Complete DNA Sequence and Characterization of a 330-kb VNTR-rich Region on Chromosome 6q27 That is Commonly Deleted in Ovarian Cancer

Takeo Minaguchi, Mieko Matsushima, Susumu Saito, Yasunobu Kanamori, Syuya Shirahama, Sachiko Okamoto, Maiko Minami, Yuji Taketani, and Yusuke Nakamura

Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, Department of Obstetrics and Gynecology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan, Center for Molecular Biology and Cytogenetics, SRL, Inc., 5-6-50 Shin-machi, Hino, Tokyo 191-0002, Japan, and Department of Human Genome Analysis, Cancer Chemotherapy Center, Japanese Foundation for Cancer Research, 1-37-1 Kami-Ikebukuro, Tokyo 170-8455, Japan

(Received 18 January 1999; revised 16 February 1999)

Abstract

We report the complete genomic DNA sequence and the characterization of a 330-kb region on chromosome 6q27 that is often deleted in ovarian cancers. Using computer programs to predict exonic sequences, we isolated four novel genes, HGC6.1-4, as well as the known AF-6 gene. None of the deduced products of the novel genes exhibited significant homology to previously known proteins. We also identified ten microsatellites and 12 different VNTR sequences within the target region. HGC6.3 contained a VNTR within a coding exon, each repeat consisting of 42 nucleotides; the predicted 14-amino-acid consensus unit is MTPTVFSSQHTAGG. At least nine different sizes of this VNTR locus were detected among 20 unrelated DNA samples from Caucasians. The polymorphic markers and the transcript map documented here may contribute to identification of novel genes or allelic aberrations associated with the development of ovarian cancers.

Key words: chromosome 6q27; DNA sequencing; VNTR; ovarian cancer

1. Introduction

A distal portion of the long arm of chromosome 6 is thought to contain one or more putative tumor suppressor genes, on the basis of loss of heterozygosity (LOH) studies in a large number of ovarian cancers. We previously performed a deletion-mapping analysis of this chromosomal arm and defined a commonly deleted region between loci defined by CI6-111 (D6S193) and CI6-24 (D6S149). Subsequently, a detailed analysis using a YAC contig map further defined the region of interest within an approximately 300-kb segment of 6q27 flanked by two cosmid loci, CI6-24 (D6S149) and A2.

By means of exon-trapping using overlapping cosmid clones, we found that the human AF-6 gene, which is disrupted in acute myeloid leukemia cells carrying a t(6;11)(q27;q23) translocation, was located within the region. We screened a large panel of ovarian cancers for mutations in AF-6, but failed to find any mutations in any of the samples examined.

As a step toward identifying other candidate genes for suppression of ovarian-cancer pathogenesis, we have sequenced the entire deletion region using a 330-kb cosmid contig extending from marker D6S149 (CI6-24) to A2. This report describes the results of genomic DNA sequencing and the identification of four novel genes.

2. Materials and Methods

2.1. DNA sequencing of cosmid clones

Isolation of YAC clones, construction of cosmid libraries, and a cosmid contig map were described in the previous paper. The DNA of each of the cosmid clones

Communicated by Yoshiyuki Sakaki

* To whom correspondence should be addressed. Tel. +81-3-5449-5372, Fax. +81-3-5449-5433, E-mail: yusuke@ims.u-tokyo.ac.jp
† Abbreviations: VNTR, variable number of tandem repeat; RT-PCR, reverse-transcribed polymerase chain reaction; RACE, rapid amplification of cDNA ends
‡ T.-M. and M. M. contributed equally to this report.
spanning the region flanked by CI6-24 and A2 was digested with EcoRI. Each EcoRI fragment was isolated, subcloned into pBluescriptII SK(-) cloning vector, and sequenced by the primer-walking method using ABI 377 autosequencers (Applied Biosystems). DNA sequences were assembled using ABI “Assembler” computer software.

2.2. Sequence interpretation

To eliminate repetitive DNA sequences contained in the 330-kb region, we filtered the nucleotide sequences by means of the REPEATMASKER program (Smit and Green at http://ftp.genome.washington.edu/RM/RepeatMasker.html). We used the BLAST8 and FASTA9 programs to search for homologies, the GRAIL10 and GENSCAN11 programs to predict exons, and the PSORT12 and SMART13 programs to predict protein motifs or domains.

2.3. Isolation of cDNAs

Messenger RNAs from various human tissues were reverse-transcribed using oligo(dT) primers or random hexamers to prepare PCR templates. RT-PCR reactions for exon-connection experiments were primed with oligonucleotides corresponding to the exonic regions predicted by GRAIL. Human cDNA libraries from fetal brain and small intestine (Stratagene) were screened using the RT-PCR products as probes. We also subjected the cDNA fragments to 5’- and 3’-rapid amplification of cDNA ends (RACE), using mRNAs from various human tissues as templates and the Marathon cDNA amplification kit (Clontech), according to the manufacturer’s instructions.

2.4. Northern blot analysis

Human multiple tissue Northern blots (Clontech) were hybridized with random-primed, 32P-labeled RT-PCR products. Prehybridization, hybridization and washing were performed according to the manufacturer’s instructions.

3. Results

3.1. Construction of a physical map of the deletion region on chromosome 6q27

On the basis of our cosmid-contig map, the entire region of interest was covered by 14 cosmids clones (Fig. 1A). Overlapping cosmid clones were aligned by restriction-digest mapping with enzymes MluI, NotI and NruI (Fig. 1C). We determined the entire DNA sequences of these 14 clones by the primer-walking method, using subclones of their EcoRI fragments, and obtained a total of 331,211-nucleotides (DDBJ/EMBL/GenBank accession No. AB016897). Computer analysis of this 331-kb sequence indicated that 76 copies of Alu (5.91%) and 46 copies of LINEs (9.4%) were present (Fig. 1J and 1K). The average GC content of this DNA segment was 43.6%. Fourteen CpG islands and ten CA-repeat microsatellites (Fig. 1E and 1H) were detected.

3.2. VNTRs identified in 330-kb region

We also found 12 VNTR loci14 in this region (Fig. 1I), whose consensus sequences are summarized in Table 1 as MM6.1-12. Approximate numbers of repeats at each VNTR locus were estimated by reference to our sequence data or through PCR experiments using cosmid clones. MM6.2 was identical with the D-segment (D6S149) used for deletion mapping in our previous study, which had detected three alleles by PvuII digestion. We detected nine alleles of MM6.7, which is part of the coding sequence of HGC6.3 (see below), among 20 unrelated DNA Caucasian samples in genomic PCR experiments (data not shown). The consensus sequence of MM6.1 (GGGCCCTTTC) resembled binding sites for the NFκB transcription factor found in immunoglobulin-κ light chain (GGGACTTTCC), the β-interferon gene enhancer κB (GGGAAATCCC), and the major histocompatibility class 1 κB element (GGGAATTCGG).15

3.3. Detection of coding elements

We analyzed genomic DNA sequences from the target region with two exon-prediction computer programs, GRAIL2 (Fig. 1F) and GENSCAN (Fig. 1G), and also searched a database of expressed sequence tags (dbEST) by BLAST (Fig. 1D). On the basis of this information, we performed exon-connection experiments. Connected sequences that were considered to be transcribed exons were then subjected to Northern analysis. Homology searches to match these fragments with known sequences in the public DNA database detected one gene, AF-6, that we had reported previously.16 Together, the analyses revealed the presence of at least four novel putative genes in the 330-kb region. The cDNAs for each transcript were obtained by means of cDNA library screening and 5’ and 3’ rapid amplification of cDNA ends (RACE). We designated the novel genes HGC6.1-4 (Table 2); their locations on the physical map are indicated in Fig. 1B.

3.4. Genes in 331-kb region

3.4.1. HGC6.1

GENSCAN analysis detected two exons of HGC6.1, and we obtained two types of HGC6.1 transcript consisting of 2785 and 1407 nucleotides respectively (DNA sequences are not shown, but are available from DDBJ/EMBL/GenBank with accession numbers AB016899 and AB016900). A 2.8-kb cDNA for HGC6.1 was isolated from a human small-intestine library, using an RT-PCR product of the connected exons as a probe, while the 1.4-kb transcript was obtained by 3’-RACE.
Figure 1. Detailed physical map and structural features of the 330-kb region at 6q27 flanked by CI6-24 and A2. (A) Locations of the overlapping cosmid clones subjected to nucleotide sequencing. (B) Location of five genes (HGC6.1-4 and AF-6). (C) Endonuclease cleavage sites. (D) Expressed sequence tags (ESTs) mapped in this area. ESTs identical to the AF-6 gene were excluded. (a) GenBank D81921 (b) R00354 (c) AA927364 (d) AI016136 (e) AA317920. (E) CpG islands predicted by the GRAIL2 program. (F) Positions of the potential exons predicted by GRAIL2. (G) Positions of the potential exons predicted by the GENSCAN program. (H) Locations of ten (CA)n microsatellites. (I) Locations of 12 VNTR (variable number of tandem repeat) sequences. (J) Locations of Alu repetitive elements. (K) Locations of LINE1 repetitive elements.

and RT-PCR analysis using mRNA from skeletal muscle. The 2.8-kb transcript (HGC6.1.1) consisted of three exons, including an open reading frame (ORF) encoding 181 amino acids; the 1.4-kb transcript (HGC6.1.2) consisted of four exons with an ORF encoding 163 amino acids (Fig. 2). These two putative proteins used alternative reading frames for their N-terminal portions, but shared 96 amino acids at their C-termini. Northern blot analysis revealed a major 1.6-kb transcript and a minor 2.5-kb transcript in all tissues examined (data not shown). We assume that HGC6.1.2 accounted for the 1.6-kb band seen on the Northern blots. Since we could not obtain cDNA clones corresponding to the 2.5-kb band, it is unclear whether the 2.5-kb band represented HGC6.1.1 or was another transcript generated by alternative splicing. The predicted amino-acid sequences of the transcripts revealed no significant similarities to any known protein, except for partial homology of HGC6.1.1 to human s-laminin (Fig. 3A),17 and of HGC6.1.2 to human CD40 (Fig. 3B)18 and human probable tumor suppressor protein MN1 (Fig. 3C).19 The SMART program predicted three potential phosphorylation sites (SARE, RRCS, and SSHE) in the deduced HGC6.1.1 product and six (RRCS, TALD, TTKD, SSHE, SSR and SLK) in the HGC6.1.2 protein sequence.

3.4.2. HGC6.2
GRAIL analysis detected two exons of HGC6.2; cDNA for the transcript was obtained by RT-PCR and 3’- and 5’-RACE experiments. The cDNA consisted of three exons and contained 1198 nucleotides, with an ORF encoding 57 amino acids (DDBJ/EMBL/GenBank accession number AB016901). RT-PCR experiments indicated that HGC6.2 was expressed in all tissues examined (data not shown). The predicted protein sequence of this gene exhibited partial similarity to human WT120 (Fig. 3D), al-
Table 1. Consensus sequence of 12 VNTR loci.

| Repeating unit | Consensus sequence | Number of repeats present in the cosmid |
|----------------|-------------------|----------------------------------------|
| MM6.1          | 37bp              | ATAGAGCCCCCAGCTGGGCGCCTTCCGGCTGGGGCC   | 7 |
| MM6.2          | 21bp              | TCTTTCCGAAACTGGAGAGGAG                 | 24 |
| MM6.3          | 39bp              | GGCTGCTTACAGGGGCCTGGGTGTGGTGATCCCA     | 15 |
| MM6.4          | 27-30bp           | CTCAGGCGAGGGGGAGGAGAAGAGGA             | 21 |
| MM6.5          | 13bp              | GTGAGGGGGGAG                        | 29 |
| MM6.6          | 58bp              | CAGACAGAGGGGAGGTGAGGAGGGAAACATCCACAGGGAGGTGGTGAGA | 25 |
| MM6.7          | 42bp              | ATCCCCCTCGAGAGGGGCAGGGCAGGGCTGGGGTC   | 12 |
| MM6.8          | 23bp              | TCCCCGGTGGAGGGAGAGGAGGAGG             | 15 |
| MM6.9          | 12-13bp           | ACAGGGGAGGGG                         | 11 |
| MM6.10         | 24-25bp           | AAGTGGCAGAGAGGGGGAGGGGAGGG            | 11 |
| MM6.11         | 31bp              | CAGAATGGGAGGGGGAGGGAGGGTGGTGGT       | 21 |
| MM6.12         | 33-35bp           | ACCACGGGCTGGGCCAGGGCGAGGGAGAAG       | 14 |
| VNTR consensus (Ref. 14) | | | |

Table 2. Genes identified from the 330-kb region.

| Gene   | Status          | Method of isolation | Transcript size (Kb) | Expression profile | Number of exons sequenced | Database hits          |
|--------|-----------------|---------------------|---------------------|-------------------|--------------------------|------------------------|
| AF-6   | Reported previously | cDNA library screening, RT-PCR | 6.0 / 8.0 | Ubiquitous (complete) | 32 | AF-6 |
| HGC6.1 | This study      | cDNA library screening, RACE, RT-PCR | 1.6 / 2.5 | Ubiquitous (complete) | 4 | - |
| HGC6.2 | This study      | RACE, RT-PCR       | - | Ubiquitous (incomplete) | 3 | - |
| HGC6.3 | This study      | RACE, RT-PCR       | 1.3 | Ubiquitous (complete) | 1 | EST (AA317920) |
| HGC6.4 | This study      | cDNA library screening | - | Ubiquitous (incomplete) | 2 | EST (AA927364) |

a) Each cDNA clone was examined by hybridization against multiple-tissue Northern blots.
b) Gene expression was examined by Northern blot or RT-PCR analysis.

though the homology occurred within neither the proline-rich transcription domain nor the zinc-finger domain of WT1.

3.4.3. HGC6.3

HGC6.3 was isolated through an EST (AA317920) which matched the genomic sequence. The transcript was further extended in the 5' direction using RT-PCR, and was also linked to a poly(A) tail by means of 3'-RACE. The isolated cDNA contained 1016 nucleotides including an ORF encoding 171 amino acids, and consisted of a single exon (DDBJ/EMBL/GenBank accession number AB016902). Northern blot analysis revealed a 1.3-kb signal in all tissues examined (data not shown). The coding region of this gene contained a tandem repeat sequence of 42 nucleotides. The predicted 14 amino acids encoded by this VNTR (consensus unit, MTPTVFSSQHTAGG) exhibited similarity to the tandem repeats of the yeast a-agglutinin attachment subunit (consensus unit, TSTSPSS). The tandem repeat sequence in the HGC6.3 coding region was identical to the VNTR identified as MM6.7 in Table 1 and as #7 in Fig. 11. MM6.7 alleles of nine different sizes were detected among 20 unrelated caucasian DNA samples in genomic PCR experiments (data not shown).

3.4.4. HGC6.4

HGC6.4 was isolated by screening a human fetal-brain cDNA library using cosmid A44 as a probe. The isolated cDNA was 2225 nucleotides long, and contained an ORF encoding 254 amino acids (DDBJ/EMBL/GenBank accession number AB016898). RT-PCR indicated that HGC6.4 was expressed in all tissues examined (data not shown). The PSORT program predicted a mitochondrial targeting sequence, ARSKWA, within the putative amino acid sequence of the gene. Four copies of a sequence encoding AGSKWA were also present in the gene.
We have sequenced and characterized a 330-kb segment of genomic DNA at 6q27 that is commonly deleted in ovarian cancers. By various approaches we identified four novel genes, termed HGC6.1-4. The putative amino acid sequence of HGC6.1.1 showed local homology to the three EGF-like domains of human s-laminin. HGC6.1.2 possessed weak similarity to the signal-peptide portion and the second TNFR-type cysteine-rich region of human CD40 receptor, and to the probable tumor suppressor MN1. The putative protein sequence encoded by HGC6.2 exhibited partial similarity to human tumor suppressor protein WT1.

We found five genes altogether in the distal portion of the 331-kb target region, but were unable to confirm the presence of any gene in the proximal 100 kilobases even though the computer programs predicted multiple candidate exons. We found no significant differences in GC content, copy numbers of repetitive sequences, or GpC islands. This 100-kb region may in fact contain no transcriptional units, or it may contain genes expressed in very low abundance or in specific tissue(s) or time periods.

We found 12 VNTR sequences in this region. One of them was a coding element within HGC6.3, and six others were located within introns or in the 5' or 3' regions flanking AF-6, HGC6.1, or HGC6.3. VNTRs were originally isolated as highly informative restriction fragment length polymorphic markers for mapping purposes, but recent evidence indicates that some VNTR sequences, or “minisatellites,” play significant roles in regulation of transcription, and that some may also influence translational efficiency or stability of mRNA, or modify the
activity of the encoded proteins by altering their structure (reviewed by YN). Furthermore, observed associations of certain alleles of VNTR sequences with personality traits or diseases have supported the possibility that these tandem repeat genomic elements possess physiological and biological importance. Hence, future investigations to correlate the lengths of the VNTR sequences with expression levels of AF-6, HGC6.1, or HGC6.3 may provide new insight into the function of minisatellites. It will be of great interest also to determine whether differences in the length of the VNTR (MM6.7) that is a part of the coding sequence of HGC6.3 affects the biological function of this gene.

Acknowledgements: This work was supported by a “Research for the Future” Program Grant (96L00102) of The Japan Society for the Promotion of Science.

Note: The total genomic sequence of this 331-kb region will appear in the DDBJ/EMBL/GenBank databases with accession number AB016897. cDNA sequence data for HGC6.1-4 are also available, with individual accession numbers AB016898-AB016902.

References

1. Saito, S., Saito, H., Nakamura, Y. et al. 1992, Fine-scale deletion mapping of the distal long arm of chromosome 6 in 70 human ovarian cancers, Cancer Res., 52, 5815-5817.

2. Cliby, W., Ritland, S., Hartmann, L. et al. 1993, Human epithelial ovarian cancer allelotype, Cancer Res., 53, 2393-2398.

3. Fouikis, W. D., Ragoussis, J., Stamp, G. W. H., Allan, G. J., and Trowsdale, J. 1993, Frequent loss of heterozygosity on chromosome 6 in human ovarian carcinoma, Br. J. Cancer, 67, 551-559.

4. Van, M., Zweizig, S., D’ablaing, G., Zheng, J., Velicescu, M., and Dubeau, L. 1994, Three distinct regions of chromosome 6 are targets of loss of heterozygosity in human ovarian carcinomas, Int. J. Oncol., 5, 1043-1048.

5. Cooke, I. E., Shelling, A. N., Meuth, V. G. L., Charnock, F. M. L., and Ganesan, T. S. 1996, Allele loss on chromosome arm 6q and fine mapping of the region at 6q27 in epithelial ovarian cancer, Genes Chrom. Cancer, 15, 223-233.

6. Saito, S., Shirahama, S., Nakamura, Y. et al. 1996, Definition of a commonly deleted region in ovarian cancers to a 300-kb segment of chromosome 6q27, Cancer Res., 56, 5586-5589.

7. Praasad, R., Gu, Y., Alder, H. et al. 1993, Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in acute myeloid leukemias with the t(6;11) chromosomal translocation, Cancer Res., 53, 5624-5628.

8. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. 1990, Basic local alignment search tool, J. Mol. Biol., 215, 403-410.

9. Pearson, W. R. and Lipman, D. J. 1988, Improved tools for biological sequence analysis, Proc. Natl. Acad. Sci. USA, 85, 1345-1414.

10. Uberbacher, E. C. and Mural, R. J. 1991, Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach, Proc. Natl. Acad. Sci. USA, 88, 11261-11265.

11. Burge, C. and Karlin, S. 1997, Prediction of complete gene structures in human genomic DNA, J. Mol. Biol., 268, 78-94.

12. Nakai, K. and Kanehisa, M. 1992, A knowledge base for predicting protein localization sites in eukaryotic cells, Genomics, 14, 897-911.

13. Ogiwara, A., Uchiyama, I., Takagi, T., and Kanehisa, M. 1996, Construction and analysis of a profile library characterizing groups of structurally known proteins, Protein Sci., 5, 1991-1999.

14. Nakamura, Y., Leppert, M., O’Connell, P. et al. 1987, Variable number of tandem repeat (VNTR) markers for human gene mapping, Science, 235, 1616-1622.

15. Ghoesh, G., Van Duyne, G., Ghoesh, S., and Sigler, P. B. 1995, Structure of NF-kB p50 homodimer bound to a κB site, Nature, 373, 303-310.

16. Saito, S., Matsuhashi, M., Nakamura Y. et al. 1998, Complete genomic structure, DNA polymorphisms, and alternative splicing of the human AF-6 Gene, DNA Res., 5, 115-120.

17. Wewer, U. M., Gerecke, D. R., Durkin, M. E. et al. 1994, Human beta 2 chain of laminin (formerly S chain): cDNA cloning, chromosomal localization, and expression in carcinomas, Genomics, 24, 243-252.

18. Stamenovic, I., Clark, E. A., and Seed, B. 1989, A B-lymphocyte activation molecule related to the nerve growth factor receptor and induced by cytokines in carcinomas, EMBO J., 8, 1403-1410.

19. Deprez, R. H. L., Riegelman, P. H. J., Groen, N. A. et al. 1995, Cloning and characterization of MNI, a gene from chromosome 22q11, which is disrupted by a balanced translocation in a meningioma, Oncogene, 10, 1521-1528.

20. Calh, K. M., Glaser, T., Ito, C. Y. et al. 1990, Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilm’s tumor locus, Cell, 60, 599-620.

21. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D. et al. 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, Nature, 377 (6547 Suppl), 3-174.

22. Roy, A., Lu, C. F., Marykwas, D. L., Lipke, P. N., and Roy, A. 1993, Human chromosome 6 are targets of loss of heterozygosity in human ovarian carcinomas, Br. J. Cancer, 67, 551-559.

23. Nakamura, Y., Koyama, K., and Matsushima M. 1998, Frequent loss of heterozygosity on chromosome 6 in 70 human ovarian cancers, Cancer Res., 58, 5217-5220.

24. Ho, J., Kurjan, J. 1991, The AGA1 gene, DNA Res., 10, 1521-1528.

25. Call, K. M., Glaser, T., Ito, C. Y. et al. 1990, Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilm’s tumor locus, Cell, 60, 509-520.

26. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D. et al. 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, Nature, 377 (6547 Suppl), 3-174.

27. Roy, A., Lu, C. F., Marykwas, D. L., Lipke, P. N., and Kurjan, J. 1991, The AGA1 product is involved in cell surface attachment of the Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin, Mol. Cell Biol., 11, 4196-4206.

28. Nakamura, Y., Koyama, K., and Matsushima M. 1998, Variable number of tandem repeat (VNTR) (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators, J. Human Genet., 43, 149-152.