

Electronic Thesis and Dissertation Repository

2-11-2019 9:00 AM

Improving the genetic diagnosis of familial hypercholesterolemia

Michael Iacocca, *The University of Western Ontario*

Supervisor: Hegele, Robert A., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biochemistry

© Michael Iacocca 2019

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons](#), [Cardiovascular Diseases Commons](#), [Medical Genetics Commons](#), and the [Medical Molecular Biology Commons](#)

Recommended Citation

Iacocca, Michael, "Improving the genetic diagnosis of familial hypercholesterolemia" (2019). *Electronic Thesis and Dissertation Repository*. 6017.

<https://ir.lib.uwo.ca/etd/6017>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Familial hypercholesterolemia (FH) is a genetic disorder of severely elevated low-density lipoprotein (LDL) cholesterol that is widely underdiagnosed and undertreated. To improve the identification of FH and initiate timely and appropriate treatment strategies, genetic testing is becoming increasingly offered worldwide as a central part of diagnosis. I describe three main ways providing a genetic diagnosis in FH can be improved. First, next-generation sequencing (NGS)-based approaches can be used to reliably identify large-scale variant types known as copy number variations (CNVs) in the LDL receptor gene (*LDLR*); second, NGS methodology can be further applied to extend CNV screening to additional FH-associated genes, which have remained uninvestigated but may harbor novel causative variation; and third, the interpretation of variants identified during the course of genetic testing can be improved with the establishment of an open-source database containing variants identified in FH patients worldwide.

Keywords

Familial hypercholesterolemia, LDL cholesterol, cardiovascular disease, next-generation sequencing, copy number variation, DNA variant pathogenicity, genetic diagnosis, genetic screening, clinical database

Co-Authorship Statement

Dr. Robert A. Hegele (supervisor) provided funding, supervision, contributed to study design, manuscript preparation, and critical revision for all Chapters.

Chapter 1 contains material from the manuscript entitled, ‘Recent advances in genetic testing for familial hypercholesterolemia’, published in *Expert Review in Molecular Diagnostics* on May 29, 2017, and co-authored by Dr. Robert A. Hegele. (PMID: 28524730).

Chapter 2 contains material from the manuscript entitled, ‘Use of next-generation sequencing to detect *LDLR* gene copy number variation in familial hypercholesterolemia’, published in the *Journal of Lipid Research* on September 5, 2017, and co-authored by Dr. Jian Wang, Ms. Jacqueline S. Dron, Mr. John F. Robinson, Mr. Adam D. McIntyre, Dr. Henian Cao, and Dr. Robert A. Hegele. (PMID: 28874442).

In Chapter 2, Dr. Jian Wang and Mr. Adam McIntyre performed the multiplex ligation-dependent probe amplification bench work and subsequent data analysis for 310 of the 388 samples. Dr. Henian Cao performed the LipidSeq next-generation sequencing bench work for all 388 samples. Dr. Jian Wang, Mr. Adam McIntyre, Ms. Jacqueline Dron, and Mr. John Robinson assisted in establishing and optimizing the bioinformatic workflow for copy number variation analysis.

Chapter 3 contains material from the manuscript entitled, ‘Whole-gene duplication of *PCSK9* as a novel genetic mechanism for severe familial hypercholesterolemia’, published in the *Canadian Journal of Cardiology* on August 4, 2018, and co-authored by Dr. Jian Wang, Ms. Samantha Sarkar, Ms. Jacqueline S. Dron, Dr. Thomas Lagace, Mr. Adam D. McIntyre, Ms. Paulina Lau, Mr. John F. Robinson, Dr. Ping Yang, Dr. Joan H. Knoll, Dr. Henian Cao, Dr. Ruth McPherson, and Dr. Robert A. Hegele. (PMID: 30269829).

In Chapter 3, Dr. Henian Cao performed the LipidSeq next-generation sequencing bench work for all 704 samples. Dr. Jian Wang assisted in the VarSeq copy number variation analysis of LipidSeq data. The London Regional Genomics Center (London, ON) performed

the whole-exome sequencing bench work (2 samples). The London Health Sciences Centre Clinical Cytogenetics laboratory at Victoria Hospital (London, ON) performed the CytoScan microarray bench work (2 samples). Dr. Thomas Lagace and Ms. Samantha Sakar performed the plasma PCSK9 analysis (1 sample). Dr. Jian Wang, Mr. Adam McIntyre, Ms. Jacqueline Dron, and Mr. John Robinson assisted in establishing and optimizing the bioinformatic workflow for copy number variation analysis.

Chapter 4 contains material from the manuscript entitled, ‘ClinVar database of global familial hypercholesterolemia-associated DNA variants’, published in *Human Mutation* on October 11, 2018, and co-authored by Ms. Joana R. Chora, Dr. Alain Carrie, Dr. Tomas Freiburger, Dr. Sarah E. Leigh, Dr. Joep C. Defesche, Dr. Lisa C. Kurtz, Dr. Marina T. DiStefano, Dr. Raul D. Santos, Dr. Steve E. Humphries, Dr. Pedro Mata, Dr. Cinthia E. Jannes, Dr. Amanda J. Hooper, Ms. Katherine A. Wilemon, Dr. Pascal Benlian, Dr. Robert O' Connor, Dr. John Garcia, Ms. Hannah Wand, Dr. Lukas Tichy, Dr. Eric J. Sijbrands, Dr. Robert A. Hegele, Dr. Mafalda Bourbon, and Dr. Joshua W. Knowles: On behalf of the ClinGen FH Variant Curation Expert Panel. (PMID: 30311388).

In Chapter 4, the following co-authors contributed variant and/or variant-level data to the ClinVar database: Dr. Sarah E. Leigh transferred 1670 variants from the Leiden Open Source Variation Database, Dr. Joep C. Defesche submitted 757 variants, Dr. Alain Carrie submitted 434 variants, Dr. Mafalda Bourbon and Ms. Joana R. Chora submitted 399 variants, Dr. Pascal Benlian submitted 344 variants, Dr. Pedro Mata submitted 331 variants, Dr. Raul D. Santos and Dr. Cinthia E. Jannes submitted 280 variants, Dr. Tomas Freiburger and Dr. Lukas Tichy submitted 197 variants, Dr. John Garcia submitted 196 variants, Dr. Amanda J. Hooper submitted 152 variants, and Dr. Robert O'Connor submitted 113 variants. Illumina Clinical Services laboratory submitted 362 variants. Ms. Joana R. Chora assisted with data analysis and contributed to study design, manuscript preparation, and critical revision. Dr. Mafalda Bourbon and Dr. C. Lisa Kurtz provided supervision and contributed to study design, manuscript preparation, and critical revision. Dr. Joshua W. Knowles provided supervision and contributed to study design and critical revision.

Dedication

For my mom, Gail.

Acknowledgements

Foremost, I would like to sincerely thank my supervisor, Dr. Robert Hegele for providing me with an opportunity to be part of your laboratory, and for your mentorship and guidance throughout all of my projects. Thank you for teaching me about all things FH, but more importantly about professionalism, sincerity, and humility; there is no one better in the world to learn from.

Thank you to my advisors Dr. Michael Boffa and Dr. Murray Huff, for providing me with excellent guidance and support in committee meetings. I valued your feedback, your genuine interest and willingness to engage, which helped me to improve and progress as a trainee. Also, a sincere thank you to Dr. Murray Huff for all your positive encouragement and kind comments that were provided during conferences and throughout my presentations.

To all members of the Hegele laboratory, thank you for always being so kind, helpful, and supportive throughout my time at Robarts. Thank you to the brilliant Dr. Jian Wang for your guidance, assistance, and expertise you provided me in all technical and bioinformatic aspects of my projects. Thank you to Mr. Adam McIntyre for your willingness to assist me with genotype information, clinic and research charts, and organization of databases, no matter how busy you would be. Thank you to Dr. Henian Cao for all of world-class bench work you perform for us graduate students, and for talking sports at any hallway encounter. Thank you Mr. Matthew Ban for your assistance with statistical analyses, and Mr. John Robinson for your help managing and organizing all things I needed to proceed with my projects. Thank you to Mrs. Brooke Kennedy and Ms. Ericka Simon for all of your assistance with clinic patient information. Thank you to my fellow graduate students and dear friends Ms. Jacqueline Dron, Ms. Rosettia Ho, and Ms. Allison Dilllott, for all of your unwavering support and encouragement you provided me throughout my time in the lab, and an extra thank you to Ms. Jacqueline Dron for all of your assistance and collaboration in experimental design and data analysis for my projects.

Thank you to my family: parents Gail and Tony, siblings Jennifer, Daniel, Emily, and Mackenzie, and grandmothers Margaret and Carmela, for your unwavering support, love, and

encouragement. Thank you to my significant other Olivia, for taking care of me all of these years, allowing me to pursue my research and career aspirations.

Thank you to the Canadian Institutes of Health Research (Foundation Grant), the Heart and Stroke Foundation of Ontario, and Genome Canada for providing funding for my research projects. And lastly, I am grateful to the Department of Biochemistry, the Schulich School of Medicine and Dentistry, and School of Post Graduate and Postdoctoral Studies of Western University, for resource, administrative, and financial support.

Table of Contents

Abstract.....	i
Co-Authorship Statement.....	ii
Dedication.....	iv
Acknowledgements.....	v
Table of Contents.....	vii
List of Tables.....	x
List of Figures.....	xi
List of Appendices.....	xii
Abbreviations.....	xiii
Chapter 1 – Introduction.....	1
1.1 LDL cholesterol.....	1
1.2 Genetic basis of familial hypercholesterolemia.....	4
1.2.1 <i>LDLR</i>	4
1.2.2 <i>APOB</i>	7
1.2.3 <i>PCSK9</i>	7
1.2.4 <i>LDLRAP1</i>	8
1.2.5 Additional FH-associated genes.....	8
1.2.6 Polygenic hypercholesterolemia.....	10
1.3 Prevalence of FH, associated risks, and treatment.....	10
1.4 Is there a need for a genetic diagnosis in FH?.....	12
1.4.1 The importance of cascade screening.....	14
1.4.2 Does knowledge of the genotype impact upon clinical management?.....	15
1.4.3 Other reasons to seek the genetic basis in patients suspected to have FH.....	18
1.5 What genetic testing is available for FH?.....	18
1.5.1 DNA hybridization assays.....	18
1.5.2 Sanger sequencing.....	19
1.5.3 Next-generation sequencing (NGS) platforms.....	20
1.5.4 Determining pathogenicity of identified variants.....	24
1.5.5 CNV analysis in the diagnosis of FH.....	25
1.5.6 Analysis of polygenic elevated LDL cholesterol.....	27
1.6 Synopsis.....	29
1.7 References.....	30
Chapter 2 – Use of next-generation sequencing to detect <i>LDLR</i> gene copy number variation in the genetic diagnosis of FH.....	39
2.1 Introduction.....	39
2.2 Materials and Methods.....	41
2.2.1 Study subjects.....	41
2.2.2 Targeted next-generation sequencing.....	41
2.2.3 CNV detection by MLPA.....	42

2.2.4	CNV detection by NGS data	42
2.2.5	CNV filtration	43
2.2.6	Statistical analyses	44
2.3	Results	44
2.3.1	Study sample demographics	44
2.3.2	CNVs detected by MLPA	45
2.3.3	CNVs detected by NGS data	45
2.4	Discussion	52
2.5	Conclusion	54
2.6	References	55

Chapter 3 – Novel copy number variation screening in secondary FH-associated genes 57

3.1	Introduction	57
3.2	Materials and Methods	58
3.2.1	Study subjects	58
3.2.2	Targeted next-generation sequencing (NGS)	58
3.2.3	Whole-exome NGS	58
3.2.4	NGS CNV analysis	59
3.2.5	CNV confirmation by microarray analysis	59
3.2.6	Plasma PCSK9 analysis	59
3.2.7	Statistical analysis	60
3.3	Results	60
3.3.1	Study sample demographics	60
3.3.2	Targeted NGS data CNV analysis	61
3.3.3	Whole-exome CNV analysis	61
3.3.4	CNV confirmation	65
3.3.5	Case presentations	66
3.3.6	Co-segregation analysis	68
3.3.7	Plasma PCSK9 levels	72
3.4	Discussion	74
3.5	Conclusion	76
3.6	References	78

Chapter 4 – ClinVar database: data sharing to improve interpretation of genetic variants identified in FH patients worldwide 80

4.1	Introduction	80
4.2	Materials and Methods	81
4.2.1	ClinGen FH Variant Curation Expert Panel	81
4.2.2	Variant submission to ClinVar	83
4.2.3	ClinVar variant analysis	84
4.3	Results	85
4.3.1	Global ClinVar submission	85
4.3.2	FH-associated variant characteristics	87
4.3.3	Variant classification methods	91
4.3.4	Variant-level data	95
4.4	Discussion	96

4.5 Conclusion	100
4.6 References.....	101
Chapter 5 – Discussion and conclusions.....	104
5.1 Overview.....	104
5.2 Study findings and implications.....	104
5.3 Future directions	107
5.4 Final conclusions	109
5.5 References.....	110
Appendices.....	111
Curriculum Vitae	113

List of Tables

Table 1.1. Commonly used criteria for the clinical diagnosis of FH.	13
Table 1.2. Targeted NGS panels commercially available for FH genetic diagnosis.	23
Table 1.3. SNPs used in reported genetic risk scores for hypercholesterolemia.	28
Table 2.1. Canadian FH cohort patient demographics.	44
Table 2.2. <i>LDLR</i> gene whole-exon CNVs identified in 388 patients with FH.	46
Table 2.3. 2 X 2 contingency analysis of CNVs called by MLPA and NGS methods.	51
Table 3.1. FH cohort patient demographics.	61
Table 4.1. ClinGen FH Variant Curation Expert Panel members.	82
Table 4.2. Number of variants submitted to ClinVar by gene.	85
Table 4.3. Centers that submitted FH-associated variants to ClinVar.	86
Table 4.4. Unique FH-associated variants submitted to ClinVar by gene and variant type. ..	87
Table 4.5. Clinical significance of <i>all</i> FH-associated variant submissions in ClinVar, independent of criteria used for classification.	90
Table 4.6. Clinical significance of <i>unique</i> FH-associated variants in ClinVar, independent of criteria used for classification.	90
Table 4.7. Criteria used for FH-associated variant classifications in ClinVar.	91
Table 4.8. Criteria for variant classification used by each submitting center.	93
Table 4.9. Number of unique variants with each variant-level data type available in ClinVar.	95

List of Figures

Figure 1.1. LDL cholesterol homeostasis via the LDL receptor.....	3
Figure 1.2. Relative proportions of <i>LDLR</i> variant types identified in FH patients.....	5
Figure 1.3. <i>LDLR</i> CNVs described in patients with FH.	6
Figure 2.1. Two methods of detection of a CNV deletion event in the <i>LDLR</i> gene in a patient with FH.	49
Figure 2.2. Two methods of detection of a CNV duplication event in the <i>LDLR</i> gene in a patient with FH.	51
Figure 3.1. NGS-based detection of a <i>PCSK9</i> CNV in a patient with FH (index case A).	63
Figure 3.2. NGS-based detection of a <i>PCSK9</i> CNV in a patient with FH (index case B).....	64
Figure 3.3. Microarray-based confirmation of a <i>PCSK9</i> CNV in two patients with FH.....	66
Figure 3.4. Family pedigree of two FH index cases with a whole-gene duplication of <i>PCSK9</i>	69
Figure 3.5. Targeted NGS-based CNV assessment of <i>PCSK9</i> in an individual with FH (affected father of index case A).....	70
Figure 3.6. Targeted NGS-based CNV assessment of <i>PCSK9</i> in an individual with FH (affected daughter of index case A).	71
Figure 3.7. Targeted NGS-based CNV assessment of <i>PCSK9</i> in an individual without FH (unaffected mother of index case A).....	71
Figure 3.8. Targeted NGS-based CNV assessment of <i>PCSK9</i> in an individual with FH (affected sister of index case B).....	72
Figure 3.9. Plasma PCSK9 level in a FH patient (index case A) with a <i>PCSK9</i> whole-gene duplication.....	73
Figure 4.1. Number of unique FH-associated variants in ClinVar by exonic location and variant type.....	89
Figure 4.2. Number of unique variants ($n=2796$) classified by different sets of criteria.....	92

List of Appendices

Appendix A - Ethics approval	111
Appendix B – Journal copyright approval	112

Abbreviations

ABCG5/8	adenosine triphosphate binding cassette subfamily G member 5/8
ACGS	Association for Clinical Genetic Science
ACMG/AMP	American College of Medical Genetics/Association of Molecular Pathologists
APOB	apolipoprotein B
APOE	apolipoprotein E
ATP	adenosine triphosphate
CABG	coronary arterial bypass graft
CAD	coronary artery disease
CHD	coronary heart disease
CVD	cardiovascular disease
DFH	definite familial hypercholesterolemia
DLCN	Dutch Lipid Clinic Network
DNA	deoxyribonucleic acid
DOC	depth of coverage
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
FH	familial hypercholesterolemia
GOF	gain of function
GRS	genetic risk score
GWAS	genome-wide association study
HDL	high-density lipoprotein
HeFH	heterozygous familial hypercholesterolemia
HMG-CoA	5-hydroxy-3-methylglutaryl-coenzyme A
HoFH	homozygous familial hypercholesterolemia
LAL	lysosomal acid lipase
LALD	lysosomal acid lipase deficiency
LDL	low-density lipoprotein
LDLR	low-density lipoprotein receptor

LDLRAP1	low-density lipoprotein receptor adapter protein 1
LOVD	Leiden Open Source Variation Database
MI	myocardial infarction
MLPA	multiplex ligation-dependent probe amplification
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
PCR	polymerase chain reaction
PCSK9	proprotein convertase subtilisin kexin type 9
RNA	ribonucleic acid
SBR	Simon Broome Register
SBS	sodium bisulfite sequencing
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
SREBP-2	sterol regulatory element binding protein 2
STAP1	signal transducer adapter protein 1
TC	total cholesterol
TG	triglycerides
UTR	untranslated region
VAF	variant allele frequency
VC-EP	Variant Curation Expert Panel
VLDL	very low-density lipoprotein
VUS	variant of unknown significance
WES	whole-exome sequencing
WGS	whole-genome sequencing

Chapter 1 – Introduction

1.1 LDL cholesterol

Cholesterol is an important biological molecule that has an essential role in cell membrane structure, and as a precursor for the biosynthesis of vitamin D, bile acid, and steroid hormones. Only ~ one-fifth of cholesterol in humans originates from dietary intake; the majority of cholesterol is synthesized by the liver (Brown & Goldstein, 1986).

As an insoluble lipid, cholesterol requires two main types of lipoproteins for transport through blood plasma: high-density lipoproteins (HDL) and low-density lipoproteins (LDL) (Hegele, 2009). Cholesterol transported within HDL particles (HDL-C) is often referred to as “good” cholesterol, as it associates with the movement of cholesterol from peripheral tissue to the liver, where it can be safely disposed of (Lagor & Millar, 2010). LDL cholesterol (LDL-C) however, is often referred to as “bad” cholesterol, as it associates with cholesterol transport from the liver to peripheral tissues, where it is deposited (Lagor & Millar, 2010). These colloquial expressions “bad” and “good” thus refer to the direction of cholesterol transport, rather than the cholesterol itself, which is qualitatively the same.

When LDL-C levels are excessively elevated, the deposition of cholesterol into peripheral tissues, such as the arterial walls, can become accelerated (Hegele, 1997). In the arterial walls, excess cholesterol is oxidatively modified by macrophages, initiating inflammation of vasculature and formation of atherosclerotic plaques (Hegele, 1997). Atherosclerosis causes a narrowing of the arteries, and is a major risk factor for the development of cardiovascular diseases (CVDs), such as coronary heart disease (CHD; restricted blood flow to the heart) and cerebrovascular disease (restricted blood flow to the brain) (Berenson et al., 1998). Alarming, for every mmol/L increment of raised LDL-C, the risk of CHD has been shown to increase by ~40% (Sharrett et al., 2001).

LDL-C homeostasis is primarily governed via regulation of the LDL receptor (LDLR) protein, and is explained briefly below in **Figure 1.1**.

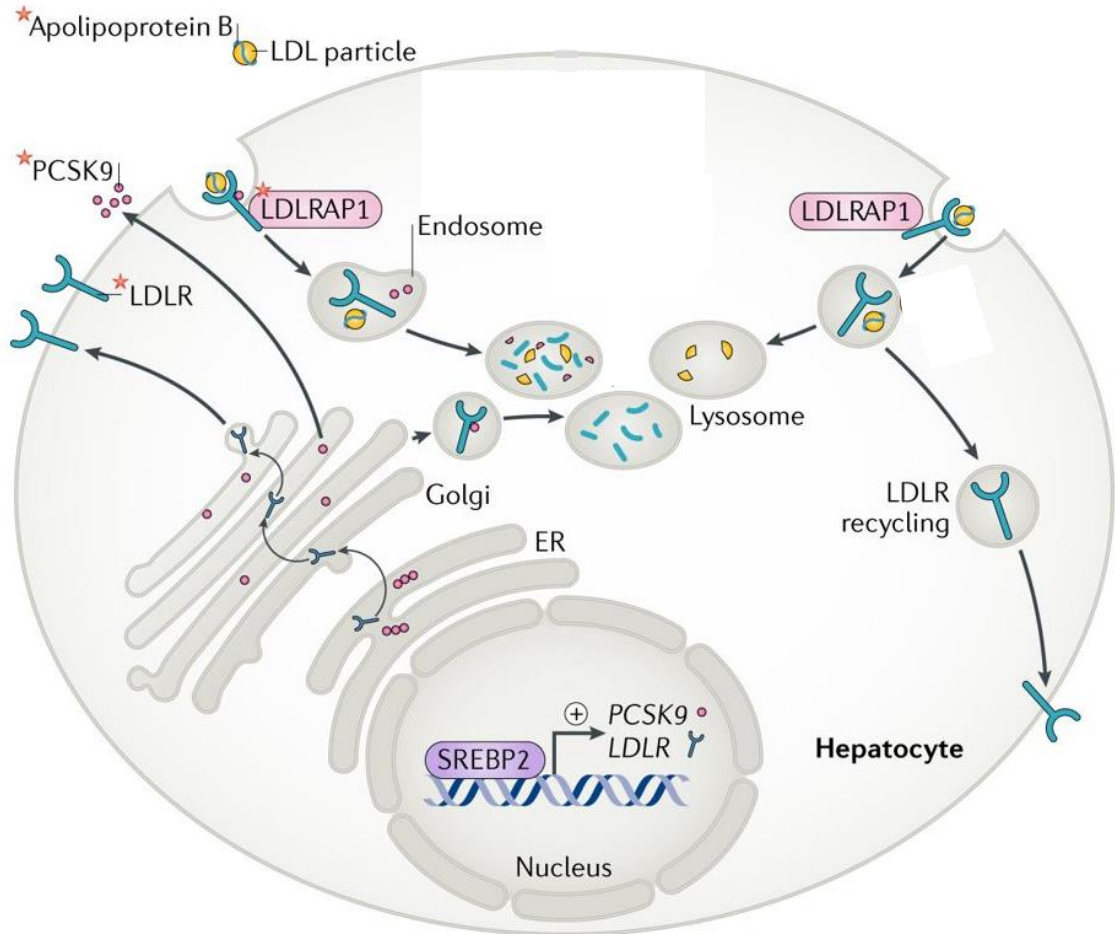


Figure 1.1. LDL cholesterol homeostasis via the LDL receptor. When intracellular cholesterol levels are low, expression of the LDLR gene (*LDLR*) is upregulated by the sterol regulator element binding protein-2 (SREBP-2) transcription factor (Horton et al., 2002). LDLR mRNA is translated in the endoplasmic reticulum and matures in the Golgi apparatus before reaching the cell surface. On the cell surface, LDLRs concentrate in clathrin-coated pits; while LDLRs are found on most cells they are most highly abundant on hepatocytes. Apolipoprotein B₁₀₀ (Apo B) is the major apolipoprotein of the LDL particle. Embedded in LDL's outer phospholipid layer, Apo B is the ligand for the LDLR (Chatterton et al., 1995). Following binding of the LDL particle to LDLR, the receptor-ligand complex is internalized by receptor-mediated endocytosis (Brown & Goldstein, 1979). The intracellular LDLR adapter protein 1 (LDLRAP1) interacts with a NPXY motif within the cytoplasmic tail of LDLR to facilitate this internalization at clathrin-coated pits (Chen, Goldstein, & Brown, 1990; Garcia et al., 2001). Clathrin-coated vesicles containing the LDLR-LDL complex bud off the cell membrane and then fuse to form the early endosome (Anderson, Brown, & Goldstein, 1977). In the late endosome, a pH change leads to the dissociation of the LDLR-LDL complex; the LDL particle is delivered to the lysosome where the protein component is hydrolyzed, liberating the cholesterol for cellular use (Goldstein et al., 1975). The LDLR is normally recycled back to the cell surface to bind more circulating LDL particles, initiating another cycle of receptor-mediated endocytosis; the LDLR can be recycled as many as 100 times (Brown et al., 1983). In response to high intracellular cholesterol, expression of *LDLR* is downregulated by inhibition of SREBP-2, and LDLRs begin to get degraded in the late endosome rather than recycled to the cell surface. Degradation of LDLR is mediated by proprotein convertase subtilisin-kexin 9 (PCSK9). PCSK9 exists both intracellularly and extracellularly, and in both cases directs the LDLR to endosome (Zhang et al., 2007). In the endosome PCSK9 disrupts the pH-induced dissociation of the LDLR-LDL complex and the complex progresses through to the lysosome where it is degraded. Interestingly, *PCSK9* expression is also upregulated by the SREBP-2 transcription factor, likely to maintain tight regulatory control of plasma LDL-C. (Figure adapted from (Berberich & Hegele, 2018)).

LDL-C levels (normal: 2-3 mmol/L) are influenced by a range of both environmental and genetic determinants, responsible for the right-skewed, normal distribution within the population (Hegele, 2009). For extreme LDL-C phenotypes however, a more prominent genetic influence is expected (Hegele, 2009). Familial hypercholesterolemia (FH) is defined as the heritable condition of elevated LDL-C, characterized by LDL-C concentrations within the 95th percentile adjusted for age and sex ($\sim \geq 5$ mmol/L in adults).

FH as a monogenic disorder has been historically well characterized in the literature. For those with monogenic FH, extreme LDL-C phenotypes can be explained by the presence of single-gene, large-effect DNA variants, capable of causing severe disruptions to LDL-C homeostasis (Rader et al., 2003).

1.2 Genetic basis of familial hypercholesterolemia

1.2.1 *LDLR*

More than 90% of genetically defined FH cases are due to autosomal codominant variants in the *LDLR* gene. Extensive study of *LDLR* in FH patients has led to the identification of >1700 unique variants prior to 2018 (Bourbon et al., 2017). Variants causing a loss of function in the encoded protein are causative for FH; such variants have been described throughout all 18 exons of *LDLR* and include single nucleotide changes, such as missense variants altering a single amino acid residue, nonsense variants introducing a premature stop codon, and splicing variants; small insertion/deletion (indel) variants which remove/add amino acids, or alter the reading frame (frameshift); and large indels or copy number variations spanning one or multiple exons (see below). The relative proportions of *LDLR* variant types identified in FH patients are illustrated in **Figure 1.2**.

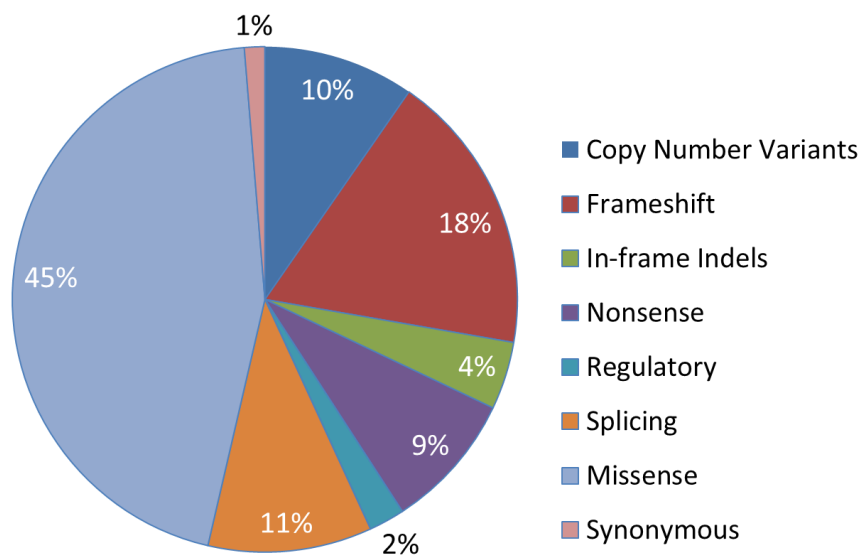


Figure 1.2. Relative proportions of *LDLR* variant types identified in FH patients.

1.2.1.1 *LDLR* copy number variations

Copy number variations (CNVs) are a form of quantitative structural rearrangement that include deletions, duplications, and higher order amplifications of DNA sequence larger than 50 bases in size (Redon et al., 2006).

The *LDLR* locus appears to be prone to acquiring CNVs across evolutionary time. The locus contains a particularly high abundance of *Alu* sequences - the most prevalent repeat element in the genome - making it especially susceptible to CNV mutagenesis via mechanisms associated with faulty DNA repair, replication, and recombination, such as non-homologous end joining (Goldmann et al., 2010), replication fork stalling and template switching (Horsthemke et al., 1987), and non-allelic homologous recombination (Lehrman et al., 1985), respectively.

Molecular biological studies over the past three decades indicate a remarkably large number of naturally occurring *LDLR* CNVs in FH probands and families: 56 unique

deletions and 27 unique duplications (**Figure 1.3**) prior to 2018 (Iacocca & Hegele, 2018). Of the 98 *Alu* repeats in *LDLR*, 95 are within the introns, and account for 85% of the intronic sequence outside of splice-site junctions (Amsellem et al., 2002), explaining why essentially all *LDLR* CNVs so far described in FH patients are in-frame, whole-exon events (i.e. breakpoints located within the introns). Despite the comparable number of unique deletions and duplications reported among the literature, in any single FH cohort the large majority (>90%) of *LDLR* CNV events detected have been heterozygous deletions, often involving multiple exons (Bertolini et al., 2017; Jannes et al., 2015; Miyake et al., 2009; Taylor et al., 2009).

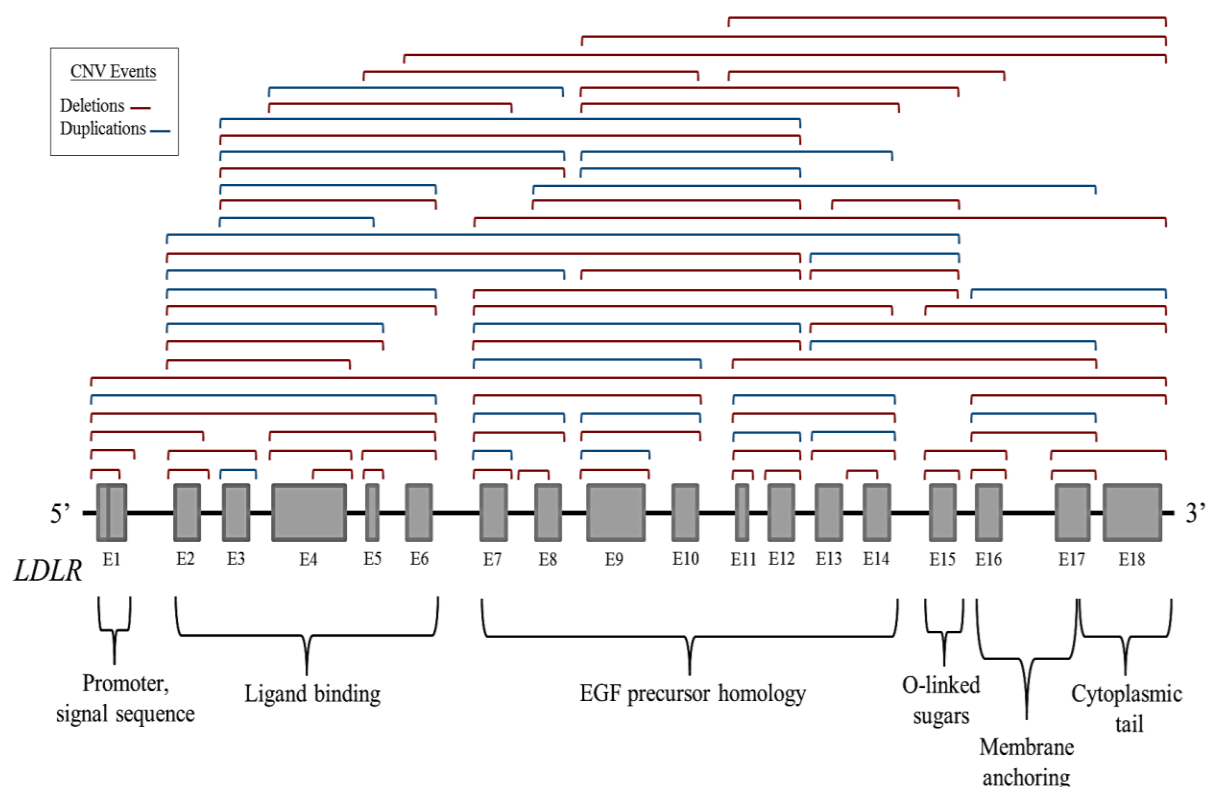


Figure 1.3. *LDLR* CNVs described in patients with FH. The *LDLR* gene structure and its 18 exons (E#) are shown from left to right (5' to 3').

1.2.2 *APOB*

Autosomal codominant variants in *APOB* underlie 5-8% of FH cases. FH-causing *APOB* variants encode a secreted Apo B that is still available for the formation of the LDL particle (more specifically its very-LDL (VLDL) particle precursor), but binds defectively to the LDL receptor. Known LDL receptor binding domains are located in *APOB* exons 26 and 29 (Kriško & Etchebest, 2007). Although there have been numerous protein-altering variants located in *APOB* exons 26 and 29 described in FH patients, the majority of *APOB* cases are due to a single amino acid substitution of arginine for glutamine at residue 3527 [exon 26; p.(Arg3527Gln)], shown to significantly reduce the affinity of Apo B for the LDLR (Innerarity et al., 1987). Interestingly, recent evidence has begun to suggest causative FH variants may also exist elsewhere throughout the *APOB* gene, although the exact mechanisms of impaired LDLR binding due to such variants are still mostly unclear (Alves et al., 2014). Causative variant types so far identified in *APOB* include frameshift, missense, nonsense, and splicing (Chora et al., 2018). Importantly, null/deleterious variants in *APOB* which prevent VLDL/LDL particle formation are known to cause hypocholesterolemia – i.e. low LDL-C.

1.2.3 *PCSK9*

Autosomal codominant variants in *PCSK9* underlie ~1% of FH cases. In *PCSK9*, gain-of-function (GOF) variants - which lead to increased degradation of the LDLR – are causative for FH. Disease-causing GOF variants have been described throughout all domains of the encoded protein, and have so far included a 5' untranslated region (UTR) substitution, missense, splicing, and small indel variant types (Dron & Hegele, 2017). Importantly, variants causing a loss of function in *PCSK9* reduce LDLR degradation and are known to cause hypocholesterolemia.

1.2.4 *LDLRAP1*

Unlike *LDLR*, *APOB*, and *PCSK9*, variants in *LDLRAP1* are recessive - i.e. an individual must carry two mutant alleles to exhibit an FH phenotype; autosomal recessive variants in *LDLRAP1* underlie <1% of FH cases, and are a form of homozygous FH (HoFH). Loss of function variants which reduce the capacity of *LDLRAP1* to interact with the cytoplasmic tail of *LDLR* in clathrin-coated pits are causative of FH (Wijers et al., 2015). Causative variant types so far described in *LDLRAP1* include frameshift, missense, nonsense, and splicing.

1.2.5 Additional FH-associated genes

Systematic studies of clinical FH patients with no apparent causative variants in *LDLR*, *APOB*, *PCSK9* (or *LDLRAP1*) have revealed a handful of ultra-rare variants in other genes that may occasionally cause FH or an FH-like phenotype (i.e. phenocopies). These include autosomal dominant variants in *APOE* and *STAP1*, and autosomal recessive variants in *LIPA* and *ABCG5/8*.

APOE encodes apolipoprotein E (Apo E). In 2012, Marduel et al. identified an in-frame deletion (c.500_502delTCC) resulting in the elimination of a leucine residue at position 167 (p.Leu167del) in *APOE* in a large French family of 14 members with clear autosomal dominant FH. This variant cosegregated with affected status in this index family, and also in two other families identified. Apo E is an apolipoprotein present on VLDL particles; functional evidence suggested that cellular VLDL uptake is increased by the p.Leu167del variant, leading to increased intracellular cholesterol and thus a down regulation of *LDLR* transcription (Cenarro et al., 2016). This down regulation of *LDLR* was thus the plausible explanation of isolated high LDL-C in FH. To date, this is the only FH-causing *APOE* variant to be reported.

STAP1 encodes signal transducer adapter protein 1 (STAP1). In 2014, Fouchier et al. described *STAP1* as a 5th putative locus for autosomal dominant FH. In 400 FH patients with no *LDLR/APOB/PCSK9* causal variants they combined parametric linkage analysis with whole-exome sequencing and identified four independent missense variants

(p.Glu97Asp, p.Leu69Ser, p.Ile71Thr, and p.Asp207Asn) in *STAP1* to associate with FH/high LDL-C in five families. *STAP1* has been suggested to act downstream of receptor tyrosine kinases (Masuhara et al., 2000), however, the function of *STAP1*, especially in relation to LDL-C, is still unknown. Association of *STAP1* to FH is still only supported by clinical data and cosegregation analysis in these five index case families, rather than by mechanistic evidence.

LIPA encodes lysosomal acid lipase (LAL). Recessive variants causing a loss of function in *LIPA* result in LAL deficiency (LALD). LAL is an enzyme that catalyzes the hydrolytic breakdown of lipids such as cholesteryl esters and triglycerides in the lysosome; the accumulation of intracellular lipids in LALD patients leads to toxic buildup of fats in multiple tissues (Goldstein et al., 1975). The severity of the disorder depends on the residual function of the mutated LAL; those with no enzymatic function develop severe multi-organ failure and rarely survive beyond the first year of life (Burton, Deegan, et al., 2015). Most LALD patients, however, experience signs and symptoms beginning in mid-childhood, and some later into adulthood. In these “later-onset” LALD cases, disease is still life-threatening and often accompanied by liver disease and high LDL-C levels (Burton, Deegan, et al., 2015). In some cases, patients with LALD may present clinically as FH (often HoFH) (Chora et al., 2017; Sjouke et al., 2016). The prevalence of LALD is still largely unknown, although most estimates suggest an incidence of ~1 in 40,000 to 1 in 300,000 (Reiner et al., 2014).

ABCG5 and *ABCG8* encode ATP-binding cassette sub-family G member 5 and 8, respectively. These proteins form an obligate heterodimer; recessive variants causing a loss of function in *ABCG5/8* result in sitosterolemia (Tada et al., 2018). Found mostly on hepatic and intestinal cells, the *ABCG5/8* transporter controls the efflux of dietary plant sterols (lipids from vegetable oils, nuts, avocados, etc.), and to a lesser extent dietary cholesterol; those with sitosterolemia have increased intestinal absorption and decreased biliary excretion of sterols. Although most patients with sitosterolemia have only mild to moderately elevated LDL-C levels, some, particularly children, may present clinically as HoFH, with xanthomas (cholesterol deposits within the skin; common in HoFH) and premature atherosclerosis (Brinton et al., 2018; Tada et al., 2018). Sitosterolemia is rare, with an estimated prevalence of ~1 in 50,000 to 1 in 200,000 (Yoo, 2016).

Together, causative variants in *APOE*, *STAP1*, *LIPA*, and *ABCG5/8* have been implicated in <1% of FH cases. However, because identifying a causative variant in one of these genes can help explain the FH phenotype in patients negative for *LDLR*, *APOB*, *PCSK9*, and *LDLRAP1* variants, and potentially impact treatment (see below), screening for them in a diagnostic context is warranted.

1.2.6 Polygenic hypercholesterolemia

Despite the progress in defining single-gene causes of FH, a substantial proportion (~20-40%) of patients with the FH phenotype validated by clinical criteria have no obvious disease-causing variant (Humphries et al., 2006; Soutar & Naoumova, 2007). Reasons for this include possible rare variants in novel genes, the confounding effects of gene-by-environment or gene-by-gene interactions, epigenetic mechanisms, and variant types that are missed by current technologies. But most importantly, polygenic inheritance explains phenotypic FH in many individuals without pathological monogenic variants (Talmud et al., 2013). Some of these patients carry a disproportionately high burden of common LDL-C raising alleles, possibly in complex interaction with environment to manifest in the clinical phenotype; a polygenic basis may explain 20% or more of patients with phenotypic FH (Wang et al., 2016).

1.3 Prevalence of FH, associated risks, and treatment

Heterozygous FH (HeFH; often just referred to as FH), which can be caused by heterozygous variants in *LDLR*, *APOB*, *PCSK9*, and *APOE* (and possibly *STAP1*), is estimated to affect 1 in 250 individuals in the general population (Akioyamen et al., 2017), making it among the most common known monogenic disorders. Further, the prevalence in certain founder populations is even higher, such as ~1 in 200 in French Canadians, ~1 in 165 in Tunisians, ~1 in 85 in Christian Lebanese, and ~1 in 72 in South African Afrikaners (Austin et al., 2004). Homozygous FH (HoFH; often specified as so)

is far more rare, with an estimated incidence of ~1 in 300,000 (Sjouke et al., 2015). HoFH can be caused by bi-allelic variants in *LDLRAP1*, bi-allelic variants in *LDLR* (either "simple homozygosity" for the identical variant or "compound heterozygosity" for two different variants), or heterozygous variants within two of the known autosomal dominant FH genes (*LDLR*, *APOB*, *PCSK9*, *APOE*, and possibly *STAP1*) - referred to as "double heterozygosity".

In HeFH, baseline LDL-C typically ranges from ~5 to 10 mmol/L in adults. If left untreated, atherosclerotic cardiovascular disease (CVD) onset occurs often by age 55 in men and 60 in women, with half of all affected men and 15% of women suffering from myocardial infarction (MI) before these ages (Nordestgaard et al., 2013). In HoFH, baseline LDL-C ranges from ~12 to upwards of >20 mmol/L. HoFH individuals can develop severe CVD and physical findings beginning in childhood; they can also suffer MI as adolescents if untreated (Cuchel et al., 2014). Given that those with FH are exposed to severely elevated LDL-C since birth, it is critical to identify and intensively treat affected individuals as early as possible (Wiegman et al., 2015).

Several lipid-lowering medications are available for treatment of FH, of which statins are the most widely used. Statins inhibit HMG-CoA (5-hydroxy-3-methylglutaryl-coenzyme A) reductase, the rate-limiting enzyme in cholesterol synthesis; this leads to an upregulation of *LDLR* gene expression by SREBP-2 and thus an increased capacity to remove LDL-C from the bloodstream (Endo, 2010). The effectiveness of statins has been supported in a number of long term clinical trials of over 50,000 individuals, shown to both reduce LDL-C levels (25-35%) and CVD risk (24-37%) (Marks et al., 2003; Stroes, 2005). In patients that do not reach the desired LDL-C targets (often 50% reduction from baseline or <2 mmol/L, if possible) on statin monotherapy alone, additional pharmacological agents may be administered. Most often, this includes ezetimibe; ezetimibe acts by decreasing cholesterol absorption in the small intestine (i.e. dietary cholesterol) (Burnett & Huff, 2006). Recently, monoclonal antibody inhibitors of PCSK9 – alirocumab and evolocumab – have become approved for the care of FH patients. Inhibition of PCSK9 reduces LDLR degradation (see **Figure 1.1** above). PCSK9 inhibitors are potent in reducing LDL-C levels, and can be especially effective in patients who have not reached desired LDL-C targets despite maximum dose of statin therapy,

and/or in those who have adverse side effects to statins. In a study of 735 HeFH patients, alirocumab reduced baseline LDL-C levels 50-60% (Kastelein et al., 2015), while evolocumab reduced LDL-C levels by an average of 53.6% in 440 HeFH patients across two clinical trials (Hovingh et al., 2017) .

Using the above prevalence figures, there are an estimated ~34 million individuals affected with FH worldwide. Early diagnosis and timely treatment can normalize life expectancy in most patients with FH (Koch et al., 2012; Versmissen et al., 2008). However, despite the fact that FH is eminently treatable, <10% of those affected worldwide have been diagnosed, and <1% have been diagnosed in most countries (Nordestgaard et al., 2013). FH is thus considered to be severely underdiagnosed and undertreated.

1.4 Is there a need for a genetic diagnosis in FH?

The diagnostic criteria for FH have evolved to reflect our increased understanding of the genetic determinants of high LDL-C. Both the widely used Dutch Lipid Clinic Network (DLCN) and Simon Broome Register (SBR) criteria use scoring systems that weigh variables such as degree of LDL-C elevation, presence of physical signs of cholesterol deposition, and family history of high LDL-C and/or premature coronary heart disease (CHD), and presence of a causative variant in *LDLR*, *APOB*, or *PCSK9*. Depending on a patient's score, the FH diagnosis is ranked as "definite", "probable" or "possible"; importantly a positive DNA result yields the highest number of points, and can be sufficient on its own to invoke a diagnosis of "definite FH". Both sets of criteria are shown in **Table 1.1**.

Table 1.1. Commonly used criteria for the clinical diagnosis of FH.

Simon Broome Register Criteria	Dutch Lipid Clinic Network Criteria
<p><u>(A)</u> Family history of myocardial infarction in 1st degree relative <60 years, or 2nd degree relative <50 years</p>	<p><u>1. Family History</u> (i) 1st degree relative with premature (<55 years men, <60 years women) CHD (1 point) (ii) 1st degree relative with LDL-C >95th percentile by age/gender for country (1 point)</p>
<p><u>(B)</u> Family history of TC >7.5 mmol/L or LDL-C >4.9 mmol/L in 1st or 2nd degree relative</p>	<p>(iii) 1st degree relative with tendon xanthoma and/or premature corneal arcus (2 points) (iv) Children <18 years with LDL-C >95th percentile by age/gender for country (2 points)</p>
<p><u>(C)</u> Adult: TC >7.5 mmol/L or LDL-C >4.9 mmol/L Child (<16 years): TC >6.7 mmol/L or LDL-C >4.0 mmol/L</p>	<p><u>2. Clinical History</u> (i) Premature CHD (2 points) (ii) Premature cerebral or peripheral vascular disease (1 point)</p>
<p><u>(D)</u> Tendon xanthoma in patient or 1st degree relative</p>	<p><u>3. Physical Signs</u> (i) Tendon xanthoma (6 points) (ii) Corneal arcus in person <45 years (4 points)</p>
<p><u>(E)</u> Causative mutation in <i>LDLR</i>, <i>APOB</i>, or <i>PCSK9</i></p>	<p><u>4. LDL-C Level</u> (i) LDL-C >8.5 mmol/L (8 points) (ii) LDL-C 6.5-8.4 mmol/L (5 points) (iii) LDL-C 5.0-6.4 mmol/L (3 points) (iv) LDL-C 4.0-4.9 mmol/L (1 point) <u>5. DNA Analysis</u> (i) Causative variant in <i>LDLR</i>, <i>APOB</i>, or <i>PCSK9</i> (8 points)</p>
<p><u>Diagnosis:</u> Definite FH requires <i>C</i> and <i>D</i>, or <i>E</i> Probable FH requires <i>A</i> and <i>C</i>, or <i>B</i> and <i>C</i></p>	<p><u>Diagnosis:</u> Definite FH > 8 points Probable FH 6-8 points Possible FH 3-5 points</p>

CHD, coronary heart disease.

It is noteworthy that the clinical features, especially the physical findings, have assumed less diagnostic importance in more recent times. Secular trends in diet and lifestyle, and widespread use of lipid-lowering therapies have altered the once "classical" FH presentation (Kindt, Mata, & Knowles, 2017). The physical findings of tendon xanthomas and corneal arcus, which carry weight in both SBR and DLCN criteria scoring (see above), are encountered far less frequently today than 20 years ago. For instance, of the 2752 patients in the SAFEHEART (Spanish Familial Hypercholesterolemia Cohort Study) registry with molecularly defined FH, only 14% had tendon xanthomas (De Isla et al., 2016), similar to findings in US (Degoma et al., 2016) and Canadian FH registries. Improvements in diet, lifestyle and medication use have also potentially delayed or attenuated CHD development in FH families. Furthermore, family history - which also carries weight in both scoring systems - may be unreliable or unascertainable in certain communities or jurisdictions. Thus, a "definite" FH diagnosis by current criteria is becoming increasingly reliant upon the identification of a causal DNA variant. Finally, the DLCN and SBR criteria are frequently used to determine eligibility for clinical trials and for insurance coverage; the increased prominence of genetic diagnosis in these criteria thus has further implications for drug development and coverage for clinical use.

Despite arguments favoring genetic diagnosis in FH, such findings must be interpreted within the entire clinical context for an individual patient or family, including the degree of LDL-C elevation. Nonetheless, documenting a causal variant is increasingly becoming a central element in FH diagnosis.

1.4.1 The importance of cascade screening

Another strategic reason to obtain a genetic diagnosis of FH is to initiate detection of affected family members. Identifying the causal variant in the index case simplifies the screening process, making it more cost-effective, as the specific variant can then be directly genotyped using a less expensive dedicated method (Knowles et al., 2017); the lipid profile can sometimes serve as a surrogate for a DNA test when the latter is not available (Williams et al., 1993). The strategy of ascertaining cases based on their familial relationship with a confirmed index case is called "cascade screening". First,

second, and third degree relatives of an index (He)FH case with a causal variant have, respectively, a 50%, 25%, and 12.5% chance of also having FH. Cascade screening has been repeatedly shown to markedly improve rates of diagnosis and treatment for FH (Hopkins, 2017). The most striking example is from the Netherlands, where cascade screening has led to >28,000 additional cases being identified, improving the rate of diagnosis to an estimated ~70% of the total FH population, which is by far the highest of any country (Nordestgaard et al., 2013). The impact of early diagnosis and treatment with statins of HeFH in the Netherlands bends the survival curve towards a normal trajectory (Koch et al., 2012); cascade screening has played a key role in this successful national prevention strategy.

In Australia and Brazil, cascade screening has also proven to be effective; each index case typically leads to the identification of FH in two affected relatives (Bell et al., 2015; Jannes et al., 2015). In addition, and perhaps most importantly, cascade screening allows for pre-symptomatic diagnosis at a young age. Although family screening can be undertaken using non-genetic phenotypic criteria alone, this approach can be inadequate since a significant proportion of affected relatives have LDL-C levels that fall below the strict diagnostic cut point (often LDL-C >90th percentile) (Leren et al., 2008). However, because their genotype predisposes to higher lifetime exposure to elevated LDL-C and CVD, causal-variant carriers who are related to an index FH case should be considered for lipid-lowering intervention as a preventative measure.

1.4.2 Does knowledge of the genotype impact upon clinical management?

It might be argued that irrespective of genotype, LDL-C level itself is the ultimate determinant of CVD risk and should primarily direct the timing and intensity of intervention (Hopkins, 2017). However, because FH is genetically heterogeneous, there are definitely instances when knowing the genotype can affect treatment decisions. The most definitive examples are HoFH individuals, whose disease course and treatment needs are dramatically different from those with HeFH (Cuchel et al., 2014). Confirming variants affecting both alleles of the *LDLR* or other causative gene has implications for

specific treatment modalities for this condition. Statins and PCSK9 inhibitors both work by increasing the presence of LDLRs on the cell surface; the effectiveness of these treatments in individuals with two (bi-allelic) causal *LDLR* variants will depend on the residual function of the mutated alleles. For instance, individuals who have at least one *LDLR* allele with predicted residual function (>2% function) show up to 25% LDL-C reduction with the PCSK9 inhibitor evolocumab, while individuals who have two null *LDLR* variants (no LDLR functionality) show no response to this treatment (Raal et al., 2015).

Two additional medications - lomitapide and mipomersen - are approved for the care of HoFH patients, which address the limitation of statins and PCSK9 inhibition in those with two mutated *LDLR* alleles (the most frequent cause of HoFH). These drugs lower LDL-C independent of LDLR function, however, have several side-effects and thus are generally only administered in these specific *LDLR* mutant HoFH cases (Gouni-Berthold & Berthold, 2015). LDL apheresis (plasma exchange to remove LDL-C; often every two weeks) is another treatment often necessary for HoFH individuals with two mutant *LDLR* alleles. On the other hand, confirming “double heterozygosity” in a HoFH patient means that at least one wild-type *LDLR* allele is present; these patients are expected to respond well to statins and/or PCSK9 inhibition (Shirahama et al., 2018), and thus, this genotypic information can help avoid any premature application of the aforementioned (often unpleasant) treatments.

Importantly, therapeutic implications from genetic testing can also occur when LALD (bi-allelic variants in *LIPA*) and sitosterolemia (bi-allelic variants in *ACBG5/8*) phenocopy cases are identified. Statins are not advisable in LALD since they burden lysosomes with even more cholesterol than what is already presently toxic in these patients – possibly worsening disease (Bernstein et al., 2013). By the same virtue of increased LDLR presence, use of PCSK9 inhibitors are also unadvised. Treatments for LALD patients are limited, although there is promise in a recently approved recombinant LAL replacement therapy called sebelipase alfa (Burton et al., 2015; Hollak & Hovingh, 2015). For those identified with sitosterolemia, dietary restriction of foods rich in plant sterols (ex. vegetable oils, wheat germs, nuts, seeds, avocado, shortening, margarine, and chocolate) can have drastic effects on lowering LDL-C, and is the

mainstay of therapy (Yoo, 2016). We have experienced this in our laboratory this year: a fellow cardiologist sent a patient's DNA sample to us for sequencing analysis; this was a 18-month old infant believed to be HoFH with a LDL-C of 18 mmol/L and tendon xanthomata. Instead of variants in *LDLR*, *APOB*, *PCSK9*, or *LDLRAP1*, we identified pathogenic compound heterozygous variants in *ABCG5*, and this infant was re-diagnosed with sitosterolemia. A small change in her diet reduced her LDL-C to 9 mmol/L almost immediately.

With respect to HeFH due to a single mutated allele, recent whole-exome sequencing data indicate that among individuals with LDL-C >5 mmol/L, the risk of early CHD rose from 6-fold to 22-fold when a heterozygous disease-causing FH variant was detected (Khera et al., 2016). While the authors suggested that more aggressive treatment is warranted in the latter instance, the clinical difference between a 6- or 22-fold increased risk seems moot: both types of patients with elevated LDL-C should be managed aggressively.

Among HeFH patients, the mutation status largely predicts the degree of LDL-C elevation: heterozygous *LDLR* variants generally underlie a more severe LDL-C phenotype, while LDL-C levels are not quite so elevated with heterozygous *APOB* or *PCSK9* variants, or among individuals with a high polygenic burden (Wang et al., 2016). The latter groups can be more often managed with statin monotherapy alone than individuals with heterozygous *LDLR* variants, who more often would require combination therapy. Further, among *LDLR* variants, LDL-C elevation is greater amongst individuals with CNVs, nonsense and splicing variants (i.e. more deleterious variant types) than those with missense variants, again correlating with the need for more aggressive treatment using combinations or the more potent PCSK9 inhibitors. On the other hand, irrespective of the specific genetic basis, the degree of LDL-C elevation and the empirical response to treatment, which itself varies widely, will help guide the "personalized" treatment plan for the patient.

1.4.3 Other reasons to seek the genetic basis in patients suspected to have FH

Mutation status information may also be important for genetic counselling. Phenotypic expression is highly penetrant in heterozygous carriers of FH-associated variants, especially in children (Sjouke et al., 2016). Although not considered to be appropriate for prenatal screening, because it is so eminently treatable, some families might appreciate knowing the discrete genetic basis for their hypercholesterolemia. A positive genetic test result can also have an impact on the patient's and healthcare provider's attitude; many patients find the information empowering and motivating, especially since much can be done to reduce development of hard cardiovascular clinical end points (Yuan, Wang, & Hegele, 2006). A genotype-based diagnosis rules out a purely environmental cause of high LDL-C, and although improvement of diet and exercise can reduce LDL-C levels (Yu-Poth et al., 1999), knowledge of an underlying genetic contribution suggests to the patient that these interventions alone are not likely to reverse the phenotype. In the Netherlands, those who received a positive test result had lower LDL-C levels compared to those without positive genetic testing, due to improved long-term treatment and lifestyle compliance (Umans-Eckenhausen et al., 2003).

1.5 What genetic testing is available for FH?

1.5.1 DNA hybridization assays

DNA hybridization arrays are designed to test only for the presence of a limited number of specific known disease-causing variants. Typically, such arrays contain the most common causal variants detected within a specific geographical location or jurisdiction; the content is necessarily limited by the selected variants and is frozen based on the knowledge at the time of design and manufacture. Examples of available arrays for FH include: 1) the Elucigene FH20 Array (Tepnel Diagnostics, Abingdon, UK) which screens for the 20 most common FH-causing variants reported in the United Kingdom; 2)

the FH Multiplex Array (Randox Laboratories, Crumlin, UK) which screens for some of the 40 most common FH variants in the UK with a special attention to those within Ireland; and 3) the LIPOchip (Progenika Biopharma, Derio, Spain) designed for detection of 251 common FH variants (242 in *LDLR*, 3 in *APOB* and 6 in *PCSK9*).

Advantages of DNA hybridization arrays include low cost and efficiency, but with caveats. The process requires very little bioinformatics processing and interpretation, with minimum required training for staff, and theoretically a rapid turn-around-time for results. However, if the sample is negative for the particular set of variants embedded in the array, clinicians or diagnostic laboratories would require more comprehensive and unbiased NGS methods to obtain a genetic diagnosis, which could have been sought initially. Furthermore, if new FH variants are introduced into a region through migration, unless the variants are already on the array, they will be missed (false-negatives); the array design would need to be periodically updated. In addition, since arrays cannot detect new variants, they are not useful for variant discovery. Other disadvantages include the lack of flexibility to cover the entire range of small-scale DNA variants such as indels that would be captured by Sanger sequencing or NGS approaches, not to mention inability to globally detect large-scale CNVs, although these could be included in the array design for already known CNVs, or if the sequence surrounding the breakpoint was known. Finally, a positive test for a single variant could be misleading in a compound or double heterozygote whose second variant was not represented on the array.

1.5.2 Sanger sequencing

Between about 1990 and 2015, Sanger re-sequencing of polymerase chain reaction (PCR)-amplified coding regions of the *LDLR* gene (18 exons plus intron-exon boundaries) and specific regions of *APOB* (e.g. exons 26 and 29 encoding receptor binding domains) and more recently the *PCSK9* gene was the most commonly used method for diagnosis and detection of new variants. However, the method is time-consuming and expensive on a per-nucleotide basis. Also, some regions, such as exon 1 of both *LDLR* and *APOB* genes are relatively refractory to reliable Sanger sequencing for

technical reasons. Despite cost and labor-intensiveness, Sanger sequencing is still commonly used by diagnostic laboratories to re-sequence NGS-targeted areas that have low or inadequate technical coverage, to confirm NGS-detected variants, and as a dedicated method to screen family members of an index case for a single specific causative variant.

1.5.3 Next-generation sequencing (NGS) platforms

"Next-generation" refers to the emergence in recent years of DNA sequencing techniques using a variety of new, non-Sanger based chemistries (Farhan & Hegele, 2014). While there are numerous manufacturers in this space, including Roche/454, Life Technologies, Novogene and Complete Genomics, Illumina's platforms have become dominant for both clinical and research applications. Most NGS methods, regardless of their proprietary sequencing chemistry, have three main components: 1) fragmentation of source DNA into millions of small random pieces, which are used to prepare sequencing libraries; 2) amplification and enrichment, usually by immobilizing the DNA, which expedites hundreds of thousands of simultaneous NGS chemical reactions; and 3) detection of the signals from this massive series of sequencing reactions. Given the vast amount of data generated, it is critical to have an integrated and validated bioinformatics pipeline, which assembles millions of overlapping "shotgun" small fragment sequencing "reads" into a string of large contiguous sequence information. The sample sequence is then compared to a reference human genome database, generating a smaller manageable list of sequence variations. Using further prioritization criteria, potentially pathogenic or causative variants related to the phenotype of interest (i.e. FH) are pulled from this list, curated and evaluated, sometimes manually.

There are several NGS approaches to clinical diagnosis of a disorder where the causative gene(s) are already known, such as FH. For instance, the sample can be processed using whole-genome or whole-exome sequencing platforms (WGS or WES, respectively), which generate the entire genomic coding plus non-coding or coding-only DNA sequence, respectively. But since the vast majority of the generated WGS or WES data are not relevant for focused diagnosis of a specific condition like FH, only a tiny

"slice" of the generated data - a few genes - is closely interrogated to detect potentially clinically relevant variants. The remaining 99.9-plus percent of the generated WES or WGS data are masked or discarded. Conceptually, the approach of "slicing" WGS or WES data to find variations in a few candidate genes seems wasteful. In addition to cost considerations, potential ethical issues also arise because incidental findings relevant to many other diseases reside within WES or WGS data. Thus, another approach is to use NGS chemistry and throughput but within a structured targeted sequencing panel, in which reagents are designed only to detect variants from the outset in a limited number of genes known to be relevant to the disease of interest.

The first example of such a targeted NGS panel for lipid disorders, including FH, is called LipidSeq, which is an Illumina-based platform that was developed in our laboratory in 2013 (Johansen et al., 2014). The LipidSeq panel screens coding regions and intron-exon boundaries of 73 lipid metabolism-related genes, including *LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, and *ABCG5/8*. The LipidSeq panel also contains reagents to detect common single nucleotide polymorphisms (SNPs) that have been shown in GWAS studies to be associated with subtle variations in plasma lipids (Kathiresan et al., 2009; Teslovich et al., 2010; Willer et al., 2013), allowing us to construct polygenic risk scores for plasma lipoprotein traits, including LDL-C (Wang et al., 2016). The advantages of a targeted NGS panel over WES or WGS for FH include: 1) lower cost; 2) greater bandwidth, allowing for more samples to be processed simultaneously; 3) greater speed of processing; 4) design that allows for simultaneous detection of monogenic rare large-effect variants and specific polygenic small-effect variants; and 5) minimizing ethical issues related to incidentally detecting disease-causing variants in genes unrelated to lipid metabolism. Finally, it is important to appreciate that targeted NGS can be used in tandem with other sequencing methods. For instance, our laboratory confirms some variants detected on LipidSeq using Sanger sequencing. Also, samples from patients with definite FH by SBR or DLCN criteria that are LipidSeq-negative for rare variants or high polygenic scores can subsequently be evaluated using WES or WGS to detect new disease-causing gene loci and variants.

Since our laboratory's initial report of the feasibility of targeted NGS for lipid disorders (Johansen et al., 2014), several diagnostic laboratories have designed

commercially available NGS panels (**Table 1.2**). In addition to complete analysis of *LDLR* coding regions, NGS panels often include the genetic screening of all coding regions of *APOB* and *PCSK9*. Because causative variants for FH are now being identified outside of the traditionally Sanger-sequenced *APOB* regions (exons 26 and 29) (Alves et al., 2014), and throughout the entire *PCSK9* gene (Dron & Hegele, 2017), the increased capacity of targeted NGS represents an important advance over Sanger sequencing. Furthermore, the flexible design of an NGS-based platform allows for future relevant genes to be included without markedly impacting upon cost. Some providers also extend their panels to include, as in LipidSeq, all minor FH-associated genes, further reducing the chances of false-negative findings.

Table 1.2. Targeted NGS panels commercially available for FH genetic diagnosis.

Service	Primary FH Genes	Additional Genes	Turn Around Time
FH Reflex Panel Mayo Clinic Rochester, MN, USA	<i>LDLR</i> , <i>APOB</i> (p.Arg3527)	None	2 weeks
FH NGS Panel DDC Clinic Middlefield, OH, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i>	4-6 weeks
FH Sequencing Panel Prevention Genetics Marshfield, WI, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i>	4 weeks
FH Panel GeneDX Gaithersburg, MD, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i>	4 weeks
FH Panel Invitae San Francisco, CA, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i>	1-3 weeks
FHNext Panel Amry Genetics Aliso Viejo, CA, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i> , <i>SLCO1B1</i> (SNP c.521T>C)	2-3 weeks
GeneSeq: CAD/FH Profile Integrated Genetics Westborough, MA, USA	<i>LDLR</i> , <i>APOB</i> (556bp of exon 26), <i>PCSK9</i>	<i>ABCA1</i> , <i>APOA2</i> , <i>APOC3</i> , <i>PON2</i> <i>SHOC2</i> (exon 2), <i>AKAP9</i> (exon 18)	6-8 weeks
FH Panel ApolloGen Irvine, CA, USA	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>ABCA</i> , <i>ABCG1/5/8</i> , <i>APOA1</i> , <i>APOE</i> <i>APOC2/3/4</i> , <i>CETP</i> , <i>LCAT</i> , <i>LIPA</i> , <i>LIPC</i> , <i>LPA</i> , <i>LPL</i> , <i>MYLIP</i> , <i>NPC1</i>	4 weeks
SEQPRO LIPO IS Progenika Biopharma San Marcos, TX, USA Derio, Spain	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i> , <i>APOE</i> , <i>STAP1</i>	4-6 weeks
Hypercholesterolemia Panel Sophia Genetics Saint-Sulpice, Switzerland	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i> , <i>APOE</i>	2-3 weeks
FH/Comprehensive Panel HealthInCode A Coruña, Spain	<i>LDLR</i> , <i>APOB</i> , <i>PCSK9</i>	<i>LDLRAP1</i> , <i>APOE</i> , <i>SLCO1B1</i> Comprehensive option includes: <i>ABCG5/8</i> , <i>LIPA</i> , <i>LPA</i> , <i>NPC1L1</i> , <i>ABCB1</i> , <i>AMPD1</i> , <i>CH25H</i> , <i>COQ2</i> , <i>CPT2</i> , <i>CYP2D6</i> , <i>CYP3A/5</i> , <i>PPARA</i> , <i>PYGM</i> , <i>RYR1</i> , <i>SLC22A8</i>	1-3 weeks

Information taken from company websites as part of review performed March 2017; not an exhaustive list.

1.5.4 Determining pathogenicity of identified variants

When any analytic method detects a variant in a FH-related gene, how do we know that the variant is clinically relevant, i.e. disease-causing? This general problem arises from the exponentially growing clinical and research applications of NGS, which can identify dozens or hundreds of rare variants (i.e. allele frequency of <1% in control population databases) per sample analyzed (dependent on the number of genes and regions assessed). The clinical and human genetics community is responding to this challenge by producing guidelines and position statements about potential pathogenicity (Green et al., 2016). An important element is the creation and maintenance of reliable databases of clinically relevant variants, of which ClinVar at the National Center for Biotechnology Information (NCBI) is becoming widely accepted (Landrum et al., 2014).

Imputing a causal role for any particular variant in FH is not an established science; it is an evolving field. When a rare DNA variant is detected in a sample from a patient with suspected FH, its "first stop" would be a comparison against entries in publicly available databases, to determine whether it has been previously reported and to gauge any additional evidence which may support or refute causality.

The "gold standard" or most undisputable form of evidence supporting a variant's pathogenicity is functional evidence demonstrating a biochemical effect of the variant on LDL-C homeostasis, often *in vitro*. However, at most ~15% of all reported *LDLR* variants have been studied functionally (Bourbon et al., 2017). A general issue in human genetics is the need to develop technologies that will scale up functional analyses of new variants in real time. At present, functional analysis is not practical in a clinical diagnostic setting. The next highest level of certainty for functional relevance involves major deleterious variants in *LDLR*, such as nonsense and frameshift variants introducing a premature stop codon, and large-scale deletions. These are generally considered to have a deleterious effect on *LDLR* activity and functional studies are often not required. Classification of CNVs or nonsense variants as "pathogenic" has traditionally been warranted without additional evidence.

The majority of variants reported as "causative" for FH do not belong to the above categories; thus a series of further analyses could help clarify their classification. These

include assessing variant frequency in control population databases, providing evidence of co-segregation of variant and affected status in families, and use of various *in silico* tools to predict impact on protein function. However, such interpretation efforts for any particular variant are often rudimentary and differ from one diagnostic laboratory to the next, leading to inaccuracies and inconsistencies in pathogenicity classification – an issue documented within the clinical genetics field (Harrison et al., 2017). In 2015, the American College of Medical Genetics and Association of Molecular Pathologists (ACMG/AMP) addressed this by publishing a set of consensus guidelines for standardized variant interpretation in Mendelian disorders (Richards et al., 2015). Within the ACMG/AMP framework, an algorithm-based scoring method classifies variants within a 5-tier system as either "pathogenic", "likely pathogenic", "variant of unknown significance (VUS)", "likely benign", or "benign" based on 28 different evidence types with varying weightings. This formal structure facilitates a more critical and consistent assessment of evidence for or against pathogenicity (Richards et al., 2015). Thus, by today's standards it is becoming more commonly suggested that only those variants classified as "pathogenic" or "likely pathogenic" by these standardized guidelines be used to confirm a clinical diagnosis of FH.

1.5.5 CNV analysis in the diagnosis of FH

Rare CNVs causing a loss of function in *LDLR* have long been recognized as a significant class of FH variants. Revisiting the pioneering work of Nobel Laureates Joseph Goldstein and Michael Brown shows that *LDLR* structural rearrangements - that we would now term CNVs - were a prevalent variant-type in some of their earliest characterized FH patients and families (Hobbs, Brown, & Goldstein, 1992), due to the use of Southern blotting as the main technology to detect DNA variation before 1988. Southern blotting was ideally suited to detect DNA fragment sizes ranging between 0.5 and ~18 kilobases, so aberrant patterns of *LDLR* gene bands in this size range due to a CNV in an affected patient were readily detectable. Some early reported *LDLR* CNVs were founder-effect variants, contributing to a high proportion of FH in distinct

populations, such as the 15-kilobase deletion spanning the promoter and exon 1, seen in 60% of French-Canadians with FH (Hobbs et al., 1987), or the 9.5-kilobase deletion spanning exons 16-18, seen in 30-40% of Finns with FH (Aalto-Setälä et al., 1987; 1992).

PCR-based methods coupled with automated Sanger sequencing became the state-of-the-art methodology to identify DNA variation around 1990. These robust tools were optimized to detect small-scale changes. Over the following 15 years, almost all reported *LDLR* variants in FH were thus rare single nucleotide variants - underlying missense, nonsense, splicing - and small indel variants resolvable by Sanger sequencing. When a large-scale CNV was present, it was not observable, analogous to ascertaining a canyon using a magnifying glass that was optimized to detect flaws on stone's surface. If an exon or exons of a mutated *LDLR* allele were missing, only the corresponding region on the normal allele remained; this would be read as normal sequence qualitatively, but there was no way to know that the read-out was based on a hemizygous genomic substrate rather than two normal diploid alleles.

This limitation was addressed around 2003, with use of a technology designed to determine DNA-dosage alterations by hybridizing specialized DNA probes directed against specific regions, namely exons of the *LDLR*. This method, multiplex ligation-dependent probe amplification (MLPA), is well-suited for *LDLR* analysis due to its dedicated exon-by-exon level resolution, relevant to the range of CNV events known to disrupt the *LDLR* gene in FH (see **Figure 1.3** above). MLPA, which has been proven ultra-reliable for over a decade, is considered the gold standard for CNV detection in *LDLR*. Because ~10% of patients with FH in various cohorts were found to have CNVs detectable by MLPA, it was necessary to set up two different methods - Sanger sequencing plus MLPA - to maximize the variant detection rate

However, because MLPA requires additional infrastructure, reagents, expertise, and analysis time, it has always been a goal to find a unified method that could both detect small-scale sequence variants - i.e. single base pair variants and small indels - and large-scale CNVs from a single chemical reaction and analysis platform, for instance, from NGS results.

1.5.6 Analysis of polygenic elevated LDL cholesterol

The presence of multiple common LDL-C raising alleles is another possible cause of hypercholesterolemia. Because 20-40% of individuals with phenotypic HeFH have no identifiable monogenic variants, incorporating polygenic analysis into the routine testing strategy has significant potential to assist in genetically defining cases. Polygenic testing in FH is done with the construction of a genetic risk score (GRS, or sometimes referred to as a polygenic trait score; PTS). Fundamentally, a GRS summarizes information across multiple SNPs into a single, predictive score. The simplest iteration sums the number of risk-alleles found at each selected trait-altering locus, either 0, 1, or 2, for a non-risk homozygous state, a risk heterozygous state, or a risk homozygous state, respectively. In this way, its calculation assumes an identical directional effect conferred by each risk allele included in the score. For complex traits such as LDL-C, however, effect sizes vary across significantly associated GWAS SNP loci (Teslovich et al., 2010; Willer et al., 2013), thus assigning each loci these weighted values is thought to improve the overall score's predictive ability.

At present, there have been multiple reports of GRSs for LDL-C in cohorts with suspected FH (Futema et al., 2014, 2015; Sjouke et al., 2016; Talmud et al., 2013; Wang et al., 2016). However, there is no consensus regarding the optimal set of SNP markers to include in such a score; different reports have used different SNPs, so the results are not directly comparable. The individual SNPs used in particular reported GRSs are shown in **Table 1.3**.

Table 1.3. SNPs used in reported genetic risk scores for hypercholesterolemia.

SNP	Locus	Risk Allele	GLGC Effect Size (mmol/L)			Talmud <i>et al.</i> (2013)	Futema <i>et al.</i> (2016)	Wang <i>et al.</i> (2016)	Sjouke <i>et al.</i> (2016)
			2008	2010	2013				
rs6511720	<i>LDLR</i>	G	0.26	0.18	0.22	x	x	x	x
rs629301	<i>SORT1</i>	T	-	0.15	0.17	x	x		x
rs1367117	<i>APOB</i>	A	-	0.10	0.12	x	x		x
rs10401969	<i>NCAN/CILP2</i>	T	0.05	0.12	0.12			x	
rs4299376	<i>ABCG8</i>	G	-	0.07	0.08	x			x
rs3846663/rs12916	<i>HMGCR</i>	T/C	0.07	0.06	0.07			x	
rs1800562	<i>HFE</i>	G	-	0.06	0.06	x			x
rs2479409	<i>PCSK9</i>	G	-	0.05	0.06	x	x		x
rs11220462	<i>ST3GAL4</i>	A	-	0.05	0.06	x			x
rs1501908/rs6882076	<i>TIMD4</i>	C/C	0.07	0.04	0.05			x	
rs1564348	<i>LPA</i>	T	-	0.01	0.05	x			x
rs2650000/rs1169288	<i>HNFI1A</i>	A/C	0.07	0.04	0.04			x	
rs3757354	<i>MYLIP</i>	C	-	0.04	0.04	x			x
rs8017377/rs2332328	<i>NYNRIN</i>	T/T	-	0.03	0.04	x			x
rs12740374	<i>SORT1</i>	G	0.23	-	-			x	
rs515135	<i>APOB</i>	C	0.16	-	-			x	
rs6544713	<i>ABCG8</i>	T	0.15	-	-			x	
rs11206510	<i>PCSK9</i>	T	0.09	-	-			x	
rs6102059	<i>MAFB</i>	C	0.06	-	-			x	
rs429358 rs7412	* <i>APOE</i>	ε2ε2	-0.90						
		ε2ε3	-0.40						
		ε2ε4	-0.20			x	x		x
		ε3ε3	0						
		ε3ε4	0.10						
		ε4ε4	0.20						

SNPs and effect sizes have been reported by the Global Lipids Genetics Consortium (GLGC), in three successive reports (Kathiresan *et al.*, 2009; Teslovich *et al.*, 2010; Willer *et al.*, 2013) meta-analyzing GWAS for lipid-associated traits. SNPs denoted by “/” are proxies (Teslovich *et al.*, 2010). **APOE* SNPs and effect sizes were reported in (Bennet *et al.*, 2007), and were not identified in the GLGC reports.

1.6 Synopsis

FH is extremely prevalent, yet severely underdiagnosed – despite being eminently treatable. Recently, FH has progressed toward the forefront of precision medicine as patients worldwide are increasingly offered genetic testing as a central part of diagnosis. Advantages of providing a genetic diagnosis for FH are manifold. They include: 1) achieving certainty in the context of incomplete clinical criteria, such as family history or typical physical findings; 2) motivating cascade screening and thus early diagnosis and preventative treatment in affected family members; 3) initiating genotype-directed treatment strategies; 4) improvement of treatment compliance; and 5) supporting insurance coverage of certain medications.

Genetic testing in FH is recommended by the Canadian Cardiovascular Society (Brunham et al., 2018), the US Centers for Disease Control Office of Public Health Genomics (Gidding et al., 2015), the United Kingdom National Institutes for Clinical Excellence (NICE), both the European and International Atherosclerosis Societies, and an international expert panel convened by the FH Foundation and American College of Cardiology (Sturm et al., 2018). The American College of Medical Genetics and Genomics also includes *LDLR*, *APOB*, and *PCSK9* among the list of 56 “medically actionable” genes for which sequence analysis can lead to direction of treatment (Green et al., 2013). Given both its prominence and importance, my thesis focuses on improving the strategies currently used to provide a genetic diagnosis of FH.

1.7 References

- Aalto-Setälä, K., Helve, E., Kovanen, P. T., & Kontula, K. (1989). Finnish type of low density lipoprotein receptor gene mutation (FH-Helsinki) deletes exons encoding the carboxy-terminal part of the receptor and creates an internalization-defective phenotype. *Journal of Clinical Investigation*, *84*(2), 499–505.
- Aalto-Setälä, K., Koivisto, U. M., Mietinen, T. A., Gylling, H., Kesäniemi, Y. A., Savolainen, M., ... Kontula, K. (1992). Prevalence and geographical distribution of major LDL receptor gene rearrangements in Finland. *Journal of Internal Medicine*, *231*(3), 227–234.
- Akiyama, L. E., Genest, J., Shan, S. D., Reel, R. L., Albaum, J. M., Chu, A., & Tu, J. V. (2017). Estimating the prevalence of heterozygous familial hypercholesterolemia: A systematic review and meta-analysis. *BMJ Open*, *7*(9), e016461.
- Alves, A. C., Atar, A., Etxebarria, A., Soutar, A. K., Atherine, Martin, C., & Bourbon, M. (2014). Novel functional APOB mutations outside LDL-binding region causing familial hypercholesterolemia. *Human Molecular Genetics*, *23*(7), 1817–1828.
- Amsellem, S., Briffaut, D., Carrié, A., Rabés, J. P., Girardet, J. P., Fredenrich, A., ... Benlian, P. (2002). Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Human Genetics*, *111*(6), 501–510.
- Anderson, R. G. W., Brown, M. S., & Goldstein, J. L. (1977). Role of the coated endocytic vesicle in the uptake of receptor-bound low density lipoprotein in human fibroblasts. *Cell*, *10*(3), 351–364.
- Austin, M. A., Hutter, C. M., Zimmern, R. L., & Humphries, S. E. (2004). Genetic causes of monogenic heterozygous familial hypercholesterolemia: A HuGE prevalence review. *American Journal of Epidemiology*, *160*(5), 407–420.
- Bell, D. A., Pang, J., Burrows, S., Bates, T. R., van Bockxmeer, F. M., Hooper, A. J., ... Watts, G. F. (2015). Effectiveness of genetic cascade screening for familial hypercholesterolemia using a centrally co-ordinated clinical service: An Australian experience. *Atherosclerosis*, *239*(1), 93–100.
- Bennet, A. M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., ... Danesh, J. (2007). Association of apolipoprotein E genotypes with lipid levels and coronary risk. *Journal of the American Medical Association*, *298*(11), 1300–1311.
- Berberich, A. J., & Hegele, R. A. (2018). The complex molecular genetics of familial hypercholesterolemia. *Nature Reviews Cardiology*, 1–12.
- Berenson, G. S., Srinivasan, S. R., Bao, W., Newman, W. P., Tracy, R. E., & Wattigney, W. A. (1998). Association between Multiple Cardiovascular Risk Factors and Atherosclerosis in Children and Young Adults. *New England Journal of Medicine*,

338(23), 1650–1656.

- Bernstein, D. L., Hülkova, H., Bialer, M. G., & Desnick, R. J. (2013). Cholesteryl ester storage disease: Review of the findings in 135 reported patients with an underdiagnosed disease. *Journal of Hepatology*, *58*(6), 1230–1243.
- Bertolini, S., Pisciotto, L., Fasano, T., Rabacchi, C., & Calandra, S. (2017). The study of familial hypercholesterolemia in Italy: A narrative review. *Atherosclerosis Supplements*, *29*, 1–10.
- Bourbon, M., Alves, A. C., & Sijbrands, E. J. (2017). Low-density lipoprotein receptor mutational analysis in diagnosis of familial hypercholesterolemia. *Current Opinion in Lipidology*, *28*(2), 120–129.
- Brown, M. S., Anderson, R. G. W., & Goldstein, J. L. (1983). Recycling receptors: The round-trip itinerary of migrant membrane proteins. *Cell*, *32*(3), 663–667.
- Brown, M. S., & Goldstein, J. L. (1979). Receptor-mediated endocytosis: insights from the lipoprotein receptor system. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(7), 3330–7.
- Brown, M. S., & Goldstein, J. L. (1986). A Receptor-Mediated Pathway for Cholesterol Homeostasis (Nobel Lecture). *Angewandte Chemie International Edition in English*, *25*(7), 583–602.
- Brunham, L. R., Ruel, I., Aljenedil, S., Rivière, J.-B., Baass, A., Tu, J. V., ... Brophy, J. (2018). Canadian Cardiovascular Society Position Statement on Familial Hypercholesterolemia: Update 2018. *Canadian Journal of Cardiology*, *34*(12), 1553–1563.
- Burnett, J. R., & Huff, M. W. (2006). Cholesterol absorption inhibitors as a therapeutic option for hypercholesterolaemia. *Expert Opinion on Investigational Drugs*, *15*(11), 1337–1351.
- Burton, B. K., Balwani, M., Feillet, F., Barić, I., Burrow, T. A., Camarena Grande, C., ... Quinn, A. G. (2015). A Phase 3 Trial of Sebelipase Alfa in Lysosomal Acid Lipase Deficiency. *New England Journal of Medicine*, *373*(11), 1010–1020.
- Burton, B. K., Deegan, P. B., Enns, G. M., Guardamagna, O., Horslen, S., Hovingh, G. K., ... Quinn, A. G. (2015). Clinical Features of Lysosomal Acid Lipase Deficiency. *Journal of Pediatric Gastroenterology and Nutrition*, *61*(6), 619–625.
- Cenarro, A., Etxebarria, A., de Castro-Orós, I., Stef, M., Bea, A. M., Palacios, L., ... Civeira, F. (2016). The p.Leu167del Mutation in *APOE* Gene Causes Autosomal Dominant Hypercholesterolemia by Down-regulation of LDL Receptor Expression in Hepatocytes. *The Journal of Clinical Endocrinology & Metabolism*, *101*(5), 2113–2121.
- Chatterton, J. E., Phillips, M. L., Curtiss, L. K., Milne, R., Fruchart, J. C., & Schumaker, V. N. (1995). Immunoelectron microscopy of low density lipoproteins yields a ribbon and bow model for the conformation of apolipoprotein B on the lipoprotein

- surface. *Journal of Lipid Research*, 36(9), 2027–37.
- Chen, W. J., Goldstein, J. L., & Brown, M. S. (1990). NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *Journal of Biological Chemistry*, 265(6), 3116–3123.
- Chora, J. R., Alves, A. C., Medeiros, A. M., Mariano, C., Lobarinhas, G., Guerra, A., ... Bourbon, M. (2017). Lysosomal acid lipase deficiency: A hidden disease among cohorts of familial hypercholesterolemia? *Journal of Clinical Lipidology*, 11(2), 477–484.e2.
- Chora, J. R., Medeiros, A. M., Alves, A. C., & Bourbon, M. (2018). Analysis of publicly available LDLR, APOB, and PCSK9 variants associated with familial hypercholesterolemia: Application of ACMG guidelines and implications for familial hypercholesterolemia diagnosis. *Genetics in Medicine*, 20(6), 591–598.
- Cuchel, M., Bruckert, E., Ginsberg, H. N., Raal, F. J., Santos, R. D., Hegele, R. A., ... Chapman, M. J. (2014). Homozygous familial hypercholesterolaemia: New insights and guidance for clinicians to improve detection and clinical management. A position paper from the Consensus Panel on Familial Hypercholesterolaemia of the European Atherosclerosis Society. *European Heart Journal*, 35(32), 2146–2157.
- De Isla, L. P., Alonso, R., Mata, N., Saltijeral, A., Muñiz, O., Rubio-Marin, P., ... Mata, P. (2016). Coronary heart disease, peripheral arterial disease, and stroke in familial hypercholesterolaemia: Insights from the SAFEHEART registry (Spanish familial hypercholesterolaemia cohort study). *Arteriosclerosis, Thrombosis, and Vascular Biology*, 36(9), 2004–2010.
- Degoma, E. M., Ahmad, Z. S., O'Brien, E. C., Kindt, I., Shrader, P., Newman, C. B., ... Knowles, J. W. (2016). Treatment gaps in adults with heterozygous familial hypercholesterolemia in the United States. *Circulation: Cardiovascular Genetics*, 9(3), 240–249.
- Dron, J. S., & Hegele, R. A. (2017). Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Current Opinion in Lipidology*, 28(2), 161–169.
- Endo, A. (2010). A historical perspective on the discovery of statins. *Proceedings of the Japan Academy, Series B*, 86(5), 484–493.
- Farhan, S. M. K., & Hegele, R. A. (2014a). Exome sequencing: New insights into lipoprotein disorders. *Current Cardiology Reports*, 16(7), 507.
- Farhan, S. M. K., & Hegele, R. A. (2014b). Sequencing: The next generation-what is the role of whole-exome sequencing in the diagnosis of familial cardiovascular diseases. *Canadian Journal of Cardiology*, 30(2), 152–154.
- Futema, M., Plagno, V., Li, K. W., Whittall, R. A., Neil, H. A. W., Seed, M., ... Humphries, S. E. (2014). Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. *Journal of Medical Genetics*, 51(8), 537–544.

- Futema, M., Shah, S., Cooper, J. A., Li, K., Whittall, R. A., Sharifi, M., ... Humphries, S. E. (2015). Refinement of variant selection for the LDL cholesterol genetic risk score in the diagnosis of the polygenic form of clinical familial hypercholesterolemia and replication in samples from 6 countries. *Clinical Chemistry*, *61*(1), 231–238.
- Garcia, C. K., Wilund, K., Arca, M., Zuliani, G., Fellin, R., Maioli, M., ... Hobbs, H. H. (2001). Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science*, *292*(5520), 1394–1398.
- Gidding, S. S., Champagne, M. A., De Ferranti, S. D., Defesche, J., Ito, M. K., Knowles, J. W., ... Wierzbicki, A. S. (2015). The Agenda for Familial Hypercholesterolemia: A Scientific Statement from the American Heart Association. *Circulation*, *132*(22), 2167–2192.
- Goldmann, R., Tichý, L., Freiburger, T., Zapletalová, P., Letocha, O., Soška, V., ... Fajkusová, L. (2010). Genomic characterization of large rearrangements of the LDLR gene in Czech patients with familial hypercholesterolemia. *BMC Medical Genetics*, *11*(1), 115.
- Goldstein, J. L., Dana, S. E., Faust, J. R., Beaudet, A. L., & Brown, M. S. (1975). Role of lysosomal acid lipase in the metabolism of plasma low density lipoprotein. Observations in cultured fibroblasts from a patient with cholesteryl ester storage disease. *Journal of Biological Chemistry*, *250*(21), 8487–8495.
- Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., ... American College of Medical Genetics and Genomics. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine*, *15*(7), 565–574.
- Green, R. C., Goddard, K. A. B., Jarvik, G. P., Amendola, L. M., Appelbaum, P. S., Berg, J. S., ... Zikmund-Fisher, B. J. (2016). Clinical Sequencing Exploratory Research Consortium: Accelerating Evidence-Based Practice of Genomic Medicine. *American Journal of Human Genetics*, *98*(6), 1051–1066.
- Harrison, S. M., Dolinsky, J. S., Knight Johnson, A. E., Pesaran, T., Azzariti, D. R., Bale, S., ... Rehm, H. L. (2017). Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. *Genetics in Medicine*, *19*(10), 1096–1104.
- Hegele, R. A. (1997). Candidate genes, small effects, and the prediction of atherosclerosis. *Critical Reviews in Clinical Laboratory Sciences*, *34*(4), 343–367.
- Hegele, R. A. (2009). Plasma lipoproteins: Genetic influences and clinical implications. *Nature Reviews Genetics*, *10*(2), 109–121.
- Hobbs, H. H., Brown, M. S., & Goldstein, J. L. (1992). Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Human Mutation*, *1*(6), 445–466.
- Hobbs, H. H., Brown, M. S., Russell, D. W., Davignon, J., & Goldstein, J. L. (1987). Deletion in the gene for the low-density-lipoprotein receptor in a majority of French Canadians with familial hypercholesterolemia. *New England Journal of Medicine*,

317(12), 734–737.

- Hollak, C. E. M., & Hovingh, G. K. (2015). Lysosomal acid lipase deficiency—a cautious leap forward. *Nature Reviews Endocrinology*, 11(12), 696–697.
- Hopkins, P. N. (2017). Genotype-guided diagnosis in familial hypercholesterolemia: Population burden and cascade screening. *Current Opinion in Lipidology*, 28(2), 136–143.
- Horsthemke, B., Beisiegel U., Dunning, A., Havinga, J. R. (1987). Unequal crossing-over between two alu-repetitive DNA sequences in the low-density-lipoprotein-receptor gene: A possible mechanism for the defect in a patient with familial hypercholesterolaemia. *European Journal of Biochemistry*, 164(1), 77–81.
- Horton, J. D., Goldstein, J. L., & Brown, M. S. (2002). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation*, 109(9), 1125–1131.
- Hovingh, G. K., Raal, F. J., Dent, R., Stefanutti, C., Descamps, O., Masana, L., ... Sullivan, D. (2017). Long-term safety, tolerability, and efficacy of evolocumab in patients with heterozygous familial hypercholesterolemia. *Journal of Clinical Lipidology*, 11(6), 1448–1457.
- Humphries, S. E., Whittall, R. A., Hubbart, C. S., Maplebeck, S., Cooper, J. A., Soutar, A. K., ... Neil, H. (2006). Genetic causes of familial hypercholesterolaemia in patients in the UK: Relation to plasma lipid levels and coronary heart disease risk. *Journal of Medical Genetics*, 43(12), 943–949.
- Iacocca, M. A., & Hegele, R. A. (2018). Role of DNA copy number variation in dyslipidemias. *Current Opinion in Lipidology*, 29(2), 125–132.
- Innerarity, T. L., Weisgraber, K. H., Arnold, K. S., Mahley, R. W., Krauss, R. M., Vega, G. L., & Grundy, S. M. (1987). Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. *Proceedings of the National Academy of Sciences of the United States of America*, 84(19), 6919–23.
- Jannes, C. E., Santos, R. D., de Souza Silva, P. R., Turolla, L., Gagliardi, A. C. M. M., Marsiglia, J. D. C. C., ... Pereira, A. C. (2015). Familial hypercholesterolemia in Brazil: Cascade screening program, clinical and genetic aspects. *Atherosclerosis*, 238(1), 101–107.
- Johansen, C. T., Dubé, J. B., Loyzer, M. N., MacDonald, A., Carter, D. E., McIntyre, A. D., ... Hegele, R. A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *Journal of Lipid Research*, 55(4), 765–772.
- Kastelein, J. J. P., Ginsberg, H. N., Langslet, G., Kees Hovingh, G., Ceska, R., Dufour, R., ... Farnier, M. (2015). ODYSSEY FH i and FH II: 78 week results with alirocumab treatment in 735 patients with heterozygous familial hypercholesterolaemia. *European Heart Journal*, 36(43), 2996–3003.
- Kathiresan, S., Willer, C. J., Peloso, G. M., Demissie, S., Musunuru, K., Schadt, E. E., ...

- Cupples, L. A. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*, *41*(1), 56–65.
- Khera, A. V., Won, H. H., Peloso, G. M., Lawson, K. S., Bartz, T. M., Deng, X., ... Kathiresan, S. (2016). Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *Journal of the American College of Cardiology*, *67*(22), 2578–2589.
- Knowles, J. W., Rader, D. J., & Khoury, M. J. (2017). Cascade screening for familial hypercholesterolemia and the use of genetic testing. *JAMA - Journal of the American Medical Association*, *318*(4), 381–382.
- Koch, S. A., Gilder, S. A., Pohl, J., & Trepmann, C. (2012). Geomagnetic field intensity recorded after impact in the Ries meteorite crater, Germany. *Geophysical Journal International*, *189*(1), 383–390.
- Kriško, A., & Etchebest, C. (2007). Theoretical model of human apolipoprotein B100 tertiary structure. *Proteins: Structure, Function and Genetics*, *66*(2), 342–358.
- Lagor, W. R., & Millar, J. S. (2010). Overview of the LDL receptor: Relevance to cholesterol metabolism and future approaches for the treatment of coronary heart disease. *Journal of Receptor, Ligand and Channel Research*, *3*, 1–14.
- Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(D1), D980-5.
- Lehrman, M. A., Schneider, W. J., Südhof, T. C., Brown, M. S., Goldstein, J. L., & Russell, D. W. (1985). Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science*, *227*(4683), 140–146.
- Leren, T. P., Finborud, T. H., Manshaus, T. E., Ose, L., & Berge, K. E. (2008). Diagnosis of familial hypercholesterolemia in general practice using clinical diagnostic criteria or genetic testing as part of cascade genetic screening. *Community Genetics*, *11*(1), 26–35.
- Marks, D., Thorogood, M., Neil, H. A. W., & Humphries, S. E. (2003). A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. *Atherosclerosis*, *168*(1), 1–14.
- Masuhara, M., Nagao, K., Nishikawa, M., Sasaki, M., Yoshimura, A., & Osawa, M. (2000). Molecular cloning of murine STAP-1, the stem-cell-specific adaptor protein containing PH and SH2 domains. *Biochemical and Biophysical Research Communications*, *268*(3), 697–703.
- Miyake, Y., Yamamura, T., Sakai, N., Miyata, T., Kokubo, Y., & Yamamoto, A. (2009). Update of Japanese common LDLR gene mutations and their phenotypes: Mild type mutation L547V might predominate in the Japanese population. *Atherosclerosis*, *203*(1), 153–160.
- Nordestgaard, B. G., Chapman, M. J., Humphries, S. E., Ginsberg, H. N., Masana, L.,

- Descamps, O. S., ... Tybjærg-Hansen, A. (2013). Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: Guidance for clinicians to prevent coronary heart disease. *European Heart Journal*, *34*(45), 3478–3490.
- Raal, F. J., Honarpour, N., Blom, D. J., Hovingh, G. K., Xu, F., Scott, R., ... Stein, E. A. (2015). Inhibition of PCSK9 with evolocumab in homozygous familial hypercholesterolaemia (TESLA Part B): A randomised, double-blind, placebo-controlled trial. *The Lancet*, *385*(9965), 341–350.
- Rader, D. J., Cohen, J., & Hobbs, H. H. (2003). Monogenic hypercholesterolemia: New insights in pathogenesis and treatment. *Journal of Clinical Investigation*, *111*(12), 1795–1803.
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., ... Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444–454.
- Reiner, Ž., Guardamagna, O., Nair, D., Soran, H., Hovingh, K., Bertolini, S., ... Ros, E. (2014). Lysosomal acid lipase deficiency – An under-recognized cause of dyslipidaemia and liver dysfunction. *Atherosclerosis*, *235*(1), 21–30.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Rehm, H. L. (2015). 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. *Genetics in Medicine*, *17*(5), 405–424.
- Sharrett, A. R., Ballantyne, C. M., Coady, S. A., Heiss, G., Sorlie, P. D., Catellier, D., & Patsch, W. (2001). Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins A-I and B, and HDL density subfractions: The Atherosclerosis Risk in Communities (ARIC) Study. *Circulation*, *104*(10), 1108–1113.
- Sjouke, B., Defesche, J. C., de Randamie, J. S. E., Wiegman, A., Fouchier, S. W., & Hovingh, G. K. (2016). Sequencing for LIPA mutations in patients with a clinical diagnosis of familial hypercholesterolemia. *Atherosclerosis*, *251*, 263–265.
- Sjouke, B., Kusters, D. M., Kindt, I., Besseling, J., Defesche, J. C., Sijbrands, E. J. G., ... Hovingh, G. K. (2015). Homozygous autosomal dominant hypercholesterolaemia in the Netherlands: Prevalence, genotype-phenotype relationship, and clinical outcome. *European Heart Journal*, *36*(9), 560–565.
- Sjouke, B., Tanck, M. W. T., Fouchier, S. W., Defesche, J. C., Hutten, B. A., Wiegman, A., ... Hovingh, G. K. (2016). Children with hypercholesterolemia of unknown cause: Value of genetic risk scores. *Journal of Clinical Lipidology*, *10*(4), 851–859.
- Soutar, A. K., & Naoumova, R. P. (2007). Mechanisms of disease: Genetic causes of familial hypercholesterolemia. *Nature Clinical Practice Cardiovascular Medicine*, *4*(4), 214–225.
- Stroes, E. (2005). Statins and LDL-cholesterol lowering: an overview. *Current Medical*

Research and Opinion, 21(sup6), S9–S16.

- Sturm, A. C., Knowles, J. W., Gidding, S. S., Ahmad, Z. S., Ahmed, C. D., Ballantyne, C. M., ... Foundation, C. by the F. H. (2018). Clinical Genetic Testing for Familial Hypercholesterolemia. *Journal of the American College of Cardiology*, 72(6), 662–680.
- Tada, H., Nomura, A., Yamagishi, M., & Kawashiri, M. (2018). First case of sitosterolemia caused by double heterozygous mutations in ABCG5 and ABCG8 genes. *Journal of Clinical Lipidology*, 12(5), 1164–1168.e4.
- Talmud, P. J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J. A., ... Humphries, S. E. (2013). Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: A case-control study. *The Lancet*, 381(9874), 1293–1301.
- Taylor, A., Martin, B., Wang, D., Patel, K., Humphries, S. E., & Norbury, G. (2009). Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia. *Clinical Genetics*, 76(1), 69–75.
- Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., ... Kathiresan, S. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466(7307), 707–713.
- Umans-Eckenhausen, M. A. W., Defesche, J. C., Van Dam, M. J., & Kastelein, J. J. P. (2003). Long-term compliance with lipid-lowering medication after genetic screening for familial hypercholesterolemia. *Archives of Internal Medicine*, 163(1), 65–68.
- Versmissen, J., Oosterveer, D. M., Yazdanpanah, M., Defesche, J. C., Basart, D. C. G., Liem, A. H., ... Sijbrands, E. J. G. (2008). Efficacy of statins in familial hypercholesterolaemia: a long term cohort study. *BMJ*, 337(nov11 1), a2423–a2423.
- Wang, J., Dron, J. S., Ban, M. R., Robinson, J. F., McIntyre, A. D., Alazzam, M., ... Hegele, R. A. (2016). Polygenic Versus Monogenic Causes of Hypercholesterolemia Ascertained Clinically. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 36(12), 2439–2445.
- Wiegman, A., Gidding, S. S., Watts, G. F., Chapman, M. J., Ginsberg, H. N., Cuchel, M., ... Watts, G. F. (2015). Familial hypercholesterolaemia in children and adolescents: Gaining decades of life by optimizing detection and treatment. *European Heart Journal*, 36(36), 2425–2437.
- Wijers, M., Kuivenhoven, J. A., & Van De Sluis, B. (2015). The life cycle of the low-density lipoprotein receptor: Insights from cellular and in-vivo studies. *Current Opinion in Lipidology*, 26(2), 82–87.
- Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., ... Abecasis, G. R. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, 45(11), 1274–1285.

- Williams, R. R., Hunt, S. C., Schumacher, M. C., Hegele, R. A., Leppert, M. F., Ludwig, E. H., & Hopkins, P. N. (1993). Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics. *The American Journal of Cardiology*, *72*(2), 171–176.
- Yu-Poth, S., Zhao, G., Etherton, T., Naglak, M., Jonnalagadda, S., & Kris-Etherton, P. M. (1999). Effects of the National Cholesterol Education Program's Step I and Step II dietary intervention programs on cardiovascular disease risk factors: A meta-analysis. *American Journal of Clinical Nutrition*, *69*(4), 632–646.
- Yuan, G., Wang, J., & Hegele, R. A. (2006). Heterozygous familial hypercholesterolemia: An underrecognized cause of early cardiovascular disease. *CMAJ*, *174*(8), 1124–1129.
- Zhang, D. W., Lagace, T. A., Garuti, R., Zhao, Z., McDonald, M., Horton, J. D., ... Hobbs, H. H. (2007). Binding of proprotein convertase subtilisin/kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation. *Journal of Biological Chemistry*, *282*(25), 18602–18612.

Chapter 2 – Use of next-generation sequencing to detect *LDLR* gene copy number variation in the genetic diagnosis of FH

Chapter 2 was adapted from the independent study “Use of next-generation sequencing to detect *LDLR* gene copy number variation in familial hypercholesterolemia”, published in the *Journal of Lipid Research* in 2017 (Iacocca et al., 2017; PMID: 28874442).

2.1 Introduction

Successful genetic diagnosis depends on the ability of the designated method to assess both locus and allele heterogeneity associated with FH. The cost-effectiveness of such methods may limit their widespread implementation and routine use. Currently, the procedure employed in “best practice” diagnostic laboratories is 1) the use of targeted next-generation sequencing (NGS) panels for the detection of small-scale DNA variants - i.e. single base pair variants and small indels – which commonly are designed to assess all coding and intronic splice-site regions in *LDLR*, *APOB*, and *PCSK9* (and less commonly also in *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, *ABCG5/8*), followed by 2) the use of multiplex-ligation dependent probe amplification (MLPA), for the detection of large-scale CNVs in *LDLR*. The latter bench method has proven to be essential in the routine screening strategy for FH as ~10% of defined FH cases have been attributed to CNVs in *LDLR* (Bourbon et al., 2017); screening for them avoids false-negative diagnoses. However, while MLPA has been considered the “gold standard” for CNV detection in *LDLR* for the last decade or more, not all diagnostic laboratories have the infrastructure, resources, time, or interest to establish this parallel system despite its importance.

With use of NGS methods, comes the opportunity to detect CNVs. To be certain about single nucleotide variation from a genomic template, the library of derived short overlapping NGS fragments must redundantly encompass any particular nucleotide multiple times, particularly to ensure that heterozygous variation can be unambiguously

detected. A useful metaphor is the flipped coin - if heads are seen with each of only 5 flips, the presence of a side with "tails" cannot be ruled out with as much certainty as when "heads" appear consistently after 30 flips. This attribute of NGS data is referred to as "depth of coverage" (DOC); the standard DOC for NGS research applications is 30-fold, i.e. each base is read independently from a minimum 30 different synthesized fragments. For clinical applications, the required DOC can be much higher, often ~300-fold.

The large number of synthesized derived overlapping fragments required to maximize DOC in NGS offers an opportunity to determine not just qualitative but also quantitative aspects of the genomic region being interrogated. In general, the number of synthesized fragments follows stoichiometrically from the quantity of starting material. If there is half as much genomic template (i.e. only one copy instead of two), the resulting DOC reflected in the number of synthesized fragments will be reduced by half. If this involves a contiguous DNA region, then the DOC across the region will be reduced by 50% compared to normal neighboring regions. Similarly, for a duplicated genomic region, an increased number of DNA fragments are available to be read. This additional information from NGS data may be quantified using specific bioinformatic tools, with potential for CNVs to be detected.

Here, we determined the potential of replacing MLPA with bioinformatic analysis applied to NGS data for the detection of *LDLR* CNVs in FH patients. In analysis of 388 FH patient samples, we demonstrated 100% concordance in *LDLR* CNV detection between these two methods: 38 reported CNVs identified by MLPA were also successfully detected by our NGS method, while 350 samples negative for CNVs by MLPA were also negative by NGS. This result suggests that MLPA can be removed from the routine diagnostic screening for FH, significantly reducing associated costs, resources, and analysis time, while promoting more widespread assessment of this important class of variation across diagnostic laboratories.

2.2 Materials and Methods

2.2.1 Study subjects

We studied 388 Canadian individuals aged ≥ 18 years who were referred to the London Health Sciences Centre, University Hospital (London, ON) for treatment of severe hypercholesterolemia. Diagnosis of at least “possible FH” was made using the Dutch Lipid Clinic Network (DLCN) criteria; all patients had untreated LDL-C ≥ 5 mmol/L, plus family history of hypercholesterolemia, plus some with either personal or family history of premature CHD. Our protocol was approved by the Western University Research Ethics Board (No. 07920E) and all participants provided informed consent for genetic analyses.

2.2.2 Targeted next-generation sequencing

Genomic DNA was isolated from whole blood using the Puregene DNA Blood Kit (Gentra Systems, Qiagen, Mississauga, ON, Canada), and was subject to targeted NGS using our LipidSeq panel (Johansen et al., 2014). With LipidSeq, each sample is sequenced for 73 key genes in lipid metabolism, including all coding regions, ~ 150 base pairs (bp) at intron-exon boundaries, and ~ 250 bp of the 5' untranslated region (UTR) of all FH-associated genes *LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, *ABCG5*, and *ABCG8*. Library preparation was performed using the Nextera Rapid Capture Custom Enrichment kit (Illumina, San Diego, CA), and enriched samples were sequenced on a MiSeq personal sequencer platform (Illumina) using 2 x 150 bp paired-end chemistry and in accordance with manufacturer instructions. MiSeq-generated .FASTQ files were downloaded and processed individually using a custom automated workflow in CLC Genomics Workbench version 8.51 (CLC Bio, Aarhus, Denmark) for sequence alignment (mapped to human genome build GRCh37/hg19), variant calling (generation of .VCF files), and target region coverage statistics (generation of .BAM files). Our LipidSeq method has an average depth of coverage of 300-fold for each base.

2.2.3 CNV detection by MLPA

The MLPA Salsa P062-D2 kit (MRC Holland, Amsterdam, Netherlands) was used for the detection of large-scale whole-exon deletion and duplication events in *LDLR*. The P062-D2 kit contains 20 probes for *LDLR* (one for each of the promoter and all 18 exons, with the exception of two for exon 15), plus one flanking probe for upstream of *LDLR* and 12 reference probes for gene loci on alternative autosomal chromosomes. The probe-mix also contains 9 control fragments that generate short products to indicate that the DNA quantity and ligation reaction are sufficient for proper analysis. The principles and stages of probe hybridization are as previously described (Schouten et al., 2002), and protocol followed the manufacturer's guide version MDP-005 (www.mrc-holland.com). PCR amplification was carried out in a Veriti thermocycler (Applied Biosystems, Foster City, CA) and products were subsequently analyzed using a 3730 Automated DNA Sequencer (Applied Biosystems). MLPA fragment analysis was performed using Coffalyser software version 140721.1958 (MRC Holland; www.coffalyser.net), where relative amounts of probe-amplified products are compared with normal controls (samples within the same run) to determine the copy number state for each target region. We used one normal control sample per 7 study samples. Ratio values < 0.75 indicating copy number loss and > 1.33 indicating copy number gain were flagged. Two-sample t-tests were used for all statistical comparisons against the profiles of normal controls ($P < 0.05$).

2.2.4 CNV detection by NGS data

The bioinformatics tool CNV Caller, an application within the variant annotation software VarSeq v1.4.3 (Golden Helix, Bozeman, MT), was used for analysis of our existing LipidSeq data set for CNV detection. VarSeq CNV Caller requires .VCF and .BAM files (generated by NGS; see above) as inputs for each sample, plus a .BED file which defines the target region chromosomal and probe start/stop coordinates for the specific NGS panel used. The VarSeq algorithm uses normalized depth of coverage (DOC) analysis as its principal method, whereby an increase in sample DOC across a

target region, when compared to reference controls, suggests a gain in genomic material; and a decrease in sample DOC suggests a loss.

To first normalize the raw coverage data, the VarSeq algorithm uses a set of matched reference controls. We provided the algorithm a population of > 100 normal controls, from which it selects 30 with the lowest percent difference in coverage data compared to the sample of interest; samples are flagged if the average percent difference is > 20%. Matched reference controls are further used to correct for GC-content bias and regions that are relatively unamenable to mapping. A ratio and z-score metric are then computed for each target region. The ratio is calculated as the sample coverage divided by the mean reference sample coverage. The z-score measures the number of standard deviations that a sample's coverage is from the mean reference sample coverage. A Bayesian frame network model then assigns CNV state based on the probability that for each target region these two metrics represent either a: 1) diploid (normal) state; 2) heterozygous deletion; 3) homozygous deletion; or 4) duplication event. Further, the algorithm also exploits SNP heterozygosity information across a target region as an additional supporting metric for assigning CNV state. Denoted as variant allele frequencies (VAF), a VAF of any non 0 or 1 value provides further evidence against deletions, whereas a VAF such as 1/3 or 2/3 provides further evidence for duplications. Finally, segmentation analysis merges multiple affected target regions to characterize contiguous CNV events; the minimum limit of CNV detection is the smallest whole-exon (lower limit ~300 base pairs) while the maximum limit is the entire *LDLR* gene (approximately 18 kilobases).

2.2.5 CNV filtration

Following CNV analysis, CNVs were filtered based on mutually inclusive ratio and z-score thresholds. A ratio threshold value of ≤ 0.7 and z-score of ≤ -5.0 were used to identify probable heterozygous deletions, whereas a ratio value of ≥ 1.30 and z-score of ≥ 5.0 were used for duplications. For further validation, evidence from target region VAF's were also manually evaluated as explained above.

2.2.6 Statistical analyses

Analyses of demographic features were performed in SAS version 9.1 (SAS Institute, Cary NC). Quantitative traits were compared using unpaired t-tests, while discrete traits were compared using chi-square analysis, typically 2 X 2 contingency analyses. The nominal level of statistical significance was set at $P < 0.05$.

2.3 Results

2.3.1 Study sample demographics

Baseline clinical and biochemical features of the individuals studied here are shown below in **Table 2.1**.

Table 2.1. Canadian FH cohort patient demographics.

	Overall (N=388)	Women (N=212)	Men (N=176)
Age, years	50.7±15.2	52.1±16.3	48.9±13.6
Body mass index, kg/m ²	27.8±5.9	27.3±6.1	28.6±5.5
Total cholesterol, mmol/L	8.94±1.91	9.13±1.94	8.66±1.83
LDL cholesterol, mmol/L	6.79±1.79	6.93±1.80	6.60±1.76
HDL cholesterol, mmol/L	1.35±0.38	1.43±0.39	1.22±0.35
Triglyceride, mmol/L	1.79±0.88	1.77±0.98	1.81±0.73
Personal history of CVD, *%	17.9	12.9	25.0
Family history of CVD, *%	40.0	44.7	50.0
Definite or probable FH (DLCN score), *%	65.5	67.1	63.3

Values are represented as mean±standard deviation. CVD indicates cardiovascular disease (onset <55 years in men, <60 years in women). DLCN; Dutch Lipid Clinic Network.

*Based on complete data from 145 individuals.

2.3.2 CNVs detected by MLPA

Thirty-eight (9.8%) of 388 FH patients were positive for whole-exon CNVs in *LDLR* detected by MLPA (**Table 2.2**). The majority (35 of 38; 92.1%) of these patients had heterozygous deletions, of which 13 spanned multiple exons. There were 3 detected duplications. The most common CNV involved a heterozygous deletion of the promoter and exon 1, found in 22 of 38 (57.9%) CNV-positive patients. Exon 6 was affected in 6 of 38 (15.8%) patients. Of the 18 exons in *LDLR*, only exons 8, 9 and 10 were unaffected by CNV events among the study sample. All control samples had normal MLPA profiles. Sample outputs from MLPA for two different types of CNVs are shown in **Figures 2.1A** and **2.2A**.

2.3.3 CNVs detected by NGS data

Thirty-eight (9.8%) of 388 FH patients were positive for CNVs in *LDLR* detected by NGS. These CNVs and their associated states were in 100% concordance with those detected by MLPA (**Table 2.2**). Furthermore, the 350 samples negative for CNVs by MLPA were also negative by NGS. Using MLPA as the “gold standard” there were no false positives and no false negatives using our bioinformatics procedure applied to NGS data, which translates to a diagnostic test specificity and sensitivity of 100% each (**Table 2.3**). Sample outputs from VarSeq CNV Caller for two different types of CNVs corresponding to MLPA tracings referred to above are shown in **Figures 2.1B** and **2.2B**.

Table 2.2. *LDLR* gene whole-exon CNVs identified in 388 patients with FH.

Sample Number	MLPA		NGS Data		
	Type	Region	Detection	Ratio	Z-score
1	Het. Deletion	Promoter-Exon 1	Yes	0.62	-6.1
2	Het. Deletion	Promoter-Exon 1	Yes	0.43	-9.8
3	Het. Deletion	Promoter-Exon 1	Yes	0.58	-5.6
4	Het. Deletion	Promoter-Exon 1	Yes	0.51	-6.2
5	Het. Deletion	Promoter-Exon 1	Yes	0.50	-8.2
6	Het. Deletion	Promoter-Exon 1	Yes	0.53	-7.9
7	Het. Deletion	Promoter-Exon 1	Yes	0.52	-8.7
8	Het. Deletion	Promoter-Exon 1	Yes	0.56	-8.7
9	Het. Deletion	Promoter-Exon 1	Yes	0.51	-8.6
10	Het. Deletion	Promoter-Exon 1	Yes	0.52	-7.9
11	Het. Deletion	Promoter-Exon 1	Yes	0.51	-7.5
12	Het. Deletion	Promoter-Exon 1	Yes	0.59	-6.6
13	Het. Deletion	Promoter-Exon 1	Yes	0.55	-5.4
14	Het. Deletion	Promoter-Exon 1	Yes	0.51	-6.2
15	Het. Deletion	Promoter-Exon 1	Yes	0.59	-6.0
16	Het. Deletion	Promoter-Exon 1	Yes	0.51	-8.0
17	Het. Deletion	Promoter-Exon 1	Yes	0.59	-6.6
18	Het. Deletion	Promoter-Exon 1	Yes	0.46	-8.7
19	Het. Deletion	Promoter-Exon 1	Yes	0.51	-8.1
20	Het. Deletion	Promoter-Exon 1	Yes	0.57	-6.2
21	Het. Deletion	Promoter-Exon 1	Yes	0.51	-9.8
22	Het. Deletion	Promoter-Exon 1	Yes	0.57	-6.8
23	Het. Deletion	Promoter-Exon 2	Yes	0.57	-5.6
24	Het. Deletion	Promoter-Exon 2	Yes	0.50	-9.5
25	Het. Deletion	Promoter-Exon 6	Yes	0.54	-7.4
26	Het. Deletion	Exons 2-3	Yes	0.56	-6.7
27	Het. Deletion	Exons 2-6	Yes	0.54	-9.7
28	Duplication	Exons 2-6	Yes	1.38	11.8
29	Het. Deletion	Exons 3-6	Yes	0.53	-9.7
30	Het. Deletion	Exons 5-6	Yes	0.54	-14.7

31	Duplication	Exon 7	Yes	1.47	7.3
32	Duplication	Exons 11-12	Yes	1.86	12.7
33	Het. Deletion	Exons 11-12	Yes	0.54	-7.8
34	Het. Deletion	Exons 13-14	Yes	0.52	-15.9
35	Het. Deletion	Exons 13-15	Yes	0.65	-8.7
36	Het. Deletion	Exons 16-18	Yes	0.53	-9.9
37	Het. Deletion	Exons 17-18	Yes	0.53	-9.3
38	Het. Deletion	Exons 17-18	Yes	0.55	-10.1

For multi-exon copy number variants the reported ratio and z-score values are averaged across each affected region. Het., heterozygous.

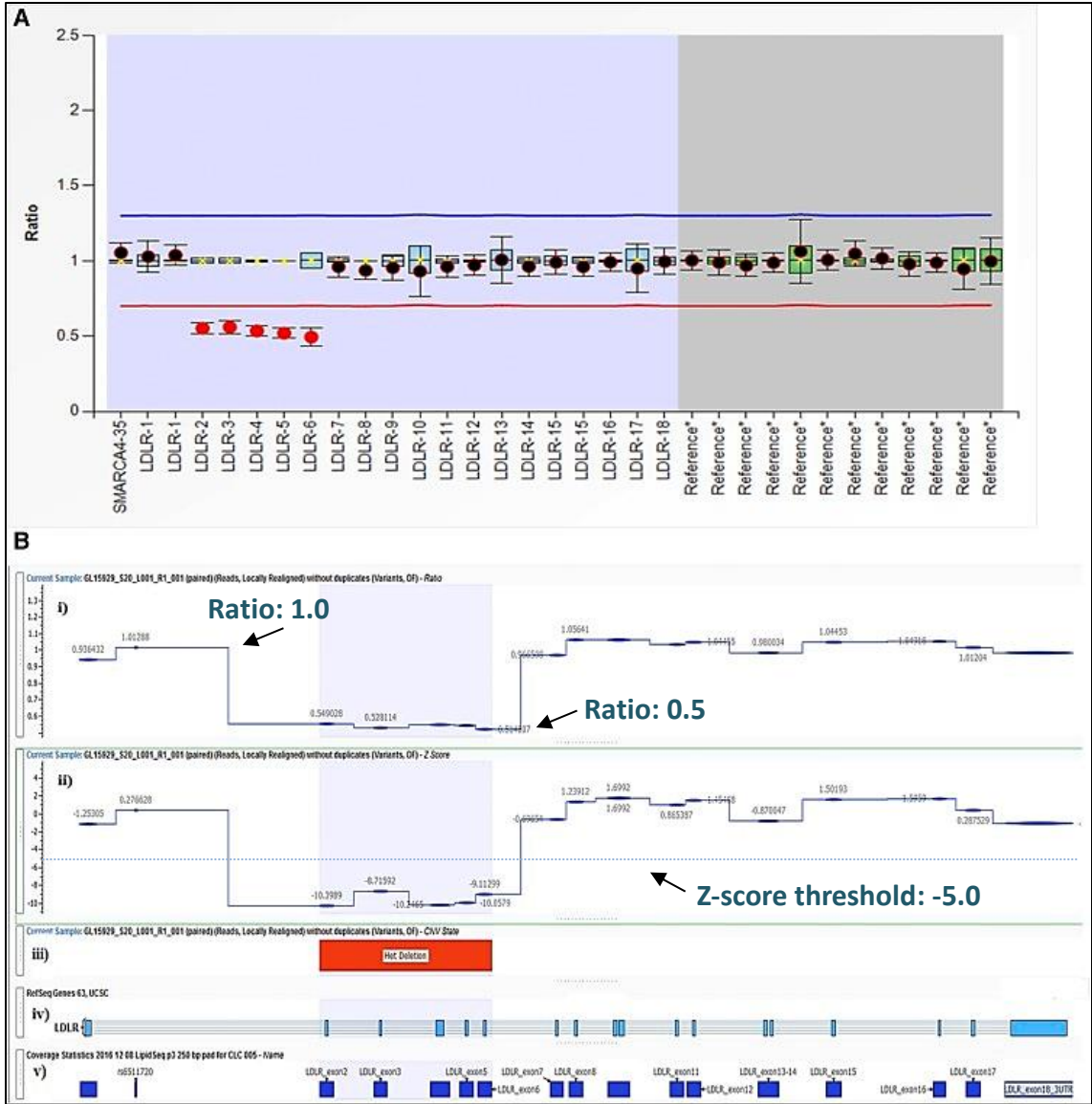


Figure 2.1. Two methods of detection of a CNV deletion event in the *LDLR* gene in a patient with FH. A) MLPA method output: heterozygous deletion in *LDLR* exons 2 – 6. Exon numbers are shown by “LDLR-N” (where N is the number of the exon, the first “LDLR-1” indicates the promoter, and “SMARCA4-35” is upstream of the promoter), and “*Reference” indicate reference probes bound to alternative chromosomes. For each probe target region, two separate plots are generated: 1) the normalized reference sample set is represented by 1-standard deviation box plots, where “X” indicates the mean and the horizontal line the median probe-signal intensity; and 2) the normalized patient sample probe-signal ratio is overlay as a dot, and is surrounded by error bars depicting the 95% confidence interval. The upper arbitrary border (blue line) and lower arbitrary border (red line) are placed +/- 0.3 from the reference sample mean of each probe. B) VarSeq CNV Caller method output: heterozygous deletion in *LDLR* exons 2 – 6. Different regions of the output are as follows: i) normalized ratio metric computed for each LipidSeq target region in *LDLR*; depth of sequence coverage comparative to reference controls where ~1.0 indicates diploid (normal) copy number state and ~0.50 indicates a heterozygous deletion event; ii) normalized z-score metric; number of standard deviations the depth of coverage is from the reference control mean coverage, where ≤ -5.0 is the threshold set to indicate a deletion event;. iii) CNV state, determined by ratio and z-score metrics together with supporting evidence from variant allele frequencies (not shown). Segmentation analysis has merged multiple affected target regions to call a contiguous heterozygous deletion event; iv) exon map of *LDLR* gene; and v) LipidSeq probe target regions.

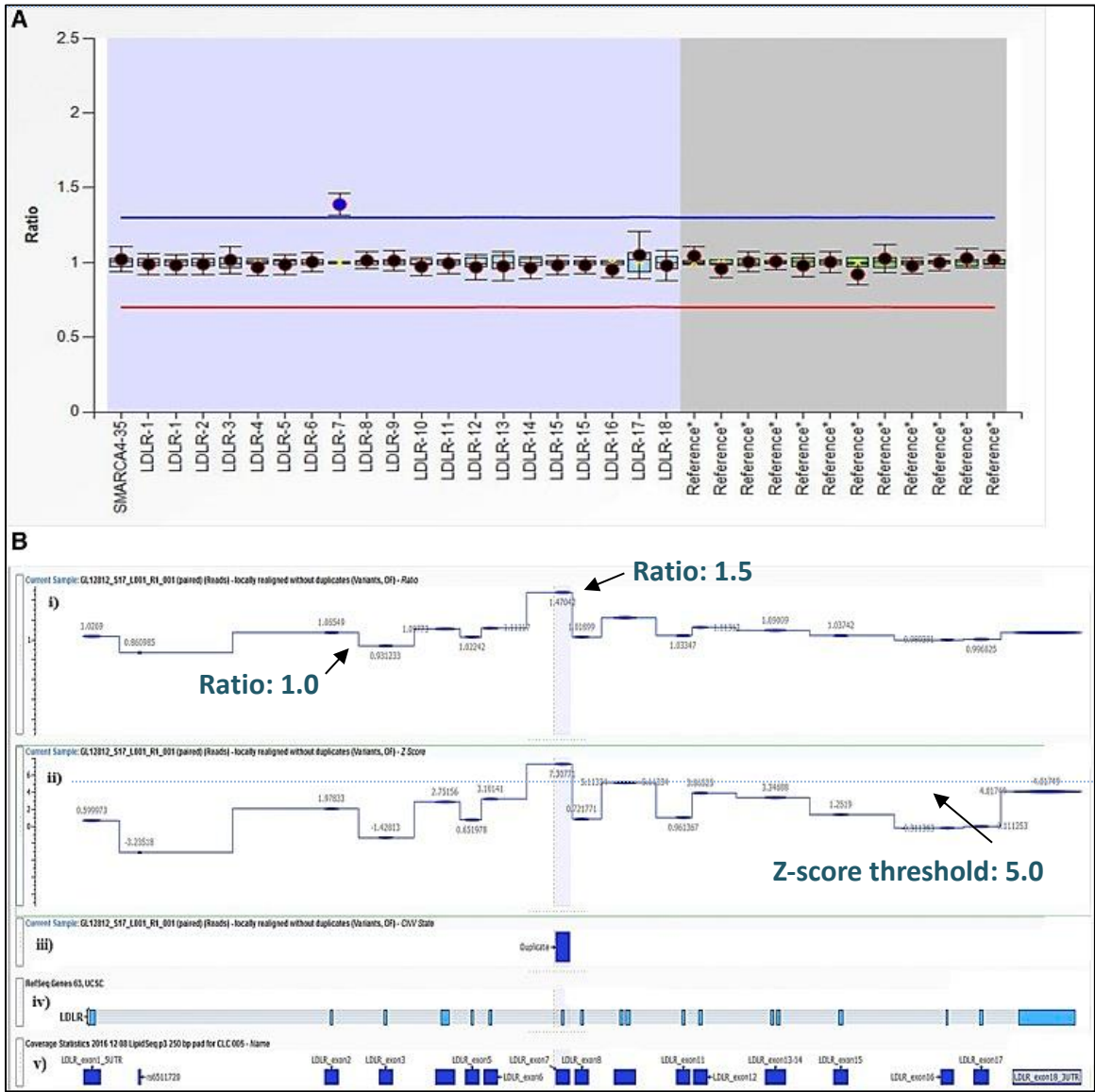


Figure 2.2. Two methods of detection of a CNV duplication event in the *LDLR* gene in a patient with FH. See legend to Figure 2.1 for overall structure of the panels. A) MLPA method output: duplication in *LDLR* exon 7. B) VarSeq CNV Caller method output: duplication in *LDLR* exon 7. Section i) normalized ratio metric computed for each LipidSeq target region in *LDLR*; depth of sequence coverage comparative to reference controls where ~1.0 indicates diploid (normal) copy number state and ~1.5 indicates a duplication event; ii) normalized z-score metric; number of standard deviations the depth of coverage is from the reference control mean coverage, where ≥ 5.0 is the threshold set to indicate a duplication event; other sections as in Figure 2.1.

Table 2.3. 2 X 2 contingency analysis of CNVs called by MLPA and NGS methods.

		<i>MLPA result</i>	
		CNV state	Diploid
<i>NGS data result</i>	Positive	True Positives 38	False Positives 0
	Negative	False Negatives 0	True Negatives 350
		Sensitivity: 100%	Specificity: 100%

2.4 Discussion

The ability to detect the full spectrum of DNA variation in *LDLR* is critical in obtaining a genetic diagnosis for FH, especially since up to 10% or more of such variants are large-scale CNVs rather than small-scale variants (Bourbon et al., 2017). The current procedure for diagnostic laboratories often includes targeted NGS followed by MLPA. Our findings suggest that the information about potential CNVs also resides within NGS data, and that MLPA is dispensable, particularly for the *LDLR* gene. NGS with appropriate bioinformatics has the potential to identify both small and large-scale variant detection in a single platform and single analytic procedure.

Specifically, in our analysis of 388 samples referred for FH diagnosis, 38 reported CNVs detected by MLPA were also successfully detected by NGS; no sample that was positive for a CNV by MLPA was missed by our bioinformatic approach. Importantly, with a specificity and sensitivity of 100%, there were no false-positive or false-negative calls derived from NGS data compared with MLPA. Furthermore, this targeted NGS method identified a wide range of CNV events, including those affecting almost all 18 exons, both single and multi-exon events, and both deletions and duplications (see **Table 2.2**).

The prevalence of whole-exon CNVs (9.8%) in FH patients is similar in our cohort compared to those previously studied (Goldmann et al., 2010; Taylor et al., 2009). The *LDLR* locus is known to have an especially high frequency of *Alu* repeat elements, making it susceptible to CNV mutagenesis (Goldmann et al., 2010; Hobbs et al., 1985). The pattern of CNV events detected across *LDLR* largely correlates with the distribution of these *Alu* repeats; sequence analysis in *LDLR* has revealed that the large majority of known CNV-breakpoints are found within introns 1-8 and 12-18 / 3'UTR, which is where *Alu* elements are most concentrated (Amsellem et al., 2002). This feature might explain why exons 8, 9, and 10 were unaffected by CNVs in our cohort. The high frequency of promoter-exon 1 heterozygous deletions can be attributed to the presence of French Canadians in our study sample. This ~15 kilobase deletion is a well-known founder-effect variant first discovered in 1987 to be present in ~60% of French Canadians with FH; presumably originating among the 8000 ancestors of the present-day French

Canadians who have traditionally had little cross-breeding with other ethnic groups (Hobbs et al., 1987). Because of the high prevalence of this specific variant, CNV analysis has long been an important component of FH screening in Canada.

For the last decade or more, MLPA has been regarded as the "gold standard" for CNV detection in *LDLR*. Prediction of CNVs from NGS data has been investigated previously; however, it remains a relatively new and challenging field. Commonly used CNV prediction programs include CoNIFER (Jiménez-Cruz et al., 2002), ExomeDepth (Plagnol et al., 2012), ExomeCopy (Love et al., 2011), XHMM (Fromer et al., 2012), and CNV-seq (Xie & Tammi, 2009), however many of these designated methods have shown high rates of false-positive CNVs, which poses a major limitation upon potential clinical use. Moreover, many of the literature reported CNV prediction programs have been designed and optimized for whole-genome or whole-exome NGS analysis, which is inherently different from targeted NGS analysis, as the latter focuses on only a few target genes, with known reference copy number counts, and provides a higher average sequence coverage per base, which in turn allows for depth of coverage (DOC) methods to be a suitable approach. The higher DOC for each particular *LDLR* base using targeted NGS versus whole-genome or whole-exome NGS potentially increases the sensitivity to detect CNVs. Finally, our study took advantage of our unique large cohort of known *LDLR* MLPA positive and negative samples as reference standards to evaluate the applicability of this bioinformatics approach to CNV detection in the clinical diagnostic context for FH.

Essential to the performance of DOC analysis is use of appropriate matched reference controls for cross-sample normalization and comparison (i.e. controls sequenced with the exact enrichment chemistry and NGS panel version design as the sample-of-interest) and quality-control thresholds set for ratio and z-score metric outputs. Although proven robust in detection, our methodology has some limitations in further defining CNVs. In the event of a called "duplication", the VarSeq CNV Caller output does not specify the exact degree of amplification. By design, this feature is a result of the difficulty in accurately differentiating DOC metrics as copy numbers incrementally increase. Another limitation is the inability to determine whole-exon CNV breakpoints as these reside in the intronic regions which are unsequenced on our LipidSeq panel.

Importantly however, although such information may be useful for research purposes, it does not affect the documentation of a CNV for the purpose of diagnosis.

2.5 Conclusion

In conclusion, we report 100% concordance for the detection of whole-exon CNVs in *LDLR* between a bioinformatics approach applied to existing NGS data and the “gold standard” reference method of MLPA. This result suggests that the latter independent bench method can be removed from the routine genetic diagnostic workup for FH, improving costs, resources, and analysis time and thus encouraging an even more commonplace assessment of this important class of variation across diagnostic laboratories in the future.

2.6 References

- Amsellem, S., Briffaut, D., Carrié, A., Rabés, J. P., Girardet, J. P., Fredenrich, A., ... Benlian, P. (2002). Intronic mutations outside of Alu-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Human Genetics*, *111*(6), 501–510.
- Bourbon, M., Alves, A. C., & Sijbrands, E. J. (2017). Low-density lipoprotein receptor mutational analysis in diagnosis of familial hypercholesterolemia. *Current Opinion in Lipidology*, *28*(2), 120–129.
- Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., ... Purcell, S. M. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *American Journal of Human Genetics*, *91*(4), 597–607.
- Goldmann, R., Tichý, L., Freiburger, T., Zapletalová, P., Letocha, O., Soška, V., ... Fajkusová, L. (2010). Genomic characterization of large rearrangements of the LDLR gene in Czech patients with familial hypercholesterolemia. *BMC Medical Genetics*, *11*(1), 115.
- Hobbs, H. H., Brown, M. S., Russell, D. W., Davignon, J., & Goldstein, J. L. (1987). Deletion in the gene for the low-density-lipoprotein receptor in a majority of French Canadians with familial hypercholesterolemia. *New England Journal of Medicine*, *317*(12), 734–737.
- Hobbs, H. H., Lehrman, M. A., Yamamoto, T., & Russell, D. W. (1985). Polymorphism and evolution of Alu sequences in the human low density lipoprotein receptor gene. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(22), 7651–5.
- Jiménez-Cruz, F., Ríos-Olivares, H., & Rubio-Arroyo, M. (2002). ¹H and ¹³C NMR investigation of 2-eq,9-ax-diaryl-azaadamantan-4-ones. *Journal of Molecular Structure*, *604*(2–3), 261–268.
- Johansen, C. T., Dubé, J. B., Loyzer, M. N., MacDonald, A., Carter, D. E., McIntyre, A. D., ... Hegele, R. A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *Journal of Lipid Research*, *55*(4), 765–772.
- Love, M. I., Myšičková, A., Sun, R., Kalscheuer, V., Vingron, M., & Haas, S. A. (2011). Modeling read counts for CNV detection in exome sequencing data. *Statistical Applications in Genetics and Molecular Biology*, *10*(1).
- Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., ... Nejentsev, S. (2012). A robust model for read count data in exome sequencing

experiments and implications for copy number variant calling. *Bioinformatics*, 28(21), 2747–2754.

Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F., & Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research*, 30(12), e57.

Taylor, A., Martin, B., Wang, D., Patel, K., Humphries, S. E., & Norbury, G. (2009). Multiplex ligation-dependent probe amplification analysis to screen for deletions and duplications of the LDLR gene in patients with familial hypercholesterolaemia. *Clinical Genetics*, 76(1), 69–75.

Xie, C., & Tammi, M. T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*, 10(1), 80.

Chapter 3 – Novel copy number variation screening in secondary FH-associated genes

Chapter 3 was adapted from the independent study “Whole-gene duplication of *PCSK9* as a novel genetic mechanism for severe hypercholesterolemia”, published in the *Canadian Journal of Cardiology* in 2018 (Iacocca et al., 2018; PMID: 30269829).

3.1 Introduction

FH is heterogeneous at the genetic level; although most cases result from inactivating variants in the *LDLR* gene, several other genes have also been implicated, including *APOB* and *PCSK9*, and less frequently *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, and *ABCG5/8*.

After demonstrating that bioinformatic tools could be applied to NGS data for the detection of CNVs in Chapter 2, we now had the ability to extend CNV analysis beyond the commonly studied *LDLR* gene and into all secondary FH-associated genes present on a given NGS panel. In the case of LipidSeq, this includes *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA* and *ABCG5/8* – genes in which only a limited number of small-scale variants have so far been described in FH cases (Calandra et al., 2017).

Although causative CNVs in these genes are expected to be rare, they have nevertheless remained completely uninvestigated in FH since MLPA methods are either not available or not applied for genes outside of *LDLR*. Extending CNV analysis to all FH-associated genes furthers our ability to account for all genetic abnormalities capable of explaining FH cases; this in turn further decreases the potential for false-negative findings during the course of genetic diagnosis.

Here, we performed novel CNV analysis in FH associated genes *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, and *ABCG5/8* to determine the potential for previously overlooked CNVs to be implicated in FH cases. In 704 FH samples, we identified a whole-gene duplication of *PCSK9* in two unrelated FH index cases; this *PCSK9* CNV

was also found to cosegregate with affected status in family members. This finding was associated with a profound FH phenotype in affected individuals, and the highest known plasma PCSK9 level reported in a human. We found no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, or *ABCG5/8* in our cohort of 704 FH individuals.

3.2 Materials and Methods

3.2.1 Study subjects

We increased our cohort size from 388 (studied in Chapter 2) to 704 patients. All patients had at least “possible FH” according to validated clinical criteria. This cohort included 429 samples from individuals referred to London Health Sciences Centre, University Hospital (London, ON, Canada) for treatment of severe hypercholesterolemia, plus 275 samples sent by collaborating physicians for genetic analyses. Our protocol was approved by the Western University Research Ethics Board (No. 07920E) and all individuals provided informed consent for genetic analyses.

3.2.2 Targeted next-generation sequencing (NGS)

Targeted NGS was performed using our LipidSeq panel, comprised of 73 lipid metabolism-related genes including all specified non-*LDLR* FH-associated genes, namely *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, and *ABCG5/8*. Details of our NGS protocol are as previously described above in Chapter 2, Methods section 2.2.2.

3.2.3 Whole-exome NGS

Whole-exome sequencing (WES) was performed in two CNV-positive index cases at the London Regional Genomics Centre (London ON, Canada). Library preparation was performed using the TruSeq Rapid Exome kit (Illumina) and enriched samples were sequenced on the Illumina NextSeq500 using 2 x 150 bp paired-end chemistry.

Bioinformatic analysis of raw sequencing data was performed using a custom automated workflow in CLC Genomics Workbench, as described above. This WES method has an average DOC of 125-fold per base.

3.2.4 NGS CNV analysis

CNV screening of NGS data was performed using the bioinformatic tool VarSeq CNV Caller (v1.4.3; Golden Helix, Bozeman, MT). The methodology and criteria used to call CNVs in genes of interest were described above in detail in Chapter 2, Methods section 2.2.4 and 2.2.5. VarSeq CNV analysis was applied to both our targeted NGS (LipidSeq) and WES data.

3.2.5 CNV confirmation by microarray analysis

Confirmation in two CNV-positive index cases was performed using the CytoScan HD Array (ThermoFisher Scientific, Waltham, MA, USA). The array has > 1.9 million non-polymorphic probes and > 750,000 single nucleotide polymorphism (SNP) probes. Only the CNV-containing genomic region was evaluated in each sample. The microarray was performed at Victoria Hospital (London ON, Canada) in accordance with manufacturer protocol. Data were analyzed using the Chromosome Analysis Suite (ChAS) version 3.2 (ThermoFisher Scientific). The signal patterns were compared with normal *in silico* reference data built in the ChAS software. Copy number loss or gain was visualized by \log_2 ratio (sample intensity/expected reference intensity).

3.2.6 Plasma PCSK9 analysis

Plasma PCSK9 levels were assessed in CNV-positive index case A by both an enzyme-linked immunosorbent (ELISA) and immunoprecipitation assay. ELISA (CircuLex) was performed on EDTA-plasma according to manufacturer's protocol (MBL International, Woburn, MA, USA), and repeated in triplicate. Immunoprecipitation was performed on

10 μ L of EDTA-plasma (case) or 50 μ L (normal control) aliquots that were diluted into 1 mL of buffer A (20 mM Hepes-KOH, pH 7.4, 100 mM NaCl, 1.5 mM MgCl₂, 1 mM CaCl₂, 1% NP-40) containing 2 mM PefaBloc (Roche, Basel, Switzerland) and precipitated overnight at 4°C using an in-house rabbit polyclonal anti-serum (Ab 1697) raised against full-length recombinant human PCSK9 and captured using goat anti-rabbit IgG-conjugated agarose beads (Rockland). The beads were washed three times with buffer A and eluted in SDS loading buffer (50 mM Tris-HCl, pH 6.8; 1% SDS; 5% glycerol; 10 mM EDTA; 0.0032% bromophenol blue, 2.5% (v/v) 2-mercaptoethanol). Immunoprecipitated proteins were subjected to 8% SDS-PAGE, transferred to nitrocellulose membrane (Bio-Rad) and incubated with primary anti-PCSK9 mouse monoclonal antibody (15A6). Infrared dye (IRDye-800)-labeled secondary antibody was used for detection on a LI-COR Odyssey infrared system (LI-COR Biosciences).

3.2.7 Statistical analysis

Analyses of demographic features were performed in SAS version 9.1 (SAS Institute, Cary NC). Discrete traits were compared using chi-square analysis, typically 2 X 2 contingency analyses, while quantitative traits were compared using unpaired t-tests. The nominal level of statistical significance was set at $P < 0.05$.

3.3 Results

3.3.1 Study sample demographics

Baseline clinical and biochemical traits of our cohort individuals are described in **Table 3.1**.

Table 3.1. FH cohort patient demographics.

	Overall (N=704)	Women (N=370)	Men (N=334)
Age, years	50.5 ± 15.9	52.6 ± 16.2	47.7 ± 15.0
Body mass index, kg/m ²	28.1 ± 5.69	27.5 ± 5.84	28.9 ± 5.39
Total cholesterol, mmol/L	8.88 ± 2.38	9.09 ± 2.63	8.59 ± 1.97
LDL cholesterol, mmol/L	6.7 ± 2.22	6.86 ± 2.47	6.48 ± 1.81
HDL cholesterol, mmol/L	1.37 ± 0.49	1.44 ± 0.38	1.29 ± 0.6
Triglyceride, mmol/L	2.02 ± 1.35	1.88 ± 1.11	2.21 ± 1.61

Values are represented as mean ± standard deviation. Complete demographic data available from 429 individuals.

3.3.2 Targeted NGS data CNV analysis

In screening our LipidSeq dataset for CNVs in FH-associated genes outside of *LDLR*, we detected a large-scale duplication encompassing the entire *PCSK9* gene in two FH index cases: hereafter termed index cases “A” and “B”. Located on chromosome 1p32, the human *PCSK9* gene is ~25 kilobases (kb) long and is comprised of 12 exons. Sample outputs for each index case are shown in **Figures 3.1A and 3.2A**. We detected no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA* and *ABCG5/8* in this cohort of 704 FH individuals.

3.3.3 Whole-exome CNV analysis

To determine whether the large-scale duplication encompassing *PCSK9* extended beyond the *PCSK9* locus and into flanking genes we performed WES in both index cases and applied subsequent bioinformatic CNV analysis. In index cases A and B the genes flanking *PCSK9* – 5’ *BSND* (upstream), and 3’ *USP24* (downstream) - were unaffected. Sample outputs are shown in **Figures 3.1B and 3.2B**.

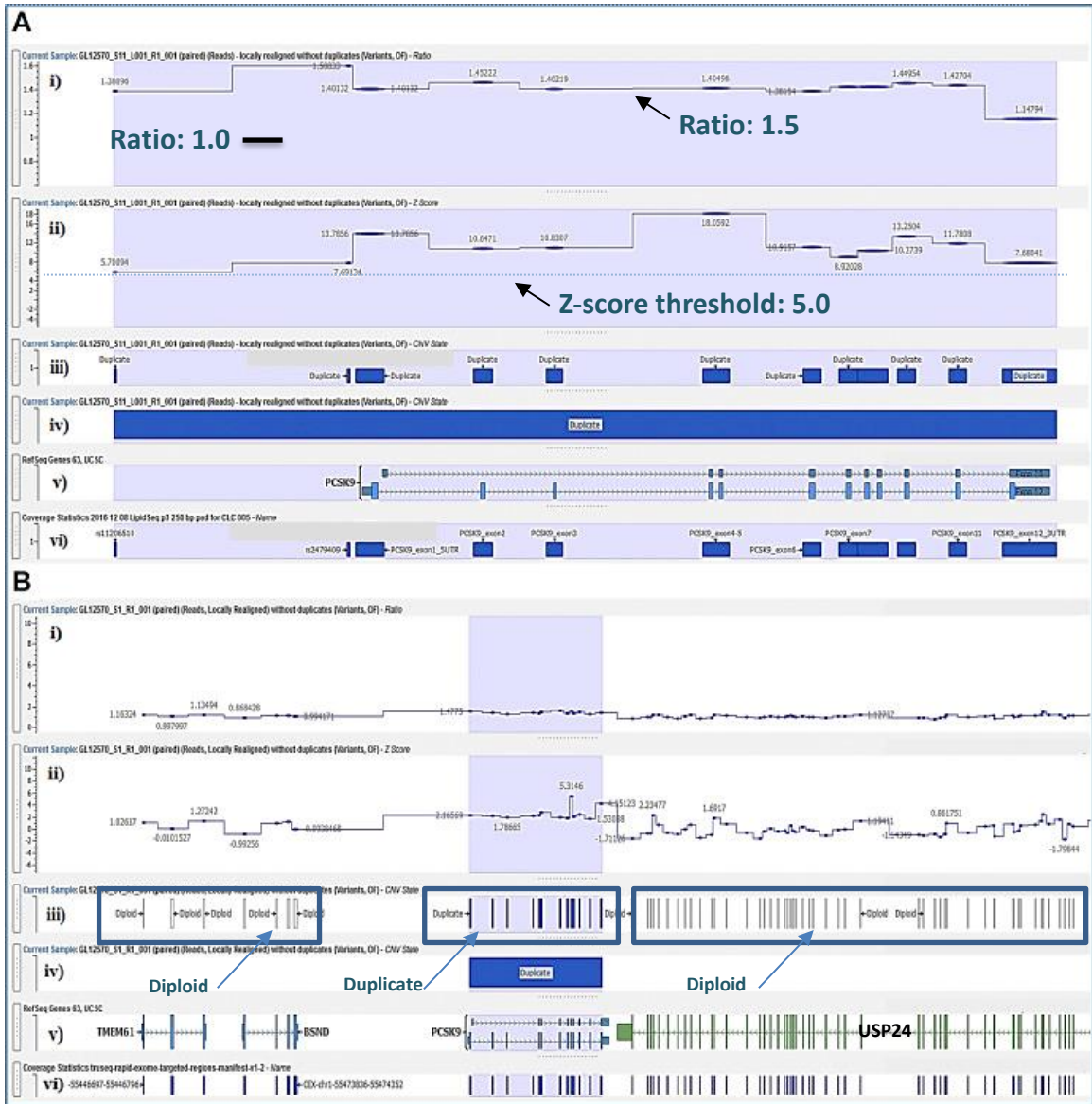


Figure 3.1. NGS-based detection of a *PCSK9* CNV in a patient with FH (index case A). (A) Targeted NGS output: duplication of all 12 exons of the *PCSK9* gene, plus rs11206510 probe 8655 bases upstream of *PCSK9*. Different regions of the output are as follows. i) Normalized ratio metric computed for each NGS probe target region in *PCSK9*; depth of sequence coverage comparative to reference controls. ii) Normalized z-score metric; number of standard deviations the depth of coverage is from the reference control mean coverage. iii) Called CNV state per probe target region, determined by ratio and z-score metrics together with supporting evidence from variant allele frequencies (not shown). iv) Multiple affected target regions merged by segmentation analysis to call a contiguous duplication event. v) Exon map of *PCSK9* gene. vi) LipidSeq probe target regions. (B) Whole-exome sequencing (WES) output: validation of *PCSK9* whole-gene duplication, plus flanking genes *BSND* (5') and *USP24* (3') unaffected (diploid). Panel regions i) – iv) are as in (A). v) Exon map of *PCSK9* and flanking genes. vi) WES probe target regions.

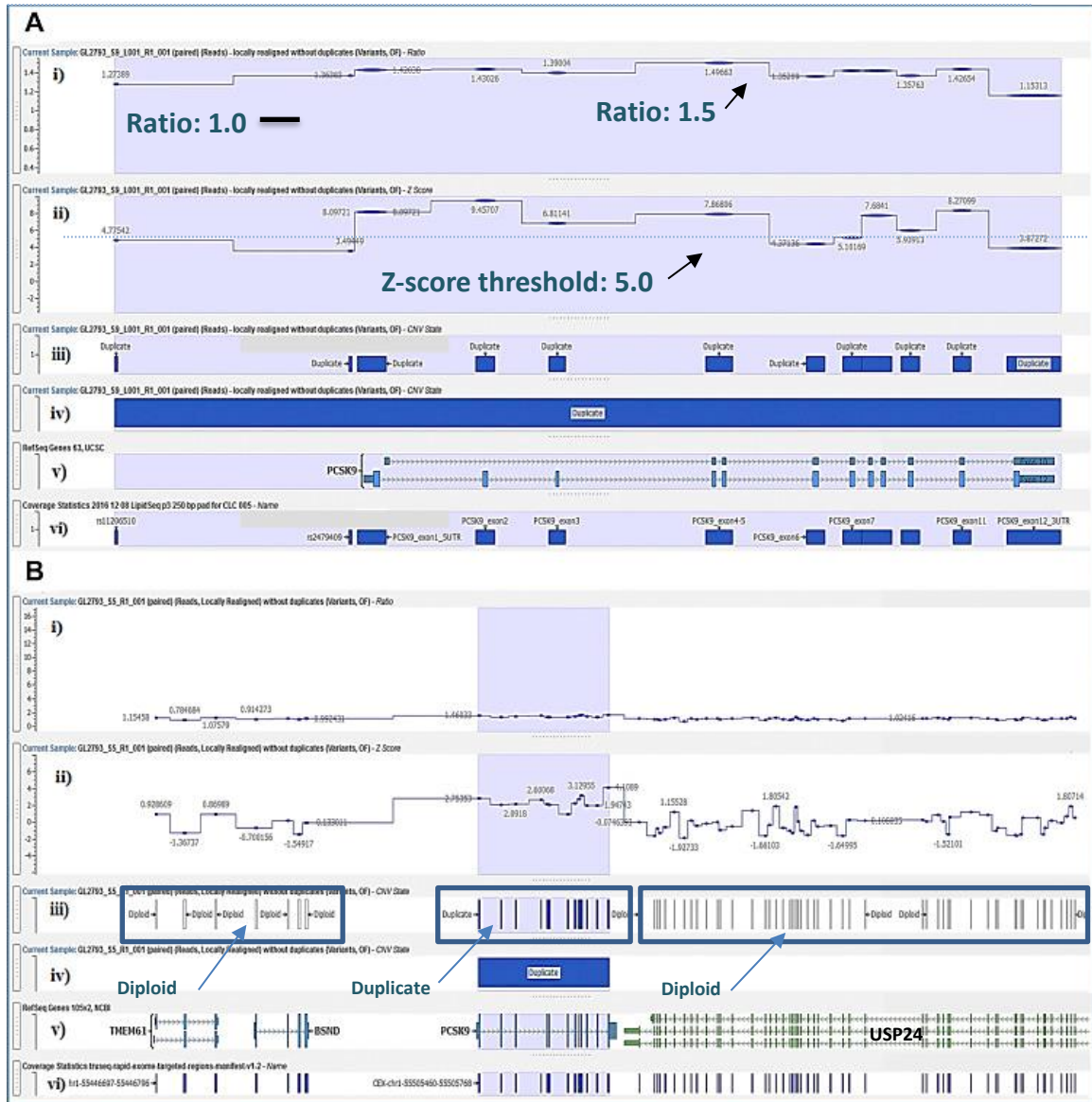


Figure 3.2. NGS-based detection of a *PCSK9* CNV in a patient with FH (index case B). (A) Targeted NGS output: duplication of all 12 exons of the *PCSK9* gene, plus rs11206510 probe 8655 bases upstream of *PCSK9*. (B) Whole-exome sequencing output: validation of *PCSK9* whole-gene duplication, plus flanking genes *BSND* (5') and *USP24* (3') unaffected (diploid). All panel regions are as in Figure 3.1.

3.3.4 CNV confirmation

Microarray-based CNV analysis performed in FH index cases A and B confirmed the presence of whole-gene *PCSK9* duplications, while adjacent genes were unaffected. The array allowed for further fine mapping; the total size of this CNV duplication was predicted to be ~ 35 kilobases. Sample outputs are shown in **Figure 3.3**.

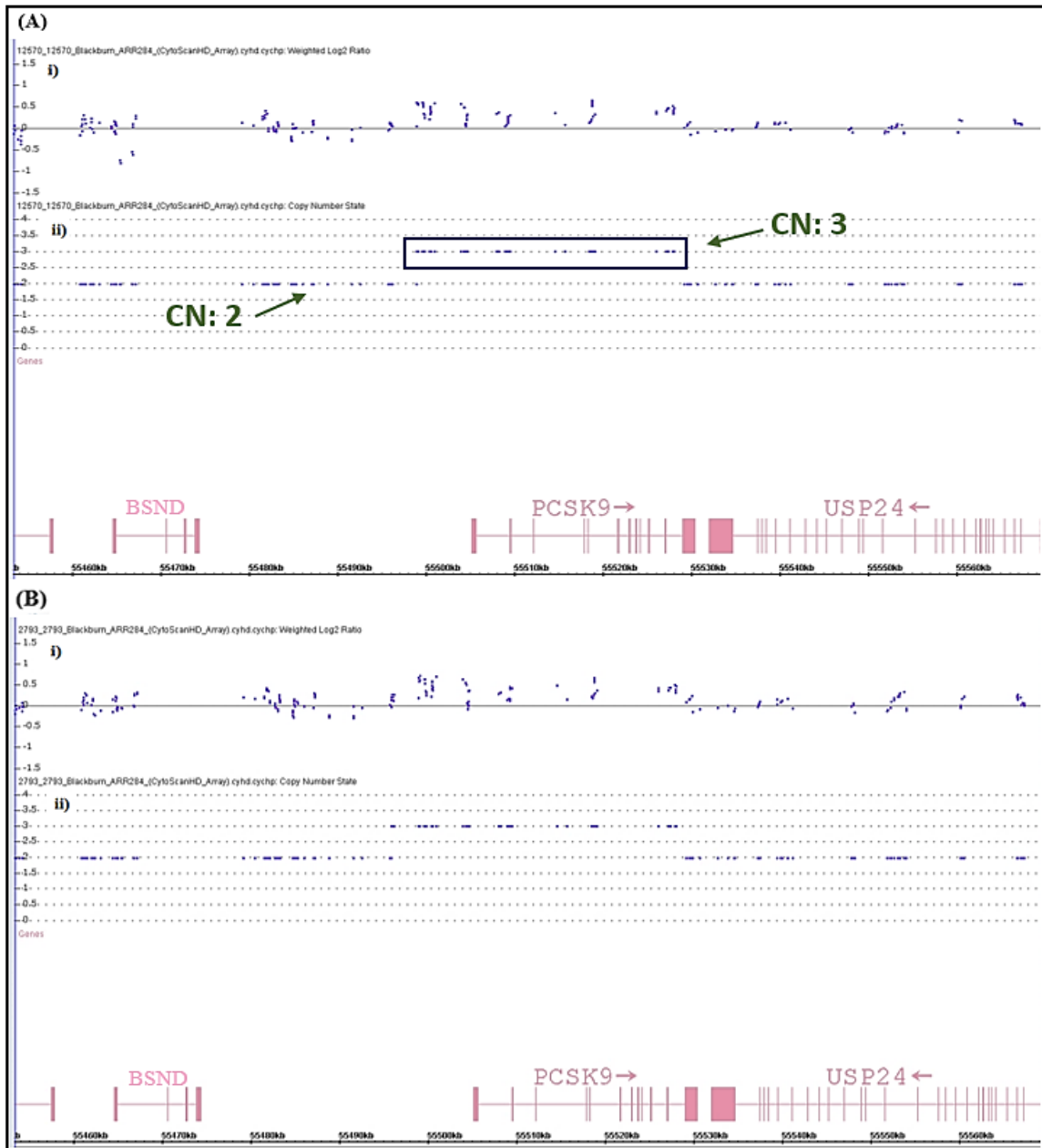


Figure 3.3. Microarray-based confirmation of a *PCSK9* CNV in two patients with FH.

(A) Index case A. Array output: Copy number of 3 for *PCSK9* gene plus probes located ~10 kilobases upstream of *PCSK9*; flanking genes *BSND* (5') and *USP24* (3') unaffected.

(B) Index case B. Array output: Copy number of 3 for *PCSK9* gene plus probes located ~10 kilobases upstream of *PCSK9*; flanking genes *BSND* (5') and *USP24* (3') unaffected.

Panel regions of the output are as follows. i) Log₂ ratio; per probe on the array, calculated as sample hybridization intensity compared to expected reference intensity. ii) Called copy number state.

3.3.5 Case presentations

3.3.5.1 Index case A

Index case A is a male of Northern European descent who was first treated for severe hypercholesterolemia at age 37 in 2013. He presented with an untreated LDL-C of 14.9 mmol/L, tendon xanthomata, and extensive atherosclerosis with angina symptoms. He was found to have severe multi-vessel coronary artery disease (CAD), with 95%, 99% and 60% occlusions in left main, circumflex and right coronary arteries, respectively, and 100% occlusion of the first diagonal branch of the left anterior descending coronary artery. He underwent urgent three-vessel coronary arterial bypass graft (CABG) surgery. His initial response to high-intensity statin therapy was poor, with < 20% LDL-C reduction from baseline values. Addition of ezetimibe 10 mg daily to atorvastatin (statin) 80 mg daily reduced the LDL-C to 9.7 mmol/L (i.e. a 34.8% reduction from baseline). Serial bi-weekly plasmapheresis treatments were more effective, with mean post-apheresis total cholesterol of 2.8 mmol/L; however, these treatments were discontinued after several months due to poor venous access. Addition of alirocumab (*PCSK9* inhibitor) 150 mg subcutaneously every 2 weeks reduced LDL-C to 6.8 mmol/L (i.e. an incremental 29.9% reduction from the value on statin plus ezetimibe), and a similar response was noted when alirocumab was switched to evolocumab (*PCSK9* inhibitor) 140 mg subcutaneously every 2 weeks. Most recently, evolocumab dose was increased to 420 mg subcutaneously every 2 weeks, reducing LDL-C to 5.5 mmol/L (a further 19.1%

reduction, or a 43.3% reduction compared to statin plus ezetimibe treatment alone). In 2015, initial LipidSeq NGS analysis followed by CNV analysis in *LDLR* by MLPA showed no causative variants.

His father had a historical untreated total cholesterol level of ~ 15 mmol/L with a similar attenuated response to statin treatment. He underwent CABG at age 50 years. Index case A also reported premature cardiovascular disease in second degree paternal relatives. His mother's untreated total cholesterol was 6.0 mmol/L. His asymptomatic 13-year-old daughter had serum total cholesterol of 9.5 mmol/L, triglycerides (TG) of 0.6 mmol/L; LDL-C of 7.5 mmol/L; and HDL-C of 1.78 mmol/L.

3.3.5.2 Index case B

Index case B is a male of Northern European descent, not known to be related to index case A, who was referred at age 40 with refractory, severe hypercholesterolemia, which was first diagnosed at age 25 years. His historical untreated LDL-C was 14.5 mmol/L. He was asymptomatic from cardiovascular and metabolic perspectives. He had diffuse xanthomatosis, involving finger extensor, Achilles and plantar flexor tendons bilaterally. With rosuvastatin 40 mg daily and ezetimibe 10 mg daily, his lowest recorded LDL-C level was 4.32 mmol/L, but typically this level ranged between 5.5 and 7.0 mmol/L on treatment. PCSK9 inhibition was never initiated before he was lost to follow-up due to work-related relocation. His family history was strongly positive for hypercholesterolemia. His father suffered a stroke at age 55 and had bilateral lower limb amputations in the seventh decade of life. His mother had hypercholesterolemia and underwent 4-vessel CABG at age 62. His maternal grandfather died at age 40 of a myocardial infarction. His 10 year-old son was reported to have hypercholesterolemia.

His older sister, younger brother and younger sister all had severe hypercholesterolemia; all received high-intensity statin and ezetimibe. His younger sister was assessed at age 38 after having been diagnosed with hypercholesterolemia at age 31. Her highest recorded untreated LDL-C level was 11.4 mmol/L. A lifelong cigarette smoker, she continued to smoke one pack daily even after her hypercholesterolemia diagnosis. On examination, she had bilateral xanthelasmas and diffuse pronounced

xanthomatosis, involving finger extensor and Achilles tendons bilaterally. At age 35 she developed left lower limb claudication, with diffuse femoral atherosclerosis demonstrated angiographically. With rosuvastatin 40 mg daily and ezetimibe 10 mg daily, her lowest recorded LDL-C level was 6.12 mmol/L, but typically this level ranged between 7 and 8 mmol/L on treatment. Her 11 year-old son was also reported to have hypercholesterolemia. Before PCSK9 inhibitors became available, she died at age 42 of a myocardial infarction. Initial LipidSeq NGS analysis of both siblings in 2014 followed by CNV analysis in *LDLR* using MLPA found no causative variants to explain their phenotype.

3.3.6 Co-segregation analysis

We next obtained DNA samples from family members of index case A and performed targeted NGS-based CNV analysis using LipidSeq. Both the affected father and affected daughter of index case A (described above) were positive for this *PCSK9* duplication, while the unaffected mother was CNV negative. The above-described affected sister of index case B was also one of our FH patients; her DNA was available for analysis and was found to be CNV positive. No additional family members of index case B were available for analysis. Pedigrees are shown in **Figure 3.4**. NGS-based sample outputs for index case family members are shown in **Figure 3.5 – 3.8**.

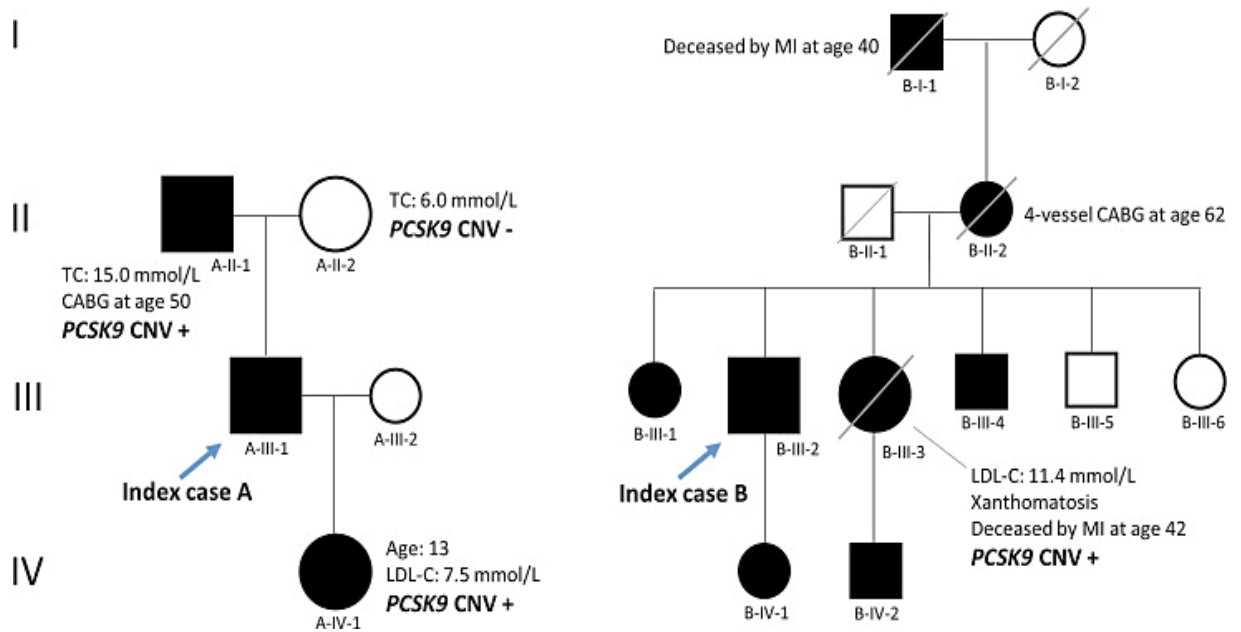


Figure 3.4. Family pedigree of two FH index cases with a whole-gene duplication of *PCSK9*. Males and females are represented as squares and circles, respectively, while black-shaded and un-shaded represents individuals with reported severe hypercholesterolemia and normal lipid profiles, respectively. Enlarged shapes refer to individuals where a DNA sample was possible to obtain and analyze for the presence (+) or absence (-) of a *PCSK9* copy number variation (CNV). Grey diagonal lines indicate deceased. Roman numerals I-IV indicate generation. MI, myocardial infarction; CABG, coronary arterial bypass graft.



Figure 3.5. Targeted NGS-based CNV assessment of *PCSK9* in an individual with FH (affected father of index case A). Targeted NGS output: duplication of all 12 exons of the *PCSK9* gene, plus rs11206510 probe 8655 bases upstream of *PCSK9*. Different regions of the output are as follows. i) Normalized ratio metric computed for each NGS probe target region in *PCSK9*; depth of sequence coverage comparative to reference controls. ii) Normalized z-score metric; number of standard deviations the depth of coverage is from the reference control mean coverage. iii) Called CNV state per probe target region, determined by ratio and z-score metrics together with supporting evidence from variant allele frequencies (not shown). iv) Multiple affected target regions merged by segmentation analysis to call a contiguous duplication event. v) Exon map of *PCSK9* gene. vi) LipidSeq probe target regions.

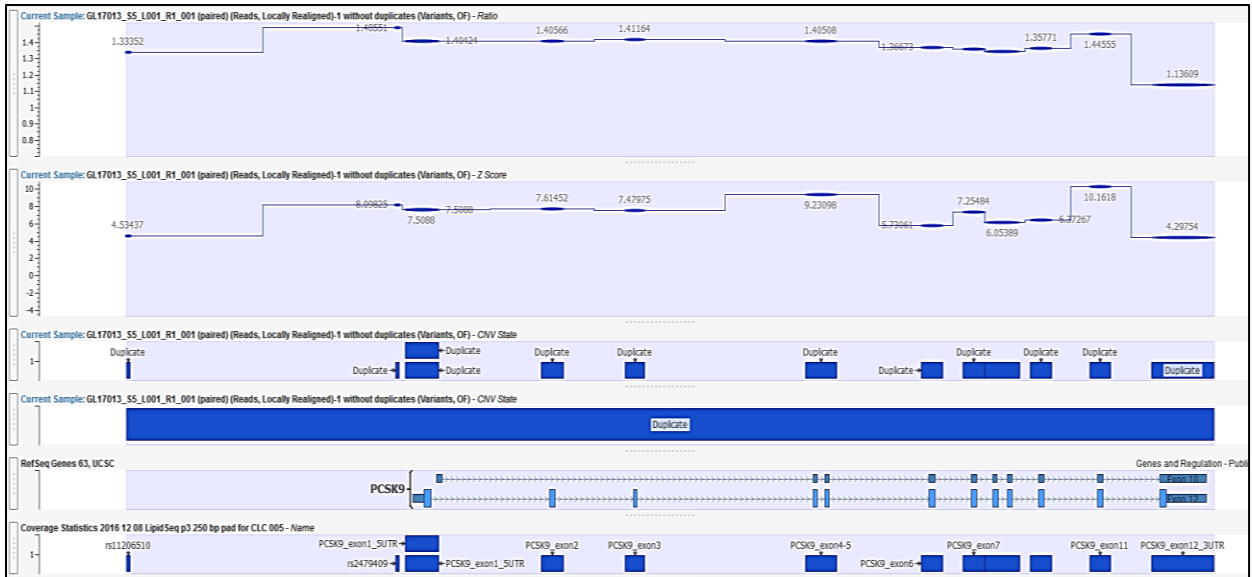


Figure 3.6. Targeted NGS-based CNV assessment of *PCSK9* in an individual with FH (affected daughter of index case A). Targeted NGS output: duplication of all 12 exons of the *PCSK9* gene, plus rs11206510 probe 8655 bases upstream of *PCSK9*.

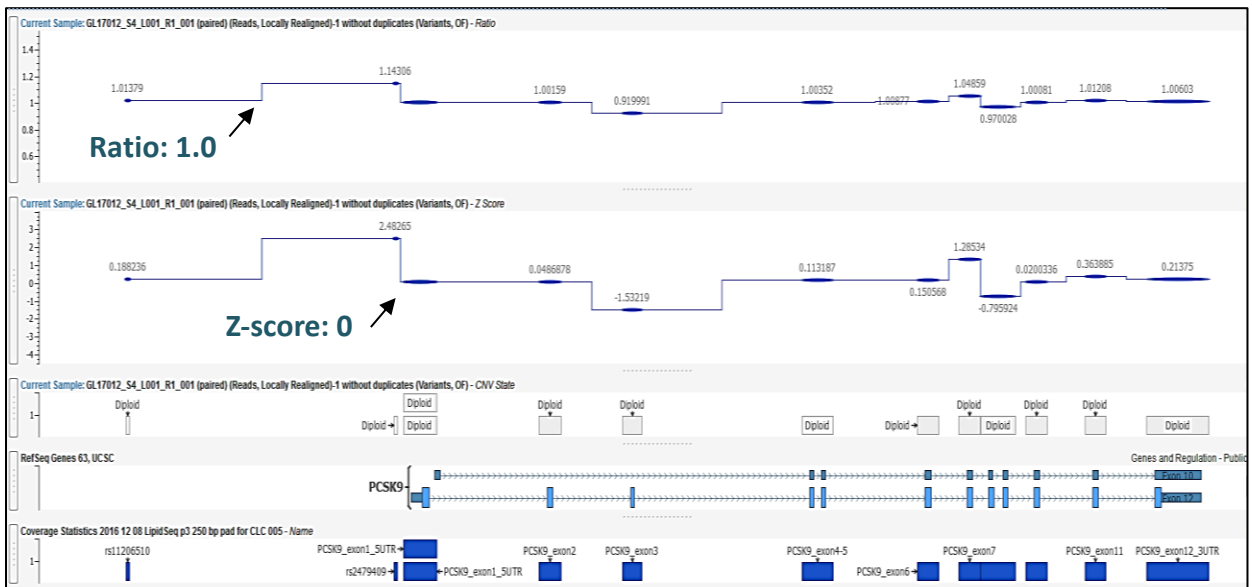


Figure 3.7. Targeted NGS-based CNV assessment of *PCSK9* in an individual without FH (unaffected mother of index case A). Targeted NGS output: unaffected (diploid) *PCSK9* gene.

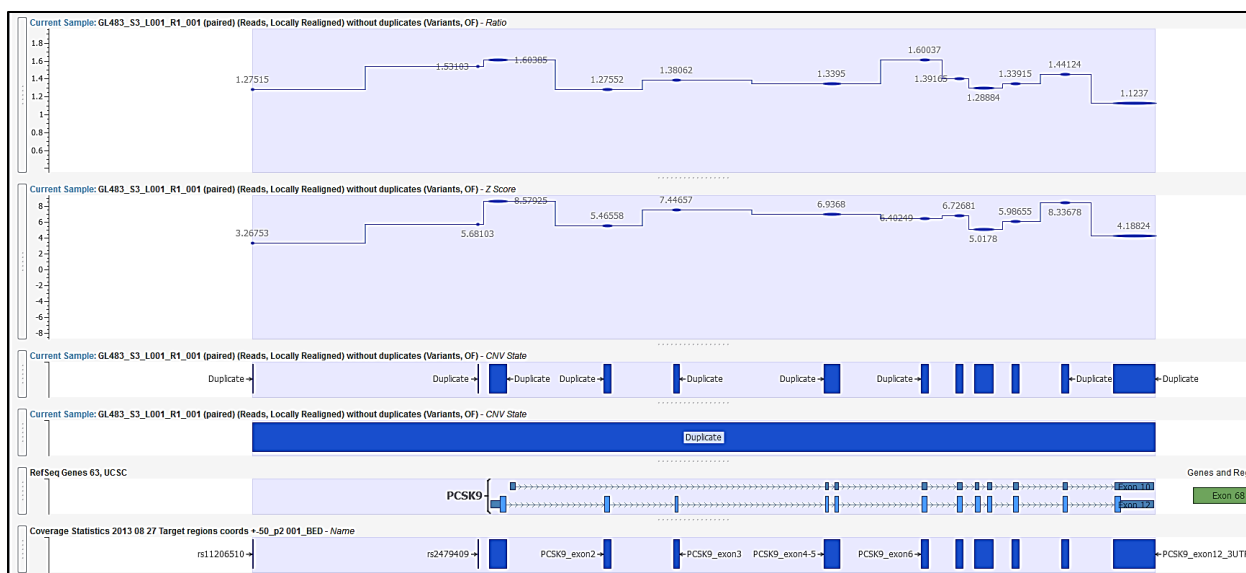


Figure 3.8. Targeted NGS-based CNV assessment of *PCSK9* in an individual with FH (affected sister of index case B). Targeted NGS output: duplication of all 12 exons of the *PCSK9* gene, plus rs11206510 probe 8655 bases upstream of *PCSK9*.

3.3.7 Plasma PCSK9 levels

Plasma PCSK9 in index case A was ~ 5000 ng/ml as determined by ELISA; this was a ~ 21-fold increase compared to a normal control (**Figure 3.9A**). Immunoprecipitation and immunoblot analysis confirmed that the observed increase corresponded to full-length PCSK9 (**Figure 3.9B**) and not a furin-cleaved inactive form (Benjannet et al., 2006). Plasma samples from index case B or affected relatives were not available for analysis.

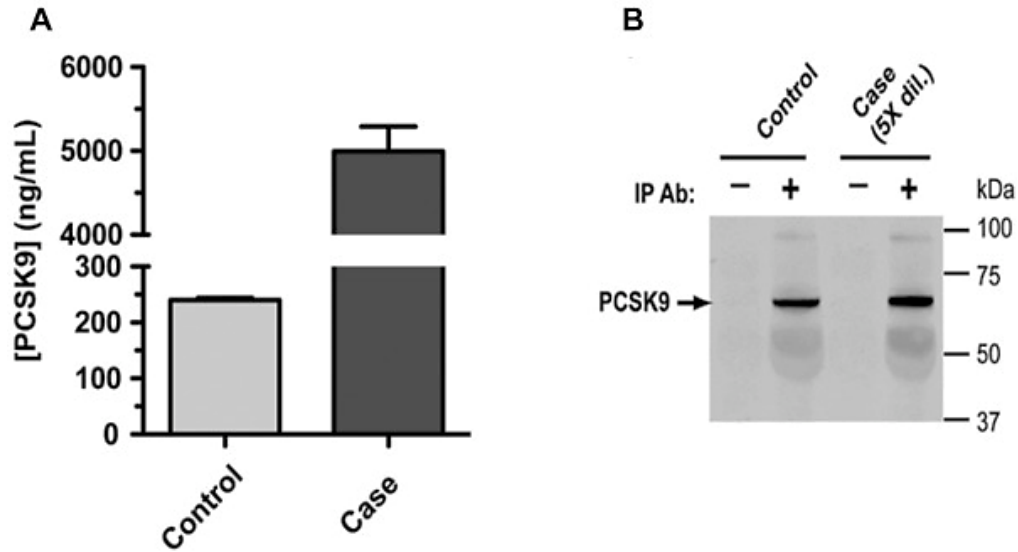


Figure 3.9. Plasma PCSK9 level in a FH patient (index case A) with a *PCSK9* whole-gene duplication. (A) Plasma PCSK9 enzyme-linked immunosorbent assay (ELISA) measurement (repeat n=3). (B) Plasma PCSK9 immunoprecipitation. PCSK9 was immunoprecipitated from EDTA-plasma using a rabbit polyclonal antibody raised against full-length recombinant human PCSK9 and detected with a monoclonal antibody (15A6). For comparison to normolipidemic control, plasma from FH patient (index case A) was diluted 5-fold prior to immunoprecipitation analysis.

3.4 Discussion

In this study, we used a NGS bioinformatic approach to perform novel CNV screening in FH-associated genes outside of the commonly studied *LDLR*. In a large cohort of 704 FH individuals we detected a whole-gene duplication of *PCSK9* in two index cases and no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, or *ABCG5/8*.

PCSK9 is a serine protease that governs net LDLR activity. Secreted mainly by the liver, *PCSK9* binds the LDLR at the cell surface, and following endocytosis of the LDLR-*PCSK9* complex, diverts LDLR toward lysosomes for degradation, thus short circuiting the normal recycling of the receptor to the cell surface (Zhang et al., 2007). Variants causing a gain of function (GOF) in *PCSK9* enhance LDLR degradation, resulting in elevated plasma LDL-C. Genetic analysis of atypical FH patients initially led to the discovery of *PCSK9*'s role in LDLR recycling and cholesterol homeostasis; in 2003 Abifadel *et al.* identified two GOF missense variants, p.Ser127Arg and p.Phe216Leu, in three French families with autosomal dominant FH (Abifadel et al., 2003). Since then, about 30 different *PCSK9* variants, many with distinct GOF mechanisms, have been described throughout all domains of the protein (Dron & Hegele, 2017). To date, however, these have all been small-scale variants – namely 24 missense, 2 splicing, 2 tri-nucleotide indels, and a 5'UTR substitution. The large-scale whole-gene duplication identified here, causing an increase in gene dosage, constitutes a novel GOF mechanism described for *PCSK9* in FH.

The severity of the FH phenotype in the index cases and their affected relatives is notable. In particular, untreated LDL-C levels ranged between 11.0 and 15.0 mmol/L, with prominent xanthomatosis and atherosclerotic cardiovascular disease presenting in the fourth decade of life, specifically 4-vessel CABG in index case A and myocardial infarction in index case B relatives. LDL-C levels elevated to this degree are more characteristic of homozygous FH, however, both NGS-based and microarray-based CNV analysis confirm that only a single *PCSK9* allele was affected in both families, with overall copy number increasing from two (diploid) to three.

Thus, a single extra copy of *PCSK9* seems to profoundly affect LDL-C homeostasis, underlying a severe form of FH. However, the phenotypic outcome of any

gene duplication depends on several factors, including the location and orientation of the duplicated genomic material. In other disorders, there is evidence to suggest most large-scale duplications occur in tandem (Newman et al. 2015), however, it is possible that the duplicated material is present elsewhere in the genome. The expression of a duplicated gene depends on genomic location and its epigenetic regulation. One caveat of using exome-based NGS CNV analysis, as well as microarray-based CNV analysis, is that while duplicated material can be detected, its precise location and orientation are not always defined.

This *PCSK9* CNV is associated with a plasma PCSK9 measurement of ~5000 ng/ml, a 21-fold increase compared to a normal control. Although variability in ELISA-based protocols make comparisons with values from other studies difficult, this is nonetheless by many-fold the highest known human level reported. This finding supports the functionality of this particular CNV event – i.e. the duplicated material is actively expressed.

Other factors might have influenced plasma PCSK9 levels in index case A, including high baseline LDL-C levels and statin treatment. For instance, plasma PCSK9 levels positively correlate with LDL-C; for LDL-C levels typically seen in HoFH (i.e. LDL-C >13.0 mmol/L, as seen in the index cases reported here), baseline plasma PCSK9 levels are 2- to 3-fold higher than normolipidemic controls (Cameron et al., 2012; Drouin-Chartier et al., 2015; Raal et al., 2013). Also, statins upregulate *PCSK9* expression; the *PCSK9* promoter contains a sterol regulator element (SRE) site and is co-expressed with *LDLR* following nuclear translocation of SREBP-2 in response to low intracellular cholesterol (Dubuc et al., 2004). Index case A remained on rosuvastatin 40 mg daily while plasma was taken for PCSK9 determination. Typically, administration of a high-intensity statin is associated with a 25–50% increase in plasma PCSK9 levels (Nozue, 2017). Raal et al. showed that rosuvastatin 40 mg daily resulted in a 37% increase in plasma PCSK9 levels (Raal et al., 2013). The effect of statins on *PCSK9* expression might be amplified when an extra copy of *PCSK9* is present. However, despite both the high background LDL-C and high-intensity statin therapy in index case A, the observed 21-fold increase in plasma PCSK9 is still disproportionately high. It is possible that the duplicated *PCSK9* gene in this patient may be located elsewhere in the

genome, perhaps driven by an unknown enhancer element, or resides in a genomic region where transcription is continually active (i.e. in a euchromatic state), resulting in an increased rate of expression and thus high PCSK9 levels.

Given the CNV-detection methodologies used here, the exact location of the *PCSK9* duplication within the genome is not known. This limits the ability to determine whether the same ancestral CNV event is present in both families, and also to speculate on possible gene expression influences which may explain the disproportionately high PCSK9 level detected.

These findings have therapeutic implications. As statin-induced upregulation of PCSK9 may be accentuated in patients with an extra copy of *PCSK9*, high-dose statin therapy may have only limited efficacy. Indeed, there was resistance to intensive statin therapy in both index cases and in several family members. With plasma PCSK9 levels increased, index case A also appeared to require a high dose of a PCSK9 inhibitor.

In addition, we found no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, or *ABCG5/8* in this cohort of 704 FH individuals. It is possible that pathogenic CNVs in these other FH-related genes exist, but may require larger FH cohorts to be detected. The potential of finding CNVs in these genes is of interest since ~20-40% of patients with suspected FH in many clinical cohorts have no apparent "typical" or obvious genetic aberrations underlying their phenotype. Evaluating SNPs to define a possible polygenic basis for hypercholesterolemia may explain an additional ~20% of clinically ascertained FH cases (Wang et al., 2016), but still leaves a substantial number of "unexplained" cases. Systematic screening in additional populations could help evaluate the possibility that CNVs in other FH-related genes may be present in some FH subjects.

3.5 Conclusion

In conclusion, we performed novel CNV screening in FH-associated genes in a large cohort of 704 FH individuals, and identified a whole-gene *PCSK9* duplication in two FH index cases and no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, or *ABCG5/8*. This is the first report of a *PCSK9* CNV associated with a severe FH phenotype and

profoundly elevated plasma PCSK9. The grossly elevated PCSK9 level may limit the efficacy of intensive statin therapy and perhaps also the efficacy of PCSK9 inhibition. These findings also highlight the potential for finding novel disease-causing variants when CNV screening is extended beyond the commonly studied *LDLR* gene, and may help to further avoid false-negative genetic diagnoses and direct treatment strategy.

3.6 References

- Abifadel, M., Varret, M., Rabès, J. P., Allard, D., Ouguerram, K., Devillers, M., ... Boileau, C. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genetics*, *34*(2), 154–156.
- Benjannet, S., Rhainds, D., Hamelin, J., Nassoury, N., & Seidah, N. G. (2006). The proprotein convertase (PC) PCSK9 is inactivated by furin and/or PC5/6A: Functional consequences of natural mutations and post-translational modifications. *Journal of Biological Chemistry*, *281*(41), 30561–30572.
- Calandra, S., Tarugi, P., & Bertolini, S. (2017). Impact of rare variants in autosomal dominant hypercholesterolemia causing genes. *Current Opinion in Lipidology*, *28*(3), 267–272.
- Cameron, J., Bogsrud, M. P., Tveten, K., Strom, T. B., Holven, K., Berge, K. E., ... Leren, T. P. (2012). Serum levels of proprotein convertase subtilisin/kexin type 9 in subjects with familial hypercholesterolemia indicate that proprotein convertase subtilisin/kexin type 9 is cleared from plasma by low-density lipoprotein receptor-independent pathways. *Translational Research*, *160*(2), 125–130.
- Dron, J. S., & Hegele, R. A. (2017). Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Current Opinion in Lipidology*, *28*(2), 161–169.
- Drouin-Chartier, J. P., Tremblay, A. J., Hogue, J. C., Ooi, T. C., Lamarche, B., & Couture, P. (2015). The contribution of PCSK9 levels to the phenotypic severity of familial hypercholesterolemia is independent of LDL receptor genotype. *Metabolism: Clinical and Experimental*, *64*(11), 1541–1547.
- Dubuc, G., Chamberland, A., Wassef, H., Davignon, J., Seidah, N. G., Bernier, L., & Prat, A. (2004). Statins upregulate PCSK9, the gene encoding the proprotein convertase neural apoptosis-regulated convertase-1 implicated in familial hypercholesterolemia. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*(8), 1454–1459.
- Newman, S., Hermetz, K. E., Weckselblatt, B., & Rudd, M. K. (2015). Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints. *American Journal of Human Genetics*, *96*(2), 208–220.
- Nozue, T. (2017). Lipid lowering therapy and circulating PCSK9 concentration. *Journal of Atherosclerosis and Thrombosis*, *24*(9), 895–907.
- Raal, F., Panz, V., Immelman, A., & Pilcher, G. (2013). Elevated PCSK9 levels in untreated patients with heterozygous or homozygous familial hypercholesterolemia

and the response to high-dose statin therapy. *Journal of the American Heart Association*, 2(2), e000028–e000028.

Wang, J., Dron, J. S., Ban, M. R., Robinson, J. F., McIntyre, A. D., Alazzam, M., ... Hegele, R. A. (2016). Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 36(12), 2439–2445.

Zhang, D. W., Lagace, T. A., Garuti, R., Zhao, Z., McDonald, M., Horton, J. D., ... Hobbs, H. H. (2007). Binding of proprotein convertase subtilisin/kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation. *Journal of Biological Chemistry*, 282(25), 18602–18612.

Chapter 4 – ClinVar database: data sharing to improve interpretation of genetic variants identified in FH patients worldwide

Chapter 4 was adapted from the independent study “ClinVar database of global familial hypercholesterolemia-associated DNA variants”, published in *Human Mutation* in 2018 (Iacocca et al., 2018; PMID: 30311388).

4.1 Introduction

Successful genetic diagnosis hinges on the ability of a diagnostic laboratory to accurately interpret detected DNA variants as “pathogenic” or “benign” – a challenging task when considering there are often dozens or more rare DNA variants identified during the course of NGS sequence analysis. Since genetic analysis in FH began in the 1970s, a vast number of DNA variants have been identified in patients, primarily within *LDLR*, and more recently in *APOB* and *PCSK9*.

Interpreting the clinical significance of genetic variants is challenging, and often multifaceted, however, can be greatly improved when there is information on variants from multiple independent sources which can be shared among laboratories. A data-sharing culture is not new among the FH field; for years the Leiden Open Source Variation Database (LOVD) has served as a publicly available FH-variant repository, hosting 1707 unique *LDLR* variants as of 2016 (Leigh et al., 2017). However, ClinVar, an NCBI-funded resource, has since emerged as the primary centralized database for archiving clinically relevant variants for many Mendelian diseases (Landrum et al., 2014). ClinVar facilitates a much more comprehensive approach to both the consolidation and presentation of patient and molecular data, and includes a multitude of interconnected resources to aid in improving variant interpretation (Harrison et al., 2016).

Here, I present the recent efforts made by the Clinical Genome (ClinGen) Resource consortium, along with various global FH researchers, to update the number and characterization of FH variants hosted by ClinVar. I specifically reviewed,

standardized, and submitted data on 369 FH-associated variants identified in Dr. Robert Hegele's laboratory to ClinVar (4th largest single-center variant contribution globally), and following submission efforts, co-lead data analysis of all 6,651 FH-associated variants submitted worldwide. In doing this, I also identified multiple areas for improvement to further reform the interpretation of FH-associated variants in the future.

4.2 Materials and Methods

4.2.1 ClinGen FH Variant Curation Expert Panel

The ClinGen FH Variant Curation Expert Panel (FH VC-EP) is composed of >20 members (**Table 4.1**). Members were selected on the basis of achieving a balanced representation of expert clinicians, clinical laboratory diagnosticians, researchers, and genomic medicine specialists. An emphasis was also placed on global representation, with members from the United States, Brazil, United Kingdom, Netherlands, France, Portugal, Czech Republic, Spain, Israel, Australia and Canada. The FH VC-EP is part of the ClinGen Cardiovascular Clinical Domain Working Group.

Table 4.1. ClinGen FH Variant Curation Expert Panel members.

Name	Institution	Area and Type of Expertise	Role
Joshua W. Knowles, MD, PhD	Stanford University & FH Foundation, USA	Clinician/Researcher/Laboratory Director	Executive Leader
Mafalda Bourbon, PhD	Instituto Nacional de Saúde Doutor Ricardo Jorge, Portugal	Researcher/Laboratory Director	Executive Leader
C. Lisa Kurtz, PhD	University of North Carolina, Chapel Hill, USA	Researcher	Coordinator
Robert A. Hegele, MD	Robarts Research Institute, Western University, Canada	Clinician/Researcher/Laboratory Director	Executive Committee
Eric J. Sijbrands, MD, PhD	Erasmus University, Netherlands	Clinician/Researcher/Laboratory Director	Executive Committee
Alain Carrie, MD, PhD	Pitié-Salpêtrière University Hospital, France	Researcher/Laboratory Director	Executive Committee
Joep C. Defesche, PhD	Academic Medical Center, University of Amsterdam, Netherlands	Researcher/Laboratory Director	Executive Committee
Tomas Freiburger, MD, PhD	University Hospital Brno, Czech Republic	Researcher/Laboratory Director	Executive Committee
Sarah E. Leigh, PhD	Genomics England, United Kingdom	Researcher	Executive Committee
Amanda J. Hooper, PhD	PathWest Laboratory Medicine, University of Western Australia	Researcher/Laboratory Director	Expert
Steve E. Humphries, PhD	University College of London, United Kingdom	Researcher/Laboratory Director	Expert
Amit V. Khera, MD	Broad Institute & Massachusetts General Hospital, USA	Clinician/Researcher	Expert
Michael Murray, MD	Geisinger, USA	Clinician/Researcher/Director of Clinical Genomics	Expert
Jean-Pierre Rabes, MD, PhD	Hôpital Ambroise Paré, France	Clinician/Researcher/Laboratory Director	Expert
Daniel Rader, MD	University of Pennsylvania, USA	Clinician/Researcher/Scientific Director	Expert
Raul Santos, MD, PhD	InCor, São Paulo University, Brazil	Clinician/Researcher/Scientific Director	Expert
Marianne Stef, PhD	Progenika/Grifols, USA	Researcher/Laboratory Director	Expert
Marina Cuchel, MD, PhD	University of Pennsylvania, USA	Clinician/Researcher/Laboratory Director	Expert

Mariko Harada-Shiba, MD, PhD	National Cerebral and Cardiovascular Center Research Institute, Japan	Clinician/Researcher/Laboratory Director	Expert
Margaret Chen, PhD, FACMG	GeneDx, USA	Researcher/Laboratory Director	Expert
Ronen Durst, MD	Hadassah Hebrew University Medical Center, Israel	Clinician/Researcher/Laboratory Director	Expert
Pedro Mata, MD, PhD	Fundacion Hipercolesterolemia Familiar, Spain	Clinician/Researcher/Laboratory Director	Expert
Joana R. Chora, MSc	Instituto Nacional de Saúde Doutor Ricardo Jorge, Portugal	PhD student	Curator
Michael A. Iacocca, BSc	Robarts Research Institute, Western University, Canada	MSc student	Curator
Lukas Tichy, PhD	University Hospital Brno, Czech Republic	Post doctorate researcher	Curator

4.2.2 Variant submission to ClinVar

Starting in 2016, several sources were recruited for consolidation of FH-associated variants into ClinVar. These efforts were facilitated by the FH Foundation working with ClinGen leadership to convene a session of interested parties, including members of the FH VC-EP at the 2016 international *FH Summit* in Dallas, USA, and 2017 in Miami, USA. First, FH VC-EP members began submitting FH-associated variants and variant-level data from their respective internal databases to ClinVar. We then encouraged global colleagues to submit internally stored FH-associated variants, with a focus on the largest remaining sequencing centers from various countries and jurisdictions. Further, we facilitated variant transfer from existing centralized databases, namely LOVD (<https://databases.lovd.nl/shared/genes/LDLR>).

Submitters followed a standard protocol for submission. They were required to register their organization/center on the ClinVar Submission Portal (<https://submit.ncbi.nlm.nih.gov/clinvar/>). Following ClinVar approval, variant submissions were performed using the Submission Template spreadsheet

(<https://www.ncbi.nlm.nih.gov/clinvar/docs/submit/>). Submitted variants required standardized annotation (HGVS expression or chromosomal coordinate change), associated condition, interpretation of clinical and/or functional significance, interpretation criteria, collection method (clinical testing or research), allele origin (germline or somatic), and individual affected status. A wide range of additional variant-level data types were optional for inclusion, such as number of variant observations, ethnicity and/or geographic origin of the individual, cosegregation/family data, functional data, phenotypic information, and/or normolipidemic screening results.

4.2.3 ClinVar variant analysis

Following submission efforts, ClinVar Miner (<https://clinvarminer.genetics.utah.edu/>) was used to extract variant-level data from the ClinVar database for *LDLR*, *APOB*, and *PCSK9*. Variants that did not have a submitted disease association of “Familial hypercholesterolemia” or accepted alternative term were removed manually, specifically: “Familial hypobetalipoproteinemia” (n=221), “Hypercholesterolemia, autosomal dominant, type B; Hypobetalipoproteinemia, familial, 1” (n=156; entry of two opposing conditions per single individual), “Low density lipoprotein cholesterol level quantitative trait locus 1” (n=3), “hypocholesterolemia” (n=2), “Hypobetalipoproteinemia, familial, 1” (n=2), “Early-onset coronary artery disease (CAD)” (n=2; removed as other dyslipidemias/morbidities can lead to CAD), “Hypobetalipoproteinemia” (n=1), “C0950123: Inborn genetic diseases” (n=1), “not specified” (n=191), and “not provided” (n=164). Variant consequences were determined manually from DNA and protein level variant information and confirmed using the Mutalyzer Name Checker batch tool v.2.0.28 (Leiden University Medical Center, Netherlands; <https://mutalyzer.nl/>).

4.3 Results

4.3.1 Global ClinVar submission

Prior to 2016, there were 242 (193 unique) *LDLR*, 63 (59) *APOB* and 26 (26) *PCSK9* variant submissions present in ClinVar. In a concerted effort to increase this number, the ClinGen FH VC-EP encouraged the submission of FH-associated variants by colleagues and sequencing centers on a global scale. As a result, the number of FH-associated submissions now residing in the ClinVar database increased ~ 18-fold and is summarized in **Table 4.2**. Additionally, there are 201 *LDLR*, 423 *APOB*, and 119 *PCSK9* variant submissions that do not have a disease association of FH and were removed from analysis. A total of 30 centers from 13 different countries have submitted FH-associated variants to ClinVar. Submitting center totals are listed per gene in **Table 4.3**.

Table 4.2. Number of variants submitted to ClinVar by gene.

	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>	Total
All variants submitted to ClinVar	5174	1003	474	6651
Variants detected in FH patients	4973	580	355	5908
Unique variants detected in FH patients	2314	353	216	2883

Table 4.3. Centers that submitted FH-associated variants to ClinVar.

Submitting Centers	Country	LDLR	APOB	PCSK9	Total
LDLR-Leiden Open Source Variation Database, British Heart Foundation	United Kingdom	1670	-	-	1670
Laboratory of Molecular Diagnostics, Vascular Medicine, Academic Medical Centre, University of Amsterdam	Netherlands	686	25	46	757
Centre of Molecular Genetics, Obesity and Dyslipidemias Unit, Pitié-Salpêtrière University Hospital	France	414	1	19	434
Cardiovascular Research Group, National Institute of Health Dr. Ricardo Jorge	Portugal	276	53	70	399
Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, Western University	Canada	202	137	30	369
Clinical Services Laboratory, Illumina	USA	97	180	85	362
Molecular Medicine of Metabolic Diseases Unit (U4M), University of Lille, Regional Hospital Center	France	344	-	-	344
Spanish Familial Hypercholesterolemia Foundation	Spain	320	10	1	331
Laboratory of Genetics and Molecular Cardiology, University of São Paulo	Brazil	201	63	16	280
Molecular Genetics Laboratory, Centre for Cardiovascular Surgery and Transplantation	Czech Republic	197	-	-	197
Invitae	USA	156	-	40	196
Cardiovascular Genetics Laboratory, PathWest Laboratory Medicine WA	Australia	152	-	-	152
Color Genomics	USA	23	65	25	113
Other	USA, Germany, Finland, India, South Korea	235	46	23	304

4.3.2 FH-associated variant characteristics

Unique FH-associated variants present on ClinVar are categorized by type for *LDLR*, *APOB* and *PCSK9* in **Table 4.4** and shown by location across all exons in **Figure 4.1**. Missense variants are the most prevalent unique variant type in each of the three genes, followed by frameshift variants in *LDLR*, and synonymous variants in both *APOB* and *PCSK9*. In *LDLR*, 18% of all unique variants are located in exon 4, in *APOB*, 41% are in exon 26 and 15% in exon 29, and in *PCSK9*, 19% are in the 3'UTR region.

Table 4.4. Unique FH-associated variants submitted to ClinVar by gene and variant type.

Variant Type	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>
3'UTR	77	9	40
5'UTR	54	4	18
Frameshift	430	12	1
In-frame indels	87	5	6
Intronic	48	3	26
Splicing	198	24	13
CNV (deletion)	100	-	-
CNV (duplication)	42	-	-
Missense	1011	218	82
Nonsense	179	4	1
Synonymous	83	74	28
<i>Cis</i> variants	5	-	-
Total	2314	353	216

In-frame indels: smaller than one exon; *cis* variants (single submission of two variants on same allele) include: three double missense, one in-frame indel + frameshift, one in-frame indel + missense; indel, insertion or deletion; UTR, untranslated region.

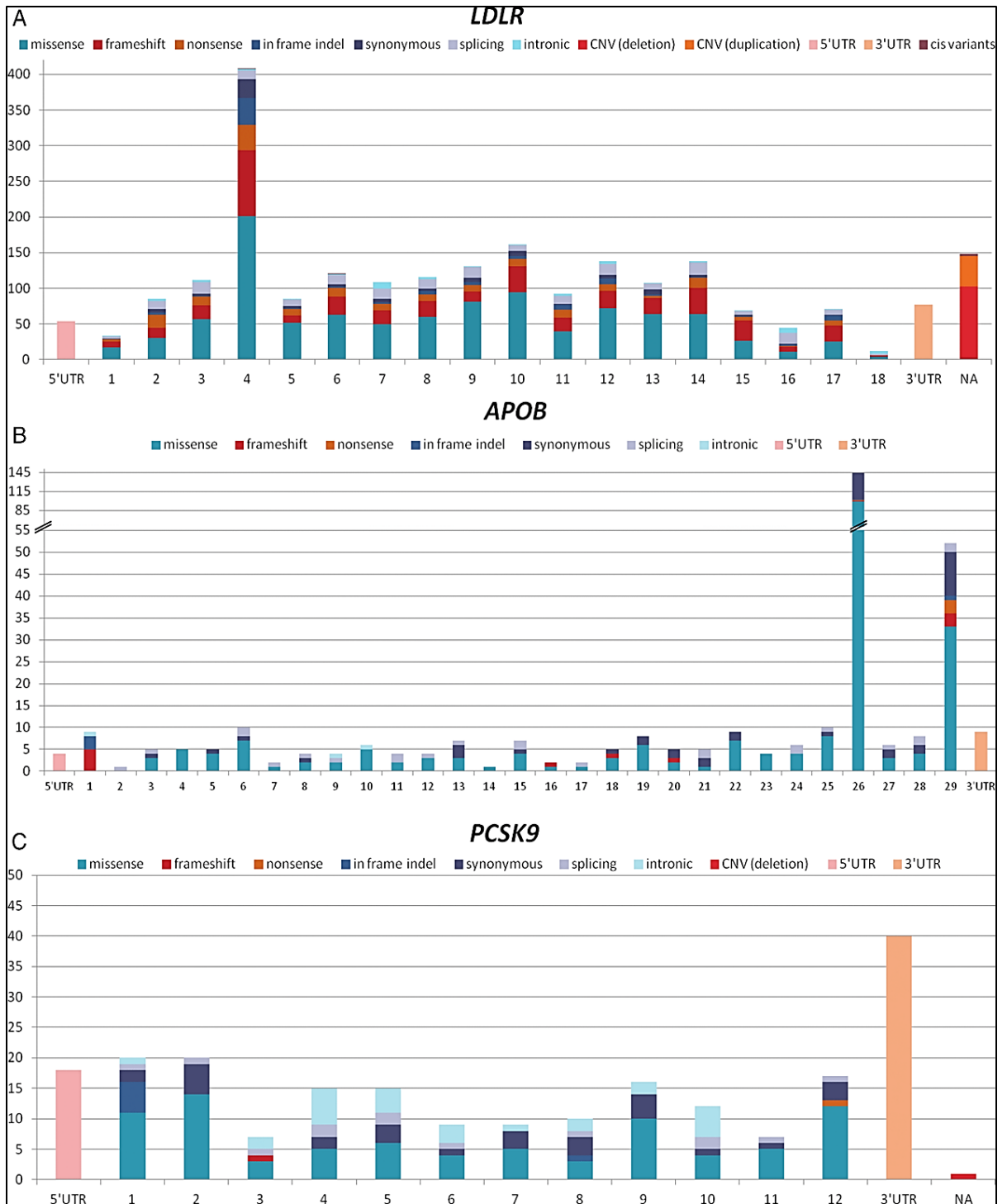


Figure 4.1. Number of unique FH-associated variants in ClinVar by exonic location and variant type. Variants in introns are represented in the closest exon. *Cis* variants: single submission with two variants in same allele; intronic: variants outside ± 15 nucleotides (nts) of intron/exon border; splicing: variants known to affect splicing or variants within ± 15 nts of intron/exon border. CNV, copy number variation; indel, insertion or deletion; NA, not applicable (variants spanning more than one exon); UTR, untranslated region.

Variants submitted to ClinVar range from benign to pathogenic or can be submitted without an assertion; with the exception of 198 FH-associated variant submissions, submitting centers provided a pathogenicity classification for their variants, found summarized by gene in **Table 4.5**. Unique variants are categorized by classification in **Table 4.6**; 57.9% (1670 of 2883) of these variants have been classified by submitters as pathogenic or likely pathogenic (or both, in cases of multiple submissions for the same variant), 15.5% (448 of 2883) have been classified as a variant of unknown significance (VUS) and 10.4% (299 of 2883) have been classified as benign or likely benign. The remaining 13.1 % of variants (379 of 2883) have conflicting classifications using a three-tier system (Benign/Likely benign + Uncertain significance; or Pathogenic/Likely pathogenic + Uncertain significance; or Benign/Likely benign + Pathogenic/Likely pathogenic).

Table 4.5. Clinical significance of *all* FH-associated variant submissions in ClinVar, independent of criteria used for classification.

Clinical significance	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>
Benign	205 (4.1%)	57 (9.8%)	88 (24.8%)
Likely benign	312 (6.3%)	97 (16.7%)	54 (15.2%)
Uncertain significance	526 (10.6%)	254 (43.8%)	132 (37.2%)
Likely pathogenic	1525 (30.7%)	10 (1.7%)	15 (4.2%)
Pathogenic	2351 (42.3%)	42 (7.2%)	42 (11.8%)
<i>Not provided</i>	<i>54 (1.1%)</i>	<i>120 (20.7%)</i>	<i>24 (6.8%)</i>
Total	4973	580	355

Table 4.6. Clinical significance of *unique* FH-associated variants in ClinVar, independent of criteria used for classification.

Clinical significance	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>
Benign/Likely benign	200 (8.7%)	44 (15.1%)	55 (26.8%)
Uncertain significance	182 (7.9%)	171 (58.6%)	95 (46.3%)
Pathogenic/Likely pathogenic	1614 (70.2%)	30 (10.3%)	26 (12.7%)
Conflicting classification	303 (13.2%)	47 (16.0%)	29 (14.2%)
<i>Not provided</i>	<i>15</i>	<i>61</i>	<i>11</i>
Total	2314	353	216

Conflicting classifications (considered for variants with multiple submissions):
 Benign/Likely benign + Uncertain significance; or Pathogenic/Likely pathogenic +
 Uncertain significance; or Benign/Likely benign + Pathogenic/Likely pathogenic.

4.3.3 Variant classification methods

A wide range of criteria have been used to classify FH-associated variants present on ClinVar. These include the general American College of Medical Genetics / Association for Molecular Pathology (ACMG/AMP) guidelines (2015), specified guidelines adhering to the ACMG/AMP framework, and a number of independent methods. The most used method was ACMG/AMP framework classification, followed by the Association for Clinical Genetic Science (ACGS) guidelines used in all LOVD transferred variants; a large number of variants (n=865) with classifications did not have indication of criteria used (**Table 4.7**). Most variants with multiple submissions have been classified using various different criteria (**Figure 4.2**). The specific criteria used by each submitter are listed in **Table 4.8**.

Table 4.7. Criteria used for FH-associated variant classifications in ClinVar.

Criteria used for classification	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>	Total
ACMG/AMP Guidelines	1144	127	99	1370
ACMG/AMP Framework	295	194	120	609
ACGS Guidelines	1669	-	-	1669
Independent methods	186	26	9	221
No criteria	793	25	47	865

ACMG/AMP framework: specified criteria adhering to the ACMG/AMP framework; No criteria: classification given but criteria used not provided. ACGS, Association for Clinical Genetic Science; ACMG/AMP, American College of Medical Genetics / Association for Molecular Pathology.

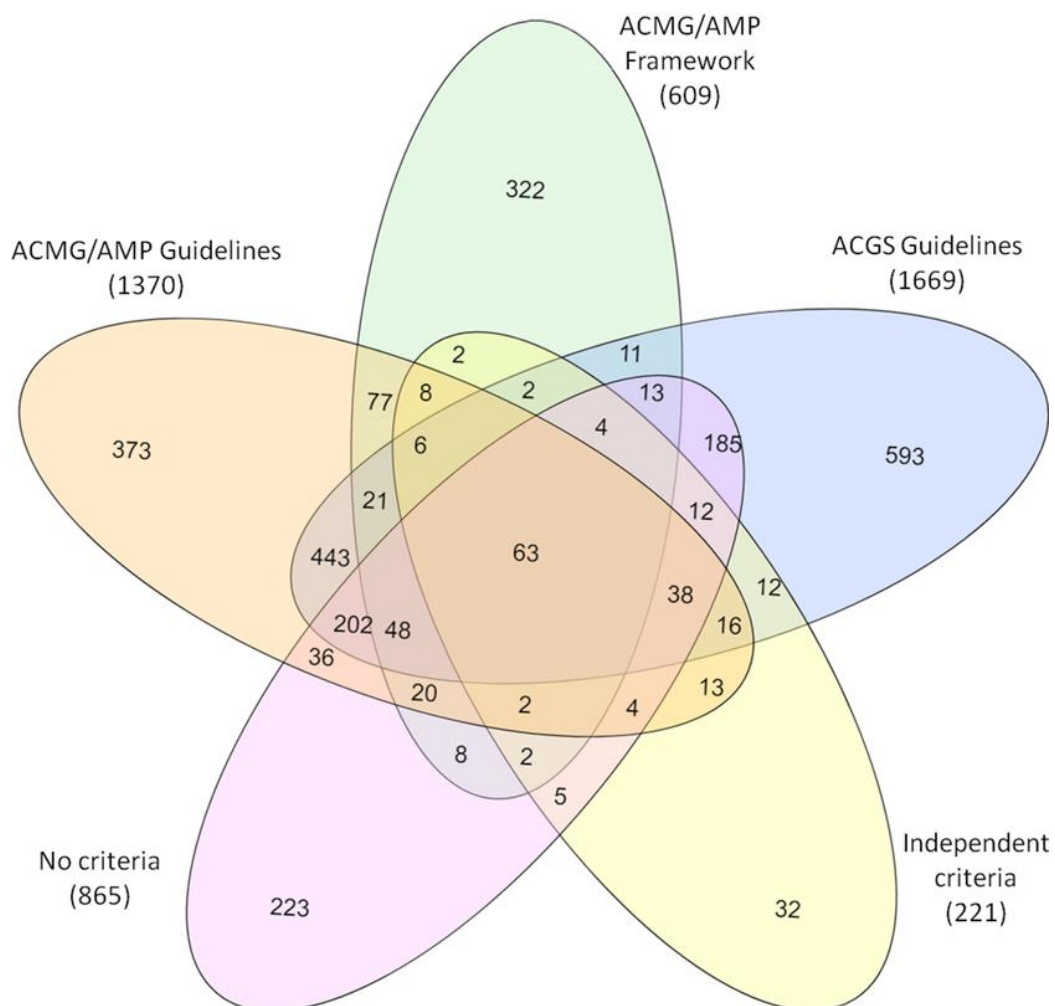


Figure 4.2. Number of unique variants ($n=2796$) classified by different sets of criteria. For 87 unique variants, no classification was submitted. ACMG/AMP framework: specified criteria adhering to ACMG/AMP framework; No criteria: classification given but criteria used not provided. ACGS, Association for Clinical Genetic Science; ACMG/AMP, American College of Medical Genetics / Association for Molecular Pathology.

Table 4.8. Criteria for variant classification used by each submitting center.

Submitting Centers	Country	Criteria
Centre of Molecular Genetics, Obesity and Dyslipidemias Unit, Pitié-Salpêtrière University Hospital	France	ACMG/AMP Guidelines (Richards et al., 2015)
Cardiovascular Research Group, National Institute of Health Dr. Ricardo Jorge	Portugal	ACMG/AMP Guidelines (Richards et al., 2015)
Molecular Medicine of Metabolic Diseases Unit (U4M), University of Lille, Regional Hospital Center	France	ACMG/AMP Guidelines (Richards et al., 2015)
Spanish Familial Hypercholesterolemia Foundation	Spain	ACMG/AMP Guidelines (Richards et al., 2015)
Laboratory of Genetics and Molecular Cardiology, University of São Paulo	Brazil	ACMG/AMP Guidelines (Richards et al., 2015)
Molecular Genetics Laboratory, Centre for Cardiovascular Surgery and Transplantation	Czech Republic	ACMG/AMP Guidelines (Richards et al., 2015)
Color Genomics	USA	ACMG/AMP Guidelines (Richards et al., 2015)
Knight Diagnostic Laboratories, Oregon Health and Sciences University	USA	ACMG/AMP Guidelines (Richards et al., 2015)
Phosphorus	USA	ACMG/AMP Guidelines (Richards et al., 2015)
Molecular Diagnostics Laboratory, Nemours Alfred I. duPont Hospital for Children	USA	ACMG/AMP Guidelines (Richards et al., 2015)
Institute of Medical Genetics and Genomics, Sir Ganga Ram Hospital	India	ACMG/AMP Guidelines (Richards et al., 2015)
Soonchunhyang University Medical Center	South Korea	ACMG/AMP Guidelines (Richards et al., 2015)
Clinical Services Laboratory, Illumina	USA	ICSL Variant Classification (ACMG/AMP framework; (Illumina Clinical Services Laboratory, 2016))
Invitae	USA	Invitae Variant Classification: Sherlock (ACMG/AMP framework; (Nykamp et al., 2017))

Laboratory Corporation of America	USA	LabCorp Variant Classification Specifications (ACMG/AMP framework; (Laboratory Corporation of America, 2015))
Division of Human Genetics & Genomic Diagnostics, Children's Hospital of Philadelphia	USA	DGD Variant Analysis Guidelines (ACMG/AMP framework; (Division of Genomic Diagnostics & The Children's Hospital of Philadelphia, 2015))
Cardiovascular Biomarker Research Laboratory, Mayo Clinic	USA	Mayo Cardiovascular Biomarkers Research Laboratory <i>LDLR</i> variant Interpretation Criteria (ACMG/AMP framework; (Kullo Laboratory, 2015))
Blueprint Genetics	Finland	Blueprint Variant Classification (ACMG/AMP framework; (Blueprint Genetics, 2016))
<i>LDLR</i> -Leiden Open Source Variation Database, British Heart Foundation	United Kingdom	ACGS Variant Guidelines (Wallis, Payne, McAnulty, & Bodmer, 2013)
Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, Western University	Canada	Independent method; Submitters publication (Wang et al., 2016)
Clinical Sequencing Exploratory Research, University of Washington	USA	Independent method; Literature (Amendola et al., 2015)
Institute for Integrative and Experimental Genomics, University of Luebeck	Germany	Independent method; Submitter's publication (Brænne et al., 2016)
Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Harvard Medical School	USA	Independent method; Submitter's publication (Duzkale et al., 2013)
SNPedia	USA	Independent method; Literature (Khera et al., 2016)
Laboratory of Molecular Diagnostics, Vascular Medicine, Academic Medical Centre, University of Amsterdam	Netherlands	None
Cardiovascular Genetics Laboratory, PathWest Laboratory Medicine WA	Australia	None

Online Mendelian Inheritance in Man (OMIM)	USA	None
GeneReviews	USA	None
Bioscience Institute for Medical Diagnostics, Sonic Healthcare	Germany	None
GenomeConnect	USA	None

ACMG/AMP, American College of Medical Genetics / Association for Molecular Pathology; ACGS, Association for Clinical Genetic Science.

4.3.4 Variant-level data

Some variants (n=1972 unique, 3435 submissions) were submitted with some kind of supporting variant-level data. This included information on patient clinical features, if there was family history of disease, the number of variant alleles or number of families with the variant identified, number of families with observed segregation, if it was an incidental finding and note of any related functional studies published (**Table 4.9**).

However, information of co-segregation was only submitted to ClinVar for eight variants, and phenotype data was only submitted for 490 unique variants (in 1043 total submissions). Functional studies were reported for 334 unique variants (437 submissions).

Table 4.9. Number of unique variants with each variant-level data type available in ClinVar.

Variant-level data submitted as evidence*	<i>LDLR</i>	<i>APOB</i>	<i>PCSK9</i>
Variant alleles/number of families with variant	1885	26	11
Clinical features/Family history	490	0	0
Incidental finding	344	0	0
Functional study	293	19	22
Number of families with observed segregation	8	0	0

*Labels extracted directly from ClinVar

4.4 Discussion

Data sharing through a centralized open-source database is essential to achieve accurate and consistent interpretation of variants identified during the course of genetic testing. Through the concerted efforts of the ClinGen FH VC-EP, submission of FH-associated variants to ClinVar from different global laboratories resulted in an increase of 10 times the number of unique variants reported during the past years. This was only possible due to a common effort and willingness to share internal data, and demonstrates the power of collaboration across patient-groups, academic labs, commercial labs and scientific funding bodies.

An extensive range of FH-associated variants are now present on ClinVar to aid with variant interpretation. The relative proportions of variants and variant-types per gene are consistent with what has been previously reported (Chora et al., 2018; Leigh et al., 2017). However, there are more known FH-associated variants identified in *LDLR*, *APOB* and *PCSK9* than previously thought. The FH literature has continued to cite a historical number of ~2000 FH-associated variants identified worldwide; however, with ~2900 presented here, this has now become outdated.

It is noteworthy that a number of variants with multiple submissions may include instances of “double counting”; a few FH centers have submitted a proportion of their variants to both the LOVD database (in the past) and ClinVar. While the exact number of these variants is presently unknown, efforts are underway to remove such cases. Secondly, the number of unique CNVs in *LDLR* (142; 100 deletions and 42 duplications) may be underestimated quite considerably. There have been 273 total CNV submissions, yet only 12 have defined breakpoints. This is a result of commonly applied detection methods such as MLPA (Wang, Ban, & Hegele, 2005), or more recently NGS depth of coverage analysis (Iacocca et al., 2017), which are limited to exon-level resolution. *LDLR* CNV submissions in ClinVar have thus largely been grouped by affected exon(s), but the likelihood of each breakpoint being identical in these “unique” CNV types is questionable. Previous breakpoint analysis has shown there are multiple unique CNV events which lead to the deletion of the 5'UTR–Exon 1 in *LDLR* (Hobbs et al., 1988) and the same may be true for other *LDLR* CNV types.

Only 10.7% of classified variants in *LDLR* have been considered as VUS by ClinVar submitters, compared to 55.2% and 39.9% VUS in *APOB* and *PCSK9*, respectively, suggesting potential pathogenicity is much more difficult to evaluate in *APOB/PCSK9* compared to *LDLR*. Because a loss-of-function in *LDLR* is a known disease mechanism of FH, any clearly deleterious variant-type in *LDLR* can be considered pathogenic. However, only very specific variants in *APOB* and *PCSK9* lead to FH. In *PCSK9*, causative variants must induce a gain of function in the encoded protein, and in *APOB*, causative variants must allow the production of the protein, but need to specifically alter the binding affinity to LDLR (known LDL binding domains are located within *APOB* exons 26 and 29). Generally, any null variant in these genes will lead to hypocholesterolemia, and thus are not expected to be identified in FH patients. This leaves most candidate *APOB* and *PCSK9* variants missense or synonymous, which pose challenges to interpretation. Further, some *APOB* variants have been shown to have low penetrance, adding another level of difficulty in interpreting variants in this gene (Alves et al., 2014). Accordingly, it is unwarranted to confidently classify variants as pathogenic in *APOB* and *PCSK9* without performing functional studies, leaving many of them as VUS.

This effort has also revealed that many different variant classification methods are being used, which is problematic since non-standardization can lead to different interpretation of identical variants. Indeed, we saw 379 variants (~15% of variants in each gene) with conflicting classifications. Use of ACMG/AMP guidelines aims to achieve greater standardization and consistency in variant interpretation (Richards et al., 2015). As we saw here, many FH research and diagnostic groups have adopted this new standard. However, the ACMG/AMP guidelines were designed to be generalizable to all Mendelian disorders, and ambiguities leave potential for differences in the application of various criteria among users, yielding inconsistent classifications. For instance, 114 unique variants have conflicting classifications despite all submitters having cited the ACMG/AMP guidelines.

Beyond a degree of inherent subjectivity, the current ACMG/AMP guidelines do not adequately address FH. In a separate study, ACMG/AMP classification of a large subset of FH-associated variants resulted in a large proportion of VUS (42% in *LDLR*, as

well as 90% in *APOB* and 92% in *PCSK9*) (Chora et al., 2018). Cases of misclassifications when compared against known pathogenic/benign variants were also found. One of ClinGen's key goals is the standardization of gene/disease-specific adjustments to the ACMG/AMP guidelines to address these issues, and to use these specified guidelines to provide a high level of confidence in ClinVar variant classifications.

Current ClinVar submissions point to specific issues that need to be addressed imminently in order to further improve the interpretation of FH-associated variants. First, clinical details accompanying a submission need to have minimum standards. Many *LDLR*, *APOB* and *PCSK9* variants were submitted without a disease association, rendering them of little value to interpretation efforts. Others were submitted with both hyper/hypocholesterolemia associations, and some had potentially incorrect disease associations – for example, deleterious/null variants in *APOB/PCSK9* submitted with a disease association of FH.

Second, richer supporting variant-level data must be submitted. Although FH centers successfully reported numerous variants, the same cannot be said concerning additional supporting variant-level data. Only eight variants had information about cosegregation, and patient phenotype descriptions were nearly nonexistent (e.g., no data on lipid profiles or cardiovascular disease). The large majority of submitters reported no functional studies for detected variants, although this is key to pathogenicity attribution, and are publicly available for more than 300 variants. The ACMG/AMP framework awards points to functional-level data, co-segregation data, normolipidemic data, and the number of observations/unrelated patients with each variant; if this information is kept stored in internal databases it will ultimately have a major negative impact on accurately interpreting ClinVar variants. Patient ethnicity would also be useful data, but was unreported.

All submitters should include supporting variant-level data for retrospective and prospective variant submission. Ideally, submissions should include a short summary of phenotype and genetic testing results for each individual, such as untreated LDL-C, the genes tested, and any other variants detected in the patient's sample. As an illustrative case, consider a patient who presents with an LDL-C value typical of heterozygous FH

and has a candidate variant in both *LDLR* and *APOB*. If the *LDLR* variant is clearly pathogenic (suggested by previous aggregate evidence) then this case-level information adds evidence to support the *APOB* variant being benign (if no other evidence is available to suggest otherwise). When two such variants are submitted separately outside the testing context, others might interpret the *APOB* variant as a VUS or perhaps even as pathogenic if it is the only variant ascertained in their patient and see it has been previously reported on the database. Such contextual interpretation is undoubtedly performed internally by diagnostic laboratories but is currently not part of any variant submission process, despite it being readily accessible at the time of submission.

Third, data submission needs to be ongoing. Although most of the world's largest laboratory repositories for FH variants have now made submissions to ClinVar, a few important populations remain outstanding; including Italy, Denmark, Norway, Germany, Israel and Japan. Efforts are underway to encourage outstanding centers to submit their variants, and it is imperative this is achieved prior to the reclassification of all variants using FH-specified ACMG/AMP criteria to ensure diverse representation is accounted for in the specification of these criteria. Further, FH-associated variants are likely being identified on an exponential scale as NGS panels are becoming increasingly implemented in routine FH diagnosis, a trend sure to continue as sequencing costs continue to plummet and awareness of FH broadens. Thus, real-time submission of variant data must be a focus for all centers.

4.5 Conclusion

Efforts of data sharing, and reliable variant interpretation, are extremely important to improve the care of FH patients. Since FH is so prominent in the population, and as educational efforts continue, more health care/family physicians can be expected to order genetic testing. As such, FH-associated variant submissions to ClinVar are likely to continue to increase. This will also increase the use of ClinVar as an essential resource for variant interpretation, ultimately affecting patient management and cascade screening. The ClinGen FH VC-EP will continue to encourage data sharing and communication between clinical and research FH experts in order to improve variant interpretation and harmonize FH diagnosis across the world.

4.6 References

- Alves, A. C., Etxebarria, A., Soutar, A. K., Martin, C., & Bourbon, M. (2014). Novel functional APOB mutations outside LDL-binding region causing familial hypercholesterolaemia. *Human Molecular Genetics*, *23*(7), 1817–1828.
- Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., ... Jarvik, G. P. (2015). Actionable exomic incidental findings in 6503 participants: Challenges of variant classification. *Genome Research*, *25*(3), 305–315.
- Blueprint Genetics. (2016). A guide to understanding variant classification. Retrieved from https://submit.ncbi.nlm.nih.gov/ft/byid/nxpnxkpc/variant_classification_wp_vara41-03.pdf
- Brønne, I., Kleinecke, M., Reiz, B., Graf, E., Strom, T., Wieland, T., ... Schunkert, H. (2016). Systematic analysis of variants related to familial hypercholesterolemia in families with premature myocardial infarction. *European Journal of Human Genetics*, *24*(2), 191–197.
- Chora, J. R., Medeiros, A. M., Alves, A. C., & Bourbon, M. (2018). Analysis of publicly available LDLR, APOB, and PCSK9 variants associated with familial hypercholesterolemia: Application of ACMG guidelines and implications for familial hypercholesterolemia diagnosis. *Genetics in Medicine*, *20*(6), 591–598.
- Division of Genomic Diagnostics, & The Children’s Hospital of Philadelphia. (2015). DGD Variant Analysis Guidelines. Retrieved from https://submit.ncbi.nlm.nih.gov/ft/byid/q5qzurm4/dgd_variant_analysis_guidelines.docx
- Duzkale, H., Shen, J., McLaughlin, H., Alfares, A., Kelly, M. A., Pugh, T. J., ... Lebo, M. S. (2013). A systematic approach to assessing the clinical significance of genetic variants. *Clinical Genetics*, *84*(5), 453–463.
- Harrison, S. M., Riggs, E. R., Maglott, D. R., Lee, J. M., Azzariti, D. R., Niehaus, A., ... Rehm, H. L. (2016). Using ClinVar as a resource to support variant interpretation. *Current Protocols in Human Genetics*, *2016*, 8.16.1-8.16.23.
- Hobbs, H. H., Leitersdorf, E., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1988). Multiple crm-mutations in familial hypercholesterolemia. Evidence for 13 alleles, including four deletions. *Journal of Clinical Investigation*, *81*(3), 909–917.
- Iacocca, M. A., Wang, J., Dron, J. S., Robinson, J. F., McIntyre, A. D., Cao, H., & Hegele, R. A. (2017). Use of next-generation sequencing to detect *LDLR* gene copy number variation in familial hypercholesterolemia. *Journal of Lipid Research*, *58*(11).

- Illumina Clinical Services Laboratory. (2016). ICSL Variant Classification. Retrieved from https://submit.ncbi.nlm.nih.gov/ft/byid/4jQgNGYk/ICSL_Variant_Classification_20161018.pdf
- Khera, A. V., Won, H. H., Peloso, G. M., Lawson, K. S., Bartz, T. M., Deng, X., ... Kathiresan, S. (2016). Diagnostic Yield and Clinical Utility of Sequencing Familial Hypercholesterolemia Genes in Patients With Severe Hypercholesterolemia. *Journal of the American College of Cardiology*, *67*(22), 2578–2589.
- Kullo Laboratory. (2015). Kullo Lab Assertion Criteria. Retrieved from https://submit.ncbi.nlm.nih.gov/ft/byid/RJ4WMVct/Kullo_Lab_Assertion_Criteria_01072016.pdf
- Laboratory Corporation of America. (2015). LabCorp: Variant Classification Specifications.
- Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(D1), D980-5.
- Leigh, S., Futema, M., Whittall, R., Taylor-Beadling, A., Williams, M., den Dunnen, J. T., & Humphries, S. E. (2017). The UCL low-density lipoprotein receptor gene variant database: Pathogenicity update. *Journal of Medical Genetics*, *54*(4), 217–223.
- Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y. Y., ... Topper, S. (2017). Sherlock: A comprehensive refinement of the ACMG-AMP variant classification criteria. *Genetics in Medicine*, *19*(10), 1105–1117.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... Rehm, H. L. (2015). 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. *Genetics in Medicine*, *17*(5), 405–424.
- Wallis, Y., Payne, S., McAnulty, C., & Bodmer, D. (2013). Practice guidelines for the evaluation of pathogenicity and the reporting of sequence variants in clinical molecular genetics. *Association for Clinical Genetic Science*, 1–16.
- Wang, J., Ban, M. R., & Hegele, R. A. (2005). Multiplex ligation-dependent probe amplification of *LDLR* enhances molecular diagnosis of familial hypercholesterolemia. *Journal of Lipid Research*, *46*(2), 366–372.
- Wang, J., Dron, J. S., Ban, M. R., Robinson, J. F., McIntyre, A. D., Alazzam, M., ...

Hegele, R. A. (2016). Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 36(12), 2439–2445.

Chapter 5 – Discussion and conclusions

5.1 Overview

FH is a highly prevalent inherited dyslipidemia characterized by a lifelong exposure to elevated LDL-C levels with increased risk of premature atherosclerosis causing cardiovascular disease (Hegele, 1997). Despite the number of effective medications widely available, FH is severely underdiagnosed (<1% in most countries) and undertreated (Nordestgaard et al., 2013).

As molecular genetic technologies – namely next-generation sequencing - have become increasingly more affordable and available in recent years, FH has moved to the very forefront of precision medicine. Genetic testing has now become a central part of FH diagnosis globally. There are many important advantages in obtaining a genetic diagnosis of FH, which include: 1) achieving certainty in the context of incomplete clinical criteria, such as reduced prevalence of typical physical findings and/or unattainable or unreliable family history; 2) motivating cascade screening in family members, which markedly improves the rate of underdiagnosis and can be seen as a cornerstone of cardiovascular prevention; 3) directing specific therapeutic strategies for improved patient management; 4) to improve treatment compliance; and 5) to support insurance coverage of certain medications.

My thesis focused on improving the strategies used to provide genetic diagnosis of FH; this was approached in three main ways as discussed below.

5.2 Study findings and implications

In chapter 2, we investigated the potential for NGS data to be bioinformatically processed for the detection of CNVs in *LDLR*. In analysis of 388 FH patient samples, we found 100% concordance in *LDLR* CNV detection between MLPA - the gold standard or

reference standard method - and our new NGS approach. This demonstrated the ability of NGS with appropriate bioinformatics to identify both small and large-scale variants in a single platform and analytic procedure.

Accurate identification of *LDLR* CNVs from NGS data is important because this class of variation comprises a significant proportion of FH cases but not all sequencing facilities have the resources, time or interest to establish a parallel MLPA system for detecting them. Our cost for MLPA analysis in *LDLR* - including reagents, controls, duplicate analyses, and labor - was ~\$110 CAD per patient sample, which totaled ~\$43,000 CAD for this cohort of 388 FH individuals. These costs are essentially eliminated when applying a bioinformatics method to NGS data; since such data are already being generated for small-scale variant analysis that precedes CNV assessment. We have found that once established, the bioinformatics workflow for CNV detection takes only an additional ~10 minutes for a set of 24 samples.

Because of the 100% sensitivity and specificity demonstrated by our method specifically, it can also be clinically reliable; this approach is now being implemented in the diagnosis of FH in Canada. Since our initial report in 2017, there have been 32 more *LDLR* CNVs identified in Canadian FH patients. Ultimately, transitioning to an NGS-based approach for CNV analysis in *LDLR* will promote more widespread assessment of this important class of variation across diagnostic laboratories in the future.

In chapter 3 we performed CNV screening in additional FH-associated genes *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, and *ABCG5/8* in a cohort of 704 FH individuals using our newly demonstrated NGS bioinformatic method. Interestingly, we identified a whole-gene duplication of *PCSK9* in two index cases, while we found no CNVs in *APOB*, *LDLRAP1*, *APOE*, *STAP1*, *LIPA*, or *ABCG5/8*.

CNVs in these seven genes had remained completely uninvestigated in FH; since collectively these genes are implicated in fewer than 10% of known monogenic FH cases, CNV analysis by traditional methods has been considered far too laborious and costly to perform, especially in large clinical cohorts. By using bioinformatic tools to evaluate existing targeted NGS data, surveying these genes for CNVs can now be considered economical. It is the goal of any diagnostic laboratory to account for all genetic abnormalities capable of explaining FH cases; this helps to avoid the possibility of false-

negative genetic diagnoses. To this point, the two index cases who were positive for a causative *PCSK9* duplication identified by this extended CNV screening, had previously no genetic aberrations to explain their phenotype, despite having genetic testing done in 2014 and 2015, respectively. This finding also has therapeutic relevance, as elevated PCSK9 levels due to an increase in gene dosage may limit the efficacy of high-dose statin therapy and also PCSK9 inhibition. Indeed, the individuals under study with this PCSK9 duplication had poor response to statin therapies and LDL-C target levels were unattainable, leaving them potentially exposed to continued progression of atherosclerosis. As demonstrated in index case A, increasing PCSK9 inhibition to a high dose was necessary for improved LDL-C management.

In chapter 4, we took part in a large global collaboration effort to establish a centralized database of clinically-relevant genetic variants identified in FH patients. Together, this effort facilitated the submission of 5487 total and 2803 unique FH-associated variants from 30 different centres in 13 countries, of which we contributed 369 variants - the 4th largest single-center contribution.

Prominent laboratories performing sequencing analysis of FH samples around the world have traditionally kept large numbers of variants and variant-level data in internal databases. This practice can lead to inconsistencies in the way two different laboratories interpret the same or similar variants; use of open-source data is essential for accurate and concordant variant classification (i.e. “pathogenic”; “unknown significance”; or “benign”). There is now an extensive range of FH-associated variants present on the ClinVar database to aid in interpretation efforts.

Analysis of all 5908 total and 2883 unique FH-variant submissions presently on ClinVar has proven that there is little standardization in the methods and criteria used amongst laboratories to classify these variants. The ClinGen FH Variant Curation Expert Panel (FH VC-EP) has been working diligently over the past 2 years to establish a set of FH-specific variant classification guidelines, which adhere to the general ACMG/AMP framework. I have been fortunate enough to be involved with the collaborative discussions, drafting, and testing of these new FH-specific ACMG/AMP guidelines over the last two years, and have been signified as a co-first author for my contributions (manuscript currently in preparation). Establishment of these guidelines will help to

achieve a more evidence-based, standardized method for the classification of variants detected in FH patients; both retrospectively and prospectively. Following the approval of these guidelines by the ClinGen Clinical Domain Working Group Oversight Committee, we will use them to re-classify all 2883 unique FH-associated variants present on the ClinVar database. FH variant classifications which have been classified by these FH-specified ACMG/AMP guidelines and approved by the ClinGen FH VC-EP will receive “3-star” status in the ClinVar database, which signifies Expert Panel-level confidence. As of December 5, 2018, ClinVar became the very first public database containing information about genes, variants, and their relationship to disease to be approved by the U.S. Food and Drug Administration (FDA). Specifically, “3-star” level variant classifications on ClinVar will now have an associated FDA tag, allowing them to be recognized as a source of valid scientific evidence that can be used to support clinical validity. This announcement has been considered an immense achievement for the ClinGen consortium due to its implications on further advancing the field of precision medicine.

5.3 Future directions

Despite advancements in CNV screening and interpretation of DNA variants identified during the course of genetic testing, multiple aspects of FH diagnosis can be further improved. Most notably, ~20% of patients with a diagnosis of FH by clinical criteria have no identifiable genetic basis (monogenic or polygenic) to explain their phenotype; it is of interest to address this remaining “missing heritability”.

First, it is possible that rare causative variants exist in novel genes. Individuals with clear-cut “definite” FH (DFH; as diagnosed by Simon Broome or Dutch Lipid Clinic Network criteria; see Chapter 1 Table 1.1) but no identifiable causal variants in canonical FH-genes are prime candidates for discovery of novel FH-associated loci. Although, it is noteworthy that such an approach has been previously unavailing. Futema et al. (2014) performed whole-exome sequencing (WES) in a large cohort of 125 DFH patients negative for *LDLR/APOB/PCSK9* variants but identified no putative FH-

associated loci, suggesting that the genetic determinants in unexplained FH cases are likely to be highly heterogeneous, complicating the gene discovery process. Systematic study of larger DFH cohorts may be required, however, these are presumably difficult to acquire since 80% of DFH patients are found to be causative variant-positive by conventional screening in known FH-genes.

Second, increased availability of whole-genome sequencing (WGS) is likely to address the possibility that causative variants in non-coding regions, such as introns, exist in some FH individuals. The first example of this was reported by Reeskamp et al. in 2018, who performed WGS in a large family with an unexplained autosomal dominant FH trait; they identified a novel single nucleotide variant (c.2140+103G>T) located deep within an intron of *LDLR* as the causal variant. The challenge with applying WGS to capture intronic and/or intergenic variation is that even when potential disease-causing variants are identified, they are often inherently difficult to interpret, at least in the absence of functional assays. Moreover, interpretation efforts are further compounded by the increased scale of variants likely to be identified, since intronic regions are generally variant-rich.

Another potential non-coding element worthy of investigation are microRNAs (miRNAs), which are a class of short regulatory RNAs (20-24 nucleotides) known to modulate mRNA levels (Wagschal et al., 2015). At least four miRNAs (miR-128-1, miR-148a, miR-130b, and miR-301b) have been predicted to regulate protein expression of *LDLR*, thus variation disrupting function in any one of these miRNAs may elicit a disease phenotype (Wagschal et al., 2015).

Third, epigenetic modifications, which can impact gene expression, are a potential mechanism that may underlie high LDL-C in variant-negative FH patients. Methods such as sodium bisulfite sequencing (SBS) have been developed to detect epigenetic modifications, however, epigenetic changes are often cell-type specific (Huang, Jiang, & Zhang, 2014; Li & Tollefsbol, 2011). With the liver being the principal organ where epigenetic effects would be believed to impact LDL-C levels most significantly, SBS and/or gene expression analysis used to reveal epigenetic modifications causing high LDL-C may require a biopsy of these cells, which would be far too invasive, especially in the context of a diagnostic assay.

5.4 Final conclusions


Advances in FH diagnosis have been propagated by an improved understanding of the underlying genetic determinants together with substantially reduced costs associated with appropriate screening strategies. Here, we have demonstrated the applicability of NGS techniques to reliably detect CNVs in *LDLR*, and to further perform cost-effective CNV screening in additional FH-associated genes, expanding the ability to account for all genetic abnormalities capable of underlying FH cases. Lastly, to improve the interpretation of variants identified during genetic diagnosis, we have made a significant contribution to establishing a centralized publicly-available database, that will serve as an essential resource for harmonizing the interpretation of variants identified in FH patients worldwide.

5.5 References

- Futema, M., Plagno, V., Li, K. W., Whittall, R. A., Neil, H. A. W., Seed, M., ... Humphries, S. E. (2014). Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations. *Journal of Medical Genetics*, *51*(8), 537–544.
- Hegele, R. A. (1997). Candidate genes, small effects, and the prediction of atherosclerosis. *Critical Reviews in Clinical Laboratory Sciences*, *34*(4), 343–367.
- Huang, B., Jiang, C., & Zhang, R. (2014). Epigenetics: The language of the cell? *Epigenomics*, *6*(1), 73–88.
- Li, Y., & Tollefsbol, T. O. (2011). DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods in Molecular Biology*, *791*, 11–21.
- Nordestgaard, B. G., Chapman, M. J., Humphries, S. E., Ginsberg, H. N., Masana, L., Descamps, O. S., ... Tybjaerg-Hansen, A. (2013). Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: Guidance for clinicians to prevent coronary heart disease. *European Heart Journal*, *34*(45), 3478–3490.
- Reeskamp, L. F., Hartgers, M. L., Peter, J., Dallinga-Thie, G. M., Zuurbier, L., Defesche, J. C., ... Hovingh, G. K. (2018). A Deep Intronic Variant in *LDLR* in Familial Hypercholesterolemia. *Circulation: Genomic and Precision Medicine*, *11*(12).
- Wagschal, A., Najafi-Shoushtari, S. H., Wang, L., Goedeke, L., Sinha, S., Delemos, A. S., ... Näär, A. M. (2015). Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. *Nature Medicine*, *21*(11), 1290–1297.

Appendices

Appendix A - Ethics approval

 Western Research
Date: 4 October 2018
To: Robert Hegele
Project ID: 0379
Study Title: Candidate gene sequencing, genetic and genomic analysis for identification of new genetic determinants of intermediate traits of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population. 07920E
Application Type: Continuing Ethics Review (CER) Form
Review Type: Delegated
REB Meeting Date: 06/Nov/2018
Date Approval Issued: 04/Oct/2018
REB Approval Expiry Date: 03/Nov/2019
<hr/>
Dear Robert Hegele,
The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.
REB members involved in the research project do not participate in the review, discussion or decision.
Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.
Please do not hesitate to contact us if you have any questions.
Sincerely,
Daniel Wyzynski, Research Ethics Coordinator, on behalf of Dr. Joseph Gilbert, HSREB Chair
<i>Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).</i>

Appendix B – Journal copyright approval

This Agreement between Mr. Michael Iacocca ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.	
License Number	4483810159580
License date	Dec 07, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cardiology
Licensed Content Title	The complex molecular genetics of familial hypercholesterolaemia
Licensed Content Author	Amanda J. Berberich et al
Licensed Content Date	Jul 4, 2018
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Improving the molecular diagnosis of familial hypercholesterolemia
Institution name	Western University
Expected presentation date	Jan 2019
Portions	Figure 1. LDL-receptor life cycle

Note: All first-authored publications included in this MSc thesis are permitted to be used by the authors with attribution to the original source.

Curriculum Vitae

Name: Michael Iacocca

Post-secondary Education and Degrees:

Bachelor of Science- Honors Specialization in Genetics 2012–2016
 Western University, London, Ontario
 University of Copenhagen, Denmark (one-term exchange program; 2013)

Masters of Science- Biochemistry 2016–present
 Western University, London, Ontario

Honors and Awards:

- 1) Top Oral Presentation Award; Canadian Lipoprotein Conference 2018/06
- 2) Top Poster Presentation Award; Robarts Research Retreat 2018/06
- 3) CIHR Circulatory and Respiratory Health Travel Award 2018/04
- 4) Canadian Society of Atherosclerosis Thrombosis & Vascular Biology Travel Award 2018/03
- 5) Golden Helix Abstract Competition 2017/2018: 1st Place 2018/02

Related Work Experience:

Teaching Assistant, Western University

- 1) Biology 1001A/1201A 2017/09-2017/12
- 2) Biology 1002B/1202B 2018/01-2018/04
- 3) Biology 1001A/1201A 2018/09-2018/12

Publications:

- 1) Recent advances in genetic testing for familial hypercholesterolemia.
Michael A. Iacocca and Robert A. Hegele
Expert Review in Molecular Diagnostics. 17:641-651. (Published 2017/05).
- 2) Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia.
Michael A. Iacocca, Jian Wang, Jacqueline S. Dron, John F. Robinson, Adam D. McIntyre, Henian Cao, and Robert A. Hegele.
Journal of Lipid Research. 58:2202-2209. (Published 2017/11).

- 3) Role of DNA copy number variation in dyslipidemias.
Michael A. Iacocca and Robert A. Hegele.
Current Opinion in Lipidology. 29:125-132. (Published 2018/04).
- 4) Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia.
Jacqueline S. Dron, Jian Wang, Amanda J. Berberich, **Michael A. Iacocca**, Henian Cao, Ping Yang, Joan H. Knoll, Karine Tremblay, Diane Brisson, Christian Netzer, Ioanna Gouni-Berthold, Daniel Gaudet, and Robert A. Hegele.
Journal of Lipid Research. 59:1529-1535. (Published 2018/08).
- 5) Whole-gene duplication of PCSK9 as a novel genetic mechanism for severe familial hypercholesterolemia.
Michael A. Iacocca, Jian Wang, Samantha Sarkar, Jacqueline S. Dron, Thomas Lagace, Adam D. McIntyre, Paulina Lau, John F. Robinson, Ping Yang, Joan H. Knoll, Henian Cao, Ruth McPherson, and Robert A. Hegele.
Canadian Journal of Cardiology. 34:1316-1324. (Published 2018/08).
- 6) ClinVar database of global familial hypercholesterolemia-associated DNA variants.
Michael A. Iacocca, Joana R. Chora, Alain Carrie, Tomas Freiburger, Sarah E. Leigh, Joep C. Defesche, Lisa C. Kurtz, Marina T. DiStefano, Raul D. Santos, Steve E. Humphries, Pedro Mata, Cinthia E. Jannes, Amanda J. Hooper, Katherine A. Wilemon, Pascal Benlian, Robert O' Connor, John Garcia, Hannah Wand, Lukas Tichy, Eric J. Sijbrands, Robert A. Hegele, Mafalda Bourbon, and Joshua W. Knowles: On behalf of the ClinGen FH Variant Curation Expert Panel.
Human Mutation. 39:1631-1640. (Published 2018/08).
- 7) Informing ASPIRE and future a student-run clinic: healthcare needs assessment of London, Ontario.
Michael A. Iacocca, Yara Abou-Hamde, Ruth Fisher, Adam Hopfgartner, and Bobbak Makooie.
University of Western Ontario Medical Journal. (Submitted 2018/08; ID:4869).
- 8) Severe hypertriglyceridemia is primarily polygenic.
Jacqueline S. Dron, Jian Wang, Henian Cao, Adam D. McIntyre, **Michael A. Iacocca**, Jyler R. Menard, Irina G. Movsesyan, Mary J. Malloy, Clive R. Pullinger, John P. Kane, and Robert A. Hegele.
Journal of Clinical Lipidology. 13:80-88. (Published 2018/10).
- 9) Progress in finding pathogenic DNA copy number variations in dyslipidemia.
Michael A. Iacocca, Jacqueline S. Dron, and Robert A. Hegele.
Current Opinion in Lipidology. (Published 2019/01; Epub ahead of print).

- 10) Efficacy of evolocumab in monogenic versus polygenic hypercholesterolemia
 Timothy Lee, **Michael A. Iacocca**, Matthew R. Ban, and Robert A. Hegele.
Canadian Journal of Cardiology Open. (Accepted 2019/02; ID:CJCO-D-19-00008).

Abstracts and Presentations at Scientific Meetings, Peer Reviewed

- 1) European Atherosclerosis Society Congress 2017, Prague, Czech Republic
 Poster: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
 Abstract published: *Atherosclerosis*. 263: e236-e236.
- 2) American Heart Association ATVB/PVD Scientific Sessions 2017, Minneapolis, USA
 Poster: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
 Abstract published: *Arteriosclerosis, Thrombosis, and Vascular Biology*. 37: A174-A174.
- 3) Curating the Clinical Genome 2017, Washington, USA
 Poster: Progress in ACMG/AMP-adapted guidelines for standardized variant curation in familial hypercholesterolemia.
- 4) Canadian Lipoprotein Conference 2017, Ottawa, Canada
 Oral: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
- 5) London Health Research Day 2017, London, Canada
 Poster: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
- 6-7) Robarts Research Retreat 2017, London, Canada
 Poster1: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
 Poster2: Progress in ACMG/AMP-adapted guidelines for standardized variant curation in familial hypercholesterolemia.
- 8) European Atherosclerosis Society Congress 2018, Lisbon, Portugal
 Science at a Glance (discussed e-Poster): DNA copy number variation screening in familial hypercholesterolemia-related genes.
 Abstract published: *Atherosclerosis*. 275: e79-e79.

- 9) Department of Medicine Research Day 2018, Western University, London, Canada
Oral: Use of next-generation sequencing to detect copy number variants in the routine molecular diagnosis of familial hypercholesterolemia.
- 10) London Health Research Day 2018, London, Canada
Poster: Adaptation of ACMG/AMP guidelines for standardized variant interpretation in familial hypercholesterolemia.
- 11) Robarts Research Retreat 2018, London, Canada
Poster: DNA copy number variation screening in familial hypercholesterolemia-associated genes.
- 12) Canadian Lipoprotein Conference 2018, Toronto, Canada
Oral: Whole-gene duplication of *PCSK9* as a novel genetic mechanism for severe familial hypercholesterolemia.
- 13–15) International Symposium on Atherosclerosis 2018, Toronto, Ontario
Oral: Whole-gene duplication of *PCSK9* as a novel genetic mechanism for severe familial hypercholesterolemia.
Poster1: Adaptation of ACMG/AMP guidelines for standardized variant interpretation in familial hypercholesterolemia.
Poster2: Genetic determinants of cardiovascular risk in familial hypercholesterolemia.
Abstracts (3) published: *Atherosclerosis Supplements*. 32: 9-10; 51-51; 53-53.

Abstracts and Presentations, Non Peer Reviewed

- 1) Robarts Molecular Medicine Data Club 2017, London, Canada
Oral: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
- 2) Biochemistry Research Showcase 2017, Western University, London, Canada
Poster: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
- 3) Western University Health and Research Conference 2017, London, Canada
Poster: Progress in ACMG/AMP-adapted guidelines for standardized variant curation in familial hypercholesterolemia.
- 4) Familial Hypercholesterolemia Global Summit 2017, Miami, USA

Oral (*FH Variant Curation Meeting): Progress in ACMG/AMP-adapted guidelines for standardized variant curation in familial hypercholesterolemia.

- 5) Biochemistry Research Symposium 2018, Western University, London, Canada
Oral: Use of next-generation sequencing to detect copy number variants in the molecular diagnosis of familial hypercholesterolemia.
- 6) Robarts Molecular Medicine Data Club 2018, London, Canada
Oral: Whole-gene duplication of *PCSK9* as a novel genetic mechanism for severe familial hypercholesterolemia.
- 7) Kiwanis Senior's Centre, London, Canada (2018)
Oral: Improving the diagnosis of familial hypercholesterolemia.