Polymorphisms in the Genes Encoding Members of the Tristetraprolin Family of Human Tandem CCCH Zinc Finger Proteins

PERRY J. BLACKSHEAR,[*,†] RUTH S. PHILLIPS,[*,†] JOHANA VAZQUEZ-MATIAS,[§] AND HARVEY MOHRENWEISER[§]

[*]*Office of Clinical Research and Laboratory of Signal Transduction, A2-05 National Institute of Environmental Health Sciences, 111 Alexander Drive, Research Triangle Park, NC 27709, USA;*
[†]*Department of Medicine and Biochemistry, Duke University Medical Center, Durham, NC 27710, USA; and*
[§]*Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94551-9900, USA*

The three known mammalian CCCH tandem zinc finger proteins of the tristetraprolin (TTP) class have recently been demonstrated to be mRNA-binding proteins. The prototype, TTP, functions in normal physiology to promote the instability of the tumor necrosis factor α (TNFα) and granulocyte-macrophage colony-stimulating factor mRNAs. Conversely, these mRNAs are stabilized in TTP-deficient mice, leading to an inflammatory phenotype characterized by overproduction of these cytokines. To explore sequence variations in TTP and its two related proteins, we sequenced genomic DNA encoding the TTP protein (*ZFP36*) and those of its two known mammalian relatives, *ZFP36L1* and *ZFP36L2*, from 72 to 92 anonymous human subjects from various geographical and ethnic backgrounds. We also sequenced *ZFP36* in genomic DNA from 92 subjects exhibiting evidence of excessive TNFα action. The resequencing strategy identified 13 polymorphisms in the protein-coding regions of these

Progress in Nucleic Acid Research and Molecular Biology, Vol. 75

ISSN: 0079-6603/2003 $35.00

three genes, of which six would result in amino acid changes; other putative polymorphisms were identified by EST searches. One mutation in *ZFP36L1* was a dinucleotide substitution that would prevent splicing of the single intron. This mutation was identified in only one allele of the original 144 sequenced from an adult female Aka Pygmy from the Central African Republic; a second individual with the same variant allele was found by genotyping 58 additional Aka DNA samples. Analysis of mRNA from one of these subject's lymphoblasts confirmed that ZFP36L1 mRNA levels were approximately 50% of those in a comparable sample without the mutation. The functional significance of this and the other polymorphisms identified remains to be determined by both biochemical and population linkage studies.

I. Introduction

The tristetraprolin (TTP) family of CCCH tandem zinc finger proteins contains three known members in mammals and a fourth in *Xenopus laevis* and fish (*1,2*). TTP, also known as Tis11, Nup475, and G0S24, is encoded by the human gene *ZFP36* (OMIM number 190900) (*3*), and was discovered through its rapid transcriptional induction in cells stimulated with insulin, serum, or phorbol esters (*4–7*). Studies in TTP-deficient mice have shown that TTP is involved in the physiological regulation of the secretion of tumor necrosis factor α (TNFα) (*8*) and granulocyte-macrophage colony-stimulating factor (GM-CSF) (*9*). TTP somehow destabilizes the mRNAs encoding these proteins after binding to AU-rich elements (ARE) located in their 3′-untranslated regions (3′UTR), leading to the accelerated destruction of these mRNAs (*1,8–11*).

TTP deficiency in mice leads to a chronic inflammatory state, characterized by weight loss, severe erosive arthritis, conjunctivitis, dermatitis, splenomegaly, adenopathy, autoimmunity, and severe medullary and extramedullary myeloid hyperplasia (*12*). Most aspects of this syndrome appear to be due to chronically elevated concentrations of TNFα (*8,12,13*), and the chronically elevated GM-CSF levels may be partly responsible for the myeloid hyperplasia (*9,14*). Thus, complete TTP deficiency in man would be expected to result in a severe inflammatory disease with myeloid hyperplasia, probably not compatible with life. The heterozygous state in the mouse is also associated with a similar syndrome, although its onset is later in life, and the syndrome is less severe (E. Carballo and P. J. Blackshear, unpublished data). Thus, it is possible that the heterozygous state in humans would be compatible with life, and that less severe mutations affecting TTP's biosynthesis, turnover or function might result in human disease.

Although the phenotypes of mouse knockouts of the two known mammalian *ZFP36* relatives, now officially designated as *ZFP36L1* (for

ZFP36-like 1) (*15*) and *ZFP36L2* (*16*), have not been published to date, we have recently shown that their encoded proteins can exert similar effects to TTP to destabilize TNFα and GM-CSF mRNAs in a cell cotransfection system (*1*). The genes encoding these proteins are regulated differently from the *ZFP36*, but given the similar mRNA binding and destabilizing properties to TTP of the encoded proteins, it seems possible that abnormalities in their expression or primary sequence might also be associated with clinical disease.

To begin to explore the natural variations in the protein-coding regions of these three genes in man, we initially sequenced genomic DNA encompassing their protein-coding regions from the 72 individuals from various ethnic groups that make up the initial study population for the Environmental Genome Project being carried out by the National Institute of Environmental Health Sciences (*17–19*). We also resequenced several sets of DNA from other pertinent groups of subjects, and searched for polymorphisms in the protein-coding regions of these genes by EST analysis. Altogether, we identified 11 polymorphisms in the protein-coding regions of these genes that would result in amino acid changes, including a dinucleotide mutation at an intron splice site that appears to lead to hemizygous expression of the ZFP36L1 mRNA.

II. Materials and Methods

A. Nomenclature

The approved nomenclature for the human gene encoding TTP (also known as TIS11, GOS24, and Nup475) is *ZFP36* (*3*) (OMIM number 190700). This stands for zinc finger protein 36. The approved nomenclature for the second human member of the family is ZFP36-like 1, *ZFP36L1* (OMIM number 601064). Other names for the encoded protein are BRF1, TIS11B, cMG1, Berg-36, and ERF1. The approved nomenclature for the third family member is ZFP36-like 2, *ZFP36L2* (no OMIM number); the encoded protein is also known as TIS11d, ERF2, and BRF2. Approved gene symbols were obtained from the HUGO Gene Nomenclature Committee (HGNC). Further details are available at URL http://www.gene.ucl.ac.uk/nomenclature/ or e-mail nome@galton.ucl.ac.uk.

B. Subjects and DNA

Lymphoblasts representing the 72 anonymous subjects in the initial arm of the Environmental Genome Project were obtained from the Coriell Collection. Details of this collection, as well as culture conditions and genomic DNA

extraction methods, are described in detail in Ref. (*17*). Briefly, the group labeled Asian (A) contained representatives from the Taiwanese Ami (5 subjects), Indo-Pakistan (4), Cambodian (3), Beijing Chinese (5), Japanese (3), and Melanesian (4) groups. The group labeled Black (B) contained African American (15) and Pygmy subjects from the Central African Republic (4) and Zaire (5). The group labeled Caucasian (C) contained Druze (5), Adygei (5), Moscow Russian (5), and Utah (9) subjects. In the case of *ZFP36*, we also sequenced the promoter, portions of the 5′- and 3′UTRs, and the single intron, from the original 72 subjects as well as from 20 additional North American individuals from the Coriell Polymorphism Discovery collection, 42 individuals that were either hypo- or hyperresponsive to inhaled endotoxin in terms of their bronchial resistance (kindly provided by Dr. David A. Schwartz, Duke University Medical Center, Durham, NC; see Ref. (*20*)), and 50 individuals with a systemic inflammatory syndrome resembling the Tumor necrosis factor Receptor-Associated Periodic Syndrome (TRAPS) (*21*) that were negative for mutations in the TNFα receptor (kindly provided by Dr. Ivona Aksentijevich and Daniel L. Kastner, National Institutes of Health, Bethesda, MD). For specific mutations, we used other genotyping strategies, as described in the next section, in DNA from 58 additional Aka Pygmy subjects (kindly provided by Dr. L. Luca Cavalli-Sforza and Alice A. Lin, Stanford University, Palo Alto, CA), 430 anonymous residents of Durham, NC of various ethnicities (kindly provided by Dr. Douglas Bell, NIEHS, Research Triangle Park, NC), and 20 DNA samples from germ cell testicular cancers (kindly provided by Dr. David Hogg, University of Toronto, Canada) (*22*).

C. Sequencing Strategy

The strategy of resequencing of PCR products containing the exons plus splice sites and the 5′ and 3′ regions of genes as an efficient strategy for identification of genetic variation, especially sequence variation resulting in amino acid substitutions, in heterozygote individuals in a population has been discussed (*23–25*). PCR primers were located so that amplification of the genomic sequence was initiated approximately 50 bp from the intron–exon boundary, and the PCR products were 350–450 bp.

D. PCR-Amplification Conditions

PCR primers were designed using the Oligo Primer Analysis Software (National Biosciences, Inc., Plymouth, MN). Appended to the 5′ end of each of the PCR primers were sequences containing the primer-binding sites for the forward or reverse energy transfer DNA-sequencing primers (Amersham Life Science, Cleveland, OH). PCR primers were matched so that the sense

and the antisense PCR primers contained different sequencing primer-binding sites. Other procedures are as described (*25*).

E. DNA Sequencing

The PCR products were diluted 10-fold with TE [10 mM Tris–HCl (pH 8.0 at 20°C), 0.1 mM EDTA] and directly sequenced in both directions as described (*25*). Sequencing reactions were performed according to the manufacturer's instructions using the DYEnamic Direct cycle sequencing kit with the DYEnamic ET primers (Amersham Life Science, Inc., Cleveland, OH).

F. Sequence Analysis

The initial data analysis (lane tracking and base calling) was performed with the ABI Prism DNA Sequence Analysis Software (version 2.1.2). The chromatograms were reanalyzed with Phred (bases called and quality of sequence values assigned, version 0.961028), assembled with Phrap (version 0.960213), and the resulting data viewed with Consed (version 4.1). Description and documentation for Phred, Phrap, and Consed may be obtained at http://www.genome.washington.edu. "PolyPhred" (version 2.1), a software package that utilizes the output from Phred, Phrap, and Consed to identify nucleotide substitutions in heterozygote individuals (*24,26*), was utilized to identify presumptive variants.

G. Other Methods

A dinucleotide polymorphism identified in one chromosome of a single subject was predicted to destroy an intron splice site within the protein-coding region of *ZFP36L1*. In this case, the DNA sequence traces in both directions were retrieved to confirm the presence of this mutation, and a novel restriction fragment length polymorphism (RFLP) was identified at the site in the form of a new site for cleavage by Fnu4HI and TseI. The subject's DNA was used to confirm the presence of this novel restriction site by amplifying genomic DNA with PCR primers based on neighboring sequences, followed by cleavage with TseI under standard conditions. The forward primer used was 5′-GATGACCACCACCCTCGTGTC-3′ (complementary to bp 650–670 of GenBank accession number X79066) and the reverse primer was 5′-CCCATTGCCTTCCAAGACCC-3′ (complementary to bp 781–800). To confirm that the splice site mutation resulted in failure of mRNA splicing, lymphoblasts from this subject were grown under standard conditions, and the cells were harvested in log phase growth and used for the preparation of total cellular RNA as described (*1*). This RNA, and a

parallel sample from another member of the same ethnic group, was used for northern blotting of ZFP36L1 mRNA, using a probe based on the protein-coding region of the orthologous rat cDNA sequence (GenBank accession number X52590, bases 110–1120) (*10*). ZFP36L1 mRNA expression was quantitated by PhosphorImager analysis, and corrected for gel loading by cohybridization with a GAPDH cDNA probe as described (*8*).

This RFLP assay was also used to screen the 58 additional Aka DNA samples described earlier.

Other genotyping was performed on various sample sets for selected polymorphisms as described (*27*) (QIAGEN Genomics, Inc., Bothell, WA). For the trinucleotide repeat polymorphism detected in *ZFP36L2*, resequencing of this region using selected sample sets was performed by methods similar to those described earlier for the primary resequencing project.

H. EST Searches

One class of EST searches focused on identifying ESTs that contained polymorphisms already documented by the current sequencing effort. In this case, approximately 30 bp on either side of the polymorphism was included in a blastn search of the current human EST database. No attempt was made to examine the original EST sequence traces for accuracy, with the following exception: in the case of one EST, a 6-bp CAG repeat polymorphism was identified in *ZFP36L2* that would result in an increase from seven to nine consecutive encoded glutamine residues; in this case, the original GenBank traces were examined, and one of the EST clones was obtained and resequenced in both directions using dRhodamine Terminator Cycle Sequencing (Perkin-Elmer, Foster City, CA) to confirm the presence of the extra CAGCAG sequence.

A second EST searching effort was directed at finding new polymorphisms that had not been discovered by the resequencing effort. To this end, the edited reference sequences (see below) spanning only the protein-coding regions of the three genes under consideration were used to search the GenBank human EST database, using the program blastn and the "flat query-anchored with identities" alignment view. Putative polymorphisms that were seen three times or more were then checked by reblasting the database with the putative polymorphism flanked by 30 bp of genomic sequence on either side. If the polymorphism occurred three or more times, in what appeared to be an otherwise reliable sequence, it was counted as a putative polymorphism and included in Tables I, II and III, along with the frequency of occurrence in the GenBank human EST collection.

TABLE I
ZFP36 POLYMORPHISMS

Polymorphism	Location	B	Base change	Sequence	Amino acid change	Variant allele frequency in 92 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
*ZFP36*1*	Promoter	316	C → A	CCCCC(C/A)ATCCG		1.8		0	0	0
*ZFP36*2*	Promoter	359	A → G	CGGTC(A/G)CGGCT		47		29	30	32
*ZFP36*3*	Promoter	490	C → A	CCGGC(C/A)CCGGC		3.1		0	100	0
*ZFP36*4*	Promoter	492	C → T	GGCCC(C/T)GGCCC		0.6		0	100	0
*ZFP36*5*	Intron	1226	G → A	GGGAA(G/A)CCGGG		0.5		0	100	0
*ZFP36*6*	Intron	1256	C → G	TAAGG(C/G)CTCGG		0.5		0	100	0
*ZFP36*7*	PCD (ex. 2)	1525	C → T	CGGGA(C/T)CCTGG	P37 → S	0.6	0/127	0	0	0
*ZFP36*8*	PCD (ex. 2)	1725	C → T	TCGCG(C/T)TACAA	R103 → R	6.2	2/127 (1.6%)	22	44	33
*ZFP36*9*	PCD (ex. 2)	2235	T → C	CCCTC(T/C)GTACA	S273 → S	4.2	1/69 (1.4%)	0	50	50
*ZFP36*10*	3′UTR	2980	Del TT	TTTTT(delTT)GTAAT		7.6	64/249 (26%)	18	27	55
*ZFP36*11*[a]	PCD (ex. 2)	1807	G → T	GCCTG(G/T)GCGAG	G131 → C		5/118 (4.2%)			
*ZFP36*12*[a]	PCD (ex. 2)	2112	C → T	GCCTT(C/T)TCTGC	F232 → F		4/97 (4.1%)			
*ZFP36*13*[a]	PCD (ex. 2)	2184	C → A	AGGGC(C/A)ACTCC	A256 → A		10/94 (11%)			
*ZFP36*14*[b]	Intron	783	C → T	TGCCT(C/T)CCGCT				0	0	0

These data are based on the following accession numbers for human *ZFP36*: gene, M92844; cDNA, M63625; protein, AAA61240. B refers to the base number in the genomic sequence M92844; AA refers to the amino acid number in AAA61240. The polymorphic changes are indicated as follows: C (original base or amino acid) → T (polymorphic base or amino acid). See Fig. 1 for a depiction of the gene, sequencing primers, and location of polymorphisms.

PCD, protein-coding domain; ex., exon; 3′UTR, 3′-untranslated region; 5′UTR, 5′-untranslated region; EGP, Environmental Genome Project. A, Asian; B, Black; C, Caucasian.

Variants shown as having frequencies of 0 for all three ethnic groups were found in the 20 added EGP samples of unknown ethnicity.

[a]Polymorphisms identified by EST searches only.

[b]Found in 1/162 (0.6%) of alleles from bronchial sensitivity and TRAPS-like syndrome subjects, not found in 92 EGP subjects.

TABLE II
ZFP36L1 Polymorphisms

Polymorphism	Location	B	Change	Sequence	Amino acid change	Variant allele frequency in 72 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
*ZFP36L1*1*	5′UTR	644	G → T	CGAAC(G/T)CACAG		12.5	6/60 (10%)	0	56	44
*ZFP36L1*2*	PCD1	706	AG → GC	ATGCA(AG/GC)GTAAA	K19 → S	0.7	0/65[a]	0	100	0
*ZFP36L1*3*	Intron 1	729	G → A	CATTT(G/A)CTTTT		0.7		0	100	0
*ZFP36L1*4*	Intron 1	772	C → CC	ACCCC(C/CC)AAAAG		18.8		7.4	56	37
*ZFP36L1*5*	Intron 1	804	A → G	GGGAA(A/G)GTGGT		6.2		0	0	100
*ZFP36L1*6*	Intron 1	845	G → C	TTTCT(G/C)CCAAG		4.2		100	0	0
*ZFP36L1*7*	PCD2	3685	G → A	ACCGG(G/A)CTGCT	G218 → G	0.7	0/40	100	0	0
*ZFP36L1*8*	PCD2	3915	C → A	CAGCC(C/A)TCAGG	P295 → H	0.7	0/46	100	0	0

These data are based on accession numbers X79066 and X79067 for the 5′ and 3′ regions of the gene, X99404 for the cDNA, and CAA67781 for the protein. The intron sequence between X79066 and X79067 was supplied by the chromosome 14 BAC clone AL 133313.2.2. The numbering system used for polymorphisms *ZFP36L1*1–6* was from X79066. The numbering system for polymorphisms *ZFP36L1*7* and P8 was based on interposing 1347 bp between X79066 and X79067, and numbering the bases consecutively starting at the 3′ end of X79066. The corresponding base numbers for polymorphisms *ZFP36L1*7* and *ZFP36L1*8* in X79067 are 1365 and 1595, respectively. See the legend to Table I for abbreviations and other details.

[a]One EST continued into the intron (BF873589).

TABLE III
ZFP36L2 POLYMORPHISMS

Polymorphism	Location	B	Change	Sequence	Amino acid change	Variant allele frequency in 72 EGP subjects (%)	Variant allele frequency in ESTs (%)	Variant allele frequency in ethnic groups from 72 EGP subjects (%)		
								A	B	C
*ZFP36L2*1*	5′UTR	1591	Del	CTGCC(GCTCCGGCCACTGCG/del)GGATC		0.7	0/11	0	100	0
*ZFP36L2*2*	Intron 1	1753	C → A	ACGCC(C/A)CTCCC		0.7		100	0	0
*ZFP36L2*3*	PCD2	2389	C → T	GCCGG(C/T)GGTCC	G93 → G	0.7	0/20	100	0	0
*ZFP36L2*4*	PCD2	2435	G → A	GCGGC(G/A)GCACA	G109 → S	2.8	1/27 (3.7%)	25	0	75
*ZFP36L2*5*	PCD2	2445	T → A	AGCCC(T/A)GCTCA	L112 → Q	1.4	0/28	0	100	0
*ZFP36L2*6*	PCD2	2719	C → T	ACCAT(C/T)GGCTT	I203 → I	6.25	20/78 (26%)	22	44	33
*ZFP36L2*7*	PCD2	2870	C → T	CCAAG(C/T)TGCAC	L254 → L	48	52/97 (54%)	20	57	20
*ZFP36L2*8*	PCD2	3055	C → T	ACGCC(C/T)TCGGG	P315 → P	2.1	0/70	33	0	67
*ZFP36L2*9*	PCD2	3323	G → T	CCCCC(G/T)CGCAG	A407 → S	1.4	0/204	0	100	0
*ZFP36L2*10*	3′UTR	3601	G → A	AGGGC(G/A)CCAGT		0.7	0/106	100	0	0
*ZFP36L2*11*[a]	PCD2	3055	C → A	ACGCC(C/A)TCGGG	P315 → P		3/69 (4.3%)			
*ZFP36L2*12*[a]	PCD2	3317	Insert (CAG)2	AGCAG(insertCAGCAG)GGCCT	Insert QQ after Q401		6/>200			
*ZFP36L2*13*[a]	PCD2	3458	TC → CT	CGGAC(TC/CT)GCTGT	S452 → L		4/>200			
*ZFP36L2*14*[a]	PCD2	3536	G → A	GCCTC(G/A)ACCCT	D478 → N		3/>150			
*ZFP36L2*15*[a]	PCD2	3579	C → G	CATCT(C/G)CGACG	S492 → C		5/>100			

These data are based on accession numbers U07802 (gene), NM_006887.3 (cDNA) and NP_008818 (protein). The numbering of DNA and protein polymorphisms is based on the numbering systems used in U07802 and NP_008818, respectively. See the legend to Table I for abbreviations and other details.

[a]Polymorphisms identified from EST searches only.

III. Results

A schematic representation of the results of this study is shown in Fig. 1. It depicts human *ZFP36*, *ZFP36L1*, and *ZFP36L2*, as well as the location of exons, introns, protein-coding regions, and the 64-amino acid tandem zinc finger domain within the genomic sequences. It also shows the orientation of sequencing primer pairs, the location of the PCR products generated, and the locations of the polymorphisms detected by the present resequencing effort. These polymorphisms are named, for example, *ZFP36*4*, to indicate polymorphism 4 within *ZFP36*; this nomenclature convention was approved by the Human Genome Variation Society http://ariel.ucs.unimelb.edu.au/~cotton/mdi.htm (S. E. Antonarakis, personal communication).

The polymorphisms determined by the resequencing effort are indicated below the line representing each gene by the arrowheads labeled, e.g., *4. The putative polymorphisms generated solely by the EST searches are indicated with an arrowhead and numbered asterisk above the gene. In the cases of *ZFP36L1* and *ZFP36L2*, our focus was on protein-coding regions, but some intronic sequence was determined as a byproduct of this effort. In the case of *ZFP36*, sequence was also determined for a regulatory region of the promoter (28); the complete intron, which also functions to regulate expression (29); and an AU-rich region in the 3′UTR, which is at least partly responsible for the instability of this message (W. S. Lai and P. J. Blackshear, unpublished data). Detailed information on the numbered polymorphisms identified in each of the three genes can be found in Tables I, II and III.

A. *ZFP36* Polymorphisms

The present resequencing effort confirmed the cDNA and protein GenBank reference sequences (RefSeq) NM_003407 and NP_003398 at every position. There were many differences between our results and the genomic sequence M92844; however, every one of the differences found in the resequencing effort agreed with the chromosome 19 sequence, AC0011500. These differences will not be described here. However, on the basis of these comparisons, we have concluded that the "wild-type" human genomic *ZFP36* sequence corresponding to part of the promoter, most of the two exons, and the single intron, is accurately represented by bp 11,119–14,350 of the chromosome 19 sequence AC0011500.

In the genomic DNA samples from the core 72 individuals, there were no polymorphisms in the *ZFP36* protein-coding region that resulted in amino acid changes. We therefore expanded the sequencing effort to include DNA from a separate collection of 20 anonymous North American subjects (Coriell Discovery Set) (Fig. 1 and Table I). The following discussion therefore applies to all 92 subjects or 184 alleles. In the key regulatory region of the promoter,

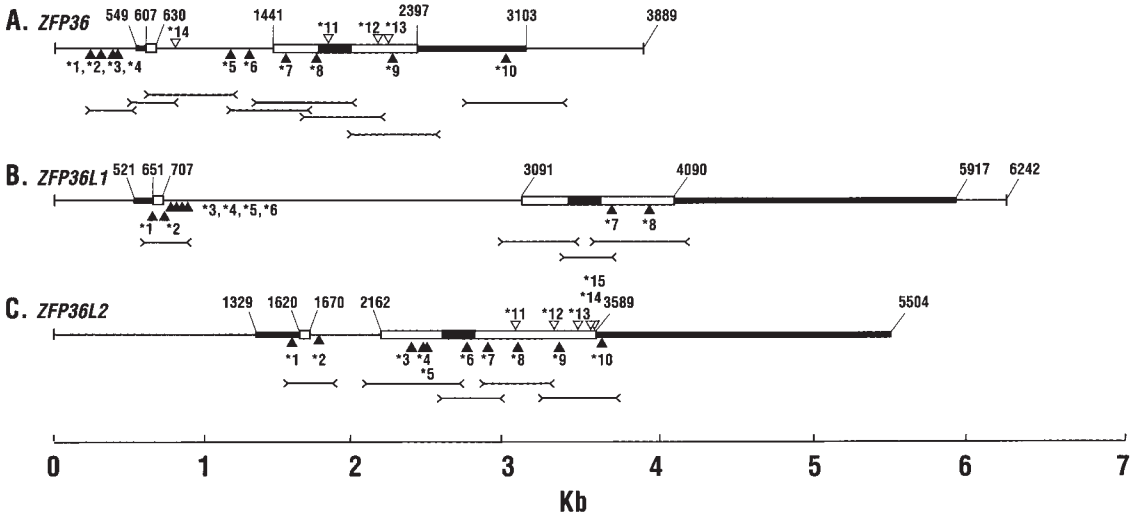


FIG. 1. Schematic representation of genes encoding human CCCH proteins. The three genes encoding the known human CCCH proteins with tandem zinc fingers are illustrated schematically. For all genes, the length in kb is indicated at the bottom of the figure. Genomic noncoding sequences, including introns, are indicated by a thin solid line; the 5′- and 3′UTRs by thick solid lines; and the protein-coding domains by thick open lines, in all three cases separated by the single intron. The gray regions within each of the second protein-coding regions represent the locations of the tandem zinc finger domains. The numbers at the top of each of gene show the location of specific bases within the primary sequence; the arrowheads labeled using numbers with asterisks below the indicated genes represent the sites of the numbered polymorphisms identified by sequence analysis, described in the text and in Tables I, II and III; the numbered arrowheads with asterisks above the genes represent the locations of putative polymorphisms based solely on EST analysis, but include those verified by direct sequencing of ESTs or by other means. The locations of the sequenced fragments are indicated below each gene by a solid line connecting two primer pairs, with > and < representing the binding sites of forward and reverse primers, respectively. The numbering systems used are: for *ZFP36* (A), the numbering system is from the genomic sequence in GenBank accession number M92844; for *ZFP36L1* (B), the numbering system begins with the genomic clone X79066, continues with 1347 bp of intronic sequence from AL132986.2, and then finishes with X79067, as described in detail in Table II; for *ZFP36L2* (C), the numbering system is from the genomic sequence U07802. See the text for further details.

four SNPs were identified (Fig. 1 and Table I). The first, polymorphism *ZFP36*1*, involved a single nucleotide polymorphism (SNP) at position 316 in GenBank accession number M92844. This was seen only once, in the additional 20 DNA samples from subjects of unknown ethnicity, and occurred at a 1.8% overall frequency. A second promoter SNP was seen at base 359 (*ZFP36*2*); this was very common, occurring in 47% of tested alleles, and was roughly equally distributed among the three major ethnic groups. It was found at a lower frequency in 15.4% of the alleles from subjects with bronchial hyperresponsiveness and in 34.1% of the alleles from patients with TRAPS-like syndromes. Two other promoter SNPs were identified, *ZFP36*3* and *ZFP36*4*, occurring at frequencies of 3.1% and 0.6%, respectively, in the 92 core samples. *ZFP36*3* was present in six alleles from Black individuals, and two alleles were from Americans of unknown ethnicity. Neither of these two uncommon SNPs was found in the bronchial sensitivity subjects or patients with TRAPS-like syndromes. None of the four promoter SNPs was in the three key enhancer elements previously identified as regulating serum-induced transcription of *ZFP36* (28).

Although the entire intron was sequenced in all of the 92 DNA samples, only two intron SNPs were identified (*ZFP36*5* and *P6*), and neither one was in the key regulatory regions that have previously been identified in the intron (29). Both were rare, occurring in only one allele of the 184 studied, in both cases in Black individuals. They were not found in the bronchial sensitivity subjects or the TRAPS-like patients.

Three SNPs occurred in the protein-coding region, in all cases within the second exon. The first, *ZFP36*7*, occurred at position 1525 in accession number M92844. This resulted in a change from a proline to a serine at position 37 in AAA61240. This was a rare polymorphism in this study sample, occurring at a frequency of 0.6% in individuals of unknown ethnicity, and was not seen in 127 ESTs spanning that region when searched on 11/15/02. Using a separate assay, we also evaluated genomic DNA from anonymous residents of Durham, NC, of varying ethnicities; of 844 alleles successfully evaluated, 834 contained C at this position, encoding a proline, and 10 contained T, encoding a serine, for an SNP frequency of 1.2% in this population. Only one instance of this SNP was found in 92 alleles successfully examined from the TRAPS-like patients (1.1%), and it was not seen in the bronchial sensitivity subjects. Examination of the same amino acid position in the protein sequences of other organisms showed that the wild-type proline was present in the bovine sequence (30) (accession number P53781), but that this was replaced by a leucine in both the mouse (10) (accession number P26651) and rat (P47973) proteins. Although TTP-like proteins have been sequenced from both *X. laevis* (2) (accession number AAD24207) and *D. rerio* (W. S. Lai and P. J. Blackshear, unpublished data), the protein sequences at this

position in these organisms were too unrelated to make similar phylogenetic comparisons.

Two other SNPs were identified in the protein-coding region, but neither resulted in an amino acid change. *ZFP36*8* resulted in no change to arginine 103; this SNP occurred roughly equally in the three major ethnic groups tested at 6.2% of alleles tested; it was not present in the samples from the bronchial sensitivity subjects or the TRAPS-like syndromes. It was seen in 2/127 (1.6%) relevant ESTs searched on 11/15/02. *ZFP36*9* occurred in 4.2% of alleles tested, and was seen in Black and Caucasian but not in Asian samples. This failed to change serine 273, and was seen in 1 of 69 ESTs (1.4%) that spanned this region when searched on 11/15/02. It was found in 3/76 (4%) of alleles successfully examined in the subjects with bronchial sensitivity, and in 1/92 (1.1%) of alleles from the patients with the TRAPS-like syndromes.

A final polymorphism was identified within the 3′UTR. Polymorphism *ZFP36*10* comprised a deletion of two T residues beginning at position 2980 in accession number M92844. This polymorphism was found in 7.6% of the alleles, in 64 of 249 (26%) ESTs in GenBank that spanned this region when searched on 11/15/02, and was roughly equally distributed among the three ethnic groups. It was also found at a frequency of 14.5% in the subjects with bronchial sensitivity, and 17.4% in the patients with TRAPS-like syndromes.

Three additional putative SNPs were identified within the protein-coding region of the second exon of *ZFP36* by the EST search. In the case of *ZFP36*11*, there was a G to T substitution at position 1807 in the genomic sequence, which resulted in a change of glycine at position 137 to a cysteine. This was found in 5/118 (4.2%) of ESTs examined on 11/15/02. This nonconservative change is between the two zinc fingers in the mRNA tandem zinc finger-binding domain of the TTP protein; however, it did not significantly affect TTP binding to an RNA probe in a cell-free gel shift assay (W. S. Lai and P. J. Blackshear, unpublished data). Two of the polymorphic ESTs (GenBank accession numbers AV661270 and AV661568) were from the same library prepared from noncancerous liver tissue from a Chinese subject, and one was from unknown tissue from a subject of unknown ethnicity (AW952247.1). Comparison of this amino acid position with TTP orthologs from other animal species revealed that glycine in this position was conserved in TTP from mouse, rat, and cow, but not in *X. laevis*.

Two other putative SNPs in the protein-coding region were identified by the EST search, but neither affected the encoded amino acid. *ZFP36*12* was a C to T substitution at position 2112, and occurred in 4/97 ESTs (4.1%), two of which were from the same human placenta library. *ZFP36*13* was a C to A substitution at position 2184, and occurred in 10/94 (11%) ESTs at this position.

A final SNP was identified in the resequencing of samples from the bronchial sensitivity subjects or those with TRAPS-like syndromes. This SNP,

*ZFP36*14*, was found in the intron at position 783 in the genomic sequence; it was found in only 1/162 alleles (0.6%) successfully examined. It was not in one of the known transcriptional regulatory elements in this intron (29).

We also analyzed the putative SNPs identified by GenBank and listed as "variants" in accession number NM_003407. Four variants have been identified and placed into dbSNP. In three of the four cases, these putative SNPs did not meet our criteria for inclusion when EST searches were conducted on 11/15/02: dbSNP # 2229272 was not seen in 121 ESTs, # 1803662 was seen in 2/250 ESTs, and #14869 was seen in 1/229 ESTs. However, # 1042905 was seen in 17/232 (7.3%) of ESTs searched on 11/15/02, and it therefore seems reasonable to include it as *ZFP36*15*. This SNP was an A to C change within the extreme 3′ end of the mRNA (b 1692 of NM_003407.1, b 3049 of M92844), and was within an AU-rich region that could be involved in the well-known instability of TTP mRNA (5). All but four of the variant transcripts at this site were from tumor specimens.

B. *ZFP36L1* Polymorphisms

We compared the current consensus sequence derived from the resequencing effort with data available in GenBank. The current consensus sequence that included the first exon corresponded exactly to bp 558–899 of X79066. The current consensus sequence differed by one nucleotide within the intron portion of *ZFP36L1*, in that G845 in X79066 was changed to C in the consensus. G845 was present in X79066.1 and XM_007458.1, but C was present in AL132986.4, Z64730, and Z61811.1. That this is a true intronic SNP is suggested by the facts that 95.8% of the 144 alleles resequenced here contained the G residue, and 4.2% contained the C residue (all Asian subjects). This is represented as SNP *ZFP36L1*6* in Table I and Fig. 1.

The consensus sequence that included the second exon spanned from bp 648 to 1844 of X79067. There were two differences between the current consensus sequence and X79067, but these have now been corrected in the RefSeq NM_004926.2 and are reflected in the current RefSeq protein sequence NP_004917.2.

A total of eight polymorphisms was found in *ZFP36L1* (Fig. 1 and Table II). A single SNP was identified in the 5′UTR (*ZFP36L1*1*); it was present in 12.5% of the alleles. Fifty-six percent of the polymorphic alleles were in Black subjects, and 44% were in Caucasians. Six of 60 (10%) of ESTs in GenBank on 8/15/02 also contained this SNP; this was also noted in dbSNP as # 1051533.

Within the protein-coding sequence, a potentially important double nucleotide mutation (*ZFP36L1*2*) was found in a single allele at the junction of the first exon with the single intron. Reexamination of the original sequence traces from this sample confirmed high-quality reads spanning this mutation in both directions (Fig. 2A). This AG to GC mutation would result in

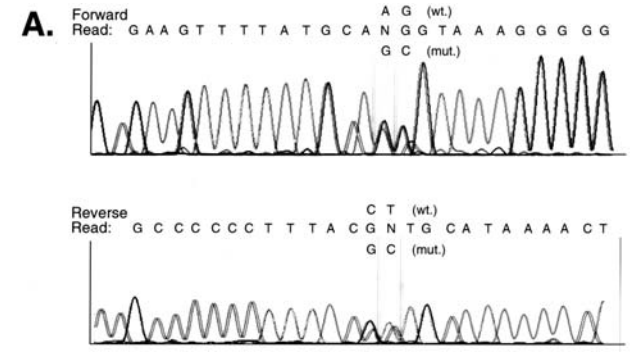
a potentially significant change of a lysine to a serine at position 19 in the protein (accession number NP_004917.2). More significantly, this change would result in the elimination of the 5′-donor portion of the splice site for the single intron in this gene. This mutation was found in a single allele, from an adult female Aka Pygmy subject from the Central African Republic (Coriell cell line AM10473A; subject X19). It was not found in 65 ESTs from GenBank that included this region on 11/15/02, although one EST (BF873589) contained unspliced 5′-intronic sequence. This mutation was not encountered in DNA from seven other Pygmies from the Central African Republic and Zaire that were in the original 72 samples.

This mutated splice site occurs at position 19 in the protein sequence, well before the tandem zinc finger domain that is the hallmark of this family of proteins. Theoretical translation of the mutated allele (Fig. 2B) indicates that after the lysine to serine conversion another 37 amino acids would be translated from intron sequence before a stop codon is reached. We suspect that the resulting hypothetical peptide, which would be 56 amino acids in length with an M_r of 6014 and a pI of 8.07, would have little biological significance, but this remains to be determined.

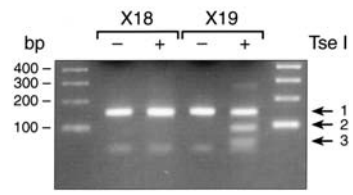
Mutation of both nucleotides in one allele would introduce two novel restriction sites into the genomic DNA, for the enzymes Fnu4HI and TseI, at position 704 of X79066. To confirm that both mutant bases were in the same allele and to devise an alternative assay for genotyping this dinucleotide mutation, we amplified a short region of DNA surrounding the mutation with appropriate PCR primers, and then cut the resulting 145 bp product with TseI. As shown in Fig. 2C, one allele from this DNA sample (X19) was cut into two fragments of approximately 50 and 95 bp, whereas the other allele was unaffected. Simultaneous digestion of a DNA sample from a subject of the same ethnic group did not reveal the presence of the mutation (Fig. 2C; X18). Thus, the presence of both mutations in the same allele was confirmed.

To determine whether this mutation would, as predicted, result in a failure to splice the mutant pre-mRNA, we subjected total lymphoblast RNA from subject X19 to northern-blotting analysis with a rat ZFP36L1 cDNA probe. As shown in Fig. 2D, lymphoblast RNA from a control subject from the same ethnic group (X15) produced the expected 3.9-kb band of the mature human ZFP36L1 mRNA. The amount of ZFP36L1 mRNA from the affected subject (X19) was decreased to 56% of that from the control cells, after PhosphorImager analysis of the northern blot and correction for GAPDH mRNA levels (Fig. 2D).

Because the two genomic sequences for *ZFP36L1* in GenBank did not meet in the middle of the intron, we searched GenBank for complete human genomic sequences in this region. This gene has been mapped previously to chromosome 14q22–24 (*31*). There was essentially a perfect match with


A.
Forward
Read: G A A G T T T T A T G C A N G G T A A A G G G G G G
A G (wt.)
G C (mut.)
Reverse
Read: G C C C C C C T T T A C G N T G C A T A A A A C T
C T (wt.)
G C (mut.)


B.
Normal allele:
ATGACCACCACCCTCGTGTCTGCCACCATCTTCGACTTGAGCGAAGTTTTATGCAAGgtaaag
M T T T L V S A T I F D L S E V L C K ------
Mutant allele:
ATGACCACCACCCTCGTGTCTGCCACCATCTTCGACTTGAGCGAAGTTTTATGCAGCgtaaag
M T T T L V S A T I F D L S E V L C S V K
ggggggcggagcatttgcttttcaaaaagtcttttcttttgccttaagaaagaaccccccaaaa
G G G A F A F Q K V F F F C L K K E P P K
gtttgggtcttggaaggcaatggggaaagtggtgcagcgagatga
V W V L E G N G E S G A A R *


C.
X18
X19
bp
– + – +
Tse I
400 –
300 –
200 –
100 –
1
2
3

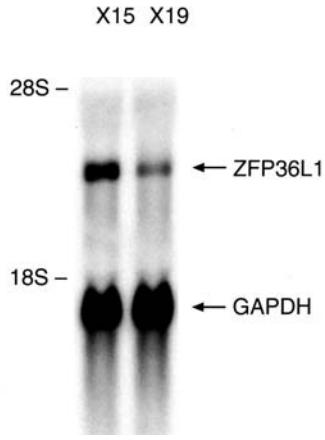

D.
X15 X19
28S –
ZFP36L1
18S –
GAPDH

FIG. 2. Analysis of a *ZFP36L1* nonsplicing allele in subject X19. In (A) are shown the sequence traces from the sequencing of genomic DNA from subject X19; the wild-type and mutant base changes are shown in both strands. In (B) is shown the conceptual translation of the normal and X19 mutant *ZFP36L1* alleles at the amino terminal end of the ZFP36L1 protein, along with the relevant DNA sequence extending into the first part of the intron. The normal DNA sequence begins with the ATG at position 651 of X79066. The corresponding DNA and amino acid sequences for the mutant allele from subject X19 are also shown; the DNA sequence was translated until a stop codon was encountered within the intron sequence. The normal and mutant dinucleotides are indicated in bold type; the intron sequence is indicated by lower case letters, and by a series of dashes at the end of the amino acid sequence; the hypothetical translated mutant "protein" sequence is indicated by the underlined amino acids; and the stop codon in the intron that terminates this hypothetical open-reading frame is indicated by an asterisk. In (C), DNA (approximately 1 μg) comprising a 145-bp PCR product from subject X19 was digested using the restriction endonuclease TseI, and was compared with the same amount of DNA from a control subject (X18) from the same ethnic group, as indicated. The arrow labeled 1 indicates the undigested DNA from both the mutant and nonmutant alleles; the arrows labeled 2 and 3 indicate the new restriction fragments of 95 and 50 bp, respectively, derived from the TseI digestion of the mutant but not the wild-type allele. In (D) is shown a northern analysis of spliced and unspliced *ZFP36L1* mRNA from subject X19. Total cellular RNA (15 μg) from lymphoblasts derived from subject X19 and a control subject from the same ethnic group (X15) was subjected to northern analysis using a ZFP36L1 cDNA probe. The arrow labeled ZFP36L1 indicates the correctly spliced ZFP36L1 mRNA with a size of approximately 3.9 kb. The arrow labeled "GAPDH" indicates the GAPDH mRNA, which served as an internal control. After correction for the GAPDH mRNA values, the amount of correctly spliced ZFP36L1 mRNA from subject X19 was 56% of the amount from control subject X15, as determined by PhosphorImager analysis. See the text for further details.

a BAC clone from chromosome 14 (GenBank accession number AL133313.2). Position 1 in X79066 is the equivalent to position 78926 in the BAC sequence; position 3922 in the 3′-genomic clone X79067 is the equivalent of 85172 in the BAC sequence. This indicates that the single intron in *ZFP36L1* is 2385 bp in size, as indicated in Fig. 1. It also indicates that a putative unspliced pre-mRNA should be about 6–6.5 kb. However, this putative unspliced pre-mRNA was not detected on the northern blot of the affected subject's RNA (X19), even after a much longer exposure of the autoradiograph depicted in Fig. 2D.

The RFLP assay described earlier was used to analyze DNA samples from 58 other Aka individuals. One other mutant allele was identified from a 54-year-old male Aka subject, also from the Central African Republic. The relationship between the two anonymous subjects with this mutation is not known. This dinucleotide variant was not found in the genotyping of DNA samples from 430 anonymous residents of Durham, NC.

Although we focused on the protein-coding regions of these genes, the sequencing strategy for *ZFP36L1* yielded several polymorphisms in the 5′ region of the single intron, and this small region was quite polymorphic. Intron SNPs were identified at positions 729, 772, 804, and 845 (polymorphisms *ZFP36L1*3–6*, respectively). *ZFP36L1*3* was present in only a single allele (Table II). *ZFP36L1*4* was found in 27 alleles, roughly distributed among the three racial groups. *ZFP36L1*5* was found in nine alleles, in every case in Caucasian subjects. *ZFP36L1*6* was found in six alleles, in every case in Asian subjects; of these, two were from the Taiwan Ami group, two were Cambodian, and one each was Beijing Chinese and Japanese.

Two other SNPs were identified in the protein-coding region of the second exon. In the first (*ZFP36L1*7*), no change in the glycine residue at position 218 in the protein would occur; this was found in a single Asian allele, and in 0/40 ESTs spanning that region on 8/15/02. *ZFP36L1*8* was also found in a single Asian allele, but not in 46 ESTs spanning that region on 11/15/02. In this case, the resulting predicted amino acid change (a proline at position 295 changed to a histidine) is nonconservative and possibly significant. This portion of the ZFP36L1 protein sequence was compared with those of the orthologous sequences from mouse (accession number P23950), rat (P17431), and *Xenopus* (AF061081). In all cases, the original proline in the human protein was present at the same position in the orthologs; it was in the middle of a highly conserved region of the proteins, with a core sequence of PSP*QD in human, mouse, and rat, and PSP*RD in *Xenopus*, with the asterisk indicating the proline changed to histidine in *ZFP36L1*8*. This amino acid change removes a potential proline-directed protein kinase phosphorylation site on the serine immediately in front of the changed proline; in the case of TTP, members of this family of protein kinases are thought to phosphorylate

the protein in intact cells (*30,32*). Whether this potential site is phosphorylated by one or more of these kinases *in vivo*, and whether this polymorphism causes any other changes in the property of the protein, are subjects for future study.

No other polymorphisms for *ZFP36L1* were identified by the EST analysis of the protein-coding regions. However, it was apparent that there were at least two highly polymorphic regions within the 3′UTR of the mRNA: the region immediately 5′ of b 2750 of NM_004926.2 contained at least five common variants of a series of A residues that was highly variable in length, and the region between 1730 and 1755 of the NM_004926.2 contained at least four common variants. These regions will require future resequencing to determine allele frequency and haplotypes of these common variants.

C. *ZFP36L2* Polymorphisms

Based on the consensus sequence for *ZFP36L2* generated by the current resequencing effort, the GenBank RefSeq NM_006887.3 now corresponds to what we believe to be the correct cDNA sequence. The corresponding protein sequence is NP_008818.3.

Polymorphisms in the *ZFP36L2* protein-coding region that were identified by the sequencing effort are indicated schematically in Fig. 1 and in more detail in Table III. A single polymorphism (*ZFP36L2*1*) was identified in the 5′UTR that comprised a 15 b deletion. This was found in a single allele, from an African-American subject, and in 0/11 EST from GenBank on 8/15/02. A single intron SNP was also identified (*ZFP36L2*2*) in one allele from an Asian subject. Seven SNPs were identified in the protein-coding domain of the second exon. The first (*ZFP36L2*3*) was found in a single allele and did not change glycine 93 of accession number AAA91778.1; it was not found in 20 relevant ESTs. A second (*ZFP36L2*4*) was found in 2.8% of alleles and in 1/27 ESTs (3.7%), and resulted in a semiconservative glycine to serine change at position 109. Examination of *ZFP36L2* sequences from other species in GenBank revealed that the glycine at position 93 was conserved in mouse (accession number C39590), and in the conceptual translations of ESTs for cow (AW465810), rat (BF522474), and pig (AW465810); in contrast, the *Xenopus* ortholog XC3H-3 contained a serine at that position (AAD24209). A third coding region SNP (*ZFP36L2*5*) occurred in two alleles and 0/28 ESTs, and resulted in a change from leucine to glutamine at position 112; the wild-type leucine residue was perfectly conserved in the five animal species listed earlier, using data from the same DNA sequences. The fourth (*ZFP36L2*6*) occurred in 6.25% of alleles and in 20/78 (26%) of ESTs; this resulted in no change to isoleucine 203. This has been identified by GenBank as dbSNP # 8098. The fifth (*ZFP36L2*7*) occurred in 48% of alleles and in 52/97 ESTs (54%), and resulted in no change to leucine 254. This SNP was found in all

three racial groups; it has been identified as dbSNP # 7933. The sixth (*ZFP36L2*8*) occurred in 2.1% of alleles and 0/70 ESTs, and resulted in no change in proline 315. The seventh (*ZFP36L2*9*) occurred in 1.4% of alleles and in 0/204 ESTs, and resulted in a conservative alanine to serine change at position 405. This SNP was in a region of the protein sequence that is relatively poorly conserved among animal species, so it was not possible to do a phylogenetic comparison at this site. Finally, a 3′UTR SNP (*ZFP36L2*10*) was identified in one allele but not in 106 ESTs; this was not in a canonical instability motif.

Five additional putative polymorphisms in the protein-coding region of *ZFP36L2* were uncovered by the EST analysis. The first, *ZFP36L2*11*, was a C to A substitution at position 3055 in the genomic sequence; this did not change proline 315, and was found in 3/69 (4.3%) of ESTs spanning this site. A C to T SNP at the same site (*ZFP36L2* 8*) had previously been identified in the resequencing effort, also with no amino acid change.

A trinucleotide repeat polymorphism was identified in 6/>200 ESTs on 7/4/01; this will be called *ZFP36L2*12*. This change increased a series of seven CAG repeats to nine, resulting in two additional glutamine residues at position 393 in the ZFP36L2 protein (NP_008818), for a total of nine consecutive glutamines. One of the ESTs with nine CAG repeats was resequenced in both directions, and the CAGCAG insert was confirmed. The five polymorphic ESTs came from two libraries. Three were from a mixed male germ cell tumor library (accession numbers AW589855.1, AI633753.1, and AI202919.1), and two were from a tonsilar germinal center B cell library (AA489889.1 and AA761736.1). A second complete cDNA and conceptually translated protein sequence for ZFP36L2 were deposited in GenBank (AAH05010), in which there were 10 CAG repeats at this site, corresponding to 10 glutamine residues. This clone was derived from a pancreatic adenocarcinoma. These expanded CAG repeat polymorphisms were not present in the 144 alleles that form the basis of this paper, all of which contained the wild-type seven consecutive CAGs. In addition, resequencing of this portion of *ZFP36L2* in DNA from 20 germ cell testicular tumors did not reveal any differences from wild type in this potential microsatellite.

Three other potential polymorphisms that would result in amino acid changes were identified by the EST analysis. *ZFP36L2*13* was a TC to CT change that would result in a change of leucine 450 to serine. Although this could be a simple sequencing error, it was found in four of the more than 200 ESTs spanning this region, two of which were from the same prostate library. *ZFP36L2*14* was a G to A substitution that would lead to a change from asparagine 476 to aspartate, and was seen in three of the more than 150 relevant ESTs, two of which were from the same germ cell tumor library. *ZFP36L2*15* was a C to G substitution that would change a serine at

position 492 to a cysteine, and occurred in six of more than 150 ESTs spanning this region.

Recently, the genomic *ZFP36L2* sequence was identified on a BAC contig from chromosome 2, accession number AC010883.5. Base 1 in U07802, the *ZFP36L2* genomic clone, aligned with base 152080 in AC010883.5; base 5503 of U07802 aligned with base 146557 in AC010883.5. Although the sequence identity between these two clones was 97%, with 79 gaps, it seems likely that this position in chromosome 2 is the locus for *ZFP36L2*. The BAC sequence AC010883.5 contained the normal seven CAG repeats at the polyglutamine polymorphic region (*ZFP36L2*12*).

IV. Discussion

These studies provide initial information on the occurrence and type of polymorphisms among the three known human genes encoding proteins of the TTP CCCH tandem zinc finger type. These analyses focused on the protein-coding regions of all three genes. Out of 13 polymorphisms identified in these protein-coding regions by the resequencing effort, seven (54%) resulted in no change in amino acid, and 6/13 (46%) resulted in amino acid changes ranging from conservative to nonconservative. Notably, there were no amino acid changes within the 64-amino acid tandem zinc finger domains of these proteins, perhaps reflecting the critical importance of many residues in this domain to mRNA binding (*1*). An additional triplet repeat polymorphism was identified in *ZFP36L2* by EST searches (*ZFP36L2*11*); this resulted in an increase from seven to nine consecutive glutamine residues. This polymorphism was found in ESTs from two different libraries, from mixed male germ cell tumors, and from nonmalignant tonsilar B cell, and an even longer repeat of 10 glutamines was found in a clone derived from a pancreatic adenocarcinoma (AAH05010). Although this type of triplet repeat polymorphism was not found in the 144 alleles resequenced for this paper, it seems possible that it represents an example of microsatellite instability in tumor tissue. Matched samples of tumor and nontumor DNA from the same individuals are currently being evaluated to determine if this is the case.

For most of these polymorphisms, further study will be required to determine if they cause any functional change in the resulting mRNAs and proteins. An exception is in the case of the dinucleotide splice site mutation found in one *ZFP36L1* allele from the Aka Pygmy subject X19, and subsequently in another individual from the same ethnic group, but not in 430 anonymous subjects from various ethnic groups in Durham, NC, or in ESTs currently in GenBank. This mutation resulted in a decrease in steady-state *ZFP36L1* mRNA levels to about 50% of its normal level of expression in cells

derived from subject X19, suggesting normal expression from the unaffected allele; the absence of a detectable unspliced form of the mRNA suggests that the mutant allele is not expressed. The function of the normal *ZFP36L1* protein in physiology is unknown. However, when it is expressed in human embryonic kidney 293 cells, the ZFP36L1 protein can, like TTP and ZFP36L2, promote the instability of ARE-containing mRNAs following direct binding to the class II ARE (*1*). In addition, its expression is increased in a form of acute myelogenous leukemia, where its overexpression is associated with enhanced myeloid cell proliferation in response to granulocyte colony-stimulating factor (*33*). Recent experiments with the mouse gene knockout indicate that total deficiency of *Zfp36L1* is lethal in early gestation, suggesting that even the hemizygous state might be associated with a disease or trait (D. J. Stumpo, R. S. Phillips, Y. Mishina and P. J. Blackshear, unpublished data). We are currently attempting to identify human kindreds with this splicing mutation, with the goal of examining them for coinheritance of a disease or trait with the mutant allele.

Another *ZFP36L1* polymorphism with potential functional consequences is *ZFP36L1*8*, which occurred in a single allele from an Indo-Pakistani subject (sample W08, Coriell cell line GM11213) and in none of the 46 ESTs from GenBank on 11/15/02 that spanned this region. This SNP would result in the removal of a potential phosphorylation site for proline-directed protein kinases, such as the mitogen-activated protein (MAP) kinases. The normal sequence in this region is sppsPqdsls, with the proline indicated by the upper case P mutated to a histidine in the polymorphic sequence. It is not known at present if this is a physiologically relevant phosphorylation site; in the case of TTP, MAP kinase phosphorylation of the protein is known to occur in intact cells (*30*), and recent evidence from our laboratory suggests that phosphorylation of TTP by the MAP kinase p38 has functional effects on mRNA binding and destabilization (*9*). The serine–proline duo at positions 294, 295 in the human sequence is conserved in all orthologues of *ZFP36L1* examined, including mouse (accession number B39590), rat (P17431), *X. laevis* (AAD24208.1), and a presumed ortholog encoded by a chicken EST (AJ399312.1, reading frame +1). Ongoing experiments in our laboratory will test the hypothesis that this site is a phosphorylation site for one of these kinases under normal circumstances, and that this mutation affects not only phosphorylation of the protein but also one or more functional properties.

Another potential site for interesting polymorphisms is the ARE within the 3′UTR of the TTP mRNA, which is responsible in part for the lability of this mRNA. Direct resequencing of this region yielded only two potential polymorphisms, one TT deletion at position 2980 in the genomic sequence, and one putative T to A change at position 3072. The former dinucleotide deletion polymorphism (*ZFP36*10*) appears to be a *bona fide*, common human

variant: it appeared in 7.6% of the alleles directly sequenced in this study, was represented in samples from all three major racial groups, and was found in 64/249 (26%) of ESTs in GenBank on 11/15/02. Although it is not within a canonical AUUUA instability motif, it is possible that it contributes in some way to TTP mRNA stability. The latter potential polymorphism identified in this region of the TTP mRNA is problematical. Specifically, it occurred in a region that could not be reliably sequenced using the present methods, and was not found in 144 ESTs in GenBank that spanned this region. At this point, its existence must be considered doubtful. A third polymorphism within the 3′UTR that was identified only by EST analysis is likely to be real; it (ZFP36*15) was found in 17/232 (7.3%) of EST examined on 11/15/02, and was within an AU-rich region of the mRNA that is likely to be involved in mRNA stability.

Although these polymorphisms in the TTP mRNA currently are of unclear significance to the function or steady-state levels of the mRNA or protein, they may be of interest in linkage studies. TTP has been localized to chromosome 19q13.1 (*3*), and its genomic sequence is contained within a BAC contig localized to chromosome 19 (accession number AC011500.6), where genomic sequence 1–3880 from accession number M92844 is equivalent to bp 45,571–41,684 of accession number AC011500.6. Based on the results to date in with the TTP knockout mice, complete TTP deficiency in man would be expected to result in a severe autoimmune and inflammatory condition with myeloid hyperplasia that would probably not be compatible with life when untreated; the hemizygous state might result in a less severe syndrome of later onset. However, as we demonstrated previously, the phenotype of the TTP knockout mice can be virtually normalized using an anti-TNFα monoclonal antibody (*12*) or interbreeding with mice deficient in both TNFα receptors (*14*). Similar treatments, using a "humanized" anti-TNFα antibody or a chimeric soluble TNFα receptor, have proven to be effective treatments for human rheumatoid arthritis and Crohn's disease, and might well be used to delay or prevent the consequences of complete or partial TTP deficiency in man. These considerations suggest that linkage studies with TTP polymorphisms might be profitable in human autoimmune, autoinflammatory, or myeloproliferative disorders, but our limited knowledge of normal TTP physiology precludes identification of more likely patient groups.

As an initial exploratory step, we resequenced the protein coding and regulatory regions of the gene encoding TTP, *ZFP36*, from two groups of subjects: one group of otherwise normal subjects that exhibited bronchial hypo- or hypersensitivity in response to inhaled endotoxin (*20*), and a second group of subjects with a chronic inflammatory syndrome resembling that caused by activating mutations in one of the TNFα receptors (*21*). Only a single new intronic SNP was identified by this expanded resequencing effort;

to date, none of the other polymorphisms in *ZFP36* exhibited major differences in frequencies when the TRAPS-like subjects were compared with the other subjects, or data from the bronchial hypersensitivity group were compared with the hyposensitivity group.

These studies have begun to describe the normal sequence variability within members of this small human gene family. Future studies will focus on identifying which of the polymorphisms described here are associated with changes in the biosynthesis, turnover, or function of the encoded CCCH proteins. In addition, we hope to link each of these polymorphisms to specific haplotypes, identify kindreds with each haplotype, and then determine the possible association of these haplotypes with heritable traits.

Acknowledgments

Supported in part by Interagency Agreement Yi-ES-8054-05, and by a Cooperative Research and Development Agreement with AstraZeneca Ltd. We are grateful to Dr. James Selkirk and Tina Xi for advice, to Mary Watson for the cultured lymphoblasts, and to Drs. Jack Taylor and Stephanie London for helpful comments on the manuscript.

References

1. W. S. Lai, E. Carballo, J. M. Thorn, E. A. Kennington and P. J. Blackshear, Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA, *J. Biol. Chem.* **275,** 17827–17837 (2000).
2. J. De, W. S. Lai, J. M. Thorn, S. M. Goldsworthy, X. Liu, T. K. Blackwell and P. J. Blackshear, Identification of four CCCH zinc finger proteins in Xenopus, including a novel vertebrate protein with four zinc fingers and severely restricted expression, *Gene* **228,** 3454 (1999).
3. G. A. Taylor, W. S. Lai, R. J. Oakley, M. F. Seldin, T. B. Shows, R. L. Eddy. Jr. and P. J. Blackshear, The human TTP protein: sequence, alignment with related proteins and chromosomal localization of the mouse and human genes, *Nucleic Acids Res.* **19,** 3454 (1991).
4. R. N. DuBois, M. W. McLane, K. Ryder, L. F. Lau and D. Nathans, A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence, *J. Biol. Chem.* **265,** 19185–19191 (1990).
5. W. S. Lai, D. J. Stumpo and P. J. Blackshear, Rapid insulin-stimulated accumulation of an mRNA encoding a proline-rich protein, *J. Biol. Chem.* **265,** 16556–16563 (1990).
6. Q. Ma and H. R. Herschman, A corrected sequence for the predicted protein from the mitogen-inducible TIS11 primary response gene, *Oncogene* **6,** 1277–1278 (1991).
7. B. C. Varnum, R. W. Lim, V. P. Sukhatme and H. R. Herschman, Nucleotide sequence of a cDNA encoding TIS11, a message induced in Swiss 3T3 cells by the tumor promoter tetradecanoyl phorbol acetate, *Oncogene* **4,** 119–120 (1989).
8. E. Carballo, W. S. Lai and P. J. Blackshear, Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin, *Science* **281,** 1001–1005 (1998).

9. E. Carballo, W. S. Lai and P. J. Blackshear, Evidence that tristetraprolin is a physiological regulator of granulocyte-macrophage colony-stimulating factor messenger RNA deadenylation and stability, *Blood* **95,** 1891–1899 (2000).
10. W. S. Lai, E. Carballo, J. R. Strum, E. A. Kennington, R. S. Phillips and P. J. Blackshear, Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA, *Mol. Cell. Biol.* **19,** 4311–4323 (1999).
11. W. S. Lai and P. J. Blackshear, Interactions of CCCH zinc finger proteins with mRNA: tristetraprolin-mediated AU-rich element-dependent mRNA degradation can occur in the absence of a poly(A) tail, *J. Biol. Chem.* **276,** 23144–23154 (2001).
12. G. A. Taylor, E. Carballo, D. M. Lee, W. S. Lai, M. S. Thompson, D. D. Patel, D. I. Schenkman, G. S. Gilkeson, H. E. Broxmeyer, B. F. Haynes and P. J. Blackshear, A pathogenetic role for TNF alpha in the syndrome of cachexia, arthritis, and autoimmunity resulting from tristetraprolin (TTP) deficiency, *Immunity* **4,** 445–454 (1996).
13. E. Carballo, G. S. Gilkeson and P. J. Blackshear, Bone marrow transplantation reproduces the tristetraprolin-deficiency syndrome in recombination activating gene-2 (−/−) mice. Evidence that monocyte/macrophage progenitors may be responsible for TNFalpha overproduction, *J. Clin. Invest.* **100,** 986–995 (1997).
14. E. Carballo and P. J. Blackshear, Roles of tumor necrosis factor alpha receptor subtypes in the pathogenesis of the tristetraprolin-deficiency syndrome. *Blood* **98**, 2389–2395 (2001).
15. Z. Q. Ning, J. D. Norton, J. Li and J. J. Murphy, Distinct mechanisms for rescue from apoptosis in Ramos human B cells by signaling through CD40 and interleukin-4 receptor: role for inhibition of an early response gene, Berg36, *Eur. J. Immunol.* **26,** 2356–2363 (1996).
16. X. F. Nie, K. N. Maclean, V. Kumar, I. A. McKay and S. A. Bustin, ERF-2, the human homologue of the murine Tis11d early response gene, *Gene* **152,** 285–286 (1995).
17. E. Fritsche, G. S. Pittman and D. A. Bell, Localization, sequence analysis, and ethnic distribution of a 96-bp insertion in the promoter of the human CYP2E1 gene [In Process Citation], *Mutat. Res.* **432,** 1–5 (2000).
18. F. P. Guengerich, The environmental genome project: functional analysis of polymorphisms, *EHP* **106,** 365–368 (1998).
19. R. R. Sharp and J. C. Barrett, The environmental genome project and bioethics, *Kennedy Inst. Ethics J.* **9,** 175–188 (1999).
20. J. G. Moreland, R. M. Fuhrman, C. L. Wohlford-Lenane, T. J. Quinn, E. Benda, J. A. Pruessner and D. A. Schwartz, TNF-alpha and IL-1 beta are not essential to the inflammatory response in LPS-induced airway disease, *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280,** L173–L180 (2001).
21. I. Aksentijevich, J. Galon, M. Soares, E. Mansfield, K. Hull, H. H. Oh, R. Goldbach-Mansky, J. Dean, B. Athreya, A. J. Reginato, M. Henrickson, B. Pons-Estel, J. J. O'shea and D. L. Kastner, The tumor-necrosis-factor receptor-associated periodic syndrome: new mutations in TNFRSF1A, ancestral origins, genotype-phenotype studies, and evidence for further genetic heterogeneity of periodic fevers, *Am. J. Hum. Genet.* **69,** 301–314 (2001).
22. B. L. King, H. Q. Peng, P. Goss, S. Huan, D. Bronson, B. M. Kacinski and D. Hogg, Repeat expansion detection analysis of (CAG)n tracts in tumor cell lines, testicular tumors, and testicular cancer families, *Cancer Res.* **57,** 209–214 (1997).
23. D. A. Nickerson, S. L. Taylor, K. M. Weiss, A. G. Clark, R. G. Hutchinson, J. Stengard, V. Salomaa, E. Vartiainen, E. Boerwinkle and C. F. Sing, DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene [see comments], *Nat. Genet.* **19,** 233–240 (1998).
24. M. J. Rieder, S. L. Taylor, V. O. Tobe and D. A. Nickerson, Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome, *Nucleic Acids Res.* **26,** 967–973 (1998).

25. M. R. Shen, I. M. Jones and H. Mohrenweiser, Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans, *Cancer Res.* **58,** 604–608 (1998).
26. D. A. Nickerson, V. O. Tobe and S. L. Taylor, PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing, *Nucleic Acids Res.* **25,** 2745–2751 (1997).
27. M. Kokoris, K. Dix, K. Moynihan, J. Mathis, B. Erwin, P. Grass, B. Hines and A. Duesterhoeft, High-throughput SNP genotyping with the Masscode system, *Mol. Diagn.* **5,** 329–340 (2000).
28. W. S. Lai, M. J. Thompson, G. A. Taylor, Y. Liu and P. J. Blackshear, Promoter analysis of Zfp-36, the mitogen-inducible gene encoding the zinc finger protein tristetraprolin, *J. Biol. Chem.* **270,** 25266–25272 (1995).
29. W. S. Lai, M. J. Thompson and P. J. Blackshear, Characteristics of the intron involvement in the mitogen-induced expression of Zfp-36, *J. Biol. Chem.* **273,** 506–517 (1998).
30. G. A. Taylor, M. J. Thompson, W. S. Lai and P. J. Blackshear, Phosphorylation of tristetraprolin (TTP), a potential zinc finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase in vitro. *J. Biol. Chem.* **270**, 13341–13347 (1995).
31. K. N. Mclean, C. G. See, I. A. McKay and S. A. Bustin, The human immediate early gene BRF1 maps to chromosome 14q22-q24. *Genomics* **30**, 89–90 (1995).
32. E. Carballo, H. Cao, W. S. Lai, E. A. Kennington, D. Campbell and P. J. Blackshear, Decreased sensitivity of tristetraprolin-deficient cells to p38 inhibitors suggests the involvement of tristetraprolin in the p38 signaling pathway, *J. Biol. Chem.* **276,** 42580–42587 (2001).
33. H. Shimada, H. Ichikawa, S. Nakamura, R. Katsu, M. Iwasa, I. Kitabayashi and M. Ohki, Analysis of genes under the downstream control of the t(8;21) fusion protein AML1-MTG8: overexpression of the TIS11b (ERF-1, cMG1) gene induces myeloid cell proliferation in response to G-CSF, *Blood* **96,** 655–663 (2000).