Family-specific aggregation of lipid GWAS variants confers the susceptibility to familial hypercholesterolemia in a large Austrian family

Elina Nikkola [a], Arthur Ko [a, b], Marcus Alvarez [a], Rita M. Cantor [a], Kristina Garske [a], Elliot Kim [a], Stephanie Gee [a], Alejandra Rodriguez [a], Reinhard Muxel [c], Niina Matikainen [d, e, f], Sanni Söderlund [e, f], Mahdi M. Motazacker [g], Jan Borén [h], Claudia Lamina [i], Florian Kronenberg [i], Wolfgang J. Schneider [j], Aarno Palotie [k, l, m], Markku Laakso [n], Marja-Riitta Taskinen [e, f], Päivi Pajukanta [a, b, o, *]

[a] Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA
[b] Molecular Biology Institute at UCLA, Los Angeles, USA
[c] General Practice Centre, Austria
[d] Endocrinology, Abdominal Centre, Helsinki University Hospital, Finland
[e] Heart and Lung Center, Helsinki University Hospital, Finland
[f] Research Programs Unit, Diabetes and Obesity, University of Helsinki, Finland
[g] Department of Clinical Genetics, Academic Medical Center at the University of Amsterdam, The Netherlands
[h] Department of Molecular and Clinical Medicine/Wallenberg Laboratory, University of Gothenburg and Sahlgrenska University Hospital, Sweden
[i] Division of Genetic Epidemiology, Medical University of Innsbruck, Austria
[j] Department Medical Biochemistry, Medical University Vienna and Max F. Perutz Laboratories, Austria
[k] Institute for Molecular Medicine, University of Helsinki, Finland
[l] The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[m] Massachusetts General Hospital, Boston, MA, USA
[n] Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
[o] Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA

A R T I C L E   I N F O

A B S T R A C T

Background and aims: Hypercholesterolemia confers susceptibility to cardiovascular disease (CVD). Both serum total cholesterol (TC) and LDL-cholesterol (LDL-C) exhibit a strong genetic component (heritability estimates 0.41—0.50). However, a large part of this heritability cannot be explained by the variants identified in recent extensive genome-wide association studies (GWAS) on lipids. Our aim was to find genetic causes leading to high LDL-C levels and ultimately CVD in a large Austrian family presenting with what appears to be autosomal dominant inheritance for familial hypercholesterolemia (FH).
Methods: We utilized linkage analysis followed by whole-exome sequencing and genetic risk score analysis using an Austrian multi-generational family with various dyslipidemias, including elevated TC and LDL-C, and one family branch with elevated lipoprotein (a) (Lp(a)).
Results: We did not find evidence for genome-wide significant linkage for LDL-C or apparent causative variants in the known FH genes rather, we discovered a particular family-specific combination of nine GWAS LDL-C SNPs ($p = 0.02$ by permutation), and putative less severe familial hypercholesterolemia mutations in the *LDLR* and *APOB* genes in a subset of the affected family members. Separately, high Lp(a) levels observed in one branch of the family were explained primarily by the LPA locus, including short (<23) Kringle IV repeats and rs3798220.
Conclusions: Taken together, some forms of FH may be explained by family-specific combinations of LDL-C GWAS SNPs.

© 2017 Elsevier B.V. All rights reserved.

* Corresponding author. Department of Human Genetics, David Geffen School of Medicine at UCLA, Gonda Center, Room 6335B, 695 Charles E. Young Drive South, Los Angeles, CA 90095-7088, USA.
   E-mail address: ppajukanta@mednet.ucla.edu (P. Pajukanta).

1. Introduction

High levels of serum total cholesterol (TC) and especially low-density lipoprotein cholesterol (LDL-C) predispose to cardiovascular disease (CVD), the major cause of death worldwide [1]. Genetics plays a major role in CVD (heritability estimates 0.38–0.57) [2,3]. However, variants identified in extensive genome-wide association studies (GWAS) explain only 6–20% of the variance in lipid traits and even less of CVD [4]. This missing heritability may partially be explained by rare and private variants, and thus large families with several affected individuals without risk variants in the known familial hypercholesterolemia (FH) genes may help identify new genes causing Mendelian forms of dyslipidemia or other inherited mechanisms contributing to high LDL-C.

FH affects 1 in 200–600 people [5]. To date there are only a handful of genes known to cause FH, including *LDLR*, *APOB*, *PCSK9* and *LDLRAP1* [6]. However, it is estimated that only approximately 20–60% of FH subjects exhibit a causal variant within these four genes [7–9], suggesting that variants in these genes do not explain all cases of FH.

To find mutations leading to high LDL-cholesterol and ultimately CVD, we systematically screened both rare coding and common genomic variants in a large Austrian dyslipidemic family exhibiting elevated TC and LDL-C levels, in addition to elevated lipoprotein a (Lp(a)) levels in one branch of the family. All affected elderly family members had suffered a cardiovascular event in the past, and the index case did not have known FH variants in *LDLR* or *APOB*.

Combining linkage analysis with whole-exome sequencing has become a common approach to pinpoint candidate chromosomal regions and specific variants for Mendelian diseases [10,11]. We first genotyped the family members using a genome-wide SNP array to cover the common variants, and then exome-sequenced the family members to capture their coding variants. We screened for mutations in the known FH genes, performed a genome-wide linkage analysis, and assessed the coding variants present predominantly in the affected individuals for functional predictions. Since no genome-wide significant linkage peaks or mutations in the known FH genes were found, we estimated genetic risk scores using all common GWAS SNPs previously associated with LDL-C [12] and identified a family-specific combination of nine LDL-C GWAS variants, contributing to the high LDL-C levels in this family.

2. Materials and methods

2.1. Overview

We first searched for a possible monogenic cause for FH in a large Austrian pedigree using a linkage analysis, followed by an exome sequencing analysis and subsequent variant screening in existing European cohorts. We also comprehensively analyzed all variants identified in the known FH genes [6]. We then searched for a potential polygenic association with FH in this family by performing a genetic risk score analysis of the known LDL-C GWAS variants [12].

2.2. Study samples

The study sample consists of 16 individuals from a large Austrian dyslipidemic family (Fig. 1; for clinical characteristics, see Table 1). The diagnostic criteria of heterozygous FH were based on the MedPED and world health organization (WHO) criteria [13,14]. The index case was a 46-year old healthy man with TC level of 11.7 mmol/l and LDL-C level of 9.2 mmol/l measured on a routine check-up. His father has had premature coronary heart disease (CHD) before the age of 55 but had died accidentally. Thus, the index had a score >8 using the Dutch Lipid Clinic Network diagnostic criteria for FH. He had a low level of serum triglycerides and an elevated level of serum high-density lipoprotein cholesterol (HDL-C). These findings initiated the family screening, as two cousins from the paternal side also had early CHD events before the age of 55. Family members with LDL-C > 4.0 mmol/l with or without statin therapy were considered affected for FH. In addition, one family branch exhibited high Lp(a) levels (>50 mg/dl). DNA was extracted from blood, and clinical phenotypes were measured using standard protocols. Fasting serum samples from all available family members were ultracentrifuged to separate lipoprotein fractions [15], and cholesterol and triglyceride concentrations were measured by automated enzymatic methods in total serum and in VLDL, LDL and HDL fractions. The number of apolipoprotein (a) Kringle IV (apo(a) KIV) repeats was measured by SDS-agarose electrophoresis followed by immunoblotting, as described previously [16]. Lp(a) concentrations were measured by ELISA, as recently described [17]. Phenotypes included age, sex, status of statin medication, total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), Lp(a), and the number of apo(a) KIV repeats. All family members gave a written informed consent, and the study was approved by the local ethic committees.

2.3. Validation cohorts

To validate our findings, we utilized genome-wide genotyping data from the METabolic Syndrome In Men (METSIM) cohort (n = 10,197) [18] and the European exome sequencing database of type 2 diabetes consortium (n~13,000) for the association with LDL-C.

2.4. Genome-wide SNP genotyping and whole exome sequencing

We performed genotyping using a genome-wide SNP panel (Illumina HumanCoreExome-12v1-1), as well as exome sequencing of all available affected and unaffected family members, using the Agilent SureSelect All exon 50-Mb capture with the Illumina Hiseq2500 platform employing 75 bp paired-end reads, resulting in a mean coverage of 75X, and capturing ~91% of the targeted regions with ≥10X coverage. We aligned and called the variants using BWA [19] and GATK [20]. We used the hg19 genomic reference sequence for the alignment and all analyses. We checked the data quality, including the call rate of the SNPs, gender check based on X chromosome SNPs, and heterozygosity rate using PLINK [21] as well as pedigree consistency using the Mendel software packages [22].

2.5. Linkage analysis

We first estimated the expected maximum LOD score (EMLOD) based on the pedigree structure and binary LDL-C affection status using fastSLINK package [23], employing the same penetrance model as in the linkage analysis. We performed a genome-wide parametric two-point linkage analysis for LDL-C and Lp(a) using Mendel [22] and utilizing ~95 K high-quality (genotyped in all family members) and informative (MAF>10% in the family (>3 carriers)) SNPs, spaced ~25 kb apart throughout the genome. For LDL-C, we employed an autosomal dominant model with a penetrance of 0.95 and phenocopy rate of 0.001. For Lp(a), we used an autosomal dominant model with penetrance of 0.99 and a phenocopy rate 0.0001 to test the variants at the *LPA* locus.
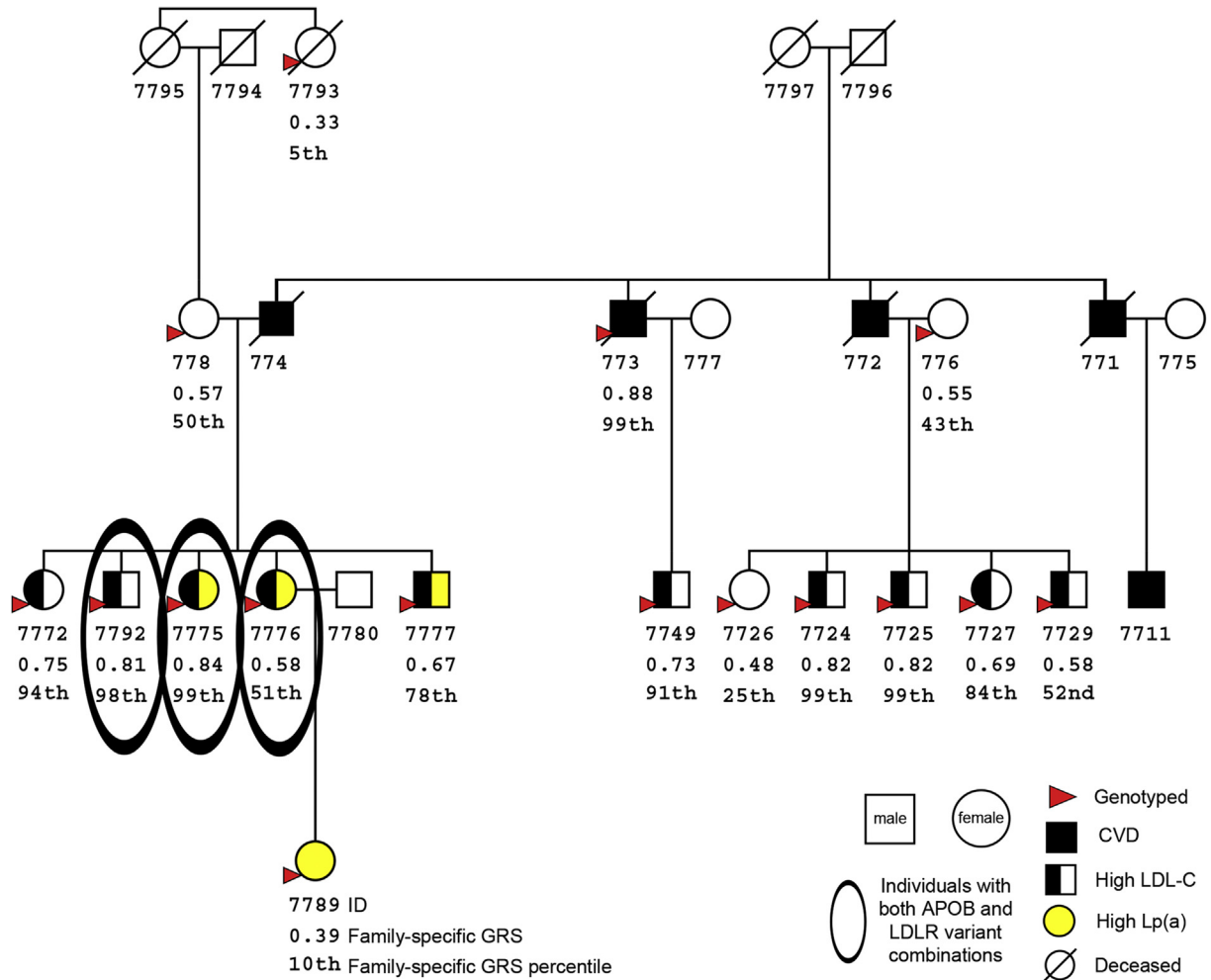
Fig. 1. The Austrian hypercholesterolemia family showing an autosomal dominant type of inheritance.
The figure includes only those family members who gave an informed consent for blood sampling and DNA analyses. The DNA sample from the affected family member 7711 did not pass the quality requirements for the genomic analyses. The right bottom corner shows the explanations of the used signs: the circles indicate a female and the square a male; the filled circle or square indicates a person who has suffered from a previous cardiovascular event; the half-filled circle or square indicates an individual with high low-density lipoprotein cholesterol (LDL-C); yellow color indicates an individual with high lipoprotein (a) (Lp(a)); the red arrow shows individuals with available DNA; and the open squares or circles indicate unaffected subjects. The family-specific genetic risk score (GRS) and percentile are given under each individual. The individuals with specific APOB and LDLR variant combinations are circled in black (for the specific APOB and LDLR combinations, see Supplementary Table 3). The pedigree was drawn using CraneFoot [36]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. Variant filtering

We focused on the potentially functional variants (non-synonymous and splice site variants) fulfilling the following criteria: minor allele frequency (MAF) $\leq 10\%$; location in the regions with an LOD score ≥ 1.0; and present predominantly in the affected individuals (high LDL-C or high Lp(a)).

2.7. Genetic risk score analysis

We calculated weighted genetic risk scores of the 65 known common GWAS LDL-C variants [12] in the METSIM cohort (n = 10,064) and family members. For each LDL-C associated locus (>1 Mb apart), we selected the SNP with the lowest p-value and weighed each risk allele with the beta effect size established by Willer et al. from ~180,000 individuals [12]. The selected SNPs, including their risk alleles and weights, are listed in Supplementary Table 1. We first calculated the risk scores for each individual in the METSIM cohort, and then compared the risk scores of the affected family members with the estimated population percentiles

obtained in the METSIM cohort.

2.8. Permutation analyses

To assess the significance of the difference in average number of risk alleles we observed between the affected and non-affected family members with the nine family-specific GWAS SNPs, we performed a permutation analysis for the nine SNPs by randomly selecting 20 individuals with LDL-C >75th percentile (the LDL-C cut off ≥ 3.9 mmol/l) and 20 individuals with LDL-C <50th percentile (the LDL-C cut off ≤ 3.5 mmol/l) from the METSIM cohort. We calculated how many times the difference in average number of risk alleles is larger for all nine SNPs between the METSIM individuals with high LDL-C and normal LDL-C than the 20% difference observed for all nine SNPs in the family.

To assess the significance of the observed risk SNP combination, we performed an additional permutation of the risk scores by randomly selecting 100 times a 9-SNP set from the 65 LDL-C increasing SNPs [12]. We constructed new risk scores weighted by beta and estimated the percentiles in the METSIM population for

Table 1
Clinical characteristics and genetic findings contributing to high LDL-C in the family members.

Individual ID	TC (mmol/l)	LDL-C (mmol/l)	Statin usage	Lp(a) (mg/dl)	Number of Kringle IV repeats, Allele 1	Number of Kringle IV repeats, Allele 2	Known LPA variant (rs3798220), risk allele C, MAF=0.02	Genetic risk score[c]	Family-specific genetic risk score of 9 SNPs (see Table 2)	Both APOB and LDLR variant combinations[d]
7711	(4.32)[a]	(2.16)[a]	Yes	4	24 (85%)[b]	36		3.84 (89th)	0.75 (94th)	
7772	7.57	5.11	No	4	26 (10%)[b]	36	T/T	3.66 (69th)	0.81 (98th)	
7792	7.77	5.45	No	7	23 (80%)[b]	36	T/T	3.69 (73th)	0.84 (99th)	Yes
7775	8.98	5.72	No	77	20	0	T/C	3.41 (34th)	0.58 (51st)	Yes
7776	6.80 (6.09)[a]	4.20 (3.56)[a]	Yes	90	20	0	T/C	3.71 (76th)	0.67 (78th)	Yes
7777	**11.70 (5.06)[a]**	**9.20 (1.94)[a]**	**Yes**	**113**	**20**	**0**	**T/C**	**3.70 (75th)**	**0.88 (99th)**	
773	(4.52)[a]	(2.96)[a]	Yes	9	36	0	T/T	3.53 (51th)	0.73 (91th)	
7749	7.60 (4.92)[a]	5.00 (2.87)[a]	Yes	4	23 (80%)[b]	36	T/T	3.84 (88th)	0.82 (99th)	
7724	6.80 (5.93)[a]	5.00 (3.90)[a]	Yes	6	36	0	T/T	3.70 (74th)	0.82 (99th)	
7725	7.00 (5.57)[a]	5.36 (3.19)[a]	Yes	2	36	0	T/T	4.01 (97th)	0.69 (84th)	
7727	6.64	4.54	No	10	36	0	T/T	3.76 (81th)	0.58 (52nd)	
7729	9.60	6.38	No	3	30	0	T/T	3.24 (15th)	0.33 (5th)	
7793	4.38	2.57	No	0	0	0	T/T	3.81 (86th)	0.57 (50th)	
778	5.03	2.40	No	5	23 (95%)[b]	26	T/T	3.68 (72nd)	0.55 (43th)	
776	7.20	3.32	No	11	30 (85%)[b]	36	T/T	3.66 (70th)	0.48 (25th)	
7726	5.92	2.82	No	7	30	0	T/T	3.42 (35th)	0.39 (10th)	
7789	4.49	2.34	No	66	20	0	T/C			

The index case (7777) is bolded and circled in red. The grey highlight indicates an individual who has high LDL-C. The affected individuals are sorted by nuclear families. TC, total cholesterol; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein (a); MAF, minor allele frequency
[a] TC or LDL-C values in parenthesis are measured while on a variable dose of different statins without any other lipid lowering drugs; [b] the value in parenthesis is the percentage of Kringle IV repeat allele 1; [c] weighted genetic risk score of the 65 LDL-C GWAS SNPs, the percentile is indicated in the parenthesis; [d] specific APOB and LDLR variants listed in Supplementary Table 3.

each of the 9-SNP sets. We then calculated how many times the average risk score of the affected individuals would be in the $\geq 90^{th}$ percentile.

2.9. Evaluation of the known FH variants and genes

We screened all individuals for variants in the four previously known FH genes, *LDLR*, *APOB*, *PSCK9*, and *LDLRAP1* [6].

3. Results

In this study, we aimed to identify the variant(s) for an autosomal dominant type of inheritance of high LDL-C levels in a large Austrian dyslipidemic family. The affected family members had an average pre-statin LDL-C level of 5.56 mmol/l (range 4.20–9.20 mmol/l), and four siblings from the first generation had suffered a cardiovascular event (Fig. 1). One branch of the family also exhibited 4 individuals with extremely high Lp(a) levels (66 mg/dl-113 mg/dl, i.e. all above the 90th percentile), a known independent risk factor for CVD [24,25]. We first performed a linkage study followed by exome-sequencing analysis to find novel variants co-segregating with the high LDL-C status in the family. Lp(a) levels were investigated for variants at the *LPA* locus. We evaluated our identified LDL-C variants for association in existing large European cohorts, and calculated the weighted genetic risk scores for LDL-C

using genome-wide significant LDL-C variants from the Willer et al. *meta*-GWAS study [12], utilizing the METSIM cohort as a reference panel. Lastly, we systematically screened all variants we identified in the known FH genes, *LDLR*, *APOB*, *PSCK9*, and *LDLRAP1*, for co-segregation with high LDL-C status among the family members.

3.1. Linkage analysis followed by a variant screening did not pinpoint a locus for high LDL-C

The estimated maximum LOD score for this family was 3.3 using fastSLINK. We observed 22 loci with an LOD score >1.0, with the highest maximum LOD score of 2.1 on chromosome 5. The fine-mapping of the chromosome 5 region, did not result in LOD scores >2.1. Within the 22 loci, we identified 6 potential functional variants (Fig. 2A and Supplementary Table 2). Even though separately none of these variants were robustly associated with quantitative LDL-C in the large European cohorts (*p*-values>0.008), four of the 11 affected had all six risk alleles and two of the four were on statin therapy.

3.2. Genetic risk score analysis of known LDL-C loci identified a family-specific combination of nine risk variants

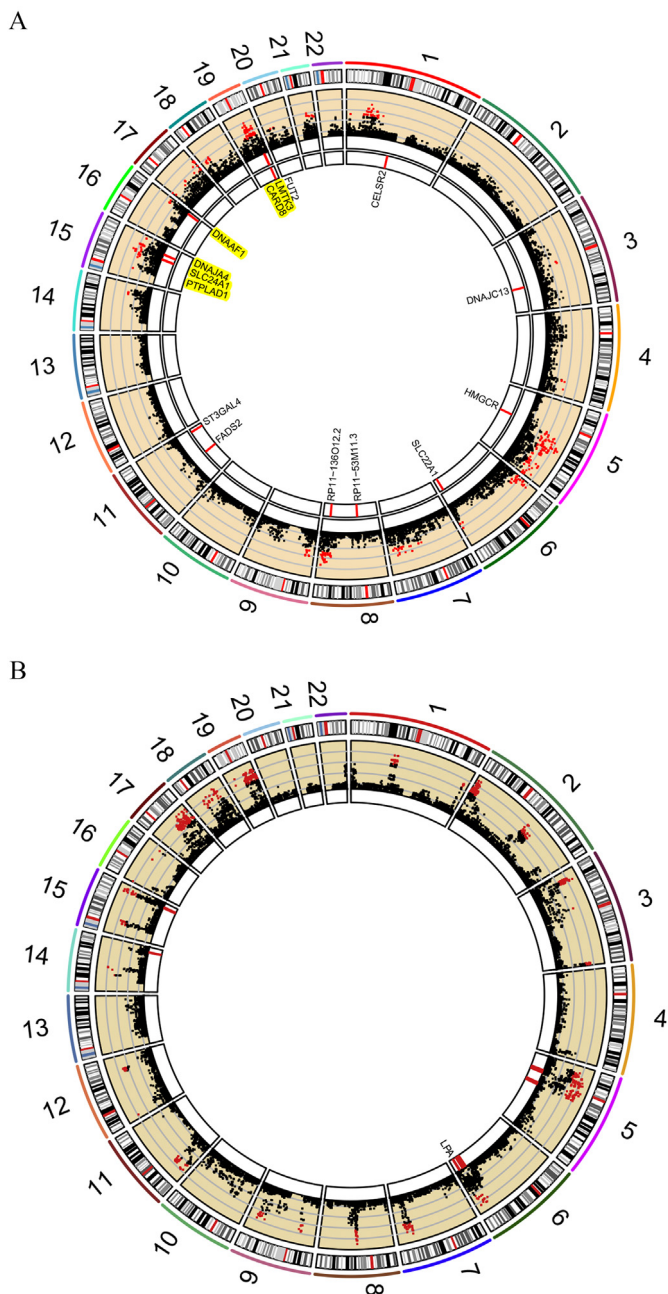Out of the 69 independent LDL-C associated GWAS risk loci

A

B

Fig. 2. Overview of the genetic results.
(A) Overlap between the 22 LDL-C regions with an LOD score >1.0, exome variants (non-synonymous or frameshift variants with a MAF<10%), and 9 family-specific LDL-C GWAS variants identified in the Austrian family members, as illustrated by rCircos [37]. The outer most track indicates the chromosome number, followed by the cytoband; the scatter plot shows the LOD scores of the ~95 K SNPs from the linkage analysis (red indicates a LOD score >1.0 using a scale of 0–2.25); next to the scatter plot are the exome variants predominantly present in the affected family members (Supplementary Table 2) that reside in the regions with LOD>1.0; the inner most track indicates the 9 family-specific GWAS SNPs (Table 2), and the gene names (or the closest gene) of the variants are shown inside of the circle; the yellow highlight indicates that the variant was identified by exome sequencing. (B) Overlap between the Lp(a) regions with an LOD score >1.0 and exome variants (potentially functional and MAF<10%) identified in the Austrian family members, as illustrated by rCircos [37]. The outer most track indicates the chromosome number, followed by the cytoband; the scatter plot shows the LOD scores of the ~95 K SNPs from the linkage analysis (red indicates a LOD >1.0, using a scale of 0–1.5); and the inner most track indicates the exome variants present only in the family members with high Lp(a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

identified by Willer et al. [12], 65 SNPs (or their linkage disequilibrium (LD) proxies, $r^2 > 0.95$) were successfully genotyped or imputed for 10,064 METSIM individuals and 16 family members. If the GWAS lead SNP or its LD proxies were not available, we selected the second lowest p-value available within the GWAS locus. We constructed the overall genetic risk scores by calculating the sum of number of risk alleles, weighted by the beta established by Willer et al. of each of the 65 SNPs for every individual. The weighted LDL-C risk score observed in the METSIM cohort was correlated with serum LDL-C levels (Pearson's correlation = 0.22, $p < 2.2 \times 10^{-16}$), after removing statin users (n = 2812). For the calculations of the population percentiles of the genetic risk scores, we included all METSIM participants (n = 10,064). The 50th percentile of the LDL-C genetic risk scores in the METSIM cohort was 3.52, whereas the average of the affected family members was 3.71 (~75th percentile), suggesting a stronger predisposition for high LDL-C in the family based on the common LDL-C GWAS variants.

To further investigate the LDL-C GWAS risk variants observed in this family, we investigated the nine variants with the highest difference in the average number of risk alleles (>0.40) between the affected and non-affected family members (Table 2) for family-specific effects. To evaluate if this 9-SNP combination was indeed family-specific, we first performed a permutation analysis by randomly selecting 20 individuals with high LDL-C (>75th percentile) and 20 individuals with LDL-C <50th percentile from the METSIM cohort, and observed no similar difference in the average number of risk alleles between all of the 9 LDL-C GWAS SNPs among the subjects with low and high LDL-C levels, using 100 permutations ($p < 0.01$). This suggests that the distinct combination with a large difference in the average number of risk alleles with these 9 SNPs is specific for this family. Next, we derived the risk scores using the sum of the weighted betas of these 9 SNPs (Table 2). The new 9-SNP-weighted LDL-C risk score of the METSIM participants was correlated with serum LDL-C levels (Pearson's correlation = 0.12, $p < 2.2 \times 10^{-16}$), after removing statin users. As above, for calculation of the population percentiles of the genetic risk scores, we included all METSIM participants (n = 10,064). The average risk score in the METSIM population sample was 0.34 (50th percentile), whereas the average risk score of the affected family members was 0.74 (>90th percentile) and of the unaffected family members 0.46 (<25th percentile) (Table 1 and Fig. 1), respectively, further suggesting that the combination of the 9 SNPs is contributing to the high LDL-C levels in this family. These nine SNPs did not have a significantly higher effect size when compared to the rest of the 56 SNPs (Wilcoxon–Mann–Whitney test $p = 0.22$), demonstrating that their effect sizes do not differ from typical GWAS variants.

To determine whether this type of aggregation would appear by chance, selecting any set of LDL-C-raising GWAS SNPs, we first calculated the risk scores using the well-established Global Lipid Genetic Consortium 12-SNP LDL-C gene score calculation [26]. Indeed, the average risk score of the affected family members was in the 70th percentile and the average of the unaffected family members was in the 40th percentile when compared to the percentiles in the Whitehall II controls [26]. When we estimated the 12- SNP GRS [26] in the METSIM study, it correlated with serum LDL-C levels (Pearson's correlation = 0.23, $p < 2.2 \times 10-16$), similarly to using all of the 65 LDL-C increasing genome-wide significant GWAS SNPs (Pearson's correlation = 0.22, $p < 2.2 \times 10-16$), indicating that both of these genetic risk scores can equally predict the LDL-C levels at the population level. However, the average 12-SNP risk score of the affected family members is in the 65th percentile, suggesting that the discovered family-specific risk score is more suitable and accurate in this particular family. Three of the SNPs overlap between the 12-SNP GRS and family specific risk

Table 2
The nine family-specific LDL-C GWAS variants.

Chr.	Genomic location in base pairs (hg19)	Variant rs number	Variant type	Gene	Risk allele	Alternative allele	Effect size of the variant given as a beta value[a]	GWAS p-value[b]	GWAS freq[c]	Family members with high LDL-C											Average # of risk alleles in affected per variant	Family members with normal LDL-C					Average # of risk alleles in unaffected per variant	Difference in average # of risk alleles per variant
										7772	7792	7775	7776	7777	773	7749	7724	7725	7727	7729		7793	778	776	7726	7789		
1	109,818,530	rs646776	Intergenic	CELSR2[d]	T	C	0.160	1.6E-272	0.79	2	2	2	2	2	2	2	2	2	2	2	2.00	1	2	2	2	1	1.60	0.40
3	132,163,200	rs17404153	Intronic	DNAJC13	G	T	0.034	1.83E-09	0.86	2	2	2	2	2	2	2	1	1	1	1	1.64	2	2	0	1	1	1.20	0.44
5	74,656,539	rs12916	3'-UTR	HMGCR	C	T	0.073	7.79E-78	0.43	2	2	2	1	0	2	1	2	2	2	2	1.64	1	1	1	1	0	0.80	0.84
6	160,578,860	rs1564348	Intronic	SLC22A1	C	T	0.048	2.76E-21	0.15	1	1	0	0	0	1	1	1	1	1	0	0.64	0	0	1	0	0	0.20	0.44
8	55,421,614	rs10102164	Intergenic	RP11-53M11.3	A	G	0.032	3.74E-11	0.17	1	2	1	1	2	0	0	1	0	0	0	0.73	0	1	0	0	0	0.20	0.53
8	126,490,972	rs2954029	Intergenic	RP11-136O12.2	A	T	0.056	2.10E-50	0.53	1	1	1	1	1	2	2	1	2	2	0	1.27	0	0	1	1	2	0.80	0.47
11	61,609,750	rs174583	Intronic	FADS2	C	T	0.052	7.00E-41	0.63	0	1	1	0	1	1	0	2	2	0	1	0.82	0	0	1	0	1	0.40	0.42
11	126,243,952	rs11220462	Intronic	ST3GAL4	A	G	0.059	6.61E-21	0.14	1	1	2	0	1	2	1	1	1	0	0	0.91	0	1	0	0	0	0.20	0.71
19	49,206,417	rs492602	Synonymous	FUT2	G	A	0.029	9.42E-14	0.43	1	0	2	1	2	1	2	1	0	1	1	1.09	1	1	0	0	1	0.60	0.49

The bold font indicates an individual with high LDL-C.

Chr., chromosome; LDL-C, low-density lipoprotein cholesterol; Risk, LDL-C increasing variant; GWAS, genome-wide association study.

[a] beta value, [b] p-value, and [c] freq, frequency of the risk variant obtained from Willer et al. [12]; [d] if intergenic, the closest gene is given.

Numbers in the matrix under the individual IDs represents number of risk alleles in that particular variant.

scores. We further performed 100 permutations by randomly selecting 9-SNP combinations from the 65 LDL-C-increasing genome-wide GWAS SNPs and calculated how many times the average risk score of the affected family members is $\geq 90^{th}$ percentile. We observed this phenomenon only with two sets ($p = 0.02$). These additional risk score permutations suggest that randomly selecting other sets of LDL-C GWAS SNPs does not present as high a risk as the actual nine family-specific SNP combination.

3.3. High lipoprotein (a) (Lp(a)) levels are likely explained by the known genetic variants in the LPA locus

Lp(a) is largely regulated by the number of Kringle (IV) repeats and two independent SNPs (rs3798220, c.5793A > G p.(I1891M) and rs10455872, NM_005577.2:c.3947 + 467T > C), which together explain 30—70% of the Lp(a) variation [25]. In the family, the individuals with high Lp(a) also had low number of Kringle (IV) repeats (<23) (Table 1). Furthermore, we identified a well-known Lp(a) variant, rs3798220, in the *LPA* locus using linkage analysis (Fig. 2B). The intronic LPA variant rs10455872 did not segregate with the high Lp(a) levels. These data suggest that the high Lp(a) levels observed in a branch of the family are likely explained by rs3798220 and the low number of Kringle (IV) repeats.

3.4. Variants in the known FH genes may explain high LDL-C levels in one family branch

The index case had been previously screened negative for the known FH variants in FH genes (*LDLR*, *APOB*, *PCSK9* and *LDLRAP1*) by DNA sequencing. We screened all family members for the known FH genes and identified a total of 87 variants, of which 19 were nonsynonymous or splice site variants (Supplementary Table 3). None of the variants fully segregated with the high LDL-C status. However, we identified two splice site variants (rs72658867, NM_000527.4:c.2140+5G > A, MAF = 0.011 and rs72658861, NM_000527.4:c.1061-8T > C MAF = 0.0085) and one nonsynonymous *LDLR* variant rs45508991 (c.2177C > T(p.(T726I), MAF = 0.0095), predicted to be deleterious by SIFT in one family branch (Supplementary Table 3). This potentially deleterious variant rs45508991 is in full LD with the splice site variant rs72658861, both present in 3 of the 11 affected family members and in one of the 5 unaffected family members. All of these variants have been previously implicated in FH, but do not consistently co-segregate with hypercholesterolemia [27], suggesting that another variant must be present for these *LDLR* variants to be pathogenic, as previously proposed [28,29].

Similarly, we identified 3 potentially pathogenic non-synonymous variants (rs1801695, c.13569G > A(p.(A4481T), MAF = 0.033; rs61742247, c.4966 G > A p.(S1613T), MAF = 0.0011; and rs1801701, c.11041G > A p.(R3638Q), MAF = 0.090) (Supplementary Table 3) in *APOB*, of which rs1801701 has been implicated for LDL-C in a previous GWAS [12]. Interestingly, these *APOB* variants appeared in the same branch as the *LDLR* variants described above, with 3 affected family members sharing a combination of these *LDLR* and *APOB* variants (Table 1, Supplementary Table 3 and Fig. 1). We postulate that in order to have an impact on the ApoB metabolism, and furthermore on TC and LDL-C levels, these *LDLR* and *APOB* variants may need to appear as a risk combination or require other GWAS LDL-C variants as a haplotypic background. For example, one of the *APOB* variants (p.R3638Q) resides in the C-terminus of apoB100, a region known to be regulating LDL receptor binding [30].

4. Discussion

Our comprehensive analysis of a large Austrian family with phenotypical familial hypercholesterolemia (FH) showed evidence of a specific polygenic contribution to high LDL-C. The linkage analysis did not pinpoint to a single genetic locus for high LDL-C; rather, we found 22 loci with an LOD score >1.0, implying that it is likely several loci contribute to the high LDL-C levels in this family. Consistent with that, our risk score analyses followed by a permutation analysis identified a combination of nine LDL-C GWAS SNPs specific for polygenic FH in this family. In addition, a systematic examination of the variants in the known FH genes resulted in the identification of possible less severe FH mutations in the *LDLR* and *APOB* genes in a subset of the affected family members, in line with the previous hypothesis [28,29] that specific *LDLR* and *APOB* coding variants may only become pathogenic in the presence of an additional risk variant in these FH genes. Because three of the affected family members carried both *LDLR* and *APOB* risk combinations, we postulate that small functional defects in both genes, whose biological functions are tightly bound, escalate the effects and contribute to the high LDL-C levels in these individuals.

Recent evidence suggests that FH is a heterogeneous disorder that can be caused by monogenic or polygenic mechanisms, including rare variants at the traditional FH genes or multiple common variants at the LDL-C GWAS loci and other genes [7,31]. We did not identify FH-causing mutations in the known FH genes, and our linkage study combined with exome sequence analysis did not pinpoint a causative variant or gene. When we evaluated the effects of the weighted risk scores of 65 LDL-C GWAS SNPs collectively present in the family members, we observed that the affected members had a significantly higher average risk score than the reference population ($p = 0.001$). However, the risk scores were not in the top 90th percentile, which has previously been used to distinguish polygenic from monogenic forms of FH [26]. Interestingly, however, we found a combination of nine variants at the LDL-C GWAS loci among the affected members of this particular family ($p < 0.01$ by permutation). The risk scores constructed using only these nine variants changed the average risk score of the affected individuals to >90th percentile of LDL-C. Based on our data, we propose that constructing family-specific risk scores may be helpful in some large families to explain high LDL-C levels.

Among the nine family-specific GWAS variants (Table 2), there is one *HMGCR* variant, rs12916. HMGCR is a rate-limiting enzyme in cholesterol biosynthesis and the main target of statin therapy. Given the relatively large effect size of this GWAS variant (beta = 0.07) for LDL-C [12] and the previous evidence that the rs12916 is a liver eQTL [32], it is likely that the elevation of serum LDL-C levels due to the C allele is caused by augmented *HMGCR* expression and the subsequent increased cholesterol synthesis in the liver. The increased cholesterol synthesis in turn activates a feedback mechanism that inhibits the uptake of LDL-C from blood via the LDL receptor. Interestingly, a recent longitudinal metabolomics study observed that the carriers of the protective T allele exhibit a similar lipidomics profile as observed in individuals who have started statin therapy [33].

Our study has several limitations. Analysis of only one family does not provide information that could be directly extrapolated to the entire Austrian population. However, our findings further demonstrate the genetic complexity of FH in individuals without the known FH mutations. This type of presentation can clearly complicate the diagnosis and identification of hypercholesterolemic individuals in early stages of disease, emphasizing the importance of family-based evaluation of FH. We showed a likely polygenic effect that included variants residing in regions with LOD scores >1.0 and a combination of nine LDL-C GWAS SNPs

aggregating in the affected family members. We hypothesize that most of the FH families without a single known pathogenic mutation will exhibit a specific combination of the LDL-C GWAS variants that can be distinguished if extensive family data are available. We recognize that it is possible that we missed the causal variant(s) since no whole genome sequencing was performed and the causal variant might reside outside the protein coding regions or be a large copy number variation, not studied here. This scenario is, however, less likely given our negative linkage screening that, based on our simulation, had adequate power to identify a single monogenic variant. Our design does not fully exclude potential low-frequency modifier variants residing outside of the coding regions, not captured either by the SNP array or exome sequencing utilized here.

The Finnish population may not be an optimal reference population for the Austrian family because the minor allele frequencies between the two European populations might differ slightly. However, the LDL-C associated SNPs from Willer et al. [12] are mostly common variants (MAF>5%), typically largely shared by the European populations [34]. Furthermore, we also calculated separate risk scores using the Global Lipid Genetic Consortium 12-SNP LDL-C risk score, and compared the risk scores of the Austrian FH family with the published results of the Whitehall II controls [26]. We obtained similar results as with METSIM, suggesting that the METSIM cohort provides a sufficient reference population.

Our study focused mainly on the genetic architecture of LDL-C, one of the major risk factors for CHD. Hence, using genetic risk scores specific for CHD such as the ones recently established by Natajran et al. [35] might help understand the overall genetic risk for CHD in this family, and further identify individuals with a high risk for CHD, who potentially benefit most from the statin therapy [35].

In summary, our linkage study followed by exome sequencing and a GWAS LDL-C risk score analysis supports a polygenic cause for hypercholesterolemia in this Austrian family. Potential cascade testing of identified variants in the third generation of this family might provide valuable information regarding who should be followed up for early treatment of hypercholesterolemia. Our study demonstrates the importance of using family-wide genetic data, when available, in future personalized medicine initiatives of complex diseases. For example, in other FH families without the known FH mutations, a similar approach could be used to establish a family-specific polygenic hypercholesterolemia diagnosis, when sufficient numbers of affected and unaffected family members are available for identification of a family-specific set of LDL-C increasing GWAS SNPs that exceed the 90th risk score percentile in the particular population. Subsequently, the family's younger generations could be tested for these variants to provide an earlier personalized diagnosis.

Conflict of interest

The authors declared they do not have anything to disclose regarding conflict of interest with respect to this manuscript.

Financial support

Author contributions

Study design: EN, PP, M-RT, and FK.
Methods development and/or statistical and computational analysis: EN, PP, RMC, AK, MA, CL, KG, EK, SG and AR.

Clinical data collection, GWAS genotyping, and/or exome sequencing: M-RT, ML, RM, AP, FK, CL, WJS, MMM, NM, SS, and JB.
Manuscript: EN, PP and M-RT wrote the manuscript and all authors read, reviewed and/or edited the manuscript.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atherosclerosis.2017.07.024.

References

[1] S. Mendis, P. Puska, B. Norrving, Global atlas on cardiovascular disease prevention and control, World Heal Organ (2011) 2—14.
[2] M. Fischer, U. Broeckel, S. Holmer, A. Baessler, C. Hengstenberg, et al., Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction, Circulation 111 (7) (2005) 855—862, http://dx.doi.org/10.1161/01.CIR.0000155611.41961.BB.
[3] S. Zdravkovic, A. Wienke, N.L. Pedersen, M.E. Marenberg, A.I. Yashin, U. De Faire, Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins, J. Intern Med. 252 (3) (2002) 247—254.
[4] P. Deloukas, S. Kanoni, C. Willenborg, M. Farrall, T.L. Assimes, et al., Large-scale association analysis identifies new risk loci for coronary artery disease, Nat. Genet. 45 (1) (2013) 25—33, http://dx.doi.org/10.1038/ng.2480.
[5] R.D. Santos, S.S. Gidding, R.A. Hegele, M.A. Cuchel, P.J. Barter, et al., Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the international atherosclerosis society severe familial hypercholesterolemia panel, Lancet Diabetes Endocrinol. Elsevier Ltd. 8587 (16) (2016) 19—21, http://dx.doi.org/10.1016/S2213-8587(16)30041-9.
[6] M. Cuchel, E. Bruckert, H.N. Ginsberg, F.J. Raal, R.D. Santos, et al., Homozygous familial hypercholesterolemia: new insights and guidance for clinicians to improve detection and clinical management. A position paper from the consensus panel on familial hypercholesterolaemia of the European atherosclerosis society, Eur. Heart J. 35 (32) (2014) 2146—2157, http://dx.doi.org/10.1093/eurheartj/ehu274.
[7] M. Futema, V. Plagnol, K. Li, R a Whittall, H.A.W. Neil, et al., Whole exome sequencing of familial hypercholesterolaemia patients negative for LDLR/APOB/PCSK9 mutations, J. Med. Genet. 51 (8) (2014) 537—544, http://dx.doi.org/10.1136/jmedgenet-2014-102405.
[8] R.E.J. Clarke, S.T. Padayachee, R. Preston, Z. McMahon, M. Gordon, et al., Effectiveness of alternative strategies to define index case phenotypes to aid genetic diagnosis of familial hypercholesterolaemia, Heart 99 (3) (2013) 175—180, http://dx.doi.org/10.1136/heartjnl-2012-302917.
[9] F. Civeira, E. Jarauta, A. Cenarro, A.L. García-Otín, D. Tejedor, et al., Frequency of low-density lipoprotein receptor gene mutations in patients with a clinical diagnosis of familial combined hyperlipidemia in a clinical setting, J. Am. Coll. Cardiol. 52 (19) (2008) 1546—1553, http://dx.doi.org/10.1136/heartjnl-2012-302917.
[10] M. Bahlo, R. Tankard, V. Lukic, K.L. Oliver, K.R. Smith, Using familial information for variant filtering in high-throughput sequencing studies, Hum. Genet. 133 (2014) 1331—1341, http://dx.doi.org/10.1007/s00439-014-1479-4.
[11] S. Eggers, K.R. Smith, M. Bahlo, L.H. Looijenga, S.L. Drop, et al., Whole exome sequencing combined with linkage analysis identifies a novel 3 bp deletion in NR5A1, Eur. J. Hum. Genet. 23 (4) (2015) 486—493, http://dx.doi.org/10.1038/ejhg.2014.130.
[12] C.J. Willer, E.M. Schmidt, S. Sengupta, G.M. Peloso, S. Gustafsson, et al., Discovery and refinement of loci associated with lipid levels, Nat. Genet. 45 (11) (2013) 1274—1283, http://dx.doi.org/10.1038/ng.2797.
[13] R.R. Williams, S.C. Hunt, M.C. Schumacher, R.A. Hegele, M.F. Leppert, et al., Diagnosing heterozygous familial hypercholesterolemia using new practical criteria validated by molecular genetics, Am. J. Cardiol. 72 (2) (1993) 171—176.
[14] A.L. Catapano, I. Graham, G. De Backer, O. Wiklund, M.J. Chapman, et al., 2016 ESC/EAS guidelines for the management of dyslipidaemias, Eur. Heart J. 37 (2016), http://dx.doi.org/10.5603/KP.2016.0157, 2999—3058l.
[15] B.R.J. Havel, H.A. Eder, J.H. Bragdon, The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum, J. Clin. Invest. 34 (9) (1955) 1345—1353, http://dx.doi.org/10.1172/JCI103182.
[16] F. Kronenberg, E. Kuen, E. Ritz, R. Junker, P. König, et al., Lipoprotein(a) serum concentrations and apolipoprotein(a) phenotypes in mild and moderate renal failure, J. Am. Soc. Nephrol. 11 (1) (2000) 105—115.
[17] A. Laschkolnig, B. Kollerits, C. Lamina, C. Meisinger, B. Rantner, et al., Lipoprotein (a) concentrations, apolipoprotein (a) phenotypes, and peripheral

arterial disease in three independent cohorts, Cardiovasc Res. 103 (1) (2014) 28—36, http://dx.doi.org/10.1093/cvr/cvu107.

[18] A. Stancakova, M. Javorsky, T. Kuulasmaa, S.M. Haffner, J. Kuusisto, M. Laakso, Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men, Diabetes 58 (5) (2009) 1212—1221, http://dx.doi.org/10.2337/db08-1607.

[19] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (14) (2009) 1754—1760, http://dx.doi.org/10.1093/bioinformatics/btp324.

[20] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, et al., The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (2010) 1297—1303, http://dx.doi.org/10.1101/gr.107524.110.

[21] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (3) (2007) 559—575, http://dx.doi.org/10.1086/519795.

[22] K. Lange, J.C. Papp, J.S. Sinsheimer, R. Sripracha, H. Zhou, E.M. Sobel, Mendel: the Swiss army knife of genetic analysis programs, Bioinformatics 29 (12) (2013) 1568—1570, http://dx.doi.org/10.1093/bioinformatics/btt187.

[23] A.A. Schäffer, M. Lemire, J. Ott, G.M. Lathrop, D.E. Weeks, Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees, Hum. Hered. 71 (2) (2011) 126—134, http://dx.doi.org/10.1159/000324177.

[24] B.G. Nordestgaard, M.J. Chapman, K. Ray, J. Borén, F. Andreotti, et al., Lipoprotein(a) as a cardiovascular risk factor: current status, Eur. Heart J. 31 (23) (2010) 2844—2853, http://dx.doi.org/10.1093/eurheartj/ehq386.

[25] F. Kronenberg, G. Utermann, Lipoprotein(a): resurrected by genetics, J. Intern Med. 273 (1) (2013) 6—30, http://dx.doi.org/10.1111/j.1365-2796.2012.02592.x.

[26] P.J. Talmud, S. Shah, R. Whittall, M. Futema, P. Howard, et al., Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case-control study, Lancet 381 (9874) (2013) 1293—1301, http://dx.doi.org/10.1016/S0140-6736(12)62127-8.

[27] I. Brænne, M. Kleinecke, B. Reiz, E. Graf, T. Strom, et al., Systematic analysis of variants related to familial hypercholesterolemia in families with premature myocardial infarction, Eur. J. Hum. Genet. (April) (2015) 1—7, http://dx.doi.org/10.1038/ejhg.2015.100.

[28] M.T. Tejedor, A. Cenarro, D. Tejedor, M. Stef, R. Mateo-Gallego, et al., Haplotype analyses, mechanism and evolution of common double mutants in the human LDL receptor gene, Mol. Genet. Genomics 283 (6) (2010) 565—574, http://dx.doi.org/10.1007/s00438-010-0541-8.

[29] K.K. Alharbi, M.A. Aldahmesh, E. Spanakis, L. Haddad, R.A. Whittall, et al., Mutation scanning by meltMADGE: validations using BRCA1 and LDLR, and demonstration of the potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population, Genome Res. 15 (7) (2005) 967—977, http://dx.doi.org/10.1101/gr.3313405.

[30] J. Borén, I. Lee, W. Zhu, K. Arnold, S. Taylor, T.L. Innerarity, Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective Apo-B100, J. Clin. Invest. 101 (5) (1998) 1084—1093, http://dx.doi.org/10.1172/JCI1847.

[31] J. Wang, J.S. Dron, M.R. Ban, J.F. Robinson, A.D. McIntyre, et al., Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically, Arter. Thromb. Vasc. Biol. (12) (2016) 36, http://dx.doi.org/10.1161/ATVBAHA.116.308027.

[32] D.I. Swerdlow, D. Preiss, K.B. Kuchenbaecker, M.V. Holmes, J.E.L. Engmann, et al., HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials, Lancet 385 (9965) (2015) 351—361, http://dx.doi.org/10.1016/S0140-6736(14)61183-1.

[33] P. Würtz, Q. Wang, P. Soininen, A.J. Kangas, G. Fatemifar, et al., Metabolomic profiling of statin use and genetic inhibition of HMG-CoA reductase, J. Am. Coll. Cardiol. 67 (10) (2016) 1200—1210, http://dx.doi.org/10.1016/j.jacc.2015.12.060.

[34] H. Chheda, P. Palta, M. Pirinen, S. McCarthy, K. Walter, et al., Whole genome view of the consequences of a population bottleneck using 2926 genome sequences from Finland and United Kingdom, Eur. J. Hum. Genet. 25 (4) (2017) 477—484, http://dx.doi.org/10.1038/ejhg.2016.205.

[35] P. Natarajan, R. Young, N.O. Stitziel, S. Padmanabhan, U. Baber, et al., Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting, Circulation 135 (18) (2017) 617—643, http://dx.doi.org/10.1161/CIRCULATIONAHA.116.024436.

[36] V.-P. Mäkinen, M. Parkkonen, M. Wessman, P.-H. Groop, T. Kanninen, K. Kaski, High-throughput pedigree drawing, Eur. J. Hum. Genet. 13 (8) (2005) 987—989, http://dx.doi.org/10.1038/sj.ejhg.5201430.

[37] H. Zhang, P. Meltzer, S. Davis, RCircos: an R package for Circos 2D track plots, BMC Bioinforma. 14 (1) (2013) 244, http://dx.doi.org/10.1186/1471-2105-14-244.