Hezhao Ji · Terry Blake Ball · Joshua Kimani
Francis Allan Plummer

Novel interferon regulatory factor-1 polymorphisms in a Kenyan population revealed by complete gene sequencing

Received: 28 May 2004 / Accepted: 2 July 2004 / Published online: 17 September 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract Variation in susceptibility to HIV-1 infection depends on numerous factors, and host genetic variation has been well-described as an important component. As a transcriptional regulator, interferon regulatory factor 1 (IRF-1) plays a key role in both innate and adaptive immunity against viral infection. IRF-1 has also been shown to directly interact with HIV-1 5' LTR and efficiently initiate or amplify HIV-1 replication. By complete gene sequencing, we investigated genetic polymorphism of the IRF-1 gene in an HIV-1-endemic Kenyan population. This population displayed extensive genetic diversity at the IRF-1 locus. Fifty-three single nucleotide polymorphisms (SNPs) were identified in this population, including 26 novel SNPs. Two insertion and one deletion polymorphisms in IRF-1 were also identified. Linkage disequilibrium (LD) among these genetic variations was shown to be common in IRF-1. The functional consequences of these mutations in the context of HIV-1/AIDS remain to be determined. We also identified 35 consistent discrepancies between IRF-1 GenBank sequences and our population based sequencing data, suggesting that the previously submit-

ted GenBank data were not representative of the majority of human IRF-1 sequences.

Keywords Interferon regulatory factor 1 · Polymorphism · Kenya · Sequencing

Introduction

Variation in susceptibility to HIV-1 infection depends on numerous factors, and host genetic variation has been well-described as an important component. Genetic variation in HIV-1 coreceptors and human leukocyte antigen genes are among the best-described host factors shown to be associated with resistance or susceptibility to HIV-1 infection (Anzala et al. 1998; Dunand et al. 1997; MacDonald et al. 2000). Little is known about the role of polymorphisms in immune regulatory genes in susceptibility to HIV-1 infection. Interferon regulatory factor 1 (IRF-1), the first-identified member of the IRF family, was originally defined as a transcription regulator of IFN- β (Fujita et al. 1989, 1988). Later studies showed that IRF-1 function was not restricted to the IFN pathway but, rather, by binding an interferon-stimulated response element (ISRE) present in the promoters of IFN-stimulated genes (ISG), IRF-1 could regulate multiple immune genes' expression and is broadly involved in both innate and adaptive immunity, especially in the regulation of Th1/Th2 differentiation, which may affect resistance to HIV-1 infection (Kroger et al. 2002; Lohoff et al. 1997; Mamane et al. 1999; Taki et al. 1997; Taniguchi et al. 2001; Trivedi et al. 2001). The human IRF-1 gene, spanning 7.72 Kb with its 495 bp immediate upstream promoter sequence, has been assigned to 5q31.1 (Harada et al. 1994; Itoh et al. 1991). By conducting a search of the most recent assembly of human genome sequence data (2003: BLAT genome search, UCSC Genome Bioinformatics), IRF-1 and its promoter mapped to nt 131902681–131894935 and 131903176–131902682 on chromosome 5, respec-

This research conformed to the guidelines of the University of Manitoba and University of Nairobi human research ethics boards. Parts of this work were presented at the 12th Annual Canadian Conference on HIV/AIDS Research, Halifax, Nova Scotia, Canada, April 2003.

H. Ji · T. B. Ball · F. A. Plummer (✉)
Department of Medical Microbiology,
University of Manitoba, Room 514, 730 William Avenue,
Winnipeg, MB, Canada, R3E 0W3
E-mail: plummerf@cc.umanitoba.ca
Tel.: +1-204-7893312
Fax: +1-204-7893926

J. Kimani
Department of Microbiology, University of Nairobi,
Nairobi, Kenya

F. A. Plummer
National Microbiology Laboratory,
Canadian Science Center for Human and Animal Health,
Winnipeg, MB, Canada

tively. They are located in the IL-4 or Th2 gene cluster, a group of critical immune genes that includes IL-3, IL-4, IL-5, IL-13, GM-CSF, and IRF-1.

Besides its importance in regulating antiviral immune responses, IRF-1 has also been shown to be able to directly regulate HIV-1 transcription and replication. A region downstream of the HIV-1 5'-LTR, spanning nt +200 to +217, was shown to be homologous to the ISRE and has been suggested to be important for HIV-1 transcription and replication (Battistini et al. 2002; Sgarbanti et al. 2002). IRF-1 was shown to be able to activate HIV-1 LTR transcription in the absence of Tat, a pivotal HIV-1 protein believed to be the key initiator and regulator of HIV-1 transcription and replication (Frankel and Young 1998). This suggests that IRF-1 plays a role in the establishment of primary HIV-1 infection, because it can be induced early upon viral entry prior to Tat. IRF-1 could also cooperate with Tat in amplifying HIV-1 gene transcription and replication, suggesting that IRF-1 is likely important in HIV-1/AIDS disease progression as well.

IRF-1 genetic polymorphism has been studied in a variety of disease contexts. Beside the well-studied different "GT" dinucleotides repeats in intron 7 of IRF-1, 31 single nucleotide polymorphisms (SNPs) in IRF-1 have been identified in human subjects or cell lines by other researchers (Database1; Database2; Donn et al. 2001; Nakao et al. 2001; Noguchi et al. 2000; Saito et al. 2001, 2002). Due to the potential role that IRF-1 may play in HIV-1 replication and its importance in the antiviral immune response, it is important to identify genetic polymorphisms in IRF-1 and determine if these mutations result in coding changes in expressed IRF-1 protein. Once identified, these polymorphisms could be examined for their relevance to susceptibility to HIV-1 infection.

Materials and methods

Subjects

Two hundred seventy-seven subjects were enrolled in this study. All individuals are from a well-characterized commercial female sex worker cohort established in Pumwani area of Nairobi, Kenya (Fowke et al. 2000, 1996). The population came from diverse regions of Kenyan and represents a variety of ethnic groups. Eighteen healthy donors without African genetic background were utilized as controls. Studies involving these sampling cohorts had been approved by both the University of Manitoba and the University of Nairobi ethics review panels.

Extraction of DNA

Peripheral blood samples were obtained from the described individuals; DNA was extracted either from

peripheral blood mononuclear cells (PBMCs) or from whole blood directly using Qiagen DNA extraction kit (Qiagen, Inc., Mississauga, ON, Canada) following the manufacturer's instructions.

Complete gene sequencing of IRF-1 and its promoter

Comparative genomic sequencing for the IRF-1 gene was conducted on all DNA samples. In total, 17 segments were PCR amplified using primers designed to cover the entire IRF-1 gene and its promoter region. All primers were designed based on GenBank sequences for IRF-1 and its immediate upstream promoter region (accession numbers: L05072 and X53095) (Table 1). Following DNA extraction, PCR amplification was conducted as follows: the 50 µl PCR reaction mixture contained 60 mM Tris-HCl (pH 9.0), 1.5 mM MgCl₂, 15 mM (NH₄)₂SO₄, 100 µM dNTP mixture, 25 pmol of each primer; 1.25 U Taq DNA Polymerase (Invitrogen Life Technologies, Burlington, ON, Canada); and 100–200 ng of DNA template. PCR was performed on a GeneAmp PCR-System 9600 thermocycler (Applied Biosystems, Foster City, CA, USA) using the following program: (1) an initial denaturation step of 3 min at 94°C; (2) 35 cycles of 30 s at 94°C, 30 s at the indicated annealing temperature (T_m) (Table 1), and 2 min elongation at 72°C followed by a final extension step of 10 min at 72°C. All PCR amplicons were purified using Millipore's Amicon Microcon-PCR Centrifugal filter devices (Millipore Bedford, MA, USA) prior to sequencing. Sequencing was conducted using ABI PRISM BigDye Terminator Version 3.0 Cycle Sequencing system (Applied Biosystems). Each reaction contained 2 µl of Big Dye (Version 3 or 3.1), 1.5 µl of 10 mM primer (forward or reverse), and 2 µl of purified PCR amplicon templates from above. Sequencing cycling reactions were as follows: (1) an initial denaturation step of 3 min at 96°C; (2) 80 reaction cycles each consisting of 30 s at 96°C, 30 s at the specified T_m (Table 1), and 4 min at 60°C. All PCR products were ethanol precipitated and then reconstituted with 20 µl formamide (Applied Biosystems) and resolved using ABI Prism 3100 Genetic Analyzer (Hitachi, Japan). SNPs in a select 16 individuals were reconfirmed using Expand High Fidelity Plus PCR system (Roche Diagnostics GmbH, Mannheim, Germany) to eliminate the possibility of errors introduced due to imperfect proof-reading in PCR amplification.

Sequence resolution and polymorphism identification

Sequence data were resolved with Sequencher (Version 4.0.5, Gene Codes Corporation, USA), and gene variations were identified by alignment with published GenBank sequences (accession numbers: L05072 or X53095). All polymorphisms were confirmed by sequencing in both directions. Genotype and allele

Table 1 Oligonucleotide primers used for IRF-1 sequencing. *T_m* annealing temperature

Locus amplified ^a	Amplification primer sequences		<i>T_m</i> (°C)	Sequencing primer sequences		<i>T_m</i> (°C)	<i>T_m</i> (°C)
	Forward	Reverse		Forward	Reverse		
-489 to -33	gagga gccag gctgc cagtc	cactt agcgg gatic cccag	54.8	ggagc cagggc tggcca gt	cactt agcgg gatic ccca	53.0	55.0
-85 to 250	cgcgc ggcig tccgg gggcg	gficc ccggc cteaa agggc	65.6	agggc tggggc aatcc cgc	gficc ccggc cteaa ag	53.0	53.0
13 to 662	tcctt agtcg agggc agac	cagta agcag cccct gcc	54.8	tagtc gaggc aagac gfg	agccc ttgcc accag caca	53.0	55.0
611 to 1262	ccacg gcgac aggtt ttg	ctcca gtcig gaatic tcttg a	50.0	caggg gacag gtttt gggct	ctgga atctg ttgaa cagta cct	53.0	51.0
1171 to 1550	tagca gctgg ctgag gtag	tcacg caagt gaggc ccagg	63.1	tagca gctgg ctgag gtag	tcacg caagt gaggc cca	53.0	51.0
1422 to 2086	cttct ggagg gacca aagct	ctgag tccag ccgtt caagt	57.0	gacca aagct tccag tccag	tcacg tccaa atctt ggctt c	53.0	51.0
2008 to 2678	cagag gcact gagag gtaga	ctgag tccag gaccc aggc	59.0	gaggc actga ggggt agatc	gctcg cccag ccaga cag	47.0	47.0
2611 to 3046	gggaa aagag gggaa gicag	ctaag atctt ggctt ttccc	55.8	gaaa aaggg gaagt cag	agatc ttggt ttctt cct	47.0	47.0
2989 to 3642	ctgga gaaga gagtc gatca tc	gcttt gctac ggctt agagg	59.0	cgatc atctc agcaa atctt c	tagag ggcag gaga aaaag agg	53.0	53.0
3591 to 4040	catgg cctcg acgat cctct	cactt tagca ttctt tccac	52.0	tggcc ctgac gatcc tct	cttag cattt ctccc actgc	53.0	53.0
4043 to 4545	ctggg cctaa gctgc ttic	acact caagc agcag gccca	61.0	tggcc ctaag ctgct ttct	agcca agcag caggc cag	53.0	53.0
4482 to 4899	tgta cccat gccct ccac	gtgga cctat ctgcc ataga g	57.0	tggcc tccac ctctg aag	ccat ctgcc ataga gatic ca	53.0	53.0
4951 to 5390	aaggg caagg ggtac tggtg	tgggg catgt tgggg cagag	64.7	agggc agggg gatct ggt	gcag ttggg gcaga ggtat	55.0	55.0
5323 to 6247	cagct ctgig agtca ggcac	ctggg gacag aagag caga	57.0	gtcag gcaat atcag cgtatg	agcaa agggc tatca actcc	55.0	55.0
6211 to 6916	gaagg gattt gatag cctct	cttgg cagtg gggic aca	55.0	ttgat agcct ctgic tcttc	cttgg ctgtt gaggg gc	53.0	53.0
6641 to 7200	cctca taca ccttc ttac	tctct tagtg cagag ccagc	54.8	catac aacct tcttc acatc	tctct tagtg cagag cca	51.0	49.0
7155 to 7705	cagaa aagca taaca ccaat	cattc ctaca agtia aattt c	44.0	aagca taaca ccaat cccag	tctca caagt taaat ttctt tt	53.0	51.0

^aNucleotide numbers indicating locus amplified are numerated according to GenBank sequence of human IRF-1 gene (accession number L05072), the immediate upstream promoter sequence (495 nucleotides) was numerated by -1 to -495 (accession number X53095)

frequencies in the sample population were then calculated. All SNPs were tested for Hardy–Weinberg equilibrium. Pairwise linkage disequilibrium (LD) analysis between SNPs was conducted using PyPop (Lancaster et al. 2003).

Results and discussion

The Kenyan population displayed extensive diversity throughout the IRF-1 locus. Fifty-three SNPs in IRF-1 were identified in both the IRF-1 gene and its promoter region (Table 2, Fig. 1) (Cha et al. 1992). The genotype and allelic frequencies of each SNP were calculated and depicted in Table 2. Twenty-seven of these SNPs have been previously reported either in publications or by the International Human Genome Sequencing Consortium through the NCBI-based SNP database (Database1; Database2; Donn et al. 2001; Nakao et al. 2001; Noguchi et al. 2000; Saito et al. 2001, 2002). A number of previously reported SNPs were not observed in the subjects from this study, including 5551T/G, 4950A/G, 5558T/C, 5636G/A, 7662C/- and 6355G/A, demonstrating the high degree of IRF-1 genetic polymorphism in different ethnic populations. Beside the 26 novel SNPs, two novel insertion mutations were identified in the second intron of IRF-1 [“CA” insertion between nt 2592 and 2593 in L05072 or nt 131900079 and 131900078 in chromosome 5 (TG → TCAG)] and in the 3' UTR [an “A or G” insertion between nt 7607 and 7608 or nt 131895046–131895044 in chromosome 5 (AG → AA/GG)] with genotype frequencies at 16.33% and 93.33%, respectively. A 16 bp deletion was identified in intron 7 (Table 3) in 28.30% of subjects tested. This mutation is completely linked with the “T” allele of SNP at 4816. Remarkably, this deletion is located quite close to the “GT” dinucleotides repeats (4907 → 28 in L05072), a putative site for the formation of Z-DNA (Cha et al. 1992). Taken together, these findings indicate a high degree of polymorphism in this particular population and suggest that there may be distinct IRF-1 genetic variations within the local Kenyan population.

All of the above variations, except for two silent mutations at 4396A/G and 4420C/T in exon 7, were located in noncoding regions (Fig. 1) (Cha et al. 1992). Of the 53 SNPs identified, 11 were located in the promoter region (-415A/C, -410G/A, -388T/C, -386C/T, -300G/A, -298G/A, -281G/A, -280G/C, -203G/A, -90G/A, and -65G/A), four in Exon 1 (5' untranslated region (UTR)) [(53C/G, 142C/T, 154C/T, 197G/A)], and seven in the 3' UTR (6936C/T, 7175C/T, 7238C/T, 7303T/C, 7311G/A, 7447C/A, 7489C/T). Another reported silent mutation, 6355G/A, was not found in any subjects (Noguchi et al. 2000). The other variations were distributed throughout introns 1, 2, 3, and 6–9. Except for the SNP at position 4816 ($P = 0.0025$), all SNPs tested were found to be in Hardy–Weinberg equilibrium. The 4816 SNP is also associated

Table 2 Polymorphisms in the IRF-1 gene and its promoter region detected in a Kenyan population

Loci ^a	Geno- type	Percent ^b	Allele	Percent ^b	Loci	Geno- type	Percent	Allele	Percent	Loci	Geno- type	Percent	Allele	Percent
–415 (n = 187) ^c	AA	18.7	A	44.4	–410 (n = 187)	GG	18.2	G	43.9	–388 (n = 188)	TC	4.3	T	2.1
	AC	51.3	C	55.6		GA	51.3	A	56.1		CC	95.7	C	97.9
	CC	29.9				AA	30.5							
–386 (n = 188)	CC	96.8	C	98.4	–300 (n = 188)	GG	26.1	G	54.0	–297 (n = 186)	GG	98.9	G	99.5
	CT	3.2	T	1.6		GA	55.8	A	46.0		GA	1.1	A	0.5
						AA	18.1							
–281 (n = 188)	GG	99.5	G	99.7	–280 (n = 188)	GG	99.5	G	99.7	–203 (n = 188)	GG	99.5	G	99.7
	GA	0.5	A	0.3		GC	0.5	C	0.3		GA	0.5	A	0.3
–90 (n = 187)	GG	98.9	G	99.5	–65 (n = 188)	GG	98.9	G	99.5	53 (n = 151)	CC	90.7	C	95.4
	GA	1.1	A	0.5		GA	1.1	A	0.5		CG	9.3	G	4.6
142 (n = 156)	CC	91.0	C	95.5	154 (n = 156)	CC	35.9	C	59.9	197 (n = 153)	GG	96.7	G	98.4
	CT	9.0	T	4.5		CT	48.1	T	40.1		GA	3.3	A	1.6
						TT	16.0							
379 (n = 105)	CC	81.9	C	91.0	619 (n = 104)	AA	9.6	A	32.7	831 (n = 48)	GG	91.7	G	95.8
	CT	18.1	T	9.0		AC	46.2	C	67.3		GC	8.3	C	4.2
						CC	44.2							
858 (n = 48)	CC	97.9	C	99.0	1970 (n = 48)	TT	35.4	T	60.4	2392 (n = 66)	AA	47.0	A	67.4
	CT	2.1	T	1.0		TG	50.0	G	39.6		AT	40.9	T	32.6
						GG	14.6				TT	12.1		
2529 (n = 56)	AA	94.6	A	97.3	2936 (n = 83)	TT	34.9	T	61.4	3168 (n = 48)	TT	97.9	T	99.0
	AG	5.4	G	2.7		TC	53.0	C	38.6		TC	2.1	C	1.0
						CC	12.1							
3267 (n = 48)	TT	33.3	T	59.4	4220 (n = 93)	TT	95.7	T	97.8	4227 (n = 93)	AA	24.7	A	51.6
	TC	52.1	C	40.6		TC	4.3	C	2.2		AG	53.8	G	48.4
	CC	14.6									GG	21.5		
4318 (n = 93)	AA	24.7	A	51.6	4379 (n = 93)	TT	24.7	T	51.6	4396 (n = 93)	AA	24.7	A	51.6
	AG	53.8	G	48.4		TC	53.8	C	48.4		AG	53.8	G	48.4
	GG	21.5				CC	21.5				GG	21.5		
4420 (n = 90)	CC	98.9	C	99.4	4583 (n = 129)	AA	28.7	A	51.6	4656 (n = 130)	GG	87.7	G	93.8
	CT	1.1	T	0.6		AC	45.7	C	48.4		GA	12.3	A	6.2
						CC	25.6							
4659 (n = 130)	GG	29.2	G	51.9	4692 (n = 130)	CC	99.2	C	99.6	4709 (n = 130)	GG	28.5	G	51.5
	GA	45.4	A	48.1		CT	0.8	T	0.4		GT	46.1	T	48.5
	AA	25.4									TT	25.4		
4816 (n = 120)	GG	34.2	G	52.9	5203 (n = 93)	CC	24.7	C	51.6	5261 (n = 93)	GG	24.7	G	51.6
	GT	37.5	T	47.1		CT	53.8	T	48.4		GA	53.8	A	48.4
	TT	28.3				TT	21.5				AA	21.5		
5617 (n = 44)	TT	31.8	T	59.1	6163 (n = 47)	GG	34.0	G	59.6	6460 (n = 277)	GG	22.7	G	48.7
	TC	54.6	C	40.9		GT	51.1	T	40.4		GA	52.0	A	51.3
	CC	13.6				TT	14.9				AA	25.3		
6467 (n = 277)	AA	22.7	A	48.7	6516 (n = 263)	GG	11.0	G	32.7	6637 (n = 46)	CC	30.4	C	56.5
	AG	52.0	G	51.3		TG	43.4	T	67.3		CT	52.2	T	43.5
	GG	25.3				TT	45.6				TT	17.4		
6639 (n = 46)	AA	43.5	A	65.2	6936 (n = 36)	CC	94.4	C	97.2	7175 (n = 36)	CC	94.4	C	97.2
	AG	43.5	G	34.8		CT	5.6	T	2.8		CT	5.6	T	2.8
	GG	13.0												
7238 (n = 90)	CC	97.8	C	98.9	7303 (n = 90)	TT	98.9	T	99.4	7311 (n = 92)	GG	23.9	G	51.6
	CT	2.2	T	1.1		TC	1.1	C	0.6		GA	55.4	A	48.4
											AA	20.7		
7447 (n = 90)	CC	98.9	C	99.4	7489 (n = 90)	CC	97.8	C	98.9					
	CA	1.1	A	0.6		CT	2.2	T	1.1					

^aNucleotide numbers indicating site of polymorphism are numerated according to the GenBank sequence of human IRF-1 (accession number L05072, corresponding to nt 131902681–131894935 of human chromosome 5), the immediate upstream promoter sequence (495 nucleotides) was numerated by –1 to –495 (accession

number X53095, corresponding to nt 131903176–131902682 of human chromosome 5)

^bGenotype or allele frequencies were calculated directly based upon number sequenced for each polymorphism

^cNumbers in brackets indicate the number of individuals tested

with the 16 bp deletion polymorphism in intron 7, suggesting this region may be undergoing some selective pressure, or there is a bias in the population under investigation in this study.

Polymorphisms in the promoter region may affect the transcriptional activation of IRF-1 and its subsequent

expression. Supporting evidence comes from a study that demonstrated that a particular promoter polymorphism (SNP at –300G/A) was shown to be correlated with cellular immunity to HIV infection (Saito et al. 2002). Our data demonstrates that promoter polymorphism is common in the Kenyan IRF-1 gene sequences.

Fig. 1 IRF-1 genomic structure, and distribution of polymorphic sites. Nucleotide numbers indicating loci are according to GenBank sequence for IRF-1 (accession number L05072) and correspond to nt 131902681–131894935 of human chromosome 5. The human IRF-1 gene, spanning 7.72 Kb, and its immediate upstream promoter sequence have been mapped to chromosome 5q31.1. It is composed of 10 exons (exon 1 is untranslated), nine introns, and one 3' UTR. Forty-two SNPs were distributed in the untranslated exon I, introns 1, 2, 3, 6, 7, 8, 9, and in the 3' UTR. Two insertion polymorphisms were identified in intron 2 and 3' UTR. One 16 bp deletion polymorphism was discovered just before the dinucleotide "GT" repeat in intron 7. Two silent mutations were also identified in exon VII (4396 A/G and 4420C/T)

Table 3 Deletion polymorphism identified in intron 7 of the IRF-1 gene

	4787 ^a →	← 4819 ^a
L05072 ^b	AGGGTGAGGAAGGAAGTAGGGTAGGGGTG G GAA	
With Deletion	AGGG*****GTAGGGGTG T GAA	

^aNucleotide deletion from 4791 to 4806 (chromosome 5: 131897874–131897859)

^aNucleotide numbers according to L05072

^bGenBank sequence for human IRF-1 (accession number L05072)

Fig. 2 IRF-1 promoter sequence and distribution of polymorphic sites. * Nucleotide numbers indicating loci are according to GenBank sequence for IRF-1 promoter (accession number X53095) and correspond to nt 131903176–131902682 of human chromosome 5. The immediate upstream region of the IRF-1 promoter is 495 bp in length and mapped to chromosome 5q31.1. The putative recognition sites for known transcription factors are indicated; GAS: IFN- γ -activated sequence. Eleven SNPs marked with * dispersed in the IRF-1 promoter region. SNPs at –386, –203 and –65 were located within three Sp1 binding sites, respectively. No SNPs were related with GAS and NF- κ B recognition. This figure was adapted from Harada et al. (1994)

We identified 11 promoter polymorphisms in the target population, seven of which are novel. Three of them are located in putative Sp1 binding sites (Fig. 2). The functional significance of these variations in IRF-1 promoter remains unknown. Introns and other noncoding regions in the genome have often been regarded as "junk DNA" with little or no apparent function. But significant evidence is arising that the noncoding regions have significant functions in gene

expression, and the “junk” is no longer considered “junk” (Herbert 1996; Mattick 1994; Moore 1996). Genetic variations in noncoding regions of cytokine genes have been correlated with multiple disease conditions (Haukim et al. 2002). A SNP at 7311 in the IRF-1, 3' UTR is correlated with occurrence of juvenile idiopathic arthritis (Donn et al. 2001). Whole allele distributions of “GT” repeats in intron 7 of IRF-1 have also been shown to be associated with onset time of childhood atopic asthma (Nakao et al. 2001). Further study on the epidemiological and functional consequences of these IRF-1 mutations on the HIV-1 epidemic in Kenya is underway.

LD between SNPs was observed throughout the IRF-1 gene. Pairwise LD analysis revealed that LD was extensive in the IRF-1 gene and was observed to some degree to be present between most SNPs (Tables 4, 5). LD exists even between SNPs separated by long intervals and distributed in different functional regions of the IRF-1 gene. As shown in Table 4, complete linkage was observed among SNPs at positions 4227, 4318, 4379,

4396, 5203, 5261, 6460, 6467, and 7311 (“4227 cluster”). Partial linkage disequilibrium is also observed between the “GT” dinucleotides repeats in intron 7, the “4227 cluster,” and other identified SNPs (data not shown). This suggests that haplotype analysis may be necessary for full evaluation of the effect of IRF-1 gene variation in its functional context or in association studies with disease conditions. We were unable to confirm a previous report that nearly complete LD existed between SNPs at -300 and 4396 (Noguchi et al. 2000).

Also remarkably, 35 consistent discrepancies were identified between the sequence data we obtained and GenBank sequences for IRF-1 and its promoter. The consistent discrepancies were defined here as the same “variation” found in all subjects but different from the GenBank submissions. All these 35 discrepancies are depicted in Table 6. These discrepancies were reconfirmed in a control population of 18 male/female individuals with different genetic backgrounds, including Asian, European, and North American, suggesting these discrepancies are consistent among ethnic groups tested.

Table 4 Complete linkage between 4227 and other SNPs within the IRF-1 gene. This linkage was observed in all tested subjects

4227 ^a	4318	4379	4396	5203	5261	6460	6467	7311
A	A	T	A	C	G	G	A	G
A/G	A/G	T/C	A/G	C/T	G/A	G/A	A/G	G/A
G	G	C	G	T	A	A	G	A

^aNucleotide numbers indicating loci according to GenBank sequence for IRF-1 (accession number L05072)

Table 5 Pairwise LD estimates for all SNPs with minor allele frequency higher than 0.1. Pairwise LD analysis were conducted for all identified SNPs with minor allele frequencies higher than 0.1 by using PyPop. The results were shown based on *P*-value. The “4227

cluster” was represented by the 4227 SNP itself. Nucleotide numbers indicating loci were numerated according to GenBank sequence for IRF-1 (accession number L05072)

	-410	-300	154	619	1970	2392	2936	3267	4227	4583	4659	4709	4816	5617	6163	6516	6637	6639
-415	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
-410		***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
-300			***	**	***	***	***	***	***	***	***	***	***	***	***	***	***	***
154				***	***	***	***	***	***	***	***	***	***	***	***	***	***	***
619					***	***	*	***	***	***	***	***	***	***	***	***	***	***
1970						***	***	***	***	***	***	***	***	***	***	***	***	***
2392							***	***	***	***	***	***	***	***	***	***	***	***
2936								***	***	***	***	***	***	***	***	***	***	***
3267									***	***	***	***	***	***	***	***	***	***
4227										***	***	***	***	***	***	***	***	***
4583											***	***	***	***	***	***	***	***
4659												***	***	***	***	***	***	***
4709													***	***	***	***	***	***
4816														***	***	***	***	***
5617															***	***	***	***
6163																***	***	***
6460																	***	***
6516																		***
6637																		***

**P* < 0.05

***P* < 0.01

****P* < 0.001

Table 6 Consistent discrepancies between standard and detected sequences

Position ^a	L05072	Variation	Position	L05072	Variation	Position	L05072	Variation
97–100	CGCG ^b	CGGCGG ^{b,c}	1751	C	CC ^b	4746	T	– ^b
111	C ^b	A ^b	1867	T	TT ^b	4782–4786	CGTGG	C–T–G ^b
153–156	TCCG	TC/TCGCGCG ^{b,c}	2044	G	A ^b	5000	G	GG ^b
242–243	CC	CGC ^b	2077–2079	CTC	TCT ^b	5160	G	– ^b
335–339	TGTCG	T–T–G ^b	2544–2545	CC	CCC ^b	5253	C	G ^b
416–417	CG	GC ^b	3212–3213	TT	TTT ^b	5260–5261	TG	TG/AGGAAG ^b
603–604	CG	GC ^b	3418–3419	CC	CCC ^b	5293–5294	GC	GGCC ^b
609–610	GC	CC ^b	3565	G	GG ^b	5552	G	GG ^b
655	G	GG ^b	4534	C	CC ^b	5685	G	GG ^b
688	C	CC ^b	4552	C	CC ^b	6081	C	CCC ^b
833	C	G ^b	4626	C	CCC ^b	6183–6184	GG	GGG ^b
1217–1219	GGG	GGGG ^b	4735	T	CT ^b			

^aAll numbers indicating positions according to GenBank sequence for IRF-1 (accession number L05072)

^bFound in all tested samples

^cSame as X53095

We also noted differences between the two GenBank sequences (X53095 and L05072) in the overlapping region (496–669 of X53095 and 1–173 of L05072), suggesting that these discrepancies might be caused by sequence inaccuracy in the available IRF-1 sequence data. This study will provide more complete data on sequence of the human IRF-1 gene.

Acknowledgements We thank Drs. Ma Luo and Xiaojuan Mao for assistance with sequencing and Dr. Keith Fowke for manuscript preparation. This work was supported by grant from the National Institute of Health (RO1 A1 56980) and the Canadian Institute of Health Research (CIHR HOP43135). F. Plummer is a Tier I CIHR Canadian Research chair. H. Ji is recipient of CIHR/International Center for Infectious Diseases doctoral training award.

References

- Anzala AO, Ball TB, Rostron T, O'Brien SJ, Plummer FA, Rowland-Jones SL (1998) CCR2-64I allele and genotype association with delayed AIDS progression in African women. University of Nairobi Collaboration for HIV Research. *Lancet* 351:1632–1633
- Battistini A, Marsili G, Sgarbanti M, Ensoli B, Hiscott J (2002) IRF regulation of HIV-1 long terminal repeat activity. *J Interferon Cytokine Res* 22:27–37
- Cha Y, Sims SH, Romine MF, Kaufmann M, Deisseroth AB (1992) Human interferon regulatory factor 1: intron–exon organization. *DNA Cell Biol* 11:605–611
- Database1. Entrez SNP database <http://www.ncbi.nlm.nih.gov/entrez>
- Database2. Functionally Annotated Mutation Database <http://mutdb.org/AnnoSNP>
- Donn RP, Barrett JH, Farhan A, Stopford A, Pepper L, Shelley E, Davies N, Ollier WE, Thomson W (2001) Cytokine gene polymorphisms and susceptibility to juvenile idiopathic arthritis. British Paediatric Rheumatology Study Group. *Arthritis Rheum* 44:802–810
- Dunand VA, L M, Wade JA et al (1997) Certain MHC Class II DR alleles are associated with decreased susceptibility to HIV-1 infection. 1997 Keystone Symposia: AIDS Pathogenesis, Keystone, Colorado
- Fowke KR, Nagelkerke NJ, Kimani J, Simonsen JN, Anzala AO, Bwayo JJ, MacDonald KS, Nguni EN, Plummer FA (1996) Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* 348:1347–1351
- Fowke KR, Kaul R, Rosenthal KL, Oyugi J, Kimani J, Rutherford WJ, Nagelkerke NJ, Ball TB, Bwayo JJ, Simonsen JN, Shearer GM, Plummer FA (2000) HIV-1-specific cellular immune responses among HIV-1-resistant sex workers. *Immunol Cell Biol* 78:586–595
- Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem* 67:1–25
- Fujita T, Sakakibara J, Sudo Y, Miyamoto M, Kimura Y, Taniguchi T (1988) Evidence for a nuclear factor(s), IRF-1, mediating induction and silencing properties to human IFN-beta gene regulatory elements. *EMBO J* 7:3397–3405
- Fujita T, Kimura Y, Miyamoto M, Barsoumian EL, Taniguchi T (1989) Induction of endogenous IFN-alpha and IFN-beta genes by a regulatory transcription factor, IRF-1. *Nature* 337:270–272
- Harada H, Takahashi E, Itoh S, Harada K, Hori TA, Taniguchi T (1994) Structure and regulation of the human interferon regulatory factor 1 (IRF-1) and IRF-2 genes: implications for a gene network in the interferon system. *Mol Cell Biol* 14:1500–1509
- Haukim N, Bidwell JL, Smith AJ, Keen LJ, Gallagher G, Kimberly R, Huizinga T, McDermott MF, Oksenberg J, McNicholl J, Pociot F, Hardt C, D'Alfonso S (2002) Cytokine gene polymorphism in human disease: on-line databases, supplement 2. *Genes Immun* 3:313–330
- Herbert A (1996) RNA editing, introns and evolution. *Trends Genet* 12:6–9
- Itoh S, Harada H, Nakamura Y, White R, Taniguchi T (1991) Assignment of the human interferon regulatory factor-1 (IRF1) gene to chromosome 5q23-q31. *Genomics* 10:1097–1099
- Kroger A, Koster M, Schroeder K, Hauser H, Mueller PP (2002) Activities of IRF-1. *J Interferon Cytokine Res* 22:5–14
- Lancaster A, Nelson MP, Meyer D, Thomson G, Single RM (2003) PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data. *Pac Symp Biocomput* 8:514–525
- Lohoff M, Ferrick D, Mittrucker HW, Duncan GS, Bischof S, Rollinghoff M, Mak TW (1997) Interferon regulatory factor-1 is required for a T helper 1 immune response in vivo. *Immunity* 6:681–689
- MacDonald KS, Fowke KR, Kimani J, Dunand VA, Nagelkerke NJ, Ball TB, Oyugi J, Njagi E, Gaur LK, Brunham RC, Wade J, Luscher MA, Krausa P, Rowland-Jones S, Nguni E, Bwayo JJ, Plummer FA (2000) Influence of HLA supertypes on susceptibility and resistance to human immunodeficiency virus type 1 infection. *J Infect Dis* 181:1581–1589
- Mamane Y, Heylbroeck C, Genin P, Algarte M, Servant MJ, LePage C, DeLuca C, Kwon H, Lin R, Hiscott J (1999) Interferon regulatory factors: the next generation. *Gene* 237:1–14
- Mattick JS (1994) Introns: evolution and function. *Curr Opin Genet Dev* 4:823–831

- Moore MJ (1996) Gene expression. When the junk isn't junk. *Nature* 379:402–403
- Nakao F, Ihara K, Kusuhara K, Sasaki Y, Kinukawa N, Takabayashi A, Nishima S, Hara T (2001) Association of IFN-gamma and IFN regulatory factor 1 polymorphisms with childhood atopic asthma. *J Allergy Clin Immunol* 107:499–504
- Noguchi E, Shibasaki M, Arinami T, Yamakawa-Kobayashi K, Yokouchi Y, Takeda K, Matsui A, Hamaguchi H (2000) Mutation screening of interferon regulatory factor 1 gene (IRF-1) as a candidate gene for atopy/asthma. *Clin Exp Allergy* 30:1562–1567
- Saito H, Tada S, Ebinuma H, Wakabayashi K, Takagi T, Saito Y, Nakamoto N, Kurita S, Ishii H (2001) Interferon regulatory factor 1 promoter polymorphism and response to type 1 interferon. *J Cell Biochem Suppl* 36(Suppl):191–200
- Saito H, Tada S, Wakabayashi K, Nakamoto N, Takahashi M, Nakamura M, Ebinuma H, Ishii H (2002) The detection of IRF-1 promoter polymorphisms and their possible contribution to T helper 1 response in chronic hepatitis C. *J Interferon Cytokine Res* 22:693–700
- Sgarbanti M, Borsetti A, Moscufo N, Bellocchi MC, Ridolfi B, Nappi F, Marsili G, Marziali G, Coccia EM, Ensoli B, Battistini A (2002) Modulation of human immunodeficiency virus 1 replication by interferon regulatory factors. *J Exp Med* 195:1359–1370
- Taki S, Sato T, Ogasawara K, Fukuda T, Sato M, Hida S, Suzuki G, Mitsuyama M, Shin EH, Kojima S, Taniguchi T, Asano Y (1997) Multistage regulation of Th1-type immune responses by the transcription factor IRF-1. *Immunity* 6:673–679
- Taniguchi T, Ogasawara K, Takaoka A, Tanaka N (2001) IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol* 19:623–655
- Trivedi HN, Plummer FA, Anzala AO, Njagi E, Bwayo JJ, Ngugi EN, Embree JE, Hayglass KT (2001) Resistance to HIV-1 infection among African sex workers is associated with global hyporesponsiveness in interleukin 4 production. *FASEB J* 15:1795–1797