

---

Electronic Thesis and Dissertation Repository

---

12-17-2020 2:00 PM

# The Genetic Determinants of Complex Lipid and Lipoprotein Phenotypes

Jacqueline S. Dron, *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 Doctor of Philosophy degree in Biochemistry

© Jacqueline S. Dron 2020

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



Part of the [Genetics Commons](#)

---

## Recommended Citation

Dron, Jacqueline S., "The Genetic Determinants of Complex Lipid and Lipoprotein Phenotypes" (2020). *Electronic Thesis and Dissertation Repository*. 7522.  
<https://ir.lib.uwo.ca/etd/7522>

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](mailto:wlsadmin@uwo.ca).

## Abstract

Cardiovascular disease (CVD) is the primary cause of death globally and is estimated to cause one-third of deaths in Canada. Each year, millions of Canadians are affected by CVD despite ongoing efforts to reduce risk through lifestyle modifications and pharmacological therapies. With the expected rise in CVD prevalence due to the obesity epidemic, we need to better understand the genetic basis of heritable, modifiable risk factors, including levels of high-density lipoprotein (HDL) cholesterol and triglyceride, for insights into future therapeutic treatments and risk prediction. Through the use of a targeted next-generation sequencing panel designed specifically to study lipid and metabolic disorders, I have explored a spectrum of genetic variation—including rare and common variants, single-nucleotide and copy-number variants—in over 3,000 DNA samples isolated from individuals with abnormal lipid phenotypes, including: (i) hypoalphalipoproteinemia; (ii) hyperalphalipoproteinemia; and (iii) hypertriglyceridemia. From my research efforts, I demonstrated that the majority of individuals with abnormal HDL cholesterol levels did not carry many phenotypically-relevant genetic factors, but in those who did, rare variants were more prevalent in individuals with extremely low HDL cholesterol levels, while both rare variants and the accumulation of common variants were approximately equal in individuals with extremely high HDL cholesterol levels. Meanwhile, hypertriglyceridemia had a stronger genetic basis, with common variant accumulation being the most prevalent genetic determinant. Further, I uncovered that genetic determinants are more prevalent as the hypertriglyceridemia phenotype becomes more severe, and a genetic locus, *CREB3L3*, may have an extremely important, previously unappreciated role in hypertriglyceridemia susceptibility. By better understanding the genetic underpinnings of abnormal levels of HDL cholesterol and triglyceride, future efforts can explore the relationship between these phenotypes and their genetic determinants, and how we might leverage this information to develop better therapeutics to lower levels of these risk factors or create screening methods to identify individuals who might be at higher risk for CVD.

## Keywords

Complex traits and diseases, dyslipidemia, genetics, high-density lipoprotein (HDL) cholesterol, hyperalphalipoproteinemia, hypertriglyceridemia, hypoalphalipoproteinemia, next-generation sequencing (NGS), polygenic risk scores, triglyceride.

## Summary for Lay Audience

Heart disease is the second leading cause of death in Canada and affects millions of individuals each year. Despite efforts to reduce disease risk with healthy lifestyles and medications, these strategies are not always successful. The variation in effectiveness may be linked to differences in an individual's genetic make-up (i.e. DNA), and how these changes in DNA might be impacting levels of well-established risk factors for heart disease, such as: high-density lipoprotein (HDL) cholesterol—also referred to as the “good” cholesterol—and triglyceride (i.e. fats). These two lipid factors have been observed to associate with heart disease risk, and medications have been designed specifically to alter these lipid levels to reduce disease risk. Here, I worked to better understand the different DNA changes, also called “genetic variants”, that can influence levels of HDL cholesterol and triglyceride, and to specifically study the genetic variants in individuals with extreme lipid disorders characterized by either: (i) extremely low HDL cholesterol levels; (ii) extremely high HDL cholesterol levels; or (iii) extremely high triglyceride levels. After studying the DNA of over 3,000 individuals, I determined that each lipid disorder has a unique combination of rare and common genetic variants that help drive the presentation of each extreme lipid trait. During this research, I was also able to create two “genetic risk scores”—a method to aggregate information from many sites of common DNA variation into a single measure of disease risk—for both HDL cholesterol and triglyceride. From my collective research efforts, we now have a better understanding of the different DNA changes that can cause or increase risk for different lipid disorders, each of which have varying degrees of heart disease risk. By understanding the relevant genetic variants underlying lipid disorders involving abnormal levels of HDL cholesterol and triglyceride, future research efforts can explore how we might be able to take advantage of this information to develop better medications and therapeutics to lower levels of these heart disease risk factors or create genetic screening methods to identify individuals at higher risk for heart disease because of different types of genetic variation.



## Co-Authorship Statement

For each manuscript incorporated into my Dissertation, I am the first-listed author and have indicated which manuscripts were used in each Chapter. In each listed research article, I contributed towards the study's design, identified all rare variants of interest, developed and calculated all polygenic risk scores (when applicable), performed all data analyses, and wrote each manuscript. In each listed review, I wrote more than 85-90% of each manuscript.

My supervisor, Dr. Robert A. Hegele provided: (i) excellent supervision and guidance; (ii) funding support; and (iii) patient samples for all research analyses. He also contributed towards: (i) study design; (ii) manuscript preparation; and (iii) manuscript revisions. These contributions were consistent across all manuscripts.

Technical assistance for each research manuscript was provided by core members of the Hegele Lab, including: (i) Adam D. McIntyre, for extracting and isolating DNA from the patient samples under study; (ii) Dr. Henian Cao, for preparing the DNA samples for sequencing; (iii) Dr. Jian Wang, for providing access to clinical databases and validating the copy-number variants (CNVs) described in [Chapter 3](#) and [Chapter 5](#); (iv) John F. Robinson, for support in project management; (v) Matthew R. Ban, for support in lab management; (vi) Ericka Simon, for providing clinical demographic information for patients from the Lipid Genetics Clinic; and (vii) Brooke Kennedy, for maintaining all ethical protocols and study approvals.

Additional support for varying research manuscripts was provided by other members of the Hegele Lab and collaborators at Western University, including: (i) Dr. Amanda J. Berberich, for characterizing the clinical features of patients in [Chapter 3](#); (ii) Michael A. Iacocca, for assisting in the analysis and interpretation of CNVs in [Chapter 3](#); (iii) Allison A. Dilliott, for assisting in principle-component analyses in [Chapter 7](#); (iv) Jyler R. Menard, for assisting in the optimization of triglyceride polygenic risk scores in [Chapter 4](#); (v) Brent D. Davis and Dr. Dan J. Lizotte, for efforts in developing and launching the Exautomate pipeline (**Appendix L**); (vi) Arden Lawson, for beta-testing the Exautomate pipeline; and (vii) Drs. P. Yang and Joan Knoll, for CNV validation using the Affymetrix CytoScan HD array described in [Chapter 3](#).

External collaborators were also instrumental for the completion of many research studies, including: (i) Drs. Cecile Low-Kam, David Rhainds, Marie-Pierre Dubé, Guillaume Lettre, Jean-Claude Tardif (Montréal Heart Institute) and Drs. Sumeet A. Khetarpal, Dan J. Rader (University of Pennsylvania), for contributing sequence data for additional patients with extreme high-density lipoprotein (HDL) cholesterol levels in Chapter 2; (ii) Drs. Karine Tremblay, Diane. Brisson, Christian Netzer, Ioanna Gouni-Berthold, Daniel Gaudet (Université de Montréal), for sharing clinical information related to a hypoalphalipoproteinemia patient carrying a CNV in Chapter 3; (iii) Drs. John P. Kane, Clive R Pullinger, Mary J Malloy, Irene Movsesyan, and James Feng (University of California, San Francisco) for providing DNA samples and clinical information for patients with severe hypertriglyceridemia in Chapter 4, Chapter 5, and Chapter 7; and (iv) Dr. P. Barton Duell (Oregon Health and Science University) and Dr. Priya Manjoo (University of British Columbia), for sharing clinical information related to severe hypercholesterolemia patients carrying CNVs in Chapter 5.

*“Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world.”*

— Louis Pasteur

# Dedication

For Mum, Dad, and Andrew.

# Acknowledgments

I would not be where I am today—or who I am today—without the support from my mentors, family, friends, and colleagues. I am grateful for all of the experiences I have had, the things I have learned, and the incredible friendships I have made along the way!

## **To Dr. Hegele:**

It has been an absolute honour and privilege to have had such an incredible PhD supervisor. I have grown and learned so much from your mentorship and guidance—I will forever be thankful! All of my successes and triumphs have been thanks to the wonderful opportunities you have provided me with. If I am ever privileged enough to have the chance to supervise my own students in the future, I hope that I can be for them what you have been for me. Your guidance, mentorship, patience, and encouragement have helped guide me towards the researcher I strive to be, and I would not be where I am today without you.

Thank you for everything!

## **To the Hegele Lab members:**

Matt, Adam, John, Jian, Henian, Ericka, Brooke, David, Jenn—thank you all so much for your support. You have each helped me in achieving my research goals in some way, and have definitely made me laugh on more than one occasion! Every time I have a Tim Horton's smile cookie, celebrate a birthday with ice cream cake, sit at the back of a lecture hall, or be the first to line up for seminar snacks, I'll be reminded of the great times I had in the lab with all of you!

## **To Sali:**

You were my rock as I began navigating the treacherous waters of graduate school! Your dedication and work ethic inspired me and helped shape the type of researcher I wanted to be. You are an inspiration and I hope to be like you when I grow up! Thanks for being there for me, always listening and providing amazing advice, and for being an all-around amazing individual. I am so thankful to have shared an office with you and am honoured to call myself your lab mate, but more importantly, your friend!

**To Alli, Rosettia, Michael, and Julie:**

I literally do not think I would have been able to stay sane the last 5 years without all of you! To Alli, Rosettia and Michael, looking back, I'm not really sure how we managed to get our research done while sharing an office, but we did... and we also managed to remain BFFs the whole time! Some of my best memories were in the office with the 3 of you—who needed windows when I worked with you 3 rays of sunshine every day?!

Julie, I am so glad that you joined our lab as well! Even though a large portion of our time as lab mates has been interrupted by COVID-19, I love that we've been able to talk science and have fantastic video chats. It is wonderful knowing that friends (and colleagues) are only a call away!

To the four of you, thank you for everything you have taught me along the way, for making me laugh, encouraging me, and being such amazing friends! I look forward to the future as we continue to be BFFs and succeed in all of our endeavours!

**To Cal and Pranee:**

Who knew I would get to work with some amazing undergraduate students along the way? Working with you both were some of the highlights of graduate school, and whether you know it or not, you both taught me many things! I am so glad I got the opportunity to help mentor you both, and I am especially glad that I can also consider you both friends, and dare I say, future colleagues? I'll always be here to cheer you both on your own journeys, where ever you may go!

**To the Huff Lab members (and some 4<sup>th</sup> Floor extras):**

Murray, Dawn, Brian, Jane, Amy, Nadya, Corey, Zainab, Matt, and Caroline—an unexpected benefit of being part of the Hegele Lab was getting to be 4<sup>th</sup> Floor neighbours with all of you. I will forever consider you all part of my scientific family, who are happy to talk about lipids, lipoproteins (except for HDL, right Dawn?), and life! The support I felt from all of you was immense and I am so profoundly grateful. My graduate experience would not have been as enjoyable without all of you in it!

As well, thank you Murray for your wonderful support as my graduate advisor. I knew that I could always count on having at least one person come to my poster, no matter what conference I was at! Thank you for your questions, your challenges, for introducing me to the wonderful individuals of the lipids community, and for welcoming me as an honorary member of your group while away at conferences—it is something I will never forget.

**To Tom, Wiggers, and Dylan:**

For all of the time you spent listening to me talk about my research (whether you understood it or not) asking me questions about it, and checking in on me... thank you for being such wonderful friends and for reminding me that a world exists outside the lab! I have learned that the secret to success is not only having support *in* research, but to have support *outside* of research as well!

**To my family:**

Mum, Dad, and Andrew, the three of you have been my biggest cheerleaders not only during graduate school, but for every challenge I have ever faced. I made it to where I am today because of the love and support I have received from all of you! I have only been able to believe in myself because you all taught me how to do it. Thank you for being with me every step of the way! I love you all so much!

**To Zach:**

Words cannot express how grateful I am to have had you support and encourage me during these last few years. I cannot imagine having gone through this experience without you—I have you to thank for becoming both the researcher and person I am today! I can't wait to see what the future has in store for this “scientific power couple”!

## Funding Acknowledgments

The research detailed in this Dissertation was supported by funds from the Canadian Institutes of Health (CIHR), the Heart and Stroke Foundation of Ontario, and Genome Canada through Genome Québec.

Further, I am personally thankful for financial support from the CIHR Fredrick Banting and Charles Best Doctoral Research Award, the CIHR Michael Smith Foreign Supplement, the Ontario Graduate Scholarship, and both the Schulich School of Medicine and Dentistry, and School of Graduate and Postdoctoral Studies at Western University.



# Table of Contents

<b>Abstract</b> .....	ii
<b>Summary for Lay Audience</b> .....	iv
<b>Co-Authorship Statement</b> .....	v
<b>Dedication</b> .....	viii
<b>Acknowledgments</b> .....	ix
<b>Funding Acknowledgments</b> .....	xii
<b>Table of Contents</b> .....	xiii
<b>List of Tables</b> .....	xviii
<b>List of Figures</b> .....	xix
<b>List of Appendices</b> .....	xxi
<b>List of Abbreviations</b> .....	xxii
<b>Chapter 1 – Introduction</b> .....	1
1.1 Overview .....	2
1.2 Human genetic variation .....	5
1.2.1 Single-nucleotide variants.....	5
1.2.2 Structural variants .....	8
1.2.2.1 Insertions and deletions .....	10
1.2.2.2 Copy-number variants .....	10
1.2.2.3 Chromosomal alterations.....	12
1.2.3 Variant frequency.....	12
1.2.3.1 Rare variants.....	14
1.2.3.2 Common variants.....	14
1.3 Genetic basis of traits and disease .....	15
1.3.1 Monogenic inheritance.....	15
1.3.1.1 Penetrance and expressivity .....	19
1.3.2 Polygenic inheritance.....	21
1.3.2.1 Heritability.....	22
1.3.3 Methods to study genetic variation .....	22
1.3.3.1 Sanger sequencing .....	22
1.3.3.2 Next-generation sequencing .....	23
1.3.3.3 Microarrays.....	26
1.3.4 Approaches to study the genetic basis of diseases.....	26
1.3.4.1 Linkage analysis in families or samples of related individuals..	26
1.3.4.2 Associating genotype with phenotype.....	28
1.3.4.3 Polygenic scores .....	33

1.4	Lipids and lipoproteins .....	37
1.4.1	Lipids .....	37
1.4.1.1	Cholesterol.....	38
1.4.1.2	Triglyceride .....	38
1.4.1.3	Plasma lipid sources .....	38
1.4.2	Lipoproteins .....	40
1.4.2.1	High-density lipoprotein.....	43
1.4.2.2	Triglyceride-rich lipoproteins.....	46
1.5	Dyslipidemia.....	51
1.5.1	Genetics of dyslipidemia.....	51
1.5.2	Abnormalities in high-density lipoprotein cholesterol levels .....	54
1.5.2.1	Hypoalphalipoproteinemia .....	56
1.5.2.2	Hyperalphalipoproteinemia .....	58
1.5.3	Abnormalities in triglyceride levels.....	59
1.5.3.1	Hypertriglyceridemia.....	61
1.5.3.2	Hypotriglyceridemia.....	67
1.6	Genetic assessment of dyslipidemia .....	68
1.6.1	LipidSeq: a targeted next-generation sequencing panel for dyslipidemia phenotypes .....	69
1.7	Thesis outline .....	72
1.7.1	Overall research aim and objectives .....	72
1.7.2	Hypothesis.....	73
1.7.3	Summary .....	73
1.8	References .....	74

## **Chapter 2 – Polygenic determinants in extremes of high-density lipoprotein**

	<b>cholesterol .....</b>	<b>94</b>
2.1	Abstract .....	95
2.2	Introduction.....	95
2.3	Materials and Methods.....	97
2.3.1	Study subjects .....	97
2.3.2	DNA preparation and targeted sequencing .....	98
2.3.3	Bioinformatic processing of sequencing data.....	98
2.3.4	Annotation and analysis of rare variants.....	99
2.3.5	Polygenic risk score for high-density lipoprotein cholesterol levels .....	102
2.3.6	Statistical analysis.....	102
2.4	Results.....	103
2.4.1	Characteristics of study subjects.....	103
2.4.2	Rare variants identified in high-density lipoprotein cholesterol-altering genes .....	103
2.4.3	Polygenic risk score development .....	106
2.4.4	Testing and validation of the polygenic risk score .....	108
2.5	Discussion .....	115
2.6	Conclusion .....	118
2.7	References.....	120

<b>Chapter 3 – Large-scale deletions of the <i>ABCA1</i> gene in patients with hypoalphalipoproteinemia</b> .....	125
3.1 Abstract.....	126
3.2 Introduction.....	126
3.3 Materials and Methods.....	128
3.3.1 Study subjects .....	128
3.3.2 DNA preparation and targeted sequencing.....	128
3.3.3 Bioinformatic processing of sequencing data.....	128
3.3.4 Detection of single-nucleotide and copy-number variants .....	128
3.3.5 Validation of partial gene deletions .....	129
3.3.5.1 Breakpoint identification .....	129
3.3.5.2 Sanger confirmation .....	129
3.3.6 Validation of full gene deletions.....	129
3.3.6.1 Exome sequencing.....	129
3.3.6.2 Microarray analysis .....	130
3.3.6.3 Breakpoint identification .....	130
3.3.6.4 Sanger confirmation .....	130
3.4 Results.....	130
3.4.1 Study subjects .....	130
3.4.2 <i>ABCA1</i> copy-number variant detection .....	131
3.4.3 Copy-number variant validation and identifying breakpoints .....	131
3.5 Discussion.....	141
3.6 Conclusion .....	143
3.7 References.....	144
<b>Chapter 4 – Severe hypertriglyceridemia is primarily polygenic</b> .....	148
4.1 Abstract.....	149
4.2 Introduction.....	150
4.3 Materials and Methods.....	152
4.3.1 Study subjects .....	152
4.3.2 DNA preparation and targeted sequencing.....	152
4.3.3 Bioinformatic processing of sequencing data.....	153
4.3.4 Annotation and analysis of rare single-nucleotide variants .....	153
4.3.5 Detection of rare copy-number variants .....	153
4.3.6 Polygenic risk score for elevated triglyceride levels .....	154
4.3.7 Statistical analysis.....	154
4.4 Results.....	154
4.4.1 Characteristics of study subjects.....	154
4.4.2 Rare variants identified in canonical triglyceride metabolism genes .....	157
4.4.3 Measuring accumulation of common triglyceride-raising alleles.....	157
4.4.4 Comparison of genetic profiles between cohorts.....	160
4.4.5 Comparison of triglyceride levels between molecular forms of hypertriglyceridemia .....	160
4.5 Discussion.....	163
4.6 Conclusion .....	167
4.7 References.....	168

<b>Chapter 5 – Partial <i>LPL</i> deletions: rare copy-number variants contributing towards the polygenic form of severe hypertriglyceridemia</b> .....	173
5.1 Abstract.....	174
5.2 Introduction.....	174
5.3 Materials and Methods.....	176
5.3.1 Study subjects .....	176
5.3.2 DNA preparation and targeted sequencing.....	176
5.3.3 Bioinformatic processing of sequencing data.....	176
5.3.4 Detection of copy-number variants.....	176
5.3.5 Validation of partial gene deletions .....	177
5.3.5.1 Breakpoint identification .....	177
5.3.5.2 Sanger confirmation .....	177
5.4 Results.....	177
5.4.1 Study subjects .....	177
5.4.2 <i>LPL</i> copy-number variant detection.....	177
5.4.3 Copy-number variant validation and identifying breakpoints .....	178
5.5 Discussion.....	183
5.6 Conclusion .....	186
5.7 References.....	187
<b>Chapter 6 – The polygenic nature of mild-to-moderate hypertriglyceridemia</b> .....	191
6.1 Abstract.....	192
6.2 Introduction.....	193
6.3 Materials and Methods.....	194
6.3.1 Study subjects .....	194
6.3.2 DNA preparation and targeted sequencing.....	195
6.3.3 Bioinformatic processing of sequencing data.....	195
6.3.4 Annotation and analysis of rare single-nucleotide variants .....	195
6.3.5 Detection of rare copy-number variants .....	195
6.3.6 Polygenic risk score for elevated triglyceride levels .....	195
6.3.7 Statistical analysis.....	196
6.4 Results.....	196
6.4.1 Characteristics of study subjects.....	196
6.4.2 Rare variants identified in canonical triglyceride metabolism genes .....	196
6.4.3 Measuring accumulation of common triglyceride-raising alleles.....	196
6.4.4 Comparison of genetic profiles between cohorts.....	199
6.5 Discussion.....	202
6.6 Conclusion .....	204
6.7 References.....	205
<b>Chapter 7 – Loss-of-function <i>CREB3L3</i> variants in patients with severe hypertriglyceridemia</b> .....	209
7.1 Abstract.....	210
7.2 Introduction.....	210
7.3 Materials and Methods.....	212
7.3.1 Study subjects .....	212
7.3.2 DNA preparation and targeted sequencing.....	213
7.3.3 Bioinformatic processing of sequencing data.....	213

7.3.4	Principal component analysis .....	213
7.3.5	Annotation and analysis of loss-of-function variants .....	214
7.3.6	Gene-based rare variant association analysis.....	214
7.3.7	Odds ratio assessment .....	215
7.4	Results.....	215
7.4.1	Characteristics of study subjects.....	215
7.4.2	Loss-of-function variants identified from the LipidSeq gene panel .....	220
7.4.3	Gene-based rare variant association study using SKAT-O.....	220
7.4.4	Gene-based odds ratio assessment .....	220
7.5	Discussion.....	224
7.6	Conclusion .....	226
7.7	References.....	228
<b>Chapter 8 – Discussion .....</b>		<b>232</b>
8.1	Overview.....	233
8.2	Summary of research findings .....	233
8.2.1	The genetic architecture of extreme high-density lipoprotein cholesterol levels .....	233
8.2.1.1	Hypoalphalipoproteinemia .....	235
8.2.1.2	Hyperalphalipoproteinemia .....	236
8.2.1.3	Genetic influences across high-density lipoprotein cholesterol levels.....	237
8.2.2	The genetic architecture of hypertriglyceridemia .....	238
8.2.2.1	Severe hypertriglyceridemia.....	238
8.2.2.2	Mild-to-moderate hypertriglyceridemia .....	241
8.2.2.3	Genetic influences across hypertriglyceridemia phenotypes ...	242
8.3	Study strengths, limitations and caveats .....	245
8.3.1	Strengths .....	245
8.3.2	Caveats.....	246
8.3.3	Limitations .....	248
8.4	Applications and future directions .....	251
8.4.1	Estimating effects of genetic determinants .....	251
8.4.2	Screening for genetic risk .....	251
8.4.3	Updating lipid-based polygenic risk scores .....	252
8.4.3.1	Genome-wide scores .....	253
8.4.4	Finding additional susceptibility genes.....	254
8.5	Risk for cardiovascular disease, and levels of high-density lipoprotein cholesterol and triglyceride .....	255
8.6	Conclusions.....	257
8.7	References.....	260
<b>Appendices.....</b>		<b>270</b>
<b>Curriculum Vitae .....</b>		<b>319</b>

## List of Tables

<b>Table 1.1</b> Dyslipidemia phenotypes and their genetic etiologies.....	53
<b>Table 2.1</b> Genes with candidate (primary) and non-candidate (secondary) influences on HDL cholesterol levels.....	101
<b>Table 2.2</b> Clinical and demographic information of patients with low HDL cholesterol levels (N=686).....	104
<b>Table 2.3</b> Clinical and demographic information of patients with high HDL cholesterol levels (N=1,165). ....	104
<b>Table 2.4</b> The 9 SNPs used in the polygenic risk score for HDL cholesterol levels. ....	107
<b>Table 3.1</b> Clinical and demographic features of subjects with <i>ABCA1</i> CNVs. ....	132
<b>Table 3.2</b> Genomic coordinates and breakpoints of <i>ABCA1</i> CNVs.....	140
<b>Table 4.1</b> The 16 SNPs used in polygenic risk score for elevated triglyceride levels. ....	155
<b>Table 4.2</b> Clinical and demographic information of severe hypertriglyceridemia cohorts (N=563).....	156
<b>Table 5.1</b> Clinical and demographic features of subjects with <i>LPL</i> CNVs.....	180
<b>Table 5.2</b> Genomic coordinates and breakpoints of <i>LPL</i> CNVs.....	181
<b>Table 6.1</b> Clinical and demographic information of the mild-to-moderate hypertriglyceridemia patient cohort (N=134). ....	197
<b>Table 7.1</b> Clinical and demographic information of patients with severe hypertriglyceridemia (N=265).....	219
<b>Table 7.2</b> Output from SKAT-O analysis between severe hypertriglyceridemia (N=265) and normolipidemic controls (N=477). ....	221
<b>Table 7.3</b> The odds of severe hypertriglyceridemia patients (N=265) carrying a loss-of-function variant in a particular gene compared to normolipidemic controls (N=477). ....	222
<b>Table 8.1</b> Distinguishing between familial chylomicronemia syndrome, multifactorial chylomicronemia, and mild-to-moderate hypertriglyceridemia. ....	244

# List of Figures

<b>Figure 1.1</b> Different types of SNVs defined by sequence ontology. ....	7
<b>Figure 1.2</b> Different types of structural variants. ....	9
<b>Figure 1.3</b> Small-scale structural variants. ....	11
<b>Figure 1.4</b> Spectrum of genetic variation related to phenotypic effect and population frequency. ....	13
<b>Figure 1.5</b> Schematic representation of Mendelian inheritance patterns. ....	18
<b>Figure 1.6</b> Variant penetrance and expressivity. ....	20
<b>Figure 1.7</b> Regression models to determine genotype-phenotype associations. ....	30
<b>Figure 1.8</b> Lipid and apolipoprotein composition of lipoprotein classes. ....	41
<b>Figure 1.9</b> Lipoprotein classes and their physical characteristics. ....	42
<b>Figure 1.10</b> Metabolic lifecycle of HDL particles. ....	45
<b>Figure 1.11</b> Metabolic lifecycle of chylomicrons and their remnant particles. ....	47
<b>Figure 1.12</b> Metabolic lifecycle of VLDL, IDL, and their remnant particles. ....	48
<b>Figure 1.13</b> The genetic architecture underlying the spectrum of measurable HDL cholesterol levels. ....	55
<b>Figure 1.14</b> The genetic architecture underlying the spectrum of measurable triglyceride levels. ....	61
<b>Figure 1.15</b> Genes targeted by the LipidSeq panel. ....	71
<b>Figure 2.1</b> Summary of rare variants identified within patients from the Lipid Genetics Clinic cohort (N=255). ....	105
<b>Figure 2.2</b> Polygenic risk score distribution of non-carrier patients between different cohorts. ....	110
<b>Figure 2.3</b> Differences in extreme polygenic risk scores between carrier and non-carrier patients with low HDL cholesterol levels. ....	111
<b>Figure 2.4</b> Differences in extreme polygenic risk scores between carrier and non-carrier patients with high HDL cholesterol levels. ....	112
<b>Figure 2.5</b> The comparison of genetic profiles of extreme HDL cholesterol phenotypes between cohorts. ....	113

<b>Figure 2.6</b> Association between polygenic risk score deciles and HDL cholesterol levels.	114
<b>Figure 3.1</b> Identification of <i>ABCA1</i> CNVs using the VarSeq-CNV® caller algorithm on targeted sequencing data.	136
<b>Figure 3.2</b> Confirmation of the full-gene <i>ABCA1</i> CNV using the VarSeq-CNV® caller algorithm on exome data.	137
<b>Figure 3.3</b> Validation of full-gene deletion of <i>ABCA1</i> in Patient 4 with hypoalphalipoproteinemia.	138
<b>Figure 3.4</b> Validation of partial gene deletions of <i>ABCA1</i> in Patients 1, 2, and 3 with hypoalphalipoproteinemia.	139
<b>Figure 4.1</b> Identification of a <i>LPL</i> CNV using the VarSeq-CNV® caller algorithm.	158
<b>Figure 4.2</b> Polygenic risk score analysis for severe hypertriglyceridemia patients.	159
<b>Figure 4.3</b> The comparison of genetic determinants of severe hypertriglyceridemia between cohorts.	161
<b>Figure 4.4</b> Differences in genetic determinants of severe hypertriglyceridemia between cohorts.	162
<b>Figure 5.1</b> Identification of <i>LPL</i> CNVs using the VarSeq-CNV® caller algorithm on targeted sequencing data.	179
<b>Figure 5.2</b> Validation of deletions disrupting <i>LPL</i> in patients with severe hypertriglyceridemia.	182
<b>Figure 6.1</b> Polygenic risk score distributions between cohorts.	198
<b>Figure 6.2</b> The comparison of genetic profiles between cohorts.	200
<b>Figure 6.3</b> Differences in genetic determinants of hypertriglyceridemia.	201
<b>Figure 7.1</b> Principal component analysis.	218
<b>Figure 7.2</b> Odds ratio of loss-of-function variants across LipidSeq genes.	223
<b>Figure 8.1</b> The updated genetic architecture underlying the spectrum of measurable HDL cholesterol levels.	234
<b>Figure 8.2</b> The updated genetic architecture underlying the spectrum of measurable triglyceride levels.	239
<b>Figure 8.3</b> The comparison of genetic profiles of different dyslipidemia phenotypes.	259



# List of Appendices

<b>Appendix A.</b> Copyright Permissions .....	270
<b>Appendix B.</b> University of Western Ontario - Ethics Approval .....	279
<b>Appendix C.</b> Summary of unique rare SNVs identified in primary HDL cholesterol genes. .....	280
<b>Appendix D.</b> Summary of unique rare SNVs identified in secondary HDL cholesterol genes. .....	281
<b>Appendix E.</b> Screening primers for <i>ABCA1</i> copy-number variations. ....	282
<b>Appendix F.</b> Summary of unique rare SNVs and CNVs identified in the Lipid Genetics Clinic cohort (N=251).....	283
<b>Appendix G.</b> Summary of unique rare SNVs and CNVs identified in the UCSF cohort (N=312).....	284
<b>Appendix H.</b> Summary of unique rare SNVs identified in the reference 1000 Genomes Project cohort (N=503). ....	285
<b>Appendix I.</b> Screening primers for <i>LPL</i> CNVs. ....	286
<b>Appendix J.</b> Summary of unique rare SNVs identified in mild-to-moderate hypertriglyceridemia patients (N=134).....	287
<b>Appendix K.</b> Summary of unique rare LOF variants identified in multifactorial chylomicronemia patients (N=265). ....	288
<b>Appendix L.</b> Exautomate: A user-friendly tool for region-based rare variant association analysis.....	289

## List of Abbreviations

A	Adenine
ABCA1	ATP-binding cassette transporter A1
ABCG5	ATP-binding cassette transporter G5
ABCG8	ATP-binding cassette transporter G8
Acetyl-CoA	Acetyl coenzyme A
ACMG	American College of Medical Genetics and Genomics
ANGPTL3	Angiopoietin-like protein 3
Apo	Apolipoprotein
APOBEC-1	Apo B mRNA editing enzyme catalytic subunit 1
ASCVD	Atherosclerotic cardiovascular disease
ASSEDA	Automated Splice Site and Exon Definition Analyses
C	Cytosine
CADD	Combined Annotation Dependent Depletion
CETP	Cholesteryl ester transfer protein
CI	Confidence interval
CNV	Copy-number variant
CREBH	Cyclic AMP-responsive element-binding protein H
CVD	Cardiovascular disease
EL	Endothelial lipase
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
F	Forward
FCS	Familial chylomicronemia syndrome
FED	Fish-eye disease
FH	Familial hypercholesterolemia
FISH	Fluorescent <i>in situ</i> hybridization
FLD	Familial LCAT deficiency
G	Guanine
GLGC	Global Lipids Genetics Consortium
gnomAD	Genome Aggregation Database

GPIHBP1	Glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HL	Hepatic lipase
HMG-CoA	$\beta$ -hydroxy $\beta$ -methylglutaryl-coenzyme A
IDL	Intermediate-density lipoprotein
Indel	Insertion or deletion
LCAT	Lecithin-cholesterol acyltransferase
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LDLR	Low-density lipoprotein receptor
LMF1	Lipase maturation factor 1
LOD	Logarithm of the odds
LPL	Lipoprotein lipase
LRP1	Low-density lipoprotein-related protein 1
MAF	Minor allele frequency
MCM	Multifactorial chylomicronemia
MHI	Montréal Heart Institute
MLPA	Multiplex ligation-dependent probe amplification
MODY	Mature-onset diabetes of the young
MTP	Microsomal triglyceride transfer protein
NGS	Next-generation sequencing
NPC1L1	Niemann-Pick C1 like 1 protein
OR	Odds ratio
PC	Principal component
PCA	Principal component analysis
PCSK9	Proprotein convertase subtilisin/kexin type 9
PLTP	Phospholipid transfer protein
PolyPhen2	Polymorphism Phenotyping version 2
R	Reverse
RVAS	Rare variant association study

SD	Standard deviation
SIFT	Sorting Intolerant from Tolerant
SKAT-O	Optimal unified sequence kernel association test
SNP	Single-nucleotide polymorphism
SNV	Single-nucleotide variant
SPANR	Splicing Based Analysis of Variants
SR-BI	Scavenger receptor class B type I
SREBP	Sterol regulatory element binding protein
SVS	SNP & Variation Suite
T	Thymine
UCSF	University of California, San Francisco
UPenn	University of Pennsylvania
UTR	Untranslated region
VLDL	Very-low-density lipoprotein

## Chapter 1 – Introduction

The text contained in this Chapter has been adapted from previously published sources for brevity and to ensure consistency throughout this Dissertation.

**Dron, J.S.**, and Hegele, R.A. (2016). Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130-141.

**Dron, J.S.**, and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.

**Dron, J.S.**, and Hegele, R.A. (2017). Molecular Genetics of Hypertriglyceridemia. In eLS, L.E. John Wiley & Sons, ed., pp. 1-8.

**Dron, J.S.**, and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.

**Dron, J.S.**, and Hegele, R.A. (2019). The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol* 30, 71-81.

**Dron, J.S.**, Wang, J., McIntyre, A.D., Iacocca, M.A., Robinson, J.F., Ban, M.R., Cao, H., and Hegele, R.A. (2020). Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Med Genomics* 13, 23.

## 1.1 Overview

Cardiovascular disease (CVD) is the primary cause of death globally and is estimated to cause one-third of deaths in Canada (Statistics Canada, 2015; World Health Organization, 2020). With an expected increase in prevalence due to the obesity epidemic, CVD will continue to strain our health-care system and economy—over \$20 billion is lost annually through hospital costs and decreased productivity (Genest et al., 2009). As such, there is a pressing need to characterize CVD risk factors for applications towards clinical risk prediction, prognosis indicators, and effective medical interventions to reduce risk.

Interdisciplinary collaborations have implicated a number of risk factors for CVD, including hypertension, obesity, diabetes, smoking, and a sedentary lifestyle (Lloyd-Jones, 2010; Lloyd-Jones et al., 2010). Importantly, one of the most heritable, modifiable identified risk factors are levels of plasma lipids, namely cholesterol and triglyceride, and their lipoprotein carriers (Castelli et al., 1986; Kannel et al., 1964; Wilson et al., 1980). Epidemiologic and clinical studies have identified three main quantitative traits associated with CVD risk, including circulating levels of: (i) low-density lipoprotein (LDL) cholesterol; (ii) high-density lipoprotein (HDL) cholesterol; and (iii) triglyceride.

Researchers and physicians have so far collected extensive and compelling evidence—genetic, mechanistic, and clinical trial data—supporting a direct causal relationship between LDL cholesterol levels and CVD (Ference et al., 2017). Several classes of drugs, including statins (Taylor et al., 2013), ezetimibe (Ballantyne et al., 2007) and proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors (Sabatine, 2019), both lower plasma LDL cholesterol levels and reduce risk for CVD events and death; this knowledge has been translated into clinical practice guidelines (Anderson et al., 2016; Grundy et al., 2019).

In contrast, the path to clinically translating our understanding of HDL cholesterol and triglyceride levels has been less straightforward. Decades' worth of observations in populations showing an inverse relationship between HDL cholesterol levels and CVD prompted clinical trials to attempt to pharmacologically increase HDL cholesterol levels with the expectation that this would lower CVD risk (Gordon et al., 1977; Sharrett et al.,

2001). Unfortunately, trials of HDL cholesterol-raising agents like niacin, fibrates and cholesteryl ester transfer protein (CETP) inhibitors all failed to reduce CVD, and in some cases even paradoxically increased all-cause mortality (Chait and Eckel, 2016; Schwartz et al., 2012). These counterintuitive findings were mirrored by more recent Mendelian randomization studies, which used genetic markers to impute lifelong levels of HDL cholesterol and found no evidence of a causal link between HDL cholesterol levels and CVD (Frikke-Schmidt et al., 2008; Haase et al., 2012; Johannsen et al., 2009; Voight et al., 2012). A more rigorous characterization of the genetics underlying HDL cholesterol levels and subsequent analyses (like Mendelian randomization studies) may have alerted researchers to some of the challenges that were faced in clinical trials.

The disappointing clinical results due to failure to show benefit with intervention on HDL cholesterol levels over the past 10 years have led to a shift in focus to triglyceride as an alternative, and perhaps more reasonable target for CVD risk reduction. While pharmacologic studies of triglyceride-lowering agents have generally been more positive than studies of HDL cholesterol reduction, there are still inconsistencies. In epidemiological studies, adjustments for confounding variables seemed to neutralize associations between triglyceride levels and CVD in epidemiological studies (Dron and Hegele, 2017a; Emerging Risk Factors Consortium et al., 2009), but Mendelian randomization studies have shown a causal relationship between triglyceride levels and CVD events, purported as being independent of other markers or variables (Allara et al., 2019; Holmes et al., 2015; Jorgensen et al., 2014; Tg et al., 2014).

Without a complete understanding of the mechanisms underlying variation in HDL cholesterol and triglyceride levels, challenges in developing strategies to target these traits for CVD risk reduction will remain. Therefore, a comprehensive assessment of the main factors driving these traits is necessary to overcome these uncertainties. With an estimated heritability for plasma lipid and lipoprotein levels ranging from 40-60% (Tada et al., 2014), it stands to reason that a thorough assessment of the genetic architecture underlying these traits will contribute towards our foundational understanding of both HDL cholesterol and triglyceride levels.

Early breakthroughs in understanding the genetic factors influencing these traits were seen with studies of relatively rare individuals who had extreme levels of the respective lipid of interest. Historically, the factors driving these phenotypes—initially characterized as rare genetic variations with large phenotypic effects—were identified through association studies and linkage analyses of affected kindreds (Breslow, 2000; Hegele, 2009). These early studies unveiled many key genes and proteins involved in the respective metabolic pathways; however, relevant causal rare variants are not observed in all phenotypically affected individuals (Candini et al., 2010; Cohen et al., 2004; Hegele, 2009; Holleboom et al., 2011; Kiss et al., 2007; Sadananda et al., 2015; Singaraja et al., 2013; Talmud et al., 2013; Tietjen et al., 2012; Wang et al., 2016). As a result, the research focus has expanded to consider genetic variations with smaller phenotypic effects, namely single-nucleotide polymorphisms (SNPs), which are more frequent in the population, but because of their small effect size, require epidemiological-scale, genome-wide association studies (GWASs) in large populations to be detected (Frazer et al., 2009; Hegele, 2009).

While individual research efforts have evaluated specific types of genetic determinants for either HDL cholesterol or triglyceride levels, there has been minimal effort or experience to date in assessing multiple types of determinants simultaneously. This is due to a reductionist focus on only one type of variation in most experimental designs, typically related to technological limitations. Heretofore separate methods have been required to study different types of genetic variation. Furthermore, practical challenges arise though the need to aggregate specialized cohorts enriched with the extreme phenotype of interest in order to achieve sufficient statistical power.

Substantial efforts are necessary to study and understand the genetic underpinnings of HDL cholesterol and triglyceride levels before these traits can be rationally targeted therapeutically for CVD risk prevention—probing the complete, holistic genetic foundation of each trait is likely required. Such holistic evaluation requires multiple components: (i) technology enabling assessment of multiple types of genetic determinants simultaneously; (ii) large cohorts of patients with extreme levels of either HDL cholesterol or triglyceride with sufficient statistical confidence; and (iii) a robust



bioinformatic process to allow assessment of phenotype-genotype relationships. Promising associations can later be followed up using laboratory experiments involving *in vitro* or *in vivo* model systems; however, genetics often generates the earliest clues and leads for mechanisms that can be evaluated by functional and mechanistic experiments.

## 1.2 Human genetic variation

The human genome is the complete set of nucleic acid sequences encoded as DNA and resides primarily within 23 chromosome pairs in the nucleus, with a small amount in the mitochondria. The total length of the human genome is more than 3 billion nucleotide base pairs, of which there are four that comprise DNA: adenine (A), cytosine (C), guanine (G), and thymine (T). Together, the human genome sequence contains all the biological information necessary to support us throughout our life cycle. The human genome map has allowed for accurate, quantitative positioning of every base pair. While humans share up to 99.9% of their genomic sequence (Feuk et al., 2006), genetic variation exists across all individuals, which manifests as nucleotide sequence differences at particular positions along the genome map. These differences between individuals are usually silent at the phenotypic level, but occasionally they may give rise to unique and distinctive phenotypic characteristics and differences, including but not limited to differences in physical appearances, metabolic and biochemical activities, and disease risk (Frazer et al., 2009; Genomes Project et al., 2010).

Different types of interindividual variations exist within the genome and these can be defined by various characteristics, including physical-chemical properties, frequency of a variant within the population, and associations of variants with differences in phenotypic outcomes.

### 1.2.1 Single-nucleotide variants

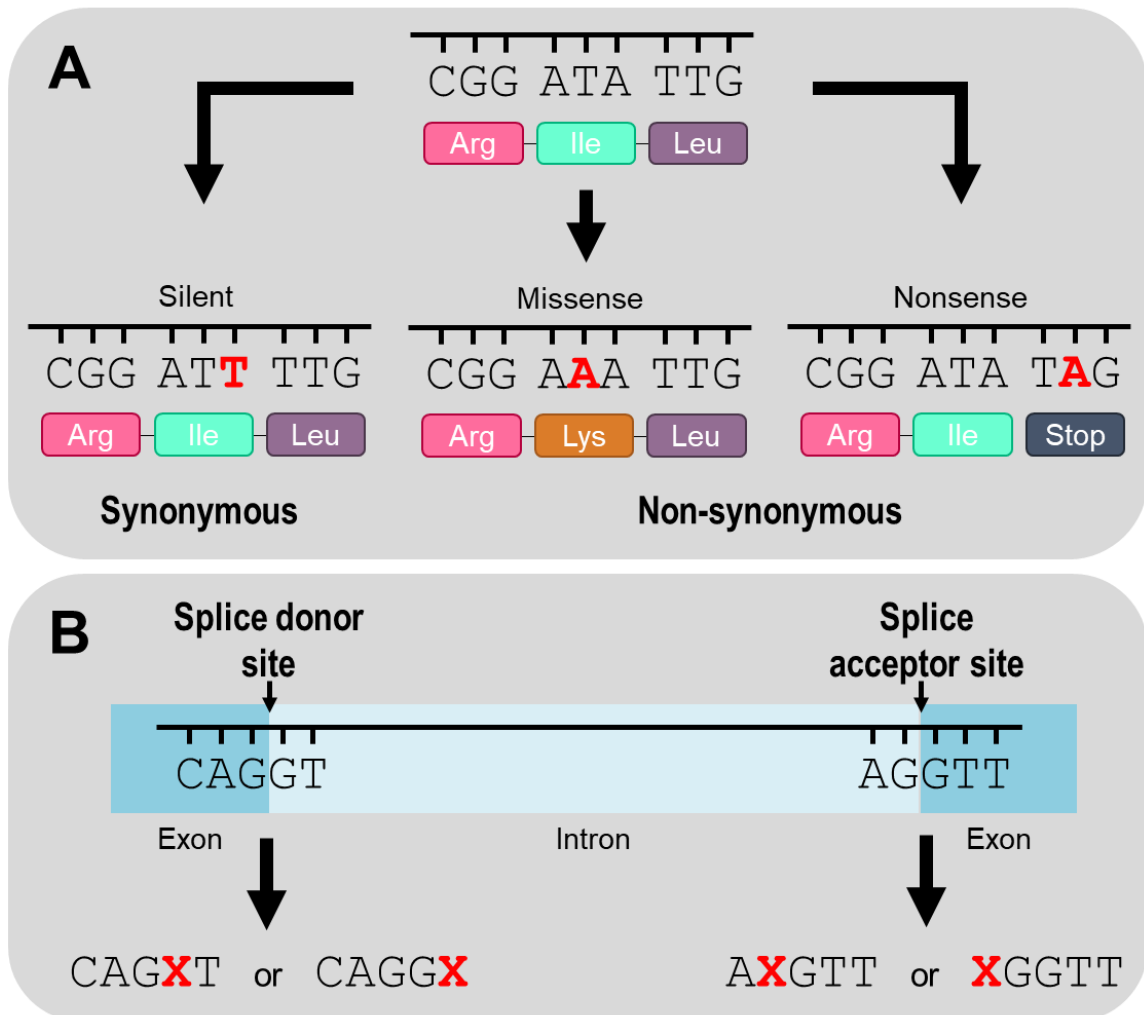
Single-nucleotide variants (SNVs) are defined as changes that involve single nucleotide positions and represent the most common form of human genetic variation (Frazer et al., 2009; Timpson et al., 2018). Substitutions that occur between the purine nucleotides (A and G), or the pyrimidine nucleotides (C and T), are called “transitions”; while substitutions from a purine to a pyrimidine or vice versa are called “transversions”. These

simple substitutions can be further characterized by their impact on sequence ontology (**Figure 1.1A**), especially when the SNV occurs within a coding gene sequence.

For instance, SNVs with no impact on the gene's encoded protein product are referred to as “synonymous” variants; because of redundancy in amino acid codons, it is possible that a change in a single nucleotide can still result in the same translated amino acid sequence. Conversely, SNVs that alter the gene's protein product are referred to as “non-synonymous” variants, and can be subclassified further as: (i) “missense” variants that lead to a codon change and result in a different translated amino acid sequence; or (ii) “nonsense” variants that lead to the inappropriate introduction of a stop codon, often creating an early truncation of the encoded protein product. Because of such consequences, nonsense variants are one type of “protein-truncating” variant. Another potential type of protein-truncating variant involve changes that affect the RNA splicing machinery, such as “splice-donor” or “splice-acceptor” variants that fall within sequences at mRNA splice-junctions at the beginning or end of an intron, respectively (**Figure 1.1B**) (Cartegni et al., 2002).

SNVs that occur outside of protein-coding regions are by definition non-coding and cannot appropriately be labeled using terms such as “synonymous” or “nonsynonymous”. Instead, they are defined according to the type of regions in which they are found, such as within introns, 5' or 3' untranslated regions (UTRs), promoters, enhancers, silencers, non-coding genes, or pseudogenes.

“SNPs” are a specific subtype of the more general “SNVs”, the latter of which is an agnostic term with respect to the variant's population frequency—it has no connotation as to whether the variant is common or rare in the population. Describing a variant as a “SNP” is conventional when considering variants that occur more frequently in the population: a “SNP” implies a relatively prevalent SNV. SNPs are the workhorses of genetic association studies, as discussed further in [Section 1.3.4.2](#)



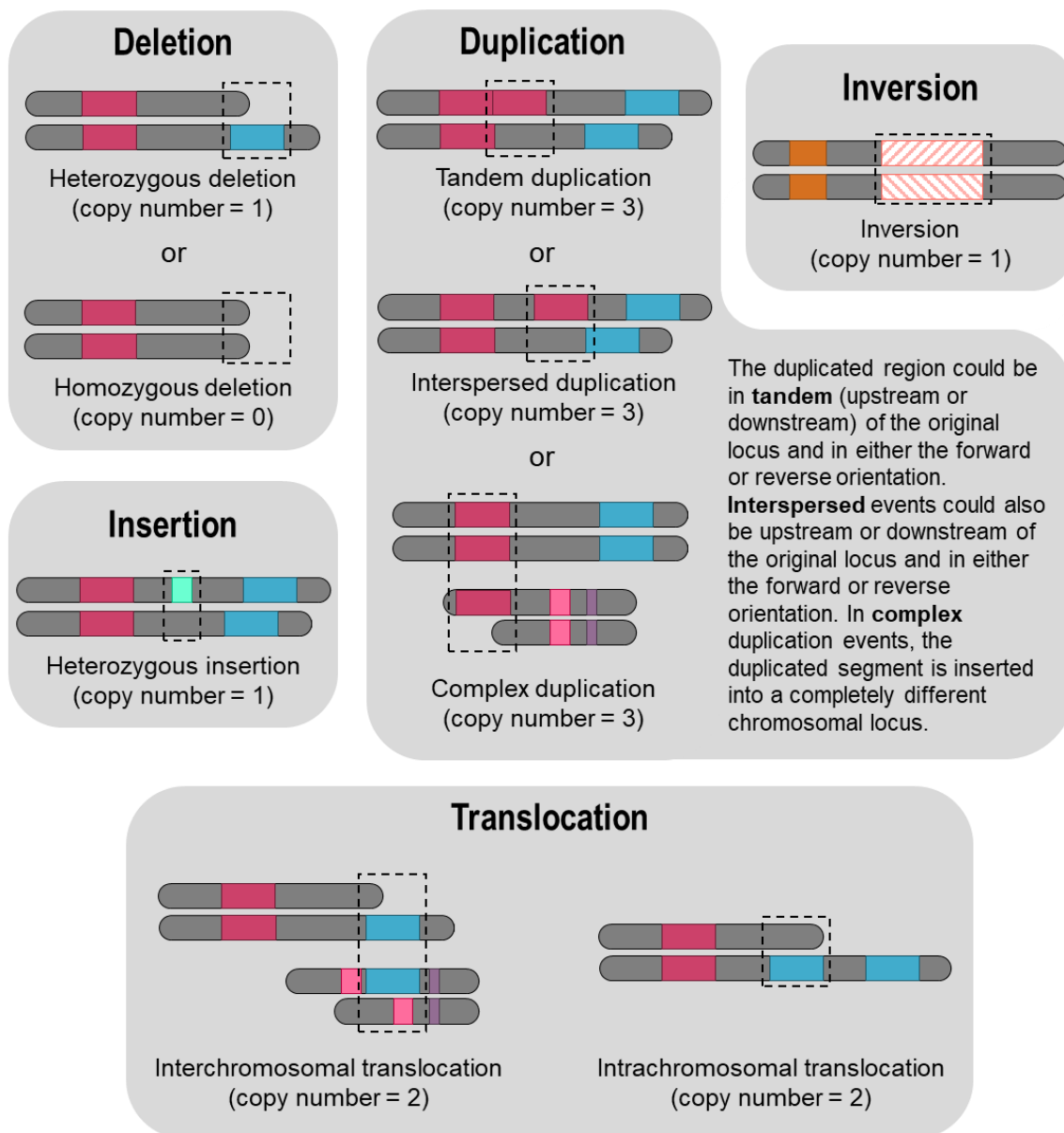
**Figure 1.1 Different types of SNVs defined by sequence ontology.**

SNVs can be characterized by their impact on sequence ontology. **A)** Single nucleotide changes that occur within protein-coding regions of genes (i.e. exons) can be “synonymous” or “silent” if the encoded amino acid does not change, or “non-synonymous” if the encoded amino acid changes. Non-synonymous variants are normally classified as “missense”, unless the SNV changes the amino acid to a stop codon, in which case it is defined as “nonsense”. **B)** An SNV disrupting an mRNA splice junction is defined as either a “splice-donor” or “splice-acceptor” variant, depending on whether the SNV occurs in the splice donor or splice acceptor site, respectively. Bolded red nucleotides reflect the SNV in each example; “X” could represent any nucleotide substitution.

### 1.2.2 Structural variants

“Structural variants” refer to a type of genetic variant larger than SNVs, ranging in size from only a few impacted nucleotides, up to full chromosomal segments (Frazer et al., 2009). Estimates have suggested that structural variants may account for 5-14% of the human genome (Conrad et al., 2010; Sudmant et al., 2015; Zarrei et al., 2015). Given the physical impact these variants can have on the genome—with the potential to encompass genes in whole or in part—many have been associated and causally linked with certain diseases. On the other hand, certain structural variants have been reported to have either no apparent phenotypic consequence or even phenotypically beneficial ones (Weischenfeldt et al., 2013).

Structural variants can be further classified into subgroups, including: insertions, deletions, duplications, inversions, and translocations (**Figure 1.2**) (Weischenfeldt et al., 2013). The molecular mechanisms leading to these events are typically due to errors in DNA recombination (unequal crossing over), replication, and/or repair (Hastings et al., 2009; Weischenfeldt et al., 2013).



**Figure 1.2 Different types of structural variants.**

Structural variant sizes can range from 50 nucleotides up to full chromosomal segments. The loss of genetic material is defined as a “deletion”, while the gain of novel genetic material is defined as an “insertion”. “Duplications” reflect the insertion of genetic material that has been duplicated from an existing genomic locus. An “inversion” indicates a genomic locus that has flipped its orientation (ex. from the forward to the reverse orientation). A “translocation” is used to describe an event in which a genomic segment has been moved to a different chromosome (“inter-”) or to the opposing allele of the same chromosome (“intra-”). Each horizontal bar reflects a chromosomal region, with each coloured block reflecting a genomic locus of interest, such as a gene. The dashed box indicates the area of interest for each structural variant. The normal state of the human genome is diploid, with a copy number of 2. This figure and legend have been adapted from (Iacocca et al., 2019).

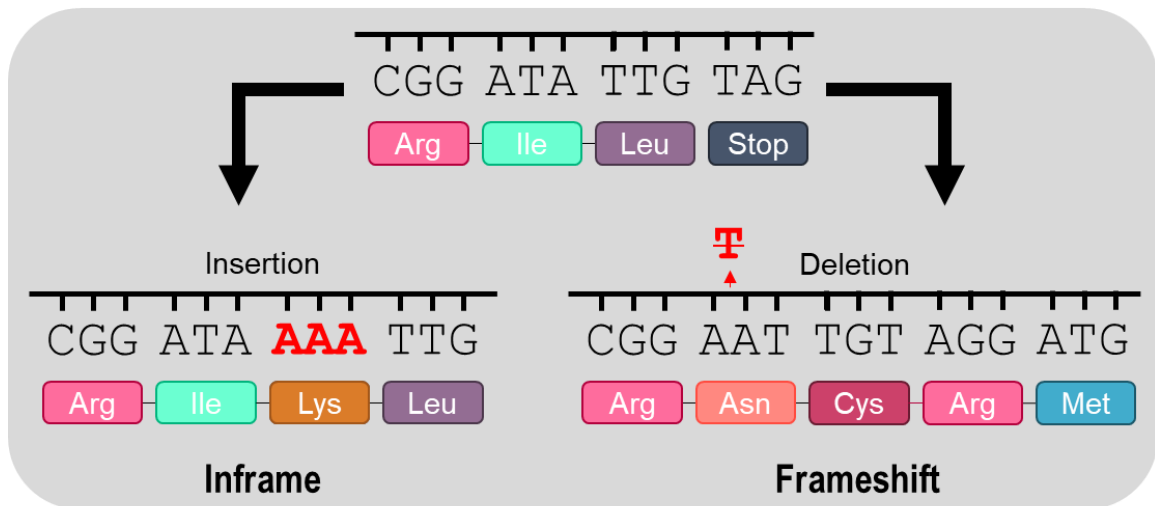
### 1.2.2.1 Insertions and deletions

As the terms suggest, insertions and deletions refer to the gain and loss of nucleotides, respectively. These types of variations are jointly referred to as “indels” and can occur anywhere throughout the genome, ranging in size from 1-50 nucleotides (Sudmant et al., 2015).

When an indel variant occurs within a protein-coding gene, it can sometimes have a substantial impact on the final protein product (**Figure 1.3**). For instance, indels that disrupt the codon reading frame are referred to as “frameshift” variants and can alter the protein’s amino acid sequence, effectively altering the originally encoded protein; as with nonsense variants, some frameshift variants can result in early termination of translation and are thus considered as protein-truncating. Indels can also shift the reading frame such that the normal stop codon is lost, and translation continues, producing a qualitatively abnormal, elongated protein product. Conversely, an “inframe” variant is due to an insertion or deletion of entire codons (i.e. in multiples of 3 nucleotides), but while slightly changing the length of the translated variant protein, they do not disrupt the overall reading frame, keeping the stop codon intact and are thus not protein-truncating.

### 1.2.2.2 Copy-number variants

As diploid organisms, humans normally have two copies of their nuclear genome—both a maternal and paternal copy. Changes to this diploid state at a particular locus or region through either duplication or deletion events are defined as changes in copy number; duplications lead to gains in copy number, while deletions lead to losses in copy number. By convention, the results of these events when spanning >50 nucleotides in length are referred to as copy-number variants (CNVs) (Redon et al., 2006; Sudmant et al., 2015). CNVs are the most common type of structural variant within the human genome (Conrad et al., 2010; Zarrei et al., 2015).



**Figure 1.3 Small-scale structural variants.**

The gain or loss of nucleotides are defined as “insertions” and “deletions”, and can lead to changes in an encoded protein product when these events occur within genes. Insertions or deletions of  $3n$  nucleotides between adjacent codons do not disrupt the reading frame and are called “inframe”; the original amino acid sequence is largely retained. Meanwhile, insertions or deletions that disrupt the original amino acid sequence due to a change in the reading frame are defined as “frameshift”. Bolded red nucleotides reflect the newly inserted or deleted nucleotide(s) in each example.

Outcomes related to a copy-number change could be beneficial, detrimental, or neutral to the organism, depending on the impacted regions of the genome (Zarrei et al., 2015); this spectrum of phenotypic impact can range from adaptive features to embryonic lethality (Conrad et al., 2010; Hastings et al., 2009; Zarrei et al., 2015).

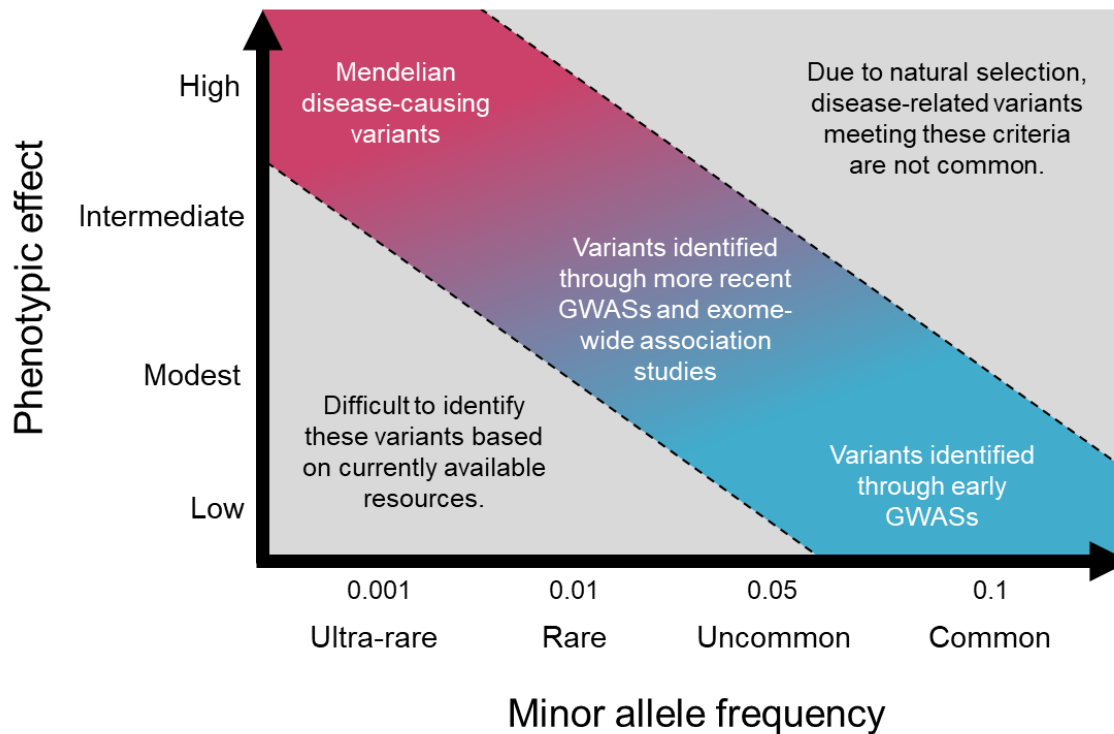
### 1.2.2.3 Chromosomal alterations

Genetic variation defined as “chromosomal alterations” are often large enough to be observed using cytogenetic techniques, such as fluorescent *in situ* hybridization (FISH); some of the earliest alterations could be simply observed cytogenetically with a light microscope (Feuk et al., 2006). Translocations and inversions are examples of intra- and interchromosomal rearrangements, respectively (Feuk et al., 2006). Even larger alterations include abnormal chromosomal counts, defined as aneuploidy, which could be considered as chromosomal-scale CNVs.

### 1.2.3 Variant frequency

Through international collaborative efforts, publicly available databases of genetic information have provided detailed information for the frequencies at which genomic variants—both SNVs and structural variants—occur within the population (Genomes Project et al., 2015; Karczewski et al., 2020; Lek et al., 2016). This variant attribute is defined as “minor allele frequency” (MAF). The terminology is a remnant from the nomenclature of classical genetics, in which “major” and “minor” allele refer to the more and less common allele at a particular variant locus, respectively. This does not necessarily correspond to the “reference” and “alternate” allele distinctions, which are specifically relevant to the human reference genome. The population through which a MAF is determined can be defined as the general global population, a particular ancestral group, or specialized cohorts (e.g. those with a particular disease). This information can reveal insights into a variant in the context of its phenotypic consequence (**Figure 1.4**), ancestral significance, and its relationship with natural selection. It is notable that the terminology and designations are relative: there are many examples of variant or polymorphic loci at which the minor (less common) allele in one particular geographical or ancestral group is the major (more common) allele in a different group.





**Figure 1.4 Spectrum of genetic variation related to phenotypic effect and population frequency.**

A genetic variant's population frequency is a function of the variant's phenotypic effect and how it impacts an organism's fitness, discussed in [Section 1.2.3](#). Rare and ultra-rare variants with large phenotypic effects are often the cause of Mendelian disorders (discussed in [Section 1.3.1](#)), while variants with smaller effects on a phenotype are more common and can be identified through population-scale association studies, like GWAS. Meanwhile, rare and ultra-rare variants with smaller phenotypic effects will only be uncovered as association study cohorts increase in size, and techniques to study rare variants improve. Figure adapted from (Assimes and Roberts, 2016). Abbreviations: GWAS = genome-wide association study.

### 1.2.3.1 Rare variants

Rare variants are almost universally defined as having a MAF of  $\leq 1\%$  (Katsanis, 2016; MacArthur et al., 2014), although, the term “ultra-rare” can be used to classify variants with stricter frequency thresholds (Katsanis, 2016).

*De novo* variants are considered to be the rarest type of genetic determinant, as they occur spontaneously in an individual and in theory would have a virtually non-existent population frequency (Ku et al., 2012). If a *de novo* variant occurs in the germline, when the variant is passed along to the individual’s offspring, the variant is then classified as “inherited” and would still have a virtually absent population frequency.

In addition to the spontaneous occurrence of *de novo* variants, variants that have been acted upon by natural selection can become rare over generations. For instance, variants that decrease an organism’s biological fitness are considered “deleterious” and undergo negative (purifying) selection, thus becoming less frequent in the population since the variant is not able to be passed along to subsequent generations (Lohmueller, 2014; Quintana-Murci, 2016).

As mentioned above, sometimes allele frequencies in different ancestral groups can differ due to founder effects and population bottlenecks (Quintana-Murci, 2016). This is an important consideration in designing research studies and deriving conclusions, especially when statistically testing for differences in variant frequencies between two distinct population samples (e.g. cases and controls) and then drawing inferences about the potential biological relevance if a statistical difference is detected. If the experiment is not properly controlled, statistical differences in allele frequencies could reflect artifactual differences in the samples related to ancestry rather than a biological impact of the variant locus.

### 1.2.3.2 Common variants

In contrast to rare variants, common variants have a MAF of  $>1\%$  (MacArthur et al., 2014). Given their extensive range of frequency, common variants are further classified as “uncommon”, with frequencies between 1-5%.

A genetic variant with only a modest phenotypic impact that does not influence an organism's fitness is unlikely to be acted upon by natural selection, and can therefore be inherited by subsequent generations largely undisturbed if the host organism survives to reproductive age. As such, changes in frequency of these variants with mild-to-neutral phenotypic effects are driven by genetic drift, and perhaps further punctuated through founder or bottleneck populations (Quintana-Murci, 2016).

When a genetic variant improves an organism's fitness, it can undergo positive selection within the population and become even more frequent since there is an increased chance of the variant being inherited by subsequent generations (Quintana-Murci, 2016).

### 1.3 Genetic basis of traits and disease

The genetic basis of a phenotype, whether it is a trait or a disease, is typically described as following either a monogenic or polygenic inheritance pattern.

#### 1.3.1 Monogenic inheritance

A phenotype driven exclusively by genetic variation in a single (i.e. "mono-") gene is defined as "monogenic". The term is used synonymously with "Mendelian", referencing the inheritance patterns described by Gregor Mendel (Abbott and Fairbanks, 2016). His observations in pea plant height and petal colour between parent and offspring eventually led to Mendel's Laws of Inheritance, the foundation from which we began to understand monogenic phenotypes.

"The Law of Segregation" states that during gamete formation in a parent, a gamete randomly receives a single gene allele, and through "The Law of Independent Assortment", these alleles segregate independently from other gene alleles (Castle, 1903). When the gametes from two parents meet during conception, "The Law of Dominance" states that between two different alleles for the same gene, the stronger (i.e. "dominant") allele will dominate the expression of the weaker (i.e. "recessive") allele (Castle, 1903).

A number of inheritance patterns exist for monogenic phenotypes and are dependent on: (i) the dominant and/or recessive nature of the alleles present; (ii) allelic zygosity; and

(iii) the area of genome under study, including autosomes (i.e. the non-sex chromosomes), sex chromosomes, or the mitochondrial genome.

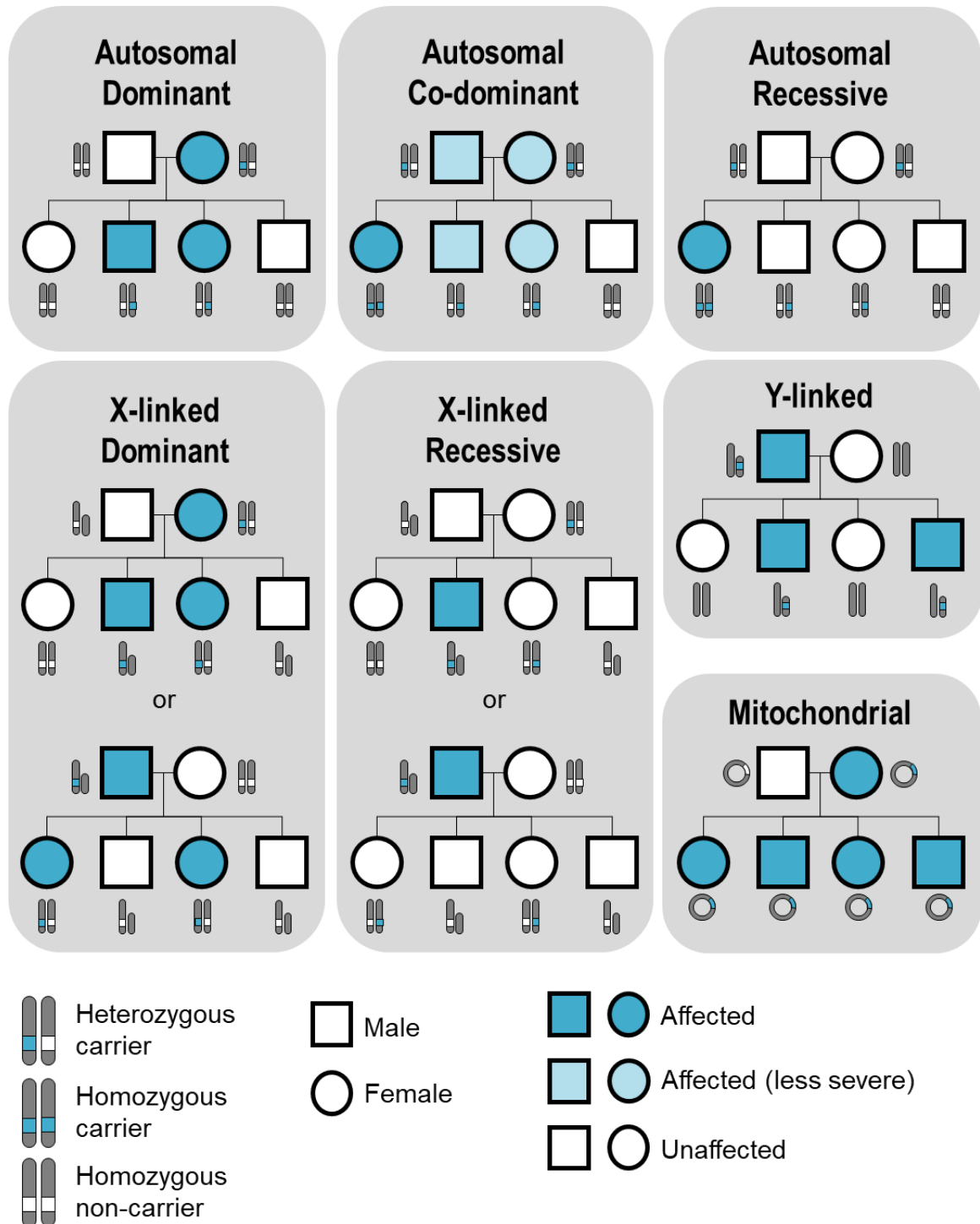
Monogenic phenotypes follow recessive, dominant, and co-dominant inheritance patterns (**Figure 1.5**). Phenotypes that only occur in the presence of two mutated alleles due to “bi-allelic” variants—either simple homozygous variants or distinct heterozygous variants on opposing alleles of the same gene—are considered to be recessive (Winsor, 1988). In contrast, autosomal dominant conditions occur in the presence of a single copy of a mutated gene allele, brought about by a heterozygous variant (Winsor, 1988). “Haploinsufficiency” is a term used in the context of autosomal dominant phenotypes to describe a gene that cannot produce a normal phenotype without two normal alleles (Deutschbauer et al., 2005), such that a heterozygous loss-of-function variant leads to half-normal net activity of the products of the gene locus, since the heterozygous normal or “wild-type” allele still functions normally. This is distinct from “dominant negative”, which describes when a mutated gene allele produces an abnormal protein that interferes with the normal functioning of the protein produced from the non-mutated allele, thus causing a dominant phenotype, but with somewhat less than half of the total possible biological activity seen in an individual with two wild-type copies of the gene.

Autosomal co-dominant phenotypes are a nuanced form of a dominant phenotype. Co-dominance is distinguished by the fact that a mutated gene allele cannot fully overcome the expression of the normal gene allele, but rather, there is co-expression of each the normal and mutated allele, resulting in an intermediate phenotype between the homozygous states for having two normal alleles or two mutated alleles.

Inheritance patterns also exist for gene variants on the X chromosome. In females, X-linked phenotypes follow the same recessive and dominant patterns as autosomal phenotypes because there are two copies of the X chromosome. However in males, due to hemizyosity for the X chromosome, a single deleterious variant will have no concomitant wild-type allele, regardless as to whether the phenotype is considered to be recessive or dominant in females who are diploid for the X chromosome. Similarly, in mutated genes that are found on the Y chromosome, the terms “dominant” and

“recessive” do not apply, since a normal diploid male will only have one Y chromosome (Winsor, 1988). If a male inherits a mutated gene allele on the X or Y chromosome, the mutated allele will be expressed by default.

Mutated genes in the mitochondrial genome follow a different inheritance pattern than those seen for autosomes and sex chromosomes. Since the mitochondrial genome exclusively follows maternal inheritance, if the mother carries a mutated mitochondrial gene and presents with a mitochondrial-related disorder, the mutation and resultant phenotype will always be present in the offspring as well (Hutchison et al., 1974).



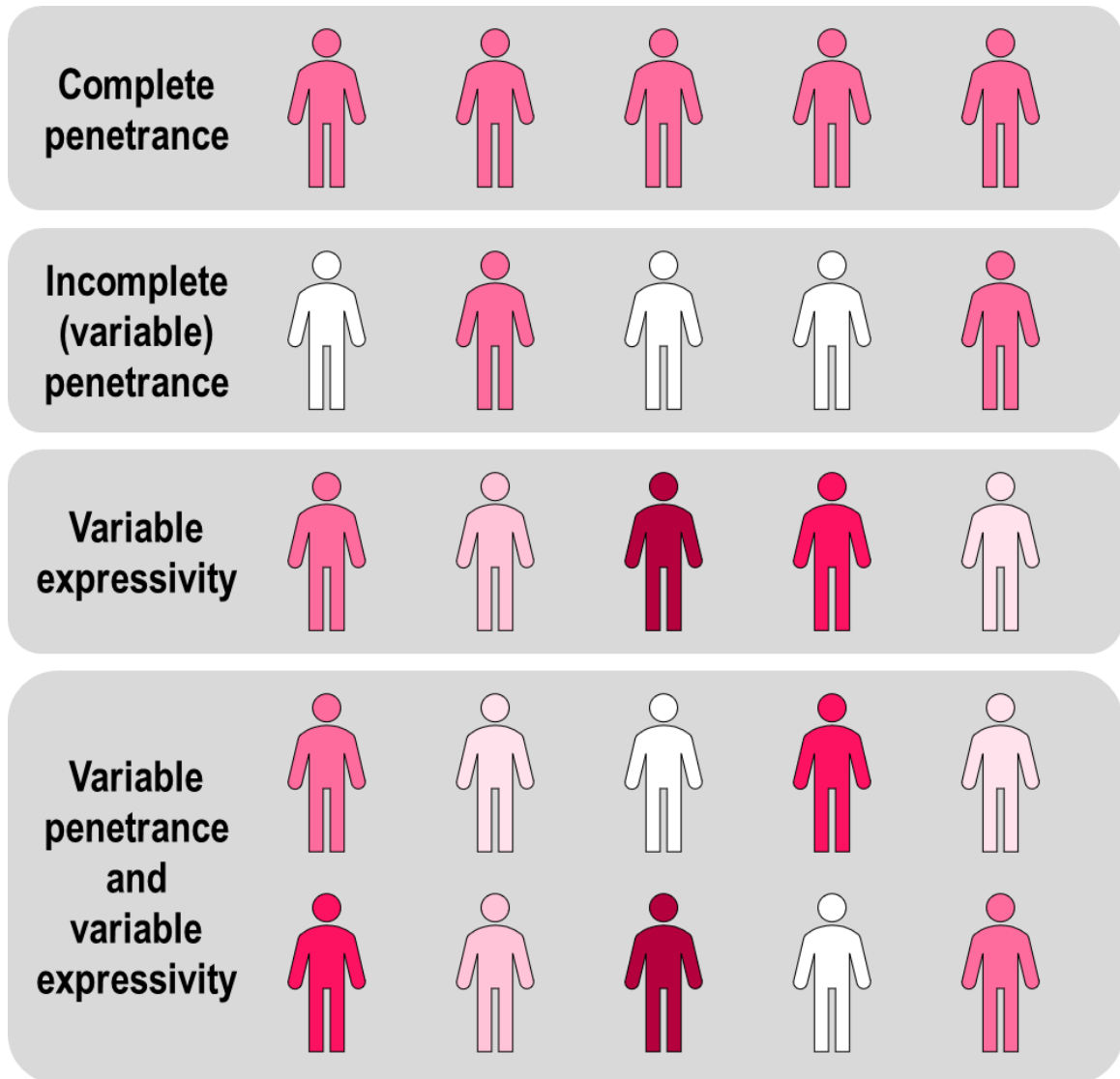
**Figure 1.5 Schematic representation of Mendelian inheritance patterns.**

The autosomal, sex chromosome, and mitochondrial inheritance patterns are provided for a single family pedigree comprised of two generations: (i) an unrelated father and mother; and (ii) four offspring: two daughters and two sons. Not shown: a father with a mitochondrial variant will not pass the variant to any offspring.

### 1.3.1.1 Penetrance and expressivity

Two important phenomena that are tied to monogenic inheritance include the “penetrance” and “expressivity” of variants (**Figure 1.6**) (Katsanis, 2016). “Penetrance” refers to the probability of carriers of the same variant expressing the same phenotype (Miko, 2008). A variant with perfect or complete penetrance would be characterized by presentation of the phenotype in 100% of carriers, while non-carriers would not express the phenotype. In contrast, incomplete penetrance refers to the situation in which carriers of the same variant do not all share the same phenotypic outcome; i.e. some proportion of carriers appear to be phenotypically normal or unaffected (Miko, 2008). Conversely, “expressivity” is a different property which refers to the situation when carriers of the same variant show differing or variable degrees of severity of a particular phenotype (Miko, 2008).

In the context of monogenic phenotypes, a highly penetrant variant with stable expressivity is typically disease-causing or phenotype-driving. However, a variant with both incomplete penetrance and variable expressivity would be difficult to classify: in some individuals, it might be disease-causing, while in others it is simply a susceptibility factor. By definition, a variant with incomplete penetrance and/or variable expressivity would be considered a “polygenic” determinant, as it alone is not enough to drive a monogenic phenotype.



**Figure 1.6 Variant penetrance and expressivity.**

Genetic variation can have differing degrees of both penetrance and expressivity. Carriers of variants with complete penetrance will always present the associated phenotype, while carriers of variants with incomplete penetrance may or may not present with the associated phenotype. Carriers of variants with variable expressivity will present with varying degrees of severity for the associated phenotype. Carriers of variants with both incomplete penetrance and variable expressivity may or may not present with some varying degree of severity for the associated phenotype. All silhouettes shown here represent carriers for a particular variant. A coloured silhouette represents an individual expressing a particular phenotype; the colour intensity reflects the severity of the phenotype.



### 1.3.2 Polygenic inheritance

In contrast to monogenic phenotypes in which a single mutated gene is the driving factor, polygenic phenotypes are the result from many (i.e. “poly-”) genetic variants found across the genome, including both common and rare variants, residing within both coding and non-coding regions (Dron and Hegele, 2018). These variants range in size from SNVs to structural variants and can have varying phenotypic impacts depending on whether the variant directly or indirectly impacts biologically relevant pathways. Variants within genes that encode proteins involved in the main mechanistic pathway tend to have larger impacts compared to variants with peripheral involvement (Boyle et al., 2017).

With a spectrum of observable variation, quantitative or continuous traits are polygenic, as they are driven by many genetic factors that differ in type, impact, and genomic location (Boyle et al., 2017; Dron and Hegele, 2018). Even extreme manifestations of quantitative traits can be polygenic in nature due to an excess of polygenic determinants with a cumulatively large phenotypic impact; however, in some instances of these extreme phenotypes, particularly those with syndromic features affecting multiple systems and organs, a monogenic basis is more likely (Frazer et al., 2009; MacArthur et al., 2014).

Because of the varying phenotypic impacts of polygenic determinants—the majority of which tend to be modest—it can be challenging in any particular individual to assign definitive causality to a set of genetic factors for an extreme quantitative trait (Marian, 2014). Rather, the accumulation of polygenic factors is described as increasing an individual’s susceptibility for the phenotype, but is not absolutely causative, deterministic or guaranteed to be associated with its expression. The degree to which these factors increase susceptibility or “risk” may also differ between individuals, as genetic variants have been shown to have varying degrees of impact (i.e. expressivity), even between family members (Wright et al., 2019).

Furthermore, polygenic phenotypes are often described as being “complex” to acknowledge the impact of not only genetic factors, but non-genetic factors as well—

such as environmental and lifestyle—on the expression of the trait. As with the small-effect genetic determinants, any of these individual components may not in and of itself be sufficient to cause phenotypic expression, but in aggregate, they act additively or synergistically. While these non-genetic factors are not a focus in the contents of this Dissertation, they should not be forgotten as important phenotypic contributors towards the presentation of quantitative traits.

### 1.3.2.1 Heritability

The term “heritability” refers to the proportion of interindividual variance observed for a particular trait that is attributed towards genetic factors (Manolio et al., 2009). For quantitative traits and diseases, the range of observable phenotypic variation suggests a spectrum of genetic factors contributing towards phenotypic presentation and susceptibility; not only does this include protein-coding variants with incomplete penetrance and variable expressivity, but intergenic variants as well.

A common challenge faced when studying polygenic phenotypes is “missing heritability”, which refers to the phenotypic expression and/or measurable variance of a particular polygenic trait that cannot fully be explained by known, associated genetic determinants (Manolio et al., 2009). In [Section 1.3.4.2](#), it is described how genotype-phenotype association methods have been utilized in an attempt to uncover additional contributory genetic factors that could help account for some instances of missing heritability for different phenotypes.

### 1.3.3 Methods to study genetic variation

#### 1.3.3.1 Sanger sequencing

The ground-breaking development of Sanger sequencing allowed researchers to effectively “read” an entire DNA sequence, which assisted in the precise identification of genetic variation in individuals. From its initial description in 1977, this sequencing method relied on DNA fragments of different lengths, generated using special chain-terminating nucleotides—one each for A, C, T, and G (Sanger et al., 1977). In the traditional Sanger method, four distinct PCR reactions were set up for each chain-

terminating nucleotide, and the randomly-sized DNA fragments from each reaction could be run on polyacrylamide gels by electrophoresis (Heather and Chain, 2016). By knowing what chain-terminating nucleotide was used for each of the samples run on the four gel lanes, the exact 5' to 3' DNA sequence could be determined by “reading” the gel from the smallest to largest DNA fragment. In the modernized version of Sanger sequencing, chain-terminating nucleotides are fluorescently labelled—one label each for A, C, T, and G—so that when the different sized DNA fragments undergo size separation by capillary gel electrophoresis, the 5' to 3' DNA sequence can also be determined based on the measured fluorescence given off by the smallest to the largest DNA fragment (Heather and Chain, 2016).

Although Sanger sequencing was a pivotal method that contributed towards the successful elucidation of the first human genome (Lander et al., 2001), it is laborious and cost-restrictive for studies that: (i) are studying larger cohorts; (ii) are interested in larger or multiple genomic areas; or (iii) are focused on gene or variant discovery.

### 1.3.3.2 Next-generation sequencing

Next-generation sequencing (NGS) techniques are an effective alternative to Sanger sequencing. NGS is a massively parallel, high-throughput sequencing approach that generates millions of sequencing reads for multiple genomic areas of interest (Shendure et al., 2017). With the high read-depth coverage generated across each sequenced nucleotide—that is, the number of times a nucleotide gets sequenced—allelic zygosity and dosage can be determined. Further, NGS can be used to sequence DNA from multiple samples simultaneously, which is revolutionary compared to what was feasible during the Sanger era. As a cost-effective method for large-scale sequencing efforts, NGS has been an incredibly useful tool in identifying phenotypically impactful variants and biologically relevant genes for both monogenic and polygenic phenotypes (Shendure et al., 2017).

A common example of NGS is whole-exome sequencing, which has been used to map disease genes and variants without the constraint for familial relationships. As the name suggests, whole-exome sequencing is a subtype of NGS that targets the exons of all

protein-coding regions of the genome (i.e. the “exome”). This technique can be utilized for studies ranging from individual assessments, small-scale family studies, and large-scale population studies (Chong et al., 2015; Cordell and Clayton, 2005; MacArthur et al., 2014; Timpson et al., 2018). With phenotype-altering SNVs and CNVs identified in almost 3000 genes and 85% of disease-causing variants being uncovered in protein-coding regions, it is unsurprising that whole-exome sequencing continues to be a successfully applied method for variant and candidate disease-gene discovery (Chong et al., 2015; Rabbani et al., 2014).

Another NGS subtype that generates data for the entire genome is aptly referred to as “whole-genome sequencing”. This method can be utilized for a range of studies, whether the focus is at an individual level or population level, or if the genetic variation of interest are protein-coding or non-coding variants, SNVs or CNVs. Although the cost to sequence a genome has dropped significantly—from over \$100,000,000 for the first human genome sequence using Sanger methods to roughly \$1000 almost 20 years later (Goodwin et al., 2016; National Human Genome Research Institute, 2020; Schwarze et al., 2020)—limitations remain. The computational resources required to bioinformatically process whole genomes are substantial and can pose as a significant barrier for both research and clinical laboratories that do not have the infrastructure to house and process the associated data files. Further, genome sequencing data is often generated with a low depth of coverage per nucleotide (2x to 4x read depth) to minimize costs and computational resources; however, this can cause an increase in incorrect genotype calls (Li et al., 2011). A generally accepted standard of 30x read depth has >99% genotype accuracy (Bentley et al., 2008), and many clinical laboratories aim for greater coverage to increase accuracy and confidence in identified variants (Rehm et al., 2013). Until these limitations are addressed, whole-exome sequencing remains a more practical NGS subtype compared to whole-genome sequencing.

#### 1.3.3.2.1 Variant interpretation

With improvements to sequencing methods, identifying variants of potential phenotypic relevance has become quite straightforward. Following the generation of NGS data, variants of interest can be identified by: (i) prioritizing those with a MAF coinciding with

the phenotype's population prevalence; (ii) considering variants with disruptive sequence ontology (i.e. nonsynonymous, frameshift); (iii) assessing predictions of how damaging a variant will be using tools that take into account nucleotide conservation between species, amino acid property changes, and protein-domain functionality; and (iv) utilizing previously published data relating to how a variant is known to alter RNA expression, protein expression, or protein function (MacArthur et al., 2014). This method typically results in a list of rare and uncommon variants. From here, when a potential variant of interest is identified, it cannot be classified or validated as “disease-causing” until: (i) functional studies have been conducted to mechanistically confirm the variant's impact through observational changes to RNA or protein expression, protein function, or protein interactions; (ii) there is a confirmed relationship between the mutated gene and the phenotype of interest; and (iii) there are statistical analyses providing evidence that the observed relationship between the variant and phenotype is not due simply to chance (MacArthur et al., 2014).

In 2015, the American College of Medical Genetics and Genomics (ACMG) published a framework to standardize the classification of identified variants of interest (Richards et al., 2015). While much of the data analysis described in this Dissertation was finalized before the wide-spread adoption of the ACMG framework<sup>1</sup>, it is worth mentioning the importance of these guidelines moving forward. The guidelines provide a number of recommendations based on categories for interpretation, such as population MAF, predictive *in silico* algorithms, functional data, segregation data, *de novo* status, and allelic data; some of these categories strongly overlap with the criteria outlined by MacArthur et al., 2014. From the ACMG guidelines, the final classification of a variant could be either: (i) pathogenic; (ii) likely pathogenic; (iii) uncertain significance; (iv) likely benign; or (v) benign (Richards et al., 2015). Importantly, a slightly altered

---

<sup>1</sup> In Chapters 2-7, when a “causal” classification cannot be assigned due to insufficient supporting functional data, variants with a high degree of evidence towards being phenotypically relevant and damaging are considered as variants that “contribute” towards disease susceptibility rather than “cause” disease. This type of consideration is particularly common when studying complex, polygenic phenotypes.

framework for classifying CNVs has also been published by the ACMG (Kearney et al., 2011).

### 1.3.3.3 Microarrays

Unlike Sanger and NGS methods that assess stretches of sequential nucleotides, microarrays genotype specific loci interspersed at relatively equal intervals across the genome (Bumgarner, 2013). These loci are typically common SNPs that fall within coding or non-coding regions; since roughly 1% of the genome encodes for proteins, the majority of microarray targets are intergenic (Bumgarner, 2013).

The general methodological overview of a microarray is relatively straightforward. First, DNA fragments containing the SNP loci of interest are captured and hybridized to a microarray chip. Subsequently, two fluorescently labelled probes are applied to the chip—one for each SNP allele—to determine the presence of different alleles at each locus, i.e. the SNP's genotype (Bumgarner, 2013).

As a relatively affordable genotyping method and the inclusion of unbiasedly selected SNPs across the genome, microarrays are a popular method in genetic association studies. While this is discussed further in the following [Section 1.3.4.2.1](#), it is important to emphasize that the SNP markers captured by microarrays are virtually never directly causative for any trait or disease. Rather, they act as an associated “tag” or “proxy” for the variant mechanistically linked to the phenotype under study.

## 1.3.4 Approaches to study the genetic basis of diseases

### 1.3.4.1 Linkage analysis in families or samples of related individuals

Large kindreds in which many members express the same disease phenotype have served as some of the original study cohorts to uncover disease-causing variants and the “disease gene” behind monogenic disorders. This is largely because a Mendelian disease's inheritance pattern can be established from a well characterized pedigree, and the presence of a sufficient number of affected and unaffected relatives allows for a statistically well-powered, case-control comparison. From this natural study design,

“linkage analyses” could be performed in an attempt to identify the genomic region containing the candidate disease-gene and variant (Teare and Barrett, 2005).

In a linkage analysis study, DNA samples from both affected and unaffected family members are obtained and DNA markers—typically polymorphic sites at SNP loci that can be detected either by restriction enzyme digestion, allele specific oligonucleotide hybridization or direct fragment sizing using gel electrophoresis—are assessed along each chromosome to establish “haplotypes” for an individual. A “haplotype” is the genetic pattern or signature of a chromosomal region or locus (International HapMap, 2005; Teare and Barrett, 2005). Due to recombination during gamete formation, haplotype patterns become increasingly diverse with each subsequent generation, as new genetic material is introduced from the biological parent external to the primary pedigree and line of descent. When haplotype markers on the same chromosome are inherited together more frequently than what would be expected by chance, the markers are said to be in “linkage disequilibrium” (LD), which refers to significant allelic association or tight correlation (Cordell and Clayton, 2005; International HapMap, 2005; Teare and Barrett, 2005).

Linkage analysis tests whether the presumed locus or variant causing a phenotype in a family is always inherited together with certain DNA markers within a region of LD. If there is no divergence between the phenotype and the DNA markers, the phenotype is considered linked to the locus. Importantly, the markers themselves are almost never the direct pathogenic cause of the disease. The metric often used to report linkage is a logarithm of the odds (LOD) score, which evaluates the probability that a phenotype and set of markers in LD are always inherited together compared to the state of complete linkage equilibrium (e.g. the DNA markers and causative variant are inherited completely independently of each other). Linkage analysis requires several variables or parameters to compute and interpret: these include the recombination fraction for the genomic area of interest, the putative inheritance pattern, the frequency of DNA marker alleles, and the structure of the chromosomal haplotype (Teare and Barrett, 2005). Traditionally, a LOD score of  $\geq 3$  (i.e. odds favoring non-random association or linkage between a DNA marker and phenotype of  $\geq 1000:1$ ) is conventionally accepted as providing strong evidence to

support co-segregation of a particular haplotype of interest and the disease under study (Teare and Barrett, 2005). LOD scores this high are obtained only if there is a large number of family members and there is not a single instance of mis-inheritance deviating from affected family members each carrying the putative linked DNA marker, while all unaffected family members lack the marker.

Typically with positive linkage, the candidate disease gene and causative variant are contained within the haplotype block. By assessing haplotype patterns across generations and between affected and unaffected relatives, a haplotype occurring exclusively in those with the phenotype is said to co-segregate with disease status, suggesting that the causative, mutated region of interest is contained within the chromosomal segment defined by the haplotype (Palmer and Cardon, 2005; Teare and Santibanez Koref, 2014). Once a region of interest has been established, it can be explored further using methods like Sanger sequencing to identify a phenotypically-disruptive variant within a biologically relevant gene (Teare and Santibanez Koref, 2014). It is also important to note that in these kindred-based studies, it is ideal that multiple, independent kindreds with the same disease phenotype can be studied and the results aggregated to build a joint LOD score, in an attempt to account for possible bias due to unmeasured genetic or environmental factors specific to a particular family (Hopper et al., 2005).

#### 1.3.4.2 Associating genotype with phenotype

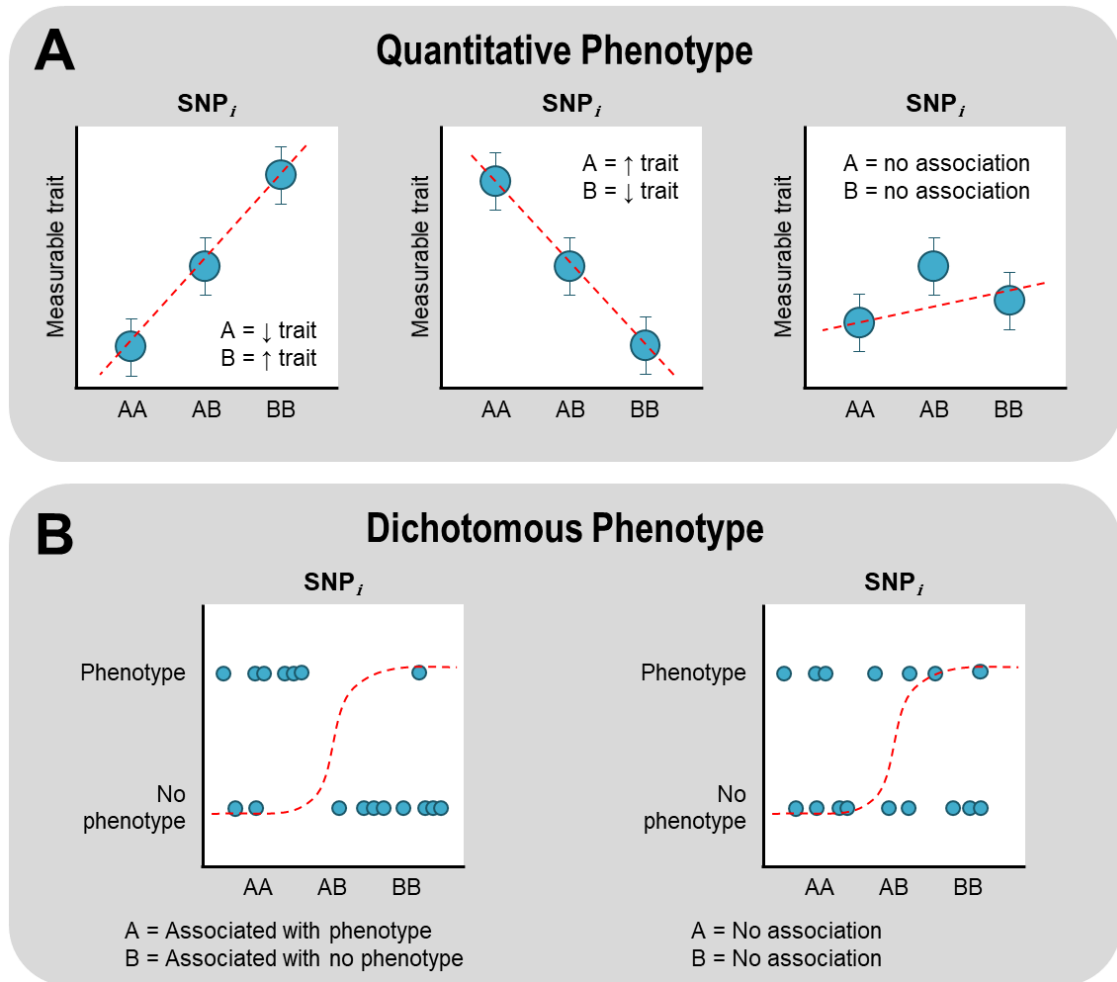
Common SNPs are incredibly informative markers of phenotypic association, serving as genetic proxies to causal variants that fall within the same LD block. Approximately 500,000 SNPs are needed to sufficiently tag all LD blocks in individuals of non-African ancestry (International HapMap, 2005; Visscher et al., 2012). With a plethora of common tag SNPs, studies have been effective in assessing whether carriers of various SNP genotypes differ statistically for a particular phenotype.

By definition, each SNP locus has two alleles and three possible genotypes (ex. AA, AB, or BB, with “A” signifying the reference allele, and “B” signifying the alternative allele). In consideration of quantitative phenotypes, linear regression is used to model the relationship between the dependent variable (i.e. phenotype of interest) and independent



variable (i.e. genotype). Since a SNP genotype may have 0, 1, or 2 alternative alleles, a linear regression model assesses how allelic dosage for the alternative allele impacts the phenotype of interest; importantly, an additive effect for each additional allele is assumed (**Figure 1.7A**). For statistically significant relationships in which the presence of an alternative allele modifies the phenotype, the beta coefficient of the regression model is designated as the “weight” or “effect” of the alternative allele. An important consideration when interpreting results from such regression models is that both variant alleles—the reference and alternative—have the same measured magnitude of effect but in different directions (Cordell and Clayton, 2005). Meanwhile, for dichotomous phenotypes, chi-squared analysis, Fisher’s exact test, or logistic regression is used to determine if there is a significant difference in the expected and observed frequencies of the phenotype under study for each SNP genotype (**Figure 1.7B**) (Cordell and Clayton, 2005). When a significant association is observed, typically the calculated odds ratio (OR) is used as the allelic weight.

For a SNP locus that is significantly associated with a quantitative trait, one allele associates with higher levels of the trait of interest, while the other allele associates with lower levels. Similarly, for a SNP locus significantly associated with a dichotomous phenotype, one allele associates with the presentation of the phenotype while the other allele associates with the absence of the phenotype. Earlier terminology such as “risk allele” or “protective allele” that was used to describe significantly associated alleles has given way to the more impartial term, “effect allele”; it is important that studies clearly indicate what allele is being considered as the “effect” allele and to what phenotypic outcome it associates with to avoid ambiguity. Importantly, effect alleles and their associated outcomes are probabilistic and not deterministic, since phenotypically normal individuals can also carry disease-associated effect alleles.



**Figure 1.7 Regression models to determine genotype-phenotype associations.**

**A)** Linear regression models can be used to determine if there is an association between variant genotypes and a quantitative, continuous phenotype. Here, examples are provided for a strong positive correlation in which “B” is associated with increasing the trait, a strong negative correlation in which “B” is associated with decreasing the trait, and no correlation between either alleles and the trait. **B)** Logistic regression models can be used to determine if there is an association between variant genotypes and a dichotomous phenotype with two outcomes. The outcome (i.e. presentation of the phenotype) could be associated with either the A or B allele, or there could be no association between either allele and the outcome. Abbreviations: SNP = single-nucleotide polymorphism.

#### 1.3.4.2.1 Genome-wide association studies

Microarrays have proven to be a very accessible method to generate genotype information for millions of SNPs across the genome. The statistical analyses described in the previous Section can be performed for each SNP captured by a microarray in large population studies, which serves as the foundation for large-scale genetic association studies such as GWASs.

GWASs allow for the simultaneous assessment of association between millions of SNPs and a particular phenotype (Tam et al., 2019). By performing GWASs in large populations that range in size from tens to hundreds of thousands of individuals, common genetic variants associated with small-to-modest effects towards a particular phenotype can be identified (Visscher et al., 2017). With simultaneous statistical testing for roughly 1 million independent SNP genotypes, a Bonferroni-corrected alpha threshold of  $5 \times 10^{-8}$  is the standard for considering whether the observed association between a SNP and phenotype meets “genome-wide significance” (Fadista et al., 2016). An important consideration is that for any significantly associated SNP, it is unlikely to be a directly causal variant for the phenotype of interest; rather, the SNP is likely tagging the truly causative variant that falls elsewhere in its LD block and was not directly genotyped by the microarray (Visscher et al., 2017).

Early GWASs successfully identified common variants with more moderately-sized phenotypic effects, and as GWAS cohorts became magnitudes larger, common variants with even smaller effects across additional loci were identified (Visscher et al., 2017). For each GWAS that is performed, related “summary statistics” are generated, detailing the genomic coordinates, reference and alternative alleles, and the estimated effect associated with the alternate alleles for the phenotype of interest.

#### 1.3.4.2.2 Rare variant association studies

Genetic association studies are often thought of in the context of common SNPs and their incremental phenotypic contributions due to the successes and discoveries of GWASs. However, rare variants with smaller phenotypic impacts that are not captured by GWASs

can also contribute towards the heritability of traits and disease (Zuk et al., 2014). This subset of rare, small-effect variants can be uncovered through “rare variant association studies” (RVASs) that also rely on large cohorts and microarray technologies, similarly to GWAS (Zuk et al., 2014).

Improvements to sequencing and genotyping technologies have provided researchers the opportunity to utilize more cost-effective and accessible methods to perform RVAS on immense populations, addressing previous challenges related to uncovering rare variant associations due to insufficiently sized cohorts and statistical power (Auer and Lettre, 2015; Lee et al., 2014). As well, modified study designs have provided additional opportunities for successful RVASs.

One RVAS design is dependent on the use of an “exome-based” microarray that specifically targets lower frequency variants within protein-coding regions, rather than the common SNPs targeted in a traditional microarray. This alternative microarray design has enabled a distinct type of RVAS to be conducted, namely an “exome-wide association study”, to assess for associations between low frequency, protein-coding variants and phenotypic traits and diseases of interest. Because of this design, there are fewer variant loci to correct for after multiple testing; Bonferroni corrections to account for exome-wide significance lead to an alpha threshold of  $5 \times 10^{-7}$  (Fadista et al., 2016). This, coupled with the ability to sequence larger cohorts due to the affordability of an exome-based microarray (compared to whole-exome sequencing), has provided opportunities for rare protein-coding variants with smaller phenotypic effects to be identified (Lee et al., 2014).

Another modified study design for an effective RVAS takes a “gene-focused” approach rather than the typical “variant-focused” approach. In a gene-focused or “gene-based” RVAS, rare variants are grouped by the gene they occur in (or some other genomic unit of consideration) and are assessed with either a burden or variance-component test (Auer and Lettre, 2015). With the underlying assumption that all rare variants have an impact towards the same phenotypic outcome, a burden test is used to determine whether carriers versus non-carriers for genetic variants are phenotypically distinct—that is, do they

significantly differ for a measurable trait mean or disease prevalence for continuous or dichotomous phenotypes, respectively (Auer and Lettre, 2015; Lee et al., 2014). Meanwhile, for variance-component tests, this method works under the assumption that variants in the same gene may have opposing effects for the same phenotype—that is, some variants could lead to an increase or decrease in a measurable trait, or could increase or decrease disease risk (Auer and Lettre, 2015; Lee et al., 2014). In a variance-component test, the measurable variance of a trait is considered between carriers and non-carriers for genetic variants rather than the mean: a larger degree of variance in carriers would suggest that the rare variants within the gene under study have measurable effects on the phenotype of interest, but in opposing directions. For a gene-based RVAS using either test method, when correcting for multiple tests under the assumption of ~20,000 genes in the human genome, the resultant alpha threshold is  $2.5 \times 10^{-6}$  (Auer and Lettre, 2015).

#### 1.3.4.3 Polygenic scores

While GWASs were useful in identifying common SNPs associated with a particular phenotype, these variants alone had limited predictive power and were not overly informative when trying to explain heritability. In 2009, the International Schizophrenia Consortium demonstrated that schizophrenia had a sizable polygenic architecture that involved thousands of common SNPs with small effects, and together, these SNPs could explain a larger degree of phenotypic variance compared to individual common variants (International Schizophrenia et al., 2009). Similarly in 2010, Yang et al. reported that the simultaneous assessment of GWAS-identified SNPs could explain a greater degree of heritability for height, another polygenic trait, compared to individual common variants (Yang et al., 2010). This method to assess the accumulation of common SNPs contributing towards a particular phenotype came to be defined as a “polygenic score” or “polygenic risk score”—the latter term being preferentially used in the context of an unfavourable disease phenotype. Specifically, polygenic scoring is used to quantify an individual's total burden of phenotype-associated effect alleles across SNP loci of interest (Choi et al., 2020).

### 1.3.4.3.1 Polygenic score development

When developing a polygenic score, the first step necessitates the selection of SNPs that will comprise the score. Early polygenic scores were constructed using a P-value thresholding approach, which involved the selection of a limited number of highly significant SNP loci, identified through GWAS for a particular phenotype. This approach was further refined, as considerations started being made to account for LD. Recall that a GWAS-identified SNP is not likely to be causal, but rather tags the causative variant that falls elsewhere in its LD block. If multiple SNPs from the same LD block are incorporated into a polygenic score, the calculation is effectively counting the same association signal multiple times, which over-inflates the score's performance (Choi et al., 2020; Prive et al., 2019). The removal of SNPs based on LD is referred to as “clumping” or “pruning”.

More recent SNP-selection methods have expanded beyond the P-value threshold approach and now consider larger numbers of SNPs, even those that are not statistically associated with the phenotype of interest. It came to be appreciated that SNPs passing genome-wide significance was somewhat arbitrary and study dependent; with sufficiently large study cohorts many previously “non-significant” loci would become nominally significant even with minimal measurable effect sizes (Dron and Hegele, 2019). This criteria liberalization and the inclusion of non-significant SNP loci allowed for orders of magnitudes of more SNPs to be considered in score development and has become popular for studies in which polygenic risk scores are being used for disease risk prediction (Choi et al., 2020).

Once the set of SNPs has been selected, the polygenic risk score calculation can be finalized. At each SNP locus, there could be 0, 1, or 2 effect alleles, depending on zygosity. Counting the total number of effect alleles ( $\omega$ ) for  $n$  SNP loci yields a maximum score of  $2n$ , indicating an individual who has inherited two effect alleles at every single locus included in the score. This provides the base equation for an unweighted polygenic risk score, which is the basic summation of effect alleles inherited by an individual:

$$\text{Polygenic risk score} = \sum_{i=1}^n (\omega_i)$$

A weighted polygenic score expands upon this base equation by integrating each effect alleles' calculated weight ( $\beta$ ) towards the phenotype of interest:

$$\text{Polygenic risk score} = \sum_{i=1}^n (\omega_i \beta_i)$$

A SNP's weight in a polygenic risk score is often the measured effect allele derived from a GWAS for the phenotype of interest. While GWAS effect estimates are widely used for polygenic risk score weights, novel statistical methods have been developed in an attempt to mitigate some of the limitations related to these estimates, including: (i) inability to adjust for LD patterns; and (ii) over-estimation of the effect for casual or tagged causal variants (i.e. Winner's Curse) (Choi et al., 2020). Different "shrinkage" methods to reduce GWAS effect estimates have been published, each with different underlying assumptions and statistical foundations; however, the polygenic risk scores described in this Dissertation use GWAS effect estimates that have not been adjusted.

When a novel polygenic score has been developed, it is crucial that: (i) the weights for each SNP were not derived from the same population in which the polygenic score is being calculated; and (ii) the score is tested and validated in two independent cohorts. These considerations are necessary to prevent overfitting of the risk score. "Overfitting" occurs when the polygenic risk score has been optimized for the cohort it was derived from; that is, if a weighted polygenic score is calculated in the same cohort from which the SNP effects were derived, then the score would perform extremely well and show strong associations between the score and phenotype of interest. However, once the score is calculated in another cohort, it would have a much poorer performance, leading to skewed results and incorrect conclusions (Choi et al., 2020). Having separate populations for weight derivation, score testing, and subsequent score validation, ensures the validity of the score and increases the confidence in any derived conclusions.

### 1.3.4.3.2 Applications of a polygenic score

Depending on the cohort under study and relevant research questions, polygenic risk scores can be used for different types of analyses. For instance, in a case-control study, polygenic risk scores can be calculated to determine differences in the accumulation of risk-associated alleles between individuals with and without the phenotype of interest. A straightforward application of this approach compares the mean calculated risk score between cases and controls; a significant difference in the mean scores indicates that cases and controls are distinct with respect to the accumulation of small-effect genetic variants. Alternatively, the proportion of cases and controls with scores above a critical threshold can be compared using chi-square analyses or Fisher's exact tests. The threshold for stratification of genetic risk is usually defined as a score percentile, determined after calculating the polygenic risk score en masse for a large population of healthy individuals and generating the distribution of scores in the general population. Individuals with an extremely high polygenic risk score—often defined as a score above the 90<sup>th</sup> percentile—are considered to have an extreme accumulation of risk-increasing alleles. This is the threshold for high polygenic risk that we have used in many studies from our laboratory. Formal evaluation tests the hypothesis that a case cohort has a much greater proportion of individuals with extreme risk scores compared to control cohorts, versus the null hypothesis that the prevalence of high score is the same in cases and controls. If statistical comparisons reject the null hypothesis, this suggests a strong polygenic component of the phenotype or disease of interest.

In other experimental situations when the study cohort is a single prospectively sampled population, instead of generating score percentiles for comparison against a different cohort, the percentiles of risk score can be determined in the single population under study. Regression models can be used to determine the association between the score and phenotypic outcome of interest and the degree of phenotypic variation that can be explained. In a regression model, the polygenic risk score can be considered a continuous independent variable—with the input either being the raw calculated score or the score's percentile—or as a binary independent variable indicating whether the score falls above or below some predetermined threshold (i.e. above or below the 90<sup>th</sup> percentile).



## 1.4 Lipids and lipoproteins

Circulating levels of lipids and lipoproteins are common examples of quantitative traits that have been heavily studied using human genetic methodologies. Concentrations of plasma lipids and lipoproteins are regulated by a complex network of genetic determinants that encode key biochemical products, including receptors, adaptor proteins, transporters, enzymes, and co-factors, each of which have distinct biological roles (Daniels et al., 2009; Dron and Hegele, 2016; Feingold and Grunfeld, 2000).

Furthermore, several secondary non-genetic factors—diet, smoking status, activity level, other medical conditions such as diabetes, obesity or hypothyroidism, and certain medications—can exacerbate the clinical presentation of lipid phenotypes and make it difficult to determine phenotypic contributions from genetic versus non-genetic sources (Brahm and Hegele, 2016; Johansen and Hegele, 2011).

Extreme deviations of lipid traits from median population levels typically suggests a more prominent, underlying genetic influence (Hegele, 2009). Relatively more common in this situation is an extreme polygenic accumulation of common variants. Less commonly, these extreme trait deviations are monogenic in nature and are driven by a single large-effect variant. Most extreme lipid phenotypes appear to have a combination of both common and rare variants comprising their underlying genetic architecture, illustrating the complexities behind understanding the genetics of lipid and lipoprotein levels. However, the precise proportion of extreme lipid phenotypes driven by common versus rare variants has not been quantified because these different types of variation have not been studied concurrently in dyslipidemic patient cohorts. Extreme deviations of plasma lipid concentrations will be explained further in [Section 1.5](#).

### 1.4.1 Lipids

Although circulating plasma lipids levels—both cholesterol and triglyceride—are recognized as risk factors for atherosclerotic CVD (ASCVD), they both also have extremely important biological roles.

### 1.4.1.1 Cholesterol

Cholesterol is an amphipathic sterol molecule made up of four hydrocarbon rings, a hydrocarbon tail and a hydroxyl group (Ikonen, 2008; Simons and Ikonen, 2000).

Cholesterol has many important physiological roles, including: (i) an integral component of all cell membranes; (ii) the backbone of steroid hormones; (iii) the precursor for bile acids; and (iv) a signalling molecule in the central nervous system (Porter and Herman, 2011; Simons and Ikonen, 2000). Our largest source of cholesterol is endogenously synthesized though the liver; only a small amount comes from exogenous, dietary origins (Feingold and Grunfeld, 2000; Iqbal and Hussain, 2009).

### 1.4.1.2 Triglyceride

Triglyceride is a non-polar lipid molecule comprised of a glycerol esterified to three fatty acid chains. These lipid molecules can be further defined by the properties of their fatty acids. Depending on the number of double-bonded carbon (C=C) molecules, triglycerides can be saturated (no C=C) or unsaturated (1 or more C=C) and be further classified depending on where the C=C occurs along the fatty acid chain.

Triglycerides are an incredibly important source of energy that are stored in adipose tissue; when metabolized, their fatty acid chains are released through hydrolysis and undergo fatty acid oxidation where they are converted into acetyl coenzyme A (acetyl-CoA) for use in the Krebs cycle and mevalonate pathway. Our primary source for triglycerides are from exogenous, dietary origins (Iqbal and Hussain, 2009).

### 1.4.1.3 Plasma lipid sources

#### 1.4.1.3.1 Exogenous

Following the ingestion of food, dietary cholesterol and triglyceride form emulsions with phospholipids, fat soluble vitamins, plant sterols and hepatically synthesized bile acids; together, these molecules form mixed micelles (Feingold and Grunfeld, 2000; Iqbal and Hussain, 2009). In the duodenum of the small intestine, micelle contents are hydrolyzed by pancreatic enzymes, resulting in free fatty acids, mono- and di-acylglycerols, and glycerols (Feingold and Grunfeld, 2000; Iqbal and Hussain, 2009). Contents of these

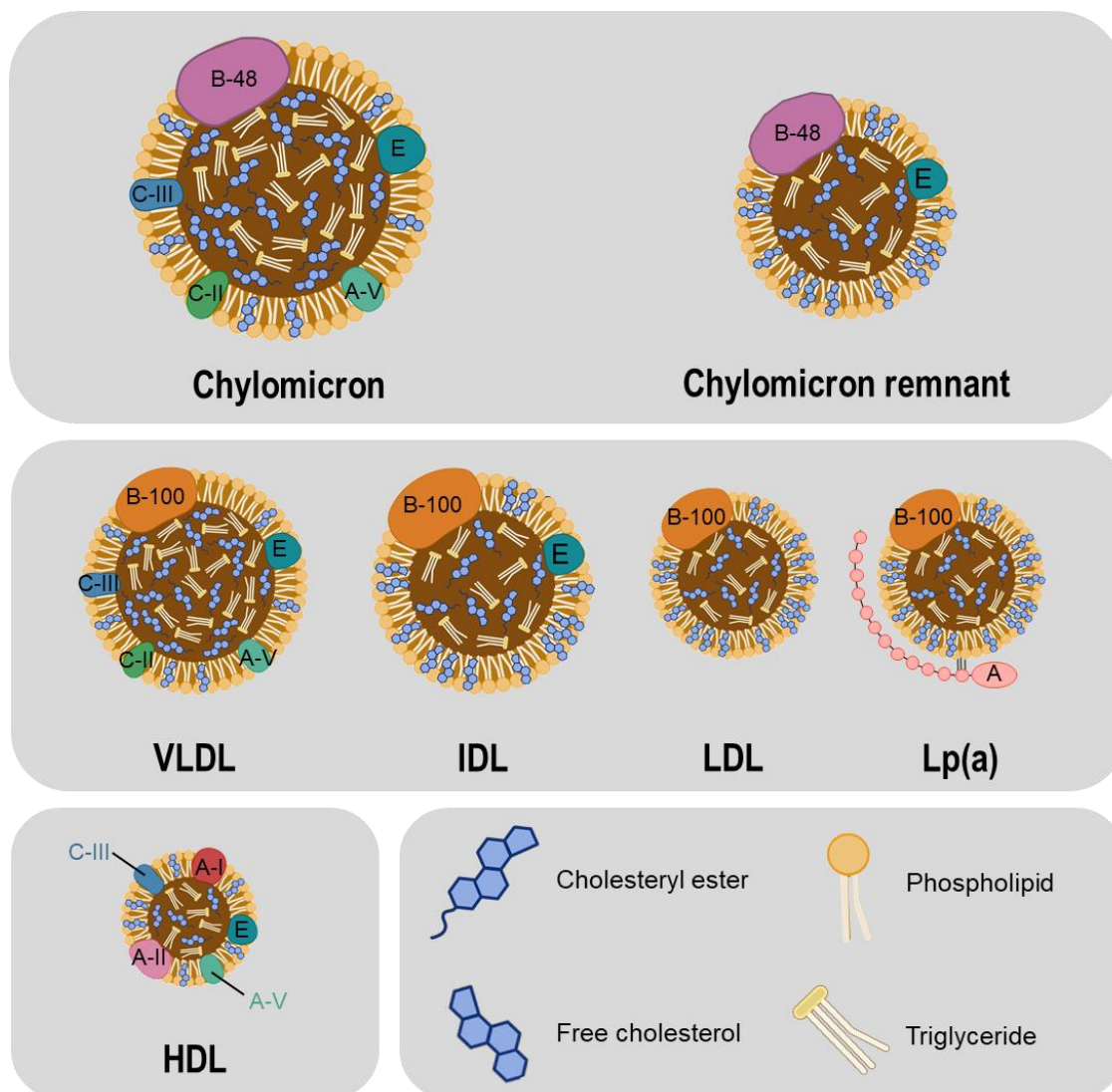
micelles can be absorbed by intestinal enterocytes in the jejunum; fatty acids and glyceride compounds are taken up through both passive and active diffusion, while cholesterol is absorbed by the Niemann-Pick C1 like 1 protein (NPC1L1) transporter (Feingold and Grunfeld, 2000; Iqbal and Hussain, 2009). Once absorbed, these molecules can be resynthesized into triglycerides and cholesteryl esters for subsequent lipoprotein assembly in the intestine, which is discussed in the upcoming [Section 1.4.2.2.1](#). Of the cholesterol that is taken up by the enterocytes, about 50% is actively transported back into the intestine by ATP-binding cassette transporter G5 (ABCG5) and by ATP-binding cassette transporter G8 (ABCG8) for excretion. The majority of remaining bile acids are reabsorbed by the terminal ileum of the small intestine and return to the liver (Feingold and Grunfeld, 2000; Iqbal and Hussain, 2009).

#### 1.4.1.3.2 Endogenous

*De novo* cholesterol synthesis can occur within hepatocytes. In low states of free cellular cholesterol, sterol regulatory element binding protein (SREBP) transcription factors become activated and upregulate a number of cholesterol metabolism regulators, including the main enzyme involved in cholesterol synthesis,  $\beta$ -hydroxy  $\beta$ -methylglutaryl-coenzyme A (HMG-CoA) reductase (Ikonen, 2008; Simons and Ikonen, 2000). This enzyme is the rate-limiting step in cholesterol synthesis; an increase in the protein's expression ultimately leads to an increase in the production of free cholesterol via the HMG-CoA reductase or mevalonate pathway, in which acetyl-CoA is the starting molecule (Ikonen, 2008). *De novo* triglyceride synthesis also occurs within hepatocytes, using free fatty acids derived from fatty acid synthesis and glycerol derived from glycolysis (Alves-Bezerra and Cohen, 2017). The newly synthesized lipids are assembled into hepatically-derived lipoproteins, which is discussed in the upcoming [Section 1.4.2.2.2](#).

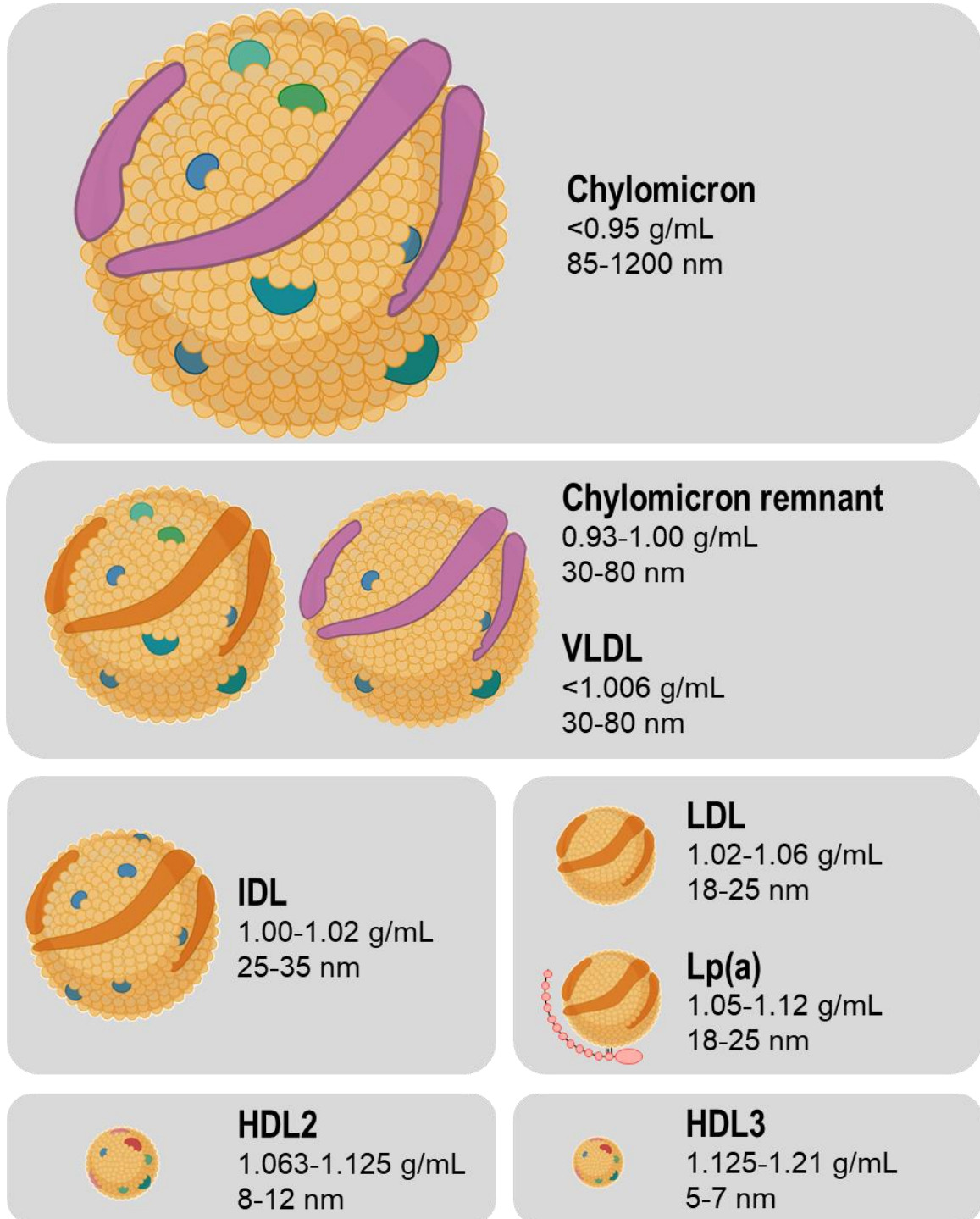
## 1.4.2 Lipoproteins

Due to the insoluble nature of cholesterol and triglyceride, lipoprotein particles are responsible for transporting these lipid molecules throughout the body. Lipoproteins are discrete macromolecular entities that vary in size, density and composition (**Figure 1.8**). These unique features arise because of qualitative and quantitative differences in their: (i) characteristic lipid-associated proteins or “apolipoproteins” (apo); (ii) amount and ratio of cholesterol and triglyceride content; and (iii) other lipids species, such as sphingolipids and phospholipids (**Figure 1.9**) (Feingold and Grunfeld, 2000; Hegele, 2009). At a first level of approximation, lipoproteins can be classified based on their cholesterol and triglyceride content. The main cholesterol-carrying lipoproteins include LDL and HDL, while chylomicrons and very-low-density lipoproteins (VLDL) are the main triglyceride-carrying lipoproteins (Feingold and Grunfeld, 2000); VLDL also carries cholesterol, whose molar concentration is about one-third that of triglyceride, meaning that it is relatively cholesterol-poor and thus less dense compared to LDL and HDL. After VLDL is secreted by the liver and remodelled through the lipolytic process (discussed further in [Section 1.4.2.2.2](#)), the resulting remnant particle, sometimes called intermediate-density lipoprotein (IDL), is smaller, more dense and more cholesterol-rich. However, IDL is not usually considered as a primary carrier of either lipid.



**Figure 1.8 Lipid and apolipoprotein composition of lipoprotein classes.**

This diagram shows the cross-sectional view of different lipoproteins. Lipoproteins are complex macromolecules made up of different combinations of lipids—free cholesterol, cholesteryl ester, phospholipid, triglyceride—and apolipoproteins. The major apolipoprotein constituents are shown for each lipoprotein. Abbreviations: apo = apolipoprotein; A = apo(a); A-I = apo A-I; A-V = apo A-V; B-48 = apo B-48; B-100 = apo B-100; C-II = apo C-II; C-III = apo C-III; E = apo E; HDL = high-density lipoprotein; IDL = intermediate-density lipoprotein; LDL = low-density lipoprotein; Lp(a) = lipoprotein(a); VLDL = very-low-density lipoprotein. Biological images adapted from <https://biorender.com/>.



**Figure 1.9 Lipoprotein classes and their physical characteristics.**

Lipoproteins are classified based on size (diameter, nm) and density (g/mL). These characteristics are driven by a particle's composition of lipids and apolipoproteins (**Figure 1.8**). Abbreviations: HDL = high-density lipoprotein; IDL = intermediate-density lipoprotein; LDL = low-density lipoprotein; Lp(a) = lipoprotein(a); VLDL = very-low-density lipoprotein. Biological images adapted from <https://biorender.com/>.

### 1.4.2.1 High-density lipoprotein

HDL is a heterogeneous particle comprised of both apolipoproteins and lipids. As the smallest and densest of the lipoproteins, more than half of an HDL particle is protein-based; apo A-I is the defining protein of HDL and is the main structural component (Fisher et al., 2012). There is substantial intra-particle variation, with differences in composition, size, and charge, prompting the further subclassification of HDL particles into HDL2 (larger and less dense) and HDL3 (smaller and more dense) (Fisher et al., 2012; Tosheska-Trajkovska and Topuzovska, 2017). Despite this distinction, when referring to circulating levels of HDL cholesterol in an individual, this measurement encompasses the total amount of cholesterol associated to all types of HDL particles. The primary physiological role of HDL is to transport cholesterol from peripheral tissues to the liver for eventual excretion in a process called “reverse cholesterol transport”; this transport pathway largely overlaps the natural HDL lifecycle (**Figure 1.10**) (Ouimet et al., 2019), which is explained in [Section 1.4.2.1.2](#).

#### 1.4.2.1.1 High-density lipoprotein lifecycle

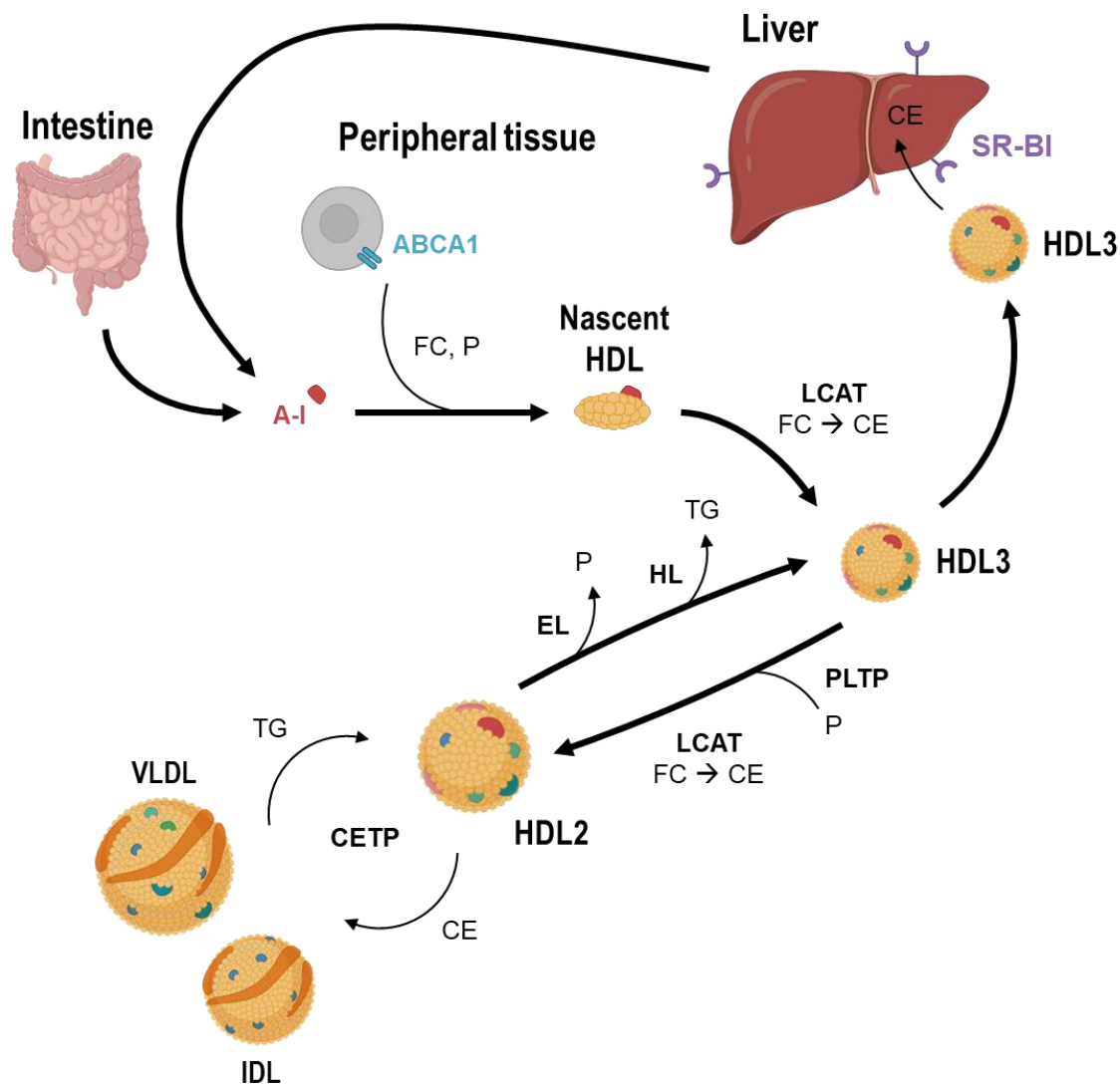
The synthesis of HDL particles begins with the production of apo A-I from hepatic and intestinal sources (Fisher et al., 2012; Tosheska-Trajkovska and Topuzovska, 2017). Free apo A-I has a high affinity for cholesterol and becomes lipidated following the expulsion of free cholesterol and phospholipids from peripheral tissues by ATP-binding cassette transporter A1 (ABCA1) (Fisher et al., 2012; Tosheska-Trajkovska and Topuzovska, 2017). The newly lipid-associated apo A-I takes on a discoidal shape and is considered a nascent HDL particle. From here, hepatically-synthesized circulating lecithin-cholesterol acyltransferase (LCAT) esterifies the free cholesterol of nascent HDL following its activation by apo A-I (Fisher et al., 2012; Tosheska-Trajkovska and Topuzovska, 2017). This prompts a structural change of the lipoprotein into a more spherical shape due to the increased cholesteryl ester content, which marks the transition of nascent HDL into HDL3 (Tosheska-Trajkovska and Topuzovska, 2017). When HDL3 undergoes further esterification by LCAT and acquires additional phospholipids via phospholipid transfer

protein (PLTP), the lipoprotein matures into an HDL2 particle (Albers et al., 2012; Daniels et al., 2009; Tosheska-Trajkovska and Topuzovska, 2017).

Following maturation into HDL2, the particle is additionally modified by CETP, which originates from both hepatocytes and adipocytes (Daniels et al., 2009; Tosheska-Trajkovska and Topuzovska, 2017). CETP is the main protein involved in the transfer of cholesteryl esters from HDL2 particles to triglyceride-rich lipoproteins in exchange for triglyceride (Tosheska-Trajkovska and Topuzovska, 2017). Because of the increased triglyceride content, this triglyceride-carrying HDL2 particle becomes a target for hepatic lipase (HL), a hepatically secreted enzyme that hydrolyzes triglyceride molecules into free fatty acids (Tosheska-Trajkovska and Topuzovska, 2017). These particles may be further targeted by endothelial lipase (EL), which works to hydrolyze the phospholipids of HDL2 (Paradis and Lamarche, 2006). Together, HL and EL help generate smaller HDL particles, often back into the HDL3 subclass (Daniels et al., 2009; Tosheska-Trajkovska and Topuzovska, 2017; Yu et al., 2018).

HDL3 particles bind with high affinity to scavenger receptor class B type I (SR-BI) located on the cell surface of many tissues, particularly the liver (Tosheska-Trajkovska and Topuzovska, 2017). Cholesteryl esters dissociate from the particle and are moved into the liver for delivery from peripheral tissues.





**Figure 1.10 Metabolic lifecycle of HDL particles.**

After lipidation, apo A-I particles and associated lipid molecules take on a discoidal shape and become nascent HDL. Further modification by LCAT leads to particle maturation into HDL3. HDL3 particles can either interact with hepatic SR-BI and lose some of its cholesteryl ester content, or mature into HDL2 after further modification by PLTP and LCAT. HDL2 particles can exchange lipid content with triglyceride-rich lipoproteins via CETP, and then be modified by HL and EL, back into HDL3 particles. Abbreviations: ABCA1 = ATP-binding cassette transporter A1; apo = apolipoprotein; A-I = apo A-I; CE = cholesteryl ester; CETP = cholesteryl ester transfer protein; EL = endothelial lipase; FC = free cholesterol; HDL = high-density lipoprotein; HL = hepatic lipase; IDL = intermediate-density lipoprotein; LCAT = lecithin-cholesterol acyltransferase; P = phospholipid; PLTP = phospholipid transfer protein; SR-BI = scavenger receptor class B type I; TG = triglyceride; VLDL = very-low-density lipoprotein. Biological images adapted from <https://biorender.com/>.

#### 1.4.2.1.2 Reverse cholesterol transport

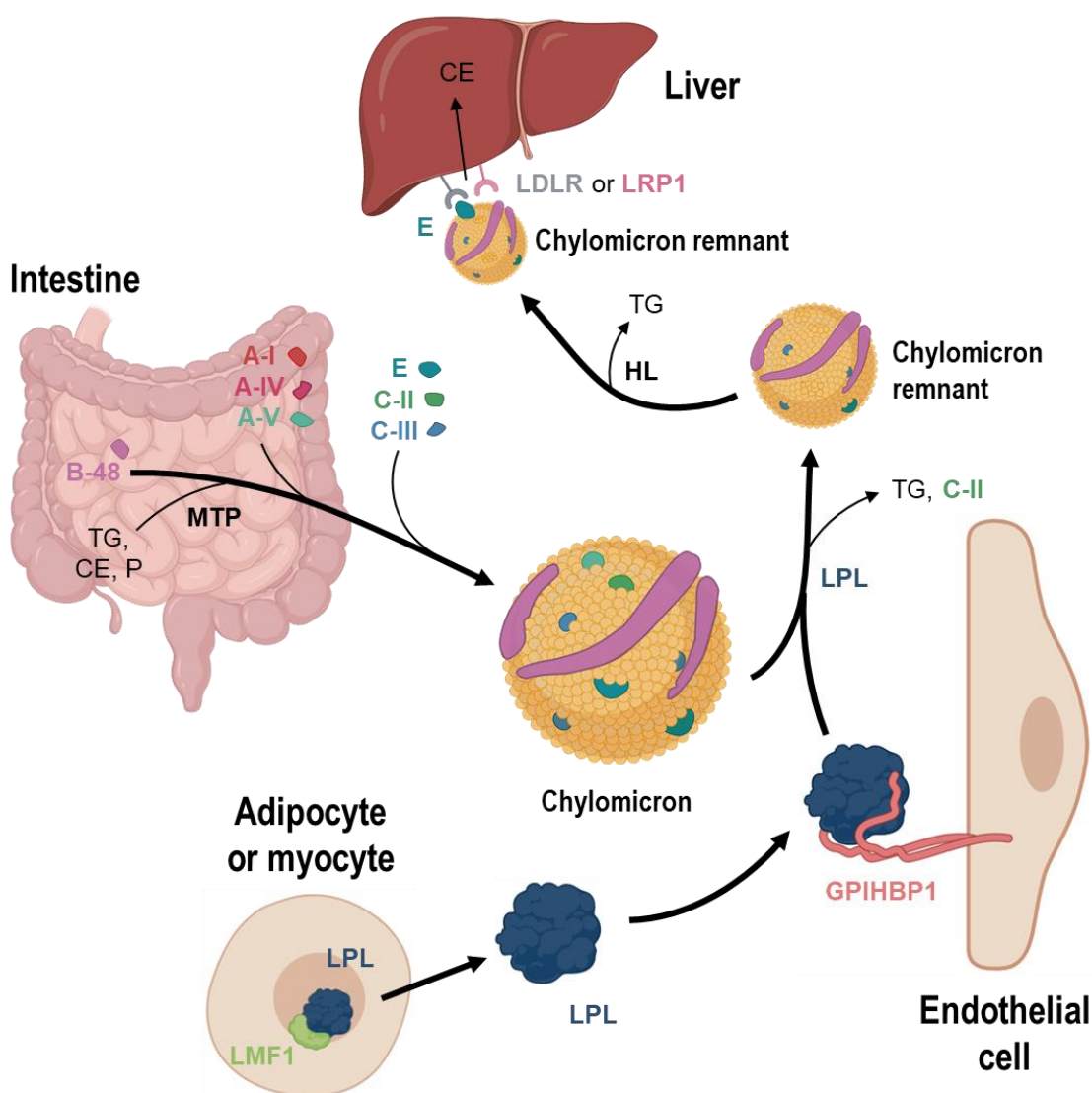
The reverse cholesterol transport pathway encompasses the movement of excess cellular cholesterol—following the HDL lifecycle—out of peripheral tissues by HDL and its delivery to the liver for excretion or recycling into bile acids and salts (Ouimet et al., 2019). The transport of cholesterol out of macrophages has been a focus of interest related to risk for ASCVD—cholesterol-laden macrophages can develop into foam cells, which are a prominent component of atherosclerotic lesions in the vascular wall.

A measure of the reverse cholesterol transport process is tied to an HDL particle's ability to accept cholesterol: this measure of HDL functionality is referred to as “cholesterol efflux”. There is a strong inverse correlation between HDL cholesterol efflux and ASCVD that is independent from HDL cholesterol levels (Khera et al., 2011); as such, this functional metric of HDL has been shown to be a better measure of ASCVD risk compared to measurable levels of HDL cholesterol (Rader and Hovingh, 2014).

#### 1.4.2.2 Triglyceride-rich lipoproteins

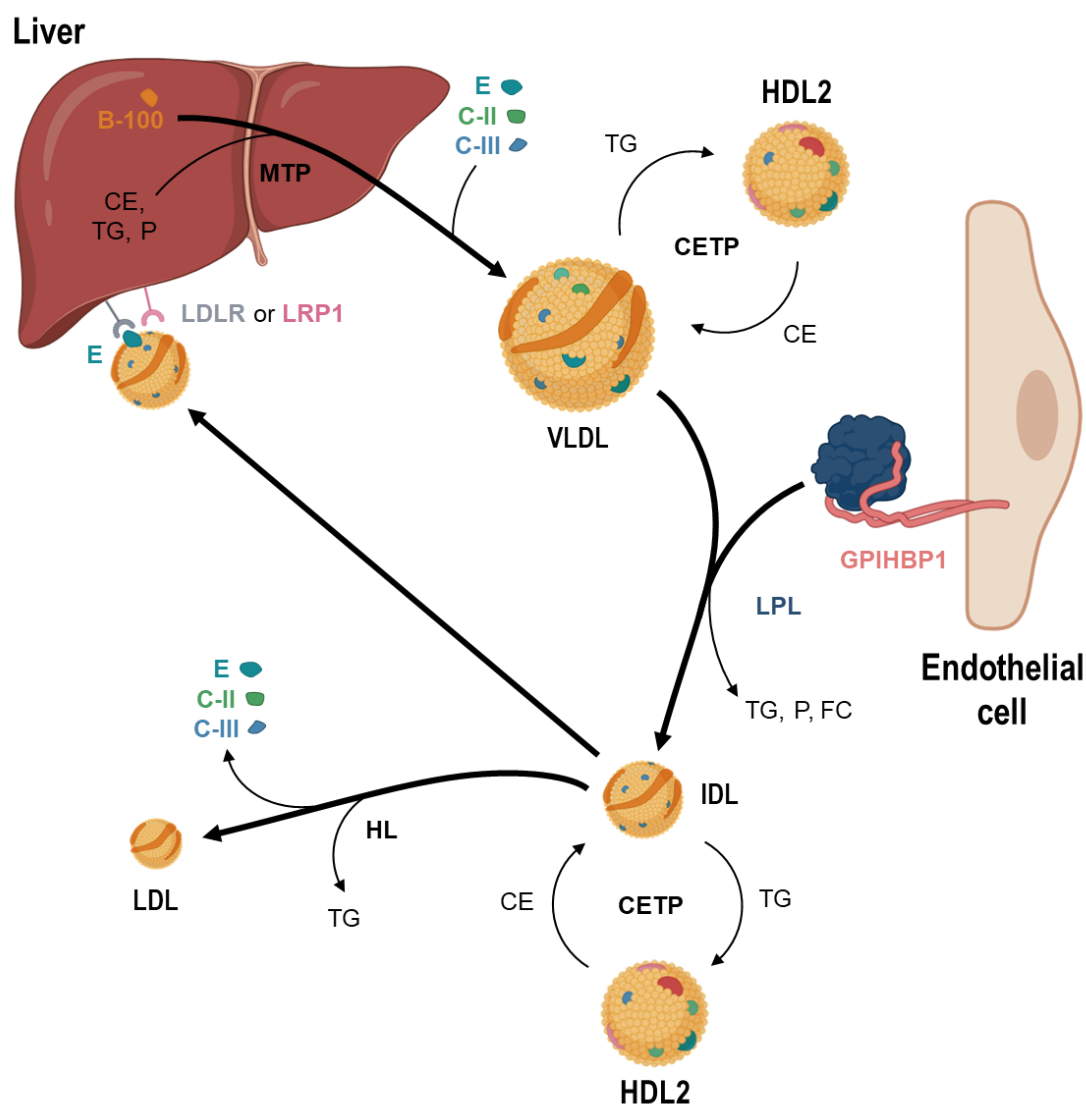
Measured circulating triglyceride levels represent the integrated measurement of triglyceride molecules carried by all circulating triglyceride-rich lipoprotein species: chylomicrons, VLDL, and their metabolic remnants, including IDL. To a much lesser degree, triglyceride is carried within the main cholesterol-carrying lipoproteins, but the contribution of their triglyceride content to the total plasma measurement is miniscule (Dron and Hegele, 2017b).

There are two distinct classes of triglyceride-rich lipoproteins: (i) those containing apo B-48 (chylomicrons and remnants) (**Figure 1.11**), and (ii) those containing apo B-100 (VLDL, IDL, and remnants) (**Figure 1.12**). While certain proteins are involved in both metabolic pathways, the lifecycle of these lipoproteins are largely independent, with intestinal and hepatic origins, respectively (Feingold and Grunfeld, 2000).



**Figure 1.11 Metabolic lifecycle of chylomicrons and their remnant particles.**

Chylomicrons are assembled in the intestine, with apo B-48 as the structural scaffold. After entering circulation, additional apolipoproteins are added to the particle. Within the vasculature, endothelial-bound LPL hydrolyzes triglycerides from circulating chylomicrons after interactions with apo C-II and apo A-V. Apo C-III inhibits the hydrolytic action of LPL. The resultant chylomicron remnant loses additional triglyceride content due to HL, and once enriched in apo E, can interact with hepatic LDLR and LRP1 for uptake into the liver. Abbreviations: apo = apolipoprotein; A-I = apo A-I; A-IV = apo A-IV; A-V = apo A-V; B-48 = apo B-48; C-II = apo C-II; C-III = apo C-III; CE = cholesteryl ester; E = apo E; GPIHBP1 = glycosylphosphatidylinositol-anchored high-density lipoprotein binding protein 1; HL = hepatic lipase; LDLR = low-density lipoprotein receptor; LMF1 = lipase maturation factor 1; LPL = lipoprotein lipase; LRP1 = LDL-related 1 protein; MTP = microsomal triglyceride transfer protein; P = phospholipid; TG = triglyceride. Biological images adapted from <https://biorender.com/>.



**Figure 1.12 Metabolic lifecycle of VLDL, IDL, and their remnant particles.**

VLDL is assembled in the liver with apo B-100 as the structural scaffold. After entering circulation, additional apolipoproteins are added to the particle. VLDL exchanges lipid content with HDL2 via CETP. Within the vasculature, endothelial-bound LPL hydrolyzes triglycerides from VLDL. The resultant IDL particles also exchange lipid content with HDL2 via CETP and can either be taken up by the liver through interactions between apo E and LDLR or LRP1, or can be further modified by HL, lose additional apolipoproteins, and become LDL. Abbreviations: apo = apolipoprotein; B-100 = apo B-100; C-II = apo C-II; C-III = apo C-III; CE = cholesteryl ester; CETP = cholesteryl ester transfer protein; E = apo E; FC = free cholesterol; GPIHBP1 = glycosylphosphatidylinositol-anchored high-density lipoprotein binding protein 1; HDL = high-density lipoprotein; HL = hepatic lipase; IDL = intermediate-density lipoprotein; LDL = low-density lipoprotein; LDLR = low-density lipoprotein receptor; LPL = lipoprotein lipase; LRP1 = LDL-related 1 protein; MTP = microsomal triglyceride transfer protein; P = phospholipid; TG = triglyceride; VLDL = very-low-density lipoprotein. Biological images adapted from <https://biorender.com/>.

#### 1.4.2.2.1 Chylomicron metabolism

Chylomicrons are the main lipoprotein responsible for delivering endogenously acquired dietary fats to different areas of the body for energy utilization (Xiao et al., 2019).

Intestinally absorbed cholesterol and fatty acids are re-esterified into cholesteryl esters and triglycerides, respectively. Meanwhile, *APOB* is transcribed within enterocytes and edited by apo B mRNA editing enzyme catalytic subunit 1 (APOBEC-1), such that the resultant mRNA is translated into a protein that is 48% of the size of the original apo B-100 protein, namely apo B-48 (Daniels et al., 2009), the main structural component of chylomicrons. A combination of triglyceride, cholesteryl esters, and phospholipids are assembled around the apo B-48 backbone by microsomal triglyceride transfer protein (MTP) to form a pre-chylomicron particle (Daniels et al., 2009). Chylomicrons do not become fully mature until they have moved from the endoplasmic reticulum to the Golgi apparatus to the cytoplasm, where additional apolipoproteins, including apo A-I, A-IV, and V, are added (Xiao et al., 2019). Fully matured chylomicrons are then able to enter the lymphatic system and eventually enter the circulatory system through the jugular vein; during this time, chylomicrons are modified through the addition of apo C-II, C-III, and E, which are relevant for downstream enzymatic interactions (Feingold and Grunfeld, 2000).

Once circulating, chylomicrons interact with lipoprotein lipase (LPL), the main enzyme responsible for hydrolyzing triglyceride-rich lipoproteins. As an extracellular lipase, LPL is anchored to the endothelial lining of vascular networks throughout adipose and muscle tissue, and interacts with circulating lipoproteins (Lambert and Parks, 2012; Zilversmit, 1995). The catabolic action of LPL removes triglyceride from the core of chylomicrons, where they can be stored as energy reserves in adipose or used for metabolic processing in muscle (Boullart et al., 2012; Lambert and Parks, 2012). Through these actions, triglyceride levels are endogenously maintained.

Apo C-II and apo A-V are important constituents of chylomicron particles that are required for LPL hydrolysis (Daniels et al., 2009; Feingold and Grunfeld, 2000; Hegele, 2016). As a co-factor for LPL, apo C-II is essential for the interaction between circulating

chylomicrons and anchored LPL (Kei et al., 2012). And although its precise mechanism of action is not clearly determined, apo A-V normally enhances LPL function indirectly by interacting with glycosylphosphatidylinositol-anchored high-density lipoprotein binding protein 1 (GPIHBP1) (Forte et al., 2016).

Other proteins are necessary for proper LPL functioning. Lipase maturation factor 1 (LMF1) is a chaperone bound to the endoplasmic reticulum that assists in the folding and maturation of LPL (Doolittle et al., 2010). Another critical protein in the early stages of the LPL life-cycle is GPIHBP1. Following its interaction with LMF1, LPL is transported and anchored to the endothelial lining of the vascular wall by GPIHBP1 (Young and Zechner, 2013). When LPL dissociates from the cell surface through indirect inhibition of angiopoietin-like protein 3 (ANGPTL3), triglyceride hydrolysis from triglyceride-rich lipoproteins stops (Tikka and Jauhiainen, 2016).

Subsequent to the hydrolyzing action of LPL, chylomicron particles become smaller remnant particles and lose apo C-II (Daniels et al., 2009; Feingold and Grunfeld, 2000). With the resultant enrichment of apo E, chylomicron remnants undergo additional remodelling by HL and are taken up by hepatocytes through endocytosis, mediated by the LDL receptor (LDLR) and LDL-related 1 protein (LRP1), both of which have a binding affinity for apo E (Daniels et al., 2009). The lipid molecules taken up by the liver are hydrolyzed and can be used in VLDL synthesis, while apo E is released back into circulation (Daniels et al., 2009).

#### 1.4.2.2.2 Very-low-density lipoprotein metabolism

Endogenously synthesized lipids are transported out of the liver by VLDL. Since APOBEC-1 is not expressed in hepatocytes, the full form of *APOB* can be produced, namely apo B-100—the main structural component of both VLDL and LDL (Daniels et al., 2009; Feingold and Grunfeld, 2000). Like in the intestine, hepatic MTP aggregates triglyceride, cholesteryl esters, and phospholipids to the apo B-100 scaffold within the rough endoplasmic reticulum to form the basis of VLDL particles (Feingold and Grunfeld, 2000). As it matures throughout the cell, VLDL is released into circulation, where these nascent particles take up apo C-II, C-III, apo E, and cholesteryl esters from

HDL particles through interactions with CETP (Tosheska-Trajkovska and Topuzovska, 2017). At this point, hydrolysis of triglyceride molecules in VLDL by LPL mimics the metabolic pathway described for chylomicrons in Section 1.4.2.2.1.

As VLDL decreases in size through catabolic interactions, it loses a number of surface constituents, including phospholipids, free cholesterol, and apolipoproteins; these remnant particles are referred to as IDL. IDL can interact further with CETP, exchanging its triglyceride content for additional cholesteryl esters from HDL (Daniels et al., 2009). At this point, IDL may either be taken up by the liver through interactions between apo E and LDLR, or may undergo further triglyceride hydrolysis by HL and become an LDL particle after losing any remaining apo E, C-II, and C-III molecules (Daniels et al., 2009). The latter pathway allows for the delivery of cholesteryl esters to peripheral tissues via LDL transport. The metabolic pathway for LDL particles will not be discussed, as it is beyond the scope of this Dissertation.

## 1.5 Dyslipidemia

“Dyslipidemia” is defined as an extreme deviation of plasma lipid concentration, which is often due to dysfunctional lipid-related biochemical products including receptors, adaptor proteins, transporters, enzymes, and co-factors that disrupt the metabolic synthesis, processing, function, or catabolism of lipoproteins (Hegele, 2009). Many genetic factors—ranging in population frequency, ontology, and functional consequence—are often responsible for these dysfunctional metabolic proteins (Hegele, 2009); however, in some scenarios, a dyslipidemic profile can also be driven or exacerbated by non-genetic factors such as lifestyle behaviours (Cole et al., 2015; Dron and Hegele, 2016; Hegele, 2009). Depending on the impacted lipoprotein(s), affected lipid trait(s), and additional phenotypic features, a more specific dyslipidemia diagnosis may be given.

### 1.5.1 Genetics of dyslipidemia

There are 24 named dyslipidemias with a variety of genetic underpinnings (**Table 1.1**) (Hegele et al., 2015). Most of these disorders were characterized at the molecular level >10 years ago using classical biochemical and genetic mapping methods, which allowed researchers to establish the important, casual genes related to each disease (Breslow,

2000; Hegele et al., 2015); the encoded protein products of these mutated genes have important roles in the metabolic pathway of relevant lipoprotein species, many of which were described in [Section 1.4](#). The rarity of these phenotypes are reflected by the infrequency in which their causative genetic factors are seen in the general population.

The majority of dyslipidemia cases are polygenic, resulting from the contributions of several types of genetic determinants that predispose an individual towards a more severe presentation of a lipid trait (Dron and Hegele, 2018; Kathiresan et al., 2009).

Incompletely penetrant, rare variants in genes encoding lipid-related biochemical products contribute to polygenic dyslipidemias by conferring a state of susceptibility in carriers (Hegele, 2009). Often, these variants are seen at an increased frequency in cohorts of dyslipidemia cases compared to cohorts of healthy controls; however, these variants do not completely co-segregate with abnormal phenotypes in pedigrees.

Nonetheless, their strong statistical relationship with perturbed lipids in dyslipidemia patients support their contributory role, although are not independently causative per se in a particular individual (Dron and Hegele, 2018). This distinction reflects the difference between determinism—i.e., rare variants directly cause specific monogenic dyslipidemia phenotypes—versus probability—i.e., rare variants act as polygenic contributors (among other factors) leading to susceptibility to dyslipidemia.



**Table 1.1 Dyslipidemia phenotypes and their genetic etiologies.**

<b>Lipid phenotype</b>	<b>Clinical diagnosis</b>	<b>Genetic basis</b>	<b>Gene(s)</b>
High LDL cholesterol	Familial hypercholesterolemia	Co-AD	<i>LDLR; APOB; PCSK9</i>
	Hypercholesterolemia	Polygenic	
	Phenocopy of familial hypercholesterolemia	AD	<i>APOE</i>
		AR	<i>LDLRAP1; LIPA</i>
Low LDL cholesterol	Sitosterolemia	AR	<i>ABCG5; ABCG8</i>
	Abetalipoproteinemia *	AR	<i>MTP</i>
	Hypobetalipoproteinemia	AR; AD	<i>APOB</i> *
		AD	<i>PCSK9</i>
		AR	<i>SAR1B</i>
Combined hypolipidemia *	AR	<i>ANGPTL3</i>	
High Lp(a)	Hyperlipoproteinemia(a)	Co-AD	<i>LPA</i>
Low HDL cholesterol	Tangier disease	AR	<i>ABCA1</i>
	Apo A-I deficiency	AR	<i>APOA1</i>
	Familial LCAT deficiency	AR	<i>LCAT</i>
	Fish-eye disease	AR	<i>LCAT</i>
	Hypoalphalipoproteinemia	Polygenic	
	CETP deficiency	AR	<i>CETP</i>
High HDL cholesterol	SR-BI deficiency	AR	<i>SCARB1</i>
	Hepatic lipase deficiency	AR	<i>LIPC</i>
	Endothelial lipase deficiency	AR	<i>LIPG</i>
	Hyperalphalipoproteinemia	Polygenic	
High triglyceride	Familial chylomicronemia syndrome	AR	<i>LPL; LMF1; GPIHBP1; APOA5; APOC2</i>
	Infantile hypertriglyceridemia	AR	<i>GPD1</i>
	Dysbetalipoproteinemia	AR; AD; polygenic	<i>APOE</i>
	Multifactorial chylomicronemia	Polygenic	
	Mild-to-moderate hypertriglyceridemia	Polygenic	
Low triglyceride	Hypotriglyceridemia	AR; AD; polygenic	<i>APOC3; ANGPTL3; ANGPTL4</i>

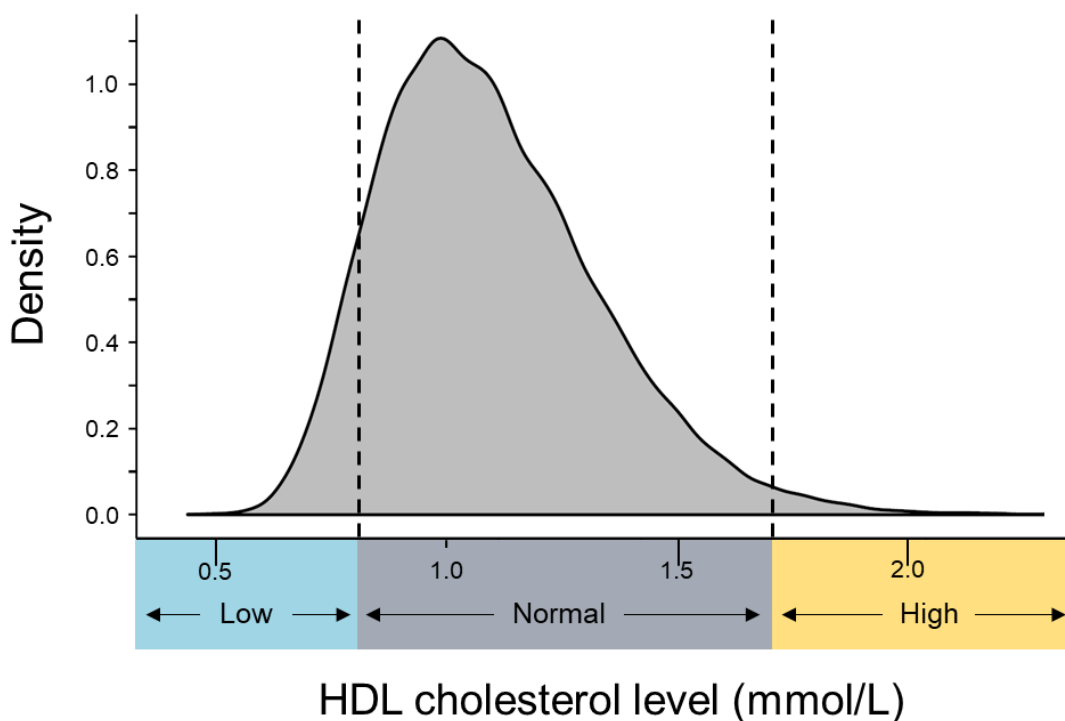
“\*” denotes a phenotype that also has a low triglyceride levels. Abbreviations: apo = apolipoprotein; AR = autosomal recessive; AD = autosomal dominant; CETP = cholesteryl ester transfer protein; co-AD = co-dominant; HDL = high-density lipoprotein; LCAT = lecithin-cholesterol acyltransferase; LDL = low-density lipoprotein; Lp(a) = lipoprotein(a); SR-BI = scavenger receptor class B type I.

The more frequent polygenic contributor are common genetic variants that have high population frequencies but individually modest influences on lipid traits. In aggregate, multiple common variants with smaller effects can together increase susceptibility towards a dyslipidemic state; this accumulation can be measured using polygenic scores, described in [Section 1.3.4.3](#). Early GWASs from the Global Lipids Genetics Consortium (GLGC) were some of the first to identify common variants governing plasma lipids and lipoproteins in essentially normolipidemic populations (Teslovich et al., 2010; Willer et al., 2013). The 157 loci identified by the GLGC explain 10-20% of the total variation in total cholesterol, LDL cholesterol, HDL cholesterol, and triglyceride levels (Willer et al., 2013). Over the last decade, >260 loci associated with blood lipid traits have been discovered using genetic association studies (Albrechtsen et al., 2013; Asselbergs et al., 2012; Below et al., 2016; Chasman et al., 2009; Liu et al., 2017; Lu et al., 2017; Peloso et al., 2014; Teslovich et al., 2010; Willer et al., 2013). A recent meta-GWAS conducted in over 600,000 participants between the Million Veteran Program and GLGC cohorts revealed an additional 118 novel loci associated with these traits (Klarin et al., 2018). In addition to GWASs, exome-wide association studies have also successfully uncovered genetic variants with small effects on blood lipid traits (Liu et al., 2017; Lu et al., 2017).

### 1.5.2 Abnormalities in high-density lipoprotein cholesterol levels

HDL cholesterol levels are normally distributed in the general population (**Figure 1.13**) (Sachdeva et al., 2009). Extreme deviations in HDL cholesterol levels are often caused by genetic determinants, while the typical variation observed for this phenotype can be due to a combination of different genetic factors.

Generally, extremely low and high levels of HDL cholesterol are diagnosed as hypoalphalipoproteinemia and hyperalphalipoproteinemia, respectively. Defining thresholds for these phenotypes are dependent on age, sex and race. A typical threshold for low HDL cholesterol levels in men and women are <1 mmol/L and <1.3 mmol/L, respectively (Schaefer et al., 2016); an extreme deficiency in HDL cholesterol is considered as levels <0.5 mmol/L (Schaefer et al., 2016). With respect to extremely high HDL cholesterol levels, levels above the 5<sup>th</sup> percentile based on age and sex are often accepted.



**Figure 1.13 The genetic architecture underlying the spectrum of measurable HDL cholesterol levels.**

The distribution of HDL cholesterol levels has a normal distribution in the general population; however, it is important to note that this distribution includes both males and females, which have different thresholds for what is considered “low” and “high”. The thresholds shown in this figure are not exact and are for illustrative purposes only. Studies tend to focus on individuals with extreme HDL cholesterol levels, falling in the tail-ends of the distribution, to better understand the genetic determinants driving these phenotypes. Abbreviations: HDL = high-density lipoprotein.

### 1.5.2.1 Hypoalphalipoproteinemia

Extremely low levels of HDL cholesterol are suggestive of metabolic issues related to the inability to synthesize HDL particles. There are a number of monogenic syndromes for hypoalphalipoproteinemia that are defined by the causative mutated gene: *ABCA1*, *APOA1*, and *LCAT*.

Tangier disease is an autosomal recessive disorder caused by rare bi-allelic variants in *ABCA1* (Schaefer et al., 2010). With substantial disruptions to both copies of the *ABCA1* gene, the first stage in the development of HDL particles through lipidation of apo A-I cannot occur. In addition to having extremely low, virtually absent HDL cholesterol levels because of this functional deficit of ABCA1, individuals with Tangier disease also have moderately elevated triglyceride levels, reduced LDL cholesterol levels, and can present with both hepatosplenomegaly and enlarged, lipid-laden tonsils (Fredrickson et al., 1961). As well, manifestations of the disease can include peripheral neuropathy, corneal opacities, and an increased risk for CVD (Bale et al., 1971; Engel et al., 1967).

Another autosomal recessive disorder with an extremely low HDL cholesterol level profile is apo A-I deficiency, caused by rare bi-allelic variants in *APOA1* (Schaefer et al., 2010). In the absence of apo A-I particles—due to either a decrease in expression or dysfunctional forms of the proteins—HDL particles cannot be synthesized, as there is no protein available for lipidation of free cholesterol exported out of cells via ABCA1. Beyond undetectable levels of apo A-I and severely decreased HDL cholesterol levels, a collection of clinical manifestations have been observed in patients, including xanthomas, cerebellar ataxia, corneal arcus and opacification, and premature CVD (Matsunaga et al., 1991; Ng et al., 1994; Santos et al., 2008c).

Familial LCAT deficiency (FLD) and fish-eye disease (FED) are two additional autosomal recessive disorders caused by rare bi-allelic variants in *LCAT* (Schaefer et al., 2016; Schaefer et al., 2010). The disorders differ depending on which lipoproteins are impacted by the dysfunctional LCAT activity: (i) FLD encompasses issues impacting HDL and apo B-containing lipoproteins, and (ii) FED encompasses issues impacting only

HDL (Schaefer et al., 2016). Given the shared molecular disease etiology, there are a number of overlapping clinical features between FLD and FED in addition to decreased HDL cholesterol levels, including corneal opacification, elevated triglyceride and LDL levels, and risk for CVD later in life (Gjone et al., 1974; Norum and Gjone, 1967; Santos et al., 2008a; Schaefer et al., 2016; Schaefer et al., 2010). Since FLD impacts more lipoproteins, additional clinical features have been observed, including anemia and proteinuria (Norum and Gjone, 1967).

Beyond the aforementioned syndromes, in non-monogenic instances of hypoalphalipoproteinemia, an increased prevalence of heterozygous rare variants in *ABCA1*, *APOA1*, and *LCAT* has been observed (Candini et al., 2010; Cohen et al., 2004; Holleboom et al., 2011; Kiss et al., 2007; Motazacker et al., 2013; Sadananda et al., 2015; Santos et al., 2008a; Santos et al., 2008b; Santos et al., 2008c; Schaefer et al., 2016; Singaraja et al., 2013; Tietjen et al., 2012; Wada et al., 2009). Damaging variants disrupting these genes may impact the synthesis and modification of HDL particles that lead to an overall lower circulating HDL cholesterol concentration. While these variants are not deterministic and are not guaranteed to cause hypoalphalipoproteinemia, they are instead probabilistic and increase an individual's susceptibility towards decreased concentrations of HDL cholesterol.

In addition to heterozygous rare variants, the polygenic aggregation of common SNPs associated with HDL cholesterol levels can modulate further an individual's susceptibility towards the hypoalphalipoproteinemia phenotype; dozens of SNPs have shown significant associations to HDL cholesterol levels across many GWASs (Albrechtsen et al., 2013; Asselbergs et al., 2012; Below et al., 2016; Chasman et al., 2009; Liu et al., 2017; Lu et al., 2017; Peloso et al., 2014; Teslovich et al., 2010; Willer et al., 2013). While each individual SNP may have only a small overall phenotypic impact, the aggregated effects from multiple SNP alleles associated with lower HDL cholesterol levels may substantially alter the HDL cholesterol phenotype; this aggregation can be quantified using a polygenic risk score (Aulchenko et al., 2009; Buscot et al., 2016; Justesen et al., 2015; Latsuzbaia et al., 2016; Lutsey et al., 2012;

Paquette et al., 2017; Piccolo et al., 2009; Raffield et al., 2013; Teslovich et al., 2010; Tikkanen et al., 2011; Zubair et al., 2014).

### 1.5.2.2 Hyperalphalipoproteinemia

Extremely elevated HDL cholesterol levels can be caused by dysfunctional proteins in the HDL metabolic pathway due to genetic variation in *CETP*, *SCARB1*, *LIPC*, and *LIPG*.

CETP deficiency was first observed in Japanese kindreds and was found to be due to loss-of-function variants in *CETP* (Brown et al., 1989; Inazu et al., 1990; Yamashita et al., 1988). Normally, CETP facilitates the exchange of cholesteryl esters for triglycerides between HDL particles and apo B-containing lipoproteins (Tosheska-Trajkovska and Topuzovska, 2017). When this process is hindered, HDL particles retain their cholesterol content and the overall concentration for HDL cholesterol begins to increase; this mechanism was pharmacologically mimicked using CETP inhibitors—small molecules that prevented the normal functioning of CETP (Tall and Rader, 2018). While both longevity and a reduction in CVD risk are phenotypic outcomes that have been associated with CETP deficiency (Milman et al., 2014), there are also reports of individuals being at an increased CVD risk despite elevations in HDL cholesterol levels (Hirano et al., 1995; Hirano et al., 2014).

Disruptions to *SCARB1*, another HDL-associated gene, have similar outcomes to what has been described for CETP deficiency. As a receptor for HDL, reductions in either the expression or activity of SR-BI decreases hepatic uptake of HDL-associated cholesteryl esters, which results in the increased plasma concentration of HDL cholesterol (Hoekstra et al., 2010). In mice, an overexpression of SR-BI leads to a decrease in HDL cholesterol levels (Ji et al., 1999; Kozarsky et al., 1997; Ueda et al., 1999; Wang et al., 1998), while deletions of the gene cause increased levels (Brundert et al., 2005; Varban et al., 1998). Importantly, a human individual with extremely high levels of HDL cholesterol was found to carry a rare, homozygous missense variant in *SCARB1* (Zanoni et al., 2016). In both the knockout mice and homozygous human carrier, atherosclerotic plaque progression was observed, despite the elevated HDL cholesterol profile.

*LIPC*, encoding HL, is also closely related to HDL cholesterol levels. SNPs both in and around the *LIPC* locus have been associated with elevations in HDL cholesterol levels and decreases in HL activity (Guerra et al., 1997; Hodoglugil et al., 2010; McCaskie et al., 2006; Zambon et al., 1998). A decrease in HL function results in a decrease in catabolism of HDL particles through reductions in triglyceride hydrolysis and phospholipid lipolysis (Feitosa et al., 2009); this decrease in function results in elevations of HDL particles and by association, HDL cholesterol levels. Similarly, EL encoded by *LIPG* is another lipase in which a reduction in its activity leads to elevated HDL cholesterol levels. Many genetic variants in *LIPG* have shown strong associations with HDL cholesterol levels (deLemos et al., 2002; Edmondson et al., 2009; Tietjen et al., 2012), and a functional analysis of loss-of-function *LIPG* variants showed that the resultant decrease in EL activity contributed towards the overall elevation in HDL cholesterol levels (Singaraja et al., 2013).

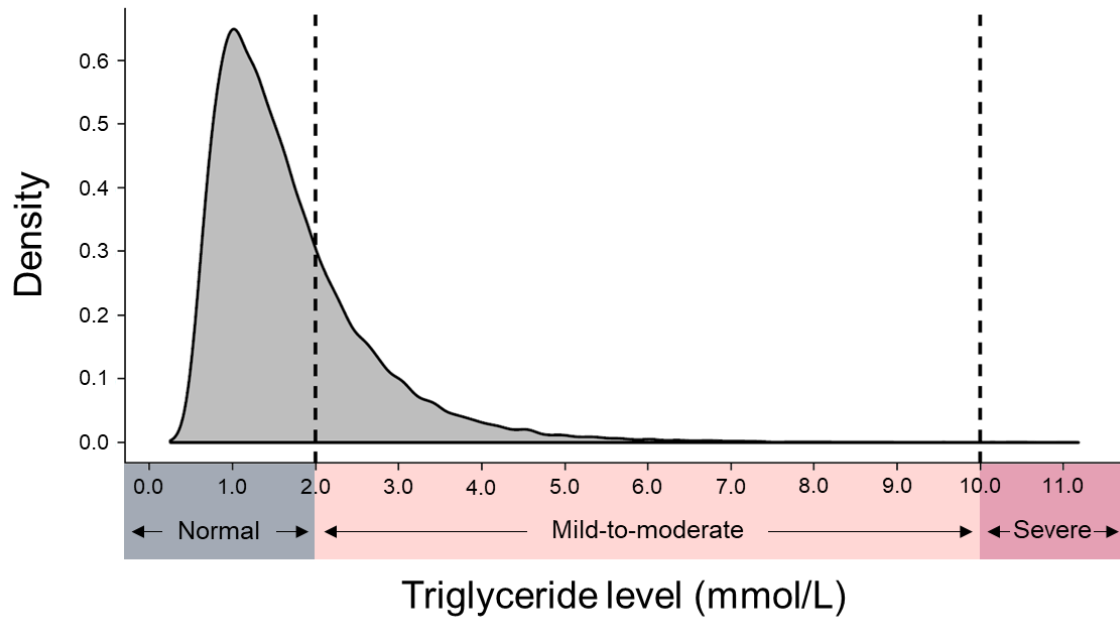
Like hypoalphalipoproteinemia, hyperalphalipoproteinemia is largely influenced by variants disrupting canonical HDL metabolism genes, both rare and common. Elsewhere in the genome, SNPs identified through GWASs have also been associated with small elevations in levels of HDL cholesterol (Albrechtsen et al., 2013; Asselbergs et al., 2012; Below et al., 2016; Chasman et al., 2009; Liu et al., 2017; Lu et al., 2017; Peloso et al., 2014; Teslovich et al., 2010; Willer et al., 2013). A polygenic basis for hyperalphalipoproteinemia has been reported (Motazacker et al., 2013), which aligns with the genetic architecture described for the opposing HDL cholesterol phenotype. Despite being on opposite ends of the phenotypic spectrum, as extremes of the same trait, it is unsurprising that similar types of genetic determinants appear to underlie both hyperalphalipoproteinemia and hypoalphalipoproteinemia.

### 1.5.3 Abnormalities in triglyceride levels

Fasting triglyceride levels follow a right-skewed distribution in the general population (**Figure 1.14**). While both environmental and genetic factors can influence triglyceride levels, the more extreme phenotypes primarily have a genetic basis, ranging from rare to common variants with varying phenotypic impacts (Hegele et al., 2009; Johansen et al., 2010; Wang et al., 2008a; Wang et al., 2008b; Wang et al., 2007). The precise assortment

of variants differ among individuals; those with a greater quantitative and qualitative burden of triglyceride-raising variants are assumed to be predisposed to more severe pathological triglyceride elevations. Conversely, individuals with an extreme absence of these triglyceride-raising variants instead possess an extreme burden of triglyceride-lowering variants and are more likely to present with very low triglyceride concentrations.





**Figure 1.14 The genetic architecture underlying the spectrum of measurable triglyceride levels.**

The distribution of triglyceride levels has a positive skew in the general population. Normal levels of triglyceride are considered to be less than 2.0 mmol/L. Individuals with triglyceride levels between 2.0 to 9.9 mmol/L are diagnosed with mild-to-moderate hypertriglyceridemia, while individuals with triglyceride levels above 10.0 mmol/L are diagnosed with severe hypertriglyceridemia. Studies tend to focus on individuals with hypertriglyceridemia to better understand the genetic determinants driving this extreme phenotype.

### 1.5.3.1 Hypertriglyceridemia

A clinical diagnosis of “hypertriglyceridemia” is usually made by applying threshold values to the distribution of plasma triglyceride levels (Hegele et al., 2014). Different consensus committees recommend various threshold values for such discrete classifications as mild-to-moderate and severe hypertriglyceridemia (Hegele et al., 2014). From the Canadian Heart Health Surveys, the mean overall triglyceride level in adults is 1.6 mmol/L (Connelly et al., 1999). A level of 2.0 mmol/L represents about the 75<sup>th</sup> percentile, while a level of 3.3 mmol/L represents the top 95<sup>th</sup> percentile (Connelly et al., 1999; Johansen et al., 2011a); however, these thresholds can vary between geographic areas and jurisdictions, and can also be dependent on age, sex, and race.

#### 1.5.3.1.1 Severe hypertriglyceridemia

Severe hypertriglyceridemia is defined as total plasma triglyceride  $\geq 10$  mmol/L (885 mg/dL). Such an extreme elevation in triglyceride levels typically signals the persistence of large intestinally-derived chylomicrons, particularly in the fasting state, when these particles otherwise should have been physiologically cleared (Lewis et al., 2015). Chylomicrons, with their high ratio of volume-to-surface area, present the most direct physical mechanism to achieve severe hypertriglyceridemia (Brahm and Hegele, 2013; Brahm and Hegele, 2015); as such, severe hypertriglyceridemia is often referred to as “chylomicronemia” to better describe this abnormal lipoprotein content. Elevations of the remaining classes of smaller triglyceride-rich lipoproteins can further augment the phenotype and may have larger roles in potential risk for CVD, discussed in the upcoming [Section 1.5.3.1.3](#).

##### 1.5.3.1.1.1 Familial Chylomicronemia Syndrome

Familial chylomicronemia syndrome (FCS) is the monogenic form of severe hypertriglyceridemia and follows a classic autosomal recessive inheritance pattern (Brahm and Hegele, 2015). FCS is extremely rare in the population, with a prevalence of 1 in 100,000-1,000,000 individuals (Brahm and Hegele, 2013; Brahm and Hegele, 2015; Gotoda et al., 2012). Clinical diagnosis can occur between infancy and early adulthood

(Brahm and Hegele, 2013; Brahm and Hegele, 2015). An accumulation of triglyceride-rich chylomicrons starting at birth may lead to manifestation of clinical features including failure to thrive, lipemia retinalis, recurrent abdominal pain, nausea, vomiting, hepatosplenomegaly, and eruptive xanthomas on the trunk, extremities and buttocks (Feoli-Fonseca et al., 1998; Rahalkar and Hegele, 2008). Of these, the most serious complication is the increased risk of acute pancreatitis and its 5-6% associated mortality rate (Brahm and Hegele, 2015). A number of less common features may also appear and include anemia, diarrhea, intestinal bleeding, irritability, seizures, and encephalopathy (Feoli-Fonseca et al., 1998; Rahalkar and Hegele, 2008).

As an autosomal recessive disease, the molecular basis underlying FCS involves the presence of rare, bi-allelic variants in the canonical triglyceride metabolism genes that exert large, disruptive effects on triglyceride hydrolysis. Specifically, variants that compromise the regulation or function of the LPL enzyme and impede the breakdown of chylomicrons, leading to extreme deviations of triglyceride levels from normal (Chokshi et al., 2014). The most common form of FCS—making up 95% or more of cases—results from bi-allelic variants within the *LPL* gene itself. Monogenic disruptions of related genes encoding factors that interact with LPL including *LMF1*, *GPIHBP1*, *APOC2*, and *APOA5*, are much less frequent than bi-allelic *LPL* variants and affect a total of <100 reported families worldwide (Brahm and Hegele, 2015).

#### 1.5.3.1.1.2 Multifactorial Chylomicronemia

In contrast to FCS, multifactorial chylomicronemia (MCM) is much more common, complicated and nuanced due to its polygenic nature (Chait and Eckel, 2019). Based on the reported Canadian prevalence of adults with severe hypertriglyceridemia, the estimated population prevalence of MCM is roughly 1 in 600-1,000 individuals (Dron and Hegele, 2020; Johansen et al., 2011a). Because lipolysis activity is only partially compromised by polygenic determinants, MCM encompasses a much broader population of elevated triglyceride-rich lipoprotein and remnant species than FCS, including chylomicrons, VLDL, IDL, and remnant particles. These mechanistic discrepancies are also reflected in different clinical features between monogenic FCS and polygenic MCM. For instance, onset of polygenic chylomicronemia typically begins in adulthood, and

while patients are likely to experience recurrent abdominal pain, nausea, and vomiting as in the monogenic form, they are less likely to present lipemia retinalis, and eruptive xanthomas (Brahm and Hegele, 2015; Chait and Eckel, 2019). In addition, while the absolute risk of acute pancreatitis is high, it occurs less commonly in MCM than in FCS; some estimates are ~10-20% over a lifetime, while rates in FCS have been estimated at ~60-80% (Baass et al., 2020; Gotoda et al., 2012). Not surprisingly, the differences in these clinical manifestations and their underlying molecular mechanisms are attributable to the complex nature of polygenic inheritance, as there is a wider range of potential permutations of genetic factors.

MCM is polygenic in nature, and unlike FCS, relevant genetic factors are probabilistic in that they increase the risk of developing MCM, but do not guarantee its clinical expression.

Rare loss-of-function variants in *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2* are important genetic contributors to MCM; however, many heterozygotes for such dysfunctional variants have normal lipid profiles (Johansen et al., 2011b; Surendran et al., 2012); a secondary factor is required to drive expression of the severe phenotype. Not only are patients with MCM more likely to carry disruptive heterozygous variants in these canonical genes, they are also more likely to carry rare variants in non-canonical genes involved in triglyceride metabolism (Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012). For instance, *CREB3L3* encoding the transcription factor cyclic AMP-responsive element-binding protein H (CREBH), is an example of a gene that impacts triglyceride levels and was discovered through the use of animal models (Lee et al., 2011). In addition, *GCKR* encoding glucokinase regulatory protein, is an example of a gene that harbors rare large-effect determinants of human triglyceride levels that was initially identified as a common locus for triglyceride levels through GWAS (Rees et al., 2014).

In addition to the accumulation of heterozygous rare variants within triglyceride-related genes, another defining genetic feature of MCM is the increased burden of triglyceride-associated SNPs (Teslovich et al., 2010; Wang et al., 2008b). Many GWASs have

successfully identified these common variants and their small phenotypic effects on triglyceride levels (Albrechtsen et al., 2013; Asselbergs et al., 2012; Below et al., 2016; Chasman et al., 2009; Liu et al., 2017; Lu et al., 2017; Peloso et al., 2014; Teslovich et al., 2010; Willer et al., 2013). Several of these SNPs are within loci already known to be involved in triglyceride metabolism, including *LPL* and *APOA5* (Kuivenhoven and Hegele, 2014). Others were found in close proximity to genes that at the time were not relevant, but were found to be in subsequent studies (i.e. *GCKR*); and many SNPs identified were intergenic and may be important in regulatory processes. When considering triglyceride-associated SNPs in MCM patients compared to normolipidemic individuals, a distinct increase in SNP accumulation in these patients has been observed (Johansen et al., 2010; Johansen et al., 2011b). Individually, each SNP has a slight influence on triglyceride levels; however, when a substantial burden of multiple small-effect variants is present in an individual, it can synergistically contribute towards an overall large phenotypic effect.

The contributory effects coming from rare heterozygous variants with larger phenotypic influences and the excessive accumulation of common variants scattered throughout the genome, all work in concert to produce polygenic MCM due to perturbations of chylomicrons, as well as other triglyceride-rich lipoproteins.

#### 1.5.3.1.2 Mild-to-moderate hypertriglyceridemia

Mild-to-moderate hypertriglyceridemia is defined as total plasma triglyceride between 2.0 and 9.9 mmol/L (Hegele et al., 2014) and most often results from elevations of liver-derived, triglyceride-rich lipoprotein species such as VLDL and their remnants, rather than chylomicrons. Thus, factors related to biosynthesis, secretion and catabolism of VLDL would be relatively more important in susceptibility for mild-to-moderate hypertriglyceridemia. In contrast, factors related to biosynthesis, secretion and catabolism of chylomicrons are relatively more important in susceptibility to severe hypertriglyceridemia, although there is considerable overlap with factors that modulate VLDL levels, particularly on the catabolic side (Dron and Hegele, 2016).

Unsurprisingly, a similar general architecture of genetic susceptibility is seen in patients with mild-to-moderate hypertriglyceridemia and in patients with polygenic MCM (Dron and Hegele, 2016; Hegele et al., 2014; Johansen et al., 2011b). This includes: (i) higher odds of carrying a heterozygous rare variant in one of the five canonical metabolism genes (Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Surendran et al., 2012); and (ii) an increased cumulative burden of small-effect SNPs (Johansen et al., 2011b). The relative burden of these factors together with secondary non-genetic factors may determine the severity of the phenotype between individuals (Dron and Hegele, 2016; Hegele et al., 2014).

#### 1.5.3.1.3 Hypertriglyceridemia and risk for cardiovascular disease

The relationship between hypertriglyceridemia and CVD stems from the disturbed lipoprotein fractions in the individual. Mild-to-moderate hypertriglyceridemia is associated with a higher risk for CVD because the predominantly disturbed lipoproteins are VLDL and IDL (Brahm and Hegele, 2015). These triglyceride-rich lipoprotein species of hepatic origin are atherogenic in nature due to their increased cholesterol content compared to chylomicrons, as this cholesterol can contribute towards the build-up of atherosclerotic lesions (Varbo and Nordestgaard, 2016). This is a modernization of the seminal Zilversmit hypothesis, an early articulation of the atherogenic role of triglyceride-rich lipoproteins (Zilversmit, 1995): according to this model, triglyceride-rich lipoproteins are metabolically independent of LDL cholesterol in atherogenesis, and act additively to further increase CVD risk.

From the distribution of triglyceride levels within the population, most patients with hypertriglyceridemia fall within the mild-to-moderate range, and thus any potential atherosclerosis risk is tied to elevations in VLDL and IDL particles. At higher strata of triglyceride levels, chylomicrons and their remnants begin to predominate. In this important but much less prevalent subgroup, it has been more or less axiomatic that chylomicrons are too large to penetrate the arterial wall (Chait and Brunzell, 1992; Lewis et al., 2015). However, chylomicron remnants, especially on the smaller end of the spectrum, may contribute to atherogenesis since they are smaller in size and may be able

to directly contribute their cholesterol towards atherosclerotic lesions after passing through the arterial wall (Lewis et al., 2015).

Thus, among the diverse range of patients with hypertriglyceridemia, those with monogenic impairment of triglyceride hydrolysis would primarily have chylomicronemia because the deficiency in LPL activity prevents their catabolism, and would be at a relatively low risk for atherosclerosis. In contrast, among individuals with the same degree of triglyceride elevation due to varied polygenic plus secondary factors, the spectrum of triglyceride-rich particles is much more diffuse, and includes many remnant particles, since lipolysis is not completely impaired (Johansen and Hegele, 2012). Here one could postulate that atherosclerosis risk is increased, due to the relative abundance of atherosclerosis-related remnants.

### 1.5.3.2 Hypotriglyceridemia

Hypotriglyceridemia is defined as very low or absent triglyceride levels. As with hypertriglyceridemia, genetic determinants of hypotriglyceridemia include ultra-rare monogenic syndromic disorders that are associated with a range of other lipoprotein, biochemical and clinical abnormalities, such as abetalipoproteinemia and homozygous hypobetalipoproteinemia, which result, respectively, from bi-allelic variants in *MTTP* and *APOB* (Hegele, 2009). Importantly, heterozygotes for *MTTP* loss-of-function variants have no obvious clinical or biochemical phenotypes.

Non-syndromic forms of hypotriglyceridemia have been reportedly driven by deficiencies of apo C-III and *ANGPTL3*, which result from bi-allelic variants in *APOC3* and *ANGPTL3*, respectively (Musunuru et al., 2010; Pollin et al., 2008). Carriers for heterozygous loss-of-function variants in these genes, as well as in the *ANGPTL4* gene, have been reported to have significantly lower triglyceride levels and a decreased risk for CVD (Dewey et al., 2016; Jorgensen et al., 2014; Myocardial Infarction et al., 2016; Pisciotta et al., 2012; Romeo et al., 2007; Stitzel et al., 2017; Tg et al., 2014).

## 1.6 Genetic assessment of dyslipidemia

Historically, genetic researchers have availed themselves of the prevalent genetic technologies of any particular era. The low-hanging fruit—rare large-effect variants (i.e. highly penetrant, disease-causing mutations)—were first identified by studying kindreds containing individuals with clinically diagnosed dyslipidemia syndromes (Dron and Hegele, 2016; Hegele, 2009). In that era, Sanger sequencing was used to identify rare variants in candidate genes—sometimes highlighted through linkage analysis—driving monogenic forms of dyslipidemia (Dron and Hegele, 2016). The effects of these variants are so strong and highly penetrant that they faithfully co-segregate with disease phenotypes across generations; their pathogenicity can be inferred by studying only a few individuals or families.

Over the last decade, the focus has shifted towards studying the influence of SNP genotypes on inter-individual variation of lipid traits in the general population. In contrast to rare large-effect variants, the weak and inconsistent effects of common SNPs are difficult to ascertain in families. Their modest phenotypic effects underlie low phenotypic penetrance, with no obvious co-segregation across generations. The inconsistent association of SNPs with lipid traits in small samples was also a feature of candidate gene-association studies performed in the 1990's and early 2000's (Hegele, 2002). More recently, aggregation and meta-analyses of large cohorts coupled with cost-effective, microarray-based, high-throughput genotyping has enabled informative GWASs that have revolutionized our understanding of the small phenotypic effects imparted by SNPs (Christoffersen and Tybjaerg-Hansen, 2015).

GWASs have allowed researchers to uncover common variants dispersed across the genome—including intergenic and intronic regions—that are associated with small but consistent phenotypic effects in essentially normolipidemic individuals. To date, over 300 SNPs with subtle effects on lipid or lipoprotein traits have been described (Albrechtsen et al., 2013; Asselbergs et al., 2012; Below et al., 2016; Chasman et al., 2009; Klarin et al., 2018; Liu et al., 2017; Lu et al., 2017; Peloso et al., 2014; Teslovich et al., 2010; Willer et al., 2013). While many of the significantly associated loci were already well-known in



the field, the majority of loci uncovered by GWAS had no previous known connection to lipoprotein metabolism.

We are now well into the “post-GWAS” era, in which NGS technologies have become more accessible in both clinical and research settings. Researchers can explore rare variants in important genetic loci that arose from candidate gene studies and GWASs, and can characterize rare large-effect variants in genes not previously known to be related to lipid traits. The present genetic technological methods have brought the field to a point where assessing multiple types of genetic factors across virtually all areas of the genome is feasible.

### 1.6.1 LipidSeq: a targeted next-generation sequencing panel for dyslipidemia phenotypes

One of the main objectives of the Hegele Lab is to uncover and understand the genetic factors underlying the phenotypes of patients from the London Lipid Genetics Clinic. Because the clinical practice encompasses all dyslipidemias and many metabolic syndromes, a primary focus has been on disease ontology (Fu et al., 2013; Hegele, 2009; Rahalkar and Hegele, 2008) and on documenting dyslipidemia-associated variants (Fu et al., 2013). With this focused interest, a targeted NGS panel, called “LipidSeq”, was designed to aid in the genetic diagnosis and research of this set of diseases and associated genetic variants (Dron et al., 2020; Johansen et al., 2014).

Unlike whole-exome sequencing in which all genes are sequenced, or whole-genome sequencing in which the entire genome is sequenced, the LipidSeq panel was designed to target a specific subset of genes (**Figure 1.15**) underlying known dyslipidemias and other disorders for which dyslipidemia is a secondary manifestation, such as inherited forms of diabetes (Hegele, 2019; Johansen et al., 2014). With a high read-depth of coverage, sequencing data generated from LipidSeq has allowed for the ability to concurrently identify CNVs along with SNVs; previously, separate dedicated methods to identify CNVs were required, such as multiplex ligation-dependent probe amplification (MLPA) or microarrays (Iacocca and Hegele, 2017, 2018). Furthermore, because of our laboratory’s longstanding interest in the polygenic basis of plasma lipids (Hegele et al.,

1995; Johansen et al., 2011b; Wang et al., 2008b), the panel was designed to simultaneously genotype 185 SNP loci that were reported from early GWASs to be associated with lipid and lipoprotein levels (Kathiresan et al., 2009; Teslovich et al., 2010; Willer et al., 2013). This focused interest on a subset of SNPs negated the need for running full microarrays in conjunction with sequencing the phenotypically relevant genes of interest.

Thus, LipidSeq is a targeted NGS panel that can be used to simultaneously detect: (i) functionally relevant rare SNVs and CNVs in genes underlying monogenic dyslipidemias; and (ii) SNPs associated with lipid and lipoprotein levels that can be used to develop lipid-specific polygenic scores (Dron et al., 2020). This method allows for the comprehensive assessment of a range of genetic determinants relevant to dyslipidemia phenotypes. Until now, the assessment of genetic factors related to dyslipidemia were dependent on the technology used: rare variants could only be identified through gene-sequencing approaches like Sanger or whole-exome sequencing, while SNPs could only be assessed using microarrays or Sanger sequencing of SNP loci. Short of performing whole-genome sequencing—which was neither economically feasible for thousands of patient samples, nor computationally feasible for large-scale bioinformatically analysis—there were no effective methodological options for the in-depth genetic assessments of dyslipidemia cohorts prior to the development of LipidSeq.

<p><b><u>HDL cholesterol</u></b></p> <p><b>ABCA1</b>      <b>LIPG</b>  <b>ABCG1</b>      <b>LIPC</b>  <b>APOA1</b>      <b>SCARB1</b>  <b>CETP</b>        <b>PLTP</b>  <b>LCAT</b></p>	<p><b><u>Triglyceride</u></b></p> <p><b>ANGPTL3*</b>    <b>GCKR</b>  <b>APOA4</b>        <b>GPD1</b>  <b>APOA5</b>        <b>GPIHBP1</b>  <b>APOC2</b>        <b>LMF1</b>  <b>APOC3</b>        <b>LPL</b>  <b>APOE*</b>        <b>MLXIPL</b>  <b>CREB3L3</b>      <b>PPARA</b>  <b>GALNT2</b>        <b>TRIB1</b></p>	<p><b><u>LDL cholesterol</u></b></p> <p><b>ABCG5</b>        <b>MTTP</b>  <b>ABCG8</b>        <b>MYLIP</b>  <b>ANGPTL3*</b>     <b>NPC1L1</b>  <b>APOB</b>         <b>PCSK9</b>  <b>APOE*</b>        <b>SAR1B</b>  <b>LDLR</b>         <b>SORT1</b>  <b>LDLRAP1</b>     <b>STAP1</b>  <b>LIPA</b></p>
<p><b><u>Lipodystrophy</u></b></p> <p><b>AGPAT2</b>      <b>LMNB2</b>  <b>AKT2</b>        <b>LPIN1</b>  <b>BSCL2</b>      <b>PLIN1</b>  <b>CAV1</b>        <b>POLD1</b>  <b>CIDEA</b>       <b>PPARG</b>  <b>DYRK1B</b>      <b>PTRF</b>  <b>LIPE</b>        <b>WRN</b>  <b>LMNA</b>        <b>ZMPSTE24</b></p>	<p><b><u>Diabetes</u></b></p> <p><b>ABCC8</b>        <b>INS</b>  <b>BLK</b>          <b>KCNJ11</b>  <b>CEL</b>          <b>KLF11</b>  <b>GCK</b>          <b>NEUROD1</b>  <b>HNF1A</b>        <b>PAX4</b>  <b>HNF1B</b>        <b>PDX1</b>  <b>HNF4A</b></p>	<p><b><u>Miscellaneous</u></b></p> <p><b>CAV2</b>        <b>PNPLA2</b></p>

**Figure 1.15 Genes targeted by the LipidSeq panel.**

The 69 genes that are targeted by LipidSeq panel, grouped by their associated lipid or metabolic phenotype. Bolded genes were included in **Table 1.1**, as they have causal or statistical associations with different named dyslipidemias. "\*" denotes genes that appear in multiple lipid categories. Abbreviations: HDL = high-density lipoprotein; LDL = low-density lipoprotein.

## 1.7 Thesis outline

### 1.7.1 Overall research aim and objectives

To date, independent studies have examined the genetic determinants underlying different dyslipidemia phenotypes—hypoalphalipoproteinemia, hyperalphalipoproteinemia, and hypertriglyceridemia. However, despite appreciation for the range of genetic variation that influence phenotypic susceptibility, the comprehensive genetic profile for each phenotype has not been objectively or rigorously quantified. It stands to reason that a more detailed characterization of multiple genetic determinants—rare SNVs, CNVs, and common SNPs—related to each dyslipidemia of interest will help improve general academic knowledge of the full range of genetic factors driving these phenotypes. With this information, concerted efforts can be made to establish methods to better determine genetic risk for each dyslipidemia, with possible downstream applications related to mitigating associated health risks like CVD.

The aim of my PhD research was to robustly characterize the genetic determinants of hypoalphalipoproteinemia, hyperalphalipoproteinemia, and hypertriglyceridemia using sequencing data generated from the targeted NGS panel, LipidSeq.

My first objective was focused on the phenotypic extremes of HDL cholesterol levels and assessing the prevalence of rare SNVs and extreme accumulation of SNPs in hypoalphalipoproteinemia and hyperalphalipoproteinemia patients compared to normolipidemic individuals. Rare SNVs were screened for following a candidate gene approach for each phenotype, and I developed a novel polygenic risk score to quantify the accumulation of HDL cholesterol-associated SNPs. The details of these efforts are provided in [Chapter 2](#). After the release of a novel CNV detection algorithm, I subsequently screened the study subjects from [Chapter 2](#) for CNVs disrupting the same candidate genes of interest. The details of this effort are provided in [Chapter 3](#).

My second objective was focused on elevations in triglyceride levels and assessing the prevalence of rare SNVs, CNVs, and the extreme accumulation of SNPs in severe hypertriglyceridemia and mild-to-moderate hypertriglyceridemia patients, compared to

normolipidemic individuals. Rare variants were screened for following a candidate gene approach, and I developed a novel polygenic risk score to quantify the accumulation of triglyceride-associated SNPs. The details of these efforts for severe hypertriglyceridemia are provided in [Chapters 4 and 5](#), while the details of these efforts for mild-to-moderate hypertriglyceridemia are provided in [Chapter 6](#).

My final objective was to employ a custom-designed bioinformatic pipeline (**Appendix L**) to perform a gene-based RVAS in an attempt to identify rare variants in non-candidate (i.e. “non-canonical”) genes that might be further contributing towards susceptibility towards extreme elevations in triglyceride levels, namely, severe hypertriglyceridemia. The details of these efforts are provided in [Chapter 7](#).

### 1.7.2 Hypothesis

Extreme levels of circulating lipids, both HDL cholesterol and triglycerides, have distinctive and genetically diverse architectures made up of discrete combinations of rare SNVs and CNVs with larger phenotypic impacts and common SNPs with smaller phenotypic effects, that cumulatively contribute towards polygenic susceptibility for (i) hypoalphalipoproteinemia; (ii) hyperalphalipoproteinemia; or (iii) hypertriglyceridemia.

### 1.7.3 Summary

This Dissertation details my research related to uncovering and understanding the comprehensive genetic profile of patients with either: (i) hypoalphalipoproteinemia; (ii) hyperalphalipoproteinemia; or (iii) hypertriglyceridemia. To achieve this, I utilized the LipidSeq targeted NGS panel to capture genetic variation—ranging from rare SNVs and CNVs to common SNPs—across metabolically relevant genetic loci in over 3,000 patient and control samples. Collectively, this work has furthered our understanding of the genetic nature of the aforementioned phenotypes of interest. Importantly, my work has highlighted a prominent polygenic underpinning for each dyslipidemia phenotype, demonstrating the importance of considering common genetic variants—despite having smaller phenotypic effects—in conjunction with heterozygous rare, large-effect variants for an improved understanding towards genetic factors contributing towards the susceptibility for extremes of either HDL cholesterol or triglyceride levels.

## 1.8 References

- Abbott, S., and Fairbanks, D.J. (2016). Experiments on Plant Hybrids by Gregor Mendel. *Genetics* 204, 407-422.
- Albers, J.J., Vuletic, S., and Cheung, M.C. (2012). Role of plasma phospholipid transfer protein in lipid and lipoprotein metabolism. *Biochim Biophys Acta* 1821, 345-357.
- Albrechtsen, A., Grarup, N., Li, Y., Sparso, T., Tian, G., Cao, H., Jiang, T., Kim, S.Y., Korneliussen, T., Li, Q., *et al.* (2013). Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 56, 298-310.
- Allara, E., Morani, G., Carter, P., Gkatzionis, A., Zuber, V., Foley, C.N., Rees, J.M.B., Mason, A.M., Bell, S., Gill, D., *et al.* (2019). Genetic Determinants of Lipids and Cardiovascular Disease Outcomes: A Wide-Angled Mendelian Randomization Investigation. *Circ Genom Precis Med* 12, e002711.
- Alves-Bezerra, M., and Cohen, D.E. (2017). Triglyceride Metabolism in the Liver. *Compr Physiol* 8, 1-8.
- Anderson, T.J., Gregoire, J., Pearson, G.J., Barry, A.R., Couture, P., Dawes, M., Francis, G.A., Genest, J., Jr., Grover, S., Gupta, M., *et al.* (2016). 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult. *Can J Cardiol* 32, 1263-1282.
- Asselbergs, F.W., Guo, Y., van Iperen, E.P., Sivapalaratnam, S., Tragante, V., Lanktree, M.B., Lange, L.A., Almoguera, B., Appelman, Y.E., Barnard, J., *et al.* (2012). Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet* 91, 823-838.
- Assimes, T.L., and Roberts, R. (2016). Genetics: Implications for Prevention and Management of Coronary Artery Disease. *J Am Coll Cardiol* 68, 2797-2818.
- Auer, P.L., and Lettre, G. (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 7, 16.
- Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., *et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genetics* 41, 47-55.
- Baass, A., Paquette, M., Bernard, S., and Hegele, R.A. (2020). Familial chylomicronemia syndrome: an under-recognized cause of severe hypertriglyceridaemia. *J Intern Med* 287, 340-348.
- Bale, P.M., Clifton-Bligh, P., Benjamin, B.N., and Whyte, H.M. (1971). Pathology of Tangier disease. *J Clin Pathol* 24, 609-616.

Ballantyne, C.M., Weiss, R., Moccetti, T., Vogt, A., Eber, B., Sosef, F., Duffield, E., and Investigators, E.S. (2007). Efficacy and safety of rosuvastatin 40 mg alone or in combination with ezetimibe in patients at high risk of cardiovascular disease (results from the EXPLORER study). *Am J Cardiol* 99, 673-680.

Below, J.E., Parra, E.J., Gamazon, E.R., Torres, J., Krithika, S., Candille, S., Lu, Y., Manichakul, A., Peralta-Romero, J., Duan, Q., *et al.* (2016). Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Sci Rep* 6, 19429.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53-59.

Boullart, A.C., de Graaf, J., and Stalenhoef, A.F. (2012). Serum triglycerides and risk of cardiovascular disease. *Biochim Biophys Acta* 1821, 867-875.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177-1186.

Brahm, A.J., and Hegele, R.A. (2013). Hypertriglyceridemia. *Nutrients* 5, 981-1001.

Brahm, A.J., and Hegele, R.A. (2015). Chylomicronaemia--current diagnosis and future therapies. *Nat Rev Endocrinol* 11, 352-362.

Brahm, A.J., and Hegele, R.A. (2016). Combined hyperlipidemia: familial but not (usually) monogenic. *Curr Opin Lipidol* 27, 131-140.

Breslow, J.L. (2000). Genetics of lipoprotein abnormalities associated with coronary artery disease susceptibility. *Annu Rev Genet* 34, 233-254.

Brown, M.L., Inazu, A., Hesler, C.B., Agellon, L.B., Mann, C., Whitlock, M.E., Marcel, Y.L., Milne, R.W., Koizumi, J., Mabuchi, H., *et al.* (1989). Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins. *Nature* 342, 448-451.

Brundert, M., Ewert, A., Heeren, J., Behrendt, B., Ramakrishnan, R., Greten, H., Merkel, M., and Rinninger, F. (2005). Scavenger receptor class B type I mediates the selective uptake of high-density lipoprotein-associated cholesteryl ester by the liver in mice. *Arterioscler Thromb Vasc Biol* 25, 143-148.

Bumgarner, R. (2013). Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol* Chapter 22, Unit 22 21.

Buscot, M.J., Magnussen, C.G., Juonala, M., Pitkanen, N., Lehtimäki, T., Viikari, J.S., Kahonen, M., Hutri-Kahonen, N., Schork, N.J., Raitakari, O.T., *et al.* (2016). The Combined Effect of Common Genetic Risk Variants on Circulating Lipoproteins Is

Evident in Childhood: A Longitudinal Analysis of the Cardiovascular Risk in Young Finns Study. *PLoS One* *11*, e0146081.

Candini, C., Schimmel, A.W., Peter, J., Bochem, A.E., Holleboom, A.G., Vergeer, M., Dullaart, R.P., Dallinga-Thie, G.M., Hovingh, G.K., Khoo, K.L., *et al.* (2010). Identification and characterization of novel loss of function mutations in ATP-binding cassette transporter A1 in patients with low plasma high-density lipoprotein cholesterol. *Atherosclerosis* *213*, 492-498.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* *3*, 285-298.

Castelli, W.P., Garrison, R.J., Wilson, P.W., Abbott, R.D., Kalousdian, S., and Kannel, W.B. (1986). Incidence of coronary heart disease and lipoprotein cholesterol levels. The Framingham Study. *JAMA* *256*, 2835-2838.

Castle, W.E. (1903). MENDEL'S LAW OF HEREDITY. *Science* *18*, 396-406.

Chait, A., and Brunzell, J.D. (1992). Chylomicronemia syndrome. *Adv Intern Med* *37*, 249-273.

Chait, A., and Eckel, R.H. (2016). Lipids, Lipoproteins, and Cardiovascular Disease: Clinical Pharmacology Now and in the Future. *J Clin Endocrinol Metab* *101*, 804-814.

Chait, A., and Eckel, R.H. (2019). The Chylomicronemia Syndrome Is Most Often Multifactorial: A Narrative Review of Causes and Treatment. *Ann Intern Med* *170*, 626-634.

Chasman, D.I., Pare, G., Mora, S., Hopewell, J.C., Peloso, G., Clarke, R., Cupples, L.A., Hamsten, A., Kathiresan, S., Malarstig, A., *et al.* (2009). Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet* *5*, e1000730.

Choi, S.W., Mak, T.S., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc.* *15*, 2759-2772.

Chokshi, N., Blumenschein, S.D., Ahmad, Z., and Garg, A. (2014). Genotype-phenotype relationships in patients with type I hyperlipoproteinemia. *J Clin Lipidol* *8*, 287-295.

Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., *et al.* (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet* *97*, 199-215.

Christoffersen, M., and Tybjaerg-Hansen, A. (2015). Novel genes in LDL metabolism--a comprehensive overview. *Curr Opin Lipidol* *26*, 179-187.



Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869-872.

Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.

Connelly, P.W., Petrasovits, A., Stachenko, S., MacLean, D.R., Little, J.A., and Chockalingam, A. (1999). Prevalence of high plasma triglyceride combined with low HDL-C levels and its association with smoking, hypertension, obesity, diabetes, sedentariness and LDL-C levels in the Canadian population. Canadian Heart Health Surveys Research Group. *Can J Cardiol* 15, 428-433.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., *et al.* (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-712.

Cordell, H.J., and Clayton, D.G. (2005). Genetic association studies. *Lancet* 366, 1121-1131.

Daniels, T.F., Killinger, K.M., Michal, J.J., Wright, R.W., Jr., and Jiang, Z. (2009). Lipoproteins, cholesterol homeostasis and cardiac health. *Int J Biol Sci* 5, 474-488.

deLemos, A.S., Wolfe, M.L., Long, C.J., Sivapackianathan, R., and Rader, D.J. (2002). Identification of genetic variants in endothelial lipase in persons with elevated high-density lipoprotein cholesterol. *Circulation* 106, 1321-1326.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169, 1915-1925.

Dewey, F.E., Gusarova, V., O'Dushlaine, C., Gottesman, O., Trejos, J., Hunt, C., Van Hout, C.V., Habegger, L., Buckler, D., Lai, K.M., *et al.* (2016). Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J Med* 374, 1123-1133.

Doolittle, M.H., Ehrhardt, N., and Peterfy, M. (2010). Lipase maturation factor 1: structure and role in lipase folding and assembly. *Curr Opin Lipidol* 21, 198-203.

Dron, J.S., and Hegele, R.A. (2016). Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130-141.

Dron, J.S., and Hegele, R.A. (2017a). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.

Dron, J.S., and Hegele, R.A. (2017b). Molecular Genetics of Hypertriglyceridaemia. In eLS, L.E. John Wiley & Sons, ed., pp. 1-8.

- Dron, J.S., and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.
- Dron, J.S., and Hegele, R.A. (2019). The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol* 30, 71-81.
- Dron, J.S., and Hegele, R.A. (2020). Genetics of Hypertriglyceridemia. *Front Endocrinol (Lausanne)* 11, 455.
- Dron, J.S., Wang, J., McIntyre, A.D., Iacocca, M.A., Robinson, J.F., Ban, M.R., Cao, H., and Hegele, R.A. (2020). Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Med Genomics* 13, 23.
- Edmondson, A.C., Brown, R.J., Kathiresan, S., Cupples, L.A., Demissie, S., Manning, A.K., Jensen, M.K., Rimm, E.B., Wang, J., Rodrigues, A., *et al.* (2009). Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest* 119, 1042-1050.
- Emerging Risk Factors Consortium, Di Angelantonio, E., Sarwar, N., Perry, P., Kaptoge, S., Ray, K.K., Thompson, A., Wood, A.M., Lewington, S., Sattar, N., *et al.* (2009). Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 302, 1993-2000.
- Engel, W.K., Dorman, J.D., Levy, R.I., and Fredrickson, D.S. (1967). Neuropathy in Tangier disease. Alpha-Lipoprotein deficiency manifesting as familial recurrent neuropathy and intestinal lipid storage. *Arch Neurol* 17, 1-9.
- Fadista, J., Manning, A.K., Florez, J.C., and Groop, L. (2016). The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* 24, 1202-1205.
- Feingold, K.R., and Grunfeld, C. (2000). Introduction to Lipids and Lipoproteins. In *Endotext*, A.B. Feingold KR, Boyce A, Chrousos G, de Herder WW, Dungan K, Grossman A, Hershman JM, Hofland HJ, Kaltsas G, Koch C, Kopp P, Korbonits M, McLachlan R, Morley JE, New M, Purnell J, Singer F, Stratakis CA, Trencle DL, Wilson DP, ed. (South Dartmouth (MA): MDText.com, Inc.).
- Feitosa, M.F., Myers, R.H., Pankow, J.S., Province, M.A., and Borecki, I.B. (2009). LIPC variants in the promoter and intron 1 modify HDL-C levels in a sex-specific fashion. *Atherosclerosis* 204, 171-177.
- Feoli-Fonseca, J.C., Levy, E., Godard, M., and Lambert, M. (1998). Familial lipoprotein lipase deficiency in infancy: clinical, biochemical, and molecular study. *J Pediatr* 133, 417-423.
- Ference, B.A., Ginsberg, H.N., Graham, I., Ray, K.K., Packard, C.J., Bruckert, E., Hegele, R.A., Krauss, R.M., Raal, F.J., Schunkert, H., *et al.* (2017). Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic,

- epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 38, 2459-2472.
- Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. *Nat Rev Genet* 7, 85-97.
- Fisher, E.A., Feig, J.E., Hewing, B., Hazen, S.L., and Smith, J.D. (2012). High-density lipoprotein function, dysfunction, and reverse cholesterol transport. *Arterioscler Thromb Vasc Biol* 32, 2813-2820.
- Forte, T.M., Sharma, V., and Ryan, R.O. (2016). Apolipoprotein A-V gene therapy for disease prevention / treatment: a critical analysis. *J Biomed Res* 30, 88-93.
- Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10, 241-251.
- Fredrickson, D.S., Altrocchi, P.H., Avioli, L.V., Goodman, D.S., and Goodman, H.C. (1961). Tangier Disease. *Ann Intern Med* 55, 1016-1031.
- Frikke-Schmidt, R., Nordestgaard, B.G., Stene, M.C., Sethi, A.A., Remaley, A.T., Schnohr, P., Grande, P., and Tybjaerg-Hansen, A. (2008). Association of loss-of-function mutations in the ABCA1 gene with high-density lipoprotein cholesterol levels and risk of ischemic heart disease. *JAMA* 299, 2524-2532.
- Fu, J., Kwok, S., Sinai, L., Abdel-Razek, O., Babula, J., Chen, D., Farago, E., Fernandopulle, N., Leith, S., Loyzer, M., *et al.* (2013). Western Database of Lipid Variants (WDLV): a catalogue of genetic variants in monogenic dyslipidemias. *Can J Cardiol* 29, 934-939.
- Genest, J., McPherson, R., Frohlich, J., Anderson, T., Campbell, N., Carpentier, A., Couture, P., Dufour, R., Fodor, G., Francis, G.A., *et al.* (2009). 2009 Canadian Cardiovascular Society/Canadian guidelines for the diagnosis and treatment of dyslipidemia and prevention of cardiovascular disease in the adult - 2009 recommendations. *Can J Cardiol* 25, 567-579.
- Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061-1073.
- Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- Gjone, E., Skarbovik, A.J., Blomhoff, J.P., and Teisberg, P. (1974). Familial lecithin:cholesterol acyltransferase deficiency. Report of a third Norwegian family with two afflicted members. *Scand J Clin Lab Invest Suppl* 137, 101-105.

- Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17, 333-351.
- Gordon, T., Castelli, W.P., Hjortland, M.C., Kannel, W.B., and Dawber, T.R. (1977). High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 62, 707-714.
- Gotoda, T., Shirai, K., Ohta, T., Kobayashi, J., Yokoyama, S., Oikawa, S., Bujo, H., Ishibashi, S., Arai, H., Yamashita, S., *et al.* (2012). Diagnosis and management of type I and type V hyperlipoproteinemia. *J Atheroscler Thromb* 19, 1-12.
- Grundy, S.M., Stone, N.J., Bailey, A.L., Beam, C., Birtcher, K.K., Blumenthal, R.S., Braun, L.T., de Ferranti, S., Faiella-Tommasino, J., Forman, D.E., *et al.* (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 73, e285-e350.
- Guerra, R., Wang, J., Grundy, S.M., and Cohen, J.C. (1997). A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci USA* 94, 4532-4537.
- Haase, C.L., Tybjaerg-Hansen, A., Qayyum, A.A., Schou, J., Nordestgaard, B.G., and Frikke-Schmidt, R. (2012). LCAT, HDL cholesterol and ischemic cardiovascular disease: a Mendelian randomization study of HDL cholesterol in 54,500 individuals. *J Clin Endocrinol Metab* 97, E248-256.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* 10, 551-564.
- Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1-8.
- Hegele, R.A. (2002). SNP judgments and freedom of association. *Arteriosclerosis, Thrombosis, and Vascular Biology* 22, 1058-1061.
- Hegele, R.A. (2009). Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* 10, 109-121.
- Hegele, R.A. (2016). Multidimensional regulation of lipoprotein lipase: impact on biochemical and cardiovascular phenotypes. *J Lipid Res* 57, 1601-1607.
- Hegele, R.A. (2019). Editorial: designing targeted sequencing panels for dyslipidemia. *Curr Opin Lipidol* 30, 53-55.
- Hegele, R.A., Ban, M.R., Cao, H., McIntyre, A.D., Robinson, J.F., and Wang, J. (2015). Targeted next-generation sequencing in monogenic dyslipidemias. *Curr Opin Lipidol* 26, 103-113.

- Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Hum Mol Genet* *18*, 4189-4194.
- Hegele, R.A., Brunt, J.H., and Connelly, P.W. (1995). Multiple genetic determinants of variation of plasma lipoproteins in Alberta Hutterites. *Arterioscler Thromb Vasc Biol* *15*, 861-871.
- Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Averna, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* *2*, 655-666.
- Hirano, K., Yamashita, S., Kuga, Y., Sakai, N., Nozaki, S., Kihara, S., Arai, T., Yanagi, K., Takami, S., Menju, M., *et al.* (1995). Atherosclerotic disease in marked hyperalphalipoproteinemia. Combined reduction of cholesteryl ester transfer protein and hepatic triglyceride lipase. *Arterioscler Thromb Vasc Biol* *15*, 1849-1856.
- Hirano, K.I., Nagasaka, H., Kobayashi, K., Yamaguchi, S., Suzuki, A., Toda, T., and Doyu, M. (2014). Disease-associated marked hyperalphalipoproteinemia. *Mol Genet Metab Rep* *1*, 264-268.
- Hodoglugil, U., Williamson, D.W., and Mahley, R.W. (2010). Polymorphisms in the hepatic lipase gene affect plasma HDL-cholesterol levels in a Turkish population. *J Lipid Res* *51*, 422-430.
- Hoekstra, M., Van Berkel, T.J., and Van Eck, M. (2010). Scavenger receptor BI: a multi-purpose player in cholesterol and steroid metabolism. *World J Gastroenterol* *16*, 5916-5924.
- Holleboom, A.G., Kuivenhoven, J.A., Peelman, F., Schimmel, A.W., Peter, J., Defesche, J.C., Kastelein, J.J., Hovingh, G.K., Stroes, E.S., and Motazacker, M.M. (2011). High prevalence of mutations in LCAT in patients with low HDL cholesterol levels in The Netherlands: identification and characterization of eight novel mutations. *Hum Mutat* *32*, 1290-1298.
- Holmes, M.V., Asselbergs, F.W., Palmer, T.M., Drenos, F., Lanktree, M.B., Nelson, C.P., Dale, C.E., Padmanabhan, S., Finan, C., Swerdlow, D.I., *et al.* (2015). Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* *36*, 539-550.
- Hopper, J.L., Bishop, D.T., and Easton, D.F. (2005). Population-based family studies in genetic epidemiology. *Lancet* *366*, 1397-1406.
- Hutchison, C.A., 3rd, Newbold, J.E., Potter, S.S., and Edgell, M.H. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature* *251*, 536-538.

Iacocca, M.A., Dron, J.S., and Hegele, R.A. (2019). Progress in finding pathogenic DNA copy number variations in dyslipidemia. *Curr Opin Lipidol* 30, 63-70.

Iacocca, M.A., and Hegele, R.A. (2017). Recent advances in genetic testing for familial hypercholesterolemia. *Expert Rev Mol Diagn* 17, 641-651.

Iacocca, M.A., and Hegele, R.A. (2018). Role of DNA copy number variation in dyslipidemias. *Curr Opin Lipidol* 29, 125-132.

Ikonen, E. (2008). Cellular cholesterol trafficking and compartmentalization. *Nat Rev Mol Cell Biol* 9, 125-138.

Inazu, A., Brown, M.L., Hesler, C.B., Agellon, L.B., Koizumi, J., Takata, K., Maruhama, Y., Mabuchi, H., and Tall, A.R. (1990). Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* 323, 1234-1238.

International HapMap, C. (2005). A haplotype map of the human genome. *Nature* 437, 1299-1320.

International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748-752.

Iqbal, J., and Hussain, M.M. (2009). Intestinal lipid absorption. *Am J Physiol Endocrinol Metab* 296, E1183-1194.

Ji, Y., Wang, N., Ramakrishnan, R., Sehayek, E., Huszar, D., Breslow, J.L., and Tall, A.R. (1999). Hepatic scavenger receptor BI promotes rapid clearance of high density lipoprotein free cholesterol and its transport into bile. *J Biol Chem* 274, 33398-33402.

Johannsen, T.H., Kamstrup, P.R., Andersen, R.V., Jensen, G.B., Sillesen, H., Tybjaerg-Hansen, A., and Nordestgaard, B.G. (2009). Hepatic lipase, genetically elevated high-density lipoprotein, and risk of ischemic cardiovascular disease. *J Clin Endocrinol Metab* 94, 1264-1273.

Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* 55, 765-772.

Johansen, C.T., and Hegele, R.A. (2011). Genetic bases of hypertriglyceridemic phenotypes. *Curr Opin Lipidol* 22, 247-253.

Johansen, C.T., and Hegele, R.A. (2012). The complex genetic basis of plasma triglycerides. *Curr Atheroscler Rep* 14, 227-234.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011a). Genetic determinants of plasma triglycerides. *J Lipid Res* 52, 189-206.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42, 684-687.

Johansen, C.T., Wang, J., Lanktree, M.B., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., Schwartz, S.M., *et al.* (2011b). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 31, 1916-1926.

Johansen, C.T., Wang, J., McIntyre, A.D., Martins, R.A., Ban, M.R., Lanktree, M.B., Huff, M.W., Peterfy, M., Mehrabian, M., Lusic, A.J., *et al.* (2012). Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 5, 66-72.

Jorgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., and Tybjaerg-Hansen, A. (2014). Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* 371, 32-41.

Justesen, J.M., Allin, K.H., Sandholt, C.H., Borglykke, A., Krarup, N.T., Grarup, N., Linneberg, A., Jorgensen, T., Hansen, T., and Pedersen, O. (2015). Interactions of Lipid Genetic Risk Scores With Estimates of Metabolic Health in a Danish Population. *Circ Cardiovasc Genet* 8, 465-472.

Kannel, W.B., Dawber, T.R., Friedman, G.D., Glennon, W.E., and McNamara, P.M. (1964). Risk Factors in Coronary Heart Disease. An Evaluation of Several Serum Lipids as Predictors of Coronary Heart Disease; the Framingham Study. *Ann Intern Med* 61, 888-899.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., *et al.* (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434-443.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41, 56-65.

Katsanis, N. (2016). The continuum of causality in human genetic disorders. *Genome Biol* 17, 233.

Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., South, S.T., and Working Group of the American College of Medical Genetics Laboratory Quality Assurance, C. (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 13, 680-685.

- Kei, A.A., Filippatos, T.D., Tsimihodimos, V., and Elisaf, M.S. (2012). A review of the role of apolipoprotein C-II in lipoprotein metabolism and cardiovascular disease. *Metabolism* 61, 906-921.
- Khera, A.V., Cuchel, M., de la Llera-Moya, M., Rodrigues, A., Burke, M.F., Jafri, K., French, B.C., Phillips, J.A., Mucksavage, M.L., Wilensky, R.L., *et al.* (2011). Cholesterol efflux capacity, high-density lipoprotein function, and atherosclerosis. *N Engl J Med* 364, 127-135.
- Kiss, R.S., Kavaslar, N., Okuhira, K., Freeman, M.W., Walter, S., Milne, R.W., McPherson, R., and Marcel, Y.L. (2007). Genetic etiology of isolated low HDL syndrome: incidence and heterogeneity of efflux defects. *Arterioscler Thromb Vasc Biol* 27, 1139-1145.
- Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., *et al.* (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* 50, 1514-1523.
- Kozarsky, K.F., Donahee, M.H., Rigotti, A., Iqbal, S.N., Edelman, E.R., and Krieger, M. (1997). Overexpression of the HDL receptor SR-BI alters plasma HDL and bile cholesterol levels. *Nature* 387, 414-417.
- Ku, C.S., Vasiliou, V., and Cooper, D.N. (2012). A new era in the discovery of de novo mutations underlying human genetic disease. *Hum Genomics* 6, 27.
- Kuivenhoven, J.A., and Hegele, R.A. (2014). Mining the genome for lipid genes. *Biochim Biophys Acta* 1842, 1993-2009.
- Lambert, J.E., and Parks, E.J. (2012). Postprandial metabolism of meal triglyceride in humans. *Biochim Biophys Acta* 1821, 721-726.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Latsuzbaia, A., Jaddoe, V.W., Hofman, A., Franco, O.H., and Felix, J.F. (2016). Associations of genetic variants for adult lipid levels with lipid levels in children. The Generation R Study. *J Lipid Res* 57, 2185-2192.
- Lee, J.H., Giannikopoulos, P., Duncan, S.A., Wang, J., Johansen, C.T., Brown, J.D., Plutzky, J., Hegele, R.A., Glimcher, L.H., and Lee, A.H. (2011). The transcription factor cyclic AMP-responsive element-binding protein H regulates triglyceride metabolism. *Nat Med* 17, 812-815.
- Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 95, 5-23.



- Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285-291.
- Lewis, G.F., Xiao, C., and Hegele, R.A. (2015). Hypertriglyceridemia in the genomic era: a new paradigm. *Endocr Rev* *36*, 131-147.
- Li, Y., Sidore, C., Kang, H.M., Boehnke, M., and Abecasis, G.R. (2011). Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res* *21*, 940-951.
- Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin, C., Alam, D., Alves, A.C., *et al.* (2017). Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet* *49*, 1758-1766.
- Lloyd-Jones, D.M. (2010). Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation* *121*, 1768-1777.
- Lloyd-Jones, D.M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L.J., Van Horn, L., Greenlund, K., Daniels, S., Nichol, G., Tomaselli, G.F., *et al.* (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation* *121*, 586-613.
- Lohmueller, K.E. (2014). The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet* *10*, e1004379.
- Lu, X., Peloso, G.M., Liu, D.J., Wu, Y., Zhang, H., Zhou, W., Li, J., Tang, C.S., Dorajoo, R., Li, H., *et al.* (2017). Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat Genet* *49*, 1722-1730.
- Lutsey, P.L., Rasmussen-Torvik, L.J., Pankow, J.S., Alonso, A., Smolenski, D.J., Tang, W., Coresh, J., Volcik, K.A., Ballantyne, C.M., Boerwinkle, E., *et al.* (2012). Relation of lipid gene scores to longitudinal trends in lipid levels and incidence of abnormal lipid levels among individuals of European ancestry: the Atherosclerosis Risk in Communities (ARIC) study. *Circ Cardiovasc Genet* *5*, 73-80.
- MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., *et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature* *508*, 469-476.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., *et al.* (2009). Finding the missing heritability of complex diseases. *Nature* *461*, 747-753.

- Marian, A.J. (2014). Causality in genetics: the gradient of genetic effects and back to Koch's postulates of causality. *Circ Res* *114*, e18-21.
- Matsunaga, T., Hiasa, Y., Yanagi, H., Maeda, T., Hattori, N., Yamakawa, K., Yamanouchi, Y., Tanaka, I., Obara, T., and Hamaguchi, H. (1991). Apolipoprotein A-I deficiency due to a codon 84 nonsense mutation of the apolipoprotein A-I gene. *Proc Natl Acad Sci USA* *88*, 2793-2797.
- McCaskie, P.A., Cadby, G., Hung, J., McQuillan, B.M., Chapman, C.M., Carter, K.W., Thompson, P.L., Palmer, L.J., and Beilby, J.P. (2006). The C-480T hepatic lipase polymorphism is associated with HDL-C but not with risk of coronary heart disease. *Clin Genet* *70*, 114-121.
- Miko, I. (2008). Phenotype Variability: Penetrance and Expressivity. *Nat Edu* *1*, 137.
- Milman, S., Atzmon, G., Crandall, J., and Barzilai, N. (2014). Phenotypes and genotypes of high density lipoprotein cholesterol in exceptional longevity. *Curr Vasc Pharmacol* *12*, 690-697.
- Motazacker, M.M., Peter, J., Treskes, M., Shoulders, C.C., Kuivenhoven, J.A., and Hovingh, G.K. (2013). Evidence of a polygenic origin of extreme high-density lipoprotein cholesterol levels. *Arterioscler Thromb Vasc Biol* *33*, 1521-1528.
- Musunuru, K., Pirruccello, J.P., Do, R., Peloso, G.M., Guiducci, C., Sougnez, C., Garimella, K.V., Fisher, S., Abreu, J., Barry, A.J., *et al.* (2010). Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* *363*, 2220-2227.
- Myocardial Infarction, G., Investigators, C.A.E.C., Stitzel, N.O., Stirrups, K.E., Masca, N.G., Erdmann, J., Ferrario, P.G., Konig, I.R., Weeke, P.E., Webb, T.R., *et al.* (2016). Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* *374*, 1134-1144.
- National Human Genome Research Institute (2020). The Cost of Sequencing a Human Genome (NIH). <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>
- Ng, D.S., Leiter, L.A., Vezina, C., Connelly, P.W., and Hegele, R.A. (1994). Apolipoprotein A-I Q[-2]X causing isolated apolipoprotein A-I deficiency in a family with analphalipoproteinemia. *J Clin Invest* *93*, 223-229.
- Norum, K.R., and Gjone, E. (1967). Familial serum-cholesterol esterification failure. A new inborn error of metabolism. *Biochim Biophys Acta* *144*, 698-700.
- Ouimet, M., Barrett, T.J., and Fisher, E.A. (2019). HDL and Reverse Cholesterol Transport. *Circ Res* *124*, 1505-1518.

- Palmer, L.J., and Cardon, L.R. (2005). Shaking the tree: mapping complex disease genes with linkage disequilibrium. *Lancet* 366, 1223-1234.
- Paquette, M., Chong, M., Theriault, S., Dufour, R., Pare, G., and Baass, A. (2017). Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *J Clin Lipidol* 11, 725-732 e725.
- Paradis, M.E., and Lamarche, B. (2006). Endothelial lipase: its role in cardiovascular disease. *Can J Cardiol* 22 *Suppl B*, 31B-34B.
- Peloso, G.M., Auer, P.L., Bis, J.C., Voorman, A., Morrison, A.C., Stitzel, N.O., Brody, J.A., Khetarpal, S.A., Crosby, J.R., Fornage, M., *et al.* (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet* 94, 223-232.
- Piccolo, S.R., Abo, R.P., Allen-Brady, K., Camp, N.J., Knight, S., Anderson, J.L., and Horne, B.D. (2009). Evaluation of genetic risk scores for lipid levels using genome-wide markers in the Framingham Heart Study. *BMC Proc* 3 *Suppl* 7, S46.
- Pisciotta, L., Favari, E., Magnolo, L., Simonelli, S., Adorni, M.P., Sallo, R., Fancello, T., Zavaroni, I., Ardigo, D., Bernini, F., *et al.* (2012). Characterization of three kindreds with familial combined hypolipidemia caused by loss-of-function mutations of ANGPTL3. *Circ Cardiovasc Genet* 5, 42-50.
- Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A., *et al.* (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322, 1702-1705.
- Porter, F.D., and Herman, G.E. (2011). Malformation syndromes caused by disorders of cholesterol synthesis. *J Lipid Res* 52, 6-34.
- Prive, F., Vilhjalmsjon, B.J., Aschard, H., and Blum, M.G.B. (2019). Making the Most of Clumping and Thresholding for Polygenic Scores. *Am J Hum Genet* 105, 1213-1221.
- Quintana-Murci, L. (2016). Understanding rare and common diseases in the context of human evolution. *Genome Biol* 17, 225.
- Rabbani, B., Tekin, M., and Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 59, 5-15.
- Rader, D.J., and Hovingh, G.K. (2014). HDL and cardiovascular disease. *Lancet* 384, 618-625.
- Raffield, L.M., Cox, A.J., Hsu, F.C., Ng, M.C., Langefeld, C.D., Carr, J.J., Freedman, B.I., and Bowden, D.W. (2013). Impact of HDL genetic risk scores on coronary artery calcified plaque and mortality in individuals with type 2 diabetes from the Diabetes Heart Study. *Cardiovasc Diabetol* 12, 95.

- Rahalkar, A.R., and Hegele, R.A. (2008). Monogenic pediatric dyslipidemias: classification, genetics and clinical spectrum. *Mol Genet Metab* 93, 282-294.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., *et al.* (2006). Global variation in copy number in the human genome. *Nature* 444, 444-454.
- Rees, M.G., Raimondo, A., Wang, J., Ban, M.R., Davis, M.I., Barrett, A., Ranft, J., Jagdhuhn, D., Waterstradt, R., Baltrusch, S., *et al.* (2014). Inheritance of rare functional GCKR variants and their contribution to triglyceride levels in families. *Hum Mol Genet* 23, 5570-5578.
- Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E., *et al.* (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 15, 733-747.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (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. *Genet Med* 17, 405-424.
- Romeo, S., Pennacchio, L.A., Fu, Y., Boerwinkle, E., Tybjaerg-Hansen, A., Hobbs, H.H., and Cohen, J.C. (2007). Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* 39, 513-516.
- Sabatine, M.S. (2019). PCSK9 inhibitors: clinical evidence and implementation. *Nat Rev Cardiol* 16, 155-165.
- Sachdeva, A., Cannon, C.P., Deedwania, P.C., Labresh, K.A., Smith, S.C., Jr., Dai, D., Hernandez, A., and Fonarow, G.C. (2009). Lipid levels in patients hospitalized with coronary artery disease: an analysis of 136,905 hospitalizations in Get With The Guidelines. *Am Heart J* 157, 111-117 e112.
- Sadananda, S.N., Foo, J.N., Toh, M.T., Cermakova, L., Trigueros-Motos, L., Chan, T., Liany, H., Collins, J.A., Gerami, S., Singaraja, R.R., *et al.* (2015). Targeted next-generation sequencing to diagnose disorders of HDL cholesterol. *J Lipid Res* 56, 1993-2001.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463-5467.
- Santos, R.D., Asztalos, B.F., Martinez, L.R., Miname, M.H., Polisecki, E., and Schaefer, E.J. (2008a). Clinical presentation, laboratory values, and coronary heart disease risk in marked high-density lipoprotein-deficiency states. *J Clin Lipidol* 2, 237-247.
- Santos, R.D., Miname, M.H., Martinez, L.R., Rochitte, C.E., Chacra, A.P., Nakandakare, E.R., Chen, D., and Schaefer, E.J. (2008b). Non-invasive detection of aortic and coronary

atherosclerosis in homozygous familial hypercholesterolemia by 64 slice multi-detector row computed tomography angiography. *Atherosclerosis* 197, 910-915.

Santos, R.D., Schaefer, E.J., Asztalos, B.F., Polisecki, E., Wang, J., Hegele, R.A., Martinez, L.R., Miname, M.H., Rochitte, C.E., Da Luz, P.L., *et al.* (2008c). Characterization of high density lipoprotein particles in familial apolipoprotein A-I deficiency. *J Lipid Res* 49, 349-357.

Schaefer, E.J., Anthanont, P., Diffenderfer, M.R., Polisecki, E., and Asztalos, B.F. (2016). Diagnosis and treatment of high density lipoprotein deficiency. *Prog Cardiovasc Dis* 59, 97-106.

Schaefer, E.J., Santos, R.D., and Asztalos, B.F. (2010). Marked HDL deficiency and premature coronary heart disease. *Curr Opin Lipidol* 21, 289-297.

Schwartz, G.G., Olsson, A.G., Abt, M., Ballantyne, C.M., Barter, P.J., Brumm, J., Chaitman, B.R., Holme, I.M., Kallend, D., Leiter, L.A., *et al.* (2012). Effects of dalcetrapib in patients with a recent acute coronary syndrome. *N Engl J Med* 367, 2089-2099.

Schwarze, K., Buchanan, J., Fermont, J.M., Dreau, H., Tilley, M.W., Taylor, J.M., Antoniou, P., Knight, S.J.L., Camps, C., Pentony, M.M., *et al.* (2020). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med* 22, 85-94.

Sharrett, A.R., Ballantyne, C.M., Coady, S.A., Heiss, G., Sorlie, P.D., Catellier, D., Patsch, W., and Atherosclerosis Risk in Communities Study, G. (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, 1108-1113.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H. (2017). DNA sequencing at 40: past, present and future. *Nature* 550, 345-353.

Simons, K., and Ikonen, E. (2000). How cells handle cholesterol. *Science* 290, 1721-1726.

Singaraja, R.R., Sivapalaratnam, S., Hovingh, K., Dube, M.P., Castro-Perez, J., Collins, H.L., Adelman, S.J., Riwanto, M., Manz, J., Hubbard, B., *et al.* (2013). The impact of partial and complete loss-of-function mutations in endothelial lipase on high-density lipoprotein levels and functionality in humans. *Circ Cardiovasc Genet* 6, 54-62.

Statistics Canada (2015). Health Fact Sheets: The 10 leading causes of death in Canada, 2013. <https://www150.statcan.gc.ca/n1/pub/82-625-x/2017001/article/14776-eng.htm>

Stitzel, N.O., Khera, A.V., Wang, X., Bierhals, A.J., Vourakis, A.C., Sperry, A.E., Natarajan, P., Klarin, D., Emdin, C.A., Zekavat, S.M., *et al.* (2017). ANGPTL3

Deficiency and Protection Against Coronary Artery Disease. *J Am Coll Cardiol* 69, 2054-2063.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H., *et al.* (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75-81.

Surendran, R.P., Visser, M.E., Heemelaar, S., Wang, J., Peter, J., Defesche, J.C., Kuivenhoven, J.A., Hosseini, M., Peterfy, M., Kastelein, J.J., *et al.* (2012). Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe hypertriglyceridaemia. *J Intern Med* 272, 185-196.

Tada, H., Won, H.H., Melander, O., Yang, J., Peloso, G.M., and Kathiresan, S. (2014). Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. *Circ Cardiovasc Genet* 7, 583-587.

Tall, A.R., and Rader, D.J. (2018). Trials and Tribulations of CETP Inhibitors. *Circ Res* 122, 106-112.

Talmud, P.J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J.A., Harrison, S.C., Li, K., Drenos, F., Karpe, F., *et al.* (2013). Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* 381, 1293-1301.

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 20, 467-484.

Taylor, F., Huffman, M.D., Macedo, A.F., Moore, T.H., Burke, M., Davey Smith, G., Ward, K., and Ebrahim, S. (2013). Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev*, CD004816.

Teare, M.D., and Barrett, J.H. (2005). Genetic linkage studies. *Lancet* 366, 1036-1044.

Teare, M.D., and Santibanez Koref, M.F. (2014). Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. *Brief Funct Genomics* 13, 378-383.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.

Tg, Hdl Working Group of the Exome Sequencing Project, N.H.L., Blood, I., Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitzel, N.O., Lange, L.A., Lu, Y., *et al.* (2014). Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 371, 22-31.

Tietjen, I., Hovingh, G.K., Singaraja, R.R., Radomski, C., Barhdadi, A., McEwen, J., Chan, E., Mattice, M., Legendre, A., Franchini, P.L., *et al.* (2012). Segregation of LIPG,

- CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol. *PLoS One* 7, e37437.
- Tikka, A., and Jauhiainen, M. (2016). The role of ANGPTL3 in controlling lipoprotein metabolism. *Endocrine* 52, 187-193.
- Tikkanen, E., Tuovinen, T., Widen, E., Lehtimäki, T., Viikari, J., Kahonen, M., Peltonen, L., Raitakari, O.T., and Ripatti, S. (2011). Association of known loci with lipid levels among children and prediction of dyslipidemia in adults. *Circ Cardiovasc Genet* 4, 673-680.
- Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., and Richards, J.B. (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nat Rev Genet* 19, 110-124.
- Tosheska-Trajkowska, K., and Topuzovska, S. (2017). High-density lipoprotein metabolism and reverse cholesterol transport: strategies for raising HDL cholesterol. *Anatol J Cardiol* 18, 149-154.
- Ueda, Y., Royer, L., Gong, E., Zhang, J., Cooper, P.N., Francone, O., and Rubin, E.M. (1999). Lower plasma levels and accelerated clearance of high density lipoprotein (HDL) and non-HDL cholesterol in scavenger receptor class B type I transgenic mice. *J Biol Chem* 274, 7165-7171.
- Varban, M.L., Rinninger, F., Wang, N., Fairchild-Huntress, V., Dunmore, J.H., Fang, Q., Gosselin, M.L., Dixon, K.L., Deeds, J.D., Acton, S.L., *et al.* (1998). Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci USA* 95, 4619-4624.
- Varbo, A., and Nordestgaard, B.G. (2016). Remnant Cholesterol and Triglyceride-Rich Lipoproteins in Atherosclerosis Progression and Cardiovascular Disease. *Arterioscler Thromb Vasc Biol* 36, 2133-2135.
- Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. *Am J Hum Genet* 90, 7-24.
- Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 101, 5-22.
- Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Holm, H., Ding, E.L., Johnson, T., *et al.* (2012). Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 380, 572-580.
- Wada, M., Iso, T., Asztalos, B.F., Takama, N., Nakajima, T., Seta, Y., Kaneko, K., Taniguchi, Y., Kobayashi, H., Nakajima, K., *et al.* (2009). Marked high density

lipoprotein deficiency due to apolipoprotein A-I Tomioka (codon 138 deletion). *Atherosclerosis* 207, 157-161.

Wang, J., Ban, M.R., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., and Hegele, R.A. (2008a). APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. *Nat Clin Pract Cardiovasc Med* 5, 730-737.

Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008b). Polygenic determinants of severe hypertriglyceridemia. *Hum Mol Genet* 17, 2894-2899.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L., and Hegele, R.A. (2007). Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol* 27, 2450-2455.

Wang, J., Dron, J.S., Ban, M.R., Robinson, J.F., McIntyre, A.D., Alazzam, M., Zhao, P.J., Dilliot, A.A., Cao, H., Huff, M.W., *et al.* (2016). Polygenic Versus Monogenic Causes of Hypercholesterolemia Ascertained Clinically. *Arterioscler Thromb Vasc Biol* 36, 2439-2445.

Wang, N., Arai, T., Ji, Y., Rinninger, F., and Tall, A.R. (1998). Liver-specific overexpression of scavenger receptor BI decreases levels of very low density lipoprotein ApoB, low density lipoprotein ApoB, and high density lipoprotein in transgenic mice. *J Biol Chem* 273, 32920-32926.

Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* 14, 125-138.

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.

Wilson, P.W., Garrison, R.J., Castelli, W.P., Feinleib, M., McNamara, P.M., and Kannel, W.B. (1980). Prevalence of coronary heart disease in the Framingham Offspring Study: role of lipoprotein cholesterols. *Am J Cardiol* 46, 649-654.

Winsor, E. (1988). Mendelian genetics. *Can Fam Physician* 34, 859-862.

World Health Organization (2020). Cardiovascular diseases.

<https://www.who.int/westernpacific/health-topics/cardiovascular-diseases>

Wright, C.F., West, B., Tuke, M., Jones, S.E., Patel, K., Laver, T.W., Beaumont, R.N., Tyrrell, J., Wood, A.R., Frayling, T.M., *et al.* (2019). Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* 104, 275-286.



Xiao, C., Stahel, P., and Lewis, G.F. (2019). Regulation of Chylomicron Secretion: Focus on Post-Assembly Mechanisms. *Cell Mol Gastroenterol Hepatol* 7, 487-501.

Yamashita, S., Matsuzawa, Y., Okazaki, M., Kako, H., Yasugi, T., Akioka, H., Hirano, K., and Tarui, S. (1988). Small polydisperse low density lipoproteins in familial hyperalphalipoproteinemia with complete deficiency of cholesteryl ester transfer activity. *Atherosclerosis* 70, 7-12.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., *et al.* (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42, 565-569.

Young, S.G., and Zechner, R. (2013). Biochemistry and pathophysiology of intravascular and intracellular lipolysis. *Genes Dev* 27, 459-484.

Yu, J.E., Han, S.Y., Wolfson, B., and Zhou, Q. (2018). The role of endothelial lipase in lipid metabolism, inflammation, and cancer. *Histol Histopathol* 33, 1-10.

Zambon, A., Deeb, S.S., Hokanson, J.E., Brown, B.G., and Brunzell, J.D. (1998). Common variants in the promoter of the hepatic lipase gene are associated with lower levels of hepatic lipase activity, buoyant LDL, and higher HDL2 cholesterol. *Arterioscler Thromb Vasc Biol* 18, 1723-1729.

Zanoni, P., Khetarpal, S.A., Larach, D.B., Hancock-Cerutti, W.F., Millar, J.S., Cuchel, M., DerOhannessian, S., Kontush, A., Surendran, P., Saleheen, D., *et al.* (2016). Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* 351, 1166-1171.

Zarrei, M., MacDonald, J.R., Merico, D., and Scherer, S.W. (2015). A copy number variation map of the human genome. *Nat Rev Genet* 16, 172-183.

Zilversmit, D.B. (1995). Atherogenic nature of triglycerides, postprandial lipidemia, and triglyceride-rich remnant lipoproteins. *Clinical Chemistry* 41, 153-158.

Zubair, N., Mayer-Davis, E.J., Mendez, M.A., Mohlke, K.L., North, K.E., and Adair, L.S. (2014). Genetic risk score and adiposity interact to influence triglyceride levels in a cohort of Filipino women. *Nutr Diabetes* 4, e118.

Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. *Proc Natl Acad Sci USA* 111, E455-464.

## Chapter 2 – Polygenic determinants in extremes of high-density lipoprotein cholesterol

The work contained in this Chapter has been edited from its original publication in the *Journal of Lipid Research* for brevity and to ensure consistency throughout this Dissertation.

JOURNAL OF LIPID RESEARCH



Supplemental Material can be found at:  
<http://www.jlr.org/content/suppl/2017/09/24/jlr.M079822.DC1>

*patient-oriented and epidemiological research*

◆ Author's Choice

### Polygenic determinants in extremes of high-density lipoprotein cholesterol<sup>®</sup>

Jacqueline S. Dron,<sup>\*,†</sup> Jian Wang,<sup>†</sup> Cécile Low-Kam,<sup>‡</sup> Sumeet A. Khetarpal,<sup>‡</sup> John F. Robinson,<sup>†</sup> Adam D. McIntyre,<sup>†</sup> Matthew R. Ban,<sup>†</sup> Henian Cao,<sup>†</sup> David Rhoads,<sup>§</sup> Marie-Pierre Dubé,<sup>§</sup> Daniel J. Rader,<sup>¶</sup> Guillaume Lettre,<sup>¶</sup> Jean-Claude Tardif,<sup>¶</sup> and Robert A. Hegele<sup>1,\*,†,‡</sup>

Department of Biochemistry,<sup>\*</sup> Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; Robarts Research Institute,<sup>†</sup> Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; Montréal Heart Institute et Faculté de Médecine,<sup>‡</sup> Université de Montréal, Montréal, Québec, Canada; Departments of Genetics and Medicine,<sup>§</sup> Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; Departments of Genetics, Medicine, and Pediatrics,<sup>¶</sup> the Cardiovascular Institute, and the Institute for Translational Medicine and Therapeutics, Perleman School of Medicine, University of Pennsylvania, Philadelphia, PA; and Department of Medicine,<sup>¶</sup> Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

ORCID ID: 0000-0002-3045-6530 (J.S.D.); 0000-0003-2861-5325 (R.A.H.)

Downloaded from www.jlr.org by guest on October 14, 2020

**Abstract** HDL cholesterol (HDL-C) remains a superior biochemical predictor of CVD risk, but its genetic basis is incompletely defined. In patients with extreme HDL-C concentrations, we concurrently evaluated the contributions of multiple large- and small-effect genetic variants. In a discovery cohort of 255 unrelated lipid clinic patients with extreme HDL-C levels, we used a targeted next-generation sequencing panel to evaluate rare variants in known HDL metabolism genes, simultaneously with common variants bundled into a polygenic trait score. Two additional cohorts were used for validation and included 1,746 individuals from the Montréal Heart Institute Biobank and 1,048 individuals from the University of Pennsylvania. Findings were consistent between cohorts: we found rare heterozygous large-effect variants in 18.7% and 10.9% of low- and high-HDL-C patients, respectively. We also found common variant accumulation, indicated by extreme polygenic trait scores, in an additional 12.8% and 19.3% of overall cases of low- and high-HDL-C extremes, respectively. Thus, the genetic basis of extreme HDL-C concentrations encountered clinically is frequently polygenic, with contributions from both rare large-effect and common small-effect variants. Multiple types of genetic variants should be considered as contributing factors in patients with extreme dyslipidemia.—Dron, J.S., Wang, C. Low-Kam, S. A. Khetarpal, J. F. Robinson, A. D. McIntyre, M. R. Ban, H. Cao, D. Rhoads, M.-P. Dubé, D. J. Rader, G. Lettre, J.-C. Tardif, and R. A. Hegele. Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J. Lipid Res.* 2017, 58: 2162–2170.

**Supplementary key words** genetics • genomics • dyslipidemias • genes in lipid dysfunction • complex trait • rare variants • common variants • next-generation sequencing • polygenic risk score

Despite apprehension over its direct causal role in atherogenesis and its value as a drug target (1), HDL cholesterol (HDL-C) remains a valid biochemical predictor of CVD risk (2–4). Understanding the full range of factors that determine plasma HDL-C concentrations, including genetics, still has relevance for epidemiology and risk projection (5). Furthermore, specific etiologies of extreme perturbations of HDL-C may have clinical importance in terms of diagnosis and directed therapies (1, 6).

Multiple genetic factors could be present in an individual, creating a polygenic network of influential determinants on HDL-C (7–9). These determinants include monogenic disorders (10, 11), such as extremely low or absent HDL-C due to rare homozygous mutations in *ABCA1*, *LCAT*, and *APOA1* (12–15), and extremely elevated HDL-C due to

**Abbreviations:** 1KG, 1000 Genomes Project; FH, familial hypercholesterolemia; GWAS, genome-wide association study; HDL-C, HDL cholesterol; LDL-C, LDL cholesterol; M+, mutation positive; M–, mutation negative; MHI, Montréal Heart Institute; OR, odds ratio; PIS, polygenic trait score; UPen, University of Pennsylvania.

<sup>†</sup>To whom correspondence should be addressed.  
 E-mail: hegele@roberts.ca

<sup>®</sup>The online version of this article (available at <http://www.jlr.org>) contains a supplement.

Copyright © 2017 by the American Society for Biochemistry and Molecular Biology, Inc. This article is available online at <http://www.jlr.org>

This work was supported by the Jacob J. Wolfe Distinguished Medical Research Chair, the Edith Schulich Vinet Research Chair in Human Genetics, and the Martha G. Blackburn Chair in Cardiovascular Research, as well as operating grants from the Canadian Institutes of Health Research (Fundation Accord), the Heart and Stroke Foundation of Ontario (G13-0009214), and Genome Canada, through Genome Quebec (award #330) (R.A.H.). This work was also supported by the Montréal Heart Institute foundations and the Demarus family, in support of Genome Quebec/Genome Canada. R.A.H. is a consultant and speaker's bureau member for Amgen, Angion, Boston Heart Diagnostics, Genis, Eli Lilly, Genphix, Pfizer, and Sanofi. The other authors have no disclosures.

\*Author's Choice—Final version free via Creative Commons CC-BY license.  
 Manuscript received 9 August 2017 and in revised form 31 August 2017.  
 Published, JLR Papers in Press, September 4, 2017  
 DOI <https://doi.org/10.1194/jlr.M079822>

2162 Journal of Lipid Research Volume 58, 2017

**Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhoads, D., Dube, M.P., et al. (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.**

## 2.1 Abstract

**Objective:** Levels of high-density lipoprotein (HDL) cholesterol remain a superior biochemical predictor of cardiovascular disease risk, but its genetic basis is incompletely defined. In patients with extreme HDL cholesterol concentrations, we concurrently evaluated the presence of rare variants and the accumulation of multiple common variants.

**Methods and Results:** In a discovery sample of 255 unrelated lipid clinic patients with extreme HDL cholesterol levels, we used a targeted next-generation sequencing panel to evaluate rare variants in known HDL metabolism genes, and simultaneously assessed the burden of common variants using a novel polygenic risk score. Two additional cohorts were used to validate our polygenic risk score, totaling 2,794 individuals. After combining cohorts, we found rare variants in 18.7% and 10.9% of low and high HDL cholesterol patients, respectively. We also found common variant accumulation—indicated by extreme polygenic risk scores—in an additional 12.8% and 19.3% of overall cases of low and high HDL cholesterol extremes, respectively.

**Conclusions:** The genetic basis of extreme HDL cholesterol concentrations encountered clinically is comprised of both rare and common variants. Multiple types of genetic variants should be considered as contributing factors in patients with extreme dyslipidemia.

## 2.2 Introduction

Despite apprehension over its direct causal role in atherogenesis and value as a drug target (Rosenson, 2016), high-density lipoprotein (HDL) cholesterol remains a valid biochemical predictor of cardiovascular disease (CVD) risk (Emerging Risk Factors Consortium et al., 2012; Parish et al., 2012; Perk et al., 2012). Understanding the full range of factors that determine plasma HDL cholesterol concentrations, including genetics, still has relevance for epidemiology and risk projection (Raffield et al., 2013). Furthermore, specific etiologies of extreme perturbations of HDL cholesterol may have

clinical importance in terms of diagnosis and directed therapies (Hovingh et al., 2015; Rosenson, 2016).

Multiple genetic factors could be present in an individual, creating a polygenic network of influential determinants on HDL cholesterol levels (Cohen et al., 2004; Hegele, 2009; Motazacker et al., 2013). These determinants include monogenic disorders (Dron and Hegele, 2016; Weissglas-Volkov and Pajukanta, 2010), such as extremely low or absent HDL cholesterol levels (i.e. “hypoalphalipoproteinemia”) due to bi-allelic rare variants in *ABCA1*, *LCAT* and *APOA1* (Brooks-Wilson et al., 1999; Kuivenhoven et al., 1996; Ng et al., 1994; Schaefer et al., 2016), and extremely elevated HDL cholesterol levels (i.e. “hyperalphalipoproteinemia”) due to rare variants in *CETP*, *LIPC*, *SCARB1*, and *LIPG* (Hegele et al., 1993; Inazu et al., 1990; Tietjen et al., 2012; Zanoni et al., 2016). In contrast, the potential role of other genetic determinants in extreme, non-monogenic HDL cholesterol phenotypes, namely common single-nucleotide polymorphisms (SNPs) (Rosenson, 2016), has not been systematically evaluated.

Polygenic factors—which can be assessed by quantifying the accumulation of SNPs with small phenotypic effects using polygenic scores—contribute to numerous medical conditions, including coronary artery disease (McPherson and Tybjaerg-Hansen, 2016) and diabetes (Bonfond and Froguel, 2015). Among dyslipidemias, polygenic factors play a substantial role in familial hypercholesterolemia (FH) (Talmud et al., 2013), which was previously considered an archetypal “monogenic” disorder. For instance, in patients referred with extremely elevated low-density lipoprotein (LDL) cholesterol, targeted next-generation sequencing (NGS) demonstrated that ~50% of individuals had heterozygous rare variants while another ~16% had an accumulation of common SNPs, identified previously from genome-wide association studies (GWASs) as determinants of LDL cholesterol (Wang et al., 2016). While earlier sequencing experiments indicate that 11-35% of patients with extremely low HDL cholesterol and 5-20% of patients with extremely high HDL cholesterol have heterozygous rare variants driving the phenotypes, the proportion of such patients with excessive GWAS-identified SNPs, as quantified using polygenic risk scores, is unknown (Candini et al., 2010; Cohen et al., 2004;

Holleboom et al., 2011; Kiss et al., 2007; Sadananda et al., 2015; Singaraja et al., 2013; Tietjen et al., 2012).

Here we used targeted NGS to robustly characterize the genetic determinants influencing HDL cholesterol levels in patients with low and high HDL cholesterol phenotypes. This allowed us to concurrently evaluate the burden of rare variants and common SNPs, the latter assessed using a polygenic score. We saw that ~30% of individuals at each HDL cholesterol extreme had an identifiable genetic determinant, with an extreme SNP accumulation being more common than the presence of a rare variant. Our findings illustrate that both types of determinants are enriched in individuals with extremely low and high HDL cholesterol levels compared to normolipidemic controls.

## 2.3 Materials and Methods

### 2.3.1 Study subjects

Patients of interest included those of European ancestry and with either low or high levels of HDL cholesterol from the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada), the Montréal Heart Institute (MHI) Biobank (Montréal, QC, Canada), or the University of Pennsylvania (UPenn) (Philadelphia, PA, USA).

Low HDL cholesterol was defined as  $\leq 0.8$  mmol/L and  $\leq 1.0$  mmol/L in males and females, respectively. High HDL cholesterol was defined as  $\geq 1.4$  mmol/L and  $\geq 1.8$  mmol/L in males and females, respectively. These thresholds adhere closely to the top and bottom 10<sup>th</sup> percentiles of HDL cholesterol levels in a population largely of European ancestry (Rifkind and Segal, 1983). The two patient exclusion criteria were: 1) triglyceride levels  $\geq 3.37$  mmol/L—as low HDL cholesterol can simply be secondary to elevated triglycerides, which have their own distinct determinants—and 2) diagnosis of monogenic syndromes of extreme HDL cholesterol (e.g. Tangier disease).

In adherence to the Declaration of Helsinki, all patients provided written, informed consent for collection of personal data and DNA with approval from the Western

University (London ON, Canada) ethics review board (no. 07290E) or the patients' originating study centre.

As a reference control cohort of normolipidemic individuals, the publicly available data pertaining to the European subgroup of the 1000 Genomes Project (N=503) were studied. An additional 1,198 normolipidemic individuals were assessed from the MHI Biobank, ascertained as previously described (Low-Kam et al., 2016).

### 2.3.2 DNA preparation and targeted sequencing

Genomic DNA was extracted from patient blood samples using the Puregene® DNA Blood Kit (Gentra Systems, Qiagen Inc., Mississauga, ON, Canada) (*Cat No. 158389*). Sequencing libraries consisting of 24 patient DNA samples were generated for indexing and enrichment with the Nextera® Rapid Capture Custom Enrichment Kit (*Cat No. FC-140-1009*) “LipidSeq” design (Johansen et al., 2014). Briefly, samples were enriched for genomic areas in accordance with our “LipidSeq” panel, which captures 69 genes (all exons, and 50 bases into the intron from each splice junction) and 185 SNPs associated with dyslipidemia and other metabolic disorders (**Figure 1.15**) (Johansen et al., 2014). These libraries were then sequenced at the London Regional Genomics Centre ([www.lrgc.ca](http://www.lrgc.ca); London ON, Canada) on an Illumina MiSeq personal sequencer (Illumina, San Diego CA, USA).

Sequencing and genotyping methods performed at the MHI Biobank (Low-Kam et al., 2016) and UPenn (Zanoni et al., 2016) are described elsewhere.

### 2.3.3 Bioinformatic processing of sequencing data

After sequencing, two FASTQ files were generated for each patient sample—one each for sequencing reads generated for forward and reverse strands—and imported into CLC Bio Genomics Workbench (version 7.5; CLC Bio, Aarhus, Denmark). For each patient sample, the sequencing reads within each FASTQ file were mapped and aligned against the human reference genome (GRCh37/hg19); a secondary local alignment was performed to control for possible misalignment due to insertions or deletions not present in the reference genome. Duplicate mapped reads due to PCR amplification from the

library preparation were removed to ensure accurate depth-of-coverage metrics for each sequenced nucleotide. From the reassembled sequencing reads, variants with a minimum 30-fold coverage and 35% variant frequency were called for each patient sample and exported into VCF files (Dilliott et al., 2018; Johansen et al., 2014).

Sequence data from the European subset of the 1000 Genomes Project were downloaded and filtered for the genomic coordinates captured by our LipidSeq panel using PLINK v1.9 (Purcell et al., 2007).

### 2.3.4 Annotation and analysis of rare variants

Variants were annotated with a customized ANNOVAR annotation pipeline (Wang et al., 2010). Annotation methods performed at the MHI Biobank (Low-Kam et al., 2016) and UPenn (Zanoni et al., 2016) are described elsewhere.

Rare variants were defined as those with a minor allele frequency of  $\leq 1\%$  or missing in the 1000 Genomes Project (<http://browser.1000genomes.org/index.html>) (Genomes Project et al., 2015), Exome Sequencing Project (ESP; <http://evs.gs.washington.edu/EVS/>), and Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org/>) (Lek et al., 2016) databases. Rare variants were considered to have large phenotypic effects if they met the following criteria: 1) sequence ontology of either missense, nonsense, deletion, insertion, splice-acceptor site, or splice-donor site; and 2) deleterious or damaging predictions in at least half of the available *in silico* prediction tools, including Polymorphism Phenotyping version 2 (PolyPhen2; <http://genetics.bwh.harvard.edu/pph2/>) (Adzhubei et al., 2013), Sorting Intolerant from Tolerant (SIFT; <http://sift.jcvi.org/>) (Kumar et al., 2009), MutationTaster (<http://www.mutationtaster.org/>), Combined Annotation Dependent Depletion (CADD; <http://cadd.gs.washington.edu/score>) (Kircher et al., 2014), Splicing Based Analysis of Variants (SPANR; <http://tools.genes.toronto.edu/>) (Xiong et al., 2015), and Automated Splice Site and Exon Definition Analyses (ASSEDA; <http://splice.uwo.ca/>) (Mucaki et al., 2013).

We also considered rare variants that did not necessarily meet the above criteria, but were previously reported in the Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/ac/all.php>) (Stenson et al., 2014) as causative for either lowering or raising levels of HDL cholesterol.

Of the variants meeting the above criteria, those within lipid-associated genes with candidate (primary) and non-candidate (secondary) effects on HDL cholesterol levels were considered for analysis (**Table 2.1**). It is important to note that since the UPenn cohort comes from an established on-going study (the UPenn High HDL Cholesterol Study), the criteria used in identifying rare variants of interest differs slightly from what was considered here (Edmondson et al., 2009; Zanoni et al., 2016). To ensure consistent filtering criteria, the UPenn cohort was excluded from the rare variant analysis and was only used in the validation of our polygenic score.



**Table 2.1 Genes with candidate (primary) and non-candidate (secondary) influences on HDL cholesterol levels.**

Influence on HDL cholesterol	Phenotype	Gene	Related disorder	
Primary	Low HDL cholesterol	<i>ABCA1</i>	Tangier disease	
		<i>APOA1</i>	Apolipoprotein A-I deficiency	
		<i>LCAT</i>	Familial LCAT deficiency	
	High HDL cholesterol		<i>LIPC</i>	Hepatic lipase deficiency
			<i>SCARB1</i>	SR-BI deficiency
			<i>CETP</i>	CETP deficiency
			<i>LIPG</i>	Hyperalphalipoproteinemia
Secondary	Low triglyceride	<i>APOC3</i>	Apolipoprotein C-III deficiency	
	High triglyceride	<i>LPL</i>	Lipoprotein lipase deficiency	
		<i>APOA2</i>	Apolipoprotein C-II deficiency	
		<i>APOA5</i>	Apolipoprotein A-V deficiency	
		<i>LMF1</i>	Lipase maturation factor deficiency	
		<i>GPIHBP1</i>	Severe hypertriglyceridemia	
		<i>GPD1</i>	Infantile hypertriglyceridemia	
<i>APOE</i>	Dysbetalipoproteinemia			

Genes of interest were selected based on phenotypic reporting by Johansen et al. (2014). Abbreviation: CETP = cholesteryl ester transfer protein; HDL = high-density lipoprotein; LCAT = lecithin-cholesterol acyltransferase; SR-BI = scavenger receptor class B type I.

### 2.3.5 Polygenic risk score for high-density lipoprotein cholesterol levels

Between the LipidSeq targets and 1000 Genomes Project variant data, genotypes for 34 HDL cholesterol-associated SNPs were available for study; these SNPs were selected from the most recent GWAS meta-analyses on blood lipids and lipoproteins, published by the Global Lipids Genetics Consortium (GLGC) (Willer et al., 2013). A polygenic score encompassing all available SNPs was calculated for patients in the discovery cohort (i.e. the Lipid Genetics Clinic cohort). In the interest of future application and usability, smaller sets of 10 SNPs or less were tested and compared to the original 34-SNP score—the aim was to select a smaller number of SNPs that were just as informative as the full set of 34. For each SNP set, SNPs could not be in linkage disequilibrium (LD) with each other.

Scores were calculated using a weighted approach; the number of alleles associated with raising HDL cholesterol at a locus (0, 1, or 2) were summed and multiplied by the reported effect size for the respective allele. The products for each locus were totalled to provide the overall polygenic risk score for an individual. The underlying assumption when calculating the polygenic risk score was that each allele had an additive effect towards their respective HDL cholesterol phenotypes. Higher scores indicated that individuals carried a greater number of alleles associated with raising HDL cholesterol levels, while lower scores indicated that individuals carried fewer alleles associated with raising HDL cholesterol, and therefore carried a greater number of alleles associated with lowering HDL cholesterol levels.

### 2.3.6 Statistical analysis

Normality was assessed using the D'Agostino and Pearson test. Differences between parametric data were assessed using an unpaired Students t-test while differences between nonparametric data were assessed using a Mann-Whitney test. Differences between mean polygenic risk scores were assessed using a one-tailed, unpaired Wilcoxon rank-sum tests. All tests were performed assuming unequal variances and are reported as the mean  $\pm$  standard deviation (SD). Odds ratios (ORs) were derived using 2-by-2

contingency tables, with Fisher's exact tests to assess significance. Statistical analyses were conducted using SAS (version 9.3; SAS Institute, Cary NC, USA). Statistical significance was defined as  $P < 0.05$ .

## 2.4 Results

### 2.4.1 Characteristics of study subjects

Two hundred and fifty-five unrelated patients were selected for study from the Lipid Genetics Clinic; 136 patients had low HDL cholesterol levels and 119 patients had high HDL cholesterol levels. An additional cohort of 201 and 347 patients with low and high HDL cholesterol levels, respectively, were selected from the MHI Biobank. Further, 349 and 699 patients with low and high HDL cholesterol levels, respectively, were selected from UPenn, ascertained as previously described (Edmondson et al., 2009; Zanoni et al., 2016).

Clinical and demographic information for patients with low and high HDL cholesterol levels from the Lipid Genetics Clinic, the MHI Biobank, and UPenn are summarized in **Table 2.2** and **Table 2.3**.

### 2.4.2 Rare variants identified in high-density lipoprotein cholesterol-altering genes

A total of 68 unique variants were identified in patients from the Lipid Genetics Clinic: 43 were in primary genes, and 10 were in secondary genes (**Figure 2.1A**, **Appendix C**, **Appendix D**). When considering variants in the primary genes, 72.1% were missense, 4.7% were splicing, 14.0% were frameshift, and 9.3% were nonsense (**Figure 2.1B**). One individual was homozygous for *ABCA1* p.G851R, and one individual was a compound heterozygote for *ABCA1* p.W590C and p.W590L. A single individual carried rare heterozygous variants in both a low and high HDL cholesterol-associated gene—i.e. *ABCA1* and *SCARB1*—and presented with low HDL cholesterol levels.

**Table 2.2 Clinical and demographic information of patients with low HDL cholesterol levels (N=686).**

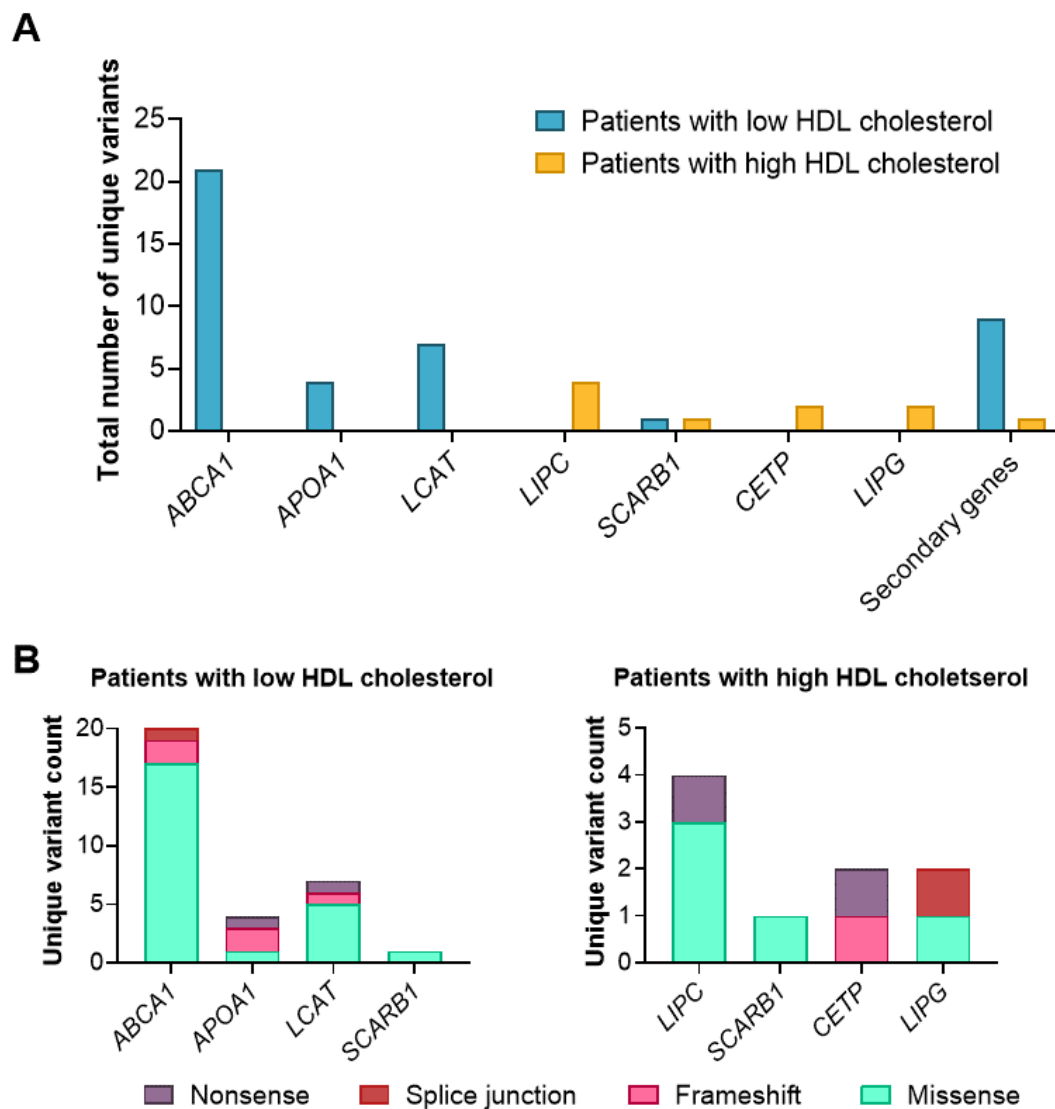
	Lipid Genetics Clinic		Montréal Heart Institute Biobank		University of Pennsylvania	
	Males	Females	Males	Females	Males	Females
<b>N</b>	90	46	131	70	202	147
<b>Age</b>	48.1 ± 16.8*	45.4 ± 12.5*	64.4 ± 10.4	68.9 ± 8.4	56.0 ± 12.2	53.2 ± 15.0
<b>BMI (kg/m<sup>2</sup>)</b>	29.0 ± 5.6*	28.8 ± 6.0*	31.0 ± 5.2*	31.4 ± 6.4*	32.4 ± 5.0*	34.5 ± 7.4*
<b>Total cholesterol (mmol/L)</b>	4.2 ± 1.4	5.8 ± 2.3	3.4 ± 1.1	3.7 ± 1.0	4.0 ± 1.1	4.5 ± 1.3
<b>HDL cholesterol (mmol/L)</b>	0.6 ± 0.2	0.8 ± 0.2	0.7 ± 0.1	0.9 ± 0.1	0.8 ± 0.2	0.9 ± 0.2
<b>LDL cholesterol (mmol/L)</b>	2.7 ± 1.3	4.0 ± 2.3	2.2 ± 1.0	2.2 ± 0.9	2.7 ± 1.3	3.2 ± 1.7
<b>Triglyceride (mmol/L)</b>	2.2 ± 1.3	2.0 ± 1.1	2.2 ± 0.7	2.3 ± 0.7	1.8 ± 0.7	1.6 ± 0.6
<b>CVD Hx</b>	45.2%*	21.9%*	60.3%	67.1%	11.9%*	12.9%*

Values are indicative of the mean ± SD. “\*” indicates means were calculated with an incomplete dataset. Abbreviations: BMI = body-mass index; CVD Hx = cardiovascular disease history; HDL = high-density lipoprotein; LDL = low-density lipoprotein.

**Table 2.3 Clinical and demographic information of patients with high HDL cholesterol levels (N=1,165).**

	Lipid Genetics Clinic		Montréal Heart Institute Biobank		University of Pennsylvania	
	Males	Females	Males	Females	Males	Females
<b>N</b>	60	59	280	67	217	482
<b>Age</b>	58.5 ± 14.2	58.6 ± 10.5	65.6 ± 10.1	71.2 ± 7.2	58.7 ± 14.9	58.2 ± 11.7
<b>BMI (kg/m<sup>2</sup>)</b>	26.5 ± 3.7	25.3 ± 3.5*	26.9 ± 4.5*	26.4 ± 6.0*	29.0 ± 5.0*	27.2 ± 7.0*
<b>Total cholesterol (mmol/L)</b>	5.7 ± 1.4	6.9 ± 1.5	4.5 ± 1.0 <sup>a</sup>	5.2 ± 1.1	6.5 ± 1.6	6.4 ± 1.2
<b>HDL cholesterol (mmol/L)</b>	2.1 ± 0.5	2.7 ± 0.7	1.7 ± 0.2	2.1 ± 0.3	2.5 ± 0.5	2.9 ± 0.5
<b>LDL cholesterol (mmol/L)</b>	3.2 ± 1.4	3.7 ± 1.5	2.4 ± 0.9 <sup>a</sup>	2.5 ± 0.9	3.0 ± 1.2	3.6 ± 2.0
<b>Triglyceride (mmol/L)</b>	1.0 ± 0.5	1.2 ± 0.6	1.4 ± 0.6	1.3 ± 0.5	0.9 ± 0.4	0.9 ± 0.4
<b>CVD Hx</b>	29.8%*	16.7%*	40.7%	31.3%	6.0%*	4.1%*

Values are indicative of the mean ± SD. “\*” indicates means were calculated with an incomplete dataset. Abbreviations: BMI = body-mass index; CVD Hx = cardiovascular disease history; HDL = high-density lipoprotein; LDL = low-density lipoprotein.



**Figure 2.1 Summary of rare variants identified within patients from the Lipid Genetics Clinic cohort (N=255).**

**A)** A total of 68 unique variants were identified: 43 were in primary genes, and 10 were in secondary genes. **B)** For each unique variant within the primary genes, breakdown by variant ontology has been presented for patients with low HDL cholesterol levels (left) and high HDL cholesterol levels (right). Abbreviations: HDL = high-density lipoprotein.

Only a few rare variants were identified in secondary genes of interest (**Appendix D**). In nine low HDL cholesterol patients, missense variants were identified in *LPL*, *APOA5*, *LMF1*, *GPD1*, and *APOE*. In two high HDL cholesterol patients, the same splicing variant was identified in *APOC3*. All variants in the secondary genes were heterozygous.

Overall, 30.1% and 12.6% of patients from the Lipid Genetics Clinic with low and high HDL cholesterol, respectively, carried at least one variant contributing towards their phenotype. In the MHI cohort, 10.9% and 10.4% of patients with of patients with low and high HDL cholesterol, respectively, carried rare variants—all were heterozygous. In the UPenn cohort, since different criteria were used in rare variant identification carriers were not considered for analysis.

### 2.4.3 Polygenic risk score development

After testing polygenic risk scores made up of 10 SNPs or less, a set of nine SNPs were selected to make up the polygenic risk score used in this study, as the score's results were the most similar to the results from the original 34-SNP score. The nine SNPs were in linkage equilibrium and showed significant primary associations with plasma levels of HDL cholesterol; some of the loci were previously implicated either directly or indirectly to HDL metabolism (**Table 2.4**). Each SNP was selected on the basis of its reported P-value; the most significantly associated SNPs were of top priority. The allele associated with higher HDL cholesterol levels was taken as the primary variable.

**Table 2.4 The 9 SNPs used in the polygenic risk score for HDL cholesterol levels.**

Chr:position	rsID	Closest gene	Effect allele	Relation with HDL metabolism	Variant ontology
1:182199750	rs1689800	<i>ZNF648</i>	A (0.034)	Mechanism is poorly characterized.	Upstream
1:230159944	rs4846914	<i>GALNT2</i>	A (0.048)	Recently confirmed as an important determinant of HDL cholesterol (Khetarpal et al., 2016).	Upstream
9:104902020	rs1883025	<i>ABCA1</i>	C (0.07)	Causative gene for Tangier disease (Hovingh et al., 2015).	Downstream
12:109562388	rs7134594	<i>MVK-MMAB</i>	T (0.035)	<i>MVK</i> encodes mevalonate kinase, which is involved in biosynthesis of cholesterol and isoprenoids (Browne and Timson, 2015), although the closely linked <i>MMAB</i> gene encoding cob(I)alamin adenosyltransferase may actually underlie the HDL cholesterol association at this locus (Fogarty et al., 2010).	Upstream
12:124777047	rs838880	<i>SCARB1</i>	C (0.048)	Causative gene for SR-BI deficiency.	Downstream
15:58391167	rs1532085	<i>LIPC</i>	A (0.107)	Causative gene for hepatic lipase deficiency.	Upstream
16:56959412	rs3764261	<i>CETP</i>	A (0.241)	Causative gene for CETP deficiency. Facilitates the transfer of lipids between HDL and triglyceride-rich lipoproteins.	Upstream
16:81501185	rs2925979	<i>CMIP</i>	C (0.035)	Mechanism is poorly characterized.	Intronic
19:8368312	rs7255436	<i>ANGPTL4</i>	A (0.032)	Regulates lipoprotein lipase with reciprocal effects on triglycerides and HDL cholesterol (Dijk and Kersten, 2014).	Upstream

Variant information related to effect size was extracted from Willer et al. (2013). Effect alleles are in reference to trait elevation; the bracketed value denotes the effect size of each allele per increase in standard deviation. Variant ontology is relative to the closest gene. Abbreviations: chr = chromosome; CETP = cholesteryl ester transfer protein; HDL = high-density lipoprotein; SR-BI = scavenger receptor class B type I.

#### 2.4.4 Testing and validation of the polygenic risk score

The polygenic score was tested in the Lipid Genetics Clinic cohort and then validated in the MHI and UPenn cohorts. The score distribution in each cohort, subdivided by phenotype and rare variant carrier status are illustrated in **Figure 2.2**. Only patients without identifiable rare variants (i.e. “non-carriers”) were considered in this analysis.

In the Lipid Genetics Clinic cohorts, neither carrier group for low nor high HDL cholesterol had mean polygenic scores significantly different from the normolipidemic controls (data not shown). Compared to the mean polygenic score for normolipidemic controls ( $0.58 \pm 0.19$ ), non-carriers with low HDL cholesterol ( $0.48 \pm 0.18$ ,  $P < 0.0001$ ) and non-carriers with high HDL cholesterol ( $0.65 \pm 0.21$ ,  $P = 0.0015$ ) had significantly lower and greater mean scores, respectively (**Figure 2.2A**). In addition, 25.3% of non-carriers with low HDL cholesterol had an excess of alleles associated with lowering HDL cholesterol levels, as defined by the bottom 10<sup>th</sup> percentile of polygenic scores in the normolipidemic controls. These patients were 3.00-times (95% CI [1.67-5.35];  $P < 0.0001$ ), as likely to have extremely low polygenic scores compared to the normolipidemic controls (**Figure 2.3**). Conversely, 20.2% of non-carriers with high HDL cholesterol had an excess of alleles associated with raising HDL cholesterol levels, as defined by the top 90<sup>th</sup> percentile of polygenic scores in the normolipidemic controls. These patients were 2.19-times (95% CI [1.21-3.96];  $P = 0.006$ ), as likely to have extremely high polygenic scores compared to the normolipidemic controls (**Figure 2.4**). Patients were defined as having a more polygenic basis for their phenotype if they had an extreme polygenic score (extremely low for patients with low HDL cholesterol, and extremely high for patients with high HDL cholesterol) (**Figure 2.5**).

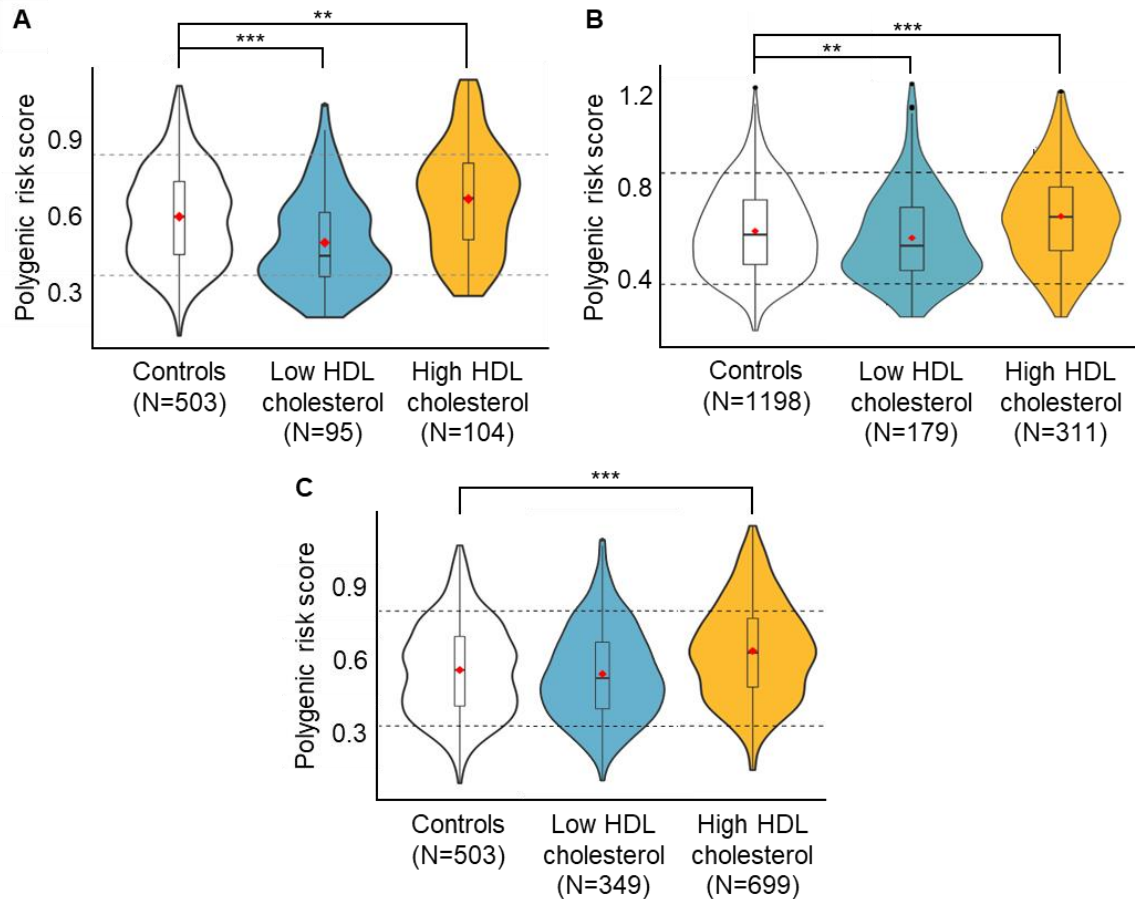
When patients were grouped by polygenic score decile, there was a strong linear relationship between increasing scores and HDL cholesterol levels (**Figure 2.6**).

Results from the MHI cohort were similar to those of the Lipid Genetics Clinic. Compared to the mean polygenic score for normolipidemic controls ( $0.58 \pm 0.19$ ), non-carriers with low HDL cholesterol ( $0.55 \pm 0.20$ ,  $P = 0.007$ ) and non-carriers with high



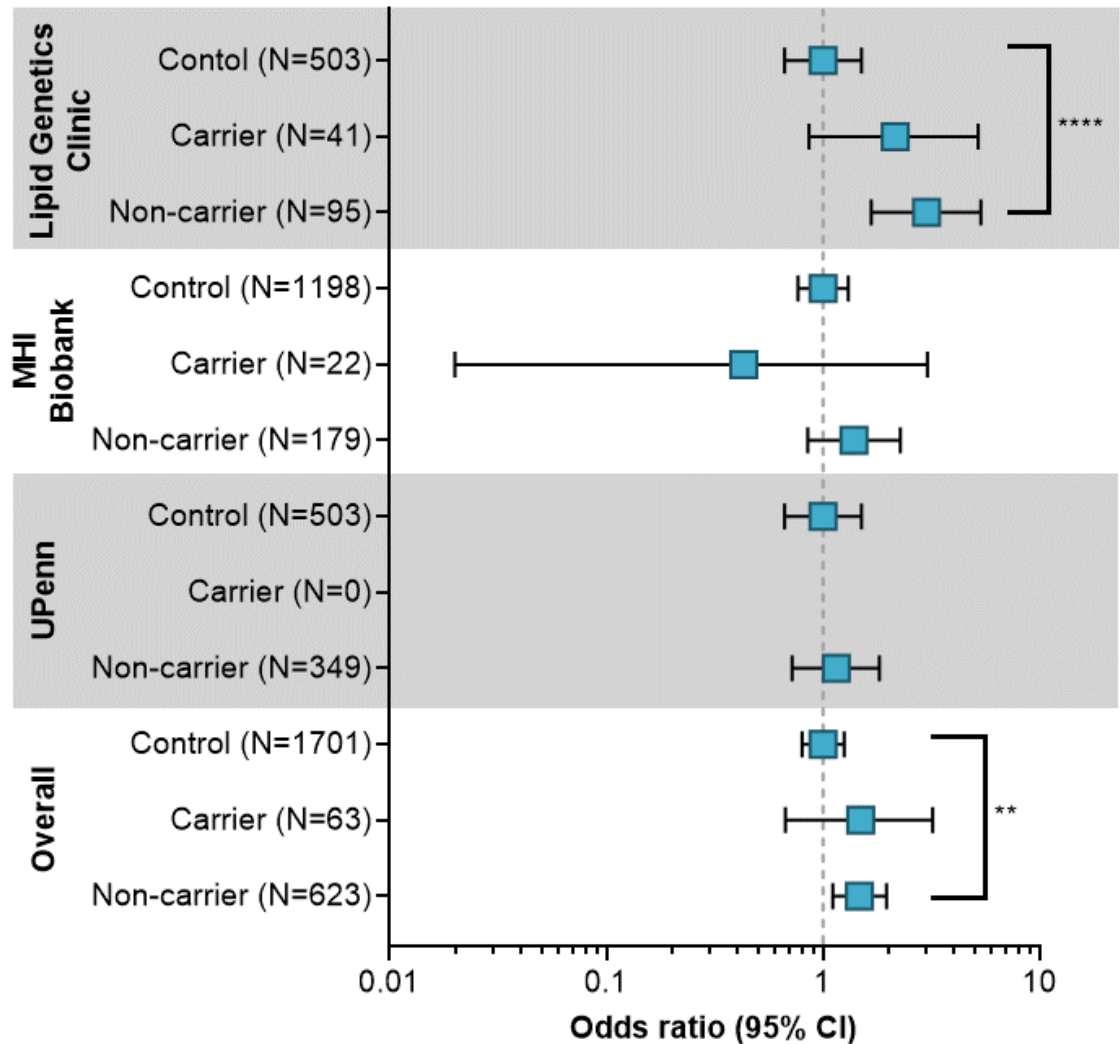
HDL cholesterol ( $0.64 \pm 0.20$ ,  $P < 0.0001$ ) had significantly lower and greater mean scores, respectively (**Figure 2.2B**). However, in contrast to the Lipid Genetics Clinic cohort, only the non-carriers with high HDL cholesterol showed a significantly increased prevalence of having extremely high polygenic scores (OR: 2.12 [95% CI: 1.48-3.02];  $P < 0.0001$ ) (**Figure 2.5**).

From the UPenn cohort, only the non-carriers with high HDL cholesterol ( $0.66 \pm 0.20$ ,  $P < 0.0001$ ) had a mean polygenic score significantly greater than the normolipidemic controls ( $0.58 \pm 0.19$ ) (**Figure 2.2C**). Similarly, only in the non-carriers with high HDL cholesterol was there was a significantly increased prevalence of having extremely high polygenic scores (OR: 2.27 [95% CI: 1.59-3.24];  $P < 0.0001$ ) (**Figure 2.6**).



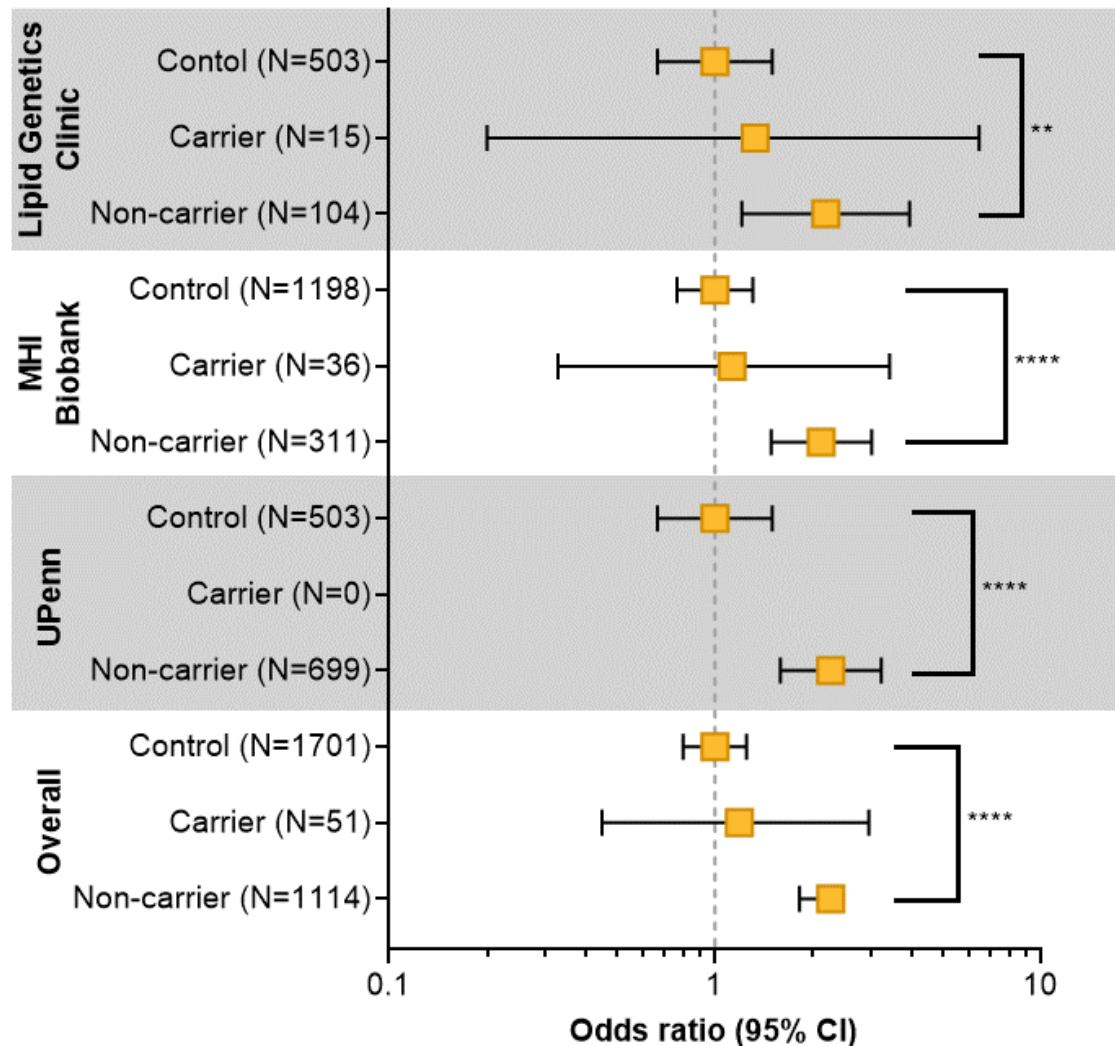
**Figure 2.2 Polygenic risk score distribution of non-carrier patients between different cohorts.**

Violin plots illustrate the distribution of polygenic risk scores in normolipidemic controls, patients with low HDL cholesterol, or patients with high HDL cholesterol in the **A**) Lipid Genetics Clinic cohort; **B**) the MHI Biobank cohort; and **C**) UPenn cohort. Red diamonds mark the mean score of the group. The top and bottom dashed lines represent the threshold for the top 90<sup>th</sup> and bottom 10<sup>th</sup> percentiles of scores in the normolipidemic controls from the 1000 Genomes Project, respectively. P-values were generated from a Kruskal-Wallis test and adjusted with Dunn's multiple comparisons based on mean polygenic risk score values between groups. P-values: \* <0.05; \*\* <0.01; \*\*\* <0.001. Abbreviations: HDL = high-density lipoprotein.



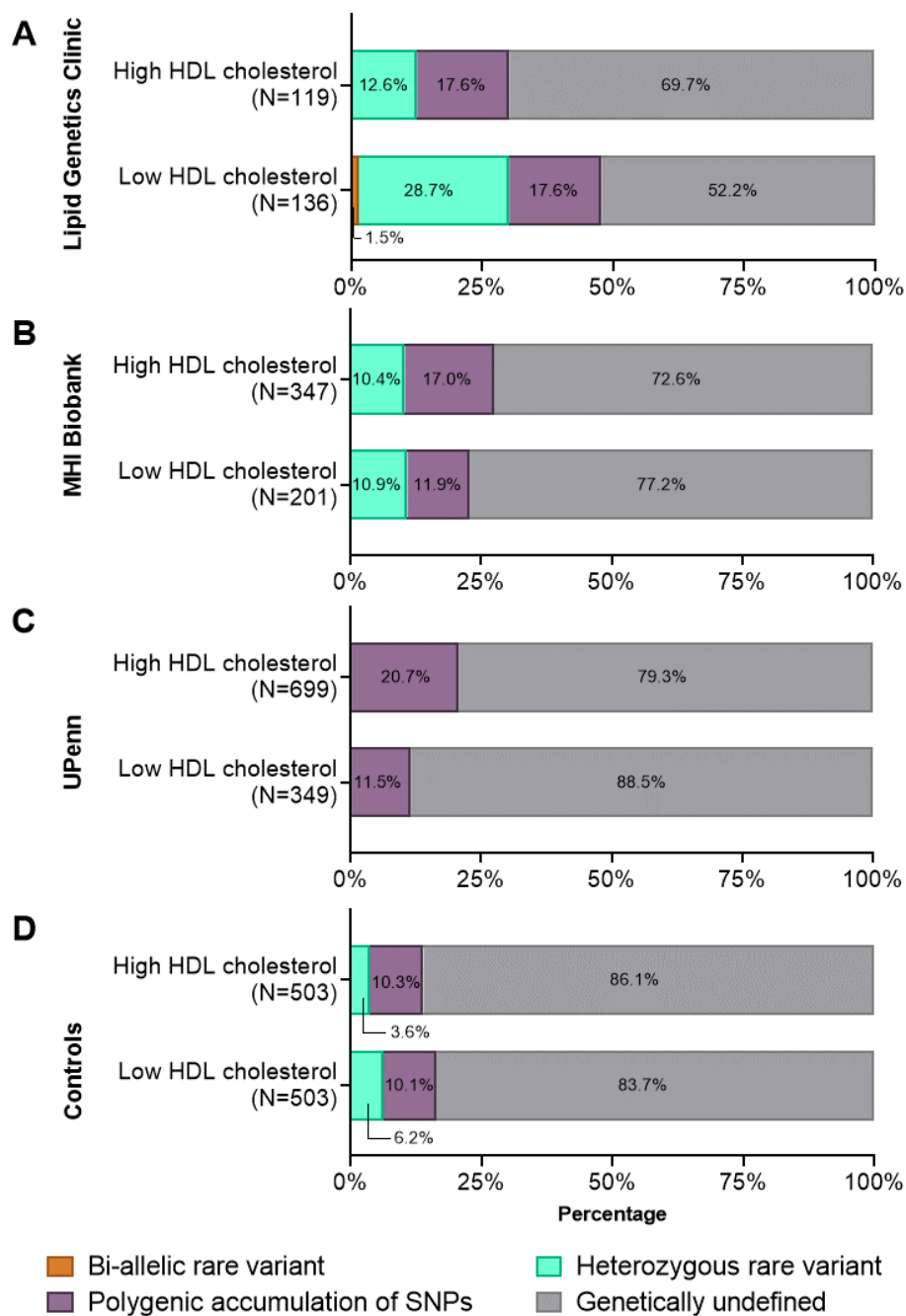
**Figure 2.3 Differences in extreme polygenic risk scores between carrier and non-carrier patients with low HDL cholesterol levels.**

This forest plot illustrates the odds ratio of carriers or non-carriers for rare variants having an extreme accumulation of common HDL cholesterol-lowering alleles (as indicated by an extremely low polygenic risk score) in patients from different cohorts, compared to normolipidemic controls. The dashed line indicates an odds ratio of 1.0. P-values were generated using a Fisher's exact tests. P-values: \*\* <0.01; \*\*\*\* <0.0001. Abbreviations: MHI = Montréal Heart Institute; UPenn = University of Pennsylvania.



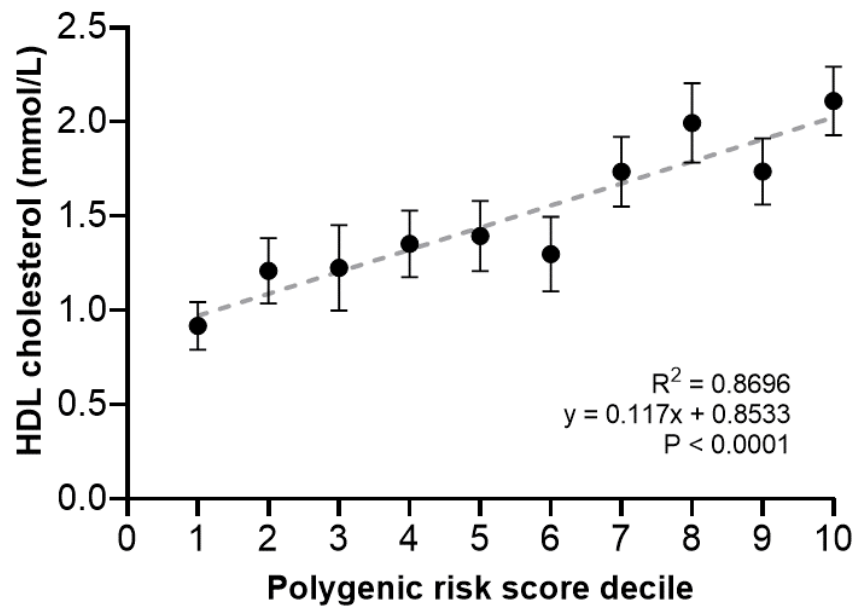
**Figure 2.4 Differences in extreme polygenic risk scores between carrier and non-carrier patients with high HDL cholesterol levels.**

This forest plot illustrates the odds ratio of carriers or non-carriers for rare variants having an extreme accumulation of common HDL cholesterol-raising alleles (as indicated by an extremely high polygenic risk score) in patients from different cohorts, compared to normolipidemic controls. The dashed line indicates an odds ratio of 1.0. P-values were generated using a Fisher's exact tests. P-values: \*\* <0.01; \*\*\*\* <0.0001. Abbreviations: MHI = Montréal Heart Institute; UPenn = University of Pennsylvania.



**Figure 2.5 The comparison of genetic profiles of extreme HDL cholesterol phenotypes between cohorts.**

Percentages were determined for individuals from the **A)** Lipid Genetics Clinic cohort; **B)** the MHI Biobank cohort; **C)** UPenn cohort; and **D)** normolipidemic controls. It must be noted that the UPenn cohort was not screened for rare variants. Abbreviations: HDL = high-density lipoprotein; MHI = Montréal Heart Institute; UPenn = University of Pennsylvania.



**Figure 2.6 Association between polygenic risk score deciles and HDL cholesterol levels.**

Mean HDL cholesterol levels for each polygenic risk score decile is shown for non-carriers from the Lipid Genetics Clinic cohort. There is a strong linear relationship between increasing polygenic scores and HDL cholesterol levels, as indicated by the  $R^2$  value of 0.8696 ( $P < 0.0001$ ). Vertical bars indicate standard error of the mean. Abbreviations: HDL = high-density lipoprotein.

## 2.5 Discussion

We report a polygenic risk score for HDL cholesterol that expands the proportion of individuals that have a relevant, identifiable genetic determinant. We first confirmed an excess of heterozygous rare variants in *ABCA1*, *LCAT* and *APOA1*, and in *CETP*, *LIPC*, *LIPG*, and *SCARB1* among individuals with extremely low and high HDL cholesterol, respectively. Overall, 18.7% and 10.9% of patients with low and high HDL cholesterol levels, were rare variant carriers, respectively. Then, among the remaining non-carriers, we showed an ~1.5- to 2-fold increased risk of having an extreme polygenic score due to an extreme accumulation of SNPs. Overall, 12.8% and 19.3% of patients with low and high HDL cholesterol levels, respectively, had an extreme polygenic score. Cumulatively, >30% of patients had either a rare variant or an extreme accumulation of SNPs associated with their respective HDL cholesterol phenotype. Our study highlights the importance of polygenic effects as determinants of extreme HDL cholesterol, and reinforces the polygenic nature of this complex trait.

From the Lipid Genetics Clinic, 47.7% and 30.2% of patients with low and high HDL cholesterol levels, respectively, had identifiable genetic factors contributing towards their phenotypes. The prevalence of rare variant carriers in the low HDL cholesterol subgroup was higher than the prevalence of rare variant carriers in the MHI cohort, perhaps reflecting ascertainment bias. Mean HDL cholesterol levels were markedly lower in the clinically ascertained low HDL cholesterol subgroup compared with the MHI and UPenn cohorts; rare variants may be more important determinants of the phenotype.

Furthermore, it appeared that when a rare variant was present, it was the main determinant of the HDL cholesterol phenotype, overriding a polygenic score favouring the opposite phenotype.

In contrast, among clinically ascertained carriers with low HDL cholesterol levels, many also had low polygenic scores. There were non-significant trends towards lower polygenic scores among non-carrier patients from MHI and UPenn. This pattern was mirrored by respective deficits of high polygenic risk scores in these cohorts (**Figure 2.3**). This demonstrates that individuals with low HDL cholesterol levels and no large-

effect variants had a more prominent polygenic contribution of small-effect variants. In the Lipid Genetics Clinic, MHI, and UPenn cohorts, among non-carriers with high HDL cholesterol levels, many had high polygenic risk scores (overall OR: 2.27 [95% CI: 1.82-2.83];  $P < 0.0001$ ). This pattern was mirrored by deficits of low polygenic risk scores in the same cohorts (**Figure 2.4**). This demonstrates that among individuals with high HDL cholesterol and no large-effect variants, there was a significant polygenic contribution from small-effect variants.

We also found that individuals carrying a rare variant and having an extreme polygenic score, both in association with the same HDL cholesterol phenotype, did not have HDL cholesterol levels that were significantly different than carriers with a normal polygenic score (data not shown). This suggests that rare variants and polygenic determinants are independent, and when present together, are not necessarily additive: rare variants appear to predominantly determine the HDL cholesterol phenotype. This contrasts with conclusions derived from a whole-genome sequence analysis of individuals with less extreme phenotypes, in whom common variants were determined to be the predominant determinants of HDL cholesterol (Morrison et al., 2013). Of course, our cohorts were still relatively small: a possible additive or synergistic relationship between rare and common variants will require evaluation in much larger samples of such extreme individuals.

Application of polygenic scores is becoming popular in the area of cardiovascular health and related complex traits (Smith et al., 2015). Mendelian randomization studies have previously evaluated these scores to infer a causal role of HDL cholesterol in CVD outcomes (Voight et al., 2012). However, until now there has been minimal to no evaluation of polygenic scores in individuals selected for extremes of HDL cholesterol levels.

Among extreme dyslipidemias, polygenic scores have been well-studied in cohorts of patients with extremely high LDL cholesterol levels, particularly FH. For instance, among clinically ascertained individuals with likely FH, 50-80% have a heterozygous rare variant in either *LDLR*, *APOB* or *PCSK9*, while another 15-20% have an extreme polygenic score comprised for LDL cholesterol (Talmud et al., 2013; Wang et al., 2016).



The exact proportions of individuals with rare and common variants differ in our cohorts with extreme HDL cholesterol levels, but the overall pattern of genetic contributors to both complex lipoprotein traits is similar. One possible difference is that syndromic FH was intentionally enriched in the extreme LDL cholesterol studies, while we excluded patients with known clinical syndromes of extreme HDL cholesterol levels.

Also, for LDL cholesterol, only individuals with extremely high levels are typically studied. In contrast, our current study assessed individuals with both extremes. The fact that our polygenic score was directionally associated with both extremes of HDL cholesterol (i.e. excessive high and low polygenic among individuals with high and low HDL phenotypes, respectively) indicates that this score applies bi-directionally for HDL cholesterol.

There may be clinical relevance in knowing the genetic basis of a patient's HDL cholesterol phenotype. For instance, in patients with high LDL cholesterol, the CVD risk compared to normolipidemic individuals was ~22-fold higher in those who carried a heterozygous rare variant versus ~6-fold higher among those who did not (Khera et al., 2016). Although polygenic effects were not evaluated, extreme LDL cholesterol in at least some individuals in the latter subgroup likely had a polygenic basis. While both groups are at high risk, having such patient-substrata can be used to generate hypotheses for different interventions under the framework of precision medicine. For instance, prospective randomized studies may show that among individuals with extremely high LDL cholesterol, carriers of a rare variant may benefit relatively more from certain treatments, such as PCSK9 inhibitors, than individuals with a stronger polygenic basis (Santos et al., 2016). By analogy, individuals with extremely low HDL cholesterol who carry a rare variant versus those who have a high polygenic burden can be studied to determine if there are differential effects of therapies targeted towards raising HDL cholesterol (Zheng et al., 2016).

This study has some limitations. First, patient ascertainment differed between the three cohorts: Lipid Genetics Clinic patients were referred because of abnormal lipid profiles, MHI Biobank participants were recruited based on cardiovascular health, and while

UPenn patients also came from lipid referrals, there was more of focus on high HDL cholesterol phenotypes. This may explain why the patients with low HDL cholesterol levels from the discovery cohort had a greater burden of rare variants: these individuals' HDL cholesterol phenotypes were more pronounced, and perhaps more likely to have a genetic basis. In contrast, since CVD was of primary interest at the MHI, abnormal HDL cholesterol profiles were less extreme and perhaps more often secondary to other, non-genetic health issues. Testing the polygenic score in other cohorts with more closely matched patient-ascertainment parameters would not only increase the power of our study, but also alleviate these biases. Second, application of the polygenic score assumes each allele has a linearly additive effect, with no epistatic interactions. Modelling epistasis could improve polygenic score accuracy and comprehension. Third, the polygenic score was tested largely in individuals of European ancestry and may not be generalizable to other ancestral groups. Also, we did not evaluate other factors—such as epigenetic regulators or large copy-number variants—as possible explanations for the extreme phenotypes. Additionally, some important variants may have been overlooked, since only genes with a known link to HDL cholesterol syndromes were screened, and only a subset of SNPs were considered; this could have led to a skew in the percentage of carriers identified or patients with an extreme SNP accumulation. Finally, given that low-pass whole-genome sequencing was used to genetically characterize participants from the MHI Biobank, it is possible that rare variants may have been missed. Despite these limitations, we have demonstrated the genetic complexity underlying extreme HDL cholesterol phenotypes by considering both rare variants and the accumulation of common SNPs simultaneously, for the first time.

## 2.6 Conclusion

In summary, we concurrently detected both rare variants and the accumulation of common SNPs using our NGS platform. In patients with both low and high HDL cholesterol extremes, we confirmed the enrichment of rare variants, while simultaneously detecting individuals with extreme polygenic scores. This substantially expanded the number of individuals with a genetic contributor towards their phenotype: about one-sixth of patients with extreme HDL cholesterol levels had an extreme polygenic score.

Loci for rare and common variants contributing to extreme HDL cholesterol levels encode products acting at all stages of the HDL lifecycle; we suggest that both rare and common variants be considered concurrently for understanding extreme HDL cholesterol levels. The large proportion of individuals still unaccounted for can be studied for additional mechanisms, such as possible new genes, gene-gene or gene-environment interactions, and non-Mendelian influences including mitochondrial or epigenetic effects. In addition to acquiring a more complete genetic picture of patients with extreme dyslipidemia, stratifying them genetically may help evaluate inter-individual differences in their clinical course or responses to interventions.

## 2.7 References

- Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7, Unit 7* 20.
- Bonnefond, A., and Froguel, P. (2015). Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metab* 21, 357-368.
- Brooks-Wilson, A., Marcil, M., Clee, S.M., Zhang, L.H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J.A., Molhuizen, H.O., *et al.* (1999). Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* 22, 336-345.
- Browne, C., and Timson, D.J. (2015). In Silico Prediction of the Effects of Mutations in the Human Mevalonate Kinase Gene: Towards a Predictive Framework for Mevalonate Kinase Deficiency. *Ann Hum Genet* 79, 451-459.
- Candini, C., Schimmel, A.W., Peter, J., Bochem, A.E., Holleboom, A.G., Vergeer, M., Dullaart, R.P., Dallinga-Thie, G.M., Hovingh, G.K., Khoo, K.L., *et al.* (2010). Identification and characterization of novel loss of function mutations in ATP-binding cassette transporter A1 in patients with low plasma high-density lipoprotein cholesterol. *Atherosclerosis* 213, 492-498.
- Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869-872.
- Dijk, W., and Kersten, S. (2014). Regulation of lipoprotein lipase by ANGPTL4. *Trends Endocrinol Metab* 25, 146-155.
- Dillio, A.A., Farhan, S.M.K., Ghani, M., Sato, C., Liang, E., Zhang, M., McIntyre, A.D., Cao, H., Racacho, L., Robinson, J.F., *et al.* (2018). Targeted Next-generation Sequencing and Bioinformatics Pipeline to Evaluate Genetic Determinants of Constitutional Disease. *J Vis Exp* 134, 57266.
- Dron, J.S., and Hegele, R.A. (2016). Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130-141.
- Edmondson, A.C., Brown, R.J., Kathiresan, S., Cupples, L.A., Demissie, S., Manning, A.K., Jensen, M.K., Rimm, E.B., Wang, J., Rodrigues, A., *et al.* (2009). Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans. *J Clin Invest* 119, 1042-1050.
- Emerging Risk Factors Consortium, Di Angelantonio, E., Gao, P., Pennells, L., Kaptoge, S., Caslake, M., Thompson, A., Butterworth, A.S., Sarwar, N., Wormser, D., *et al.* (2012). Lipid-related markers and cardiovascular disease prediction. *JAMA* 307, 2499-2506.

- Fogarty, M.P., Xiao, R., Prokunina-Olsson, L., Scott, L.J., and Mohlke, K.L. (2010). Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. *Hum Mol Genet* *19*, 1921-1929.
- Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* *526*, 68-74.
- Hegele, R.A. (2009). Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* *10*, 109-121.
- Hegele, R.A., Little, J.A., Vezina, C., Maguire, G.F., Tu, L., Wolever, T.S., Jenkins, D.J., and Connelly, P.W. (1993). Hepatic lipase deficiency. Clinical, biochemical, and molecular genetic characteristics. *Arterioscler Thromb Vasc Biol* *13*, 720-728.
- Holleboom, A.G., Kuivenhoven, J.A., Peelman, F., Schimmel, A.W., Peter, J., Defesche, J.C., Kastelein, J.J., Hovingh, G.K., Stroes, E.S., and Motazacker, M.M. (2011). High prevalence of mutations in LCAT in patients with low HDL cholesterol levels in The Netherlands: identification and characterization of eight novel mutations. *Hum Mutat* *32*, 1290-1298.
- Hovingh, G.K., Rader, D.J., and Hegele, R.A. (2015). HDL re-examined. *Curr Opin Lipidol* *26*, 127-132.
- Inazu, A., Brown, M.L., Hesler, C.B., Agellon, L.B., Koizumi, J., Takata, K., Maruhama, Y., Mabuchi, H., and Tall, A.R. (1990). Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* *323*, 1234-1238.
- Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* *55*, 765-772.
- Khera, A.V., Won, H.H., Peloso, G.M., Lawson, K.S., Bartz, T.M., Deng, X., van Leeuwen, E.M., Natarajan, P., Emdin, C.A., Bick, A.G., *et al.* (2016). Diagnostic Yield of Sequencing Familial Hypercholesterolemia Genes in Patients with Severe Hypercholesterolemia. *J Am Coll Cardiol* *67*, 2578-2589.
- Khetarpal, S.A., Schjoldager, K.T., Christoffersen, C., Raghavan, A., Edmondson, A.C., Reutter, H.M., Ahmed, B., Ouazzani, R., Peloso, G.M., Vitali, C., *et al.* (2016). Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans, Nonhuman Primates, and Rodents. *Cell Metab* *24*, 234-245.
- Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* *46*, 310-315.

Kiss, R.S., Kavaslar, N., Okuhira, K., Freeman, M.W., Walter, S., Milne, R.W., McPherson, R., and Marcel, Y.L. (2007). Genetic etiology of isolated low HDL syndrome: incidence and heterogeneity of efflux defects. *Arterioscler Thromb Vasc Biol* 27, 1139-1145.

Kuivenhoven, J.A., Stalenhoef, A.F., Hill, J.S., Demacker, P.N., Errami, A., Kastelein, J.J., and Pritchard, P.H. (1996). Two novel molecular defects in the LCAT gene are associated with fish eye disease. *Arterioscler Thromb Vasc Biol* 16, 294-303.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4, 1073-1081.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291.

Low-Kam, C., Rhainds, D., Lo, K.S., Provost, S., Mongrain, I., Dubois, A., Perreault, S., Robinson, J.F., Hegele, R.A., Dube, M.P., *et al.* (2016). Whole-genome sequencing in French Canadians from Quebec. *Hum Genet* 135, 1213-1221.

McPherson, R., and Tybjaerg-Hansen, A. (2016). Genetics of Coronary Artery Disease. *Circ Res* 118, 564-578.

Morrison, A.C., Voorman, A., Johnson, A.D., Liu, X., Yu, J., Li, A., Muzny, D., Yu, F., Rice, K., Zhu, C., *et al.* (2013). Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet* 45, 899-901.

Motazacker, M.M., Peter, J., Treskes, M., Shoulders, C.C., Kuivenhoven, J.A., and Hovingh, G.K. (2013). Evidence of a polygenic origin of extreme high-density lipoprotein cholesterol levels. *Arterioscler Thromb Vasc Biol* 33, 1521-1528.

Mucaki, E.J., Shirley, B.C., and Rogan, P.K. (2013). Prediction of mutant mRNA splice isoforms by information theory-based exon definition. *Hum Genet* 34, 557-565.

Ng, D.S., Leiter, L.A., Vezina, C., Connelly, P.W., and Hegele, R.A. (1994). Apolipoprotein A-I Q[-2]X causing isolated apolipoprotein A-I deficiency in a family with analphalipoproteinemia. *J Clin Invest* 93, 223-229.

Parish, S., Offer, A., Clarke, R., Hopewell, J.C., Hill, M.R., Otvos, J.D., Armitage, J., Collins, R., and Heart Protection Study Collaborative, G. (2012). Lipids and lipoproteins and risk of different vascular events in the MRC/BHF Heart Protection Study. *Circulation* 125, 2469-2478.

Perk, J., De Backer, G., Gohlke, H., Graham, I., Reiner, Z., Verschuren, M., Albus, C., Benlian, P., Boysen, G., Cifkova, R., *et al.* (2012). European Guidelines on cardiovascular disease prevention in clinical practice (version 2012). The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular

Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts). *Eur Heart J* 33, 1635-1701.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575.

Raffield, L.M., Cox, A.J., Hsu, F.C., Ng, M.C., Langefeld, C.D., Carr, J.J., Freedman, B.I., and Bowden, D.W. (2013). Impact of HDL genetic risk scores on coronary artery calcified plaque and mortality in individuals with type 2 diabetes from the Diabetes Heart Study. *Cardiovasc Diabetol* 12, 95.

Rifkind, B.M., and Segal, P. (1983). Lipid Research Clinics Program reference values for hyperlipidemia and hypolipidemia. *JAMA* 250, 1869-1872.

Rosenson, R.S. (2016). The High-Density Lipoprotein Puzzle: Why Classic Epidemiology, Genetic Epidemiology, and Clinical Trials Conflict? *Arterioscler Thromb Vasc Biol* 36, 777-782.

Sadananda, S.N., Foo, J.N., Toh, M.T., Cermakova, L., Trigueros-Motos, L., Chan, T., Liany, H., Collins, J.A., Gerami, S., Singaraja, R.R., *et al.* (2015). Targeted next-generation sequencing to diagnose disorders of HDL cholesterol. *J Lipid Res* 56, 1993-2001.

Santos, R.D., Gidding, S.S., Hegele, R.A., Cuchel, M.A., Barter, P.J., Watts, G.F., Baum, S.J., Catapano, A.L., Chapman, M.J., Defesche, J.C., *et al.* (2016). Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the International Atherosclerosis Society Severe Familial Hypercholesterolemia Panel. *Lancet Diabetes Endocrinol* 4, 850-861.

Schaefer, E.J., Anthanont, P., Diffenderfer, M.R., Polisecki, E., and Asztalos, B.F. (2016). Diagnosis and treatment of high density lipoprotein deficiency. *Prog Cardiovasc Dis* 59, 97-106.

Singaraja, R.R., Sivapalaratnam, S., Hovingh, K., Dube, M.P., Castro-Perez, J., Collins, H.L., Adelman, S.J., Riwanto, M., Manz, J., Hubbard, B., *et al.* (2013). The impact of partial and complete loss-of-function mutations in endothelial lipase on high-density lipoprotein levels and functionality in humans. *Circ Cardiovasc Genet* 6, 54-62.

Smith, J.A., Ware, E.B., Middha, P., Beacher, L., and Kardia, S.L. (2015). Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. *Curr Epidemiol Rep* 2, 180-190.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 133, 1-9.

- Talmud, P.J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J.A., Harrison, S.C., Li, K., Drenos, F., Karpe, F., *et al.* (2013). Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* *381*, 1293-1301.
- Tietjen, I., Hovingh, G.K., Singaraja, R.R., Radomski, C., Barhdadi, A., McEwen, J., Chan, E., Mattice, M., Legendre, A., Franchini, P.L., *et al.* (2012). Segregation of LIPG, CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol. *PLoS One* *7*, e37437.
- Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Holm, H., Ding, E.L., Johnson, T., *et al.* (2012). Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* *380*, 572-580.
- Wang, J., Dron, J.S., Ban, M.R., Robinson, J.F., McIntyre, A.D., Alazzam, M., Zhao, P.J., Dilliot, A.A., Cao, H., Huff, M.W., *et al.* (2016). Polygenic Versus Monogenic Causes of Hypercholesterolemia Ascertained Clinically. *Arterioscler Thromb Vasc Biol.* *36*, 2439-2445.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* *38*, e164.
- Weissglas-Volkov, D., and Pajukanta, P. (2010). Genetic causes of high and low serum HDL-cholesterol. *J Lipid Res* *51*, 2032-2057.
- Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* *45*, 1274-1283.
- Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., *et al.* (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* *347*, 1254806.
- Zanoni, P., Khetarpal, S.A., Larach, D.B., Hancock-Cerutti, W.F., Millar, J.S., Cuchel, M., DerOhannessian, S., Kontush, A., Surendran, P., Saleheen, D., *et al.* (2016). Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* *351*, 1166-1171.
- Zheng, K.H., van der Valk, F.M., Smits, L.P., Sandberg, M., Dasseux, J.L., Baron, R., Barbaras, R., Keyserling, C., Coolen, B.F., Nederveen, A.J., *et al.* (2016). HDL mimetic CER-001 targets atherosclerotic plaques in patients. *Atherosclerosis* *251*, 381-388.



## Chapter 3 – Large-scale deletions of the *ABCA1* gene in patients with hypoalphalipoproteinemia

The work contained in this Chapter has been edited from its original publication in the *Journal of Lipid Research* for brevity and to ensure consistency throughout this Dissertation.

JOURNAL OF LIPID RESEARCH



Supplemental Material can be found at:  
<http://www.jlr.org/content/suppl/2018/05/04/jlr.P086280.DC1>

◆ Author's Choice

patient-oriented and epidemiological research

### Large-scale deletions of the *ABCA1* gene in patients with hypoalphalipoproteinemia<sup>®</sup>

Jacqueline S. Dron,<sup>\*,2</sup> Jian Wang,<sup>\*,8</sup> Amanda J. Berberich,<sup>\*,7,8</sup> Michael A. Iacocca,<sup>\*,1</sup> Henian Cao,<sup>\*,9</sup> Ping Yang,<sup>\*,8</sup> Joan Knoll,<sup>\*\*,8</sup> Karine Tremblay,<sup>11</sup> Diane Brisson,<sup>11</sup> Christian Netzer,<sup>18</sup> Ioanna Gouni-Berthold,<sup>\*\*\*</sup> Daniel Gaudet,<sup>11</sup> and Robert A. Hegele<sup>1,\*,7,8</sup>

Robarts Research Institute,<sup>\*,8</sup> Department of Biochemistry,<sup>1</sup> Department of Medicine,<sup>3</sup> and Department of Pathology and Laboratory Medicine,<sup>\*\*</sup> Schulich School of Medicine and Dentistry, Western University, London ON, Canada; <sup>7</sup> Lipidology Unit,<sup>11</sup> Community Genomic Medicine Centre and ECOGENE-21, Department of Medicine, Université de Montréal, Saguenay QC, Canada; Institute for Human Genetics,<sup>18</sup> and Polyclinic for Endocrinology, Diabetes and Preventive Medicine,<sup>\*\*\*</sup> University of Cologne, Germany

ORCID IDs: 0000-0002-9045-6530 (J.S.D.); 0000-0003-2861-5325 (R.A.H.)

ASBMB

**Abstract** Copy-number variations (CNVs) have been studied in the context of familial hypercholesterolemia but have not yet been evaluated in patients with extreme levels of HDL cholesterol. We evaluated targeted, next-generation sequencing data from patients with very low levels of HDL cholesterol (i.e., hypoalphalipoproteinemia) with the VarSeq-CNV<sup>®</sup> caller algorithm to screen for CNVs that disrupted the *ABCA1*, *LCAT*, or *APOA1* genes. In four individuals, we found three unique deletions in *ABCA1*: a heterozygous deletion of exon 4, a heterozygous deletion that spanned exons 8 to 31, and a heterozygous deletion of the entire *ABCA1* gene. Breakpoints were identified with Sanger sequencing, and the full-gene deletion was confirmed by using exome sequencing and the Affymetrix CytoScan HD array.<sup>®</sup> Previously, large-scale deletions in candidate HDL genes had not been associated with hypoalphalipoproteinemia; our findings indicate that CNVs in *ABCA1* may be a previously unappreciated genetic determinant of low levels of HDL cholesterol. By coupling bioinformatic analyses with next-generation sequencing data, we can successfully assess the spectrum of genetic determinants of many dyslipidemias, including hypoalphalipoproteinemia.—Dron, J.S., J. Wang, A. J. Berberich, M. A. Iacocca, H. Cao, P. Yang, J. Knoll, K. Tremblay, D. Brisson, C. Netzer, I. Gouni-Berthold, D. Gaudet, and R. A. Hegele. Large-scale deletions of the *ABCA1* gene in patients with hypoalphalipoproteinemia. *J. Lipid Res.* 2018. 59: 1529–1535.

**Supplementary key words:** ATP-binding cassette subfamily A member 1 • bioinformatic analysis • copy-number variation • high density lipoprotein cholesterol • next-generation sequencing • diagnostic tools • genetic testing • dyslipidemia

Extremely low levels of HDL cholesterol, clinically characterized as “hypoalphalipoproteinemia”, can result from various molecular etiologies. DNA sequencing of candidate genes has shown that between ~10–35% of affected individuals have rare heterozygous missense, nonsense, or splicing variants in *ABCA1*, *APOA1*, and *LCAT* genes, encoding ABCA1, apo A-I, and lecithin:cholesterol acyl transferase, respectively (1–6). We recently found that another ~18% of affected individuals have an extreme polygenic accumulation of common variants, as quantified by a polygenic trait score that considers several common SNPs associated with HDL cholesterol levels (1). However, the genetic basis of low HDL cholesterol in the majority of individuals with hypoalphalipoproteinemia remains to be characterized.

Copy-number variations (CNVs) are deletions and duplications of genomic material that are much larger than single nucleotide variations (SNVs); by convention, “CNVs”

R.A.H. is supported by the Jacob J. Wolfe Distinguished Medical Research Chair, the Edith Schuchk Visot Research Chair in Human Genetics, and the Martha G. Blackburn Chair in Cardiovascular Research. R.A.H. has also received operating grants from the Canadian Institutes of Health Research (Foundation award), the Heart and Stroke Foundation of Ontario (G15-0009214), and Genome Canada through Genome Quebec (award 4530). R.A.H. is a consultant and speaker's bureau member for Amgen, Angen, Boston Heart Diagnostics, Genentech, Eli Lilly, Genpharm, Pfizer and Sanofi. J.C.R. is a consultant and speaker's bureau member for Amgen, Angen, Sanofi, Regeneron and Alkermes. H.C. received research grant support from FH Canada, Canadian Cardiovascular Research Network, Apoliron (Novartis Biotech), Alkermes, Amgen, AstraZeneca, Genentech, Dalcor Pharma, Esperion, Genpharm, GlaxoSmithKline, Kowa, HDL Therapeutics, Jans, Institut de cardiologie de Montréal, Pfizer, Regeneron, The Medicine Company, UroDare and served as a consultant for Amgen, Alkermes, HDL Therapeutics, Novartis, Regeneron, Sanofi. The other authors have no disclosures.

<sup>®</sup> Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

Manuscript received 23 April 2018 and in revised form 21 May 2018.

Published, JLR Papers in Press, June 4, 2018.

DOI: <https://doi.org/10.1194/jlr.P086280>

Copyright © 2018 Dron et al. Published by The American Society for Biochemistry and Molecular Biology, Inc.

This article is available online at <http://www.jlr.org>

JOURNAL OF LIPID RESEARCH

Journal of Lipid Research Volume 59, 2018 1529

Downloaded from www.jlr.org by guest on October 14, 2020

**Dron, J.S., Wang, J., Berberich, A.J., Iacocca, M.A., Cao, H., Yang, P., Knoll, J., Tremblay, K., Brisson, D., Netzer, C., et al. (2018). Large-scale deletions of the *ABCA1* gene in patients with hypoalphalipoproteinemia. *J Lipid Res* 59, 1529-1535.**

### 3.1 Abstract

**Objective:** Copy-number variants (CNVs) have been studied in the context of familial hypercholesterolemia but have not yet been evaluated in patients with extreme levels of high-density lipoprotein (HDL) cholesterol.

**Methods and Results:** We evaluated targeted next-generation sequencing (NGS) data from patients with extremely low levels of HDL cholesterol (i.e., hypoalphalipoproteinemia) with the VarSeq-CNV<sup>®</sup> caller algorithm to screen for CNVs disrupting the *ABCA1*, *LCAT*, or *APOA1* genes. In four individuals, we found three unique deletions in *ABCA1*: a heterozygous deletion of exon 4, a heterozygous deletion that spanned exons 8 to 31, and a heterozygous deletion of the entire *ABCA1* gene. Breakpoints were identified with Sanger sequencing, and the full-gene deletion was confirmed using exome sequencing and the Affymetrix CytoScan HD array.

**Conclusion:** Previously, large-scale deletions in candidate HDL genes had not been associated with hypoalphalipoproteinemia; our findings indicate that CNVs in *ABCA1* may be a previously unappreciated genetic determinant of low levels of HDL cholesterol. By coupling bioinformatic analyses with NGS data, we can successfully assess the spectrum of genetic determinants of many dyslipidemias, including hypoalphalipoproteinemia.

### 3.2 Introduction

Extremely low levels of high-density lipoprotein (HDL) cholesterol, clinically characterized as “hypoalphalipoproteinemia”, can result from various molecular etiologies. DNA sequencing of candidate genes has shown that between ~10-35% of affected individuals have rare heterozygous missense, nonsense or splicing variants in *ABCA1*, *APOA1* and *LCAT* genes, encoding ATP-binding cassette subfamily member A1 (ABCA1), apolipoprotein (apo) A-I and lecithin cholesterol acyl transferase (LCAT), respectively (Candini et al., 2010; Cohen et al., 2004; Dron et al., 2017; Holleboom et al., 2011; Kiss et al., 2007; Sadananda et al., 2015). We recently found that another ~18% of affected individuals have an extreme polygenic accumulation of common variants, as

quantified by a polygenic risk score that considers several common single-nucleotide polymorphisms (SNPs) associated with HDL cholesterol levels (Dron et al., 2017). However, the genetic basis of low HDL cholesterol in the majority of individuals with hypoalphalipoproteinemia remains to be characterized.

Copy-number variants (CNVs) are deletions and duplications of genomic material that are much larger than single-nucleotide variations (SNVs); by convention, “CNVs” are deletions or duplications >50 bp in size (Zarrei et al., 2015). While CNVs have been commonly identified throughout the genome, there has been a surging focus on CNVs that are rare within the population, and their relationship to certain phenotypes and diseases (Iacocca and Hegele, 2018). This redefined focus has been due to improvements in bioinformatic tools, and targeted next-generation sequencing (NGS) panels designed for clinical utility. Previously, specialized molecular methods, such as multiplex ligation-dependent probe amplification (MLPA), have been required to detect CNVs, and had to be performed concurrently to other genetic methods. Now, through the development of new bioinformatic methods, CNVs can be easily screened for in patient groups using data generated by a single genetic approach, namely, NGS. We recently reported that data generated with a targeted NGS panel designed to detect SNVs in genes related to familial hypercholesterolemia (FH) could be processed with dedicated bioinformatic tools to diagnose the presence of CNVs in *LDLR*, encoding the low-density lipoprotein (LDL) receptor. Results of our NGS-based CNV detection method showed 100% concordance with traditional MLPA of *LDLR*, with no false negative or false positive results (Iacocca et al., 2017).

CNVs disrupting *ABCA1*, *APOA1*, or *LCAT* in individuals with hypoalphalipoproteinemia have not yet been reported. Here, we applied our novel bioinformatic approach on previously generated targeted NGS data from patients with hypoalphalipoproteinemia, with particular interest in patients without rare variants in HDL-associated genes or without an extreme polygenic accumulation of common variants (Dron and Hegele, 2018). Out of 288 patients screened, four carried one of three novel heterozygous CNVs within the *ABCA1* gene; the variants were confirmed using independent methods. Our findings not only demonstrate the usefulness of applying

bioinformatically-based CNV calling algorithms to NGS data, but we also provide the first example of large-scale CNV deletions that may be contributing towards the hypoalphalipoproteinemia phenotype.

### 3.3 Materials and Methods

#### 3.3.1 Study subjects

Patients who were referred to the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada) for “low HDL cholesterol” or “hypoalphalipoproteinemia” were considered for this screening study. Patients provided signed consent with approval from the Western University ethics review board (no. 07290E).

#### 3.3.2 DNA preparation and targeted sequencing

DNA isolation and preparation for targeted sequencing follows the same methodology as described in [Chapter 2, Section 2.3.2](#).

#### 3.3.3 Bioinformatic processing of sequencing data

The bioinformatic processing of sequencing data follows the same methodology as described in [Chapter 2, Section 2.3.3](#); however, an updated version of CLC Bio Genomics Workbench (version 8.5; CLC Bio, Aarhus, Denmark) was used. In addition to CLC Bio Genomics Workbench generating VCF files containing variant information for each patient, depth of coverage for a patient’s sequencing data was also exported as a BAM file.

#### 3.3.4 Detection of single-nucleotide and copy-number variants

The BAM and VCF files generated for each patient were imported into VarSeq® (version 1.4.8; Golden Helix, Inc., Bozeman MT, USA) for annotation of each genetic variant. SNVs were identified following methods described in [Chapter 2, Section 2.3.4](#). Assessment of CNVs in *ABCA1*, *APOA1*, and *LCAT* was performed using the VarSeq-CNV® caller algorithm. To identify CNVs, the depth-of-coverage information contained within each subject’s BAM file was compared against the average coverage information

from a set of samples that have been confirmed to not carry CNVs. Increases and decreases in read-depth indicate a duplication or deletion of genetic material, respectively. The exact criteria used to identify CNVs has been previously described (Iacocca et al., 2017).

### 3.3.5 Validation of partial gene deletions

#### 3.3.5.1 Breakpoint identification

To identify the presence of partial gene deletions, primers were designed to flank regions surrounding the putative deletions and were used for PCR amplification (Expand 20 kb<sup>plus</sup> PCR System, Roche, Mannheim, Germany) (*Cat No. 11811002001*). Forward (F) and reverse (R) primers flanking the deletion junctions were: F1 5'-AGCACGATAGGAAGCATCTTC-3' and R1 5'-ATCACTGTCTGTGGCAACCAG-3' (exon 4 deletion); F2 5'-GACCCAGCTTCCAATCTTCATAA-3' and R2 5'-TAGACAGAATCAGGCCATAATCTG-3' (exons 8-31 deletion). Gel electrophoresis of the PCR products was used as a visual confirmation of the mutant alleles. Sanger sequencing and primer-walking of the PCR products were performed to identify the deletion breakpoints.

#### 3.3.5.2 Sanger confirmation

Once deletion breakpoints were identified, screening primers spanning the upstream or downstream breakpoint were designed for PCR and Sanger sequencing (**Appendix E**) to confirm the deletion breakpoint sequences for the wild-type or deleted alleles.

### 3.3.6 Validation of full gene deletions

#### 3.3.6.1 Exome sequencing

Patients with expected full-gene deletions had their DNA samples indexed and pooled using the TruSeq Rapid Exome Kit (Illumina, San Diego CA, USA) (*Cat No. 20020616*) in preparation for exome sequencing. Sequencing was then performed at the London Regional Genomics Centre ([www.lrgc.ca](http://www.lrgc.ca); London ON, Canada), using a NextSeq 500 (Illumina, San Diego CA, USA). The same bioinformatic approach described above was used to replicate the CNV call made by the VarSeq-CNV<sup>®</sup> caller algorithm.

### 3.3.6.2 Microarray analysis

Patients with expected full-gene deletions had their DNA samples assessed with the Affymetrix CytoScan™ HD Array (Thermo Fisher Scientific, Waltham MA, USA) for the genomic region containing the CNV. With >2 million probes on the array, deletions >25 kb can be detected. The microarray was performed following the manufacturer's instructions at Victoria Hospital (London ON, Canada), and the resultant data were analyzed using the Chromosome Analysis Suite (version 3.2; Thermo Fisher Scientific, Waltham MA, USA). The regions between adjacent probes that differed in copy-number state were marked as containing the approximate breakpoints of the CNV and were used to gauge the approximate size of the deletion.

### 3.3.6.3 Breakpoint identification

Once establishing the magnitude of the deletion, the approximate locations of each breakpoint were estimated. Primers flanking the deletion junction were: F3 5'-CCTGGCTGCTTCTAAGAGCCTATGATC-3' and R3 5'-TGTCTCTACATGGTCCTCCTTCTGTGC-3', and were used for PCR amplification (Expand 20 kb<sup>plus</sup> PCR System, Roche, Mannheim, Germany) (*Cat No. 11811002001*). Gel electrophoresis of the PCR products was used as a visual confirmation of the mutant allele. Sanger sequencing and primer-walking of the PCR product were performed to identify the deletion breakpoints.

### 3.3.6.4 Sanger confirmation

Once deletion breakpoints were identified, screening primers spanning the upstream or downstream breakpoint were designed for PCR and Sanger sequencing (**Appendix E**) to confirm the deletion breakpoint sequences for the wild-type or deleted allele.

## 3.4 Results

### 3.4.1 Study subjects

A total of 288 patients with “low HDL cholesterol” or “hypoalphalipoproteinemia” were sequenced with LipidSeq and screened for CNVs disrupting *ABCA1*, *APOA1*, and *LCAT*.

Clinical and biochemical characteristics of the four patients identified as carriers for CNVs are shown in **Table 3.1**.

### 3.4.2 *ABCA1* copy-number variant detection

Analysis of LipidSeq output with the VarSeq-CNV® caller algorithm identified four hypoalphalipoproteinemia patients as carriers of large-scale deletions in *ABCA1* (**Figure 3.1**). Patient 1 had a heterozygous deletion spanning exon 4; Patient 2 and Patient 3, a pair of siblings had a heterozygous deletion spanning exons 8 to 31; and Patient 4 had a heterozygous deletion spanning the entire *ABCA1* gene. None of these patients carried rare SNVs in *ABCA1*, *APOA1* or *LCAT*. There were no CNVs detected in *APOA1* or *LCAT* for any patients in this study.

To determine the size of the deletion in Patient 4, the VarSeq-CNV® caller algorithm was used on exome data to confirm the heterozygous absence of *ABCA1* (**Figure 3.2**), while the CytoScan™ analysis confirmed and replicated the heterozygous nature of this CNV (**Figure 3.3A**). Exome sequencing and Cytoscan™ revealed that the CNV was ~2 Mb in length, and encompassed six additional protein-coding genes, including *SMC2*, *NIPSNAP3A*, *NIPSNAP3B*, *SLC44A1*, *FSDIL*, and *FKTN*.

### 3.4.3 Copy-number variant validation and identifying breakpoints

Sanger sequencing across the CNV breakpoints in Patient 1 (**Figure 3.4A**), Patients 2 and 3 (**Figure 3.4B**), and Patient 4 (**Figure 3.3B**) revealed the genomic coordinates involved in the deletion event and allowed us to determine the exact size of the CNV (**Table 3.2**). Screening primers spanning breakpoints were used to distinguish between wild-type and deleted alleles, as indicated in **Figure 3.3C**, **Figure 3.4C** and **Figure 3.4D**.

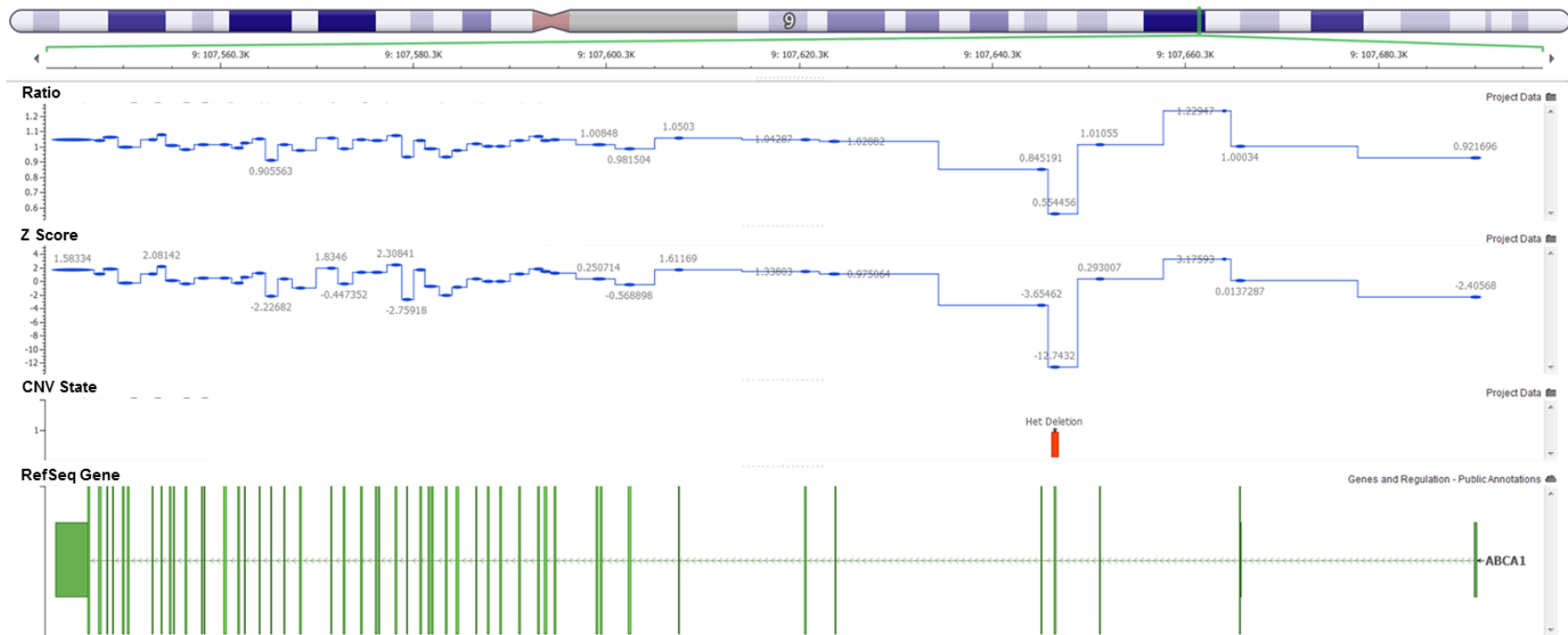
**Table 3.1 Clinical and demographic features of subjects with *ABCA1* CNVs.**

	<b>Patient 1</b>	<b>Patient 2</b>	<b>Patient 3</b>	<b>Patient 4</b>
<b>Age</b>	37	34	59	40
<b>Sex</b>	Female	Female	Male	Female
<b>BMI (kg/m<sup>2</sup>)</b>	23.5	29.2	-	24.1
<b>Ancestry</b>	European	European	European	European
<b>Total cholesterol (mmol/L)</b>	8.16	3.30	5.46	4.71
<b>Triglyceride (mmol/L)</b>	2.52	1.01	4.48	5.13
<b>HDL cholesterol (mmol/L)</b>	0.81	0.56	0.47	0.03
<b>LDL cholesterol (mmol/L)</b>	6.20	2.42	3.28	3.54
<b>apo A-I (g/L)</b>	-	0.59	0.60	0.09
<b>apo B (g/L)</b>	-	0.81	1.33	-
<b>Creatine kinase (U/L)</b>	79	-	78	113
<b>Fasting glucose (mmol/L)</b>	4.0	5.4	5.3	4.7
<b>Lp(a) (nmol/L)</b>	-	-	290	363
<b>Co-morbidities</b>	Heterozygous FH ( <i>LDLR</i> NM_000527 p.V523M); minor carotid intimal thickening	Obesity	Hypertension; TIA; aortic valvular stenosis	TIA; cerebral arteriosclerotic microangiopathy; hypertension; juvenile myoclonic epilepsy; diffuse non-Hodgkin's lymphoma stage III

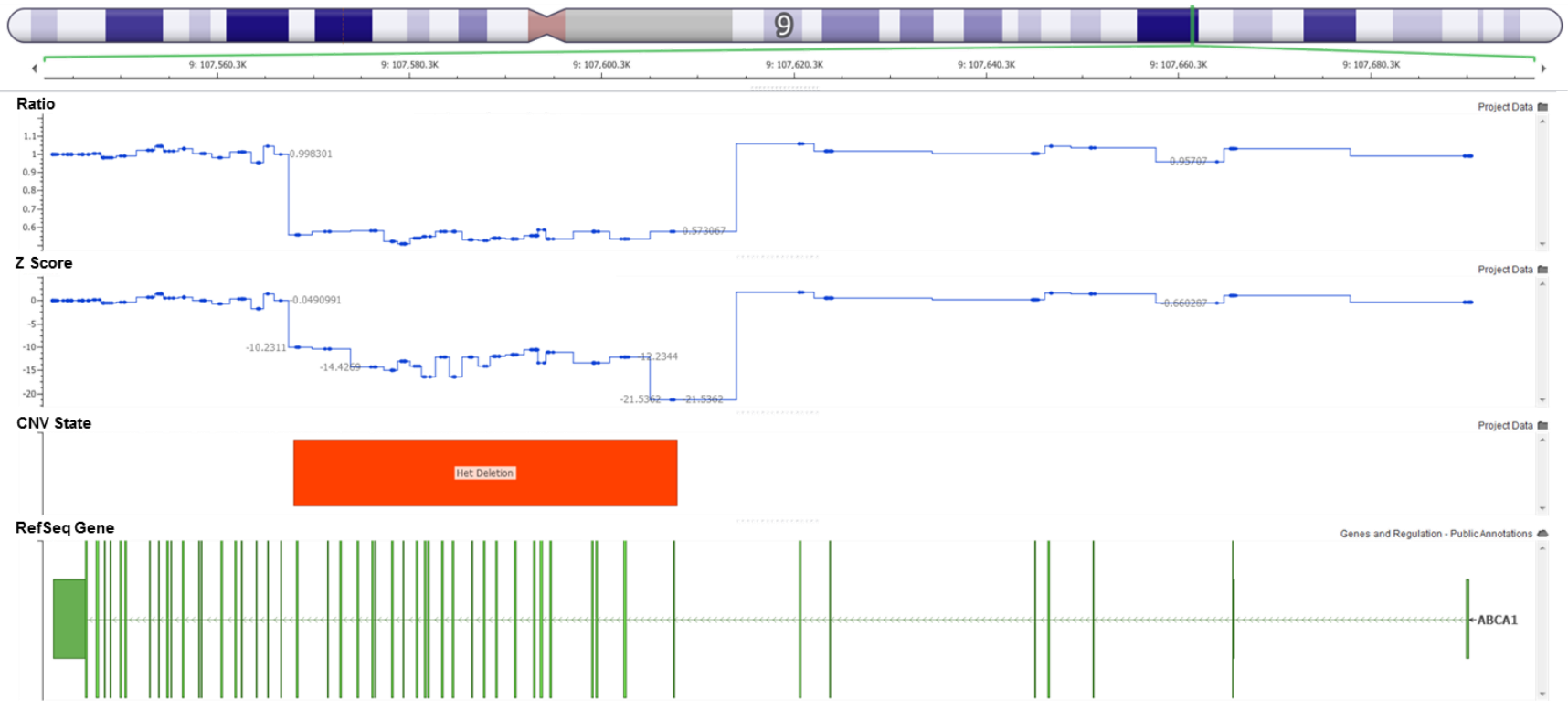
Values provided are from first presentation to specialist lipid clinic, or date first obtained. Lp(a) conversions from g/L to nmol/L were done following the conversion factor described by Brown WV et al. (2010). Abbreviations: apo = apolipoprotein; BMI = body-mass index; FH = familial hypercholesterolemia; HDL = high-density lipoprotein; LDL = low-density lipoprotein; Lp(a) = lipoprotein(a); TIA = transient ischemic attack.



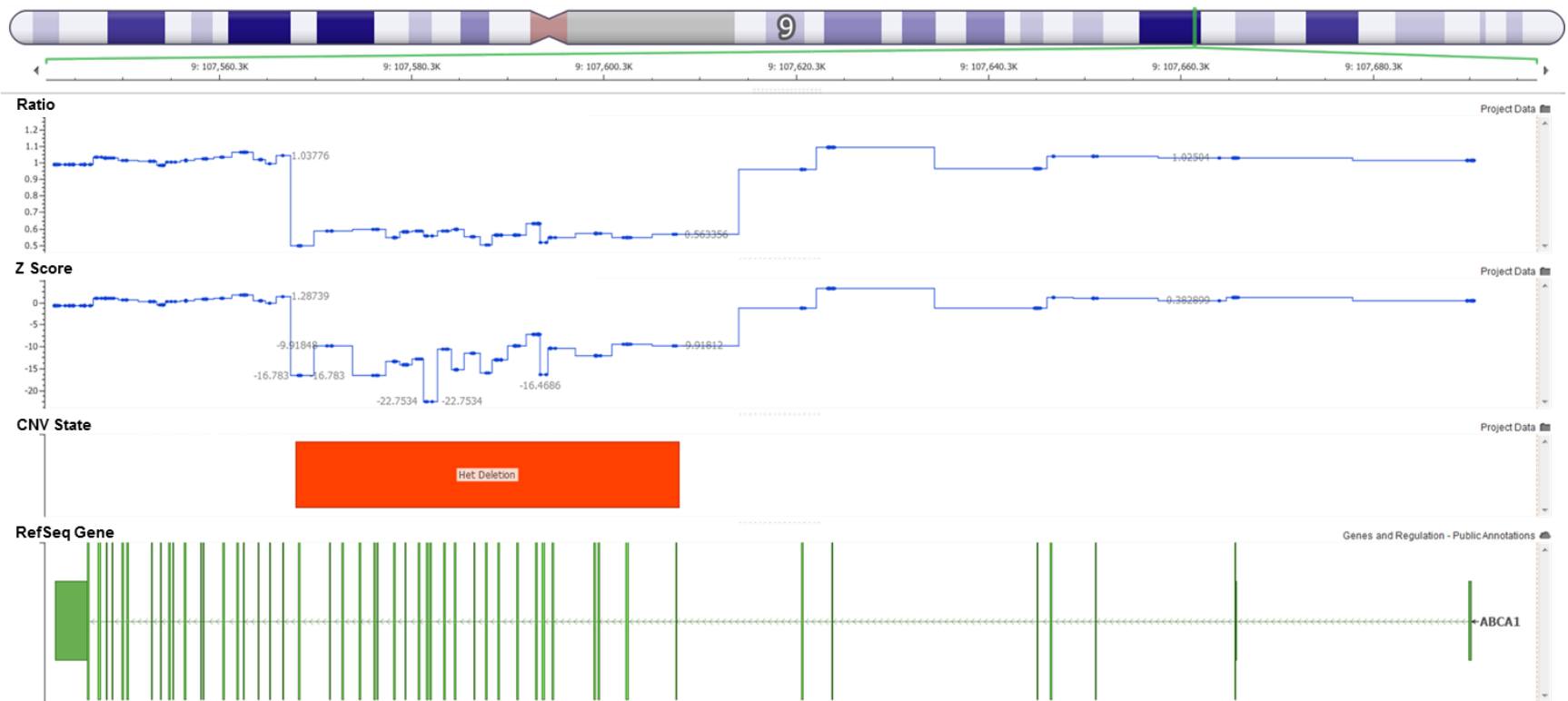
A



B



C



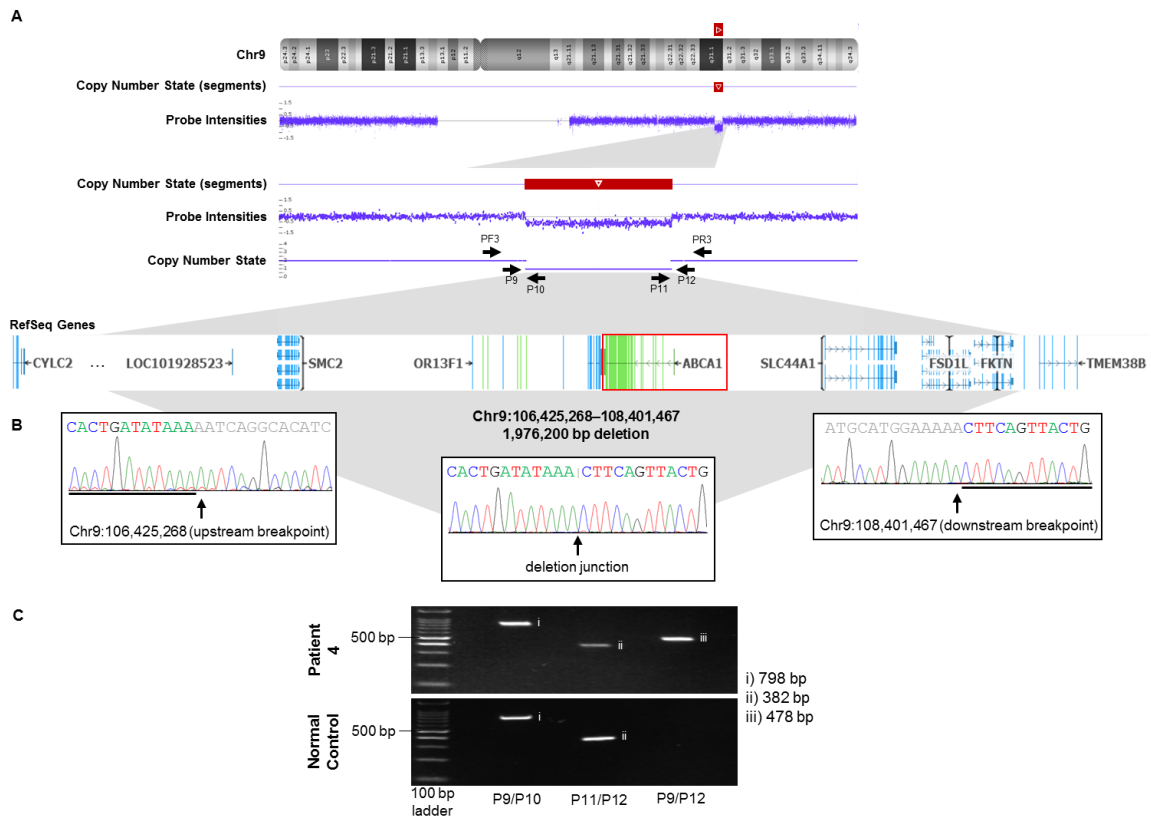


**Figure 3.1 Identification of *ABCA1* CNVs using the VarSeq-CNV® caller algorithm on targeted sequencing data.**

Chr9:107,542,273–107,697,356 (hg19 genome build) is the region visualized in each panel, with the CNV “ratio”, “Z score”, and “state” available for each subject. **A)** Subject 1, carrier of a heterozygous deletion of exon 4. **B)** Subject 2, carrier of a heterozygous deletion spanning exons 8 to 31. **C)** Subject 3, carrier of a heterozygous deletion spanning exons 8 to 31. **D)** Subject 4, carrier of a heterozygous deletion of the entire *ABCA1* gene. Abbreviations: CNV = copy-number variant.

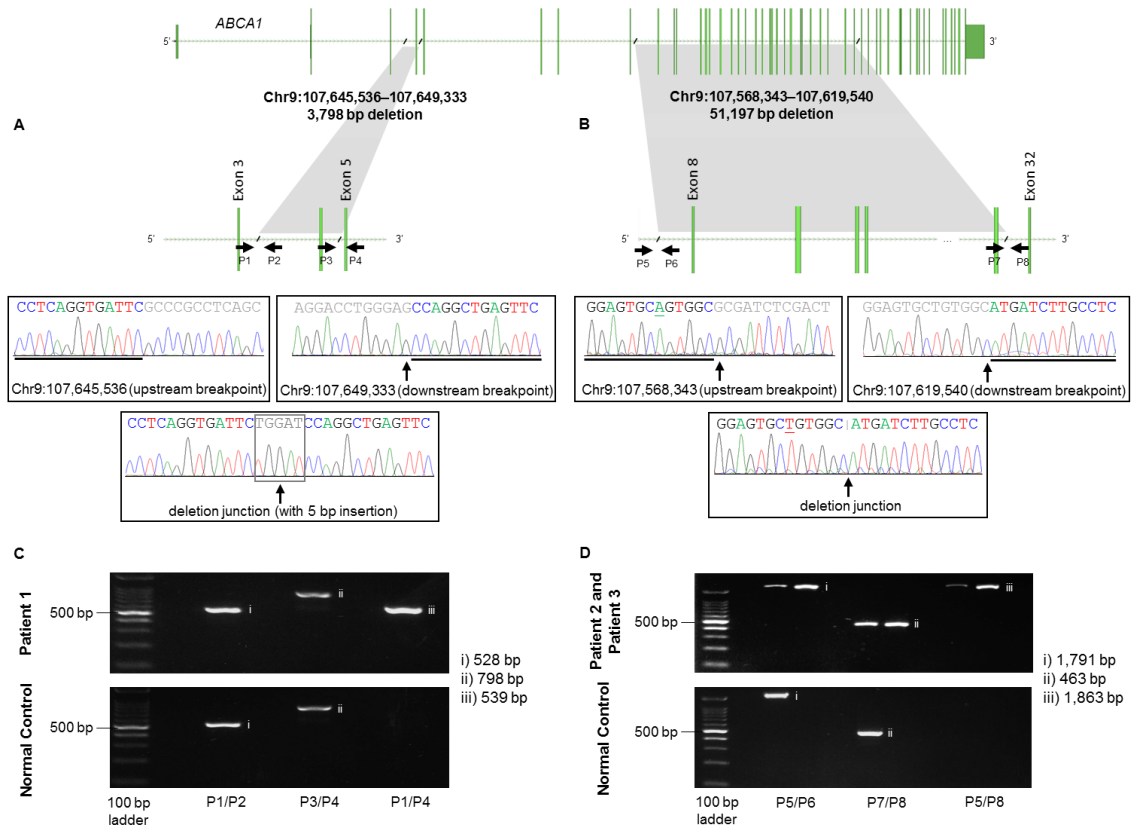


**Figure 3.2 Confirmation of the full-gene *ABCA1* CNV using the VarSeq-CNV® caller algorithm on exome data.** Chr9:105,295,869–109,769,141 (hg19 genome build) is the region visualized, with the CNV “ratio”, “Z score”, and “state” available for the subject. Subject 4, carrier of a heterozygous deletion of the entire *ABCA1* gene and surrounding loci. Abbreviations: CNV = copy-number variation.



**Figure 3.3 Validation of full-gene deletion of *ABCA1* in Patient 4 with hypoalphalipoproteinemia.**

**A)** Results of the CytoScan™ HD Array, visualized using Chromosome Analysis Suite “Copy Number State (segments)” identifies the region containing the CNV. “Probe Intensities” show a drop in signal, indicating a decrease in copy number at that position, evident under “Copy Number State”. The black arrows demonstrate the position and orientation of primers used in breakpoint identification and Sanger sequencing. The genes, both coding and non-coding, encompassed by the deletion are evident under “RefSeq Genes”; the image was taken and modified from and VarSeq®. **B)** Sanger sequencing results for the forward strand across upstream and downstream breakpoints, and the deletion junction. **C)** Gel electrophoresis of PCR products across upstream and downstream breakpoints, and deletion junction. Results from Patient 4 are presented on the top, with results from a normal control on the bottom. Lane 1 contains 100bp ladder, lane 2 contains products across the upstream breakpoint, lane 3 contains products across the downstream breakpoint, and lane 4 contains products across the deletion junction. Abbreviations: bp = base pair; chr = chromosome; F = forward strand; P = primer; R = reverse strand.



**Figure 3.4 Validation of partial gene deletions of *ABCA1* in Patients 1, 2, and 3 with hypolipoproteinemia.**

Sanger sequencing results for the reverse strand across upstream and downstream breakpoints, and the deletion junctions for **A**) Patient 1 and **B**) Patients 2 and 3. Underlined bases represent polymorphic sites between subjects. The black slashes indicate the sequence breakpoints, while the arrows demonstrate the position and orientation of primers used in breakpoint identification and Sanger sequencing. The gene transcript image was taken and modified from and VarSeq®. Gel electrophoresis of PCR products across upstream and downstream breakpoints, and deletion junction for **C**) Patient 1 and **D**) Patients 2 and 3. Results from each patient are presented on the top, with results from a normal control on the bottom. Lane 1 contains 100bp ladder, lane 2 contains products across the upstream breakpoint, lane 3 contains products across the downstream breakpoint, and lane 4 contains products across the deletion junction. Abbreviations: bp = base pair; chr = chromosome; P = primer.

**Table 3.2 Genomic coordinates and breakpoints of *ABCA1* CNVs.**

CNV	Zygoty state	Breakpoint		
		Genomic coordinates	Length (bp)	HGVS notation
Exon 4	Het	chr9:107645536 to chr9:107649333	3,798	g.107645536_107649333delinsATCCA c.160_301del p.Cys54LeufsTer22
Exons 8 to 31	Het	chr9:107568343 to chr9:107619540	51,197	g.107568343_107619540del c.720_4463del p.Arg241_Gln1488del
Full deletion	Het	chr9:106425268 to chr9:108401467	1,976,200	g.106425268_108401467del

The sequences are in the forward-strand orientation, with genomic coordinates based on the hg19 human genome reference build. Abbreviations: bp = base pair; chr = chromosome; CNV = copy-number variation; het = heterozygous; HGVS = Human Genome Variation Society.



### 3.5 Discussion

In 288 patients with hypoalphalipoproteinemia, we identified three rare, large-scale deletions in *ABCA1* in four individuals by applying specialized bioinformatic tools to NGS data. While it is not the first time CNVs have been observed in *ABCA1* (Abecasis et al., 2012; Ahn et al., 2009; Alsmadi et al., 2014; Boomsma et al., 2014; Conrad et al., 2010; Cooper et al., 2011; Itsara et al., 2009; Kidd et al., 2008; Mills et al., 2011; Park et al., 2010; Shaikh et al., 2009; Suktitipat et al., 2014; Teague et al., 2010; Tuzun et al., 2005; Wong et al., 2013), it is the first report of *ABCA1* CNVs being found specifically in patients with hypoalphalipoproteinemia, and may be large contributors towards the low HDL cholesterol phenotype.

*ABCA1* is a critical player in the reverse cholesterol transport pathway. Found on the surface of macrophages, *ABCA1* mediates the transport of free cholesterol out of the cell, where it can be picked up by apo A-I, leading to the generation of nascent HDL particles (Lewis and Rader, 2005). Disruptions to this protein can alter its function and lead to problems with cholesterol efflux and the generation of circulating HDL particles. Rare homozygous variants in this gene have been shown to cause Tangier Disease (Bodzioch et al., 1999; Brooks-Wilson et al., 1999; Rust et al., 1999), while heterozygous mutations can lead to less severe forms of hypoalphalipoproteinemia (Brooks-Wilson et al., 1999; Marcil et al., 1999). Given the sizes of our identified CNVs and their predicted consequences on the protein product, they are likely loss-of-function, leading to a decrease in the generation of HDL particles and an overall decrease in circulating HDL cholesterol.

The smallest CNV deletion is 3,798 bp in size, with its breakpoints in introns 3 and 4, causing a partial loss of both introns, and a full loss of exon 4. The deletion of the coding sequence caused a frameshift and a premature truncation of the protein at the 76<sup>th</sup> amino acid: 96.7% of the protein is lost. Since our study is limited in that we did not test mRNA levels, protein levels, or protein function, we cannot comment on the exact mechanism by which this *ABCA1* CNV leads to low HDL cholesterol levels; however, given that the

CNV produces a premature stop codon, the truncated mRNA could be degraded through the nonsense-mediated decay pathway (Brognia and Wen, 2009).

The intermediate CNV deletion is 51,197 bp size, with its breakpoints in introns 7 and 31, causing a partial loss of both introns, and a full loss of 23 exons. Since the deletion is in-frame, there is no introduction of a premature stop codon, but 1248 out of 2261 amino acids are lost, accounting for 55.2% of the protein. The lost amino acids span from the first extracellular domain to the second, and include the intracellular nucleotide-binding domain, the first regulatory domain, and six transmembrane domains (Qian et al., 2017). Given the size of the deletion, there are many possibilities for mechanistic dysfunction. One possibility is that apo A-I is unable to interact with ABCA1 through its extracellular domains, while an alternative possibility is that cholesterol cannot be transported out the cell (Fitzgerald et al., 2004; Nagao et al., 2012; Vedhachalam et al., 2007; Wang et al., 2000).

The full-gene CNV deletion is ~2 Mb and encompasses seven protein-coding genes, including *ABCA1*. In contrast to the previous two CNVs, due to the complete loss of a functional allele, the mechanism of decreased HDL cholesterol may simply be based on a decrease in *ABCA1* expression. As the largest CNV out of all four patients, it is also interesting to note that the patient carrying this deletion has the most severely decreased levels of HDL cholesterol, at 0.03 mmol/L.

When considering the magnitude of each CNV, the size of the genomic deletion correlates to the severity of the HDL phenotype for each patient; however, the corresponding loss of amino acids does not. The patient with the smallest CNV had an HDL cholesterol level of 0.81 mmol/L, while the patients with the intermediate CNV had HDL cholesterol levels of 0.56 mmol/L and 0.47 mmol/L. Additional studies are necessary to fully understand the mechanistic consequences of each CNV—particularly the partial deletions—and how they impact each patients' HDL phenotype. As well, the severity of each patients' phenotype may not solely be due to the CNV, but may be influenced by additional genetic or environmental determinants (Cole et al., 2015). Others have noted a wide range in HDL cholesterol levels, ranging from ~15 to 70% of

normal values among heterozygous carriers of *ABCA1* nonsense mutations resulting in premature protein truncation (Pisciotta et al., 2004); this inter-individual variation in HDL cholesterol reduction echoes the range of biochemical disturbances seen in the small patient sample studied here. Difficulty in attributing quantitative or pathogenic impact is also encountered in research on heterozygous *ABCA1* SNVs that affect HDL cholesterol; functional studies may help understand the mechanistic impact of a SNV, but even between individuals who share the same genetic variant, there can be substantial differences in HDL cholesterol levels (Brunham et al., 2006). Such differences might result from unmeasured gene-gene interactions, unmeasured gene-environment interactions, epigenetic, mitochondrial or microbiome effects.

### 3.6 Conclusion

Our findings implicate a novel form of genetic variation that is likely impacting HDL cholesterol levels, and further emphasizes the complex genetic architecture underlying HDL phenotypes. Understanding that levels of HDL cholesterol can be influenced by rare SNVs, accumulation of common SNPs, and now the presence of rare CNVs, will influence future screening of individuals with extreme HDL phenotypes. Systematic screening for CNVs until recently had heretofore not been feasible due to time-consuming and costly methods (Iacocca and Hegele, 2018); improvements to bioinformatic tools have enabled robust analysis of NGS data, leading to comprehensive, simultaneous assessment of multiple types of genetic determinants. These tools will likely reveal further diversity of the genetic basis for other dyslipidemia and metabolic phenotypes. Given their low frequency in our patient cohort, we anticipate that large-scale CNVs, either deletions or insertions, will likely be infrequent among patients with dyslipidemias, but will nonetheless still need to be considered, in addition to small-scale rare genetic variants and polygenic risk.

### 3.7 References

- Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65.
- Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C., *et al.* (2009). The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* 19, 1622-1629.
- Alsmadi, O., John, S.E., Thareja, G., Hebbar, P., Antony, D., Behbehani, K., and Thanaraj, T.A. (2014). Genome at juncture of early human migration: a systematic analysis of two whole genomes and thirteen exomes from Kuwaiti population subgroup of inferred Saudi Arabian tribe ancestry. *PLoS One* 9, e99069.
- Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., *et al.* (1999). The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat Genet* 22, 347-351.
- Boomsma, D.I., Wijmenga, C., Slagboom, E.P., Swertz, M.A., Karsen, L.C., Abdellaoui, A., Ye, K., Guryev, V., Vermaat, M., van Dijk, F., *et al.* (2014). The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 22, 221-227.
- Brogna, S., and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Biol* 16, 107-113.
- Brooks-Wilson, A., Marcil, M., Clee, S.M., Zhang, L.H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J.A., Molhuizen, H.O., *et al.* (1999). Mutations in ABCA1 in Tangier disease and familial high-density lipoprotein deficiency. *Nat Genet* 22, 336-345.
- Brunham, L.R., Singaraja, R.R., and Hayden, M.R. (2006). Variations on a gene: rare and common variants in ABCA1 and their impact on HDL cholesterol levels and atherosclerosis. *Annu Rev Nutr* 26, 105-129.
- Candini, C., Schimmel, A.W., Peter, J., Bochem, A.E., Holleboom, A.G., Vergeer, M., Dullaart, R.P., Dallinga-Thie, G.M., Hovingh, G.K., Khoo, K.L., *et al.* (2010). Identification and characterization of novel loss of function mutations in ATP-binding cassette transporter A1 in patients with low plasma high-density lipoprotein cholesterol. *Atherosclerosis* 213, 492-498.
- Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869-872.
- Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.

- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., *et al.* (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704-712.
- Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., *et al.* (2011). A copy number variation morbidity map of developmental delay. *Nat Genet* 43, 838-846.
- Dron, J.S., and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.
- Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhoads, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.
- Fitzgerald, M.L., Morris, A.L., Chroni, A., Mendez, A.J., Zannis, V.I., and Freeman, M.W. (2004). ABCA1 and amphipathic apolipoproteins form high-affinity molecular complexes required for cholesterol efflux. *J Lipid Res* 45, 287-294.
- Holleboom, A.G., Kuivenhoven, J.A., Peelman, F., Schimmel, A.W., Peter, J., Defesche, J.C., Kastelein, J.J., Hovingh, G.K., Stroes, E.S., and Motzack, M.M. (2011). High prevalence of mutations in LCAT in patients with low HDL cholesterol levels in The Netherlands: identification and characterization of eight novel mutations. *Hum Mutat* 32, 1290-1298.
- Iacocca, M.A., and Hegele, R.A. (2018). Role of DNA copy number variation in dyslipidemias. *Curr Opin Lipidol* 29, 125-132.
- Iacocca, M.A., Wang, J., Dron, J.S., Robinson, J.F., McIntyre, A.D., Cao, H., and Hegele, R.A. (2017). Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia. *J Lipid Res* 58, 2202-2209.
- Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., *et al.* (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84, 148-161.
- Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* 55, 765-772.
- Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., *et al.* (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56-64.
- Kiss, R.S., Kavaslar, N., Okuhira, K., Freeman, M.W., Walter, S., Milne, R.W., McPherson, R., and Marcel, Y.L. (2007). Genetic etiology of isolated low HDL syndrome: incidence and heterogeneity of efflux defects. *Arterioscler Thromb Vasc Biol* 27, 1139-1145.

Lewis, G.F., and Rader, D.J. (2005). New insights into the regulation of HDL metabolism and reverse cholesterol transport. *Circ Res* 96, 1221-1232.

Marcil, M., Brooks-Wilson, A., Clee, S.M., Roomp, K., Zhang, L.H., Yu, L., Collins, J.A., van Dam, M., Molhuizen, H.O., Loubster, O., *et al.* (1999). Mutations in the ABC1 gene in familial HDL deficiency with defective cholesterol efflux. *Lancet* 354, 1341-1346.

Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., *et al.* (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* 470, 59-65.

Nagao, K., Kimura, Y., and Ueda, K. (2012). Lysine residues of ABCA1 are required for the interaction with apoA-I. *Biochim Biophys Acta* 1821, 530-535.

Park, H., Kim, J.I., Ju, Y.S., Gokcumen, O., Mills, R.E., Kim, S., Lee, S., Suh, D., Hong, D., Kang, H.P., *et al.* (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat Genet* 42, 400-405.

Pisciotta, L., Hamilton-Craig, I., Tarugi, P., Bellocchio, A., Fasano, T., Alessandrini, P., Bon, G.B., Siepi, D., Mannarino, E., Cattin, L., *et al.* (2004). Familial HDL deficiency due to ABCA1 gene mutations with or without other genetic lipoprotein disorders. *Atherosclerosis* 172, 309-320.

Qian, H., Zhao, X., Cao, P., Lei, J., Yan, N., and Gong, X. (2017). Structure of the Human Lipid Exporter ABCA1. *Cell* 169, 1228-1239 e1210.

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J.C., Deleuze, J.F., Brewer, H.B., Duverger, N., Deneffe, P., *et al.* (1999). Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nat Genet* 22, 352-355.

Sadananda, S.N., Foo, J.N., Toh, M.T., Cermakova, L., Trigueros-Motos, L., Chan, T., Liany, H., Collins, J.A., Gerami, S., Singaraja, R.R., *et al.* (2015). Targeted next-generation sequencing to diagnose disorders of HDL cholesterol. *J Lipid Res* 56, 1993-2001.

Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., *et al.* (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res* 19, 1682-1690.

Suktitipat, B., Naktang, C., Mhuantong, W., Tularak, T., Artiwet, P., Pasomsap, E., Jongjaroenprasert, W., Fuchareon, S., Mahasirimongkol, S., Chantratita, W., *et al.* (2014). Copy number variation in Thai population. *PLoS One* 9, e104355.

Teague, B., Waterman, M.S., Goldstein, S., Potamouisis, K., Zhou, S., Reslewic, S., Sarkar, D., Valouev, A., Churas, C., Kidd, J.M., *et al.* (2010). High-resolution human

genome structure by single-molecule analysis. *Proc Nat Acad Sci USA* *107*, 10848-10853.

Tuzun, E., Sharp, A.J., Bailey, J.A., Kaul, R., Morrison, V.A., Pertz, L.M., Haugen, E., Hayden, H., Albertson, D., Pinkel, D., *et al.* (2005). Fine-scale structural variation of the human genome. *Nat Genet* *37*, 727-732.

Vedhachalam, C., Duong, P.T., Nickel, M., Nguyen, D., Dhanasekaran, P., Saito, H., Rothblat, G.H., Lund-Katz, S., and Phillips, M.C. (2007). Mechanism of ATP-binding cassette transporter A1-mediated cellular lipid efflux to apolipoprotein A-I and formation of high density lipoprotein particles. *J Biol Chem* *282*, 25123-25130.

Wang, N., Silver, D.L., Costet, P., and Tall, A.R. (2000). Specific binding of ApoA-I, enhanced cholesterol efflux, and altered plasma membrane morphology in cells expressing ABC1. *J Biol Chem* *275*, 33053-33058.

Wong, L.P., Ong, R.T., Poh, W.T., Liu, X., Chen, P., Li, R., Lam, K.K., Pillai, N.E., Sim, K.S., Xu, H., *et al.* (2013). Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet* *92*, 52-66.

Zarrei, M., MacDonald, J.R., Merico, D., and Scherer, S.W. (2015). A copy number variation map of the human genome. *Nat Rev Genet* *16*, 172-183.

## Chapter 4 – Severe hypertriglyceridemia is primarily polygenic

The work contained in this Chapter has been edited from its original publication in the *Journal of Clinical Lipidology* for brevity and to ensure consistency throughout this Dissertation.

Journal of Clinical Lipidology (2019) 13, 80–88



### Severe hypertriglyceridemia is primarily polygenic

Check for updates

**Jacqueline S. Dron, BSc, Jian Wang, MD, Henian Cao, MD, PhD, Adam D. McIntyre, BSc, Michael A. Iacocca, BSc, Jyler R. Menard, BSc, Irina Movsesyan, PhD, Mary J. Malloy, MD, Clive R. Pullinger, PhD, John P. Kane, MD, Robert A. Hegele, MD\***

*Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada (Drs Dron, Wang, Cao, McIntyre, Iacocca and Hegele); Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada (Drs Dron, Iacocca and Hegele); Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada (Dr Menard); Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA (Drs Movsesyan, Malloy, Pullinger and Kane); and Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada (Dr Hegele)*

**KEYWORDS:**  
Copy-number variants (CNVs); Familial chylomicronemia syndrome (FCS); Hypertriglyceridemia (HTG); Monogenic; Next-generation sequencing (NGS); Polygenic; Polygenic risk score; Rare variants; Single-nucleotide polymorphism (SNP)

**BACKGROUND:** Hypertriglyceridemia (HTG) is a complex trait defined by elevated plasma triglyceride levels. Genetic determinants of HTG have so far been examined in a piecemeal manner; understanding of its molecular basis, both monogenic and polygenic, is thus incomplete.

**OBJECTIVE:** The objective of this study was to characterize genetic profiles of patients with severe HTG, and quantify the genetic determinants and molecular contributors.

**METHODS:** We concurrently assessed rare and common variants in two independent cohorts of 251 and 312 Caucasian patients with severe HTG. DNA was subjected to targeted next-generation sequencing of 73 genes and 185 SNPs associated with dyslipidemia. *LPL*, *APOC2*, *GPIIIBP1*, *APOA5*, and *LMF1* genes were screened for rare variants, and a polygenic risk score was used to assess the accumulation of common variants.

**RESULTS:** As there were no significant differences in the prevalence of genetic determinants between cohorts, data were combined for all 563 patients: 1.1% had biallelic (homozygous or compound heterozygous) rare variants, 14.4% had heterozygous rare variants, 32.0% had an extreme accumulation of common variants (ie, high polygenic risk), and 52.6% remained genetically undefined. Patients with HTG were 5.77 times (95% CI [4.26–7.82],  $P < .0001$ ) more likely to carry one of these types of genetic susceptibility compared with controls.

**CONCLUSIONS:** We report the most in-depth, systematic evaluation of genetic determinants of severe HTG to date. The predominant feature was an extreme accumulation of common variants (high polygenic risk score), whereas a substantial proportion of patients also carried heterozygous rare

Declarations of interest: R.A.H. is a consultant and speakers' bureau member for Amgen, Alkermes/Innis, Amgen, Cerenis, Genzyme, and Sanofi. The other authors have no disclosures.

\* Corresponding author. Robarts Research Institute, Western University, 4288A-1151 Richmond Street North, London, N6A 5C1, Ontario, Canada.

E-mail address: hegele@robarts.ca  
Submitted September 22, 2018. Accepted for publication October 16, 2018.

1933-2874/© 2018 National Lipid Association. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).  
<https://doi.org/10.1016/j.jacl.2018.10.006>

**Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., et al. (2019). Severe hypertriglyceridemia is primarily polygenic. J Clin Lipidol 13, 80–88.**



## 4.1 Abstract

**Objective:** Hypertriglyceridemia is a complex trait defined by elevated plasma triglyceride levels. Genetic determinants of hypertriglyceridemia have so far been examined in a piecemeal manner; understanding of its molecular basis, both monogenic and polygenic, is thus incomplete. Here, we characterize genetic profiles of severe hypertriglyceridemia patients and quantify their genetic determinants and molecular contributors.

**Methods and Results:** We concurrently assessed rare and common variants in two independent cohorts of 251 and 312 severe hypertriglyceridemia patients of European ancestry. DNA was subjected to targeted next-generation sequencing of 69 genes and 185 SNPs associated with dyslipidemia. *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2* genes were screened for rare variants, and a polygenic risk score was used to assess the accumulation of common variants. As there were no significant differences in the prevalence of genetic determinants between cohorts, data were combined for all 563 patients: 1.1% had bi-allelic (homozygous or compound heterozygous) rare variants, 14.4% had heterozygous rare variants, 32.0% had an extreme accumulation of common variants (i.e. high polygenic risk), and 52.6% remained genetically undefined. Hypertriglyceridemia patients were 5.77-times (95% CI [4.26-7.82];  $P < 0.0001$ ) more likely to carry one of these types of genetic susceptibility compared to normolipidemic controls.

**Conclusions:** We report the most in-depth, systematic evaluation of genetic determinants of severe hypertriglyceridemia to date. The predominant feature was an extreme accumulation of common variants (high polygenic risk score), while a substantial proportion of patients also carried heterozygous rare variants. Overall, 46.3% of patients had polygenic hypertriglyceridemia (i.e. multifactorial chylomicronemia), while only 1.1% had monogenic hypertriglyceridemia (i.e. familial chylomicronemia syndrome).

## 4.2 Introduction

Elevated fasting levels of plasma triglyceride is a common dyslipidemia that is clinically identified as hypertriglyceridemia. Depending on the degree of elevation, hypertriglyceridemia can be mild-to-moderate ( $\geq 2$ - $9.9$  mmol/L) or severe ( $\geq 10$  mmol/L) (Hegele et al., 2014). Of particular importance is severe hypertriglyceridemia: with a prevalence of  $\sim 1$  in 600 (Johansen et al., 2011a), affected individuals are at risk of several clinical manifestations, the most serious being acute pancreatitis (Brahm and Hegele, 2015; Dron and Hegele, 2017). Although it is relatively prevalent in the population and can lead to life-threatening medical emergencies, there remains substantial unfamiliarity with the molecular genetic determinants of severe hypertriglyceridemia, as well as the role of genetic testing in its diagnosis.

Severe hypertriglyceridemia very often results from chylomicronemia, defined as the pathological accumulation of circulating chylomicrons (Brahm and Hegele, 2015). While abnormalities in the catabolic processing of other triglyceride-rich lipoproteins—namely, very-low-density lipoprotein (VLDL), intermediate-density lipoproteins (IDL), and remnant particles—can also contribute to the severe hypertriglyceridemia phenotype, chylomicrons are usually considered to be the primary lipoprotein of concern. Biochemically quantifying chylomicron concentration can be inconvenient and difficult, so a diagnosis of “chylomicronemia” is often based on the fasting triglyceride concentration and the presence of other suggestive features, such as a milky appearance of the patient’s blood plasma (Brahm and Hegele, 2013; Brahm and Hegele, 2015). Clinically, patients are at risk of pancreatitis and physical signs such as eruptive xanthomas, lipemia retinalis and hepatosplenomegaly can be observed. Due to the challenges in quantifying the abnormal lipoprotein fractions in patients with elevated triglyceride, we focus on the generalized “severe hypertriglyceridemia” phenotype, rather than subtypes defined by the presence of particular abnormal lipoprotein particles, as seen in the Fredrickson classification of hyperlipidemias (Beaumont et al., 1970).

Severe hypertriglyceridemia is considered to have both monogenic and polygenic determinants (Brahm and Hegele, 2015). A subset of this patient group has familial chylomicronemia syndrome (FCS), a rare form of monogenic hypertriglyceridemia that

has textbook estimates of a prevalence of ~1 to 10 in a million (Johansen et al., 2011a). As an autosomal recessive disorder, definitive molecular diagnosis of FCS hinges on detection of rare, bi-allelic (homozygous or compound heterozygous) variants in the same gene (Johansen et al., 2011a); the canonical triglyceride metabolism genes found to be mutated in FCS include *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*. Genetic assessment has superseded biochemical assays of plasma post-heparin lipolytic activity as the current gold standard for diagnosis of deficiency of lipoprotein lipase (LPL), encoded by *LPL*, and related factors (Brahm and Hegele, 2015).

The remainder of genetically-based, non-FCS cases of severe hypertriglyceridemia are considered to be polygenic in nature and can be referred to as “multifactorial chylomicronemia” (Brahm and Hegele, 2015). Here, several different genetic factors contribute to disease susceptibility, including rare heterozygous variants in canonical triglyceride genes, common variants associated with elevated triglyceride levels, and/or variants in non-canonical triglyceride genes (Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Kathiresan et al., 2009; Surendran et al., 2012; Teslovich et al., 2010; Wang et al., 2008; Willer et al., 2013). As well, certain environmental factors can interact with this assortment of polygenic determinants to force expression of severely elevated triglyceride levels. Despite the detailed documentation of similarities and differences between monogenic and polygenic hypertriglyceridemia (i.e. FCS vs. multifactorial chylomicronemia), in practice, there is a tendency to equate “severe hypertriglyceridemia” with FCS (Brahm and Hegele, 2015). Usually, when triglyceride levels exceed 10 mmol/L, there is no monogenic cause identified (Brahm and Hegele, 2015), making LPL deficiency or FCS a highly unlikely cause of severe hypertriglyceridemia.

In our experience, a molecular diagnosis for a patient with severe hypertriglyceridemia requires simultaneous assessment of all possible genetic determinants—both common and rare variants. Here, we sought to systematically evaluate the genetic profiles of almost 600 severe hypertriglyceridemia patients to provide an updated and comprehensive description of the genetic landscape of this complex phenotype. With our custom-designed, targeted next-generation sequencing (NGS) panel, “LipidSeq”, and

bioinformatic tools, we can concurrently measure multiple genetic factors, including rare variants—both single-nucleotide variants (SNVs) and copy-number variants (CNVs)—together with the accumulation of common variants (i.e. single-nucleotide polymorphisms [SNPs]) within a polygenic risk score, thus directly evaluating prevalence of each type of genetic determinant in severe hypertriglyceridemia. We demonstrate that severe hypertriglyceridemia in adults is most often associated with polygenic factors (either heterozygous rare variants or high polygenic risk scores), and that FCS due to monogenic bi-allelic variants (i.e. homozygous or compound heterozygous) is very uncommon in these patient cohorts.

## 4.3 Materials and Methods

### 4.3.1 Study subjects

Patients of interest included those of European ancestry with triglyceride levels  $\geq 10$  mmol/L; they were defined as having “severe” hypertriglyceridemia.

In adherence to the Declaration of Helsinki, all patients provided written, informed consent for collection of personal data and DNA with approval from either the Western University (London ON, Canada) ethics review board (no. 07290E) or the Committee on Human Research of the University of California, San Francisco (UCSF).

As a reference control cohort of normolipidemic individuals, the publicly available data pertaining to the self-reported healthy individuals from the European subgroup of the 1000 Genomes Project (N=503) were studied.

### 4.3.2 DNA preparation and targeted sequencing

DNA isolation and preparation for targeted NGS follows the same methodology as described in [Chapter 2, Section 2.3.2](#).

From the UCSF cohort, genomic DNA was isolated as described elsewhere (Pullinger et al., 2015).

### 4.3.3 Bioinformatic processing of sequencing data

The bioinformatic processing of sequencing data follows the same methodology as described in [Chapter 3, Section 3.3.3](#); however, an updated version of CLC Bio Genomics Workbench (version 10.0; CLC Bio, Aarhus, Denmark) was used.

### 4.3.4 Annotation and analysis of rare single-nucleotide variants

The SNVs contained within each patients' VCF file were annotated using VarSeq® (version 1.4.8; Golden Helix, Inc., Bozeman MT, USA). Variants of interest within *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2* were identified following a “rare variant” model. SNVs were identified as having a minor allele frequency of  $\leq 1\%$  or missing in the Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org/>) (Lek et al., 2016) and 1000 Genomes Project (<http://browser.1000genomes.org/index.html>) (Genomes Project et al., 2015) databases. Rare missense, nonsense, deletion, insertion, splice-acceptor, and splice-donor variants were retained. *In silico* prediction algorithms were then used to select SNVs with likely large phenotypic effects. The Combined Annotation Dependent Depletion (CADD; <http://cadd.gs.washington.edu/score>) (Kircher et al., 2014) PHRED-scaled score was the primary metric considered for variant deleteriousness. Variants were required to have a CADD PHRED-scaled score  $\geq 10$ , and be predicted to be deleterious or damaging by at least one additional prediction tool—Polymorphism Phenotyping version 2 (PolyPhen2; <http://genetics.bwh.harvard.edu/pph2/>) (Adzhubei et al., 2013), Sorting Intolerant From Tolerant (SIFT; <http://sift.jcvi.org/>) (Kumar et al., 2009), and MutationTaster (<http://www.mutationtaster.org/>)—when classifications were available. SNVs with a read-depth of  $< 30$  were excluded.

### 4.3.5 Detection of rare copy-number variants

CNVs were identified following the same methodology as described in [Chapter 3, Section 3.3.4](#). The genes in which CNVs were screened for included *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*.

CNV analysis could not be performed on the 1000 Genomes Project data, as BAM files were not available.

### 4.3.6 Polygenic risk score for elevated triglyceride levels

We created a weighted polygenic risk score consisting of 16 SNPs associated with triglyceride levels, as reported by the Global Lipids Genetics Consortium (GLGC) genome-wide association study (GWAS) (Willer et al., 2013), and calculated it for each patient (**Table 4.1**). The number of triglyceride-raising alleles at a locus (either 0, 1, or 2) was counted and multiplied by its beta coefficient, or phenotypic “effect size” as reported in the GLGC GWAS summary statistics. The products for each SNP locus were then totalled for the overall weighted polygenic risk score for each patient.

### 4.3.7 Statistical analysis

Normality was assessed using the D’Agostino and Pearson test. Differences between parametric data were assessed using an unpaired, one-tailed Students t-test while differences between nonparametric data were assessed using a Mann-Whitney test. Differences between mean polygenic risk scores and mean triglyceride levels across molecular hypertriglyceridemia cohorts were assessed using a Kruskal-Wallis test followed by a Dunn’s multiple comparison. All tests were performed assuming unequal variances and are reported as the mean  $\pm$  standard deviation (SD). Odds ratios (ORs) were derived using 2-by-2 contingency tables, with one-tailed Fisher’s exact tests to assess significance. Statistical analyses were conducted using GraphPad Prism for Windows (version 7.04; GraphPad Software, La Jolla CA, USA). Statistical significance was defined as  $P < 0.05$ .

## 4.4 Results

### 4.4.1 Characteristics of study subjects

Two-hundred and fifty-one patients were selected for study from the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada). An additional 312 patients were also selected from the Genomic Resource in Arteriosclerosis and Metabolic Disease who were recruited at the Lipid, Diabetes, or Cardiology Clinics at UCSF (San Francisco, CA, USA). Clinical and demographic characteristics of both cohorts are defined in **Table 4.2**.

**Table 4.1 The 16 SNPs used in polygenic risk score for elevated triglyceride levels.**

Chr:position	rsID	Closest gene	Effect allele	Relation with triglyceride metabolism	Variant ontology
1:63025942	rs2131925	<i>ANGPTL3</i>	T (0.066)	ANGPTL3 inhibits LPL and reduces triglyceride hydrolysis (Tikka and Jauhiainen, 2016).	Upstream
1:230295691	rs4846914	<i>GALNT2</i>	G (0.04)	Impacts triglyceride clearance through interactions with apo C-III (Holleboom et al., 2011).	Intronic
4:88030261	rs442177	<i>KLHL8, AFF1</i>	T (0.031)	Mechanism is poorly characterized.	Downstream
5:55861786	rs9686661	<i>MAP3K1</i>	T (0.038)	Mechanism is poorly characterized.	Upstream
7:72982874	rs17145738	<i>MLXIPL</i>	C (0.115)	Helps regulate glycolysis, gluconeogenesis and lipogenesis (Nakayama et al., 2011).	Downstream
8:18272881	rs1495741	<i>NAT2</i>	G (0.04)	Role in insulin sensitivity (Knowles et al., 2015).	Downstream
8:19844222	rs12678919	<i>LPL</i>	A (0.17)	Hydrolyzes triglyceride from triglyceride-rich lipoproteins (Lambert and Parks, 2012).	Downstream
8:126490972	rs2954029	<i>TRIB1</i>	A (0.076)	Regulates expression of lipogenic genes (Douvris et al., 2014).	Downstream
10:65027610	rs10761731	<i>JMJD1C</i>	A (0.031)	Regulates expression of lipogenic genes (Viscarra et al., 2020).	Intronic
11:61569830	rs174546	<i>FADS1-S2-S3</i>	T (0.045)	Modification of dietary fatty acids (Mathias et al., 2014).	3'UTR, intronic, downstream
11:116648917	rs964184	<i>APOA1-C3-A4-A5</i>	G (0.234)	Involved in the structure of triglyceride-rich lipoproteins and regulation of triglyceride hydrolysis (Feingold and Grunfeld, 2000).	Downstream, upstream, downstream, downstream
15:42683787	rs2412710	<i>CAPN3</i>	A (0.099)	Mechanism is poorly characterized.	Intronic
15:44245931	rs2929282	<i>FRMD5</i>	T (0.072)	Mechanism is poorly characterized.	Intronic
16:56993324	rs3764261	<i>CETP</i>	C (0.04)	Facilitates the transfer of lipids between HDL and triglyceride-rich lipoproteins (Daniels et al., 2009).	Upstream
19:19407718	rs10401969	<i>CSPG3, CILP2</i>	T (0.121)	Mechanism is poorly characterized.	Intronic
20:44554015	rs6065906	<i>PLTP</i>	C (0.053)	Moves phospholipids between lipoproteins (Daniels et al., 2009).	Upstream

Variant information related to effect size was extracted from Willer et al. (2013). Effect alleles are in reference to trait elevation; the bracketed value denotes the effect size of each allele per increase in standard deviation. Variant ontology is relative to the closest gene. Abbreviations: ANGPTL3 = angiotensin-like protein 3; apo = apolipoprotein; chr = chromosome; HDL = high-density lipoprotein; LPL = lipoprotein lipase; UTR = untranslated region.

**Table 4.2 Clinical and demographic information of severe hypertriglyceridemia cohorts (N=563).**

	Lipid Genetics Clinic		UCSF	
	Males	Females	Males	Females
<b>N</b>	171	80	203	109
<b>Age</b>	50.9 ± 11.2*	49.0 ± 15.0*	47.1 ± 11.7	50.3 ± 13.7
<b>BMI (kg/m<sup>2</sup>)</b>	30.6 ± 4.42*	30.8 ± 6.03*	29.7 ± 4.86*	28.2 ± 4.68*
<b>Total cholesterol (mmol/L)</b>	11.0 ± 5.59*	11.9 ± 7.37*	11.1 ± 5.44*	12.3 ± 6.58*
<b>Triglyceride (mmol/L)</b>	26.6 ± 20.9	30.3 ± 26.3	22.1 ± 17.9	29.4 ± 32.5
<b>HDL cholesterol (mmol/L)</b>	0.83 ± 0.45*	0.77 ± 0.27*	0.74 ± 0.32*	0.81 ± 0.41*
<b>LDL cholesterol (mmol/L)</b>	3.24 ± 2.12*	3.67 ± 5.00*	2.11 ± 1.25*	2.09 ± 1.10*
<b>Diabetes</b>	37.7%*	37.5%*	36.3%*	47.5%*

Values are indicative of the mean ± SD. “\*” indicates means were calculated with an incomplete dataset. Abbreviations: BMI = body-mass index; HDL = high-density lipoprotein; LDL = low-density lipoprotein; UCSF = University of California, San Francisco.



#### 4.4.2 Rare variants identified in canonical triglyceride metabolism genes

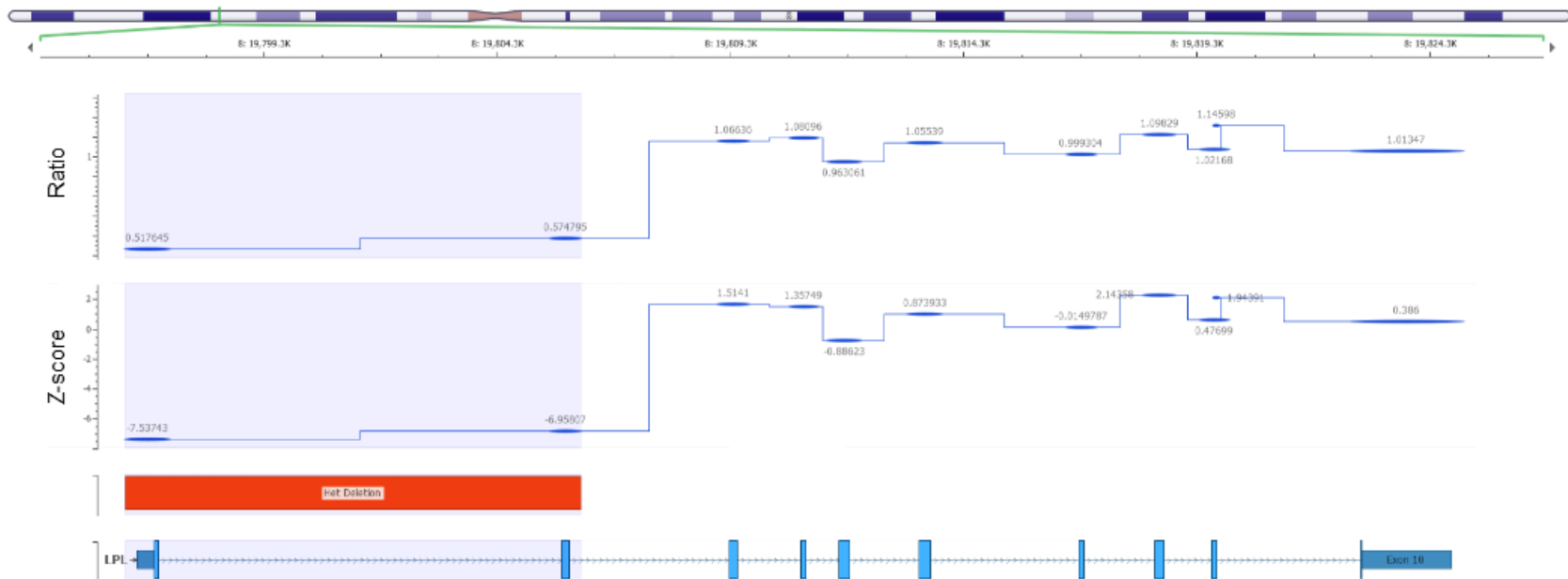
We assessed patients for rare SNVs or CNVs in the primary triglyceride-related genes (*LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*). From the Lipid Genetics Clinic and UCSF, 15.9% (40/251) and 15.0% (47/312) of patients carried rare variants, respectively, compared to only 4.0% (20/503) of individuals from the normolipidemic controls. A total of 71 unique variants were present across all five genes in these samples.

Across all three cohorts, the majority of rare SNVs were heterozygous and only a few were bi-allelic. Both bi-allelic SNVs and CNVs were exclusive to the patient cohorts. From the Lipid Genetics Clinic, three patients carried bi-allelic SNVs, and 37 patients carried heterozygous SNVs (**Appendix F**). From UCSF, three patients carried bi-allelic SNVs, 43 patients carried heterozygous SNVs, and one patient carried a CNV—a partial deletion of *LPL* (**Figure 4.1** and **Appendix G**). Twenty individuals from the 1000 Genomes cohort carried heterozygous SNVs (**Appendix H**).

#### 4.4.3 Measuring accumulation of common triglyceride-raising alleles

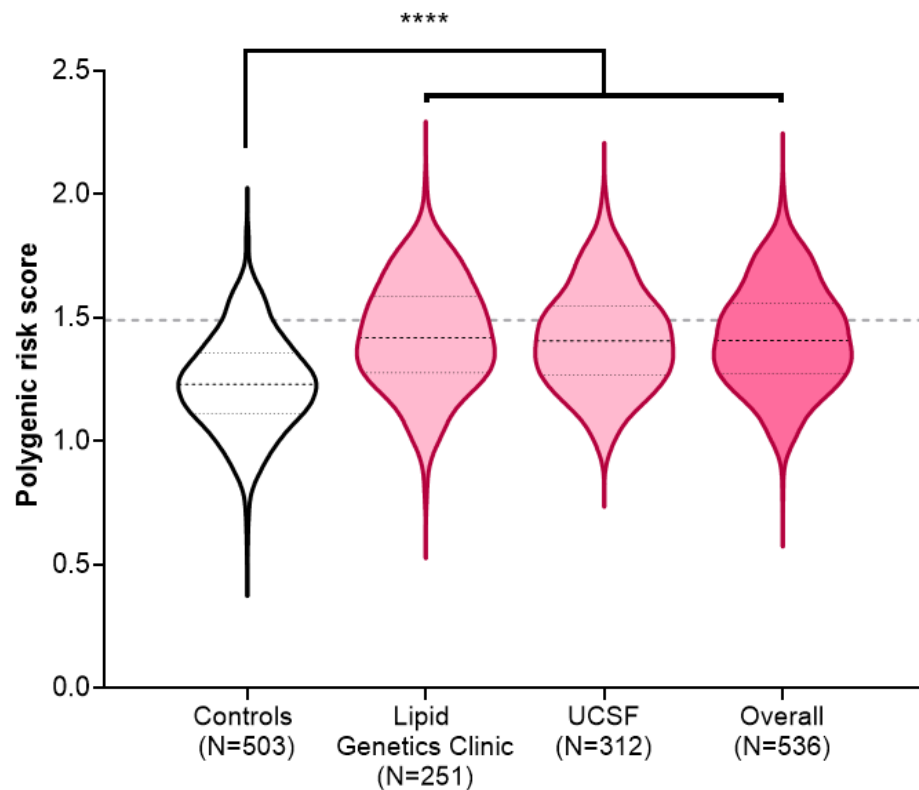
When assessing our polygenic risk score, higher scores reflect increased accumulations of triglyceride-raising alleles. We considered scores  $\geq 1.49$  (90<sup>th</sup> percentile in accordance with scores from the normolipidemic controls) as “extreme” risk scores, indicating an extreme accumulation of triglyceride-raising alleles.

From the Lipid Genetics Clinic and UCSF cohorts, 41.2% (87/211) and 35.0% (93/265) of patients without rare variants, respectively, had extreme risk scores, compared to only 9.5% (48/473) of individuals without rare variants from the 1000 Genomes cohort. When considering all individuals from the Lipid Genetics Clinic and UCSF, 34.7% (87/251) and 29.8% (93/312) of patients had extreme polygenic risk scores, respectively, indicating a polygenic basis for their severe hypertriglyceridemia phenotype (**Figure 4.2**).



**Figure 4.1 Identification of a *LPL* CNV using the VarSeq-CNV® caller algorithm.**

A patient from UCSF was found to carry a heterozygous deletion of *LPL*. This figure has been taken and modified from the VarSeq program. Chr8:19,794,505–19,826,742 (hg19 genome build) is the region visualized in each panel, with the CNV “ratio”, and “Z-score” for the deleted region. Spanning across two exons, the deleted region has an average target depth of 129.753, an average Z-score of -7.248, and an average ratio of 0.546. Abbreviations: CNV = copy-number variant.



**Figure 4.2 Polygenic risk score analysis for severe hypertriglyceridemia patients.**

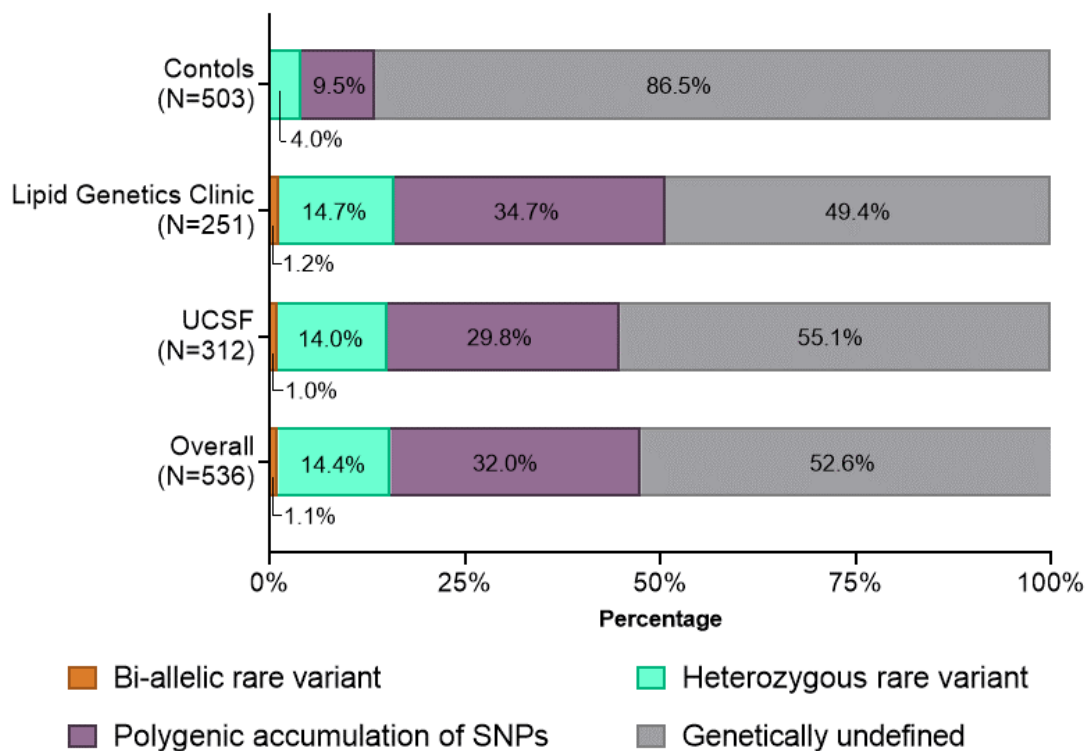
Violin plots illustrate the distribution of polygenic risk scores in normolipidemic controls from the 1000 Genomes cohort and severe hypertriglyceridemia patients from the Lipid Genetics Clinic cohort and UCSF. Patients with an “extreme” accumulation of triglyceride-raising SNP alleles are defined as having scores above the 90<sup>th</sup> percentile threshold ( $\geq 1.49$ ) in the 1000 Genomes cohort, which is illustrated by the grey hashed line. The hashed lines within each violin plot represent the median and interquartile ranges. P-values were generated from a Kruskal-Wallis test and adjusted with Dunn’s multiple comparisons based on mean polygenic risk score values between groups. P-values: \*\*\*\*  $<0.0001$ . Abbreviations: UCSF = University of California, San Francisco.

#### 4.4.4 Comparison of genetic profiles between cohorts

The genetic profiles for the Lipid Genetics Clinic, UCSF and 1000 Genomes Project are presented in **Figure 4.3**. The ORs comparisons for each type of genetic variant are detailed in **Figure 4.4** and demonstrate that there is no genetic difference between the two patient groups, indicating a successful validation of the observations from the Lipid Genetics Clinic analysis. Overall, hypertriglyceridemia patients are 5.77-times (95% CI [4.26-7.82];  $P < 0.0001$ ) more likely to carry one of the three types of genetic determinants linked to hypertriglyceridemia, compared to normolipidemic controls. There is a striking difference in the genetic profiles between patients with severe hypertriglyceridemia, and normolipidemic controls (see **Figure 4.3** and **Figure 4.4**).

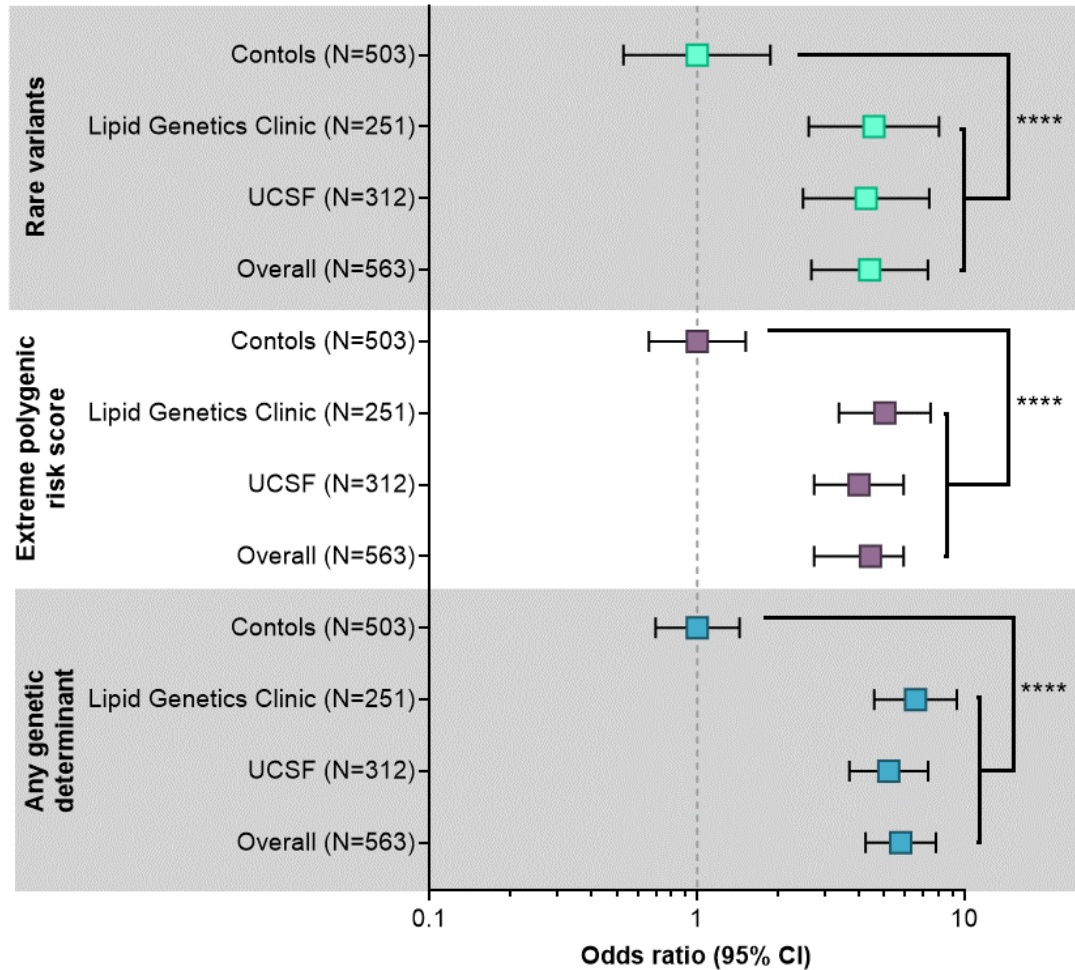
#### 4.4.5 Comparison of triglyceride levels between molecular forms of hypertriglyceridemia

There was a nonsignificant trend towards elevated mean triglyceride levels in patients with FCS (N=6;  $34.8 \pm 13.8$  mmol/L), compared to patients with polygenic (N=261;  $25.5 \pm 19.8$  mmol/L;  $P=0.153$ ) or genetically undefined (N=296;  $26.3 \pm 26.8$  mmol/L;  $P=0.077$ ) hypertriglyceridemia.



**Figure 4.3 The comparison of genetic determinants of severe hypertriglyceridemia between cohorts.**

Percentages were determined from individuals from the Lipid Genetics Clinic, UCSF cohort, combined patient cohort and the 1000 Genomes Project. Only the patient cohorts contain bi-allelic variants, which are the molecular hallmark of monogenic FCS. Abbreviations: SNPs = single-nucleotide polymorphism; UCSF = University of California, San Francisco.



**Figure 4.4 Differences in genetic determinants of severe hypertriglyceridemia between cohorts.**

Each forest plot illustrates the odds ratio of patients from the Lipid Genetics Clinic and UCSF cohorts having rare variants (including SNVs and CNVs), the extreme accumulation of common triglyceride-raising alleles (as indicated by an extreme polygenic risk score), or either type of genetic determinant, compared to normolipidemic controls from the 1000 Genomes Project. The dashed line indicates an odds ratio of 1.0. P-values were generated from one-tailed Fisher's exact tests. P-values: \* <0.05; \*\*\*\* <0.0001. Abbreviations: CI = confidence interval; UCSF = University of California, San Francisco.

## 4.5 Discussion

Here we report a comprehensive genetic analysis using NGS and bioinformatic tools to simultaneously assess multiple types of genetic variants in patients with severe hypertriglyceridemia. From the Lipid Genetics Clinic cohort of 251 patients, we identified ~50% of individuals with a genetic factor likely contributing towards their severe hypertriglyceridemia phenotype. Of importance was the virtually identical replication of this genetic profile in an independent cohort from UCSF of 312 severe hypertriglyceridemia patients. Across all 563 severe hypertriglyceridemia patients, 1.1% had bi-allelic rare SNVs, 14.2% had heterozygous rare SNVs, and 0.2% had heterozygous rare CNVs, 32.0% had extreme polygenic risk scores, and 52.6% were genetically undefined. In consideration of the genetic classifications of hypertriglyceridemia, 1.1% of patients had monogenic hypertriglyceridemia (defined as having bi-allelic variants in the same gene; i.e. FCS), 46.6% of patients had polygenic hypertriglyceridemia (defined as either a heterozygous mutant rare allele or high polygenic risk scores; i.e. multifactorial chylomicronemia), while the remaining 52.6% of patients had genetically uncharacterized hypertriglyceridemia.

The presence of bi-allelic, loss-of-function variants in canonical triglyceride genes causes FCS; however only 1.1% of patients across both cohorts carried these variants. Our findings strengthen previous reports that bi-allelic variants and FCS are actually an extremely rare subset of the entirety of severe hypertriglyceridemia. Our findings confirm that polygenic hypertriglyceridemia or multifactorial chylomicronemia is overwhelmingly the most common form of this phenotype in adults. We note that the six hypertriglyceridemia individuals with bi-allelic variants tended to have higher triglyceride levels (by about 20%) than individuals with other defined genetic forms of hypertriglyceridemia. The impulse to think first of the ultra-rare monogenic explanation versus the more likely polygenic explanation when confronted with a patient whose triglyceride level exceeds 10 mmol/L should be restrained; although these rare patients with bi-allelic variants exist they constitute a tiny minority of adult patients with severe hypertriglyceridemia (Hegele, 2018).

Heterozygous SNVs were the most frequent type of rare variant identified and were extremely prevalent in hypertriglyceridemia patients compared to normolipidemic controls. Increased heterozygous rare variant frequency in patients compared to healthy individuals has been shown previously for hypertriglyceridemia and other dyslipidemias (Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Motazacker et al., 2013). Interestingly, 4.0% of normolipidemic controls also carried heterozygous SNVs. Studies have reported that individuals sharing the same rare, heterozygous variant can have a wide range of triglyceride levels (Babirak et al., 1989; Hegele et al., 1991; Nordestgaard et al., 1997); secondary factors likely underlie these differences and help explain the presence of heterozygous rare variants with no apparent clinical consequences in some healthy individuals. Indeed, this is commonly seen in complex traits; while a rare heterozygous variant may not be sufficient to drive the severe hypertriglyceridemia phenotype, in concert with other genetic and environmental influences, it can act as a strong polygenic contributor that increases susceptibility to high triglyceride levels (Dron and Hegele, 2018; Hegele et al., 2014).

We observed that CNVs in canonical triglyceride genes were the rarest of all genetic determinants. Although less frequent than bi-allelic variants, the identified CNV was heterozygous, and like heterozygous SNVs, is also insufficient to be considered a driver of FCS. While the *LPL* CNV deletion almost certainly resulted in no functional protein from the mutant allele, the patient was heterozygous, meaning that they potentially had one fully functional *LPL* allele. However, total potential lipolytic capacity would be diminished for this patient, creating vulnerability to the effects of a secondary factor that further compromised LPL activity.

The most prevalent genetic feature underlying severe hypertriglyceridemia here was the polygenic accumulation of common variants—more specifically, the accumulation of triglyceride-raising alleles across multiple SNP loci. While it has been appreciated that SNPs with small phenotypic effects are enriched in hypertriglyceridemia patients (Hegele et al., 2009; Johansen et al., 2011b; Piccolo et al., 2009; Teslovich et al., 2010; Wang et al., 2008), and triglyceride-based risk scores have explained a portion of the variance in triglyceride levels (Aulchenko et al., 2009; Justesen et al., 2015; Latsuzbaia et al., 2016;



Lutsey et al., 2012; Tikkanen et al., 2011), direct comparisons between triglyceride polygenic risk scores in large severe hypertriglyceridemia cohorts and normolipidemic controls are scarce. For the first time, we illustrate that an extreme accumulation of common variants is the most predominant genetic determinant, present in >30% of severe hypertriglyceridemia patients, demonstrating that a large proportion of cases result from the accumulation of multiple small effects originating from numerous small-effect triglyceride-raising loci from across the genome. Our OR calculations confirm that an extreme accumulation of common variants is strongly associated with severe hypertriglyceridemia. Our results further emphasize the importance of considering “polygenic” hypertriglyceridemia as the most common type of genetically-derived severe hypertriglyceridemia.

For the remaining 52.6% of patients without the above genetic determinants, there remain several possible factors that may contribute towards their phenotype. They may have rare variants with clinically relevant effect sizes within certain non-canonical hypertriglyceridemia genes that were not assayed in this study, such as *GALNT2* or *CREB3L3*. This would also include genes involved in pathways that are secondarily associated with elevated triglyceride levels, such as diabetes, insulin resistance and hepatosteatosis. Furthermore, while variants in such genes are not directly associated with extremely elevated triglyceride levels like what is seen in patients with severe hypertriglyceridemia, perhaps they contribute to this phenotype in conjunction with other factors, such as environmental and lifestyle determinants. Studies have started to consider complex gene-environment interactions (Cole et al., 2015), and could guide future analyses in severe hypertriglyceridemia. Certain genotypes alone are likely insufficient to cause extreme elevations in triglyceride levels, but in the presence of certain environmental triggers such as poor diet, obesity, stress or alcohol use, these could contribute to phenotypic changes. For example, adiposity was shown to almost double the impact of triglyceride-associated SNPs incorporated into a weighted risk score (Cole et al., 2014).

Compared to genetic analyses performed in individuals with extremes of low-density lipoprotein (LDL) cholesterol (Wang et al., 2016) and high-density lipoprotein (HDL)

cholesterol (Dron et al., 2017), the severe hypertriglyceridemia phenotype shows marked differences in prevalence and predominance of genetic determinants. Different extreme lipid phenotypes appear to have different underlying genetic architecture. For example, extremely high levels of LDL cholesterol, presenting as “suspected familial hypercholesterolemia” (FH), has a large monogenic component. When studying patients with extreme deviations of HDL cholesterol and triglyceride, researchers have often assumed an analogy with FH and have imposed a monogenic framework on their experiments to define genetic determinants of these complex dyslipidemias. However, a substantial proportion of patients with extreme lipid phenotypes have a primarily polygenic basis, even for many cases of FH (Futema et al., 2015; Talmud et al., 2013; Wang et al., 2016). Furthermore, with clearly defined monogenic dyslipidemias, there can be phenotypic differences depending on the underlying genetic basis, such as in the case of FCS patients who have bi-allelic *LPL* variants versus those with bi-allelic variants in the four minor canonical genes (Hegele et al., 2018).

Another tendency when dealing with quantitative traits is to assume that more extreme deviations reflect stronger genetic components. This has been observed in FH, where patients with higher LDL cholesterol levels were more likely to have monogenic FH (Wang et al., 2016). However, this is not the case for hypertriglyceridemia. Among a subgroup of nine of our patients with triglyceride  $\geq 100$  mmol/L, only one had a rare heterozygous *LPL* variant, while another had an extremely high polygenic risk score, and the remaining seven had no defined genetic determinant. The initial clinical intuition might be that these patients must have monogenic FCS, and that the extreme deviation is due to bi-allelic, large-effect variants (i.e. FCS analogous to homozygous FH). However, this is not the case; hypertriglyceridemia is a volatile trait with genetics that are not analogous to FH or other dyslipidemias.

Our study has some limitations. First, we have no triglyceride measurements for the normolipidemic controls from the 1000 Genomes Project. Since they were self-reported as healthy, we assumed this cohort followed the general distribution of triglyceride levels in a European population. With a prevalence of 1 in 600 individuals having severe hypertriglyceridemia, it is unlikely that any affected individuals were included. Second,

as previously mentioned, neither variants in non-canonical triglyceride genes nor gene-environment interactions were considered. It is very likely that some of the remaining genetically undefined patients may carry one of these alternative genetic influences. To build on this study, future steps could incorporate these factors for an even more detailed look into the genetic landscape of severe hypertriglyceridemia. Also, to broaden our understanding of hypertriglyceridemia, genetic analysis could be extended to include individuals with mild-to-moderate hypertriglyceridemia. Finally, our study was limited to individuals of European ancestry, and may not be generalizable to other geographical ancestries, a shortcoming that is not unique to this study (Need and Goldstein, 2009; Popejoy and Fullerton, 2016). Given the emerging challenge of dyslipidemia in the developing world, it is crucial to evaluate hypertriglyceridemia patients of different geographical ancestries.

## 4.6 Conclusion

Here, we assessed genetic profiles of severe hypertriglyceridemia patients using our targeted NGS panel and bioinformatic tools. We report the most comprehensive and in-depth portrait of genetic determinants of severe hypertriglyceridemia to date. After a concurrent assessment of rare variants, both SNVs and CNVs, and the accumulation of common variants, we found that the accumulation of common variants was the most predominant genetic feature, and almost half of the patients had some type of polygenic determinant. Patients with bi-allelic rare variants (i.e. FCS) are a very rare subset of this phenotype. Nonetheless, there is a very strong genetic component underlying severe hypertriglyceridemia; this is clearly polygenic in large proportion of patients with severe hypertriglyceridemia.

## 4.7 References

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7, Unit 7* 20.

Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., *et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41, 47-55.

Babirak, S.P., Iverius, P.H., Fujimoto, W.Y., and Brunzell, J.D. (1989). Detection and characterization of the heterozygote state for lipoprotein lipase deficiency. *Arteriosclerosis* 9, 326-334.

Beaumont, J.L., Carlson, L.A., Cooper, G.R., Fejfar, Z., Fredrickson, D.S., and Strasser, T. (1970). Classification of hyperlipidaemias and hyperlipoproteinaemias. *Bull World Health Organ* 43, 891-915.

Brahm, A.J., and Hegele, R.A. (2013). Hypertriglyceridemia. *Nutrients* 5, 981-1001.

Brahm, A.J., and Hegele, R.A. (2015). Chylomicronaemia--current diagnosis and future therapies. *Nat Rev Endocrinol* 11, 352-362.

Cole, C.B., Nikpay, M., Lau, P., Stewart, A.F., Davies, R.W., Wells, G.A., Dent, R., and McPherson, R. (2014). Adiposity significantly modifies genetic risk for dyslipidemia. *J Lipid Res* 55, 2416-2422.

Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.

Daniels, T.F., Killinger, K.M., Michal, J.J., Wright, R.W., Jr., and Jiang, Z. (2009). Lipoproteins, cholesterol homeostasis and cardiac health. *Int J Biol Sci* 5, 474-488.

Douvris, A., Soubeyrand, S., Naing, T., Martinuk, A., Nikpay, M., Williams, A., Buick, J., Yauk, C., and McPherson, R. (2014). Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. *J Am Heart Assoc* 3, e000884.

Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.

Dron, J.S., and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.

Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhoads, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.

Feingold, K.R., and Grunfeld, C. (2000). Introduction to Lipids and Lipoproteins. In Endotext, A.B. Feingold KR, Boyce A, Chrousos G, de Herder WW, Dungan K, Grossman A, Hershman JM, Hofland HJ, Kaltsas G, Koch C, Kopp P, Korbonits M, McLachlan R, Morley JE, New M, Purnell J, Singer F, Stratakis CA, Trence DL, Wilson DP, ed. (South Dartmouth (MA): MDText.com, Inc.).

Futema, M., Shah, S., Cooper, J.A., Li, K., Whittall, R.A., Sharifi, M., Goldberg, O., Drogari, E., Mollaki, V., Wiegman, A., *et al.* (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. *Clin Chem* 61, 231-238.

Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68-74.

Hegele, R.A. (2018). Learning From Patients With Ultrarare Conditions: Cholesterol Hoof Beats. *J Am Coll Cardiol* 71, 289-291.

Hegele, R.A., Ban, M.R., Hsueh, N., Kennedy, B.A., Cao, H., Zou, G.Y., Anand, S., Yusuf, S., Huff, M.W., and Wang, J. (2009). A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Hum Mol Genet* 18, 4189-4194.

Hegele, R.A., Berberich, A.J., Ban, M.R., Wang, J., Digenio, A., Alexander, V.J., D'Erasmus, L., Arca, M., Jones, A., Bruckert, E., *et al.* (2018). Clinical and biochemical features of different molecular etiologies of familial chylomicronemia. *J Clin Lipidol* 12, 920-927 e924.

Hegele, R.A., Breckenridge, W.C., Cox, D.W., Maguire, G.F., Little, J.A., and Connelly, P.W. (1991). Interaction between variant apolipoproteins C-II and E that affects plasma lipoprotein concentrations. *Arterioscler Thromb Vasc Biol* 11, 1303-1309.

Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Aversa, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* 2, 655-666.

Holleboom, A.G., Karlsson, H., Lin, R.S., Beres, T.M., Sierts, J.A., Herman, D.S., Stroes, E.S., Aerts, J.M., Kastelein, J.J., Motazacker, M.M., *et al.* (2011). Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man. *Cell Metab* 14, 811-818.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011a). Genetic determinants of plasma triglycerides. *J Lipid Res* 52, 189-206.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in

genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* *42*, 684-687.

Johansen, C.T., Wang, J., Lanktree, M.B., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., Schwartz, S.M., *et al.* (2011b). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* *31*, 1916-1926.

Johansen, C.T., Wang, J., McIntyre, A.D., Martins, R.A., Ban, M.R., Lanktree, M.B., Huff, M.W., Peterfy, M., Mehrabian, M., Lusic, A.J., *et al.* (2012). Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* *5*, 66-72.

Justesen, J.M., Allin, K.H., Sandholt, C.H., Borglykke, A., Krarup, N.T., Grarup, N., Linneberg, A., Jorgensen, T., Hansen, T., and Pedersen, O. (2015). Interactions of Lipid Genetic Risk Scores With Estimates of Metabolic Health in a Danish Population. *Circ Cardiovasc Genet* *8*, 465-472.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* *41*, 56-65.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* *46*, 310-315.

Knowles, J.W., Xie, W., Zhang, Z., Chennamsetty, I., Assimes, T.L., Paananen, J., Hansson, O., Pankow, J., Goodarzi, M.O., Carcamo-Orive, I., *et al.* (2015). Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. *J Clin Invest* *125*, 1739-1751.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* *4*, 1073-1081.

Lambert, J.E., and Parks, E.J. (2012). Postprandial metabolism of meal triglyceride in humans. *Biochim Biophys Acta* *1821*, 721-726.

Latsuzbaia, A., Jaddoe, V.W., Hofman, A., Franco, O.H., and Felix, J.F. (2016). Associations of genetic variants for adult lipid levels with lipid levels in children. The Generation R Study. *J Lipid Res* *57*, 2185-2192.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., *et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* *536*, 285-291.

Lutsey, P.L., Rasmussen-Torvik, L.J., Pankow, J.S., Alonso, A., Smolenski, D.J., Tang, W., Coresh, J., Volcik, K.A., Ballantyne, C.M., Boerwinkle, E., *et al.* (2012). Relation of

lipid gene scores to longitudinal trends in lipid levels and incidence of abnormal lipid levels among individuals of European ancestry: the Atherosclerosis Risk in Communities (ARIC) study. *Circ Cardiovasc Genet* 5, 73-80.

Mathias, R.A., Pani, V., and Chilton, F.H. (2014). Genetic Variants in the FADS Gene: Implications for Dietary Recommendations for Fatty Acid Intake. *Curr Nutr Rep* 3, 139-148.

Motazacker, M.M., Peter, J., Treskes, M., Shoulders, C.C., Kuivenhoven, J.A., and Hovingh, G.K. (2013). Evidence of a polygenic origin of extreme high-density lipoprotein cholesterol levels. *Arterioscler Thromb Vasc Biol* 33, 1521-1528.

Nakayama, K., Yanagisawa, Y., Ogawa, A., Ishizuka, Y., Munkhtulga, L., Charupoonphol, P., Supannatas, S., Kuartei, S., Chimedregzen, U., Koda, Y., *et al.* (2011). High prevalence of an anti-hypertriglyceridemic variant of the MLXIPL gene in Central Asia. *J Hum Genet* 56, 828-833.

Need, A.C., and Goldstein, D.B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends Genet* 25, 489-494.

Nordestgaard, B.G., Abildgaard, S., Wittrup, H.H., Steffensen, R., Jensen, G., and Tybjaerg-Hansen, A. (1997). Heterozygous lipoprotein lipase deficiency: frequency in the general population, effect on plasma lipid levels, and risk of ischemic heart disease. *Circulation* 96, 1737-1744.

Piccolo, S.R., Abo, R.P., Allen-Brady, K., Camp, N.J., Knight, S., Anderson, J.L., and Horne, B.D. (2009). Evaluation of genetic risk scores for lipid levels using genome-wide markers in the Framingham Heart Study. *BMC Proc* 3 Suppl 7, S46.

Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. *Nature* 538, 161-164.

Pullinger, C.R., Stock, E.O., Movsesyan, I., Malloy, M.J., Frost, P.H., Tripuraneni, R., Quinn, A.G., Ishida, B.Y., Schaefer, E.J., Asztalos, B.F., *et al.* (2015). Identification and metabolic profiling of patients with lysosomal acid lipase deficiency. *J Clin Lipidol* 9, 716-726 e711.

Surendran, R.P., Visser, M.E., Heemelaar, S., Wang, J., Peter, J., Defesche, J.C., Kuivenhoven, J.A., Hosseini, M., Peterfy, M., Kastelein, J.J., *et al.* (2012). Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe hypertriglyceridaemia. *J Intern Med* 272, 185-196.

Talmud, P.J., Shah, S., Whittall, R., Futema, M., Howard, P., Cooper, J.A., Harrison, S.C., Li, K., Drenos, F., Karpe, F., *et al.* (2013). Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study. *Lancet* 381, 1293-1301.

- Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* *466*, 707-713.
- Tikka, A., and Jauhiainen, M. (2016). The role of ANGPTL3 in controlling lipoprotein metabolism. *Endocrine* *52*, 187-193.
- Tikkanen, E., Tuovinen, T., Widen, E., Lehtimäki, T., Viikari, J., Kahonen, M., Peltonen, L., Raitakari, O.T., and Ripatti, S. (2011). Association of known loci with lipid levels among children and prediction of dyslipidemia in adults. *Circ Cardiovasc Genet* *4*, 673-680.
- Viscarra, J.A., Wang, Y., Nguyen, H.P., Choi, Y.G., and Sul, H.S. (2020). Histone demethylase JMJD1C is phosphorylated by mTOR to activate de novo lipogenesis. *Nat Commun* *11*, 796.
- Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008). Polygenic determinants of severe hypertriglyceridemia. *Hum Mol Genet* *17*, 2894-2899.
- Wang, J., Dron, J.S., Ban, M.R., Robinson, J.F., McIntyre, A.D., Alazzam, M., Zhao, P.J., Dillio, A.A., Cao, H., Huff, M.W., *et al.* (2016). Polygenic Versus Monogenic Causes of Hypercholesterolemia Ascertained Clinically. *Arterioscler Thromb Vasc Biol* *36*, 2439-2445.
- Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* *45*, 1274-1283.



## Chapter 5 – Partial *LPL* deletions: rare copy-number variants contributing towards the polygenic form of severe hypertriglyceridemia

The work contained in this Chapter has been edited from its original publication in the *Journal of Lipid Research* for brevity and consistency throughout this Dissertation.

JOURNAL OF LIPID RESEARCH



Supplemental Material can be found at:  
<http://www.jlr.org/content/suppl/2019/09/13/jlr.P119000335>

*patient-oriented and epidemiological research*

### Partial *LPL* deletions: rare copy-number variants contributing towards severe hypertriglyceridemia<sup>§</sup>

Jacqueline S. Dron,<sup>\*,1</sup> Jian Wang,<sup>\*</sup> Adam D. McIntyre,<sup>\*</sup> Henian Cao,<sup>\*</sup> John F. Robinson,<sup>\*</sup> P. Barton Duell,<sup>§</sup> Priya Manjoo,<sup>\*\*</sup> James Feng,<sup>††</sup> Irina Movsesyan,<sup>††</sup> Mary J. Malloy,<sup>††</sup> Clive R. Pullinger,<sup>††</sup> John P. Kane,<sup>††</sup> and Robert A. Hegele<sup>\*,4,7,§§</sup>

Robarts Research Institute<sup>\*</sup> and Departments of Biochemistry<sup>†</sup> and Medicine,<sup>§§</sup> Schulich School of Medicine and Dentistry, Western University, London, ON N6A 5B7, Canada; Knight Cardiovascular Institute,<sup>§</sup> Oregon Health and Science University, Portland, OR 97239; Department of Medicine,<sup>\*\*</sup> Gordon and Leslie Diamond Centre, University of British Columbia, Vancouver, BC V5Z 1M9, Canada; and Cardiovascular Research Institute,<sup>††</sup> University of California San Francisco, San Francisco, CA 94158

ORCID IDs: 0000-0002-3045-6530 (J.S.D.); 0000-0003-2861-5325 (R.A.H.)

**Abstract** Severe hypertriglyceridemia (HTG) is a relatively common form of dyslipidemia with a complex pathophysiology and serious health complications. HTG can develop in the presence of rare genetic factors disrupting genes involved in the triglyceride (TG) metabolic pathway, including large-scale copy-number variants (CNVs). Improvements in next-generation sequencing technologies and bioinformatic analyses have better allowed assessment of CNVs as possible causes of or contributors to severe HTG. We screened targeted sequencing data of 632 patients with severe HTG and identified partial deletions of the *LPL* gene, encoding the central enzyme involved in the metabolism of TG-rich lipoproteins, in four individuals (0.63%). We confirmed the genomic breakpoints in each patient with Sanger sequencing. Three patients carried an identical heterozygous deletion spanning the 5' untranslated region (UTR) to *LPL* exon 2, and one patient carried a heterozygous deletion spanning the 5'UTR to *LPL* exon 1. All four heterozygous CNV carriers were determined to have multifactorial severe HTG. The predicted null nature of our identified *LPL* deletions may contribute to relatively higher TG levels and a more severe clinical phenotype than other forms of genetic variation associated with the disease, particularly in the polygenic state.<sup>§§</sup> The identification of novel CNVs in patients with severe HTG suggests that methods for CNV detection should be included in the diagnostic workup and molecular genetic

**evaluation of patients with high TG levels.**—Dron, J. S., Wang, A. D., McIntyre, H. Cao, J. F. Robinson, P. B. Duell, P. Manjoo, J. Feng, I. Movsesyan, M. J. Malloy, C. R. Pullinger, J. P. Kane, and R. A. Hegele. Partial *LPL* deletions: rare copy-number variants contributing towards severe hypertriglyceridemia. *J. Lipid Res.* 2019, 60: 1953–1958.

**Supplementary key words** bioinformatic analysis • diagnostic tools • dyslipidemia • genetic testing • human genetics • next-generation sequencing • triglycerides • lipoprotein lipase

Elevations in fasting plasma triglyceride (TG) levels are diagnosed as hypertriglyceridemia (HTG). TG levels  $\geq 10$  mmol/L (885 mg/dl) are classified as severe HTG (1) and are seen in  $\sim 1$  in 600 individuals (2). As a relatively common form of dyslipidemia with serious health complications that include pancreatitis (2, 3), there is a focus on identifying and understanding factors that can increase susceptibility to or cause severe HTG.

A combination of rare single-nucleotide variants and common SNPs can contribute permissively or causally toward the presentation of this complex disease (3). The monogenic form of severe HTG, also referred to as familial chylomicronemia syndrome (FCS), is caused by bi-allelic variants disrupting canonical genes involved in TG metabolism, such as *LPL*, *APOC2*, *APOA5*, glycosylphosphatidylinositol-anchored HDL-binding protein 1 (*GPIHBP1*), or lipase

*J.S.D.* is supported by the Canadian Institutes of Health Research (Doctoral Research Award); *R.A.H.* is supported by the Jacob J. Wolfe Distinguished Medical Research Chair, the Edith Schulich Vice-Chancellor Research Chair in Human Genetics, the Martha C. Blackburn Chair in Cardiovascular Research, and operating grants from the Canadian Institutes of Health Research (Translational Grant) and the Heart and Stroke Foundation of Ontario (G-18-002147). The work at University of California San Francisco was supported by the Joseph Drown Foundation, the Campbell Foundation, and gifts from Peter Head, Harold Dittmer, Susan Hoang, and Donald Yellon. *R.A.H.* reports consulting fees from Acosta, Argenson, Alkermes/Lonis, Amgen, HHS Therapeutics, and Sanofi. *P.B.D.* reports consulting fees and/or institutional research grants from Abnova, Astra Zeneca, Esperion, Regeneron, RegenXBio, and Retimpin. The other authors have no disclosures.

Manuscript received 11 August 2019 and in revised form 9 September 2019.  
Published, JLR Papers in Press, September 13, 2019  
DOI <https://doi.org/10.1194/jlr.P119000335>

Copyright © 2019 Dron et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.  
This article is available online at <http://www.jlr.org>

Journal of Lipid Research Volume 60, 2019 1953

Downloaded from [www.jlr.org](http://www.jlr.org) at Univ Western Ontario - OJDL on October 14, 2020

**Dron, J.S., Wang, J., McIntyre, A.D., Cao, H., Robinson, J.F., Duell, P.B., Manjoo, P., Feng, J., Movsesyan, I., Malloy, M.J., et al. (2019). Partial *LPL* deletions: rare copy-number variants contributing towards severe hypertriglyceridemia. *J Lipid Res* 60, 1953–1958.**

## 5.1 Abstract

**Objective:** Severe hypertriglyceridemia is a relatively common form of dyslipidemia with a complex pathophysiology and serious health complications. Hypertriglyceridemia can develop in the presence of rare genetic factors disrupting genes involved in the triglyceride metabolic pathway, including large-scale copy-number variants (CNVs). Improvements in next-generation sequencing (NGS) technologies and bioinformatic analyses have better allowed assessment of CNVs as possible causes of or contributors to severe hypertriglyceridemia.

**Methods and Results:** We screened targeted NGS data of 632 patients with severe hypertriglyceridemia and identified partial deletions of the *LPL* gene, encoding the central enzyme involved in the metabolism of triglyceride-rich lipoproteins, in four individuals (0.63%). We confirmed the genomic breakpoints in each patient with Sanger sequencing. Three patients carried an identical heterozygous deletion spanning the 5' untranslated region (UTR) to *LPL* exon 2, and one patient carried a heterozygous deletion spanning the 5'UTR to *LPL* exon 1. All four heterozygous CNV carriers were determined to have the polygenic form of severe hypertriglyceridemia (i.e. multifactorial chylomicronemia).

**Conclusion:** The predicted null nature of our identified *LPL* deletions may contribute to relatively higher triglyceride levels and a more severe clinical phenotype than other forms of genetic variation associated with the disease, particularly in the polygenic state. The identification of novel CNVs in patients with severe hypertriglyceridemia suggests that methods for CNV detection should be included in the diagnostic workflow and genetic evaluation of patients with high triglyceride levels.

## 5.2 Introduction

Elevations in fasting plasma triglyceride levels are diagnosed as hypertriglyceridemia; triglyceride levels  $\geq 10$  mmol/L are classified as “severe” hypertriglyceridemia (Hegele et al., 2014) and are seen in ~1 in 600 individuals (Dron and Hegele, 2017). As a relatively common form of dyslipidemia with serious health complications that include pancreatitis

(Brahm and Hegele, 2015; Dron and Hegele, 2017), there is a focus on identifying and understanding factors that can increase susceptibility or cause severe hypertriglyceridemia.

A combination of rare single-nucleotide variants (SNVs) and common single-nucleotide polymorphisms (SNPs) can contribute permissively or causally towards the presentation of this complex disease (Brahm and Hegele, 2015). The monogenic form of severe hypertriglyceridemia—also referred to as familial chylomicronemia syndrome (FCS)—is caused by bi-allelic variants disrupting canonical genes involved in triglyceride metabolism, such as *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2* (Johansen et al., 2011a). Conversely, increased susceptibility for the polygenic form of severe hypertriglyceridemia, called “multifactorial chylomicronemia”, is due to heterozygous rare variants, common triglyceride-raising alleles at certain SNP loci, or a combination of both (Dron et al., 2019; Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Kathiresan et al., 2009; Surendran et al., 2012; Teslovich et al., 2010; Wang et al., 2008; Willer et al., 2013).

Previously, it has been shown that copy-number variants (CNVs) are an additional type of genetic variation that can markedly contribute to extreme perturbations of triglyceride levels (Benlian et al., 1995; Devlin et al., 1990; Langlois et al., 1989; Okubo et al., 2007), as well as other lipid traits and disorders (Dron et al., 2018; Iacocca et al., 2018a; Iacocca et al., 2019; Iacocca and Hegele, 2018; Iacocca et al., 2018b). Assessment of CNVs is becoming easier due to improvements in sequencing technologies and bioinformatic analysis tools (Iacocca and Hegele, 2018; Valsesia et al., 2013). Because of this, it is possible to screen for CNVs in patient samples concurrently with rare SNVs and SNPs (Iacocca et al., 2019), and assess them as possible causes or contributors towards severe hypertriglyceridemia.

A previous study of 563 patients with severe hypertriglyceridemia led to the identification of one individual who was likely carrying a heterozygous CNV deletion in *LPL* (Dron et al., 2019). From our next-generation sequencing (NGS) method and data archive (Johansen et al., 2014), we expanded our search for additional *LPL* CNVs that

might be contributing towards the presentation of severe hypertriglyceridemia in a larger cohort of patients. We discovered a total of four out of 632 patients with severe hypertriglyceridemia who were heterozygous carriers for one of two novel CNV deletions disrupting *LPL*. We molecularly confirm and characterize each deletion and discuss their likely contribution to severe hypertriglyceridemia.

## 5.3 Materials and Methods

### 5.3.1 Study subjects

Severe hypertriglyceridemia patients (defined as triglyceride levels  $\geq 10$  mmol/L on at least one occasion) from the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada), the Genomic Resource in Arteriosclerosis and Metabolic Disease recruited at the Lipid, Diabetes, or Cardiology Clinics (University of California, San Francisco CA, USA), or patient samples directly from collaborating research centres were screened for CNVs. Patients provided signed consent with approval from the Western University ethics review board (no. 07290E) or from the originating institution.

### 5.3.2 DNA preparation and targeted sequencing

DNA isolation and preparation for targeted NGS follows the same methodology as described in [Chapter 2, Section 2.3.2](#).

### 5.3.3 Bioinformatic processing of sequencing data

The bioinformatic processing of sequencing data follows the same methodology as described in [Chapter 3, Section 3.3.3](#); however, an updated version of CLC Bio Genomics Workbench (version 12.0; CLC Bio, Aarhus, Denmark) was used.

### 5.3.4 Detection of copy-number variants

CNVs were detected following the same methodology described in [Chapter 3, Section 3.3.4](#); however, an updated version of VarSeq® (version 2.1.0; Golden Helix, Inc., Bozeman MT, USA) was used. The *LPL* gene was specifically screened for CNVs.

### 5.3.5 Validation of partial gene deletions

#### 5.3.5.1 Breakpoint identification

To confirm each deletion, we designed primers to flank the regions likely to contain the deletions and used them for PCR amplification (Expand 20 kb<sup>plus</sup> PCR System, Sigma-Aldrich St. Louis MO, USA, cat. No. 11811002001). The forward (F) and reverse (R) primers used were: F1 5'-TACAAGACGGTGTGTTGTGTTGTGGCACGG-3' and R1 5'-GTGACTTGATCCACAGCACAGAGCTGGAG-3' (5' untranslated region [UTR] – exon 1 deletion); F2 5'-AAGCTAGCTAGCTAGCTGGCTGGCCAG-3' and R2 5'-GGGTCTCTTGCAGCTAAGTCAGAACTCCAG-3' (5'UTR – exon 2 deletion). PCR products were run on a gel for visual confirmation of the mutant alleles. Sanger sequencing and primer-walking of the PCR products were performed to identify the deletion breakpoints.

#### 5.3.5.2 Sanger confirmation

After identifying deletion breakpoints by primer-walking the PCR products, screening primers spanning the proximal or distal breakpoint were designed for PCR and Sanger sequencing (**Appendix I**).

## 5.4 Results

### 5.4.1 Study subjects

A total of 632 patients with severe hypertriglyceridemia were screened for CNVs disrupting *LPL*. We identified four individuals (**Table 5.1**) who were carriers for partial deletions in *LPL* using the VarSeq-CNV® caller algorithm (**Figure 5.1**).

### 5.4.2 *LPL* copy-number variant detection

Subject 1 was detected as carrying a heterozygous deletion of the 5'UTR to exon 1. From our LipidSeq panel, the CNV was detected to cover a single probe, and had an average ratio of 0.504 and average Z score of -13.030.

Subjects 2, 3, and 4 were all detected as carrying a heterozygous deletion of the 5'UTR to exon 2; the observation of Subject 4's CNV was first reported by our group earlier this

year (Dron et al., 2019). From our LipidSeq panel, the CNV was detected to cover two probes. Subject 2 had an average ratio of 0.566 and average Z score of -7.058. Subject 3 had an average ratio of 0.542 and average Z score of -9.713. Subject 4 had an average ratio of 0.546 and average Z score of -7.248.

#### 5.4.3 Copy-number variant validation and identifying breakpoints

A combination of PCR primer-walking upstream and downstream of the putative CNVs and gel electrophoresis validated the deletions and allowed for their characterization (**Table 5.2**). The deletion in Subject 1 was found to be 5,917 bp in size. This deletion began 1,038 bp upstream of *LPL*, covered the 5'UTR and exon 1, and ended 4,420 bp downstream of the splice donor site in intron 1 (**Figure 5.2**). Subjects 2, 3, and 4 were found to have the exact same deletion, which was 11,598 bp in size. This deletion began 1,432 bp upstream of *LPL*, covered the 5'UTR, exon 1 and exon 2, and ended 895 bp downstream of the splice donor in intron 2 (**Figure 5.2**). We currently do not have any information suggesting that these three individuals are related.



**Figure 5.1 Identification of *LPL* CNVs using the VarSeq-CNV® caller algorithm on targeted sequencing data.**

Chr8:19,795,931-19,829,369 (hg19 genome build) is the region visualized in each panel. **A)** Subject 1, carrier of a heterozygous deletion spanning the 5'UTR and exon 1 of *LPL*. **B)** Subject 2, 3, and 4, carriers of a heterozygous deletion spanning the 5'UTR, exon 1 and exon 2 of *LPL*. Abbreviations: chr = chromosome; CNV = copy-number variant; het = heterozygous.

**Table 5.1 Clinical and demographic features of subjects with *LPL* CNVs.**

	<b>Subject 1</b>	<b>Subject 2</b>	<b>Subject 3</b>	<b>Subject 4</b>
<b>Age</b>	53	48	64	46
<b>Sex</b>	Male	Male	Male	Female
<b>BMI (kg/m<sup>2</sup>)</b>	30.3	28.9	31.2	34.6
<b>Race</b>	White and Hispanic	White	White	White
<b>Total cholesterol (mmol/L)</b>	10.1	4.98	7.54	17.4
<b>Triglyceride (mmol/L)</b>	36.1	16.7	35.9	36.4
<b>HDL cholesterol (mmol/L)</b>	0.59	0.76	0.45	0.39
<b>LDL cholesterol (mmol/L)</b>	1.76	-	-	1.24
<b>apo B (g/L)</b>	1.28	0.69	0.84	4.44
<b>Fasting glucose (mmol/L)</b>	10.0	6.3	10.0	11.0
<b>Co-morbidities</b>	Acute pancreatitis x3; pancreatic pseudocyst; type 2 diabetes; carotid and aortoiliac plaque; hepatic steatosis, gout; historically highest triglyceride was 102 mmol/L	Herpes zoster; impaired glucose tolerance; hepatosteatorosis	Acute pancreatitis x3; type 2 diabetes; CABG, MI x2; gout	Acute pancreatitis; gallstones (cholecystectomy)
<b>Identified genetic factors</b>	<i>LPL</i> exon 1 deletion (het); common <i>LPL</i> p.D36N variant (het); normal polygenic risk score (<64 <sup>th</sup> percentile)	<i>LPL</i> exon 1-2 deletion (het); normal polygenic risk score (<43 <sup>rd</sup> percentile)	<i>LPL</i> exon 1-2 deletion (het); normal polygenic risk score (<77 <sup>th</sup> percentile)	<i>LPL</i> exon 1-2 deletion (het); normal polygenic risk score (<31 <sup>st</sup> percentile)

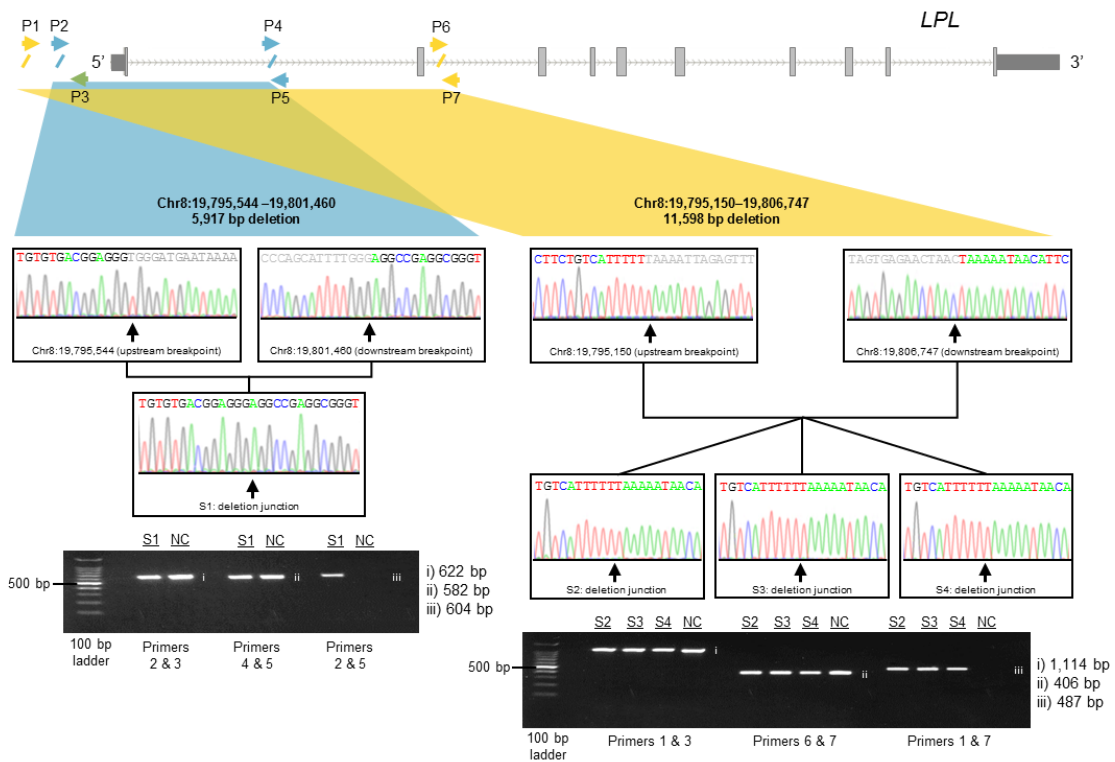
Values provided are from first presentation to specialist lipid clinic, or date first obtained. Abbreviations: apo = apolipoprotein; BMI = body-mass index; CABG = coronary artery bypass graft; HDL = high-density lipoprotein; het = heterozygous; LDL = low-density lipoprotein; MI = myocardial infarction.



**Table 5.2 Genomic coordinates and breakpoints of *LPL* CNVs.**

CNV	Zygoty state	Breakpoint		
		Genomic coordinates	Length (bp)	HGVS notation
5'UTR to exon 1	Het	chr8:19,795,544 to chr8:19,801,460	5,917	g.19795544-19801460del c. 1_88del p. Met1?
5' UTR to exon 2	Het	chr8:19,795,150 to chr8:19,806,747	11,598	g.19795150-19806747del c.1_249del p.Met1?

The sequences are in the forward-strand orientation. Abbreviations: bp = base pair; chr = chromosome; CNV = copy-number variation; het = heterozygous; HGVS = Human Genome Variation Society.



**Figure 5.2 Validation of deletions disrupting *LPL* in patients with severe hypertriglyceridemia.**

The *LPL* gene transcript with the approximate breakpoints of the smaller CNV deletion encompassing exon 1 (left) and the larger CNV deletion encompassing exons 1 and 2 (right) are indicated in blue and yellow, respectively. The diagonal slashes along the transcript indicate sequence breakpoints, while the arrows demonstrate the position and orientation of primers used in breakpoint identification and Sanger sequencing. Gel electrophoresis of PCR products across upstream and downstream breakpoints, and deletion junctions for each deletion are shown for Subjects 1 to 4. The primer pairs used for each PCR are indicated underneath the corresponding gel lanes. Abbreviations: bp = base pair; chr = chromosome; NC = normal control; P = primer; S = subject.

## 5.5 Discussion

Out of 632 patients with severe hypertriglyceridemia, four (0.63%) were identified as carriers of one of two unique, partial gene deletions in *LPL*. CNVs involving *LPL*—both deletions and duplications—have previously been identified using older methods (Benlian et al., 1995; Devlin et al., 1990; Langlois et al., 1989; Okubo et al., 2007), but to our knowledge this is one of the first few reports identifying and characterizing *LPL* CNVs using an NGS-based bioinformatic method, with confirmation of the genomic breakpoints. A recent study identified an *LPL* CNV deletion in an individual with severe hypertriglyceridemia last year using different NGS-based methods (Marmontel et al., 2018).

LPL is the primary enzyme responsible for the hydrolysis of triglyceride-rich lipoproteins, such as chylomicrons and very-low-density lipoproteins (VLDL) (Olivecrona, 2016; Young and Zechner, 2013). After being chaperoned by lipase maturation factor 1 (LMF1) from parenchymal cells to endothelial cells, LPL is anchored to the vascular lumen by glycosylphosphatidylinositol-anchored high-density lipoprotein-binding protein 1 (GPIHBP1) (Young and Zechner, 2013). From there, LPL binds to the apolipoprotein (apo) C-II component of circulating triglyceride-rich lipoproteins to initiate the catabolism of their triglyceride-rich cores (Young and Zechner, 2013). Molecular disruptions that impair LPL mobilization or activity lead to an overall decrease in the hydrolysis of triglyceride. With fewer triglyceride-rich lipoproteins being catabolized, there is a resultant increase in the circulating concentration of triglyceride, which is the defining feature of hypertriglyceridemia.

Considering the two identified CNVs spanning the 5'UTR to exon 1 and the 5'UTR to exon 2 both delete the initiator codon, it is almost certain that these CNVs are null mutations (Walter et al., 2005). However, the exact molecular consequences of these partial gene deletions cannot be confirmed without functional data related to mRNA expression, protein expression, or protein function. Since only heterozygous deletions were found, each patient can be classified as presenting with the polygenic form of hypertriglyceridemia (Brahm and Hegele, 2015; Dron et al., 2019; Wang et al., 2008)

with additional factors—either genetic or environmental or both—contributing to their clinical phenotype (Cole et al., 2015); this can also be referred to as “multifactorial chylomicronemia”. Interestingly, none of these patients have a high polygenic risk score or any other rare variants in canonical triglyceride metabolism genes. Although the sizes of the CNVs are quite large and the reported triglyceride levels are extremely high, these patients are not considered to have FCS, which refers specifically to a highly penetrant autosomal recessive disease. Only individuals with bi-allelic variants disrupting one of the canonical triglyceride metabolism genes can be diagnosed with FCS.

To our knowledge, these particular *LPL* CNV deletions have never been reported. Overall, publications on *LPL* CNVs have been infrequent. In 1989, Langlois et al. identified several *LPL*-deficient individuals with either a 2-kb insertion or a 6-kb deletion in *LPL* using Southern blotting (Langlois et al., 1989). The next year, Devlin et al. further characterized the insertion and showed that it was a 2-kb tandem duplication event disrupting exon 6 of *LPL* (Devlin et al., 1990). Some years later, the first report of a homozygous CNV deletion in *LPL* was reported by Benlian et al. who used a PCR-based approach to define a 2.1-kb deletion encompassing exon 9 and flanking intronic sequence in a patient with *LPL* deficiency (Benlian et al., 1995). The next report on a *LPL* CNV was published more than a decade later, when Okubo et al. described a complex deletion-insertion event (Okubo et al., 2007). By using both Southern blot analysis and PCR, they found their *LPL*-deficient proband was a homozygous carrier for a 2.3-kb deletion across exon 2 and 150 bp insertion at the break junction (Okubo et al., 2007). When considering more modern detection methods, a recent study by Marmontel et al. identified a heterozygous *LPL* deletion of exons 3 to 7 in a young patient with severe hypertriglyceridemia (triglyceride = 87 mmol/L); this individual also carried a heterozygous SNV in *LPL* (c.642A>C) and was classified as having a bi-allelic variant, and thus was diagnosed with FCS (Marmontel et al., 2018).

Given the rarity of *LPL* CNVs, it was interesting to find the same deletion in three of our patients, Subjects 2 to 4, who have no known relationship between them. Since their deletion breakpoints are identical by sequencing, it is possible that these individuals have a distant common ancestor who carried the CNV. Although these patients presently live

in different geographical locations, they all self-report similar ancestry. An alternative—albeit far less likely—explanation is that the exact same CNV event occurred independently in each separate patient lineage. For this deletion, there is sequence homology and repeated sequence around the breakpoint junction, which increases the likelihood of slippage, replication errors, and CNV events (Hastings et al., 2009). Despite having features that promote CNV events, the rarity of this *LPL* deletion in the literature and public databases suggests that this CNV is more likely shared by a common ancestor, rather than a reoccurring, independent deletion.

When considering triglyceride levels in these patients, we noted that Subject 1 who had the smallest CNV also had the highest measured triglyceride levels at 102 mmol/L, while Subjects 2 to 4 who shared the larger CNV had somewhat lower triglyceride measurements ranging between 16.7 mmol/L and 36.4 mmol/L. It is unclear as to whether CNV size corresponds to magnitude in triglyceride elevation, or if it gives any indication for the function of the resultant protein product. Overall, the patients ranged from 48 to 53 years old and presented with a variety of co-morbidities. Interestingly, Subjects 1, 3 and 4 had reported past instances of acute pancreatitis requiring hospitalization; Subjects 1 and 3 each had three reported episodes. Acute pancreatitis has heretofore been a more frequent manifestation among individuals with the monogenic form of severe hypertriglyceridemia (Paquette et al., 2019). Given that these four patients almost certainly have polygenic hypertriglyceridemia, we speculate these predicted null mutations may have predisposed to relatively higher triglyceride levels than other types of genetic variation. Without functional studies and larger cohorts, it is difficult to isolate the CNV-specific effects. Disparities in genotype-phenotype relationships have previously been observed with CNVs underlying depressed high-density lipoprotein (HDL) cholesterol levels, in which the same genetic variants were found in individuals with variable lipid profiles (Brunham et al., 2006; Dron et al., 2018).

The reported non-genetic factors, including co-morbidities such as diabetes (**Table 5.1**), are likely contributing towards the overall severity of these patients' hypertriglyceridemia phenotypes, and in turn may help to explain the frequency of acute pancreatitis episodes in these individuals. By considering these additional pieces of information, we can more

specifically diagnose our patients with multifactorial chylomicronemia. As stated previously, a single heterozygous variant is not enough to cause hypertriglyceridemia; given that the patients did not have any additional related genetic factors identified as contributing towards their phenotypes, these non-genetic factors must be considered as likely contributory factors.

Future studies are required to characterize the functional impact of our identified CNVs on LPL activity and triglyceride clearance pathways. Moving forward, it is also important to screen for CNVs in the other canonical triglyceride metabolism genes, as they have been previously identified in individuals with hypertriglyceridemia, such as CNVs disrupting *GPIHBP1* and *APOC2* (Hegele et al., 2018; Patni et al., 2016; Rios et al., 2012).

## 5.6 Conclusion

In summary, although they are relatively infrequent, *LPL* CNVs are an important type of genetic variation that should be screened for when establishing the genetic basis of hypertriglyceridemia, given their disruptive nature. With developments and improvements to NGS techniques and more accessible CNV detection methods, CNV assessment can be easily incorporated into routine screens of rare SNVs and polygenic risk score calculations (Iacocca et al., 2019; Iacocca and Hegele, 2018). Efforts must be taken to carefully characterize different determinants, including CNVs, SNVs, and the accumulation of SNPs. By assessing a larger spectrum of genetic factors, we can achieve a more comprehensive understanding of the genetic etiology underlying severe hypertriglyceridemia.

## 5.7 References

- Benlian, P., Etienne, J., de Gennes, J.L., Noe, L., Brault, D., Raisonnier, A., Arnault, F., Hamelin, J., Foubert, L., Chuat, J.C., *et al.* (1995). Homozygous deletion of exon 9 causes lipoprotein lipase deficiency: possible intron-Alu recombination. *J Lipid Res* 36, 356-366.
- Brahm, A.J., and Hegele, R.A. (2015). Chylomicronaemia--current diagnosis and future therapies. *Nat Rev Endocrinol* 11, 352-362.
- Brunham, L.R., Singaraja, R.R., and Hayden, M.R. (2006). Variations on a gene: rare and common variants in ABCA1 and their impact on HDL cholesterol levels and atherosclerosis. *Annu Rev Nutr* 26, 105-129.
- Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.
- Devlin, R.H., Deeb, S., Brunzell, J., and Hayden, M.R. (1990). Partial gene duplication involving exon-Alu interchange results in lipoprotein lipase deficiency. *Am J Hum Genet* 46, 112-119.
- Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.
- Dron, J.S., Wang, J., Berberich, A.J., Iacocca, M.A., Cao, H., Yang, P., Knoll, J., Tremblay, K., Brisson, D., Netzer, C., *et al.* (2018). Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia. *J Lipid Res* 59, 1529-1535.
- Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., *et al.* (2019). Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol* 13, 80-88.
- Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhinds, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* 10, 551-564.
- Hegele, R.A., Berberich, A.J., Ban, M.R., Wang, J., Digenio, A., Alexander, V.J., D'Erasmus, L., Arca, M., Jones, A., Bruckert, E., *et al.* (2018). Clinical and biochemical features of different molecular etiologies of familial chylomicronemia. *J Clin Lipidol* 12, 920-927 e924.
- Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Aversa, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* 2, 655-666.

Iacocca, M.A., Chora, J.R., Carrie, A., Freiberger, T., Leigh, S.E., Defesche, J.C., Kurtz, C.L., DiStefano, M.T., Santos, R.D., Humphries, S.E., *et al.* (2018a). ClinVar database of global familial hypercholesterolemia-associated DNA variants. *Hum Mutat* 39, 1631-1640.

Iacocca, M.A., Dron, J.S., and Hegele, R.A. (2019). Progress in finding pathogenic DNA copy number variations in dyslipidemia. *Curr Opin Lipidol* 30, 63-70.

Iacocca, M.A., and Hegele, R.A. (2018). Role of DNA copy number variation in dyslipidemias. *Curr Opin Lipidol* 29, 125-132.

Iacocca, M.A., Wang, J., Dron, J.S., Robinson, J.F., McIntyre, A.D., Cao, H., and Hegele, R.A. (2017). Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia. *J Lipid Res* 58, 2202-2209.

Iacocca, M.A., Wang, J., Sarkar, S., Dron, J.S., Lagace, T., McIntyre, A.D., Lau, P., Robinson, J.F., Yang, P., Knoll, J.H., *et al.* (2018b). Whole-Gene Duplication of PCSK9 as a Novel Genetic Mechanism for Severe Familial Hypercholesterolemia. *Can J Cardiol* 34, 1316-1324.

Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* 55, 765-772.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011a). Genetic determinants of plasma triglycerides. *J Lipid Res* 52, 189-206.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42, 684-687.

Johansen, C.T., Wang, J., Lanktree, M.B., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., Schwartz, S.M., *et al.* (2011b). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 31, 1916-1926.

Johansen, C.T., Wang, J., McIntyre, A.D., Martins, R.A., Ban, M.R., Lanktree, M.B., Huff, M.W., Peterfy, M., Mehrabian, M., Lusk, A.J., *et al.* (2012). Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 5, 66-72.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41, 56-65.



Langlois, S., Deeb, S., Brunzell, J.D., Kastelein, J.J., and Hayden, M.R. (1989). A major insertion accounts for a significant proportion of mutations underlying human lipoprotein lipase deficiency. *Proc Natl Acad Sci USA* 86, 948-952.

Marmontel, O., Charriere, S., Simonet, T., Bonnet, V., Dumont, S., Mahl, M., Jacobs, C., Nony, S., Chabane, K., Bozon, D., *et al.* (2018). Single, short in-del, and copy number variations detection in monogenic dyslipidemia using a next-generation sequencing strategy. *Clin Genet* 94, 132-140.

Okubo, M., Horinishi, A., Saito, M., Ebara, T., Endo, Y., Kaku, K., Murase, T., and Eto, M. (2007). A novel complex deletion-insertion mutation mediated by Alu repetitive elements leads to lipoprotein lipase deficiency. *Mol Genet Metab* 92, 229-233.

Olivecrona, G. (2016). Role of lipoprotein lipase in lipid metabolism. *Curr Opin Lipidol* 27, 233-241.

Paquette, M., Bernard, S., Hegele, R.A., and Baass, A. (2019). Chylomicronemia: Differences between familial chylomicronemia syndrome and multifactorial chylomicronemia. *Atherosclerosis* 283, 137-142.

Patni, N., Brothers, J., Xing, C., and Garg, A. (2016). Type 1 hyperlipoproteinemia in a child with large homozygous deletion encompassing GPIHBP1. *J Clin Lipidol* 10, 1035-1039 e1032.

Rios, J.J., Shastry, S., Jasso, J., Hauser, N., Garg, A., Bensadoun, A., Cohen, J.C., and Hobbs, H.H. (2012). Deletion of GPIHBP1 causing severe chylomicronemia. *J Inherit Metab Dis* 35, 531-540.

Surendran, R.P., Visser, M.E., Heemelaar, S., Wang, J., Peter, J., Defesche, J.C., Kuivenhoven, J.A., Hosseini, M., Peterfy, M., Kastelein, J.J., *et al.* (2012). Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe hypertriglyceridaemia. *J Intern Med* 272, 185-196.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.

Valsesia, A., Mace, A., Jacquemont, S., Beckmann, J.S., and Kutalik, Z. (2013). The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation. *Front Genet* 4, 92.

Walter, M.C., Witt, T.N., Weigel, B.S., Reilich, P., Richard, P., Pongratz, D., Bonne, G., Wehnert, M.S., and Lochmuller, H. (2005). Deletion of the LMNA initiator codon leading to a neurogenic variant of autosomal dominant Emery-Dreifuss muscular dystrophy. *Neuromuscul Disord* 15, 40-44.

Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008). Polygenic determinants of severe hypertriglyceridemia. *Hum Mol Genet* 17, 2894-2899.

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.

Young, S.G., and Zechner, R. (2013). Biochemistry and pathophysiology of intravascular and intracellular lipolysis. *Genes Dev* 27, 459-484.

## Chapter 6 – The polygenic nature of mild-to-moderate hypertriglyceridemia

The work contained in this Chapter has been edited from its original publication in the *Journal of Clinical Lipidology* for brevity and consistency throughout this Dissertation.



**Dron, J.S., Wang, J., McIntyre, A.D., Cao, H., and Hegele, R.A. (2020).** The polygenic nature of mild-to-moderate hypertriglyceridemia. *J Clin Lipidol* 14, 28-34 e22.

## 6.1 Abstract

**Objective:** Patients with mild-to-moderate hypertriglyceridemia are thought to share specific genetic susceptibility factors that are also present in severe hypertriglyceridemia patients, but no data have been reported on this issue. Here, we characterized genetic profiles of mild-to-moderate hypertriglyceridemia patients and compared them to patients with severe hypertriglyceridemia.

**Methods and Results:** DNA from patients with mild-to-moderate hypertriglyceridemia was sequenced using our targeted sequencing panel, “LipidSeq”. For each patient, we assessed: 1) rare variants disrupting five triglyceride metabolism genes; and 2) the accumulation of 16 common single-nucleotide polymorphisms (SNPs) using a polygenic risk score. The genetic profiles for these patients were then compared to normolipidemic controls and to patients with severe hypertriglyceridemia. Across 134 mild-to-moderate hypertriglyceridemia patients, 9.0% carried heterozygous rare variants and 24.6% had an excess accumulation of common SNPs. Mild-to-moderate hypertriglyceridemia patients were 2.38-times (95% CI [1.13-4.99];  $P=0.021$ ) more likely to carry a rare variant and 3.26-times (95% CI [2.02-5.26];  $P<0.0001$ ) more likely to have an extreme polygenic risk score compared to normolipidemic controls from the 1000 Genomes Project. In addition, severe hypertriglyceridemia patients were 1.86-times (95% CI [0.98-3.51];  $P=0.032$ ) more likely to carry a rare variant and 1.63-times (95% CI [1.07-2.48];  $P=0.013$ ) more likely to have an extreme polygenic risk score compared to mild-to-moderate hypertriglyceridemia patients.

**Conclusions:** We report an increased prevalence of genetic determinants in patients with an increased severity of the hypertriglyceridemia phenotype when considering either rare variants disrupting triglyceride metabolism genes or an excess accumulation of common SNPs. As well, the findings confirm that the most prevalent genetic contributor to hypertriglyceridemia, regardless of severity, is polygenic SNP accumulation.

## 6.2 Introduction

As a common dyslipidemia encountered in the clinic, hypertriglyceridemia is defined by elevated fasting triglyceride levels. Depending on the degree of elevation, individuals can present with mild-to-moderate hypertriglyceridemia (range 2-9.9 mmol/L) or severe hypertriglyceridemia ( $\geq 10$  mmol/L) (Hegele et al., 2014). Numerous genetic determinants contribute to susceptibility and presentation of hypertriglyceridemia (Brahm and Hegele, 2015). In severe hypertriglyceridemia patients with the monogenic recessive form of the disease—familial chylomicronemia syndrome (FCS)—bi-allelic rare variants disrupting canonical genes in the triglyceride metabolic pathway—including *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*—are casual factors (Johansen et al., 2011a). There is no reported autosomal dominant form of hypertriglyceridemia. By contrast, most individuals with severe hypertriglyceridemia have a complex polygenic predisposition; genetic susceptibility results from either incompletely penetrant heterozygous rare variants disrupting the aforementioned triglyceride metabolism genes, or the incremental effects from the accumulation of common triglyceride-associated single-nucleotide polymorphisms (SNPs), or a combination of these genetic factors (Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Kathiresan et al., 2009; Surendran et al., 2012; Teslovich et al., 2010; Wang et al., 2008; Willer et al., 2013). This is referred to as multifactorial chylomicronemia.

Previously, we characterized the genetic determinants underlying severe hypertriglyceridemia in a cohort of 563 patients and found: 1) 1.1% of cases were monogenic due to bi-allelic rare variants; and 2) 14.4% of cases carried heterozygous rare variants of variable penetrance; and 3) 32.0% had an excess accumulation of common SNPs (Dron et al., 2019). While that study advanced our understanding of the genetic profiles of severe hypertriglyceridemia patients, the genetic profiles of mild-to-moderate hypertriglyceridemia patients have not been examined.

Despite having lower triglyceride levels compared to those with severe hypertriglyceridemia, patients with mild-to-moderate hypertriglyceridemia also have health concerns, including an increased risk for cardiovascular disease (Dron and Hegele, 2017). The main disturbance in mild-to-moderate hypertriglyceridemia patients is an

excess of very-low-density lipoproteins (VLDL) and their remnants, including intermediate-density lipoproteins (IDL) (Brahm and Hegele, 2015; Varbo and Nordestgaard, 2016), which are considered atherogenic. Fasting chylomicrons are usually absent in mild-to-moderate hypertriglyceridemia patients, while in severe hypertriglyceridemia patients, chylomicrons are present typically together with excess VLDL and remnant particles (Chait and Brunzell, 1992; Dron et al., 2017; Lewis et al., 2015). Some patients with mild-to-moderate hypertriglyceridemia can deteriorate into severe hypertriglyceridemia when excess VLDL saturates triglyceride removal mechanisms, such that incoming chylomicrons cannot be cleared and thus accumulate pathologically (Chait and Eckel, 2019). It has been assumed that patients with mild-to-moderate hypertriglyceridemia share particular genetic susceptibility factors with severe hypertriglyceridemia patients. The phenotype can be further worsened by secondary factors such as diabetes, obesity, poor diet or alcohol use. Genetically characterizing mild-to-moderate hypertriglyceridemia patients may clarify potential underlying similarities and differences with severe hypertriglyceridemia.

In our clinic, we routinely perform next-generation sequencing (NGS) on all consenting patients and obtain a complete profile of both rare variants of large effect and common variants of small effect underlying dyslipidemias, including mild-to-moderate and severe hypertriglyceridemia. Here, we report the use of our well-established sequencing panel to assess the genetic profiles of 134 mild-to-moderate hypertriglyceridemia patients and compare these with reported findings from patients with severe hypertriglyceridemia.

## 6.3 Materials and Methods

### 6.3.1 Study subjects

Patients of interest included those of European ancestry with earliest reported triglyceride levels  $\geq 3.3$  mmol/L and  $< 10$  mmol/L, and a total cholesterol of  $< 5$  mmol/L. Patients with triglyceride levels ever reported as  $\geq 10$  mmol/L were excluded from consideration.

In adherence to the Declaration of Helsinki, all patients provided written, informed consent for collection of personal data and DNA with approval from the Western University (London ON, Canada) ethics review board (no. 07290E).

As a reference cohort of normolipidemic controls, we used the European subset of the 1000 Genomes Project (N=503) (Genomes Project et al., 2015). For additional comparison, we also utilized our cohort of 563 severe hypertriglyceridemia patients, in which all patients had triglyceride levels  $\geq 10$  mmol/L (Dron et al., 2019); this cohort is described in [Chapter 4](#).

### 6.3.2 DNA preparation and targeted sequencing

DNA isolation and preparation for targeted NGS follows the same methodology as described in [Chapter 2, Section 2.3.2](#).

### 6.3.3 Bioinformatic processing of sequencing data

The bioinformatic processing of sequencing data follows the same methodology as described in [Chapter 5, Section 5.3.3](#).

### 6.3.4 Annotation and analysis of rare single-nucleotide variants

The annotation and analysis of rare single-nucleotide variants (SNVs) follows the same methodology as described in [Chapter 4, Section 4.3.4](#); however, an updated version of VarSeq® (version 2.1.1; Golden Helix, Inc., Bozeman MT, USA) was used.

### 6.3.5 Detection of rare copy-number variants

Copy-number variants (CNVs) were identified following the same methodology as described in [Chapter 4, Section 4.3.5](#); however, an updated version of VarSeq® (version 2.1.1; Golden Helix, Inc., Bozeman MT, USA) was used.

### 6.3.6 Polygenic risk score for elevated triglyceride levels

The polygenic risk score used to assess for single-nucleotide polymorphisms (SNPs) associated with triglycerides is described in [Chapter 4, Section 4.3.6](#). This score was calculated in all patients and controls assessed in this study.

### 6.3.7 Statistical analysis

Statistical analyses were performed as described in [Chapter 4, Section 4.3.7](#); however, an updated version of GraphPad Prism for Windows (version 8.0.2; GraphPad Software, La Jolla CA, USA) was used.

## 6.4 Results

### 6.4.1 Characteristics of study subjects

One hundred and thirty-four patients were selected for study from the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada). The demographic and clinical characteristics of the 134 mild-to-moderate hypertriglyceridemia patients are summarized in **Table 6.1**.

### 6.4.2 Rare variants identified in canonical triglyceride metabolism genes

Mild-to-moderate hypertriglyceridemia patients were assessed for rare variants—both SNVs and CNVs—in the five canonical triglyceride metabolism genes. Overall, 9.0% (12/134) of patients were carriers for heterozygous rare SNVs, representing 11 unique variants (**Appendix J**). Neither bi-allelic SNVs nor CNVs were identified. Of interest, one patient carried two rare heterozygous SNVs: one in *LPL* and one in *APOA5* (i.e. “double heterozygosity”).

### 6.4.3 Measuring accumulation of common triglyceride-raising alleles

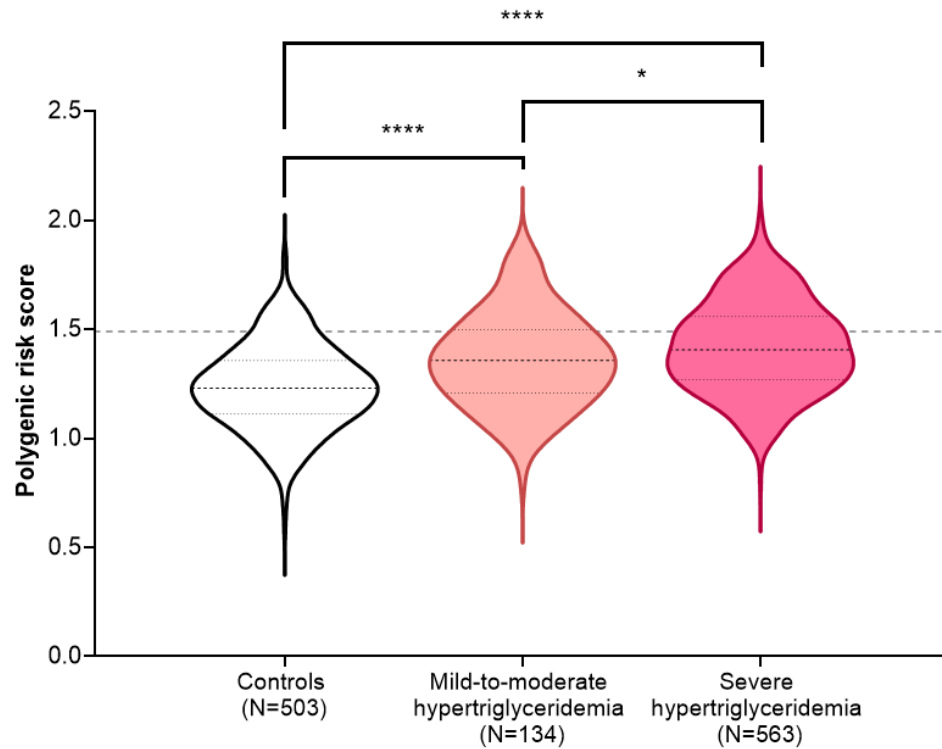
An extreme accumulation of triglyceride-raising alleles at common SNP sites was defined as a polygenic risk score  $\geq 1.49$  ( $>90^{\text{th}}$  percentile) (Dron et al., 2019). An extreme score was identified in 26.9% (36/134) of mild-to-moderate hypertriglyceridemia patients. The distribution of polygenic risk scores for each cohort are visualized in **Figure 6.1**.



**Table 6.1 Clinical and demographic information of the mild-to-moderate hypertriglyceridemia patient cohort (N=134).**

	Mild-to-moderate			Severe	P-value
	Males	Females	Total	Total	
<b>N</b>	93	41	134	563	
<b>Age</b>	53.2 ± 11.8*	54.9 ± 12.6*	53.68 ± 12.0	49.12 ± 12.5*	0.0006
<b>BMI</b> (kg/m <sup>2</sup> )	31.18 ± 2.8*	30.87 ± 5.9*	30.07 ± 4.0	29.86 ± 5.0*	0.2201
<b>Total cholesterol</b> (mmol/L)	4.06 ± 0.6	4.14 ± 0.6	4.08 ± 0.6	11.3 ± 6.1*	<0.0001
<b>Triglyceride</b> (mmol/L)	4.65 ± 1.3	4.73 ± 1.4	4.68 ± 1.3	26.03 ± 23.7	<0.0001
<b>HDL cholesterol</b> (mmol/L)	0.79 ± 0.2	0.86 ± 0.3	0.81 ± 0.2	0.81 ± 0.4*	0.0231

Values are indicative of the mean ± SD. “\*” indicates means that were calculated with an incomplete dataset. All reported P-values are two-tailed and show comparisons between the total mild-to-moderate hypertriglyceridemia and total severe hypertriglyceridemia cohort. Abbreviations: BMI = body-mass index; HDL = high-density lipoprotein.



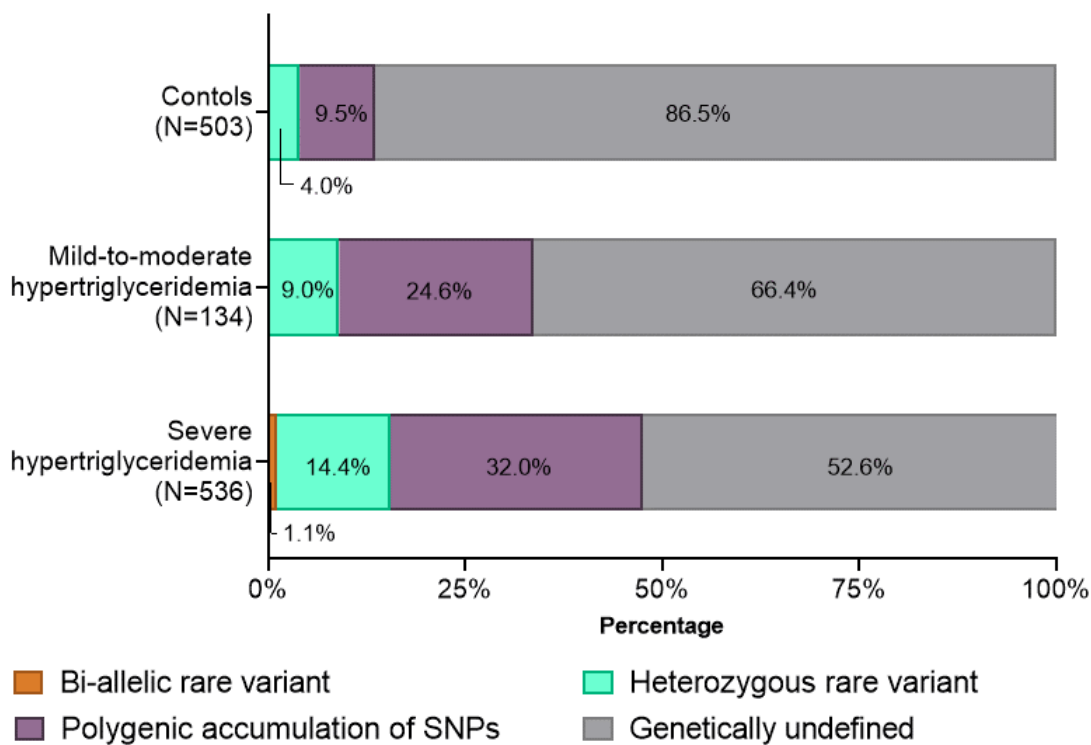
**Figure 6.1 Polygenic risk score distributions between cohorts.**

Each violin plot illustrates the distribution of the 16-SNP polygenic risk score calculated in the 1000 Genomes Project, mild-to-moderate hypertriglyceridemia, and severe hypertriglyceridemia cohorts. The 90<sup>th</sup> percentile is denoted by the hashed black line (1.49). The lines within each plot represent the median and quartiles for each cohort. P-values were generated from a Kruskal-Wallis test and adjusted with Dunn's multiple comparisons based on mean polygenic risk score values between groups. P-values: \* <0.05; \*\*\*\* <0.0001.

#### 6.4.4 Comparison of genetic profiles between cohorts

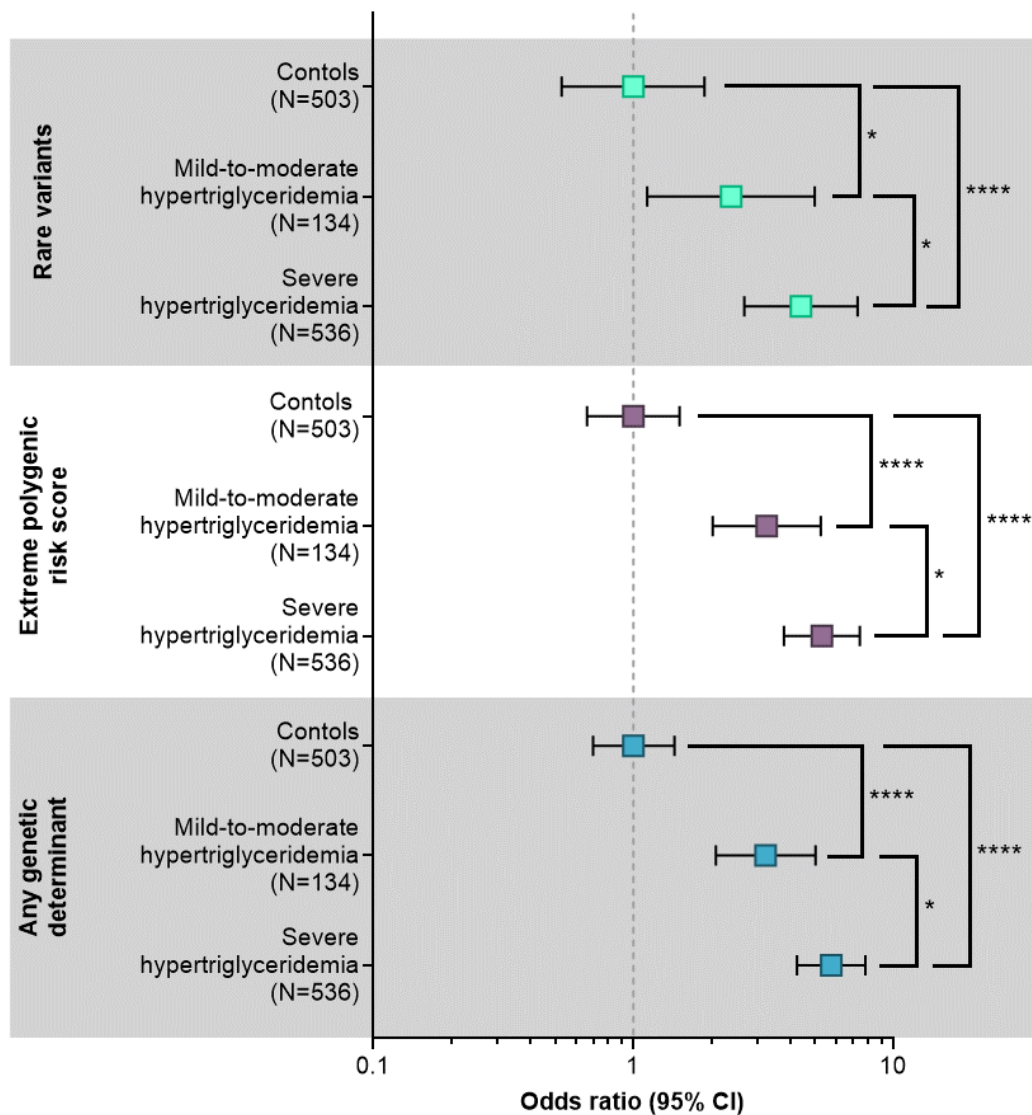
The genetic profile of the mild-to-moderate hypertriglyceridemia cohort was compared to individuals from the previously published 1000 Genomes Project and severe hypertriglyceridemia cohorts (Dron et al., 2019) (**Figure 6.2**). When considering rare variants only, 4.0% (20/503), 9.0% (12/134), and 15.5% (87/563) of individuals in the 1000 Genomes Project, mild-to-moderate hypertriglyceridemia, and severe hypertriglyceridemia cohorts were carriers, respectively. When considering the accumulation of triglyceride-raising alleles, 10.1% (51/503), 26.9% (36/134), and 37.5% (211/563) of individuals in the 1000 Genomes Project, mild-to-moderate hypertriglyceridemia, and severe hypertriglyceridemia cohorts had extreme polygenic risk scores, respectively. When considering both rare variants or accumulated triglyceride-raising alleles (i.e. an extreme polygenic risk score), 13.5% (68/503), 33.6% (45/134), and 47.4% (267/563) of individuals in the 1000 Genomes Project, mild-to-moderate hypertriglyceridemia, and severe hypertriglyceridemia cohorts had either genetic determinant, respectively.

The forest plots in **Figure 6.3** highlight differences in prevalence of genetic determinants between mild-to-moderate and severe hypertriglyceridemia patient cohorts, compared to the 1000 Genomes Project. Mild-to-moderate hypertriglyceridemia patients are 2.38-times (95% CI [1.13-4.99];  $P=0.021$ ) more likely to carry a rare variant and 3.26-times (95% CI [2.02-5.26];  $P<0.0001$ ) more likely to have an extreme polygenic risk score compared to the 1000 Genomes Project. Severe hypertriglyceridemia patients are 1.86-times (95% CI [0.98-3.51];  $P=0.032$ ) more likely to carry a rare variant and are 1.63-times (95% CI [1.07-2.48];  $P=0.013$ ) more likely to have an extreme polygenic risk score compared to mild-to-moderate hypertriglyceridemia patients. These comparisons show a significant stepwise increase in the prevalence of genetic determinants—both rare and common—from control to mild-to-moderate to severe hypertriglyceridemia patients. Finally, in our mild-to-moderate hypertriglyceridemia cohort, only 3/12 patients (25%) with a rare variant also had an extreme polygenic risk score, while in our severe hypertriglyceridemia cohort, 30/87 patients (34.5%) with a rare variant also had an extreme polygenic risk score.



**Figure 6.2 The comparison of genetic profiles between cohorts.**

Percentages were determined from normolipidemic controls and patients with either mild-to-moderate or severe hypertriglyceridemia. The prevalence of genetic determinants in the 1000 Genomes Project cohort and severe hypertriglyceridemia cohort have been previously reported by Dron et al. (2019). Abbreviations: SNPs = single-nucleotide polymorphisms.



**Figure 6.3 Differences in genetic determinants of hypertriglyceridemia.**

Each forest plot illustrates the odds ratio of mild-to-moderate or severe hypertriglyceridemia patients having rare variants (including SNVs and CNVs), the extreme accumulation of common triglyceride-raising alleles (as indicated by an extreme polygenic risk score), or either type of genetic determinant, compared to the normolipidemic controls of the 1000 Genomes Project. The dashed line indicates an odds ratio of 1.0. P-values were generated from one-tailed Fisher's exact tests. P-values: \* <0.05; \*\*\*\* <0.0001. Abbreviations: CI = confidence interval.

## 6.5 Discussion

In this study, we performed a comprehensive assessment of the genetic determinants underlying mild-to-moderate hypertriglyceridemia. Using the same rare variant and polygenic risk score analysis as we reported previously in our severe hypertriglyceridemia cohort (Dron et al., 2019), we compared genetic profiles of patients who had varying hypertriglyceridemia severity. Across 134 mild-to-moderate hypertriglyceridemia patients, 9.0% carried heterozygous rare variants and 24.6% had an excess accumulation of common SNPs, totalling 33.6% of the study sample with an identifiable genetic determinant, while 66.4% were genetically undefined.

Next, after comparing the genetic profiles of mild-to-moderate and severe hypertriglyceridemia patients (Dron et al., 2019), we noted a stepwise increase in the prevalence of genetic determinants as the hypertriglyceridemia phenotype became more severe. These differences were significant between pairwise comparisons. Cumulatively, our data show that mild-to-moderate hypertriglyceridemia patients have a genetic burden that is intermediate between normolipidemic controls and patients with severe hypertriglyceridemia.

Despite the significantly increased overall prevalence of rare variants in severe hypertriglyceridemia patients compared to mild-to-moderate hypertriglyceridemia patients (Dron et al., 2019), certain rare variants were shared between individuals in the two patient groups, underscoring that a single rare variant is insufficient to distinguish between the hypertriglyceridemia sub-phenotypes (Babirak et al., 1989; Hegele et al., 1991; Nordestgaard et al., 1997). These determinants might affect individuals differently or could have variable penetrance, making it challenging to isolate a single genetic factor responsible for causing hypertriglyceridemia, except for bi-allelic rare variants in monogenic FCS (Brahm and Hegele, 2015). Such underlying similarities are consistent with the complex nature of the disease. Not only might there be genetic determinants beyond what we assessed in this study, there are also environmental influences such as smoking, activity level, and diet that increase disease susceptibility and modulate triglyceride levels (Brahm and Hegele, 2015; Cole et al., 2015; Dron and Hegele, 2018;

Hegele et al., 2014; Johansen et al., 2010; Johansen et al., 2011b; Johansen et al., 2012; Kathiresan et al., 2009; Surendran et al., 2012; Teslovich et al., 2010; Wang et al., 2008; Willer et al., 2013). These additional factors may contribute to phenotype severity, perhaps determining whether a patient develops mild-to-moderate versus severe hypertriglyceridemia.

Genetic factors increase risk of developing hypertriglyceridemia, but are not absolute indicators for causality, except for bi-allelic rare variants that cause monogenic FCS. Additionally, phenotypic severity can depend on a myriad of exposures that may change over time, possibly blurring the prediction at any time point of whether an individual will express the mild-to-moderate or severe form of the disease. Among patients with severe hypertriglyceridemia, those with a rare variant or extreme SNP accumulation had significantly more metabolic risk factors—including higher body-mass index, blood pressure, and fasting glucose—compared to patients with FCS (Paquette et al., 2019). Individuals may develop this phenotype due to a greater overall burden of both genetic and environmental risk factors that contribute towards disease presentation. In contrast, individuals with mild-to-moderate hypertriglyceridemia have an intermediate burden of genetic predisposing factors. But in the presence of secondary non-genetic factors, the milder genetic susceptibility in mild-to-moderate hypertriglyceridemia can be overcome, and patients could slip metabolically into a severe hypertriglyceridemia phenotype.

To predict future risk of developing hypertriglyceridemia, either mild-to-moderate or severe, assessing genetic and non-genetic determinants simultaneously would seem logical. Genetic analysis could be broadened to include a genome-wide polygenic score that concurrently assesses millions of SNPs contributing towards hypertriglyceridemia susceptibility. Several such large-scale, genome-wide polygenic scores have been created for other complex diseases, including coronary artery disease, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer (Khera et al., 2018). In large populations, Khera et al. found that individuals with a high genome-wide polygenic score were at an equivalent risk for disease as individuals carrying a single pathogenic variant related to the disease (Khera et al., 2018). In addition, rare variants disrupting non-canonical triglyceride genes could be examined for a better-defined genetic profile, as

other genes beyond the triglyceride metabolic pathway have reported relationships with triglyceride levels, such as *CREB3L3*, *MLXIPL* and *GCKR* (Kathiresan et al., 2009; Santoro et al., 2012; Willer et al., 2008).

Finally, environmental factors must be considered in conjunction with the genetic profile when diagnosing and defining hypertriglyceridemia. Future research could study differences in environmental factors between individuals with similar genetic profiles yet differing severities of hypertriglyceridemia. We did not systematically record baseline environmental factors, which is a limitation as we would anticipate a lesser burden of non-genetic stressors among patients with mild-to-moderate hypertriglyceridemia versus severe hypertriglyceridemia. Larger and more systematically defined hypertriglyceridemia cohorts would be ideal going forward, for instance, when developing hypertriglyceridemia patient registries.

## 6.6 Conclusion

With more extreme hypertriglyceridemia severity, we see an increased prevalence of genetic determinants, including both variably penetrant heterozygous rare variants disrupting a triglyceride metabolism gene—including *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*—and an extreme accumulation of 16 common triglyceride-associated SNPs. Genetic testing alone cannot be used to accurately predict hypertriglyceridemia severity for any single patient at any particular time point. At present, we have no evidence that clinical outcomes or interventions differ according to the genotype, although its potential use in prognostication and predicting response to treatment should be evaluated. Additional research is required to consider environmental risk factors in conjunction with our established method of ascertaining genetic profiles related to hypertriglyceridemia. At present, we recommend that most clinical decisions—diet, statin, fibrate, new biologics (Laufs et al., 2020)—can be based on the biochemical severity of the lipid disturbance, without the need for extensive genetic testing.



## 6.7 References

- Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7, Unit 7* 20.
- Babirak, S.P., Iverius, P.H., Fujimoto, W.Y., and Brunzell, J.D. (1989). Detection and characterization of the heterozygote state for lipoprotein lipase deficiency. *Arteriosclerosis* 9, 326-334.
- Brahm, A.J., and Hegele, R.A. (2015). Chylomicronaemia--current diagnosis and future therapies. *Nat Rev Endocrinol* 11, 352-362.
- Chait, A., and Brunzell, J.D. (1992). Chylomicronemia syndrome. *Adv Intern Med* 37, 249-273.
- Chait, A., and Eckel, R.H. (2019). The Chylomicronemia Syndrome Is Most Often Multifactorial: A Narrative Review of Causes and Treatment. *Ann Intern Med* 170, 626-634.
- Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.
- Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.
- Dron, J.S., and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.
- Dron, J.S., Wang, J., Berberich, A.J., Iacocca, M.A., Cao, H., Yang, P., Knoll, J., Tremblay, K., Brisson, D., Netzer, C., *et al.* (2018). Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia. *J Lipid Res* 59, 1529-1535.
- Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., *et al.* (2019). Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol* 13, 80-88.
- Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhainds, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.
- Hegele, R.A., Breckenridge, W.C., Cox, D.W., Maguire, G.F., Little, J.A., and Connelly, P.W. (1991). Interaction between variant apolipoproteins C-II and E that affects plasma lipoprotein concentrations. *Arterioscler Thromb Vasc Biol* 11, 1303-1309.
- Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Averna, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The

polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* 2, 655-666.

Holmes, M.V., Asselbergs, F.W., Palmer, T.M., Drenos, F., Lanktree, M.B., Nelson, C.P., Dale, C.E., Padmanabhan, S., Finan, C., Swerdlow, D.I., *et al.* (2015). Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* 36, 539-550.

Iacocca, M.A., Wang, J., Dron, J.S., Robinson, J.F., McIntyre, A.D., Cao, H., and Hegele, R.A. (2017). Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia. *J Lipid Res* 58, 2202-2209.

Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* 55, 765-772.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011a). Genetic determinants of plasma triglycerides. *J Lipid Res* 52, 189-206.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., *et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 42, 684-687.

Johansen, C.T., Wang, J., Lanktree, M.B., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E., Schwartz, S.M., *et al.* (2011b). An increased burden of common and rare lipid-associated risk alleles contributes to the phenotypic spectrum of hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 31, 1916-1926.

Johansen, C.T., Wang, J., McIntyre, A.D., Martins, R.A., Ban, M.R., Lanktree, M.B., Huff, M.W., Peterfy, M., Mehrabian, M., Lusic, A.J., *et al.* (2012). Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 5, 66-72.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41, 56-65.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., *et al.* (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 50, 1219-1224.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46, 310-315.

- Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4, 1073-1081.
- Laufs, U., Parhofer, K.G., Ginsberg, H.N., and Hegele, R.A. (2020). Clinical review on triglycerides. *Eur Heart J* 41, 99-109c.
- Lewis, G.F., Xiao, C., and Hegele, R.A. (2015). Hypertriglyceridemia in the genomic era: a new paradigm. *Endocr Rev* 36, 131-147.
- Nordestgaard, B.G., Abildgaard, S., Wittrup, H.H., Steffensen, R., Jensen, G., and Tybjaerg-Hansen, A. (1997). Heterozygous lipoprotein lipase deficiency: frequency in the general population, effect on plasma lipid levels, and risk of ischemic heart disease. *Circulation* 96, 1737-1744.
- Paquette, M., Bernard, S., Hegele, R.A., and Baass, A. (2019). Chylomicronemia: Differences between familial chylomicronemia syndrome and multifactorial chylomicronemia. *Atherosclerosis* 283, 137-142.
- Santoro, N., Zhang, C.K., Zhao, H., Pakstis, A.J., Kim, G., Kursawe, R., Dykas, D.J., Bale, A.E., Giannini, C., Pierpont, B., *et al.* (2012). Variant in the glucokinase regulatory protein (GCKR) gene is associated with fatty liver in obese children and adolescents. *Hepatology* 55, 781-789.
- Surendran, R.P., Visser, M.E., Heemelaar, S., Wang, J., Peter, J., Defesche, J.C., Kuivenhoven, J.A., Hosseini, M., Peterfy, M., Kastelein, J.J., *et al.* (2012). Mutations in LPL, APOC2, APOA5, GPIHBP1 and LMF1 in patients with severe hypertriglyceridaemia. *J Intern Med* 272, 185-196.
- Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.
- Varbo, A., and Nordestgaard, B.G. (2016). Remnant Cholesterol and Triglyceride-Rich Lipoproteins in Atherosclerosis Progression and Cardiovascular Disease. *Arterioscler Thromb Vasc Biol* 36, 2133-2135.
- Wang, J., Ban, M.R., Zou, G.Y., Cao, H., Lin, T., Kennedy, B.A., Anand, S., Yusuf, S., Huff, M.W., Pollex, R.L., *et al.* (2008). Polygenic determinants of severe hypertriglyceridemia. *Hum Mol Genet* 17, 2894-2899.
- Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M., *et al.* (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40, 161-169.

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.

## Chapter 7 – Loss-of-function *CREB3L3* variants in patients with severe hypertriglyceridemia

The work contained in this Chapter has been edited from its original publication in *Arteriosclerosis, Thrombosis, and Vascular Biology* for brevity and consistency throughout this Dissertation.



**Dron, J.S.,** Dillott, A.A., Lawson, A., McIntyre, A.D., Davis, B.D., Wang, J., Cao, H., Movsesyan, I., Malloy, M.J., Pullinger, C.R., *et al.* (2020). Loss-of-Function *CREB3L3* Variants in Patients With Severe Hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 40, 1935–1941.

## 7.1 Abstract

**Objective:** Significant genetic determinants of severe hypertriglyceridemia include both common variants with small effects (assessed using polygenic risk scores) plus heterozygous and homozygous rare variants in canonical genes directly affecting triglyceride metabolism. Here we broadened our scope to detect statistical associations of rare loss-of-function variants in genes affecting non-canonical pathways, including those known to affect triglyceride metabolism indirectly.

**Methods and Results:** From targeted next-generation sequencing of 69 metabolism-related genes in 265 patients of European descent with severe hypertriglyceridemia ( $\geq 10$  mmol/L) and 477 normolipidemic controls, we focused on the association of rare heterozygous loss-of-function variants in individual genes. We observed that compared to controls, severe hypertriglyceridemia patients were 20.2-times (95% CI [1.11-366.1];  $P=0.03$ ) more likely than controls to carry a rare loss-of-function variant in *CREB3L3*, which encodes a transcription factor that regulates several target genes with roles in triglyceride metabolism.

**Conclusions:** Our findings indicate that rare variants in a non-canonical gene for triglyceride metabolism, namely *CREB3L3*, contribute significantly to severe hypertriglyceridemia. Secondary genes and pathways should be considered when evaluating the genetic architecture of this complex trait.

## 7.2 Introduction

Severe hypertriglyceridemia is defined as triglyceride levels  $\geq 10$  mmol/L in the fasted state (Hegele et al., 2014). With a population prevalence of about 1 in 600, individuals with severely elevated triglyceride levels are at risk for several health complications, the most serious being acute pancreatitis (Brahm and Hegele, 2015; Dron and Hegele, 2017). As a multifactorial disease, severe hypertriglyceridemia can be caused by various genetic determinants, environmental factors, or some combination of both, which strongly reflects the phenotype's complexity (Brahm and Hegele, 2015; Hegele et al., 2014).

When studying genetic influences on triglyceride levels, canonical genes involved in triglyceride metabolism—*LPL*, *LMF1*, *GPIHBP1*, *APOA5*, and *APOC2*—are screened for rare variants that disrupt the catabolism of triglyceride-rich lipoproteins and lead to marked elevations in plasma triglyceride concentration (Johansen and Hegele, 2012; Johansen et al., 2011). Additionally, we consider the accumulation of common triglyceride-associated single-nucleotide polymorphisms (SNPs) using polygenic scores; although individual SNPs have small phenotypic impacts, in aggregate they impart a larger cumulative increase in triglyceride levels (Dron et al., 2019). Assessing these different types of genetic determinants simultaneously is necessary to understand the broader genetic basis of hypertriglyceridemia.

We recently studied rare variants and common SNPs in a cohort of 563 severe hypertriglyceridemia patients. We found that 1.1% of patients carried bi-allelic rare variants in canonical triglyceride metabolism genes (Dron et al., 2019). By definition, these patients were diagnosed with familial chylomicronemia syndrome (FCS), the monogenic form of severe hypertriglyceridemia. As an autosomal recessive disorder, FCS is the only instance in which a patient's hypertriglyceridemia phenotype is driven exclusively by a single genetic factor. All other hypertriglyceridemia cases are non-monogenic and follow a multifactorial model; this distinction of severe hypertriglyceridemia is defined as “multifactorial chylomicronemia”. From the same severe hypertriglyceridemia cohort, 46.4% of patients were found to carry genetic determinants contributing towards their phenotypes, including either heterozygous rare variants disrupting canonical metabolism genes or an excess accumulation of triglyceride-associated SNPs (Dron et al., 2019). The remaining 52.6% of patients did not carry any identified genetic determinant. With a multifactorial nature and strong polygenic background, many of these multifactorial patients likely have additional genetic factors contributing towards their severe hypertriglyceridemia phenotype that were not assessed in our initial study.

Uncovering new genetic determinants contributing towards disease can be challenging, particularly when focusing on rare variants in non-canonical genes; large patient cohorts are required to provide adequate power to show associations between variants and the

phenotype of interest (Auer and Lettre, 2015; Lee et al., 2014; Michailidou, 2018). In smaller cohorts, gene-based rare variant association studies (RVAS) are an alternative method that can be used to boost statistical power by grouping variants according to function(s) of the gene product (Auer and Lettre, 2015; Ionita-Laza, 2013; Lee et al., 2014). If a gene has an increased prevalence of rare variants in cases versus controls, it suggests that the gene may play some role in driving or influencing the phenotype of interest. This methodology has been successful in uncovering new gene associations in complex traits and diseases such as body-mass index, height, Alzheimer's disease, and lipid levels (Marouli et al., 2017; Nho et al., 2016; Pirim et al., 2015; Turcot et al., 2018).

Since the severe hypertriglyceridemia patients were previously sequenced using our targeted LipidSeq panel design, in addition to the five aforementioned canonical genes for FCS, sequencing data were also generated for 64 other genes associated with lipid traits and metabolic disorders. To better define the full spectrum of genetic determinants underlying severe hypertriglyceridemia, we evaluated the multifactorial and undefined hypertriglyceridemia patients using gene-based RVAS to identify rare loss-of-function variants in secondary or non-canonical genes. We thus explored a diverse range of genetic determinants across non-canonical triglyceride genes to further tease out the genetic underpinnings of severe hypertriglyceridemia and define in greater detail the multifactorial and polygenic nature of this phenotype.

## 7.3 Materials and Methods

### 7.3.1 Study subjects

Patients of interest included those of European ancestry with triglyceride levels  $\geq 10$  mmol/L. Importantly, patients with FCS, diagnosed by the presence of bi-allelic rare variants in canonical triglyceride metabolism genes (Dron et al., 2019) were excluded from consideration, since these patients have a clear, genetic explanation for their severe hypertriglyceridemia phenotype.

Five hundred and fifty-seven patients met our criteria and were selected for study from either the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada), or the Lipid, Diabetes, and Cardiology Clinics at the



University of California, San Francisco. In adherence to the Declaration of Helsinki, all patients provided written, informed consent for collection of personal data and DNA with approval from either the Western University (London ON, Canada) ethics review board (no. 07290E) or the Committee on Human Research of the University of California, San Francisco.

Clinical and demographic information for each patient were collected at the time of their first clinic visit. Fasting lipid profiles were measured according to clinical standards of care using the Roche Cobas C502 Analyzer (Hoffmann La Roche, Mississauga, Ontario).

As a reference control cohort, we used the European subset of the 1000 Genomes Project (N=503) (Genomes Project et al., 2015). While phenotype information is not available for these individuals, we make the assumption that their triglyceride levels follow the typical distribution seen in a population with similar ancestral background (Castelli et al., 1977; National Cholesterol Education Program Expert Panel on Detection and Treatment of High Blood Cholesterol in Adults, 2002). Further, severe hypertriglyceridemia has a population prevalence of ~1 in 600 individuals, which suggests that it is very unlikely that anyone in the 1000 Genomes Project has extremely elevated triglyceride levels. For these reasons, we refer to this cohort as “normolipidemic”.

### 7.3.2 DNA preparation and targeted sequencing

DNA isolation and preparation for targeted next-generation sequencing (NGS) follows the same methodology as described in [Chapter 2, Section 2.3.2](#).

### 7.3.3 Bioinformatic processing of sequencing data

The bioinformatic processing of sequencing data follows the same methodology as described in [Chapter 5, Section 5.3.3](#).

### 7.3.4 Principal component analysis

To account for differential ancestry and batch effects of the patients and normolipidemic controls, a principal component analysis (PCA) was performed. VCF files were merged and filtered to include only single-nucleotide variants (SNVs) appearing within the

exonic and splice regions captured by the LipidSeq panel with a minor allele frequency of  $>0.5\%$  in the Genome Aggregation Database (gnomAD; <https://gnomad.broadinstitute.org/>) (Karczewski et al., 2020). The merged VCF was then processed with Exautomate (<https://github.com/exautomate/Exautomate-Core>) (Davis et al., 2019) to produce MAP and PED files. Linkage disequilibrium pruning at a threshold of 0.5 and PCA were executed within SNP & Variation Suite v8.8.3 (SVS; Golden Helix Inc., Bozeman MT, USA). Significant principal components were identified using logistic regression within R. Multidimensional outlier detection (multiplier = 1.5) was performed using significant components within SVS v8.8.3.

### 7.3.5 Annotation and analysis of loss-of-function variants

Variant annotation using VarSeq® (Golden Helix, Inc., Bozeman MT, USA) was described in [Chapter 6, Section 6.3.4](#).

Loss-of-function variants in patients and normolipidemic controls were identified using the following criteria: 1) minor allele frequency of  $<1\%$  or missing in gnomAD; 2) sequence ontology of nonsense, frameshift, splice donor, splice acceptor, or copy-number variant (CNV) deletion; 3) CADD PHRED-scaled score of  $\geq 10$ ; and 4) an American College of Medical Genetics (ACMG) classification of pathogenic, likely pathogenic, or uncertain significance (Richards et al., 2015).

### 7.3.6 Gene-based rare variant association study

The optimal unified sequence kernel association test (SKAT-O) (Lee et al., 2012a; Lee et al., 2012b)—a combination of burden and variance-component tests—was used to perform a gene-based RVAS between our patients and normolipidemic controls. To enrich for variants that likely have the largest phenotypic impact, only loss-of-function variants were considered in this analysis. We performed SKAT-O using a linear weighted kernel and the optimal adjustment method through the use of Exautomate (<https://github.com/exautomate/Exautomate-Core>) (Davis et al., 2019) (**Appendix L**). P-values generated by SKAT-O were adjusted with the Bonferroni correction for multiple comparisons. Significant results were considered as  $P < 0.05$ .

### 7.3.7 Odds ratio assessment

Since only SNVs can be considered using SKAT-O, we followed up by generating 2-by-2 contingency tables for each gene by counting carriers versus non-carriers of loss-of-function variants, including SNVs, frameshifts, and CNVs. After determining the odds ratios (ORs) and 95% confidence intervals (CIs) for each gene, P-values were generated using Fisher's exact tests and adjusted with the Bonferroni correction for multiple comparisons. One-tailed P-values were generated if the gene was considered to be a canonical metabolism gene, and two-tailed P-values were generated for the remaining genes. To calculate ORs with cell counts of zero, the Haldane-Anscombe correction method was applied by adding 0.5 to each cell of the contingency table.

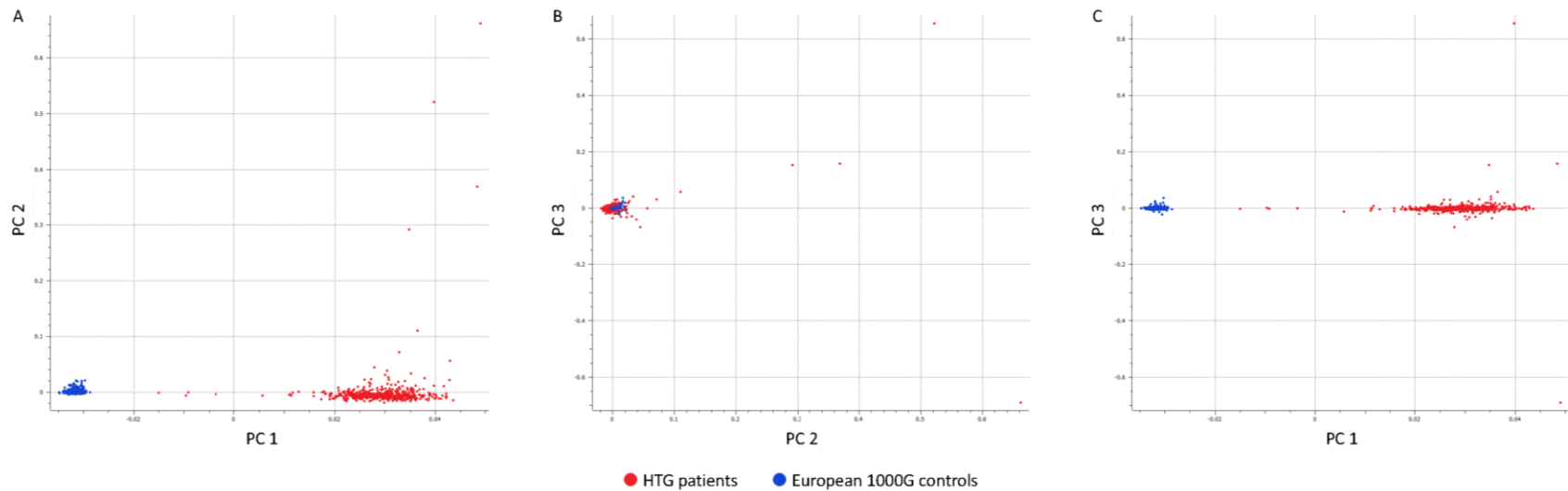
Statistical analyses were performed using GraphPad Prism v8.0.1 (GraphPad Software, La Jolla, CA, USA).

## 7.4 Results

### 7.4.1 Characteristics of study subjects

We performed a PCA and multidimensional outlier detection to remove samples that may have been affected by batch effects or were population outliers (**Figure 7.1**). Following outlier removal, the final dataset consisted of 265 patients and 477 normolipidemic controls.

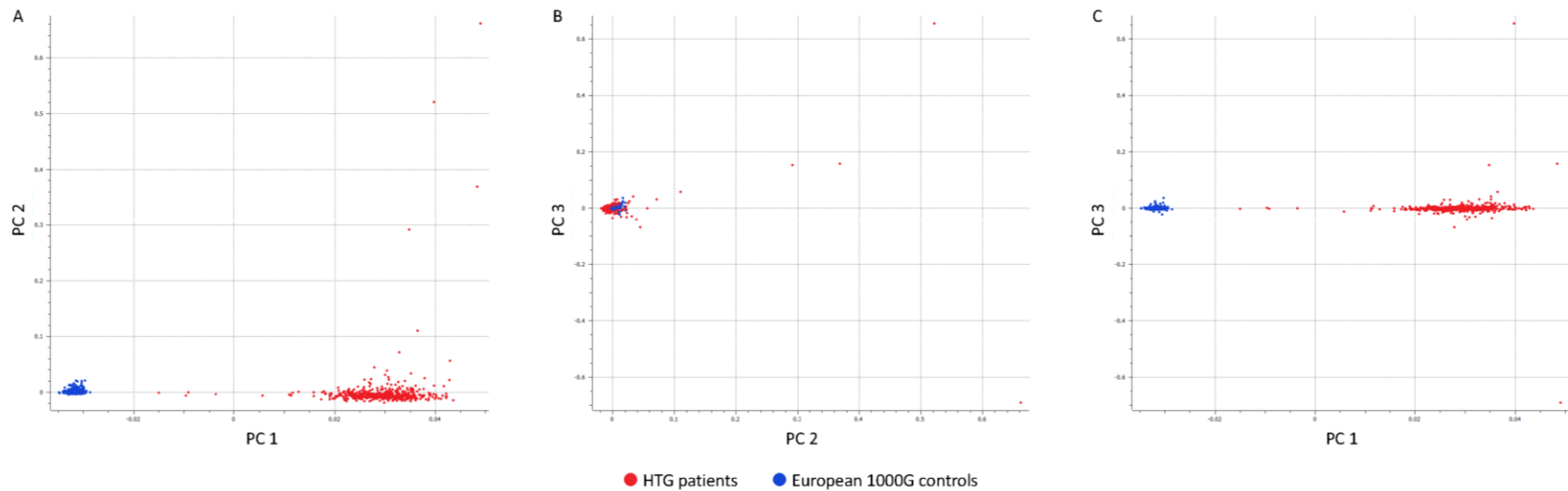
The demographic and clinical characteristics of the 265 severe hypertriglyceridemia patients are summarized in **Figure 7.1**



**Figure 7.1 Principal component analysis.**

Principal component (PC) analysis of patients with severe hypertriglyceridemia (N=557) and the normolipidemic controls from the European subset of the 1000 Genomes Project (N=503). **A)** PC1 and PC2 of the HTG patients and normolipidemic controls. **B)** PC2 and PC3 and the severe hypertriglyceridemia patients and normolipidemic controls. **C)** PC1 and PC3 of the severe hypertriglyceridemia patients and normolipidemic controls. Abbreviations: 1000G = 1000 Genomes Project; HTG = hypertriglyceridemia; PC = principal component.

Table 7.1. Across the total patient cohort, the mean  $\pm$  standard deviation (SD) triglyceride concentration and age were  $26.7 \pm 25.2$  mmol/L and  $50.9 \pm 12.3$  years, respectively. Most patients were male (64.2%) and of the 216 patients with data available, 44.4% had diabetes.



**Figure 7.1 Principal component analysis.**

Principal component (PC) analysis of patients with severe hypertriglyceridemia (N=557) and the normolipidemic controls from the European subset of the 1000 Genomes Project (N=503). **A)** PC1 and PC2 of the HTG patients and normolipidemic controls. **B)** PC2 and PC3 and the severe hypertriglyceridemia patients and normolipidemic controls. **C)** PC1 and PC3 of the severe hypertriglyceridemia patients and normolipidemic controls. Abbreviations: 1000G = 1000 Genomes Project; HTG = hypertriglyceridemia; PC = principal component.

**Table 7.1 Clinical and demographic information of patients with severe hypertriglyceridemia (N=265).**

	<b>Males</b>	<b>Females</b>
<b>N</b>	170	95
<b>Age</b>	50.0 ± 10.9	52.6 ± 14.3
<b>BMI (kg/m<sup>2</sup>)</b>	29.8 ± 4.4*	29.4 ± 5.2*
<b>Total cholesterol (mmol/L)</b>	11.3 ± 6.4	12.0 ± 6.8*
<b>Triglyceride (mmol/L)</b>	25.0 ± 22.6	29.8 ± 29.2
<b>HDL cholesterol (mmol/L)</b>	0.77 ± 0.4*	0.82 ± 0.4*
<b>Diabetes</b>	41.0%*	50.0%*

Values are indicative of the mean ± SD. “\*” indicates an incomplete dataset. Abbreviations: BMI = body-mass index; HDL = high-density lipoprotein.

#### 7.4.2 Loss-of-function variants identified from the LipidSeq gene panel

In 37 carriers, we identified 26 unique loss-of-function variants across 15 different genes between our patients and normolipidemic controls (**Appendix K**). Of these unique variants, 10 (38.5%) were frameshifts, 9 (34.6%) were nonsense, two (7.7%) were splice donors, and five (19.2%) were CNV deletions.

*GCKR* had the highest number of unique variants (4; 15.4%) and the highest number of carriers (10; 27.0%).

#### 7.4.3 Gene-based rare variant association analysis using SKAT-O

From our SKAT-O analysis, only SNVs from 14 genes were considered. With such a small working dataset, there were no genes that had a significantly different number of variants between our patients and normolipidemic controls (**Table 7.2**). *CREB3L3*, *APOA5*, *LIPC*, and *PLIN1* were the only genes that had a non-one P-value, although they were not significant.

#### 7.4.4 Gene-based odds ratio assessment

In order to consider frameshift variants and CNVs along with SNVs, we performed OR assessments on 15 genes (**Table 7.3**). *CREB3L3* had a significant increase in the prevalence of patients carrying loss-of-function variants compared to normolipidemic controls (**Figure 7.2**). Our severe hypertriglyceridemia patients were 20.2-times (95% CI [1.11-366.1]; two-tailed P=0.03) more likely to carry a rare loss-of-function variant in *CREB3L3* compared to normolipidemic controls.



**Table 7.2 Output from SKAT-O analysis between severe hypertriglyceridemia (N=265) and normolipidemic controls (N=477).**

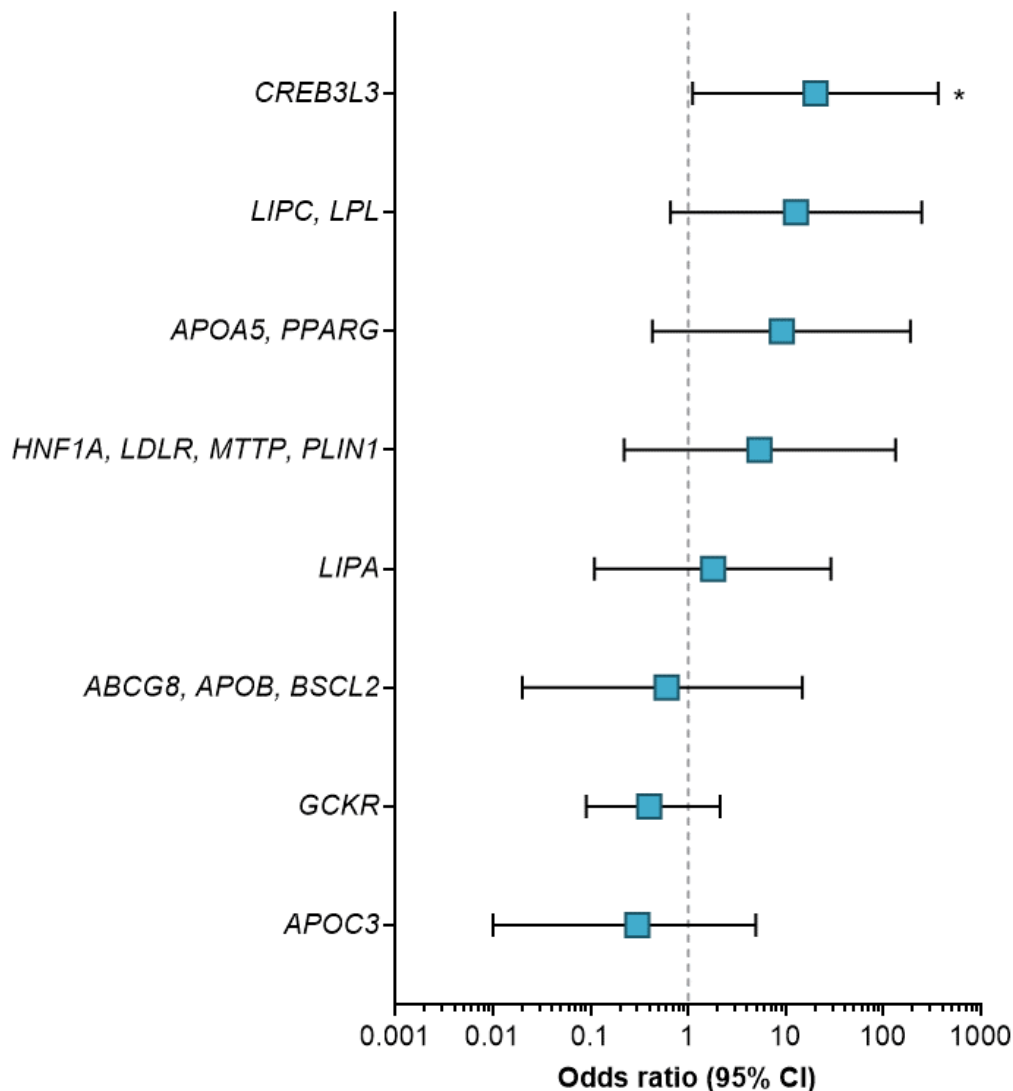
<b>Gene</b>	<b>Adjusted P-value</b>	<b>Number of variants considered in analysis</b>
<i>CREB3L3</i>	0.090735	4
<i>APOA5</i> *	0.090735	3
<i>LIPC</i>	0.090735	1
<i>PLIN1</i>	0.090735	1
<i>GCKR</i>	1	2
<i>KLF11</i>	1	1
<i>ABCG5</i>	1	1
<i>ABCG8</i>	1	1
<i>NPC1L1</i>	1	1
<i>BLK</i>	1	1
<i>LPL</i> *	1	1
<i>WRN</i>	1	1
<i>PYGM</i>	1	1
<i>LMF1</i> *	1	1

P-values were adjusted using the Bonferroni correction method. “\*” indicates canonical genes involved in the triglyceride metabolic pathway.

**Table 7.3** The odds of severe hypertriglyceridemia patients (N=265) carrying a loss-of-function variant in a particular gene compared to normolipidemic controls (N=477).

Gene	OR (95% CI)	Adjusted P-value	Cases with variants	Controls with variants
<i>CREB3L3</i>	20.2 (1.11-366.1)	0.03	5	0
<i>LPL</i> *	12.7 (0.66-247.5)	0.24	3	0
<i>LIPC</i>	12.7 (0.66-247.5)	0.24	3	0
<i>APOA5</i> *	9.1 (0.43-189.4)	0.675	2	0
<i>PPARG</i>	9.1 (0.43-189.4)	0.675	2	0
<i>HNFA1A</i>	5.4 (0.22-133.4)	1	1	0
<i>MTTP</i>	5.4 (0.22-133.4)	1	1	0
<i>PLIN1</i>	5.4 (0.22-133.4)	1	1	0
<i>LDLR</i>	5.4 (0.22-133.4)	1	1	0
<i>LIPA</i>	1.8 (0.11-28.94)	1	1	1
<i>ABCG8</i>	0.6 (0.02-14.74)	1	0	1
<i>APOB</i>	0.6 (0.02-14.74)	1	0	1
<i>BSCL2</i>	0.6 (0.02-14.74)	1	0	1
<i>GCKR</i>	0.4 (0.09-2.12)	1	2	8
<i>APOC3</i> *	0.3 (0.01-4.96)	1	0	3

“\*” indicates canonical genes involved in the triglyceride metabolic pathway. P-values were generated using Fisher’s exact test and adjusted using the Bonferroni correction for multiple comparisons. Canonical genes have one-tailed P-values listed, while the remaining genes have two-tailed P-values listed. Abbreviations: CI = confidence interval; OR = odds ratio.



**Figure 7.2 Odds ratio of loss-of-function variants across LipidSeq genes.**

Each forest plot illustrates the odds ratio of severe hypertriglyceridemia patients (N=265) carrying a loss-of-function variant in one particular gene compared to normolipidemic controls from the 1000 Genomes Project (N=477). The dashed line indicates an odds ratio of 1.0. Canonical genes involved in the triglyceride metabolic pathway include *APOA5*, *APOC3*, *LMF1*, and *LPL*. P-values were generated using Fisher's exact test and adjusted using the Bonferroni correction for multiple comparisons. Canonical genes have one-tailed P-values listed, while the remaining genes have two-tailed P-values listed. P-value: \* <0.05. Abbreviations: CI = confidence interval.

## 7.5 Discussion

In this study, we considered genetic determinants beyond the canonical triglyceride metabolism genes in 265 patients of European descent with non-FCS, severe hypertriglyceridemia to determine whether rare variants in other genes contribute towards the phenotype. By analyzing this subset of patients from the initial severe hypertriglyceridemia cohort, we enriched for individuals likely carrying non-canonical, polygenic variants. Moreover, by focusing solely on loss-of-function variants in our analyses, we enriched our dataset for variants with likely larger impacts on triglyceride levels.

Our initial gene-based RVAS using SKAT-O did not reveal any significant results. This was unsurprising, since we had two relatively small cohorts; if causative genes harbored only a few variants with modest influences on triglyceride levels affecting a few patients, these individual gene signals would be difficult to detect with such small sample sizes. From our SKAT-O analysis, *CREB3L3* was one of four the genes that produced P-values not equal to one. It was interesting to note that *CREB3L3* appeared as the most significant result in subsequent analyses that aggregated variants.

SKAT-O is limited in that it cannot be used to consider multi-nucleotide variants, such as frameshifts and CNVs. Therefore, gene-specific 2-by-2 contingency tables for carriers of any type of loss-of-function variants were used to determine gene-specific ORs. With this approach, *CREB3L3* was shown to have a significant enrichment for such variants in our patients compared to normolipidemic controls. The genes *LPL* and *APOA5* appearing among the top most enriched genes, albeit not significantly, provided positive validation for our analysis, since both genes encode proteins directly involved in triglyceride metabolism (Brahm and Hegele, 2015; Hegele et al., 2014). We previously showed that 46.4% of the initial severe hypertriglyceridemia cohort carried heterozygous rare variants in the canonical genes, including *LPL* and *APOA5* (Dron et al., 2019). The present novel findings indicate that *CREB3L3*, a non-canonical gene, is associated with the severe hypertriglyceridemia phenotype with at least similar strength or magnitude as *LPL* and *APOA5*.

Severe hypertriglyceridemia patients were 20.2-times ( $P=0.03$ ) more likely to carry rare loss-of-function variants in *CREB3L3* compared to normolipidemic controls. *CREB3L3* encodes cAMP-responsive element-binding protein H (CREBH), a transcription factor primarily expressed in the liver and small intestine (Nakagawa and Shimano, 2018). CREBH has been shown to regulate apo C-II and A-IV expression, which helps activate triglyceride hydrolysis through its transfer from triglyceride-rich lipoproteins to high-density lipoprotein (HDL) particles (Goldberg et al., 1990; Nakagawa and Shimano, 2018; Weinberg and Spector, 1985; Xu et al., 2014). Previous reports have shown an excess of *CREB3L3* rare variants in hypertriglyceridemia patients (Johansen et al., 2012); a very recent study noted a number of *CREB3L3* variants in patients with multifactorial chylomicronemia (D'Erasmus et al., 2019). Furthermore, in an *in vivo* model, *Creb3l3*<sup>-/-</sup> mice had significantly higher plasma triglyceride levels compared to wild-type littermates (Lee et al., 2011), and when bred onto a full *Ldlr*<sup>-/-</sup> background, mice had increased very-low-density lipoprotein (VLDL) levels and decreased hepatic apo A-I production when fed a Western diet (Park et al., 2016). Although bi-allelic, loss-of-function variants in *CREB3L3* have not yet been found to cause FCS—these variants were absent in our clinical database and publicly available databases—the excess variants found in our patients and the mechanistic relationship with triglyceride metabolism demonstrate that *CREB3L3* is an important non-canonical triglyceride gene in the context of hypertriglyceridemia. Previous studies have even suggested that certain loss-of-function variants with variable penetrance lead to severe hypertriglyceridemia (Cefalu et al., 2015).

After *CREB3L3*, *LPL* and *APOA5* had the next highest prevalence of loss-of-function variants in severe hypertriglyceridemia patients, at 12.7-times (95% CI [0.66-247.5]; two-tailed  $P=0.24$ ) and 9.1-times (95% CI [0.43-189.4]; two-tailed  $P=0.675$ ) compared to normolipidemic controls, respectively. *LPL* encodes lipoprotein lipase (LPL), the main enzyme involved in triglyceride hydrolysis (Boullart et al., 2012; Lambert and Parks, 2012), while *APOA5* encodes apo A-V, an apolipoprotein that assists in enhancing the function of LPL (Forte et al., 2016). With rare bi-allelic variants in these genes causing FCS, heterozygous loss-of-function variants likely lead to large elevations in triglyceride levels through partial disruptions in the metabolic pathway, contributing towards the

severe hypertriglyceridemia phenotype (Brahm and Hegele, 2015; Dron et al., 2019; Hegele et al., 2014).

Although not significant, we observed that certain loss-of-function variants only occurred in normolipidemic controls and were absent in our patient cohort. Apo C-III, encoded by *APOC3*, is found on triglyceride-rich lipoproteins and inhibits LPL-mediated triglyceride hydrolysis by opposing the stimulatory action between apo C-II and LPL (Brahm and Hegele, 2015; Johansen et al., 2011; van Dijk et al., 2004). *In vivo* models have shown that mice overexpressing apo C-III have hypertriglyceridemia (Ito et al., 1990), while *APOC3*-deficient mice have hypotriglyceridemia (Maeda et al., 1994). Further, many studies have found human carriers of *APOC3* loss-of-function variants to have reduced triglyceride levels (Kohan, 2015). Our results are in line with these findings, supporting the conclusion that lost or reduced apo C-III function reduces circulating triglyceride levels and “protects” against the hypertriglyceridemia phenotype.

Despite the strengths of our study design, some limitations remain. By stringently only considering rare loss-of-function variants, their infrequency constrained the number of genetic determinants that could be considered for analysis, such as deleterious missense variants. Our study was also limited in that we were unable to discover new gene relationships since the LipidSeq panel was designed to target genes already known to be involved in dyslipidemic phenotypes and metabolic disorders. To address these limitations, a larger sample size would increase the statistical power and likelihood of identifying more variants of interest. An increased sample size would also be necessary if an even larger gene set were to be analyzed in the hopes of identifying genes with novel or unexplored relationships with the hypertriglyceridemia phenotype.

## 7.6 Conclusion

When evaluating the genetic determinants contributing towards a complex phenotype, it can be challenging to identify genes carrying variants with milder phenotypic impacts compared to genes associated with monogenic forms of disease. As such, rare variant association methods that group variants by gene can help to increase power and identify such additional genetic influencers. From our gene-based analysis, we found that the non-

canonical triglyceride gene, *CREB3L3*, is an important contributor towards the severe hypertriglyceridemia phenotype. Importantly, the associated loss-of-function variants were more prevalent in *CREB3L3* compared to both *LPL* and *APOA5*, both of which are well-established genes involved in triglyceride metabolism, and in which homozygous rare variants can cause FCS. Our findings suggest that searching beyond the canonical triglyceride metabolism genes may help better understand the genetic basis of severe hypertriglyceridemia. Future studies should widen the range of secondary factors and pathways for which genetic determinants may contribute to the pool of patients with severe hypertriglyceridemia.

## 7.7 References

- Auer, P.L., and Lettre, G. (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Med* 7, 16.
- Boullart, A.C., de Graaf, J., and Stalenhoef, A.F. (2012). Serum triglycerides and risk of cardiovascular disease. *Biochim Biophys Acta* 1821, 867-875.
- Brahm, A.J., and Hegele, R.A. (2015). Chylomicronaemia--current diagnosis and future therapies. *Nat Rev Endocrinol* 11, 352-362.
- Castelli, W.P., Cooper, G.R., Doyle, J.T., Garcia-Palmieri, M., Gordon, T., Hames, C., Hulley, S.B., Kagan, A., Kuchmak, M., McGee, D., *et al.* (1977). Distribution of triglyceride and total, LDL and HDL cholesterol in several populations: a cooperative lipoprotein phenotyping study. *J Chronic Dis* 30, 147-169.
- Cefalu, A.B., Spina, R., Noto, D., Valenti, V., Ingrassia, V., Giammanco, A., Panno, M.D., Ganci, A., Barbagallo, C.M., and Aversa, M.R. (2015). Novel CREB3L3 Nonsense Mutation in a Family With Dominant Hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 35, 2694-2699.
- D'Erasmus, L., Di Costanzo, A., Cassandra, F., Minicocci, I., Polito, L., Montali, A., Ceci, F., and Arca, M. (2019). Spectrum of Mutations and Long-Term Clinical Outcomes in Genetic Chylomicronemia Syndromes. *Arterioscler Thromb Vasc Biol* 39, 2531-2541.
- Davis, B.D., Dron, J.S., Robinson, J.F., Hegele, R.A., and Lizotte, D.J. (2019). Exautomate: A user-friendly tool for region-based rare variant association analysis (RVAA). *BioRxiv*.
- Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.
- Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., *et al.* (2019). Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol* 13, 80-88.
- Forte, T.M., Sharma, V., and Ryan, R.O. (2016). Apolipoprotein A-V gene therapy for disease prevention / treatment: a critical analysis. *J Biomed Res* 30, 88-93.
- Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- Goldberg, I.J., Scheraldi, C.A., Yacoub, L.K., Saxena, U., and Bisgaier, C.L. (1990). Lipoprotein ApoC-II activation of lipoprotein lipase. Modulation by apolipoprotein A-IV. *J Biol Chem* 265, 4266-4272.



- Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Averna, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* 2, 655-666.
- Ionita-Laza, I., Cho, M.H., Laird, N.M. (2013). Statistical Challenges in Sequence-Based Association Studies with Population- and Family-Based Design. *Stat Biosci* 5, 54-70.
- Ito, Y., Azrolan, N., O'Connell, A., Walsh, A., and Breslow, J.L. (1990). Hypertriglyceridemia as a result of human apo CIII gene expression in transgenic mice. *Science* 249, 790-793.
- Johansen, C.T., and Hegele, R.A. (2012). The complex genetic basis of plasma triglycerides. *Curr Atheroscler Rep* 14, 227-234.
- Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011). Genetic determinants of plasma triglycerides. *J Lipid Res* 52, 189-206.
- Johansen, C.T., Wang, J., McIntyre, A.D., Martins, R.A., Ban, M.R., Lanktree, M.B., Huff, M.W., Peterfy, M., Mehrabian, M., Lusic, A.J., *et al.* (2012). Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 5, 66-72.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., *et al.* (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434-443.
- Kohan, A.B. (2015). Apolipoprotein C-III: a potent modulator of hypertriglyceridemia and cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes* 22, 119-125.
- Lambert, J.E., and Parks, E.J. (2012). Postprandial metabolism of meal triglyceride in humans. *Biochim Biophys Acta* 1821, 721-726.
- Lee, J.H., Giannikopoulos, P., Duncan, S.A., Wang, J., Johansen, C.T., Brown, J.D., Plutzky, J., Hegele, R.A., Glimcher, L.H., and Lee, A.H. (2011). The transcription factor cyclic AMP-responsive element-binding protein H regulates triglyceride metabolism. *Nat Med* 17, 812-815.
- Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 95, 5-23.
- Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Team, N.G.E.S.P.-E.L.P., Christiani, D.C., Wurfel, M.M., and Lin, X. (2012a). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 91, 224-237.
- Lee, S., Wu, M.C., and Lin, X. (2012b). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13, 762-775.

- Maeda, N., Li, H., Lee, D., Oliver, P., Quarfordt, S.H., and Osada, J. (1994). Targeted disruption of the apolipoprotein C-III gene in mice results in hypotriglyceridemia and protection from postprandial hypertriglyceridemia. *J Biol Chem* 269, 23610-23616.
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., *et al.* (2017). Rare and low-frequency coding variants alter human adult height. *Nature* 542, 186-190.
- Michailidou, K. (2018). Meta-Analysis of Common and Rare Variants. *Methods Mol Biol* 1793, 73-88.
- Nakagawa, Y., and Shimano, H. (2018). CREBH Regulates Systemic Glucose and Lipid Metabolism. *Int J Mol Sci* 19.
- National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 106, 3143-3421.
- Nho, K., Horgusluoglu, E., Kim, S., Risacher, S.L., Kim, D., Foroud, T., Aisen, P.S., Petersen, R.C., Jack, C.R., Jr., Shaw, L.M., *et al.* (2016). Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease. *BMC Med Genomics* 9 Suppl 1, 30.
- Park, J.G., Xu, X., Cho, S., and Lee, A.H. (2016). Loss of Transcription Factor CREBH Accelerates Diet-Induced Atherosclerosis in *Ldlr*<sup>-/-</sup> Mice. *Arterioscler Thromb Vasc Biol* 36, 1772-1781.
- Pirim, D., Wang, X., Radwan, Z.H., Niemsiri, V., Bunker, C.H., Barmada, M.M., Kamboh, M.I., and Demirci, F.Y. (2015). Resequencing of LPL in African Blacks and associations with lipoprotein-lipid levels. *Eur J Hum Genet* 23, 1244-1253.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (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. *Genet Med* 17, 405-424.
- Turcot, V., Lu, Y., Highland, H.M., Schurmann, C., Justice, A.E., Fine, R.S., Bradfield, J.P., Esko, T., Giri, A., Graff, M., *et al.* (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* 50, 26-41.
- van Dijk, K.W., Rensen, P.C., Voshol, P.J., and Havekes, L.M. (2004). The role and mode of action of apolipoproteins CIII and AV: synergistic actors in triglyceride metabolism? *Curr Opin Lipidol* 15, 239-246.

Weinberg, R.B., and Spector, M.S. (1985). Human apolipoprotein A-IV: displacement from the surface of triglyceride-rich particles by HDL2-associated C-apoproteins. *J Lipid Res* 26, 26-37.

Xu, X., Park, J.G., So, J.S., Hur, K.Y., and Lee, A.H. (2014). Transcriptional regulation of apolipoprotein A-IV by the transcription factor CREBH. *J Lipid Res* 55, 850-859.

## Chapter 8 – Discussion

The text contained in this Chapter has been adapted from previously published sources for brevity and to ensure consistency throughout this Dissertation.

**Dron, J.S.**, and Hegele, R.A. (2016). Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130-141.

**Dron, J.S.**, and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.

**Dron, J.S.**, and Hegele, R.A. (2018). Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.

**Dron, J.S.**, and Hegele, R.A. (2019). The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol* 30, 71-81.

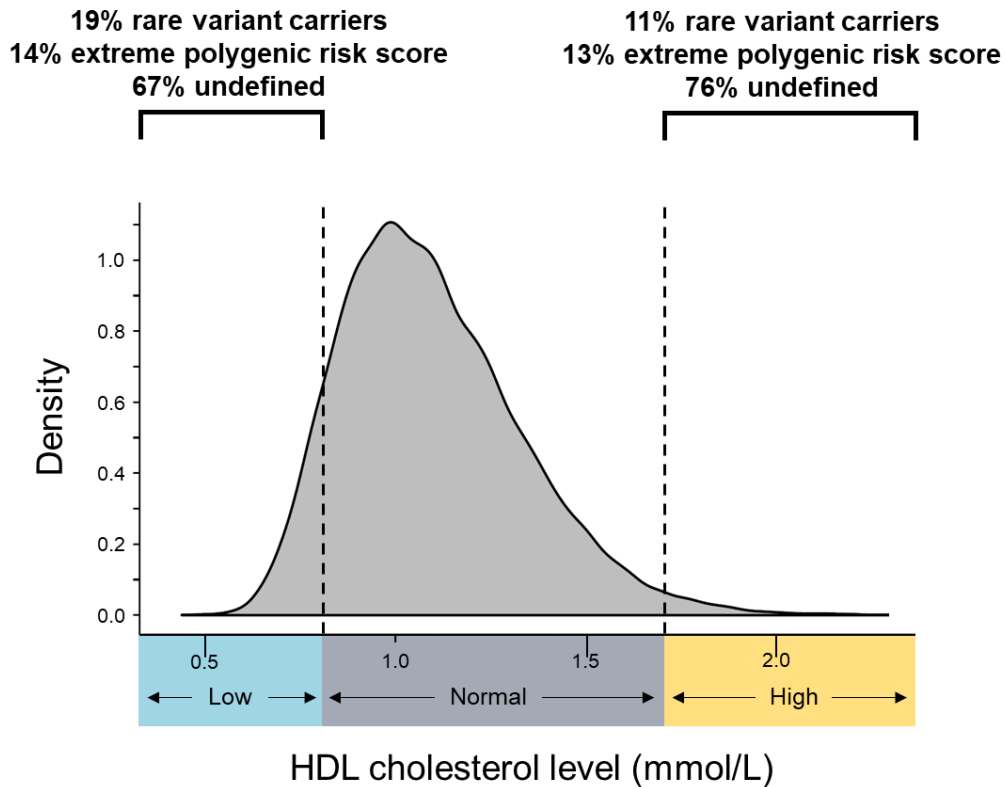
## 8.1 Overview

In this Dissertation, I have described in detail my efforts to comprehensively evaluate the genetic determinants underlying three dyslipidemia phenotypes: (i) hypoalphalipoproteinemia; (ii) hyperalphalipoproteinemia; and (iii) hypertriglyceridemia. By leveraging data produced by our laboratory's targeted next-generation sequencing (NGS) panel, LipidSeq, I have assessed a range of genetic factors—rare single-nucleotide variants (SNVs), copy-number variants (CNVs), and common single-nucleotide polymorphisms (SNPs)—across metabolically relevant genetic loci that heretofore required separate, dedicated methods for identification. With the sequencing data generated using LipidSeq, I have successfully analyzed the genetic factors of over 3,000 dyslipidemia patients and have detailed the genetic nature of each phenotype.

## 8.2 Summary of research findings

### 8.2.1 The genetic architecture of extreme high-density lipoprotein cholesterol levels

A summary for the genetic architecture of extreme deviations in high-density lipoprotein (HDL) cholesterol levels is depicted in **Figure 8.1**. In order to establish this genetic summary, DNA samples collected from individuals with hypoalphalipoproteinemia and hyperalphalipoproteinemia across North America—including at the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada), the Montréal Heart Institute (MHI) Biobank (Montréal, QC, Canada), and the University of Pennsylvania (UPenn) (Philadelphia, PA, USA)—were carefully evaluated for various genetic determinants.



**Figure 8.1 The updated genetic architecture underlying the spectrum of measurable HDL cholesterol levels.**

The distribution of HDL cholesterol levels has a normal distribution in the general population; however, it is important to note that this distribution includes both males and females. For the research described in this Dissertation, low levels of HDL cholesterol (i.e. “hypoalphalipoproteinemia”) were diagnosed for males and females with levels below 0.8 and 1.0 mmol/L, respectively. High levels of HDL cholesterol (i.e. “hyperalphalipoproteinemia”) were diagnosed for males and females with levels above 1.4 and 1.8 mmol/L, respectively. Normal levels were considered between these thresholds. The thresholds shown in this figure are not exact and are for illustrative purposes. More extreme phenotypes that fall at the tails of the distribution are more likely to have a genetic factor contributing towards the phenotype; monogenic syndromes of HDL cholesterol have either virtually non-existent levels of HDL cholesterol, or extremely high levels. The prevalence of heterozygous rare variants in genes involved in HDL metabolism was slightly higher in individuals with low HDL cholesterol levels. The accumulation of common SNPs had a similar prevalence between both extremes of HDL cholesterol. Abbreviations: HDL = high-density lipoprotein.

### 8.2.1.1 Hypoalphalipoproteinemia

The results described in [Chapters 2 and 3](#) are the summation of the first comprehensive assessment of rare SNVs, CNVs, and common variant accumulation in individuals with extremely low levels of HDL cholesterol (Dron et al., 2018; Dron et al., 2017).

A total of 686 DNA samples from patients with HDL cholesterol levels  $\leq 0.8$  mmol/L and  $\leq 1.0$  mmol/L in males and females, respectively, were collected from the Lipid Genetics Clinic, MHI Biobank, and UPenn. Initially, rare variants disrupting candidate genes with primary effects on HDL cholesterol levels were screened for in *ABCA1*, *APOA1*, and *LCAT*; rare variants were also screened for in non-candidate genes with secondary effects on HDL cholesterol. Across cohorts, it was identified that 18.7% of patients carried at least one variant likely contributing towards their hypoalphalipoproteinemia phenotype. The difference in rare variant carriers between cohorts was discussed in [Chapter 2, Section 2.5](#), and is likely due to differences in patient ascertainment and sequencing methods.

With the majority of patients lacking an identifiable rare variant, we sought to determine whether there was an excess accumulation of common small-effect SNP alleles contributing towards the hypoalphalipoproteinemia phenotype. To achieve this, we developed a novel polygenic risk score using 9 SNPs identified from previous genome-wide association studies (GWASs) that were highly associated with HDL cholesterol levels (Willer et al., 2013). The score was calculated for all 686 patients, and 12.8% had extremely low scores—this reflected a severe absence of SNP alleles associated with raising HDL cholesterol levels. Collectively, rare variant non-carriers were 1.47-times (95% CI [1.11-1.96]; one-tailed  $P < 0.01$ ) more likely to have an extremely low polygenic risk score compared to normolipidemic controls. When considering patients from the Lipid Genetics Clinic cohort alone, rare variant non-carriers were 3.00-times (95% CI [1.67-5.35]; one-tailed  $P < 0.0001$ ) more likely to have an extremely low polygenic risk score compared to normolipidemic controls.

In the Lipid Genetics Clinic cohort (N=136), 47.8% of patients had an identifiable genetic determinant likely contributing towards their phenotypic presentation of hypoalphalipoproteinemia. For the remaining 52.2% of individuals, it is possible that they carried a contributory genetic determinant that was not captured by the rare variant assessment or 9-SNP polygenic risk score. Subsequent to the publication of these results, a new bioinformatic tool became available, allowing us to leverage read-depth coverage information generated by our LipidSeq panel to identify CNVs in our sequencing data.

Including the 136 patient samples described in [Chapter 2](#), a total of 288 hypoalphalipoproteinemia patients from the Lipid Genetics Clinic were screened for CNVs in *ABCA1*, *APOA1*, and *LCAT* (Dron et al., 2018). Three unique deletions in *ABCA1* were identified across four individuals, including: (i) a heterozygous deletion of exon 4; (ii) a heterozygous deletion that spanned exons 8 to 31; and (iii) a heterozygous deletion of the entire *ABCA1* gene. These results presented in [Chapter 3](#) were the first reported instance of hypoalphalipoproteinemia patients carrying CNVs in *ABCA1* or any other candidate low HDL cholesterol gene, as the main genetic determinant for the phenotype (Dron et al., 2018).

Together, the assessment of rare SNVs, CNVs, and polygenic risk scores allowed for the most comprehensive understanding to date about the genetic determinants underlying low HDL cholesterol levels, and highlighted the polygenic component of this phenotype.

### 8.2.1.2 Hyperalphalipoproteinemia

The results presented in [Chapter 2](#) also highlight the polygenic nature of extremely high levels of HDL cholesterol through the presence of both rare SNVs and accumulation of common genetic SNPs (Dron et al., 2017).

DNA samples from 1,165 patients with HDL cholesterol levels  $\geq 1.4$  mmol/L and  $\geq 1.8$  mmol/L in males and females, respectively, were collected from the Lipid Genetics Clinic, MHI Biobank, and UPenn. Initially, rare variants disrupting candidate genes with primary effects on HDL cholesterol were screened for in genes previously linked to high HDL cholesterol phenotypes, including *LIPC*, *SCARB1*, *CETP*, and *LIPG* (Hegele et al.,



1993; Inazu et al., 1990; Tietjen et al., 2012; Zanoni et al., 2016). Rare variants in a non-candidate gene with secondary effects on HDL cholesterol, were also screened for. It was identified that 10.9% of patients carried at least one variant likely contributing towards their hyperalphalipoproteinemia phenotype.

Following the rare variant assessment, the 9-SNP polygenic risk score was calculated in all 1,165 study participants to determine common SNP accumulation. It was identified that 10.3% of individuals had extremely high scores reflecting an excess of SNP alleles associated with raising HDL cholesterol levels. Collectively, rare variant non-carriers were 2.27-times (95% CI [1.82-2.83]; one-tailed  $P < 0.0001$ ) more likely to have an extremely high polygenic risk score compared to normolipidemic controls. When considering the Lipid Genetics Clinic cohort alone, rare variant non-carriers were 2.19-times (95% CI [1.21-3.96]; one-tailed  $P < 0.01$ ) more likely to have an extremely high polygenic risk score compared to normolipidemic controls.

Between rare variants and the extreme accumulation of SNP alleles in the Lipid Genetics Clinic cohort, 30.3% of patients had an identifiable genetic determinant contributing towards their phenotypic presentation of hyperalphalipoproteinemia. The subsequent assessment for rare CNVs in the candidate genes associated with elevated levels of HDL cholesterol did not reveal any changes in copy-number.

### 8.2.1.3 Genetic influences across high-density lipoprotein cholesterol levels

Collectively, my research has illustrated the prevalence of polygenic determinants across extremes of HDL cholesterol. Rare variants—both SNVs and CNVs—are more prevalent in individuals with hypoalphalipoproteinemia compared to hyperalphalipoproteinemia. Although the polygenic accumulation of SNPs is similar between HDL cholesterol extremes, there is a slight increase of extreme polygenic risk scores in patients with hyperalphalipoproteinemia.

Across both extreme HDL cholesterol cohorts, more than half of the patients under study did not have an identifiable genetic factor relevant to their phenotype. This could suggest that in those patients, either: (i) they carry genetic factors that were not screened for;

and/or (ii) non-genetic factors—diet, medications and activity levels—may be influencing the HDL cholesterol phenotype.

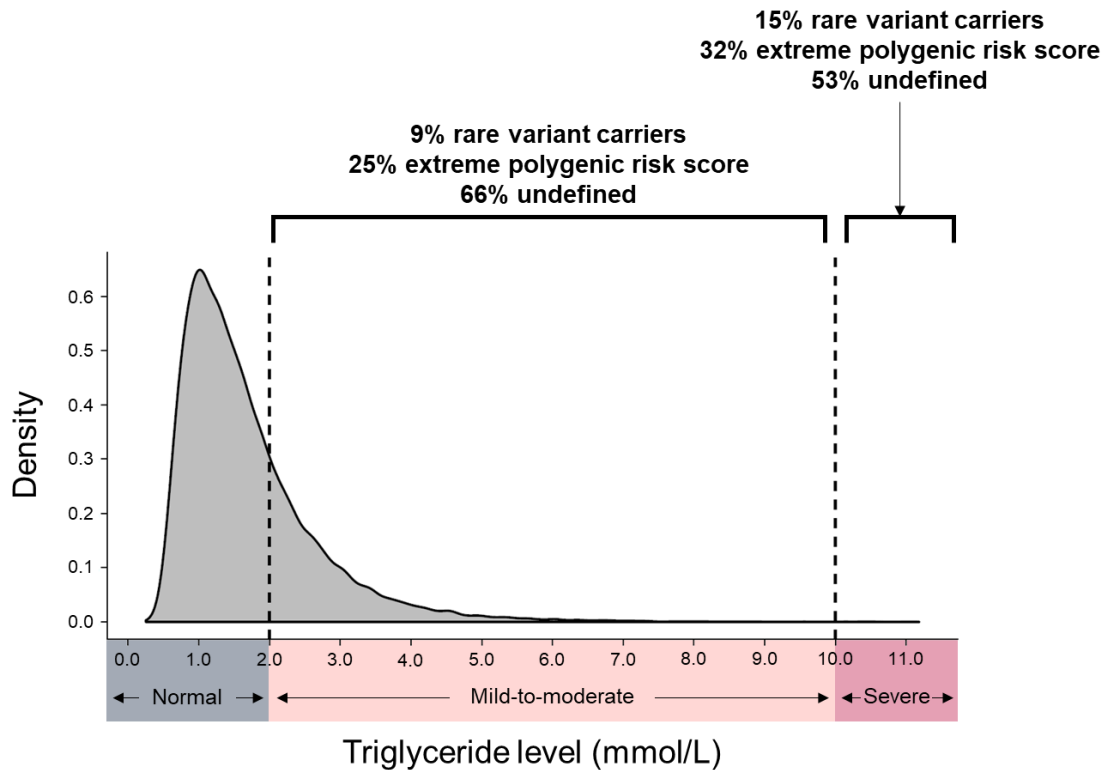
## 8.2.2 The genetic architecture of hypertriglyceridemia

A summary for the genetic architecture of hypertriglyceridemia is depicted in **Figure 8.2**. In order to establish this genetic summary, DNA samples collected from individuals with varying severities of hypertriglyceridemia across North America—including at the Lipid Genetics Clinic at the London Health Sciences Centre, University Hospital (London ON, Canada) and the Lipid, Diabetes, or Cardiology Clinics at the University of California, San Francisco (UCSF) (San Francisco, CA, USA)—were carefully evaluated for various genetic determinants.

### 8.2.2.1 Severe hypertriglyceridemia

The research described in [Chapters 4, 5, and 7](#) has culminated in the most comprehensive assessment of genetic factors in the largest cohort of severe hypertriglyceridemia patients to date (Dron et al., 2020a; Dron et al., 2019a; Dron et al., 2019b). Prior to this, studies focused on single types of genetic determinants at a time, effectively missing the overall spectrum of genetic variation contributing towards extreme elevations in triglyceride levels.

A total of 563 individuals with severe hypertriglyceridemia (triglycerides  $\geq 10$  mmol/L) were screened for rare variants disrupting canonical triglyceride metabolism genes (i.e. *LPL*, *LMF1*, *GPIHBP1*, *APOA5*, *APOC2*). We identified only a small subset of patients with the monogenic, autosomal recessive disorder, familial chylomicronemia syndrome (FCS); this highlighted the rarity of FCS, since even in a specialized cohort enriched for individuals with extremely elevated triglyceride levels, only 1.1% (6/563) of patients had FCS due to the presence of bi-allelic rare variants in a canonical triglyceride metabolism gene (Dron et al., 2019a). When considering heterozygous rare variants, 14.4% (81/563) of individuals were carriers, and were thus considered to have multifactorial chylomicronemia, a polygenic form of severe hypertriglyceridemia.



**Figure 8.2 The updated genetic architecture underlying the spectrum of measurable triglyceride levels.**

The distribution of triglyceride levels has a positive skew in the general population. Normal levels of triglyceride are considered to be less than 2.0 mmol/L. Individuals with triglyceride levels between 2.0 to 9.9 mmol/L are diagnosed with mild-to-moderate hypertriglyceridemia, while individuals with triglyceride levels above 10.0 mmol/L are diagnosed with severe hypertriglyceridemia. Although not focused on in this Dissertation, individuals with extremely low levels of triglyceride are diagnosed with hypotriglyceridemia (not shown in diagram). Severe hypertriglyceridemia cases caused by monogenic determinants (i.e. bi-allelic rare variants in triglyceride metabolism genes) are defined as familial chylomicronemia syndrome (FCS) and are extremely rare in the population, while cases driven by polygenic determinants (i.e. heterozygous rare variants in triglyceride metabolism genes and/or the extreme accumulation of SNPs) are defined as multifactorial chylomicronemia and are far more common relative to FCS.

In addition to rare variants, we sought to assess whether an excess of common SNPs with smaller phenotypic effects might also be contributing towards the hypertriglyceridemia phenotype. After developing a polygenic risk score comprised of 16 SNPs significantly associated to triglyceride levels (Willer et al., 2013), we calculated it in all patients under study and identified that 32.0% (180/563) of severe hypertriglyceridemia patients had extremely high polygenic risk scores. When considering both types of genetic determinants simultaneously, 30/87 patients (34.5%) with a rare variant also had an extreme polygenic risk score.

When considering all types of genetic determinants, severe hypertriglyceridemia patients were 4.41-times (95% CI [2.67-7.29]; one-tailed  $P < 0.0001$ ) more likely to carry a rare variant compared to normolipidemic controls, and were 4.45-times (95% CI [3.15-6.30]; one-tailed  $P < 0.0001$ ) more likely to have an extremely high polygenic risk score compared to normolipidemic controls. Overall, severe hypertriglyceridemia patients were 5.77-times (95% CI [4.26-7.82]; one-tailed  $P < 0.0001$ ) more likely to carry any type of genetic determinant linked to hypertriglyceridemia, compared to normolipidemic controls.

As part of our rare variant screening, we identified and characterized novel CNV deletions disrupting *LPL* in a single individual; in [Chapter 5](#), after further screening of 69 severe hypertriglyceridemia patients, three additional individuals were found to carry CNVs in *LPL*. Collectively, the CNVs included: (i) a heterozygous deletion spanning the 5'UTR to exon 2; and (ii) a heterozygous deletion spanning the 5'UTR to exon 1 (Dron et al., 2019b). Similarly to what has been observed for SNVs in canonical metabolism genes, the impact of CNVs on the processing of triglyceride-rich lipoproteins likely increases susceptibility for hypertriglyceridemia. Although CNVs as drivers of hypertriglyceridemia are not frequently reported, they are important phenotypic contributors that should be screened for (Iacocca et al., 2019).

[Chapter 7](#) describes further efforts to uncover genetic contributions towards severe hypertriglyceridemia susceptibility. A subset of 265 multifactorial chylomicronemia patients were screened for rare loss-of-function variants across all genes included on the

LipidSeq panel (**Figure 1.15**). Specifically, a gene-based rare variant association study (RVAS) using a variance-component test was performed between severe hypertriglyceridemia patients and normolipidemic controls to determine if rare loss-of-function variants in non-canonical triglyceride metabolism genes were susceptibility factors towards the hypertriglyceridemia phenotype (Dron et al., 2020a). We identified that multifactorial chylomicronemia patients were 20.2-times (95% CI [1.11-366.1]; two-tailed  $P=0.03$ ) more likely to carry a rare loss-of-function variant in *CREB3L3* compared to normolipidemic controls, suggesting that this gene has an important role in influencing measurable triglyceride levels and is important in the context of hypertriglyceridemia. *CREB3L3* encodes cAMP-responsive element-binding protein H (CREBH), a transcription factor expressed in the liver and small intestine, that upregulates genes involved in the hydrolysis of triglyceride-rich lipoproteins (Goldberg et al., 1990; Nakagawa and Shimano, 2018; Weinberg and Spector, 1985; Xu et al., 2014). Since an enrichment of rare variants was more substantial than what was observed in the canonical triglyceride metabolism genes, our findings suggest that screening *CREB3L3* for loss-of-function variants may be incredibly useful in identifying individuals with increased susceptibility for extremely elevated triglyceride levels.

Between rare variants in triglyceride metabolism genes, extremely high polygenic risk scores, and loss-of-function variants in *CREB3L3*, there is a variety of genetic determinants underlying severe hypertriglyceridemia. The collective findings here emphasize that the majority of severe hypertriglyceridemia cases are polygenic in nature, can be further classified as “multifactorial chylomicronemia”, and likely come about through the increased accumulation of genetic determinants that increase phenotypic susceptibility.

### 8.2.2.2 Mild-to-moderate hypertriglyceridemia

My research described in [Chapter 6](#) details the genetic profile of patients with mild-to-moderate hypertriglyceridemia and provides a clearer understanding behind the genetic architecture of this phenotype (Dron et al., 2020b).

Following the study design established for severe hypertriglyceridemia in [Chapter 4](#), rare variants disrupting the canonical triglyceride metabolism genes and a triglyceride-specific polygenic risk score were assessed in 134 individuals with mild-to-moderate hypertriglyceridemia (triglyceride between 2-9.9 mmol/L). It was determined that 9.0% (12/134) of patients were heterozygous rare variant carriers, while 24.6% (36/134) of patients had extremely high polygenic risk scores, reflecting an excess of SNP alleles associated with elevated triglyceride levels. When considering both types of genetic determinants simultaneously, only 3/12 patients (25%) with a rare variant also had an extreme polygenic risk score.

Mild-to-moderate hypertriglyceridemia patients were 2.38-times (95% CI [1.13-4.99]; one-tailed  $P=0.021$ ) more likely to carry a rare variant and 3.26-times (95% CI [2.02-5.26]; one-tailed  $P<0.0001$ ) more likely to have an extreme polygenic risk score compared to normolipidemic controls. Overall, these patients were 3.23-times (95% CI [2.08-5.02]; one-tailed  $P<0.0001$ ) more likely to carry any type of genetic determinant linked to hypertriglyceridemia, compared to normolipidemic controls.

Although the prevalence of genetic factors in patients with mild-to-moderate hypertriglyceridemia was not as high as patients with severe hypertriglyceridemia—33.6% compared to 47.4%, respectively—the overall pattern remained the same: the most common genetic determinant was an increased accumulation of common variants (as denoted by a high polygenic risk score), followed by the presence of rare variants.

### 8.2.2.3 Genetic influences across hypertriglyceridemia phenotypes

Collectively, my research has demonstrated that hypertriglyceridemia—along its spectrum of severity—is largely polygenic, with both common and rare genetic susceptibility components; except for cases of FCS, which is monogenic in nature (**Table 8.1**). Furthermore, clinical expression of the hypertriglyceridemia phenotype is likely related to qualitative and quantitative differences in the precise combination of variants in an individual's genome. A higher burden of both rare and common triglyceride-raising variants likely associates with a more extreme phenotype, such as multifactorial chylomicronemia. Additional genetic factors not considered in the contents of this

Dissertation might also contribute towards differences in phenotypic presentation as well: this includes variation impacting other genomic loci beyond what is captured by LipidSeq, as well as concepts like variant penetrance and expressivity that were not accounted for here. Importantly, secondary non-genetic factors—including diet, alcohol intake, obesity, diabetes control, liver and renal disease—are important in determining the final quantitative triglyceride phenotype, although are not discussed here (Hegele et al., 2014). These additional considerations could be used to tease apart key differences in what drives a mild-to-moderate versus severe form of hypertriglyceridemia.

**Table 8.1 Distinguishing between familial chylomicronemia syndrome, multifactorial chylomicronemia, and mild-to-moderate hypertriglyceridemia.**

	Severe hypertriglyceridemia		Mild-to-moderate hypertriglyceridemia
	Familial chylomicronemia syndrome	Multifactorial chylomicronemia	
<b>Triglyceride range (mmol/L)</b>	≥10.0		2.0 to 9.9
<b>Primarily disturbed lipoprotein fractions</b>	Chylomicrons	Chylomicrons and remnants VLDL IDL	VLDL (IDL)
<b>Genetic basis</b>	Monogenic (autosomal recessive)		Polygenic
<b>Relevant genetic determinants</b>	- Bi-allelic (homozygous or compound heterozygous) rare variants in canonical triglyceride metabolism genes ( <i>LPL</i> , <i>LMF1</i> , <i>GPIHBP1</i> , <i>APOA5</i> , <i>APOC2</i> )	- Heterozygous rare variants in canonical triglyceride metabolism genes ( <i>LPL</i> , <i>LMF1</i> , <i>GPIHBP1</i> , <i>APOA5</i> , <i>APOC2</i> ) - The accumulation of common SNPs associated with small elevations in triglyceride concentration - Rare variants in non-canonical genes peripherally involved in triglyceride metabolism (ex. <i>CREB3L3</i> )	
<b>Is there an impact from environmental determinants?</b>	- Severity of the phenotype may be exacerbated by environmental factors, but the phenotype is driven by bi-allelic variants	- Since these phenotypes are complex, a combination of genetic and environmental factors lead to the phenotype's presentation. Gene-environment interactions may account for phenotypic variability/severity	

Abbreviations: IDL= intermediate-density lipoprotein; SNPs = single-nucleotide polymorphisms; VLDL = very-low-density lipoprotein.



## 8.3 Study strengths, limitations and caveats

In Chapters 2-7 of this Dissertation, the strengths and limitations specific to each study, along with caveats for consideration, were described in the relevant Discussion sections. Here, I will describe the overarching considerations—specifically related to study design, methodologies, and technological resources—that apply to my collective research efforts.

### 8.3.1 Strengths

The targeted nature of the LipidSeq panel was the main strength of my research, as it allowed for the assessment of: (i) rare and common variants; (ii) both SNVs and CNVs; and (iii) biologically relevant genomic loci for multiple dyslipidemia phenotypes. Separate methodologies were previously required to study rare SNVs (ex. Sanger sequencing, whole-exome sequencing), CNVs, (ex. multiplex ligation-dependent probe amplification, microarray), and common variants (ex. microarray, Sanger sequencing, TaqMan genotyping); with LipidSeq, these genetic variations can be studied using the same dataset. Further, because the LipidSeq panel was designed specifically for the patients of the Lipid Genetics Clinic, the genomic loci that the panel targets are relevant to the patients' dyslipidemia and metabolic phenotypes (Dron et al., 2020c). Overall, LipidSeq led to the generation of a single dataset for robust assessment of multiple types of genetic factors and multiple phenotypes.

With respect to the identification of CNVs, this was only made possible due to the high read-depth generated by the LipidSeq panel, at almost 300-times coverage (Johansen et al., 2014). Within the last decade, computational algorithms have been developed to leverage read-depth information from NGS runs to uncover genomic areas with an enrichment or a depletion of sequencing reads, which signals the presence of a CNV (Iacocca et al., 2019). Between the depth-of-coverage of the LipidSeq panel—greater coverage provides greater confidence in identifying CNVs—and the development of the VarSeq-CNV® caller algorithm (Golden Helix, Inc., Bozeman MT, USA), we could screen the LipidSeq sequencing data for each individual and uncover CNVs disrupting phenotypically relevant genes (Iacocca et al., 2019; Iacocca et al., 2017). This provided us the opportunity to perform one of the first large-scale, NGS screening efforts for

CNVs in dyslipidemia cohorts (Dron et al., 2020c). Many novel CNVs were identified in the process (Berberich et al., 2019a; Berberich et al., 2019b; Dron et al., 2018; Dron et al., 2019b; Iacocca et al., 2017; Iacocca et al., 2018).

Beyond the benefits of LipidSeq, another strength of my research is attributed towards the number of patient DNA samples I had access to. For over 25 years, between our laboratory's research efforts and the Lipid Genetics Clinic, we have collected and sequenced DNA from over 3,000 individuals with a variety of dyslipidemia and metabolic phenotypes (Dron et al., 2020c). Because of the samples obtained through our referral clinic and from external research collaborators, our studies often boast some of the largest specialized dyslipidemia study cohorts in the field. For example, although severe hypertriglyceridemia has a population prevalence of ~1 in 600, my study cohort was comprised of over 500 patient samples. The benefit here is that extreme phenotypes—both Mendelian disorders and extreme manifestations of quantitative traits—are more likely to have a genetic basis (MacArthur et al., 2014; Panarella and Burkett, 2019), so studying cohorts enriched for individuals with extreme dyslipidemia phenotypes increases the likelihood of uncovering relevant genetic factors contributing towards disease susceptibility. Notably, the size of the dyslipidemia cohorts I had access to was directly responsible for the success of my gene-based RVAS described in [Chapter 7](#).

### 8.3.2 Caveats

When considering the data presented in this Dissertation, certain caveats should be considered for appropriate interpretation of the conclusions and implications.

The main results reported in this Dissertation were the prevalence of different genetic determinants within the study cohorts of interest—I did not quantify the estimated effects of each determinant towards the phenotype of interest (i.e. impact on disease susceptibility or effects on HDL cholesterol or triglyceride levels). Despite this, my results provide a sense of what types of genetic factors are the most common in a particular disease cohort. This information could help guide screening strategies to identify individuals at an increased genetic risk for hypoalphalipoproteinemia,

hyperalphalipoproteinemia, or hypertriglyceridemia, or could assist researchers in prioritizing what types of genetic variation should be studied further to better understand genetic-specific effects on these phenotypes (Bookman et al., 2006; Kwon and Goate, 2000). In order to quantify effect estimates of genetic variation, linear or logistic regression models could be used to assess the effect of variants towards changes in measurable lipid levels or disease presentation, respectively; however, additional information would be required as covariates to adjust these models for other variables impacting lipid phenotypes (Cole et al., 2015; Heller et al., 1993; Johnson et al., 2004).

Additional caveat considerations are in relation to the measured lipid concentrations for our study participants. Although blood samples are requested to be taken after a 12-hour fast, mechanisms were not in place to systematically confirm adherence. Individuals who did not fast likely had higher measurable lipid levels given the recent exogenous lipid source, particularly triglycerides. It is possible that non-fasting individuals may have had higher lipid measurements than normal; however, this is unlikely to have impacted patient recruitment for either extreme HDL cholesterol cohort. It is possible that non-fasting individuals could have passed the lower bound of inclusion for the mild-to-moderate hypertriglyceridemia cohort, but for the severe hypertriglyceridemia cohort, the inclusion criteria was so high, that even individuals non-compliant with the fasting recommendation would not have likely had triglyceride levels surpass that threshold (Nordestgaard et al., 2016). Although, if non-fasting individuals did present with a triglyceride profile surpassing our threshold of 10 mmol/L, that might be indicative of dysfunctional clearance of triglyceride-rich lipoprotein particles, potentially due to genetic factors—in which case, these individuals would be of interest to study. To address this potential issue of fasting vs. non-fasting in future studies, additional inclusion criteria could necessitate multiple triglyceride measurements above 10 mmol/L to ensure the severe hypertriglyceridemia phenotype is neither transient nor driven exclusively by non-genetic factors.

Another caveat is in reference to the disease study cohorts. While a huge strength of this work is related to cohorts enriched for extreme dyslipidemia phenotypes, it also means that the results are not directly translatable to a general population (Panarella and Burkett,

2019). Further, if I had determined effect estimates for the genetic determinants under study, the estimates would be inflated and have a larger magnitude compared to if I had calculated effect estimates in a cohort more representative of the general population (Panarella and Burkett, 2019).

Lastly, the polygenic risk scores developed for HDL cholesterol and triglyceride were constrained to the SNPs targeted by the LipidSeq panel. Having been designed in 2014, LipidSeq only captures the lipid-related SNP loci identified by GWAS published by that point—it does not include SNPs identified in more recent GWASs. Fortunately, because the SNPs targeted by LipidSeq were among the original loci found to be associated with lipid traits, these SNPs have larger phenotypic impacts compared to more recently uncovered loci, since they were identified in smaller study cohorts (Visscher et al., 2012). So while the polygenic risk scores used in this Dissertation were not large in terms of the number of SNPs that were included, they did include SNPs with larger phenotypic impacts.

### 8.3.3 Limitations

Clinical and biochemical variables were not systematically available for all study subjects, including: ancestry, body-mass index, diabetes status, smoking status, fasting status, diet, alcohol intake, activity level, etc. Without these additional data points, I was unable to assess environmental factors that may have been contributing towards the phenotypes under study (Cole et al., 2015; Heller et al., 1993; Johnson et al., 2004). As discussed in the previous Section, these variables could have been used as covariates in models to better estimate the effects of the genetic determinants and to uncover gene-environment or gene-gene interactions. Interestingly, recent studies have modeled how polygenic determinants alter the penetrance of a rare variant for many phenotypes, including different lipid disorders, breast cancer, Huntington's disease, and glaucoma (Craig et al., 2020; Fahed et al., 2020; Jong-Min Lee et al., 2015; Oetjens et al., 2019). This is an extremely important demonstration of the interplay between genetic determinants that was not assessed in my work, but could be in the future by following a rigorous standardization of covariate data for cases and controls, the latter for which we are extremely limited as we had no phenotypic information available.

Another limitation to consider is related to the polygenic risk scores developed in Chapters 2 and 4. The underlying assumption in the score's calculation is that the cumulative effect from different alleles works in a linearly additive manner; however, this assumption may be invalid considering the complexities and non-linearity of pathways and networks in lipid metabolism. Further, allele effect estimates are also derived under this assumption of a simple additive effect (de Vlaming and Groenen, 2015). Taken together, the method employed here for polygenic risk score derivation and calculation cannot account for potential non-linear epistatic effects that might occur in the presence of a certain combination of risk alleles. Additional work is needed to advance polygenic risk scores, specifically focusing on the linear additive assumption and assessing whether new frameworks can be established to better reflect the genetic complexities underlying different traits and disease. While there has been some effort in this area, polygenic risk scores out-perform other non-linear, machine-learning methods, suggesting that either polygenic risk scores and their linear assumptions are valid, or we have not developed the proper statistical methods to adequately address this research question (Gola et al., 2020; Vivian-Griffiths et al., 2019).

Additionally, another limitation that is applicable to almost all polygenic risk scores studies is that these scores are tailored towards European populations (Martin et al., 2019); across a 10-year span, more than 60% of studies using polygenic risk scores were made up exclusively of individuals of European ancestry (Duncan et al., 2019). This is arguably one of the biggest limitations not only in this Dissertation, but in the genetics community as a whole, as there is substantial bias and inequality in research towards Black, Indigenous and people of colour (Cell Editorial, 2020). The polygenic risk score bias is a product of selecting SNPs and their weights from GWAS that have been performed in individuals of European ancestry (Asselbergs et al., 2012; Aulchenko et al., 2009; Chasman et al., 2009; Chasman et al., 2008; Duncan et al., 2019; Kathiresan et al., 2007; Kathiresan et al., 2008; Kathiresan et al., 2009; Martin et al., 2019; Sabatti et al., 2009; Surakka et al., 2015; Teslovich et al., 2010; Willer et al., 2013; Wu et al., 2013). Because GWASs rely on linkage disequilibrium (LD) blocks and “tag SNPs” to identify SNP associations with the nearby causal variant, differences in ancestral-specific LD patterns alter association signals; a tag SNP may be associated with a phenotype in one

ancestral group but not the other, simply because the SNP does not tag the same LD blocks between groups. Since LD blocks are larger in Europeans (Shifman et al., 2003): (i) more tag SNPs are required to effectively capture these additional LD blocks in non-Europeans; and (ii) it is easier to identify an association signal in European cohorts because tag SNPs cover larger genomic regions (i.e. LD blocks) that might harbour the causative variant (Martin et al., 2019). The need for additional SNP genotypes to get the same amount of information between ancestral groups, coupled with the fact that there are fewer non-Europeans being included in GWASs, further impacts the bias (Martin et al., 2019). Fortunately, there have been efforts to increase the number of large-scale sequencing projects in non-European cohorts to identify ancestral-specific SNP associations and ancestral-specific effect estimates (Below et al., 2016; Kim et al., 2011; Liu et al., 2017; Takeuchi et al., 2012; Wu et al., 2013), which should allow for ancestral-specific polygenic risk scores. There are also efforts to develop methods for trans-ancestry polygenic risk scores, which could be applied to individuals of different ancestral groups (Wang et al., 2020b).

As an extension of the aforementioned point, another limitation in this Dissertation was that the study cohorts were made up of individuals of European ancestry, due to our geographic location in Southern Ontario; to match our ancestry breakdown, collaborators could only provide European patient samples of as well. This prevented us from determining if the genetic determinants underlying different dyslipidemia phenotypes were consistent across ancestral groups, or if the genetic profile varied.

Lastly, due to the LipidSeq panel design, novel gene discovery was not a feasible component of the research described in this Dissertation. While the targeted design provided huge strengths in terms of studying candidate genes related to each phenotype, this constraint prevented the discovery of genes with previously unknown links to HDL cholesterol or triglyceride metabolism; whole-exome sequencing would have provided this discovery opportunity. The gene-based RVAS that included non-candidate genes described in [Chapter 7](#) was an effective alternative to novel gene discovery. *CREB3L3* had been reported previously in the literature with links to triglyceride levels, but there had not been strong evidence in human subjects linking it to severe hypertriglyceridemia

until recently (D'Erasmus et al., 2019). Rather than discovering a new gene, I instead provided additional evidence to support and further substantiate the importance of *CREB3L3* in the context of hypertriglyceridemia.

## 8.4 Applications and future directions

With my research helping to enhance the foundational understanding of the genetic basis of extreme lipid disorders, there are now avenues to further explore the complex network of contributory factors towards these dyslipidemia phenotypes, and areas where this information could be applied towards more clinically relevant applications.

### 8.4.1 Estimating effects of genetic determinants

Future studies could quantify the effect estimates of specific genetic determinants towards either: (i) measurable changes to HDL cholesterol or triglyceride levels; and/or (ii) susceptibility for hypoalphalipoproteinemia, hyperalphalipoproteinemia, or hypertriglyceridemia. This work could be further expanded to consider interactive effects between rare variants and the accumulation of common SNPs, similarly to what was done in previous studies that assessed how a polygenic background could modify variant penetrance (Craig et al., 2020; Fahed et al., 2020; Jong-Min Lee et al., 2015; Oetjens et al., 2019).

Assessing how the penetrance and expressivity of rare variants is polygenically modified through the use of polygenic risk scores is an area of extreme interest, as it is a relatively unexplored area in the lipids field. If researchers are able to quantify genetic effects and determine which factors have the largest contributions towards a particular disease state, then this information could be utilized in genetic screening endeavours to identify individuals, for example, at high risk for hypertriglyceridemia.

### 8.4.2 Screening for genetic risk

Screening individuals earlier in life for genetic factors increasing their risk for dyslipidemia provides an opportunity to proactively alter lifestyle behaviours to more aggressively combat negative genetic influences towards lipid profiles (Khera et al., 2016). An early indication of being at high risk for a lipid disorder could also prompt

individuals to have their blood lipid profile monitored more frequently to catch when their lipids exceed a particular threshold, warranting medical attention and treatment.

There is also the possibility that future studies could reveal therapeutic treatments tailored towards individuals with a particular genetic determinant (e.g. variants disrupting a particular gene) driving their phenotype or individuals who fall within a certain stratification of genetic risk (e.g. top 95<sup>th</sup> percentile of a polygenic risk score) (Mars et al., 2020). For example, inhibitors of proteins with key roles in different lipoprotein processing pathways have been of great clinical benefit, including: (i) evolocumab to inhibit proprotein convertase subtilisin/kexin type 9 (PCSK9) and lower levels of low-density lipoprotein (LDL) cholesterol in individuals with increased genetic risk for cardiovascular disease (CVD) (Marston et al., 2020); (ii) volanesorsen to inhibit apolipoprotein (apo) C-III and lower levels of triglyceride in individuals with FCS (Witztum et al., 2019); and (iii) evinacumab to inhibit angiopoietin-like protein 3 (ANGPTL3) and lower levels of LDL cholesterol in individuals with homozygous familial hypercholesterolemia (FH) (Raal et al., 2020). This level of specificity between an individual and therapy—down to the genetic level—is considered precision or personalized medicine, in which a therapeutic treatment is completely tailored towards the individual and their phenotype etiology. In the future, additional therapies may become available that are particularly effective for individuals with a high polygenic risk score for a particular dyslipidemia.

### 8.4.3 Updating lipid-based polygenic risk scores

With each additional lipid-centric GWAS, larger cohorts have revealed a larger number of significantly associated SNPs. In 2010, a GWAS of ~100,000 people identified 95 SNP loci significantly associated at genome-wide levels with at least one plasma lipid trait (Teslovich et al., 2010). In 2013 and 2018, when the sample sizes increased to ~188,000 and >600,000 people, respectively, an additional 62 (Willer et al., 2013) and 118 (Klarin et al., 2018) new SNPs reached genome-wide levels of significance. The effect sizes of the newly associated SNP loci were very small—larger sample sizes permit the identification of SNPs with very small effects (Visscher et al., 2012). With additional statistically significant loci, researchers can incorporate more SNPs into their



risk scores when using the P-value threshold and pruning method, described in [Chapter 1, Section 1.3.4.3](#). Further, improved methods to derive SNP weights have been developed by accounting for LD patterns and adjusting for the underlying genetic architecture of the phenotype of interest (Choi et al., 2020; Ge et al., 2019; Vilhjalmsón et al., 2015; Wang et al., 2020b); these effect weights can be incorporated into weighted polygenic scores for a more accurate measure of polygenic determinants for a particular phenotype. Concerted efforts must also be made to derive risk scores that can be utilized effectively in populations of non-European ancestry, through the use of SNPs and estimated effects derived from non-European populations and methods that account for ancestral LD patterns (Wang et al., 2020b).

#### 8.4.3.1 Genome-wide scores

As polygenic scores and risk scores grew to encompass millions of SNP loci—the majority of them with non-significant trait effects—they were defined as “genome-wide scores”. These scores came to the forefront of polygenic research when Khera et al. described five different scores for five common diseases: coronary artery disease, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer (Khera et al., 2018a). In this study, the prevalence of individuals with extremely high genome-wide risk scores was compared to the prevalence of individuals carrying rare variants that conferred similar degrees of risk. Specifically for coronary artery disease, when considering genetic determinants that conferred a 3-fold increased risk for disease, individuals with high genome-wide risk scores were 20-fold as frequent in the population compared to rare variant carriers (Khera et al., 2018a). This incredible finding not only demonstrated the importance of genome-wide scores and using them to find more individuals at risk for disease, but it also demonstrated that considering the polygenic nature of common diseases and complex traits was extremely informative, despite the smaller associated effects from common SNPs. Genome-wide scores have since been used to consider early-onset myocardial infarction (Khera et al., 2018b), weight and obesity trajectories (Khera et al., 2019), ischemic stroke (Hachiya et al., 2020), severe hypercholesterolemia (Natarajan et al., 2018; Ripatti et al., 2020), and hypertriglyceridemia (Ripatti et al., 2020).

Specifically related to the hypertriglyceridemia genome-wide score, its degree of association was assessed against both triglyceride measurements and coronary artery disease risk in the Finnish National FINRISK Study population cohort and FinnGen project cohort, respectively (Ripatti et al., 2020). The authors demonstrated that the score could explain 5.1% of variation in triglyceride levels, and individuals with scores in the 90<sup>th</sup> percentile had a 1.3-fold increased risk for coronary artery disease (Ripatti et al., 2020).

With respect to genome-wide risk scores, efforts have already been made in assessing their practicality in non-European groups in a concerted effort to deal with ancestral biases related to polygenic-based methodologies, as discussed in the previous section (Wang et al., 2020a).

#### 8.4.4 Finding additional susceptibility genes

Gene-based RVAS may help to uncover genes with previously unappreciated or unreported links to HDL cholesterol and triglyceride metabolism. In consideration of the data presented in [Chapter 7](#), with a larger sample cohort and exome-level data, a similar gene-based approach could be used to determine if there are non-candidate genes beyond what is targeted by LipidSeq that are enriched for rare variants and driving dyslipidemia phenotypes. Since there was a large proportion of study subjects in my Dissertation without an identifiable genetic factor related to their phenotype, a gene-based RVAS using exome sequencing data in those individuals might uncover genes with some currently unappreciated link towards HDL cholesterol or triglyceride metabolism, or a novel mechanistic pathway all together. For example, a recent study performed gene-based RVASs for over 4,000 phenotypes using almost 50,000 exomes from the UK Biobank (Cirulli et al., 2020). Their findings related to HDL cholesterol and triglyceride levels showed a number of known metabolic genes, as well as genes not directly implicated with these two lipid traits, which provide new avenues of exploration. Although *CREB3L3* did not appear in their top results related to triglyceride levels, the UK Biobank is made up largely healthy volunteers, which is distinct from the cohort of patients with severe hypertriglyceridemia studied in [Chapter 7](#).

## 8.5 Risk for cardiovascular disease, and levels of high-density lipoprotein cholesterol and triglyceride

As we deepen our understanding of the genetic underpinnings of extreme circulating levels of HDL cholesterol and triglyceride, it becomes more feasible to critically assess these traits and their relationship with CVD, as well as develop a better appreciation for previous studies in this space.

While HDL cholesterol levels remain a widely used risk predictor for CVD (Anderson et al., 2016; Grundy et al., 2019), bypassing this metabolic measurement and instead relying on associated genetic factors has not been an effective predictor. CVD was not a consistent outcome in individuals with monogenic forms of hypoalphalipoproteinemia. Despite understanding the genetic cause and mechanism leading to Tangier disease, apo A-I deficiency, familial LCAT deficiency (FLD) and fish-eye disease (FED), premature CVD was not explicitly shown to associate with these syndromes (Rader and Hovingh, 2014). Even in cases of extremely high HDL cholesterol levels due to cholesteryl ester transfer protein (CETP) deficiency, there was no clear consensus on whether there was protection against CVD (Rader and Hovingh, 2014). Further, Mendelian randomization studies—an epidemiological approach that relies on genetic variants to assess causality of a modifiable exposure (i.e. lipids) on a particular phenotypic outcome (i.e. CVD), by leveraging the understanding of genetic variation with known associations to the modifiable exposure (Emdin et al., 2017)—demonstrated that genetic variants associated with HDL cholesterol levels beyond *ABCA1*, *APOA1*, *LCAT* and *CETP* failed to show causal links to CVD outcomes (Burgess and Thompson, 2015; Do et al., 2013; Voight et al., 2012). In many of these studies, only a small subset of relevant genetic factors were considered. As demonstrated by the findings presented in this Dissertation, multiple types of genetic determinants are responsible for driving HDL cholesterol levels, particularly towards extremes of the distribution (Dron et al., 2017). It remains to be seen whether a collective assessment of multiple genetic determinants, both rare and common, related to HDL cholesterol levels would associate with CVD. As mentioned in [Chapter 1, Section 1.1](#), it has been shown that the functionality of HDL or the number of HDL particles are better metrics to assess CVD risk compared to measurable HDL cholesterol levels

(Mackey et al., 2012; Mora et al., 2013). Perhaps future studies should focus on the genetic determinants related to HDL functionality and cholesterol efflux, rather than measurable levels of HDL cholesterol.

In contrast, genetic variants associated with triglyceride concentration have shown stronger associations with CVD risk. For example, a number of loss-of-function variants in *APOC3* were shown to reduce triglyceride levels and coronary artery disease risk (Jorgensen et al., 2014; Pollin et al., 2008; Tg et al., 2014). Further, a genome-wide score of ~6 million SNPs showed an association to coronary artery disease as well (Ripatti et al., 2020); however, adjustments were not made for HDL cholesterol levels. A challenge the field has faced in this space has been disentangling the joint, inverse association between HDL cholesterol and triglyceride levels, as both traits are often simultaneously abnormal when CVD associations are observed (see Table 1 from Dron and Hegele, 2017) (Clee et al., 2001; Dewey et al., 2016; Do et al., 2015; Jorgensen et al., 2014; Mailly et al., 1996; Myocardial Infarction et al., 2016; Nordestgaard, 2016; Teslovich et al., 2010; Tg et al., 2014; Triglyceride Coronary Disease Genetics et al., 2010). There are few studies that show an association with only one of the aforementioned lipid traits and CVD risk. For example, after model adjustments, Do et al. identified that genetic determinants with predominantly triglyceride-related effects were correlated with increased coronary heart disease risk, while genetic determinants with predominantly HDL cholesterol-related effects were not (Do et al., 2013). This triglyceride-specific association might be related to the cholesterol content of triglyceride-rich lipoprotein particles and their remnants, specifically very-low-density lipoproteins (VLDL) and intermediate-density lipoprotein (IDL) (FERENCE et al., 2019; Varbo et al., 2013). This aligns with the association between CVD risk and mild-to-moderate hypertriglyceridemia: disturbances in levels of VLDL and IDL lead to elevations in triglyceride levels, and due to their cholesterol content, are likely also contributing towards CVD risk through atherosclerotic plaque development (Dron and Hegele, 2017). This is in contrast to severe hypertriglyceridemia, in which chylomicrons—with a very small cholesterol content—are the main lipoprotein disturbance, and the overall CVD risk is almost negligible. In considering differing severities of hypertriglyceridemia and the

associated health complications, it is evident that knowing what lipoprotein fractions are disturbed is useful in assessing CVD risk.

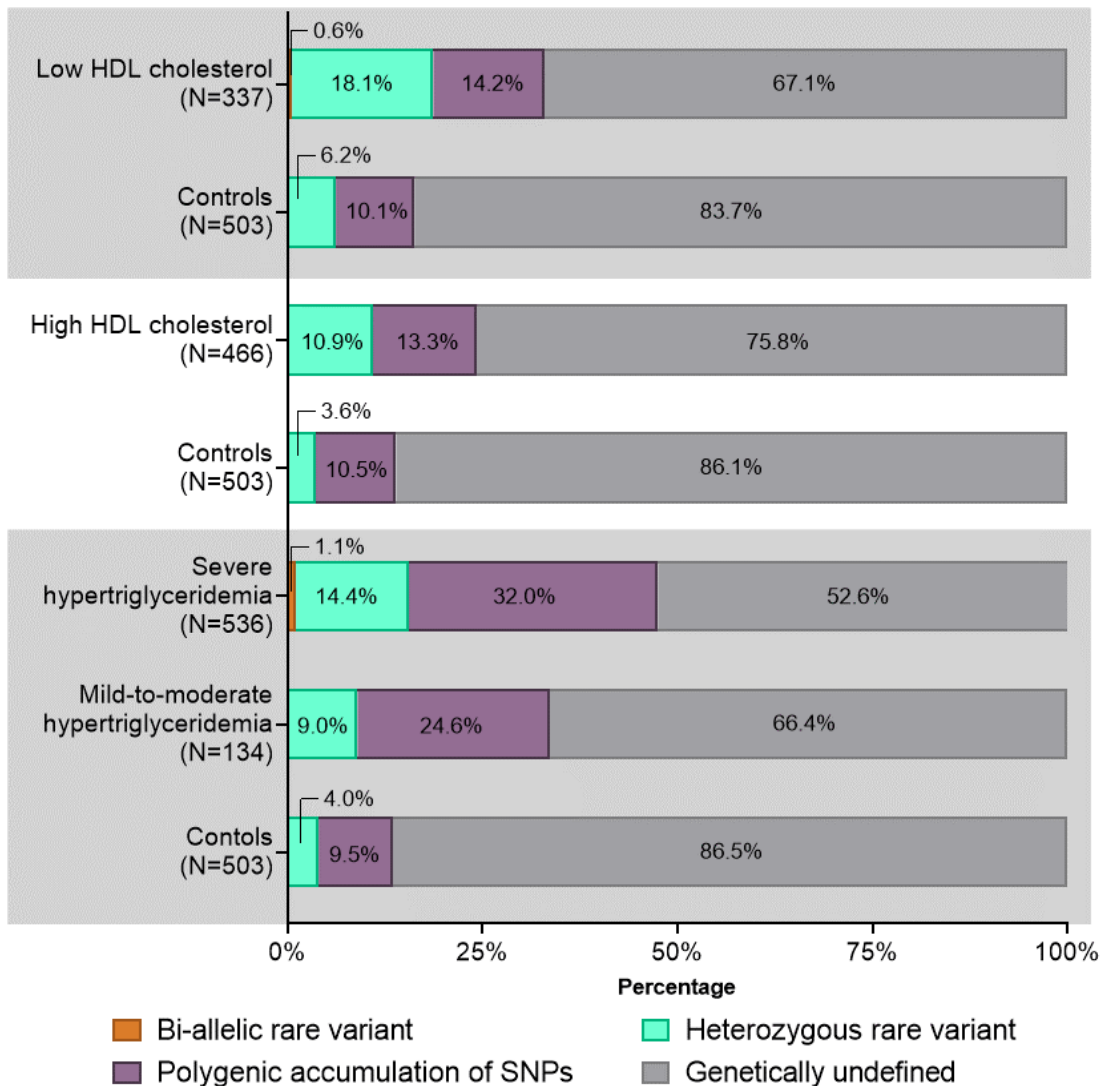
With respect to both HDL particles and triglyceride-rich lipoproteins, it can be challenging to assess fraction breakdown, lipid-content per fraction, and lipoprotein particle numbers on a large-scale because of more involved techniques and assays; however, it might provide better risk predictions for CVD compared to the measurable circulating concentration of HDL cholesterol and triglyceride. If this holds true, then future genetic studies could look at associations between genetic variants and these measurements (i.e. fraction breakdown, lipoprotein functionality) to eventually work towards a genetic test for CVD prediction earlier in life. Or, findings from these studies may reveal an area of lipid and lipoprotein metabolic pathways that might be an attractive target for future therapies attempting to reduce CVD risk.

To summarize, without a clear causal relationship or independent association, assessing the predictive power of single genetic variants related to either HDL cholesterol or triglyceride levels and CVD is not an ideal course of action. Either a collective genetic assessment spanning multiple types of determinants or coupling genetic data with functional information on lipoprotein fraction-specific data, might prove more useful in CVD risk prediction and possibly narrowing the focus towards mechanistically impactful metabolic areas that are therapeutically targetable for CVD risk reduction.

## 8.6 Conclusions

Fully understanding the genetic architecture of dyslipidemia is challenging. The perturbed lipid traits defining these phenotypes—cholesterol and triglyceride—are influenced by a complex network of genetic determinants that differ in population frequency, physical size, sequence ontology, and phenotypic impact. Throughout my Dissertation, I have assessed the diverse spectrum of genetic determinants present in groups of patients with different dyslipidemia phenotypes, including: (i) hypoalphalipoproteinemia ([Chapters 2-3](#)); (ii) hyperalphalipoproteinemia ([Chapter 2](#)); and (iii) hypertriglyceridemia ([Chapters 4-7](#)). This was made possible using the targeted NGS panel, LipidSeq, that produced a single dataset from which I could perform a robust

set of genetic analyses. From my research, I have demonstrated that despite being jointly considered as “lipid disorders”, each phenotype studied has a distinct genetic profile (**Figure 8.3**). By better understanding the genetic underpinnings of HDL cholesterol, triglyceride, and their dyslipidemic counterparts, future efforts can explore the relationship between these phenotypes and their co-morbidities, such as CVD. As demonstrated previously, genetics often provides invaluable insights into the biological mechanisms driving health and disease.



**Figure 8.3 The comparison of genetic profiles of different dyslipidemia phenotypes.**

The percentage of individuals in each cohort that carried a particular type of genetic determinant relevant to the phenotype under study, either hypoalphalipoproteinemia (i.e. low HDL cholesterol), hyperalphalipoproteinemia (i.e. high HDL cholesterol), severe hypertriglyceridemia (including familial chylomicronemia syndrome and multifactorial chylomicronemia), or mild-to-moderate hypertriglyceridemia. Abbreviations: HDL = high-density lipoprotein; SNPs = single-nucleotide polymorphisms.

## 8.7 References

Anderson, T.J., Gregoire, J., Pearson, G.J., Barry, A.R., Couture, P., Dawes, M., Francis, G.A., Genest, J., Jr., Grover, S., Gupta, M., *et al.* (2016). 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult. *Can J Cardiol* 32, 1263-1282.

Asselbergs, F.W., Guo, Y., van Iperen, E.P., Sivapalaratnam, S., Tragante, V., Lanktree, M.B., Lange, L.A., Almqguera, B., Appelman, Y.E., Barnard, J., *et al.* (2012). Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet* 91, 823-838.

Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., *et al.* (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 41, 47-55.

Below, J.E., Parra, E.J., Gamazon, E.R., Torres, J., Krithika, S., Candille, S., Lu, Y., Manichakul, A., Peralta-Romero, J., Duan, Q., *et al.* (2016). Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Sci Rep* 6, 19429.

Berberich, A.J., Huot, C., Cao, H., McIntyre, A.D., Robinson, J.F., Wang, J., and Hegele, R.A. (2019a). Copy Number Variation in GCK in Patients With Maturity-Onset Diabetes of the Young. *J Clin Endocrinol Metab* 104, 3428-3436.

Berberich, A.J., Mokashi, A., McIntyre, A.D., Robinson, J.F., Cao, H., Wang, J., and Hegele, R.A. (2019b). Bioinformatic detection of copy number variation in HNF4A causing maturity onset diabetes of the young. *Clin Genet* 96, 376-377.

Bookman, E.B., Langehorne, A.A., Eckfeldt, J.H., Glass, K.C., Jarvik, G.P., Klag, M., Koski, G., Motulsky, A., Wilfond, B., Manolio, T.A., *et al.* (2006). Reporting genetic results in research studies: summary and recommendations of an NHLBI working group. *Am J Med Genet A* 140, 1033-1040.

Burgess, S., and Thompson, S.G. (2015). Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 181, 251-260.

Cell Editorial Team. (2020). Science Has a Racism Problem. *Cell* 181, 1443-1444.

Chasman, D.I., Pare, G., Mora, S., Hopewell, J.C., Peloso, G., Clarke, R., Cupples, L.A., Hamsten, A., Kathiresan, S., Malarstig, A., *et al.* (2009). Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLoS Genet* 5, e1000730.



Chasman, D.I., Pare, G., Zee, R.Y., Parker, A.N., Cook, N.R., Buring, J.E., Kwiatkowski, D.J., Rose, L.M., Smith, J.D., Williams, P.T., *et al.* (2008). Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet* 1, 21-30.

Choi, S.W., Mak, T.S., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc.* 15, 2759-2772.

Cirulli, E.T., White, S., Read, R.W., Elhanan, G., Metcalf, W.J., Tanudjaja, F., Fath, D.M., Sandoval, E., Isaksson, M., Schlauch, K.A., *et al.* (2020). Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun* 11, 542.

Clee, S.M., Loubser, O., Collins, J., Kastelein, J.J., and Hayden, M.R. (2001). The LPL S447X cSNP is associated with decreased blood pressure and plasma triglycerides, and reduced risk of coronary artery disease. *Clin Genet* 60, 293-300.

Cole, C.B., Nikpay, M., and McPherson, R. (2015). Gene-environment interaction in dyslipidemia. *Curr Opin Lipidol* 26, 133-138.

Craig, J.E., Han, X., Qassim, A., Hassall, M., Cooke Bailey, J.N., Kinzy, T.G., Khawaja, A.P., An, J., Marshall, H., Gharahkhani, P., *et al.* (2020). Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nat Genet* 52, 160-166.

D'Erasmus, L., Di Costanzo, A., Cassandra, F., Minicocci, I., Polito, L., Montali, A., Ceci, F., and Arca, M. (2019). Spectrum of Mutations and Long-Term Clinical Outcomes in Genetic Chylomicronemia Syndromes. *Arterioscler Thromb Vasc Biol* 39, 2531-2541.

de Vlaming, R., and Groenen, P.J. (2015). The Current and Future Use of Ridge Regression for Prediction in Quantitative Genetics. *Biomed Res Int* 2015, 143712.

Dewey, F.E., Gusarova, V., O'Dushlaine, C., Gottesman, O., Trejos, J., Hunt, C., Van Hout, C.V., Habegger, L., Buckler, D., Lai, K.M., *et al.* (2016). Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N Engl J Med* 374, 1123-1133.

Do, R., Stitzel, N.O., Won, H.H., Jorgensen, A.B., Duga, S., Angelica Merlini, P., Kiezun, A., Farrall, M., Goel, A., Zuk, O., *et al.* (2015). Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* 518, 102-106.

Do, R., Willer, C.J., Schmidt, E.M., Sengupta, S., Gao, C., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., *et al.* (2013). Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet* 45, 1345-1352.

Dron, J.S., Dillio, A.A., Lawson, A., McIntyre, A.D., Davis, B.D., Wang, J., Cao, H., Movsesyan, I., Malloy, M.J., Pullinger, C.R., *et al.* (2020a). Loss-of-Function CREB3L3

- Variants in Patients With Severe Hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* *40*, 1935-1941.
- Dron, J.S., and Hegele, R.A. (2017). Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* *19*, 31.
- Dron, J.S., Wang, J., Berberich, A.J., Iacocca, M.A., Cao, H., Yang, P., Knoll, J., Tremblay, K., Brisson, D., Netzer, C., *et al.* (2018). Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia. *J Lipid Res* *59*, 1529-1535.
- Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., *et al.* (2019a). Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol* *13*, 80-88.
- Dron, J.S., Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhinds, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* *58*, 2162-2170.
- Dron, J.S., Wang, J., McIntyre, A.D., Cao, H., and Hegele, R.A. (2020b). The polygenic nature of mild-to-moderate hypertriglyceridemia. *J Clin Lipidol* *14*, 28-34 e22.
- Dron, J.S., Wang, J., McIntyre, A.D., Cao, H., Robinson, J.F., Duell, P.B., Manjoo, P., Feng, J., Movsesyan, I., Malloy, M.J., *et al.* (2019b). Partial LPL deletions: rare copy-number variants contributing towards severe hypertriglyceridemia. *J Lipid Res* *60*, 1953-1958.
- Dron, J.S., Wang, J., McIntyre, A.D., Iacocca, M.A., Robinson, J.F., Ban, M.R., Cao, H., and Hegele, R.A. (2020c). Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Med Genomics* *13*, 23.
- Duncan, L., Shen, H., Gelaye, B., Meijssen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* *10*, 3328.
- Emdin, C.A., Khera, A.V., and Kathiresan, S. (2017). Mendelian Randomization. *JAMA* *318*, 1925-1926.
- Fahed, A.C., Wang, M., Homburger, J.R., Patel, A.P., Bick, A.G., Neben, C.L., Lai, C., Brockman, D., Philippakis, A., Ellinor, P.T., *et al.* (2020). Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* *11*, 3635.
- Ference, B.A., Kastelein, J.J.P., Ray, K.K., Ginsberg, H.N., Chapman, M.J., Packard, C.J., Laufs, U., Oliver-Williams, C., Wood, A.M., Butterworth, A.S., *et al.* (2019). Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. *JAMA* *321*, 364-373.

- Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10, 1776.
- Gola, D., Erdmann, J., Muller-Myhsok, B., Schunkert, H., and Konig, I.R. (2020). Polygenic risk scores outperform machine learning methods in predicting coronary artery disease status. *Genet Epidemiol* 44, 125-138.
- Goldberg, I.J., Scheraldi, C.A., Yacoub, L.K., Saxena, U., and Bisgaier, C.L. (1990). Lipoprotein ApoC-II activation of lipoprotein lipase. Modulation by apolipoprotein A-IV. *J Biol Chem* 265, 4266-4272.
- Grundy, S.M., Stone, N.J., Bailey, A.L., Beam, C., Birtcher, K.K., Blumenthal, R.S., Braun, L.T., de Ferranti, S., Faiella-Tommasino, J., Forman, D.E., *et al.* (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol* 73, e285-e350.
- Hachiya, T., Hata, J., Hirakawa, Y., Yoshida, D., Furuta, Y., Kitazono, T., Shimizu, A., and Ninomiya, T. (2020). Genome-Wide Polygenic Score and the Risk of Ischemic Stroke in a Prospective Cohort: The Hisayama Study. *Stroke* 51, 759-765.
- Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Averna, M., Boren, J., Bruckert, E., Catapano, A.L., Descamps, O.S., *et al.* (2014). The polygenic nature of hypertriglyceridaemia: implications for definition, diagnosis, and management. *Lancet Diabetes Endocrinol* 2, 655-666.
- Hegele, R.A., Little, J.A., Vezina, C., Maguire, G.F., Tu, L., Wolever, T.S., Jenkins, D.J., and Connelly, P.W. (1993). Hepatic lipase deficiency. Clinical, biochemical, and molecular genetic characteristics. *Arterioscler Thromb Vasc Biol* 13, 720-728.
- Heller, D.A., de Faire, U., Pedersen, N.L., Dahlen, G., and McClearn, G.E. (1993). Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med* 328, 1150-1156.
- Iacocca, M.A., Dron, J.S., and Hegele, R.A. (2019). Progress in finding pathogenic DNA copy number variations in dyslipidemia. *Curr Opin Lipidol* 30, 63-70.
- Iacocca, M.A., Wang, J., Dron, J.S., Robinson, J.F., McIntyre, A.D., Cao, H., and Hegele, R.A. (2017). Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia. *J Lipid Res* 58, 2202-2209.
- Iacocca, M.A., Wang, J., Sarkar, S., Dron, J.S., Lagace, T., McIntyre, A.D., Lau, P., Robinson, J.F., Yang, P., Knoll, J.H., *et al.* (2018). Whole-Gene Duplication of PCSK9 as a Novel Genetic Mechanism for Severe Familial Hypercholesterolemia. *Can J Cardiol* 34, 1316-1324.

Inazu, A., Brown, M.L., Hesler, C.B., Agellon, L.B., Koizumi, J., Takata, K., Maruhama, Y., Mabuchi, H., and Tall, A.R. (1990). Increased high-density lipoprotein levels caused by a common cholesteryl-ester transfer protein gene mutation. *N Engl J Med* *323*, 1234-1238.

Johansen, C.T., Dube, J.B., Loyzer, M.N., MacDonald, A., Carter, D.E., McIntyre, A.D., Cao, H., Wang, J., Robinson, J.F., and Hegele, R.A. (2014). LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* *55*, 765-772.

Johnson, J.L., Slentz, C.A., Duscha, B.D., Samsa, G.P., McCartney, J.S., Houmard, J.A., and Kraus, W.E. (2004). Gender and racial differences in lipoprotein subclass distributions: the STRRIDE study. *Atherosclerosis* *176*, 371-377.

Jong-Min Lee, Vanessa C. Wheeler, Michael J. Chao, Jean Paul G. Vonsattel, Ricardo Mouro Pinto, Diane Lucente, Kawther Abu-Elneel, Eliana Marisa Ramos, Jayalakshmi Srinidhi Mysore, Tammy Gillis, *et al.* (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* *162*, 516-526.

Jorgensen, A.B., Frikke-Schmidt, R., Nordestgaard, B.G., and Tybjaerg-Hansen, A. (2014). Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* *371*, 32-41.

Kathiresan, S., Manning, A.K., Demissie, S., D'Agostino, R.B., Surti, A., Guiducci, C., Gianniny, L., Burt, N.P., Melander, O., Orho-Melander, M., *et al.* (2007). A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Medical Genom* *8 Suppl 1*, S17.

Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burt, N.P., Rieder, M.J., Cooper, G.M., Roos, C., Voight, B.F., Havulinna, A.S., *et al.* (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* *40*, 189-197.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., *et al.* (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* *41*, 56-65.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., *et al.* (2018a). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* *50*, 1219-1224.

Khera, A.V., Chaffin, M., Wade, K.H., Zahid, S., Brancale, J., Xia, R., Distefano, M., Senol-Cosar, O., Haas, M.E., Bick, A., *et al.* (2019). Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* *177*, 587-596 e589.

Khera, A.V., Chaffin, M., Zekavat, S.M., Collins, R.L., Roselli, C., Natarajan, P., Lichtman, J.H., D'Onofrio, G., Mattered, J.A., Dreyer, R.P., *et al.* (2018b). Whole Genome

Sequencing to Characterize Monogenic and Polygenic Contributions in Patients Hospitalized with Early-Onset Myocardial Infarction. *Circulation*. 139, 1593-1602.

Khera, A.V., Emdin, C.A., Drake, I., Natarajan, P., Bick, A.G., Cook, N.R., Chasman, D.I., Baber, U., Mehran, R., Rader, D.J., *et al.* (2016). Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. *N Engl J Med* 375, 2349-2358.

Kim, Y.J., Go, M.J., Hu, C., Hong, C.B., Kim, Y.K., Lee, J.Y., Hwang, J.Y., Oh, J.H., Kim, D.J., Kim, N.H., *et al.* (2011). Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet* 43, 990-995.

Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., *et al.* (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet* 50, 1514-1523.

Kwon, J.M., and Goate, A.M. (2000). The candidate gene approach. *Alcohol Res Health* 24, 164-168.

Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin, C., Alam, D., Alves, A.C., *et al.* (2017). Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet* 49, 1758-1766.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., *et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature* 508, 469-476.

Mackey, R.H., Greenland, P., Goff, D.C., Jr., Lloyd-Jones, D., Sibley, C.T., and Mora, S. (2012). High-density lipoprotein cholesterol and particle concentrations, carotid atherosclerosis, and coronary events: MESA (multi-ethnic study of atherosclerosis). *J Am Coll Cardiol* 60, 508-516.

Maily, F., Fisher, R.M., Nicaud, V., Luong, L.A., Evans, A.E., Marques-Vidal, P., Luc, G., Arveiler, D., Bard, J.M., Poirier, O., *et al.* (1996). Association between the LPL-D9N mutation in the lipoprotein lipase gene and plasma lipid traits in myocardial infarction survivors from the ECTIM Study. *Atherosclerosis* 122, 21-28.

Mars, N., Koskela, J.T., Ripatti, P., Kiiskinen, T.T.J., Havulinna, A.S., Lindbohm, J.V., Ahola-Olli, A., Kurki, M., Karjalainen, J., Palta, P., *et al.* (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 26, 549-557.

Marston, N.A., Kamanu, F.K., Nordio, F., Gurmu, Y., Roselli, C., Sever, P.S., Pedersen, T.R., Keech, A.C., Wang, H., Lira Pineda, A., *et al.* (2020). Predicting Benefit From Evolocumab Therapy in Patients With Atherosclerotic Disease Using a Genetic Risk Score: Results From the FOURIER Trial. *Circulation* 141, 616-623.

Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 51, 584-591.

Mora, S., Glynn, R.J., and Ridker, P.M. (2013). High-density lipoprotein cholesterol, size, particle number, and residual vascular risk after potent statin therapy. *Circulation* 128, 1189-1197.

Myocardial Infarction, G., Investigators, C.A.E.C., Stitzel, N.O., Stirrups, K.E., Masca, N.G., Erdmann, J., Ferrario, P.G., Konig, I.R., Weeke, P.E., Webb, T.R., *et al.* (2016). Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* 374, 1134-1144.

Nakagawa, Y., and Shimano, H. (2018). CREBH Regulates Systemic Glucose and Lipid Metabolism. *Int J Mol Sci* 19.

Natarajan, P., Peloso, G.M., Zekavat, S.M., Montasser, M., Ganna, A., Chaffin, M., Khera, A.V., Zhou, W., Bloom, J.M., Engreitz, J.M., *et al.* (2018). Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat Commun* 9, 3391.

Nordestgaard, B.G. (2016). Triglyceride-Rich Lipoproteins and Atherosclerotic Cardiovascular Disease: New Insights From Epidemiology, Genetics, and Biology. *Circ Res* 118, 547-563.

Nordestgaard, B.G., Langsted, A., Mora, S., Kolovou, G., Baum, H., Bruckert, E., Watts, G.F., Sypniewska, G., Wiklund, O., Boren, J., *et al.* (2016). Fasting Is Not Routinely Required for Determination of a Lipid Profile: Clinical and Laboratory Implications Including Flagging at Desirable Concentration Cutpoints-A Joint Consensus Statement from the European Atherosclerosis Society and European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem* 62, 930-946.

Oetjens, M.T., Kelly, M.A., Sturm, A.C., Martin, C.L., and Ledbetter, D.H. (2019). Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nat Commun* 10, 4897.

Panarella, M., and Burkett, K.M. (2019). A Cautionary Note on the Effects of Population Stratification Under an Extreme Phenotype Sampling Design. *Front Genet* 10, 398.

Pollin, T.I., Damcott, C.M., Shen, H., Ott, S.H., Shelton, J., Horenstein, R.B., Post, W., McLenithan, J.C., Bielak, L.F., Peyser, P.A., *et al.* (2008). A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* 322, 1702-1705.

Raal, F.J., Rosenson, R.S., Reeskamp, L.F., Hovingh, G.K., Kastelein, J.J.P., Rubba, P., Ali, S., Banerjee, P., Chan, K.C., Gipe, D.A., *et al.* (2020). Evinacumab for Homozygous Familial Hypercholesterolemia. *N Engl J Med* 383, 711-720.

- Rader, D.J., and Hovingh, G.K. (2014). HDL and cardiovascular disease. *Lancet* 384, 618-625.
- Ripatti, P., Ramo, J.T., Mars, N.J., Fu, Y., Lin, J., Soderlund, S., Benner, C., Surakka, I., Kiiskinen, T., Havulinna, A.S., *et al.* (2020). Polygenic Hyperlipidemias and Coronary Artery Disease Risk. *Circ Genom Precis Med* 13, e002725.
- Sabatti, C., Service, S.K., Hartikainen, A.L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C.G., Zaitlen, N.A., Varilo, T., Kaakinen, M., *et al.* (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 41, 35-46.
- Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., and Darvasi, A. (2003). Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12, 771-776.
- Surakka, I., Horikoshi, M., Magi, R., Sarin, A.P., Mahajan, A., Lagou, V., Marullo, L., Ferreira, T., Miraglio, B., Timonen, S., *et al.* (2015). The impact of low-frequency and rare variants on lipid levels. *Nat Genet* 47, 589-597.
- Takeuchi, F., Isono, M., Katsuya, T., Yokota, M., Yamamoto, K., Nabika, T., Shimokawa, K., Nakashima, E., Sugiyama, T., Rakugi, H., *et al.* (2012). Association of genetic variants influencing lipid levels with coronary artery disease in Japanese individuals. *PLoS One* 7, e46385.
- Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-713.
- Tg, Hdl Working Group of the Exome Sequencing Project, N.H.L., Blood, I., Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitzel, N.O., Lange, L.A., Lu, Y., *et al.* (2014). Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* 371, 22-31.
- Tietjen, I., Hovingh, G.K., Singaraja, R.R., Radomski, C., Barhdadi, A., McEwen, J., Chan, E., Mattice, M., Legendre, A., Franchini, P.L., *et al.* (2012). Segregation of LIPG, CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol. *PLoS One* 7, e37437.
- Triglyceride Coronary Disease Genetics, C., Emerging Risk Factors, C., Sarwar, N., Sandhu, M.S., Ricketts, S.L., Butterworth, A.S., Di Angelantonio, E., Boekholdt, S.M., Ouwehand, W., Watkins, H., *et al.* (2010). Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet* 375, 1634-1639.
- Varbo, A., Benn, M., Tybjaerg-Hansen, A., Jorgensen, A.B., Frikke-Schmidt, R., and Nordestgaard, B.G. (2013). Remnant cholesterol as a causal risk factor for ischemic heart disease. *J Am Coll Cardiol* 61, 427-436.

- Vilhjalmsson, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindstrom, S., Ripke, S., Genovese, G., Loh, P.R., Bhatia, G., Do, R., *et al.* (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet* 97, 576-592.
- Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. *Am J Hum Genet* 90, 7-24.
- Vivian-Griffiths, T., Baker, E., Schmidt, K.M., Bracher-Smith, M., Walters, J., Artemiou, A., Holmans, P., O'Donovan, M.C., Owen, M.J., Pocklington, A., *et al.* (2019). Predictive modeling of schizophrenia from genomic data: Comparison of polygenic risk score with kernel support vector machines approach. *Am J Med Genet B Neuropsychiatr Genet* 180, 80-85.
- Voight, B.F., Peloso, G.M., Orho-Melander, M., Frikke-Schmidt, R., Barbalic, M., Jensen, M.K., Hindy, G., Holm, H., Ding, E.L., Johnson, T., *et al.* (2012). Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. *Lancet* 380, 572-580.
- Wang, M., Menon, R., Mishra, S., Patel, A.P., Chaffin, M., Tanneeru, D., Deshmukh, M., Mathew, O., Apte, S., Devanboo, C.S., *et al.* (2020a). Validation of a Genome-Wide Polygenic Score for Coronary Artery Disease in South Asians. *J Am Coll Cardiol* 76, 703-714.
- Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P.M., and Yengo, L. (2020b). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat Commun* 11, 3865.
- Weinberg, R.B., and Spector, M.S. (1985). Human apolipoprotein A-IV: displacement from the surface of triglyceride-rich particles by HDL2-associated C-apoproteins. *J Lipid Res* 26, 26-37.
- Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013). Discovery and refinement of loci associated with lipid levels. *Nat Genet* 45, 1274-1283.
- Witztum, J.L., Gaudet, D., Freedman, S.D., Alexander, V.J., Digenio, A., Williams, K.R., Yang, Q., Hughes, S.G., Geary, R.S., Arca, M., *et al.* (2019). Volanesorsen and Triglyceride Levels in Familial Chylomicronemia Syndrome. *N Engl J Med* 381, 531-542.
- Wu, Y., Waite, L.L., Jackson, A.U., Sheu, W.H., Buyske, S., Absher, D., Arnett, D.K., Boerwinkle, E., Bonnycastle, L.L., Carty, C.L., *et al.* (2013). Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* 9, e1003379.
- Xu, X., Park, J.G., So, J.S., Hur, K.Y., and Lee, A.H. (2014). Transcriptional regulation of apolipoprotein A-IV by the transcription factor CREBH. *J Lipid Res* 55, 850-859.



Zanoni, P., Khetarpal, S.A., Larach, D.B., Hancock-Cerutti, W.F., Millar, J.S., Cuchel, M., DerOhannessian, S., Kontush, A., Surendran, P., Saleheen, D., *et al.* (2016). Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science* 351, 1166-1171.

## Appendices

### Appendix A. Copyright Permissions

**Dron, J.S., and Hegele, R.A. (2016).** Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130-141.

#### Rights and permissions

This article is published under an open access license. Please check the 'Copyright Information' section for details of this license and what re-use is permitted. If your intended use exceeds what is permitted by the license or if you are unable to locate the licence and re-use information, please contact the [Rights and Permissions team](#).

#### About this article



Check for updates

#### Cite this article

Dron, J.S., Hegele, R.A. Genetics of Lipid and Lipoprotein Disorders and Traits. *Curr Genet Med Rep* 4, 130–141 (2016). <https://doi.org/10.1007/s40142-016-0097-y>

[Download citation](#)

Published	Issue Date
07 June 2016	September 2016

DOI  
<https://doi.org/10.1007/s40142-016-0097-y>

#### Author reuse

Please check the Copyright Transfer Statement (CTS) or Licence to Publish (LTP) that you have signed with Springer Nature to find further information about the reuse of your content.

Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published article in their thesis according to current citation standards.

**Dron, J.S., and Hegele, R.A. (2017).** Genetics of Triglycerides and the Risk of Atherosclerosis. *Curr Atheroscler Rep* 19, 31.

#### Genetics of Triglycerides and the Risk of Atherosclerosis

**Author:** Jacqueline S. Dron et al  
**Publication:** Current Atherosclerosis Reports  
**Publisher:** Springer Nature  
**Date:** May 23, 2017

Copyright © 2017, Springer Nature

**SPRINGER NATURE**

#### Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.  
To request permission for a type of use not listed, please contact [Springer Nature](#)

**Dron, J.S., and Hegele, R.A. (2018).** Polygenic influences on dyslipidemias. *Curr Opin Lipidol* 29, 133-143.



Wolters Kluwer

### Polygenic influences on dyslipidemias

Author: Jacqueline S. Dron and Robert A. Hegele

Publication: Current Opinion in Lipidology

Publisher: Wolters Kluwer Health, Inc.

Date: Apr 1, 2018

*Copyright © 2018, Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.*

#### License Not Required

Wolters Kluwer policy permits only the final peer-reviewed manuscript of the article to be reused in a thesis. You are free to use the final peer-reviewed manuscript in your print thesis at this time, and in your electronic thesis 12 months after the article's publication date. The manuscript may only appear in your electronic thesis if it will be password protected. Please see our Author Guidelines here: [https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?\\_mzts=636410951730000000](https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?_mzts=636410951730000000).

BACK

CLOSE WINDOW

**Dron, J.S., and Hegele, R.A. (2019).** The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol* 30, 71-81.



Wolters Kluwer

### The evolution of genetic-based risk scores for lipids and cardiovascular disease

Author: Jacqueline S. Dron and Robert A. Hegele

Publication: Current Opinion in Lipidology

Publisher: Wolters Kluwer Health, Inc.

Date: Apr 1, 2019

*Copyright © 2019, Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.*

#### License Not Required

Wolters Kluwer policy permits only the final peer-reviewed manuscript of the article to be reused in a thesis. You are free to use the final peer-reviewed manuscript in your print thesis at this time, and in your electronic thesis 12 months after the article's publication date. The manuscript may only appear in your electronic thesis if it will be password protected. Please see our Author Guidelines here: [https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?\\_mzts=636410951730000000](https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?_mzts=636410951730000000).

BACK

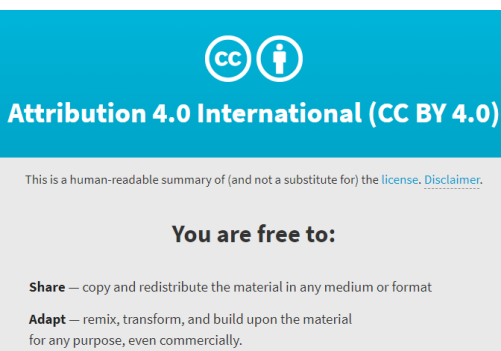
CLOSE WINDOW

**Dron, J.S.,** Wang, J., Low-Kam, C., Khetarpal, S.A., Robinson, J.F., McIntyre, A.D., Ban, M.R., Cao, H., Rhinds, D., Dube, M.P., *et al.* (2017). Polygenic determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res* 58, 2162-2170.

Received August 9, 2017.  
Revision received August 31, 2017.

Copyright © 2017 by the American Society for Biochemistry and Molecular Biology, Inc.

**Author's Choice**—Final version free via [Creative Commons CC-BY license](#).

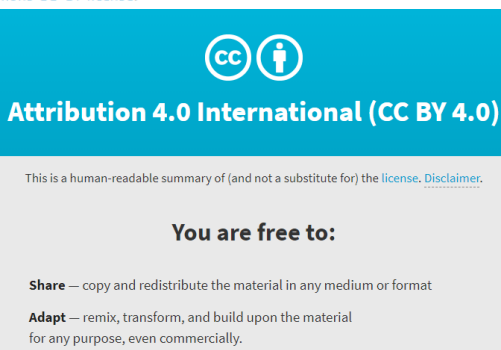


**Dron, J.S.,** Wang, J., Berberich, A.J., Iacocca, M.A., Cao, H., Yang, P., Knoll, J., Tremblay, K., Brisson, D., Netzer, C., *et al.* (2018). Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia. *J Lipid Res* 59, 1529-1535.

Received April 23, 2018.  
Revision received May 21, 2018.

Copyright © 2018 Dron et al. Published by The American Society for Biochemistry and Molecular Biology, Inc.

**Author's Choice**—Final version open access under the terms of the [Creative Commons CC-BY license](#).



**Dron, J.S.**, Wang, J., McIntyre, A.D., Cao, H., Robinson, J.F., Duell, P.B., Manjoo, P., Feng, J., Movsesyan, I., Malloy, M.J., *et al.* (2019). Partial LPL deletions: rare copy-number variants contributing towards severe hypertriglyceridemia. *J Lipid Res* 60, 1953-1958.

Journal of lipid research : JLR

E-JOURNAL

ISSN: 1539-7262

Publication year(s): 1959 - present

Author/Editor: American Society for Biochemistry and Molecular Biology

Publisher: AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, INC.

Rightsholder: AMERICAN SOC FOR BIOCHEMISTRY & MOLECULAR BIOLOGY

Language: English

Country of publication: United States of America



**!** Journal of Lipid Research permissions policies: • If you are an author of the content for which you are seeking permission, or if you are not an author but are requesting permission to copy, distribute, transmit and adapt the work for noncommercial purposes (e.g. reproduction of a figure for educational purposes such as schoolwork, or appending a reprinted article to a PhD dissertation), you do not need to seek permission using the options listed below, as long as any reuse includes the credit line in the reuse policies listed above. • Parties who are not authors on the article who wish to reuse content for commercial purposes such as reproducing a figure in a book, journal, or coursepack published by a commercial publisher, do need permission and should request permission by completing the form below. For more information please see Journal of Lipid Research: <http://www.jlr.org/site/misc/policy/edpolicy.xhtml#Copyright>

**Dron, J.S.**, Wang, J., McIntyre, A.D., Iacocca, M.A., Robinson, J.F., Ban, M.R., Cao, H., and Hegele, R.A. (2020). Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Med Genomics* 13, 23.

Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias

Author: Jacqueline S. Dron et al

**SPRINGER NATURE**

Publication: BMC Medical Genomics

Publisher: Springer Nature

Date: Feb 10, 2020

Copyright © 2020, Springer Nature

#### Creative Commons

This is an open access article distributed under the terms of the [Creative Commons CC BY](#) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

You are not required to obtain permission to reuse this article.

CC0 applies for supplementary material related to this article and attribution is not required.

**Dron, J.S., Wang, J., Cao, H., McIntyre, A.D., Iacocca, M.A., Menard, J.R., Movsesyan, I., Malloy, M.J., Pullinger, C.R., Kane, J.P., et al. (2019). Severe hypertriglyceridemia is primarily polygenic. J Clin Lipidol 13, 80-88.**



#### Severe hypertriglyceridemia is primarily polygenic

**Author:**

Jacqueline S. Dron, Jian Wang, Henian Cao, Adam D. McIntyre, Michael A. Iacocca, Jyler R. Menard, Irina Movsesyan, Mary J. Malloy, Clive R. Pullinger, John P. Kane, Robert A. Hegele

**Publication:** Journal of Clinical Lipidology

**Publisher:** Elsevier

**Date:** January-February 2019

© 2018 National Lipid Association. Published by Elsevier Inc.

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

**Dron, J.S., Wang, J., McIntyre, A.D., Cao, H., and Hegele, R.A. (2020). The polygenic nature of mild-to-moderate hypertriglyceridemia. J Clin Lipidol 14, 28-34 e22.**



#### The polygenic nature of mild-to-moderate hypertriglyceridemia

**Author:** Jacqueline S. Dron, Jian Wang, Adam D. McIntyre, Henian Cao, Robert A. Hegele

**Publication:** Journal of Clinical Lipidology

**Publisher:** Elsevier

**Date:** January-February 2020

© 2020 National Lipid Association. All rights reserved.

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

**Dron, J.S.,** Dilliot, A.A., Lawson, A., McIntyre, A.D., Davis, B.D., Wang, J., Cao, H., Movsesyan, I., Malloy, M.J., Pullinger, C.R., *et al.* (2020). Loss-of-Function CREB3L3 Variants in Patients With Severe Hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* 40, 1935-1941.



Wolters Kluwer

#### Loss-of-Function CREB3L3 Variants in Patients With Severe Hypertriglyceridemia

Author: Robert A. Hegele, John P. Kane, Clive R. Pullinger, et al

Publication: ATVB

Publisher: Wolters Kluwer Health, Inc.

Date: Jun 25, 2020

Copyright © 2020, Wolters Kluwer Health

#### License Not Required

Wolters Kluwer policy permits only the final peer-reviewed manuscript of the article to be reused in a thesis. You are free to use the final peer-reviewed manuscript in your print thesis at this time, and in your electronic thesis 12 months after the article's publication date. The manuscript may only appear in your electronic thesis if it will be password protected. Please see our Author Guidelines here: [https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?\\_mzts=636410951730000000](https://cdn-tp2.mozu.com/16833-m1/cms/files/Author-Document.pdf?_mzts=636410951730000000).

BACK

CLOSE WINDOW

**Dron, J.S., and Hegele, R.A. (2017). Molecular Genetics of Hypertriglyceridemia. In eLS, L.E. John Wiley & Sons, ed., pp. 1-8.**

10/21/2020

RightsLink Printable License

JOHN WILEY AND SONS LICENSE  
TERMS AND CONDITIONS

Oct 21, 2020

This Agreement between Ms. Jacqueline Dron ('You') and John Wiley and Sons ('John Wiley and Sons') consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 4933751420524

License date Oct 21, 2020

Licensed Content Publisher John Wiley and Sons

Licensed Content Publication Wiley Books

Licensed Content Title Molecular Genetics of Hypertriglyceridemia

Licensed Content Author Robert A Hegele, Jacqueline S Dron

Licensed Content Date Mar 6, 2017

Licensed Content Pages 8

Type of use Dissertation/Thesis

Requestor type University/Academic

Format Print and electronic

Portion Text extract

Number of Pages 2

Will you be translating? No

Title The genetic determinants of complex lipid and lipoprotein phenotypes

Institution name Western University

Expected presentation date Dec 2020

Portions Under the subheading "Genetic Determinants of Hypertriglyceridemia" Select sentences starting on page 2 to page 5. Not consecutive text blocks.

Requestor Location

Publisher Tax ID EUS26007151

Billing Type Invoice

Billing Address



10/21/2020

RightsLink Printable License

Total 0.00 CAD

## Terms and Conditions

## TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

## Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, and any CONTENT (PDF or image file) purchased as part of your order, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of reproduction or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgment given to the author, title of the material/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. For STM Signatory Publishers clearing permission under the terms of the STM Signatory Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest in any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.
- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES, AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

<https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=0154bd01-2d4d-4f2a-b7fb-1986f4b70922>

2/3

10/21/2020

RightsLink Printable License

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing, signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the OCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supercedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regard to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

#### WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the Wiley Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

##### The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

##### Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (see below)

##### Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution-Non-Commercial-NoDerivs License \(CC-BY-NC-ND\)](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

##### Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library  
<http://about.wiley.com/WileyCDA/Section/id-410695.html>

##### Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

## Appendix B. University of Western Ontario - Ethics Approval



# Western Research

**Date:** 15 October 2020

**To:** Dr. 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:** November 3, 2020

**Date Approval Issued:** 15/Oct/2020 06:13

**REB Approval Expiry Date:** 03/Nov/2021

---

Dear Dr. 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,

The Office of Human Research Ethics

*Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).*

### Appendix C. Summary of unique rare SNVs identified in primary HDL cholesterol genes.

Gene	Nucleotide change	Amino acid change	Sequence Ontology	Allele count	HGMD	<i>In silico</i> predictions*	Carrier's HDL cholesterol phenotype	
ABCA1	c.103A>G	p.I35V	Missense	1		22.9   Benign   Damaging   Damaging	Low	
	c.206G>T	p.W69L	Missense	2		26.3   Probably damaging   Damaging   Damaging	Low	
	c.208delG	p.V70fsX53	Frameshift	2		22.9   NA   NA   NA	Low	
	c.688C>T	p.R230C	Missense	1	Yes	19.22   Benign   Tolerated   Tolerated	Low	
	c.1660T>C	p.Y554H	Missense	1		27.5   Possibly damaging   Damaging   Damaging	Low	
	c.1770G>C	p.W590C	Missense	1	Yes	24.7   Possibly damaging   Damaging   Damaging	Low	
	c.1769G>T	p.W590L	Missense	1	Yes	24.7   Possibly damaging   Tolerated   Damaging	Low	
	c.2270T>C	p.L757P	Missense	2		28   Probably damaging   Damaging   Damaging	Low	
	c.2328G>C	p.K776N	Missense	2	Yes	27.6   Probably damaging   Damaging   Damaging	Low	
	c.2551G>A	p.G851R	Missense	2		34   Probably damaging   Damaging   Damaging	Low	
	c.2819C>T	p.T940M	Missense	1	Yes	32   Probably damaging   Damaging   Damaging	Low	
	c.3191A>G	p.D1064G	Missense	1		29.5   Probably damaging   Damaging   Damaging	Low	
	c.3343_3344delTC	p.S1115PfsX31	Frameshift	1	Yes	35   NA   NA   NA	Low	
	c.3544G>A	p.A1182T	Missense	1		20.9   Benign   Tolerated   Damaging	Low	
	c.4156G>A	p.E1386K	Missense	1		24.7   Benign   Tolerated   Damaging	Low	
	c.4430G>T	p.C1477F	Missense	1	Yes	34   Probably damaging   Damaging   Damaging	Low	
	c.5398A>C	p.N1800H	Missense	2	Yes	26.6   Possibly damaging   Damaging   Damaging	Low	
	IVS42+1G>A			Splicing	1		27.2   NA   NA   NA	Low
	c.5672A>C	p.E1891A	Missense	1		25.2   Probably damaging   Damaging   Damaging	Low	
c.5774G>A	p.R1925Q	Missense	1		24.8   Benign   Tolerated   Damaging	Low		
APOA1	c.85dupC	p.Q29PfsX29	Frameshift	2	Yes	28.3   NA   NA   NA	Low	
	c.535delC	p.H179MfsX45	Frameshift	1		24.8   Probably damaging   Damaging   Damaging	Low	
	c.566C>G	p.P189R	Missense	1	Yes	24.8   Probably damaging   Damaging   Damaging	Low	
	c.718C>T	p.Q240X	Nonsense	1	Yes	35   NA   NA   Damaging	Low	
CETP	c.164delC	p.S56AfsX11	Frameshift	1		36   NA   NA   Damaging	High	
	c.976C>T	p.Q326X	Nonsense	1	Yes	32   NA   NA   NA	High	
LCAT	c.109_110delAC	p.T37AfsX3	Frameshift	1		37   NA   NA   Damaging	Low	
	c.321C>A	p.Y107X	Nonsense	1	Yes	31   Probably damaging   Tolerated   Damaging	Low	
	c.491G>A	p.R164H	Missense	2	Yes	27.8   Probably damaging   Damaging   Damaging	Low	
	c.892A>C	p.T298P	Missense	1		31   Probably damaging   Damaging   Damaging	Low	
	c.997G>A	p.V333M	Missense	1	Yes	31   Probably damaging   Damaging   Damaging	Low	
	c.1039C>T	p.R347C	Missense	1	Yes	18.3   Benign   Damaging   Tolerated   Damaging	Low	
LIPC	c.1244A>G	p.N415S	Missense	1	Yes	22.8   Possibly damaging   Tolerated   Damaging	Low	
	c.193C>T	p.R65X	Nonsense	1		35   NA   NA   Damaging	High	
	c.866C>T	p.S289F	Missense	1	Yes	25.5   Probably damaging   Damaging   Damaging	High	
	c.1214C>T	p.T405M	Missense	3	Yes	24.6   Probably damaging   Damaging   Damaging	High	
LIPG	c.1231G>C	p.G411R	Missense	1		27.5   Probably damaging   Damaging   Damaging	High	
	IVS5+1G>T			Splicing	1		23.2   NA   NA   Damaging	High
SCARB1	c.1187A>G	p.N396S	Missense	4		23.8   Probably damaging   Tolerated   Damaging	High	
	c.520C>T	p.R174C	Missense	1		32   Probably damaging   Damaging   Damaging	High	
	c.1258G>T	p.G420W	Missense	1		31   Probably damaging   Damaging   Damaging	Low	

“\*” The order of prediction tool outcomes is: CADD PHRED Score, PolyPhen2, SIFT, and MutationTaster. Abbreviations: NA = not available; del = deletion; dup = duplication; ext = extension; fs = frameshift; HDL = HDL cholesterol; HGMD = Human Gene Mutation Database; SNV = single-nucleotide variant.

### Appendix D. Summary of unique rare SNVs identified in secondary HDL cholesterol genes.

Gene	Nucleotide change	Amino acid change	Sequence Ontology	Allele count	HGMD	<i>In silico</i> predictions*	Carrier's HDL cholesterol phenotype
<i>APOA5</i>	c.944C>T	p.A315V	Missense	1	Yes	28   Probably damaging   Damaging   Tolerated	Low
<i>APOC3</i>	IVS2+1G>A		Splicing	2	Yes	25.1   NA   NA   Damaging	High
<i>APOE</i>	c.433G>C	p.G145R	Missense	1		28.4   Probably damaging   Damaging   Tolerated	Low
	c.805C>G	p.R269G	Missense	2	Yes	25.6   Benign   Damaging   Damaging	Low
<i>GPD1</i>	c.208C>A	p.P70T	Missense	1		31   Probably damaging   Damaging   Damaging	Low
	c.760G>A	p.E254K	Missense	1		27.7   Probably damaging   Damaging   Damaging	Low
<i>LMF1</i>	c.1351C>T	p.R451W	Missense	2	Yes	24.9   Probably damaging   Damaging   Tolerated	Low
	c.1405G>A	p.A469T	Missense	2	Yes	29.7   Probably damaging   Damaging   Damaging	Low
<i>LPL</i>	c.644G>A	p.G215E	Missense	1	Yes	22   Probably damaging   Tolerated   Damaging	Low
	c.701C>T	p.P234L	Missense	1	Yes	34   Probably damaging   Damaging   Damaging	Low

“\*” The order of prediction tool outcomes is: CADD PHRED Score, PolyPhen2, SIFT, and MutationTaster. Abbreviations: NA = not available; del = deletion; dup = duplication; ext = extension; fs = frameshift; HDL = HDL cholesterol; HGMD = Human Gene Mutation Database; SNV = single-nucleotide variant.

**Appendix E. Screening primers for *ABCA1* copy-number variations.**

CNV	Breakpoint	Primer direction	Primer sequence (5' to 3')	Annealing temperature (°C)	Primer labels in Figure 3.3 and Figure 3.4
Exon 4	Upstream	FWD	CCAAATAGCTGAGACTACAGGCATG	60	P1
		REV	GTGATGGTGAAGGTATTTTCAG	60	P2
	Downstream	FWD	CATGACTGCATTGGTATAAAGATG	60	P3
		REV	ATCACTGTCTGTGGCAACCAG	60	P4
Exons 8 to 31	Upstream	FWD	GACCCAGCTTCCAATCTTCATAATCCTC	60	P5
		REV	GGTTGCAAAGATCCCTGTAGAG	60	P6
	Downstream	FWD	GAGATATCATGTTGGGAGGGTCTG	60	P7
		REV	GCCACAGTCTGTCCTGTGACTTTAC	60	P8
Full deletion	Upstream	FWD	TATCATGCTACTCAGAACAGCATG	60	P9
		REV	TGGTGATTCTTGTGTGCACAAAG	60	P10
	Downstream	FWD	CAGGATATTACATAGGTAAGCAGG	60	P11
		REV	CTTAATGATAGTGAAGACAAGGAG	60	P12

The primers listed were designed to flank the two breakpoints for each CNV. The “Breakpoint” listed is relative to the deletion section of the gene. The sequence orientation for P1-P8 are relative to the *ABCA1* gene, while the sequence orientation for P9-P12 are relative to the full chromosome. Abbreviations: CNV = copy-number variant; FWD = forward; REV = reverse.

### Appendix F. Summary of unique rare SNVs and CNVs identified in the Lipid Genetics Clinic cohort (N=251).

Variant information		RefSeq gene information				Minor allele frequencies			In silico predictions			Allele Counts
Gene	Chr:Pos	Ref/Alt	HGVS c.	HGVS p.	Sequence Ontology	ExAC	1KG	CADD PHRED	SIFT	PolyPhen2	Mutation Taster	
APOA5	11:116661290	C/G	c.655G>C	p.Ala219Pro	Missense			22.9	T	PoD	T	1
	11:116661305	C/G	c.640G>C	p.Ala214Pro	Missense			24.2	D	ProD	D	2
	11:116661335	G/A	c.610C>T	p.Arg204Cys	Missense	8.27E-06	0.000199681	33	D	PoD	D	2
	11:116661656	G/A	c.289C>T	p.Gln97Ter	Nonsense			36			D	1
	11:116663095	C/T	c.-33+1G>A		Splice donor			24.4				1
APOC2	19:45452024	A/C	c.122A>C	p.Lys41Thr	Missense	0.0008731	0.000399361	15	D	B	D	4
LMFI	16:904642	C/T	c.1594G>A	p.Gly532Ser	Missense			29.7	D	ProD	D	1
	16:919894	C/T	c.1405G>A	p.Ala469Thr	Missense	0.0008543	0.000599042	29.7	D	ProD	D	1
	16:919912	C/T	c.1387G>A	p.Asp463Asn	Missense	3.31E-05		28.3	D	ProD	D	1
	16:919948	G/A	c.1351C>T	p.Arg451Trp	Missense	0.004243	0.00319489	24.9	D	ProD	T	3
	16:929680	G/A	c.787C>T	p.His263Tyr	Missense	2.48E-05		24.9	D	ProD	D	2
	16:943023	A/G	c.713T>C	p.Met238Thr	Missense			26.8	D	ProD	D	1
	16:1004615	CT/-	c.244_245delAG	p.Arg82Glyfs	Frameshift	2.48E-05		22.9				1
LPL	8:19796997	CA/-	c.46_47delCA	p.Gln16Glufs	Frameshift			35				2
	8:19805715	A/G	c.113A>G	p.Glu38Gly	Missense	8.24E-06		24.3	D	PoD	D	1
	8:19805736	C/A	c.134C>A	p.Thr45Asn	Missense	0.0001071		22.7	D	PoD	D	1
	8:19805756	G/C	c.154G>C	p.Asp52His	Missense			28.9	D	ProD	D	1
	8:19809302	G/A	c.272G>A	p.Trp91Ter	Nonsense			38			D	1
	8:19809316	G/C	c.286G>C	p.Val96Leu	Missense	0.0001235		28.2	D	ProD	D	1
	8:19811733	G/A	c.644G>A	p.Gly215Glu	Missense	0.0001318	0.000199681	22	T	ProD	D	9
	8:19811769	T/C	c.680T>C	p.Val227Ala	Missense	3.30E-05	0.000199681	27	D	PoD	D	1
	8:19811790	C/T	c.701C>T	p.Pro234Leu	Missense	4.94E-05		34	D	ProD	D	1
	8:19811806	T/A	c.717T>A	p.Phe239Leu	Missense			26.5	D	ProD	D	1
	8:19813405	G/A	c.829G>A	p.Asp277Asn	Missense	8.24E-06		32	D	B	D	1
	8:19813411	C/G	c.835C>G	p.Leu279Val	Missense	0.0001318		25.7	D	ProD	D	1
	8:19813474	G/C	c.898G>C	p.Gly300Arg	Missense			31	D	ProD	D	1
	8:19813594	G/A	c.1018G>A	p.Val340Ile	Missense			23.3	D	B	D	1
	8:19816785	G/A	c.1033G>A	p.Val345Ile	Missense	1.65E-05		18.9	T	B	D	1
	8:19816892	G/A	c.1139+1G>A		Splice donor	8.24E-06		26.5			D	1

Abbreviations: 1KG = 1000 Genomes Project; Alt = alternate; B = benign; CADD = Combined Annotation Dependent Depletion; CNV = copy-number variant; D = damaging; del = deletion; ExAC = Exome Aggregation Consortium; HGVS = Human Genome Variation Society; fs = frameshift; ins = insertion; PoD = possibly damaging; ProD = probably damaging; Ref = reference; SNV = single-nucleotide variant; T = tolerated.

### Appendix G. Summary of unique rare SNVs and CNVs identified in the UCSF cohort (N=312).

Variant information			RefSeq gene information		Minor allele frequencies			In silico predictions				Allele Counts
Gene	Chr:Pos	Ref/Alt	HGVS c.	HGVS p.	Sequence Ontology	ExAC	1KG	CADD PHRED	SIFT	PolyPhen2	Mutation Taster	
APOA5	11:116660857	A/T	c.1088T>A	p.Leu363Gln	Missense			10.58	D	B	T	1
	11:116661001	G/A	c.944C>T	p.Ala315Val	Missense	0.000626	0.0003993	28	D	ProD	T	7
	11:116661057	T/-	c.888delA	p.Ile296Metfs	Frameshift del			24.4				1
	11:116661062	G/A	c.883C>T	p.Gln295Ter	Nonsense	8.24E-06		37			D	1
	11:116661656	G/A	c.289C>T	p.Gln97Ter	Nonsense	8.24E-05		36			D	1
	11:116661734	G/-	c.211delC	p.Leu71Trpfs	Frameshift del			17.42				1
	11:116662386	C/-	c.77delG	p.Gly26Alafs	Frameshift del			24				1
APOC2	19:45451743	C/T	c.8C>T	p.Thr3Ile	Missense	0.0003377	0.0007987	13.79	D	B	T	1
	19:45451745	C/G	c.10C>G	p.Arg4Gly	Missense	9.88E-05		23.6	D	ProD	T	1
	19:45452024	A/C	c.122A>C	p.Lys41Thr	Missense	0.0008731	0.0003994	15	D	B	D	1
GPIHBP1	8:144297206	G/A	c.368G>A	p.Gly123Glu	Missense	0.0004616		17.35		B	T	1
	8:144297361	G/C	c.523G>C	p.Gly175Arg	Missense	0.0008799	0.0021965	14.89		PoD	T	1
LMF1	16:919894	C/T	c.1405G>A	p.Ala469Thr	Missense	0.0008543	0.0005990	29.7	D	ProD	D	2
	16:919908	C/T	c.1391G>A	p.Trp464Ter	Nonsense			39			D	1
	16:919948	G/A	c.1351C>T	p.Arg451Trp	Missense	0.004243	0.0031949	24.9	D	ProD	T	4
	16:919982	G/C	c.1317C>G	p.Tyr439Ter	Nonsense			36			D	2
	16:920817	C/T	c.1144G>A	p.Val382Met	Missense	6.61E-05	0.0001997	27.3	D	ProD	D	1
	16:921293	T/C	c.946A>G	p.Met316Val	Missense	8.28E-06		15.58	T	B	D	1
	16:929617	C/T	c.850G>A	p.Gly284Ser	Missense	2.48E-05		24.2	T	PoD	D	1
	16:943023	A/G	c.713T>C	p.Met238Thr	Missense			26.8	D	ProD	D	1
	16:943053	C/G	c.683G>C	p.Gly228Ala	Missense			24	D	PoD	D	1
	16:943053	C/T	c.683G>A	p.Gly228Glu	Missense	0.000198		28.8	D	PoD	D	1
	8:19796331-19806101 (approximate)				Large-scale deletion							1
	8:19805730	-/T	c.127_128insT	p.Arg44Lysfs	Frameshift ins			33				1
	8:19805815	C/G	c.213C>G	p.His71Gln	Missense	0.001128	0.0037939	12.52	T	B	D	1
	8:19809322	G/A	c.292G>A	p.Ala98Thr	Missense	0.0001647		33	D	ProD	D	1
	8:19811679	G/A	c.590G>A	p.Arg197His	Missense	4.12E-05		29.3	D	ProD	D	1
	8:19811710	C/G	c.621C>G	p.Asp207Glu	Missense			21.6	D	ProD	D	2
	8:19811711	G/A	c.622G>A	p.Val208Ile	Missense			28.4	D	ProD	D	1
	8:19811733	G/A	c.644G>A	p.Gly215Glu	Missense	0.0001318	0.0001997	22	T	ProD	D	8
	8:19811751	T/C	c.662T>C	p.Ile221Thr	Missense	8.24E-06		27.1	D	ProD	D	2
	8:19811844	T/C	c.755T>C	p.Ile252Thr	Missense	1.65E-05		24.8	D	PoD	D	1
	8:19811864	G/A	c.775G>A	p.Asp259Asn	Missense			11.9	T	B	D	1
	8:19813465	T/C	c.889T>C	p.Phe297Leu	Missense			29	D	PoD	D	1
	8:19813501	C/T	c.925C>T	p.Arg309Cys	Missense	1.65E-05		35	D	ProD	D	1
	8:19813528	A/G	c.952A>G	p.Asn318Asp	Missense			23.6	T	B	D	1
	8:19816785	G/A	c.1033G>A	p.Val345Ile	Missense	1.65E-05		18.9	T	B	D	1
	8:19816888	C/T	c.1136C>T	p.Thr379Ile	Missense	0.001911		20.2	D	B	T	1
	8:19818574	A/TTTT	c.1302delinsTTTT	p.Lys434delinsAsnPhe	Inframe ins			18.857				1

Abbreviations: 1KG = 1000 Genomes Project; Alt = alternate; B = benign; CADD = Combined Annotation Dependent Depletion; CNV = copy-number variant; D = damaging; del = deletion; ExAC = Exome Aggregation Consortium; HGVS = Human Genome Variation Society; fs = frameshift; ins = insertion; PoD = possibly damaging; ProD = probably damaging; Ref = reference; SNV = single-nucleotide variant; T = tolerated; UCSF = University of California, San Francisco.



### Appendix H. Summary of unique rare SNVs identified in the reference 1000 Genomes Project cohort (N=503).

Variant information			RefSeq gene information				Minor allele frequencies		In silico predictions				Allele Counts
Gene	Chr:Pos	Ref/Alt	rsID	HGVS c.	HGVS p.	Sequence Ontology	ExAC	1KG	CADD PHRED	SIFT	PolyPhen2	Mutation Taster	
<i>APOA5</i>	11:116660983	T/A	rs201201147	c.962A>T	p.His321Leu	Missense	0.001095	0.000599042	25.7	D	PoD	D	3
	11:116661335	G/A	rs546060544	c.610C>T	p.Arg204Cys	Missense	8.27E-06	0.000199681	33	D	PoD	D	1
<i>APOC2</i>	19:45452024	A/C	rs120074114	c.122A>C	p.Lys41Thr	Missense	0.0008731	0.000399361	15	D	B	D	2
<i>GPIHBP1</i>	8:144297142	C/T	rs200196582	c.304C>T	p.Leu102Phe	Missense		0.000199681	13.7		B	T	1
<i>LMF1</i>	16:904561	G/A	rs199544373	c.1675C>T	p.Arg559Cys	Missense	3.33E-05	0.000199681	25.1	D	ProD	D	1
	16:919883	C/G	rs200876477	c.1416G>C	p.Gln472His	Missense		0.000199681	23.3	T	PoD	D	1
	16:919894	C/T	rs181731943	c.1405G>A	p.Ala469Thr	Missense	0.0008543	0.000599042	29.7	D	ProD	D	2
	16:919948	G/A	rs138205062	c.1351C>T	p.Arg451Trp	Missense	0.004243	0.00319489	24.9	D	ProD	T	4
	16:920733	C/T	rs199713950	c.1228G>A	p.Gly410Arg	Missense	0.0008925	0.000599042	26.4	D	ProD	D	1
	16:929692	G/C	rs564167344	c.775C>G	p.Pro259Ala	Missense	8.28E-06	0.000199681	23.5	D	ProD	D	1
	16:1004447	G/C	rs200382562	c.413C>G	p.Ser138Cys	Missense	0.0004132	0.000399361	10.68	T	PoD	T	1
<i>LPL</i>	8:19805792	G/A	rs114101772	c.190G>A	p.Val64Met	Missense	2.47E-05	0.000199681	24.8	D	PoD	T	1
	8:19811631	G/A	rs191402029	c.542G>A	p.Gly181Asp	Missense	8.24E-06	0.000199681	27	D	ProD	D	1
	8:19819628	T/G	rs116403115	c.1325T>G	p.Val442Gly	Missense	0.0004036	0.000199681	25.6	D	PoD	D	1

Abbreviations: 1KG = 1000 Genomes Project; Alt = alternate; B = benign; CADD = Combined Annotation Dependent Depletion; CNV = copy-number variant; D = damaging; del = deletion; ExAC = Exome Aggregation Consortium; HGVS = Human Genome Variation Society; fs = frameshift; ins = insertion; PoD = possibly damaging; ProD = probably damaging; Ref = reference; SNV = single-nucleotide variant; T = tolerated.

**Appendix I. Screening primers for *LPL* CNVs.**

<b>CNV</b>	<b>Breakpoint</b>	<b>Primer direction</b>	<b>Primer sequence (5' to 3')</b>	<b>Annealing temperature (°C)</b>	<b>Primer labels in Figure 5.2</b>
5'UTR – exon 1	Upstream	F	TTGTAGGTTAGAGTGAACGTGCACAG	60	P2
		R	CATTATGCTGATGCTGCACAACTCTG	60	P3
	Downstream	F	TTCACACTTGATGGTCTCATTTCAGTGG	60	P4
		R	GATCAGACTGAATTGATTGGTCTGTTTCAG	60	P5
5' UTR – exon 2	Upstream	F	CTCTATTGGACGTGCTAATGGCACAG	60	P1
		R	CATTATGCTGATGCTGCACAACTCTG	60	P3
	Downstream	F	ACTGACATGCTGACATGCCAGATG	60	P6
		R	CATCTGTGTGAATTCTGTTAGTAGTAG	60	P7

The primers listed were designed to flank the two breakpoints for each CNV. The “Breakpoint” listed is relative to the deleted section of the gene. The sequence orientation for P1-P7 are relative to *LPL*. Highlighted primer sequences are the same. Abbreviations: CNV = copy-number variant; F = forward; R = reverse; UTR = untranslated region.

### Appendix J. Summary of unique rare SNVs identified in mild-to-moderate hypertriglyceridemia patients (N=134).

Gene	Variant Information					Minor allele frequencies			<i>In silico</i> predictions				Allele Counts
	Chr:Pos	Ref/Alt	HGVS c.	HGVS p.	Sequence Ontology	gnomAD	ExAC	1KG	CADD PHRED	SIFT	PolyPhen2	Mutation Taster	
<i>LPL</i>	8:19811733	G/A	c.644G>A	p.Gly215Glu	Missense	6.46E-05	0.0001318	0.0001997	23.3	T	ProD	D	2
	8:19811864	G/A	c.775G>A	p.Asp259Asn	Missense				22.9	T	B	D	1
	8:19813384 *	C/T	c.808C>T	p.Arg270Cys	Missense		1.65E-05		34	D	ProD	D	1
	8:19813481	G/T	c.905G>T	p.Cys302Phe	Missense				32	D	ProD	D	1
<i>APOA5</i>	11:116661001 *	G/A	c.944C>T	p.Ala315Val	Missense	0.000581546	0.000626	0.0003994	24.4	D	ProD	T	1
	11:116661392	C/A	c.553G>T	p.Gly185Cys	Missense	0.00531122	0.006132	0.0113818	22.7	D	ProD	T	2
	11:116661653	C/A	c.292G>T	p.Glu98Ter	Nonsense				36			D	1
	11:116661656	G/A	c.289C>T	p.Gln97Ter	Nonsense	6.46E-05	8.24E-05		37			D	1
<i>LMF1</i>	16:919894	C/T	c.1405G>A	p.Ala469Thr	Missense	0.000258415	0.0008543	0.0005990	24.8	D	ProD	D	1
	16:921323	C/T	c.916G>A	p.Gly306Arg	Missense		5.80E-05		27.5	D	ProD	D	1
	16:929650	T/G	c.817A>C	p.Ile273Leu	Missense		4.13E-05		10.72	T	B	D	1

“\*” indicates the variants that occur in the same patient. Abbreviations: 1KG = 1000 Genomes Project; Alt = alternate; B = benign; CADD = Combined Annotation Dependent Depletion; D = damaging; del = deletion; ExAC = Exome Aggregation Consortium; HGVS = Human Genome Variation Society; fs = frameshift; ins = insertion; PoD = possibly damaging; ProD = probably damaging; Ref = reference; SNVs = single-nucleotide variants; T = tolerated.

### Appendix K. Summary of unique rare LOF variants identified in multifactorial chylomicronemia patients (N=265).

Variant Info					MAF		in silico Prediction Tools					Total Allele Counts	
Chr: Position	Ref/Alt	Gene	HGVS c.	HGVS p.	Sequence Ontology	gnomAD All Pops	CADD	SIFT	Poly Phen2	Mutation Taster	ACMG	# Het	# Hom
2:21231489	C/A	<i>APOB</i>	NM_000384.2:c.8251G>T	NP_000375.2:p.Glu2751Ter	Nonsense		35			D	LP	1	0
2:27726415	C/T	<i>GCKR</i>	NM_001486.3:c.679C>T	NP_001477.2:p.Arg227Ter	Nonsense	0.0003185	35			D	LP	1	0
2:27730170	-/A	<i>GCKR</i>	NM_001486.3:c.1135dupA	NP_001477.2:p.Thr379Asnfs*36	Frameshift	0.0012737	22.8				LP	3	0
2:27745372	C/T	<i>GCKR</i>	NM_001486.3:c.1618C>T	NP_001477.2:p.Arg540Ter	Nonsense	0.0008602	35			D	LP	5	0
2:27746184	G/T	<i>GCKR</i>	NM_001486.3:c.1756G>T	NP_001477.2:p.Glu586Ter	Nonsense	0.0010832	35			D	VUS	1	0
2:44102404	G/A	<i>ABCG8</i>	NM_022437.2:c.1608G>A	NP_071882.1:p.Trp536Ter	Nonsense	3.18E-05	51			D	P	1	0
3:12458392	CTT GA/-	<i>PPARG</i>	NM_015869.4:c.1014_1018delC TTGA	NP_056953.2:p.Asp338Glu fs*25	Frameshift		35				P	1	0
3:12475485	C/-	<i>PPARG</i>	NM_015869.4:c.1361delC	NP_056953.2:p.Pro454Leu fs*14	Frameshift		35				P	1	0
4:100485266	-/GA	<i>MTTP</i>	NM_000253.3:c.-229_-228dupAG		Frameshift		14.44				LP	1	0
8:19795150-19806747		<i>LPL</i>	NM_001715.2:c.1_249del	NP_006184.2:p.Met1?	CNV del							1	0
8:19796997	CA/-	<i>LPL</i>	NM_000237.2:c.46_47delCA	NP_000228.1:p.Gln16Glu fs*24	Frameshift		34				LP	1	0
8:19805730	-/T	<i>LPL</i>	NM_000237.2:c.128dupT	NP_000228.1:p.Arg44Lys fs*4	Frameshift		28.1				LP	1	0
10:91005432	C/T	<i>LIPA</i>	NM_000235.3:c.229+1G>A		Splice Donor		34			D	LP	1	0
10:90987706-90988405		<i>LIPA</i>	Deletion encompassing: Exon 4		CNV del							1	0
11:62458267	G/A	<i>BSCL2</i>	NM_032667.6:c.953C>T	NP_001124174.2:p.Gln271Ter	Nonsense	9.56E-05	17.53	T	B	T	LP	1	0
11:116661062	G/A	<i>APOA5</i>	NM_052968.4:c.883C>T	NP_443200.2:p.Gln295Ter	Nonsense	3.18E-05	39			S	LP	1	0
11:116661734	G/-	<i>APOA5</i>	NM_052968.4:c.211delC	NP_443200.2:p.Leu71Trp fs*4	Frameshift		22.4				LP	1	0
11:116701354	G/A	<i>APOC3</i>	NM_000040.1:c.55+1G>A		Splice Donor	0.0013398	31			D	P	3	0
12:121432117	-/C	<i>HNF1A</i>	NM_000545.6:c.863_864insC	NP_000536.5:p.Pro289Ala fs*27	Frameshift		26.2				LP	1	0
15:58723924-58724569		<i>LIPC</i>	Deletion encompassing: Exon 1		CNV del							2	0
15:58838104	-/CG	<i>LIPC</i>	NM_000236.2:c.738_739dupCG	NP_000227.2:p.Gly247Ala fs*12	Frameshift	0.0001912	35				VUS	1	0
15:90213298	G/A	<i>PLIN1</i>	NM_002666.4:c.511C>T	NP_002657.3:p.Arg171Ter	Nonsense		36			D	LP	1	0
19:4168357	C/T	<i>CREB3L3</i>	NM_032607.2:c.724C>T	NP_115996.1:p.Arg242Ter	Nonsense	9.57E-05	39			D	VUS	1	0
19:4168365	-/G	<i>CREB3L3</i>	NM_032607.2:c.732dupG	NP_115996.1:p.Lys245Glu fs*130	Frameshift	0.0003190	35				VUS	3	0
19:4153347-4155274		<i>CREB3L3</i>	Deletion encompassing: 5'UTR to exon 2		CNV del							1	0
19:11241706-11244755		<i>LDLR</i>	Deletion encompassing: Exon 18 to the 3'UTR		CNV del							1	0

Abbreviations: Alt = alternate; B = benign; CADD = Combined Annotation Dependent Depletion; Chr = chromosome; CNV = copy-number variant; D = damaging; del = deletion; ExAC = Exome Aggregation Consortium; HGVS = Human Genome Variation Society; het = heterozygous; fs = frameshift; ins = insertion; LOF = loss-of-function; LP = likely pathogenic; MAF = minor allele frequency; P = pathogenic; pops = populations; Ref = reference; SNVs = single-nucleotide variants; T = tolerated; VUS = variant of uncertain significant.

**Appendix L. Exautomate: A user-friendly tool for region-based rare variant association analysis**

**Exautomate: A user-friendly tool for region-based rare variant association analysis (RVAA)**

Brent D. Davis<sup>1\*</sup>, Jacqueline S. Dron<sup>2,3\*</sup>, John F. Robinson<sup>2</sup>, Robert A. Hegele<sup>2,3,4</sup>, Dan J. Lizotte<sup>1,5</sup>

<sup>1</sup> Department of Computer Science, Western University, London, Canada

<sup>2</sup> Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Canada

<sup>3</sup> Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Canada

<sup>4</sup> Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Canada

<sup>5</sup> Department of Epidemiology & Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, Canada

\* Authors contributed equally to this work.

### **Abstract**

Region-based rare variant association analysis (RVAA) is a popular method to study rare genetic variation in large datasets, especially in the context of complex traits and diseases. Although this method shows great promise in increasing our understanding of the genetic architecture of complex phenotypes, performing a region-based RVAA can be challenging. The sequence kernel association test (SKAT) can be used to perform this analysis, but its inputs and modifiable parameters can be extremely overwhelming and may lead to results that are difficult to reproduce. We have developed a software package called "Exautomate" that contains the tools necessary to run a region-based RVAA using SKAT and is easy-to-use for any researcher, regardless of their previous bioinformatic experiences. In this report, we discuss the utilities of Exautomate and provide detailed examples of implementing our package. Importantly, we demonstrate a proof-of-principle analysis using a previously studied cohort of 313 familial hypercholesterolemia (FH) patients. Our results show an increased burden of rare variants in genes known to cause FH, thereby demonstrating a successful region-based RVAA using Exautomate. With our easy-to-use package, we hope researchers will be able to perform reproducible region-based RVAA to further our collective understanding behind the genetics of complex traits and diseases.

## Introduction

Understanding the genetic architecture of complex traits and diseases is an active area of research, driven largely by massive next-generation sequencing (NGS) efforts and the availability of public repositories containing genotypic and phenotypic data on hundreds of thousands of individuals<sup>1-3</sup>. While the previous obstacle in genetics research was acquiring such datasets, we are now faced with the challenge of figuring out how to effectively and appropriately study these data<sup>4,5</sup>.

Bioinformatic analyses are necessary to help identify unknown genetic determinants and explore these datasets. Rare variant association analysis (RVAA) is an increasingly popular approach to study rare variants in this context<sup>6-10</sup>; however, their frequency in the population makes it difficult to attain large enough sample sizes to detect significant relationships between variants and disease<sup>11-13</sup>. As such, researchers often perform region-based RVAA by grouping or collapsing rare variants together—typically by gene—to increase statistical power for association testing<sup>12-14</sup>.

One method frequently used for region-based RVAA is the sequence kernel association test (SKAT)<sup>15-18</sup>. While the method has been useful in revealing rare variants and genomic loci of interest in complex traits and diseases—such as cardiovascular disease, body-mass index, height, and neurodegenerative diseases<sup>19-22</sup>—reproducing these results can be difficult. SKAT is challenging to implement for exome- and genome-scale analyses as it requires significant data preprocessing involving additional software dependencies and variables that can complicate reproducibility<sup>23</sup>. On top of that, few published studies report the exact preprocessing steps and SKAT parameters used in their analysis. With an increase in the availability of genetic and phenotypic data, there has been a surge in exploratory analyses of the large NGS datasets, making SKAT a very popular tool. To assist with transparency in research and encouraging reproducibility of results, easily accessible and user-friendly bioinformatic tools are necessary.

We have created an open-access, modular script package built from pre-existing tools to automate data handling, processing, and perform a region-based RVAA using SKAT. As a proof of principle, we utilized publicly available data from the 1000 Genomes Project and data from a well-characterized lipid disorder in which the disease-causing genes are known, to test our script package. We also outline precautions needed when performing a region-based RVAA and adjusting SKAT parameters.

Our “Exautomate” package is user-friendly, designed with genetic researchers in mind, and generates a detailed methods-log to be utilized in publishing efforts. Our automated



analysis package is our standardization attempt to ensure consistent, reproducible region-based RVAA results, and to further the validity of this modern, exploratory genetic method.

## Methods & Recommended Script Usage

### Operating system compatibility

Our packages have been tested on the following platforms: 1) Windows Subsystem for Linux (WSL, Bash on Windows) I; 2) Ubuntu 16.04 and 18.04; and 3) Mac OS X 10.13.1.

### Software dependencies and installation

The Exautomate package is presented as a series of bash scripts that can be found on Github (<https://github.com/exautomate/Exautomate-Core>). An installer script [`Installer.sh` or `mac-installer.sh`] is available to download and install the following dependencies:

bedtools<sup>24</sup>, BWA<sup>25</sup>, Genome Analysis Toolkit (GATK) including Picard<sup>26</sup>, SAMtools<sup>27</sup>, tabix<sup>28</sup>, and VCFtools<sup>29</sup>. The following programming languages are required: C, Java, Perl, Python, and R.

These dependencies are installed using a combination of browserless downloads, `wget`, and the `apt-get` package manager. The included Mac OS X installer requires `homebrew`. When using R, the R-specific libraries of SKAT<sup>15</sup>, `ggplot2`<sup>30</sup>, and `reshape2`<sup>31</sup> are installed inside `RunSkat.R` if they have not been installed previously. At the time of this publication, all packages listed are open-source and freely available. The authors of PLINK<sup>32</sup> and ANNOVAR<sup>33</sup> would prefer that their tools be registered before being downloaded and are therefore not included in our installers. Details on how to download ANNOVAR and PLINK for use by Exautomate can be found in the **Supplemental Materials (Methods)**.

### Performing the region-based RVAA

Options '1' and '2' from the Exautomate main menu allow the user to: 1) perform SKAT or SKAT-O on a pre-merged `.vcf` file containing variant data on both controls and cases, or 2) merge a `.vcf` file containing variant data on controls with a `.vcf` file containing variant data on cases, from which the resultant `.vcf` file is used for SKAT or SKAT-O. Exautomate is set up such that the user inputs the necessary pieces of information (i.e. number of cases vs. controls, file names, kernel option, SKAT method) at the start of the workflow. There are two subsequent instances where the user will need to interact with the terminal before script completion. First, the user must encode the newly generated `.fam` file with control/case information by assigning the "phenotype" column variable as 1 (unaffected) or 2 (affected). Second, the user must modify the newly generated `.SetID` file; the "sets" in the first column of the file cannot be greater than 50

characters. Exautomate is set up under the assumption that the user wishes to group variants into gene sets, and therefore generates a gene-based .SetID file from ANNOVAR output. This step in the workflow is where the user may alter their sets as required. Considerations for preparing .SetID files can be found in the **Supplemental Materials (Additional Information)**.

#### Retrieval of 1000 Genomes data

Option '3' from the Exautomate main menu allows the user to download data from the 1000 Genomes FTP site. There are options to filter the downloaded data based on ethnicity and genomic sites of interest, as specified by a .bed file. An example of using this option can be found in **Supplemental Materials (Methods)**. An important note on this option is that it does not download information related to the sex chromosomes or mitochondrial DNA, and it does not take relatedness of the 1000 Genomes participants into account.

#### Creating a synthetic dataset

Option '4' from the Exautomate main menu allows the user to perform SKAT or SKAT-O analysis with a synthetic dataset in the form of a .sim file, generated using PLINK. This option was largely used to test the functionality of the Exautomate package and has been included as an extra feature for users.

#### Proof-of-principle demonstration analysis

The European subset of the 1000 Genomes cohort (N=503) was retrieved using option '3' of Exautomate and was filtered to contain sites covered by the LipidSeq targeted NGS panel<sup>34</sup>. Patients diagnosed with familial hypercholesterolemia (FH) (N=313), which is defined as having severely elevated low-density lipoprotein (LDL) cholesterol levels, were used as a case cohort. Since these patients have been genetically diagnosed in a previous study<sup>35</sup> and the FH phenotype itself is well characterized, we have an a priori expectation of what the analysis should reveal, making it an ideal cohort for a proof of principle.

To highlight usage of the Exautomate package, we document each step of the workflow—from installation to output—for our proof-of-principle analysis using data from the 1000 Genomes and the FH patient cohort in the **Supplemental Materials (Methods)**. We strongly encourage users to consult this document first for a complete understanding of the tool, as we outline important considerations for each stage of the workflow. We also provide a general overview in **Figure 1** regarding the flow of information between file types.

## Results

### Region-based RVAA on FH patient cohort

The .vcf files from all FH patients and 1000 Genomes controls were merged together and filtered on the following parameters: biallelic sites, minor allele frequency  $\leq 1\%$  (based on the gnomAD database<sup>36</sup>), sequence ontology (insertion, deletion, missense, splice acceptor, splice donor, nonsense), and *in silico* predictions (CADD Phred<sup>37,38</sup>  $\geq 10$ ). This filtered .vcf was used as the input .vcf for option '1' of Exautomate. We selected options for a linear weighted, gene-based SKAT-O analysis and a Bonferroni adjustment of P-values. From start to finish—including manually editing the .fam and .SetID files—Exautomate ran for 5 minutes and 13 seconds on an 8-cores @ 2.33GHz, 64GB RAM Ubuntu 18.04 Server machine.

The gene sets and their adjusted P-values are detailed in **Table S1**, while the distribution of adjusted P-values is illustrated in **Figure 2**. Overall, 19 genes met statistical significance when using an  $\alpha$ -threshold of 0.05 (**Table 1**). Of importance, neither the SKAT nor SKAT-O analysis indicate which study group (i.e. the cases or the controls) has the increased burden of rare variants driving the statistical association; therefore, additional downstream analysis is required to determine if the increase of rare variants is specific to the cases or controls.

## Discussion

Through the development and implementation of our open-access, user-friendly Exautomate package, we have made possible the ability to conduct a region-based RVAA following a reliable, reproducible, and transparent method. As a proof-of-principle, we utilized Exautomate to perform a region-based RVAA on a previously studied disease cohort of 313 FH patients with an a priori understanding of the genetic factors causing their phenotypes<sup>39</sup>. As one of the most commonly inherited types of metabolic disease, the molecular basis and mechanisms leading to FH are extremely well characterized<sup>39</sup>.

After performing optimally adjusted SKAT-O, our resultant output suggested a reliable analysis. The gene with the most significant prevalence of rare variants in cases compared to controls was *LDLR*, encoding the LDL receptor (LDLR). The LDLR is the primary receptor responsible for the removal of LDL particles from the blood; an extreme accumulation of LDL particles leads to an extreme elevation in LDL cholesterol levels, which is the main phenotypic feature of FH<sup>40</sup>. Given that mutations in *LDLR* account for >90% of FH cases<sup>41</sup>—with over 2000 mutations reported to cause FH<sup>39</sup>—it is unsurprising that our analysis revealed *LDLR* to have the greatest prevalence of rare variants. It should be noted that in the previously described FH



study, 105 unique *LDLR* variants were reported to explain the FH phenotype of 53.7% of patients, while our region-based RVAA only utilized 70 variants for analysis. This difference is because our analysis only considers single-nucleotide variants, and if there is a single missing allele call at any position, that entire genomic coordinate will be excluded from analysis. These are points that should be considered when running Exautomate on any dataset.

Two other genes known to cause ~8% and ~2% of FH cases include *APOB* and *PCSK9*, encoding apolipoprotein (apo) B and proprotein convertase subtilisin/kexin type 9 (PCSK9) respectively<sup>41</sup>; both genes were present on our list of significant results. Apo B is the main protein constituent of LDL particles and serves as the primary ligand for LDLR binding, allowing for the clearance of LDL particles from circulation<sup>42</sup>. Disruptions to the LDLR-binding site causes disruptions in the uptake of LDL, leading to elevations in circulating levels of LDL cholesterol<sup>39</sup>. Conversely, PCSK9 is a circulating protein that directly interacts with LDLR. When bound to LDLR, PCSK9 prevents the cell-surface recycling of the receptor following its internalization, and instead targets it for lysosomal degradation<sup>43</sup>. Rare gain-of-function variants in PCSK9 direct more receptors towards degradation<sup>44</sup>; fewer available LDLRs leads to a decrease in LDL particle clearance and an increase in LDL cholesterol levels.

Understanding and correctly interpreting our results required an intimate understanding of our study cohorts and NGS panel. For example, we were left to consider why *APOB* and *PCSK9* did not rank as the second and third most significant hits from our RVAA, respectively. One of the biggest considerations of this proof-of-principle analysis was that each cohort was sequenced using different methods. Regarding our second most significant gene output, *CEL*, we observed that the majority of rare variants appeared in our FH cohort. Some individuals might interpret this to mean that *CEL* is related to FH; however, through prolonged use and familiarity of our LipidSeq panel, we know that *CEL* is often met with sequencing artifacts due to a neighboring pseudogene, which we have observed to harbour a large number of variants<sup>45</sup>. Since our control cohort was sequenced using a different method, these artifacts do not appear in the control dataset, explaining the apparently artefactual statistical association of *CEL* with FH in this analysis. Had our control dataset been sequenced with LipidSeq, we anticipate this would have corrected the issue. It may be the case that a few of our significant gene hits are false positives due to this cohort-sequencing bias. When applying a region-based RVAA to any dataset of interest, it is imperative to understand possible differences in sequencing methods, be familiar with the nature of genes, and consider inherent characteristics of both case and control cohorts—this will help in remaining mindful of the high false positive rate and will assist with correctly interpreting results.

We recommend performing a region-based RVAA using our Exautomate package as an early-stage analysis for exploratory or observational purposes. Given the considerations discussed above, and reports of this analysis having a higher false positive rate and other limitations<sup>13,46,47</sup>, potentially interesting results should be followed up with more stringent approaches including segregation analyses and functional studies. An attractive application for a region-based RVAA may be to serve as a guide for gene exploration: if the objective is to identify a previously unreported gene related to a disease or phenotype of interest, the significant output may be used as a starting point to help narrow the focus on potential genes of interest. This could be particularly helpful when dealing with exome data, with over 20,000 possible genes to consider.

The potential for variable results is large, given the number of parameters that could be adjusted prior to performing a region-based RVAA. When developing Exautomate, we took into consideration the importance of reproducible and transparent results, so we created a detailed method log output to assist in reporting. Particularly with SKAT and SKAT-O, even slight parameter adjustments at the outset can produce significant differences in output. In a substantial search of published articles reporting the use of SKAT or SKAT-O for different phenotypes—including cardiovascular disease, body-mass index, height, amyotrophic lateral sclerosis, red blood cell traits, Alzheimer's disease, Parkinson's disease, lipid traits, and blood pressure—there are virtually no statements or reporting of the specific parameters used, other than perhaps the variant frequency threshold<sup>19-22,48-54</sup>. Unfortunately, this trend of minimal methodological information neither instills confidence in the research, nor does it facilitate replications of results. As a powerful analysis, researchers using SKAT-related tools must provide the appropriate information in published works; one might argue that these methodological details are more important than the results themselves.

While the possibility for false positives remains and efforts to explore potential biologically relevant results requires careful consideration and subsequent analysis, a region-based RVAA still remains an attractive method to set the stage for in-depth studies of rare variants influencing complex phenotypes. In order to successfully perform these analyses, researchers must have easily accessible tools that support the idea of transparency and reproducibility in research. With our Exautomate package and its implementation of SKAT and SKAT-O, we hope that researchers will utilize this tool to assist in their efforts to publish well-documented methods, correctly interpret results, and make new discoveries that will continue to add to our growing understanding of the genetic architecture underlying complex traits and disease.

bioRxiv preprint first posted online May 24, 2019; doi: <http://dx.doi.org/10.1101/649368>. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

9

### **Acknowledgements**

We would like to thank the individuals that independently tested the Exautomate package and provided feedback on areas requiring improvement: Allison Dilliot, Arden Lawson, and Julieta Lazarte.

## References

1. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
2. Hebring, S. Genomic and Phenomic Research in the 21st Century. *Trends Genet* **35**, 29-41 (2019).
3. Hirata, M. *et al.* Overview of BioBank Japan follow-up data in 32 diseases. *J Epidemiol* **27**, S22-S28 (2017).
4. Schmidt, B. & Hildebrandt, A. Next-generation sequencing: big data meets high performance computing. *Drug Discov Today* **22**, 712-717 (2017).
5. Bien, S.A. *et al.* The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE. *Annu Rev Genomics Hum Genet* (2019).
6. Cohen, J.C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72 (2004).
7. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* **37**, 161-5 (2005).
8. Cohen, J.C., Boerwinkle, E., Mosley, T.H., Jr. & Hobbs, H.H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* **354**, 1264-72 (2006).
9. Romeo, S. *et al.* Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet* **39**, 513-6 (2007).
10. Ji, W. *et al.* Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* **40**, 592-599 (2008).
11. Michailidou, K. Meta-Analysis of Common and Rare Variants. *Methods Mol Biol* **1793**, 73-88 (2018).
12. Auer, P.L. & Lettre, G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* **7**, 16 (2015).
13. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014).
14. Ionita-Laza, I., Cho, M.H., Laird, N.M. Statistical Challenges in Sequence-Based Association Studies with Population- and Family-Based Design. *Statistics in Biosciences* **5**, 54-70 (2013).
15. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
16. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* **91**, 224-37 (2012).
17. Derkach, A., Lawless, J.F. & Sun, L. Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. *Genet Epidemiol* **37**, 110-21 (2013).
18. Hsieh, A.R. *et al.* A non-threshold region-specific method for detecting rare variants in complex diseases. *PLoS One* **12**, e0188566 (2017).
19. Pirim, D. *et al.* Resequencing of LPL in African Blacks and associations with lipoprotein-lipid levels. *Eur J Hum Genet* **23**, 1244-53 (2015).
20. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
21. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190 (2017).

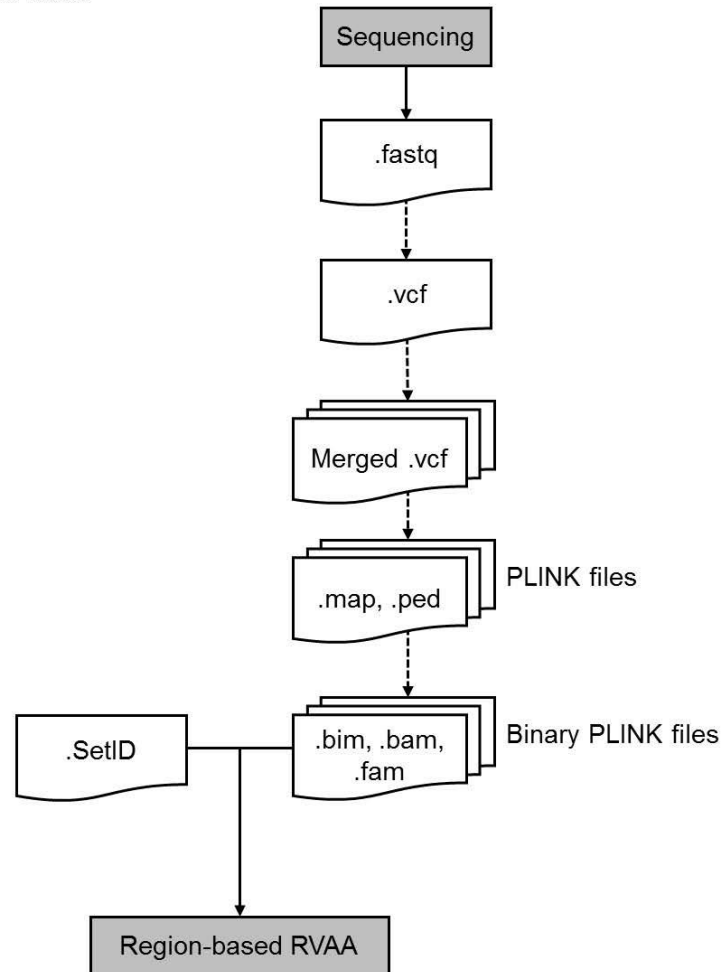


22. Nho, K. *et al.* Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease. *BMC Med Genomics* **9** Suppl 1, 30 (2016).
23. Mangul, S. *et al.* Challenges and recommendations to improve installability and archival stability of omics computational tools. *bioRxiv* (2019).
24. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
26. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
27. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-93 (2011).
28. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718-9 (2011).
29. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
30. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* (2016).
31. Wickham, H. Reshaping Data with the reshape Package. *Journal of Statistical Software* **21**, 1-20 (2007).
32. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
33. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
34. Johansen, C.T. *et al.* LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *J Lipid Res* **55**, 765-72 (2014).
35. Wang, J. *et al.* Polygenic Versus Monogenic Causes of Hypercholesterolemia Ascertained Clinically. *Arterioscler Thromb Vasc Biol* **36**, 2439-2445 (2016).
36. Karczewski, K.J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210 (2019).
37. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
38. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
39. Defesche, J.C. *et al.* Familial hypercholesterolaemia. *Nat Rev Dis Primers* **3**, 17093 (2017).
40. Berberich, A.J. & Hegele, R.A. The complex molecular genetics of familial hypercholesterolaemia. *Nat Rev Cardiol* **16**, 9-20 (2019).
41. Iacocca, M.A. *et al.* ClinVar database of global familial hypercholesterolemia-associated DNA variants. *Hum Mutat* **39**, 1631-1640 (2018).
42. Hegele, R.A. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet* **10**, 109-21 (2009).
43. Burke, A.C., Dron, J.S., Hegele, R.A. & Huff, M.W. PCSK9: Regulation and Target for Drug Development for Dyslipidemia. *Annu Rev Pharmacol Toxicol* **57**, 223-244 (2017).
44. Dron, J.S. & Hegele, R.A. Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Curr Opin Lipidol* **28**, 161-169 (2017).
45. Fjeld, K. *et al.* A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. *Nat Genet* **47**, 518-522 (2015).



46. Jeng, X.J., Daye, Z.J., Lu, W. & Tzeng, J.Y. Rare Variants Association Analysis in Large-Scale Sequencing Studies at the Single Locus Level. *PLoS Comput Biol* **12**, e1004993 (2016).
47. Moutsianas, L. *et al.* The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet* **11**, e1005165 (2015).
48. Gaastra, B. *et al.* Rare genetic variation in UNC13A may modify survival in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener* **17**, 593-599 (2016).
49. Chami, N. *et al.* Exome Genotyping Identifies Pleiotropic Variants Associated with Red Blood Cell Traits. *Am J Hum Genet* **99**, 8-21 (2016).
50. Nettiksimmons, J., Tranah, G., Evans, D.S., Yokoyama, J.S. & Yaffe, K. Gene-based aggregate SNP associations between candidate AD genes and cognitive decline. *Age (Dordr)* **38**, 41 (2016).
51. Tran, N.T. *et al.* PCSK9 variation and association with blood pressure in African Americans: preliminary findings from the HyperGEN and REGARDS studies. *Front Genet* **6**, 136 (2015).
52. Lin, Y.C. *et al.* Identifying rare and common disease associated variants in genomic data using Parkinson's disease as a model. *J Biomed Sci* **21**, 88 (2014).
53. Pirim, D. *et al.* Lipoprotein lipase gene sequencing and plasma lipid profile. *J Lipid Res* **55**, 85-93 (2014).
54. Pirim, D. *et al.* Resequencing of the CETP gene in American whites and African blacks: Association of rare and common variants with HDL-cholesterol levels. *Metabolism* **65**, 36-47 (2016).

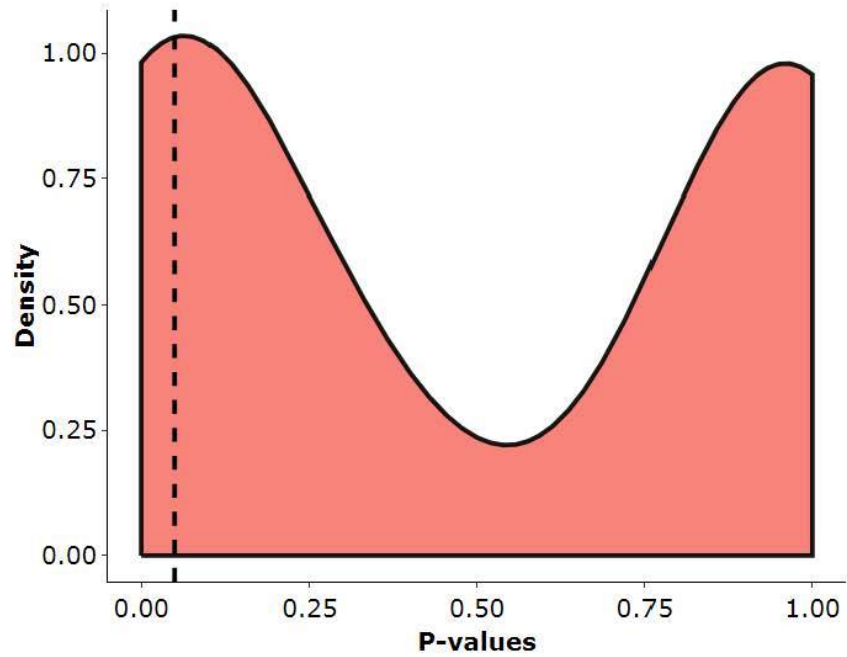
### Figures and Tables



**Figure 1.** The flow of information between file formats from sequencing to a region-based RVAA. Our automation pipeline aims to reduce the need for user input and minimize potential sources of variability. Abbreviations: RVAA = rare variant association analysis.

bioRxiv preprint first posted online May 24, 2019; doi: <http://dx.doi.org/10.1101/649368>. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

14



**Figure 2.** Kernel density plot of SKAT-O output, after performing a region-based RVAA on FH patients (N=313) vs. controls from the 1000 Genomes cohort (N=503). The dashed line represents a P-value of 0.05. P-values have been adjusted using the Bonferroni correction method.

**Table 1.** Output from proof-of-principle SKAT-O analysis demonstrating genes with a significant burden of rare variants in one study cohort compared to the other.

Gene group	Adjusted P-value	Number of variants used in test
<i>LDLR</i> *	1.32E-52	70
<i>CEL</i>	1.04E-31	23
<i>APOB</i> *	6.66E-13	19
<i>CYP2D6</i>	2.84E-07	20
<i>ABCA1</i>	3.24E-06	13
<i>ABCG8</i>	0.000582935	9
<i>PYGM</i>	0.001242416	10
<i>LIPE</i>	0.002173174	9
<i>CPT2</i>	0.002959799	9
<i>ABCG5</i>	0.003864299	9
<i>LMNB2</i>	0.006346106	3
<i>POLD1</i>	0.008366644	5
<i>LMF1</i>	0.011220071	7
<i>NPC1L1</i>	0.011858161	8
<i>PCSK9</i> *	0.015456462	7
<i>SORT1</i>	0.023792097	6
<i>ABCC8</i>	0.030397477	7
<i>HNF4A</i>	0.031537345	4
<i>PNPLA2</i>	0.040399011	5

P-values have been adjusted using the Bonferroni correction method. "\*" indicate genes known to cause autosomal dominant FH.

## **Supplemental Materials**

### **Dependencies, Future Proofing, and Open-Source Software**

Much of modern research software development is supported by the open-source community. More often, research efforts seem to require cutting-edge technology, such as machine-learning approaches and advanced, computationally intensive statistical methods. Programming languages that support rapid software updates and the speed behind novel tool development are key to the fast-paced research environment; however, languages come in and out of vogue. While Python and R are currently popular in the research community—our Exautomate packages utilizes tools requiring both languages—new trends may lead to drastic changes in dependencies, which leaves a constant issue of not always being able to ensure consistent or reproducible methods.

We want to further emphasize that our pipeline has been built entirely on open-source software (although some of it requires registration). Eventually, the software we rely on may undergo a combination of interface changes, functionality changes, abandonment issues or other software daemons that are going to result in a broken pipeline. To prepare for this, we strongly recommend that users of Exautomate investigate the possibility of creating code containers, such as Docker or Singularity, which support their code in a reproducible, maintained environment. As long as the containers produced by Docker or Singularity are available, users will be able to run the software contained therein.

### **Issues, Bugs, and Feature Requests**

We ask that all users who encounter issues, bugs or desire new features added to the Exautomate core, post their issues or requests at <https://github.com/exautomate/Exautomate-Core/issues>.



```
Welcome to Exautomate.

Main Menu:

1: Pre-merged .vcf for analysis.
2: Merge case and control .vcf for analysis.
3: Retrieve 1000 Genomes, no analysis.
4: Synthetic run.
5: Exit.

Please select an option to run (1-5): 3

##### OPTION 3: 1000 Genomes Utility Suite #####

Option for input .bed files:
../input/lipidseq.bed  ../input/exome.bed

Enter the name of the .bed file to filter by: ../input/lipidseq.bed

Ethnicities in the 1000 Genomes cohort:
EUR (includes: CEU, FIN, GBR, IBS, TSI)
EAS (includes: CDX, CHB, CHS, JPT, KHV)
AMR (includes: CLM, MXL, PEL, PUR)
SAS (includes: BEB, GIH, ITU, PJL, STU)
AFR (includes: ACB, ASW, ESN, GWD, LWK, MSL, YRI)
CUSTOM (user-specified file, must be named 'custom.txt' in the src directory)
ALL (the entire 1000 Genomes dataset)

Please select which population group (3-letter code only, ALL, or CUSTOM) you'd
like to download from the 1000 Genomes database: EUR

# The download and processing for the 1000 Genomes can take some time. As
Exautomate continues, it will occasionally print notices of how far along it is.

Finished 1000 Genomes retrieval.

Finished concatenation of 1000 Genomes files.

Filtering by ethnicity on 1000 Genomes files.

Finished filtering 1000 Genomes file. Ensure that your final 1000 Genomes .vcf file
of interest is in the output directory.

# Exautomate requires two additional pieces of user input. They are related to
whether the user wishes to keep the original 1000 Genomes files or any intermediate
files. If the user chooses to keep the original 1000 Genomes files, then the user
will not have to re-download them in the future.
```



```
Delete original 1000 Genomes files? (y/n): n
Delete bed filtered chromosome files? (y/n): y
```

#### Example: Proof-of-principle demonstration analysis

For this proof-of-principle analysis, all scripts were run using Ubuntu 18.04 unless otherwise specified. We specifically used 'Option 1' from the Exautomate menu, which required us to have already generated a case/control merged .vcf file.

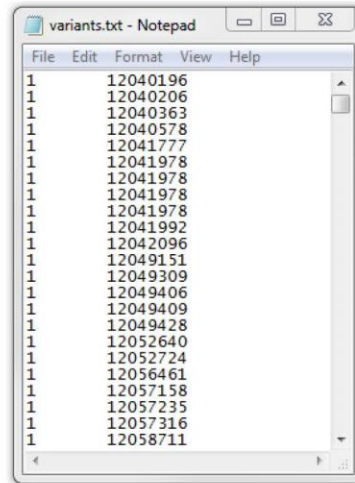
- 1) **Gathering .vcf files** → Gather all of the .vcf files needed for this study. For organizational purposes, we recommend having a folder for case .vcf files, a folder for control .vcf files, and a folder for case and control .vcf files together.
- 2) **Merging .vcf files** → Use your favourite method of merging .vcf files. After merging, ensure that the cases and controls are grouped together within the merged .vcf file (i.e. All of the cases appear first, followed by all of the controls. Or, all of the controls appear first, followed by all of the cases).

```
# We used a customize, in-house script that utilizes vcf-merge (part of VCFtools)
for merging.
> ./merge_vcfs.sh

# The output from this script is 'merged.vcf'.
```



- 3) **Selecting variants of interest** → This step may be dependent on the tools and resources available to the user. Ensure you have a tab-delimited .txt file with one variant of interest per line in the format of: 'chromosome number<tab>scaffold position'. This file does not require a header. An example of what this file should look like can be found in **Figure A**.



**Figure A.** In `variants.txt`, the first column contains the chromosome number, while the second column contains the scaffold position. Each line is a unique variant of interest.

- 4) **Pre-processing the merged .vcf file** → This step can be tailored towards the .vcf file(s) in use, as it will largely depend on how the .vcf files were generated. Below is an example of a minimal number of pre-processing steps.

```
# To select for rare variants with our desired sequence ontology (see main
manuscript), we utilized --positions to select only the variant positions of
interest to be output into a filtered down .vcf file.
> vcftools --vcf merged.vcf --positions variants.txt --recode --out merged_filtered

# Any other steps needed to fix the final .vcf file. For us, we needed to fix the
sample header names. Other users may require different fixes.
> vcf-sort -c merged_filtered.recode.vcf > merged-case_control.vcf
> dos2unix merged-case_control.vcf
```

- 5) **Running Exautomate** → Once the .vcf file has been properly modified, it must be placed in `/input/`.



```
Choose SKAT or SKAT-0: SKAT-0

Multiple comparisons options: holm, hochberg, hommel, bonferroni, BH, BY, fdr, none
Enter the multiple comparison option to be used in the analysis: bonferroni

# At this point, Exautomate begins its own processing of the input .vcf file.
Eventually, the user will be prompted to edit the .fam file that has been generated
in /output/. The only column in the .fam file that needs adjusting is column F.
Stop and edit the .fam file (must be the same name as what was entered at the
beginning + .adj.fam). Finished? (y/n): y
# From this step, the user should have created a file called "FH_1000G.adj.fam".

# Shortly after, the user will be prompted a final time to edit the .SetID file
that has been generated in /output/. There are a few considerations for modifying
.SetID files that we have detailed in this document under "Additional Information".
We encourage the user to consult this information before continuing on through
Exautomate.
Stop and edit the .SetID file (must be the same name as what was entered at the
beginning + .adj.SetID). Finished? (y/n): y
# From this step, the user should have created a file called "FH_1000G.adj.SetID".
```

- 6) **Finishing Exautomate** → All intermediate and final files, along with the EXAUTOMATEmethods.log, will be in /output/. We recommend viewing the .log file using Microsoft Excel or a similar program for readability.

**Figures and Tables**

**Table S1.** Output from proof-of-principle SKAT-O analysis.

Gene group	Adjusted P-value	Number of variants used in test
<i>LDLR</i> *	1.32E-52	70
<i>CEL</i>	1.04E-31	23
<i>APOB</i> *	6.66E-13	19
<i>CYP2D6</i>	2.84E-07	20
<i>ABCA1</i>	3.24E-06	13
<i>ABCG8</i>	0.000582935	9
<i>PYGM</i>	0.001242416	10
<i>LIPE</i>	0.002173174	9
<i>CPT2</i>	0.002959799	9
<i>ABCG5</i>	0.003864299	9
<i>LMNB2</i>	0.006346106	3
<i>POLD1</i>	0.008366644	5
<i>LMF1</i>	0.011220071	7
<i>NPC1L1</i>	0.011858161	8
<i>PCSK9</i> *	0.015456462	7
<i>SORT1</i>	0.023792097	6
<i>ABCC8</i>	0.030397477	7
<i>HNF4A</i>	0.031537345	4
<i>PNPLA2</i>	0.040399011	5
<i>MLXIPL</i>	0.057246197	5
<i>PLIN1</i>	0.073372526	5
<i>LPIN1</i>	0.094274082	5
<i>BSCL2</i>	0.094407912	5
<i>APOA4</i>	0.09445813	5

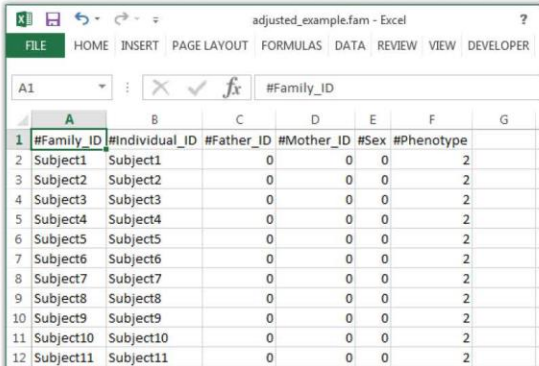
<i>CIDEA</i>	0.123915735	2
<i>WRN</i>	0.127442247	6
<i>AMPD1</i>	0.15541434	3
<i>BLK</i>	0.156022806	4
<i>PLTP</i>	0.290540089	4
<i>ABCG1</i>	0.29216291	4
<i>CETP</i>	0.292540617	4
<i>SCARB1</i>	0.293263721	4
<i>GPIHBP1</i>	0.338799533	3
<i>LIPC</i>	0.342994911	3
<i>LPL</i>	0.348510034	3
<i>PPARA</i>	0.805521573	3
<i>MTTP</i>	0.817677583	3
<i>TRIB1</i>	0.821329205	2
<i>HNF1A</i>	0.833601165	3
<i>LDLRAP1</i>	0.84751109	3
<i>COQ2</i>	0.853822722	2
<i>PAX4</i>	0.857245649	3
<i>AKT2</i>	0.867023187	2
<i>ANGPTL3</i>	0.877304003	2
<i>MFN2</i>	1	2
<i>GALNT2</i>	1	2
<i>KLF11</i>	1	2
<i>GCKR</i>	1	2
<i>NEUROD1</i>	1	1
<i>PPARG</i>	1	2
<i>STAP1</i>	1	1

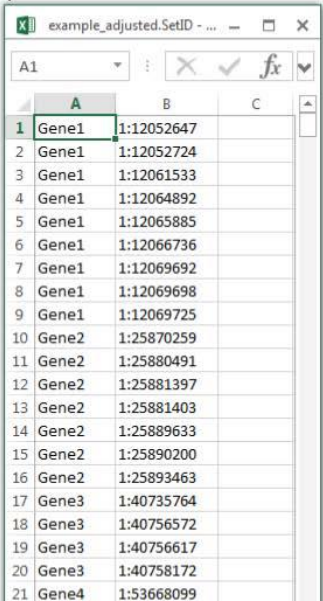
<i>GCK</i>	1	1
<i>CAV2</i>	1	1
<i>AGPAT2</i>	1	2
<i>LIPA</i>	1	1
<i>KCNJ11</i>	1	2
<i>SLC22A8</i>	1	2
<i>APOA5</i>	1	2
<i>GPD1</i>	1	1
<i>PDX1</i>	1	2
<i>LCAT</i>	1	1
<i>HNF1B</i>	1	2
<i>LIPG</i>	1	1
<i>CREB3L3</i>	1	2
<i>DYRK1B</i>	1	2

P-values have been adjusted using the Bonferroni correction method. "\*" indicate genes known to cause familial hypercholesterolemia.

### Additional Information

#### File-type Glossary

File Type	Extension	Additional Notes
PLINK binary biallelic genotype table	.bed	<ul style="list-style-type: none"> <li>Described <a href="#">here</a></li> <li>Not to be confused with the UCSC Genome Browser's BED format</li> </ul>
PLINK extended .map file	.bim	<ul style="list-style-type: none"> <li>Described <a href="#">here</a></li> </ul>
PLINK sample information file	.fam	<ul style="list-style-type: none"> <li>Described <a href="#">here</a></li> </ul>  <ul style="list-style-type: none"> <li>Headers are included in this image for clarity but should not be present when running Exautomate. As well, for our purposes, Exautomate does not require information in columns C to E, so they can remain coded as "0".</li> <li>When utilizing PLINK for relatedness applications, columns A and B may not be identical, and columns C and D would have additional information.</li> </ul>
FASTQ	.fastq	<ul style="list-style-type: none"> <li>Contains both sequencing reads and the base Phred qualities</li> </ul>
PLINK text fileset variant information file	.map	<ul style="list-style-type: none"> <li>Described <a href="#">here</a></li> </ul>
PLINK/MERLIN/Haploview text pedigree +	.ped	<ul style="list-style-type: none"> <li>Described <a href="#">here</a></li> </ul>

genotype table		
Set identified file	.SetID	<ul style="list-style-type: none"> <li>• This file defines variant sets               <ul style="list-style-type: none"> <li>◦ In Exautomate, the assumption is that the sets of interest are by gene</li> </ul> </li> <li>• The first column contains the name of the set (i.e. the name of the gene), and the second column contains the variant position</li> </ul>  <ul style="list-style-type: none"> <li>• Important considerations are required when dealing with .SetID files               <ul style="list-style-type: none"> <li>◦ The name of the set must be less than 50 characters long</li> <li>◦ Duplicate entries (i.e. matching set and variant position) are not allowed</li> <li>◦ ANNOVAR generates extra transcript information that may need to be manually removed following a prompt from Exautomate</li> </ul> </li> </ul>
Variant Call Format file	.vcf	<ul style="list-style-type: none"> <li>• Contains basic information on genetic variants               <ul style="list-style-type: none"> <li>◦ Chromosome</li> <li>◦ Scaffold position</li> <li>◦ rsID (if applicable)</li> <li>◦ Reference and alternate allele</li> </ul> </li> <li>• May also include additional information, such as</li> </ul>



		<p>sequencing depth, genotype, and genotype quality</p> <ul style="list-style-type: none"><li>○ If there are errors when running Exautomate, the user may consider using bcftools to remove some of this additional information that is not necessary for RVAA</li></ul>
--	--	--

**References**

1. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
2. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

# Curriculum Vitae

## JACQUELINE DRON

H-index: 13 | Citations: 626

### EDUCATION

#### Western University, London ON

#### Doctor of Philosophy (PhD)

September 2015 – December 2020

*Biochemistry*

- Thesis title: *The genetic determinants of complex lipid and lipoprotein phenotypes*
- Supervisor: Robert A. Hegele, MD

#### Bachelor of Science, Honours (BSc)

September 2010 – April 2015

*Honours Specialization in Genetics and Microbiology and Immunology Major*

### LABORATORY AND RESEARCH EXPERIENCE

#### Postdoctoral Research Scholar

January 2021 – TBD

*Precision Medicine Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston MA, United States | Medical and Population Genetics Group, Broad Institute, Cambridge MA, United States*

- Supervisor: Amit V. Khera, MD MSc

#### PhD Candidate

September 2015 – December 2020

*Blackburn Cardiovascular Genetics Lab, Robarts Research Institute, Western University, London ON, Canada*

- Supervisor: Robert A. Hegele, MD

#### Visiting Scholar

January 2020 – June 2020

*Precision Medicine Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston MA, United States | Medical and Population Genetics Group, Broad Institute, Cambridge MA, United States*

- Supervisor: Amit V. Khera, MD MSc

#### Undergraduate Honours Research Thesis

September 2014 – April 2015

*Blackburn Cardiovascular Genetics Lab, Robarts Research Institute, Western University, London ON, Canada*

- Supervisor: Robert A. Hegele, MD

### FELLOWSHIPS (1)

#### NHLBI BioData Catalyst Fellow

September 2020 – September 2021

Project Title: *Modeling lifetime coronary artery disease trajectories of naturally occurring DNA variation linked to low circulating triglycerides*

## HONOURS, ACHIEVEMENTS AND AWARDS (18)

### 2019

---

- **Michael Smith Foreign Study Supplement** | CIHR
- **Norman E. Nixon Marie Rämö Nixon Award** | The Schulich School of Medicine and Dentistry
- **The Lucille & Norton Wolf London Health Research Day Trainee Publication Award** | The London Health Research Day Academic Committee
- **Rising Star Award in the Lipids Field** | The Journal of Lipid Research

### 2018

---

- **Academic Achievement Scholarship** | The Public Service Alliance of Canada, Local 610 at Western University
- **Cobban Student Award in Heart and Stroke Research** | The Schulich School of Medicine and Dentistry
- **Nellie Farthing Fellowship in Medical Sciences Research** | The Schulich School of Medicine and Dentistry
- **London Health Research Day Oral Presentation Award** | The London Health Research Day Academic Committee
- **The Lucille & Norton Wolf London Health Research Day Trainee Publication Award** | The London Health Research Day Academic Committee
- **Queen Elizabeth II Graduate Scholarship in Science and Technology – DECLINED** (1 year) | The School of Graduate and Postdoctoral Studies, Western University
- **CIHR Doctoral Award – Fredrick Banting and Charles Best Canada Graduate Scholarship** (3 years) | CIHR
- **CIHR Travel Award – Institute Community Support** | CIHR – Institute of Circulatory and Respiratory Health
- **Canadian Society of Atherosclerosis, Thrombosis, and Vascular Biology Trainee Travel Award** | Canadian Society of Atherosclerosis, Thrombosis, and Vascular Biology

### 2017

---

- **Academic Achievement Scholarship** | The Public Service Alliance of Canada, Local 610 at Western University
- **Canadian Society of Atherosclerosis, Thrombosis, and Vascular Biology Trainee Travel Award** | Canadian Society of Atherosclerosis, Thrombosis, and Vascular Biology
- **Ontario Graduate Scholarship** (1 year) | The School of Graduate and Postdoctoral Studies, Western University
- **CIHR Travel Award – Institute Community Support** | CIHR – Institute of Circulatory and Respiratory Health

### 2016

---

- **Outstanding Research Contributions Scholarship** | The Public Service Alliance of Canada, Local 610 at Western University

## RESEARCH ARTICLES (17)

1. Lazarte J, **Dron JS**, McIntyre AD, Skanes AC, Gula LJ, Tang AS, Tadros R, Hegele RA, Roberts JD. The role of common genetic variation in “lone” atrial fibrillation. (2020) . *Circ Genom Precis Med* [invited revisions]
2. Lazarte J, Laksman Z, Wang J, Robinson JF, **Dron JS**, Liew J, McIntyre AD, Skanes AC, Gula LJ, Leong-Sit P, Cao H, Trost B, Scherer SW, Hegele RA, Roberts JD. Enrichment of loss-of-function and copy-number variants in ventricular cardiomyopathy genes in “lone” atrial fibrillation. (2020) *EP Europace*. [under review]
3. Gill PK\*, **Dron JS\***, Dillioott AA, McIntyre AD, Cao H, Wang J, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. Ancestry-specific profiles of genetic determinants of severe hypertriglyceridemia. (2020) *J Clin Lipidol*. [doi: 10.1016/j.jacl.2020.11.007]
4. Gill PK\*, **Dron JS\***, Berberich AJ, Wang J, McIntyre AD, Cao H, Hegele RA. Combined hyperlipidemia is similar to isolated hypertriglyceridemia. (2020) *J Clin Lipidol*. [doi: 10.1016/j.jacl.2020.11.006]
5. **Dron JS**, Lawson A, Davis BD, Wang J, Cao H, McIntyre AD, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. (2020) Loss-of-function *CREB3L3* variants in patients with severe hypertriglyceridemia (HTG). *Arterioscler Thromb Vasc Biol*. 40(8):1935-1941
6. **Dron JS**, Wang J, McIntyre AD, Iacocca MA, Robinson JF, Ban MR, Cao H, Hegele RA. (2020) Six years' experience with LipidSeq: clinical and research learnings from a hybrid, targeted sequencing panel for dyslipidemias. *BMC Med Gen*. 13(1):23
7. **Dron JS**, Wang J, McIntyre AD, Cao H, Hegele RA. (2020) The polygenic nature of mild-to-moderate hypertriglyceridemia. *J Clin Lipidol*. 14(1):28-34
8. **Dron JS**, Wang J, McIntyre AD, Cao H, Robinson JF, Duell PB, Manjoo P, Feng J, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. (2019) Partial LPL deletions: rare copy-number variants contributing towards the polygenic form of severe hypertriglyceridemia. *J Lipid Res*. 60(11):1953-1958
9. Berg MD, Giguere DJ, **Dron JS**, Lant JT, Genereaux J, Liao C, Wang J, Robinson JF, Gloor GB, Hegele RA, O'Donoghue P, Brandl CJ. (2019) Targeted sequencing reveals expanded genetic diversity of human transfer RNAs. *RNA Biol*. 16(11):1574-1585
10. **Dron JS\***, Davis BD\*, Robinson JF, Hegele RA, Lizotte DJ. (2019) Exautomate: A user-friendly tool for region-based rare variant association analysis (RVAA). *BioRxiv*. [DOI: 10.1101/649368]
11. Gulilat M, Lamb T, Teft WA, Wang J, **Dron JS**, Robinson JF, Tirona RG, Hegele RA, Kim RB, Schwarz UI. (2019) Targeted next generation sequencing as a tool for precision medicine. *BMC Medical Gen*. 12:81
12. **Dron JS**, Wang J, Cao H, McIntyre AD, Iacocca MA, Menard JR, Movsesyan I, Malloy MJ, Pullinger CR, Kane JP, Hegele RA. (2019) Severe hypertriglyceridemia is primarily polygenic. *J Clin Lipidol*. 13(1):80-88
13. Iacocca MA, Wang J, Sarkar S, **Dron JS**, Lagace T, McIntyre AD, Lau P, Robinson JF, Yang P, Knoll JH, Cao H, McPherson R, Hegele RA. (2018) Whole-gene duplication of PCSK9 as a novel genetic mechanism for severe familial hypercholesterolemia. *Can J Cardiol*. 34(10):1316-1324
14. **Dron JS**, Wang J, Berberich AJ, Iacocca MA, Cao H, Yang P, Knoll J, Tremblay K, Brisson D, Netzer C, Gouni-Berthold I, Gaudet D, Hegele RA. (2018) Large-scale deletions of the ABCA1 gene in patients with hypoalphalipoproteinemia. *J Lipid Res*. 59(8):1529-1535
15. **Dron JS**, Wang J, Low-Kam C, Khetarpal SA, Robinson JF, McIntyre AD, Ban MR, Cao H, Rhainds D, Dubé M-P, Rader DJ, Lettre G, Tardif J-C, Hegele RA. (2017) Polygenic

- determinants in extremes of high-density lipoprotein cholesterol. *J Lipid Res.* 58(11):2162-2170
16. Iacocca MA, Wang J, **Dron JS**, Robinson JF, McIntyre AD, Cao H, Hegele RA. (2017) Use of next-generation sequencing to detect LDLR gene copy number variation in familial hypercholesterolemia. *J Lipid Res.* 58(11):2202-2209
  17. Wang J, **Dron JS**, Ban MR, Robinson JF, McIntyre AD, Alazzam M, Zhao PJ, Dilliot AA, Cao H, Huff MW, Rhoads D, Low-Kam C, Dube M-P, Lettre G, Tardif J-C, Hegele RA. (2016) Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. *Arterioscler Thromb Vasc Biol.* 36(12):2439-2445

\* Co-first authors.

## REVIEW ARTICLES (12)

1. Gill PK, **Dron JS**, Hegele RA. (2020) Genetics of hypertriglyceridemia and atherosclerosis. *Curr Opin Cardiol.* [in press]
2. **Dron JS** and Hegele RA. (2020) Genetics of hypertriglyceridemia. *Front Endocrinol.* [doi: 10.3389/fendo.2020.00455]
3. Hegele RA and **Dron JS**. (2020) 2019 George Lyman Duff Memorial Lecture: Three decades of examining DNA in patients with dyslipidemia. *Arterioscler Thromb Vasc Biol.* 40(9):1970-1982.
4. Iacocca MA, **Dron JS**, Hegele RA. (2019) Progress in finding pathogenic DNA copy number variations in dyslipidemia. *Curr Opin Lipidol.* 30(2):63-70
5. **Dron JS** and Hegele RA. (2019) The evolution of genetic-based risk scores for lipids and cardiovascular disease. *Curr Opin Lipidol.* 30(2):71-81
6. **Dron\* JS**, Lazarte\* J, Hegele RA. (2018) Recent highlights of ATVB – Lipoproteins. *Arterioscler Thromb Vasc Biol.* 38(11):133-143
7. **Dron JS**, and Hegele RA. (2018) Polygenic influences on dyslipidemias. *Curr Opin Lipidol.* 29(2):133-143
8. **Dron\* JS**, Ho\* R, Hegele RA. (2017) Recent advances in the genetics of atherothrombotic disease and its determinants. *Arterioscler Thromb Vasc Biol.* 37(10):e158-166
9. **Dron JS** and Hegele RA. (2017) Genetics of triglycerides and the risk of atherosclerosis. *Curr Athero Rep.* 19(7):31 [DOI: 10.1007/s11883-017-0667-9]
10. **Dron JS**, and Hegele RA. (2017) Complexity of mechanisms among human proprotein convertase subtilisin-kexin type 9 variants. *Curr Opin Lipidol.* 28(2):161-169
11. Burke\* AC, **Dron\* JS**, Hegele RA, and Huff MW. (2017) PCSK9: Regulation and target for drug development for dyslipidemia. *Ann Rev Pharmacol Toxicol.* 57:223-244
12. **Dron JS** and Hegele RA. (2016) Genetics of lipid and lipoprotein disorders and traits. *Curr Genet Med Rep.* 4(3):130-141

## BOOK CHAPTERS (1)

1. **Dron JS**, and Hegele RA. (March 2017) Molecular genetics of hypertriglyceridemia. In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net> [doi: 10.1002/9780470015902.a0026710]

## LETTERS, EDITORIALS AND PERSPECTIVES (4)

1. Hannah W, Lambert SA, ... **Dron JS** [9 out of 33] ... Wojcik G. (2020) Improving reporting standards for polygenic scores in risk prediction studies. *Nature* [invited revisions]

2. **Dron JS\***, Wang W\*, Patel AP, Ng K, Hegele RA, Khera AV. (2020) A genetic predictor to identify individuals with high lipoprotein(a) concentrations. *Circ Genom Precis Med* [in press]
3. **Dron JS**, Hegele RA. (2019) Ischemic event reduction and triglycerides. *J Am Coll Cardiol.* 74(14):1848-1849
4. **Dron JS**, Hegele RA. (2019) A tip of the CAP1 to cholesterol metabolism. *Euro Heart J.* [DOI: 10.1093/eurheartj/ehz771]

### INVITED CONFERENCE PRESENTATIONS AND SEMINARS (2)

1. **European Atherosclerosis Society Congress** (International; Helsinki, Finland; May 30-June 2, 2021), “Novel insight into the genetics of hypertriglyceridemia”
2. **Experimental Biology Conference; American Society of Biochemistry and Molecular Biology** (International; Orlando, FL; April 6-9, 2019), “Complex genetic determinants of hypertriglyceridemia”

### PLATFORM CONFERENCE PRESENTATIONS (7)

1. † **Canadian Lipoprotein Conference/Vascular Lipid Summit** (National; Banff, AB; October 3-6, 2019)
2. **International Symposium on Atherosclerosis** (International; Toronto, ON; June 9-12, 2018)
3. † **Canadian Lipoprotein Conference** (National; Toronto, ON; June 7-9, 2018)
4. † **London Health Research Day** (Regional; London, ON; May 10, 2018)
5. † **Canadian Lipoprotein Conference** (National; St. John's, NL; September 15-18, 2016)
6. **Canadian Human and Statistical Genetics Meeting** (National; Halifax, NS; April 16-19, 2016)
7. † **Schulich School of Medicine and Dentistry, Department of Medicine Resident Research Day** (Institutional; London, ON; May 26, 2016)

† Award-winning

### POSTER CONFERENCE PRESENTATIONS (14)

1. **American Society of Human Genetics** (International; Houston, TX; October 15-19, 2019)
2. **European Society of Human Genetics** (International; Gothenburg, Sweden; June 15-18, 2019)
3. **Schulich School of Medicine and Dentistry, Department of Medicine Resident Research Day** (Institutional; London, ON; May 10, 2019)
4. **London Health Research Day** (Regional; London, ON; April 30, 2019)
5. **American Society of Human Genetics** (International; San Diego, CA; October 15-20, 2018)
6. **International Symposium on Atherosclerosis** (International; Toronto, ON; June 9-12, 2018)
7. **Canadian Lipoprotein Conference** (National; Toronto, ON; June 7-9, 2018)
8. **Schulich School of Medicine and Dentistry, Department of Medicine Resident Research Day** (Institutional; London, ON; May 11, 2018)
9. **Canadian Lipoprotein Conference** (National; Ottawa, ON; October 19-22, 2017)
10. **Arteriosclerosis, Thrombosis and Vascular Biology | Peripheral Vascular Disease Conference** (International; Minneapolis, MN; May 4-6, 2017)
11. **London Health Research Day** (Regional; London, ON; March 28, 2017)
12. **Arteriosclerosis, Thrombosis and Vascular Biology | Peripheral Vascular Disease Conference** (International; Nashville, TN; May 5-7, 2016)
13. **London Health Research Day** (Regional; London, ON; March 19, 2016)
14. **Canadian Lipoprotein Conference** (National; Toronto, ON; October 15-18, 2015)

## TEACHING EXPERIENCE

### GRADUATE TEACHING ASSISTANTSHIPS

- |  |                                |
|--|--------------------------------|
| ▪ Med Sci 4900FG: Advanced Medical Sciences Lab                | <b>Sept. 2018 – Dec. 2019</b>  |
| ▪ Med Sci 4930FG: Selected Topics in Medical Sciences          | <b>Sept. 2018 – Dec. 2019</b>  |
| ▪ Med Sci 4300FG: Healthcare Challenges and Scientific Inquiry | <b>July 2017 – August 2017</b> |
| ▪ Med Sci 3900FG: Interdisciplinary Medical Sciences Lab       | <b>Sept. 2016 – April 2019</b> |

## MENTORING EXPERIENCE

- Undergraduate Thesis Students, Western University [8]
- Summer Work Students, Western University [4]
- Co-op Students, University of Waterloo [3]

## INTERVIEWS AND MEDIA RELATIONS

- ASBMB Today, JLR Early-Career Researcher – [“Investigating genetic variants that underlie abnormal lipid phenotypes”](#) (March 2019)
- The Collaborator, Schulich School of Medicine and Dentistry – [“On a mission to increase diversity and inclusion in the world of science”](#) (April 2019)
- CTV News London – [“Robarts unveils new information on cholesterol that could save lives”](#) (October 2016)
- Robarts Discovery Newsletter, Trainee Profile – [“Genetic discovery”](#) (October 2016)
- CHRW 94.9 FM, Gradcast – [“High-density lipoprotein cholesterol”](#) (September 2015)