SHORT COMMUNICATION

Aritoshi Iida · Yozo Ohnishi · Kouichi Ozaki · Yoko Arijii
Yusuke Nakamura · Toshihiro Tanaka

High-density single-nucleotide polymorphism (SNP) map in the 96-kb region containing the entire human DiGeorge syndrome critical region 2 (*DGCR2*) gene at 22q11.2

Received: July 4, 2001 / Accepted: July 11, 2001

Abstract We constructed a high-density single-nucleotide polymorphism (SNP) map in the 96-kb region containing the DiGeorge syndrome critical region 2 (*DGCR2*) gene at chromosome 22q11.2, a human counterpart of mouse seizure-related gene *SEZ-12*. A total of 102 SNPs were isolated from the region by systematic screening among 48 Japanese individuals: 9 SNPs in the 5' flanking region, 3 in the 5' untranslated region, 2 in the coding regions, 77 in introns, 7 in the 3' untranslated region, and 4 in the 3' flanking region. By a comparison of our data with SNPs deposited in the dbSNP database in the National Center for Biotechnology Information, 80 SNPs (78.4%) were considered to be novel. The ratio of transition to transversion was 3.08:1. In addition, eight other types of genetic variations (one GA dinucleotide polymorphism and seven insertion/deletion polymorphisms) were discovered. The high-resolution map that we constructed will be a useful resource for analyzing gene scans of complex diseases mapped to this local segment on chromosome 22.

Key words Single-nucleotide polymorphisms (SNPs) · High-density SNP map · Japanese population · Human *DGCR2* gene · Coding SNPs · Nonsynonymous substitutions

Introduction

Deletions at 22q11 (OMIM #188400, #192430, #217095) are associated with a wide range of developmental defects embraced by the acronym CATCH22, including DiGeorge syndrome, Shprintzen syndrome (velocardiofacial syndrome), and congenital heart disease. The DiGeorge syndrome critical region 2 gene (*DGCR2*, also referred to as *IDD* and *LAN*) is a putative membrane glycoprotein isolated from the vicinity of the t(2;22)(q14.1;q11.2) translocation breakpoint associated with DiGeorge syndrome (Budarf et al. 1995; Demczuk et al. 1995; Gong et al. 1996; Wadey et al. 1995). In addition, the predicted amino acid sequence of *DGCR2* revealed several structural domains such as a cysteine-rich repeat domain, a C-type lectin domain, and a transmembrane domain corresponding to those of mouse seizure-related gene, *SEZ-12* (Kajiwarra et al. 1996; Taylor et al. 1997). These reports suggested that *DGCR2* plays a crucial role in embryogenesis and that haploinsufficiency of this gene may be partly related to the etiology of DiGeorge syndrome. The gene spans approximately 96 kb of genomic DNA and is composed of ten exons; exon 1 includes the translation-initiation codon, and the stop codon is located in the last exon.

In this study, we report the construction of a fine-scale variation map of the entire *DGCR2* gene region at 22q11, containing 102 SNPs, one dinucleotide polymorphism, and seven insertion/deletion polymorphisms in the Japanese population.

Materials and methods

Blood samples were obtained with written informed consent from 48 healthy Japanese volunteers for this study, which was approved by the ethical committee of the RIKEN SNP Research Center. We obtained genomic sequences containing the *DGCR2* gene from the GenBank database (accession numbers L77570.1 and AC004461.3), and then designed primer sets to amplify the *DGCR2* gene

A. Iida · Y. Nakamura
Laboratory for Genotyping, RIKEN SNP Research Center, c/o
Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Y. Nakamura
Laboratory of Molecular Medicine, Human Genome Center,
Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Y. Ohnishi · K. Ozaki · Y. Arijii · T. Tanaka (✉)
Laboratory for Cardiovascular Diseases, RIKEN SNP Research
Center, c/o Institute of Medical Science, The University of Tokyo, 4-
6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
Tel. +81-35449-5675; Fax +81-35449-5674
e-mail: toshitan@ims.u-tokyo.ac.jp

Fig. 1. A high-density single-nucleotide polymorphism (SNP) map spanning the 96-kb region containing the *DGC2* gene. Detailed information of variations isolated in this study is shown in Table 1. Exons and introns are represented by rectangles and horizontal lines, respectively. Hatched boxes are shown next to the regions containing repetitive sequences. In addition, polymerase chain reaction (PCR) amplicons are represented by horizontal lines. The 102 SNPs are indicated above the gene (designations correspond to those in the left-most column of Table 1); locations of other types of polymorphisms listed in Table 2 are also indicated

in its entirety and up to 2 kb upstream of the first exon and downstream of the last exon. A total of 41 211 bp (42.9%) in the 96 070-bp region of the *DGCR2* gene was screened, except the regions containing repetitive sequences, which were predicted by the RepeatMasker Program. Each polymerase chain reaction (PCR) was carried out using 20 ng of DNA pooled from three individuals. Amplification of genomic DNA fragments by PCR and DNA sequencing of the amplified fragments was performed according to methods described previously (Ohnishi et al. 2000). All variations detected by the PolyPhred computer program (Nickerson et al. 1997) were confirmed by sequencing both strands of each PCR product.

Results and discussion

A total of 102 SNPs were identified in the 96-kb region containing an entire *DGCR2* gene: 9 in the 5' flanking regions, 3 in the 5' untranslated region, 2 in the coding re-

gions, 77 in introns, 7 in the 3' untranslated region, and 4 in the 3' flanking region. The exon organization of the *DGCR2* gene and the locations of identified SNPs are illustrated schematically in Fig. 1 (also see Table 1). The density of SNPs in the sequenced 41.2-kb region was 1 in 404 bp on average. By a comparison of our data with SNPs deposited in the dbSNP database in the National Center for Biotechnology Information, 80 SNPs (78.4%) were considered to be novel (Table 1). The frequencies of the substitutions were 38.2% for C/T, 37.2% for A/G, 6.9% for A/C, 6.9% for G/T, 5.9% for C/G, and 4.9% for A/T. The ratio of transition to transversion was 3.08:1. In general, the CpG dinucleotide is a hotspot for variation in vertebrate genomes and its variation rate is known to be 8.5 times higher than the average (Cooper et al. 1995). Analysis of the nucleotide neighbors of the proposed variations reveals no bias for C/T (G/A) substitutions to occur at CpG dinucleotides (37.7% of C/T [G/A] SNPs were at CpG dinucleotides).

In addition, only 2 (1.96%) of the 102 SNPs were nonsynonymous substitutions; 1 is a 1324G>A (G442S) in exon 9 and the other is a 1418T>C substitution (V473A) in

Table 1. Characterization of single-nucleotide polymorphisms (SNPs) from the *DGCR2* gene locus

ID	Location	Exon	SNP (5' to 3')	Substitution	Identify to dbSNP
RSC22-1	5' flanking region	-4635	tccccactgggagctgggag	C/T	caggctggtggggagctggct
RSC22-2	5' flanking region	-3955	tctgatggacagagcacaat	G/T	aacgagttgcccagcagggt
RSC22-3	5' flanking region	-3781	tctacaggggtacctcgga	T/C	gcttcttggtcctgtcctct
RSC22-4	5' flanking region	-3752	tgggatgtgttttcccaagc	A/G	aacgaggaatcattgcacaa
RSC22-5	5' flanking region	-3383	aacagcggtcctccgggac	C/T	gtcttctctctctgtgttgt
RSC22-6	5' flanking region	-3123	ctcgggctgtgtctgttccc	C/T	aggtctcagcttgggtggcag
RSC22-7	5' flanking region	-3110	tgttccccagggctccagctt	G/A	gtggcagacctcttggcaga
RSC22-8	5' flanking region	-2580	acacttgagggtctgtgca	T/C	gatgtccaggctcgtggcagt
RSC22-9	5' flanking region	-2280	tctgtgtccagcactgttcc	C/T	gggctatgccctaggggagt
RSC22-10	5' untranslated region	-93	gcgcctcgcccagggtcagc	T/C	ccgcctgacctcccgcttgg
RSC22-11	5' untranslated region	-45	ccctgacctcagctccggcc	T/C	ggctccggccttccagccgcc
RSC22-12	5' untranslated region	-34	agtcggcctgggtccggcct	T/C	ccagccgccccgggagcagat
RSC22-13	intron1	1543	tgggtgatagaggtgactga	C/A	gtccccgactccttttctct
RSC22-14	intron1	5066	tctgtgtgtgtcctcttagc	A/G	gtaagggcaggtgatctagc
RSC22-15	intron1	5106	ctgtgcacgtgtcctcttag	C/A	catggtgggaggtgatctgg
RSC22-16	intron1	5317	ctagtattgactctctcct	G/A	ctggacctgagcctttgaca
RSC22-17	intron1	5455	acatttaaggaccctcgtaga	T/A	agctcttgactgattagatg
RSC22-18	intron1	5664	tgggtctgggtatggttatt	A/G	tgtgatagtctgtgtaacacc
RSC22-19	intron1	6294	ccctcaggaggtggcagagt	A/G	cagtggctctgaggagtgtc
RSC22-20	intron1	6343	agagagtgtcttctgtgtt	A/C	tgatgggatcagagggactg
RSC22-21	intron1	6371	atcagagggtgctcgtatga	C/T	ttgctgagagccccgaaac
RSC22-22	intron1	6554	atatcttgacataaagagca	A/C	cagctacttttagccagcagg
RSC22-23	intron1	7058	cttctctctctctctctgtg	C/G	ggtagctgtgtcttcagtea
RSC22-24	intron1	7588	agtttaacctacagaactctg	T/C	ctatggctcttagaataagag
RSC22-25	intron1	7661	atggcaggccttctgtgcag	T/C	tcttggaggttctctggcca
RSC22-26	intron1	15308	gccccaccagggtctctgc	T/A	ggcttcagggtatggggattt
RSC22-27	intron1	15470	tcaatacatagaacagcttc	G/A	ggccttctctcacattttgc
RSC22-28	intron1	16550	acagtgtgccttccagagcg	T/C	gcacgtctccttccccact
RSC22-29	intron1	16588	actgggcagcactgttcag	A/G	atgggtgcttggcagaggcc
RSC22-30	intron1	16602	gttcagaatgggtgcttgcc	A/G	gaggcctgccatggccctt
RSC22-31	intron2	84	gcctcttctgtagggtggag	G/T	tgtgtgtgtgtcttaacacat
RSC22-32	intron2	281	gcactcctggaacttggatg	A/G	tcttgagtgtgaacctctcc
RSC22-33	intron2	2128	gggacagtctcttctgtggg	A/G	gcctctctgtgtgggtctttg
RSC22-34	intron2	3794	tgggtttcagaggctggaat	G/A	ctggagcagtagcactcaga
RSC22-35	intron2	18332	accaaaggctggccctccgc	G/A	ggaagaggacattctctac
RSC22-36	intron2	18733	caactgggtctccacccacg	G/A	gggcagatgacttccactcag
RSC22-37	intron2	18993	cagtgcattctccagccac	A/G	ctcaggcctggcctcagggtt
RSC22-38	intron2	19020	cctggcctcagggttatccc	G/A	gatatgtctgcctttgcctt
RSC22-39	intron2	19156	caagtgtgatgcccatcc	T/C	agtggctgcacccctgcac

Table 1. Continued

ID	Location	Exon	SNP (5' to 3')	Substituion	Identiy to dbSNP
RSC22-40	intron2	19691	aaccatagagatttatgggga	T/C	gaaagaaatccatttgtgga
RSC22-41	intron2	19900	catagttaaatatggccttt	G/A	ttgtttatgtcttttcaaa
RSC22-42	intron2	20860	caggaccctttgtgtgcta	G/T	caatgacctgttagaaaaatg
RSC22-43	intron2	21042	actctgtgtctacagggcccc	C/G	ctccggccttgacggcctga
RSC22-44	intron3	107	aggggtttctcaaacgagcc	C/T	ttggaggccagttaggggaa
RSC22-45	intron3	154	aggctgcatgatgaccacc	C/T	gcagtcaccagcgctggag
RSC22-46	intron3	1319	cttctgcctgggtagtaagc	A/G	actccaccaagttcagaaca
RSC22-47	intron3	2205	actagtgcacacagacatag	C/A	tgcactttaaaaatggcatt
RSC22-48	intron3	2321	agggttaaagattctataaaa	G/C	aggtcactgggtttacctaa
RSC22-49	intron3	2595	gtggagggttgccagccccc	T/C	actgcaaatgctcatgtaag
RSC22-50	intron3	2639	aaaaggaggagaggggagtg	A/G	gctagctggcagcgggagcc
RSC22-51	intron4	980	cctgtggagtgcctgtcct	A/G	tggttttagcatagtcaccg
RSC22-52	intron4	988	ctatggtttttagcatagtca	C/T	cgagttgtattcatggggtt
RSC22-53	intron5	80	ctctccctctctccctcttac	C/T	cagtgtgctgctggccctag
RSC22-54	intron5	605	ggacatactccagcagtgcc	T/C	tctccagctgcccaggggt
RSC22-55	intron5	705	gagctaggatttggttctt	T/C	tcgtaaacacatcccagttcc
RSC22-56	intron5	708	ctaggattttgggttcttttc	G/A	taaacacatcccagctctgc
RSC22-57	intron5	844	ctggagaaaaggaggcccca	C/T	agaggcggggccatgcctgg
RSC22-58	intron5	1058	agattgctattgactaatca	A/G	tggtctctgggtttccaggc
RSC22-59	intron5	1113	ttgcctcatagagcataatt	A/C	tcctggtagcttgtcctcct
RSC22-60	intron5	1437	cacggaggtgaccatgtgcc	A/G	ctgcctgcttgtgcccagtt
RSC22-61	intron5	2027	acatgggtgcaaagcagaggg	G/C	tcattgggggtctgaggtgtt
RSC22-62	intron5	2052	tggggctctgaggtgttgcc	C/A	ccccaaaaatacccaaaaa
RSC22-63	intron5	2068	ggcccccccaaaaaaccaca	A/G	aaaacccatgggaaaatggatg
RSC22-64	intron5	2209	aaatgtttcttctcatctaca	T/A	gttgctcatctttctgtgt
RSC22-65	intron5	2266	gattggggtgtgagtatac	C/T	atggtacatagacgcttact
RSC22-66	intron5	2302	ttactatgccccatgtgtcg	T/C	accctgcacacacatttgaa
RSC22-67	intron5	2937	cgctgaggccagcagtcagg	G/T	ctgtgggtcctgggccagcct
RSC22-68	intron5	3134	ggggagacagaccctgagtg	G/A	ggcagctgccaccgtggtgg
RSC22-69	intron5	4778	tcagcagtgagaccagccc	T/A	ccaggtggtagtgggaggtg
RSC22-70	intron5	4926	ctgtgagctgcccgtggcc	T/C	ggcacaggtctctgcaggtct
RSC22-71	intron6	237	gcttcttttccacggaaga	G/A	acaatcaccctgtcctccc
RSC22-72	intron6	7745	gaetgcttcccaagcagaa	C/T	tggcagcttctttacacaca
RSC22-73	intron6	7760	cagaactggcagcttctta	C/G	acacattctgattttgaaat
RSC22-74	intron6	8020	atccatggcattccagccct	G/A	ggaagactctcagcacaatc
RSC22-75	intron7	881	cagctctgtgtgaattcaga	T/C	ctcaagagaaatgtaaatac
RSC22-76	intron7	1120	aatgttgctcagtgccctcc	A/G	cactggggggctgctgggtg
RSC22-77	intron7	1124	ttgtcagtgctccctccacac	T/C	gggggggctgctgggtggag
RSC22-78	intron7	4335	ttgggtttattttagaacgg	A/G	ttaaaggttagtcattgcaa
RSC22-79	intron7	4841	ggagctgatgggtcactcct	C/T	gttttgaagcctccctggcc
RSC22-80	intron7	5117	aggggtgtgctccctctggcg	G/C	tggtgacatttgaccagat
RSC22-81	intron7	5227	gcattccagccaagtgggtg	T/C	cacgaaaagcgtgaaatggt
RSC22-82	intron7	5601	cgcccaaccacagcagaggg	A/G	catttgtgaaatgatacctg
RSC22-83	intron7	5613	gcagaggcacatttgtgaaa	C/T	gatacctgagttgaccgtgc
RSC22-84	intron7	5836	cccttgacagctgtcctccc	T/C	acggcagcgcggcagccat
RSC22-85	intron7	5942	cctggccaaacctggcctcc	T/C	accaggaaagggtgtcctc
RSC22-86	coding region	1324	exon9 catacaagtaccggacatc	G/A	gccagcccagcaccctccg Gly 442 Ser
RSC22-87	intron9	91	ttgccagctgcagaggcct	A/G	gggacctcaactcttcccg
RSC22-88	intron9	95	ccagctgcagaggcctaggg	A/G	cctcacttcttccgcttt
RSC22-89	intron9	974	ggggctgtgagggcctggcc	G/A	ggaggaggtgcatcaccgg
RSC22-90	intron9	1694	gtgtctgcctttgtgagatg	G/T	gccctgcaccttgcatatac
RSC22-91	coding region	1418	exon10 cgatgatgcttttgagcctg	T/C	ggagggtcagcctgccagccc Val 473 Ala
RSC22-92	3' untranslated region	2006	exon10 ctgatgcgtttttgggtggc	T/C	tacttgggtcagtgctgtgt
RSC22-93	3' untranslated region	2008	exon10 gatgcgtttttgggtggc	A/T	cttgggtcagtgctgtctcc
RSC22-94	3' untranslated region	2134	exon10 tggaccgtcagtggtcgta	G/A	gcagtggtgtctgtccaggg
RSC22-95	3' untranslated region	2265	exon10 gggagctcccttgggctct	G/A	gaacctgaagccctgagaag
RSC22-96	3' untranslated region	2650	exon10 agtccactccttacatcacc	G/A	ccactactgcagctcagtc
RSC22-97	3' untranslated region	2701	exon10 catgcacacctcccgccccc	T/G	cctcagcagcgcaggggctg
RSC22-98	3' untranslated region	2744	exon10 ggccttctctggcaaccttt	T/G	gcctgacggacctgttgcc
RSC22-99	3' flanking region	404	acaaggaagtctctggccaca	G/A	gagtcctctactggctgcag
RSC22-100	3' flanking region	1416	agagctgagcctggagccccc	T/C	aggtggagtgtagcagggaga
RSC22-101	3' flanking region	1921	gggagcctcgaggatgtggc	C/T	cctcattttttatcccatct
RSC22-102	3' flanking region	3236	gagctctgggcctgctgtgat	T/C	gccgggtcccagtcagtcgaa

Nucleotide numbering is according to the mutation nomenclature (Dunnen and Antonarakis 2000)

Table 2. Characterization of one dinucleotide and seven insertion/deletion polymorphisms from the *DGCR2* gene locus

ID	location		variation (5' to 3')		
RSC22-103	5' flanking region	-2456	tttttctgcccctggctggg	C/del	tcccagggcagatctggggc
RSC22-104	intron5	114	cccgtaggggtgctgtgcag	GGGACAT/ins	ggagtgggtgagccccagtt
RSC22-105	intron5	395	aaggaaactatcaggcagat	(AG) ₁₄₋₃₃	tgagcatgcatgtgtgtgtg
RSC22-106	intron5	5433	gactgccttgtgcctggctt	AGAC/ins	aggcaggagctgagtgggtg
RSC22-107	intron7	4733	ccttggggcaggcagtaggg	G/del	tcctgcctgccgagattgtt
RSC22-108	intron7	5026	cctttcatgcctgtaccac	AC/ins	tcatgtcatctctgcctccc
RSC22-109	3' flanking region	80	ctcagtcctttccagaccagc	A/ins	ggggcagcacctgggagggc
RSC22-110	3' flanking region	3607	cactggaagaagctggcccc	C/del	tgctgtgggggcccaggcc

Nucleotide numbering is according to the mutation nomenclature (Dunnen and Antonarakis 2000)
del, deletion polymorphisms; ins, insertion polymorphisms

exon 10. The SNPs are distributed nonrandomly, spanning the entire *DGCR2* gene locus. The density of SNPs in the noncoding regions was found to be 50 times higher than that in the coding regions. Hence, it is suggested that *DGCR2* is under huge selection pressure to conserve the sequence during evolution. It is noted that 9 SNPs within the 5' flanking region (putative promoter region) may affect the quantity of the gene product. We also identified a total of eight other types of variations at *DGCR2* gene locus (Table 2).

We present here 102 SNPs in the human *DGCR2* region among 48 Japanese individuals. The SNP collection represents a useful resource for studies on susceptibility to common diseases mapped to this local segment on chromosome 22q11 and is adequate for experimental characterization of linkage disequilibrium and haplotype-based disease association.

Acknowledgments We thank Akemi Inose and Maki Takahashi for technical assistance.

References

Budarf ML, Collins J, Gong W, Roe B, Wang Z, Bailey LC, Sellinger B, Michaud D, Driscoll DA, Emanuel BS (1995) Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. *Nat Genet* 10:269–278
Cooper DN, Krawczak M, Antonarakis SE (1995) The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL,

Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease, 7th edn. McGraw-Hill, New York, pp 259–291
Demczuk S, Aledo R, Zucman J, Delattre O, Desmaze C, Dauphinot L, Jalbert P, Rouleau GA, Thomas G, Aurias A (1995) Cloning of a balanced translocation breakpoint in the DiGeorge syndrome critical region and isolation of a novel potential adhesion receptor gene in its vicinity. *Hum Mol Genet* 4:551–558
Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15:7–12
Gong W, Emanuel BS, Collins J, Kim DH, Wang Z, Chen F, Zhang G, Roe B, Budarf ML (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum Mol Genet* 5:789–800
Kajiwar K, Nagasawa H, Shimizu-Nishikawa K, Ookura T, Kimura M, Sugaya E (1996) Cloning of SEZ-12 encoding seizure-related and membrane-bound adhesion protein. *Biochem Biophys Res Commun* 222:144–148
Nickerson DA, Tobe VO, Taylor SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25:2745–2751
Ohnishi Y, Tanaka T, Yamada R, Suematsu K, Minami M, Fujii K, Hoki N, Kodama K, Nagata S, Hayashi T, Kinoshita N, Sato H, Sato H, Kuzuya T, Takeda H, Hori M, Nakamura Y (2000) Identification of 187 single nucleotide polymorphisms (SNPs) among 41 candidate genes for ischemic heart disease in the Japanese population. *Hum Genet* 106:288–292
Taylor C, Wadey R, O'Donnell H, Roberts C, Mattei MG, Kimber WL, Wynshaw-Boris A, Scambler PJ (1997) Cloning and mapping of murine *Dgcr2* and its homology to the *Sez-12* seizure-related protein. *Mamm Genome* 8:371–375
Wadey R, Daw S, Taylor C, Atif U, Kamath S, Halford S, O'Donnell H, Wilson D, Goodship J, Burn J, Scambler P (1995) Isolation of a gene encoding an integral membrane protein from the vicinity of a balanced translocation breakpoint associated with DiGeorge syndrome. *Hum Mol Genet* 4:1027–1033