SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.)

Characterization of the Hungarian Reference DNA Biobank. The role of ACE I/D polymorphism in susceptibility to metabolic syndrome among Hungarians

by Szilvia Fiatal MD

Supervisor: Prof. Róza Ádány MD, PhD, DSc

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF HEALTH SCIENCES

DEBRECEN, 2011

Introduction

1. Significance of population based genetic data for preventive medicine

In the 21st century, genetic studies based on human genome play a central role in public health by providing genetic information for disease prediction and prevention. Translation of these results into meaningful actions in order to improve health and prevent diseases depends on scientific information from medical and public health studies. Most of the 'gene discoveries' are based on studies of high-risk families or selected population groups. After a candidate gene is found, well-designed epidemiological studies are needed to quantify the population impact of gene variants as well as to identify and measure the impact of modifiable environmental risk factors that interact with gene variants.

Family-based or population-based genetic association studies have been performed to determine association between a given genetic variant and a disease, i.e. whether an individual carrying one or two copies of a high-risk variant is at increased risk of developing a disease. Case-control study is the most frequently a design for association studies that compares the frequency of SNP alleles in two well-defined groups of individuals: a) cases who have been diagnosed with the disease under study in a defined population (should be representative of all cases in the population), b) and controls, who are either known to be unaffected, or who have been randomly selected from the same underlying population. If the controls do not represent the source population that gave raise the cases, selection bias may occur and as a consequence distribution of genetic or environmental factors of interest among case and control subject is not comparable.

In the genetic epidemiology literature most of the studies published have compared genotypes of patients in a clinical case series with those of a non-diseased sample of control subjects and consequently they provide little basis - if any - for risk estimation at population level. From the public health perspective, a population-based estimate of relative risk is important because it provides the basis for estimating population

attributable fraction i.e. the proportion of cases that would not occur in the absence of a particular genotype in the population.

Recognizing the power of a population-based approach to search for complex disease susceptibility genes many countries have established population-based biobanks. Population-based biobanks can be utilized to estimate allele frequencies in the population because they are comprised of large numbers of randomly selected participants being representative of the population by age, sex and geographical distribution.

2. Need for Hungarian Reference DNA Biobank

At population level, attributable-risk is different by alleles, and aside from the co-effects of other alleles and environmental factors, the AR is a function of population allele prevalence. Frequencies of gene variants vary by populations therefore the allele spectrum that is present in a certain population cannot predict directly the risk of a given disease in another population. Thus, in order to estimate the role of disease-specific susceptibility gene variants in the Hungarian population, it is important to know the prevalence of risk alleles in the general population.

Recently, we have performed a structured literature search to give an overview on the occurrence of risk alleles responsible for development of chronic diseases with high public health importance in the Hungarian population. The Hungarian genetic epidemiological studies are typically association studies using case-control design with only some hundreds of cases. The control groups used are mainly sets of non-diseased individuals typically non-randomly selected and consequently not representative for the Hungarian general population. These studies have not assigned control groups required by exact methodology, thus their conclusions can be accepted with severe restrictions. In fact, the allele frequencies calculated on the basis of results obtained in these studies mentioned do not describe the allele-distribution in the general population and their results cannot be applied to define the possible relationship between a given mutation and a disease at population level. It has became obvious that future genetic epidemiological studies require population-based DNA biobank collected by respecting well-defined selection criteria to be representative for the Hungarian general population. It is important

to highlight that Hungarian national population-based DNA biobank has been never established before.

2.1. Recruitment of the samples to the reference DNA biobank

Population-based disease registry such as General Practitioners' Morbidity Sentinel Stations Program (GPMSSP) provides the most accomplished way for generating the reference sample by a cost-effective way. The Hungarian GPMSSP was established in 1998 by the School of Public Health, University of Debrecen and the National Public Health and Medical Officer Service in order to monitor the prevalence and incidence of chronic non-communicable diseases of great public health importance.

Recently, based on the infrastructure of GPMSSP many epidemiological studies have been performed. Development of the reference DNA biobank, being representative for the Hungarian population by age, gender and geographical distribution, took place in several stages within the framework of these epidemiological studies.

The fundamental purpose of DNA sample collection was to create a reference sample suitable for characterizing the frequency of genetic variations predisposing diseases with high public health importance. The individuals of the reference population were randomly chosen from the source-population by the general practicioners (GPs) by using a predetermined algorithm. In the first step of the multi-stage sampling the sample sizes were defined on the basis of age and sex distribution of the participating counties population. In the second step, the practice specific sample sizes were calculated in a given county on the basis of the participating practices' age and sex distribution. In the third step individuals were chosen from the source-population by systematic random sampling.

The general practitioners also have collected data on demographic characteristics, anthropometric parameters and have performed complete physical examinations and collected blood for laboratory investigations. Blood samples were also collected in EDTA anticoagulant tubes for DNA analysis.

2.2. Construction requirements of the biobank

2.2.1. Representativeness

The study cohorts of epidemiological studies used for development of reference DNA biobank were created on the basis of the population size of the counties, and the age and sex distribution of the county in the given year (details see below – Description of biobank). Participating counties were selected in a way to represent both the eastern and the western part of the country. In recruiting family practitioners for the GPMSSP, participating practices were required to be representative of all practices in their home counties in terms of both settlement size and geographical location.

2.2.2. Sufficient Sample Size and Power

Chronic diseases are caused by a mixture of common gene variants with small effects. For common diseases individual contributions of these genes would have odds ratios less than two, even odds ratios of 1.2 to 1.5 can be important to detect and quantify. Thus in designing biobanks, it is essential to ensure that they are capable of detecting the small effect sizes that are plausible under this hypothesis. Therefore, accurate power calculations are needed to determine sample size requirements.

2.3. Handling whole blood samples collected

Blood samples anticoagulated by EDTA were collected and transferred within 3 hours into the county institutions of NPHMOS. The collected blood samples were stored in polypropylene centrifuge tubes in freezer, mostly at -86 °C (at least under -20 °C). Then, for DNA extraction and genetic analysis, blood samples were transferred from the county institution of NPHMOS to the Department of Preventive Medicine, University of Debrecen. DNA extraction was carried out by using of MagNA Pure LC DNA Isolation Kit–Large Volume (Roche Diagnostics GmbH, Mannheim, Germany) according to the

manufacturer's instructions. The quality of extracted DNA was assessed by its yield and purity. After extraction specimens were divided into aliquots to prevent freeze-thaw cycles. Long-term storage is done at -80^{o}C and short term storage is done of 4 ^{o}C.

2.4. Standardizing measurement conditions

Standard implementation of blood pressure measurement, blood tests and physical examination are important parts of a valid (genetic)epidemiological study. Consequently in order to eliminate possible bias general practitioners and theirs assistants had to undergo training where they acquired basic methodological requirements. Clinical parameters were measured in only one (reference) laboratory in each case.

3. Description the biobank

3.1. Population based biobank

One major class of biobanks is designated as "population-based" biobanks in which subjects are sampled from a defined target population with no explicit attempt to over- or under sample subjects based on a current disease status, enrolling large number of individuals. This set of controls is representative of the general population as a whole may therefore be used as a reference or general control series and compared to sets of cases.

3.1.1. Establishment of reference control for genetic epidemiological studies mapping genetic differences in susceptibility to cardiovascular diseases

The source population of the study involved males and females older than 20 years from districts of Győr-Moson-Sopron, Hajdú-Bihar, Szabolcs-Szatmár-Bereg and Zala counties participating in GPMSSP in 2001. Sample size was determined by the population size of the counties, and their distribution by age and gender of the year 1999: 1196 individuals were drawn from source population. This sample size was enough to

evaluate the prevalence of risk factors' genes involved in cardiovascular diseases with high accuracy. DNA preparation could be accomplished technically in 1184 cases: Győr-Moson-Sopron: 127 male, 157 female; Hajdú-Bihar: 171 male, 187 female; Szabolcs-Szatmár-Bereg: 168 male, 187 female; Zala: 86 male, 101 female. Demographical data and body weight, body height, lipid profile of the individuals/ participants are available.

3.1.2. Control group collected during a cross-sectional study of "The occurrence and clinical characteristics of metabolic syndrome among the adult Hungarian population "

In this study the source population involved males and females between the age of 20 and 69 from districts of 8 counties (Baranya, Bács-Kiskun, Győr-Moson-Sopron, Hajdú-Bihar, Heves, Komárom-Esztergom, Szabolcs-Szatmár-Bereg és Zala) participating in GPMSSP in the year of 2005. The planned sample size was 2006 individuals aged 20-69 years. Beside the determination of alleles occurrence involved in metabolic syndrome (according to the criteria of International Diabetes Federation (IDF)) among Hungarian adults between the age of 20 and 69 years , this sample size provided the detection of the indicator parameter, pathological HDL-cholesterol level with 20-25% deviation (previously estimated occurrence was 4.8 %); this parameter has the lowest frequency among the factors according to the criteria of National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP-III) accepted also by the Metabolic Workgroup of Hungarian Diabetes Society. One thousand eight hundred and nineteen persons (91%) from the invited volunteered to participate in the study, DNA preparation could be accomplished technically in 1783 cases: Baranya: 102 male, 120 female; Bács-Kiskun: 149 male, 164 female; Győr-Moson-Sopron: 109 male, 134 female; Hajdú-Bihar: 117 male, 140 female; Heves: 86 male, 95 female; Komárom-Esztergom: 69 male, 80 female; Szabolcs-Szatmár-Bereg: 137 male, 138 female; Zala: 72 male, 71 female. Available data were demographical data, lipid profile, body weight and height, waist circumference, blood pressure, medical treatment/drug administration record, serum levels of creatinin, urea, liver enzymes and glucose, verified presence of hypertension, obesity.

3.2. Disease specific biobank

Other major class of biobanks is designated as “disease specific“, which includes samples and information from a large number of cases of a specific disease. “Disease specific” biobanks may be combined with appropriate set of controls to provide a foundation for powerful case-control studies.

3.2.1. Group established during the study of " Mapping the care of type II diabetes mellitus in GP practice"

This study implied the population of 11 counties’ practices (Baranya, Borsod-Abaúj-Zemplén, Bács-Kiskun, Győr-Moson-Sopron, Hajdú-Bihar, Heves, Komárom-Esztergom, Nógrád, Szabolcs-Szatmár-Bereg, Szolnok, Zala) participating in GPMSSP program as a source population; patients involved were above 35 years, suffered from type 2 diabetes mellitus (based on database of 31st December, 2006), but were not hospitalized or had been not living in community home at time of the study. DNA preparation could be accomplished technically in 1300 cases. Distribution of individuals by age and gender are as follows: Baranya: 88 male, 77 female; Borsod-Abaúj-Zemplén: 35 male, 38 female; Bács-Kiskun: 59 male, 53 female; Győr-Moson-Sopron: 35 male, 42 female; Hajdú-Bihar: 28 male, 30 female; Heves: 50 male, 49 female; Komárom-Esztergom: 27 male, 38 female; Nógrád: 41 male, 47 female; Szabolcs-Szatmár-Bereg: 116 male, 158 female; Szolnok: 41 male, 38 female ; Zala: 110 male, 100 female. Available data were demographical data, lipid profile, body weight and height, waist circumference, blood pressure, HbA1c blood level, serum levels of creatinin, urea blood level, CRP, ALP, liver enzymes, glucose (fasting), verified presence of hypertension, obesity.

4. Benefits and challenges of the Hungarian Reference DNA Biobank

Potential benefits of population-based Biobanks in Hungary can be summarized as follows:

a. Establishing a well-defined biobank according to strict criteria can provide the most acceptable estimation on allele frequencies and genotype prevalence of human genetic variants for the Hungarian population. Population-based allele frequencies and genotype prevalence are necessary for measuring the contribution of genetic variation to human disease susceptibility and for assessing population attributable risk.

b. It can also facilitate future research on gene–disease associations of genetic variants by facilitating estimation of genetic susceptibilities that lead to the development of multifactorial diseases and the quantification of identified major genetic risk and environmental susceptibility factors for common diseases having multifactorial origin; and description of major gene-environment and gene-gene interactions

c. Make this resource available to biomedical research community as a whole. In order to reach this goal recently we plan to join the pan-European network of *Biobanking and Biomolecular Resources Research Infrastructure*.

d. Despite the health and research benefits the biobank – as other repositories - has a number of ethical, legal, and social implications (ELSI). Some of these very important issues include proper informed consent procedures, protection of donor confidentiality, and regular public consultation, etc.

5. Genetic epidemiological studies based on the Hungarian Reference DNA Biobank

The following genetic epidemiological studies have used DNA samples from the Hungarian Reference DNA Biobank as control group until now:

1. Study on the association if insertion/deletion polymorphism of angiotensin-1 converting enzyme and metabolic syndrome in Hungarian adults.
2. Identification of multiple common variants for celiac disease influencing immune gene expression.

Recently, second-generation genome wide association study of 4,533 celiac disease cases and 10,750 controls has been performed in international collaboration. Furthermore in the second follow-up stage 113 selected SNPs with *P*GWAS<10^{-4}, and 18 SNPs from 14 known loci, in 7 independent follow-up cohorts comprising of 4,918 celiac cases and 5,684 controls were genotyped. All individuals were of European ancestry. As a member of this collaborative work we have participated in forming a group of 1067 population controls selected from our DNA biobank. As a result of this study variants from 13 new regions reached genome wide significance (*P* combined<5x10-8), most contain immune function genes (*BACH2, CCR4*, *CD80*,*CIITA*/*SOCS1*/*CLEC16A, ICOSLG*, *ZMIZ1)* with *ETS1, RUNX3, THEMIS* and *TNFRSF14* playing key roles in thymic T cell selection. A further 13 regions had suggestive association evidence.

3. Fine mapping of the CELIAC2 locus on chromosome 5q31-q33 in the Finnish and Hungarian populations.
4. Cost-effective HLA typing with tagging SNPs predicting celiac disease risk haplotypes in the Finnish, Hungarian, and Italian populations.
5. IL23R in the Swedish, Finnish, Hungarian and Italian populations: association with IBD and psoriasis, and linkage to celiac disease.
6. Linkage and association study of FcgammaR polymorphisms in celiac disease.
7. Association study of the IL18RAP locus in three European populations with coeliac disease.
8. The shared CTLA4-ICOS risk locus in celiac disease, IgA deficiency and common variable immunodeficiency.
9. Myosin IXB gene region and gluten intolerance: linkage to celiac disease and a putative dermatitis herpetiformis association.
10. Association of the 8.1 ancestral MHC haplotype with colorectal cancer risk.
11. Modulation of the risk of coronary sclerosis/myocardial infarction by the interaction between factor XIII subunit A Val34Leu polymorphism and fibrinogen concentration in the high risk Hungarian population.

6. Aims of the studies

In our studies we aimed to characterize the Hungarian Reference DNA Biobank representing the Hungarian population by age, gender and geographic distribution and to demonstrate the role of ACE I/D and AGT M235T polymorphisms in susceptibility to metabolic syndrome in the Hungarian population. The questions addressed by our work:

1. What is the frequency of ACE I/D, AGT M235T gene polymorphisms in the Hungarian population?
2. Does remarkable association exist between these polymorphisms and MS?
3. Do any of these polymorphisms correlate with metabolic variables in a representative sample of Hungarians?

Materials and methods

1. Subjects

In the Hungarian population recently a cross sectional study was performed based on GPMSSP to define the prevalence of metabolic syndrome among Hungarians. In the source-population of the study all 20-69 year old Hungarian citizens registered with the 59 general practitioners of the eight participant counties (Baranya, Bács-Kiskun, Hajdú-Bihar, Heves, Győr-Moson-Sopron, Komárom-Esztergom, Szabolcs-Szatmár-Bereg and Zala) were included. The counties were selected to represent both the eastern and the western part of the country and the planned sample size was 2006 people. It makes possible to detect with a 20-25% divergence the frequency of the pathologic HDL-cholesterol level estimated to be the less frequent alteration among MS criteria indicators on the basis of earlier surveys carried out in the Hungarian adult (aged between 20 and 69) population.

The study population representing the Hungarian adult population on the basis of geographic, age and sex distribution was chosen from the source-population by coincidence. The sampling frame of the survey was formed by volunteers aged between 20 and 69 (based on data on 31st December 2005) who don't live in social institutions are) from the eight participating counties' practice. The general practitioners excluded the

sick living in institutions (social services etc.) prior to the sampling. As the first step of the multi-stage sampling the participating counties population's sample sizes on the basis of age and sex distribution were defined. As the second step the practice specific sample sizes in a given county on the basis of the participating practices' age and sex distribution were calculated. As the third step study subjects from the source-population were chosen by concentric systematic random sampling.

2. Data collection

The GPs have collected data on demographic characteristics (age and sex), anthropometric parameters (body weight, height, waist circumference), performed complete physical examinations and collected blood for laboratory testing to define the levels of blood components taken into consideration for the diagnosis of MS. Blood samples were also collected in EDTA anticoagulant tubes for DNA analysis. Standard instruments were provided for GPs to measure waist circumference and blood pressure and they were instructed to follow standard protocols.

MS was defined according to the latest and widely used diagnostic criteria proposed by IDF. The study participants were first divided into two groups as with MS (MS +) and without MS (MS-). The MS + group was classified into subgroups on the basis of the combination of different pathological parameters meeting MS criteria.
The study protocol was approved by the Ethic Committee of the University of Debrecen and informed consent was obtained from all participants.

3. DNA extraction and genotyping

DNA was extracted from peripheral blood samples of the subjects using MagNA Pure LC DNA Isolation Kit–Large Volume (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturer's instructions. PCR and genotyping with melting curve analysis were performed by LightCycler real time PCR system (Roche Diagnostics GmbH, Mannheim, Germany). As reaction buffer in the PCR, the LightCycler DNA Master Hybridization Probes 10x buffer (Roche Molecular

Biochemicals) with a final Mg^{2+} concentration of 3.2 mmol/L (ACE) and 2.4 mmol/L (AGT) were used. PCR was performed in a reaction volume of 20 µL with 0.5 µmol/L of each primer (ACE: 5'CTG GAG ACCACTCCC ATC CTT TCT3' and 5'GAT GTG GCC ATC ACA TTC GTC AGA3' , AGT: 5'CTC TAT CTG GGA GCC TT 3' and 5'GTT TGC CTT ACC TTG GAA3') and 0.2 µmol/L (in case of AGT 0,1µmol/L) of the anchor and detection probes (TIB MolBiol). The detection probe (ACE: 5'CGT GAT ACA GTC ACT TTT ATG—*FL3*' AGT: 5'CCC TGA TGG GAG CCA GTG—*FL3*') was labeled at the 3' end with fluorescein; the anchor probe (ACE: *5'LC Red640*-GGT TTC GCC AAT TTT AT CCA GCT CTG--*PH3',* AGT: 5'*LC Red640*-GAC AGC ACC CTG GCT TTC AAC AC—*PH3'*) was labeled with LightCycler Red 640 at its 5' end and modified at the 3' end by phosphorylation to block extension.

Cycling and melting curve analysis conditions for AGT gene were the following: after initial denaturation at 94 °C for 45s, 55 cycles were performed in a Light Cycler™ using a temperature profile of 94 °C 0 s, 57 °C 15 s and 72 °C 25 s with a temperature transition rate of 20 °C/s. Cycling and melting curve analysis conditions for ACE gene were the following: after initial denaturation at 95 °C for 60s 50 cycles were performed using a temperature profile of 95 °C 3 s, 61 °C 20 s and 72 °C 30 s with a temperature transition rate of 20 °C/s. Fluorescence was measured at the end of each annealing phase. After the amplification, melting curves were generated with heating the samples at 95 °C for 60 s using a ramping rate of 20 °C/s, holding them at 40 °C for 1 min in order to get maximum hybridization, then slowly heating the reaction mixtures up to 85 °C with a temperature transition rate of 0.4 °C/s.

4. Statistical analysis

Normally distributed parameters are presented as means (and ±SD) for different groups and subgroups in both genders and were compared with Student t test. Logarithmic transformation of BMI and reciprocal square root transformation of average systolic blood pressure were used to achieve normal distribution. The overall allele frequencies of the study populations are calculated from the overall genotype distributions according to gene counting method. Deviations from Hardy-Weinberg

proportions were tested with a chi-square goodness-of-fit approach. Statistical difference in genotype distribution and also allele frequencies among study groups/subgroups were assessed by $\chi 2$ test. A multiple logistic regression model was used to estimate the effect of genotypes on the likelihood of MS and its different components. Age and sex do not influence the genetics of an individual, thus matching cases and controls by these properties were not necessary. A value of $p<0.05$ was considered statistically significant. STATA software 9.2 (College Station, Texas) was used to perform statistical analysis.

Results

One thousand eight hundred and nineteen persons (91%) from the invited volunteers to participate in the study; 1762 blood samples were technically suitable for DNA isolation and genotyping. Six hundred and forty-one persons had MS whereas 1121 did not meet the above-mentioned IDF criteria.

The prevalence of MS in the two sex groups was slightly different ($p = 0.02$). Beside central obesity as the essential requirement of the MS diagnosis the most frequent finding among the metabolic components was the raised blood pressure or treatment of previously diagnosed hypertension (87%); 50% (n=320) of the patients had two and 35% (n=224) of the patients had three metabolic abnormalities besides central obesity. All metabolic components were observed in 15% (n=97) of the patients. The most frequent component combination was central obesity, dislipidaemia and hypertension (21%).

The frequency of ACE II, ID and DD genotypes were 19.5% (n= 125), 49.1% (n=315) and 31.4% (n=201) in the group of patients with MS and 23.1% (n=259), 51.5% (n=577) 25.4% (285) in patients without MS, respectively. The frequency of DD genotype (31.36% vs. 25.42%, p=0.006) and the frequency of D allele (0.56 vs. 0.51, p=0.006) were significantly higher in the group of MS patients than in the MS – group. The distribution of AGT M235M, M235T and T235T genotypes in the MS + group were 27.5% (n=176), 48.8% (n=313), 23.7% (n=152) while 25.4% (n=285), 51.4% (n=576), 23.2% (n=260) in the group without MS, respectively. Similar distribution of AGT genotypes was found in the two study groups.

The overall allele frequencies in the sample representative for the Hungarian adult population were 0.51 and 0.49 for AGT allele M235 and allele 235T, 0.47 and 0.53 for ACE allele I and allele D, respectively, and similar to those published in the literature for Caucasians. Genotype frequencies were all in accordance with the Hardy-Weinberg equilibrium (AGT: chi^2=0.39, p=0.53; AGT: chi^2=0.6, p=0.43). The genotype and allele distribution were found not significantly different between males and females.

Logistic regression model revealed association of DD genotype with MS (crude OR=1.46, 95% CI=1.10-1.93, p=0.008). This result indicates that subject with DD genotype compared to II genotype has 46% higher risk for developing MS. Adjustment for the confounding effect of obesity did not modify these values.

No significant relationship with any of the metabolic components and ACE genotypes were found. Among patients having central obesity, elevated TG and low HDL-cholesterol level significant association with ACE DD genotype was shown. The DD genotype subjects in this subgroup have 2.7 times higher risk (OR: II vs. DD; 95% CI: 1.14-6.31, p=0.024) for developing MS compared to those with II genotype. Furthermore, significant association was found when dyslipidaemia was combined with high blood pressure. Both ID (OR=1.7, 95% CI: 1.02 - 2.9, p=0.038) and DD (OR=1.9, 95% CI: 1.1-3.3, p=0.022) genotypes increased the risk of developing MS. Although when central obesity was combined with elevated TG, low HDL-cholesterol level, high blood pressure and glucose intolerance, the odds ratio did not show elevated risk (II vs. ID: OR=1, 95% CI:0,59-1,69 , II vs. DD: OR=1, OR=1, 95% CI:0,57-1,83). The AGT M235T gene polymorphism was not associated with MS as an entity or any of its components.

Discussion

Estimating the risk for complex disorders (without a clear pattern of inheritance), describing the distribution of genetic traits in populations and evaluating the role of genetic factors in disease occurrence requires ‘larger’ studies and genetic information of unrelated people. Information on prevalence of gene variants, genotype-phenotype correlations, and gene-gene and gene-environment interactions can be collected

systematically by epidemiologic studies conducted in populations resembling those to which inferences will be drawn. The Hungarian Reference DNA Biobank is the first relatively large-scale, population-based sample collection in Hungary to obtain such a data. The GMSSP is the only nationally representative, population-based sample survey that systematically collects morbidity data on a large number of individuals in Hungary. The GPMSSP based DNA bank will be an important next step toward identifying genetic variants that can help to estimate disease susceptibility.

We have examined the distribution ACE I/D and AGT M235T gene polymorphisms in a sample representative for Hungarian adult population (n=1762) consisting of 641 patients with MS according to the IDF criteria and 1121 subjects without MS in a cross sectional study.

Many studies have investigated associations between these polymorphisms and separate components of MS as well as with MS as defined on the basis of ATP III criteria. Our results are in accordance with those of some previous studies in other non-Caucasian populations. Alvarez-Aguilar *et al* reported that the DD genotype of ACE gene increases susceptibility to MS in Mexican population. Lee and Tsai have showed that DD genotype of ACE confers an increased risk factor for MS with dyslipidaemia and albuminuria in the Chinese population. In contrast, no significant associations between ACE gene I/D polymorphism and MS were found in Brazilian population of European ancestries.

The effect of the AGT M235T polymorphism on metabolism has also been studied. Pollex *et al* investigated the possible genetic determinants of metabolic syndrome in aboriginal Caucasians. They have chosen only polymorphisms that were previously found to be associated with individual components of MS in this population but they have not found any association between AGT M235T polymorphism and MS. Thomas *et al* also investigated this AGT polymorphism for a possible role in modulating various components of the metabolic syndrome in Chinese subjects. There was limited evidence to suggest that the AGT M235T polymorphism may be associated with certain disturbances in lipid metabolism. On the basis of the contradictory findings it was supported that the allele frequency of ACE gene varies according to ethnic group.

Both the ACE D allele and AGT M235T polymorphism were reported formerly to influence susceptibility to hypertension. Therefore we have expected an association between the polymorphisms and elevated blood pressure but our investigation failed to detect any significant association of the polymorphism with this disorder.

We have increased the homogeneity of study groups with MS by sub-grouping them according to clinical phenotype. This breakdown resulted in smaller sample sizes. Nevertheless some of the subgroups were large enough to detect significant differences in genotype distribution of ACE I/D polymorphism. Subjects with DD and ID genotype were found in this study to have significantly higher prevalence of dyslipidaemia. It can be interpreted as genes coding all components of RAS are found to be expressed in adipose tissue, which may have multiple functions including prostaglandin production and adipose tissue lipolysis. ACE inhibitor therapy has been shown to induce improvements in atherosclerosis and insulin resistance and by this means lowering the risk of developing diabetes, which suggests that the RAS pathway has significant impact not only on vascular functions but also on metabolic traits, and the activation of RAS may promote the development of dislipidaemia and diabetes.

In population genetics, an association may be due to confounders including population stratification (or confounding by ethnicity). Differing SNP allele frequencies in ethnically different strata of a population lead to a spurious association or 'mask' an association by artificially modifying allele frequencies in cases and controls when there is no real association. In our study the cases and controls do not differ in their ethnic backgrounds thus population stratification does not arise. We tested our study population for Hardy Weinberg equilibrium and we did not get any departure from it i.e. the genotypes are in HW equilibrium. It implies also that population stratification do/may not arise in our population.

Another key determinant of quality in an association study is sample size. Finding common genes with small effect size (OR of 1.2 - 1.5) and detecting it with high power requires relatively large sample size. It was calculated in advance that our sample size of approximately 1800 was required large enough to obtain power 80 % to detect an effect of size 1.2 or greater. One of the weaknesses of our study that it was a so called „single gene study”. However the characterization of the genetics of complex traits such as MS is

much more complicated. Many combinations of susceptibility alleles at several loci in a particular individual are possible, thus the risk of MS can not be predicted from the separate effect of each variant. Because of the above mentioned genetic heterogenity an individual polymorphisms may have low predictive value for complex human diseases. Beside construction of prediction models that estimate the effects of multiple susceptible gene variants, the investigation of gene-gene and gene-environment interactions also must be considered.

Summary

Translation of the results of genetic studies into meaningful actions in order to improve health and prevent diseases depends on scientific information from medical and public health studies. After a candidate gene is found, well-conducted epidemiological studies are needed to quantify the population impact of the gene variant, as well as to identify and measure the impact of modifiable environmental risk factors on the expression of its effect. Most of the genetic epidemiology studies have compared genotypes of patients in a clinical case series with those of a non-diseased sample of control subjects, so they provide little basis - if any - for risk estimation at population level. In order to characterize the role of disease-specific susceptibility gene variants in the Hungarian population, it is important to know the prevalence of risk alleles in the general population, i.e. genetic epidemiological studies require population-based DNA biobank collected according to well-defined selection criteria. Recently, based on the infrastructure of General Practitioners' Morbidity Sentinel Stations Program (GPMSSP) many epidemiological studies have been performed and in the framework of these studies the Reference DNA Biobank representative for the Hungarian population by age, gender and geographical distribution could be developed. The Hungarian Reference DNA Biobank – presently consists of 4267 samples – has provided population control samples for several national as well as international genetic epidemiological studies.

In the Hungarian population recently a cross sectional study was performed based on GPMSSP to define the prevalence of metabolic syndrome in the Hungarian adult population (20-69 years) as well as to estimate the frequency of ACE I/D and AGT M235T gene polymorphisms and define whether any association exists between metabolic syndrome and these gene polymorphisms in Hungarians. MS was defined according to the latest diagnostic criteria proposed by the International Diabetes Federation. The study population (n = 1762) representing the age and sex distribution of the general Hungarian population were recruited from the GPMSSP. The overall allele frequencies in the representative sample were 0.51 and 0.49 for AGT allele M235 and allele 235T, 0.47 and 0.53 for ACE allele I and allele, respectively, similar to those published in the literature for Caucasians. The frequency of DD genotype (31.36% vs.

25.42%, p = 0.006) and the frequency of D allele (0.56 vs. 0.51, p = 0.006) were significantly higher in the group of patients with metabolic syndrome than in the group free from the syndrome. The distribution of the AGT M235T polymorphism was similar in each group investigated. No relationship was found with any of the separate metabolic components and ACE genotypes, but in case of patients in whom central obesity was combined with elevated TG and low HDL-cholesterol level significant association was shown (p = 0.024 and p = 0.022), which suggests that ACE I/D polymorphism is likely to be involved in lipid metabolism.

Register Number: DEENKÉTK /207/2011.
Item Number:
Subject: Ph.D. List of Publications

Candidate: Szilvia Fiatal

Neptun ID: QBII75

Doctoral School: Doctoral School of Health Sciences

List of publications related to the dissertation

1. **Fiatal, S.**, Szigethy, E., Széles, G., Tóth, R., Ádány, R.: Insertion/deletion polymorphism of angiotensin-1 converting enzyme is associated with metabolic syndrome in Hungarian adults.
J. Renin-Angiotensin-Aldosterone Syst. Epub ahead of print (2011)
DOI: http://dx.doi.org/10.1177/1470320310394231
IF:1.6 (2010)

2. Dubois, P.C.A., Trynka, G., Franke, L., Hunt, K.A., Romanos, J., Curtotti, A., Zhernakova, A., Heap, G.A.R., Ádány, R., Aromaa, A., Bardella, M.T., van, d.B.L.H., Bockett, N.A., Concha, E.G.d.l., Dema, B., Fehrmann, R.S.N., Fernández-Arquero, M., **Fiatal, S.**, Grandone, E., Green, P.M., Groen, H.J.M., Gwilliam, R., Houwen, R.H.J., Hunt, S.E., Kaukinen, K., Kelleher, D., Korponay-Szabó, I., Kurppa, K., MacMathuna, P., Mäki, M., Mazzilli, M.C., McCann, O.T., Mearin, M.L., Mein, C.A., Mirza, M.M., Mistry, V., Mora, B., Morley, K.I., Mulder, C.J., Murray, J.A., Núnez, C., Oosterom, E., Ophoff, R.A., Polanco, I., Peltonen, L., Platteel, M., Rybak, A., Salomaa, V., Schweizer, J.J., Sperandeo, M.P., Tack, G.J., Turner, G., Veldink, J.H., Verbeek, W.H.M., Weersma, R.K., Wolters, V.M., Urcelay, E., Cukrowska, B., Greco, L., Neuhausen, S.L., McManus, R., Barisani, D., Deloukas, P., Barrett, J.C., Saavalainen, P., Wijmenga, C., Heel van, D.A.: Multiple common variants for celiac disease influencing immune gene expression.
Nature Genet. 42 (4), 295-302, 2010.
DOI: http://dx.doi.org/10.1038/ng.543
IF:36.377

3. **Fiatal S.**, Ádány R.: A népegészségügyi szempontból jelentős betegségekre hajlamosító genetikai mutációk előfordulása a magyar populációban.
Népegészségügy 87 (3), 185-194, 2009.

List of other publications

4. Tóth, R., **Fiatal, S.**, Petrovsky, B.É., McKee, M., Ádány, R.: Combined effect of ADH1B rs1229984, rs2066702 and ADH1C rs1693482/ rs698 alleles on alcoholism and chronic liver diseases.
   Dis. Markers. "accepted by publisher", 2010.
   IF:1.723

5. Tóth, R., Pocsai, Z., **Fiatal, S.**, Széles, G., Kardos, L., Petrovsky, B.É., McKee, M., Ádány, R.: ADH1B*2 allele is protective against alcoholism but not chronic liver disease in the Hungarian population.
   Addiction. 105 (5), 891-896, 2010.
   DOI: http://dx.doi.org/10.1111/j.1360-0443.2009.02876.x
   IF:4.145

Total IF: 43,845

Total IF (publications related to the dissertation): 37,977

The Candidate's publication data submitted to the Publication Database of the University of Debrecen have been validated by Kenezy Life Sciences Library on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

05 October, 2011

Relevant lectures and poster:

1. Fiatal Szilvia, Pocsai Zsuzsanna, Széles György, Ádány Róza: Magasvérnyomásra hajlamosító angiotenzinogén és angiotenzin konvertáz enzim genotípusok gyakorisága a magyar populációban. Népegészségügyi Tudományos Társaság XIV. Nagygyűlése, Szeged, 2005. április 20-22. (előadás)

2. Szilvia Fiatal, György Széles, Endre Szigethy, Róza Ádány: Insertion/deletion polymorphism of angiotensin -1 converting enzyme is associated with metabolic syndrome in Hungarian adults. A Debreceni Egyetem Orvos-és Egészségtudományi Centrum és az MTA DAB Orvostudományi és Biológiai Szakbizottságának tudományos ülése, 2008. (poszter)

3. Fiatal Szilvia, Szigethy Endre, Széles György, Tóth Réka, Ádány Róza: Az angiotenzin-konvertáz enzim ins/del polimorphizmusa fokozza a metabolikus szindrómára való fogékonyságot a magyar felnőtt lakosság körében DEOEC Egészségtudományok Doktori Iskola Ph.D. hallgatóinak 2009.évi szimpóziuma, (előadás)

5. Fiatal Szilvia, Szigethy Endre, Széles György, Tóth Réka, Ádány Róza: Az angiotenzin-konvertáz enzim ins/del polimorphizmusa fokozza a metabolikus szindrómára való fogékonyságot a magyar felnőtt lakosság körében, Népegészségügyi Képző- és Kutatóhelyek Országos Egyesületének III. Konferenciája, 2009.szeptember 24-26.(előadás)

6. Fiatal Szilvia, Ádány Róza: A népegészségügyi szempontból jelentős betegségekre hajlamosító genetikai mutációk előfordulása a magyar populációban DE OEC Egészségtudományok Doktori Iskola Ph.D. hallgatóinak 2010.évi szimpóziuma, (előadás)