Allele-specific transcript quantification detects haplotypic variation in the levels of the SDF-1 transcripts†

Ryosuke Kimura1,2, Tomoki Nishioka2, Augustinus Soemantri3 and Takafumi Ishida2,*

1Department of Human Genetics, Graduate School of Medicine and 2Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan and 3Department of Pediatrics, Faculty of Medicine, Diponegoro University, Semarang, Indonesia

Received February 2, 2005; Revised April 2, 2005; Accepted April 14, 2005

It has been suggested that SDF1-G801A, a single nucleotide polymorphism (SNP) in the 3′ untranslated region (UTR) of the SDF1 gene, is associated with susceptibility to diseases such as AIDS and type-I diabetes. However, experimental studies examining the effect of SDF1-G801A on SDF-1 expression have not supported its functional importance. In this study, to examine whether other polymorphisms have a cis-acting effect on SDF1 expression, we carried out haplotype analyses of the SDF1 gene and the allele-specific transcript quantification utilizing Epstein–Barr virus-transformed lymphoblastoid cell lines with heterozygous genotype for SDF1-G801A. Haplotype-based analyses on the proportion of the allele-specific transcripts revealed the presence of haplotypes associated with a decreased amount of the transcripts. In addition, we observed haplotypic variation in response to dibutyl cyclic AMP and tetradecanoyl phorbol acetate that enhances the levels of SDF-1 transcripts probably through activation of transcription factors. Showing evidence that polymorphisms other than the SDF1-G801A have a cis-acting effect on expression of SDF-1 transcripts, the results of this study contribute to the interpretation of previous disease-association studies and to the selection of SNP markers for future studies. As shown in this study, allele-specific transcript quantification coupled with haplotype analyses can be an effective tool for detecting cis-acting polymorphisms in expressional regulation.

INTRODUCTION

Stromal cell-derived factor-1 (SDF1/CXCL12) is a chemokine that plays an important role in the regulation of migration, proliferation and differentiation of hematopoietic cells (1–4). There are at least two isoforms of SDF-1, SDF-1α and -β, that result from alternative splicing (5). The physiological receptor of SDF-1 is CXCR4, a membrane protein known to serve as a co-receptor for the entry of T-cell-tropic HIV-1 into CD4+ lymphocytes (6). It has been demonstrated that SDF-1 inhibits infection of HIV-1 into target cells, although it enhances HIV-1 replication through intracellular signals (7–9).

As the functions of SDF-1 are thought to be important for the pathogenesis of certain diseases, several studies have been carried out to examine an association between disease susceptibility and a single nucleotide polymorphism (SNP) located at nucleotide position (np) 801 in the 3′ untranslated region (3′-UTR) of the SDF-1β transcript, SDF1-G801A (SDF1-3′G and -3′A alleles). In these studies, it has been suggested that SDF1-3′A is associated with delayed progression of AIDS and early onset of type-I diabetes (10,11). However, an in vitro transfection study did not support the functional importance of SDF1-G801A in expression of SDF-1 transcripts and proteins (12). Our previous study using Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) did not show evidence for the allelic difference in the expression of the SDF1 gene as well (13). There still remains the possibility of the presence of other functionally important polymorphisms in linkage disequilibrium (LD) with SDF1-G801A.

*To whom correspondence should be addressed at: Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Tel: +81 385414633; Fax: +81 338187547; Email: tishida@biol.s.u-tokyo.ac.jp
†This study constitutes part of a doctoral dissertation (R. Kimura) submitted to the University of Tokyo.

© The Author 2005. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oupjournals.org
To detect an effect of polymorphisms on synthesis of gene products, specimens from individuals with different genotypes are usually compared in the amount of the gene products such as the secreted protein or intracellular protein or mRNA. Despite their simplicity, these methods have a disadvantage that individual backgrounds confound the levels of gene products and reduce detectability of the effect of polymorphisms. To solve this problem, it is effective to employ allele-specific transcript quantification (14–16), in which results are not affected by trans-acting factors as either of the alleles can serve as an internal control for the other allele in each individual sample. Although the method requires a polymorphism in exons as a tag for measuring the relative amount of the allele-specific transcripts, it can be applied for investigating the effect of polymorphisms located outside exons when haplotype information is available. To examine the presence of polymorphisms cis-acting on SDF1 expression, allele-specific transcript quantification was carried out utilizing SDF1-G801A as a tag for the allele-specific transcripts. We used a number of EBV-transformed LCLs derived from Indonesian islanders (Sulawesi, Flores and Timor) with a high heterozygosity at SDF1-G801A (17), which is appropriate to this method.

RESULTS

Haplotypes and haplotype frequencies

Seven polymorphisms were detected by sequencing analysis of the SDF1 5′ flanking region, and an unpublished SNP C-668G was newly identified in Indonesian individuals. From the polymorphisms detected and published in databases, we chose 14 SNPs and an insertion–deletion polymorphism (indel) for genotyping of 105 individuals from the Timorese Islanders. According to haplotype frequencies and LD coefficients \((D'\) and \(r^2\)) estimated from the genotype data (Fig. 1), 10 haplotype-tagging SNPs (htSNPs) for major haplotypes were determined. Then, we genotyped an additional 243
proportion of Msp\(^{+}\) (mean SDF1) in the three populations are shown in Figure 2. Here, np 801 in the SDF-1 gene including the intron sequences. The 3’-A-bearing haplotypes were named A1–A4 and the 3’-G-bearing haplotypes were named G1–G9.

### Haplotypic variation in SDF1 transcription

Sixty EBV-transformed LCLs with the SDF1-3’-G/3’-A heterozygous genotype were subjected to allele-specific transcript quantification with RT–PCR and MspI digestion that recognizes the sequence of the SDF1-3’-G. In the PCR products of the DNA samples derived from the cell lines (n = 60), the proportion of MspI-digested fragments was \(0.255 \pm 0.012\) (mean \(\pm SD\)) (Fig. 3). When standard mixtures of 3’G/3’G and 3’A/3’A DNA samples (n = 2 each) were examined, the proportion of MspI-digested fragments was quadratically increased depending on proportion of SDF1-3’G, which suggests that heteroduplexes undigested by MspI were randomly formed during later cycles of the PCR. Therefore, the proportion of SDF1-3’G was theoretically calculated regarding the proportion in the heterozygous DNA samples as 0.5 and considering allelic difference in PCR efficiency and formation of heteroduplexes undigested by MspI. As the result of the calculation, the proportion of SDF1-3’G in the heterozygous DNA samples was \(0.500 \pm 0.012\) (mean \(\pm SD\)).

To examine allelic imbalances in transcript production, we performed analyses based on the SDF1 haplotypes. In stepwise multiple regression analyses, it was found that a high proportion of 3’G transcripts was significantly associated with haplotypes A2 (\(P = 0.041\)) and G1 (\(P = 0.0044\)). Among the cells not carrying haplotype G1 (G1-negative), the cells carrying haplotype A2 (A2-positive) showed a significantly higher proportion of 3’G transcripts than did the A2-negative cells (\(P = 0.0090\), Mann–Whitney U-test) (Fig. 4), which means that A2 with \(-668G\) was a lower expressing haplotype when compared with the other 3’A-bearing haplotypes (Fig. 5). In addition, the proportion of 3’A transcripts was significantly lower than 0.5 in the G1-negative/A2-positive cells (\(P = 0.034\), t-test). Among the A2-negative cells, likewise, the G1-positive cells had a significantly higher proportion of 3’G transcripts than did the G1-negative cells (\(P = 0.0065\), Mann–Whitney U-test) (Fig. 4). Only two cell lines were A2-and-G1-double-positive and this genotype did not show any significant difference from the others. In the A2-and-G1-double-negative cells, the expression of 3’G transcripts was significantly lower than the expression of 3’A transcripts (\(P = 0.00011\), t-test). The 3’G-bearing haplotypes other than G1 were segregated by A6201G from the other haplotypes (Fig. 5). The haplotype bearing \(-668G\) (A2) expressed 0.63 times and the haplotypes bearing 6201G expressed 0.76 times of the transcripts when compared with the haplotypes bearing \(-668C\) and 6201A.

### Enhancement of the SDF-1 mRNA by dbcAMP and TPA

The EBV-transformed LCLs were stimulated with dibutylyl cyclic AMP (dbcAMP) or tetradeacnyl phorbol acetate (TPA) treatment. Twenty-four hours after the treatment, the expression of SDF-1 transcripts was significantly increased with large individual variation (Fig. 6). In the dbcAMP-treated cells, the levels of SDF-1 transcripts were significantly increased and this genotype did not show any significant difference from the others. In the A2-and-G1-double-negative cells, the expression of 3’G transcripts was significantly lower than the expression of 3’A transcripts (\(P = 0.00011\), t-test). The 3’G-bearing haplotypes other than G1 were segregated by A6201G from the other haplotypes (Fig. 5). The haplotype bearing \(-668G\) (A2) expressed 0.63 times and the haplotypes bearing 6201G expressed 0.76 times of the transcripts when compared with the haplotypes bearing \(-668C\) and 6201A.

![Figure 2. SDF1 haplotypes estimated from the genotype data for the 10 hSNPs in the three Indonesian populations.](image)

![Figure 3. Standard curve for the proportion of SDF1-3’G. Gray dots, 3’G/3’A heterozygous DNA samples (n = 60); black bar, mean ± SD; black dots, standard mixtures of homozygous DNA samples (n = 2 each); white squares, theoretical values taking the allelic difference in PCR efficiency into consideration; gray squares, theoretical values taking heteroduplex formation as well as the different PCR efficiency into consideration.](image)
transcripts than the untreated cells \( (P = 0.028, t\)-test), but the dbcAMP-treated cells did not \( (P = 0.065, t\)-test) (Fig. 7B). However, when compared with the A2-negative cells, the increase in 3’A contribution reached a significant difference in dbcAMP treatment \( (P = 0.043, \text{Mann–Whitney } U\)-test), but not in TPA treatment \( (P = 0.14, \text{Mann–Whitney } U\)-test). Neither treatment significantly changed the proportion of the allele-specific transcripts in cells without haplotypes G1 and A2.

**DISCUSSION**

This study demonstrates the presence of haplotypic variation in the expression of SDF-1 transcripts. For haplotype A2, we may need to consider type I error because the number of the samples was limited; nonetheless, we emphasize that advantages of the allele-specific transcript quantification are high detectability of the allelic difference because of the elimination of trans-acting factors and high applicability to rare haplotypes because of the use of heterozygous samples. Indeed, we have previously exhibited a clear difference in expression of IL-1β transcripts between −31T and C alleles using a small number of cell lines (16).

Minimum mutation networks of the SDF1 gene (Fig. 5), showing a clusterization of the 3’G-bearing haplotypes other than G1, indicate that A6201G and C − 668G are good markers to identify the low-expressing haplotypes. However, A6201G and C − 668G themselves are not necessarily responsible for the altered levels of the transcripts. Because most of the examined polymorphisms in the SDF1 gene region are in complete LD \( (|D'| = 1) \) (Fig. 1), other functionally important polymorphisms that are in absolute LD \( (r^2 = 1) \) with A6201G and C − 668G may exist elsewhere in this region. To verify which polymorphisms are substantially important, it is indispensable to perform further haplotype analyses throughout the gene and experiments that investigate effects of candidate polymorphisms on DNA–protein interaction, alternative splicing and mRNA stability.

DbcAMP is a cell-permeable analog of an intracellular second messenger, cAMP; TPA is a potent activator of an important signaling enzyme, protein kinase C (PKC). As this study showed that dbcAMP and TPA stimulations enhance SDF1 transcription in the EBV-transformed LCLs (Fig. 6), it is suggested that cAMP- and PKC-mediated intracellular signals are involved in SDF1 gene regulation. As the large individual variation in the degree of the enhancement cannot be explained only by the haplotypic variation, epigenetic conditions must be related. Moreover, the haplotypic variation in response to dbcAMP or TPA treatment (Fig. 7) may be attributed to the polymorphisms in the binding sites of the transcription factors. Searching for potential binding sites of
transcription factors with a database, MatInspector V2.2 (18), we found that \(-668C\) and \(-1652C\) constitute the consensus sequence of a potential binding site for activator protein (AP)-1 and AP-2, respectively. Because EBV proteins can activate transcription from AP-1 binding sites (19–21), the low levels of haplotype A2 transcripts (Fig. 4) may result from the disruption of the AP-1 binding site. Then, activation of several transcription factors by dbcAMP and TPA stimulations may reduce the relative contribution of AP-1 and result in increasing the proportion of haplotype A2 transcripts (Fig. 7B). The slight but certain decrease in the proportion of haplotype G1 transcripts by dbcAMP treatment (Fig. 7A) may be caused by the loss of the AP-2 binding site as cAMP strongly enhances AP-2 activity through cAMP-dependent protein kinase A (22–24).

A genetic epidemiological study has shown a protective effect of \(\text{SDF1}-3'\) homozygosity on AIDS progression (10), but later studies have disagreed with this preceding study (25–27). In this study, we demonstrate that other polymorphisms in LD with the \(\text{SDF1}-G801A\) (\(\text{G12197A}\)), rather than \(\text{SDF1}-G801A\) itself, are responsible for the altered levels of transcripts. Therefore, the discrepancy among the previous epidemiological studies may be attributed to the haplotype structures and frequencies in the subjected populations; the \(\text{SDF1}-3'\) may show a strong LD with functionally important polymorphisms in certain populations, but may not in other populations. Therefore, the \(\text{SDF1}\) haplotypes in a variety of populations need to be investigated in detail and disease-association studies need to be performed on the basis of haplotype information.

As technologies for genome analyses have developed, evidence for an association between genetic polymorphisms and susceptibility to diseases has increased. However, studies on a direct effect of polymorphisms on expression are left behind when compared with a great number of association studies. As shown in this study, the allele-specific transcript quantification coupled with haplotype analyses can be an effective tool for detecting polymorphisms \(\text{cis}\)-acting on the transcript levels. This method enables us to select candidates of functionally important polymorphisms in expression regulation without efficient genetic epidemiological studies. Such a strategic approach should contribute to the investigation of the functional implication of human genome diversity.

**MATERIALS AND METHODS**

**DNA samples**

We used DNA samples extracted from 348 anonymized EBV-transformed LCLs that are derived from healthy, unrelated volunteers of three Indonesian islands (Sulawesi, Flores and Timor).

**Sequencing**

For screening of genetic variations, we sequenced the 5' flanking region of the \(\text{SDF1}\) gene using PCR-direct sequencing. Twelve DNA samples from the Indonesian populations were subjected to PCR using a primer set, \(21934F\) and \(230R\) (Table 1), that amplifies \(2\,\text{kb}\) of this region. The PCR products were then used as templates for the nested PCRs using six primer sets, \(21843F\) and \(21608R\), \(21649F\) and \(21226R\), \(21302F\) and \(2860R\), \(21018F\) and \(2598R\), \(2755F\) and \(2205R\) and \(2313F\) and 91R. These products were sequenced by using ABI genetic analyzer 310 (Applied Biosystems, Foster City, CA, USA).

**Haplotyping**

Using PCR–single strand conformation polymorphism (SSCP) and PCR–restriction fragment length polymorphism (RFLP) assays, 105 Timorese samples were genotyped for the following 15 polymorphisms (Table 2): \(G-1882G\) (\(\text{rs2839682}\)), \(G-1750A\) (\(\text{rs2839683}\)), \(G-1748A\) (\(\text{rs2839684}\)), \(C-1652T\) (\(\text{rs11239027}\)), \(T-1185C\) (\(\text{rs2839685}\)), \(C-668G\), \(T-504G\)
Reverse primers

SDF1

Sixty EBV-transformed LCLs heterozygous for the (rs1801157) and C14246G (rs266093). Then, 10 hSNPs were (rs2236533), 6461–6462 indel (rs2839694), G12197A C6145A (rs2236534), G6201A (rs266085), C6238T (rs17156283), T5587C (rs2839692), G5887A (rs2839693), 5' flanking 5' flanking 5' flanking

Forward primers

−1934F AGCCGATGGCCTTTGCTT
−1843F TACCAGGCTGTTAGATTTCC
−1649F CGACTGCTCCTCTCACTGG
−1302F GAACGGGCGACGTAACC
−1018F GGTTGCGGAGTTCTTCTAG
−755F AGCCACCAGTCAGTCAAGAA
−313F CGACAGCTGAAAGGTCAAA
5419F GCACCCCTCTCCACTAAATGG
6055F CCGAAAGACTCTCTTCATCACAT
6322F CGCAGACTCCTAAGGATCC
11957F GCCTCTCTCGAATCCAGG
14023F CCTTGGCCTCCTGCTATACCT

Re reverse primers

−1737R GACGCCCCCTCTGCTG
−1608R ACTGACCTTGTGGCTGCTTTTG
−1226R CTTTGCGCTGCTTTTG
−860R GTGGAGGGATGATTCAAG
−596R GTGCGTGGGGGCTCAAG
−391R CTGAGGCGCCCTATTTCT
−205R GTGGCCGCGTTCTCTGAC
91R CGGAGCCCGAGCCTC
230R ACCCTCGGCGAAAATAAGT
5957R CGCAAGTTAAGAAATCGGAG
6316R CTATGACTGGCCGATCTG
6527R GACGCTTACATCTTTCAACACT
12243R CCTTGGCTCTCAACTCTG
12351R GGGCTAGAGAAGCTGTT
14477R TGGGGCTCCTTCAGAATGAG

(rs17156283), T5587C (rs2839692), G5887A (rs2839693), C6145A (rs2236534), G6201A (rs266085), C6238T (rs2236533), 6461–6462 indel (rs2839694), G12197A (rs1801157) and C14246G (rs266093). Then, 10 hSNPs were selected and used for haplotyping of 151 Sualanesian and 92 Florenian samples.

Cell culture and RNA extraction

Sixty EBV-transformed LCLs heterozygous for the SDF1-G801A (G12197A) were selected from our cell bank for this experiment. These cell lines carry haplotypes of which frequencies are >1% (A1–A3 and G1–G7). Cells were cultured in RPMI1640 medium supplemented with 10% fetal calf serum and antibiotics. Each cell line was untreated or treated with 1 mM dbcAMP (Sigma, St Louis, MO, USA) or 20 ng/ml TPA (Sigma). Twenty-four hours after the treatment, we extracted total RNA from 5 × 10^6 cells using GenElute Mammalian Total RNA Kit (Sigma).

SDF1 mRNA quantification

To examine an effect of dbcAMP or TPA treatment on the amount of SDF1 transcripts, we quantified the relative amount of the SDF1α, SDF1β and β-actin mRNA in 12 cell lines by means of real-time PCR using Light Cycler (Roche, Indianapolis, IN, USA) as described elsewhere (13). The levels of SDF1 transcripts were standardized with the levels of β-actin transcripts.

<table>
<thead>
<tr>
<th>PCR primer set</th>
<th>Method: polymorphism</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>−1934F and −1737R</td>
<td>SSCP: −1882G/C, −1750G/A, −1748G/A</td>
<td>5' flanking</td>
</tr>
<tr>
<td>−1843F and −1608R</td>
<td>SSCP: −1750G/A, −1750G/A, −1652C/T</td>
<td>5' flanking</td>
</tr>
<tr>
<td>−1302F and −860R</td>
<td>RFLP: −115T/C (HpyCH4IV)</td>
<td>5' flanking</td>
</tr>
<tr>
<td>−755F and −391R</td>
<td>RFLP: −668G/C, −59T/G (HpyCH4IV)</td>
<td>5' flanking</td>
</tr>
<tr>
<td>5419F and 5957R</td>
<td>RFLP: 5887T/C (Snd), 5887G/A (Rad)</td>
<td>Intron 2</td>
</tr>
<tr>
<td>6055F and 6316R</td>
<td>SSCP: 6145C/A, 6201G/A, 6238C/T</td>
<td>Intron 2</td>
</tr>
<tr>
<td>6322F and 6527R</td>
<td>SSCP: 6461–6462 indel</td>
<td>α mRNA 3'-UTR</td>
</tr>
<tr>
<td>11957F and 12351R</td>
<td>RFLP: 12197G/A (MspI)</td>
<td>β mRNA 3'-UTR</td>
</tr>
<tr>
<td>14023F and 14477R</td>
<td>RFLP: 14246G/C (HindIII)</td>
<td>β mRNA 3'-UTR</td>
</tr>
</tbody>
</table>

Nucleotide positions of the SNPs are counted from the start codon of the SDF1 gene based on NC_000010.

Quantification of the allele-specific transcripts

The RNA samples from the 60 cell lines were subjected to RT–PCR using SuperScript One-Step RT–PCR with Platinum Taq (Invitrogen, Carlsbad, CA, USA). The selected primer set, 6322F and 12351R (Table 1), spans an intron so as not to amplify the genomic DNA sequence. The RT–PCR was duplicated for each RNA sample. Then, we performed the second nested PCR using 11957F labeled with FITC and 12243R, both of which are located in exon 4.

To prepare standards, we used DNA samples with the 3'G/3'T (n = 1), 3'A/3'A (n = 1) and 3'G/3'A (n = 60) genotypes as templates for the PCR with 11957F and 12351R primers. The PCR products from the 3'G/3'G and the 3'A/3'A homozygotes were mixed at ratios ranging from 1:9 to 9:1. The preparation of the standard mixtures was duplicated. Then, the semi-nested PCR using the FITC-labeled primer set was performed.

These FITC-labeled PCR products from the samples and the standards were treated with MspI, a restriction nuclease that digests the SDF1-3'G sequence, and separated using 10% polyacrylamide gel electrophoresis. Fluorescence intensities of the digested (239 bp) and the undigested (304 bp) fragments were measured with LAS-1000plus image analyzer (Fuji film, Tokyo, Japan). The standards were compared with theoretical values considering only allelic difference in PCR efficiency [I_{239}(I_{239} + P_{239}) = P_{3'G}(P_{3'G} + kP_{3'A})], where I_{239} and I_{304} are, respectively, the intensities of the 239 and 304 bp fragments, and P_{3'G} and P_{3'A} are, respectively, the proportion of the 3'G and 3'A before the PCR and k is the ratio of PCR amplification efficiency, or values additionally considering formation of heteroduplexes undigested by MspI [I_{239} - I_{304}] = (P_{3'G}(P_{3'G} + kP_{3'A})^2]. Finally, proportion of the allele-specific transcripts was calculated by the use of the latter model.

Statistics

Haplotype phase and frequencies were estimated by using a software PHASE version 2.0 (28). Transforming haplotypes
to dummy variables, we used the stepwise multiple regression to detect the significant effect of haplotypes on the proportion of the allele-specific transcripts. To compare the proportion of the allele-specific transcripts and its alteration by dbcAMP and TPA stimulations between two groups, the Mann–Whitney U-test was used. The t-test was performed to examine the deviation from a 1:1 ratio of the allele-specific transcripts. Effects of dbcAMP and TPA stimulations on the SDF-1 mRNA levels and on the proportion of the allele-specific transcripts were also examined with the t-test.

Conflict of Interest statement. None declared.

REFERENCES


