VILNIUS UNIVERSITY

INGRIDA DOMARKIENĖ

INVESTIGATION OF THE GENETIC STRUCTURE OF LITHUANIAN POPULATION, BASED ON THE ANALYSIS OF DISEASE-ASSOCIATED SINGLE NUCLEOTIDE POLYMORPHISMS

Summary of doctoral dissertation

Biomedical Sciences, Medicine (06 B)

Vilnius, 2014

The dissertation has been prepared at Vilnius University in the period 2010–2014.

Scientific supervisor:

Prof. Habil. Dr. Vaidutis Kučinskas (Vilnius University, Biomedical Sciences, Medicine – 06 B).

The defence of the dissertation will take place at Vilnius University, Academic Board of Examiners in Medicine:

Chairperson: Prof. Dr. Algirdas Utkus (Vilnius University, Biomedical Sciences, Medicine – 06 B).

Members:

Prof. Habil. Dr. Limas Kupčinskas (Lithuanian University of Health Sciences, Biomedical Sciences, Medicine – 06 B);

Prof. Dr. Gražina Slapšytė (Vilnius University, Biomedical Sciences, Biology – 01 B);

Prof. Dr. Vytautas Kasiulevičius (Vilnius University, Biomedical Sciences, Medicine – 06 B);

Prof. Dr. Andres Metspalu (Tartu University, Biomedical Sciences, Biology – 01 B).

The defence of the dissertation will be public at the meeting of the Academic Board of Examiners in Medicine on the 21st of November at 12 p.m. in the 1st Auditorium of the VUH Santariškių Klinikos (Audiences housing).

Address: Santariškių Str. 2, Vilnius, Lithuania

The summary of the doctoral dissertation was distributed on the 21st of October, 2014.

The doctoral dissertation is available at the Vilnius University Library and Vilnius Universitety website **http://www.vu.lt/lt/naujienos/ivykiu-kalendorius.**

VILNIAUS UNIVERSITETAS

INGRIDA DOMARKIENĖ

LIETUVOS POPULIACIJOS GENETINĖS STRUKTŪROS TYRIMAS REMIANTIS VIENO NUKLEOTIDO POLIMORFIZMŲ ASOCIACIJOS SU LIGA ANALIZE

Daktaro disertacijos santrauka

Biomedicinos mokslai, Medicina (06 B)

Vilnius, 2014

Disertacija rengta 2010–2014 metais Vilniaus universitete

Mokslinis vadovas:

prof. habil. dr. Vaidutis Kučinskas (Vilniaus universitetas, biomedicinos mokslai, medicina – 06 B).

Disertacija ginama Vilniaus universiteto Medicinos mokslo krypties taryboje:

Pirmininkas: prof. dr. Algirdas Utkus (Vilniaus universitetas, biomedicinos mokslai, medicina – 06 B).

Nariai:

prof. habil. dr. Limas Kupčinskas (Lietuvos sveikatos mokslų universitetas, biomedicinos mokslai, medicina – 06 B);

prof. dr. Gražina Slapšytė (Vilniaus universitetas, biomedicinos mokslai, biologija – 01 B);

prof. dr. Vytautas Kasiulevičius (Vilniaus universitetas, biomedicinos mokslai, medicina – 06 B);

prof. dr. Andres Metspalu (Tartu universitetas, biomedicinos mokslai, biologija – 01 B).

Disertacija bus ginama viešame Medicinos mokslo krypties tarybos posėdyje 2014 m. lapkričio mėn. 21 d. 12 val. VšĮ VUL Santariškių klinikų pirmoje auditorijoje (Auditorijų korpusas).

Adresas: Santariškių g. 2, Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2014 m. spalio mėn. 21 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir Vilniaus universiteto interneto svetainėje adresu: **http://www.vu.lt/lt/naujienos/ivykiu-kalendorius**.

CONTENT

ABBREVIATIONS 7
INTRODUCTION 9
1. METHODS 13
1.1. Strategy 13
1.2. Study population 14
1.2.1. Individuals' group for genome-wide genotyping. Genetic profile analysis of known selected SNPs in the Lithuanian population 14
1.2.2. Individuals' group for genome-wide genotyping. Pilot GWAS for new genetic loci identification 14
1.3. Workflow and methods 16
1.3.1. DNA extraction 16
1.3.2. *Illumina Infinium® HD* genotyping 17
1.3.3. PCR and Sanger sequencing 17
1.3.4. Statistical and bioinformatical analysis 17
1.3.4.1. Methods for the analysis of genetic variation of disease risk SNPs in Lithuanian population 17
1.3.4.2. Pilot GWAS analysis methods for new loci identification............18
2. RESULTS AND DISCUSSION 20
2.1. Selection of genes and SNPs 20
2.2. Determination of genetic structure and diversity according to the selected risk SNPs of CHD 21
2.2.1. Intra-population genetic structure and diversity 21
2.2.1.1. Allele frequencies 21
2.2.1.2. Haplotype blocks 26
2.2.1.3. Evaluation of CHD risk in the Lithuanian population 29
2.2.2. Inter-population genetic structure and diversity 31
2.3. Identification of new genetic loci. Association analysis 37
2.3.1. Family GWAS 37
2.3.2. Analysis of the potential candidate genes *RTN4* and *FBXL17* 42
CONCLUSIONS 46
REFERENCES 48
SANTRAUKA 52

APPROBATION 53
ABOUT THE AUTHOR 55
APIE AUTORĘ 56

ABBREVIATIONS

ATHEROGEN – VU research project "Pathogenetic mechanisms of atherosclerosis determined by human genome diversity", 2004–2006

BMI – body mass index

CEU – Utah Residents (CEPH) with Northern and Western European ancestry

CH – the abbreviation of Switzerland (according to the international ISO standard 3166-1)

CHD – coronary heart disease

CI –confidence interval (in statistics)

CRP – C reactive protein

D'– the deviation of the observed frequency of a haplotype from the expected and divided by the theoretical maximum of the observed allele frequencies ($D_{AB} = p_{AB} - p_A p_B$, $D' = D/D_{max}$, if $D > 0$; $D' = D/D_{min}$, if $D < 0$; A and B different genetic loci, p – frequency)

denom df – denominator degree of freedom (in statistics)

df – degree of freedom (in statistics)

DHMG – Department of Human and Medical Genetics

F – test statistic to compare two variances (in statistics)

FI – the abbreviation of Finland (according to the international ISO standard 3166-1)

FIN – Finnish in Finland

GBR – British in England and Scotland

gDNA – genomic DNA

GWAS – genome wide association study

HCC – healthcare centre

HDL – high density lipoprotein

LDL – low density lipoprotein

LIGENKOR – VU MF research project "Genomic diversity of the Lithuanian population and the possibility for the personalized treatment of coronary heart disease", 2010–2011

LITGEN – VU MF research project "Genetic diversity of the population of Lithuania and changes of its genetic structure associated with evolution and common diseases", 2011–2015

LOD – LOD score (logarithm (base 10) of odds) is a statistical test for linkage (Z), which compares the likelihood of obtaining the test data if the two loci are indeed linked, to the likelihood of observing the same data purely by chance

LTU – general population of Lithuanians from Lithuania (individuals representing six ethno-linguistic groups)

MAF – minor allele frequency

MF – Faculty of Medicine

MI – myocardial infarction

num df – numerator degree of freedom (in statistics)

OR – odds ratio (in statistics)

p – the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone, if the null hypothesis H_0, is true (in statistics)

PCR – polymerase chain reaction

RA – risk allele

RAF – risk allele frequency

RR – risk ratio (in statistics)

SNP – single nucleotide polymorphism

t – statistic for the differences between the two independent sample means (in statistics)

TSI – Toscani in Italia

USA – United States of America

UTR – untranslated region

VRBREC – Vilnius Regional Biomedicine Research Ethics Committee

VU – Vilnius University

VUH SK CMG – Vilnius University Hospital Santariškių Klinikos Centre for Medical Genetics

χ^2 – chi square statistic for the test of independence (in statistics)

INTRODUCTION

Topic, relevance, novelty, importance

The title of the dissertation "Lithuanian population genetic structure definition on the basis of analysis of disease associated single nucleotide polymorphisms" is complex and involves three important scientific fields – population genetic structure, genetic association studies, and disease. In this work, these fields are combined in order to find substantial, clinically significant and specific for Lithuanian population genomic markers of the disease. Coronary heart disease (CHD) was chosen for the study as a model / example of a multifactorial disease, thus, many other multifactorial diseases could be analysed according to this research strategy.

Genetic population structure, profile and diversity are the objects of the **genetic disease architecture** research. Holistically, genetic disease architecture is the structure of the genotype–phenotype relation. It includes the genetic nature of the phenotype diversity which is the characteristic of the genetic loci, i.e. the number of alleles in a particular locus, the effect size on the phenotype, the phenomena of pleiotropy and epistasis, dominance, additive and epigenetic effects, etc. Genetic architecture could be defined also as the distribution of the alleles in time and space and, if to be more precise, the evolutionary and biogeografic nature of the phenotypic difference according to the tendencies of the diversity of alleles. The deconstruction of the genetic disease architecture requires identifying the genetic constituents of the phenotypic diversity, i.e. the knowledge of how the diversity of the molecular structural and functional units forms a particular phenotypic feature and its variance. This was, and still is, one of the exceptional and fundamental problems of genetics. **Genetic profile** and its evaluation is one of the genetic disease architecture units that we have discussed before. Genetic profile in a strict sense could be understood directly as a technical genotyping for the evaluation of the alleles, genotype, and their combination. In a broad sense,

genetic profile means the frequency and diversity analysis of the alleles, genotypes as well as their combinations. Genetic profile could be evaluated in order not only to identify individuals, species, genetic similarities or differences, but also the qualitative and quantitative content of the known and unknown genetic disease predisposing factors. Genetic profile is usually understood in the context of the multifactorial diseases; therefore, we should think not about the pattern of the disease inheritance, but about the **predisposition** to the disease.

There are many genome-wide association studies (GWAS) performed in the field of multifactorial diseases; however, the majority of them were done in the **admixture** populations. This has led us to the knowledge of what is common for the humans in the sense of analysed genetic traits. Nevertheless, performing those studies, do we not miss what is specific and unique for the different populations? There is a growing number of publications emphasizing the need of analysing different / ethnic populations for better understanding the genetic disease architecture. Moreover, despite the aspect of scientific knowledge, there are some practical issues. According to the GWAS findings, there are plenty of commercial tests created and available for the prognosis of multifactorial diseases. Still we have no complete view of the GWAS results and associated genomic markers, and thus we cannot use them for medical practice. Therefore, an additional research is mandatory for a better understanding of human population genetics and improvement of the existing risk algorithms.

Coronary heart disease is multifactorial. Its morbidity, mortality, and costs of medical care are increasing not only in Lithuania, but also in the major part of the world. There are many known CHD risk factors, among which we can find the genomic factors as well. The molecular mechanisms of genomic factors are still unknown, and the need for their further research remains.

In Lithuania, the genome-wide studies are at an early stage, and this study is the first attempt to perform a genome-wide analysis in the field of CHD genomics. This work should be the ice breaker for the further research in the genetic CHD architecture of Lithuania. According to the strategy of this work, many multifactorial diseases could be explored at the population level. This work could also serve the clinical practice not only in creating a specific CHD risk evaluation algorithm but also better diagnostics, prognostics as well as prevention system. Therefore, this work is valuable from both the fundamental and applied sciences perspectives.

Aim, goals, statements to defend

The aim of this study is to explore the genetic multifactorial disease architecture in Lithuanian population by analysing common genomic markers – SNPs.

The main **goals** to achieve the aim are:

1. To select the disease as an appropriate model for the multifactorial disease analysis.
2. To compile the list of known common SNPs associated with the selected multifactorial disease in order to analyse the genetic diversity and genetic disease architecture in the Lithuanian population.
3. To perform the analysis of the genetic diversity of particular SNPs in the Lithuanian population and to compare with other populations:
    a. Determine the genetic structure of common SNPs associated with the disease by using the genome-wide genotyping data;
    b. Evaluate the prevalence of the common risk alleles and their combinations in the context of the heritable predisposition of multifactorial disease.
4. To define the new potential risk alleles for the multifactorial disease by performing the pilot GWAS in the Lithuanian population:
    a. Perform the GWAS in the CHD families;

b. Validate the potential candidate genes that were found to be significantly associated with CHD in the pilot GWAS by performing a case–control association study.

Following the study results, the **statements to defend** are:

1. According to the investigated risk SNPs of CHD, the Lithuanian population as compared with other populations is significantly different by the different genetic variants and the combinations of the allele frequencies. This leads to the hypothesis approval that there exist both common and unique genetic disease architecture variants for the different populations.
2. Regarding the genetic structure and diversity of the risk SNPs of CHD, the Lithuanian population as compared with the European populations falls into the Northern-Southern gradient. The potential explanation is that the Lithuanian gene pool could have experienced the geographical climate and ecological influence during the evolution process.
3. Considering the risk allele number per person of the investigated risk SNPs of CHD in the Lithuanian population, the majority of population individuals have a relatively average risk of developing CHD.
4. Genes *ITPR2* and *FBXL17* are the putative candidate genes that could participate in the pathogenesis of CHD and atherosclerosis.

1. METHODS

1.1. Strategy

This dissertation is part of several scientific projects (ATHEROGEN, LIGENKOR, LITGEN) of VU MF DHMG. All projects were approved by the VRBREC. Written informed consents were obtained from all of the participants of the projects.

The strategy and model of the dissertational work is shown in **Figure 1.1**. There are two parts of the study – the analysis of known genetic loci associated with the relevant disease in the Lithuanian population (I) and the search of the new genetic loci associated with the relevant disease in the Lithuanian population (II).

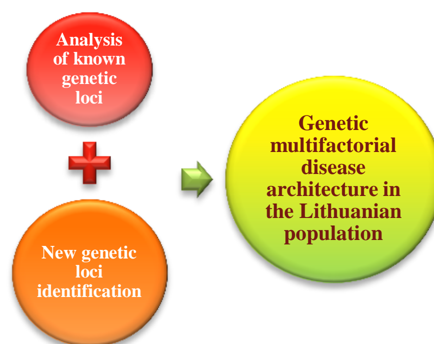

Figure 1.1. The strategy of the genetic multifactorial disease architecture analysis.

(I) Firstly, after the data mining, the list of known SNPs associated with CHD has been compiled. After the genotyping, the intra-population and inter-population analysis has been performed by the frequency analysis of targeted SNPs' alleles as well as genotypes and determination of the haplotypes, genetic profile and its diversity.

(II) The pilot GWAS has been performed in the Lithuanian CHD families. After the identification, the validation analysis of the new genetic loci followed. The case-control analysis has been performed.

1.2. Study population

1.2.1. Individuals' group for genome-wide genotyping. Genetic profile analysis of known selected SNPs in the Lithuanian population

Samples were collected during the scientific project LITGEN in the period 2011–2013. For the genetic analysis, in order to represent the Lithuanian population it is reasonable to divide the population into Aukštaičiai and Žemaičiai and further into the six ethno-linguistic groups according to the dialects: Eastern, Western, Southern Aukštaičiai and Northern, Southern, Western Žemaičiai (**Figure 1.2**) (*Girdenis A, Zinkevičius Z, 1966*).

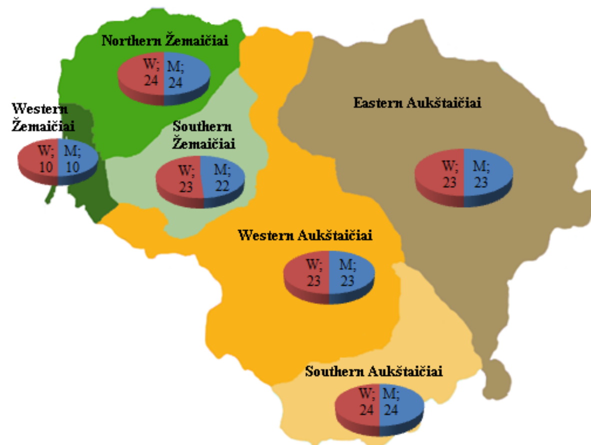

Figure 1.2. Study population representing general Lithuanian population. The number of DNA samples of women and men from the ethno-linguistic regions of Lithuania are shown in the pie charts (ethno-linguistic regions of Lithuania are shown according to the map No. 1, *Morkūnas K, 1977*).

All in all, there were 253 unrelated individuals recruited from the six particular areas – 113 Žemaičiai (20 Western, 48 Northern, 45 Southern) and 140 Aukštaičiai (46 Eastern, 46 Western, 48 Southern), or 127 women and 126 men.

1.2.2. Individuals' group for genome-wide genotyping. Pilot GWAS for new genetic loci identification

The premature coronary heart disease (further CHD) phenotype was selected for the analysis.

There are two more parts of this stage of the study. Actually, two groups of individuals are needed: CHD families group (CHD proband and his parents) for family GWAS and the case-control group for the validation study (**Figure 1.3**).

CHD families. There are 32 probands (1 woman, 31 men) and 64 parents.

CHD patients (cases). This group consists of the same 32 CHD probands and additional 61 patient (7 women and 86 men).

Control group. The control group is represented by 48 healthy individuals from the Lithuanian population (24 men and 24 women or 8 individuals per each ethno-linguistic region).

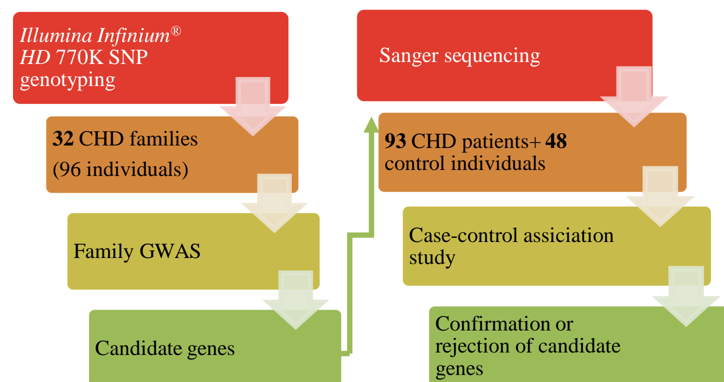

Figure 1.3. The example model of the identification of new risk alleles of multifactorial disease. In the scheme there are methods, number of analysed samples and study designs provided. Arrows denote the sequence of the experiment.

CHD phenotype. Patients were clinically examined at the Clinic of Cardiology and Angiology, MF VU, Vilnius, Lithuania. The patients' clinical phenotype was evaluated upon assessing anthropometrical measurements, clinical and instrumental examination, and laboratory biochemical testing. Information about CHD risk factors, other diseases and treatment was obtained during the conventional anamnesis.

Patient recruitment criteria were as follows: men aged 35–55, women aged 35–65; individuals who were experiencing acute coronary

syndrome for the first time in their lives, who were hospitalized at an intensive cardiology unit for myocardial infarction (MI) with or without Q wave or unstable angina pectoris (confirmed by examination of common electrocardiographic and / or coronographic changes, assessment of cardio-specific markers); any previous or current evidence of significant atherosclerotic CHD (MI, percutaneous coronary angioplasty, coronary artery bypass graft or coronary angiography with hemodynamically significant stenosis). Anatomical vascular changes were confirmed by non-invasive methods evaluating the presence of the atherosclerotic plaques as well as examining the arterial stiffness and endothelial function. Patient exclusion criteria were as follows: diabetes treated with insulin, kidney function deficiency, III–IV functional class of heart deficiency, tumours (except skin basalioma), alcoholism and other social factors that may influence the study results.

The evaluation of biochemical phenotype was performed at the Centre of Laboratory Medicine of Vilnius University Hospital Santariškių Klinikos and included inflammatory and metabolic markers, participating in the pathogenesis of atherosclerosis: C-reactive protein (CRP), Hb A_{1c}, lipoprotein (Lp(a)), apolipoprotein A1 (ApoA1), apolipoprotein B (ApoB), ratio ApoB/A1, lipids, oxidized low-density lipoprotein (oxLDL), homocysteine, interleukin-6 (IL-6), fasting glucose in plasma, potassium (K), sodium (Na), urea, creatinin.

1.3. Workflow and methods

1.3.1. DNA extraction

The gDNA, extracted from venous blood in VU MF DHMG, was used during all stages of this study. There were two methods of gDNA extraction: either phenol-chloroform (according to the approved diagnostic protocol of the VUH SK CMG, *DBR išskyrimas fenolio chloroformo metodu*) or using the magnetic beads (for the protocol, refer the reference list) on the

automatic robotic system *TECAN Freedom EVO® 200* (manufacturer *Tecan Schweiz AG,* CH).

1.3.2. *Illumina Infinium® HD* genotyping

The genotyping was performed on the *Illumina HiScan™SQ* instrument using the *Illumina Infinium® HD* SNP assays *HumanOmniExpress-12 v1.0* (733,202 SNPs) and *HumanOmniExpress-12 v1.1* (719,666 SNPs).

The workflow of the genotyping was performed according to the protocols provided by the manufacturer: *Infinium® HD Assay Ultra Manual Experienced User Card*, *Infinium® HD Assay Ultra Protocol Guide*, *Illumina Infinium® Assay Lab Set Up and Procedures*.

The genotyping by using *Illumina Infinium® HD* SNP assay *HumanOmniExpress-12 v1.0* was performed at the Tartu University Institute of Molecular and Cell Biology (Estonia) according to the *Infinium® Multi-Use Assay, Manual Protocol, 15013850 Rev. A* protocol which is almost identical to the protocol that was used for genotyping with *HumanOmniExpress-12 v1.1* assay in VU MF DHMG.

1.3.3. PCR and Sanger sequencing

Sanger sequencing was used after the pilot GWAS for the hypothesis-driven analysis of the potential candidate genes.

1.3.4. Statistical and bioinformatical analysis

1.3.4.1. Methods for the analysis of genetic variation of disease risk SNPs in Lithuanian population

Genotyping data were analysed according to the relevant *Genome Studio™ Genotyping Module v1.0 User Guide* and *GenomeStudio™ Software v2 011.1 Release Notes*.

The Hardy–Weinberg equilibrium, frequency of genotypes and alleles were calculated as well as input files for *Haploview v4.2* were generated

with the *PLINK v1.07* software (*Purcell S et al., 2007*; http://pngu.mgh.harvard.edu/purcell/plink/).

The calculations made with the *R v3.0.3* software (*R Development Core Team, 2008;* http://www.R-project.org) are described in **Table 1.1**. Graphical visualisation was also done with the *R* and Microsoft Office Excel 2007 software.

Table 1.1. Calculations made with *R v3.0.3* software

Statistic		Remarks
Descriptive statistics	Mean	-
	Mode	-
	Median	-
	Quartiles	-
χ^2	Pearson's	-
	Fisher's	-
Distribution normality test	Graphical method	For the distribution of the RA number per person
***t* test**	Comparison of means of two independent samples	The mean of the RA number per person in the groups of either men and women or Aukštaičiai and Žemaičiai
***F* test**	Comparison of two variances	The variance of the RA number per person in the groups of either men and women or Aukštaičiai and Žemaičiai

1.3.4.2. Pilot GWAS analysis methods for new loci identification

The data of genotyped CHD families were analysed with the *PLINK v1.07* software integrated in the *BC|Gene* platform (*Biocomputing Platforms Ltd.,* FI). Data quality control was performed according to the criteria provided in **Table 1.2**.

Table 1.2. Criteria of genotyping data quality control

Criterion	Abbreviation	Threshold value	Desired value
Minor allele frequency	MAF	0.01	> 0.01
Missingness per marker	GENO	0.1	< 0.1
Significance level of Hardy-Weinberg equilibrium	-	0.0005	> 0.0005
Missingness per individual	MIND	0.05	< 0.05

After the quality control procedure, the *Family GWAS* test was performed. The adaptive permutation (maximum number of permutations 10^6) procedure was performed as the multiple comparison correction. The significance level for the analysis was 10^{-4}.

The empirical power of the test, OR and 95% CI for the significant SNPs were calculated with the R *v2.15.3* software (*R Development Core Team, 2008*). The statistical empirical power was calculated according to the transmitted alleles by using the approximation of the binomial distribution (*Germanas Š, Jakaitienė A, 2012*). The odds ratio was calculated as a number of transmitted risk alleles ratio with the number of alternative alleles. Confidence intervals were calculated according to the formula (L, U) = (exp{T-1.96SE}, exp{T+1.96SE}), where L and U are the border values of the 95% CI, T – logarithmic value of the OR, SE – the standard error of the logarithmic OR calculated according to the formula SE = $\sqrt{(1/b + 1/c)}$, where b and c are the numbers of the transmitted alleles, respectively.

After the pilot family GWAS, the confirmation of the loci and potential candidate genes was performed in the case–control association study evaluating the χ^2 (Pearson's or Fisher's) calculated with the *R v3.0.3* software (*R Development Core Team, 2008*). Before the association analysis, all tested SNPs identified during the Sanger sequencing were tested for the Hardy–Weinberg equilibrium, and the frequencies of the alleles were determined by using online software (*Rodriguez S et al., 2009*; *http://www.oege.org/software/hardy-weinberg.html*). The significance level was set at 0.001. The frequencies of genotypes were determined by direct count. OR and RR with the 95% CI were calculated according to the *Altman DG, 1991* by using the *MedCalc v13.1.2* software (*MedCalc Software, Belgium*; *http://www.medcalc.org*).

2. RESULTS AND DISCUSSION

2.1. Selection of genes and SNPs

For the analysis of the genetic multifactorial disease architecture, the CHD was chosen as a disease model. Referring the scientific literature and *Ensembl*, *NCBI Gene*, *dbSNP*, *1000 Genomes* databases, the list of SNPs associated with CHD was compiled (**Table 2.1**).

Table 2.1. The list of SNPs and risk alleles associated with the CHD

Genes or loci	SNP	RA*	Reference	Genes or loci	SNP	RA*	Reference
APOB	rs1042031	A	*3*	***FURIN/FES***	rs17514846	A	*2*
9p21	rs10757274	G	*1*	***MTHFR***	rs1801133	A	*3*
RXRA	rs11185660	A	*6*	***HDAC9***	rs2023938	G	*2*
LIPA	rs11203042	A	*2*	***SLC22A3/LPAL2/LPA***	rs2048327	G	*2*
PCSK9	rs11206510	A	*2; 7*	***TOMM40 (APOE/APOC1)***	rs2075650	G	*2*
ZC3HC1	rs11556924	G	*2; 7*	***LPL***	rs2197089	A	*4*
PCSK9	rs11591147	A	*4*	***ZEB2***	rs2252641	G	*2*
GCKR	rs1260326	A	*4*	***HNF1A/C12orf43***	rs2259816	A	*5*
LIPA	rs1412444	A	*7*	***LCAT***	rs2271293	A	*4*
ABCC6	rs150468	A	*6*	***MRAS***	rs2306374	G	*7*
UBE2Z	rs15563	G	*2*	***9p21***	rs2383206	G	*1;8*
MLXIPL	rs17145738	A	*4*	***9p21***	rs2383207	G	*1*
SMAD3	rs17228212	G	*5*	***LIPC***	rs261332	A	*4*
FADS1	rs174547	A	*4*	***USF1***	rs2774279	G	*1*
MIA3	rs17464857	A	*2*	***SH2B3***	rs3184504	A	*2; 7*
MIA3	rs17465637	C	*5; 7*	***CDKN2BAS1***	rs3217992	A	*2*
CETP	rs3764261	A	*4*	***SORT1***	rs599839	A	*5; 4; 7*
PLG	rs4252120	A	*2*	***CELSR2/SORT1***	rs646776	A	*5; 4*
APOE/APOC4	rs4420638	A	*4*	***PON1***	rs662	G	*3*
APOE	rs445925	G	*2*	***LDLR***	rs688	A	*1*
COL4A1/COL4A 2	rs4773144	G	*2; 7*	***C6orf105***	rs6903956	A	*7*
LIPC	rs4775041	C	*4*	***APOB***	rs693	A	*4*
LIPC	rs4775065	A	*6*	***ADAMTS7***	rs7173743	A	*2*
IL6R	rs4845625	A	*2*	***SELP***	rs732314	A	*6*
CDKN2A/CDKN 2B	rs4977574	G	*7; 1*	***GCKR***	rs780094	A	*4*
CXCL12	rs501120	A	*5; 2*	***CUBN***	rs7893395	G	*6*
APOB	rs562338	A	*4*	***PHACTR1***	rs9369640	A	*2*
ABO	rs579459	G	*2; 7*	***COL4A1/COL4A2***	rs9515203	A	*2*
CETP	rs5882	A	*6*	***PDGFD***	rs974819	A	*2; 7*
ITGB3	rs5918	G	*3*	***MRAS***	rs9818870	A	*5; 2*

*The provided risk alleles could not correspond with those given in the literature as far as it depends on the DNA strand (forward or reverse) which is chosen as a reference. (1) *Correia M et al., 2013;* (2) *Deloukas P et al., 2013;*(3) *Drenos F et al., 2007;* (4) *Dumitrescu L et al., 2011;* (5) *Humphries SE et al., 2010;* (6) *Peloso GM et al., 2010;* (7) *Roberts R, Stewart AFR, 2012;* (8) *Preuss M, CARDIoGRAM et al., 2010*)

Coronary heart disease is complex and occurs as an atherosclerosis complication. The etiopathogenesis of atherosclerosis is also a multifactorial process. Thus, there is no single biological pathway which determines the mechanism of the disease. This makes it difficult to define the precise phenotype of the disease for the association study. That is way we have selected both genetic loci associated directly and indirectly with CHD.

2.2. Determination of genetic structure and diversity according to the selected risk SNPs of CHD

After the genotyping of 253 individuals, 60 SNPs were selected for the further evaluation. All samples and markers met the quality criteria. All the SNPs analysed were in the Hardy–Weinberg equilibrium ($p \geq 0.001$).

2.2.1. Intra-population genetic structure and diversity

2.2.1.1. Allele frequencies

Intra-population genetic structure and diversity were evaluated according to the frequency analysis of alleles and genotypes of selected SNPs in the general Lithuanian population and by genders (men and women) as well as ethno-linguistic (Aukštaičiai and Žemaičiai) groups. Moreover, the haplotypes, haplotype blocks, and the number of risk alleles per person were analysed.

The distribution of frequencies of the selected SNPs' risk alleles in the Lithuanian population is shown in **Figure 2.1**. The rarest allele (of rs11591147 SNP) is in the coding sequence of the *PCSK9* gene (1st chromosome). The most common allele is of rs445925 SNP in the *APOE* gene locus (19th chromosome). The frequency of 25 risk alleles out of 60 analysed (i.e. almost 42% of alleles) was more than 50%.

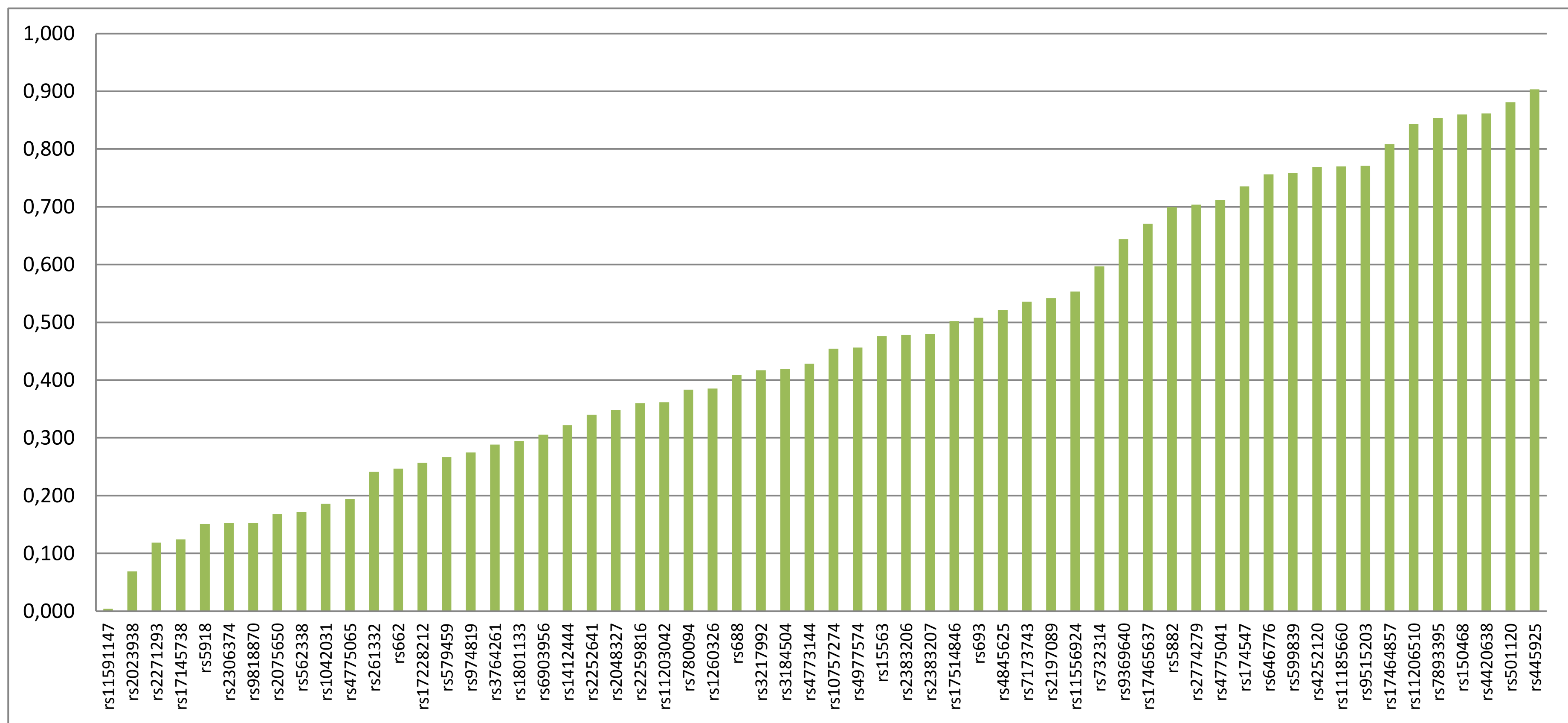

Figure 2.1. Frequencies of selected SNPs' risk alleles in the study population (LTU). X axis – SNP, Y axis – risk allele frequency.

The distribution of frequencies of the selected risk alleles in the gender groups of the Lithuanian population is shown in **Figure 2.2**. The rarest allele in both groups (men and women) is the same as in the general population – rs11591147, as well as the most common allele – rs445925.

The distribution of frequencies of the selected risk alleles in the ethno-linguistic groups of Lithuanian population is shown in **Figure 2.3**. The rarest allele in both groups (Aukštaičiai and Žemaičiai) is the same as in the general population – rs11591147; the most common allele in the Aukštaičiai group is rs445925, and in the Žemaičiai group the most common is the allele rs501120 SNP which is associated with the *CXCL12* gene (10th chromosome).

The rarest rs11591147 (A = 0.004) and the most frequent rs445925 (G = 0.903) alleles in the Lithuanian population correspond with the information provided in the NCBI dbSNP database where the frequencies are similar – 0.009 and 0.869, respectively. After comparing the groups of population, there were no statistically significant differences in the frequencies of analysed alleles among gender groups ($p > 0.05$), but there were two statistically significant ($p < 0.05$) differences among Aukštaičiai and Žemaičiai (**Table 2.2**).

Table 2.2. Statistically significantly different allele frequencies found between the Aukštaičiai and the Žemaičiai groups

SNP	RA	RAF		χ^2	*p*
		AUK	ŽEM		
rs11206510	A	**0.814**	**0.881**	4.166	0.041
rs15563	G	**0.522**	**0.420**	5.121	0.024

AUK – Aukštaičiai group, ŽEM – Žemaičiai group.

Considering the frequency differences of those alleles (rs11206510 SNP and rs15563 SNP), the assumption could be that there might be a difference in the CHD risk pattern according to the ethno-linguistic group in the Lithuanian population.

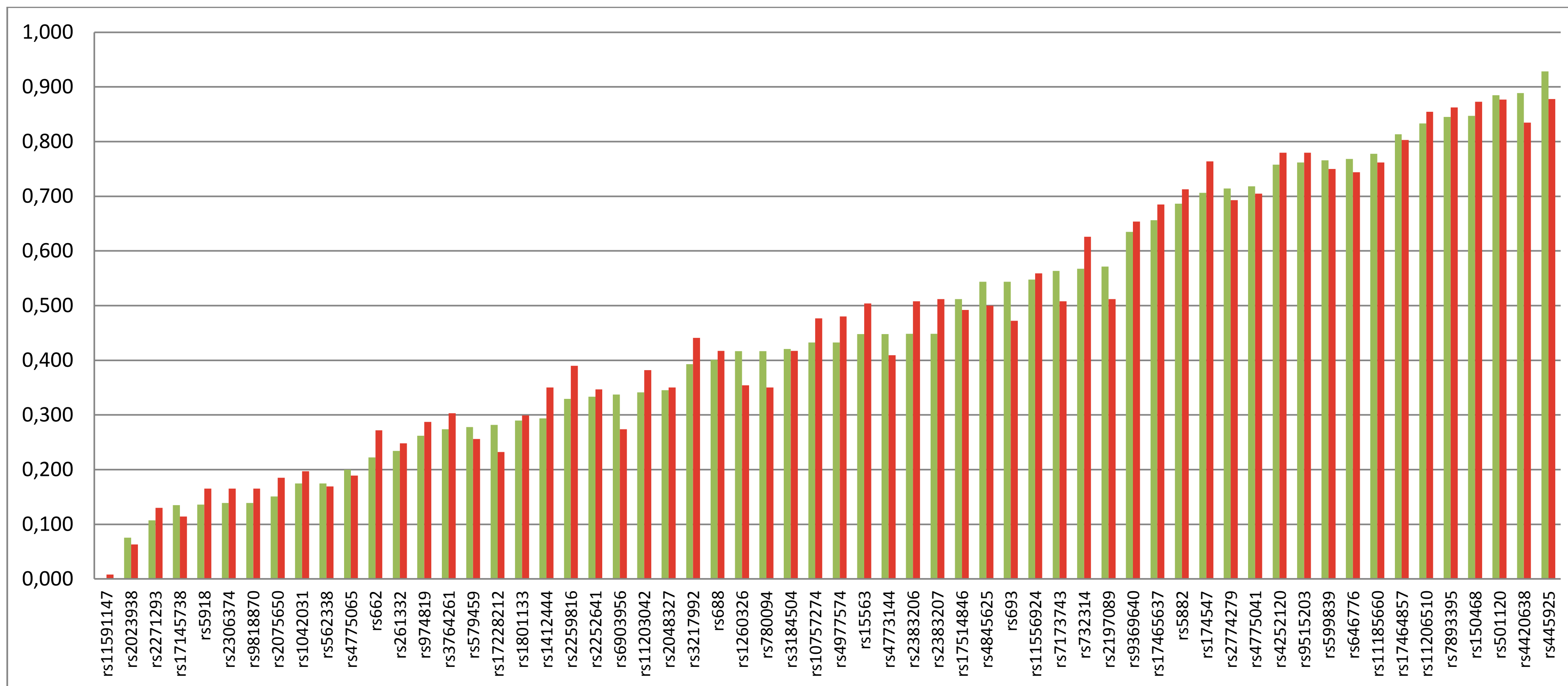

Figure 2.2. Frequencies of selected SNPs' risk alleles in the groups of men and women. X axis – SNP, Y axis – risk allele frequency. Green bars – men red bars – women.

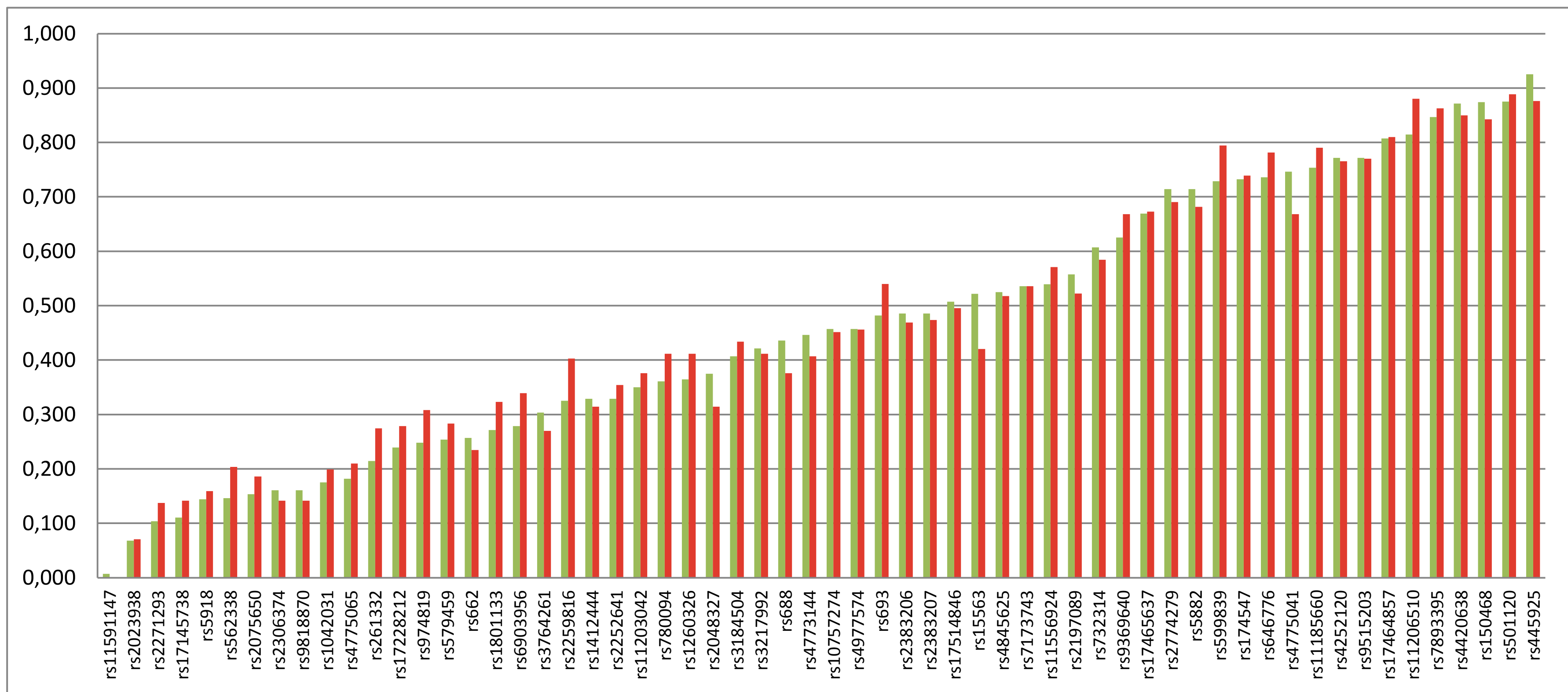

Figure 2.3. Frequencies of selected SNPs' risk alleles in the groups of Aukštaičiai and Žemaičiai. X axis – SNP, Y axis – risk allele frequency. Green bars – Aukštaičiai, red bars – Žemaičiai.

2.2.1.2. Haplotype blocks

In order to define the haplotypes and haplotype blocks of the selected SNPs in the Lithuanian population, we have divided 60 SNPs into the groups by chromosomes. There were 14 groups in total (rs2197089 was excluded from the analysis because it was sole in the 8th chromosome). The haplotype blocks, haplotypes and their frequencies were derived with the *Haploview v4.2* software (*Barrett JC et al., 2005*). The results are shown in **Table 2.3**.

Table 2.3. The haplotype groups determined with the *Haploview* software in the Lithuanian population according to the risk SNPs of CHD

Chr	SNPs	#SNP	# H-type blocks, SNPs in the haplotype		H-type	Frequency of the haplotype				
						G	M	W	A	Ž
1	rs646776, rs599839, rs17465637, rs732314, rs4845625, rs2774279, rs1801133, rs17464857, rs11591147, rs11206510	10	1	rs646776, rs599839	AA	0.75	0.77	0.74	0.73	0.78
					GG	0.24	0.23	0.25	0.26	0.21
2	rs780094, rs693, rs562338, rs2252641, rs1260326, rs1042031	6	2	rs1042031, rs693	GA	0.51	0.54	0.47	0.48	0.54
					GG	0.31	0.28	0.33	0.34	0.26
					AG	0.19	0.18	0.2	0.18	0.2
				rs1260326, rs780094	GG	0.62	0.58	0.65	0.64	0.59
					AA	0.38	0.42	0.35	0.36	0.41
3	rs9818870, rs2306374	2	1	rs9818870, rs2306374	AG	0.85	0.86	0.84	0.84	0.86
					GA	0.15	0.14	0.16	0.16	0.14
9	rs2383207, rs2383206, rs11185660, rs579459, rs4977574, rs3217992, rs10757274	7	1	rs1075727, rs4977574, rs2383206, rs2383207	GGGG	0.52	0.55	0.49	0.51	0.53
					AAAA	0.46	0.43	0.48	0.46	0.45
					AAGG	0.02	0.02	0.03	0.03	0.01
19	rs688, rs445925, rs4420638, rs2075650	4	1*	rs2075650, rs445925	AG	-	-	0.69	-	-
					GG	-	-	0.19	-	-
					AA	-	-	0.12	-	-

*Found only in the women's group; Chr – chromosome, # – number, H-type – haplotype, G – general Lithuanian population, M – the group of men in the Lithuanian population, W – the group of women in the Lithuanian population, A – Aukštaičiai group in the Lithuanian population, Ž – Žemaičiai group in the Lithuanian population. Risk alleles are shown in red.

The largest haplotype block (consisting of four SNPs) was identified in the 9th chromosome while the haplotype blocks identified in other chromosomes had only two SNPs each. The haplotype block of the 9th chromosome does not differ from the blocks derived in other populations (*Correia M et al., 2013*). This, though, confirms the hypothesis of common genetic risk factors for the general population of humans (*Visscher PM et al., 2012 (1)*). However, we

should not forget about the possibility of different combinations of these factors in different populations.

Analysing the haplotypes comprised only of the risk alleles, it was found that in all cases they were rarer than other haplotypes, except in the haplotype block of the 1st chromosome. The particular haplotype of the 1st chromosome (rs599839 and rs646776 risk alleles (AA)) was more frequent than the alternative one. This particular haplotype had the similar frequency in the Asian-Indian cohort (0.72) and found to be associated with CHD and potentially able to influence the plasma cholesterol level (*Arvind P et al., 2014*).

In the 2nd chromosome, there were two haplotype blocks identified (rs1042031 and rs693; rs1260326 and rs780094). All other chromosomes with the identified haplotype blocks had single ones. The haplotype block of rs1260326 and rs780094 was also identified in Danish and French populations, and the haplotype of the risk alleles (AA) is associated with the risk factors of CHD (*Tam CH et al., 2009*). Between those two identified haplotype blocks, the linkage disequilibrium as well as the recombination are possible. According to the D' statistics, the haplotype linkage in the groups of women and Žemaičiai (D' = 0.13) showed a higher linkage equilibrium than the groups of men (D' = 0.28) and Aukštaičiai (D' = 0.30). This might in part substantiate the statement that Žemaičiai compared with Aukštaičiai were former residents of a particular region (*Kučinskas V, 2004*). As far as it is known, the longer the population exists, the more time it had for recombination events to occur, and the genetic diversity increases while its linkage declines (*The International HapMap Consortium, 2005*). As for the women group, the explanation could be the biological phenomenon of the more frequent recombination during the chromosomal crossover (*Coop G, Przeworski M, 2007*). What is more, the haplotype block (rs2075650 (gene *TOMM40*) and rs445925 (gene *APOC1*)) in the 19th chromosome is identified only in the women's group. It might be that the linkage disequilibrium (D' = 1)

of those SNPs occurred only in the group of women non-accidentally. The linkage of those SNPs was also identified in the CEU cohort (*Bekris LM et al., 2012*). There is a known association of the rs2075650 in the *TOMM40* gene with the quantitative traits of Alzheimer's disease. Moreover, the SNP of the *TOMM40* is in the linkage disequilibrium with the loci of the *APOE* region (the second SNP (rs445925) of the haplotype block from our study also falls into that region). According to the performed functional studies, the *TOMM40* and *APOE* expression could be regulated by the regional haplotypes which may have the biological effect on the pathogenesis of Alzheimer's disease (*Bekris LM et al., 2012*). The link between Alzheimer's disease and the gender was already noticed. Experiments with mice showed that depending on the tissue, oestrogens could regulate the *apoE* expression (*Srivastava RA et al., 1996*). Recently, it has been shown that the gender modifies the susceptibility to the *APOE*-associated Alzheimer's disease, and this explains way Alzheimer's disease affects women more often than men (*Altmann A et al., 2014*).

The identified haplotype block in the 3rd chromosome (SNPs rs9818870 and rs2306374) consists of two SNPs that were associated with CHD in other studies. The first variant associated with CHD was rs9818870, and it was shown that the risk allele A *in cis* works as a silencer for the *MRAS* expression in the vascular tissue (*Folkersen L et al., 2010*). Nevertheless, in the *CARDIoGRAM* study, the major association with CHD was shown for the other SNP – rs2306374, and the risk allele (G) has been shown to enhance the expression of *FAIM* (*Fas apoptotic inhibitory molecule*). These results suggest that either one or the other, or even maybe both variants could mediate the susceptibility risk for CHD (*Maouche S, Schunkert H, 2012*). Our study results suggest that both variants could play an important role in the predisposition to CHD risk.

2.2.1.3. Evaluation of CHD risk in the Lithuanian population

According to the selected 60 CHD-associated SNPs, the number of risk alleles per person in the Lithuanian population as well as in the intra-population groups was determined. Theoretically, if there would be a person homozygous for each of the 60 SNPs' risk alleles, the maximum number of risk alleles would be 120. The distribution of the number of risk alleles per person in the Lithuanian population is provided in the **Figure 2.4**. After the graphical testing (performed with *R* software), our data fit the normal distribution. The descriptive statistics of the number of risk alleles per person is given in **Table 2.4**.

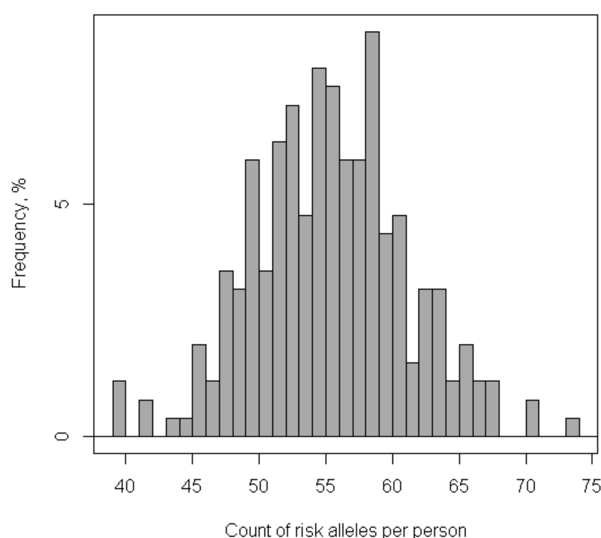

Figure 2.4. The distribution of the number of risk alleles per person in the Lithuanian population (%).

Table 2.4. Statistics of the number of risk alleles per person in the Lithuanian population and itra-population groups: men and women, Aukštaičiai and Žemaičiai

Statistics	Value	Value, women	Value, men	Value, Aukštaičiai	Value, Žemaičiai
The mean of the number of risk alleles per person	55.81	56.10	55.51	55.45	56.25
Median	56	56	55	55.5	56
Mode	59	59	55	59	56
Minimum value	39	39	42	39	42
1-quartile (25%)	52	53	51	52	52
2-quartile (50%)	56	56	55	55.5	56
3-quartile (75%)	59	60	59	59	60
Maximum value	74	71	74	71	74

The minimum and maximum numbers of risk alleles per person were 39 and 74, respectively (observed in 0.4% of individuals in both cases). The mean value of the number of risk alleles per person in the population is approximately ~56, i.e. almost half of alleles of all the 120 possible. The value of the 3rd quartile shows that 25% of individuals of the Lithuanian population had the greater number of risk alleles than the other 75%. After dividing individuals into three groups by the number of risk alleles (39–50 (I); 51–62 (II); 63–74 (III)), the majority of individuals (68%) fell into the 2nd and the minority into the 3rd group (13%). Referring the number of risk alleles per person in the groups, we could presume that individuals in these groups could be relatively divided into those having small, medium or large number of risk alleles. This brings us to the assumption about the CHD risk in the population. It is not so easy to evaluate it when it comes to the genetic factors. Many authors assert that risk alleles of SNPs themselves have a small effect on the multifactorial phenotype, i.e. the OR is between 1 and 2, but the additive effect of multiple risk alleles could be undeniable (*Manolio TA et al., 2009*). And if the hypothesis about the additive effect of multiple risk alleles is correct, we can conclude that, regarding the number of risk alleles per person, the majority of Lithuanian population individuals are at an average CHD risk. This statement could be justified by other prospective study results when according to the number of risk alleles the majority of individuals belonged to the group of average risk (*Humphries SE et al., 2010*).

A comparison of the mean and variance of the number of risk alleles per person in the intra-population groups (men–women; Aukštaičiai–Žemaičiai) has revealed no statistically significant differences. Thus, the population according to the number of risk alleles per person is homogeneous.

The existing CHD risk algorithms are non-universal and limited. A combination of both the genetic risk prediction algorithm and that which evaluates conventional risk factors could be created as a more accurate risk

prognosis algorithm than the separate ones. This particular study could be the background for creating the algorithm specific of the Lithuanian population.

2.2.2. Inter-population genetic structure and diversity

The genetic profile among the compared populations was determined by evaluating the risk allele frequencies. The Lithuanian population was compared with the populations that met the criteria of ancestry, geographical position, the size of the study population, and data availability. Thus, there were four European ancestry populations selected from the database of the 1000 Genomes project: Utah residents with Northern and Western European ancestry (CEU), British in England and Scotland (GBR), Finnish in Finland (FIN), and Toscani in Italy (TSI), i.e. populations representing three geographical poles – North, West, and South. The direct allele count was used to determine the frequencies of alleles. There were 506 chromosomes analysed in the Lithuanian population, 170 in CEU, 178 in GBR, 186 in FIN, and 196 in TSI.

The distribution and tendency of frequencies of risk alleles in the Lithuanian population as compared with CEU, FIN, GBR, and TSI is shown in **Figure 2.5.** The frequency of 25 risk alleles (i.e. almost 42% of alleles) was higher than 0.5 in the Lithuanian population. According to the 1000 Genomes Project respective data in the general population of humans, the tendency of these allele frequencies corresponds to them. The rarest allele in the compared populations was of the same SNP rs11591147 in the *PCSK9* gene. The most common allele in all compared populations, except Finnish, was the same allele of SNP rs445925 in the *APOE* gene region. The statistically significant differences of risk allele frequencies in the Lithuanian population as compared with CEU, FIN, GBR and TSI populations are presented in **Table 2.5**. It is obvious that the frequencies of risk alleles are different among populations, and there is a frequency variation in different populations. Despite this fact, **Figure 2.5** shows a visible tendency in the curve poles. The rarest and the most

common alleles are distributed more evenly if compared with those within the range. This means that alleles in the middle of the frequency distribution tend to be more scattered. That could be the general characteristic of a population and multifactorial diseases' genomics, i.e. both rare and common risk alleles are general for human populations, and the genetic diversity which lies in between the rare and common is this particularly unique combination of alleles which is specific to each ethnic population.

Table 2.5. Frequencies of alleles found to be statistically significantly different in the LTU as compared with CEU, FIN, GBR or TSI

SNP	RA	RAF		χ^2	*p*
		LTU	**CEU**		
rs11203042	A	0.362	0.447	3.927	0.048
rs15563	G	0.476	0.576	5.113	0.024
rs17228212	G	0.257	0.353	5.807	0.016
rs174547	A	0.735	0.653	4.219	0.040
rs2023938	G	0.069	0.124	4.949	0.026
rs2252641	G	0.340	0.441	5.626	0.018
rs2774279	G	0.704	0.594	6.959	0.008
rs732314	A	0.597	0.459	9.847	0.002
		LTU	**FIN**		
rs1042031	A	0.186	0.258	4.358	0.037
rs11203042	A	0.362	0.452	4.644	0.031
rs11556924	G	0.553	0.742	20.222	6.897×10^{-6}
rs11591147	A	0.004	0.038	12.020	5.264×10^{-4}
rs1260326	A	0.385	0.296	4.740	0.029
rs1412444	A	0.322	0.425	6.295	0.012
rs15563	G	0.476	0.624	11.830	5.829×10^{-4}
rs174547	A	0.735	0.532	25.764	3.859×10^{-7}
rs17464857	A	0.808	0.935	16.578	4.669×10^{-5}
rs17514846	A	0.502	0.387	7.191	0.007
rs2048327	G	0.348	0.269	3.860	0.049
rs261332	A	0.241	0.129	10.262	0.001
rs4420638	A	0.862	0.758	10.534	0.001
rs4775065	A	0.194	0.312	10.699	0.001
rs562338	A	0.172	0.242	4.317	0.038
rs693	A	0.508	0.376	9.445	0.002
rs974819	A	0.275	0.177	6.905	0.009
		LTU	**GBR**		
rs11203042	A	0.362	0.506	11.379	7.429×10^{-4}
rs174547	A	0.735	0.640	5.745	0.017
rs2048327	G	0.348	0.466	7.855	0.005
rs4845625	A	0.522	0.404	7.243	0.007
rs662	G	0.247	0.343	6.086	0.014
rs732314	A	0.597	0.449	11.603	6.586×10^{-4}
		LTU	**TSI**		
rs10757274	G	0.455	0.561	6.438	0.011
rs11203042	A	0.362	0.469	6.881	0.009
rs1260326	A	0.385	0.515	9.780	0.002
rs150468	A	0.860	0.796	4.345	0.037
rs17145738	A	0.125	0.066	4.953	0.026
rs17465637	C	0.671	0.781	8.150	0.004
rs1801133	A	0.295	0.474	20.269	6.729×10^{-6}
rs2023938	G	0.069	0.117	4.326	0.038
rs2075650	G	0.168	0.097	5.650	0.017
rs2252641	G	0.340	0.490	13.446	2.456×10^{-4}
rs2259816	A	0.360	0.449	4.760	0.029
rs2383206	G	0.478	0.571	4.906	0.027
rs2383207	G	0.480	0.571	4.700	0.030
rs3184504	A	0.419	0.531	7.114	0.008
rs4773144	G	0.429	0.515	4.285	0.038
rs4775065	A	0.194	0.281	6.135	0.013
rs4845625	A	0.522	0.434	4.383	0.036
rs4977574	G	0.457	0.561	6.200	0.013
rs6903956	A	0.306	0.388	4.322	0.038
rs780094	A	0.383	0.510	9.333	0.002

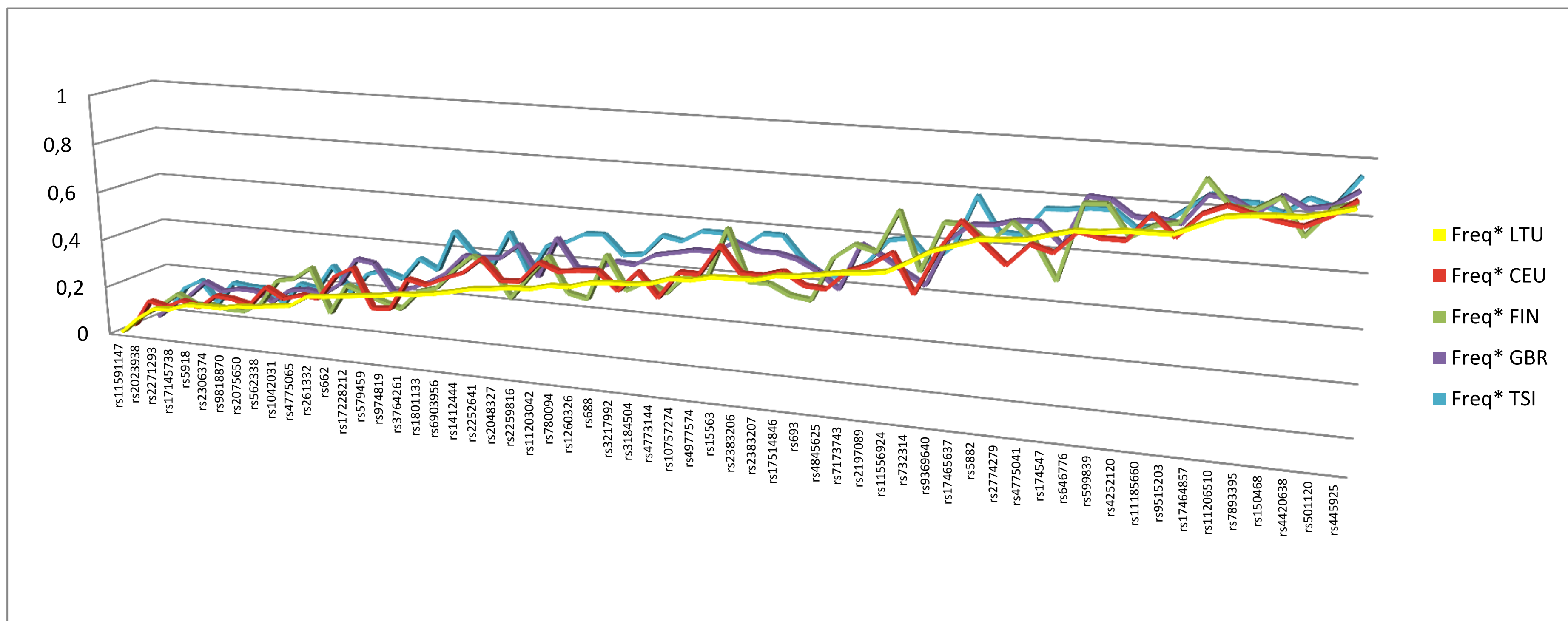

Figure 2.5. The frequency distribution of risk alleles in the Lithuanian population as compared with other populations.* – frequencies of risk alleles.

After comparing the frequencies of risk alleles in the Lithuanian population with those of other populations, the differences were found mainly between the Lithuanian and the Toscani populations (20 differences), and the least differences were found between the Lithuanian and the British populations (6 differences). The counts of identified differences in the Lithuanian population as compared with other populations are provided in **Figure 2.6**.

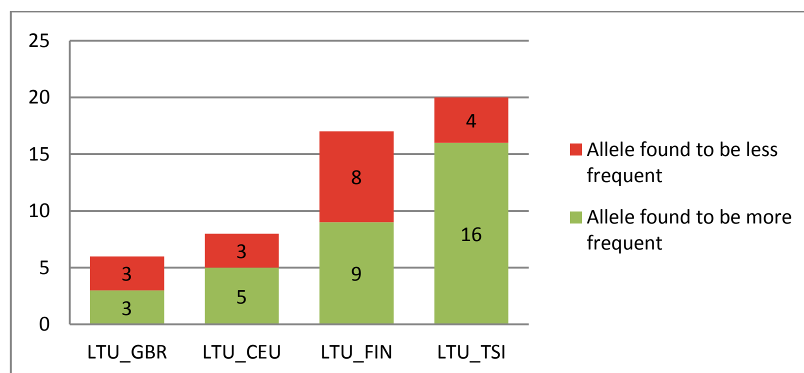

Figure 2.6. The number and type of statistically significant differences of allele frequencies revealed by comparing the Lithuanian population with other populations.

All in all, there were 51 differences in allele frequencies, and 19.61% of these differences overlapped, i.e. the difference of the same allele frequency in the Lithuanian population was found in comparison with more than one population (**Table 2.6**). The types of frequency difference were in some cases rarer and in some cases more common than those compared with. Out of 51 frequency difference cases, in 18 cases (31.29%) risk alleles were more common and in 33 cases (64.71%) rarer in the Lithuanian population than in other populations. The allele rs11203042 in *LIPA* gene was rarer in the Lithuanian population as compared with CEU, FIN, GBR, and TSI populations, and the allele of the rs174547 SNP in the *FADS1* gene was found to be more common in the Lithuanian population than in CEU, GBR, and FIN. This analysis again illustrates the genetic diversity and uniqueness of each population.

Table 2.6. Ten statistically significant differences of risk alleles found in more than one population of those that Lithuanian population was compared with. Statistically significant values are shown in red.

Gene (-s)	SNP	RA	LTU *	CEU *	*p*	FIN *	*p*	GBR *	*p*	TSI*	*p*
LIPA	rs11203042	A	**0.36**	0.45	0.05	0.45	0.03	0.51	7.43×10^{-4}	0.47	0.01
GCKR	rs1260326	A	**0.39**	0.43	0.31	0.30	0.03	0.39	0.85	0.52	1.76×10^{-3}
UBE2Z	rs15563	G	**0.48**	0.58	0.02	0.62	5.83×10^{-4}	0.55	0.09	0.51	0.42
FADS1	rs174547	A	**0.74**	0.65	0.04	0.53	3.86×10^{-7}	0.64	0.02	0.76	0.59
HDAC9	rs2023938	G	**0.07**	0.12	0.03	0.09	0.45	0.07	0.94	0.12	0.04
SLC22A3/ LPAL2/LPA	rs2048327	G	**0.35**	0.36	0.79	0.27	0.05	0.47	0.01	0.33	0.59
ZEB2	rs2252641	G	**0.34**	0.44	0.02	0.39	0.25	0.40	0.12	0.49	2.46×10^{-4}
LIPC	rs4775065	A	**0.19**	0.22	0.51	0.31	1×10^{-3}	0.18	0.67	0.28	0.01
IL6R	rs4845625	A	**0.52**	0.45	0.09	0.54	0.62	0.40	0.01	0.43	0.04
SELP	rs732314	A	**0.60**	0.46	2×10^{-3}	0.52	0.08	0.45	6.59×10^{-4}	0.52	0.05

As regards the number of allele frequency differences identified in the Lithuanian population in comparison with other populations, we can find a geographical gradation. The major number of identified differences (LTU *vs* TSI, LTU *vs* FIN) falls into the North–South axis (vertical) and the minor (LTU *vs* GBR, LTU *vs* CEU) into the West axis (horizontal). This pattern corresponds with the geographical latitude. Referring to the analysis of the genetic history of European populations by the genetic component analysis performed by L. L. Cavalli-Sforza and co-authors, the gradient from the North to the South was found. The difference in the climatic and ecologic conditions depending on the geographical latitude could be the best explanation of this finding (*Cavalli-Sforza LL et al., 1994*). In another genetic study of European populations, including Lithuanians (*Nelis M et al., 2009*), a correlation between the genetic structure of European populations and their geographical position was shown. The genetic diversity is situated gradiently from the European northwest to the southeast. The resulting genetic map forms a triangular structure with a) Finland, b) the Baltic region, Poland, and Western Russia, and c) Italy as its vertexes, and with d) Central and Western Europe in its centre (*Nelis M et al., 2009*). Even though our results also support this positioning, it must be stressed that in our study the Lithuanian population differed from the other populations by various combinations of allele frequencies. This makes the genetic variation in the Lithuanian population both similar and different

from that in the populations of European ancestry. Also if the genetic variation corresponds to a geographical position of a population, then there exists a particular combination of genetic variation, which could define the coordinates of the population location more accurately. It is likely that the risk alleles predisposing to multifactorial diseases are general for human populations. What makes the difference among the populations is the size of the effect and the combinations of allele frequencies (*Visscher PM et al., 2012 (1)*). Our study strongly supports this statement. Allelic heterogeneity emphasizes the importance of the studies of different populations, especially when the analysis includes multifactorial traits such as lipid concentrations (*Coram MA et al., 2013*).

2.3. Identification of new genetic loci. Association analysis

2.3.1. Family GWAS

After the genotyping quality control out of selected 32 families, one CHD patient (the only woman) was excluded, and there were 646,445 SNPs left for the analysis (out of 731,412 SNPs). Only the autosomal markers were analysed.

After the family GWAS analysis there were 9 statistically significantly associated SNPs found (**Table 2.7**). The Manhattan plot of associated SNPs could be found in **Figure 2.7**. There are several SNPs (rs12734338, rs3883013, rs3853444) that do not show a correlation with the neighbouring SNPs and are quite distant from the other SNPs. These are the artefacts, and they were not included in the further analysis. The power test values were more than 0.65 and even close to the desired ones between 0.8 and 0.95 (*Whitley E, Ball J, 2002*). The most promising OR values were of two SNPs: rs17046570 in the *RTN4* gene and rs11743737 in the *FBXL17* gene.

Table 2.7. Statistically significant SNPs associated with CHD

Chr	Gene	Function	SNP	Al.	χ^2	*p*	OR (95% CI)	Power*
2p16.1	***RTN4***	***Neuroendocrinous secretion; apoptosis processes***	***rs17046570***	***A***	***15.21***	9.62×10^{-5}	***18 (2.4–134.8)***	***0.74***
3p23	–	–	rs294314	A	15.38	8.77×10^{-5}	0.13 (0.04–0.43)	0.65
5q31.1	*LOC340073; LOC100996485*	Uncharacterized	rs1346440	G	18.24	1.946×10^{-5}	0.12 (0.03–0.38)	0.82
5q31.1	*LOC340073; LOC100996485*	Uncharacterized	rs2019973	G	16.33	5.312×10^{-5}	0.13 (0.04–0.42)	0.68
5q21.3	***FBXL17***	***SCF complex; protein ubiquitination***	***rs11743737***	***A***	***15.38***	8.77×10^{-5}	***7.67 (2.30–25.53)***	***0.65***
9q22.23	–	–	rs10819695	G	15.70	7.439×10^{-5}	0.10 (0.02–0.41)	0.69
12p11.23	*ITPR2*	Intracellular Ca^{2+} relaxation	rs11048567	A	17.19	3.38×10^{-5}	0.05 (0.01–0.37)	0.81
20q11.21	–	–	rs6141273	A	16.20	5.70×10^{-5}	0.05 (0.01–0.39)	0.77
20p12	–	–	rs1321936	G	16.03	6.23×10^{-5}	0.18 (0.07–0.46)	0.64

Chr – chromosome, Al. – transmitted allele, * – empirical power; green cells denote genes selected for the further candidate gene analysis.

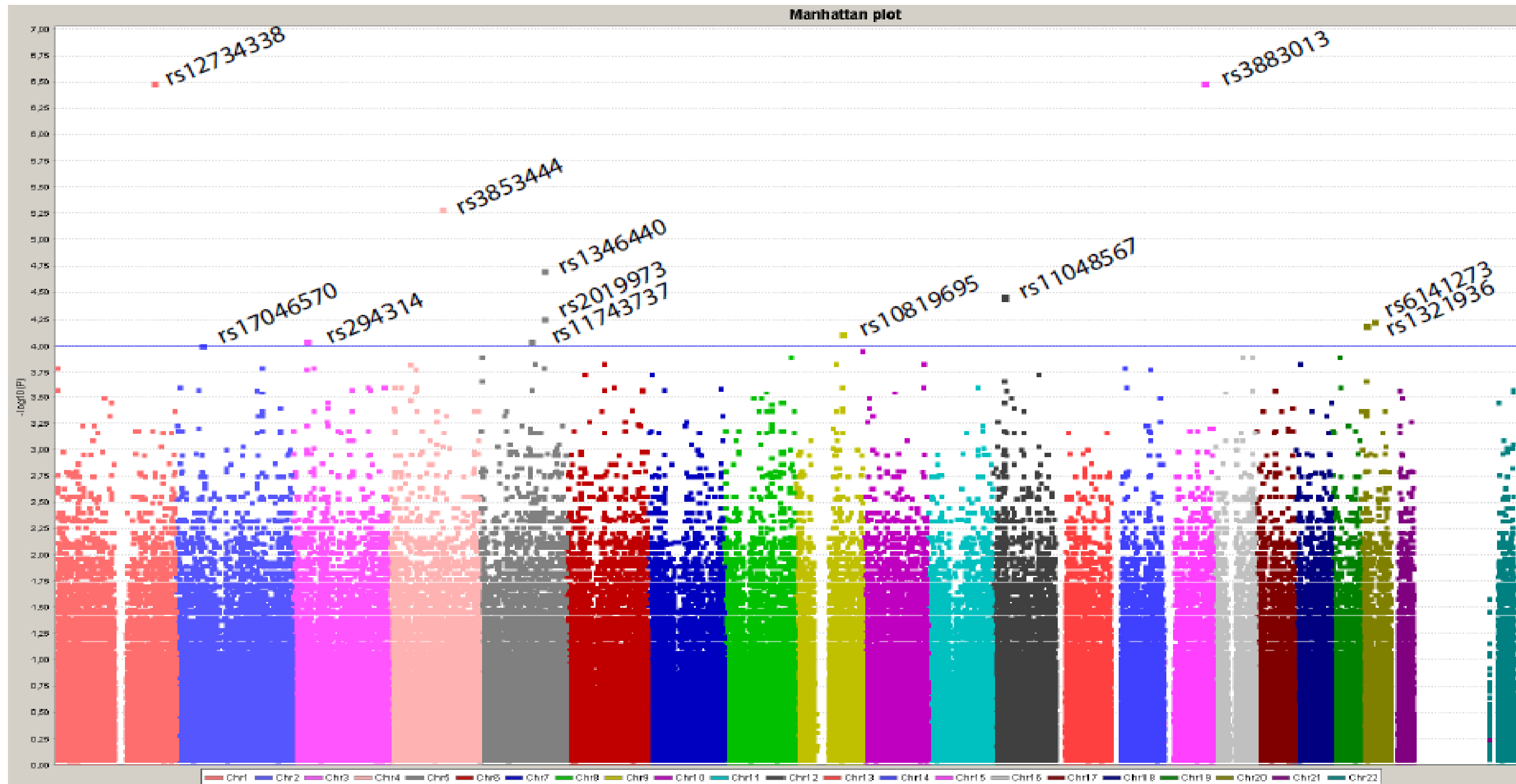

Figure 2.7. The Manhattan plot of SNPs associated with CHD. The horizontal line depicts the significance level ($\alpha = 0.0001$). The X-axis represents the positions of SNPs according to each chromosome, and the Y-axis shows $-\log10(p)$ values of the SNPs.

Additionally, the case-control association analysis was performed in order to check the validity of previous association results. There were 29 CHD patients compared for the eight allele frequencies with 253 individuals from the Lithuanian population. The results are shown in **Table 2.8**. There were four associations confirmed (SNPs rs1346440 and rs2019973 in the locus 5q31.1, rs11048567 in the locus 12p11.23, and rs6141273 in the locus 20q11.21). Associations of these SNPs are new, not reported elsewhere. Out of the three loci confirmed as associated, one locus (20q11.21) had no genes near it. Other associated SNPs are in the genes: rs11048567 – *ITPR2*; rs1346440 and rs2019973– *LOC340073*; *LOC100996485*.

Table 2.8. The additional comparison of the allele frequencies of associated SNPs (from the family GWAS) in the case-control group

Chr	Gene	SNP	Al.	LTU	Cases	χ^2	df1. *p*	OR (95 % CI)
2p16.1	*RTN4*	rs17046570	A	0.39	0.43	0.46	0.50	0.83 (0.46–1.50)
3p23	-	rs294314	A	0.47	0.41	0.77	0.38	1.28 (0.71–2.32)
5q21.3	*FBXL17*	rs11743737	A	0.45	0.54	1.41	0.24	0.72 (0.40–1.29)
5q31.1	*LOC340073; LOC100996485*	rs1346440	G	0.54	0.29	13.06	3.01×10^{-4}	2.87 (1.54–5.53)
		rs2019973	G	0.51	0.26	13.57	2.30×10^{-4}	3.02 (1.60–6.02)
9q22.23	-	rs10819695	G	*Not found in the Illumina Infinium® HD HumanOmniExpress-12 v1.1 SNP assay*				
12p11.23	*ITPR2*	rs11048567	A	0.23	0.05	10.01	1.56×10^{-3}	5.50 (1.73–28.02)
20p12	-	rs1321936	G	0.39	0.29	2.17	0.14	1.56 (0.84–3.01)
20q11.21	-	rs6141273	A	0.12	0.02	5.81	0.02	7.94 (1.32–324.51)

Chr – chromosome, Al. – transmitted allele, LTU – allele frequency determined in the Lithuanian population (controls), Cases – allele frequency determined in the patient group. Statistically significant differences are shown in red.

LOC340073 and *LOC100996485* are non-coding RNAs that can have an impact on the regulation of biological pathways, i.e. can act indirectly through the regulation of gene expression. It is too early for specific speculations about the role of RNA genes in the CHD pathogenesis as *LOC340073* and *LOC100996485* genes are not fully defined, and the information in the databases is obscure.

The *ITPR2* gene codes the inositol 1,4,5-trisphosphate receptor type 2. In the cell, the ITPR2 functions as a transmembrane carrier of Ca^{2+}, as an inositol 1,4,5-trisphosphate dependent calcium relaxation channel and interacts with phosphatidil inositol (*The Gene Ontology Database*). There is an association found between *ITPR2* and inflammation processes as well as the endothelium function. It is considered that *ITPR2* could be involved in the processes of cardiovascular diseases (*Wilker EH et al., 2009*). Recently, it has been shown that ITPR2 takes part in the response to oxidative stress and regulation of the cell cycle (*Warnatz HJ et al., 2011*), and it is already known that these processes are associated with atherosclerosis and CHD.

The other four associations were not replicated (3p23, 20p12, 2p16.1, 5q21.3). Despite this fact, after the data mining, we have considered *RTN4* and *FBXL17* (loci 2p16.1 and 5q21.3, respectively) as the genes that potentially could have a role in the atherosclerosis process and CHD predisposition. The *RTN4* gene encodes the reticulon 4 protein. Reticulons are associated with the endoplasmic reticulum and involved in neuroendocrine secretion or in membrane trafficking in neuroendocrine cells. RTN4 is a potential inhibitor of neurone growth which could suppress the regeneration of the central nervous system in the higher vertebrates (*NCBI database, Chen Y et al., 2010*). RTN4 is a candidate mediator of the vascular cell apoptosis and atherosclerosis process (*Chen Y et al., 2010*).

The *FBXL17* gene encodes the F-box and the leucine-rich repeat protein 17. The F-box is a sequence motif of 40 amino acids. F-box proteins, together with proteins SKP1 (S-phase kinase-associated protein 1) and CUL1 (culin1), form the SCF complex which is protein ubiquitin ligase (*Jin J et al., 2004*). The complex is one of those most important in the ubiquitin-proteosome system (UPS), which participates in about 70–90% of the degradation processes of the cell. Among degraded proteins, there are important cardiovascular system proteins. Moreover, UPS controls the cell cycle of endothelium cells. The effect of oxidative stress on the SCF complex affects

the UPS and at the same time the proper functioning of the endothelium, which is so important for the state of vessels (*Depre C et al., 2010*).

So, these two genes (*RTN4* and *FBXL17*) become the object of the association analysis under hypothesis – the potential candidate gene analysis. There were not enough scientific proof for the hypothesis formulation of two other loci that were not confirmed repeatedly (3p23 and 20p12). Thus, we have proceeded with the validation analysis of the *RTN4* and *FBXL17* by the candidate genes approach.

2.3.2. Analysis of the potential candidate genes *RTN4* and *FBXL17*

The additional case-control study was performed by Sanger sequencing of the *RTN4* and *FBXL17* coding regions in the previous CHD group of probands (31 men and 1 woman), additional 61 CHD patients (55 men and 6 women) and 48 control individuals from the Lithuanian population (randomly selected 24 men and 24 women). The first step was to perform the pilot sequencing for the search of variants in the probands' group. Subsequently, targeted sequencing was performed in the groups of patients and controls. The first exon of both genes was not analysed for technical reasons.

The identified variants of the *RTN4* and *FBXL17* genes are presented in **Table 2.9**. There were two variants found in the *FBXL17* and nine in the *RTN4*. After sequencing, the genotype and allele frequencies of the variants were determined. All variants were in the Hardy–Weinberg equilibrium.

The association analysis of *RTN4* variants showed no significant differences in genotype frequencies among the cases and controls, whereas a statistically significant association was found with the rs34990078 and rs61749621 of *FBXL17*. The OR and RR were also significant for the SNPs of *FBXL17*.

Table 2.9. Variants identified in the *RTN4* and *FBXL17* genes and results of the association study

Gene	SNP	Location of SNP	MAF_ patients	MAF_ controls	χ^2	*p*	OR (95 % CI)	*p*	RR	*p*
RTN4*	rs13002841	2 intron	0.13	0.22	3.93	0.14	0.47 CI (0.22–1.00)	0.05	0.6 CI (0.36–0.99)	0.05
	rs11677099	3 exon	0.04	0.01	2.25	0.13	4.42 CI (0.54–36.46)	0.17	4.13 CI (0.53–32.06)	0.18
	rs80121116	3 exon	0.04	0	3.80	0.05	8.41 CI (0.47–150.47)	0.15	7.82 CI (0.46–134.08)	0.16
	rs13401937	3 intron	0.08	0.03	2.78	0.10	2.89 CI (0.79–10.51)	0.11	2.58 CI (0.79–8.48)	0.12
	rs2580770	3 intron	0.23	0.28	1.11	0.58	0.69 CI (0.34–1.39)	0.29	0.81 CI (0.55–1.19)	0.28
	rs2580769	4 intron	0.26	0.31	1.82	0.40	0.64 CI (0.32–1.23)	0.21	0.79 CI (0.56–1.13)	0.2
	rs2255026	6 intron	0.23	0.29	1.73	0.42	0.63 CI (0.31–1.26)	0.2	0.77 CI (0.53–1.13)	0.19
	rs201547975	6 intron	*Found in only one patient, not chosen for further analysis*							
	rs2255112	7 intron	0.35	0.36	1.14	0.57	1.13 CI (0.56–2.3)	0.73	1.05 CI (0.79–1.4)	0.74
FBXL17*	rs34990078	9 exon	0.2	0.11	4.38	0.04	2.31 CI (1.04–5.09)	0.04	1.77 CI (1.0–3.15)	0.05
	rs61749621	3' UTR	0.2	0.11	4.37	0.04	2.31 CI (1.04–5.09)	0.04	1.77 CI (1.0–3.15)	0.05

* – 93 CHD patients and 48 control individuals analysed; ** – 91 CHD patients and 48 control individuals analysed. Statistically significant differences are shown in red.

For a better understanding of how the *FBXL17* gene and the FBXL17 protein could be involved in CHD and atherosclerosis process, we have performed a bioinformatic analysis. The functional partners of FBXL17 were determined by using the *STRING v9.1* software (*Franceschini A et al., 2013*), and the network of proteins is shown in **Figure 2.8**. According to the Gene Ontology database, the FBXL17 molecular function is binding to other proteins. As mentioned above, FBXL17, together with CUL1 and SKP1, forms the SCF complex which is important in the ubiquitination of proteins. The other FBXL17 functional partners are the BTBD17 (BTB (POZ) domain containing 17), FAM179A (family with sequence similarity 179, member A, 179), KIAA2013 and the KITLG (KIT ligand). The KITLG stimulates the proliferation of mast cells and is able to augment the proliferation of both myeloid and lymphoid hematopoietic progenitors in the bone marrow culture. The KITLG mediates also the cell–cell adhesion and acts synergistically with other cytokines, probably with interleukins.

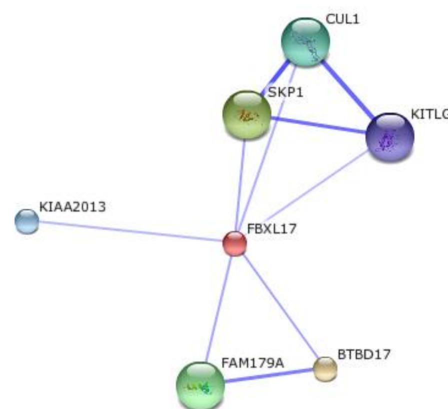

Figure 2.8. Functional partners of FBXL17. Proteins are shown as circles, connecting blue lines denote the association strength among proteins (the thicker the line the stronger the association).

There is few information about FBXL17; the gene was mapped and the protein was defined in 2004 (*Jin J et al., 2004*). The recent finding is that FBXL17 controls the transcription of the NRF2 (nuclear factor, erythroid 2-like 2) target via the turnover of the transcriptional repressor BACH1 (BTB and CNC homology 1, basic leucine zipper transcription factor 1) (*Tan MK et al., 2013*). The *NRF2* encoded transcription factor regulates the genes which contain antioxidant response elements (ARE) in their promoters; many of these genes encode proteins involved in response to injury and inflammation which includes the production of free radicals (*NCBI database*

information). BACH1 regulates the previously mentioned *ITPR2* gene expression and affects the oxidative stress response, cell cycle, and apoptosis (*Warnatz HJ et al., 2011*). In mice, Bach 1 is involved in the cardiovascular system disease processes. The disruption of the Bach1 gene in the Apo E knocked-out mice that were fed a high-fat diet, which caused an inhibition of atherosclerosis through the up regulation of HO-1 (heme oxygenase 1). Moreover, the treatment of double knocked-out mice with stannum protoporphyrin, an inhibitor of HO, abolished the suppressive effects of Bach1 ablation on plaque formation (*Watari Y et al., 2008*). These results illustrate the importance of FBXL17-BACH1-NRF2 and proteins of oxidative stress in both atherosclerosis and CHD processes (*Igarashi K, Watanabe-Matsui M, 2014*). The further investigation is crucial for the specification of a particular biological pathway of FBXL17-BACH1-NRF2.

During the *FBXL17* sequencing and variant association with CHD analysis there were two variants, rs34990078 and rs61749621, of the same genotype frequency identified. Those variants are known to be in linkage disequilibrium D' = 0.96 (1000 Genomes project first phase results of the CEU population analysis). There were no homozygous individuals for both minor alleles of SNPs found. The SNP rs34990078 or c.2432G>A is in the 9th exon of *FBXL17* and is classified as a synonymous variant (p.Thr675=). The other SNP rs61749621 or c.1*G>A of *FBXL17* is in the 3'UTR and is classified as a 3'UTR variant. Referring to the Ensembl database, both variants are in the regulatory domain of the protein and thus could not be rejected as not important for FBXL17 functionality.

Despite the fact that none of the identified *RTN4* SNPs were associated, *RTN4* could not be excluded as a potential candidate gene, because the first exons of both genes (*RTN4* and *FBXL17*) were not examined. Moreover, the sample size could influence the results, as it is known that the greater the number of samples, the more accurate the analysis is. All in all, the *FBXL17* could be recognised as a potential candidate gene participating in the pathogenesis of atherosclerosis and CHD.

CONCLUSIONS

1. Referring to the coronary heart disease as an instance, the genetic multifactorial disease architecture research strategy was modelled. This strategy could be applied in numerous other multifactorial disease studies as well.
2. A list of 60 SNPs associated with coronary heart disease was compiled.
3. According to the selected 60 SNPs, the Lithuanian population genetic structure and diversity was revealed:
    a. An intra-population allele frequency study showed no differences between Lithuanian men and women, whereas there were two differences found between Aukštaičiai and Žemaičiai;
    b. The analysis of haplotype blocks demonstrated the existing differences in men and women groups as well as in Aukštaičiai and Žemaičiai groups: there were different haplotype frequencies and linkage disequilibrium determined, and the haplotype block in the 19th chromosome was found only in the women's group;
    c. Regarding the number of risk alleles per person, the Lithuanian population could be divided into the three groups: 39–50 alleles (1st group), 51–62 alleles (2nd group), 63–74 alleles (3rd group). Most of individuals (68%) fall into the 2nd group and least (13%) into the 3rd one. Therefore, most of Lithuanian population individuals according to the number of risk alleles per person are at a relatively average coronary heart disease risk.
    d. Referring to the genetic profile of 60 risk alleles in the Lithuanian population as compared to other European ancestry populations (CEU, FIN, GBR, TSI), most of allele frequency differences were found on the North-South geographical axis (LTU–FIN 17 and LTU–TSI 20 differences), and least differences were found with Western populations (LTU–CEU 8 and LTU–GBR 6 differences). This illustrates the influence of geographical latitude on the genetic diversity and also the uniqueness of each population in the context of multifactorial diseases.
4. A genome-wide association study revealed new loci associated with the coronary heart disease:

a. After the pilot family GWAS, we have identified eight new loci associated with the coronary heart disease. Subsequently three of them were validated in the case-control analysis: 5q31.1 (rs1346440, rs2019973 in the RNA genes *LOC340073* and *LOC10099648*), 12p11.23 (rs11048567 in the gene *ITPR2*) and 20q11.21 (rs6141273).
b. After further hypothesis driven candidate genes association analysis for two more loci (*RTN4* and *FBXL17*), *RTN4* was not validated but could not be excluded as a potential candidate gene for the coronary heart disease, whereas the *FBXL17* gene variants were found to be associated with the coronary heart disease phenotype and thus could be confirmed as a potential candidate gene participating in the pathogenesis of atherosclerosis and coronary heart disease.

REFERENCES

1. Altman DG. Practical Statistics for Medical Research. Chapman and Hall, 1991, London, UK.
2. Altmann A, Tian L, Henderson VW, Greicius MD. Alzheimer's disease neuroimaging initiative investigators. Sex modifies the APOE-related risk of developing Alzheimer’s disease. Ann Neurol. 2014 Mar 13. doi: 10.1002/ana.24135. (Epub ahead of print)
3. Arvind P, Nair J, Jambunathan S, Kakkar VV, Shanker J. CELSR2-PSRC1-SORT1 gene expression and association with coronary artery disease and plasma lipid levels in an Asian Indian cohort. J Cardiol. 2014 Mar 24. pii: S0914-5087(14)00067-7. doi:10.1016/j.jjcc.2014.02.012. (Epub ahead of print)
4. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15; 21(2): 263–5.
5. Bekris LM, Lutz F, Yu CE. Functional analysis of APOE locus genetic variation implicates regional enhancers in the regulation of both TOMM40 and APOE. J Hum Genet. 2012 Jan; 57(1): 18–25.
6. Cavalli-Sforza LL, Menozzi P, Piazza A. The History ang Geography of Human Genes. Princeton University Press, 1994, Princeton, US.
7. Chen Y, Zhao S, Xiang R. RTN3 and RTN4: Candidate modulators in vascular cell apoptosis and atherosclerosis. J Cell Biochem. 2010 Nov 1; 111(4): 797–800.
8. Coop G, Przeworski M. An evolutionary view of human recombination. Nat Rev Genet. 2007 Jan; 8(1): 23–34.
9. Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW et al. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. Am J Hum Genet. 2013 Jun 6; 92(6): 904–16.
10. Correia M, Branco CC, Bruffaerts N, Balagué A, Mota-Vieira L. Genetic risk assessment for cardiovascular disease in Azoreans (Portugal): a general population-based study. Gene. 2013 Dec 10; 532(1): 132–9.

11. Deloukas P, The CARDIoGRAMplusC4D Consortium et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2013 Jan; 45(1): 25–33.
12. Depre C, Powell SR, Wang X. The role of the ubiquitin-proteasome pathway in cardiovascular disease. Cardiovasc Res. 85(2) (2010): 251–2.
13. DNA extraction technical note: MagneSil® Genomic, Large Volume System, Instructions for Use of Products A4080, A4082, A4085 Technical Bulletin.
14. Drenos F, Whittaker JC, Humphries SE. The use of meta-analysis risk estimates for candidate genes in combination to predict coronary heart disease risk. Ann Hum Genet. 2007 Sep; 71(Pt 5): 611–9.
15. Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA et al. Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. PLoS Genet. 2011 Jun; 7(6): e1002138.
16. Folkersen L, van't Hooft F, Chernogubova E, Agardh HE, Hansson GK et al. Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. Circ Cardiovasc Genet. 2010; 3: 365–73.
17. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan; 41 (Database issue): D808–15.
18. Germanas Š, Jakaitiene A. Power approximation of the transmission disequilibrium test. Lithuanian Mathemat J. 2012; 53 (Series B): 56–61.
19. Girdenis A, Zinkevičius Z. On the Lithuanian dialect classification. Kalbotyra. 1966; 14: 139–47 (in Lithuanian).
20. Humphries SE, Drenos F, Ken-Dror G, Talmud PJ. Coronary heart disease risk prediction in the era of genome-wide association studies: current status and what the future holds. Circulation. 2010 May 25; 121(20): 2235–48.
21. Igarashi K, Watanabe-Matsui M. Wearing red for signaling: the heme-bach axis in heme metabolism, oxidative stress response and iron immunology. Tohoku J Exp Med. 2014; 232(4): 229–53.

22. Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic analysis and nomenclature of mammalian F-box proteins. Genes Dev. 2004 Nov 1; 18(21): 2573–80.
23. Kučinskas V. Genomo įvairovė: Lietuviai Europoje. Spalvų šalis, 2004, Vilnius.
24. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8; 461(7265): 747–53.
25. Maouche S, Schunkert H. Strategies beyond genome-wide association studies for atherosclerosis. Arterioscler Thromb Vasc Biol. 2012 Feb; 32(2): 170–81.
26. Morkūnas K. Lithuanian Language Atlas I. Mokslas, 1977, Vilnius, Map No.1, http://www.tarmes.lt/images/tarmiu_skirstymas/zem_1_nauja%20klasifikacija.gif (in Lithuanian)
27. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, et al. Genetic Structure of Europeans: A View from the North–East. PLoS ONE. 2009; 4(5): e5472.
28. Peloso GM, Demissie S, Collins D, Mirel DB, Gabriel SB et al. Common genetic variation in multiple metabolic pathways influences susceptibility to low HDL-cholesterol and coronary heart disease. J Lipid Res. 2010 Dec; 51(12): 3524–32.
29. Preuss M, König IR, Thompson JR, Erdmann J, Absher D et al.; CARDIoGRAM Consortium. Design of the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study: A Genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls. Circ Cardiovasc Genet. 2010 Oct; 3(5): 475–83.
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep; 81(3): 559–75.
31. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
32. Roberts R, Stewart AFR. 9p21 and the genetic revolution for coronary artery disease. Clin Chem. 2012 Jan; 58(1): 104–12.
33. Rodriguez S, Gaunt TR, Day INM. Hardy–Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. Am J Epidemiol. 2009 Feb 15; 169(4): 505–14.

34. Srivastava RA, Bhasin N, Srivastava N. Apolipoprotein E gene expression in various tissues of mouse and regulation by estrogen. Biochem Mol Biol Int. 1996 Feb; 38(1): 91–101.
35. Tam CH, Ma RC, So WY, Wang Y, Lam VK et al. Interaction effect of genetic polymorphisms in glucokinase (GCK) and glucokinase regulatory protein (GCKR) on metabolic traits in healthy Chinese adults and adolescents. Diabetes. 2009 Mar; 58(3): 765–9.
36. Tan MK, Lim HJ, Bennett EJ, Shi Y, Harper JW. Parallel SCF adaptor capture proteomics reveals a role for SCFFBXL17 in NRF2 activation via BACH1 repressor turnover. Mol Cell. 2013 Oct 10; 52(1): 9–24.
37. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005 Oct 27; 437(7063): 1299–320.
38. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012 Jan 13; 90(1): 7–24. (1)
39. Warnatz HJ, Schmidt D, Manke T, Piccini I, Sultan M et al. The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. J Biol Chem. 2011 Jul 1; 286(26): 23521–32.
40. Watari Y, Yamamoto Y, Brydun A, Ishida T, Mito S et al. Ablation of the bach1 gene leads to the suppression of atherosclerosis in bach1 and apolipoprotein E double knockout mice. Hypertens Res. 2008 Apr; 31(4): 783–92.
41. Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care. 2002; 6(4): 335–341.
42. Wilker EH, Alexeeff SE, Poon A, Litonjua AA, Sparrow D et al. Candidate genes for respiratory disease associated with markers of inflammation and endothelial dysfunction in elderly men. Atherosclerosis. 2009 Oct; 206(2): 480–5.
43. Zhang K, Sun F, Waterman MS, Chen T. Haplotype block partition with limited resources and applications to human chromosome 21 haplotype data. Am J Hum Genet. 2003 Jul; 73(1): 63–73.

SANTRAUKA

Genetinės architektūros mįslė yra viena iš svarbiausių genetikoje, kurią pasaulio mokslas bando įminti. Siekiant papildyti žinias apie genetinę požymių architektūrą, neišvengiamas yra ir jau žinomų genetinių vienetų tyrimas tiek bendroje žmonių populiacijoje, tiek ir etninėse populiacijose. Šiam tikslui pasiekti, šio tyrimo metu, pasirinkta strategija, darbo modelis, kuris gali būti taikomas daugumai kitų daugiaveiksnių ligų tirti. Modelio atraminiai taškai yra žinomų genetinių sričių siejamų su liga analizė ir naujų genetinių sričių siejamų su liga paieška Lietuvos populiacijoje.

Pasirinkta daugiaveiksnė liga – koronarinė širdies liga (KŠL), atrinktos žinomos genetinės sritys – 60 VNP, lemiančių riziką susirgti KŠL. Tiriamiems asmenims nustatyti genotipų, alelių ir haplotipų dažniai, sukonstruoti haplotipų blokai, atliktas alelių dažnių ir haplotipų blokų palyginimas populiacijos viduje (palygintos vyrų ir moterų, aukštaičių ir žemaičių grupės), alelių dažniai palyginti su kitomis populiacijomis, nustatytas rizikos alelių tenkančių vienam asmeniui pasiskirstymas populiacijoje ir tarp vidupopuliacinių grupių. Taip pat atlikta naujų genetinių veiksnių paieškos šeiminė plataus masto genomo asociacijų analizė ir atvejo-kontrolės genetinės asociacijos analizė nustatytų genetinių sričių patvirtinimui.

Pagal tirtus genetinius KŠL rizikos žymenis (VNP), bendra Lietuvos lietuvių populiacija nuo kitų populiacijų, su kuriomis palyginta, statistiškai reikšmingai skyrėsi skirtingais variantais ir jų alelių dažnių deriniais, tuo įrodant, kad skirtingoms populiacijoms yra būdingi tiek bendri, tiek ir unikalūs genetinės architektūros vienetai. Bendros Lietuvos lietuvių populiacijos genetinę struktūrą ir įvairovę pagal tirtus genetinius KŠL rizikos žymenis galima orientuoti gradientiškai Europos Šiaurės-Pietų kryptimi, tai aiškinant geografinių platumų lemiama klimato ir ekologijos įtaka genofondui evoliucijos eigoje. Pagal tirtų genetinių KŠL žymenų rizikos alelių skaičių tenkantį asmeniui, tirtoje bendroje Lietuvos lietuvių populiacijoje, daugiausia asmenų turi vidutinę riziką susirgti koronarine širdies liga. Atliekant naujų KŠL genetinių veiksnių paiešką, nustatytos aštuonios potencialios kandidatinės genetinės sritys, kurios iki šiol dar nebuvo asocijuotos su KŠL, iš jų patvirtintos – keturios, o *ITRP2* ir *FBXL17* genai yra galimi genai kandidatai dalyvaujantys KŠL ir aterosklerozės patogenezėje.

APPROBATION

Publications (2)

1. **Pepalytė I**, Kučinskienė ZA, Grigalionienė K, Petrulionienė Ž, Dženkevičiūtė V, Bagdonaitė L, Kučinskas V. Genetic variants that participate in oxidation processes and/or oxidative stress and are associated with atherosclerosis. European Medical, Health and Pharmaceutical Journal. 2012; 3: 13–16. ISSN 1804-5804.
2. **Domarkienė I**, Pranculis A, Germanas S, Jakaitienė A, Vitkus D, Dženkevičiūtė V, Kučinskienė Z, Kučinskas V. *RTN4* and *FBXL17* Genes are Associated with Coronary Heart Disease in Genome-Wide Association Analysis of Lithuanian Families. Balkan Journal of Medical Genetics. 2013; 16(2): 17–22. doi: 10.2478/bjmg-2013-0026.

Oral presentations (3)

1. **Pepalytė I**, Pranculis A, Dženkevičiūtė V, Bagdonaitė L, Kučinskas V, Kučinskienė ZA. Defining coronary heart disease genomics in the Lithuanian population. 11th Baltic Congress in Laboratory Medicine, 2012, Vilnius, Lithuania.
2. Pranculis A, **Pepalytė I**, Arasimavičius J, Kučinskienė ZA, KučinskasV. Pharmacogenomics of CVD treatment in patients from the Lithuanian population. International conference “Evolutionary medicine: new solutions for old problems”, 2012, Vilnius, Lithuania.
3. **Domarkienė I**, Pranculis A, Germanas Š, Jakaitienė A, Dženkevičiūtė V, Kučinskas V, Kučinskienė ZA. Are we ready to incorporate genes in risk estimation? Lithuanian way towards genomics of CHD. 6th Baltic Atherosclerosis Congress, 2013, Riga, Latvia.

Poster presentations (7)

1. **Domarkienė I**, Pranculis A, Germanas Š, Jakaitienė A, Dženkevičiūtė V, Kučinskas V, Kučinskienė ZA. Lithuanian way towards genomics of coronary heart disease. 6th Baltic Atherosclerosis Congress, 2013, Riga, Latvia.
2. Pranculis A*, **Pepalytė I***, Dženkevičiūtė V, Bagdonaitė L, Kučinskas V, Kučinskienė ZA. A family-based genomics and pharmacogenomics study of coronary artery disease in the patient group from the Lithuanian population. 11th Baltic Congress in Laboratory Medicine, 2012, Vilnius, Lithuania.
   *Authors contributed equally.
3. Kučinskas V, **Pepalytė I**, Pranculis A, Bagdonaitė L, Dženkevičiūtė V, Kučinskienė ZA. Coronary artery disease genomics – CNV findings in the group of patients from Lithuania. Conference of ASHG, 2012, San Francisco, USA.
4. Pranculis A, **Pepalytė I**, Arasimavičius J, Kučinskienė ZA, Kučinskas V. Pharmacogenomics of cardiovascular disease treatment in subjects from the Lithuanian population. Conference of ASHG, 2012, San Francisco, USA.
5. **Pepalytė I**, Uktverytė I, Dirsė V, Kučinskas V. Characteristics of the genomic structural variation in the Lithuanian population. International conference “Evolutionary medicine: new solutions for old problems”, 2012, Vilnius, Lithuania.
6. Pranculis A, **Pepalytė I**, Arasimavičius J, Kučinskienė ZA, Kučinskas V. Pharmacogenomics of CVD treatment in patients from the Lithuanian population. International conference “Evolutionary medicine: new solutions for old problems”, 2012, Vilnius, Lithuania.
7. **Pepalytė I**, Grigalionienė K, Petrulionienė Ž, Bagdonaitė L, Kučinskienė ZA, Kučinskas V. Common genetic variants in *ITGA2*, *IL1B*, *ALOX5AP*, *OR13G1*, *MMP9* genes that participate in oxidative stress and influence atherosclerosis phenotype formation. Conference of ESHG, 2011, Amsterdam, Netherlands.

ABOUT THE AUTHOR

Personal information: Ingrida Domarkienė, born March 1st, 1985

Contact information: Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariškių Str. 2, Vilnius, LT-08661

E-mail address: ingrida.domarkiene@mf.vu.lt

Education:

1992–2004 Grigiškės "Šviesa" secondary school (B No. 078252)

2004–2008 Bachelor of Biology (BA No. 0714685), Molecular Biology study program, Faculty of Natural Sciences, Vilnius University

2008–2010 Master of Biology (GM No. 0640084), Genetics study program, Faculty of Natural Sciences, Vilnius University

2010–2014 Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, PhD student

Work experience:

2008–2010 Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, technician;
2010 – Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, junior research fellow; Vilnius University Hospital Santariškių Klinikos Centre for Medical Genetics, medical geneticist.

Membership in professional societies:

Lithuanian Society of Human Genetics
European Society of Human Genetics

APIE AUTORĘ

Asmeniniai duomenys: Ingrida Domarkienė, gimusi 1985 03 01.

Kontaktinė informacija: Vilniaus universiteto Medicinos fakulteto Žmogaus ir medicininės genetikos katedra, Santariškių g. 2, Vilnius, LT-08661

El. pašto adresas: ingrida.domarkiene@mf.vu.lt

Išsilavinimas:

1992 m. – 2004 m. Vilniaus miesto savivaldybės Grigiškių „Šviesos“ vidurinė mokykla (B Nr. 078252)

2004 m. – 2008 m. Vilniaus universitetas, Gamtos mokslų fakultetas, molekulinės biologijos studijų programa, biologijos bakalauro kvalifikacinis laipsnis (BA Nr. 0714685)

2008 m. – 2010 m. Vilniaus universitetas, Gamtos mokslų fakultetas, genetikos studijų programa, biologijos magistro kvalifikacinis laipsnis (GM Nr. 0640084)

2010 m. – 2014 m. Vilniaus universitetas, Medicinos fakultetas, Žmogaus ir medicininė genetikos katedra, doktorantė

Darbinė patirtis

2008 m. – 2010 m. VU MF Žmogaus ir medicininės genetikos katedra, pareigos – laborantė;
Nuo 2010 m. VU MF Žmogaus ir medicininės genetikos katedra, pareigos – jaunesnioji mokslo darbuotoja; VšĮ VUL SK Medicininės genetikos centras, pareigos – medicinos genetikė.

Narystė profesinėse organizacijose:

Lietuvos Žmogaus genetikos draugijos narė
Europos Žmogaus genetikos draugijos narė