Population genetic and Alzheimer's disease related gene-interaction studies on 17q21.31 genomic inversion

Ph.D. Thesis

Péter Zoltán Álmos, M.D.

Doctoral School of Clinical Medicine

Department of Psychiatry

Faculty of Medicine

University of Szeged

Supervisor: Zoltán Janka, M.D., Ph.D., D.Sc.

Szeged

2013

Papers the thesis is based on:

I. **Péter Zoltán Álmos**, Szatmár Horváth, Ágnes Czibula, István Raskó, Botond Sipos, Péter Bihari, Judit Béres, Anna Juhász, Zoltán Janka, János Kálmán
H1 tau haplotype-related genomic variation at 17q21.3 as an Asian heritage of the European Gypsy population
Heredity, 2008; 5:416-9 IF 3.823

II. Ágnes Fehér, Anna Juhász, Ágnes Rimanóczy, **Péter Álmos**, Judit Béres, Zoltán Janka, János Kálmán
Dopamine metabolism-related gene polymorphisms in Roma (Gypsy) and Hungarian populations
Journal of Genetics, 2011; 90:e72-5 IF 1.086

III. **Péter Zoltán Álmos**, Szatmár Horváth, Ágnes Czibula, István Raskó, Nóra Domján, Anna Juhász, Zoltán Janka, János Kálmán
Tau haplotypes and APOE4 do not act in synergy on Alzheimer's disease
Psychiatry Research, 2011; 186:448-50 IF 2.524

Cumulative impact factor: 7.433

Selected abstracts closely related to the thesis:

Péter Álmos, Ágnes Czibula, István Raskó, Judit Béres, Aranka László, Emőke Endreffy, Anna Juhász, Ágnes Rimanóczy, Zoltán Janka, János Kálmán
Tau gene as a population genetic marker and risk factor of tauopathies in a Hungarian Roma population
6th Congress of the Hungarian Human Genetic Association, Győr, Hungary, 2006. Abstract book, 76.p.

Ágnes Fehér, Judit Béres, Anna Juhász, Ágnes Rimanóczy, **Péter Álmos**, János Kálmán, Zoltán Janka
Investigation of dopamine system related genetic polymorphisms in Roma and non-Roma populations
7th Congress of the Hungarian Human Genetic Association, Pécs, Hungary, 2008. Abstract book, 51.p.

Péter Álmos, Szatmár Horváth, Nóra Domján, Anna Juhász, Zoltán Janka, János Kálmán
Examining gene–gene interactions between tau haplotypes and APOE4 in Alzheimer's disease
9th World Congress of Biological Psychiatry, Paris, France, 2009. Abstract book, 271.p.

TABLE OF CONTENTS

ABBREVIATIONS

AD: Alzheimer's disease

APOE, APOE4: apolipoprotein E, epsilon 4 allele

APP: amyloid precursor protein

Aβ: amyloid-β

DAT: dopamine transporter

DRD3: dopamine receptor D3

GWAS: genome wide association study

HWE: Hardy–Weinberg equilibrium

MAPT: microtubule associated protein tau

MMSE: Mini-Mental State Examination

NAHR: non-allelic homologous recombination

NINCDS/ADRDA: National Institute of Neurological and Communicative Disorders and Stroke /Alzheimer's Disease and Related Disorders Associations

OR: odds ratio

PCR: polymerase chain reaction

RFLP: restriction fragment length polymorphism

SFA: synergy factor analysis

SNP: single nucleotide polymorphism

SLC6A3: solute carrier family 6 (neurotransmitter transporter), member 3

SV: structural variant

VNTR: variable number of tandem repeats

BRIEF SUMMARY

The emerging research field in molecular genetics which studies structural genomic variants as significant contributors of genomic landscape is dating back 15 years. This work puts emphasis on the recently discovered genomic inversion at chromosome locus 17q21.31 from the following population genetic and medical points of views:

(1) The first observation on a structural variants' distinctive carrier rate in the Roma founder population is presented here, reflecting the inversion haplotype distribution as a hallmark of Asian ancestry in the Roma ethnicity. Furthermore, our data provide evidence that the similar heritage is diminished in the Hungarian population.

(2) These records are supported by a study focusing on the distribution of dopaminergic gene variants in the populations above.

(3) Beside the population genetic question a medical genetic aspect is involved by investigating the 17q21.31 variant in the light of its potential role in Alzheimer's disease. We present a case–control study which contributes to the growing research area of genomic disorders by examining the inversion from a gene–gene interaction point of view. Synergism of the inversion related microtubule associated protein tau genetic variant with apolipoprotein E status is analyzed. We support previous findings that latter is a risk factor of Alzheimer's disease in the Hungarian population. On the other hand the disorder was found independent from the inversion, revealing its variability as a risk factor among different populations. At last, we demonstrate that carefully chosen statistical methods can uncover false positive epistatis in genetic interaction research.

INTRODUCTION

Heritable biological features which define us as individuals or compose us to groups are examined by natural sciences. The blueprint of these factors is our genome, a system forming dynamically on an evolutionary timescale. Research on genome variants in the past decades revealed that significant individual specific changes – sometimes making up a couple of hundred kilobases in size – can turn to cumulative within families. Alterations in the genome, the rise of a new variant can help the survival of the individual and may have beneficial consequences on the adaptability and reproduction. This advantage can increase the variant's occurrence through generations in the population. On the other hand, since the genome is an open system, benefits or drawbacks of a variant always depend on boundary conditions. If those (e.g. environmental factors) change, the same variant can turn out as a shortcoming property, limitating the carriers in terms of fitness.

Considering from a medical genetic aspect the term variant may stand for a genetic alteration involved in disease development. More and more detailed knowledge on human genome made it possible to obtain data regarding probable factors of disease. This research field evolved parallel but independently with the molecular biological quest for variants which build up and characterize populations. In the past few years these two fields merged when research projects showed that certain variants are associated to disease only in certain populations. Studies now set sight on the possibility to uncover how population related genomic background contribute to development of pathology. In the following we provide an outlook to studies on mixed populations and the interplay between networks within the genome. Considering these aspects in medical research may facilitate to rise above the classic resolution of association studies.

Structural variants in the human genome

The genome's large scale structural rearrangements are comprehensively known since they are easily recognized by traditional cytogenetic methods as fluorescent in situ hybridization. Their discovery was early since they result in significant consequences regarding phenotype. Chronologically research focus then jumped to the other end of the variant size scale: single nucleotide variability guided the focus of interest since sequencing of human genome started. These two, especially single nucleotide polymorphisms (SNP) were studied extensively in the past decades, and majority of whole genome association studies were also based on single nucleotide variants (Sullivan et al., 2012).

There were great expectations to reveal the genetic background of common disorders by defining their genetic architecture. However the germline variants discovered by genome wide association studies (GWAS) explained only a small fraction of the traits which led to the issue of "missing heritability" (Manolio et al., 2009). It became obvious that common SNPs – variants which are present in the population with at least 5% frequency – give only a fraction of variability in the genome. Mapping all of the alternations is required to carry out a comprehensive study of genetic basis of phenotype. From 2006 it has turned out to be clear that variants of the genome build up a continuum from SNPs to larger rearrangements (Raphael, 2012).

Because of technological gaps, structural variants (SVs) ranging from 50 basepairs to megabases in size remained hard to find. Until 2007 genomic technologies had a bias toward typing unique tags (Baker, 2012), only after the development of array detection methods as microarray-based comparative genome hybridization and high-throughput pair-end sequencing was it possible to identify structural variants including insertions, duplications, deletions, inversions, recurring mobile elements covering more than 50 base pairs (see Table 1)

Although structural variants were first assumed as rare elements (e.g. classic cytogenetics identified 9 inversions distinguishing humans and chimpanzees), in the past few years it was

revealed that the major contributors to human genomic variation are structural variants (actually there is an order of 1500 inversions between humans and chimps, shown by Feuk et al., 2005). In contrast to SNPs which account for 0.1% of heritable nucleotide differences between individuals, SVs are responsible for 0.5-1%. This totals circa 50 megabases per genome. Furthermore, investigations considering evolutionary perspective clarified that *de novo* development of structural variants accelerated in primates, and this is clearly outstanding in chimpanzee and humans.

Table 1. Spectrum of variations (modified after Sharp, 2006)

Variation	Rearrangement type	Size range
Single base pair	Single nucleotide polymorphism, point mutation	1 bp
Small insertion/ deletion	Binary insertion/ short sequence deletion	1 – 50 bp
Short tandem repeat	Microsatellites, simple repeats	1 – 500 bp
Fine-scale structural variation	Deletions, duplications, tandem repeats, inversions	50bp – 5Kb
Retroelement insertion	Interspersed elements, long terminal repeat, endogenous repeat virus	300 bp – 10 Kb
Intermediate-scale structural variation	Deletions, duplications, tandem repeats, inversions	5 Kb – 50 Kb
Large scale structural variation	Deletions, duplications, tandem repeats, inversions	50 KB – 5 Mb
Chromosomal variation	Euchromatic variants, deletions, duplications, translocations, inversions, aneuploidy	5 MB – entire chromosomes

The high rate of their presence in the genome can be the result of so called genomic hotspots which are involved in the development of structural variants. SVs can emerge as a consequence of DNA recombination, replication and repair associated processes.

In Figure 1 non-allelic homologous recombination (NAHR) is illustrated. This is the major mechanism involved in the development of genomic rearrangements of human genome.

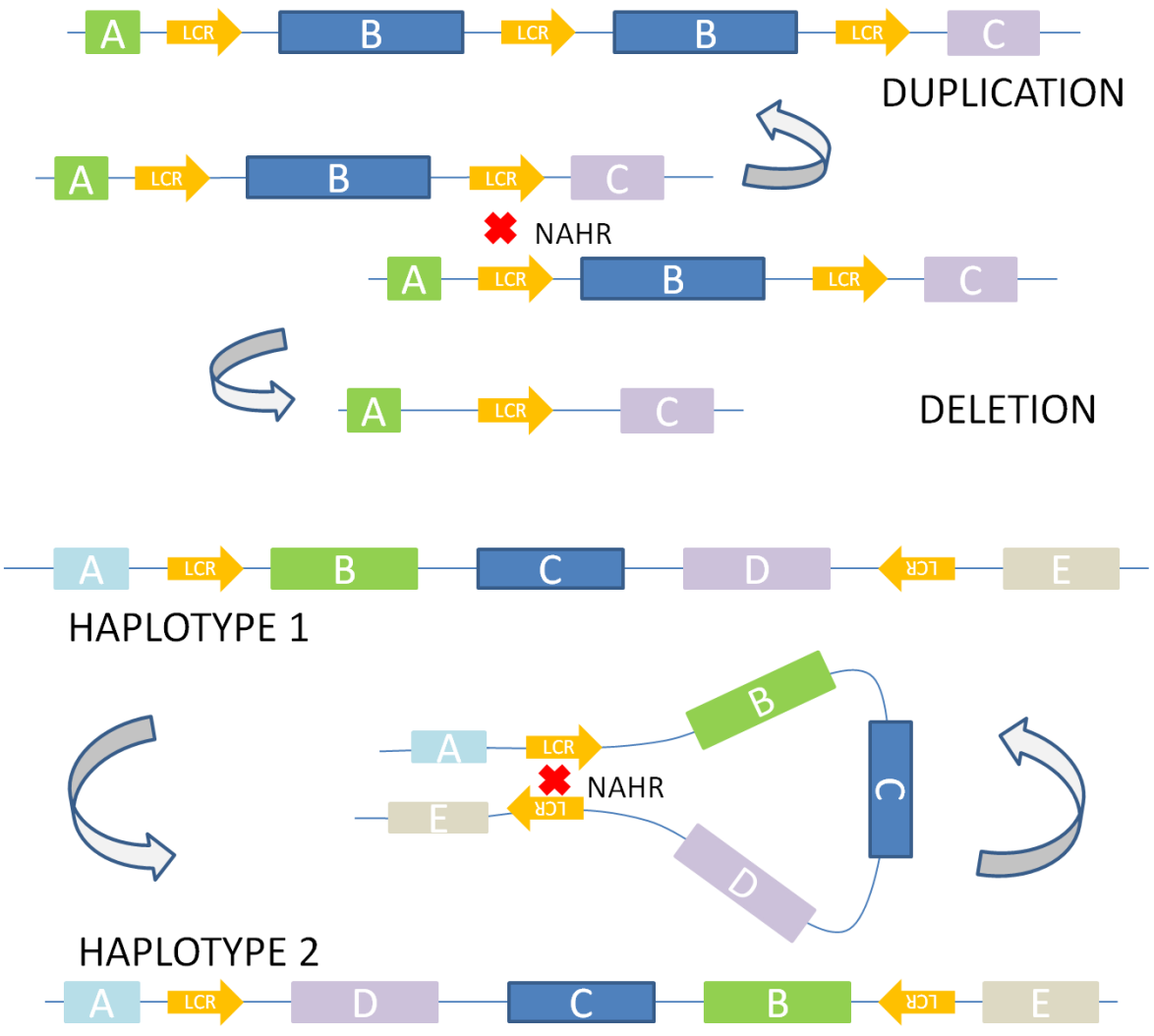

Figure 1. Deletion, duplication and inversion evolved via NAHR

In some cases non-homologous end-joining, mobile element insertion and Fork Stalling and Template Switching models shape these rearrangements (Weischenfeldt et al., 2013).

Depending on the recombination process structural variants can be differentiated as unbalanced variations characterized by quantitative change in genome material or balanced variations which do not result in genomic mass alteration.
Inversions are balanced structural variations and evolve if NAHR take place between segmental duplications or highly identical sequences of inverted orientation. These structural variants are extremely complicated to detect. Although the first evidence on a chromosomal inversion was published in 1921 by Alfred Sturtevant, uncovering submicroscopic inversions is still a challenge, as inversions cannot be detected via arrays. Recently, with pair end sequencing their discovery has accelerated but the number of inversions found is still insignificant compared to CNVs (Baker, 2012).

After overcoming the difficulties of detection, it is even harder to interpret them in respect to their functional consequences. The first notion that genomic structural variants are involved in common diseases dates back to 1998, when the term “genomic disease” was used for the first time (Lupski, 1998). Since then and especially in the past 5 years several studies testified the role of CNVs in complex genomic disorders (with a prominence in mental disorders). Despite this progress, our knowledge on inversion related phenotypic variants is still very restricted. Beside the detection-bias this can be partially because in contrast to CNVs even large inversions can remain neutral on phenotype since there are no dosage imbalances. On the other hand, there is an intriguing question on inversions which clearly points toward their significance: if an inverted chromosome has the same genetic information as its pair, why does it spread in the population (Kirkpatrick, 2010)?

A key point to consider is that inversions evolve by leaving breakpoints. Several examples show that inversions can lead to the genetic consequences by their positional effect (see Figure 2). By disturbing the architecture of the genome by their breakpoint, they can affect

coding regions or interrupt transcriptional regulation, even inducing over- or ectopic expression and accelerate the emergence of further variants.

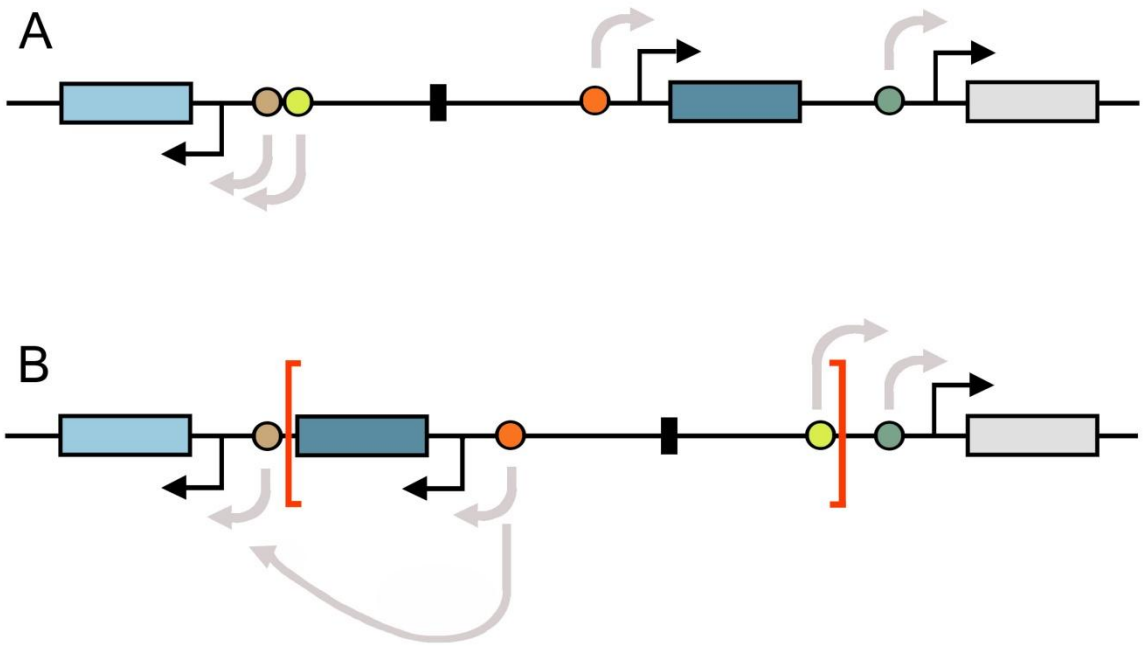

Figure 2. Functional consequences of an inversion through positional effect

A: Normal genomic landscape where gene-expression is tuned by regulatory elements

B: After an inversion event all coding regions may remain intact without any quantitative imbalances. However the inversion tumbles up the sequence leading to altered expression

Similarly, the role of inversions in disease is sometime not directly causative, rather increase the risk of further rearrangements that cause disease. Disease associated genomic rearrangements can be recurrent with fixed breakpoint; or non-recurrent with a minimal

region of overlap which is associated strongly to the locus conveying the disorder (Feuk, 2010). Furthermore, inversions also differ from other structural variants as recombination is suppressed among heterokaryotypes.

Research on 17q21.31 in Alzheimer's disease

A good example regarding the multifaceted nature of inversions is demonstrated by the research on one of the most dynamic and complex region of the human genome the 17q21.31 locus. Among other genes this region codes microtubule-associated protein tau (MAPT), which is widely studied, as it contributes to several human diseases (Hardy et al., 2006). The main function of microtubule associated protein tau (MAPT) is to maintain the cellular structure and morphology (Avila, 2006) in neurons. Beside physiological function, tau protein is much more investigated as the core element of neurofibrillary tangles, the major hallmark of neurodegenerative disorders, especially Alzheimer's disease (AD). Certain variants and mutations of MAPT gene are more likely disposed to tau protein hyperphosphorylation, leading to the development of tauopathies as a consequence of neurofibrillary tangle formation. The 900 Kb inversion at 17q21.3 is one of the most notable structural variants found to date. Since its identification in 2005 (Stefansson et al., 2005) it is in the centre of research interest, as it encompasses several genes (Kalinderi et al., 2009). By this inversion two non-recombining major MAPT allele forming haplotypes (H1 and H2) can be differentiated (see Figure 3) which affect MAPT related pathomechanisms in distinctive manners.

The extensive investigations revealed that out of the two main non-recombining MAPT locus haplotypes H1 plays a role in the development of sporadic tauopathies (Laws et al., 2007) while H2 is involved in a neurodevelopmental disorder.

H2 can lead to a disorder-related phenotype with germline breaking mechanisms which were clarified in the past 3 years. The most studied H2 haplotype associated neurodevelopmental disorder is Koolen De Vries syndrome. In this case the H2 haplotype (H2D subhaplotype)

provide possibility to the development of a causative microdeletion event, encompassing MAPT and leading to mental retardation.

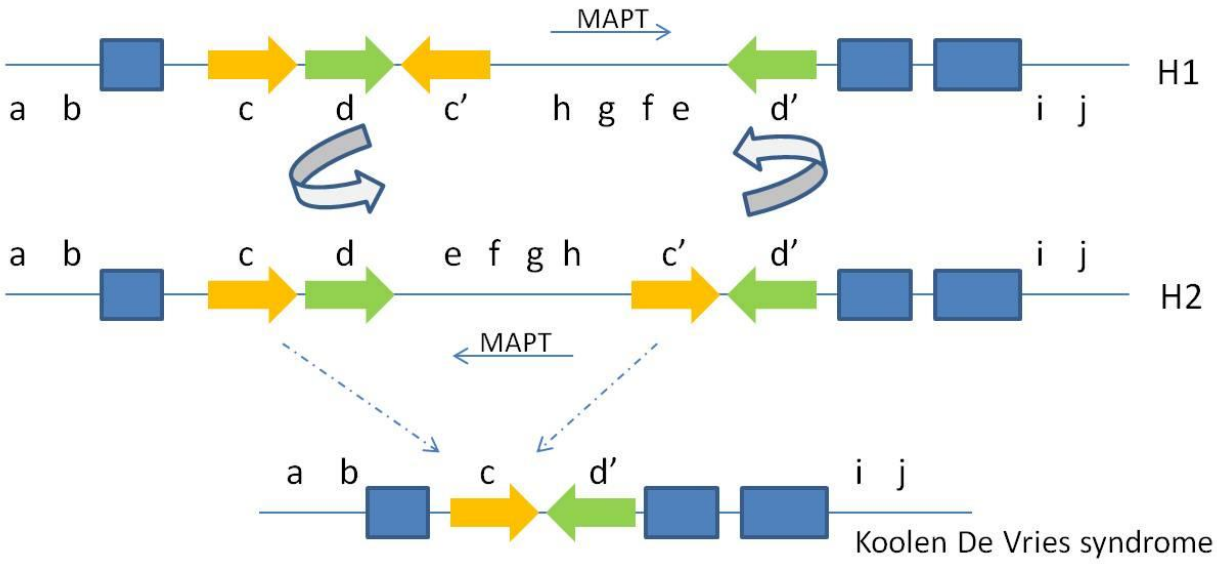

Figure 3. The architecture of H1 and H2 17q21.31 regions

The H2D inversion result in a genomic architecture which give rise to NAHR (Shaw-Smith et al., 2006) and consecutive microdeletion. The disorder therefore limited to H2 and cannot appear in H1 carriers.

According to recent publications H2 was found to be the more ancient haplotype (Zody et al., 2008) and in spite of its association to the Koolen De Vries syndrome it is a target of positive selection in Europe since H2 carrier women have higher recombination rate with higher reproductive success (Stefansson et al., 2005).

Regarding H1 other effects can take effect making H2 as the protective haplotype. The H1 clade is involved in pathophysiologic processes probably as a consequence of more various alternative splicing and expression with clear genetic association to Parkinson's disease (Zabetian et al., 2007), progressive supranuclear palsy (Pittman et al., 2004), argyrophilic grain disease (Fujino et al., 2005), corticobasal degeneration (Pittman et al., 2005), frontotemporal dementia (Verpillat et al., 2002) and lower regional cerebral gray matter volume in healthy individuals (Canu et al., 2009). Its association to Alzheimer's disease is also supported by many findings (Myers et al., 2007) but there are controversial results too (Caffrey and Wade-Martins, 2012). The profound question is, however, why some studies have refuted this association or found carriage of allele H2 to be a risk factor for neurodegenerative disorders (Ghidoni et al., 2006, Russ et al., 2001)?

An explanation for the controversy could be the possible epistatis or interactions of different disease specific susceptibility genes. Inversions affecting regulatory element can lead to loss of autoregulation or disturbed interaction with other genes (see Figure 4 on next page).

In complex disorders such as tauopathies, single gene association studies often lead to controversial results because they are not sensitive enough to reveal the role of a gene with limited effect. As an example, leaving out of consideration the interaction of the major pathways implicated in the pathogenesis of Alzheimer's disease may lead to type 2 errors (Combarros et al., 2009). In order to find a suitable model for sporadic complex disorders (as late onset Alzheimer's disease), genetic association studies with special emphasis on the synergistic effects of disease associated genes should be performed (Corder et al., 2006).

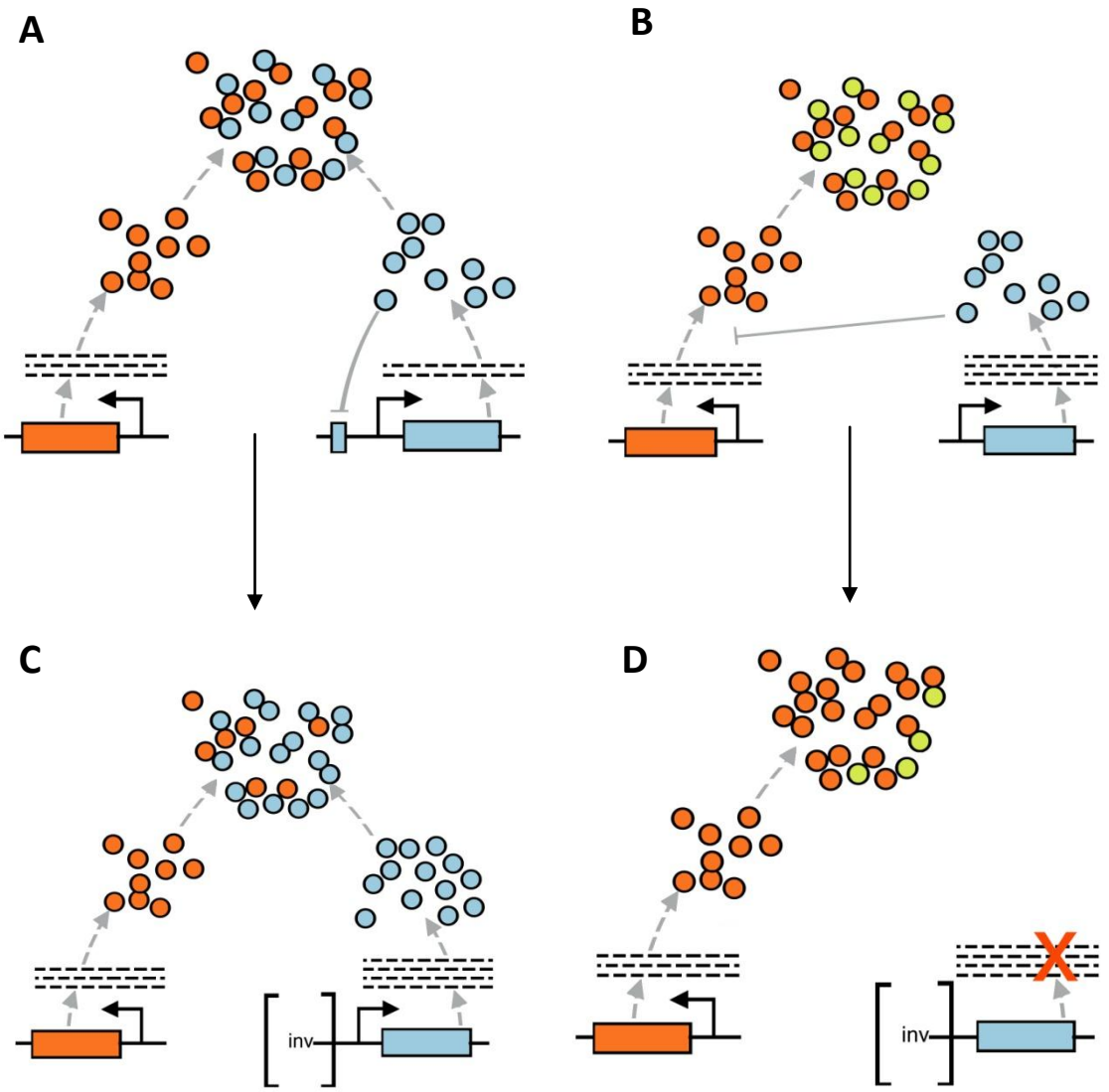

Figure 4. Positional effect resulting in impaired autoregulation and loss of epistasis

Normal gene autoregulation (A) or genetic epistasis (B) results in a ratio of proteins which form complexes in a balanced manner. Inversions (C and D) disintegrate the genomic architecture disturbing regulatory elements and initiating cascades in biological pathways.

Another answer to the issue might lie in the notable difference in the ethnical distribution of the two main haplotypes. H2 haplotype is rare in Africans, and almost absent in East Asians and Native Americans, but very frequent (20–30%) in populations of European Caucasian origin (Evans et al., 2004, Stefansson et al., 2005). While in the beginning it has even been postulated that the H2 haplotype was contributed to the human genome by *Homo neanderthalensis* (Hardy et al., 2005), broad evidence support now that H2 emerged in Eastern or Central Africa and was replaced by H1 cca. 2.3 million years ago in the *Homo* ancestral populations. H2 later expanded exclusively and rapidly in the European out-of-Africa populations and became Caucasian-specific (Donelly et al., 2010, Steinberg et al., 2013).

A few centuries ago humans opened a new, exciting chapter in their genetic history by leaving their geographical environment with accelerated migration and changed it to an unfamiliar one in evolutionary extremely short time. In the new milieu the genomic architecture which was adapted and fine tuned to the environmental triggers for thousands of years might face novel triggers which induced detoriation in homeostasis. The consequence can be population specific risk factor of disorders or altered response to treatment (Yang et al., 2011).

The new developments in large scale genome sequencing (e.g. 1000 Genomes Project, http://www.1000genomes.org) provided an opportunity to generate geographical maps of the frequency of structural variants. Some of them are typical for different ethnic groups or for certain populations (Gu et al., 2007, O'Hara, 2007, Spielman et al., 2007). The presence of these variations is thought to be an important contributor to the evolution in human genetic diversity and can generate difference in disease susceptibility (Feuk et al., 2006). Thus, medical genetic studies with a special focus on population genetics started to examine admixed population as a form of disease associated gene-discovery (Seldin et al., 2011).

The intriguing genetic history of two populations in the Carpathian basin

In this work two historically and ethnically different populations (Roma/Gypsies and Caucasian Hungarians) are examined from the same geographical area. The Caucasian Hungarians belong to the Uralic linguistic family, a diverse group of people related by an ancient common linguistic heritage, distinct from that of the Indo-European speakers who surround them. Of the approximately 25 million Finno-Ugrians, the best known are the Estonians, the Finns and the Hungarians. Around the 5th century BC, the ancient Hungarians were caught up in a wave of migrations that swept the steppes and were displaced from their western Siberian homeland. Migrating westwards, the Hungarians arrived in 895 in the Carpathian Basin, an area where the overwhelming majority of the indigenous population was Slavic. Various genetic appraisals have estimated that the newly arrived Hungarians accounted for 10–50% of the total population of the Carpathian Basin (Cavalli-Sforza et al., 1994). During the turbulent history of present-day Hungary, the mixing process has continued, and Hungarians can now be regarded as members of a mixed European population (Semino et al., 2000). In contrast to Hungarians, Roma are a conglomerate founder population with Asian Caucasian roots, imbedded in a genetically different European Caucasian population. The social sciences and comparative linguistic studies have hinted at the Asian origin, and this has been supported by population genetic studies of single-locus polymorphism, of multi-locus STR Y chromosome haplotypes and of mtDNA haplotypes (Gresham et al., 2001, Kalaydjieva et al., 2001a, Morar et al., 2004, Rai et al., 2012, Mendizabal et al., 2012). The most recent study investigated Roma SNP data in 6 populations (Moorjani et al., 2013). They revealed that in present-day Roma populations' characteristics of Eastern-European and North-Western Indian heritage can be revealed. The estimated time of the founder event occurred about 27 generations (~800 years) ago. The combined evidence suggests that Roma migrated from Punjab region of Northwest India 1000–1500 years ago and traveled through Asia (along Persia, today's Armenia and Turkey). The main stream moved into the Balkans and Greece and some of them into Eastern Europe ahead of the Turks. Early diaspora appeared in western Europe around the period from the fourteenth to the

fifteenth century, and another wave of migrations to western Europe started after the abolition of serfdom in the Habsburg Empire in 1841, and recently from 1989 after the disappearance of the Iron Curtain (Kalaydjieva et al., 2005).

At present, 8–10 million Roma live in fragmented subisolates in Europe, approximately 600000 of them in Hungary. In Roma society, the primary unit is the group, and groups are members of metagroups. They live in a closed society structure, with rare admixture with other populations, and a relatively high rate of consanguinity (Assal et al., 1991). There appears to have been population bottlenecks, both when they left India and during the European segregation. A high intragroup diversity can be observed (Gresham et al., 2001). Hungarian Roma were not classified in previous publications or were included among western European Roma/ Gypsies (Morar et al., 2004). However, we think that the comparison of the Hungarian Roma population is an adequate choice for genetic investigations because the ethnic diversity in Hungary is not as high as in the Balkans, and it is possible to distinguish three well-described metagroups among Hungarian Roma. Carpathian Roma or Romungros are the least characterized and intact metagroup. Their language consists elements from Beas, Lovari and Hungarian. They represent the 70% of the Roma living in Hungary.

The two smaller metagroups are more closed and cohesive; they live typically in separated parts of smaller villages or towns. They preserve their traditions and language; as a consequence, the assimilation with other metagroups or with the Caucasoid Hungarian population is low. Beas represents 10% of the Hungarian Roma population; their migration to the Carpathian Basin came from the Central-West Balkans. They speak the Beas language. The Olahs, with a proportion of 20% from the Hungarian Roma population, arrived at the Carpathian Basin from the territory of today's Romania and they speak the Lovari language. They are the descendants of the Valachian/Vlax Roma, the most studied Roma population (Kalaydjieva et al., 2001a).

Genetic variants in the Roma population

We have limited information on the spectrum of genetic variants in the Hungarian Roma population. Most of data available focuses on SNPs and there are several founder effect associated, clinically relevant findings from Hungarian research groups (e.g. Sipeky et al., 2009, 2013). On the other hand, SVs are barely investigated. Although the published genetic research on Roma populations is fragmentary so far, it indicates that medical genetics can have an important role in improving the health conditions and health statistics of the Roma population. Several mendelian disorders and private mutations have been identified, but the distribution of alleles that lead to genetically complex diseases have not been studied systematically in the Roma (Kalaydjieva et al., 2001b). There is a need for further research, because no exact data are available on the prevalence of psychiatric diseases or the genetic background of these disorders in the Roma population.

In this work primarily we aimed to investigate the 17q21.31 structural variant. In a satellite study, supporting our goal, variants representing other rearrangement types are also included. Since complex mental disorders make up our main area of interest, candidate genes of dopaminergic pathways were investigated.

Dysfunctions of the dopaminergic system occur in several neuropsychiatric disorders, such as schizophrenia, bipolar affective disorder, drug abuse and Parkinson's disease (Cousins et al., 2009, Halliday and McCann, 2010, Lodge and Grace, 2011). The susceptibility to these disorders can be mediated by variants of genes involved in dopaminergic transmission, i.e. dopamine transporter and dopamine receptors (Hoenicka et al., 2007). The dopamine transporter (DAT) gene (*SLC6A3*) 40 bp variable number tandem repeat (VNTR) and the dopamine D3 receptor (*DRD3*) Ser9Gly polymorphisms have been widely studied for population variations, but until now the Roma population was not examined for these markers. DAT is responsible for the presynaptic reuptake of dopamine and it is also the target of several psychoactive drugs (Kang et al., 1999). The human SLC6A3 gene is located on chromosome 5p15.3 and a 40 bp VNTR polymorphism has been identified in the 3'

untranslated region (Sano et al., 1993). The diverse physiological functions of dopamine are mediated by five different dopamine receptors. The D1 and D5 receptors are members of the D1-like family of dopamine receptors, whereas the D2, D3 and D4 receptors are members of the D2-like family. DRD3 is predominantly expressed in limbic brain areas which are altered in several psychiatric disorders (Bouthenet et al., 1991). The DRD3 gene has been mapped to chromosome 3q13.3. A single-nucleotide polymorphism (SNP) in the 5' part of the DRD3 gene producing a non-conservative amino acid substitution at codon 9 (Ser/Gly) has been identified (Lannfelt et al., 1992).

AIMS

I

The first aim of this work was to investigate the allocation of 17q21.31 related genomic inversion haplotypes in Hungarian Roma populations. Since they are Asian in origin our hypothesis was that the frequency of Caucasian-specific H2 haplotype is low as a result of closed Roma societies. The control population was Hungarians where previous studies showed lack of genetic heritage reflecting the Asian roots.

II

The second goal was to carry out an independent satellite investigation on a larger group and study well-characterized genomic variants to support our findings in the previous study. The present study provides the first data about the SLC6A3 40 bp VNTR and the DRD3 Ser9Gly polymorphisms in Roma population in Hungary.

III

Third, the region of our interest is controversially related to complex psychiatric disorders, that is, tauopathies. Since the region's structure show great variability in different populations, we found it important to examine its relation to Alzheimer's disease in the Hungarian population. Moreover, we extended this study to examine the genetic interaction with the widely replicated apolipoprotein E epsilon 4 (APOE4) allele. Considering the convergence of their biological pathways (Adalbert et al., 2007), we examined the possible interaction of tau H1 haplotype and APOE4 in the Caucasian Hungarian population.

METHODS

Study I

Sample characteristics

In this study, 118 healthy Roma of the Olah/Vlax metagroup and 184 healthy Caucasian Hungarians were genotyped. The Roma participants were recruited from three villages in the same geographical area in northeastern Hungary. The Hungarians were employees and students of Department of Psychiatry, University of Szeged, and Department of Hungarian Congenital Abnormality and Rare Disease Registry of the National Centre For Healthcare Audit and Improvement and their acquaintances, who were matched with the Roma volunteers for age and gender. After complete description of the study to the subjects, written informed consent was obtained.

DNA isolation

The genomic DNA of Roma and control subjects was extracted from peripheral blood according to a standard method (Davies, 1993).

Genotyping

In this study, our goal was to evaluate the H1–H2 haplotype frequencies in the populations mentioned above by using a polymorphism of MAPT gene as a marker. The selected region was amplified by the means of the PCR. The inverted chromosome region was screened by applying the standardly used biallelic intron 9 deletion-inversion polymorphism (Baker et al., 1999). The following primer pairs were used: forward: 5'-GAAGACGTTCTCACTGATCTG-3'; reverse: 5'-AGGAGTCTGGCTTCAGTCTC-3'. Polymerase chain reaction amplification was carried out in 20 µl reaction volume containing 2 µl of 10xZenonBio, 10x reaction buffer, 50 nM of each of the primers, 0.5 mM of each of the dNTPs, 4 mM MgCl2, 100 ng of DNA extract and 0.3U of ZenonBio TaqPolymerase. The amplification protocol was as follows: 3 min at 93 °C, 30 cycles of 93 °C for 60 s, 60 °C for

60 s and 72 °C for 60 s, and final extension at 75 °C for 5 min. A volume of 7 µl of PCR product was run on 6% native polyacrylamide gel and visualized after ethidium bromide staining by UV transillumination, and the size of the products was determined with the gelBase gel documentation system (UVP).

Statistical analysis

The departure from the Hardy–Weinberg equilibrium was tested by using the 'HWE.test' function (P-value calculated by the exact method) of the genetics R package (R version 2.4.0, R Development Core Team, 2006; Warnes, 2008). Fisher's exact tests carried out in R were used to determine the significance of differences in genotype and allele frequencies.

Study II

Sample characteristics

The study included 189 Olah Roma and 189 Hungarian Caucasian healthy probands from Hungary. The mean age (±SD) for the Olah Roma group was 38.3 (±13.2) years, and for the Hungarian group 45.1 (±16.1) years. The male/female ratio was 76/113 in the Olah Roma and 82/107 in the Hungarian group. Informed consent was obtained from the subjects and all protocols were approved by the Ethics Committee of the University of Szeged.

DNA isolation

DNA was extracted from peripheral blood leucocytes according to a standard procedure using the Roche High Pure PCR Template Preparation Kit (Roche Applied Science, Basel, Switzerland).

Genotyping

SLC6A3 40 bp polymorphism genotyping was made by polymerase chain reaction (PCR) as described earlier (Sano et al., 1993). Genotyping of the DRD3 Ser9Gly polymorphism was conducted by PCR amplification and then enzymatic digestion with the restriction enzyme

*Msc*I, followed by polyacrylamide gel electrophoresis with ethidium bromide staining (Lannfelt et al., 1992).

Statistical analysis

The statistical analyses were performed by using SPSS 15.0 for Windows software (SPSS, Chicago, USA). The significance level for all statistical tests was set at 0.05. The Pearson's χ^2 test was used to compare the SLC6A3 and DRD3 genotypes and the SLC6A3 allele frequencies between the Olah Roma and Hungarian groups, while Fisher's exact test was applied to assess DRD3 allelic differences between the two investigated populations. Hardy–Weinberg equilibrium (HWE) for the distribution of SLC6A3 and DRD3 genotypes was estimated by Pearson's χ^2 test.

Study III

Sample characteristics

One hundred and seventy-four Caucasian probands participated in our study from the Southern Hungarian Region. The 91 AD (mean age / SD, 69.5 / 11.9 years; male/female 40/51; MMSE score / SD, 14.1/ 4.9 points) patients were randomly selected from the outpatient population of the University of Szeged Memory Clinic. The clinical diagnosis of late onset sporadic AD was based on the ICD-10 and the generally accepted criteria of the National Institute for Neurological and Communication Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA). The AD probands were considered sporadic type, because none of them had a family history of dementia. All patients underwent CT and MRI studies (in some dubious cases diagnose was confirmed by SPECT) in order to exclude any other neurological disorder. The 83 healthy control persons (mean age / SD, 67.4 / 12.3 years; male/female 42/41; Mini-Mental State Exam (MMSE) score $\geq$ 28 points) were spouses of the AD probands and none of them had verified symptoms of dementia. All the participants gave their informed consent to the study, which was approved

by the local ethics committee. The study was conducted according to the Declaration of Helsinki and subsequent revisions.

DNA isolation

The genomic DNA of AD and control subjects was extracted from peripheral blood according to a standard method (Davies, 1993).

Genotyping

MAPT genotyping is described in Study I (Almos et al., 2008). APOE alleles (E2, E3, and E4) were determined by previously described polymerase chain reaction-based strategy (Kalman et al., 1997). Briefly, PCR reaction was performed in a PTC 100, Thermal Controller MJ Res. Inc. thermal cycler. The final volume of PCR solution was 25 µl, containing 20 µM of two primers (5'-TCCAAGGAGCTGCAGGCGGCGCA-3', and 5'-ACAGAATTCGCCCCGGCCTGGTACACTGCCA-3'), 1.25 µl from each, 50-300 ng of genomic DNA, 1.25 µl of dNTPs (20 mM), consisting a mix of 5 mM of each, 1.5 µl (25 mM) MgCl2, 2.5 µl (5%) dimethyl sulfoxide, 0.5 U of TaqDNA polymerase (Promega), in 67 mM TRIS-HCl buffer (pH 8.8). The initial denaturation was 5 min at 95°C, followed by 30 cycles of 30 s at 94°C denaturation, 22 s at 63°C annealing and extension for 30 s at 72°C. A final extension for 3 min at 72°C completed the amplification procedure. The amplified DNA was digested with 5 U CfoI (Promega) overnight at 37°C, and the DNA fragments (91, 81, 72, 48 base pairs) were separated on 8% non-denaturing acrylamide gel. The gel was stained with 0.5 µg/ml ethidium bromide, and APOE genotype was determined by the pattern of DNA fragments present.

Statistical analysis

Age and MMSE scores were normally distributed (according to Kolmogorov-Smirnov test) and compared by independent samples t-test. Variances of MMSE scores were unequal (according to Levene's test), therefore it was compared with Welch's t-test. Alleles and genotypes were counted and their distribution between the groups was compared by Pearson

χ^2 test. The data were analyzed using SPSS for Windows (version 12.0). χ^2 effect sizes and power calculation were estimated by PASS 2008 software. Since small sample size is limiting the power of the study, the examination of gene interaction was carried out by the mean of synergy factor analysis (SFA) (Combarros et al., 2009). Based on logistic regression, SFA can be used as a method which provides statistically reliable data in case of limited number of participants. Power calculations for expected synergy factors were estimated by the statistical software R (v. 2.8.1) with the script "SFProgrammes.r", provided by Mario Cortina-Borja (Cortina-Borja et al., 2009).

RESULTS

Study I

The MAPT allele frequencies in the Caucasian sample were in Hardy–Weinberg equilibrium (P=0.842). A deviation from the Hardy–Weinberg equilibrium was observed in the Roma population sample (P=0.017). The distribution of MAPT genotypes are presented in Figure 5. The MAPT H1 homozygote haplotype is seen to be overrepresented in the Roma as compared with the Caucasians (83.0% (n=98) vs. 56.5% (n=104) one-tailed P<0.001). H1/H2 heterozygotes prevail in the Caucasian population (38.0%; n=70 in the Caucasians vs. 13.6%; n=16 in the Roma) (one-tailed P<0.001). The calculated frequency of the H1 allele in the Roma population was greater than that in the Caucasians (89.8% (n=212) vs. 75.5% (n=278) (one-tailed P<0.001), whereas H2 allele was more dominant in the Caucasian population.

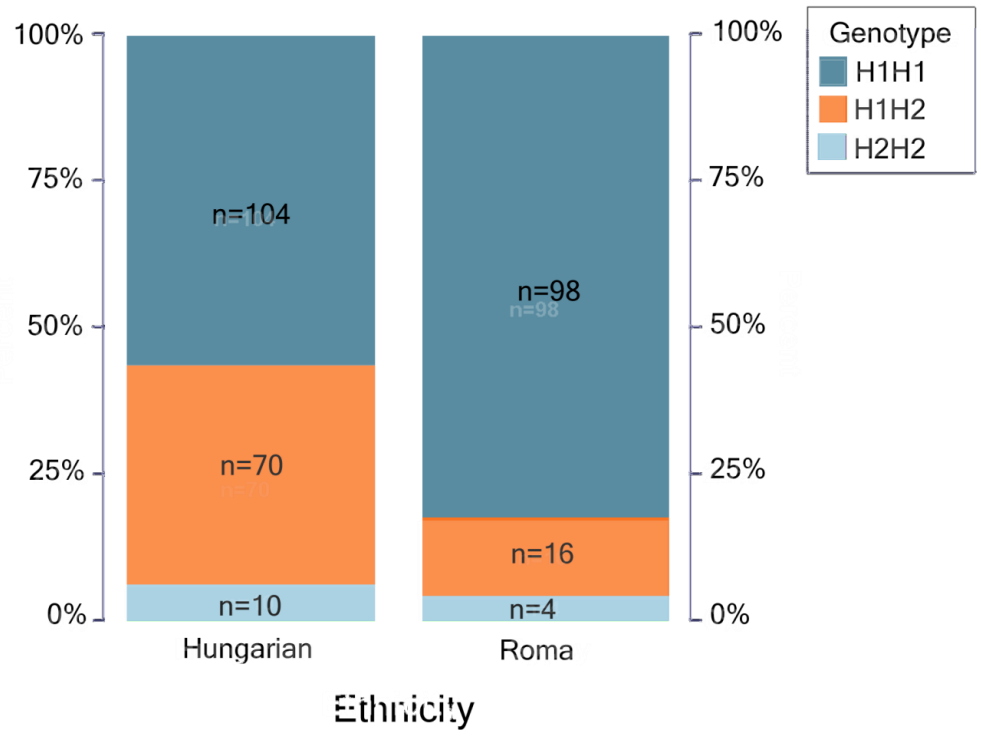

Figure 5. Ethnicity dependent MAPT genotype distribution

Study II

Genotype and allele distributions of *SLC6A3* 40 bp VNTR polymorphism are shown in Table 2 and 3. In this polymorphism the different alleles are determined by the copy number of a 40 bp long DNA segment in the 3' untranslated region of the *SLC6A3* gene. Four types of *SLC6A3* alleles were found in this study: the eight-repeated (A8), the nine-repeated (A9), the ten-repeated (A10) and the eleven-repeated (A11) alleles. In the Olah Roma group no A8 allele carriers were detected, while we found only one person with A8/A10 genotype among the Hungarians. The frequency of the A9/A10 genotype was significantly higher in the Hungarian population as compared to the Olah Roma group (Roma, 23.8%; Hungarian, 43.9%). The frequency of the A10/A11 genotype was significantly higher in the Olah Roma population than in Hungarians (Roma, 8.5%; Hungarian, 1.6%). The A10 allele occurred with similar frequency in the two populations (Roma, 72.2%; Hungarian, 70.6%). In contrast, the occurrence of the A9 allele was significantly lower, whereas the A11 frequency was significantly higher in the Olah Roma population than in the Hungarian probands (A9: Roma, 20.4%; Hungarian, 28.0%; A11: Roma, 7.4%; Hungarian, 1.1%).

The DRD3 Ser9Gly genotype and allele frequencies are presented in Table 4 and 5. Comparison of DRD3 genotype frequencies between the Olah Roma and Hungarian groups showed no significant difference, although the frequency of the Ser9Ser homozygous genotype was numerically lower and the frequency of the Ser9Gly genotype was numerically higher in the Olah Roma than in the Hungarian population (Ser9Ser: Roma, 42.9%; Hungarian, 50.3%; Ser9Gly: Roma, 52.9%; Hungarian, 45.0%). The Gly9Gly genotype occurred with similar frequency in the two populations (Roma, 4.2%; Hungarian, 4.7%). Similarly, there were no statistical differences in the occurrence of DRD3 alleles in Olah Roma population as compared to the Hungarians (Ser9: Roma, 69.3%; Hungarian, 72.8%; Gly9: Roma, 30.7%; Hungarian, 27.2%). The SLC6A3 and the DRD3 genotype frequencies were in HWE in the Hungarian group (SLC6A3, P=0.548; DRD3, P=0.065), while a deviation from the HWE was detected in the Olah Roma population (SLC6A3: P<0.001; DRD3: P=0.001).

Table 2. Dopamine transporter genotype frequencies

Genotypes*	Roma *n*=189	Hungarian *n*=189
A8/A10	-	1 (0.5%)
A9/A9	15 (7.9%)	11 (5.9%)
A9/A10	45 (23.8%)	83 (43.9%)
A9/A11	2 (1.1%)	1 (0.5%)
A10/A10	106 (56.1%)	90 (47.6%)
A10/A11	16 (8.5%)	3 (1.6%)
A11/A11	5 (2.6%)	-

*χ^2=28.431 (6), *P*<0.0001

Table 3. Dopamine transporter allele frequencies

Alleles**	Roma *n*=189	Hungarian *n*=189
A8	-	1 (0.3%)
A9	77 (20.4%)	106 (28.0%)
A10	273 (72.2%)	267 (70.6%)
A11	28 (7.4%)	4 (1.1%)

** χ^2=23.662 (3), *P*<0.0001

Table 4. Dopamine D3 receptor genotype frequencies

Genotypes*	Roma *n*=189	Hungarian *n*=189
Ser9Ser	81 (42.9%)	95 (50.3%)
Ser9Gly	100 (52.9%)	85 (45.0%)
Gly9Gly	8 (4.2%)	9 (4.7%)

*χ^2=2.389 (2), *P*=0.303

Table 5. Dopamine D3 receptor genotype frequencies

Alleles**	Roma *n*=189	Hungarian *n*=189
Ser9	262 (69.3%)	275 (72.8%)
Gly9	116 (30.7%)	103 (27.2%)

** Fisher's exact *P*=0.336

Table 6. Dopamine transporter gene and dopamine D3 receptor allele frequencies in different populations

	European Caucasian			Indian origin	Indian
SLC6A3	Swedish*	French Canadian*	Hungarian*	Roma*	North Indian#
DRD3	Jönsson et al., 1993	Joober et al., 2000	Present work	Present work	Prasad et al., 2008
A8	-	-	0.3%	-	-
A9	-	25.5%	28.0%	20.4%	-
A10	-	74.5%	70.6%	72.2%	-
A11	-	-	1.1%	7.4%	-
Ser9	72%	71%	73%	69%	64%
Gly9	28%	29%	27%	31%	36%

* healthy probands, # type-2 diabetes subjects

Study III

There was no significant difference between the mean ages of the two groups (t=-0.424, df=158, P=0.672). MMSE scores of the patient group were significantly lower than those of the control group *(Welch's* t=25.948, df=79.829, P<0.0001). MAPT and APOE allele frequencies were in Hardy–Weinberg equilibrium both in control (MAPT χ^2=0.009, df=1, P=0.99), (APOE P=0.93, χ^2=1.36, df=5) and AD samples (MAPT χ^2=0.06, df=1, P=0.97), (APOE χ^2=5.33, df=5, P=0.38). χ^2 effect sizes were small regarding all MAPT genotypes (χ^2=1.03, df=2, P=0.59), MAPT alleles (χ^2=0.97, df=1, P=0.32) and APOE4 state (χ^2=4.01, df=1, P=0.03). Our sample size of 91 patients and 83 controls achieves 65% power to detect the widely accepted effect size of 0.2 using a 2 degrees of freedom Chi-square Test with a significance level of 0.05 (Cohen, 1988).

Table 7. MAPT and APOE genotype and allele frequencies in patients and controls

	MAPT			MAPT		APOE4	
	H1/H1	H1/H2	H2/H2	H1	H2	+	-
AD	60.4%	35.2%	4.4%	78%	22%	34.1%	65.9%
(n=91)	*(55)*	*(32)*	*(4)*	*(142)*	*(40)*	*(31)*	*(60)*
Ctrl	54.2%	38.6%	7.2%	73.5%	26.5%	20.5%	79.5%
(n=83)	*(45)*	*(32)*	*(6)*	*(122)*	*(44)*	*(17)*	*(66)*
	χ^2=1.03,df=2, P=0.59			χ^2=0.97, df=1, P=0.32		χ^2=4.01, df=1, P=0.03	

However the SFA can be applied to datasets of any size, its power calculation could be carried out only on H1 homozygotes with the preliminary script (see Table 9 and Figure 6). Table 7 represents allele and genotype frequencies of MAPT and proportions of APOE4 carriers. No significant differences can be observed in the distribution of MAPT genotypes or allele frequencies in AD patients as compared with control individuals. The calculated frequency of the H1 allele in the AD population did not differ significantly from that in the controls. The allele frequency of APOE4 was significantly higher in the AD sample. The individuals with the combination of at least one APOE4 allele with at least one H1 allele were overrepresented in the AD sample if compared to other participants of the group (30.8%, n=28 vs. 14.5%, n=12, χ^2=6.52, df=1, P=0.011).

Table 8 represents synergy factor analysis. *SF* values are 1.02 if H1/H1 genotype and 5.42 if H1 allele is considered to act as a risk factor: neither H1 allele, nor H1/H1 genotype obtain statistically significant interaction with APOE4.

Table 8. Synergy factor analysis of MAPT and APOE4 in different constellations

MAPT H1/H1	APOE4	Controls	AD	OR	*SF, Z, p*
-	-	31	25	Reference	
+	-	36	35	1.205	
-	+	7	12	2.126	
+	+	9	19	2.618	*SF*=1.02, *Z*=0.03, *P*=0.98

Table 8 (cont'). Synergy factor analysis of MAPT and APOE4 in different constellations

MAPT H1	APOE4	Controls	AD	OR	*SF, Z, p*
-	-	2	2	Reference	
+	-	65	58	0.89	
-	+	4	2	0.5	
+	+	12	29	2.41	*SF*=5.42, *Z*=1.23, *P*=0.22

Table 9. SFA power values at different sample sizes

SF	*n*=83	*n*=87	*n*=91	*n*=100
1	0.03959	0.04648	0.05084	0.04835
1.5	0.14111	0.14048	0.14826	0.15167
2	0.25060	0.25080	0.27140	0.27718
2.5	0.36201	0.35936	0.38309	0.40357
3	0.45463	0.45478	0.48740	0.50839
3.5	0.53686	0.54299	0.57182	0.60346
4	0.60659	0.61699	0.64347	0.67547
4.5	0.67363	0.68858	0.68587	0.74047
5	0.74067	0.76017	0.72827	0.80546
5.5	0.80772	0.83176	0.77067	0.87045
6	0.87476	0.90335	0.81307	0.93545

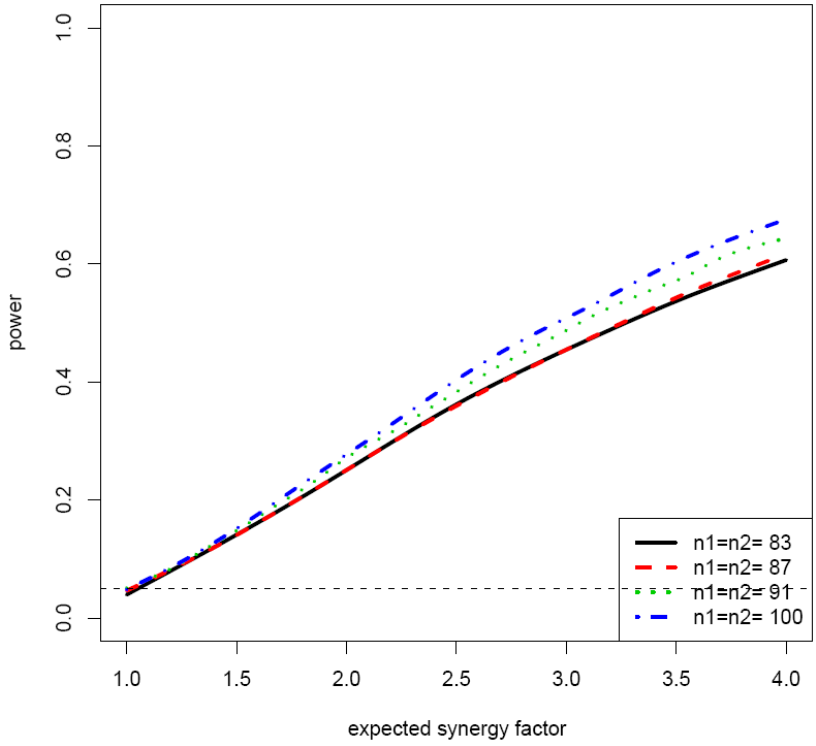

Figure 6. SFA power curves at different sample sizes

DISCUSSION

Our results indicate a different proportion of the inversion at 17q21.3 in Olah Roma as compared with Caucasian Hungarians. This study has revealed that Olah Roma, who are related to the Asian population, carry the H1 allele at a higher proportion than European Caucasian populations. This supports the notion that 17q21.3 structural variation and tau haplotypes are suitable markers for the demonstration of the degree of admixture in a well-characterized non–European population. The 24.5% H2 allele frequency in the Hungarian population accords well with the frequency of ~25% in Middle Eastern and European populations (Evans et al., 2004). The previously reported 8% of H2 allele frequency (Evans et al., 2004) in the Finnish population stands closer to the Asian genotype distribution. These results suggest that the Finnish population experienced less admixture than the population of Hungary, and the Asian descent of the latter is not detectable by this method.

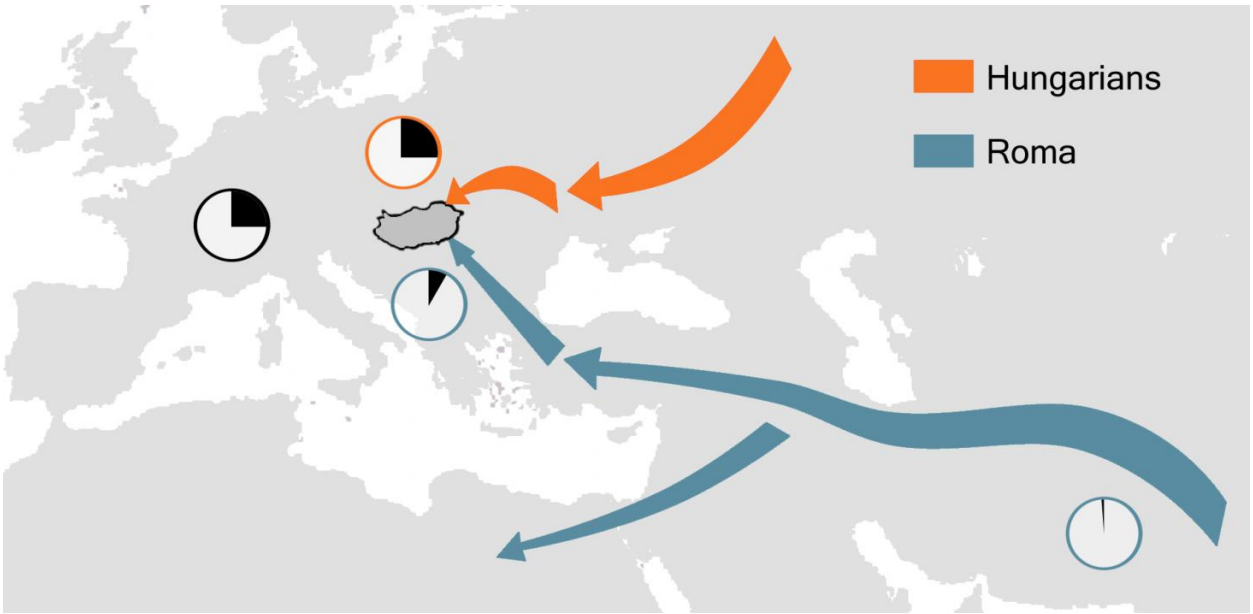

Figure 7. Routes of migration and frequency of 17q21.31 inversion

In our Roma sample, the frequency of H1 allele was lower than previous estimates from populations of Asian origin (only populations from South Pakistan were similar) (Evans et al., 2004). Lower frequency of the H1 haplotype in the Roma population may be a consequence of their coexistence for centuries and partial admixture with H2 carrier Caucasian populations. This effect is likely to have been strengthened by the fact that the Olah/Vlax metagroup traditionally tolerates marriages with non-Roma women, whereas some other Roma groups do not. The deviation from the Hardy–Weinberg equilibrium in the Roma group can be explained by the population genetic effect of their closed society structure and the higher rate of consanguineous mating.

The Roma ethnic group was ignored for centuries by Western society and medicine. The United Nations Development Programme (www.undp.org) and the Decade of Roma Inclusion 2005–2015 (www.romadecade.org) recognized the importance of medical and social studies. In the past decade, various mendelian diseases with a carrier rate of 5–15% have been identified in the Roma population (Kalaydjieva et al., 2001b), but multifactorial tauopathies have not been well described in Roma. This can be explained by their social and medical neglect and the fact that tauopathies are typically late-onset neurodegenerative diseases, although the average life expectancy of Roma is 10–15 years lower than the European standard (Sepkowitz, 2006).

Population specific inversion contributes to phenotypic variability and adaptation

The clearest example regarding an inversion's effect on phenotype can be observed in local adaptation and speciation of a plant, the yellow monkeyflower, *Mimulus guttatus*. This species exists in two ecotypes which show distinct differences on flowering time. The one which is annual is habituated to dry inlands and flower early, while the other is perennial and adapted to moist and cool weather at the coast with flowering later in the year. This ends up in premating isolation and also a postzygotic in hybrids. These phenotypic variations are attributed to an inversion which suppresses recombination in hybrids and contribute to

reproductive isolation between the forms. This is a compelling example on the local adaptation hypothesis for inversions (Kirkpatrick, 2010; Lowry & Willis, 2010).

As described earlier, the 17q21.31 inversion haplotypes show distinct effects on populations. H2 carriers are at risk to develop a microdeletion event leading to neurodevelopmental delay while in H1 homozygotes this form of replication event is not possible. Therefore, Koolen De Vries syndrome (which is accounting for 1% of mental retardations) is exclusively related to the Caucasian genome and absent from Asians. Now, first in the literature it is shown that Roma populations are at risk to develop this disorder, since owing to the admixture effect they carry the H2 allele.

On the other hand, regarding H1 related pathology the case is not so evident. It was shown that H1 carriers are under a negative selection in Europe since H2 carrier women have more children (Stefansson et al., 2005, Voight et al., 2006) and because of the possible role of H1 allele in tauopathies. Alzheimer's disease (Laws et al., 2007, Myers et al., 2005), Parkinson's disease (Skipper et al., 2004), progressive supranuclear palsy (Pittman et al., 2004), argyrophilic grain disease (Fujino et al., 2005), corticobasal degeneration (Buee and Delacourte, 1999) and the Parkinson–dementia complex of Guam (Sundar et al., 2007) are all associated with MAPT H1 in certain populations. It seems that carrying H1 allele influence disease onset via influencing gene expression or alternative splicing. Both may lead to enhanced tangle formation and the development of the disease (Avila, 2006, Caffrey et al., 2006, Hardy et al., 2006, O'Hara, 2007).

It is well known that exposure to different (that is, European) environmental factors may lead to differences in epigenetic effects on gene expression (Spielman et al., 2007). A recent study (Winkler et al., 2007) demonstrated H1/H1 genotype as an ethnically dependent risk factor of Parkinson's disease, and another one raised further remarkable suggestions on this field (Fung et al., 2005). An early work also observed association regarding tau variants and Asian versus Caucasian populations in progressive supranuclear palsy (Conrad et al., 1998). Thus, higher H1 frequency in Roma might be a risk factor of multifactorial disorders and be manifested as

an elevated susceptibility to tauopathies among the Roma population in Europe. Further investigations are needed in populations with high H1 frequency where the social and medical aspects and the average life expectancy are better.

When non-European genetic variants meet European environment

Our study focusing on previously not examined SNP and VNTR variants in Roma populations supported our findings on genetic heritage. The results revealed a statistically significant difference between Olah Roma and Hungarian populations in the distribution of SLC6A3 alleles. The frequency of the A9 allele was significantly lower whereas the occurrence of the A11 allele was significantly higher in the Olah Roma group as compared to the Hungarian population. However, the comparison of the frequencies of the A10 allele showed no significant difference. While this is the first report on SLC6A3 polymorphism in the Roma population and no data are available from North India where the Roma originate from, several other populations have been studied (Joober et al., 2000, Kang et al., 1999, Mitchell et al., 2000). Kang and co-workers summarize and compare the SLC6A3 allele frequencies of more than 1500 individuals from 30 populations in a meta-analysis (Kang et al., 1999). The observed alleles show a range from 3 to 12 repeats, but the three-, seven-, eight- and twelve-repeat alleles occurred only with very low frequency and no four-, five- or six-repeat alleles were detected. The A10 allele is the most frequent with some variation in the different populations. The second most frequent allele is A9 (Kang et al., 1999, Mitchell et al., 2000). These findings are in agreement with our results in the Olah Roma and Hungarian populations, as well as with another study investigating the SLC6A3 allele distribution in the French–Canadian population (Joober et al., 2000). See Table 6 for comparison.

The A9 allele was associated with severity of alcohol withdrawal symptoms (Sander et al., 1997) and reduced risk of tobacco smoking (Lerman et al., 1999) while the A10 allele was linked to attention deficit hyperactivity disorder (Cook et al., 1995, Gill et al., 1997) A significant genotypic effect on DAT levels was found in a large sample of healthy subjects:

the A9 carriers had a significantly higher striatal DAT availability compared to the A10/A10 homozygotes (van Dyck et al., 2005). On the other hand biological data regarding A11 allele is fragmentary so far. These observations reveal that the genetic variants of SLC6A3 show a remarkable difference which may point toward certain neuropsychiatric and addictive disorders in the Roma population. Therefore these findings should be considered once interventions programs are developed to battle high rates of alcohol and nicotine misuse in Roma populations.

The role of DRD3 Ser9Gly polymorphism is not entirely clarified, but it has been extensively investigated and a correlation was found between the Ser9 allele and the response to typical antipsychotics, and between the Gly9 allele and the response to atypical antipsychotics in schizophrenic patients (Scharfetter, 2004). Another study from our laboratory reported that Ser9Ser genotype is associated with worse therapeutic response and more severe dysfunctions in schizophrenic patients (Szekeres et al., 2004). There were no statistical differences in the occurrence of DRD3 alleles in the Olah Roma population as compared to the Hungarian population in our study. Association studies investigating European Caucasian and North Indian populations also reported data similar to the pattern observed in the Olah Roma and Hungarian populations (Jonsson et al., 1993, Prasad et al., 2008). Only a small difference within the limits of the statistical error was found between Europeans and North Indians (Prasad et al., 2008). The frequency of the Ser9 allele seems slightly lower in Olah Roma people and in the North Indian population as compared to European Caucasians, although this difference proved to be statistically non–significant.

In summary, our results provide evidences about the polymorphisms of the dopamine-related genes in a Roma population which deserves further characterization.

17q21.31 and APOE4 do not act in synergy in AD

The third study contributing to this work is a case–control study, examining the distribution of MAPT and APOE4 alleles and the combination of those in Hungarian Caucasian AD samples

and healthy controls. The results indicate that in both groups the representation of H1 haplotype accords well with the frequency of ~75% in Middle–Eastern and European populations which was determined earlier (Evans et al., 2004). In this manner MAPT H1 haplotype can not be identified as a risk factor of AD in the Hungarian population.

It should be considered why several studies found association to AD and other tauopathies (Baker et al., 1999, Fujino et al., 2005, Pittman et al., 2005, Togo et al., 2002, Zabetian et al., 2007) while ours pertain to those which disprove this association (Russ et al., 2001).

First, a possible explanation could be that only more specific subhaplotypes of H1 clade may play a major role in tauopathies. Regarding AD, recent studies indicated that the promoter polymorphism rs242557 delineates a subhaplotype (H1c) which could be the risk factor for developing AD (Myers et al., 2007). Though, there are negative replications with H1c too (Mukherjee et al., 2007). Furthermore, cumulating evidences suggest that disorders where the diagnosis is based on tau-pathology related stable biomarkers are associated to this genetic background. It is well-known that late onset AD is a complex disorder, where tau pathology is an important, but not sole contributor to disease process (Caffrey et al., 2012).

The second issue deserving interest is the interaction of major pathways and biological networks implicated in tauopathies. MAPT haplotypes do affect gene expression in tissue specific manners (de Jong et al., 2012). Recent studies indicate that hyperphosphorylated tau is overrepresented in Parkinson's disease patients with H1/H1 alleles (Kwok et al., 2005), and the activity of tau kinases which are responsible for the phosphorylation of the protein are increased in AD (Leroy et al., 2007). The kinases of tau phosphorylation and the connection with other major pathways together may constitute the genetic basis of AD. A synergistic effect between glycogen synthase kinase-3beta and tau genes was found recently (Kwok et al., 2008).

In our study, we examined MAPT haplotypes and APOE4 state as elements of converging pathways in the development of AD. This supposition was based on the fact that APOE has a broad role in AD pathology with several synergistic connections (Combarros et al., 2009) and

has influence on tau phosphorylation (Tesseur et al., 2000). Recently further evidence showed that APOE status comprises a network of connections with APP and MAPT predisposing to a molecular prodrome that result in clinical AD (Conejero-Goldberg et al., 2011).

The broadly supported finding that carrying APOE4 allele is a risk factor of AD was replicated. Nevertheless, it was shown here that APOE4 and MAPT haplotypes do not act on synergy in Alzheimer's disease in the Hungarian population. This part of the work also draws attention to the importance of validation of true epistasis. As it was discussed earlier by Combarros et al., and can be observed in this study too, a combined analysis based solely on χ^2 test could have led to a false positive façade of gene interaction. However, the conclusions that can be drawn from this study must be tempered with the limitations imposed by statistical power arising from sample size and haplotype frequency.

As tauopathies are multifactorial disorders, the role of environmental factors and epigenetic effects on the genome are also considerable (McCulloch et al., 2008). These can influence gene expression (Caffrey et al., 2006), alternative splicing (Andreadis, 2005) or both and may lead to enhanced tangle formation and disease development. The above mentioned population genetic effect is particularly interesting since the frequency of tauopathies is not elevated in East Asians or Africans where the H1 haplotype is almost obligate.

CONCLUSIONS

This work encompasses studies which were born together with the research field of genomic inversion behind phenotypic variance. An outlook to populations characterized by a founder effect can be a medical researchers' tool to shine more light on complex nature of multifactorial disorders. In Hungary, studying genetic variants related to mendelian disorders already shown success within Roma populations. Their unique genomic heritage could provide medical information to help studying structural variants. This project demonstrated that as the result of genetic admixture, 17q21.31 H2 haplotype appeared in Roma populations, and spread to ~10% even in the closed societies. Roma population therefore carries the Koolen De Vries syndrome associated genomic variant and Roma individuals are at risk to develop the disease. In our second study we demonstrated that dopamine transporter VNTR variants which were shown to be associated to addictive behavior (among other neuropsychiatric disturbances) are present with notably different frequencies in Hungarians and Roma. This result may have biological implications regarding therapeutic interventions for addictive disorders in Roma populations. Our third study examined H1 haplotype in Hungarian Alzheimer's disease patients and studied genetic interactions with APOE4. H1 is not associated to AD in Hungarian populations while APOE4 is confirmed again. Together with other findings this result suggests the notion that 17q21.31 inversion H1 haplotype should be further studied in disorders where tauopathy is the major pathomechanism.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Zoltán Janka for his continuous support during my research and clinical activities. He gave the possibility to carry out these projects and opened space to a broad range of scientific interests outside the fields of genetics.

I am much obliged to István Raskó and Ágnes Czibula whose attitude to science shaped my future.

I would like to give credit to all participants and co-authors of this project. I wish to thank Szatmár Horváth for the fruitful discussions and his contribution to the publications, Professor János Kálmán for providing the samples of the Alzheimer's disease research group and Bálint Andó for the cooperation in the past years.

As a researcher I am embedded in clinical background. My views as a clinician were determined by working on the "4/B Unit" with Zoltán Ambrus Kovács, György Szekeres, István Szendi, Szatmár Horváth, Csongor Cimmer and Gábor Csifcsák.

Last, but not least I would like to thank my wonderful family and friends for their endless support.

My doctoral studies were supported by SCHIZO-08 project. This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP-4.2.4.A/ 2-11/1-2012-0001 'National Excellence Program'.

REFERENCES

ALMOS, P. Z., HORVATH, S., CZIBULA, A., RASKO, I., SIPOS, B., BIHARI, P., BERES, J., JUHASZ, A., JANKA, Z. & KALMAN, J. 2008. H1 tau haplotype-related genomic variation at 17q21.3 as an Asian heritage of the European Roma population. *Heredity (Edinb),* 101**,** 416-9.

ANDREADIS, A. 2005. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. *Biochim Biophys Acta,* 1739**,** 91-103.

ASSAL, S., SUSANSZKY, E. & CZEIZEL, A. 1991. High consanguinity rate in Hungarian gipsy communities. *Acta Paediatr Hung,* 31**,** 299-304.

AVILA, J. 2006. Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett,* 580**,** 2922-7.

BAKER, M., 2012. Structural variation: the genome's hidden architecture. *Nat Methods*, 2, 133-137

BAKER, M., LITVAN, I., HOULDEN, H., ADAMSON, J., DICKSON, D., PEREZ-TUR, J., HARDY, J., LYNCH, T., BIGIO, E. & HUTTON, M. 1999. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum Mol Genet,* 8**,** 711-5.

BOUTHENET, M. L., SOUIL, E., MARTRES, M. P., SOKOLOFF, P., GIROS, B. & SCHWARTZ, J. C. 1991. Localization of dopamine D3 receptor mRNA in the rat brain using in situ hybridization histochemistry: comparison with dopamine D2 receptor mRNA. *Brain Res,* 564**,** 203-19.

BUEE, L. & DELACOURTE, A. 1999. Comparative biochemistry of tau in progressive supranuclear palsy, corticobasal degeneration, FTDP-17 and Pick's disease. *Brain Pathol,* 9**,** 681-93.

CAFFREY, T. M., JOACHIM, C., PARACCHINI, S., ESIRI, M. M. & WADE-MARTINS, R. 2006. Haplotype-specific expression of exon 10 at the human MAPT locus. *Hum Mol Genet,* 15**,** 3529-37.

CAFFREY, T. M. & WADE-MARTINS, R. 2012. The role of MAPT sequence variation in mechanisms of disease susceptibility. *Biochem Soc Trans,* 40, 687–692.

CANU, E., BOCCARDI, M., GHIDONI, R., BENUSSI, L., TESTA, C., PIEVANI, M., BONETTI, M., BINETTI, G. & FRISONI, G. B. 2009. H1 haplotype of the MAPT gene is associated with lower regional gray matter volume in healthy carriers. *Eur J Hum Genet,* 17**,** 287-94.

CAVALLI-SFORZA, L. L., MENOZZI, P. & PIAZZA, A. 1994. *The history and geography of human genes,* Princeton, N.J., Princeton University Press.

COMBARROS, O., CORTINA-BORJA, M., SMITH, A. D. & LEHMANN, D. J. 2009. Epistasis in sporadic Alzheimer's disease. *Neurobiol Aging,* 30**,** 1333-49.

CONEJERO-GOLDBERG, C., HYDE, T. M., CHEN, S., DRESES-WERRINGLOER, U., HERMANN, M. M., KLEINMAN, J. E., DAVIES, P. & GOLDBERG T. E. 2011. Molecular signatures in post-mortem brain tissue of younger individuals at high risk for Alzheimer's disease as based on APOE genotype. *Mol Psychiatry*, 16, 836-847.

CONRAD, C., AMANO, N., ANDREADIS, A., XIA, Y., NAMEKATAF, K., OYAMA, F., IKEDA, K., WAKABAYASHI, K., TAKAHASHI, H., THAL, L. J., KATZMAN, R., SHACKELFORD, D. A., MATSUSHITA, M., MASLIAH, E. & SAWA, A. 1998. Differences in a dinucleotide repeat polymorphism in the tau gene between Caucasian and Japanese populations: implication for progressive supranuclear palsy. *Neurosci Lett,* 250**,** 135-7.

COOK, E. H., JR., STEIN, M. A., KRASOWSKI, M. D., COX, N. J., OLKON, D. M., KIEFFER, J. E. & LEVENTHAL, B. L. 1995. Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet,* 56**,** 993-8.

CORDER, E. H., HUANG, R., CATHCART, H. M., LANHAM, I. S., PARKER, G. R., CHENG, D., SMITH, S. & PODUSLO, S. E. 2006. Membership in genetic groups predicts Alzheimer disease. *Rejuvenation Res,* 9**,** 89-93.

CORTINA-BORJA, M., SMITH, A. D., COMBARROS, O. & LEHMANN, D. J. 2009. The synergy factor: a statistic to measure interactions in complex diseases. *BMC Res Notes,* 2**,** 105.

COUSINS, D. A., BUTTS, K. & YOUNG, A. H. 2009. The role of dopamine in bipolar disorder. *Bipolar Disord,* 11**,** 787-806.

DAVIES, K. E. 1993. *Human genetic disease analysis : a practical approach,* Oxford, New York, IRL Press.

DE JONG, S., CHEPELEC, I., JANSON, E., STRENGMAN, E., VAN DEN BERG, L. H., VELDINK, J. H. & OPHOFF R. A. 2012. Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics*, 13, 458.

DONELLY, M. P., PASCHOU, P., GRIGORENKO, E., GURWITZ, D., MEHDI, S. Q., KAJUNA, S. L., BARTA, C., KUNGULILO, S., KAROMA, N. J., LU, R. B., ZHUKOVA, O. V., KIM, J. J., COMAS, D., SINISCLACO, M., NEW, M., LI, P., LI, H., MANOLOPOULOS, V. G., SPEED, W. C., RAJEEVAN, H., PAKSTIS, A. J., KIDD, J. R. & KIDD, K. K. 2010. The distribution and most recent common ancestor of the 17q21 inversion in humans. *Am J Hum Genet,* 86, 161-71.

EVANS, W., FUNG, H. C., STEELE, J., EEROLA, J., TIENARI, P., PITTMAN, A., SILVA, R., MYERS, A., VRIEZE, F. W., SINGLETON, A. & HARDY, J. 2004. The tau H2 haplotype is almost exclusively Caucasian in origin. *Neurosci Lett,* 369**,** 183-5.

FEUK, L. 2010. Inversion variants in the human genome: role in disease and genome architecture. *Genome Med,* 2, 11.

FEUK, L., CARSON, A. R. & SCHERER, S. W. 2006. Structural variation in the human genome. *Nat Rev Genet,* 7**,** 85-97.

FEUK, L., MACDONALD, J. R., TANG, T., CARSON, A. R., LI, M., RAO, G., KHAJA, R. & SCHERER S. W. 2005. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *Plos Genet,* 1, e56

FUJINO, Y., WANG, D. S., THOMAS, N., ESPINOZA, M., DAVIES, P. & DICKSON, D. W. 2005. Increased frequency of argyrophilic grain disease in Alzheimer disease with 4R tau-specific immunohistochemistry. *J Neuropathol Exp Neurol,* 64**,** 209-14.

FUNG, H. C., EVANS, J., EVANS, W., DUCKWORTH, J., PITTMAN, A., DE SILVA, R., MYERS, A. & HARDY, J. 2005. The architecture of the tau haplotype block in different ethnicities. *Neurosci Lett,* 377**,** 81-4.

GHIDONI, R., SIGNORINI, S., BARBIERO, L., SINA, E., COMINELLI, P., VILLA, A., BENUSSI, L. & BINETTI, G. 2006. The H2 MAPT haplotype is associated with familial frontotemporal dementia. *Neurobiol Dis,* 22**,** 357-62.

GILL, M., DALY, G., HERON, S., HAWI, Z. & FITZGERALD, M. 1997. Confirmation of association between attention deficit hyperactivity disorder and a dopamine transporter polymorphism. *Mol Psychiatry,* 2**,** 311-3.

GRESHAM, D., MORAR, B., UNDERHILL, P. A., PASSARINO, G., LIN, A. A., WISE, C., ANGELICHEVA, D., CALAFELL, F., OEFNER, P. J., SHEN, P., TOURNEV, I., DE PABLO, R., KUCINSKAS, V., PEREZ-

LEZAUN, A., MARUSHIAKOVA, E., POPOV, V. & KALAYDJIEVA, L. 2001. Origins and divergence of the Roma (Roma). *Am J Hum Genet,* 69**,** 1314-31.

GU, S., PAKSTIS, A. J., LI, H., SPEED, W. C., KIDD, J. R. & KIDD, K. K. 2007. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. *Eur J Hum Genet,* 15**,** 302-12.

HALLIDAY, G. M. & MCCANN, H. 2010. The progression of pathology in Parkinson's disease. *Ann N Y Acad Sci,* 1184**,** 188-95.

HARDY, J., PITTMAN, A., MYERS, A., FUNG, H. C., DE SILVA, R. & DUCKWORTH, J. 2006. Tangle diseases and the tau haplotypes. *Alzheimer Dis Assoc Disord,* 20**,** 60-2.

HARDY, J., PITTMAN, A., MYERS, A., GWINN-HARDY, K., FUNG, H. C., DE SILVA, R., HUTTON, M. & DUCKWORTH, J. 2005. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. *Biochem Soc Trans,* 33**,** 582-5.

HOENICKA, J., ARAGUES, M., PONCE, G., RODRIGUEZ-JIMENEZ, R., JIMENEZ-ARRIERO, M. A. & PALOMO, T. 2007. From dopaminergic genes to psychiatric disorders. *Neurotox Res,* 11**,** 61-72.

JONSSON, E., LANNFELT, L., SOKOLOFF, P., SCHWARTZ, J. C. & SEDVALL, G. 1993. Lack of association between schizophrenia and alleles in the dopamine D3 receptor gene. *Acta Psychiatr Scand,* 87**,** 345-9.

JOOBER, R., TOULOUSE, A., BENKELFAT, C., LAL, S., BLOOM, D., LABELLE, A., LALONDE, P., TURECKI, G. & ROULEAU, G. A. 2000. DRD3 and DAT1 genes in schizophrenia: an association study. *J Psychiatr Res,* 34**,** 285-91.

KALAYDJIEVA, L., CALAFELL, F., JOBLING, M. A., ANGELICHEVA, D., DE KNIJFF, P., ROSSER, Z. H., HURLES, M. E., UNDERHILL, P., TOURNEV, I., MARUSHIAKOVA, E. & POPOV, V. 2001a. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet,* 9**,** 97-104.

KALAYDJIEVA, L., GRESHAM, D. & CALAFELL, F. 2001b. Genetic studies of the Roma (Roma): a review. *BMC Med Genet,* 2**,** 5.

KALAYDJIEVA, L., MORAR, B., CHAIX, R. & TANG, H. 2005. A newly discovered founder population: the Roma/Roma. *Bioessays,* 27**,** 1084-94.

KALINDERI, K., FIDANI, L. & BOSTANTJOPOULOU, S. 2009. From 1997 to 2007: a decade journey through the H1 haplotype on 17q21 chromosome. *Parkinsonism Relat Disord,* 15**,** 2-5.

KALMAN, J., JUHASZ, A., CSASZAR, A., KANKA, A., MAGLOCZKY, E., BENCSIK, K., JANKA, Z. & RASKO, I. 1997. Apolipoprotein E allele frequencies in patients with late-onset sporadic Alzheimer's dementia in Hungary. *Acta Neurol Scand,* 95**,** 56-9.

KANG, A. M., PALMATIER, M. A. & KIDD, K. K. 1999. Global variation of a 40-bp VNTR in the 3'-untranslated region of the dopamine transporter gene (SLC6A3). *Biol Psychiatry,* 46**,** 151-60.

KIRKPATRICK, M. 2010. How and why chromosome inversions evolve. *PLOS Biol,* 9, e1000501.

KWOK, J. B., HALLUPP, M., LOY, C. T., CHAN, D. K., WOO, J., MELLICK, G. D., BUCHANAN, D. D., SILBURN, P. A., HALLIDAY, G. M. & SCHOFIELD, P. R. 2005. GSK3B polymorphisms alter transcription and splicing in Parkinson's disease. *Ann Neurol,* 58**,** 829-39.

KWOK, J. B., LOY, C. T., HAMILTON, G., LAU, E., HALLUPP, M., WILLIAMS, J., OWEN, M. J., BROE, G. A., TANG, N., LAM, L., POWELL, J. F., LOVESTONE, S. & SCHOFIELD, P. R. 2008. Glycogen synthase kinase-3beta and tau genes interact in Alzheimer's disease. *Ann Neurol,* 64**,** 446-54.

LANNFELT, L., SOKOLOFF, P., MARTES M. P., PILON, C., GIROS, B., JÖNSSON, E., SEDVALL, G. & SCHWARTZ, J-C. 1992. Amino acid substitution in the dopamine D3 receptor as a useful polymorphism for investigating psychiatric disorders. *Psychiatr Genet* 2, 249–256.

LAWS, S. M., FRIEDRICH, P., DIEHL-SCHMID, J., MULLER, J., EISELE, T., BAUML, J., FORSTL, H., KURZ, A. & RIEMENSCHNEIDER, M. 2007. Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease. *Mol Psychiatry,* 12**,** 510-7.

LERMAN, C., CAPORASO, N. E., AUDRAIN, J., MAIN, D., BOWMAN, E. D., LOCKSHIN, B., BOYD, N. R. & SHIELDS, P. G. 1999. Evidence suggesting the role of specific genetic factors in cigarette smoking. *Health Psychol,* 18**,** 14-20.

LEROY, K., YILMAZ, Z. & BRION, J. P. 2007. Increased level of active GSK-3beta in Alzheimer's disease and accumulation in argyrophilic grains and in neurones at different stages of neurofibrillary degeneration. *Neuropathol Appl Neurobiol,* 33**,** 43-55.

LODGE, D. J. & GRACE, A. A. 2011. Developmental pathology, dopamine, stress and schizophrenia. *Int J Dev Neurosci,* 29**,** 207-13.

LOWRY, D. B. & WILLIS, J. H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLOS Biol,* 8, e10005000.

LUPSKI, J. R. 1998. Genomic disorders: structural features of genome can lead to DNA rearrangements and human disease traits. *Trends Genet,* 14, 417-422.

MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R., CHAKRAVARTI, A., CHO, J. H., GUTTMACHER, A. E., KONG, A., KRUGLYAK, L., MARDIS, E., ROTIMI, C. N., SLATKIN, M., VALLE, D., WHITTEMORE, A. S., BOEHNKE, M., CLARK, A. G., EICHLER, E. E., GIBSON, G., HAINES, J. L., MACKAY, T. F., MCCAROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.

MCCULLOCH, C. C., KAY, D. M., FACTOR, S. A., SAMII, A., NUTT, J. G., HIGGINS, D. S., GRIFFITH, A., ROBERTS, J. W., LEIS, B. C., MONTIMURRO, J. S., ZABETIAN, C. P. & PAYAMI, H. 2008. Exploring gene-environment interactions in Parkinson's disease. *Hum Genet,* 123**,** 257-65.

MENDIZABAL, I., LAO, O., MARIGORTA, U.M., WOLLSTEIN, A., GUSMAO, L., FERAK, V., IOANA, M., JORDANOVA, A., KANEVA, R., KOUVATSI, A., KUCINSKAS, V., MAKUKH, H., METSPALU, A., NETEA, M. G., DE PABLO, R., PAMJAV, H., RADOJKOVIC, D., ROLLESTON, S. J., SERTIC, J., MACEK, M. JR., COMAS, D. & KAYSER, M. 2012. Reconstructing the population history of European Romani from genome-wide data. *Curr Biol,* 24, 2342-2349

MITCHELL, R. J., HOWLETT, S., EARL, L., WHITE, N. G., MCCOMB, J., SCHANFIELD, M. S., BRICENO, I., PAPIHA, S. S., OSIPOVA, L., LIVSHITS, G., LEONARD, W. R. & CRAWFORD, M. H. 2000. Distribution of the 3' VNTR polymorphism in the human dopamine transporter gene in world populations. *Hum Biol,* 72**,** 295-304.

MOORJANI, P., PATTERSON, N., LOH, P-R., LIPSON, M., KISFALI, P., MELEGH, B. I., BONIN, M., KADASI, L., RIESS, O., BERGER, B., REICH, D. & MELEGH, B. 2013. Reconstructing Roma History from Genome-Wide Data. *PLoS ONE,* 3, e58633

MORAR, B., GRESHAM, D., ANGELICHEVA, D., TOURNEV, I., GOODING, R., GUERGUELTCHEVA, V., SCHMIDT, C., ABICHT, A., LOCHMULLER, H., TORDAI, A., KALMAR, L., NAGY, M., KARCAGI, V., JEANPIERRE, M., HERCZEGFALVI, A., BEESON, D., VENKATARAMAN, V., WARWICK CARTER, K.,

REEVE, J., DE PABLO, R., KUCINSKAS, V. & KALAYDJIEVA, L. 2004. Mutation history of the roma/Roma. *Am J Hum Genet,* 75**,** 596-609.

MUKHERJEE, O., KAUWE, J. S., MAYO, K., MORRIS, J. C. & GOATE, A. M. 2007. Haplotype-based association analysis of the MAPT locus in late onset Alzheimer's disease. *BMC Genet,* 8**,** 3.

MYERS, A. J., KALEEM, M., MARLOWE, L., PITTMAN, A. M., LEES, A. J., FUNG, H. C., DUCKWORTH, J., LEUNG, D., GIBSON, A., MORRIS, C. M., DE SILVA, R. & HARDY, J. 2005. The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. *Hum Mol Genet,* 14**,** 2399-404.

MYERS, A. J., PITTMAN, A. M., ZHAO, A. S., ROHRER, K., KALEEM, M., MARLOWE, L., LEES, A., LEUNG, D., MCKEITH, I. G., PERRY, R. H., MORRIS, C. M., TROJANOWSKI, J. Q., CLARK, C., KARLAWISH, J., ARNOLD, S., FORMAN, M. S., VAN DEERLIN, V., DE SILVA, R. & HARDY, J. 2007. The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts. *Neurobiol Dis,* 25**,** 561-70.

O'HARA, R. B. 2007. Human expression patterns: genetic differences between populations. *Heredity (Edinb),* 98**,** 245-6.

PITTMAN, A. M., MYERS, A. J., ABOU-SLEIMAN, P., FUNG, H. C., KALEEM, M., MARLOWE, L., DUCKWORTH, J., LEUNG, D., WILLIAMS, D., KILFORD, L., THOMAS, N., MORRIS, C. M., DICKSON, D., WOOD, N. W., HARDY, J., LEES, A. J. & DE SILVA, R. 2005. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet,* 42**,** 837-46.

PITTMAN, A. M., MYERS, A. J., DUCKWORTH, J., BRYDEN, L., HANSON, M., ABOU-SLEIMAN, P., WOOD, N. W., HARDY, J., LEES, A. & DE SILVA, R. 2004. The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum Mol Genet,* 13**,** 1267-74.

PRASAD, P., KUMAR, K. M., AMMINI, A. C., GUPTA, A., GUPTA, R. & THELMA, B. K. 2008. Association of dopaminergic pathway gene polymorphisms with chronic renal insufficiency among Asian Indians with type-2 diabetes. *BMC Genet,* 9**,** 26.

RAPHAEL, B. J. 2012. Chapter 6: Structural variation and medical genomics. *PLOS Comput Biol*, 12, e1002821.

R DEVELOPMENT CORE TEAM 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

RAI, N., CHAUBEY, G., TAMANG, R., PATHAK, A.K., SINGH, V.K., RANI, D. S., ANUGULA, S., YADAV, B. K., SINGH, A., SRINIVASAGAN, R., YADAV, A., KASHYAP, M., NARVARIYA, S., REDDY, A. G., VAN DRIEM, G., UNDERHILL, P. A., VILLEMS, R., KIVISILD, T., SINGH, L. & THANGARAJ, K. 2012. The phylogeography of Y-chromosome haplogroup H1a1a-M82 reveals the likely Indian origin of the European Romani populations. *PLOS ONE* 11, e48477.

RUSS, C., POWELL, J. F., ZHAO, J., BAKER, M., HUTTON, M., CRAWFORD, F., MULLAN, M., ROKS, G., CRUTS, M. & LOVESTONE, S. 2001. The microtubule associated protein Tau gene and Alzheimer's disease--an association study and meta-analysis. *Neurosci Lett,* 314**,** 92-6.

SANDER, T., HARMS, H., PODSCHUS, J., FINCKH, U., NICKEL, B., ROLFS, A., ROMMELSPACHER, H. & SCHMIDT, L. G. 1997. Allelic association of a dopamine transporter gene polymorphism in alcohol dependence with withdrawal seizures or delirium. *Biol Psychiatry,* 41**,** 299-304.

SANO, A., KONDOH, K., KAKIMOTO, Y. & KONDO, I. 1993. A 40-nucleotide repeat polymorphism in the human dopamine transporter gene. *Hum Genet,* 91**,** 405-6.

SCHARFETTER, J. 2004. Pharmacogenetics of dopamine receptors and response to antipsychotic drugs in schizophrenia--an update. *Pharmacogenomics,* 5**,** 691-8.

SELDIN, M. F., PASANIUC, B. & PRICE, A. L. 2011 New approaches to disease mapping in admixed populations. *Nat Rev Genet*, 12, 523–528.

SEMINO, O., PASSARINO, G., QUINTANA-MURCI, L., LIU, A., BERES, J., CZEIZEL, A. & SANTACHIARA-BENERECETTI, A. S. 2000. MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. *Eur J Hum Genet,* 8**,** 339-46.

SEPKOWITZ, K. A. 2006. Health of the world's Roma population. *Lancet,* 367**,** 1707-8.

SIPEKY, C., CSONGEI, V., JAROMI, L., SAFRANY, E., POLGAR, N., LAKNER, L., SZABO, M., TAKACS, I. & MELEGH, B. 2009. Vitamin K epoxide reductase complex 1 (VKORC1) haplotypes in average Hungarian and in Roma population samples. *Pharmacogenomics*, 6, 1025-32.

SIPEKY, C., WEBER, A., SZABO, M., MELEGH, B. I., JANICSEK, I., TARLOS, G., SZABO, I., SUMEGI, K. & MELEGH, B. 2013. High prevalence of CYP2C19*2 allele in Roma samples: study on Roma and Hungarian population samples with review of the literature, *Mol Biol Rep,* 40, 4727-4735

SHARP, A. J., CHENG, Z. & EICHLER, E. E. 2006. Structural Variation of the Human Genome. *Annu Rev Genomics Hum Genet*, 7, 407–42.

SHAW-SMITH, C., PITTMAN, A. M., WILLATT, L., MARTIN, H., RICKMAN, L., GRIBBLE, S., CURLEY, R., CUMMING, S., DUNN, C., KALAITZOPOULOS, D., PORTER, K., PRIGMORE, E., KREPISCHI-SANTOS, A. C., VARELA, M. C., KOIFFMANN, C. P., LEES, A. J., ROSENBERG, C., FIRTH, H. V., DE SILVA, R. & CARTER, N. P. 2006. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet,* 38**,** 1032-7.

SKIPPER, L., WILKES, K., TOFT, M., BAKER, M., LINCOLN, S., HULIHAN, M., ROSS, O. A., HUTTON, M., AASLY, J. & FARRER, M. 2004. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. *Am J Hum Genet,* 75**,** 669-77.

SPIELMAN, R. S., BASTONE, L. A., BURDICK, J. T., MORLEY, M., EWENS, W. J. & CHEUNG, V. G. 2007. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet,* 39**,** 226-31.

STEFANSSON, H., HELGASON, A., THORLEIFSSON, G., STEINTHORSDOTTIR, V., MASSON, G., BARNARD, J., BAKER, A., JONASDOTTIR, A., INGASON, A., GUDNADOTTIR, V. G., DESNICA, N., HICKS, A., GYLFASON, A., GUDBJARTSSON, D. F., JONSDOTTIR, G. M., SAINZ, J., AGNARSSON, K., BIRGISDOTTIR, B., GHOSH, S., OLAFSDOTTIR, A., CAZIER, J. B., KRISTJANSSON, K., FRIGGE, M. L., THORGEIRSSON, T. E., GULCHER, J. R., KONG, A. & STEFANSSON, K. 2005. A common inversion under selection in Europeans. *Nat Genet,* 37**,** 129-37.

STEINBERG, K. M., ANTONACCI, F., SUDMANT, P. H., KIDD, J. M., CAMPBELL, C. D., VIVES, L., MALIG, M., SCHEINFELDT, L., BEGGS, W., IBRAHIM, M., LEMA, G., NYAMBO T. B., OMAR, S. A., BODO J-M., FROMENT, A., DONELLY, M. P., KIDD, K. K., TISHKOFF, S. A. & EICHLER, E. E. 2013. Structural Diversity and African Origin of the 17q21.31 Inversion Polymorphism *Nat Genet,* 44, 8, 872–880.

SULLIVAN, P. F., DALY, M. J. & O'DONOVAN, M. 2012. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Rev Genet* 13, 537-551.

SUNDAR, P. D., YU, C. E., SIEH, W., STEINBART, E., GARRUTO, R. M., OYANAGI, K., CRAIG, U. K., BIRD, T. D., WIJSMAN, E. M., GALASKO, D. R. & SCHELLENBERG, G. D. 2007. Two sites in the MAPT region confer genetic risk for Guam ALS/PDC and dementia. *Hum Mol Genet,* 16**,** 295-306.

SZEKERES, G., KERI, S., JUHASZ, A., RIMANOCZY, A., SZENDI, I., CZIMMER, C. & JANKA, Z. 2004. Role of dopamine D3 receptor (DRD3) and dopamine transporter (DAT) polymorphism in cognitive

dysfunctions and therapeutic response to atypical antipsychotics in patients with schizophrenia. *Am J Med Genet B Neuropsychiatr Genet,* 124B**,** 1-5.

TESSEUR, I., VAN DORPE, J., SPITTAELS, K., VAN DEN HAUTE, C., MOECHARS, D. & VAN LEUVEN, F. 2000. Expression of human apolipoprotein E4 in neurons causes hyperphosphorylation of protein tau in the brains of transgenic mice. *Am J Pathol,* 156**,** 951-64.

TOGO, T., SAHARA, N., YEN, S. H., COOKSON, N., ISHIZAWA, T., HUTTON, M., DE SILVA, R., LEES, A. & DICKSON, D. W. 2002. Argyrophilic grain disease is a sporadic 4-repeat tauopathy. *J Neuropathol Exp Neurol,* 61**,** 547-56.

VAN DYCK, C. H., MALISON, R. T., JACOBSEN, L. K., SEIBYL, J. P., STALEY, J. K., LARUELLE, M., BALDWIN, R. M., INNIS, R. B. & GELERNTER, J. 2005. Increased dopamine transporter availability associated with the 9-repeat allele of the SLC6A3 gene. *J Nucl Med,* 46**,** 745-51.

VERPILLAT, P., CAMUZAT, A., HANNEQUIN, D., THOMAS-ANTERION, C., PUEL, M., BELLIARD, S., DUBOIS, B., DIDIC, M., MICHEL, B. F., LACOMBLEZ, L., MOREAUD, O., SELLAL, F., GOLFIER, V., CAMPION, D., CLERGET-DARPOUX, F. & BRICE, A. 2002. Association between the extended tau haplotype and frontotemporal dementia. *Arch Neurol,* 59**,** 935-9.

VOIGHT, B. F., KUDARAVALLI, S., WEN, X. & PRITCHARD, J. K. 2006. A map of recent positive selection in the human genome. *PLoS Biol,* 4**,** e72.

WARNES, G., Gorjanc, G., Leisch, F. & MAN., M. 2008. Genetics: Population Genetics, R package version 1.3.3.

WINKLER, S., KONIG, I. R., LOHMANN-HEDRICH, K., VIEREGGE, P., KOSTIC, V. & KLEIN, C. 2007. Role of ethnicity on the association of MAPT H1 haplotypes and subhaplotypes in Parkinson's disease. *Eur J Hum Genet,* 15**,** 1163-8.

YANG, J. J., CHENG, C., DEVIDAS, M., CAO, X., FAN, Y., CAMPANA, D., YANG, W., NEALE, G., COX, N., SCHEET, P., BOROWITZ, M. J., WINICK, N. J., MARTIN, P. L., WILLMAN, C. L., BOWMAN, W. P., CAMITTA, B. M., CARROLL, A., REAMAN, G. H.,CARROLL, W. L., LOH, M., HUNGER, S. P.,PUI, C-H., EVANS, W. E. & RELLING, M. V. 2011 Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nature Genet*; 43, 237-241.

ZABETIAN, C. P., HUTTER, C. M., FACTOR, S. A., NUTT, J. G., HIGGINS, D. S., GRIFFITH, A., ROBERTS, J. W., LEIS, B. C., KAY, D. M., YEAROUT, D., MONTIMURRO, J. S., EDWARDS, K. L., SAMII, A. & PAYAMI, H. 2007. Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann Neurol,* 62**,** 137-44.

ZODY, M. C., JIANG, Z., FUNG, H. C., ANTONACCI, F., HILLIER, L. W., CARDONE, M. F., GRAVES, T. A., KIDD, J. M., CHENG, Z., ABOUELLEIL, A., CHEN, L., WALLIS, J., GLASSCOCK, J., WILSON, R. K., REILY, A. D., DUCKWORTH, J., VENTURA, M., HARDY, J., WARREN, W. C. & EICHLER, E. E. 2008. Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat Genet,* 40**,** 1076-83.

APPENDIX

I. Péter Zoltán Álmos, Szatmár Horváth, Ágnes Czibula, István Raskó, Botond Sipos, Péter Bihari, Judit Béres, Anna Juhász, Zoltán Janka, János Kálmán
H1 tau haplotype-related genomic variation at 17q21.3 as an Asian heritage of the European Gypsy population
Heredity, 2008; 5:416-9

II. Ágnes Fehér, Anna Juhász, Ágnes Rimanóczy, Péter Álmos, Judit Béres, Zoltán Janka, János Kálmán
Dopamine metabolism-related gene polymorphisms in Roma (Gypsy) and Hungarian populations
Journal of Genetics, 2011; 90:e72-5

III. Péter Zoltán Álmos, Szatmár Horváth, Ágnes Czibula, István Raskó, Nóra Domján, Anna Juhász, Zoltán Janka, János Kálmán
Tau haplotypes and APOE4 do not act in synergy on Alzheimer's disease
Psychiatry Research, 2011; 186:448-50