

**Evolution, biochemical properties and single-molecule
dynamics of transposon-encoded anti-silencing factor**

Saito, Raku

Doctor of Philosophy

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies,

SOKENDAI

Epigenomes Laboratory

National Institute of Genetics

Table of Contents

Summary	3-4
Introduction	5-9
Results	10-21
Discussion	22-26
Figures and Tables	27-61
Materials and Methods	62-67
Reference	68-76
Acknowledgement	76

Summary

Transposable Elements (TEs) have a wide variety of survival strategies for each TE. Because their transposition and insertion are potentially harmful to the host, the host silences their transcriptional activity by epigenetic systems such as DNA methylation. On the other hand, our understanding about how TEs counteract host silencing, i.e., "anti-silencing" by TEs against the host, is less advanced than the control of TEs by the host. One of the few known examples of anti-silencing by TEs is caused by VANC21 protein, which is encoded in *VANDAL21* family TEs in *Arabidopsis thaliana*. Exogenous expression of *VANC21* gene induces loss of DNA methylation specifically in *VANDAL21* family TEs, which results in transcriptional re-activation of *VANDAL21*-encoded genes. Although VANC21 protein binds to the specific region of *VANDAL21* sequence, DNA de-methylation spreads over the entire *VANDAL21* sequence, which leads to robust loss of silencing. Unraveling the mechanisms behind these phenomena will advance our understanding of how TEs have proliferated in the genomes of diverse organisms. However, the evolutionary aspect, and the biochemical process from DNA-binding to de-methylation, are entirely unknown. In this doctoral thesis, I first report the evolutionary conservation of VANC protein. I focus on VANC-like genes carried by other *VANDAL1~22* families that are phylogenetically closely related to *VANDAL21*. By analyzing these several candidates of active VANCs, I found that sequence-specific anti-silencing function is conserved in one gene of them. This gene, which I named *VANC6*, induced loss of DNA methylation specifically in *VANDAL6* and closely related TE families. The target sequence spectra of hypomethylated TEs were distinct between *VANC21* and *VANC6*, suggesting that these VANC genes have evolved to induce anti-silencing to the specific TEs of their own or close relatives. I next report the biochemical properties and

real-time dynamics of DNA-binding of VANC21. Using High Speed Atomic Force Microscopy (HS-AFM), I was able to visualize the single-molecule and real-time dynamics of VANC21 protein's specific binding to *VANDAL21* sequence. VANC21 bound to DNA with positive cooperativity, and moved dynamically to form multimers by electric charge bias of its disordered regions. The electric charge bias is also conserved in the disordered regions of VANC6. I propose a model that positive cooperative effects and subsequent multimerization of VANC protein contribute to its specific binding to rapidly-evolving arrays of short target motifs. This model may explain the spreading of DNA de-methylation across the entire *VANDAL21* sequence. These findings on biochemical property of the VANC proteins would pave the way towards understanding the molecular basis of the elaborate survival strategy of the *VANDAL* TEs.

Introduction

Transposable Elements (TEs), also known as parasitic DNA sequences, are DNA sequences that can jump and change their genomic position within the host genomes¹⁻⁴. Based on the mode of transposition, TEs are generally classified into two types; Class 1 and Class 2¹⁻⁴. Class 1 is called retrotransposon, which is proliferated by copy-and-paste manner. Transcribed RNA of retrotransposon is reverse-transcribed to DNA by reverse transcriptase encoded by the TE itself, then the copied DNA is inserted into a new genomic position in the host genome¹⁻⁴. Class 2 is called DNA transposon, which is copied by cut-and-paste manner¹⁻⁴. DNA transposon encodes transposase protein, which recognizes terminal sequence ends of TE itself and catalyze the excision and insertion. Both classes of TEs are universally found in the genomes of all eukaryotes and accounts for the majority of the genome in some species, e.g., 45% in the human genome and 80% in the maize genome⁵⁻⁶. This suggests that TEs have behaved as driving factor of the evolution of host genome. Indeed, several cases have been found where novel insertion events of TEs create adaptive systems for the host organism⁷⁻¹¹. However, TE insertions in gene-rich regions are generally eliminated by natural selection because TE insertion events can disrupt the genome and produce various threat to the fitness of host⁷⁻⁸. Although TEs can have positive effects on the host during evolution, they are also potentially deleterious to the host¹⁻⁴. To overcome this harmfulness, host typically silences the transcriptional activity of TEs and regulates their transposition^{1-4, 12-13}. Silencing often involves epigenetic systems such as DNA methylation and histone modifications^{1-4, 11,13-16}. In land plants, DNA methylation occurs in three different cytosine contexts, CG, CHG, and CHH (H stands for A, C or T)¹⁴⁻¹⁸. In *Arabidopsis thaliana*, a model organism whose regulation of TEs has been extensively studied, TE sequences

are robustly DNA-methylated and transcriptionally repressed^{1-2, 14}. Due to high DNA methylation levels, no TE transpositions can be observed by simply growing *Arabidopsis* wild-type strain Col-0 under laboratory conditions^{8, 17}. On the other hand, many TEs are transcriptionally activated by mutations of factors required for DNA methylation, which loses DNA methylation in TE sequences¹⁸⁻¹⁹. For example, in a mutant called *ddm1*, in which DNA methylation in the genome is lost by approximately 70%, transpositions of various TE families are observed regardless of taxonomic class²⁰⁻²³. This genetic result indicates that DNA methylation is important for the silencing of TE transpositions.

The host organism silences the activity of TEs, but TEs are also the major component of the genome⁵⁻⁶. This suggests that TEs somehow circumvent the host's silencing system and successfully transpose. Interestingly, however, the number of reports about “anti-silencing” system of TEs against the host is surprisingly few compared to that about TE silencing system by host²⁴⁻²⁶. Only a few examples have been reported²⁴⁻²⁶. One of these few examples, *VANC21*, is a gene encoded by *Arabidopsis* *VANDAL21* TE family which is class 2 TE found to jump in *ddm1* mutant²⁶. *VANC21* has function to induce loss of DNA methylation of sequences in *VANDAL21* TE family²⁶. This de-methylation results in the transcriptional re-activation of *VANDAL21*-encoded genes²⁶. This activity is thought to counteract the silencing by host and is involved in the efficient proliferation of *VANDAL21* itself²⁶. Thus, dissecting the mechanism of anti-silencing action of *VANC21* provides invaluable insight into understanding the survival strategy of how TEs have proliferated in the eukaryotic genomes.

An interesting aspect of *VANC21* is its target specificity. The *Arabidopsis* genome contains many TE families that are similar in sequence to *VANDAL21* TE family²⁶⁻²⁷. These *VANDAL* TE families are phylogenetically related to *VANDAL21* TE families that share a common ancestor, but *VANC21*

has no effect on these families (Figures for introduction a)²⁶. Furthermore, *VANC21*-like genes are also found in these other *VANDAL* TE families (hereafter, I call *VANC21*-like gene of *VANDAL1* family as *VANC1*, and that of *VANDAL2* family as *VANC2*, and so on)²⁶. These findings raise the possibility that each *VANDAL* TE family has acquired a *VANC* gene that targets only its own family and differentiates its target specificity so that *VANCs* do not recognize other families. Despite this possibility of rapid evolution, the study of *VANCs* remain at the stage of discovery and the following important questions have not been fully answered; Whether the *VANC* genes carried by *VANDAL* TE families other than *VANDAL21* have a DNA de-methylation function, and whether the de-methylation targets by other *VANCs* are highly specific and how it differs from the target specificity of *VANC21*.

In addition, recent study has revealed that *VANC21* gene works in the protein state to generate sequence specificity in its target.²⁸ *VANC21* protein, encoded by *VANC21* gene, localizes specifically to the non-coding regions of *VANDAL21* in the genome (Figures for introduction b)²⁸. At the localization site of *VANC21* protein, a unique 9-base sequence, YAGTATTAY (Y stands for C or T), was found²⁸. This motif is densely aligned, tandemly in some cases, at the *VANC21* localization site²⁸. Although many YAGTATTAY sequences are found throughout the genome, *VANC21* protein does not localize at any YAGTATTAY sequences outside of *VANDAL21*²⁸. Repetitive form of this motif on *VANDAL21* sequence may contribute effective recruit and specific binding of *VANC21* protein within *VANDAL21* non-coding region. On the other hand, despite the specific localization of *VANC21* protein in the non-coding region of *VANDAL21*, loss of methylation effect by *VANC21* occurs across entire *VANDAL21* sequence, including coding region^{26, 28}. While de-methylation occurs strongly in all cytosine contexts at the *VANC21* localization site, de-methylation spreads over the entire *VANDAL21* sequence at non-CG site^{26, 28}. Elucidating the

mechanism behind the spread effect of de-methylation will promote our understanding of sequence-specific de-methylation of *VANDAL21* sequence and provide important insights into the application of *VANC21* as a tool for epigenome editing. However, the mechanism of action, such as the binding kinetics, mode, structure, and real-time dynamics of *VANC21* protein binding to DNA, remains completely unclear. Accordingly, the molecular basis which links between the binding of *VANC21* protein to YAGTATTAY motif and high target specificity, or spreading effect of de-methylation at non-CG context is unknown.

In this doctoral thesis, I focus on following three points about *VANC* protein; evolution, biochemical properties, and real-time dynamics of DNA-binding.

To investigate the evolutionary aspect of *VANC* genes, I first focused on *VANC* genes of other *VANDAL* TE families that were particularly similar to *VANC21* in amino acid sequence. I expressed these *VANC* genes as transgenes and analyzed DNA methylome by whole-genome bisulfite sequencing (WGBS), and confirmed that one of *VANC* gene has an anti-silencing function.

Next, I was interested in the biochemical properties of *VANC* protein. Previous study revealed that *VANC21* protein is localized in *VANDAL21* region. Despite the fact that *VANC21* gene functions as a protein, little progress has been made in the biochemical analysis of *VANC* protein. For biochemical analysis, a single-band, nucleic acid-free pure *VANC* protein is required. I have successfully purified a pure *VANC21* protein using ion-exchange column chromatography and gel filtration chromatography. I used this protein to investigate the biochemical properties such as kinetics of the reaction in binding to DNA.

Since I had successfully purified the pure *VANC21* protein, I wondered if I could use high-speed atomic force microscopy (HS-AFM) to observe the

real-time, real-scale dynamics of VANC21-DNA binding. HS-AFM is the powerful and only tool available for observing the real-time dynamics of proteins²⁹⁻³³. Recently, HS-AFM has been used to reveal the dynamics of interaction between DNA and protein, such as Cas9 protein³⁴⁻³⁵. By applying HS-AFM to VANC21 protein, I was able to confirm that VANC21 protein did indeed bind specifically to the target DNA motif. These observations also visualize its binding mode and real-time dynamics.

Based on the above different three points, I discuss the mechanism of how VANCs protein realizes high specificity for target sequences, and how VANCs protein effectively induce loss of methylation in the entire *VANDAL* sequences.

Results

VANC6 of VANDAL6 TE family has an anti-silencing function

Previous studies have reported that *VANC21* transgene could de-methylates the endogenous *VANDAL21* TE sequences^{26, 28}. There are several other *VANDAL* TE families in the *Arabidopsis* genome that are phylogenetically closely related to the *VANDAL21* TE family, but *VANC21* targets only *VANDAL21*. On the other hand, ORFs that are similar in amino acid sequence to *VANC21* are encoded in other *VANDAL* TE families. To investigate whether the *VANC* genes carried by *VANDAL*s other than *VANDAL21* have de-methylation activity and, if so, how specific they are, I selected *VANC2*, *VANC6*, *VANC14* genes that are particularly similar in amino acid sequence to *VANC21* as candidates of active *VANC*s. I then amplified them by PCR, introduced them as transgene into the *Arabidopsis* wild-type Col-0 ecotype and expressed them (Fig1). Using Whole Genome Bisulfite Sequencing (WGBS), I checked if there are sequences that were uniquely hypo-methylated. WGBS data showed that *VANC6* gene (At4g09370) carried by a member of *VANDAL6* TE family (AT4TE25050) has an anti-silencing function, although other *VANC* genes did not induce significant hypomethylation compared to the wild-type (Fig 2a-c).

VANC6 induces loss of DNA methylation in VANDAL6 and closely related TE families

In the *VANC6* transgenic plants, several copies of *VANDAL6* were DNA de-methylated (Table1). Interestingly, transposons other than *VANDAL6*, such as *VANDAL8*, are also de-methylated (Table1, Fig 3a-b). These de-methylated TE families are the most closely related to *VANDAL6* among the *VANDAL* TE families in the *Arabidopsis* genome²⁶. This result indicates that *VANC6* specifically de-methylates *VANDAL6* and related *VANDAL* TE

members. In TEs de-methylated in *VANC6* transgenic plant, CG sites were locally de-methylated and non-CG sites were de-methylated throughout the entire TE length (Fig 3a-b). The characteristics of this de-methylation are similar to that of *VANC21*²⁸.

VANC6 has the unique binding motifs AGTTGTMC and AGTTGGCC

Previous study revealed the genomic localization of *VANC21* protein and showed that a unique 9-base sequence, YAGTATTAY (where Y stands for C or T), was present at the localization site²⁸. *VANC21* de-methylated the entire length of *VANDAL21* in non-CG site, but the de-methylation of CG site was local and closely correlated with the localization site of *VANC21* protein²⁸. Because *VANC6* targets were also de-methylated along the entire TE sequence in non-CG site but were de-methylated locally in CG site (Fig 3), I wondered if it would be possible to identify *VANC6* protein binding sequence motifs from WGBS data as highly enriched sequences in CG de-methylated regions in *VANC6* transgenic plant. I statistically analyzed the whole genome in a window of 100 bp each, in which CG site methylation is reduced by more than 50% compared to the wild-type, and identified specific sequence motifs AGTTGTMC and AGTTGGCC (M means C or A) (Fig 4a-b). This sequence appears specifically in the de-methylated TE family in *VANC6* transgenic plants, suggesting its potential as a scaffold for localization (Fig 4c). In addition, *VANC6* protein bound to the DNA probe with high density of the identified sequence motif *in vitro* (Fig 4d). Interestingly, these *VANC6*-binding motifs are also found in AT9TSD1 TE family, which lose DNA methylation in *VANC6* transgenic plant (Fig 2a, d, Fig 3c, Fig 4e-f), even though AT9TSD1 is not related to *VANDAL* TE families. This result supports that *VANC6* uses AGTTGTMC and AGTTGGCC as binding target to induce anti-silencing in TE

sequence. It is interesting how *AT9TSD1* TE family acquired the *VANC6* binding motif.

VANC6 and *VANC21* are closely related but have different target specificities

VANC21 (*VANDAL21*) and *VANC6* (*VANDAL6*) share a common ancestor²⁶. Interestingly, the spectra of hypomethylated TEs were completely different between *VANC21* and *VANC6* transgenic plants (Fig 4g). Differences in the sequence specificity of *VANC21* and *VANC6* targets could also be confirmed at the *in vitro* level (Fig 4h). These results indicate that the *VANC* gene has evolved to specifically de-methylate TE families similar to the TEs that encode it. Differences of target sequence specificity are produced at the protein level and are determined by unknown DNA-binding domains present within *VANC* proteins. *VANDAL* TE families share a high degree of sequence similarity with each other, and *VANDAL* TE families have only been diverged relatively recently during evolution. For long target sequences (8 kb), de-methylation occurs not only at the localization site but also throughout the entire length of the sequence. It is interesting to note that each *VANC* gene efficiently differentiates its target specificity for such a long sequence. An unknown DNA-binding system may exist in *VANC* protein that enables such differentiation.

The naturally disordered regions at both ends of VANC21 protein are not required for DNA binding

Considering its function, VANC should have domains involved in two functions, DNA binding and anti-silencing. However, there are no functionally known domains of Pfam database in VANC²⁶. There should be an unknown DNA binding system in the VANC gene that allows for efficient differentiation of target specificity. To understand the binding system of VANC proteins to DNA, biochemical analysis using purified proteins are essential. Despite the importance of biochemical assays for understanding VANC21 function, previous studies were mainly carried out using genetics and genomics^{26, 28}. Although VANC21 protein and VANC6 protein were used for electrophoretic mobility shift assay (EMSA) in Fig 4h and previous research²⁸, they were only purified by affinity chromatography against their tags and they contained other protein bands, nucleic acids from *E. coli*. Our understanding of the biochemical properties of VANC proteins is completely lacking, so that I first remove the nucleic acid by ion-exchange column chromatography and then used gel filtration chromatography to obtain the pure, single-band VANC21 protein. In addition, to simplify biochemical studies, I also attempted to identify regions that are not required for binding to DNA. 280-684aa region of VANC21 protein, where the amino acid sequence is highly conserved among *VANDAL6*, *VANDAL7*, *VANDAL8*, *VANDAL17*, *VANDAL21* families, is annotated with the functionally unknown domains of DUF1985 and DUF287 in the Pfam database (Fig 5a)²⁶. Since the other regions, i.e., the N- and C-terminal regions, were likely to be disordered regions (Fig 5a-b). I prepared VANC21 protein with both terminal regions removed. Gel filtration chromatography showed that full-length VANC21 protein was separated into the void fraction regardless of its tag using affinity chromatography (Fig 5c, GST tag and 8x His tag). On the other hand,

VANC21 protein with deletion of the N- and C-terminal regions (hereafter deletion VANC21 protein), resulted in the fractionation of VANC21 protein to its original molecular weight and retained its DNA-binding activity (Fig 5c-e). These results raise the possibility that full-length VANC21 protein may aggregate into multimeric forms, and its aggregation requires the N- and C-terminal regions. Since deletion VANC21 protein does not aggregate, this may be more suitable for biochemical analysis than full-length VANC21 protein.

VANC21 protein binds to DNA with positive cooperativity

VANC21 protein specifically localizes to the intergenic region of the *VANDAL21* sequence²⁸. On the other hand, VANC21-binding DNA motif, YAGTATTAY, is universally found in the *Arabidopsis* genome²⁸. Previous study suggested that the high frequency of this sequence motif on the *VANDAL21* sequence is the reason why VANC21 protein can be localized specifically to *VANDAL21*²⁸. However, even though there is a correlation between the high frequency of motif sequence per unit length and the specific localization of VANC21 proteins, the mechanism for the efficient binding of VANC21 proteins to *VANDAL21* sequences is not clear. One hypothesis to explain this phenomenon is that a positive cooperation of VANC21 and the DNA motif, in addition to one-to-one binding, further promotes VANC21 loading onto the motif, and the loading is enhanced by high density of the motifs. In this hypothesis, the binding of VANC21 protein to DNA is more cooperative when the density of motifs is high. To test this hypothesis, I performed EMSA of VANC21 protein using DNA probes with and without the motif, and examined the kinetics of the binding reaction to DNA by applying a concentration gradient to VANC21 protein (Fig 6a). The binding reaction curves show sigmoidal curve, suggesting that the reaction is not a simple binding reaction but with positive cooperativity (Fig 6b-c). Here, VANC21 protein is denoted as P and the DNA

probe as D. I first assumed the following simple reaction equation for the VANC21-DNA binding reaction.



Next I assumed the following equation that includes cooperative effect into simple binding reaction equation (1).



Since how many VANC21 proteins actually bind to a single motif is unknown, I put the above equations (1)(2) together and assumed the following equation.



Summarizing the above equation (3), I considered the following equation as whole VANC21-DNA binding reaction.



In this equation (4), if I define the increase in DP_n per unit of time as V , following equation (5) can be established using K_d as the concentration of P when half of all D s are DP_n ³⁶⁻³⁸.

$$(5) \quad V = \frac{d[DP_n]}{dt} = \frac{[DP_n]}{([D] + [DP_n])} = \frac{[P]^n}{(K_d^n + [P]^n)}$$

This binding reaction equation can be approximated to Hill equation by assuming that D is the enzyme and P is the substrate³⁶⁻³⁸. Therefore, I fitted the result of VANC21-DNA binding reaction curve to Hill equation by using K_d and n as the Hill coefficient, which is a measure of cooperativity. Assuming V_{max} to be 1, Hill coefficients were 2.89 (95% confidence interval 2.47 to 3.42) when the motif is included and 2.66 (95% confidence interval 2.43 to 2.90) when motif is not included (Fig 6c-f). These results indicate that VANC21 protein binds to DNA with positive cooperativity regardless of target sequence, and its cooperativity is stronger if YAGTATTAY motif exists. Interestingly, the kinetics

of probe without motifs also show sigmoidal curve, suggesting the cooperative effects of VANC21 protein does not use motif as reaction scaffold. Since YAGTATTAY motif forms tandem repeat on *VANDAL21* sequence, DNA-binding system with positive cooperativity may induce anti-silencing in target TE sequences more effective than in the other TE sequences. Therefore, deciphering how VANC21 protein binds to YAGTATTAY motif at real-scale, real-time levels may provide important insight into understanding the mechanism of anti-silencing by *VANC21*.

HS-AFM observation reveals that full-length VANC21 protein forms multimer

Although it has been shown that VANC21 proteins have positive cooperative effect for DNA-binding, the DNA-binding system of VANC21 protein has not been fully clarified. Since it is not clear how VANC21 protein bind to the motif, the biological significance of the YAGTATTAY motif's repetitive presence on *VANDAL21* sequence is unknown. The repetitive form of YAGTATTAY motif raises some possibilities, such as facilitating interactions between VANC21 proteins on the *VANDAL21* sequence or the movement of VANC21 proteins to adjacent motifs. To elucidate the biological significance, I used high-speed atomic force microscopy (HS-AFM) to directly observe the binding of DNA to VANC21 protein. Using 3-aminopropyl-triethoxysilane treated mica (AP-mica), I was able to visualize the full-length and deletion VANC21 protein on the mica surface (Fig 7a, c). Deletion VANC21 protein has round shape with the height of 1.3-1.6 nm (Fig 7d). This is consistent with the fact that mass of deletion VANC21 is 44,289 Da. On the other hand, the observed height of full-length VANC21 varied from about 2.5 nm to over 20 nm (Fig 7a-b). This is much larger than expected height considering that the mass of full-length VANC21 is 79,027 Da. This observation is consistent with the gel filtration chromatography data, suggesting that full-length VANC21 protein forms a multimer on the mica surface. Full-length VANC21 protein retains its sequence-specific DNA-binding activity, suggesting that it is properly folded. While the height of deletion VANC21 protein was constant, full-length VANC21 protein is various without being biased towards a specific size (Fig 7b, d), indicating that the N- and C-terminal disorder region may contribute to various type of multimer formation. Given that the YAGTATTAY motif is repeatedly present on *VANDAL21* sequence, the multimeric nature of VANC21 proteins presents the possibility

that VANC21 protein bound to one motif may interact with other VANC21 proteins bound to adjacent motifs to form multimer.

VANC21 protein forms multimer in a DNA-independent manner

On the surface of normal mica without treatment with 3-aminopropyltriethoxysilane, VANC21 proteins actively moved around on the mica surface and behaved like an oil droplet floating in water (Fig 8). Most importantly, I was also able to take advantage of this property to observe VANC21 forming multimers independent of DNA binding. During the observation, VANC21 proteins were attracted to each other and binds when they were in close position (Fig 8, 20.57s). The newly created dimer of VANC21 was sometimes separated into two monomers (Fig 8, 21.57s, the height was used to measure the number of VANC21 proteins aggregated). I also observed the separation of monomer from one multimer, and the separated monomer was bound to another multimer (Fig 8a, 3.59s~20.57s). This movie demonstrates the ability of full-length VANC21 protein to form multimers in a DNA-independent manner by real-time scale.

VANC21 protein binds specifically to YAGTATTAY on the mica surface

To elucidate the binding mode of VANC21 protein to DNA and the biological significance of multimer formation, I first visualized the binding of VANC21 protein to a YAGTATTAY motif using HS-AFM. For observation, I focused on the genomic region downstream of the *VANA21* gene, which has four repetitive motifs, and prepared two 500 bp DNA sequence fragment with this region in the central (central-type) or the end (end-type) (Fig 9a-b). Using AP-mica, I was able to observe DNA and binding of VANC21 to these DNA sequences (Fig 9d-k). The observed length of both type of DNA was about 180nm (Fig 9c). I found VANC21 protein binds to the region corresponding to

the position of the motif of each DNA-type. (Fig 9h-k). These results indicate that the dynamics of VANC21 protein and its specific binding to the YAGTATTAY motif can be directly observed in real-time by using HS-AFM. Although non-specific binding was found for any combination of protein and DNA type, much of the binding occurred in the region where the motif was present (Fig 9h-k).

VANC21 protein binds to DNA in a dimeric form

Because DNA motif YAGTATTAY is repetitively present on *VANDAL21* sequence, it is possible that VANC21 protein binds to the *VANDAL21* sequence in a multimeric form. I was able to observe full-length VANC21 protein binding to DNA in a dimeric form (Fig 10a-c). VANC21 protein had already formed a dimer prior to binding, and the binding of one monomer triggered the binding of the other monomer to the DNA (Fig 10a, 2.7s~38.7s). The binding point is area with motifs. (Fig 10b). 70 seconds after binding to DNA, one of monomers forming dimer moved on DNA and the dimer split (Fig 10a, 107.1s~144.9s). Each of the monomers was separated from DNA (Fig 10a, 180.9s~303s). Each of the monomers in the dimer has binding activity (Fig 10a, 38.7s~63.3s). These results indicate that VANC21 retains binding activity to motif even in dimeric form.

VANC21 binds to DNA in any form of multimer

In addition to the movie of dimer binding to DNA, I was able to observe the scene in which VANC21 multimer forms an even larger multimer through the binding to motif (Fig 10d-f). Another multimer was attracted to and bound to VANC21 multimer which was already bound to a motif region of DNA (Fig 10d, 49.2s). The two assembled multimers dynamically changed their formation of monomers and were stable on the DNA (Fig 9d, 49.2s ~ 243.6s). These images

show that the multimeric form retains its binding activity to DNA and the multimeric form does not have a specific formation.

VANC21 protein monomer bind dynamically to VANDAL21 sequence by moving from motif to neighboring motif

Most of the observed binding was stable on the DNA, but I was able to observe the unstable movement of the full-length VANC21 protein across the DNA (Fig 11a). The bound VANC21 seemed to be a monomer (Fig 11a-b). At the beginning of the observation, full-length VANC21 protein was in the central part of the DNA, but then continuously moved around on the DNA, leaving the motif region after 28.2s and dissociating from the DNA after 32.4s (Fig 11c). This movie raises the possibility that binding to motif with monomeric form may not provide sufficient affinity to ensure target specificity.

The N-terminal disorder region of VANC21 protein is extremely acidic

The fact that full-length VANC21 protein formed a multimer while the deletion VANC21 protein was a monomer suggests that the N- and C-terminal disorder regions may contribute to multimer formation. The N-terminal disorder region of VANC21 contains an extremely acidic amino acid-rich region (Fig 12a-b, 139-242 aa). The amino acid composition of this N-terminal 139-242 aa region is not conserved in VANC6, but interestingly, the feature of acidic amino acid richness is conserved (Fig 12a, c). The 280-648 aa of VANC21 protein, a region of deletion VANC21, show no extreme charge bias (Fig 12b, c). This order region is interspersed with slightly acidic and basic biases, and this feature was similar in VANC6 (Fig 12b, c). In addition, an extremely basic-rich regions are found in disorder region, from N-terminal 1 aa to 90 aa in VANC21 and from 600 aa to C-terminal 666 aa end in VANC6. Therefore, the

multimerization of VANC21 may be caused by the interaction of two different disorder regions that are extremely negative or positive charged.

Discussion

In this doctoral thesis, I first reported the evolutionary conservation of *VANC* genes and then reported the biochemical properties and single-molecule dynamics of their DNA binding. These results will provide insight into the binding mechanism and action dynamics of *VANC21* proteins *in vivo*, as well as the significance of the repetitive nature of the binding motifs, and will enhance our understanding of transposon survival strategies against host silencing systems.

Firstly, I found that the anti-silencing function of *VANC* is also conserved in another *VANDAL* family, *VANDAL6* (Fig 2a). *VANDAL6* family and its closely related *VANDAL* TE families were specifically DNA de-methylated, whereas the *VANDAL21* family was unaffected (Fig 4g). Although *VANDAL21* and *VANDAL6* TE families are closely related²⁶, the targets for de-methylation of their respective *VANC* proteins were different and did not affect each other (Fig 4g). These results suggest that all *VANDAL* TE families may have rapidly evolved to acquire *VANCs* with different target specificities. Genetic evidence that multiple *VANDAL* families are hypomethylated in F1 plants between DNA-hypomethylated mutant and wild-type strongly supports this idea³⁹. Since the evolution of tandem repeats can occur rapidly, the tandem repeat formation of *VANC*-targeting motif may contribute this rapid evolution⁴⁰⁻⁴¹. Although several horizontal transfers have been reported for TEs since their discovery, these events are rare compared to viruses⁴²⁻⁴⁴. Unlike viruses that can efficiently spread horizontally, TEs cannot proliferate when their hosts die. Therefore, they cannot employ a survival strategy that reduces host fitness globally, as viruses do with RNA silencing suppressors (RSSs)⁴⁶⁻⁴⁷. Therefore, the differentiation of target specificity of *VANC* among all *VANDAL* TE families has an advantage for their survival.

A unique motif AGTTGTMC/AGTTGGCC frequently appears in differentially hypomethylated regions in *VANC6* transgenic plants, and this sequence is not found in *VANDAL21* (Fig 4a-c). High target specificity of *VANCs* are also determined at the DNA sequence level, where each *VANDAL* family may have acquired different specific DNA binding motif for its own *VANC* protein. These motifs, which consist of about 8 to 9 bases, are located in non-coding regions of each *VANDAL* sequence and in some cases form tandem repeats²⁸. These results suggest that these motifs aid in anti-silencing of the *VANDAL* family as a scaffold of *VANC* protein and lead to efficient proliferation of *VANDALS*.

The kinetics of the complex of deletion *VANC21* protein and DNA showed a sigmoidal curve as the concentration of deletion *VANC21* protein increased, indicating that *VANC21* proteins binds to its target DNA motif with positive cooperativity (Fig 6c-e). The kinetics curve is sigmoid regardless of the presence or absence of the motif, suggesting that positive cooperative effect occur even in the absence of motifs (Fig 6c-e). The fact that *VANC21* has a positive cooperative effect regardless of the presence of the motifs indicates that the cooperative second binding reaction followed by the first binding of *VANC21* protein is not DNA sequence-dependent. This result suggests that the cooperative second binding may occur at the sequence in which the first binding occurs, regardless of DNA sequence. One AFM observation that *VANC21* protein bound to the motif in a dimeric form supports this idea (Fig 10a-c). Moreover, I found the YAGTATTAY motif is a partial palindromic sequence when the 9th "Y (C or T)" is "C" and 3' following nucleotides are TR (where R stands for G or A) (Fig 13). Given that *VANC21* protein has some binding affinity to probe with SNPs in the central nucleotide of YAGTATTAY motif²⁸, it is possible that both strands of the partial palindromic sequence of the motif are involved in binding affinity. These findings suggest the *VANC21*-

binding mode in which two VANC21 proteins bind to both strands of one single motif to form a dimer.

Full-length VANC21 protein formed multimer in DNA-independent manner on mica surface (Fig 7, 8). Multimerization of full-length VANC21 also occur with DNA-binding manner (Fig 10a-f). If VANC21 proteins are brought into close proximity with each other through binding to DNA, they are more likely to form multimeric forms. These results suggest that the local and repetitive presence of YAGTATTAY motif on *VANDAL21* sequence leads the efficient multimerization of VANC21 protein. Interestingly, when the N- and C-terminal disorder regions on VANC21 were deleted, VANC21 did not form multimers (Fig 7c-d). The N-terminal 139-242 aa disordered region of VANC21 is extremely acidic, which was also conserved in VANC6 (Fig 12a-c). In addition, extremely basic region is also found in disorder regions of both VANC21 and VANC6 (Fig 12a-c). These findings strongly suggest that the multimer formation was due to electric charge bias and multimerization may be essential for their target specificity. Since positive cooperative effects was found in deletion VANC21 protein, multimerization occurs independently of cooperativity. Therefore, the following model is likely to explain the biological significance of the local and repetitive form of YAGTATTAY motif on *VANDAL21* sequence. One VANC21 protein first bind to a motif in the form of a monomer (Fig 14a1), and then another monomer cooperatively binds to same motif sequence (Fig 14a2). This binding occurs in adjacent motif. And the last, VANC21 proteins in closed position form multimer using extreme electric charge bias of the disorder regions, with the help of the repetitive form of target motif (Fig 14a3). The localization of VANC21 protein is specifically stabilized only on *VANDAL21* sequence due to its multimerization.

The multimeric form of VANC21 protein does not take a constant shape (Fig 7a-b). In addition, VANC21 proteins dynamically attach to and

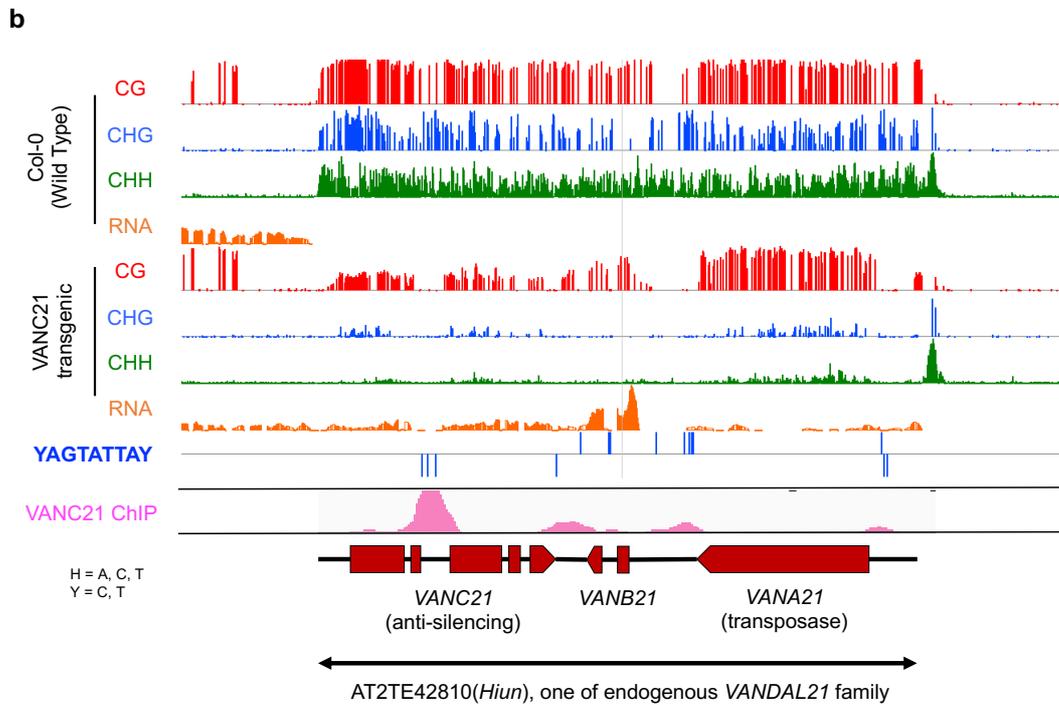
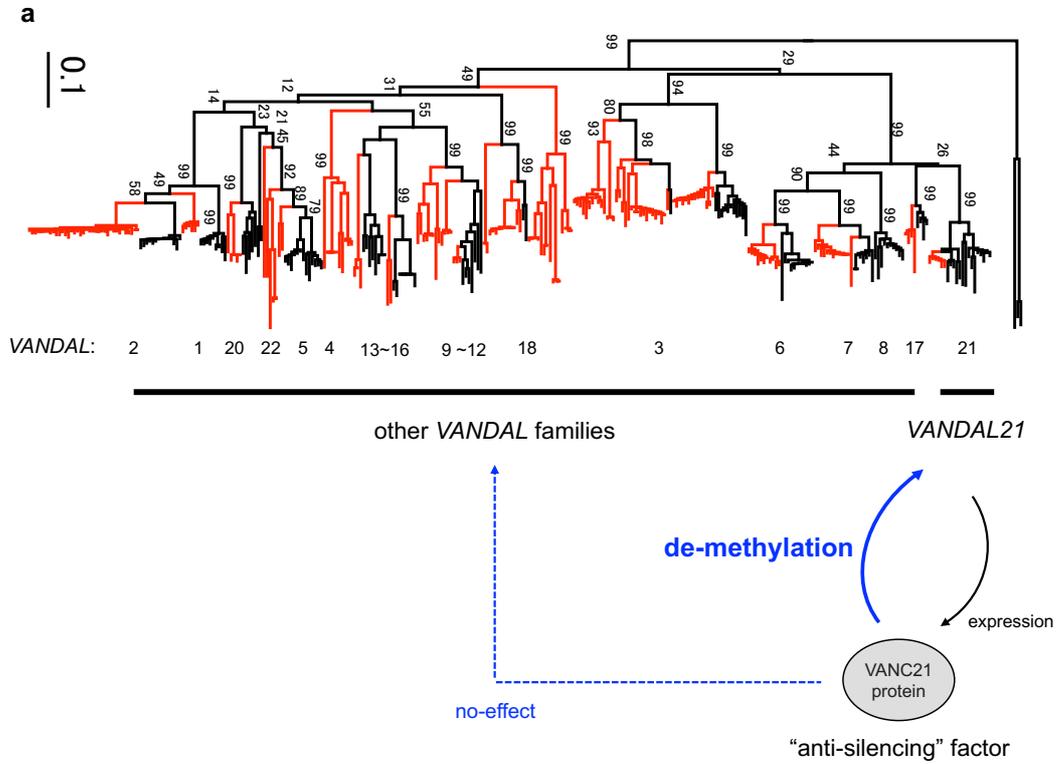
detach from each other (Fig 8). Thus, VANC21 multimers formed on *VANDAL21* sequence may move dynamically, even *in vivo*. Since VANC21 protein has some affinity for DNA without motifs (Fig 6a, c), it is possible that VANC21 multimer localized at the non-coding region of *VANDAL21* due to its strong affinity for motifs could dynamically move and detach, and bind to motif-free regions of *VANDAL21* sequence. This may explain the phenomenon of de-methylation at non-CG sites by VANC21 protein, which spreads over from the localization site to the entire *VANDAL21* sequence. Based on the above findings, I propose a model that VANC21 protein induce compaction of *VANDAL21* sequence in three dimensions, and creates a unique space (Fig 14b). In fact, one AFM observation show an image that VANC21 multimer bundle together two different DNA fragments (Fig 15a). On the other hand, the question remains as to why the spreading of de-methylation by VANC21 is limited to non-CG sites. Spreading of de-methylation across the entire sequence of *VANDAL21* is restricted to non-CG sites, whereas methylation at VANC21-localized sites is lost at all contexts, including CG sites. The difference in de-methylation between localized and non-localized sites in *VANDAL21* sequence may be determined by a quantitative issue, namely the amount of VANC21 protein binding.

Recently, many studies showed that intracellular separation of proteins by liquid-liquid phase separation is important for the proper functioning of proteins, especially in the nucleus⁴⁶⁻⁵¹. For example, the transcription factor EBF1 regulates chromatin remodeling through phase separation by naturally disordered region of its C-terminal domain⁵². VANC21 protein behaves like an oil droplet (Fig 8), which is reminiscent of phase separation. The large unusual structures observed in AP-mica may be the result of VANC21 assembly by phase separation (Fig 15b). Moreover, although the half of VANC21 has disorder domain, string-like structures are not observed in full-length VANC21

protein with any form (Fig 7a). This suggests that extremely negative and positive charged disorder regions of VANC21 protein interacts each other and well-folded. It is possible that VANC21 induces appropriate anti-silencing by liquid-liquid phase separation, but validating this will require more biochemical experiments. If the crystal structure of VANC21 protein is clarified, more important interpretations can be gleaned from the AFM data.

Combining all results, the following explanations are possible for the biological significance of biochemical properties and rapid evolution. VANDAL TEs contain the VANC-binding motifs in their non-coding regions as tandem repeat formation. This may enable the rapid evolution of the VANDAL families, and leads to a diversification and differentiation of the target specificity of VANC proteins. On the other hand, VANC protein needs to bind efficiently and specifically to motifs that form tandem repeats in each VANDAL family. To achieve this, VANC has function of positive cooperativity for DNA-binding. These three points influence each other and result in an efficient proliferation for each VANDAL TE family. These results suggests that positive cooperative effect and multimerization of VANC protein contribute to its specific binding to short target motifs, and rapid evolution, as well as spread of the effects on de-methylation to entire VANDAL21.

Figures and Tables



Figures for introduction. Function of VANC21 mentioned in previous research.

a Phylogenetic tree among *VANDAL* families in genomes of *A. thaliana* (black line) and *A. lyrata* (red line) described in Fu et al²⁶. *VANC21* induces loss of methylation in *VANDAL21* but no effects in other *VANDAL* families, although these other *VANDAL* families are closely related *VANDAL21*.

b DNA methylation, transcriptomic and *VANC21* protein localization status of *Hiun*, one copy of *VANDAL21* family described in previous research^{26, 28}. Scales of DNA methylation are percentages from 0 to 100, and that of RNA indicates the number of reads of RNA-seq with the maximum being 700. *VANC21* induces loss of methylation in all contexts and transcriptional re-activation at *Hiun*. *VANC21* protein is localized in non-coding region of *VANDAL21* sequence²⁸. A specific motif, YAGTATTAY (Y stands for C or T), was densely packed at *VANC21*-localized site and in some cases forms a tandem repeat²⁸. Motifs present in watson chain are shown as + and those present in crick chain as -.

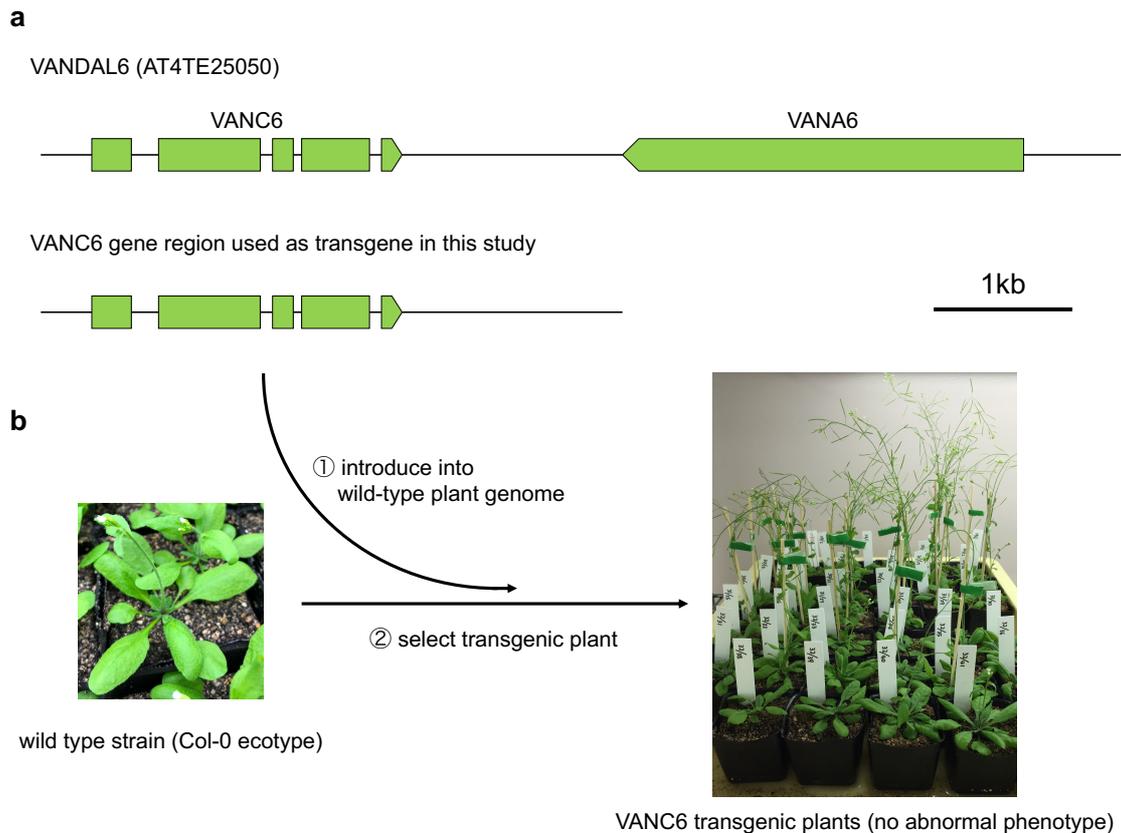


Fig 1. Schematic diagram on generation of VANC6-transgenic plants

a Schematic diagram of VANC6 gene used for transformation. Boxes indicate exons. The transgene of VANC6 include upstream region, which is the edge of VANDAL6.

b Process of generating VANC6 transgenic plants. The transgene was introduced in wild-type ecotype strain Col-0, and then transgenic plants are selected by antibiotic. All transgenic plants did not show abnormal phenotype as well as VANC21 transgenic plants.

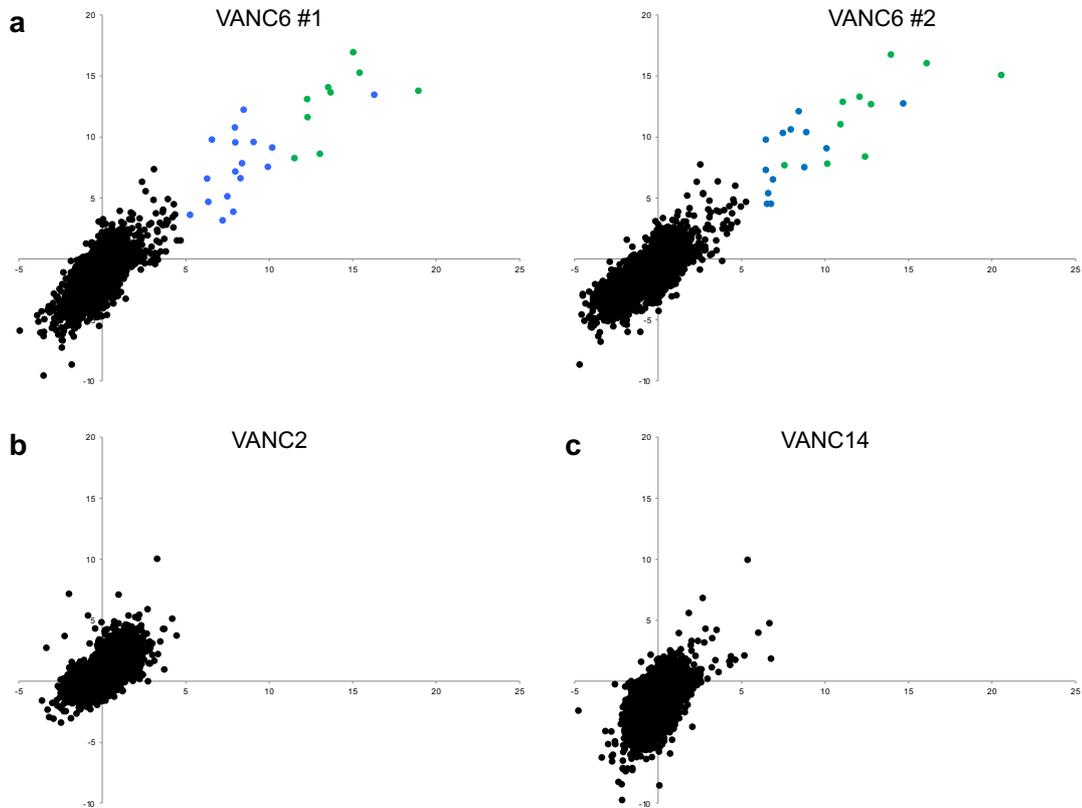
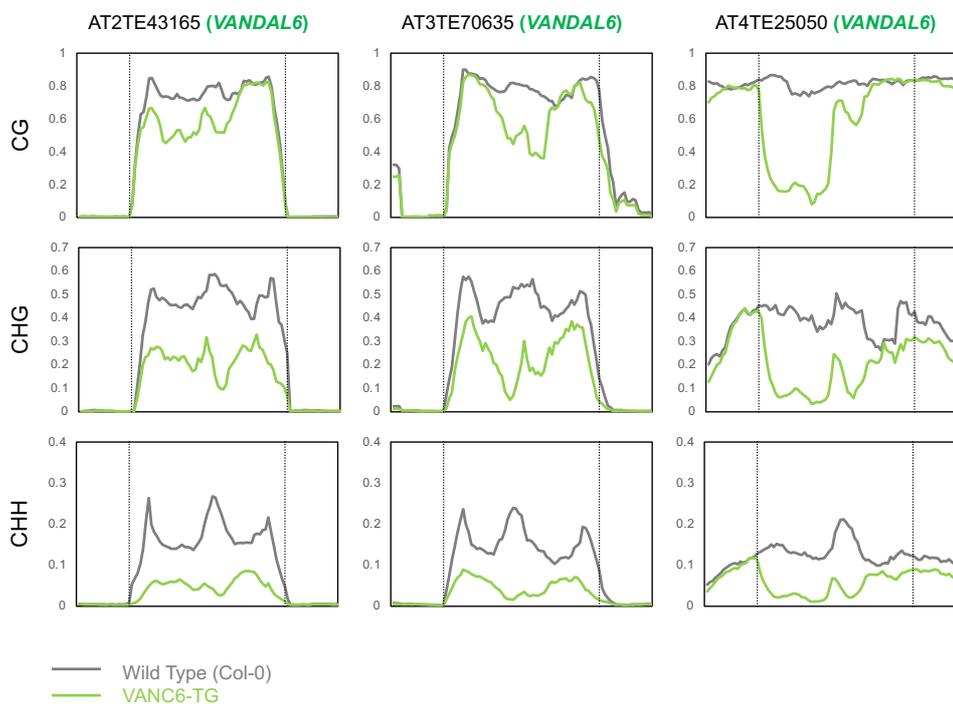
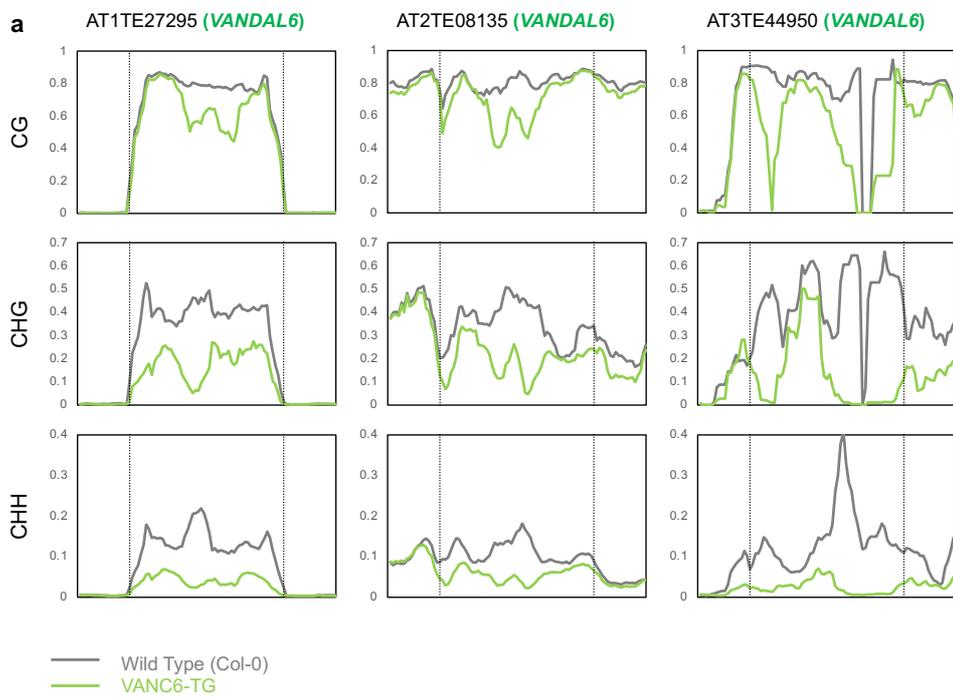
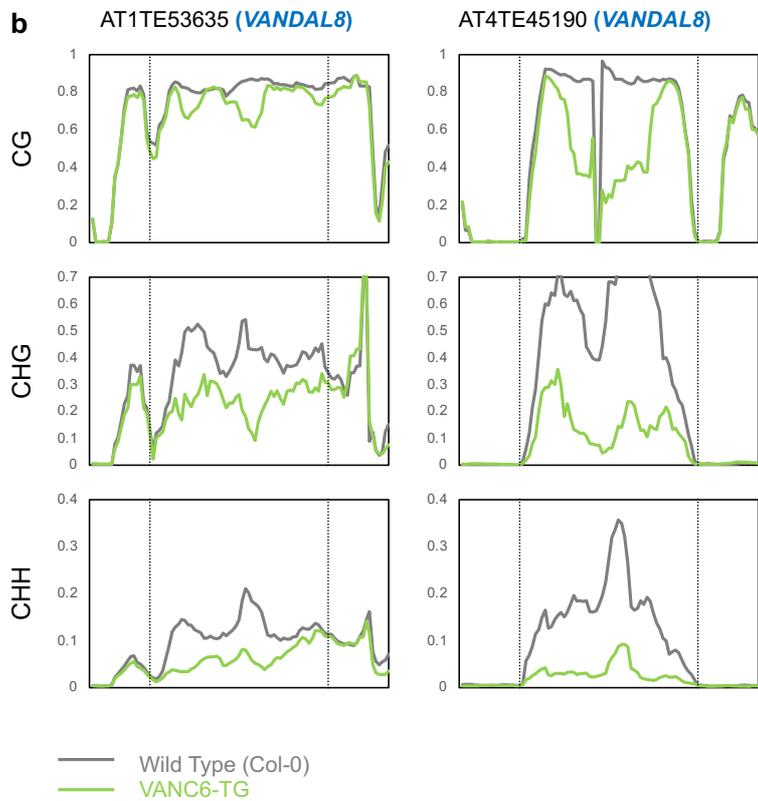
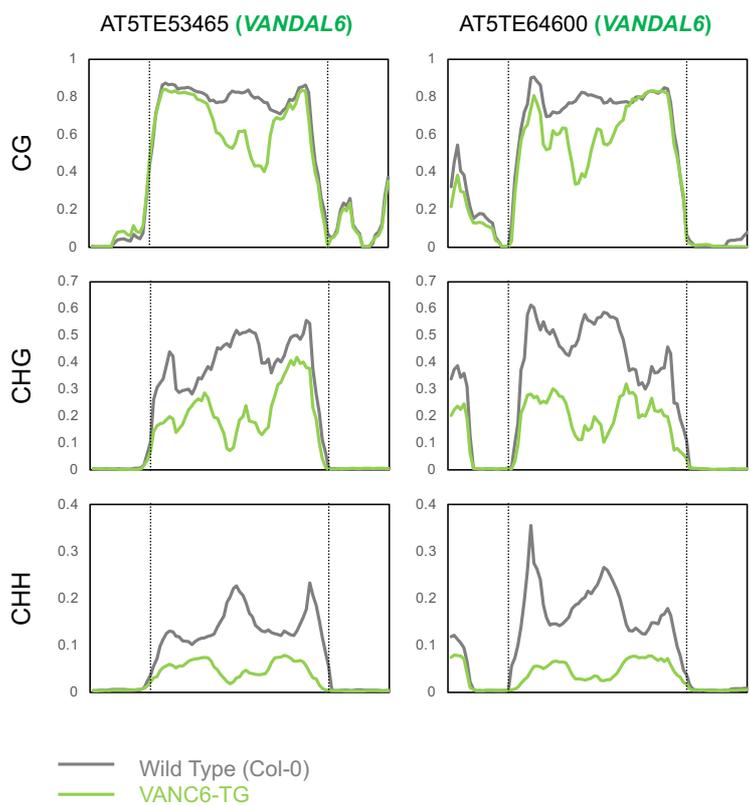


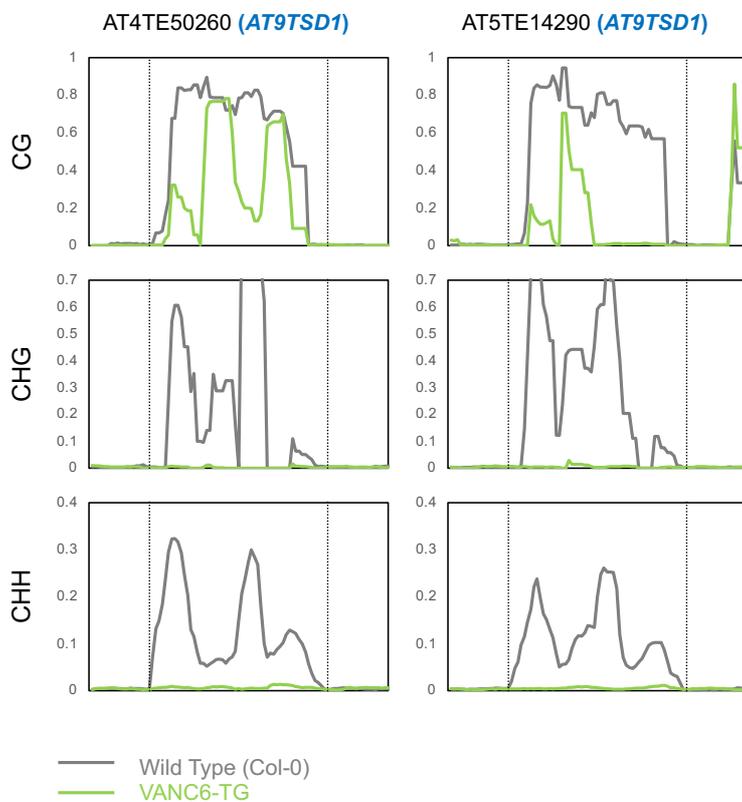
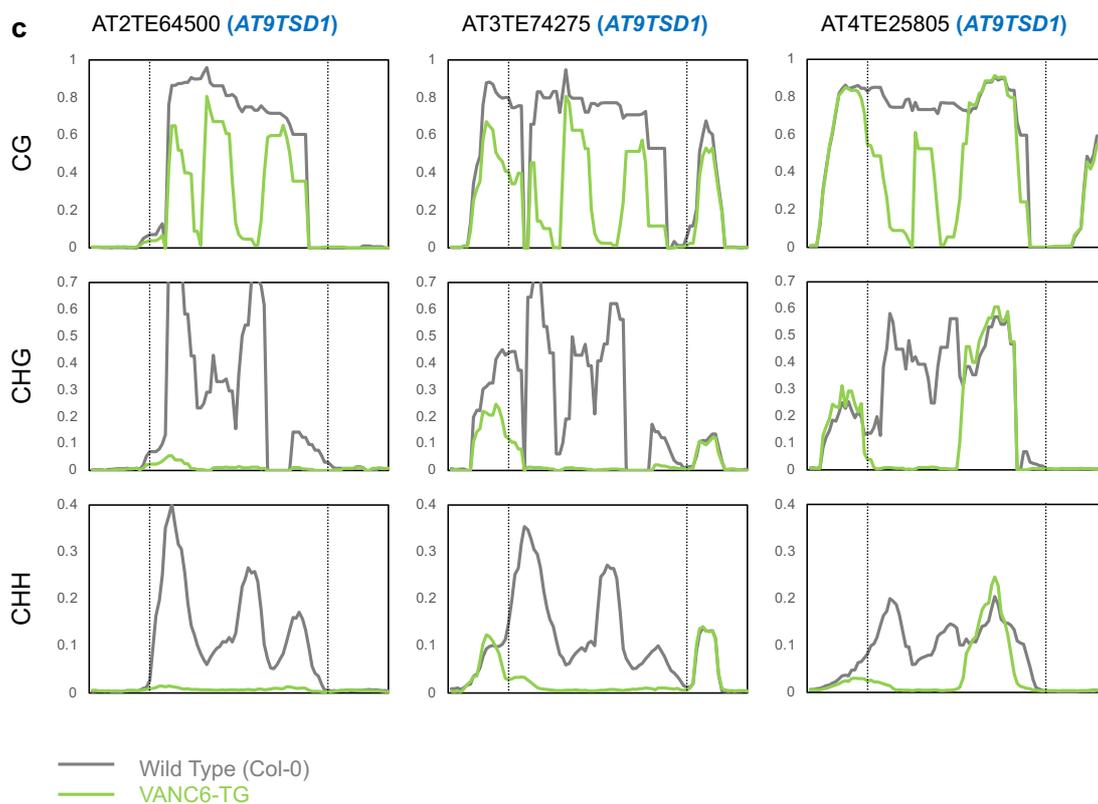
Fig 2. TEs with reduced methylation in *VANC* transgenic plants.

a-c DNA hypomethylation for each TE at CHG sites and CHH sites in the *VANC6* transgenic plant (**a**), *VANC2* transgenic plant (**b**), and *VANC14* transgenic plant (**c**). X and Y axes indicates changes in methylation at CHG and CHH sites, respectively. For each of them, significance of decrease in methylation was assessed by the value $(Mn/Cn - Mt/Ct) / (1/\sqrt{Cn} + 1/\sqrt{Ct})$, where Mn, Cn, Mt and Ct are methylated cytosine (M) and total cytosine (C) counts mapped for each TE in the non-transgenic (n) and transgenic (t) plants, respectively. These panels don't include TE copies which either Cn or Ct was zero because the above formula cannot be calculated. In panels (**a**), hypomethylated copies including *VANDAL6* are colored green, *VANDAL7*,

VANDAL8, *VANDAL17*, *AT9TSD1* are blue, unaffected copies are black, respectively.







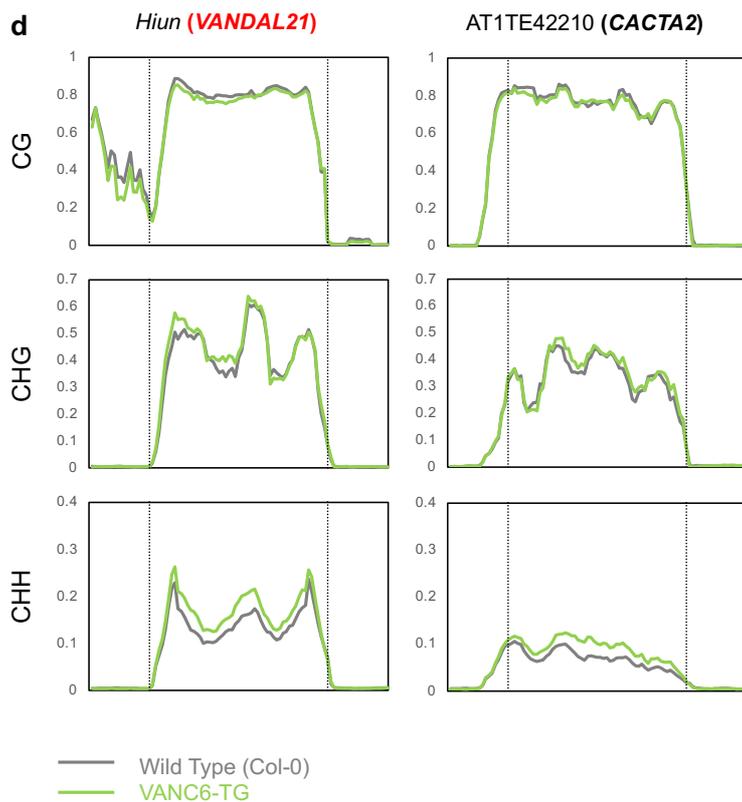
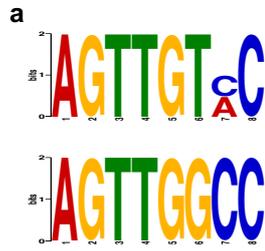


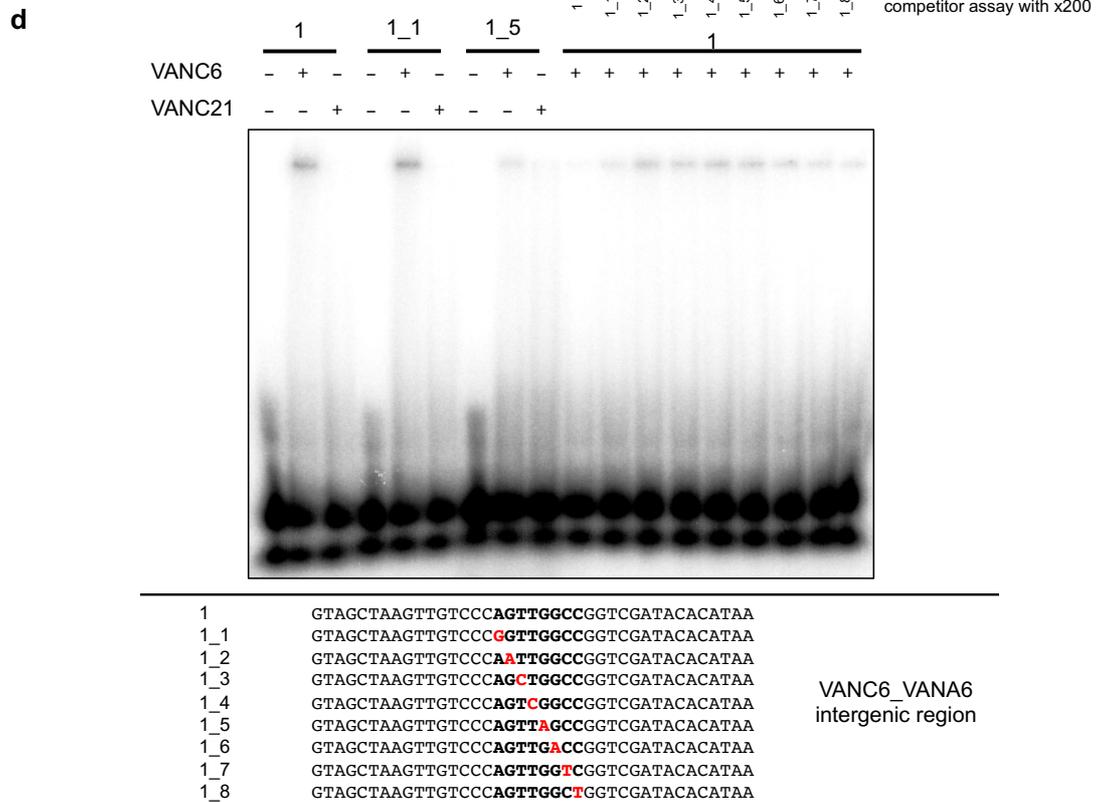
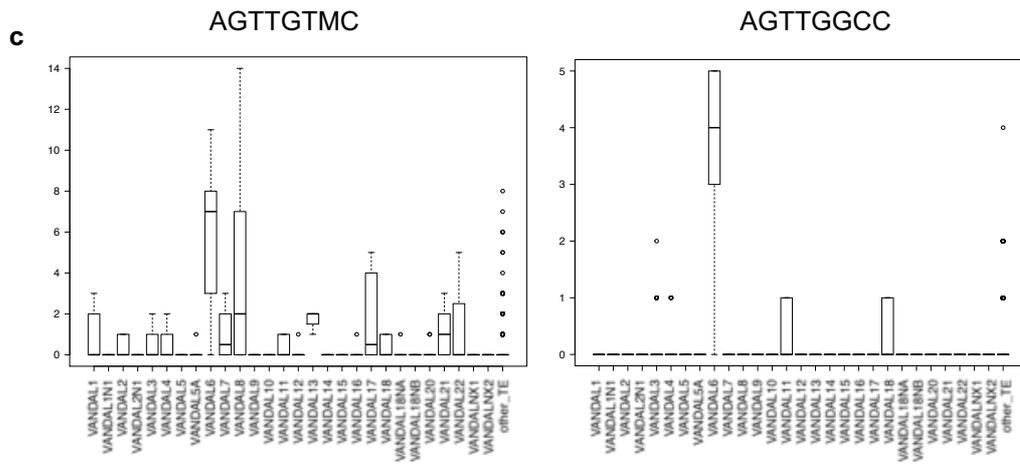
Fig 3. The methylation status of TEs in VANC6 transgenic plants.

DNA methylation percentage levels of *VANDAL6* copies (a), *VANDAL8* copies (b), *AT9TSD1* copies (c), and other TE copies (d) in *VANC6* transgenic plants and wild-type plants. Broken lines show TE ends. Each point represents proportion of methylated cytosine for a sliding window with seven fractions after separating each TE for 100 fractions. Right and left flanking regions are also analyzed by the same conditions.

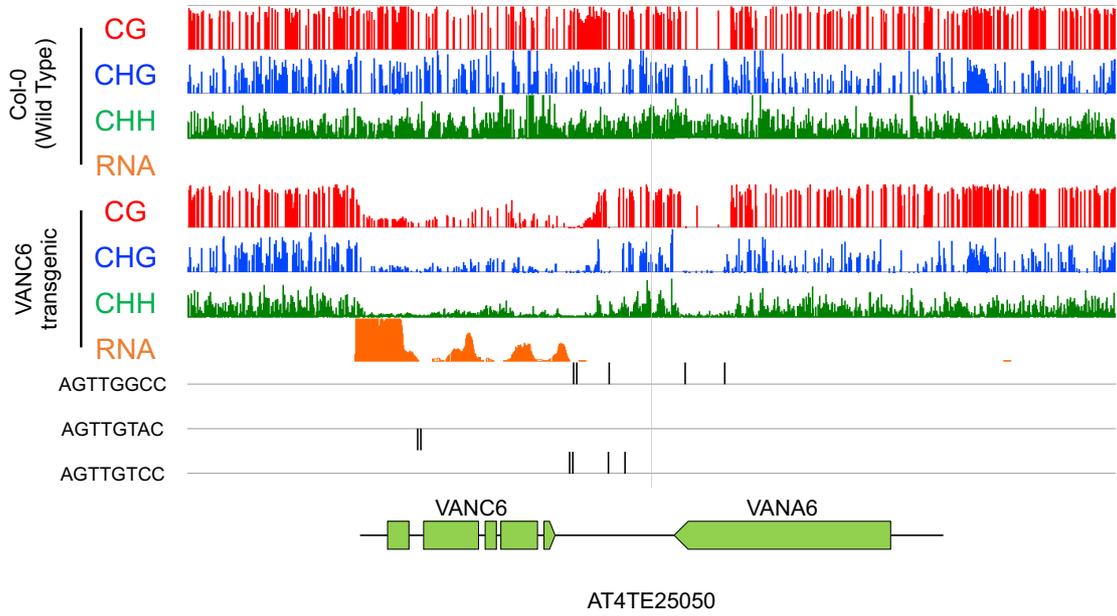


b

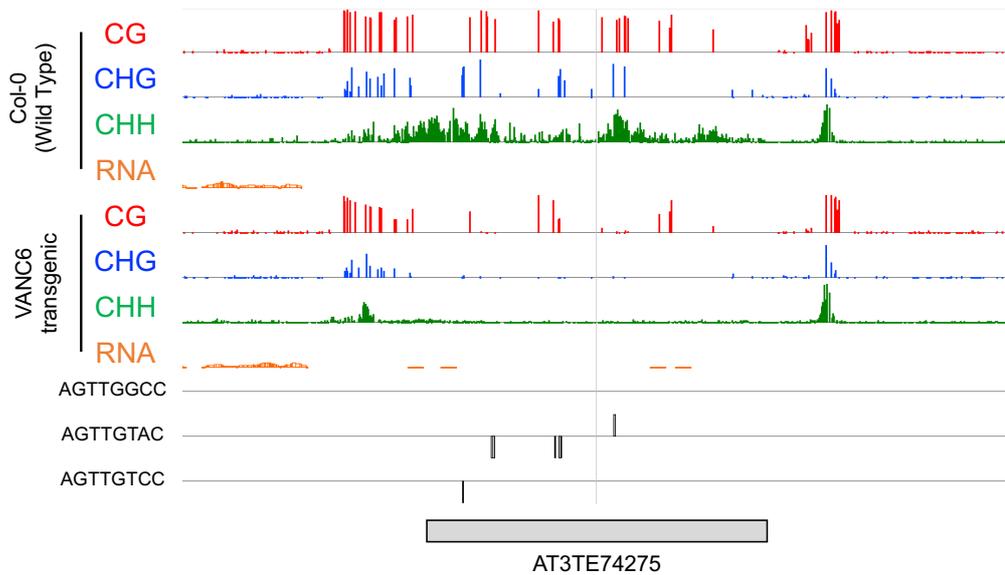
identified motif	pos	neg	p-value	e-value
AGTTGTMC	64	8	1.6e-012	5.7e-008
AGTTGGCC	34	0	4.7e-011	1.7e-006
TACAHATA	79	21	1.1e-009	4.0e-005
CCAWC	391	278	2.8e-007	1.0e-002
AAAACAHA	142	73	5.5e-007	1.9e-002



e AT4TE25050, VANDAL6



f AT3TE74275, AT9TSD1



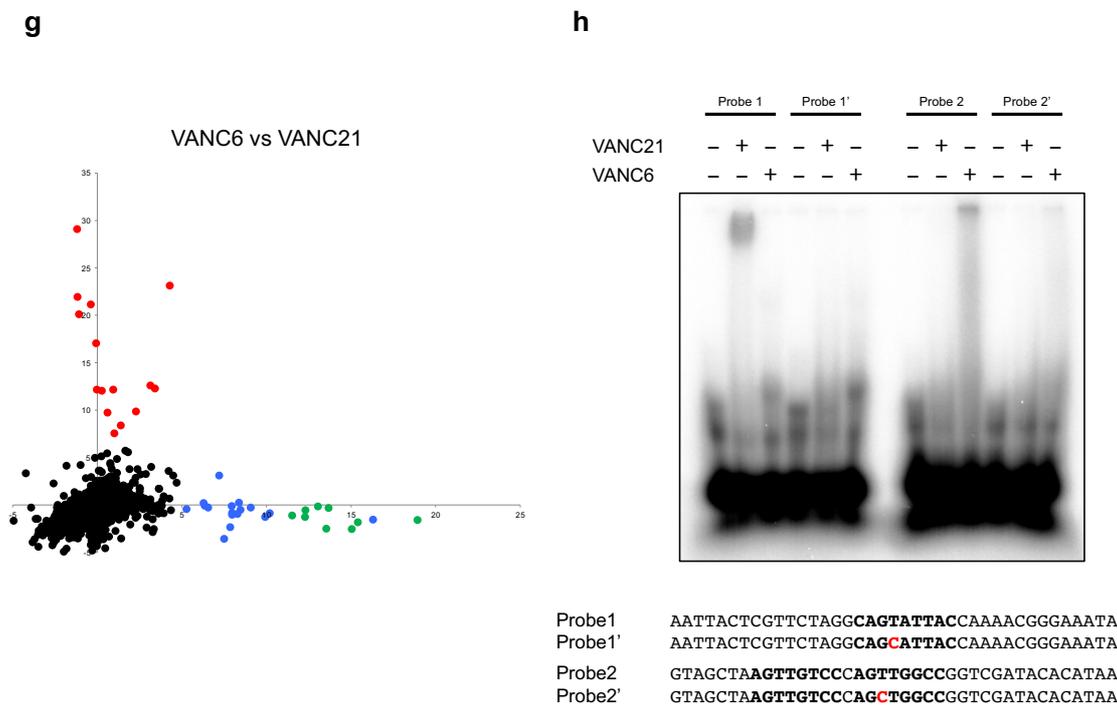


Fig 4. VANC6 has a unique DNA- binding motif.

- a** Overrepresented motif found in differentially-hypomethylated regions at CG-sites (CG-hypoDMRs) induced by *VANC6* genes using DREME software.
- b** List of motifs identified by DREME software. Pos/neg means positive or negative sequences matching the motif in CG-hypoDMR candidate sequences, respectively. The p-value means Fisher's Exact Test for enrichment of the motif in the positive sequences. E-value means the motif p-value times the number of candidate motifs tested.
- c** Number of "AGTTGTMC" (M=A or C) and AGTTGGCC motif in *VANDAL* family members. TEs longer than 1kb were characterized.
- d** EMSA with *VANC6* proteins. Sequences of the dsDNA probe used in EMSA are listed in the bottom. Competitor assay consists of labeled probe 1 with non-labeled probe1-1~1-8 by concentration x 200.

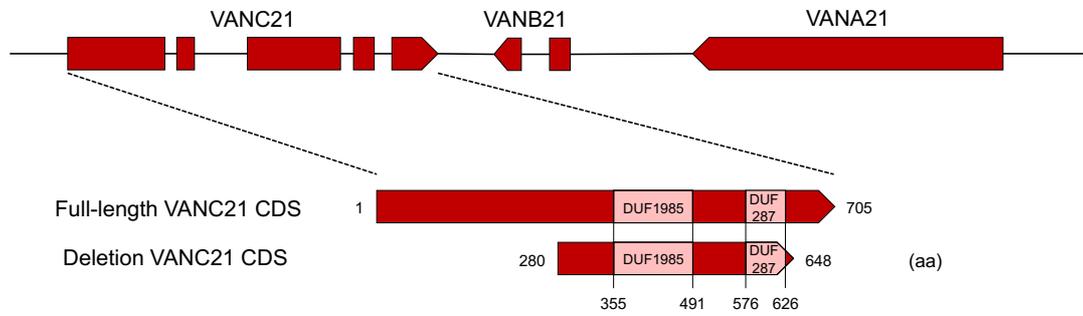
e-f DNA methylation (CG, CHG, CHH) and transcription status image described by Integrated Genomic Browser (IGB) in one copy of *VANDAL6* (AT4TE25050, **e**), and one copy of *AT9TSD1* (AT3TE74275, **f**). Scales of DNA methylation are percentages from 0 to 100, and that of RNA indicates the number of reads of RNA-seq with the maximum being 500. Motifs present in watson chain are shown as + and those present in crick chain as -. Identified motifs in CG-hypoDMRs in *VANC6* transgenic plants were also found in *AT9TSD1*.

g Comparison of DNA hypomethylation between *VANC21* and *VANC6* transgenic plants at CHG sites. Hypomethylated copies including *VANDAL21* are colored red, *VANDAL6* are green, *VANDAL7*, *VANDAL8*, *VANDAL17*, *AT9TSD1* are blue, unaffected copies are black, respectively.

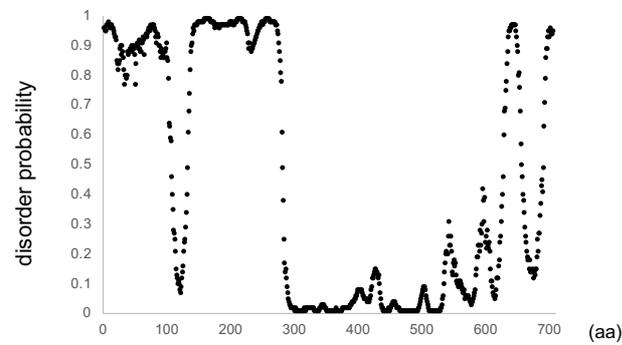
h EMSA with *VANC21* and *VANC6* proteins. Sequences of the dsDNA probe used in EMSA are listed in the bottom. The probe 2 has two motifs of *VANC6*, and the binding signal became weaker when one of the motifs has a single base substitution.

a

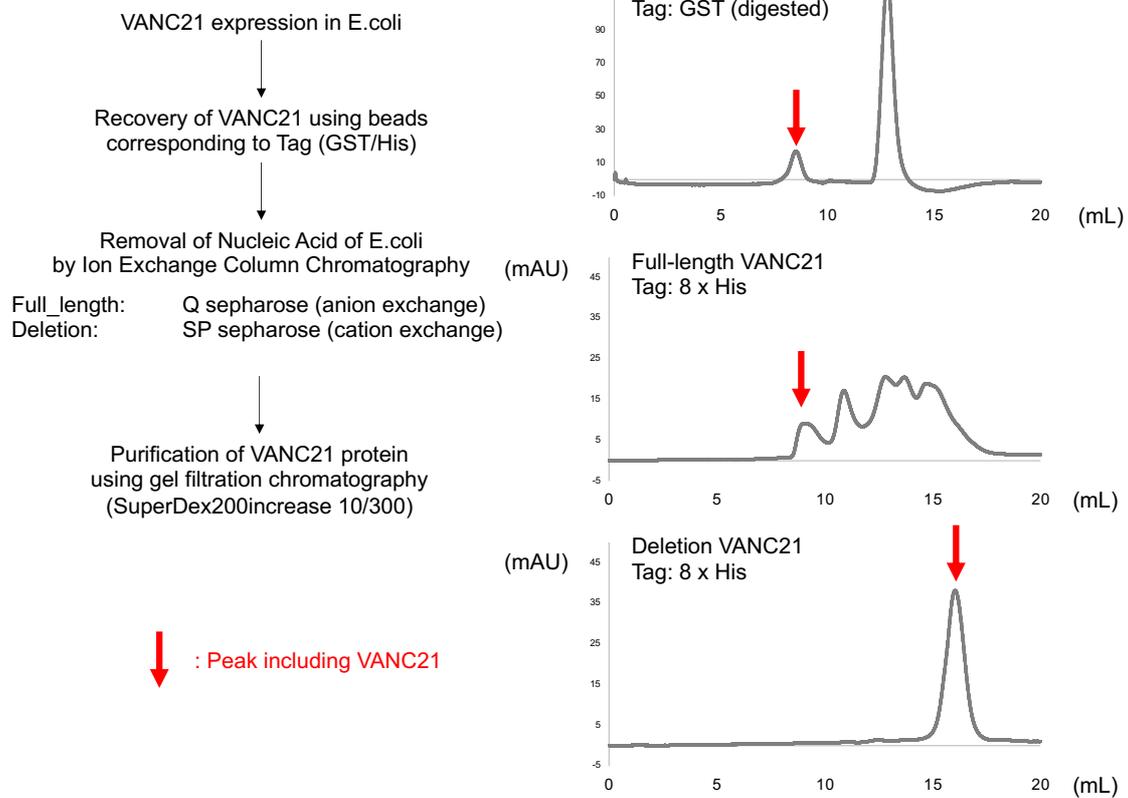
Hiun (AT2TE42810) VANDAL21

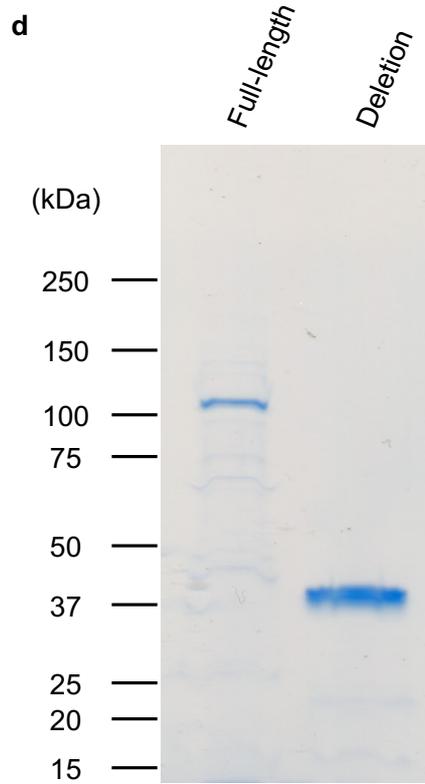


b



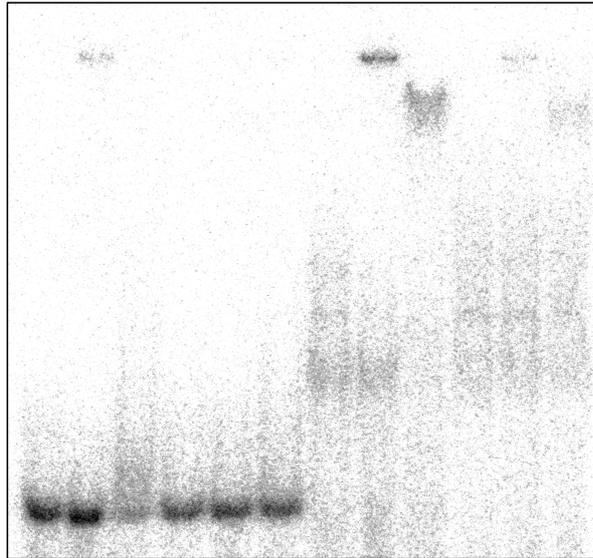
c





e

	Probe 1			Probe 1'			Probe 2			Probe 2'		
Full-length VANC21	-	+	-	-	+	-	-	+	-	-	+	-
Deletion VANC21	-	-	+	-	-	+	-	-	+	-	-	+



Probe 1 5-AATTACTCGTTC TAGGCAGTATTACCAAAACGGGAAATA-3

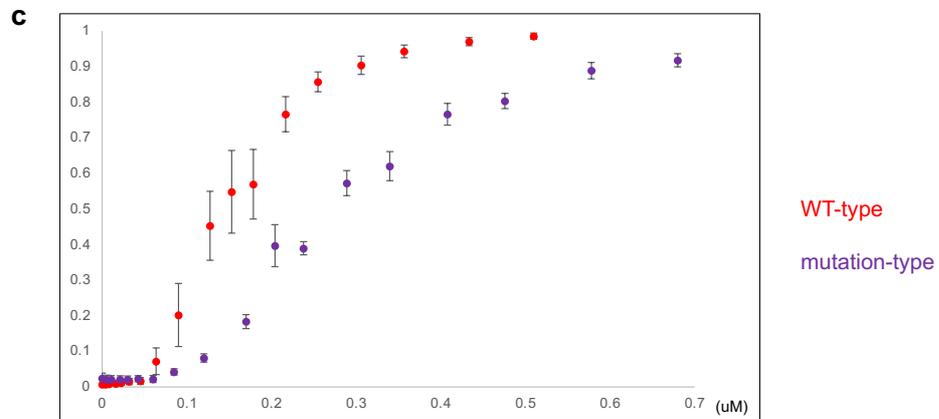
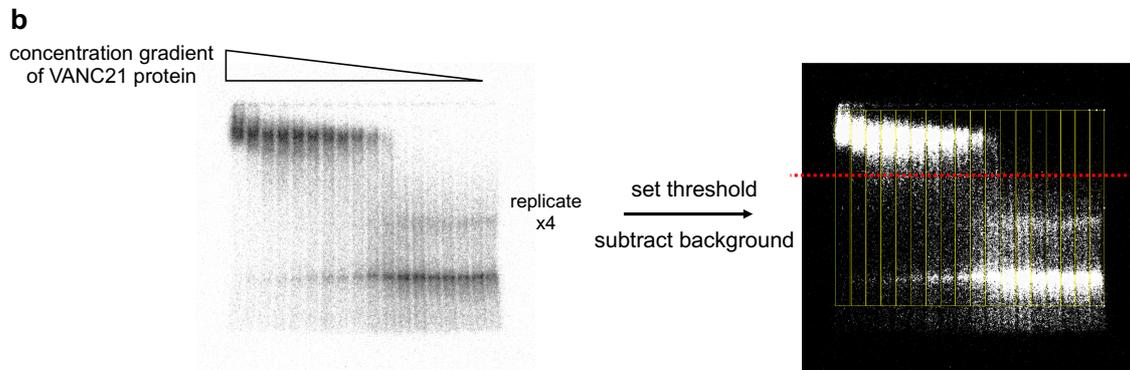
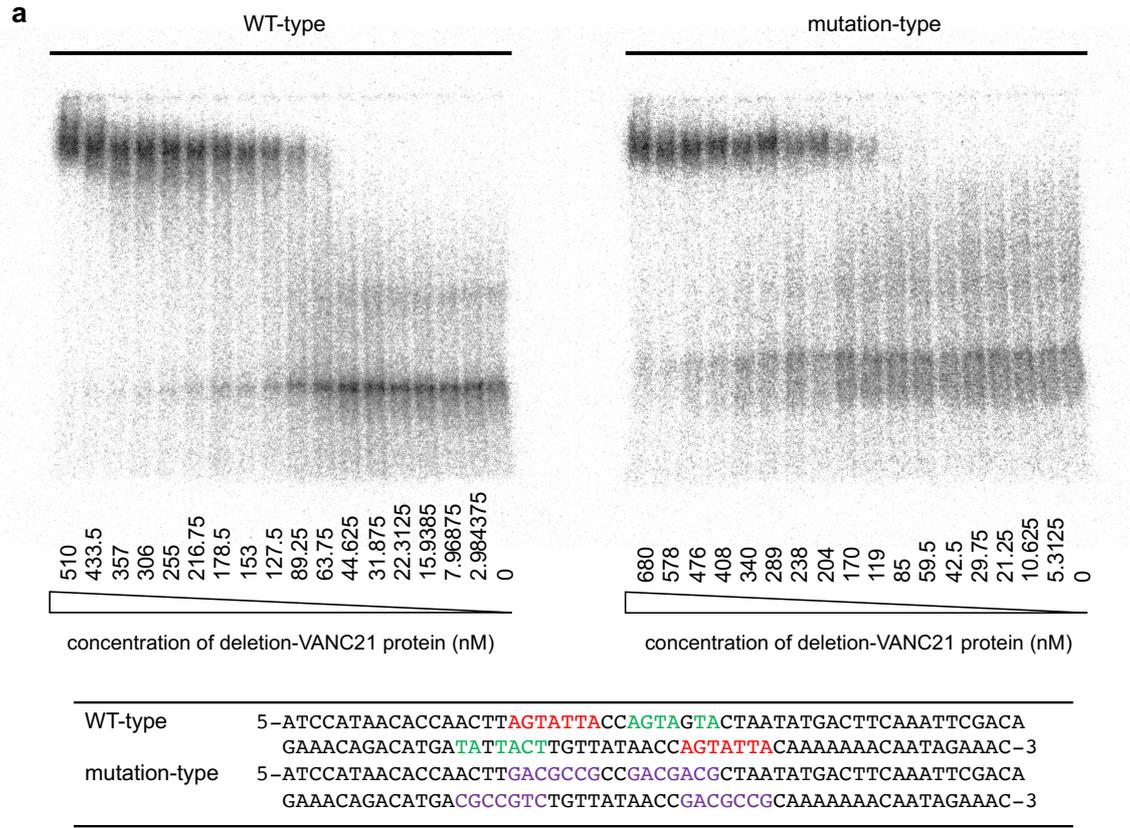
Probe 1' 5-AATTACTCGTTC TAGGCAGCATTA CCAAAACGGGAAATA-3

Probe 2 5-ATCCATAACCAACTTAGTATTACCAGTAGTACTAATATGACTTCAAATTCGACA
GAAACAGACATGATATTACTTGT TATAACCGATATTACAAAAACAATAGAAAC-3

Probe 2' 5-ATCCATAACCAACTTGACGCCGCCGACGACGCTAATATGACTTCAAATTCGACA
GAAACAGACATGACGCCGTCGT TATAACCGACGCCGCAAAAAACAATAGAAAC-3

Fig 5. VANC21 without disorder region also retains its DNA-binding activity

- a** Schematic diagram of full-length and deletion VANC21 gene. Boxes indicate exons.
- b** Disorder prediction of full-length VANC21 CDS. For prediction, DISOPRED3 in PSIPRED software ver 4.0 was used⁵³⁻⁵⁴.
- c** Process of VANC21 protein purification and results of gel filtration chromatography. Peaks where VANC21 was identified are indicated by red arrows.
- d** SDS-PAGE of full-length and deletion VANC21 protein.
- e** EMSA with full length and deletion VANC21 protein. Sequences of the dsDNA probe used in EMSA are listed in the bottom. The free probe 2 show smear band because of its secondary structure caused by long length. Each type of VANC21 protein retain DNA-binding activity. DNA-binding specificity of VANC21 protein became weaker when YAGTATTAY motif was mutagenized.



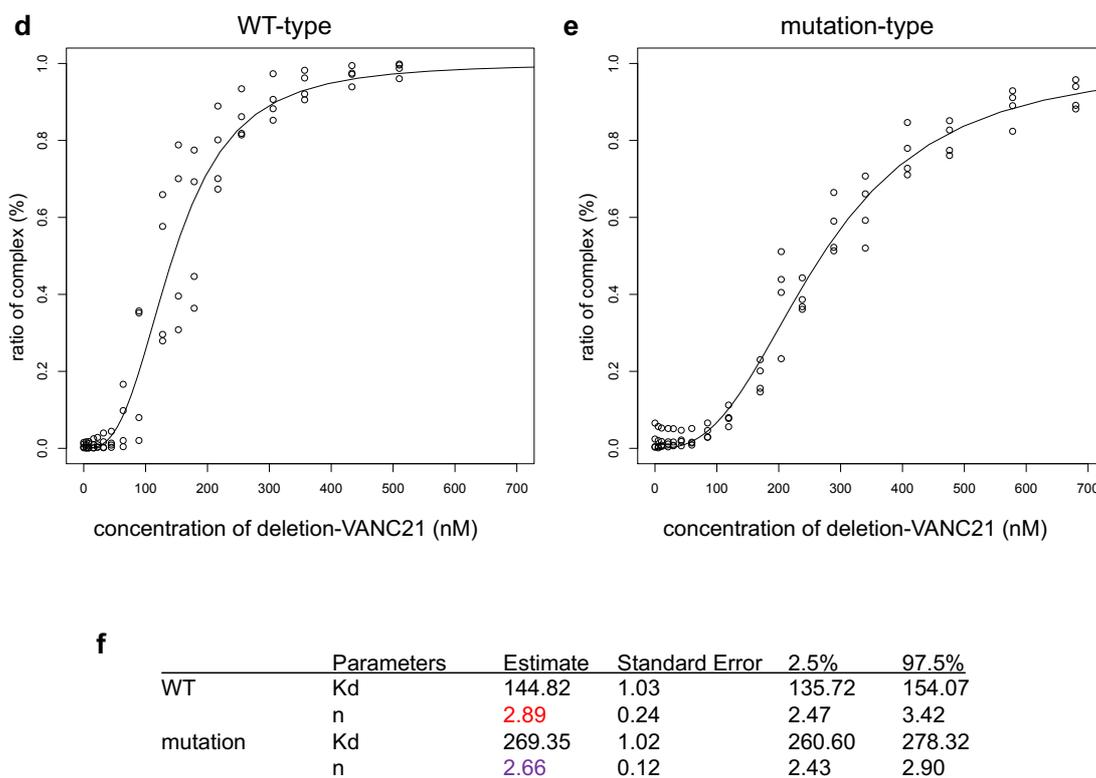


Fig 6. VANC21 protein binds to DNA with positive cooperativity.

a EMSA of deletion VANC21 protein with concentration gradient. Non-specific competitor was not included because of high purity of deletion VANC21 protein. Sequences of the dsDNA probe used are listed in the bottom.

b Process for calculating kinetics; performed EMSA four times as biological replicate by applying gradient to the concentration of VANC21 protein. EMSA data were subtracted background to calculate the ratio of free probe to complex.

c DNA-binding kinetics of deletion VANC21 protein. The percentage of complexes in each lane was calculated.

d-e Fitting of kinetics to a Hill plot using the nonlinear least squares method for WT-type probe (**c**) and mutation-type probe (**d**). R was used for the fitting (see Methods).

f Estimates and standard errors for each parameter obtained by fitting.
2.5% and 97.5% means 95% confidence interval.

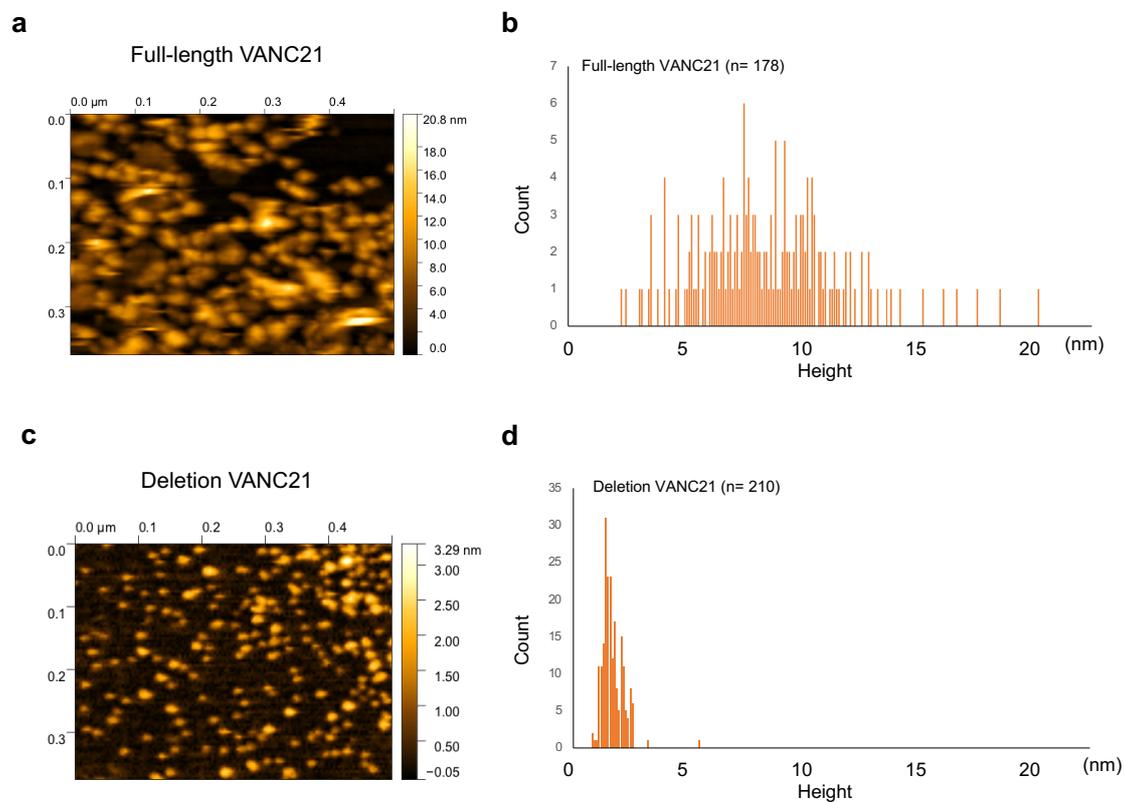


Fig 7. Full-length and deletion VANC21 protein on the AP-mica surface.

a, c Full-length VANC21 protein (**a**) or deletion VANC21 protein (**c**) observed in AP-mica.

b, d Maximum height of full-length VANC21 proteins observed in AP-mica (**b**) or deletion VANC21 protein (**d**). Gwyddion software was used to measure the height (see Methods).

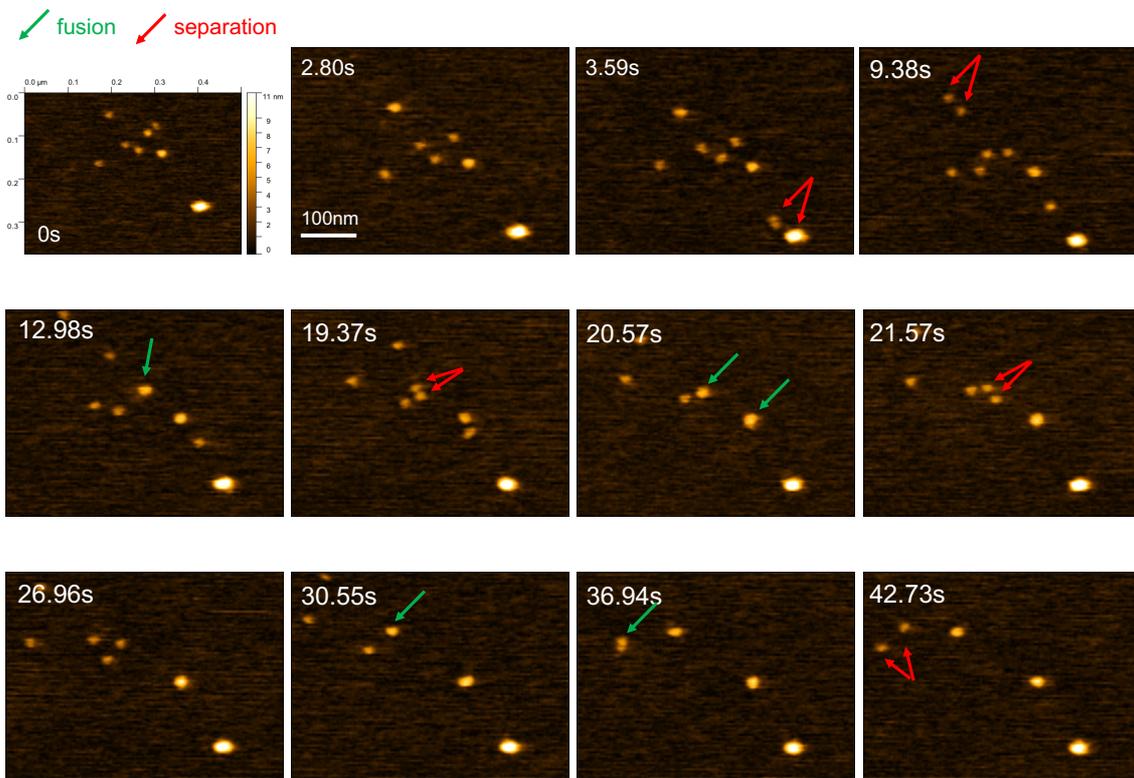
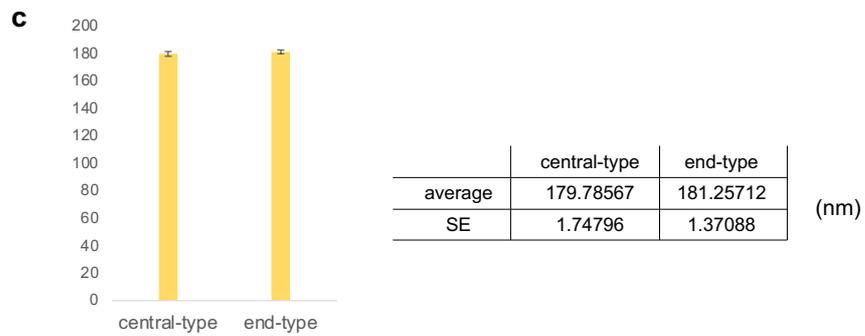
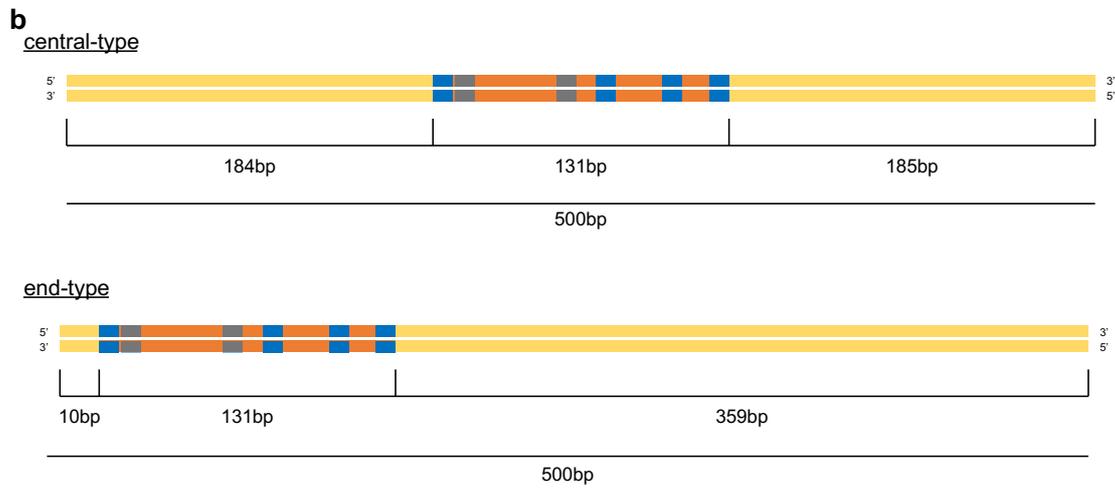
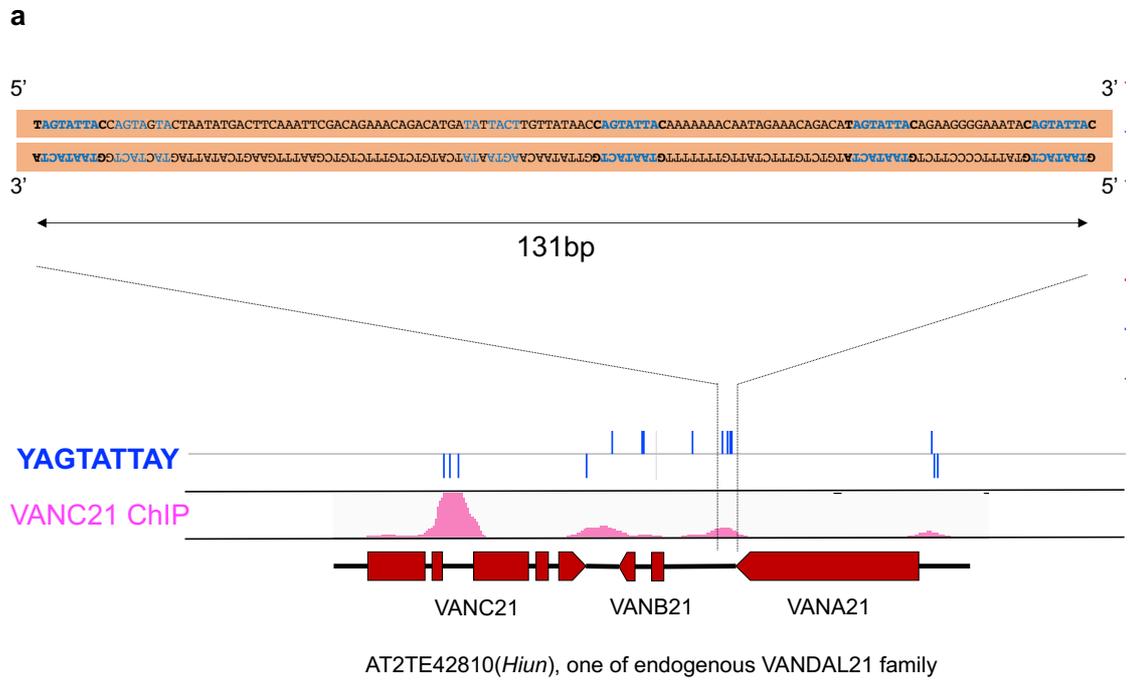
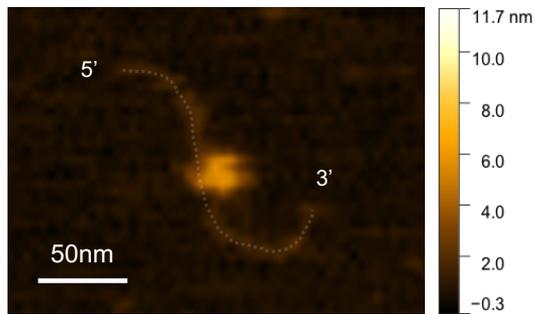


Fig 8. Full-length VANC21 protein form multimer independent of DNA-binding.

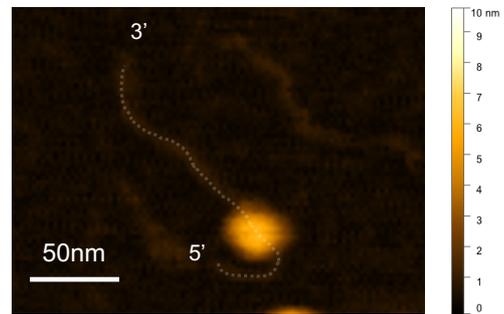
Images of full-length VANC21 protein in dynamic motion on the surface of mica without treatment of 3-aminopropyltriethoxysilane. Green and red arrows indicate scenes of VANC21 protein binding or separation, respectively.



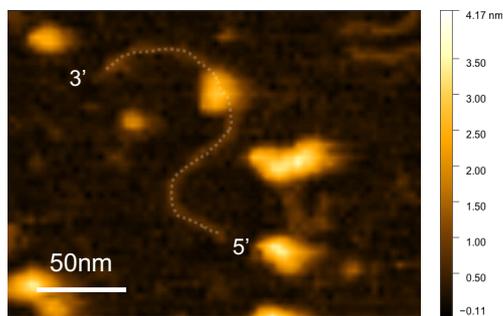
d Full-Length VANC21 + central-type DNA



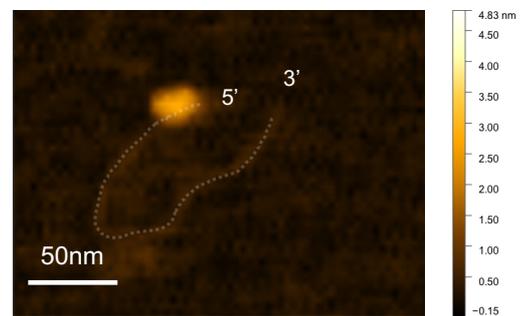
e Full-Length VANC21 + end-type DNA



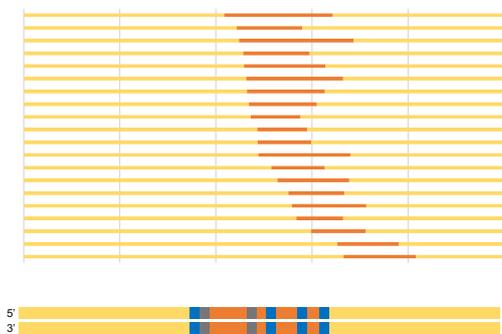
f Deletion VANC21 + central-type DNA



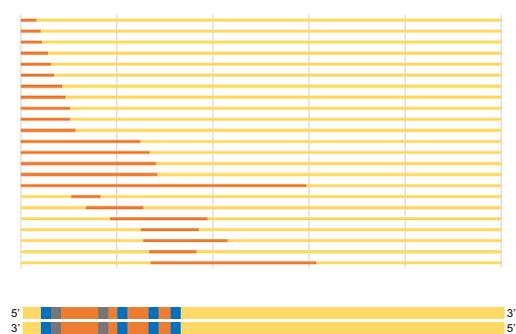
g Deletion VANC21 + end-type DNA



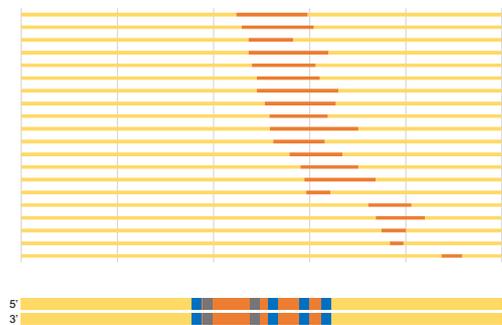
h Full-Length VANC21 + central-type DNA (n=18)



i Full-Length VANC21 + end-type DNA (n=23)



j Deletion VANC21 + central-type DNA (n=20)



k Deletion VANC21 + end-type DNA (n=21)

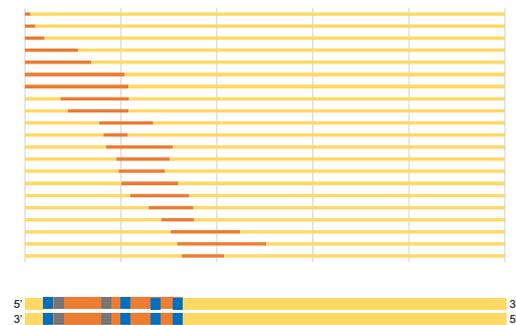


Fig 9. VANC21 protein binds to YAGTATTAY motif on the AP-mica surface.

a Motif-rich region adopted for HS-AFM observation, with four motifs and two pseudo motifs in the 131 bp region. Motif sequences and motif with SNP are colored by blue and pale blue, respectively.

b Schematic diagram of the DNA sequence used for HS-AFM observation. Two DNA-types were prepared to evaluate the specificity of DNA-binding.

c Observed DNA length of central-type and end-type on AP-mica (n=12).

d-g One HS-AFM observation image to which full-length VANC21 protein and deletion VANC21 protein binds to DNA on the AP-mica surface.

h-k Binding sites on the DNA of full-length VANC21 protein or deletion VANC21 protein, where the length of DNA was set to 180 nm and the length of flanking DNA was measured (see Methods).

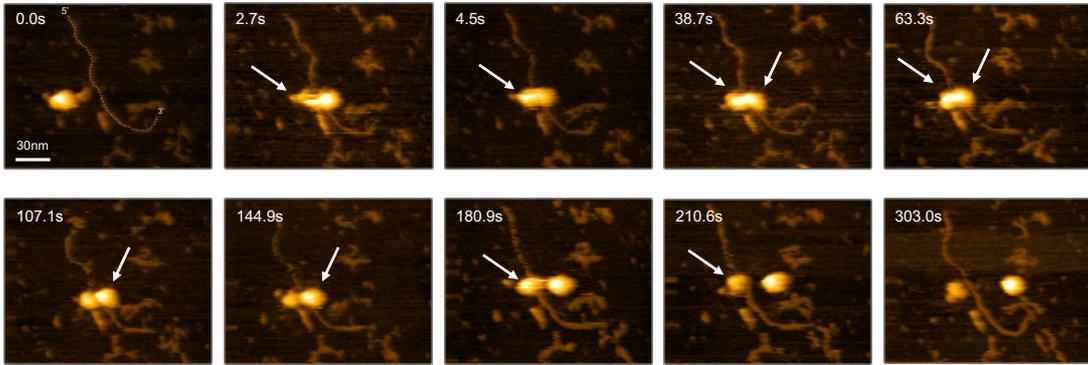
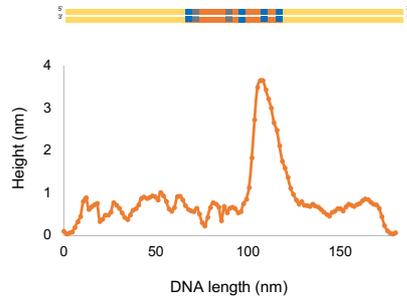
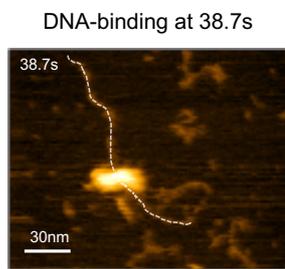
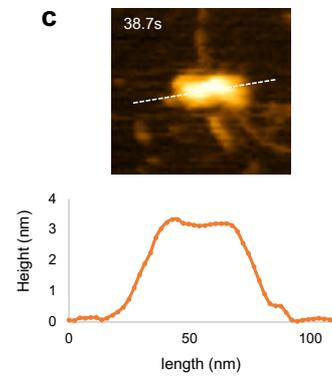
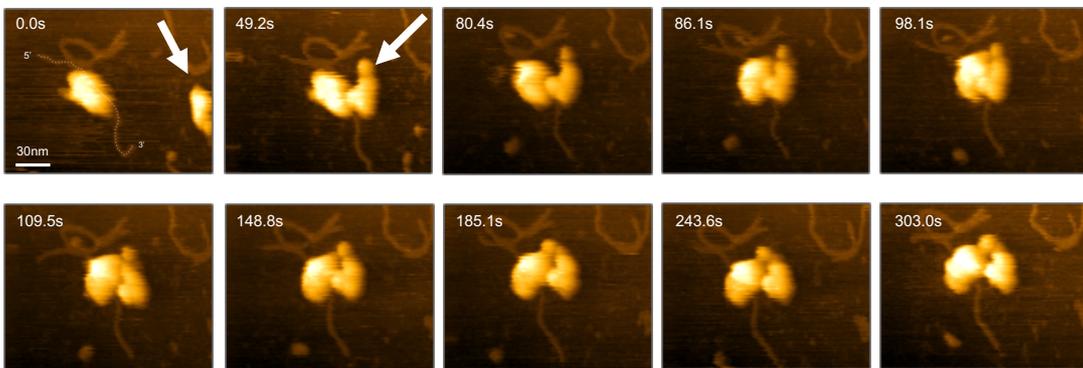
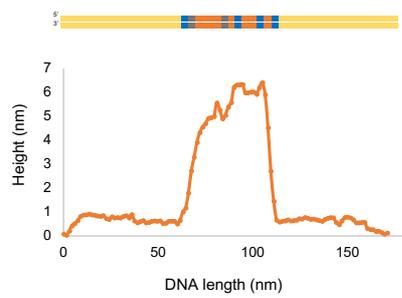
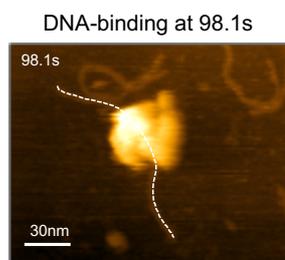
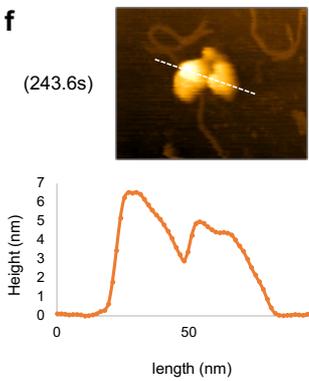
a**b****c****d****e****f**

Fig 10. VANC21 protein binds to DNA with multimeric forms.

- a** Images of full-length VANC21 protein binding to DNA in dimeric form. VANC21 protein bound to DNA is indicated by white arrows.
- b** Height of VANC21 protein with DNA at 38.7s in (a).
- c** Height of VANC21 protein at 38.7s in (a).
- d** Images of full-length VANC21 protein binding to DNA in multimeric form. The newly bound multimer is indicated by white arrows.
- e** Height of VANC21 protein with DNA at 98.1s in (d).
- f** Height of VANC21 protein at 243.6s in (d).

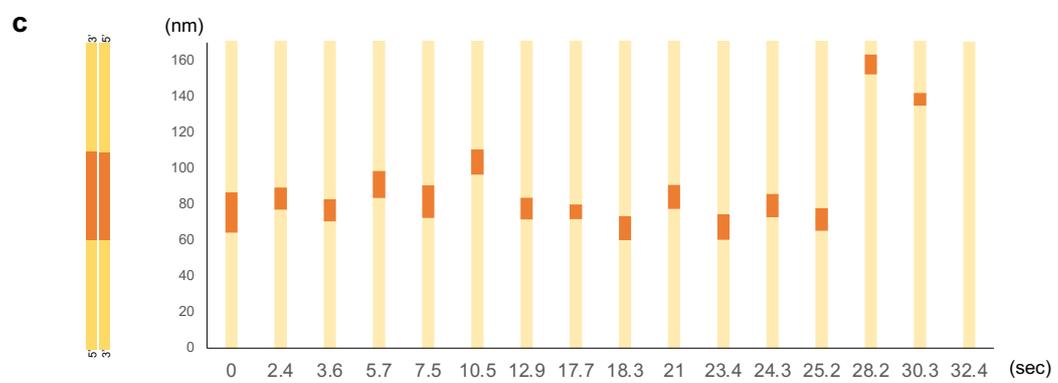
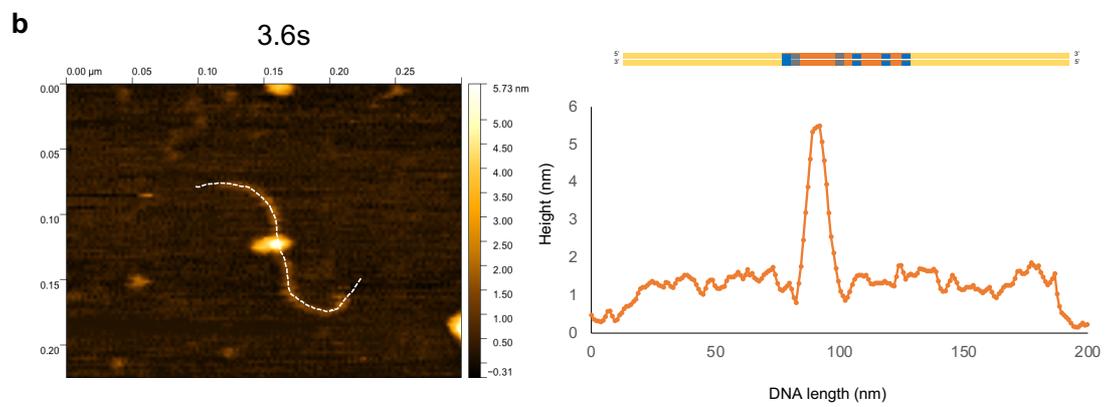
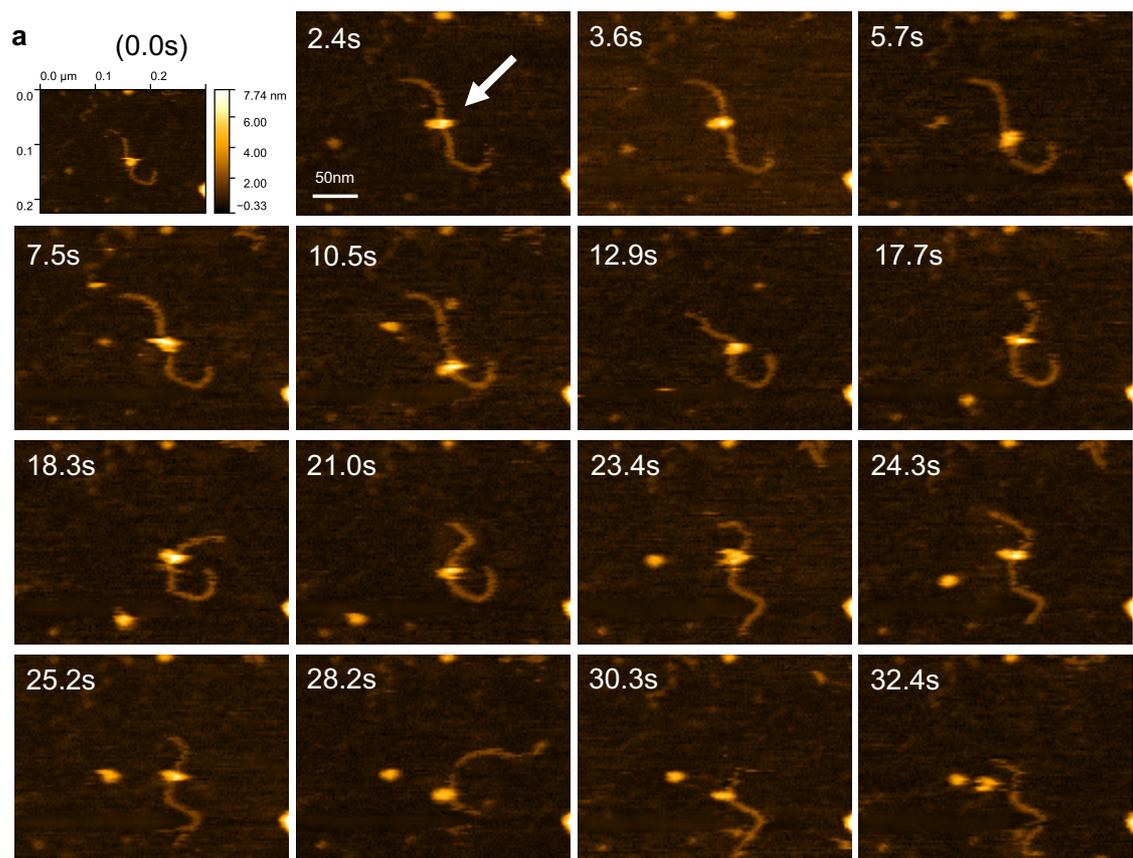


Fig 11. Full-length VANC21 monomer fluctuates on DNA.

a Images of full-length VANC21 protein fluctuating around as it binds to DNA with monomers.

b Height of VANC21 protein at 3.6s in (a).

c Binding sites on the DNA of full-length VANC21 protein, where the length of DNA was set to 180 nm and the length of flanking DNA was measured (see Methods, like Fig 8d).

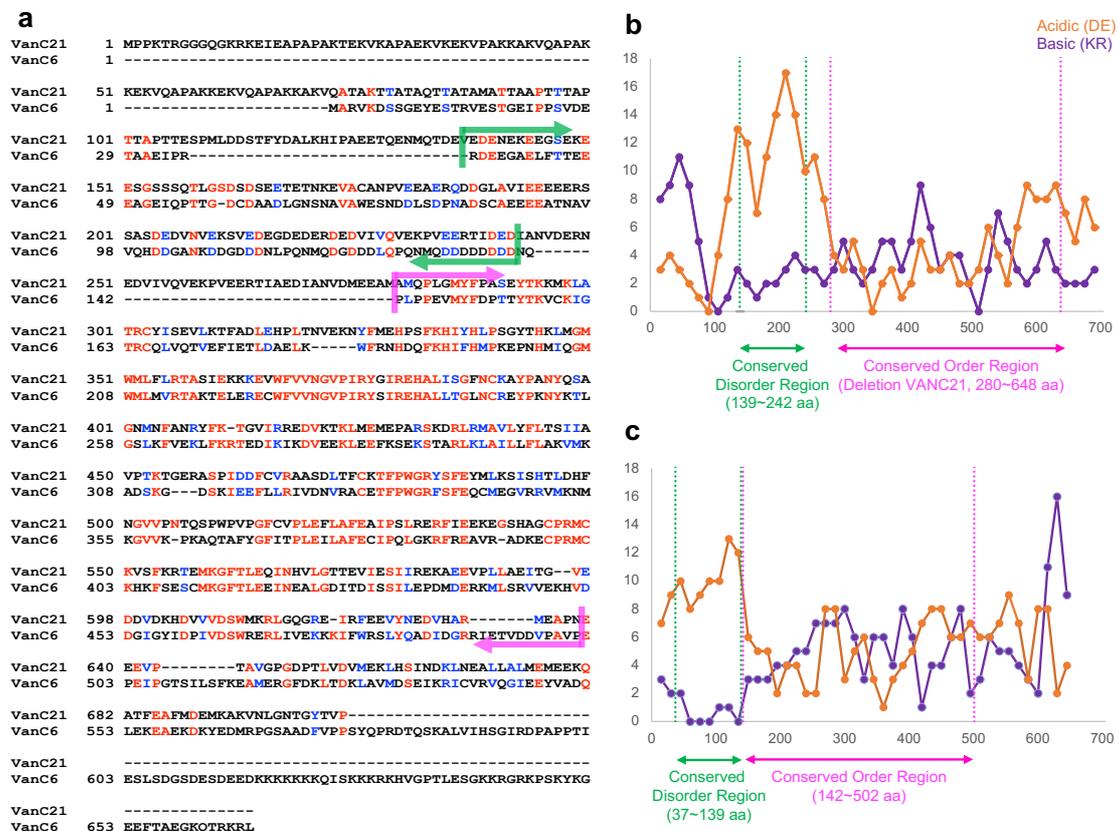


Fig 12. Amino acid composition of VANC proteins and its comparison.

a Comparison of amino acids between VANC21 and VANC6. Identical, similar, and different amino acids are shown in red, blue, and black, respectively. Green and pink arrows indicate the Conserved Disorder Region and Conserved Order Region, respectively (b and c).

b-c Distribution of acidic and basic amino acids in VANC21 protein (b) and VANC6 protein (c). Each point represents the number of acidic (D and E) or basic (K and R) amino acids at 30 amino acids with 15 amino acids allowed to overlap.

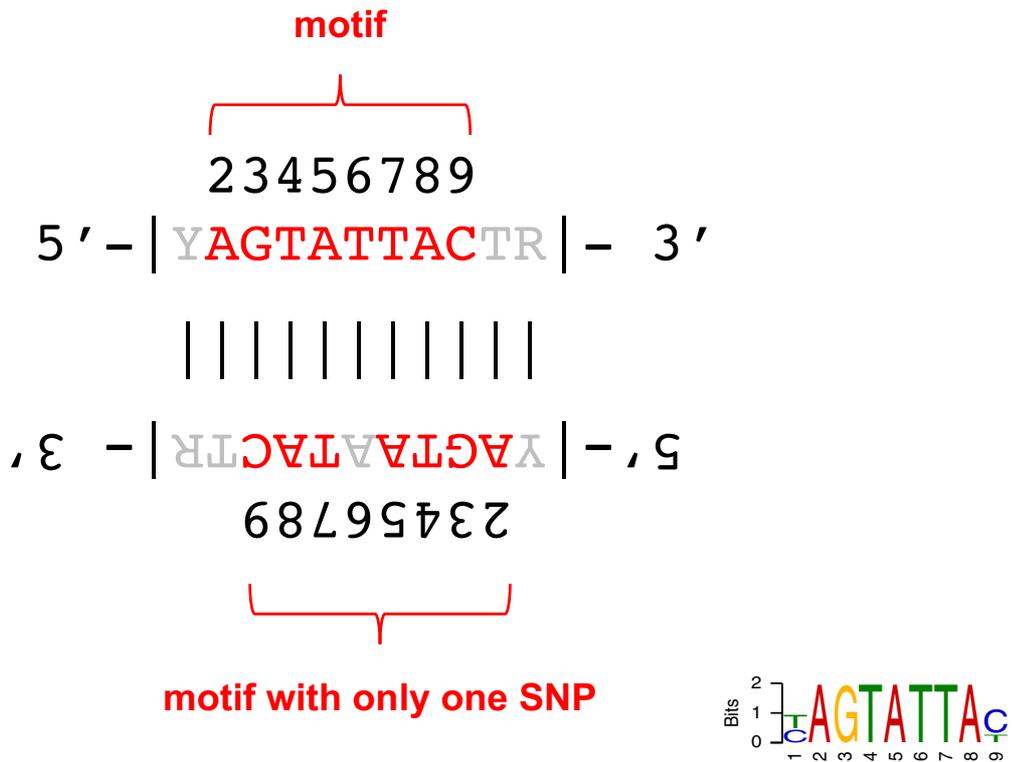
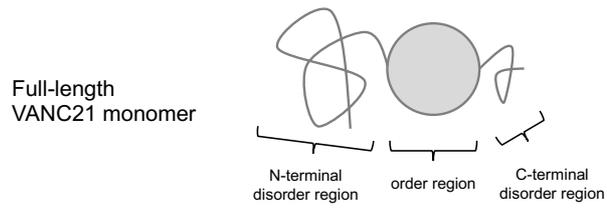


Fig 13. The YAGTATTAY motif is a partial palindromic sequence.

Schematic diagram of the both strand of YAGTATTAY sequence when the motif is a “C-type motif”²⁸. When the nucleotide following the 3' end of the “C-type motif” is T, the YAGTATTAY motif becomes a partial palindromic sequence with one SNP allowed.

a



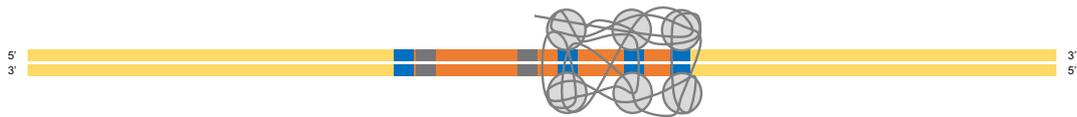
1. one monomer binds to one motif



2. another monomer come to same motif, and be stable on DNA with help of positive cooperative effect



3. form various types of multimer using disorder region



b

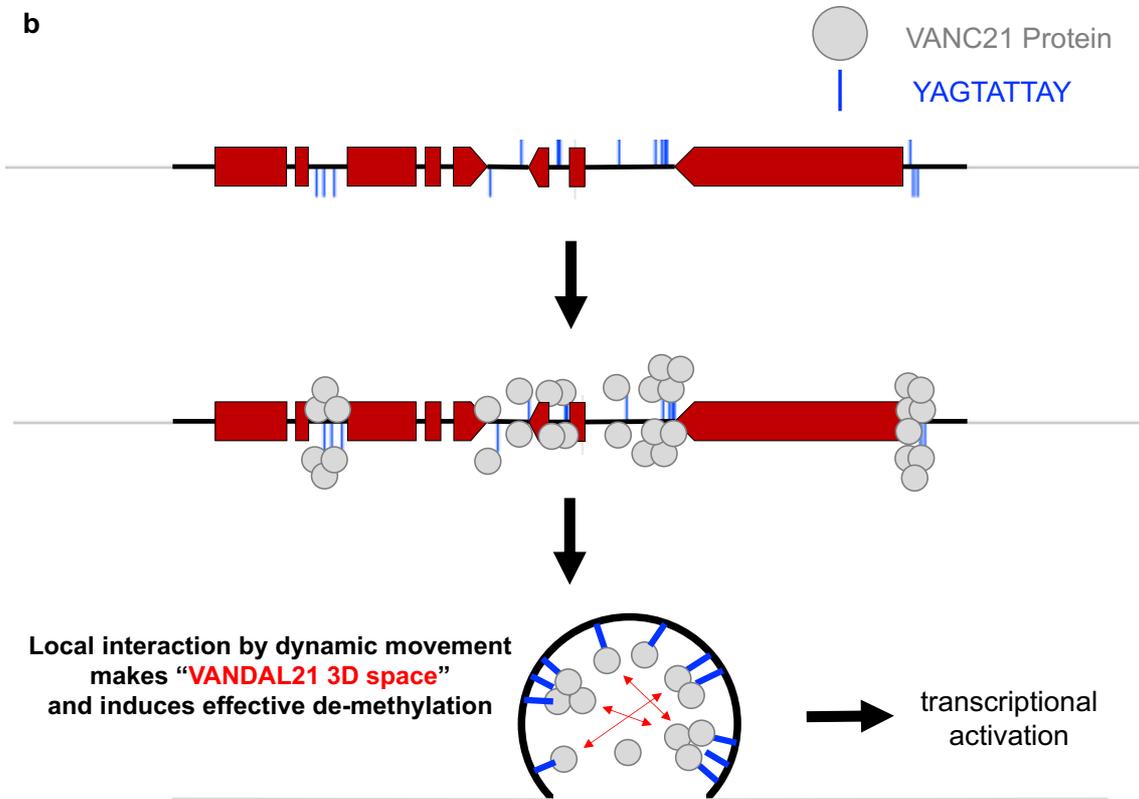
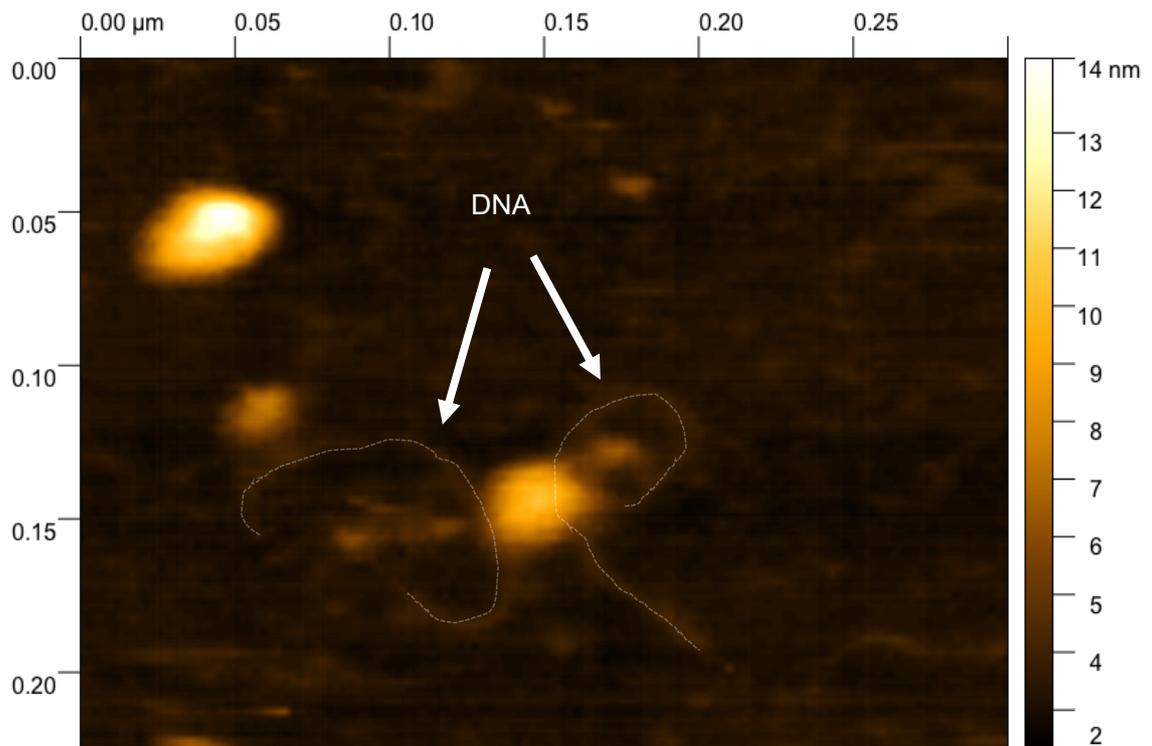
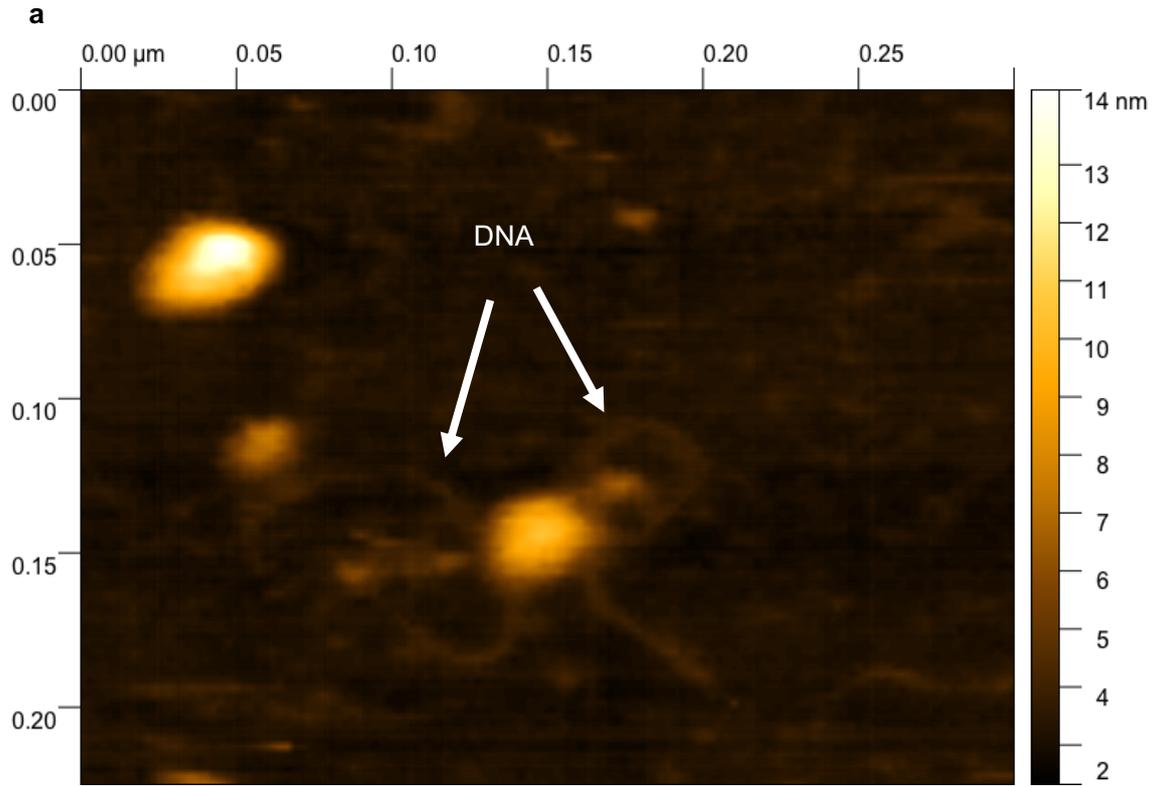


Fig 14. Model of DNA-binding mode of VANC21 protein

- a** DNA binding model for VANC21 protein in the motif-rich region.
- b** Model for the action of VANC21 protein at full length of *VANDAL21* sequence.



b

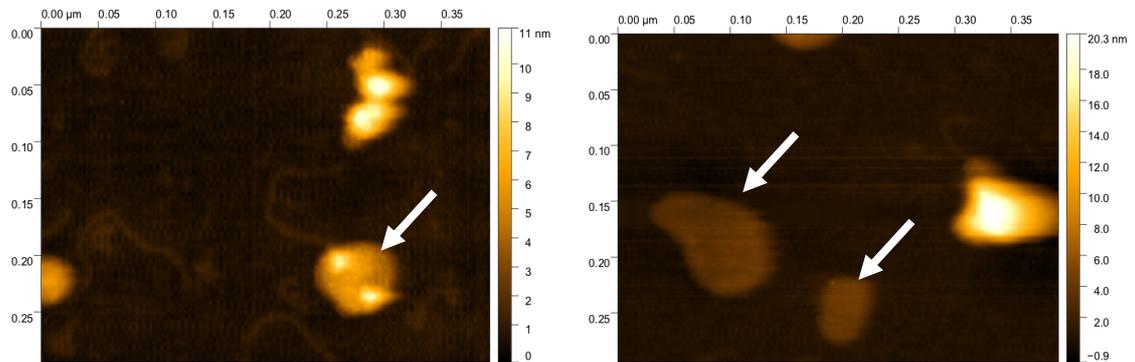


Fig 15. The images VANC21 protein

a The image that VANC21 multimer bundle together two different DNA fragments.

b Unusual structure of full-length VANC21 protein like oil droplet.

VANC6-TG replicate #1

	ID	family	super_family	ΔCG	ΔCHG	ΔCHH
1	AT4TE25050	VANDAL6	DNA/MuDR	68.18035859	78.53861513	79.54629675
2	AT4TE45190	VANDAL8	DNA/MuDR	22.74859111	71.04233457	80.97639791
3	AT2TE43165	VANDAL6	DNA/MuDR	18.78943407	55.39237751	70.62598179
4	AT5TE64600	VANDAL6	DNA/MuDR	17.36728646	54.8789873	73.9575903
5	AT4TE67560	VANDAL6	DNA/MuDR	14.29248902	53.42903421	68.78565594
6	AT3TE70635	VANDAL6	DNA/MuDR	16.02997267	46.81902598	68.23760986
7	AT3TE44950	VANDAL6	DNA/MuDR	37.4374756	80.15719091	85.29862006
8	AT1TE27295	VANDAL6	DNA/MuDR	15.96263833	54.43284901	70.54935026
9	AT5TE53465	VANDAL6	DNA/MuDR	15.98678367	45.28012637	65.01144902
10	AT2TE08135	VANDAL6	DNA/MuDR	13.76006272	47.00003332	51.38375859
11	AT5TE14290	AT9TSD1	DNA/MuDR	85.01101881	98.58117784	95.07756604
12	AT5TE34730	VANDAL8	DNA/MuDR	6.409252479	40.27223244	48.92691361
13	AT1TE53635	VANDAL8	DNA/MuDR	7.669681044	39.23311362	55.49476684
14	AT2TE64500	AT9TSD1	DNA/MuDR	46.75220041	96.6704014	94.88064758
15	AT2TE08100	VANDAL7	DNA/MuDR	43.94842203	97.36534297	85.65792701
16	AT4TE22970	VANDAL8	DNA/MuDR	4.462223283	38.18364997	43.82703825
17	AT1TE44125	VANDAL6	DNA/MuDR	14.93990283	38.33039061	50.86734725
18	AT1TE53660	VANDAL8	DNA/MuDR	11.81626332	40.94838183	56.44835548
19	AT3TE74275	AT9TSD1	DNA/MuDR	63.52785541	98.03123454	93.3305491
20	AT3TE67445	VANDAL7	DNA/MuDR	4.23005794	23.68946981	20.68138111
21	AT4TE25065	VANDAL7	DNA/MuDR	5.743930336	30.01127098	33.42069101
22	AT3TE22200	VANDAL17	DNA/MuDR	2.329971396	23.08720193	15.37739283
23	AT4TE50260	AT9TSD1	DNA/MuDR	49.99885867	98.98633946	94.56093226
24	AT5TE44235	VANDAL6	DNA/MuDR	14.45745931	38.72325478	45.11566932
25	AT4TE25805	AT9TSD1	DNA/MuDR	39.06562677	50.90437647	59.50450117

VANC6-TG replicate #2

	ID	family	super_family	ΔCG	ΔCHG	ΔCHH
1	AT4TE25050	VANDAL6	DNA/MuDR	60.39530839	32.99369353	11.8689597
2	AT2TE43165	VANDAL6	DNA/MuDR	15.57221896	28.80250318	13.03635088
3	AT4TE45190	VANDAL8	DNA/MuDR	14.61646128	34.95408992	12.72597647
4	AT5TE64600	VANDAL6	DNA/MuDR	11.24972489	24.77077722	13.49591857
5	AT4TE67560	VANDAL6	DNA/MuDR	7.621541377	25.40629236	11.1896663
6	AT3TE44950	VANDAL6	DNA/MuDR	22.12788387	39.42401782	10.5554723
7	AT3TE70635	VANDAL6	DNA/MuDR	6.862846256	20.27692145	9.945731275
8	AT5TE53465	VANDAL6	DNA/MuDR	10.41818616	18.03180896	8.933016116
9	AT1TE27295	VANDAL6	DNA/MuDR	9.358064539	20.56308657	9.245053794
10	AT2TE08135	VANDAL6	DNA/MuDR	5.864472912	15.72373666	4.898866019
11	AT5TE14290	AT9TSD1	DNA/MuDR	66.61026192	44.02828199	9.978808065
12	AT1TE53635	VANDAL8	DNA/MuDR	4.74058947	15.65764243	7.024524539
13	AT5TE34730	VANDAL8	DNA/MuDR	2.20896784	13.35869447	4.87027411
14	AT2TE64500	AT9TSD1	DNA/MuDR	44.45509906	39.78387405	13.64431241
15	AT3TE74275	AT9TSD1	DNA/MuDR	51.34387149	37.69731654	11.75444229
16	AT1TE44125	VANDAL6	DNA/MuDR	8.703131314	13.29580468	5.626970974
17	AT1TE53660	VANDAL8	DNA/MuDR	7.076729194	16.58500133	8.917452841
18	AT4TE22970	VANDAL8	DNA/MuDR	0.682308921	10.98855201	3.688924573
19	AT3TE22200	VANDAL17	DNA/MuDR	1.757592982	7.989705842	1.839163859
20	AT4TE25065	VANDAL7	DNA/MuDR	1.647974875	9.362072832	2.930194871
21	AT3TE67445	VANDAL7	DNA/MuDR	0.774133239	8.263142431	2.225966003
22	AT4TE50260	AT9TSD1	DNA/MuDR	40.95822416	30.85154687	11.5520799
23	AT4TE25805	AT9TSD1	DNA/MuDR	27.32274457	22.94684076	6.862101074
24	AT5TE44235	VANDAL6	DNA/MuDR	6.739763001	10.06087662	3.230791867
25	AT4TE18470	VANDAL8	DNA/MuDR	2.023140037	7.657601679	2.257124118

Table1. Top 25 copies with a high degree of hypomethylation at CHH sites in VANC6 transgenic plants.

The order was based on the results of the formula described in (a-c) at CHG site. The percentage decrease in methylation (Mn/Cn - Mt/Ct) x100 across the entire sequence of each TE is shown in Δ symbols. TEs longer than 1kb were selected.

Materials and Methods

Plant materials

A. thaliana strain Columbia-0 (Col-0) was used as “wild-type”. Transgenic lines with *VANC2*, *VANC6*, and *VANC14* were made as follows. Genomic DNA of Col-0 was extracted by the Illustra Nucleon Phytopure genomic DNA extraction kit (GE Healthcare). Each construct including *VANC6*, *VANC2*, and *VANC14* was generated by two rounds of PCR from genomic DNA and cloned into pPZP2H-lac vector⁵⁵ after digestion by *SpeI* and *XhoI*. *A. thaliana* wild-type strain was transformed by the standard floral dip method using each construct⁵⁶. The T-DNAs transferred to *A. thaliana* carried a hygromycin B phosphotransferase gene that was used for the selection marker. Primer sequences for this and other constructions are available upon request.

Whole-genome bisulfite sequencing

Mature rosette leaves were used for genomic DNA extraction. Bisulfite treatments and library preparations were performed by following method like described previously²⁶. Sequencing libraries (insert size: 300–400 bp) were first prepared using TruSeq DNA LT Sample Prep Kit (Illumina) and subjected to bisulphite conversion using MethylCode Bisulfite Conversion Kit (Life Technologies). Bisulphite-treated DNA molecules were amplified by PCR with 10 cycles using KAPA HiFi HotStart Uracil β ReadyMix (2x) (Kapa Biosystems) and purified with Agencourt AMPure XP (Beckman Coulter). Paired-end reads were qualified using Trimmomatic- 0.33 software with following options “ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36”⁵⁷. Qualified reads were mapped using the “bismark” command of bismark (0.14.3) software with following options “-n 1 -l 20”. PCR duplicates were removed from mapped bam files by

“deduplicate_bismark” command⁵⁸. Base resolution of read counts of methylated and unmethylated cytosines were obtained as CX_reports files by “bismark_methylation_extractor” command with following options “–bedGraph –CX –cytosine_report”. Reads from previous study were used for the wild-type data²⁸. Differentially hypomethylated regions at CG sites (CG-hypoDMRs) were defined as previously described⁵⁹. Briefly, in each 100-bp window, DMRs were defined when a difference of methylation level at CG sites was 0.5 or more. Multiple DMRs were merged if they were adjacent to each other or there was only one gap of the 100-bp window. DNA sequences of CG-hypoDMRs in the VANC6 transgenic plant (N= 1347) were used for identifying statistically enriched short motifs by DREME software ver 4.11.0⁶⁰.

Purification of VANC21 and VANC6 proteins (Fig 4d, h)

VANC proteins were synthesized in *E. coli* by following method like described previously²⁸. For VANC6 and VANC21, total RNA was isolated from *dmm1-1* plants by TRIzol (Thermo Fisher Scientific). About 1ug of total RNA was used for cDNA synthesis with PrimeScript RT-PCR Kit (Takara) using random 6-mer primers. VANC6 cDNA was amplified by Q5 High Fidelity DNA Polymerase (NEB) (98°C 5s, 58°C 10s, 72°C 3min; 40 cycles) and A-tailed by ExTaq (Takara). The cDNA was TA-cloned into pGEM T-easy vector (Promega). Cloned full length of VANC cDNAs were amplified with Q5 and primers with AttB1 and AttB2 sequences. The PCR fragment was cloned into pDEST15 vector by one-tube BP and LR Gateway reaction system following manufacturer’s protocol (Thermo Fisher Scientific). After cloning VANC cDNAs into pDEST15 vector, PreScission Protease (GE) recognition sequence was added to the region between GST tag and VANC cDNA by inverse PCR. Each pDEST15 vector containing VANC cDNA was transformed into *E. coli* of BL21-AI strain. Cells were pre-cultured for 8h, 250rpm in 5mL of LB liquid medium

and 4mL of the culture was inoculated in 400mL of LB liquid medium. After 2h of incubation at 37°C, 170rpm, the culture was incubated at 25°C, 170rpm for 30min. Expression of N-terminus-GST-tagged VANC protein was induced for 24h, 170rpm at 25°C by adding up to 0.2% of L-arabinose. Cells were harvested by centrifugation (4500rpm, 15min, 4°C) and lysed by 8mL of BugBuster Master Mix (Merck Millipore). The lysed solution was incubated with 0.15g of glutathione–agarose beads (Sigma-Aldrich, equilibrated with PBS-T buffer) for 1hr at 4°C. The beads were washed 6 times by 50mL of PBS-T, then incubated with 0.1mg of Turbo3C protease (Wako) in 3mL PBS-T for overnight at 4°C. The beads were centrifuged (100rpm, 1min, 4°C). The supernatant (about 500µl) was used as VANC protein. EMSA for binding specificity (Fig 4h) was carried out as described previously²⁸.

Purification of full-length or deletion VANC21 protein (Fig 5~11)

GST-PreScission site-VANC21 cDNA sequence in pDEST15 vector described above was amplified by Q5 High Fidelity DNA Polymerase (98°C 5s, 58°C 10s, 72°C 3min; 40 cycles). The DNA fragment was ligated in pCold I vector (Takara) using NEBuilder (NEB). pCold I vector containing full-length VANC21 cDNA was transformed into *E. coli* of BL21-AI strain. Cells were pre-cultured for 8h, 250rpm in 150mL of LB liquid medium and 120mL of the culture was inoculated in 12L of LB liquid medium. After 2h of incubation at 37°C, 170rpm, the culture was incubated at 16°C, 170rpm for 30min. Expression of N-terminus-GST-tagged VANC protein was induced for 24 h, 170rpm at 16 °C by adding up to 0.5mM of IPTG. Cells were harvested by centrifugation (3000rpm, 15min, 4°C) and lysed by 300mL of BugBuster Master Mix. The lysed solution was centrifuged (4°C 4500g, 30min). The supernatant was incubated with 2.5g of glutathione–agarose beads (Sigma-Aldrich, equilibrated with PBS buffer) for 1hr at 4°C with NaCl of final concentration 1M. The beads were washed 6 times

by 50mL of PBS-T including 1M NaCl, then incubated with 1mg of Turbo3C protease (Wako) in 90mL PBS-T for overnight at 4°C. The beads were centrifuged (100rpm, 1min, 4°C). The supernatant (about 20mL) was further purified using ion-exchange column chromatography.

pET16-8xHis-(full-length or deletion) VANC21 was kindly provided by Osamu Nureki Lab. pET16 vector containing VANC21 cDNA was transformed into *E.coli* of Rosetta2 (DE3) strain. Cells were pre-cultured for 8h, 250rpm in 150mL of LB liquid medium and 120mL of the culture was inoculated in 12L of LB liquid medium. After 2h of incubation at 37°C, 170rpm, the culture was incubated at 25°C, 170rpm for 30min. Expression of N-terminus-GST-tagged VANC protein was induced for 17 h, 170rpm at 25 °C by adding up to 1mM of IPTG. Cells were harvested by centrifugation (3000rpm, 15min, 4°C) and lysed by 300mL of lysis buffer (50 mM HEPES-NaOH, 1 M NaCl, 0.05 % tween20, 1 mM 2-me, 0.1 mM PMSF, 20 mM Imidazol, pH 8.0). The lysed solution was sonicated by Branson Sonifier 250D with the conditions of DUTY=30 %, OUTPUT=5, 10 mins. The sonicated solution was centrifuged (4500g, 30min, 4°C). The supernatant was incubated with 10mL of Ni-NTA Agarose 50% slurry (QIAGEN, equilibrated with lysis buffer) for 5min at 4°C. The beads were washed by 200mL of wash buffer (50 mM HEPES-NaOH, 1 M NaCl, 0.05 % tween20, 1 mM 2-me, pH 8.0) using open column. VANC21 protein was eluted by 24mL elution buffer (50 mM HEPES-NaOH, 500 mM NaCl, 1 mM 2-me, 5 % Glycerol, 300 mM Imidazol). Eluted fraction was further purified using ion-exchange column chromatography.

For ion-exchange column chromatography, HiTrap Q XL and HiTrap SP HP (GE Healthcare) attached to an ÄKTA Pure liquid chromatography system (GE Healthcare) was used for full-length VANC21 protein and deletion VANC21 protein, respectively. Eluted fraction including VANC21 protein was

further purified by gel filtration chromatography using SuperDex increase 10/300 (GE Healthcare).

EMSA of full-length or deletion VANC21 for kinetics (Fig 6)

EMSA of full-length or deletion VANC21 for kinetics was carried out as follows²⁸. The protein concentration was quantified using Bradford method. About 10pmol of double-stranded DNAs (dsDNAs) was radiolabeled with 1 U of T4PNK (Takara) by incubation with 0.5 MBq of [γ -³²P] ATP for 1 h at 37 °C in 10 μ l of reaction solution. The radiolabeled dsDNAs were filled up to 100ul, and then purified using MicroSpin G-25 Columns (GE). Any concentration of VANC21 protein was incubated at 4 °C for 30 min in 10 μ l of reaction solutions with 0.1 pmol of radiolabeled dsDNA in buffer containing 15 mM Tris-HCl (pH 7.5), 300 mM NaCl, 3 mM MgCl₂, 0.04% Triton X, 4% glycerol, 0.5 mM DTT. The reaction solutions were separated on 4% non-denaturing polyacrylamide gel with 1 \times TBE buffer by 50mA, for 45 min at 4°C. Radioactivity signals were detected using a FLA-9000 (GE Healthcare).

Fitting Sigmoidal Curve to Hill plot

To fit the sigmoidal curve of deletion VANC21 kinetics to Hill plot, a nonlinear least-squares method was performed using the statistical software R. For the parameters (Kd, n), the initial value of Kd was set to 140 nM (i.e. $\log_{10}Kd = 2.146$), and that of n was set to 1 for the WT probe. For the mutation-type probe, the initial value of Kd was input as 265 nM (i.e. $\log_{10}Kd = 2.423$), and that of n was set to 1. The Hill plot equation $[P]^n / (Kd^n + [P]^n)$ was calculated as $V_{min} = 0.00$ and $V_{max} = 1.00$. For nonlinear least-squares method, the nls function of the statistical software R was used.

HS-AFM observations on AP-mica surface (Fig 7-11)

HS-AFM Nano Live vision (RIBM) was used. For HS-AFM observations of VANC21-DNA complex, a mica surface was treated for 3 min with 1/8000 diluted 3-aminopropyltriethoxysilane (Sigma-Aldrich). 20nM of full-length VANC21 or 150nM of deletion VANC21 and 0.5ng/uL of DNA was pre-assembled in condition of 10mM HEPES, 300mM KCl, 3mM MgCl₂, pH 8.0. All HS-AFM observations were performed in condition with 10mM HEPES, 100 mM KCl, 3mM MgCl₂, pH 8.0. All HS-AFM experiments were performed at room temperature. For Fig 10, the laboratory-built high-speed AFM (Kanazawa University) was used in the tapping mode as described previously^{34, 61}.

Quantification of the height of VANC21 protein on AP-mica surface

The height of VANC21 protein was measured by Gwyddion software ver 2.55⁶². The background of HS-AFM images was first subtracted by “Level data by mean plane subtraction” and “Align rows using various methods”. The height of VANC21 was measured by watershed method with marking grains.

Quantification of binding points of VANC21 protein on AP-mica surface

The distance from both ends of the DNA to VANC21 protein was measured in HS-AFM videos. To eliminate arbitrary manipulations, the DNA length was automatically quantified using the Image J plugin “simple neurite tracer”.

Reference

1. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285 (2007).
2. Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* 33, 479–532 (1999).
3. Muñoz-López M, García-Pérez JL. DNA transposons: nature and applications in genomics. *Curr Genomics.* 11(2):115-28 (2010).
4. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767 (2012).
5. Pace JK 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 4, 422-32 (2007).
6. Vicient CM. Transcriptional activity of transposable elements in maize. *BMC Genomics.* 11, 601 (2010).
7. Quadrana, L. et al. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* 5, 1-25 (2016).
8. Quadrana, L. et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nature Communications* 10, E3421 (2019).

9. Ito, H. et al. A Stress-Activated Transposon in *Arabidopsis* Induces Transgenerational Abscisic Acid Insensitivity. *Sci Rep.* 15, 6:23181 (2016).
10. Ono, R. et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat Genet* 38, 101-106 (2006).
11. Suzuki, S. et al. Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PloS Genetics* 3, e55 (2007).
12. Chandler VL, Walbot V DNA modification of a maize transposable element correlates with loss of activity. *Proc Natl Acad Sci USA* 83: 1767–1771 (1986).
13. Hiroshika H, Okamoto H, Kakutani T. Silencing of Retrotransposons in *Arabidopsis* and Reactivation by the *ddm1* mutation. *The Plant Cell* 12: 357-368 (2000).
14. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408 (2014).
15. Kim, M. Y. & Zilberman, D. DNA methylation as a system of plant genomic immunity. *Trends Plant. Sci.* 19, 320–326 (2014).
16. Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* 152, 352–364 (2013).

17. Becker, C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*. 480(7376):245-9 (2011)
18. Cokus, S. J. et al. Shotgun bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219 (2008).
19. Zemach, A. et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205 (2013).
20. Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. *Arabidopsis thaliana* DNA methylation mutants. *Science* 260, 1926–1928 (1993).
21. Miura, A. et al. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* 411, 212–214 (2001).
22. Singer, T., Yordan, C. & Martienssen, R. A. Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev.* 15, 591–602 (2001).
23. Tsukahara, S. et al. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* 461, 423–426 (2009).
24. Cui H, Fedoroff NV. Inducible DNA demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein. *Plant Cell* 14: 2883–2899 (2002).

25. Nosaka M, Itoh J, Nagato Y, Ono A, Ishiwata A, Sato Y. Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS Genet.* 9,e1002953.(2012)
26. Fu, Y. et al. Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J.* 32, 2407–2417 (2013).
27. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815 (2000).
28. Hosaka A, Saito R, Takashima K, Sasaki T, Fu Y, Kawabe A, et al. Evolution of sequence-specific anti-silencing systems in *Arabidopsis*. *Nat Commun.* 8:1–10. (2017)
29. Ando, T., Uchihashi, T. & Scheuring, S. Filming biomolecular processes by high-speed atomic force microscopy. *Chem. Rev.* 114, 3120–3188 (2014).
30. Shibata, M., Yamashita, H., Uchihashi, T., Kandori, H. & Ando, T. High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin. *Nat. Nanotech.* 5, 208–212 (2010).
31. Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature* 468,72–76 (2010).
32. Igarashi, K. et al. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* 333, 1279–1282 (2011).

33. Uchihashi, T., Iino, R., Ando, T. & Noji, H. High-speed atomic force microscopy reveals rotary catalysis of rotorless F1-ATPase. *Science* 333, 755–758 (2011).
34. Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, Uchihashi T, Nureki O. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Commun.* 8(1):1430 (2017).
35. Cho C, Jang J, Kang Y, Watanabe H, Uchihashi T, Kim SJ, Kato K, Lee JY, Song JJ. Structural basis of nucleosome assembly by the Abo1 AAA+ ATPase histone chaperone. *Nat Commun.* 10(1):5764. (2019)
36. Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, Maire P. The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol.* 6:633-48 (2008).
37. Eliaš J, Clairambault J. Reaction-diffusion systems for spatio-temporal intracellular protein networks: A beginner's guide with two examples. *Comput Struct Biotechnol J.* 10(16):12-22 (2014).
38. Asmari M, Ratih R, Alhazmi HA, El Deeb S. Thermophoresis for characterizing biomolecular interaction. *Methods.* 146:107-119 (2018).
39. Rigal M, Becker C, Pélissier T, Pogorelnik R, Devos J, Ikeda Y, Weigel D, Mathieu O. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc Natl Acad Sci U S A.* 113(14): E2083-92. (2016)

40. Drinnenberg, I. A., Henikoff, S. & Malik, H. S. Evolutionary turnover of kinetochore proteins: a ship of theseus? *Trends Cell Biol.* 26, 498–510 (2015).
41. Henikoff, S. & Malik, H. S. Centromeres: selfish drivers. *Nature* 417, 227 (2002).
42. McClintock B. Chromosome organization and genic expression. *Cold Spring Harbor Symp Quant Biol* 16: 13–47 (1951).
43. McClintock B. The suppressor-mutator system of control of gene action in maize. *Carnegie Institution of Washington Year Book Vol.* 57: 415–429 (1958).
44. Yu, Z., Wright, S. I. & Bureau, T. E. Mutator-like elements in *Arabidopsis thaliana*: structure, diversity and evolution. *Genetics* 156, 2019–2031 (2000).
45. Peccoud, J., Loiseau, V., Cordaux, R. & Gilbert, C. Massive horizontal transfer of transposable elements in insects. *Proc. Natl Acad. Sci. USA* 114, 4721–4726 (2017).
46. Zhao, J. H., Hua, C. L., Fang, Y. Y. & Guo, H. S. The dual edge of RNA silencing suppressors in the virus-host interactions. *Curr. Opin. Virol.* 17,39–44 (2016).
47. Bivalkar-Mehla, S. et al. Viral RNA silencing suppressors (RSS): novel strategy of viruses to ablate the host RNA interference (RNAi) defense system. *Virus Res.* 155,1–9 (2011).

48. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* 547(7662):236-240 (2017).
49. Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, Sharp PA, Chakraborty AK, Young RA. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 361(6400): eaar3958 (2018).
50. Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, Gerlich DW, Redding S, Rosen MK. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell*. 179(2):470-484.e21 (2019).
51. Ide S, Imai R, Ochi H, Maeshima K. Transcriptional suppression of ribosomal DNA with phase separation. *Sci Adv*.6(42):eabb5953 (2020).
52. Wang Y, Zolotarev N, Yang CY, Rambold A, Mittler G, Grosschedl R. A Prion-like Domain in Transcription Factor EBF1 Promotes Phase Separation and Enables B Cell Programming of Progenitor Chromatin. *Immunity*. S1074-7613(20)30450-7 (2020).
53. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*. 47(W1):W402-W407 (2019).
54. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 31(6):857-63. (2015).

55. Fuse T, Sasaki T, Yano M. Ti-plasmid vectors useful for functional analysis of rice genes. *Plant Biotechnol* 18: 219–222 (2001).
56. Clough SJ, Bent AF. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735–743 (1998).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).
58. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572 (2011).
59. Ito, T. et al. Genome-wide negative feedback drives transgenerational DNA methylation dynamics in *Arabidopsis*. *PLoS Genet.* 11, e1005154 (2015).
60. Timothy L. Bailey, "DREME: Motif discovery in transcription factor ChIP-seq data", *Bioinformatics*, 27(12):1653-1659 (2011).
61. M. Uchihashi, T., Kodera, N. & Ando, T. Guide to video recording of structure dynamics and dynamic processes of proteins by high-speed atomic force microscopy. *Nat. Protoc.* 7, 1193–1206 (2012).
62. Nečas, D., Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *centr.eur.j.phys.* 10, 181–188 (2012).

Acknowledgement

I thank Mr. Kazuya Takashima and Ms. Akiko Terui for technical assistance; Ms. Sachiko Sakamoto and Ms. Shizuko Endo for technical advice about protein purification; Dr. Yumiko Kurokawa, Dr. Ayumi Sumino, and Dr. Mikihiro Shibata for technical advice about HS-AFM observation; Mr. Yusaku Tanaka and Dr. Taku Sasaki for sharing vector materials; Dr. Kazuhiro Maeshima, Dr. Hiroyuki Araki, Dr. Ken-ichi Nonomura, Dr. Naruya Saitou, Dr. Yasuto Murayama, Dr. Yoshihisa Oda, Dr. Yasushi Hiromi for valuable discussion as SOKENDAI progress committee. I am grateful to All Kakutani-Lab members for valuable discussion. Finally, I am most grateful to Dr. Soichi Inagaki, Dr. Yoshiaki Tarutani, Dr. Tetsuji Kakutani for various help. Computations were partially performed on the NIG supercomputer at NIG, Japan. This study was supported by grants from JSPS research fellow DC2. For learning HS-AFM observation technique, I participated in the 9th Bio-SPM Summer School at Kanazawa University (2019 Aug. 17-22).