npg

## ARTICLE

# Detection of putative functional angiotensinogen (AGT) gene variants controlling plasma AGT levels by combined segregation-linkage analysis

Eva Brand[1,5], Nathalie Chatelain[1], Françoise Paillard[2], Laurence Tiret[1], Sophie Visvikis[1], Mark Lathrop[3], Florent Soubrier*[,1] and Florence Demenais[4]

[1]Inserm Unité 525, Paris, France; [2]Inserm Unité 489, Hôpital Tenon, Paris; [3]Centre National de Genotypage, Evry, France; [4]Inserm EMI 00-06, Evry, France

Previous studies have suggested that angiotensinogen (AGT) gene variants are associated with increased plasma AGT levels, and may also contribute towards the inherited component of predisposition to essential hypertension in humans. To explore the potential functionality of several *AGT* polymorphisms and estimate their effects, together with other sources of familial correlations, on plasma AGT, we undertook a large study involving 545 healthy French volunteers in 130 nuclear families that include 285 offspring. Plasma AGT levels were measured in all participants, and bi-allelic *AGT* variants were analysed as candidate functional variants at three sites in the 5′-flanking region (C-532T, A-20C, G-6A), two sites in exon 2 (M235T, T174M) and two newly identified variant sites in the untranslated sequence of exon 5 and the 3′-flanking region (C+2054A, C+2127T) of the gene. Analysis with the class D regressive model showed significant effects influencing plasma AGT levels of all *AGT* polymorphisms tested, with the exception of T174M. The most significant result was found at C-532T ($P=0.000001$), which accounts for 4.3% of total plasma AGT variability in parents and 5.5% in offspring, with substantial residual familial correlations. Maximum likelihood estimates of haplotype frequencies and tests of linkage disequilibrium between each *AGT* polymorphism and a putative QTL are in agreement with a complete confounding of C-532T with the QTL, when taking into account sex and generation specific effects of the QTL. However, further combined segregation-linkage analyses showed significant evidence for additional effects of G-6A, M235T and C+2054A polymorphisms after accounting for C-532T, which supports a complex model with at least two functional variants within the *AGT* gene controlling AGT levels.
*European Journal of Human Genetics* (2002) **10**, 715–723. doi:10.1038/sj.ejhg.5200874

## Introduction

Angiotensinogen (AGT) is the inactive precursor of the potent vasoactive and salt-retaining hormone angiotensin II, and thus constitutes a major component of the renin angiotensin system which controls blood pressure and

*Correspondence: Florent Soubrier; Inserm Unité 525, Faculté de Méde-cine Pitié-Salpétrière, 91, Bd de l'hôpital, 75013 Paris, France
Tel.: +33 1 40779725; Fax: +33 1 40779728;
E-mail: florent.soubrier@chups.jussieu.fr
5Current address: Benjamin Franklin Medical Center, Dept. of Endocrinology and Nephrology, Freie Universität Berlin, Germany

salt-water homeostasis. Plasma AGT level is rather stable in one individual, but is under the long-term control of several hormones, such as glucocorticoids, oestrogens, thyroid hormones, and angiotensin II, which are known to induce *AGT* expression.[1] Beside the hormonal control of *AGT* expression, a genetic influence on AGT plasma level has been documented since polymorphisms of *AGT* have been shown to be associated with significant effects on plasma AGT concentration.[2] Interestingly, alleles associated with higher levels of plasma AGT have an increased frequency in hypertensive subjects as compared to normotensive controls, suggesting that *AGT* variants might be

predisposing to essential hypertension. However, positive results found in some association studies[2–5] were not replicated in others.[6–9] Linkage studies of hypertension with *AGT* in various ethnic groups, using sib-pair methods, failed to replicate[10,11] initial positive results.[2] Given the lack of power of linkage studies to detect quantitative trait loci (QTL) with small effects, *AGT* may still contribute as a minor factor to essential hypertension, through changes in its basal expression or regulation.

Finding the molecular variants within *AGT* which are responsible for genetically determined differences in expression would be a major step for elucidating regulation of *AGT* expression and its genetic variability, and for further genetic studies. The M235T polymorphism, which is strongly associated with plasma AGT levels, was definitively excluded as the functional variant by biochemical analysis of recombinant $AGT_{M235}$ and $AGT_{T235}$ molecules.[12] Haplotype-based association studies with hypertension designated the G-6A polymorphism, located in the proximal 5′-flanking region of the gene, as a good candidate.[13] *In vitro* studies showed a difference in the level of transcription induced on reporter genes by *AGT* promoters containing the A or the G allele, although differences due to the alleles were slight as compared to variability in the results.[12]

In contrast to previous studies which were based on case-control comparisons, we designed a family study to explore the potential functionality of *AGT* polymorphisms and estimate their effects together with residual familial correlations. Combined segregation-linkage analysis was carried out, using the regressive models, which make it possible to search for the effect of polymorphisms on AGT variability in the presence of other sources of familial covariation and to test whether these polymorphisms are in complete or incomplete linkage disequilibrium with the yet-unknown functional variant. To conduct this analysis, we measured several anthropometric variables in a series of 130 nuclear families, determined their plasma AGT concentration and genotypes for already described and newly identified polymorphisms in regions known to contain transcriptional regulatory elements within exon 5 and the 3′-flanking region of *AGT*.

## Materials and methods
### Family data
From 1993 to 1994, 130 healthy, nuclear families of Caucasian origin, composed of both parents aged <60 years and at least two offspring aged >9 years, were recruited in the Centre for Preventive Medicine in Vandœuvre-lès-Nancy (France) to set up a study on familial risk factors for hypertension. The study was approved by the appropriate institutional review board and informed consent was obtained from all subjects. Subjects with acute or chronic disease, body mass index (BMI) >28 kg/m$^2$, a history of alcohol intake >50 g/24 h, and gamma-glutamyl-transferase >50 IU/l, taking oral contraceptives or antihypertensive-drug therapy, were excluded. Blood pressure was measured in the recumbent position with an automatic device (Dinamap, Critikon). Seven measurements were taken at 3-min intervals, and repeated 15 days later. Application of these criteria led to the inclusion of 130 apparently healthy nuclear families with a total of 545 subjects, including 285 offspring (average number of offspring 2.2) (Table 1).

### Measurement of plasma AGT and DNA extraction
Plasma AGT was measured as the generation of angiotensin I by radioimmunoassay.[14] Genomic DNA from included subjects were extracted by standard techniques.[15]

### Identification of polymorphisms
***Search for new di-allelic polymorphisms by PCR/SSCP or REF*** From the known genomic structure of the human AGT gene[16] several overlapping sets of oligonucleotides (sequences available upon request) were designed to perform a systematic search for new polymorphisms by Polymerase-Chain Reaction/Single-Strand Conformation Analysis (PCR/SSCA)[17,18] on a 1155-bp region containing exon 5 and a part of the 3′-flanking region with two downstream enhancer core elements (+1399 to +1478 and +2191 to +2214).[19,20]

Restriction endonuclease fingerprinting (REF)[21] was performed as a modification of the SSCA method to detect genetic variants in a 395-bp fragment (Table 2) containing the 24-bp enhancer core element.

***Direct sequencing of electrophoretic variants*** DNA from samples presenting mobility shifts as variant electrophoretic patterns was reamplified by PCR with unlabelled primers and subsequently sequenced, using the chain-termination

**Table 1** Mean (standard deviation) of measured phenotypes in family members

| Study variables[a] | Fathers (n=130) | Mothers (n=130) | Sons (n=135) | Daughters (n=150) |
|---|---|---|---|---|
| Plasma AGT (ng/ml) | 2114 (420) | 2254 (620) | 1832 (369) | 2056 (513) |
| ln(AGT) | 7.6 (0.2) | 7.7 (0.3) | 7.5 (0.2) | 7.6 (0.2) |
| Age (years) | 42.6 (4.6) | 40.4 (4.6) | 14.7 (3.5) | 15.1 (3.7) |
| BMI (kg/m$^2$) | 25.6 (3.4) | 23.2 (3.4) | 19.7 (3.5) | 20.0 (2.4) |
| SBP (mmHg) | 120.92 (12.0) | 107.93 (11.0) | 107.32 (11.4) | 104.14 (10.6) |
| DBP (mmHg) | 73.77 (8.0) | 66.38 (7.77) | 58.13 (6.02) | 59.46 (7.33) |

[a]AGT=Angiotensinogen; BMI=body mass index; SBP=systolic blood pressure; DBP=diastolic blood pressure.

method (Sequenase Version 2.0 DNA Sequencing Kit, Amersham Life Science).[22]

## Genotyping

Genotyping of the whole study population for three di-allelic polymorphisms in the 5′-flanking region (C-532T, A-20C, G-6A), two in exon 2 (M235T, T174M) and two newly identified variants in the untranslated sequence of exon 5 and the 3′-flanking region (C+2054A, C+2127T) of the *AGT* gene was performed by hybridisation of PCR products (Table 2) using allele specific oligonucleotides (ASO) (Table 3) as formerly described.[23]

## Statistical methods

*Analysis of the effects of covariates on AGT* Since the AGT values were skewed, a log transformation of the data was performed which reduced the skewness from 0.85 to 0.09. Prior to combined segregation-linkage analysis, associations of AGT levels with covariates including sex, age, generation (parents/offspring) effects and body mass index (BMI) were investigated by analyses of variance and regression. Correlations of ln(AGT) with systolic (SBP) and diastolic blood pressure (DBP) were also estimated. For SBP and DBP, we used the mean of the 14 measurements (see above). All these analyses considered the observations on family members as independent and were conducted with the BMDP package.

*Preliminary analysis of the effect of AGT polymorphisms on AGT levels* The association of each polymorphism with AGT levels was tested by one-way analysis of variance, sepa-

rately in parents and offspring. The Kruskal-Wallis non-parametric test was used when the sample sizes were too small. This analysis was carried out as a first approach, assuming offspring as independent. The proportion of AGT variance attributable to each polymorphism was calculated by the ratio ($R^2$) of the sum of squares due to the polymorphism to the total sum of squares.

*Marker allele frequencies and linkage disequilibrium* - Marker allele frequencies and pairwise haplotype frequencies were estimated using the whole family information with the ILINK program of the LINKAGE package.[24] Hardy–Weinberg equilibrium of each polymorphism was tested in parents by $\chi^2$ analysis. Linkage disequilibrium between any two polymorphisms was tested by a likelihood ratio test. Pairwise linkage disequilibrium coefficients were expressed in terms of D.[25]

*Combined segregation-linkage analysis* Combined segregation-linkage analysis was conducted by use of the class D regressive model[26,27] extended to take into account linked marker loci[28] with possible linkage disequilibrium. When considering the co-segregation of a trait (Y) and a marker (M), the likelihood of a family is:

$$L\ \text{family} = \sum_{g_Y\ g_M} P(g_Y,\ g_M)\ P(Y,\ M|g_Y,\ g_M,\ X) = \sum_{g_Y\ g_M} P(g_Y,\ g_M)\ f(Y|g_Y,\ X)$$

where $g_Y$ is the vector of underlying genotypes at the unobserved quantitative trait locus (QTL), $g_M$, the vector of genotypes at the observed marker locus and X is a vector of measured covariates. $P(g_Y,\ g_M)$ is the joint probability of QTL and marker genotypes and $f(Y|g_Y,\ X)$ is the penetrance function. The probabilities of the Y and M phenotypes given the genotypes depend on $g_Y$ or $g_M$, respectively and $P(M|g_M,Y,\ X)$ is unity since M is completely determined given $g_M$. The QTL was assumed bi-allelic (d/D), with d and D alleles being responsible for low and high levels of plasma AGT, respectively. All marker polymorphisms are also bi-allelic, with the frequent and rare alleles being denoted generally 1 and 2. For individuals with no ancestors in the pedigree, $P(g_Y,\ g_M)$ is function of the haplotype frequencies (summing to one) which are: p(d-1), p(d-2), p(D-1), p(D-2). Complete linkage disequilibrium (i.e.

**Table 2** Primers used for amplification of the regions containing the analysed polymorphisms

| Location | Primer | Product size (bp) | Temp amplification |
|---|---|---|---|
| 5′-region Exon 1 | 5′-TTCCAGAAGGCACTTTTCAC-3′ 5′-TAGTACCCAGAACAACGGCA-3′ | 594 | 56°C |
| Exon 2 Intron 2 | 5′-GATGCGCACAAGGTCCTGTC-3′ 5′-GCCAGCAGAGAGGTTTGCCT-3′ | 354 | 60°C |
| Exon 5 3′-region | 5′-AATACGTGAAAGATGCAAGC-3′ 5′-GGGGCACAATACCAGCAGCA-3′ | 395 | 55°C |

**Table 3** Primers used for detection of *AGT* polymorphisms by specific oligonucleotide hybridisation

| Location | Primer 1 (Tm) | Primer 2 (Tm) |
|---|---|---|
| −532 | 5′-TGTGTTTT**C**CCCAGTGT-3′ (50°C) | 5′-ACACTGGG**A**AAAACACA-3′ (48°C) |
| −20 | 5′-ATAGGGC**A**TCGTGAC-3′ (46°C) | 5′-ATAGGGC**C**TCGTGAC-3′ (48°C) |
| −6 | 5′-ACCCGGCC**A**GGGGAAGA-3′ (58°C) | 5′-TCTTCCCCC**G**GCCGGGT-3′ (60°C) |
| +521 (T174M) | 5′-CTGTCCA**C**GGTGGTG-3′ (50°C) | 5′-CTGTCCA**T**GGTGGTG-3′ (48°C) |
| +704 (M235T) | 5′-GCTCCTCGA**T**GGGAGCC-3′ (58°C) | 5′-GCTCCCTGA**C**GGGAGCC-3′ (60°C) |
| +2054 | 5′-TGCGGAAC**C**ATAGCTGG-3′ (54°C) | 5-TGCGGAAC**A**ATAGCTGG-3′ (52°C) |
| +2127 | 5′-TTATCTTT**C**ATTTAGCA-3′ (42°C) | 5′-TTATCTTT**T**ATTTAGCA-3′ (40°C) |

Amplification of PCR products were performed with primers shown on Table 2, as described in the Method section.

complete identity between the marker locus and the QTL) corresponds to $p(d-2)=p(D-1)=0$. Intermediate models can be also considered where one of the two marker alleles is in complete gametic disequilibrium with only one QTL allele. When there is no linkage disequilibrium, the haplotype frequencies are the product of allele frequencies. For individuals with ancestors in the pedigree, $P(g_Y, g_M)$ is function of the Mendelian probabilities and the recombination fraction between the 2 loci which is zero here. The penetrance function $f(Y|g_Y, X)$ is a multivariate normal density decomposed in a product of univariate densities by regressing each person's residual phenotype on the residual phenotypes of preceding relatives.[26] Residual phenotypes are AGT levels adjusted for the effects of the QTL and covariates.

The penetrance function is expressed in terms of the following parameters: the three genotype-specific means at the QTL ($\mu_{dd}$, $\mu_{Dd}$, $\mu_{DD}$), the residual variance ($\sigma^2$), assumed to be the same for each genotype but possibly differing by sex (males/females) and generation (parents/offspring), the four phenotypic correlations, father-mother ($\rho_{FM}$), father-offspring ($\rho_{FO}$), mother-offspring ($\rho_{MO}$) and sib-sib ($\rho_{SS}$) and the regression coefficients on covariates which can be genotype-dependent ($\beta_g$'s).

The covariates incorporated in the analysis were the four types of family members (fathers, sons, mothers, daughters) to take into account sex- and/or generation-specific effects of the QTL (i.e. QTL genotype-specific means possibly differing in fathers, mothers, sons and daughters). At a later stage, they also included the polymorphism with the largest effect on plasma AGT to search for additional effects of the remaining polymorphisms.

Parameter estimation and test of hypotheses were carried out by use of maximum-likelihood methods as implemented in the computer program REGRESS,[29] which incorporates the regressive approach in the ILINK program of the LINKAGE package[24] and includes the GEMINI optimization procedure.[30] The likelihood of the family sample was maximised under different models, as summarised in Table 4. In a first step (step A), we assumed complete identity between the QTL and marker locus to test and estimate the effect of the marker on the quantitative trait, which is equivalent to the measured genotype analysis described by Boerwinkle *et al*.[31]

Four classes of models were considered: the first class of model (A.I) is a sporadic model with no marker effect and no familial correlation (FC); the second class of model (A.II) includes FC but no marker effect; the third class of model (A.III) includes only a marker effect without FC; the fourth class of model (A.IV) includes both a marker effect and FC. Evidence for a marker effect is obtained by rejecting model A.II against model A.IV and presence of residual familial correlations is tested by comparing model A.III to model A.IV. If there is a significant effect of the marker on the quantitative trait, combined-segregation analysis is pursued (step B) to test the type of linkage disequilibrium between the marker and QTL. Mainly, two models, complete linkage disequilibrium (B.I) and absence of linkage disequilibrium between the QTL and the marker (B.II), are tested against the general model (B.III) where all haplotype frequencies are estimated together with the other parameters. Additional models including sex- (males/females) and generation- (parents/offspring) specific effects as well as dominance/recessivity of the QTL were also considered. Interactions between the QTL and covariates were tested by comparing sub-models where the regression coefficients, $\beta_g$'s, were set equal to the same estimate $\beta$ (no interaction) versus models in which three (or two) $\beta_g$'s were estimated (interaction).

**Table 4** Sub-models and corresponding parameters considered in measured genotype analysis and combined segregation-linkage analysis under the general class D regressive model

| Model | $p(d-1)$ | $p(d-2)$ | $p(D-1)^b$ | $\mu_{dd}$ | $\mu_{Dd}$ | $\mu_{DD}$ | $\sigma^2$ | $\rho_{FM}$ | $\rho_{FO}$ | $\rho_{MO}$ | $\rho_{SS}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *A. Measured genotype analysis*[c] | | | | | | | | | | | |
| I. Sporadic model | * | [0] | [0] | * | $=\mu_{dd}$ | $=\mu_{dd}$ | * | [0] | [0] | [0] | [0] |
| II. No marker effect, FC[d] | * | [0] | [0] | * | $=\mu_{dd}$ | $=\mu_{dd}$ | * | * | * | * | * |
| III. Marker effect, no FC | * | [0] | [0] | * | * | * | * | [0] | [0] | [0] | [0] |
| IV. Marker effect+FC | * | [0] | [0] | * | * | * | * | * | * | * | * |
| | | | | | | | | | | | |
| *B. Segregation-linkage analysis* | | | | | | | | | | | |
| I. Complete LD[e] between QTL and marker | * | [0] | [0] | * | * | * | * | * | * | * | * |
| II. Absence of LD between QTL and marker | [f] $p_d \times p_1$ | $p_d \times p_2$ | $p_D \times p_1$ | * | * | * | * | * | * | * | * |
| III. General model | * | * | * | * | * | * | * | * | * | * | * |

[a]All parameters defined in the text; parameters in square brackets are fixed at a value specified by the model, the asterisk * indicates the parameters which are estimated. Covariates can also be included in the model with corresponding regression coefficients. [b]Three of the four haplotype frequencies are estimated (or fixed) since their sum is one. [c]Measured genotype analysis is a particular case of combined segregation-linkage analysis where there is identity between the marker and the QTL. [d]FC=familial correlations in addition to the QTL effect expressed in terms of four parameters: spouse ($\rho_{FM}$), father-offspring ($\rho_{FO}$), mother-offspring ($\rho_{MO}$), and sib-sib ($\rho_{ss}$) correlations. [e]LD=linkage disequilibrium. [f]$p_d$, $p_D$ are the QTL allele frequencies ($p_d+p_D=1$); $p_1$, $p_2$ are the marker allele frequencies ($p_1+p_2=1$).

## Results

### Polymorphism detection

Two new polymorphisms (C+2054A and C+2127T) were identified near an enhancer element (+2191 to +2214)[19] within the untranslated sequence of exon 5 and the 3′-flanking region at position +2054 and +2127, respectively (Table 3, Figure 1). The C+2054A polymorphism is located within a putative AP-4 consensus site. Since, both polymorphisms were completely associated, the analysis was only performed on the C+2054A variant.

### Analysis of variance and correlation analyses

The clinical characteristics of the family members are shown in Table 1, including means and standard deviations of plasma AGT levels, age, body mass index (BMI), systolic (SBP) and diastolic blood pressure (DBP). AGT means are significantly higher in parents than in offspring ($P<0.0001$) and in daughters than in sons ($P<0.0001$). Means of SBP and DBP are both higher in parents than in children ($P<0.0001$) and in fathers than in mothers ($P<0.0001$). In children, SBP mean is significantly higher in sons than in daughters ($P=0.01$), whereas there is no difference for DBP. In the whole sample, plasma AGT is significantly correlated with SBP ($r=0.22$, $P<0.001$) and DBP ($r=0.25$, $P<0.001$). These correlations are of the same order of magnitude and significant in each group of family members defined by sex and generation, except that, in mothers and sons, the correlation between AGT and DBP does not differ significantly from zero ($r=0.14$ in each of these two groups).

Given the sex and generation differences of ln(AGT), association of AGT levels with the relevant covariates, age and BMI, were investigated separately in the four groups of family members: fathers, mothers, sons and daughters. No significant effect of age and BMI was found in either

group of subjects. Thus, ln(AGT) was adjusted for the sex- and generation-specific means and standardised within each of the four groups of family members.

### Linkage disequilibrium between *AGT* variants

None of the *AGT* gene polymorphisms showed significant deviation from Hardy–Weinberg expectations. As shown in Table 5, all polymorphisms are in strong linkage disequilibrium ($P<0.0001$) with each other, the pairwise linkage disequilibrium coefficients D being all greater than 0.80.

### Effects of AGT polymorphisms by analysis of variance

Associations of each polymorphism with the adjusted ln(AGT) levels are shown in Table 6. C-532T is the only polymorphism significantly associated with AGT in both parents ($P<0.005$) and offspring ($P<0.005$) and explaining respectively 5.1% and 7.0% of the variance.

The G-6A and and M235T polymorphisms have a significant effect on AGT in parents only, accounting for 3.5% and 4.1% of the variance, respectively. However, there is no significant association of T174M and C+2054A polymorphisms with AGT. We also checked that all these marker effects were similar with raw AGT values.

### Combined segregation-linkage analysis

The main outcomes of combined segregation-linkage analyses of AGT levels with each of the five polymorphisms are summarised in Table 7. These included the test of the marker effect in measured genotype analysis (model A.II vs model A.IV of Table 4) and the tests of absence of disequilibrium and complete linkage disequilibrium against the general model (models B.II and B.I vs B.III of Table 4).

The measured genotype analysis showed first that familial correlations of AGT levels without a marker effect were highly significant when compared to a sporadic model
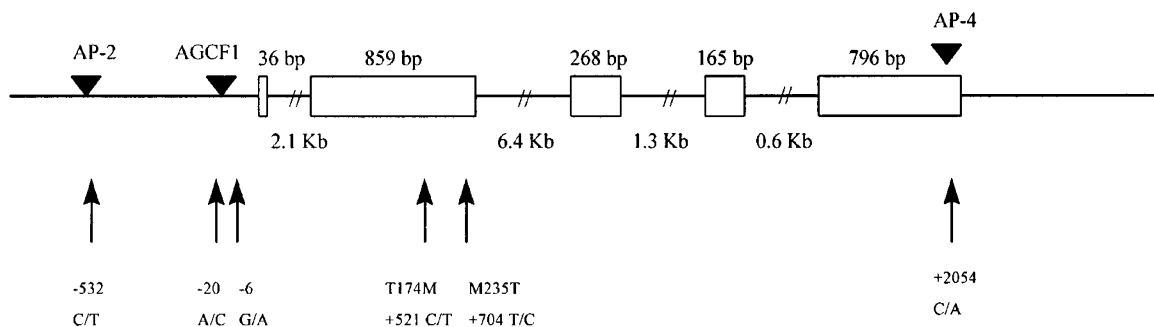


**Figure 1** Schematic representation of the human angiotensinogen gene (12 kb) and position of analysed polymorphisms. The localisation of the variants in the 5′-flanking region (C-532T, A-20C, G-6A) are numbered by reference to the transcription-initiation site as defined previously.[37] The two polymorphisms of exon 2 (T174M, M235T) are at amino acid residues 174 and 235 and nucleotide position +521 and +704. The newly identified polymorphisms in exon 5 and the 3′-flanking region are located at position +2054 and +2127, respectively. Exon 1 to 5 are represented by open boxes. Their sizes are marked on the top of the boxes. Introns and flanking regions are shown by thin lines. Intron sizes are shown below the line. Only putative regulatory elements located on polymorphic sites are indicated. Triangles indicate consensus sites for transcriprion factor AP-2, AP-4, and the human angiotensinogen core promoter binding factor 1 (AGCF1).

($P<0.000001$) with estimates of father-offspring correlation, $\rho_{FO}=0.36$, mother-offspring correlation, $\rho_{MO}=0.28$, and sib-sib correlation, $\rho_{SS}=0.42$ (results not shown). In the presence of these correlations, all polymorphisms, except T174M, had a significant effect on plasma AGT (second column of Table 7). The C-532T variant had the highest effect ($P=0.000001$), followed by M235T ($P=0.00002$), G-6A ($P=0.0002$) and C+2054A ($P=0.0005$). The proportion of AGT variance due to each polymorphism is of the same order of magnitude as that obtained previously by analysis of variance, being the highest for C-532T (4.3% in parents and 5.5% in offspring). We noted that the C+2054A polymorphism had an inverse effect as compared to the others, with the frequent allele being associated with high AGT levels rather than the rarer allele. The effect of each polymorphism did not differ significantly in fathers, mothers, sons and daughters and dominant and recessive modes of inheritance for this effect were rejected against a more general codominant model. Moreover, residual familial correlations were still highly significant ($P<0.000001$) with estimates similar to those obtained under a model without a polymorphism effect, which is due to the small effect of these genetic variants.

As compared with the general linkage disequilibrium model the hypothesis of no linkage disequilibrium between the QTL and each polymorphism was strongly rejected

whereas the hypothesis of complete identity between the QTL and either polymorphism is never rejected (columns 3 and 4 of Table 7). However, estimates of the haplotype frequencies (right end of Table 7) indicate that C-532T is the only polymorphism almost confounded with the putative QTL, the C allele being almost exclusively associated with putative low-AGT level allele (d) and the T allele with putative high-level allele (D). Moreover, under a more general model which specified that the means of AGT levels could differ not only according to the QTL genotype but also according to the sex and generation of family members (fathers, mothers, sons and daughters) within each genotype, the estimates of haplotype frequencies are in agreement with complete identity between the C-532T and the QTL. The G-6A and M235T variants have only their rarer allele never associated with the putative low-level (d) allele. Comparison of these haplotype frequencies to those estimated between the C-532T variant and each other polymorphism suggests there may be more than one polymorphism within the *AGT* gene controlling AGT levels. Indeed, the C-532 allele is almost exclusively associated with the putative (d) allele, whereas the -6A allele is only associated with the putative D allele while the estimated frequency of (-532C/-6A) haplotype is 0.33 in this study.

To test for additional effects of polymorphisms besides that of C-532T, combined segregation-analysis was repeated with G-6A, M235T and C+2054A by either adjusting AGT levels for the effect of C-532T prior to the analysis or including C-532T as a covariate in the model (Table 8). The effects of these three variants are still significant but in a lesser extent than before (at about 5% level for G-6A and at 1% level for M235T and C+2054A) and residual familial correlations are still highly significant ($P<0.000001$). Again, the hypothesis of no linkage disequilibrium is rejected and a model of complete linkage disequilibrium fits well with the data, although the haplotype frequencies show similar patterns as before, with -6A and 235T alleles being never associated with the putative

**Table 5** Allele frequencies (rarer alleles) and pairwise linkage disequilibrium coefficients (D′) between *AGT* polymorphisms

| Polymorphism | Allele frequency | G-6A | C-532T | T174M | M235T |
|---|---|---|---|---|---|
| G-6A | 0.43 | – | – | – | – |
| C-532T | 0.10 | 0.96 | – | – | – |
| T174M | 0.12 | 0.96 | 0.83 | – | – |
| M235T | 0.43 | 0.94 | 0.95 | 1.0 | – |
| C+2054A | 0.32 | 0.85 | 0.94 | 0.82 | 0.91 |

**Table 6** Association between ln(AGT) and polymorphisms of *AGT* in parents and offspring assuming observations are independent

| Polymorphism | | Genotypic mean (SD)/no. of subjects | | | % of variance | |
| | | 11[a] | 12 | 22[b] | $R^2$ | P value |
|---|---|---|---|---|---|---|
| G-6A | Parents | −0.18 (1.0)/71 | 0.01 (0.99)/115 | 0.37 (0.83)/40 | 3.5 | 0.02 |
| | Offspring | −0.11 (1.08)/69 | −0.03 (0.96)/140 | 0.31 (1.09)/33 | 1.7 | 0.14 |
| C-532T | Parents | −0.11 (0.99)/182 | 0.44 (0.92)/41 | 0.82 (−)[c]/2 | 5.1 | 0.005[d] |
| | Offspring | −0.08 (0.96)/203 | 0.26 (1.05)/40 | 2.57 (−)[c]/2 | 7.0 | 0.005[d] |
| T174M | Parents | 0.06 (1.01)/172 | −0.17 (0.98)/49 | −0.48 (−)[c]/4 | 1.3 | 0.19[d] |
| | Offspring | −0.91 (1.04)/182 | 0.25 (0.88)/61 | −0.01 (−)[c]/3 | 2.1 | 0.06[d] |
| M235T | Parents | −0.25 (1.06)/72 | 0.04 (0.97)/114 | 0.34 (0.85)/38 | 4.1 | 0.01 |
| | Offspring | −0.17 (1.07)/66 | 0.02 (0.95)/140 | 0.32 (1.07)/34 | 2.2 | 0.07 |
| C+2054A | Parents | 0.09 (0.97)/98 | −0.02 (1.03)/111 | −0.30 (0.96)/17 | 1.1 | 0.30 |
| | Offspring | 0.15 (1.04)/108 | −0.06 (0.96)/119 | −0.35 (1.01)/19 | 2.0 | 0.08 |

[a]1 denotes presence of the frequent allele. [b]2 denotes presence of the rare allele. [c]Standard deviation is irrelevant given the small sample size. [d]Kruskal-Walllis non-parametric test.

**Table 7** Combined segregation-linkage analysis of ln(AGT) and five polymorphisms of the *AGT* locus

| | Chi-squares values corresponding to the tests of different hypotheses[a] | | | Haplotype frequencies[c] | | | |
|---|---|---|---|---|---|---|---|
| Polymorphism | No polymorphism effect vs effect ($\chi^2_2$) | Absence of LD[b] vs general LD model ($\chi^2_1$) | Complete LD[b] vs general LD model ($\chi^2_2$) | d-1 | d-2 | D-1 | D-2 |
| G-6A | 17.49 (P=0.0002) | 17.65 (P=0.00003) | 1.81 (P=0.40) | 0.28 | 0.0 | 0.29 | 0.43 |
| | | | | *0.38* | *0.0* | *0.18* | *0.43* |
| C-532T | 27.98 (P=0.000001) | 22.39 (P=0.000002) | 0.27 (P=0.87) | 0.89 | 0.02 | 0.01 | 0.08 |
| | | | | *0.90* | *0.00* | *0.00* | *0.10* |
| T174M | 4.13 (P=0.13) | – | – | | | | |
| M235T | 21.98 (P=0.000002) | 19.93 (P=0.000008) | 2.48 (P=0.29) | 0.33 | 0.0 | 0.24 | 0.43 |
| | | | | *0.42* | *0.00* | *0.15* | *0.43* |
| C+2054A | 15.31 (P=0.0005) | 13.43 (P=0.0003) | 1.20 (P=0.55) | 0.50 | 0.12 | 0.18 | 0.20 |
| | | | | *0.25* | *0.25* | *0.42* | *0.08* |

[a]All tests shown here were conducted in presence of residual familial correlations which were highly significant. [b]LD=linkage disequilibrium. [c]d and D are the alleles of the putative functional variant responsible for low and high AGT levels for all polymorphisms, except for C+2054A where it is the contrary; 1 and 2 represent the frequent and rare alleles of each polymorphism, respectively. The first line shows the estimates of haplotype frequencies under a sub-model with no sex- and generation-specific effects of the QTL, which fits the data, whereas the second line (in italics) is the more general model including these effects (i.e. effect of QTL differing in fathers, mothers, sons and daughters within each genotype).

**Table 8** Combined segregation-linkage analysis of ln(AGT) and polymorphisms of the *AGT* gene when adjusting for the effect of the C-532T polymorphism prior to the analysis (1st line) or introducing it as a covariate in the model of analysis (2nd line)

| | Chi-squares values corresponding to the tests of different hypotheses[a] | | |
|---|---|---|---|
| Polymorphism | No polymorphism effect vs effect ($\chi^2_2$) | Absence of LD[b] vs general LD model ($\chi^2_1$) | Complete LD vs general LD model ($\chi^2_2$) |
| G-6A | 6.89 (P=0.032) | 6.61 (P=0.011) | 3.12 (P=0.21) |
| | 6.13 (P=0.047) | 9.11 (P=0.003) | 3.92 (P=0.14) |
| M235T | 10.14 (P=0.006) | 12.17 (P=0.0005) | 5.17 (P=0.08) |
| | 9.72 (P=0.008) | 13.23 (P=0.0003) | 5.10 (P=0.08) |
| C+2054A | 10.09 (P=0.006) | 9.79 (P=0.0002) | 1.99 (P=0.37) |
| | 9.24 (P=0.009) | 8.00 (P=0.005) | 0.94 (P=0.63) |

[a]All tests shown here were conducted in presence of residual familial correlations which were highly significant. [b]LD=linkage disequilibrium.

functional allele for low AGT levels. When C-532T was included as a covariate in the analysis, we checked that its effect in presence of the tested polymorphism remained always significant. Interactions between the effects of C-532T and either tested variant were not significant.

## Discussion

Because AGT level is a stable quantitative trait which is the most closely related to the *AGT* gene, it was used as a phenotype in our study which was aimed at searching for candidate functional variant(s) within *AGT*, which may regulate AGT expression. In these healthy nuclear families, typically composed of two parents and two children, plasma AGT was significantly higher in the parental generation, both in men and women. However, within generation there was no correlation with age. A positive correlation between plasma AGT and both DBP and SBP in parents and children was found in our study. This is in agreement with previous results from Walker *et al.*,[32] who found a significant correlation between recumbent DBP and AGT concentration in subjects with a mean age of $31.3 \pm 0.4$ years.

All polymorphisms we studied were located in the *AGT* gene, within a region of 12 kb between position -532 upstream from the transcription start to position +2127 in the 3'-flanking region. This explains the strong linkage disequilibrium found between these polymorphisms.

Preliminary analyses of variance showed a strong association of the C-532T polymorphism with plasma AGT in both parents and children, whereas G-6A and M235T had less significant effects only in parents and no association was detected for the two remaining markers. The G-6A polymorphism, previously considered as a candidate functional variant,[12] explained a smaller part of the AGT variance than C-532T, although the frequency of the allele responsible for high AGT levels is much higher (0.43 *vs* 0.10).

Combined segregation-linkage analysis shows highly significant effects of all polymorphisms, except for T174M. This confirms the high power of this approach to detect candidate genes with small effects underlying a complex trait.[33] Indeed, the proportion of total phenotypic variance explained by these different polymorphisms varies from 0.2% to 4.3% in parents and from 1.1% to 5.5% in

offspring. Moreover, the estimates of residual familial correlations, which are highly significant, underline the substantial role of other genes and/or shared environmental factors influencing AGT levels. We noted, that the pattern of these correlations did not differ significantly from that specified by a polygenic model where the parent-offspring and sib-sib correlations are equal ($\rho_{FO}=\rho_{MO}=\rho_{SS}=0.33$).

The hypothesis of complete confounding of either polymorphism with the putative functional variant is never rejected, although estimates of haplotype frequencies do not show such confounding. This may be due to a lack of power, as found in a simulation study.[34] The C-532T polymorphism shows the strongest effect and is confounded with the putative functional variant when sex- and generation-specific effects of this variant are taken into account. This designates C-532T as the best candidate for further functional studies. The C allele is associated with the putative low-level allele and the T allele with the putative high-level allele. Taking into account the effect of this variant by adjusting AGT levels prior to the analysis or including C-532T as a covariate in the model leads to residual significant effects of G-6A, M235T and C+2054A, the two latter being more significant. Interestingly, the two approaches lead to similar results. The rarer alleles of G-6A and M235T are never associated with the putative functional allele responsible for low AGT levels. Interactive effects of C-532T with any of the polymorphisms are not significant, which may be also due to a lack of power. Indeed, including an interaction in the model, leads to an effect of either tested polymorphism, G-6A, M235T or C+2054A, which is higher in subjects bearing at least one -532T allele as opposed to those bearing only C-532 alleles. A polymorphism (A-20C), located in the proximal 5′-flanking region, was shown to be associated with differences in plasma AGT in a Japanese population, and to be associated with *in vitro* changes in transcription level induced by *AGT* promoters.[35] This polymorphism is also in linkage disequilibrium with the M235T polymorphism.[36] Therefore, this polymorphism was also tested by analysis of variance and by segregation-linkage analysis and no significant effect was found to be associated with this polymorphism. Altogether, these results indicate that there are at least two functional variants within the *AGT* gene controlling part of AGT variation with other yet-unknown familial factors playing an important role.

In view of the correlation between plasma AGT and blood pressure, genes affecting plasma AGT may affect blood pressure. The effect of the C-532T polymorphism was tested on SBP and DBP by analysis of variance and indicated only a borderline significant association with DBP. Segregation-linkage analysis carried out on DBP, adjusted for relevant covariates, did not show any significant effect. Since C-532T explains about 5% of plasma AGT variance and the correlation between DBP and AGT is at most 0.21 in fathers and daughters, there may not be enough power

to detect a significant effect of this polymorphism on blood pressure. Alternatively, these polymorphisms may have no direct effect on blood pressure but may act indirectly by controlling AGT variability.

This study clearly designates the C-532T polymorphism as a strong candidate for having a functional role on genetic determination of AGT expression and accounting for approximately 5% of AGT variability. However, our analyses suggest a more complex model than a single functional diallelic variant, involving more likely a combination of variants within the *AGT* gene modulating gene expression. Combined segregation-linkage analysis based on regressive models appears to be a powerful tool to detect candidate genetic variants involved in a complex trait and urge to test *in vitro* the functional role of candidate polymorphisms, alone and in combination.

## References

1 Brasier AR, Li J: Mechanisms for inducible control of angiotensinogen gene transcription. *Hypertension* 1996; **27**: 465–475.

2 Jeunemaitre X, Soubrier F, Kotelevtsev YV *et al*: Molecular basis of human hypertension: role of angiotensinogen. *Cell* 1992; **71**: 169–180.

3 Hata A, Namikawa C, Sasaki M *et al*: Angiotensinogen as a risk factor for essential hypertension in Japan. *J Clin Invest* 1994; **93**: 1285–1287.

4 Iwai N, Shimoike H, Ohmichi N, Kinoshita M: Angiotensinogen gene and blood pressure in the Japanese population. *Hypertension* 1995; **25**: 688–693.

5 Johnson AG, Simons LA, Friedlander Y, Simons J, Davis DR, MaCallum J: M235– >T polymorphism of the angiotensinogen gene predicts hypertension in the elderly. *J Hypertens* 1996; **14**: 1061–1065.

6 Caulfield M, Lavender P, Farrall M *et al*: Linkage of the angiotensinogen gene to essential hypertension. *N Engl J Med* 1994; **330**: 1629–1633.

7 Rotimi C, Morrison L, Cooper R *et al*: Angiotensinogen gene in human hypertension. Lack of an association of the 235T allele among African Americans. *Hypertension* 1994; **24**: 591–594.

8 Fornage M, Turner ST, Sing CF, Boerwinkle E: Variation at the M235T locus of the angiotensinogen gene and essential hypertension: a population-based case-control study from Rochester, Minnesota. *Hum Genet* 1995; **96**: 295–300.

9 Hingorani AD, Sharma P, Jia H, Hopper R, Brown MJ: Blood pressure and the M235T polymorphism of the angiotensinogen gene. *Hypertension* 1996; **28**: 907–911.

10 Brand E, Chatelain N, Keavney B *et al*: Evaluation of the angiotensinogen locus in human essential hypertension: an European study. *Hypertension* 1998; **31**: 725–729.

11 Niu T, Xu X, Rogus J *et al*: Angiotensinogen gene and hypertension in Chinese. *J Clin Invest* 1998; **101**: 188–194.

12 Inoue I, Nakajima T, Williams CS *et al*: A nucleotide substitution in the promoter of human angiotensinogen is associated with essential hypertension and affects basal transcription in vitro. *J Clin Invest* 1997; **99**: 1786–1797.

13 Jeunemaitre X, Inoue I, Williams C *et al*: Haplotypes of angiotensinogen in essential hypertension. *Am J Hum Genet* 1997; **60**: 1448–1460.

14 Plouin PF, Chatellier G, Guyene TT, Vincent N, Corvol P: Recent advances in the clinical study of the renin system. Reference values and conditions of validity. *Presse Med* 1989; **18**: 917–921.

15 Marcadet A, O'Connell P, Cohen D: Standardized southern blot workshop technique; in Dupont B (ed). *HistocompatibilityTesting*. New York: Springer, 1986, vol 1 pp 553–560.

16 Fukamizu A, Takahashi S, Seo MS *et al*: Structure and expression of the human angiotensinogen gene. Identification of a unique and highly active promoter. *J Biol Chem* 1990; **265**: 7576–7582.

17 Orita M, Suzuki Y, Sekiya T, Hayashi K: Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 1989; **5**: 874–879.

18 Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM: The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* 1993; **16**: 325–332.

19 Nibu Y, Takahashi S, Tanimoto K, Murakami K, Fukamizu A: Identification of cell type-dependent enhancer core element located in the 3′-downstream region of the human angiotensinogen gene. *J Biol Chem* 1994; **269**: 28598–28605.

20 Nibu Y, Tanimoto K, Takahashi S, Ono H, Murakami K, Fukamizu A: A cell type-dependent enhancer core element is located in exon 5 of the human angiotensinogen gene. *Biochem Biophys Res Commun* 1994; **205**: 1102–1108.

21 Liu Q, Sommer SS: Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long, contiguous segments of DNA. *Biotechniques* 1995; **18**: 470–477.

22 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463–5467.

23 Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA: Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 1986; **324**: 163–166.

24 Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984; **81**: 3443–3446.

25 Thompson EA, Deeb S, Walker D, Motulsky AG: The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am J Hum Genet* 1988; **42**: 113–124.

26 Bonney GE: On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am J Med Genet* 1984; **18**: 731–749.

27 Demenais FM, Murigande C, Bonney GE: Search for faster methods of fitting the regressive models to quantitative traits. *Genet Epidemiol* 1990; **7**: 319–334.

28 Bonney GE, Lathrop GM, Lalouel JM: Combined linkage and segregation analysis using regressive models. *Am J Hum Genet* 1988; **43**: 29–37.

29 Demenais F, Lathrop M: REGRESS: a computer program including the regression approach into the linkage package. *Genet Epidemiol* 1994; **11**: 291.

30 Lalouel JM: GEMINI-a computer program for optimization of non-linear assumptions. Department of Medical Biophysics and Computing, University of Utah, Technical Report # 14, 1979.

31 Boerwinkle E, Chakraborty R, Sing CF: The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 1986; **50**: 181–194.

32 Walker WG, Whelton PK, Saito H, Russell RP, Hermann J: Relation between blood pressure and renin, renin substrate, angiotensin II, aldosterone and urinary sodium and potassium in 574 ambulatory subjects. *Hypertension* 1979; **1**: 287–291.

33 Martinez M, Abel L, Demenais F: How can maximum likelihood methods reveal candidate gene effects on a quantitative trait?. *Genet Epidemiol* 1995; **12**: 789–794.

34 Chatelain N, Rosenberg-Bourguin M, Demenais F: Effect of family selection on the detection of a candidate gene in a complex quantitative trait. *Genet Epidemiol* 1996; **13**: 311.

35 Zhao YY, Zhou J, Kumar A: Molecular mechanism of human essential hypertension. A mutation in the proximal promoter of the angiotensinogen gene from hypertensive patients increases its transcription in human hepatoma cells. *Circulation* 1996; **94**: I–693.

36 Sato N, Katsuya T, Rakugi H *et al*: Association of variants in critical core promoter element of angiotensinogen gene with increased risk of essential hypertension in Japanese. *Hypertension* 1997; **30**: 321–325.

37 Gaillard I, Clauser E, Corvol P: Structure of human angiotensinogen gene. *DNA* 1989; **8**: 87–99.