



# Obstetric complication-associated *ANXA5* promoter polymorphisms may affect gene expression via DNA secondary structures

Hidehito Inagaki<sup>1</sup> · Sayuri Ota<sup>1</sup> · Haruki Nishizawa<sup>2</sup> · Hironori Miyamura<sup>2</sup> · Kumiko Nakahira<sup>3</sup> · Machiko Suzuki<sup>1</sup> · Sachie Nishiyama<sup>2</sup> · Takema Kato<sup>1</sup> · Itaru Yanagihara<sup>3</sup> · Hiroki Kurahashi<sup>1</sup>

Received: 25 October 2018 / Revised: 1 February 2019 / Accepted: 4 February 2019 / Published online: 22 February 2019  
© The Author(s) under exclusive licence to The Japan Society of Human Genetics 2019

## Abstract

Recent findings have highlighted the possibility that polymorphisms within the annexin A5 gene (*ANXA5*) promoter contribute to the etiology of various obstetric complications. However, the underlying mechanisms are unknown. The M2 haplotype of the *ANXA5* shows lower activity and less expression of *ANXA5* mRNA. This gene promoter region has a motif that potentially forms a G-quadruplex structure. In vitro G-quadruplex propensity estimated by circular dichroism indicated that the M2 haplotype oligonucleotide manifested a decreased potential for G-quadruplex formation. In addition, in vivo G-quadruplex formation of the promoter region was evidenced by the presence of single-stranded DNA shown by sodium bisulfite treatment of placental genomic DNA. Comparative analysis indicated less potential in the M2 allele than the major allele. Promoter activity of the two haplotypes determined by luciferase reporter analysis correlated with the estimated G-quadruplex propensity. Our data lend support to the developing paradigm that genomic variation affects gene expression levels via DNA secondary structures leading to the disease susceptibility.

## Introduction

Many common human diseases are believed to be polygenic disorders associated with several genetic and environmental factors [1]. Based on the “common disease-common variant” hypothesis, genome-wide association studies (GWAS) using common single nucleotide polymorphisms (SNPs) have identified hundreds of genetic variants that are statistically associated with different target diseases. However, most of the polymorphisms that have been identified by

GWAS to date are not deleterious variants and confer relatively small increases in disease risk. In addition, the functional impact of the majority of these SNPs on gene expression has not yet been validated since these sequence variations are mostly located at non-coding or intergenic regions [2]. The processes by which these SNPs confer a higher risk of disease thus remain an enigma.

Previously published results suggest that polymorphisms within the annexin A5 gene (*ANXA5*) are associated with common obstetric complications, such as recurrent pregnancy loss (RPL), pre-eclampsia, and pregnancy-related thrombophilic disorder [3–8]. The *ANXA5* gene upstream region contains four common variations, i.e., SNP1 (g. –467G>A, rs112782763) and SNP2 (g. –448A>C, rs28717001) in the untranscribed promoter region, and SNP3 (g. –422T>C, rs28651243) and SNP4 (g. –373G>A, rs113588187) near and downstream of the transcription start points, respectively. These four SNPs manifest strong linkage disequilibrium, generating two major haplotypes: the N haplotype with all major alleles and an M2 haplotype with all minor alleles. The frequency of the M2 allele in the general Japanese population was reported to be 5.4%, lower than that in the western countries (~16%) [6].

The M2 haplotype has been associated with various disorders. Annexin A5 is known as a placental

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s10038-019-0578-4>) contains supplementary material, which is available to authorized users.

✉ Hiroki Kurahashi  
kura@fujita-hu.ac.jp

- <sup>1</sup> Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan
- <sup>2</sup> Department of Obstetrics and Gynecology, Fujita Health University School of Medicine, Toyoake, Aichi 470-1192, Japan
- <sup>3</sup> Department of Developmental Medicine, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Osaka 594-1101, Japan

anticoagulation factor that shields the apical surface of the syncytiotrophoblasts covering the placental villi [9]. Hence, the low expression of the *ANXA5* gene might reasonably account for a higher susceptibility to obstetric complications. Indeed, lower expression of the *ANXA5* gene from the M2 allele has been reported [10, 11]. However, it remains unclear how these polymorphisms affect the *ANXA5* gene expression levels and thereby lead to disease. For example, there was no association found between the M2 haplotype and RPL risk in a previous northern European study [12]. It is therefore critically important to elucidate the biological impact of the different *ANXA5* haplotypes.

With regard to the effects of gene polymorphisms on gene expression, the contribution of DNA secondary structures has been highlighted previously [13, 14]. A considerable body of evidence now indicates that certain SNPs located within the transcribed region of a gene can affect mRNA stability or translation efficiency via the propensity for stem-loop formation. Guanine-rich DNA can fold into a non-canonical DNA structure known as a G-quadruplex [15]. This structure comprises intrastrand interactions of G-tetrads paired by Hoogsteen bonds. G-quadruplexes are often identified in and around the untranslated region of genes and are potentially associated with gene regulation [16, 17]. The association between the G-quadruplex structure and transcriptional regulation has been extensively characterized for oncogenes such as *MYC*, *KIT*, or *KRAS* [18–20]. Recently, it has also been reported that some polymorphisms disrupt the formation of G-quadruplex structures, leading to alterations in the expression of nearby genes [21]. In our present study, we examined the *ANXA5* gene promoter variants in terms of the association between the potential for G-quadruplex formation and the *ANXA5* gene expression levels.

## Methods

### Samples and ethical approval

Four placental samples were obtained from women with an uncomplicated pregnancy at the Department of Obstetrics and Gynecology, Fujita Health University Hospital. Informed consent was obtained from each patient and the study protocol was approved by the Ethical Review Board for Human Genome Studies at the Fujita Health University. The genotypes of the SNPs at the *ANXA5* promoter region were determined by sequencing as previously described [6]. All methods were performed in accordance with the relevant guidelines and regulations including a biosafety regulation in Japan, the Act on the Conservation and Sustainable Use of Biological Diversity through Regulations on the Use of Living Modified Organisms.

### Circular dichroism

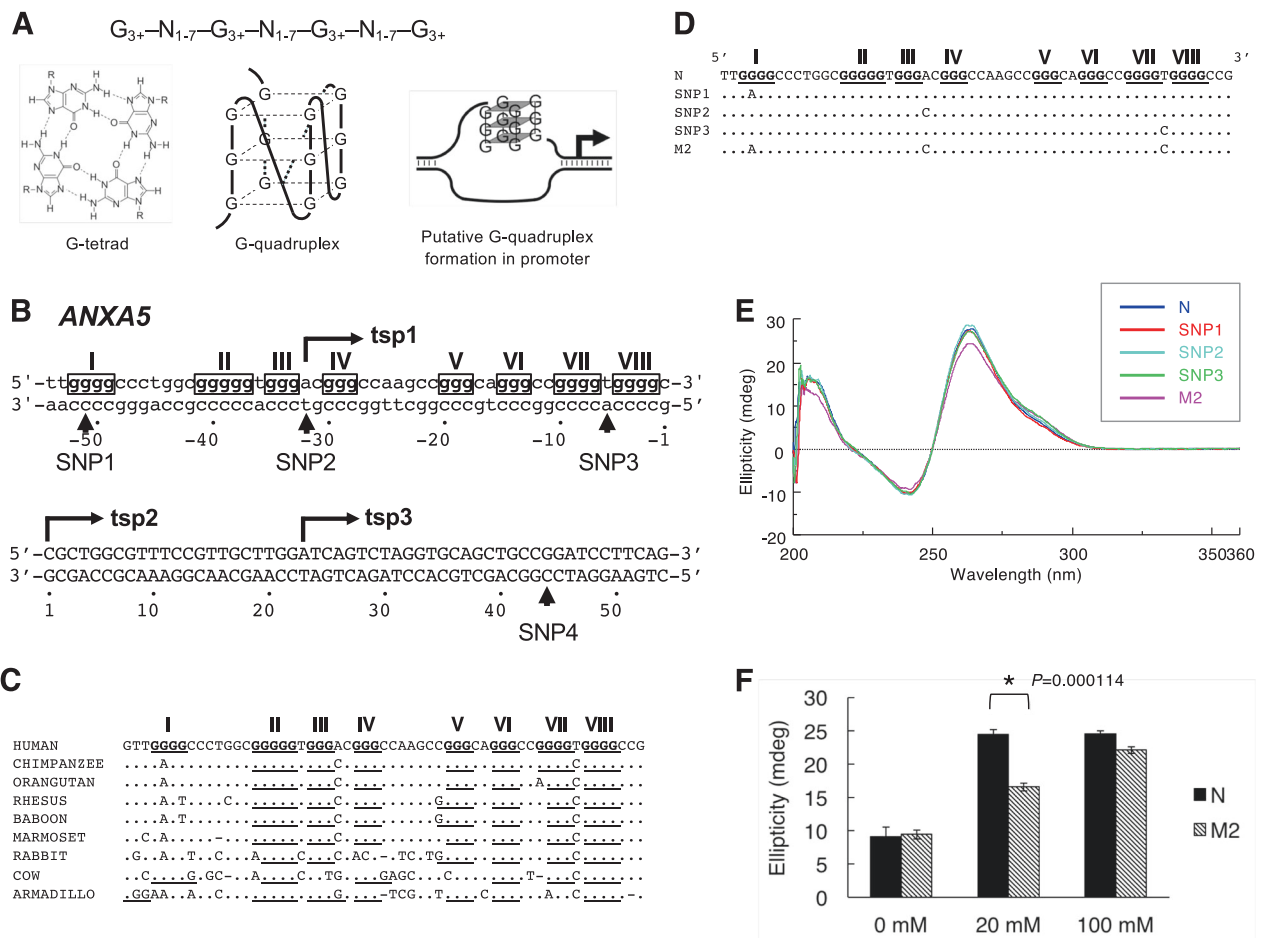
Circular dichroism (CD) experiments were performed using a J-720 spectropolarimeter (JASCO). Oligonucleotides were synthesized and diluted to 50 ng/μl with a buffer containing 100 mM potassium chloride and 10 mM Tris-HCl (pH 7.4). Where indicated, the potassium chloride concentration was decreased, and lithium chloride was added to adjust the salt concentration to 100 mM. The samples were next heat-denatured at 95 °C for 5 min and then cooled slowly for 6 h to 25 °C. Scans were performed at 25 °C using a 1 cm cuvette over a 200–360 nm range. CD spectra were recorded from the average of five scans at 50 nm/min, with a 2 s response time, 1 nm bandwidth, and 0.1 nm resolution. The molar ellipticity was then plotted. The 56 oligonucleotides used in this CD analysis are described in Fig. 1c.

### Methylation analysis

Bisulfite sequencing was performed to determine the methylation status of the *ANXA5* promoter region. One N/N homozygous and three N/M2 heterozygous samples were analyzed. Bisulfite conversion was performed using an EpiTect Bisulfite Kit (Qiagen, Tokyo, Japan) in accordance with the manufacturer's protocol. Bisulfite-treated DNA was amplified using an uracil stalling-free polymerase, EpiTaq HS (Takara bio, Kusatsu, Japan) and the primers: 5'-GGTTATA-GAGGGTAGGGAGGTTTAA-3' and 5'-CACCCAACTA-TAAAACCCAAATAC-3'. The ~300 bp resulting products were then cloned into the pT7Blue T-vector (Merck, Darmstadt, Germany). Colonies were subsequently selected, and the plasmids were isolated for sequencing.

### Bisulfite sequencing for detection of DNA secondary structure

Bisulfite treatment was applied to genomic DNAs of N/M2 heterozygous samples purifying under mild conditions. Briefly, placental tissues were powdered under liquid nitrogen and treated with proteinase K at 37 °C. Genomic DNAs were then column-purified using DNeasy Blood & Tissue Kit (Qiagen, Tokyo, Japan) in accordance with the manufacturer's instructions. Approximately 500 ng of DNA was used for sodium bisulfite treatment using an EpiTect Bisulfite Kit (Qiagen) in general accordance with the manufacturer's protocol except that the conditions for the sodium bisulfite reaction involved a constant temperature at 37 °C for 16 h. The resulting converted DNAs were used as PCR templates. Primers were designed for the regions containing few T nucleotides to normalize the annealing efficiency between the converted and unconverted DNAs. The 5' end of each primer was designed for use with the Nextera system (Illumina). The primers are listed in Supplementary Table S1.



**Fig. 1** G-quadruplex structure formation at the *ANXA5* promoter region in vitro. **a** A square planar structure of a G-quartet (left), possible form of an intramolecular G-quadruplex structure (middle), and possible G-quadruplex formation at a gene promoter. **b** Genomic structure of the *ANXA5* gene promoter region in terms of its G-quadruplex structure. Eight runs of three or four guanines are indicated by the boxed regions. Transcription start points (tsp) are indicated by arrows. The nucleotide at tsp2 is numbered +1, and the nucleotides within the transcript are capitalized. Polymorphisms are indicated by arrowheads. **c** Conservation of the *ANXA5* gene promoter region among mammalian species. Dots indicate conserved nucleotides, and

conserved guanines are underlined. Dashes indicate the absence of nucleotides. **d** Oligonucleotides used for CD spectra analysis. **e** The propensity for G-quadruplex structure formation at the *ANXA5* gene promoter region in vitro evaluated by CD spectra. Each polymorphic allele was analyzed. The vertical bars indicate the levels of ellipticity estimated by CD spectra. The ellipticity was measured four times and the curves indicate the mean data values. **f** Propensity for G-quadruplex structure formation at various concentrations of potassium chloride evaluated by CD spectra. The ellipticity was measured at 260 nm, and the error bars represent the S.D. ( $n = 3$ )

PCR amplifications were carried out using EpiTaq HS (Takara bio, Kusatsu, Japan). The reaction conditions were 94 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min. There was a final incubation step at 72 °C for 10 min. The resulting products were ligated into pBluescriptII (Agilent, Tokyo, Japan) and plasmids were grown in *E. coli*, purified and sequenced. These sequences were designated as the upper or lower strand in accordance with the conversion of the nucleotides (C>T or G>A) and as N or M2 alleles according to the haplotype. The PCR products were also analyzed by massive parallel sequencing using Nextera index primers (Illumina) following a second round of PCR. After purification and quantification, the products were applied to a Miseq

sequencer (Illumina) at a 250 bp single-end read. The sequence data were quality filtered using FASTX-Toolkit 0.0.14 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) trimming with a cut-off of 20 (phred score) and a minimum length of 12 and filtering with a cut-off of 20 and minimum percentage of 80. The resulting data was further processed by filtering-out PCR duplicates and sequences with indels. After conversion of the fastq to fasta format, clustering analysis of the sequences was carried out using MAFFT [22].

**qRT-PCR**

Quantitative real-time RT-PCR (qRT-PCR) analysis was performed using the TaqMan System. The Superscript First-

strand Synthesis System for RT-PCR (Invitrogen) with random primers was used to produce single-stranded cDNA templates. TaqMan probes and primers for the *ANXA5* gene (Hs00134054\_m1) were obtained commercially (Applied Biosystems). A housekeeping gene, *18S RNA* (Hs99999901\_s1), was used to normalize the mRNA levels because the expression levels among the samples were stable. All qRT-PCR reactions were performed in triplicate in a final volume of 25  $\mu$ l. The cycling conditions used were 2 min at 50 °C, 30 min at 60 °C, and 1 min at 95 °C for RT, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C for PCR amplification.

Allele-specific qRT-PCR was carried out as previously described [11]. The transcript from N allele was amplified with the primers 5'-CAGTCTAGGTGCAGCTGCCG-3' and 5'-GGTGAAGCAGGACCAGACTGT-3', and that from the M2 allele was amplified with 5'-CAGTC-TAGGTGCAGCTGCCA-3' and 5'-GGTGAAGCAG-GACCAGACTGT-3'. The product levels were quantified using SYBR Premix Ex Taq II (Takara BIO) and the 7300 real-time PCR system (Applied Biosystems). The *TBP* gene was used as an internal control.

### Promoter assay

Luciferase reporter constructs were kindly provided by Dr. Arseni Markoff (University of Muenster, Germany), and the assay was performed as previously described [3]. Briefly, the promoter region and exon 1 encompassing SNP1 to 4 was amplified by PCR and the ~450 bp products were cloned immediately upstream of luciferase initiation codon in the pGL3-Basic Vector (Promega) using the *Mlu*I and *Xho*I sites. Each polymorphic nucleotide change was introduced to the vector by means of site-directed mutagenesis. The resulting constructs were co-transfected with pRL-TK into the HeLa cell line using Lipofectamine 2000 (Invitrogen). Luciferase activity was measured at 48 h after transfection using a Dual-Luciferase Reporter Assay System (Promega).

### Statistical analysis

Statistical significance was determined using the Student *t*-test and one-way analysis of variance (ANOVA). *P*-values of <0.05 were considered statistically significant. Data are reported as the mean  $\pm$  SD for each group.

## Results

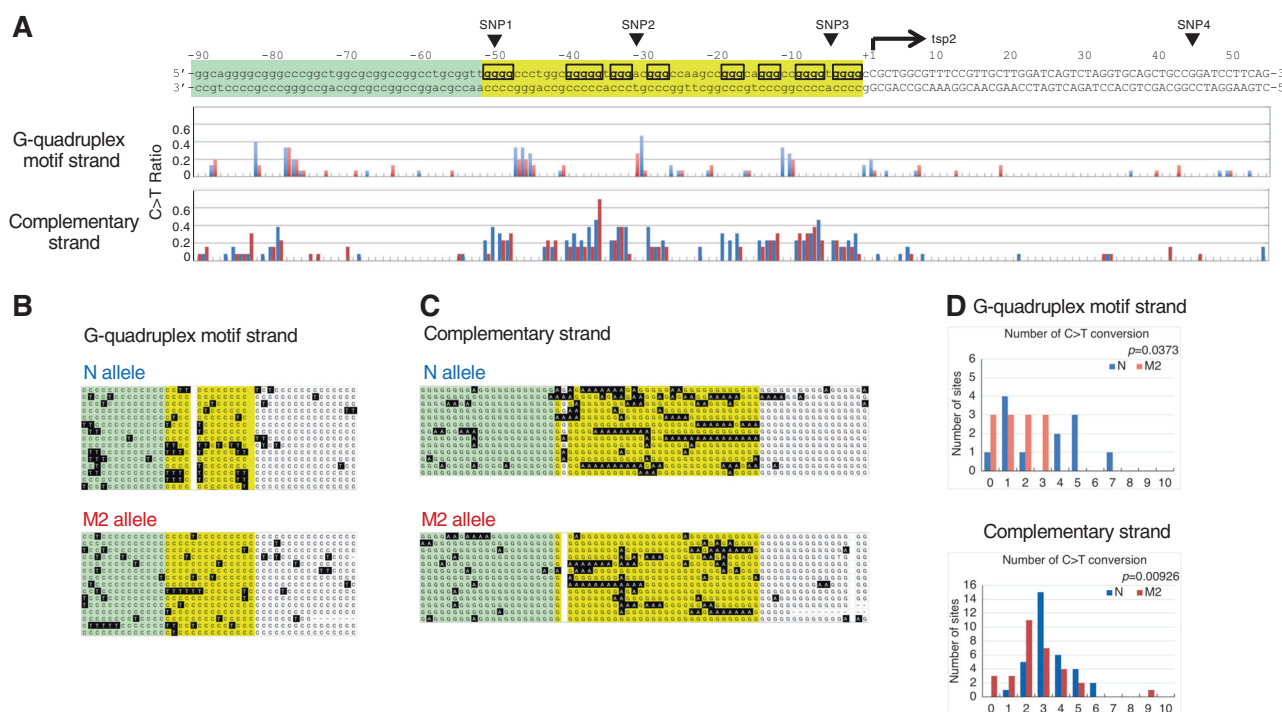
We identified eight runs of three or four guanines with 1–7 nucleotide intervals upstream of the *ANXA5* gene that corresponded to a consensus sequence of a potential

G-quadruplex forming motif,  $G_{3+}-N_{1-7}-G_{3+}-N_{1-7}-G_{3+}-N_{1-7}-G_{3+}$  (Fig. 1a, b) [15]. The *ANXA5* gene has multiple transcription start points (tsp) [23], and although transcripts from tsp1 include 5 runs of guanines at the 5' region, all eight runs of guanines are located 5' upstream of the non-transcribed region in cases of transcription starting from tsp2 and tsp3. The first run of guanines is unique to humans, but all of other seven runs are highly conserved among primates. Other mammalian species also carry at least six runs of guanines, suggesting that this G-rich region plays an important role in gene regulation (Fig. 1c).

To evaluate the G-quadruplex structure forming propensity of this region in vitro, we performed a CD spectroscopy experiment using synthesized oligonucleotides that included the consensus sequence of the potential G-quadruplex forming motif. Typically, parallel form G-quadruplexes display a characteristic positive peak at 260 nm and a negative peak at 240 nm, whereas the anti-parallel form of these structures displays a positive peak at 295 nm and a negative peak at 265 nm on the CD spectra [24]. The oligonucleotide with the N allele sequence produced a positive peak at 260 nm and a trough at 240 nm with a small additional positive peak at 290 nm. This indicates that the N allele DNA forms a G-quadruplex structure in vitro, adopting a mixture of parallel and anti-parallel forms. When SNP1, SNP2, and SNP3 were separately introduced into the N allele, no remarkable change was observed in the CD spectra. However, when these variations were combined to form the M2 allele, the CD spectra showed a reduction in the positive peak at 260 nm, suggesting that the potential for G-quadruplex formation had been decreased (*t*-test,  $P < 0.01$ ) (Fig. 1d, e). To evaluate the propensity for G-quadruplex structure formation in vitro, we analyzed the CD spectra for the oligonucleotides under various potassium ion concentrations (Fig. 1e). The M2 allele showed lower positive peaks at 260 nm in 100 mM, and especially in 20 mM, of potassium chloride ( $P < 0.01$ ), again suggesting that the potential for G-quadruplex formation had been decreased.

It was of interest to us to determine how the M2 haplotype, which appears to possess less potential for G-quadruplex structure formation than the N haplotype, impacted on the promoter activity of the *ANXA5* gene. The restricted methylation of a G-quadruplex structured DNA region was reported previously [25]. On the assumption that *ANXA5* polymorphisms would affect the methylation status of the gene promoter via G-quadruplex structures, and thereby alter gene expression, we performed bisulfite sequencing to locate methylated cytosines in this region. We found that the CpG islands of the *ANXA5* gene upstream region were hypomethylated in placental DNA. However, we did not observe any allele-specific alteration of the *ANXA5* promoter methylation status (Supplementary Fig. S1).





**Fig. 2** Bisulfite modification analysis that could reflect the DNA secondary structure status within the *ANXA5* gene promoter region. The presented data were obtained by gentle bisulfite treatment of placental DNA harboring various *ANXA5* promoter genotypes containing different polymorphisms. Cytosines on each strand of the double helix were evaluated for C-to-T bisulfite modification by cloning and sequencing of the PCR products. **a** The upper panel shows the *ANXA5* promoter sequence. The upstream promoter region, G-quadruplex motif region and transcribed region are depicted in green, yellow and white, respectively. The nucleotide at *tsp2* is numbered +1, and the

transcribed nucleotides are capitalized. Polymorphisms are indicated by arrowheads. Lower panels show the ratios of modified cytosines in the G-quadruplex motif strand and its complementary strand. Blue indicates N allele data and red indicates those for the M2 allele. **b** Sanger sequencing results for each clone derived from the products of N or M2 alleles of the G-quadruplex motif strand, or **c** the complementary strand. **d** Histograms showing the number distribution of modified cytosines at the G-quadruplex motif region (G-quadruplex motif strand,  $n = 13$ ; complementary strand,  $n = 15$

The question that therefore arose from our current findings was the actual cause of the differential gene expression between the M2 and N haplotype alleles. Prior experimental findings had suggested the existence of a G-quadruplex structure at the *ANXA5* promoter in vivo, which could be detected by an antibody specific for this structure [26]. In addition, this G-quadruplex remained in the genomic DNA after purification from cells [27]. We therefore next tested these G-quadruplex structures in genomic DNA using sodium bisulfite modification assays. In the bisulfite modification reaction, cytosines in a single-stranded DNA are converted to uracils but not those in a double-stranded DNA. It is expected therefore that most of the cytosines in a double-strand helix would not change, whereas those that form secondary structure, e.g., the dissociated strand, would be converted (Fig. 1a). Hence, the secondary structure status of a DNA region could be reflected by the bisulfite conversion rate of the cytosines.

Genomic DNAs from N and M2 heterozygous placentas were treated with sodium bisulfite under mild conditions that would not dissociate the strands during the treatment. This DNA was then used as a PCR template to amplify the

*ANXA5* promoter region using specific primers. Sanger sequencing of the resulting cloned PCR products demonstrated a C-to-T conversion by the sodium bisulfite treatment that was specific to the G-quadruplex motif region (Fig. 2a, yellow). The results indicated that a small subset (5–10%) of the molecules had indeed formed the single-stranded DNA as evidenced by successive converted nucleotides that extended across a region of around 30 bp (Fig. 2b, c). Similar results were obtained by massive parallel sequencing of the PCR products (Supplementary Fig. S2). The observed clusters at the complementary strand of the G-quadruplex motif region suggested that G-quadruplex formation occurs in vivo since the formation of the G-quadruplex structure on its own strand may inhibit the C-to-T conversion whilst the single-strandedness of the complementary strand may manifest a higher conversion rate. Notably, the proportion of converted Cs was found to be higher in N allele than in the M2 allele on the complementary strand ( $P = 0.00926$ ) and the G-quadruplex motif strand ( $P = 0.0373$ ), indicating that G-quadruplex formation in vivo might be affected by SNPs, and might contribute to the upregulation of *ANXA5* gene expression (Fig. 2d).

Hence, we examined the expression levels of the *ANXA5* gene against the various SNP genotypes in its promoter. As the *ANXA5* gene is abundantly expressed in the human placenta, we examined the expression effects of its promoter SNPs in this tissue. As homozygotes for high-risk M2 alleles are rare, we compared the *ANXA5* expression levels in M2/N-heterozygous and N-homozygous placentas. The *ANXA5* transcripts were detected at significantly lower levels in the M2/N heterozygote (Supplemental Fig. S3A). To exclude the possible effects of various confounders, we examined allele-specific expression in each M2/N-heterozygous placenta and compared the levels of expression from M2 and N alleles. As was expected, this was lower from the M2 allele (Supplemental Fig. S3B, C). In addition, we examined the effects of the SNPs to the *ANXA5* gene promoter activity using the luciferase reporter system. We amplified the ~450 bp region upstream of the *ANXA5* gene incorporating SNP1 to 4, which was cloned into upstream of the luciferase reporter vector. The M2 haplotype was found to have lower promoter activity (Supplemental Fig. S3D). Thus, at least one of the four SNPs within the M2 haplotype appears to affect *ANXA5* promoter activity leading to its low levels of expression in placental tissues. These results suggest that the M2 haplotype is associated with obstetric complications that arise via altered expression of the *ANXA5* gene, and the expression of the gene might be regulated via G-quadruplex formation in vivo.

## Discussion

In our current study, we show from both in silico and in vitro experiments that the *ANXA5* promoter has the potential for G-quadruplex formation. CD analyses further indicated that the G-rich region of this promoter forms a mixture of parallel and anti-parallel G-quadruplexes in vitro. On the other hand, G-quadruplex formation at this gene promoter in vivo is still somewhat controversial. The formation of these structures requires a long single-stranded DNA at the G-rich region, but this is unlikely to occur upstream of a transcription start point (tsp). However, it is possible that G-quadruplex formation upstream of a tsp might be facilitated by the negative supercoiling induced by transcription [28]. Our current data demonstrated that clustering of the bisulfite modification on the complementary strand of the G-quadruplex motif region, and differences were observed between the N and M2 allele, suggesting that G-quadruplex structures form in vivo at the *ANXA5* promoter and impact its transcription regulation.

The question arises as to the underlying mechanism that drives transcriptional activation when the *ANXA5* upstream region forms a G-quadruplex. Several lines of evidence have suggested that G-quadruplex formation in transcribed

RNA molecules likely contributes to gene regulation at the translational level [29, 30]. It has emerged also that RNA polymerase pausing may contribute to transcription down-regulation of *ANXA5* [31]. These could not apply to the G-quadruplex at the *ANXA5* untranscribed promoter region. Another intriguing hypothesis is that methylation restriction in regions with the potential for G-quadruplex formation affects gene expression [25]. However, we did not observe any methylation differences between the N and M2 alleles in our present analysis. On the other hand, the placenta is a hypomethylated organ, suggesting that G-quadruplex formation might be facilitated in the context of a hypomethylation phenomenon. To shed further light on the relationship between G-quadruplex formation and placental environment, additional analyses will be necessary such as an evaluation of G-quadruplex formation at the *ANXA5* promoter in blood cells. These investigations will provide new insights into the connection between pregnancy success and maternal or placental *ANXA5* haplotypes.

A relatively recent genome-wide surveillance of G-quadruplex structures unveiled high G-quadruplex density in functional regions such as 5' untranslated regions and splicing sites [32]. Enrichment of the promoter regions of highly transcribed genes was observed in another study using an antibody-based G-quadruplex chromatin immunoprecipitation technique [33]. G-quadruplex helicases, XPB and XPD, are enriched near the transcription start site, especially highly transcribed genes [34]. These observations raise the possibility of a regulatory role of G-quadruplex formation on transcriptional regulation. In addition, small molecules or oligonucleotides that target possible G-quadruplex motifs in the promoters of genes responsible for embryonic development have been shown to decrease their expressions in zebrafish [35]. This suggests a role for G-quadruplex formation in the regulation of the gene expression at specific developmental stages.

It is possible that polymorphisms affect the affinity of transcriptional regulatory proteins for a G-quadruplex and thereby lead to the change in transcriptional efficiency [36]. On the other hand, this G-rich region of the *ANXA5* gene includes consensus motifs for transcription factors such as MTF-1, HNF-3, and Sp1 [3], and polymorphisms in the *ANXA5* gene region may possibly alter the binding affinity for these molecules. Recently, a differential impact of the SNPs on *ANXA5* promoter activity was shown using a luciferase promoter assay and electrophoretic mobility shift assay (EMSA) [8]. These findings indicate an effect on gene regulation through the combination of the SNPs, irrespective of whether it is through an alteration of the primary sequence or secondary structure. Theoretically, a conventional method such as EMSA using short DNA duplexes does not necessarily reflect the G-quadruplex formation in long double-strand DNAs. Other genome-wide profiling

such as ChIP-seq technology described above can be used to quantify the propensity for the G-quadruplex formation of each *ANXA5* allele *in vivo*.

The M2 haplotype is common to mammals other than humans, indicating an ancestral lineage. In this regard, the genomic data of Denisova hominin was recently made available on the UCSC genome browser (<http://genome.ucsc.edu/>). The haplotype of the sequenced Denisovan was shown to be the same as the M1 allele, which is another haplotype of the *ANXA5* promoter in the modern human population that shares SNP2 and SNP3 with the M2 allele. The promoter activity of M1 is weaker than that of the N allele but stronger than the M2 allele [3]. An evolutionary advantage of the N allele over the ancestral M2 or M1 alleles could explain majority of the N allele in the modern humans, which express more anticoagulation factor on placental villi, reducing the risk of pre-eclampsia or other obstetric complications. Investigation of the relationship between the evolution of the human phenotype and that from the M2 to N haplotype may provide new insights into the functional propensities of these polymorphisms.

Many common human diseases are believed to be polygenic disorders associated with several genetic and environmental factors [1]. However, most of the disease-susceptible SNPs that have been identified to date confer relatively small increments only in disease risk. Mechanisms by which these SNPs affect gene expression and confer a higher risk of disease thus remains somewhat of an enigma. Our current findings however highlight the contribution of the DNA secondary structure to the fine tuning of the gene expression regulation. A more thorough analysis of G-quadruplexes combined with genome-wide association analyses would likely reinforce the hypothesis that the DNA secondary structure has a fine tuning role in controlling gene expression and thereby has an effect on the susceptibility to common diseases.

**Acknowledgements** We thank Ms. Y. Nakura (Osaka Women's and Children's Hospital, Japan) for technical assistance, and Drs. H. Kogo, M. Tsutsumi, T. Ohye, M. Shimada (Fujita Health University, Japan), A. Markoff (University of Muenster, Germany), K. Hata (National Research Institute for Child Health and Development, Japan), and K. Shiraki (University of Tsukuba) for helpful discussions. This study was supported by grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (HK and IY), grants-in-aid for Scientific Research from the Ministry of Health, Labor, and Welfare of Japan (HK and IY), and AMED under Grant Number JP17gk0110018 (HK).

## Compliance with ethical standards

**Conflict of interest** The authors that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–53.
- Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13:135–45.
- Bogdanova N, Horst J, Chlystun M, Croucher PJ, Nebel A, Bohring A, et al. A common haplotype of the annexin A5 (*ANXA5*) gene promoter is associated with recurrent pregnancy loss. *Hum Mol Genet*. 2007;16:573–8.
- Tiscia G, Colaizzo D, Chinni E, Pisanelli D, Sciannamè N, Favuzzi G, et al. Haplotype M2 in the annexin A5 (*ANXA5*) gene and the occurrence of obstetric complications. *Thromb Haemost*. 2009;102:309–13.
- Grandone E, Tiscia G, Colaizzo D, Chinni E, Pisanelli D, Bafunno V, et al. Role of the M2 haplotype within the annexin A5 gene in the occurrence of pregnancy-related venous thromboembolism. *Am J Obstet Gynecol*. 2010;203:461.e1–5.
- Miyamura H, Nishizawa H, Ota S, Suzuki M, Inagaki A, Egusa H, et al. Polymorphisms in the annexin A5 gene promoter in Japanese women with recurrent pregnancy loss. *Mol Hum Reprod*. 2011;17:447–52.
- Ota S, Miyamura H, Nishizawa H, Inagaki H, Inagaki A, et al. Contribution of fetal *ANXA5* gene promoter polymorphisms to the onset of pre-eclampsia. *Placenta*. 2013;34:1202–10.
- Tiscia GL, Dørum E, Myklebust CF, Grandone E, Sandset PM, Stretting G. Functional characterization of annexin A5 gene promoter allelic variants. *Thromb Res*. 2016;144:93–9.
- Rand JH. The annexinopathies: a new category of diseases. *Biochim Biophys Acta*. 2000;1498:169–73.
- Chinni E, Tiscia GL, Colaizzo D, Vergura P, Margaglione M, Grandone E. Annexin V expression in human placenta is influenced by the carriership of the common haplotype M2. *Fertil Steril*. 2009;91:940–2.
- Markoff A, Gerdes S, Feldner S, Bogdanova N, Gerke V, Grandone E. Reduced allele specific annexin A5 mRNA levels in placentas carrying the M2/*ANXA5* allele. *Placenta*. 2010;31:937–40.
- Nagirnaja L, Nõmmemees D, Rull K, Christiansen OB, Nielsen HS, Laan M. Annexin A5 promoter haplotype M2 is not a risk factor for recurrent pregnancy loss in northern Europe. *PLoS ONE*. 2015;10:e0131606.
- Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011;12:32–42.
- Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet*. 2011;12:683–91.
- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*. 2006;34:5402–15.
- Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res*. 2008;36:1321–33.
- Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res*. 2008;36:6260–8.
- Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci USA*. 2002;99:11593–8.
- Fernando H, Reszka AP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, et al. A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*. 2006;45:7854–60.

20. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* 2006;34:2536–49.
21. Baral A, Kumar P, Halder R, Mani P, Yadav VK, Singh A, et al. Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals. *Nucleic Acids Res.* 2012;40:3800–11.
22. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
23. Carcedo MT, Iglesias JM, Bances P, Morgan RO, Fernandez MP. Functional analysis of the human annexin A5 gene promoter: a downstream DNA element and an upstream long terminal repeat regulate transcription. *Biochem J.* 2001;356:571–9.
24. Balagurumoorthy P, Brahmachari SK, Mohanty D, Bansal M, Sasisekharan V. Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* 1992;20:4061–7.
25. Halder R, Halder K, Sharma P, Garg G, Sengupta S, Chowdhury S. Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol Biosyst.* 2010;6:2439–47.
26. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem.* 2013;5:182–6.
27. Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun.* 2013;4:1796.
28. Kouzine F, Sanford S, Elisha-Feil Z, Levens D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol.* 2008;15:146–54.
29. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol.* 2007;3:218–21.
30. Beaudoin JD, Perreault JP. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 2010;38:7022–36.
31. Eddy J, Vallur AC, Varma S, Liu H, Reinhold WC, Pommier Y, et al. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res.* 2011;39:4975–83.
32. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol.* 2015;33:877–81.
33. Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet.* 2016;48:1267–72.
34. Gray LT, Vallur AC, Eddy J, Maizels N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol.* 2014;10:313–8.
35. David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res.* 2016;44:4163–73.
36. Simonsson T. G-quadruplex DNA structures--variations on a theme. *Biol Chem.* 2001;382:621–8.