THÈSE

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
UFR de médecine et de pharmacie
Cibles moléculaires et thérapeutique de la maladie d'Alzheimer - CiMoTheMA (Poitiers)
(Diplôme National - Arrêté du 25 mai 2016)

École doctorale : Sciences Biologiques et Santé (Limoges)
Secteur de recherche : Aspects moléculaires et cellulaires de la Biologie

Présentée par :
Gara Samara Brajadenta

Development of a functional assay for CHD7, a protein involved in CHARGE syndrome

Directeur(s) de Thèse :
Alain Kitzis, Vincent Thoreau

Soutenue le 14 juin 2019 devant le jury

Jury :

Président	**Thierry Bergès**	Professeur, Université de Poitiers
Rapporteur	**Patrick Vourc'h**	Professeur, Université de Tours
Rapporteur	**Veronique Pingault**	Professeur et praticien hospitalier, Université Paris Descartes
Membre	**Alain Kitzis**	Professeur et praticien hospitalier, Université de Poitiers
Membre	**Vincent Thoreau**	Maître de conférences, Université de Poitiers
Membre	**Sultana M. H. Faradz**	Professeur, Université Diponegoro, Semarang, Indonésie

Pour citer cette thèse :
Gara Samara Brajadenta. *Development of a functional assay for CHD7, a protein involved in CHARGE syndrome* [En ligne]. Thèse Aspects moléculaires et cellulaires de la Biologie. Poitiers : Université de Poitiers, 2019. Disponible sur Internet <http://theses.univ-poitiers.fr>

THESE

Pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITÉ DE POITIERS

(Faculté Médecine et Pharmacie)
(Diplôme National – Arrêté du 25 mai 2016)

École Doctorale : Sciences Biologiques et Santé

Secteur de Recherche : Aspects Moléculaires et

Cellulaires de la Biologie

Présenté par :

Gara Samara Brajadenta

****

Mise au Point d'un Test Fonctionnel pour la Protéine CHD7
Impliquée dans le Syndrome CHARGE

(Development of a Functional Assay for CHD7, a Protein Involved in CHARGE Syndrome)

****

Directeur de Thèse : Professeur Alain Kitzis

Co-directeur de Thèse : Docteur Vincent Thoreau

****

Soutenue le 14 juin 2019

devant la Commission d'Examen

****

JURY

Professeur	Thierry Bergès	Université de Poitiers	Président
Professeur	Patrick Vourc'h	Université de Tours	Rapporteur
Docteur	Veronique Pingault	Université Paris Descartes	Rapporteur
Professeur	Sultana MH Faradz	Université Diponegoro	Examinateur
Professeur	Alain Kitzis	Université de Poitiers	Directeur de thèse
Docteur	Vincent Thoreau	Université de Poitiers	Co-Directeur de thèse

Preface

This thesis is written as partial fulfillment of the requirements to obtain a Ph.D. degree at the Faculty of Medicine and Pharmacy, the University of Poitiers, France. The work was carried out from October 2015 to June 2019 in the laboratory EA-3808 Neurovascular Unit and Cognitive Impairments, University of Poitiers, France under the supervision of Professor Alain Kitzis and Doctor Vincent Thoreau. The work was funded by University of Poitiers and was also supported by Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia.

Gara Samara Brajadenta

Poitiers, June 2019

ACKNOWLEDGMENTS

Many people have contributed their skills into this research; this work would have been impossible without their help and assistance. I would like to express my sincere gratitude to my supervisor, Prof. Alain Kitzis, MD, PhD, for his guidance, time to teach me and endless encouragement.

I am also thankful to my co-supervisor Dr. Vincent Thoreau, for his guidance, knowledge sharing, teaching me basic techniques of the molecular biology, his close supervision in almost all my days in the laboratory, for editing my thesis and for widening my research from various perspectives.

I would like to thank all staff of laboratory of genetics and EA-3808 Neurovascular Unit and Cognitive Impairments, Faculty of Medicine and Pharmacy, the University of Poitiers for their cooperation and friendship they have shared over the years. Particularly, I would like to thank Dr. Frédéric Bilan, Dr. Sylvie Patri, and Dr. Montserrat Rodriguez-Ballesteros for their scientific support during my research, especially for teaching me how to interpret the molecular analysis.

My sincere gratitude goes to Prof. Patrick Vourc'h from the University of Tours and Dr. Veronique Pingault from the University of Paris Descartes for accepting to evaluate my PhD report.

I would like to thank Prof. Sultana MH Faradz, MD, PhD from the Diponegoro University, Semarang, Indonesia and Prof. Thierry Bergès from the University of Poitiers, for honoring me by their presence in the jury of my thesis.

My grateful thank also goes to Prof. Guylène Page, who provided me an opportunity to join EA-3808 team, and who gave me access to the laboratory and research facilities.

Extraordinary gratitude goes to all down at Indonesian Endowment Fund for Education (LPDP) Ministry of Finance, the Republic of Indonesia for helping and providing the funding for the study and the research.

I would like to express my special thanks to Prof. Gérard Mauco, MD, PhD and Prof. Sultana MH Faradz, MD, PhD, who have given me recommendations to continue my PhD at the University of Poitiers.

Last but not the least, I would like to thank Dr. Affandi (former dean), Dr. Catur Setiya Sulistiyana, the dean of Faculty of Medicine, and Dr. Mukarto Siswoyo, the rector of Swadaya Gunung Jati University, Cirebon, Indonesia for giving me the golden opportunity to do this excellent PhD thesis in France.

TABLE OF CONTENTS

CHAPTER VII. CONCLUSIONS AND FUTURE PERSPECTIVES

Abstract

CHARGE syndrome (CS) is a rare genetic disease characterized by numerous congenital abnormalities, mainly caused by *de novo* alterations of the *CHD7* gene. It encodes a chromodomain protein, involved in the ATP-dependent remodeling of chromatin. The vast majority of *CHD7* alterations consists in null alleles like deletions, non-sense substitutions or frameshift-causing variations. We report the first molecular diagnosis of an Indonesian CS patient by a targeted NGS (next-generation sequencing) gene panel (*CHD7*, *EFTUD2*, and *HOXA1*). We identified a novel heterozygous nonsense mutation in exon 34 of *CHD7* (c.7234G>T or p.Glu2412Ter). Functional analyses to confirm the pathogenicity of *CHD7* variants are lacking and urgently needed. Therefore, the aim of this study was to establish a functional test for wild-type (WT) or variants of CHD7 protein found in CS patients. Using an expression vector encoding CHD7, three variants harboring an amino acid substitution and one variant with a five-amino acid insertion were generated via site-directed mutagenesis. Then CHD7 proteins, either wild-type (WT) or variants, were overexpressed in HeLa cell line. Protein expression was highlighted by Western blot and immunofluorescence. We then used real-time RT-PCR to study CHD7 functionality by evaluating the transcript amounts of five genes whose expression is regulated by *CHD7* according to the literature. These reporter genes are *45S rDNA*, *SOX4*, *SOX10*, *ID2,* and *MYRF*. We observed that, upon WT-CHD7 expression, the reporter gene transcriptions were downregulated, whereas the four variant alleles of *CHD7* had no impact. This suggests that these alleles are not polymorphisms because the variant proteins appeared non-functional. Furthermore, we applied our biological assay in SH-SY5Y cell line in which endogenous *CHD7* gene was mutated using the CRISPR/Cas9 technique. Then, we observed that when a *CHD7* missense variant was expressed, the transcription levels of the five reporter genes were non-significantly different, compared with the cells in which both C*HD7* alleles were knocked-out. Therefore, the studied variants can be considered as disease-causing of CS.

Keywords: CHARGE syndrome, *CHD7*, functional test, targeted gene panel.

Résumé

Le syndrome CHARGE (CS) est une maladie génétique rare caractérisée par de nombreuses anomalies congénitales, majoritairement causées par des altérations *de novo* du gène *CHD7*. Celui-ci code pour une protéine à chromodomaines, impliquée dans le remodelage ATP-dépendant de la chromatine. La grande majorité des altérations de *CHD7* consiste en allèles nuls tels que des délétions, des substitutions non-sens ou des décalages du cadre de lecture. Nous avons réalisé le premier diagnostic moléculaire d'un patient Indonésien atteint du CS, en étudiant un panel de gènes (*CHD7*, *EFTUD2*, et *HOXA1*) par NGS (next-generation sequencing). Nous avons identifié une nouvelle mutation non-sens hétérozygote dans l'exon 34 du gène *CHD7* (c.7234G>T ou p.Glu2412Ter). Par ailleurs, il n'existe pas d'analyse fonctionnelle qui permettrait de caractériser la pathogénicité des variants de la protéine CHD7 rencontrés chez des patients. C'est pourquoi l'objectif de ce travail est de mettre au point un test fonctionnel de la protéine CHD7, sous forme sauvage ou mutée. Pour cela, nous avons généré par mutagénèse dirigée des vecteurs codant pour trois variants faux-sens de CHD7 et le variant présentant une insertion de cinq acides aminés. Ensuite, les protéines CHD7, sous forme sauvage ou variante, ont été surexprimées dans la lignée HeLa. L'expression des protéines a été mise en évidence par Western blot et par immunofluorescence. Pour étudier la fonctionnalité de CHD7, nous avons quantifié par RT-qPCR les transcrits de cinq gènes (l'*ADNr 45S*, *SOX4*, *SOX10*, *MYRF*, et *ID2*), dont la transcription est selon la littérature régulée par CHD7. Nous avons observé que l'expression de CHD7 sauvage entraînait une diminution significative et reproductible des quantités de transcrits correspondant à tous les gènes rapporteurs. Par contre, l'expression des quatre allèles variants de CHD7 n'avait aucun impact, ce qui suggère que ces variants ne sont pas fonctionnels. Par ailleurs, nous avons appliqué notre test biologique dans des cellules de la lignée SH-SY5Y, pour lesquelles nous avons introduit une mutation faux-sens dans le génome en utilisant la technique CRISPR/Cas9. Lorsque ce variant était exprimé, les niveaux de transcription des cinq gènes rapporteurs n'étaient pas significativement différents de ceux observés dans les cellules où les deux allèles de *CHD7* avaient été invalidés. Par conséquent, les variants étudiés peuvent être répertoriés comme résultant de mutations causales du CS.

Mots clés: *CHD7*, syndrome CHARGE, test fonctionnel, panel de gènes.

List of Abbreviations

ATP	Adenosine triphosphate
ASD	Atrial septal defect
AVSD	Atrioventricular septal defects
BRK	Brahma and Kismet
cDNA	Complementary deoxyribonucleic acid
CGH	Comparative genomic hybridization
CHARGE	Coloboma, Heart disease, Choanal atresia, Retardation of growth and/or development, Genital hypoplasia and Ear abnormalities with or without deafness
Cas9	CRISPR-associated protein 9
CHD7	Chromodomain Helicase DNA-binding 7
CLP	Cleft lip and/or palate
CRISPR	Clustered regularly interspaced short palindromic repeats
CrRNA	CRISPR RNA
DMEM	Dulbecco's modified eagle's medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DORV	Double outlet right ventricle
DSB	Double strand break
FISH	Fluorescence in situ hybridization
HDR	Homology-directed repair
Indel	Insertions and deletions
LB	Luria Bertani
Mb	Mega base
MCA	Multiple congenital abnormalities
MLPA	Multiplex ligation-dependent probe amplification

M-MLV	Moloney murine leukemia virus
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDA	Patent ductus arteriosus
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SANT	Switching-defective protein 3, Adaptor 2, Nuclear receptor compressor, Transcription factor IIIB
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sgRNA	Single guide RNA
SpCas9	Streptococcus pyogenes Cas9
SSB	Single strand break
ssODN	Single stranded donor oligonucleotide
SNP	Single nucleotide polymorphism
TALENs	Transcription activator-like effector nucleases
TGA	Transposition of the great arteries
trascrRNA	Trans-activating CrRNA
TOF	Tetralogy of Fallot
UV	Unclassified variant
VSD	Ventricular septal defect
WES	Whole exome sequencing
WT	Wild type
ZFNs	Zinc-finger nucleases

List of Figures

List of Tables

List of Publications

Brajadenta GS. Gangguan Akibat Kekurangan Iodium (GAKI). *Tunas Medika*. 1(1):69-77 (2012). e-ISSN 2579-7514.

Brajadenta GS. Gangguang Somatisasi. *Tunas Medika*. 2(2):1-9 (2015). ISSN 9-772089-604004. e-ISSN 2579-7514.

Brajadenta GS, Adyanti RD. Pengaruh Pemberian Jus Daun Lidah Buaya terhadap Kadar Gula Darah Tikus Jantan Albino yang diinduksi Aloksan. *Tunas Medika*. 2(2):15-20 (2015). e-ISSN 2579-7514.

Rudianto CC, Brajadenta GS, Fitrikasari A, Winarni TI. Correlation of SNP8NRG433E1006 Polymorphism Neuregulin 1 (NRG1) Gene with Schizophrenia in Java Ethnic. *Global Medical and Health Communication*. 6(1):49-56 (2018). DOI: http://dx.doi.org/10.29313/gmhc.v6i1.2658.

Brajadenta GS, Laksana ASD, Peramiarti IDSAP. Faktor Risiko Tuberkulosis Paru Anak: Studi pada Balai Kesehatan Paru Masyarakat (BKPM) Purwokerto. *STRADA Jurnal Ilmiah Kesehatan*. 7(2):1-6 (2018). DOI: https://doi.org/10.30994/sjik.v7i2.160.

Brajadenta GS, Sumaerah DS. Therapeutic Effect of Anti Tuberculosis Drugs on Erythrocyte Sedimentation Rate: Study of Childhood Tuberculosis in Prof. Dr. Margono Soekarjo Hospital. *Jurnal Ilmiah Ilmu Kesehatan: Wawasan Kesehatan*. 5:1-9 (2019). DOI: https://doi.org/10.33485/jiik-wk.v5i2.96.

Brajadenta GS, Utari A, Patri S, Bilan F, Faradz SMH, Kitzis A, Thoreau V. Identification of a Novel CHD7 Mutation in a CHARGE Syndrome Patient in Indonesia. *Ann Lab Med*. 39:503-506 (2019). DOI: https://doi.org/10.3343/alm.2019.39.5.503.

Brajadenta GS, Kitzis A, Thoreau V. Localization Study of CHD7 Protein Expressed in HeLa and HEK293 Cells. *J. Phys.: Conf. Ser*. 1360 012002 (2019). DOI: 10.1088/1742-6596/1360/1/012002

Brajadenta GS, Sari AIP, Nauphar D, Pratamawati TM, Thoreau V. Molecular Analysis of Exon 7 of the Fibroblast Growth Factor Receptor 2 (*FGFR2*) gene: Case Report of an Indonesian Patient with Apert Syndrome. *J Med Case Rep*. 13(1):244 (2019). DOI: 10.1186/s13256-019-2173-x.

Brajadenta GS, Bilan F, Gilbert-Dussardier B, Kitzis A, Thoreau V. A Functional Assay to Study the Pathogenicity of CHD7 Protein Variants Encountered in CHARGE Syndrome Patients. *Eur J Hum Genet*. 27(11):1683-1691 (2019). DOI: 10.1038/s41431-019-0465-7.

Brajadenta GS, Létienne L, Charraud V, Pratamawati TM, Nauphar D, Sari AIP, Kitzis A, Thoreau V, Bilan F, Faradz SMH. Identification of Genetic Causes of Unexplained Intellectual Disability by Array-CGH in Indonesian High-Risk Populations. In preparation.

Mise au Point d'un Test Fonctionnel pour la Protéine CHD7 Impliquée dans le Syndrome CHARGE

Synopsis

Le syndrome CHARGE (SC) est une maladie génétique rare caractérisée par de nombreuses anomalies congénitales, majoritairement causées par des altérations de novo du gène *CHD7*. Celui-ci code pour une protéine à chromodomaines, impliquée dans le remodelage ATP-dépendant de la chromatine. La grande majorité des altérations de *CHD7* consiste en allèles nuls tels que des délétions, des substitutions non-sens ou des décalages du cadre de lecture. Des cas sporadiques de SC sont liés aux altérations d'autres gènes. Par conséquent, le séquençage ciblé de panels de gènes est devenu la plus simple stratégie d'analyse de mutations à grande échelle.

Le diagnostic du SC est exclusivement basé sur des caractéristiques cliniques. Les patients présentent généralement divers signes majeurs ou une combinaison de signes majeurs et mineurs. Cependant, le test moléculaire est essentiel pour confirmer le diagnostic clinique. Il aide les cliniciens à identifier et à confirmer la cause des multiples anomalies congénitales chez les patients. Cette connaissance leur permet de fournir les meilleurs soins à ces patients et également de donner des informations sur le risque de récurrence.

On suppose que l'haploinsuffisance pour *CHD7* est le mécanisme pathogène responsable du SC. L'interprétation des mutations tronquées non-sens ou des décalages du cadre de lecture est claire. Les variants faux-sens et d'épissage de *CHD7* sont rares dans la population et ne représentent respectivement que 8% et 11% des cas du SC. Ces variants restent de signification clinique inconnue (VUS) en raison du manque d'informations génétiques qui pourraient aider à déterminer leur pathogénicité. De plus, l'interprétation de l'effet d'un faux-sens du gène *CHD7* reste un défi. En raison de la grande taille du gène (188 kb, avec 38 exons) et donc de la protéine (340 kDa), le test fonctionnel de CHD7 est difficile à réaliser.

Par ailleurs, les analyses fonctionnelles déjà publiées pour confirmer la pathogénicité des variants sont trop complexes pour une utilisation en routine. C'est pourquoi l'objectif principal de ce travail est de mettre au point un test fonctionnel simple de la protéine CHD7, sous forme sauvage ou mutée. Pour cela, nous avons généré par mutagénèse dirigée des variants

faux-sens de CHD7 et un variant présentant une insertion de cinq acides aminés. Nous avons d'abord développé un système de test fonctionnel basé sur la surexpression de CHD7 et avons également développé notre système pour tester l'effet des variants faux-sens de CHD7 avec un niveau d'expression endogène dans lequel la substitution de nucléotide a été introduite dans le génome par modification génomique avec la technique CRISPR/Cas9.

L'objectif secondaire de cette thèse est de réaliser le premier diagnostic moléculaire d'un patient Indonésien précédemment diagnostiqué comme un SC typique, en étudiant un panel de gènes (*CHD7*, *EFTUD2*, et *HOXA1*) par NGS (next-generation sequencing).

Dans le chapitre 2, nous présentons une revue de la littérature sur le SC, comprenant une brève histoire du SC et de l'association au syndrome, l'évolution du spectre phénotypique, la mise à jour des critères de diagnostic du SC et les causes génétiques impliquées dans le SC. Nous décrivons également les mutations de *CHD7*, l'organisation des domaines de la protéine CHD7 et sa fonction. De plus, nous résumons le mécanisme pathologique du SC et un nouveau système de classification permettant de prédire l'effet pathogène des variants faux-sens de *CHD7*.

Le chapitre 3 décrit les matériels et les méthodes utilisés pour identifier la cause génétique du CS chez un patient Indonésien à l'aide d'un panel de gènes ciblé par NGS. En outre, nous décrivons les techniques utilisées pour développer un test fonctionnel permettant d'étudier la pathogénicité des variants de CHD7 en utilisant une approche par surexpression puis en utilisant la technologie CRISPR/Cas9 pour obtenir l'expression endogène des variants de CHD7. Nous fournissons également des analyses de la littérature de plusieurs gènes cibles de CHD7 que nous avons utilisés dans notre système pour étudier la fonctionnalité des variants de CHD7. Ces gènes cibles sont *l'ADNr 45S*, *SOX4*, *SOX10*, *ID2* et *MYRF*. Dans la dernière partie, nous décrivons la technologie d'édition du génome CRISPR/Cas9 et expliquons comment nous avons développé notre système pour tester l'effet des variants faux-sens CHD7 avec un niveau d'expression endogène, pour étudier leur fonctionnalité dans un contexte plus physiologique.

Le chapitre 4 porte sur un diagnostic clinique et moléculaire du SC. Nous décrivons le premier patient Indonésien atteint du SC qui a été précédemment diagnostiqué selon une combinaison de critères cliniques. Par ailleurs, nous avons confirmé ce diagnostic au niveau moléculaire. Nous avons identifié une nouvelle mutation non-sens hétérozygote dans l'exon 34 du gène CHD7 (c.7234G>T ou p.Glu2412Ter), qui a ensuite été

confirmée par séquençage selon la technique de Sanger. Le dépistage génétique parental a confirmé l'origine *de novo* de la mutation.

Dans le chapitre 5, nous exposons les résultats. Le sous-chapitre 5.1 présente les résultats du développement d'un test fonctionnel des variants de CHD7. En utilisant un vecteur d'expression codant pour CHD7, trois substitutions d'acides aminés p.(Gly1982Trp) ou G1982W, p.(Arg2319Cys) ou R2319C, p.(Leu1302Pro) ou L1302P et une insertion de cinq acides aminés p.(His1801_Gly1802insAspGlyHisGlyThr) ou 1801insHGHG, qui ont été découverts chez des patients par notre laboratoire, ont été générés par mutagenèse dirigée. Ensuite, les protéines CHD7, de type sauvage (WT) ou variants, ont été surexprimées dans la lignée cellulaire HeLa. L'expression des protéines a été mise en évidence par Western Blot et par immunofluorescence. Nous avons ensuite utilisé la RT-PCR en temps réel pour étudier la fonctionnalité de CHD7 en évaluant les quantités de transcription de cinq gènes cibles. Nous avons observé que, lors de l'expression de WT-CHD7, les transcriptions des gènes cibles étaient diminuées. En revanche, lorsque nous avons exprimé l'un ou l'autre des quatre allèles variants de CHD7, la quantité de transcription de chaque gène cible était non affectée par rapport à la condition de contrôle transfectée par un vecteur vide (mock). Ce résultat suggère que la capacité normale de la protéine CHD7 à modifier le niveau de transcription de plusieurs gènes a été altérée par les trois substitutions d'acides aminés et par l'insertion de cinq acides aminés que nous avons étudiées. Par conséquent, ces variations peuvent être considérées comme résultant de mutations causales du SC.

Pour valider ces résultats, nous avons appliqué notre test biologique à deux variants de la protéine CHD7, p.(Thr894Ala) ou T894A et p.(Ala2160Thr) ou A2160T, classés comme non pathogènes avec l'outil de prédiction GnomAD, et ayant ont été rapportés comme des polymorphismes dans la littérature. Ceci a été confirmé par notre test. Nous avons conclu que la réduction de la quantité de transcription des gènes cibles devait résulter de la fonctionnalité normale de la protéine CHD7 dans les cellules HeLa.

Le sous-chapitre 5.2 présente un test fonctionnel de variants faux-sens de CHD7 avec un niveau endogène d'expression en utilisant la technique CRISPR/Cas9. Afin de perturber la fonction du gène *CHD7*, un ARN guide ciblant l'exon 3 du gène *CHD7* a été dessiné. Un exon précoce a été choisi afin d'augmenter les chances d'occurrence d'une perte totale de fonction (knock-out). Dans cette expérience, nous avons induit l'activation de la voie de réparation de l'ADN (NHEJ) non-homologue, afin

de créer des insertions ou délétions (indels), de manière à obtenir des mutations par décalage du cadre de lecture. Nous avons d'abord évalué le niveau d'expression de l'ARNm des gènes cibles de CHD7 dans les cellules HeLa : l'expression endogène de CHD7 n'intervient pas dans la régulation de la transcription de ces gènes. Par contre, dans la lignée SH-SY5Y, l'expression endogène de CHD7 contribue à inhiber la transcription de ces gènes cibles. Pour générer des variants faux-sens de CHD7, un plasmide codant pour Cas9 et les ARN guides ciblant les exons correspondant à chaque variant ont été choisis. Nous avons ensuite co-transfecté dans les cellules SH-SY5Y une matrice de recombinaison homologue, sous forme de plasmide double brin linéarisé. Nous avons ainsi obtenu un clone contenant la mutation homozygote L1302P-CHD7. Malheureusement, nous n'avons pas encore obtenu les deux autres variants (G1982W et R2319C). Dans les cellules SH-SY5Y éditées, la quantité d'ARN transcrite à partir de ces gènes était régulée positivement, par rapport à la lignée parentale. De plus, aucune différence statistiquement significative n'a été observée par rapport aux cellules où les deux allèles de *CHD7* ont été invalidés. Ce résultat confirme que la capacité normale de la protéine CHD7 à modifier le niveau de transcription de plusieurs gènes est altérée par le variant L1302P.

Le chapitre 6 propose une discussion générale. Dans ce chapitre, nous analysons, interprétons et décrivons de manière critique l'importance de nos résultats. La mutation récemment découverte c.7234G>T ou p.(Glu2412Ter) crée un codon de terminaison de traduction prématurée, conduisant vraisemblablement à la suppression de l'ARNm via un processus de dégradation (NMD). Par conséquent, nous supposons que cette mutation représente un allèle nul, causant le SC par haploinsuffisance. Cette nouvelle mutation a été soumis à la base de données CHD7 (https://www.chd7.org). De plus, nous discutons des raisons pour lesquelles il est essentiel d'établir un diagnostic moléculaire chez les patients SC diagnostiqués cliniquement et fournissons des arguments en faveur de l'utilisation d'un panel ciblé de gènes impliqués dans le SC.

Nous avons développé un test simple par deux approches. En utilisant l'approche par surexpression des variants faux-sens de CHD7 dans la lignée HeLa, nous avons observé que l'expression de CHD7 sauvage entraînait une diminution significative et reproductible des quantités de transcrits correspondant à tous les gènes cibles. Par contre, l'expression des quatre allèles variants de CHD7 n'avait aucun impact, ce qui suggère que ces variants ne sont pas fonctionnels. Par ailleurs, nous avons appliqué

notre test biologique dans des cellules de la lignée SH-SY5Y, pour lesquelles nous avons introduit une mutation faux-sens dans le génome en utilisant la technique CRISPR/Cas9. Lorsque ce variant était exprimé, les niveaux de transcription des cinq gènes cibles n'étaient pas significativement différents de ceux observés dans les cellules où les deux allèles de CHD7 avaient été invalidés. Par conséquent, les variants étudiés peuvent être répertoriés comme résultant de mutations causales du SC.

Le chapitre 7 fournit des conclusions et discute des perspectives futures. Le chapitre 7.1 résume les résultats décrits dans cette thèse. Dans le chapitre 7.2, une réflexion est donnée sur ce que nous avons réalisé et sur les connaissances que nous avons ajoutées au domaine du diagnostic moléculaire du SC et de l'analyse fonctionnelle de la protéine CHD7. Les résultats sont discutés dans une perspective plus large par rapport aux développements récents concernant le test fonctionnel de CHD7 et le système de classification des mutations, afin de déterminer la pathogénicité des variants faux-sens de CHD7.

CHAPTER I

INTRODUCTION

1.1 General Introduction

CHARGE syndrome (CS; OMIM 214800) is a congenital disorder that arises during early fetal development and affects many areas of the body and organ systems. The term CHARGE itself is an abbreviation for clinical features commonly found in the patients: **C**oloboma, **H**eart defect, **A**tresia choanae, **R**etarded growth and development, **G**enital anomalies, and **E**ar anomalies (Pagon *et al*., 1981).

CS is a rare disease in which the gene encoding CHD7, a chromatin remodeling protein, has been identified as a genetic cause in more than half of all CS cases (Vissers *et al*., 2004; Aramaki *et al*., 2006; Félix *et al*., 2006, Vuorela *et al*., 2007). In addition, sporadic CS cases are linked to alterations in other genes (Lalani *et al*., 2004; Legendre *et al*., 2017). *CHD7* is one of the largest human genes. Therefore, as a large-scale mutation scanning strategy, targeted gene panel sequencing has become the most straightforward approach used in the genetics laboratories (van Ravenswaaij and Martin, 2017). To date, clinical diagnostic criteria are still used in the clinical setting to establish a clinical diagnosis of CS (Verloes, 2005; Lalani *et al*., 2012; Hale *et al*., 2016). Affected individuals usually have various major signs or a combination of major and minor signs.

1.2 Background

1.2.1 Molecular Diagnosis of CHARGE Syndrome

In developing countries with a lack of molecular diagnostic facilities, the diagnosis of CS is still based exclusively upon clinical features.

However, molecular testing is essential to establish a molecular diagnosis. It helps clinicians to identify and to confirm the definitive cause of multiple congenital anomalies in clinically-diagnosed CS patients. This knowledge allows them to provide these patients with the best care and also to give the information on recurrence risk. Finally, the ascertainment of the causal nature of CS-causing gene sequence variants creates many opportunities to perform new research on CS and its phenotypes.

There are several approaches to establish a molecular diagnosis of clinical suspicion of CS. The first options are *CHD7* Sanger sequencing and chromosomal microarray or multiplex ligation-dependent probe amplification (MLPA) (Bergman *et al.*, 2008; Lalani *et al.*, 2012; Ravenswaaij *et al.*, 2015). Currently, sequencing of panels including genes involved in intellectual disability (ID) is widely used (Vissers *et al.*, 2004; Legendre *et al.*, 2017; Grozeva *et al.*, 2015; Villate *et al.*, 2018). Occasionally, CS-causing *CHD7* alterations are accidentally identified in the sequencing of large panels of ID-related genes (Grozeva *et al.*, 2015; Kahmoto *et al.*, 2016; Villate *et al.*, 2018) since ID has been detected in almost all CS cases (Zentner *et al.*, 2010; Bergman *et al.*, 2011; Hale *et al.*, 2016). However, the cost-benefit balance of these methods is questionable, especially in developing countries with limited funding. Genetic analysis requires cost-effective and time-consuming strategies. In this thesis, we therefore favor a strategy for molecular diagnosis of CS using targeted next-generation sequencing (NGS) gene panel.

1.2.2 Challenges in the Development of Functional Assay of CHD7

Haploinsufficiency for CHD7 is hypothesized as the pathogenic mechanism causing CS (Delehaye *et al*., 2007). The interpretation of truncating nonsense and frameshift mutations is often clear. *CHD7* missense and splice sites variants are rare in the population and account for only 8% and 11-12% of CS cases, respectively (Jongmans *et al*., 2006; Zentner *et al*., 2010; Jenssen *et al*., 2012; Stenson *et al*., 2017). These variants often remain of unknown clinical significance because of the lack of genetic information that may help to determine their pathogenicity. Moreover, interpreting the effect of a missense or other variant in the CHD7 protein amino acid sequence remains a challenge. Due to the large size of the gene (188 kb, with 38 exons) and the protein (340 kDa), the functional assay of CHD7 is challenging to be conducted.

Although computational tools have been refined to predict putative deleterious effects of missense and splice variants, functional analyses confirming the pathogenicity of the variants are lacking (Bergman *et al*., 2012; Villate *et al*., 2018). A biochemical method has been described to test in vitro the function of the CHD7 protein (Bouazoune and Kingstone, 2012). Two years later, another approach has been developed in a zebrafish model. *Chd7* knock-down was achieved by injection of splice-blocking morpholino (MO) against *chd7* (*chd7*-Mo). Then, phenotype rescue was studied following mRNA injection of WT-CHD7 or mutated alleles (Balasubramanian *et al*., 2014). Moreover, DNA methylation signature was

recently proposed as a diagnosis tool for disorders involving proteins that attend the epigenetic machinery (Butcher *et al*., 2017; Aref-Eshghi *et al*., 2018). However, the complexity of these systems is incompatible with routine use. Therefore, alternative ways of predicting the pathogenicity of these variants are urgently needed. In this thesis, we developed a functional test system using techniques commonly used in most cellular and molecular biology laboratories, to study the impact upon protein functionality of the missense variants and of other mutations leading to change in CHD7 amino acid sequence found in CS patients.

1.3 Objectives

The primary objective of this thesis was to develop a functional test for wild-type (WT) or variants of CHD7, a protein involved in CS. In our laboratory, in the case of patients for whom CS has been diagnosed, the exploration of *CHD7* consists in the search for mutations by analysis of the 38 exons and flanking intronic sequences. If no mutation is detected, Comparative Genomic Hybridization (CGH) array experiment is performed with the aim to discover other regions containing genes whose alterations may be responsible for this syndrome. Mutations in the *CHD7* gene such as deletion, nonsense, and those with an impact on RNA splicing have predictable consequences: the development of CS. However, the impact of missense variants is unclear: are these substitutions mutations or polymorphisms? They may have an influence upon splicing of the gene (which is studied in the laboratory by the minigene technique) or upon the

function of the CHD7 protein? Furthermore, by using our system, we performed a functionality test for a CHD7 protein variant resulting from an intronic mutation, to determine the pathogenicity of this variant. We first developed a functional assay system based on overexpression of CHD7.

We also developed our system to test the effect of CHD7 missense variants with an endogenous expression level in cell lines in which the nucleotide substitution has been introduced into the genome by genetic engineering with CRISPR/Cas9 technique.

The secondary objective of this thesis was to investigate the molecular diagnosis by NGS targeted gene panel of an Indonesian patient who was previously diagnosed as typical CS. We determined the whole exonic and flanking intronic sequences of two genes involved in CS, *CHD7* (OMIM 608892) and *EFTUD2* (OMIM 603892) (Vissers *et al*., 2004; Aramaki *et al*., 2006; Legendre *et al*., 2017). *HOXA1* (OMIM 142955) had been included in the panel since alterations of this gene are involved in overlapping features with CS (Tischfield *et al*., 2005).

CHAPTER II

LITERATURE REVIEW

2.1 Overview of CHARGE Syndrome

2.1.1 CHARGE: From Association to Syndrome

CHARGE syndrome (CS; OMIM 214800) is a rare genetic disease characterized by a combination of various congenital malformations. It was first recognized in 1979 independently by a pediatrician Dr. Bryan Hall and an ophthalmologist Dr. Helen Hittner and colleagues (Hall, 1979; Hittner *et al*., 1979). In 1979, Hall reported a cohort of 17 patients with various congenital malformations including congenital heart defect, ear abnormality, coloboma, hypogenitalism, and bilateral or unilateral posterior choanal atresia. They assumed that there was an association between these clinical manifestations and suggested a broader clinical spectrum since they also identified patients with the same association of this clinical manifestations but without choanal atresia (Hall, 1979). In the same year, Hittner and colleagues studied the association that appeared in ten patients with colobomatous microphthalmia, congenital heart defects, ear abnormalities (including hearing loss), facial paralysis, and intellectual disability (Hittner *et al*., 1979). Hereafter in 1981, Roberta Pagon and her colleagues realized that this ‘’association’’ marked by Hall and Hittner *et al*. formed a phenotype group specific of a disorder. They also included 21 other patients with the same manifestations to support the association. Finally, to make this association better distinguishable and to create awareness in the clinical setting, Pagon coined the term CHARGE, as an acronym for **C**oloboma, **H**eart defects, **A**tresia of choanae, **R**etardation of

growth and/or development, **G**enital hypoplasia and **E**ar abnormalities and/or deafness (Pagon *et al.*, 1981).

Furthermore, CHARGE is now accepted as a genetic syndrome, differentiated by the fact that one common pathologic anomaly causes all manifestations. This distinction was brought through the work of a group of geneticists from the Netherlands who published a study linking CS to a microdeletion on the chromosome 8q12 and a balanced translocation between chromosomes 6 and 8 (Vissers *et al.*, 2004). With this new information, the medical community then reconsidered the use of term CHARGE syndrome instead of association.

The discovery of the *CHD7* gene as the major cause of CS resulted in a renewal of interest to this syndrome. Until 2018, there are 268 publications indexed in www.pubmed.org using “CHARGE syndrome and CHD7” as a keyword (Figure 1).

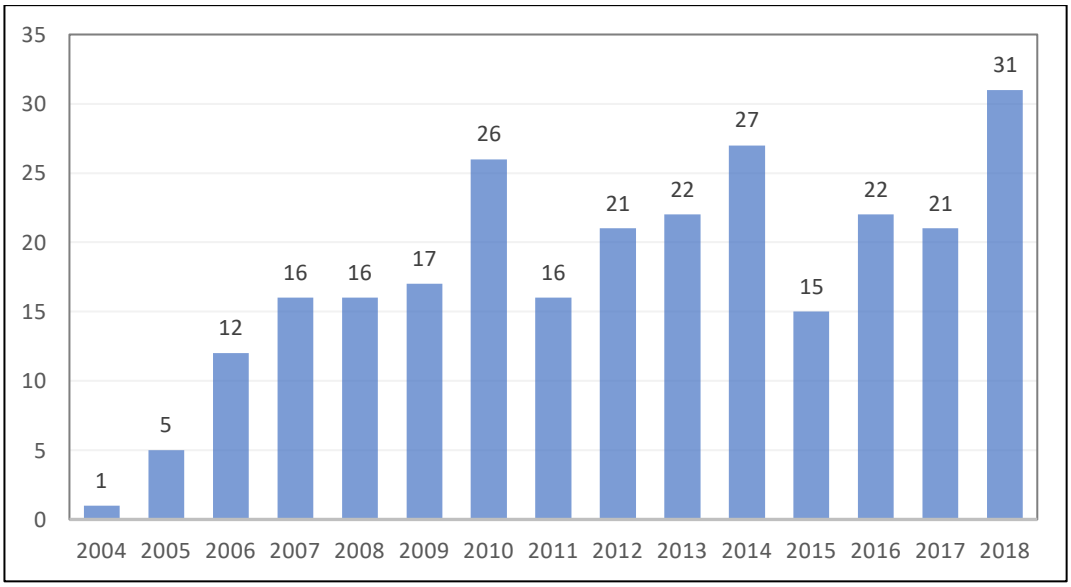

Figure 1. PubMed search on CHARGE syndrome or CHD7

2.1.2 Prevalence and Demographics

The precise incidence of CS is likely difficult to define since there is a lack of strict diagnostic parameters and since the cost of molecular analyses of a wide population is prohibitive. However, the prevalence is estimated between 1 in 8,500-12,000 live births (Kallen *et al*., 1999; Lalani *et al*., 2012; Issekutz *et al*., 2005). Several studies of individuals with genetic CS have shown a slight female predominance (59%:41%). However, larger studies are needed to establish a definitive female-to-male ratio (Jongmans *et al*., 2006; Aramaki *et al*., 2006).

2.1.3 Inheritance Pattern

The CS-causing mutations mostly occur *de novo*, which means that the mutation occurs sporadically and is not inherited from the parents. Approximately 97% of CS cases were *de novo*, and only 3% were inherited from a parent by autosomal dominant transmission (Vissers *et al*., 2004; Jongmans *et al*., 2006; Wincent *et al*., 2008; Lalani *et al*., 2006; Vuorela *et al*., 2008). Moreover, advanced paternal age has been reported as a contributing factor in *de novo* cases (Goldmann *et al*., 2016).

In CS cases, autosomal dominant inheritance rarely occurs due to infertility problems and delayed or absent puberty. A case of autosomal dominant inheritance has been reported in a mildly affected mother (Lalani *et al*., 2006). In another case, this type of inheritance also occurred in a mother with a mild presentation attributable to somatic cell mosaicism. Both

of her sons were affected (Vissers *et al*., 2004). This condition shows that mosaicism is responsible for some CS cases. Besides, it has been described that a parent with one affected child has an empiric recurrence risk of 1-2% due to germline mosaicism (Rothlisberger and Kotzot, 2007; Lalani *et al*., 2012).

2.1.4 Variability in CHARGE Clinical Features

2.1.4.1 Coloboma and ophthalmic features

Coloboma is a fissure or segmental defect resulting from a failure to close the eyeball during fetal development (Schneiderman and Balogun, 2000). Coloboma and other ocular malformations have been reported in 80-90% of CS patients (Lalani *et al*., 2012; Blake *et al*., 1998). The presence of these rare anomalies has previously been perceived as a pathognomonic sign of clinical CS. Recent studies with confirmed genetic CS patients have proven a strong association between coloboma and CS (Jongmans *et al*., 2006; Aramaki *et al*., 2006; Wincent *et al*., 2008; Lalani *et al*., 2006). This coloboma may occur bilaterally or unilaterally, affecting the eyelid, iris, retina, choroid, optic disc or macula. One typical coloboma is chorioretinal that predisposes to retinal detachment and has an impact on visual impairment (McMain *et al*., 2008). The variation in visual capability depends on the site and nature of the malformation, ranging from absent to near-normal vision (Nishina *et al*., 2012).

In addition, anterior segment anomalies can also be involved, including microphthalmia, microcornea, and cataracts. Other ophthalmic features include refractive errors, strabismus, and ptosis which had been reported but less frequently (Onwochei *et al*., 2000). Microphthalmia is commonly identified in clinical CS as an isolated condition or in conjunction with coloboma (Blake and Prasad, 2006; Lalani *et al*., 2012; Sanlaville and Verloes, 2007). Therefore, suspected CS patients should be evaluated by an ophthalmologist.

2.1.4.2 Cardiac malformations

Cardiac defects have also been reported in the majority (75-85%) of clinical CS patients (Jongmans *et al*., 2006). Previous investigations, performed before *CHD7* identification, showed that conotruncal heart anomalies such as Tetralogy of Fallot (ToF), aortic arch interruption, double outlet right ventricle (DORV), arch vessel anomalies, and atrioventricular septal defects (AVSD) were commonly detected (Wyse *et al*., 1993; Corsten-Jansen *et al*., 2013). Recent studies, since the discovery of the *CHD7* gene, have identified broader phenotype with every type of cardiac defect (except heterotaxy and cardiomyopathy). However, AVSDs and conotruncal are remaining the majority (Corsten-Jansen and Scambler, 2017).

2.1.4.3 Choanal atresia and other upper airway abnormalities

Choanal atresia is a narrowing or obstruction between the nasal cavity and the nasopharynx. This obstruction may be bony or membranous and occurs in one or both nasal cavities (Blake and Prasad, 2006). This abnormality has been previously identified as a reliable criterion for clinical CS, with an incidence of 50-60% (Blake and Prasad, 2006; Lalani *et al*., 2012). At birth, this condition results in breathing difficulty due to obstruction, and this can be detected when a nasogastric tube cannot pass through the nose and into the pharynx. Moreover, children with bilateral choanal atresia require endotracheal intubation or management with an oral airway until a surgical correction can be performed.

Airway obstruction below the choanae is also present in 70% of CS patients: laryngomalacia (40%), tracheomalacia (20%), and subglottic stenosis (10%) (Morgan *et al*., 1993). Moreover, cleft lip and palate are present in 15-20% of CS cases, and tracheoesophageal fistula (TOF) is also reported in the same percentage (Houben and Curry, 2008).

2.1.4.4 Growth and developmental retardation

Children with CS usually have a normal birth weight and length. They are appropriate for gestational age (AGA); however, in the majority, after birth, they then fail to achieve optimal growth (Husu *et a*l., 2013; Blake and Prasad, 2006; Lalani *et al*., 2012). The incidence of growth retardation in CS patients is estimated at 70-80% (Lalani *et al*., 2012). A study has

evaluated the growth hormone secretion by provocation testing in 25 CS children. The result revealed that ninety percent had a normal level of growth hormone and ten percent of children revealed low peak hormone level which was significantly shorter (height ≤2.5 standard deviation score) (Pinto *et al*., 2005). Until now, there are no published data on recombinant human growth hormone therapy in CS children and short stature.

Cognitive ability is also affected in CS children. A study to access cognitive ability has been performed using the Adaptive Behavior Evaluation Scale (ABES): 50% of subjects had scores <70 (reference means 100, with standard deviation 15). CS children who walked earlier had fewer health problems, better hearing and vision, and higher ABES scores (Salem-Hartshorne and Jacob, 2005). Moreover, communication and language abilities are also often impaired. It has been described that only half of 123 CS children more than 4-year-old used verbal or sign language in complete sentences (Hartshorne, 2012).

2.1.4.5 Genitourinary problems

Genital hypoplasia is a common and well-known phenotype of CS (Pinto *et al*., 2005; Wheeler *et al*., 2000; Asakura *et al*., 2008). In most cases, this anomaly is easily recognized in boys (micropenis/cryptorchidism), but females with reduced clitoral size had also been reported (Jongmans *et al*., 2006). In addition, ultrasound examination of females with genetic CS sometimes reveals a hypoplastic uterus

(Jongmans *et al*., 2008). In male patients with clinical CS, the manifestation of micropenis or cryptorchidism is approximately 50-60% (Lalani *et al*., 2012). Genital hypoplasia is suggested to be secondary to hypogonadotropic hypogonadism (Wheeler *et al*., 2000), which during adolescence may result in pubertal delay or arrest in pubertal development. As a consequence, the testes or ovaries do not function properly due to an insufficient level of both luteinizing hormone (LH) and follicle-stimulating hormone (FSH) (Dauber *et al*., 2010).

Moreover, renal anomalies such as renal dysgenesis, hydronephrosis, solitary kidney, and duplex kidneys have been reported and occurred in approximately 25-40% of CS patients (Blake *et al*., 1998; Ragan *et al*., 1999).

2.1.4.6 Ear and hearing problems

The main inner ear abnormality in CS patients is the absence of the lateral semicircular canals. Besides, dysplasia of both the vestibular and semicircular canals, as well as the cochlear parts (Mondini malformation) can also occur, resulting in varying degrees of sensorineural hearing loss (Holcomb *et al*., 2013). These malformations can be detected using computed tomographic (CT) scan or magnetic resonance imaging (MRI) of the temporal bone. The semicircular canals abnormalities are highly suggestive of genetic CS patients. In different studies, 83 out of 85 genetically confirmed patients had external ear malformations (98%), and

56 of 72 patients (78%) had a variable degree of hearing loss (Jongmans *et al*., 2006; Wincent *et al*., 2008; Lalani *et al*., 2006).

External ear abnormalities often involve an abnormal shape and position of the pinnae, such as a wide helix, preauricular tag, reduced vertical height of the pinna and cup-shaped ears (Sanlaville and Verloes, 2007). The ears can protrude from the head and be asymmetric (Lalani *et al*., 2012).

2.1.4.7 Central nervous system and cranial nerve dysfunctions

Central nervous system (CNS) abnormalities have been reported including arrhinencephaly, corpus callosum agenesis and posterior fossa anomalies (Tellier *et al*., 1998). Moreover, abnormalities in CNS can vary, resulting from an alteration of one or more cranial nerves. The alteration of cranial nerve I, resulting in anosmia due to the absence or hypoplasia of the olfactory bulbs. An absent sense of smell is present in almost all CS patients. Abnormality of cranial nerve VII results in facial palsy (reported in 50% of CS cases) and alterations of cranial nerves IX, X, and XI may result in swallowing and aspiration problems (reported in 70-90% of CS cases), and/or gastroesophageal reflux.

Feeding and swallowing difficulties are present in almost all CS cases. Almost 90% of CS children need tube feeding at some time (Dobblesteyn *et al*., 2005). Deformity of cranial nerve VIII may cause hearing loss (Blake and Prasad, 2006; Lalani *et al*., 2012). Among CS patients, 78% had some

form of hearing loss. However, that hearing loss may also result from a variety of inner ear, including deformity of the cochlea, aplasia of ossicles, or absence of the oval window among others (Lalani *et al*., 2012).

2.1.4.8 Behavioral phenotype and sleep-related issues

Several studies have been published concerning the behavioral phenotype of CS patients. Attention-deficit hyperactivity disorders (ADHD), obsessive-compulsive, aggressive, goal-directed persistent, self-abusive, repetitive motor behaviors have been reported in many CS patients (Bernstein and Denno, 2005; Hartshorne *et al*., 2005, Smith *et al*., 2005, Hartshorne and Cypher, 2004).

Sleep cycles in CS children are often disturbed, including problems with initiating/maintaining sleep, breathing, arousal, transition, somnolence, and hyperhidrosis (Hartshorne and Cypher, 2004). Obstructive sleep apnea was found to affect 65% of CS patients (Trider *et al*., 2012).

2.1.4.9 Other potential endocrine problems

It has been reported that one out of nine CS patients had hypothyroidism. It was not possible to ascertain whether it was due to hypothalamic-pituitary dysfunction (Asakura *et al*., 2008). Moreover, secondary hypoadrenalism has been reported in one CS patient (James *et al*., 2003) but was not observed in larger cohorts of subjects (Asakura *et al*., 2008; Khadilkar *et al*., 1999). However, routine analysis of adrenal function

in CS patients has not yet been recommended since there is a lack of evidence.

2.1.4.10 Infections and immune deficiency

Immune deficiency related to dysplasia of the thymus has been reported in CS patients (Corsten-Jansen *et al*., 2013). Recent studies have shown that immune dysfunction is an often-missed complication of CS. A study reported that approximately 60% of CS patients had immune compromise presenting as lymphopenia (Jyonouchi et al., 2009). Other studies have demonstrated similar results in CS patients, displaying T-cell lymphopenia, impaired T-cell function, low immunoglobulins, and the severe T-cell deficiency (Theodoropoulos, 2003). Even though the immune compromise is not considered as a criterion of clinical CS, clinicians should be aware of the high risk of cell-mediated and humoral immunity defects in CS patients. CS patients with severe compromise problem should receive irradiated blood to avoid graft versus host reaction and should not receive live vaccines (Jyonouchi *et al*., 2009).

2.1.5 Clinical Criteria Diagnosis of CHARGE Syndrome

Clinical diagnosis of CS was firstly established by the observation of an association of choanal atresia with anomalies of the heart, eyes, and gastrointestinal tract (Hall, 1979). Two years later, Pagon *et al*. proposed the term CHARGE to describe six cardinal features in CS. The patient may

have CS if displaying at least four anomalies represented in this term (Table 1) (Pagon *et al*., 1981).

Table 1. Pagon's Diagnostic Criteria (Pagon *et al*., 1981)

Clinical Features
Coloboma
Heart Anomalies
Atresia choanae
Postnatal growth deficiency
Retarded development and/or CNS (central nervous system) anomalies
Genital hypoplasia
Ear Abnormalities
Diagnosis of CHARGE if patient has 4 out of 7 criteria
Must have coloboma and/or choanal atresia

In 1998 Blake *et al*. proposed further refinements of the original diagnostic criteria (Table 2) (Blake *et al*., 1998). Although clinical diagnostic criteria have undergone some revision, identification of CS has remained largely subjective, and the probability of finding a mutation depends on the accuracy of the initial clinical diagnosis (Basson and van Ravenswaaij, 2015).

Table 2. Blake's Diagnosis Criteria (Blake *et al*., 1998)

Major	Minor	Occasional
• Coloboma of iris, retina, choroid, disc; microphthalmia • Choanal atresia • Characteristic ear abnormalities • Cranial nerve dysfunction	• Genital hypoplasia • Development delay • CV (cardiovascular) malformation • Growth deficiency • Orofacial cleft • TEF (tracheoesophageal fistula) • Characteristic face	• Thymic/parathyroid hypoplasia • Renal anomalies • Hand anomalies • General appearance • Abdominal defect • Spine anomalies
Diagnosis of CHARGE if patient has: - All 4 major criteria - 3 major and 3 minor criteria - 2 major criteria and several minor criteria		

2.2 Cloning of *CHD7* gene and Mutations in the *CHD7* Gene

To identify the molecular abnormalities in patients with CS, Vissers *et al*. have used array CGH (Comparative Genomic Hybridization) technique to identify microdeletions or duplications underlying CS. Their results evidenced a patient with a microdeletion occurring *de novo* in the 8q1.2 band. This microdeletion was subsequently confirmed by FISH (Fluorescent in Situ Hybridization) technique. Other patients had no deletion in this region, so they sequenced the nine genes located in this region. In these patients, they found the presence of heterozygous point mutations in the *CHD7* gene. The majority of the mutations were nonsense, as well as intron-exon junction mutations with a minority of missense mutations, all appearing *de novo*. Finally, they concluded that the *CHD7* gene is responsible for most cases of CS (Vissers *et al*., 2004).

CHD7 gene (OMIM 608892) is located in chromosome 8 (8q12.2) starting at 61.59 Mb from the p-arm telomere. The gene has a genomic size of 188 Kb and consists in 38 exons, of which the first is non-coding (Blake and Prasad, 2006; Sanlaville and Verloes, 2007; Lalani *et al*., 2012) (Figure 2).

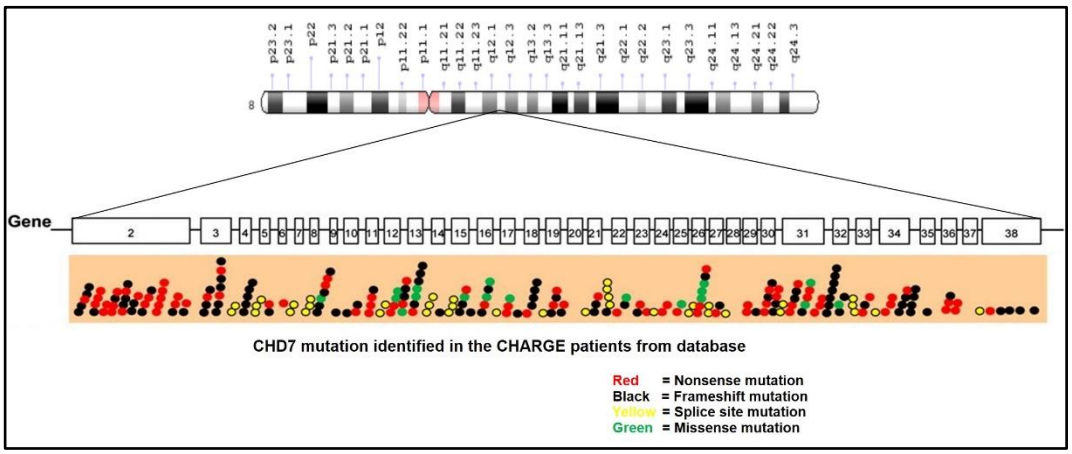

Figure 2. Location of the *CHD7* gene and the mutations identified in CS (Adapted from Balasubramanian *et al*., 2014; https://www.chd7.org).

According to subsequent studies, mutations in *CHD7* gene have been found in about 2/3 of CS cases (Vissers *et al*., 2004; Aramaki *et al*., 2006; Félix *et al*., 2006, Vuorela *et al*., 2007). Therefore, even today, 10-20% of clinical CS patients are not carriers of a mutation in this gene. *CHD7* mutations have also been identified in individuals with diseases that show clinical sign overlapping with CS, including Kallmann syndrome, Omenn-like syndrome, and 22q11.2 deletion syndrome (Ogata *et al*., 2006; Gennery *et al*., 2008; Jongmans *et al*., 2009; Jyonouchi *et al*., 2009).

In the CHD7 mutation database (https://www.chd7.org, last accessed May 03, 2019), 554 pathogenic CHD7 mutations are listed in CS. The majority of the pathogenic CHD7 variants is intragenic mutations. Nonsense and frameshift mutation have found in over 75% of the CS patients. Missense and splice site mutations occur in 19%, while large deletions/duplications, translocation, and small in-frame deletion rarely occur (Figure 3) (Jongmans *et al.*, 2006; Zentner *et al.*, 2010; Jenssen *et al.*, 2012).

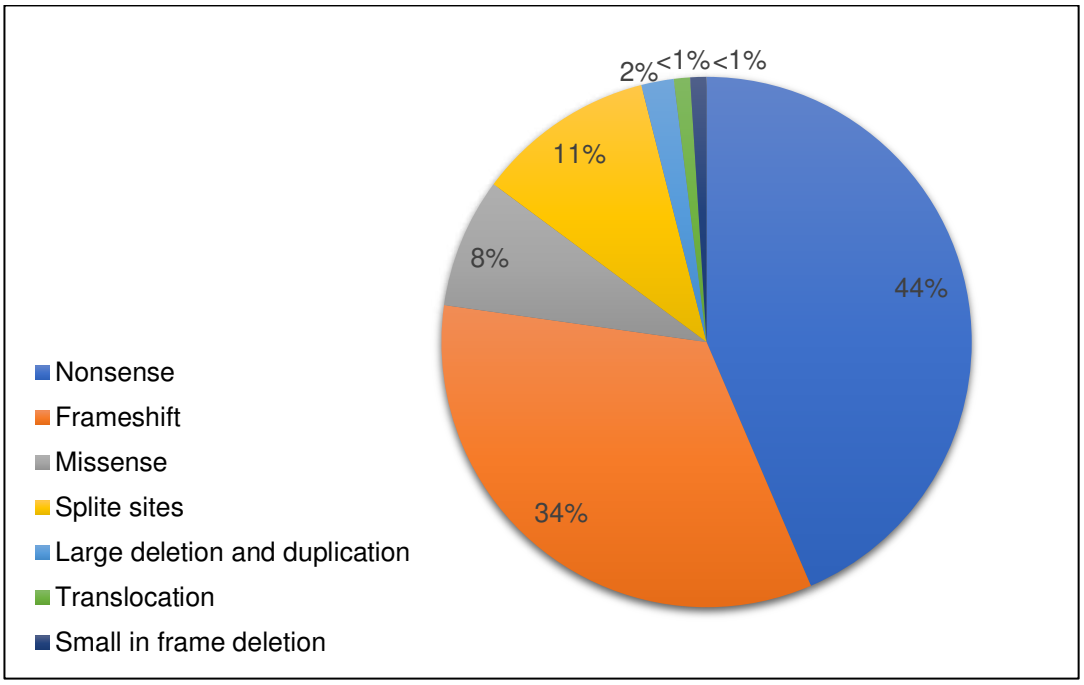

Figure 3. Distribution of pathogenic mutation types in the *CHD7* gene (Adapted from https://www.chd7.org).

Most of the *CHD7* mutations are nonsense or frameshift. They are predicted to elicit the loss of function, and haploinsufficiency is hypothesized to be the pathogenic defect (Bosman *et al.*, 2005; Hurd *et al.*,

2007). Invalidation studies of *CHD7* gene in mouse were made: when the two alleles are deleted, embryonic mortality was observed, while the deletion of a single allele entailed the 'CHARGE-like' symptoms in mouse (Bosman *et al*., 2005; Hurd *et al*., 2007; Bergman *et al*., 2010). There is a statement that both haploinsufficiency and complete loss of CHD7 expression leads to increased DNA methylation of the rRNA promoter, resulting in decreased rRNA expression (Zentner *et al*., 2010).

Even though different types of mutations have been observed in CS patients, it remains unclear whether the type of mutation affects phenotype, but future studies may reveal significance (Lalani *et al*., 2012; Jongmans *et al*., 2006). It has been reported that monozygotic twins with an identical nonsense mutation in exon 16 of the *CHD7* gene exhibit different phenotypes (Lalani *et al*., 2006).

2.3 Updated Diagnosis Criteria for CHARGE Syndrome

After the discovery of *CHD7* mutations as the cause of CS, many groups have proposed some revision and updated criteria for diagnosing CS. In 2005, Verloes introduced renewal clinical criteria: coloboma and choanal atresia were still used as major features, in addition semicircular canal defect was added as a third major item, and gave a formal definition for partial and atypical CHARGE syndromes (Verloes, 2005). Furthermore, to establish a clinical diagnosis of CS, two sets of Blake's and Verloes' diagnostic criteria are still used in a clinical setting with some update and

revision by Blake in 2006, Sanlaville and Verloes in 2007, and by consortium in 2012 (Table 3) (Blake *et al*., 2006; Sanlaville and Verloes, 2007; Lalani *et al*., 2012).

Table 3. Diagnosis Criteria for CS defined by Verloes (2005) updated by Blake (2006), Sanlaville and Verloes (2007), and by Hale (2016)

Criteria	Verloes	Hale
Major	• Coloboma • Choanal atresia and/or cleft lip or palate • Semicircular canals agenesis/hypoplasia • Arhinencephaly and/or anosmia	• Coloboma • Choanal atresia or cleft palate • Abnormal external, middle or inner ears, including hypoplastic semicircular canals • Pathogenic CHD7 variant
Minor	• Cranial nerves VII to XII palsy • Hypothalamo-hypophyseal dysfunction • External- or middle-ear anomalies • Heart defects or esophageal anomalies • Intellectual disability	• Cranial nerve dysfunction • Dysphagia/feeding difficulties • Structural brain anomalies • Developmental delay/ID/Autism • Hypotalamo-hypophyseal dysfunction (gonadotropin or growth hormone deficiency) and genital anomalies • Heart or esophagus malformation • Renal anomalies skeletal/limb anomalies
Inclusion rule	• Typical CHARGE: 3 major or 2 major + 2 minor • Partial CHARGE: 2 major + 1 minor • Atypical CHARGE: 2 major + 0 minor or 1 major + 3 minor	2 major + any number of minor

Recently, Hale *et al*. proposed to include the pathogenic *CHD7* variant to major criteria and skeletal/limb anomalies to the minor criteria (Table 3) (Hale *et al*., 2016).

2.4 Genetic Causes of CHARGE Syndrome

The genetic cause of CS remains unclear in 5-10% of typical CHARGE patients and 40-60% of patients suspected of CS (Bergman *et al*., 2011; Jongmans *et al*., 2006; Lalani *et al*., 2006). Non-detectable rearrangements in *CHD7* gene (e.g., deep intronic mutations that affect splicing, intra-genic rearrangements or mutations in regulatory regions), and whole gene or exon deletions/duplications may explain why CS occurred in some of these patients. Another possibility is that there are other genes responsible for CS. One of the genes that have been proven involved in the pathogenesis of CS is *SEMA3E* gene (OMIM 608166) (Ufartes *et al*., 2018). This gene was found to be mutated in one CS patient and disrupted in another patient with a *de novo* chromosomal translocation between chromosomes 2 and 7. Moreover, there was no *CHD7* mutation detected in these patients (Lalani *et al*., 2004). Thus far, no additional *SEMA3E* mutation has been reported in CS patients. Furthermore, *EFTUD2* gene (OMIM 603892) has been recommended to be analyzed in individuals with atypical CS who do not carry *CHD7* mutation (Luquetti *et al*., 2013). In 2017 Legendre *et al*. detected rare CS cases that were linked to alterations in *EFTUD2* gene (Legendre *et al*., 2017). Other candidate genes have also been studied without revealing any pathogenic mutations, e.g. *PITX2* (OMIM 601542)

and *PAX2* (OMIM 167409) in 29 and 34 patients with CS, respectively (Martin *et al*., 2002; Tellier *et al*., 2000).

Phenocopies due to chromosomal imbalances have been reported in patients with a CHARGE-like phenotype. Unfortunately, most cases were published before 2004 so that *CHD7* analysis was not performed. These chromosomal aberrations are duplication in 1(q25q32) and deletion in 4(q31qter) (Dev *et al*., 1985; Shroff *et al*., 1981).

In contrast to the unique chromosomal cases, a recurrent clinical overlap has been reported for 22q11.2 deletion syndrome and CS (Randall *et al*., 2009; Bergman *et al*., 2011; Gennery *et al*., 2008; Inoue *et al*., 2010; Sanka *et al*., 2007; Writzl *et al*., 2007; Devriendt *et al*., 1998; Emanuel *et al*., 1992). The overlapping clinical features include cleft palate, cardiac malformations, ear abnormalities, hearing loss, growth deficiency, developmental delay, renal abnormalities, hypocalcaemia and immune deficiency (Randall *et al*., 2009; Inou *et al*., 2010; Sanka *et al*., 2007; Writzl *et al*., 2007; Devriendt *et al*., 1998; Lonlay-Debeney *et al*., 1997; Jyonouchi *et al*., 2009).

CHD7 mutations are more often, but not exclusively, associated with coloboma, choanal atresia, facial nerve palsy, trachea-esophageal fistula, and micropenis compared to 22q11.2 deletions (Jyonouchi *et al*., 2009). Hypoplastic semicircular canals are suggestive for CS, as they are present in almost all CS patients (Bergman *et al*., 2011; Verloes, 2005; Admiraal *et al*., 1998; Amiel *et al*., 2001). However, semicircular canal abnormalities

cannot exclude 22q11.2 deletion syndrome since this feature has also been described in patients with a 22q11 deletion, albeit very rarely (Bergman et *al*., 2011; Hopsu *et al*., 2007). Defects of the lateral semicircular canals were also noted in a mouse model for 22q11.2 deletion syndrome, the *Tbx1* (+/-) mouse (Randall *et al*., 2009).

In summary, *CHD7* is the major causative gene in CS. If sequence analysis does not reveal a *CHD7* mutation, genome-wide array studies should be performed in patients suspected of CS.

2.5 CHD7 Protein and Its Function

The chromodomain helicase DNA-binding (CHD) proteins are involved in ATP-dependent chromatin remodeling. It is a group of nuclear proteins with nine members in vertebrates (Micucci *et al*., 2015). Generally, CHD proteins are classified into three subfamilies based on their constituent domains: subfamily I (CHD1 and CHD2), subfamily II (CHD3, CHD4, and CHD5) and subfamily III (CHD6, CHD7, CHD8, and CHD9) (Hall and Georgel, 2007; Marfella and Imbalzano, 2007).

CHD7 protein, consisting in 2997 amino acids with a molecular weight of 340 kDa, is localized in both nucleoplasm and nucleolus (Zentner *et al*., 2010; Kita *et al*., 2012). The CHD7 protein is characterized by two chromodomains in N-terminal, a SNF2/SWI domain, a helicase domain, a SANT-like domain (Switching-defective protein 3, Adaptor 2 Nuclear receptor corepressor, Transcription factor IIIB) and two paired BRK

(Brahma and Kismet) domains in C-terminal (Figure 4) (Hall and Georgel, 2007; Marfella and Imbalzano, 2007). Chromodomains are involved in the recognition of lysine-methylated histone tails and DNA (and RNA) targets. This domain plays a role by mediating chromatin interactions in a variety of different protein context. SWI/SNF2 domains are characterized by DNA-dependent ATPase activity, involved in regulating the structure of chromatin. Helicase domains are critical in DNA strand separation during replication, repair, recombination, and transcription. It has been proposed that the SANT domain plays a crucial role in the interaction between the connection ends of the histones and enzymatic catalysis involved in nucleosome remodeling. The function of BRK domains is unknown, but they are usually found in association with chromodomains (Dirscherl and Krebs, 2004).

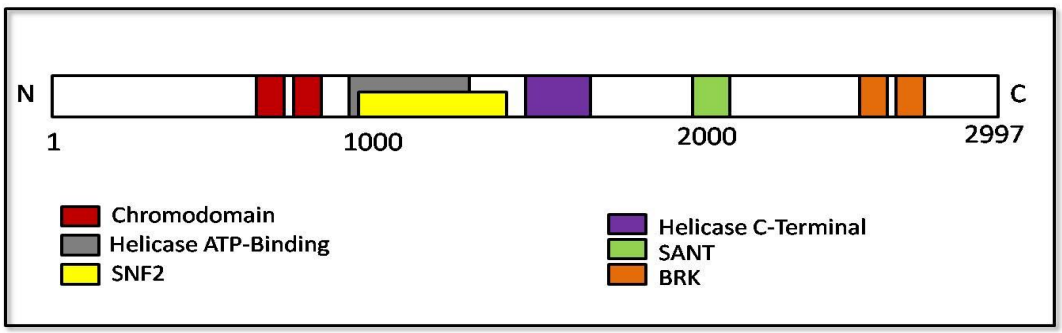

Figure 4. Schematic representation of the domain organization of the CHD7 protein (Adapted from http://www.ebi.ac.uk/interpro/protein/Q9P2D1).

The CHD7 protein is highly conserved in other species and several orthologs have been found in xenopus, zebrafish, mouse, and chicken (Aramaki *et al*., 2007; Bajpai *et al*., 2010; Bosman *et al*., 2005). Several

studies had been conducted in homozygous *Chd7* mutant mice; the results showed that the mice do not survive beyond an early embryonic stage (Hurd *et al*., 2007; Randal *et al*., 2009). This showed that CHD7 function is critical. Furthermore, recent studies have suggested that CHD7 plays a role in controlling gene expression program by ATP-dependent chromatin remodeling (Schnetz *et al*., 2009; Layman *et al*., 2010; Bajpai *et al*., 2010).

These studies have discussed *in vitro* experiment insights into CHD7 function and its mechanism. In 2009, Schnetz *et al*. mapped the distribution of CHD7 protein using chromatin immunoprecipitation coupled with microarray (ChIP-chip) approach. They used human colorectal carcinoma cells, human neuroblastoma cells, and mouse embryonic stem (ES) cells before and after differentiation into neural precursor cells. They described that CHD7 localizes to discrete locations along chromatin that are specific to each cell type, and that the cell-specific binding of CHD7 was shown to match the binding sites of histone H3 methylated on lysine 4 (H3K4me). One year later, Zentner *et al*. using ChIP-chip or massively parallel DNA sequencing (ChIP-seq) analyses demonstrated that CHD7 specifically associated with hypomethylated, active rDNA. They showed that siRNA-mediated depletion of CHD7 results in hypermethylation of the rDNA promoter and in a concomitant reduction of 45S pre-rRNA levels. Moreover, Kita *et al*. confirmed that overexpression of full-length protein CHD7L, as well as of CHD7S, a short isoform of CHD7 that is generated by alternative splicing of exon 6, resulted in an increase in 45S pre-rRNA level (Kita *et al*.,

2012). These results indicate that CHD7 protein functions as a transcriptional regulator in the nucleoplasm and that the protein is constitutively localized to the nucleolus, the site of rRNA transcription (Schnetz *et al.*, 2009; Zentner *et al.*, 2010).

Several studies described that CHD7 cooperates with other transcription factors such as *Oct3/4*, *Sox2*, *Nanog*, and *p300* for the development of neural stem cells in mouse ES cells (Engelen *et al.*, 2011; Schnetz *et al.*, 2010). Moreover, interactions between Chd7 and Chd8 have been demonstrated using a yeast two-hybrid library screen: Chd8 was found as a binding partner of Chd7 and disruption of this interaction could be involved in a pathomechanism of CS (Batsukh *et al.*, 2010). Although CHD7 has been involved in transcriptional activation or suppression of tissue-specific genes throughout differentiation, it has remained unclear whether changes in such gene expression are sufficient to give rise to CHARGE syndrome (Layman *et al.*, 2009; Bajpai *et al.*, 2010; Hurd *et al.* 2010).

Studies on the function of CHD7 have suggested a role in controlling the programming of gene expression by ATP-dependent chromatin remodeling in embryonic stem cells and other cell types (Schnetz *et al.*, 2009; Bajpai *et al.*, 2010; Basson and van Ravenswaaij, 2015). CHD7 intervenes at two levels: first, by binding to histone H3 methylated at lysine 4 in the promoter regions of genes, depending on the developmental stage and cell type, and it would then play a role of transcription activator (Schnetz

et al., 2009). Second, it is involved in the upregulation of rRNA synthesis in the nucleolus in a tissue-specific manner (Zentner *et al*., 2010). An explanation for CHD7 influence upon rRNA concentrations is as follows: CHD7 is assumed to have an activity of chromatin remodeling because of its SNF2/helicase type domains. It was studied *in vitro* by Bouazoune and Kingston in 2012. Thus, a likely scenario is that CHD7 initiates or maintains the opening of chromatin regions containing repeated sequences that are transcribed into rRNA, to promote the association of factors involved in the transcription of the rDNA. Moreover, the latest studies have shown that overexpression of the wild-type CHD7 protein causes an increase in the expression of the 45S rRNA precursor (Zentner *et al*., 2010; Kita *et al*., 2012).

Another study in *Xenopus* embryos has shown that CHD7 cooperates with transcription factors Sox9, Twist, and Slug in the migration of neural crest cells. Moreover, CHD7 and PBAF (polybromo- and BRG1-associated factor containing complex) bind together to the neural crest cell-specific regulatory elements of *TWIST1* (OMIM 601622) and *SOX9* (608160). This cooperation promotes neural crest gene expression and cell migration (Bajpai *et al*., 2010).

Other studies in a mouse model described that Chd7 is required for the differentiation of neural stem cells in the basal olfactory epithelium (Layman *et al*., 2009). Chd7 is also necessary for the proliferation of neuroblasts and the expression of several genes that are involved in inner

ear morphogenesis such as *Ngn1*, *Itx2*, and *Fgf10* (Hurd et al., 2010). Another study has suggested that Chd7 affects GnRH (Gonadotropin-releasing hormone) neurogenesis and signaling by influencing the transcriptional regulation of target genes involved in the BMP and FGF pathways (Layman *et al*., 2011). Furthermore, several studies have shown that Chd7 cooperates with Sox2 in activating the expression of Sonic Hedgehog (*Gli2*, *Gli3*, *Mycn*, and T*ulp3*) and Notch pathway (*Jag1*, *Rbpj*, *Hes5*) (Engelen *et al*., 2011; Puc and Rosenfeld, 2011). Moreover, CHD7 protein regulates the expression of genes during embryonic development in a tissue-specific and stage-specific manner (Layman *et al*., 2010) (Figure 5). Together, these studies allow the emergence of a global pattern concerning the function of CHD7 and its interaction with other proteins.

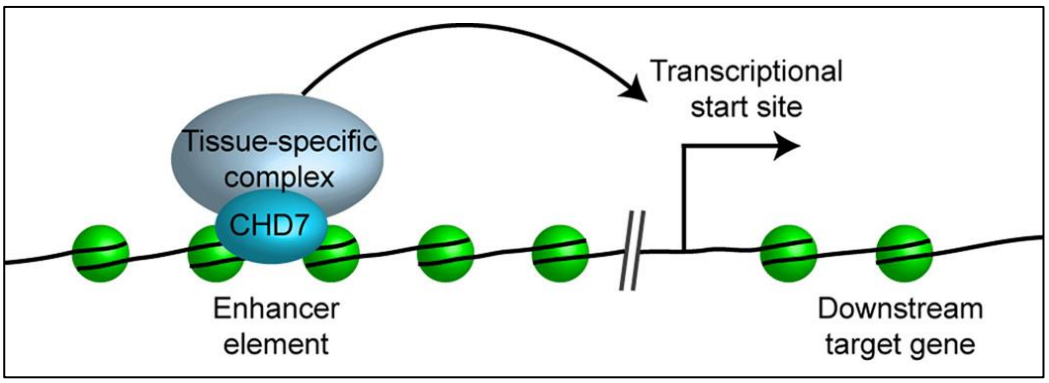

Figure 5. The function of CHD7. According to recent studies, CHD7 protein has been shown to have a tissue-specific and stage-dependent function in enhancer-mediated transcription (Layman *et al*., 2010).

2.6 Pathomechanism of CHARGE Syndrome

The reason for the phenotypic variation in CS remains unknown. A hypothesis is that the *CHD7* gene influence over a broad variety of other genes may explain the variability of clinical features in CS (Basson and van Ravenswaaij, 2015). CHD7 have an effect on several signaling pathways in control development. The developmental alterations involve the midline structures of the body and affect the craniofacial structures. Despite the fact that the roles of *CHD7* gene in the development of the eye, olfactory epithelium, inner ear, and vascular tissues have been demonstrated, the variability of expression of this gene is not yet fully understood (Sanlaville and Verloes, 2007). In a fetus harboring a *CHD7* mutation, the problems begin with arrest in embryologic differentiation at the first trimester (between the third- and ninth-week post conception) (Blake and Prasad, 2006; Verloes, 2005). Several malformations including eye, ear, and cranial nerve occur between days 33-34 of gestation. Moreover, conotruncal heart malformations and abnormalities in cephalic neural crest cell migration occur between the fourth and the fifth weeks post-conception. Failure of the primitive bucconasal membrane to rupture between the fifth- and sixth-weeks post conception results in choanal atresia (Blake and Prasad, 2006).

The pathomechanisms underlying CS are briefly suggested: (1) disruption of the neural crest cells development, (2) alteration of the interaction between neural crest cells and mesoderm, and (3) alteration of

the interaction between mesenchymal and epithelial cells (Williams, 2005; Moccia *et al.*, 2018).

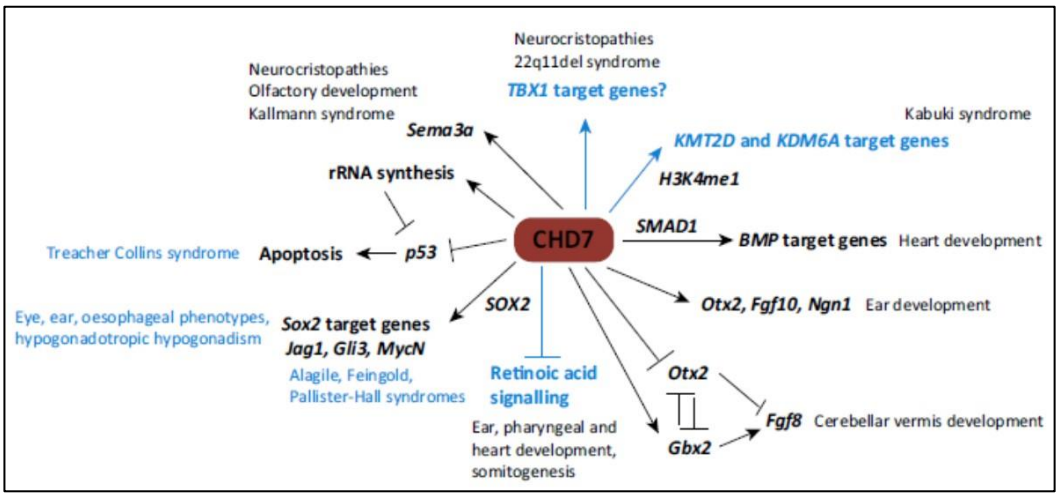

Figure 6. Developmental and disease-associated pathways regulated by *CHD7*. Studied interactions are in black and presumed hypothetical associations are in blue (Basson and van Ravenswaaij, 2015).

CHD7 is a key regulator of some developmental transcription factors (TFs) genes and control the activity of several signaling pathways (Figure 6). CHD7 controls *BMP4* expression or cooperate with *SMAD1* to control growth factor BMP pathway genes involved in heart development (Liu *et al.*, 2014; Jiang *et al.*, 2012). CHD7 protein plays a role as a positive regulator of transcription factor *Otx2* expression during ear development, a decrease of expression of this gene resulting in downregulation of growth factor *Fgf8* expression and cerebellar vermis hypoplasia during early cerebellar development (Hurd *et al.*, 2010; Yu *et al.*, 2013).

CHD7 inhibits retinoic acid signaling in neuronal progenitor cells, so the loss of function of CHD7 leads to hyperactive retinoic acid signaling

resulting in inner ear defects (Micuci *et al*., 2010). Moreover, CHD7 interacts with transcription factor SOX2 in neural stem cells. The *SOX2* gene controls some genes that regulate notch and hedgehog signaling pathways and cell proliferation pathways such as *Jag1*, *Gli3*, and *Mycn* (Engelen *et al*., 2011). A study reported that *de novo* mutations in *SOX2* gene results in some CS phenotypes including eye, external ear abnormalities or deafness, esophageal atresia, and hypogonadotropic hypogonadism (Numakura *et al*., 2010).

A recent study described that transcriptionally inactive variant of p53 during mouse development resulted in typical CHARGE phenotypes (Van Nostrand *et al*., 2014). Moreover, they demonstrated evidence that CHD7 can repress *p53* gene expression, otherwise in *Chd7*-null mouse neural crest cells and fibroblast from CS patients showed increased p53 signaling. Accordingly, haploinsufficiency of CHD7 leads to defects in ribosome biogenesis via p53 pathway hyperactivation and affects rDNA transcription (Van Nostrand *et al*., 2014; Zentner *et al*., 2010). CHD7 also regulates SEMA3E expression, which may be involved in neural crest cell migration and olfactory development (Schutz *et al*., 2014).

CHARGE and 22q11deletion syndromes have been known to show clinical overlap, presumably due to the interaction between CHD7 and TBX1 that involved in a molecular pathway or both CHD7 and TBX1 controlling the same genes in affected organs (Randal *et al*., 2009). CHARGE and Kabuki syndromes also show significant clinical overlap.

Mutation of *KMT2D* or *KDM6A* gene has been associated with Kabuki syndrome (KS) (Wessels *et al*., 2002). These genes encode proteins that play a role as histone methyltransferase and as histone demethylase, respectively. KS shows significant phenotype overlapping with CHARGE because both KS-associated genes might regulate CHD7 recruitment to histone H3 methylated at lysine 4 (H3K4me1)-marked enhancer regions. In mammalian cells, H3K4me1 is an epigenetic mark that plays a role at enhancers to regulate the gene expression (Local *et al*., 2018). Therefore, *CHD7*, *KMT2D*, and *KDM6A* taken together are assumed to control the same target genes (Schulz *et al*., 2014).

2.7 Novel Classification System to Predict the Pathogenicity of *CHD7* Missense Variants and Prospective

Although the missense mutations in the *CHD7* gene occurs only in 8% of all CS cases, their challenging interpretation results in difficulties for genetic counseling. Furthermore, Bergman *et al*. have developed a novel algorithm to predict the pathogenicity of *CHD7* missense variants. They combined the results of two computational algorithms (PolyPhen-2 and Align-GVGD), prediction of the structural model of CHD7 protein and phenotypic data (Table 4).

Table 4. A novel algorithm to predict the pathogenicity of *CHD7* missense variants (Bergman *et al.*, 2012)

Computational algorithms (summed score between 0 and +2)
Polyphen-2: benign = 0, possibly damaging = +0.5 and probably damaging = +1
Align-GVGD: C0 = 0, C15/C25/C35 = +0.5 and C45/C55/C65 = +1
Structural model (summed score between -1 and +1)
Minor effect = -1, undetermined effect = 0, detrimental effect or located close to the ATP binding site = +1
Segregation analysis (summed score between -10 and +4)
Variant occurred *de novo* in one patient with features of CHARGE syndrome = +3 OR Variant occurred at least twice *de novo* in patients with features of CHARGE syndrome = +4
Asymptomatic carrier of the variant[a] = -2
Variant found in a homozygous state = -5
Variant found in combination with a pathogenic *CHD7* mutation[b] = -3
Prediction based on total summed score (total score between -11 and +7)
Probably benign: total score 0 or less **Unknown:** total score between 0 and +4 **Probably pathogenic:** total score +4 or more

[a] This means that the variant is present in one or more clinically well-characterized persons without features of CHARGE syndrome, or the variant is found in two or more persons reported to be normal, but for whom no detailed clinical information is available (e.g., controls reported in the NCBI SNP database or not thoroughly investigated family members).

[b] A pathogenic *CHD7* mutation is defined as a truncating, missense, or splice site mutation in the *CHD7* gene that is clearly pathogenic (this category does not include unclassified variants or benign variants).

The combination of these variables leads to a more confident pathogenicity prediction than previously available methods (Bergman *et al*., 2012). They used this system to classify 145 *CHD7* missense variants. Furthermore, they compared the clinical phenotypes of patients harboring the missense variants that were classified as probably pathogenic with the phenotypes of patients harboring truncating mutations. They concluded that *CHD7* missense alterations are generally associated with a milder phenotype than truncating alteration (Bergman *et al*., 2012).

Although computational tools have been developed to classify the pathogenicity of missense variants, functional analyses to confirm their pathogenicity are lacking (Bergman *et al*., 2012; Villate *et al*., 2018). Interpreting the effect of amino acid variations in CHD7 protein remains a challenge. Moreover, the expression of the large CHD7 protein (2997 amino acids, 340 kDa) to perform functional assay may be difficult due to an increased potential of degradation.

A biochemical method has been described to test *in vitro* the function of the CHD7 protein: first, the authors purified CHD7 protein expressed in insect cells using a baculoviral vector. Then, they used these recombinant proteins to study *in vitro* their chromatin remodeling properties (Bouazoune and Kingstone, 2012). Two years later, another approach has been

developed in a zebrafish model. *Chd7* knock-down was achieved by injection of splice-blocking morpholino (MO) against *chd7* (*chd7*-Mo). Then, phenotype rescue was studied following injection of mRNA encoding WT-CHD7 or mutated alleles (Balasubramanian *et al*., 2014). Moreover, DNA methylation signature was recently proposed as a diagnosis tool for disorders involving proteins that attend the epigenetic machinery (Butcher *et al*., 2017; Aref-Eshghi *et al*., 2018). However, the complexity of these systems is incompatible with routine use. Therefore, in this thesis, we developed a novel and simple approach to accurately assess the functional impact of CHD7 variants.

CHAPTER III

MATERIALS, PATIENTS AND METHODS

3.1 Molecular Diagnosis of CHARGE Syndrome

3.1.1 Chromosome Analysis

Chromosome analysis was conducted at the laboratory of Central Biomedical Research (CEBIOR), Diponegoro University, Semarang, Indonesia using GTG-banding technique from peripheral blood lymphocytes. Ten drops of heparinized blood were cultured under an atmosphere of 5% CO_2 at 37°C for 72 hours in two different 5 mL media, TC199 and Minimum Essential Media (MEM) (Gibco, Carlsbad, California, USA), supplemented with 5% Fetal Bovine Serum (FBS) and 0.025 mL Phytohemaglutinin-P (Gibco). In a MEM media tube, 0.1 mL thymidine (final concentration of 0.3 µg/mL) and 3 drops of colchicine (concentration of 1 µg/mL) were added and incubated for 24 hours and 25 minutes before cell harvesting. A TC199 tube was treated with colchicine (concentration of 1 µg/mL) but without thymidine. After the culture process was finished, the culture tube was centrifuged at 1,000 rpm for 10 minutes, and the supernatant was removed. A warm hypotonic solution (0.075 M KCl) was added to the cell pellet and then resuspended until homogeneity and subsequently incubated at 37°C in a water bath for 15-30 minutes. Afterward, the cell suspension was centrifuged at 1,000 rpm for 10 minutes, the supernatant was removed, and 5 mL Carnoy's solution (3:1 methanol:glacial acetic acid) was added slowly through the tube wall, and shaken well. These steps were repeated constantly until clear precipitation was obtained and fresh Carnoy's solution was added to suspended

residues. Subsequently, two drops of cell suspension were released onto a glass slide from a height of around 20 cm.

Finally, the slides were stored for approximately 3 days. After that, the slides were rinsed with water and put into warm Hanks solution, put into 0.1% trypsin (in warm Hanks buffer) for 10-25 seconds, depending on the sensitivity and slide age, and again rinsed with water. After that, the slide was flooded with 10% Giemsa staining in phosphate buffer PH 6.8 for 1 minute.

3.1.2 DNA Extraction

DNA extraction was performed at the laboratory of Central Biomedical Research, Diponegoro University, Semarang, Indonesia using a salt saturation method. Genomic DNA was isolated from peripheral blood of patient. Two mL of blood were resuspended in 5-10 mL lysis buffer (155 mM NH_4Cl, 10 mM NH_4CO_3, 1 mM EDTA), incubated for 10-30 minutes at room temperature and centrifuged at 3,500-5,000 rpm for 5 minutes. The supernatant was removed, and lysis buffer was added again. These steps were repeated three times. Pellets were resuspended with 2 mL of TE buffer, 10mg/mL Proteinase-K and 100 µL of 10% sodium dodecyl sulfate (SDS) and then incubated overnight in a water bath at 50°C. Subsequently, 6 M NaCl approximately one-third of the volume of the tube was added to the suspension and centrifuged at 4,000 rpm for 10 minutes. DNA was precipitated by adding absolute ethanol. DNA that looked like white

substance was removed by a fine needle. To wash the DNA, 70% of ethanol drops were added and left to dry out. The DNA was transferred to a sterile Eppendorf tube with 250 µL of TE buffer and let to dissolve overnight at room temperature. Purified DNA was stocked at 4°C.

3.1.3 Targeted NGS Gene Panel

The amount of DNA and its purity were measured by spectrophotometric absorbance reading at both 260 nm and 280 nm using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed on the instrument PGM Hi-Q View Sequencing (Thermo Fisher Scientific). To achieve targeted gene panel sequencing, we used an Ion AmpliSeq™ (Thermo Fisher Scientific) *CHD7*, *HOXA1*, and *EFTUD2* gene panel, containing 209 primer pairs in two pools. Multiplex polymerase chain reaction (PCR) was performed using 10 ng/µL genomic DNA in a final volume of 10 µL, with a premixed primer pool and Ion AmpliSeq™ HiFi master mix (Ion AmpliSeq™ Library Kit 2.0) for 2 minutes at 99°C, followed by 19 cycles at 99°C for 15 seconds and 60°C for 4 minutes, ending with a holding period at 10°C. The PCR amplicons were treated with 2 µL of FuPa reagent to partially digest primer sequences and phosphorylate the amplicons at 50°C for 10 minutes, followed by 55°C for 10 minutes, then 60°C for 20 minutes. The amplicons were ligated to adapters with the diluted barcodes of the Ion Xpress™ Barcode Adapters kit (Life Technologies) for 30 minutes at 22°C, then 72°C for 10 minutes. Adaptor-ligated amplicon libraries were purified using Agencourt AMPure

XP reagents (Beckman Coulter). Next, we quantified the library on Qubit instrument (Thermo Fisher Scientific), and then realized dilutions at 100 pM in TE buffer. After obtaining the library, the next step was performed on the One Touch 2 device. PCR was performed using the Ion OneTouch™ System and Ion OneTouch™ 200 Template Kit v2 (Life Technologies) according to the manufacturer's instructions. Template-positive Ion Sphere™ Particles were then enriched with Dynabeads MyOne™ Streptavidin C1 Beads (Life Technologies) using an Ion OneTouch™ ES system (Life Technologies). Purified Ion Sphere particles were loaded on Ion 316 Chip V2. Sequencing was carried out on a Personal Genome Machine (PGM) sequencer (Ion Torrent™) using the Ion PGM™ Sequencing 200 Kit V2 according to the manufacturer's instructions. The variants were evaluated using the Alamut Visual 2.11 software (Interactive Biosoftware, Rouen, France).

3.1.4 Mutation Confirmation by Sequencing Analysis

The method used is derived from that described by Sanger in 1977. The incorporation, during the reaction, of four dideoxyribonucleotides (ddNTPs) labeled with four different fluorochromes results in stopping of the elongation. Sequencing analysis of *CHD7*, targeting the exon 34, was performed using the primers *CHD7*-F 5'-GCCAGCCCATATAGCAGTAC-3' and *CHD7*-R 5'-AACACAGCCCAGCATCGTGA-3'. Approximately 20 ng of DNA solution (2.5 µL) was added to 22.5 µL of PCR mixture. This PCR mixture contained 0.25 µL of 25 mM deoxyribonucleotide triphosphates

(dNTPs), 3 µL of 25 mM $MgCl_2$, 0.25 µL of 20 µM for each primer, 2.5 µL of 10X PCR buffer, 0.125 µL of 5 U/µL Diamond® high fidelity *Taq* DNA polymerase (Eurogentec, Seraing, Belgium), and 16.13 µL of H_2O. PCR was initiated with denaturation at 94°C for 5 minutes, followed by 35 PCR cycles (at 94°C for 30 seconds, 58°C for 30 seconds, and 72°C for 30 seconds) and 7 minutes final elongation at 72°C. The amplified products were detected by electrophoresis on a 1.5% agarose gel with 0.5 mg/mL ethidium bromide and visualized under ultraviolet (UV) light. Furthermore, 5 µL of the PCR product was cleaned-up with 2 µL ExoSAP reagent (ThermoFisher) to remove excess primers and unincorporated nucleotides enzymatically. Hereafter, samples are incubated at 37°C and 80°C for 15 minutes, consecutively. Finally, 2 µL of the PCR product was used for the sequence reaction (BigDye Terminator Cycle Sequencing Kit Version 3.3; Applied Biosystems), on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems), following the manufacturer's directions. Sequencing was performed bidirectionally using the forward and reverse PCR primers. The obtained sequence was compared to a published reference nucleotide sequence (RefSeq CHD7 NM_017780.3) using Chromas software 2.6.4.

3.2 Development of Functional Assay of CHD7 Variants

3.2.1 Patients and Bioinformatic Prediction Tools

The index cases were referred to our laboratory because of clinical features characteristic of CS according to Verloes' updated criteria. *CHD7* mutation screening was realized as previously described (Bilan *et al.*, 2012;

Legendre *et al*., 2017). Nucleotide RefSeq CHD7 NM_017780.3 and exon numbering NG_007009.1 were used. Pathogenicity of *CHD7* variants was classified according to the American College of Medical Genetics (ACMG) recommendations (Richards *et al*., 2015). To investigate the potential impact of missense mutations upon CHD7 protein, we have used 10 software embedded in the VarSome tool (https://varsome.com/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), and Align-GVGD (http://agvgd.hci.utah.edu). To establish the potential impact on the splicing mechanism, missense variants were studied by using Human Splicing Finder (HSF) software (http://www.umd.be/HSF3/).

Familial CHARGE with c.5944G>T or p.(Gly1982Trp) variant: the female index patient (9 years old) presented with a typical form of CS (notably semi-circular canal hypoplasia, atresia of the choanae, tetralogy of Fallot and typical cup-shaped ears). This variant was found in her affected brother (intellectual deficiency, unilateral deafness, cryptorchidism) and was not found in her mother. Her late father presented with unilateral deafness; unfortunately, *CHD7* molecular analysis had never been done. Nevertheless, her paternal uncle (53 years old) who also harbor this variation presented with a typical CS form (complete clinical features were previously reported by Legendre *et al*., 2017). Gly1982Trp variation affects the SANT (switching-defective protein 3, adaptor 2, nuclear receptor co-repressor, transcription factor IIIB) domain of the protein and is predicted pathogenic by all software. The variation was not found in any public

database. Taken together, these data sustain a class 3 (uncertain significance) variation. It could be noticed that HSF predicted no significant impact upon the splicing mechanism.

Patient harboring c.6955C>T or p.(Arg2319Cys): this young male baby (9 months old) presented with a partial CS form with a bilateral choanae atresia, right semi-circular canal hypoplasia, and typical dysplastic ears. Ophthalmologic, cardiac and abdominal examinations were normal. This variant has been previously published in two studies (Félix *et al*., 2006; Jongmans *et al*., 2006). This missense variant is predicted pathogenic by all software and arose *de novo*. It is reported one time as pathogenic in ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), five times in the CHD7 database (http://www.chd7.org) and is not reported in control databases like Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS/) or gnomAD (http://gnomad.broadinstitute.org/). Moreover, another variant affecting the same amino acid p.(Arg2319Ser) is also reported twice as pathogenic in ClinVar and twice in CHD7 database. Taken together, these data are consistent with p.(Arg2319Cys) belonging to class 5 (disease-causing) variation. HSF predicted no significant impact upon the splicing mechanism.

Patient harboring c.3905T>C or p.(Leu1302Pro): this male patient (13 years old) presented with a typical form of CS including semi-circular canal hypoplasia, cleft lip and/or palate, heart defect, intellectual disability, pituitary defect, hypogonadism, genital anomaly, deafness, ear anomaly,

and kidney anomaly. Clinical features of the patient have been previously reported (Legendre *et al*., 2017). This variant, localized in Helicase C domain, is predicted pathogenic by all software and arose *de novo*. It was reported once as pathogenic in the CHD7 database and was not reported in the control databases like EVS or gnomAD. All these data are consistent with a class 5 (disease-causing) variation. HSF predicted no significant impact upon the splicing mechanism.

Our laboratory has previously described familial cases displaying c.5405-17G>A variation, which lies in a recurrent hotspot of mutation in intron 25 of *CHD7* (Legendre *et al*., 2018). Family one: the male index patient (5 years old) born from unrelated parents. He had a typical CS with 2 major criteria (uveo-retinal coloboma and semi-circular canal agenesis) and 3 minor criteria. His affected mother had unilateral ptosis and moderate myopia, mild facial asymmetry, lachrymal canal anomaly, and partial syndactyly on the left foot. Family two: the male patient born from unrelated parents. Typical CS was diagnosed in early childhood as he had 2 major criteria (microphthalmia with coloboma and semi-circular canal hypoplasia) and 4 minor criteria. The alteration was inherited from a mildly affected mother who had unilateral deafness, uropathy and congenital dislocation of the hip (complete clinical features were previously published by Legendre *et al*., 2018). This variant is also reported twice as pathogenic in ClinVar and five times in the CHD7 database. We showed by minigene assay that

it elicits a splicing defect, leading to the synthesis of a protein harboring a five-amino acid insertion, p.(His1801_Gly1802insAspGlyHisGlyThr).

3.2.2 Plasmids

In this study, we transfected various cell lines with different plasmids to overexpress CHD7. The pCI-neo, plasmid encoding untagged wild-type CHD7 protein, and pCIneo-CHD7-HA, plasmid encoding CHD7 tagged with HA at the C-terminus, are a generous gift from Cynthia Bartels and Peter Scacheri (Case Western Reserve University, Cleveland, Ohio, USA). The other plasmid, pcDNA3-FLAG-CHD7, given by Dr. Nakayama (Kyushu University, Fukuoka, Japan), encodes CHD7 tagged with FLAG at the N-terminus. Plasmid DNA is purified using the QIAprep Spin Miniprep or Maxiprep Kits (Qiagen) according to the supplier's instructions. Plasmids were assayed by measuring their absorbance at 260 nm with a spectrophotometer BIOMATE 3 (Thermo Fisher Scientific).

3.2.3 Site-Directed Mutagenesis

To go further in the understanding of amino acid change consequence upon CHD7 function, we modified *CHD7* cDNA sequence in order to analyze the effect of these amino acid changes by introducing the mutated gene in cell lines, which then will express the variant protein. We used a site-directed mutagenesis method called circular mutagenesis to create mutations exactly where we want for each mutation. We first designed pairs of 25-45 bases oligonucleotides that were complementary to the target

sequence, with the changed nucleotide in the center position. We then selected the plasmid with the mutated sequence with the restriction enzyme *Dpn I* which will only digest methylated parental plasmid DNA.

The nucleic acid substitutions were generated using the Quick-Change II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, California, USA). We conducted the reaction mixture according to the instructions but doubling the amount of DNA template (20 ng) due to the large size of the plasmid (16 Kb). The DNA sequences of the oligonucleotides (Eurogentec) are listed in table 5.

Table 5. Sequences of Site-Directed Mutagenesis Primers

Variant	Oligonucleotide sequences
G1982W	Forward: CGTGTGGTATCCACCTTTTGGGTTATTTTTGACCCTG Reverse: CAGGGTCAAAAATAACCCAAAAGGTGGATACCACACG
R2319C	Forward: CCTAAGGATAGAGTAATGATAAACTGCTTAGACAACATCTGTGAAGCAG Reverse: CTGCTTCACAGATGTTGTCTAAGCAGTTTATCATTACTCTATCCTTAGG
L1302P	Forward: CTGCTGCCAAAACCGAAGGCTGGTGGC Reverse: GCCACCAGCCTTCGGTTTTGGCAGCAG
T894A	Forward: GACTTTGCACGTAGCGCAGATGACCGGGGAG Reverse: CTCCCCGGTCATCTGCGCTACGTGCAAAGTC
A2160T	Forward: CTCCTCCAGTCATCTCATCTACTCATATTCAAGATGAGAGG Reverse: CCTCTCATCTTGAATATGAGTAGATGAGATGACTGGAGGAG
1801insD GHGT	Forward: ATGGGCACGGCACAGGCTATGAGAAGTACAACTCCATG Reverse: CCTGTGCCGTGCCCATCATGTTTGAACACTCCAATTAAG

Substituted or inserted nucleotides are indicated in red.

The reaction was performed in a ProFlex PCR system (Thermo Fisher Scientific). The reaction was initiated by preheating the mixture reaction at 95°C for 1 minute and then 18 cycles were carried out under the following conditions: denaturation at 95°C for 50 seconds, annealing at 60°C for 50

seconds and extension at 68°C for 20 minutes (1 minute and 15 seconds per kb of plasmid), followed by incubation at 68ºC for 7 minutes.

To generate a five-amino acid insertion, we applied a modified mutagenesis protocol (Liu *et al.*, 2008). The reaction was initiated by preheating the mixture reaction at 94°C for 7 minutes and then was carried out in two PCR steps consisting in 12 and 3 cycles respectively, under the following conditions: 12 cycles of denaturation at 94°C for 1 minute, annealing at 57°C for 1 minute and extension at 68°C for 34 minutes, 3 cycles of denaturation at 95°C for 1 minute, annealing at 49°C for 1 minute and extension at 68°C for 34 minutes, followed by incubation at 68ºC for 20 minutes. The amplification products are then treated with 1 µL of *Dpn I* restriction enzyme for 1 hour at 37°C.

In each case, the mutagenesis reaction product was precipitated using ethanol then dissolved in 5 µL of water. The total was used to transform One Shot TOP10 chemically competent *E. coli* bacteria (Thermo Fisher). The plasmids contained in the bacterial colonies were isolated using Miniprep (Qiagen) and then subjected to restriction analysis using the *Bgl I* enzyme. After electrophoresis on a 1% agarose gel, the clones with the same restriction profile as the original plasmid and appearing in sufficient concentration were then sequenced on both strands.

3.2.4 Confirmation of Mutated Variants by Sequencing Analysis

The sequencing reactions used for this analysis are as follows: 2 µL of BigDye® Terminator Mix (Applied Biosystems, Foster City, California, USA), 1 µL of 3.2 µM primer, 200 ng of purified plasmid DNA in a total volume of 5 µL. The selected primers allowed to verify whether the mutagenesis worked well, and the plasmids contained the desired variants. The DNA sequencing primers (Eurogentec) are listed in table 6.

Table 6. Forward and Reverse Sequencing Primers

Variant	Primer Sequence
G1982W	Forward: GCTATAAACGCCAACTCACTGA Reverse: ATTGCTGTTTCACAGGGTCAA
R2319C	Forward: CCCGAAGCAGGAGCTGTCT Reverse: TCTGCATTCTTTGGTACATAACTT
L1302P	Forward: GGGAAAAACTATCCAGTCCATTA Reverse: TATCAGAATCAGGTTTGGAGAAT
T894A	Forward: TAAGAAACCTGACTCAGAAGCAA Reverse: CCATGATACACAACCACGTTC
A2160T	Forward: CCCGAAGCAGGAGCTGTCT Reverse: TCATGAAGGAGCTGAGCTACT
1801insDGHGT	Forward: GCTATAAACGCCAACTCACTGA Reverse: ATTGCTGTTTCACAGGGTCAA

The reaction carried out in a thermocycler comprises 25 cycles under the following conditions: denaturation at 95°C for 10 seconds, annealing at 50°C for 5 seconds and elongation at 60°C for 4 minutes. After 25 cycles of elongation, the amplified products were purified using the X-Terminator purification kit (Applied Biosystems), then the sequences are analyzed with the sequencer ABI 3130 Genetic Analyzer (Applied Biosystems). The

forward and reverse sequences were analyzed using the Chroma Lite version 2.6.4 software and compared with the reference sequence of the *CHD7* gene (GenBank number NM_017780).

3.2.5 Cell Culture and Transfection

In this research, we used various human cell lines: HeLa (derived from a carcinoma cervix uterus), HEK293 (embryonic kidney cells), and SH-SY5Y (neuroblastoma) cells. These cells were cultured under an atmosphere of 5% CO_2 at 37ºC in Dulbecco's Modified Eagles Medium (DMEM) for HeLa and HEK cells, and DMEM/F-12 for SH-SY5Y cells (Invitrogen, Carlsbad, California, USA), supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin-streptomycin (10,000 U/mL, Invitrogen).

The culture medium was changed every two days to prevent impoverishment and to remove dead cells. When cells are reaching confluency, they must be subcultured: they are detached from the bottom of the dish by treatment with trypsin. The culture medium was removed, and the cells were washed with PBS (Phosphate Buffered-Saline: 140 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4 pH 7.3), thereby removing the serum elements which can inhibit the action of trypsin. The cells were then covered with 2 mL of 0.5 g/L trypsin 0.2 g/L EDTA (Invitrogen) and were incubated at 37°C for two minutes. The reaction was then stopped by adding 8 mL of culture medium. The suspension was

centrifuged for 5 minutes at 1,500 rpm. The cells were subcultured in culture medium and redistributed in culture plats at the appropriate dilution. The number of viable cells is determined by counting after staining with trypan blue.

For transitory transfection, HeLa, HEK293, and SH-SY5Y cells were seeded at a density of $4x10^5$, $1.6x10^5$, $8x10^5$ cells per well in 6-well culture plates, respectively. Then, the cells were incubated overnight at 37°C to obtain 70-80% confluence. Transfection was performed with 5 µL of Lipofectamine 2000 (Invitrogen) mixed with 4 µg of plasmid DNA. On the one hand, we tested a range of 1, 2, 4 and 8 µg of plasmid with 5 or 10 µL of Lipofectamine 2000 (Invitrogen) reagent per well; on the other hand, we tried 1 or 2 µg of the plasmid with 3 µL of FuGENE (Roche, Basel, Switzerland) per well. Two different exposure times of cells to the complexes were applied: 48 hours and 72 hours.

3.2.6 Cell Lysis and Protein Assay

Forty-eight hours after transfection, the cells were washed three times with PBS to remove the culture medium and dead cells. They were lysed using lysis buffer containing 10 mM Tris-HCl pH 7.5, 1% NP-4O and 0.5% DOC (Sodium Deoxycholate) and supplemented with 1 mM AEBSF (4-benzenesulfonyl fluoride hydrochloride) (Sigma-Aldrich, St. Louis, Missouri, USA) and a protease inhibitor cocktail (Roche). The lysate was clarified by centrifugation at 14,000 rpm for 5 minutes at 4°C. Total protein

concentration obtained after extraction was measured by the BCA method: 50 volumes of bicinchoninic acid (BCA, Sigma-Aldrich) were mixed with 1 volume of 4% copper sulfate solution (Sigma-Aldrich). A reference range was also performed to obtain a correspondence between absorbance and amount of protein contained in the sample: 5 tubes were made with range 0, 2, 5, 10 and 20 µg of bovine serum albumin (BSA) (total volumes 50 µL). An assay of the proteins in the range and in the samples was performed using spectrophotometer BIOMATE 3 at 562 nm. Then, the next step was denaturation of protein lysates, wherein a mixture of 1 volume of Laëmmli buffer (62.5 mM Tris-HCl pH 6.8; 2.3% SDS (v/v), 5% α-monothioglycerol (v/v), 15% glycerol (v/v), 0.001% bromophenol blue (v/v)) and 3 volumes of lysate were incubated at 95°C for 5 minutes.

3.2.7 Western blot Analysis

The samples were separated on SDS-polyacrylamide gel (SDS-PAGE) prepared according to the proportions described in annex 1. The concentration in acrylamide of the gel depends on the relative molecular mass (Mr) of the studied proteins. The migration of the proteins was carried out at a current of 34 mA in the SDS-PAGE buffer (SDS-10TGX, Amresco). The progress of the migration can be followed by a colored mass marker (Kaleidoscope, Bio-Rad, Hercules, California, USA). After that, the separated proteins were transferred from the gel onto a Hybond ECL nitrocellulose membrane (GE Healthcare, Chicago, Illinois, USA) in the presence of transfer buffer (ethanol 1 L; TG SDS- 10X 250 mL; H_2O 3.75

L). The transfer takes place at 200 mA for a period adapted to the relative molecular mass of the studied proteins (2 hours for CHD7 and 30 minutes for actin) at 4°C. After rinsing the membrane in 0.1% PBS-Tween (v/v), a step of blocking was performed. The membrane was covered with 10 mL of 0.1% PBS-Tween with 5% lyophilized skimmed milk at 4°C with stirring to prevent antibody binding to non-specific sites. Then three 5 minutes washes were carried out in 0.1% PBS-Tween. The primary antibody diluted in PBS Tween supplemented with 5% skimmed milk was applied to the membrane at 4°C with stirring overnight. After three times of washing in 0.1% PBS-Tween, the membrane was incubated with secondary antibody coupled to peroxidase, diluted in 0.1% PBS-Tween supplemented with 5% skimmed milk for one hour. Then three washes are made in the same buffer. Finally, the bindings of antibody to the membranes were recognized using chemiluminescence detection with ECL Plus Western blotting Detection System (GE Healthcare), and the proteins of interest were detected by exposure to photographic films Hyperfilm ECL (GE Healthcare).

3.2.8 Immunofluorescence

The cells were cultured on a glass coverslip. Forty-eight hours after transfection, they were washed three times in PBS to remove the culture medium and then fixed for 10 minutes at room temperature with 3.7% paraformaldehyde in PBS. The cells were rinsed two times with PBS, and permeabilized in PBS containing 0.5% Triton X-100 (Sigma-Aldrich) for 10

minutes to permit entry of antibodies into cells. Unspecific antigen sites were blocked with PBS containing 0.5% BSA (Sigma-Aldrich) for one hour at room temperature. The cells were incubated overnight with 30 µL primary antibody diluted 1/100 in PBS-BSA at room temperature in a humid chamber. The next day, the coverslips were washed three times in PBS-BSA and incubated 1 hour in the presence of 30 µL of secondary antibody (diluted 1/100 in PBS-BSA) supplemented with DRAQ5 (eBioscience, Waltham, Massachusetts, USA) to mark the nucleus. The coverslips were washed two times with PBS and once with distilled water. Coverslips were mounted with Mowiol (Sigma-Aldrich) mounting medium. Immunofluorescent localization of the proteins was visualized using a spectral confocal microscope Olympus FV1000 (Shinjuku, Tokyo, Japan).

3.2.9 Antibodies

The following primary antibodies were used in this study: rabbit monoclonal antibody anti-CHD7 (clone D3F5, Cell Signaling, Danvers, Massachusetts, USA); mouse monoclonal antibodies anti-HA, anti-FLAG, anti-β-actin (clone AC-74, Sigma-Aldrich) and anti-UBF, anti-nucleolin (C23, clone MS-3, Santa Cruz Biotechnology, Dallas, Texas, USA). Peroxidase-linked secondary antibodies for Western blot were obtained from Sigma-Aldrich: goat anti-mouse IgG (Fc-specific) and goat anti-rabbit IgG (whole molecule). For fluorescent imaging, we used RRX-conjugated goat anti-mouse IgG (H+L) and FITC-conjugated donkey anti-rabbit IgG

(H+L) (Jackson ImmunoResearch Laboratories, Cambridgeshire, United Kingdom).

3.2.10 RNA Extraction and cDNA Synthesis

RNA extraction was performed using the procedure of Chomczynski with the RNABle reagent (Eurobio, Courtaboeuf, France). Attached cells in 6-well culture plates were lysed by adding 1 mL of RNABle directly into the culture dish. The cells were homogenized by pipetting and transferred to a 1.5 mL Eppendorf tube. The cell homogenate was stored for 5 minutes at room temperature to permit the complete dissociation of nucleoprotein complexes. Next, 100 µL of chloroform was added to 1 mL of RNABle homogenate, shaken vigorously for 30 seconds and the resulting mixture was stored at room temperature for 5-10 minutes and centrifuged at 14,000 rpm for 15 minutes at 4°C. The aqueous phase was transferred into a fresh tube and RNA was precipitated by mixing with 500 µL of isopropanol. Samples were incubated at room temperature for 5-10 minutes and centrifuged at 14,000 rpm for 5 minutes at 4°C. After centrifugation, the supernatant was removed, and RNA pellet was washed once with 500 µL of 75% ethanol and subsequently centrifuged at 7,500 rpm for 5 minutes at 4°C. After removal of the supernatant, RNA pellet was air-dried for 5 minutes, and RNA was dissolved in H_2O by passing the solution a few times through a pipette tip. The RNA concentration was measured with a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

Reverse transcription was performed by mixing the following reagents: 2 µg of RNA, 5 µL of 5X buffer (Thermo Fisher Scientific), 2.5 µL of 100 mM DTT (Thermo Fisher Scientific), 1.5 µL of 1,600 ng/µL random hexanucleotides (Invitrogen), 1 µL of 25 mM dNTPs (Promega, Madison, Wisconsin, USA), 40 U of RNAguard (Euromedex, Strasbourg, France) and 400 U of M-MLV RT (Promega). cDNA synthesis was performed at 37°C for one hour, then denaturation of the enzyme was carried out by heating at 100°C for 2 minutes. The cDNA solution was stored at -20ºC. Finally, the cDNA was diluted 1/10 for subsequent use in quantitative PCR.

3.2.11 Quantitative RT- PCR

3.2.11.1 General Protocol

The PCR reactions were performed in 96-well plates containing a final volume of 15 µL: 7.5 µL of TaqMan Universal PCR Master Mix (Applied Biosystems), 0.675 µL of 20 µM forward and reverse primers, 0.2 µL of 10 µM specific TaqMan probe, 0.95 µL of distilled water and 5 µL of cDNA diluted in 1:10. The PCR amplification program is successively composed of activation of Taq polymerase for 10 minutes at 95°C followed by 40 PCR cycles: 15 seconds at 95ºC and 1 minute at 60°C. PCR reactions were performed in triplicate on a 7500 Fast Real-Time PCR System (Applied Biosystems). The sequence of oligonucleotides (Eurogentec) used for real-time PCR are listed in table 7.

Table 7. Sequences of qRT-PCR Oligonucleotides

Gene	Primers	Sequences
GAPDH	Forward	GCTCCTCCTGTTCGACAGTCA
	Reverse	ACCTTCCCCATGGTGTCTGA
	Internal probe	CCGCATCTTCTTTTGCGTCGCC
45S rDNA	Forward	GGAAGGAAGGAGGTGGGT
	Reverse	CGGTACGAGGAAACACCTG
	Internal probe	CCTCGAGCGTTCGCGTTCAG
SOX4	Forward	AGGCGAATTCCCGTTTGG
	Reverse	TTCCTAGCGCCGGTCACA
	Internal probe	TTTTCCTCCCTCTTTTCCCCTTGCCC
SOX10	Forward	TCTGAAGGCAGGAAGGAGTTG
	Reverse	TCTCAGACAAAGAATGAGGTTATTGG
	Internal probe	CACAGAGGCCCCCTGATCCAATTCTG
ID2	Forward	CAACACGGATATCAGCATCC
	Reverse	CGCTTATTCAGCCACACAGT
	Internal probe	TCCTTGCAGGCTTCTGAATTCCCT
MYRF	Forward	AAAGACACCGGAGACATGGT
	Reverse	GTTGTCTGTCAGCTTGCACA
	Internal probe	CAAGGAGCGCATCTTCATGG

Internal probes are labeled with 6-FAM in 5' and with TAMRA in 3'.

A comparative threshold cycle (C_T) was used to determine relative gene expression. The concept of the threshold cycle (C_T), corresponding to the middle of the exponential amplification phase, is the basis of accurate and reproducible quantification. Then, to compare the different values of transcript levels of RNA, a ΔC_T between the studied gene and the housekeeping gene is calculated:

$$\Delta C_T = C_{T\ target} - C_{T\ \textit{GAPDH}}$$

Then the ΔC_T of the samples and the reference cDNA are compared. The expression ratio (R) is determined using the formula below:

$$\Delta\Delta Ct = \Delta C_{T\ samples} - \Delta C_{T\ ref}$$

$$R = 2^{-\Delta\Delta CT}$$

Hence, the value used to plot relative gene transcription level was determined using the expression $2^{-\Delta\Delta CT}$ (Livak and Schmittgen, 2001). RNA amounts were expressed as an n-fold difference relative to the mock-transfected condition. Results are displayed as the means and standard deviation corresponding to three independent transfections. To compare sets of data, we used the Student's t-test. All statistical tests were performed using GraphPad Prism version 7.0 software for Windows (GraphPad Software Inc., La Jolla, California, USA).

3.2.11.2 Choice of Reporter Genes

In this thesis, we propose a simple assay, in which CHD7 protein variants are expressed in cell models. Then, the impact upon the transcription of several genes, whose expression is controlled by CHD7 protein according to the literature, is evaluated by qRT-PCR.

As we mentioned previously, the *CHD7* gene is involved in the fine-tuning of gene transcription in the early steps of the development of various tissues. It has been described that CHD7 protein plays a role as a positive regulator of the nucleolar expression of the 45S ribosomal RNA precursor in a colorectal cancer cell line. In mouse embryonic stem cells (ES), the

overexpression of wild-type CHD7 shows an increased expression levels of *45S pre-rRNA* compared with control. Moreover, depletion of CHD7 also reduced cell proliferation and protein synthesis. Last, compared with wild-type ES cells, the levels of *45S pre-rRNA* are reduced in both *Chd7* (+/-) and *Chd7* (-/-) mouse ES cells (Zentner *et al*., 2010).

The rRNA genes are transcribed by RNA polymerase I, and each gene produces a primary RNA transcript. In humans, the RNA transcript known as 45S rRNA, is about 13,000 nucleotides. Before it leaves the nucleus in assembled ribosomal particles, the 45S rRNA is cleaved to give the 28S rRNA (about 5,000 nucleotides), the 18S rRNA (about 2,000 nucleotides), and the 5.8S rRNA (about 160 nucleotides) of the final ribosome. The derivation of these three rRNAs from the same primary transcript ensures that they will be made in equal quantities. The remaining part of each primary transcript (about 6,000 nucleotides) is degraded in the nucleus (Figure 7). Some of these extra RNA sequences are thought to play a transient part in ribosome assembly, which begins immediately as specific proteins bind to the growing 45S rRNA transcripts in the nucleus (Albert *et al*., 1994). The oligonucleotides and the internal probe hybridize on the 45S rRNA cDNA upstream of that corresponding to the 28S rRNA. This fragment of the 45S rRNA is degraded during the processing of 28S, 18S and 5.8S rRNA: thus, the amplified product corresponding to the precursor rRNA but not to the mature rRNAs.

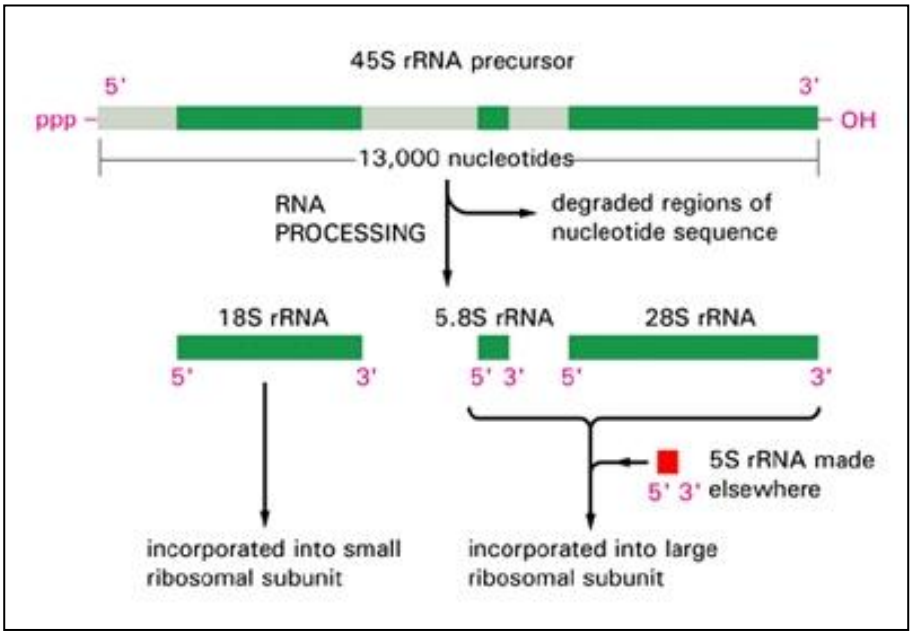

Figure 7. Processing of a 45S rRNA precursor molecule into three separates ribosomal RNAs. Nearly half of the nucleotide sequences in the primary RNA transcript is degraded in the nucleus (Alberts *et al*., 2002).

In another study in mice, genome-occupancy analyses coupled with transcriptome profiling, revealed that Chd7 interacted with Sox10 and targeted the enhancers of key myelinogenic genes such as *myrf* and *Id2/4*. So, *CHD7* coordinates with *Sox10* to regulate the initiation of myelinogenesis and acts as a molecular nexus of regulatory networks in the development**.** Quantitative real-time PCR (qRT-PCR) showed an increased expression of *Id2/4* and a decreased expression of *Myrf* or *Sox10* genes in *Chd7* knock-out mouse (He *et al*., 2016). Among them, we have chosen as reporter genes *MYRF* and *SOX10* itself, whose transcription was decreased in the spinal cord of Chd7 knock-out mice, and *ID2* which was upregulated.

Furthermore, it has been studied in mouse neural stem cells (NSCs) that CHD7 also stimulates the expression of *Sox4* and *Sox11* genes via the remodeling of their promoters to an open chromatin state (Feng *et al*., 2013). Therefore, we selected *SOX4* as a reporter gene. From these shreds of evidence, it would be possible to study the relationship between overexpression of *CHD7* gene variant alleles and the transcription variations of these reporter genes: *45S rDNA*, *SOX4*, *SOX10*, *ID2,* and *MYRF.*

3.2.12 Sensitivity and Specificity Test

The accuracy of this functional test system was determined by calculating sensitivity and specificity values. Sensitivity and specificity were calculated from a cross classification table (Table 8). For our study, sensitivity was defined as the proportion of *CHD7* pathogenic variants (CS-causing) who were tested positive by the functional assay and were proven as CS-causing as well. Specificity was defined as the proportion of *CHD7* non-pathogenic variants which were tested negative by the functional assay and were proven as functional as wild-type CHD7.

Table 8. Sensitivity and Specificity Test

	Pathogenic variant	**Non-pathogenic variant**
Functional assay positive	A True Positive	B False Positive
Functional assay negative	C False Negative	D True Negative
Total	A + C	B + D

$$\text{Sensitivity} = \frac{A}{A + C} \times 100\% \qquad \text{Specificity} = \frac{D}{B + D} \times 100\%$$

3.3 Genome Modification Technologies

Genome modification is a milestone of modern research in genetics and molecular biology. These technologies were used widely to create cell models (Bauer *et al*., 2014; Cong *et al*., 2013; Jiang *et al*., 2013; Maeder *et al*., 2013), plants (Miao *et al*., 2013) and animals (Jao *et al*., 2013; Li *et al*., 2013) genetically modified by inserting, deleting, modifying or replacing the DNA in their genomes for research purposes. Recently, by making model organisms regarding biological pathways in disease and health, it appeared potentially suitable for gene therapy in humans (Xue *et al*., 2014; Yin *et al*., 2014; Schwank *et al*., 2013; Xie *et al*., 2014).

Various genome editing technologies have evolved rapidly in recent years, including zinc-finger nucleases (ZFNs) (Porteus and Baltimore, 2003; Miller *et al*., 2007; Sander *et al*., 2011; Wood *et al*., 2011), transcription activator-like effector nucleases (TALENs) (Wood *et al*., 2011; Christian *et al*., 2010; Zhang *et al*., 2011; Hockemeyer *et al*., 2011; Reyon *et al*., 2012; Boch *et al*., 2009; Moscou and Bogdanove, 2009; Sanjana *et al*., 2012) and the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) technology. The ZFN and TALEN techniques use an approach of tethering endonuclease catalytic domain to modular DNA-binding proteins for inducing targeted DNA double-stranded breaks (DSBs) at specific genomic loci. By contrast, CRISPR/Cas9 involves a nuclease guided by small RNAs pairing with target DNA (Garneau *et al*., 2010; Jinek *et al*., 2012; Gasiunas *et al*., 2012; Ran *et al*., 2013; Doudna and

Charpentier, 2014). These technologies enabled the direct study of protein function by modulating or editing the expression of its endogenous genes.

In this study, we aimed to develop a biological functional test of CHD7 protein with an endogenous expression level, both in wild-type form and missense variants. For this purpose, we implement the CRISPR/Cas9 technique for genome modification due to its design simplicity, specificity, and efficiency compared to ZFN and TALENT (Ran *et al*., 2013).

3.3.1 CRISPR/Cas9 System

Recent advances in molecular techniques and bioinformatic analysis have allowed for the isolation and the sequencing of a large number of prokaryotic genomes which has led to the discovery of many unknown genes. First discovered in 1987 in the K12 strain of *Escherichia coli*, the CRISPR term was previously identified as an independent family of DNA repetitive regions. These regions are interspaced by similar-sized non-repetitive (spacer) DNA, associated with Cas genes (Figure 8) (Ishino *et al*. 1987, Jansen *et al*., 2002; Mojica *et al*., 2005). The CRISPR and the CRISPR associated (Cas) genes are found in approximately 40% of bacterial genomes and in nearly 90% of sequenced archaeal genomes (Zhang *et al*., 2014). This finding elicits the breakthrough of this technology and its rapid evolution within the scientific community.

The CRISPR/Cas locus has a unique architecture. Firstly, the AT-rich leader sequence contains a promoter and different sites for several

regulatory elements (Jansen *et al*., 2002). Secondly, a specific set of Cas genes are located upstream to each CRISPR locus. Thirdly, repeated sequences are interspaced by specific DNA sequences named spacers. The palindromic repeat sequences are 21-48bp in length identical through each CRISPR/Cas locus and separated by 26- to 72-bp spacers (Wiedenheft *et al*., 2012). The DNA spacer unique sequences have been shown to be virus-derived and to protect the host against an invading organism containing the foreign nucleic acid called protospacers (Cady *et al*., 2012; Zhang *et al*., 2014).

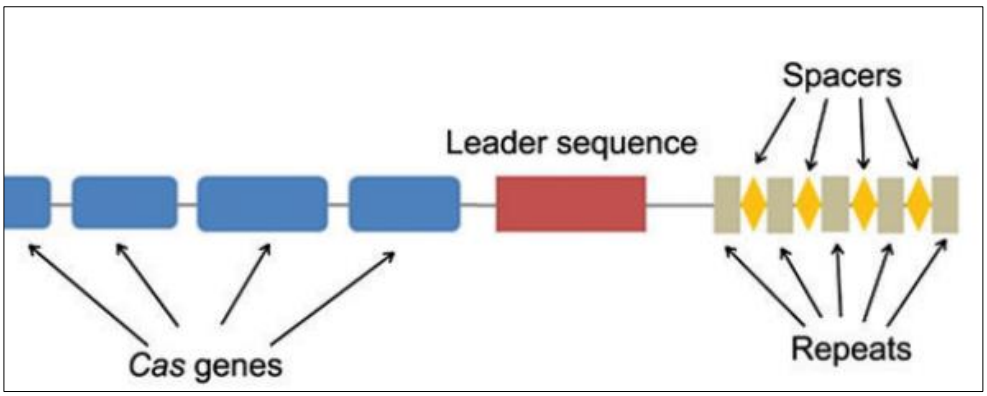

Figure 8. Architecture of the CRISPR/Cas locus. The locus consists of a leader sequence, identical repeat sequences, unique spacer regions and Cas genes (Zhang *et al*., 2014).

3.3.2 Stages and Types of the CRISPR/Cas

Due to the rapid evolution of this system, it has been a challenge to group CRISPR/Cas systems, however three major types have been established: type I (6 subtypes A-F), II (2 subtypes, A-B), and III (2 subtypes, A-B) (Makarova *et al*., 2011). The type II system has fewest components,

making it the simplest system for use in genetic engineering (Gasiunas *et al*., 2012). Type II from *S. pyogenes* requires only the Cas9 protein compared to type I (cascade proteins) and III (Csm), has been adapted for targeted genome editing (Sander and Joung, 2014). This is the CRISPR system described in this study.

CRISPR/Cas immunity as described in *S. pyogenes* involves three following stages: acquisition, expression, and interference (Figure 9) (Bikard *et al*., 2012).

Stage 1 - CRISPR adaptation: *S. pyogenes* integrates a new spacer sequence from the foreign DNA into its chromosome at one end of the CRISPR locus (Barrangou *et al*., 2007). The integrated sequence needs to contain a short sequence known as Protospacer Adjacent Motif (PAM) to be recognized by the host cell (Zhang *et al*., 2014). SpCas9 recombinant Cas9 protein is the most widely used which requires a 5´-NGG-3´ PAM sequence and resulting in a blunt-ended double-strand break (DSB) (Anders et al., 2014). The DSB has typically created three base pairs 5' of the PAM sequence. This guide sequence is coupled to the scaffold sequence (tracrRNA) to form the sgRNA (Ran *et al*., 2013). The researchers are still investigating the mechanism of spacer integration (Levy *et al*., 2015).

Stage 2 – CRISPR expression: The CRISPR/Cas9 system includes two RNA called CRISPR RNA (crRNA) and the trans-activated crRNA (tracrRNA). During this process, the CRISPR locus is transcribed by an

RNA polymerase into a long primary transcript, the premature CRISPR RNA (pre-crRNA) made up of multiple spacer-repeat units in tandem. It is then cleaved and processed into small mature crRNAs. The tracrRNA forms a duplex structure with the crRNA which anneals with the target DNA and guide the Cas9 endonuclease to cleave the invader's DNA (Makarova *et al*., 2011; Bikard *et al*., 2012).

Stage 3 – CRISPR interference: In the last stage, the foreign DNA is destroyed using the processed crRNA complexed with Cas proteins (Bikard *et al*., 2012). The CRISPR interference stage involves the cleavage and the degradation of the target nucleic acid. The mature crRNA is loaded into the Cas9 endonuclease which is then activated. The complex crRNA/Cas9 targets complementary DNA and Cas9 perform the cleavage of the complementary and non-complementary strands of the DNA target using its two nuclease motifs: RuvC at the N-terminal and an HNH-domain at the C-terminal (Gasiunas *et al*., 2012, Jinek *et al*., 2012). The binding of the complex to the protospacer sequence is mediated by the recognition of an obligatory PAM sequence by Cas9 located immediately downstream of the protospacer. In this way, the bacteria acquire an efficient system to defend itself against future invaders (Karvelis *et al*., 2013).

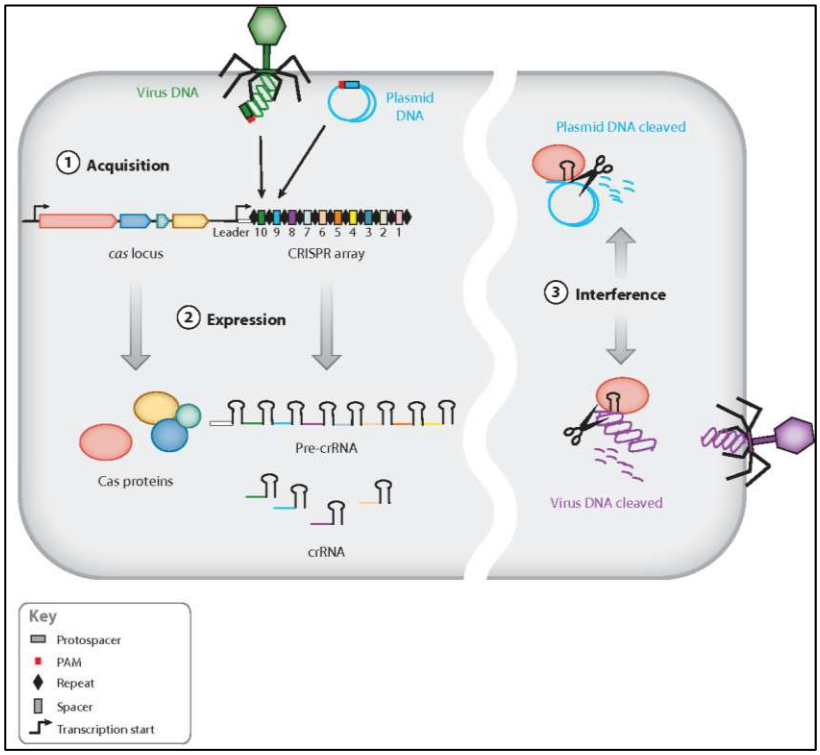

Figure 9. CRISPR mechanism of action overview. Here are graphically represented the three stages of CRISPR mechanism of action; (i) acquisition, (ii) crRNA processing and (iii) interference in the three major types (Bhaya *et al.*, 2011).

3.3.3 Application of CRISPR/Cas9 System for Genome Modification in Functional Study

The CRISPR/Cas9 system is a powerful tool to create mutations or to make deletion and/or insertions at specific sites in the genome. Therefore, this technique had known rapid development these past few years in the scientific community. The simplicity of the *Streptococcus pyogenes* CRISPR/Cas9 type II system (Figure 10), has allowed scientists to engineer it for research purposes. In this aim, they constructed two component genome editing tool consists of 1) a guide RNA (gRNA), a single sequence combination of crRNA and tracrRNA (scaffold); and 2) NLS sequences and

Cas9 endonuclease which cleaves both stands of the DNA target using its two nuclease motifs, RuvC and HNH (Jinek *et al*., 2012; Li *et al*., 2013).

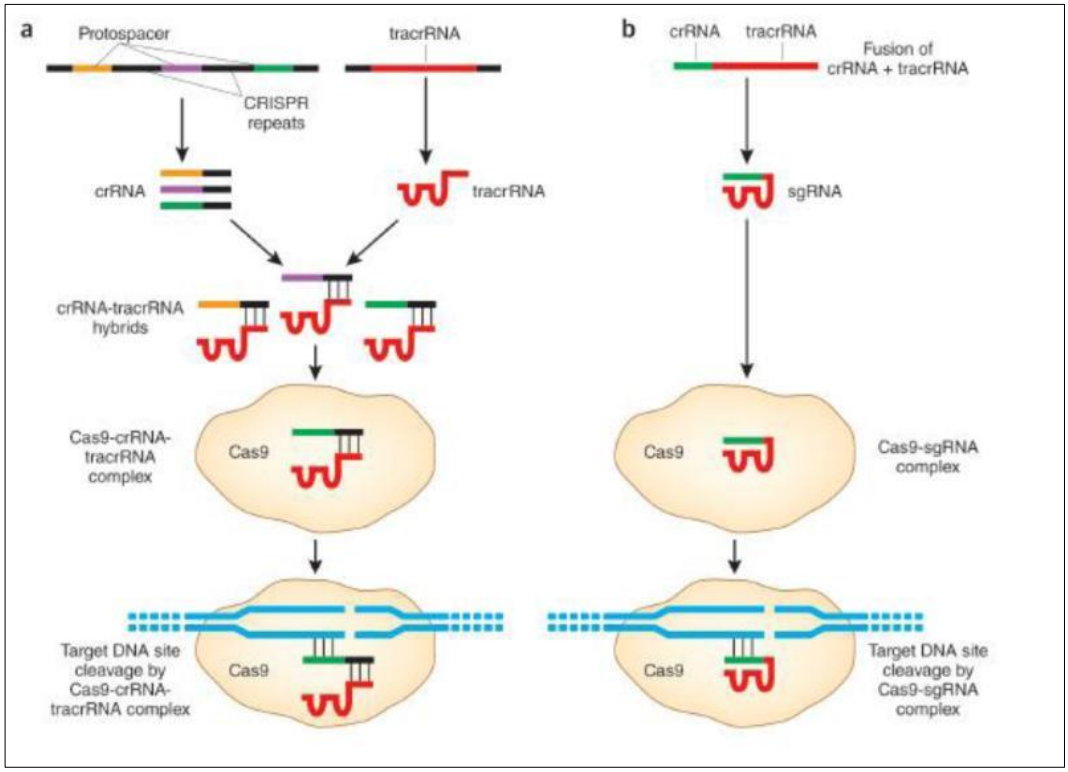

Figure 10. Natural vs. engineered CRISPR systems. A) Natural CRISPR pathway starting from the transcription of each pre-crRNA and tracrRNA. Next, tracrRNA binds to pre-crRNA followed by cleavage of guide RNA from pre-crRNA. Finally, to produce the active Cas9 nuclease, inactive Cas9 nuclease binds to the guide RNA. B) Engineered CRISPR start from the transcription of guide RNA as a single sequence. After that, transcription and translation of Cas9 nuclease occur followed by binding of guide RNA to Cas9 that permits activation of Cas9 (Sander and Joung, 2014).

The CRISPR/Cas system has the following features 1) target design simplicity, 2) assembly speed, 3) efficiency and 4) multiplexed mutations (Ran *et al*., 2013). Online web-based software provides CRISPR target selection and off-target prediction. The scientific continues to investigate

the biochemistry and the mechanism of Cas9 and guide RNA (Hsu *et al*., 2013).

In this thesis, the main workflow involves the perturbation of *CHD7* gene function (knock-out or knock-in the desired variants) to investigate the consequence on the transcription of our reporter genes. For this application, guide sequences can be designed using various on-line design tools (Cui *et al*., 2018). Designing a guide sequence require several information including the type of Cas protein used, the target sequence and the organism that is targeted. To achieve the best guide sequence, the GC content of a guide must be balanced: low or high numbers of G or C especially proximal to the PAM will decrease the activity of the guide (Ran *et al*., 2013). Finally, in order to edit the genome, a vector carrying the Cas9 gene and engineered sequences which will be transcribed to produce specifically targeted RNA guides should be delivered into the cells (Thurtle-Schmidt and Lo, 2018). However, expression from a DNA vector may cause increased off-target effects due to overexpression of the Cas9 components (Jacobi *et al*., 2017).

3.3.4 DNA Reparation Systems

When the Cas9/sgRNA are delivered and have successfully identified the target, Cas9 will break the DNA strand. There are two possible repair pathways (Figure 9). The first choice is the Non-Homologous End Joining (NHEJ). The NHEJ is an error-prone repair in which the cells simply ligate the two strands together via the end-joining repair system (Wilson and

Lieber, 1999). However, this mechanism often fails to join the DNA without modification of the sequence. This results in random insertions or deletions (indels) of nucleotides in the targeted sequence (Rodgers and Mcvey, 2016). For this reason, in genome engineering, the mistakes produced by NHEJ are used for generating targeted gene knock-out. When random indels, with a length that is not multiple of three, are produced in genes especially at the early exons, it generally results in frame-shift which leads to the creation of premature stop codon, hence loss of function.

The second possible option is to rely on Homology-Directed Repair (HDR). A homologous DNA sequence will be used as a template to mediate a custom repair of the DNA strand. By this process, introducing desired sequence modification can be performed by delivering an engineered double-stranded donor template with homologous flanking regions to the DSB region. Alternatively, it is possible to use an oligonucleotides (ssODNs) harbors desired modifications as a template for HDR (Jasin and Rothstein, 2013). The frequency of these modifications is various, generally higher than 1% and can even reach 50% (Sander and Joung, 2014). A recent study described that approximately about 15% gene editing using CRISPR/Cas9 fails due to persistent binding of the Cas9 protein to the DNA at the cut site, which blocks the DNA repair enzymes from accessing the cut (Figure 11) (Clarke *et al*., 2018).

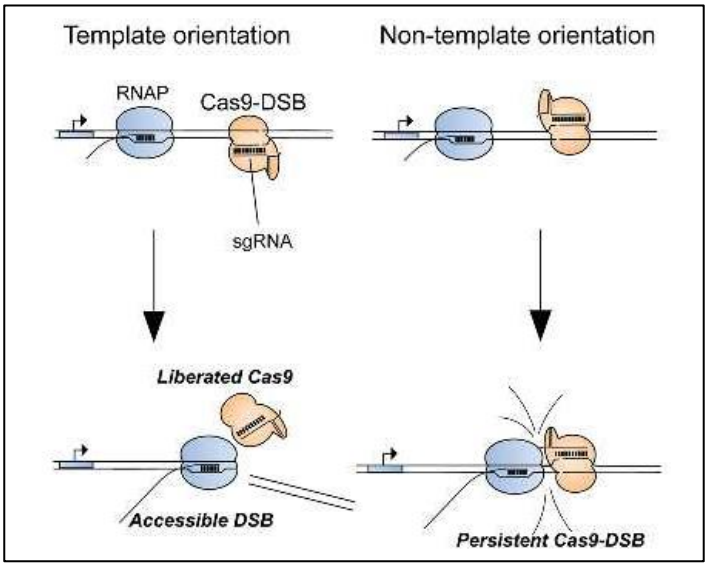

Figure 11. Persistent Cas9 binding to double-strand DNA breaks (DSBs) blocks DNA break repair (Clarke *et al.*, 2018).

Several studies of the HDR process with different kinds of donor templates have shown that asymmetrical single-stranded DNA (ssDNA) donors of 127 nucleotides are highly efficient (Richardson *et al.*, 2016). These donors need to be complementary to the strand that is released from the Cas9 complex first following DNA cleavage (Figure 12). However, it should be noticed that the use of a DNA template does not restrict the cell to use HDR alone. As a result, a mixture of reparations by the two different mechanisms sometimes arises in the cell.

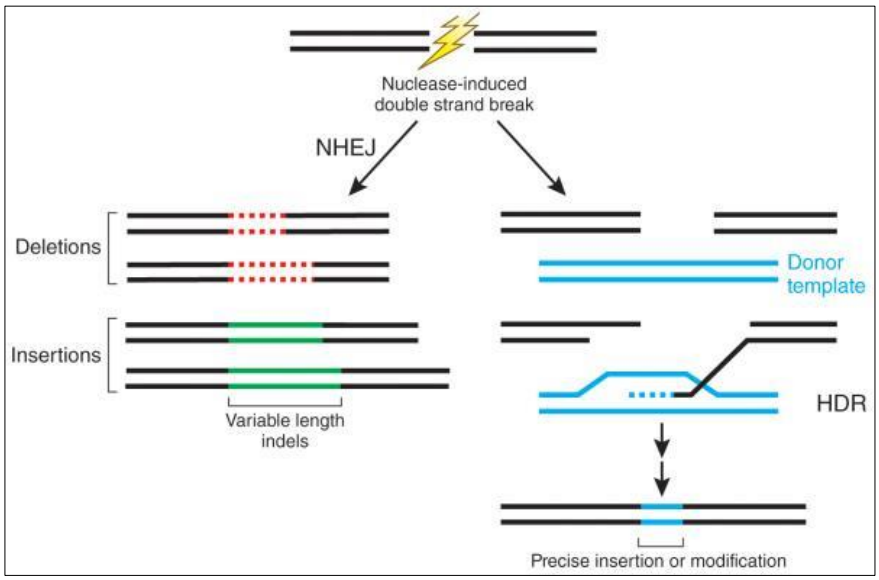

Figure 12. Fixing the DSB: NHEJ vs. HDR. Non-Homologous End Joining (NHEJ) is the primary DNA repair mechanism which may cause gene disruption. Homology-Directed Repair (HDR) will mediate a strand-exchange process to repair DNA damage accurately, based on the availability of the homologous DNA template (Sander and Joung, 2014).

Validation of mutations and analysis of off-target effects are challenging because Cas9 targets single cells and a population of cells will thus contain a collection of different edits in the target gene. Even within the cell, it is likely that two alleles are mutated differently. To verify the desired modifications, PCR amplification of the target sequence and subsequent cloning of this product are required.

A genetically mosaic population of cells can also be analyzed with T7 Endonuclease I (T7E1). The method consists of 4 steps: genomic DNA isolation, PCR on the desired locus, denaturation and reannealing, and T7 endonuclease I cleavage. This assay detects heteroduplex DNA that results from annealing of DNA strands that have been modified after a

sgRNA/Cas9-mediated cut to DNA strands without modifications. Besides, there are multiple other assays available that do essentially the same thing (e.g., Surveyor assay) (Guschin *et al.*, 2010).

3.4 Functional Assay of Endogenously Expressed CHD7 Missense Variants using CRISPR/Cas9 System

3.4.1 Choice of *CHD7* Gene Sequence to Target with the CRISPR/Cas9

The exons chosen to be targets were: exon 3 (for generating knock-out), exon 16, 30, and 33 (for generating missense variants) of the *CHD7* gene. Blast analysis of the chosen targets was performed against the human genome to ensure specificity. The sequence of each exon was input into the online software CRISPR Design (http://tools.genome-engineering.org), to search for protospacer target sequences with the form 5′(N)20-NGG-3′, NGG being the PAM sequence necessary for the Cas9 to cut the protospacer DNA. The output included several 20 bp target options, with different specificity values, based on a statistical logarithm of off-target hits. The option with the best specificity was chosen in order to minimize the chances of off-target binding of the sgRNA and therefore to cut by the nuclease Cas9.

3.4.2 Cas9 Nuclease and Single Guide RNA (sgRNA) Constructs

The plasmid pSpCas9(BB)-2A-Puro was purchased from Addgene (#62988). Spacer selection for targeting by Cas9 and subsequent generation of PCR amplicon was performed as described in Ran *et al.*,

2013. The vector harbors the sequences encoding NLS-Cas 9, puromycin resistant, and the crRNA: tracrRNA backbone, preceded by a multiple cloning site, to make it possible to insert the specific spacer for each target region, downstream of a T7 promoter. The specific protospacer sequence was generated by annealing short oligonucleotides which were designed to be complementary to the *CHD7* gene target in the human genome. The design of the primers consisted in leaving overhangs to enable cloning into the vector (Figure 13). The oligonucleotides are indicated in table 9.

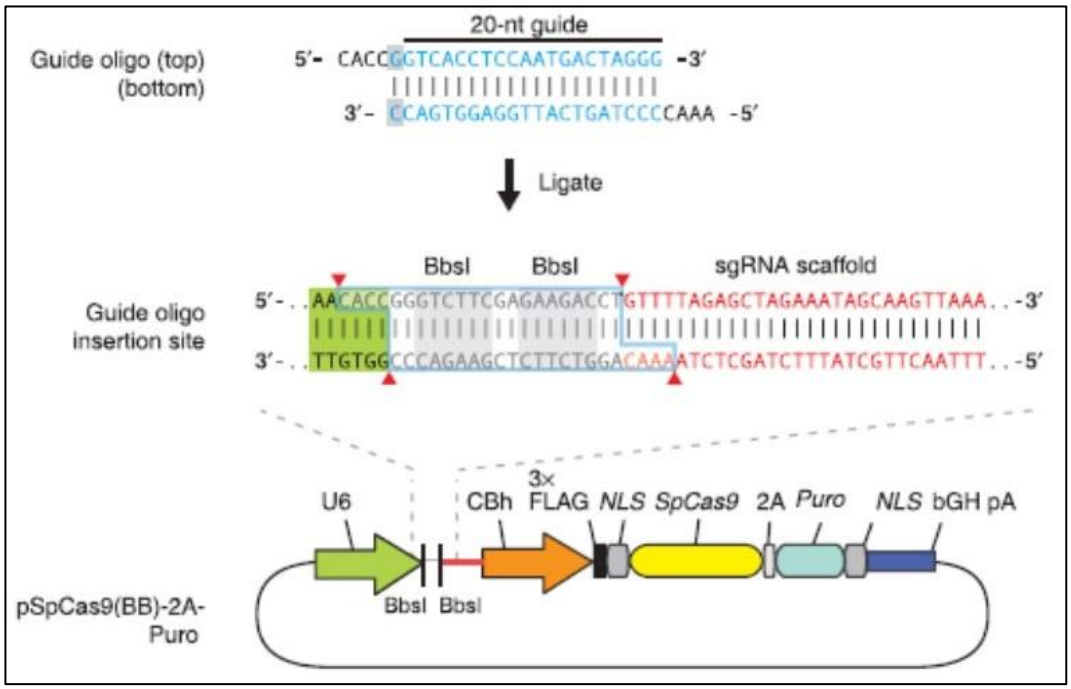

Figure 13. Schematic for cloning of the guide sequence oligos into pSpCas9(BB)-2A-Puro plasmid containing the sequence allowing the expression of Cas9 and the sgRNA scaffold. The guide oligos for the top and bottom strands example (blue) contain overhangs for ligation into the pair of *BbsI* sites in the plasmid. Plasmid digestion with *BbsI* permits the replacement of the type II restriction sites (blue outline) with direct insertion of annealed oligos (Ran *et al*., 2013).

Table 9. Target sequences of the three sgRNA and complementary oligo sgRNAs

Exon	Target Site	PAM	Oligonucleotides
Exon 3	ACCTGAGTCATATCCGGCAC	TGG	Top: CACCGACCTGAGTCATATCCGGCAC Bottom: AAACGTGCCGGATATGACTCAGGTC
Exon 16	TGGCCACCAGCCTTCAGTTT	TGG	Top: CACCGTGGCCACCAGCCTTCAGTTT Bottom: AAACAAACTGAAGGCTGGTGGCCAC
Exon 30	ACCGTGTGGTATCCACCTTT	GGG	Top: CACCGACCGTGTGGTATCCACCTTT Bottom: AAACAAAGGTGGATACCACACGGTC
Exon 33	CTTCACAGATGTTGTCTAAG	CGG	Top: CACCGCTTCACAGATGTTGTCTAAG Bottom: AAACCTTAGACAACATCTGTGAAGC

The annealing of the primers occurred in a total volume of 20 µL with 5µL of 200 µM top and bottom oligos, 2 µL of 10x oligo annealing buffer (Invitrogen) under the following conditions: incubation at 95°C for 4 minutes and store at room temperature for 5-10 minutes. Finally, the annealed products were diluted 1/200 for subsequent use in the ligation process. The cloning of the oligonucleotides in the vector consisted of a one-step digestion and ligation reaction as follow: pSpCas9(BB)-2A-Puro plasmid was digested with *BbsI* (New England Biosciences). 50 ng of linearized plasmid was used for subsequent ligation. The extremities of the plasmid were ligated with 250 nM annealed oligonucleotides by T4 DNA ligase (Thermo Fisher Scientific) by incubation overnight at 15°-4°C. The ligated plasmid was transformed into TOP10 chemically competent *E. coli* bacteria

(Thermo Fisher Scientific) by using 2 µL of ligation product and plating 50 µL in an LB-agarose plate with ampicillin for selection. For identification of the correct clones, 15 colonies of each construct were selected and cultured to extract the plasmid DNA by using the QIAGEN MiniPrep kit according to the manufacturer's instructions. Finally, the plasmids from each construct were sequenced with the sequence primers to confirm the presence of the insert (Table 5). The correct clones were then extracted using QIAGEN MidiPrep, according to the manufacturer's instructions.

3.4.3 Repair-Template Design: Single-stranded DNA Oligo-nucleotides (ssODNs) and Double-stranded Targeting Plasmid

For a targeted DNA base change purpose, HDR must be the repair pathway. For each of the three different sgRNA targeted sites (located in the exon 16, exon 30, and exon 33 of the *CHD7* gene), we designed three single-stranded DNA oligonucleotides (ssODNs) repair templates. These templates contained symmetrical homology arms, flanking the theoretical Cas9 cut site (located 3 base pairs upstream of the PAM), with total lengths of 108, 110, and 110 nucleotides, respectively, and sharing sequence identity with the sgRNA-binding strands. The substituted nucleotides are listed in table 10.

Table 10. List of single-stranded oligo deoxynucleotides (ssODNs) used as HDR templates

Variant	ssODN
G1982W	5'ATAATCTTTCTAACAGGTGGACAAGAAGAGAAGAGGCTGATTTTTACCGTGTGGTATCCACCTTTTGGGTTATTTTTGACCCTGTGAAACAGCAATTTGACTGGAACCAA-3'
R2319C	5'CTTGTGACTTTTCTTCTCCCTCCAGGATAGAGTAATGATAAACTGCTTAGACAACATCTGTGAAGCAGTGTTGAAAGGCAAATGGCCAGTAAATAGGCGCCAGATGTTTG-3'
L1302P	5'TCCAGGCTGCTGGCAAGCTAGTGCTGATTGACAAGCTGCTGCCAAAACCGAAGGCTGGTGGCCACAGGGTGCTTATCTTTTCCCAGATGGTGCGCTGCTTGGACATAC-3'

Substituted nucleotides are indicated in red.

As an alternative, we constructed HDR template plasmids, a conventional double-stranded DNA with homology arms flanking the mutated sequence. For L1302P variant, the exon 16 of *CHD7* gene from patient's DNA harboring this variant was amplified and ligated with pGEM-T Easy vector using T4 DNA ligase by incubation overnight at 4°C. The ligated plasmid was transformed into TOP10 chemically competent *E. coli* bacteria kit (Thermo Fisher Scientific) by using 2 µL of ligation product and plating 50 µL in an LB-agarose plate with Isopropyl β-D-1-thiogalactopyranoside (IPTG) along with X-gal for blue-white screening and 100 µg/mL ampicillin. For identification of the correct clones, 5 white colonies were selected for Miniprep (Qiagen). The plasmids were sequenced using T7 universal primer 5′-d(TAATACGACTCACTATAGGG)-3′ and SP6 universal primer 5′-d(TATTTAGGTGACACTATAG)-3′ to confirm the presence of the desired mutation. The correct plasmid was then linearized using *Pst1* enzyme.

For G1982W and R2319C variants, the constructions were made with the same procedure as described above, but the DNA templates for amplification of exon 30 and exon 33 were obtained from healthy human DNA. The desired mutations were introduced into the ligated plasmids using site-directed mutagenesis. The DNA sequences of the oligonucleotides (Eurogentec) are listed in table 5. Finally, plasmids with desired mutations were linearized using *Pst1* enzyme.

3.4.4 Transfection and Clonal Cell Isolation

In this experiment, we used HeLa and SH-SY5Y cells. The cells were cultured under steady-state culture conditions as described above. Twenty-four hours before transfection, the cells seeded at a density of $8x10^5$ cells per well into 6-well culture plates. On the day of transfection, 70-90% confluency is optimal. Transfection was performed with 5 µL of Lipofectamine 2000 (Invitrogen) mixed with 4 µg of pSpCas9(sgRNA) to generate CHD7 knock-out, and co-transfected with 80 µM of ssODN or 4 µg of linearized HDR template plasmids to obtain desired missense variants. Forty-eight hours after transfection, Puromycin selection (Invivogen, Toulouse, France) was applied at a concentration of 10 µg/mL and 4 µg/mL for HeLa and SH-SY5Y cells, respectively. Furthermore, the cells were incubated for 72 hours with puromycin and then grown in a selection-free medium.

Isolation of clonal cell lines with specific modifications is often desired. Isolating single cells were achieved through serial dilution to a final concentration of 0.5 cells per well in 96-well plate to reduce the likelihood of having multiple cells per well, followed by an expansion period to establish a new clonal cell line.

3.4.5 T7 Endonuclease I Assay

The T7E1 assay was performed as previously described (Guschin *et al.*, 2010). We used this assay to get a first estimate of whether our targeting was successful or not. Genomic DNA was extracted by using QIAamp DNEasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. The DNA was assayed by measuring the absorbance at 260 nm using spectrophotometer Nanodrop ND 1000 instrument (Thermo Fisher Scientific). PCR amplification of exon 3 was performed using the intronic primers *CHD7*-F 5'-GCCAGCCCA TATAGCAGTAC-3' and *CHD7*-R 5'-AACACAGCCCAGCATCGTGA-3'. Approximately 100 ng of DNA solution (1 µL) was added to 49 µL of PCR mixture. This PCR mixture contained 0.5 µL of 25 mM deoxyribonucleotide triphosphates (dNTPs), 4 µL of 25 mM $MgCl_2$, 0.5 µL of each 20 µM primer, 5 µL of 10X PCR buffer, 0.25 µL of 5 U/µL Diamond® high fidelity Taq DNA polymerase (Eurogentec), and 38.25 µL of H_2O. PCR was initiated with denaturation at 94°C for 5 minutes, followed by 35 PCR cycles (at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 45 seconds) and 10 minutes final elongation at 72°C. The amplified products were detected by

electrophoresis on a 1.5% agarose gel with 0.5 mg/mL ethidium bromide and visualized under ultraviolet (UV) light. Next, PCR products were purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions. The PCR products were eluted in 50 µL of water, then quantified using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific).

T7 Endonuclease I assay was performed using 200 ng of purified PCR products and 2 µL 10X NE Buffer in 19 µL reaction mixture. The hybridization of PCR products conducted in conditions as follows: initial denaturation at 95°C for 5 minutes, denaturation from 95° to 85°C with -2°C/second ramp rate, followed by incubation from 85° to 25°C with -0.1°C/second ramp rate. Afterward, 1 µL of T7 *Endonuclease I* (New England Biolabs) was added to the annealed PCR products and incubated at 37°C for 15 minutes. The reaction was stopped by adding 1.5 µL of 0.25 M EDTA. Finally, the reaction products were analyzed by electrophoresis on a 1.5% agarose gel to resolve full-length DNA and cleavage products (Figure 14). Bands were quantified by densitometry using Scion Image software. The percentage of nuclease-specific cleavage products (fraction cleaved) was determined.

The estimation of gene modification was calculated using the following formula:

$$\% \text{ gene modification} = 100 \times [1 - (1 - \text{fraction cleaved})^{1/2}]$$

$$\text{Fraction cleaved} = \frac{[\text{sum of cleaved band intensities}]}{[\text{sum of the cleaved band + parental band intensities}]}$$

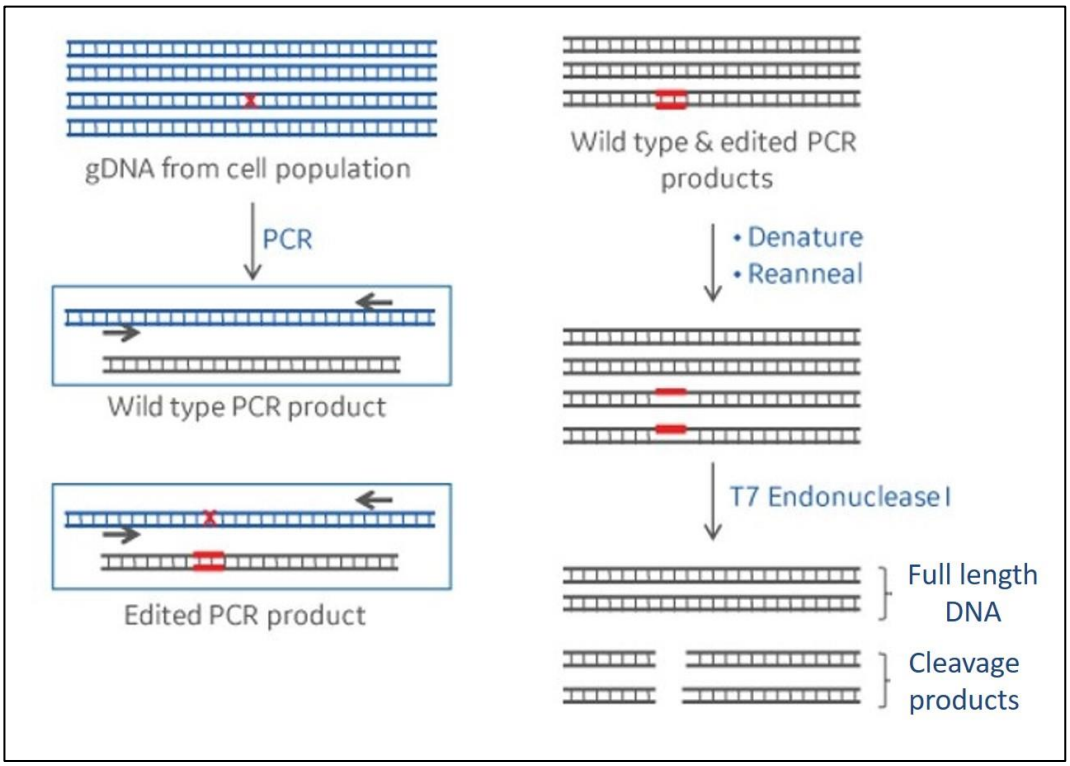

Figure 14. Schematic of the T7EI assay. Genomic DNA (blue) from treated cells containing WT and modified DNA (mutation in red). PCR amplification generates WT and modified PCR products (black). Denaturing and reannealing of these PCR products generates mismatch are cleaved by T7E1 nuclease (Adapted from https://dharmacon.horizondiscovery.com).

3.4.6 Verification of the Mutations

The genomic DNA of 12-24 isolated clones was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instruction. DNA concentrations were measured by NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). The desired missense variants were detected by PCR followed by Sanger sequencing using sequencing primers (Table 5). To confirm the *CHD7* knock-out, exon 3 of the *CHD7* gene was amplified and ligated in pGEM-T Easy vector as described previously. The PCR products of targeted exon were cloned into the vector in order to verify the modification occurred in both alleles. In general, each allele carrying a different mutation. For identification, 16 white colonies of each clone were selected for Miniprep. The plasmids were sequenced with the T7 and SP6 universal primers to confirm the indel mutations.

CHAPTER IV

IDENTIFICATION OF A NOVEL *CHD7* MUTATION IN AN INDONESIAN CHARGE SYNDROME PATIENT

4.1 Case Presentation

In Indonesia, only one case of clinically diagnosed CS has been reported to date, without further investigation (Pramudita *et al*., 2017). We provide a retrospective report of the molecular diagnosis of this patient based on next-generation sequencing (NGS). To our knowledge, this is the first report on the molecular diagnosis of CS in an Indonesian patient.

A five-year-old boy with multiple deformities was born from non-consanguineous Indonesian healthy parents. The pregnancy was without complications, and delivery was spontaneous. The mother was 37-year-old, and the father was 36-year-old. His birth weight was 2600 gram. He had two healthy brothers, and there was no family history of similar complaints, nor any other congenital abnormality in the three-generation family pedigree (Figure 15).

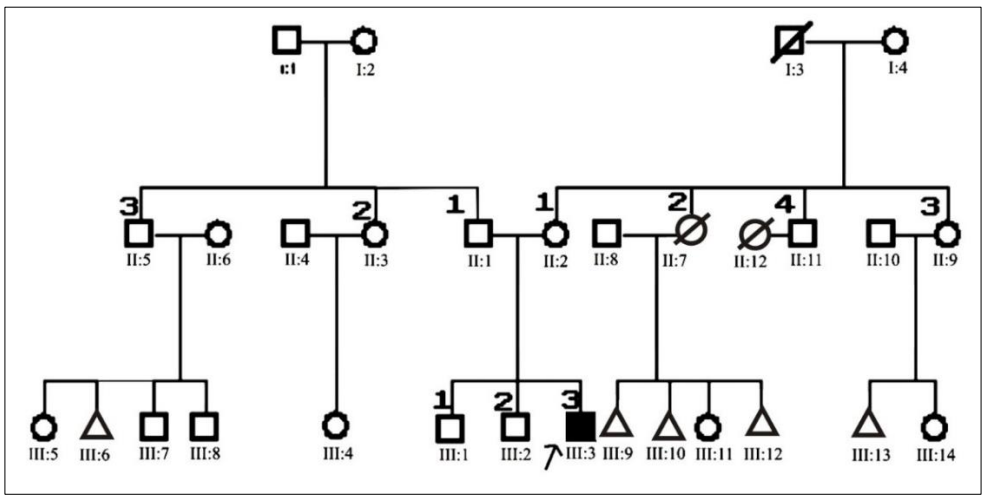

Figure 15. Pedigree of an Indonesian CS patient. The patient (indicated with black arrow) is the only affected individual in a three-generation pedigree. He has two healthy siblings from healthy parents; thus, the pattern of inheritance is *de novo*.

Since the first week of life, the infant had feeding problems, short stature, and failure to thrive. A feeding tube was placed through the patient's nose into the stomach in order that he would receive adequate nutrition. Subsequent examinations revealed multiple malformations such as iris coloboma on the left eye, microphthalmia and bulbous atrophy on the right eye, hypoplasia and abnormal pinnae on both ear cartilages, small nostril on the right side, and high arched palate (Figure 16).

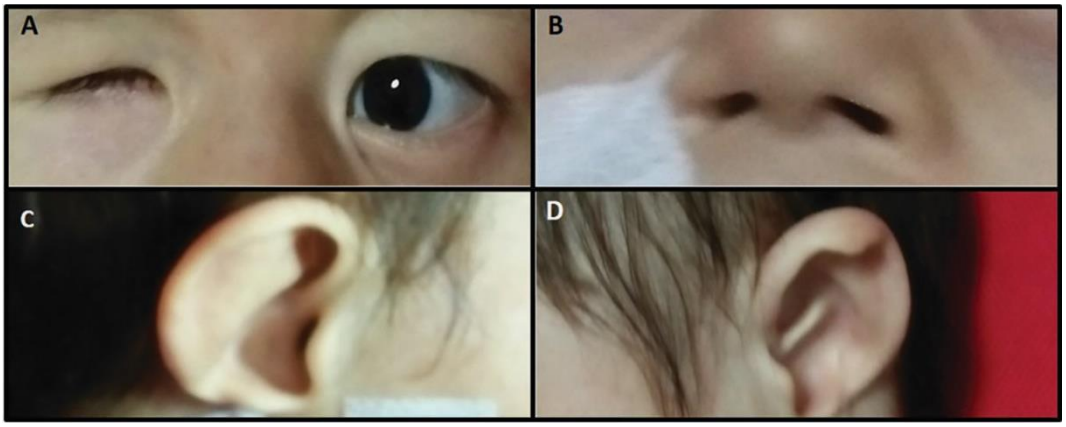

Figure 16. Facial gestalts of an Indonesian CS patient. The 5-year-old patient showed striking facial gestalts of CS such as (A) microphthalmia on the right eye, (B) small nostril especially on the left side, (C) prominent anti-helix discontinuous with antitragus, rudimentary right ear lobe and (D) thin and unfold helix, prominent anti-helix discontinuous with antitragus, and rudimentary left lobe.

In addition, he had type 1 laryngomalacia and neurogenic dysphagia. An audiologic BERA (brainstem evoked response audiometry) test was applied and revealed severe hearing loss in his right ear and profound hearing loss (deafness) in his left ear. The cardiac defect showed cardiomegaly, DORV (double outlet right ventricle) and severe pulmonary

stenosis. Other abnormalities were found, including horseshoe kidney, micropenis, and absence of testicles in the scrotum and inguinal region. Cerebral scan using MSCT (multiple slice computer tomography) showed arachnoid cyst in the sellar region.

4.2 Clinical Diagnosis and Chromosome Analysis

The clinical diagnosis of CS is based on combinations of phenotypic signs and symptoms that are sorted as major and minor criteria. The patient met the diagnostic criteria of CS defined by Hale *et al*. presenting both major and minor criteria, including 2 major and 7 minor characteristics (Table 11). The clinical diagnosis of CS was established in the first year by pediatricians at Diponegoro National Hospital, Semarang, Indonesia. Furthermore, to exclude large chromosomal aberrations, the patient's karyotype was analyzed, revealing a normal male karyotype of 46, XY (Figure 17).

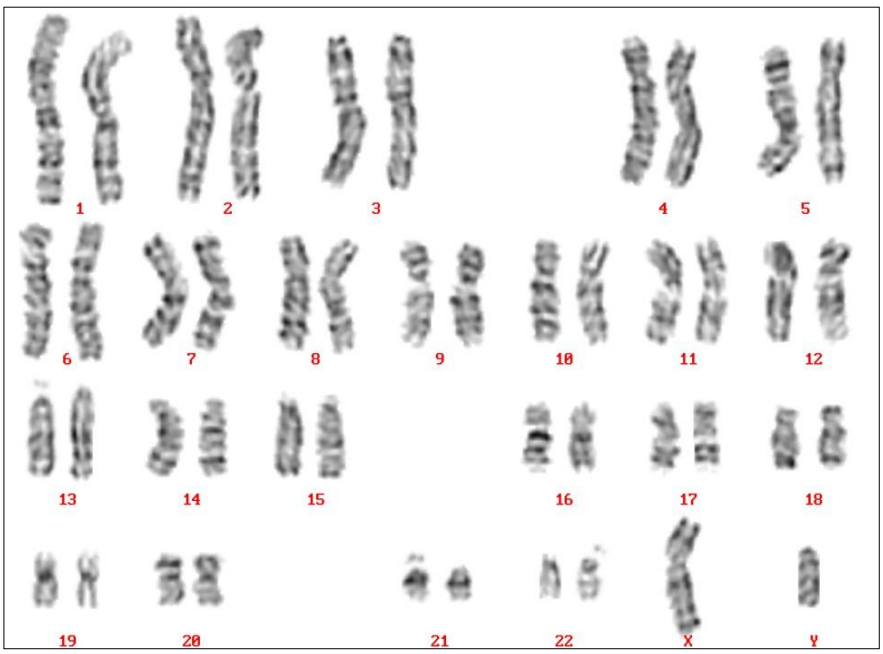

Figure 17. Patient's karyotype showed a normal male karyotype of 46, XY.

Table 11. Clinical features of the patient

Diagnostic criteria	Clinical features of the patient	Frequency in CS patients (Hale *et al.*, 2016)
Major		
- Coloboma	+ (iris coloboma, microphthalmia)	75%
- Choanal atresia or cleft lip or palate	-	44%
- Abnormal external, middle or inner ears, including hypoplastic semicircular canals	+ (hypoplasia and abnormal pinnae on both ear cartilage)	100%
- Pathogenic *CHD7* variant	+	67-90%
Minor		
- Cranial nerve dysfunction including hearing loss	+ (IX/X)	70-90% (IX/X)
- Dysphagia/feeding difficulties	+	80%
- Structural brain anomalies	+ (arachnoid cyst in the sellar region)	50%
- Developmental delay/intellectual disabilities/autism	+	100%
- Hypothalamo-hypophyseal dysfunction (gonadotropin or growth hormone deficiency) and genital anomalies	+ (micropenis, cryptorchidism)	64-100%
- Heart or esophagus malformation	+ (DORV)	75%
- Renal anomalies	+ (horseshoe kidney)	46%
- Skeletal/limb anomalies	-	80%

Abbreviations: DORV, double outlet right ventricle

4.3 Molecular Analysis by Targeted NGS Gene Panel

We determined the whole exonic and flanking intronic sequences of two genes involved in CS, *CHD7* and *EFTUD2*. *HOXA1* (OMIM 142955) was included in the panel as alterations in this gene are involved in features

overlapping with CS (Tischfield *et al*., 2005). These analyses identified two non-pathogenic variants in the *HOXA1* gene, eleven non-pathogenic heterozygous variants in the *EFTUD2* gene, and detected sixteen non-pathogenic heterozygous variants in the *CHD7* gene (Table 12). Moreover, we identified a heterozygous nonsense variant in exon 34 of the *CHD7* gene: NM_017780.3:c.7234G>T or p.(Glu2412Ter) (Figure 18A). We confirmed this finding by PCR amplification of exon 34 followed by Sanger sequencing. This variant was confirmed as *de novo* because we did not detect it in the DNA of either of his parents (Figure 18B). As this novel variant has not been reported before, we deposited it in the CHD7 database (https://www.chd7.org, last accessed October 02, 2018).

Table 12. Detected nonpathogenic variants in *HOXA1*, *CHD7* and *EFTUD2* genes in the patient

No.	Gene	Variant	Location	Allele Call	RefSNP (rs)
1.	*HOXA1*	c.907T>C	Exon 2	Heterozygous	rs76826100
2.	*HOXA1*	c.218C>T	Exon 1b	Homozygous	rs10951154
3.	*CHD7*	c.1665+34G>A	Intron 2	Homozygous	rs7836586
4.	*CHD7*	c.1666-147C>A	Intron 2	Homozygous	rs6471901
5.	*CHD7*	c.1666-103_1666-96del	Intron 2	Heterozygous	rs142885557
6.	*CHD7*	c.2238+39G>A	Intron 4	Homozygous	rs4540437
7.	*CHD7*	c.2376+42_2376+48ins	Intron 5	Homozygous	rs201756536
8.	*CHD7*	c.2442-93A>G	Intron 6	Homozygous	rs10448027
9.	*CHD7*	c.2614-45A>G	Intron 8	Homozygous	rs6471902
10.	*CHD7*	c.3779-217G>A	Intron 15	Homozygous	rs6471905
11.	*CHD7*	c.3989+51C>A	Intron 16	Homozygous	rs7005873
12.	*CHD7*	c.4186-225A>G	Intron 17	Homozygous	rs11986059
13	*CHD7*	c.4533+46A>G	Intron 19	Homozygous	rs7844902
14.	*CHD7*	c.5666-189G>T	Intron 28	Heterozygous	rs3763591
15.	*CHD7*	c.*1189A>C	Exon 38	Heterozygous	rs4738834
16.	*CHD7*	c.*1444T>G	Exon 38	Heterozygous	rs4237040
17.	*CHD7*	c.*1728T>C	Exon 38	Heterozygous	rs2280917
18.	*CHD7*	c.*1797C>T	Exon 38	Heterozygous	rs6985168
19.	*EFTUD2*	c.1014A>G	Exon 28	Heterozygous	rs2231647
20.	*EFTUD2*	c.588T>C	Exon 28	Heterozygous	rs2304986
21.	*EFTUD2*	c.2562-40T>C	Intron 25	Heterozygous	rs2289673
22.	*EFTUD2*	c.2046-102A>G	Intron 20	Heterozygous	rs16971033
23.	*EFTUD2*	c.2045+71A>G	Intron 20	Heterozygous	rs6503405
24.	*EFTUD2*	c.2045+28T>C	Intron 20	Heterozygous	rs58089352
25.	*EFTUD2*	c.1059-31A>G	Intron 12	Heterozygous	rs2120276
26.	*EFTUD2*	c.620-133C>T	Intron 8	Heterozygous	rs2289676
27.	*EFTUD2*	c.619+67A>G	Intron 8	Heterozygous	rs1443463
28.	*EFTUD2*	c.272-91T>C	Intron 3	Heterozygous	rs75343286
29.	*EFTUD2*	c.-735T>G	5'UTR	Heterozygous	rs2277617

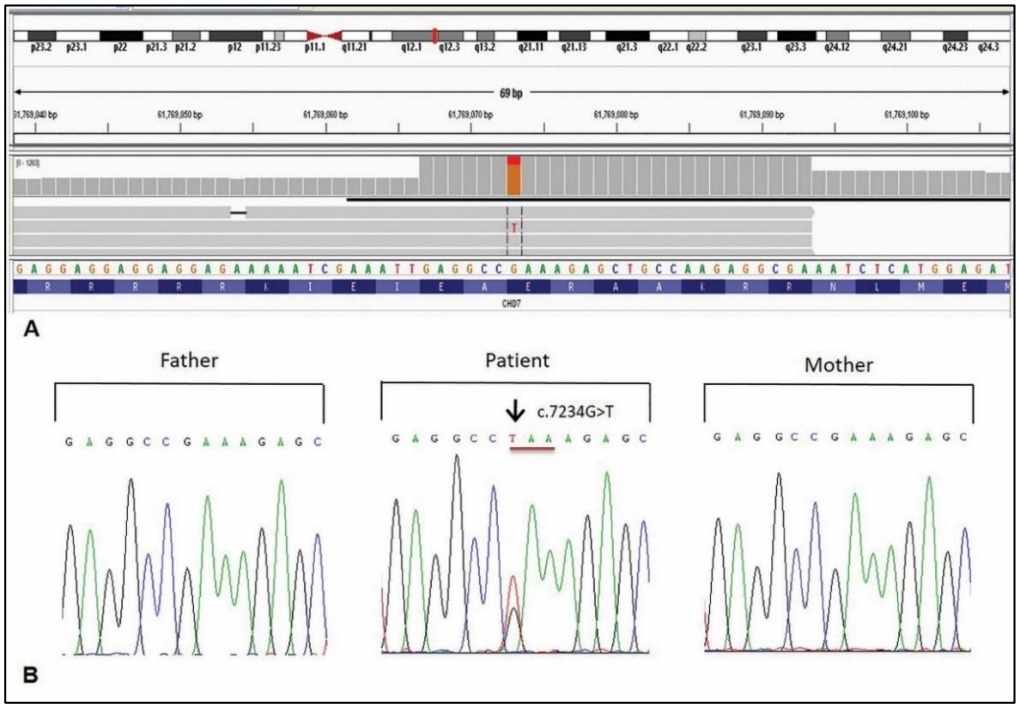

Figure 18. Pathogenic *CHD7* variant identified as the genetic cause of CS in an Indonesian patient. (A) Heterozygous nucleotide substitution at NM_017780.3:c.7234G>T identified by next-generation sequencing. (B) Sanger sequencing confirmed a *de novo* heterozygous nucleotide substitution (c.7234G>T) in exon 34 of *CHD7*. This change resulted in a premature stop codon p.(Glu2412Ter).

According to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) Standards and Guidelines for the Interpretation of Sequence Variants (Richards *et al.*, 2015), this variant is classified as pathogenic: PVS1 (very strong evidence of pathogenicity because it is a null variant), PS2 (*de novo*, both paternity and maternity confirmed), PM2 (variant absent in control population), PP3 (deleterious effect confirmed by multiple computational evidence). Thus, we identified the genetic cause of CS in this patient, and we added this disease-causing *CHD7* variant as a third major criterion for CS diagnosis (Table 11).

CHAPTER V

RESULTS

5.1 Development of a Functional Assay for CHD7 Protein

5.1.1 Site-Directed Mutagenesis

The objective of this study was the development of a functional test for CHD7 protein involved in the CHARGE syndrome. For this purpose, we have expressed in human cell lines the CHD7 protein, either wild-type or missense or any amino acid sequence variant found in CHARGE patients, to look for functional consequences of the expression of these variants. In the initial experiment, we have used two plasmids for expressing CHD7: pCIneo-CHD7-HA, expressing CHD7 HA-labeled in C-terminal and pcDNA3-FLAG-CHD7, expressing CHD7 FLAG-labeled in N-terminal. Briefly, using both plasmids, three missense variants (G1982W, R2319C, L1302P) that found in patients by our laboratory were generated. We also generated these variants, an insertion variant (1801insDGHGT), and two other CHD7 protein variants (T894A and A2160T) using pCIneo-CHD7, encoding untagged native CHD7. The two latter variants presenting one amino acid substitution but classified as benign using GnomAD software predictive tool and have been reported as polymorphisms (Vuorela *et al*., 2007; Wincent *et al*., 2008; Bartels *et al*., 2010). All of the studied CHD7 variants are summarized in table 13.

Table 13. Denomination of the Studied CHD7 Variants

Nucleotide Change	Amino Acid Change	Abbreviation
c.5944G>T	p.(Gly1982Trp)	G1982W
c.6955C>T	p.(Arg2319Cys)	R2319C
c.3905T>C	p.(Leu1302Pro)	L1302P
c.2680A>G	p.(Thr894Ala)	T894A
c.6478G>A	p.(Ala2160Thr)	A2160T
c.5405-17G>A	p.(His1801_Gly1802insAspGlyHisGlyThr)	1801insDGHGT

To confirm the desired *CHD7* variants, the plasmids were subsequently sequenced. We found that in each mutagenesis reaction, all transformants contained the desired variants (Figure 19-21).

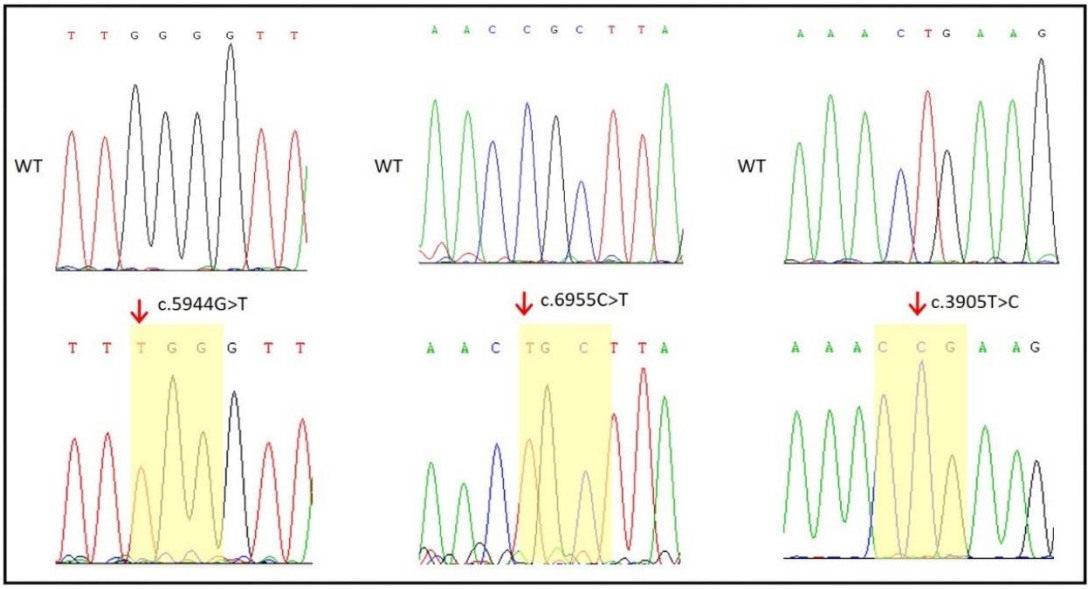

Figure 19. Partial electropherogram results of site-directed mutagenesis for missense variants. The arrow indicates the nucleotide change and the yellow shading indicates the impacted codon. The missense variants obtained are as follows: p.(Gly1982Trp) (or G1982W), p.(Arg2319Cys) (or R2319C) and p.(Leu1302Pro) (or L1302P).

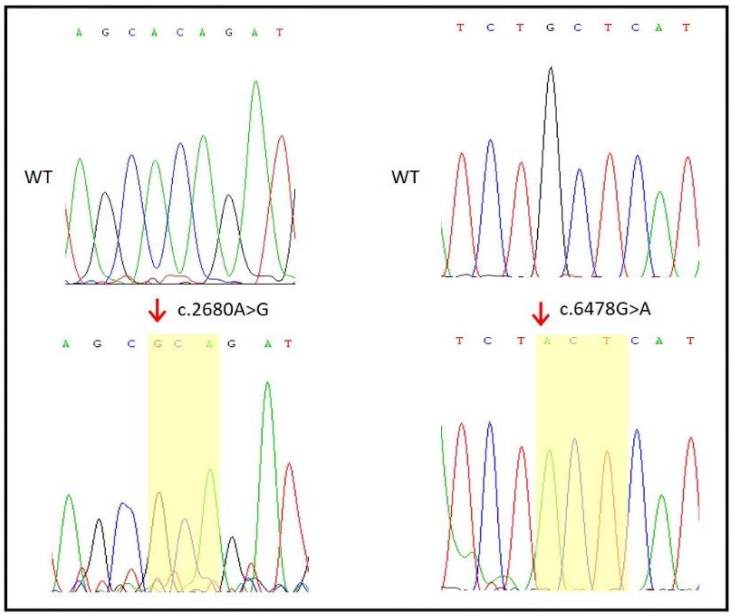

Figure 20. Partial electropherogram results of site-directed mutagenesis for polymorphism variants. The arrow indicates the nucleotide change and the yellow shading indicates the impacted codon. The obtained missense variants are as follows: p.(Thr894Ala) (or T894A) and p.(Ala2160Thr) (or A2160T). These nucleotide changes have been classified as non-pathogenic variants.

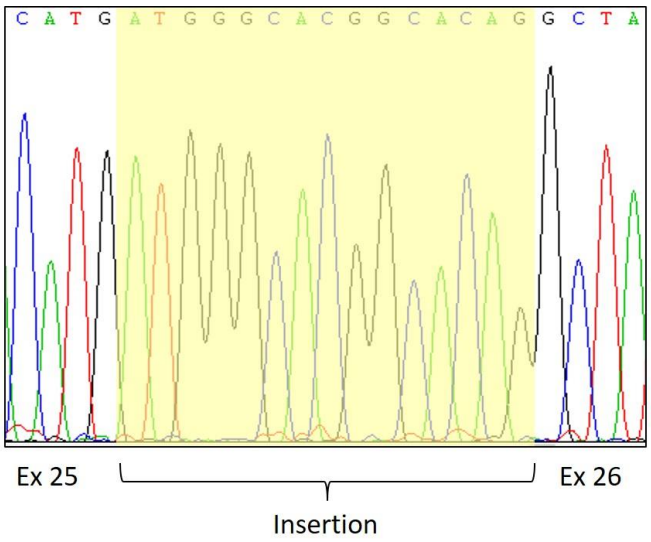

Figure 21. Partial electropherogram result of site-directed mutagenesis for the insertion variant. The yellow shading indicates the 15-intronic-nucleotide-insertion resulting from c.5405-17 G>A mutation in intron 25.

5.1.2 Optimal Transfection Condition

In the initial experiments, we studied several conditions to establish an optimal transfection condition in 6-well culture plates. The plasmid encoding CHD7-HA was transiently transfected into HeLa cells. As a negative control, an expression vector that does not contain the desired gene (mock) was used. Various plasmid DNA concentrations, volumes of lipofection reagent, and incubation times post-transfection were tested. On the one hand, we tested a range of 1, 2, 4 and 8 µg plasmid with 5 or 10 µL of Lipofectamine reagent per well; on the other hand, we tried 1 or 2 µg of the plasmid with 3 µL of FuGENE per well. Two different exposure times of cells to the complexes were applied: 48 hours and 72 hours.

The transfection efficiency was assessed in HeLa cells by Western blot using an anti-HA primary antibody to test CHD7 protein expression (expected molecular weight of 340 kDa). We subjected the protein lysates on a 5% SDS-PAGE gel to study the protein of interest, CHD7. In parallel, the detection of β-actin protein was also performed on a 10% SDS-PAGE gel to verify the quality of the protein lysate and the uniformity of the protein amount loaded into each well. Indeed, we noted the presence of a band above 250 kDa that corresponds to the CHD7 protein (340 kDa), this band is absent from the control transfection (mock). This protein is therefore expressed, and we concluded that HeLa cells are transfected under these conditions. The highest transfection efficiency in HeLa cells was obtained within 48 hours by using 4 µg of plasmid DNA with 10 µL of Lipofectamine

2000 per well (Figure 22). On the contrary, the transfection using FuGENE was unsuccessful, although it had been repeated in various concentrations of plasmids. As a conclusion, the transfection of cultured HeLa cell is effective only when using Lipofectamine 2000 reagent.

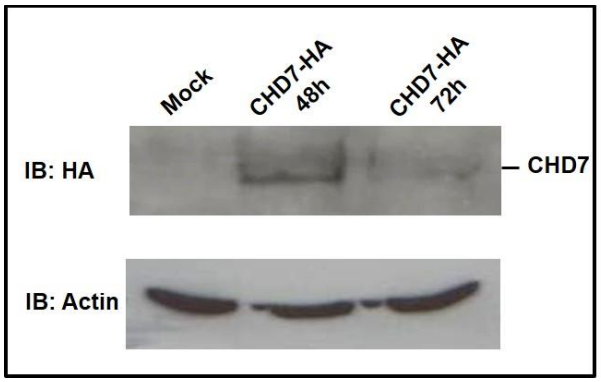

Figure 22. CHD7-HA protein expression. HeLa cells were transfected with an empty vector (mock) or with a plasmid encoding CHD7-HA (4 µg). 100 µg of protein lysates were separated by SDS-PAGE and analyzed by immunoblotting with an anti-HA antibody and anti-actin as a loading control. Maximum expression was obtained 48 hours after transfection.

In the next experiment, we tested FLAG-CHD7 encoding plasmid to overexpress CHD7 in HeLa, HEK293, and SH-SY5Y cells using the optimal transfection conditions established for the preceding plasmid. We analyzed the expression of the CHD7 protein by Western blot using anti-FLAG primary antibody. As predicted, FLAG-CHD7 was also evidence by Western blot in all cell models (Figure 23).

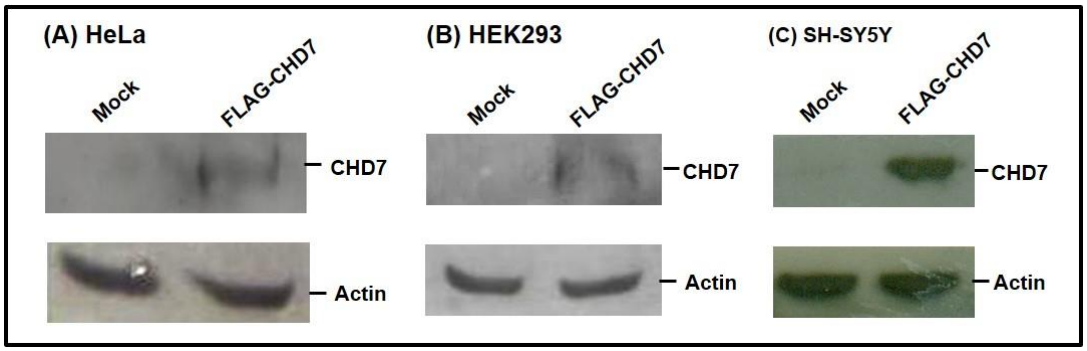

Figure 23. FLAG-CHD7 protein expression. HeLa (A), HEK293 (B), and SH-SY5Y (C) cells were transfected with an empty vector (mock) or with a plasmid encoding FLAG-CHD7. 100 µg (or 30 µg for actin control) of protein lysates were separated by SDS-PAGE and analyzed by immunoblotting with an anti-FLAG antibody or anti-actin as a control.

According to these results, in the further transfection experiments, we always used 4 µg of plasmid DNA with 10 µL of Lipofectamine 2000 followed by incubation of cells for 48 hours.

5.1.3 Localization of Overexpressed Wild-Type CHD7 Protein

To investigate the subcellular localization of the protein, HA-tagged CHD7 was expressed in HeLa and HEK293 cells. Double immunostaining was performed against HA and endogenous UBF, a protein expressed in the nucleoplasm and highly concentrated in the nucleolus (Figure 24).

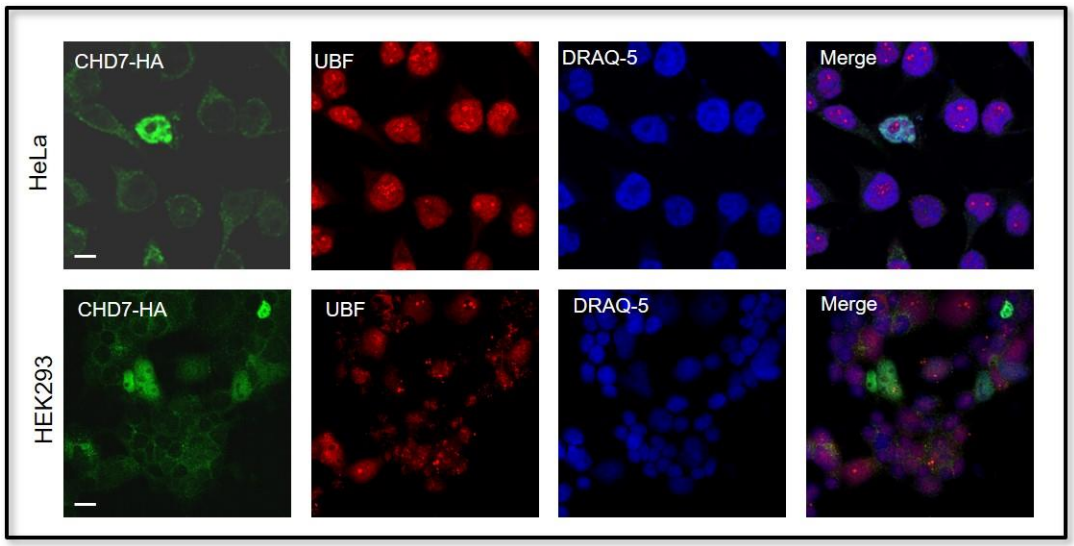

Figure 24. Localization of CHD7-HA protein. CHD7-HA was expressed in HeLa (top) and HEK293 cells (bottom). Immunostaining was visualized using confocal microscopy. The CHD7 protein was detected in green and UBF protein in red. The DRAQ5 dye was used to color the nuclei (blue). CHD7-HA protein was identified only in the nucleoplasm. Bar: 10 µm.

We observed that CHD7-HA was present throughout the nucleoplasm but absent of the nucleolus. As it was contradictory to some data of the literature, we hypothesized that the C-terminus tag might interfere with the trafficking of CHD7. Therefore, we performed in both cells a co-transfection experiment, to express FLAG-tagged CHD7 and a fusion protein between nucleolin (localized exclusively in the nucleolus) and the red fluorescent protein mCherry (Figure 25). This result confirmed that, in our hands, tagged CHD7 proteins localize exclusively in the nucleoplasm, but not in the nucleolus.

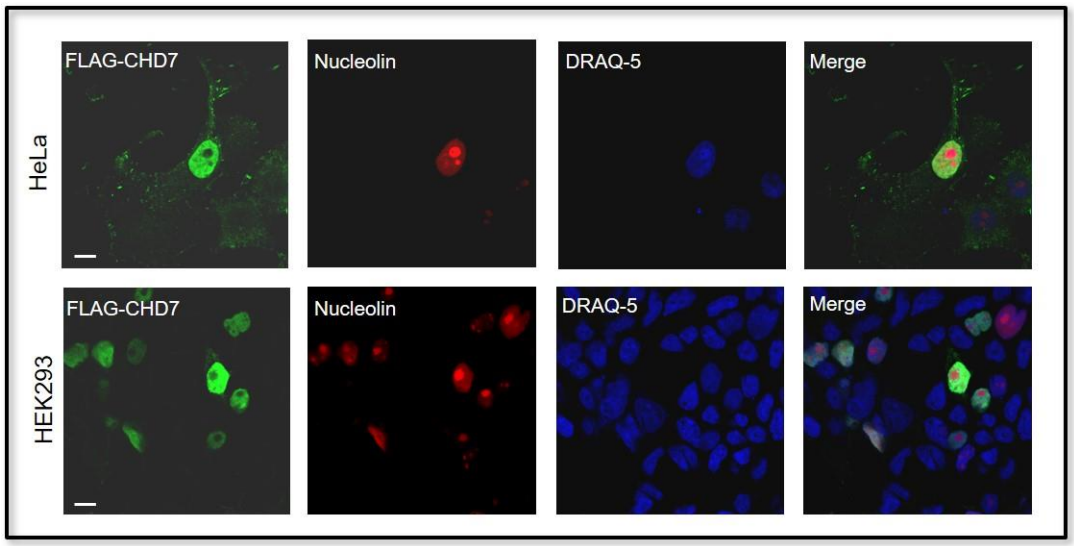

Figure 25. Localization of FLAG-CHD7 protein. HeLa (top) and HEK293 cells (bottom) were co-transfected to express FLAG-CHD7 and mCherry-nucleolin. The fluorescence was visualized using confocal microscopy. FLAG-CHD7 protein was detected in green and nucleolin in red. DRAQ5 dye was used to color the nuclei (blue). FLAG-CHD7 protein was localized only in the nucleoplasm. Bar: 10 µm.

Moreover, we continued to investigate the subcellular localization of the CHD7 protein in SH-SY5Y cells since CHD7 protein is highly expressed in neuronal progenitor cells (Feng *et al*., 2013). HA- and FLAG-tagged CHD7 were co-transfected with a fusion protein between nucleolin and the red fluorescent protein mCherry (Figure 26). However, HA- and FLAG-CHD7 protein were identified exclusively in the nucleoplasm. We concluded that the position of the tag was not involved in the localization of CHD7.

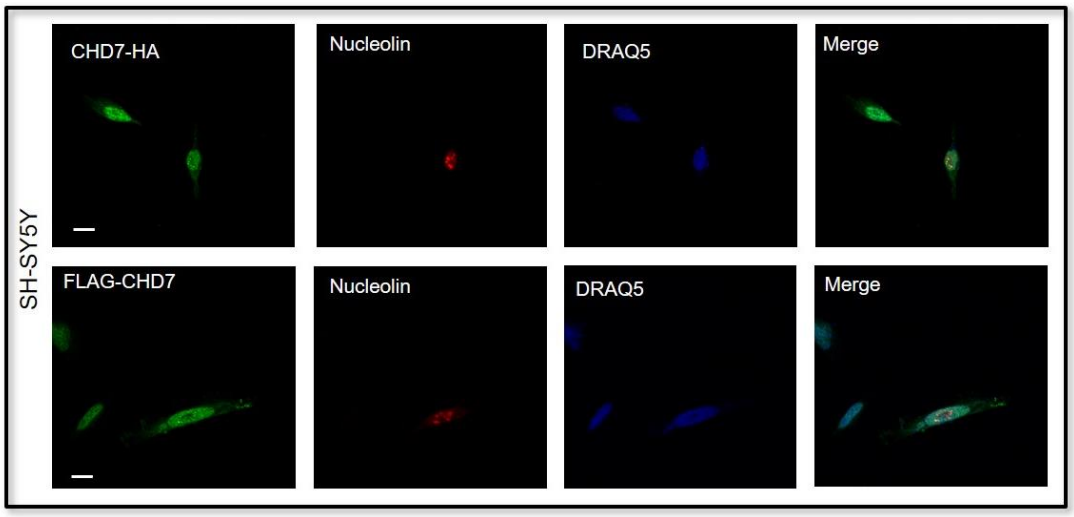

Figure 26. Localization of CHD7-HA and FLAG-CHD7 proteins. SH-SY5Y were co-transfected to express CHD7-HA or FLAG-CHD7 and mCherry-nucleolin. The fluorescence was visualized using confocal microscopy. CHD7 protein was detected in green and nucleolin in red. DRAQ5 dye was used to color the nuclei (blue). CHD7 protein was identified only in the nucleoplasm. Bar: 10 µm.

To check whether the tags were responsible of the results, we also performed immunostaining of endogenous CHD7 protein compared with untagged-CHD7 protein that we co-transfected with a fusion protein between nucleolin and the red fluorescence protein mCherry in HeLa cells (Figure 27). Contrary to the literature, repeated experiments performed the same results showing that, in our hands, CHD7 proteins localize exclusively in the nucleoplasm, but not in the nucleolus.

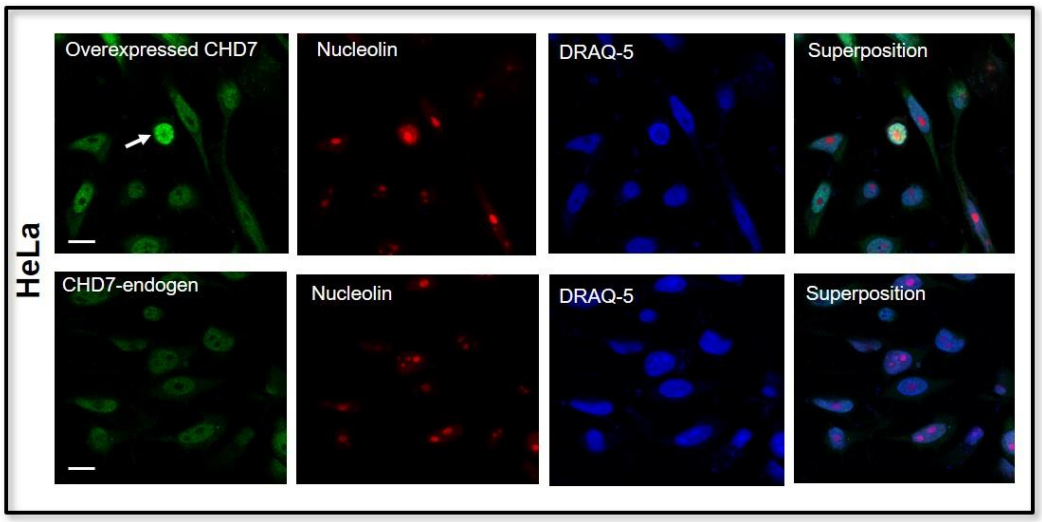

Figure 27. Localization of endogenous and overexpressed wild-type CHD7 protein. Immunostaining was visualized using confocal microscopy. CHD7 protein was detected in green and nucleolin in red. Arrow indicates a cell that overexpresses. CHD7 protein is localized mostly in the nucleoplasm, but not in the nucleolus. Bar: 10µm.

We have performed the experiments under the following conditions: we expressed untagged (UT)-, C- or N-terminal labeled CHD7 proteins in three cell lines (HeLa, HEK293, and SH-SY5Y) and used two different nucleolar markers (UBF or mCherry-nucleolin). In conclusion, we did not find in any case colocalization of CHD7 with a nucleolar protein, but the CHD7 protein was localized in the nucleoplasm. For further experiments, we have chosen pCIneo-UT-CHD7 plasmid to perform CHD7 functional analysis.

5.1.4 Expression and Localization of CHD7 Missense Variants

After plasmid transfection in HeLa cells, we evaluated the efficiency of CHD7 overexpression by Western blot using CHD7 rabbit monoclonal

antibody; then we analyzed the expression of the five missense CHD7 variants: T894A, A2160T, G1982W, R2319C, and L1302P (Figure 28). As expected, for WT and all variant forms of CHD7 protein, a specific protein band appeared at a position corresponding to a relative molecular mass of approximately 340 kDa. Hence, each of the plasmids that we used elicited the expression of the corresponding allele of CHD7 protein.

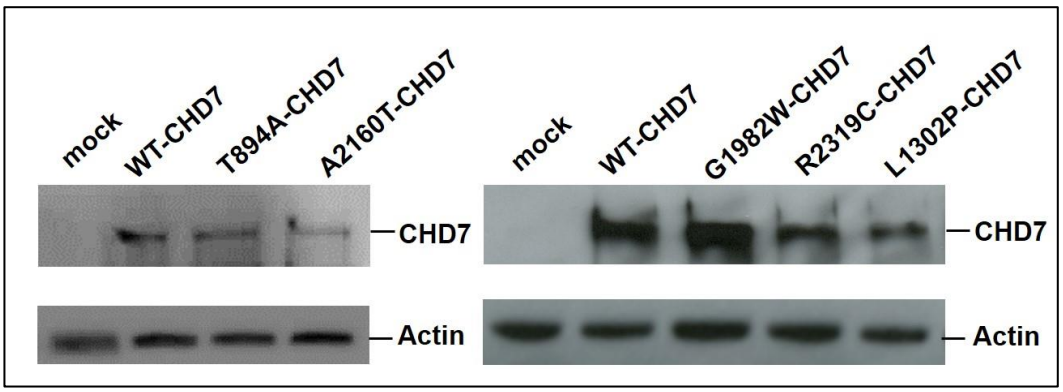

Figure 28. Expression of wild-type and variant forms of CHD7 protein in HeLa cells. For each transfection, 100 µg of protein lysates were separated by SDS-PAGE and analyzed by Western blot with anti-CHD7 antibody, or anti-actin as a loading control. The expected relative molecular mass of the CHD7 protein is 340 kDa.

To investigate the subcellular localization of the different forms (WT or missense variants) of CHD7 protein expressed in HeLa cells, we performed immunofluorescent staining followed by confocal microscopy imaging. We observed that the five missense variants, T894A-, A2160T-, G1982W-, R2319C- and L1302P-CHD7, exhibited the same localization as the WT form of the protein (Figure 29). In each case, CHD7 appeared localized exclusively in the nucleoplasm. In several cells, the intensity of CHD7

staining appeared higher but the localization remained confined to the nucleoplasm: therefore, higher overexpression level did not alter the observed localization. Moreover, the three amino acid substitutions found in patients had no impact upon CHD7 trafficking.

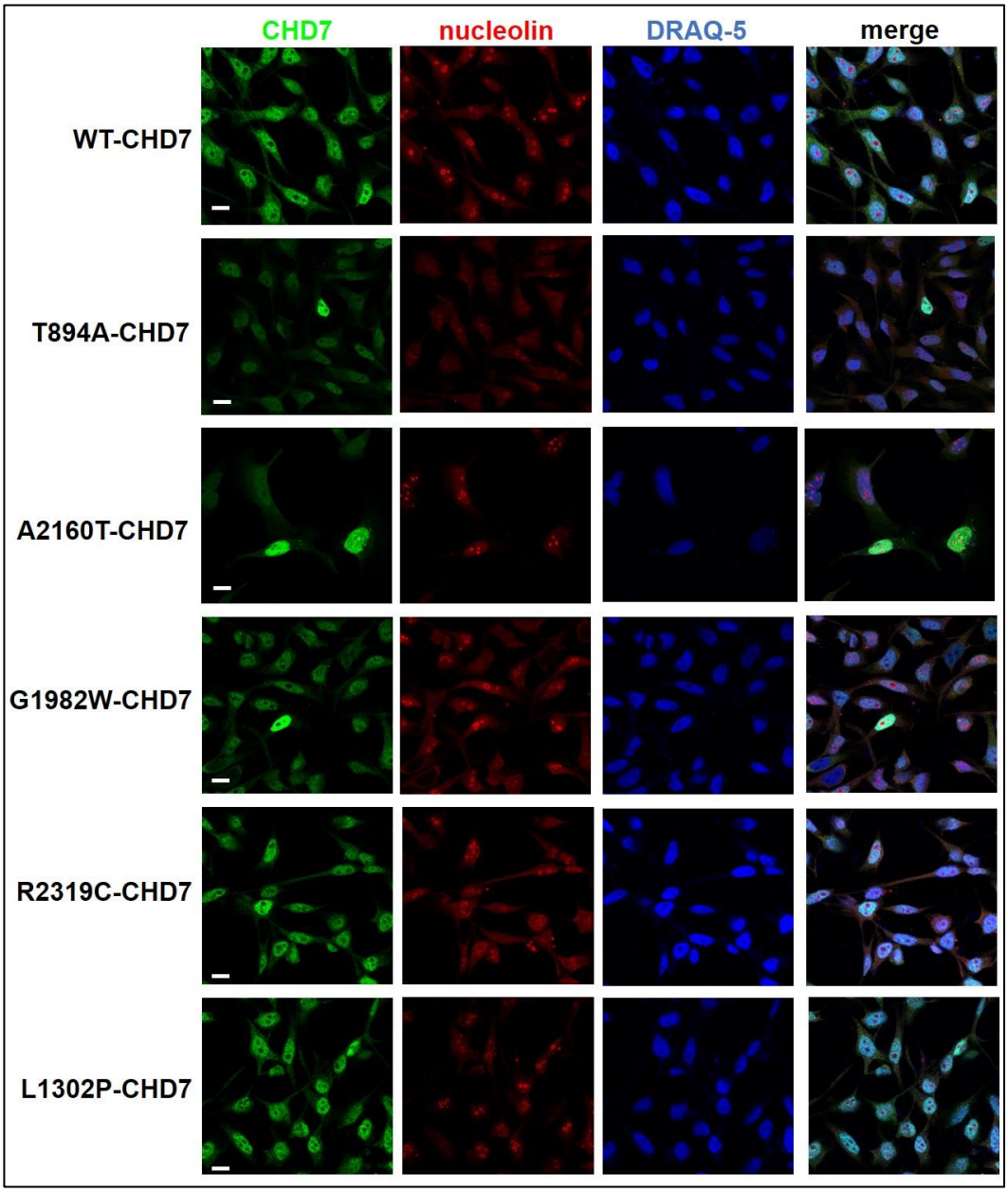

Figure 29. Nucleoplasmic localization of wild-type and variant forms of CHD7 protein expressed in HeLa cells. Immunostaining was visualized using confocal microscopy. CHD7 protein was detected in green and nucleolin in red. DRAQ5 dye was used to color the nuclei (blue). We observed no colocalization between both proteins, suggesting that none of the CHD7 protein variants are nucleolar. Bar: 10 µm.

5.1.5 Expression and Localization of CHD7 Insertion Variant

We previously described the finding in several CS patients of the c.5405-17G>A splicing mutation, localized in a hotspot of intronic mutation (Legendre *et al.*, 2017). This substitution leads to create a new 3' splice site (3'ss) (Figure 30). We assessed by minigene assay that it results in a fifteen intronic nucleotide insertion in the mRNA, presumably leading to the synthesis of a protein with five-amino acid insertion: p.(His1801_Gly1802insAspGlyHisGlyThr) or 1801insDGHGT.

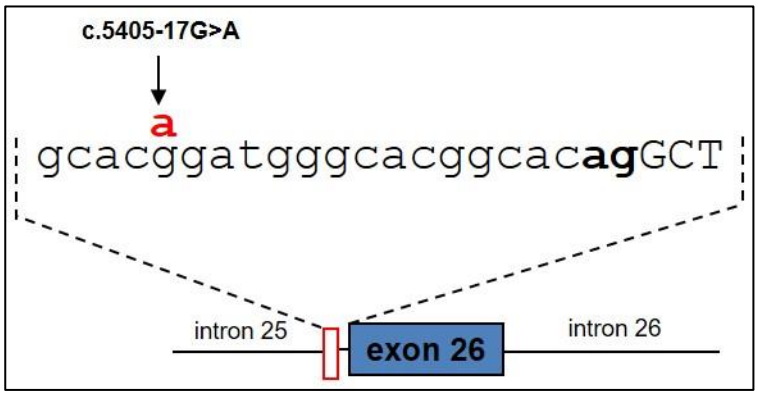

Figure 30. Mutation of c.5405-17G>A. The arrow indicates the nucleotide change form G>A in intron 25 resulting in a 15 intronic nucleotides insertion (Adapted from Legendre *et al.*, 2018).

We generated a plasmid encoding this variant, and we used it to transfect HeLa cells. We checked by Western blot that it elicited overexpression of the protein (Figure 31A). Furthermore, we also performed immunofluorescent staining followed by confocal microscopy imaging and revealed the same localization as the WT or missense variants of CHD7, localized exclusively in the nucleoplasm (Figure 31B).

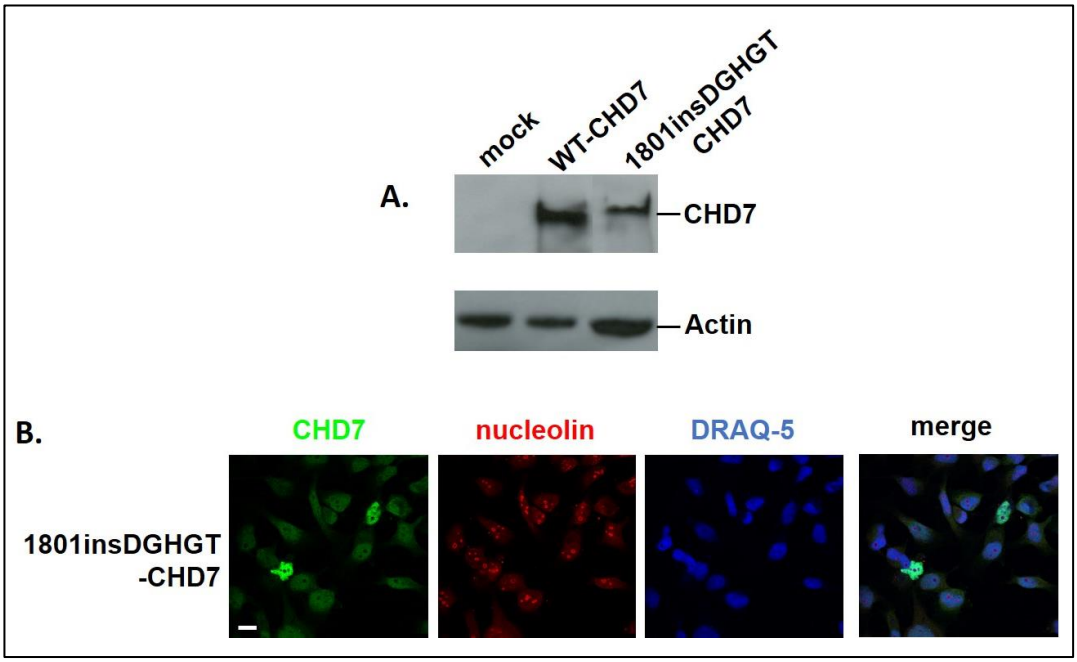

Figure 31. Expression and localization of an insertion variant of CHD7 protein in HeLa cells. (A) For each transfection, 100 µg of protein lysates were separated by SDS-PAGE and analyzed by Western blot with anti-CHD7 antibody, or anti-actin as a loading control. (B) Immunostaining was visualized using confocal microscopy. CHD7 protein was detected in green and nucleolin in red. DRAQ5 dye was used to color the nuclei (blue). CHD7 protein was localized only in the nucleoplasm. Bar: 10 µm.

5.1.6 Assessment of Wild-Type CHD7 Protein Functionality

First, we aimed to establish a functional test to evidence the functionality of normal CHD7 protein and subsequently to observe putative dysfunctions of CHD7 variants expressed in CS patients. CHD7 was proposed to regulate the transcription of numerous genes throughout development, by binding regulatory regions of the chromatin prior to remodeling. To investigate the functionality of *CHD7* alleles, we measured by qRT-PCR the consequences of *CHD7* overexpression upon the RNA

transcript amount for several genes whose transcription can be modulated by CHD7 according to the literature. We quantified 45S pre-ribosomal RNA (Zentner *et al*., 2010), and the mRNA of *SOX4*, *SOX10*, *ID2*, and *MYRF* genes (Feng *et al*., 2013; He *et al*., 2016).

When we overexpressed WT-CHD7 in HeLa cells, the amount of RNA transcribed from these genes were downregulated, compared to the mock-transfected cells (Figure 32): the transcript amounts were reduced to 39.4±4.5% of control for *45S rDNA*, 60.5±1.5% for *SOX4*, 48.9±2.7% for *SOX10*, 23.2±4.2% for *ID2* and 40.6±1.3% for *MYRF* gene.

To validate these results, we applied our biological assay to two CHD7 protein variants, T894A and A2160T, that are classified as non-pathogenic variant. After transfection in HeLa cells, we quantified the transcript amounts of the five reporter genes: in both cases, mRNA amounts were significantly decreased compared with the mock-transfection (Figure 32). Moreover, no statistically significant difference was observed compared with WT-CHD7 expression condition. Overall, we concluded that the reduction of the transcript amount of the reporter genes should result from the normal functionality of CHD7 protein in HeLa cells.

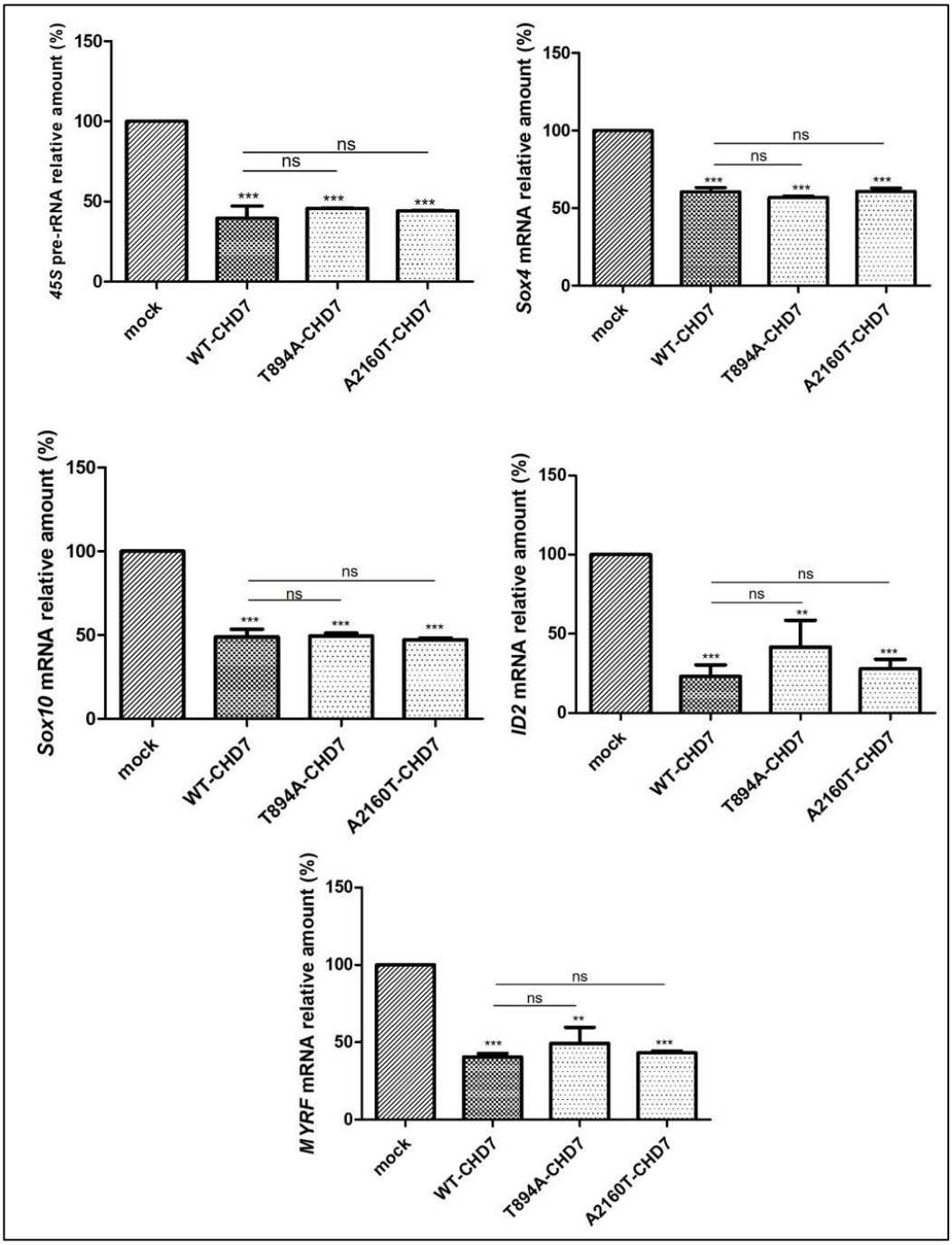

Figure 32. Impact of functional CHD7 alleles expression upon the transcription of five reporter genes. After transfection of wild-type or variant forms of CHD7, we performed qRT-PCR to evaluate the impact upon the transcription of genes regulated by CHD7: *45s rDNA*, *SOX4*, *SOX10*, *ID2*, and *MYRF*. RNA amounts are displayed as an n-fold difference relative to the mock-transfected condition. Data represent means and standard

deviation corresponding to at least three independent transfections. ns: non-significant, **P < 0.01, ***P < 0.001.

5.1.7 Functional Assay of CHD7 Missense Variants

When we expressed WT-CHD7 in HeLa cells, the amounts of RNA transcribed from reporter genes were downregulated, compared to the mock-transfected cells. By contrast, when we expressed either one of the three missense *CHD7* variant alleles, the transcript amount of each reporter gene was unaffected, compared with mock-transfected control condition (Figure 33). This result suggests that the normal ability of CHD7 protein to modify the transcription level of several genes was impaired by the three amino acid substitutions that we investigated.

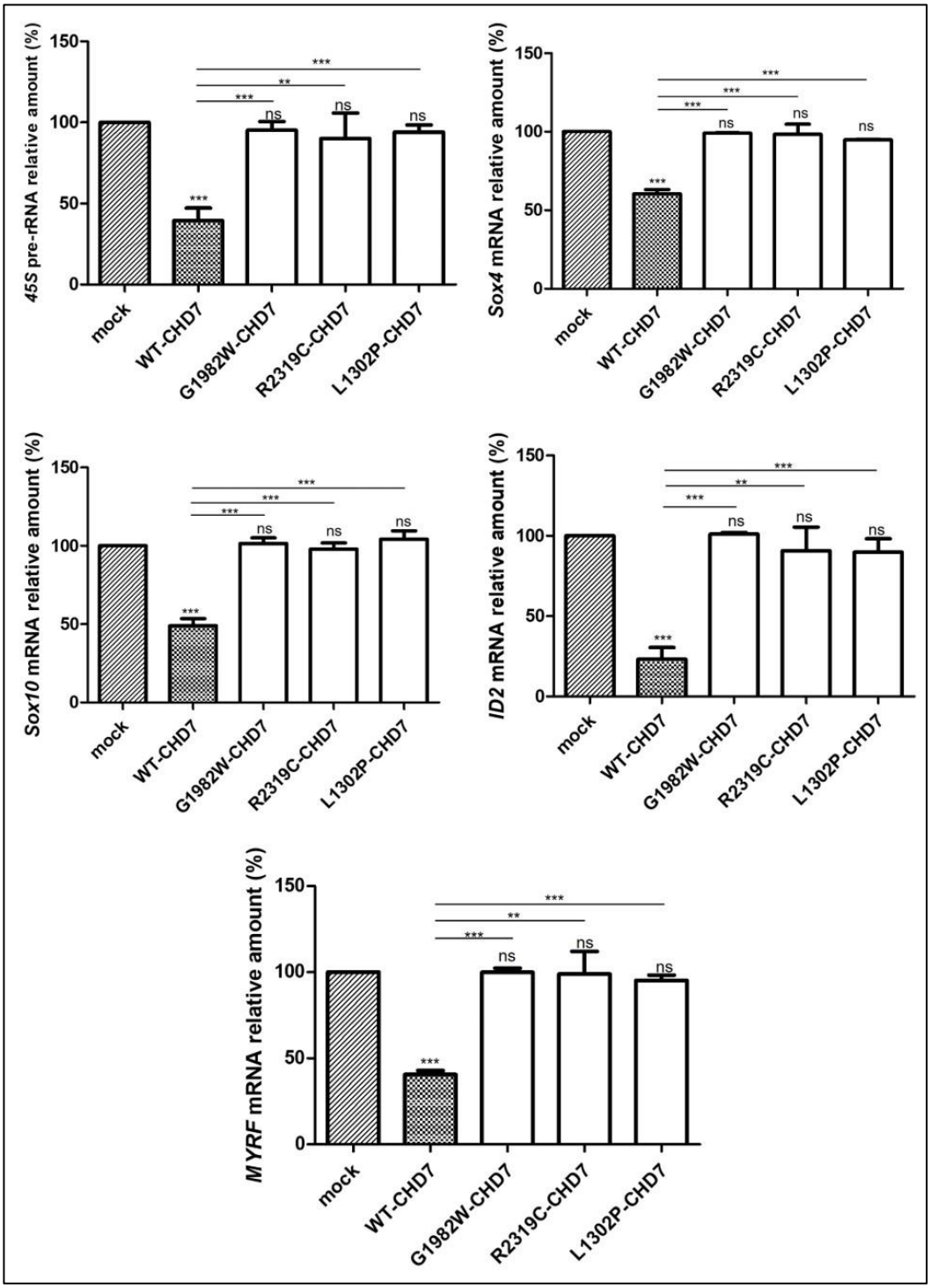

Figure 33. Impact of missense variants expression upon the transcription of five reporter genes. After transfection of wild-type or variant forms of CHD7, we performed qRT-PCR to evaluate the impact upon the transcription of genes regulated by CHD7: *45s rDNA*, *SOX4*, *SOX10*, *ID2*,

and *MYRF*. RNA amounts are displayed as an *n*-fold difference relative to the mock-transfected condition. Data represent means and standard deviation corresponding to at least three independent transfections. ns: non-significant, **P < 0.01, ***P < 0.001.

5.1.8 Functionality of CHD7 Insertion Variant

We evaluated the impact of 1801insDGHGT-CHD7 overexpression upon the transcription of the five reporter genes: we observed that their mRNA level was not significantly different from the mock-transfection condition (Figure 34). This suggests that the 1801insDGHGT-CHD7 protein variant is non-functional.

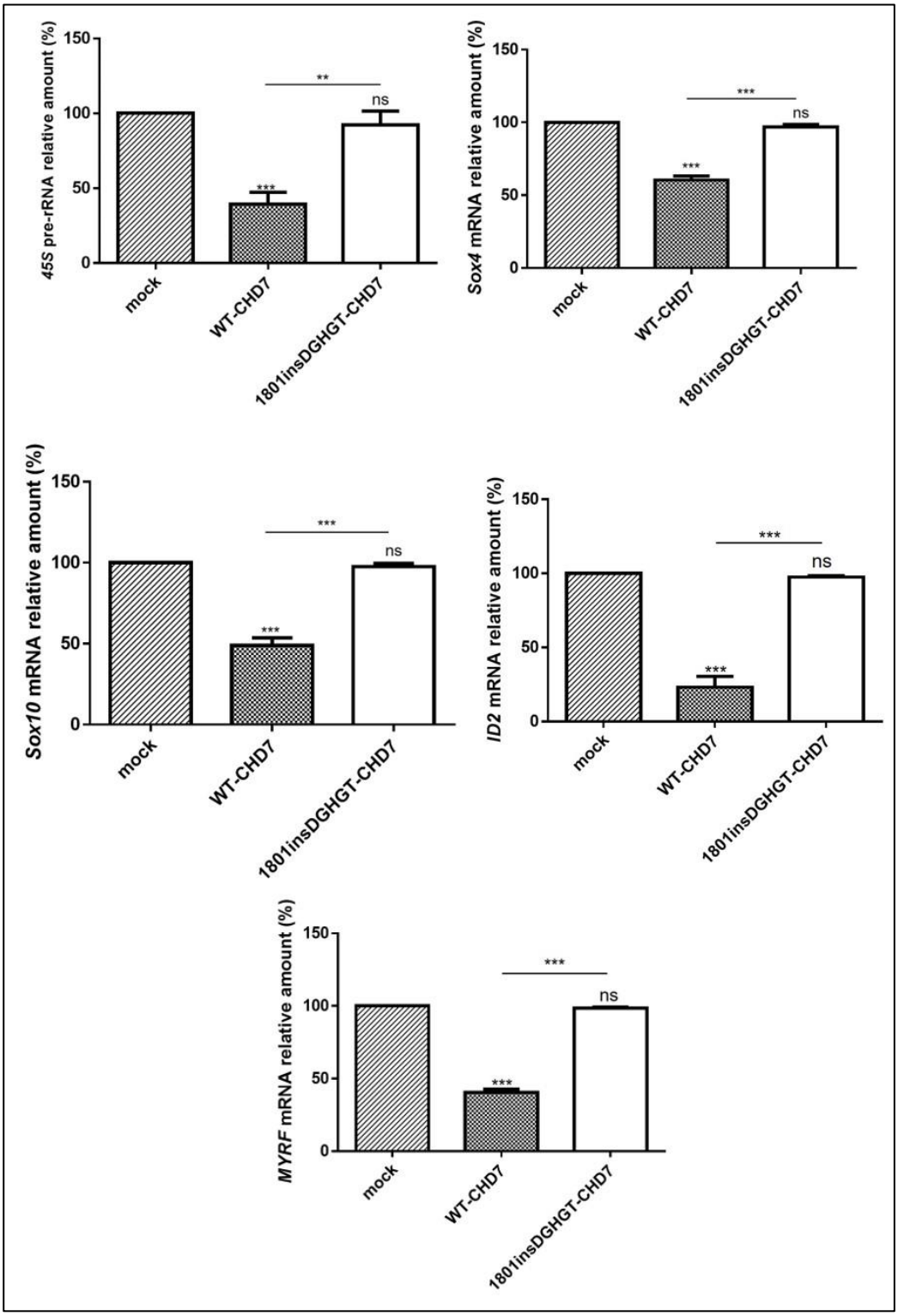

Figure 34. Impact of insertion variant expression upon the transcription of five reporter genes. After transfection of wild-type or insertion variant of CHD7, we performed qRT-PCR to evaluate the impact upon the

transcription of genes regulated by CHD7: *45s rDNA*, *SOX4*, *SOX10*, *ID2*, and *MYRF*. RNA amounts are displayed as an *n*-fold difference relative to the mock-transfected condition. Data represent means and standard deviation corresponding to at least three independent transfections. ns: non-significant, **P < 0.01, ***P < 0.001.

5.1.9 Sensitivity and Specificity Test

The sensitivity and specificity of the functional assay for CHD7 variants has been calculated using a set of 3 known pathogenic and 2 known non-pathogenic variants. Sensitivity has been estimated at 100% and specificity at 100%. However, the number of tested variants is still small.

$$\text{Sensitivity} = \frac{3}{3+0} \times 100\% = 100\%$$

$$\text{Specificity} = \frac{2}{0+2} \times 100\% = 100\%$$

5.2 Functional Assay of Endogenously Expressed CHD7 Missense Variants using CRISPR/Cas9 System

5.2.1 Generation of *CHD7* Gene Knock-out using CRISPR/Cas9 System

To disrupt the function of *CHD7* gene, a sgRNAs targeting the third exon of the *CHD7* gene was designed. This early exon was chosen in order to increase the chance of full loss-of-function occurrence (knock-out). In this experiment, we induced the activation of the NHEJ pathway to create small insertions or deletions (indels), so that frameshift mutations were obtained. The target sites are shown in figure 35.

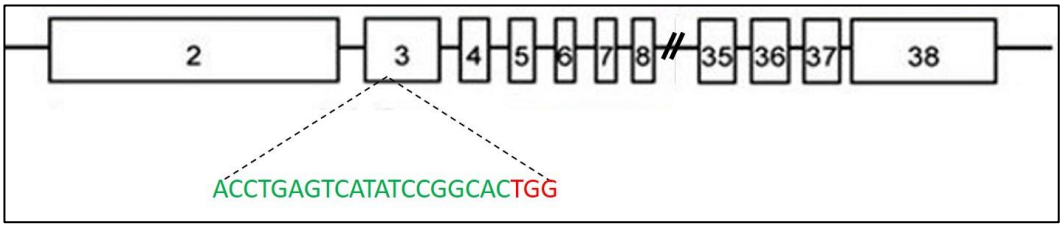

Figure 35. Generation of the *CHD7* gene knock-out using the CRISPR/Cas9 system. Schematic diagram of the sgRNA target site located in the exon 3 of the *CHD7* locus. *CHD7* exons are indicated by rectangles; target sites of the sgRNA sequence is highlighted in green; protospacer-adjacent motif (PAM) sequence is highlighted in red.

Furthermore, to generate *CHD7* knock-out, a plasmid encoding Cas9 and the sgRNA was transfected firstly into HeLa cells and then SH-SY5Y cells, subsequently. Puromycin selection was applied for 72 hours and isolated single cells were obtained through serial dilution.

5.2.2 Determining Genome Targeting Efficiency using T7 Endonuclease I

The genome targeting efficiency or mutation hit-rate at a specific locus needs to be evaluated. For this analysis, we used the enzyme mismatch cleavage method using T7 endonuclease I (T7EI). The enzyme recognizes and cleaves structural deformities in DNA heteroduplexes.

In this experiment, HeLa cells first transfected with Cas9/sgRNA plasmid and then puromycin selection was applied for 72 hours. Afterward, the cells were cultured several days with culture medium without puromycin to obtain a sufficient number of cells. Genomic DNA of a population of HeLa cells was extracted and amplified by PCR surrounding the CRISPR guide RNA target site. The PCR products were then annealed and digested with

T7 Endonuclease I. This enzyme recognized and cleaved DNA mismatches in those heteroduplexes. By running the cleavage products on a 1.5% agarose gel, full-length and cleavage products were resolved (Figure 36).

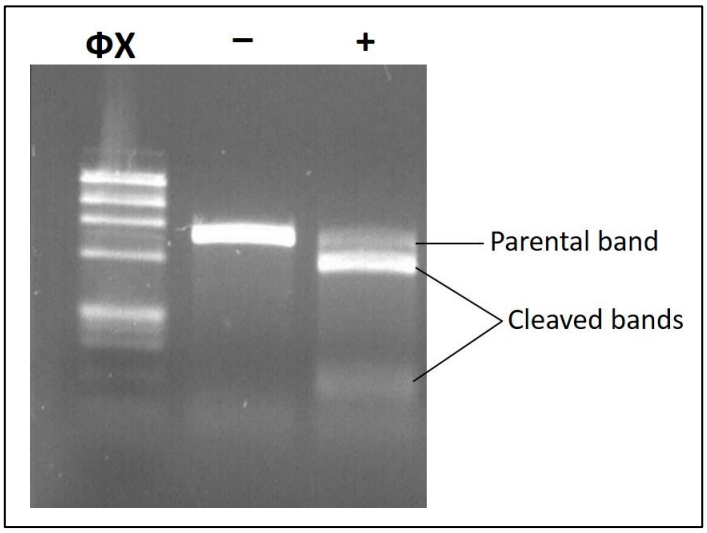

Figure 36. DNA mismatch detection assay. PCR products were treated with T7EI nuclease. Then, full-length and cleavage products were resolved by agarose gel electrophoresis. Gel shows DNA marker (ΦX174 DNA-Hae III), untreated (-), and edited cells (+).

The intensity of the respective bands allowed to calculate the gene editing percentage that has occurred as follows:

$$\text{Fraction cleaved} = \frac{128{,}32}{128{,}32 + 39{,}25} = 0.766$$

$$\% \text{ gene modification} = 100 \times [1 - (1 - 0.766)^{1/2}]$$
$$= 51.6\%$$

As a conclusion, the genome targeting by CRISPR/Cas9 system functioned properly, and the efficiency rate of this system was coherent with the literature, that is 51.6% (Sander and Joung, 2014). Therefore, the cells were

subsequently cloned to verify the genome modification in the specific targeted locus.

5.2.3 Verification of the *CHD7* Knock-Out in HeLa Cells

Clonal selection of edited HeLa cells was performed. Genomic DNA from 12 clonal cells was extracted, and the exon 3 of *CHD7* was amplified. Five of 12 PCR products were ligated with pGEM-T Easy vector to be transformed into *E. coli*. Finally, 16 independent colonies of each clone were selected and sequenced. These results are summarized in table 14.

Table 14. Nucleotide Changes in Exon 3 of *CHD7* in Edited HeLa Cells

Clones	Nucleotide Changes
3.2	- 21 nucleotides deletion - 6 nucleotides deletion
3.4	204 nucleotides insertion
5.8	- 11 nucleotides deletion - 2 nucleotides insertion
5.9	- 30 nucleotides deletion - 8 nucleotides deletion
5.11	2 nucleotides deletion

Briefly, in HeLa cell we have chosen clones 5.8 and 5.11 since they have 11-nucleotide deletion (c.1702_1712delCCGGATATGAC) and 2-nucleotide insertion (c.1702_1703insTC) for the first one, and homozygous 2-nucleotide deletion (c.1702_1703delCC) for the second one, in exon 3 of *CHD7*. These variations would change the reading frame and cause premature termination of translation at a new nonsense codon (Figure 37).

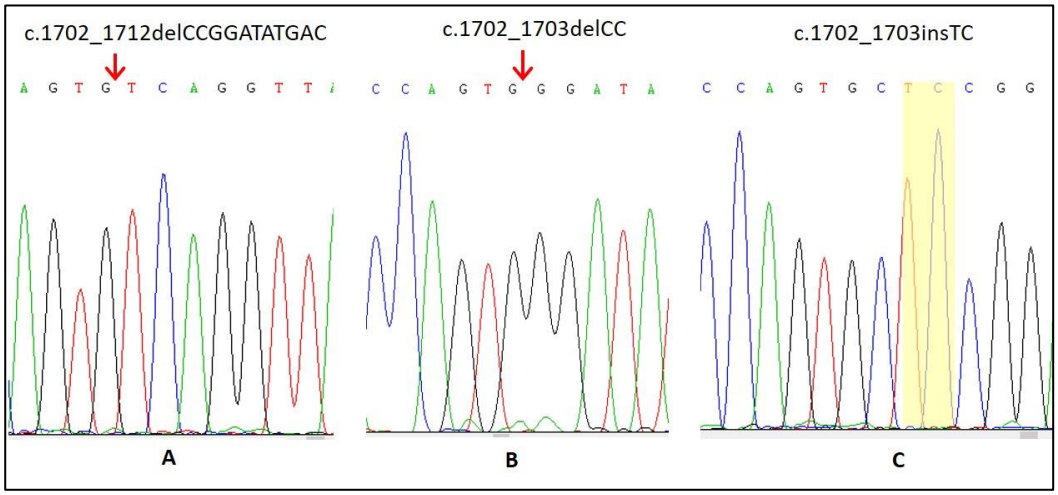

Figure 37. Insertion deletion (indels) identification in exon 3 of *CHD7* by Sanger sequencing. Deletions (A), (B), and insertion (C) were detected in HeLa cells. Arrow indicates the deletion site and yellow shading indicates the nucleotides insertion. This insertion or deletions will lead to a frameshift.

5.2.4 Functionality of CHD7 Knock-out in HeLa Cells

After having successfully generated *CHD7* (-/-) HeLa cells, we first evaluated the *CHD7* mRNA expression level by qRT-PCR to validate the two clones (5.8 and 5.11). As expected, compared with the parental HeLa cell line, the CHD7 mRNA level was reduced to 34.2±2.4% in both *CHD7* (-/-) HeLa clones (Figure 38). The decrease of mRNA amount is likely caused by indel that created a premature stop codon, presumably leading to mRNA removing via nonsense-mediated mRNA decay (NMD) process.

We then quantified the transcript amounts of the five reporter genes in both *CHD7*(-/-) clones. Unexpectedly, in both clones, the mRNA amount of each reporter gene appeared unaffected and the expression levels were not different from the parental HeLa cells (Figure 38). As a conclusion, in the HeLa cell line, the endogenous expression of CHD7 is not involved in

regulating the transcription of these reporter genes. From this result, we assumed that HeLa cells model may not suitable for functional assay of endogenous CHD7. We then tried generating *CHD7* knock-out in SH-SY5Y neuroblastoma cells.

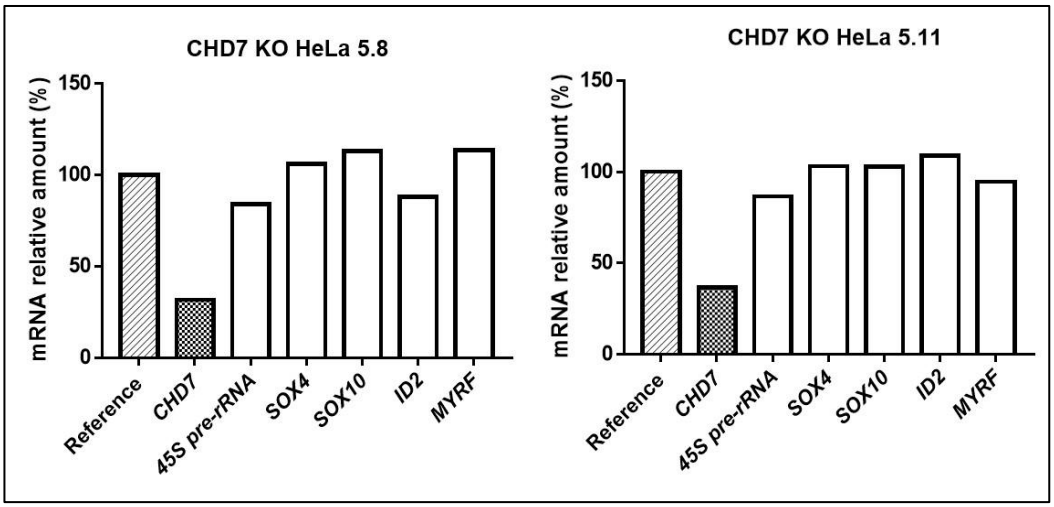

Figure 38. Impact of *CHD7* knock-out in HeLa cells upon the transcription of *CHD7* and five reporter genes. We performed qRT-PCR using clones 5.8 and 5.11 of HeLa cells in which we invalidated *CHD7* gene to evaluate the impact upon the transcription of *CHD7* and of genes regulated by CHD7: *45s rDNA*, *SOX4*, *SOX10*, *ID2*, and *MYRF*. RNA amounts are displayed as an *n*-fold difference relative to the 100% reference levels measured in the parental HeLa cells.

5.2.5 Verification of the *CHD7* Knock-Out in SH-SY5Y Cells

Clonal selection of edited SH-SY5Y cells was performed. Genomic DNA from 12 clonal cells was extracted and the exon 3 of *CHD7* was amplified. Three of 12 PCR products were ligated with pGEM-T Easy vector to be transformed into *E. coli*. Finally, 16 independent colonies of each

clone were selected and sequenced. The results are summarized in table 15.

Table 15. Nucleotide Changes in Exon 3 of *CHD7* in Edited SH-SY5Y Cells

Clones	Nucleotide Changes
1	2 nucleotides deletion
2	1 nucleotides deletion
3	- 2 nucleotides deletion - 9 nucleotides deletion

We have selected clones 1 and 2 since they have homozygous deletion: 2 nucleotides (c.1702_1703delCC) and 1 nucleotide deletions (c.1701delG), respectively (Figure 39).

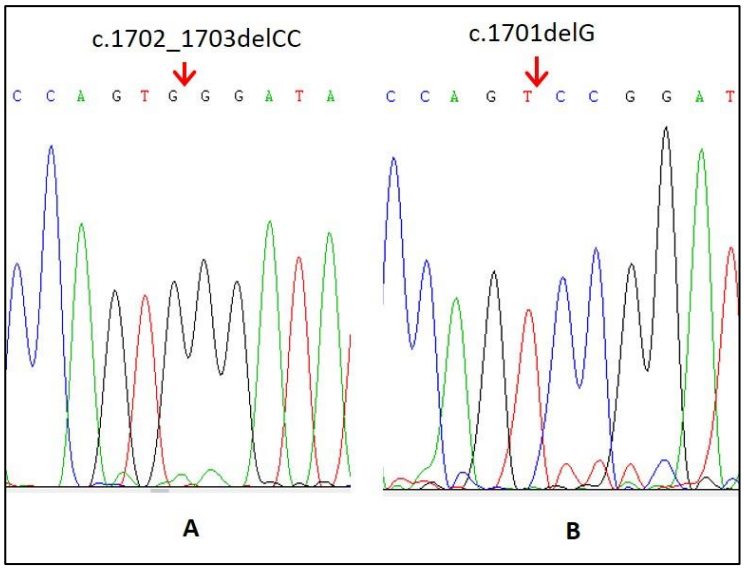

Figure 39. Deletions identification in exon 3 of *CHD7* by Sanger sequencing. Deletions (A) and (B) were detected in edited SH-SY5Y cells. Arrow indicates the deletion site. These deletions will lead to a premature termination codon.

5.2.6 Functionality of CHD7 Knock-out in SH-SY5Y Cells

After having successfully generated *CHD7* invalidation in SH-SY5Y cells, we further measured the *CHD7* mRNA expression level by qRT-PCR using clone 1. As expected, compared with the parental SH-SY5Y cell line, the *CHD7* mRNA level was reduced to 27.2±0.6% in *CHD7* (-/-) SH-SY5Y cells (Figure 40). We tested the transcript amounts of the five reporter genes in this *CHD7* (-/-) SH-SY5Y cells. The amount of RNA transcribed from these genes was upregulated, compared to parental SH-SY5Y cells (Figure 40). The transcript amounts were increased to 253±9.1% of reference for *45S rDNA*, 194±3.3% for *SOX4*, 240±23.4% for *SOX10*, 231±0.4% for *ID2* and 544±6.4% for *MYRF* gene.

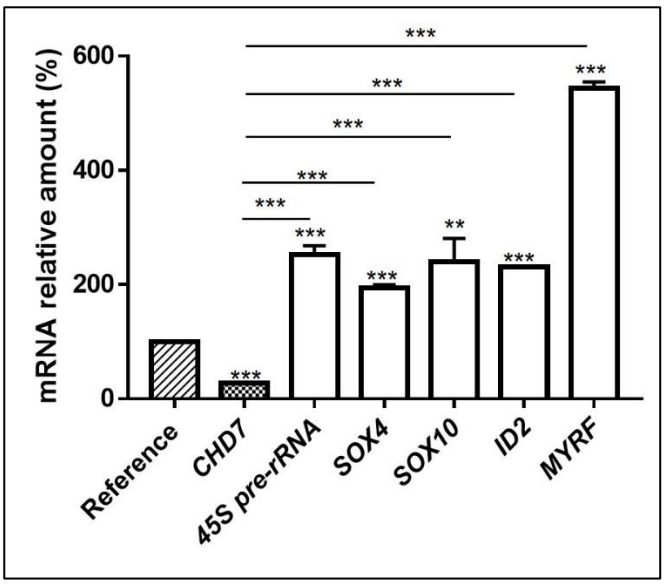

Figure 40. Impact of *CHD7* knock-out in SH-SY5Y cell upon the transcription of *CHD7* and five reporter genes. After invalidation of *CHD7* gene in SH-SY5Y cells, we performed qRT-PCR using clone 1 to evaluate the impact upon the transcription of *CHD7* and genes regulated by CHD7: *45s rDNA*, *SOX4*, *SOX10*, *ID2* and *MYRF*. RNA amounts are displayed as

an *n*-fold difference relative to the 100% reference levels measured in the parental SH-SY5Y cells. Data represent means and standard deviation corresponding to at least three independent transfections. **P < 0.01, ***P < 0.001.

According to these results, we assumed that in SH-SY5Y cell line, the endogenous expression of CHD7 contributes to inhibiting the transcription of these reporter genes. Therefore, to generate an endogenous expression of CHD7 missense variants, we determined to use the SH-SY5Y neuroblastoma cell line.

5.2.7 Generation of Endogenously Expressed *CHD7* missense variants in SH-SY5Y Cells using CRISPR/Cas9 System

Having successfully used CRISPR/Cas9 to knock-out *CHD7* gene in SH-SY5Y cells, we further performed CRISPR/Cas9 for targeted *CHD7* gene modification to obtain the missense variants that we studied previously. The target sites are shown in figure 41.

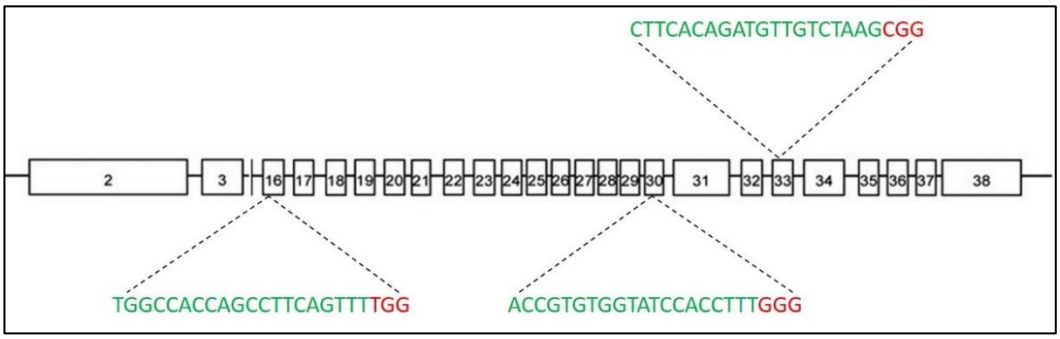

Figure 41. Generation of the *CHD7* missense variants using the CRISPR/Cas9 system. Schematic diagram of the sgRNA target site located in the exon 16, 30, and 33 of the *CHD7* locus. *CHD7* exons are indicated by rectangles; target sites of the sgRNA sequence are highlighted in green; protospacer-adjacent motif (PAM) sequences are highlighted in red.

To generate *CHD7* missense variants, in each case, a plasmid encoding Cas9/sgRNA and ssODN, used as HDR template, were co-transfected into SH-SY5Y cells. Puromycin selection was applied for 72 hours and isolated single cells were obtained through serial dilution. The genomic DNA from 16 cell clones was isolated. We performed PCR amplification using these genomic DNAs as a template. Finally, to verify the results, each PCR product was sequenced. However, repeated experiments using ssODNs shown that the cell clones did not contain the desired variants. Therefore, alternatively we performed these experiments with the same protocol but using linearized double-stranded targeting plasmid as a repair template. By using this type of HDR, we obtained a clone containing the homozygous mutation L1302P (Figure 42).

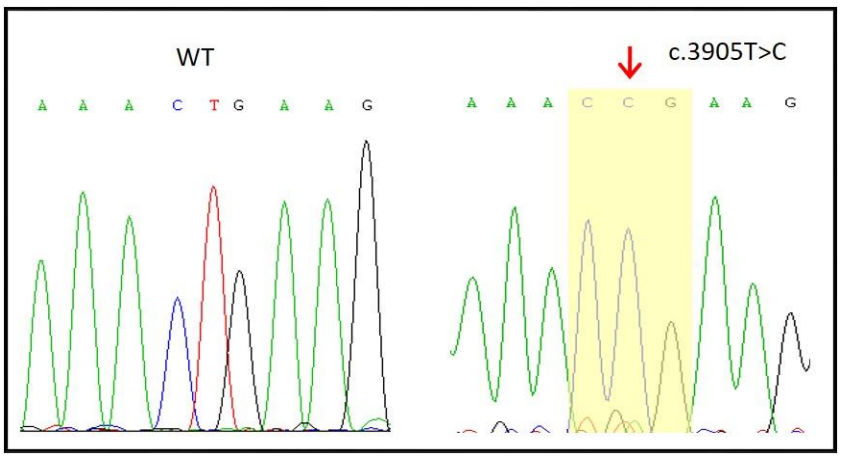

Figure 42. Partial electropherogram result of targeted DNA base change using CRISPR/Cas9. The arrow indicates the nucleotide change and the yellow shading indicates the changed codon p.(Leu1302Pro) or L1302P.

Unfortunately, until now we have not yet obtained the two other variants (G1982W and R2319C) even though we repeated transfections twice to generate these variants.

5.2.8 Functional Assay of Endogenously Expressed CHD7 Missense Variant

To investigate the functionality of L1302P-CHD7 missense variant, we measured by qRT-PCR the RNA transcript amount for several genes (*45S rDNA*, *SOX4*, *SOX10*, *ID2*, and *MYRF*) as we previously described. In edited SH-SY5Y cells, the amount of RNA transcribed from these genes were upregulated to 235±38.4% compared to WT-CHD7 for *45S rDNA*, 190.2±13.8% for *SOX4*, 191.1±8.5% for *SOX10*, 232.4±1.4% for *ID2*, and 544.1±5.5% for *MYRF* gene (Figure 43).

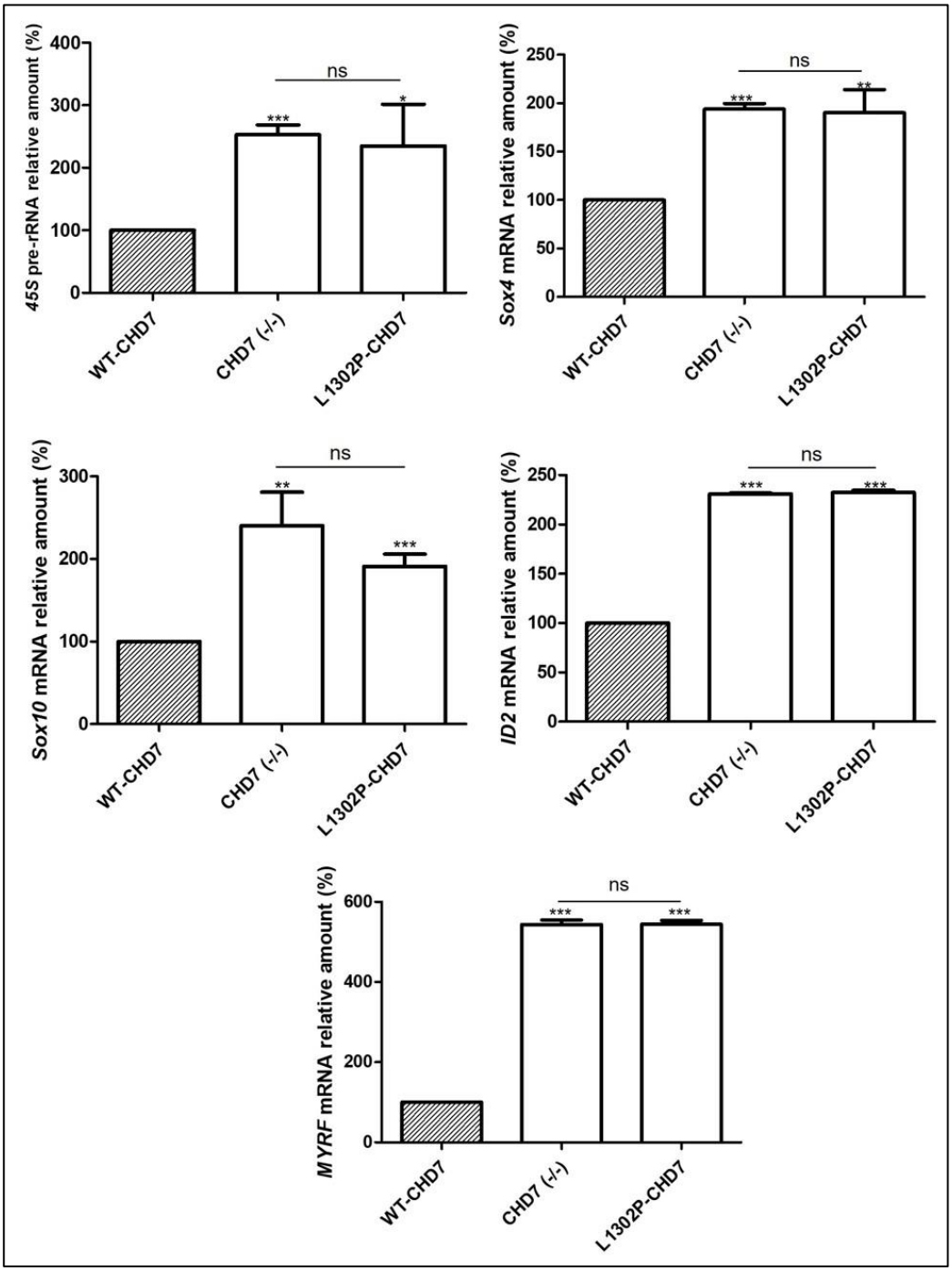

Figure 43. Impact of endogenously expressed L1302P-CHD7 variant upon the transcription of five reporter genes. We performed qRT-PCR to evaluate the impact upon the transcription of genes regulated by CHD7: *45S rDNA*, *SOX4*, *SOX10*, *ID2*, and *MYRF*. RNA amounts are displayed as an n-fold difference relative to the WT-expression condition. Data represent means

and standard deviation corresponding to at least three independent transfections. ns: non-significant, *P < 0.05, **P < 0.01, ***P < 0.001.

After having successfully modified genomically both alleles of *CHD7* gene to encode the L1302P variant, we quantified the transcript amounts of the five reporter genes: mRNA amounts were significantly increased compared with the WT-*CHD7* (Figure 43). Furthermore, no statistically significant difference was observed compared with *CHD7* (-/-) cells. This result suggests that the normal ability of CHD7 protein to modify the transcription level of several genes was impaired by the L1302P variant.

CHAPTER VI

GENERAL DISCUSSION

6.1 Novel *CHD7* Mutation in an Indonesian CHARGE Syndrome Patient

The CS diagnosis is primarily based on combinations of clinicopathological findings that are sorted as major and minor criteria. Our patient presented both major and minor characteristics of CS, including 4 major and 7 minor characteristics (Table 11). The patient was therefore clinically diagnosed as CS according to the novel clinical criteria (Hale *et al*., 2016). Most CS patients have a normal karyotype; however, a cytogenetic analysis must be performed first to exclude chromosomal abnormalities and other syndromes overlapping CS. Although this patient showed similar dysmorphism to previously reported cases of CS, we resorted to molecular analysis to establish a conclusive diagnosis.

The newly found mutation c.7234G>T or p.(Glu2412Ter) resulted in a premature stop codon, presumably leading to mRNA removing via nonsense-mediated mRNA decay (NMD) process. Therefore, we assume that this mutation represents a null allele, causing the disorder due to haploinsufficiency. It has been noted that nine other nonsense mutations have been found in exon 34 of *CHD7* so far. However, the mutation p.(Glu2412Ter) was not present in the CHD7 database, suggesting it to be novel. This novel mutation was subsequently submitted to the CHD7 database (https://www.chd7.org). This novel mutation was sporadic, as in most CS cases described to date (Lalani *et al*., 2006; Jongmans *et al*., 2006).

The birth of a child with a severe genetic disease poses considerable psychological, social, and economic problems. A genetic counselor should provide recurrence-risk estimates to the parents in such cases. Even if the parents have had a child affected by apparently *de novo CHD7* alterations, there is still a recurrence risk of 1-2% because of germline mosaicism (Lalani *et al*., 2012). Therefore, prenatal genetic testing should be recommended to the mother during subsequent pregnancies to determine whether the fetus has a risk of having CS.

There are several approaches to establish a molecular diagnosis of clinically suspected CS. Sanger sequencing of 38 exons and flanking intronic sequences of the *CHD7* gene is considered obsolete, as it is more time-consuming and eventually more expensive than NGS (Frank *et al*., 2013). Currently, sequencing of large panels of genes involved in intellectual disability (ID) is widely used, occasionally leading to the identification of CS-causing *CHD7* mutations (Grozeva *et al*., 2015). However, as the clinical diagnosis of CS relies on well-established criteria, sequencing of CS-targeted gene panels appears more straightforward. Moreover, as this approach requires the use of lower-capacity NGS equipment, it allows testing at a lower cost than sequencing of large gene panels or genome-wide analysis. Therefore, this approach could be applied in low- and middle-income countries as a cost-effective strategy for routine diagnosis of CS.

6.2 Functional Assay of CHD7 Variants

6.2.1 Patients and CHD7 Alterations

The G1982W variant results from a substitution in exon 30 of the *CHD7* gene. This variant was identified in a rare familial CHARGE with multiple affected members due to a segregating *CHD7* mutation. We identified this missense variant c.5944G>T or p.(Gly1982Trp) in two siblings with typical CS. Molecular analysis of their mother shows normal *CHD7* sequence. Their late father had unilateral deafness, and it is unfortunate that the *CHD7* molecular analysis had never been done. Nevertheless, an uncle from the father side also harbors this variant. From the pedigree analysis, we assumed that the variant must have been inherited from the father side (Figure 44).

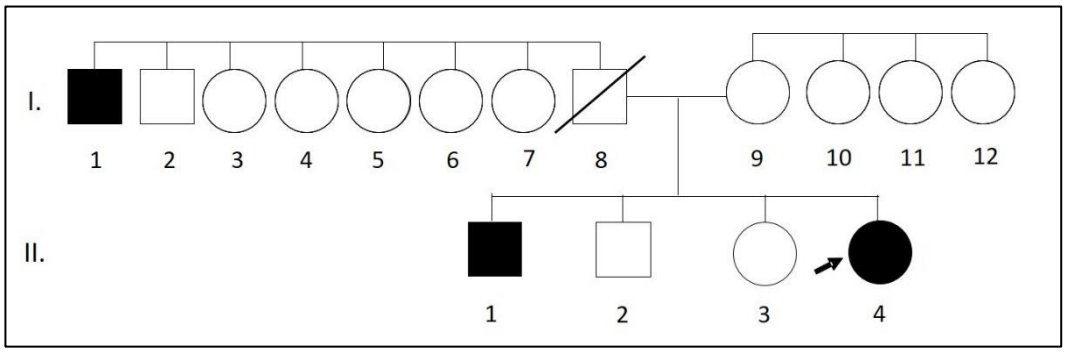

Figure 44. Pedigree of a familial CHARGE with G1982W-CHD7 alteration. Healthy members are illustrated with open symbols, established cases of CS harboring c.5944G>T variant marked with filled symbols (I-1, II-1, II-4). The patient (index case) is indicated with black arrow.

The G1982W variant is located in the SANT-like domain (Figure 45), which is present in many subunits of the chromatin remodeling complexes.

Its precise function is unclear, but it could be involved in binding to DNA and histones (Hall and Georgel, 2007).

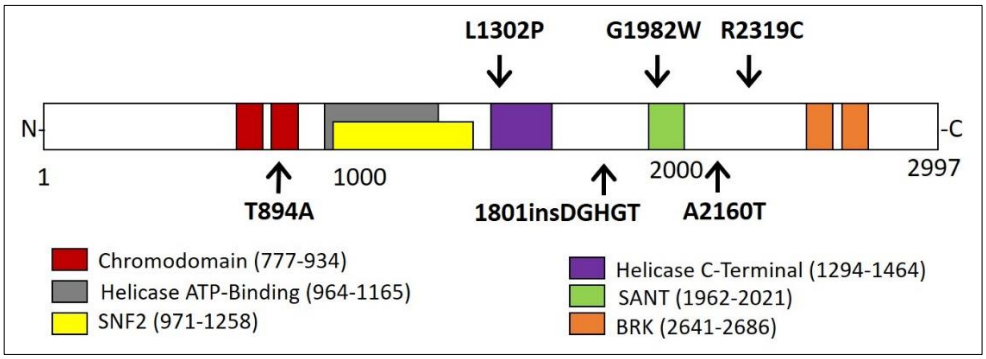

Figure 45. Position of studied amino acid sequence variants. Arrows indicate the positions of the amino acid sequence (Adapted from http://www.ebi.ac.uk/interpro/protein/Q9P2D1).

The amino acid in position 1982 is changed from a glycine (a nonpolar-neutral amino acid) to a tryptophan (a polar-neutral amino acid). In addition, the substitution of a glycine, the smallest amino acid, to a tryptophan, the largest amino acid, is probably deleterious because it may destabilize the folding of the SANT domain (Figure 46).

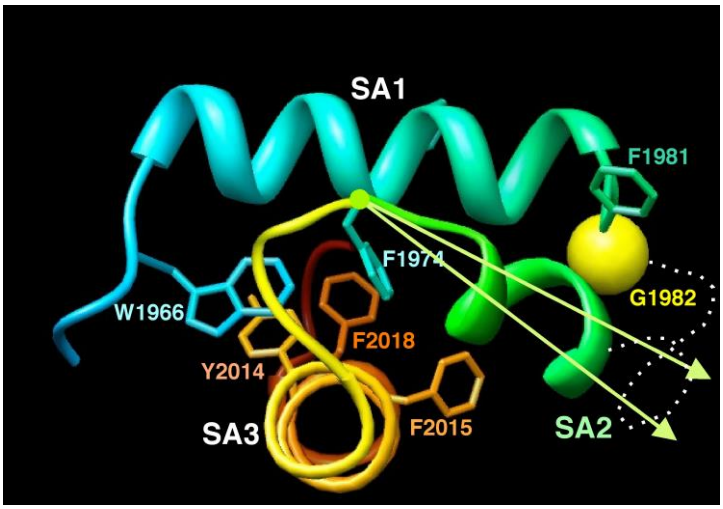

Figure 46. Computational modeling of SANT domain of CHD7 (amino acid position: 1962-2021) (Isabelle Callebault, personal communication, 2012).

The R2319C variant was identified in a case of partial CHARGE. The nucleotide substitution in the variant is CG to TG, in which is a most frequent substitution in the human genome due to deamination of 5' methyl-cytosine (Antonarakis *et al.*, 2000). The substitution affects amino acid 2319, from an arginine (a polar-basic amino acid) to a cysteine (a polar-neutral amino acid). These two amino acids having different properties, it may result in a modification of the three-dimensional conformation of CHD7. However, this mutation lies in a CHD7 region whose sequence presents no homology with any consensus domain (Figure 45).

The L1302P missense variant was identified in a patient with typical CS. The amino acid 1302 is changed from a leucine to a proline, which may cause a strain in the folding of the protein. Moreover, this substitution is localized in a helicase domain (Figure 45), whose integrity may be essential for the normal function of CHD7. The pathogenicity of this variant is consistent with a previous report stating that pathogenic missense mutations were mainly found in the highly conserved middle exons including the chromo-, helicase- and SANT domains (Bergman *et al.*, 2012).

The variation c.5405-17G>A, lies in a recurrent hotspot of intronic mutations. This substitution leads to the creation of a new 3' acceptor splice site (3'ss) that can overcome the weakness of the natural acceptor site. This elicits the maturation of an mRNA encoding the 1801insDGHGT-CHD7 protein variant, presenting a five-amino acid insertion between the helicase and the SANT domains (Figure 45). This nucleotide substitution has been

found in some individuals, who exhibited either atypical or typical CS with various combinations of clinical anomalies (Jongmans *et al*., 2006; Legendre *et al*., 2018).

In our laboratory, the c.5405-17G>A variant has been found in two familial CHARGE (Figure 47). In both families, the variant was previously reported with a maternal inheritance from their mildly affected mother that failed to fulfill Verloes' diagnostic criteria (Legendre *et al*., 2018). However, when we applied an updated Hale's criteria, these mothers were diagnosed to have CS. The mother in the family one had two major criteria (inner ears anomaly and pathogenic CHD7 variant) and two minor criteria (unilateral deafness and limb anomaly). The mother in the family two had two major criteria (bilateral coloboma and pathogenic CHD7 variant) and three minor criteria (developmental delay, unilateral deafness, and kidney anomaly). Her daughter was studied and concluded to have CS based on Hale's criteria. Her sons were completely asymptomatic (according to the French law, *CHD7* analysis was not performed for these asymptomatic minor children). Moreover, a new pregnancy began and prenatal diagnosis showed that the fetus inherited the same mutation.

These familial cases suggested that a splicing defect could occur with various intensities according to the patients, leading to differences in variant protein amount during early development. This may account for the diversity of clinical features in patients presenting this variant.

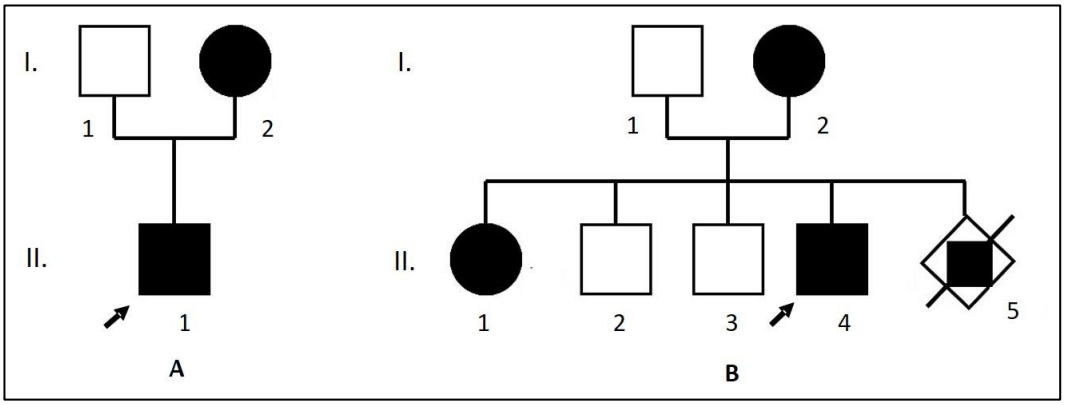

Figure 47. Pedigrees of two familial CHARGE with CHD7:c.5405-17G>A alteration. Family one (A) and family two (B) inherited the variant from their mildly affected mothers (I-2). Both index cases, indicated by a black arrow, had typical CS (II-1, II-4) (Legendre *et al*., 2018).

6.2.2 Localization of Overexpressed Wild-Type and Variant CHD7

Several teams have shown in various cell types that CHD7 protein is localized in the nucleoplasm but is essentially concentrated in the nucleolus (Zentner *et al*., 2010; Kita *et al*., 2012). We have also performed some experiments to point out the localization of CHD7 protein by indirect immunofluorescence. Therefore, we transfected HeLa and HEK293 cells with a plasmid encoding CHD7-HA, then performed an immunostaining of CHD7 and UBF, a transcription factor which is involved in the expression of the rRNAs and whose localization is nucleolar. A merge of the images makes it possible to see the putative co-localization of CHD7 and UBF. These experiments have always revealed a nucleoplasmic but non-nucleolar localization of CHD7-HA (Figure 24). We hypothesized that the HA tag we used could disrupt CHD7 addressing.

Further, we performed the same experiments in HeLa and HEK293 cells using a plasmid encoding FLAG-CHD7 (labeled at the other end of the protein), co-transfected with a fusion protein between the nucleolin (located exclusively in the nucleolus) and mCherry (a red fluorescent protein). However, we obtained the same result as before: the CHD7 protein was localized exclusively in the nucleoplasm (Figure 25).

In brief, we expressed in HeLa and HEK239 cells two forms of CHD7 protein, tagged in N- or C-terminus, and we used two distinct nucleolar markers, but we did not observe any nucleolar localization for CHD7 protein. Nevertheless, it may be more appropriate to carry out this study in cellular models in which CHD7 is endogenously expressed: for example, CHD7 is selectively expressed in neuronal stem cells (NSC) and neuronal progenitor cells (Feng *et al*., 2013). However, the use of primary cultures of this type of cells would make our technique too cumbersome to use since primary mammalian neurons derived from embryonic central nervous system present the limitation that, once terminally differentiated into mature neurons, the cells can no longer be propagated. Thus, it may be more suitable to use neuronal cell lines for further development. That is why we determined to use SH-SY5Y neuroblastoma cells for further experiments. To confirm this finding, we expressed in SH-SY5Y cells two forms of CHD7 protein, tagged in N- or C-terminus, however the localization of CHD7 protein is identical to the previous results (Figure 26).

Moreover, N- or C-terminus fusion tags may interfere with the addressing of CHD7. Therefore, it may be more appropriate to express the CHD7 protein using a plasmid encoding untagged CHD7 and perform immunostaining with an anti-CHD7 antibody. So, to verify these results, we performed immunostaining of endogenous CHD7 protein compared with untagged-CHD7 protein that we co-transfected with a fusion protein between nucleolin and the mCherry in HeLa cells. Contrary to the literature, repeated experiments gave the same results showing that endogenous or overexpressed CHD7 proteins localize exclusively in the nucleoplasm, but not in the nucleolus (Figure 27). We concluded that the tag in either position was not involved in the localization of CHD7 protein. It is still unclear why CHD7, in our hands, was always localized exclusively in the nucleoplasm and not in the nucleolus. In our functional test system, we chose to express untagged CHD7 in HeLa cells to compare the localization of CHD7 protein in WT or variant forms, and next to assess their functionality.

Whatever the structural consequence of these changes in amino acid sequence upon CHD7 folding, we designed a simple functional assay based on CHD7 ability to regulate gene transcription. First, we checked the overexpression of wild-type and variants form of CHD7 by Western blot and immunofluorescence. We observed that all variant forms of CHD7 exhibit the same nucleoplasmic localization compared with the wild-type. This suggests that these changes in CHD7 amino acid sequence have no consequence upon the intracellular trafficking of the protein.

6.2.3 Evaluation of Pathogenicity by Computational Tools for CHD7 Missense Variants

Bioinformatic computational tools have been widely used to assign the functionality of a protein, including CHD7 protein variants (Bergman *et al*., 2012; Yao *et al*., 2018; Gonçalves *et al*., 2019). Moreover, the reported information about CHD7 variations and their pathogenicity have been archived in several databases such as CHD7 database, ClinVar, VarSome, EVS, gnomAD. It allows to easily identify the pathogenicity of a CHD7 variant found in a patient. However, interpreting *de novo* CHD7 variants (missense or other variants in the CHD7 protein amino acid sequence) that have not yet been reported before remains a challenge. Sometimes, *in-silico* CHD7 pathogenicity analysis using several tools even give different results. To resolve this problem, Bergman *et al*. had proposed an algorithm that combines two computational tools, structural model and phenotypic data (Bergman *et al*., 2012). Furthermore, to confirm the bioinformatic prediction of CHD7 missense variants that we studied, we applied this algorithm to evaluate their pathogenicity using published data (Table 16). Unfortunately, we had to classify the G1982W variant as a Variant remain of Unknown clinical Significance (VUS) due to a lack of segregation data and structural model. However, this classification is coherent with what was previously described in chapter 3. Therefore, for this VUS variant, a biological functional test is highly required and must be conducted to confirm its pathogenicity.

Table 16. Pathogenicity of five studied missense variants according to Bergman *et al*.

CHD7 Variant	Poly Phen-2	Align-GVGD	Structural Model	Segregation Analysis	Total Score; Classification	Ref.
c.5944G>T; G1982W	Prob/1	C65	Undeterm	N/A	+2; VUS	
c.6955C>T; R2319C	Prob/1	C65	Detrimental	>1 *de novo*	+7; Pathogenic	1, 2, 3
c.3905T>C; L1302P	Prob/1	C65	Detrimental	*de novo*	+6; Pathogenic	3, 4
c.2680A>G; T894A	Benign/ 0.126	C55	Minor	parent carrier	-2; Benign	3, 5
c.6478G>A; A2160T	Benign/ 0.002	C0	Minor	parent carrier, homozygous	-8; Benign	3,6

PolyPhen-2 output: benign; prob, probably damaging. Align-GVGD output: C0, benign; C55/C65, probably pathogenic. Structural model: minor, minor effect; detrimental, detrimental effect; undeterm, undetermined.
1 Félix *et al*., 2006; 2 Jongmans *et al ;* 3 Bergman *et al*., 2012 ; 4 Wincent *et al*., 2008; 5 Bertels *et al*., 2010 ; 6 Vuorela *et al*., 2007.

6.2.4 Functionality of CHD7 Variants

In this work, we describe a straightforward approach to evaluate the impact of amino acid sequence variations upon the functionality of CHD7 protein. We applied this method to six variants that found in CS patients: five variants harboring an amino acid substitution, T894A, A2160T, G1982W, R2319C and L1302P, and one variant with a five-amino acid insertion, 1801insDGHGT.

To compare the functionality of the variants with that of WT-CHD7, we quantified the consequence of CHD7 overexpression in HeLa cells upon the transcript amounts of genes whose transcription was published to be

regulated by CHD7. When we transiently expressed WT-CHD7 in HeLa cells, the transcription of the five reporter genes was downregulated: the mRNA amount levels were decreased to 23.2-48.9% of the levels quantified in mock-transfected cells. When we expressed T894A and A2160T variants, the transcript amounts of the five reporter genes were decreased and no statistically significant difference was observed compared with WT-CHD7 (Figure 32). These results confirmed that both variants are polymorphism. In contrast, when we expressed any of the four *CHD7* variants: G1982W, R2319C, L1302P, and 1801insDGHGT, the transcription levels of the five reporter genes were non-significantly different from the control levels. This suggests that these CHD7 variants have lost their functionality of chromatin remodeler and gene transcription regulator. Hence, these amino acid changes are not polymorphisms but are likely loss-of-function variants, leading to haploinsufficiency in patients.

The replicates of each sample (WT and variants) were included within each independent experiment, and each independent transfection was repeated three times. The sensitivity and specificity of this functional assay system have been evaluated using a set of 3 pathogenic and 2 non-pathogenic variants. Sensitivity has been estimated at 100% and specificity at 100%. These results suggest that this functional assay system is reliable to predict the pathogenicity of CHD7 variants. However, our study had certain limitations due to the low number of studied variants. Therefore, evaluation and validation of our functional assay system should be

improved by performing this assay using a larger set of either various pathogenic and non-pathogenic, as well as VUS variants that are reported in the literature or in the CHD7 database.

6.2.5 Technical Consideration for Functional Assay using Overexpression Approach

In overexpression approach, the modulation of transcription elicited by *CHD7* appeared sometimes contradictory with previously published data obtained in various *ex vivo* or *in vivo* models, in which *CHD7* was either overexpressed or depleted. However, in a transformed cell line like HeLa cells, the chromatin packing status and epigenetic tagging may differ largely compared to these models.

What are the expected clinical consequences of studying the functionality of CHD7 missense or other variants using our system? This system could determine whether a substitution is a polymorphism or a disease-causing alteration. Moreover, if we detect functional alterations of different magnitudes between several variants, we will have to consider correlations with the different phenotypes of patients with CS or other diseases involving the *CHD7* gene.

The functional analysis may also be useful for genetic counseling in some cases. Indeed, even if the majority of *CHD7* mutations appear *de novo*, there are rare cases of family mutations. This is the case of the nucleotide substitution giving rise to the variant G1982W. This is why, especially in this case, genetic counseling is really necessary to explain the

probabilities of transmitting the CS to the offspring or to inform the parents of the possibility to perform a prenatal diagnosis.

The G1982W and L1302P variants were found in typical CS patients, while the R2319C variant was identified in a partial CHARGE case. It would therefore be consistent to observe different functional consequences of studied variants expression by our system. Therefore, the use of *in vitro* assays that evaluate the consequence of CHD7 variants encountered in CS patients on the structure and function of the protein is an essential alternative approach to predict the pathogenicity of CHD7 variants. The results of functional assay in G1982W and L1302P variants (typical CS) and R2319C variant (partial CS) are similar in our system by using overexpression approach. As a perspective, to validate these results, the functional assay in genomically modified cell line should be realized and compared with the previous results.

6.3 Reclassification of CHD7 Variants

Previous studies have shown that the pathogenicity of variants may be reclassified when additional information becomes available (Mersch *et al.*, 2018). According to the functional tests performed using our system, we therefore recommend updating the classification of G1982W-, R2319C-, L1302P-, and 180insDGHGT-*CHD7* variants according to the International Agency for Research on Cancer (IARC) and American College of Medical Genetics (ACMG) classifications. Based on IARC classification, for R2319C, L1302P, and 1801insDGHGT variants, our data confirm the *in-*

silico predictions that they should be classified as class 5 (disease-causing) variations. As for G1982W, we propose that this variation should be upgraded from class 3 (variant of uncertain significance) to class 4 (probably affecting function).

Furthermore, according to ACMG classification, R2319C is classified as pathogenic: PS1 (same amino acid change as a previously established pathogenic variant), PP3 (multiple lines of computational evidence support a deleterious effect on the gene). L1302P and 1801insDGHGT variants are also classified as pathogenic: PS1, PP3, and PM1 (located in a mutational hot spot and/or critical and well-established functional domain). G1982W variant can be reclassified as likely-pathogenic: PS1, PM1, PP4 (patient's phenotype or family history is highly specific for a disease with a single genetic etiology).

It is important to conduct a functional test and to communicate its results, mainly for class 3 (uncertain significance) variants. It will allow discussing the collection of additional information and material with the patient and their family that can eventually lead to a better assessment of the variants (for example, in our case of G1982W variant with a familial mutation). A consequence of not communicating this variant is that potentially pathogenic variant would go unrecognized and remain classified as class 3.

6.4 Functional Assay of Endogenously Expressed CHD7 Missense Variants Using CRISPR/Cas9 System

Having successfully developed a biological functional test using overexpression approach, we further developed our system using CRISPR/Cas9 technology to obtain genomically expressed CHD7 variants and study their functionality in a more physiological context. We successfully applied the CRISPR/Cas9 genome editing system to target genome location within the cell models and to generate *CHD7* knock-out cells. Therefore, we showed in this study that the customized sgRNA/Cas9 system could be used to induce *CHD7* gene mutations in cell models. The results shown an achievement of sgRNA/Cas9 mediated mutagenesis with an efficiency of 51.6% for a knock-out in HeLa cells.

HeLa cells are frequently used in cell manipulation since there are many advantages, such as a rapid growing, easy transfection, easiness and speed in generating mutant lines. However, HeLa cells cannot be used as a physiologic model of differentiated cells. On the other hand, CHD7 is highly expressed in neuronal progenitor cells such as SH-SY5Y cells (Feng *et al.*, 2013). By using CRISPR/Cas9, we generated *CHD7* (-/-) cells in both HeLa and SH-SY5Y lines. We then evaluated the *CHD7* mRNA expression level by qRT-PCR using verified clones. In both cell models, the *CHD7* mRNA expression level was reduced compared with the parental cell line confirming the knocking-out of the gene. We further measured the transcript amounts of the five reporter genes in both *CHD7* (-/-) cell models. In the

HeLa cell line, the endogenous expression of CHD7 is not involved in regulating the transcription of these reporter genes. However, in SH-SY5Y cells, the amount of RNA transcribed from these genes were upregulated when *CHD7* was knocked-out. From this result, we assumed that CHD7 plays a role as a regulator of transcription in SH-SY5Y cells, but not in the HeLa cells. This result is consistent with previous description that CHD7 is highly expressed in neuronal cells and that the selective expression of CHD7 may be involved in the development of the neurons (Feng et *al*., 2017; Micucci et *al*., 2014). Therefore, to introduce the *CHD7* missense variant into the genome, we determined to use SH-SY5Y neuroblastoma cells.

In the first experiment, we used ssODNs as an artificial repair template for HDR repair, which was homologous to the target sequence and containing base-pair substitutions of interest (to encode G1982W, R2319C, or L1302P variants). Several researchers have described their preference to use ssODNs for genomic modification due to its simplicity, for being less time- and less cost-consuming than using a double-stranded template (Leonetti *et al*., 2016; Mikuni *et al*., 2016). Recently, several methods such as ‘easi-CRISPR’ have been developed. This method permits to generate long HDR template (more than 1 kb of sequence) and enables to insert longer sequences (e.g. reporters or gene tags) (Miura *et al*., 2018). Moreover, using the ssODNs for HDR repair could prevent random integration of the exogenous DNA products into the organism's genome

(Würtele *et al*., 2003; Zorin *et al*., 2005; Won and Dawid, 2017). Briefly, single-stranded template has shown higher efficiency than dsDNA template (Beumer *et al*., 2013; Miura *et al*., 2015; Yoshimi *et al*., 2016).

However, in our experiments, ssODNs did not work as expected. It might be because its efficiency can vary widely depending on the cell type and status, as well as on the genomic locus and repair template, even if we targeted three different loci. Previous studies have discovered the impact of adaptations to the length, symmetry and strand complementarity of the ssODN repair template on genome-editing efficiency, although the consensus concerning the impact of these different adaptations is lacking (Richardson *et al*., 2016; Yang *et al*., 2013). Alternatively, by using linearized double-stranded targeting plasmid as HDR template, we obtained a clonal cell line containing the L1302P variant. Unfortunately, the two other missense variants (G1982W and R2319C) have not yet been obtained, although repeated transfections to generate these variants had been performed.

For further experiments, the CRISPR/Cas9 plasmid transfection efficiency should be improved. CRISPR/Cas9 plasmid containing green fluorescence protein (GFP) could be used to detect the transfected cells and to make a selection by cytometry before cloning. Moreover, some modifications should be considered, such as adding chemicals (SCR7, NU7441 and KU0060648) that could block the NHEJ pathway (Ma *et al*., 2016; Maruyama *et al*., 2015; Singh *et al*., 2015; Srivastava *et al*.,

2012) or stimulate the HDR pathway (Jayathilaka *et al*., 2008; Pinder *et al*., 2015; Song *et al*., 2016; Yu *et al*., 2015) to improve the success rate of HDR.

A study has shown that various cells have different abilities to repair DSBs using either NHEJ or HDR. The phase of the cell cycle is involved in the choice of pathway: NHEJ dominates DNA repair during G1, S and G2 phases, while HDR occurs only in the late S and G2 phases (DNA replication is completed, and sister chromatids are available to be used as repair templates) (Heyer *et al*., 2010). Therefore, a strategy combining well-established synchronization methods with direct nucleofection of pre-assembled Cas9 ribonucleoprotein (RNP) complexes could be performed to increase the efficiency of HDR (Lin *et al*., 2015).

In *CHD7* knock-out and in the L1302P variant SH-SY5Y cells, the transcription amount of the five reporter genes was increased to 253-544% of the levels quantified in parental SH-SY5Y cells. The modulation of transcription elicited by endogenous *CHD7* is different with the one that we observed previously using overexpression approach. This condition can be explained as follows. The cell lines that we used in each approach are different. In overexpression approach, we transiently transfected the plasmids encoding WT-CHD7 or variants into HeLa cells. On the other hand, by using CRISPR/Cas9, SH-SY5Y cells were genomically modified to obtain desired missense variants. The latter approach allows to obtain *CHD7* missense variant at the genome level. As we mentioned previously,

CHD7 is highly expressed in neuronal cell lines; for this reason, endogenous level expression of WT or variant CHD7 in SH-SY5Y cells may have different consequence compared with overexpression approach in HeLa cells. However, the use of both approaches in different cell lines revealed similar results: for L1302P variant, the transcription levels of the five reporter genes were non-significantly different of the control (mock) or *CHD7* knock-out levels. This confirms that L1302P-*CHD7* variant is not polymorphism but is likely a loss-of-function variant, leading to haploinsufficiency in patients.

The genomic modification approach may allow to observe the functionality of CHD7 when heterozygous substitution occurred in CS patients. For that purpose, it would be interesting to obtain cell lines containing heterozygous mutations, to compare with homozygous one. However, this approach has certain limitations. It will be difficult to obtain the heterozygous genomically modified cell to study the functionality in the same conditions as in patients. In 2016, a method for introducing mutation at single alleles to obtain a model of genetic disease caused by heterozygous mutations had been described, however, the complexity of this method will require months of work to complete (Paquet *et al.*, 2016).

In our experiments, after the SH-SY5Y cells were treated by CRISPR/Cas9 complex and incubated with puromycin, the cells require more time than usual to grow. Long waiting time in this approach could make difficult to use it as a routine assay. Therefore, we recommend using

the overexpression approach to investigate the functionality of CHD7 variants.

It has been described in previous studies that the expression of CHD7 is selective in different types of mature neuronal cells. For example, CHD7 is tuned off in most of the mature neurons in the brain during neurogenesis. In contrast, CHD7 is highly expressed in some interneuron cells in the olfactory bulb and in cerebellar granule neurons of adult mouse and human brain (Feng *et al*., 2017; Micucci *et al*., 2014). A remaining question concerning the function of CHD7 is whether CHD7 is required for the differentiation of neuronal cells. Therefore, we propose to use SH-SY5Y neuroblastoma cell line that we generated to investigate the role of CHD7 in the differentiation of neurons. This future study is expected to advance our understanding of neuronal behavior abnormality that is frequently found in CS patients.

CHAPTER VII

CONCLUSIONS AND FUTURE PERSPECTIVES

7.1 Conclusions

This thesis favors a strategy for molecular diagnosis of CS using next-generation sequencing (NGS) of targeted gene panel and provided new insights on developing a functional test to better interpret some *CHD7* variants.

7.1.1 Novel CHD7 Mutation in an Indonesian CHARGE Syndrome Patient

We report a novel heterozygous nonsense mutation of the *CHD7* gene c.7234G>T or p.(Glu2412Ter) in an Indonesian CS patient. Clinical stigmata coupled with molecular analysis is important to determine the disease-causing gene and establish a definitive diagnosis of CS.

7.1.2 Functional Assay of CHD7 Protein

We have developed a novel and simple assay to assess accurately the functional impact of amino acid sequence CHD7 variants using overexpression approach. These results suggest that the four *CHD7* studied variants (G1982W, R2319C, L1302P, 1801insDGHGT) are not polymorphisms but that the variant proteins are not functional. In conclusion, these variants can be considered as disease-causing of CS.

We successfully applied the CRISPR/Cas9 genome editing system to invalidate the *CHD7* gene in HeLa and SH-SY5Y cells. This system was also used to introduce one (L1302P-*CHD7*) out of three missense variants in the SH-SY5Y cells. We then investigated the functionality of this variant

using our system. This result also confirms that L1302P variant is pathogenic.

7.2 Future Perspectives

7.2.1 Molecular Diagnosis of CHARGE Syndrome

Mutation in the *CHD7* gene is the major cause of CS. However, after molecular analysis using Sanger sequencing, no *CHD7* mutation is identified in 5-10% of typical CS patients (Bergman *et al*., 2008). Identification of other candidate genes has evolved rapidly. So far, only two other genes have been proven to be implicated in CS, the *EFTUD2* and *SEMA3E* genes (Lalani et *al*., 2004; Legendre *et al*., 2017). Recently, other candidate genes *RERE* and *KMT2D* have been reported causing CHARGE-like attributes (Jordan *et al*., 2018; Badalato *et al*., 2017). Therefore, to develop and to improve the targeted gene NGS panel, these genes should be considered to be included in these panels.

7.2.2 Development of Functional Assay for CHD7 Protein

Currently, 93 unclassified variants have been reported in the CHD7 database. These Variants remain of Unknown clinical Significance (VUS) because of the lack of genetic information that may help to establish the pathogenicity. Therefore, it would be helpful if we could extend our biological assay system to other CHD7 missense variants that are currently classified as VUS.

The functional analysis may also identify variants with an intermediate function that may have intermediate or moderate effects on pathogenicity

(Lovelock *et al*., 2013). In this work, the studied variants were determined easily as functional or non-functional since the transcript amount of each reporter gene was statistically not different from the wild-type or control (mock). The next challenge that should be addressed is how to determine accurately the pathogenicity of variants in which the transcript amount of the reporter genes is between wild-type and control (mock). It would be possible to calculate the cut-off point of the expression level by conducting several experiments with a larger variant sample.

Several experiments using Cas9/sgRNA plasmids and HDR that did not result in desired modification should be repeated. Some techniques should be modified and improved, including transfection efficiency. Another transfection technique such as nucleofection using the Nucleofector II (Amaxa Biosystem) should be applied. Furthermore, if necessary, sgRNA and ssODNs should be redesigned to ensure the generation of the G1982W and R2319C variants. Moreover, using CRISPR/Cas9 technology, it will be interesting to study the functionality of heterozygous *CHD7* variants by introducing mutation at single alleles to model CS caused by heterozygous alterations (Paquet *et al*., 2016). We believe that the CRISPR/Cas9 derived genome-editing technique will enable the performance of functional studies and help us to understand the role of *CHD7* gene and mechanisms underlying their physiological and pathophysiological effects.

REFERENCES

Admiraal RJ, Joosten FB, Huygen PL. Temporal bone CT findings in the CHARGE association. *Int J Pediatr Otorhinolaryngol*. 1998;45:151-162.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. RNA synthesis and RNA processing in molecular biology of the cell. 4th edition. New York: Garland Science; 2002.

Amiel J, Attiee-Bitach T, Marianowski R, Cormier-Daire V, Abadie V, Bonnet D, et al. Temporal bone anomaly proposed as a major criteria for diagnosis of CHARGE syndrome. *Am J Med Genet*. 2001;99:124-127.

Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014;513:569-573.

Antonarakis SE, Krawczak M, Cooper DN. Disease-causing mutations in the human genome. *Eur J Pediatr*. 2000;159 Suppl 3:S173-S178.

Aramaki M, Udaka T, Kosaki R, Makita Y, Okamoto N, Yoshihashi H, et al. Phenotypic spectrum of CHARGE syndrome with CHD7 mutations. *J Pediatr*. 2006;148:410-414.

Aramaki M, Kimura T, Udaka T, Kosaki R, Mitsuhashi T, Okada Y, et al. Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome. *Birth Defects Res A Clin Mol Teratol*. 2007;7950-57.

Aref-Eshghi E, Rodenhiser DI, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. Genomic DNA methylation signatures enable concurrent diagnosis and clinical genetic variant classification in neurodevelopmental syndromes. *Am J Hum Genet*. 2018;102:156–174.

Asakura Y, Toyota Y, Muroya K, Kurosawa K, Fujita K, Aida N, et al. Endocrine and radiological studies in patients with molecularly confirmed CHARGE syndrome. *J Clin Endocrinol Metab*. 2008;93:920-924.

Badalato L, Farhan SM, Dilliott AA, Care4Rare Canada Consortium, Bulman DE, Hegele RA, et al. KMT2D p.Gln3575His segregating in a family with autosomal dominant choanal atresia strengthens the Kabuki/CHARGE connection. *Am J Med Genet A*. 2017;173:183-189.

Bajpai R, Chen DA, Rada-Iglesias A, Zhang J, Xiong Y, Helms J, et al. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*. 2010;463:958-962.

Balasubramanian R, Choi JH, Francescatto L, Willer J, Horton ER, Asimacopoulos EP, et al. Functionally compromised CHD7 alleles in patients with isolated GnRH deficiency. *Proc Natl Acad Sci USA*. 2014;111:17953-17958.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007;345:1709-1712.

Bartels CF, Scacheri C, White L, Scacheri PC, Bale S. Mutations in the CHD7 gene: The experience of a commercial laboratory. *Genet Test Mol Biomarkers*. 2010;14:881-891.

Basson MA, van Ravenswaaij-Arts C. Functional insights into chromatin remodeling from studies on CHARGE syndrome. *Trends Genet*. 2015;31:600-611.

Batsukh T, Pieper L, Koszucka AM, von Velsen N, Hoyer-Fender S, Elbracht M, et al. CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Hum Mol Genet*. 2010;19:2858-2866.

Bergman JE, Bosman EA, van Ravenswaaij CM, Steel KP. Study of smell and reproductive organs in a mouse model for CHARGE syndrome. Eur. *J Hum Genet*. 2010;18:171-177.

Bergman JE, de Wijs I, Jongmans MC, Admiraal RJ, Hoefsloot LH, Ravenswaaij-Arts CM. Exon copy number alterations of the CHD7 gene are not a major cause of CHARGE and CHARGE-like syndrome. *Eur J Med Genet*. 2008;51:417-425.

Bergman JE, Janssen N, Hoefsloot LH, Jongmans MCJ, Hofstra RMW, van Ravenswaaij-Arts CMA. CHD7 mutations and CHARGE syndrome: the

clinical implications of an expanding phenotype. *J Med Genet*. 2011;48:334–342.

Bergman JE, Janssen N, van der Sloot AM, de Walle HE, Schoots J, Rendtorff ND, et al. A novel classification system to predict the pathogenic effects of CHD7 missense variants in CHARGE syndrome. *Hum Mutat*. 2012;33:1251-1260.

Bernstein V, Denno LS. Repetitive behaviors in CHARGE syndrome: differential diagnosis and treatment options. *Am J Med Genet A*. 2005;133A:232-239.

Beumer KJ, Trautman JK, Mukherjee K, Carroll D. Donor DNA utilization during gene targeting with zinc-finger nucleases. *G3*. 2013;3:657-664.

Bhaya D., Davison M., Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet*. 2011;45:273-297.

Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe*. 2012;12:177-186.

Bilan F, Legendre M, Charraud V, Manière B, Couet D, Gilbert-Dussardier B, et al. Complete screening of 50 patients with CHARGE syndrome for anomalies in the CHD7 gene using a denaturing high-performance liquid chromatography-based protocol: new guidelines and a proposal for routine diagnosis. *J Mol Diagn*. 2012;14:46-55.

Blake KD, Davenport SL, Hall BD, Hefner MA, Pagon RA, Williams MS, et al. CHARGE association: an update and review for the primary pediatrician. *Clin Pediatr (Phila)*. 1998;37:159-173.

Blake KD, Prasad C. CHARGE syndrome. *Orphanet J Rare Dis*. 2006;1:34.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009;326:1509-1512.

Bosman EA, Penn AC, Ambrose JC, Kettleborough R, Stemple DL, Steel KP. Multiple mutations in mouse Chd7 provide models for CHARGE syndrome. *Hum Mol Genet*. 2005;14:3463-3476.

Bouazoune K, Kingston RE. Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human development disorders. *Proc Natl Acad Sci USA*. 2012;47:19238-19243.

Butcher DT, Cytrynbaum C, Turinsky AL, Siu MT, Inbar-Feigenberg M, Mendoza-Londono R, et al. CHARGE and Kabuki syndromes: gene-specific DNA methylation signatures identify epigenetic mechanisms linking these clinically overlapping conditions. *Am J Hum Genet*. 2017;100:773-788.

Cady KC, Bondy-Denomy J, Heussler GE, Davidson AR, O'Toole GA. The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *J Bacteriol*. 2012;194:5728-5738.

CHD7 database. NL Radboudumc, Groningen. 2018. https://molgenis51.gcc.rug.nl/. Accessed 15 March 2018.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*. 2010;186:757-761.

Clarke R, Heler R, MacDougall MS, Yeo NC, Chavez A, Regan M, et al. Enhanced bacterial immunity and mammalian genome editing via RNA-polymerase-mediated dislodging of Cas9 from double-strand DNA breaks. Mol Cell. 2018;71:42-55.e8

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819-823.

Corsten-Janssen N, Kerstjens-Frederikse WS, du Marchie Sarvaas GJ, Baardman ME, Bakker MK, Bergman JE, et al. The cardiac phenotype in patients with a CHD7 mutation. *Circ Cardiovasc Genet*. 2013;6:248-254.

Corsten-Janssen N, Saitta SC, Hoefsloot LH, McDonald-McGinn DM, Driscoll DA, Derks R, et al. More clinical overlap between 22q11.2 deletion syndrome and CHARGE syndrome than often anticipated. *Mol Syndromol*. 2013;4:235-245.

Corsten-Janssen N, Scambler PJ. Clinical and molecular effects of CHD7 in the heart. *Am J Med Genet*. 2017;175C:487-495.

Cui Y, Xu J, Cheng M, Liao X, Peng S. Review of CRISPR/Cas9 sgRNA design tools. *Interdiscip Sci*. 2018;10:455-465.

Dauber A, Hirschhorn JN, Picker J, Maher TA, Milunsky A. Delayed puberty due to a novel mutation in CHD7 causing CHARGE syndrome. *Pediatrics*. 2010;126:e1594-e1598.

Delehaye A, Sznajer Y, Lyonnet S, Elmaleh-Bergès M, Delpieree I, Audollent S, et al. Familial CHARGE syndrome because of CHD7 mutation: clinical intra and interfamilial variability. *Clin Genet*. 2007;72:112-121.

Dev VG, Butler MG, Phillips JA. 1q duplication due to inequal crossover in a patient with CHARGE association and di george sequence. *Am J Hum Genet*. 1985;37:A90.

Devriendt K, Swillen A, Fryns JP. Deletion in chromosome region 22q11 in a child with CHARGE association. *Clin Genet*. 1998;53:408-410.

Dirscherl SS, Krebs JE. Functional diversity of ISWI complexes. *Biochem Cell Biol*. 2004;82:482-489.

Dobbelsteyn C, Marche DM, Blake K, Rashid M. Early oral sensory experiences and feeding development in children with CHARGE syndrome: a report of five cases. *Dysphagia*. 2005;20:89-100.

Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014; 346:1258096.

Emanuel BS, Budarf ML, Sellinger B, Goldmuntz E, Driscoll DA. Detection of microdeletions of 22q11.2 with fluorescence in situ hybrydization (FISH): Diagnosis of DiGeorge (DGS), velo-cardial-facial syndrome, CHARGE association and conotruncal cardiac malformations. *Am J Hum Genet*. 1992;51:A80.

Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet*. 2011;43:607-611.

Félix T, Hanshaw B, Mueller R, Bitoun P, Murray J. CHD7 gene and nonsyndromic cleft lip and palate. *Am J Med Genet Part A*. 2006;140:2110-2114.

Feng W, Khan MA, Bellvis P, Zhu Z, Bernhardt O, Herold-Mende C, et al. The chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription factors. *Cell Stem Cell*. 2013;13:62-72.

Frank M, Prenzler A, Eils R, Graf von der Schulenburg JM. Genome sequencing: a systematic review of health economic evidence. *Health Econ Rev*. 2013;3:29.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010;468:67-71.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA*. 2012;109:E2579–2586.

Gennery AR, Slatter MA, Rice J, Hoefsloot LH, Barge D, McLean-Tooke A, et al. Mutations in CHD7 in patients with CHARGE syndrome cause T-B + natural killer cell + severe combined immune deficiency and may cause omenn-like syndrome. *Clin Exp Immunol*. 2008;153:75-80.

Goldmann JM, Wong WS, Pinelli M, Farrah T, Bodian D, Stittrich AB, et al. Parent of origin specific signatures of de novo mutations. *Nat Genet*. 2016;48:935-939.

Gonçalves CI, Patriarca FM, Aragüés JM, Carvalho D, Fonseca F, Martins S, et al. High frequency of CHD7 mutations in congenital hypogonadotropic hypogonadism. *Sci Rep*. 2019;9:1597.

Grozeva D, Carss K, Spasic-Boskovic O, Tejada MI, Gecz J, Shaw M, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Hum Mutat*. 2015;36:1197-1204.

Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol*. 2010;649:247-256.

Hale CL, Niederriter AN, Green GE, Martin DM. Atypical phenotypes associated with pathogenic CHD7 variants and a proposal for broadening CHARGE syndrome clinical diagnostic criteria. *Am J Med Genet Part A*. 2016;170A:344-354.

Hall BD. Choanal atresia and associated multiple anomalies. *J Pediatr*. 1979;95:395-398.

Hall JA, Georgel PT. CHD proteins: a diverse family with strong ties. *Biochem Cell Biol*. 2007;85:463-476.

Hartshorne TS. Australasian CHARGE Conference, Perth, WA: CHARGE syndrome Association of Australasia, 2012.

Hartshorne TS, Hefner MA, Davenport SL. Behavior in CHARGE syndrome: introduction to the special topic. *Am J Med Genet A*. 2005;133A:228-231.

Hartshorne TS, Cypher AD. Challenging behavior in CHARGE syndrome. *Mental Health Aspects Dev Dis*. 2004;7:41-52.

He D, Marie C, Zhao C, Kim B, Wang J, Deng Y, et al. Chd7 cooperates with Sox10 and regulates the onset of CNS myelination and remyelination. *Nat Neurosci*. 2016;19:678-689.

Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. *Ann Rev Genet*. 2010;44:113-139.

Hittner HM, Hirsch NJ, Kreh GM, Rudolph AJ. Colobomatous microphthalmia, heart dis-ease, hearing loss, and mental retardation--a syndrome. *J Pediatr Ophthalmol Strabismus*. 1979;16:122-128.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol*. 2011;29:731-734.

Holcomb MA, Rumboldt Z, White DR. Cochlear nerve deficiency in children with CHARGE syndrome. *Laryngoscope*. 2013;123:793-796.

Hopsu E, Markkola A, Pitkaranta A. Labyrinthine malformation in the 22q11.2 deletion syndrome. *Clin Dysmorphol*. 2007;16:67-68.

Houben CH, Curry JI. Current status of prenatal diagnosis, operative management and outcome of esophageal atresia/tracheo-esophageal fistula. *Prenat Diagn*. 2008;28:667-675.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827-832.

Hurd EA, Capers PL, Blauwkamp MN, Adams ME, Raphael Y, Poucher HK, et al. Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defect in multiple developing tissues. *Mamm Genome*. 2007;18:94-104.

Hurd EA, Poucher HK, Cheng K, Raphael Y, Martin DM. The ATP-dependent chromatin remodeling enzyme CHD7 regulates pro-neural gene expression and neuro-genesis in the inner ear. *Development*. 2010;137:3139-3150.

Husu E, Hove HD, Farholt S, Bille M, Tranebjærg L, Vogel I, et al. Phenotype in 18 Danish subjects with genetically verified CHARGE syndrome. *Clin Genet*. 2013;83:125-134.

Inoue H, Takada H, Kusuda T, Goto T, Ochiai M, Kinjo T, et al. Successful cord blood transplantation for a CHARGE syndrome with CHD7 mutation showing DiGeorge sequence including hypoparathyroidism. *Eur J Pediatr*. 2010;169:839-844.

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol*. 1987;169:5429-5433.

Issekutz KA, Graham JM, Prasad C, Smith IM, Blake KD. An epidemiological analysis of CHARGE syndrome: preliminary results from a Canadian study. *Am J Med Genet*. 2005;133A:309-317.

Jacobi AM, Rettig GR, Turk R, Collingwood MA, Zeiner SA, Quadros RM, et al. Simplified CRISPR tools for efficient genome editing and streamLined protocols for their delivery into mammalian cells and mouse zygotes. *Methods*. 2017;121-122:16-28.

James PA, Aftimos S, Hofman P. CHARGE association and secondary hypoadrenalism. *Am J Med Genet A*. 2003;117A:177-180.

Janssen N, Bergman JE, Swertz MA, Tranebjaerg L, Lodahl M, Schoots J, et al. Mutation update on the CHD7 gene involved in CHARGE

syndrome. *Hum Mutat*. 2012;33:1149-1160.

Jansen R, Embden JDA van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002;43:1565-1575.

Jao LE, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci USA*. 2013;110:13904-13909.

Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol*. 2013;5:a012740.

Jayathilaka K, Sheridan SD, Bold TD, Bochenska K, Logan HL, Weichselbaum RR, Bishop DK, et al. A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc Natl Acad Sci USA*. 2008;105:15848-15853.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31:233-239.

Jiang X, Zhou Y, Xian L, Chen W, Wu H, Gao X. The mutation in Chd7 causes misexpression of Bmp4 and developmental defects in telencephalic midline. *Am J Pathol*. 2012;181:626-641.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816-821.

Jordan VK, Fregeau B, Ge X, Giordano J, Wapner RJ, Balci TB, et al. Genotype-phenotype correlations in individuals with pathogenic RERE variants. *Hum Mutat*. 2018;39:666-675.

Jongmans MC, Admiraal R, van der Donk K, Vissers L, Baas A, Kapusta L, et al. CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. *J Med Genet*. 2006;43:306-314.

Jongmans MC, Hoefsloot LH, van der Donk KP, Admiraal RJ, Magee A, van de Laar I, et al. Familial CHARGE syndrome and the CHD7 gene: a recurrent missense mutation, intrafamilial recurrence and variability. *Am J Med Genet A*. 2008;146A:43-50.

Jongmans MC, van Ravenswaaij-Arts CM, Pitteloud N, Ogata T, Sato N, Claahsen-van der Grinten HL, et al. CHD7 mutations in patients initially diagnosed with Kallmann syndrome-the clinical overlap with CHARGE syndrome. *Clin Genet*. 2009;75:65-71.

Jyonouchi S, McDonald-McGinn DM, Bale S, Zackai EH, Sullivan KE. CHARGE (coloboma, heart defect, atresia choanae, retarded growth and development, genital hypoplasia, ear anomalies/deafness) syndrome and chromosome 22q11.2 deletion syndrome: a comparison of immunologic and nonimmunologic phenotypic features. *Pediatrics*. 2009;123:e871-877.

Kahmoto T, Shono M, Naruto T, Watanabe M, Suga K, Nakagawa R, et al. A novel frameshift mutation of CHD7 in a Japanese patient with CHARGE syndrome*. Hum Genome Var*. 2016;3:16004.

Kallen K, Robert E, Mastroiacovo P, Castilla EE, Kallen B. CHARGE Association in newborns: a registry-based study. *Teratology*. 1999;60:334-343.

Karvelis T, Gasiunas G, Siksnys V. Programmable DNA cleavage in vitro by Cas9. *Biochem Soc Trans*. 2013;41:1401-1406.

Khadilkar VV, Cameron FJ, Stanhope R. Growth failure and pituitary function in CHARGE and VATER associations. *Arch Dis Child*. 1999;80: 167-170.

Kita Y, Nishiyama M, Nakayama KI. Identification of $CHD7_S$ as a novel splicing variant of CHD7 with functions similar and antagonistic to those of the full-length $CHD7_L$. *Genes Cells*. 2012;17:536-547.

Lalani SR, Safiullah AM, Molinari LM, Fernbach SD, Martin DM, Belmont JW. SEMA3E mutation in a patient with CHARGE syndrome. *J Med Genet*. 2004;41:e94.

Lalani SR, Safiullah AM, Fernbach SD, Harutyunyan KG, Thailer C, Peterson LE, et al. Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am J Hum Genet*. 2006;78:303-314.

Lalani SR, Hefner MA, Belmont JW, Davenport SLH. CHARGE syndrome. http://www.ncbi. nlm.nih.gov/books/NBK1117/. Updated 2012.

Layman WS, Hurd EA, Martin DM. Chromodomain proteins in development: Lessons from CHARGE syndrome. *Clin Genet*. 2010;78:11-20

Layman WS, Hurd EA, Martin DM. Reproductive dysfunction and decreased GnRH neurogenesis in a mouse model of CHARGE syndrome. *Hum Mol Genet*. 2011;20:3138-3150.

Layman WS, McEwen DP, Beyer LA, Lalani SR, Fernbach SD, Oh E, et al. Defects in neural stem cell proliferation and olfaction in Chd7 deficient mice indicate a mechanism for hyposmia in human CHARGE syndrome. *Hum Mol Genet*. 2009;18:1909-1923.

Legendre M, Abadie V, Attié-Bitach T, Philip N, Busa T, Bonneau D, et al. Phenotype and genotype analysis of a French cohort of 119 patients with CHARGE syndrome. *Am J Med Genet*. 2017;175C:417-430.

Legendre M, Rodriguez-Ballesteros M, Rossi M, Abadie V, Amiel J, Revencu N, et al. CHARGE syndrome: a recurrent hotspot of mutations in CHD7 IVS25 analyzed by bioinformatic tools and minigene assays. *Eur J Hum Genet*. 2018;26:287-292.

Leonetti MD, Sekine S, Kamiyama D, Weissman JS, Huang B. A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc Natl Acad Sci USA*. 2016;113:E3501-3508.

Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*. 2015;520:505-510.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, et al. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat Biotechnol*. 2013;31:688-691.

Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*. 2015;3:e04766.

Liu Y, Harmelink C, Peng Y, Chen Y, Wang Q, Jiao K. CHD7 interacts with BMP R-SMADs to epigenetically regulate cardiogenesis in mice. *Hum Mol Genet*. 2014;23:2145-2156.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real time quantitative PCR and the 2 $^{-\Delta\Delta Ct}$ method. *Methods*. 2001;25:402-408.

Local A, Huang H, Albuquerque CP, Singh N, Lee AY, Wang W, et al. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nature Genet*. 2018;50:73-82.

Lonlay-Debeney P, Cormier-Daire V, Amiel J, Abadie V, Odent S, Paupe A, et al. Features of DiGeorge syndrome and CHARGE association in five patients. *J Med Genet*. 1997;34:986-989.

Lovelock PK, Spurdle AB, Mok MT, Farrugia DJ, Lakhani SR, Healey S, et al. Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants? *Breast Cancer Res*. 2007;9:R82.

Luquetti DV, Hing AV, Rieder MJ, Nickerson DA, Turner EH, Smith J, et al. "Mandibulofacial dysostosis with microcephaly" caused by EFTUD2 mutations: expanding the phenotype. *Am J Med Genet A*. 2012;161A:108-113.

Ma Y, Chen W, Zhang X, Yu L, Dong W, Pan S, et al. Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA Biol.* 2016;13:605-612.

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013;10:977-999.

Makarova KS, Haft DH, Barrangou R, Bruons SJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*. 2011;9:467-477.

Marfella CG, Imbalzano AN. The Chd family of chromatin remodelers. *Mutat Res*. 2007;618:30-40.

Martin DM, Probst FJ, Fox SE, Schimmenti LA, Semina EV, Hefner MA, et al. Exclusion of PITX2 mutations as a major cause of CHARGE association. *Am J Med Genet*. 2002;111:27-30.

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-

Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 2015;33:538-542.

McMain K, Robitaille J, Smith I, Johnson J, Wood E, Tremblay F, et al. Ocular features of CHARGE syndrome. *J AAPOS*. 2008;12:460-465.

Mersch J, Brown N, Pirzadeh-Miller S, Mundt E, Cox HC, Brown K, et al. Prevalence of Variant Reclassification Following Hereditary Cancer Genetic Testing. *JAMA.* 2018;320:1266-1274.

Miao J, Guo D, Zhang J, Huang Q, Qin G, Zhang X, et al. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res*. 2013;23:1233-1236.

Micucci JA, Layman WA, Hurd EA, Sperry ED, Frank SF, Durham MA, et al. CHD7 and retinoic acid signaling cooperate to regulate neural stem cell and inner ear development in mouse models of CHARGE syndrome. *Hum Mol Genet*. 2014;23:434-448.

Micucci JA, Sperry ED, Martin DM. Chromodomain helicase DNA-binding proteins in stem cells and human developmental diseases. *Stem Cells Dev*. 2015;24:917–926.

Mikuni T, Nishiyama J, Sun Y, Kamasawa N, Yasuda R. High-throughput, high-resolution mapping of protein localization in mammalian brain by in vivo genome editing. *Cell*. 2016;165:1803-1817.

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*. 2007;25:778-785.

Miura H, Gurumurthy CB, Sato T, Sato M, Ohtsuka M. CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Sci Rep.* 2015;5:12799.

Miura H, Quadros RM, Gurumurthy CB, Ohtsuka M. Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nat Protoc.* 2018;13:195-215.

Moccia A, Srivastava A, Skidmore JM, Bernat JA, Wheeler M, Chong JX, et al. Genetic analysis of CHARGE syndrome identifies overlapping molecular biology. *Genet Med*. 2018;20:1022-1029.

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*. 2005;60:174-182.

Morgan D, Bailey M, Phelps P, Bellman S, Grace A, Wyse R. Ear-nose-throat abnormalities in the CHARGE association. *Arch Otolaryngol Head Neck Surg*. 1993;119:49-54.

Moscou MJ, Bogdanove AJ. A simple chiver governs DNA recognition by TAL effectors. *Science*. 2009;236:1501.

Numakura C, Kitanaka S, Kato M, Ishikawa S, Hamamoto Y, Katsushima Y, et al. Supernumerary impacted teeth in a patient with SOX2 anophthalmia syndrome. *Am J Med Genet*. 2010;152A:2355-2359.

Nishina S, Kosaki R, Yagihashi T, Azuma N, Okamoto N, Hatsukawa Y, et al. Ophthalmic features of CHARGE syndrome with CHD7 mutations. *Am J Med Genet A*. 2012;158A:514-518.

Ogata T, Fujiwara I, Ogawa E, Sato N, Udaka T, Kosaki K. Kallman syndrome phenotype in a female patient with CHARGE syndrome and CHD7 mutation. *Endocr J*. 2006;53:741-3.

Onwochei BC, Simon JW, Bateman JB, Couture KC, Mir E. Ocular colobomata. *Surv Opthalmol*. 2000;45:175-194.

Pagon RA, Graham JM, Jr., Zonana J, Yong SL. Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *J Pediatr*. 1981;99:223-227.

Paquet D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, et al. Efficient introduction of specific homozygous and heterozygous mutation using CRISPR/Cas9. *Nature*. 2016;533:125-129.

Pinder J, Salsman J, Dellaire G. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* 2015;43:9379-9392.

Pinto G, Abadie V, Mesnage R, Blustajn J, Cabrol S, Amiel J, et al. CHARGE syndrome includes hypogonadotropic hypogonadism and abnormal olfactory bulb development. *J Clin Endocrinol Metabol*. 2005;90:5621-6526.

Porteus MH, Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. *Science*. 2003;300:763.

Pramudita J, Utari A, Winarni T, Faradz S. CHARGE syndrome: an Indonesian case report. *J Biomed Transl Res*. 2017;1:23-25.

Puc J, Rosenfeld MG. SOX2 and CHD7 cooperatively regulate human disease genes. *Nat Genet*. 2011;43:505-506.

Ragan DC, Casale AJ, Rink RC, Cain MP, Weaver DD. Genitourinary anomalies in the CHARGE association. *J Urol*. 1999;161:622-625.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013;154:1380-1389.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281-2308.

Randall V, McCue K, Roberts C, Kyriakopoulou V, Beddow S, Barrett A, et al. Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. *J Clin Invest*. 2009;119:3301-3310.

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*. 2012;30:460-465.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405-424.

Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol*. 2016;34:339-344.

Rodgers K, McVey M. Error-prone repair of DNA double-strand breaks. *J Cell Physiol*. 2016;231:15-24.

Rothlisberger B, Kotzot D. Recurrence risk in de novo structural chromosomal rearrangements. *Am J Med Genet*. 2007;143:1708-14.

Salem-Hartshorne N, Jacob S. Adaptive behavior in children with CHARGE syndrome. *Am J Med Genet A*. 2005;133A:262-267.

Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res*. 2004;32:3683-3688.

Sander JD, Dahlborg EJ, Goodwin MJ, Cade L, Zhang F, Cifuentes D, et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods*. 2010;8:67-69.

Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32:347-355.

Sanjana NE, Cong L, Zhou Y, Cunniff MM, Feng G, Zhang F. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc*. 2012;7:171-192.

Sanka M, Tangsinmankong N, Loscalzo M, Sleasman JW, Dorsey MJ. Complete DiGeorge syndrome associated with CHD7 mutation. *J Allergy Clin Immunol*. 2007;120:952-954.

Sanlaville D, Etchevers H, Gonzales M, Martinovic J, Clément-Ziza M, Delezoide A, et al. Phenotypic spectrum of CHARGE syndrome in fetuses with CHD7 truncating mutations correlates with expression during human developmant. *J Med Genet*. 2006;43:211-7.

Sanlaville D, Verloes A. CHARGE syndrome: An update. *Eur J Hum Genet*. 2007;15:389-399.

Schneiderman H, Balogun S. What's your diagnosis?: simple iris coloboma. *Consultant*. 2000;40:2093-2096.

Schnetz MP, Bartels CF, Shastri K, Balasubramanian D, Zentner GE, Balaji R, et al. Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Res*. 2009;19:590-601.

Schnetz MP, Handoko L, Akhtar-Zaidi B, Bartels CF, Pereira CF, Fisher AG, et al. CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genet*. 2010;6:e1001023.

Schulz Y, Fresse L, Mänz J, Zoll B, Völter C, Brockmann K, et al. CHARGE and Kabuki syndromes: a phenotypic and molecular link. *Hum Mol Genet*. 2014;23:4396-4405.

Schulz Y, Wehner P, Opitz L, Salinas-Riester G, Bongers EM, van Ravenswaaij-Arts CM, et al. CHD7, the gene mutated in CHARGE syndrome, regulates genes involved in neural crest cell guidance. *Hum Genet*. 2014;133:997-1009.

Schwank G, Koo BK, Sasselli V, Dekkers JF, Heo I, Demircan T, et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*. 2013;13:653-658.

Shroff M, Israel J, Rosenthal F. Congenital anomalies associated with partial deletion of the long arm of chromosome 4 46, XY, del (4) (q31). *Am J Hum Genet*. 1981;33:122A.

Singh P, Schimenti JC, Bolcun-Filas E. A mouse geneticist's practical guide to CRISPR applications. *Genetics*. 2015;199:1-15.

Smith IM, Nichols SL, Issekutz K, Blake K. Behavioral profiles and symptoms of autism in CHARGE syndrome: preliminary Canadian epidemiological data. *Am J Med Genet A*. 2005;133A:248-256.

Song J, Yang D, Xu J, Zhu T, Chen YE, Zhang J. RS-1 enhances CRISPR/Cas9-and TALEN-mediated knock-in efficiency. *Nat. Commun.* 2016;7:10548.

Srivastava M, Nambiar M, Sharma S, Karki SS, Goldsmith G, Hegde M, et al. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell*. 2012;151:1474-1487.

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136:665–677.

Tellier AL, Amiel J, Delezoide AL, Audollent S, Auge J, Esnault D, et al. Expression of the PAX2 gene in human embryos and exclusion in the CHARGE syndrome. *Am J Med Genet*. 2000;93:85-88.

Tellier AL, Cormier-Daire V, Abadie V, Amiel J, Sigaudy S, Bonnet D, et al. CHARGE syndrome: report of 47 cases and review. *Am J Med Genet*. 1998;76:402-409.

Theodoropoulos D. Immune deficiency in CHARGE association. *Clin Med Res*. 2003;1:43-48.

Thurtle-Schmidt DM, Lo TW. Molecular biology at the cutting edge: A review on CRISPR/CAS9 gene editing for undergraduates. *Biochem Mol Biol Educ*. 2018;46:195-205.

Tischfield MA, Bosley TM, Salih MAM, Alorainy IA, Sener EC, Nester MJ, et al. Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. *Nature Genet.* 2005;37:1035-1037.

Trider C-L, Corsten G, Morrison D, Hefner M, Davenport S, Blake K. Understanding obstructive sleep apnea in children with CHARGE syndrome. *Int J Pediatr Otorhinolaryngol*. 2012;76:947-953.

Ufartes R, Schwenty-Lara J, Freese L, Neuhofer C, Möller J, Wehner P, et al. Sema3a plays a role in the pathogenesis of CHARGE syndrome. *Hum Mol Genet*. 2018;27:1343-1352.

Van Nostrand JL, Brady CA, Jung H, Fuentes DR, Kozak MM, Johnson TM, et al. Inappropriate p53 activation during development induces features of CHARGE syndrome. *Nature*. 2014;514:228-232.

Van Ravenswaaij-Arts C, Martin DM. New insights and advances in CHARGE syndrome: Diagnosis, etiologies, treatments, and research discoveries. *Am J Med Genet*. 2017;175C:397–406.

Van Ravenswaaij-Arts CM, Blake K, Hoefsloot L, Verloes A. Clinical utility gene card for: CHARGE syndrome - update 2015. *Eur J Hum Genet*. 2015;23(11).

Verloes A. Updated diagnostic criteria for CHARGE syndrome: A proposal. *Am J Med Genet A*. 2005;133:306-308.

Villate O, Ibarluzea N, Fraile-Bethencourt E, Valenzuela A, Velasco EA, Grozeva D, et al. Functional analyses of a novel splice variant in the CHD7 gene, found by next generation sequencing, confirm its

pathogenicity in a Spanish patient and diagnose him with CHARGE syndrome. *Front Genet*. 2018;9:7.

Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, et al. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*. 2004;36:955-957.

Vuorela P, Ala-Mello S, Saloranta S, Penttinen M, Poyhonen M, Huoponen K, et al. Molecular analysis of the CHD7 gene in CHARGE syndrome : identification of 22 novel mutations and evidence for a low contribution of large CHD7 deletions. *Genet Med*. 2007;9:690-694.

Vuorela P, Penttinen M, Hietala M, Laine J, Huoponen K, Kaariainen H. A familial CHARGE syndrome with a CHD7 nonsense mutation and new clinical features. *Clin Dysmorphol*. 2008;17:249-53.

Wessels MW, Brooks AS, Hoogeboom J, Niermeijer MF, Willems PJ. Kabuki syndrome: a review study of three hundred patients. *Clin Dysmorph*. 2002;11:95-102.

Wheeler PG, Quigley CA, Sadeghi-Nejad A, Weaver DD. Hypogonadism and CHARGE association. *Am J Med Genet*. 2000;94:228–231.

Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012;482:331-338.

Williams MS. Speculations on the pathogenesis of CHARGE syndrome. *Am J Med Genet A*. 2005;133A:318-325.

Wilson TE, Lieber MR. Efficient Processing of DNA Ends during Yeast Nonhomologous End Joining. *J Biol Chem*. 1999;274:23599-23609.

Wincent J, Holmberg E, StrömLand K, Soller M, Mirzaei L, Djureinovic T, et al. CHD7 mutation spectrum in 28 Swedish patients diagnosed with CHARGE syndrome. *Clin Genet*. 2008;74:31-38.

Won M, Dawid IB. PCR artifact in testing for homologous recombination in genomic editing in zebrafish. *PLos ONE.* 2017;12:e0172802.

Wood AJ, Lo TW, Zeitler B, Pickle CS, Ralston EJ, Lee AH, et al. Targeted genome editing across species using ZFNs and TALENs. *Science*. 2011;333:307.

Writzl K, Cale CM, Pierce CM, Wilson LC, Hennekam RC. Immunological abnormalities in CHARGE syndrome. *Eur J Med Genet*. 2007;50:338-345.

Würtele H, Little KCE, Chartrand P. Illegitimate DNA integration in mammalian cells. *Gene Ther.* 2003;10:1791-1799.

Wyse RK, al-Mahdawi S, Burn J, Blake K. Congenital heart disease in CHARGE association. *Pediatr Cardiol*. 1993;14:75–81.

Xie F, Ye L, Chang JC, Beyer AI, Wang J, Muench MO, et al. SeamLess gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res*. 2014;24:1526-1533.

Xue W, Chen S, Yin H, Tammela T, Papagiannakopoulus T, Joshi NS, et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*. 2014;514:380-384.

Yang L, Guell M, Byrne S, Yang JL, De Los Angeles A, Mali P, et al. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.* 2013;41:9049-9061.

Yao H, Hill SF, Skidmore JM, Sperry ED, Swiderski DL, Sanchez GJ, et al. CHD7 represses the retinoic acid synthesis enzyme ALDH1A3 during inner ear development. *JCI Insight*. 2018;3:e97440.

Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M, et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol*. 2014;32:551-553.

Yoshimi K, Kunihiro Y, Kaneko T, Nagahora H, Voigt B, Mashimo T. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun.* 2016;7:0431.

Yu C, Liu Y, Ma T, Liu K, Xu S, Zhang Y, et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell*. 2015;16:142-147.

Yu T, Meiners LC, Danielsen K, Wong MT, Bowler T, Reinberg D, et al. Deregulated FGF and homeotic gene expres-sion underlies cerebellar vermis hypoplasia in CHARGE syndrome. *Elife*.2013;2:e01305.

Zentner GE, Hurd EA, Schnetz MP, Handoko L, Wang C, Wang Z, et al.

CHD7 function in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Hum Mol Genet*. 2010;19:3491-3501.

Zentner GE, Layman WS, Martin DM, Scacheri PC. Molecular and phenotypic aspect of CHD7 mutation in CHARGE syndrome. *Am J Med Genet*. 2010;152: 674-86.

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol*. 2011;29:149-153.

Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum Mol Genet*. 2014;23:R40-46.

Zorin B, Hegemann P, Sizova I. Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in Chlamydomonas reinhardtii. *Eukaryot Cell*. 2005;4:1264-1272.

ANNEX

Annex 1. SDS Polyacrylamide Gel Electrophoresis

5% gradient gel composition		10% gradient gel composition	
Lower gel (separation gel):		**Lower gel (separation gel):**	
Water	2.5 mL	Water	2 mL
40% Acrylamide	0.5 mL	40% Acrylamide	1 mL
Lower Buffer	1 mL	Lower Buffer	1 mL
Temed	4 µL	Temed	4 µL
APS	40 µL	APS	40 µL
Upper gel (concentration gel):		**Upper gel (concentration gel):**	
Water	1.3 mL	Water	1.3 mL
40% Acrylamide	0.2 mL	40% Acrylamide	0.2 mL
Lower Buffer	0.5 mL	Lower Buffer	0.5 mL
Temed	2 µL	Temed	2 µL
APS	20 µL	APS	20 µL

*5% SDS-PAGE gel is used for analysis a protein with molecular weight upper 70 kDa

*10% SDS-PAGE gel is used for analysis a protein with molecular weight between 20 and 70 kDa

Composition of acrylamide/bis-acrylamide gel:

Lower buffer pH 8.8: 1.5 M Tris; 0.4% SDS (w/v)

Upper buffer pH 6.8: 0.5 M Tris; 0.4% SDS (w/v)

APS: 10% ammonium persulfate (p/v)

Developer and fixer solutions: Kodak X-OMAT EX II

Molecular weight β-actin: 42 kDa

Molecular weight CHD7: 340 kDa

Annex 2. Composition of Buffers and Solutions

Electrophoresis Buffer:

100 mL TG-SDS 10X (0.25 M Tris; 1.92 M Glycine, 1% SDS)

dH_2O to 1 L

PBS-Tween Solution:

100 mL PBS 10X

1 mL Twin (Tween 20)

dH_2O to 1 L

Lysis buffer, pH 7.5:

10 mM Tris

0.5% DOC (w/v)

1% NP-40 (v/v)

1 mM Pefabloc SC (Fluka Analytical)

Laëmmli buffer 2X, pH 6.8:

125 mM Tris

4.5% SDS (v/v)

30% glycerol (v/v)

0.002% bromophenol blue (w/v)

5% 1-Thioglycerol (v/v)

Annex 3. Antibodies for Western blot and Immunofluorescence

Primary antibodies:

- Rabbit monoclonal antibody anti-CHD7 (1/1000, Cell Signaling)
- Mouse monoclonal antibody anti-HA (1/1000, Sigma-Aldrich)
- Mouse monoclonal antibody anti-FLAG (1/2000, Sigma-Aldrich)
- Mouse monoclonal antibody anti-β-actin (1/10000, Sigma-Aldrich)
- Rabbit polyclonal antibody anti-UBF (1/100, Santa Cruz)
- Mouse monoclonal antibody anti-nucleolin (C23; 1/100, Santa Cruz)

Secondary antibodies:

- Polyclonal antibody goat anti-mouse IgG peroxidase-linked (1/10000, Sigma-Aldrich)
- Polyclonal antibody goat anti-rabbit IgG peroxidase-linked (1/5000, Sigma-Aldrich)
- RRX-conjugated goat anti-mouse IgG (H+L) (1/100, Jackson ImmunoResearch Laboratories)
- FITC-conjugated donkey anti-rabbit IgG (H+L) (1/50, Jackson ImmunoResearch Laboratories)