CURRICULUM VITAE

Megan Lynch, Ph.D.
Program in Epidemiology and Human Genetics
University of Maryland School of Medicine
Email: megantherese314@gmail.com

Education

PhD University of Maryland School of Medicine December 2021
Human Genetics and Genomic Medicine

BS Lehigh University January 2015
Biomedical Engineering

Research Experience

Dissertation, University of Maryland School of Medicine, Baltimore MD 2016-Present

- Performed statistical analysis of disease-susceptibility genes in an Amish community and identified pathogenic variants in clinically actionable genes.
- Mapped epigenetic modification in familial hyperlipidemia and postprandial hyperlipidemia.
- Performed population admixture analysis for ancestry estimation.
- Created a novel method of mapping trait association with regional autozygosity.

Research Assistant, Lehigh University, Bethlehem Pennsylvania 2013 to 2015

- Investigated the 3D structure of protein binding pockets
- Measured steric and electric influences on the interaction between inhibitors and binding pockets using python, linux/unix, and C++

Teaching Service

Teaching Assistant, Computer Science Spring 2015

Lehigh University, Bethlehem PA

- Lead students through three projects in which they were asked to apply computational techniques to principles of structural biology, examining molecular structure, function, and evolution.
- 18 students, 8 hrs/wk

Publications

Journal Papers

Streeten, E.A., See, V., Jeng, L.B. J., Maloney, K.A., **Lynch, M.T.,** Glazer, A.M., Yang, T., Roden, D., Pollin, T.I., Daue, M., Ryan, K.A., Perry, J.A., O'Connell, J., Beitelshees, A., Palmer, K., Mitchell, B.D., Shuldiner, A.R. "KCNQ1 and Long QT Syndrome in 1/45 Amish: The road from identification to implementation of culturally appropriate precision medicine", Circulation Precision Medicine, 2020

Guo L., Torii S., Fernandez R., Braumann R.E., Fuller D.T., Hyun Paek K., Gadhoke N.V., Harris K., Mayhew C.M., Zarpak R., Jinnouchi H., Sakamoto A., Sato Y., Mori H., Kutyna M.D., Lee P.J., Weinstein L.M., Collado-Rivera C.J., Ali B.B., Atmakuri D.R., Dhingra R., Finn E.L.B., Bell M.W., Maloney K.A., **Lynch M.T.**, Cornelissen A., Kuntz S.H., Park J., Kutys R., Wang L., Hong S.N., Gupta A., Kolodgie F.D., Romero M.E., Jeng L.JB., Mitchell B.D., Surve D., Fowler D.R., Hong C.C., Virmani R., Finn A.V., "Genetics of Unexplained Sudden Cardiac Death in Adult Caucasian and African American Individuals", JAMA Cardiology, May 2021

Lynch M.T., Maloney K.A., Pollin T.I., Streeten E.A., Xu H., Shuldiner A.R., Van Hout C.V., Gonzaga-Jauregui C., Mitchell B.D., "The Burden of Pathogenic Variants in Clinically Actionable Genes in a Founder Population", AJMG, August 2021

Abstracts

Lynch M.T., Mitchell, B.D., Xu, H., "Abstract MP151: Post-prandial Lipemia Acutely Changes Epigenetic Regulation of Inflammatory Genes Over 4 Hr", American Heart Association: Arteriosclerosis, Thrombosis, and Vascular Biology, June 2020

Presentations and Invited Lectures

Poster Presentation, "High Frequency of Loss of Function Variants in ACMG Clinically Actionable Genes in the Amish", CHARGE, October 2018.

Poster Presentation, "Enrichment of Variants in Medically Actionable Genes in the Amish", ACMG, April 2019.

Poster Presentation, "Burden of Medically Actionable Variants in the Amish, a Founder Population," Amish Research Conference, June 2019.

Platform Presentation, "Quantifying the Burden of Pathogenic Variants in ACMG Clinically Actionable genes in the Amish and non-founder populations," Keystone Symposia, January 2020.

Short Talk "Kinship in a Founder Population Impacts Reproductive Health and Other Health Outcomes", University of Maryland Baltimore 3MT, November 2020

Platform Presentation "Impact of Genetic Relatedness of Parents on Reproductive Outcomes", ACMG, April 2021

Invited Lecture "Impact of Consanguinity on Fitness Related Traits in a Founder Population", Johns Hopkins University, April 2021

Professional Training

B. Braun Medical, Product Design Intern, 2013-2014

- Team member of engineering students who designed an intravenous fluid warming device for hospital use
- Created an accurate and consistent temperature read-out using an Arduino

Professional Affiliations

President, Graduate Student Association, UMD, May 2018-2020
Graduate Council Representative, Graduate Student Association, UMB, May 2017-2018
Chair, Graduate Assistant Advisory Committee, UMB, 2018-2020
Senator, University Student Government Association, UMB, 2017-2019

Professional Service

Research Conference Co-Organizer, University of Maryland Baltimore Graduate Research Conference, 2019, 2020

Community Service

Member, Junior League of Baltimore, Baltimore MD, 2021-Present
Child Care Assistant, House of Ruth Maryland, Baltimore MD, 2017-2019

Honors and Awards

Nutrigenomics, Nutrigenetics and Precision Nutrition Scholarship 2016
Donald and Dorothy Stabler Scholarship 2011-2014

Skills

- Strong linux/unix skills and programming ability in Python, R, SAS, Perl
- Advanced knowledge of statistical methodologies
- Development and optimization of pipelines for quality control, processing, and analyzing genetic data
- Integrating human genomic and disease phenotypic data to predict associations
- Proficiency in the use of bioinformatics/statistical software including BCFtools, Blast, and Plink

Abstract
Title of Dissertation: Identification of Actionable Variants, Reproductive Health and Other Health Outcomes in a Founder Population

Megan Lynch, Doctor of Philosophy, 2021

Dissertation directed by:

Braxton D. Mitchell, PhD, M.P.H
Professor
Epidemiology & Public Health
Vice Chair for Research, Division of Endocrinology, Diabetes & Nutrition, Dept. of Medicine

Genetically isolated populations that have arisen due to recent bottleneck events have reduced genetic variation that reflect the common set of founders. Many genetic variants that are rare in the general population drift to a high frequency in isolate populations, including some variants causing rare Mendelian diseases. The Old Order Amish of Lancaster County, PA, are a population isolate that are largely descendants of <200 of the original settlers. The unique genetic, cultural, and environmental homogeneity of the Amish is reinforced by the cultural proclivity to marry within the community. The overall theme of this thesis is to evaluate the impact of limited genetic diversity on reproductive outcomes and other traits of biomedical importance. This theme is addressed through three complementary sub-studies, each considering a different aspect of the impact of genetic homogeneity on the health of the Amish.

In the first sub-study, seven pathogenic/likely pathogenic (P/LP) variants were identified within genes deemed clinically actionable by the American College of Medical Genetics and Genomics (ACMG) or Geisinger's My Code Health Initiative. In total, 14.7% of Lancaster Amish individuals carry at least one of these variants, largely explained by the 13% who harbor a copy of a single variant in *APOB*. The Amish harbor

fewer actionable variants compared to similarly characterized non-founder populations but have a higher frequency of each variant identified. The second sub-study assessed the burden of parental relatedness on reproductive health. Amish couples who were both heterozygous carriers of a highly penetrant P/LP variant experienced fewer number of miscarriages than non-carrier couples from the same Amish community. In addition, overall genetic relatedness between spouses was highly correlated with number of live births ($p < 0.0001$), pregnancies ($p < 0.0001$), and stillbirths ($p = 0.03$). The third sub-study evaluated autozygosity (F_{ROH}), the portion of the genome that is homozygous by descent. Measurements of genome-wide and regional F_{ROH} were used as the primary predictors of trait variation in association analysis of 96 traits. No associations were identified when assessing genome-wide autozygosity measurements, but regional F_{ROH} estimation revealed two regions in which increased autozygosity was associated with increased trait values.

Identification of Actionable Variants, Reproductive Health and Other Health Outcomes in a Founder Population

by
Megan T. Lynch

Dissertation submitted to the faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Dedication

This dissertation is dedicated to my parents. Thank you for your endless support which gives me the confidence to follow my dreams.

Marie-Thérèse Lynch
Richard G. Lynch

Acknowledgments

Foremost, I would like to express my gratitude to my dissertation committee chair Dr. Braxton Mitchell. I am very fortunate to have had your guidance, insight, and advice. Thank you for your encouragement and support.

I want to give a special thank you to Kristin Maloney and Dr. Huichun Xu who were invaluable mentors and who gifted me with their perspective and direction.

I also wish to thank the other members of my dissertation committee, Drs. Toni Pollin, Elizabeth Streeten, and John Cole for their availability, expertise, and input into this dissertation.

Table of Contents

List of Tables

Chapter 3

Chapter 4

List of Figures

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations

P/LP	Pathogenic or likely pathogenic
VUS	Variant of Uncertain Significance
ES	Exome sequencing
ACMG	American College of Medical Genetics and Genomics
ROH	Run of homozygosity
GWAS	Genome wide association study
R	Coefficient of relationship
F	Coefficient of inbreeding
IBD	Identical by descent
LOF	Loss of function
Kb	Kilobase

I. Background

Founder population genetics

Isolated founder populations can offer insights into the genetic basis of disease susceptibility because they tend to have unique genetic backgrounds due to a bottleneck event.[1] Genetic diversity is reduced when a population remains closed following a recent bottleneck event in which a small group of individuals migrate from a larger population and carry only a subset of the gene pool with them.[2,3] In contrast to parent populations from which a founder group diverged, different spectra and frequencies of variants are seen because genetic drift has a profound impact on the gene pool.[4] Variants that are rare amongst the small number of founders of an isolated population, but rise to high frequency through random genetic drift, are called founder variants and some of them may have disease implications.[2] The lack of genetic variation observed in founder populations increases the impact of variants in potentially actionable genes and greatly increases the incidence rate of a small number of recessive conditions due to endogamy.[5]

Old Order Amish of Lancaster, PA

The >40,000 Amish currently residing in Lancaster County, Pennsylvania are a founder population that expanded from a bottleneck event when ~500 settlers immigrated to this area in the early 1700s.[6] They are one of several populations that belong to the more broad category of Anabaptists with similar history and lifestyle, including the Old Order Mennonites and Hutterites.[7] These Anabaptist communities have remained largely closed and maintain their cultural and genetic homogeneity.[4]

The Anabaptist movement began in 1525 when a group of people refused to conform to established religions and fled Switzerland to escape religious persecution.[8,9] This group was named Anabaptist Christians because they re-baptized their adult followers.[10] They believe that baptism should be delayed until the candidate freely confesses their faith as a consenting adult.[11] A major leader of this group was Menno Simons and his followers were named Mennonites.[10] The movement spread to Germany, France, the Netherlands, and to America in the 18th century.[12] By the time the Amish settled in a Pennsylvania colony, they were considered a more conservative group that had split from the Mennonites and the leader of this group was named Jacob Ammann.[13,14] There are no longer Amish communities in Europe partially due to persecution and other hardships that remaining Amish individuals faced.[14] There are three major Amish groups and the Old Order Amish is the largest and the most conservative.[14]

The Amish have self-governing church districts that are established based on geographical location.[9] The size of each district is typically about 20 to 40 families, determined by how many can fit in the farm dwelling or home.[9] Diversity in beliefs exists across locations and church districts and culture should not be generalized in a singular image.[10] Certain core values held by most Amish individuals include a strong emphasis on family and community, integrity, order, responsibility, obedience, and consideration of the human body as God's temple.[12] The Amish are generally not accessible to outsiders joining the sect and believe in conserving their traditions and powerful religious ideology.[9,15]

In the Lancaster Amish community, all individuals can be connected by a single 14-generation pedigree descending from 554 founders.[6] Analyses of the founders'

expected genetic contribution showed that only 128 founders account for over 95 % of the average extant genome.[6] Genetic drift has a great impact on the genetic architecture of the present generation because some founders have more descendants than others, increasing their contribution to the gene pool of the current population.

The increased frequency of certain genetic conditions, along with several other characteristics including the Amish community's receptiveness to medical studies, makes them an excellent community to make discoveries regarding genetic traits.[4] Initial research endeavors focused on autosomal recessive conditions and several of these recessive conditions were originally identified and described in the Amish.[16–19] The community also has several attributes conducive to research of complex disease and initial research efforts focused on diabetes and obesity.[15,20] The large family sizes, low relocation rates, and uniform socio-economic status and lifestyle of the Amish community support examining disorders with multifactorial phenotypes.[4,15] Another key advantage is the availability of essentially complete genealogies that date back to the early 1700s which have been compiled in a computer-searchable genealogy database for use by geneticists.[7,15]

Variation enriched in Amish: Recessive conditions

Because of the small number of founders, lack of new genetic influx, and random genetic drift, certain recessive variants that are rare in the general population are at a relatively higher frequency amongst Amish communities. While many rare genetic disorders may be absent in founder populations (e.g., phenylketonuria (PKU) and cystic fibrosis, are absent in the Lancaster County Amish), a small number of diseases are found

at much greater relative frequency. For example, the Lancaster Mennonite population has an increased frequency of Maple Syrup Urine Disease (MSUD)[21] and the Lancaster Amish have increased cases of glutaric aciduria type I (GAI).[22]

Observing the increased frequency of early onset disorders with full penetrance highlighted a need to ensure that care is provided for children with autosomal recessive conditions. The Clinic for Special Children (CSC), a medical facility, was founded to bring pediatric care to the Amish and Mennonite children with genetic diseases. Part of the CSC's mission is to strategically allocate pediatric resources to children and families needing the most support[5]. Throughout their history, the CSC has innovated creative ways to provide faster and cheaper genetic diagnoses and have recently developed a sequencing assay that enables them to screen individuals for all known pathogenic variants seen in the Amish and Mennonite populations in a single test, called the Plain Insight Panel (PIP).[5]

The PIP screening panel contains pathogenic variations that cause 162 genetic diseases affecting two large founder populations in southeastern Pennsylvania, the Lancaster Old Order Amish and the Old Order Mennonites.[5] The Mendelian recessive conditions on this panel are of varying severity, including some fully penetrant disorders with congenital or infantile onset. Couples testing positive in carrier screening (i.e., both identified as carrying a pathogenic /likely pathogenic variant in the same recessive gene) may choose to ask for newborn screening to enable earlier detection, delivery at a hospital with additional resources, or, in some cases, withholding intervention for a uniformly fatal disease. Early diagnosis of a genetic disease, and preparation by the family, can have a huge positive impact on health outcomes and palliative care.

Variation enriched in Amish: Complex phenotypes

The genetic homogeneity of the Amish population might allow for easier discernment of genetic effects on complex phenotypes including quantifying disease penetrance, identifying gene interactions, or increasing understanding of disease etiology. Amish individuals have been recruited to participate in population-based research studies to address genetic effects on complex phenotypes. Participants have been enrolled at the University of Maryland's Amish Research Clinic (ARC) in Lancaster County, PA over the past two decades and studies were open to all community members 18 years of age or older to assess a variety of diseases.[23–25] Enrolled participants filled out detailed medical and family history questionnaires underwent physical examinations and gave blood samples. Over the past 26 years, this data has been used in numerous studies including genome wide association studies (GWAS), identifying many associations between complex traits and genetic variation.

Examples of causal variants in which previous GWAS studies identified variant and phenotype associations include an association between a null *APOC3* variant, p.Arg19*, and decreased triglyceride levels and apparent cardioprotection[26], a highly significant association between an *APOB* variant, p. Arg3527Gln, and increased low-density lipoprotein cholesterol (LDL-C) levels[27], and longer QTc intervals/long QT syndrome caused by the *KCNQ1* variant p.Thr224Met[28]. Discovering these variants is important for clinical risk prediction. Within the Amish and other founder populations, high effect size variants that have been pinpointed through GWAS are typically applicable to a larger proportion of the population than in outbred populations.

Levels of LDL-C are highly variable with many genetic and environmental influencers and elevated levels increase cardiovascular disease risk.[29,30] A GWAS of Amish participants revealed the presence of a known *APOB* p.Arg3527Gln variant that is a major predictor of LDL-C levels and 13% of Amish individuals are carriers of this variant. Variants in the *APOB* gene cause familial hypercholesterolemia, a common autosomal dominant disorder.[31] It is known that *APOB* genetic variants have incomplete penetrance and carriers have a more mild phenotype compared to carriers of variants in other genes causing familial hyperlipidemia (*LDLR* and *PCSK9*).[32–34] Amish carriers of the p.Arg3527Gln variant did not exhibit other metabolic abnormalities, providing a unique opportunity to understand the impact of LDL-C elevation in the absence of other factors.[35]

Another large effect causal variant, *KCNQ1* p.Thr224Met, was found to be enriched in the Amish population (Amish MAF = 1.2%). This variant was previously identified in two individuals with long QT syndrome[36,37] but, due to insufficient clinical and functional evidence, was classified as a variant of unknown significance. Because this is a high frequency variant in the Amish population (carrier frequency of 1/45), further clinical phenotyping and in vitro functional studies were performed to better assess its possible health risk to the Amish community. Through this effort, the variant was reclassified as pathogenic.[28] The enrichment of this variant, the *APOB* p.Arg3527Gln, and the *APOC3* p.Arg19* variants have likely been maintained in the population by genetic drift. This offered an opportunity to further evaluate the disease mechanism in a relatively high frequency of individuals.[27]

The *APOB* and the *KCNQ1* gene are both on the consensus panel of 78 clinically actionable genes identified by ACMG and Geisinger's MyCode Health Initiative.[38–40] Their inclusion on this list indicates that P/LP variants in these genes should be reported when identified during clinical exome or genome sequencing because individuals may be able to work with their healthcare provider to manage risk through established interventions.[38] In population-based research exome sequencing, there is currently no legal obligation to return results. However, because of the high prevalence of the *KCNQ1* p.Thr224Met variant in the community, the relationship between the Lancaster Amish and the University of Maryland, and the geographic localization of this community, carriers were notified. They were given the option to obtain additional information including home visits, repeat EKG, and other relevant medical tests to ensure comprehensive treatment recommendations.[28] Almost all individuals who harbor one or two copies of the *APOB* p.Arg3527Gln variant have not been notified of carrier status; however, everyone enrolled in a study at the ARC is screened for LDL-C levels and notified of their results.

Actionable variants and databases

An important component of genetic research in Amish participants, in addition to health benefits to the local community, is the inclusion of deidentified findings in publicly available datasets for use by future researchers. For example, Amish genomes are included in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program and disease-causing variants have been reported in the ClinVar database which contains reports of interpretations of genomic variation and their

relationship to human health. When submitting a variant to the ClinVar database, researchers should thoroughly annotate the variants using the ACMG/AMP recommended guidelines to interpret results and use the databases five-tier system from pathogenic to benign.[41] Pathogenic variants almost certainly increase an individual's predisposition to a disorder and benign variants almost certainly do not effect disease.[42] The likely pathogenic and likely benign categories are reserved for variants where data supports a high likelihood that a variant is pathogenic or benign.[42] Various types of evidence are evaluated in the variant curation process, including population frequency and predictive assessment data, published case-control data, the presence of functional domains, and the pathogenicity of other variants at the same locus. Because many genetic findings in Amish participants are rare in the general population, including Amish findings in public databases could potentially alert future health care providers of the pathogenicity of previously unknown finding.[41]

Kinship: Reproductive outcomes

Many pregnancy losses are due to known chromosomal abnormalities or variations at specific loci, but a large portion have an unknown etiology.[43]An estimated 70% of all pregnancies in humans, including those that have not yet been recognized, are lost prior to live birth.[44] In order to assess possible risk factors of pregnancy loss, previous studies assessed populations with consanguineous unions and measured the elevated risk that highly related parents may have. To do so, they used several measurements to quantify the degree of relatedness between parents. Kinship, the proportion of genes that two individuals have in common, is measured using the

coefficient of relationship (R), which is twice the coefficient of inbreeding (F). These measures of relatedness can be estimated by pedigrees or by using genomic markers to calculate the identical-by-descent (IBD) status of two alleles in an individual.

The Pakistan Risk of Myocardial Infarction Study (PROMIS) recruited Pakistani participants, in which the median coefficient of inbreeding was fourfold higher than outbred populations, and sought to quantify the increased burden of LOF variation associated with highly related unions.[45] Among study participants, twice as many variants were identified that are embryonically lethal recessive than those that result in fetal or infant death. A typical individual from this population was estimated to carry 1.6 recessive LOF lethal-equivalent variants in the heterozygous state, 1.1 of these are embryonically lethal and the remaining 0.5 result in miscarriage, stillbirth, or infant mortality.[45]

As quantified by the PROMIS study, unions between individuals with elevated kinship, which are more common in founder populations, increase the probability of a zygote receiving two copies of a disease-causing recessive allele with potentially harmful impacts on fertility. Overall genetic homogeneity between parents is measured by the coefficient of relationship (R), which is the proportion of genes that two individuals have in common and is twice the coefficient of inbreeding (F). F coefficient estimates in randomly mating outbred populations are close to zero (< .005)[46] but founder populations have much higher F values.[47,48]

Conflicting reports have been published regarding the relationship between the R coefficient and adverse reproductive outcomes. Many of these studies have been conducted in founder populations because of their elevated kinship. In a broad population

of Anabaptists, including individuals from Lancaster Amish communities, a strong positive association was identified between consanguinity and family size and shorter interbirth intervals.[7] In contrast, a study conducted in a Hutterite population identified a reduction of fecundity for highly related couples which was shown by longer inter-birth intervals and longer intervals to a recognized pregnancy. However, they did not see increased rates of fetal loss, and family sizes did not differ depending on the inbreeding coefficient of Hutterite women.[49]

The reproductive risks associated with consanguineous relationships have also been studied in populations that are not comprised of Plain people. These studies utilized pedigree data or genetic kinship values to measure their influence on reproductive outcomes. Through these studies, a positive correlation was detected between kinship and number of surviving children[50] and total number of pregnancies.[51] A study of Icelandic couples measured kinship values using pedigrees going back 10 generations and found greater fertility and reproductive success with increasing kinship between spouses with a notable deficit in the reproductive success of couples related at the level of second cousins or closer.[52]

Other studies report no relationship between consanguinity and fertility measured by recurrent miscarriage and fetal loss.[53,54] In a sample population of >200,000 women from various Western countries, an association between fetal loss and kinship values was diminished when confounders including illiteracy, age at marriage, and contraceptive uptake were included in the multivariate analysis.[53] Qatari women in first cousin marriages showed no difference in the rates of previous pregnancy loss and retrospective investigation showed no difference in the rate of maternal disorders, median gestational

age, or fetal weight compared to non-consanguineous women and no familial clustering of recurrent miscarriage was seen in either group.[54]

High kinship has also been identified as a risk factor for stillbirth and miscarriage.[55,56] In an Australian population in which consanguinity was self-reported by the mother as a bivariate metric, the consanguineous group had significantly higher number of stillbirths. Additional risks including fetal anomalies and perinatal mortality were also significantly more common in the consanguineous group included in this study.[55] An increased rate of miscarriage and stillbirths was also seen in consanguineous couples using data from a survey conducted in the Palestinian Territories. Using incidence risk ratios, this study tested consanguinity among other lifestyle factors and identified consanguinity as the strongest risk factor for stillbirth and miscarriage.[56]

Kinship: Inbreeding depression

Inbreeding depression is the reduced fitness seen in offspring with closely related parents. Homozygosity, which is increased in founder populations, can have harmful effects when deleterious recessive alleles combine in continuous segments of homozygous alleles (regions of homozygosity, or ROHs), directly contributing to phenotypic variance of complex traits. Evidence suggests that inbreeding depression is broad in scope and associated with not only reproductive traits, but additional complex traits closely linked to evolutionary fitness.[57] Although consanguinity is most commonly associated with inborn errors of metabolism, which are usually recessive conditions, there is also an increased incidence or earlier onset of complex disorders including diabetes[58], cardiovascular disorders[59], waist to hip ratio[60], and certain types of cancers[61,62].

Kinship: Autozygosity

Autozygosity measurements are an alternative method to quantify the risks associated with consanguineous relationships using a trait of each individual rather than a trait of a spouse pair. Regions of homozygosity arise when IBD haplotypes are inherited from each parent.[57] The amount of ROH present in an individual's genome, F_{ROH}, is correlated with the number of homozygous loss of function (LOF) variants present in an individual.[46] The net directional effect of all homozygous variants on phenotypes can be studied using genome-wide autozygosity estimates, revealing the full deleterious effects of recessive or partially recessive alleles.[63,64] Individuals with increased autozygosity are more likely to exhibit deficits in traits associated with fitness.[63] Autozygosity mapping can identify segments of the genome that harbor low-frequency and rare genetic variants with a recessive effect on traits.[65,66]

Previous studies have shown conflicting results regarding a correlation between individual autozygosity levels and reproductive traits. A study conducted on a population isolate living in a Swiss village estimated both parental relatedness and coefficients of inbreeding for each parent using genealogical data. Although parental relatedness had no effect on the number of children that a couple had, mothers with higher autozygosity were found to have significantly fewer children with inbred fathers being unaffected.[67] In contrast, other studies have found that the inbreeding level of the father negatively impacts family size.[68,69] For example, inbreeding depression, first speculated by Charles Darwin, had adverse effects on male fertility in subjects of his own lineage. The

inbreeding coefficient of the father was significantly associated with decreased family size.[69]

Additional associations have been identified between autozygosity and complex traits. One study found that ROH was significantly associated with apparently deleterious changes in 32 out of 100 traits analyzed.[47] An earlier study identified a statistically significant negative association between F_{ROH} and height, forced expiratory lung volume in one second, general cognitive ability and educational attainment.[64] These negative associations were confirmed by a second study that also identified a negative association between autozygosity and income and between autozygosity and grip strength. This second study also observed significant positive associations between F_{ROH} and age at first sexual intercourse and religious group participation.[63] Confounding variables may influence these findings including assortative mating and geographic mobility which are both generally seen for cognitive ability and socioeconomic status.[63,64] However, body mass index is also generally associated with assortative mating, but no effect was seen in these studies.[64]

The study of F_{ROH} in a founder population offers the benefit of using a smaller sample of individuals for the same predictive power as studies in non-founder populations with large sample sizes.[65,70] In fact, cohorts with high levels of homozygosity provide up to 100 times greater per-sample statistical power.[47] This is because of the parental relatedness in founder populations. They often have both a greater quantity of ROH and longer ROH segments.[57] A measure called F_{IS}, calculated as the mean individual departure from Hardy-Weinberg equilibrium, is the level of inbreeding of the most recent generation reflective of instances of non-random mating. Populations with

small F_{IS} values, including the Amish and Hutterite populations, tend to promote marriages between individuals who are a certain degree of relation or less to avoid unions within closely related individuals.[57] In these populations, autozygosity is primarily caused by small effective population size rather than preferential consanguineous unions.

Kinship: Autozygosity Mapping

Although most LOF variants do not occur in an autozygous state, nearly all rare homozygous predicted LOF variants are found within autozygous segments and the mean number of rare homozygous LOF variants per individual is proportional to the rate of autozygosity.[45] Therefore, deleterious changes in traits could be associated with ROH and not with common variant homozygosity, and hotspots along the genome that contain ROH may drive associations seen between autozygosity and complex trait outcomes.

The initial utilization of autozygosity mapping was for recessive disorders to localize potential disease-causing haplotypes shared by both parents.[71] More recently, the approach has been applied to complex disease and other multifactorial traits in case-control studies.[72,73] These efforts have helped to pinpoint regions of the genome in which multiple recessive variants contribute to complex traits. For example, autozygosity mapping methods have been used to localize broad signals from GWAS by pinpointing regions of autozygosity that are associated with schizophrenia and height.[65,66]

II. Specific Aims

The Amish are a population isolate with unique genetic, cultural, and environmental homogeneity. Genetic diversity of the population was reduced after a bottleneck event in which ~500 founders immigrated from Europe to Eastern Pennsylvania. Because of this bottleneck, many genetic variants that are rare in the general population have drifted to a high frequency in the Amish, including some associated with rare Mendelian diseases. Founder populations that remain largely closed communities, as the Amish have, are classified by increased levels of consanguineous marriages since all members derive from a small number of founding individuals. The overall theme of this thesis is to address the impact that genetic homogeneity may have on disease state and other traits of biomedical importance including reproductive outcomes. Through three complementary sub-studies, each considering a different aspect of the reduced genetic diversity of the Amish, I evaluate the impact that their unique genetic architecture may have had on disease susceptibility and trait outcomes.

The first sub-study addressed the burden of pathogenic and likely pathogenic (P/LP) variants in clinically actionable genes. These genes were identified by the American College of Medical Genetics and Genomics (ACMG) who published recommendations for return of secondary findings for clinically significant variants in 59 medically actionable genes when they are identified during clinical exome or genome sequencing.[38] The Geisinger Medical Center has published a similar panel of 76 medically actionable genes.[39] Using a "consensus panel" of 78 genes from both ACMG and Geisinger, disease-causing variants were annotated in accordance with ACMG and the Association for Molecular Pathology's (AMP) guidelines for variant interpretation. We showed that,

although the Amish population harbored fewer unique actionable variants compared to similarly characterized non-founder populations, they had a higher frequency of individuals harboring each variant identified.

The second sub-study measured the impact of known, P/LP recessive variants on several reproductive outcomes. Many of these variants cause early onset recessive disorders, including some that are fully penetrant with palliative diagnoses.[5] In addition, we evaluated whether overall genetic relatedness of parents across the genome was associated with reproductive health outcomes. Associations between the overall genetic relatedness of spouse pairs and reproductive outcomes have previously been identified in Anabaptist populations. For example, a strong negative association between inbreeding and family size was identified in Hutterite populations.[49] In Amish and Mennonite populations the opposite effect was seen and highly related couples were reported to have increased family size and inter-birth intervals.[7] The second sub-study confirmed the finding that highly related couples have increased family size, expands this result to include additional reproductive outcomes, and evaluates whether other genetic markers were also associated with reproductive outcomes.

Autozygosity measurements were used in the third sub-study to quantify the risks associated with consanguineous relationships using a trait specific to the individual rather than relatedness between couple pairs. Inbreeding depression occurs when offspring of closely relative individuals have higher rates of congenital disorders and lower fitness.[74] It is associated with complex traits closely linked to evolutionary fitness. Because inbreeding depression is not only associated with reproductive traits, but additional complex traits closely linked to evolutionary fitness[47], we expanded the phenotypic

measures used in this analysis. To evaluate the contribution of autozygosity to complex diseases and traits, we estimated both genome-wide levels of autozygosity and the probability of autozygosity at 10 Kilobase (Kb) average intervals throughout the genome. These measurements were used in association analysis and trait mapping to assess evidence for association of genome-wide and locus-specific association of autozygosity with 96 different phenotypes.

Specific Aim 1:

Compare the burden of pathogenic and likely pathogenic actionable variants between the Lancaster Amish population and European ancestry non-founder populations.

a) Annotate variants in accordance with ACMG/AMP guidelines
b) Compare allele frequencies of pathogenic and likely pathogenic variants between Amish and the following non-founder, mostly European populations: Geisinger and The National Heart, Blood and Lung Initiative (NHLBI) Exome Sequencing Project (ESP), and the UK biobank.

Hypothesis 1:

By virtue of their reduced allelic diversity, the Amish population harbors a lower number of unique pathogenic or likely pathogenic variants compared to non-founder populations, although the Amish have a higher carrier frequency for those variants that are found.

Specific Aim 2:

Evaluate the association of reproductive outcomes with (a) carrier status of parents for known pathogenic variants in the same gene and (b) overall genetic relatedness between the parents.

a) Quantify the potential impact of carrier testing and assess the association of reproductive outcomes with carrier status of parents for the known highly penetrant recessive variants.
b) Assess the association of reproductive outcomes with genetic relatedness between the parents.

Hypothesis 2:

The Lancaster Amish population is enriched for homozygous variants in recessive disorders covered by the Plain Insight Panel (PIP). We will estimate the frequency of homozygosity and compound heterozygosity for variants that cause these disorders. We hypothesize that carrier status of parents for PIP variants in the same gene and overall genetic relatedness between the parents is associated with adverse reproductive outcomes.

Specific Aim 3:

Evaluate the association of biomedically important trait outcomes with global level of autozygosity of individuals across the genome and utilize homozygosity mapping to identify the specific chromosomal regions that may drive the association with trait outcomes.

a) Assess the association between autozygosity of each individual and biomedically important trait outcomes.
b) Utilize homozygosity mapping to uncover regions of the genome that are associated with trait outcomes.

Hypothesis 3:

Genome wide levels of autozygosity in individuals will be associated with reproductive outcomes and other trait outcomes of biomedical importance. Performing an ROH mapping analysis across the genome will detect certain genomic locations in which autozygous regions are responsible for this influence on reproductive and other health and disease-related outcomes. Certain loci in the homozygous state may be less tolerated than other locations and influence complex phenotypic traits that make up an individual's fitness.

III. Significance and Rationale

Specific Aim 1:

The goal of this first aim was to uncover genetic findings of potential clinical importance (i.e., "actionable"), including novel variation and quantify the impact of these variants within this population in terms of number of unique variants segregating in the population and the number of carriers of each since through genetic drift, even a small number of variants could have many carriers each. Annotating P/LP variants and identifying individuals who are heterozygous carriers is important because knowledge that an individual carries one of these variants enables individuals, even when asymptomatic, to seek risk management through established interventions which reduce morbidity and mortality. This can also benefit participants' families by informing genetic counseling for family screening. Quantifying the burden of P/LP variants segregating in a population can aid in this effort. Although there is currently no obligation to report clinical findings from research studies[75], this is increasingly being encouraged.[39,76,77]

In the Amish and other founder, identifying P/LP variants can ensure that population specific genetic screens and return of results can be efficiently managed. Additionally, the enrichment of variants in founder populations provides unique opportunities to uncover novel disease-causing variants that are either at low frequency or absent from outbred populations and can be added to public databases. Identification of already known pathogenic variants enriched in a founder population can also increase understanding of the disease mechanism.

Specific Aim 2:

The variants of interest for the second sub-study were previously curated highly penetrant variants leading to recessive and x-linked diseases. All of these variants are known to have carriers in the Amish or Mennonite communities of Lancaster, PA.[5] Retrospectively identifying couples who are both heterozygous for a recessive, highly penetrant, disease-causing variant and quantifying how many such couples are present in the population may help with planning for future genetic testing. Analysis of reproductive outcomes associated with carrier couples may help to uncover risks previously not associated with some monogenic diseases and allow healthcare providers to give more complete risk assessment information to such couples. Couples with a high degree of relatedness have an increased probability of producing a zygote that has two copies of a pathogenic recessive allele, which may negatively impact reproductive outcomes.4 Looking more broadly at genetic relatedness between couples may uncover a risk for pregnancy loss associated with highly related couples, indicating that being a carrier for a recessive condition may simply be a marker for higher relatedness across the entire genome.[78] Looking more broadly at genetic relatedness between couples may uncover a risk for pregnancy loss associated with highly related couples, indicating that being a carrier for a recessive condition may simply be a marker for higher relatedness across the entire genome.

Specific Aim 3:

The net directional effect of all recessive variants on phenotypes can be studied using a genome-wide autozygosity calculation.[46] The final aim assessed the effect of consanguinity on complex traits in the Amish population using an individual measure of autozygosity rather than as a trait of couple pairs. In well-studied recessive disorders, homozygous state has clear disease implications. An additional way to evaluate homozygous state and its impact on disease, may be to look at autozygosity, measured by runs of homozygosity (ROHs). Homozygosity directly contributes to phenotype variance of some complex traits[61] and the net directional effect of all recessive variants on phenotypes can be studied using autozygosity calculations. Autozygosity mapping can identify segments of the genome that harbor low-frequency and rare genetic variants with a recessive effect on traits.[65,66] This analysis in a founder population could help to identify instances where ROHs are associated with deleterious complex traits.

Since founder populations have closer and more recent inbreeding and increased lengths and number of ROHs, they are an ideal population for this study.[57,70] Finally, this study has broad applications because 10.4% of people worldwide live in populations where consanguineous marriages are common.[79]

IV. The burden of pathogenic variants in clinically actionable genes in a founder population[1]*

Abstract

Founder populations may be enriched with certain genetic variants of high clinical impact compared to non-founder populations due to bottleneck events and genetic drift. Using exome sequencing (ES), we quantified the load of pathogenic variants that may be clinically actionable in 6,136 apparently healthy adults living in the Old Order Amish settlement of Lancaster County, PA. We focused on variants in 78 genes deemed clinically actionable by the American College of Medical Genetics and Genomics (ACMG) or Geisinger's MyCode Health Initiative. ES revealed 3191 total variants in these genes including 480 nonsynonymous variants. After quality control and filtering, we applied the ACMG/AMP guidelines for variant interpretation and classified seven variants, in seven genes, as either pathogenic or likely pathogenic. Through genetic drift, all seven variants are highly enriched in the Amish compared to non-founder European populations. In total, 14.7% of Lancaster Amish individuals carried at least one of these variants, largely explained by the 13% who harbored a copy of a single variant in *APOB*. Other studies report combined frequencies of pathogenic/likely pathogenic (P/LP) variants in actionable genes between 2.0% and 6.2% in non-founder populations.[39,80,81] The Amish population harbors fewer actionable variants compared to similarly characterized non-founder populations but have a higher frequency of each variant identified, offering opportunities for efficient and cost-effective targeted precision medicine.

[1] Lynch, M. T., Maloney, K. A., Pollin, T. I., Streeten, E. A., Xu, H., Regeneron Genetics Center, Shuldiner, A. R., Van Hout, C. V., Gonzaga-Jauregui, C., & Mitchell, B. D. (2021). The burden of pathogenic variants in clinically actionable genes in a founder population. American journal of medical genetics. Part A, 185(11), 3476–3484. https://doi.org/10.1002/ajmg.a.62472

*MTL analyzed data, generated tables and figures, and drafted the manuscript.

Introduction

The increasing use of exome sequencing in research studies has prompted much discussion about the ethical responsibility of research institutions to return clinically actionable genetic results to study participants. The American College of Medical Genetics and Genomics (ACMG) has published recommendations for return of secondary findings involving pathogenic (P) and likely pathogenic (LP) variants in medically actionable genes when identified during clinical exome or genome sequencing.[38] This list of clinically actionable genes recommended by ACMG includes 59 genes that are associated with highly penetrant medical conditions in which individuals, even when asymptomatic, may be able to manage risk through established interventions. The MyCode Health Initiative from Geisinger, a precision medicine project, developed a similar list expanding on the ACMG recommendations. This list contains 76 genes that account for 27 highly penetrant and monogenic medical conditions.[39] The union of the two lists contains 78 genes. A recent publication recommended that additional genes are added to the list of clinically actionable genes (ACMG SF v3.0).[82] The v3.0 list, containing 73 genes, was published after the conclusion of the current study.

Since the publication of the ACMG return of results guidelines and list of medically actionable genes, efforts have been made to estimate the number of individuals in the general population harboring pathogenic variants in at least one of these clinically actionable genes. Quantifying the prevalence of seemingly healthy individuals harboring pathogenic variants in these genes in the general population is important because of the possibility for medical interventions and for planning for a future in which genome

sequencing is more widespread. Population level secondary finding rates may also help to support informed decision making regarding the rationale for or against the return of results from research studies.

The ACMG Working Group on Incidental Findings in Clinical Exome and Genome Sequencing published an initial estimation that ~1% of individuals carry at least one incidental clinically actionable variant.[41] Shortly thereafter, an estimate of an approximately 1% secondary finding yield was made using Phase 1 of the 1000 Genomes dataset.[83]

Within a sample of 50,726 predominantly European ancestry participants from the Geisinger MyCode Health Initiative who underwent exome sequencing (ES) through the Geisinger-Regeneron DiscovEHR collaboration, it was found that 3.5% of individuals carried a pathogenic or likely pathogenic variant within their medically actionable gene panel.[39] The National Heart, Blood and Lung Initiative (NHLBI) Exome Sequencing Project (ESP) estimated the prevalence of secondary findings in a study of 6503 (4300 European and 2203 African ancestry) and found 1.7% of individuals with European descent and 1.2% of individuals with African descent had findings in 68 clinically actionable genes[81]. In a third study, utilizing 2,091 physician referred clinical cases (of which 17.2% were not trios), 6.2% (n=129) of individuals had reportable secondary findings in the ACMG medically actionable gene panel.[84] More recently, analyses of the ES data for the initial 49,960 sequenced individuals from the UK Biobank identified medically actionable variants in the ACMG recommended genes in 2% of the participants[80] (Table 3.1, Figure 3.1).

	Amish (n=6136)	Geisinger (n=1,415)	ESP (n=6503)	UK Biobank (n=49,960)
Clinically Actionable Variants	**7**	**43**	**67**	**548**
Genes in Panel	**78**	**76**	**68**	**59**
Individuals with Reportable results (%)	**900 (14.7%)**	**49 (3.5%)**	**99 (1.5%)**	**992 (2.0%)**

Table 3.1. Frequency of P/LP variants identified in Amish and reported in non-founder populations. Geisinger includes individuals sequenced as part of Geisinger's MyCode Community Health Initiative[39], NHLBI Exome Sequencing Project (ESP) includes individuals from six NHLBI cohorts in the HeartGO Consortium[81], UK Biobank data includes UKB participants with whole-exome sequencing generated by the Regeneron Genetics Center, as part of a collaboration with GlaxoSmithKline[80].

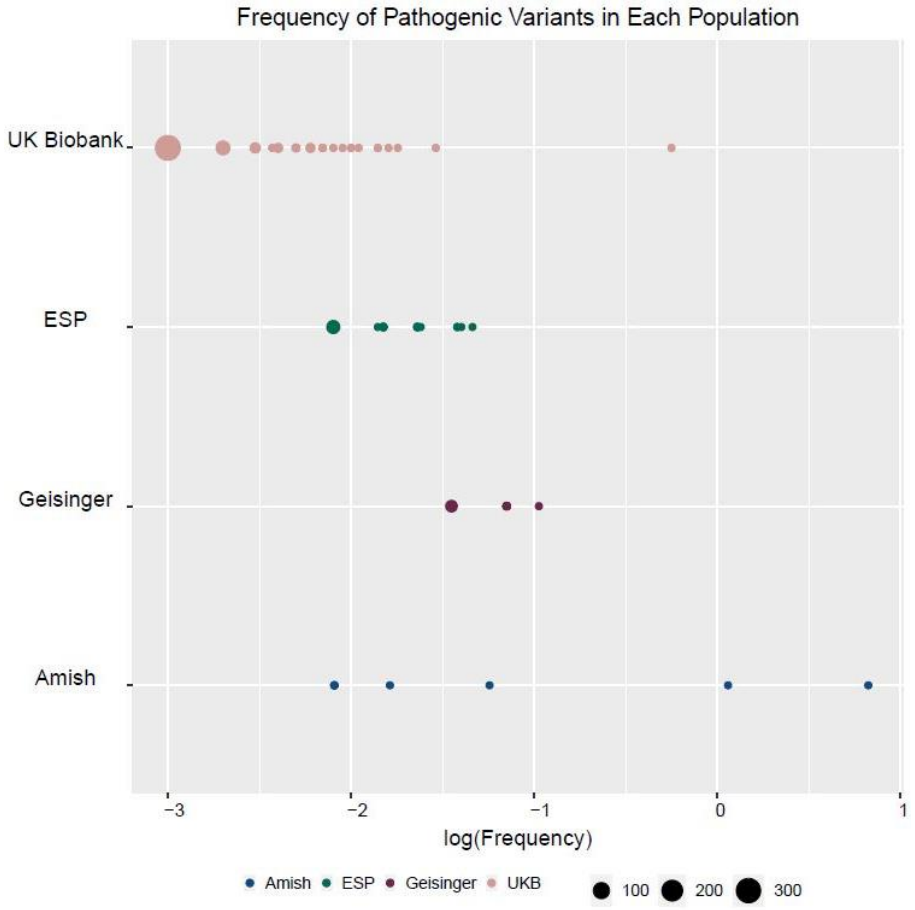


Figure 3.1 Burden of disease-causing variants in the Amish versus two outbred populations. Size of dot indicates the number of variants at the allele frequency.

In contrast to outbred populations, isolated founder populations may have different spectra and frequencies of clinically significant variants because of their population demographic history.[35] When a population remains closed following a recent bottleneck event, genetic diversity is reduced and genetic drift may have a profound impact on the allele frequency spectrum, causing inflation of the frequency of certain variants and reduction in the frequency of others.[1] Rare alleles present in the founding group of the population that rise to high frequency through genetic drift are called founder variants and some of them may have disease implications.[2] Autosomal recessive conditions, which are generally individually very rare in cosmopolitan populations, occur at increased frequency due to endogamy. The spectrum of these disorders in the Amish have been reported elsewhere.[29,66–69]

The >40,000 Amish currently residing in Lancaster County, Pennsylvania are a founder population that expanded from a bottleneck event when ~500 settlers immigrated to this area in the early 1700s.[6] Analyzing a sample of 6,136 apparently healthy individuals from this population, our goal was to estimate the frequency of pathogenic/likely pathogenic (P/LP) variants in 78 clinically actionable genes. In this report, we quantify the impact of these variants in the Amish population in terms of number of variants and number of individuals harboring a pathogenic variant. We hypothesize that through genetic drift, a small number of medically relevant variants could each have many heterozygotes in Amish.

Methods

Study population:

As part of the collaboration between the Amish Research Clinic and the Regeneron Genetics Center (RGC), we performed exome sequencing (ES) in 6,136 Amish individuals from Lancaster County, PA over 18 years of age, representing approximately 18% of the adult population in the Lancaster Amish settlement. All study participants were recruited through prior community-wide studies by our group over the past 25 years.[89] Studies were open to all community members to assess a variety of diseases. No community members were recruited based on a phenotype or family history associated with the 78 genes assessed here.

Exome Sequencing (ES) and Genome Informatics:

ES was conducted by the RGC, using a high-throughput automated pipeline developed in-house utilizing a slightly modified version of the xGen (Integrated DNA Technologies) capture reagent. Sequencing was performed on the Illumina NovaSeq platform using 75 bp paired-end reads. Captured fragments were sequenced to achieve a minimum of 85% of the target bases covered at 20x or greater coverage. A cloud-based pipeline that uses standard bioinformatics tools was used for sample-level data production and analysis. Raw sequence data was uploaded to DNAnexus for automated production including sample de-multiplexing using Illumina's CASAVA software. Sequence reads were mapped and aligned to the GRCh38/hg38 human genome reference assembly using BWA-mem. SNP and INDEL variants and genotypes were called using GATK's HaplotypeCaller. QC metrics were applied as previously described.[90] Project-

level and sample-level VCFs for downstream analyses were generated as the pipeline output.

Selection of 'clinically actionable' genes:

We analyzed sequence variation in the recommended 59 ACMG clinically actionable genes[38,41] plus an additional 17 genes considered medically actionable to be returned to participants in the MyCode Health Initiative from Geisinger. [39] Pathogenic and likely pathogenic variants in the 78 genes on this "Consensus Panel" are associated with 29 mendelian conditions with moderate to high penetrance.

Filtering Pipeline:

We first identified all genetic variants within the 78 genes on the Consensus Panel by querying the ES data for all Amish participants. Of these, we focused on splice site, stop gain/loss, indels and missense variants. Initial quality control parameters included phred-scaled genotype likelihood (> 10^{10}), allelic balance (0.2-0.8), coverage depth (>10), and pored quality score (>30). For the purposes of identifying potentially actionable variants, we increased the stringency of QC parameters to avoid an abundance of novel nonsynonymous SNPs appearing as apparent de novo variants. Following stringent filtering on VCF files, validity of sequencing was evaluated by IGV confirmation including allelic balance, presence of positive and negative strands, or the variants in the center of a strand rather than only on the end of read. We checked pedigrees for Mendel errors by verifying that a variant in an individual could reasonably have been received from either of their parents.

We filtered out all singleton variants except those that had a previous ClinVar annotation. All variants having at least two heterozygotes in our Amish population cohort and a minor allele frequency <0.001 in five non-founder gnomAD populations, or singleton variants with a previous P/LP ClinVar annotation, were subject to manual curation.

Manual Curation:

All identified variants were reviewed in accordance with the ACMG/AMP guidelines for variant interpretation, including a comprehensive assessment of population frequency, predicted variant effect, functional evidence, statistical support for association, functional domains, as well as the presence of other variants at the same position.[41] Published gene-specific ACMG/AMP rules from ClinGen Variant Curation Expert Panels (VCEPs) were employed for appropriate genes and were also used as guides for genes in similar clinical domains.[91–95] We referenced the PVS1 flowchart which contains further guidance published by the ClinGen SVI which recommends the application of relative strengths for different types of potentially null mutations.[96] Additional ACMG/AMP evidence codes were used with modified strength when appropriate, including PM2, PM3, and PS3, to distinguish moderate and supporting level evidence in favor pathogenicity.[97] Assignment of pathogenicity was made according to the ClinVar 5-tier classification system (pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, benign).

Sanger Sequencing:

Guidelines issued by the Centers for Disease Control (CDC) state that any genetic results to be used for clinical purposes must be CLIA-certified.[98] We considered this to be especially important for our study because the exome data used was initially processed using quality control filters designed for research and discovery purposes where some false positives are tolerated. To address the concern for clinical use of false positive results, we performed more stringent filtering to increase the specificity of identifying true pathogenic variants and ensured that no Mendel errors were present in pedigrees of all P/LP identified variants. Sanger sequencing of any variant showing poor evidence for Mendelian segregation and all variants identified as P/LP was performed by the University of Maryland School of Medicine Center for Innovative Biomedical Resources, Genomics Core.

Results

Amish Demographics:

The mean age (± SD) of the 6,136 Amish participants included in this study was 49.8 ± 17 years and 43% were male. Nearly all members of this community can be connected into a single 14-generation pedigree descending from 554 individuals.[6]

Identification of P/LP variants: Pre-manual curation:

The 78 genes in our panel account for five disease categories: inborn errors of metabolism, cardiac, cancer/tumor, connective tissue disorder, and malignant hyperthermia susceptibility. Within these genes, we identified 3,191 variants, 480 of which are nonsynonymous. Following quality control parameters, our pipeline identified 40 variants that were either classified previously in ClinVar as P/LP at least once (n=28) or were novel but likely disruptive variants based on their classification as splicing, nonsense, and indels (n=12) (Figure 3.2). These 40 variants, herein referred to as "potentially pathogenic variants" [41], are dispersed among four disease categories (Figure 3.3). An additional 84 missense variants were also identified through our filtering pipeline and curated to determine pathogenicity.

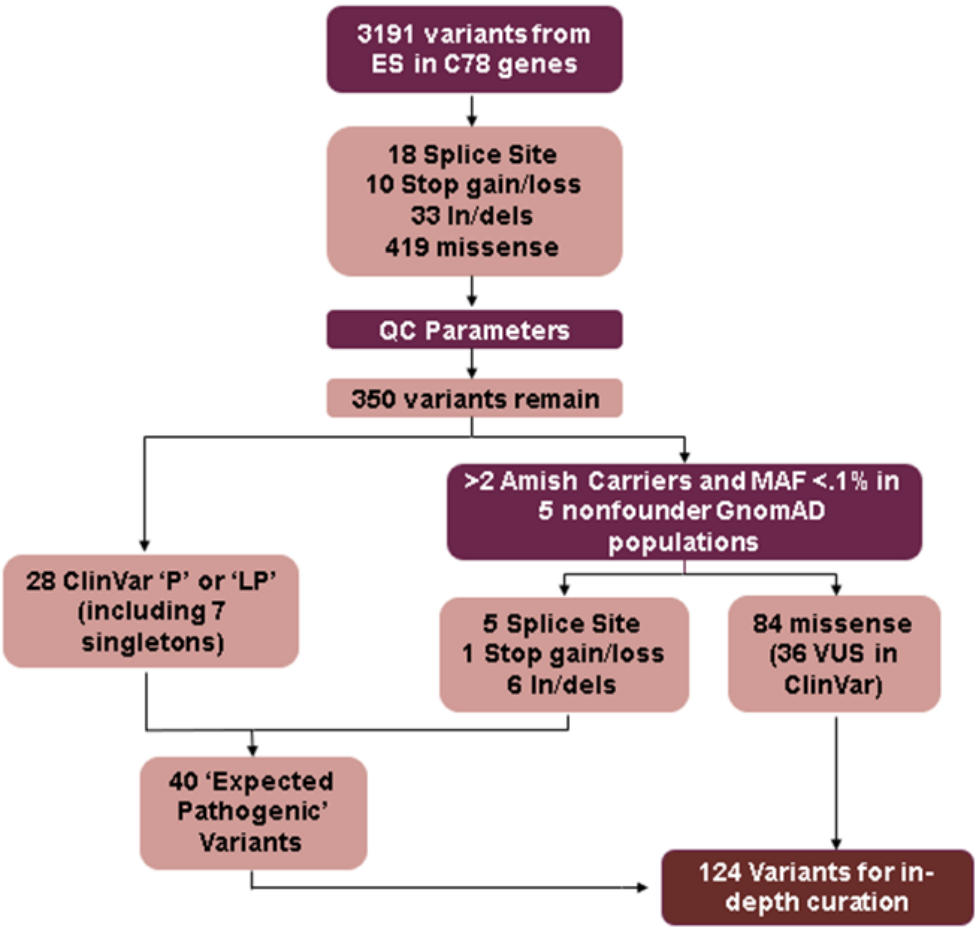


Figure 3.2. Variant filtering pipeline including quality control. We identified 3191 variants within the Consensus 78 genes. We chose to focus on 480 nonsynonymous variants. After applying QC parameters, variants previously classified in ClinVar as Pathogenic or Likely Pathogenic were added to the analysis. If variants were not seen previously, they underwent allele frequency filtering.

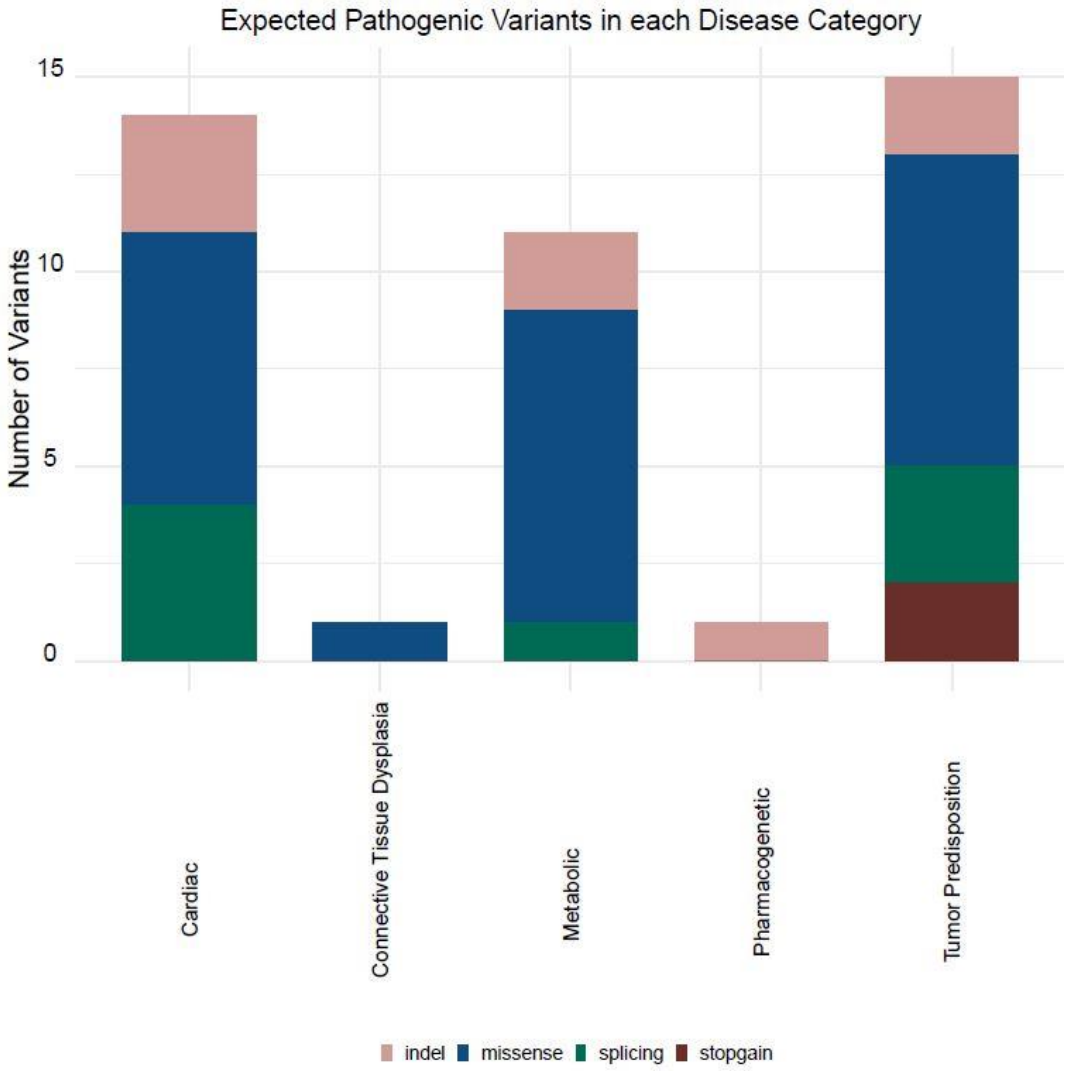


Figure 3.3. Potentially pathogenic variants by disease category. Colored by function.

Identification of P/LP variants: Manual curation:

We then performed a thorough review of the above identified 124 (40 "potentially pathogenic variants" plus 84 missense) and classified per ACMG/AMP guidelines for variant interpretation and updated guidelines from ClinGen.[96,97] From these, 12 met criteria for P/LP, although 5 of these were suspicious due to Mendel errors and the number of apparent 'de novo' mutations (Table 3.2). We performed Sanger sequencing on representative heterozygotes for each of the 12 variants except two that we have previously shown to be real (*APOB* p.Arg3527Gln[27] and *KCNQ1* p.Thr224Met[28]). The five variants showing poor evidence for Mendelian segregation were determined to be artifacts. The remaining seven variants considered to be P/LP were confirmed by Sanger sequencing. Additionally, 104 (83.9%) variants in 49 genes were designated as VUS (Table 3.3). Missense variants accounted for 90% of VUS.

Table 3.2. Variants classified as P/LP by manual curation. Five variants were determined to be artifacts (red).

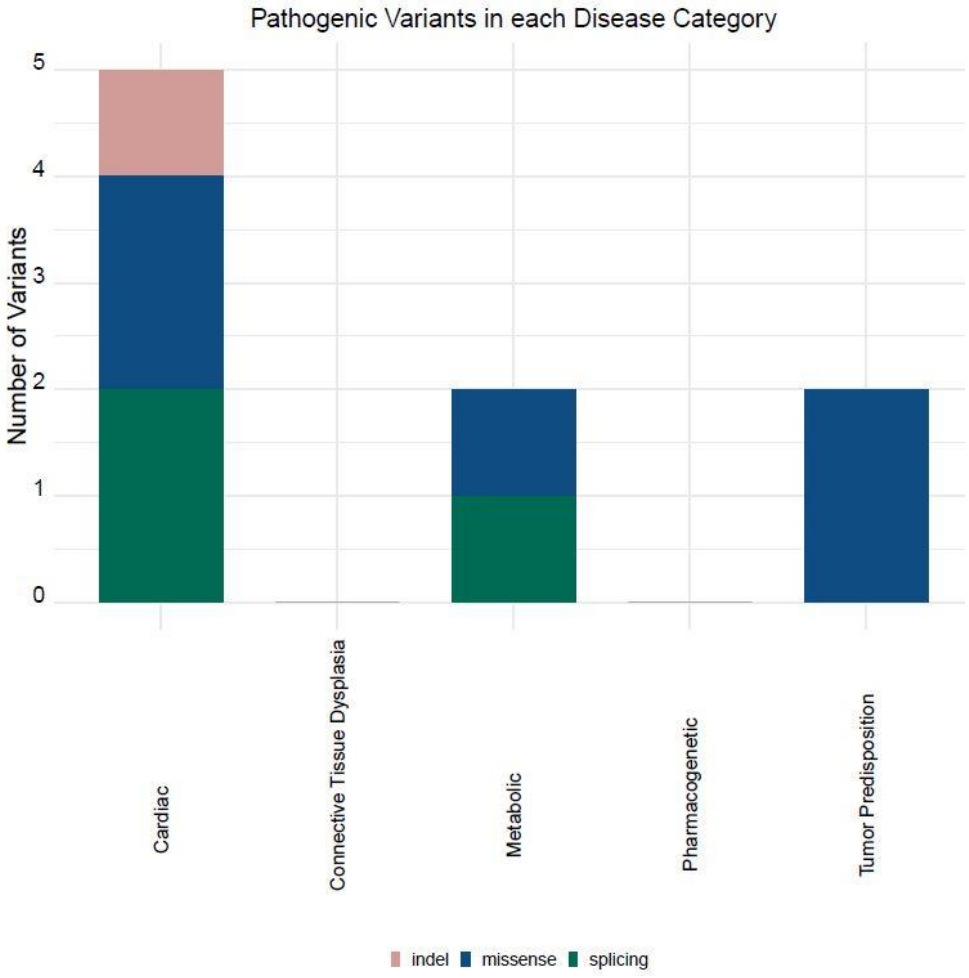
chr (hg38)	pos (hg38)	AAchg	Gene	Disease	Amish Freq	Annotation	Evidence
11	2571391	p.T224M	KCNQ1	Long QT syndrome	0.011423	P	PMID: 33141630
7	5997424		PMS2	Lynch Syndrome	0.001403	LP	PVS1, PM2_Supporting
23	38403694	p.M206fs	OTC	Ornithine carbamoyltransferase deficiency	0.000165	LP	PVS1, PM2
17	41763323		JUP	Cardiomyopathy, dilated	0.001437	LP	PVS1, PM2
7	150947431	p.A677fs	KCNH2	Long QT syndrome	0.000668	LP	PVS1, PM2
2	21006288	p.R3527Q	APOB	Hypercholesterolemia, familial	0.066596	P	PS4, PP1-strong, PS3, PP3
10	43118458	p.L790F	RET	Multiple endocrine neoplasia	8.10E-05	P	PS1, PP1-strong, PP3
11	47333192		MYBPC3	Familial Hypertrophic Cardiomyopathy	0.000163	P	PVS1, PS4, PP1
10	87960892		PTEN	PTEN hamartoma tumor syndrome	0.003509	P	PVS1, PM2, PS4_Supporting
1	156134457	p.R78W	LMNA	Cardiomyopathy, dilated	0.00057	P	PS3, PS4, PM2_Supporting, PM5_Supporting, PP1_Strong, PP3, PM1
17	7674220	p.R116Q	TP53	Li-Fraumeni syndrome 1	0.000081	P	PS3 , PS4, PP3, PM1, PM5, PM6
23	38369878		OTC	Ornithine carbamoyltransferase deficiency	0.000081	P	PVS1, PM2_Supporting, PP4

Table 3.3. Disease-causing pathogenic variants present in the Amish population within the 78 Consensus genes

Disorder	Gene	dbSNP ID	Nuc change	AA change	Amish carrier freq (x-fold enrichment in paratheses)
Dilated cardiomyopathy	*LMNA*	rs59026483	c.232C>T	p.R78W	0.0011 (∞)
Dilated cardiomyopathy, Familial hypertrophic cardiomyopathy	*MYBPC3*	rs387906397	c.3330+2A>C		0.0003 (4.9)
Familial hypercholesterolemia	*APOB*	rs5742904	c.10580C>T	p.R3527Q	0.13 (234.5)
Familial medullary thyroid carcinoma	*RET*	rs75030001	c.2370G>T	p.L790F	0.0002 (3.9)
Li-Fraumeni syndrome 1	*TP53*	rs11540652	c.347C>T	p.R116Q	0.0002 (3.9)
Ornithine carbamoyltransferase deficiency	*OTC*	rs68058881	c.298+1G>A		0.0002 (∞)
Long QT syndrome	*KCNQ1*		c.671C>T	p.T224M	0.023 (1283.3)

The minor allele frequencies of the 7 pathogenic variants in the Amish range between 0.00008 and 0.067, the highest is the missense change p.Arg3527Gln in *APOB* (carrier frequency 13.01% in Amish; 0.11% in European). The *APOB* p.Arg3527Gln variant, pathogenic for familial hypercholesterolemia (FH, MIM #144010), is strongly associated with elevated LDL-C levels and coronary artery calcification.27 We have recently reported the *KCNQ*1 p.Thr224Met (carrier frequency 2.28% in Amish; 0.0009% in European) variant to be pathogenic for long QT syndrome and associated with longer QT interval and increased risk of syncope and unexplained sudden death.[28] In total, 14.7% of our study population harbored at least one of these 7 clinically actionable pathogenic variants; the percentage was 1.7% if all individuals harboring at least one *APOB* p.Arg3527Gln variant were excluded (Figure 3.4).

Figure 3.4. Confirmed P/LP variants by disease category. Colored by function.



Comparison with non-Amish: number of variants and number of individuals harboring these variants:

We compared the burden of actionable variants in the Amish with that in three different non-founder populations (DiscovEHR[40], ESP[81], and UK Biobank[80]) .Each study used similar panels of genes categorized as clinically actionable. The Amish population harbored the least number of reportable variants but the highest frequency of individuals

that harbor one of these variants (Table 3.1). This result is expected for a founder population in which there is decreased genetic diversity and increased frequency of some disease-causing alleles not under purifying selection due to genetic drift. In each population, the disease class in which actionable variants are found varies. A large portion of variants identified in Amish individuals cause metabolic or cardiac disorders. Variants within genes on the panel that cause connective tissue dysplasias and pharmacogenetic disorders were absent in the Amish population but present in cosmopolitan populations (Figure 3.5).

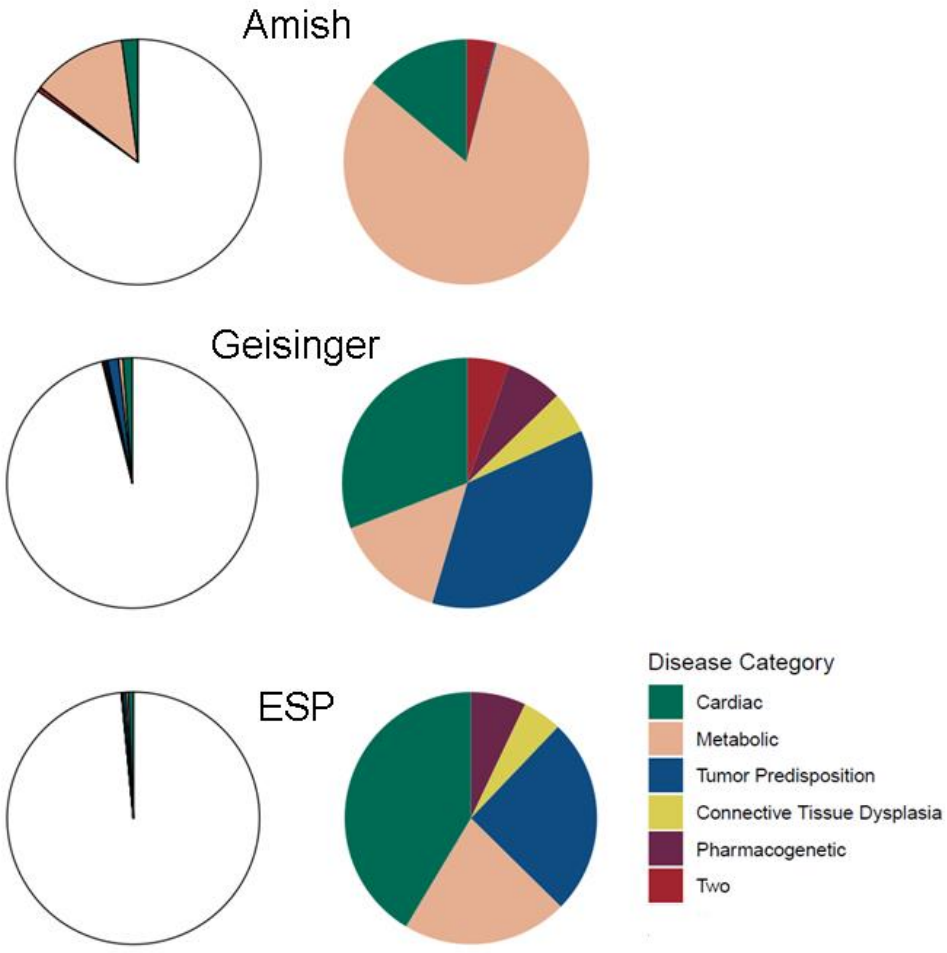


Figure 3.5. Frequency of P/LP variants by disease category in each population. All subjects (left) and only individuals harboring at least one copy of a pathogenic variant (right).

Discussion

Employing a rigorous annotation and curation process, we found that ~14.7% of all Lancaster County Amish carry at least one of seven pathogenic variants in seven clinically actionable genes. In comparison, similar studies performed in non-founder populations report that between 2.0% and 6.2% of individuals harbor actionable variants. Notably, this increase is largely explained by the high prevalence of the *APOB* p.Arg3527Gln FH variant, omitting this variant, 1.7% of Amish carried at least one of the remaining six P/LP variants.

ES performed as part of a research study can uncover findings of potential clinical importance. Although there is no systemic legal obligation to report any medically relevant findings from research studies, institutions are increasingly choosing to offer participants validated genetic results that are of high clinical significance and actionability based on ethical considerations.[75] There is significant advocacy for disclosing individual research results contingent that researchers clearly communicate that findings require CLIA confirmation which introduces philosophical, legal, funding, and practical challenges for investigators.[99–101] Many of these challenges can be lessened or circumvented in founder populations where there may be a smaller number of clinically actionable variants, each with more individuals harboring a variant, making genetic screening and return of results more straightforward. Additionally, many concerns associated with follow-up can be alleviated through active engagement with the community, as with the ongoing relationship between the Lancaster Amish, the University of Maryland, referral patterns to local clinicians, and the geographic localization of this community.

Cascade screening, performed in founder populations where large multi-generational families are prevalent, increases the likelihood of identifying additional individuals harboring a pathogenic variant, creating opportunities to identify key factors determining disease penetrance, expressivity, and genetic and environmental modifiers of risk. In addition, cascade screening has the potential to diagnose individuals with a monogenic disorder while they are asymptomatic, after which risk reduction strategies can be offered. Targeted detection of additional heterozygotes through genetic testing is currently underway in the Amish community for the *KCNQ1* p.Thr224Met variant, which causes long QT syndrome.[28] Most individuals harboring at least one copy of the *APOB* p.Arg3527Gln variant have not been notified of their genetic results, but LDL-C levels were reported back to all study subjects with alerts to those with elevated levels, including *APOB* p.Arg3527Gln carriers.[27] We are currently returning results for the five other pathogenic variants identified in the current study, including the *MYBPC3* splice site variant c.3330+2T>G which has been a focus in other Amish populations.[102,103] These studies benefit the Amish population through population-based precision health initiatives and provide an opportunity to better understand disease prognosis using a population with many heterozygotes and a limited number of environmental influences.

The knowledge that a founder variant exists within a population enables more accurate estimates of prior probability of carrying a mutation and enables targeted genetic testing and family counseling. Universal screens may lack variants that are enriched or novel in founder populations or known variants with varying penetrance dependent on environmental factors present in specific populations. ACMG emphasizes the importance of further assessment of penetrance and expressivity of disease in particular populations,

rather than implementing general population screening.[104] Designing a founder population-specific panel for future genetic screening, targeted to certain variants, would be more rapid and less expensive. The Plain Insight Panel is an example of a genotyping array adapted for genetic testing that leverages population specific information, and is applicable to the Lancaster Amish population.[5]

We encountered several challenges during manual curation and interpretation of findings. First, the majority of current ClinVar annotations have not been curated by an expert panel and must undergo manual review. Moreover, establishing the pathogenicity of clinically actionable variants is often challenging due to small sample size of rare disease, insufficient evidence in the literature, and incomplete knowledge of disease penetrance and variable expressivity. Without such information, we were unable to gather sufficient evidence to reclassify most VUS. Reclassification of missense variants provides an especially difficult challenge because their functional consequence is difficult to predict. Second, our filtering pipeline eliminated from analysis the singletons that did not have previous ClinVar annotations. This was done to prioritize variants that were likely real but is a limitation to the study because we could have missed disease causing de novo mutations. An additional limitation is that we did not analyze CNVs from the sequence data.

Utilizing guidelines outlined by the ACMG/AMP and ClinGen with high stringency ensured that we were taking a conservative approach and not overestimating the number of pathogenic variants in this population. Using data from a founder population, we were able to uncover rare variants with clinical actionability and many novel VUS. Of the 124 variants annotated, 58 do not have previous ClinVar submissions.

Of the 28 variants previously reported in ClinVar at least once as either pathogenic or likely pathogenic, only six were classified as such by our team. This result highlights the importance of following the ACMG/AMP guidelines including up to date recommendations from the ClinGen SVI to improve accuracy of variant classification.[96,97] We agreed that an important piece of the variant classification procedure was to incorporate all rule updates and employ additional comprehensive review of variants not yet curated by ClinGen Variant Curation Expert Panels (VCEP).

The filtering steps to prioritize variants for curation include a minor allele frequency cutoff in reference populations. Several VUS were low frequency in gnomAD but enriched in Amish. Given the known penetrance of cancer predisposition genes, we could conclude that some of these variants are not pathogenic due to frequency in the Amish population. However, for many genes associated with cardiac disorders characterized by reduced penetrance, we did not have enough data to conclude that high frequencies in the Amish equate to a lack of contribution to disease (Figure 3.6). There are several key pieces of information needed to determine that the variants on the higher end of this frequency spectrum are definitively not causing disease. Our current phenotyping data does not include comprehensive family history in cancer and tumor related disorders, inborn errors of metabolism, or malignant hyperthermia sensitivity. Future studies could gather additional phenotype data from heterozygotes and family members that might enable reclassification of the identified VUS.

A potential follow-up to this study would include targeted recruitment of participants and deep phenotyping of individuals who carry a VUS and their family members, family data questionnaires, echocardiograms, and other disease-specific testing

to identify functional roles and calculations of locus heterogeneity and variant penetrance. For example, the p.Arg78Trp *LMNA* variant has eight heterozygotes in a single Amish pedigree and its current annotation is VUS. *LMNA* variants are a known cause of a range of disorders known as laminopathies, including dilated cardiomyopathy and sudden cardiac death (MIM#115200), as well as familial partial lipodystrophy type 2 (MIM #151660) characterized by features of metabolic syndrome.[105,106] Obtaining additional and deep phenotype data on the members of this pedigree harboring pathogenic variants could help aid in the re-classification of this VUS.

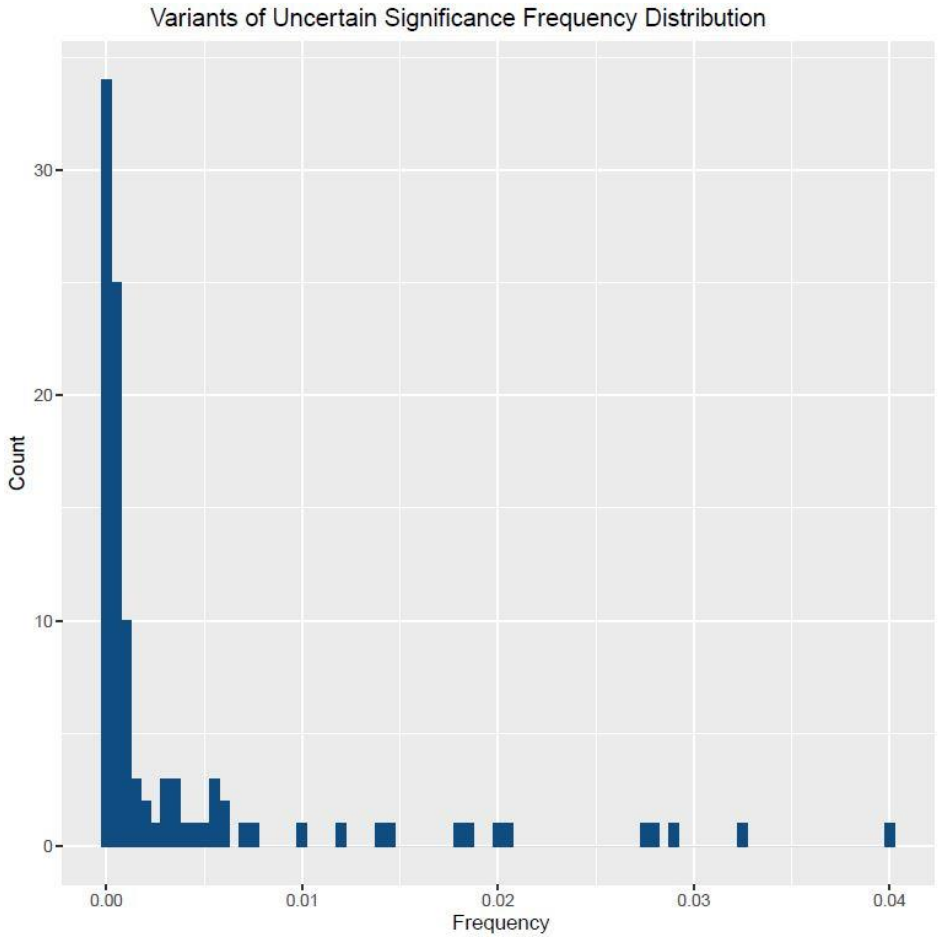


Figure 3.6. Allele frequency spectrum of VUS. There were 104 variants classified as VUS

Summary

We have assessed the burden of pathogenic and likely pathogenic clinically actionable variants in the Lancaster County Amish population versus three non-Amish European populations. As expected, the Amish have fewer total actionable variants identified in comparison to the cosmopolitan populations studied. However, the Amish have a higher frequency of individuals harboring the identified pathogenic variants. Following careful quality control and analysis, we identified seven variants carried in 900 of 6,136 individuals (14.7%), including the *APOB* p.Arg3527Gln variant in which 13% of Amish individuals harbor at least one copy. Of these seven variants meeting the criteria for pathogenicity as described in the current analysis, two have been previously well studied in this population. The high frequency of these two clinically actionable variants offers opportunities for targeted precision medicine in an efficient and cost-effective way. These results highlight a different pattern of enrichment of clinically actionable variants in founder populations compared to outbred cosmopolitan populations.

V. Impact of parental relatedness on reproductive outcomes[2]*

Abstract

Genetically isolated populations that have arisen due to recent bottleneck events have reduced genetic variation that reflects the common set of founders. The increased genetic relatedness among members of isolated populations puts them at increased risk for some recessive disorders that are rare in the general population. To assess the burden on reproductive health in a population enriched for recessive conditions, we compared the frequency of adverse reproductive outcomes between Amish couples who were both heterozygous carriers of a highly penetrant pathogenic or likely pathogenic variant and non-carrier couples from the same Amish community. In addition, we evaluated whether overall genetic relatedness of parents across the genome was associated with reproductive health outcomes. Of the 1824 couples included in our study, 11.1% were at risk of producing a child with an autosomal recessive disorder. Carrier couples reported a lower number of miscarriages compared to noncarrier couples (p=0.02), although the number of stillbirths (p=0.3), live births (p=0.9) and number of pregnancies (p=0.9) did not differ significantly between groups. In contrast, higher overall relatedness between spouses was positively correlated with number of live births ($p < 0.0001$), pregnancies ($p < 0.0001$), and stillbirths (p=0.03), although not with the number of miscarriages (p=0.4). These results highlight a complex association between relatedness of parents and reproductive health outcomes in this founder community.

[2]Megan T. Lynch, Kristin A. Maloney, Toni I. Pollin, Elizabeth A. Streeten, Erik G. Puffenberger, Kevin A. Strauss, Regeneron Genetics Center, Alan R. Shuldiner, Braxton D Mitchell. "Impact of Parental Relatedness on Reproductive Outcomes." In preparation for submission.
*MTL performed statistical analysis, drafted the manuscript, and generated tables and figures.

Introduction

Approximately 10-20% of recognized pregnancies end in miscarriage, leading to significant psychological impact on the parents.[107,108] Although most pregnancy losses are unexplained[43,109], many are associated with chromosomal abnormalities[110,111] or other known genetic abnormalities. Certain single gene recessive disorders (e.g. Smith-Lemli-Optiz syndrome[112], glycogen storage diseases[110,113], among others[109,114]) may also result in stillbirth or miscarriage when the fetus is affected. The burden of pregnancy loss may be particularly high in populations with high degrees of consanguinity, due to higher frequencies of recessive disorders affecting pregnancy outcomes or limited genetic diversity in general.[115–118]

Founder effects, which occur following the isolation of a subpopulation, reduce genetic variation, and genetic drift may inflate the incidence of some recessive disorders. In founder populations, marriages between individuals with increased relatedness (elevated kinship) is common. These marriages are identified as consanguineous if they share a common biological ancestor usually up to second cousins, equivalent to a coefficient of inbreeding in their children of $F \geq 0.0156$.[78,119] Consanguinity has been defined elsewhere as first-degree cousin marriages.[120] The incidence of congenital malformations among offspring of consanguineous unions has been estimated between 1.3% to 4.5%, compared to ≤1% in non-consanguineous unions.[78,121] Unions occurring between individuals with elevated kinship increase the probability of a zygote receiving two copies of a disease-causing recessive allele and could potentially have a negative impact on reproductive outcomes. It is also possible that genetic relatedness of the parents in general may contribute to pregnancy loss, and that being a carrier for a disease-

causing recessive condition may simply be a marker for higher relatedness across the whole genome. For example, high kinship among parents has been identified as a risk factor for stillbirth[55] and miscarriage.[56]. Conversely, some[7,49,50,52], but not all[53,54], studies have reported correlations between higher consanguinity and better fertility outcomes.

The Amish population of Lancaster County, PA is a genetic isolate with all members deriving from a common set of founders and 128 of the founding members account for over 95% of the average extant genome.[6] Although unions between closely related individuals (e.g., first cousins) are rare, the small effective population size and reduced genetic diversity of this community leads to unions between couples who are at least moderately related. The average spouse pair does not have one relationship that's as close as first cousins once removed, but equivalent kinship values are seen because they are more distant cousins but related in multiple ways. Consequently, the risk of producing offspring with rare Mendelian recessive variants is increased. Some recessive disorders that are enriched in the Amish have previously been linked to pregnancy loss[122–127] or fetal structural abnormalities.[128]

Through our long-standing population health studies in the Lancaster Amish[15,23,129,130], we retrospectively identified 1,824 couple pairs in whom exome sequencing (ES) has been performed and detailed information about pregnancy outcomes was available. In this study, we evaluate the impact of known highly penetrant recessive variants[5] on reproductive outcomes. We also assessed the association of pregnancy outcomes with overall relatedness between parents.

Methods

Participants and reproductive health information:

This study was based on 6,136 Amish individuals from Lancaster, PA over 18 years of age and who were recruited through prior community-wide studies conducted over the past 20 years.[89,130] Study subjects had previously filled out detailed surveys regarding their prior pregnancy and reproductive histories including number of pregnancies, number of children, birth date of each of their children, number of stillbirths, and number of miscarriages. Additional lifestyle information obtained through surveys included church district, occupation, date of marriage, and date of menopause for female subjects. Calculated variables from survey phenotype data included interval between births, interval from marriage to first birth, and age at last birth (Table 4.1).

Table 4.1. Reproductive outcomes summary data. None of these variables differed significantly by t-test between carriers and non-carriers (p val > .4 for all), except R Coefficient (p=.003), mother's age at marriage (p=0.002), and number of miscarriages which was not adjusted for age.

Variable	Range	Mean ± SD	
		NonCarrier (n = 1531)	Carrier (n =202)
Mother Age	22-86	47 ± 14	47 ± 15
Father Age	22-88	49 ± 14	49 ± 14
Stillborn Number	0-2	0.08 ± 0.3	0.1 ± 0.3
Number of Children	0-16	6.3 ± 2.8	6.3 ± 2.9
Number of Pregnancies	0-22	7.6 ± 3.2	7.5 ± 3.1
Number of Miscarriage	0-10	1.0 ± 1.4	0.7 ± 1.0
R Coefficient	0-0.15	0.07 ± 0.02	0.08 ± 0.02
Mother Age First Birth	18-39	22 ± 2	22 ± 2
Mother Age Last Birth	21 - 48	35 ± 5	35 ± 5
Mother Age at Marriage	17 - 39	21 ± 2	20 ± 2
Mother Age at Menopause	27-61	49 ± 5	48 ± 6

Exome sequencing (ES) and genome informatics:

ES was conducted by Regeneron Genetics Center (RGC), using a high-throughput automated in-house pipeline utilizing a slightly modified version of the xGen (Integrated DNA Technologies) capture. Sequencing was performed on the Illumina NovaSeq platform using paired-end 75 bp reads. A cloud-based pipeline of standard tools was used for sample-level data production and analysis. Raw sequence data were uploaded to DNAnexus which triggered the automated production analysis pipeline. Key steps were sample de-multiplexing using Illumina software, alignment to the GRCh38 reference sequence, post-alignment BAM processing (e.g. marking of duplicate reads and other read mapping evaluation), and SNP and intra-read INDEL calling with genotyping software.

Evaluation of carrier status of pathogenic and likely pathogenic recessive variants:

Previously, a large panel of pathogenic or likely pathogenic (P/LP) variants segregating in the Lancaster Amish and another large founder population, the Old Order Mennonites, was curated from comprehensive clinical testing of children affected with presumed recessive disorders. In order to facilitate carrier screening and diagnostic testing in these communities, a novel sequencing assay was previously developed by the Clinic for Special Children (the Plain Insight Panel (PIP)) which enables population-wide carrier screening of all known P/LP variants in a single test.[5] Using exome sequencing data in 6,136 Amish, which represents approximately one-sixth of the Lancaster Amish population, we focused on variants causing the Mendelian recessive conditions included on the PIP and used the carrier frequencies from this sample to predict population level frequencies for expected number of children with one of these recessive disorders due to

homozygosity or compound heterozygosity. To find the frequency of carrier couples, we concentrated on the 1824 couples with ES data. For each of the 74 P/LP recessive variants across 70 genes carried by the Lancaster Amish, genotype status was extracted, and we identified couples in whom both members carried the same or different P/LP recessive variants in the same gene. This data was obtained retrospectively and was not provided to the Amish individuals analyzed.

Coefficient of relationship:

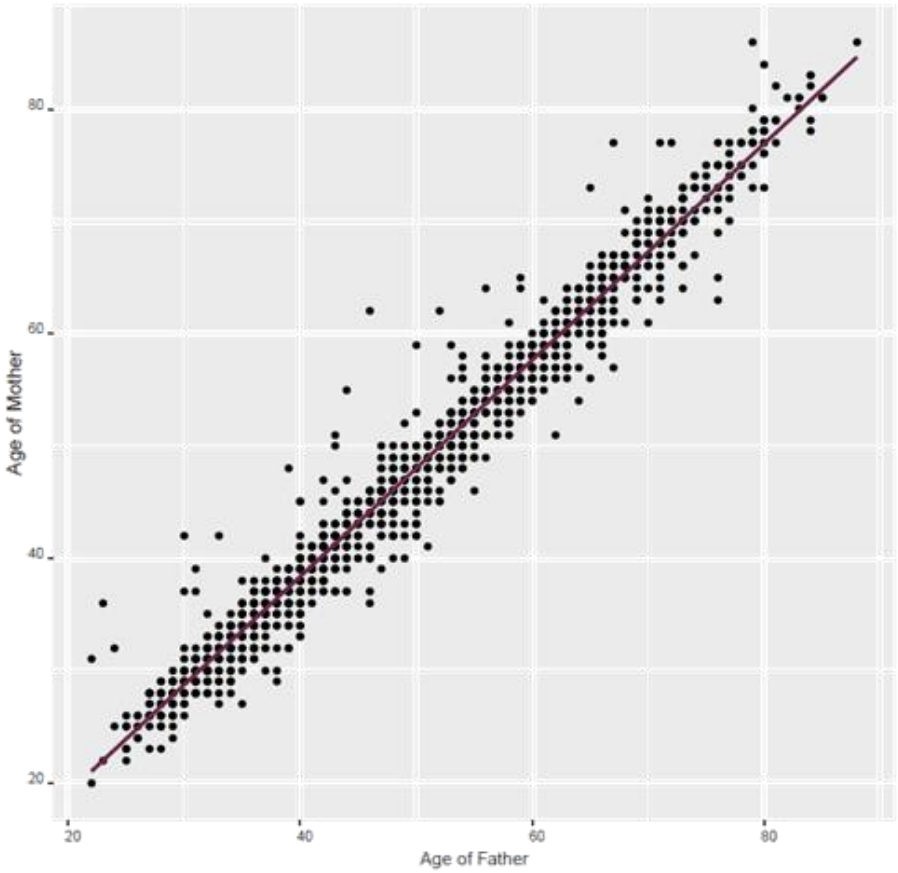
We measured degree of relatedness within spouse pairs using the coefficient of relationship (R), which is the proportion of genes that two individuals have in common and is twice the coefficient of inbreeding (F). We used our in-house MMAP software[131] to compute relationship matrices using genetic markers to estimate the relationship coefficients from SNP data and extracted the values within spouse pairs for the 1824 couples with ES data.

Statistical Analysis:

The primary predictor of reproductive outcomes was either carrier couple classification, both members being heterozygous for a P/LP variant in the same gene, or the R coefficient of the couple. Reproductive outcomes included number of stillborn, number of children, number of pregnancies, number of miscarriages, birth interval, and time to first birth. To perform stratified analysis of reproductive outcomes, couples were grouped by age of mother and kinship value quartiles were used. The age of the mother at the time of recruitment (obtained from reproductive survey) was used as a covariate in all association analysis. Age of father was highly correlated with age of mother (r=0.99)

therefore only one needed to be included in the model (Figure 4.1). Both generated the same result from statistical association analysis. For binary traits, logistic regression models were run using the same covariates as in the linear regression models for quantitative traits. Confounding factors tested included age at marriage, age at first birth, age at last birth (of individuals who were older than 45 at time of survey), and age at menopause. When evaluating the number of miscarriages, total number of pregnancies was included as a confounding variable. Couples with zero children were included in the analysis. Removing them did not alter the strength of associations.

Figure 4.1 Age of mother and father were highly correlated (r=0.99, p <.0001)



Results

The 6,136 Amish participants carried 74 pathogenic and likely pathogenic variants in genes associated with recessive conditions. Allele frequencies of variants ranged from 0.0001 to 0.08, and the highest frequency variant was rs121918299 in *TJP2* (Figure 4.2, Table 4.2).

Figure 4.2. Allele frequency spectrum of P/LP recessive variants. There was a total of 74 variants identified in the 6,136 Amish with exome sequencing data.

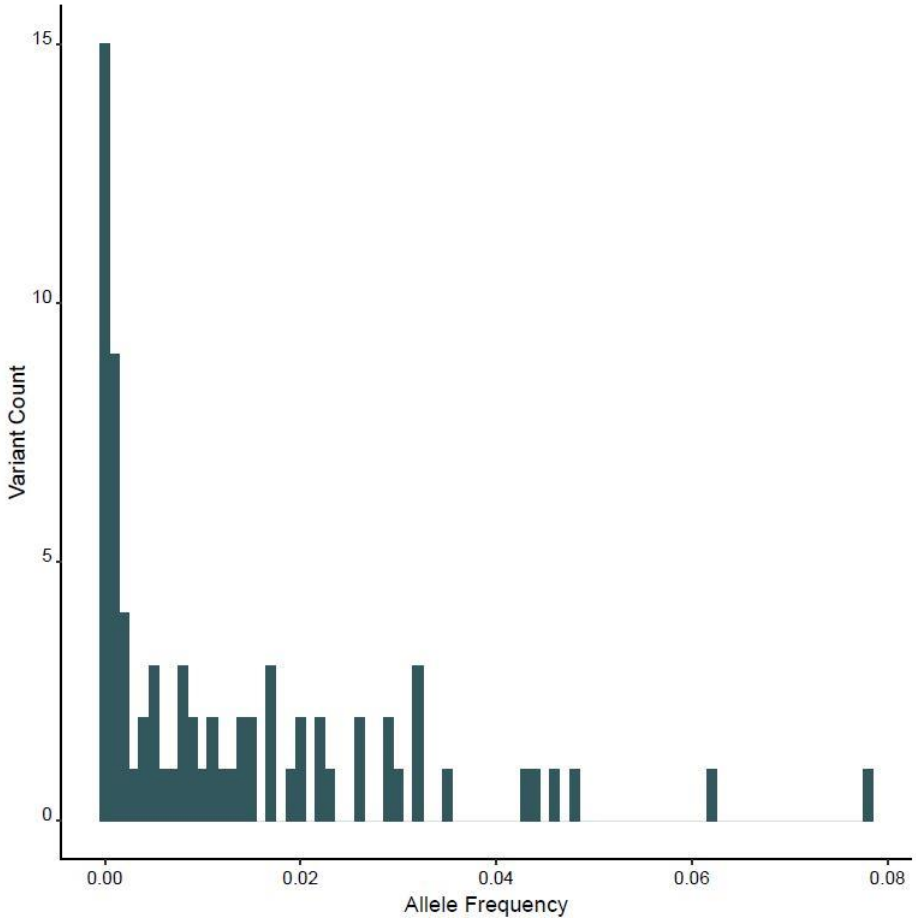


Table 4.2. P/LP variants in which at least one carrier was identified within 1824 couples. Frequency refers to the entire Amish population (n=6136).

Gene	Disorder	Sequence Variants	p.	Freq	Couples (N)
ABCG8	Sitosterolemia	c.1720G>A	p.Gly574Arg	0.0181	1
ACADM	Medium-chain acyl-CoA dehydrogenase deficiency	c.199T>C	p.Tyr67His	0.0250	4
BAAT	Familial hypercholanemia	c.226A>G	p.Met76Val	0.0280	2
BBS1	Bardet-Biedl syndrome	c.1169T>G	p.Met390Arg	0.0205	10
BRAT1	Lethal neonatal rigidity and multifocal epilepsy	c.638dupA	p.Val214Glyfs	0.0141	3
CNTNAP2	Cortical dysplasia and focal epilepsy	c.3709delG	p.Asp1237fs	0.0217	1
CYP24A1	Infantile hypercalcemia	c.428_430delAAG	p.Glu143del	0.0294	4
DUOXA2	Thyroid dyshormonogenesis 5	c.829_830delTG	p.Cys277fs	0.0260	4
EVC	Ellis-van Creveld syndrome	c.1886+5G>T		0.0613	23
GBA	Gaucher disease	c.1226A>G	p.Asn409Ser	0.0216	3
GCDH	Glutaric aciduria, type 1	c.1262C>T	p.Ala421Val	0.0314	7
GJC2	Hypomyelinating leukodystrophy	c.203A>G	p.Tyr68Cys	0.0087	1
HFE	Hereditary hemochromatosis	c.845G>A	p.Cys282Tyr	0.0440	17
HSD3B2	3-[f-Hydroxysteroid dehydrogenase deficiency	c.35G>A	p.Gly12Glu	0.0261	4
KPTN	Macrocephaly, neurodevelopmental delay, and seizures	c.776C>A	p.Ser259Ter	0.0080	1
LONP1	CODAS syndrome	c.2161C>G	p.Arg721Gly	0.0469	12
MKKS	McKusick-Kauffman syndrome	[c.250C>T + c.724G>T]	p.His84Tyr/p.Ala242Ser	0.0483	12
MYORG	Basal ganglia calcification, idiopathic, 7	c.1967T>C	p.Ile656Thr	0.0321	8
PAH	Phenylketonuria	c.1066-11G>A		0.0004	7
		c.782G>A	p.Arg261Gln	0.0148	
		c.284_286delTCA	p.Ile95del	0.0164	
PCCB	Propionic acidemia	c.1606A>G	p.Asn536Asp	0.0161	1
SERPINA1	Alpha-1 antitrypsin deficiency	c.1096G>A	p.Glu342Lys	0.0083	2
SLC25A19	Amish microcephaly	c.530G>C	p.Gly177Ala	0.0361	12
SLITRK6	Deafness and myopia	c.1240C>T	p.Gln414Ter	0.0296	8
SYNE1	Spinocerebellar ataxia type 8	c.17905C>T	p.Gln5898Ter	0.0165	3
TJP2	Familial hypercholanemia	c.143T>C	p.Val48Ala	0.0778	46
TMCO1	Craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome	c.292_293delAG	p.Ser98fs	0.0204	3
TNNT1	Nemaline rod myopathy	c.538G>T	p.Glu180Ter	0.0328	13

We identified 197 out of 1824 (10.18%) couples who were carriers for the same P/LP variant and were at risk of producing a child with an autosomal recessive disorder. An additional five couples were carriers of different variants within the same gene and are at risk of producing a compound heterozygous child (202 out of 1824; 11.1%). Most couples each had a disease-causing variant in just one gene, eleven couples each had disease-causing variants in more than one gene (Figure 4.3). In total, the 202 Amish carrier couples shared 29 disease-causing variants in 27 genes (Table 4.3).

Figure 4.3 Number of P/LP variants that both members of a couple pair had in common.

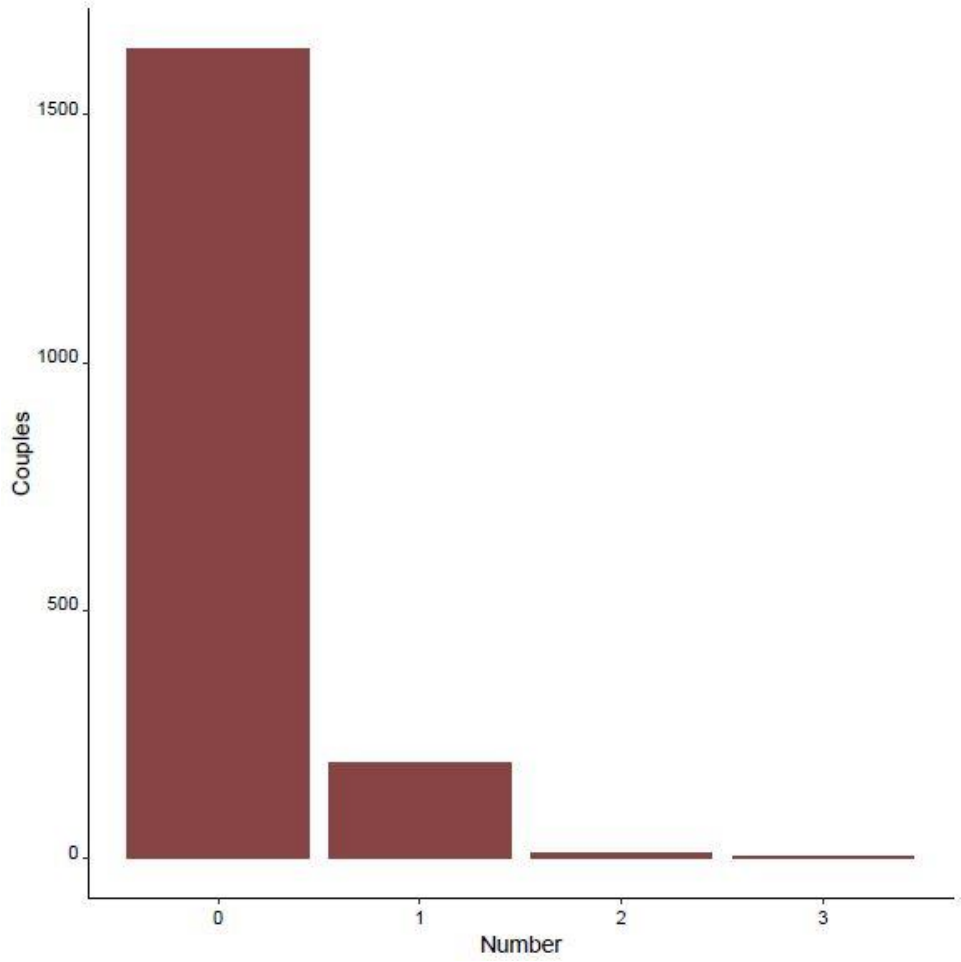


Table 4.3. Disorders in which couple pairs were each heterozygous for a P/LP variant.

Disorder	Couples (N)
Sitosterolemia	1
Medium-chain acyl-CoA dehydrogenase deficiency	4
Familial hypercholanemia	2
Bardet-Biedl syndrome	10
Lethal neonatal rigidity and multifocal epilepsy	3
Cortical dysplasia and focal epilepsy	1
Infantile hypercalcemia	4
Thyroid dyshormonogenesis 5	4
Ellis-van Creveld syndrome	23
Gaucher disease	3
Glutaric aciduria, type 1	7
Hypomyelinating leukodystrophy	1
Hereditary hemochromatosis	17
3-├ƒ-Hydroxysteroid dehydrogenase deficiency	4
macrocephaly, neurodevelopmental delay, and seizures	1
CODAS syndrome	12
McKusick-Kauffman syndrome	12
basal ganglia calcification, idiopathic, 7	8
phenylketonuria	7
propionic acidemia	1
alpha-1 antitrypsin deficiency	2
Amish microcephaly	12
deafness and myopia	8
spinocerebellar ataxia type 8	3
familial hypercholanemia	46
craniofacial dysmorphism, skeletal anomalies, and mental retardation syndrome	3
nemaline rod myopathy	13

Due to the large range in age of participants and the known associations between parental age and many of the reproductive outcomes, we adjusted for parental age in all analyses evaluating the association between coefficient of relationship between couple pairs and reproductive outcomes. Maternal and paternal ages were highly correlated in this population (r^2=.99), so we adjusted only for age of the mother, acknowledging that adjustment for paternal age produced similar results. As expected in a closed founder population, most (1822 out of 1824; 99.9%) unions had a coefficient of relationship (R) estimate >0.005, which is the estimated value in randomly mating outbred populations.[46] The mean age of the mother at the time of marriage was younger in couples who each carry a P/LP recessive variant, compared to non-carrier couples and this difference was statistically significant by t-test (20.3 vs. 20.8 years, p = 0.002) and carrier couples had a higher R coefficient (0.08 vs. 0.07, p = 0.0003) (Table 4.1). Carrier couples had a lower total number of miscarriages (0.73 vs. 1.03, p = 0.003), although the number of stillbirths (0.09 vs. 0.08, p = 0.5), number of pregnancies (7.5 vs. 7.6, p = 0.6), and number of live births (6.3 vs. 6.3, p = 0.8) did not differ significantly between groups (Table 4.4, 4.5, Figure 4.4)

Table 4.4. Regression association analysis of couple carrier status and R coefficient with reproductive outcomes.

	Miscarriage (n = 1154)		Stillborn (n = 1717)		Children (n = 1655)		Pregnancy (n = 1677)	
	Beta	P Val	Beta	P Val	Beta	P Val	Beta	P Val
Mother Age	0.004 ± 0.006	0.21	0.001 ± 0.0008	0.0027	0.08 ± 0.009	<.0001	0.09 ± 0.01	<.0001
Carrier Status	-0.3 ± 0.2	0.016	0.02 ± 0.03	0.302	0.03 ± 0.4	0.89	-0.02 ± 0.5	0.95

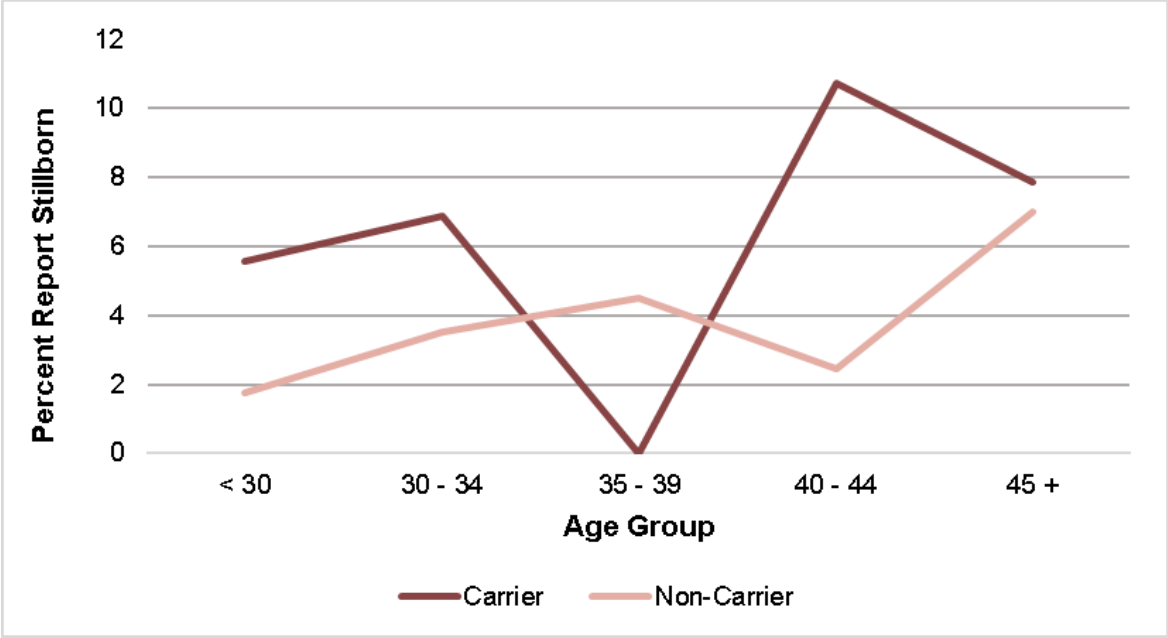
	Miscarriage (n = 1154)		Stillborn (n = 1717)		Children (n = 1655)		Pregnancy (n = 1677)	
	Beta	P Val	Beta	P Val	Beta	P Val	Beta	P Val
Mother Age	0.003 ± 0.005	0.17	0.002 ± 0.001	0.002	0.08 ± 0.009	<.0001	0.09 ± 0.01	<.0001
Kinship	3.4 ± 4.5	0.14	0.80 ± 0.62	0.011	23.6 ± 7.3	<.0001	26 ± 8.3	<.0001

Table 4.5. Stratified association analysis of couple carrier status and R coefficient with reproductive outcomes. Association analysis Mendel Hanzael chi square test

Mother Age	% Report Stillborn (n)				Mean Number Pregnancy (n)				Mean Number Children (n)			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
< 30	0 (16)	14.3 (14)	0 (19)	6.7 (15)	3.3 (27)	3.6 (30)	3.9 (42)	3.7 (34)	2.9 (26)	2.8 (31)	3.0 (42)	3.1 (33)
30-34	2.9 (35)	7.3 (41)	7.3 (41)	4.8 (42)	5.4 (50)	5.9 (55)	5.6 (58)	6.0 (56)	4.7 (50)	4.7 (54)	4.3 (55)	5.0 (56)
35-39	0 (37)	4.1 (49)	8.5 (47)	10.3 (39)	6.9 (53)	7.1 (62)	7.7 (67)	7.8 (53)	5.6 (54)	6.1 (60)	6.0 (68)	6.4 (54)
40-44	4.8 (42)	4.2 (48)	7.7 (39)	2.5 (40)	7.0 (59)	8.0 (61)	8.3 (47)	8.2 (60)	5.7 (58)	6.9 (57)	7.3 (40)	6.9 (57)
45+	9.7 (165)	5.6 (142)	10.1 (149)	13.3 (158)	8.2 (221)	8.4 (202)	8.7 (196)	9.6 (208)	6.8 (217)	7.0 (204)	7.4 (199)	8.2 (204)
Total:	6.4 (295)	5.8 (294)	8.5 (295)	9.9 (294)	7.2 (410)	7.5 (410)	7.6 (410)	8.2 (411)	6.0 (405)	6.2 (406)	6.3 (404)	7.0 (404)
			χ^2: 3.55				χ^2: 33.08				χ^2: 35.46	
			Prob: 0.06				Prob: <.0001				Prob: <.0001	

Mother Age	% Report Stillborn				Mean # Pregnancy				Mean # Children			
	Carrier		Non-Carrier		Carrier		Non-Carrier		Carrier		Non-Carrier	
	%	(n)	%	(n)	#	(n)	#	(n)	#	(n)	#	(n)
< 30	5.6	18	1.7	115	4.12	24	3.56	109	2.96	21	2.94	111
30 - 34	6.9	29	3.5	200	5.67	39	5.74	180	4.92	37	4.62	178
35 - 39	0.0	26	4.5	222	7.22	36	7.42	199	5.47	36	6.12	200
40 - 44	10.7	28	2.5	203	8.61	36	7.71	191	7.06	34	6.55	178
45 +	7.9	89	7.0	787	8.83	128	8.7	699	7.25	126	7.32	698
Total:	6.84	190	5.17	1527	7.68	263	7.58	1378	6.28	254	6.34	1365
	χ^2: 1.5351				χ^2: 0.1379				χ^2: 0.1515			
	Prob: 0.2154				Prob: 0.7103				Prob: 0.6971			

Figure 4.4. Reproductive outcomes for noncarrier and carrier groups stratified by age of mother.



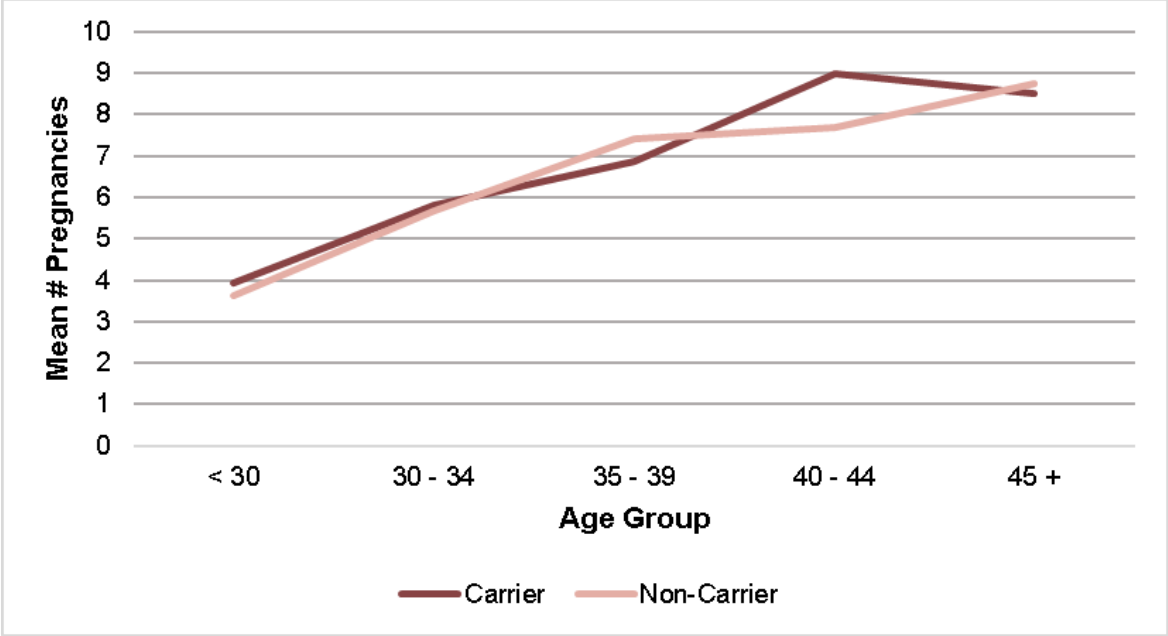
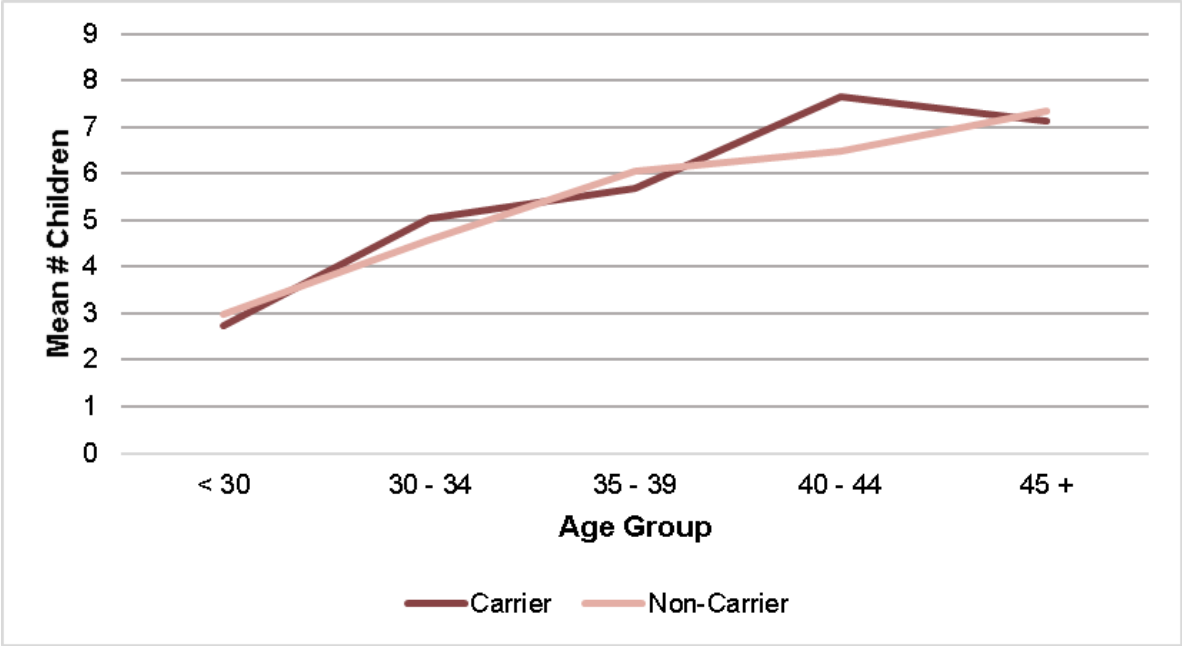
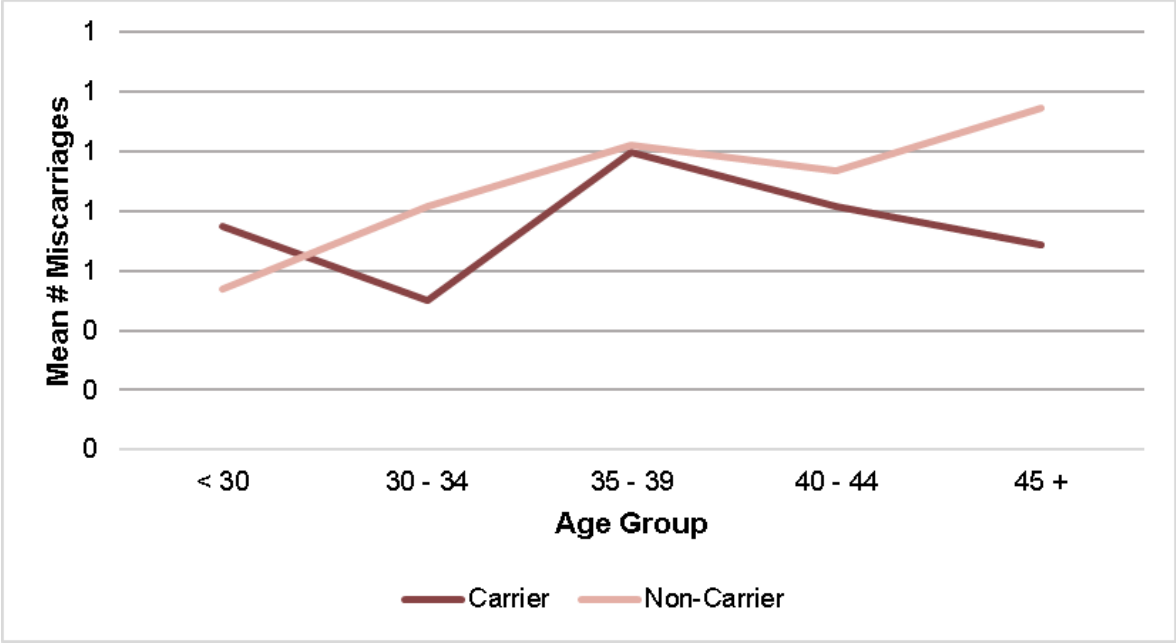


Figure 4.4. continued





Analysis of the 27 genes individually revealed multiple genes significantly associated with reporting any stillbirths. However, sample size for single gene analysis was small (Table 4.6). To further explore whether specific genes were statistically associated with reported stillbirth instance, we annotated genes according to prior knowledge of stillbirth association and created categories to group carrier couples. There was no association between carrier status and stillbirth after removing 23 couple pairs in which both members were heterozygous carriers for mutations in a gene already known to be associated with stillbirths (*EVC*; p = 0.13, OR=1.02), and there was no association present after removing an additional 74 couple carriers for mutations in genes suggested as being possibly associated with stillbirths (e.g. *GBA*, *HFE*, *LONP1* and *TJP2*) (p = 0.35, OR=1.02).

Table 4.6. Reported instance of stillbirths by carrier status for each gene.

Gene	carrier couples (n)	Proportion of couples reporting >= 1 SB		Disease Name
		carrier	non-carrier	
ABCG8	1	0	92/1716	Sitosterolemia
ACADM	4	0	92/1713	Medium-chain acyl-CoA dehydrogenase deficiency
BAAT	2	0	92/1715	Familial hypercholanemia
BBS1	10	1/10	91/1707	Bardet-Biedl syndrome
BRAT1	3	0	92/1714	Lethal neonatal rigidity and multifocal epilepsy
C8orf37	2	0	92/1715	Bardet-Biedl syndrome 21
CNTNAP2	1	0	92/1716	Cortical dysplasia and focal epilepsy
CYP24A1	4	0	92/1713	Infantile hypercalcemia
DUOXA2	4	0	92/1713	Thyroid dyshormonogenesis 5
EVC	21	4/21	88/1696	Ellis-van Creveld syndrome
GBA	3	0	92/1714	Gaucher disease
GCDH	7	0	92/1710	Glutaric aciduria, type 1
GJC2	1	0	92/1716	Hypomyelinating leukodystrophy
HFE	16	1/16	91/1701	Hereditary hemochromatosis
HSD3B2	4	0	92/1713	3- ├f-Hydroxysteroid dehydrogenase deficiency
KPTN	1	0	92/1716	Macrocephaly, neurodevelopmental delay, and seizures
LONP1	11	1/11	91/1706	CODAS syndrome
MKKS	12	2/12	90/1705	McKusick-Kauffman syndrome
MYORG	8	0/8	92/1709	Basal ganglia calcification, idiopathic, 7
PAH	7	1/7	91/1710	Phenylketonuria
PCCB	1	0	92/1716	Propionic acidemia
SERPINA1	2	0	92/1715	Alpha-1 antitrypsin deficiency
SLC25A19	12	0	92/1705	Amish microcephaly
SLITRK6	8	1/8	91/1709	Deafness and myopia
SYNE1	3	0	92/1714	Spinocerebellar ataxia type 8
TJP2	39	3/39	89/1678	Familial hypercholanemia
TMCO1	3	0	92/1714	Craniofacial dysmorphism, skeletal and mental retardation syndrome
TNNT1	12	0	92/1705	Nemaline rod myopathy

After observing a negative association between carrier status of couples and reported number of miscarriages, we evaluated more broadly whether couples who were more related to each other in general (i.e. had higher R coefficient values) were more likely to have adverse reproductive outcomes. This enabled us to test whether carrier status for a specific locus was a marker for an association that was due to higher relatedness across the genome. In the overall population of 6136 Amish participants, the mean *R* coefficient (± SD) between spouses was 0.073 ± 0.017 (Figure 4.5), a value almost 15-fold higher than randomly mating outbred populations (*R* <0.005)[132]. On average, the coefficient of relationship between spouse pairs is equivalent to fifth degree relatives including first cousins once removed. We found that a higher relatedness between spouses was correlated with a larger number of children ($p < 0.0001$), increased number of pregnancies ($p < 0.0001$), and increased instance of stillbirths ($p = 0.03$), although not with the number of reported miscarriages ($p = 0.4$) (Table 4.4, 4.5, Figure 4.6).

We also observed the R coefficient to be negatively associated with age at first birth ($p < 0.0001$) and negatively associated with age at marriage ($p < 0.0001$). These two variables were highly correlated with each other ($r^2 = 0.83$) and including either of them in the multivariate analysis did not reduce the strong association with increased number of children ($p < 0.0001$), increased pregnancies ($p < 0.0001$), or increased stillbirths ($p = 0.01$) (Table 4.5). When age of father was included in the model, the strong positive association did not change.

Figure 4.5. Distribution of Relationship coefficients in couples

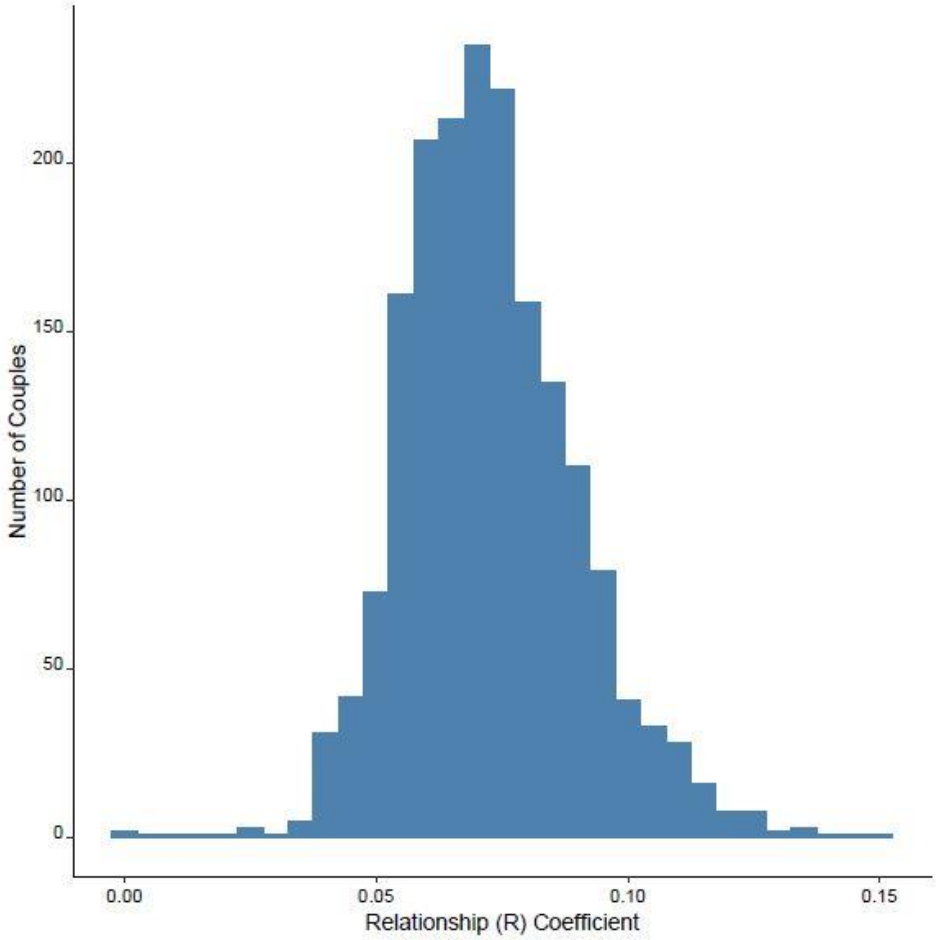


Figure 4.6. Kinship is positively associated with number of children and number of pregnancies. Kinship values in quartiles and stratified by age of mother. Mean number of pregnancies (top), children (middle) or stillbirths (bottom) reported.

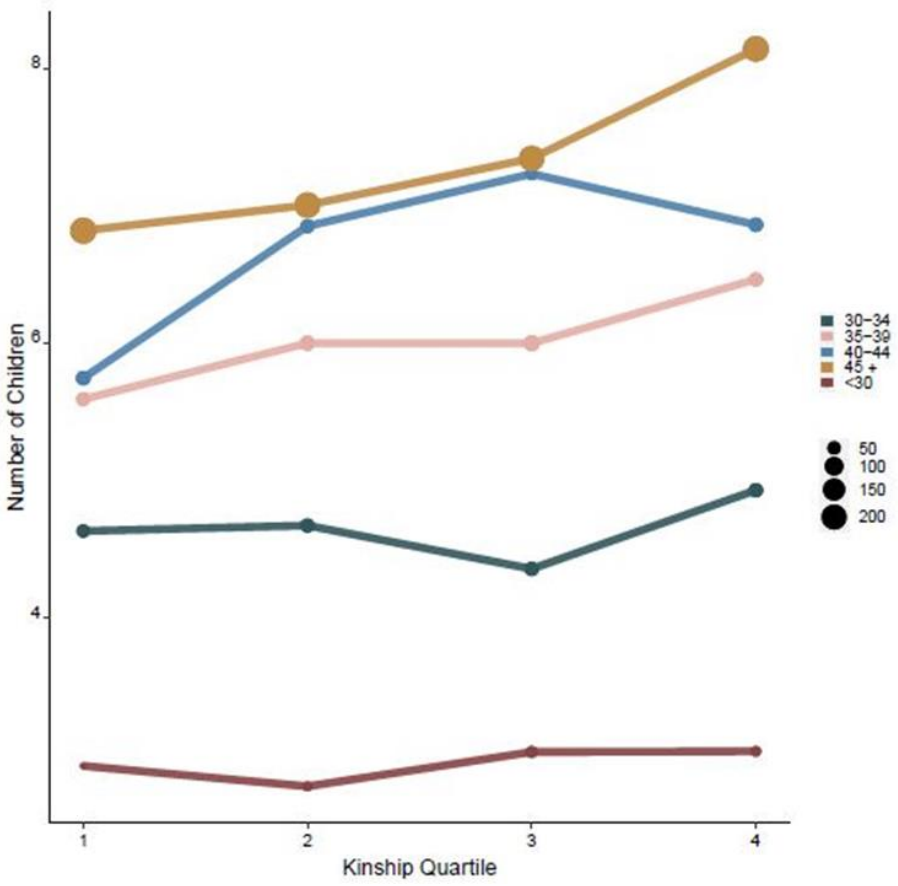
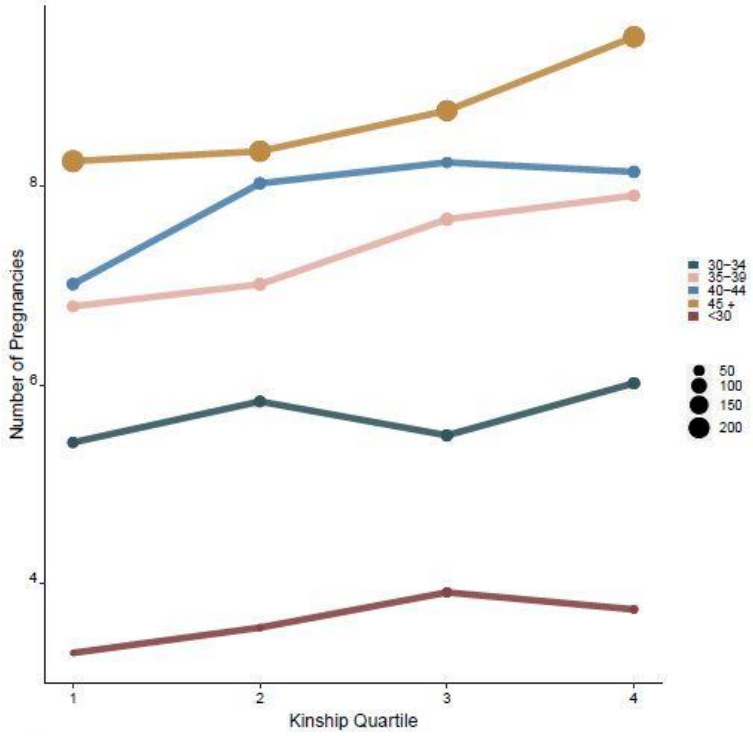
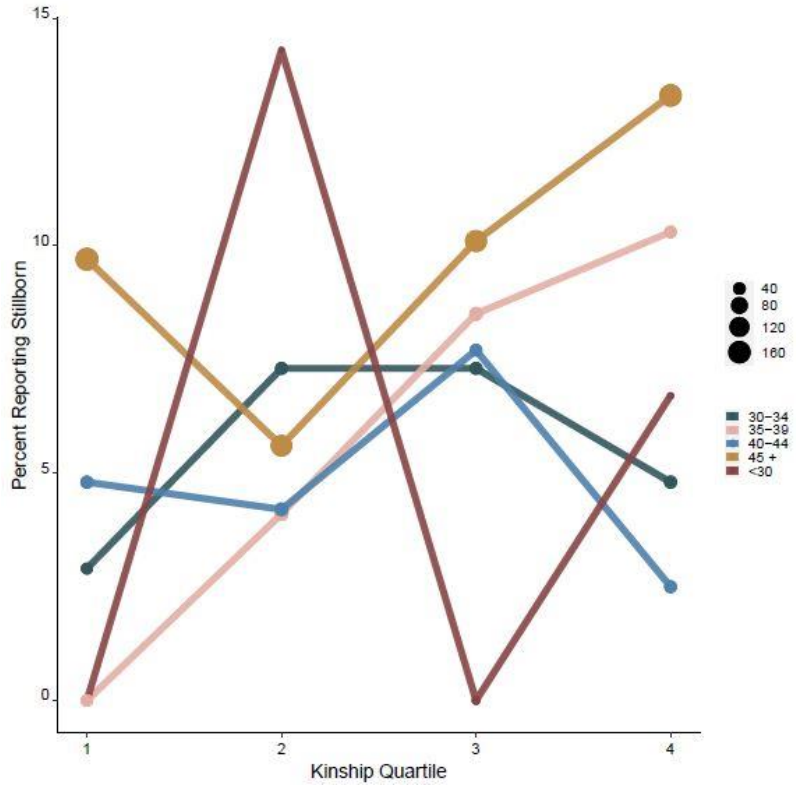


Figure 4.6. continued





Discussion

We found that carrier couples for highly penetrant autosomal recessive disorders that are common in the Amish experienced fewer miscarriages compared to noncarrier couples and that overall genetic relatedness among parents was associated with higher number of children, pregnancies, and instance of stillbirth.

Our findings align with a previous study in Amish and other Anabaptist founder populations with similar genetic backgrounds.[7] This study of a broader population of North American Anabaptist, including individuals from Lancaster Amish communities, reported that consanguineous marriage frequencies have increased over time with pedigree-based estimates of the median kinship coefficient, rising in the 20th century.[7] Highly related couples were reported to have increased family size and decreased inter-birth intervals.[7] We have replicated this result in an independent sample of this cohort and extended these results by also examining the number of pregnancies, number of miscarriages and stillborn, and age of mother at first birth.

The reasons underlying these observations are unclear. It is possible that higher relatedness among couples leads to more pregnancies to compensate for increased risk of postnatal loss, enhanced maternal-fetal genetic compatibility[62], or is a marker for social factors that influence family size. In the Amish population, one social factor that may influence family size is church district, because substantial genetic substructure exists based on geography. Additionally, carrier couples had fewer instance of miscarriage but no change in the number of pregnancies or number of children. This result could indicate that pregnancies are under-reported.

Previous studies in populations with consanguineous unions have quantified the increased risk of producing offspring with recessive disorders. The PROMIS study recruited Pakistani participants, in which the median coefficient of inbreeding was fourfold higher than in outbred populations.[45] Twice as many variants were identified that cause embryonically lethal recessive disorders than those that result in fetal or infant death. A typical individual from this population was estimated to carry on average 1.6 recessive loss of function (LOF) lethal-equivalent variants in the heterozygous state, but only about one-third of these resulted in miscarriage, stillbirth, or infant mortality.[45] Therefore, it is very important to identify risk loci in these populations and quantify the risk associated with adverse reproductive outcomes. Quantifying the impact that carrier screening would provide is an important step in planning future genetic testing and analysis of reproductive outcomes associated with carrier couples will help to describe the scope of the problem faced by such couples.

Our study evaluated reproductive outcome risks to carrier couples retrospectively, in some cases after childbearing was completed. For this reason, carrier status was not reported to couples in this study. Although Amish couples do not currently use genetic information to make reproductive decisions, they are increasingly receptive to learning their carrier status for recessive conditions using genetic testing, as by the PIP. Couples may use this knowledge to inform themselves and their health care providers of the potential need for emergency care after the birth of a baby. Our study adds additional data to support the efforts of the Clinic for Special Children and other clinics who care for the Amish, to strategically allocate pediatric resources to children and families needing the most support.

This study provides insight into potential reproductive outcomes of highly related parents. Although consanguinity is most commonly associated with recessive conditions, many of which are inborn errors of metabolism, consanguinity is also associated with an increased incidence of later onset disorders including diabetes, cardiovascular disorders[59], obesity, and some cancers.[61,62] These conditions may affect pregnancy and reproductive outcomes[78] and have an adverse effect on fertility through an increased rate of miscarriage, infant mortality, and morbidity. Therefore, future studies should assess the effect of consanguinity on complex traits in the Amish population. The etiology of stillbirth, miscarriage, and other reproductive outcomes for consanguineous unions should be further investigated to determine preventative strategies to identify which variants are driving these associations and through what mechanism.

Summary

We have assessed reproductive outcomes in couples who are both carriers for a P/LP variant that causes a recessive, highly penetrant condition. We identified an association in which carrier couples have less miscarriages. We also have considered the genetic relationship within couples, the R coefficient, and identified a positive correlation with three reproductive outcomes. Highly related couples had increased number of pregnancies, increased number of children, and increased instance of stillbirths.

VI. Assessing the impact of individual autozygosity on complex traits[3]*

Abstract

Autozygosity, the proportion of the genome that is homozygous by descent, has been associated with variation in traits of biomedical importance impacting evolutionary fitness in some populations. Autozygosity (F_{ROH}) is typically measured from runs of homozygosity (ROHs) that arise when identical-by-descent (IBD) haplotypes are inherited from each parent. Population isolates with a small set of common founders have elevated autozygosity relative to outbred populations. In this study, we examined whether the degree of autozygosity was associated with variation in 96 complex traits among 7221 Old Order Amish individuals residing in Lancaster County, PA. The Amish are a relatively recent founder population, having emigrated from Europe to Lancaster 14-15 generations ago. In this population, the average length of an ROH segment is 6350 KB and the average number of segments 1.5 KB or larger per individual is 20, which in aggregate span ~3.7% of the genome. Measurements of genome-wide and regional F_{ROH} were used as the primary predictors of trait variation in association analysis. We analyzed 96 traits including basic anthropometrics, blood pressure, fasting blood lipids, glucose, insulin, HbA1c, basic medical blood chemistry measurements, and medical histories. We did not identify any associations that withstood Bonferroni-correction ($p = 0.0005$), but the lead association ($p = 0.0036$) for genome-wide F_{ROH} was with longer EKG QT interval. Regional F_{ROH} estimation revealed two significant trait associations after Bonferroni-correction. Serum bilirubin levels were significantly associated ($p = 1x10^{-43}$)

[3]Megan T. Lynch, Kristin A. Maloney, Huichun Xu , Regeneron Genetics Center,
Alan R. Shuldiner, Braxton D. Mitchell. "Assessing Individual Autozygosity by Mapping Trait Associations". In preparation for submission.
*MTL performed statistical analysis, drafted the manuscript, and generated tables and figures.

with a region on chromosome two localized to a region surrounding the *UGT1A10* gene and HbA1c levels were significantly associated ($p = 8x10^{-10}$) with a region on chromosome eight localized to a region that includes the *CHRNB3* gene. We also observed an association between autozygosity at *APOB* and LDL-cholesterol. This association was markedly diminished after adjusting for a well-studied LDL-C associated *APOB* variant, p.Arg3527Gln.

Introduction

Autozygosity, which is defined as the probability that a region is homozygous due to the inheritance of alleles identical-by-descent (IBD), is determined by the presence of extended homozygosity in that region. Since runs of homozygosity (ROHs) are a recognized signature of recessive inheritance, homozygosity mapping using ROH has commonly been used to map recessive disorders when there is suspicion that parents may share a common haplotype.[71]

More recently, studies have evaluated the impact of overall autozygosity on variation in common complex traits, where overall autozygosity is estimated as the proportion of ROH across the entire genome (genome-wide autozygosity). Genetic homogeneity, estimated as overall autozygosity, has previously been linked to adverse health outcomes in multiple traits impacting evolutionary fitness. Populations with increased autozygosity are more likely to experience inbreeding depression, or reduced fitness, which has been linked to a range of phenotypic consequences including cardiovascular disease,[59,72] shorter stature,[64,65] lower general cognitive ability,[64] decreased fertility,[55,56] and higher hip-to-waist ratio.[60] In case-control study designs involving outbred populations, higher genome-wide autozygosity has also been associated with coronary artery disease[72] and amyotrophic lateral sclerosis.[73] These studies raise the potential for employing ROH mapping to identify hotspots along the genome that contain multiple rare, recessive loci that influence health and disease. Autozygosity mapping may be a particularly powerful tool when applied to founder populations where consanguinity is high.[70]

Elevated autozygosity measurements are seen in population isolates with a small set of common founders.[47] These populations often have an increased burden of recessive disorders that are rare in the general population and recent inbreeding magnifies the occurrence of mildly deleterious variants.[133] The Amish of Lancaster County, PA are a relatively recent founder population who emigrated from Europe to Lancaster 14-15 generations ago. In this study, we examined whether the degree of autozygosity was associated with variation in among 96 different complex traits among 7221 Old Order Amish individuals with genotyping residing in Lancaster County, PA. In these subjects we estimated genome-wide levels of autozygosity as the proportion of the autosomal genome in runs of homozygosity > 1.5 Mb. We also estimated the probability of autozygosity at 10 Kb average intervals throughout the genome. We used these measures in association analysis to assess evidence for association of genome-wide and locus-specific association of autozygosity with 96 different phenotypes.

Methods

Participants:

Subjects in this study were recruited through multiple protocols carried out between 2003 and 2019 as part of the Amish Research Program through the University of Maryland Amish Research Clinic (ARC) in Lancaster, PA. These studies were open to volunteers throughout the Lancaster Amish community, and were designed to assess primarily cardio-metabolic and bone health.[89,130] Recruitment for most protocols was phenotype agnostic; that is, subjects were not recruited for particular diseases or health conditions. This report is based on 7,220 apparently healthy Amish individuals 18 years of age or older recruited from the community and in whom we obtained genotyping data.

Phenotyping:

Clinical examinations were performed by trained nursing staff at the Amish Research Clinic in Lancaster, PA or in the homes of study participants. For this report, we restricted analysis to a set of phenotypes measured in common across all (or nearly all) protocols. These include basic anthropometrics, blood pressure, fasting blood lipids, glucose and insulin, hemoglobin a1c (glycated hemoglobin), basic blood chemistries, and medical histories. A complete list of the 96 phenotypes analyzed is provided in Appendix 5.1.

Genotyping:

Genotyping was performed at the Regeneron Genetics Center using the Illumina Global Screening Array, which included 490k single nucleotide polymorphisms (SNPs) that passed quality control parameters. Variants and subjects with >10% missingness were removed, a Mendel error threshold was applied, and a 1^{-10} minor allele frequency threshold was used. Genotype data was not imputed.

Genome-wide autozygosity estimation:

Genome-wide autozygosity was estimated as the proportion of the genome that fell into ROHs. SNPs with more than 3% missingness across individuals and with a minor allele frequency less than 5% were excluded from ROH calculations. ROH estimates were made using the IBD tool implemented in Plink which has several built-in arguments.[134] The key parameters to identify truly autozygous segments are minimum length (kilobase, Kb) needed for a tract to qualify as homozygous, number of contiguous homozygous SNPs, and minimum tract density requirements. To avoid very short and common strands of homozygosity that occur throughout the genome due to linkage disequilibrium (LD), we used a minimal length of ROH of 1.5 Mb[64,135]. The fraction of each autosomal genome in ROH > 1.5 Mb correlates well with pedigree-based estimates of inbreeding[135]. All other parameters were the default parameters set by Plink except for a decrease in the number of contiguous homozygous SNPs to 50, consistent with previous studies[47,64]. We used the PLINK default settings to allow each autozygous segment to include up to five missing SNPs and up to one heterozygous SNP. Since study

individuals were genotyped with the same assay and processed with the same QC, ROH was defined as the total kilobase (Kb) included in a ROH.

Regional autozygosity estimation:

Regional autozygosity computation was performed using the *GARLIC* software which outputs ROH and length class information in UCSC's plain-text BED format.[136] Autozygosity was estimated at 10 kb average intervals using genotype error rate and population-specific allele frequencies and a window size of 50 SNPs. To format individual level predictor files for statistical analysis, a python script was created to extract ROH inclusion status for each SNP on the genotyping array.

Statistical analysis:

We tested for association of the autozygosity estimates (genome-wide and regional) with trait variation using a linear mixed model with phenotype as the outcome and autozygosity as the independent variable. As fixed covariates for each trait analysis, we included age, sex, and two high effect size SNPs (*APOB* p.Arg3527Gln and *APOC3* p.Arg19Ter) previously identified in this Amish community and as a random effect to account for covariation among related individuals we used a relatedness matrix. The analysis was performed using our in-house MMAP software and the polygenic component was modeled using a genomic relationship matrix (GRM) constructed by the MMAP program using the Amish Illumina Global Screening Array genotypes.[131] Genome wide or regional F_{ROH} was used as the primary predictor of the traits of interest. The genome wide significance threshold after Bonferroni correction was $p = 0.0005$. As a

secondary analysis, we tested for sex-specific effects of autozygosity on trait variation using an autozygosity*sex interaction term. For regional analysis, an independent association analysis was performed for each genomic location using ROH status as the primary predictor of the 96 traits of interest with Bonferroni adjusted significance threshold of $p = 1.7x10^{-9}$.

Results

ROH across the genome:

The mean ROH segment length was 6350 KB among the 7220 Amish participants. On average, each participant had 20 segments of >1.5 Mb, which in aggregate spanned 110 Mb which is ~3.8% of the genome (Figure 5.1). ROH were present on each autosome and were widely distributed across the genome. We assessed the genome-wide distribution of ROHs by the frequency of entire ROH segments and by the frequency in which SNPs are included in an ROH segment.

Across the genome, shorter ROH segments were more common, especially in two instances where short segments had notably high frequency (Figure 5.2). The highest frequency segment is located at chr1: [145927328: 148353534] and is 2.4 Mb long, occurring in 184 individuals. The segment with the second highest frequency, occurring in 173 individuals, is located at chr10: [42113412: 46332633], and is 4.2 Mb long.

Figure 5.1. Histograms of ROH segment length and number. Length of ROH segments (tops) and number of ROH segments (bottom).

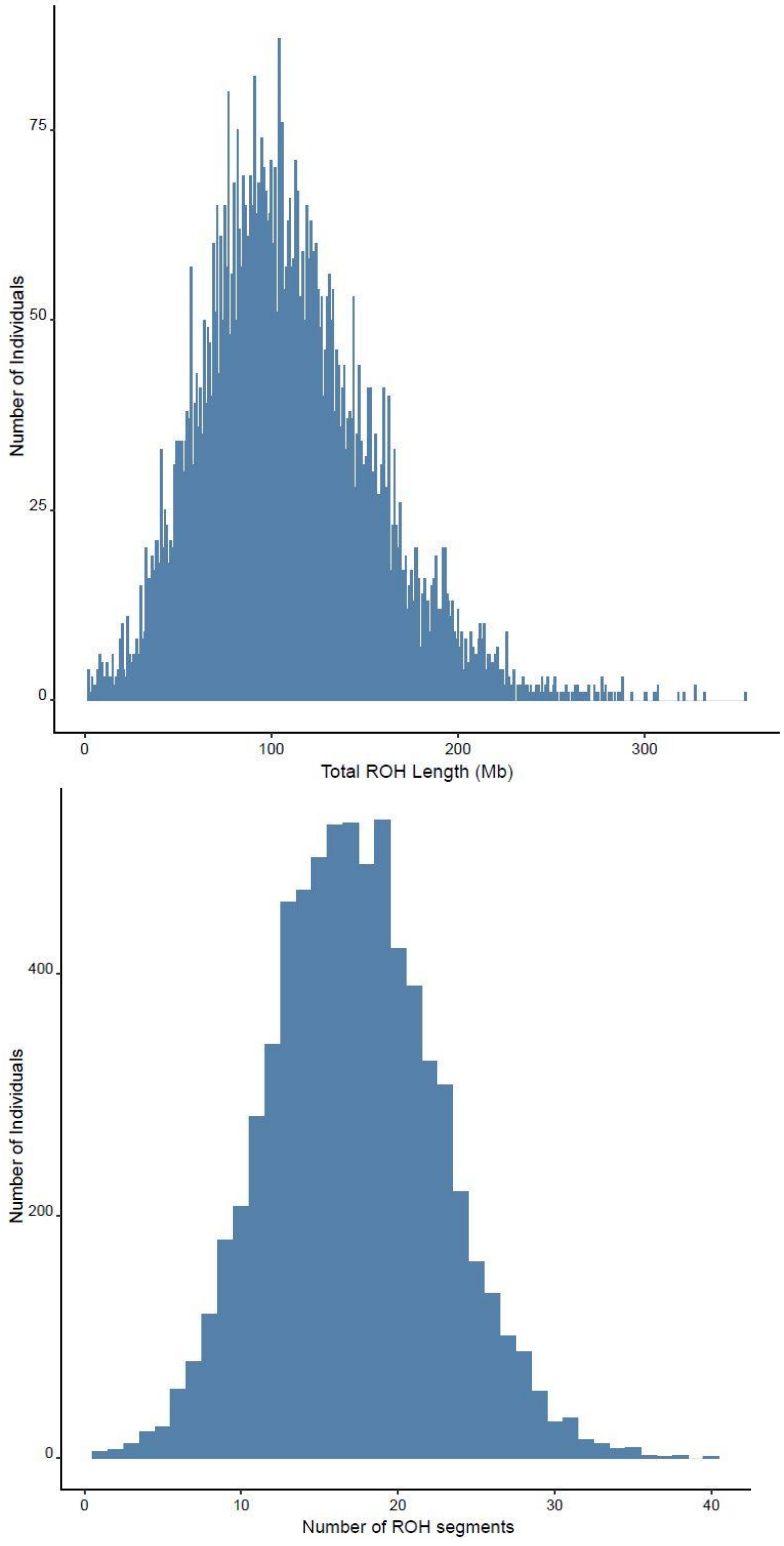
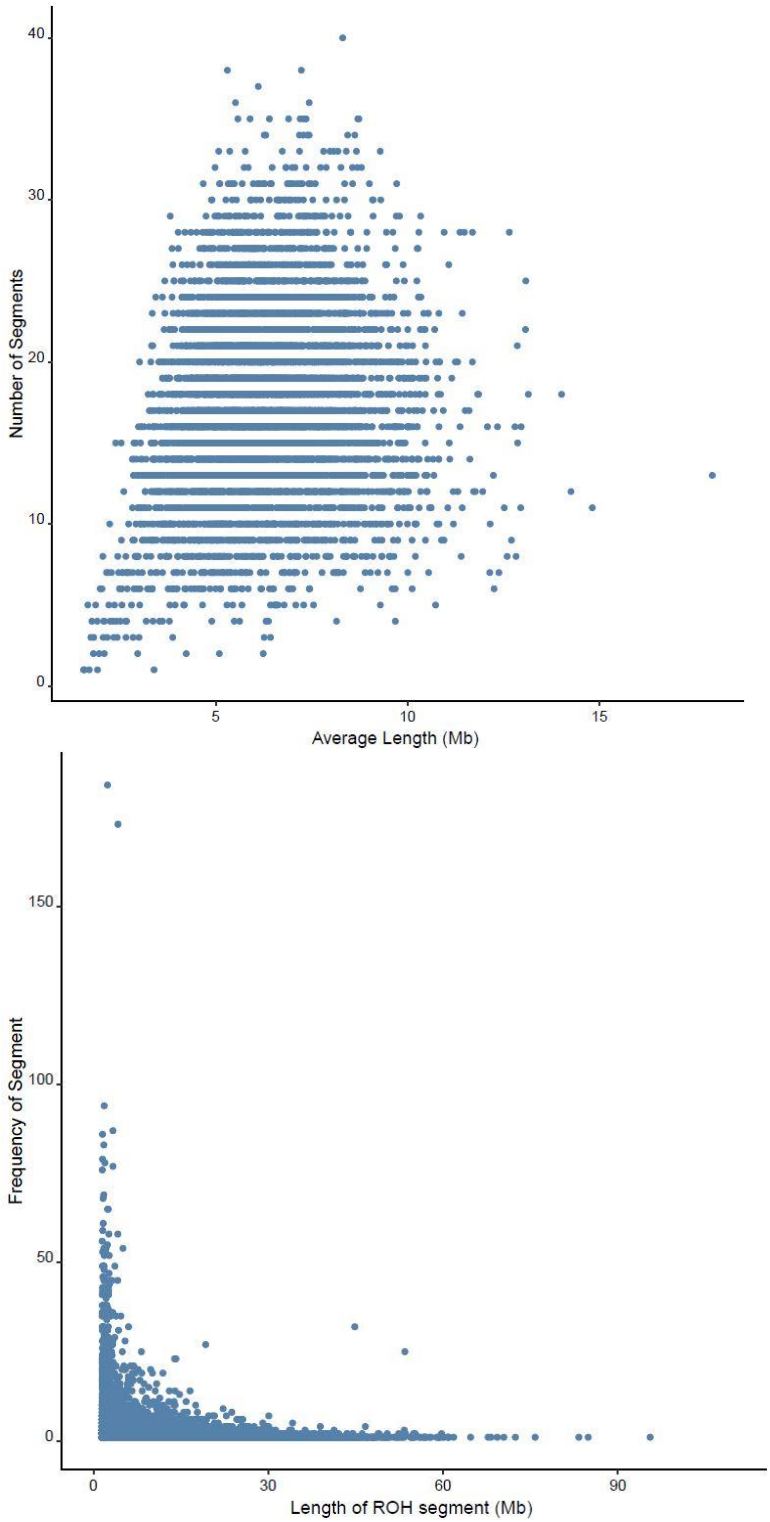
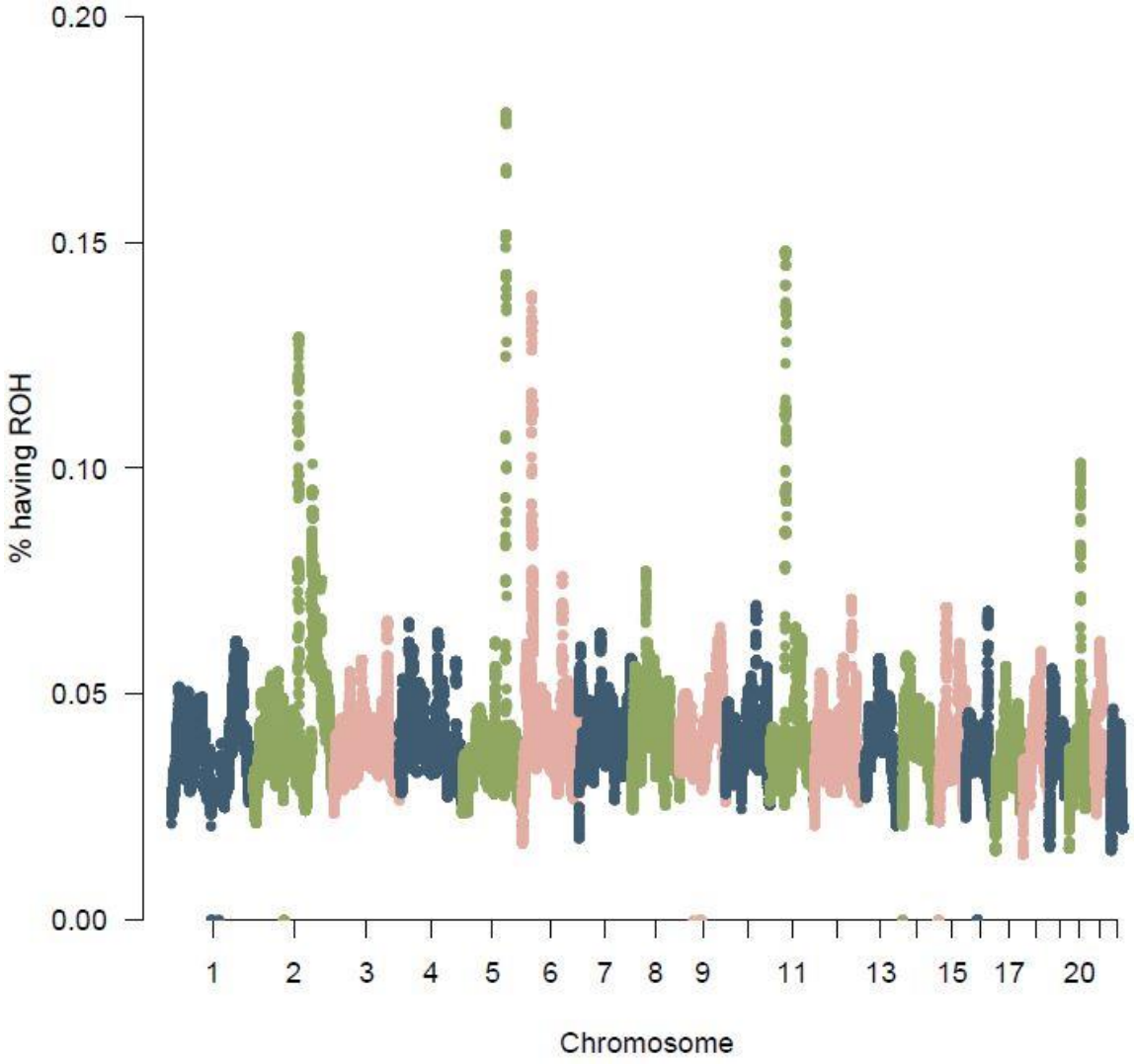


Figure 5.2. Size distribution of ROH segments. Average number and average length of autozygous segments in individuals (top) and frequency of autozygous segments by length (bottom).



There are several regions with a notable increase in the frequency of individuals with SNPs in ROH segments (chr 2: [134600135:136231566]; chr 2: [176901247: 178324285]; chr 5: [130334251:132873169]; chr 6: [27512529: 29306571]; chr 11: [49758165:47253513]; chr 20: [34813904:35819776]) (Figure 5.3, Appendix 5.2). The highest regional ROH frequency for a SNP was 18% which encompasses a region on chromosome 5 containing the genes *RAPGEF6*, *FNIP1*, and *ACSL6*. This region has previously been identified as having enriched ROH frequency for SNPs.[137] In chromosome 2, the highest ROH frequencies were found in *LCT*, known to have recent positive selection in European ancestry individuals[138], and two genes in close proximity; *UBXN4* and *R3HDM1* which have not previously been noted in ROH studies. Our finding that *FOLH1* and *OR4A47*, located in chromosome 11, both had SNPs with increased ROH frequency, is consistent with previous findings.[137]

Figure 5.3. Frequency of SNPs in ROH segments across the autosome.



Almost all SNPs fell into an ROH at least once, except in six instances across the genome in which small clusters of SNPs did not belong in any ROH segment (Figure 5.3). This may indicate that diversity is favored and ROHs are not well tolerated in these regions.

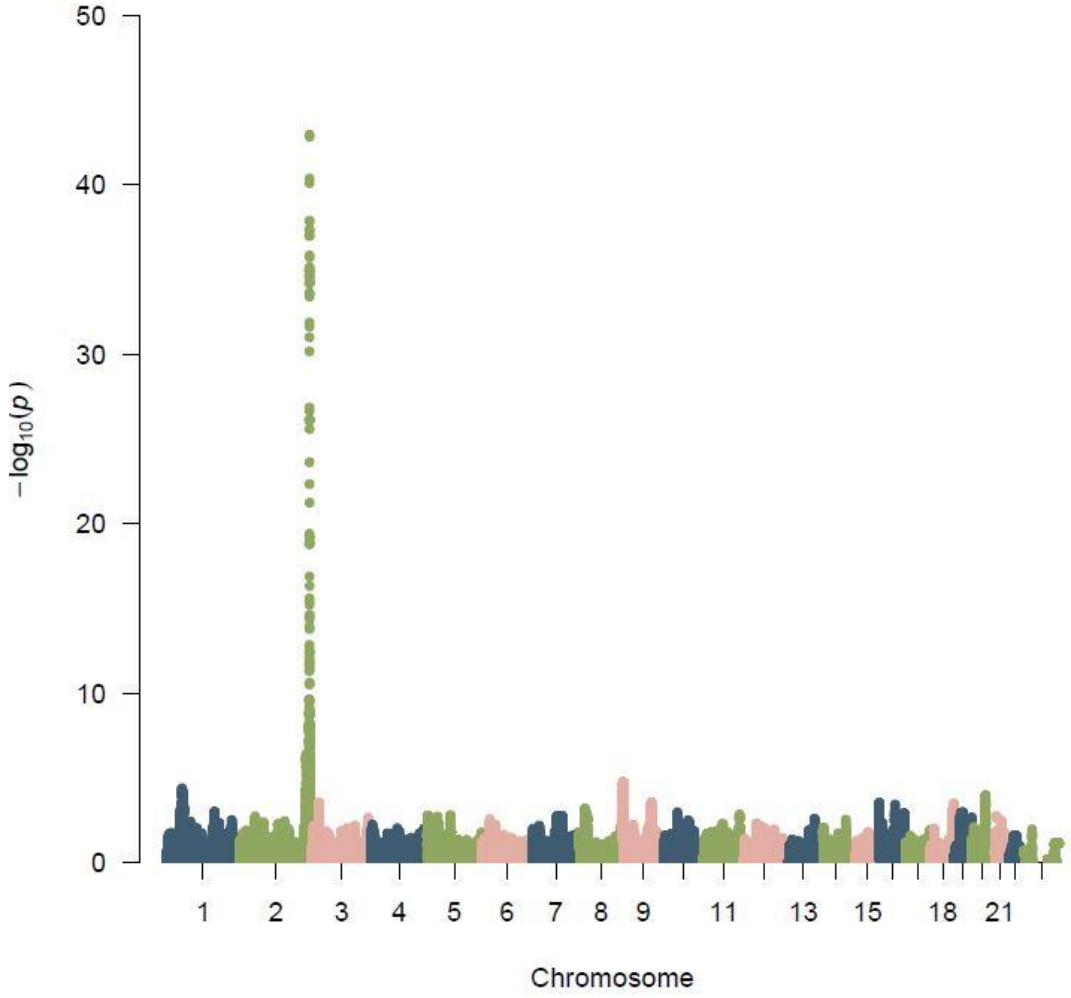
Association of genome-wide ROH with health and disease-related traits:

Measurements of F_{ROH} were then used as the primary predictor of phenotypic variation. We analyzed 96 traits including basic anthropometrics, blood pressure, fasting blood lipids, glucose, insulin, HbA1c, basic medical blood chemistry measurements, and medical histories. We did not identify any associations that withstood Bonferroni-correction ($p < .0005$), but the lead association ($p = 0.0036$) with genome-wide F_{ROH} was with EKG QT interval, followed by Serum CO_2 level ($p = 0.03$), urea nitrogen level ($p = 0.04$), and serum thyroid hormone measurement (tsh) ($p = 0.04$). In a sex-stratified analysis, we did not identify any sex-specific effects of $F_{ROH.}$

Association of regional autozygosity with health and disease-related traits:

We analyzed 96 traits for associations with regional autozygosity levels and identified two traits associations at genome-wide significance following Bonferroni correction ($p = 1.7x10^{-9}$) (Figure 5.4a, Figure 5.4b). Increased serum bilirubin levels were significantly associated ($p = 1x10^{-43}$) with increased F_{ROH} at a region of chromosome two that includes the *UGT1A10* gene. Increased levels of HbA1c were significantly associated ($p = 8x10^{-10}$) with increased F_{ROH} at a region on chromosome eight surrounding the *CHRNB3* gene.

Figure 5.4a. Autozygosity mapping Manhattan plots. Serum bilirubin levels (top) and HbA1c levels (bottom)



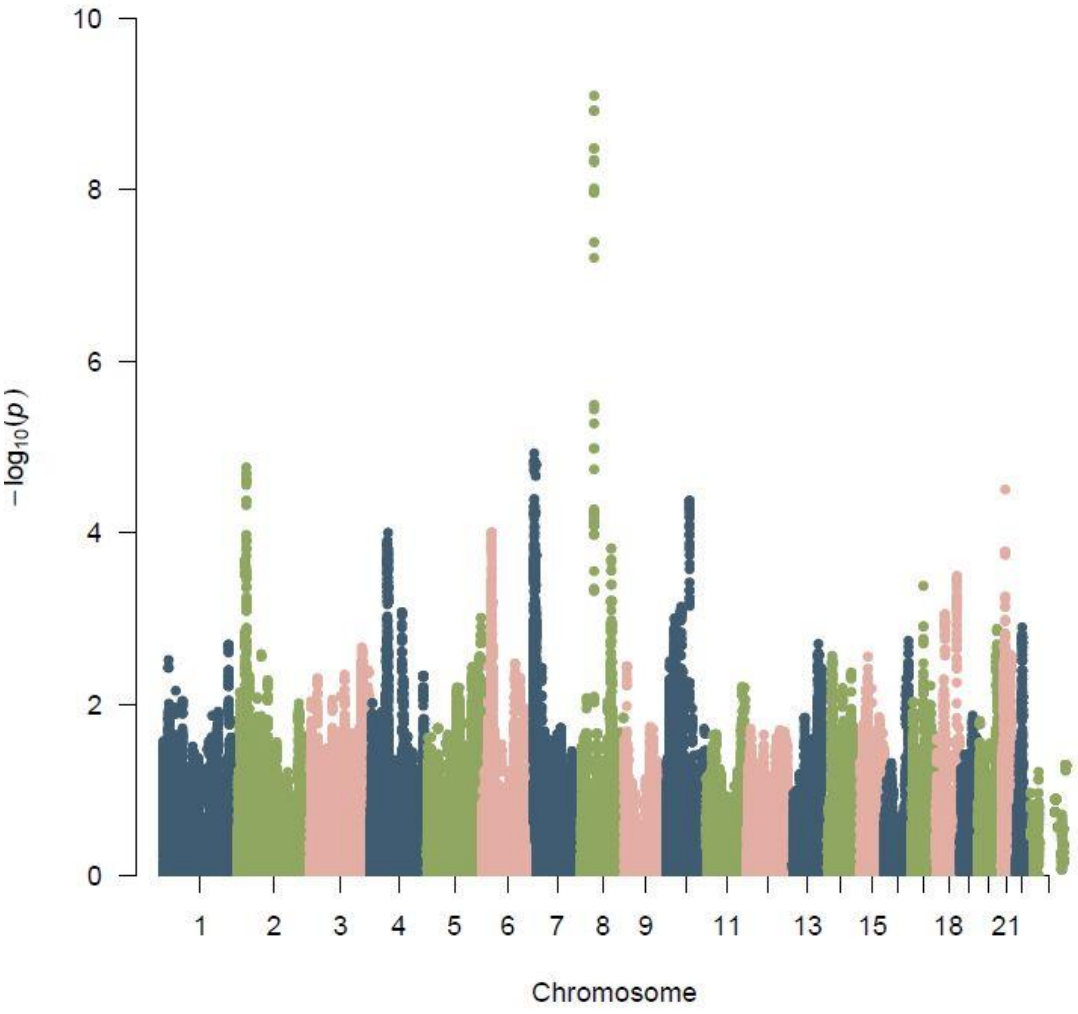
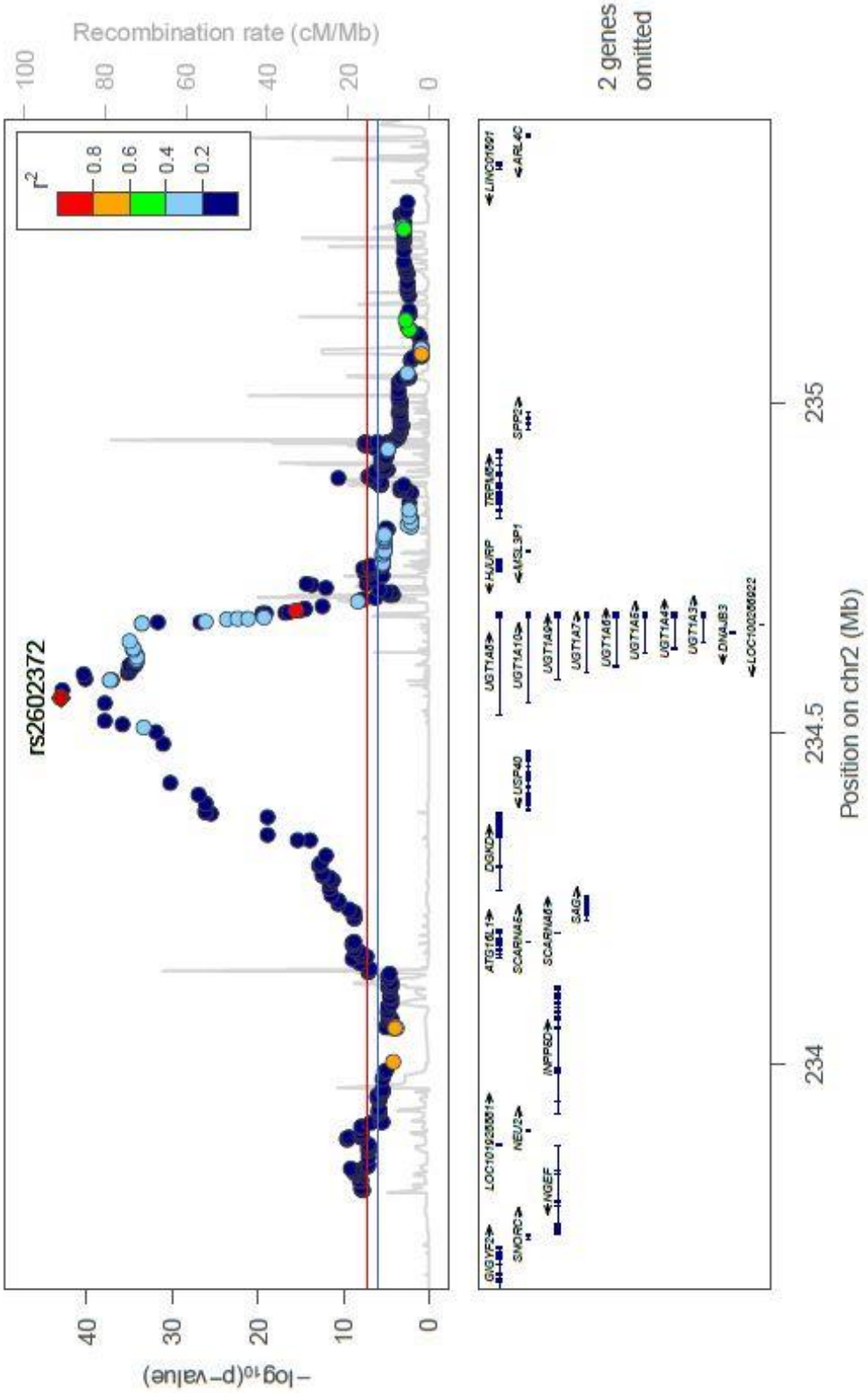
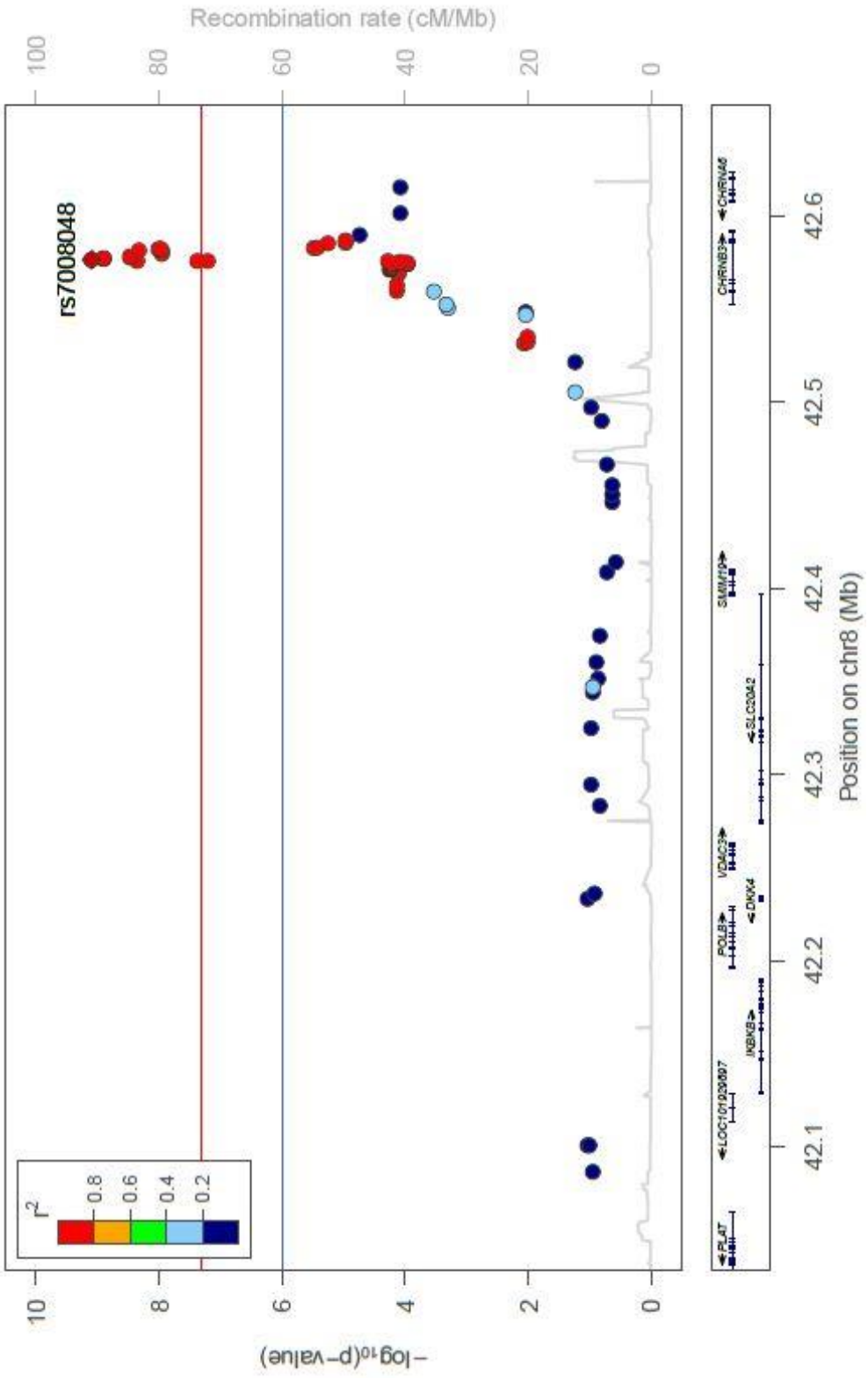

10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
11
13
15
18
21
Chromosome

Figure 5.4b. Autozygosity mapping LocusZoom plots. Serum bilirubin levels (top) and HbA1c levels (bottom). Red line indicates P < 5 x 10-8. Blue line indicates P < 1 x 10-6.




Recombination rate (cM/Mb)
100
80
60
40
20
0
rs7008048
r²
0.8
0.6
0.4
0.2
10
8
6
4
2
0
−log10(p−value)
PLAT
LOC101929697
IKBKB
POLB
DKK4
VDAC3
SLC22A2
SMARCD3
CHRNB3
CHRNA6
42.1
42.2
42.3
42.4
42.5
42.6
Position on chr8 (Mb)

Association of ROH with ApoB region:

We also observed an association between an autozygous segment on chromosome 2 and LDL-cholesterol levels (Figure 5.5a). This autozygous segment included the APOB gene, for which the p.Arg3527Gln variant (rs5742904) has previously been associated with LDL-C and total cholesterol levels.[27] We tested if this association could in fact be attributed to the p. Arg3527Gln variant by adjusting for this variant as a fixed effect in the analysis and the association was markedly diminished (Figure 5.5a). Individuals with no copies and with two copies of the variant had an enrichment of autozygosity within the *APOB* gene compared to heterozygotes (Figure 5.5b). Interestingly, the association was not diminished compared to the initial result when homozygous carriers of the variant were removed from analysis. Therefore, it is the heterozygous carriers with lower autozygosity measurements that drive the negative association between lower autozygosity and increased cholesterol levels.

Figure 5.5a. Manhattan plot of regional autozygosity association analysis of chromosome 2 with cholesterol. Without control for the *APOB* p. Arg3527Gln variant (top) and after controlling for carrier status (bottom).

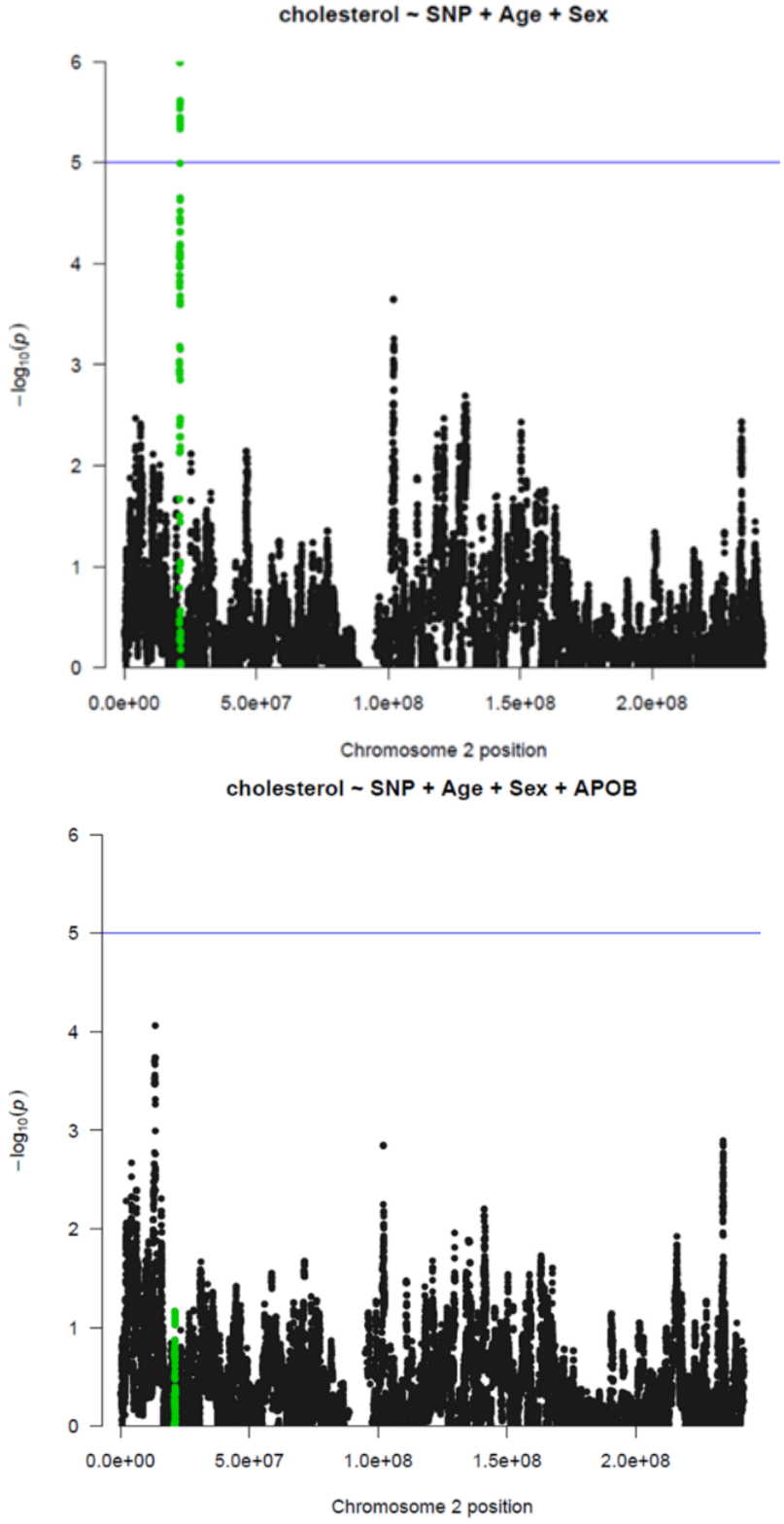
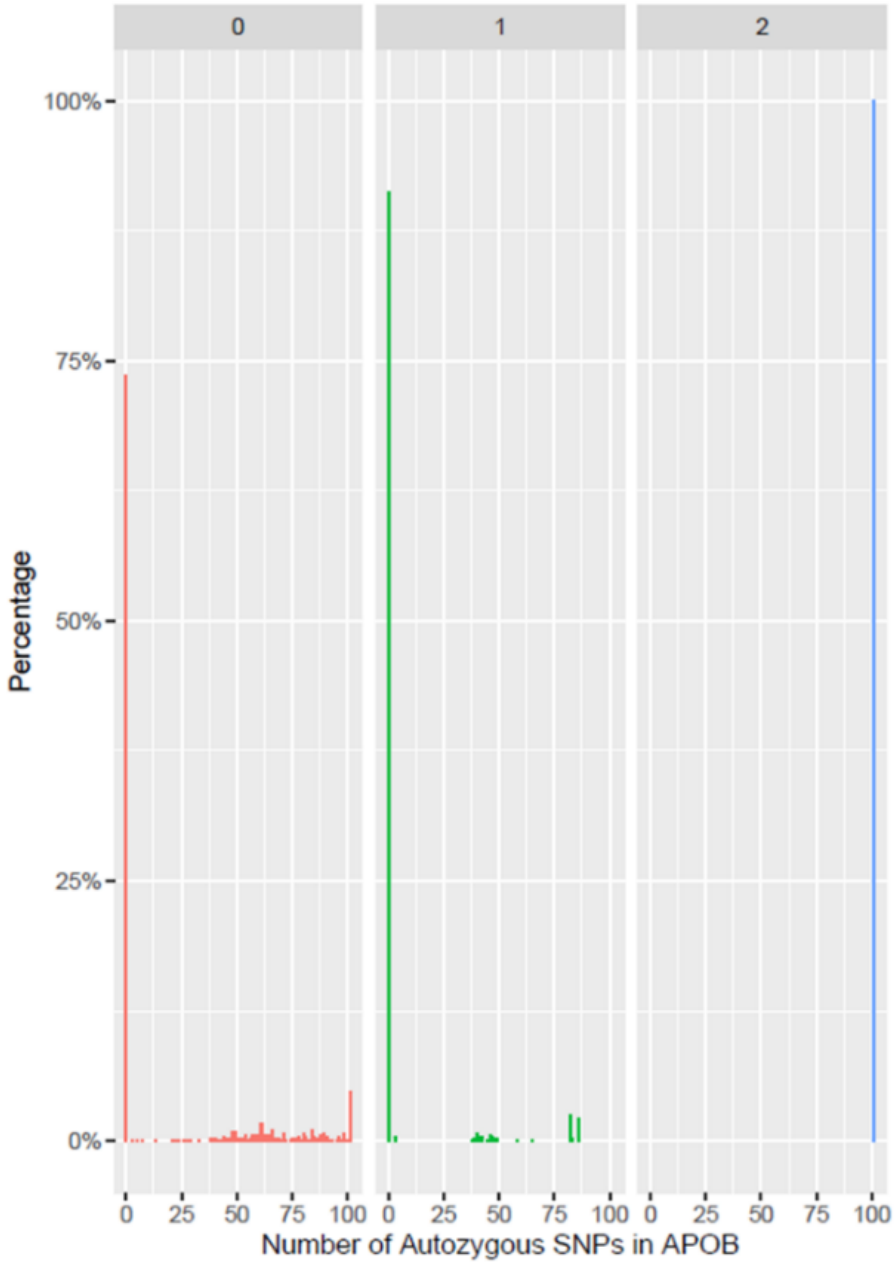


Figure 5.5b. SNPs in the *APOB* gene determined to be within autozygous segments separated by *APOB* p. Arg3527Gln carrier status.



Discussion

ROH across the genome:

The average Amish individual had autozygous segments spanning ~3.7% of their genome. In comparison, UK biobank individuals with European ancestry had ~0.4%.[47] Amish participants harbored 19.5 ROH segments on average with an average length of 5700 Kb. In contrast, non-founder European populations harbored 8.02 segments on average with an average length of 1421 Kb.[139] The observation that ROHs are longer and more numerous in this founder population compared to outbred populations is not surprising given that longer shared haplotypes are inherited from recent common ancestors and reduced effective population size, due to a population bottlenecks, increases the number of ROH present.[57]

ROHs were widely distributed across the genome and were present on each autosome (Figure 5.3). The region with the highest frequency of SNPs in an ROH was identified in chromosome 5. The *LCT* gene in chromosome 2 had an enrichment of frequency of ROH which is consistent with previous studies.[138] Although an enrichment was also seen on chromosome 6, it was not localized to the MHC locus as seen previously.[66,139] Across the genome, there were just six instances in which small clusters of SNPs were not included in any ROH segments. Five of these regions contain pseudogenes and noncoding RNA (Appendix 5.3). This may indicate that diversity is well tolerated, and the regions are not under selective pressure. The sixth instance is along chromosome 14, in a region containing the gene *OR11H12*. This gene is a member of the olfactory receptor proteins and is known to have copy number variations.[140]

Association of genome-wide ROH with traits of interest:

We have tested for associations between autozygosity, as measured by ROH, and a large panel of complex traits in a founder population of Amish individuals. We did not identify associations of genome-wide autozygosity with any traits that withstood Bonferroni-correction. The lead association was between higher autozygosity measurements and QTc interval. Because the Amish study population is enriched for a known, high effect size, *KCNQ1* pathogenic variant, p.Thr244Met, that increases QTc intervals,[28] we removed the *KCNQ1* genomic region from the analysis and re-assessed the association. The association, although not statistically significant after Bonferroni correction, was still present and did not diminish.

Association of regional autozygosity with traits of interest:

Previously, ROH mapping has been used to detect loci with a segregating recessive variant. This method is regarded as more powerful than traditional single-marker association studies using a recessive model because there is more certainty of the haplotype on which the two alleles appear and greater viability at extreme minor allele frequencies.[73] *GARLIC* and other ROH prediction tools have been used for case-control studies and whole genome F_{ROH} associations with complex phenotypes. Here, we implemented *GARLIC* to map regional associations for complex trait discovery.

We detected associations of increased F_{ROH} with higher levels of serum bilirubin levels at a region on chromosome 2 with increased HbA1c at a region on chromosome 8. The mapping analysis also identified an association with cholesterol clustering around the *APOB* region on chromosome 2 which contains a high effect size variant

p.Arg3527Gln.[27] This association was between increased LDL-C levels and decreased autozygosity. Therefore, heterozygotes in this region are driving the association.

Strengths and limitations

Population isolates have an increased burden of ROH and can uncover genomic regions associated with complex phenotypes that may not have been identified in populations with more distant parental relatedness. Due to low levels of genome-wide homozygosity commonly seen in modern human populations, very large numbers of study subjects are required to provide sufficient statistical power.[135] Although no complex traits in the present study were associated with genome-wide autozygosity levels, this could potentially be because many examined phenotypes were cardio-metabolic which are risk factors for late-onset conditions and may not be under evolutionary selective pressure.[64] Also, genome-wide autozygosity measurements may not predict some polygenic traits influenced by dominant alleles at different loci, each influencing the trait in opposite directions. To identify genomic regions with dominant or additive alleles driving trait associations, GWAS are a better method than regional autozygosity mapping. For example, GWAS identified a statistically significant association between *UGT1A10* and bilirubin levels ($p= 2x10^{-38}$ for *UGT1A10* rs17854828 in 5830 subjects), although an association between HbA1c levels and chromosome 8 has not previously been reported.

Summary

Amish have larger ROH segments and more of the Amish genome included in autozygous regions compared to outbred populations. Genome-wide summed autozygosity was not significantly associated with any of the 96 traits tested in this study. By regional autozygosity mapping methods, we identified two traits associated with regional levels of autozygosity. We found that increased serum bilirubin levels were associated with increased autozygosity on chromosome two, localized to the *UGT1A10* gene and that increased HbA1c levels were associated with increased autozygosity on chromosome eight, localized to the *CHRNB3* gene.

VII. Discussion

Conclusions

The three aims that comprise this thesis project focus on the impact of the genetic background of a founder population, leveraging the genetic and environmental homogeneity of the Amish population to make new discoveries regarding genetic predictors of trait outcomes. The major areas of interest are genetic predisposition to complex disease, genetic risk factors for reproductive outcomes including pathogenic variant carrier status and relatedness between couples, and the impact that increased levels of homogeneity in individual genomes, measured by autozygosity, may have on complex traits.

In the first aim we addressed the enrichment of variants associated with complex disorders, specifically focusing on medically actionable variants. Seven P/LP variants were identified, each associated with a gene and disorder on the ACMG or Geisinger panel of genes recommended for return of results when identified during a clinical test. These findings are evidence of the enrichment of clinically actionable variants in the Amish populations compared to the frequency occurring in European non-founder groups. Among the seven identified pathogenic variants, two were previously well-studied and were responsible for a large portion of individuals who harbored one of two clinically actionable variants; the *APOB* p.Arg3527Gln variant and the *KCNQ1* p.Thr244Met variant. Overall, we found that ~14.7% of all Amish harbor at least one pathogenic variant within seven clinically actionable genes, compared to frequencies of 2.0% to 6.2% that have been previously reported by large studies in non-founder

populations. This is particularly important when considering the prevalence of seemingly healthy individuals who harbor one of these variants and could benefit from genetic screening and medical interventions. Populations in which there are known enriched variants may benefit from the creation of customized screening panels that cater to their unique genetic background.

In the second aim, we identified couple pairs who were carriers for pathogenic recessive variants to determine whether carrier status influenced reproductive health outcomes. We found that couple pairs who were both carriers for pathogenic variants in the same gene were at an increased risk of reporting a stillbirth. However, carrier status did not have an impact on family size measured by pregnancies and number of children. Carrier status also did not have an impact on miscarriages, which may indicate that the pathogenic recessive variants in this study have detrimental effects either to late-stage pregnancies, resulting in a stillbirth, or, prior to a recognized pregnancy.

Next, we sought to distinguish whether genetic relatedness of couple pairs across the entire genome, rather than the sharing of a particular pathogenic allele, would influence reproductive outcomes. We found a positive association between couples with a high coefficient of relationship and higher number of pregnancies, higher number of children, and increased instance of stillbirths. Again, we did not identify an association with miscarriages, but highly related couples could still share embryonic lethal variants which have a detrimental impact prior to recognized pregnancy. These results may indicate a mechanism in which highly related couples have more pregnancies to compensate for early pregnancy loss or, a yet to be identified social factor is influencing family size in these couples.

In the final aim, we evaluated the associations of autozygosity with a large panel of biomedically important traits. We assessed genomic homozygosity, measured as a sum of genome-wide autozygosity and measured regionally within each 10 Kb segment throughout the genome. Genome-wide levels of autozygosity were not significantly associated with any traits.

Evaluation of regional autozygosity revealed that almost all SNPs fell into an ROH at least once, except in six instances across the genome. These regions may have a functional importance that does not tolerate homozygosity. The mapping analysis also identified two trait associations. The first was an association that was highly statistically significant after Bonferroni correction. Increased serum bilirubin levels were associated with increased levels of autozygosity in a region of chromosome 2 that surrounds the *UGT1A10* gene. The pharmacogenetic locus UGT1A1*28 is known to elevate bilirubin levels. Subjects from the Amish population used in analysis are known carriers of this variant (frequency ~ 0.43).[141] Accounting for this locus in conditional analysis did not diminish the regional autozygosity association. The second identified statistically significant association was between increased HbA1c levels and increased levels of autozygosity in a region of chromosome eight surrounding the *CHRNB3* gene.

These two findings suggest that autozygosity may not be well tolerated in these regions. High levels of serum bilirubin levels may indicate different types of liver problems and *UGT1A10* codes a protein that is one of several UDP-glucuronosyltransferases expressed in the liver that are responsible for metabolizing bilirubin.[142,143] HbA1c levels are associated with blood sugar and elevated levels are diagnostic of diabetes. In a meta-analysis GWAS study with a large cohort of >150,000

individuals analyzing HbA1c levels, no association was seen in the *CHRNB3* region identified here.[144]

The mapping analysis also identified an association with cholesterol clustering around the *APOB* region on chromosome 2. This association was diminished after controlling for the *APOB* p.Arg3527Gln variant that increases cholesterol levels. This result suggests that heterozygous carriers with decreased autozygosity were driving the association.

Ethical Considerations

It is important to be culturally appropriate and sensitive of the Amish community's lifestyle and beliefs. There are several important considerations that arise regarding cultural differences. The first consideration is to maintain appropriate terminology and cultural competency which will help to increase trust between the Amish community and research institution.[10] This includes an understanding that Amish culture is diverse and should not be grouped into a stereotypical image.[10] The second consideration is to maintain sensitivity to the Amish community's belief system regarding modern healthcare. This aspect is an especially important consideration when examining the potential return of genetic findings. The Amish value health and believe that it can be achieved through working hard, living simply and cleanly, and eating a well-balanced diet.[12] Amish individuals rely on the opinion of friends and family when making important health decisions and will seek professional care in some circumstances.[10] They also face additional barriers including affordability issues for those without insurance and lower vaccination rates.[10]

Limitations

Many of the variants in clinically actionable genes that were identified by the analysis of the first sub-study were assigned a VUS classification. Although in depth phenotyping has been performed in Amish individuals, many of the clinically actionable diseases had phenotypes that were not included such as cancer predisposition, malignant hypothermia, and some inborn errors of metabolism. Additional phenotyping data would have allowed for segregation analysis to find genotype-phenotype correlations. While it would have been beneficial if additional evidence was available, it is expected in variant annotation studies that most variants are designated VUS. Also, our filtering procedure eliminated singletons from analysis if they did not have previous ClinVar annotations. This was done to prioritize variants for classification but potentially pathogenic de novo mutations may have been excluded. We also did not include synonymous variations in analysis and CNVs were not available in the sequence data.

In sub-study two, we relied on questionnaire data from participants. Inaccurate information could have harmed the integrity of the study and missing data eliminated couples from being including in the analysis, decreasing sample size. Although we analyzed confounder variables including church district, age of marriage, and age of menopause, additional confounding variables could have been present. If present, they could have caused a bias in the analysis. Additionally, some genes with pathogenic recessive variants in which carrier status was determined, had very few carriers each. This did not inhibit the major analysis presented in sub-study 2 but limited the range of analysis that could be performed on the data. To identify specific alleles that may be associated with each of the four reproductive outcomes (i.e. stillbirths, number of

children, number of pregnancies, and number of miscarriages), a larger sample size would be necessary. Sample size could be increased by further exome sequencing, or, targeting individuals in whom their spouse has already been sequenced to gain an additional couple pair.

In the sub-study 3 analysis, PLINK software was used to estimate F_{ROH}. To find an accurate sum of F_{ROH} genome wide, it was crucial to identify truly autozygous segments while still maintaining genome coverage. This required the specification of various parameters in Plink software and was guided by literature evidence. However, regional analysis presented a unique challenge because there was no previously outlined computational method to identify highly autozygous segments across the genome. To circumvent this obstacle, Plink and *GARLIC* output files were manipulated to create subject level predictor files that span the GWAS array data. There are additional ways in which this method could be validated. One example would be to utilize a third approach, the BCFtools/RoH software, which uses a hidden Markov model[45], and re-run the analysis to generate the correlation of these methods. A final limitation present in sub-study three was a lack of control data. The *APOB* p.Arg3527Gln variant was used to validate the procedure, but, it is known to have an additive effect on LDL-C values. Therefore, without controlling for the presence of the *APOB* variant, LDL-C values in the region were highly associated with decreased heterozygosity, indicating that the *APOB* heterozygous carriers were driving the association. Due to the low frequency of *APOB* homozygous carriers, this result was expected. Nonetheless, an example in which the homozygous state of a variant, with a recessive inheritance pattern, drives the trait association, would have added validation to this study.

There are many benefits to conducting genetic study in the Lancaster Amish and other founder populations including their genetic and environmental homogeneity. However, the high degree of relatedness between Amish individuals that is leveraged in sub-studies two and three, decreases the total number of new variants discovered. This is not a disadvantage to the Amish community or for studies conducted on variants that are present in the Amish. In fact, the high number of Amish carriers is a great advantage when studying a variant's role in disease but there are a limited number of total variants discovered and causal variants present in the general outbred population may not be present. This could result in decreased ability to utilize findings outside of the Amish community.

Future directions

Through these three complementary sub-studies, we have identified P/LP variants in clinically actionable genes, identified reproductive health outcomes associated with homogeneity of couples using both carrier status and relatedness between couple pairs, and assessed complex traits associated with genome-wide and regional autozygosity. In addition to the work conducted here, many other aspects of these research topics are worth investigating.

There are several ways to expand on the results of sub-study one. First, the filtering pipeline eliminated singletons that did not have previous ClinVar annotations. These variants were not prioritized for annotations but are worth further investigating and, in addition, future exome sequencing data may identify more carriers in the Lancaster Amish community which would further help to annotate these variants. In addition, many of the genes on the consensus 78 panel are associated with cancer or tumor predisposition, inborn errors of metabolism, or malignant hyperthermia. Currently, we do not have sufficient phenotype data relevant to these disorders. Since our annotations identified many VUS, a future study could gather relevant phenotype data and attempt to reclassify some of these VUS as either P/LP, benign, or likely benign. These efforts could target known carriers and could include offering cascade genetic screening and phenotyping to their family members. Also, because of advances in medicine and future iterations of the ACMG guidelines, there will likely be additional genes and disorders added to the recommended list for return of results. A future pipeline could include monitoring this list and continuing variant annotation when new genes arise.

The second sub-study could be extended by investigating additional carriers of pathogenic recessive variants to increase sample size and allow for the detection of associations between reproductive outcomes and individual variants. This analysis could also be extended to test additional confounding variables if further survey data becomes available. There may be additional lifestyle choices made by parents who are both carriers of a pathogenic recessive variant that influence the number of children and number of pregnancies they have. For example, pathogenic recessive variant carriers may have a child affected with the recessive condition and therefore, alter their future decision making regarding reproductive health.

Autozygosity measurements, both genome wide and regional, could be used in multiple ways for future studies. For example, an analysis could be conducted in which the ROH distribution patterns of the Lancaster Amish are compared to the ROH distribution in populations with different history and demographics. Also, the association analysis could be expanded to look at traits in different ways. Though the current phenotype data does not have dichotomous traits of interest, grouping individuals by extreme phenotypes and looking at regions with differential ROH patterns could be an interesting future study. Case controls studies have used similar methods to find 'consensus regions' amongst cases.[72] In both autozygosity mapping and genome-wide analysis, size restrictions could be placed on ROHs in order to identify traits that are detrimentally impacted by ROHs with different size specifications. Frequency of long runs versus short runs across individuals with extreme phenotypes could be examined as well. Lastly, study subjects could be placed into groups determined by summed ROH length across the genome and trait variation could be measured in groups that fall into

extreme ends of summed ROH length. This analysis could also be performed using smaller regions of the genome. An example of an expected finding would be that trait variation diminishes with increased ROH length.

Potential Impact

This work has assessed the scope of genetic contribution to disease in the Amish community. Identification of P/LP variants in clinically actionable genes can help to better document disease burden in the population. Consistent with ACMG's intent for the use of the panel, individuals who are informed of their carrier status will be encouraged to work with their healthcare provider to make more informed health care decisions. Using data from a founder population, we were able to uncover enriched and novel P/LP variants that can be added to current databases including ClinVar. This will add to the current estimates on the frequency of the population that harbors one of these variants. There is a lot of interest in quantifying the disease burden in diverse populations which help create public health policies.[82]

Identifying anticipated population frequencies of Mendelian recessive disorders could inform the future development of novel types of carrier screening efforts in the Lancaster Amish community. The highly penetrant and recessive variants assessed in sub-study two were indicated to increase the risk of stillbirths. This knowledge can also aid in the efforts of the CSC. Evaluating the impact of relatedness across the genome as well as individual homozygosity suggested that relatedness between couples is associated with increased number of children and number of pregnancies and suggested that there are no specific traits in which genome-wide autozygosity has a detrimental impact. The regional mapping analysis identified an association with bilirubin that has a biological explanation. The second association identified through regional mapping does not have a biological explanation and was not identified during a large GWAS meta-analysis. This finding may spark additional interest in the region to evaluate the mechanism in which

HbA1c levels are increased. Although a high effect size variant may not be present in the region, there could be multiple recessive alleles driving this association. Autozygosity mapping has found a region worth further exploration.

Appendix 5.1. List of 96 phenotypes analyzed.

TRAIT	DESCRIPTION
adiponectin	Adiponectin values measured in mg/ml. Measured by Hopkins Bayview.
albumin	Screening blood test measured in g/dL. Serum albumin level. Measured by Quest PA.
alkalinePhosphatase	Screening blood test measured in U/L. Liver function test. Measured by Quest PA.
alt	Screening blood test measured in U/L. Liver function test. Measured by Quest PA.
armMidBMD	Forearm mid radius + ulna bone mineral density in grams / cm2 measured by DXA
armMidZScore	Forearm mid radius + ulna z score
armOneThirdBMD	Forearm one-third radius + ulna bone mineral density in grams / cm2 measured by DXA
armOneThirdZScore	Forearm one-third radius + ulna z score
armTotalBMD	total Forearm radius + ulna bone mineral density in grams / cm2 measured by DXA
armTotalZScore	total Forearm radius + ulna z score
armUltradistalBMD	Forearm ultradistal radius + ulna bone mineral density in grams / cm2 measured by DXA
armUltradistalZScore	Forearm ultradistal radius + ulna z score
ast	Screening blood test measured in U/L. Liver function test. Measured by Quest PA.
bilirubin	Screening blood test measured in mg/dL. Serum bilirubin level. Measured by Quest PA.
BMI	Calculated BMI in kg/m2
bodyBMD	whole body Bone mineral density in grams / cm2 measured by DXA in grams / cm2 measured by DXA
calcium	Screening blood test measured in mg/dL. Serum calcium level. Measured by Quest PA.
carbonDioxide	Screening blood test measured in mmol/L. Serum CO2 level. Measured by Quest PA.
carotidIMT	Common carotid artery Intimal Medial Thickness in mm
cfPWV	Pulse wave velocity in the femoral carotid
chloride	Screening blood test measured in mmol/L. Serum chloride level. Measured by Quest PA.
cholesterol	Fasting Total Cholesterol measured at Study Visit in mg/dl units. Measured by Quest.
coronaryCalcification	
creatinine	Screening blood test measured in mg/dL. Kidney function test. Measured by Quest PA.

Appendix 5.1. continued

CRP	C-reactive protein measured by Quest. Units are mg/dL
crPWV	Pulse wave velocity in the radial carotid
diastolicBloodPressure	Diastolic blood pressure measured during the initial clinic visit
eGFR	Estimated glomerular filtration rate= eGFR= 186 x (creatinine)^-1.154 x (age)^-0.203 x (0.742 if female) x (1.210 if African American). GFR is expressed in mL/min/1.73m^2
epicardialFat	Adipose tissue between the surface of the heart and the visceral epicardium, cm3
farWallMax	Common carotid artery Intimal Far Wall Max Thickness in mm (mean of 4 measures-2 right and 2 left)
fatMass	Fat mass in grams measured by the DXA
globulin	Screening blood test measured in g/dL. Serum globulin level. Calculated and measured by Quest PA.
Glucose	fasting glucose measured in mg/dl. Measured by the University of MD or Quest (Wellness samples).
HbA1c	
HDL	Fasting HDL Cholesterol measured at Study Visit in mg/dl units. Measured by Quest.
HDL2	Fasting HDL subfraction 2 mg/dl
HDL3	Fasting HDL subfraction 3 mg/dl
height	Height of individual in centimeters
hematocrit	Cell blood count measured in %. Hematocrit. Measured by Quest PA.
hemoglobin	Cell blood count measured in g/dL. Hemoglobin. Measured by Quest PA.
hipCircumference	Circumference of individual's hip in centimeters
hipFemoralNeckBMD	Hip Femoral Neck bone mineral density in grams / cm2 measured by DXA
hipFemoralNeckZScore	Hip Femoral Neck z score
hipIntertrochanterBMD	Hip Intertrochanter bone mineral density in grams / cm2 measured by DXA
hipIntertrochanterZScore	Hip Intertrochanter z score
hipTotalBMD	total Hip bone mineral density in grams / cm2 measured by DXA
hipTotalZScore	total Hip z score
hipTrochanterBMD	Hip Trochanter bone mineral density in grams / cm2 measured by DXA
hipTrochanterZScore	Hip Trochanter z score
HOMAB	
HOMAIR	
homocysteine	Homocysteine measured in micromol/L

Appendix 5.1. continued

IDL	Fasting intermediate density lipoprotein mg/dl: note: zero and negative values were changed to 1 per Dr. Kris Kulkarni @Atherotech
Insulin	Fasting insulin measured in microU/ml. Measured by Hopkins Bayview
LDL	Fasting LDL Cholesterol measured at Study Visit in mg/dl units. Measured by Quest.
leanMass	Lean mass in grams measured by the DXA
lipoproteinA	Fasting lipoprotein (a)
liverAverage	Average of liver attenuation
liverSpleenRatio	Average of liver attenuation/spleen attenuation
LuminalDiameter	Diameter of the common carotid artery at the end diastole, measured in mm
meanAnkleBrachialIndex	Calculated variable: average of right and left ABI measurements
meanArterialPressure	Mean arterial pressure calculated as 2/3DBP + 1/3SBP
meanCorpuscularHGB	Mean corpuscular hemoglobin = HB/RBC (picograms/cell)
meanCorpuscularHGBConc	Mean corpuscular hemoglobin concentration = MCH/MCV (grams/decilitier)
meanCorpuscularVolume	Mean corpuscular volume Hematocrit/rbc (femotliters/cell)
medSternalFat	Intrathoracic adipose tissue, the combination of the epicardial and pericardial adipose tissue from the right pulmonary artery to the diaphragm and the anterior chest wall to the vertebral column, cm3
percentFat	% body fat by dexa
pericardialFat	Adipose tissue around the heart but outside of the fibrous pericardium, cm3
platelets	Cell blood count measured in thous/mcl. Platelet count. Measured by Quest PA.
potassium	Screening blood test measured in mmol/L. Serum potassium level. Measured by Quest PA.
PRinterval	PR Interval from the EKG
protein	Screening blood test measured in g/dL. Serum protein level. Measured by Quest PA.
QTcInterval	QTc Interval from the EKG
QTInterval	QT Interval from the EKG
rbcDistributionWidth	Cell blood count measure in %. Red blood cell distribution width. Measured by Quest PA.
redBloodCellCount	Cell blood count measured in mill/mcl. Red blood cell count. Measured by Quest PA.
RelativeWallThickness	Calculated variable: CommonCarotidIMT / LuminalDiameter
remnantLipoprotein	Fasting remnant lipoprotein cholesterol
sodium	Screening blood test measured in mmol/L. Serum sodium level. Measured by Quest PA.
spineTotalBMD	total Spine bone mineral density in grams / cm2 measured by DXA

Appendix 5.1. continued

spineTotalZScore	total Spine z score
systolicBloodPressure	Systolic blood pressure measured during the initial clinic visit
totalMass	Total mass in grams measured by DXA
TotalNonHDL	Fasting non-HDL cholesterol mg/dl
TotalVLDL	Fasting very low density lipoprotein cholesterol mg/dl
triglycerides	Fasting triglycerides measured at Study Visit in mg/dl units. Measured by Quest.
tsh	Thyroid hormone measured at screening blood draw in miu/L. Measured by Quest PA.
tTriglycerides	log transformed: Fasting triglycerides measured at Study Visit in mg/dl units. Measured by Quest.
ureaNitrogen	Screening blood test measured in mg/dL. Kidney function test. Measured by Quest PA.
uricAcid	Uric Acid measured in mg/dL. Measured by Quest PA.
VascularMass	Calculated variable: 1.06*pi*((LuminaDiameter/2 + CommonCarotidIMT)^2 - (LuminalDiameter/2)^2) Note: diameters are first converted to cm
VentRate	Heart rate measured by EKG
VLDL3	Fasting VLDL subfraction 3 mg/dl
waistCircumference	Circumference of individual's waist in centimeters
weight	Weight of individual in kilograms
whiteBloodCellCount	Cell blood count measured in thous/mcl. White blood cell count. Measured by Quest PA.

Appendix 5.2. Regions with increased frequency of individuals with SNPs in ROH segments.

rsNum	Pos38	Chr38	A2	A1	Gene	SNP Freq	ROH Freq
rs6707326	134600135	2	G	A	TMEM163	0.61	0.11
rs6745632	134642751	2	G	T	TMEM163	0.85	0.11
rs6747870	134712199	2	A	G	TMEM163	0.20	0.11
rs35853750	134712458	2	C	A	TMEM163	0.20	0.11
rs6430538	134782397	2	C	T	ACMSD	0.36	0.11
rs79472142	134852437	2	G	A	ACMSD	0.05	0.12
rs6430553	134873830	2	C	T	CCNT2-AS1	0.39	0.12
rs3769027	134919465	2	T	C	CCNT2	0.28	0.12
rs10166142	134952752	2	A	G	CCNT2	0.39	0.12
rs12463963	134961486	2	C	T	CCNT2	0.11	0.12
rs58663889	134998695	2	G	T	MAP3K19	0.08	0.12
rs16831256	135023465	2	C	T	MAP3K19	0.08	0.12
rs4954218	135045855	2	G	T	RAB3GAP1	0.65	0.12
rs7570971	135080336	2	C	A	RAB3GAP1	0.28	0.12
rs7575424	135099463	2	G	A	RAB3GAP1	0.09	0.12
rs76333606	135230665	2	C	T	ZRANB3	0.03	0.12
rs935613	135265228	2	C	T	ZRANB3	0.23	0.12
rs10928541	135635587	2	G	A	R3HDM1	0.07	0.13
rs12619365	135640604	2	C	T	R3HDM1	0.21	0.13
rs60558655	135752732	2	C	A	UBXN4	0.05	0.13
rs2322659	135798089	2	T	C	LCT	0.77	0.13
rs2304371	135803987	2	G	A	LCT	0.84	0.13
rs3754689	135833176	2	C	T	LCT	0.16	0.13
rs11693502	136051379	2	C	T	DARS-AS1	0.70	0.12
rs749873	136059518	2	C	T	DARS-AS1	0.69	0.12
rs6430600	136078795	2	G	A	CXCR4	0.37	0.12
rs12991970	136089261	2	G	A	CXCR4	0.48	0.12
rs7569447	136106246	2	C	A	CXCR4	0.85	0.12
rs2228014	136115514	2	G	A	CXCR4	0.08	0.11
rs6595989	131023522	5	T	C	HINT1	0.96	0.18
rs200148803	131107489	5	A	G	HINT1	0.05	0.18
rs6596004	131211472	5	G	T	LYRM7	0.22	0.18
rs3822311	131535600	5	T	C	RAPGEF6	0.14	0.18
rs2042252	131644334	5	G	A	FNIP1	0.72	0.18
rs10066812	131661220	5	T	G	FNIP1	0.73	0.18
rs6870930	131774432	5	C	A	FNIP1	0.72	0.18
rs2074808	131993898	5	T	C	ACSL6	0.14	0.18
rs17132283	131995966	5	G	A	ACSL6	0.04	0.18
rs440970	132000594	5	T	G	ACSL6	0.25	0.18

Appendix 5.2. continued

rs247004	132008665	5	T	C	ACSL6	0.82	0.18
rs247011	132023921	5	G	A	ACSL6	0.15	0.18
rs11575022	132066322	5	A	C	IL3	0.03	0.18
rs3091338	132067045	5	C	T	IL3	0.56	0.18
rs31474	132069334	5	T	C	CSF2	0.17	0.18
rs25887	132080368	5	C	A	CSF2	0.74	0.18
rs657075	132094425	5	G	A	CSF2	0.12	0.18
rs10065787	132100793	5	G	T	CSF2	0.57	0.18
rs45236	132104927	5	G	T	CSF2	0.16	0.18
rs13358835	132138477	5	A	G	P4HA2-AS1	0.12	0.18
rs156033	132191628	5	C	T	P4HA2-AS1	0.16	0.18
rs2077380	132232924	5	T	C	P4HA2	0.21	0.18
rs2136187	132242201	5	T	C	P4HA2	0.79	0.18
rs6871350	132244527	5	T	C	PDLIM4	0.79	0.18
rs7727544	132254841	5	C	T	PDLIM4	0.75	0.18
rs3900945	132257177	5	T	C	PDLIM4	0.87	0.18
rs162904	132260091	5	T	C	PDLIM4	0.87	0.18
rs28532117	132287998	5	C	T	SLC22A4	0.12	0.18
rs419291	132297662	5	T	C	SLC22A4	0.73	0.18
rs13179900	132311897	5	G	A	LOC553103	0.11	0.18
rs11950562	132316836	5	A	C	LOC553103	0.65	0.18
rs273914	132324738	5	A	T	LOC553103	0.73	0.18
rs273913	132325463	5	T	C	LOC553103	0.73	0.18
rs272889	132329685	5	A	G	LOC553103	0.73	0.18
rs273909	132331660	5	A	G	LOC553103	0.09	0.18
rs1050152	132340627	5	C	T	SLC22A4	0.60	0.18
rs272869	132342304	5	A	G	LOC553103	0.73	0.18
rs10058074	132350453	5	G	A	LOC553103	0.65	0.18
rs274546	132364175	5	A	G	LOC553103	0.73	0.18
rs2631372	132367886	5	G	C	LOC553103	0.16	0.18
rs2631367	132369766	5	C	G	LOC553103	0.35	0.18
rs274567	132378717	5	C	T	SLC22A5	0.27	0.18
rs2073643	132387596	5	T	C	SLC22A5	0.76	0.18
rs274551	132391988	5	T	C	SLC22A5	0.87	0.18
rs2405522	132406004	5	A	G	C5orf56	0.87	0.18
rs1016988	132408882	5	T	C	C5orf56	0.12	0.18
rs4451042	132419785	5	G	T	C5orf56	0.95	0.18
rs11242111	132420366	5	A	G	C5orf56	0.95	0.18
rs2106854	132433482	5	C	T	C5orf56	0.08	0.18
rs2188962	132435113	5	C	T	C5orf56	0.60	0.18
rs7730247	132445596	5	A	C	C5orf56	0.08	0.18
rs12521868	132448701	5	G	T	C5orf56	0.60	0.18

Appendix 5.2. continued

rs77450473	132449844	5	G	A	C5orf56	0.08	0.18
rs1023518	132458080	5	G	T	C5orf56	0.16	0.18
rs2522051	132461886	5	T	C	C5orf56	0.24	0.18
rs2522056	132466034	5	G	A	C5orf56	0.08	0.18
rs10072700	132481211	5	A	C	IRF1	0.20	0.18
rs78467594	27896050	6	C	T	HIST1H2BO	0.07	0.14
rs9461425	27951910	6	G	A	OR2B6	0.17	0.14
rs149971	28014374	6	G	A	OR2B6	0.41	0.14
rs149990	28030480	6	G	A	ZNF165	0.12	0.14
rs9468287	28111963	6	C	A	ZSCAN16-AS1	0.12	0.14
rs16893731	28162416	6	G	A	ZNF192P1	0.07	0.13
rs1150674	28166590	6	A	G	ZNF192P1	0.08	0.13
rs1150678	28174498	6	C	T	ZNF192P1	0.38	0.13
rs9380064	28175340	6	A	G	ZNF192P1	0.14	0.13
rs9380069	28235522	6	A	G	ZSCAN9	0.12	0.13
rs1150711	28240757	6	T	C	ZKSCAN4	0.31	0.13
rs13408	28244970	6	G	A	ZKSCAN4	0.41	0.13
rs1635	28259826	6	C	A	NKAPL	0.08	0.13
rs3800326	28296940	6	C	T	PGBD1	0.08	0.13
rs6938961	28312520	6	G	A	PGBD1	0.29	0.13
rs17312661	28332559	6	A	G	ZSCAN31	0.14	0.13
rs77983937	28340115	6	A	G	ZSCAN31	0.08	0.13
rs6903823	28354519	6	A	G	ZKSCAN3	0.24	0.13
rs78512949	28360516	6	G	A	ZKSCAN3	0.10	0.13
rs2232422	28398485	6	G	A	ZSCAN12	0.03	0.13
rs1005125	28399578	6	G	A	ZSCAN12	0.44	0.13
rs9468379	28463407	6	C	T	ZSCAN23	0.09	0.13
rs392471	28591470	6	A	C	ZBED9	0.09	0.13
rs6901724	28598050	6	T	C	ZBED9	0.03	0.13
rs9501180	28611694	6	T	C	ZBED9	0.49	0.13
rs116591906	28615615	6	A	G	ZBED9	0.15	0.13
rs62402400	28634012	6	A	G	ZBED9	0.24	0.13
rs7772289	28706545	6	G	T	ZBED9	0.36	0.13
rs9368570	28708416	6	G	A	ZBED9	0.48	0.13
rs9468413	28721895	6	C	A	ZBED9	0.86	0.13
rs2893970	28735504	6	T	C	LOC401242	0.43	0.13
rs1233578	28744470	6	A	G	LOC401242	0.14	0.13
rs57558053	28786002	6	C	A	LOC401242	0.04	0.13
rs7739915	28802906	6	G	T	LOC401242	0.16	0.13
rs4324798	28808340	6	G	A	LOC401242	0.13	0.13
rs172329	28821489	6	T	C	LOC401242	0.16	0.13
rs9257425	28968779	6	A	G	LINC01556	0.47	0.13

Appendix 5.2. continued

rs4713190	28975056	6	G	A	ZNF311	0.53	0.13
rs4713191	28976092	6	T	C	ZNF311	0.05	0.13
rs9257446	28981752	6	A	G	ZNF311	0.17	0.13
rs7484151	49671741	11	G	A	LOC440040	0.90	0.13
rs139682087	49667208	11	T	C	LOC440040	0.05	0.13
rs1814175	49537620	11	T	C	LOC440040	0.78	0.13
rs12098986	49525356	11	C	T	LOC440040	0.10	0.13
rs115251079	49505525	11	A	C	LOC440040	0.59	0.13
rs7102702	49379573	11	T	G	FOLH1	0.82	0.14
rs1164660	49280089	11	G	T	FOLH1	0.74	0.14
rs7117025	49275273	11	T	C	FOLH1	0.10	0.14
rs202712	49177372	11	C	T	FOLH1	0.14	0.14
rs11039885	48625555	11	A	G	OR4A47	0.65	0.14
rs10160716	48607931	11	C	T	OR4A47	0.91	0.14
rs10838956	48560213	11	T	C	OR4A47	0.77	0.14
rs114343057	48536620	11	C	T	OR4A47	0.10	0.14
rs7395662	48497341	11	A	G	OR4A47	0.77	0.14
rs636696	48414196	11	T	G	OR4C45	0.90	0.14
rs1483121	48311808	11	G	A	OR4S1	0.14	0.14
rs1566729	48143169	11	T	C	PTPRJ	0.93	0.14
rs1503185	48125070	11	G	A	PTPRJ	0.11	0.14
rs10838801	48076728	11	G	A	PTPRJ	0.88	0.14
rs10838798	48069751	11	T	G	PTPRJ	0.87	0.14
rs7130876	48029443	11	A	G	PTPRJ	0.09	0.14
rs1681630	47947600	11	T	C	PTPRJ	0.86	0.15
rs10838790	47910005	11	C	T	NUP160	0.96	0.15
rs1471712	47890598	11	G	A	NUP160	0.77	0.15
rs12364432	47881331	11	G	A	NUP160	0.67	0.15
rs11039377	47768224	11	C	T	FNBP4	0.23	0.15
rs10838739	47651191	11	C	T	AGBL2	0.23	0.15
rs10838738	47641497	11	A	G	MTCH2	0.67	0.15
rs3817334	47629441	11	C	T	MTCH2	0.69	0.15
rs10838725	47536319	11	T	C	CELF1	0.18	0.15
rs9666924	47518455	11	C	T	CELF1	0.92	0.15
rs7124681	47508395	11	C	A	CELF1	0.69	0.15
rs34312154	47448793	11	G	A	RAPSN	0.04	0.15
rs12419342	47446993	11	C	T	RAPSN	0.87	0.15
rs45617144	47437918	11	G	A	RAPSN	0.04	0.14
rs11604680	47435988	11	A	G	RAPSN	0.58	0.14
rs10838708	47419962	11	G	A	PSMC3	0.71	0.14
rs1377416	47395195	11	C	T	SLC39A13	0.29	0.14
rs4752990	47388842	11	G	A	SPI1	0.58	0.12

Appendix 5.2. continued

rs4911502	35475209	20	C	T	CEP250	0.20	0.10
rs3748433	35502691	20	G	A	CEP250	0.06	0.10
rs2236164	35509524	20	T	C	CEP250	0.20	0.10
rs7261862	35528453	20	T	C	C20orf173	0.13	0.10
rs41293076	35560555	20	T	G	FER1L4	0.07	0.10
rs2277862	35564866	20	C	T	FER1L4	0.11	0.10
rs6088887	35583402	20	G	A	FER1L4	0.11	0.10
rs6058303	35702359	20	A	G	RBM39	0.13	0.10
rs1890470	35769086	20	C	T	PHF20	0.08	0.10
rs12481545	35819776	20	C	T	PHF20	0.08	0.10

Appendix 5.3. Regions with decreased frequency of individuals with SNPs in ROH segments.

rsNum	Pos38	Chr38	A2	A1	Gene	SNP Freq	ROH Freq
rs6683357	121422747	1	G	A	SRGAP2-AS1	0.49	0.00
rs4844381	121443040	1	C	T	SRGAP2-AS1	0.49	0.00
rs61803643	121455021	1	G	A	SRGAP2-AS1	0.06	0.00
rs115706173	121538687	1	A	G	EMBP1	0.05	0.00
rs11249433	121538815	1	A	G	EMBP1	0.30	0.00
rs78847106	121569869	1	G	A	EMBP1	0.10	0.00
rs114551810	121588919	1	T	C	EMBP1	0.03	0.00
rs1851250	121607560	1	A	G	EMBP1	0.14	0.00
rs116029751	143692183	1	C	T	NBPF25P	0.11	0.00
rs145974645	143837623	1	G	T	FCGR1C	0.07	0.00
rs59942232	143842894	1	G	A	FCGR1C	0.22	0.00
rs58198317	91748437	2	G	A	GGT8P	0.05	0.00
rs56033681	91763846	2	G	A	GGT8P	0.09	0.00
rs1179068	91944868	2	C	T	ACTR3BP2	0.04	0.00
rs117534047	91974615	2	C	T	ACTR3BP2	0.03	0.00
rs28429086	92072753	2	C	T	ACTR3BP2	0.05	0.00
rs11794599	61549722	9	G	T	FAM27C	0.07	0.00
rs11263320	61655601	9	G	A	FAM27C	0.07	0.00
rs11263321	61655793	9	A	G	FAM27C	0.07	0.00
rs10121167	61657188	9	G	T	FAM27C	0.96	0.00
rs35478119	61663338	9	C	T	FAM27C	0.07	0.00
rs11263339	61667843	9	G	C	FAM27C	0.56	0.00
rs13291409	61672104	9	C	T	FAM27C	0.07	0.00
rs118106267	63840173	9	A	G	LOC642236	0.06	0.00
rs61969218	18385175	14	G	T	OR11H12	0.53	0.00
rs61970557	18488398	14	G	A	OR11H12	0.35	0.00
rs117764863	19839089	15	G	A	CHEK2P2	0.05	0.00
rs10163108	19946357	15	C	T	CHEK2P2	0.03	0.00
rs12906138	19979347	15	G	A	CHEK2P2	0.37	0.00
rs7496731	19983994	15	C	T	CHEK2P2	0.34	0.00
rs7179358	19986488	15	A	G	CHEK2P2	0.39	0.00
rs4932079	19986758	15	C	A	CHEK2P2	0.78	0.00
rs11853381	20645098	15	A	G	NBEAP1	0.03	0.00
rs369761757	20666219	15	T	C	NBEAP1	0.03	0.00
rs372974967	20672738	15	C	T	NBEAP1	0.04	0.00
rs117277427	33758179	16	T	G	RNU6-76P	0.03	0.00
rs8047974	34987328	16	C	A	UBE2MP1	0.57	0.00
rs117807171	35188590	16	C	T	UBE2MP1	0.04	0.00
rs75974209	35378992	16	G	A	LINC01566	0.06	0.00

Appendix 5.3. continued

rs74987014	35397706	16	A	G	LINC01566	0.12	0.00
rs8048302	35452681	16	G	A	FRG2DP	0.04	0.00
rs62057643	35479538	16	A	G	FRG2DP	0.04	0.00
rs76316557	35485726	16	G	A	FRG2DP	0.03	0.00
rs118035966	35506209	16	C	T	LOC100130700	0.07	0.00
rs12929970	35569018	16	C	T	LOC100130700	0.19	0.00
rs74756439	35641318	16	A	G	FLJ26245	0.06	0.00
rs55872932	35643344	16	C	T	FLJ26245	0.07	0.00
rs79185219	35729568	16	C	T	FLJ26245	0.08	0.00
rs72814856	35760765	16	A	C	FLJ26245	0.06	0.00
rs79247636	35814989	16	T	G	FLJ26245	0.04	0.00
rs7204407	35877150	16	G	A	FLJ26245	0.04	0.00
rs7187627	35908045	16	G	A	FLJ26245	0.15	0.00
rs72814889	36022890	16	C	T	FLJ26245	0.06	0.00
rs79971566	92127	18	G	A	ROCK1P1	0.04	0.01
rs113299446	122517	18	C	T	ROCK1P1	0.19	0.01
rs4797697	139767	18	C	T	ROCK1P1	0.21	0.01
rs72856217	155280	18	G	A	USP14	0.09	0.01
rs545684	159382	18	G	A	USP14	0.51	0.01
rs1274275	188526	18	C	T	USP14	0.42	0.01
rs681760	195644	18	T	G	USP14	0.12	0.01
rs9951575	213592	18	A	G	USP14	0.13	0.01
rs8094221	220341	18	C	A	THOC1	0.14	0.01
rs71352929	227116	18	A	G	THOC1	0.03	0.01
rs7226677	278796	18	A	G	THOC1	0.15	0.01
rs116907095	283900	18	T	C	THOC1	0.05	0.01
rs620837	286971	18	C	A	THOC1	0.55	0.01
rs585955	295046	18	G	T	COLEC12	0.21	0.01
rs503436	297805	18	A	G	COLEC12	0.40	0.01
rs116858912	298506	18	G	T	COLEC12	0.05	0.01
rs16942188	303811	18	A	G	COLEC12	0.41	0.01
rs74585125	308782	18	C	T	COLEC12	0.17	0.01
rs668340	312221	18	C	T	COLEC12	0.43	0.01
rs60318638	316349	18	T	C	COLEC12	0.24	0.01
rs8094638	332469	18	G	T	COLEC12	0.22	0.01
rs2305027	334742	18	C	T	COLEC12	0.67	0.01
rs77962671	338054	18	C	T	COLEC12	0.09	0.01

References

1. Kääriäinen H, Muilu J, Perola M, Kristiansson K. Genetics in an isolated population like Finland: a different basis for genomic medicine? *J Community Genet*. 2017;8(4):319-326. doi:10.1007/s12687-017-0318-4

2. Neuhausen SL. Founder populations and their uses for breast cancer genetics. *Breast Cancer Res*. 2000;2(2):77-81. doi:10.1186/bcr36

3. Crow EJF, Dove WF. Anecdotal, Historical and Critical Commentaries on Genetics. :6.

4. Francomano CA, McKusick VA, Biesecker LG. Medical genetic studies in the Amish: historical perspective. *Am J Med Genet C Semin Med Genet*. 2003;121C(1):1-4. doi:10.1002/ajmg.c.20001

5. Crowgey EL, Washburn MC, Kolb EA, Puffenberger EG. Development of a Novel Next-Generation Sequencing Assay for Carrier Screening in Old Order Amish and Mennonite Populations of Pennsylvania. *J Mol Diagn*. 2019;21(4):687-694. doi:10.1016/j.jmoldx.2019.03.004

6. Lee W-J, Pollin TI, O'Connell JR, Agarwala R, Schäffer AA. PedHunter 2.0 and its usage to characterize the founder structure of the Old Order Amish of Lancaster County. *BMC Med Genet*. 2010;11:68. doi:10.1186/1471-2350-11-68

7. Agarwala R, Schäffer AA, Tomlin JF. Towards a complete North American Anabaptist Genealogy II: analysis of inbreeding. *Hum Biol*. 2001;73(4):533-545. doi:10.1353/hub.2001.0045

8. Umble J. The Old Order Amish, Their Hymns and Hymn Tunes. *The Journal of American Folklore*. 1939;52(203):82. doi:10.2307/536013

9. Hostetler J. *Amish Society*. 4th ed. Johns Hopkins University Press; 1993.

10. Crawford SY, Manuel AM, Wood BD. Pharmacists' considerations when serving Amish patients. *Journal of the American Pharmacists Association*. 2009;49(1):86-97. doi:10.1331/JAPhA.2009.07160

11. Huffman jasper. *History of the Mennonite Brethren in Christ Church*. Bethel Publishing Company; 1920.

12. Wenger AFZ. Cultural Context, Health and Health Care Decision Making. *J Transcult Nurs*. 1995;7(1):3-14. doi:10.1177/104365969500700102

13. Beachy A, Hershberger E, Davidhizar R, Giger JN. Cultural implications for nursing care of the Amish. *J Cult Divers*. 1997;4(4):118-126; quiz 127-128.

14. Weyer SM, Hustey VR, Rathbun L, et al. A Look Into the Amish Culture: What Should We Learn? *J Transcult Nurs*. 2003;14(2):139-145. doi:10.1177/1043659602250639

15. Hsueh WC, Mitchell BD, Aburomia R, et al. Diabetes in the Old Order Amish: characterization and heritability analysis of the Amish Family Diabetes Study. *Diabetes Care*. 2000;23(5):595-601. doi:10.2337/diacare.23.5.595

16. Mckusick VA, Eldridge R, Hostetler JA, Ruangwit U, Egeland JA. DWARFISM IN THE AMISH. II. CARTILAGE-HAIR HYPOPLASIA. *Bull Johns Hopkins Hosp*. 1965;116:285-326.

17. Kelley RI, Robinson D, Puffenberger EG, Strauss KA, Morton DH. Amish lethal microcephaly: A new metabolic disorder with severe congenital microcephaly and 2-ketoglutaric aciduria. *Am J Med Genet*. 2002;112(4):318-326. doi:10.1002/ajmg.10529

18. Cross HE. The Troyer Syndrome: A Recessive Form of Spastic Paraplegia With Distal Muscle Wasting. *Arch Neurol*. 1967;16(5):473. doi:10.1001/archneur.1967.00470230025003

19. Cross HE. The Mast Syndrome: A Recessively Inherited Form of Presenile Dementia With Motor Disturbances. *Arch Neurol*. 1967;16(1):1. doi:10.1001/archneur.1967.00470190005001

20. Hsueh W-C, Mitchell BD, Schneider JL, et al. Genome-Wide Scan of Obesity in the Old Order Amish. 2001;86(3):7.

21. Strauss KA, Carson VJ, Soltys K, et al. Branched-chain α-ketoacid dehydrogenase deficiency (maple syrup urine disease): Treatment, biomarkers, and outcomes. *Molecular Genetics and Metabolism*. 2020;129(3):193-206. doi:10.1016/j.ymgme.2020.01.006

22. Morton DH, Bennett MJ, Seargeant LE, Nichter CA, Kelley RI. Glutaric aciduria type I: A common cause of episodic encephalopathy and spastic paralysis in the Amish of Lancaster County, Pennsylvania. *Am J Med Genet*. 1991;41(1):89-95. doi:10.1002/ajmg.1320410122

23. Mitchell BD, McArdle PF, Shen H, et al. The genetic response to short-term interventions affecting cardiovascular function: Rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *American Heart Journal*. 2008;155(5):823-828. doi:10.1016/j.ahj.2008.01.019

24. Shuldiner AR. Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy of Clopidogrel Therapy. *JAMA*. 2009;302(8):849. doi:10.1001/jama.2009.1232

25. Markon AO, Ryan KA, Wadhawan A, et al. Risk factors for *Toxoplasma gondii* seropositivity in the Old Order Amish. *Epidemiol Infect*. 2021;149:e89. doi:10.1017/S0950268820002897

26. Pollin TI, Damcott CM, Shen H, et al. A Null Mutation in Human APOC3 Confers a Favorable Plasma Lipid Profile and Apparent Cardioprotection. *Science*. 2008;322(5908):1702-1705. doi:10.1126/science.1161524

27. Shen H, Damcott CM, Rampersaud E, et al. Familial defective apolipoprotein B-100 and increased low-density lipoprotein cholesterol and coronary artery calcification in the old order amish. *Arch Intern Med*. 2010;170(20):1850-1855. doi:10.1001/archinternmed.2010.384

28. Streeten EA, See VY, Jeng LBJ, et al. KCNQ1 and Long QT Syndrome in 1/45 Amish: The Road From Identification to Implementation of Culturally Appropriate Precision Medicine. *Circ Genom Precis Med*. 2020;13(6):e003133. doi:10.1161/CIRCGEN.120.003133

29. Bhatnagar A. Environmental Determinants of Cardiovascular Disease. *Circ Res*. 2017;121(2):162-180. doi:10.1161/CIRCRESAHA.117.306458

30. Brown M, Goldstein J. A receptor-mediated pathway for cholesterol homeostasis. *Science*. 1986;232(4746):34-47. doi:10.1126/science.3513311

31. Innerarity TL, Weisgraber KH, Arnold KS, et al. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. *Proceedings of the National Academy of Sciences*. 1987;84(19):6919-6923. doi:10.1073/pnas.84.19.6919

32. Alves AC, Benito-Vicente A, Medeiros AM, Reeves K, Martin C, Bourbon M. Further evidence of novel APOB mutations as a cause of familial hypercholesterolaemia. *Atherosclerosis*. 2018;277:448-456. doi:10.1016/j.atherosclerosis.2018.06.819

33. Lacaze P, Sebra R, Riaz M, et al. Familial Hypercholesterolemia in a Healthy Elderly Population. *Circ: Genomic and Precision Medicine*. 2020;13(4). doi:10.1161/CIRCGEN.120.002938

34. Motazacker MM, Pirruccello J, Huijgen R, et al. Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia. *European Heart Journal*. 2012;33(11):1360-1366. doi:10.1093/eurheartj/ehs010

35. Mitchell BD, Schäffer AA, Pollin TI, et al. Mapping Genes in Isolated Populations: Lessons from the Old Order Amish. In: Duggirala R, Almasy L, Williams-Blangero S, Paul SFD, Kole C, eds. *Genome Mapping and Genomics in Human and Non-Human Primates*. Springer Berlin Heidelberg; 2015:141-153. doi:10.1007/978-3-662-46306-2_10

36. Schwartz PJ, Stramba-Badiale M, Crotti L, et al. Prevalence of the Congenital Long-QT Syndrome. *Circulation*. 2009;120(18):1761-1767. doi:10.1161/CIRCULATIONAHA.109.863209

37. Kapplinger JD, Tester DJ, Salisbury BA, et al. Spectrum and prevalence of mutations from the first 2,500 consecutive unrelated patients referred for the FAMILION® long QT syndrome genetic test. *Heart Rhythm*. 2009;6(9):1297-1303. doi:10.1016/j.hrthm.2009.05.021

38. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):249-255. doi:10.1038/gim.2016.190

39. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science*. 2016;354(6319). doi:10.1126/science.aaf6814

40. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med*. 2016;18(9):906-913. doi:10.1038/gim.2015.187

41. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565-574. doi:10.1038/gim.2013.73

42. ; on behalf of the ACMG Laboratory Quality Assurance Committee, Richards S, Aziz N, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-423. doi:10.1038/gim.2015.30

43. Fretts RC. Etiology and prevention of stillbirth. *American Journal of Obstetrics and Gynecology*. 2005;193(6):1923-1935. doi:10.1016/j.ajog.2005.03.074

44. Macklon NS. Conception to ongoing pregnancy: the “black box” of early pregnancy loss. *Human Reproduction Update*. 2002;8(4):333-343. doi:10.1093/humupd/8.4.333

45. Narasimhan VM, Hunt KA, Mason D, et al. Health and population effects of rare gene knockouts in adult humans with related parents. *Science*. 2016;352(6284):474-477. doi:10.1126/science.aac8624

46. Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017;544(7649):235-239. doi:10.1038/nature22034

47. Clark DW. Associations of autozygosity with a broad range of human phenotypes. :17.

48. Agarwala R, Schaffer AA, Tomlin JF. Towards a Complete North American Anabaptist Genealogy II: Analysis of Inbreeding. *Human Biology*. 2001;73(4):533-545. doi:10.1353/hub.2001.0045

49. Ober C, Hyslop T, Hauck WW. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am J Hum Genet*. 1999;64(1):225-231. doi:10.1086/302198

50. Hosseini-Chavoshi M, Abbasi-Shavazi MJ, Bittles AH. Consanguineous marriage, reproductive behaviour and postnatal mortality in contemporary Iran. *Hum Hered*. 2014;77(1-4):16-25. doi:10.1159/000358403

51. Shami SA, Schmitt LH, Bittles AH. Consanguinity, spousal age at marriage and fertility in seven Pakistani Punjab cities. *Ann Hum Biol*. 1990;17(2):97-105. doi:10.1080/03014469000000842

52. Helgason A, Pálsson S, Gudbjartsson DF, Kristjánsson T, Stefánsson K. An association between the kinship and fertility of human couples. *Science*. 2008;319(5864):813-816. doi:10.1126/science.1150232

53. Bittles AH, Grant JC, Sullivan SG, Hussain R. Does inbreeding lead to decreased human fertility? *Ann Hum Biol*. 2002;29(2):111-130. doi:10.1080/03014460110075657

54. Saad FA, Jauniaux E. Recurrent early pregnancy loss and consanguinity. *Reprod Biomed Online*. 2002;5(2):167-170. doi:10.1016/s1472-6483(10)61620-3

55. Kapurubandara S, Melov S, Shalou E, Alahakoon I. Consanguinity and associated perinatal outcomes, including stillbirth. *Aust N Z J Obstet Gynaecol*. 2016;56(6):599-604. doi:10.1111/ajo.12493

56. Assaf S, Khawaja M, DeJong J, Mahfoud Z, Yunis K. Consanguinity and reproductive wastage in the Palestinian Territories. *Paediatric and Perinatal Epidemiology*. 2009;23(2):107-115. doi:10.1111/j.1365-3016.2008.00988.x

57. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet*. 2018;19(4):220-234. doi:10.1038/nrg.2017.109

58. Gosadi IM, Goyder EC, Teare MD. Investigating the Potential Effect of Consanguinity on Type 2 Diabetes Susceptibility in a Saudi Population. *Hum Hered*. 2014;77(1-4):197-206. doi:10.1159/000362447

59. Ismail J, Jafar TH, Jafary FH, White F, Faruqui AM, Chaturvedi N. Risk factors for non-fatal myocardial infarction in young South Asian adults. *Heart*. 2004;90(3):259-263. doi:10.1136/hrt.2003.013631

60. Yengo L, Zhu Z, Wray NR, et al. Detection and quantification of inbreeding depression for complex traits from SNP data. *Proc Natl Acad Sci USA*. 2017;114(32):8602-8607. doi:10.1073/pnas.1621096114

61. Liede A, Malik IA, Aziz Z, Rios P de los, Kwan E, Narod SA. Contribution of BRCA1 and BRCA2 Mutations to Breast and Ovarian Cancer in Pakistan. *The American Journal of Human Genetics*. 2002;71(3):595-606. doi:10.1086/342506

62. Bittles A. Consanguinity, genetic drift and genetic diseases in populations with reduced numbers of founders. In: *Human Genetics - Principles and Approaches*. 4th ed. ; 2009:507-528.

63. Johnson EC, Evans LM, Keller MC. Relationships between estimated autozygosity and complex traits in the UK Biobank. Browning SR, ed. *PLoS Genet*. 2018;14(7):e1007556. doi:10.1371/journal.pgen.1007556

64. The BioBank Japan Project, Joshi PK, Esko T, et al. Directional dominance on stature and cognition in diverse human populations. *Nature*. 2015;523(7561):459-462. doi:10.1038/nature14618

65. McQuillan R, Eklund N, Pirastu N, et al. Evidence of Inbreeding Depression on Human Height. Gibson G, ed. *PLoS Genet*. 2012;8(7):e1002655. doi:10.1371/journal.pgen.1002655

66. Mukherjee S, Guha S, Ikeda M, et al. Excess of homozygosity in the major histocompatibility complex in schizophrenia. *Human Molecular Genetics*. 2014;23(22):6088-6095. doi:10.1093/hmg/ddu308

67. Postma E, Martini L, Martini P. Inbred women in a small and isolated Swiss village have fewer children: Inbred women have fewer children. *Journal of Evolutionary Biology*. 2010;23(7):1468-1474. doi:10.1111/j.1420-9101.2010.02013.x

68. Robert A, Toupance B, Tremblay M, Heyer E. Impact of inbreeding on fertility in a pre-industrial population. *Eur J Hum Genet*. 2009;17(5):673-681. doi:10.1038/ejhg.2008.237

69. Álvarez G, Ceballos FC, Berra TM. Darwin was right: inbreeding depression on male fertility in the Darwin family: Inbreeding Depression in the Darwin Family. *Biol J Linn Soc Lond*. 2015;114(2):474-483. doi:10.1111/bij.12433

70. Wang S, Haynes C, Barany F, Ott J. Genome-wide autozygosity mapping in human populations. *Genet Epidemiol*. 2009;33(2):172-180. doi:10.1002/gepi.20344

71. Lander E, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. 1987;236(4808):1567-1570. doi:10.1126/science.2884728

72. Christofidou P, Nelson CP, Nikpay M, et al. Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages. *The American Journal of Human Genetics*. 2015;97(2):228-237. doi:10.1016/j.ajhg.2015.06.001

73. McLaughlin RL, Kenna KP, Vajda A, et al. Homozygosity mapping in an Irish ALS case–control cohort describes local demographic phenomena and points towards potential recessive risk loci. *Genomics*. 2015;105(4):237-241. doi:10.1016/j.ygeno.2015.01.002

74. Keller MC, Simonson MA, Ripke S, et al. Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. Gibson G, ed. *PLoS Genet*. 2012;8(4):e1002656. doi:10.1371/journal.pgen.1002656

75. Wolf SM. Return of individual research results and incidental findings: facing the challenges of translational science. *Annu Rev Genomics Hum Genet*. 2013;14:557-577. doi:10.1146/annurev-genom-091212-153506

76. Fabsitz RR, McGuire A, Sharp RR, et al. Ethical and Practical Guidelines for Reporting Genetic Research Results to Study Participants: Updated Guidelines From a National Heart, Lung, and Blood Institute Working Group. *Circ Cardiovasc Genet*. 2010;3(6):574-580. doi:10.1161/CIRCGENETICS.110.958827

77. Bombard Y, Brothers KB, Fitzgerald-Butt S, et al. The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results. *The American Journal of Human Genetics*. 2019;104(4):578-595. doi:10.1016/j.ajhg.2019.02.025

78. Oniya O, Neves K, Ahmed B, Konje JC. A review of the reproductive consequences of consanguinity. *Eur J Obstet Gynecol Reprod Biol*. 2019;232:87-96. doi:10.1016/j.ejogrb.2018.10.042

79. Bittles AH, Black ML. Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences*. 2010;107(suppl_1):1779-1786. doi:10.1073/pnas.0906079106

80. Van Hout CV, Tachmazidou I, Backman JD, et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. 2020;586(7831):749-756. doi:10.1038/s41586-020-2853-0

81. Amendola LM, Dorschner MO, Robertson PD, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res*. 2015;25(3):305-315. doi:10.1101/gr.183483.114

82. Miller DT, Lee K, Gordon AS, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23(8):1391-1398. doi:10.1038/s41436-021-01171-4

83. Olfson E, Cottrell CE, Davidson NO, et al. Identification of Medically Actionable Secondary Findings in the 1000 Genomes. *PLoS One*. 2015;10(9):e0135193. doi:10.1371/journal.pone.0135193

84. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2016;18(7):696-704. doi:10.1038/gim.2015.148

85. Strauss KA, Williams KB, Carson VJ, et al. Glutaric acidemia type 1: Treatment and outcome of 168 patients over three decades. *Molecular Genetics and Metabolism*. 2020;131(3):325-340. doi:10.1016/j.ymgme.2020.09.007

86. Bowser LE, Young M, Wenger OK, et al. Recessive GM3 synthase deficiency: Natural history, biochemistry, and therapeutic frontier. *Molecular Genetics and Metabolism*. 2019;126(4):475-488. doi:10.1016/j.ymgme.2019.01.013

87. Fox MD, Carson VJ, Feng H-Z, et al. TNNT1 nemaline myopathy: natural history and therapeutic frontier. *Human Molecular Genetics*. 2018;27(18):3272-3282. doi:10.1093/hmg/ddy233

88. Morton DH, Morton CS, Strauss KA, et al. Pediatric medicine and the genetic disorders of the Amish and Mennonite people of Pennsylvania: AMERICAN JOURNAL OF MEDICAL GENETICS (SEMIN. MED. GENET.). *Am J Med Genet*. 2003;121C(1):5-17. doi:10.1002/ajmg.c.20002

89. Xu H, Ryan KA, Jaworek TJ, et al. Familial Hypercholesterolemia and Type 2 Diabetes in the Old Order Amish. *Diabetes*. 2017;66(7):2054-2058. doi:10.2337/db17-0173

90. Montasser ME, Hout CVV, McFarland R, et al. *Genetic and Functional Evidence Relates a Missense Variant in* B4GALT1 *to Lower LDL-C and Fibrinogen*. Genetics; 2019. doi:10.1101/721704

91. Kelly MA, Caleshu C, Morales A, et al. Adaptation and validation of the ACMG/AMP variant classification framework for MYH7-associated inherited cardiomyopathies: recommendations by ClinGen's Inherited Cardiomyopathy Expert Panel. *Genet Med*. 2018;20(3):351-359. doi:10.1038/gim.2017.218

92. Oza AM, DiStefano MT, Hemphill SE, et al. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat*. 2018;39(11):1593-1613. doi:10.1002/humu.23630

93. Mester JL, Ghosh R, Pesaran T, et al. Gene-specific criteria for PTEN variant curation: Recommendations from the ClinGen PTEN Expert Panel. *Hum Mutat*. 2018;39(11):1581-1592. doi:10.1002/humu.23636

94. Johnston JJ, Dirksen RT, Girard T, et al. *Variant Curation Expert Panel Recommendations for RYR1 Pathogenicity Assertions in Malignant Hyperthermia Susceptibility*. Genetics; 2020. doi:10.1101/2020.11.29.402768

95. Fortuno C, Lee K, Olivier M, et al. Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. *Hum Mutat*. Published online December 10, 2020. doi:10.1002/humu.24152

96. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018;39(11):1517-1524. doi:10.1002/humu.23626

97. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12(1):3. doi:10.1186/s13073-019-0690-2

98. *US Congress (1988). Clinical Laboratory ImprovementAmendments of 1988. Public Law 100–578, 102 Stat. 2903;Codified as Amended at 42 USCx263a (2006); See Regulationsat 42 CFR 493. Available from: Https://Www.Cdc.Gov/Clia/Law-Regulations.Html.* https://www.cdc.gov/clia/law-regulations.html

99. Jarvik GP, Amendola LM, Berg JS, et al. Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *Am J Hum Genet*. 2014;94(6):818-826. doi:10.1016/j.ajhg.2014.04.009

100. Burke W, Evans BJ, Jarvik GP. Return of results: ethical and legal distinctions between research and clinical care. *Am J Med Genet C Semin Med Genet*. 2014;166C(1):105-111. doi:10.1002/ajmg.c.31393

101. Sobel ME, Dreyfus JC, Dillehay McKillip K, et al. Return of Individual Research Results: A Guide for Biomedical Researchers Utilizing Human Biospecimens. *Am J Pathol*. 2020;190(5):918-933. doi:10.1016/j.ajpath.2020.01.014

102. Wang H, Xin B. Hypertrophic cardiomyopathy in the Amish community — What we may learn from it. *Progress in Pediatric Cardiology*. 2011;31(2):129-134. doi:10.1016/j.ppedcard.2011.02.011

103. Xin B, Puffenberger E, Tumbush J, Bockoven JR, Wang H. Homozygosity for a novel splice site mutation in the cardiac myosin-binding protein C gene causes severe neonatal hypertrophic cardiomyopathy. *Am J Med Genet A*. 2007;143A(22):2662-2667. doi:10.1002/ajmg.a.31981

104. ACMG Board of Directors. The use of ACMG secondary findings recommendations for general population screening: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2019;21(7):1467-1468. doi:10.1038/s41436-019-0502-5

105. Gonzaga-Jauregui C, Ge W, Staples J, et al. Clinical and Molecular Prevalence of Lipodystrophy in an Unascertained Large Clinical Care Cohort. *Diabetes*. 2020;69(2):249-258. doi:10.2337/db19-0447

106. Hershberger RE, Morales A. LMNA-Related Dilated Cardiomyopathy. :17.

107. Curtin SC. Pregnancy Rates for U.S. Women Continue to Drop. 2013;(136):8.

108. Cohain JS, Buxbaum RE, Mankuta D. Spontaneous first trimester miscarriage rates per woman among parous women with 1 or more pregnancies of 24 weeks or more. *BMC Pregnancy Childbirth*. 2017;17(1):437. doi:10.1186/s12884-017-1620-1

109. Stanley KE, Giordano J, Thorsten V, et al. Causal Genetic Variants in Stillbirth. *N Engl J Med*. 2020;383(12):1107-1116. doi:10.1056/NEJMoa1908753

110. Wapner RJ. Genetics of Stillbirth. *Clinical Obstetrics & Gynecology*. 2010;53(3):628-634. doi:10.1097/GRF.0b013e3181ee2793

111. Korteweg FJ, Bouman K, Timmer A, Holm JP. Cytogenetic Analysis After Evaluation of 750 Fetal Deaths. 2008;111(4):10.

112. Wilkins-Haug L. Genetic innovations and our understanding of stillbirth. *Hum Genet*. 2020;139(9):1161-1172. doi:10.1007/s00439-020-02146-2

113. Butler DC, Glen WB, Schandl C, Phillips A. Glycogen Storage Disease Type IV Diagnosed at Fetal Autopsy. *Pediatr Dev Pathol*. 2020;23(4):301-305. doi:10.1177/1093526619890224

114. Zhao C, Chai H, Zhou Q, et al. Exome sequencing analysis on products of conception: a cohort study to evaluate clinical utility and genetic etiology for pregnancy loss. *Genet Med*. 2021;23(3):435-442. doi:10.1038/s41436-020-01008-6

115. Martin HC, Jones WD, McIntyre R, et al. Quantifying the contribution of recessive coding variation to developmental disorders. Published online 2018:5.

116. Abouelhoda M, Sobahy T, El-Kalioby M, et al. Clinical genomics can facilitate countrywide estimation of autosomal recessive disease burden. *Genet Med*. 2016;18(12):1244-1249. doi:10.1038/gim.2016.37

117. Nakatsuka N, Moorjani P, Rai N, et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat Genet*. 2017;49(9):1403-1407. doi:10.1038/ng.3917

118. Xiao Q, Lauschke VM. The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders. *npj Genom Med*. 2021;6(1):41. doi:10.1038/s41525-021-00203-x

119. Bittles A. Consanguinity and its relevance to clinical genetics: Consanguinity and its relevance to clinical genetics. *Clinical Genetics*. 2001;60(2):89-98. doi:10.1034/j.1399-0004.2001.600201.x

120. Tamim H. Consanguinity and Apnea of Prematurity. *American Journal of Epidemiology*. 2003;158(10):942-946. doi:10.1093/aje/kwg226

121. Chakraborty R, Chakravarti A. On consanguineous marriages and the genetic load. *Hum Genet*. 1977;36(1):47-54. doi:10.1007/BF00390435

122. Vora KS, Modi MP, Butala BP, Shah VR. Anesthetic management of two cases of Bardet-Biedl syndrome for renal transplantation. *Saudi J Kidney Dis Transpl*. 2017;28(2):384-387. doi:10.4103/1319-2442.202775

123. Saldino RM, Noonan CD. Severe thoracic dystrophy with striking micromelia, abnormal osseous development, including the spine, and multiple visceral anomalies. *Am J Roentgenol Radium Ther Nucl Med*. 1972;114(2):257-263. doi:10.2214/ajr.114.2.257

124. Akbar AZ. Second reported case of Short-Rib-Polydactyly type III in China with a novel mutation of DYNC2H1 gene detected by Next Generation Sequencing. 2019;10(1):4.

125. Qiao Y, Wen J, Tang F, et al. Whole exome sequencing in recurrent early pregnancy loss. *Mol Hum Reprod*. 2016;22(5):364-372. doi:10.1093/molehr/gaw008

126. Li CC. Segregation of the Ellis-van Creveld Syndrome as Analyzed by the First Appearance Method. :9.

127. Dixon PH, Sambrotta M, Chambers J, et al. An expanded role for heterozygous mutations of ABCB4, ABCB11, ATP8B1, ABCC2 and TJP2 in intrahepatic cholestasis of pregnancy. *Sci Rep*. 2017;7(1):11823. doi:10.1038/s41598-017-11626-x

128. Lord J, McMullan DJ, Eberhardt RY, et al. Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. *The Lancet*. 2019;393(10173):747-757. doi:10.1016/S0140-6736(18)31940-8

129. Streeten EA, Mcbride DJ, Lodge AL, et al. Reduced Incidence of Hip Fracture in the Old Order Amish. *J Bone Miner Res*. 2003;19(2):308-313. doi:10.1359/JBMR.0301223

130. He S, Ryan KA, Streeten EA, et al. Prevalence, control, and treatment of diabetes, hypertension, and high cholesterol in the Amish. *BMJ Open Diab Res Care*. 2020;8(1):e000912. doi:10.1136/bmjdrc-2019-000912

131. O'Connell J. *Mixed Model Analysis for Pedigrees and Populations*. https://mmap.github.io./

132. Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017;544(7649):235-239. doi:10.1038/nature22034

133. Szpiech ZA, Xu J, Pemberton TJ, et al. Long Runs of Homozygosity Are Enriched for Deleterious Variation. *The American Journal of Human Genetics*. 2013;93(1):90-102. doi:10.1016/j.ajhg.2013.05.003

134. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaSci*. 2015;4(1):7. doi:10.1186/s13742-015-0047-8

135. McQuillan R, Leutenegger A-L, Abdel-Rahman R, et al. Runs of Homozygosity in European Populations. *The American Journal of Human Genetics*. 2008;83(3):359-372. doi:10.1016/j.ajhg.2008.08.007

136. Szpiech ZA, Blant A, Pemberton TJ. GARLIC: Genomic Autozygosity Regions Likelihood-based Inference and Classification. Berger B, ed. *Bioinformatics*. 2017;33(13):2059-2062. doi:10.1093/bioinformatics/btx102

137. Nothnagel M, Lu TT, Kayser M, Krawczak M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics*. 2010;19(15):2927-2935. doi:10.1093/hmg/ddq198

138. Beja-Pereira A, Luikart G, England PR, et al. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet*. 2003;35(4):311-313. doi:10.1038/ng1263

139. Howrigan DP, Simonson MA, Davies G, et al. Genome-wide autozygosity is associated with lower general cognitive ability. *Mol Psychiatry*. 2016;21(6):837-843. doi:10.1038/mp.2015.120

140. Sheth F, Ewers E, Kosyakova N, et al. A small supernumerary marker chromosome present in a Turner syndrome patient not derived from X- or Y-chromosome: a case report. *Mol Cytogenet*. 2009;2(1):22. doi:10.1186/1755-8166-2-22

141. McArdle PF, Whitcomb BW, Tanner K, Mitchell BD, Shuldiner AR, Parsa A. Association between bilirubin and cardiovascular disease risk factors: using Mendelian randomization to assess causal inference. *BMC Cardiovasc Disord*. 2012;12(1):16. doi:10.1186/1471-2261-12-16

142. Strassburg CP, Oldhafer K, Manns MP, Tukey RH. Differential Expression of the *UGT1A* Locus in Human Liver, Biliary, and Gastric Tissue: Identification of *UGT1A7* and *UGT1A10* Transcripts in Extrahepatic Tissue. *Mol Pharmacol*. 1997;52(2):212-220. doi:10.1124/mol.52.2.212

143. Lv X, Wang X-X, Hou J, et al. Comparison of the inhibitory effects of tolcapone and entacapone against human UDP-glucuronosyltransferases. *Toxicology and Applied Pharmacology*. 2016;301:42-49. doi:10.1016/j.taap.2016.04.009

144. Wheeler E, Leong A, Liu C-T, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. Gregg E, ed. *PLoS Med*. 2017;14(9):e1002383. doi:10.1371/journal.pmed.1002383