

Electronic Thesis and Dissertation Repository

8-2-2018 1:30 PM

Genetic determinants underlying rare diseases identified using next-generation sequencing technologies

Rosettia Ho, *The University of Western Ontario*

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

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

© Rosettia Ho 2018

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



Part of the [Medical Genetics Commons](#)

Recommended Citation

Ho, Rosettia, "Genetic determinants underlying rare diseases identified using next-generation sequencing technologies" (2018). *Electronic Thesis and Dissertation Repository*. 5497.
<https://ir.lib.uwo.ca/etd/5497>

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

Abstract

Rare disorders affect less than one in 2000 individuals, placing a huge burden on individuals, families and the health care system. Gene discovery is the starting point in understanding the molecular mechanisms underlying these diseases. The advent of next-generation sequencing has accelerated discovery of disease-causing genetic variants and is showing numerous benefits for research and medicine. I describe the application of next-generation sequencing, namely LipidSeq™ – a targeted resequencing panel for the identification of dyslipidemia-associated variants – and whole-exome sequencing, to identify genetic determinants of several rare diseases. Utilization of next-generation sequencing plus associated bioinformatics led to the discovery of disease-associated variants for 71 patients with lipodystrophy, two with early-onset obesity, and families with brachydactyly, cerebral atrophy, microcephaly-ichthyosis, and widow's peak syndrome. Understanding these variants and their contribution to disease can increase understanding of disease mechanisms and help with the development of therapeutic interventions in the future.

Keywords

Next-generation sequencing; targeted sequencing; whole-exome sequencing; Mendelian disease; genetic disease; rare disease; metabolic disease; lipodystrophy; extreme early-onset obesity; brachydactyly; cerebral atrophy; CARE for RARE; FORGE Canada; autosomal recessive; autosomal dominant; variant discovery; copy-number variation

Co-Authorship Statement

Dr. Robert A. Hegele provided all the supervision and funding, and his input for the conceptualization, design and protocols of this Thesis project. Dr. Hegele also provided all lipodystrophy, two extreme early-onset obesity, and cerebral atrophy samples.

Chapters 1, 3 and 4 contain modified text from a submitted manuscript entitled, “Whole-exome sequencing identifies a novel *IHH* insertion in an Ontario family with brachydactyly type A1”, co-authored by Adam D. McIntyre, Brooke A. Kennedy, and Dr. Robert A. Hegele.

Dr. Hegele, Dr. Kathleen A. Hill and Dr. C. Anthony Rugar provided critical revision for all Chapters.

Dr. Victoria M. Siu provided three of the extreme early-onset obesity samples discussed in the Thesis, and the CARE for RARE and FORGE Canada consortiums provided samples for the CARE for RARE/FORGE Canada projects. Dr. Siu, Dr. Rugar and Dr. Rana Chakrabarti provided clinical information, additional samples and support.

Dr. Jian Wang, Dr. Henian Cao, and Adam D. McIntyre each contributed their technical expertise to the design of this Thesis project. More specifically, Dr. Wang helped with primer design and conceptualization, and greatly assisted with the validation of copy-number variation events, Dr. Cao performed all LipidSeq™ targeted sequencing, and Adam C. McIntyre provided clinical information and technical support. John F. Robinson provided assistance and support for each CARE for RARE and FORGE Canada project. Brooke A. Kennedy provided clinical information for the brachydactyly case.

In addition to the co-authors listed above, several students were involved in the CARE for RARE and FORGE Canada projects. Calwing Liao provided expertise and technical support for the microcephaly-ichthyosis, heterotaxy, and widow’s peak cases; Cory Soinen worked on the dystonia case; Gagandeep Singh worked on sequence analysis for the craniorachischisis case; Jing (Tony) Yao worked on the craniosynostosis case; and Xingyu (Robin) Liu helped with the sequencing analysis of the cryptophthalmos case.

Dedication

In loving memory of my Grandma – thank you for supporting me, loving me unconditionally and raising me to be the person I am today. I love and miss you more than words can say.

Acknowledgements

The past few years in the Hegele lab have been some of the most enriching years of my life. I have learned more than I thought possible from extraordinary individuals and created great friendships with my exceptional colleagues. Here, I would like to acknowledge the fantastic individuals that made my Thesis such a rewarding experience.

First and foremost, I would like to thank my supervisor, Dr. Rob Hegele, for his supervision, support and mentorship. You have been a guiding light throughout my entire Masters, and I will be forever grateful for the time I spent under your supervision. In addition to your knowledge, thank you for bringing me laughter on Friday afternoons, and sharing your love of dogs. Speaking on behalf of all the graduate students past-and-present and myself, we are so thankful for everything you do. Thank you for taking a chance on me in third year. It has been a privilege to be your student.

To my advisory committee, Dr. Kathleen Hill and Dr. Tony Rugar, thank you for your outstanding mentorship. I am so appreciative of all the feedback and support you have provided, not only during committee meetings but elsewhere as well.

An additional thank you to Dr. Hill for being so supportive these past few years. It has been an absolute honour to have you as an advisor for both my undergraduate and graduate Thesis projects.

Thank you to Dr. Victoria Siu for being a mentor to me. It has been a pleasure to work with you on this Thesis project, and to volunteer in the clinic with you. Your wisdom and graciousness are inspiring.

To the entirety of the Hegele lab, thank you so much for being such excellent colleagues and listening to me rant throughout my years in the lab. Thank you to Adam McIntyre for all your support throughout my Thesis – from helping me find charts, office supplies or DNA samples – you have been so helpful throughout it all. Thank you to John for your help along the way. Also, thank you so much to Dr. Henian Cao for all your technical assistance and mentorship. Furthermore, thank you, Dr. Jian Wang, not only for your

mentorship but also for recommending the best Chinese restaurants and being such a great colleague – I hope that one day we finally get our trio sequenced.

To the Department of Biochemistry, thank you for providing administrative and research support. Thank you to Barb Green for being a friendly face in the department and providing administrative assistance.

Thank you to the friends I have made along the way – Emily, Cal, Amy, Nadya, Freda – you are the best. Thank you so also to the brilliant Sali for your mentorship, kindness, and guidance.

To my main pals – Alli, Jacqueline and Michael – thank you so much for being such remarkable human beings and for all the memories. I would not be where I am without you and your humour, laughter, and support of my eccentric personality. I look forward to staying in contact and watching you all succeed in the future. More specifically, Alli – thank you for being my twin. We are the same person and I would not have it any other way. I could not imagine having a better, more beautiful or generous neighbour than you. Jacqueline – you are an angel on earth. Thank you for your kindness and thoughtfulness; you inspire me every day. Thank you for being the light in the darkest of times. Michael – thank you for being such a great friend. You and I are two peas in a pod; if I had to choose anyone to be incarcerated with, it would be you.

To my family – I love you with all my heart. Thank you to my Mom and Dad for your encouragement, support and unconditional love. I know how much you have sacrificed for Neuton and me. I could not imagine better, more generous, caring parents than you. To my sidekick Neuton (hey, brother!), you are not just my sibling but also my best friend. To Shirley, you are a sister to me, and I am so grateful for our lifelong relationship. Thank you, Neu and Shin, for keeping me young and fun and being there to talk to about anything. To my Grandma, I will cherish our memories forever. To Jayvee, thank you for opening my heart and coming into my life when I needed you the most. Lastly, to my partner Alex, thank you for indulging in Junior Chickens with me, going to every restaurant, letting me vent, and being there for the last six-and-a-half years and counting. You are the definition of a partner-in-crime, and I am so grateful for you.

Table of Contents

Abstract	i
Keywords	ii
Co-Authorship Statement	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Figures	xiii
List of Tables	xiv
List of Appendices	xv
List of Abbreviations	xvi
Chapter 1 – Introduction	1
1.1 Rare disease burden in Canada.....	1
1.1.1 Many rare diseases are of genetic origin.....	2
1.1.2 Research problem and motivation	2
1.2 Rare human genetic diseases.....	3
1.2.1 Inherited lipodystrophies	3
1.2.1.1 Congenital generalized lipodystrophy.....	3
1.2.1.2 Familial partial lipodystrophy	4
1.2.2 Extreme early-onset obesity.....	5
1.2.3 Isolated brachydactyly	6
1.2.4 Cerebral atrophy.....	8
1.2.5 FORGE Canada and CARE for RARE.....	8
1.2.5.1 Project overview.....	8
1.2.5.2 Project significance	9
1.2.5.3 FORGE Canada and CARE for RARE cases.....	9
1.3 The human genome sequence	13
1.3.1 Variation within the human genome.....	13
1.3.1.1 Single nucleotide variation.....	14
1.3.1.1.1 Splicing variation.....	15
1.3.1.1.2 Insertions and deletions	15
1.3.1.2 Repetitive elements	16

1.3.1.3	Structural variation	16
1.3.1.3.1	Copy-number variation.....	17
1.3.2	Germline versus somatic variation.....	17
1.3.3	Inherited versus <i>de novo</i> variation	17
1.4	Human genetic diseases	18
1.4.1	Heritability of genetic diseases	18
1.4.1.1	Monogenic diseases.....	19
1.4.1.2	Digenic diseases	20
1.4.1.3	Polygenic diseases.....	21
1.4.1.4	Disease penetrance and expressivity.....	21
1.5	Genetic variation and rare diseases	22
1.5.1	The genetics of lipodystrophy.....	22
1.5.1.1	Congenital generalized lipodystrophy.....	23
1.5.1.2	Familial partial lipodystrophy	24
1.5.1.3	Unexplained cases of lipodystrophy	25
1.5.2	The genetics of obesity	25
1.5.2.1	Unexplained cases of early-onset obesity	27
1.5.3	The genetics of brachydactyly	27
1.5.3.1	Brachydactyly type A.....	29
1.5.3.1.1	Brachydactyly type A1 (BDA1).....	29
1.5.3.1.2	Brachydactyly type A1, B (BDA1B).....	29
1.5.3.1.3	Brachydactyly type A1, C (BDA1C).....	30
1.5.3.1.4	Brachydactyly type A1, D (BDA1D)	30
1.5.3.1.5	Brachydactyly type A2 (BDA2).....	30
1.5.3.1.6	Brachydactyly type A3 (BDA3).....	31
1.5.3.1.7	Brachydactyly type A4 (BDA4).....	31
1.5.3.1.8	Unexplained cases of brachydactyly	31
1.5.4	The genetics of autosomal recessive cerebral atrophy.....	34
1.5.4.1	Unexplained cerebral atrophy	34
1.5.5	Genetically unexplained consortium cases	36
1.6	Next-generation sequencing and rare diseases.....	36
1.6.1	Targeted resequencing	37

1.6.1.1	LipidSeq™	37
1.6.2	Whole-exome sequencing.....	45
1.6.2.1	Whole-exome sequencing successes.....	45
1.6.3	Whole-genome sequencing.....	45
1.7	Thesis project outline	46
1.7.1	Project overview	46
1.7.2	Project hypothesis	47
1.7.3	Project goal and aims	47
Chapter 2	– Materials and Methods.....	48
2	Overview.....	48
2.1	Ethics approval.....	48
2.2	Study subjects.....	50
2.2.1	Clinical information.....	50
2.3	DNA extraction and quality control.....	50
2.4	SNP genotyping and autozygosity mapping	50
2.5	Next-generation sequencing.....	51
2.5.1	Patient selection for LipidSeq™.....	51
2.5.1.1	LipidSeq™ targeted sequencing	51
2.5.2	Patient selection for whole-exome sequencing.....	52
2.5.2.1	Whole-exome sequencing	53
2.6	Bioinformatics processing.....	54
2.7	Variant annotation	56
2.8	Variant prioritization.....	56
2.8.1	Zygoty	58
2.8.2	Read depth	58
2.8.3	Sequence ontology.....	58
2.8.4	Minor allele frequency.....	59
2.8.5	<i>In silico</i> analyses	59
2.8.6	Gene and protein analysis and literature review	59
2.8.7	Association with the disease phenotype	59
2.9	Copy-number variation detection.....	60
2.9.1	Copy-number variation prioritization	61

2.10	Variant validation and co-segregation analysis.....	61
2.10.1	PCR.....	61
2.10.2	PCR imaging and purification	61
2.10.3	Sanger sequencing and analysis.....	62
2.11	Validation of copy-number variation events.....	62
2.12	Gene enrichment analysis.....	63
2.13	Statistical analysis: Brachydactyly case.....	63
2.14	Box plot generation: Brachydactyly case.....	63
Chapter 3 – Results		64
3	Overview.....	64
3.1	Study subjects and clinical features	64
3.2	LipidSeq™ variant detection	64
3.3	Sanger sequencing of <i>TMPRSS4</i>	66
3.4	Whole-exome sequencing	66
3.4.1	Lipodystrophy	69
3.4.2	Extreme early-onset obesity.....	69
3.4.3	Brachydactyly	72
3.4.4	Cerebral atrophy-like	72
3.4.5	CARE for RARE and FORGE Canada.....	73
3.5	Shared sequence variation.....	75
3.6	Copy-number variation discovery.....	78
3.6.1	LipidSeq™ copy-number variation detection.....	78
3.6.2	Exome copy-number variation detection.....	80
3.7	Validation of candidate variants and copy-number variations.....	80
3.7.1	<i>IHH</i> validation (brachydactyly).....	80
3.7.2	<i>VPS52</i> validation (microcephaly-ichthyosis).....	83
3.7.3	<i>KDM6A</i> validation (widow’s peak).....	83
3.7.4	<i>AR</i> validation (lipodystrophy and obesity)	83
3.7.5	Copy-number variation validation.....	86
3.7.5.1	Patient 1099: <i>PLIN1</i> exon 2 deletion	86
3.7.5.2	Patient 1120: <i>CREB3L3</i> duplication	88
3.8	Gene enrichment analysis.....	90

3.9	Statistical analysis: Brachydactyly case	90
Chapter 4 – Discussion	92
4	Summary of findings.....	92
4.1	LipidSeq™	93
4.1.1	Unexpected variation in <i>POLD1</i>	93
4.1.2	Copy-number variation findings	94
4.1.2.1	Patient 1099.....	94
4.1.2.2	Patient 2108.....	95
4.1.2.3	Patient 15529.....	95
4.1.2.4	Patient 4192	95
4.1.2.5	Patient 14557.....	96
4.1.2.6	Patient 15678.....	97
4.1.2.7	Patient 1120	97
4.1.3	Benefits of LipidSeq™	98
4.1.4	Limitations of LipidSeq™	98
4.2	Whole-exome sequencing of patients with rare metabolic diseases	99
4.2.1	Lipodystrophy	100
4.2.1.1	Patient 1804.....	100
4.2.1.2	Patient 8615.....	100
4.2.1.3	Patient 1088	101
4.2.1.4	Patient 3458.....	102
4.2.1.5	Patient 3969.....	102
4.2.1.6	Patient 4116.....	103
4.2.1.7	Patient 4665.....	103
4.2.1.8	Patient 5662.....	104
4.2.1.9	Patient 2025.....	105
4.2.1.10	Patient 3011	105
4.2.1.11	Patient 14318	106
4.2.1.12	Patient 4194	107
4.2.1.13	Patient 1120	108
4.2.2	Extreme early-onset obesity.....	109
4.2.2.1	Patient 15544.....	109

4.2.2.2	Patient 15858.....	109
4.2.3	Lipodystrophy-obesity connection.....	110
4.3	Whole-exome sequencing of rare disease families	112
4.3.1	Brachydactyly type A1	112
4.3.2	Cerebral atrophy-like	114
4.3.2.1	CASC5 and primary microcephaly-4.....	114
4.4	Whole-exome sequencing for CARE4RARE and FORGE Canada projects..	116
4.4.1	Microcephaly-ichthyosis in the Old Order Amish.....	116
4.4.2	Widow’s peak syndrome.....	117
4.4.3	CARE for RARE and FORGE summary.....	118
4.5	Gene enrichment analysis.....	118
4.6	Clinical utility of LipidSeq™ and whole-exome sequencing	120
4.6.1	Comparison of LipidSeq™ and whole-exome sequencing to whole-genome sequencing	120
4.7	Limitations of the project	121
4.7.1	Study considerations	121
4.7.2	General methodological limitations.....	122
4.8	Future directions.....	123
4.9	Conclusions	129
4.10	Project significance	130
	References	132
	Appendices	150
	Curriculum Vitae	160

List of Figures

Figure 1.5.3.1.8 Pedigree of a Canadian autosomal dominant brachydactyly-affected family	33
Figure 1.5.4.1 Pedigree of a family with cerebral atrophy	35
Figure 2. Flowchart of steps taken in order to identify the genetic bases of rare disorders	49
Figure 2.6. Next-generation sequencing bioinformatics workflow	55
Figure 3.2. Variant breakdown for patients with lipodystrophy sequenced using the LipidSeq™ targeted gene panel	65
Figure 3.5. Androgen receptor gene and protein structure	77
Figure 3.7.1A Pedigree and <i>IHH</i> genotype information for a Canadian autosomal dominant brachydactyly type A1-affected family.....	81
Figure 3.7.1B Chromosome location, gene structure, sequence analysis and protein structure of Indian Hedgehog	82
Figure 3.7.2 Sanger sequencing confirmation of the <i>VPS52</i> gene variant in a family with a history of microcephaly-ichthyosis	84
Figure 3.7.3 Sanger sequencing confirmation of the <i>KDM6A</i> gene variant in a family with a history of widow’s peak syndrome	85
Figure 3.7.5.1 Validation of a <i>PLINI</i> exon 2 deletion in a patient with familial partial lipodystrophy	87
Figure 3.7.5.2 Detection of a duplication event from LipidSeq™ versus whole-exome sequencing data	89
Figure 3.9 Comparison of anthropometric measures between those positive for the <i>IHH</i> mutation and those negative for the <i>IHH</i> mutation	91

List of Tables

Table 1.2.3 Clinical features of isolated brachydactyly	7
Table 1.2.5.3 Clinical description and pedigrees for all CARE for RARE and FORGE Canada cases	10
Table 1.5.2 Genes involved in the pathophysiology of obesity	26
Table 1.5.3 The genetics of isolated brachydactyly	28
Table 1.6.1.1 Details regarding the genes targeted by the LipidSeq™ gene panel relating to lipodystrophy and associated metabolic comorbidities	38
Table 2.8 Criteria used to prioritize candidate variants and infer them as disease causing	57
Table 3.4A Features of all patients with lipodystrophy in which whole-exome sequencing was performed	67
Table 3.4B Features of all patients with extreme early-onset obesity in which whole- exome sequencing was performed	68
Table 3.4.1 Candidate variants identified in patients with lipodystrophy using whole- exome sequencing	70
Table 3.4.2. Candidate variants identified in patients with extreme early-onset obesity using whole-exome sequencing	71
Table 3.4.5 Status of all CARE for RARE and FORGE Canada cases	74
Table 3.5 Candidate variants identified in the androgen receptor gene	76
Table 3.6.1 Candidate copy-number variants identified in patients with lipodystrophy using the LipidSeq™ targeted gene panel	79

List of Appendices

Appendix I: Ethics Approval	150
Appendix II: Consent Form	151
Appendix III: Primer List	156
Appendix IV: PANTHER gene lists	157

List of Abbreviations

A, additive effects

ABCA1, ATP-binding cassette A1

ABCA4, ATP binding cassette A4

ABCG1, ATP binding cassette subfamily G member 1

ABCG8, sterolin 2

ABI, Applied Biosystems

ACMG, The American College of Medical Genetics and Genomics

AGPAT2, 1-acylglycerol-3-phosphate O-acyltransferase 2

AKT2, AKT serine/threonine kinase 2

ALD, acquired lipodystrophy

ANGPTL3, angiopoietin like 3

ANKRD24, ankyrin repeat domain 24

ANLN, anillin

ANNOVAR, Annotate Variation

APOA1, apolipoprotein A1

APOA5, apolipoprotein A-V

APOB, apolipoprotein B

APOC2, apolipoprotein C-II

APOC3, apolipoprotein C-III

AR, androgen receptor

ASPM, Abnormal Spindle Microtubule Assembly

BAT, brown adipose tissue

BBS10, Bardet-Biedl Syndrome 10

BCYM4, brachyolmia type 4

BDA, brachydactyly type A

BDNF, brain-derived neurotrophic factor

bHLH, basic helix-loop-helix

BLK, B-lymphocyte specific kinase

bp, base pair

BRCA1, breast cancer, DNA repair associated 1

BRCA2, breast cancer, DNA repair associated 2

BSCL2, seipin lipid droplet biogenesis associated

BUBR1, mitotic spindle checkpoint protein BUBR1

CAD, coronary artery disease

CADD, Combined Annotation Dependent Depletion

CARE4RARE, CARE for RARE

CASC5, kinetochore scaffold 1

CAV1, caveolin 1

CDK5RAP2, CDK5 Regulatory Subunit Associated Protein

cDNA, complementary deoxyribonucleic acid

CEL, carboxyl-ester lipase

CEP, centrosomal protein

CFTR, cystic fibrosis transmembrane conductance regulator

CGL, congenital generalized lipodystrophy

ChAS, Chromosome Analysis Suite

CIDEC, cell death inducing DFFA like effector C

CIAP, calf intestinal alkaline phosphatase

cm, centimeter

CMT, Charcot-Marie Tooth

CNV, copy-number variation

co-IP, co-immunoprecipitation

CRE, cyclic AMP response element

CREB3L3, cAMP responsive element-binding protein 3-like 3

DGKD, Diacylglycerol Kinase Delta

DHAP, dihydroxyacetone phosphate

DIDA, Digenic Diseases Database

DNA, deoxyribonucleic acid

DVL3, Dishevelled Segment Polarity Protein 3

E, environment

ER, endoplasmic reticulum

ESP, Exome Sequencing Project

ExAC, Exome Aggregation Consortium

F8, coagulation factor VIII

FBN1, fibrillin 1

Fe-S, iron-sulfur clusters

FH, familial hypercholesterolemia

FHL1, four and a half LIM domains 1

FLNC, gamma filamin

FORGE Canada, Finding of Rare Disease Genes in Canada

FPLD, familial partial lipodystrophy

G, genotype

G3P, glycerol-3-phosphate

GALC, galactosylceramidase

GALNT2, polypeptide N-acetylgalactosaminyltransferase 2

GARP, Golgi associated retrograde protein

GCK, glucokinase

GCKR, glucokinase receptor

gDNA, genomic DNA

gnomAD, Genome Aggregation Database

GPD1, glycerol-3-phosphate dehydrogenase 1

GPIHBP1, glycosylphosphatidylinositol anchored high-density lipoprotein binding protein 1

GSDV, glycogen storage disease type V

GWAS, genome-wide association study

HAART, highly active antiretroviral therapy

HDL, high-density lipoprotein

HEXA, hexosaminidase A

HGMD, Human Gene Mutation Database

HIV, human immunodeficiency virus

HLA, human leukocyte antigen

HNF1A, hepatocyte nuclear factor-1-alpha

HNF1B, hepatocyte nuclear factor-1-beta

HNF4A, hepatocyte nuclear factor-4-alpha

HR, homologous recombination

HSPG2, heparin sulfate proteoglycan 2

HTG, hypertriglyceridemia

IFRD1, Interferon Related Developmental Regulator 1

IHH, indian hedgehog

INS, insulin

ISCU, iron-sulfur cluster assembly enzyme

kb, kilobase

KDM6A, lysine-specific demethylase 6A

KEGG, Kyoto Encyclopedia of Genes and Genomes

KLF11, Kruppel-like factor 11

KMT2C, lysine-specific methyltransferase 2C

LCAT, lecithin-cholesterol acyltransferase

LDL, low-density lipoprotein

LDLR, low-density lipoprotein receptor

LEP, leptin

LEPR, leptin receptor

LFNG, LFNG O-Fucosylpeptide 3-Beta-N-Acetylglucosaminyltransferase

LIPE, lipoprotein E

LMF1, lipase maturation factor 1

LMNA, lamin A

LMNB2, lamin B

LOD, logarithm of the odds

LPL, lipoprotein lipase

LRGC, London Regional Genomics Centre

MADB, mandibuloacral dysplasia with type B lipodystrophy

MAF, minor allele frequency

MAP2K2, mitogen-activated protein kinase kinase 2

MC4R, melanocortin 4 receptor

MCPH, microcephaly, primary hereditary

MDPL, mandibular hypoplasia, deafness, progeroid features and lipodystrophy

MFN2, mitofusin 2

MLXIPL, MLX interacting protein like

MME, MatchMaker Exchange

MODY, mature-onset diabetes of the young

mRNA, messenger ribonucleic acid

MSL, multiple symmetric lipomatosis

MVA, Mosaic Variegated Aneuploidy

NADH, nicotine adenine dinucleotide

NASH, non-alcoholic steatohepatitis

NCBI, National Centre for Biotechnology Information

NEUROD1, neurogenic differentiation 1

NGS, next-generation sequencing

NHEJ, non-homologous end joining

NTD, N-terminal domain

NTRK2, neurotrophic tyrosine kinase receptor type 2

NUDCD2, NudC Domain Containing 2

OMIM, Online Mendelian Inheritance in Man

P, phenotype

PANTHER, protein annotation through evolutionary relationship

PAPSS2, 3'-phosphoadenosine 5'-phosphosulfate synthase 2

PAX4, paired box gene 4

PCNA, proliferating nuclear cell antigen

PCOS, polycystic ovary syndrome

PCR, polymerase chain reaction

PCSK1, prohormone convertase 1

PCSK9, prohormone convertase 9

PDRM16, PR domain containing 16

PDX1, pancreatic and duodenal homeobox 1

PIAS4, protein inhibitor of activated STAT 4

PLIN1, perilipin 1

POLD1, polymerase (DNA-directed) d1

PolyPhen-2, Polymorphism Phenotyping, v2

POMC, proopiomelanocortin

PPAR γ , peroxisome proliferator-activated receptor gamma

PTH, parathyroid hormone

PTRF, polymerase I and transcript release factor

PYGM, glycogen phosphorylase, muscle associated

RBM, reducing-body myopathy

RDS, peripherin 2

RFC, replication factor C

ROM1, retinal outer segment membrane protein 1

rRNA, ribosomal ribonucleic acid

RT-PCR, reverse transcription polymerase chain reaction

RT-qPCR, reverse transcription quantitative polymerase chain reaction

SETD8, SET domain containing (lysine methyltransferase) 8

SIFT, Sorting Intolerant From Tolerant

SIRT6, sirtuin-6

SNP, single nucleotide polymorphism

SNV, single nucleotide variation

SPANR, Splicing Based Analysis of Variants

SULT, sulfotransferase

SV, structural variation

T2D, type 2 diabetes

TE, transposable element

TEAD4, TEA Domain Transcription Factor 4

TG, triglycerides

TMPRSS4, transmembrane protease, serine 4

TRAF7, TNF Receptor Associated Factor 7

TRE, triplet repeat expansion

TRIB1, Tribbles Pseudokinase 1

UTR, untranslated region

VAF, variant allele frequency

VCF, variant calling format

V_G, genetic variation

VLDL, very low-density lipoprotein

VNTR, variable number tandem repeat

VP, phenotypic variation

VPS52, vacuolar protein sorting-52

WES, whole-exome sequencing

WGS, whole-genome sequencing

WRN, Werner

ZBTB33, zinc finger and BTB domain containing 33

ZBTB7A, zinc finger and BTB domain containing 7A

ZMPSTE24, zinc metalloproteinase STE24

Chapter 1 – Introduction

1.1 Rare disease burden in Canada

In Canada, rare diseases are defined as conditions affecting less than one in every 2000 individuals¹. Rare disorders place a significant burden on individuals, their families and the health care system, as there is often a lack of specialized services, social awareness, and clinical recognition. Furthermore, Canadian families with rare diseases face extraordinary challenges, including countless futile medical tests, unnecessary surgeries, financial hardships, social isolation, shortages in treatment options, and early mortality.

In the case of rare diseases, the term “diagnostic odyssey” describes the burden of these diseases on the individual, familial and systemic levels. The diagnostic odyssey details the time between the first concerns of a rare disorder to the time when a diagnosis is made². In the diagnostic odyssey, more than half of the affected individuals with rare diseases are children, with 30% not living to see their fifth birthday. Further, 40% of individuals with rare diseases will receive an incorrect diagnosis, 25% will wait 5–25 years for a diagnosis, and 50% will never be diagnosed in the course of their lifetime³.

Ultra-rare diseases – those limited to a handful of individuals – are often referred to as “orphan diseases”. The term orphan disease describes an exceptionally rare disease historically neglected by the pharmaceutical industry, due to the assumed low financial incentive in devising proper treatments or cures. Because orphan diseases may affect only single individuals, families or communities, these individuals face even greater challenges than those with more prevalent rare diseases. Furthermore, these diseases are often chronically debilitating and life threatening⁴. With the hardships facing the individuals, their families and the health care system, and the lack of treatment options available, there is a pressing need to study rare diseases, and develop novel therapeutic interventions.

1.1.1 Many rare diseases are of genetic origin

To date, there are an estimated 7000 rare diseases. As most of these 7000 rare diseases are predicted to be of genetic origin due to severity and familial recurrence, gene discovery is the critical starting point for the understanding of the molecular mechanisms underlying these diseases, the development of targeted treatments, and the end of the diagnostic odyssey for these patients. Developing therapies is crucial, as fewer than five-percent of the 7000 rare diseases known-to-date offer effective therapeutic interventions⁵.

Previously, rare disease research was limited to standard genetic techniques including homozygosity mapping and candidate gene sequencing⁶. Although these were largely successful in identifying the molecular bases of >3500 rare disorders, limitations including time requirements and high costs have led to the advent of new and improved technologies, and the use of next-generation sequencing (NGS) – deoxyribonucleic acid (DNA) sequencing technologies capable of rapidly generating a massive amount of DNA sequence data – in rare disease studies⁷.

1.1.2 Research problem and motivation

Although the genetic basis of many rare diseases has been clarified using genetic techniques, many rare diseases, rare disease subtypes and orphan diseases remain genetically unexplained. With the success of NGS in elucidating the genetic basis of numerous rare diseases, I propose the use of NGS technologies in rare disease cases not previously subjected to NGS. By using these modern NGS technologies, my goal is to identify genetic determinants underlying several genetically unexplained rare diseases. Understanding these genetic variants and their contribution to disease can increase understanding of disease mechanisms and pave the way for the development of novel treatments in the future.

1.2 Rare human genetic diseases

There are many illustrative examples of rare human genetic diseases, notably lipodystrophy, extreme early-onset obesity, brachydactyly, and autosomal recessive cerebral atrophy. Further, research consortiums such as CARE for RARE (CARE4RARE) and Finding of Rare Disease Genes in Canada (FORGE Canada) collectively pursue research on the molecular basis of rare diseases.

1.2.1 Inherited lipodystrophies

Lipodystrophies are rare disorders characterized by selective and variable loss of adipose tissue, with excess accumulation of fat in otherwise unusual sites including the liver and muscle⁸. This results in a generalized muscular appearance with hypertrophy often arising in the extremities, more specifically, in the calves and arms. The genetic form of lipodystrophy is familial, and includes congenital generalized lipodystrophy (CGL) and familial partial lipodystrophy (FPLD). Affected individuals are predisposed to insulin resistance, which often progresses to other metabolic complications including diabetes, hypertension, metabolic syndrome and hypertriglyceridemia⁹. Furthermore, individuals with lipodystrophy are vulnerable to severe metabolic derangements including acute pancreatitis, hepatic cirrhosis and premature cardiovascular disease¹⁰. Women with severe lipodystrophy often acquire polycystic ovary disease and suffer from infertility, highlighting the aggressiveness of the disorder¹¹. Although the global prevalence is less than one in a million, milder forms of lipodystrophy are commonly misdiagnosed as common central obesity or metabolic syndrome, suggesting that lipodystrophy is an underestimated condition¹². This is concerning as a clear diagnosis of lipodystrophy can allow for earlier care, better management, and improved prognosis.

1.2.1.1 Congenital generalized lipodystrophy

First reported by Berardinelli and Seip in 1954 and 1959, CGL, also known as complete lipodystrophy or Berardinelli-Seip syndrome, is an extremely rare (less than one in 12 million people) disorder characterized by the absence of metabolically active adipose

tissue from birth^{13,14}. Due to the absence of functioning adipocytes, lipids are stored in unusual sites in the body such as the liver and muscle, subsequently leading to a marked muscular appearance.

As the more severe form of the inherited lipodystrophies, individuals with complete lipodystrophy generally present with insulin resistance, rapid aging of bones, hepatic steatosis, hypertriglyceridemia, early-onset diabetes and high blood pressure, often developing into metabolic syndrome. In addition, individuals with CGL acquire hepatomegaly and skeletal muscle hypertrophy; hypertrophic cardiomyopathy is also reported in 25% of individuals and is a common cause of mortality⁸.

To date, there is no cure for CGL. Treatment is through cosmetic procedures, diet, hypoglycemic drugs and lipid-lowering drugs. With regards to medication, the management of CGL is through the medicated treatment of each symptom. Although there has been recent success in CGL treatment with Metreleptin – a leptin replacement that aids in the control of metabolic processes in the body, such as triglyceride and blood sugar management – this drug has yet to be approved in Canada.

1.2.1.2 Familial partial lipodystrophy

The second, milder form of lipodystrophy is FPLD, characterized by the selective loss of adipose tissue from the body (most often the limbs) during childhood or puberty.

Although affected regions may include the face, neck and extremities, the body fat in the intra-abdominal or truncal region are often spared, and some individuals present with excess fat deposition in the face and neck¹⁵. Most individuals presenting with FPLD in the clinic are females, with the metabolic complications more severe in women. Most women remain fertile; however, some women present with hirsutism and polycystic ovary syndrome, leading to fertility issues. Additionally, some patients develop cardiomyopathy, myopathy, and features of multisystem dystrophy^{16,17}.

Similar to CGL, the treatment for FPLD includes lifestyle modifications including diet changes and exercise. Again, the treatment revolves around treating each individual

symptom (i.e. Metformin for diabetes, statins for high cholesterol) rather than the disease as a whole¹⁸.

1.2.2 Extreme early-onset obesity

Extreme early-onset obesity, or excess adiposity, is defined as a body mass index [(weight in kilograms ÷ height in meters)²] greater than or equal to the 85th and 95th percentile for gender and age, respectively. In Canada, childhood-onset obesity is a pressing health concern with the rates having tripled over the last three decades.

Currently, approximately one in seven children are obese, with rates depending on factors such as age, biological sex, location of residence, and socioeconomic status¹⁹. As obesity is a significant risk factor for coronary heart disease, high blood pressure, ischemic stroke and diabetes, it places a major toll on children and their families. This burden is especially apparent in the health care system, as early-onset obesity is associated with increased incidence of adult obesity, hepatic steatosis, and cardiovascular disease. For the first time in history, the complications associated with obesity are leading to a decrease in life expectancy of 2–5 years²⁰.

In the past, obesity was mainly attributed to environmental changes, specifically the intake of high-energy food in combination with a sedentary lifestyle. In recent years, the heritability of obesity has been established, with genetic factors explaining most differences between populations²¹. In fact, obesity risk in children is 10-fold greater when both parents are obese than when both parents are of a normal weight²¹. Aside from lifestyle alterations and medications for the management of associated co-morbidities (i.e. high blood pressure, diabetes), there is no direct treatment or cure for extreme early-onset obesity.

1.2.3 Isolated brachydactyly

Brachydactyly is an inherited malformation characterized by the shortening of the fingers and toes due to deficiencies in the development of the phalanges, metacarpals, and/or metatarsals. Although brachydactyly can be a common feature in complex malformation syndromes, there are many cases of congenital, isolated forms of brachydactyly. These isolated brachydactyly syndromes represent one of the ten categories of hand malformations as classified by Temtamy and McKusick²². The prevalence of isolated brachydactyly is rare, with the exception of types A3 and D, where prevalence varies between populations, reportedly averaging around 2%²³. Aside from these exceptions, most forms of brachydactyly are limited to a few pedigrees. In addition to shortened fingers and toes, other features of brachydactyly include short stature, syndactyly (fused digits), polydactyly (extra digits), and symphalangism (stiff digits) depending on the subtype (Table 1.2.3). Although brachydactyly is best diagnosed using genetic testing, it can also be diagnosed using clinical, anthropometric and radiological means. Treatment for brachydactyly is strictly cosmetic, through elective surgeries.

Table 1.2.3. Clinical features of isolated brachydactyly.

	Subtypes of isolated brachydactyly												Brachymetatarsus	Sugarman brachydactyly	Kirner deformity		
	BDA1	BDA1B	BDA1C	BDA1D	BDA2	BDA3	BDA4	BDB1	BDB2	BDC	BDD	BDE1				BDE2	IV
Gene identified	+		+	+	+			+	+	+	+	+	+				
First digit (thumb) affected	+	+	+	+			+	+			+	+					+
Second digit affected	+	+	+	+	+		+	+	+	+		+					+
Third digit affected	+	+	+	+			+	+	+			+					+
Fourth digit affected	+	+	+				+	+	+			+					+
Fifth digit affected	+	+	+	+	+	+	+	+	+	+		+					+
Proximal phalanges affected	+	+		+	+			+		+							+
Middle phalanges affected	+	+	+	+	+	+	+	+	+	+		+					
Distal phalanges affected		+	+	+			+	+	+		+		+				+
Abnormal phalanx shape/orientation					+	+	+										+
Metacarpals affected	+	+			+			+	+	+		+	+				+
Metatarsals affected								+	+			+		+			
Carpals affected									+								
Tarsals affected									+								
First toe (big toe) affected				+	+			+									+
Second toe affected				+	+		+										
Third toe affected							+	+									
Fourth toe affected							+							+			
Fifth toe affected							+										
Nail dysplasia							+			+							
Symphalangism	+							+	+	+							+
Syndactyly				+				+	+								
Shortened stature	+	+	+				+					+	+				
Learning difficulties													+				

1.2.4 Cerebral atrophy

Neurodegenerative diseases are diseases characterized by the progressive loss of neural cells from the brain. Although some neurodegenerative diseases are more prevalent, such as Alzheimer disease, many neurodegenerative diseases are rare, such as the ataxias or spastic paraplegias^{24,25}. Cerebral atrophy, the loss of neurons and neuronal connections in the brain, is a common feature of many rare neurodegenerative diseases. In 2013, my laboratory described a case of autosomal recessive cerebral atrophy, a novel neurodegenerative disorder in an Old Order Amish family. In this specific case of cerebral atrophy, the affected individuals featured additional symptoms including microcephaly, seizures and psychomotor retardation²⁶. Cerebral atrophy is diagnosed through radiological imaging of the brain.

1.2.5 FORGE Canada and CARE for RARE

In Canada, two Consortiums study the genetic basis of over 200 rare diseases with pediatric onset. These coalitions, FORGE Canada and CARE4RARE, represent the first and second phases of a nationwide project led by Dr. Kym Boycott and Dr. Alex MacKenzie of the Children's Hospital of Eastern Ontario Research Institute.

1.2.5.1 Project overview

The goal of the previous FORGE Canada and current CARE4RARE project was to study the genetic causes of over 200 rare diseases affecting children. In order to do this, the organizers formed a consortium composed of 21 research sites and hospitals, 80 physicians, and 50 scientists. The approach of the project was to use high power DNA sequencing in order to identify novel disease genes. The recruitment of this project, which ended in March 2017, included over 4000 patients with 950 different rare diseases²⁷.

Several criteria required satisfaction in order to qualify for either the FORGE Canada or CARE4RARE Projects. These criteria included the following:

- 1) Unrelated patients and/or families with the same disorder *OR*,
- 2) Single families with multiple affected family members (+/- consanguinity) *AND*,
- 3) Meet the following inclusion sub-criteria:
 - a. Onset is in childhood or adolescence
 - b. The disease is most likely due to a single gene (monogenic)
 - c. DNA is accessible for the proband and available family members
 - d. Clinical material is of high quality

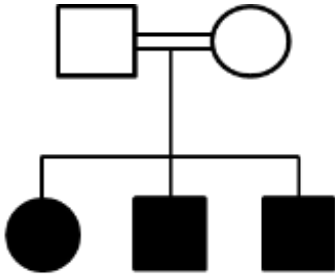
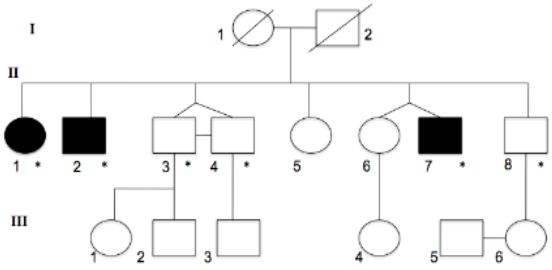
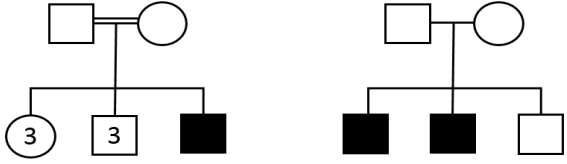
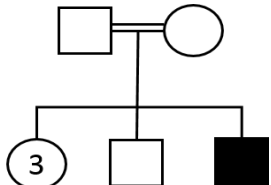
1.2.5.2 Project significance

In sum, the FORGE Canada Project has identified mutations in disease genes for 105 of 362 families. Further, since 2011, the CARE4RARE Project has successfully resulted in the publication of 139 manuscripts on the discovery of 135 novel disease-causing genes²⁷. These results have provided closure and ended the diagnostic odyssey for the individuals and the families through the provision of a diagnosis.

1.2.5.3 FORGE Canada and CARE for RARE cases

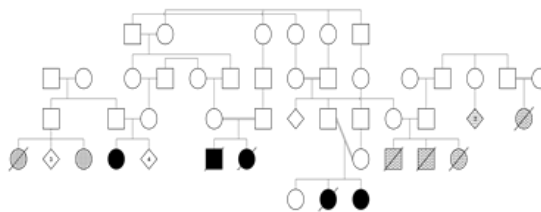
The Blackburn Cardiovascular Genetics Laboratory has been involved in several FORGE Canada and CARE4RARE Projects. There are currently 11 projects in progress, namely: FORGE 214 (Angelman-like syndrome), FORGE 122 (Dystonia, ataxia and intellectual disability), FORGE 11 (Osteopetrosis), FORGE 352 (Microcephaly-ichthyosis), C4R 572 (Autosomal dominant isolated cryptophthalmos), C4R 573 (Widow's peak syndrome), C4R 575 (Amyoplasia), C4R 592 (Metaphyseal chondrodysplasia), C4R 664 (Craniosynostosis), C4R 907 (Heterotaxy disorder), and C4R 924 (Craniorachischisis). The details of each case are listed in Table 1.2.5.3.

Table 1.2.5.3. Clinical description and pedigrees for all CARE for RARE and FORGE Canada cases.

Case	Clinical Description	Pedigree
FORGE 214, Angelman-like syndrome	<ul style="list-style-type: none"> Three children equally affected, parents are healthy first cousins with Iraqi ancestry Features include: significant global developmental delay, delayed speech, repetitive behaviour and similar physical features such as triangular face, wide mouth, pointed chin and fifth finger clinodactyly 	
FORGE 122, Dystonia, ataxia and intellectual disability*	<ul style="list-style-type: none"> Three affected siblings with a progressive, childhood-onset ataxia syndrome Not known to be consanguineous Features include: progressive gait ataxia beginning between ages 2-5, wheelchair-bound by early adolescence and dysarthria Two oldest siblings are non-verbal with limited responsiveness and minimal spontaneous movement 	
FORGE 11, Osteopetrosis	<ul style="list-style-type: none"> Old Order Mennonite and Low-German Speaking Mennonite families Features include: optic atrophy, macrocephaly, mid-facial hypoplasia, micrognathia, abnormal eruption of teeth, broad halluces and short stature 	 

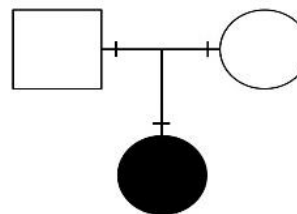
FORGE 352, Microcephaly-ichthyosis*^A

- Old Order Amish families
- Additional features include: failure to thrive, feeding difficulties, developmental delay, lethality in the first year of life, enlarged ventricles and early delivery around 33 weeks gestation
- Recurrence in the family



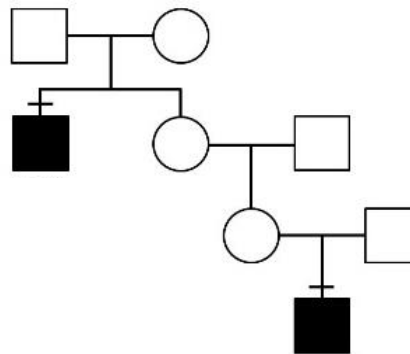
C4R 572, Autosomal dominant isolated cryptophthalmos

- Female child born to non-consanguineous parents
- Female displays complete bilateral isolated cryptophthalmos, a condition resulting in the overgrowth of skin over the eye



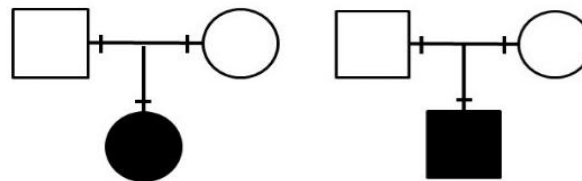
C4R 573, Widow's peak syndrome*

- Non-consanguineous family
- Widow's peak syndrome features: a characteristic widow's peak hairline, delayed development, inability to touch the shoulders, dysmorphic face and the inability to supinate the arms



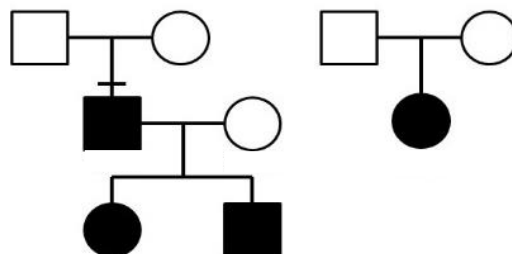
C4R 575, Amyoplasia

- Three unrelated families with amyoplasia, characterized by a generalized lack of muscular development and growth in newborns
- Both probands present with severe muscle hypoplasia



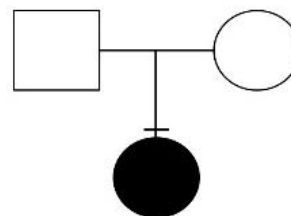
**C4R 592,
Metaphyseal
chondrodysplasia**

- Metaphyseal chondrodysplasia features short stature with abnormally short arms and legs
- No abnormal facial features were present in the patients of interest



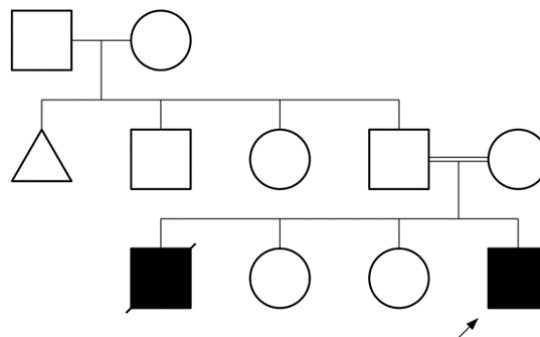
**C4R 664,
Craniosynostosis***

- Non-consanguineous family
- Craniosynostosis is a birth defect resulting in the abnormal development of the skull
- Digital anomalies also present



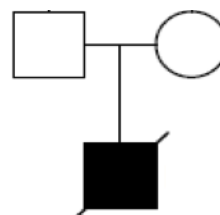
**C4R 907, Heterotaxy
disorder**

- Two affected children born to consanguineous parents of Lebanese descent
- Disorder features: heterotaxy with holoprosencephaly, bilateral ectopic thymus, fused kidneys, annular pancreas, bi-lobed right lung, absent olfactory nerves, dysplastic optic nerves and chiasm and a single umbilical artery



**C4R 924,
Craniorachischisis***

- Stillborn baby presented with craniorachischisis, a severe neural tube defect with anencephaly and spina bifida



For each pedigree, biological males are represented as squares and biological females as circles. Shaded shapes indicate the individual is affected with the disorder. Double lines indicate consanguineous relationships. In some pedigrees, hashed lines represent individuals in which sequencing has been performed. When shown, asterisks represent patients in which DNA is available.

Abbreviations are as follows: FORGE, Finding of Rare Disease Genes Canada project.

*The red asterisk represents the projects I was directly involved in or was working with undergraduate students on. The remaining cases without an asterisk are those in which I assume an administrative role.

[^]Note: for the microcephaly-ichthyosis pedigree, the shaded shapes are individuals clinically diagnosed with microcephaly-ichthyosis, the shapes with vertical curved lines are those with histidyl-tRNA synthetase disorder, the horizontal curved lines are individuals with congenital hypoventilation syndrome, and the shape filled with the plus sign is an individual clinically diagnosed with cystinosis.

1.3 The human genome sequence

In the year 1990, the International Human Genome Sequencing Consortium, composed of public and private researchers, began the publically funded Human Genome Project, a 13-year project with the goal of determining the DNA sequence of the human genome. With the sequencing of over three billion base pairs and approximately 21,000 genes, the group of researchers began a revolution in the fight against human disease by introducing a trove of genetic information applicable for numerous fields including developmental biology, physiology, evolution and medicine^{28,29}.

1.3.1 Variation within the human genome

Human genetic variation refers to the genetic differences within and among populations. A variant form of a gene is an allele, where humans inherit one allele from each parent at each genetic locus, or position. Each pair of alleles represent the genotype of a gene, which can be classified as homozygous or heterozygous, referring to two identical alleles or two different alleles, respectively. Genetic variation is manifested as single nucleotide changes, insertions and deletions, repetitive elements and structural rearrangements of the DNA sequence³⁰. A mutation is a permanent change in the DNA sequence. The term genetic “variant” has been gaining popularity over the term “mutation” in the field, and refers to a benign or pathogenic alteration, or one of unknown significance.

Many decades prior to the availability of DNA sequencing technology or the sequence of the human genome, researchers first became aware of differences in the genome through the study of chromosome number and structure³¹. This led to the advent of molecular biology, a multifaceted field in biology with one branch focused on the structure and function of nucleic acids and proteins essential for life. Research into molecular biology gave way to the understanding of the variation present in the human genome, notably the implications of this variation and relationship to disease. Now, with the sequence of the human genome available, researchers are able to study the impact of variation in the genome, where the variation is composed of single nucleotide changes, insertions and deletions, copy-number variations, repeat variations, and structural variations, all with

potential disruptive effects on the genome. Although some variation can be benign, some defects in the genome can cause genetic diseases, health conditions caused by variation in the genome.

1.3.1.1 Single nucleotide variation

Single nucleotide variations (SNVs) are the most common class of genome variation, with single nucleotide changes occurring approximately once every 500–1000 nucleotide bases in the human genome sequence³². Each SNV represents a change in a single DNA building block or a nucleotide base – composed of Thymine, Adenine, Cytosine and Guanine, often shortened to T, A, C and G – commonly referred to as point mutations. An example of a SNV may be a replacement of a Thymine with a Cytosine (T>C) in a specific stretch of DNA. When these SNVs occur in a region coding for an amino acid (three nucleotides, also referred to as a codon), these SNVs can result in either synonymous (no change in amino acid) or non-synonymous changes (change from one amino acid to another).

The wobble effect is the effect caused by the redundancy of the genetic code. More specifically, each amino acid (with the exception of methionine and tryptophan) may be coded by multiple codons, or multiple three letter codes composed of A, T, C or G. As there are more codons than amino acids, the third base is not as discriminatory for the amino acid as the first two bases. Due to the redundancy in the DNA code, many SNV changes are tolerated or inconsequential. However, in certain cases, these changes can lead to pathogenic missense or nonsense changes, referring to a change in one amino acid to another, or a change in one amino acid to one coding for the truncation of a protein downstream.

Due to the high frequency of these SNVs in the genome and the fact that some SNVs are seen in approximately 1% of the population, SNVs often act as biological markers or contribute to phenotypic traits (observable and measurable features)³³. For example, personal differences in SNPs may cause alterations at the level of the cell, or influence drug metabolism. In some circumstances, SNVs are associated with disease and act by

driving a disease phenotype (the observable characteristics of an individual) or contributing to phenotypic traits. These disease-causing SNVs are most often exonic, non-synonymous SNVs – those in the coding region causing a substitution in amino acid.

1.3.1.1.1 Splicing variation

Splicing is the editing of precursor messenger ribonucleic acid (mRNA) to mature mRNA. During splicing, introns (nucleotide sequences in a gene discarded during splicing) are removed, and exons (the part of the gene that codes for the final mRNA, which is translated into protein) are ligated together. Splice site mutations refer to genetic variants that insert, delete or change the nucleotides in the specific region where splicing takes place. As these variants can result in the retention of intronic DNA in the mRNA or result in the loss of exons from mature mRNA, variants in the splice site can have large effects on human health and disease. In fact, an estimated 15-50% of human disease variants are those that alter splicing elements³⁴. An example of a mutation in a splice site leading to disease is seen in hypoparathyroidism, where a G to C substitution in the first nucleotide base of intron 2 in the parathyroid hormone gene (*PTH*) causes exon skipping (the loss of an exon), and parathyroid hormone deficiency³⁵.

1.3.1.1.2 Insertions and deletions

Another form of variation present at the nucleotide level are micro insertions or deletions affecting 1–50 base pairs, collectively known as “INDELs”. In terms of occurrence, INDELs represent a common alteration in the genome, second only to SNVs³⁶. These INDELs alter the number of DNA bases in a gene by either adding or removing a piece of DNA. When occurring in functionally important sites in the human genome, these INDELs are likely to influence human traits or drive disease³⁶. A notable example is Tay-Sachs disease in the Ashkenazi Jewish, where a four base pair insertion (c.1278insTATC) alters the reading frame of the hexosaminidase A gene (*HEXA*)³⁷. This mutation leads to an infantile form of Tay-Sachs disease where the most common features include deterioration of both mental and physical abilities, and mortality approximately at the age of four. INDELs can either occur inframe when in a multiple of three (deletion or gain in

amino acids) or out-of-frame, where there is a change in amino acids downstream of the insertion or deletion, with a possible inframe occurrence of a stop codon. This is referred to as a frameshift and may alter or abolish the function of the resulting protein. Further, INDELS can result in aberrant splicing when occurring in splice sites.

1.3.1.2 Repetitive elements

The human genome consists of millions of copies of transposable elements (TE) and other repetitive sequences, altogether contributing up to two-thirds of the human genome³⁸. These repetitive elements – most commonly tandem repetitive elements or interspersed repetitive elements that do not encode proteins – are labelled “junk” DNA. Although considered junk or inconsequential, a growing body of evidence has been accumulating in recent years clarifying the role of these repetitive regions in the regulation of gene expression³⁹. Continued research into these repeat sequences is exploiting the role of these elements in gene expression.

Although most repetitive elements in the genome are harboured outside of the coding region, some repeat sequences lie within genes and are subject to mutation. For example, triplet repeat expansions (TREs) – trinucleotide repeats (i.e. CAGCAGCAG) abnormally expanded in certain genes or introns – are dynamic mutations that occur in regions with trinucleotide repeats. These TREs are the cause of TRE disorders such as Fragile X syndrome or Huntington’s disease.

1.3.1.3 Structural variation

Structural variation (SV) affects a region of DNA approximately one kilo base (1-kb) or larger⁴⁰. SVs, including inversions, balanced translocations or genomic imbalances, have long been associated with chromosomal rearrangements and genetic disorders. SVs can arise through various mechanisms including non-allelic homologous recombination (HR), non-homologous end joining (NHEJ), or replication-error⁴¹.

1.3.1.3.1 Copy-number variation

Copy-number variation (CNV) is a form of structural variation where there is a change in copy-number involving a DNA fragment over 1-kb⁴⁰. CNVs can include deletions, insertions or duplications present in variable copy-number in comparison to the reference genome. As CNVs are regions of variable copy number, they may change gene expression levels, and alter transcription levels to values higher or lower than expected in a reference genome. Due to the change in gene expression, CNVs may play a large role in phenotypic variation. Further, since CNVs encompass significantly more nucleotide content per genome than single nucleotide polymorphisms (SNPs; up to 10-fold greater or up to 12% of the genome), CNVs play an important role in genetic diversity, disease and evolution^{42,43}.

1.3.2 Germline versus somatic variation

Germline variation is defined as variation present in the germline, or the tissue involved in forming sex cells. As a mutant sex cell may participate in fertilization, germline variation may be passed on from parent to offspring⁴⁴. In contrast, somatic variation occurs in a single cell in developing somatic tissue. By definition, somatic cells are those that are never transmitted from parent to progeny⁴⁴. As cells are constantly dividing and genetic material is continually replicated over the course of a lifespan, mistakes in replication can accumulate in somatic cells through a lifetime⁴⁵. Although many of these variants do not have a phenotypic effect, some may alter key cellular functions by affecting normal protein function and drive certain human diseases. Early somatic variants – those occurring during embryogenesis – can also contribute to developmental disorders, cancer or aging.

1.3.3 Inherited versus *de novo* variation

Variation can arise in the genome through two mechanisms. Inherited variation are DNA changes transmitted from parent to offspring. Specifically, each egg and sperm cell contribute half of the genetic information needed to produce an offspring. When the egg

and sperm join during fertilization, a new cell is formed, composed of all the necessary genetic information needed to form an individual. This leads to variation passed on from parent to progeny.

On the other hand, *de novo* variation are DNA changes that are spontaneous (through DNA replication errors or spontaneous lesions) or induced (caused by mutagens such as radiation or chemical substances) in an offspring's genome, not inherited from either parent. These *de novo* genetic alterations may be acquired any time post-zygotically (during the lifespan). This variation is *de novo* as the variation is not present in either parental genome.

Both inherited and *de novo* variation may provide meaningful contributions to disease. More specifically, diseases may be caused by inherited variation, or *de novo* variation. An example of inherited variation causing disease is hemophilia A, where multiple generations in a family inherit a variant in the coagulation factor VIII gene (*F8*) and present with a severe bleeding disorder. In contrast, *de novo* variation is best exemplified in severe early-onset pediatric diseases, where the causative mutations only arise post-zygotically given that they are lethal in the germline and/or individuals with these mutations do not reproduce⁴⁶.

1.4 Human genetic diseases

Although there are diseases that occur through bacterial infection or exposure to viruses, it is estimated that most human diseases have a genetic component. A genetic disease is any disease attributed to abnormality in the human genome. These genetic diseases can range from minor to severe and can be due to any of the aforementioned variation. In this section, I describe the heritability of human genetic diseases.

1.4.1 Heritability of genetic diseases

Heritability refers to the ratio of genetic variation (V_G) to phenotypic variation (V_P); a fundamental tool used in the study of the inheritance of human disease that is able to partition the phenotypic variation into both genetic and environmental components⁴⁷. Put

simply, heritability is a statistic that is used to determine the proportion of variation in a trait within a population that can be explained by genetics (rather than environment or random chance). This is modelled through the following equation:

$$\text{Phenotype } (P) = \text{Genotype } (G) + \text{Environment } (E)$$

There are two definitions of heritability⁴⁸. The first, H^2 or broad-sense, refers to the genetic contributions to phenotypic variance through additive (phenotypic contributions from more than one gene or alleles of a single gene, where the sum of the effects of both alleles is equal to the sum of their effect individually), dominance (the relationship between alleles of one gene where the effect of one allele masks the other, leading to a phenotypic effect) and epistatic genetic effects (the phenomenon where one gene relies on other modifier genes, where the sum of the effects in combination do not equal the sum of the effects individually).

$$H^2 = \frac{\text{Var}(G)}{\text{Var}(P)}$$

The second definition, h^2 or narrow sense, captures the proportion of genetic variation due to additive genetic effects (A; contributions from more than one gene or alleles of a single gene acting synergistically).

$$h^2 = \frac{\text{Var}(A)}{\text{Var}(P)}$$

Heritability of disease can also refer to the mode in which transmission occurs. Diseases can be monogenic, digenic or polygenic, and involve variable penetrance, each defined in the following sections.

1.4.1.1 Monogenic diseases

Monogenic diseases, or single gene disorders, are those caused by variants within a single gene. These variants can be inherited in a family; can be dominant or recessive, and autosomal or sex-linked. These modes of inheritance refer to the manner in which genetic traits are passed from one generation to the next. The modes of inheritance include: 1)

autosomal dominant; 2) autosomal recessive; 3) X-linked; 4) Y-linked; 5) codominant or; 6) mitochondrial. Their definitions are as follows:

- 1) Autosomal dominant: when one copy of a gene in a cell harbours a variant. This may be transmitted from one parent to an offspring or inherited *de novo*.
- 2) Autosomal recessive: when both copies of a gene in a cell carry variants. One copy is transmitted paternally and one copy is transmitted maternally to the offspring.
- 3) X-linked: when variants are harboured in genes on the X chromosome(s), one of the two sex chromosomes that can be present in the genome. Note that normally biological males have one X and one Y chromosome, and biological females have two X chromosomes.
- 4) Y-linked: when variants are harboured on the Y chromosome, with transmission only possible from the paternal parent to a male offspring.
- 5) Codominant: when two different alleles of a gene are expressed and each influence the genetic trait.
- 6) Mitochondrial: when genes in mitochondrial DNA carry variants and are transmitted through maternal inheritance.

Monogenic diseases can also be considered Mendelian diseases, defined as diseases where alternative genotypes fall into distinct phenotypic classes, according to Gregor Mendel's laws of inheritance⁴⁹. One example of a monogenic disease is cystic fibrosis, caused by a variant in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*) with an autosomal recessive mode of inheritance.

1.4.1.2 Digenic diseases

Digenic diseases are those where the disease phenotypes are caused by a combination of variants in two genes, or two separate phenotypes caused by variants in two different genes acting synergistically⁵⁰. To cause disease, these genes may interact directly,

indirectly, through a common pathway, through co-expression, or through similar function. Due to the prevalence of digenic diseases, a common resource known as the Digenic Diseases Database (DIDA) is accessible, containing information on over 40 digenic diseases and over 200 digenic combinations⁵¹. The first reported example of a digenic disease was retinitis pigmentosa, where double heterozygotes with variants in the peripherin 2 (*RDS*) and retinal outer segment membrane protein 1 (*ROM1*) loci displayed the disorder⁵².

1.4.1.3 Polygenic diseases

Polygenic diseases are defined as genetic disorders caused by a combination of abnormalities in more than one gene. Due to the fact that they are attributed to variants in more than a single gene, they are inherited differently than monogenic disorders (as described in Section 1.4.1.1). In humans, polygenic diseases are more frequent than monogenic disorders, and have a greater economic burden. Due to the multiple gene (multigenic) nature of these diseases, different methods have been developed to study these diseases. These approaches include: 1) the use of *a priori* data or a set number of genetic loci, or 2) the complete analysis of all available loci⁵³. These may take the form of case-control studies or genome-wide association studies (GWASs), respectively, where case-control studies refer to the comparison of affected and healthy subgroups, and GWASs refer the analysis and determination of similarities in genome-wide data of an affected cohort. Association studies require the selection of several candidate genes in order to carry out association analysis⁵³. Illustrative examples of polygenic disease are the familial dyslipidemias, such as coronary heart disease and diabetes. These are polygenic as they are associated with multiple genes, and are not inherited as simply as single gene disorders⁵⁴.

1.4.1.4 Disease penetrance and expressivity

Disease penetrance and expressivity refer to factors that influence genetic changes. Explicitly, penetrance refers to the proportion of individuals with a genetic change who phenotypically exhibit a genetic disorder. When some individuals with the genetic variant

do not display all or some of the features of a disorder, the condition is said to have incomplete or reduced penetrance. Incomplete penetrance is demonstrated in cases of breast cancer, where variants in the breast cancer, DNA repair associated 1 or 2 genes (*BRCA1* or *BRCA2*) may cause cancer in some individuals over the course of a lifetime, but not others⁵⁵.

On the other hand, expressivity is defined as the degree in which the expression of a trait differs among individuals. This is exhibited when individuals with the same phenotype display varying severity of a disease; or in other words, the disease exists on a spectrum. For example, variable expressivity is seen in Marfan syndrome where some individuals have mild symptoms (tall and thin features with slender fingers), while some experience additional features such as life-threatening heart and blood vessel complications. Although this disorder displays variable expressivity, most individuals with Marfan syndrome carry variants in the fibrillin 1 gene (*FBNI*)⁵⁶.

1.5 Genetic variation and rare diseases

For the rare genetic diseases described earlier – namely, lipodystrophy, extreme early-onset obesity, brachydactyly, autosomal recessive cerebral atrophy, and those ascertained from the CARE4RARE and FORGE Canada consortiums – although many cases of these diseases have been genetically explained with causative variants identified, the genetic etiology of many remain unknown. In this section, I describe the known genetics of these diseases and unexplained cases in my laboratory.

1.5.1 The genetics of lipodystrophy

Over the years, many genetic variants have been associated with lipodystrophy. As the defective gene often causes a distinct lipodystrophy phenotype, both CGL and FPLD have been divided into subtypes.

1.5.1.1 Congenital generalized lipodystrophy

There are four subtypes of CGL based on previously identified genes; denoted as CGL1–4. CGL commonly follows an autosomal recessive mode of inheritance. Although each subtype features the complete loss of adipose tissue, different subtypes present with distinct characteristics⁵⁷:

- 1) CGL1: characterized by the absence of metabolically active adipose tissue from birth, CGL1 is associated with variants in the 1-acylglycerol-3-phosphate O-acyltransferase 2 gene (*AGPAT2*), a gene encoding an enzyme required for phospholipid and triglyceride biosynthesis.
- 2) CGL2: featuring the lack of adipose tissue from birth, mild mental disability and cardiomyopathy, CGL2 is associated with variants in the seipin lipid droplet biogenesis associated gene (*BSCL2*) encoding seipin, a protein that may play a role in adipocyte differentiation.
- 3) CGL3: reported in a single patient, CGL3 features the absence of body fat, shortened stature and vitamin D resistance. CGL3 is correlated with variants in the caveolin 1 gene (*CAVI*) encoding the caveolin 1 protein, an integral component of caveolae on adipocyte membranes.
- 4) CGL4: featuring the absence of body fat with congenital myopathy and cardiomyopathy, CGL4 is associated with variants in the polymerase I and transcript release factor gene (*PTRF*; also referred to as cavin) associated with the biogenesis of caveolae.

Lipid droplets are defined as organelles responsible for the storage of triglycerides. Playing a fundamental role in metabolism, the reduction or lack of lipid droplets cause the lipodystrophies⁵⁸. With each form of CGL, although each are attributed to distinct genes, all gene products are involved in adipocyte lipid droplet formation. Further, each form of CGL is associated with low leptin (the hormone of energy balance) as a result of the lack of adipose tissue. This results in excess triglycerides and high blood sugar.

1.5.1.2 Familial partial lipodystrophy

As in the case of CGL, FPLD also exists in various forms (FPLD1–5) with subtypes defined based on their associated gene(s). Generally, FPLD follows an autosomal dominant mode of inheritance. Similar to CGL, each subtype features a similar pattern in adipose tissue loss but various additional characteristics⁵⁷:

- 1) FPLD1: a genetically unexplained form of FPLD, characterized by the loss of subcutaneous fat from the extremities.
- 2) FPLD2: a form of FPLD featuring the loss of subcutaneous fat from the extremities and truncal region from puberty. FPLD2, or Dunnigan variety, is associated with variants in the lamin A gene (*LMNA*) encoding the lamin A protein, where deleterious mutations result in the premature death of adipocytes.
- 3) FPLD3: characterized by the loss of subcutaneous fat from the extremities, principally from the distal regions. FPLD3 is associated with variants in the peroxisome proliferator-activated receptor gamma gene (*PPAR γ*), encoding a transcription factor required for adipogenesis.
- 4) FPLD4: reported in a single family, FPLD4 features the loss of subcutaneous fat from the extremities. This subtype of FPLD4 is associated with variants in the AKT serine/threonine kinase 2 gene (*AKT2*) encoding protein kinase B, involved in adipocyte differentiation and insulin receptor signaling.
- 5) FPLD5: a subtype of FPLD featuring fibrosis of adipose tissue and the loss of subcutaneous fat from the extremities. FPLD5 is attributed to variants in the perilipin 1 gene (*PLIN1*) encoding an integral constituent of lipid droplet membranes, involved in lipid storage. FPLD5 has also been associated with homozygous variants the cell death inducing DFFA-like effector C gene (*CIDEA*), encoding a protein required for lipid droplet formation and energy storage in adipose tissue.

1.5.1.3 Unexplained cases of lipodystrophy

With the familial lipodystrophies, although many causative gene mutations have been identified in affected individuals as described (i.e. *LMNA*, *PPARG*, *AGPAT2* and *BSCL2* variants), many individuals remain genetically unexplained^{59,60}. Over the years, the Blackburn Cardiovascular Genetics Laboratory (Robarts Research Institute, London ON, Canada) has been successful in identifying the genetic basis of 120 lipodystrophy-affected patients using Sanger sequencing. Although this resulted in 44% of 267 lipodystrophy patients characterized by mutations in lipodystrophy-causing genes, 56% of patients remained genetically unexplained after Sanger sequencing.

1.5.2 The genetics of obesity

To date, there have been many genes identified that underlie monogenic and polygenic obesity. The rare, monogenic forms of human obesity have been critical in the elucidation of the pathogenic mechanisms causing obesity. These rare forms of monogenic obesity are associated with the following genes: *LEP*, *LEPR*, *POMC*, *PCSK1*, *MC4R*, *BDNF*, and *NTRK2*, where all aforementioned genes and protein products are involved in body weight management (described in detail in Table 1.5.2)⁶¹⁻⁶⁷. Further, all genes encode either hormones or neurotransmitters/receptors belonging to the highly conserved leptin-melanocortin pathway, a critical pathway in the regulation of food intake and body weight.

In contrast to the monogenic forms of obesity where obesity is the main feature, the syndromic forms of early-onset obesity are Mendelian disorders with obesity as one clinical feature. The additional clinical features may include intellectual disability, dysmorphic countenances and developmental abnormalities, generally organ-specific. To date, over 30 syndromes have been identified, with Prader-Willi, Bardet-Biedl, Cohen, and Alström syndrome as classic examples⁶⁸⁻⁷¹.

Table 1.5.2. Genes involved in the pathophysiology of obesity.

Gene name	Symbol	Function
Leptin	<i>LEP</i>	Encoding the hormone leptin, this hormone is involved in the regulation of body weight. Leptin works by attaching to the leptin receptor on the surface of organs and tissues, including the hypothalamus. The hypothalamus controls states including hunger and thirst, and the release of hormones through the body. The binding of leptin in the hypothalamus triggers a feeling of satiety.
Leptin receptor	<i>LEPR</i>	The leptin receptor is a Type I cytokine receptor that binds with the hormone leptin. The leptin receptor is responsible for the regulation of energy metabolism and body weight through interactions with gangliosides in the cell membrane.
Proopiomelanocortin	<i>POMC</i>	The <i>POMC</i> gene encodes the proopiomelanocortin protein, which is cleaved into small peptides in the body. These peptides then bind to several proteins in the body to trigger signaling pathways responsible for functions including energy homeostasis, melanocyte stimulation and immune modulation.
Prohormone convertase 1	<i>PCSK1</i>	Encoding a member of the subtilisin-like proprotein convertase family, this protein is involved in the secretory pathway. More specifically, it functions in the proteolytic activation of polypeptide hormones and neuropeptide precursors.
Melanocortin 4 receptor	<i>MC4R</i>	The <i>MC4R</i> gene encodes a membrane-bound receptor belonging to the melanocortin receptor family. This protein interacts with melanocyte-stimulating hormones and adrenocorticotrophic hormones and plays a role in energy homeostasis and somatic growth.
Brain-derived neurotrophic factor	<i>BDNF</i>	This gene encodes a protein found in the brain and spinal cord that plays a role in growth, differentiation, and maintenance of nerve cells. The BDNF protein is localized in brain regions responsible for controlling eating, drinking and body weight.
Neurotrophic tyrosine kinase receptor type 2	<i>NTRK2</i>	The <i>NTRK2</i> gene encodes a membrane-bound receptor kinase that phosphorylates both itself and members of the mitogen-activated protein kinase pathway. Signaling through NTRK2 results in cell differentiation.

1.5.2.1 Unexplained cases of early-onset obesity

As in the case of the lipodystrophy patients in the laboratory, the prevalence of genetically unexplained cases is also reflected in the early-onset obesity cohort, where known obesity-causing genes, *POMC*, *LEP*, *LEPR*, *PCSK1* and *MC4R*, cannot explain the phenotype of five patients and/or families with extreme early-onset obesity ascertained in the Blackburn Cardiovascular Genetics Laboratory (Robarts Research Institute, London ON, Canada)²⁰. In some instances, these patients display variable co-morbidities (i.e. hyperphagia, intellectual disability) in addition to extreme early-onset obesity.

1.5.3 The genetics of brachydactyly

Most isolated brachydactyly syndromes display an autosomal dominant mode of inheritance, with variable penetrance. To date, several gene mutations have been designated as causative for each subtype (Table 1.5.3). Here, I describe subtype brachydactyly type A.

Table 1.5.3. The genetics of isolated brachydactyly.

The genetics of isolated brachydactyly							
Name	Subtype (Abbreviation)	OMIM	Synonym(s)	Gene Symbol	Gene Locus	Mutations	Reference
Brachydactyly type A	Brachydactyly type A1 (BDA1)	112500	Farabee type brachydactyly	<i>IHH</i>	2q35	p.E95G p.E95K,p.D100E, p.E131K p.D100N p.R128Q, p.T130N p.T154I p.R158C c.283_285delGAG p.Glu95_Asn96insLys	Kirkpatrick et al. (2003) Gao et al. (2001) McCready et al. (2002) Byrnes et al. (2009) Liu et al. (2006) Stattin et al. (2009) Lodder et al. (2008) Ho et al. (2018)
	Brachydactyly type A1, B (BDA1B)	607004					
	Brachydactyly type A1, C (BDA1C)	615072		<i>GDF5</i>	20q11.22	p.R399C	Byrnes et al. (2010)
	Brachydactyly type A1, D (BDA1D)	616849		<i>BMPR1B</i>	4q22.3	c.447-1G>A, p.K325N	Racacho et al. (2015)
	Brachydactyly type A2 (BDA2)	112600	Brachymesophalangy II; Mohr-Wriedt type brachydactyly	<i>BMPR1B</i>	4q22.3	p.I200K, p.R486Q, p.R486W p.R486L	Lehmann et al. (2003) Badura-Stronka et al. (2015)
				<i>BMP2</i>	20q12.3	dup 4663 bp ~110 kb downstream dup 4671 bp downstream dup 5547 bp and 5895 bp, ~110 kb downstream	Su et al. (2011) Liu et al. (2014) Dathe et al. (2009)
				<i>GDF5</i>	20q11.22	p.R380Q	Ploger et al. (2008)
		Brachydactyly type A3 (BDA3)	112700	Brachymesophalangy V; Brachydactyly-clindodactyly			
		Brachydactyly type A4 (BDA4)	112800	Brachymesophalangy II and V; Temtamy type brachydactyly			
	Brachydactyly type B	Brachydactyly type B1 (BDB1)	113000		<i>ROR2</i>	9q22.31	p.W749*, p.Y755*, c.2249delG p.S758* p.Q760*, c.1321_1325delCGGCG, c.1398dupA, c.1386+3_1386+5delCTCinsTACATT GTGTAAGCACAAA c.1397_1398delAA c.2244delC
Brachydactyly type C	Brachydactyly type B2 (BDB2)	611377		<i>NOG</i>	17q22	p.PE5A, p.A36P, p.R167G, p.P187S	Lehmann et al. (2007)
	Brachydactyly type C (BDC)	113100	Brachydactyly, Haws type	<i>GDF5</i>	20q11.22	p.M173V p.L176P p.T201P, p.L263P p.T203N, p.V286M p.S204R, p.C498S, c.158delT, c.830delT, c.157dupC, ins 23 bp nt 811 p.R301*, p.R438C, c.122delG, c.760delG, c.205dupG p.L373R, p.Y487* c.631+1G>T c.498delC c.498dupC	Schwabe et al. (2004) Farooq et al. (2013) Stange et al. (2014) Al-Qattan et al. (2015) Everman et al. (2002) Polinkovsky et al. (1997) Yang et al. (2008) Carli et al. (2013) Galijaard et al. (2001) Holder-Espinasse et al. (2004)
Brachydactyly type D	Brachydactyly type D (BDD)	113200	Stub thumb	<i>HOXD13</i>	2q31.1	p.S316C, p.I322L	Johnson et al. (2003)
Brachydactyly type E	Brachydactyly type E1 (BDE1)	113300		<i>HOXD13</i>	2q31.1	p.R274*	Jamsheer et al. (2012)
	Brachydactyly type E2 (BDE2)	613382		<i>PTHLH</i>	12p11.22	p.L44P, p.L60P, p.K120*, p.*178W, delKlopocki et al. (2010) ~907 kb incl. gene and 6 others c.101+3_101+6delAAGT, del 128 bp c.47_101+73	Thomas-Teinturier et al. (2015)
Brachymetatarsus IV		113475	Metatarsus IV, short; Toe, fourth, short				
Sugarman brachydactyly		272150	Brachydactyly with major proximal phalangeal shortening				
Dystelephalangy		128000	Kirner deformity				

1.5.3.1 Brachydactyly type A

As described by Bell in 1951, the hallmark feature of brachydactyly type A is the shortening of the middle phalanges⁷². Type A has been further divided into four subtypes, A1 to A4, with each subtype further subdivided into types A to D.

1.5.3.1.1 Brachydactyly type A1 (BDA1)

In the early 1900s, Farabee and Drinkwater characterized the first disorder that could be explained by Mendelian autosomal dominant inheritance – brachydactyly. In 1963, Haws and McKusick continued the analysis on Farabee's family and associated the condition to BDA1⁷³. In BDA1, the middle phalanges of all the digits are absent or shortened, with shortening of the thumb and big toe. In some instances, the middle and terminal phalanges are fused (terminal symphalangism, or fusion of the terminal phalanges), with shortening of the metacarpals. Drinkwater found that the shortened phalanges were due to the absence of the epiphyses, which subsequently lead to overall shortened stature in adulthood^{22,74,75}. BDA1 has been attributed to mutations in the Indian Hedgehog gene (*IHH*), where mutations affect endochondral ossification of the developing bones⁷⁶⁻⁸³.

1.5.3.1.2 Brachydactyly type A1, B (BDA1B)

Armour *et al.* described a family with a mild form of brachydactyly that did not fit the clinical findings of other families with BDA1⁸⁴. In their family, Armour and colleagues described shortened middle and terminal phalanges, with all the digits slightly shortened. The feet of this family were also affected, with the children displaying multiple coned epiphyses. From this research, Armour *et al.* highlighted the benefits of using a metacarpophalangeal profile in designating clinical status. In 2002, Armour *et al.* performed a linkage study, and mapped the condition to the cytogenetic bands 5p13.3-p13.2⁸⁵. However, the causative gene and putative mutations have not yet been identified.

1.5.3.1.3 Brachydactyly type A1, C (BDA1C)

Type A1C was first described by Byrnes and colleagues in a consanguineous French-Canadian family⁸⁶. In this subtype of BDA1, individuals displayed markedly shortened middle phalanges of the second through fifth digits, and the first terminal phalanx. There was also shortening in the first metacarpals resulting in the appearance of a shortened thumb. The third to fifth metacarpals were slightly affected, and the feet were affected in this family. BDA1C is attributed to a mutation in Growth Differentiation Factor 5 (*GDF5*), associated with a decreased effectiveness of chondrogenesis and abnormal skeletal morphogenesis⁸⁷. BDA1C is inherited in an autosomal recessive or dominant manner, with recessive mutations causing a more severe phenotype⁸⁶.

1.5.3.1.4 Brachydactyly type A1, D (BDA1D)

In 2015, Racacho *et al.* described patients with BDA1D, a “BDA1-like” brachydactyly⁸⁸. In the first proband, features included short second digits, curved fifth digits, and shortened first and second toes. Upon radiographic analysis, it was found that there was absence of the middle phalanges in the index fingers, with shortened middle phalanges in the fifth digits. The terminal phalanges of the thumbs were also affected. In the second proband, brachydactyly and clinodactyly of the fifth digits and right thumb were reported, with abnormally-shaped middle phalanges of the second digit and fifth digit. This condition was designated as BDA1-arachnodactyly. Type BDA1D has been associated with mutations in Bone Morphogenetic Protein Receptor Type 1B (*BMPRI1B*)⁸⁸. Where the p.K325N mutation acts in a dominant-negative manner. Mutations in *BMPRI1B* result in brachydactyly by affecting normal chondrogenesis.

1.5.3.1.5 Brachydactyly type A2 (BDA2)

BDA2 is characterized by the absence or shortening of the middle phalanx of the index finger and second toe. Rarely, this feature can extend to the fifth digit. In severe cases, the middle phalanx is a rhomboid or triangular shape causing the index finger to deviate radially. BDA2 is inherited in an autosomal dominant manner, with mutations identified

in Bone Morphogenetic Protein 2 (*BMP2*), *BMPRI1B*, and *GDF5*⁸⁹⁻⁹⁴. The reported families are of German, Norwegian, Swiss, Chinese and Polish descent²².

1.5.3.1.6 Brachydactyly type A3 (BDA3)

In BDA3, the most common features include the shortening and slanting of the middle phalanx of the fifth digit. Due to the characteristic rhomboid or triangular shape of the middle phalanx, clinodactyly of the fifth finger is present²². The prevalence of BDA3 is quite common, often seen in syndromic forms. The frequency of BDA3 can vary from 3.4% to 21% in various, isolated ethnic populations²². Although there are multiple definitions of BDA3, Hertzog defined BDA3 as requiring the middle phalanx of the fifth finger to be less than half the size of the middle phalanx of the fourth digit⁹⁵. Williams *et al.* also found that cone-shaped epiphyses in BDA3 are twice as likely in females than males, suggesting a strong sex-bias in BDA3 inheritance⁹⁶. Although there is no candidate locus or gene identified to date, BDA3 is inherited in an autosomal dominant manner with variable penetrance²².

1.5.3.1.7 Brachydactyly type A4 (BDA4)

In 1978, Temtamy and McKusick studied a family with four generations of BDA4²². The features of this family included the shortening of the second and fifth digits due to brachymesophalangy – an abnormal shortening of the middle phalanges. Although the fourth finger was not always affected, it displayed an irregularly shaped middle phalanx resulting in radial deviation of the terminal phalanx when afflicted. The feet were also affected²². BDA4 has been reported multiple times in the literature with added features including nail dysplasia, clinodactyly, short stature, and a bifid appearance of the thumb⁹⁷⁻⁹⁹. No causative locus or gene for this sub-phenotype has been identified.

1.5.3.1.8 Unexplained cases of brachydactyly

To date, various forms of brachydactyly have been characterized and numerous causative genes have been found (Table 1.5.3). Although there is much known about brachydactyly, many subtypes remain genetically undefined. We introduce a 31-year-old

female from Ontario referred to Dr. Hegele's Endocrinology Clinic (London Health Sciences Centre, University Hospital, London ON, Canada) for medical consultation for a lifelong history of short stature and shortening of all digits of the upper and lower extremities. There was no history of other medical conditions, specifically no nystagmus, musculoskeletal abnormalities, developmental delay, or scoliosis. Upon examination, all digits of upper and lower extremities were noted to be short, with missing middle phalanges and splayed fingers, similar to a brachydactyly type A phenotype. The proband's parents were both of short stature, but only her mother shared identical features including very short fingers and toes. Eight maternal aunts and one uncle had all reported abnormally shortened fingers of variable severity. One brother and her daughter had similarly shortened digits (Figure 1.5.3.1.8).

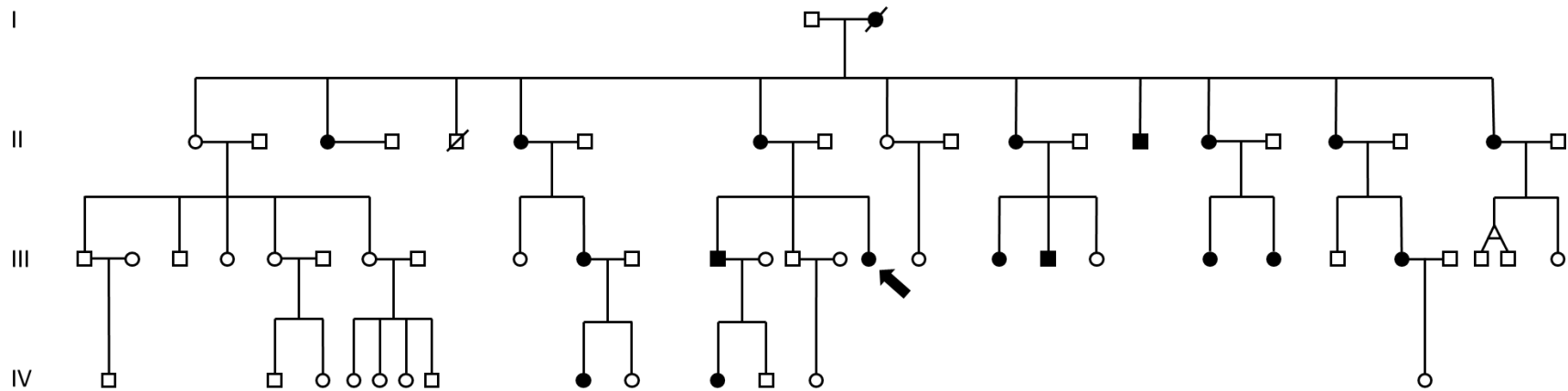


Figure 1.5.3.1.8. Pedigree of a Canadian autosomal dominant brachydactyly-affected family. The black arrow indicates the proband. All shaded shapes represent those clinically affected. Biological females are represented by circles and biological males are represented by squares.

1.5.4 The genetics of autosomal recessive cerebral atrophy

In 2013, our laboratory (Blackburn Cardiovascular Genetics Laboratory) elucidated the genetic basis of a novel pediatric neurodegenerative disorder deemed, autosomal recessive cerebral atrophy²⁶. Using a combination of homozygosity mapping and exome sequencing, Lahiry *et al.* identified four individuals with a c.995C>T (p.Thr332Met) variant in the *TMPRSS4* gene encoding transmembrane protease, serine 4, a protein involved in numerous biological processes²⁶.

1.5.4.1 Unexplained cerebral atrophy

Recently, a second family with a history of disease resembling autosomal recessive cerebral atrophy was referred to the Blackburn Cardiovascular Genetics Laboratory from India (Figure 1.5.4.1). In this case, the proband presented with microcephaly, severe cortical atrophy and widened lateral ventricles. The parents were both clinically unaffected with normal brain structure and function. The sister of the proband was also affected but passed away prior to genetic testing.

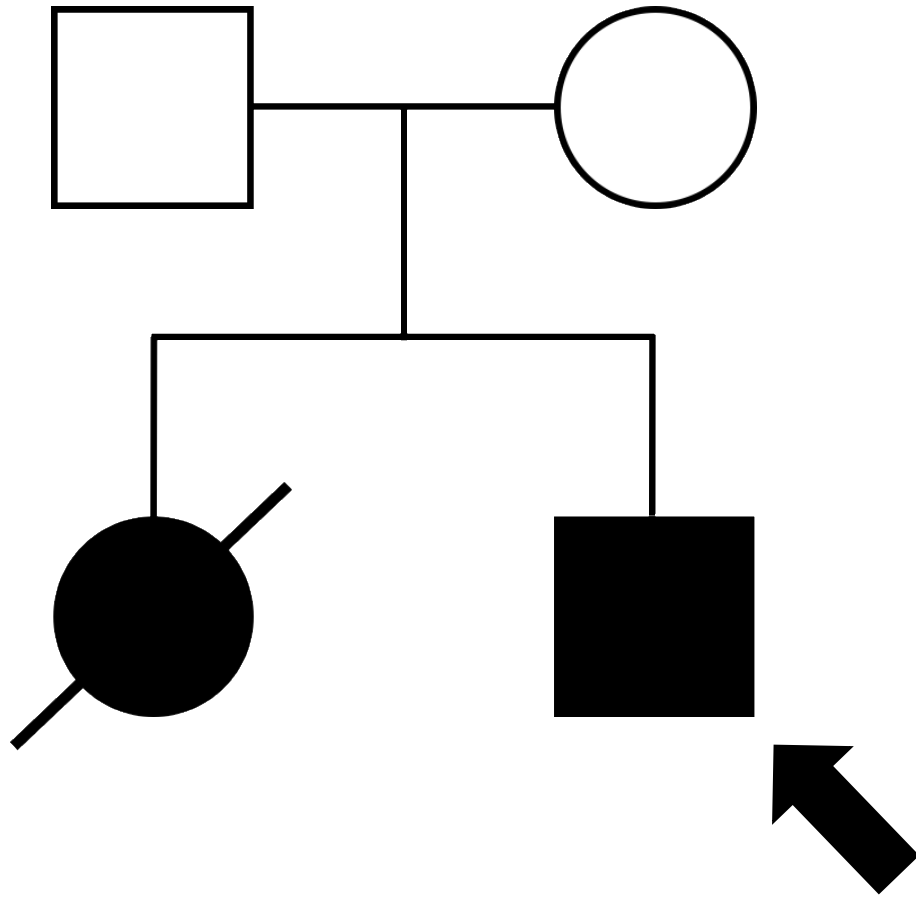


Figure 1.5.4.1. Pedigree of a family with cerebral atrophy. Biological males are squares and biological females are circles. The arrow indicates the proband, whereas the shaded shapes represent those affected by the cerebral atrophy, microcephaly and widened lateral ventricles.

1.5.5 Genetically unexplained consortium cases

The Blackburn Cardiovascular Genetics Laboratory is actively working on 11 FORGE Canada and CARE4RARE Projects. The rare diseases in the 11 families previously listed are all genetically unexplained and require a genetic diagnosis. By using NGS technologies, I seek to identify genetic determinants underlying all the genetically unexplained rare disease cases ascertained in the Blackburn Cardiovascular Genetics Laboratory.

1.6 Next-generation sequencing and rare diseases

Since the 1970s, Sanger sequencing has been the gold standard in genetic studies with several historical accomplishments, notably the sequencing of the first human genome²⁹. However, limitations including time requirements and high costs have led to the advent of new and improved technologies, designating Sanger sequencing as first generation sequencing¹⁰⁰. Over the past five years, we have been in a new era of genomics, where NGS methods have become mainstay in genomic research. NGS is an umbrella term used to describe all the newer methods, capable of generating a colossal amount of data rapidly and efficiently¹⁰⁰. It includes three dominant methods used to identify genetic variants, namely targeted resequencing, whole-exome sequencing (WES), and whole-genome sequencing (WGS).

Today, many platforms exist to perform NGS. Although each company offers a distinct method, the general workflow is as follows: 1) template preparation, 2) sequencing and imaging, 3) genome alignment and assembly, and 4) genome enrichment. Regardless of the company chosen, each method provides a rapid and inexpensive genome-wide sequence readout as an output, and allows for characterization of evolutionary relationships and the elucidation of the DNA sequence for studies in health and disease¹⁰¹.

1.6.1 Targeted resequencing

Targeted resequencing is a method of NGS that works by using NGS chemistry to isolate and sequence any number of pre-selected genes from a sample library. With the ability to rapidly sequence target genes, these targeted gene panels are able to identify causative variants in genomic loci. Targeted resequencing is one of the fastest growing applications for NGS technologies.

1.6.1.1 LipidSeq™

An excellent example of a NGS clinical resequencing panel is LipidSeq™, a targeted gene panel designed by, and used in, the Blackburn Cardiovascular Genetics Laboratory¹⁰². The objective of the pilot LipidSeq™ project was to develop a NGS-based method to facilitate in the clinical diagnosis of dyslipidemias. Using Illumina technologies, this pilot study on LipidSeq™ demonstrated a 95% concordance rate with the gold standard, Sanger sequencing, highlighting its diagnostic utility in the clinic. To date, over 3,000 DNA samples have been processed using LipidSeq™.

Currently, the LipidSeq™ panel targets ~73 genes and >180 SNPs associated with monogenic dyslipidemias. The LipidSeq™ targets genes involved in lipid metabolism as well as genes associated with lipodystrophy and its associated metabolic comorbidities (Table 1.6.1.1).

Table 1.6.1.1. Details regarding the genes targeted by the LipidSeq™ gene panel relating to lipodystrophy and associated metabolic comorbidities.

Phenotype	Gene	Function	RefSeq (NM_)	Chromosome: Start-Stop/ Location
Lipodystrophy	<i>LMNA</i>	<i>LMNA</i> encodes lamin A and lamin C proteins, which contribute to the structural network of the inner nuclear membrane of cells.	170707	1:156052369-156109880/ 1q22
	<i>PPARG</i>	The <i>PPARG</i> gene encodes peroxisome proliferator-activated receptor-gamma (PPAR γ). This is a nuclear receptor, which regulates adipocyte differentiation.	015869	3:12329349-12475855/ 3p25.2
	<i>LMNB2</i>	<i>LMNB2</i> encodes lamin B, a component of the structural network of nuclear membranes in vertebrates. It is important for mitotic processes, nuclear stability, chromatin structure, and gene expression.	032737	19:2428163-2456966/ 19p13.3
	<i>CIDEA</i>	<i>CIDEA</i> is a DNA fragmentation factor (DFF). In adipocytes, the encoded protein contributes to lipid droplet formation and plays a major role in promoting apoptosis.	0011996 23	3:9908394-9921938/ 3p25.3
	<i>AKT2</i>	<i>AKT2</i> , an oncogene, encodes a serine/threonine kinase with Src Homology 2-like domains. It is highly expressed in insulin-sensitive tissues and regulates hepatic metabolism during fasting.	001626	19:40736224-40791443/ 19q13.2
	<i>CAVI</i>	<i>CAVI</i> encodes caveolin-1, the main component in the caveolae plasma membranes of most cell types. Caveolin-1 is an integral membrane protein, which links integrin subunits to proto-oncogene tyrosine kinase, FYN, promoting cell cycle progression.	001753	7:116164839-116201239/ 7q31.2
	<i>AGPAT2</i>	<i>AGPAT2</i> encodes a 1-acylglycerol-3-phosphate O-acyltransferase protein, which converts lysophosphatidic acid to phosphatidic acid during phospholipid biosynthesis. This protein is localized to the ER.	006412	9:139567595-139581911/ 9q34.3

	<i>BSCL2</i>	<i>BSCL2</i> encodes seipin, a transmembrane protein that is activated during pre-adipocyte differentiation and contributes to lipid droplet morphology. Seipin is localized to the ER.	0011229 55	11:62457734- 62477091/ 11q12.3
	<i>PTRF</i>	<i>PTRF</i> encodes cavin, a protein that regulates rRNA transcription and plays an essential role in caveolae biogenesis. Caveolae, in turn, are crucial for signal transduction and lipid transport in many cell types.	012232	17:40554467- 40575506/ 17q21.2
	<i>LIPE</i>	<i>LIPE</i> encodes a hormone-sensitive lipase, which exists in two forms, depending on the translational start codon. Present in steroidogenic tissues, the longer form contributes to cholesterol processing for steroid hormone production. Expressed in adipose, the shorter form hydrolyzes TGs, thereby regulating the mobilization of free fatty acids.	005357	19:42905664- 42931578/ 19q13.2
Werner's Syndrome	<i>WRN</i>	The <i>WRN</i> gene encodes the Werner protein, which shows dominant nucleolar localization, possesses DNA helicase activity (3' to 5') and also acts as an exonuclease (3' to 5'). This protein may also be involved with DNA double strand break repair.	000553	8:30890778- 31031277/ 8p12
Mature Onset Diabetes of the Young (MODY)	<i>HNF4A</i>	Hepatocyte nuclear factor-4-alpha (HNF4A) is a transcription factor belonging to the nuclear receptor family. HNF4A is the most prominent DNA-binding protein in the liver, where it is known to regulate genes involved in lipid metabolism. The encoded protein also controls the expression of genes including hepatocyte nuclear factor 1 alpha. This gene may also play a role in liver, kidney and intestine development.	000457	20:42984441- 43061485/ 20q13.12
	<i>GCK</i>	<i>GCK</i> is expressed in the liver and pancreatic islet beta cells and plays a significant role in glucose metabolism. Glucokinase (GCK) is a unique member of the hexokinase family where, in metabolic pathways, hexokinases	033508	7:44183870- 44229022/ 7p13

	phosphorylate glucose to produce glucose-6-phosphate.		
<i>HNF1A</i>	Hepatocyte nuclear factor-1-alpha (<i>HNF1A</i>) or HNF1 homeobox A is thought to aid in the orderly and sequential activation of genes during development at the level of transcription. HNF1 protein is a transcription factor that binds to a variety of promoters in the liver and is required for the expression of several liver-specific genes.	000545	12:121415861-121440315/12q24.31
<i>PDX1</i>	Pancreatic and duodenal homeobox 1 (<i>PDX1</i>) acts as a transactivator that binds the TAAT element in the promoter region of target genes involved in pancreas development. The protein encoded by <i>PDX1</i> is a transcriptional activator of several genes and plays a significant role in glucose-dependent regulation of insulin gene expression.	000209	13:28494168-28500451/13q12.2
<i>HNF1B</i>	Hepatocyte nuclear factor-1-beta (<i>HNF1B</i>) is a transcription factor (otherwise known as transcription factor-2) that is a member of the homeodomain-containing superfamily of transcription factors. This protein can bind as either a homodimer or a heterodimer in conjunction with HNF1A. This gene has been shown to take part in nephron development and regulates the development of the embryonic pancreas.	000458	17:36046434-36105096/17q12
<i>NEUROD1</i>	Neurogenic differentiation 1 (<i>NEUROD1</i>) is a basic helix-loop-helix (bHLH) protein, which acts as a transcription factor involved in determining cell type during development. <i>NEUROD1</i> (after heterodimerization with the ubiquitous helix-loop-helix protein E47, regulates insulin gene expression by binding to a critical E-box motif on the insulin promoter.	002500	2:182540833-182545392/2q31.3
<i>KLF11</i>	Kruppel-like factor 11 (<i>KLF11</i>) encodes a zinc finger transcription factor that binds to specificity protein 1-like sequences in both epsilon- and gamma-globin gene	003597	2:10170776-10194963/2q25.1

		promoters. Their binding inhibits cell growth and results in apoptosis.		
	<i>CEL</i>	Carboxyl-ester lipase (<i>CEL</i>) is a major component of pancreatic juice and is in charge of the hydrolysis of cholesterol and other dietary esters. The protein encoded by <i>CEL</i> promotes large chylomicron production in the intestine and is also thought to interact with cholesterol and oxidized lipoproteins to modulate the progression of atherosclerosis.	001807	9:135936741-135947250/ 9q34.2
	<i>PAX4</i>	Paired box gene 4 (<i>PAX4</i>) is a member of the PAX family of transcription factors. This gene plays an important role during fetal development and cancer growth. In addition, it has been found that <i>PAX4</i> is essential for the differentiation of insulin-producing beta cells in the mammalian pancreas.	006193	7:127250346-127255982/ 7q32.1
	<i>INS</i>	Consisting of two dissimilar polypeptide chains, A and B, and linked by two disulfide bonds, insulin is synthesized by beta cells of the islets of Langerhans. Insulin is under the control of a single gene locus and binds the insulin receptor, stimulating glucose uptake.	0011850 97	11:2181009-2182439/ 11p15.5
	<i>BLK</i>	B-lymphocyte specific kinase (<i>BLK</i>) encodes a non-receptor tyrosine kinase. These genes are typically involved in cell proliferation and differentiation. This protein stimulates insulin synthesis and secretion in response to glucose levels and enhances expression levels of several pancreatic beta-cell transcription factors.	001715	8:11351521-11422108/ 8p23.1
Hypertriglyceridemia (HTG)	<i>LPL</i>	<i>LPL</i> is responsible for the breakdown of chylomicron triglycerides into two fatty acids and a monoacylglycerol along the luminal surface of capillaries in the heart, skeletal muscle, and adipose tissue. It functions as a homodimer and requires APOC2 as a cofactor.	000237	8:19796582-19824770/ 8p22

<i>APOC2</i>	APOC2 is a lipid-binding protein secreted in the plasma as a component of chylomicrons and VLDLs. It is essential for lipoprotein lipase activation and TG breakdown activity.	000483	19:45449239-45452822/ 19q13.2
<i>APOC3</i>	Apolipoprotein C-III is a component of HDL and ApoB containing lipoprotein particles. It impairs hepatic uptake and catabolism of ApoB-containing lipoproteins.	000040	11:116700624-116703787/ 11q23.3
<i>APOA5</i>	APOA5 is secreted in the plasma as a component of HDL, and to a lesser extent, VLDL and chylomicrons. It is a strong stimulator of LPL-mediated TG hydrolysis and an activator of LPL.	052968	11:116660086-116663136/ 11q23
<i>LMF1</i>	The LMF1 protein resides in the ER and is vital for LPL maturation and transport through the secretory pathway. It is therefore essential for functioning circulating LPL in the blood.	022773	16:903634-1031318/ 16q13.3
<i>GPIHBP1</i>	LPL and chylomicrons bind to <i>GPIHBP1</i> in the capillary endothelial cells of the heart, skeletal muscle and adipose tissue, making it an important component of chylomicron TG catabolism.	178172	8:144295068-144299044/ 8q24.3
<i>GCKR</i>	<i>GCKR</i> is expressed in the liver and pancreatic islet beta cells and plays a significant role in glucose metabolism. It is a unique member of the hexokinase family where, in metabolic pathways, hexokinases phosphorylate glucose to produce glucose-6-phosphate.	033508	7:44183870-44229022/ 7p13
<i>CREB3L3</i>	cAMP responsive element-binding protein 3-like 3 (<i>CREB3L3</i>) is a transcription factor specific to the liver. The encoded protein is localized to the ER and is activated by cyclic AMP stimulation. The encoded protein binds the cyclic AMP response element (CRE) and box-B element.	032607	19:4153598-4173051/ 19p13.3
<i>TRIB1</i>	By controlling the differentiation of tissue-resident M2-like macrophages, <i>TRIB1</i> is considered critical for adipose tissue	025195	8:126442563-126450647/ 8q24.13

		maintenance and the suppression of metabolic disorders.		
	<i>GPD1</i>	<i>GPD1</i> encodes an important protein involved in lipid metabolism. The encoded protein catalyzes the conversion of dihydroxyacetone phosphate (DHAP) and reduced nicotinic adenine dinucleotide (NADH) to glycerol-3-phosphate (G3P) and NAD ⁺ .	005276	12:50497602-50505103/12q13.12
	<i>GALNT2</i>	<i>GALNT2</i> encodes a membrane-bound glycosyltransferase protein that catalyzes the first step of O-glycosylation of peptides in the Golgi.	004481	1:230202956-230417875/1q42.13
	<i>MLXIPL</i>	<i>MLXIPL</i> encodes a leucine zipper transcription factor, which belongs to the Myc/Max/Mad family. In a glucose-dependent manner, it activates the promoters of genes coding for TG synthesis.	032951	7:73007524-73038903/7q11.23
Low high-density lipoprotein (LHDL)	<i>ABCA1</i>	<i>ABCA1</i> is part of the ATP-binding cassette (ABC) superfamily of transporters. <i>ABCA1</i> is a cholesterol efflux pump, which removes cholesterol from cells.	005502	9:107543283-107690527/9q31.1
	<i>APOA1</i>	Apolipoprotein 1 is synthesized in the liver and small intestines. It is the main protein of plasma HDL, promotes cholesterol efflux for excretion and acts as a cofactor to LCAT.	000039	11:116706467-116708338/11q23-q24
	<i>LCAT</i>	Lecithin-cholesterol acyltransferase (LCAT) esterifies cholesterol with a lecithin molecule, reducing its solubility and aiding in the formation of pseudo-HDL/LDL/VLDL. LCAT has a high affinity for HDL and plays an important role in the formation of HDL cholesterol.	000229	16:67973787-67978656/16q22.1
	<i>ANGPTL3</i>	Angiotensin-like 3, expressed predominantly in the liver, is in a class of lipid metabolism modulators that regulate VLDL levels in the plasma. It acts as a dual inhibitor of both lipoprotein lipase and endothelial lipase.	014495	1:63063158-63071976/1p31.1-p22.3

	<i>ABCG1</i>	<i>ABCG1</i> encodes an ATP-binding cassette transporter protein which regulates macrophage cholesterol and phospholipid transport. It also contributes to cholesterol homeostasis in several cell types.	004915	21:43619799-43724497/21q22.3
High low-density lipoprotein (HDL)	<i>APOB</i>	Apolipoprotein B is a hydrophilic protein on chylomicrons, thereby allowing for transport of LDL to the tissues of the body. ApoB-48 is synthesized in the intestines and ApoB-100 is synthesized in the liver.	000384	2:21224301-21266945/2p24-23

Abbreviations: HDL, high density lipoprotein; VLDL, very low density lipoprotein; LPL, lipoprotein lipase; TG, triglyceride; ER, endoplasmic reticulum; LDL, low density lipoprotein; rRNA, ribosomal RNA.

1.6.2 Whole-exome sequencing

WES is a method of NGS capable of sequencing all protein-coding regions of the genome (approximately 1–2%). Using a library preparation kit that captures all exonic content of the genome (known as the exome); WES allows identification of all gene variants, rather than the select few possible in the scaled-down approach of targeted resequencing. WES is a feasible and rapid approach to sequencing a large amount of DNA, and a valuable tool for rare disease research.

1.6.2.1 Whole-exome sequencing successes

WES has had momentous success as a tool for rare disease variant discovery. Although it only targets a significantly small proportion of the human genome, it stands as the method-of-choice for rare disease research because 85% of disease causing mutations found-to-date have been harboured in exons¹⁰³. Further, the use of WES has had much success in identifying the genetic basis of rare diseases, with the identification of causative genes for both Schinzel-Giedion syndrome and Miller syndrome as notable examples of the first cases of WES success^{104,105}. In sum, the genetic basis of >200 disorders have been identified using WES, making the success and efficiency of WES preferred over WGS^{100,106}. Moreover, due to the declining cost of WES, the high diagnostic yield, and the general acceptance of WES in the health marketplace, it has become the preferred diagnostic method for rare disease patients on the diagnostic odyssey who have not been genetically explained using standard genetic techniques¹⁰⁷.

1.6.3 Whole-genome sequencing

The third and most comprehensive form of NGS is WGS, a method reporting almost every nucleotide (~85%) in the human genome. As WGS targets most nucleotides, it is capable of determining variations in any part of the genome, including regulatory regions. These variations detected through WGS have significant implications on an individual's phenotype, ranging from protective against disease, to causative of disease, to neutral or benign. In recent years, WGS has become a widely used tool in cancer research and the

analysis of somatically acquired variants in tumour tissues¹⁰⁸. Despite the utility of WGS, there are many caveats to the widespread implementation of WGS. For example, factors including misdiagnosis, risk of incidental findings, and the gap in literature on non-exonic regions are all major considerations¹⁰⁸. Because of these drawbacks, WES remains the preferred sequencing method for rare disease research.

1.7 Thesis project outline

With a staggering number of rare diseases without a genetic explanation, there is a substantial need to investigate these rare disorders and end the diagnostic odyssey for these individuals and their families. In the Blackburn Cardiovascular Genetics Laboratory, we have been referred patients with a variety of rare diseases including lipodystrophy, extreme-early onset obesity, brachydactyly, cerebral atrophy, and those through the FORGE Canada and CARE4RARE Projects. Generally, when familial recurrence of a rare phenotype is present, the likelihood that the disease is monogenic is high. Further, extreme severity of a disease with high penetrance also suggests the possibility of a single gene explanation. Therefore, it is predicted that these rare diseases are attributed to novel, single gene, genetic variations. Through NGS methods such as LipidSeq™ and WES, I seek to identify the molecular basis of each rare disorder.

1.7.1 Project overview

Overall, we enrolled 153 patients with lipodystrophy, 5 patients with extreme early-onset obesity, one family with brachydactyly, one family with cerebral atrophy, and 11 families associated with FORGE Canada or CARE4RARE. To start, I began by using the LipidSeq™ targeted gene panel for the lipodystrophy cohort, and Sanger sequencing for all other rare disease samples to determine if known genes could explain these disorders. In the cases where no mutations were found from Sanger sequencing or LipidSeq™, I sought to determine the genetic basis of these unexplained diseases using WES and subsequent bioinformatics analysis. In the event that candidate variants were identified, I performed variant validation and co-segregation analysis. Further, when possible, I used NGS data to identify novel CNVs.

1.7.2 Project hypothesis

Given the evidence of a genetic basis for lipodystrophy, extreme early-onset obesity, brachydactyly, autosomal recessive cerebral atrophy, and past CARE4RARE/ FORGE Canada cases, I hypothesize that these aforementioned diseases are caused by rare genetic variants with functional relevance, following monogenic inheritance. Furthermore, I hypothesize that these variants will co-segregate with disease status in families, and be absent from unaffected family members and control populations.

1.7.3 Project goal and aims

Using LipidSeq™ and WES, the overall goal was to utilize these NGS technologies to identify a molecular basis for each patient/family disease of interest. In order to do this, my aims were as follows:

- 1) To use NGS methods and bioinformatics processing, namely LipidSeq™ and WES, to detect genetic variants or CNVs.
- 2) To identify rare or uncommon variants that are likely causative or associated with the disease phenotype of interest.
- 3) To determine the functional relevance of these genetic variants using *in silico* tools and literature reviews.
- 4) To perform variant validation studies to resolve whether these newly discovered variants are co-segregating with the disease phenotype in the family or an unaffected control population.

Chapter 2 – Materials and Methods

2 Overview

The following section provides an overview of all the Materials used and Methods performed in this study (Figure 2). Although the study relied on several external factors, including clinical diagnoses, the processing of the LipidSeq™ targeted gene panel and WES; I was involved in patient selection and developing eligibility criteria, and responsible for all sample preparation, bioinformatics processing and data analysis, variant validation and co-segregation analysis. Most importantly, I was responsible for all variant prioritization, consisting of literature review and analysis of hundreds-to-thousands of variants per case. Furthermore, I was responsible for performing all analytics, reporting of the data, and manuscript preparation. For the CARE4RARE and FORGE Canada cases, I was significantly involved in the microcephaly-ichthyosis and widow's peak cases (responsible for patient selection, bioinformatics processing and analysis, variant validation, and co-segregation analysis). For the remaining CARE4RARE and FORGE Canada cases, I was responsible for managing each case, organizing and chairing meetings with CARE4RARE/FORGE Canada collaborators, mentoring each student involved, and re-analyzing and confirming all results identified by students in these cases.

2.1 Ethics approval

Ethics approval was obtained for all projects in accordance with the Ethics Review Board at Western University (Certificate Number 07920E) in London, Ontario, Canada (Appendix I). For each study, informed consent and ethical approval was obtained from each participant, or in the event of a pediatric case, from a parent or guardian (the consent form can be found in Appendix II).

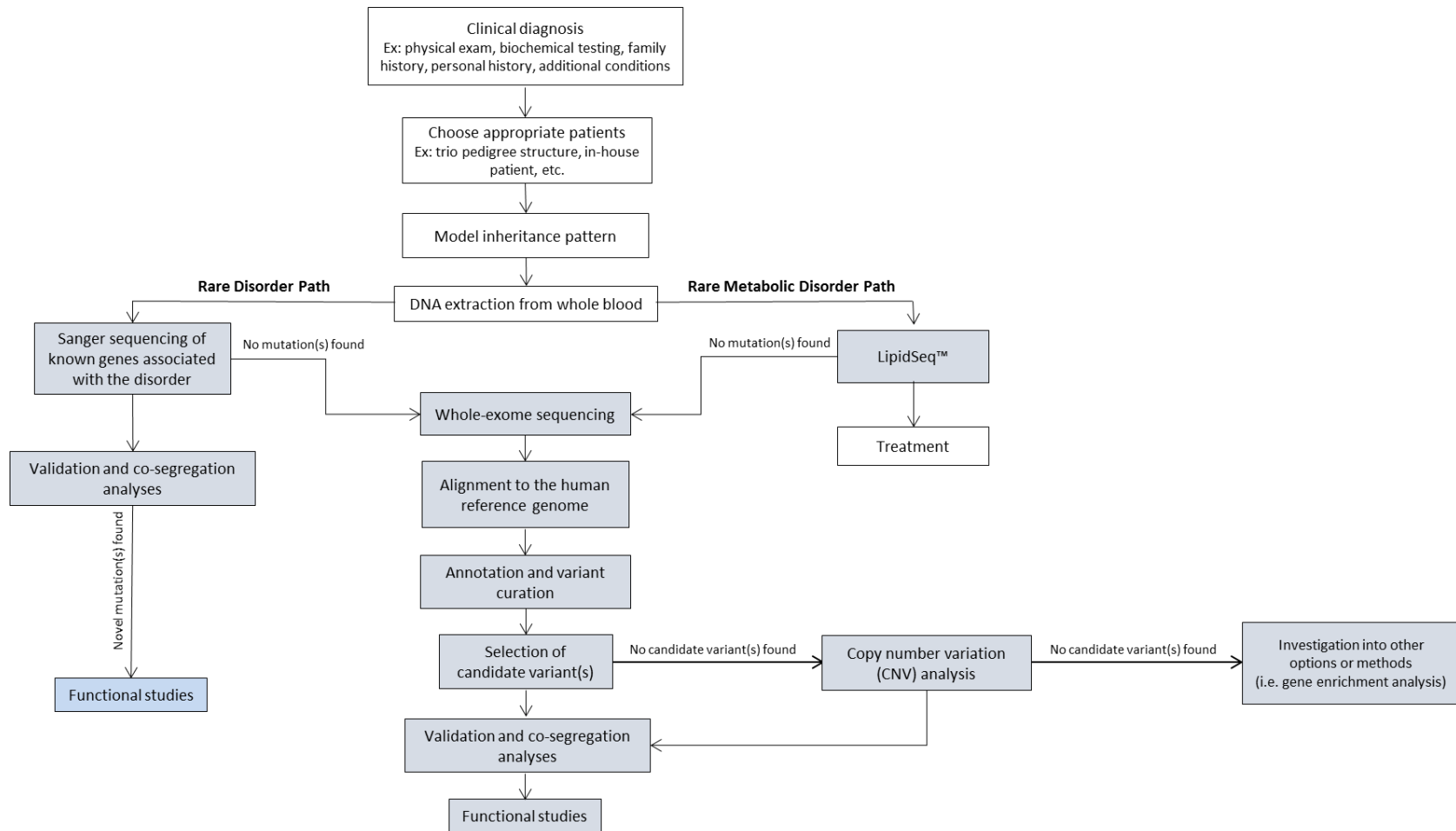


Figure 2. Flowchart of steps taken in order to identify the genetic bases of rare disorders. All steps shaded in blue represent steps where I was involved.

2.2 Study subjects

All study participants were recruited from Dr. Hegele's Endocrinology Clinic (London Health Sciences Centre, University Hospital, London, ON, Canada), the Blackburn Cardiovascular Genetics Laboratory (Robarts Research Institute, London, ON, Canada), the Medical Genetics Program of Southwestern Ontario (London Health Sciences Centre, Victoria Hospital, London, ON, Canada), and the Children's Hospital of Western Ontario (London Health Sciences Centre, Victoria Hospital, London, ON, Canada). For the CARE4RARE and FORGE Canada projects, patients were recruited from areas throughout Canada.

2.2.1 Clinical information

For each study subject, clinical information was provided from Dr. Hegele or the referring physician. For some CARE4RARE and FORGE Canada cases, clinical features were obtained from the CARE4RARE and FORGE Canada Consortia. In the event that detailed clinical information was unavailable, the most common features and co-morbidities of the diagnosed disease were assumed present.

2.3 DNA extraction and quality control

DNA was isolated from whole blood (3 mL) using the Puregene® Blood Extraction Kit (Cat. No. 158389, Gentra Systems, Qiagen Inc., Mississauga, ON, Canada). Both the quality and quantity of the DNA sample were measured using the NanoDrop 1000 spectrophotometer with target ratios of 260:280 and 260:230 of approximately 1.8 meeting the threshold for the study. DNA was purified and diluted to 5 ng/μL for both targeted and WES.

2.4 SNP genotyping and autozygosity mapping

For certain CARE4RARE and FORGE cases, namely CARE4RARE 907 (Heterotaxy) and FORGE 11 (Osteopetrosis), 122 (Dystonia, ataxia and intellectual disability) and 352 (Microcephaly-ichthyosis), genomic DNA (gDNA) was extracted and genotyped for

SNPs using the Affymetrix® Genome-Wide Human SNP Array 6.0 (Santa Clara, California, United States). With the data as input, GeneSpring GT v2.0 (Agilent Technologies, Santa Clara, California, United States) software was used to identify areas of homozygosity that were identical-by-descent within family members. The SNP allele frequencies of healthy individuals were used to calculate a logarithm of the odds (LOD) score for each SNP. Homozygous regions in affected individuals with high LOD scores were prioritized when identifying candidate variants.

2.5 Next-generation sequencing

Two methods of NGS were used for the study. One technique used was targeted sequencing, namely LipidSeq™, and the second was WES, discussed in Sections 2.5.1 and 2.5.2, respectively.

2.5.1 Patient selection for LipidSeq™

One-hundred-and-fifty-three patients referred to the Blackburn Cardiovascular Genetics Laboratory with a clinical diagnosis of lipodystrophy were included in this portion of the study. Patients were prioritized based on two criteria: 1) the absence of mutations in known lipodystrophy-causing genes as previously determined from Sanger sequencing; and 2) disease severity, where those with severe phenotypic characteristics (i.e. the partial or complete absence of adipose tissue from the body) and metabolic comorbidities (i.e. insulin resistance, diabetes, hypertriglyceridemia) were selected.

2.5.1.1 LipidSeq™ targeted sequencing

After dilution and purification to 5 ng/μL, all DNA samples were indexed and pooled using the Nextera® Rapid Capture Custom Enrichment Kit, namely the “LipidSeq™” targeted sequencing design. Using LipidSeq™, 73 genes and >180 SNPs associated with lipid metabolism were sequenced for each sample. More specifically, all coding regions and a ~150-base pair (bp) pad at intron-exon boundaries were sequenced. Following indexing and pooling, library preparation was performed with the Nextera® Rapid Capture Custom Enrichment Kit, and enriched DNA samples were sequenced on an

Illumina MiSeq personal desktop sequencer (Illumina, San Diego, California, United States) using 2×150-bp paired-end chemistry in accordance to manufacturer instructions. DNA samples were sequenced in-house at the London Regional Genomics Centre (LRGC; <http://lrgc.ca/>) in London, Ontario with an average of approximately 300× depth-of-coverage.

Briefly, the extracted and purified DNA was first tagged with adapters. Reduced cycle amplification added additional motifs including unique primer binding sites, indices, and complementary regions to the oligonucleotides on the flow cell. The library was then quantified and measured using the Agilent® 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, United States) prior to being processed on the Illumina MiSeq. Once on the MiSeq, the DNA was immobilized onto a glass flow cell using the oligonucleotide complementary adapters. Bridge amplification occurred where all the strands were clonally amplified. This process occurred in parallel to millions of reactions, and resulted in cluster generation and amplification of all the fragments. After bridge amplification, sequencing was performed by way of sequencing-by-synthesis. More specifically, each nucleotide base possessed a unique fluorescence and competed for addition to the template strand, and the MiSeq instrument detected the fluorescent light emitted during synthesis. This resulted in complete sequencing of the DNA in a base-by-base fashion¹⁰⁹.

2.5.2 Patient selection for whole-exome sequencing

Twenty patients with lipodystrophy and five children with extreme early-onset obesity from Dr. Hegele's Endocrinology Clinic were selected for this portion of the study. These patients were chosen based on the following criteria: 1) the absence of mutations in known disease-associated genes as determined from Sanger sequencing or LipidSeq™, and 2) disease severity, where those with severe characteristics (i.e. complete absence of adipose tissue or extreme excess of adiposity) and several confounding conditions (i.e. diabetes, insulin resistance or overgrowth) were prioritized. From the group of lipodystrophy patients deemed mutation negative from Sanger sequencing or LipidSeq™, patients of Dr. Hegele's (rather than referrals from external physicians) were prioritized

for WES to ensure consistent clinical diagnoses and tests were performed. The parents of the children with extreme early-onset obesity were also approached for participation in the study. This resulted in the recruitment of two sets of parents, or two sets of trios (Mother, Father and proband) to the study.

Families with rare genetic diseases (namely brachydactyly, type A1 and cerebral atrophy) referred to Dr. Hegele's clinic or the Blackburn Cardiovascular Genetics Laboratory, and cases ascertained through the CARE4RARE and FORGE Canada projects were also selected for this portion of the study. These families were chosen based on similar criteria: lack of mutations in known disease-associated genes, disease severity, and rarity.

2.5.2.1 Whole-exome sequencing

After DNA extraction, quality control and dilution as described previously in Section 2.3, the DNA samples were indexed and pooled using the Illumina TruSeq Rapid Capture Exome Library Prep Kit in preparation for WES. The patient samples were sequenced in accordance to the WES protocol on the Illumina NextSeq 500 (Illumina, San Diego, California, United States) benchtop sequencer with 2×100-bp paired-end chemistry in accordance to the protocol used at the LRGC. Again, sequencing was performed using the sequencing-by-synthesis method, as described previously in Section 2.5.1.1. The average depth-of-coverage was approximately 100×.

The lipodystrophy samples, all early-onset obesity samples, the autosomal recessive cerebral atrophy-like trio, and the brachydactyly proband were sequenced in-house at the LRGC. For the remaining cases, WES was generously funded by the FORGE Canada and CARE4RARE Research Programs, and performed at either the Genome Québec Innovation Centre (<http://gqinnovationcenter.com/>) in Montreal, Québec or The Centre of Applied Genomics (<http://www.tcag.ca/>) in Toronto, Ontario. At both locations, the samples were run on an Illumina HiSeq 2000 (Illumina, San Diego, California, United States) sequencer using an Agilent® SureSelect v5 Target Enrichment Kit.

2.6 Bioinformatics processing

Following sequencing, both the LipidSeq™ and WES raw data were returned in the form of patient-specific, paired FASTQ files. FASTQ files contain FASTQ format, a text-based format with nucleotide sequence and corresponding quality scores. The paired FASTQ files were input into CLC Bio Genomics Workbench v10 (CLC Bio, Aarhus, Denmark) to align the data to the human reference genome (build Hg19). Sequence alignment was performed using a custom automated workflow (Figure 2.6).

Succinctly, the workflow began by mapping the reads to the reference genome. From there, the software performed local realignment, a “proofreading” step that used alignment information of other nearby reads in order to re-align each given read. A second “proofreading” step was then executed, where the tool effectively removed bias by discarding duplicate reads that may have arose from PCR amplification during sample preparation. Fixed ploidy variant detection was performed, using The Error Model estimation in order to estimate the probability of having a certain base in a read, but calling a different base. This tool allowed variants to be called or identified. The software then performed two steps: 1) filtered based on overlap, or kept the reads overlapping with the BED file (tab-delimited text file with information on the regions targeted using the LipidSeq™ or WES kit); and 2) created statistics for target regions, which provided a summary of the local and overall coverage. After the alignment and variant calling steps, variant data were exported in the form of a variant calling format or VCF (.vcf) file (a tab-delimited text file containing the variant information), and coverage statistics were exported in the form of a BAM file (.bam), a binary version of a SAM file (a tab-delimited text file containing the sequence alignment data).

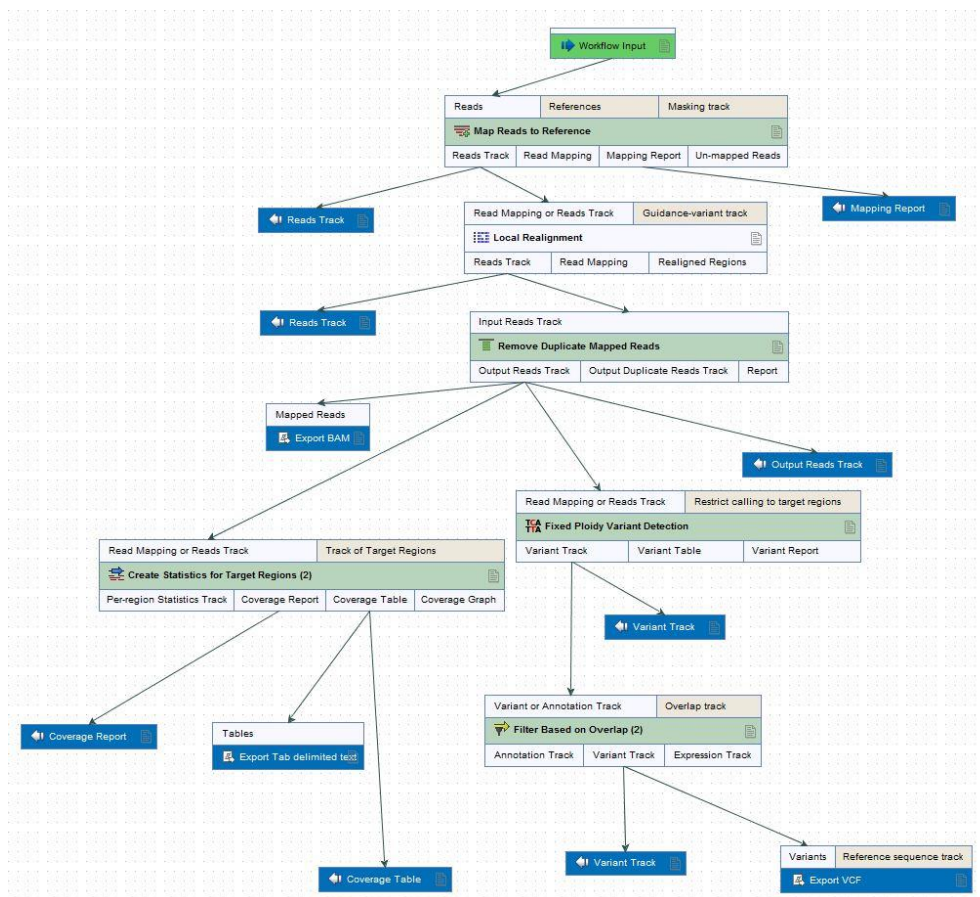


Figure 2.6. Next-generation sequencing bioinformatics workflow. Custom workflow designed in CLC Genomics Workbench v10. For each individual case, the workflow uses two batched FASTQ files and generates a VCF and BAM file as output. Note that the “Filter Based on Overlap” step utilizes the BED file for the LipidSeq™ or whole-exome sequencing kit.

2.7 Variant annotation

After variant calling, the VCF files were utilized as input for variant annotation. For the LipidSeq™ data, ANNOVAR (<http://annovar.openbioinformatics.org/>), a software tool that functionally annotates genetic variants, was used to annotate the variants within the VCF files using a customized script. For the WES data, the Golden Helix VarSeq® software (Golden Helix Inc., Bozeman, Massachusetts, United States), a variant annotation and filtration software, was used to fully annotate the variants. ANNOVAR and VarSeq® both annotated each variant with predictions from *in silico* tools including: Polymorphism Phenotyping, v2 (PolyPhen-2; <http://genetics.bwh.harvard.edu/pph2/>)¹¹⁰, Sorting Intolerant From Tolerant (SIFT) predictor (<http://sift.jcvi.org/>)¹¹¹, MutationTaster (<http://www.mutationtaster.org/>)¹¹², and Combined Annotation Dependent Depletion (CADD; <http://cadd.gs.washington.edu/score>)¹¹³. For splicing mutations, concordance with *in silico* pathogenic predictions from the Human Splicing Finder (<http://www.umd.be/HSF3/HSF.shtml>)¹¹⁴ and the Splicing Based Analysis of Variants (SPANR; <http://tools.genes.toronto.edu/>)¹¹⁵ databases were used. For each variant, the minor allelic frequency (MAF) was also extracted from the following databases: 1000 Genomes (<http://www.internationalgenome.org/1000-genomes-browsers/>)¹¹⁶, Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>), Exome Aggregation Consortium (ExAC) browser (<http://exac.broadinstitute.org/>)¹¹⁷, and Genome Aggregation Database (gnomAD) (<http://gnomad.broadinstitute.org/>)¹¹⁷. Annotated variants harboured within genes associated with the disease of interest were prioritized.

2.8 Variant prioritization

Once annotated, variants were prioritized based on several criteria for causality. These stringent criteria were used to ensure the candidate variants identified for each case had a high likelihood of being disease-causing (Table 2.8). Note that given the complexities of, and differences between each case, some variant prioritization criteria varied between patient sample or family (i.e. autosomal dominant versus recessive mode of inheritance, minor allele frequency [MAF] of >1% versus 0%, etc.), and these criteria were customized for each case.

Table 2.8. Criteria used to prioritize candidate variants and infer them as disease causing.

Criteria considered to predict a variant as disease-causing:

- I. The variant is predicted to be pathogenic by the following analyses:
 - A. Multiple *in silico* prediction tools agree that the variant has a high likelihood of being disease-causing
 - B. Gene function, expression, and pathway analyses are in line or associated with the phenotype of interest
Using the following databases, but not limited to: GeneCards (<http://www.genecards.org/>)¹¹⁸, OMIM (<https://www.omim.org/>), NCBI Gene (<https://www.ncbi.nlm.nih.gov/gene/>), UniProt (<http://www.uniprot.org/>)¹¹⁹, Reactome Pathway Database (<http://reactome.org/>)¹²⁰, Human Protein Atlas (<http://www.proteinatlas.org/>)¹²¹, KEGG (<http://www.genome.jp/kegg/>)¹²²
 - C. Disease databases show a relation between the gene or variant and the disorder of interest
Using the following databases, but not limited to: HGMD (<http://www.hgmd.cf.ac.uk/ac/all.php>)¹²³, NCBI ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), Orphanet (<http://www.orpha.net/consor/cgi-bin/index.php>)¹²⁴, DECIPHER (<https://decipher.sanger.ac.uk/>)¹²⁵
 - D. Literature searches and guidelines suggest the variant has a high likelihood of being disease-causing
Using PubMed, Web of Science, and the ACMG standards and guidelines¹²⁶
- II. Genotype analysis shows the mutation is rare (at least <1%) or absent from a large population*
- III. The mutation co-segregates in the family in accordance with the predetermined mode of inheritance, and can be validated through an independent method
- IV. Functional assessment – including *ex vivo*, *in vitro* and *in vivo* bench studies – show the mutant gene product is dysfunctional

*Note that the control databases (1000 Genomes, ExAC, ESP, gnomAD) were first checked to ensure their control cohorts did not include the phenotype of interest. In the event that the phenotype of interest was part of the control cohort in the database, the frequency was used with caution.

The database abbreviations are as follows: OMIM, Online Mendelian Inheritance in Man; KEGG, Kyoto Encyclopaedia of Genes and Genomes; HGMD, Human Gene Mutation Database; ACMG, The American College of Medical Genetics and Genomics.

2.8.1 Zygoty

Given the predicted mode of inheritance of the disease from the family history, zygoty was used to prioritize specific variants. More specifically, if the disease was expected to follow an autosomal dominant mode of inheritance, heterozygous variants were prioritized; if the disease was expected to follow an autosomal recessive mode of inheritance, either homozygous variants or compound heterozygous variants were prioritized. In the event that there was no family history available, the predicted zygoty of the disease of interest was used (i.e. all known cases of congenital general lipodystrophy are associated with an autosomal recessive inheritance pattern, therefore homozygous variants were prioritized).

2.8.2 Read depth

The read depth represents the number of unique reads that include a given nucleotide in a sequence. In order to discard sequencing artifacts or Mendelian errors, variants are prioritized based on read depth. A read depth criterion of ≥ 30 was used for quality control. Note that the read depth is indicated as a quality control measurement and this step is capable of removing ~95% of Mendelian errors¹²⁷.

2.8.3 Sequence ontology

Sequence ontology refers to the structured vocabulary used for genomic annotation. More specifically, it offers a common, universal set of terms and definitions for genomic data. In this project, the sequence ontology terms – disruptive inframe insertion/deletion, frameshift variant, inframe insertion/deletion, initiator codon variant, missense variant, splice acceptor/donor/region variant, stop gain/loss/retained variant – were prioritized. These non-synonymous and splicing variants were prioritized as they are more likely to be pathogenic, destabilize the protein structure, and result in dysfunctional interactions between interacting molecules¹²⁸⁻¹³¹.

2.8.4 Minor allele frequency

Minor allele frequency refers to the frequency at which the second most prevalent allele occurs in a population. In this case, the MAFs were extracted from 1000 Genomes, Exome Sequencing Project, ExAC and gnomAD, as mentioned in Section 2.7. A MAF of <0.01 was used for all rare diseases with an autosomal recessive mode of inheritance, and a MAF of zero was used for rare diseases with a suspected autosomal dominant mode of inheritance.

2.8.5 *In silico* analyses

As specified in Section 2.7, *in silico* tools were used to predict the biological effect or functional relevance of the variants on protein function. Variants were prioritized if the tools predicted that the variant has a high-likelihood of being disease causing.

2.8.6 Gene and protein analysis and literature review

To narrow down the list of candidate variants, analysis into gene function, expression and pathway were performed, and disease databases were analyzed to determine a relationship between the variant and disease of interest. A comprehensive literature review was performed for each variant, and the American College of Medical Genetics (ACMG) standards and guidelines were implemented and considered. See Table 2.8 for more details on the prioritization method.

2.8.7 Association with the disease phenotype

In the event that a variant was identified in a known gene previously associated with the disorder, that variant was prioritized if it satisfied all aforementioned criteria. With regard to the LipidSeq™ data, the lipodystrophy-associated genes were selectively analyzed for potentially causative variants.

2.9 Copy-number variation detection

All patients processed using LipidSeq™ or with WES performed in-house were subject to CNV detection. The CNV Caller tool, an application within the VarSeq® software, was used to call CNVs with sequencing depth-of-coverage data from the previously generated BAM file. The VCF file and BED file (tab-delimited text file with information on the targeted chromosomal regions) for either LipidSeq™ or WES were also input to VarSeq®, as the program uses a set of matched control samples prepared using the same library-preparation chemistry or sequencing method (i.e. LipidSeq™ or WES kit). The CNV Caller tool is capable of identifying four CNV states: 1) wild-type diploid state; 2) heterozygous deletion; 3) homozygous deletion; and 4) duplication.

In short, the CNV calling algorithm relied on the coverage information computed from the BAM files. The algorithm inspected for a change in coverage compared to a set of reference samples as evidence of a CNV event. More specifically, the program selected the top 30 reference samples with the least percent-difference (<20%) in coverage data compared to the sample of interest. These reference samples were used to identify differences in coverage, and correct for GC-content bias and regions that were unamenable to mapping. Using the reference samples, the algorithm computed two evidence metrics: Z-score and ratio threshold value. The Z-score is a measurement referring to the number of standard deviations from the reference sample mean, whereas the ratio refers to the normalized mean for the sample divided by the average normalized mean of all the reference samples. Using these two metrics, the algorithm was able to assign a CNV state (deletion or duplication) to each target region. In addition to these metrics, a third metric, the Variant Allele Frequency (VAF), was calculated to be used as evidence for a CNV. Any VAF value other than zero or one were evidence against deletions, whereas VAF values of 1/3 and 2/3 represented supporting evidence for duplications.

In addition to evidence metrics, quality control flagging was also performed by the software. These flags indicated whether a CNV event was unreliable, specifically if there

was a low reference sample read depth or high variation between reference samples in the region of interest. These flags were used to filter out unlikely CNV events.

2.9.1 Copy-number variation prioritization

After CNV calling in VarSeq®, CNV events were prioritized based on both ratio-threshold value (≤ 0.7 for heterozygous deletions, ≥ 1.3 for duplications) and Z-score (≤ -3.0 for heterozygous deletions, ≥ 3.0 for duplications). The VAF value was considered as secondary validation of evidence. In the event that a CNV event was flagged with a quality control flag, the CNV was no longer considered a candidate CNV associated with the disease of interest.

2.10 Variant validation and co-segregation analysis

In the event that candidate variants were identified, validation and co-segregation analyses were performed using an independent method (i.e. Sanger sequencing or microarray analysis).

2.10.1 PCR

Genomic DNA was ascertained from the proband and all other family members (if available) and amplified using polymerase chain reaction (PCR) and a custom set of primers (see Appendix III for primer sequences). Unique primers were designed for each variant of interest.

2.10.2 PCR imaging and purification

After PCR, the PCR products were loaded onto a 1.5% agarose gel with RedSafe™ Nucleic Acid Staining Solution (iNtRON Biotechnology Inc., Gyeonggi Province, South Korea) and run with a 100-bp DNA Ladder (ThermoFisher Scientific, Waltham, Massachusetts, United States). The gel was run for 45 minutes at 145 volts using a Gibco BRL Horizontal 11-4 Gel Electrophoresis Apparatus (Life Technologies, Carlsbad, California, United States). The gel was subsequently imaged using the Azure Biosystems

c150 Gel Documentation Platform (ThermoFisher Scientific, Waltham, Massachusetts, United States). If the PCR products were of expected size when compared to the ladder fragment size, the PCR products were purified using calf intestinal alkaline phosphatase (CIAP) and exonuclease I (ThermoFisher Scientific, Waltham, Massachusetts, United States).

2.10.3 Sanger sequencing and analysis

After PCR, the cleaned and purified DNA product was sent for Sanger sequencing at the LRGC using an Applied Biosystems (ABI) 3730 Automated DNA Sequencer (Applied Biosystems, Foster City, California, United States). Electropherograms were generated and analyzed using the ABI SeqScape® Software v2.6 (Life Technologies, Carlsbad, California, United States) where it was aligned to the corresponding reference gene sequence from the National Centre for Biotechnology Information (NCBI) GenBank. If the variants were present in other clinically affected individuals in the family (when available) following the same mode of inheritance, they were considered to co-segregate with disease status in the family. In the event that the variant was absent from affected family members, the candidate variant was deemed not causative of the disease of interest.

2.11 Validation of copy-number variation events

For the validation of candidate CNVs, either Sanger sequencing as described or the CytoScan™ HD Array Kit and Reagent Kit Bundle (ThermoFisher Scientific, Waltham, Massachusetts, United States) were used. Briefly, the CytoScan™ HD Solution provides the broadest coverage for detecting human chromosomal aberrations; the CytoScan™ HD Array can reliably detect over 750,000 SNPs and 2.6 million copy-number probes (with 25-50 kilo-base copy-number changes). Due to the high SNP coverage, the CytoScan™ HD Array allows a user to accurately estimate breakpoint locations, a necessary feat in validating CNV events. The CytoScan™ HD Array was performed according to manufacturer's instructions at Victoria Hospital in London, Ontario (London Health Sciences Centre, Victoria Hospital, London, ON, Canada). Following sequencing, the

data were analyzed using the Chromosome Analysis Suite (ChAS) v3.2 (ThermoFisher Scientific, Waltham, Massachusetts, United States). The regions lying between probes differing in copy-number state were used to predict the approximate size of the copy-number variant and were predicted to harbour the breakpoint of the candidate CNV.

2.12 Gene enrichment analysis

For patients with lipodystrophy or extreme early-onset obesity in which no candidate variant or multiple candidate variants were identified, PANTHER (protein annotation through evolutionary relationship) classification system (<http://www.pantherdb.org/>) was used to determine if any protein families (and their genes) were under-or-overrepresented using Fisher's Exact Test in comparison to a reference database¹³². Put simply, PANTHER was used to determine whether certain genes or pathways were affected more-or-less often than expected by chance. For the cases in which PANTHER was run, the candidate variant list used as input was the list of variants remaining after filtering based on zygosity, ontology, and minor allele frequency. Note that proteins were classified according to family and subfamily, molecular function, biological process, and pathway.

2.13 Statistical analysis: Brachydactyly case

Statistical analysis and comparison of the anthropometric data between *IHH* mutation positive and *IHH* mutation negative individuals in the brachydactyly case was done using the Wilcoxon-Mann-Whitney test on SAS v.9.4 (SAS Institute, Cary, North Carolina, United States). The Wilcoxon-Mann-Whitney test is used when sample size is small, and data are not normally distributed¹³³. For statistical analysis in the brachydactyly case, eighteen family members were assessed. Fourteen of these 18 family members were clinically affected, and four were clinically unaffected.

2.14 Box plot generation: Brachydactyly case

Box plots for the brachydactyly case were generated using R: The R Project for Statistical Computing software and the ggPlot2 package¹³⁴.

Chapter 3 – Results

3 Overview

One-hundred-and-fifty-three patients with lipodystrophy, five children with extreme early-onset obesity, one family with a history of brachydactyly, one family with a history of cerebral atrophy, and 11 families with several rare diseases (as ascertained from FORGE Canada and CARE4RARE) participated in this study. Here, I discuss all the results and outcomes of NGS.

3.1 Study subjects and clinical features

For all cases studied – namely the lipodystrophy, early-onset obesity, brachydactyly, cerebral atrophy, CARE4RARE and FORGE Canada projects – patient files, clinical notes, and/or hospital records were thoroughly reviewed to develop accurate clinical descriptions and select appropriate patients for sequencing. All cases were chosen based on disease severity. In the event that limited information about the phenotype was available, the phenotype described in the OMIM database was used.

3.2 LipidSeq™ variant detection

After sequencing using the LipidSeq™ targeted gene panel, the data for all 153 patients with lipodystrophy were analyzed for potential disease-causing variants. These variants were prioritized based on zygosity, read depth, sequence ontology and minor allele frequency as discussed in Chapter 2, Section 2.8. Of the 153 total patients, 81 variants were identified within 14 lipodystrophy-associated genes: *LMNA*, *PPARG*, *PLIN1*, *CIDEA*, *LIPE*, *LPIN1*, *AGPAT2*, *BSCL2*, *CAVI*, *AKT2*, *PTRF*, *DYRK1B*, *POLD1* and *WRN* (Figure 3.2). These variants explain the clinical phenotype or secondary complications of 58 patients (38% of the patient cohort) processed using the LipidSeq™ gene panel, as all genes are either associated with lipodystrophy or with common features often implicated in lipodystrophy (i.e. high triglycerides, high cholesterol). The 95 patients classified as mutation negative after LipidSeq™ became candidates for the WES portion of the study.

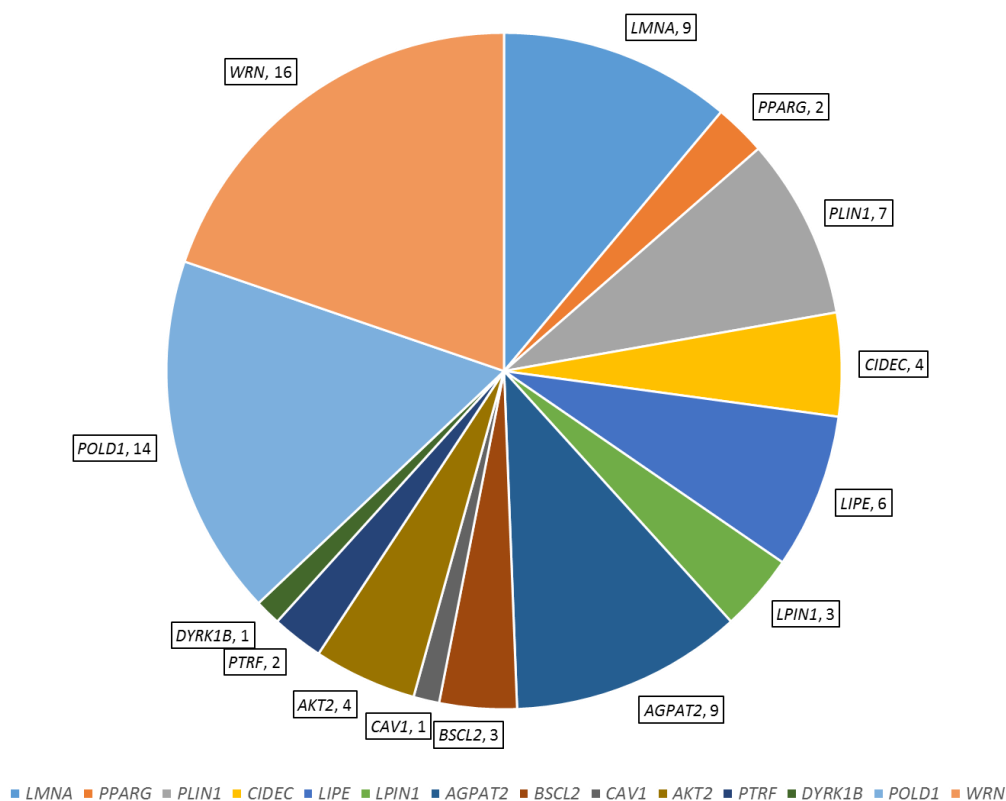


Figure 3.2. Variant breakdown for patients with lipodystrophy sequenced using the LipidSeq™ targeted gene panel. The DNA samples of Dr. Hegele’s patients with lipodystrophy were ascertained and sequenced using LipidSeq™. Fifty-eight mutation-positive patients were classified according to lipodystrophy-causing gene. Each region in the graph represents the gene implicated, and the number of patients with a variant in the gene. Note that some patients harboured multiple variants in more than one gene, which likely contributed to their phenotype.

3.3 Sanger sequencing of *TMPRSS4*

For the proband featuring a cerebral atrophy-like phenotype, Sanger sequencing of the expected *TMPRSS4* variant region using a custom set of primers (see Appendix III) was performed to determine whether the case shared the same molecular etiology as the previous case in the Blackburn Cardiovascular Genetics Laboratory (as described in Chapter 1, Section 1.5.4). From Sanger sequencing, no variants were identified in the *TMPRSS4* region of interest designating the proband eligible for the WES part of the study.

3.4 Whole-exome sequencing

Twenty patients with lipodystrophy, five children with extreme early-onset obesity, one proband with brachydactyly, one proband with a cerebral atrophy-like phenotype, and 11 CARE4RARE and FORGE Canada cases were selected for the whole-exome portion of the study. To start, patient data were collected in order to fully understand the patient phenotype. The clinical features of the lipodystrophy and extreme early-onset obesity cases are described in Table 3.4A and Table 3.4B, respectively. The clinical description of the brachydactyly, cerebral atrophy-like, CARE4RARE and FORGE Canada cases are listed in Chapter 1 (Sections 1.5.3.1.8, 1.5.4.1 and Table 1.2.5.3, sequentially). After WES, bioinformatics processing and *in silico* analyses were performed on the WES data to identify potential disease-causing variants. Variants were prioritized if they satisfied criteria I-III in Table 2.8 (Chapter 2).

Table 3.4A. Features of all patients with lipodystrophy in which whole-exome sequencing was performed^A.

Patient ID ^B	Biological Sex ^C	Classification	Additional Clinical Features (if present)
1804	Female	Total lipodystrophy	T2D, insulin resistance, hyperlipidemia, exercise intolerance
8615	Female	Total lipodystrophy	T2D, iron-deficient anemia, diabetic neuropathy, history of bronchitis, class 3 obesity, CAD
1088	Female	FPLD	N/A
3458	Female	FPLD	Hypothyroidism, diabetes
3969	Female	FPLD	Peripheral neuropathy (motor/axonal), Charcot Marie Tooth, T2D, constant migraines, NASH
4116	Female	FPLD	Hypertension, T2D, chronic renal failure, hypothyroidism
4665	Female	FPLD	Neuropathy, T2D, Behçet syndrome, pyoderma gangrenosum, hidradenitis
5662	Female	FPLD	N/A
2025	Female	FPLD	Profound insulin resistance, decreased creatinine levels, proteinuria, dermatomyositis
3011	Female	FPLD	Insulin resistance, dyslipidemia, hirsutism, high triglycerides
11618	Female	FPLD	Hypertension, T2D, hypercholesterolemia, obesity
14318	Female	FPLD	Severe insulin resistance
794	Female	FPLD	Hepatosplenomegaly, cirrhosis and NASH, high triglycerides, high cholesterol, insulin-requiring T2D, liver failure
14106	Female	FPLD	N/A
1120	Male	FPLD	N/A
4162	Female	Acquired lipodystrophy	Hypercholesterolemia, lower motor neuron weakness, hair loss, osteoarthritis
4164	Female	Acquired lipodystrophy	Insulin-requiring T2D
4194	Female	Acquired lipodystrophy	T2D, diabetic neuropathy, osteoarthritis
4226	Female	Acquired lipodystrophy	N/A
12395	Female	PCOS/FPLD	T2D, neuropathy, hypertension

The abbreviations are as follows: ID, identification; N/A, not applicable; T2D, Diabetes mellitus, Type 2; CAD, coronary artery disease; FPLD, familial partial lipodystrophy; NASH, non-alcoholic steatohepatitis; PCOS, polycystic ovary syndrome.

^APatients were selected based on their diagnosis with a rare metabolic disorder, poor lipid profile, and lack of mutations in known disease-related genes (as determined through Sanger sequencing). Patients of Dr. Hegele's (London Health Sciences Centre) were prioritized to ensure consistent diagnosis, high sample quality, and detailed, accurate patient information. After whole-exome sequencing, the data were run through CLC Bio Genome Workbench v10 and annotated using the curation and annotation program, Golden Helix VarSeq®.

^BThe sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

^CReferring to the biological sex assigned at birth.

Table 3.4B. Features of all patients with extreme early-onset obesity in which whole-exome sequencing was performed^A.

Patient ID ^B	Biological Sex ^C	Classification	Additional Clinical Features (if present)
12340	Male	Early-onset obesity	Insulin resistance, hyperphagia
15544	Male	Early-onset obesity	Autism spectrum disorder and developmental delay, overgrowth
15858 ^D	Female	Early-onset obesity	Hyperphagia, overgrowth, intellectual disability
12424	Male	Early-onset obesity	N/A
14504 ^D	Female	Early-onset obesity	N/A

The abbreviations are as follows: ID, identification; N/A, not applicable.

^A Patients were selected based on their diagnosis with a rare metabolic disorder, poor lipid profile, and lack of mutations in known disease-related genes (as determined through Sanger sequencing). After whole-exome sequencing, the data was run through CLC Bio Genome Workbench v10 and annotated using the curation and annotation program, Golden Helix VarSeq®.

^B The sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

^C Referring to the biological sex assigned at birth.

^D Trio sequencing and analysis was performed with this patient and their unrelated parents.

3.4.1 Lipodystrophy

From WES and bioinformatics analysis, over one-million variants were identified among the 20 patients with lipodystrophy. After variant prioritization, candidate variants were identified for 13 of the 20 patients with lipodystrophy within the following genes: *ISCU*, *HSPG2*, *LDLR*, *PDRM16*, *ZMPSTE24*, *FLNC*, *GALC*, *ABCA4*, *WRN*, *ABCG8*, *FHL1*, *ANLN*, *LPL*, *FNI*, *PAPSS2*, *NEUROD1*, *BSCL2*, *APOB*, *PCSK9*, and *PYGM* (Table 3.4.1). These variants partially explain the lipodystrophy phenotype or associated comorbidities displayed by these patients.

3.4.2 Extreme early-onset obesity

After WES, data processing and bioinformatics analysis, over 100,000 variants were called in the exomes of five children with extreme early-onset obesity. After variant prioritization, candidate variants were identified in two of the probands (Table 3.4.2). These variants are predicted to explain the extreme early-onset obesity phenotype or the associated conditions (i.e. intellectual disability) described in these patients.

Table 3.4.1. Candidate variants identified in patients with lipodystrophy using whole-exome sequencing^A.

Patient ID ^B	Gene	Chromosome: Position	DNA change	Amino acid change	Zygoty/Effect	Frequency
1804 (CGL)	<i>ISCU</i>	12:108956417	c.19_20delinsGG	p.Phe7Gly	Homo/Missense	0
8615 (CGL)	<i>HSPG2</i>	1:22200454	c.3707C>A	p.Ala1236Glu	Homo/Missense	1×10 ⁻³
1088 (FPLD)	<i>LDLR</i>	19:11221435	c.1048C>T	p.Arg350Ter	Het/Nonsense	8×10 ⁻⁶
	<i>PRDM16</i>	1:3329229	c.2468G>C	p.Arg823Pro	Het/Missense	7×10 ⁻⁴
3458 (FPLD)	<i>ZMPSTE24</i>	1:40756665	c.1199A>G	p.Asn400Ser	Het/Missense	0
3969 (FPLD)	<i>FLNC</i>	7:128494727	c.6988G>A	p.Gly2330Ser	Het/Missense	7×10 ⁻⁴
	<i>GALC</i>	14:88452941	c.334A>G	p.Thr112Ala	Het/Missense	2×10 ⁻³
4116 (FPLD)	<i>LDLR</i>	19:11233961	c.2252G>A	p.Arg751Gln	Het/Missense	8×10 ⁻⁵
		19:11233940	c.2231_2232delins AG	p.Arg744Gln	Het/Missense	0
4665 (FPLD)	<i>ABCA4</i>	1:94528818	c.1610G>A	p.Arg537His	Het/Missense	1×10 ⁻³
		1:94510250	c.2969G>A	p.Gly990Glu	Het/Missense	0
	<i>WRN</i>	8:30998961	c.2983G>A	p.Ala995Thr	Het/Missense	7×10 ⁻⁴
		8:30938692	c.1149G>T	p.Leu383Phe	Het/Missense	4×10 ⁻⁴
5662 (FPLD)	<i>ABCG8</i>	2:44101538	c.1412-8delinsTT	N/A	Homo/Splicing	0
2025 (FPLD)	<i>FHL1</i>	X:135288637	c.46G>T	p.Gly16Trp	Het/Missense	0
	<i>ANLN</i>	7:36450231	c.1205C>T	p.Pro402Leu	Het/Missense	0
3011 (FPLD)	<i>LPL</i>	8:19811864	c.775G>A	p.Asp259Asn	Het/Missense	0
	<i>PAPSS2</i>	10:89501019	c.1114G>T	p.Val372Phe	Het/Missense	8×10 ⁻⁶
14318 (FPLD)	<i>NEUROD1</i>	2:182543418	c.170G>A	p.Gly57Glu	Het/Missense	0
	<i>BSCL2</i>	11:62457948	c.1280T>C	p.Leu427Pro	Het/Missense	2×10 ⁻³
	<i>APOB</i>	2:21232803	c.6936_6937delins TG	p.Ile2313Val	Homo/Missense	0
	<i>PCSK9</i>	1:55505647	c.137G>T	p.Arg46Leu	Het/Missense	8×10 ⁻³
	<i>PYGM</i>	11:64527223	c.148C>T	p.Arg50Ter	Het/Nonsense	1×10 ⁻⁴
4194 (ALD)	<i>WRN</i>	8:30921838	c.243T>G	p.Phe81Leu	Het/Missense	0

Abbreviations are as follows: CGL, congenital generalized lipodystrophy; FPLD, familial partial lipodystrophy; ALD, acquired lipodystrophy; Het, heterozygous; Homo, homozygous.

^AThe candidate genes and variants were selected based on: low frequency in multiple databases, deleterious *in silico* predictions, relevance to the disease phenotype, and support from numerous papers in the literature. frequency listed for each candidate variant is the frequency from gnomAD, also confirmed from the databases: 1000 Genomes, Exome Aggregation Consortium, and the Exome Sequencing Project.

^BThe sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

Table 3.4.2. Candidate variants identified in patients with extreme early-onset obesity using whole-exome sequencing^A.

Patient ID ^B	Gene	Chromosome: Position	DNA change	Amino acid change	Zygoty/Effect	Frequency
15544	<i>BLK</i>	8:11405609	c.244G>C	p.Gly82Arg	Het/Missense	1×10 ⁻⁵
	<i>MC4R</i>	18:58038832	c.751A>C	p.Ile251Leu	Het/Missense	6×10 ⁻³
15858	<i>BBS10</i>		c.424G>A	p.Asp142Asn	Het/Missense	0
	<i>SETD8</i>		c.995T>C	p.Leu332Pro	Het/Missense	0
			c.713G>C	p.Arg238Pro	Het/Missense	0
	<i>KMT2C</i>		c.2645T>C	p.Ile882Thr	Het/Missense	0

Abbreviations are as follows: Het, heterozygous; Homo, homozygous; Hemi, hemizygous.

^AThe candidate genes and variants were selected based on: low frequency in multiple databases, deleterious *in silico* predictions, relevance to the disease phenotype, and support from numerous papers in the literature. frequency listed for each candidate variant is the frequency from gnomAD, also confirmed from the databases: 1000 Genomes, Exome Aggregation Consortium, and the Exome Sequencing Project.

^BThe sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

3.4.3 Brachydactyly^A

From WES and bioinformatics processing of the DNA sample from the proband, over 37,000 variants were identified in the proband's exome. After a non-synonymous rare variant prioritization approach and literature review, I was left with one candidate variant correlating with the brachydactyly phenotype. This variant, a heterozygous disruptive inframe insertion (c.285_287dupGAA, p.Glu95_Asn96insLys), was detected in the Indian hedgehog gene (*IHH*). This *IHH* variant is predicted to exert a damaging effect on protein function from multiple *in silico* prediction tools, is considered novel in multiple control population databases, and has not been previously reported in the literature. Dominant variants in the *IHH* gene are associated with brachydactyly type A1, correlating with the phenotype and inheritance pattern seen in this family⁸³.

3.4.4 Cerebral atrophy-like

Due to the absence of the expected variant in *TMPRSS4*, WES and bioinformatics analysis was performed on the DNA of the proband featuring a cerebral atrophy-like phenotype and his parents in order to perform trio analysis and identify causative genetic factors. From this analysis, one homozygous candidate variant (from 52,958 total variants) in the Kinetochores Scaffold 1 or *CASC5* gene was identified (c.6485A>G, p.Glu2162Gly), associated with primary microcephaly-4. This gene variant is predicted to cause a pathogenic effect on protein function from *in silico* tools and is present at a minor allele frequency of 6×10^{-6} in control population databases. Note that the frequency in the normal population is associated with heterozygotes.

^A A version of this section has been submitted for publication⁸³.

3.4.5 CARE for RARE and FORGE Canada

From WES and bioinformatics processing on the exomes of the affected probands, candidate variants were identified in eight of the 11 CARE4RARE or FORGE Canada cases, namely microcephaly-ichthyosis, autosomal dominant isolated cryptophthalmos, widow's peak syndrome, amyoplasia, metaphyseal chondrodysplasia, craniosynostosis, heterotaxy disorder, and craniorachischisis (Table 3.4.5).

Here, I focus on the microcephaly-ichthyosis and widow's peak cases. With the microcephaly-ichthyosis case, the DNA from two affected infants were subject to WES. From WES and data processing, a total of 121,280 variants were identified in both exomes. After prioritization, this was narrowed down to one candidate variant in the vacuolar protein sorting-52 gene (*VPS52*; c.57delG, p.Thr20fs). The *VPS52* gene is associated with pluripotent differentiation for ectodermal cell interactions. Further, mutations in the *VPS52* complex are associated with microcephaly, a clinical feature of the affected probands.

For the widow's peak case, two affected individuals had their DNA extracted and WES performed. From WES and bioinformatics analysis, a total of 1,873,928 variants were identified in the two exomes. After prioritizing based on variants on the X chromosome (due to the predicted X-linked dominant inheritance), frequency, read depth and sequence ontology, one candidate variant in the lysine-specific demethylase 6a gene (*KDM6A*) (c.2703-6_2703-5delTT) remained. Variation in the *KDM6A* gene is associated with Kabuki Syndrome-2, matching the phenotype of the proband.

Table 3.4.5. Status of all CARE for RARE and FORGE Canada cases.

Case	Status
FORGE 214, Angelman-like syndrome	<ul style="list-style-type: none"> Nothing significant found with the functional studies Input into MatchMaker Exchange
FORGE 122, Dystonia, ataxia and intellectual disability*	<ul style="list-style-type: none"> Compound heterozygous variants previously found in the NudC Domain Containing 2 (<i>NUDC2</i>) Variants Sanger sequenced again, and are found to not co-segregate in the family Rainfall plots generated to determine mutational burden; gene enrichment analysis performed
FORGE 11, Osteopetrosis	<ul style="list-style-type: none"> Exome analysis phase – looking for candidate variants Waiting for DNA on an additional case
FORGE 352, Microcephaly-ichthyosis*	<ul style="list-style-type: none"> Candidate variant (p.Thr20fs) found in the vacuolar protein sorting 52 (<i>VPS52</i>) gene co-segregates in the pedigree Waiting on DNA for additional affected family members
C4R 572, Autosomal dominant isolated cryptophthalmos	<ul style="list-style-type: none"> Investigating a candidate variant in the Diacylglycerol Kinase Delta (<i>DGKD</i>) gene
C4R 573, Widow's peak syndrome*	<ul style="list-style-type: none"> Candidate variant (c.2703-6_2703-5delTT) identified in Lysine Demethylase 6A (<i>KDM6A</i>) RNA extraction performed on the samples of both affected individuals RT-PCR to be performed to confirm the variant
C4R 575, Amyoplasia	<ul style="list-style-type: none"> New mutations identified in affected patients and not in parents: heparin sulfate proteoglycan 2 (<i>HSPG2</i>; p.Val3507Leu) in affected male, TEA Domain Transcription Factor 4 (<i>TEAD4</i>; p.Val212Met) in affected female Currently looking to recruit additional patients
C4R 592, Metaphyseal chondrodysplasia	<ul style="list-style-type: none"> Candidate variant identified in the Dishevelled Segment Polarity Protein 3 (<i>DVL3</i>) gene (p.Gln301Ter) Currently investigating other variants and considering functional studies
C4R 664, Craniosynostosis*	<ul style="list-style-type: none"> Candidate variant identified in the TNF Receptor Associated Factor 7 (<i>TRAF7</i>) gene (c.1673C>T, p.S558F)
C4R 907, Heterotaxy disorder	<ul style="list-style-type: none"> Candidate variant identified in the Interferon Related Developmental Regulator 1 (<i>IFRD1</i>) gene (c.568-4_568-3delTT) Waiting on DNA to confirm the variant of interest
C4R 924, Craniorachischisis*	<ul style="list-style-type: none"> Variants of interest identified in the Zinc Finger And BTB Domain Containing 33 (<i>ZBTB33</i>; p.Asp189_Asp190insAla) and the LFNG O-Fucosylpeptide 3-Beta-N-Acetylglucosaminyltransferase (<i>LFNG</i>; p.Trp49Argfs) genes Confirmed to co-segregate with disease status in the pedigree

Abbreviations are as follows: FORGE, Finding of Rare Disease Genes Canada project.

Results listed are those up until July 1, 2018.

*The red asterisk represents the projects I was directly involved in or working with undergraduate students on. The remaining cases without an asterisk were those in which I assumed an administrative role.

3.5 Shared sequence variation

From over one million total variants across the 25 rare metabolic disease (lipodystrophy and early-onset obesity) exomes, six ultra-rare androgen receptor (*AR*) variants were identified in four patients with partial lipodystrophy and one with extreme early-onset obesity (Table 3.5). Of these six identified variants, three have never been reported in the literature or in any database. For the variants that are not novel, note that a frequency greater than zero-percent was prioritized to account for heterozygotes in the normal control population.

The markedly high prevalence of rare *AR* variants in this small patient cohort is unexpected; furthermore, *AR* variants have never been described in either of these two disorders characterized by abnormal tissue distribution. All variants detected were harboured within the N-terminal domain of the *AR* protein product (Figure 3.5)

Table 3.5. Candidate variants identified in the androgen receptor gene.

Patient ID (Phenotype) ^A	Amino Acid Change	Sequence Ontology	Zygoty/Effect	Frequency
2025 (FPLD)	Gly470_Gly473del	In-frame deletion	Homozygous	5×10^{-3}
3011 (FPLD)	Gln80_Glu81insGlnGlnGln	In-frame insertion	Heterozygous	0
4116 (FPLD)	His384Gln	Missense	Heterozygous	0
14318 (FPLD)	Gly470_Gly473del	In-frame deletion	Homozygous	5×10^{-3}
	Gln62Leu	Missense	Heterozygous	4×10^{-5}
15544 (Obesity)	Gln69fs	Frameshift	Hemizygous	0

The candidate genes and variants were selected based on: low frequency in multiple databases, deleterious *in silico* predictions, relevance to the disease phenotype, and support from numerous papers in the literature.

The frequencies listed for each candidate variant were extracted from gnomAD and verified using independent databases, ExAC, 1000 Genomes and ESP.

Abbreviations are as follows: FPLD, familial partial lipodystrophy.

^AThe sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

AR gene



AR protein



Figure 3.5. Androgen receptor gene and protein structure. Amino acid changes of interest in the androgen receptor (AR) protein structure are represented by the yellow stars.

3.6 Copy-number variation discovery

All patients processed using LipidSeq™ (individuals with lipodystrophy) or with WES performed in-house (lipodystrophy and obesity cases) were subject to CNV detection. To call CNVs, the CNV Caller tool within the VarSeq® software was used. CNVs were called using sequencing depth-of-coverage data from each BAM file. As discussed in Chapter 2, CNVs were prioritized based on both ratio-threshold value (≤ 0.7 for heterozygous deletions, ≥ 1.3 for duplications) and Z-score (≤ -3.0 for heterozygous deletions, ≥ 3.0 for duplications).

3.6.1 LipidSeq™ copy-number variation detection

For each patient with lipodystrophy processed on the LipidSeq™, the CNV Caller tool within VarSeq™ was used to call CNVs. After CNV calling and prioritization, seven candidate CNVs were identified within seven patients and predicted to be causative of the lipodystrophy phenotypes (Table 3.6.1).

In total, three heterozygous deletions and four duplications were detected and prioritized as causative. The deletions were present in *AGPAT2*, *PLIN1* and *CIDEA*, where *AGPAT2* is associated with *CGL1*, and both *PLIN1* and *CIDEA* are associated with *FPLD5*. Each deletion is expected to cause or be associated with the lipodystrophy phenotype of these patients. The duplications identified in four patients were present in *PLIN1*, *MFN2*, *APOA5/APOA4* and *CREB3L3* associated with *FPLD5*, neuropathy, high triglycerides, and hyperlipidemia/hypertriglyceridemia, respectively. While the *PLIN1* variant is expected to cause the lipodystrophy phenotype, the *MFN2*, *APOA5/APOA4* and *CREB3L3* duplication CNVs are thought to be associated with the secondary complications of lipodystrophy (i.e. high triglycerides).

Table 3.6.1. Candidate copy-number variants identified in patients with lipodystrophy using the LipidSeq™ targeted gene panel.

Patient ID ^A	Region (affected probes)	Gene(s), Exon(s) ^B	CNV State
1120 ^C	19:4153347-4173301	<i>CREB3L3</i> , 5'UTR – 3'UTR	Duplication
4192	11:116659835-116694261	<i>APOA5</i> , <i>APOA4</i> , 5'UTR – 3'UTR	Duplication
14557	1:12052361-12057728	<i>MFN2</i> , exons 4 – 6	Duplication
15678	15:90207349-90216895	<i>PLIN1</i> , exon 3 – 3'UTR	Duplication
1099 ^D	15:90220425-90220484	<i>PLIN1</i> , exon 2	Deletion/Heterozygous
15529	3:9921657-9922188	<i>CIDEA</i> , alternative non-coding exon 1	Deletion/Heterozygous
2108	9:139581377-139582161	<i>AGPAT2</i> , 5'UTR – exon 1	Deletion/Heterozygous

After sequencing using the LipidSeq™ panel, the data was run through CLC Bio Genome Workbench v10 and annotated using the curation and annotation program, Golden Helix VarSeq®. Copy-number variation was called using the depth of coverage and the CNV caller algorithm embedded in the Golden Helix VarSeq® software.

Abbreviations are as follows: ID, identification; CNV, copy-number variation; UTR, untranslated region.

^AThe sample ID number represents a random, integral laboratory number assigned upon the receipt of the patient sample to allow for the privacy and respect of the patient.

^BAll known genes and exons/regions affected.

^CCopy-number alteration confirmed using the independent method, CytoScan™ HD.

^DCopy-number alteration confirmed using Sanger sequencing.

3.6.2 Exome copy-number variation detection

For the 25 lipodystrophy and obesity exomes, the CNV Caller tool in VarSeq® was utilized to call CNVs. After prioritization based on ratio threshold value and Z-score, no CNVs were predicted to be causative from WES data.

3.7 Validation of candidate variants and copy-number variations

When a candidate variant or CNV is discovered, it is imperative to ensure the variant or CNV is not a sequencing artifact, and to determine whether the variant or CNV co-segregates with disease status in the family (if available). For select cases, variants were confirmed through Sanger sequencing, and CNVs were confirmed through Sanger sequencing, WES, or the CytoScan™ HD microarray.

3.7.1 *IHH* validation (brachydactyly)^A

The *IHH* p.Glu95_Asp96insLys variant was validated using PCR and Sanger sequencing. Primers were designed (see Appendix III) for the region of interest (c.285_287dupGAA, p.Glu95_Asn96insLys) and genomic DNA from the family was genotyped to confirm the presence of the mutation. Seventeen family members were genotyped for the mutation using direct Sanger sequencing, and the variant was determined to co-segregate with disease status in the family. Figure 3.7.1A shows the same family pedigree as Chapter 1, Figure 1.5.3.1.8 with additional genotyping information, Figure 3.7.1B illustrates the results from Sanger sequencing.

^A A version of this section has been submitted for publication⁸³.

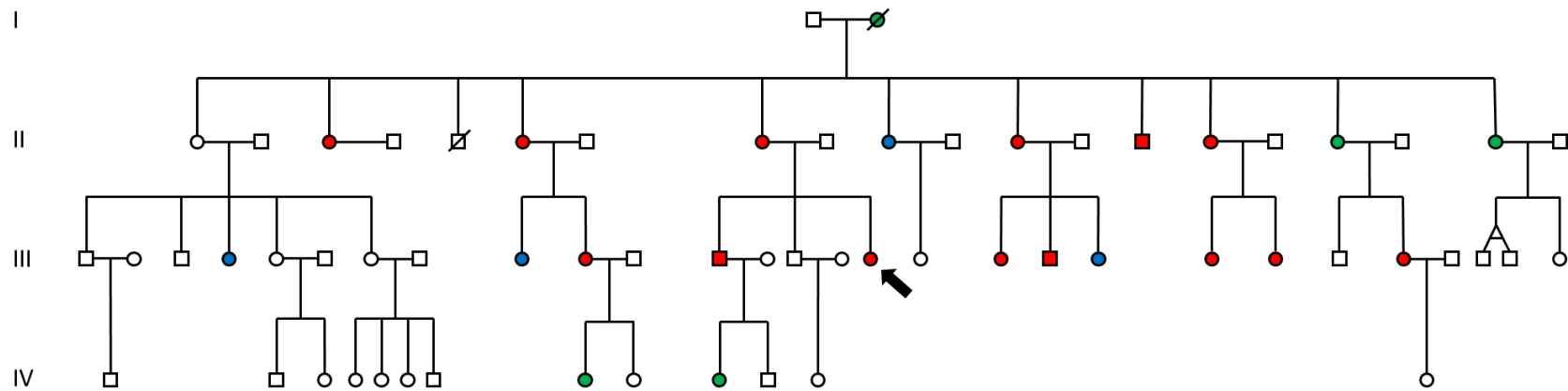


Figure 3.7.1A. Pedigree and *IHH* genotype information for a Canadian autosomal dominant brachydactyly type A1-affected family.

The proband, as indicated by the black arrow, was the first patient ascertained. All red and blue shaded individuals represent those whose DNA samples were Sanger sequenced for the *IHH* variant (exon 1, c.285_287dupGAA, p.Glu95_Asn96insLys). The reddened squares (males) and circles (females) represent affected individuals with brachydactyly type A1 who carry the *IHH* variant. The blue coloured squares and circles indicate those who are unaffected by brachydactyly and do not carry the *IHH* variant. The green coloured squares and circles represent the individuals clinically diagnosed with brachydactyly whose DNA was not obtained or sequenced. Diagonal lines indicate deceased individuals.

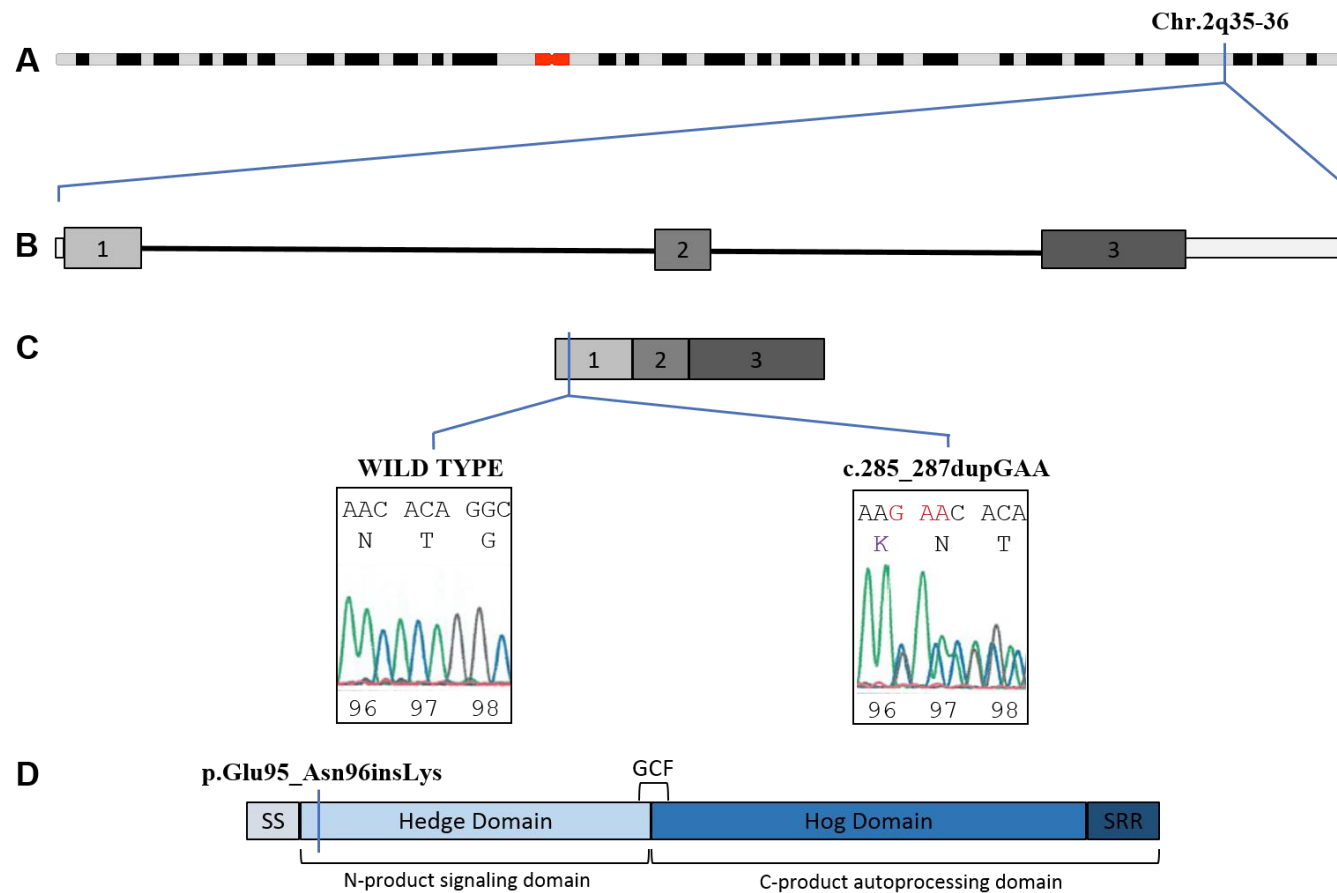


Figure 3.7.1B. Chromosome location, gene structure, sequence analysis and protein structure of Indian Hedgehog. A shows chromosome 2 and the location of the *IHH* gene drawn to scale. B shows the genomic DNA for the *IHH* gene. C shows the mRNA for *IHH* with the mutation indicated. Electrophoretic tracings for the *IHH* variant for both unaffected and affected family members are also shown. D shows the IHH protein structure with key domains, regions, and the mutation indicated.

3.7.2 *VPS52* validation (microcephaly-ichthyosis)

Using PCR followed by Sanger sequencing, the *VPS52* (c.57delG, p.Thr20fs) variant was validated in the two affected individuals. Additionally, the DNA of an available, unaffected sibling was Sanger sequenced to confirm the absence of the variant. The *VPS52* variant was found to co-segregate with disease status in the pedigree (Figure 3.7.2).

3.7.3 *KDM6A* validation (widow's peak)

The *KDM6A* variant (c.2703-6_2703-5delTT) was subjected to PCR and Sanger sequencing to confirm the validity of the candidate variant. The variant was present in both affected individuals, and absent in an unaffected family member, confirming co-segregation with disease status in the family (Figure 3.7.3).

3.7.4 *AR* validation (lipodystrophy and obesity)

All six *AR* gene variants (p.Gly470_Gly473del, p.Gln80_Glu81insGlnGlnGln, p.His384Gln, p.Gly470_Gly473del, p.Gln62Leu, p.Gln69fs) were subject to PCR and Sanger sequencing for validation. Due to the high GC-content of the *AR* gene, the sequencing results were inconclusive and validation of the *AR* variants is ongoing.

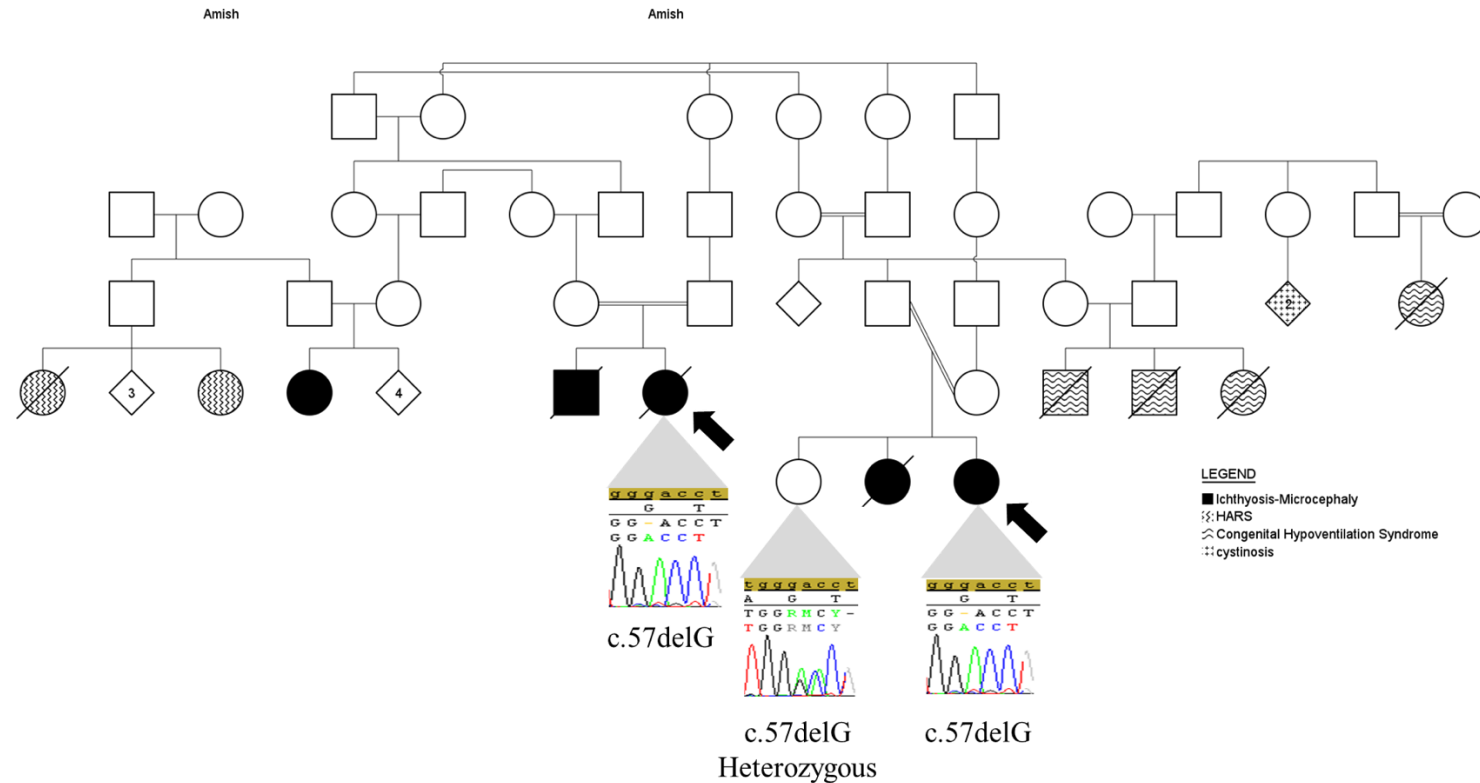


Figure 3.7.2. Sanger sequencing confirmation of the *VPS52* gene variant in a family with a history of microcephaly-ichthyosis. From whole-exome sequencing, one candidate variant (c.57delG) in the vacuolar protein sorting-52 gene (*VPS52*) was predicted to be causative of the microcephaly-ichthyosis phenotype in the family. The arrows indicate the probands. After whole-exome sequencing, Sanger sequencing was utilized to confirm the variant in the two affected probands and confirm the absence of the variant in an available unaffected family member. Squares indicate biological males, circles indicate biological females, and diagonal lines represent individuals who are deceased.

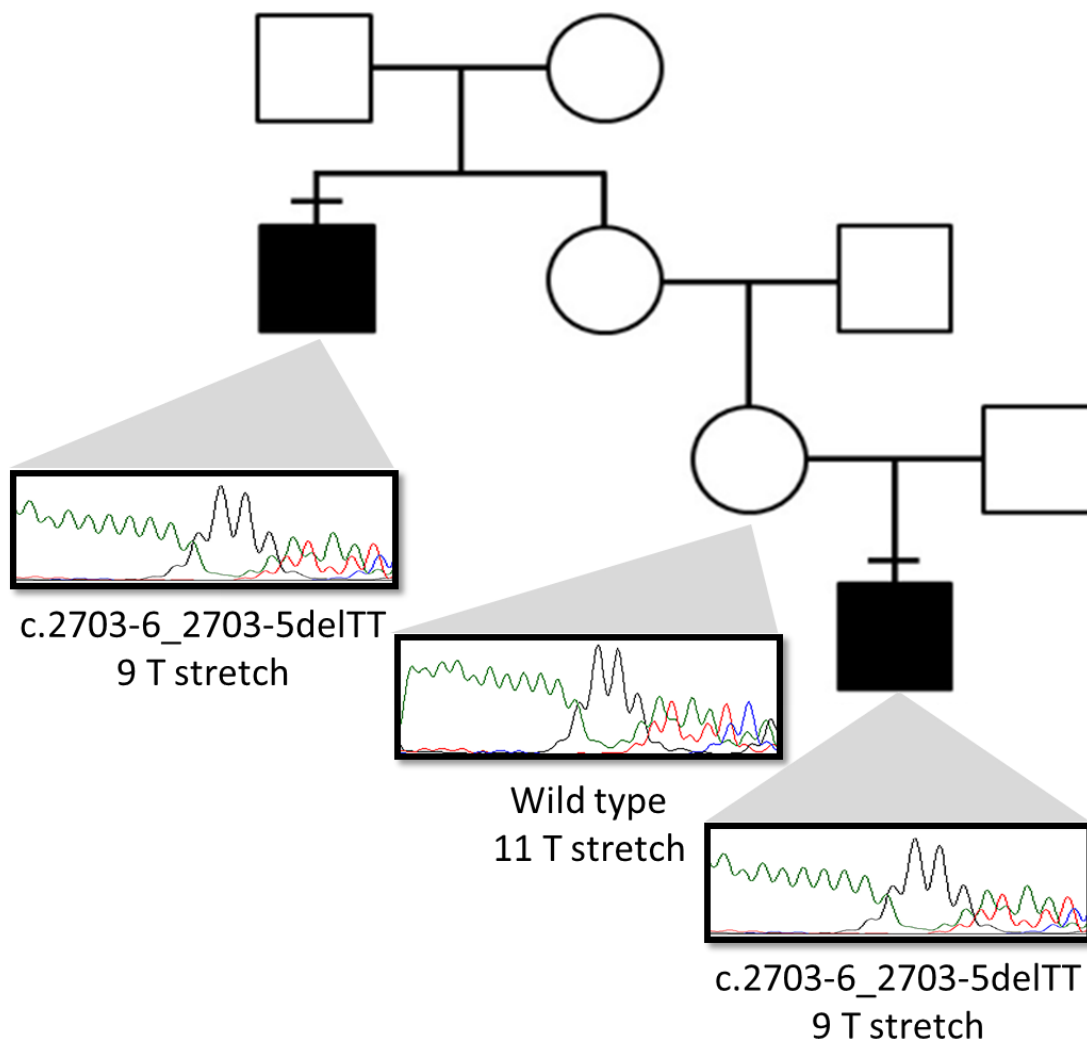


Figure 3.7.3. Sanger sequencing confirmation of the *KDM6A* gene variant in a family with a history of widow's peak syndrome. From whole-exome sequencing, one candidate variant (c.2703-6_2705-5delTT) in the lysine-specific demethylase 6a gene (*KDM6A*) was predicted to be causative of the widow's peak syndrome phenotype in the family. After whole-exome sequencing, Sanger sequencing was utilized to confirm the variant in the two affected probands and confirm the absence of the variant in an available unaffected family member. Squares indicate biological males, and circles indicate biological females. Horizontal lines over the squares represent the samples subjected to whole-exome sequencing.

3.7.5 Copy-number variation validation

After CNV detection, validation was necessary to ensure that the CNV events called were not artifacts or false positives. In order to validate the CNVs, three independent methods were used, namely Sanger sequencing, WES and the CytoScan™ HD array. To date, two CNVs in patients 1099 and 1120 have been confirmed while five are ongoing.

3.7.5.1 Patient 1099: *PLIN1* exon 2 deletion

From the VarSeq® CNV Caller tool and LipidSeq™ data, an exon 2 deletion was identified in the *PLIN1* gene. This CNV event was validated using PCR, gel electrophoresis and Sanger sequencing on the patient and normal control samples (no lipodystrophy) with primers flanking exons 1 and 3 (Figure 3.7.5.1; see Appendix III for primer details).

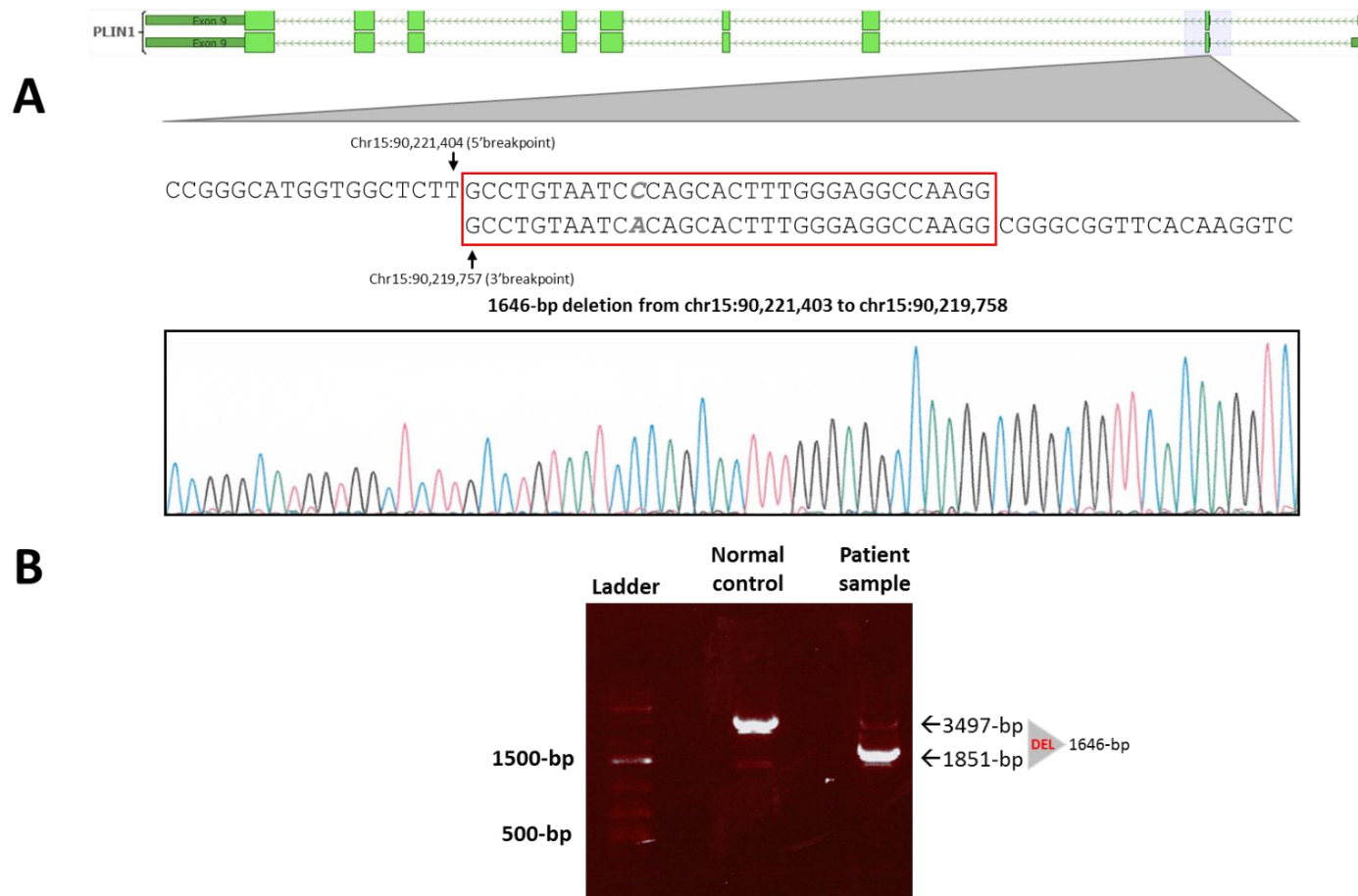


Figure 3.7.5.1. Validation of a *PLIN1* exon 2 deletion in a patient with familial partial lipodystrophy. Panel A: Sanger sequencing results showing the deletion of exon 2. Panel B: PCR and agarose gel electrophoresis across the breakpoints to confirm a *PLIN1* exon 2 deletion in a patient with familial partial lipodystrophy. Lane 1 shows a 1-kb ladder, lane 2 shows a normal control, and lane 3 shows the patient sample. The patient is heterozygous for both the wild type sequence and a 1646-bp deletion.

3.7.5.2 Patient 1120: *CREB3L3* duplication

From the VarSeq® CNV Caller tool and LipidSeq™ data, a full gene deletion of *CREB3L3* was detected. To confirm this CNV event and determine whether the neighbouring genes were affected, WES was performed. From WES, it was determined that several neighbouring genes were also duplicated (Figure 3.7.5.2). More specifically, the *PIAS4*, *ZBTB7A*, *MAP2K2*, *CREB3L3*, *SIRT6*, and *ANKRD24* genes were fully duplicated (5'UTR–3'UTR). Notably, the *SIRT6* gene encodes the sirtuin-6 protein, a nuclear protein that catalyzes deacylation and mono-ADP ribosylation. Deficiency of *SIRT6* results in accelerated aging, suggesting a link between lamin A and *SIRT6*, as lamin A defects are associated with progeroid features and lipodystrophy. Lamin A deficiency is the main cause of FPLD2, matching the phenotype of the patient.

To further validate the CNV event detected from WES, the DNA of the patient was subjected to genotyping using the CytoScan™ HD array. The microarray demonstrated the same results as the WES data, confirming the validity of the multi-gene duplication.

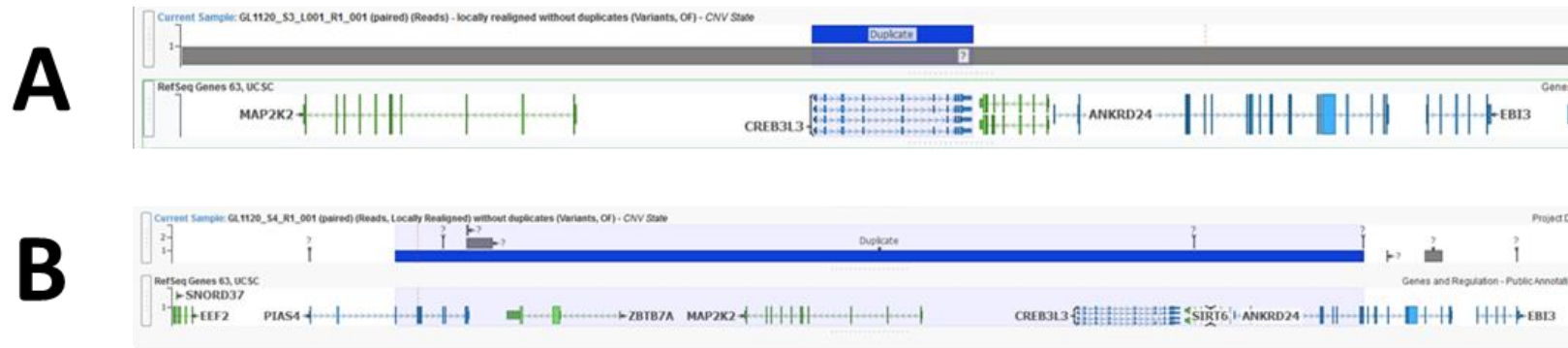


Figure 3.7.5.2. Detection of a duplication event from LipidSeq™ versus whole-exome sequencing data. Panel A shows the CNV event called by the VarSeq® CNV Caller tool using LipidSeq™ data. Panel B illustrates the CNV event called by the VarSeq® CNV Caller tool using exome data. In each panel, the top region shows the CNV state as determined from both ratio and Z-score metrics, and the bottom region shows the exon map of the affected gene(s). The regions highlighted in blue as “duplicate” are the regions where a duplication event occurred.

3.8 Gene enrichment analysis

In the event that multiple candidate variants, no candidate variants or no candidate CNVs were identified, PANTHER was used to analyze gene lists and determine whether there was overrepresentation of certain protein families, molecular functions, biological processes, or cellular components and pathways. The list of genes remaining after initial rare non-synonymous filtering (non-synonymous sequence ontology, Read Depth ≥ 30 , Minor Allele Frequency $\leq 1\%$) were used as input. From analysis on the gene lists of patients with lipodystrophy and early-onset obesity, two partial lipodystrophy patients were found to harbour variants with significant overrepresentation in specific biological pathways. More specifically, for patient 14318, gene overrepresentation was identified in the biological pathways of *O*-glycan processing and tissue homeostasis. For patient 11618, overrepresentation was significant in the interferon-gamma-mediated signaling pathway.

3.9 Statistical analysis: Brachydactyly case

To strengthen the association of the *IHH* variant with the brachydactyly phenotype displayed in the family, family-based association analyses were performed using the Wilcoxon-Mann-Whitney test and anthropometric measurements. From these tests, the *IHH* variant was associated with shortened length of middle phalanges by 21.1% ($P < 0.001$), of palms by 13.8% ($P < 0.01$), of middle digit to palm ratio by 6.8% ($P < 0.03$) and of stature by 9.5% ($P < 0.001$), consistent with the relative mild clinical phenotype displayed by the patient (Figure 3.9).

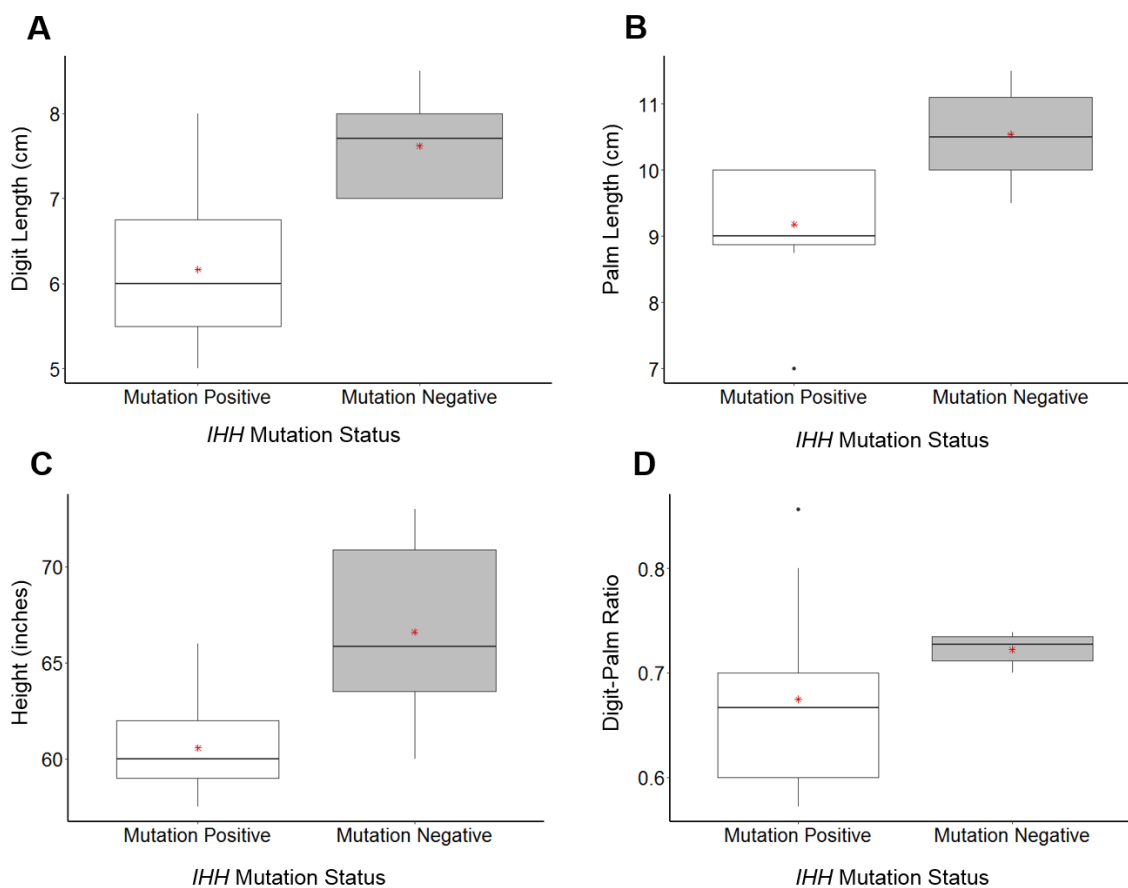


Figure 3.9. Comparison of anthropometric measures between those positive for the *IHH* mutation and those negative for the *IHH* mutation. Physical measurements were taken for all available family members with brachydactyly. Validation of the *IHH* mutation was performed using Sanger sequencing. Statistical analysis was performed using the Wilcoxon-Mann-Whitney test on SAS v9.4. Box plots were generated using the ggPlot2 package in R project v3.4.3. Panel A represents the digit length (cm) for *IHH* mutation positive and mutation negative individuals taken from the base of the third finger to the tip. B shows the palm length (cm) for *IHH* mutation positive and negative individuals taken from the base of the palm to the top of the palm. C displays the height in inches for mutation positive and negative individuals. D shows the digit-palm ratio for mutation positive and negative individuals, where the digit and palm lengths were taken as described previously.

Chapter 4 – Discussion

4 Summary of findings

In this MSc Thesis, I have identified genetic variants underlying several rare diseases using NGS technologies. More specifically, each case was subjected to either LipidSeq™ targeted sequencing or WES. Bioinformatics processing and *in silico* analyses were then utilized to identify potential disease-causing variants from the LipidSeq™ and WES data.

From LipidSeq™ targeted sequencing of 153 patients clinically diagnosed with lipodystrophy, 58 patients were found to harbour 81 variants within 14 lipodystrophy-associated genes. Furthermore, seven patients were found to harbour CNVs in lipodystrophy-associated genes. This is the first report of systematically screening for CNVs in patients with lipodystrophy, and these are the first CNVs reported in lipodystrophy patients in the literature.

From WES and bioinformatics analysis, candidate variants were identified in 13 of the 20 patients with lipodystrophy, and two of the five patients with extreme early-onset obesity. For the rare disease cases ascertained through the Blackburn Cardiovascular Genetics Laboratory, namely the brachydactyly and autosomal recessive cerebral atrophy-like, candidate variants were identified in both cases. Last, for the rare disease cases ascertained through the CARE4RARE and FORGE Canada Consortiums, eight out of the 11 cases had candidate variants identified.

Of the candidate variants found, several were validated using Sanger sequencing and many were found to co-segregate with disease status in the pedigrees. These variants have been prioritized and will be considered for future functional studies. Of the candidate CNVs identified, the CNV in patient 1120 (*CREB3L3* and surrounding genes) was confirmed using the independent methods of WES and CytoScan™ HD. Secondary validation (through Sanger sequencing) of the *PLIN1* CNV was also performed using Sanger sequencing.

The context of these findings are discussed within the subsequent sections in this Chapter. Additionally, I discuss the implications, limitations, and future directions for this study, and the benefits and caveats of using NGS for identifying the genetic basis of rare diseases.

4.1 LipidSeq™

LipidSeq™, the custom-designed targeted NGS panel, was utilized for the sequencing of lipodystrophy-associated genes in patients with various forms of lipodystrophy. From the DNA of 153 patients with lipodystrophy, LipidSeq™ targeted sequencing, bioinformatics processing and variant curation were performed, and 81 variants harboured within 14 genes were identified among the 58 patients. Furthermore, from CNV calling and prioritization, seven novel CNVs were identified within seven patients.

4.1.1 Unexpected variation in *POLD1*

From the LipidSeq™ data, despite the fact that many of the causative genetic variants were harboured within common lipodystrophy-associated genes (i.e. *LMNA*, *PPARG*, *PLIN1*, etc.); an unexpected 14 variants were detected within *POLD1*. The *POLD1* gene, encoding polymerase (DNA-directed), $\delta 1$, is a catalytic and proofreading subunit of DNA polymerase δ (Pol δ), involved in lagging strand DNA synthesis during replication^{135,136}. In order to aid in high fidelity DNA replication, synthesis and repair, *POLD1* works with accessory proteins *POLD2*, *POLD3* and *POLD4*. Together with these subunits, the heterotetramer interacts with Proliferating Nuclear Cell Antigen (PCNA) and the clamp loader, Replication Factor C (RFC) to form the polymerase holoenzyme. In addition to its role in lagging strand synthesis, Pol δ plays a role in proofreading (through exonuclease activity), and lesion repair resulting from mutagenic exposure. Furthermore, Pol δ is important in DNA repair mechanisms including base excision repair, mismatch repair, double strand break repair, and nucleotide excision repair. Mutations in Pol δ have been associated with genomic instability and tumourigenesis. Variants have also been identified in the developmental disorders mandibular hypoplasia, deafness, progeroid features and lipodystrophy (MDPL), and Werner's syndrome¹³⁷.

MDPL is an extremely rare, heterogeneous, multi-system disorder featuring premature aging of the tissues and organs. Most importantly, MDPL features variable lipodystrophy. In 2013, Weedon *et al.* reported *POLD1* variants in four individuals featuring a prominent, atypical lipodystrophy phenotype¹³⁶. Using deletion mutants, they showed that *POLD1* variants abolished polymerase activity while maintaining exonuclease activity.

Due to the similar adipose tissue distribution of MDPL and the lipodystrophy seen in our patient cohort, I encourage LipidSeq™ targeted sequencing to differentiate lipodystrophy associated with *POLD1* versus lipodystrophy caused by genes such as *LMNA*, *PPARG* and *PLINI*. Differentiation may change treatment, thereby altering patient prognosis. Moreover, due to the higher than expected prevalence of *POLD1* variants in our cohort, I suggest that lipodystrophy associated with *POLD1* variants may be underestimated, further encouraging investigation of *POLD1* in addition to classic lipodystrophy-associated genes in the future.

4.1.2 Copy-number variation findings

The advent of LipidSeq™ targeted sequencing has shown numerous benefits for patients with metabolic disease and clinical medicine, notably for the clarification of disease phenotypes caused by pathogenic variants. Understudied however, was the contribution of CNVs to lipodystrophy phenotypes. Using LipidSeq™ data and the CNV Caller Tool in VarSeq®, I detected CNVs by analysing depth of coverage readings from next-generation sequencing results. From the 153 uncharacterized lipodystrophy patients who were evaluated for potential CNVs, I identified seven novel CNVs that have never been reported in the literature or in any control database. Three individuals carried heterozygous deletions (patients 1099, 2108 and 15529), while four carried heterozygous duplications (patients 1120, 4192, 14557 and 15678).

4.1.2.1 Patient 1099

In a biological male with FPLD, a *PLINI* heterozygous deletion of exon 2 was identified. Using Sanger sequencing, this *PLINI* deletion was confirmed to be a 1646-bp deletion.

Perilipin-1, encoded by the *PLIN1* gene, is a lipid droplet coat protein expressed in adipocytes, involved in the facilitation of lipolysis. Loss-of-function variants in *PLIN1* are associated with autosomal dominant FPLD4, a subtype of partial lipodystrophy featuring variable loss of adipose tissue from the extremities¹³⁸. As mutant *PLIN1* proteins have an increased propensity for degradation and heterozygous knockout mice display significantly reduced fat mass, I suggest that the *PLIN1* exon 2 deletion is causing the lipodystrophy in this patient¹³⁹.

4.1.2.2 Patient 2108

In patient 2108 – a biological female clinically diagnosed with autosomal dominant Dunnigan-Köbberling syndrome – a heterozygous deletion of the 5'UTR–exon 1 region of *AGPAT2* was identified. Dunnigan-Köbberling syndrome is a form of lipodystrophy characterized by the loss of adipose tissue from the body, sparing the face¹⁴⁰. As homozygous variants of *AGPAT2* are associated with CGL (a more severe, recessive form of complete lipodystrophy), I suggest that this heterozygous CNV event may be causing the milder, partial lipodystrophy phenotype in this patient.

4.1.2.3 Patient 15529

In a biological female featuring FPLD, a heterozygous deletion of alternative non-coding exon 1 in *CIDEA* was detected. Generally, recessive variants in *CIDEA* are associated with FPLD5. As this is a heterozygous deletion, it is predicted that this *CIDEA* CNV is contributing to the FPLD phenotype rather than causing it. As alternative non-coding exons play a large role in alternative splicing, I suggest that this CNV affects the splicing of the *CIDEA* gene. Further, as a candidate variant (c.T53G, p.V18G) was identified in the *AGPAT2* gene through LipidSeq™ sequencing, I propose that the *CIDEA* CNV and *AGPAT2* variant are acting synergistically to cause the partial lipodystrophy phenotype.

4.1.2.4 Patient 4192

In another biological female with FPLD, I detected a duplication of the entire *APOA5* and *APOA4* genes. *APOA5* encodes apolipoprotein A-V, a low abundance plasma protein that

plays a role in triacylglycerol homeostasis¹⁴¹. *APOA4* encodes apolipoprotein A-IV, a component of chylomicrons and high-density lipoproteins that plays a protective role in reverse-cholesterol transport¹⁴². GWASs have identified multiple SNPs in *APOA5* that are associated with elevated TGs. Furthermore, studies on mutant *Apoa5*^{-/-} mice demonstrate aberrant lipoprotein binding and TG modulation functions¹⁴³. Due to the duplication of the entire *APOA5* coding gene, I suggest that the duplication may be contributing towards the lipodystrophy phenotype in this patient. Furthermore, I suggest the duplication may be synergistically working with the *APOA4* duplication and an additional aberrant gene product in order to cause the lipodystrophy present in this patient. In the future, I suggest WES to determine added genetic factors causing the lipodystrophy phenotype.

4.1.2.5 Patient 14557

In a patient (14557) with FPLD, I identified a novel duplication event in the *MFN2* gene. The *MFN2* gene encodes the mitofusin 2 protein, involved in the morphology of mitochondria. As the mitochondria require flexibility in morphology to maintain mitochondrial function, the mitofusin 2 protein is generated in many tissues and cells including muscle, the spinal cord, nerves, and peripheral nerves. Variants in *MFN2* are the primary cause of autosomal dominant Charcot-Marie-Tooth (CMT) disease, type 2A¹⁴⁴. CMT is a heterogeneous disorder, characterized by peripheral neuropathy¹⁴⁵. An additional clinical disease associated with *MFN2* mutations is multiple symmetric lipomatosis (MSL), a mitochondrial disorder associated with abnormal brown fat metabolism. Patients with MSL feature lipomatosis (multiple lipomas or benign fat tissue tumours on the body) as well as lipodystrophy. MSL has been attributed to a p.R707W (c.2119C>T) mutation¹⁴⁶. As our patient was found to carry a heterozygous p.R707W mutation in *MFN2* in addition to a heterozygous duplication, I suggest this variant and CNV are cooperatively contributing to the lipodystrophy phenotype. Additionally, due to the overlap between FPLD and MSL in terms of lipoatrophy, I suggest that the CNV duplication and variant in *MFN2* are causing an atypical form of lipodystrophy. Follow-up studies are required to confirm the pathogenicity of the CNV event and the effects of the CNV combined with the p.R707W variant.

4.1.2.6 Patient 15678

From CNV detection, a biological female with FPLD was found to harbour a heterozygous duplication of *PLIN1* (exon 3 – 3'UTR). As in the case with patient 1099, I propose that the FPLD phenotype in this patient can be explained by the *PLIN1* duplication. As both excess and insufficient lipid storage in white adipose tissue lipid droplets can result in dyslipidemia, insulin resistance and diabetes, I suggest the *PLIN1* duplication is contributing to the lipodystrophic phenotype of this patient¹⁴⁷. Furthermore, as a novel heterozygous *AGPAT2* variant (c.T53G, p.V168G) was also detected in this patient's genome from LipidSeq™, I suggest this *AGPAT2* variant and *PLIN1* duplication are both contributing to the patient's lipodystrophy.

4.1.2.7 Patient 1120

In a biological male clinically diagnosed with FPLD, a duplication of the *CREB3L3* gene was identified. The cyclic AMP-responsive element-binding protein 3-like protein 3 (*CREB3L3*) gene encodes a transcription factor that activates unfolded protein response target genes during endoplasmic reticulum stress. Activated by cAMP stimulation, CREB3L3 binds the cAMP response element (CRE) and box-B element to activate transcription. In addition, CREB3L3 is involved in fatty acid oxidation and ketogenesis, and induced during fasted or insulin-resistant states¹⁴⁸. Variations in *CREB3L3* cause hypertriglyceridemia, where *Creb3l3*^{-/-} mice have elucidated the mechanism as due to inefficient catalysis of triglyceride clearance¹⁴⁹. As the entire *CREB3L3* gene was duplicated, I speculated that neighbouring genes (such as the *SIRT6* gene associated with lipid metabolism) were also affected. To confirm this prediction, I pursued WES. The results of the CNV detection from WES data are further discussed in Section 4.2.13.

In addition to the *CREB3L3* CNV, patient 1120 also harbours two missense mutations, each in *PLIN1* (c.G814A, p.V272M) and *PTRF* (c.C168A, p.S56R) predicted as deleterious from *in silico* prediction tools. Moreover, *PLIN1* and *PTRF* are associated with FPLD5 and CGL4, respectively, each associated with lipodystrophy. All together, I suggest that the *CREB3L3* duplication, along with the variants in *PLIN1* and *PTRF*, are

modulating or causing the lipodystrophy phenotype and abnormal lipid profile present in this patient.

4.1.3 Benefits of LipidSeq™

The use of LipidSeq™ targeted sequencing offers several benefits. First, LipidSeq™ is a targeted panel, with the flexibility to adapt with the growing body of literature. As novel lipodystrophy genes are continually being discovered (i.e. adrenoceptor alpha 2A [*ADRA2A*] mutations in FPLD¹⁵⁰), the LipidSeq™ has the ability to stay current and include all known genes associated with dyslipidemia. Second, as LipidSeq™ is restricted to specific genes, it lessens the probability of incidental findings associated with other disorders. With concerns regarding the proper disclosure of incidental findings, the use of LipidSeq™ decreases the possibility of these findings and the potential for genetic discrimination in the workplace or from insurers. Third, in comparison to Sanger sequencing, LipidSeq™ can rapidly provide a higher diagnostic yield at a cost-effective rate. With LipidSeq™ data generation taking less than a week, LipidSeq™ offers a quick turnaround from bench-to-bedside. Last and most importantly, LipidSeq™ allows for the genetic diagnosis of several metabolic disorders. Notably in the 153-person lipodystrophy cohort, LipidSeq™ sequencing led to the genetic diagnosis of 40% of patients through variants or CNVs, speaking to the utility of LipidSeq™ in the clinic. For these patients, the results may have actionable consequences in the realms of treatment, management or predictive recurrence risks, altogether improving patient outcome and prognosis.

4.1.4 Limitations of LipidSeq™

Although I was able to use the LipidSeq™ targeted gene panel to identify the genetic basis (variants and CNVs) of 40% of patients with lipodystrophy, 60% of cases remain genetically unexplained, speaking to the limitations of LipidSeq™. This gap in variant and/or CNV discovery is due to the inherited constraints of LipidSeq™ targeted sequencing. First, with LipidSeq™ there is low-to-no coverage of the untranslated and non-coding regions, making the identification of candidate variants in those regions

impossible. This is a drawback of LipidSeq™ as disease-associated variants have been known to be harboured in the UTR. For example, in 2017 Wang *et al.* identified a *TBX5* 3'UTR variant driving congenital heart disease in a Han Chinese population¹⁵¹. Second, with targeted NGS-based approaches, although there can be identification of CNV events, many CNVs can be missed depending on the location of the event. More specifically, numerous CNV events occur in highly repetitive regions through genomic instability, non-allelic HR or NHEJ¹⁵². Due to these areas being relatively unamenable to mapping (i.e. due to guanine-cytosine rich regions), these areas are either missed or not targeted through LipidSeq™ sequencing. Last, despite the fact that the LipidSeq™ is flexible and can be modified to incorporate any gene of interest, variant discovery from LipidSeq™ is restricted to *a priori* knowledge. Therefore, it is impossible for LipidSeq™ sequencing to allow for the identification of novel disease-associated genes. Albeit the utility of LipidSeq™ in identifying variants in known lipodystrophy-associated genes, it is necessary to use broader methods such as WES in order to identify novel pathogenic genes and variants, not previously implicated in lipodystrophy.

4.2 Whole-exome sequencing of patients with rare metabolic diseases

In patients with prior Sanger sequencing or LipidSeq™ sequencing in which no candidate variants were found, WES was pursued in order to identify novel disease-causing variants. Rare metabolic disease patients – namely 20 patients with lipodystrophy and five with extreme early-onset obesity – were subjected to the WES process. Following WES, data processing and bioinformatics analysis were performed to identify and prioritize putative disease-causing variants. After individual analysis, the exomes of individuals with lipodystrophy and extreme early-onset obesity were evaluated for shared variation. From these analyses, 13 of the 20 patients with lipodystrophy and two of five patients with extreme early-onset obesity had candidate variants identified. Furthermore, variation in the *AR* gene was detected in four patients with partial lipodystrophy and one patient with extreme early-onset obesity, creating a connection between the two opposite phenotypes.

4.2.1 Lipodystrophy

WES was performed for 20 patients with variable forms of lipodystrophy. WES identified variants within 13 patients discussed herein.

4.2.1.1 Patient 1804

In a biological female clinically diagnosed with CGL, a homozygous variant in the iron-sulfur cluster assembly enzyme (*ISCU*, c.19_20delinsGG, p.Phe7Gly) was identified. The *ISCU* protein is a scaffold protein required for the *de novo* synthesis of iron-sulfur clusters (Fe-S) in mitochondria. Specifically, *ISCU* is responsible for the maturation of mitochondrial and cytoplasmic proteins¹⁵³. These Fe-S clusters are required for the function of numerous proteins, including those required for iron regulation and DNA repair. Proteins with these Fe-S clusters are also required for energy production in the mitochondria¹⁵⁴. As Fe-S clusters are required cofactors for proteins involved in several biological processes, it is likely that dysfunction of the iron-sulfur cluster biogenesis pathway underlies several human diseases.

Mutations in *ISCU* are causative of iron-sulfur deficiency myopathy, or myopathy with exercise intolerance. This disease features severe, progressive muscle weakness, severe exercise intolerance and cardiomyopathy^{155,156}. As patient 1804 exhibits exercise intolerance, specifically exertion-related chest complications, I suggest that this novel *ISCU* variant is contributing towards the phenotype present in this patient. As a candidate variant has not been identified as causative of her lipodystrophy, I suggest further analysis of the exome data using less stringent criteria (i.e. lower read depth) to determine the genetic basis of the lipodystrophy in this patient. I also suggest considering other genetic factors (discussed in Section 4.8) as causing the lipodystrophy in this patient.

4.2.1.2 Patient 8615

From WES, patient 8615 – a biological female featuring general lipodystrophy, diabetes, iron-deficient anemia, diabetic neuropathy, central obesity and coronary artery disease –

was found to harbour a homozygous variant in the heparin sulfate proteoglycan 2 gene or *HSPG2* (c.3707C>A, p.Ala1236Glu) encoding for the perlecan protein. The perlecan protein belongs to the extracellular matrix basement membrane, where it supports cells in many tissues. Additionally, perlecan is found in cartilage and plays a role in cell signaling, angiogenesis and muscle contraction¹⁵⁷. Perlecan also plays a large role in the cardiovascular system, where it plays a role in lipid metabolism by interacting with lipoprotein lipase and apolipoprotein B. Variants in *HSPG2* have been linked to coronary artery disease, insulin sensitivity, adipose tissue defects and metabolic syndrome, all secondary features of the proband^{158,159}. Therefore, I predict that the *HSPG2* variants are contributing towards the lipodystrophy phenotype of this patient. As the definitive cause of the lipodystrophy present in this patient has not been elucidated, I suggest further analysis of the exome of this patient using less stringent criteria, as discussed prior.

4.2.1.3 Patient 1088

In patient 1088, a biological female featuring FPLD, high cholesterol, diabetes and high blood pressure, two candidate variants were detected and are predicted to cause the phenotype of the patient. The first is a missense variant (c.2468G>C, p.Arg823Pro) in *PRDM16* encoding the PR domain containing 16 protein, a transcription co-regulator that binds and activates PPAR γ to develop brown adipocytes from brown adipose tissue (BAT). When functioning normally, BAT oxidizes chemical energy to produce heat in protection against obesity. An association between *PRDM16* variants and a metabolic syndrome phenotype in a Uygur Chinese population has been reported, which suggests it may be involved in the development of lipodystrophy, as lipodystrophy is classified as a metabolic disorder¹⁶⁰. The second candidate variant (c.1048C>T, p.Arg250Ter) is a classified pathogenic variant within *LDLR*, encoding the low-density lipoprotein receptor. LDL receptors are cell surface proteins involved in the receptor-mediated endocytosis of LDL cholesterol. Under normal circumstances, LDL is transported into the cell through receptor binding, and carried to the lysosome where it is subsequently degraded. Mutations in *LDLR* cause familial hypercholesterolemia (FH), an autosomal dominant disorder characterized by elevation of serum cholesterol bound to LDL, promoting the deposit of cholesterol in the arteries, tendons, and skin¹⁶¹.

In sum, these findings suggest that patient 1088 is not only affected with FPLD but may also have FH, explaining the high cholesterol levels. Considering the chance of developing familial partial lipodystrophy is approximately one in 50,000 and the prevalence of FH is one in 300, this makes the phenotype of patient 1088 one in 15 million. This illustrates the power of WES in the clinic, as it would be difficult for a clinician to diagnose a disorder with such low and improbable statistical odds.

4.2.1.4 Patient 3458

In patient 3458, a female clinically diagnosed with FPLD, a novel, heterozygous variant in the *ZMPSTE24* gene was detected (c.1199A>G, p.Asn400Ser). The *ZMPSTE24* protein is a zinc metalloproteinase STE24 with a role in processing farnesylated proteins¹⁶². *ZMPSTE24* also ensures the correct processing and maturation of lamin A. Mutations in *ZMPSTE24* are known to cause mandibuloacral dysplasia with type B lipodystrophy (MADB). This condition features bone abnormalities, patchy skin colouring, and loss of fat tissue across the body¹⁶³. Mutations in *ZMPSTE24* are also associated with severe progeria and restrictive dermopathy (thin, easily eroded skin and joint contracture). Complete loss-of-function alleles are associated with restrictive dermopathy, whereas residual or partial activity of *ZMPSTE24* correlates with increased disease severity and MADB or progeroid features¹⁶⁴. We suggest that this *ZMPSTE24* variant is contributing towards the lipodystrophy phenotype of this patient. As a note, Sanger sequencing performed 12 years ago missed the variant in *ZMPSTE24*. The recent identification of this variant using WES highlights the power of NGS in variant identification.

4.2.1.5 Patient 3969

In patient 3969, a biological female featuring FPLD, candidate variants were identified in *FLNC* (c.6988G>A, p.Gly2330Ser) and *GALC* (c.334A>G, p.Thr112Ala). *FLNC* encodes gamma filamin, a protein that aids in the crosslinking of actin filaments in the cortical cytoplasm, involved in the anchoring of membrane proteins for the actin cytoskeleton. As gamma filamin is muscle-specific, dominant mutations in *FLNC* cause myofibrillar

myopathy, a myopathy characterized by muscle weakness and the atrophy of the upper and lower extremities¹⁶⁵. *GALC* encodes galactosylceramidase, an enzyme involved in breaking down galactolipid fats in the nervous system, epithelial cells of the small intestine and colon, and kidneys¹⁶⁶. Defects in *GALC* are associated with Krabbe disease, a disease featuring mobility restrictions due to damage to myelin-forming cells in the nervous system. Due to the loss of myelin, nerves in the brain cannot properly transmit signals, leading to Krabbe disease^{167,168}. As the primary health concern of patient 3969 is peripheral neuropathy, I suggest that the variants in *FLNC* and *GALC* are contributing towards the neuropathy phenotype. Additional analysis should be done to determine whether these genetic variants interact to amplify the neuropathy in this patient. Furthermore, future directions for this patient should include continued analysis of her exome data using less stringent criteria to determine the genetic basis of her lipodystrophy.

4.2.1.6 Patient 4116

After WES and bioinformatics analysis, patient 4116 – a biological female clinically diagnosed with FPLD – had two *LDLR* candidate variants (c.2252G>A, p.Arg751Gln; c.2231_2232delinsAG, p.Arg744Gln) identified in her exome. Similar to patient 1088 as discussed in Section 4.2.1.3, the *LDLR* variants are likely contributing to the high cholesterol levels in this patient. In addition to the *LDLR* variants, a novel variant in the *AR* gene was also detected (c.1152C>G, p.His384Gln) in the patient's exome (discussed in Section 4.2.3). Together, the *LDLR* and *AR* variants are likely contributing to the abnormal lipid profile of the patient. Future work should focus on identifying the genetic basis of the lipodystrophy phenotype of the patient. This can be done by using less stringent criteria when analysing the exome data or considering alternative genetic factors (as discussed in Section 4.8).

4.2.1.7 Patient 4665

In a biological female with FPLD (patient 4665), four candidate variants were identified from WES. Two variants were detected in *ABCA4* (c.1610G>A, p.Arg537His;

c.2969G>A, p.Gly990Glu) and two in *WRN* (c.2983G>A, p.Ala995Thr; c.1149G<T, p.Leu383Phe). *ABCA4* encodes a member of the ATP binding cassette transporters. Involved in lipid transport, *ABCA4* is associated with the import of lipid substrates such as vitamins, phospholipids, glycolipids, fatty acids, cholesterol and bile salts¹⁶⁹. On the other hand, *WRN* encodes the Werner protein, where variants in *WRN* cause an aging disease known as Werner syndrome, featuring abnormal fat deposition including the loss of adipose from the distal limbs¹⁷⁰. As Werner syndrome is caused by pathogenic recessive variants in the *WRN* gene and this patient harbours two heterozygous variants in *WRN*, I suggest that patient 4665 may feature an atypical form of Werner syndrome. I also propose that the *ABCA4* gene variants are contributing to the lipodystrophy phenotype of the patient. Additional studies should be done to confirm the pathogenicity of the variants and their role in the lipodystrophy phenotype (see Section 4.8 for details).

4.2.1.8 Patient 5662

The DNA of patient 5662, a female featuring FPLD was subjected to WES and bioinformatics processing. After analysis, a homozygous splicing variant was identified in *ABCG8* (c.1412-8delinsTT). This *ABCG8* splicing variant is predicted to alter the wild type splice acceptor site from multiple *in silico* splicing prediction tools, is considered novel in multiple control databases, and has not been previously reported in the literature.

The *ABCG8* gene encodes sterolin-2, half of the protein, sterolin. In the body, sterolin is involved in eliminating fatty components of plant-based foods that cannot be used in human cells¹⁷¹. Mutations in *ABCG8* have been identified in individuals with sitosterolemia, a condition characterized by the accumulation of plant sterols and the impaired ability to remove cholesterol from the body. Furthermore, individuals with sitosterolemia feature atherosclerosis, xanthomas, and blood abnormalities¹⁷². I propose that the *ABCG8* gene is contributing to the abnormal lipid profile of the patient; however, further analysis of the exome data is required to determine the genetic cause of the patient's lipodystrophy.

4.2.1.9 Patient 2025

Patient 2025 is a biological female that was clinically diagnosed with lipodystrophy. Upon review of the patient's clinical data, it was discovered that the patient also has dermatomyositis – a rare genetic inflammatory disease causing severe myopathy, painful skin rashes, sensitivity to light and calcinosis (calcium deposits in the muscle). Furthermore, the patient suffers from severe insulin resistance and proteinuria. Due to the fact that this patient has dermatomyositis in addition to lipodystrophy, variants associated with dermatomyositis were also prioritized during exome analysis.

In this patient, three candidate variants were identified within three candidate genes: *FHL1* (c.46G>T, p.Gly16Trp), *ANLN* (c.1205C>T, p.Pro402Leu) and *AR* (c.1369_1380delGGCGGCGGCGGC, p.Gly470_Gly473del). The *FHL1* gene (four and a half LIM domains 1) encodes three isoforms that play critical roles in skeletal muscle and cardiac muscle movement. More specifically, FHL1A plays a key role in sarcomere assembly, FHL1B plays a role in the nuclear envelope, and FLH1C is thought to play a role in the function of skeletal and cardiac muscles¹⁷³. Dominant mutations in *FHL1* have been linked to reducing-body myopathy (RBM), a rare myopathy featuring the progressive weakness of the muscles, a rigid spine, and a progressive disease course¹⁷⁴. The *ANLN* gene encodes anillin, an actin binding protein with a role in cell growth and migration. Variants in *ANLN* are causative of focal segmental glomerulosclerosis-8, a renal disease featuring proteinuria and the progressive decline in renal function¹⁷⁵. The *AR* variant is discussed in Section 4.2.3.

Due to the association with myopathy, proteinuria and renal function and the severe myopathy and renal dysfunction in the patient, I suggest that the novel *FHL1* and *ANLN* variants are contributing towards the disease phenotype of the patient. Furthermore, I suggest that the *AR* gene variant is contributing to the metabolic profile of patient 2025.

4.2.1.10 Patient 3011

Patient 3011 is a biological female featuring FPLD, insulin resistance, dyslipidemia, hirsutism (overgrowth of hair) and hypertriglyceridemia. From WES, I identified three

candidate gene variants contributing to the phenotype of the patient: *LPL* (c.775G>A, p.Asp259Asn), *PAPSS2* (c.1114G>T, p.Val372Phe), and *AR* (c.170_171insGCAGCAGCA, p.Gln80_Glu81insGlnGlnGln). The *LPL* gene encodes lipoprotein lipase, an enzyme found within cells that line capillaries in muscles and adipose tissue. LPL plays a key role in breaking down fat into triglycerides, which are then transferred from organs to the blood through lipoproteins. Mutations in *LPL* have been associated with many conditions, including atherosclerosis, insulin resistance, and dyslipidemia associated with diabetes¹⁷⁶. The *PAPSS2* gene encodes 3'-phosphoadenosine 5'-phosphosulfate synthase 2, a sulfate donor co-substrate required for sulfotransferase (SULT) enzymes. These SULT enzymes catalyze many endogenous and exogenous compounds, such as drugs and xenobiotics¹⁷⁷. Mutations in *PAPSS2* are associated with brachyolmia type 4 with mild epiphyseal changes (BCYM4). Furthermore, variants in *PAPSS4* have been associated with hirsutism, androgen excess and polycystic ovary syndrome, all clinical features of patient 3011¹⁷⁸. The *AR* variant is discussed in detail in Section 4.2.3.

All together, I predict that the *LPL*, *PAPSS2* and *AR* gene variants are all contributing to the patient's phenotype. More specifically, it is likely that the *LPL* variant is contributing to the insulin resistance, dyslipidemia and hypertriglyceridemia, whereas the *PAPSS2* variant is contributing to the polycystic ovary syndrome and hirsutism. The *AR* variant is likely also contributing to the hirsutism and involved in the metabolic profile of the patient. Further analysis is required to confirm the pathogenicity of the novel variants and identify the genetic basis of the patient's lipodystrophy (as discussed in Section 4.8).

4.2.1.11 Patient 14318

Patient 14318 is a biological female in her teens, clinically diagnosed with FPLD. From WES and bioinformatics analysis, several candidate variants were detected. These include variants in *NEUROD1* (c.170G>A, p.Gly57Glu), *BSCL2* (c.1280T>C, p.Leu427Pro), *APOB* (c.6936_6937delinsTG, p.Ile2313Val), *PCSK9* (c.137G>T, p.Arg46Leu), and *PYGM* (c.148C>T, p.Arg50Ter) and *AR* (c.185A>T, p.Gln62Leu; c.1369_1380delGGCGGCGGCGGC, p.Gly470_Gly473del).

NEUROD1, encoding the neurogenic differentiation 1 protein, is a transcription factor involved in cell type determination during development. *NEUROD1* is also important in the regulating insulin gene expression. Mutations in *NEUROD1* are associated with mature-onset diabetes of the young 6 (MODY6) and autosomal dominant diabetes mellitus¹⁷⁹. The *BSCL2* gene encodes seipin, an endoplasmic reticulum protein involved in late pre-adipocyte differentiation and lipid droplet formation¹⁸⁰. Mutations in *BSCL2* are associated with CGL2, as discussed in Chapter 1. *APOB* encodes apolipoprotein B, the main apolipoprotein on low density lipoproteins (LDLs) and chylomicrons¹⁸¹. Defects in *APOB* are associated with autosomal dominant hypercholesterolemia. The prohormone convertase 9 (*PCSK9*) gene encodes a serine protease involved in the elevation of plasma LDL cholesterol and reduction of hepatic and extrahepatic LDL receptor levels¹⁸². Mutations in *PCSK9* are causative of autosomal dominant hypercholesterolemia, 3, featuring significantly high cholesterol levels. The glycogen phosphorylase, muscle associated (*PYGM*) gene or the muscle isoform of glycogen phosphorylase is involved in the breakdown of glycogen during glycogenolysis¹⁸³. Mutations in *PYGM* are associated with glycogen storage disease type V (GSDV), a disease caused by the impaired breakdown of glycogen. Due to this defective enzyme being unable to break down glycogen, muscle cells cannot produce energy and become easily fatigued¹⁸⁴. The *AR* variant is further discussed in Section 4.2.3.

Together, I suggest that the variants within *NEUROD1*, *BSCL2*, *APOB*, *PCSK9*, *PYGM* and *AR* are all contributing to the severe lipodystrophy phenotype of this young patient. Further research is required to determine the pathogenicity of the novel variants and determine whether these variants work synergistically (see Section 4.8 for details).

4.2.1.12 Patient 4194

In patient 4194, a female clinically diagnosed with acquired lipodystrophy, a variant in the *WRN* gene (c.243T>G, p.Phe81Leu) was detected from WES. Variants in the *WRN* gene are associated with Werner syndrome or adult progeria, characterized by dramatic aging, abnormal fat deposition, diabetes mellitus, diminished infertility and atherosclerosis¹⁸⁵. More specifically, patients with Werner syndrome experience fat loss

from the extremities with excess fat in the truncal region, similar to patients with FPLD¹⁷⁰. As Werner syndrome is generally caused by homozygous variants, I suggest that this heterozygous variant is causing the milder phenotype seen in this patient.

As patient 4194 has a form of acquired lipodystrophy generally caused by environmental effects, I also propose the possibility that this heterozygous variant in *WRN* could have been activated by environmental factors, leading to the onset of lipodystrophy in this patient. Note that this patient was not processed on the LipidSeq™ gene panel prior to WES.

4.2.1.13 Patient 1120

After identifying a *CREB3L3* duplication from LipidSeq™ targeted sequencing, the DNA of patient 1120 was subjected to WES to investigate whether additional neighbouring genes were also duplicated. From WES data and CNV analysis, the duplication in patient 1120 was found to span six genes: protein inhibitor of activated STAT 4 (*PIAS4*), zinc finger and BTB domain containing 7A (*ZBTB7A*), mitogen-activated protein kinase kinase 2 (*MAP2K2*), *CREB3L3*, *SIRT6*, and ankyrin repeat domain 24 (*ANKRD24*). Interestingly, one of the duplicated genes is the *SIRT6* gene, encoding the sirtuin-6 protein.

The sirtuin-6 or SIRT6 protein is a protein deacetylase and mono-ADP ribosyltransferase enzyme activated under stress conditions¹⁸⁶. SIRT6 is a major player in many pathways associated with aging, such as glycolysis, lipid metabolism, inflammation and DNA repair¹⁸⁷. Mutations in *SIRT6* are a major cause of progeroid diseases, which feature lipodystrophy, kyphosis (excessive convex curvature of the spine), neurodegeneration, immune deficiency, and metabolic defects¹⁸⁸. Specific to metabolism, the SIRT6 protein is considered the epigenetic gatekeeper of glucose metabolism, playing a key role in glucose homeostasis. This role was demonstrated in SIRT6 knockout mice, where the complete loss of subcutaneous fat and insulin hypersensitivity led to early mortality^{189,190}. As our patient harbours a duplication of *SIRT6* in addition to *CREB3L3*, I propose that

the lipodystrophy phenotype and metabolic abnormalities featured by this patient are caused by this large CNV event.

4.2.2 Extreme early-onset obesity

The DNA of five children with extreme early-onset obesity were subjected to WES. Candidate variants were identified in two patients from WES and bioinformatics processing.

4.2.2.1 Patient 15544

In a male child with early-onset obesity, features of autism spectrum disorder, developmental delay and overgrowth, candidate variants were detected in three genes: *BLK* (c.244G>C, p.Gly82Arg), *MC4R* (c.751A>C) and *AR* (c.204_205insA, p.Gln69fs). Further discussion on the *AR* gene is in Section 4.2.3.

The *BLK* gene encodes a non-receptor tyrosine kinase involved in cell proliferation and differentiation. Playing a role in B-cell receptor signaling and development, this protein is also known to stimulate insulin synthesis and secretion¹⁹¹. Mutations in *BLK* are associated with mature-onset diabetes of the young (MODY11), characterized by autosomal dominant diabetes and insulin-dependence¹⁹². As discussed in Chapter 1, the *MC4R* gene is a member of the melanocortin receptor family that interacts with adrenocorticotrophic hormones. Defects in this gene are known to cause autosomal dominant obesity. The *AR* variant is discussed further in Section 4.2.3. Together, I suggest that the defects in *BLK*, *MC4R* and *AR* are contributing to the extreme early-onset obesity phenotype of this patient by modulating normal lipid metabolism.

4.2.2.2 Patient 15858

In a female child featuring extreme early-onset obesity, hyperphagia and intellectual disability, a novel heterozygous missense variant (c.2645T>C, p.Ile882Thr) was identified within the *KMT2C* gene. *KMT2C* encodes the lysine-specific methyltransferase 2C, a histone methyltransferase that regulates gene transcription through modification of

chromatin structure. Particularly, KMT2C mediates the mono- and tri-methylation of histone H3 at lysine 4¹⁹³. In 2012, Kleefstra *et al.* discovered a dominant mutation in *KMT2C* as causative for Kleefstra syndrome-2, a neurodevelopmental disease characterized by delayed development, intellectual disability, childhood obesity, and prominent facial features¹⁹⁴. As the description of intellectual disability, obesity, and facial features (hypertelorism, synophrys, carp mouth with macroglossia) of Kleefstra syndrome matches the phenotype of the patient, I suggest that this child has a form of Kleefstra syndrome-2¹⁹⁵.

In addition to the dominant *KMT2C* variant, additional variants were identified in the Bardet-Biedl Syndrome 10 (*BBS10*) and SET domain containing (lysine methyltransferase) 8 (*SETD8*) gene. As variants in *BBS10* are associated with Bardet-Biedl syndrome and variants in *SETD8* are associated with adipogenesis and rodent obesity, I suggest that these gene variants are also modulating the phenotype and causing the extreme overgrowth and obesity features of this patient^{196,197}. Further studies (such as those outlined in Section 4.8) are necessary to confirm the pathogenicity of the *KMT2C* variant and determine whether the *SETD8* or *BBS10* variants are involved in the disease phenotype of this patient.

4.2.3 Lipodystrophy-obesity connection

With lipodystrophy and obesity, despite the fact that they sit on opposite ends of the spectrum (excess versus absent adiposity), they share many metabolic complications including diabetes mellitus, hypertriglyceridemia and hepatic steatosis. Furthermore, although both disorders are widely heterogeneous in terms of underlying genes and pathways, the diseases both feature severe insulin resistance, suggesting a commonality between the two disorders and the potential for shared genetic variation¹⁹⁸. To determine whether there was a connection between lipodystrophy and obesity, I analyzed all exomes for shared variation.

Altogether, over one million variants were identified in the 25 rare metabolic disease exomes of interest. From the 25 exomes, ultra-rare androgen receptor gene (*AR*)

mutations were detected in 20%. These variants include missense mutations in two familial partial lipodystrophy females, small deletions in three insulin resistant partial lipodystrophy females, and a frameshift in a male child with extreme early-onset obesity. Of these identified variants, four have never been recorded in the literature. Due to the burden of ultra-rare *AR* mutations in these rare metabolic phenotypes, I suggest that the *AR* gene may be the common link between lipodystrophy and extreme early-onset obesity.

The *AR* gene encodes a DNA-binding transcription factor involved in the regulation of gene expression. Activated by androgens, these androgen receptors allow the body to appropriately respond to these hormones, playing critical roles in the reproductive, muscular, skeletal, cardiovascular and central nervous systems. In the literature, dysfunctional androgen levels have been known to predispose individuals to metabolic abnormalities. More specifically, hyperandrogenism in women has shown to predispose to T2D, and hypoandrogenism in men has been linked to insulin resistance and metabolic syndrome^{199,200}. Furthermore, androgen receptor-deficient mice have been shown to develop obesity and metabolic dysfunction, further demonstrating the association between aberrant *AR* and metabolic disease features²⁰¹. Due to the high frequency of *AR* variants in this cohort, I suggest that *AR* mutations are contributing to these metabolic phenotypes by modulating the metabolic dysfunction in these patients.

Although these effects on metabolic syndrome are well established, our study is the first to suggest a direct association between *AR* mutations, lipodystrophy and obesity. In addition, because the burden of ultra-rare *AR* variants is specific to the N-terminal domain (NTD), I suggest an important role of the NTD in metabolic function. Identifying a common genetic factor for both lipodystrophy and extreme early-onset obesity can allow for the understanding of the pathways controlling insulin sensitivity and glucose uptake and allow for the development of therapeutic interventions for adipose tissue disorders. In the future, I suggest assessing all lipodystrophy and obesity patients for *AR* gene variation, and functional studies on the AR protein (additional details in Section 4.8).

4.3 Whole-exome sequencing of rare disease families

For families with a history of rare disease referred to the Blackburn Cardiovascular Genetics Laboratory (namely, BDA1 and cerebral atrophy-like), WES was performed on the DNA of the proband followed by bioinformatics processing and variant curation. From thousands of variants, one candidate variant was identified for each case. The successful identification of candidate variants for both cases highlights the capability of WES in identifying novel variants. Notably, the BDA1 case demonstrates the ability of WES to clarify the genetic basis of “cold cases”.

4.3.1 Brachydactyly type A1^A

Fifteen years ago, a 31-year-old female from Ontario was referred for medical consultation at Dr. Hegele’s clinic for a lifelong history of short stature and shortening of all digits of the upper and lower extremities. After discussion regarding her family history, it was discovered that the brachydactyly phenotype co-segregated in her family. After Sanger sequencing of all known brachydactyly-associated genes was performed at the Blackburn Cardiovascular Laboratory, the family remained genetically unexplained and the case was left untouched for 15 years. One patient sample was recently processed using WES and analyzed for variants of interest. In this family, we report autosomal dominant BDA1, characterized by short stature and shortened digits. Affected family members carry a heterozygous in-frame insertion in *IHH*, designated c.285_287dupGAA, p.Glu95_Asn96insLys, explaining their brachydactyly phenotype. This *IHH* variant is predicted to exert a damaging effect on protein function from multiple *in silico* prediction tools, co-segregates with disease status in the family, is considered novel in multiple control population databases, and has not been previously reported in the literature.

^A A version of this section has been submitted for publication⁸³.

4.3.1.1 *IHH* and brachydactyly type A1

Previously, a genetic basis of BDA1 had been identified, with causative mutations harboured in *IHH*^{77,202}. Located at the chromosomal position 2q35-36, *IHH* encodes the IHH protein, a member of the hedgehog family of proteins. Along with sonic and desert hedgehog, IHH acts as a signaling molecule that regulates patterning processes in both vertebrate and invertebrate development²⁰³. In vertebrates, the family is involved in limb polarity and chondrogenesis, with IHH playing a critical role in human skeletal development. IHH mutants show decreased chondrocyte maturation and proliferation, and the failure of osteoblast development in endochondral bones, highlighting the role that IHH plays in bone formation and cartilage differentiation²⁰⁴. The connection between skeletogenesis and BDA1 has been solidified in *Ihh*^{-/-} mice studies, where the loss of IHH signaling leads to reduction defects in the forelimbs and digits²⁰⁵.

Several mutations have been identified in the *IHH* gene. Dominant mutations in *IHH* are causative of BDA1, whereas recessive mutations in *IHH* have been linked to an extended phenotype, acrocapitofemoral dysplasia, which features short stature, short limbs and cone-shaped epiphyses²⁰⁶. All mutations responsible for BDA1 have been limited to the N-terminal active fragment of *IHH*⁷⁹. These mutations have been restricted to codon positions 95, 100, 131 and 154, with ours as no exception. The prevalence of BDA1 mutations at these codon positions suggests they are mutation hotspots, with high conservation of these amino acid positions existing between humans and species including mice, chicken, frog and zebrafish²⁰⁷. This feature solidifies the importance of the N-terminal region in bone development and differentiation, especially since mutations causing acrocapitofemoral dysplasia differ and are exclusively at the distal N and C terminal regions.

The IHH protein works through a feedback control mechanism. IHH binds to the PTC1/patched receptor, which in turn, binds to SMO/smoothened to activate the Gli complex of transcription factors. From there, these transcription factors continue to signal and regulate downstream genes affecting patterning. Because IHH acts later in

endochondral bone formation, mutations in *IHH* lead to the delay in bone growth and subsequent digit abnormality. Studies on *IHH*-null mice uncovered the association between *IHH* mutations and *BDA1*, as loss of *IHH* signaling leads to shortened forelimbs and abnormal digits. We propose that our novel variant is acting in the same manner, leading to digit reduction and shortened stature. More specifically, although our variant is the first insertion identified in *BDA1*, we suggest that our variant acts similar to the p.Glu95Lys missense mutation identified by Ma *et al.*, where the mutation results in the conversion of a negatively charged area to a positive area in a critical calcium-binding groove²⁰⁷. In turn, this leads to instability of the N-terminal fragment, and the potential increased intracellular degradation through the lysosome.

4.3.2 Cerebral atrophy-like

In 2013, our laboratory used homozygosity mapping and WES to identify the genetic basis of a novel pediatric neurodegenerative disorder, autosomal recessive cerebral atrophy. This phenotype is caused by a missense variant (c.995C>T, p.Thr332Met) in the *TMPRSS4* gene. Recently, our laboratory was referred a second case from India with a high resemblance to the initial autosomal recessive cerebral atrophy case. More specifically, the patient displayed microcephaly and cerebral atrophy. Using Sanger sequencing of the region containing the previously identified *TMPRSS4* variant, I sought to determine whether this phenotype shared the same genetic basis. When no variants were identified from Sanger sequencing, I pursued WES to identify novel disease-associated variants. From WES and bioinformatics processing of the proband's exome, I identified a variant in the kinetochore scaffold 1 gene (*CASC5*; c.6485A>G, p.Glu2162Gly) matching the phenotype of the proband.

4.3.2.1 *CASC5* and primary microcephaly-4

During cell division, genetic material must be accurately transmitted from parent to daughter cells. To ensure this, eukaryotes couple the movement of spindle microtubules to replicated chromosomes through a multi-protein attachment complex termed the kinetochore. The kinetochore is a disc-shaped protein that holds the sister chromatids

together during cell division. In total, kinetochores perform four functions: 1) create a sturdy interface with centromeric chromatin, 2) contribute a microtubule-binding interface, 3) generate the spindle assembly checkpoint, and 4) correct improper attachments²⁰⁸. Even the simplest kinetochores consist of over 19 proteins, with all proteins highly conserved among eukaryotes. One such protein that localizes to the kinetochore is the kinetochore scaffold 1 protein, encoded by the *CASC5* gene. This protein is pertinent for two events: the accurate attachment of centromeres to the microtubules, and spindle-assembly checkpoint signaling²⁰⁹.

Microcephaly, Primary Hereditary (MCPH) is a neurodevelopmental disorder featuring intellectual disability coupled with microcephaly from birth, where the microcephaly is caused by the reduction of the cerebral cortex. To date, several causative genes have been identified, including: *microcephalin*, CDK5 Regulatory Subunit Associated Protein 2 (*CDK5RAP2*), Abnormal Spindle Microtubule Assembly (*ASPM*), Centrosomal Protein 135 (*CEP135*), Centrosomal Protein 152 (*CEP152*) and *CASC5*²¹⁰. Although each of these genes play a unique role in the pathogenesis of MCPH, each is predicted to cause MCPH by disturbing mitotic spindle orientation, causing premature chromosomal condensation, or affecting microtubule dynamics²¹¹.

Specifically, variants in the *CASC5* gene cause microcephaly 4, primary, autosomal recessive – a form of MCPH featuring brain reduction and mental handicap²⁰⁹. Although the pathophysiological mechanism has not been resolved, it is thought that variants in *CASC5* affect the ability for *CASC5* to bind the mitotic spindle checkpoint protein, *BUBR1*, an essential component of the mitotic checkpoint²¹². Variants in *BUBR1* are associated with Mosaic Variegated Aneuploidy (MVA) Syndrome 1, a phenotype attributed to a mitotic chromosomal segregation defect leading to aneuploidy²¹³. Due to the required direct interaction between *CASC5* and *BUBR1*, it is suggested that defects in *CASC5* are pathogenic by way of a common mechanism involving chromosome missegregation. As our variant is in the same exon as known, classified microcephaly-4 variants, I suggest that our variant acts in a similar fashion, causing chromosome missegregation, and allowing premature entry into mitosis.

4.4 Whole-exome sequencing for CARE for RARE and FORGE Canada projects

For families ascertained through clinicians as part of the CARE4RARE and FORGE Canada projects, WES was performed on the DNA of the probands followed by bioinformatics processing and variant curation. From thousands of variants, candidate variants were narrowed down for eight of the 11 total cases. The successful identification of candidate variants for these rare disease cases demonstrates the utility of WES in identifying novel variants. Furthermore, as this was the effort of numerous clinicians and scientists, it emphasizes the importance of collaboration in the discovery of the genetic basis of rare diseases. Herein, I discuss two cases, microcephaly-ichthyosis and widow's peak syndrome, and briefly summarize the project.

4.4.1 Microcephaly-ichthyosis in the Old Order Amish

The Old Order Amish are a North American religious group from Anabaptist origins. In Canada, the majority of the Old Order settlements are within Ontario, mainly in the Norfolk and Bruce counties. Due to social isolation, the founder effect (loss of genetic variation in an area established by a small number of individuals), and the many genetic bottleneck events in Amish history, there is a high incidence of particular genetic disorders among the Old Order Amish²¹⁴. This is due to the autozygosity – chromosomal segments or alleles that are identical-by-descent due to consanguineous mating – often present in the Old Order Amish population.

In an Old Order Amish family with a history of microcephaly and ichthyosis, WES, bioinformatics processing and variant detection was performed. From WES on the DNA of the proband, a novel frameshift insertion was identified in the *VPS52* gene (c.57delG, p.Thr20fs). This variant results in a premature stop codon 30 amino acids from the start site, ultimately leading to deletion of 90% of the protein. Furthermore, this frameshift variant co-segregates with disease status in the pedigree, is predicted to be pathogenic from *in silico* prediction tools, is not present in any disease database, and has not been reported previously in the literature. This variant is also harboured within an autozygous

block in the family, 6p22.2p21.2. Due to the severity of the frameshift and the association of the gene defects with the phenotype, I propose that the *VPS52* variant is causing the recurring microcephaly-ichthyosis phenotype in this family.

The *VPS52* gene encodes the vacuolar protein sorting-52 protein, a component of the Golgi associated retrograde protein (GARP) complex. The GARP complex is a large complex involved in the retrograde transport of proteins from endosomes to the trans-Golgi network²¹⁵. Moreover, the GARP complex is a tetrameric complex consisting of four proteins, VPS51, VPS52, VPS53 and VPS54. These four proteins tether together to fulfil retrograde vesicle trafficking.

Vps52 acts to support the growth and pluripotent differentiation of the embryonic ectoderm through cell-cell interactions in extraembryonic tissues. Null mutations in Vps52 have been shown to cause early embryonic lethality, ectoderm death, and gastrulation defects in murine models²¹⁶. Furthermore, Vps52 knockouts result in the complete loss of GARP function²¹⁵. Due to the similarity between the consequences of null Vps52 and the microcephaly-ichthyosis phenotype in this family, I predict that the variants in *VPS52* are causative of the disease. More specifically, as the ectoderm is responsible for skin formation and there is embryonic lethality in the affected individuals, I suggest that the *VPS52* frameshift variant is causative of the microcephaly-ichthyosis lethal phenotype. Functional studies are required to confirm that the variant is indeed causing the microcephaly-ichthyosis phenotype in this family.

4.4.2 Widow's peak syndrome

Widow's peak syndrome is an ultra-rare X-linked dominant disease featuring a prominent widow's peak (pointed frontal hairline), facial abnormalities and skeletal anomalies. Described by Kapur *et al.* in 1989, widow's peak syndrome has only been reported once previously in the literature²¹⁷. Through CARE4RARE, we report a second case of widow's peak syndrome with matching features, namely, a pointed frontal hairline, epiphyseal dysplasia (disorder of bone and cartilage development affecting the long bones or epiphyses in the arms and legs), rotated ears, bilateral ptosis (drooping of both

eyelids) and knee joint issues. From WES, bioinformatics processing and variant detection, I report a novel splicing variant in *KDM6A* (c.2703-6_2703-5delTT), predicted to be pathogenic from *in silico* prediction tools. Using Sanger sequencing, this variant was found to co-segregate with disease status in the family.

Mutations in *KDM6A* are causative of Kabuki syndrome-2, a congenital syndrome featuring intellectual disability, facial anomalies (long palpebral fissures and eversion of the eyelids), prominent ears, scoliosis, and radiographic abnormalities^{218,219}. Due to the overlapping clinical features of widow's peak syndrome and Kabuki syndrome-2, I suggest that this family features a novel disease overlapping with these syndromes. Furthermore, I propose a potential link between these exclusive disorders. Due to the presence of the variant in a polyT region, it is necessary to quantitatively confirm the *KDM6A* variant. SNaPshot Multiplex sequencing was unsuccessful in confirming the variant, therefore, it is necessary to perform reverse transcription polymerase chain reaction (RT-PCR) in the future. To date, RNA extraction has been done and RT-PCR is ongoing.

4.4.3 CARE for RARE and FORGE summary

In summary, eight of 11 CARE4RARE and FORGE Canada Cases in the Blackburn Cardiovascular Genetics Laboratory were solved using WES and bioinformatics analysis, highlighting the utility of WES in gene discovery. The cases with candidate variants identified are microcephaly-ichthyosis, autosomal dominant isolated cryptophthalmos, widow's peak syndrome, amyoplasia, metaphyseal chondrodysplasia, craniosynostosis, heterotaxy disorder, and craniorachischisis. Each aforementioned case requires functional assessment in order to confirm the pathogenicity of the variants of interest. For the remaining cases, namely, Angelman-like syndrome, dystonia, ataxia and intellectual disability, and osteopetrosis, it is necessary to consider factors beyond the exome.

4.5 Gene enrichment analysis

From PANTHER computing and analysis on data from patients with lipodystrophy and extreme early-onset obesity, two patients with partial lipodystrophy were found to

harbour variants with significant overrepresentation in specific biological pathways. Namely, patient 14318 had overrepresentation of genes involved in *O*-glycan processing and tissue homeostasis, and patient 11618 had overrepresentation of genes involved in interferon-gamma-mediated signaling (PANTHER outputs in Appendix IV).

The *O*-glycan processing pathway is involved in adding a carbohydrate or carbohydrate derivative to an *O*-linked residue to form a core *O*-glycan structure. As these genes are fundamental, defects in this pathway often result in severe metabolic perturbations or embryonic lethality²²⁰. For patient 14318, there was a 5.11-fold enrichment of genes involved in this biological process. Therefore, variants in genes involved in *O*-glycan processing were considered for association with lipodystrophy. From this gene list associated with *O*-glycan processing, most of the genes involved were those belonging to the mucin gene family. As mucin genes are highly polymorphic with variable number tandem repeat (VNTR) length polymorphisms, these variants were excluded as they are likely unassociated with the lipodystrophy in this patient^{221,222}.

Genes involved in maintaining tissue homeostasis are those that regulate somatic stem cell activity²²³. In patient 14318 there was a 3.35-fold enrichment of genes involved in this biological process. For the list of overrepresented genes in patient 14318 involved in tissue homeostasis, each gene was examined for association with lipodystrophy. As there was no literature supporting an association between each gene and/or variant with lipodystrophy, these genes were not prioritized as candidates. Although there is no available literature supporting these genes now, future work should continue to investigate the variants within these genes for their potential contribution to lipodystrophy.

For patient 11618, genes involved in interferon-gamma mediated signaling were overrepresented. More specifically, there was 6.54-fold enrichment of the genes in this pathway. As most of these genes were human leukocyte antigen (HLA) genes with a high degree of polymorphism, they were deemed unassociated with the lipodystrophy in this patient.

In sum, although PANTHER was helpful in identifying overrepresented biological processes in these patients, no gene lists generated from PANTHER are predicted to contribute to the partial lipodystrophy phenotypes of these patients.

4.6 Clinical utility of LipidSeq™ and whole-exome sequencing

From the utilization of the LipidSeq™ resequencing gene panel, 40% of a 153-person lipodystrophy cohort was, at least partially, genetically explained. From WES, 65% of lipodystrophy patients, 40% of extreme early-onset obesity patients, two rare disease cases (brachydactyly type A1 and cerebral atrophy-like), and 73% of CARE4RARE/FORGE Canada cases had candidate variants identified. Together, these statistics indicate the clinical utility of NGS, especially in the context of variant discovery in rare disease research.

4.6.1 Comparison of LipidSeq™ and whole-exome sequencing to whole-genome sequencing

As some cases remained genetically unexplained following the implementation of LipidSeq™ and WES, it begs the question whether WGS would have been preferred. Here, I argue that although WGS is capable of sequencing almost all three-billion base pairs of the human genome, compared to 2% with WES and even less with targeted sequencing, LipidSeq™ and WES are still superior to WGS for rare disease research. With WGS, there are several implications affecting clinicians, researchers and patients. These implications refer to data yield, interpretation and ethics.

Although the diagnostic yield is greater with WGS, there are several issues concerning amassing such a wealth of data. In terms of computational storage, it is impractical for most laboratories to store such data. Furthermore, the processing of the data require high computational and statistical power, often unavailable at a cost-efficient rate. As large cohorts would need to be processed and screened to result in a clinical outcome, it is unfeasible for most laboratories to advance towards the sole use of WGS. In addition, the

body of literature regarding intergenic regions and their role in disease is sparse. Without an abundance of information on these regions or a universal standard for their interpretation, it is difficult for providers to assess risk without creating false reassurance or needless worry. Last, there are several ethical implications regarding the availability of such data. Without distinct laws in place on genetic discrimination, issues may arise when patients with WES data try to obtain insurance. Additionally, with an almost guaranteed likelihood of incidental findings, there is apprehension to progress towards WGS. As WGS has shown to only offer limited clinical utility compared to WES (only 3% increased detection in a recent study), WES is recommended over WGS²²⁴.

4.7 Limitations of the project

Although I demonstrated the clinical utility of NGS in solving or clarifying the molecular basis of rare disorders, there are limitations and study caveats to consider. These caveats are divided to those specific to this project, and those specific to the use of targeted panels and WES technology.

4.7.1 Study considerations

1) Acquired lipodystrophy cases

In this study, I relied on an assumption that all the lipodystrophy cases were genetic. For many patients in our lipodystrophy cohort, it is likely that some of these lipodystrophy cases are acquired with no inherited component. The presence of acquired lipodystrophy cases was demonstrated when it was determined that four patients subjected to WES had acquired lipodystrophy. Acquired lipodystrophy refers to lipodystrophy that sporadically appears during a random point in life. As up to 70% of human immunodeficiency virus (HIV)-infected individuals on highly active antiretroviral therapy (HAART) therapy report to have HIV-associated lipodystrophy, it is likely that some forms of lipodystrophy are drug-related or environmental²²⁵. The study of these cases may have caused an underestimation of the clinical utility of LipidSeq™ and WES. In the future, it is necessary to consider these cases separately in order to refrain from any bias in the results.

2) Limited clinical information

For some of the cases listed in this MSc Thesis, clinical information was limited. Moreover, some of the cases were referred from external physicians to the laboratory leading to inconsistencies in clinical diagnosis. As additional clinical information would have streamlined the variant prioritization process (i.e. if information was available on mode of inheritance, secondary complications, etc.) and potentially changed the outcome of cases, this was an inherent limitation of the study.

3) Lack of trios and families

For many of the pediatric cases and those affecting multiple members in a family, one frequent challenge was the inaccessibility of patient samples. Regardless of whether it was due to lack of proximity of the patients or lack of willingness to participate in genetic research, the inaccessibility of patient samples made it difficult to confirm variants, co-segregation and pathogenicity. To account for the unavailable DNA samples, I used numerous databases; however, it would have been beneficial to ascertain these samples to perform trio and family analysis.

Furthermore, as rare diseases are so uncommon it can often be more challenging to study these diseases due to small sample size. In this study, some cases were limited to a single affected individual. This n-of-1 problem also contributed to the difficulty in confirming variants and pathogenicity.

4.7.2 General methodological limitations

In general, there are numerous limitations with targeted sequencing and WES. With these methods, several genetic factors are not considered. These include large rearrangements, mitochondrial gene mutations, epigenetic factors, uniparental disomy, environmental effects, and mutations in repetitive and GC-rich regions. As these were not investigated, I may be lacking all the information necessary to genetically explain each patient or rare disease case.

4.8 Future directions

From this MSc Thesis, there are several considerations in terms of future directions. These include those specific to this project and those necessary for the rare disease research field. These are separated into their respective categories below.

Future work for the project

From here, future work for this project will require focus on confirming the WES results and ensuring the disease-causing criteria for candidate variants are fulfilled. For cases with a candidate variant or multiple variants identified, validation and co-segregation analyses should be performed if not done so already. For all candidate CNVs identified, secondary validation using an independent method (i.e. Sanger sequencing or CytoScan™ HD) should be done. Ultimately, functional studies are required to confirm the pathogenicity of each CNV in causing the disease of interest. These studies are recommended for all novel variants, particularly the CARE4RARE and FORGE Canada cases, and are listed in the following sub-section. For those cases in which a candidate variant (or multiple candidate variants) was not identified, the exome data should be revisited and re-evaluated using less stringent criteria (i.e. higher minor allele frequencies and lower read depth stringency) and other factors, as discussed in the following section, should be considered.

Functional assessments to be considered for this project

For those cases in which a candidate variant was identified using NGS – namely the lipodystrophy, extreme early-onset obesity, brachydactyly, cerebral atrophy-like and solved CARE4RARE/FORGE Canada cases – it is suggested that functional assessments be undertaken to confirm the validity and pathogenicity of the candidate variants. These experiments may include, but are not limited to the following:

- 1) Western blotting

A Western blot is an analytic method used in molecular biology to identify specific protein molecules from a mixture of proteins extracted from cells. By separating based on

size, transferring to a solid support, and using a primary and secondary antibody to visualize a target protein, Western blotting can elucidate the molecular weight of a protein of interest²²⁶.

For this project, Western blotting can be used as a first tier assessment of whether a protein product is present or absent. More specifically, Western blotting can allude to whether the defective gene results in abnormal protein expression. If Western blotting demonstrates a change in expression, additional studies may be pursued to determine if the protein product is dysfunctional.

2) mRNA quantification (RT-qPCR)

Reverse transcription quantitative polymerase chain reaction (RT-qPCR) is a technology used to analyze RNA expression. By converting labile RNA to more stable complementary DNA (cDNA) using a reverse transcriptase and using real-time fluorescence to quantify the quantity of DNA present at each cycle during PCR, RT-qPCR can be used for measuring gene expression and copy-number variation²²⁷. In this study, RT-qPCR can be used to determine differences in mRNA expression between patients with genetic variants and control samples. Additionally, this technique can allow for independent confirmation of NGS results²²⁸.

3) Co-immunoprecipitation

In a cell, two or more proteins can interact and lead to biological effects. Co-immunoprecipitation (co-IP) is a common method used to analyze protein-protein interactions. Briefly, in co-IP the protein of interest is isolated using a specific antibody, and interacting partners are identified by subsequent Western blotting²²⁹. In this project, co-IP may be a useful technique in determining AR protein interactions. Furthermore, with the lipodystrophy patients harbouring variants in multiple genes, co-IP may be useful in elucidating protein interactions.

4) Molecular studies

In molecular biology, molecular studies and model systems are common methods used to study the biological consequences of the genetic variation underlying human diseases.

These studies may include *in vitro* assays, *ex vivo* studies, or *in vivo* models.

In vitro assays are those used to determine whether gene products are defective and/or disrupt downstream pathways. These assays typically introduce the variant into the DNA of an artificial system to assess the biological effects on protein function and connecting pathways. *Ex vivo* studies are those that are generally performed on extracted patient tissues. Experiments to determine gene expression or protein interactions are pursued with patient-derived cells to investigate the pathogenicity of a candidate variant. *In vivo* experiments are those done within a model system. Common model organisms include *Drosophila melanogaster* (fruit flies), *Caenorhabditis elegans* (roundworms), *Danio rerio* (zebrafish) and *mus musculus* (mice). Model organisms are an efficient and effective strategy, as you can manipulate the genome of the animal and investigate the resultant phenotype.

In this project, molecular studies would be useful to confirm that the novel candidate variants discovered are indeed pathogenic.

Further considerations for the project

For those cases in which a candidate variant was not identified – namely Angelman-like syndrome; osteopetrosis; dystonia, ataxia and intellectual disability; remaining lipodystrophy cases; and the remaining extreme early-onset obesity cases – it is necessary to consider other genetic factors or determinants of disease. Here I list some avenues that are being pursued, and others that are options for the future:

1) Mosaicism

Mosaicism describes a situation where different cells in an individual carry a distinct number of chromosomes or chromosomal arrangements. This is demonstrated as the

presence of two genotypes in a single individual from a single egg²³⁰. Mosaicism is often the cause of sporadic disease in affected individuals with unaffected parents²³¹.

In one study, I took mosaicism into account when analysing the LipidSeq™ results of a child with prenatal hypoglycemia. This child could not be explained solely by a heterozygous mutation in *KCNJ11*. After consideration of mosaicism, it was determined through surgery and DNA testing that the child harboured somatic *ABCC8* mutations in her pancreas. Together, the *KCNJ11* and *ABCC8* variants explained her phenotype.

2) Splicing

Splice variants play a central role in the diversification of genomes and in cellular function. Furthermore, splice sites are associated with disease states and may drive disease phenotypes²³². As exonic regions were prioritized more in this study, it is necessary to retroactively consider splicing variants, especially those more distal to the exons.

Recently, my laboratory began a collaboration with Dr. Brendan Frey of the University of Toronto in order to classify splice variants and use algorithms to predict the pathogenicity of each splice variant. Continued work on understanding these splice region variants and their contribution to disease will be pursued in the future.

3) Chromosomal rearrangements

Chromosomal rearrangements involve large structural changes including deletions, duplications, inversions and translocations. As the DNA is sheared prior to WES and the WES kit only captures ~150-bp reads, it is necessary to consider cytogenetic techniques to determine whether chromosomal rearrangements may be involved.

4) Regulatory regions

Regulatory regions are one of the mechanisms through which protein levels are controlled. Regulatory regions include promoters, enhancers, silencers and insulators²³³. As much of the regulatory content is harboured within intergenic DNA not sequenced through WES, it is imperative that the sequence of these regions be considered.

5) Digenic disorders

Digenic disorders refer to those diseases where mutations on two genes interact to cause the phenotype²³⁴. In this MSc Thesis, I consistently use knowledge of candidate genes and protein-protein interactions in order to consider the possibility of digenic inheritance. Future work will continue to explore the prospect of digenic inheritance by keeping up with current literature and continually checking for newly elucidated interactions between candidate genes.

6) Polygenic disorders

Polygenic diseases are those caused by the contribution of numerous distinct or interacting genes⁵³. Generally, in these disorders each variant contributes a small-to-moderate effect on the phenotype, cumulatively causing the disease. To determine whether the genetic basis of any of the unsolved cases are polygenic, it would be necessary to perform a GWAS to identify genetic loci associated with the disease. From there, it would be possible to develop a polygenic risk score – a metric that summarizes all genome-wide genotype data to predict the genetic risk of a disorder – to determine risk for disease²³⁵.

7) Matchmaker Exchange

Due to the scarcity of rare diseases and the small datasets in each laboratory for each rare disease case, it is necessary for data sharing in the research community. The Matchmaker Exchange (MME) platform was designed to provide a robust approach to gene discovery through the creation of a network connecting databases of rare phenotypes and genotypes using a programming interface²³⁶. Put simply, MME is a platform allowing clinicians, physicians and patients to connect with the hope for serendipitous occasions where two investigators learn that they share similar rare disease cases (i.e. cases with similar features or symptoms). So far, the Angelman-like syndrome case has been input into MME, with other CARE4RARE and FORGE Canada cases to follow if necessary.

8) Epigenetics

The term “epigenetics” refers to a change in gene expression without a change in DNA sequence. These modifications in gene expression may be through changes in methylation or acetylation. As epigenetics has been shown to play a large role in phenotypic traits and disease, it is necessary to consider epigenetic mechanisms of disease. In the future, it may be necessary to explore epigenetic changes. For example, future work could include the use of bisulfite genomic sequencing – a method to detect changes in DNA methylation through the conversion of genomic DNA with sodium bisulfite – to evaluate the epigenetic landscape of each unsolved disease.

Future directions for the rare disease research field

For the rare disease research field, there is an ample amount of work that needs to be done. Three main points that should be focused on are: 1) the standardization of variant interpretation and analysis; 2) the conversion of all data to open-source; and 3) the increase in awareness of rare diseases. With candidate variant interpretation, it is necessary for all clinicians and researchers to conform to a standard set of guidelines to ease communication and promote collaboration. In this study, I considered the ACMG standards and guidelines for variant interpretation¹²⁶. I suggest the adoption of these guidelines across the field to 1) promote a universal standard of variant interpretation and 2) reduce any possible contradictions regarding variant interpretation. I also encourage that all data be open-source and accessible. As rare diseases are extremely difficult to study due to small sample sizes and low statistical power, it is necessary that all investigators make their data publically available to allow for further and faster advancements within the field. Last, I advocate for increased awareness of rare diseases. As individuals with rare diseases face a diagnostic odyssey, it is necessary to increase the awareness of these diseases to pave the way towards the end of the diagnostic odyssey for these patients. By raising awareness with policy makers, researchers, physicians and the public, it may translate to greater research funding and research interest, more robust diagnoses in the clinic, and ultimately, the elucidation of molecular mechanisms and potentially, novel treatments or cures in the future.

4.9 Conclusions

NGS technologies, including targeted resequencing panels and WES, are highly robust, efficient and cost-effective approaches in the diagnosis of rare disorders. In this MSc Thesis, I have effectively achieved my goal: to demonstrate the utility of NGS in rare disease research, the laboratory and the clinic, and identify a molecular basis for rare disease patients and families of interest. More specifically, I have fulfilled my aims to apply NGS technologies to detect genetic variants and CNVs, identify rare variants underlying disease phenotypes, determine the functional relevance of these variants using *in silico* tools, and perform variant validation studies to determine whether the variants co-segregate with disease status in families.

More specifically, by processing 153 patients clinically diagnosed with lipodystrophy through the LipidSeq™, a custom gene panel used for the clinical diagnosis of monogenic dyslipidemias, 81 variants in 14 lipodystrophy-associated genes (*LMNA*, *PPARG*, *PLIN1*, *CIDEC*, *LIPE*, *LPINI*, *AGPAT2*, *BSCL2*, *CAVI*, *AKT2*, *PTRF*, *DYRK1B*, *POLD1* and *WRN*) were identified within 58 patients. Furthermore, an additional seven CNV events were detected in seven patients with lipodystrophy, namely duplications in *PLIN1*, *MFN2*, *APOA5/APOA4*, *CREB3L3* and deletions in *AGPAT2*, *PLIN1* and *CIDEC*. This is the first report of systematically screening for CNV events, and these represent the first CNVs identified in patients with lipodystrophy.

Twenty patients with lipodystrophy and five patients with extreme early-onset obesity were subjected to WES and bioinformatics analysis. From WES, candidate variants were discovered in 13 patients with lipodystrophy and two with extreme early-onset obesity in the following genes: *ISCU*, *HSPG2*, *LDLR*, *PDRM16*, *ZMPSTE24*, *FLNC*, *GALC*, *AR*, *ABCA4*, *WRN*, *ABCG8*, *FHL1*, *ANLN*, *LPL*, *FN1*, *PAPSS2*, *NEUROD1*, *BSCL2*, *APOB*, *PCSK9*, *PYGM*, *BLK*, *MC4R*, *BBS10*, *SETD8*, and *KMT2C*. Furthermore, two of two rare disease cases (brachydactyly type A1, *IHH*; cerebral atrophy-like, *CASC5*) and eight of 11 CARE4RARE/FORGE Canada had candidate variants identified from WES data and bioinformatics processing. The successful identification of candidate genes and the clarification of the genetic determinants underlying many of these rare disease cases

demonstrate the power of WES in rare disease research. Additionally, as many of these genetic variants have never been reported, these results highlight the effectiveness of NGS in detecting novel genetic variation.

In sum, in my MSc Thesis project I have identified novel disease-associated genes, characterized novel variants, and demonstrated the utility of targeted sequencing and WES in clarifying the genetic basis of rare genetic diseases. Furthermore, I have provided instructions on how to perform NGS and bioinformatics in a robust, efficient manner, and effectively ended the diagnostic odyssey for some patients through the identification of candidate variants. Understanding these genetic variants and their association with each will lead to many positive implications, most importantly, benefits to families, changes in treatment options for genetically explained patients, and added knowledge for the field.

4.10 Project significance

Broadly, many significant outcomes may result from this MSc Thesis. These outcomes take three forms: benefits to the family, therapeutic benefits, and scientific benefits. On the family level, identifying the molecular basis of these rare diseases can allow for several positive outcomes. First, it can allow the physician to assign a definitive name of the condition. Although this may not lead to altered management, this may allow the individual or the family to find support in community groups or the scientific literature. Second, it marks the end of the diagnostic odyssey. As a tumultuous journey for individuals, the end of the diagnostic odyssey may prevent unnecessary tests, prescriptions or consultations. Third, it may open the door for preventative medicine. For example, one clinical story describes a lipodystrophy patient whose *LMNA* mutation discovery encouraged physicians to approach her case differently. This enabled the detection of a severe cardiac rhythm disorder, and allowed subsequent life-saving treatment²³⁷. Last, it may provide recurrence risk for future pregnancies. The understanding of recurrence may allow for advanced family planning and preparation on the psychosocial level.

The second significant outcome of this research is the effect that these results have on treatment options for an affected individual. Through a genetic diagnosis, a physician may be able to offer a change in surveillance, or modification in medication. These alterations may provide an improved prognosis for an individual.

The third outcome of genetic variant discovery are the benefits that these results have on the rare disease research field. By identifying a genetic basis of a rare disorder, the results may provide insight on both normal and disease pathways and pave the way towards the development of novel therapeutic interventions.

In addition to the positive implications on rare disease research, the application of NGS technologies in rare disease research also has significant implications on the field. More specifically, by demonstrating the clinical utility of LipidSeq™ and WES for rare disease cases, we can encourage other researchers or clinicians to use these methods when investigating the genetic basis of rare disease phenotypes.

In sum, identifying the molecular bases of rare disorders using NGS technologies demonstrated the power of NGS as a clinical diagnostic tool, and contributed the first step towards the long-term goal – to develop novel treatments and end the diagnostic odyssey for these rare disease patients.

References

1. MacKenzie, A. & Boycott, K.M. The future is now for rare genetic diseases. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne* **184**, 1603 (2012).
2. Basel, D. & McCarrier, J. Ending a Diagnostic Odyssey: Family Education, Counseling, and Response to Eventual Diagnosis. *Pediatric Clinics of North America* **64**, 265-72 (2017).
3. Shashi, V. *et al.* The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* **16**, 176-82 (2014).
4. Zhang, M., Zhu, C., Jacomy, A., Lu, L.J. & Jegga, A.G. The orphan disease networks. *American Journal of Human Genetics* **88**, 755-766 (2011).
5. Sawyer, S.L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clinical Genetics* **89**, 275-84 (2016).
6. Lander, E.S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567-70 (1987).
7. Bamshad, M.J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* **12**, 745-55 (2011).
8. Garg, A. Lipodystrophies. *The American Journal of Medicine* **108**, 143-52 (2000).
9. Garg, A. Acquired and inherited lipodystrophies. *The New England Journal of Medicine* **350**, 1220-34 (2004).
10. Handelsman, Y. *et al.* The clinical approach to the detection of lipodystrophy - an AACE consensus statement. *Endocrine Practice* **19**, 107-16 (2013).
11. Vantyghem, M.C. *et al.* Fertility and obstetrical complications in women with LMNA-related familial partial lipodystrophy. *Journal of Clinical Endocrinology and Metabolism* **93**, 2223-9 (2008).
12. Cortes, V.A. & Fernandez-Galilea, M. Lipodystrophies: adipose tissue disorders with severe metabolic implications. *Journal of Physiology and Biochemistry* **71**, 471-8 (2015).
13. Berardinelli, W. An undiagnosed endocrinometabolic syndrome: report of 2 cases. *The Journal of Clinical Endocrinology and Metabolism* **14**, 193-204 (1954).

14. Seip, M. Lipodystrophy and gigantism with associated endocrine manifestations. A new diencephalic syndrome? *Acta Paediatrica* **48**, 555-74 (1959).
15. Garg, A., Peshock, R.M. & Fleckenstein, J.L. Adipose tissue distribution pattern in patients with familial partial lipodystrophy (Dunnigan variety). *The Journal of Clinical Endocrinology and Metabolism* **84**, 170-4 (1999).
16. Garg, A., Speckman, R.A. & Bowcock, A.M. Multisystem dystrophy syndrome due to novel missense mutations in the amino-terminal head and alpha-helical rod domains of the lamin A/C gene. *The American Journal of Medicine* **112**, 549-55 (2002).
17. Subramanyam, L., Simha, V. & Garg, A. Overlapping syndrome with familial partial lipodystrophy, Dunnigan variety and cardiomyopathy due to amino-terminal heterozygous missense lamin A/C mutations. *Clinical Genetics* **78**, 66-73 (2010).
18. Fiorenza, C.G., Chou, S.H. & Mantzoros, C.S. Lipodystrophy: pathophysiology and advances in treatment. *Nature Reviews Endocrinology* **7**, 137-50 (2011).
19. Rao, D.P., Kropac, E., Do, M.T., Roberts, K.C. & Jayaraman, G.C. Childhood overweight and obesity trends in Canada. *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice* **36**, 194-8 (2016).
20. Choquet, H. & Meyre, D. Genomic insights into early-onset obesity. *Genome Medicine* **2**, 36 (2010).
21. Bouchard, C. Childhood obesity: are genetic differences involved? *The American Journal of Clinical Nutrition* **89**, 1494S-1501S (2009).
22. Temtamy, S.A. & McKusick, V.A. The genetics of hand malformations. *Birth Defects Original Article Series* **14**, i-xviii, 1-619 (1978).
23. Temtamy, S.A. & Aglan, M.S. Brachydactyly. *Orphanet Journal of Rare Diseases* **3**, 15 (2008).
24. Bertram, L. & Tanzi, R.E. The genetic epidemiology of neurodegenerative disease. *The Journal of Clinical Investigation* **115**, 1449-57 (2005).
25. Matilla-Duenas, A. *et al.* Rare Neurodegenerative Diseases: Clinical and Genetic Update. *Advances in Experimental Medicine and Biology* **1031**, 443-496 (2017).
26. Lahiry, P. *et al.* A mutation in the serine protease *TMPRSS4* in a novel pediatric neurodegenerative disorder. *Orphanet Journal of Rare Diseases* **8**, 126 (2013).
27. CHEO Research Institute. CARE for RARE. (2018). at <<http://care4rare.ca/>>

28. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
29. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
30. Feero, W.G., Guttmacher, A.E. & Collins, F.S. Genomic medicine - an updated primer. *The New England Journal of Medicine* **362**, 2001-11 (2010).
31. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nature Reviews Genetics* **7**, 85-97 (2006).
32. Sherry, S.T., Ward, M. & Sirotkin, K. dbSNP - database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Research* **9**, 677-9 (1999).
33. Bessenyei, B., Marka, M., Urban, L., Zeher, M. & Semsei, I. Single nucleotide polymorphisms: aging and diseases. *Biogerontology* **5**, 291-303 (2004).
34. Baralle, D. & Buratti, E. RNA splicing in human disease and in the clinic. *Clinical Science* **131**, 355-68 (2017).
35. Parkinson, D.B. & Thakker, R.V. A donor splice site mutation in the parathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *Nature Genetics* **1**, 149-52 (1992).
36. Mullaney, J.M., Mills, R.E., Pittard, W.S. & Devine, S.E. Small insertions and deletions (INDELs) in human genomes. *Human Molecular Genetics* **19**, R131-6 (2010).
37. Frisch, A. *et al.* Origin and spread of the 1278insTATC mutation causing Tay-Sachs disease in Ashkenazi Jews: genetic drift as a robust and parsimonious hypothesis. *Human Genetics* **114**, 366-76 (2004).
38. de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A. & Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics* **7**, e1002384 (2011).
39. Liang, K.C., Tseng, J.T., Tsai, S.J. & Sun, H.S. Characterization and distribution of repetitive elements in association with genes in the human genome. *Computational Biology and Chemistry* **57**, 29-38 (2015).
40. Freeman, J.L. *et al.* Copy number variation: new insights in genome diversity. *Genome Research* **16**, 949-61 (2006).
41. Stankiewicz, P. & Lupski, J.R. Structural variation in the human genome and its role in disease. *Annual Review of Medicine* **61**, 437-55 (2010).

42. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444-54 (2006).
43. Hastings, P.J., Lupski, J.R., Rosenberg, S.M. & Ira, G. Mechanisms of change in gene copy number. *Nature Reviews Genetics* **10**, 551-64 (2009).
44. Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., *et al.* Somatic versus germinal mutation. *An Introduction to Genetic Analysis* (W. H. Freeman, New York, 2000).
45. Martincorena, I. & Campbell, P.J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483-9 (2015).
46. Acuna-Hidalgo, R., Veltman, J.A. & Hoischen, A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biology* **17**, 241 (2016).
47. Tenesa, A. & Haley, C.S. The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics* **14**, 139-49 (2013).
48. Visscher, P.M., Hill, W.G. & Wray, N.R. Heritability in the genomics era-- concepts and misconceptions. *Nature Reviews Genetics* **9**, 255-66 (2008).
49. Antonarakis, S.E. & Beckmann, J.S. Mendelian disorders deserve more attention. *Nature reviews. Genetics* **7**, 277-82 (2006).
50. Gazzo, A. *et al.* Understanding mutational effects in digenic diseases. *Nucleic Acids Research* **45**, e140 (2017).
51. Gazzo, A.M. *et al.* DIDA: a curated and annotated digenic diseases database. *Nucleic Acids Research* **44**, D900-7 (2016).
52. Kajiwara, K., Berson, E.L. & Dryja, T.P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and *ROM1* loci. *Science* **264**, 1604-8 (1994).
53. Lvovs, D., Favorova, O.O. & Favorov, A.V. A Polygenic Approach to the Study of Polygenic Diseases. *Acta Naturae* **4**, 59-71 (2012).
54. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. Lessons from studying monogenic disease for common disease. *Human Molecular Genetics* **15 Spec No 1**, R67-74 (2006).
55. McCoy, M.L., Mueller, C.R. & Roskelley, C.D. The role of the breast cancer susceptibility gene 1 (*BRCA1*) in sporadic epithelial ovarian cancer. *Reproductive Biology and Endocrinology: RB&E* **1**, 72 (2003).
56. Ambani, L.M., Gelehrter, T.D. & Sheahan, D.G. Variable expression of Marfan syndrome in monozygotic twins. *Clinical Genetics* **8**, 358-63 (1975).

57. Garg, A. Clinical review: Lipodystrophies: genetic and acquired body fat disorders. *The Journal of Clinical Endocrinology and Metabolism* **96**, 3313-25 (2011).
58. Krahmer, N., Farese, R.V., Jr. & Walther, T.C. Balancing the fat: lipid droplets and human disease. *EMBO Molecular Medicine* **5**, 973-83 (2013).
59. Garg, A. & Agarwal, A.K. Lipodystrophies: disorders of adipose tissue biology. *Biochimica et Biophysica Acta* **1791**, 507-13 (2009).
60. Agarwal, A.K. & Garg, A. Genetic basis of lipodystrophies and management of metabolic complications. *Annual Review of Medicine* **57**, 297-311 (2006).
61. Souren, N.Y. *et al.* Common SNPs in *LEP* and *LEPR* associated with birth weight and type 2 diabetes-related metabolic risk factors in twins. *International Journal of Obesity* **32**, 1233-9 (2008).
62. Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401 (1998).
63. Challis, B.G. & Millington, G.W.M. Proopiomelanocortin Deficiency. *GeneReviews(R)* (eds. Adam, M.P. *et al.*) (Seattle (WA), 1993).
64. Benzinou, M. *et al.* Common nonsynonymous variants in *PCSK1* confer risk of obesity. *Nature Genetics* **40**, 943-5 (2008).
65. Mergen, M., Mergen, H., Ozata, M., Oner, R. & Oner, C. A novel melanocortin 4 receptor (*MC4R*) gene mutation associated with morbid obesity. *The Journal of Clinical Endocrinology and Metabolism* **86**, 3448 (2001).
66. Rosas-Vargas, H., Martinez-Ezquerro, J.D. & Bienvenu, T. Brain-derived neurotrophic factor, food intake regulation, and obesity. *Archives of Medical Research* **42**, 482-94 (2011).
67. Gray, J. *et al.* Functional characterization of human *NTRK2* mutations identified in patients with severe early-onset obesity. *International Journal of Obesity* **31**, 359-64 (2007).
68. Gunay-Aygun, M., Schwartz, S., Heeger, S., O'Riordan, M.A. & Cassidy, S.B. The changing purpose of Prader-Willi syndrome clinical diagnostic criteria and proposed revised criteria. *Pediatrics* **108**, E92 (2001).
69. Muller, J. *et al.* Identification of 28 novel mutations in the Bardet-Biedl syndrome genes: the burden of private mutations in an extensively heterogeneous disease. *Human Genetics* **127**, 583-93 (2010).
70. Seifert, W. *et al.* Mutational spectrum of *COH1* and clinical heterogeneity in Cohen syndrome. *Journal of Medical Genetics* **43**, e22 (2006).

71. Marshall, J.D. *et al.* New Alstrom syndrome phenotypes based on the evaluation of 182 cases. *Archives of Internal Medicine* **165**, 675-83 (2005).
72. Bell, J. *Treasury of Human Inheritance Vol. 5* (Cambridge University Press, London, 1951).
73. Haws, D.V. & McKusick, V.A. Farabee's Brachydactylous Kindred Revisited. *Bulletin of the Johns Hopkins Hospital* **113**, 20-30 (1963).
74. Drinkwater, H. An account of a brachydactylous family. *Proceedings of the Royal Society of Edinburgh* **28**, 35-57 (1908).
75. Drinkwater, H. Account of a family showing minor-brachydactyly. *Journal of Genetics* **2**, 21-40 (1912).
76. Kirkpatrick, T.J. *et al.* Identification of a mutation in the Indian Hedgehog (*IHH*) gene causing brachydactyly type A1 and evidence for a third locus. *Journal of Medical Genetics* **40**, 42-4 (2003).
77. Gao, B. *et al.* Mutations in *IHH*, encoding Indian hedgehog, cause brachydactyly type A-1. *Nature Genetics* **28**, 386-8 (2001).
78. McCready, M.E. *et al.* A novel mutation in the *IHH* gene causes brachydactyly type A1: a 95-year-old mystery resolved. *Human Genetics* **111**, 368-75 (2002).
79. Byrnes, A.M. *et al.* Brachydactyly A-1 mutations restricted to the central region of the N-terminal active fragment of Indian Hedgehog. *European Journal of Human Genetics : European Journal of Human Genetics* **17**, 1112-20 (2009).
80. Liu, M. *et al.* A novel heterozygous mutation in the Indian hedgehog gene (*IHH*) is associated with brachydactyly type A1 in a Chinese family. *Journal of Human Genetics* **51**, 727-31 (2006).
81. Stattin, E.L., Linden, B., Lonnerholm, T., Schuster, J. & Dahl, N. Brachydactyly type A1 associated with unusual radiological findings and a novel Arg158Cys mutation in the Indian hedgehog (*IHH*) gene. *European Journal of Medical Genetics* **52**, 297-302 (2009).
82. Lodder, E.M., Hoogeboom, A.J., Coert, J.H. & de Graaff, E. Deletion of 1 amino acid in Indian hedgehog leads to brachydactylyA1. *American Journal of Medical Genetics Part A* **146A**, 2152-4 (2008).
83. Ho, R., McIntyre, A.D., Kennedy, B.A., & Hegele, R.A. Whole-exome sequencing identifies a novel *IHH* insertion in an Ontario family with brachydactyly type A1. *SAGE Open Medicine Case Reports* **Submitted** (2018).

84. Armour, C.M., Bulman, D.E. & Hunter, A.G. Clinical and radiological assessment of a family with mild brachydactyly type A1: the usefulness of metacarpophalangeal profiles. *Journal of Medical Genetics* **37**, 292-6 (2000).
85. Armour, C.M., McCready, M.E., Baig, A., Hunter, A.G. & Bulman, D.E. A novel locus for brachydactyly type A1 on chromosome 5p13.3-p13.2. *Journal of Medical Genetics* **39**, 186-8 (2002).
86. Byrnes, A.M. *et al.* Mutations in *GDF5* presenting as semidominant brachydactyly A1. *Human Mutation* **31**, 1155-62 (2010).
87. Chang, S.C. *et al.* Cartilage-derived morphogenetic proteins. New members of the transforming growth factor-beta superfamily predominantly expressed in long bones during human embryonic development. *The Journal of Biological Chemistry* **269**, 28227-34 (1994).
88. Racacho, L. *et al.* Two novel disease-causing variants in *BMPRI1B* are associated with brachydactyly type A1. *European Journal of Human Genetics* **23**, 1640-5 (2015).
89. Ploger, F. *et al.* Brachydactyly type A2 associated with a defect in proGDF5 processing. *Human Molecular Genetics* **17**, 1222-33 (2008).
90. Lehmann, K. *et al.* Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12277-82 (2003).
91. Badura-Stronka, M. *et al.* Novel mutation in the *BMPRI1B* gene (R486L) in a Polish family and further delineation of the phenotypic features of *BMPRI1B*-related brachydactyly. *Birth Defects Research. Part A, Clinical and Molecular Teratology* **103**, 567-72 (2015).
92. Su, P. *et al.* A 4.6 kb genomic duplication on 20p12.2-12.3 is associated with brachydactyly type A2 in a Chinese family. *Journal of Medical Genetics* **48**, 312-6 (2011).
93. Liu, X. *et al.* Identification of duplication downstream of *BMP2* in a Chinese family with brachydactyly type A2 (BDA2). *PloS One* **9**, e94201 (2014).
94. Dathe, K. *et al.* Duplications involving a conserved regulatory element downstream of *BMP2* are associated with brachydactyly type A2. *American Journal of Human Genetics* **84**, 483-92 (2009).
95. Hertzog, K.P. Shortened fifth medial phalanges. *American Journal of Physical Anthropology* **27**, 113-18 (1967).

96. Williams, K.D. *et al.* Heritability of brachydactyly type A3 in children, adolescents, and young adults from an endogamous population in eastern Nepal. *Human Biology* **79**, 609-22 (2007).
97. Bass, H.N. Familial absence of middle phalanges with nail dysplasia: a new syndrome. *Pediatrics* **42**, 318-23 (1968).
98. Cuevas-Sosa, A. & Garcia-Segur, F. Brachydactyly with absence of middle phalanges and hypoplastic nails: a new hereditary syndrome. *The Journal of Bone and Joint Surgery, British Volume* **53**, 101-5 (1971).
99. Ohzeki, T. *et al.* Brachydactyly type A-4 (Temtamy type) with short stature in a Japanese girl and her mother. *American Journal of Medical Genetics* **46**, 260-2 (1993).
100. Bick, D. & Dimmock, D. Whole exome and whole genome sequencing. *Current Opinion in Pediatrics* **23**, 594-600 (2011).
101. Metzker, M.L. Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46 (2010).
102. Johansen, C.T. *et al.* LipidSeq: a next-generation clinical resequencing panel for monogenic dyslipidemias. *Journal of Lipid Research* **55**, 765-72 (2014).
103. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19096-101 (2009).
104. Ng, S.B. *et al.* Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics* **42**, 30-5 (2010).
105. Hoischen, A. *et al.* *De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nature Genetics* **42**, 483-5 (2010).
106. Farhan, S.M. & Hegele, R.A. Exome sequencing: new insights into lipoprotein disorders. *Current Cardiology Reports* **16**, 507 (2014).
107. Lazaridis, K.N. *et al.* Outcome of Whole Exome Sequencing for Diagnostic Odyssey Cases of an Individualized Medicine Clinic: The Mayo Clinic Experience. *Mayo Clinic Proceedings* **91**, 297-307 (2016).
108. Berberich, A.J., Ho, R. & Hegele, R.A. Whole genome sequencing in the clinic: empowerment or too much information? *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne* **190**, E124-E125 (2018).
109. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-41 (2008).

110. Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics* **Chapter 7**, Unit7 20 (2013).
111. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073-81 (2009).
112. Schwarz, J.M., Cooper, D.N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* **11**, 361-2 (2014).
113. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310-5 (2014).
114. Desmet, F.O. *et al.* Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* **37**, e67 (2009).
115. Xiong, H.Y. *et al.* RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
116. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
117. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
118. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database: the Journal of Biological Databases and Curation* **2017**(2017).
119. Pundir, S., Martin, M.J. & O'Donovan, C. UniProt Protein Knowledgebase. *Methods in Molecular Biology* **1558**, 41-55 (2017).
120. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **44**, D481-7 (2016).
121. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
122. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353-D361 (2017).
123. Stenson, P.D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics* **133**, 1-9 (2014).

124. Pavan, S. *et al.* Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS One* **12**, e0170365 (2017).
125. Firth, H.V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics* **84**, 524-33 (2009).
126. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* **17**, 405-24 (2015).
127. Patel, Z.H. *et al.* The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors. *Frontiers in Genetics* **5**, 16 (2014).
128. Sunyaev, S., Ramensky, V. & Bork, P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends in Genetics* **16**, 198-200 (2000).
129. Wang, Z. & Mout, J. SNPs, protein structure, and disease. *Human Mutation* **17**, 263-70 (2001).
130. Ferrer-Costa, C., Orozco, M. & de la Cruz, X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *Journal of Molecular Biology* **315**, 771-86 (2002).
131. Kucukkal, T.G., Petukh, M., Li, L. & Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Current Opinion in Structural Biology* **32**, 18-24 (2015).
132. Mi, H., Muruganujan, A., Casagrande, J.T. & Thomas, P.D. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* **8**, 1551-66 (2013).
133. Hart, A. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ* **323**, 391-3 (2001).
134. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 1 edn (Springer-Verlag New York, New York, 2009).
135. Palles, C. *et al.* Germline mutations affecting the proofreading domains of *POLE* and *POLD1* predispose to colorectal adenomas and carcinomas. *Nature Genetics* **45**, 136-44 (2013).
136. Weedon, M.N. *et al.* An in-frame deletion at the polymerase active site of *POLD1* causes a multisystem disorder with lipodystrophy. *Nature Genetics* **45**, 947-50 (2013).

137. Nicolas, E., Golemis, E.A. & Arora, S. *POLD1*: Central mediator of DNA replication and repair, and implication in cancer and other pathologies. *Gene* **590**, 128-41 (2016).
138. Kozusko, K. *et al.* Clinical and molecular characterization of a novel *PLIN1* frameshift mutation identified in patients with familial partial lipodystrophy. *Diabetes* **64**, 299-310 (2015).
139. Tansey, J.T. *et al.* Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6494-9 (2001).
140. Jackson, S.N., Howlett, T.A., McNally, P.G., O'Rahilly, S. & Trembath, R.C. Dunnigan-Kobberling syndrome: an autosomal dominant form of partial lipodystrophy. *QJM: Monthly Journal of the Association of Physicians* **90**, 27-36 (1997).
141. Pennacchio, L.A. *et al.* An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294**, 169-73 (2001).
142. Wang, F. *et al.* Apolipoprotein A-IV: a protein intimately involved in metabolism. *Journal of Lipid Research* **56**, 1403-18 (2015).
143. Sharma, V. *et al.* Aberrant hetero-disulfide bond formation by the hypertriglyceridemia-associated p.Gly185Cys *APOA5* variant (rs2075291). *Arteriosclerosis, Thrombosis, and Vascular Biology* **34**, 2254-60 (2014).
144. Zuchner, S. *et al.* Mutations in the mitochondrial GTPase mitofusin 2 cause Charcot-Marie-Tooth neuropathy type 2A. *Nature Genetics* **36**, 449-51 (2004).
145. Hikiami, R. *et al.* Charcot-Marie-Tooth disease type 2A with an autosomal-recessive inheritance: the first report of an adult-onset disease. *Journal of Human Genetics* **63**, 89-92 (2018).
146. Sawyer, S.L. *et al.* Homozygous mutations in *MFN2* cause multiple symmetric lipomatosis associated with neuropathy. *Human Molecular Genetics* **24**, 5109-14 (2015).
147. Konige, M., Wang, H. & Sztalryd, C. Role of adipose specific lipid droplet proteins in maintaining whole body energy homeostasis. *Biochimica et Biophysica Acta* **1842**, 393-401 (2014).
148. Nakagawa, Y. *et al.* CREB3L3 controls fatty acid oxidation and ketogenesis in synergy with PPARalpha. *Scientific Reports* **6**, 39182 (2016).
149. Zhang, C. *et al.* Endoplasmic reticulum-tethered transcription factor cAMP responsive element-binding protein, hepatocyte specific, regulates hepatic

- lipogenesis, fatty acid oxidation, and lipolysis upon metabolic stress in mice. *Hepatology* **55**, 1070-82 (2012).
150. Garg, A., Sankella, S., Xing, C. & Agarwal, A.K. Whole-exome sequencing identifies *ADRA2A* mutation in atypical familial partial lipodystrophy. *JCI Insight* **1**(2016).
 151. Wang, F. *et al.* A *TBX5* 3'UTR variant increases the risk of congenital heart disease in the Han Chinese population. *Cell Discovery* **3**, 17026 (2017).
 152. Cardoso, A.R., Oliveira, M., Amorim, A. & Azevedo, L. Major influence of repetitive elements on disease-associated copy number variants (CNVs). *Human Genomics* **10**, 30 (2016).
 153. Tong, W.H. & Rouault, T. Distinct iron-sulfur cluster assembly complexes exist in the cytosol and mitochondria of human cells. *The EMBO Journal* **19**, 5692-700 (2000).
 154. Li, K., Tong, W.H., Hughes, R.M. & Rouault, T.A. Roles of the mammalian cytosolic cysteine desulfurase, ISCS, and scaffold protein, ISCU, in iron-sulfur cluster assembly. *The Journal of Biological Chemistry* **281**, 12344-51 (2006).
 155. Kollberg, G. *et al.* Clinical manifestation and a new *ISCU* mutation in iron-sulphur cluster deficiency myopathy. *Brain: a Journal of Neurology* **132**, 2170-9 (2009).
 156. Mochel, F. *et al.* Splice mutation in the iron-sulfur cluster scaffold protein ISCU causes myopathy with exercise intolerance. *American Journal of Human Genetics* **82**, 652-60 (2008).
 157. Arikawa-Hirasawa, E., Wilcox, W.R. & Yamada, Y. Dyssegmental dysplasia, Silverman-Handmaker type: unexpected role of perlecan in cartilage development. *American Journal of Medical Genetics* **106**, 254-7 (2001).
 158. Cai, H., Wang, X.L. & Wilcken, D.E. Genetic polymorphism of heparan sulfate proteoglycan (perlecan, HSPG2), lipid profiles and coronary artery disease in the Australian population. *Atherosclerosis* **148**, 125-9 (2000).
 159. Yamashita, Y., Nakada, S., Yoshihara, T., Nara, T., Furuya, N., Miida, T., Hattori, N., and Arikawa-Hirasawa, E. Perlecan, a heparan sulfate proteoglycan, regulates systemic metabolism with dynamic changes in adipose tissue and skeletal muscle. *Scientific Reports* **8**(2018).
 160. Zhang, J.H. *et al.* Association of genetic variations of *PRDM16* with metabolic syndrome in a general Xinjiang Uygur population. *Endocrine* **41**, 539-41 (2012).
 161. Marais, A.D. Familial hypercholesterolaemia. *The Clinical Biochemist, Reviews* **25**, 49-68 (2004).

162. Freije, J.M. *et al.* Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins. *Genomics* **58**, 270-80 (1999).
163. Ahmad, Z., Zackai, E., Medne, L. & Garg, A. Early onset mandibuloacral dysplasia due to compound heterozygous mutations in *ZMPSTE24*. *American Journal of Medical Genetics, Part A* **152A**, 2703-10 (2010).
164. Barrowman, J., Wiley, P.A., Hudon-Miller, S.E., Hrycyna, C.A. & Michaelis, S. Human *ZMPSTE24* disease mutations: residual proteolytic activity correlates with disease severity. *Human Molecular Genetics* **21**, 4084-93 (2012).
165. Duff, R.M. *et al.* Mutations in the N-terminal actin-binding domain of filamin C cause a distal myopathy. *American Journal of Human Genetics* **88**, 729-740 (2011).
166. Chen, Y.Q., Rafi, M.A., de Gala, G. & Wenger, D.A. Cloning and expression of cDNA encoding human galactocerebrosidase, the enzyme deficient in globoid cell leukodystrophy. *Human Molecular Genetics* **2**, 1841-5 (1993).
167. Spratley, S.J. *et al.* Molecular Mechanisms of Disease Pathogenesis Differ in Krabbe Disease Variants. *Traffic* **17**, 908-22 (2016).
168. Wenger, D.A., Rafi, M.A. & Luzi, P. Krabbe disease: One Hundred years from the bedside to the bench to the bedside. *Journal of Neuroscience Research* **94**, 982-9 (2016).
169. Molday, R.S., Zhong, M. & Quazi, F. The role of the photoreceptor ABC transporter ABCA4 in lipid transport and Stargardt macular degeneration. *Biochimica et Biophysica Acta* **1791**, 573-83 (2009).
170. Mori, S. *et al.* Enhanced intra-abdominal visceral fat accumulation in patients with Werner's syndrome. *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* **25**, 292-5 (2001).
171. Kidambi, S. & Patel, S.B. Cholesterol and non-cholesterol sterol transporters: ABCG5, ABCG8 and NPC1L1: a review. *Xenobiotica; the Fate of Foreign Compounds in Biological Systems* **38**, 1119-39 (2008).
172. Niu, D.M. *et al.* Clinical observations, molecular genetic analysis, and treatment of sitosterolemia in infants and children. *Journal of Inherited Metabolic Disease* **33**, 437-43 (2010).
173. Cowling, B.S. *et al.* Four and a half LIM protein 1 gene mutations cause four distinct human myopathies: a comprehensive review of the clinical, histological and pathological features. *Neuromuscular Disorders* **21**, 237-51 (2011).

174. Malfatti, E. *et al.* Skeletal muscle biopsy analysis in reducing body myopathy and other *FHLL1*-related disorders. *Journal of Neuropathology and Experimental Neurology* **72**, 833-45 (2013).
175. Gbadegesin, R.A. *et al.* Mutations in the gene that encodes the F-actin binding protein anillin cause FSGS. *Journal of the American Society of Nephrology* **25**, 1991-2002 (2014).
176. Mead, J.R., Irvine, S.A. & Ramji, D.P. Lipoprotein lipase: structure, function, regulation, and role in disease. *Journal of Molecular Medicine* **80**, 753-69 (2002).
177. Xu, Z.H., Freimuth, R.R., Eckloff, B., Wieben, E. & Weinshilboum, R.M. Human 3'-phosphoadenosine 5'-phosphosulfate synthetase 2 (*PAPSS2*) pharmacogenetics: gene resequencing, genetic polymorphisms and functional characterization of variant allozymes. *Pharmacogenetics* **12**, 11-21 (2002).
178. Oostdijk, W. *et al.* *PAPSS2* deficiency causes androgen excess via impaired DHEA sulfation - *in vitro* and *in vivo* studies in a family harboring two novel *PAPSS2* mutations. *The Journal of Clinical Endocrinology and Metabolism* **100**, E672-80 (2015).
179. Fajans, S.S., Bell, G.I. & Polonsky, K.S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *The New England Journal of Medicine* **345**, 971-80 (2001).
180. Cui, X. *et al.* Seipin ablation in mice results in severe generalized lipodystrophy. *Human Molecular Genetics* **20**, 3022-30 (2011).
181. Law, S.W. *et al.* Human apolipoprotein B-100: cloning, analysis of liver mRNA, and assignment of the gene to chromosome 2. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8340-4 (1985).
182. Schmidt, R.J. *et al.* Secreted proprotein convertase subtilisin/kexin type 9 reduces both hepatic and extrahepatic low-density lipoprotein receptors *in vivo*. *Biochemical and Biophysical Research Communications* **370**, 634-40 (2008).
183. Gautron, S. *et al.* Molecular mechanisms of McArdle's disease (muscle glycogen phosphorylase deficiency). RNA and DNA analysis. *The Journal of Clinical Investigation* **79**, 275-81 (1987).
184. Bruno, C. *et al.* McArdle disease: the mutation spectrum of *PYGM* in a large Italian cohort. *Human Mutation* **27**, 718 (2006).
185. Simha, V. & Garg, A. Phenotypic heterogeneity in body fat distribution in patients with congenital generalized lipodystrophy caused by mutations in the *AGPAT2* or seipin genes. *The Journal of Clinical Endocrinology and Metabolism* **88**, 5433-7 (2003).

186. Frye, R.A. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochemical and Biophysical Research Communications* **273**, 793-8 (2000).
187. Lombard, D.B., Schwer, B., Alt, F.W. & Mostoslavsky, R. SIRT6 in DNA repair, metabolism and ageing. *Journal of Internal Medicine* **263**, 128-41 (2008).
188. Rodgers, J.T. & Puigserver, P. Certainly can't live without this: SIRT6. *Cell Metabolism* **3**, 77-8 (2006).
189. Zhong, L. & Mostoslavsky, R. SIRT6: a master epigenetic gatekeeper of glucose metabolism. *Transcription* **1**, 17-21 (2010).
190. Mostoslavsky, R. *et al.* Genomic instability and aging-like phenotype in the absence of mammalian SIRT6. *Cell* **124**, 315-29 (2006).
191. Islam, K.B., Rabbani, H., Larsson, C., Sanders, R. & Smith, C.I. Molecular cloning, characterization, and chromosomal localization of a human lymphoid tyrosine kinase related to murine Blk. *Journal of Immunology* **154**, 1265-72 (1995).
192. Borowiec, M. *et al.* Mutations at the *BLK* locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 14460-5 (2009).
193. Koemans, T.S. *et al.* Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. *PLoS Genetics* **13**, e1006864 (2017).
194. Kleefstra, T. *et al.* Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *American Journal of Human Genetics* **91**, 73-82 (2012).
195. Neas, K.R. *et al.* Three patients with terminal deletions within the subtelomeric region of chromosome 9q. *American Journal of Medical Genetics, Part A* **132A**, 425-30 (2005).
196. Stoetzel, C. *et al.* BBS10 encodes a vertebrate-specific chaperonin-like protein and is a major BBS locus. *Nature Genetics* **38**, 521-4 (2006).
197. Okamura, M., Inagaki, T., Tanaka, T. & Sakai, J. Role of histone methylation and demethylation in adipogenesis and obesity. *Organogenesis* **6**, 24-32 (2010).
198. Chehab, F.F. Obesity and lipodystrophy - where do the circles intersect? *Endocrinology* **149**, 925-34 (2008).
199. Navarro, G., Allard, C., Xu, W. & Mauvais-Jarvis, F. The role of androgens in metabolism, obesity, and diabetes in males and females. *Obesity* **23**, 713-9 (2015).

200. Kelly, D.M. *et al.* Testosterone differentially regulates targets of lipid and glucose metabolism in liver, muscle and adipose tissues of the testicular feminised mouse. *Endocrine* **54**, 504-515 (2016).
201. Sato, T. *et al.* Late onset of obesity in male androgen receptor-deficient (AR KO) mice. *Biochemical and Biophysical Research Communications* **300**, 167-71 (2003).
202. Yang, X. *et al.* A locus for brachydactyly type A-1 maps to chromosome 2q35-q36. *American Journal of Human Genetics* **66**, 892-903 (2000).
203. Hammerschmidt, M., Brook, A. & McMahon, A.P. The world according to hedgehog. *Trends in Genetics* **13**, 14-21 (1997).
204. Gao, B. *et al.* A mutation in *Ihh* that causes digit abnormalities alters its signalling capacity and range. *Nature* **458**, 1196-200 (2009).
205. St-Jacques, B., Hammerschmidt, M. & McMahon, A.P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes and Development* **13**, 2072-86 (1999).
206. Hellemans, J. *et al.* Homozygous mutations in *IHH* cause acrocapitofemoral dysplasia, an autosomal recessive disorder with cone-shaped epiphyses in hands and hips. *American Journal of Human Genetics* **72**, 1040-6 (2003).
207. Ma, G. *et al.* Indian hedgehog mutations causing brachydactyly type A1 impair Hedgehog signal transduction at multiple levels. *Cell Research* **21**, 1343-57 (2011).
208. Santaguida, S. & Musacchio, A. The life and miracles of kinetochores. *The EMBO Journal* **28**, 2511-31 (2009).
209. Genin, A. *et al.* Kinetochores KMN network gene *CASC5* mutated in primary microcephaly. *Human Molecular Genetics* **21**, 5306-17 (2012).
210. Woods, C.G., Bond, J. & Enard, W. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *American Journal of Human Genetics* **76**, 717-28 (2005).
211. Faheem, M. *et al.* Molecular genetics of human primary microcephaly: an overview. *BMC Medical Genomics* **8 Suppl 1**, S4 (2015).
212. Bolanos-Garcia, V.M. *et al.* Structure of a Blinkin-BUBR1 complex reveals an interaction crucial for kinetochores-mitotic checkpoint regulation via an unanticipated binding site. *Structure* **19**, 1691-700 (2011).
213. Hanks, S. *et al.* Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in *BUB1B*. *Nature Genetics* **36**, 1159-61 (2004).

214. Payne, M., Rupar, C.A., Siu, G.M. & Siu, V.M. Amish, Mennonite, and Hutterite genetic disorder database. *Paediatrics & Child Health* **16**, e23-4 (2011).
215. Karlsson, P. *et al.* Loss of vps54 function leads to vesicle traffic impairment, protein mis-sorting and embryonic lethality. *International Journal of Molecular Sciences* **14**, 10908-25 (2013).
216. Sugimoto, M. *et al.* Molecular identification of t(w5): Vps52 promotes pluripotential cell differentiation through cell-cell interactions. *Cell Reports* **2**, 1363-74 (2012).
217. Kapur, S., Swinford, A., Freimanis, A.K. & Lachman, R.S. Apparently previously undescribed X-linked dominant syndrome with facial and skeletal anomalies. *American Journal of Medical Genetics* **33**, 357-63 (1989).
218. Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T. & Kajii, T. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *The Journal of Pediatrics* **99**, 565-9 (1981).
219. Lederer, D. *et al.* Deletion of *KDM6A*, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *American Journal of Human Genetics* **90**, 119-24 (2012).
220. Yarema, K.J. & Bertozzi, C.R. Characterizing glycosylation pathways. *Genome Biology* **2**, (2001).
221. Rose, M.C. & Voynow, J.A. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiological Reviews* **86**, 245-78 (2006).
222. Fowler, J., Vinall, L. & Swallow, D. Polymorphism of the human muc genes. *Frontiers in Bioscience: a Journal and Virtual Library* **6**, D1207-15 (2001).
223. Biteau, B., Hochmuth, C.E. & Jasper, H. Maintaining tissue homeostasis: dynamic control of somatic stem cell activity. *Cell Stem Cell* **9**, 402-11 (2011).
224. Alfares, A. *et al.* Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* (2018).
225. Mallewa, J.E. *et al.* HIV-associated lipodystrophy: a review of underlying mechanisms and therapeutic options. *The Journal of Antimicrobial Chemotherapy* **62**, 648-60 (2008).
226. Mahmood, T. & Yang, P.C. Western blot: technique, theory, and trouble shooting. *North American Journal of Medical Sciences* **4**, 429-34 (2012).

227. Klein, D. Quantification using real-time PCR technology: applications and limitations. *Trends in Molecular Medicine* **8**, 257-60 (2002).
228. Jozefczuk, J. & Adjaye, J. Quantitative real-time PCR-based analysis of gene expression. *Methods in Enzymology* **500**, 99-109 (2011).
229. Yaciuk, P. Co-immunoprecipitation of protein complexes. *Methods in Molecular Medicine* **131**, 103-11 (2007).
230. Foulkes, W.D. & Real, F.X. Many mosaic mutations. *Current Oncology* **20**, 85-7 (2013).
231. Biesecker, L.G. & Spinner, N.B. A genomic view of mosaicism and human disease. *Nature Reviews Genetics* **14**, 307-20 (2013).
232. Furnham, N., Ruffle, S. & Southan, C. Splice variants: a homology modeling approach. *Proteins* **54**, 596-608 (2004).
233. Riethoven, J.J. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods in Molecular Biology* **674**, 33-42 (2010).
234. Schaffer, A.A. Digenic inheritance in medical genetics. *Journal of Medical Genetics* **50**, 641-52 (2013).
235. Lewis, C.M. & Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Medicine* **9**, 96 (2017).
236. Philippakis, A.A. *et al.* The Matchmaker Exchange: a platform for rare disease gene discovery. *Human Mutation* **36**, 915-21 (2015).
237. Vouillarmet, J. & Laville, M. A Case of Familial Partial Lipodystrophy: From Clinical Phenotype to Genetics. *Canadian Journal of Diabetes* **40**, 376-378 (2016).

Appendices

Appendix I: Ethics Approval



Date: 23 April 2018

To: Robert Hegele

Project ID: 0379

Study Title: Candidate gene sequencing, genetic and genomic analysis for identification of new genetic determinants of intermediate traits of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population. 07920E

Reference Number/ID: N/A

Application Type: HSREB Amendment Form

Review Type: Delegated

Full Board Reporting Date: 15MAY2018

Date Approval Issued: 23/Apr/2018 16:05

REB Approval Expiry Date: 03/Nov/2018

Dear Robert Hegele ,

The Western University Health Sciences Research Ethics Board (HSREB) has reviewed and approved the WREM application form for the amendment, as of the date noted above.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
18 April 2018, REB #379 Delegated HSREB Form, addition of Sub Study, Ethics revisions	Western Protocol	18/Apr/2018	18/April2018
27 Feb 2018 Tissue Sub Study	Consent Form	27/Feb/2018	27 Feb 2018
27 Feb 2018, Email recruitment script, rough draft	Recruitment	27/Feb/2018	27 Feb 2018
27 Feb 2018, Telephone_Script (Receiving Calls from Potential Participants)	Consent Form	27/Feb/2018	27 Feb 2018
27 Feb 2018, Universal Genetics DNA Consent Form - Local Participants	Consent Form	27/Feb/2018	27 Feb 2018
27 Feb 2018, Universal Genetics DNA Consent Form - Non-local Participants, REB revisions	Consent Form	27/Feb/2018	27 Feb 2018

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB 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 HSREB 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,

Nicola Geoghegan-Morphet, Ethics Officer on behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix II: Consent Form



LETTER OF INFORMATION

Study Title: Candidate gene sequencing, phenomics, genetic and genomic analysis for identification of new genetic determinants of intermediate traits and of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population.

STUDY INVESTIGATOR: Dr. Robert Hegele

Purpose of the Study

We are conducting a research study looking at **genetic** contributions to atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy - a condition that affects where fat is deposited in the body and other rare metabolic or cardiovascular disorders in the human population. Earlier studies have suggested that some **genetic** differences in special populations are linked to these conditions. Understanding the **genetic** basis of these conditions will be helpful for developing new treatments and preventative measures. Genetic material is contained in every cell including blood cells and is a unique fingerprint of all body activities and tissues.

During the course of your treatment at the Lipid Genetics Clinic you will have a number of tests and treatments done as part of your regular care and a great deal of information about your past and current medical history will also be collected. This is all done as part of your standard care to help us understand how to treat you best.

The physicians and staff at the Lipid Genetics Clinic are engaged in ongoing research to better understand how to treat your medical condition. We are asking for your permission to use the information collected during your visits at the Lipid Genetics Clinic for research purposes. All patients who attend our clinic will be asked to participate.

It is important for you to know why the data is being collected and the research is being conducted and what we are asking you to do. Please take time to read this carefully and feel free to ask questions if anything is unclear.

The pronouns 'you' and 'your' in the letter should be read as referring to the patient rather than the parent/guardian who is signing the consent form for the patient.

Procedures

If you agree to participate, you will provide a single blood sample of 8ml or approximately 1.5 teaspoons. We will use this blood and extract your genetic material (DNA or deoxyribonucleic acid) from it for the purpose of diagnosis/research/storage. The procedure of taking blood may have to be repeated if results are not clear. You will be provided with the results of your blood tests.

Genetic analysis may yield information on other diseases. It may also reveal information on paternity if more than one family member is tested. If such information should arise, it would not be pursued further and you would not be informed unless you wish to be.

Some of the blood taken during the study will be saved and stored (archived) by Dr. R. Hegele until the sample is all used up. Two tubes of blood (20mL or 4 tsp) will be taken from your arm. It will be used only for extra tests related to the study to understand its results. The information will be used to research how blood components are related to cardiovascular disease and discrete genetic conditions. No genetic or DNA testing will be done on these blood samples.

Medical and Family History:

If you agree to participate, data relating to your health history and current care will be copied from the Lipid Genetics Clinic's clinical database to a separate database. All identifying information such as your name, address and OHIP number will be removed. The information in the research database will be identified by a unique code number that will link the research record with your health record and will allow the information in the research database to be updated during your time with the clinic. The master list that contains the link to the code number and your name and other identifying information will be kept in a very secure location under the control of Dr. Hegele. The research database will be owned by Dr. Hegele and it will be stored in a secure location on a computer in the Blackburn Cardiovascular Genetics Laboratory. The data will be retained or to allow us to look at trends over time.

The database will also help us identify those patients who if you agree may be eligible to participate in future research projects that involve more than just an analysis of existing data. In the future you may be approached to participate in other research projects in the Lipid Genetics Clinic. In those instances you will be given detailed information describing the project and you will have the opportunity to decide at that time, whether or not you want to participate in the new project.

You may be asked to release hospital medical records to the Dr. Hegele and his designated research team pertaining to the associated medical condition. It may be necessary to contact you in the future to obtain future additional information if necessary.

Risks and Discomforts

One risk of the study may arise from the discomfort associated with having your blood taken. This amount of blood is about 2 teaspoons, and the withdrawal of blood will not harm you in any way but there is a very small risk of bruising and infection at the needle site. If you are already participating in another research project at this time, please inform Dr. Hegele promptly to determine if it is appropriate to begin participation in this study.

The only other known risk to your participation in this study is the possibility that, because the research database is linked to our clinical database, someone may be able to identify you. However the research database is secured in the same manner as our clinical records and access is limited to authorized personnel only.

Benefits

There is no evidence indicating that your condition will be improved by providing us with a blood sample for genetic analysis, although there is a possibility that what we learn may help you, your family and other people with your genetic makeup. Genetic Counselling is available if you request it.

You will not benefit directly from participation in this research however the results of our research may help other patients in the future who suffer from problems similar to yours.

Withdrawal

Participation in the study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no effect on your future care. You may have your stored sample of DNA destroyed at any time. If you decide to withdraw from the study sometime in the future, your DNA and research data will be destroyed, whenever this is possible. Information that has already been transferred to the research database will not be able to be withdrawn as it would compromise the findings of the research that had already been completed. To withdraw your consent, notify Dr. Robert Hegele or his staff in writing.

Regardless of your decision to participate you can still receive continuing care through this clinic.

What are the costs of participating?

There will be no charge to you, your private medical insurance (if any) or the provincial health insurance plan for procedures given to you for this study. If you have any questions, please ask the study doctor or a member of his staff. You will not receive compensation for your participation.

Will the information collected be confidential?

All your information from this study will be kept confidential, available only to Dr. Hegele and his designated research team. You will be identified by a study number, with your name and other data kept confidential. When results are published your name will not be used. Your DNA will be stored at the Blackburn Cardiovascular Genetics Laboratory at Robarts Research Institute, directed by Dr. Robert Hegele in London Ontario in a vial. It will be stored there for 15 years and then destroyed. DNA analysis for individual or family studies will be done at the Blackburn Cardiovascular Genetics Laboratory.

Representatives of The University of Western Ontario Health Sciences Research Ethics Board may require access to your study-related records or may follow-up with you to monitor the conduct of your research.

Your study number will be linked to a locked electronic master database that contains your information, so that Dr. Hegele can review your results and advise medical treatment if needed.

If you have provided us with blood, it is your responsibility to inform the Blackburn Cardiovascular Genetics Laboratory at the Robarts Research Institute of any change of address. To obtain a renewed consent the laboratory is obliged to inform you, of any changes in policy, inability to maintain sample quality or any possible treatments. If you fail to inform this Centre of your change of address, then the present consent form is valid and final until indicated otherwise.

What happens now?

You are free to choose whether or not you want to take part in this study. You may talk to family members or friends before you make your decision. If you do not wish to take part in this study, tell your study doctor and he will make all the arrangements. Please ask the study doctor or his staff any questions about the study, to make sure you fully understand what will happen if you agree to take part.

Voluntary Participation

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no effect on your future care, or academic or employment status. You do not waive any legal rights by signing the consent form.

Alternatives to Study Participation

If you decide not to participate in this study, this will not affect the standard of care you receive.

Consent Form

Study Title: Candidate gene sequencing, phenomics, genetic and genomic analysis for identification of new genetic determinants of intermediate traits and of atherosclerosis, dyslipidemia, diabetes, obesity, hypertension, lipodystrophy and other rare metabolic or cardiovascular disorders in the human population.

I have read the accompanying letter of information, have had the nature of the study explained to me, and I agree to participate. All questions have been answered to my satisfaction. A copy of the information and signed consent form has been given to me to keep.

If research with your DNA reveals some other medical condition relating to you, do you wish your family doctor to be informed? Yes _____ No _____

If you answered yes to the question above, please fill in your family doctor's name and address below.

_____ Yes, I do / _____ No, I do not wish to be contacted in the future in order to provide additional information.

_____ I do not have a family doctor

Printed name of patient

Signature of patient

Date

Signature of parent/guardian
(if applicable)

Date

Printed Name of Study Physician or Investigator

Signature of Study Physician or Investigator

Date

Printed name of Person Responsible for obtaining consent of Participant

Signature of Person Responsible for obtaining consent of Participant

Date

Appendix III: Primer List

Gene	Primer Names	Primer Sequence (5'-3')
<i>IHH</i>	Forward	CTCGCCTACAAGCAGTTCAGC
	Reverse	AAAATGTGCCAGGGGGTGGGTAG
<i>TMPRSS4</i>	Forward	CCGGGAAAGGCTGAGTAACAAAAC
	Reverse	GATAGTCTTAGCCCAGGTTTCTC
<i>AR*</i>	AR1-F1	GTTGAACTCTTCTGAGCAAGAG
	AR1-F2	GGATGAGGAACAGCAACCTTCAC
	AR1-R5	GTGAAGGTTGCTGTTCCATCC
	AR1-F3	CCAAGGACAATTACTTAGGGG
	AR1-R4	CCCCTAAGTAATTGTCCTTGG
	AR1-F4	CAAAGGGCTAGAAGGCGAGAGCC
	AR1-R3	GGCTCTCGCCTTCTAGCCCTTTG
	AR1-F5	GAAGAAGGCCAGTTGTATGG
	AR1-R2	CCATACAACCTGGCCTTCTTC
	AR1-R1	GAACACAGAGTGACTCTGCCCTG
	AR2-FWD	GACCTGAGACTTCACTTGCC
	AR2-REV	CCCTGAAAGGTTAGTGTCTCTC
	AR3-FWD	CCCGAAGAAAGAGACTCTGGAAAC
	AR3-REV	GACTAGAAAATGAGGGAGAAGGGG
	AR4-FWD	CTATGTCACCCAGGCTAGAGC
	AR4-REV	GGATTTACCAGGCAAAGGTG
	AR5-FWD	CTCAACCCGTCAGTACCCAG
	AR5-REV	CCATCACCACCAACCAGGTCTG
	AR6-FWD	GAGGGATGGCAATCAGAGAC
	AR6-REV	GTCCTCTCTGAATCTCTGTGC
AR7-FWD	CTAATGCTCCTTCGTGGGCATG	
AR7-REV	CATTGGCTCTATCAGGCTGTTC	
AR8-FWD	GAAGAGGCTAGCAGAGGCCAC	
AR8-REV	CAGGCAGAAGACATCTGAAAGG	
<i>PLIN1</i>	Forward	CTTCAGCTCTGCCACCTACC
	Reverse	AAGAGGAGCCTGTTGTGTGG
<i>VPS52</i>	Forward	GCTAAGAGAACAGCGAAGGTTCCG
	Reverse	CAGCTCAGGTCTTTCTGAAGCTAGG
<i>KDM6A</i>	Forward	CATCTCATGGACTTGTGCAAATGCC
	Reverse	GTGGTCTTGGAGGTGGACATTTATC


*AR primer naming method: geneexon-**F**orward/**R**everse(number if necessary)

Appendix IV: PANTHER gene lists

Patient 14318, *O*-glycan processing

New! [PANTHER Generic Mapping Package](#)

PANTHER GENE LIST [Customize Gene list](#)

Convert List to: -Select- Send list to: -Select- 

Display: 30 items per page [Refine Search](#)

Hits 1-11 of 11 [page: (1)] Number of mapped ids found 11

<input type="checkbox"/>	<input type="checkbox"/>	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily <input checked="" type="checkbox"/>	PANTHER Protein Class <input checked="" type="checkbox"/>	Species <input checked="" type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	1. HUMAN HGNC=7514 UniProtKB=Q99102	MUC4	Mucin-4 MUC4 ortholog	MUCIN-4 (PTHR13802:SF53)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	2. HUMAN HGNC=7512 UniProtKB=Q02817	MUC2	Mucin-2 MUC2 ortholog	MUCIN-2 (PTHR11339:SF261)	protease inhibitor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	3. HUMAN HGNC=26639 UniProtKB=Q8NBI6	XXYL1	Xyloside xylosyltransferase 1 XXYL1 ortholog	XYLOSIDE XYLOSYLTRANSFERASE 1 (PTHR46612:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	4. HUMAN HGNC=7510 UniProtKB=Q9UKN1	MUC12	Mucin-12 MUC12 ortholog	MUCIN-12 (PTHR24041:SF25)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	5. HUMAN HGNC=15582 UniProtKB=Q8WXI7	MUC16	Mucin-16 MUC16 ortholog	MUCIN-16 (PTHR14672:SF0)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	6. HUMAN HGNC=16800 UniProtKB=Q685J3	MUC17	Mucin-17 MUC17 ortholog	MUCIN-17 (PTHR37999:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	7. HUMAN HGNC=23282 UniProtKB=Q8N307	MUC20	Mucin-20 MUC20 ortholog	MUCIN-20 (PTHR37358:SF1)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	8. HUMAN HGNC=7516 UniProtKB=Q9HC84	MUC5B	Mucin-5B MUC5B ortholog	MUCIN-5B (PTHR11339:SF269)	protease inhibitor	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	9. HUMAN HGNC=27482 UniProtKB=Q4G148	GXYLT1	Glucoside xylosyltransferase 1 GXYLT1 ortholog	GLUCOSIDE XYLOSYLTRANSFERASE 1 (PTHR46012:SF3)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	10. HUMAN HGNC=4130 UniProtKB=Q9NY28	GALNT8	Probable polypeptide N-acetylgalactosaminyltransferase 8 GALNT8 ortholog	POLYPEPTIDE N-ACETYL GALACTOSAMINYLTRANSFERASE 8-RELATED (PTHR11675:SF50)	glycosyltransferase	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	11. HUMAN HGNC=7517 UniProtKB=Q6W4X9	MUC6	Mucin-6 MUC6 ortholog	MUCIN-6 (PTHR11339:SF264)	protease inhibitor	Homo sapiens

Patient 14318, tissue homeostasis

	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily	PANTHER Protein Class	Species
<input type="checkbox"/>	1. HUMAN HGNC=7514 UniProtKB=Q99102	MUC4	Mucin-4 MUC4 ortholog	MUCIN-4 (PTHR13802-SF53)	-	Homo sapiens
<input type="checkbox"/>	2. HUMAN HGNC=2548 UniProtKB=Q60484	CUBN	Cubilin CUBN ortholog	CUBILIN (PTHR47937-SF2)	-	Homo sapiens
<input type="checkbox"/>	3. HUMAN HGNC=14637 UniProtKB=Q86UK0	ABCA12	ATP-binding cassette sub-family A member 12 ABCA12 ortholog	ATP-BINDING CASSETTE SUB-FAMILY A MEMBER 12 (PTHR19229-SF29)	ATP-binding_cassette_(ABC)_transporter	Homo sapiens
<input type="checkbox"/>	4. HUMAN HGNC=37093 UniProtKB=POCG38	POTE1	POTE ankyrin domain family member 1 POTE1 ortholog	POTE ANKYRIN DOMAIN FAMILY MEMBER E-RELATED (PTHR24118-SF56)	-	Homo sapiens
<input type="checkbox"/>	5. HUMAN HGNC=7512 UniProtKB=Q02817	MUC2	Mucin-2 MUC2 ortholog	MUCIN-2 (PTHR11339-SF261)	protease inhibitor	Homo sapiens
<input type="checkbox"/>	6. HUMAN HGNC=393 UniProtKB=Q9Y243	AKT3	RAC-gamma serine/threonine-protein kinase AKT3 ortholog	RAC-GAMMA SERINE/THREONINE-PROTEIN KINASE (PTHR24356-SF190)	annexin calmodulin non-receptor serine/threonine protein kinase transfer/carrier protein	Homo sapiens
<input type="checkbox"/>	7. HUMAN HGNC=2536 UniProtKB=P43235	CTSO	Cathepsin K CTSK ortholog	CATHEPSIN K (PTHR12411-SF55)	cysteine protease protease inhibitor	Homo sapiens
<input type="checkbox"/>	8. HUMAN HGNC=952 UniProtKB=Q99728	BARD1	BRCA1-associated RING domain protein 1 BARD1 ortholog	BRCA1-ASSOCIATED RING DOMAIN PROTEIN 1 (PTHR24171-SF8)	-	Homo sapiens
<input type="checkbox"/>	9. HUMAN HGNC=381 UniProtKB=P15121	AR	Aldose reductase AKR1B1 ortholog	ALDOSE REDUCTASE (PTHR11732-SF294)	reductase	Homo sapiens
<input type="checkbox"/>	10. HUMAN HGNC=12601 UniProtKB=Q75445	USH2A	Usherin USH2A ortholog	USHERIN (PTHR10574-SF274)	extracellular matrix linker protein protease inhibitor receptor	Homo sapiens
<input type="checkbox"/>	11. HUMAN HGNC=990 UniProtKB=P10415	BCL2	Apoptosis regulator Bcl-2 BCL2 ortholog	APOPTOSIS REGULATOR BCL-2 (PTHR11256-SF11)	signaling molecule	Homo sapiens
<input type="checkbox"/>	12. HUMAN HGNC=33895 UniProtKB=Q65813	POTE6	POTE ankyrin domain family member 6 POTE6 ortholog	POTE ANKYRIN DOMAIN FAMILY MEMBER E-RELATED (PTHR24118-SF56)	-	Homo sapiens
<input type="checkbox"/>	13. HUMAN HGNC=3473 UniProtKB=Q99718	ESRRB	Steroid hormone receptor ERR2 ESRRB ortholog	STEROID HORMONE RECEPTOR ERR2 (PTHR24084-SF23)	C4 zinc finger nuclear receptor nucleic acid binding receptor	Homo sapiens
<input type="checkbox"/>	14. HUMAN HGNC=17416 UniProtKB=Q8WXG9	GPR98	G-protein coupled receptor 98 GPR98 ortholog	G-PROTEIN COUPLED RECEPTOR 98 (PTHR46682-SF1)	-	Homo sapiens
<input type="checkbox"/>	15. HUMAN HGNC=19104 UniProtKB=Q75161	NPHP4	Nephrocystin-4 NPHP4 ortholog	NEPHROCYSTIN-4 (PTHR31043-SF3)	-	Homo sapiens
<input type="checkbox"/>	16. HUMAN HGNC=30695 UniProtKB=Q8NI35	WDR36	WD repeat-containing protein 36 WDR36 ortholog	WD REPEAT-CONTAINING PROTEIN 36 (PTHR22840-SF12)	-	Homo sapiens
<input type="checkbox"/>	17. HUMAN HGNC=3956 UniProtKB=Q14623	IHH	Indian hedgehog protein IHH ortholog	INDIAN HEDGEHOG PROTEIN (PTHR11889-SF39)	-	Homo sapiens
<input type="checkbox"/>	18. HUMAN HGNC=15946 UniProtKB=Q81WN7	RP111	Retinitis pigmentosa 1-like 1 protein RP111 ortholog	RETINITIS PIGMENTOSA 1-LIKE 1 PROTEIN (PTHR23005-SF3)	-	Homo sapiens
<input type="checkbox"/>	19. HUMAN HGNC=7517 UniProtKB=Q6W4X9	MUC6	Mucin-6 MUC6 ortholog	MUCIN-6 (PTHR11339-SF264)	protease inhibitor	Homo sapiens
<input type="checkbox"/>	20. HUMAN HGNC=18608 UniProtKB=Q38502	LRRK1	Leucine-rich repeat serine/threonine-protein kinase 1 LRRK1 ortholog	LEUCINE-RICH REPEAT SERINE/THREONINE-PROTEIN KINASE 1 (PTHR45752-SF49)	-	Homo sapiens
<input type="checkbox"/>	21. HUMAN HGNC=6412 UniProtKB=P04264	KRT1	Keratin, type II cytoskeletal 1 KRT1 ortholog	KERATIN, TYPE II CYTOSKELETAL 1 (PTHR45616-SF33)	-	Homo sapiens

Patient 11618, interferon-gamma-mediated signaling

<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> all	Gene ID	Mapped IDs	Gene Name Gene Symbol Ortholog	PANTHER Family/Subfamily	PANTHER Protein Class	Species
<input type="checkbox"/>	<input type="checkbox"/>	1. HUMAN HGNC=4940 UniProtKB=P04440	HLA-DPB1	HLA class II histocompatibility antigen, DP beta 1 chain HLA-DPB1 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DP BETA 1 CHAIN (PTHR19944:SF46)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	2. HUMAN HGNC=4953 UniProtKB=Q30154	HLA-DRB5	HLA class II histocompatibility antigen, DR beta 5 chain HLA-DRB5 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR BETA 4 CHAIN-RELATED (PTHR19944:SF56)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	3. HUMAN HGNC=4944 UniProtKB=P01920	HLA-DQB1	HLA class II histocompatibility antigen, DQ beta 1 chain HLA-DQB1 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ BETA 1 CHAIN (PTHR19944:SF85)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	4. HUMAN HGNC=7656 UniProtKB=P13591	NCAM1	Neural cell adhesion molecule 1 NCAM1 ortholog	NEURAL CELL ADHESION MOLECULE 1 (PTHR12231:SF239)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	5. HUMAN HGNC=4948 UniProtKB=P01911	HLA-DRB1	HLA class II histocompatibility antigen, DRB1-15 beta chain HLA-DRB1 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR BETA 4 CHAIN-RELATED (PTHR19944:SF56)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	6. HUMAN HGNC=4932 UniProtKB=Q31612	HLA-B	HLA class I histocompatibility antigen, B-73 alpha chain HLA-B ortholog	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-73 ALPHA CHAIN (PTHR16675:SF248)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	7. HUMAN HGNC=4943 UniProtKB=P01906	HLA-DQA2 HLA-DQA1	HLA class II histocompatibility antigen, DQ alpha 2 chain HLA-DQA2 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ ALPHA 2 CHAIN (PTHR19944:SF54)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	8. HUMAN HGNC=4933 UniProtKB=Q95604	HLA-C	HLA class I histocompatibility antigen, Cw-17 alpha chain HLA-C ortholog	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, CW-17 ALPHA CHAIN (PTHR16675:SF247)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	9. HUMAN HGNC=4931 UniProtKB=P01891	HLA-A	HLA class I histocompatibility antigen, A-68 alpha chain HLA-A ortholog	HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A-68 ALPHA CHAIN (PTHR16675:SF245)	-	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	10. HUMAN HGNC=4938 UniProtKB=P20036	HLA-DPA1	HLA class II histocompatibility antigen, DP alpha 1 chain HLA-DPA1 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DP ALPHA 1 CHAIN (PTHR19944:SF64)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	11. HUMAN HGNC=4942 UniProtKB=P01909	HLA-DQA1	HLA class II histocompatibility antigen, DQ alpha 1 chain HLA-DQA1 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ ALPHA 1 CHAIN (PTHR19944:SF59)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	12. HUMAN HGNC=4945 UniProtKB=P05538	HLA-DQB2	HLA class II histocompatibility antigen, DQ beta 2 chain HLA-DQB2 ortholog	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ BETA 2 CHAIN (PTHR19944:SF68)	major histocompatibility complex antigen	Homo sapiens
<input type="checkbox"/>	<input type="checkbox"/>	13. HUMAN HGNC=6120 UniProtKB=Q13568	IRF5	Interferon regulatory factor 5 IRF5 ortholog	INTERFERON REGULATORY FACTOR 5 (PTHR11949:SF10)	nucleic acid binding winged helix/forkhead transcription factor	Homo sapiens

Curriculum Vitae

Rosettia Ho

EDUCATION

Master of Science 2018

Supervisor: Dr. Robert A. Hegele

Department of Biochemistry, Schulich School of Medicine and Dentistry

Robarts Research Institute, Western University, London ON, Canada

Bachelor of Science, Honours Specialization Genetics and Biochemistry 2016

Western University, London ON

EMPLOYMENT

Graduate Teaching Assistant January – April 2018

Biochemistry 4455G: Translation in Cancer Biology, Western University, London ON

Graduate Teaching Assistant September – December 2016 and 2017

Biochemistry 4450A: Molecular Genetics of Human Cancer, Western University, London ON

Graduate Teaching Assistant January – April 2017

Biochemistry 3380G: Biochemistry Laboratory, Western University, London ON

Research Assistant (Part-time) September 2014 – July 2016

Dr. Kwok and Dr. Tam, Department of Social Work, King's University College, London ON

- Involved in the following research studies: Chinese Youth in Conflict with the Law, Violence Against Women in China, and Examining Gatekeeping and Assessment of Professional Suitability in Canadian Social Work Education
- Responsible for transcribing qualitative interviews, attending workshops, attending qualitative interviews, and generating bibliographies for grant proposals

Research Assistant (Full-time Summer) May – September 2014

Dr. Hegele, Blackburn Cardiovascular Genetics Lab, Robarts Research Institute, London ON

- Projects included: Identifying the Molecular Basis of Lipodystrophy Through Next-Generation Sequencing, Absence of Protective *APOC3* Mutations in Dyslipidemic Patients, *DYRK1B* R102C Mutations in Uncharacterized Lipodystrophy Patients, and assisting with Provisional Patent Application – LipidSeq™ and ONDRISeq™

LABORATORY EXPERIENCE: TECHNIQUES/SKILLS

- Primer design
- Gel electrophoresis
- Protein purification
- Whole-exome sequencing preparation
- DNA gel isolation
- Polymerase chain reaction
- Sanger sequencing
- DNA/RNA extraction
- Sequence alignment and analysis (DNA, RNA, protein)
- Bioinformatics processing
- Bioinformatics and statistical tools

ADDITIONAL SKILLS

Languages – English, Cantonese

Computer Software – Microsoft Excel, Publisher, Access, PowerPoint; CLC Bio Genomics Workbench; ANNOVAR; SeqScape; SAS; Golden Helix VarSeq®

HONOURS, AWARDS AND ACHIEVEMENTS

CIHR International Symposium on Atherosclerosis Travel Award 2018
 Canadian Institutes of Health Research (CIHR)
 Institute of Circulatory and Respiratory Health (ICHR)
 Funding Amount: \$1000

Dean's Honor List 2012 – 2016
 Western University, London ON

Western Scholarship of Excellence 2011
 Western University, London ON

TRAINING AND CERTIFICATION

Teaching Assistant Training Program September 2017
 Western University, London ON

PUBLICATIONS

Ho, R. and Hegele, R.A. (2018). Complex effects of laminopathy mutations on nuclear structure and function. *Clinical Genetics* Submitted: CGE-00672-2018.

Ho, R., Wang, J. Cao, H., and Hegele, R.A. The identification of novel CNVs in patients with inherited lipodystrophy. In: *Atherosclerosis Supplements*. June 9–12, 2018; International Symposium on Atherosclerosis, Toronto, ON. Abstract C2-2.06

Ho, R., Wang, J. McIntyre, A.D., and Hegele, R.A. Connecting extreme ends of the spectrum: ultra-rare androgen receptor mutations in metabolic phenotypes. In: *Atherosclerosis Supplements*. June 9–12, 2018; International Symposium on Atherosclerosis, Toronto, ON. Abstract P2.043.

Ho, R., McIntyre, A.D., Kennedy, B.A., and Hegele, R.A. (2018). Whole-exome sequencing identifies a novel *IHH* insertion in an Ontario family with brachydactyly type A1. *SAGE Open Medical Case Reports* Submitted: SOMCR-18-0068.

Berberich A.J., **Ho, R.**, and Hegele, R.A. (2018). Whole-genome sequencing in the clinic: empowerment or too much information? *Canadian Medical Association Journal* 190.

Ho, R.*, Dron, J.S.* and Hegele, R.A. (2017). Recent Advances in the Genetics of Atherothrombotic Disease and Its Determinants. *Arteriosclerosis, Thrombosis, and Vascular Biology* 37(10), e158-e166.

*These authors contributed equally to this article.

POSTERS

Connecting Extreme Ends of the Spectrum: Ultra-rare Androgen Receptor Mutations in Metabolic Phenotypes (June 9–12, 2018) → Presented at the International Symposium on Atherosclerosis at the Toronto Metro Convention Centre, Toronto ON

Identification of a novel *IHH* insertion causative of brachydactyly type A1 (May 11, 2018) → Presented at Schulich School of Medicine and Dentistry Department of Medicine Resident Research Day at the Lamplighter, Best Western, London ON

Identification of a novel *IHH* insertion causative of brachydactyly type A1 (May 10, 2018) → Presented at London Health Research Day 2018 at the Convention Centre, London ON

Determining the genetic basis of rare diseases using whole-exome sequencing: Brachydactyly type A1 (November 16, 2017) → Presented at the Western University Health and Research Conference at Western University, London ON

Dual genetic diagnoses identified in a large family with brachydactyly type A1 and insulin resistance using whole-exome sequencing (October 20, 2017) → Presented at the American Society of Human Genetics 2017 Annual Meeting at the Orange County Convention Center, Orlando USA

Determining the genetic basis of rare diseases using whole-exome sequencing: Brachydactyly type A1 (June 20, 2017) → Presented at the Robarts Research Retreat at Kings University College, London ON

*Won the “Best Poster Award” at the conference

Ultra rare androgen receptor mutations in metabolic phenotypes (June 20, 2017) → Presented at the Robarts Research Retreat at Kings University College, London ON

Determining the genetic basis of rare diseases using whole-exome sequencing: Brachydactyly type A1 (April 23, 2017) → Presented at the Canadian Human and Statistical Genetics Meeting at Le Chateau Frontenac, Quebec City QC

Ultra rare androgen receptor mutations in metabolic phenotypes (March 28, 2017) → Presented at London Health Research Day 2017 at the Convention Centre, London ON

Determining the molecular basis of rare disorders using whole-exome sequencing (January 20, 2017) → Presented at the Harold B. Stewart Memorial Lecture at Western University, London ON

Absence of Protective *APOC3* Mutations in Dyslipidemic Patients (August 2014) → Presented at Western University’s Biochemistry Undergraduate Summer Research Program (BUSRP) Presentation Day

ORAL PRESENTATIONS

Novel CNVs in patients with inherited lipodystrophy (June 13, 2018) → Presented at the Department of Biochemistry Graduate Student Spring Symposium at Western University, London ON

The identification of novel CNVs in patients with inherited lipodystrophy (June 11, 2018) → Presented at the International Symposium on Atherosclerosis at the Toronto Metro Convention Centre, Toronto ON

The identification of novel CNVs in patients with inherited lipodystrophy (June 1, 2018) → Presented at the Robarts Research Retreat at King's University College, London ON

* Won a monetary award (\$100) and my name engraved on a plaque

Connecting extreme ends of the spectrum: obesity and lipodystrophy (January 10, 2018) → Presented at the Molecular Data Club at Robarts Research Institute, London ON

Genetic testing: finding the needle in a haystack (September 29, 2017) → Presented with Strong Brains, Strong Minds, Strong Muscles at the Kiwanis Senior's Centre, London ON

Determining the genetic basis of rare diseases using whole-exome sequencing: Brachydactyly type A1 (June 20, 2017) → Presented at the Robarts Research Retreat at King's University College, London ON

Determining the genetic basis of rare diseases using whole-exome sequencing (May 31, 2017) → Presented at the Molecular Data Club at Robarts Research Institute, London ON

Determining the genetic basis of rare diseases using whole-exome sequencing: Brachydactyly type A1 (April 22-25, 2017) → Selected to present at the Canadian Human and Statistical Genetics Meeting at Le Chateau Frontenac, Quebec City QC

Gene hunting: determining the molecular basis of lipodystrophy using whole-exome sequencing (March 21-22, 2015) → Presented at Ontario Biology Day at Carleton University, Ottawa ON

CONFERENCES/EDUCATIONAL EVENTS

American Society of Human Genetics 2017 Annual Conference (*presented*)

October 2017

Hosted by the American Society of Human Genetics at the Orange County Convention Centre, Orlando USA

Robarts Research Retreat (*presented*)

June 2017

Hosted by Robarts Research Institute at King's University College, London ON

Sixth Annual Canadian Human and Statistical Genetics Meeting (*presented*)

April 2017

Hosted by the Canadian Institute of Health Research, Le Chateau Frontenac, Quebec City QC

London Health Research Day (*presented*)

March 2017

Hosted by Schulich School of Medicine and Dentistry, at the Convention Centre, London ON

Harold B. Stewart Memorial Lecture (*presented*)

January 2017

Schulich School of Medicine and Dentistry, Western University, London ON

Genetics in Your Practice Education Day (*participated*)

November 2016

Hosted by the Medical Genetics Program of Southwestern Ontario, at the Sheraton, London ON

Ontario Biology Day (*presented*)

March 2015

Carleton University, Ottawa ON

Biology and Environmental Sciences Research Day (*presented*)

March 2015

Department of Biology, Western University, London ON

VOLUNTEERING

Schulich Graduate Student Council Representative July 2017 – *present*
LHSC University Hospital, London ON

- Representative for the Biochemistry Department in the council
- The council is responsible for meeting every month, informing every department of the changes happening within Schulich School of Medicine and Dentistry, and representing all departmental graduate students during voting and decision-making
- The Representative is responsible for relaying all information to the Department and responding to student inquiries

Crisis and Support Line (Counselling) September 2014 – *present*
ANOVA, formerly known as the Sexual Assault Centre London and Women's
Community House, London ON

- Volunteer on the Crisis and Support Line providing non-directive, non-judgmental support and information to anyone who has been directly or indirectly impacted by sexual violence or domestic violence

Big Bike Participant/Volunteer June 20, 2017
Western University, London ON

- Raised funds and participated in the annual Big Bike event, where all proceeds go to the Heart and Stroke Foundation
- Also assisted at the Waffle Breakfast event hosted by Robarts Research Institute to raise additional funds

Discovery Day 2018 Host May 4, 2018
Western University, London ON

- Helped run the Biochemistry Laboratory Workshop for high school students at Discovery Day at Western University

Medical Genetics Sponsored Studentship November 2017 – April 2018
Medical Genetics, LHSC Victoria Hospital, London ON

- Responsibilities include: assisting genetic counsellors with generating pedigrees, assessing lifetime risks of cancer using IBIS and BOADICEA, and prioritizing patients for appointments
- Opportunities also include shadowing the Genetic Counsellors during new clinic appointments and follow-ups
- Currently involved in updating the website content

Food Services: Menu Pick Up/Patient Visiting January 2017 – November 2017
LHSC University Hospital, London ON

- Duties included: picking up menus from all inpatient floors in the hospital, going room-to-room to visit patients and pick up outstanding menus, delivering cards, and helping patients complete their daily menus

Panel Interviewee June 19, 2017
London Central Secondary School, London ON

- Attended Grade 12 English classes to answer any questions about life beyond high school
- Sat on a panel with other volunteers and provided guidance and support

Big Bike Participant/Volunteer June 14, 2017
Western University, London ON

- Raised funds and participated in the annual Big Bike event, where all proceeds go to the Heart and Stroke Foundation
- Also assisted at the Waffle Breakfast event hosted by Robarts Research Institute to raise additional funds

Magic Wand Planning Committee January 2017 – June 2017
Ronald McDonald House of Southwestern Ontario, London ON

- On the organizing committee for the 2nd biennial Storybook Ball, a night of dinner and dancing and magical fun to raise money for the Ronald McDonald House
- Responsibilities included: obtaining sponsorships and donations, planning and setting up the event, and running the event

Discovery Day 2017 Host May 12, 2017
Western University, London ON

- Helped run the Biochemistry Laboratory Workshop for high school students at Discovery Day at Western University

President September 2016 – April 2017
Western Aikido Club, Western University, London ON

- As the Sport Western Aikido Club's President, I planned, organized and carried out daily presidential responsibilities such as financial accounting, email communication to other executive members, and weekly communication with the Western Rec Centre
- I also assisted in planning club events, and attended Rec Centre Presidential Meetings

Unit Support (Surgical Outpatients) April 2012 – January 2017
LHSC University Hospital, London ON

- As a unit support volunteer in surgical outpatients, I assisted with greeting and welcoming patients, visitors and staff
- I also provided assistance to surgeons, residents, nurses, and occupational therapists by putting together operation room kits, wound kits and breast surgery care kits

LAMP Presenter (Graduate Student Meet and Greet) November 2016
Western University, London ON

- For the LAMP meet and greet, I sat on the panel and answered questions from undergraduate students pertaining to research and graduate school

Registration Volunteer (Genetics in Your Practice Education Day) November 2016
Hosted by the Medical Genetics Program of Southwestern Ontario, at the Sheraton,
London ON

- For the Medical Genetics Education Day, I was in charge of registration for all health care professionals and attendees, and organizing event handouts and CME certificates

Vice President/Communications September 2015 – April 2016
Western Aikido Club, Western University, London ON

- As the Vice President, I was responsible for planning, organizing and carrying out daily club responsibilities, and maintaining club communications

Research Assistant September 2015 – April 2016
Dr. Hegele, Blackburn Cardiovascular Genetics Lab, Robarts Research Institute, London ON

- I helped run various scientific experiments, assisted in the writing of a provisional patent, and assisted the Medical Administrator with clerical duties

Child Mentor (Stand by Me Program) September – April 2013 and 2014
Thames Valley District School Board, London ON

- Acted as a caring adult mentor for at-risk children under the age of 12
- Committed two hours a week to my mentee for one-on-one meetings, and provided friendship, reinforcement, academic support and encouragement

Daffodil Sales April 2011, 2012, 2013
Canadian Cancer Society, London ON

- Involved in the Daffodil Month sales to raise money for the Canadian Cancer Society