Novel polymorphic *AluYb8* insertion in the *WNK1* gene is associated with blood pressure variation in Europeans

Running title: Polymorphic *AluYb8* in *WNK1*

Margus Putku[1], Katrin Kepp[1], Elin Org[1], Siim Sõber[1], David Comas[2], Margus Viigimaa[3], Gudrun Veldre[1,4], Piret Kelgo[1], Pille Hallast[1], Neeme Tõnisson[1], Peeter Juhanson[1], Elza Khusnutdinova[5], Andreas Kindmark[6], Victor Kozich[7], Maris Laan[1]*

[1]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

[2]Institute of Evolutionary Biology (UPF-CSIC), CEXS-UPF-PRBB, Universitat Pompeu Fabra, Barcelona, Spain

[3]Centre of Cardiology, North Estonia Medical Centre, Tallinn, Estonia

[4]Department of Cardiology, University of Tartu, Tartu, Estonia

[5]Institute of Biochemistry and Genetics, Ufa Science Center, Russian Academy of Sciences, Ufa, Bashkortostan, Russia,

[6] Department of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden

[7]Institute of Inherited Metabolic Diseases, Charles University – First Faculty of Medicine, Prague, Czech Republic,

* Correspondence and reprint requests to: Maris Laan, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia; Telephone: +372-7375008, Fax: +372-7420286, Email: maris.laan@ut.ee

ABSTRACT (250 words): Mutations in *WNK1* and *WNK4* cause monogenetically inherited hypertension, the Gordon syndrome. The coded proteins contribute to the regulation of renal salt homeostasis, and thereby to the determination of blood pressure (BP). We have conducted a polymorphism screening in *WNK1* and *WNK4* conserved non-coding regions and identified an undescribed polymorphic *AluYb8* insertion in *WNK1* intron 10. Screening in primates revealed that this *Alu*-insertion has probably occurred in human lineage. Genotyping in 22 populations from Europe, Asia and Africa (854 individuals) indicated an expansion of the *WNK1 AluYb8* bearing chromosomes out-of-Africa. The allele frequency of this *Alu*-insertion in Sub-Saharan Africa was ~3.3 times lower than in other studied populations (4.8% versus 15.8%; P=9.7x10^{-9}). The carrier-status of the *WNK1 AluYb8* insertion was explored for the association with cardiovascular traits and the effect on gene expression profile. The *Alu*-insertion was significantly associated with SBP and DBP (P=0.0126 and P=0.0304, respectively; HYPEST study, n=1211). Gender-specific analysis revealed a more pronounced effect among women (SBP: P=0.013; DBP: P=0.006). In both, essential hypertension (17.7%) and CAD (17.23%) patients the allele frequency of the *AluYb8* insertion was higher compared to controls (14.51% and 15.32%, respectively). In leucocytes, the expression of *WNK1* splice-form excluding exon 11 was reduced in *AluYb8* carriers compared to wildtype subjects ($P\leq$0.002). Interestingly, the expression of the full-length *WNK1* transcript was up-regulated among *AluYb8* homozygous carriers (P=0.036) and down-regulated among heterozgotes (P=0.006). In summary, the identified *WNK1 AluYb8* insertion may have a functional effect on *WNK1* transcription profile and consequently on human BP determination.

KEY WORDS: *WNK1*; polymorphism screening; polymorphic *AluYb8*; blood pressure and cardiovascular disease; gene transcription profile; alternative splicing

Introduction

Essential hypertension is a complex disease promoted by an unfavorable combination of a person's life-style and heritable factors. It is a significant health risk leading to other cardiovascular and renal diseases. Genetic studies of monogenic, Mendelian forms of hypo- and hypertension have identified ~20 rare mutations in blood pressure regulating genes with a strong effect on the phenotype (Lifton, et al., 2001; Vehaskari, 2007). Although these rare mutations do not explain blood pressure variation in general population, the identified genes are promising targets for functional, physiological and genetic studies of essential hypertension (Ji, et al., 2008; REF).

Serine/threonine protein kinase family members WNK1 and WNK4 (With No K (lysine); Xu, et al., 2000) are involved in the development of a Mendelian form of hypertension, pseudohypoaldosteronism type II or the Gordon syndrome (Wilson, et al., 2001; Wilson, et al., 2003). The syndrome is caused either by a large deletion (two identified variants: 22 kb and 42 kb) in the first intron of *WNK1* or by a non-synonymous substitution in *WNK4* (four described mutations). Although *WNK1* and *WNK4* are expressed in multiple tissues, their major role is to regulate the transport of sodium and potassium ions in distal convoluted tubule and cortical collecting duct of nephrons, and thereby to contribute to blood pressure determination (Verissimo and Jordan, 2001; Wilson, et al., 2001). The human *WNK4* gene (19 exons) spans ~16 kb on chromosome 17q21.31. The human *WNK1* gene (28 exons+1 alternative exon) covers ~160 kb on chromosome 12p13 and codes for multiple transcripts initiated by alternative promoters (Wilson, et al., 2001; Xu Q, et al., 2002; Delaloy, et al., 2003;). Two major WNK1 isoforms have been described, a long isoform with complete kinase domain (L-WNK1) and a short kidney-specific

isoform, which is kinase-deficient (KS-WNK1) (Xu, Q et al., 2002). Although multiple alternative splice-forms of *WNK1* have been identified, the function of individual transcripts is yet to be determined. In addition to the identification of rare variants in *WNK1* and *WNK4* responsible for the Gordon syndrome, common single nucleotide polymorphisms (SNPs) in these genes have been associated with blood pressure variation and susceptibility to hypertension in general population among adults as well as children (Kokubo, et al., 2004; Newhouse, et al., 2009; Newhouse, et al., 2005; Osada, et al., 2009; Tobin, et al., 2005; Tobin, et al., 2008). SNPs in *WNK1* also affect the response of thiazide diuretics treatment on patient's blood pressure (Turner, et al., 2005).

While monogenic diseases are usually caused by rare variants located in the coding sequence of a gene, common diseases are rather considered to result from genetic variation in gene regulatory elements altering the expressional profile of the locus (Pastinen and Hudson, 2004; Visel, et al., 2009). As gene regulatory elements tend to map within evolutionarily conserved segments of the genome (Elgar and Vavouri, 2008; Hardison, 2000), these regions have a potential to harbor polymorphisms affecting common traits.

The aim of the current study was to screen the evolutionarily conserved non-coding regions of *WNK1* and *WNK4* to identify novel polymorphisms potentially affecting blood pressure in general population. Mutation screening resulted in the identification of a novel human-specific polymorphic *AluYb8* insertion in *WNK1* intron 10. This *Alu*-insertion was targeted to further evolutionary and population genetic analysis, as well as was also explored for association with blood pressure and its effect on the transcriptional profile of the *WNK1* gene in leucocytes.

Materials and Methods

In silico analysis of conserved non-coding regions in *WNK1* and *WNK4*

Conserved Non-coding Regions (CNRs) in *WNK1* and *WNK4* were screened using the web-based VISTA software (http://genome.lbl.gov/vista/index.shtml) with the proposed default parameters (cutoff criteria: 100 bp sliding window; sequence identity 70%; comparison with rat and mouse). The analyzed loci spanned from 10 kb upstream to 10 kb downstream of *WNK1* (12p13.3; coordinates 722,486-900,879, NCBI Build 36.1, hg18) and *WNK4* (17q21.31; coordinates 38,176,222- 38,212,587, NCBI Build 36.1, hg18). All VISTA regions that had any overlap with annotated genes track at UCSC Genome Browser (http://genome.ucsc.edu/) were excluded as potential coding regions. Polymorphism discovery was targeted to CNRs with sequence identity >70% between human and both rodents, length of the region 50-300 bp, and location >200bp from the nearest exon (**Supplementary Table S1**).

Screening for novel polymorphisms in *WNK1* and *WNK4* conserved non-coding regions

In total 40 (n=29 in *WNK1*; n=11 in *WNK4*) CNRs were selected for polymorphism screening, which was conducted either by *Denaturing Gradient Gel Electrophoresis* (DGGE; INGENYphorU-2x2 system, Ingeny International BV, Goes, The Netherlands) and/or *Denaturing High-Performance Liquid Chromatography* method (DHPLC; Wave Technologies Inc. USA). In the design of the DGGE and DHPLC assays and in establishing the experimental conditions, the manufacturers' recommendations were followed. Details of the assays are given in **Supplementary Text S1.** The design of both

DGGE and DHPLC assays was unsuccessful for 7 CNRs in *WNK1* and 2 CNR in *WNK4* due to failure in primer design or a negative result in the genome test (**Supplementary Table S1**). The rest of the 31 selected CNRs were screened for polymorphisms either by DHPLC (7 regions in *WNK1*; 5 in *WNK4*), by DGGE (7 in *WNK1*; 1 in *WNK4*) or by both assays (8 in *WNK1*; 3 in *WNK4*; **Supplementary Table S1).**

The average length of the analyzed CNR segments was 145 bp (range: 68-291 bp) and PCR fragments was 358 bp (range: 245-487 bp). Genomic DNAs of cardiovascular disease patients from two Eastern European studies (n=22 from HYPEST; n=24 from CADCZ; detailed description below) were targeted to polymorphism screening by DGGE (individual DNAs) and/or DHPLC (pools of DNA from three patients). PCR products exhibiting evidence for the presence of a polymorphism were sequenced at least twice on both forward and reverse orientations. Polymorphisms were identified using BioEdit Sequence Alignment Editor (Hall T., Department of Microbiology, North Carolina State University).

Genotyping of the *WNK1 AluYb8* in human populations

For large scale genotyping of *WNK1 AluYb8* in humans PCR followed by standard agarose gel (3% in 0.5X TBE buffer) electrophoresis was used. The primer design (WNK1_Alu_F: 5`-GGGTAACCAACCCTTGAAGTAGG-3`; WNK1_Alu_R: 5`-GGGTACTTCTCAAGTGATTAGGAGGA-3`) was carried out using the web-based program Primer3 (Rozen and Skaletsky, 2000). The distribution of the *WNK1 AluYb8* insertion was studied in six European (Estonians, n=50; Czech, n=50; CEPH, n=30; the Basque, n=50; Catalans, n=41; Spanish Gypsies, n=50), four Asian (Koreans, n=43;

Chinese Han, n=25; Tatars, n=47; Bashkir, n=47) and eight African populations (Tunisians, n=48; Algerians, n=48; Moroccans, n=84; Mandenkalu, n=24; Saharawi, n=50; Gabon Bantus, n=50; Gabon Pygmies, n=50; Tanzanians, n=17).

Conservation of the *WNK1 AluYb8* insertion in primates

The presence of the *WNK1 AluYb8* insertion was ascertained for a gorilla (*Gorilla gorilla;* primary cell line AG05251B), for an orangutan (*Pongo pygmaeus*; primary cell line AG12256) and for eleven western chimpanzees (*Pan troglodytes verus*) using identical PCR setup as in human genotyping. DNA sample of one chimpanzee originates from a wild-born male specimen (Pino) from Tallinn Zoo, Estonia. Ten samples of wild caught and unrelated animals (Annaclara, Frits, Hilko, Louise, Marco, Oscar, Regina, Socrates, Sonja and Yoran) are from the collection stored at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany and were kindly shared by Dr. Svante Pääbo. This sample collection is described in detail in Becquet, et al., 2007 and Ptak, et al., 2004.

Ancestral sequence of the targeted genomic region (*WNK1* exon10-intron10-exon11) was assessed by the comparative sequencing of the genomic DNA from *WNK1 AluYb8* insertion non-carrying (-/-) and carrying (Alu/Alu) human homozygotes as well as from a chimpanzee (Pino). Sequencing primers are listed in **Supplementary Table S3**. PCR cycling conditions, product purification and sequencing have been described elsewhere (Hallast, et al., 2005). Sequences were aligned using web-based global alignment program ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/).

Subjects for association studies with cardiovascular traits

Two Eastern-European sample collections, HYPEST (HYPertension in ESTonia) and CADCZ (Coronary Artery Disease in Czech) were used to conduct association analysis of the *WNK1 AluYb8* insertion with blood pressure and cardiovascular disease (**Table 1**). These sample collections have been recruited to target the genetic-epidemiological component of cardiovascular disease in Estonian and Czech populations, respectively. HYPEST subjects were recruited across Estonia during 2004-2007 (1823 individuals, age range 18-85 years) with the aim to evaluate risk factors for essential hypertension and related cardiovascular disease. Details of the recruitment are given in **Supplementary Text S1.** The CADCZ study subjects (n=877; n=296 coronary artery disease patients, n=581 controls) were recruited by the Cardiology Department of the 2nd Clinic of Internal Medicine, Faculty Hospital Královské Vinohrady in Prague and Czech heath clinics in years 1998-2000 (Janosikova, et al., 2003).

For the HYPEST subjects resting blood pressure was measured by trained clinicians during recruitment. Blood pressure (BP) measurements per subject were obtained using a standard mercury column sphygmomanometer and size-adjusted cuffs. All HYPEST individuals possessed a documented history of multiple systolic (SBP) and diastolic (DBP) blood pressure readings (on average 4.31 readings per individual, range 2–29) during mean 3.17 years (range 1–17 years). For association analysis with SBP and DBP we used the median across the longitudinal BP readings, and we excluded subjects receiving antihypertensive medication and BMI > 36. Thus, from among the genotyped HYPEST subjects (n=1747), 1211 HYPEST individuals (803 women, 408 men) qualified for the association study with BP (**Table 1**). In the case-control analysis the subjects with

essential hypertension (n=673) were defined as individuals under antihypertensive treatment or untreated individuals with SBP≥160 and/or DBP≥100; and controls (n=601) as individuals with SBP≤140 and DBP≤90.

Coronary artery disease (CAD) in CADCZ study was diagnosed according to WHO criteria, and one or more large stenosis of a major coronary vessel was confirmed by coronarography in all patients. Carotid wall intima media thickness (IMT) and the presence of carotid plaque, recorded in the CADCZ subjects were determined as described previously (Janosikova, et al., 2003). The CADCZ controls had no personal history of CAD, essential hypertension, MI, peripheral arterial disease, or stroke, and had never been prescribed any related medications.

RNA extraction and cDNA synthesis

EDTA-blood (9 ml) was collected from nine female subjects selected based on their alternative genotypes: three heterozygotes and three homozygotes for the *WNK1 AluYb8* insertion; and three subjects with wild-type sequence. Total RNA from leucocytes was extracted using LeukoLOCK™ Total RNA Isolation System (Ambion Inc, Austin, Texas, USA) including an optional TURBO™ DNase treatment to degrade the genomic DNA. Quantity and quality of extracted RNA was assessed with NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, LLC, Wilmington, Delaware, USA).

RNA was reverse transcribed using SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Life Technologies Corporation, Carlsbad, California, USA) according to the manufacturer's instructions (details in **Supplementary Text S1**).

Quantification of *WNK1* transcripts by real-time PCR

Relative expression analysis of three *WNK1* splice-forms (ex+11+12, ex–11+12 and ex–11–12; **Figure 3A)** was performed with real-time PCR. Primer-probe mix of the *WNK1* transcript including exon 11 (ex+11+12; Hs01018312_m1, amplicon size 78bp) and selected reference gene *HPRT1* (Human HPRT1 (HGPRT) Endogenous Control (VIC / MGB Probe, Primer Limited, amplicon size 100 bp)) were purchased from Applied Biosystems, Inc. Primers and probes for the *WNK1* transcripts lacking exon 11 (ex–11+12) and both exons 11 & 12 (ex–11–12) were designed using Primer Express version 3.0 (Applied Biosystems Inc, Foster City, CA, USA). Oligonucleotide sequences are given in **Supplementary Table S3**.

The real-time RT-PCRs were performed using Applied Biosystems 7900HT Fast Real-time PCR system in 96 micro-well plates. Target region and endogenous control were amplified in the same well. The experimental conditions for the real-time PCR are given in detail in **Supplementary Text S1**. In total six replicate analyses of each of the nine extracted RNA samples were conducted: two independently synthesized cDNAs were assayed by RT-PCR reactions in triplicate.

Statistical analysis

Statistical differences in allele frequencies between populations were calculated using the web-based Fisher´s Exact Test calculator (http://www.langsrud.com/fisher.htm). The significance of the associations between the *WNK1 AluYb8* insertion and blood pressure (systolic blood pressure, SBP; diastolic blood pressure, DBP) was tested using linear regression (additive genetic model) with age and gender as covariates. Additive genetic

model assumes a trend per copy of the minor allele to contribute to the trait or disease susceptibility on genotype categories. Association with the diagnosis of cardiovascular disease (essential hypertension; coronary artery disease, CAD) was assessed by Cochran-Armitage trend test. Association testes were implemented in the PLINK software, ver. 1.04 (http://pngu.mgh.harvard.edu/~purcell/plink/).

Normalized expression values of target regions were calculated using Microsoft® Excel®-based software Q-Gene (Muller, et al., 2002). Q-Gene calculates the normalized expression values of the target gene based on the Ct values and the reaction efficiencies of the target and the reference gene (here *HPRT*). For every study subject six replicate values of relative expression per each alternative *WNK1* splice-form (ex+11+12, ex–11+12 and ex–11–12) was calculated. As each of the genotypes (*Alu/Alu, Alu/–; –/–*) was represented by three individuals, in total 18 data-points were collected per transcript within a genotype group. The most outlier Ct value within the respective genotype group was excluded from the statistical testing. Differences of normalized expression values between alternative genotype groups were estimated by Wilcoxon rank sum test implemented in R, ver. 2.7.2 (R Development Core Team 2008, http://www.r-project.org/). For all tests p-value <0.05 was considered as statistically significant.

Results

Polymorphism screening in *WNK1* and *WNK4* conserved non-coding regions

DHPLC and/or DGGE assays were designed for screening novel polymorphisms in conserved non-coding regions (CNRs) of the *WNK1* (29 targeted CNRs based on criteria

outlined in Materials and Methods) and the *WNK4* (11 CNRs) genes (**Supplementary Table S1**). Based on the *in silico* quality control criteria for assay design, nine regions were excluded from the wet-lab analysis. Finally, 31 CNRs entered mutation detection in cardiovascular disease patients (HYPEST, 22 hypertensives; CADCZ, 24 CAD patients). Among the screened 31 CNRs, one single nucleotide polymorphism (SNP) was identified in the *WNK4* and six SNPs in the *WNK1* gene (**Supplementary Table S2**). All but one (rs36052085) of the detected SNPs were rare (minor allele frequency <10%), including three singletons (two novel). The functional effect of these SNPs was addressed neither by association studies due to large sample size requirements nor by gene expression analysis due to unavailability of minor allele homozygotes.

In addition, in one of the *WNK1* CNRs, a novel unreported common indel (~ 300 bp) was detected (**Supplementary Table S2**). Sequence analysis of this variant revealed a polymorphic insertion of an *AluYb8* element (288 bp without flanking T nucleotides) into a poly-T tract within *WNK1* intron 10 (**Figure 1, Supplementary Figure S1**). This *Alu*-insertion was targeted to further evolutionary and population genetic analysis as well as was explored for the association with cardiovascular disease and the effect on the gene expression profile.

Alu distribution in human populations

The *WNK1 AluYb8* insertion was genotyped in 22 populations from Europe, Asia and Africa (**Supplementary Table S4**). Frequency of the *AluYb8* insertion in human populations differed based on their geographic affiliation (**Table 2**). The proportion of

WNK1 AluYb8 carriers in Sub-Saharan Africa was significantly lower (average allele frequency 4.8%; range 2.1–7.0%) compared to North-African (mean 16.4%, range 10.4–25.0%; Fisher`s Exact Test, P=2.2x10^{-6}), European (mean 15.1%, range 12.0–16.5%; P=8.7x10^{-9}) and Asian (mean 15.9%, range 11.6%–22%; P=9.4x10^{-6}). On average, the allele frequency of the *WNK1 AluYb8* in Sub-Saharan Africa was ~3.3 times lower than in other studied populations (P=9.7x10^{-9}).

Alu insertion in primates and conservation around insertion site

The analysis of the *WNK1* intron 10 in eleven chimpanzees, one gorilla and one orangutan revealed that the *WNK1 AluYb8* insertion has most probably occurred in human lineage. No *AluYb8* insertion was detected in the *WNK1* intron 10 of the studied primate genomes (**Figure 1**). Comparative sequencing of the *WNK1* genomic fragment (exon10/ intron10/ exon11) amplified from a chimpanzee and from human wildtype as well as *AluYb8* insertion carrying chromosomes revealed high conservation of intron 10 (**Supplementary Figure S1)**. The substitution divergence between human wildtype and chimpanzee *WNK1* was 0%, 0.2% and 1.1% for exon 10 (150 bp), exon 11 (459 bp) and intron 10 (1211 bp), respectively.

Association studies of the *WNK1 AluYb8* insertion and cardiovascular traits

Associations of the *WNK1 AluYb8* insertion with systolic (SBP) and diastolic (DBP) blood pressure were tested by linear regression under additive genetic model (**Table 3, Figure 2**). In HYPEST study subjects (n=1211), the *AluYb8* insertion was associated with significantly higher SBP (P=0.0126; beta=2.23) and DBP (P=0.0304; beta=1.22).

Gender-specific analysis revealed that the effect of the *AluYb8* insertion on blood pressure is more pronounced in women (SBP: *P*=0.013, beta=2.72; DBP: *P*=0.006, beta= 1.84), while no significant association was detected in men (**Table 3**). *AluYb8*-carrying homozygote HYPEST women had an average 4.5 mmHg higher median SBP and 3.5 mmHg higher median DBP compared to wildtype homozygotes (**Figure 2**). Association of the *WNK1 AluYb8* insertion with cardiovascular disease was studied in two Eastern European study samples: essential hypertension (HYPEST, Estonia; cases n=673/ controls n=601) and coronary artery disease, CAD (CADCZ, Czech; cases n=296/ controls n=581). Significant association was detected with clinically diagnosed essential hypertension (Cochran-Armitage trend test, *P*=0.029) and the association was stronger with female hypertension (*P*=0.016; **Table 4**). Although association testing with CAD did not reach statistically significant *P*-values, the differences in allele frequencies between cases and controls were consistent with the results for essential hypertension (**Table 4**). In hypertensives, the frequency of the *AluYb8* insertion (17.7%) was >3 % higher compared to normotensives (14.51%). In CAD patients (17.23%) the difference from control group (15.32%) was ~ 2%. In both studies, the frequency of the *AluYb8* insertion in female patients was ~19% compared to <15% in female controls.

The impact of the *AluYb8* insertion on the expression profile of *WNK1* in leukocytes

In order to explore the functional effect of the *AluYb8* insertion on the expression profile of *WNK1*, we quantified the gene transcripts in RNA extracted from the leukocytes of nine women with alternative genotypes. The study subjects included three individuals heterozygous and three homozygous for the *AluYb8* insertion, as well as three wildtype

genotype carriers. The expression of three *WNK1* splice-forms was addressed using relative quantification method based on real-time RT-PCR assays. The studied splice-forms differed by alternative inclusion/exclusion of exon 11 and exon 12 (**Figure 3A**). Compared to the subjects with the wild-type genotype, the heterozygous *AluYb8* carriers had significantly lower expression of all three splice-forms (Wilcoxon rank sum test; ex+11+12: P=0.006; ex–11+12: P=6.78x10^{-7}; ex–11–12: P=0.005; **Figure 3B**). Consistently, homozygous *AluYb8* carriers showed lower expression level of splice forms ex–11+12 (P=0.002) and ex–11–12 (P=0.182; **Figure 3C**). However, the expression of the full-length *WNK1* transcript, which includes both exon 11 and 12 (ex+11+12) was up-regulated among *AluYb8* homozygous carriers (P=0.036). We conclude that the carrier-status of the *AluYb8* insertion may have an impact on the profile in *WNK1* transcript in leukocytes.

Discussion

We targeted conserved non-coding regions in hypertension candidate genes *WNK1* and *WNK4* to polymorphism screening in order to identify functional variants potentially contributing to blood pressure determination. We identified a novel human-specific polymorphic *AluYb8* insertion in *WNK1* intron 10.

The *AluYb8* insertion belongs to a young *Alu* subfamily represented with ~2200 copies in the human genome compared to only nine insertions detected in chimpanzee (Gibbons, et al., 2004). Consistently, we were unable to detect the studied *WNK1 AluYb8* insertion in chimpanzee, gorilla and orangutan (**Figure 1**). As *AluYb8* elements are relatively mobile *Alu*-repeats, they represent together with *AluYa5* subfamily ~58% of the polymorphic

Alu-s in the human genome (Bennett, et al., 2008). The increased carrier frequency of *WNK1 AluYb8* insertion out of Africa is consistent with recent studies showing that the allele frequencies of polymorphic *Alu*-insertions tend to be lowest in Sub-Saharan populations (**Table 2**; (Watkins, et al., 2001 and 2003). In Africa, the fraction of carriers of polymorphic *Alu*-s increase with sharp cline in the north of Sahara compared to the populations living south of the desert (Comas, et al., 2000).

Our study identified a significant association between *WNK1 AluYb8* insertion and blood pressure in HYPEST sample ($P<0.05$) and suggested a potential functional effect of this *Alu*-insertion on the expressional profile of the *WNK1* gene in leukocytes. In clinically diagnosed essential hypertension patients the allele frequency of the *AluYb8* insertion (17.7%) was significantly ($p<0.05$) higher compared to normotensive controls (14.51%; **Table 4**). Consistently, a non-significant trend for association was observed with coronary artery disease (CAD) (cases 17.23%, controls 15.32%; **Table 4**) in CADCZ study. However, the association testing with CAD was limited by an insufficient sample size.

In the current study the *WNK1 AluYb8* insertion was associated with blood pressure only among women (**Table 3**). Consistently, it showed increased frequency among female, but not among male cardiovascular disease (essential hypertension, HYPEST study; CAD, CADCZ study) patients compared to healthy controls (**Table 4**). Similar gender-specific effects have been reported for the polymorphic *Alu*-insertion (rs4646994) located in intron 16 of the *ACE* (*angiotensin converting enzyme*) gene (Rigat, et al., 1990). Three independent studies showed that *ACE Alu* I/D variant is associated with the hypertension risk only in men and not in women (Higaki, et al., 2000; O'Donnell, et al., 1998;

Stankovic, et al., 2002). Another explanation for the lack of significant association between *WNK1 AluYb8* insertion and male blood pressure in the HYPEST study could be reduced power as this sample includes twice as many women compared to men. Interestingly, 597 bp downstream from *AluYb8* insertion lies a C/T SNP (rs880054), the C allele (frequency 43.8% in British; 36.1% in Japanese) of which has been associated with reduced blood pressure (Osada, et al., 2009; Tobin, et al., 2005). Sequence comparison of the wild-type *WNK1* and the *AluYb8*-carrying chromosomes revealed that the *Alu*-insertion had probably occurred on rs880054 T-allele carrying gene variant. Consistent with the effect of the *Alu*-insertion, the rs880054 T-allele has been associated with increased risk to hypertension.

Although a majority of *Alu* elements are considered to be neutral residents of the human genome, an inserted copy of an *Alu*-repeat could interrupt structurally or functionally important genomic regions and consequently affect the expression of a locus (Batzer and Deininger, 2002; Callinan and Batzer, 2006). *Alu* elements may alter gene expression through modulating alternative splicing, RNA editing, epigenetic regulation and translation regulation (Cordaux and Batzer, 2009; Hasler and Strub, 2006). So far, thirty-three diseases directly caused by novel *Alu* insertions have been identified (Belancio, et al., 2008). Our study using mRNA extracted from human leucocytes indicated a potential effect of the presence of the *AluYb8* insertion in *WNK1* intron 10 on the expressional profile of *WNK1* alternative transcripts. Splicing is an incompletely understood process carried out by large macromolecular complex spliceosome and directed by numerous regulatory elements located within exonic and intronic sequence (Black, 2003). The size of the *WNK1* intron 10 (human wild type 1211 bp) is remarkably increased by the ~300

bp *AluYb8* insertion (human variant >1500 bp). Thus, we hypothesize that the presence of the *AluYb8* insertion may disrupt the spatial intronic structure and/or disarrange the possible splicing regulatory sequences within *WNK* intron 10. Consequently, it may affect the splicing efficiency of the down-stream exons 11 and 12. As alternative splicing tends to be a tissue and developmental stage specific process (Xu Q, et al., 2002), the impact of *AluYb8* insertion on the expressional profile of *WNK1* may vary in different tissues. The current study design was limited to addressing the effect of *AluYb8* insertion on *WNK1* expressional profile in leucocytes. Further *in vitro* and *in vivo* studies using renal tissues would reveal the potential effect of this *Alu*-insertion on *WNK1* expressional profile in kidneys, where it plays an important role in contributing to the regulation of ion transport.

In summary, we identified a novel human-specific polymorphic *AluYb8* insertion in *WNK1*. This *AluYb8* insertion showed significant association with blood pressure and an effect on the expressional profile of alternative *WNK1* transcripts in leukocytes.

Acknowledgements

We express our appreciation to all the HYPEST and the CADCZ participants. We thank Dr. Bohumila Janosikova for technical assistance in recruitment of the CADCZ sample collection. We thank the personnel of the Centre of Cardiology, North Estonia Medical Centre and the Department of Cardiology, University of Tartu for assistance in subject recruitment and data collection for the Estonian HYPEST sample collection. We also thank Lolke van der Veen, Institut des Sciences de l'Homme, Lyon, for sharing Gabonese samples. The study was supported by Wellcome Trust International Senior Research Fellowships (grant no. 070191/Z/03/Z, to ML; no. 070255/Z/03/Z to VK) in Biomedical Science in Central Europe. Additionally, the study has been supported by Estonian Ministry of Education and Science core grant no. 0182721s06, HHMI International Scholarship #55005617 (to ML), Estonian Science Foundation grants no ETF7471 (to ML; stipend for MP), travel-stipend from Kristjan-Jaak foundation (KK), and by an institutional Research project from Ministry of Education of the Czech Republic MSM0021620806 (to VK).

References

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3(5):370-9.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D. 2007. Genetic structure of chimpanzee populations. PLoS Genet 3(4):e66.

Belancio VP, Hedges DJ, Deininger P. 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18(3):343-58.

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. 2008. Active Alu retrotransposons in the human genome. Genome Res 18(12):1875-83.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72:291-336.

Callinan PA, Batzer MA. 2006. Retrotransposable Elements and Human Disease. Genome and Disease 1:104-115.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum Genet 107(4):312-9.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10(10):691-703.

Delaloy C, Lu J, Houot AM, Disse-Nicodeme S, Gasc JM, Corvol P, Jeunemaitre X. 2003. Multiple promoters in the WNK1 gene: one controls expression of a kidney-specific kinase-defective isoform. Mol Cell Biol 23(24):9208-21.

Elgar G, Vavouri T. 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. Trends Genet 24(7):344-52.

Gibbons R, Dugaiczyk LJ, Girke T, Duistermars B, Zielinski R, Dugaiczyk A. 2004. Distinguishing humans from great apes with AluYb8 repeats. J Mol Biol 339(4):721-9.

Hallast P, Nagirnaja L, Margus T, Laan M. 2005. Segmental duplications and gene conversion: Human luteinizing hormone/chorionic gonadotropin beta gene cluster. Genome Res 15(11):1535-46.

Hardison RC. 2000. Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet 16(9):369-72.

Hasler J, Strub K. 2006. Alu elements as regulators of gene expression. Nucleic Acids Res 34(19):5491-7.

Higaki J, Baba S, Katsuya T, Sato N, Ishikawa K, Mannami T, Ogata J, Ogihara T. 2000. Deletion allele of angiotensin-converting enzyme gene increases risk of essential hypertension in Japanese men : the Suita Study. Circulation 101(17):2060-5.

Janosikova B, Pavlikova M, Kocmanova D, Vitova A, Vesela K, Krupkova L, Kahleova R, Krijt J, Kraml P, Hyanek J and others. 2003. Genetic variants of homocysteine metabolizing enzymes and the risk of coronary artery disease. Mol Genet Metab 79(3):167-75.

Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40(5):592-9.

Kokubo Y, Kamide K, Inamoto N, Tanaka C, Banno M, Takiuchi S, Kawano Y, Tomoike H, Miyata T. 2004. Identification of 108 SNPs in TSC, WNK1, and WNK4 and their association with hypertension in a Japanese general population. J Hum Genet 49(9):507-15.

Lifton RP, Gharavi AG, Geller DS. 2001. Molecular mechanisms of human hypertension. Cell 104(4):545-56.

Muller PY, Janovjak H, Miserez AR, Dobbie Z. 2002. Processing of gene expression data generated by quantitative real-time RT-PCR. Biotechniques 32(6):1372-4, 1376, 1378-9.

Newhouse S, Farrall M, Wallace C, Hoti M, Burke B, Howard P, Onipinla A, Lee K, Shaw-Hawkins S, Dobson R and others. 2009. Polymorphisms in the WNK1 gene are associated with blood pressure variation and urinary potassium excretion. PLoS One 4(4):e5003.

Newhouse SJ, Wallace C, Dobson R, Mein C, Pembroke J, Farrall M, Clayton D, Brown M, Samani N, Dominiczak A and others. 2005. Haplotypes of the WNK1 gene associate with blood pressure variation in a severely hypertensive population from the British Genetics of Hypertension study. Hum Mol Genet 14(13):1805-14.

O'Donnell CJ, Lindpaintner K, Larson MG, Rao VS, Ordovas JM, Schaefer EJ, Myers RH, Levy D. 1998. Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study. Circulation 97(18):1766-72.

Osada Y, Miyauchi R, Goda T, Kasezawa N, Horiike H, Iida M, Sasaki S, Yamakawa-Kobayashi K. 2009. Variations in the WNK1 gene modulates the effect of dietary

intake of sodium and potassium on blood pressure determination. J Hum Genet 54(8):474-8.

Pastinen T, Hudson TJ. 2004. Cis-acting regulatory variation in the human genome. Science 306(5696):647-50.

Ptak SE, Roeder AD, Stephens M, Gilad Y, Paabo S, Przeworski M. 2004. Absence of the TAP2 human recombination hotspot in chimpanzees. PLoS Biol 2(6):e155.

Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F. 1990. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. J Clin Invest 86(4):1343-6.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-86.

Stankovic A, Zivkovic M, Alavantic D. 2002. Angiotensin I-converting enzyme gene polymorphism in a Serbian population: a gender-specific association with hypertension. Scand J Clin Lab Invest 62(6):469-75.

Tobin MD, Raleigh SM, Newhouse S, Braund P, Bodycote C, Ogleby J, Cross D, Gracey J, Hayes S, Smith T and others. 2005. Association of WNK1 gene polymorphisms and haplotypes with ambulatory blood pressure in the general population. Circulation 112(22):3423-9.

Tobin MD, Timpson NJ, Wain LV, Ring S, Jones LR, Emmett PM, Palmer TM, Ness AR, Samani NJ, Smith GD and others. 2008. Common variation in the WNK1 gene and blood pressure in childhood: the Avon Longitudinal Study of Parents and Children. Hypertension 52(5):974-9.

Turner ST, Schwartz GL, Chapman AB, Boerwinkle E. 2005. WNK1 kinase polymorphism and blood pressure response to a thiazide diuretic. Hypertension 46(4):758-65.

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. Am J Hum Genet 68(3):738-52.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV and others. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13(7):1607-18.

Vehaskari VM. 2009. Heritable forms of hypertension. Pediatr Nephrol 24(10):1929-37.

Verissimo F, Jordan P. 2001. WNK kinases, a novel protein kinase subfamily in multi-cellular organisms. Oncogene 20(39):5562-9.

Wilson FH, Disse-Nicodeme S, Choate KA, Ishikawa K, Nelson-Williams C, Desitter I, Gunel M, Milford DV, Lipkin GW, Achard JM and others. 2001. Human hypertension caused by mutations in WNK kinases. Science 293(5532):1107-12.

Wilson FH, Kahle KT, Sabath E, Lalioti MD, Rapson AK, Hoover RS, Hebert SC, Gamba G, Lifton RP. 2003. Molecular pathogenesis of inherited hypertension with hyperkalemia: the Na-Cl cotransporter is inhibited by wild-type but not mutant WNK4. Proc Natl Acad Sci U S A 100(2):680-4.

Visel A, Rubin EM, Pennacchio LA. 2009. Genomic views of distant-acting enhancers. Nature 461(7261):199-205.

Xu B, English JM, Wilsbacher JL, Stippec S, Goldsmith EJ, Cobb MH. 2000. WNK1, a novel mammalian serine/threonine protein kinase lacking the catalytic lysine in subdomain II. J Biol Chem 275(22):16795-801.

Xu Q, Modrek B, Lee C. 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res 30(17):3754-66.

Table 1. Phenotypic parameters of HYPEST and CADCZ study subjects

Parameter		HYPEST[1]		CADCZ[1]	
	Cohort	Cases	Controls	Cases	Controls
No. of individuals	1211	673	601	296	581
(male/female)	(408/803)	(228/445)	(162/439)	(243/53)	(279/302)
Age at recruitment (years)	44.8	56.0	38.9	55.6	47.1
mean±SD	±12.5	±9.5	±9.0	±6.7	±11.0
Age at onset of disease (years)		44.0		51.4	
mean±SD	na	±12.7	na	±7.3	NA
Body mass index (kg/m2)	26.7	30.3	24.6	28.5	26.3
mean±SD	±4.8	±5.1	±4.3	±4.1	±4.4
Systolic blood pressure (mmHg)	141.0	161.4	127.8	127.4	137.4
mean±SD	±19.0	±20.7	±8.0	±15.8	±18.8
Diastolic blood pressure (mmHg)	87.2	97.5	80.8	85.8	81.2
mean±SD	±11.0	±12.6	±6.4	±9.6	±10.1
Antihypertensive treatment					
(% of subjects)	0.0%	78.5%	0.0%	55.6%	10.8%
Myocardial infarction					
(% of subjects)	0.0%	11.9 %	0.0%	70.3 %	0.0%

[1] Recruitment in HYPEST and CADCZ studies is described in **Supplementary Text S1** and definition of cases/controls of essential hypertension (HYPEST) and coronary artery disease (CADCZ) is given in Materials and Methods; na – not applicable

Table 2. *WNK1* intron 10 *AluYb8* allele frequencies in population groups

Group	Group size	Allele frequency	Population composition
Eastern Europe	150	16.3%	Estonians, Czech
Western Europe	121	14.1%	CEPH/Utah families, Basques, Catalans
Gypsies	50	12.0%	Spanish Gypsies
Volga-Ural	94	17.0%	Tatars, Bashkirs
Eastern Asia	68	16.8%	Chinese Han, Koreans
North Africa	230	16.4%	Moroccans, Saharawi, Algerians, Tunisians
Sub-Saharan Africa	141	**4.8%**	Mandenkalu, Tanzanians, Gabon Bantus, Gabon Pygmies

Table 3. Association of the *WNK1* intron 10 *AluYb8* with blood pressure

HYPEST sample		Systolic blood pressure			Diastolic blood pressure		
Subjects	*N*	p-value[1]	beta[2]	s.e.	p-value[1]	beta[2]	s.e.
All	1211	**0.0126**	2.23	0.89	**0.0304**	1.22	0.56
women	803	**0.0132**	2.72	1.10	**0.0062**	1.84	0.67
men	408	0.5972	0.78	1.48	0.8672	-0.17	1.03

[1]p-value for linear regression (additive model), age and sex as covariates; $p<0.05$ in bold

[2]effect on blood pressure (mmHg)

N – sample size; s.e.- standard error

Table 4. Association of the *WNK1* intron 10 *AluYb8* with cardiovascular disease

Cardiovascular disease[1] (study sample)	Sample size case/control	*WNK1 AluYb8* allele frequency (%)		Association testing[2]
		Cases[1]	Controls	p-value
Essential hypertension (HYPEST sample)				
all	673/601	17.70	14.51	**0.029**
women	445/439	19.28	14.97	**0.016**
men	228/162	14.63	13.27	0.588
Coronary artery disease (CADCZ sample)				
all	296/581	17.23	15.32	0.303
women	53/302	18.87	14.90	0.315
men	243/279	16.87	15.77	0.624

[1]Definition of essential hypertension and CAD is given in Materials and Methods

[2] Cochran-Armitage trend test; $p<0.05$ in bold

Figure legends

Figure 1. Detection of the presence of *WNK1* intron 10 *AluYb8* insertion in primates. Agarose gel (3%) electrophoresis of *WNK1* intron 10 PCR products amplified from human, chimpanzee, gorilla and orangutan genomic DNAs. In humans, alternative genotype carriers are shown: wild-type homozygote without *AluYb8* insertion (-/-, PCR product 353 bp); heterozygous (A/-) and homozygous (A/A, PCR product 660 bp) carriers of the insertion.

Figure 2. Box-plot diagram for the distribution of diastolic (DBP) and systolic (SBP) blood pressure values in HYPEST sample (all, n=1211; women, n=803; men, n=408). The study subjects were sub-grouped based on the carrier status of *WNK1* intron 10 *AluYb8*. The boxes represent the 25th and 75th percentiles; whiskers are lines extending from each end of the box covering the extent of the data on 1.5 x inter-quartile range. The median value is denoted as the line that bisects the boxes and circles represent the outlier values. On each box-plot p-values from linear regression testing (additive model, age and sex as covariates) are shown.

Figure 3. Expression of **(A)** three *WNK1* alternative splice-forms in blood leukocytes obtained from **(B)** heterozygous (Alu/-) and **(C)** homozygous (Alu/Alu) carriers of the *WNK1 AluYb8* insertion in comparison with the wild-type homozygote without the insertion.

(A) Alternative splicing of *WNK1* exons 10-13 is presented schematically according to Delaloy et al., 2003. Black numbered boxes and horizontal lines represent exons and introns, respectively and dotted lines indicate splicing events. **(B, C)** Relative mRNA quantification of the targeted *WNK1* splice-forms in leukocytes was

performed with real-time RT-PCR (Taqman assay, *HPRT* as a reference gene). Relative expression of each targeted *WNK1* splice-form in subjects with the *AluYb8* insertion (Alu/Alu homozygotes, Alu/– heterozygotes) is shown using the quantity of the transcript in wild-type homozygotes (–/–) as a reference value (wt=1). The presented relative expression levels represent the mean values of the three study subjects within the genotype group (each individual represented by six data-points from replicate experiments). Bars represent standard error of the relative expression. *P*-values reflecting the differences between groups were estimated by Wilcoxon rank sum test.

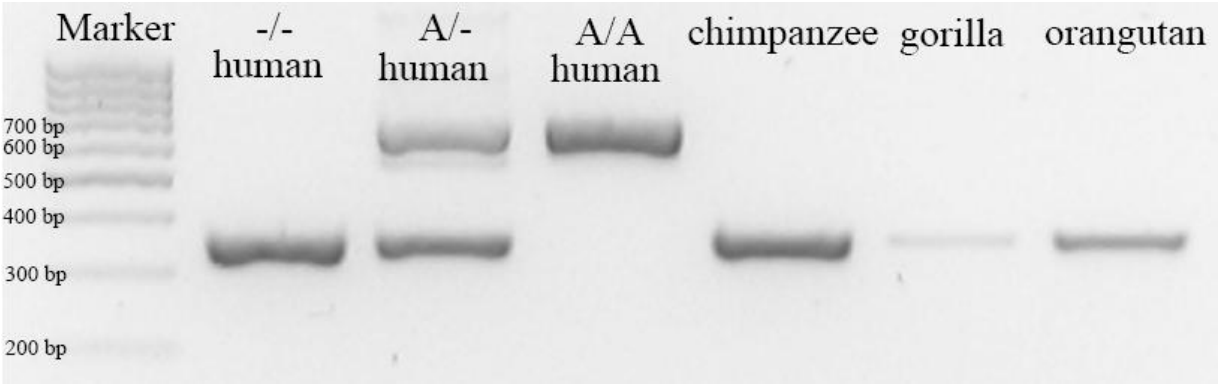
Marker
-/-
human
A/-
human
A/A
human
chimpanzee
gorilla
orangutan
700 bp
600 bp
500 bp
400 bp
300 bp
200 bp

Figure 1

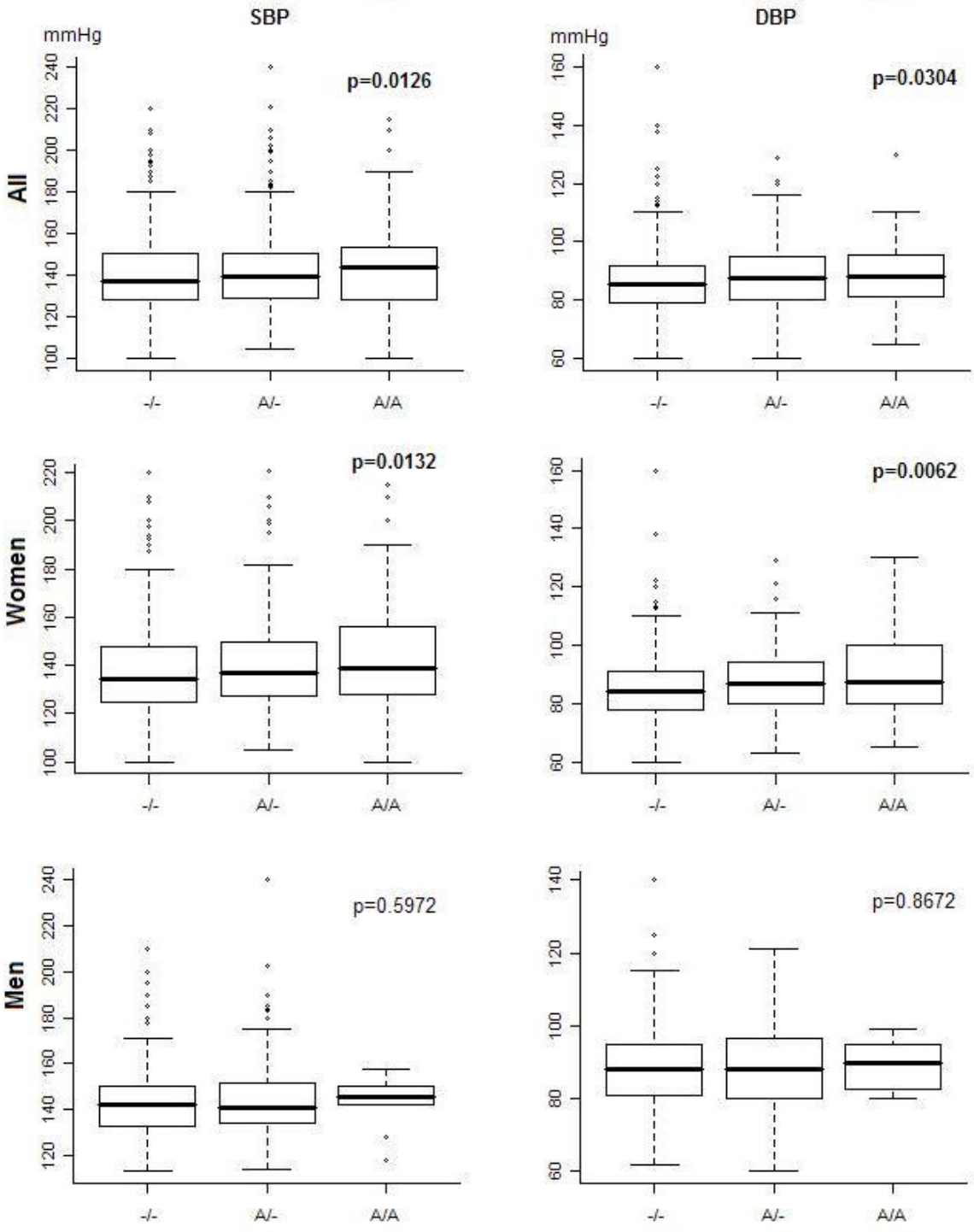
SBP
DBP
mmHg
mmHg
All
Women
Men
p=0.0126
p=0.0304
p=0.0132
p=0.0062
p=0.5972
p=0.8672
-/-
A/-
A/A

Figure 2

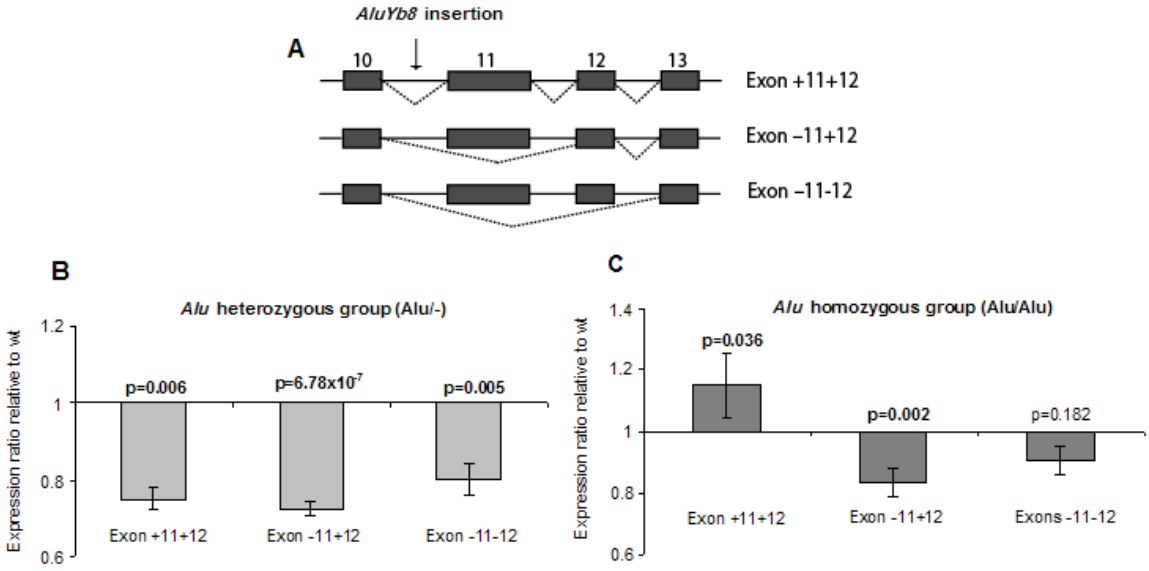
AluYb8 insertion
A
10
11
12
13
Exon +11+12
Exon –11+12
Exon –11-12
B
Alu heterozygous group (Alu/-)
Expression ratio relative to wt
1.2
1
0.8
0.6
p=0.006
p=6.78x10-7
p=0.005
Exon +11+12
Exon -11+12
Exon -11-12
C
Alu homozygous group (Alu/Alu)
Expression ratio relative to wt
1.4
1.2
1
0.8
0.6
p=0.036
p=0.002
p=0.182
Exon +11+12
Exon -11+12
Exons -11-12

Figure 3