

MINIREVIEW

Yusuke Nakamura · Kumiko Koyama
Mieko Matsushima

VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators

Received: April 11, 1998 / Accepted: May 2, 1998

Abstract VNTR (variable number of tandem repeat) markers, also called single-copy minisatellites, were originally isolated from human DNA as highly informative restriction fragment length polymorphisms for mapping purposes. Evidence has lately emerged that some VNTR sequences play significant roles in the regulation of transcription, and that some may also influence the translational efficiency or stability of mRNA, or modify the activity of proteins by altering their structure. Some apparent associations of VNTR sequences with personality traits or with susceptibility to diseases have strengthened the likelihood that these tandemly-repeated genomic elements are of physiological and biological importance. In this review, we summarize recent progress in efforts to clarify mechanisms involving VNTR sequences.

Key words VNTR (variable number of tandem repeat) · Transcriptional regulator · Translational regulator · Personality traits · Disease susceptibility

Introduction

VNTR (variable number of tandem repeat) markers, also called single-copy minisatellites, were originally developed for linkage-mapping purposes, as a high degree of polymorphism is often present among individuals with respect to the number of tandemly repeated units of a nucleotide sequence at a particular locus (Nakamura et al. 1987; 1988). Their highly polymorphic nature makes VNTRs very informative as a class of markers for linkage studies to map disease loci in families, and also makes them helpful in forensic applications (Odelberg et al. 1989; Kasai et al. 1990).

Y. Nakamura (✉) · K. Koyama · M. Matsushima
Laboratory of Molecular Medicine, Human Genome Center,
Institute of Medical Science, University of Tokyo, 4-6-1,
Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
Tel. +81-3-5449-5372; Fax +81-3-5449-5433
e-mail: yusuke@ims.u-tokyo.ac.jp

Recent reports have indicated that some VNTR sequences may function as transcriptional or translational regulators, or that they can modify the function of a protein when the tandemly repeated region lies within the coding region of the gene. In this review, we would like to introduce newly discovered aspects of VNTR elements within the human genome.

VNTRs as transcriptional regulators

Some transcriptional factors can bind VNTR sequences. For example, members of the rel/NF- κ B family of transcriptional regulators bind VNTR sequences present in the 3' flanking region of the *HRAS* gene; this process has been shown to activate transcription of a reporter gene in some cell lines, including cells derived from the FJ bladder carcinoma (Trepicchio and Krontiris 1992). Part of the consensus sequence of the 28 base pairs (bp) repeating unit (AGGGGACGCCACACTGCCCTCTCTCC) of the *HRAS* VNTR is highly homologous to the target-binding sequence (GGGGAC(G/T)(C/T)(C/T)CC) of rel/NF- κ B proteins. This VNTR varies in size from 1000 to 2500 bp (35–90 repeats); several of its alleles are common in the population but many rare alleles exist also. The presence of one or two of the rare alleles in a single individual appears to be associated with a higher than normal risk of cancers in multiple tissues, such as the colon, breast, and urinary bladder (Krontiris et al. 1993). Although several groups of investigators have reported similar results, others have failed to confirm this association. Hence, further genetic epidemiological studies involving a larger set of clinical cases will be required to clarify whether this VNTR in fact influences the occurrence of cancer. However, Phelan et al. (1996) suggested that the status of the *HRAS* VNTR may influence the penetrance of ovarian cancer in individuals who carry a mutation in the breast cancer susceptibility gene *BRCA1*, on the basis of evidence that the risk for ovarian cancer was more than twice as high in mutated-*BRCA1* carriers with one or two rare alleles of the *HRAS* VNTR than in women

with only common alleles ($P = 0.015$). However, they saw no effect of allelic status on the occurrence of breast cancer.

The insulin (*INS*) gene contains a VNTR, in the form of a 14-bp repeating unit, in its 5' control region. *INS* VNTRs are divided into three classes according to the number of repeating units present. Lucassen et al. (1993) first indicated an association of the length of this VNTR with susceptibility to insulin-dependent diabetes mellitus (IDDM). Subsequent analyses by two groups (Pugliese et al. 1997; Vafiadis et al. 1997) disclosed that the *INS* VNTR influences the expression of insulin transcript in fetal thymus: *INS* alleles with a class III VNTR (140–210 repeats) express the insulin transcript more abundantly than those with class I VNTRs (26–63 repeats). The higher level of insulin expression in thymus is considered to promote negative selection of insulin-specific T-lymphocytes, and thereby play a critical role in the pathogenesis of IDDM. Furthermore, Bennett et al. (1997) found that the insulin VNTR allele-specific effect in IDDM depended on the identity of the untranslated paternal allele.

A third VNTR that may influence transcription is the one found within the DH-JH interval of the human immunoglobulin heavy chain (*IGH*). The 50-bp consensus sequence of the *IGH* VNTR contains a sequence similar to the motif for binding myc/HLH transcription factor; in particular, 15 of the 17 nucleotides contained in this consensus sequence, GTGGGCCACGTGAGCGG, match a myc/HLH-binding site of the adenovirus major late promoter (MLP), GTAGGCCACGTGACCGG. Trepicchio and Krontiris (1993) reported that although the *IGH* VNTR has no enhancer activity, it is able to significantly suppress transcriptional activity of the MLP. This VNTR also suppressed MLP activation by the murine heavy chain gene enhancer, E μ , in murine myeloma cells.

Although no clear effect on transcription was demonstrated, Ogilvie et al. (1996) reported an association between a VNTR polymorphism within the second intron of the serotonin transporter gene and susceptibility to major depression. They suggested that the amount of serotonin transporter, the primary target for selective antidepressant drugs, might be related to this affective disorder. Among the three known alleles (9, 10, or 12 repeats of a 17-bp consensus unit, GGCTGYGACCYRGRRTG), the allele containing 9 repeats is significantly more common in patients with affective disorder than in control individuals. However, this association may simply reflect linkage disequilibrium between the 9-repeat allele and some other genetic change that influences the biological function of the serotonin transporter. Hence, it is certainly too early to conclude that this VNTR functions as a transcriptional regulator and that the shorter allele of the transporter gene is not transcribed to the same degree as the other two alleles. In addition to the intronic VNTR sequence in the serotonin transporter gene, Heils et al. (1996) found an insertion/deletion polymorphism within a series of 17 repeated sequences, each consisting of 20–23 nucleotides (consensus sequence, CCCCCCTGCA CCCCCCCCAGCAT), in the 5' promoter region. Lesch et al. (1996) subsequently showed that an allele carrying a 44-

bp deletion involving repeating units 6–8 was significantly more common in individuals with anxiety-related personality traits such as neuroticism, tension, and harm-avoidance than in the general population. They also revealed that the region containing the sequence lacking these 44 bp was less active in promoting transcription of the reporter gene than the sequence without the 44-bp deletion.

Another example of correlation between VNTR status and genetic disease occurs in progressive myoclonus epilepsy of the Unverricht-Lundborg type (EPM1). A candidate-gene approach to positional cloning identified alteration of the cystatin B (*CSTB*) gene as the cause of EPM1. In addition to missense point mutations in the coding region, Lalioti et al. (1997) and Lafreniere et al. (1997) found expansion of a 12-nucleotide repeat located about 70 bp upstream of the transcription start site in the majority of EPM1 alleles, and also revealed that expansion of this VNTR reduced levels of *CSTB* mRNA in blood. Meloni et al. (1998) provided further evidence that minisatellites can play significant roles in the regulation of transcription when they reported that a polymorphic tetranucleotide repeat in an intron of the tyrosine hydroxylase gene functions as an enhancer element. They suggested that proteins of the Fos-Jun family form a complex with this repeating sequence.

VNTRs as translational regulators

In our own investigations, we found that the *SEC14L* gene, showing a partial homology to yeast SEC14 and to retinal-binding protein (RALBP) of the Japanese flying squid, carries a VNTR in its 3' untranslated region, and we characterized this VNTR in the Japanese population. Most Japanese alleles (more than 95%) carried 40–60 repeats of the 13-bp consensus sequence, GGTAGGG (C/T)T(A/C/G)GTA, and fewer than 5% carried an allele with only 14 repeats (Chinen et al. 1996). To investigate whether this VNTR affects RNA stability or translation, we cloned DNA sequences corresponding to short or long VNTR alleles into the 3' non-coding region of the luciferase expression vector. As shown in Fig. 1, luciferase activity was suppressed to 60%–70% when a long allele was cloned into the expression vector, but no significant reduction occurred with the shorter allele. As the same amount of mRNA was present among cell lines transfected with plasmids containing the larger VNTR allele, the shorter VNTR allele, or vector alone, we assumed that the reduction of enzyme activity reflected reduction of translation, and that the VNTR itself had influenced translational efficiency.

VNTRs within coding sequences

Some genes are known to contain a VNTR within their coding sequences. The human epithelial mucin gene contains an expressed VNTR, which consists of a variable number of repeats of a 60-bp unit (Lancaster et al. 1990).

However, no evidence has been documented as to whether a difference in the number of repeats of the 20-amino-acid sequence may influence physiological function of the gene product, or whether some alleles may play pathogenic roles.

The D4 dopamine receptor (D4DR) contains a polymorphic 16-amino-acid (a.a.) repeat, PAPGLP(Q/R/P)(D/G)PCGPDCAP, in the putative third cytoplasmic loop (Van Tol et al. 1992; Lichter et al. 1993). As this gene is expressed at a high level in limbic areas of the brain, its product may well be associated with cognitive and emo-

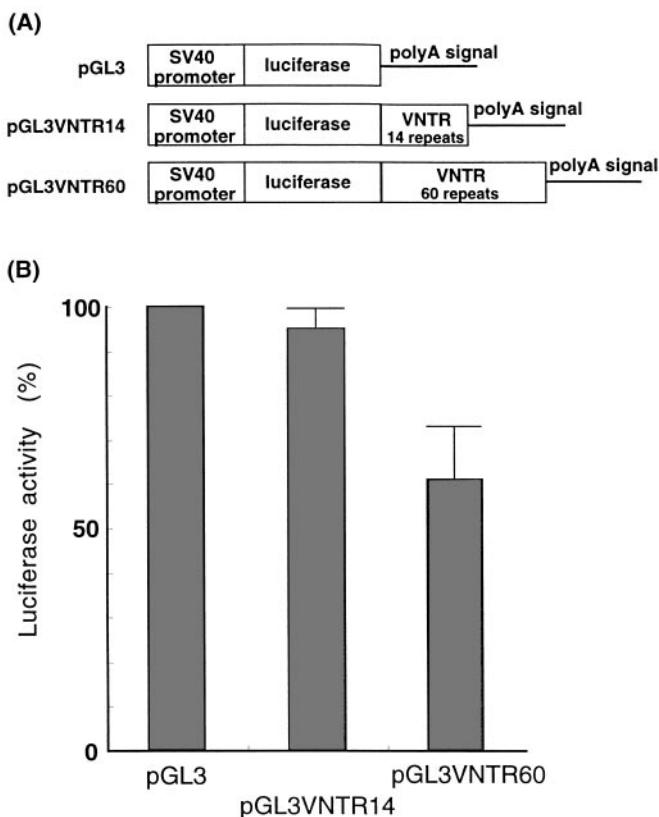


Fig. 1 A Construction of expression vectors including 14 (pGL3VNTR14) or 60 (pGL3VNTR60) repeats of a GGTAGGG(C/T)T(A/C/G)GTA sequence. The VNTR sequences were cloned separately into the 3' non-translated region of the luciferase reporter gene. **B** Relative luciferase activities in cells transfected with three expression vectors (pGL3, pGL3VNTR14, pGL3VNTR60). VNTR, Variable number of tandem repeat

Table 1 Repetitive peptide sequences in mammals

Gene or protein	Repeating unit	Consensus sequence
Human Acrosomal protein sp-10 precursor	5 a.a.	SGE(Q/H)(A/P)
Human Cholesterol esterase	11 a.a.	PVPPTGDSGAP
Human Neurofilament triplet H protein	14 a.a.	SPVK(E/A)FAKSPEKAK
Human D4 Dopamine receptor (D4DR)	16 a.a.	PAPGLP(Q/R/P)(D/G)PCGPDCAP
Human Epithelial mucin gene	20 a.a.	VTSAPD(T/N)RP(A/P)(P/L)GSTAPPA(H/Q)(D/G)
Human Intestinal mucin (MUC2)	23 a.a.	VTPTPTGTQTPTTTPITTTTT
Human Coagulation factor V	9 a.a.	LS(P/L)(E/A/D)(L/F)(S/G)(Q/H)(T/M)(T/N/P)
Pig Cathelin-like antibacterial peptide	10 a.a.	RFPPNNFPGP
Mouse Spliceosome associated protein 62 (SAP62)	7 a.a.	VHPP(A/T)(P/S)(G/V)

a.a., Amino acid

tional behaviors. Differences in ligand-binding affinity have been observed between D4DR molecules containing seven or four repeats of the 16-a.a. sequence (Van Tol et al. 1992; Asghari et al. 1994). On the basis of these observations, Benjamin et al. (1996) and Ebstein et al. (1996) examined the relationship between *D4DR* VNTRs and various personality traits; they found an interesting association between the length of the VNTR in exon 3 (corresponding to the putative third cytoplasmic loop of D4DR) and novelty-seeking behavior. Since differences in the number of repeated sequences in the coding region of a gene possibly alter protein structures, it is likely that transcribed VNTRs bear a tight correlation to the physiological activity of the gene product. In addition, the finding that mice lacking D4DR are supersensitive to ethanol, cocaine, and methamphetamine (Rubinstein et al. 1997) implies that the *D4DR* VNTR may be related to normally coordinated and drug-stimulated motor behaviors. Including the two mentioned above (i.e., human epithelial mucin and D4DR), nearly a dozen proteins of humans or other vertebrates are known to contain repeated peptide sequences (Table 1), although not all of them are known to be polymorphic.

The information reviewed here strongly implies that some, if not all, fractions of untranscribed VNTR sequences may function as transcriptional regulators. Moreover, transcribed VNTRs affect the stability and/or translational efficiency of mRNA, and if they lie within coding regions they are likely to affect the activity of the gene product. Hence, the isolation of more human genes containing VNTR within exons, introns, or flanking regions will be important for an increased understanding of complex human traits, including personality and susceptibility to disease.

References

- Asghari V, Schoots O, van Kats S, Ohara K, Jovanovic V, Guan HC, Bunzow JR, Petronis A, Van Tol HH (1994) Dopamine D4 receptor repeat: Analysis of different native and mutant forms of the human and rat genes. *Mol Pharmacol* 46: 364-373
- Benjamin J, Li L, Patterson C, Greenberg BD, Murphy DL, Hamer DH (1996) Population and familial association between the D4 dopamine receptor gene and measures of novel seeking. *Nature Genet* 12: 81-84
- Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undelien DE, Cucca F, Nistico L, Buzzetti R, the IMDIAB group, Bosi E, Pociot F, Nerup J, Cambon-Thomsen A, Pugliese A, Shield JPH, McKinney

- PA, Bain SC, Polychronakos C, Todd JA (1997) Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untranslated paternal allele. *Nature Genet* 17: 350–353
- Chinen K, Takahashi E, Nakamura Y (1996) Isolation and mapping of a human gene (SEC14L), partially homologous to yeast SEC14, that contains a variable number of tandem repeats (VNTR) site in its 3' untranslated region. *Cytogenet Cell Genet* 73: 218–223
- Ebstein RP, Novik O, Umansky R, Priel B, Osher Y, Blaine D, Bennett ER, Nemanov L, Katz M, Belmaker RH (1996) Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of novel seeking. *Nature Genet* 12: 78–80
- Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, Lesch KP (1996) Allelic variation of human serotonin transporter gene expression. *J Neurochem* 66: 2621–2624
- Kasai K, Nakamura Y, White R (1990) Amplification of a VNTR locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J Forensic Sci* 35: 1196–1200
- Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N Engl J Med* 329: 517–523
- Lafreniere RG, Rochefort DL, Chretien N, Rommens JM, Cochius JI, Kalvainen R, Nousiainen U, Patry G, Farrell K, Soderfeldt B, Federico A, Hale BR, Cossio OH, Sorensen T, Pouliot MA, Kmiec T, Uldall P, Janszky J, Pranzatelli MR, Andermann F, Andermann E, Rouleau GA (1997) Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nature Genet* 15: 298–302
- Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 386: 847–850
- Lancaster CA, Peat N, Duhig T, Wilson D, Taylor-Papadimitriou J, Gendler SJ (1990) Structure and expression of the human polymorphic epithelial mucin gene: an expressed VNTR unit. *Biochem Biophys Res Com* 173: 1019–1029
- Lesch K-P, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S, Benjamin J, Muller CR, Hamer DH, Murphy DL (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274: 1527–1531
- Lichter JB, Barr CL, Kennedy JL, Van Tol HHM, Kidd KK, Livak KJ (1993) A hypervariable segment in the human dopamine receptor D4 (D4DR) gene. *Hum Mol Genet* 2: 767–773
- Lucassen AM, Julier C, Beressi J-P, Boitard C, Froguel P, Lathrop M, Bell JI (1993) Susceptibility to insulin dependent diabetes mellitus maps to a 4.1kb segment of DNA spanning the insulin gene and associated VNTR. *Nature Genet* 4: 305–310
- Meloni R, Albanese V, Ravassard P, Treihou F, Mallet J (1998) A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element in vitro. *Hum Molec Genet* 7: 423–428
- Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable Number of Tandem Repeat (VNTR) markers for human gene mapping. *Science* 235: 1616–1622
- Nakamura Y, Carlson M, Krapcho K, Kanamori M, White R (1988) A new approach for isolating VNTR markers. *Am J Hum Genet* 43: 854–859
- Odelberg SJ, Plaetke R, Eldridge JR, Ballard L, O'Connell P, Nakamura Y, Leppert M, Lalouel J-M, White R (1989) Characterization of VNTR loci by high-resolution agarose gel electrophoresis: Implications for parentage testing and forensic individualization. *Genomics* 5: 915–924
- Ogilvie AD, Battersby S, Bubb VJ, Fink G, Harmar AJ, Goodwin GM, Smith CAD (1996) Polymorphism in serotonin transporter gene associated with susceptibility to major depression. *Lancet* 347: 731–733
- Phelan CM, Rebbeck TR, Weber BL, Devilee P, Rutledge MH, Lynch HT, Lenoir GM, Stratton MR, Easton DF, Ponder BAJ, Cannon-Albright L, Larsson C, Goldgar DE, Narod SA (1996) Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable number of tandem repeat (VNTR) locus. *Nature Genet* 12: 309–314
- Pugliese A, Zeller M, Fernandez A, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD (1997) The insulin gene is transcribed in the human thymus and transcription levels correlate with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nature Genet* 15: 293–296
- Rubinstein M, Phillips TJ, Bunzow JR, Falzone TL, Dziewczapolski G, Zhang G, Fang Y, Larson JL, McDougall JA, Chester JA, Saez C, Pugsley TA, Gershmanik O, Low MJ, Grandy DK (1997) Mice lacking dopamine D4 receptors are supersensitive to ethanol, cocaine, and methamphetamine. *Cell* 90: 991–1001
- Trepicchio WL, Krontiris TG (1992) Members of the rel/NF- κ B family of transcriptional regulatory proteins bind the HRAS minisatellite DNA sequence. *Nucl Acids Res* 20: 2427–2434
- Trepicchio WL, Krontiris TG (1993) IGH minisatellite suppression of USF-binding-site- and E μ -mediated transcriptional activation of the adenovirus major late promoter. *Nucl Acids Res* 21: 977–985
- Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C (1997) Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM locus. *Nature Genet* 15: 289–296
- Van Tol HHM, Wu CM, Guan HC, Ohara K, Bunzow JR, Civelli O, Kennedy J, Seeman P, Niznik HB, Jovanovic V (1992) Multiple dopamine D4 receptor variants in the human population. *Nature* 358: 149–152