained by RNA-seq, we found the majority of the genes expressed (RPKM > 1) in the right ventricle of TOF patients as well as in normal adult hearts (Fig. 2B and Supplementary Material, Table S11). Taken together, this underlines the function of the TOF genes during cardiac development, promotes their causative role for TOF and suggests their potential clinical relevance during adulthood, which needs to be addressed in further genotype–phenotype studies.

Affected genes coincide in a network also showing expression disturbances

We show that combinations of private and rare deleterious mutations in multiple genes build the genetics of TOF. The different TOF genes can be classified in three main functional categories such as (i) factors for DNA repair or gene transcription either as DNA-binding transcription factors or via chromatin alterations, (ii) genes coding for proteins involved in cardiac and developmental signaling pathways, or (iii) structural components of the sarcomere (Fig. 2A). We hypothesized that these genes are functionally related and constructed an interaction network based on known protein–protein interactions. Based on the TOF genes, we expanded the network for other functionally related genes (Fig. 3A, references are given in the Supplementary Material, Table S12). This shows that several TOF genes directly interact with each other or are connected by only one intermediate gene, which provides valuable information for follow-up studies. Moreover, a number of network genes show an altered expression in particular TOF cases compared with normal heart controls (mRNA-seq). Taken together, this promotes the hypothesis that isolated TOF is caused by a set of different genes building a functional network such that alterations at the edges (affected genes) could lead to a network imbalance with the phenotypic consequence of TOF. Thus, one would expect that patients sharing affected network genes also share network disturbance. To elaborate on this, we focused on the three TOF cases TOF-04, TOF-09 and TOF-12, all harboring deleterious mutations in the gene *MYOM2*. We studied the expression profiles of the network genes in the right ventricle of these TOF cases in comparison with right ventricle samples of normal hearts (Supplementary Material, Fig. S4). We also

Table 2. SNVs found in TOF genes

GMF ratio (Ø EMF ratio ^a)	Gene	Samples	Nucleotide change	Amino acid change	MAF EA controls	Sanger validation
36.8	<i>BARX1</i>	TOF-10	c.632C>T	p.Thr211Ile	0.0009	
24.5	<i>BCCIP</i>	TOF-07 TOF-14	c.106G>A c.902T>A	p.Asp36Asn p.Met301Lys	0.0006 private	
14.1	<i>DAG1</i>	TOF-13 TOF-18	c.359T>A c.2151G>C	p.Leu120His p.Gln171His	private private	
60.1	<i>EDN1</i>	TOF-02 TOF-11	c.354G>C c.570T>G	p.Lys118Asn p.Phe190Leu	0.0001 private	
11.8	<i>FANCL</i>	TOF-18 TOF-14	c.112C>T c.685A>G	p.Leu38Phe p.Thr229Ala	0.0047 0.0007	
6.1	<i>FANCM</i>	TOF-06 TOF-09	c.3676G>A c.5101C>T	p.Asp1226Asn p.Gln1701Ter	private 0.0006	
82.7	<i>FMR1</i>	TOF-13	c.1732C>T	p.Leu578Phe	private	
5.7	<i>FOXX1</i>	TOF-06, TOF-14	c.2080G>A	p.Ala694Thr	0.0076	
30.3	<i>HCN2</i>	TOF-09	c.979C>T	p.Arg327Cys	private	
4.2	<i>MYOM2</i>	TOF-11 TOF-04	c.590C>T c.2119G>A	p.Ala197Val p.Ala707Thr	0.0016 private	
		TOF-09 TOF-12	c.3320G>C c.3904A>G	p.Gly1107Ala p.Thr1302Ala	0.0069 0.0009	
6.2	<i>PEX6</i>	TOF-14 TOF-13	c.488G>C c.1718C>T	p.Arg163Pro p.Thr573Ile	private 0.0019	
32.8	<i>ROCK1</i>	TOF-02	c.2000A>T	p.Asn667Ile	private	
9.6	<i>TCEB3</i>	TOF-18 TOF-09	c.373C>T c.1939G>A	p.Arg125Trp p.Glu647Lys	0.0002 0.0059	
14.2	<i>TP53BP2</i>	TOF-11 TOF-06	c.919A>G c.1405G>A	p.Met307Val p.Val469Ile	0.0007 0.0008	
36.2 ^a	<i>TTN</i>	TOF-01, TOF-14	c.9359G>A	p.Arg3120Gln	0.0044	
		TOF-04 TOF-02	c.30389G>A c.49150A>C	p.Arg10130His p.Thr16384Pro	0.0002 private	
		TOF-02 TOF-10	c.52852C>T c.64987C>T	p.Arg17618Cys p.Pro21663Ser	0.0019 private	
		TOF-11 TOF-13	c.65047C>G c.75035G>A	p.Pro21683Ala p.Arg25012Gln	private private	
		TOF-01, TOF-14	c.98242C>T	p.Arg32748Cys	0.0041	
		TOF-04	c.100432T>G	p.Trp33478Gly	0.0002	
110.3	<i>WBSCR16</i>	TOF-11	c.43C>T	p.Arg15Trp	0	

SNVs not seen in any cohort are marked as private. Note that *WBSCR16* is not seen in the EA controls but has an rsID in dbSNP.

^aFor TTN, the average EMF ratio of all significantly overmutated exons is given. EMF, exon mutation frequency; GMF, gene mutation frequency; MAF, minor allele frequency.

found a *MYOM2* mutation in TOF-11; however, the respective RNA-seq data had to be omitted from the analysis (see Materials and Methods). In the three TOF cases, we observe shared differential expression of *MYOM2*, *HES1*, *FANCL* and *SPI* (Fig. 3B).

When analyzing gene expression profiles of cardiac tissue samples obtained from patients with CHDs, one needs to consider that the expression profile obtained represents a postnatal status and not a developmental profile. However, the majority

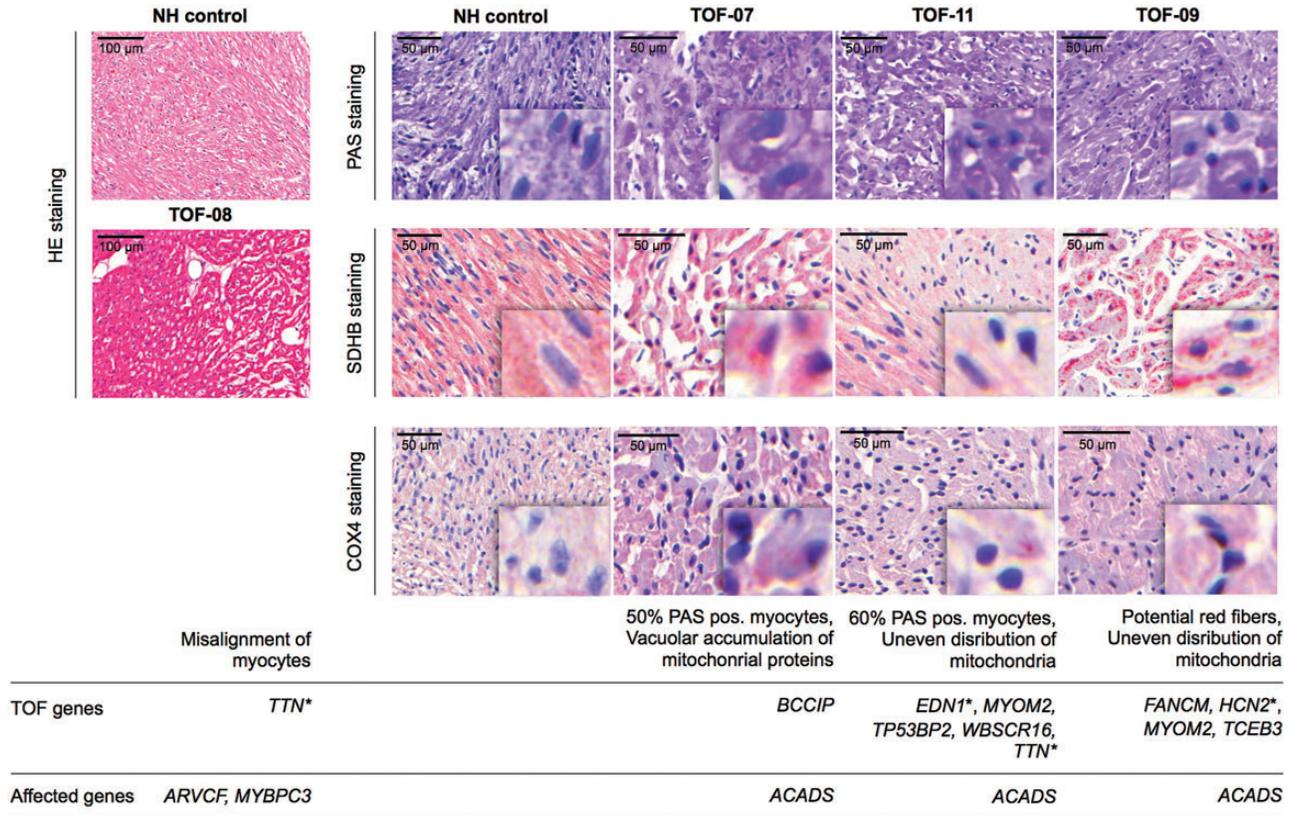


Figure 4. Genetic variations correlate with histological findings in cardiac sections of TOF patients. Histopathological assessment of right ventricular biopsies from selected TOF cases shows misalignment of the cardiac myocytes, altered PAS staining (increase of PAS-positive granules) and altered distribution of mitochondrial proteins. The image sections show 4 × magnified details of the respective pictures. Related mutations in TOF genes and affected genes of potential relevance to the phenotype are listed for each subject. Private mutations are marked with an asterisk. NH, normal heart; HE, hematoxylin and eosin; PAS, periodic acid-Schiff; SDHB, succinate dehydrogenase complex, subunit B; COX4, cytochrome c oxidase subunit IV.

of TOF genes show expression during the developmental period as well as postnatal and during adulthood (Fig. 2B). Thus, a reflection of the alterations in the protein function of respective genes in form of differential gene expression should also be detected after the developmental period. For example, a functionally relevant mutation in a transcription factor should lead to altered expression of target genes at any stage when the factor is expressed.

Genetic alterations correlate with histological findings in cardiac tissue of TOF cases

In addition to our TOF genes, other genes are affected by deleterious mutations, which are either potential modifier genes or which cannot be assessed due to insufficient genotype information in the controls at present. However, these genes might also play a role for the TOF phenotype. To assess a pathological relevance of these genes, we studied histological endomyocardial biopsy specimens of related TOF cases (Fig. 4).

TOF-08 harbors heterozygous deleterious mutations in *ARVCF* and *MYBPC3*, which have a GMF ratio of 3.1 or cannot be assessed in EA controls, respectively. The variations are located in crucial protein domains such as the Armadillo repeat region of *ARVCF* which targets the protein to the cadherin-based cellular junctions (24) and the C6 domain of

MYBPC3, which is part of a mid-region of the protein that binds to the thick filament (25). *MYBPC3* is well-known for causing cardiomyopathy (26) and knockout of *MYBPC3* in mouse results in abnormal myocardial fibers with myofibrillar disarray (27). Applying hematoxylin and eosin (HE) staining, we found a comparable disarray with an abnormal configuration of myocyte alignment with branching fibers in TOF-08 (Fig. 4), which highly promotes the causative role of these genes.

Several TOF patients show a common mutation in the mitochondrial short-chain specific acyl-CoA dehydrogenase (*ACADS*, Gly209Ser, rs1799958), which reduces enzymatic activity down to 86% but does not lead to clinically relevant deficiency on its own. However, it has been suggested that in combination with other genetic factors, this enzymatic activity could drop below the critical threshold needed for healthy functions (28,29) and thus it represents a potential modifier gene. For three of the affected patients, endomyocardial biopsies were available. All three cases show altered periodic acid-Schiff (PAS) staining, caused by an increased number of PAS-positive granules (carbohydrate macromolecules). This could be explained either by an increased glycogen storage as a result of insufficient mitochondrial activity or by an accumulation of non-degraded proteins. The latter could be caused by accumulation of the non-functional proteins in the related cases (30). Immunohistochemical stainings for mitochondrial proteins (subunit B of the succinate

dehydrogenase complex, SDHB and subunit IV of the cytochrome C oxidase, COX4) indicate loss of normal cellular distribution of mitochondria and show a similar distribution as assessed by the PAS staining (Fig. 4). Thus, our results provide evidence that variations in *ACADS* and altered mitochondrial function may modify the phenotype in these TOF cases.

DISCUSSION

We focused on a clinically in-depth characterized TOF cohort showing a homogenous phenotype and provide strong evidence that isolated TOF has a polygenic origin. To discriminate disease-related genes, we developed the novel concept of the GMF and evaluated its suitability on previously reported genomic variations in HCM patients. Applying the GMF approach to our TOF cohort resulted in 48 genes with an at least 5-fold higher GMF (EMF for *TTN*) in TOF cases than in EA controls (Supplementary Material, Fig. S2) with on average four affected genes per patient. Applying Fisher's exact test, we found 16 genes also being significantly over-mutated in TOF cases (Fig. 2A). The reduced number of genes reaching statistical significance reflects the limitation of our study by focusing on a distinct set of cases. Additional consortia studies of whole exomes in large patient collections are needed to explore the full set of variations (31).

For controls, individual genotype information is rarely available and therefore, we established the GMF_{MAX} . However, this might be higher than the real GMF (Fig. 1D) and thus relevant genes could be missed. Moreover, some genes have a low sequencing rate or quality in the controls and thus, no GMF_{MAX} could be calculated (Supplementary Material, Table S5). Especially for very long genes, both issues are problematic and thus, we developed an exon-wise approach (EMF) and show that *TTN* is also significantly altered in isolated TOF. *TTN* is a previously well-known gene for cardiomyopathy (32,33), which is of particular interest with respect to the long-term clinical outcome of patients after corrective surgery. Moreover, a recent publication for the first time showed an association of *TTN* mutations with a congenital cardiac malformation (septal defects) and the authors speculate that titin defects underlie an unsuspected number of CHD cases (34).

Our applied concept of the GMF does not weight homozygous mutations stronger than heterozygous ones as strand-specific sequence information is currently not available for most cohorts. Also, the majority of complex polygenic disorders is postulated to be caused by heterozygous mutations. However, we developed a simplified version of a chromosome-wise GMF model considering zygosity and identified exactly the same 16 significantly over-mutated genes (data not shown).

We show that individual cases harbor combinations of deleterious variations being private or rare in different genes; and different genes are affected in different cases even though they all share a well-defined coherent phenotype. The latter is frequently found in genetic disorders, examples are dilated and hypertrophic cardiomyopathy (22,35). The different genes affected in our TOF cohort can be grouped in three main functional categories and combined in an interaction network mainly built by genetically affected or differentially expressed genes (Fig. 3A). When focusing on three TOF cases sharing an affected TOF gene, we

show that this network is disturbed in a comparable manner between these cases and genes are significantly differentially expressed in comparison with healthy hearts (Fig. 3B). Thus, different genetic alterations might lead to distinct disturbances of a common interaction network, which concur to the phenotypic expression of isolated TOF. The assumption that network disturbances in general are a cause of CHD is a widely supported hypothesis (3,14,36–38).

Most of the genes in the molecular network underlying TOF are either ubiquitously expressed or characterize the two cell types contributing to the development of the right ventricle and its outflow tract, namely the NC cells and the secondary heart field (SHF) (Fig. 3A) (39–41). Notch signaling in the SHF mediates migration of the cardiac NC (42), which is crucial for appropriate outflow tract development. We show that a key member of the Notch pathway (*NOTCH1*) is affected in TOF cases. An accumulation of risk factors like local and structural variations in the molecular network underlying the outflow tract development has already been shown in CHD patients (37). Thus, the involvement of gene mutations interfering with normal development of the outflow tract is an intriguing hypothesis for the etiology of TOF, which should be further analyzed. An open hypothesis for the development of a ventricular septal defect is a premature stop of cellular growth. It is speculative if this is promoted by the involvement of genes like *TP53BP2*, *BCCIP*, *FANCM* or *FANCL* playing central roles in regulation of the cell cycle and apoptosis (43,44). The network consists of genes harboring genetic alterations as well as genes showing differential expression, such as *HES1*. *HES1* is activated by *TBX1* in the SHF (45), and we show its up-regulation in cases with deleterious mutations in *MYOM2*. The interpretation of this finding is speculative and it might be a compensatory mechanism or a primary one. Altered gene dosages as observed in CNVs are causative for a panel of developmental defects. An example is *TBX1* affected in the 22q11 deletion syndrome accounting for 15% of TOF cases (46). We did not observe genomic alterations of *TBX1* and it is not differentially expressed in our TOF cohort, which suggests that alterations in *TBX1* lead to a broader phenotype involving other organs beside the heart as described previously.

Of course, the actual disease causing effect of the disturbance of the network and the role of sequence variations and expression alterations involved await confirmation in future studies. Large-scale sequencing projects are essential to prove the network and expand it with additional genes of importance. The final functional proof will need novel techniques to be developed. The differentiation of patient-specific induced pluripotent stem cells into cardiomyocytes might be a starting point that takes into account the complex genetic background. However, the study of the process of cardiac development is only feasible with animal models; here, the different genetic background needs consideration.

Based on our findings that cardiomyopathy genes are one genetic basis of TOF, we are convinced that correlating the genetic background of TOF patients with their clinical long-term outcome harbors the opportunity to identify predictive genetic markers, which would open novel medical opportunities. Finally, we believe that the GMF is a versatile measure to identify disease causative genes and might be particularly useful to unravel complex genetic diseases.

MATERIALS AND METHODS

Subjects

Studies on patients were performed according to the institutional guidelines of the German Heart Institute in Berlin, with approval of local ethics committee, and written informed consent of patients and/or parents. Cardiac tissue samples (right ventricle) of isolated sporadic TOF cases and normal hearts as well as blood samples of TOF cases were collected in collaboration with the German Heart Institute in Berlin.

DNA was extracted from blood samples if not stated differently. Cardiac biopsies were taken from the right ventricle of patients with TOF as well as from normal human hearts during cardiac surgery after short-term cardioplegia. Samples for sequence analysis were directly snap-frozen in liquid nitrogen after excision and stored at -80°C ; samples for histology were embedded in paraffin.

DNA targeted resequencing

Three to five micrograms of gDNA were used for Roche NimbleGen sequence capturing using 365 K arrays. For array design, 867 genes and 167 microRNAs (12 910 exonic targets representing 4 616 651 target bases) were selected based on several sources as well as knowledge gained in various projects (Supplementary Material, Tables S1 and S2) (12,16,17,47). DNA enriched after NimbleGen sequence capturing was pyrosequenced for 10 TOF patients using the Genome Sequencer (GS) FLX instrument from Roche/454 Life Sciences using Titanium chemistry (~ 430 bp reads), while the remaining three samples were sequenced by Illumina Genome Analyzer (GA) Iix (36 bp paired-end reads). Sequencing was performed in-house at the Max Planck Institute for Molecular Genetics and by Atlas Biolabs according to manufacturers' protocols.

On average sequencing resulted in $\sim 14\,065\,000$ read pairs and $\sim 759\,000$ single-end reads per sample for Illumina and Roche/454, respectively. Reads resulting from Illumina sequencing were mapped to the human reference genome (GRCh37/hg19) using the BWA (48) tool v0.6.2 with 'sampe' command and default parameters. PCR duplicates were removed using Picard v1.79 (<http://picard.sourceforge.net>, accessed 10/2012). Alignments were recalibrated using GATK v2.2.2 (49). InDel realignments and base alignment quality adjustment were applied. SNV and InDel calling were performed using VarScan v2.3.2 (50) with a minimum of three supporting reads, a minimum base quality of 20 (Phred score) and a minimum variant allele frequency threshold of 0.2. Mapping as well as SNV and InDel calling for reads resulting from Roche/454 sequencing were performed using the Roche GS Reference Mapper (Newbler) v2.7.0 with default parameters resulting in high-confidence differences. On average, $\sim 12\,821\,000$ read pairs and $\sim 755\,000$ single-end reads per sample for Illumina and Roche/454, respectively, were mapped to the human reference genome (GRCh37/hg19), with high average base quality and read coverage (Supplementary Material, Fig. S5). Additional filtering of found local variations (SNVs and InDels) was performed for both techniques to ensure a minimum variant allele frequency threshold of 0.2 and a minimal coverage of 5 and 10 sequenced reads for Roche/454 and Illumina, respectively.

SNV and InDel filtering

SNVs and InDels gathered from re-sequencing and SNVs from exomes of 4300 European-American unrelated individuals (EA controls) sequenced within the Exome Sequencing Project at the National Heart, Lung, and Blood Institute [Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (<http://evs.gs.washington.edu/EVS/>, accessed 10/2012)] as well as 200 Danish controls (19) were annotated using SeattleSeqAnnotation137 (51) and PolyPhen-2 (52). We filtered for local variations predicted to be missense, non-sense, frame-shifting or affecting splice sites. Only those missense SNVs were retained which were predicted to be damaging, while tolerated variations were discarded. The filtered variations were subsequently reduced to novel variations or variations with a MAF of ≤ 0.01 in dbSNP (v137), UCSC 'snp137' track (MAF extrapolated by dbSNP from submitted frequencies), 498 parents sequenced within the Genome of the Netherlands (GoNL, release 2), one of the projects within Biobanking and Biomolecular Research Infrastructure-Netherlands (BBMRI-NL) (53) and NHLBI-ESP-EA controls. Known disease-associated variations present in the OMIM database were retained. Individual filtering steps are described in the Supplementary Material, Figure S1.

Statistical assessment of TOF relevant genes—'TOF genes'

The majority of our samples were sequenced using Roche's platform; however, three samples were sequenced with Illumina's GAIIx. Since these platforms show differences in the detection of InDels, we only focused on SNVs for the statistical assessment of TOF-relevant genes. Genes showing a significantly higher SNV rate in TOF subjects compared with controls were assessed using a one-sided Fisher's exact test without correction for multiple testing, meaning that the observed ratio of each gene's mutation frequency in TOF cases compared with controls was computed. Genes with a minimal *P*-value of 0.05 in TOF cases versus EA controls were defined as 'TOF genes'. For TOF cases and Danish controls, the GMF was calculated based on the number of individuals harboring SNVs in relationship to the total number of individuals with sufficient genotype information (Fig. 1C). Reasons for insufficient genotype information about wild-type, homozygous SNV or heterozygous SNV at a particular base are low-sequencing coverage and low-sequencing quality. For EA controls, no individual genotype information was provided and therefore, the maximal GMF (GMF_{MAX}) was calculated, based on the maximal number of individuals with SNVs (Fig. 1D). For *TTN*, the EMF was calculated using the GMF formula with two adjustments, i.e. instead of the whole gene, the calculation is based on single exons and instead of a kilobase-scaling, the EMF is 100 bp scaled accounting for the shorter size of exons compared with genes.

mRNA sequencing

mRNAs were isolated from total RNA and prepared for sequencing using the Illumina Kit RS-100-0801, according to the manufacturer's protocol. Sequencing libraries were generated using a non-strand specific library construction method. Purified DNA fragments were used directly for cluster generation, and 36 bp

single-end read sequencing was performed using Illumina Genome Analyzer. Sequencing reads were extracted from the image files using the open source Firecrest and Bustard applications (Solexa pipeline 1.5.0). Deep sequencing of mRNA libraries produced ~19 224 000 reads per sample on average.

mRNA reads were mapped to the human reference genome (NCBI v36.1; hg18) using RazerS (54) allowing at most 10 equally best hits and two mismatches (no InDels) per read. Finally, ~14 736 000 single-end reads per sample for mRNA were mapped on average to the whole human reference genome. On average ~9 431 000 (64%) reads per sample could be mapped to unique genomic locations and ~5 304 000 (36%) reads matched to multiple regions (2–10 genomic locations). Multi-matched reads were proportionally assigned to each of their mapping locations using MuMRescueLite (55) with a window size of 200 bp. Reads were assigned to genes and transcripts if their mapped location is inside of exon boundaries as defined by ENSEMBL (v54). To further assign unmapped reads, a gene-wise splice junction sequence library was produced from pairwise connection of exon sequences corresponding to all known 5' to 3' splice junctions (supported by the analysis of aligned EST and cDNA sequences). For transcripts, the read counts were adjusted using the proportion estimation method in the Solas package (56). For quality assessment, manual inspection of multi-dimensional scaling plots and existence of pile-up effects were performed, leading to the exclusion of four samples (TOF-11, TOF-14, TOF-18, TOF-19) for gene expression analysis. The read counts were RPKM (reads per kilobase transcript per million reads) normalized. To define differential expression between healthy and affected individuals, a *t*-test based on the RPKM normalized gene expression levels was performed.

Validation of genomic variations by Sanger sequencing

PCR reactions were carried out using gDNA templates and standard protocols (primer sequences are available on request) and Sanger sequenced in-house at the Experimental and Clinical Research Center.

Histopathology

Paraffin-embedded right ventricular biopsies of TOF cases were subjected to histochemical HE and PAS stainings. In addition, immunohistochemical stainings for two components of the mitochondrial respiratory chain (SDHB and COX4) were performed for selected samples with the use of rabbit polyclonal antibodies from LifeSpan Biosciences (LS-C143581 and LS-C119480, respectively). As a control, a normal homograft heart of a 4-month-old infant who died of a non-cardiac cause was used. All stainings were carried out using standard protocols and 3- μ m tissue slices.

Statistics

General bioinformatics and statistical analyses were conducted using R (including Bioconductor packages) and Perl. Given *P*-values are nominal (not adjusted for multiple testing). Multiple correction is only needed if thousands of hypotheses are tested simultaneously (multiplicity problem) because this significantly increases the chance of false positives. As we performed targeted resequencing of 867 genes instead of whole

exome sequencing (~25 000 genes) and furthermore, only 121 genes are affected by SNVs and were tested for significantly higher GMFs; no correction for multiple testing is needed.

Accession numbers

mRNA-seq data are available from the Gene Expression Omnibus (GEO) repository at NCBI (accession number GSE36761).

SUPPLEMENTARY MATERIAL

Supplementary material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are deeply grateful to the patients and families for their cooperation. We thank Katherina Bellmann for the help of assessing local variations found in the TOF patients and Andrea Behm for technical assistance. We thank Robert Kelly for his intellectual input regarding the molecular network. The authors would like to thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010).

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the European Community's Sixth and Seventh Framework Programme contracts ('HeartRepair') LSHM-CT-2005-018630 and ('CardioGeNet') 2009-223463 and ('CardioNet') People-2011-ITN-289600 (all to S.R.S.) a PhD scholarship to C.D. by the Studienstiftung des Deutschen Volkes, and the German Research Foundation (Heisenberg professorship and grant 574157 to S.R.S.).

REFERENCES

- Hoffman, J.I.E. and Kaplan, S. (2002) The incidence of congenital heart disease. *J. Am. Coll. Cardiol.*, **39**, 1890–1900.
- Nora, J.J. (1968) Multifactorial inheritance hypothesis for the etiology of congenital heart diseases. The genetic-environmental interaction. *Circulation*, **38**, 604–617.
- Fahed, A.C., Gelb, B.D., Seidman, J.G. and Seidman, C.E. (2013) Genetics of congenital heart disease: the glass half empty. *Circ. Res.*, **112**, 707–720.
- Ferencz, C., Rubin, J.D., McCarter, R.J., Brenner, J.I., Neill, C.A., Perry, L.W., Hepner, S.I. and Downing, J.W. (1985) Congenital heart disease: prevalence at livebirth. The Baltimore-Washington Infant Study. *Am. J. Epidemiol.*, **121**, 31–36.
- Apitz, C., Webb, G.D. and Redington, A.N. (2009) Tetralogy of Fallot. *Lancet*, **374**, 1462–1471.
- Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., Ichida, F., Joo, K., Kimura, M., Imamura, S.-I. *et al.* (2003) Role of TBX1 in human del22q11.2 syndrome. *Lancet*, **362**, 1366–1373.
- Korenberg, J.R., Chen, X.N., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, C., Dignan, P. and Disteche, C. (1994) Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl Acad. Sci. USA*, **91**, 4997–5001.

8. McDaniell, R., Warthen, D.M., Sanchez-Lara, P.A., Pai, A., Krantz, I.D., Piccoli, D.A. and Spinner, N.B. (2006) NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am. J. Hum. Genet.*, **79**, 169–173.
9. Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soultis, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J. *et al.* (1997) Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.*, **15**, 30–35.
10. Michielon, G., Marino, B., Formigari, R., Gargiulo, G., Picchio, F., Digilio, M.C., Anaclerio, S., Oricchio, G., Sanders, S.P. and Di Donato, R.M. (2006) Genetic syndromes and outcome after surgical correction of tetralogy of Fallot. *Ann. Thorac. Surg.*, **81**, 968–975.
11. Eldadah, Z.A., Hamosh, A., Biery, N.J., Montgomery, R.A., Duke, M., Elkins, R. and Dietz, H.C. (2001) Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum. Mol. Genet.*, **10**, 163–169.
12. Greenway, S.C., Pereira, A.C., Lin, J.C., DePalma, S.R., Israel, S.J., Mesquita, S.M., Ergul, E., Conta, J.H., Korn, J.M., McCarroll, S.A. *et al.* (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat. Genet.*, **41**, 931–935.
13. Soemedi, R., Töpf, A., Wilson, I.J., Darlay, R., Rahman, T., Glen, E., Hall, D., Huang, N., Bentham, J., Bhattacharya, S. *et al.* (2011) Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. *Hum. Mol. Genet.*, **10.1093/hmg/ddr589**.
14. Sperling, S.R. (2011) Systems biology approaches to heart development and congenital heart disease. *Cardiovasc. Res.*, **91**, 269–278.
15. Cohen, J.C., Kiss, R.S., Pertsemliadis, A., Marcel, Y.L., McPherson, R. and Hobbs, H.H. (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*, **305**, 869–872.
16. Kaynak, B., von Heydebreck, A., Mebus, S., Seelow, D., Hennig, S., Vogel, J., Sperling, H.-P., Pregla, R., Alexi-Meskishvili, V., Hetzer, R. *et al.* (2003) Genome-wide array analysis of normal and malformed human hearts. *Circulation*, **107**, 2467–2474.
17. Toenjes, M., Schueler, M., Hammer, S., Pape, U.J., Fischer, J.J., Berger, F., Vingron, M. and Sperling, S. (2008) Prediction of cardiac transcription networks based on molecular data and complex clinical phenotypes. *Mol. Biosyst.*, **4**, 589–598.
18. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J. *et al.* (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, **493**, 216–220.
19. Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliusson, T. *et al.* (2010) Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat. Genet.*, **42**, 969–972.
20. Marth, G.T., Yu, F., Indap, A.R., Garimella, K., Gravel, S., Leong, W.F., Tyler-Smith, C., Bainbridge, M., Blackwell, T., Zheng-Bradley, X. *et al.* (2011) The functional spectrum of low-frequency coding variation. *Genome Biol.*, **12**, R84.
21. Sperling, S., Grimm, C.H., Dunkel, I., Mebus, S., Sperling, H.-P., Ebner, A., Galli, R., Lehrach, H., Fusch, C., Berger, F. *et al.* (2005) Identification and functional analysis of CITED2 mutations in patients with congenital heart defects. *Hum. Mutat.*, **26**, 575–582.
22. Maron, B.J. and Maron, M.S. (2013) Hypertrophic cardiomyopathy. *Lancet*, **381**, 242–255.
23. Siddiqui, A.S., Khattra, J., Delaney, A.D., Zhao, Y., Astell, C., Asano, J., Babakaiff, R., Barber, S., Beland, J., Bohacec, S. *et al.* (2005) A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proc. Natl Acad. Sci. USA*, **102**, 18485–18490.
24. Kaufmann, U., Zuppinger, C., Waibler, Z., Rudiger, M., Urbich, C., Martin, B., Jockusch, B.M., Eppenberger, H. and Starzinski-Powitz, A. (2000) The armadillo repeat region targets ARVCF to cadherin-based cellular junctions. *J. Cell. Sci.*, **113**, 4121–4135.
25. Moolman-Smook, J., Flashman, E., de Lange, W., Li, Z., Corfield, V., Redwood, C. and Watkins, H. (2002) Identification of novel interactions between domains of Myosin binding protein-C that are modulated by hypertrophic cardiomyopathy missense mutations. *Circ. Res.*, **91**, 704–711.
26. Marston, S., Copeland, O., Gehmlich, K., Schlossarek, S., Carrier, L. and Carrier, L. (2012) How do MYBPC3 mutations cause hypertrophic cardiomyopathy? *J. Muscle Res. Cell. Motil.*, **33**, 75–80.
27. McConnell, B.K., Jones, K.A., Fatkin, D., Arroyo, L.H., Lee, R.T., Aristizabal, O., Turnbull, D.H., Georgakopoulos, D., Kass, D., Bond, M. *et al.* (1999) Dilated cardiomyopathy in homozygous myosin-binding protein-C mutant mice. *J. Clin. Invest.*, **104**, 1235–1244.
28. Pedersen, C.B., Kølvrå, S., Kølvrå, A., Stenbroen, V., Kjeldsen, M., Ensenauer, R., Tein, I., Matern, D., Rinaldo, P., Vianey-Saban, C. *et al.* (2008) The ACADS gene variation spectrum in 114 patients with short-chain acyl-CoA dehydrogenase (SCAD) deficiency is dominated by missense variations leading to protein misfolding at the cellular level. *Hum. Genet.*, **124**, 43–56.
29. Corydon, M.J., Vockley, J., Rinaldo, P., Rhead, W.J., Kjeldsen, M., Winter, V., Riggs, C., Babovic-Vuksanovic, D., Smeitink, J., De Jong, J. *et al.* (2001) Role of common gene variations in the molecular pathogenesis of short-chain acyl-CoA dehydrogenase deficiency. *Pediatr. Res.*, **49**, 18–23.
30. Rauch, R., Hofbeck, M., Zweier, C., Koch, A., Zink, S., Trautmann, U., Hoyer, J., Kaulitz, R., Singer, H. and Rauch, A. (2010) Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot. *J. Med. Genet.*, **47**, 321–331.
31. Pediatric Cardiac Genomics Consortium. (2013) The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. *Circ. Res.*, **112**, 698–706.
32. Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitás, K., Sasse-Klaassen, S., Seidman, J.G., Seidman, C., Granzier, H., Labeit, S. *et al.* (2002) Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat. Genet.*, **30**, 201–204.
33. Herman, D.S., Lam, L., Taylor, M.R.G., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S.R., McDonough, B., Sparks, E. *et al.* (2012) Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.*, **366**, 619–628.
34. Chauveau, C., Bonnemant, C.G., Julien, C., Kho, A.L., Marks, H., Talim, B., Maury, P., Arne-Bes, M.C., Uro-Coste, E., Alexandrovich, A. *et al.* (2014) Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. *Hum. Mol. Genet.*, **23**, 980–991.
35. McNally, E.M., Golbus, J.R. and Puckelwartz, M.J. (2013) Genetic mutations and mechanisms in dilated cardiomyopathy. *J. Clin. Invest.*, **123**, 19–26.
36. Andersen, T.A., de Troelsen, K.L.L. and Larsen, L.A. (2013) Of mice and men: molecular genetics of congenital heart disease. *Cell. Mol. Life Sci.*, **10.1007/s00018-013-1430-1**.
37. Lage, K., Greenway, S.C., Rosenfeld, J.A., Wakimoto, H., Gorham, J.M., Segrè, A.V., Roberts, A.E., Smoot, L.B., Pu, W.T., Pereira, A.C. *et al.* (2012) Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proc. Natl Acad. Sci. USA*, **109**, 14035–14040.
38. Rana, M.S., Christoffels, V.M. and Moorman, A.F.M. (2013) A molecular and genetic outline of cardiac morphogenesis. *Acta Physiol. (Oxf.)*, **207**, 588–615.
39. Hutson, M.R. and Kirby, M.L. (2003) Neural crest and cardiovascular development: a 20-year perspective. *Birth Defects Res. C Embryo Today*, **69**, 2–13.
40. Thomas, T., Kurihara, H., Yamagishi, H., Kurihara, Y., Yazaki, Y., Olson, E.N. and Srivastava, D. (1998) A signaling cascade involving endothelin-1, dHAND and msx1 regulates development of neural-crest-derived branchial arch mesenchyme. *Development*, **125**, 3005–3014.
41. Kelly, R.G. (2012) The Second Heart Field. In *Heart Development, Current Topics in Developmental Biology*. Elsevier, Vol. 100, pp. 33–65.
42. Jain, R., Engleka, K.A., Rentschler, S.L., Manderfield, L.J., Li, L., Yuan, L. and Epstein, J.A. (2011) Cardiac neural crest orchestrates remodeling and functional maturation of mouse semilunar valves. *J. Clin. Invest.*, **121**, 422–430.
43. Lu, H., Huang, Y.-Y., Mehrotra, S., Droz-Rosario, R., Liu, J., Bhaumik, M., White, E. and Shen, Z. (2011) Essential roles of BCCIP in mouse embryonic development and structural stability of chromosomes. *PLoS Genet.*, **7**, e1002291.
44. Vives, V., Su, J., Zhong, S., Ratnayaka, I., Slee, E., Goldin, R. and Lu, X. (2006) ASPP2 is a haploinsufficient tumor suppressor that cooperates with p53 to suppress tumor growth. *Genes Dev.*, **20**, 1262–1267.
45. Vincent, S.D. and Buckingham, M.E. (2010) How to make a heart: the origin and regulation of cardiac progenitor cells. *Curr. Top. Dev. Biol.*, **90**, 1–41.
46. Goldmuntz, E. (2005) DiGeorge syndrome: new insights. *Clin. Perinatol.*, **32**, 963-78-ix-x.
47. Schlesinger, J., Schueler, M., Grunert, M., Fischer, J.J., Zhang, Q., Krueger, T., Lange, M., Tönjes, M., Dunkel, I. and Sperling, S.R. (2011) The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS Genet.*, **7**, e1001313.

48. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
49. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
50. Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K. and Ding, L. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, **25**, 2283–2285.
51. Cooper, G.M., Goode, D.L., Ng, S.B., Sidow, A., Bamshad, M.J., Shendure, J. and Nickerson, D.A. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods*, **7**, 250–251.
52. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
53. Boomsma, D.I., Wijmenga, C., Slagboom, E.P., Swertz, M.A., Karssen, L.C., Abdellaoui, A., Ye, K., Guryev, V., Vermaat, M., van Dijk, F. *et al.* (2014) The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.*, **22**, 221–227.
54. Weese, D., Emde, A.-K., Rausch, T., Döring, A. and Reinert, K. (2009) RazerS—fast read mapping with sensitivity control. *Genome Res.*, **19**, 1646–1654.
55. Hashimoto, T., de Hoon, M.J.L., Grimmond, S.M., Daub, C.O., Hayashizaki, Y. and Faulkner, G.J. (2009) Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRRescueLite. *Bioinformatics*, **25**, 2613–2614.
56. Richard, H., Schulz, M.H., Sultan, M., Nürnberger, A., Schrinner, S., Balzereit, D., Dagand, E., Rasche, A., Lehrach, H., Vingron, M. *et al.* (2010) Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res.*, **38**, e112.