J Hum Genet (2002) 47:330–332 

SHORT COMMUNICATION

**Haruhito Harada · Akinori Kimura · Koichi Fukino**
**Seikoh Yasunaga · Hirofumi Nishi · Mitsuru Emi**

# Genomic structure and eight novel exonic polymorphisms of the human N-cadherin gene

Received: January 23, 2002 / Accepted: March 12, 2002

**Abstract** Analysis of the detailed genomic structure of human N-cadherin revealed that the 16-exon gene is more than 72 kb in length and that it consists of a mosaic of exons. Five repeated cadherin domains, a transmembrane domain, and a cytoplasmic domain are encoded by exons 4 to 13, 13 and 14, and 14 to 16, respectively. A search for molecular variants in the entire coding region in 96 Japanese individuals resulted in the identification of eight sequence polymorphisms including three CCT- or GCC-type trinucleotide repeat polymorphisms adjacent to the initiation codon and five other novel single-nucleoticle polymorphisms (SNPs) in the coding region. Three of the five SNPs accompanied an amino acid substitution: Ala118Thr, Ala826Thr, and Asn845Ser. Knowlege of the fine gene structure and eight novel polymorphisms will be useful for the genetic study of the role of N-cadherin in diseases involving cell adhesion in the brain and in cardiomyocytes.

**Key words** N-cadherin · Brain tumor · Hypertrophic cardiomyopathy · Single-nucleotide polymorphism (SNP) · Trinucleotide repeats

H. Harada · K. Fukino · M. Emi (✉)
Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, 1-396 Kosugi-cho, Nakahara-ku, Kawasaki 211-8533, Japan
Tel. +81-44-733-5230; Fax +81-44-733-5192
e-mail: memi@nms.ac.jp

A. Kimura
Department of Molecular Pathogenesis, Division of Adult Diseases, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

S. Yasunaga
Tsukumi Central Hospital, Tsukumi, Japan

H. Nishi
Cardiovascular Center, Kurume University, Kurume, Japan

## Introduction

N-cadherin is expressed predominantly in brain and cardiac muscles and functions to mediate cell adhesion events required for heart formation in vertebrate development (Garcia-Castro et al. 2000). The gene was mapped to 18q1.2 and shown to consist of 16 exons on a yeast artificial chromosome (YAC) (Wallis et al. 1994).

Decreased expression of the N-cadherin protein in a brain tumor was associated with tumor invasion and dissemination into cerebrospinal fluid during clinical recurrence (Asano et al. 1997, 2000). It is notable that N-cadherin is also a mediator of neurite outgrowth in astrocytes (Blaschuk et al. 1990). These findings suggested that N-cadherin might be another tumor suppressor gene, especially of brain tumors. It is also a candidate gene for familial hypertrophic cardiomyopathy, which is characterized by cardiac myofibril disarray and derangement in cell–cell contacts. Given the potential role of N-cadherin in the pathophysiology of the brain and cardiac myocytes, molecular variants in the human N-cadherin gene would likely result in derangement in diseases within these tissues. To establish a genetic basis for investigation of these hypotheses, we determined the structure of the human N-cadherin gene and found eight novel polymorphisms.

## Subjects and methods

Phage clones containing the N-cadherin (NCAD) gene were completely digested by *Eco*RI and subcloned, and the plasmid was hybridized with N-cadherin cDNA oligonucleotides (NM_001792) and sequenced. Sequences around the first and second exons were obtained from the draft sequence (GenBank; Z27420, Z27413, Z27421, and NT_011065). Blood samples from 96 Japanese volunteers and 96 unrelated patients with hypertrophic cardiomyopathy and 43 brain tumor specimens obtained during operations were used in the study. Written informed consent was

**Table 1.** Primer sequence for the PCR-SSCP analysis

| Exon | PCR length | Forward primer (5′ to 3′) | Reverse primer (5′ to 3′) |
|---|---|---|---|
| 1 | 197 | TGGAAACTGCCTGGAGCCGTT | GGACCGCCGCGTACCTGAAG |
| 2 | 146 | TTTTTTTTCTTGTGTTGCAGG | TTTTGAGATAGTACCATTGAG |
| 3 | 379 | CCTAAGCAGGATATAGGTTTAA | TAGTTCAAACTGTATAAAGCTGAT |
| 4 | 239 | TGTGTTTCCTAGTACTCAGGA | ATACATTTGTCTTGTGGTATGA |
| 5 | 267 | CCATCTCTGATAATGAGCTCAT | GAGGCAATGGTAAACTCTGC |
| 6 | 250 | TGTATTTATCTTCTCACAGACAT | ATACTTTCCTCAAGTCATCTTC |
| 7 | 277 | AGGGTAACCATTCATTTTCCAT | ATGTGATATGATATTGTGCACC |
| 8 | 248 | AACTGTTATGTCTTCTTGAGGA | AGTAACGAACTACAGACCCAA |
| 9 | 306 | CAAGTTGTCCTTCAGTGGTC | GGACATATATTGGTCCTTGCT |
| 10 | 345 | TTTCCTGCAAAAATATGGCTTC | GCACAGCATAGAACATAAGTAA |
| 11 | 272 | AGCCAGTTGCATTTGGAACTA | CTACTGTCTTTCATCAACATAC |
| 12 | 332 | ATTGGAGAGAACTGTTCAGC | TTCAAAATCTCTCAACTGCACA |
| 13 | 384 | ATTAGGTAGTAAATACATCCTGA | TGCTGAACATCCTAGAAGCC |
| 14 | 295 | AGGCCAGGCTTTAAAAATGAC | TGATGTATTAACACTTCCCACT |
| 15 | 281 | TTCTATCTTTGTGCCCATCTC | TACAGAGATCTAACCATTAGCA |
| 16 | 292 | TTTTGTAAAGTGTAACCCTCTC | CAAAGTTAAAGCCTAGCTTCTG |

PCR-SSCP, polymerase chain reaction–single-strand conformation polymorphism

**Table 2.** Exon–intron boundary sequences of the N-cadherin gene

| Exon number | Exon length (bp) | Splice acceptor | Splice donor |
|---|---|---|---|
| 1 | 60 | | CCTGCTTCAG**gt**acgcggcg |
| 2 | 112 | tgtgttgc**ag**GCGTCTGTAG | CTTCTCAATG**gt**actatctc |
| 3 | 227 | tcttttac**ag**TGAAGTTTAG | GTCAGTGAAG**gt**atagtatc |
| 4 | 147 | taatgtaa**ag**GAGTCAGCAG | GCTTGTCAGG**gt**aaggtggc |
| 5 | 156 | tgatcctc**ag**ATCAGGTCTG | CCGGTTTCAT**gt**aagattcc |
| 6 | 145 | acttttgc**ag**TTGAGGGCAC | TCAAAGCCTG**gt**aagttta |
| 7 | 173 | ctgtctct**ag**GAACATATGT | TGATCGAGAA**gt**aagccaac |
| 8 | 138 | gtctttttt**ag**AAAGTGCAAC | TGCCATGACG**gt**gagtacag |
| 9 | 186 | tttctcgc**ag**TTTTATGGTG | CGTGGTCAAA**gt**aagtgttc |
| 10 | 254 | tattttcc**ag**CCAATCGACT | AAAATATTAG**gt**atgaaatg |
| 11 | 143 | cattttgt**ag**ATACACTAAA | TCTGACAATG**gt**atgttcgc |
| 12 | 234 | ctcttttc**ag**GAATTCCTCC | CGGCTTAATG**gt**aagaacag |
| 13 | 234 | tttcatat**ag**GTGATTTTGC | ATCCTGCTTA**gt**gagtactt |
| 14 | 140 | ttctacct**ag**TCCTTGTGCT | AGAAGACCAG**gt**gagcagtg |
| 15 | 165 | cttttgc**ag**GACTATGACT | CATTAATGAG**gt**acagagag |
| 16 | 207 | caaattgc**ag**GGCCTTAAAG | |

obtained. A polymerase chain reaction–single-strand conformation polymorphism (PCR-SSCP) analysis for each exon was carried out as described previously (Harada et al. 2001) by using the PCR primer sets shown in Table 1.

## Results and discussion

The human N-cadherin gene consisted of 16 exons spanning more than 72 kb. Exons 4 to 13 encoded five repeated cadherin domains. Exons 13 and 14 encoded the transmembrane domain. Exons 14 to 16 encoded the cytoplasmic domain. The Shc-binding site was located in the middle of Exon 15 within the cytoplasmic region. The exon/intron boundaries are shown in Table 2. Further intron sequences can be obtained upon request.

We found a total of eight sequence polymorphisms within the *NCAD* gene: five single-nucleotide polymorphisms (SNPs) and three trinucleotide repeat polymorphisms. Among the five SNPs, the three at codons 118, 826, and 845 were nonsynonymous substitutions and the two at codons 111 and 816 were synonymous substitutions. The Ala826Thr substitution in exon 15 was located in the binding region for the adaptor protein Shc, which has been shown to interact with N-cadherin (Xu et al. 1997) and may affect the association between Shc and N-cadherin. Allele frequencies in the 96 normal Japanese individuals are displayed in Table 3. No germline mutation or somatic muta-

**Table 3.** Sequence variations in the N-cadherin gene

| Exon | Position | Polymorphism | Allele frequencies |
|---|---|---|---|
| 1 | 5′-UTR | (CCT) 4 vs (CCT) 3 | 0.98 vs 0.02 |
| 1 | 5′-UTR | (GCC) 8 vs (GCC) 9 | 0.98 vs 0.02 |
| 1 | 5′-UTR | (GCC) 8 vs (GCC) 6 | 0.97 vs 0.02 |
| 3 | codon 111 | ACC (Thr) vs ACG (Thr) | 0.87 vs 0.13 |
| 3 | codon 118 | GCA (Ala) vs ACA (Thr) | 0.87 vs 0.13 |
| 15 | codon 816 | GCT (Ala) vs GCC (Ala) | 0.42 vs 0.58 |
| 15 | codon 826 | GCC (Ala) vs ACC (Thr) | 0.99 vs 0.01 |
| 16 | codon 845 | AAT (Asn) vs AGT (Ser) | 0.98 vs 0.02 |

The digit immediately after the parenthesis shows the repeat number of the trinucleotides

tion was detected among the patients with familial hypertrophic cardiomyopathy or brain tumors.

The three trinucleotide repeat polymorphisms were all found in the 5′ untranslated region directly upstream of the initiation codon. We did not detect any trinucleotide repeat expansion in the N-cadherin polymorphisms; thus, it is unlikely that they cause a loss of transcription as seen in triplet-repeat diseases associated with neurological disorders (Cummings and Zoghbi 2000). The detailed exon–intron boundary sequences and novel polymorphisms will be useful for genetic studies of N-cadherin in diseases involving cell adhesion in the brain and in cardiomyocytes.

## References


Asano K, Kubo O, Tajika Y, Huang MC, Takakura K, Ebina K, Suzuki S (1997) Expression and role of cadherins in astrocytic tumors. Brain Tumor Pathol 14:27–33

Asano K, Kubo O, Tajika Y, Takakura K, Suzuki S (2000) Expression of cadherin and CSF dissemination in malignant astrocytic tumors. Neurosurg Rev 23:39–44

Blaschuk OW, Sullivan R, David S, Pouliot Y (1990) Identification of a cadherin cell adhesion recognition sequence. Dev Biol 139:227–229

Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genomics Hum Genet 1:281–328

Garcia-Castro MI, Vielmetter E, Bronner-Fraser M (2000) N-Cadherin, a cell adhesion molecule involved in establishment of embryonic left–right asymmetry. Science 288:1047–1051

Harada H, Nagai H, Tsuneizumi M, Mikami I, Sugano S, Emi M (2001) Identification of DMC1, a novel gene in the TOC region on 17q25.1 that shows loss of expression in multiple human cancers. J Hum Genet 46:90–95

Wallis J, Fox MF, Walsh FS (1994) Structure of the human N-cadherin gene: YAC analysis and fine chromosomal mapping to 18q11.2. Genomics 22:172–179

Xu Y, Guo DF, Davidson M, Inagami T, Carpenter G (1997) Interaction of the adaptor protein Shc and the adhesion molecule cadherin. J Biol Chem 272:13463–13466