# ORIGINAL ARTICLE

Shawkat Haider · Rie Matsumoto Nobuyuki Kurosawa · Keiko Wakui Yoshimitsu Fukushima · Masaharu Isobe

# Molecular characterization of a novel translocation t(5;14)(q21;q32) in a patient with congenital abnormalities

Received: 7 November 2005 / Accepted: 18 December 2005 / Published online: 24 February 2006 © The Japan Society of Human Genetics and Springer-Verlag 2006

Abstract Chromosomal translocations are frequently found to be associated with various malignant disorders as well as congenital abnormalities. We report the characterization of a novel reciprocal translocation t(5;14)(q21;q32) in a patient with congenital abnormalities manifested by severe mental retardation, athetotic tetraplegia, microcephaly, peculiar facies (upward slanting of palpebral fissures), clinodactyly of the fifth fingers, and overlapping toes. Using a JHGP24 lymphoblast cell line derived from this patient, metaphase fluorescence in situ hybridization with bacterial artificial chromosome and cosmid probes and subsequent molecular analysis mapped the translocation breakpoint to the nucleotide level. Sequence analysis of the breakpoint junctions revealed the presence of a homologous sequence, GTGGC, along with a single nucleotide substitution and an insertion in der(14), and a single nucleotide deletion in the der(5) chromosome. We also attempted to identify and characterize the transcripts near the breakpoint by 5' and 3' rapid amplification of cDNA ends. Although we found several transcripts near the breakpoint of chromosome 14, the lack of significant ORFs within these transcripts suggests they are likely to

S. Haider · R. Matsumoto · N. Kurosawa · M. Isobe (⊠) Laboratory of Molecular and Cellular Biology, Department of Materials and Biosystem Engineering, Faculty of Engineering, Toyama University, 3190 Gofuku, Toyama 930-8555, Japan E-mail: isobe@eng.toyama-u.ac.jp Tel.: +81-76-4456872 Fax: +81-76-4456874

K. Wakui · Y. Fukushima Department of Medical Genetics, Shinshu University School of Medicine, 3-1-1 Asahi, Matsumoto 390-8621, Japan be non-coding RNAs. These transcripts may have an important role in the neurogenesis or differentiation.

**Keywords** Reciprocal translocation · Congenital abnormalities · Non-coding RNA · Fluorescence in situ hybridization

### Introduction

Congenital abnormalities contribute to a significant proportion of infant morbidity and mortality, as well as fetal mortality, accounting for 30-50% of post-neonatal deaths (Hoekelman and Pless 1988). Such anomalies often have an underlying genetic basis, ranging from single genes playing a dominant or recessive role in Mendelian disorders (e.g., achondroplasia or Holt-Oram syndrome) to a combined contribution from multiple genes and environmental triggers in complex traits (Lidral and Murray 2004). Congenital malformations exhibit a wide spectrum of phenotypic manifestations and may occur as an isolated malformation or as part of a syndrome. They are individually rare, but due to their overall frequency and severity they are of clinical relevance. About 3-5% of all births result in congenital malformations (Robinson and Linden 1993), with genetic causes accounting for around 25% of that total (Czeizel 2005). Chromosome abnormalities have an incidence of 1 in 200 live births (Robinson and Linden 1993), and have been attributed as the cause of a large number of genetic diseases. Balanced rearrangements such as translocations and inversions occur in about 1 in 500 individuals, and unbalanced translocations occur in about 1 in 2.000.

Susceptibility to disease is frequently genetically determined, even for diseases that appear to have an explicit environmental etiology. Many hereditary and non-hereditary diseases are associated with chromosomal abnormalities. Chromosomal abnormalities, especially translocations, are a crucial etiological com-

The breakpoint sequences of der(5) and der(14) reported here are available in the GenBank database (accession numbers DQ269476 and DQ269477); sequence accession numbers of non-coding RNA genes BR514A and BR514B are DQ279400 and DQ304541, respectively.

336 ponent of many

ponent of many diseases, and are frequently involved in the pathogenesis of leukemias, lymphomas and sarcomas. The biological relevance of these translocations is underscored by the fact that a number of known and putative genes have been identified by the study of genes located at these breakpoints. For example, translocations have facilitated positional cloning of diseasecausing genes in patients with Duchenne muscular dystrophy (DMD) (Monaco et al. 1985), neurofibromatosis type 1 (NF1) (Fountain et al. 1989; Ledbetter et al. 1989), Wilms tumor/aniridia (Pritchard-Jones et al. 1990), and Alagille syndrome (Spinner et al. 1994). In addition, translocations have also provided an insight into mechanisms related to chromosomal rearrangement and genomic architecture (Kurahashi et al. 2000).

Here we report a patient with congenital anomalies manifested by multiple malformations, including severe mental retardation, athetotic tetraplegia, microcephaly, peculiar facies (upward slanting of palpebral fissures), clinodactyly of the fifth fingers, and overlapping toes. Cytogenetic analysis revealed the presence of a reciprocal translocation t(5;14)(q21;q32) as the sole chromosome abnormality in this patient. The translocation breakpoint was localized by means of fluorescence in situ hybridization (FISH) and, following molecular cloning, the breakpoint was further characterized to gain insights into the translocation mechanism, as well as to identify the gene associated with the disease. We could not find any ORF within the transcripts near the breakpoint. Interestingly, we found two evolutionarily conserved regions within these transcripts, which strongly suggests that they have an important developmental function.

### **Materials and methods**

### Patient and cell line

The JHGP24 cell line was derived from a 27-year-old woman with severe mental retardation, athetotic tetraplegia, microcephaly, peculiar facies (upward slanting of palpebral fissures), clinodactyly of the fifth fingers, and overlapping toes. This cell-line was provided by a research group of Mendelian cytogenetics granted from the Ministry of Education, Culture, Sports and Science.

## Fluorescence in situ hybridization

In situ hybridization of metaphase slides from the patient was performed with labeled probes. Probes were labeled with biotin by nick translation as described (Narducci et al. 1995). Briefly, 300 ng labeled probe was used for each experiment, and hybridization was performed at 37°C in a total volume of 10  $\mu$ l containing 2× SSC, 50% formamide, 10 mM Na<sub>3</sub>PO<sub>4</sub>, 10% dextran sulfate, 0.5  $\mu$ g Cot1 DNA (Invitrogen, Carlsbad, CA) and 20  $\mu$ g sonicated salmon sperm DNA and tRNA each. Post-hybridization washing was in 2× SSC containing 50% formamide at 37°C for 15 min followed by 2× SSC for 5 min at 37°C. Immunological detection of biotin-labeled DNA was performed with fluorescein isothiocyanate (FITC)-conjugated avidin (green signal). Band localization of the probe was carried out by fluorescence microscopy after counterstaining with DAPI (0.5  $\mu$ g/ $\mu$ l) in 2× SSC.

### Southern blot

Genomic DNA from patient cell line JHGP24 and from normal placenta was digested with the restriction enzyme *Eco*RI and Southern hybridization was performed as described (Southern 1975). Probe A and probe B, both *Hind*III–*Eco*RI fragments, correspond to UCSC genome browser nucleotide coordinates chr14:98453228-98453776 and chr14:98458069-98458636, respectively. Probes were labeled with <sup>32</sup>P-dCTP using a Megaprime DNA labeling system (Amersham Biosciences, Piscataway, NJ) according to the manufacturer's specifications.

# Bacterial artificial chromosome and cosmid clones isolation

Polymerase chain reaction (PCR) screening of the Keio bacterial artificial chromosome (BAC) library (Asakawa et al. 1997) with primers derived from STS marker D14S627 at 14q32.1 resulted in the identification of BAC clone 1343F05K (160 kb) spanning the breakpoint. DNA from BAC1343F05K was partially digested with *Sau*3AI and the resulting DNA fragments were ligated to the cosmid vector pMFG2 as described previously (Sugimoto et al. 1999). A contig was made from these clones by cross hybridizing them after suppressing repetitive sequences within each cosmid by using Cot I DNA.

### Phage library construction

A complete genomic library was constructed in phage vector EMBL3 from *Sau*3AI partially digested DNA from the patient cell line as described (Gauwerky et al. 1989); the 1 million phage clones thus obtained using *Escherichia coli* NM538 as host bacteria were further screened with probe A.

### Rapid amplification of cDNA ends

Rapid amplification of cDNA ends (RACE) was performed using the Marathon cDNA amplification kit (Clontech), according to the manufacturer's protocol. Specific cDNA templates were amplified in two reactions: the first amplification used primer AP1 bound to the adapter and a gene-specific primer GSP1 (derived from predicted exon or EST), followed by nested PCR with a similar set of primers AP2 and GSP2. AP1 and AP2 as well as the complete list of gene specific primers are available upon request. PCR products were analyzed by direct sequencing or sequencing after subcloning into pBluescript vector.

### Results

Isolation of BAC and cosmid clones and construction of a cosmid-contig spanning the breakpoint

Cytogenetic analysis with metaphase FISH of patient cell line JHGP24 revealed the apparently reciprocal translocation between the long arms of chromosomes 14 and 5. A set of BAC clones was isolated by PCR-based screening of the Keio BAC library using STS markers derived from the 14q32 region, and the localization of each clone on translocated chromosomes was determined by FISH. A single BAC clone, 1343F05K, thus identified showed a split signal on der(14) and der(5) chromosome 14. To further narrow the region of the breakpoint to a precise location, a contig of six cosmid clones (cos8, cos10, cos12, cos38, cos47, cos66) was constructed from BAC clone 1343F05K; a contig consisting of overlapping BAC clones and cosmids spanning

the breakpoint is shown in Fig. 1a. Upon FISH of patient cell line JHGP24, hybridization signal from each of these cosmids was found either distal or proximal to the breakpoint except for clone cos66, which showed a split signal from der(14) and der(5) (Fig. 1b). Thus, the cosmid clone cos66 spanning the breakpoint was identified.

Detection of rearrangement by Southern blot analysis

Southern hybridization of *Eco*RI-digested DNA from the JHGP24 cell line was carried out with probe A and probe B as shown (Fig. 2a) using placenta DNA as a control. While both probes generated a germline 5.5 kb band in normal and patient chromosomes, probes A and B detected rearranged 10 kb and 6.5 kb *Eco*RI bands, respectively, in the patient (Fig. 2b). Together with FISH results, Southern blot analysis showed that the translocation involved chromosomes 14 and 5, and that the breakpoint is located between probes A and B.

Cloning and sequence analysis of the breakpoint

A total of 1 million phage clones were constructed using DNA from the JHGP24 cell line for further screening with probe A. Nine positive clones, including

Fig. 1 a Physical map of the bacterial artificial chromosome (BAC)-cosmid contig spanning the 14q32 breakpoint. The relative size and location of the BAC and cosmid clones are shown, along with the ESTs and STS markers near the breakpoint. b Representative fluorescence in situ hybridization (FISH) and DAPI-stained G-band images obtained with metaphase spreads from patient cell line JHGP24 using the cosmid66 probe. Hybridization resulted in a split signal, and signal was detected from der(14) and der(5) as well as from normal chromosome 14





Fig. 2 a Restriction map of breakpoint region with respect to germline sequences of chromosomes 14 and 5. Probes A and B, both EcoRI–HindIII restriction fragments (*boxes*) detected rearranged EcoRI bands of 10 and 6.5 kb, respectively. Primers derived from both chromosomes near the breakpoint are indicated with arrows. Restriction sites: H HindIII, E EcoRI. **b** Southern hybridization of EcoRI-digested JHGP24 DNA using placental DNA as a control. The germline 5.5 kb band was detected in normal and patient with both probes; rearranged bands of 10 and 6.5 kb were detected in the patient with probes A and B, respectively. **c** Comparison of germline sequences from chromo-

both germline and rearranged chromosomes, were identified by plaque hybridization. Phage DNA from purified positive plaques was extracted using standard methods, and restriction map analysis revealed a single phage clone, p20-1-cl1, containing the breakpoint. A search of the NCBI database showed phage clone p20-1-cl1 to have 10.3 kb homology with the sequence of 14q32; the remaining nonhomologous sequence was found to be derived from chromosome 5. PCR was then performed with primers derived from the corre-

somes 14 and 5, with the breakpoint regions of der(14) and der(5). Germline chromosome 14 and 5 sequences are indicated with *capital* and *small letters*, respectively. The homologous sequence (GTGGC) is *boxed* and *capitalized*. In the der(14) chromosome, C was substituted by G (*asterisk*), and a T of unknown origin was inserted (+). The deletion of a single nucleotide A in der(5) is also indicated (-). **d** Transcripts derived from brain (BR514A) and leukocyte (BR514B) with respect to breakpoint of chromosome 14; mouse-human conserved between human and mouse (X, Y, and Z) are indicated (*asterisks*)

sponding chromosome 14 and 5 sequences (5'-ATG CCCTGTTCCCTGTGTTC-3' and 5'-CTCCCTGCC TTCCTTTCAAC-3') to isolate the reciprocal breakpoint. Comparative sequence analysis of germline chromosome 14 and 5 with the breakpoint region of derivative chromosomes der(14) and der(5) revealed a 5-bp homologous sequence, GTGGC, at the breakpoint junction (Fig. 2c), which suggested that the translocation was mediated by homologous recombination. The der(14) allele showed a C > G substitution

and a single T insertion while the der(5) had a single A deletion just 5' of GTGGC.

Searching for gene near the translocation breakpoint

The web-based program Repeat Masker (http:// www.ftp.genome.washington.edu/cgi-bin/RepeatMasker) was used to obtain repeat-free sequence of the BAC clone 1343F05K from chromosome 14. A database (NCBI-EST) search with these repeat-free sequences near the breakpoint region identified two ESTs-AV262687 and BB013354-6 kb telomeric and 3 kb centromeric to the breakpoint, respectively-both derived from Mus musculus testis, and conserved between human and mouse. However, we could not find any EST near the region of the chromosome 5 breakpoint. Using these EST sequences at the 14q32 breakpoint, we performed both 5' and 3' RACE to further extend the length of our transcripts. We obtained cDNA containing two exons from brain RNA using 5' RACE, but there was no significant ORF in this sequence. We also isolated cDNA from other tissues, namely testis, spleen and leukocyte; however, we found no ORF within any of these cDNAs. The relative size and location of the transcripts derived from brain and leukocyte as well as the conserved regions (X, Y, and Z) with respect to the breakpoint are shown in Fig. 2d.

### Discussion

A central goal of genome analysis is the comprehensive identification of all human genes along with their functions in pathophysiological processes. Chromosomal translocations are found in a number of congenital abnormalities and study of these translocation breakpoints greatly helps identify the causative gene as well as giving insight into the mechanism involved in the rearrangement. We have reported the characterization of a novel translocation t(5;14)(q21;q32) in a patient with multiple congenital anomalies. We mapped the site of breakpoint by FISH analysis of metaphase spreads from the patient cell line JHGP24, and the analysis of breakpoint junctions showed this rearrangement to have been mediated by homologous recombination. Our cytogenetic analysis found the reciprocal translocation t(5;14)(q21;q32) to be the sole chromosome abnormality in this patient. Multiple congenital malformation syndromes, however, are due not only to visible chromosomal imbalance but also to submicroscopic or subtelomeric rearrangements; spectral karyotyping and CGH array may help detect such rearrangements, if any.

Although we found several predicted exons near the breakpoint using the GrailExp program, no tissue expression was detected using RT-PCR. We performed RACE with primers derived from the conserved ESTs AV262687 and BB013354 near the breakpoint to isolate transcripts that mapped up to 15 kb proximal to the

breakpoint, spanning the region both telomeric and centromeric to it. Transcripts derived from brain (BR514A) and leukocyte (BR514B) are shown in Fig. 2d. We also isolated several transcripts within the breakpoint locus from testis, leukocyte, and spleen (data not shown) but they could not be connected through splicing between regions Y and Z, although this region is found connected in mouse testis. In mouse, the EST AK015510 was found to span all three conserved regions X, Y, and Z, suggesting that if such a transcript exists in human, perhaps expressed in specific tissue other than the tissues used in our study, the translocation would likely interrupt correct splicing and alter transcript function. Alternatively, translocation may affect the regulatory region of these transcripts, altering the level of expression. We repeatedly tried to extend the transcripts further upstream by 5' RACE, but achieved no extension. The absence of any significant ORF in these transcripts could be due to the partial characterization analyzing just the 3' UTR of a coding gene, although this seems unlikely. Our results strongly suggest these transcripts actually lack an ORF and hence they are most likely to be non-coding RNAs (ncRNAs), possibly a riboregulator, associated with gene translation. ncR-NAs, which produce functional RNA instead of encoding a protein, have been implicated in numerous biological processes including transcriptional regulation, chromosome replication, RNA processing and modification, messenger RNA stability and translation, and even protein degradation and translocation (Storz 2002). Although usually short in length, several extremely long ncRNAs detected in mammalian cells have been implicated in silencing genes and changing chromatin structure across large chromosomal regions (Erdmann et al. 2000; Sleutels et al. 2002; Le Meur et al. 2005). Examples include the human Xist RNA required for X chromosome inactivation, LNCAT in gene regulation and the mouse Air RNA required for autosomal gene imprinting. A number of ncRNAs have been reported to be linked to diseases ranging from malignancies to psychiatric illnesses and neuro-developmental disorders. These include: miR-15 and 16, BCMS (B cell chronic lymphocytic leukemia), TTY1 and 2 (gonadoblastoma), NCRMS (rhabdomyosarcoma), BIC (ALV-induced B cell lymphoma), H19 (breast/colon/bladder/Wilms' tumor), MALAT-1 (non-small cell lung cancer), DISC2 (schizophrenia), SCA8 (spinocerebellar ataxia). ST7OT1-4 (autism), IPW (Prader-Willi syndrome), LIT-1 (Beckwith Wiedemann syndrome), RMRP (cartilage hair hypoplasia), and UBE3A antisense (Angelman syndrome) (Pang et al. 2005). The ncRNAs near the breakpoint may thus play an important role in our patient phenotype.

As shown in Fig. 2d, using Harrplot software, comparative human-mouse sequence analysis of the100 kb region proximal to the breakpoint revealed only three conserved regions: two regions—19.7 kb (X) and 5.8 kb (Y)—telomeric to the breakpoint, and a region 2.7 kb (Z) centromeric to the breakpoint. Interestingly, two of these conserved regions (Y and Z) were found to be located within our characterized transcripts BR514A and BR514B. Such conservation of sequences, despite the fact that they lack an ORF, strongly suggests that they have an important developmental or regulatory function. These transcripts may have a small window of expression within a given tissue, regulating very specific responses. Further studies in the mouse system may provide some useful insights into their physiological functions.

Acknowledgements We thank Drs. Nobuyoshi Shimizu and Shuichi Asakawa of Department of Molecular Biology, Keio University School of Medicine for kindly providing us with the Keio BAC clones. We would also like to thank Ms. Takako Maeda for technical assistance. This research was partially supported by a grantin-aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan.

### References

- Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N (1997) Human BAC library: construction and rapid screening. Gene 19:69–79
- Czeizel AE (2005) Birth defects are preventable. Int J Med Sci 2:91– 92
- Erdmann VA, Szymanski M, Hochberg A, Groot N, Barciszewski J (2000) Non-coding, mRNA-like RNAs database Y2K. Nucleic Acids Res 28:197–200
- Fountain JW, Wallace MR, Bruce MA, Seizinger BR, Menon AG, Gusella JF, Michels VV, Schmidt MA, Dewald GW, Collins FS (1989) Physical mapping of a translocation breakpoint in neurofibromatosis. Science 244:1085–1087
- Gauwerky CE, Huebner K, Isobe M, Nowell PC, Croce CM (1989) Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. Proc Natl Acad Sci USA 86:8867–8871
- Hoekelman RA, Pless IB (1988) Decline in mortality among young Americans during the 20th century: prospects for reaching national mortality reduction goals for 1990. Pediatrics 82:582–595
- Kurahashi H, Shaikh TH, Hu P, Roe BA, Emanuel BS, Budarf ML (2000) Regions of genomic instability on 22q11 and 11q23 as

the etiology for the recurrent constitutional t(11;22). Hum Mol Genet 9:1665–1670

- Le Meur E, Watrin F, Landers M, Sturny R, Lalande M, Muscatelli F (2005) Dynamic developmental regulation of the large non-coding RNA associated with the mouse 7C imprinted chromosomal region. Dev Biol 286:587–600
- Ledbetter DH, Rich DC, O'Connell P, Leppert M, Carey JC (1989) Precise localization of NF1 to 17q11.2 by balanced translocation. Am J Hum Genet 44:20–24
- Lidral AC, Murray JC (2004) Genetic approaches to identify disease genes for birth defects with cleft lip/palate as a model. Birth Defects Res A Clin Mol Teratol 70:893–901
- Monaco AP, Bertelson CJ, Middlesworth W, Colletti CA, Aldridge J, Fischbeck KH, Bartlett R, Pericak-Vance MA, Roses AD, Kunkel LM (1985) Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. Nature 316:842–845
- Narducci MG, Virgilio L, Isobe M, Stoppacciaro A, Elli R, Fiorilli M, Carbonari M, Antonelli A, Chessa L, Croce CM, Russo G (1995) TCL1 oncogene activation in preleukemic T cells from a case of ataxia-telangiectasia. Blood 86:2358–2364
- Pang KC, Stephen S, Engstrom PG, Tajul-Arifin K, Chen W, Wahlestedt C, Lenhard B, Hayashizaki Y, Mattick JS (2005) RNAdb—a comprehensive mammalian noncoding RNA database. Nucleic Acids Res 33:D125–D130
- Pritchard-Jones K, Fleming S, Davidson D, Bickmore W, Porteous D, Gosden C, Bard J, Buckler A, Pelletier J, Housman D et al (1990) The candidate Wilms' tumour gene is involved in genitourinary development. Nature 346:194–197
- Robinson A, Linden MG (1993) Clinical genetics handbook, 2nd edn. Blackwell Scientific Publications, Oxford
- Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415:810–813
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517
- Spinner NB, Rand EB, Fortina P, Genin A, Taub R, Semeraro A, Piccoli DA (1994) Cytologically balanced t(2;20) in a twogeneration family with alagille syndrome: cytogenetic and molecular studies. Am J Hum Genet 55:238–243
- Storz G (2002) An expanding universe of noncoding RNAs. Science 296:1260–1263
- Sugimoto J, Hatakeyama T, Narducci MG, Russo G, Isobe M (1999) Identification of the TCL1/MTCP1-like 1 (TML1) gene from the region next to the TCL1 locus. Cancer Res 59:2313–2317