J Hum Genet (2005) 50:42–45
DOI 10.1007/s10038-004-0218-4

ORIGINAL ARTICLE

Aritoshi Iida · Kouichi Ozaki · Toshihiro Tanaka
Yusuke Nakamura

# Fine-scale SNP map of an 11-kb genomic region at 22q13.1 containing the *galectin-1* gene

Received: 26 October 2004 / Accepted: 1 November 2004 / Published online: 10 December 2004


**Abstract** Galectins, a family of animal lectins that bind β-galactoside sugar chains, are thought to have a variety of intra- and extracellular functions. Through a case-control study in the Japanese population and subsequent functional analyses, we previously showed that a functional single nucleotide polymorphism (SNP) in the gene encoding galectin-2 (lectin, galactoside-binding, soluble, 2; *LGALS2*) is associated with susceptibility to myocardial infarction. As an addition to the genetic information about *LGALS2* reported earlier, we provide here a map of polymorphic sites within an 11-kb region containing the gene encoding a closely related molecule, galectin-1 (lectin, galactoside-binding, soluble, 1; *LGALS1*). The map includes 14 SNPs and two genetic variations of other types detected in a Japanese population sample. Five of the 14 SNPs were not among those deposited in the dbSNP database in the US National Center for Biotechnology Information and appeared to be novel. We also analyzed linkage disequilibrium (LD) using the 12 SNPs in which minor-allele frequencies were $>0.20$. Investigation of haplotype structure within the *LGALS1* locus revealed five common haplotypes covering more than 95% of the test population. One, or a pair, of the SNPs described here might serve as a "tag" for detecting associations between complex diseases and genes in this local segment of chromosome 22q13.1.

**Keywords** Galectin-1 · Galectin-2 · Single nucleotide polymorphism (SNP) · Japanese population · SNP discovery · SNP map · Linkage disequilibrium (LD)

A. Iida (✉) · Y. Nakamura
Laboratory for Pharmacogenetics,
Research Group of Personalized Medicine,
RIKEN SNP Research Center,
c/o RIKEN Yokohama Institute,
1-7-22 Suehiro-cho, Tsurumi, Yokohama,
Kanagawa 230-0045, Japan
E-mail: iidaari@src.riken.go.jp
Tel.: +81-45-5039599
Fax: +81-45-5039326

K. Ozaki · T. Tanaka
Laboratory for Cardiovascular Diseases,
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

## Introduction

Galectins are a family of animal lectins that bind β-galactoside sugar chains and appear to have a variety of intra- and extracellular functions (Barondes et al. 1994; Hughes 1997; Perillo et al. 1998). Galectin-1 is a homodimer with a subunit molecular weight of 14.5 kDa; it is expressed in many tissues and in activated T cells and macrophages at inflammatory sites (Barondes et al. 1994; Blaser et al. 1998; Rabinovich et al. 1998). Involvement of galectin-1 in the regulation of cell adhesion, cell proliferation, and apoptosis has been revealed by in vitro studies involving various cell types including lymphocytes, thymocytes, or vascular cells (Perillo et al. 1998). Human galectin-1 has greater similarity to human galectin-2 than to any other galectins (43% amino-acid sequence identity; Cooper and Barondes 1999). The gene encoding galectin-1 (lectin, galactoside-binding, soluble, 1; *LGALS1*) lies approximately 150-kb distal to the lectin, galactoside-binding, soluble, 2 (*LGALS2*) locus on chromosome 22q13.1.

Myocardial infarction (MI), one of the leading causes of death in the world, is characterized by abrupt occlusion of coronary arteries that results in irreversible damage to cardiac muscle. A number of molecules are likely to be involved in the pathogenesis of MI, including some that mediate inflammation (Ross 1993; van der Wal et al. 1994). For example, we previously reported that functional single nucleotide polymorphisms (SNPs) in the lymphotoxin-α gene (*LTA*) are associated with susceptibility to MI (Ozaki et al. 2002). Furthermore, it

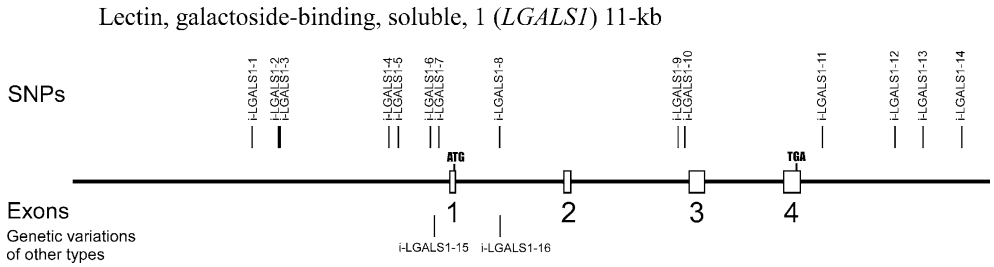


**Fig. 1** Genomic organization and locations of 16 genetic variations within the *LGALS1* locus. The four exons are represented by *rectangles*. SNPs are indicated above the "chromosome" line (designations correspond to the left-most column of Table 1). Locations of two genetic variations of other types are indicated below the map

was recently revealed that LTA protein physically interacts with galectin-2 (Ozaki et al. 2004). During a case-control study in the Japanese population and subsequent functional analyses, we found that one SNP in *LGALS2* was also associated with MI (Ozaki et al. 2004). In the course of that study, we also investigated SNPs in the *LGALS1* locus since galectin-1 is structurally related to galectin-2.

In this report, we provide a fine-scale map of the 11-kb region containing *LGALS1*, which includes 14 SNPs and two genetic variations of other types detected in a Japanese population sample.

## Subjects and methods

Specimens of peripheral blood were obtained with written informed consent from 128 healthy Japanese volunteers. Polymerase chain reaction (PCR) experiments and DNA sequencing were performed according to methods described previously (Iida et al. 2003). In brief, on the basis of the genomic sequence from the Genbank database (accession number z83844.5), we designed primer sets to amplify the entire *LGALS1* locus, excluding most regions corresponding to repetitive sequences. Each PCR experiment was performed using 20 ng of a mixture of genomic DNAs from two individuals. For genotyping, we used 5 ng of each genomic DNA. The samples were amplified in the GeneAmp PCR system 9700 (Applied

**Table 2** Distributions of genotypes and minor-allele frequencies of 14 SNPs in the *LGALS1* locus. *MAF* minor allele frequency, *HWE* Hardy–Weinberg equilibrium

| Gene symbol | Genotype distributions | | | | MAF | HWE test (*P*) |
|---|---|---|---|---|---|---|
| | 11 | 12 | 22 | Sum | | |
| i-LGALS1-1 | 37 | 42 | 17 | 96 | 0.396 | 0.403 |
| i-LGALS1-2 | 56 | 37 | 2 | 95 | 0.216 | 0.142 |
| i-LGALS1-3 | 89 | 5 | 0 | 94 | 0.027 | 0.791 |
| i-LGALS1-4 | 38 | 50 | 8 | 96 | 0.344 | 0.130 |
| i-LGALS1-5 | 37 | 42 | 17 | 96 | 0.396 | 0.403 |
| i-LGALS1-6 | 56 | 37 | 2 | 95 | 0.216 | 0.142 |
| i-LGALS1-7 | 37 | 42 | 17 | 96 | 0.396 | 0.403 |
| i-LGALS1-8 | 37 | 41 | 17 | 95 | 0.395 | 0.345 |
| i-LGALS1-9 | 36 | 46 | 14 | 96 | 0.385 | 0.911 |
| i-LGALS1-10 | 32 | 53 | 11 | 96 | 0.391 | 0.118 |
| i-LGALS1-11 | 38 | 43 | 15 | 96 | 0.380 | 0.627 |
| i-LGALS1-12 | 31 | 49 | 16 | 96 | 0.422 | 0.650 |
| i-LGALS1-13 | 78 | 17 | 1 | 96 | 0.099 | 0.945 |
| i-LGALS1-14 | 27 | 53 | 12 | 92 | 0.418 | 0.078 |

**Table 1** Characterization of 16 genetic variations in the *LGALS1* locus found among 64 Japanese chromosomes

| Gene symbol | Location | Position[a] | 5′ Flanking sequence[b] | Variation[c] | 3′ Flanking sequence[b] | dbSNP/report |
|---|---|---|---|---|---|---|
| i-LGALS1-1 | 5′ Flanking region | −2338 | agccacttggtctggaagga | G/A | gggtaaggcactcacagggg | |
| i-LGALS1-2 | 5′ Flanking region | −2021 | cctccaccaggttcctgagt | G/A | taaggtgagcacc(C/A)ctccca | rs13057866 |
| i-LGALS1-3 | 5′ Flanking region | −2007 | ctgagt(G/A)taaggtgagcacc | C/A | ctcccacccaccctcaattc | |
| i-LGALS1-4 | 5′ Flanking region | −722 | ccaccctgggccatctccta | G/A | gcttgacactctcaggcctg | rs4820293 |
| i-LGALS1-5 | 5′ Flanking region | −600 | agcctcaggcttggcgcccg | G/A | cccagcctttctttagcctt | rs4820294 |
| i-LGALS1-6 | 5′ Flanking region | −223 | gtggagagggatgaaggaga | A/C | taggaggtggagcttggact | rs2071769 |
| i-LGALS1-7 | 5′ Flanking region | −132 | ggtcacctctgctccaaaac | T/C | ggcttcaaaattccacggac | rs929039 |
| i-LGALS1-8 | Intron 1 | 514 | ggttgtgtccaggctggtgg | C/G | ((G)6–8)cgggggaaattcccttcca | |
| i-LGALS1-9 | Intron 2 | 1271 | agaggggcaggagcaggtgg | C/T | atggccagagctagaatcca | |
| i-LGALS1-10 | Intron 2 | 1362 | tactgagtgacagattagtc | A/G | gtcagtggggctggagctgg | rs9622682 |
| i-LGALS1-11 | 3′ Flanking region | 260 | ctgaaactgtttttagccta | G/A | tgggccctagcacacagcag | |
| i-LGALS1-12 | 3′ Flanking region | 1114 | cagatgcccccaccccaaca | G/A | tgggtttgggtgatgaccct | rs11914203 |
| i-LGALS1-13 | 3′ Flanking region | 1447 | ctgctggggcctgagtggac | G/A | gctagggtgccctgcagact | rs9610830 |
| i-LGALS1-14 | 3′ Flanking region | 1915 | gggctcccaggtggccaccc | G/A | gaaacagcttcagaagcctg | rs2899292 |
| i-LGALS1-15 | 5′ Flanking region | −174 to −173 | ggcggcggggtgaggggggg | G/ins | cagcagctcgccactctgat | |
| i-LGALS1-16 | Intron 1 | 515–522 | gttgtgtccaggctggtgg(C/G) | (G)6–8 | cgggggaaattcccttccac | |

[a]Nucleotide numbering is according to the mutation nomenclature (den Dunnen and Antonarakis 2000)
[b]Both 5′ and 3′ flanking sequences to each variation are denoted by small letters
[c]Variant nucleotides are indicated by capital letters

**Table 3** Pairwise linkage disequilibrium (LD) coefficients among 12 SNPs having minor-allele frequencies >0.20 in the test population

| | i-LGALS1-1 | i-LGALS1-2 | i-LGALS1-4 | i-LGALS1-5 | i-LGALS1-6 | i-LGALS1-7 | i-LGALS1-8 | i-LGALS1-9 | i-LGALS1-10 | i-LGALS1-11 | i-LGALS1-12 | i-LGALS1-14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i-LGALS1-1 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.93 | 1.00 | 0.93 | 0.68 |
| i-LGALS1-2 | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 |
| i-LGALS1-4 | | | 1.00 | 0.95 | 1.00 | 0.95 | 0.95 | 1.00 | 0.99 | 1.00 | 1.00 | 0.87 |
| i-LGALS1-5 | | | | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.93 | 1.00 | 0.93 | 0.68 |
| i-LGALS1-6 | | | | | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 |
| i-LGALS1-7 | | | | | | 1.00 | 1.00 | 0.98 | 0.93 | 1.00 | 0.93 | 0.68 |
| i-LGALS1-8 | | | | | | | 1.00 | 0.97 | 0.93 | 1.00 | 0.93 | 0.67 |
| i-LGALS1-9 | | | | | | | | 1.00 | 1.00 | 0.99 | 0.96 | 0.77 |
| i-LGALS1-10 | | | | | | | | | 1.00 | 1.00 | 0.76 | 0.92 |
| i-LGALS1-11 | | | | | | | | | | 1.00 | 0.97 | 0.73 |
| i-LGALS1-12 | | | | | | | | | | | 1.00 | 0.53 |
| i-LGALS1-14 | | | | | | | | | | | | 1.00 |

**Table 4** Haplotype structure of *LGAS1* in 96 Japanese samples

| Haplotype ID | Frequency | SNP ID | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | i-LGALS1-1 | i-LGALS1-2 | i-LGALS1-4 | i-LGALS1-5 | i-LGALS1-6 | i-LGALS1-7 | i-LGALS1-8 | i-LGALS1-9 | i-LGALS1-10 | i-LGALS1-11 | i-LGALS1-12 | i-LGALS1-14 |
| Haplotype A | 0.44 | G | G | A | G | A | T | C | C | G | G | G | A |
| Haplotype B | 0.23 | A | G | G | A | A | C | G | T | A | A | A | G |
| Haplotype C | 0.20 | G | A | G | G | C | T | C | C | A | G | G | G |
| Haplotype D | 0.05 | G | G | G | G | A | T | C | C | G | G | A | A |
| Haplotype E | 0.04 | A | G | G | A | A | C | G | T | A | A | A | A |

Biosystems, Foster City, CA, USA) under the following conditions: initial denaturation at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 2 min, and postextension at 72°C for 7 min. Products obtained from the PCR experiments served as templates for direct sequencing and for detection of SNPs using the fluorescent dye-terminator cycle-sequencing method. All SNPs detected by the Polyphred computer program were confirmed by sequencing both strands of each PCR product. Calculation of Hardy–Weinberg equilibrium LD coefficients (D′) have been described elsewhere (Yamada et al. 2001). Haplotype frequencies were estimated using SNPAlyze software (DYNACOM, Chiba, Japan).

## Results and discussion

We defined the exon–intron organization of *LGALS1* by comparing its cDNA with the genomic sequence. By direct sequencing of genomic DNA from 32 Japanese individuals, we explored SNPs in the 11-kb region containing *LGALS1* except for repetitive sequences. A total of 7.87-kb was screened in which we identified 14 SNPs distributed an average of 562 nucleotides apart. The exon organization of *LGALS1* and the locations of identified SNPs are illustrated schematically in Fig. 1, and detailed information for each of the genetic variations is provided in Table 1. After comparing our data with the dbSNP database in the National Center for Biotechnology Information, USA, we considered five of the 14 SNPs to be novel as of the middle of October 2004. Subregional distributions of SNPs were as follows: seven in the 5′ flanking region, three in introns, and four in the 3′ flanking region. The overall frequencies of nucleotide substitutions were counted as 64.3% for A/G, 14.3% for A/C, 14.3% for C/T, and 7.1% for C/G; transitions occurred 3.7 times more frequently than transversions. We also found one insertion–deletion polymorphism in the 5′ flanking region and one mononucleotide polymorphism in intron 1.

Since selection of informative SNPs for association studies depends on allele frequencies and on linkage disequilibrium (LD) among them, we genotyped an additional 96 individuals from the general Japanese population for all SNPs discovered in the *LGALS1* region. The distributions of genotypes and minor-allele frequencies for all 14 SNPs are summarized in Table 2. All genotype distributions were in Hardy–Weinberg equilibrium ($P > 0.05$). We selected for LD analyses the 12 SNPs in which minor-allele frequencies in the test population were above 0.2 and found that all of them were in strong LD with each other (Table 3). Using the same 12 SNPs, we constructed haplotypes in this locus and found that five common haplotypes covered more than 95% of the test population (Table 4).

Altogether, we constructed a fine-scale map of the 11-kb region containing *LGALS1*, showing locations of all 16 genetic variations found among the Japanese volunteers. One or a pair of the SNPs described here might serve as a "tag" for detecting associations between complex diseases and genes in this local segment of chromosome 22q13.1.

## References


Barondes SH, Cooper DN, Gitt MA, Leffler H (1994) Galectins. Structure and function of a large family of animal lectins. J Biol Chem 269:20807–20810

Blaser C, Kaufmann M, Muller C, Zimmermann C, Wells V, Mallucci L, Pircher H (1998) Beta-galactoside-binding protein secreted by activated T cells inhibits antigen-induced proliferation of T cells. Eur J Immunol 28:2311–2319

Cooper DN, Barondes SH (1999) God must love galectins; he made so many of them. Glycobiology 9:979–984

den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12

Hughes RC (1997) The galectin family of mammalian carbohydrate-binding molecules. Biochem Soc Trans 25:1194–1198

Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2003) Catalog of 668 SNPs detected among 31 genes encoding potential drug targets on the cell surface. J Hum Genet 48:23–46

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654

Ozaki K, Inoue K, Sato H, Iida A, Ohnishi Y, Sekine A, Sato H, Odashiro K, Nobuyoshi M, Hori M, Nakamura Y, Tanaka T (2004) Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. Nature 429:72–75

Perillo NL, Marcus ME, Baum LG (1998) Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death. J Mol Med 76:402–412

Rabinovich GA, Iglesias MM, Modesti NM, Castagna LF, Wolfenstein-Todel C, Riera CM, Sotomayor CE (1998) Activated rat macrophages produce a galectin-1-like protein that induces apoptosis of T cells: biochemical and functional characterization. J Immunol 160:4831–4840

Ross R (1993) The pathogenesis of atherosclerosis: a perspective for the 1990s. Nature 362:801–809

van der Wal AC, Becker AE, van der Loos CM, Das PK (1994) Site of intimal rupture or erosion of thrombosed coronary atherosclerotic plaques is characterized by an inflammatory process irrespective of the dominant plaque morphology. Circulation 89:36–44

Yamada R, Tanaka T, Unoki M, Nagai T, Sawada T, Ohnishi Y, Tsunoda T, Yukioka M, Maeda A, Suzuki K, Tateishi H, Ochi T, Nakamura Y, Yamamoto K (2001) Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. Am J Hum Genet 68:674–685