Dissertation

Submitted to the

Combined Faculty of Natural Sciences and Mathematics

Of the Ruperto Carola University Heidelberg, Germany

For the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Simon Alexander Sumer

Born in: Augsburg, Germany

Oral examination: 29.09.2020

Deciphering the role of *SHOX2* in atrial fibrillation and sinus node dysfunction

Referees: Prof. Dr. Gudrun A. Rappold
Prof. Dr. Johannes Backs

Publications, Presentations & Awards

The work carried out during my PhD has resulted in the following publications

- Hoffmann, S., Paone, C., **Sumer, S.**, Diebold, S., Weiss, B., Roeth, R., Clauss, S., Klier, I., Kääb, S., Schulz, A., Wild, P. S., Ghrib, A., Zeller, T., Schnabel, R. B., Just, S. & Rappold, G. A. 2019. Functional Characterization of Rare Variants in the SHOX2 Gene Identified in Sinus Node Dysfunction and Atrial Fibrillation. *Front Genet,* 10**,** 648.

- Hoffmann, S., Schmitteckert, S., Griesbeck, A., Preiss, H., **Sumer, S.**, Rolletschek, A., Granzow, M., Eckstein, V., Niesler, B. & Rappold, G. A. 2017. Comparative expression analysis of Shox2-deficient embryonic stem cell-derived sinoatrial node-like cells. *Stem Cell Res,* 21**,** 51-57.

- Hoffmann, S., Paone, C., **Sumer, S.A.**, Diebold, S., Weiss, B., Roeth, R., Clauss, S., Klier, I., Kääb, S., Schulz, A., Wild, P. S., Ghrib, A., Zeller, T., Schnabel, R. B., Just, S. & Rappold, G. A. 2019. Functional Characterization of Rare Variants in the SHOX2 Gene Identified in Sinus Node Dysfunction and Atrial Fibrillation. *Front Genet,* 10**,** 648.

- **Sumer S.A.**, Hoffmann S., Laue S., Campbell B., Raedecke K., Frajs V., Clauss S., Kääb S. Janssen W.G., Jauch A, Laugwitz K.-L. Dorn T., Moretti A., Rappold G.A. Precise correction of heterozygous *SHOX2* mutations in hiPSCs derived from patients with atrial fibrillation via genome editing and sib-selection (under revision)

Different parts of this work have been presented in form of oral presentations and posters on annual local retreats, national and international conferences, for example:

- Herztage, German Society of Cardiology (DGK) – Poster presentation
  October 6-8, 2016 – Berlin (Germany)

- 29th Annual Meeting of the German Society of Human Genetics – Poster presentation
  March 14- 16, 2018 - Münster (Germany)

- 7th Annual German Stem Cell Network (GSCN) Conference – Invited speaker
  September 23-25, 2019 – Berlin (Germany)

- European Society of Human Genetics (ESHG) Virtual Conference- Poster presentation
  June 6-9, 2020 – Virtual Conference

The following price was won with this work:

Cardiomyopathy poster award + selected talk
DZHK partner site retreat 2019 Heidelberg/Mannheim

Acknowledgements

The experimental data presented in this thesis was a collaborative effort with colleagues, other institutions, and core facilities.

Functional Characterization of Rare Variants in the SHOX2 Gene Identified in Sinus Node Dysfunction and Atrial Fibrillation, as published[1]:

Sandra Hoffmann[(1)(2)] designed and performed the experiments, analyzed the data, and wrote the manuscript; Christoph Paone[(3)], Simon Alexander Sumer[(1)(2)], Sabrina Diebold[(3)], Birgit Weiß[(1)], and Ralph Roeth[(1)] performed experiments; Sebastian Clauss[(4)(5)], Ina Klier[(4)(5)], Stefan Kääb[(4)(5)], Tanja Zeller[(7)(8)], and Renate B. Schnabel[(7)(8)] provided material or support; Andreas Schulz[(6)], Phillip S. Wild[(6)], Adil Ghrib[(7)(8)], and Sebastian Clauss[(4)(5)] analyzed the data; Steffen Just[(3)] and Gudrun A. Rappold[(1)(2)] designed the study, analyzed the data, and wrote the manuscript.

Precise correction of heterozygous SHOX2 mutations in hiPSCs derived from patients with atrial fibrillation via genome editing and sib-selection:

Simon Alexander Sumer[(1)(2)] and Sandra Hoffmann[(1)(2)] conceived the project, performed the experiments and analyzed the data along with Svenja Laue[(5)(9)], Birgit Campbell[(9)], Kristin Raedecke[(1)(2)], Viktoria Frajs[(1)] and Tatjana Dorn[(5)(9)]; Sebastian Clauss[(4)(5)] and Stefan Kääb[(4)(5)] recruited the patients and provided clinical records; Johannes W.G. Janssen[(1)] and Anna Jauch[(1)] performed karyotyping; Gudrun A. Rappold[(1)(2)] and , Alessandra Moretti[(5)(9)] supervised the project and helped with data interpretation, Simon Alexander Sumer[(1)(2)] wrote the manuscript with support from Sandra Hoffmann[(1)(2)], Tatjana Dorn[(5)(9)], Karl-Ludwig Laugwitz[(5)(9)], Alessandra Moretti[(5)(9)], Gudrun A. Rappold[(1)(2)] and revisions from all co-authors; funding was acquired by Sandra Hoffmann[(1)(2)], , Alessandra Moretti[(5)(9)], Gudrun A. Rappold[(1)(2)], and Karl-Ludwig Laugwitz[(5)(9)].

Technical support for ddPCR™ was given by Dr. Kathleen Börner and Prof. Dirk Grimm (Virology, Department of Infectious Diseases, Heidelberg University Hospital, Heidelberg, Germany).

[(1)] Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany

[(2)] DZHK (German Centre for Cardiovascular Research), Partner site Heidelberg/Mannheim, Heidelberg, Germany

[(3)] Department of Internal Medicine II, University of Ulm, Ulm, Germany

[(4)] Department of Medicine I, Klinikum Grosshadern, University of Munich (LMU), Munich, Germany

[(5)] DZHK (German Centre for Cardiovascular Research), Partner site Munich, Munich, Germany

[(6)] Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

[(7)] Department of General and Interventional Cardiology, University

Heart Center Hamburg (UHZ), University Hospital Hamburg/Eppendorf, Hamburg, Germany

[(8)] DZHK (German Centre for Cardiovascular Research), Partner site Hamburg/Kiel/Luebeck, Hamburg, Germany

[(9)] First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany

Abstract

The sinoatrial node (SAN) is the natural pacemaker of the heart and initiates the rhythmic contractions of this organ. Its unique genetic profile is mediated by a network of transcriptional regulators. Among them is the homeodomain transcription factor *SHOX2*, which plays a major role in maintaining the phenotypic border between the SAN and the surrounding tissue. Mutations in this gene have been associated with early-onset and familial forms of Atrial Fibrillation (AF). AF is the most common cardiac rhythm disorder, affecting 1-2% of the general population. In the clinical context, it often co-exists with malfunctions of the sinus node (sinus node dysfunction, SND), however, it is unknown if both diseases interact, perpetuate, or initiate each other.

In the first part of this project, a candidate gene study was combined with functional analyses to identify a causal relationship between novel *SHOX2* gene variants and the development of AF and SND. Screening 98 SND patients and 450 individuals with AF led to the identification of four heterozygous variants in *SHOX2* (p.P33R in the SND cohort and p.G77D, p.L129=, p.L130F, p.A293= in the AF cohort). We selected mutations based on their *in silico* predicted pathogenic potential and overexpressed them in embryonic zebrafish hearts. A dominant-negative effect leading to bradycardia and pericardial edema was detected for p.G77D, while no effect was revealed for the p.P33R and p.A293= variants. A significantly impaired transactivation activity for both missense variants p.P33R and p.G77D was demonstrated by *in vitro* reporter assays. Moreover, upon overexpression of the p.P33R mutant in zebrafish hearts, a reduced *Bmp4* target gene expression was revealed. This study demonstrated for the first time a genetic link between SND and AF involving *SHOX2*.

Patient-specific human induced pluripotent stem cells (iPSCs) harboring putative disease-causing variants offer unprecedented opportunities for the investigation of cardiovascular diseases. We generated and characterized iPSCs from patients with previously identified heterozygous *SHOX2* mutations (*SHOX2* c.849C>A and *SHOX2* c.*28T>C). To establish an isogenic control, we developed a novel strategy for the scarless correction of heterozygous mutations. Patient-derived iPSCs were gene-edited with the CRISPR/Cas system and subdivided into small cell pools (sib-selection). We quantified wildtype and mutant alleles via digital PCR and next generation sequencing to detect shifts in the wildtype/mutant allele ratio that indicated the presence of gene-corrected cells. Using this method, we managed to enrich our target cells 8-10-fold before generating a monoclonal cell population via single-cell cloning.

The recharacterization of the new lines confirmed a preserved pluripotency and a normal karyotype. Future electrophysiological and molecular analysis will give further insights into the contribution of *SHOX2* to onset and progression of AF.

Zusammenfassung

Der sinoatriale Knoten (SAN) ist der natürliche Schrittmacher des Herzens und leitet die rhythmischen Kontraktionen dieses Organs ein. Sein einzigartiges genetisches Profil wird durch ein Netzwerk von Transkriptionsregulatoren vermittelt. Zu ihnen gehört der Homöodomänen-Transkriptionsfaktor *SHOX2*, der eine wichtige Rolle bei der Aufrechterhaltung der phänotypischen Grenze zwischen dem SAN und dem umgebenden Gewebe spielt. Mutationen in diesem Gen wurden mit früh einsetzenden und familiären Formen des Vorhofflimmerns (AF) in Verbindung gebracht. Vorhofflimmern ist die häufigste Herzrhythmusstörung und betrifft 1-2% der Allgemeinbevölkerung. Im klinischen Zusammenhang existiert sie häufig zusammen mit Fehlfunktionen des Sinusknotens (Sinusknotendysfunktion, SND), es ist jedoch unbekannt, ob beide Erkrankungen interagieren, sich gegenseitig aufrechterhalten oder auslösen.

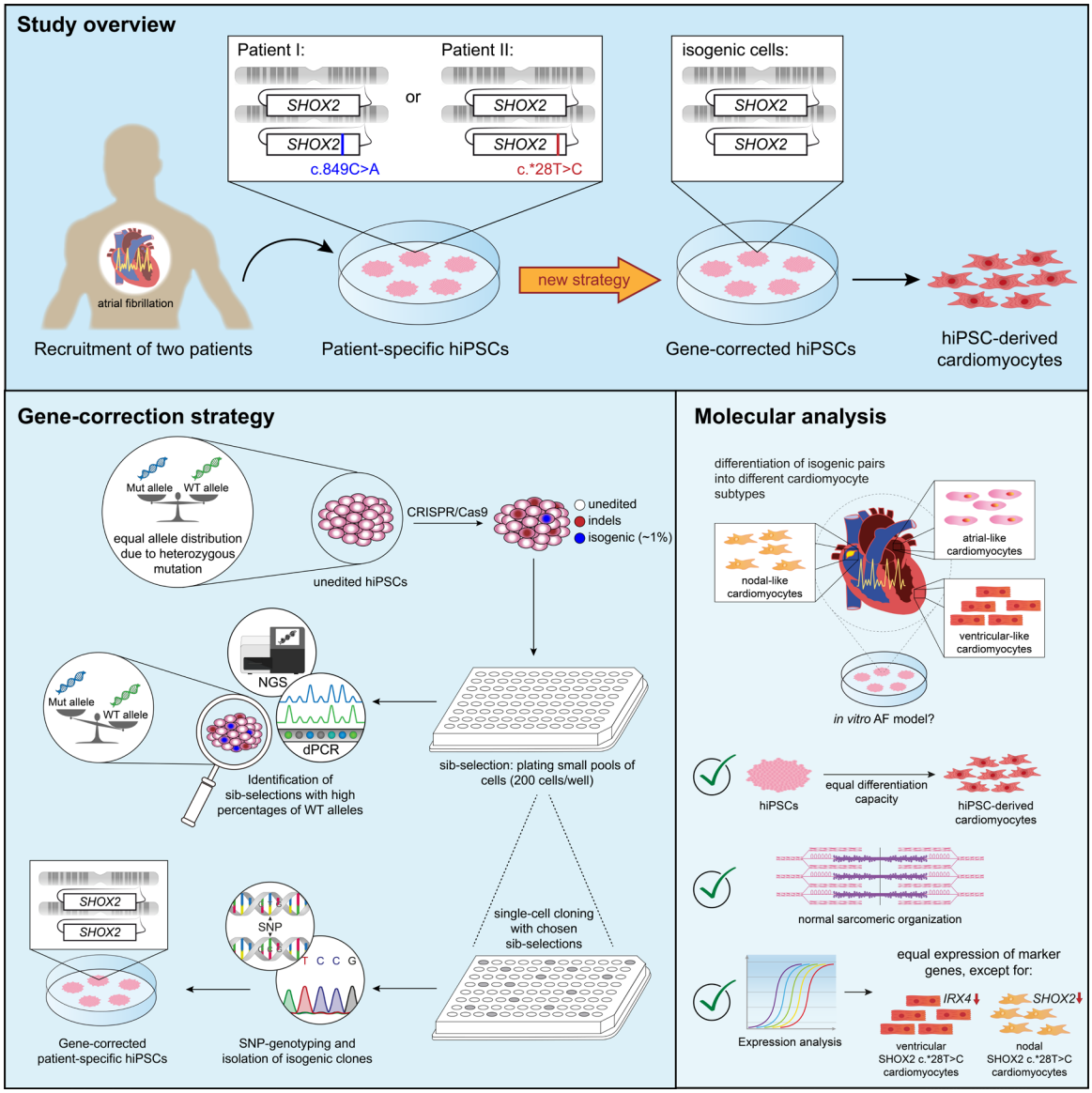
Im ersten Teil dieses Projekts wurde eine Kandidatengenstudie mit funktionellen Analysen kombiniert, um einen kausalen Zusammenhang zwischen neuartigen *SHOX2*-Genvarianten und der Entwicklung von AF und SND zu identifizieren. Bei der Überprüfung von 98 SND Patienten und 450 Individuen mit VHF wurden vier heterozygote Varianten in *SHOX2* detektiert (p.P33R in der SND-Kohorte und p.G77D, p. L129=, p.L130F, p.A293= in der Vorhofflimmerkohorte). Wir wählten Mutationen basierend auf *in silico* Vorhersagen über ihr pathogenes Potential aus und überexprimierten sie in embryonalen Zebrafischherzen. Ein dominant-negativer Effekt, der zur Bradykardie und perikardialen Ödemen führte, wurde für p.G77D gefunden, während p.P33R und p.A293= keinen Effekt zeigten. Eine signifikant beeinträchtigte Transaktivierungsaktivität für die beiden Missense-Varianten p.P33R und p.G77D wurde durch *in vitro* Reporter Assays nachgewiesen. Darüber hinaus wurde bei Überexpression der p.P33R-Mutante in Zebrafischherzen eine reduzierte Expression des *Bmp4* Zielgens festgestellt. Diese Studie zeigt zum ersten Mal eine genetische Verbindung zwischen SND und AF unter Beteiligung von *SHOX2*.

Patientenspezifische humane induzierte pluripotente Stammzellen (iPSCs), die potenziell krankheitsverursachende Varianten besitzen, bieten unzählige Möglichkeiten für die Untersuchung von Herz-Kreislauf-Erkrankungen. Wir generierten und charakterisierten iPSCs von Patienten mit zuvor identifizierten heterozygoten *SHOX2*-Mutationen (*SHOX2* c.849C>A und *SHOX2* c.*28T>C). Um eine isogene Kontrolle zu etablieren, entwickelten wir eine neuartige Strategie zur narbenlosen Korrektur heterozygoter Mutationen. Von Patienten stammende iPSCs wurden mit dem CRISPR/Cas-System geneditiert und in kleine Zellpools unterteilt (Sib-Selektion). Wir quantifizierten Wildtyp- und mutierte Allele mittels digitaler PCR und Next-Generation-Sequenzierung, um Verschiebungen im Wildtyp/Mutanten-Allel-Verhältnis zu erkennen, die auf das Vorhandensein von genkorrigierten Zellen hindeuteten. Mit dieser Methode gelang es uns,

unsere Zielzellen 8- bis 10-fach anzureichern, bevor wir durch Einzelzellklonierung eine monoklonale Zellpopulation erzeugten.

Die Re-Charakterisierung der neuen Linien bestätigte eine erhaltene Pluripotenz und einen normalen Karyotyp. Zukünftige elektrophysiologische und molekulare Analysen werden weitere Einblicke in den Beitrag von *SHOX2* zur Entstehung und Progression von Vorhofflimmern geben.

Graphical abstract



Sumer et al. (under revision)

Table of Contents

List of Figures

List of Tables

Abbreviations

°C *Degree Celsius*
3'UTR *3' untranslated region*
AADs *Antiarrhythmic Drugs*
ACM *Arrhythmogenic Cardiomyopathy*
AF *Atrial Fibrillation*
ALP *Alkaline Phosphatase*
AVB *Atrioventricular Bundle*
AVN *Atrioventricular Node*
BBs *Bundle Branches*
BTHS *Barth Syndrome*
CADD *Combined Annotation Dependent Depletion*
CCS *Cardiac Conduction System*
CNV *Copy Number Variation*
CRISPR *Clustered Regularly Interspaced Short Palindromic Repeats*
dCas9 *deactivated Cas9*
DCM *Dilated Cardiomyopathy*
ddH_2O *Double-Distilled Water*
DMEM *Dulbecco's Modified Eagle's Medium*
DMSO *Dimethylsulfoxide*
DNA *Desoxyribonucleic Acid*
dNTP *Deoxynucleotide triphosphate*
DPBS *Dulbecco's Phosphate Buffered Saline*
dPCR *digital PCR*
DSB *(DNA) Double-Strand Break*
EB *Embryoid Body*
ECG *Electrocardiogram*
EDTA *Ethylenediaminetetraacetic Acid*
EF *Ejection Fraction*
EtOH *Ethanol*
FBS *Fetal Bovine Serum*
g *Gram*
gDNA *Genomic DNA*
gnomAD *genome Aggregation Database*
gRNA *(single-)guide RNA*
GWAS *Genome-Wide Association Studies*
HCM *Hypertrophic Cardiomyopathy*
HCN4 *Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 4*
HDR *Homology-Directed Repair*
hESCs *Human Embryonic Stem Cells*
hpf *Hours Post Fertilization*
HPRT *Hypoxantin-guanin- Hypoxantin-guanin-phosphoribosyltransferase*
Indels *Insertions and Deletions*
iPSC *Induced Pluripotent Stem Cell*
iPSC-CMs *iPSC-derived Cardiomyocytes*
KCl *Potassium Chloride*
LA *Left atrium*
MCA *Multiplex Cell Authentication*
$MgCl_2$ *Magnesium Chloride*
$MgSO_4$ *Magnesium Sulfate*
N/A *non-assignable*
Na Citrate *Sodium Citrate*
NaCl *Sodium Chloride*
NGS *Next Generation Sequencing*
NHEJ *Non-Homologous End-Joining*
P/S *Penicillin-Streptomycin*
PBMCs *Peripheral Blood Mononuclear Cells*
PCR *Polymerase Chain Reaction*
PEI *Polyethylenimine*
PFA *Paraformaldehyde*
Poly-HEMA *Poly(2-Hydroxyethyl Methacrylate)*
qPCR *(Real-time) quantitative PCR*
qRT-PCR *Quantitative Reverse Transcription PCR*
RNA *Ribonucleic Acid*
RNP *Ribonucleoprotein*
rpm *rounds per minute*
RT *Room Remperature*
SAN *Sinoatrial Node*
SHOX2 *Short Stature Homeobox 2 transcription factor*
SND *Sinus Node Dysfunction*
SNP *Single Nucleotide Polymorphism*
ssODNs *Single-Stranded Oligodeoxynucleotides*
SV *Sinus Venosus*
TAE *TRIS-Acetate-EDTA*
TALEN *Transcription Activator-Like Effector Nuclease*
TBX3 *T-Box Transcription Factor 3*
TBX5 *T-Box Transcription Factor 5*
Tris *TRIS-(hydroxylmethyl)-Aminomethan*
ZFN *Zinc Finger Nuclease*

Introduction

1.1. The heart – setting the pace of life

The heart is a story of evolutionary success. From the most primitive tubular structures to the sophisticated four-chambered mammalian and avian organs, the rhythmic contractions of the heart ensure the transport of nutrients, metabolites, and oxygen throughout the body.[2] The complete separation of oxygenated and deoxygenated blood circulations in birds and mammals enable high metabolic turnovers [3] and might well have laid the foundation for the development of higher cognitive processes that are made possible by a large, energy-demanding brain. Interestingly, the cephalopods (e.g. octopuses and squids), being the only invertebrates known to have evolved some form of higher intelligence, rely on three hearts and high cardiac outputs to meet their metabolic rates.[4]

In literature and culture, the heart is often depicted as the seat of emotions, passion, and love. While this has been debunked as a myth by our modern understanding of neurobiology, love might indeed be the last chapter of cardiac evolution: An anatomical feature specific for mammals is the development of a myelinated vagus nerve that connects the heart to medullary cranial nerves. The heart rate can be rapidly slowed and sped up by changes in the vagal tone. In the brain, input from the frontal cortex allows the vagal tone to be influenced by perceptions signaling safety or by recognition of situations that require calm and prosocial behavior. The ability to slow down the heart rate enables human behavior critical for courtship and long-lasting bonding, giving the original idea of a love-heart relationship a surprisingly profound biological base.[5,6]

Anatomically, the inhibitory vagus nerves innervate the cardiac atria. A substantial part of them ends in a specialized tissue in the right atrium, the so-called ‘sinoatrial node’ (SAN). The SAN is the re-entry point for contractions in the healthy heart and therefore sets the pace. Its role is established during early embryonic development, when the heart forms as the first organ and stays active for a lifetime until its ceased activity is associated with the final stage of life – death.

1.2. The mammalian heart

1.2.1. The cardiac conduction system (CCS)

The sequential contraction of different parts in the heart is essential for ensuring an adequate blood circulation. It is orchestrated by the depolarization of specialized cardiomyocytes in the atria and ventricles. These cells are organized in electrically coupled clusters, which together form the CCS. A distinct feature of CCS components are their different conduction velocities. In a healthy heart, the SAN is the dominant pacemaker controlling the rate of contraction. It is located at the junction of the superior caval vein and the right atrium. Automatically generated electrical impulses are rapidly propagated through the atria, leading to their synchronized contraction. The signal arrives at the atrioventricular node (AVN), which is positioned dorsally at the base of the interatrial septum, where it is delayed due to the slow conduction velocity within the tissue. This ensures the complete blood flow into the ventricles. The impulse then rapidly propagates through the atrioventricular bundle (or bundle of His, AVB) into the right and left bundle branches (BBs) in the ventricular septum. The bundle of His and the BBs are the only electrical connection between atria and ventricles, which are insulated from each other by the annulus fibrosus. Once the propagation of action potentials reaches the Purkinje fiber network, it is rapidly distributed throughout the ventricles, causing them to contract and pump the blood into the vascular system.[7]

Different features of the CCS components, like the rate of spontaneous differentiation, the degree of intercellular coupling and action potential morphology suggest multiple CCS lineages.[8] During heart development, atrial and ventricular chambers are formed from fast-conducting myocardium, which differentiates within the primitive heart tube. Most of the CCS develops from areas in between these regions that maintain embryonic, slow-conducting properties (**Figure 1A**). On the molecular level, the distinct electrophysiological differences are mediated by transcriptional regulators specific for their respective CCS component (**Figure 1B**).[7]

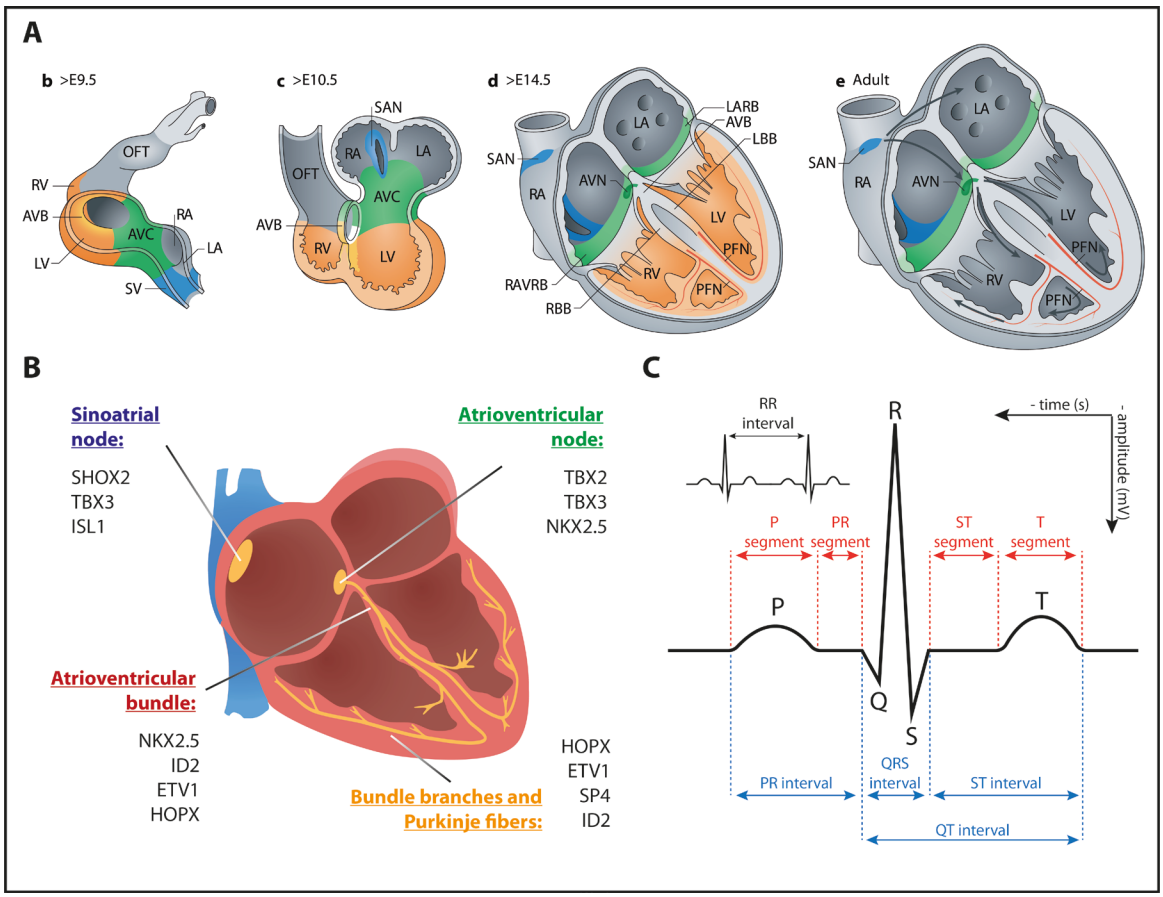


Figure 1 Mammalian heart development and electrocardiogram. (A) "From E9.5 onwards, atrioventricular canal (AVC) precursors (green) located in the IFT of the early heart tube begin to develop into the atrioventricular node (AVN, shown in parts d and e) and left and right atrioventricular ring bundles (LAVRB and RAVRB, shown in parts d and e). The sinus venosus (SV, blue) is added to the heart tube from the caudal second heart field pool of progenitors. **c|** From E10.5 onwards, the sinoatrial node (SAN) develops in the right sinus horn. From the outer curvatures of the heart tube, the myocardium expands, yielding the four heart chambers: left atrium (LA), right atrium (RA), left ventricle (LV), and right ventricle (RV). The septal part of the interventricular ring (yellow) begins to form the atrioventricular bundle (AVB). From early fetal stages onwards, compact working myocardium (grey) forms at the epicardial side of the ventricles. **d|** From E14.5 onwards, the septal trabeculations form the left and right bundle branches (LBB and RBB, yellow). The Purkinje fiber network (PFN, red) is derived from early ventricular chamber myocardium. **e|** The cardiac conduction system in the adult heart. Action potentials are initiated in the SAN and propagate (grey arrows) to the atria, AVN, AVB, BBs, PFN, and the ventricular cardiomyocytes. The LAVRB and RAVRB prevent direct action potential propagation from the atria to the ventricles." Figure and legend taken from van Eif et al., Cardiac Development 2018. **(B)** Schematic drawing showing the different CCS components and their respective key transcription factors that play a role in the development, specification, and function of the tissue. A combination of general and tissue-specific transcription factors mediates the distinct electrophysiological properties. Picture based on licensed stock photo AdobeStock_163503654 **(C)** Example of a normal electrocardiogram. Abnormalities in electrocardiogram traces can be indicative of cardiac conduction system (CCS) disorders. For example, an increased PP interval but normal PR interval indicates sinus bradycardia (sinoatrial node dysfunction). Progressive lengthening of the PR interval followed by a QRS indicates a first-degree atrioventricular block.

The sequential impulse propagation in the heart can be visualized in an electrocardiogram (ECG) (**Figure 1C**). It plots electrical activity in the heart in a graph of voltage versus time. The signaling propagation from the SAN through the atria causes cardiomyocytes to depolarize resulting in the P wave. In contrast, the ventricular depolarization is reflected by the QRS complex, and the T wave represents ventricular repolarization. The PR interval, defined as the time from the onset of the P wave to the beginning of the QRS complex indicates the time that the impulse requires to travel through the atria, the AVN and AVBs. The ventricular conduction system is assessed by the duration of the QRS complex and the QT interval measures the duration of ventricular depolarization and repolarization.

Abnormalities in the ECG, such as prolonged intervals and segment durations can be indicative of CCS disorders and are associated with cardiac arrhythmias.[7,9] Due to the non-invasiveness of electrodes that are planted on a patient's chest and limbs, electrocardiography is a routinely used examination method for assessing the cardiac function in the clinical context.

1.2.2. The sinoatrial node

The SAN is a heterogenous tissue that comprises different cell populations, including fibroblasts endothelial cells, connective tissue, and specialized cardiomyocytes. These pacemaker cells are capable of spontaneous depolarization without external stimulation and thereby initiate the cardiac cycle. Compared to cardiomyocytes of the working myocard, they contain poorly developed sarcomeres and different electrical activity. In the center of the SAN, cardiac cells are connected by slow conducting connexins, such as Cx30.2 (*Gjd3*) and Cx45 (*Gjc1*), which prevent them from being externally stimulated.[10] The automatic electrical activity is mediated by an interplaying combination of membrane ion channels, ('membrane clock') and intracellular Ca^{2+} handling mechanisms ('Ca^{2+} clock'), causing a non-stable membrane potential.[11] The SAN is innervated by the autonomic nervous system enabling a regulation of the heart beat to change cardiac output and metabolic demand.[7]

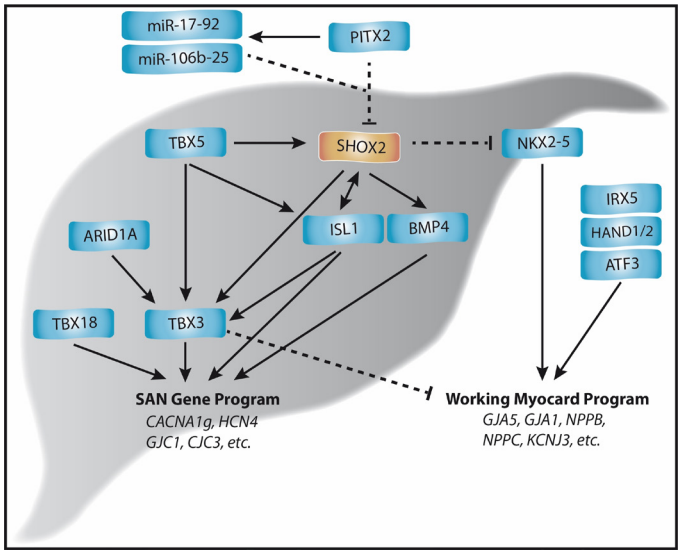


Figure 2 The genetic network of the SAN. *SHOX2* maintains the phenotypic border between working myocard and SAN tissue by antagonizing *NKX2.5* to prevent the expression of fast-conducting connexins and atrial ion channels. It synergizes with *TBX5* and *TBX18* to induce the SAN genetic program by activating the expression of *ISL1*, *BMP4* and *TBX3*. This ultimately leads in the activation of genes, such as *CACNA1g* and *HCN4*, which mediate the unique pacemaker properties. Outside of the SAN, *SHOX2* is repressed by *PITX2*. Figure and legend adapted from van Eif et al., Cardiac development 2018.

During early mouse embryogenesis (~E8), the heart tube is formed by the fusion of two heart-forming regions in the lateral plate mesoderm. At this point, all myocardial cells are slow-conducting and automatic, however dominant automaticity is already located caudally at the inflow tract. The addition of cardiomyocytes derived from the second heart field elongates the primitive heart tube at both poles (**Figure 1A b**). This gives rise to the sinus venosus (SV), which becomes the dominant re-entry point for contractions. From E10.5 onwards, the SAN develops from the right sinus horn (**Figure 1A c-e**), to which the pacemaker activity finally gets restricted.[7,9,12]

The genetic network which regulates SAN development and mediates its unique properties is highly conserved.[2] The Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 4

(*Hcn4*), is crucial for pacemaker activity. Its expression is initially global, but ultimately only maintained in pacemaker cells of the SAN and AVN.[13] In contrast, the expression of *Nkx2.5* remains high in cells of the working myocard but is downregulated in pacemaker cells which express *Tbx18* instead.[14,15] In the atria, *Nkx2.5* initiates the expression of chamber-specific genes, such as fast-conducting connexins Cx40 (*Gja5*) and Cx43 (*Gja1*) or ion channels (e.g. *Scn5a*).[16] The asymmetrical formation of the SAN is ensured by *Pitx2*, which represses left-sided pacemaker development. This is achieved by inhibition of the Short Stature homeobox 2 transcription factor (*Shox2*), T-Box Transcription Factor 3 (*Tbx3*) and presumably other mechanisms.[7,17,18] T-Box Transcription Factor 5 (*Tbx5*), which is expressed in the SV and precursor tissues, activates the pacemaker genes *Bmp4*, *Shox2* and *Tbx3*.[19,20] *Shox2* maintains the phenotypic border between SAN and working myocard by a common antagonistic mechanisms with *Nkx2.5*.[21] Additionally, it activates the SAN genetic program including *Hcn4*, *Tbx3* and its direct targets *Isl1* and *Bmp4*.[19,22] *Tbx3* supports the activation of pacemaker genes and suppresses the expression of atrial genes such as *Gja1*, *Gja5* and *Scn5a* (**Figure 2**).[23] Disturbances in this fine-tuned network of transcriptional regulators can lead to the expression of "pacemaker genes" outside the SAN or genes of the working myocard on the inside. This can affect the electrophysiological properties of the different tissues resulting in impaired electrical transmission or altered sequential contractions, which clinically manifest as arrhythmias.

1.3. Disturbances in the CCS

1.3.1. Atrial fibrillation (AF)

AF is the most common cardiac arrhythmia with a worldwide prevalence of 37 million cases (0.51% of the world population). With an ageing population, the substantial socioeconomic burden and incidence are expected to further increase.[24] The hallmark of AF is the rapid and disorganized electrical activation of the atria, leading to uncoordinated contractions. This is caused by ectopic re-entry points that reside outside of the SAN. Structural changes such as fibrosis and tissue scarring upon heart strokes or congenital abnormalities in Ca^{2+} handling (early afterdepolarizations and delayed afterdepolarizations) are the underlying mechanisms for the initial formation of this ectopic activity. This results in a progressive electrical and structural remodeling of the atrial tissue which sustains the ectopic re-entry and ultimately stabilizes it (**Figure 3**). In the clinical context, the atrial flattering is initially self-terminating (paroxysmal) but progresses to a persistent and ultimately to

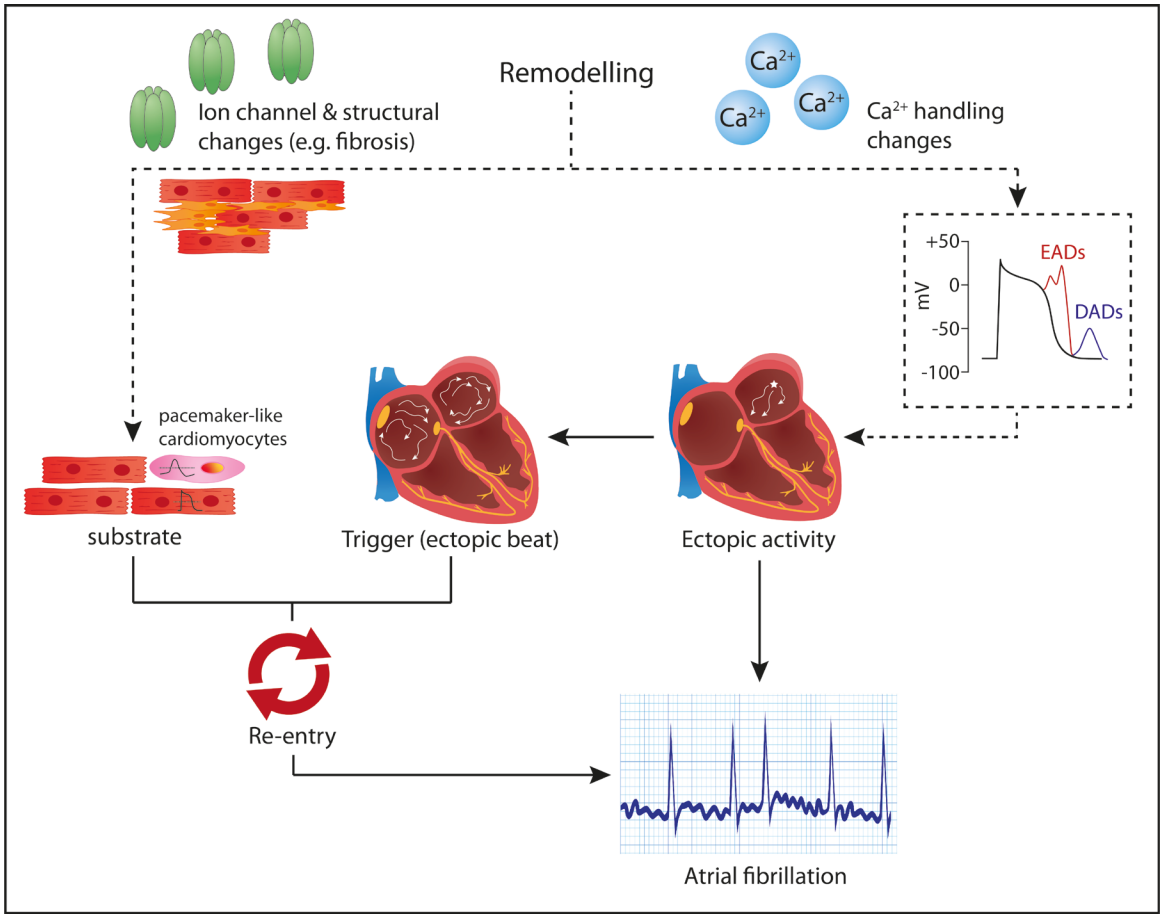


Figure 3 Key mechanisms of atrial fibrillation. Rapid/Sustained ectopic or re-entrant activity can maintain AF. An ectopic beat acting on vulnerable substrate functions as a trigger and leads to the development of a re-entry point. Atrial ectopy is rare in healthy hearts with atrial electrical properties being unsuitable for maintenances of AF. However, atrial remodeling by altering ion-channel function or inducing tissue fibrosis creates a substrate for re-entrant AF. Remodeling can also produce changes in cell Ca2+-handling, thus making ectopic activity more likely. Legend and figure concept are based on Dobrev D, Mattel S, Lancet 2010
Abbreviations: EADs=early afterdepolarizations. DADs=delayed afterdepolarizations.

a permanent type with no spontaneous termination if not treated correctly.[25] The loss of a synchronized electrical conduction leads to blood stasis in the atria and predisposes patients to the development of atrial thrombi and stroke.[26] AF is epidemiologically associated with other conditions such as coronary artery disease, cardiomyopathies, hyperthyroidism, arterial hypertension, obesity and diabetes mellitus.[24] The lack of a risk factor in some cases and the often observed familial aggregation, however, suggests underlying genetic predispositions to develop the disease. Large genome-wide association studies (GWAS) and meta-analyses have identified more than hundred loci that are significantly associated with AF.[27] Additionally, rare variants identified in AF patients were analyzed in both cell and animal models.[26] The majority of risk genes encode ion channels, but structural proteins, signaling molecules and transcription factors have been linked to this disease as well.[26] Among the transcription factors are the *SHOX2*-inducing T-Box Transcription Factor 5, the *SHOX2* antagonist *NKX2*.5, the *SHOX2* repressor *PITX2* and *SHOX2* itself.[26,28-30]

1.3.2. Sinus node dysfunction (SND)

SND, also referred to as sick sinus syndrome, includes a spectrum of symptoms and heart rhythm dysfunctions caused by an abnormal sinus impulse formation and/or propagation. This heterogenous clinical entity presents in different electrocardiographic manifestations, such as sinus arrest (no electric impulse generation for >2 s), sinoatrial exit block (sinus impulse is delayed/blocked in atria) and bradycardia (<60 heart beats per minute).[31] Electrophysiological characteristics of SND are an increased sinus node recovery time (time until first automated impulse after stimulation), a delayed conduction along the crista terminalis and areas of low voltage within the right atrium.[11,32] The disease is often asymptomatic in early phases; in progressed stages, symptoms can include fatigue and dizziness due to reduced cerebral perfusion, effort dyspnea, syncope and presyncope. Patients are at risk of developing major cardiovascular events, including chronotropic incompetence (inadequate leading pacemaker response to exercise/stress) and thromboembolism. Depending on how SND manifests clinically, it is treated by the administration of antiarrhythmic drugs (AADs) and pacemaker implantation.[31] Currently 30-50% of pacemaker implantations in the US are primarily due to SND and the incidence of the disease is projected to double until 2060.[33]

The etiology of the disease can be extrinsic, intrinsic or a mixture of the two. In a retrospective study of patients presenting with bradycardia, half of the cases were attributable to a treatable extrinsic cause such as electrolyte imbalance (e.g. Hyperkalemia, hypocalcemia), adverse drug reaction or acute myocardial infarction while the other half were assumed to be intrinsic/idiopathic.[11] The pathophysiology of idiopathic SND or 'primary SND' is still not fully understood. Idiopathic SND generally occurs in the elderly population, but familial forms have also been reported at earlier stages.[11]

1.3.3. SND and arrhythmias

Atrial arrhythmias and SND frequently coexist in 40-70% of patients suffering from AF at the time of diagnosis of SND. The combination of SND with AF is the basis of the 'tachycardia-bradycardia syndrome', where atrial flutter or fibrillation episodes alternate with sinus pauses at their termination.[34] Both disorders interact to initiate and perpetuate each other, yet, their complex relationship remains elusive.[35] The connecting paradigm between the two clinical entities is likely the anatomical and electrophysiological remodeling of the atria that is visible on the structural, cellular and genomic levels.[35] SND and AF are associated with senescence. Age-induced degenerative fibrosis leads to slower intrinsic heart rates and increased sinus node conduction time, both of which favor the manifestation of SND and AF.[31] The declined sinus node impulse formation can allow ectopic atrial foci to mature and fire. The stabilization of such re-entry points outside the SAN ultimately results in AF. In turn, AF can shut down the sinus node activity by overdrive

suppression.[34] Persistent atrial arrhythmias result in cardiomyocyte apoptosis, electric decoupling, progressive fibrosis and dilatation of the atrial and SAN tissue.[36,37] It manifests on the molecular level in the downregulation of ion channels and connexins.[35]

Under these aspects it might not be surprising that most SND-related genes are also associated with AF. For example, several mutations and deletions within the *HCN4* gene were found in patients suffering from paroxysmal AF or bradycardia.[38-40] *HCN* genes encode hyperpolarization-activated cyclic nucleotide-gated channels, which are highly expressed in the CCS, where they are responsible for the automated generation of action potential. *SCN5A* is expressed in the atria but absent in the SAN.[11] Mutations in this gene result in highly variable clinical presentations, including AF, Brugada syndrome, long-QT syndrome but also SND, which especially manifests as sinus exit block.[41,42]

Medical treatment of AF in the presence of SND and reverse is challenging. Application of rate control agents and AADs can worsen SND-related bradycardia leading to life-threatening symptoms such as long sinus pauses. On the other hand, treating one condition can ameliorate the symptoms of the other, for example by reverting detrimental tissue remodeling.[43]

1.4. *SHOX2* gene, expression, and function

1.4.1. *SHOX2* gene

During embryogenesis, the homeobox gene family controls many developmental processes that are highly conserved among species.[44] Mutations in several members have been described in human diseases including Léri-Weill dyschondrosteosis (*SHOX*) and Rieger syndrome (*PITX2*).[45,46] The short-stature homeobox containing gene family comprises *SHOX* and its related homologue *SHOX2.* The encoded amino acid sequence and common homology domains share 99% and 83% similarity, respectively.[47] *SHOX2* is localized on 3q25.32 (chr3:157,813,800-157,824,292, minus strand (GRCh37/hg19)) and consists of 6 exons. Nucleic acid sequencing revealed two isoforms, *SHOX2a* and *SHOX2b*. Both isoforms contain the homeodomain, the SH3 domain and an OAR domain. It is expressed in developing limbs, the nasal cavity, branchial arches, embryonic reproductive nodules, the central nervous system, and the heart. In the latter, *Shox2* expression is restricted to the CCS, especially the SAN and AVN and the adjacent venus valves.[48] Its central role in heart development (as described above) was confirmed in various animal models. Reduced pacemaker proliferation induces lethal SV and SAN hypoplasia and bradycardia in *Shox2*$^{-/-}$ mice between E11.5 and E17.5. In the SAN region, a switch to the genetic program of the working myocard can be observed with an ectopic expression of *Nkx2.5*, *Nppa* and *Cx40* and a downregulation of *Tbx3*, *Hcn4* and *Isl1*.[48,49] In zebrafish, the morpholino-mediated knockdown of endogenous *zShox2* leads to severe bradycardia and pericardial edema, which can be rescued by the overexpression of its direct target *Isl1*.[22]

1.4.2. *SHOX2* and AF

The first heterozygous missense mutations were identified in a cohort of 378 patients with early-onset AF.[28] From these, a coding variant (*SHOX2* c.849C>A, SHOX2 p.H283Q) and a 3' untranslated region (3'UTR) variant (*SHOX2* c.*28T>C) were shown to affect the regulation and function of *SHOX2*. Mechanistically, the coding variant impeded the transactivational activity of the SHOX2 protein, as seen in reporter assays where it failed to activate both *BMP4* and *ISL1* target genes. Furthermore, if endogenous *Shox2* was substituted for mutated *Shox2* in embryonic zebrafish hearts via morpholino injection, it could not rescue the pericardial edema and bradycardia phenotype that is observed upon loss of *Shox2* via MO-mediated knockdown (**Figure 4A**). The 3'UTR variant was present in the general population without known AF, but significantly enriched in AF patients compared to control cohorts. The base substitution led to a novel *miR-92b-5p* binding site that was shown to be functional in *in vitro* reporter assays. This miRNA was also significantly lower expressed in carrier AF patients when compared to affected non-carriers; with a more pronounced effect, when a subset of carrier patients with a PR interval >200 ms was compared to non-carriers with a PR interval of <200 ms (**Figure 4B**).[28] The association of *SHOX2* with AF was further confirmed by the identification of a heterozygous nonsense mutation (p.R194X) in a cohort of unrelated patients with familial AF that was absent in the control group. This variant co-segregated with the disease phenotype in the patient's family as well as in all affected members with complete penetrance. The nonsense mutation resulted in a premature stop codon, truncating the homeodomain of *SHOX2* and therefore leading to a complete loss of function as a transcriptional activator (**Figure 4C**).[50]

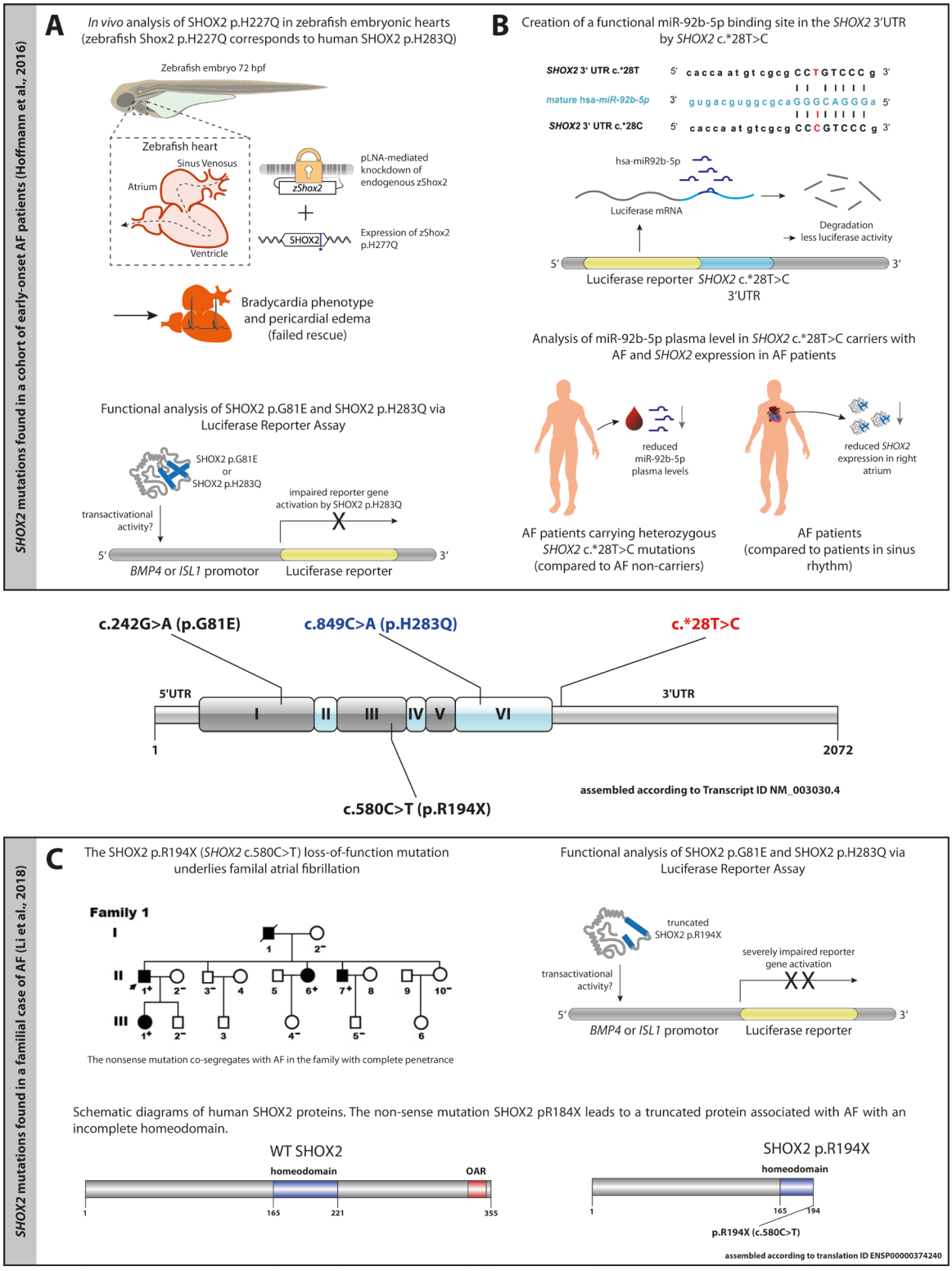


Figure 4 *SHOX2* and atrial fibrillation. (A) Three mutations (p.G81E, p.H283Qand c.*28T>C) were found in a cohort of patients with an early onset of atrial fibrillation (<65 years). zShox2 p.H277Q was overexpressed in embryonic zebrafish hearts (72 hpf) after pLNA-mediated knockdown of endogenouszShox2. It failed to rescue the bradycardia and pericardial edema phenotype that is observed after the loss of Shox2. An impaired transactivational activity was shown for SHOX2 p.H283Q but not SHOX2 p.H81E in a Dual-Luciferase Reporter Assay. (B) *SHOX2* c.*28T>C generates a novel miR-92b-5p binding site that was shown to be functional in a Luciferase Reporter Assay. AF patients carrying the heterozygous *SHOX2* c.*28T>C mutation had significantly reduced miR-92b-5p plasma levels when compared to AF non-carriers. AF patients in general showed a reduced *SHOX2* expression in right atrium tissue. (C) The SHOX2 p.R194X (*SHOX2* c.580C>T) mutation was associated with a familial form of AF. The nonsense mutation co-segregated with AF in the family with complete penetrance. SHOX2 p.R194X is a truncated protein lacking the homeodomain that is crucial for its function as a transcription factor. A severely reduced transactivational activity was shown for SHOX2p.R194X in a Dual-Luciferase Reporter Assay.

1.5. The induced pluripotent stem cell (iPSC) model

1.5.1. iPSCs in cardiac disease models

Research on the genetic basis of human cardiovascular disease has mainly relied on animal models in the past due to the limited access of human primary cardiac tissue. However, species-specific electrophysiological and transcriptional differences in cardiomyocytes can confound the translation of these findings to clinical relevance.[51,52] iPSCs help to overcome these limitations, as they provide an unlimited source of cardiomyocytes that allow for close genotype-phenotype associations if they originate from patients. This forms the basis for the investigation of underlying pathogenic mechanisms and the discovery of promising therapeutic targets.[53] In addition, they offer a personalized drug-screening platform for precision medicine and individualized therapy.[54]

In the last decade after the first description of the methods to reprogram human somatic cells into a pluripotent state[55], iPSC models for several classes of cardiac disorders have been developed. Primary arrhythmic diseases caused by mutations in ion channels and characterized by abnormal ECG, have been extensively studied in iPSCs. For example, iPSC-derived cardiomyocytes (iPSC-CMs) with mutations in *KCNQ1*, *KCNH2* or *SCNA5*, modelling 3 subtypes of Long-QT syndrome, recapitulated the electrophysiological phenotypes observed in patients and responded to anti-arrhythmic or proarrhythmic drugs in a comparable manner. Molecular analysis of these cells provided insights into pathological mechanisms, such as aberrant protein trafficking, abnormal ion currents or altered channel activation.[56-58]

Cardiomyopathies, such as dilated (DCM), hypertrophic (HCM) and arrhythmogenic (ACM) cardiomyopathies were mimicked in iPSCs harboring mutations in risk genes like *TNNT2*, *LMNA* and *TTN* (DCM), *MYH7* (HCM) or *PKP2* (ACM). These models often showed structural disorganizations including sarcomeric disarray and abnormal cell size as well as reduced contractility and aberrant Ca^{2+} handling.[59-61] The cardiovascular effects of metabolic and other disorders, for example Barth syndrome (BTHS) or Duchenne muscle dystrophy, were also analyzed in iPSC models. BTHS iPSC-CMs with mutations in the *TAZ* gene exhibited mitochondrial dysfunction and sarcomeric disorganization.[62] Patient-specific iPSC-CMs harboring deletions in the *DMD* gene, show abnormal Ca^{2+} handling and arrhythmogenic susceptibility, which can be ameliorated with somatic gene editing.[63]

AF has been mimicked in human embryonic stem cells (hESCs). Artificial tissue consisting of atrial-like CMs was generated, in which stable re-entrant rotor patterns could be induced. These rotors are typical for AF, however, treatment with flecainide (class Ic anti-arrhythmic drug) did not influence their pattern or occurence.[64] The first iPSC-based model for AF, in which the triggering cellular mechanisms of a familial form of AF were investigated, was recently described. Longer action potential durations, aberrant Ca^{2+} streams and enhanced pacemaker currents I_f were described

in iPSC-CMs with the common patient genetic background.[65] Several limitations of this study remain with a missing defined genetic causality and the lack of a state-of-the-art differentiation to atrial-like iPSC-CMs. Yet, these results are encouraging that iPSC-CMs can be used to identify AF mechanisms, to screen for therapeutic targets or to predict responses of AF patients to certain medication.

1.5.2. Genome-editing in iPSCs

The applicability and versatility of iPSCs has been further potentiated by recent advances in gene transduction and editing technologies.[66] Genome-targeting is feasible due to the development of site-specific nucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 as these enzymes helped to overcome the low frequency of random homology mediated gene targeting in stem cells.[67-70]

Following a DNA double-strand break (DSB), endogenous repair mechanisms through non-homologous end joining (NHEJ) or homology-directed repair (HDR) are activated by the cell. NHEJ does not require a homologous template for repair but often leads to the insertion or deletion of nucleotides (Indels), potentially resulting in the disruption of the gene function due to frameshift or loss-of-function mutations. This mechanism can be utilized if the loss of one or two gene copies is desired, as in the generation of knockout cell lines or animals. In contrast, during HDR a sequence of (partially) homologous DNA, naturally the undamaged sister chromosome, serves as a repair template. This normally leads to a complete restoration of the DNA sequence.[71] In the presence of an exogenously introduced template, the alteration or insertion of specific DNA sequences, ranging from single nucleotide exchanges to whole transgenes, can be achieved.[72]

While the sequence specificity of ZFNs is mediated by a sequential fusion of several zinc finger domains binding to defined base pair triplets, TALENs are recruited to their target loci by their TALE DNA-binding domain, which consists of modularly assembled amino acids recognizing dinucleotide sequences.[68,73-77] Despite concerns regarding undesired DNA cleavage at other genomic sites than the target locus (off-targets), ZFNs and TALENs have been successfully used in genome-editing of iPSCs.[78-81] However, additional technical aspects, such as restrictions in targetable DNA sequences, laborious cloning efforts and the complex protein engineering further limit a broader application of ZFNs and TALENs.[82-85]

The field of genome editing has been revolutionized by the CRISPR/Cas9 system. Originally a component of the adaptive immune response in *Streptococcus pyogenes*, the CRISPR-Cas9 nuclease has been repurposed for simple and versatile gene engineering.[86,87] The targeting of specific genomic loci is mediated by single-guide RNA (gRNA), which consists of an RNA scaffold enabling the formation of Cas9 ribonucleoprotein (RNP)/gRNA complexes and a ~20 nt

"protospacer" complementary to a genomic sequence that is adjacent to a three base pair NGG motif (PAM motif).[88-90] A significant technical advantage of CRISPR/Cas9 over TALENS and ZFNs is that this protospacer sequence can be easily exchanged to target any part of the genome. Yet, off-targeting also remains a safety issue with this method.[91]

Utilizing DSBs to introduce sequence alteration by error-prone NHEJ-mediated DNA repair has led to numerous knockout models in iPSCs.[90] However, even though the disruption of genes is a practical tool for basic research and might even be a feasible therapeutic option in certain cases, the true potential of genome engineering lies within precise modifications.[92,93] A plethora of possibilities results from wanted sequence alterations including the generation of gene knock-ins, reporter cell lines and the insertion or correction of patient-specific mutations: Knock-in and reporter cell lines can facilitate the monitoring of gene expression *in vitro* and tracing of transplanted cells *in vivo* or enable the selection and purification of certain subtypes in mixed populations of differentiated cells.[94-97] Insertion of disease-linked variants into a wildtype background helps to interrogate the influence of these mutations on the onset and progression of a disease by direct comparison of mutated cells to their isogenic counterpart. Yet, this approach is limited to monogenic diseases or variances with a high impact on the phenotype. In sporadic or idiopathic diseases, a combination of multiple risk alleles with low effect size is thought to be the genetic basis. Individual risk variants might not be sufficient to cause a disease-associated phenotype. Fortunately, optimizations in iPSC technology and commercially available tools have streamlined iPSC derivation, so that the generation of patient-specific iPSCs harboring all risk alleles has become broadly accessible.[67] The correction of putative disease-contributing variants in these cells could also unravel subtle phenotypic changes when comparing patient cells to their isogenic controls, even if the disease phenotype overall persists. Gene correction also holds potential for human gene therapy approaches, in which somatic cells are reprogrammed to iPSCs and the detrimental mutation is corrected before the cells are differentiated to the desired cell type and re-transplanted into the patient for a beneficial effect.[98]

However, a key limitation in precise genome editing remains indicating that mammalian cells preferentially choose the error-prone NHEJ pathway rather than a homology-directed approach using exogenous templates to repair DSBs.[99] Under normal circumstances, frequencies of ~1% for single-base substitutions or deletions are often reported.[90,100-102] In recent years, substantial effort has been undertaken to develop strategies for improved HDR efficiency:

Culturing iPSCs at 32°C for 24-48h after Cas9 transfection, was reported to increase the efficiency of single nucleotide exchanges.[99] Selection genes are often used to enrich for successfully edited cells. However, these genes sometimes need to be incorporated into the genome and either remain or must be removed in a second step, eventually leaving a genomic 'scar' in the process. Other approaches were based on enhancing either the NHEJ or the HDR pathway via the application of small molecules.[103] Similarly, it was shown that the inhibition of NHEJ key molecules via gene

silencing improved HDR efficiency and abolished NHEJ activity.[104] Cells preferentially choose NHEJ or HDR dependent on the cell cycle with HDR being restricted to the late S and G2 phases while NHEJ pathway is dominating during G1, S and G2 phases.[105] On this base, strategies were developed to increase HDR by fusing Cas9 to Geminin which resulted in a cell-cycle-tailored expression or by arresting cells in S phase before delivery of Cas9.[106,107] The HDR templates were optimized in regard to design (single-stranded vs double-stranded), homology arm length and symmetry around Cas9 cut site for improved incorporation of sequence changes.[108-110] Also the Cas9 enzyme itself was extensively modified to improve its specificity and efficiency: By catalytical inactivation of one of the nuclease domains, a Cas9 nickase was generated which led to single-strand DNA breaks that were preferentially repaired via HDR if provided with the right template.[111] Recent approaches focus on avoiding DSBs by coupling catalytically inactive, 'deactivated' Cas9 (dCas9) with enzymes: In base-editing, dCas9 linked with deaminases is used to achieve transition mutations (purine to pyrimidine and reverse) in a narrow window of DNA base pairs.[112] Prime-editing fuses dCas9 to a reverse transcriptase that is capable of achieving transition and transversion mutations as well as insertions and deletions.[113]

An alternative approach was used for detection of rare single-base editing events in iPSCs: Here, iPSCs were subcloned in small cell pools (termed 'sib-selection') to randomly enrich cells of interest. Subsequently digital PCR was applied to detect the correctly edited cell subpopulations in these pools. With subsequent subcloning rounds, always using the sib-selection with the highest percentage of cells of interest, the target population was enriched until single cell-cloning and screening of correctly edited clones could be performed with relative ease.[102]

1.6. Aim of the thesis

SHOX2 is a key regulator in cardiac development and specifically the SAN as shown in various animal models. Mutations in this gene have been identified in patients with early-onset and familial AF. However, the molecular mechanisms that involve *SHOX2* and lead to the onset and the progression of the disease in humans have not been fully elucidated yet. In addition, *SHOX2* has not been linked to SND despite the high gene expression in this tissue and the close relationship between this disease and AF.

The aim of the first part of this PhD thesis was to find a causal relationship between novel *SHOX2* variants and the development of SND and AF. For this, cohorts of AF and SND patients should be screened and newly identified mutations analyzed *in vivo* and *in vitro*.

In the second part, an *in vitro* AF model should be established based on patient-specific iPSCs. Cells harboring heterozygous *SHOX2* mutations have been previously described. Two patients were planned to be recruited and clinically examined. A reprogramming of somatic cells and characterization of several iPSC lines per patient was planned. To generate a suitable control for

these lines, the heterozygous *SHOX2* mutations needed to be precisely corrected and novel methods developed along the way, to allow for an easier and scarless correction of heterozygous mutations with a gene-editing tool of choice.

2. Material and Methods

1.1. Materials

1.1.1. Cell culture reagents

Basal media	
Dulbecco's Phosphate Buffered Saline (DPBS) w/o Ca^{2+}/Mg^{2+}	Thermo Fisher Scientific
DPBS /w Ca^{2+}/Mg^{2+}	Thermo Fisher Scientific
Dulbecco's Modified Eagle's Medium (DMEM)/F12 (1:1)	Thermo Fisher Scientific
DMEM (1x), high glucose	Thermo Fisher Scientific
Distilled water	Thermo Fisher Scientific

Stem cell media	
mTeSR™1	STEMCELL Technologies
Essential 8™	Thermo Fisher Scientific
StemFit® Basic02	Nippon Genetics

Media supplements	
Non-essential Amino Acid Supplement ('NEAA') (100x)	Thermo Fisher Scientific
GlutaMAX™ Supplement (100x)	Thermo Fisher Scientific
Penicillin-Streptomycin (P/S) 10.000 U (100x)	Thermo Fisher Scientific
Normocin™ (used 1:1000)	InvivoGen
Plasmocin™ (used 1:1000)	InvivoGen
Fetal bovine serum (FBS)	Thermo Fisher Scientific
Fetal bovine serum (LOT# 044M3395)	Sigma-Aldrich
2-Mercaptoethanol	Sigma-Aldrich
Recombinant Human FGF-basic (154 a.a.)	PeproTech
L-Ascorbic acid	Sigma-Aldrich
Y-27632 hydrochloride (ROCK inhibitor)	Cayman Chemicals

Cell dissociation reagents	
0.05% Trypsin-EDTA (1x), phenol red	Thermo Fisher Scientific
TrypLE™ Express	Thermo Fisher Scientific
Accumax™	Merck Millipore
Dispase (1 U/ml) in DMEM/F12	STEMCELL Technologies

Freezing media	
mFreSR™	STEMCELL Technologies
Bambanker™	Nippon Genetics

Coating reagents	
Geltrex™ LDEV-Free, hESC-Qualified, Reduced Growth Factor Basement Membrane Matrix	Thermo Fisher Scientific
Fibronectin bovine plasma	Sigma-Aldrich
Gelatin from porcine skin	Sigma-Aldrich

Transfection reagents	
Polyethyleneimine (PEI)	Sigma-Aldrich
Lipofectamine® 2000 Transfection Reagent	Thermo Fisher Scientific
FuGENE® HD	Promega
P3 Primary Cell 4D-Nucleofector™ X Kit S	Lonza

Luciferase assay	
Dual-Luciferase® Reporter Assay System	Promega

Other reagents	
KaryoMAX™ Colcemid™ Solution in PBS	Thermo Fisher Scientific
Paraformaldehyde (PFA)	Sigma-Aldrich
Dimethyl sulfoxide (DMSO	Sigma-Aldrich
UltraPure™ Ethylenediaminetetraacetic acid (EDTA), 0.5M, pH 8.0	Thermo Fisher Scientific
Poly(2-hydroxyethyl methacrylate) (Poly-HEMA)	Sigma-Aldrich

HEK293TN medium	DMEM (high glucose) 10% FBS 1% P/S
EB-20 medium	DMEM/F12 20% FBS (LOT# 044M3395) 1% GlutaMAX™ Supplement 1% NEAA Supplement 0.1 mM 2-Mercaptoethanol 0.5% P/S
EB-2 medium	DMEM/F12 2% FBS (LOT# 044M3395) 1% GlutaMAX™ Supplement 1% NEAA Supplement 0.1 mM 2-Mercaptoethanol 0.5% P/S

1.1.2. Bacteriological Media and Supplements

Reagents	
Ampicillin	Roth
Bacto™ Agar	BD
Bacto™ Tryptone	BD
Chloramphenicol	Sigma-Aldrich
Glucose	Merck
Glycerol	Sigma-Aldrich
Kanamycin	Roth
Magnesium chloride ($MgCl_2$)	Merck
Magnesium sulfate ($MgSO_4$)	Merck
Potassium chloride (KCl)	Merck
Sodium chloride (NaCl)	Sigma-Aldrich
Yeast extract	BD

Media for bacterial cultures	
LB medium	1% Bacto™ Tryptone 0.5% Yeast Extract 1% NaCl ddH_2O ad 1L

LB agar	1.5% Bacto™ Agar LB medium ad 1L
SOB medium	2% Bacto™ Tryptone 0.5% yeast extract 10 mM NaCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ ddH_2O ad 1L
SOC medium	1L SOB medium 20% glucose
Ampicillin	50 mg/ml in 70% ethanol, diluted 1:1000 in LB medium/agar
Kanamycin	30 mg/ml in ddH_2O, diluted 1:1000 in LB medium/agar

1.1.3. Vector constructs

Vectors	
pDestTol2CG2	Provided by Prof. Steffen Just
p5Ecmlc	Provided by Prof. Steffen Just
p3EpolyA	Provided by Prof. Steffen Just
pDONR™221	Thermo Fisher Scientific
pDest27	Thermo Fisher Scientific
pGL3basic	Promega
pRL-TK	Promega

1.1.4. Reagents and buffers for molecular analyses

Reagents	
Acetic acid	Merck
LE Agarose	Biozym
Chloroform	Sigma-Aldrich
Distilled Water	Thermo Fisher Scientific
Ethanol (EtOH)	Sigma-Aldrich
EDTA	Sigma-Aldrich
GeneRuler DNA Ladder (100 bp+ /1 kb+)	Thermo Fisher Scientific
Glycine	Sigma-Aldrich
Isopropanol	Sigma-Aldrich
Midori Green DNA stain	Nippon Genetics
KCl	Merck
Sodium citrate (Na Citrate)	AppliChem
TRIS-(hydroxyl methyl)-aminomethane (Tris)	Carl Roth

50x TRIS-Acetate-EDTA (TAE) buffer	2 M TRIS 1 M acetic acid 50 mM EDTA H_2O ad 10 l

1% NaCl (w/v)	5 g NaCl ddH_2O ad 500 ml
0.55% KCl (w/v)	2.75 g KCl ddH_2O ad 500 ml

1.1.5. Kits

DNA and RNA extraction, purification & quantification	
QIAquick® PCR & Gel Cleanup Kit	Qiagen
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific
ZymoPURE™ II Plasmid Midiprep Kit	Zymo Research
Quick-DNA™ Miniprep Plus Kit	Zymo Research
Quick-DNA™ 96 Plus Kit	Zymo Research
GeneElute™ 96 Well Tissue Genomic DNA Purification Kit	Merck
Qubit™ dsDNA BR Assay Kit	Thermo Fisher Scientific
Qubit™ RNA BR Assay Kit	Thermo Fisher Scientific
directPCR® lysis reagent	PEQLAB
TRIzol™ Reagent	Thermo Fisher Scientific
Direct-zol RNA Microprep	Zymo Research

Conventional PCR	
HotStarTaq DNA Polymerase	Qiagen
Q5® High-Fidelity DNA Polymerase	New England Biolabs
Deoxynucleotide triphosphate (dNTP) Set 100 mM Solutions	Thermo Fisher Scientific

Cloning	
FastDigest™ Restriction Enzymes	Thermo Fisher Scientific
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific
T4 DNA Ligase	Thermo Fisher Scientific
Gateway® Gene Cloning	Thermo Fisher Scientific

***In vitro* RNA synthesis**	
Precision gRNA Synthesis Kit	Thermo Fisher Scientific

Droplet Digital™ PCR (ddPCR™) or digital PCR (dPCR)	
Pierceable Foil Heat Seal	Bio-Rad
ddPCR™ 96-Well Plates	Bio-Rad
Droplet Generation Oil for Probes	Bio-Rad
DG8™ Cartridges for QX100/QX200 Droplet Generator	Bio-Rad
DG8™ Gaskets for QX100/QX200 Droplet Generator	Bio-Rad
ddPCR™ Buffer control kit	Bio-Rad
ddPCR™ Supermix for probes (no dUTP)	Bio-Rad
ddPCR™ Droplet Reader Oil	Bio-Rad

cDNA synthesis and qRT-PCR	
SuperScript™ III First-Strand Synthesis System	Thermo Fisher Scientific
qPCRBIO SyGreen Mix Lo-ROX	Nippon Genetics

Cloning & Mutagenesis	
FastDigest™ Restriction enzymes	Thermo Fisher Scientific
T4 DNA Ligase	Thermo Fisher Scientific

Mix & Go *E.coli* Transformation Kit and Buffer Set	Zymo Research
QuikChange™ Lightning Site-Directed Mutagenesis Kit	Agilent

Other

TaqPath™ ProAmp™ Master Mix	Thermo Fisher Scientific
GeneArt™ Platinum™ Cas9 Nuclease	Thermo Fisher Scientific
TrueCut™ Cas9 Protein v2	Thermo Fisher Scientific
Venor®GeM Classic Mycoplasma Detection Kit	Minerva Biolabs

1.1.6. Bacteria strains and cell lines

Bacterial strains

DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdr17*
XL10-Gold® Ultracompetent Cells	*Tet^{r}Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte *[F´ proAB lacI^q ZΔM15* Tn*10 (Tet^r) Amy Cam^r]*
One Shot™ *ccd*B Survival™ 2 T1^R Competent Cells	F$^-$*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *fhu*A::*IS2*

Cell lines

HEK293TN	BioCat
SHOX2.2.6 (= *SHOX2* c.849C>A_1)	Reprogrammed somatic cells to iPSCs from Patient I in collaboration
SHOX2.2.10 (= *SHOX2* c.849C>A_2)	Reprogrammed somatic cells to iPSCs from Patient I in collaboration
SHOX2 B7 (= *SHOX2* c.28*T>C_1)	Reprogrammed somatic cells to iPSCs from Patient II in collaboration
SHOX2 Klon 2 (= *SHOX2* c.28*T>C_2)	Reprogrammed somatic cells to iPSCs from Patient II in collaboration
SHOX2 Klon Z (= *SHOX2* c.28*T>C_3)	Reprogrammed somatic cells to iPSCs from Patient II in collaboration
SHOX2.2.6-1D6-1H10 (= *SHOX2*_c849C>A_isoWT_1)	isogenic control line from SHOX2.2.6 with corrected mutation
SHOX2 B7-2F4-2D2 (= *SHOX2*_c.*28T>C_isoWT_1)	isogenic control line from SHOX2 B7 with corrected mutation
HANS6 (= control_1)	control iPSC clone derived from an individual with no known cardiovascular diseases

1.1.7. Oligonucleotides

Oligonucleotides (Primers) for cloning, sequencing and expression analysis were designed with Primer3, Primer3plus[114,115] and tested for specificity with *UCSC in silico PCR*. TaqMan probes for dPCR and genotyping were designed with Primer3plus and OligoAnalyzer 3.1. By default, primers were designed for an annealing temperature of 60°C and TaqMan probes for 65°C. For quantitative PCR, primers were designed with the *Universal ProbeLibrary Assay Design Center* (Roche).

Primers for mutagenesis were designed according to the recommendations in the manual *QuikChange Lightning Site-directed Mutagenesis Kit*.

Mutagenesis Primers

Name	Sequence (5' to 3')	Purpose
SHOX2_98_C_to_G	CGTACCGGGAGGTGCTGGAGAGCGG GCGGCTGCGCGGGGCCAAGGAGCCG	introducing the mutation c.98C>G into *SHOX2*
SHOX2_230_G_to_A	CGGAGGCGGCGGCGGAGGAGGCGGA GGAGATGTAGGAGGAGGAGGAGCAG	introducing the mutation c.230G>A into *SHOX2*
SHOX2_879_C_to_T	ACGAAAGCTGAGGTCCAGGCTACGCT GTTTCTCCCGGGCGAGGCGTTTCG	introducing the mutation c.879C>T into *SHOX2*

Sequencing primers

Primer name	Sequence (5' to 3')	Purpose
SHOX2 Ex1.1 for	TGTAAAACGACGGCCAGTAGAGGTTGAGCGCC GGGCTGACGT	Sequencing
SHOX2 Ex1.1 rev	GGATAACAATTTCACACAGGCCTACACCTCCTC CGCCTCCTCCG	
SHOX2 Ex 1.2 for	TGTAAAACGACGGCCAGTCGCAGCAGCCCGGC AGTCCGGGC	
SHOX2 Ex 1.2 rev	GGATAACAATTTCACACAGGCTGCCGGGGGTC AGTCAGGTCGT	
SHOX2 Ex2+ for	CCGAGTACTGGGTGATTG	
SHOX2 Ex2+ rev	GCCAAGACCCCTCGAACT	
SHOX2 Ex2 for	TCCACGAGGGGGAAGGATTC	
SHOX2 Ex2 rev	CACCAGACACTAGAAGCACCA	
SHOX2 Ex3 for	TGTAAAACGACGGCCAGTTTGCTTGCTGTATCT CCCAAT	
SHOX2 Ex3 rev	GGATAACAATTTCACACAGGTTTGCTCAGACTA TCAAATGTTCC	
SHOX2 Ex4 for	TGTAAAACGACGGCCAGTTTTGGAACCCTGAA AAATGC	
SHOX2 Ex4 rev	GGATAACAATTTCACACAGGGGCTCAGAGACA GGTGATG	
SHOX2 Ex5 for	TGTAAAACGACGGCCAGTCCCAAACACAACCC AACTCT	
SHOX2 Ex5 rev	GGATAACAATTTCACACAGGGCTGGGAACATC ATGTAGGG	
SHOX2 Ex6 for	AGGATAGTCATTGCAACGTGA	
SHOX2 Ex6 rev	TCTCAAAGGGGTAACGGAGA	

nCounter probes for Zebrafish gene expression analysis

Probe name	Sequence (5' to 3')	Purpose
actb1	GTGCTTCTAAACAGAACTGTTGCCACCTTAAATG GCCTAGCAATGAGATTCAAACGAACGACCAACC TAAACTCTCGAACAGAACAAGATGACATCAGCA	nCounter Control probe
eef1a1l1	GAAGGCTGCCAAGACCAAGTGAATTTCCCTCAA TCACACCGTTCCAAAGGTTGCGGCGTGTTCTTCC CAACCTCTTGGAATTTCTCTAAACCTGGGCACT	nCounter Control probe
rpl13a	AAGAGAAAGGAAAAGGCCAAGCTGCGCTATTCC AAGAAGAAAGTTGAGATGAAGCTGACTAAGCAG GCTGAAAAGAACGTTGAGAGCAAGATCGCAGTA T	nCounter Control probe
rps18	GTACAAAATCCCAGACTGGTTCCTGAACAGACA GAAGGACATAAAAGATGGGAAATACAGCCAGGT CCTTGCTAATGGTCTGGACAATAAACTGAGAGA A	nCounter Control probe
b2m	TACTTTCGATATCAACTGCTGTTGTCCTGAATGC TGAAGGATTGTCTGCTTGGCTCTCTCGAATAAAA CGGCCACAATGAGAGCACTCATCACTTTTGCA	nCounter Control probe
hsp90ab1	CTCACAGTCCGGCGACGAGATGACCTCCCTCAC AGAATACGTCAGCCGTATGAAGGAGAACCAAAA GTCCATCTATTACATCACTGGTGAGAGCAAAGA C	nCounter Control probe
bmp4	GAGCCAACACCGTGAGAGGATTCCATCATGAAG AGCACCTGGAGGAGCTGCAGTCAGACGGCTCCC AGGAGACGCCTCTGCGATTCGTTTTTAATCTCAG	nCounter Shox2 target gene probe

Primers for genome-editing methods

Primer name	Sequence (5' to 3')	Purpose
SHOX2_C849A_gRNA_1_fw	TAATACGACTCACTATAGCTGCA GCTGGACAGCG	*in vitro* synthesis *SHOX2* c.849C>A gRNA-1
SHOX2_C849A_gRNA_1_rev	TTCTAGCTCTAAAACCAGCGCTG TCCAGCTGCAG	
SHOX2_C849A_gRNA_2_fw	TAATACGACTCACTATAGCCAGG TGCGGATGCAG	

SHOX2_C849A_gRNA_2_rev	TTCTAGCTCTAAAACCCACCTGCATCCGCACCTG	*in vitro* synthesis *SHOX2* c.849C>A gRNA-2
SHOX2_C28T_gRNA_1_fw	TAATACGACTCACTATAGCGTGCAGGCTGAGTGC	*in vitro* synthesis *SHOX2* c.*28T>C gRNA-1
SHOX2_C28T_gRNA_1_rev	TTCTAGCTCTAAAACGCGGCACTCAGCCTGCACGC	
SHOX2_C28T_gRNA_2_fw	TAATACGACTCACTATAGGCGTGCAGGCTGAGTG	*in vitro* synthesis *SHOX2* c.*28T>C gRNA-2
SHOX2_C28T_gRNA_2_rev	TTCTAGCTCTAAAACCGGCACTCAGCCTGCACGCG	
SHOX2_C849A_ddPCR_fw	CTTGTCCTTTCAGGTTCAGG	dPCR primer
SHOX2_C849A_ddPCR_rev	GCGGATGCAGGTGGT	
SHOX2_T28C_ddPCR_fw	TGAAAGCCAAAAAGCACGCC	dPCR primer
SHOX2_T28C_ddPCR_rev	GGCGTGCAGGCTGAGTG	
SHOX2_849A_ddPCR_FAM	[FAM]TGTGGCGCAAGCGCAC[BHQ1]	*SHOX2* c.849A detection
SHOX2_849C_ddPCR_HEX	[HEX]TGTGGCGCACGCGCAC[BHQ1]	*SHOX2* c.849C detection
SHOX2_28C_ddPCR_FAM	[6FAM]CCAATGTCGCGCCCGTCCCGC[BHQ1]	*SHOX2* c.*28C detection
SHOX2_28T_ddPCR_HEX	[HEX]CACCAATGTCGCGCCTGTCCCGC[BHQ1]	*SHOX2* c.*28T detection
SHOX2_C849A_NGS_fw	AACCCAACTCTCTCTCTGGC	Primer for *SHOX2* c.849C>A NGS amplicon
SHOX2_C849A_NGS_rev	GTTCTTGCTGGTGGTCTTGG	
SHOX2_T28C_NGS_fw	GACCACCAGCAAGAACTCCA	Primer for *SHOX2* c.*28T>C NGS amplicon
SHOX2_T28C_NGS_rev	GTCTGGCTTTCCGAGTCCAA	
SHOX2_C849A_T7_fw	ATTTCAGCCCAGTTCCCAGG	Sanger sequencing *SHOX2* c.849C>A region
SHOX2_C849A_T7_rev	AGAGTCCATCGTTGCAGCTT	
SHOX2_C28T_Ex6_fw	AGGATAGTCATTGCAACGTGA	Sanger sequencing *SHOX2* c.*28T>C region
SHOX2_C28T_Ex6_rev	TCTCAAAGGGGTAACGGAGA	

Primers for gRNA off-target sequencing

Primer name	Sequence (5' to 3')	Purpose
KCNMA1_fw	AATGCCACAATCACGGATGC	*SHOX2* c.849C>A gRNA-1 off-target 1
KCNMA1_rev	GATTTCCTGACCCCTGACCC	

RNA5SP462_fw	GAATGTTCTCTCCGGCTCCC	*SHOX2* c.849C>A
RNA5SP462_rev	GAGAGCCATGAGACTGGAGC	gRNA-1 off-target 2
LINC00315_fw	GTGAGAGGTGAGGGTCAAGC	*SHOX2* c.849C>A
LINC00315_rev	AGGGTGGTGTTGAGCTAACG	gRNA-1 off-target 3
RP11-422N16.3_fw	GGGTTCCGGGATCTGAAAGG	*SHOX2* c.849C>A
RP11-422N16.3_rev	CCTCACCTAACCCTCACAGC	gRNA-1 off-target 4
RAB10_fw	AAGCAGGTTCTCTGAGCACC	*SHOX2* c.849C>A
RAB10_rev	AATGCTCCTGTCTCCAAGCC	gRNA-1 off-target 5
SUN5_fw	AGGCGTCAGATACAAGCAGG	*SHOX2* c.849C>A
SUN5_rev	TTACCTGTCCAACCTCACGC	gRNA-1 off-target 6
ING5_fw	GACCTGAAGTGATCCACCCG	*SHOX2* c.849C>A
ING5_rev	TTGGGTCCAAACTACAGCCC	gRNA-1 off-target 7
LINGO4_fw	ATTGTTAGCAGTGTGGGGGC	*SHOX2* c.849C>A
LINGO4_rev	TGCTATTGCATGACCCCTCC	gRNA-1 off-target 8
SYNDIG1_fw	TTCACCAAGGCCTTTCCAGG	*SHOX2* c.849C>A
SYNDIG1_rev	TACAGGAAGAGGGAGAGGCC	gRNA-1 off-target 9
FCN2_fw	CCTGGAGATGATCTCGCACC	*SHOX2* c.849C>A
FCN2_rev	GCTTATCCCCACCTCACACC	gRNA-1 off-target 10
CPSF2_fw	CATGTGGAGTCTCAGGTCCC	*SHOX2* c.849C>A
CPSF2_rev	GTAGGACATAGCCCCACTGC	gRNA-1 off-target 11
POLR1A_fw	GGCCAAGAAGTGAATGCTGC	*SHOX2* c.849C>A
POLR1A_rev	ATTGAGACCAGGCCTGTTGG	gRNA-1 off-target 12
TSHZ3_fw	AATGCTGGGCTGAGAAGTCC	*SHOX2* c.*28T>C
TSHZ3_rev	CCTTTGTTCCTTCAAGCAGGC	gRNA-1 off-target 1
ABCF1_fw	CTGTTCTCTCCTGGCAGTGG	*SHOX2* c.*28T>C
ABCF1_rev	CAGGTGGTCAAAGAGGTCCC	gRNA-1 off-target 2
UPK2_fw	AGGCAGAGGAAATTCCAGGC	*SHOX2* c.*28T>C
UPK2_rev	CTTGGGTGCTGAGAGTGAGG	gRNA-1 off-target 3
TEF_fw	CTTGCTTTAGGGGAGCCTCC	*SHOX2* c.*28T>C
TEF_rev	TTTCGCTTTTGTTGCCCAGG	gRNA-1 off-target 4
TRIM71_fw	GTTCTGTGCCTCCAGAGTCC	*SHOX2* c.*28T>C
TRIM71_rev	GGCGGGCTACAATGTTTTCC	gRNA-1 off-target 5
TMEM121_fw	CCGTTCCGTTTCCTTCCTGG	*SHOX2* c.*28T>C
TMEM121_rev	AGAGCTTGATCTCCAGCACG	gRNA-1 off-target 6
LANCL2_fw	GGGCGTTCGGTTTTCTTTGG	*SHOX2* c.*28T>C
LANCL2_rev	AACACAGCCTGTCTCTTCGG	gRNA-1 off-target 7
SAMD4A_fw	GCAGGAGTAGGAAAGGCTGG	*SHOX2* c.*28T>C
SAMD4A_rev	CACTTGCTGGGATGGGTTCC	gRNA-1 off-target 8
POU5F1P2_fw	TTGGAAACCACAGGCAGAGG	*SHOX2* c.*28T>C
POU5F1P2_rev	TCTCAGTGACACACCACACG	gRNA-1 off-target 9
TPRN_fw	CTCCATAACTGGCTTGGGGG	*SHOX2* c.*28T>C
TPRN_rev	CGTCCTCATCATCGAGACGG	gRNA-1 off-target 10
PLCH2_fw	ATTCAGGCTGAGCTGTCACG	*SHOX2* c.*28T>C
PLCH2_rev	TCCCATCCCCACCAGATAGG	gRNA-1 off-target 11
KRT16_fw	GTGAAGCTTGCAGTGAACCG	*SHOX2* c.*28T>C
KRT16_rev	CCTCACACCCCATCAACTCC	gRNA-1 off-target 12

Primers for iPSC (re-)characterization

Primer name	Sequence (5' to 3')	Purpose
SEV_transgene_fw	GGATCACTAGGTGATATCGAGC	RT-PCR, Sendai virus transgene expression
SEV_ transgene_rev	ACCAGACAAGAGTTTAAGAGATATGTATC	
SOX2 trans_fw	ATGCACCGCTACGACGTGAGCGC	RT-PCR, Sendai virus transgene expression
SOX2 trans_rev	AATGTATCGAAGGTGCTCAA	
KLF4 trans_fw	TTCCTGCATGCCAGAGGAGCCC	RT-PCR, Sendai virus transgene expression
KLF4 trans_rev	AATGTATCGAAGGTGCTCAA	
cMYC trans_fw	TAACTGACTAGCAGGCTTGTCG	RT-PCR, Sendai virus transgene expression
cMYC trans_rev	AATGTATCGAAGGTGCTCAA	
OCT3/4 trans_fw	CCCGAAAGAGAAAGCGAACCAG	RT-PCR, Sendai virus transgene expression
OCT3/4 trans_rev	AATGTATCGAAGGTGCTCAA	
OCT3/4_fw	GACAGGGGGAGGGGAGGAGCTAGG	qRT-PCR, iPSC characterization
OCT3/4_rev	CTTCCCTCCAACCAGTTGCCCCAAAC	
SOX2_fw	GGGAAATGGGAGGGGTGCAAAAGAGG	qRT-PCR, iPSC characterization
SOX2_rev	TTGCGTGAGTGTGGATGGGATTGGTG	
NANOG_fw	TGCAAGAACTCTCCAACATCCT	qRT-PCR, iPSC characterization
NANOG_rev	ATTGCTATTCTTCGGCCAGTT	
REX1_fw	ACCAGCACACTAGGCAAACC	qRT-PCR, iPSC characterization
REX1_rev	TTCTGTTCACACAGGCTCCA	
TDGF1_fw	CCCAAGAAGTGTTCCCTGTG	qRT-PCR, iPSC characterization
TDGF1_rev	ACGTGCAGACGGTGGTAGTT	
PDX1_fw	AAGCTCACGCGTGGAAAG	qRT-PCR, iPSC characterization
PDX1_rev	GGCCGTGAGATGTACTTGTTG	
SOX7_fw	TGAACGCCTTCATGGTTTG	qRT-PCR, iPSC characterization
SOX7_rev	AGCGCCTTCCACGACTTT	
AFP_fw	GTGCCAAGCTCAGGGTGTAG	qRT-PCR, iPSC characterization
AFP_rev	CAGCCTCAAGTTGTTCCTCTG	
CD31_fw	ATGCCGTGGAAAGCAGATAC	qRT-PCR, iPSC characterization
CD31_rev	CTGTTCTTCTCGGAACATGGA	
DESMIN_fw	GTGAAGATGGCCCTGGATGT	qRT-PCR, iPSC characterization
DESMIN_rev	TGGTTTCTCGGAAGTTGAGG	
ACTA2_fw	GTGATCACCATCGGAAATGAA	qRT-PCR, iPSC characterization
ACTA2_rev	TCATGATGCTGTTGTAGGTGGT	
SCL_fw	CCAACAATCGAGTGAAGAGGA	qRT-PCR, iPSC characterization
SCL_rev	CCGGCTGTTGGTGAAGATAC	
MYL2_fw	TACGTTCGGGAAATGCTGAC	qRT-PCR, iPSC characterization
MYL2_rev	TTCTCCGTGGGTGATGATG	
KRT14_fw	CACCTCTCCTCCTCCCAGTT	qRT-PCR, iPSC characterization
KRT14_rev	ATGACCTTGGTGCGGATTT	
NCAM1_fw	CAGATGGGAGAGGATGGAAA	qRT-PCR, iPSC characterization
NCAM1_rev	CAGACGGGAGCCTGATCTCT	
TH_fw	TGTACTGGTTCACGGTGGAGT	qRT-PCR, iPSC characterization
TH_rev	TCTCAGGCTCCTCAGACAGG	

GABRR2_fw	CTGTGCCTGCCAGAGTTTCA	qRT-PCR, iPSC characterization
GABRR2_rev	ACGGCCTTGACGTAGGAGA	
SOX17_fw	ACGCCGAGTTGAGCAAGA	qRT-PCR, iPSC characterization
SOX17_rev	TCTGCCTCCTCCACGAAG	
GATA6_fw	ACCACCTTATGGCGCAGAAA	qRT-PCR, iPSC characterization
GATA6_rev	ATAGCAAGTGGTCTGGGCAC	
CDH5_fw	TGTCCTTGTCTATTGCGGAGA	qRT-PCR, iPSC characterization
CDH5_rev	CCTACCAGCCCAAAGTGTGT	
DES_fw	GTGAAGATGGCCCTGGATGT	qRT-PCR, iPSC characterization
DES_rev	GGGCTGGTTTCTCGGAAGTT	
TH_fw	GCCCTACCAAGACCAGACGTA	qRT-PCR, iPSC characterization
TH_rev	CGTGAGGCATAGCTCCTGA	
GABRR2_fw	TCACTGGGTATCACGACGGTG	qRT-PCR, iPSC characterization
GABRR2_rev	CAGCACCGAGAGGAACACGA	
KRT14_fw	CCATTGAGGACCTGAGGAAC	qRT-PCR, iPSC characterization
KRT14_rev	CAATCTGCAGAAGGACATTGG	
hPAX6_fw	TCAGAGCCCCATATTCGAGC	qRT-PCR, iPSC characterization
hPAX6_rev	CAAAGACACCACCGAGCTGA	
NESTIN_fw	GAGGTGGCCACGTACAGG	qRT-PCR, iPSC characterization
NESTIN_rev	AAGCTGAGGGAAGTCTTGGA	
hWNT3_fw	ATCTACGACGTGCACACCTG	qRT-PCR, iPSC characterization
hWNT3_rev	TGCTTCCCATGAGACTTCGC	

Single-stranded oligodeoxynucleotides (ssODNs)

ssODN name	Sequence (5' to 3')	Purpose
SHOX2_C28T_ssODN	ATCGCCGATCTCAGACTGAAAGCCAAAAAGCACGCCGCAGCCCTGGGTCTGTGACGCCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGGCACTCAGCCTGCACGCCCTCCGCGCCCCGCTGCTTC	ssODN to correct SHOX2 c.*28T>C
SHOX2_T28C_ssODN	ATCGCCGATCTCAGACTGAAAGCCAAAAAGCACGCCGCAGCCCTGGGTCTGTGACGCCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGGCACTCAGCCTGCACGCCCTCCGCGCCCCGCTGCTTC	ssODN to insert SHOX2 c.*28T>C
SHOX2_A849C_ssODN	CACCCTAGGATAGTCATTGCAACGTGACGCCCTTGTCCTTTCAGGTTCAGGCGCAGCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCCGCGCA	ssODN to correct SHOX2 c.849C>A
SHOX2_C849A_ssODN	CACCCTAGGATAGTCATTGCAACGTGACGCCCTTGTCCTTTCAGGTTCAGGCGCAGCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCCGCGCA	ssODN to insert SHOX2 c.849C>A

Antibodies

Target	Host	Source	Catalog number	Dilution
anti-rabbit IgG-AlexaFluor488	goat	Thermo Fisher Scientific	A32731	1:500

anti-rabbit IgG-AlexaFluor594	goat	Thermo Fisher Scientific	A32740	1:500
NANOG	rabbit	Abcam	ab21624	1:500
OCT4	rabbit	Abcam	ab19857	1:500
SOX2	rabbit	Abcam	ab137385	1:200
TRA1-81-AlexaFluor488	mouse	BD Pharmingen	560174	1:20

1.1.8. Databases

Database name	
1000 genomes	http://www.1000genomes.org
CCTop CRISPR/Cas9 target online predictor	http://crispr.cos.uni-heidelberg.de/
Cas-Analyzer	http://www.rgenome.net/cas-analyzer/#!
CHOPCHOP	http://chopchop.cbu.uib.no/
Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine program	https://www.nhlbiwgs.org/
Ensembl Genome Browser	http://ensembl.org/index.html
Genecards	https://www.genecards.org/
gnomAD	https://gnomad.broadinstitute.org/
GTEx portal:	https://gtexportal.org/home/index.html
GWAS Catalog	https://www.ebi.ac.uk/gwas/
NCBI	http://www.ncbi.nlm.nih.gov/
OligoAnalyzer 3.1	https://www.idtdna.com/pages/tools/oligoanalyzer
Primer3	http://primer3.ut.ee/
Primer3plus	https://primer3plus.com/
QuikChange Primer Design	https://www.agilent.com/store/primerDesignProgram.jsp
SNV lookup	https://cadd.gs.washington.edu/snv
The Human Protein Atlas	https://www.proteinatlas.org/
UCSC	http://genome.ucsc.edu/
UCSC in silico PCR	http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr
Universal Probe Library Assay Design Center	https://lifescience.roche.com/en_de/brands/universal-probe-library.html
Pixabay	https://pixabay.com/de/

1.1.9. Instruments

Instrument	**Company**
Automated Inverted Microscope DMI4000B	Leica
DS-11 FX spectrophotometer	DeNovix
Maxisafe 2020 Laminar flow hood	Thermo Fisher Scientific
Microbiological incubator	WTB Binder
QuantStudio3	Applied Biosystems
QUANTUM Gel Documentation System	Peqlab
QX200 Droplet Generator	Bio-Rad
QX200 Droplet Reader	Bio-Rad
Steri Cult CO_2 incubator	Thermo Fisher Scientific
Thermal Cycler Mastercycler pro-vapo protect	Eppendorf
Thermomixer	Eppendorf

1.2. Zebrafish-based Methods

Zebrafish experiments were performed in collaboration with AG Just by Dr. Christoph Paone and Sabrina Diebold (Department of Internal Medicine II, University of Ulm, Ulm, Germany).

Care and breeding of zebrafish, *Danio rerio*, was done as described previously.[116] The TE4/6 wildtype strain was used for all injection procedures. 20 ng/µl of pDestTol2CG2-empty, pDestTol2CG2-SHOX2 wildtype, or pDestTol2CG2-SHOX2 mutant was microinjected into one-cell-stage embryos for overexpression experiments as previously described.[22,28] Morphological analysis and heart rates were determined 72 hours post fertilization (hpf). For target gene expression using nCounter technology, hearts were isolated 72 hpf as well.[1]

1.3. DNA-based Methods

1.3.1. Polymerase chain reaction (PCR)

For subsequent applications such as cloning, sequencing and expression analysis, DNA amplification was performed with PCR according to manufacturer's instructions. See **Table A** and **Table B** for reaction composition.

Component	**Final concentration**	**Program**		
10x PCR Buffer	1x			
dNTP mix (10 mM each)	200 µM of each dNTP	95°C	15 min	x30
Fw+rev Primer (10 µM each)	0.1-0.5 µM	94°C	30 s	
HotStarTaq DNA Polymerase	2.5 units/reaction	60°C	30 s	
Template DNA	50-200 ng	72°C	1 min	
H_2O ad total volume	variable	72°C	10 min	
Total volume	Variable (25 or 100 µl)			

Table A: standard reaction composition using the HotStarTaq DNA Polymerase according to the *HotStarTaq® PCR Handbook* (10/2010 ©2007-2010 QIAGEN)

Component	**Final concentration**	**Program**		
5x Q5 Reaction Buffer	1x			
dNTP mix (10 mM each)	200 µM of each dNTP	98°C	30 sec	x30
Fw+rev Primer (10 µM each)	0.5 µM	98°C	10 s	
Q5 High-Fidelity DNA Polymerase	0.02 U/µl	60°C	30 s	
Template DNA	50-200 ng	72°C	30 s/kb	
H2O ad total volume	variable	72°C	2 min	
Total volume	Variable (25 or 50 µl)			

Table B: standard reaction composition using the Q5® High-Fidelity Polymerase according to the *Datasheet for Q5® High-Fidelity DNA Polymerase* (M0491)

1.3.2. Agarose gel electrophoresis

Agarose gels were run for size separation and gel extraction of PCR product and digested DNA vectors. Depending on the expected fragment sizes, 1-2% agarose gels were run at 80-120 V for 30-120 min. For size comparison, GeneRuler™ 100 bp or 1 kb Plus DNA Ladders were loaded on

the gel in addition to the samples. DNA was stained with Midori Green advanced or ethidium bromide and visualized under UV light.

1.3.3. Plasmid DNA and PCR fragment purification

PCR fragments were purified with the QIAquick® PCR Purification Kit as described in the *QIAquick® Spin Handbook 03/2008*. Small scale extraction of plasmid DNA was performed from 5 ml bacteria LB culture using the GeneJET Plasmid Miniprep Kit (for details see *Thermo Scientific GeneJET Plasmid Miniprep Kit Quick protocol*, ©2012 Thermo Fisher Scientific Inc.). Large scale extraction of plasmid DNA was performed from 50 ml bacteria LB culture using the ZymoPURE™ II Plasmid Midiprep Kit with the vacuum protocol (for details see *Instruction Manual ZymoPURE™ II Plasmid Midiprep Kit*, version 1.2.0). The DNA was eluted in H_2O or the included elution buffer and subsequently stored at -20°C.

1.3.4. Restriction digestion

DNA vectors and PCR fragments were digested with FastDigest® restriction enzymes according to the recommendations.

1.3.5. Dephosphorylation

Digested DNA vectors were dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase to prevent re-ligation according to manufacturer's instruction (for details see manual *Thermo Scientific FastAP Thermosensitive Alkaline Phosphatase*, MAN0012876, Rev. B.00).

1.3.6. Ligation

Purified inserts and vectors were mixed in a molar ratio of 1:3 and ligated using T4 DNA Ligase according to manufacturer's instructions (for details see manual *DNA Insert Ligation (sticky-end and blunt-end) into Vector DNA*, Thermo Fisher Scientific).

1.3.7. Gateway® gene cloning

Cloning of *SHOX2* into the pDONR™221 vector and subsequent generation of *SHOX2* overexpression constructs for zebrafish was done by Sandra Hoffmann (Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany) using the Gateway® cloning system. For details to reaction setups, transformation and clone selection see User Guide *Gateway® pDONR™ vectors*, MAN0000291, Publication Part number 25-0531, Rev 29 March 2012.

1.3.8. Site-directed mutagenesis

Patient-specific mutations were introduced in a pDest27 or pDestTol2CG2 vector containing the *SHOX2* wildtype sequence with the QuikChange Lightning Site-Directed Mutagenesis Kit. Primer design and reaction setup were carried out following the manufacturer's manual (for details see *QuikChange Lightning Site-Directed Mutagenesis Kit*, Instruction Manual 210518-12 Revision F.0).

1.3.9. Transformation

Chemically competent cells were prepared with the Mix & Go *E.coli* Transformation Kit and Buffer Set following the provided protocol (for details see *Mix & Go E.coli Transformation Kit and Buffer Set*, Instruction Manual, Ver. 1.0.7). E. coli were transformed with plasmids and ligations mixes:

- Add 1-2 µl ligation or <1 ng plasmid DNA to 100 µl of thawed cells
- Mix carefully and incubate 5-10 min on ice
- Add 500 µl of pre-warmed SOC medium
- Incubate cells 30-60 min at 37°C in a bench shaker at 300 rpm
- Distribute 10-100 µl of bacteria on pre-warmed LB-agar plates with glass beads containing appropriate selection antibiotics and grow overnight at 37°C.

1.3.10. Genomic DNA (gDNA) extraction and purification

Genomic DNA was extracted from pelleted cells either directly after centrifugation or from snap-frozen cell pellets using the *Quick*-DNA™ Miniprep Plus Kit according to manufacturer's instructions (for details see *Instruction Manual Quick-DNA™ Miniprep Plus Kit*, version 1.2.1). For high throughput (96-well) DNA purification from iPSC sib-selections (see chapter 1.3.10), the *Quick*-DNA™ 96 Plus Kit was used (for details see *Instruction Manual Quick-DNA™ 96 Plus Kit*, version 1.1.0). The purified DNA was eluted in TE buffer or H_2O and stored at -20°C.

1.3.11. DNA quantification

Double-stranded DNA was quantified and tested for purity with the NanoDrop 1000 system or the DeNovix system according to manufacturer's instructions. For precise measurements, DNA was analyzed with the Qubit™ dsDNA BR Assay Kit according to manufacturer's instructions (see *Qubit® dsDNA BR Assay* Kits, MAN0002325, Revision A.0) on the Qubit® 2.0 Fluorometer using 1 µl DNA solution as input.

1.3.12. Digital PCR (dPCR)

Relative allele quantification in sib-selections was performed with dPCR, here ddPCR™. Allele-specific TaqMan probes were used to discriminate wildtype from mutant alleles in *SHOX2* c.*28T>C and *SHOX2* c.849C>A iPSC clones. *SHOX2 c.*28T* and *SHOX2 c.849C* alleles were referred to as "WT alleles" and *SHOX2 c.*28C* and *SHOX2 c.849A* as "Mut alleles", respectively. The dPCR reaction was prepared with the ddPCR™ Supermix for probes as described in the official protocol in PCR 8-tube strips (for details see *ddPCR™ Supermix for Probes*, 10026235 Rev C). 1 µl FastDigest HindIII enzyme was included in the PCR reaction to improve droplet generation. The composition of the Primer/Probe mix and the dPCR reaction is listed in **Table C** and **Table D** respectively.

Component	Volume per reaction [µl]	Final concentration
dPCR primer fw [100 µM]	18	18 µM
dPCR primer rev [100 µM]	18	18 µM
FAM TaqMan probe [100 µM]	5	5 µM
HEX TaqMan probe [100 µM]	5	5 µM
H_2O ad total volume	54	-
Total volume	100	-

Table C: Primer/Probe mix for 100 µl dPCR reactions

Component	Volume per reaction [µl]	Final concentration	Program	
2x ddPCR™ Supermix for probes	11	1x		
20x target primer/probe mix (FAM/HEX)	1.1	900 nM/250 nM	95°C 10 min	
FastDigest HindIII	1.1	1x	94°C 30 sec	x40
Template DNA (50-150 ng/reaction)	1.1 - 8.8	2.5 - 7.5 ng/µl	65°C 1 min	
H_2O ad total volume	variable	variable	98°C 10 min	
Total volume	22	-		

Table D: standard dPCR reaction composition. The final volume of 22 µl includes 10% excess. 20 µl of the PCR mix are used for the droplet generation. A ramp rate of 2°C/sec was set for the PCR program.

The dPCR reaction was pipetted under sterile conditions in a PCR UV cabinet. The preparation of PCR samples for droplet generation took place at the QX200™ Droplet Generator. The dPCR droplet generation was performed at the Dept. of Infectious Diseases, Virology, Heidelberg University Hospital (under supervision of Dr. Kathleen Börner and Prof. Dirk Grimm). Rainin LTS pipets and tips were used during this process exclusively in accordance with manufacturer's recommendations. The dPCR setup in detail:

- Mount DG8™ cartridge in DG8™ cartridge holder
- Pipet 70 µl of droplet generation oil for probes in the respective wells of the DG8™ cartridge
- Vortex and spin down prepared dPCR samples in PCR 8-tube strips
- Pipet 20 µl of each dPCR sample in the respective wells of the DG8™ cartridge
- Span DG8™ gasket over DG8™ cartridge and load into QX200™ Droplet generator
- After droplet generation, load 42 µl of droplet suspension into dPCR 96-well plate
- Seal plate with pierceable foil heat seal for 5 sec at 180°C in PX1™ PCR plate sealer
- Immediately load plate into C1000 Touch™ Thermal Cycler and run PCR program
- After end-point PCR, span the dPCR 96-well plate into QX200™ Droplet Reader and load prepared template into *QuantaSoft*™ Software (version 1.7.4.0917)
- Start dPCR Droplet Readout

For details to dPCR template generation and analysis of the results see chapter 1.6.1

1.3.13. Sanger Sequencing

Sanger Sequencing of cloning constructs and PCR products was carried out by GATC (Ebersberg), EUROFINS (Ebersberg) and GENEWIZ (Leipzig). Samples were prepared according to company's guidelines.

1.3.14. Next generation sequencing (NGS) (sample preparation of sib-selections

The sib-selections with the highest abundance of WT alleles, as detected by dPCR, were analyzed in depth via next generation sequencing. DNA was extracted from thawed sib-selections. A PCR was done in duplicates using the Q5® high fidelity DNA Polymerase with 100-200 ng gDNA input and 30 reaction cycles. The PCR products were pooled, purified with the QIAquick® PCR Purification Kit and sent for NGS to GENEWIZ (Leipzig).

1.3.15. directPCR® lysis and genotyping

iPSC colonies derived from single cell-cloning were screened for *SHOX2 c.*28T* homozygosity with a TaqMan probe-based genotyping assay. The protocol was adapted from the *TaqPath*™ *ProAmp*™ *Master Mixes User Guide*. 50% iPSCs of a 96-well were lysed in 70 µl of directPCR® lysis reagent and with 0.2 mg/ml Proteinase K at 55°C for 16h. The Proteinase K was subsequently inactivated at 85°C for 45 min. The cell lysate was used directly for genotyping without further dilution. The 20x target primers/probes mix was generated as described in Table 1.3. The genotyping reaction composition is listed in **Table E**. Samples were run in duplicates; gDNA from

homozygous wildtype iPSC clones (control clones) and heterozygous *SHOX2* c.*28T>C iPSC and *SHOX2* c.849C>A clones (patient-specific clones) were used as positive and negative control, respectively.

Component	Volume per reaction [µl]	Final concentration	Program	
2x TaqPath™ ProAmp™ Master Mix	5	1x	60°C 30 sec	
20x target primers/probes (FAM/HEX)	0.5	900 nM/250 nM	95°C 5 min	
directPCR® cell lysate	2.5	≤25%	95°C 5 sec	x40
H_2O ad total volume	2	-	60°C 30 sec	
Total volume	10	-	60°C 30 sec	

Table E: standard genotyping reaction composition for the 96-well format.

1.3.16. The Multiplex Human Cell Line Authentication Testing

To exclude cross contaminations with other cell lines and to verify a common origin of all patient-specific iPSC clones and their respective isogenic controls, cell lines were authenticated using Multiplex Cell Authentication (MCA) by Multiplexion (Heidelberg, Germany) as previously described.[117] 15 µl of genomic DNA with a concentration of 15-30 ng/µl was provided as requested by the company. Single Nucleotide Polymorphism (SNP) genotyping was performed on 48 loci and the resulting SNP profile compared to the company's reference database.

1.4. RNA-based methods

1.4.1. *In vitro* gRNA synthesis

The gRNAs for the electroporation of Cas9 RNP/gRNA complexes was synthesized *in vitro* with the *GeneArt™ Precision gRNA Synthesis Kit* according to the manufacturer's manual (for details see *GeneArt™ Precision gRNA Synthesis Kit user guide)*. Primers were designed as described. The lyophilized and desalted Primers were diluted in Nuclease-free water. The purified gRNA was diluted in nuclease-free water to a concentration of 500-1000 ng/µl and measured with the Qubit™ RNA BR Assay Kit on the Qubit™ 2.0 Fluorometer using a 1 µl of a 1:100 diluted gRNA as input (for details see *Qubit® dsDNA BR Assay* Kits, MAN0002325, Revision A.0).

1.4.2. RNA isolation from zebrafish hearts

Total RNA from Zebrafish hearts 72 hpf was isolated with the Direct-zol™ RNA Microprep Kit (Zymo Research) according to manufacturer's instructions (for details see instruction manual *Direct-zol™ RNA MicroPrep* Ver.1.0.0).[1]

1.4.3. RNA isolation from iPSCs

For RNA extraction from cultured iPSCs, cells were pelleted, resuspended in TRIzol™ and either stored at -80°C or immediately processed. After adding 200 µl of chloroform per 1 ml TRIzol™, samples were vortexed thoroughly and incubated at RT for 5 min. The aqueous phase containing RNA was separated from the organic phase by centrifugation for 5 min and 12,000 g at 4°C and transferred into a new 1.5 ml tube. 10 µg of glycogen and 500 µl of isopropanol per 1 ml TRIzol™ were added and the solution was mixed by inversion. The samples were incubated for 60 min at RT and centrifuged for 10 min and 12,000 g at 4°C. After removal of the supernatant, the RNA pellet was washed twice by addition of 1 ml 75% EtOH (-20°C), vortexing and subsequent centrifugation for 5 min and 7,500 g at 4°C. The pellet was air-dried for 10 min and resuspended in 20 µl RNAse-free water. RNA concentration and purity were determined on a DeNovix DS-11 spectrophotometer.

For the characterization of patient-specific iPSC lines, the RNA was extracted with TRIzol™ reagent or Absolutely RNA Miniprep Kit in collaboration with Dr. Svenja Laue, Birgit Campbell, Dr. Tatjana Dorn and Alessandra Moretti (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany).

1.4.4. qRT-PCR

cDNA reverse transcription was performed using the SuperScript III First-Strand Synthesis Kit (Invitrogen) with 1 µg of RNA as input. For semi-quantitative analysis, 1 µl cDNA was subjected to PCR reaction using Taq polymerase (Thermo Fisher Scientific). For qRT-PCR analysis, 1-2 µl cDNA and the Power SYBR Green PCR Master Mix (Applied Biosystems) or the qPCRBIO SyGreen Mix (Nippon Genetics) were used. Real-time quantitative PCR (qPCR) was conducted on a QuantStudio3 System (Thermo Fisher Scientific). All samples were measured in duplicates and the relative gene expression levels were normalized to the reference genes *SDHA* and *HPRT1* or *GAPDH* using the Relative Standard Curve Method or semi-quantitively.

1.4.5. nCounter expression analysis

For *Bmp4* expression analysis in zebrafish, 40 hearts from two independent experiments were pooled per condition to obtain 50 ng of input material. The procedure was performed by the nCounter Core Facility Heidelberg using the nCounter SPRINT Profiler. The workflow is described in detail at http://www.nanostring.com/elements/workflow. Background correction and normalization of data was performed using the nSolver Analysis Software 4.0 (NanoString Technologies). The most stable expressed genes were chosen for normalization based on the geNorm method.[1]

1.5. Cell-based Methods

1.5.1. Cultivation and passaging of HEK293T cells

HEK293TN cells (BioCat) were maintained in DMEM high glucose medium supplemented with 10% FCS and 100 U/ml Pen-Strep at 37°C, 5% CO_2 in a humidified atmosphere. Cells were passaged every 3-4 days by aspiration of culturing medium, washing cells once in pre-warmed PBS and adding 0.05% Trypsin-EDTA. Cells were incubated until detachment from the culturing flask. Trypsinization was stopped by addition of pre-warmed culturing medium and a fraction of the cell suspension was seeded on new flasks.

1.5.2. PEI transfection for luciferase assays

HEK293TN cells were co-transfected with the BMP4 pGL3-basic reporter construct (1 µg) together with SHOX2 WT or mutant expression constructs (1 µg) and TK-Renilla Firefly (500 ng) using PEI at a DNA:volume ratio of [1µg]:[3µl]. Transfections were performed in the 6-well format. Cells were subjected to passive cell lysis 24h after transfection.[19]

1.5.3. Luciferase Assay

Luciferase reporter assays were performed with the Dual-Luciferase® Reporter Assay System (Promega) according to the manufacturer's protocol (for details see Technical Manual, TM040, Revised 06/15). In brief: 24h after transfection, cells were washed twice with PBS and lysed with 150 µl (6-well format) of Passive Lysis Buffer per well. For lysis, cells were incubated for 15 min at RT and subjected to one freeze/thaw cycle at -80°C before analysis. Cell lysates were scraped off the culturing plate with a pipette tip and aliquoted in white 96-well polypropylene flat bottom plates (Greiner Bio-Rad). Luciferase activity was measured in a Berthold Centro LB 960 Luminometer. Experiments were performed independently four times with technical triplicates for each sample.

1.5.4. Generation of patient-specific iPSC lines

AF patients carrying the *SHOX2* mutations were recruited by Prof. Stefan Kääb and Dr. Sebastian Clauss (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany). Peripheral blood mononuclear cells (PBMCs) were isolated and reprogrammed into iPSCs in collaboration with Dr. Svenja Laue, Dr. Tatjana Dorn, Birgit Campbell, Prof. Alessandra Moretti and Prof. Karl-Ludwig Laugwitz (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany) using the CytoTune-iPS 2.0 Sendai Reprogramming Kit (Life Technologies), as previously described.[56]

1.5.5. Conditions for standard iPSC cultivation and differentiation

Standard cultivation of iPSCs was performed with 5% CO_2 at 37°C in a humidified atmosphere. Cells were regularly tested for mycoplasma contamination by colleagues using the Venor®GeM Classic Kit according to manufacturer's instructions (for details see *Venor®GeM Classic Instructions for Use*, Document Version 32). Contaminated cell lines were treated with Plasmocin™ (1:1,000) in culturing medium for 2-3 weeks and re-tested before usage in differentiation and gene editing.

1.5.6. Colony-based cultivation and small clump-passaging of iPSCs with Essential 8™ and EDTA

Colony-based cultivation of iPSCs in Essential 8™ on Geltrex™ LDEV-free, hESC-qualified, Reduced Growth Factor Basement Membrane matrix (1:100) was done following manufacturer's recommendations (for details see *Essential 8™ Medium*, Publication Number MAN0007569, Revision 3.0). Every 3-4 days or upon 70-80% confluency, cells were washed twice with pre-warmed DPBS and incubated in 0.5 mM UltraPure™ EDTA at room temperature (RT) for 7-10 min. After aspiration of EDTA, cells were resuspended with a 5 ml Stripette™ in 1 ml Essential 8™ supplemented with 10 µM Y-27632 and split 1:6-1:20 into new 3.5 cm dishes. Y-27632 was removed 24h after splitting.

1.5.7. Single cell-based cultivation and passaging of iPSCs with StemFit®

Cells were grown in StemFit® supplemented with 100ng/ml bFGF on Geltrex™ LDEV-free, hESC-qualified, Reduced Growth Factor Basement Membrane matrix (1:100) according to manufacturer's instructions. After reaching 70-85% confluency, cells were passaged following this protocol:

- Wash 2x with pre-warmed DPBS
- Add pre-warmed TrypLE™ Express
- Incubate at 37°C, 5% CO_2 for 5-10 min
- Resuspend cells in TrypLE™ Express and transfer to 50 ml conical Falcon™
- Dilute cell suspension 1:10 with pre-warmed DMEM/F12(1:1)
- Centrifuge at 200 g for 5 min in swing-bucket rotor centrifuge
- Remove supernatant and resuspend cells in pre-warmed StemFit® supplemented with 100 ng/ml bFGF and 10 µM Y-27632
- Count cells in Neubauer counting chamber and seed 1,000-2,000 cells/cm^2 on a new Geltrex™-coated dish
- Cultivate cells for ≥24h in StemFit® 100 ng/ml bFGF and 10 µM Y-27632 before removing the ROCK inhibitor

1.5.8. Freezing of iPSCs

iPSCs were frozen at 70-85% confluency. The standard splitting protocol was used to generate the cell suspension. After centrifugation, the supernatant was removed, and the cell pellet was resuspended in 1 ml of Bambanker™ (single cell suspensions) or mFreSR™ (colony fragment suspensions). The freezing suspension was transferred to a 1.5 ml cryogenic vial and gradually frozen (-1 °C/min) in a Nalgene freezing container at -80°C. 24-48h after freezing, the vial was transferred to liquid nitrogen tanks.

1.5.9. Thawing of iPSCs

For thawing of iPSCs, cryogenic vials were taken from the liquid nitrogen tank and thawed in a 37°C water bath until only small ice crystals remained. The cell suspension was transferred to a 50 ml conical Falcon™. 5 ml of DMEM/F12(1:1) supplemented with 100 ng/ml bFGF and 10 µM Y-27632 was added dropwise to the suspension while gently shaking the tube. After centrifugation at 200g for 5 min, the supernatant was removed, and the cell pellet was resuspended in in culturing medium and 10 µM Y-27632. For cultivation in Essential 8™, all cells were seeded on a Geltrex™-coated dish. For cultivation in StemFit®, cells were counted and three times the number of cells that are used for routine splitting (3,000-18,000 cells/cm^2) were seeded on a Geltrex™-coated dish.

1.1.1. Spontaneous differentiation of iPSCs

Spontaneous differentiation of iPSCs into cells of all three germ layers was induced by embryoid body (EB) formation, as previously reported.[56] In brief: Cells were washed twice with pre-warmed DPBS and incubated with 0.6 ml/plate Dispase (1 U/ml) for 5 min at 37°C. Upon removal of the enzyme, cells ware washed with 3 ml/plate DMEM/F12 (1:1). Cells were detached from the plate in 2 ml DMEM/F12 (1:1) with cell scrapers, transferred to a Falcon™ tube and centrifuged at 200 g for 5 min. The colony fragments were resuspended using 5 ml Stripette™ tips in Essential 8™ containing 10 µM Y-27632 and transferred to a Poly-HEMA-coated dish to prevent adherence (day 0 of differentiation). Floating EBs were incubated for 72h under standard conditions. On day 3 of differentiation, medium was changed to EB-20 medium containing 50 µg/ml L-ascorbic acid. For this, EBs were pelleted by gravity in a Falcon™ tube and washed in 2 ml EB-20 medium. The washing medium was removed after cell clumps had sunk to the bottom and replaced by fresh EB-20 for continued cultivation in the Poly-HEMA-coated dishes. The medium was changed on day 5. On day 7, floating embryoid bodies were plated on gelatin-coated 4-well plates at densities 15-25 EBs per well in EB-20 supplemented with 50 µg/ml L-ascorbic acid (for coating purposes, plates were incubated with 0.5 ml/well 0.1 % gelatin in DPBS for ≥30 min). From day 12 onwards, differentiations were screened for the appearance of spontaneously beating cardiomyocyte clusters that were manually excised using fine cannulas. For RNA extraction and qPCR analysis of germ

layer markers during the (re-)characterization of iPSC lines, cells were harvested on day 21 in TRIzol™ Reagent (See Master Thesis Viktoria Frajs, *SHOX2 in atrial fibrillation disease modelling using induced pluripotent stem cells*)

1.5.10. Transfection of iPSCs

iPSCs were transfected with plasmid DNA, gRNA, RNPs and ssODNs with the 4D-Nucleofector™ System according to Lonza's official guidelines, as described in the *4D-Nucleofector™ System Manual* (CD-MN025 04/16, ©Lonza 2016) and the *Genome Editing using Nucleofector™ Technology Technical Reference Guide* (CD-DS021 02/15, ©Lonza 2015). In detail, the following protocol was used:

GFP Control	Cas9 (plasmids)	Cas9 (RNP/gRNA)
500 ng of pmaxGFP® (provided in the kit)	500 ng Cas9/gRNA vector + 50-200 pmol ssODN	1 µg of Platinum™ Cas9 RNP + 250 ng synthesized gRNA + 50-200 pmol ssODN

- Treat cells ≥1h before transfection with 10 µM Y-27632
- Wash 2x with pre-warmed DPBS
- Add pre-warmed TrypLE™ Express
- Incubate at 37°C, 5% CO_2 for 5-10 min
- Resuspend cells carefully in TrypLE™ Express (check complete dissociation to single cells under microscope) and transfer to 50 ml Falcon™
- Dilute cell suspension 1:10 with pre-warmed DMEM/F12(1:1)
- Centrifuge at 200 g for 5 min in swing-bucket rotor centrifuge
- In case of RNP/RNA transfection: Meanwhile prepare Cas9 RNP/gRNA complexes by mixing 1 µg of GeneArt™ Platinum™ Cas9 Protein with 250 ng of synthesized gRNA in 5 µl of P3 transfection buffer in 1.5 ml Eppendorf tubes (complexes are formed after 5 min and stable for 30 min). For the transfection control condition prepare 500ng of pmaxGFP® vector in 5 µl of P3 transfection buffer in a 1.5 ml Eppendorf tube.
- Remove supernatant and resuspend cells in pre-warmed StemFit supplemented with 100 ng/ml bFGF and 10 µM Y-27632
- Count cells in Neubauer counting chamber and transfer 2.0-5.0*10^5 cells per condition to 15 ml Falcon™
- Centrifuge at 200 g for 5 min in swing-bucket rotor centrifuge
- Remove supernatant completely and resuspend cells in 15 µl P3 transfection buffer per condition

- Mix 15 µl of cell suspension with prepared Cas9 RNP/gRNA complexes or plasmid DNA and immediately transfer to one well of a 16-stripe (avoid air bubbles!)
- Transfer cells into the 4D-Nucleofector™ X Unit and start program DN-100
- Add StemFit® supplemented with 100 ng/ml bFGF and 10 µM Y-27632 and transfer whole cell suspension into one well of a 24-well plate coated with Geltrex™ (1:100)
- Check for transfection rate 24-48h after transfection

1.5.11. Sib-selection of transfected iPSCs

Transfected iPSCs were seeded in small pools of 200 cells per well (sib-selection) on 96-well plates 48h after transfection with Cas9 RNP/gRNA complexes, as previously described[102]. The standard splitting protocol was used to generate a cell suspension of 4000 cells/ml in bFGF containing 100 ng/ml bFGF and 10 µM Y-27632. 50 µl cell suspension per well (200 cells/well) were plated with a multichannel pipet on a Geltrex™-coated 96-well plate already containing 50 µl culturing media. The first medium change without Y-27632 was carried out after 2 days. Subsequently, the cells were fed every 2-3 days until they reached the desired confluency.

1.5.12. Freezing/DNA extraction of sib-selections

After reaching 70-90% confluency (~7-9 days after seeding), sib-selections were further processed. Multichannel pipets were used for the pipetting steps. Cells were washed 2x with 100 µl/well DPBS and treated with 30 µl/well TrypLE™ Express. After incubation for 5-10 min at 37°C and 5% CO_2, 15 µl of TrypLE™ cell suspension was mixed with 35µl of DPBS and then used for DNA extraction with the *Quick*-DNA™ 96 Plus Kit (see chapter 1.3.10), while the remaining half of the cell suspension was mixed with 85 µl Bambanker™ in the original plate and frozen at -80°C in a Styrofoam container.

1.5.13. Thawing of sib-selections

Upon identification of sib-selections with a high abundance of WT alleles via dPCR, frozen cells in 96-well plates were thawed in the incubator at 37°C, 5% CO_2 for 10-15 min. The 100 µl cell suspensions were transferred to a 15 ml conical Falcon™ and diluted with 1 ml of DMEM/F12(1:1) containing 10 µM Y-27632. After centrifugation (200 g, 5 min), the cell pellets were resuspended in 100 µl StemFit supplemented with 100 ng/ml bFGF and 10 µM Y-27632 and completely transferred to a Geltrex™-coated 96-well plate.

1.5.14. Single cell-cloning of iPSCs

iPSC sib-selections, in which a subpopulation of isogenic WT cells (homozygous for *SHOX2* c.*28T or *SHOX2* c.849C)) was identified via NGS and dPCR, were used for single cell-cloning.

The protocol was modified from the manufacturer's *StemFit® Technical tips* manual. The method is based on limited dilution. In detail:

- Thaw 200 µl Geltrex™ per 96-well plate to be seeded on ice
- Wash 2x with pre-warmed DPBS
- Add pre-warmed TrypLE™ Express
- Incubate at 37°C, 5% CO_2 for 5-10 min
- Resuspend cells carefully in TrypLE™ Express (check complete dissociation to single cells under microscope) and transfer to 50 ml Falcon™
- Dilute cell suspension 1:10 with pre-warmed DMEM/F12(1:1)
- Centrifuge at 200 g for 5 min in swing-bucket rotor centrifuge
- Remove supernatant and resuspend cells in pre-warmed StemFit® supplemented with 100 ng/ml bFGF and 10 µM Y-27632
- Prepare 10 ml of a 10 cells/ml cell suspension by serial dilution with StemFit® supplemented with 100 ng/ml bFGF and 10 µM Y-27632
- Pipet 200 µl ice-cold Geltrex™ to cell suspension and immediately transfer whole mixture into pipetting reservoir (Geltrex™ polymerizes within 5 min at ≥15°C)
- Plate 100 µl (= 1 cell/well) into each well of an uncoated 96-well plate
- Change medium after 72h for the first time and subsequently every other day until cells are ready for passaging
- Split cells 1:1 into a new Geltrex™-coated 96-well after 7-8 days
- After reaching 70-85% confluency, 50% of the cells were split onto a Geltrex™-coated 24-well plate while the other 50% was subjected to directPCR lysis and subsequent genotyping (see chapter 1.3.15)

1.5.15. Karyotyping

For karyotyping of the patient-derived iPSC lines, cells were arrested in metaphase with N-desacetyl-N-methylocolchicine (KaryoMAX™ Colcemid™ solution). The metaphase preparation was provided to Brigitte Schoell (Department of Human Genetics, Institute of Human Genetics, University of Heidelberg, 69120 Heidelberg, Baden-Wuerttemberg, Germany) for multiplex in situ hybridization (M-FISH) and imaging. Ten metaphases were analyzed, and representative pictures were prepared by Prof. Anna Jauch (Department of Human Genetics, Institute of Human Genetics, University of Heidelberg, 69120 Heidelberg, Baden-Wuerttemberg, Germany). The protocol for chromosomal preparation in detail:

- Culture iPSCs in 3-4 wells of a 6-well plate for each metaphase preparation
- Change culturing medium 24h before starting Colcemid™ treatment

- Treat cells with 0.08 µg/ml Colcemid™ (1:125 dilution of 10 µg/ml stock) for 17h
- Remove medium and wash cells 1x in prewarmed DPBS (collect DPBS afterwards in 15 ml Falcon™)
- Add 1 ml of Accumax™ per well and incubate cells at 37°C for 5 min
- Collect Accumax™ solution in the 15 ml Falcon™
- Wash wells with 2 ml DPBS each and add to the 15 ml Falcon™
- Centrifuge cell suspension at 300 g for 10 min, RT
- remove supernatant, leaving about 0.5 ml above the cell pellet, re-suspend cell pellet by tapping the tube
- Add 10 ml of hypotonic solution (1:1 mixture of 0.55% KCL and 1% Na Citrate) slowly while vortexing carefully and mix by inverting the tube
- Incubate at 37°C for 20 min
- Add 2 ml of fixative (methanol / glacial acetic acid, 3:1 v:v, precooled at -20°C) to hypotonic solution
- Centrifuge at 300 g for 10 min, RT
- Remove supernatant down to 3-5 ml, and resuspend cell pellet
- Add 10 ml of fixative slowly while vortexing carefully and mix by inverting the tube
- Centrifuge at 300 g for 10 min, RT
- Repeat fixation procedure 3 times using the same volume fixative
- Transfer suspension to a new 15 ml Falcon™ tube
- Store cells in fixative at -20°C

For karyotyping by G banding, chromosomes were obtained according to routine procedures and based on previously published protocols by Karin Hüllen and Alexandra Köppel (Department of Human Genetics, Institute of Human Genetics, University of Heidelberg, 69120 Heidelberg, Baden-Wuerttemberg, Germany). Karyograms were made of trypsin giemsa-stained metaphases and analyzed by Prof. Hans Janssen (Department of Human Genetics, Institute of Human Genetics, University of Heidelberg, 69120 Heidelberg, Baden-Wuerttemberg, Germany).

1.5.16. Immunofluorescence staining

For the expression analysis of pluripotency markers during iPSC line (re-)characterization, cells were grown on Geltrex™-coated coverslips. Upon colony forming, they were fixed with 4% PFA for 15 min at RT and washed 3 times shortly in DPBS afterwards.

For the initial characterization, cells were subjected to simultaneous permeabilization and blocking with 10% normal goat serum in DPBS/0.1% Triton-X-100 for 1 h at 37°C. Cells were stained with

primary antibodies for NANOG, TRA1-81, OCT4 and SOX2 in PBS/0.1% Triton-X-100 containing 1% goat serum overnight at 4°C. After 5 washes in DPBS (3x 3 min, 2x shortly), incubation with AlexaFluor488- and AlexaFluor-594-conjugated secondary antibodies (Thermo Fisher Scientific) specific to the appropriate species was done in PBS/0.1% Triton-X-100 containing 1% goat serum for 1 h at 37°C. Nuclei were detected with 1 µg/ml Hoechst 33258.

For the re-characterization of isogenic clones, cells were permeabilized for 10 min at 4°C with 0.1% Triton X-100 in DPBS. Subsequently, they were blocked in 1% BSA in DPBS for 1 h at RT and incubated with primary antibodies in blocking solution for 1 h at RT. After 3 washes in DPBS for 5 min each, cells were incubated with respective secondary antibodies in blocking solutions for 1 h at RT. Nuclei were counterstained with Hoechst 33342 (1:5000).

The Immunofluorescence staining for the initial characterization was done in collaboration with AG Moretti by Dr. Svenja Laue, Dr. Tatjana Dorn and Birgit Campbell (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany). The immunofluorescence staining for the re-characterization was done by Viktoria Frajs under supervision (See Master Thesis Viktoria Frajs, *SHOX2 in atrial fibrillation disease modelling using induced pluripotent stem cells*). Images were taken on a Leica DMI4000 B fluorescence microscope with a Leica DFC3000 G camera system using 10x, 20x and 40x objectives.

1.5.17. Alkaline Phosphatase (ALP) staining

The newly generated iPSC lines were stained for alkaline phosphatase activity using the NBT/BCIP alkaline phosphatase blue substrate according to manufacturer's instructions in collaboration with AG Moretti by Dr. Svenja Laue, Dr. Tatjana Dorn and Birgit Campbell (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany).

For the re-characterization of the isogenic lines, iPSCs were cultured in StemFit® until reaching about 40% confluency. Cells were washed once in DPBS and fixed with 4% PFA for 20 min at RT followed by 3 washes with DPBS for 10 min each. Fixed colonies were stained with 3 ml/3.5 cm plate of ALP staining solution for 20 min at room temperature and washed twice in DPBS before imaging on a stereomicroscope. The ALP staining was done by Viktoria Frajs under supervision (See Master Thesis Viktoria Frajs, *SHOX2 in atrial fibrillation disease modelling using induced pluripotent stem cells*).

1.6. In silico methods, bioinformatics, and ethical statements

1.6.1. dPCR template generation and analysis

All dPCR experiments were run using *QuantaSoft* (version 1.7.4.0917). Templates were generated with the following settings: ABS (experiment), ddPCR™ Supermix for probes (supermix), Mut (Ch1, FAM) and WT (Ch2, HEX). For analysis, automatically set thresholds were used to define positive/negative droplets whenever applicable or set manually with the lasso function.

1.6.2. gRNA design

gRNA design was performed with the *CCTop - CRISPR/Cas9 target online predictor*.[118] As input, the genomic DNA sequence ~200bp up- and downstream of the mutation of interest (*SHOX2 c.849C>A* or *SHOX2 c.*28T>C*) was used. The PAM type was set to NGG (*Streptococcus pyogenes*). For target selection, the target site length was set to 20 nt and the target site 5' or 3' limitations were set to NN (standard settings). For off-target prediction, the maximum of total mismatches was set to 4 nt, the core length was set to 12 nt with a maximum of total mismatches of 2 (standard settings). As a reference genome, the human genome (assembly homo sapiens GRCh37/hg19) was used. The three gRNAs closest to the mutation were selected. The chosen gRNAs were re-evaluated with the CHOPCHOP web tool with regard to efficiency and off-targets.[119]

1.6.3. Off-target analysis

The top 20 off-target sites predicted by *CCTop* were further evaluated. Exonic off-target sites were automatically included into downstream analysis. For intronic and intergenic off-target sites, the target sequence coordinates were analyzed in the *UCSC Genome Browser*. The inclusion criteria for downstream analysis were: Conservation among species, DNase clustering, expressed sequence tags and active chromatin marks. If a combination of these criteria indicated a potential regulatory relevance of this DNA segment, the off-target site was sequenced.

1.6.4. ssODN design

ssODNs were designed as previously described[109]: The template spanned the Cas9 cut site asymmetrically, reaching from -82 nt to +35 nt for gRNAs targeting the *SHOX2* c.849C>A locus and -81 nt to +45 nt for gRNAs targeting the *SHOX2 c.*28T>C* locus. The DNA oligo was synthesized and purified via desaltation by Integrated DNA Technologies (IDT).

1.6.5. Analysis of iPSC sib-selections via NGS

The raw files generated after NGS (Gzipped .fastq files read 1 and read 2) were uploaded to the CRISPR RGEN Tool Cas-Analyzer[120] for the detection of potential isogenic subpopulations within the sib-selections. The following settings were used for the analysis:

Full reference sequence (5' to 3')	The full sequence of the PCR product without NGS-specific tags
Nuclease Type	Single nuclease
Selected Nuclease	SpCas9 from Streptococcus pyogenes: 5'-NGG-3'
Target DNA sequence	The 20 nt of DNA targeted by the respective gRNA
Donor DNA sequence for HDR	[no donor sequence]
Comparison range (R)	100 nt
Minimum frequency (n)	Variable, 0.00025%-0.25% of the total reads with indicator sequences
WT marker (r)	[no WT marker]

The minimum frequency of alleles to be considered in the analysis was calculated from the total amount of reads that contained both indicator sequences (the first and last 12 nt of the comparison range). The threshold was set to $0.25*10^{-3}$ % of total reads for cell pools or 0.25% of total reads for sib-selections (minimum allele frequency in $200*10^3$ or 200 cells, respectively). The frequency of WT alleles was calculated as a range (from minimum to maximum) with the following formulas:

$$minimum\ \%\ of\ WT\ alleles = \frac{\#[WT\ alleles, unedited] + \#[WT\ alleles, edited]}{\#[total\ reads]} * 100$$

$$maximum\ \%\ of\ WT\ alleles = \frac{\#[WT\ alleles, unedited] + \#[WT\ alleles, edited] + \#[alleles, ?]}{\#[total\ reads]} * 100$$

Definitions:

#[*WT alleles, unedited*]	Sum of all alleles with a detectable WT base with no additional mutations within the comparison range
#[*WT alleles, edited*]	Sum of all WT alleles with a detectable WT base with additional mutations within the comparison range (point mutations, insertions, or deletions)
#[*alleles,?*]	Sum of all alleles without a detectable *SHOX2* c.849C>A or *SHOX2* c.*28T>C base due to a base-spanning deletion
#[*total reads*]	Number of reads containing left and right indicator sequences

The percentage of isogenic cells in the sib-selection was calculated from these results with the following formulas:

$$minimum\ \%\ of\ isogenic\ cells = ([minimum\ \%\ of\ WT\ alleles] - 50\%) * 2$$

$$maximum\ \%\ of\ isogenic\ cells = ([maximum\ \%\ of\ WT\ alleles] - 50\%) * 2$$

The average of the minimum and maximum percentage of isogenic cells was then used to calculate how many single cell clones would need to be screened to find at least one isogenic clone with a self-chosen probability *P*.

$$n = \frac{\log(1 - P)}{\log q}$$

Definitions:

n	Number of cells that need to be screened to find at least one isogenic clone with the probability P
P	Probability to find at least one isogenic clone. Value can be chosen arbitrarily (e.g. 95%)
q	Probability that analyzed clone is not isogenic. Value is calculated by subtracting the percentage of isogenic cells in the sib-selection from 1

From the calculated number of cells *x*, the amount of 96-well plates that needed to be used for single cell-cloning was determined:

$$\#[96\ well\ plates] = \frac{x}{clonability[\%] * 96}$$

The clonability was defined as the percentage of wells on a 96-well plate from which an iPSC clone could be derived (survived the first split and did not differentiate). This was empirically determined in pilot experiments and found to be 10-20%.

1.6.6. Illustrations

Figures were prepared using Adobe Illustrator and Photoshop 2020. DNA schemes were drawn using Illustrator of Biological Sequences.[121] Some figure elements were taken from Pixabay (www.pixabay.com) and modified (citation or permission required).

1.7. Patient information and study design

1.7.1. Ethics statement for the genetic analysis of AF and SND patients

All patients recruited for this study were German individuals. The AF patient cohort (n = 450) consisted of participants from the Gutenberg Health Study [122] enrolled in the period from 2007 to 2012 at the University Medical Center of the Johannes Gutenberg University Mainz, Germany. The SND patient cohort (n = 98) was recruited from the Department of Medicine I of the Ludwig Maximilians University Hospital Grosshadern, Munich, from 2013 to 2014. For at least 95% of all cases, clinical parameters were available. The study was approved by the ethical commission of the Medical Faculty, University of Heidelberg, Heidelberg, Germany (S-104/2010 "Molekulare Grundlagen SHOX2-bedingter Herzfehlbildungen" 17.03.2010) and was performed in accordance

with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. A written informed consent was given by every participant prior to the inclusion in the study, including the consent to use their DNA for genetic analyses.[1]

1.7.2. Ethics Statement for the generation of iPSCs

The study was approved by the ethical commission of the Medical Faculty, Technical University of Munich, Munich, Germany ("Induzierbare pluripotente Stammzellen als innovatives Patienten-basiertes *in vitro* Modell für Vorhofflimmern" als Teilprojekt 5 des Hauptprojekts "Erzeugung und Charakterisierung patientenspezifischer induzierter, pluripotenter Stammzellen" 2109/08) and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Every participant gave written informed consent including the consent to use their blood samples prior to the inclusion in the study.

1.8. Statistics

Statistics were performed using GraphPad Prism version 8 for Windows (Prism 8 for Windows, Software MacKiev, LLC, USA).

1.8.1. Analysis of heart rate and luciferase activity

Differences in zebrafish embryonic heart rates were tested for their significance with a one-way ANOVA combined with Tukey's honestly significant difference *post hoc* test for multiple comparisons. Differences in luciferase activities were tested by a one-way ANOVA followed by uncorrected Fisher's least significant difference test.[1]

1.8.2. Expression analysis

Differences in expression were tested by a ratio t-test.[1]

2. Results

2.1. Functional characterization of rare variants in the *SHOX2* gene identified in SND and AF

Mutations in *SHOX2* have been found in patients with early-onset and familiar forms of AF.[28,50] To identify novel variants in *SHOX2* and elucidate the causal relationship between them and the development of AF and SND, a candidate gene study was combined with functional analyses. The AF cohort included 450 participants from the Gutenberg Health Study enrolled in the period from 2007 to 2012 at the University Medical Center of the Johannes Gutenberg University Mainz, Germany.[122] The SND cohort comprised 98 individuals recruited from the Department of Medicine I of the Ludwig Maximilians University Hospital Grosshadern, Munich between 2013 and 2014. Patient characteristics and clinical parameters were available for more than 95% of the candidates (See **Table S1-4** in Appendix). Recruitment was done in collaboration with Dr. Sebastian Clauss, Dr. Tina Klier, Prof. Stefan Kääb (Department of Medicine I, Klinikum Grosshadern, University of Munich (LMU), 81675 Munich, Bavaria, Germany), Tanja Zeller and Renate B. Schnabel (Department of General and Interventional Cardiology, University Heart Center Hamburg (UHZ), University Hospital Hamburg/Eppendorf, Hamburg, Germany).

2.1.1. Identification and evaluation of *SHOX2* variants in AF and SND cohorts

The six coding exons of the longest *SHOX2* isoform (NM_003030.4), including the primate specific exon 2+, were screened for variants via Sanger Sequencing in a joint effort with Birgit Weiss and Sandra Hoffmann (Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany). Values and allele frequencies are reported from the given databases as of March 2020. In total, four heterozygous synonymous and non-synonymous variants (c.230G>A/p.G77D, c.387G>A/p.L129=, c.388C>T/p.L130F, and c.879C>T/p.A293=) were identified in patients suffering from AF (4/450) and one heterozygous missense mutation (c.98C>G/p.P33R) was identified in an individual from the SND cohort (1/98). The SND patient carrying this variant also presented with AF. *In silico* prediction based on "combined annotation dependent depletion" (C-scores or CADD) [123,124] was used to predict the pathogenic potential of these variants and to select all variants with C-scores ≥20 for functional analyses. This included p.G77D (C-score = 22.2) and p.A293= (C-score = 20.6) from the AF study cohort and the SND variant p.P33R (C-score = 25.8). The *in silico* analysis was expanded using a set of publicly accessible prediction tools that are based on deep-neural network classification or machine learning approaches (DANN, FATHM-MKL, GWAVA), empirical scoring systems

(FunSeq2), or a tool that combines these classifications (PredictSNP2). Except for p.G77D, the all variants with a CADD score ≥20 were predicted to be deleterious by a majority of the tools, while the rest was mainly classified as neutral.

None of the variants were present in the European Non-Finnish population of the 1,000 genomes project.[125] From the AF variants, p.A293= was reported with an allele frequency of 0.0004578% (58/126696) in the European Non-Finnish population of genome Aggregation Database (gnomAD), while p.G77D was not present. For p.P33R, gnomAD reported an allele frequency of 0.000009694% (1/103152) (**Table 1**). In addition, the recently released DZHKomics database which comprises approximately 1,150 genomes from unrelated healthy individuals of 6 German population cohorts (GHS, Gutenberg-Gesundheitsstudie; HCHS, Hamburg City Health Study; NOKO, Heidelberg Normal Kontrollen; IKMB, Institut für Klinische Molekularbiologie Kiel; KORA, Kooperative Gesundheitsforschung in der Region Augsburg; SHIP, Study of Health in Pomerania) were included in the analysis (unpublished data, https://ihg4.helmholtz-muenchen.de/cgi-bin/DZHKomics/search.pl). Only the AF variant c.387G>A was present in this cohort with an allele frequency of 0.00433526 (9/1150).

In summary, this data expands the list of *SHOX2* variants associated with AF, confirms previous findings [28,50] and indicates a novel association of *SHOX2* with SND.

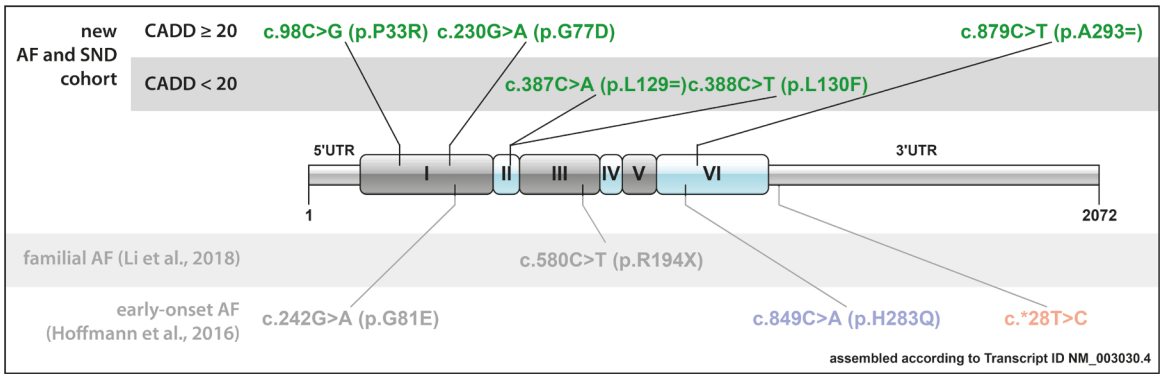


Figure 5 Identified *SHOX2* variants in patients with sinus node dysfunction (SND) and atrial fibrillation (AF). Schematic drawing showing all identified variants within the *SHOX2* gene in patients with AF and SND. The novel identified variants and the respective cohorts are highlighted in green. The previously published variants are greyed out. Figure and legend based on Hoffmann S, [...], Sumer SA et al. 2019

Patient cohort	AF (n = 450)				SND (n = 98)
Genomic position (GRCh37)	3:157823584 C>T	3:157822878 C>T	3:157822877 G>A	3:157816005 G>A	3:157823716 G>C
Reference number	N/A	rs753047735	N/A	rs150931445	rs768138092
Transcript consequence	c.230G>A	c.387G>A	c.388C>T	c.879C>T	c.98C>G
Protein consequence	p.G77D	p.L129=	p.L130F	p.A293=	p.P33R
Patient cohort allele frequency	0.0011111 (1/450)	0.0033333 (3/450)	0.0011111 (1/450)	0.0011111 (1/450)	0.0051020 (1/98)
TGP allele frequency	-	-	-	-	-
gnomAD v2.1.1 (non-TOPMed) allele frequency	-	0.0003456 (25/72338)	0.00001750 (1/57154)	0.0004578 (58/126696)	0.000009694 (1/103152)
DZHKomics allele frequency	-	0.00433526 (9/1150)	-	-	-
CADD Score	22.2	10.68	9.02	20.6	25.8
PredictSNP2	Neutral	Neutral	Neutral	Deleterious	Deleterious
DANN	Neutral	Neutral	Neutral	Deleterious	Deleterious
FATHMM	Neutral	Neutral	Neutral	Deleterious	Deleterious
FunSeq2	Deleterious	Neutral	Deleterious	Neutral	Deleterious
GWAVA	Deleterious	Neutral	Deleterious	Neutral	N/A

Table 1. Overview of SHOX2 variants identified in AF and SND patients and control databases: 1000 gnomes project (TGP) and the genome Aggregation Database (gnomAD); only the European Non-Finnish populations were considered. Novel variants were selected for further functional studies based on the predicted pathological potential via Combined Annotation Dependent Depletion (CADD). Table and legend modified from Hoffmann, […], Sumer et al., 2019.[1]

2.1.2. Functional characterization of SHOX2 p.P33R, p.PG77D and p.A293= *in vivo* and *in vitro*

The functional relevance of the three selected variants was tested by investigating their pathogenic potential in the zebrafish as a model system in collaboration with Christoph Paone, Sabrina Diebold and Steffen Just (Department of Internal Medicine II, University of Ulm, Ulm, Germany). As shown before, a morpholino-mediated knockdown of endogenous Shox2 in zebrafish embryos leads to pericardial edema and severe bradycardia.[22,48] The phenotype can be rescued by the ectopic expression of (human) Shox2. The missense mutation p.H283Q, identified in a previous study, affects a highly conserved amino acid, also present in the zebrafish genome (corresponding mutation Shox2 p.H277Q). Upon expression in Shox2-depleted embryo hearts, the mutated version failed to restore a normal heart rate and was therefore demonstrated to affect the pacemaker function of the gene.[28] Here, the SHOX2 p.H283Q was used as a positive control in a modified experiment, where this variant or the newly identified variants were overexpressed in a cardiac-specific manner. A dominant-negative effect could be observed upon overexpression of SHOX2 p.H283Q that

resulted in pericardial edema and significantly reduced heart rates, similar to what was previously found for the corresponding zebrafish variant (**Figure 5A**).[28] Similarly, the missense variant SHOX2 p.G77D, identified in the AF cohort, also revealed pericardial edema upon overexpression (**Figure 5A**). In addition, a bradycardia phenotype (significantly reduced heart rate) could be observed for p.G77D, while the synonymous AF variant p.A293= and the SND variant p.P33R showed no difference (**Figure 5B**). To investigate if the identified missense variants impede the transactivation activity of SHOX2, *in vitro* dual luciferase reporter assays were performed by Sandra Hoffmann (Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany). The reporter gene under the control of the *BMP4* promotor, a direct SHOX2 target, can be activated by the WT protein, while p.H283Q severely affects this ability.[19,28] In the current study, both missense variants p.P33R and p.G77D were found to be unable to activate the *BMP4* promotor when compared with WT SHOX2 (**Figure 5C**). These results were in accordance with the *in vivo* phenotypic changes for P.G77D and further indicated a functional consequence of this AF-associated mutation. However, as this was the first time that an observable phenotype for the p.P33R SND mutation could be shown, the target gene expression was tested *in vivo*. WT and p.P33R SHOX2 were overexpressed in the hearts of zebrafish embryos. Cardiac RNA was isolated 72 h post fertilization and subjected to comparative gene expression analysis in collaboration with Ralph Roeth (Department of Human Molecular Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany). *Bmp4* RNA levels were significantly downregulated upon overexpression of the p.P33R mutant (**Figure 5D**), confirming the impaired Bmp4 reporter activation that had been observed in the luciferase assay.

Thus, the functional investigations in zebrafish and the in vitro studies demonstrated a pathological potential for two of the three novel identified SHOX2 variants.[1]

In vitro analysis and assays as well as the use of model organisms are valuable tools to investigate the pathogenic influence of mutations on gene function and expression. A lot of the current knowledge on the molecular pathways in the heart relies on heterologous expression systems, such as mouse and zebrafish. However, distinct species-specific electrophysiological and transcriptional properties do exist, resulting in considerable functional differences. This includes features like a different resting heart rate or the sheer physiological size of the heart leading to different signal transduction times. On the other hand, access to primary human tissue, especially from disease-bearing patients, is limited in quantity and in time due to the difficulties to propagate these cells in culture.[51] These limitations can be overcome by iPSC-based disease models that offer an unlimited source of patient-specific cardiomyocytes with putative disease-causing mutations. Therefore, two patients harboring *SHOX2* variants with a functional impact on the gene's function were recruited.

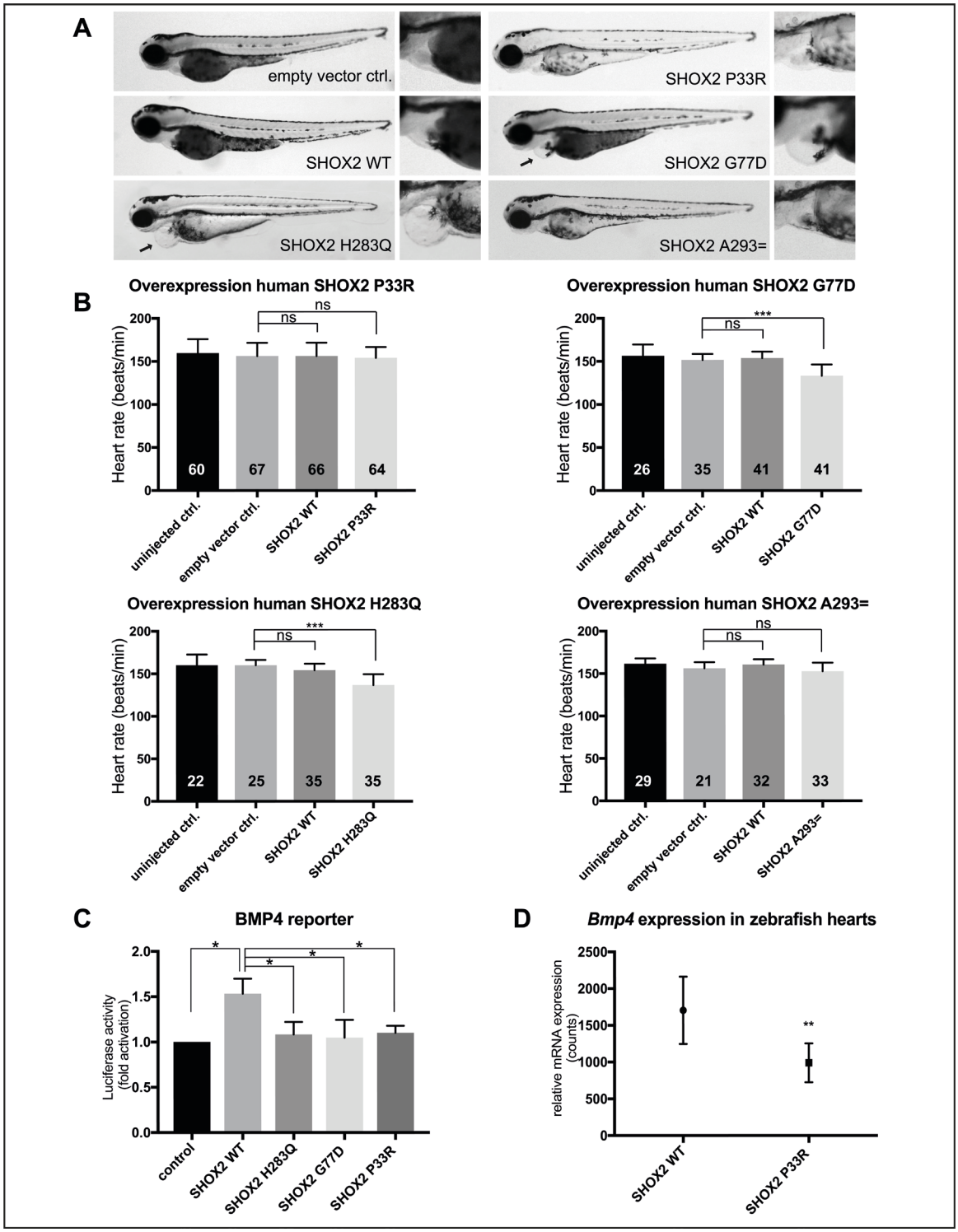


Figure 6 Functional characterization of SHOX2 variants *in vivo* and *in vitro*. (A) Cardiac-specific overexpression of SHOX2 mutants compared with SHOX2 WT (wild-type) leads to pericardial edema (arrow) for p.G77D and p.H283Q but not for p.P33R and p.A293= in zebrafish 72hpf. (B) The heart rate of zebrafish embryos was significantly reduced upon overexpression of p.G77D and p.H283Q but normal for p.P33R and p.A293= 72hpf compared with empty vector-injected embryos. SHOX2 WT overexpression showed no effect. Data are expressed as mean ± SD of three to six independent experiments. All P-values were determined by one-way ANOVA with Tukey's multiple comparison test (*P < 0.05; ***P < 0.001; ns: not significant). (C) Luciferase activity of BMP4 reporter construct co-expressed withSHOX2 wild-type (WT) or SHOX2 mutants (p.H283Q, p.G77D, p.P33R) in HEK293T cells 24 h after transfection. All values are normalized to the empty pGL3 basic vector co-transfected with the respective expression constructs. Data are expressed as mean ± SEM of four independent experiments performed in triplicate. All P-values were determined by one-way ANOVA followed by uncorrected Fisher's least significant difference test (*P < 0.05). (D) nCounter analysis revealed downregulation of Bmp4 expression in zebrafish hearts (72 hpf) overexpressing the mutant p.P33R compared with wild-type SHOX2 (n = 40 hearts per condition from two independent experiments). Statistical differences were determined by a ratio paired t-test (**P < 0.01). Figure and legend taken from Hofmann S, [...], Sumer SA et al. Front. Genet. 2019.{Hoffmann, 2019 #230}

2.2. Clinical analysis of patients recruited for iPSC reprogramming

Two AF patients harboring the previously described heterozygous *SHOX2* mutations (*SHOX2* c.849C>A and *SHOX2* c.*28T>C) [28] were recruited for the reprogramming of PBMCs to iPSCs by Prof. Stefan Kääb and Dr. med. Sebastian Clauss (Department of Medicine I, Klinikum Grosshadern, University of Munich (LMU), Munich, Germany). A blood sample was provided for the extraction of PBMCs. The cardiological conditions of the two patients were analyzed based on available information from clinical records.

2.2.1. Patient I (*SHOX2* c.849C>A, SHOX2 p.H283Q)

The male patient with the *SHOX2* c.849C>A mutation was born in 1949 and developed AF at the age of 56 years. When he was recruited for study enrollment in 2009 at the age of 62, his AF had become persistent. An echocardiography at that time revealed a slightly enlarged left atrium (LA diameter = 43 mm) and a reduced EF of 43 %. In addition, the patient was diagnosed with arterial hypertension. The patient's ECG showed a sinus rhythm with normal heartbeat rate and unremarkable P-, PR-, QRS- and QT-durations (**Figure 6A**). No other family member had developed AF (parents, one brother, one sister, one son) (**Figure 6A**). Except for hypertension in the patient's brother and mother no other cardiovascular disease was present within the family (including coronary artery disease, stroke, (dilated/hypertrophic/restrictive/arrhythmogenic) cardiomyopathy, other forms of arrhythmias, or sudden (cardiac) death) and no other family member had or has an implanted cardioverter defibrillator/pacemaker. The patient himself had no pacemaker or cardioverter/defibrillator implanted (Patient's characteristics are summarized in **Table 2**). In March 2016, at the patient's age of 66, blood samples were obtained for iPSC reprogramming purposes during a regular clinical follow-up visit.

2.2.2. Patient II (*SHOX2* c.*28T>C)

The male patient harboring the *SHOX2* c.*28T>C mutation was born in 1970 and developed AF at the age of 37 years. At the time of study enrollment in early 2011, his ECG showed a prolonged PR interval (first-degree atrioventricular block), a clinically unremarkable QRS duration (no bundle branch block) and normal QT durations (**Figure 6B**). He was further diagnosed with DCM (left ventricular end diastolic diameter of 67 mm in February 2011). For primary prevention purposes, a single chamber cardioverter-defibrillator had been implanted. The cardiovascular risk factors of this patient included: Diabetes mellitus type II, hypercholesterolemia, ex-smoker status. Additionally, the patient suffered from chronic kidney disease and Crohn's disease. No other family member developed AF (both parents, six brothers and sisters of the patient's father, six brothers and sisters of the patient, three children of the patient). The patient's father who had a known dilated

cardiomyopathy died from myocardial infarction at the age of 46 and one of the father's brothers suffers from coronary artery disease and received surgical revascularization (coronary artery bypass). Another brother of the patient's father died from cardiac arrest at the age of 59. Several relatives (both parents, one brother of the patient's father, two brothers and one sister of the patient) suffer from hypertension. Besides that, no other cardiovascular disease was present within the family (including stroke, (hypertrophic/restrictive/arrhythmogenic) cardiomyopathy, or other forms of arrhythmias) (**Figure 6B**) and no other family member had or has an implanted cardioverter defibrillator/pacemaker. In March 2011, the patient suffered from progressive congestive heart failure with an ejection fraction (EF) of 25% presumably due to the patient's DCM and the AF that progressed to a persistent type. Consequently, the patient received a heart transplant in October 2011. At that time, the patient was treated with Amiodarone (class III anti-arrhythmic drug) and Dobutamine (emergency medication for decompensated heart failure) besides the standard treatment with diuretics, while beta-blockers (class II anti-arrhythmic drug) and ACE inhibitors were stopped (Patient's characteristics are summarized in **Table 2**). In January 2015, at the patient's age of 45, blood samples were obtained for iPSC reprogramming purposes during a routine follow-up visit after heart transplantation.

		AF patient I (*SHOX2* c.849C>A)	**AF patient II (*SHOX2* c.*28T>C)**
	age at disease onset [years]	56	37
	age at recruitment [years]	66	41
	AF disease progression (at time of enrollment)	persistent	persistent
	heart transplantation	no	yes
Echocardiography (at time of enrollment)	**ejection fraction**	43%	25%
	structural	LA diameter = 43 mm	LVEED = 67 mm
	medication	Pantozol, Marcumar	Amiodarone, Dobutamine, β-blockers (paused), ACE inhibitors (paused)
	cardioverter-defibrillator [yes/no]	Yes	No
	additional cardiovascular diseases	arterial hypertension	dilated cardiomyopathy
	cardiovascular risk factors	N/A	Diabetes mellitus II, hypercholesterolemia ex-smoker
	secondary diagnosis	N/A	renal insufficiency, Morbus Crohn

Table 2: Patient characteristics

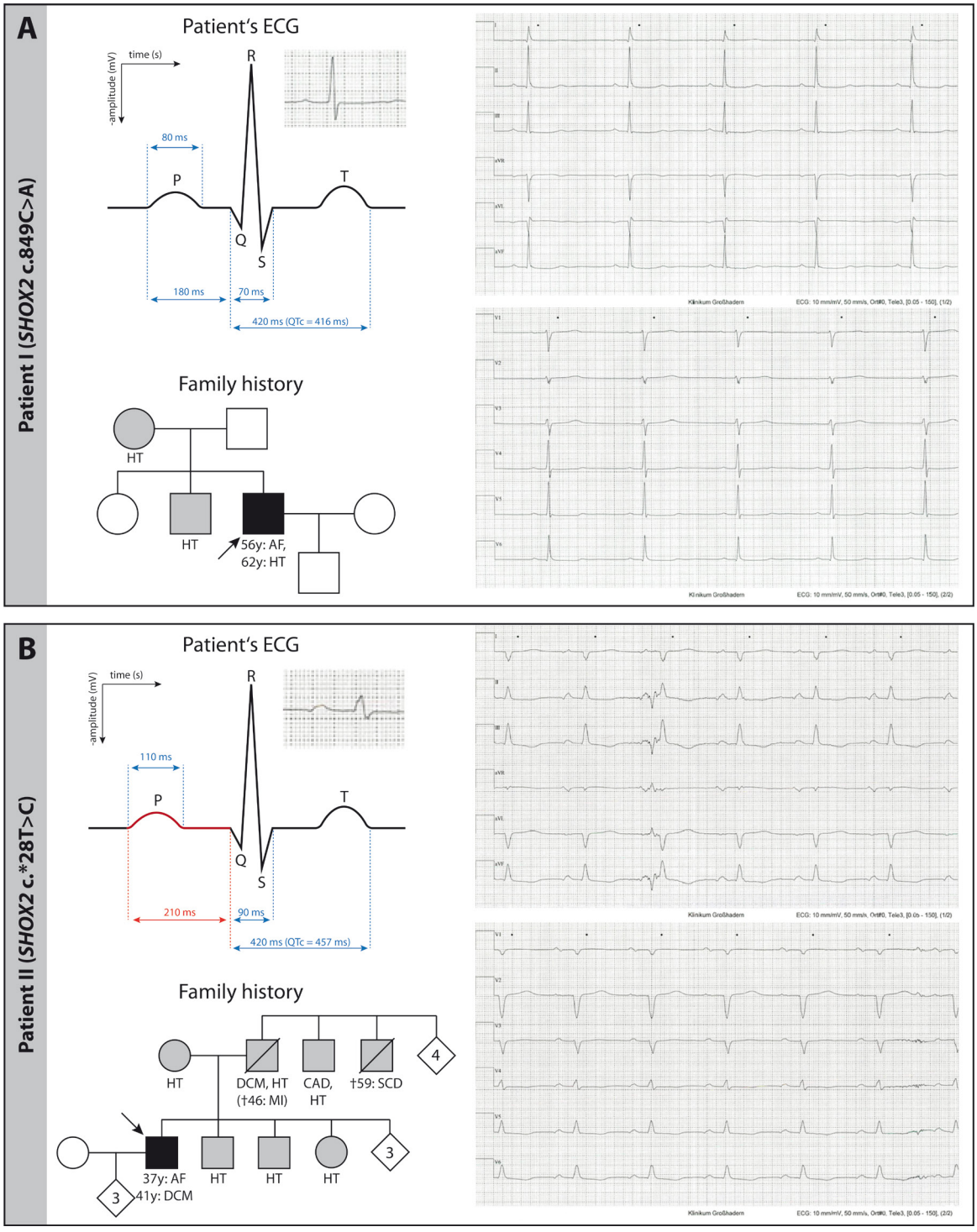


Figure 7 Patients' ECG and family pedigree with cardiovascular diseases. (A) Patient I (*SHOX2* c.849C>A) suffered from HT and AF but showed an unremarkable ECG. Except for HT in the mother and the patient's brother, no other cardiovascular disease* was present. (B) Patient II (*SHOX2* c.*28T>C) suffered from AF and DCM. His ECG revealed a slightly prolonged PR interval (first degree atrioventricular block). A substantial amount of cardiovascular diseases was present in family members, including HT in both parents, two of the father's brothers and three siblings. Like the patient, the father suffered from DCM and died from MI at the age of 46 years. One of the father's brothers suffered from CAD and another brother died from SCD, when he was 59 years old. Other cardiovascular diseases* were excluded. (Sumer et al., for revision only)

*cardiovascular diseases: Stroke, coronary artery disease, (dilated/hypertrophic/restrictive/arrhythmogenic) cardiomyopathy, other forms of arrhythmias, or sudden (cardiac) death

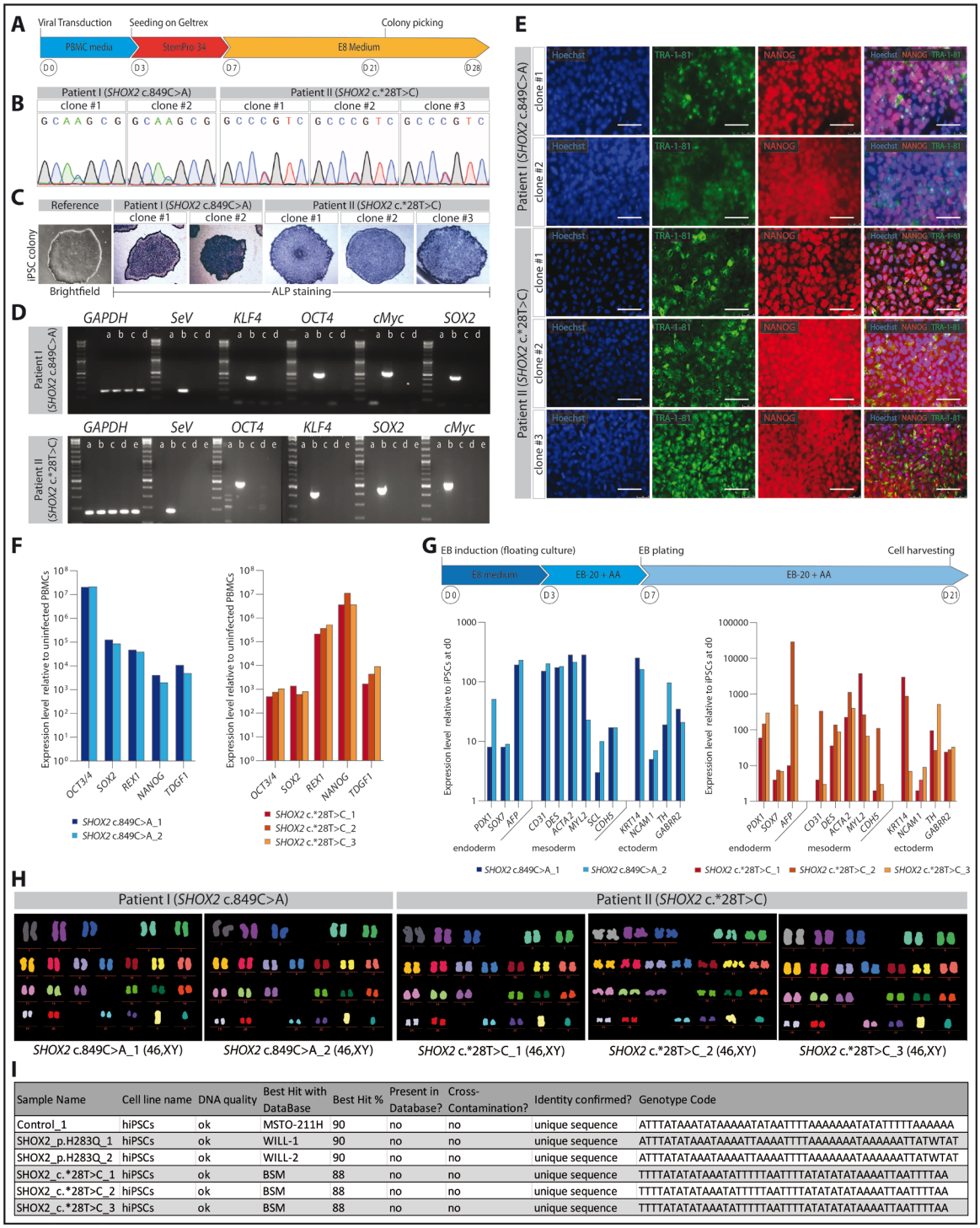
Abbreviations: ECG = electrocardiogram, HT = Hypertension, AF = Atrial fibrillation, DCM = Dilated cardiomyopathy, MI = Myocardial infarction, CAD = Coronary artery disease, SCD = sudden cardiac death

arrow = index patient, black = atrial fibrillation, grey = other cardiovascular diseases, white = no known cardiovascular diseases

2.3. Generation and characterization of *SHOX2* c.849C>A and *SHOX2* C.*28T>C iPSC lines

PBMCs were reprogrammed to iPSCs by Dr. Svenja Laue, Birgit Campbell and Dr. Tatjana Dorn (First Department of Medicine, Cardiology, Klinikum rechts der Isar – Technical University of Munich, 81675 Munich, Bavaria, Germany) with non-integrating Sendai viruses coding KLF4, OCT3/4, SOX2, c-MYC as previously described (**Figure 7A**).[56]

Two iPSC clones generated from Patient I (*SHOX2* c.849C>A) and three iPSC clones generated from patient II (*SHOX2* c.*28T>C) were selected for detailed characterization by me and in collaboration with AG Moretti and AG Jauch (Department of Human Genetics, Institute of Human Genetics, University of Heidelberg, 69120 Heidelberg, Baden-Wuerttemberg, Germany): The presence of the patient-specific heterozygous mutations was confirmed by Sanger Sequencing (**Figure 7B**). All clones displayed stem cell-like colony-forming properties and high alkaline phosphatase activity (**Figure 7C**). The loss of viral transgene expression after 10 to 20 passages was confirmed via RT-PCR (**Figure 7D**). All lines expressed the pluripotency markers NANOG and TRA1-81 on protein level, as detected by immunofluorescence staining (**Figure 7E**). In addition, qRT-PCR analysis of *OCT3/4*, *SOX2*, *REX1*, *NANOG* and *TDGF1* indicated that these iPSCs had reactivated their endogenous pluripotency genes (**Figure 7F**). Furthermore, the capacity to generate derivates of all three germ layers *in vitro* was demonstrated by spontaneous differentiation as EBs (**Figure 7G**). Metaphase analysis by M-FISH revealed no chromosomal aberrations and a normal karyotype for all 5 clones (**Figure 7H**). MCA testing confirmed a shared origin for all *SHOX2* c.849C>A and *SHOX2* c.*28T>C clones, respectively, which also differed from the control line that was cultured in parallel (**Figure 7I**).



I

Sample Name	Cell line name	DNA quality	Best Hit with DataBase	Best Hit %	Present in Database?	Cross-Contamination?	Identity confirmed?	Genotype Code
Control_1	hiPSCs	ok	MSTO-211H	90	no	no	unique sequence	ATTTATAAATATAAAAATATAATTTTAAAAAAATATATTTTTAAAAAA
SHOX2_p.H283Q_1	hiPSCs	ok	WILL-1	90	no	no	unique sequence	ATTTATATAAATAAAATTAAAATTTTAAAAAAATAAAAAATTATWTAT
SHOX2_p.H283Q_2	hiPSCs	ok	WILL-2	90	no	no	unique sequence	ATTTATATAAATAAAATTAAAATTTTAAAAAAATAAAAAATTATWTAT
SHOX2_c.*28T>C_1	hiPSCs	ok	BSM	88	no	no	unique sequence	TTTTATATATAAATATTTTTAATTTTATATATATAAAATTAATTTTAA
SHOX2_c.*28T>C_2	hiPSCs	ok	BSM	88	no	no	unique sequence	TTTTATATATAAATATTTTTAATTTTATATATATAAAATTAATTTTAA
SHOX2_c.*28T>C_3	hiPSCs	ok	BSM	88	no	no	unique sequence	TTTTATATATAAATATTTTTAATTTTATATATATAAAATTAATTTTAA

Figure 8 Generation and characterization of patient-specific iPSCs from patient I (*SHOX2* c.849C>A) and patient II (*SHOX2* c.*28T>C). iPSCs were generated by reprogramming peripheral blood mononuclear cells (PBMCs) using the CytoTune-iPSC 2.0 Sendai Reprogramming Kit (Life Technologies). (A) Integration-free reprogramming scheme with Sendai viruses, leading to the generation of two iPSC clones from patient I and three iPSC clones from patient II. (B) Sequencing of *SHOX2* c.849C>A (= SHOX2 p.H283Q) and *SHOX2* c.*28T>C mutations in patient-derived iPSC clones. (C) Detection of alkaline phosphatase activity as a pluripotency marker, representative bright field image of an iPSC colony stained for alkaline phosphatase. (D) Confirmation of loss of Sendai viral transgenes after reprogramming by RT-PCR, a = uninfected PBMCs (negative control), b = infected PBMCs (positive control), c-e = iPSC clones from patient I and patient II. (E) Immunofluorescence detection of the endogenous pluripotency markers NANOG and TRA-1-81 in patient-specific iPSCs at passage 17 or 18 after reprogramming; scale bar, 50 µm. (F) Analysis of expression profile of endogenous pluripotency genes in patient-specific iPSCs by qRT-PCR. Values are normalized to *GAPDH* and relative to uninfected parental patient PBMCs. (G) Schematic of the spontaneous differentiation of iPSCs into EBs. qRT-PCR analysis of lineage markers specific for each of the three embryonic germ layers after 21 days of spontaneous EB differentiation in patient-specific iPSCs. Values are normalized to *GAPDH* and relative to iPSCs harvested at day 0. (H) M-FISH analysis of the two *SHOX2* c.849C>A clones and three *SHOX2* c.*28T>C clones showing no chromosomal abnormalities. (I) Multiplex Human Cell Line Authentication Testing confirmed a shared origin for all SHOX2 2 c.849C>A and SHOX2 c.*28T>C clones, respectively, and excluded cross contaminations with other known cell lines. In addition, the origin from the healthy donor control line (Control_1) used in the lab was excluded. (Sumer et al., under revision)

2.4. Generation of isogenic controls for *SHOX2* c.849C>A and *SHOX2* c.*28T>C iPSC lines

Many iPSC models use cells generated from age-matched healthy donors or unaffected family members as controls.[56,57,60] Late age-onset diseases, such as AF, are characterized by long latency and slow progression, often resulting in subtle phenotypes *in vitro*.[101] To avoid that observable differences between patient-specific and control lines were masked by a variable genetic background or disease confounding factors, isogenic iPSC lines were generated for one of the *SHOX2* c.849C>A and *SHOX2* c.*28T>C clones, respectively. In these control clones the putative disease-causing mutation was corrected to the WT sequence using the CRISPR/Cas system. This genome-editing was performed in a scarless manner to exclude any potential impact on an observable phenotype.[126] Consequently, the insertion of selection markers or silent mutations to prevent re-cutting by the Cas9 enzyme was avoided. A novel method for enriching gene-corrected iPSCs prior to single-cell cloning was developed based on a previously reported strategy, which proposed the use of sib-selection to capture rare editing events.[102]

2.4.1. gRNA design and validation

To target the heterozygous *SHOX2* c.849C>A and *SHOX2* c.*28T>C mutations, gRNAs with cutting sites close to the mutation of interest were selected using *CCTop - CRISPR/Cas9 target online predictor*.[118] The gRNAs were predicted to range from moderately to highly efficient with a substantial number of off-targets. However, as gene conversion tracks are relatively short in mammalian cells,[127] the distance between Cas9 cut site and targeted DNA sequence had to be minimized for high HDR efficiency. One of the gRNAs targeting the *SHOX2* c.*28T>C region (gRNA-3) utilized the PAM site created by the T>C mutation and was therefore presumed to bind in an allele-specific manner.

Cells were transfected with Cas9 ribonucleoprotein RNP/gRNA complexes and a ssODN as HDR template (**Figure 8A**). After 48h, genome-targeting efficiency of each gRNA was determined by NGS for a precise estimation of indel size, frequency and sequence identity.[128] For the *SHOX2* c.849C>A locus, the two selected gRNAs were moderately effective, producing indel frequencies of 36% and 21%, respectively. For *SHOX2* c.*28T>C, gRNA-1 and gRNA-2 were less effective, producing indel frequencies of 17% in iPSCs, while the amount of detectable indels in gRNA-3-transfected cells barely exceeded the negative control (**Figure 8B**). gRNA-3 was therefore deemed to be non-functional and was excluded from further applications. Due to the heterozygous nature of the *SHOX2* mutations, the frequency of HDR events in these large cell pools could not be determined precisely.

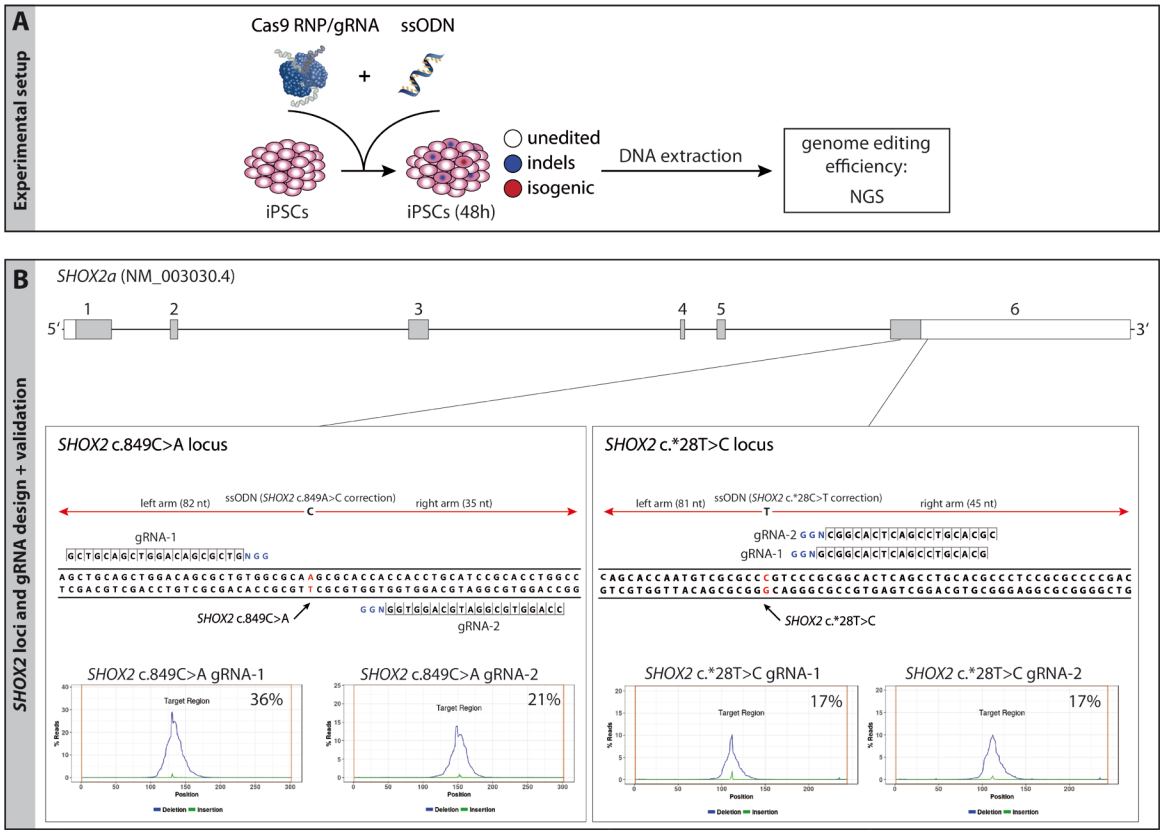


Figure 9 gRNA design and validation for the *SHOX2* c.849C>A and the *SHOX2* c.*28T>C locus. (A) Experimental overview for gRNA validation. hiPSCs were transfected with Cas9 RNP/gRNA complexes and ssODNs. After 48h, the Cas9-targeted region was amplified and analyzed via NGS. (B) Box, upper half: The *SHOX2* c.849C>A and *SHOX2* c.*28T>C locus with selected gRNA binding sites. PAM sequences are depicted in blue, the mutated base pair in red. Box, lower half: Indel frequency for each gRNA in hiPSCs 48h after transfection. The targeted region is centered. The cumulative frequency of deletions (blue) and insertions (green) at each position is depicted in percentage of reads (% Reads). (Sumer et al., under revision)
Abbreviations: RNP = ribonucleoprotein, gRNA = single guide RNA, ssODN = single-stranded oligodesoxynucleotide, NGS = Next Generation Sequencing.

2.4.2. Sib-selection and allele quantification with dPCR

As previously described, the fractionation of a heterogenous population of genome-edited stem cells can randomly enrich desired sub-populations, such as isogenic cells.[102] A population of cells containing a small number of cells of interest is subdivided into small pools ('sib-selection'). From these, the one with the highest percentage of target cells is selected and subjected to a new round of subdivision. After the enrichment of target cells to a reasonable amount, single-cell cloning can be performed to achieve a pure cell population. It was hypothesized that - by quantifying the ratio of WT and Mut alleles in these sib-selections – cells which had precisely corrected the heterozygous mutation would be detectable. A present sub-population of isogenic cells contributes two WT alleles to the DNA pool leading to a shift in the WT/Mut allele ratio. Sib-selections with an over-average abundance of WT alleles potentially contain enriched amounts of isogenic cells and can therefore be selected for further analysis and single-cell cloning.

The massive sample partitioning in a digital PCR system (such as ddPCR™), which allows tens of thousands of individual PCR reactions from single molecule templates, makes this method suitable for precise quantification and identification of DNA strands. Utilizing this property, it has been

successfully applied in low gene expression quantification, copy number variation (CNV) analysis and detection of rare genome editing events.[102,129,130] Here, it was utilized to compare WT/Mut allele ratios within individual sib-selections. Primers and TaqMan probes specific for the respective WT alleles (*SHOX2* c.849C and *SHOX2* c.*28T, labelled with HEX fluorophore) and Mut (*SHOX2* c.849C and *SHOX2* c.*28C labelled with FAM fluorophore) alleles were designed to be used in the dPCR reaction (**Figure 9A**).

For a reliable quantification, the sensitivity and specificity of probes and primers had to be tested. The TaqMan probe specificity for their respective alleles was confirmed with plasmids containing a mutated or unmutated *SHOX2* gene (**Figure 9B**). The sensitivity and specificity of the system was analyzed by mixing different ratios of genomic DNA from healthy donors with DNA from patient-specific iPSCs and comparing detected allele ratios to actual ratios. The strong correlation between calculated and measured Mut alleles indicated a high precision of allelic quantification as well as a specificity of WT and Mut TaqMan probes for their respective alleles (**Figure 9C**).

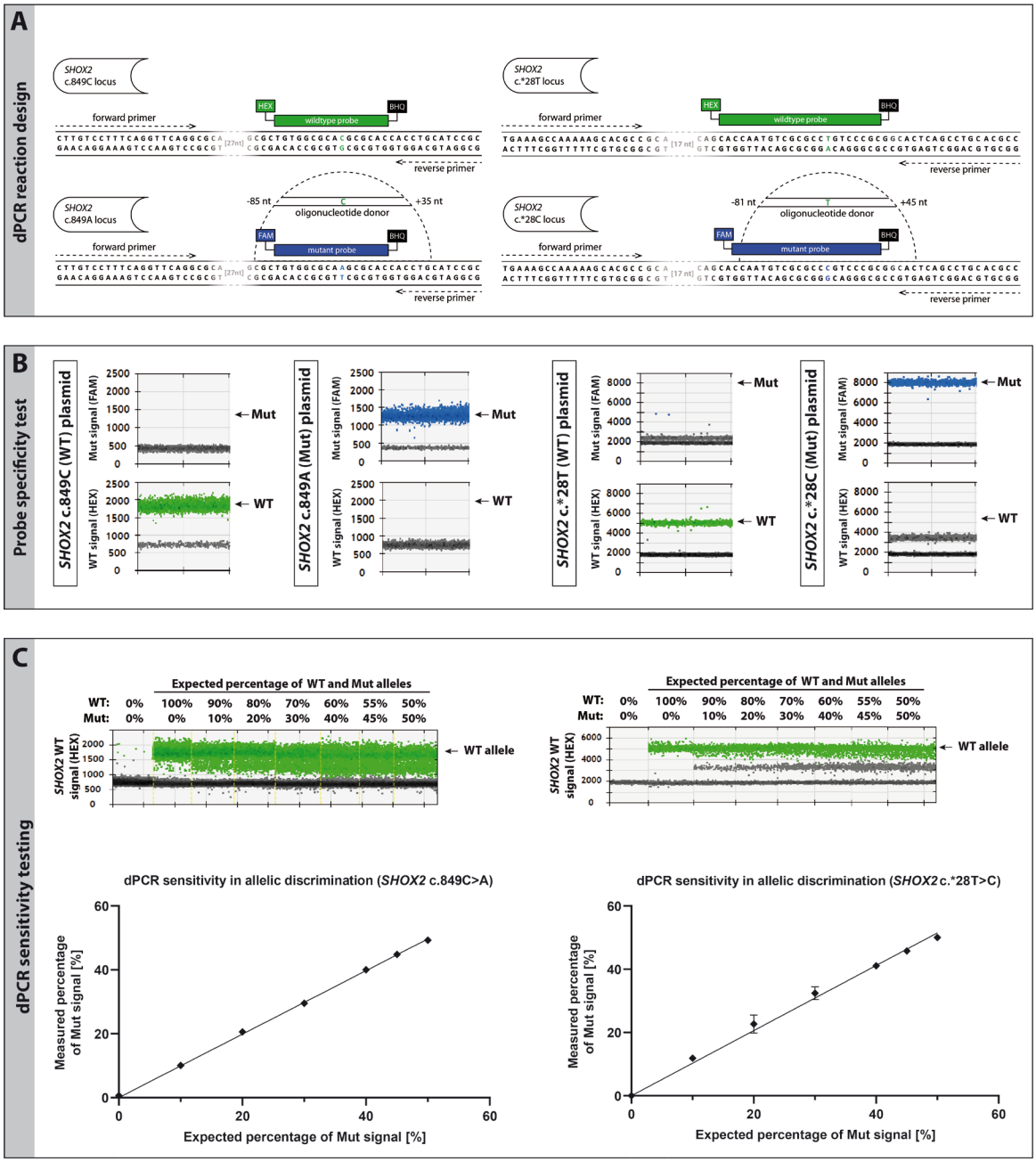


Figure 10 Pretests for allele quantification via dPCR. (A) Primer/probe design for detection of WT and Mut alleles with specific probes. The mutation-spanning oligonucleotide donor is depicted above the mutant allele. (B) dPCR result for different ratios of control and patient genomic DNA represented as HEX channel 1D amplitude (upper panel) and plotted against the expected percentage (lower panel); n=3, error bars represent ± SD of the mean. (C) Probe specificity test with plasmids containing WT (*SHOX2* c.849C, *SHOX2* c.*28T) and Mut (*SHOX2* c.849A, *SHOX2* c.*28C) alleles. (Sumer et al., under revision)
Abbreviations: dPCR = digital PCR, WT = wildtype, Mut = mutant, gDNA = genomic DNA, FAM = 6-Carboxyfluorescein, HEX = Hexachloro-fluorescein, BHQ = black hole quencher.

Cas9 RNP/gRNA/ssODN-transfected iPSCs were seeded into small cell pools of 200 cells per well on a 96-well plate, 48 hours after transfection and grown until confluency (~8-10 days) (**Figure 11A**). From each well, 50% of the cells were cryopreserved while the other 50% were subjected to DNA extraction for subsequent allele quantification with dPCR (**Figure 11B**). While most sib-selections still showed a nearly equal allele distribution of 50:50 (dotted line), which is characteristic for a heterozygous mutation and indicates no enrichment of isogenic cells, a higher

abundance of WT alleles was seen in some of them. These cell pools were therefore selected for further analysis.

The major limitation of allele quantification via dPCR is its dependence on a functional PCR reaction. Mutations or the complete loss of primer/probe binding sites in alleles can prevent a successful amplification of the DNA strands and therefore their identification. If large populations of iPSCs in a sib-selection contain WT alleles and non-detectable mutant alleles, it will generally lead to a shift in the WT/Mut allele ratio similar to what is caused by isogenic cells (**Figure 11C**). Consequently, sequences had to be analyzed in detail via NGS to confirm the presence of isogenic cells in the chosen sib-selections.

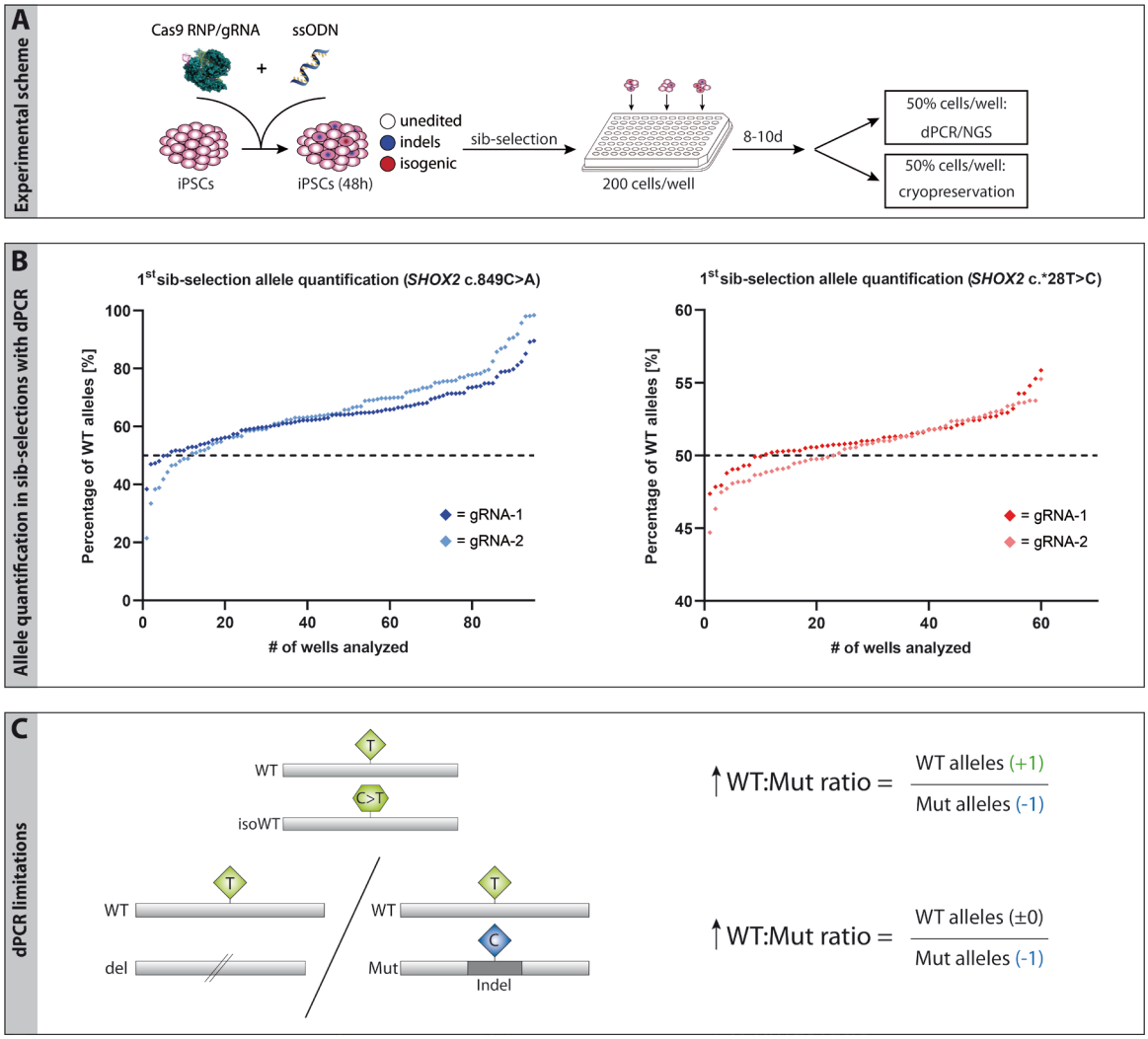


Figure 11 Allele quantification in sib-selections via dPCR and NGS. (A) Experimental scheme of gene-editing approach. (B) WT and Mut alleles were quantified in each sib-selection 10 days after transfection via dPCR. Each dot represents the result for one sib-selection. The dotted line marks the 50% WT allele percentage expected in unedited hiPSCs. Sib-selections with the highest abundance of WT alleles were thawed and re-analyzed. (C) Limitations of dPCR in allele quantification: Due to the PCR-based allele detection, non-amplifiable Mut alleles can lead to shifts in the WT:Mut ratio similar to what is caused by isogenic subpopulations. (Sumer et al., under revision)
Abbreviations: Indel = Insertions/Deletions

2.4.3. Allele quantification in sib-selections via NGS

For *SHOX2* c.849C>A, three cell pools transfected with gRNA-1 and four cell pools transfected with gRNA-2 were subjected to NGS, for *SHOX2* c.*28T>C, DNA of two cell pools transfected with gRNA-1 or gRNA-2 were deep-sequenced. To determine the percentage of targeted cells from WT:Mut allele ratios, two assumptions were made: First, the probability for *SHOX2* CNVs, for example due to a trisomy or gene duplications, was considered to be very small. CNVs could lead to changes in allele ratios that are also not caused by isogenic subpopulations. Second, the probability of an imprecise correction of the *SHOX2* mutations (mutation lost, but other mutations introduced) was neglected for that moment. Under normal circumstances, HDR leads to a complete restoration of the sequence. With these prerequisites, the percentage of isogenic subpopulations could be directly calculated from the WT and Mut alleles.

NGS reads were classified into three categories: WT alleles (reads with *SHOX2* c.849C or *SHOX2* c.*28T), Mut alleles (reads with *SHOX2* c.849A or *SHOX2* c.*28C) and non-assignable (N/A) alleles (reads with *SHOX2* c.849 or *SHOX2* c.*28 base-spanning deletions). This classification was carried out independently of additional mutations in the respective alleles. N/A alleles caused uncertainty in allele quantification, as their origin from either WT or Mut alleles could not be determined. This uncertainty was addressed by defining all N/A alleles as either WT or Mut alleles when calculating the WT/Mut allele ratio. The result was a range of possible ratios, spanning from the two extreme scenarios where N/A alleles were counted as either all WT or all Mut (**Figure 12**). Using gRNA-1 to correct the *SHOX2* c.849C>A mutation, led to a high percentage of reads with deletions spanning the mutation of interest, which made the precise quantification of isogenic sib-selections impossible (**Figure 12A**). For the *SHOX2* c.849C>A sib-selections gRNA-2 #2-4, a large fraction of N/A alleles led to a wide range of potential allele ratios. As this included a scenario, in which the higher percentage of WT alleles could solely be explained by a loss of detectable Mut alleles, a sub-population of isogenic cells was possible, but not guaranteed. On the other hand, sib-selection *SHOX2* c.849C>A gRNA-2 #1 had a strong and robust increase of WT alleles, indicating a large fraction of isogenic clones (**Figure 12B**). For *SHOX2* c.*28T>C sib-selections gRNA-1 #1 and gRNA-2 #1 showed a similar shift and were therefore selected for single-cell cloning together with *SHOX2* c.849C>A gRNA-2 #1. Allelic distributions were used to calculate the percentage of isogenic cells in these cell pools, estimating frequencies of 20-26% for *SHOX2* c.849C>A gRNA-2 #1 as well as 7-9% and 5-9% in sib-selection gRNA-1 #2 and gRNA-1 #1, respectively (**Figure 12B, red arrows**). With a supposed initial efficiency of ~1% for precise genome-editing, the sib-selection process led to a significant enrichment of target cells.

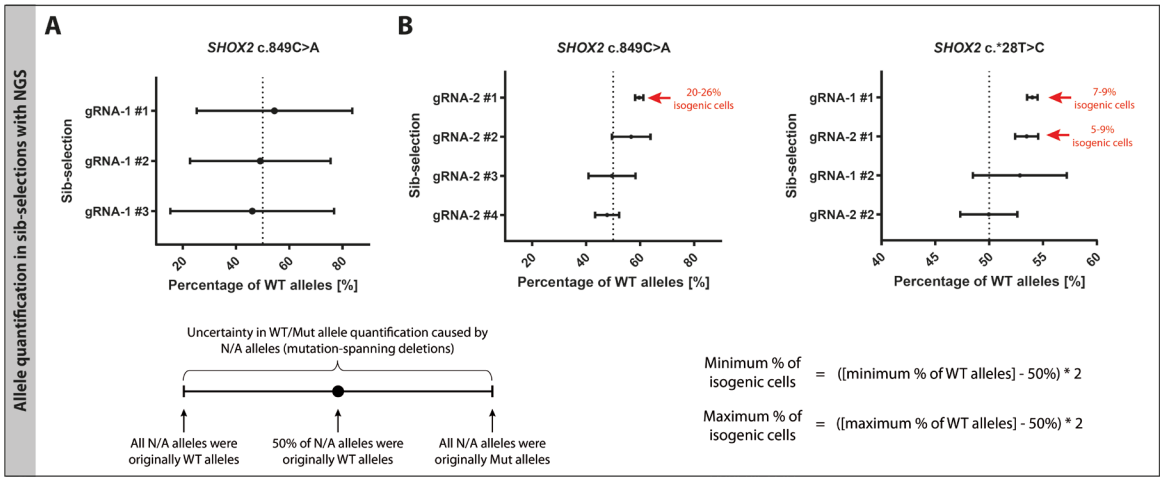


Figure 12 Allele quantification via NGS. Alleles with *SHOX2* c.849- and *SHOX2* c.*28-spanning deletions cause an uncertainty in allele quantification that is addressed by defining those alleles as all WT or all Mut. The resulting span of possible allele ratios is represented as error bars. (A) Alleles with SHOX2 c.849-spanning deletions cause an uncertainty in allele quantification that is addressed by defining those alleles as all WT or all Mut. The resulting span of possible allele ratios is represented as error bars. However, due to the high numbers of reads containing large deletions, the ratios cannot be determined precisely enough. (B) Sib-selections in which an increased WT:Mut allele ratio is not solely explicable by a loss of detectable Mut alleles were chosen for single-cell cloning. The percentage of isogenic cells was calculated with the given formula. (Sumer et al., under revision)
Abbreviations: Indel = Insertions/Deletions, N/A = non-assignable.

2.4.4. Single cell-cloning and screening

The number of cells that had to be screened to find at least one isogenic/corrected clone with a given probability was determined by negative binominal distribution. The applied parameters were the calculated frequency of target cells (~23% or ~8%), the desired number of positive clones to be found (≥ one clone) and a self-defined chance of success to find one (95%) (**Figure 13A**). Twelve clones for sib-selection *SHOX2* c.849C>A gRNA-2 #1 and 35-40 clones for *SHOX2* c.*28T>C sib-selections had to be generated to find at least one clone of interest with a 95% probability. Compared to the nearly 300 cells that would have to be screened for the same chances of success - if the target cell population was only ~1% - this option represented a substantial reduction of screening workload.

Monoclonal cell populations were obtained via limiting dilution cloning (**Figure 13B**). The clonability, which is defined as the success rate for obtaining a single cell-derived iPSC clone from every well seeded, was determined to be 10-20%. Using three 96-well plates per cell line resulted in 24, 33 and 57 analyzable clones for sib-selection *SHOX2* c.849C>A gRNA-2, *SHOX2* c.*28T>C gRNA-1 #2, and *SHOX2* c.*28T>C gRNA-2 #2, respectively (**Figure 13C**). From each well, 50% of the cells were cryopreserved while the other 50% were subjected to DNA extraction for subsequent genotyping.

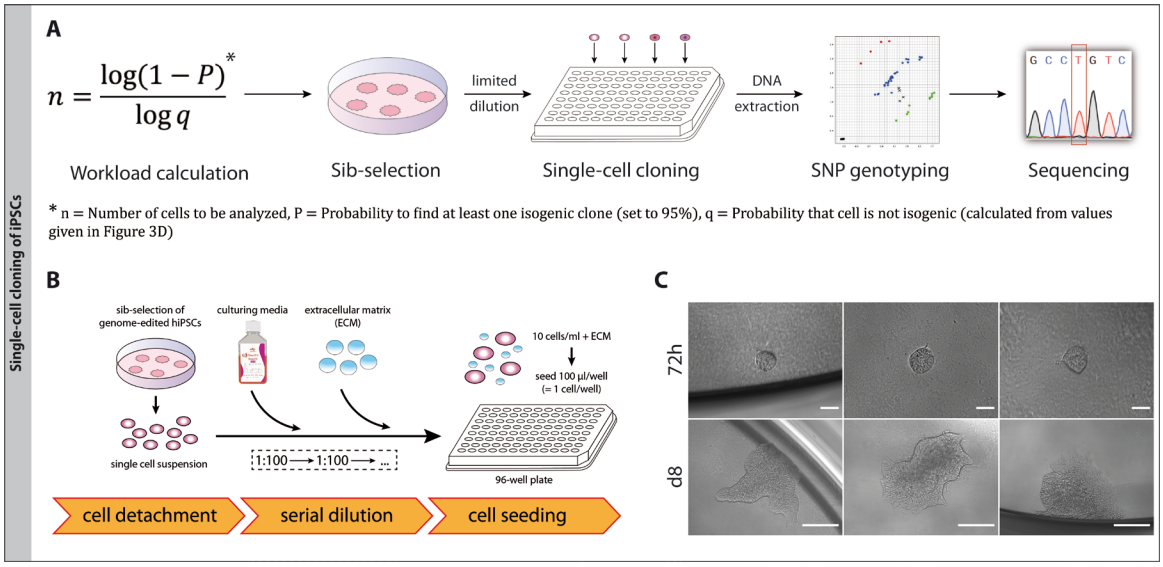


Figure 13 Single cell-cloning of iPSCs (A) Coating free method for single-cell cloning via limited dilution. A single-cell suspension is generated from the chosen sib-selection and diluted to 10 cells/ml. The extracellular matrix is added to the cell suspension and 1 cell/well is seeded on 96-well plates (100 µl). (B) Single-cell derived hiPSC colony after 72h (upper row; scale bar, 50 µm) and 8 days (lower row; scale bar, 200 µm). (Sumer et al., under revision)

Both digital PCR primers and probes were reused for TaqMan SNP genotyping of *SHOX2* c.849C>A and *SHOX2* c.*28T>C to pre-select potentially homozygous WT clones. DNA extracted from an unrelated control line (control_1) was used as a positive control for homozygous WT and patient-derived DNA as positive control for heterozygous WT. All annotated homozygous WT clones were identified from the allele discrimination plot. The automated classification could not precisely annotate the different allele combinations due to the lack of a positive control for a homozygous mutation. The melting curves of single-cell clones were therefore manually compared to the WT positive controls (**Figure 14A**). Subsequent sequencing confirmed the precise correction of the heterozygous *SHOX2* c.849C>A or *SHOX2* c.*28T>C mutation and to screen for additional mutations up- and downstream of the Cas9 target site. For *SHOX2* c.849C>A, 5 out of 24 sequenced clones (~21%) were confirmed to have lost the patient mutation, matching the expected frequency of 20-26%. However, in two of them, additional mutations were introduced during the editing process. For *SHOX2* c.*28T, 5 of the genotyped WT clones (1x from sib-selection gRNA-1 #1 and 4x gRNA-2 #1) were confirmed to be isogenic with no additional detectable mutations, neither ~600 bp upstream nor ~260 bp downstream of the Cas9 cut site. For the rest of the clones that were predicted to be homozygous WT via genotyping, a loss of primer/probe binding sites on the mutant allele explained the false annotation (**Figure 14B**).

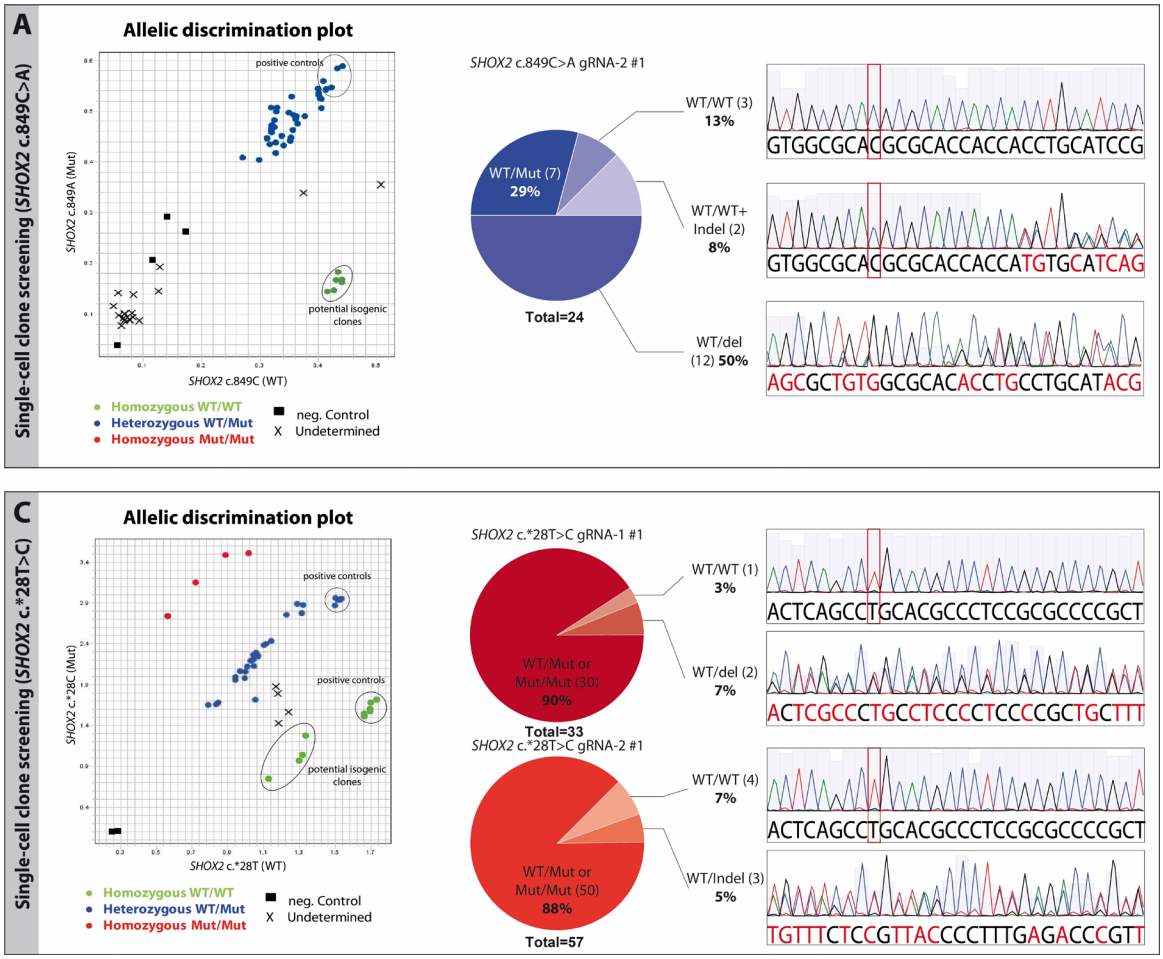
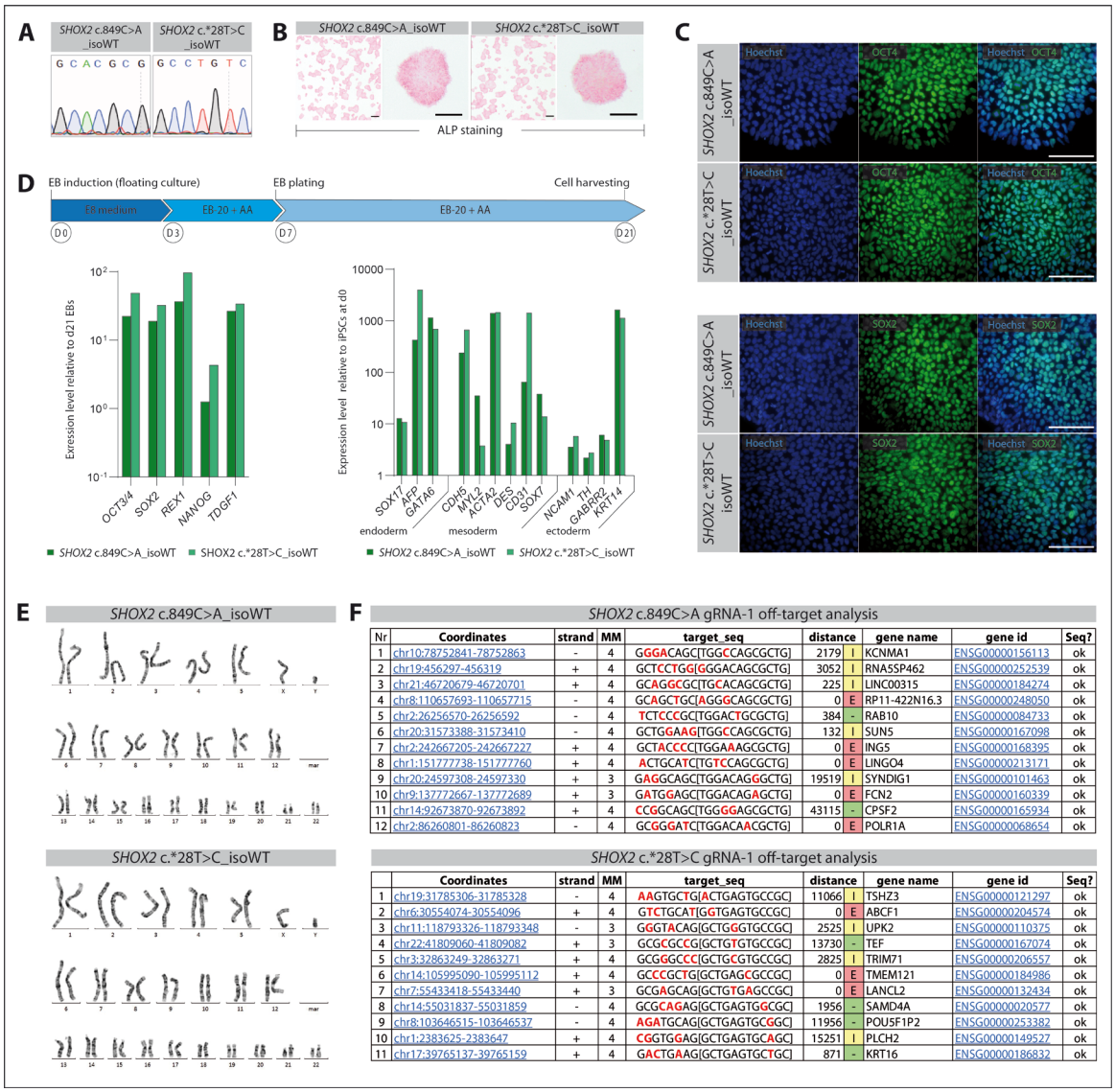


Figure 14 Single cell-cloning with sib-selections and screening for isogenic clones. (A) Experimental scheme for single-cell cloning and screening: The number of clones needed to be analyzed was calculated with binomial distribution function. Sib-selections with isogenic subpopulations were thawed for single-cell cloning via limited dilution. Clones were screened via TaqMan Probe-based SNP genotyping and potential homozygous WT clones were confirmed with Sanger Sequencing. (B) Screening for isogenic clones derived from heterozygous*SHOX2* c.849C>A cells: 24 single-cell clones were genotyped and sequenced. In 5/24 clones (21%) the mutation was corrected back to wildtype, with 3 clones showing no additional mutations several hundred nucleotides up- and downstream. (C) Screening for isogenic clones derived from heterozygous *SHOX2* c.*28T>C cells: Single-cell derived clones were genotyped. 10 annotated homozygous WT clones were sequenced to confirm the loss of the *SHOX2* c.*28T>C mutation. In 5/10 clones the mutation was repaired precisely back to WT, in the other 5/10 clones, deletions on the Mut allele explained the false annotation. (Sumer et al., under revision)
Abbreviations: SNP = Single nucleotide polymorphism, here: *SHOX2* c.849C>A and *SHOX2* c.*28T>C, Indel = Insertions/deletions.

2.4.5. Re-characterization of isogenic control lines for *SHOX2* c.849C>A and *SHOX2* c.*28T>C

For subsequent detailed re-characterization, one isogenic clone for *SHOX2* c.*28T>C isoWT and *SHOX2* c.849C>A isoWT were selected. Both clones had maintained their stem-cell like morphology and pluripotent capacity and also exhibited high ALP activity, expression of pluripotency markers on RNA/protein level and spontaneously differentiated into derivates from all three germ layers (**Figure 14A-D**) (work performed by Viktoria Frajs under supervision, also see Master Thesis Viktoria Frajs *SHOX2 in atrial fibrillation disease modelling using induced pluripotent stem cells*). Classical cytogenetic analysis on Giemsa stained chromosomes revealed in each of the 30 investigated metaphases (**Figure 14E**). Cell line authentication confirmed their patient-specific origin and excluded a cross-contamination with the control line (**Figure 13I**).

Twelve and eleven highly scored off-targets for *SHOX2* c.849C>A gRNA-2 and *SHOX2* c.*28T>C gRNA-1 consisting of exonic regions as well as intronic and intergenic regions with potential regulatory relevance were sequenced, but no additional mutations were found to be introduced in these sites (**Figure 14F**).



Nr	Coordinates	strand	MM	target_seq	distance		gene name	gene id	Seq?
1	chr10:78752841-78752863	-	4	GGGACAGC[TGGCCAGCGCTG]	2179	I	KCNMA1	ENSG00000156113	ok
2	chr19:456297-456319	+	4	GCTCCTGG[GGGACAGCGCTG]	3052	I	RNA5SP462	ENSG00000252539	ok
3	chr21:46720679-46720701	+	4	GCAGGCGC[TGCACAGCGCTG]	225	I	LINC00315	ENSG00000184274	ok
4	chr8:110657693-110657715	-	4	GCAGCTGC[AGGGCAGCGCTG]	0	E	RP11-422N16.3	ENSG00000248050	ok
5	chr2:26256570-26256592	-	4	TCTCCCGC[TGGACTGCGCTG]	384	-	RAB10	ENSG00000084733	ok
6	chr20:31573388-31573410	-	4	GCTGGAAG[TGGCCAGCGCTG]	132	I	SUN5	ENSG00000167098	ok
7	chr2:242667205-242667227	+	4	GCTACCCC[TGGAAAGCGCTG]	0	E	ING5	ENSG00000168395	ok
8	chr1:151777738-151777760	+	4	ACTGCATC[TGTCCAGCGCTG]	0	E	LINGO4	ENSG00000213171	ok
9	chr20:24597308-24597330	+	3	GAGGCAGC[TGGACAGGGCTG]	19519	I	SYNDIG1	ENSG00000101463	ok
10	chr9:137772667-137772689	+	3	GATGGAGC[TGGACAGAGCTG]	0	E	FCN2	ENSG00000160339	ok
11	chr14:92673870-92673892	+	4	CCGGCAGC[TGGGGAGCGCTG]	43115	-	CPSF2	ENSG00000165934	ok
12	chr2:86260801-86260823	-	4	GCGGGATC[TGGACAACGCTG]	0	E	POLR1A	ENSG00000068654	ok

	Coordinates	strand	MM	target_seq	distance		gene name	gene id	Seq?
1	chr19:31785306-31785328	-	4	AAGTGCTG[ACTGAGTGCCGC]	11066	I	TSHZ3	ENSG00000121297	ok
2	chr6:30554074-30554096	+	4	GTCTGCAT[GGTGAGTGCCGC]	0	E	ABCF1	ENSG00000204574	ok
3	chr11:118793326-118793348	-	3	GGGTACAG[GCTGGGTGCCGC]	2525	I	UPK2	ENSG00000110375	ok
4	chr22:41809060-41809082	+	3	GCGCGCCG[GCTGTGTGCCGC]	13730	-	TEF	ENSG00000167074	ok
5	chr3:32863249-32863271	+	4	GCGGGCCC[GCTGCGTGCCGC]	2825	I	TRIM71	ENSG00000206557	ok
6	chr14:105995090-105995112	+	4	GCCCGCTG[GCTGAGCGCCGC]	0	E	TMEM121	ENSG00000184986	ok
7	chr7:55433418-55433440	+	3	GCGAGCAG[GCTGTGAGCCGC]	0	E	LANCL2	ENSG00000132434	ok
8	chr14:55031837-55031859	-	4	GCGCAGAG[GCTGAGTGGCGC]	1956	-	SAMD4A	ENSG00000020577	ok
9	chr8:103646515-103646537	-	4	AGATGCAG[GCTGAGTGCGGC]	11956	-	POU5F1P2	ENSG00000253382	ok
10	chr1:2383625-2383647	+	4	CGGTGGAG[GCTGAGTGCAGC]	15251	I	PLCH2	ENSG00000149527	ok
11	chr17:39765137-39765159	+	4	GACTGAAG[GCTGAGTGCTGC]	871	-	KRT16	ENSG00000186832	ok

Figure 15 Re-characterization of *SHOX2* c.849C>A_isoWT and *SHOX2* c.*28T>C_isoWT. Isogenic control lines were generated as described above and re-characterized to confirm the preservation of pluripotency capacity and karyotype. (A) Loss of heterozygous *SHOX2* c.849C>A (= SHOX2 p.H283Q) and *SHOX2* c.*28T>C mutations in isogenic control clones. (B) Detection of alkaline phosphatase activity in isogenic hiPSCs as a pluripotency marker. (C) Immunofluorescence detection of the endogenous pluripotency markers OCT4 and SOX2 in isogenic hiPSCs; scale bar, 100 µm. (D) qRT-PCR analysis of pluripotency genes at d0 and germ layer markers after 21 days of EB differentiation (see the schematic above). (E) Giemsa banding of isogenic control lines revealing no chromosomal aberrations. (Sumer et al., under revision)

2.4.6. Estimation of the HDR frequency via reverse genome-editing

When calculating the enrichment of homozygous wildtype cells and the reduction in workload, the reported HDR frequency of ~1% was assumed.[66,101,131] To determine if the sib-selection process indeed enriched targeted cells, the initial HDR efficiency had to be quantified. For an approximation of the real HDR efficiency in these specific settings, the *SHOX2* c.849C>A and *SHOX2* c.*28T>C mutations were re-introduced into the isogenic clones using the same Cas9 enzyme batch and gRNAs but replacing the ssODN by analogous versions, which led to the insertion of the point mutation rather than to its repair. Seventy-two hours after transfection, the frequency of HDR events was determined by NGS (**Figure 15A**). The *SHOX2* c.849C>A mutation was detected in 0.70% of the reads about half of them with additional mutations such as deletions introduced (**Figure 15B**). The *SHOX2* c.*28T>C mutation was found in 0.85% of the reads, but in nearly all cases the base substitution had happened scarlessly (**Figure 15C**).

In conclusion, the HDR events had happened in a frequency of ~1% correlating with previous reports, and the sib-selection process with the detection via allele quantification led to an 8-20-fold enrichment of isogenic cells.

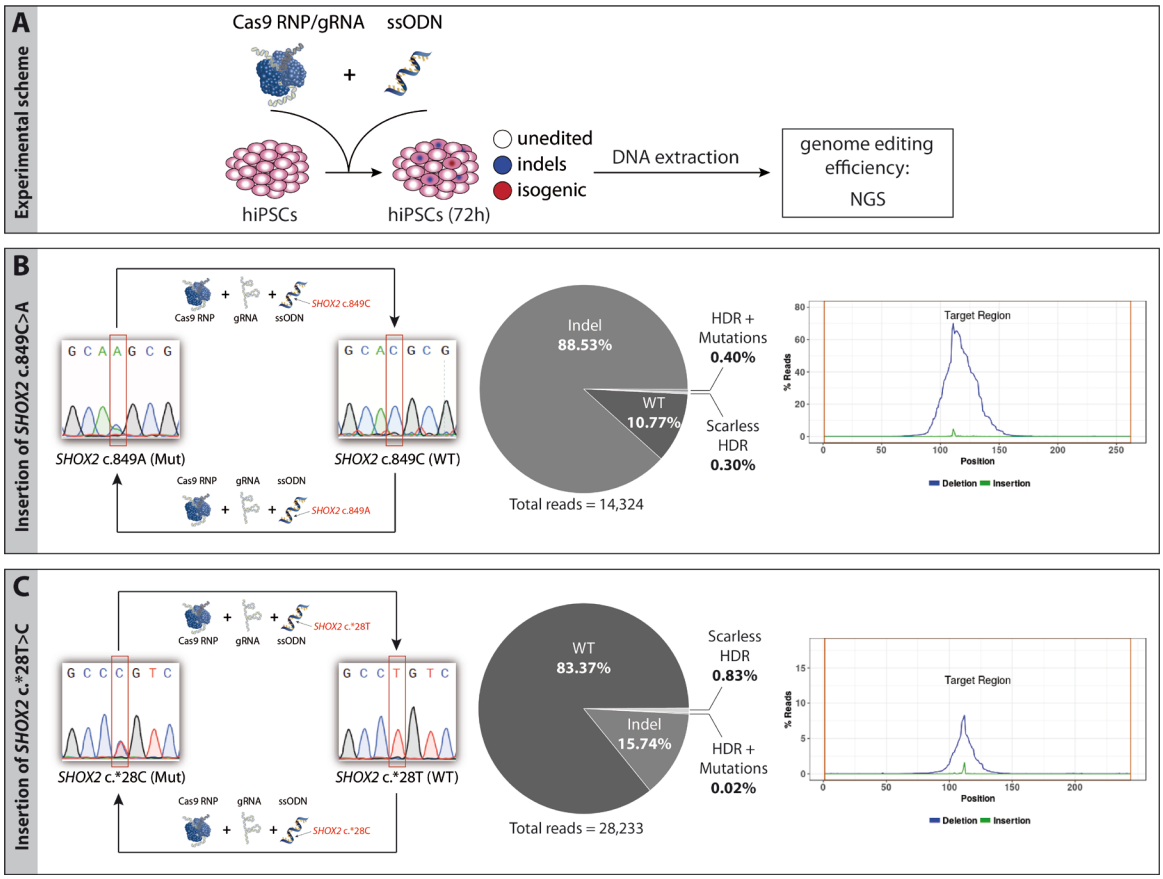


Figure 16 Reverse genome-editing to determine the initial HDR efficiency. (A) Experimental scheme for reverse gene-editing. The newly generated isogenic clones were used to determine the initial HDR efficiency. For this, the same Cas9 RNP and gRNA were used in combination with an oligonucleotide donor differing only in the one base analogous to *SHOX2* c.849 (B) or *SHOX2* c.*28 (C). In consequence, HDR events would lead to the introduction of the *SHOX2* c.849C>A or *SHOX2* c.*28T>C mutation. The frequency of HDR was determined by NGS 72h after isogenic cells were transfected with Cas9/gRNA/ssODNs. (Sumer et al., under revision)
Abbreviations: HDR = homology-directed repair

3. Discussion

3.1. Functional characterization of rare variants in the *SHOX2* gene identified in SND and AF

The proper function of the cardiac conduction system is mediated by a highly conserved and complex network of transcriptional regulators.[2,9] Arrhythmias and other cardiac conduction defects are often polygenic with thousands of common risk alleles underlying the disease onset, but they can also be caused by haploinsufficiency of single cardiac transcription factors.[132,133] AF, the most prevalent cardiac rhythm disorder, has a strong genetic component as seen by a significant familial aggregation of disease cases. Research into the genetic basis of AF has led to the discovery of several transcription factors as potential contributors to arrhythmia susceptibility. These include *TBX5*, an activator of *SHOX2*[29], *NKX2.5*, an antagonist of *SHOX2,*[30] and *PITX2*, a repressor of *SHOX2*.[17] Further evidence to support the role of transcription factors in the manifestation of AF has come from transgenic and pacing-induced AF animal models.[132] In addition, mutations in several SHOX2 regulated genes have been causally linked to AF, including the CCS-specific gene *HCN4* or the suppressed working myocard genes *GJA5* (*CX40*) and *GJA1* (*CX43*).[134-136] SND is often the consequence of structural degeneration or remodeling processes due to ageing or pathological conditions such as AF, infarction or heart failure. Additionally, idiopathic degeneration of the SAN with familial inheritance has been described.[137] AF and SND often coexist in a clinical context, but the causality between the two conditions and shared molecular mechanisms are poorly understood.[35,43] Yet,, there is a proposed substantial overlap between candidate genes for SND and AF.[137]

Genome-wide association studies did not provide evidence for a link between *SHOX2* and AF or SND so far.[27] More recent data, however, has given first clues that certain variants in this gene were linked to AF.[28,50] In the present study, the mutational analysis has been expanded to investigate if *SHOX2* also represents a common susceptibility gene for SND and general AF, and to unravel a putative shared genetic etiology that underlies both conditions. A total of four heterozygous variants were detected in a cohort of 450 AF patients comprising two synonymous (p.L192=; p.A293=) and two non-synonymous (p.G77D; p.L130F) variants. In the cohort of 98 SND patients, one heterozygous missense mutation (p.P33R) was identified. All variants resided outside the DNA-binding homeodomain. For a first estimation of the pathogenic potential, several prediction tools based on machine learning approaches or empirical scoring systems and CADD scoring was used. CADD combines diverse genomic features, such as derived evolutionary constraint, functional predictions, epigenetic measurements, surrounding sequence context and gene model annotations. All of these annotations are integrated into a single CADD or C-score for any given variant.[123,124]

The C-scores of the identified variants ranged from 9.02 to 25.8. A scaled C-score ≥20 of a variant indicates it to be within the 1% most deleterious substitutions in the human genome. None of the variants with a CADD score above 20 (p.G77D, p.A293=, p.P33R) were present in the European Non-Finnish population of the 1,000 genomes project or the DZHKomics database, indicating very rare events, and only p.P33R (1/50770) and p.A293= (55/60141) were reported in the gnomAD database. However, late-onset cardiovascular traits cannot be entirely excluded in the gnomAD database even though participants from the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine program (www.nhlbiwgs.org), were excluded in this analysis.

Subsequent *in vivo* studies in the zebrafish revealed a functional consequence only for p.G77D. Upon cardiac-specific overexpression of this missense mutation, pericardial edema and significantly reduced heart rates could be observed. No phenotype resulted from the overexpression of the synonymous variant p.A293=, which had the lowest C-score (20.6) and highest allele frequency in the control population. The SND missense variant p.P33R with the highest C-score (25.8) and a consistent classification as deleterious variant had a relatively low allele frequency in the control population but did not show obvious phenotypic effects in the zebrafish model. However, *in vitro* reporter assays demonstrated an impaired transactivation activity for this variant as well as for the variant p.G77D. The result for p.P33R could also be confirmed in zebrafish, were molecular changes resulting in an altered target gene expression of *Bmp4*.

In the first study that linked *SHOX2* variants to AF, mutated Shox2 was ectopically expressed in embryonic zebrafish hearts after morpholino-induced knockdown of endogenous *Shox2*.[28] Here, the variants were tested with human *SHOX2* and without alternating the intrinsic *Shox2* expression. p.G77D and p.P33R therefore have a dominant negative effect on SHOX2 function. SHOX2, like its paralog SHOX, is capable of forming homo- and hetero-dimers. Given that cooperative dimerization of paired-related homeodomains to DNA increases the transactivation efficiency,[138] it is possible that these variants negatively affect the DNA binding capacity of dimers if one of the SHOX2 proteins is mutated.

SHOX2 is a key transcription factor in the development of the SAN and could contribute to an increased susceptibility to SND and AF in different direct and indirect ways:

Impaired expression within the SAN has been shown in *Shox2*$^{-/-}$ mice, who die from severe SV and SAN hypoplasia between E11.5 and E17.5 due to reduced pacemaker cell proliferation and a switch to the genetic program of a working myocard.[48,49] The replacement of *Shox2* with human *SHOX* plus a *PGK-neo* cassette (*Shox2*$^{KI+Neo/KI+Neo}$) in the mouse heart results in a hypomorphic allele with reduced gene expression. *Shox2*$^{KI+Neo/KI+Neo}$ mice exhibit arrhythmias and severe bradycardia leading to death within days after birth, which supports the hypothesis of a genetically compromised *SHOX2* copy in patients resulting in AF.[139] Similarly, the deregulation of Shox2 antagonists and repressors for example due to mutations in gene expression regulation sites, could have the same impact. When the antagonist *Nkx2.5* is overexpressed, it causes the same SAN and sinus valve

hypoplasia and dysregulation of the SAN genetic network that is observed upon Shox2 knockout. This indicates that *Nkx2.5* activity is detrimental to SAN development by de-repressing the atrial myocardium development.[140] Physiologically, the loss of functional *SHOX2* can lead to the reduced expression of crucial ion channels and electrical conductors such as *HCN4*, *CX30.2* and *CX45* impeding the role of the SAN as the dominant pacemaker. While it might still be a functional tissue under normal conditions, it could facilitate the stabilization of ectopic re-entry points in the atrium making the carriers of *SHOX2* mutations prone to the manifestation of AF.

Another potential mechanism of how *SHOX2* could influence the development of AF, is the abnormal expression of this gene outside the SAN. Unlike *SHOX2*, *PITX2* has been strongly associated with AF in GWAS studies.[141,142] In mice, this gene suppresses *Shox2* by direct interaction with its promotor and indirect regulation through the inhibitory effects of miRNA miR-1792 and miR-106-25.[143] $Pitx^{+/-}$ mice develop atrial arrhythmias upon pacing and show increased expression of SAN related genes, including *Shox2* and *Tbx3*, as well as other AF-related genes such as the potassium-channel *Kcnq1*. It was hypothesized that this generated an arrhythmogenic substrate due to dysregulated gene expression, which would enhance other pathological triggers for AF.[17] Similarly, reduced activity of the *SHOX2* antagonist *NKX2.5* could lead to the same effect. In the past, several mutations in the transcriptional activation domain or the homeodomain of NKX2.5 have been described in patients suffering from lone and familial forms of AF, which lead to a loss-of-function of the protein.[30,144,145] Considering that the mutually exclusive expression patterns of *SHOX2* and *NKX2.5* are mediated through binding of the transcription factors on regulatory elements of the antagonist, the detrimental effects of these mutations could be caused by ectopic expression of *SHOX2*. Especially pulmonary veins, which have been implicated in the initiation and maintenance of AF, are prone to develop pacemaker-like structures. In a *Nkx2.5* hypomorphic model, pulmonary cardiomyocytes developed pacemaker activity by upregulation of *Hcn4* and downregulation of *Cx40*.[16] In reverse, an impaired regulation of *SHOX2* due to pathogenic cis-regulatory variants in enhancer regions could lead to the same effect. However, to date no mutations in enhancer regions of *SHOX2* in the context of cardiovascular diseases have been described and previous studies showed that the ectopic expression of Shox2 alone did not induce SAN-like foci in atria and ventricles in mice, indicating that additional factors would have to play a role.[140]

The interpretation of variants is a major challenge in clinical genetics. Here it could be demonstrated that *in silico* prediction tools alone are not sufficient to determine the pathogenicity of a genomic variant; further assays are needed to determine functionality. The SND cohort size (n=98) may have been too small to detect rare *SHOX2* variants with strong functional consequences. Assuming similar frequencies for rare pathogenic *SHOX2* variants in AF and SND (3/990; 0.3%) larger cohorts are required to detect further SHOX2 variants and decipher *SHOX2* deficiency in SND. As sequencing was restricted to exonic parts of *SHOX2*, noncoding variants affecting regulatory elements could also be addressed.

It will be interesting to complement these findings with more functional investigations *in vitro* and *in vivo* to distinguish pathogenic from nonpathogenic variants. The species-specific electrophysiological differences, the poorly understood mechanisms of atrial remodeling that maintain AF and the complex genetic causes that contribute to this disease make the development of a human AF model crucial. AF has been simulated in artificial tissue consisting of hESC-derived atrial-like cardiomyocytes that reacts to anti-arrhythmic drugs.[64] Recently, the first iPSC-based model for familial AF has been established where iPSC-CMs from patients showed abnormal ion currents and prolonged APDs compared to unrelated controls.[65] However, the genetic cause of the AF was not determined in this study and immature types of CMs were used rather than generating atrial tissue through subtype-specific differentiation. In addition, the analysis of patient-derived SAN-like CMs has not been done yet despite recent advances in differentiation approaches. Establishing an *in vitro* AF model with a clear genetic background opens exciting new possibilities for direct phenotype to genotype comparisons.

3.2. Generation and correction of *SHOX2* c.849C>A and *SHOX2* c.*28T>C iPSC lines

The *SHOX2* c.849C>A or *SHOX2* c.28T>C mutations were the first variants in this gene to be associated with AF. They were demonstrated to have a negative impact on the function of SHOX2 as seen in *in vitro* and *in vivo* studies.[28] However, the molecular basis for these findings can only be insufficiently addressed in non-human model organisms or reporter assays due to the lack of a complete genetic network. The newly generated patient-specific iPSC lines offer unprecedented opportunities to investigate the role of *SHOX2* in the genetic network of the SAN and its influence on the onset and progression of AF. Nevertheless, it should be noted that the genetic analysis of the patients should be expanded as especially patient II (*SHOX2* c.*28T>C) presented with substantial comorbidities that most likely confounded the emergence of arrhythmias. The unusually young age of 37 years at onset, however, suggests a genetic predisposition that could be explained by *SHOX2* c.*28T>C. To address this question and to generate an optimal control for the disease model, these variants were corrected precisely and scarlessly in patient-derived iPSCs with a novel approach.
The concept of using stochastic enrichment of cells by sib-selection to introduce precise mutations into iPSCs, derived from a healthy donor, has been proposed before.[102] The insertion of disease-linked variants into a wildtype background helps to interrogate the influence of these mutations on the onset and progression of a disease by direct comparison of mutated and wildtype cells. Yet, this approach is limited to monogenic diseases or variances with a high impact on the phenotype. Here, it was demonstrated that this approach can be used to correct heterozygous mutations as well. This is of particular interest when using patient-derived iPSCs that already harbor putative disease-

causing variants. Recent and future optimizations in iPSC technology and commercially available tools will further facilitate iPSC derivation, thus making the generation of patient-specific iPSCs more accessible.[67] These patient models play a central role in the investigation of sporadic or idiopathic diseases, where a combination of multiple risk alleles with low effect size is thought to be the genetic base and individual risk variants might not be sufficient to cause a disease-associated phenotype.

AF is a multifactorial disease with a strong genetic component and a complex heritability.[146] Multiple genetic loci have been associated with this disease [27] and mutations in potassium and sodium channels as well as mutations in transcription factors and structural proteins have been identified.[26] However, linking specific mutations to an AF phenotype in iPSC-CMs has not been achieved yet. The correction of putative disease-contributing variants in iPSCs could unravel subtle phenotypic changes when comparing patient cells to their isogenic controls, even if the disease phenotype overall persisted.

Genome-editing also holds potential for human gene therapy approaches, in which somatic cells are reprogrammed to iPSCs and the detrimental mutation is corrected before the cells are differentiated to the desired cell type and re-transplanted into the patient for a beneficial effect.[98] The proposed strategy can be applied to a precise repair of heterozygous mutations that does not require the use of selection marker integration, its transient expression or an enrichment of nuclease-expressing cells by fluorescence-activated cell sorting.[147-149] This does not only abolish the need to optimize the selection or sorting process, but also allows the use of unlabeled nuclease proteins or gene constructs. The enrichment of cells with a high expression of Cas9 via puromycin or cell sorting has led to concerns regarding off-targeting, as prolonged expression or high concentrations of nucleases tend to increase the probability of unwanted DSBs.[150,151] This can be countered by the use of Cas9 RNPs, which reportedly show less off-targeting due to the shorted activity span by immediate DNA cleavage after delivery into cells, followed by a rapid degradation.[152,153] In this study, gene-editing was analyzed via NGS 48h after transfection unravelling substantial targeting efficiencies for some gRNAs. Despite that, no additional mutations were found to be introduced in highly scoring off-target regions.

Avoiding delivery vectors with a potential to integrate into the genome also opens possibilities for gene-editing in clinical settings under good manufacturing practice conditions. In addition, high-fidelity gRNAs with few predicted off-targets should be preferentially chosen.[154] However, to increase the frequency of HDR events, gRNAs must be selected according to the distance between the induced DSB to the HDR-targeted DNA section, rather than off-target scores or predicted efficiency. This is mainly because of a decrease in HDR frequency with increasing distances between Cas9 cut site to the targeted nucleotides, presumably due to short gene conversion tracts in mammalians.[127] Nevertheless, to completely rule out additional editing, a genome-wide analysis via whole genome sequencing would be required. Enriching cells of interest before single-cell

seeding is especially beneficial for cell lines that behave poorly during clonal expansion. Keeping cells even in small pools of 200 cells per well of a 96-well plate greatly enhances survival rates upon splitting. This avoids the selective enrichment of cell populations with abnormal survival advantages or growth rates caused by chromosomal aberrations. Subsequently, low clonability rates (defined as the success chance to isolate a single-cell derived population from each well seeded) do not lead to an immense increase in time and material consumption due to large scale cloning efforts as only a handful of clones have to be analyzed to find an isogenic one. In fact, even with the low clonability rates of 10-20% achieved with the patient lines, only two to three 96-well plates per line were sufficient to find several isogenic populations.

In regions that are difficult to target, one might find the initial HDR efficiency to still be low even after incorporation of improvement strategies such as optimizing the DNA template[108-110], modifying the Cas9 enzyme itself[112,113] or applying small molecules to increase HDR.[103] The proposed strategy can be combined with any of these approaches, as it is locus-independent and solely relies on a stochastic enrichment rather than the modulation of biological processes. Especially emerging techniques like base-editing[155] and prime editing[113] can profit from this approach by complementing their increased efficiency in targeted editing with an additional enrichment of correctly edited cells. Quantifying isogenic subpopulations with NGS enabled predictions for the single cell cloning workload with relative precision. While only one round of sib-selection was used to enrich the cells of interest, several subsequent sib-selections are also possible to further increase the fraction of precisely edited cells and to decrease the required cloning effort. The insertion of blocking mutations together with the mutation correction in the ssODN sequence was purposely avoided. These (silent) mutations are introduced to prevent re-cutting by Cas9 after HDR-mediated editing and reportedly increase the efficiency of precise DNA modifications.[108,156] Yet, their application should be restricted to coding mutations, as unwanted side-effects in UTRs or non-coding regulatory regions by altered posttranscriptional regulation or transcription factor binding cannot be completely ruled out. An example for this is the *SHOX2* c.*28T>C mutation which resides within the 3'UTR and presumably mediates its detrimental effect by the generation of a novel miRNA binding site.[28] A thorough analysis before additional mutations are introduced is highly recommended by examining evolutionary sequence conservation or using prediction tools for miRNA and transcription factor binding.

However, despite the clear advantages of this method, several limitations remain. Laboratories are required to have fast and easy access to NGS and potentially dPCR. dPCR was used to preselect sib-selections with potentially high fractions of isogenic cells despite the stated limitations of this method to quantify such subpopulations. This allowed the sending of single samples for NGS that could be analyzed with freely available online tools such as Cas-analyzer.[120] Nevertheless, a more straightforward approach would be to quantify alleles in all sib-selections at once with deep-sequencing. Although this is more cost-intense and requires amplicon-NGS analysis knowledge due

to sample multiplexing, it would speed up the process significantly and allow to find the one sib-selection that truly has the highest percentage of isogenic cells. With recent and future advances in NGS, this method can be expected to become more feasible and affordable.[157,158] The additional passaging and cryopreservation required in the sib-selection process is a potential source for acquiring mutations and chromosomal translocations in the extended culturing periods.[159,160] On the other hand, and as mentioned before, stressful processes like cell sorting and antibiotic selection are not required. Yet, its applicability to primary cells and cell types that cannot be extensively passaged is limited. Furthermore, the additional culturing periods are time-consuming. Even under ideal circumstances, the isolation of isogenic clones takes several weeks to months. However, most of the time is spent waiting for the cells to grow and during allele quantification procedures, sib-selections are cryopreserved, thus making the process easily interruptible. The quantification of HDR events after each sib-selection allows precise workload calculations for the next step and prevents tedious single-cell cloning even if no isogenic cells are present.

In conclusion, this represents a novel strategy for the scarless correction of heterozygous mutations by random enrichment of precisely edited cells and their detection via allele quantification. It was proposed that the frequency of isogenic cells can be determined by comparing WT:Mut allele ratios with two assumptions: no copy number variation and an error-free correction of the mutation via HDR. This approach can facilitate the generation of isogenic control cells, which represent the gold standard of controls when investigating the influence of putative disease-causing variants on the disease phenotype. Finally, a cardiac disease model for AF with isogenic controls was established, which will allow direct genotype-phenotype comparison to further elucidate the role of *SHOX2* in the genetic network of atrial and sinoatrial cardiomyocytes and in the development of the disease. Compared to already existing AF models [65,161], this would be the first one using patient-specific and gene-corrected iPSCs.

4. Outlook

Detailed analyses of the generated cell lines are currently ongoing. A strategy to generate pure SAN- and atrial-like cardiomyocyte populations is the use of fluorescent proteins with subtype-specific expression to identify and sort out cells of interest (**Figure 17A**). For this, lentiviral constructs were cloned and validated in this project; however, spontaneous or directed cardiac differentiation mainly produced ventricular-like cells and did not yield enough cardiomyocytes of the right subtype for molecular analysis. Yet, this approach can still be used to purify cells that are generated in other ways.

The central role in the SAN genetic network has also established *SHOX2* as a candidate for heterologous expression in transgene-based differentiation strategies. *SHOX2* overexpression in

embryonic stem cells reportedly favors the differentiation into cardiac pacemaker cells and can influence mesenchymal stem cell fate in co-culturing models with neonatal rat cardiomyocytes.[162,163] A newer approach uses combinations of key transcription factors (e.g. *SHOX2*, *HCN2*, *TBX5*) to drive the differentiation of cardiac progenitor cells into pacemaker-like cells with functional characteristics.[164] However, detrimental effects of the *SHOX2* mutations would most likely be mediated by disturbances in development or specification of cardiac cells. Therefore, manipulation of the genetic network by transgene expression during *in vitro* differentiation could mask a phenotype through bypassing or overcoming an impaired *SHOX2* function. In fact, the original plan of using reporter constructs to isolate SAN- and atrial-like cardiomyocytes from spontaneous differentiation was based on the idea of deriving cells that had not been substantially manipulated during their specification.

Monolayer-based cardiomyocyte differentiation by modulating Wnt/β-Catenin signaling yields mainly ventricular-like cells for disease-modelling (**Figure 17B**). Substantial effort has been undertaken in recent years to develop subtype-specific differentiation protocols for stem cells by investigating embryonic developmental pathways. A key finding towards generating atrial-like cardiomyocytes was the fact that retinoic acid signaling is essential for the specification into heart cells with high expression of atrial markers (*COUPTFII*, *NPPA*, *SLN*, *PITX2*) and characteristic action potentials.[165] An advanced version of this differentiation protocol [166] was recently optimized for these lines and yielded first insights into a dysregulated expression of *SHOX2* c.*28T>C and its target genes (**Figure 17C**) (for details see Master Thesis Viktoria Frajs, *SHOX2 in atrial fibrillation disease modelling using induced pluripotent stem cells*). More detailed molecular analyses are pending and electrophysiological profiling using voltage-sensitive fluorescent proteins will be performed. Regarding SAN morphogenesis, several signaling pathways including RA and BMP have been implicated in this process.[96] However, recent insights through lineage tracing in zebrafish provide a refined paradigm as to how canonical Wnt signaling can establish/influence pacemaker cell fate.[14,167] These findings have been successfully translated into modified differentiation approaches where mesoderm induction is achieved by Wnt signaling activation without subsequent strong Wnt inhibition, which normally defines the cardiac cell fate (**Figure 17D**). The resulting cardiomyocytes show increased expression of pacemaker marker genes (*Hcn4*, *Tbx18*, *Shox2*), as well as electrophysiological features and an intrinsic automaticity typical for SAN cells.[168] Currently ongoing pilot experiments with patient and control lines have yielded promising results showing the same impaired *SHOX2* c.*28T>C that had been observed in atrial-like cells. The combination of subtype-specific differentiation protocols with lentiviral reporter constructs is likely to produce pure cardiomyocyte populations for electrophysiological and molecular investigations. A detailed analysis of this *in vitro* disease model will aid in refining the paradigm of how *SHOX2* mutations influence the onset and progression of AF and even has the potential to serve as a platform for novel drug discovery and personalized medicine.

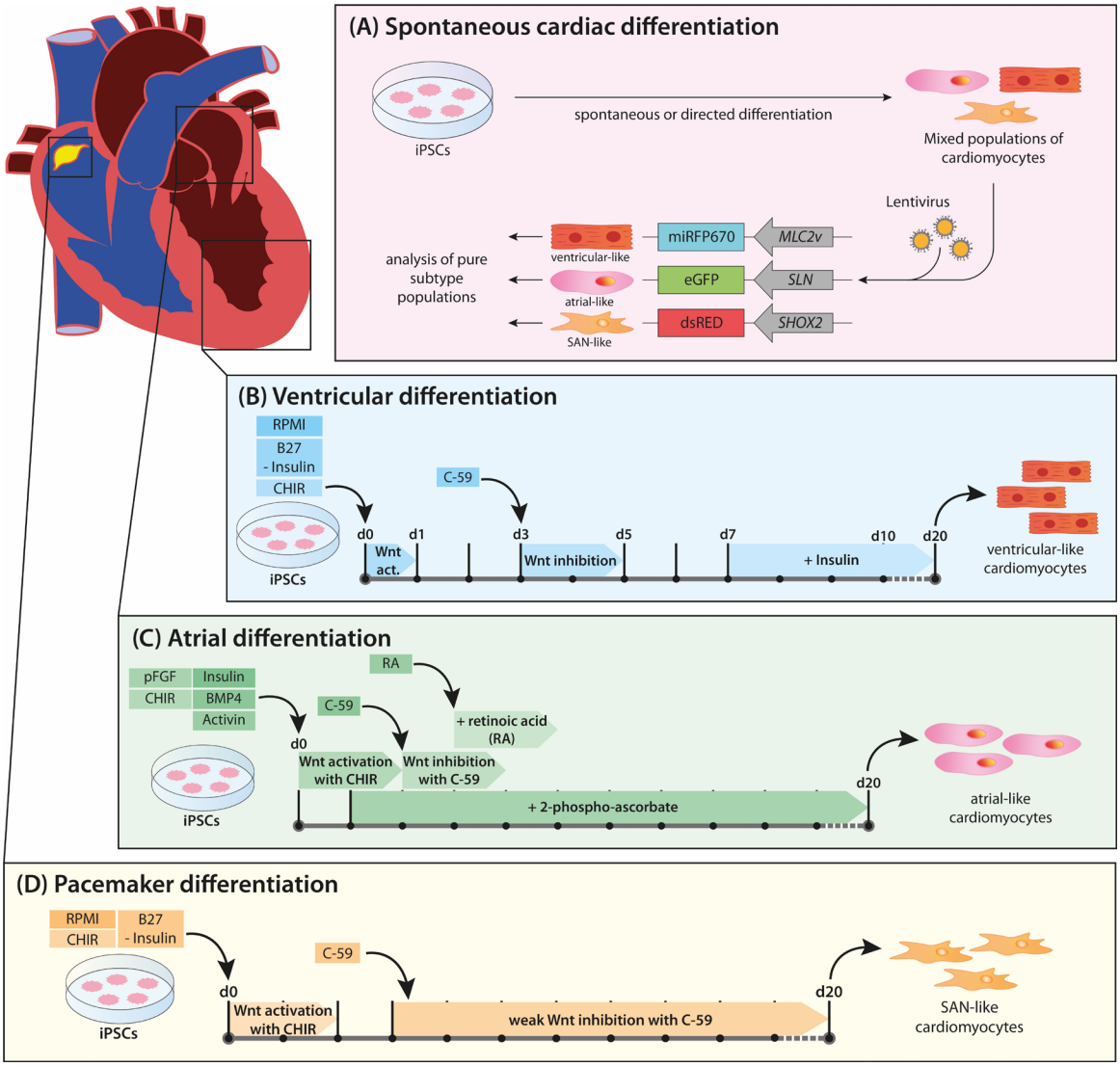


Figure 17 Generation of subtype-specific cardiomyocytes for molecular and electrophysiological analysis. (A) Labelling of CM subtypes with lentiviral constructs. iPSCs are differentiated spontaneously or with directed differentiation protocols to generate mixed populations with different CM subtypes. These cells are subsequently transduced with lentiviral constructs consisting of different fluorescent proteins (miRFP670, eGFP, dsRED) under the control of a subtype-specific promotors (*SHOX2* = CSS-specific promotor, *SLN* = atrial-specific promotor, *MLC2v* = ventricular-specific promotor). For electrophysiological analysis, voltage-sensitive fluorescent proteins can be used. Certain CM subtypes can be sorted by FACS to generate pure cell populations. This strategy is based on Chen et al., 2016. (B) Mesoderm induction by Wnt signaling activation followed by Wnt signaling inhibition for cardiac determination leads to mainly ventricular-like cardiomyocytes. (C) Pure atrial-like cardiomyocyte populations can be generated by Wnt signaling alteration and retinoic acid. This approach is based on Zhang et al. 2015, scheme modified from Frajs V, master thesis. (D) CMs with pacemaker properties can be generated by activating Wnt signaling to induce mesoderm differentiation, followed by weak inactivation of Wnt signaling to induce cardiac differentiation. This strategy is based on Liang et al. 2019

5. References

1. Hoffmann, S. *et al.* Functional Characterization of Rare Variants in the SHOX2 Gene Identified in Sinus Node Dysfunction and Atrial Fibrillation. *Front Genet* **10**, 648 (2019).
2. Burkhard, S., van Eif, V., Garric, L., Christoffels, V.M. & Bakkers, J. On the Evolution of the Cardiac Pacemaker. *J Cardiovasc Dev Dis* **4**(2017).
3. Yu, Y., Karbowski, J., Sachdev, R.N. & Feng, J. Effect of temperature and glia in brain size enlargement and origin of allometric body-brain size scaling in vertebrates. *BMC Evol Biol* **14**, 178 (2014).
4. Wells, M.J. The cephalopod heart: The evolution of a high-performance invertebrate pump. *Experientia* **48**, 800-808 (1992).
5. Porges, S.W. Love: an emergent property of the mammalian autonomic nervous system. *Psychoneuroendocrinology* **23**, 837-61 (1998).
6. Bishopric, N.H. Evolution of the heart from bacteria to man. *Ann N Y Acad Sci* **1047**, 13-29 (2005).
7. van Eif, V.W.W., Devalla, H.D., Boink, G.J.J. & Christoffels, V.M. Transcriptional regulation of the cardiac conduction system. *Nat Rev Cardiol* **15**, 617-630 (2018).
8. Mohan, R., Boukens, B.J. & Christoffels, V.M. Lineages of the Cardiac Conduction System. *J Cardiovasc Dev Dis* **4**(2017).
9. van Weerd, J.H. & Christoffels, V.M. The formation and function of the cardiac conduction system. *Development* **143**, 197-210 (2016).
10. van Eif, V.W.W. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells reveals a conserved genetic program. *Development* **146**(2019).
11. Choudhury, M., Boyett, M.R. & Morris, G.M. Biology of the Sinus Node and its Disease. *Arrhythm Electrophysiol Rev* **4**, 28-34 (2015).
12. Christoffels, V.M., Smits, G.J., Kispert, A. & Moorman, A.F. Development of the pacemaker tissues of the heart. *Circ Res* **106**, 240-54 (2010).
13. Garcia-Frigola, C., Shi, Y. & Evans, S.M. Expression of the hyperpolarization-activated cyclic nucleotide-gated cation channel HCN4 during mouse heart development. *Gene Expr Patterns* **3**, 777-83 (2003).
14. Ren, J. *et al.* Canonical Wnt5b Signaling Directs Outlying Nkx2.5+ Mesoderm into Pacemaker Cardiomyocytes. *Dev Cell* (2019).
15. Christoffels, V.M. *et al.* Formation of the venous pole of the heart from an Nkx2-5-negative precursor population requires Tbx18. *Circ Res* **98**, 1555-63 (2006).
16. Mommersteeg, M.T. *et al.* Pitx2c and Nkx2-5 are required for the formation and identity of the pulmonary myocardium. *Circ Res* **101**, 902-9 (2007).
17. Wang, J. *et al.* Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc Natl Acad Sci U S A* **107**, 9753-8 (2010).
18. Wang, J. *et al.* Pitx2-microRNA pathway that delimits sinoatrial node development and inhibits predisposition to atrial fibrillation. *Proc Natl Acad Sci U S A* **111**, 9181-6 (2014).
19. Puskaric, S. *et al.* Shox2 mediates Tbx5 activity by regulating Bmp4 in the pacemaker region of the developing heart. *Hum Mol Genet* **19**, 4625-33 (2010).
20. Mori, A.D. *et al.* Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev Biol* **297**, 566-86 (2006).
21. Ye, W. *et al.* A common Shox2-Nkx2-5 antagonistic mechanism primes the pacemaker cell fate in the pulmonary vein myocardium and sinoatrial node. *Development* **142**, 2521-32 (2015).
22. Hoffmann, S. *et al.* Islet1 is a direct transcriptional target of the homeodomain transcription factor Shox2 and rescues the Shox2-mediated bradycardia. *Basic Res Cardiol* **108**, 339 (2013).
23. Hoogaars, W.M. *et al.* Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. *Genes Dev* **21**, 1098-112 (2007).

24. Lippi, G., Sanchis-Gomar, F. & Cervellin, G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. *Int J Stroke*, 1747493019897870 (2020).
25. Wijesurendra, R.S. & Casadei, B. Mechanisms of atrial fibrillation. *Heart* **105**, 1860-1867 (2019).
26. Feghaly, J., Zakka, P., London, B., MacRae, C.A. & Refaat, M.M. Genetics of Atrial Fibrillation. *J Am Heart Assoc* **7**, e009884 (2018).
27. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* **50**, 1225-1233 (2018).
28. Hoffmann, S. *et al.* Coding and non-coding variants in the SHOX2 gene in patients with early-onset atrial fibrillation. *Basic Res Cardiol* **111**, 36 (2016).
29. Ma, J.F. *et al.* TBX5 mutations contribute to early-onset atrial fibrillation in Chinese and Caucasians. *Cardiovasc Res* **109**, 442-50 (2016).
30. Yu, H. *et al.* Mutational spectrum of the NKX2-5 gene in patients with lone atrial fibrillation. *Int J Med Sci* **11**, 554-63 (2014).
31. De Ponti, R., Marazzato, J., Bagliani, G., Leonelli, F.M. & Padeletti, L. Sick Sinus Syndrome. *Card Electrophysiol Clin* **10**, 183-195 (2018).
32. Sanders, P. *et al.* Electrophysiological and electroanatomic characterization of the atria in sinus node disease: evidence of diffuse atrial remodeling. *Circulation* **109**, 1514-22 (2004).
33. Jensen, P.N. *et al.* Incidence of and risk factors for sick sinus syndrome in the general population. *J Am Coll Cardiol* **64**, 531-8 (2014).
34. Kezerashvili, A., Krumerman, A.K. & Fisher, J.D. Sinus Node Dysfunction in Atrial Fibrillation: Cause or Effect? *J Atr Fibrillation* **1**, 30 (2008).
35. John, R.M. & Kumar, S. Sinus Node and Atrial Arrhythmias. *Circulation* **133**, 1892-900 (2016).
36. Hirsh, B.J., Copeland-Halperin, R.S. & Halperin, J.L. Fibrotic atrial cardiomyopathy, atrial fibrillation, and thromboembolism: mechanistic links and clinical inferences. *J Am Coll Cardiol* **65**, 2239-51 (2015).
37. Iwasaki, Y.K., Nishida, K., Kato, T. & Nattel, S. Atrial fibrillation pathophysiology: implications for management. *Circulation* **124**, 2264-74 (2011).
38. Milanesi, R., Baruscotti, M., Gnecchi-Ruscone, T. & DiFrancesco, D. Familial sinus bradycardia associated with a mutation in the cardiac pacemaker channel. *N Engl J Med* **354**, 151-7 (2006).
39. Nof, E. *et al.* Point mutation in the HCN4 cardiac ion channel pore affecting synthesis, trafficking, and functional expression is associated with familial asymptomatic sinus bradycardia. *Circulation* **116**, 463-70 (2007).
40. Duhme, N. *et al.* Altered HCN4 channel C-linker interaction is associated with familial tachycardia-bradycardia syndrome and atrial fibrillation. *Eur Heart J* **34**, 2768-75 (2013).
41. Ruan, Y., Liu, N. & Priori, S.G. Sodium channel mutations and arrhythmias. *Nat Rev Cardiol* **6**, 337-48 (2009).
42. Butters, T.D. *et al.* Mechanistic links between Na+ channel (SCN5A) mutations and impaired cardiac pacemaking in sick sinus syndrome. *Circ Res* **107**, 126-37 (2010).
43. Jackson, L.R., 2nd *et al.* Sinus Node Dysfunction and Atrial Fibrillation: A Reversible Phenomenon? *Pacing Clin Electrophysiol* **40**, 442-450 (2017).
44. Liu, H. *et al.* The role of Shox2 in SAN development and function. *Pediatr Cardiol* **33**, 882-9 (2012).
45. Semina, E.V. *et al.* Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nat Genet* **14**, 392-9 (1996).
46. Belin, V. *et al.* SHOX mutations in dyschondrosteosis (Leri-Weill syndrome). *Nat Genet* **19**, 67-9 (1998).
47. Blaschke, R.J. *et al.* SHOT, a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development. *Proc Natl Acad Sci U S A* **95**, 2406-11 (1998).
48. Blaschke, R.J. *et al.* Targeted mutation reveals essential functions of the homeodomain transcription factor Shox2 in sinoatrial and pacemaking development. *Circulation* **115**, 1830-8 (2007).

49. Espinoza-Lewis, R.A. *et al.* Shox2 is essential for the differentiation of cardiac pacemaker cells by repressing Nkx2-5. *Dev Biol* **327**, 376-85 (2009).
50. Li, N. *et al.* A SHOX2 loss-of-function mutation underlying familial atrial fibrillation. *Int J Med Sci* **15**, 1564-1572 (2018).
51. Moretti, A., Laugwitz, K.L., Dorn, T., Sinnecker, D. & Mummery, C. Pluripotent stem cell models of human heart disease. *Cold Spring Harb Perspect Med* **3**(2013).
52. Brandao, K.O., Tabel, V.A., Atsma, D.E., Mummery, C.L. & Davis, R.P. Human pluripotent stem cell models of cardiac disease: from mechanisms to therapies. *Dis Model Mech* **10**, 1039-1059 (2017).
53. Abou-Saleh, H. *et al.* The march of pluripotent stem cells in cardiovascular regenerative medicine. *Stem Cell Res Ther* **9**, 201 (2018).
54. van Mil, A. *et al.* Modelling inherited cardiac disease using human induced pluripotent stem cell-derived cardiomyocytes: progress, pitfalls, and potential. *Cardiovasc Res* **114**, 1828-1842 (2018).
55. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-72 (2007).
56. Moretti, A. *et al.* Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med* **363**, 1397-409 (2010).
57. Itzhaki, I. *et al.* Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* **471**, 225-9 (2011).
58. Ma, D. *et al.* Modeling type 3 long QT syndrome with cardiomyocytes derived from patient-specific induced pluripotent stem cells. *Int J Cardiol* **168**, 5277-86 (2013).
59. Sun, N. *et al.* Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Sci Transl Med* **4**, 130ra47 (2012).
60. Lan, F. *et al.* Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. *Cell Stem Cell* **12**, 101-13 (2013).
61. Kim, C. *et al.* Studying arrhythmogenic right ventricular dysplasia with patient-specific iPSCs. *Nature* **494**, 105-10 (2013).
62. Wang, G. *et al.* Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat Med* **20**, 616-23 (2014).
63. Moretti, A. *et al.* Somatic gene editing ameliorates skeletal and cardiac muscle failure in pig and human models of Duchenne muscular dystrophy. *Nat Med* (2020).
64. Laksman, Z. *et al.* Modeling Atrial Fibrillation using Human Embryonic Stem Cell-Derived Atrial Tissue. *Sci Rep* **7**, 5268 (2017).
65. Benzoni, P. *et al.* Human iPSC modeling of a familial form of atrial fibrillation reveals a gain of function of If and ICaL in patient-derived cardiomyocytes. *Cardiovasc Res* (2019).
66. Brookhouser, N., Raman, S., Potts, C. & Brafman, D.A. May I Cut in? Gene Editing Approaches in Human Induced Pluripotent Stem Cells. *Cells* **6**(2017).
67. Hockemeyer, D. & Jaenisch, R. Induced Pluripotent Stem Cells Meet Genome Editing. *Cell Stem Cell* **18**, 573-86 (2016).
68. Carroll, D. Genome engineering with targetable nucleases. *Annu Rev Biochem* **83**, 409-39 (2014).
69. Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-51 (2005).
70. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-78 (2014).
71. Jackson, S.P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-8 (2009).
72. Seah, Y.F., El Farran, C.A., Warrier, T., Xu, J. & Loh, Y.H. Induced Pluripotency and Gene Editing in Disease Modelling: Perspectives and Challenges. *Int J Mol Sci* **16**, 28614-34 (2015).
73. Segal, D.J., Crotty, J.W., Bhakta, M.S., Barbas, C.F., 3rd & Horton, N.C. Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol* **363**, 405-21 (2006).

74. Mani, M., Smith, J., Kandavelou, K., Berg, J.M. & Chandrasegaran, S. Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. *Biochem Biophys Res Commun* **334**, 1191-1197 (2005).
75. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
76. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-8 (2011).
77. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-53 (2011).
78. Ru, R. *et al.* Targeted genome engineering in human induced pluripotent stem cells by penetrating TALENs. *Cell Regen (Lond)* **2**, 5 (2013).
79. Zou, J. *et al.* Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97-110 (2009).
80. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**, 731-4 (2011).
81. Collin, J. & Lako, M. Concise review: putting a finger on stem cell biology: zinc finger nuclease-driven targeted genetic editing in human pluripotent stem cells. *Stem Cells* **29**, 1021-33 (2011).
82. Ramirez, C.L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5**, 374-5 (2008).
83. Isalan, M. Zinc-finger nucleases: how to play two good hands. *Nat Methods* **9**, 32-4 (2011).
84. Cox, D.B., Platt, R.J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat Med* **21**, 121-31 (2015).
85. Briggs, A.W. *et al.* Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* **40**, e117 (2012).
86. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-12 (2007).
87. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-21 (2012).
88. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
89. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-2 (2013).
90. Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393-4 (2013).
91. Deleidi, M. & Yu, C. Genome editing in pluripotent stem cells: research and therapeutic applications. *Biochem Biophys Res Commun* **473**, 665-74 (2016).
92. Young, C.S. *et al.* A Single CRISPR-Cas9 Deletion Strategy that Targets the Majority of DMD Patients Restores Dystrophin Function in hiPSC-Derived Muscle Cells. *Cell Stem Cell* **18**, 533-40 (2016).
93. Ye, L. *et al.* Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proc Natl Acad Sci U S A* **111**, 9591-6 (2014).
94. Li, S. *et al.* Human Induced Pluripotent Stem Cell NEUROG2 Dual Knockin Reporter Lines Generated by the CRISPR/Cas9 System. *Stem Cells Dev* **24**, 2925-42 (2015).
95. Funakoshi, S. *et al.* Enhanced engraftment, proliferation, and therapeutic potential in heart using optimized human iPSC-derived cardiomyocytes. *Sci Rep* **6**, 19111 (2016).
96. Protze, S.I. *et al.* Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. *Nat Biotechnol* **35**, 56-68 (2017).
97. Zhang, J.Z. *et al.* A Human iPSC Double-Reporter System Enables Purification of Cardiac Lineage Subpopulations with Distinct Function and Drug Response Profiles. *Cell Stem Cell* **24**, 802-811 e5 (2019).
98. Jang, Y.Y. & Ye, Z. Gene correction in patient-specific iPSCs for therapy development and disease modeling. *Hum Genet* **135**, 1041-58 (2016).

99. Guo, Q. *et al.* 'Cold shock' increases the frequency of homology directed repair gene editing in induced pluripotent stem cells. *Sci Rep* **8**, 2080 (2018).
100. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**, 424-9 (2011).
101. Soldner, F. *et al.* Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* **146**, 318-31 (2011).
102. Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat Methods* **11**, 291-3 (2014).
103. Yu, C. *et al.* Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142-7 (2015).
104. Chu, V.T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-8 (2015).
105. Heyer, W.D., Ehmsen, K.T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**, 113-39 (2010).
106. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep* **14**, 1555-1566 (2016).
107. Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
108. Okamoto, S., Amaishi, Y., Maki, I., Enoki, T. & Mineno, J. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Sci Rep* **9**, 4811 (2019).
109. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-44 (2016).
110. Zhang, J.P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol* **18**, 35 (2017).
111. Bothmer, A. *et al.* Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* **8**, 13905 (2017).
112. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-4 (2016).
113. Anzalone, A.V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157 (2019).
114. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. **23**, 1289-1291 (2007).
115. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).
116. Westerfield, M. *THE ZEBRAFISH BOOK: A guide for the laboratory use of zebrafish (Danio rerio)*, (University of Oregon Press, Eugene, 2007).
117. Castro, F. *et al.* High-throughput SNP-based authentication of human cell lines. *Int J Cancer* **132**, 308-14 (2013).
118. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J.L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* **10**, e0124633 (2015).
119. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174 (2019).
120. Park, J., Lim, K., Kim, J.S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286-288 (2017).
121. Liu, W. *et al.* IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**, 3359-61 (2015).
122. Wild, P.S. *et al.* [The Gutenberg Health Study]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **55**, 824-9 (2012).

123. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
124. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
125. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
126. Ikeda, K. *et al.* Efficient scarless genome editing in human pluripotent stem cells. *Nat Methods* **15**, 1045-1047 (2018).
127. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. & Jasin, M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol* **18**, 93-101 (1998).
128. Sentmanat, M.F., Peters, S.T., Florian, C.P., Connelly, J.P. & Pruett-Miller, S.M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep* **8**, 888 (2018).
129. Taylor, S.C., Laperriere, G. & Germain, H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Sci Rep* **7**, 2409 (2017).
130. Mazaika, E. & Homsy, J. Digital Droplet PCR: CNV Analysis and Other Applications. *Curr Protoc Hum Genet* **82**, 7 24 1-13 (2014).
131. Ding, Q. *et al.* A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* **12**, 238-51 (2013).
132. Mahida, S. Transcription factors and atrial fibrillation. *Cardiovasc Res* **101**, 194-202 (2014).
133. Seidman, J.G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* **109**, 451-5 (2002).
134. Ishikawa, T. *et al.* Sick sinus syndrome with HCN4 mutations shows early onset and frequent association with atrial fibrillation and left ventricular noncompaction. *Heart Rhythm* **14**, 717-724 (2017).
135. Thibodeau, I.L. *et al.* Paradigm of genetic mosaicism and lone atrial fibrillation: physiological characterization of a connexin 43-deletion mutant identified from atrial tissue. *Circulation* **122**, 236-44 (2010).
136. Gollob, M.H. *et al.* Somatic mutations in the connexin 40 gene (GJA5) in atrial fibrillation. *N Engl J Med* **354**, 2677-88 (2006).
137. Milanesi, R., Bucchi, A. & Baruscotti, M. The genetic basis for inherited forms of sinoatrial dysfunction and atrioventricular node dysfunction. *J Interv Card Electrophysiol* **43**, 121-34 (2015).
138. Aza-Carmona, M. *et al.* NPPB and ACAN, two novel SHOX2 transcription targets implicated in skeletal development. *PLoS One* **9**, e83104 (2014).
139. Liu, H. *et al.* Functional redundancy between human SHOX and mouse Shox2 genes in the regulation of sinoatrial node formation and pacemaking function. *J Biol Chem* **286**, 17029-38 (2011).
140. Espinoza-Lewis, R.A. *et al.* Ectopic expression of Nkx2.5 suppresses the formation of the sinoatrial node in mice. *Dev Biol* **356**, 359-69 (2011).
141. Christophersen, I.E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet* **49**, 946-952 (2017).
142. Olesen, M.S. *et al.* Genetic loci on chromosomes 4q25, 7p31, and 12p12 are associated with onset of lone atrial fibrillation before the age of 40 years. *Can J Cardiol* **28**, 191-5 (2012).
143. Viereck, J. & Thum, T. Long Noncoding RNAs in Pathological Cardiac Remodeling. *Circ Res* **120**, 262-264 (2017).
144. Huang, R.T., Xue, S., Xu, Y.J., Zhou, M. & Yang, Y.Q. A novel NKX2.5 loss-of-function mutation responsible for familial atrial fibrillation. *Int J Mol Med* **31**, 1119-26 (2013).
145. Xie, W.H. *et al.* Prevalence and spectrum of Nkx2.5 mutations associated with idiopathic atrial fibrillation. *Clinics (Sao Paulo)* **68**, 777-84 (2013).

146. Lubitz, S.A. *et al.* Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA* **304**, 2263-9 (2010).
147. Mitzelfelt, K.A. *et al.* Efficient Precision Genome Editing in iPSCs via Genetic Co-targeting with Selection. *Stem Cell Reports* **8**, 491-499 (2017).
148. Steyer, B. *et al.* Scarless Genome Editing of Human Pluripotent Stem Cells via Transient Puromycin Selection. *Stem Cell Reports* **10**, 642-654 (2018).
149. Lonowski, L.A. *et al.* Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat Protoc* **12**, 581-603 (2017).
150. Chen, Y. *et al.* A Self-restricted CRISPR System to Reduce Off-target Effects. *Mol Ther* **24**, 1508-10 (2016).
151. Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S. & Yang, S.H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
152. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012-9 (2014).
153. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res* **24**, 1020-7 (2014).
154. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-32 (2013).
155. Chang, Y.J. *et al.* CRISPR Base Editing in Induced Pluripotent Stem Cells. *Methods Mol Biol* **2045**, 337-346 (2019).
156. Idoko-Akoh, A., Taylor, L., Sang, H.M. & McGrew, M.J. High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells. *Sci Rep* **8**, 15126 (2018).
157. Park, S.T. & Kim, J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *Int Neurourol J* **20**, S76-83 (2016).
158. Levy, S.E. & Myers, R.M. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet* **17**, 95-115 (2016).
159. Martins-Taylor, K. & Xu, R.H. Concise review: Genomic stability of human induced pluripotent stem cells. *Stem Cells* **30**, 22-7 (2012).
160. Merkle, F.T. & Eggan, K. Modeling human disease with pluripotent stem cells: from genome association to function. *Cell Stem Cell* **12**, 656-68 (2013).
161. Marczenke, M. *et al.* Generation and cardiac subtype-specific differentiation of PITX2-deficient human iPS cell lines for exploring familial atrial fibrillation. *Stem Cell Res* **21**, 26-28 (2017).
162. Ionta, V. *et al.* SHOX2 overexpression favors differentiation of embryonic stem cells into cardiac pacemaker cells, improving biological pacing ability. *Stem Cell Reports* **4**, 129-142 (2015).
163. Feng, Y. *et al.* Shox2 influences mesenchymal stem cell fate in a co-culture model in vitro. *Mol Med Rep* **14**, 637-42 (2016).
164. Raghunathan, S. *et al.* Conversion of human cardiac progenitor cells into cardiac pacemaker-like cells. *J Mol Cell Cardiol* **138**, 12-22 (2020).
165. Devalla, H.D. *et al.* Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Mol Med* **7**, 394-410 (2015).
166. Zhang, M. *et al.* Universal cardiac induction of human pluripotent stem cells in two and three-dimensional formats: implications for in vitro maturation. *Stem Cells* **33**, 1456-69 (2015).
167. Burns, C.G. & Burns, C.E. Canonical Wnt Signaling Sets the Pace. *Dev Cell* **50**, 675-676 (2019).
168. Liang, W. *et al.* Canonical Wnt signaling promotes pacemaker cell specification of cardiac mesodermal cells derived from mouse and human embryonic stem cells. *Stem Cells* **38**, 352-368 (2020).

6. Appendix

Variable	All (450)	Male (311)	Female (139)
Sex (Female)	30.9% (139/450)	0% (0/311)	100.0% (139/139)
Age at study enrollment [y]	64.4 (8.4)	64.5 (8.4)	64.2 (8.7)
CVRFs:			
Diabetes (yes)	22.4% (101/450)	23.5% (73/311)	20.1% (28/139)
Obesity (yes)	37.1% (167/450)	35.0% (109/311)	41.7% (58/139)
Smoking (yes)	13.6% (61/448)	13.9% (43/310)	13.0% (18/138)
Hypertension (yes)	71.6% (322/450)	71.7% (223/311)	71.2% (99/139)
Dyslipidemia (yes)	69.8% (314/450)	73.6% (229/311)	61.2% (85/139)
FH of MI/Stroke (yes)	24.9% (112/450)	23.5% (73/311)	28.1% (39/139)
Comorbidities:			
MI (yes)	15.1% (67/443)	16.3% (50/306)	12.4% (17/137)
CAD (yes)	23.0% (99/430)	24.9% (75/301)	18.6% (24/129)
Stroke (yes)	8.5% (38/446)	9.8% (30/307)	5.8% (8/139)
AF (yes)	100.0% (450/450)	100.0% (311/311)	100.0% (139/139)
PAD (yes)	8.1% (36/443)	8.2% (25/306)	8.0% (11/137)
CHF (yes)	12.4% (56/450)	10.0% (31/311)	18.0% (25/139)
Echo:			
LA Volume (Biplane) [cm³]	89.2 (34.1)	95.3 (34.8)	76.1 (28.5)
EF [%]	60.3 (8.6)	59.9 (8.7)	61.3 (8.4)
E/E´	8.95 (3.96)	8.65 (3.93)	9.62 (3.94)
ECG:			
RR Interval [ms]	936 (220)	926 (229)	957 (197)
PQ Interval [ms]	170 (28)	172 (29)	165 (27)
QRS Duration	102.6 (20.3)	104.6 (20.3)	98.0 (19.6)
QT Interval	417 (46)	414 (46)	424 (43)

Table S1 Clinical characteristics of AF cohort. Data are presented as mean ±SD. Table and legend taken from Hoffmann et al., 2019.[1]

Variable	All (98)	Male (62)	Female (36)
Sex (Female)	37% (36/98)	0% (0/62)	100% (36/36)
Age at study enrollment [y]	75.7 (10.9)	76.1 (9.6)	75.0 (12.9)
Age at diagnosis [y]	71.0 (13.4)	71.3 (10.9)	70.5 (16.4)
CVRFs:			
Diabetes (yes)	27.6% (27/98)	25.8% (16/62)	30.6% (11/36)
Obesity (yes)	n.a.	n.a.	n.a.
Smoking (yes)	20.4% (20/98)	29.0% (18/62)	5.6% (2/36)
Hypertension (yes)	94.9 % (93/98)	96.8% (60/62)	91.7% (33/36)
Dyslipidemia (yes)	56.1% (55/98)	64.5% (40/62)	41.7% (15/36)
FH of MI/Stroke (yes)	15.3% (15/98)	19.4% (12/62)	8.3% (3/36)
Comorbidities:			
MI (yes)	21.4% (21/98)	24.2% (15/62)	16.7% (6/36)
CAD (yes)	54.1% (53/98)	62.9% (39/62)	38.9% (14/36)
Stroke (yes)	14.3% (14/98)	12.9% (8/62)	16.7% (6/36)
AF (yes)	68.4% (67/98)	59.7% (37/62)	83.3% (30/36)
PAD (yes)	14.3% (14/98)	19.4% (12/62)	5.6% (2/36)
CHF (yes)	33.7% (33/98)	38.7% (24/62)	25.0% (9/36)
AF (yes)	61.2% (60/98)	54.8% (34/62)	72.2% (26/36)
Echo:			
LA Diameter [mm]	45.3 (10.7)	46.0 (10.7)	44.3 (10.9)
EF [%]	54.6% (13.7)	54.3% (15.0)	55.0% (11.0)
E/E´	n.a.	n.a.	n.a.
ECG:			
RR Interval [ms]	919.4 (215.1)	928.2 (194.6)	903.8 (249.6)
PQ Interval [ms]	190.0 (44.4)	190.4 (47.5)	189.1 (38.0)
QRS Duration	122.4 (38.5)	124.4 (37.5)	118.9 (40.6)
QT_c Interval	438.3 (46.7)	440.3 (46.1)	434.9 (48.5)

Table S2 Clinical characteristics of SND cohort. Data are presented as mean ±SD. n.a. = not available. Table and legend taken from Hoffmann et al., 2019.[1]

SHOX2 variant	Sex	Age [y]	LA Volume (Biplane) [cm³]	EF [%]	E/E´	RR Interval [ms]	PQ Interval [ms]	QRS Duration	QT_c Interval	Patient cohort
G77D	Male	52	89	58	6	490	n.a.	100	376	AF
L129=	Male	67	102	47	7	682	n.a.	138	398	AF
L129=	Female	70	96	65	5	1088	140	80	424	AF
L129=	Male	74	78	56	11	640	n.a.	86	330	AF
L130F	Male	60	76	68	5	872	138	90	408	AF
A293=	Male	72	104	67	6	880	190	102	422	AF
P33R	Female	88	n.a.	65	n.a.	968	n.a.	78	405	SND
G81E	Female	51	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	AF
H283Q	Male	56	43	n.a.	n.a.	400	n.a.	n.a.	480	AF
R194X	Male	41	n.a.	62	n.a.	n.a.	>200	100	447	AF
R194X	Female	43	n.a.	60	n.a.	n.a.	>200	116	528	AF
R194X	Male	39	n.a.	63	n.a.	n.a.	>200	102	450	AF
R194X	Female	20	n.a.	65	n.a.	n.a.	>200	96	435	AF

Table S3 Summary of all identified SHOX2 variants in SND and AF patients. Novel identified variants and phenotypic features are highlighted in blue, the previously identified variants are highlighted in grey. n.a. = not available. Data are presented as mean ±SD. Blue: age at study enrollment; light blue: age at diagnosis; grey: age at diagnosis. Table and legend taken from Hoffmann et al., 2019.[1]

#ID	RGEN Treated Sequence	Count	Type	Allele
1	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	3225	WT	WT
2	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	2559	Mut	Mut
3	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCAC----------CACCTGGCC	1206	del	WT
4	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCATGTGGCGCATCCGCACCTGGCC	340	Ins	WT
5	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	323	Sub	WT
6	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCG---CACCACCTGCATCCGCACCTGGCC	307	del	Mut
7	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	272	Sub	Mut
8	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	171	Ins	WT
9	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCAC----------CC	136	del	Mut
10	GCTGCAGCTGGACAGCGCTG----------CCTGCATCCGCACCTGGCC	109	del	N/A
11	GCTGCAGCTGGACAGCGCTGTGG----------CCGCACCTGGCC	100	del	N/A
12	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	96	Sub	WT
13	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCG------CACCTGCATCCGCACCTGGCC	86	del	WT
14	GCTGCAGCTGGACAGCGCTGTGGCG----------CACCTGCATCCGCACCTGGCC	81	del	N/A
15	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	77	Sub	WT
16	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACC-----CATCCGCACCTGGCC	69	del	WT
17	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	63	Sub	Mut
18	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	62	Sub	Mut
19	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	54	Sub	WT
20	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	54	Sub	Mut
21	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCG---CACCACCTGCATCCGCACCTGGCC	33	del	Mut

Table S4. Sequences of allele-quantification in sib-selection *SHOX2* c.849C>A gRNA-2 #1

#ID	RGEN Treated Sequence	Count	Type	Allele
1	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	1118	Mut	Mut
2	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCC	1014	WT	WT
3	GCTGCAGCTGGACAGCGCTGTGGCGCAAG----------CGCACCTGGCC	616	del	Mut
4	GCTGCAGCTGGACAGCGCTGTGGCGCACGCG--------CCTGCATCCGCACCTGGCC	387	del	WT
5	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACC----------	342	del	WT
6	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCG------CACCTGCATCCGCACCTGGCC	260	del	Mut
7	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCAC----------CACCTGGCC	255	del	WT
8	GCT----------GCACCTGGCC	217	del	N/A
9	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCAC------CATCCGCACCTGGCC	166	del	Mut
10	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCA----CATCCGCACCTGGCC	158	del	Mut
11	GCTGCAGCTGGACAGCGCTGTGGCGCACGCG------CACCTGCATCCGCACCTGGCC	128	del	WT
12	GCTGCAGCTGGACAGCGCTGTGGCGCA----------CACCTGGCC	70	del	N/A
13	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCCTGCATCCGCACCTGGCC	46	Ins	Mut
14	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCATGCACCTGCATCCGCACCTGGC	41	Ins	WT
15	GCTGCAGCTGGACAGCGCTGTGGCG----------CACCTGCATCCGCACCTGGCC	40	del	N/A
16	GCTGCAGCTGGACAGCGCTGTGG----------	38	del	N/A
17	GCTGCAGCTGGACAGCGCTGT-GCG----------CGCACCTGGCC	37	del	N/A
18	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCTACATCCTGCATCCGCACCTACACCT	33	Ins	WT
19	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCA-CAC----------CCTGGCC	31	del	Mut
20	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCATCCGCACCTGCACCTGGCC	30	Sub	Mut
21	GCTGCAGCTGGACAGCG----------CACCTGCATCCGCACCTGGCC	27	del	N/A
22	----------CCTGGCC	27	del	N/A
23	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCAC----------	19	del	Mut
24	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCA--TG-AT--GGGTCTGGCC	18	del	WT

Table S5. Sequences of allele-quantification in sib-selection *SHOX2* c.849C>A gRNA-2 #2

#ID	RGEN Treated Sequence	Count	Type	Allele
1	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	1122	WT	WT
2	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCACCTGCATCCGCACCTGGCC	795	Mut	Mut
3	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCG---CACCACCTGCATCCGCACCTGGCC	540	del	WT
4	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACC----CCTGCATCCGCACCTGGCC	280	del	WT
5	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCAC----------CACCTGGCC	273	del	WT
6	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCG------CACCTGCATCCGCACCTGGCC	234	del	Mut
7	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCAC------CATCCGCACCTGGCC	215	del	Mut
8	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCAACCTGCATCCGCACCTGGCC	211	Ins	Mut
9	GCTGCAGCTGGACAGCGCTG--------------------------------TGGCC	211	del	N/A
10	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCAC----------CCTGGCC	183	del	WT
11	GCTGCAGCTG----------**C**CTGCATCCGCACCTGGCC	158	del	WT
12	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCG---CACCACCTGCATCCGCACCTGGCC	139	del	Mut
13	GCTGCAGCTGGACAGCGCTGTGGCGCA----------AGCATCCGCACCTGGCC	137	del	N/A
14	GCTGCAGCTGGACAGCGCTGTGGCG--------------------------------	121	del	N/A
15	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCAC----------CACCTGGCC	114	del	WT
16	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCAAGCGCACCTGCATCCGCACCT	106	Ins	Mut
17	GCTGCAGCTGGA----------CATCCGCACCTGGCC	104	del	N/A
18	GCTGCAGCTGGACAGCGCTGT----------CCTGCATCCGCACCTGGCC	98	del	N/A
19	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCA----------CACCTGGCC	96	del	Mut
20	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCAC----------CACCTGGCC	96	del	Mut
21	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACC-CCTGCATCCGCACCTGGCC	94	del	Mut
22	GCTGCAGCTGGACAGCGCTGTGGC----------GCATCCGCACCTGGCC	86	del	N/A
23	GCTGCAG----------CTGGCC	62	del	N/A
24	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCAC------CATCCGCACCTGGCC	57	del	WT
25	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCATGCAGGTGCCTGCATCCGCAC	44	Ins	Mut
26	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCG------CACCTGCATCCGCACCTGGCC	32	del	WT
27	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCATCCGCACCTGCATCCGCACCTG	31	Ins	WT
28	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCACCA---------GCACCTGGCC	29	del	Mut
29	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCCTGCATCCGCACCTGGCC	28	Ins	WT
30	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCAGGCGCATCCGCACCTGGCC	27	Sub	WT
31	GCTGCAGCTGG----------**A**CCTGCATCCGCACCTGGCC	19	del	Mut

Table S6. Sequences of allele-quantification in sib-selection *SHOX2* c.849C>A gRNA-2 #3

#ID	RGEN Treated Sequence	Count	Type	Allele
1	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	730	Mut	Mut
2	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCC	630	WT	WT
3	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCAC------CATCCGCACCTGGCC	197	del	Mut
4	GCTGCAGCTGGACAGCGCTGTGGCGCACGCG---CACCACCTGCATCCGCACCTGGCC	175	del	WT
5	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCG---CACCACCTGCATCCGCACCTGGCC	172	del	Mut
6	GCTGCAGCTGGACA----------GGCC	155	del	N/A
7	GCTGCAGCTGGACAGCGCTGTGGC----------GCATCCGCACCTGGCC	153	del	N/A
8	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACC-CCTGCATCCGCACCTGGCC	137	del	WT
9	GCTGCAGCTGGACAGCGCTG----------TGGCC	126	del	N/A
10	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCA---GCATCCGCACCTGGCC	86	del	Mut
11	GCTGCAGCTGGACAGCGCTGTGGCGCACGCG------CACCTGCATCCGCACCTGGCC	80	del	WT
12	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCCTGCATCCGCACCTGGCC	72	Ins	WT
13	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	56	Mut	Mut
14	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCAC------CATCCGCACCTGGCC	50	del	WT
15	GCTGCAGCTGGACAGCGCTGT----------CCTGCATCCGCACCTGGCC	48	del	N/A
16	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCAC----------CACCTGGCC	47	del	WT
17	GCTGCAGCTGGACAGCGCTGTGGCGCA----------CCGCACCTGGCC	40	del	N/A
18	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCA----AATCCGCACCTGGCC	39	del	WT
19	GCTGCAGCTGGACAGCGCTGTGGCGCACG----------CGCACCTGGCC	36	del	WT
20	GCTGCAGCTGGACAGCGCTGTGGCGCAAG----------CGCACCTGGCC	34	del	Mut
21	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCTGCATCCGCCTGCATCCGCACCTGGCC	33	Ins	WT
22	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	32	Mut	Mut
23	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	31	Mut	Mut
24	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCC	30	WT	WT
25	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACCACCTGCATCCGCACCTGGCC	29	WT	WT
26	GCTGCAGCTGGACAGCGCTGTG----------GCGCACGCGCC	27	del	Mut
27	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCAC----------CACCTGGCC	25	del	Mut
28	GATAGTCATTGCAACGTGACGCCCTTTTCCTTTCAGGTTCAGGCGCAGCTGCAGCTGGA	22	Sub	N/A
29	GCTGCAGCTGGACAGCGCTGTGGCGCACGCGCACCACATGATGCATCCGCACCTGGCC	22	WT	WT
30	GCTGCAGCTGGACAGCGCTGTGGCG----------CTGGCC	22	del	N/A
31	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCAC----------CACCTGGCC	21	del	Mut
32	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCA----------CGCACCTGGCC	21	del	Mut
33	GCTGCAGCTGGACAGCGCTGTGGCGCAAGCGCACCACCACCTGCATCCGCACCTGGCC	20	Mut	Mut

34	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	19	WT	WT
35	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCGCACCACCACCTGCATCCGCACCTGGCC	19	WT	WT
36	GCTGCAGCTGGACAGCGCTGTGGCGCA**C**GCG---CACCACCTGCATCCGCACCTGGCC	19	del	WT
37	GCTGCAGCTGGACAGCGCTGTGGCGCA----------CACCTGGCC	16	del	N/A
38	GCT----------GCACCTGGCC	16	del	N/A
39	GCTGCAGCTGGACAGCGCTGTGGCGCA**A**GCGCACCA-CAC---CATCCGCACCTGGCC	15	del	Mut
40	GCTGCAGCTGGACAGCGCTGTGG----------CGCACCTGGCC	14	del	N/A

Table S7. Sequences of allele-quantification in sib-selection *SHOX2* c.849C>A gRNA-2 #4

#ID	RGEN Treated Sequence	Counts	Type	Allele
1	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	55614	WT	WT
2	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	48571	Mut	Mut
3	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCC---GCACTCAGCCTGCACGCCCTCCGC	3043	del	WT
4	CCAACGCCAGCACCAATGTCGCGCC**C**GTC----------CCTGCACGCCCTCCGC	1533	del	Mut
5	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCC---GCACTCAGCCTGCACGCCCTCCGC	1507	del	Mut
6	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGC-GCACTCAGCCTGCACGCCCTCCGC	1239	del	WT
7	CCAACGCCAG----------CAGCCTGCACGCCCTCCGC	1124	del	N/A
8	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCG----------TGCACGCCCTCCGC	1014	del	WT
9	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCG-----CAGCCTGCACGCCCTCCGC	694	del	Mut
10	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGTGCACTCAGCCTGCACGCCCTCCGC	613	Ins	WT

Supplementary Table 8. Sequence of allele-quantification in sib-selection *SHOX2* c.*28T>C gRNA-1 #1

#ID	RGEN Treated Sequence	Counts	Type	Allele
1	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	47314	WT	WT
2	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	42198	Mut	Mut
3	CCAACGCCAGCAC----------CACGCCCTCCGC	4312	del	N/A
4	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCCTGCACTCAGCCTGCACGCCCTCCGC	2618	Ins	WT
5	CCAACGCCAGCACCAATGTCGC----------GCCTGCACGCCCTCCGC	2227	del	N/A
6	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCG---CTCAGCCTGCACTGCACGCCCTCCGC	2036	Ins	Mut
7	CCAACGCCAGCACCAATGTCGCGCC**C**GTCC----------CGCCCTCCGC	1953	del	Mut
8	CCAACGCCAGCACCAATGTCGCGCC**T**GTCC-------CTCAGCCTGCACGCCCTCCGC	1898	del	WT
9	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGACAGGCACTCAGCCTGCACGCCCTCCGC	1702	Ins	WT
10	CCAACGCCAGCACCAATGTCGC----------GCACGCCCTCCGC	1649	del	N/A
11	CCAACGCCAGCACCAATGTCGC----------GCACTCAGCCTGCACGCCCTCCGC	1577	del	N/A
12	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGGCACTCAGCCTGCACGCCCTCCGC	1218	Ins	Mut
13	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCC---GCACTCAGCCTGCACGCCCTCCGC	818	del	WT
14	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCC---GCACTCAGCCTGCACGCCCTCCGC	589	del	Mut
15	CCAACGCCAGCACCAATGTCGCGCC**T**GTCC----------CGCCCTCCGC	64	del	WT
16	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	45	del	Mut
17	CCAACGCCAGCACCAATGT----------CGC	45	del	N/A
18	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	41	del	WT
19	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGC-GCACTCAGCCTGCACGCCCTCCGC	35	del	WT

Supplementary Table 9. Sequences of allele-quantification in sib-selection *SHOX2* c.*28T>C gRNA-1 #2

#ID	RGEN Treated Sequence	Counts	Type	Allele
1	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	49134	WT	WT
2	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	40933	Mut	Mut
3	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGACTCAGCCTGCACTCAGCCTGCACGC	3056	Ins	WT
4	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGGCCTGCACTCAGCCTGCACGCCCTCCGC	2408	Ins	WT
5	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGGGCTGCACTCAGCCTGCACGCCCTCCGC	2363	Ins	Mut
6	CCAACGCCAGCACCAATGT----------CTCAGCCTGCACGCCCTCCGC	1916	del	N/A
7	CCAACGCCAGCACCAATGTCGCGCCCG----------CTGCACGCCCTCCGC	1554	del	Mut
8	CCAACGCCAGCACCAATGTCGCGCCCGTCC-----CACTCAGCCTGCACGCCCTCCGC	1487	del	Mut
9	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCG-------GCCTGCACGCCCTCCGC	1344	del	Mut
10	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCG--ACTCAGCCTGCACGCCCTCCGC	948	del	WT
11	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGGGCACTCAGCCTGCACGCCCTCCGC	761	Ins	Mut
12	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGC-GCACTCAGCCTGCACGCCCTCCGC	682	del	WT
13	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGG------GCCCTGCACGCCCTCCGC	371	del	Mut
14	C----------GCCTGCACGCCCTCCGC	363	del	N/A
15	CCAACGCCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGG------GCCTGCACGCCCTCCGC	177	del	Mut
16	CCAACGCCAGCACCAATGTCGCGCCCGTCCC----------GCACGCCCTCCGC	144	del	Mut
17	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGG----------GCGCCCTCCGC	125	del	Mut
18	CCAACGCCAGCACCAATGTCGCGCCTGTCCC---GCACTCAGCCTGCACGCCCTCCGC	74	del	WT
19	G----------	53	del	N/A
20	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGACTCAGCCTGCACTCAGCCTGCACGC	38	Ins	Mut
21	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGTGCACTCAGCCTGCACGCCCTCCGC	38	Ins	WT
22	CCAACGCCAGCACCAATGTCGCGCC---CC-----CACTCAGCCTGCACGCCCTCCGC	37	del	N/A
23	CCAACGCCAGCACCAATGTCGCGCCCGTCCCGCGG------GCCTGCACGCCCTCCGC	36	del	Mut
24	CCAACGCCAGCACCAATGTCGC----------GCCTGCACGCCCTCCGC	34	del	N/A
25	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCG-------GCCTGCACGCCCTCCGC	34	del	WT
26	CCAACGCCAGCACCAATGTCGCGCCTGTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	33	del	WT
27	CCAACGCCAGCACCAATGTCGCGCCT----------	32	del	WT
28	CCAACGCCAGCACCAATGTCGCGCCCGTCCC---GCACTCAGCCTGCACGCCCTCCGC	31	del	Mut
29	CCAACGCCAGCACCAATGTCGCGCCTGTCC-----CACTCAGCCTGCACGCCCTCCGC	31	del	WT

Supplementary Table 10. Sequences of allele-quantification in sib-selection *SHOX2* c.*28T>C gRNA-2 #1

#ID	RGEN Treated Sequence	Counts	Type	Allele
1	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	51685	WT	WT
2	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	42258	Mut	Mut
3	CCAACGCCAGCACCAATGTC----------GCACTCAGCCTGCACGCCCTCCGC	2985	del	N/A
4	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCG----------	2679	del	Mut
5	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGG----------ACGCCCTCCGC	1855	del	Mut
6	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGTGCAGGCATTCTCAGCCTGCACGCCCTCCGC	1532	Ins	Mut
7	CCAACGCCAGCACCAATGTCGCG------ACGCCAGCACTCAGCCTGCACGCCCTCCGC	1083	del	N/A
8	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCGCTCAGCCTGCACGCCCTCCGC	1071	Ins	Mut
9	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCACGACACTCAGCCTGCACGCCCTCCGC	1016	Ins	Mut
10	CCAACGC----------CAGCCTGCACGCCCTCCGC	920	del	N/A
11	CCAACGCCAGCACCAAT----------CCTGCACGCCCTCCGC	899	del	N/A
12	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCA--CAGCCTGCACGCCCTCCGC	710	del	WT
13	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGGCGCCACTCAGCCTGCACGCCCTCCGC	540	Ins	Mut
14	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCGCGCGACACTCAGCCTGCACGCCCTCCGC	533	Ins	Mut
15	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGGGTGCACTCAGCCTGCACGCCCTCCGC	432	Ins	Mut
16	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGTCCCG--TCTCAGCCTGCACGCCCTCCGC	344	Ins	Mut
17	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGCCCGTGCACTCAGCCTGCACGCCCTCCGC	330	Ins	Mut
18	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCG----------	139	del	WT
19	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCATTGCACTCAGCCTGCACGCCCTCCGC	110	Ins	WT
20	CTAACGCCACCACCAATGTCGCGCC**T**G----------	79	del	WT
21	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGG--------CTGCACGCCCTCCGC	78	del	WT
22	CCAACGCCAGCACCAATGTCGCGCC**T**----------	55	del	WT
23	CCAACGCCAGCACCAATG----------	40	del	N/A
24	CCAACGCCAGCACCAATGTCGCGCC**C**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	37	del	Mut
25	CCAACGCCAGCACCAATGTCGCGCC**T**G----------	37	del	WT
26	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGG----------ACGCCCTCCGC	35	del	WT
27	CCAACGCCAGCACCAATGTCGCGCC**T**GTCCCGCGGCACTCAGCCTGCACGCCCTCCGC	33	del	WT

Supplementary Table 11. Sequences of allele-quantification in sib-selection *SHOX2* c.*28T>C gRNA-2 #2