Polygenic Hypercholesterolemia: Examples of GWAS Results and Their Replication in the Czech-Slavonic Population

J. A. HUBACEK[1], V. ADAMKOVA[2], V. LANSKA[3], D. DLOUHA[1]

[1]Centre for Experimental Medicine, Institute for Clinical and Experimental Medicine, Prague, Czech Republic, [2]Department of Preventive Cardiology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic, [3]Statistical Unit, Institute for Clinical and Experimental Medicine, Prague, Czech Republic

Summary

Since 2007, the year of their first widespread use, genome-wide association studies (GWAS) have become the "gold standard" for the detection of causal genes and polymorphisms in all fields of human medicine. Cardiovascular disease (CVD), one of the major causes of morbidity and mortality, is no exception. The first GWAS focused on hypercholesterolemia and dyslipidemia as the major CVD determinants. GWAS confirm the importance of most of the previously identified genes (e.g. *APOE*, *APOB*, *LDL-R*) and recognize the importance of new genetic determinants (e.g. within the *CILP2* or *SORT1* gene clusters). Nevertheless, the results of GWAS still require confirmation by independent studies, as interethnic and interpopulation variability of SNP effects have been reported. We analyzed an association between eight variants within seven through GWAs detected loci and plasma lipid values in the Czech post-MONICA population sample (N=2,559). We confirmed an association (all P<0.01) between plasma LDL-cholesterol values and variants within the *CILP2* (rs16996148), *SORT1* (rs646776), *APOB* (rs693), *APOE* (rs4420638) and *LDL-R* (rs6511720) genes in both males (N=1,194) and females (N=1,368). In contrast, variants within the *APOB* (rs515135), *PCSK9* (rs11206510) and *HMGCoAR* (rs12654264) genes did not significantly affect plasma lipid values in Czech males or females. Unweighted gene score values were linearly associated with LDL-cholesterol values both in males (P<0.0005) and females (P<0.00005). We confirmed the effects of some, but not all analyzed SNPs on LDL-cholesterol levels, reinforcing the necessity for replication studies of GWA-detected gene variants.

Corresponding author

J. A. Hubacek, IKEM-CEM-LMG, Videnska 1958/9, 140 21 Prague 4, Czech Republic. E-mail: jahb@ikem.cz

Introduction

Cardiovascular disease (CVD), a leading cause of morbidity and mortality in developed countries, has a significant genetic background. Although monogenic Mendelian diseases leading to CVD, such as familial hypercholesterolemia (Wong *et al.* 2016) or Tangier disease (Bodzioch *et al.* 1999), are generally well recognized and their genetic backgrounds have been explained, such diseases only account for a very low proportion of CVD and do not explain the high prevalence of CVD within the general population.

The vast majority of CVD is of a polygenic character, it means that the disease has a late onset, a mild genotype, and inheritance based on the presence of risky alleles of many common polymorphisms with relatively minor effects.

The first GWAS to focus on CVD risk factors were performed roughly ten years ago and have significantly moved forward our understanding of the genetic architecture of CVD. GWA technology (Dubé and Hegele 2013, Uitterlinden 2016) facilitates the analysis of an enormous number of single nucleotide polymorphisms

(SNPs) in one individual over a very short period of time.

SNPs are the most common genetic variants. It is estimated that there are about 10 million common SNPs (with a frequency of minor alleles at over 1 % in the population) within the human genome. They are supposed to have a major genetic influence on disease heritability and are known to be the best markers of different chromosomal loci due to their very high numbers across the genome.

In GWAS, SNPs are analyzed using microarray "SNP-chips". One chip is used to analyze DNA from one subject and the results can be used to calculate risk for arbitrary known and available anthropometric, biochemical or clinical characteristics. Indeed, they can be used to analyze any type of disease in which at least some part of the genetic background is expected to have an influence, simply by comparing genotype frequencies in patients and healthy subjects.

For the first chips, SNPs have been selected primarily according to their prevalence (at least 5 % for minor alleles) and in order to mark as many chromosomal loci as possible. The potential functional importance has not been an issue in this type of SNP selection. As individual neighbouring SNPs are often linked together (linkage disequilibrium), only one of these clustered (and simultaneously inherited) SNPs needs to be included on the chip in order to characterize the entire region.

As hundreds of thousands of SNPs are analyzed simultaneously, the statistical threshold used needs to be very stringent. Based on simple Bonferroni correction, the GWAS significance threshold is set at $P<10^{-8}$. P values are calculated for each SNP individually and the results are then summarized into one single figure, which is called a Manhattan plot.

Despite the overwhelming efforts made, GWAS also present some pitfalls (Table 1). Frequently, the most important functional SNPs are not included on the chip. In fact, detailed analysis of the identified loci often leads to the detection of variants with much higher effects than originally described (Sanna *et al.* 2011). Another weak point pertains to SNP frequency, since some rare variants with strong effects are also omitted simply because they have not been included on the chip. The last issue not to be overlooked is the necessity for high quality controls. Not all SNPs included on chips pass through quality controls, which can also potentially lead to some important variants being neglected.

As mentioned above, dyslipidemia (elevated plasma lipids) is associated with increased risk of atherosclerosis development and cardiovascular disease. However, its role in total mortality determination is controversial (Anderson *et al.* 1987, Arsenault *et al.* 2011, Hubáček 2015, Liu *et al.* 2013, Pikhart *et al.* 2015, Piťha *et al.* 2016, van Wijk *et al.* 2009).

Table 1. Possible advantages and limitations of genome-wide association studies. These points apply for medical research generally.

Advantages
– large numbers of analyzed polymorphisms
– opportunity to analyze all available traits
– detecting risky factors which are stable throughout the entire lifespan
– detecting important non-coding variants with unknown mechanisms of influence

Limitations
– rare variants are omitted from the analyses
– heterogeneity of the analyzed populations is often low
– detected significant variants are not every times causal
– statistical significance does not necessarily equate to biological significance
– even results from large GWAS are not undoubtedly reproduced
– selection of SNPs was based on frequency, not on functionality

Lipids (mainly cholesterol and triglycerides) are transported through the body in heterogeneous complexes along with proteins. As lipoproteins play an important role in coronary artery disease development, their values have been the subject of intensive investigations over the last five decades. Plasma lipid concentrations are highly heritable and it is estimated that even 60-70 % of total plasma lipid values are attributable to these genetic factors. Before GWAS, our knowledge of those variants with strong effects on plasma lipids was limited and the reproducibility of results was low. But even during the pre-GWA era, recognition was given to the importance of variants within apolipoprotein E (as predictors of plasma total- and LDL-cholesterol values) (Eichner *et al.* 2002) and apolipoprotein A5 (as determinants of plasma triglycerides) (Pennacchio *et al.* 2001, Hubacek 2016).

The first four GWAS on this topic, all involving thousands of individuals, were published almost simultaneously and delivered surprisingly uniform results (Kathiresan *et al.* 2008, Sandhu *et al.* 2008, Willer *et al.*

2008, Chasman *et al.* 2008). They identified several new genes or loci that contribute to plasma lipid concentrations and confirmed a number of previously reported associations (Table 2). Another "batch" of studies followed very soon after, including more samples and descriptions of more variants, and typically documenting the more minute effects of plasma lipids (Kathiresan *et al.* 2009, Aulchenko *et al.* 2009, Teslovich *et al.* 2010, Global Lipids Genetics Consortium 2013). These results were mostly summarized by Christoffersen and Tybjærg-Hansen (2015) and by Schwarzova *et al.* (2015).

Table 2. Selected examples of genes/variants with strong effects on plasma lipid parameters based on results from four GWAS (Kathiresan *et al.* 2008, Sandhu *et al.* 2008, Willer *et al.* 2008, Chasman *et al.* 2008). Variants used for the confirmatory study are in bold.

Gene for	Variant			Influenced parameter		
Loci with prior evidence from association studies						
LDL-receptor	**rs6511720**			TC	LDL-C	
Apolipoprotein B	**rs515135**	**rs693**		TC	LDL-C	
PCSK9	**rs11206510**	rs11591147		TC	LDL-C	
HMGCoA reductase	**rs12654264**	rs3846662	rs2654264	TC	LDL-C	
APOE/CI/CIII cluster	**rs4420638**			TC	LDL-C	TG
ABCA1	rs3890182			TC	LDL-C	HDL-C
Lipoprotein lipase	rs506696			TG		
CETP	rs3764261			HDL-C		
Apolipoprotein A5	rs964184			TG		
New, GWAS-identified loci						
CILP2/PBX4 cluster	**rs16996148**			TC	LDL-C	TG
SORT-1	**rs646776**			TC	LDL-C	
MLXIPL	rs3812316			TG		
GCKR	rs1260326			TG		
ANGPTL3	rs12130333			TG		
TRIB1	rs17321515			TG		
GALNT2	rs48469114			TG	HDL-C	
TOMM40	rs2075650			TC	LDL-C	

As expected and based on the knowledge of lipid transport and metabolism (van Wijk *et al.* 2009), variants with strong effects have been detected within the enzyme HMGCoA reductase (a key enzyme of cholesterol biosynthesis), apolipoprotein B (a major structural protein of LDL particles), apolipoprotein E (maintains the structural integrity of lipoproteins and acts as a ligand for lipoprotein receptors), apolipoprotein A5 (an activator of lipoprotein lipase, which additionally reduces hepatic production of VLDL-particles) and the LDL receptor (a key receptor that is important in LDL particle assembly) (Table 2). It is important to note that new variants within these genes have been commonly found to be of higher importance in comparison with those variants analyzed during the pre-GWAS era.

More importantly, however, unexpectedly high numbers of significant variants have been detected within new, as yet unknown genes/loci (Table 2). One of the strongest GWA signals associated with plasma lipids is localized within the cluster of *CELSR2/PSRC1/SORT1* genes. Other studies have identified the gene for sortilin as being responsible for the regulation of plasma cholesterol levels. Sortilin has been recognized as an intracellular apoB receptor, which facilitates the intrahepatic formation and transport of apoB-containing lipoproteins (Kjolby *et al.* 2010).

Despite the undoubted advancements made by GWAS, these results still need to be replicated in independent populations, especially if they are to be used in clinical praxis.

The substantial lack of replication studies (for GWAS-obtained results) having been performed or published has possibly led to some bias, which means that the real effects on different populations may be over- or under-estimated (Munafò 2009). The major issue is that the original studies mostly include West-European populations

only. Detailed analysis of interethnic variability, or even the differences between nationalities or possible gene-gene and gene-environment interactions, is entirely absent from these studies. Although the different effects of SNPs on males and females have been detected (Hubacek *et al.* 2008a, Hubacek *et al.* 2009, Heo *et al.* 2014, Kulminski *et al.* 2015) and notwithstanding that GWAS usually comprise both males and females, the potential sex-specific effects have often not been analyzed/presented.

Based on the results of the first GWAS, we selected variants within seven genes (for more details, see Table 2) in order to replicate these results in the Czech-Slavonic population and to confirm the potential universality of the effects. In addition to the individual gene effects, we constructed a risk score as a potentially more powerful predictor of plasma lipid values in the general population.

Methods

Analyzed subjects

We examined one group of adults (1,191 males and 1,368 females, aged 25-64 years; mean age of 49.0±10.7 years) (Hubacek *et al.* 2009, Hubacek *et al.* 2015a, Hubacek *et al.* 2016). The individuals had participated in the post-MONICA study (Cifkova *et al.* 2010) and were selected according to the WHO MONICA project protocol (Tunstall-Pedoe *et al.* 2003), which was designed to examine the risk factors of cardiovascular disease development, including plasma lipids. The subjects were examined in nine Czech districts in 2000/2001. Written informed consent was given by all individuals and the study was approved by the Ethics Committee of the Institute for Clinical and Experimental Medicine, Prague in agreement with the Helsinki Declaration of 1975.

Genotype analysis

DNA was isolated using the standard salting-out method from whole EDTA blood (Miller *et al.* 1988). Individual variants within the seven genes (rs646776 – *CELSR2/PSRC1/SORT1* gene cluster, rs16996148 – *CILP2/PBX4* gene cluster, rs693 and rs515135 – *APOB,* rs4420638 – *APOE/C1/C4* gene cluster, rs12654264 – *HMGCoA reductase*, rs11206510 – *PCSK9* and rs6511720 – *LDL receptor*) were genotyped using the polymerase chain reaction (PCR) and restriction analysis (Hubacek *et al.* 2015b). The PCR device DYAD (MJ Research, Waltham, MA) and chemicals from Fermentas, Burlington, Canada were used. For more detailed PCR conditions, oligonucleotides and restriction

enzymes, see Vrablik *et al.* (2012).

Analysis of plasma lipids

Lipoprotein parameters (total cholesterol, HDL cholesterol and triglycerides) in fasting plasma samples were assessed using auto-analyzers and conventional enzymatic methods with reagents from Boehringer Mannheim Diagnostics and Hoffmann-La Roche in a CDC Atlanta-accredited local laboratory. LDL-cholesterol levels were calculated using the Friedewald formula: LDL-C = TC − HDL-C − (TG/2.2) (Friedewald *et al.* 1972) in subjects with TG values below 4.5 mmol/l. Dyslipidemic treatment was defined as the self-reported use of lipid-lowering drugs and/or diet.

Statistical analysis

The Hardy-Weinberg test (http://www.tufts.edu/ ~mcourt01/Documents/Court%20lab%20-%20HW%20 calculator.xls) was applied to confirm the independent segregation of alleles. ANOVA was used for statistical analysis. In cases where there was less than 5 % carriers of one genotype, these individuals were pooled with heterozygotes and analyzed together.

For polygenic analysis, an unweighted gene score was constructed according to the number of genotypes associated with higher plasma levels of LDL-cholesterol. Based on the number of negative genotypes from five variants within *SORT1*, *CILP2*, *APOE*, *APOB* and *LDL-R*, subjects were divided into six groups (from 0 up to 5 negative genotypes). We pooled groups with 0 and 1 and with 4 and 5 negative genotypes, respectively, as the numbers of individuals within the groups "0" and "5" were relatively low. The scores were used to analyze LDL-cholesterol values.

Due to the high number of performed comparisons, a significance level of P<0.01 was considered significant.

Results

Basic characteristics

The basic characteristics of the examined individuals are summarized in Table 3. The achieved call rate for the examined polymorphisms varied between 91.8 % and 99.9 %. The frequencies of individual genotypes were similar to the frequencies detected within West-European Caucasians (Kathiresan *et al.* 2008, Sandhu *et al.* 2008, Willer *et al.* 2008, Chasman *et al.* 2008) and did not differ significantly between males and females (for more details, see Table 4).

Table 3. Basic characteristics of analyzed subjects.

	Males	**Females**
Number	1,191	1,368
Age (years)	49.0 ± 10.8	48.8 ± 10.6
Total cholesterol (mmol/l)	5.75 ± 1.06	5.80 ± 1.15
LDL-cholesterol (mmol/l)	3.56 ± 0.99	3.45 ± 1.02
HDL-cholesterol (mmol/l)	1.26 ± 0.33	1.50 ± 0.36
Triglycerides (mmol/l)	1.97 ± 1.28	1.46 ± 0.85
BMI (kg/m^2)	28.2 ± 4.0	27.6 ± 5.5
Never-smokers (%)	41.3	63.8
T2DM (%)	8.9	6.8
Hypertension (%)	40.7	33.1
Dyslipidemic treatment (%)	7.3	5.6

Single nucleotide polymorphisms

For most of the analyzed SNPs, we detected a significant influence on plasma lipid values in both males (Table 5) and females (Table 6).

We observed the strongest effects on plasma lipid values (P<0.005 for LDL-C values) in the cases of the rs646776 variant within the *SORT1* gene, the rs6511720 variant in the *LDL-R* gene and the rs693 variant within the *APOB* gene. And although adjustments for dyslipidemic treatment and age reduced this significance, all associations remained significant.

In contrast, we did not confirm an association between variants within the *PCSK9* and *HMGCoAr* genes and plasma lipid values (not shown in detail).

Table 4. Genotype distributions among males and females analyzed.

CILP2/PBX4	**GG**		**GT**		**TT**	
rs16996148	N	%	N	%	N	%
Males	979	82.5	193	16.3	15	1.3
Females	1175	86.1	180	13.2	10	0.7
Apolipoprotein B	**CC**		**CT**		**TT**	
rs693	N	%	N	%	N	%
Males	327	28.0	554	47.4	287	24.6
Females	385	29.3	655	49.9	272	20.7
Apolipoprotein B	**GG**		**GA**		**AA**	
rs515135	N	%	N	%	N	%
Males	699	66.0	317	29.9	43	4.1
Females	880	68.2	368	28.5	43	3.3
SORT1	**AA**		**AG**		**GG**	
rs646776	N	%	N	%	N	%
Males	688	60.0	394	34.4	65	5.6
Females	813	61.5	458	34.6	52	3.9
PCSK9	**TT**		**TC**		**CC**	
rs11206510	N	%	N	%	N	%
Males	794	66.7	357	30.0	39	3.3
Females	915	67.6	408	30.1	31	2.3
HMG-CoA r	**AA**		**AT**		**TT**	
rs12654264	N	%	N	%	N	%
Males	421	35.6	577	48.8	184	15.6
Females	484	35.6	671	49.3	206	15.1
LDL-receptor	**GG**		**GT**		**TT**	
rs6511720	N	%	N	%	N	%
Males	982	82.7	194	16.3	11	0.9
Females	1108	80.9	244	17.8	16	1.2
APOE	**AA**		**AG**		**GG**	
rs4420638	N	%	N	%	N	%
Males	843	72.2	308	26.4	17	1.5
Females	975	74.1	304	23.1	37	2.8

Table 5. Analyzed polymorphisms with significant effects on plasma lipid values in males.

CILP2/PBX4 rs16996148	**GG**		**GT + TT**	**P**
Total cholesterol	5.78 ± 1.16		5.54 ±1.09	0.01
LDL-cholesterol	3.59 ± 0.99		3.40 ± 0.88	0.01
HDL-cholesterol	1.26 ± 0.13		1.29 ± 0.35	n.s.
Triglycerides	2.13 ± 1.16		2.00 ± 1.81	0.01
APOB rs693	**CC**	**CT**	**TT**	**P**
Total cholesterol	5.71 ± 1.05	5.72 ± 1.21	5.82 ± 1.19	n.s.
LDL-cholesterol	3.51 ± 0.84	3.53 ± 1.00	3.68 ± 1.03	0.005
HDL-cholesterol	1.29 ± 0.34	1.27 ± 0.32	1.24 ± 0.30	n.s.
Triglycerides	2.13 ± 1.80	2.11 ± 1.74	2.11 ± 1.45	n.s.
SORT1 rs646776	**AA**	**AG**	**GG**	**P**
Total cholesterol	5.84 ± 1.16	5.65 ± 1.18	5.54 ± 1.00	0.005
LDL-cholesterol	3.63 ± 0.97	3.45 ± 0.99	3.42 ± 0.91	0.005
HDL-cholesterol	1.23 ± 0.31	1.31 ± 0.34	1.35 ± 0.35	0.005
Triglycerides	2.13 ± 1.43	2.13 ± 2.04	1.85 ± 1.35	n.s.
LDL-R rs6511720	**GG**		**GT + TT**	**P**
Total cholesterol	5.74 ± 1.14		5.73 ± 1.22	n.s.
LDL-cholesterol	3.58 ± 0.96		3.43 ± 0.98	0.005
HDL-cholesterol	1.27 ± 0.32		1.25 ± 0.34	n.s.
Triglycerides	2.05 ± 1.16		2.42 ± 2.00	n.s.
APOE rs4420638	**AT + TT**		**AA**	**P**
Total cholesterol	5.85 ± 1.12		5.69 ± 1.17	0.01
LDL-cholesterol	3.67 ± 0.96		3.50 ± 0.97	0.01
HDL-cholesterol	1.23 ± 0.32		1.28 ± 0.32	n.s.
Triglycerides	2.22 ± 2.03		2.06 ± 1.52	n.s.

Gene score

The gene score was calculated from five SNPs within the genes for CILP2 (rs16996148), sortilin (rs646776), the LDL-receptor (rs6511720), APOE (rs4420638) and APOB (rs693). Both in males (N=993 with all five SNPs genotyped and valid LDL-C value) and females (N=1,155 with all five SNPs genotyped and valid LDL-C value), we were able to detect subjects across the entire span of the gene score.

Among the group with the lowest score, subjects with a maximum of one risky genotype were included. They exhibited LDL-C values of 3.35±1.01 mmol/l (males) or 3.17±1.12 mmol/l (females) in contrast with the group of subjects with at least four negative genotypes, where LDL-C values of 3.77±1.01 mmol/l (males) and 3.75±1.08 mmol/l (females) were detected. There was a significant linear trend (Fig. 1) between these groups and the differences were highly significant for both males (P<0.0005) and females (P<0.00005). After adjustments for dyslipidemic treatment and age, P values remained unchanged.

Discussion

In a large group of Slavonic Caucasian adults, we geographically extended the validity of previous findings (Kathiresan *et al.* 2008, Sandhu *et al.* 2008, Willer *et al.* 2008, Chasman *et al.* 2008) focused on the genetic determination of plasma lipid levels. We confirmed strong significant effects of the common SNPs within *SORT1, CILP2/PBX4, APOE,* the *LDL receptor* and *APOB* on total plasma, LDL-cholesterol and plasma TG levels (*CILP2/PBX4* loci only). These results were confirmed for both males and females, and age adjustment did not change the results significantly. We did not have information on female menopausal status, but it has been reported that normal physiological menopausal alterations do not significantly influence plasma lipid homeostasis (Wang *et al.* 2011). Further, performed age adjustment could partially replace the adjustment for menopausal status and in our study, no major differences have been observed after the age adjustment.
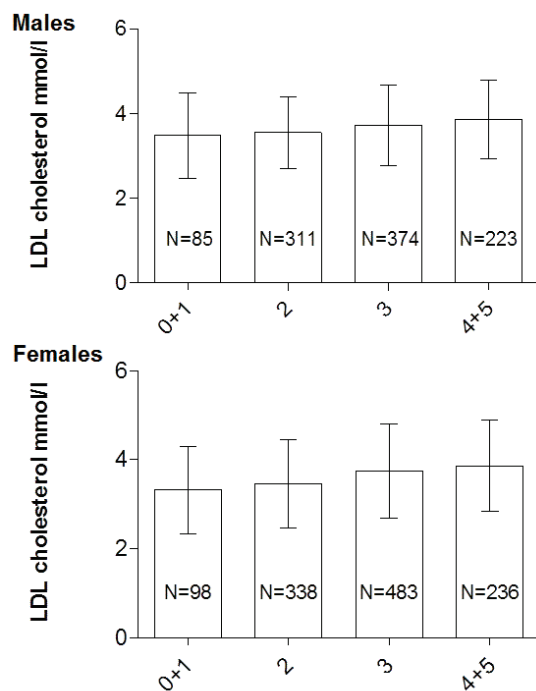
Fig. 1. Plasma LDL-cholesterol values (mmol/l) in groups with different numbers of risky genotypes of polymorphisms within the genes for APOB, APOE, LDL-receptor, sortilin-1 and CILP2/PBX4 genes.

In contrast, we did not observe an association between the *PCSK9* and *HMGCoA reductase* variants and plasma lipids in Czechs. In the case of the rs11206510 variant within the *PCSK9* gene, the differences for LDL cholesterol were nominally high (about a 0.25 mmo/l difference for LDL-cholesterol), but the relatively low number of individuals with the protective CC genotype (~3 %) may explain why the differences did not reach significance.

For the polymorphism within the *HMGCoA reductase* gene, this approach could not explain the negative results (all LDL-cholesterol differences between homozygotes were below 0.09 mmol/l) and no association was detected even when all three possible models of the analysis (dominant, co-dominant and recessive) were used. The relatively low number of subjects included (in comparison with GWAS) may have led to this unexpected difference, but we cannot exclude also the possibility, that this SNP has no biologically significant effect on plasma lipids in Czechs.

Table 6. Analyzed polymorphisms with significant effects on plasma lipid values in females.

CILP2/PBX4 rs16996148	**GG**		**GT + TT**	**P**
Total cholesterol	5.63 ± 1.22		5.50 ± 1.15	n.s.
LDL-cholesterol	3.47 ± 1.07		3.36 ± 1.06	0.01
HDL-cholesterol	1.50 ± 0.35		1.52 ± 0.39	n.s.
Triglycerides	1.48 ± 0.87		1.36 ± 0.81	0.01
APOB rs693	**CC**	**CT**	**TT**	**P**
Total cholesterol	5.57 ± 1.05	5.55 ± 1.21	5.77 ± 1.23	n.s.
LDL-cholesterol	3.40 ± 1.01	3.41 ± 1.08	3.62 ± 1.10	0.001
HDL-cholesterol	1.51 ± 0.36	1.50 ± 0.36	1.50 ± 0.35	n.s.
Triglycerides	1.45 ± 0.85	1.44 ± 0.86	1.49 ± 0.79	n.s.
SORT1 rs646776	**AA**	**AG**	**GG**	**P**
Total cholesterol	5.68 ± 1.19	5.51 ± 1.23	5.22 ± 1.03	0.001
LDL-cholesterol	3.51 ± 1.06	3.34 ± 1.08	3.08 ± 0.88	0.001
HDL-cholesterol	1.49 ± 0.35	1.52 ± 0.37	1.51 ± 0.30	n.s.
Triglycerides	1.46 ± 0.86	1.45 ± 0.85	1.38 ± 0.77	n.s.
LDL-R rs6511720	**GG**		**GT + TT**	**P**
Total cholesterol	5.66 ± 1.21		5.42 ± 1.21	n.s.
LDL-cholesterol	3.49 ± 1.00		3.27 ± 1.09	0.005
HDL-cholesterol	1.50 ± 0.30		1.50 ± 0.35	n.s.
Triglycerides	1.46 ± 0.85		1.43 ± 0.91	n.s.
APOE rs4420638	**AT + TT**		**AA**	**P**
Total cholesterol	5.85 ± 1.21		5.52 ± 1.19	0.01
LDL-cholesterol	3.68 ± 1.07		3.36 ± 1.07	0.01
HDL-cholesterol	1.48 ± 0.35		1.52 ± 0.36	n.s.
Triglycerides	1.50 ± 0.84		1.43 ± 0.86	n.s.

GWA screenings of plasma lipids have undoubtedly contributed to recent huge advancements in our knowledge in this field. However, despite the increasing popularity, the genome-wide association studies (Rosenberg *et al.* 2010) also have some pitfalls and limitations. Despite the extremely high P values obtained, these results still need to be replicated in order to negate the "winner's curse" effect. Further, it is well known that there are significant differences between populations and ethnic groups in the allelic frequencies of detected SNPs, in addition to environmental factors. These facts must surely impact on the global validity of the results obtained through such GWAS.

The results we have produced over recent years clearly show that GWAS results do not have absolute validity. We have focused on the replications of different GWAS findings in Czech-Slavonic population and, as expected, replicated some (Hubacek *et al.* 2008b, 2012, 2016) but not all of the associations (Vrablik *et al.* 2008, Hubacek *et al.* 2013, 2015c).

Our results point to a further important issue. In line with confirmation of the associations between single SNPs and some biochemical/anthropometrical markers, a much higher predictive value can be obtained when a polygenic score is created. Even though we used a simple and less precise unweighted score, the differences reached 12 % in males and 18 % in females between the extreme groups in the case of LDL-cholesterol.

Much recent discussion has focused on the issue of whether gene analysis can provide any extra benefit to the "traditionally" calculated risk score. Indeed, the conclusions drawn often dispute its value (Talmud *et al.*

2010, Weijmans *et al.* 2015, Morris *et al.* 2016). However, these studies neglect to construct an overall view of the potential for the genetic prediction of disease. Using a gene profile to improve the risk calculation of cardiovascular disease in a 65-year-old man who is a lifelong smoker, obese and insulin-resistant is unlikely to be useful.

But it is of more benefit to determine genetic information for the timely estimation of disease risk in (very) young subjects (maybe even in children) to apply personalized and potentially long-term focused environmental intervention in order to postpone or even prevent the outbreak of disease.

Conclusions

We confirmed strong significant effects of common SNPs within the *SORT1, CILP2/PBX4, APOE, LDL receptor* and *APOB* (but not *PCSK9* and *HMGCoA-R*) genes on plasma LDL-cholesterol levels and plasma TG levels (*CILP2/PBX4* loci only) in both males and females. The unweighted gene score was linearly associated with plasma LDL-cholesterol and the effect of this score was much higher in comparison with individually analyzed polymorphisms.

Conflict of Interest

There is no conflict of interest.

Acknowledgement

References

ANDERSON KM, CASTELLI WP, LEVY D: Cholesterol and mortality. 30 years of follow-up from the Framingham study. *JAMA* **257**: 2176-2180, 1987.

ARSENAULT BJ, BOEKHOLDT SM, KASTELEIN JJ: Lipid parameters for measuring risk of cardiovascular disease. *Nat Rev Cardiol* **8**: 197-206, 2011.

AULCHENKO YS, RIPATTI S, LINDQUIST I, BOOMSMA D, HEID IM, PRAMSTALLER PP, PENNINX BW, JANSSENS AC, WILSON JF, SPECTOR T, MARTIN NG, PEDERSEN NL, KYVIK KO, KAPRIO J, HOFMAN A, ET AL.: Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* **41**: 47-55, 2009.

BODZIOCH M, ORSÓ E, KLUCKEN J, LANGMANN T, BÖTTCHER A, DIEDERICH W, DROBNIK W, BARLAGE S, BÜCHLER C, PORSCH-OZCÜRÜMEZ M, KAMINSKI WE, HAHMANN HW, OETTE K, ROTHE G, ASLANIDIS C, LACKNER KJ, SCHMITZ G: The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat Genet* **22**: 347-351, 1999.

CHASMAN DI, PARÉ G, ZEE RY, PARKER AN, COOK NR, BURING JE, KWIATKOWSKI DJ, ROSE LM, SMITH JD, WILLIAMS PT, RIEDER MJ, ROTTER JI, NICKERSON DA, KRAUSS RM, MILETICH JP, RIDKER PM: Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet* **1**: 21-30, 2008.

CHRISTOFFERSEN M, TYBJÆRG-HANSEN A: Novel genes in LDL metabolism - a comprehensive overview. *Curr Opin Lipidol* **26**: 179-187, 2015.

CÍFKOVÁ R, SKODOVÁ Z, BRUTHANS J, ADÁMKOVÁ V, JOZÍFOVÁ M, GALOVCOVÁ M, WOHLFAHRT P, KRAJCOVIECHOVÁ A, POLEDNE R, STÁVEK P, LÁNSKÁ V: Longitudinal trends in major cardiovascular risk factors in the Czech population between 1985 and 2007/8. Czech MONICA and Czech post-MONICA. *Atherosclerosis* **211**: 676-681, 2010.

DUBÉ J, HEGELE RA: Genetics 100 for cardiologists: Basics of the genome-wide association studies. *Can J Cardiol* **29**: 10-17, 2013.

EICHNER JE, DUNN ST, PERVEEN G, THOMPSON DM, STEWART KE, STROEHLA BC: Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am J Epidemiol* **155**: 487-495, 2002.

FRIEDEWALD WT, LEVY RI, FREDRICKSON DS: Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* **18**: 499-502, 1972.

GLOBAL LIPIDS GENETICS CONSORTIUM, WILLER CJ, SCHMIDT EM, SENGUPTA S, PELOSO GM, GUSTAFSSON S, KANONI S, GANNA A, CHEN J, BUCHKOVICH ML, MORA S, BECKMANN JS, BRAGG-GRESHAM JL, CHANG HY, DEMIRKAN A, ET AL.: Discovery and refinement of loci associated with lipid level. *Nat Genet* **45**: 1274-1283, 2013.

HEO SG, HWANG JY, UHMN S, GO MJ, OH B, LEE JY, PARK JW: Male-specific genetic effect on hypertension and metabolic disorders. *Hum Genet* **133**: 311-319, 2014.

HUBÁČEK JA: High cholesterol is not associated with increased mortality (in Czech). *Hyper Kardiovask Prevence* **3**: 58-62, 2015.

HUBACEK JA: Apolipoprotein A5 fifteen-years anniversary: Lessons from genetic epidemiology. *Gene* **592**: 193-199, 2016.

HUBACEK JA, LÁNSKÁ V, SKODOVÁ Z, ADÁMKOVÁ V, POLEDNE R: Sex-specific interaction between APOE and APOA5 variants and determination of plasma lipid levels. *Eur J Hum Genet* **16**: 135-138, 2008a.

HUBACEK JA, BOHUSLAVOVA R, KUTHANOVA L, KUBINOVA R, PEASEY A, PIKHART H, MARMOT MG, BOBAK M: The FTO gene and obesity in a large Eastern European population sample: the HAPIEE study. *Obesity (Silver Spring)* **16**: 2764-2766, 2008b.

HUBACEK JA, PITHA J, ADAMKOVA V, LANSKA V, POLEDNE R: A common variant in the FTO gene is associated with body mass index in males and postmenopausal females but not in premenopausal females. Czech post-MONICA and 3PMFs studies. *Clin Chem Lab Med* **47**: 387-390, 2009.

HUBÁČEK JA, STANĚK V, GEBAUEROVÁ M, POLEDNE R, ASCHERMANN M, SKALICKÁ H, MATOUŠKOVÁ J, KRUGER A, PĚNIČKA M, HRABÁKOVÁ H, VESELKA J, HÁJEK P, LÁNSKÁ V, ADÁMKOVÁ V, PITHA J: Association between a marker on chromosome 9 and acute coronary syndrome. Confirmatory study on Czech population. *Folia Biol (Praha)* **58**: 203-208, 2012.

HUBÁČEK JA, NEŠKUDLA T, KLEMENTOVÁ M, ADÁMKOVÁ V, PELIKÁNOVÁ T: Tagging rs10811661 variant at CDKN2A/2B locus is not associated with type 2 diabetes mellitus in Czech population. *Folia Biol (Praha)* **59**: 168-171, 2013.

HUBACEK JA, DLOUHA D, LANSKA V, ADAMKOVA V: Strong gender-specific effects of the NYD-SP18 and FTO variants on BMI values. *Physiol Res* **64** (Suppl 3): S419-S426, 2015a.

HUBÁČEK JA, PIKHART H, PEASEY A, KUBÍNOVÁ R, BOBÁK M: Nobody is perfect: Comparison of the accuracy of PCR-RFLP and KASP™ method for genotyping. ADH1B and FTO polymorphisms as examples. *Folia Biol (Praha)* **61**: 156-160, 2015b.

HUBACEK JA, STANĚK V, GEBAUEROVÁ M, POLEDNE R, ASCHERMANN M, SKALICKÁ H, MATOUŠKOVÁ J, KRUGER A, PĚNIČKA M, HRABÁKOVÁ H, VESELKA J, HÁJEK P, LÁNSKÁ V, ADÁMKOVÁ V, PIŤHA J: Rs6922269 marker at the MTHFD1L gene predict cardiovascular mortality in males after acute coronary syndrome. *Mol Biol Rep* **42**: 1289-1293, 2015c.

HUBACEK JA, VRABLIK M, DLOUHA D, STANEK V, GEBAUEROVA M, ADAMKOVA V, CESKA R, DOSTÁLOVÁ G, LINHART A, VITEK L, PITHA J: Gene variants at FTO, 9p21, and 2q36.3 are age-independently associated with myocardial infarction in Czech men. *Clin Chim Acta* **454**: 119-123, 2016.

KATHIRESAN S, MELANDER O, GUIDUCCI C, SURTI A, BURTT NP, RIEDER MJ, COOPER GM, ROOS C, VOIGHT BF, HAVULINNA AS, WAHLSTRAND B, HEDNER T, CORELLA D, TAI ES, ORDOVAS JM, ET AL.: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**: 189-197, 2008.

KATHIRESAN S, WILLER CJ, PELOTO GM, DEMISSIE S, MUSUNURU K, SCHADT EE, KAPLAN L, BENNETT D, LI Y, TANAKA T, VOIGHT BF, BONNYCASTLE LL, JACKSON AU, CRAWFORD G, SURTI A, ET AL.: Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* **41**: 56-65, 2009.

KJOLBY M, ANDERSEN OM, BREIDERHOFF T, FJORBACK AW, PEDERSEN KM, MADSEN P, JANSEN P, HEEREN J, WILLNOW TE, NYKJAER A: Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metab* **12**: 213-223, 2010.

KULMINSKI AM, CULMINSKAYA I, ARBEEV KG, ARBEEVA L, UKRAINTSEVA SV, STALLARD E, WU D, YASHIN AI: Birth cohort, age, and sex strongly modulate effects of lipid risk alleles identified in genome-wide association studies. *PLoS One* **10**: e0136319, 2015.

LIU J, ZENG FF, LIU ZM. ZHANG CX, LING WH, CHEN YM: Effect of blood triglycerides on cardiovascular and all-cause mortality: a systematic review and meta-analysis of 61 prospective studies. *Lipids Health Dis* **12**: 159, 2013.

MILLER SA, DYKES DD, POLESKY HF: A simple salting out procedure for DNA extraction from human nucleated cells. *Nucleic Acid Res* **16**: 1215, 1988.

MORRIS RW, COOPER JA, SHAH T, WONG A, DRENOS F, ENGMANN J, McLACHLAN S, JEFFERIS B, DALE C, HARDY R, KUH D, BEN-SHLOMO Y, WANNAMETHEE SG, WHINCUP PH, CASAS JP, ET AL.: Marginal role for 53 common genetic variants in cardiovascular disease prediction. *Heart* **102**: 1640-1647, 2016.

MUNAFÒ MR: Reliability and replicability of genetic association studies. *Addiction* **104**: 1439-1440, 2009.

PENNACCHIO LA, OLIVIER M, HUBACEK JA, COHEN JC, COX DR, FRUCHART JC, KRAUSS RM, RUBIN EM: An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science* **294**: 169-173, 2001.

PIKHART H, HUBÁČEK JA, PEASEY A, KUBÍNOVÁ R, BOBÁK M: Association between fasting plasma triglycerides, all-cause and cardiovascular mortality in Czech population. Results from the HAPIEE study. *Physiol Res* **64** (Suppl 3): S355-S361, 2015.

PIŤHA J, KOVÁŘ J, BLAHOVÁ T: Fasting and nonfasting triglycerides in cardiovascular and other diseases. *Physiol Res* **64** (Suppl 3): S323-S330, 2016.

ROSENBERG NA, HUANG L, JEWETT EM, SZPIECH ZA, JANKOVIC I, BOEHNKE M: Genome-wide association studies in diverse populations. *Nat Rev Genet* **11**: 356-366, 2010.

SANDHU MS, WATERWORTH DM, DEBENHAM SL, WHEELER E, PAPADAKIS K, ZHAO JH, SONG K, YUAN X, JOHNSON T, ASHFORD S, INOUYE M, LUBEN R, SIMS M, HADLEY D, McARDLE W, BARTER P, ET AL.: LDL-cholesterol concentrations: a genome-wide association study. *Lancet* **371**: 483-491, 2008.

SANNA S, LI B, MULLAS A, SIDORE C, KANG HM, JACKSON AU, PIRAS MG, USALA G, MANINCHEDDA G, SASSU A, SERRA F, PALMAS MA, WOOD WH 3RD, NJØLSTAD I, LAAKSO M, ET AL.: Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double explained heritability. *PLoS Genet* **7**: e1002198, 2011.

SCHWARZOVA L, HUBACEK JA, VRABLIK M: Genetic predisposition of human plasma triglyceride concentrations. *Physiol Res* **64** (Suppl 3): S341-S354, 2015.

TALMUD PJ, HINGORANI AD, COOPER JA, MARMOT MG, BRUNNER EJ, KUMARI M, KIVIMÄKI M, HUMPHRIES SE: Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* **340**: b4838, 2010.

TESLOWICH TM, MUSUNURU K, SMITH AV, EDMONDSON AC, STYLIANOU IM, KOSEKI M, PIRRUCCELLO JP, RIPATTI S, CHASMAN DI, WILLER CJ, JOHANSEN CT, FOUCHIER SW, ISAACS A, PELOSO GM, BARBALIC M, ET AL.: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**: 707-713, 2010.

TUNSTALL-PEDOE H, KUULASMAA K, TOLONEN H, DAVIDSON M, MENDIS S: MONICA monograph and multimedia sourcebook 2003. TUNSTALL-PEDOE H (ed.), World Health Organisation, Geneva, 2003.

UITTERLINDEN AG: An introduction to Genome-wide association studies: GWAS for dummies. *Semin Reprod Med* **34**: 196-204, 2016.

VAN WIJK DF, STROES ES, KASTELEIN JJ: Lipid measures and cardiovascular disease prediction. *Dis Markers* **26**: 209-216, 2009.

VRABLIK M, CESKA R, ADAMKOVA V, PEASEY A, PIKHART H, KUBINOVA R, MARMOT M, BOBAK M, HUBACEK JA: MLXIPL variant in individuals with low and high triglyceridemia in white population in Central Europe. *Hum Genet* **124**: 553-555, 2008.

VRABLÍK M, HUBÁČEK JA, DLOUHÁ D, LÁNSKÁ V, RYNEKROVÁ J, ZLATOHLÁVEK L, PRUSÍKOVÁ M, ČEŠKA R, ADÁMKOVÁ V: Impact of variants within seven candidate genes on statin treatment efficacy. *Physiol Res* **61**: 609-617, 2012.

WANG X, MAGKOS F, MITTENDORFER B: Sex differences in lipid and lipoprotein metabolism: it's not just about sex hormones. *J Clin Endocrinol Metab* **96**: 885-893, 2011.

WEIJMANS M, DE BAKKER PI, VAN DER GRAAF Y, ASSELBERGS FW, ALGRA A, JAN DE BORST G, SPIERING W, VISSEREN FL; SMART STUDY GROUP: Incremental value of a genetic risk score for the prediction of new vascular events in patients with clinically manifest vascular disease. *Atherosclerosis* **239**: 451-458, 2015.

WILLER CJ, SANNA S, JACKSON AU, SCUTERI A, BONNYCASTLE LL, CLARKE R, HEATH SC, TIMPSON NJ, NAJJAR SS, STRINGHAM HM, STRAIT J, DUREN WL, MASCHIO A, BUSONERO F, MULAS A, ET AL.: Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* **40**: 161-169, 2008.

WONG B, KRUSE G, KUTIKOVA L, RAY KK, MATA P, BRUCKERT E: Cardiovascular disease risk associated with familial hypercholesterolemia: a systematic review of the literature. *Clin Ther* **38**: 1696-1709, 2016.